UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Toward red emitting nucleosides: design, synthesis, photophysical properties, oligonucleotide incorporation and applications

Permalink

https://escholarship.org/uc/item/8tq8h27v

Author

Hopkins, Patrycja

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Toward red emitting nucleosides: design, synthesis, photophysical properties,
oligonucleotide incorporation and applications

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Patrycja Anna Hopkins

Committee in Charge:

      Professor Yitzhak Tor, Chair
      Professor Steven Dowdy
      Professor Thomas Hermann
      Professor Douglas Magde
      Professor Emmanuel Theodorakis

2016

The dissertation of Patrycja Anna Hopkins is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

<div align="right">Chair</div>

University of California, San Diego

2016

DEDICATION

to my parents Alicja and Grzegorz Janaszkiewicz
and my husband Alec Hopkins

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABELS

LIST OF SPECTRA

ACKNOWLEDGMENTS

I would like to thank my parents Alicja and Grzegorz Janaszkiewicz for their unconditional love and constant support thought out the journey of graduate school. I would also want to thank my husband Alec Hopkins for his continues words of encouragement and his endless patience.

Also I would like to acknowledge my laboratory partners: Richard Fair, Kaivin Hadidi, Kristina Hamill, Yao Li, Lisa McCoy, Andro Rios, Alexander Rovira and Ryan Weiss as well as post-dorctoral fellows: Dr. Andrea Fin, Dr. Noam Freeman, Dr. Dongwon Shin, Dr. Sinkeldam and Dr. Ezequiel Wexselblatt. Especially, I would like to thank my partner in crime Lisa McCoy for many scientific conversations, her constant support, advice and encouragement. I would also like to thank Lisa Adamiak, who has started this journey with me and who always has had a kind word for me.

I would like to acknowledge Professor Yitzhak Tor for providing the environment, excellent scientific advice and support. In addition, I thank him for guiding me throughout this process, for teaching me how to questions results and overall for facilitating a path in becoming a scientist. He has helped me tremendously with my research for which I am extremely grateful.

I would to thank my committee for their stimulation classes and support during my graduate work. I would also like to acknowledge Professor Arnold Rheingold and Dr. Curtis Moore (UCSD Crystallography Facility) for their crystallography work, Dr. Yongxuan Su (UCSD MS Facility) for the mass

spectroscopy measurement and Dr. Anthony Mrse (UCSD NMR Facility) for his availability and assistance with the NMR instruments.

In regards to the material in this dissertation, please note the following:

Portion of chapter 1 (subchapter 1.6) has been previously published in the following article: "Emissive 5-Substituted Uridine Analogues" *In Modified Nucleic Acids*, Springer: 2016; 31, 1-26 by Fin, A.; Rovira, A. R.; Hopkins, P. A.; Tor. Y.. The dissertation author is one of the primary authors of this work.

Chapter 2 is in full a reprint of: Sinkeldam, R. W.; Hopkins, P. A; Tor, Y. "Modified 6-Aza Uridines: Highly Emissive pH-sensitive Fluorescent Nucleosides" *ChemPhysChem* **2012**, 13, 3350. The dissertation author was one of the primary investigators and authors of this paper.

Chapter 3 is in full a reprint of: Hopkins, P. A.; Sinkeldam, R. W.; Tor, Y. "Visibly Emissive and Responsive Extended 6–Aza-Uridines" *Org. Lett.*, **2014**, *16*, 5290. The dissertation author is the main author and researcher for this work.

Chapter 4 ESIPT, the dissertation author is the main author and researcher for this work.

Chapter 5 is a full reprint from: Hopkins, P. A.; McCoy, L.; Tor, Y. "Enzymatic Incorporation and Utilization of an Emissive 6-Aza Uridine." *submitted*. The dissertation author is the primary investigator and author of this article.

Chapter 6 is a full reprint from: Hopkins, P. A.; McCoy, L.; Fin, F.; Tor, Y. " Enzymatic synthesis of partially and fully modified RNAs." *-in preparation*. The dissertation author is the primary investigator and author of this article.

VITA

Education:

University of California, San Diego - La Jolla, CA
                    Ph.D. Chemistry - December 2016

University of Lodz - Lodz, Poland
                    M. S. Chemistry - June 2008

Honors and Awards:

GAANN (Graduate Assistance in Areas of National Needs) Fellowship 2015/2016

GAANN (Graduate Assistance in Areas of National Needs) Fellowship 2014/2015

Teaching Assistant Excellence Award 2013

Fall Teaching Assistant Excellence Award - September 2013

PUBLICATIONS

1. Hopkins, P. A.; McCoy, L.; Fin, A.; Tor, Y.: " Enzymatic Synthesis of Partially and Fully Modified RNAs." – *in preparation*.

2. Hopkins, P. A.; McCoy, L.; Tor, Y. "Enzymatic Incorporation and Utilization of an Emissive 6-Aza Uridine." – *submitted*.

3. Fin, A.; Rovira, A. R.; Hopkins, P. A.; Tor. Y.: "Emissive 5-Substituted Uridine Analogues" *In Modified Nucleic Acids*, Springer: 2016; 31, 1-26.

4. Hopkins, P. A.; Sinkeldam, R. W.; Tor, Y. "Visibly Emissive and Responsive Extended 6‑Aza-Uridines" *Org. Lett.*, 2014, *16*, 5290.

5. Sinkeldam, R. W.; Hopkins, P. A.; Tor, Y. "Modified 6-Aza Uridines: Highly Emissive pH-Sensitive Fluorescent Nucleosides" *ChemPhysChem* 2012, *13*, 3350.

ABSTRACT OF THE DISSERTATION

Toward red emitting nucleosides: design, synthesis, photophysical properties,

oligonucleotide incorporation and applications

by

Patrycja Anna Hopkins

Doctor of Philosophy in Chemistry

University of California, San Diego, 2016

Professor Yitzhak Tor, Chair

As our understanding of complex cellular processes advances, nucleic acids, the biomolecules of life, consistently appear as key players. The ability to explore the structure, dynamics, and recognition of nucleic acids has been accessible due to many techniques. Specifically, the sensitivity of fluorescent spectroscopy has made it one of the most useful tools for the detailed study of macromolecules. The development of synthetically modified nucleosides has been driven by the lack of naturally occurring nucleosides with favorable photophysical features that allow for probing of nucleic acid-containing systems. The design of

the ideal probe should ensure high structural resemblance to the native nucleosides, minimizing the perturbation of the biomolecular architecture, a feature commonly referred to as isomorphicity. This is particularly challenging when one aims at shifting the emission bands further into the red spectral domains.

Three related structural designs of synthetic 6-aza-uridines were implemented to facilitate a bathochromic shift of emission bands. First, the electronics of 5-substituted uridines were altered by replacing the pyrimidine core with the more electron withdrawing 6-aza-uridine, yielding a red shift as well as hyperchromic effect and higher brightness. In the second design, to further shift the emission into the red region, the polarization of this conjugated electron-poor/electron-rich biaryl system was enhanced by directly conjugating donor groups through an extended aromatic system. The last design resulted in a synthetic uridine surrogate, which can undergo a proton transfer in the excited state (a phenomenon known as ESIPT), thus facilitating a relatively large bathochromic shift while fulfilling the requirements of isomorphic nucleosides.

Furthermore, due to their structural similarity, these isomorphic mimics should fabricate functions and maintain bio-compatibility similar to native building blocks. It has been shown that T7 RNA polymerase accepts the triphosphate-functionalized thiophene 6-aza-uridines, which can determine the affinity of aminoglycoside antibiotics for the singly-modified fluorescent A-site. This finding is important to the biochemical community as it introduces a new attractive pyrimidine analogue, which can be used for probing RNA targets with therapeutic potentials.

To investigate more than uridine analogues, the affinity of T7 RNA polymerase for synthetic pyrimidine and purine mimics was compared and closely evaluated. Subsequently, the partially and fully modified transcripts were successfully synthesized and utilized in a ligation reaction, thereby showing that they can moderately mimic functions and bio-compatibility of the native strands. These results point to intriguing future applications of all four mimics, either site specific modification can be envisioned or partially and fully modified constructs can be utilized for biochemical assays.

Chapter 1   Introduction

1.1 Central Dogma

The central dogma describes flow of the genetic information in living organisms, involving three key macromolecules, deoxyribonucleic acids (DNA), ribonucleic acids (RNA) and proteins (Figure 1.1).[1] The flow of genetic information is rather complex and involving many cofactors, enzymes and complexes. Herein, we present a very simplistic view, and encourage the interested reader to explore other resources, which describe these processes in greater detail.[2-4]

Replication is a fundamental process involved in this path, resulting in two identical replicas of DNA from the original DNA molecule in order to pass on genetic material to new cells.[3, 5-6] Specifically, two DNA strands held together by base stacking between adjacent bases and base paring between complementary constructs are unwound by helicase and topoisomerase. The two separated strands act as templates, however their opposite orientation leads to different replication mechanisms (continuous and discontinuous). In the continuous replication, only one RNA primer binds to the end of the leading strand, which is recognized by DNA polymerase responsible for the synthesis of the complementary daughter strand. The lagging strand receives several RNA primers and is replicated discontinuously forming Okazaki fragments (DNA) for each primer. First the RNA primers are removed and replaced with DNA pieces which are then joined by DNA ligase into one strand.

The second process is called transcription, which describes production of the ribonucleic acid (RNA).[5-6] Enzymes (RNA polymerases) facilitating the production of RNA use sequence of DNA as a template to convert the information into an antiparallel complementary RNA strand.[7-9] Although there are several types of ribonucleic acids, only mRNA, produced by transcription from DNA, serves as template for the translation to amino acids sequences.[2] Ribosomes are the cellular machines responsible for the synthesis of polypeptides, which can fold into three dimensional structures that are determined by the sequence of amino acids. Protein's function is dependent on their complex structures. They function as catalysts of metabolic reactions, they transport and store small molecules, provide immune protection and control growth and differentiation. [2, 10-12]
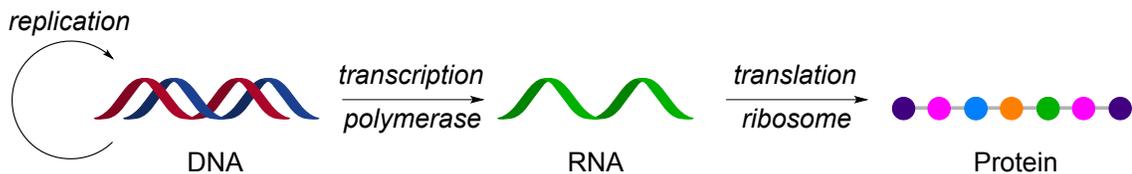
Figure 1.1 Central dogma; flow of genetic information.

1.2 Ribonucleic acid (RNA)

For many years it was believed that RNA had only one role as the messenger of genetic information (mRNA).[13-16] RNA can also transfer a specific amino acid to growing polypeptide (tRNA)[17-18] or serve as a ribosome in the translation process (rRNA).[19-20] Furthermore, researches have shown that RNA can have catalytic functions as ribozymes and regulatory functions such as riboswitches.[10, 13, 21-25] The family of non-coding RNAs is growing rapidly, some of

them are involved in post-transcriptional modification (splicing) such as a small nuclear RNA (snRNA), or others such as microRNA (miRNA) and small interfering RNA (siRNA) play a role in gene regulation and in silencing RNA.[26-28] RNA is now recognized as a multifunctional biomolecule thus, studying its dynamic and structural features should advance knowledge of RNA's roles in key cellular events.

RNA usually exists in a single polynucleotide chain, which can adopt three dimensional structures mostly consisting of small double helical parts connected with single stranded regions.[29] RNA can fold into variety of secondary structures (hairpins, bulges, internal loops) and tertiary structures (junctions, pseudoknots) (Figure 1.2),[30-31] The most common secondary structure is the A-form duplex with a shallow, wide minor groove and deep, narrow major groove, allowing for binding of small and large molecules (such as antibiotics and proteins).[32-37]
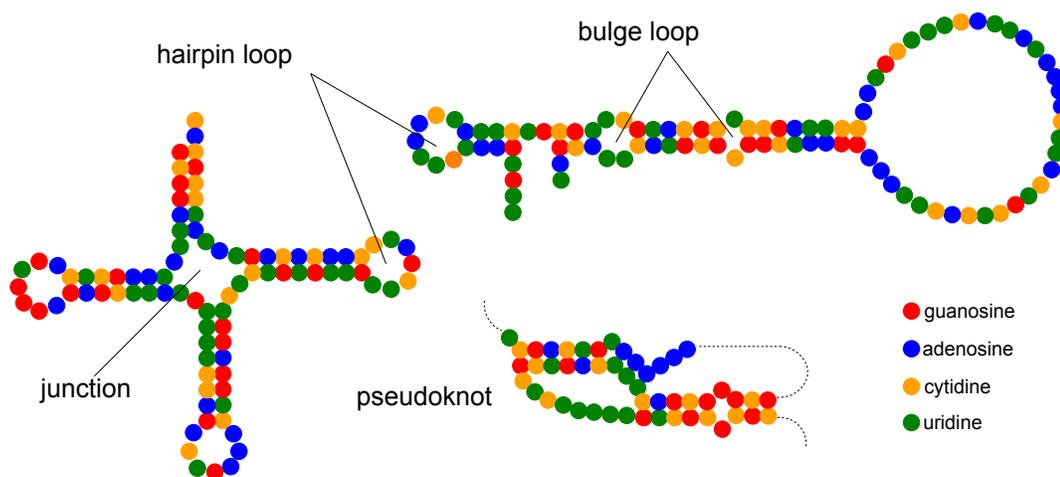
Figure 1.2 Examples of hairpin loops, junctions and bulges in ribonucleic acid.

All nucleotides are constructed out of three components: a nucleobase, a D-ribose and phosphate group (Figure 1.3). There are two major classes of bases:

heterocycle purines and pyrimidines. Found in RNA are guanine (G, **1**), adenine (A, **2**), cytosine (C, **3**) and uracil (U, **4**) (two purines and two pyrimidines, respectively). The base stacking and hydrogen bonding between bases play a major part in conformational preferences of the secondary structures. Recent advancement of computation calculations provides a qualitative description of the aromatic base stacking phenomenon, which is determined by three molecular interactions: dispersion attraction, electrostatic interaction and short-range repulsion.[38-41] In hydrogen bonding, NH groups of the bases are good hydrogen bond donors (**d**), while the lone pairs on the oxygen atom of carbonyl groups and on the nitrogen atom of the heterocycle are good acceptors (**a**) (Figure 1.3). The most dominant base pair pattern is known as Watson-Crick (WC) pairing.[29] In RNA molecules, the mutual recognition of guanine and cytosine (G•C) as well as adenine and uracil (A•U) is particularly important for the formation of vast range of double-stranded oligonucleotides. Pyrimidines are complementary with purines and purine-purine or pyrimidine-pyrimidine pairing is energetically unfavorable. Although the WC base paring is the dominant pattern other pairing is identified as Hoogsteen and "wobble" pairs.[42-44] These non-canonical hydrogen bonding interactions often introduce additional stability to loops and bulges, giving rise to the complex secondary and tertiary RNA structures.[37]
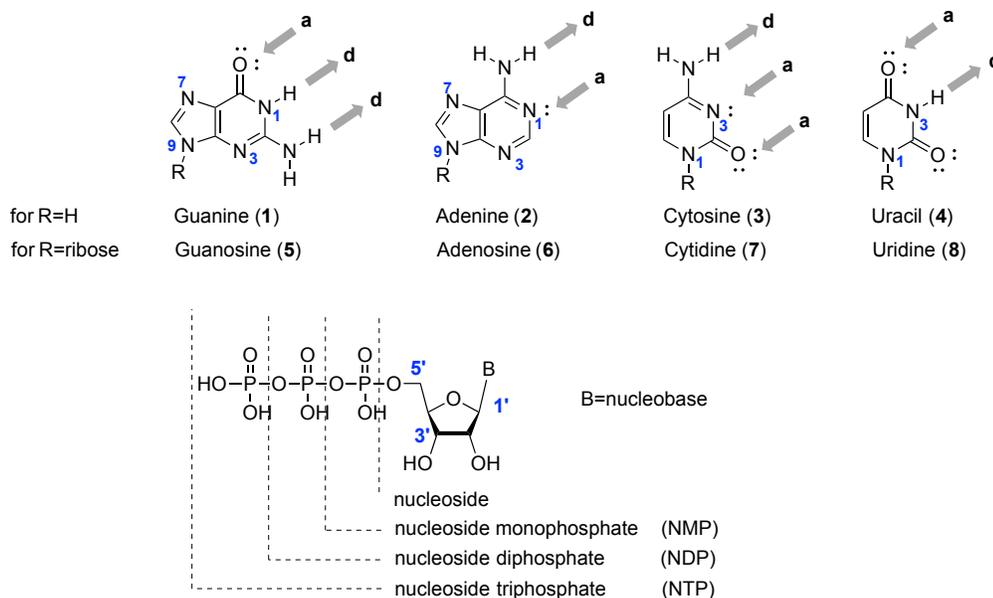
Figure 1.3 Structures of nucleobases, nucleosides and nucleotides depicted for RNA.

The nucleoside consists of a nucleobase connected with ribose via a glycosidic bond. More precisely a nitrogen atom (N9 and N1 for purines and pyrimidines, respectively) from the nucleobase is joined with the C1' of ribose at as the beta anomer. RNA nucleosides **5**–**8** having one, two or three repeating units of 5' phosphate esters are called nucleotides, also referred to as mono-, di- and triphosphate abbreviated NMP, NDP and NTP.

1.3 Photophysical properties of native nucleosides

Neither the sugar nor the phosphate moiety has any significant UV absorption above 230 nm. However, nucleobases absorb strongly at wavelength $\lambda_{max} \approx 260$ nm with molar absorptivity around 10^4 (Table 1.1).[45-47] This suggests that isolated nucleobases, nucleosides and nucleotides show very similar absorption spectra. Noteworthy, the intensity of absorption is significantly reduced

in oligonucleotides due to base stacking interactions. Typically, the decrease of the UV intensity is up to 30%; this phenomenon is called hypochromicity.

Table 1.1 Photophysical properties of nucleosides and nucleotides in water.[45-46]

Name	Absorption		Emission	
	λ_{max} [nm]	ε [$\times$ 10^3 M^{-1} cm^{-1}]	λ_{max} [nm]	ϕ
G (**1**)	253	13.6		
GMP	252	13.7	330	0.00008
A (**2**)	260	14.9	310	0.00005
AMP	259	15.4	312	0.00005
C (**3**)	271	9.1	324	0.00007
CMP	271	9.1	330	0.00012
U (**4**)	262	10.1	309	0.00005
UMP	262	10.0	320	0.00003

Despite early nucleic acid research, which suggested that nucleobases are fluorescent, the naturally occurring bases are essentially non-emissive.[48] Their fluorescence can be described by very low values of quantum yield, which is the fraction of excited molecules that return to the ground state with emission of fluorescence. (Table 1.1).[49] The properties of oligonucleotides presented a major challenge in the investigation of the nucleic acids' folding patterns, dynamics and interactions. The scientific community recognized the necessity for the synthetic modifications of nucleosides. Numerous paths have been taken to introduce desired properties into nucleoside.[49-51] The sensitivity and simplicity of fluorescence spectroscopy has made it one of the most invaluable techniques in exploring nucleic acids and their interactions. Fluorescent sensitivity is typically up to a 1000-fold higher than in absorption spectroscopy. A lot of materials absorb

light, making it difficult to recognize signals of the isolated biomolecules. On a contrary, fluorescence signal can be very selective because far fewer molecules can emit light. Selective excitation makes this technique a very effective tool for *in vitro* and *in vivo* bioassays. It is rather rare that non-targeted molecules will emit at the same wavelength as fluorescent probes. Additionally, emission intensity is usually proportional to the concentration of the targeted molecule (at low concentrations). Fluorescent spectroscopy is also easily accessible, fast and relatively simple as a technique. Most importantly it allows monitoring of real-time processes and is not invasive as a tool. Furthermore, if a synthetic emissive nucleoside is strategically placed within the strand, it can act as a reporter of nucleic acid's conformation changes, hybridization, base flipping and folding events.[52-62]

1.4 Fluorescent nucleosides

The common building blocks of nucleic acids lack useful fluorescent properties. To display favorable properties, the non-emissive native nucleosides need to be modified. Creative design can provide probes with desired excitation and emission wavelengths. Diverse approaches have resulted in countless fluorescent nucleosides utilized as vast variety of reporters.[49, 63-69] They can be classified into five major families of probes: chromophic analogues, pteridines, as well as expanded, extended and isomorphic nucleosides.

Efforts of many researchers produced a family of chromophoric base analogues, in which the natural heterocycle is replaced with a fluorescent aromatic residue. Features typically include high quantum yields ($\phi \approx 1$) and lack the hydrogen-bonding face (Figure 1.4). Some of the polycyclic aromatic hydrocarbons (PAH) such as in compounds **9** and **10** display complex electronic interactions upon base stacking in oligonucleotides.[70-77] Native nucleobase can be replaced not only by new PAH analogues but also with known fluorophores, such as coumarin 102 (**11**).[78-82]

The next class, highly emissive pteridines, was utilized mostly by Hawkins group. These fluorescent nucleosides are analogues of purines, namely of adenosine and guanosine, 6-MAP (**12**) and 3-MI (**13**) respectively.[63, 83-84] Commercial availability and high quantum yields are features making pteridines an attractive emissive family, and in contrast to chromophoric probes, their hydrogen bonding face is restored.
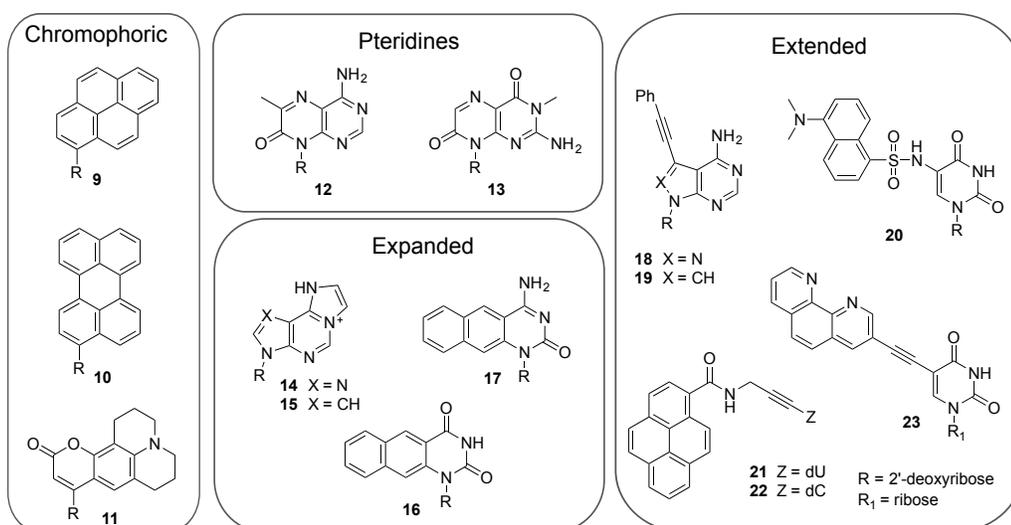
Figure 1.4 Examples of chromophoric base analogues, pteridines, expanded and extended nucleosides.

Non-emissive native bases when fused with additional aromatic rings can yield desired fluorescent properties. One of the first molecules from the expanded nucleoside family was etheno-A (**14**) investigated by Leonard group.[85-86] Both fluorescent nucleoside **14** and its 7-deaza analog (**15**) display an intense emission band in the visible.[87] A quite recent addition to this class is naptho-expended nucleosides **16** and **17**, which display high quantum yields (0.82, 0.62, respectively).[88-89] Particularly in this class, photophysical properties can be tuned by altering the electronic character by conjugating electron rich or poor substituents.

Another design, the extended nucleosides, is based on non-conjugated linkage of known fluorophores, like in structure **20**. The Saito group has linked pyrene to U or C core (**21** and **22**, respectively) and utilized them as very successful single nucleotide polymorphism (SNP) probes.[65] An advantage of such design is that fluorescent features are usually very similar to the parent chromophore. Alternatively, when a fluorophore is conjugated to the nucleobase (**23**) as opposed to being just linked, photophysical properties can be somewhat unpredictable, but more precise information about hydrogen bonding face and environmental changes near the nucleobase can be monitored by following changes in the fluorescence intensity.[90]

To ensure a biologically meaningful read-out a designer probe should closely resemble its natural counterpart in size and shape. Generally including

aromatic rings or extending conjugation will increase steric bulk and therefore can affect the surrounding interactions.

1.5 Fluorescent isomorphic nucleosides

A family of molecules resembling the native building blocks closely and maintaining their distinct biological compatibility is isomorphic nucleosides. Only a limited fraction of fluorescent nucleosides meets the design criteria of isomorphicity. Such modified nucleobases need to closely resemble the corresponding native bases with respect to their size, shape, hydrogen bonding face, and ability to form Watson-Crick paring. The primary design entails simple and minimally perturbing emissive building blocks, which maintain original bio-compatibility and function. The majority of the chemical modifications were introduced at the 7-position of purines and the 5-position of pyrimidines. In our group a successful probe displays emission at long wavelengths (preferably in the visible range) and absorption maxima beyond 300 nm. Additionally, it is desired for the designed fluorescent nucleosides to have high quantum efficiency and retain it upon incorporation into oligonucleotides. Lastly, environmental sensitivity to changes in polarity, viscosity and acidity is preferred as well.

One of the first and most widely employed isomorphic nucleosides is 2-aminopurine (2-AP, **24**, Figure 1.5). Its ability to form Watson-Crick paring, high quantum yield, sensitivity to environmental polarity and commercial availability contribute to its extensive use.[91] Researchers have exploited quenching of
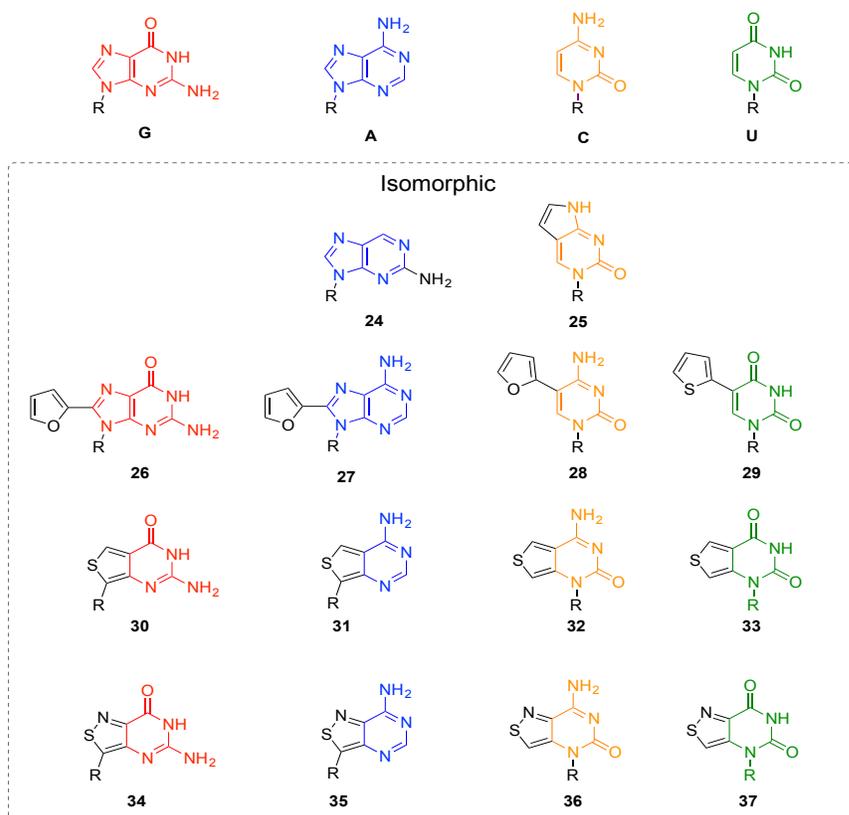
Figure 1.5 Isomorphic fluorescent nucleosides.

emissions intensity upon incorporation into oligonucleotides, however this property significantly limits its utility.[92] PyrroloC (pC, **25**) is another vastly utilized probe. It was discovered as a side product in Sonogashira coupling reactions.[93-94] Its potential as a cytidine surrogate was quickly realized after recognizing its emissive properties. PyrroloC absorbs beyond 340nm and emits in the visible spectrum. Additionally, the minimally perturbing structure has been utilized in many applications.[53, 95] Guanosine and adenosine analogs (**26** and **27**, respectively) were investigated by our group.[96] Both probes display high quantum efficiency (0.69 and 0.57, respectively) and average Stoke shifts (70-80 nm), which describe the difference between absorption and emission bands maxima. Pyrimidine

analogues **28** and **29** isomorphic nucleosides have been incorporated into partially and fully modified oligonucleotides, this study will be covered in detail in chapter 5.

The first fluorescent RNA alphabet was recently synthesized and published by our group in 2011.[97] Both purine and pyrimidine analogues were derived from a single heterocyclic core (**30**–**33**, Figure 1.5). These highly emissive analogs have been utilized for investigation of RNA-related processes, including probing the cleavage of the Hammerhead ribozyme.[62, 98] This study suggested that N-7 position of guanosine is essential for the cleavage reaction of the ribozyme, therefore, in the second generation alphabet, N-7 has been restored (**34**–**37**, Figure 1.5).[99]

1.6 Fluorescent isomorphic uridine analogues

Often emissive isomorphic uridine analogues are modified by conjugating aromatic five-membraned heterocycles to the pyrimidine core. Small heterocyclic rings, such as furan-, thiopene-, oxazole-, thiazole- can be introduced by one-step palladium cross coupling reactions (Figure 1.6).[100-101] The emission properties were shown to be sensitive to changes in environmental polarity while the absorption spectra were almost unchanged, likely indicating a significant component of charge transfer and a twisted conformation of the excited state.[55, 102]

Table 1.2 Photophysical properties of isomorphic uridine analogues in water.[55, 103-106]

	Abs (ε)	Em (ϕ)		Abs (ε)	Em (ϕ)
29	314 (9.0)	434 (0.01)	**41**	325 (9.2)	454 (0.01)
38	316 (11.0)	431 (0.03)	**42**	322 (16.1)	446 (0.19)
39	296 (10.0)	400 (<0.01)	**43**	293 (-)	484 (<0.01)
40	316 (11.5)	404 (<0.01)	**44**	318 (9.3)	458 (0.03)

Absorption and emission wavelengths are reported in nm, the molar extinction factor is relative to water solution and is reported in 10^3 M^{-1} cm^{-1}.

While **39** and **40** were characterized by emission bands at 400 and 404 nm, respectively, and low quantum yields in water, **38** showed an intense red-shifted emission at 430 nm, decaying deeply into the visible range (> 550 nm) (Table 1.2). The absorption and emission spectra remained unchanged upon single incorporation of **38** into single strand ODN. Upon hybridization to a perfect complement, the emission was significantly quenched but easily detectable. Interestingly, the melting curve of a **38**-containing duplex was characterized by identical T_m as compared to the unmodified duplex. When hybridized to a single strand containing an abasic site, the emission of **38** was enhanced 7-fold compared to the perfect matched duplex. The T_m increased by 4 °C, suggesting a flip of the nucleobase and a intrahelical base stacking of the furan moiety, stabilizing the DNA duplex. Moreover, the emission decayed sharper above 500 nm compared to the matched duplex indicating flatter orientation of the pending furan with restricted rotational freedom. The molecular rotor character of **38** was further supported by a remarkable emission intensity enhancement in a binary solvent mixture with increased viscosity as well as by lowering the temperature of

a pure **38** solution in glycerol. These experiments substantiated the hypothesized nucleoside flip and subsequent enhanced emission observed for **38** in an abasic site-containing duplex.[107]

The fluorescence features of **38** were not only sensitive to environmental viscosity but also strongly affected by the polarity of the media. This property, coupled with the proximity of the furan moiety to the nucleobase and its placement in the major groove, prompted its use as a probe for the groove's micropolarity. Different from related investigations carried out with extended probes, **38**, located deeper in the major groove, yielded a rather apolar environmental readout in agreement with theoretical models suggesting a steep increase in the dialectric-constant upon moving away from the groove wall toward the groove exterior.[102]

The advantages of **38** were not limited to its simple preparation or the straightforward conversion to the cytidine analogue,[59] but included enzymatic compatibility and potential applications in biochemical and medicinal chemistry. The uridine derivative was converted to the corresponding 5'-triphosphate and subsequently enzymatically incorporated into oligonucleotides by T7 RNA polymerase with an overall efficiency of 78% compared to the natural UTP. Interestingly, T7 RNA polymerase showed a slight preference for the **38** triphosphate in comparison to the UTP when the transcription reaction was carried out in equimolar concentrations of both triphosphates. Following the same enzymatic procedure, a bacterial A-site labeled with **38** in close proximity of the binding site was prepared to monitor aminoglycoside antibiotics binding by

monitoring fluorescence changes.[56] A similar design was used to monitor the HIV-1 TAR-Tat interaction.[57]

R: deoxyribose or ribose   R: deoxyribose   R: deoxyribose   R: deoxyribose   R: ribose

29   **38**   **39**   **40**   **41**
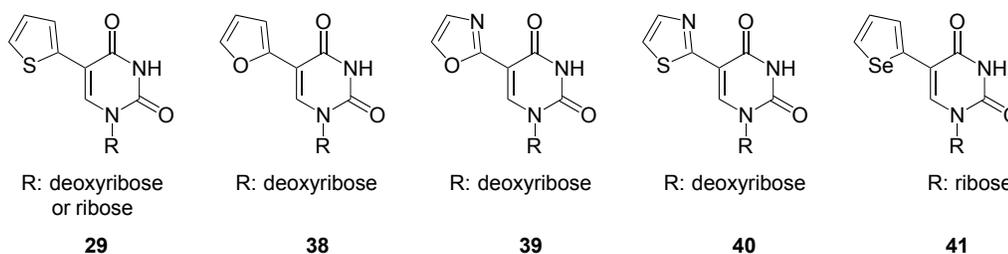
Figure 1.6 Structure isomorphic 5-substituted uridine with single aromatic 5-member ring.

The thiophene analogue **29** showed similar features to that of **38**.[57, 108] Recently, oligonucleotides containing single and multiple-labeling **29** in alternating or neighboring positions were prepared by solid-state incorporation to investigate its effect on duplex stability and the fluorescence response to the duplex formation or dissociation. Multiple incorporations of **29** did not affect the stability or the conformation of the DNA duplexes. Interestingly, an additional Cotton effect in the duplex containing three neighboring **29** was noticed around 320 nm, indicating a likely alignment of the adjacent thiophene units. The fluorescence intensity of the single-labeled, single stranded ODN was slightly quenched upon hybridization to the duplex as is the case for the majority of emissive nucleoside analogues. The oligonucleotide containing three alternating **29** analogues was characterized by a similar emission intensity of the single strand while the fluorescence was dramatically increased upon duplex formation, suggesting a chromophore–chromophore quenching due to the high flexibility in the single stranded oligomer. The same quenching phenomena could also explain the overall lower emission intensity and further quenching in the oligonucleotide bearing three adjacent

modified nucleosides. Such a remarkable variation of the photochemical properties upon hybridization allowed the determination of the duplex melting temperature by monitoring the fluorescence intensity, providing results in agreement with classical T_m measurements.[109]

Recently, a selenophene-based 5-substituted uridine **41** was synthesized and shown comparable fluorescence properties and sensitivity towards environmental polarity (437–453 nm) relative to **38**. The presence of selenium and its X-ray radiation anomalous scattering properties can facilitate the use of this nucleoside as a dual probe for combining solution-based fluorescence measurements and solid-state X-ray studies. This nucleoside was also converted to the triphosphate following standard protocol and enzymatically incorporated into oligonucleotides. Neither the emission intensity nor the helix stability was affected by the presence of **41** as a uridine analogue along the strand. Similar to **38**, the selenophene analogue was used to monitor the binding of aminoglycosides to an A-site RNA with comparable results.[103]

R: deoxyribose	R: deoxyribose	R: ribose	R: deoxyribose
42	**43**	**44**	**45**
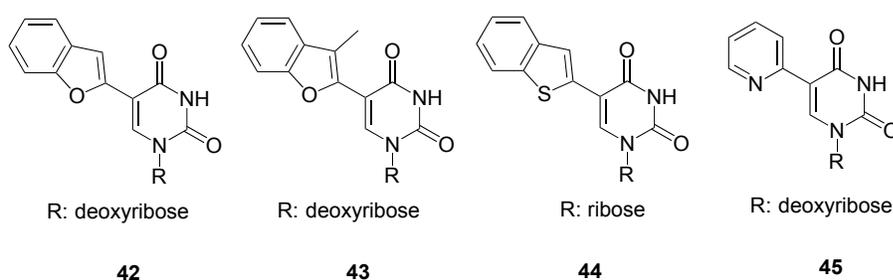
Figure 1.7 Structure isomorphic 5-substituted uridine with extended aromatic rings.

The extension of the electronic conjugation by fusing a phenyl ring to the furano or thiophene moiety of **38** and **29** is a recent addition to this class of isomorphic uridine analogues providing probes **42**, **43** and **44** (Figure 1.7). The

larger π–conjugated system of **42** was shown to impart a minimal red-shift of 6 nm in the absorption maximum (332 nm), while the emission maximum (451 nm) was bathochromically shifted by 15 nm compared to **38**. Despite these unimpressive variations of the photophysical properties, the solvatochromism and the sensitivity towards viscosity were reported to be comparable to that of the furano precursor and were exploited to prove the utility of **42** in the discrimination of flanking base and the detection of abasic sites upon incorporation into oligonucleotides.[104] The introduction of a methyl group in the β–position of the furan ring increased the molecular rotor characteristic of probe **43**. The remarkable sensitivity towards viscosity made this probe useful for the investigation of DNA triplex formation due to its higher emission quantum yield in viscous or apolar media. A 16-fold emission enhancement was reported in response to triplex formation with hairpin duplexes. Finally, the on/off emission of **43** upon triplex formation was applied for the detection of miRNA in conjunction with rolling-cycle amplification.[105]

Similar synthetic pathways and photophysical properties were reported for **44** in which the thiophene ring was extended by a phenyl moiety. The ability of **44** to respond to variation in polarity and viscosity was used to follow the formation of micelles and reverse micelle in good agreement with standard methods. Moreover, the ODNs containing an enzymatically incorporated **44**, were able to report the formation and variation in the environment of micelles, showing its promise for investigating cell-like confined environments.[106, 110]

Historically, some of the isomorphic uridine derivatives were first introduced for entirely different applications and their remarkable environmentally sensitive photophysical properties were unnoticed.[111] Recently, the multisensing character of an emissive deoxyuridine derivative bearing a conjugated pyridine ring **45** was investigated in detail, reporting the ability to tune its photophysical properties in response to change in pH, viscosity and polarity of the media (Figure 1.7). The absorption red-shift, upon protonation of the pyridine nitrogen, was shown to be as remarkable as the emission intensity enhancement of the protonated form, likely due to an intramolecular hydrogen-bonding mediated planarization and rigidification of the extended fluorophore. This finding was in agreement with the previously reported crystal structures for the neutral and protonated molecules containing pyridine fragment in proximity to carbonyl groups.[112] Similar locked conformation between the pyridine and the pyrimidine rings was also achieved by increasing the viscosity in a binary solvent mixture, with a 10-fold increased between pure methanol to pure glycerol. In addition, **45** was also characterized by a high sensitivity towards solvent polarity changes depicted by a Stokes shift variation of around 3500 cm^{-1}, between dioxane (36.0 kcal mol^{-1}) and water (63.1 kcal mol^{-1}).[113]

1.7 Incorporation of modified nucleosides in oligonucleotides (T7 RNA polymerase-mediated *in vitro* transcription reactions)

Fluorescent nucleosides have been valuable reporters used to investigate physical and biological properties of DNA and RNA that would be intractable otherwise. A number of methods including enzymatic incorporation, solid-phase synthesis and post-synthetic modification are available for introducing emissive modification into oligonucleotides.[8-9, 114-115] The first two approaches are most used to incorporate fluorescent nucleosides. Although both methods have their advantages, polymerase-mediated transcription in many cases is more beneficial for the synthesis of longer biologically relevant constructs therefore, it will be discussed herein.

Enzymatic incorporation of the modified nucleosides can be achieved through polymerase-mediated transcription. Although there are few enzymes which have been widely used: T3, T7 and SP6 RNA polymerases we will concentrate on T7.[116-119] This template-dependent enzyme is known to tolerate unnatural bases; therefore, it is extremely useful for fluorescent nucleosides incorporations.[62, 120-123]

In order to site-specifically introduce modifications, a distinctive base-pairing system is needed. Precisely, a short DNA promoter-template duplex directs T7 RNA polymerase to initiate synthesis at a unique position (Figure 1.8).[124] While the transcription reaction usually is initiated with nucleosides triphosphate, other
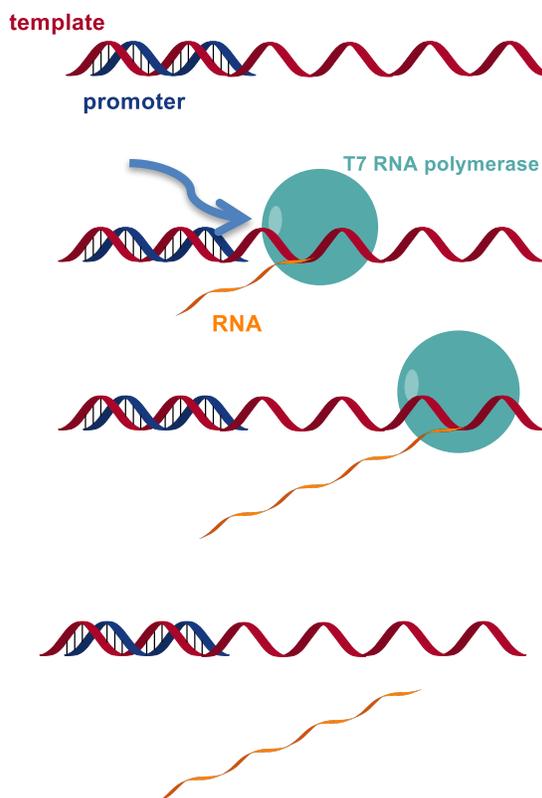
Figure 1.8 In vitro transcription utilizing T7 RNA polymerase.

compounds such as native and modified monophosphates, 5'-azides and 5'-hydroxyl nucleosides can be used as well.[125-131] During the initiation phase the polymerase goes through several elongation attempts, resulting in short abortive transcripts usually 2-8 nucleotides long.[7, 9, 117] Once the polymerase reaches +6 position, the initiation complexes convert to more stable elongation complexes. Further, the transcription proceeds and the enzyme reads DNA template until a full length RNA construct is transcribed. Subsequently, during the *in vitro* run-off transcription T7 RNA polymerase falls off of the template and the transcribed RNA dissociates from the DNA strand.[132-133] Noteworthy, in most of the reactions, the expected full length construct (*N*), which corresponds precisely to the template

sequence is not the only synthesized product. Transcripts having one or two additional residues (*N*+1*, N*+2) added to the 3' end are also obtained.[7, 9]

1.8 Aims of the thesis

Despite numerous technological and instrumental advances, the quality of fluorescence-based analyses ultimately relies on the fluorescent probes themselves. Countless fluorescent tags, labels and probes have been designed over the years, but only a limited fraction meets the stringent design criteria of isomorphicity.[49, 134] The fundamental challenges result from both structural and electronic dilemmas, where any modification aimed at enhancing the electronic features favoring fluorescence can hamper the WC face and its tautomeric preferences, as well as the hybridization and folding features of the resulting oligomers. This issue is particularly challenging when one aims at shifting the emission bands further into the visible and red spectral domains. Such low energy emission is frequently associated with relatively large chromophores with physical footprints much larger than the native nucleobases.[135] The design of isomorphic probes is further challenged by the fact that the relationship between structure and photophysical properties is hard to deduce, and can typically only be elucidated experimentally. This thesis will demonstrate three related structural designs of synthetic uridine surrogates facilitating the bathochromic shift of emission bands (Figure 1.9).[136-137] A small synthetic modification of the parent molecule (**U**) can result in new valuable surrogate, whose photophysical features make them useful as probes, namely for the detection of RNA-ligand interactions.[138]
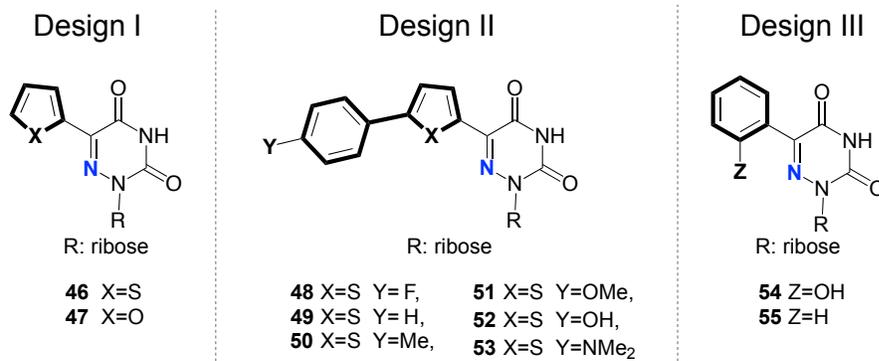
Figure 1.9 Evolution of the design for visibly emitting nucleosides.

Due to the structural similarity, the isomorphic mimics should fabricate functions and bio-compatibility of native building blocks. In order to address this hypothesis, a study of the enzymatic incorporation of isomorphic surrogates (**28**–**31**) into RNA constructs was conducted in a systematic way.[139] Furthermore, transcription efficiency of single/double/triple/quadruple incorporations of modified surrogates, allowing for multiple modifications in the same oligonucleotide is compared. Consequently, synthesis of partially and fully modified RNAs is introduced, which to our knowledge have not yet been reported. To further investigate the impact of synthetic modification on RNA function, the capability of efficient phosphodiester bond formation in 9DB1 DNA enzyme-mediated ligation will be tested.

Acknowledgment

Portion of introduction (subchapter 1.6) has been previously published in the following article: "Emissive 5-Substituted Uridine Analogues" *In Modified Nucleic Acids*, Springer: 2016; 31, 1-26 by Fin, A.; Rovira, A. R.; Hopkins, P. A.; Tor. Y.. The dissertation author is one of the primary authors of this work.

1.9 References

1.   F. Crick, *Nature*, 1970, **227**, 561-563.

2.   J. A. Shapiro, *Natural Genetic Engineering and Natural Genome Editing*, 2009, **1178**, 6-28.

3.   B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Garland Science, New York, Fourth Edition edn., 2002.

4.   L. Stryer, J. L. Tymoczko and B. J. M., *Biochemistry*, W. H. Freeman and Company., 2002.

5.   H. J. Muller, *Am Nat*, 1966, **100**, 493-517.

6.   K. Dose, *J Biol Phys*, 1994, **20**, 181-192.

7.   C. Cazenave and O. C. Uhlenbeck, *P Natl Acad Sci USA*, 1994, **91**, 6972-6976.

8.   J. F. Milligan and O. C. Uhlenbeck, *Method Enzymol*, 1989, **180**, 51-62.

9.   J. F. Milligan, D. R. Groebe, G. W. Witherell and O. C. Uhlenbeck, *Nucleic Acids Res*, 1987, **15**, 8783-8798.

10.   S. A. Strobel and J. C. Cochrane, *Curr Opin Chem Biol*, 2007, **11**, 636-643.

11.   H. D. Madhani, *Cell*, 2013, **155**, 1213-1215.

12.   B. L. Bass and T. R. Cech, *Biochemistry-Us*, 1986, **25**, 4473-4477.

13.   A. Serganov and D. J. Patel, *Nat Rev Genet*, 2007, **8**, 776-790.

14.   F. Jacob and J. Monod, *J Mol Biol*, 1961, **3**, 318-356.

15.   S. Brenner, M. Meselson and F. Jacob, *Nature*, 1961, **190**, 576-581.

16.   F. Gros, C. G. Kurland, H. Hiatt, J. D. Watson, W. Gilbert and R. W. Risebrough, *Nature*, 1961, **190**, 581-585.

17.   P. Zamecnik, M. Hoagland, M. Stephenson and J. Scott, *Acta Union Int Contr*, 1958, **14**, 63-63.

18.   F. Chapeville, G. V. Ehrenstein, S. Benzer, B. Weisblum, W. J. Ray and F. Lipmann, *P Natl Acad Sci USA*, 1962, **48**, 1086-1092.

19. A. Claude, *Science*, 1940, **91**, 77-78.

20. J. Brachet, *Sym Soc Exp Biol*, 1947, **1**, 207-224.

21. K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech, *Cell*, 1982, **31**, 147-157.

22. T. R. Cech, *Annu Rev Biochem*, 1990, **59**, 543-568.

23. J. A. Doudna and T. R. Cech, *Nature*, 2002, **418**, 222-228.

24. S. Altman, *Nat Struct Biol*, 2000, **7**, 827-828.

25. T. R. Cech, *Biochem Int*, 1989, **18**, 7-14.

26. S. Thore, C. Mayer, C. Sauter, S. Weeks and D. Suck, *J Biol Chem*, 2003, **278**, 1239-1247.

27. S. L. Lin, J. D. Miller and S. Y. Ying, *J Biomed Biotechnol*, 2006.

28. K. Ahmad and S. Henikoff, *Cell*, 2002, **111**, 281-284.

29. J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 737−738.

30. M. H. Wilkins, A. R. Stokes and H. R. Wilson, *Nature*, 1953, **171**, 738−740.

31. W. Saenger, *Principles of nucleic acid structure*, Springer-Verlag, New York, 1984.

32. Y. Tor, *Angew Chem Int Edit*, 1999, **38**, 1579-1582.

33. Y. Tor, *Chembiochem*, 2003, **4**, 998-1007.

34. T. Hermann and Y. Tor, *Expert Opin Ther Pat*, 2005, **15**, 49-62.

35. Y. Tor, *Biochimie*, 2006, **88**, 1045-1051.

36. Y. Tor, *Pure Appl Chem*, 2009, **81**, 263-272.

37. G. M. Blackburn and M. J. Gait, *Nucleic acids in chemistry and biology*, Oxford University Press, Oxford, 2nd edn., 1996.

38. J. Sponer, J. E. Sponer, A. Mladek, P. Jurecka, P. Banas and M. Otyepka, *Biopolymers*, 2013, **99**, 978-988.

39. J. Sponer, J. Leszczynski and P. Hobza, *Biopolymers*, 2001, **61**, 3-31.

40. J. Sponer, I. Berger, N. Spackova, J. Leszczynski and P. Hobza, *J Biomol Struct Dyn*, 2000, 383-407.

41. C. R. Martinez and B. L. Iverson, *Chem Sci*, 2012, **3**, 2191-2201.

42. K. Hoogsteen, *Acta Crystallogr*, 1959, **12**, 822-823.

43. K. Hoogsteen, *Acta Crystallogr*, 1963, **16**, 907-916.

44. F. H. C. Crick, *J Mol Biol*, 1966, **19**, 548-555.

45. C. A. Sprecher and W. C. Johnson, *Biopolymers*, 1977, **16**, 2243-2264.

46. P. R. Callis, *Annual Review of Physical Chemistry*, 1983, **34**, 329–357.

47. J. Peon and A. H. Zewail, *Chem Phys Lett*, 2001, **348**, 255-262.

48. S. M. M. Stimson and S. M. A. Reuter, *J Am Chem Soc*, 1941, **63**, 697-699.

49. R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

50. M. Sameiro and T. Goncalves, *Chem Rev*, 2009, **109**, 190-212.

51. A. A. Tanpure, M. G. Pawar and S. G. Srivatsan, *Isr J Chem*, 2013, **53**, 366-378.

52. S.-i. Nakano, Y. Uotani, K. Uenishi, M. Fujii and N. Sugimoto, *Nucl. Acids Res.*, 2005, **33**, 7111-7119.

53. R. H. E. Hudson and A. G. Choghamarani, *Nucleosides, Nucleotides and Nucleic Acids*, 2007, **26**, 533 - 537.

54. K. S. Yang and R. J. Stanley, *Photochem Photobiol*, 2008, **84**, 741-749.

55. N. J. Greco and Y. Tor, *J Am Chem Soc*, 2005, **127**, 10784-10785.

56. S. G. Srivatsan and Y. Tor, *J Am Chem Soc*, 2007, **129**, 2044-2053.

57. S. G. Srivatsan and Y. Tor, *Tetrahedron*, 2007, **63**, 3601-3607.

58. S. G. Srivatsan, N. J. Greco and Y. Tor, *Angew Chem Int Edit*, 2008, **47**, 6661-6665.

59. N. J. Greco, R. W. Sinkeldam and Y. Tor, *Org Lett*, 2009, **11**, 1115-1118.

60. Y. Xie, A. V. Dix and Y. Tor, *J Am Chem Soc*, 2009, **131**, 17605-17614.

61.   Y. Xie, T. Maxson and Y. Tor, *J Am Chem Soc*, 2010, **132**, 11896-11897.

62.   L. S. McCoy, D. Shin and Y. Tor, *J Am Chem Soc*, 2014, **136**, 15176-15184.

63.   M. E. Hawkins, *Cell Biochem Biophys*, 2001, **34**, 257-281.

64.   M. J. Rist and J. P. Marino, *Current Organic Chemistry*, 2002, **6**, 775-793.

65.   A. Okamoto, Y. Saito and I. Saito, *Journal of Photochemistry and Photobiology C*, 2005, **6**, 108-122.

66.   J. N. Wilson and E. T. Kool, *Org Biomol Chem*, 2006, **4**, 4265-4274.

67.   D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378-391.

68.   L. M. Wilhelmsson, *Quarterly reviews of biophysics*, 2010, **43**, 159-183.

69.   M. Kimoto, R. S. I. Cox and I. Hirao, *Expert Review of Molecular Diagnostics*, 2011, **11**, 321-331.

70.   C. Strässler, N. E. Davis and E. T. Kool, *Helv Chim Acta*, 1999, **82**, 2160-2171.

71.   J. M. Gao, C. Strassler, D. Tahmassebi and E. T. Kool, *J Am Chem Soc*, 2002, **124**, 11590-11591.

72.   J. M. Gao, S. Watanabe and E. T. Kool, *J Am Chem Soc*, 2004, **126**, 12748-12749.

73.   R. X. F. Ren, N. C. Chaudhuri, P. L. Paris, S. Rumney and E. T. Kool, *J Am Chem Soc*, 1996, **118**, 7671-7678.

74.   J. N. Wilson, J. M. Gao and E. T. Kool, *Tetrahedron*, 2007, **63**, 3427-3433.

75.   Y. N. Teo and E. T. Kool, *Bioconjugate Chem*, 2009, **20**, 2371-2380.

76.   Y. N. Teo, J. N. Wilson and E. T. Kool, *Chem-Eur J*, 2009, **15**, 11551-11558.

77.   L. H. Yuen, R. M. Franzini, S. L. Wang, P. Crisalli, V. Singh, W. Jiang and E. T. Kool, *Angew Chem Int Edit*, 2014, **53**, 5361-5365.

78.   M. L. Horng, J. A. Gardecki, A. Papazyan and M. Maroncelli, *J Phys Chem-Us*, 1995, **99**, 17311-17337.

79.   R. S. Coleman and M. L. Madaras, *J Org Chem*, 1998, **63**, 5700-5703.

80. R. S. Moog, W. W. Davis, S. G. Ostrowski and G. L. Wilson, *Chem Phys Lett*, 1999, **299**, 265-271.

81. D. Andreatta, J. L. Perez Lustres, S. A. Kovalenko, N. P. Ernsting, C. J. Murphy, R. S. Coleman and M. A. Berg, *J Am Chem Soc*, 2005, **127**, 7270–7271.

82. N. Kitamura, T. Fukagawa, S. Kohtani, S. Kitoh, K. K. Kunimoto and R. Nakagaki, *J Photoch Photobio A*, 2007, **188**, 378-386.

83. M. E. Hawkins, W. Pfleiderer, A. Mazumder, Y. G. Pommier and F. M. Falls, *Nucleic Acids Res*, 1995, **23**, 2872-2880.

84. M. E. Hawkins, L. Brand and M. L. Johnson, *Method Enzymol*, 2008, **450**, 201-231.

85. J. A. Secrist, G. Weber, N. J. Leonard and J. R. Barrio, *Biochemistry-Us*, 1972, **11**, 3499-3506.

86. J. A. Secrist, J. R. Barrio and N. J. Leonard, *Science*, 1972, **175**, 646-647.

87. F. Seela, E. Schweinberger, K. Y. Xu, V. R. Sirivolu, H. Rosemeyer and E. M. Becker, *Tetrahedron*, 2007, **63**, 3471-3482.

88. F. Godde, J. J. Toulme and S. Moreau, *Biochemistry-Us*, 1998, **37**, 13765-13775.

89. F. Godde, J. J. Toulme and S. Moreau, *Nucleic Acids Res*, 2000, **28**, 2977-2985.

90. D. J. Hurley, S. E. Seaman, J. C. Mazura and Y. Tor, *Org Lett*, 2002, **4**, 2305-2308.

91. D. C. Ward, E. Reich and L. Stryer, *J Biol Chem*, 1969, **244**, 1228-1237.

92. A. A. Marti, S. Jockusch, Z. M. Li, J. Y. Ju and N. J. Turro, *Nucleic Acids Res*, 2006, **34**, e50.

93. M. J. Robins and P. J. Barr, *Tetrahedron Lett*, 1981, **22**, 421-424.

94. M. J. Robins and P. J. Barr, *J Org Chem*, 1983, **48**, 1854-1862.

95. R. H. E. Hudson and A. Ghorbani-Choghamarani, *Synlett*, 2007, 870-873.

96. N. J. Greco and Y. Tor, *Tetrahedron*, 2007, **63**, 3515-3527.

97. D. Shin, R. W. Sinkeldam and Y. Tor, *J Am Chem Soc*, 2011, **133**, 14912-14915.

98. D. Shin, P. Lonn, S. F. Dowdy and Y. Tor, *Chem Commun*, 2015, **51**, 1662-1665.

99. A. R. Rovira, A. Fin and Y. Tor, *J Am Chem Soc*, 2015, **137**, 14602-14605.

100. A. J. Gutierrez, T. J. Terhorst, M. D. Matteucci and B. C. Froehler, *J Am Chem Soc*, 1994, **116**, 5540-5544.

101. P. Wigerinck, C. Pannecouque, R. Snoeck, P. Claes, E. Declercq and P. Herdewijn, *J Med Chem*, 1991, **34**, 2383-2389.

102. R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chembiochem*, 2008, **9**, 706-709.

103. M. G. Pawar, A. Nuthanakanti and S. G. Srivatsan, *Bioconjugate Chem*, 2013, **24**, 1367-1377.

104. A. A. Tanpure and S. G. Srivatsan, *Chembiochem*, 2012, **13**, 2392-2399.

105. T. Kanamori, H. Ohzeki, Y. Masaki, A. Ohkubo, M. Takahashi, K. Tsuda, T. Ito, M. Shirouzu, K. Kuwasako, Y. Muto, M. Sekine and K. Seio, *Chembiochem*, 2015, **16**, 167-176.

106. M. G. Pawar and S. G. Srivatsan, *Org Lett*, 2011, **13**, 1114-1117.

107. R. W. Sinkeldam, A. J. Wheat, H. Boyaci and Y. Tor, *Chemphyschem*, 2011, **12**, 567-570.

108. M. S. Noe, A. C. Rios and Y. Tor, *Org Lett*, 2012, **14**, 3150-3153.

109. M. S. Noe, R. W. Sinkeldam and Y. Tor, *J Org Chem*, 2013, **78**, 8123-8128.

110. M. G. Pawar and S. G. Srivatsan, *J Phys Chem B*, 2013, **117**, 14273-14282.

111. A. J. Gutierrez, T. J. Terhorst, M. D. Matteucci and B. C. Froehler, *J Am Chem Soc*, 1994, **116**, 5540-5544.

112. K. M. Clapham, A. S. Batsanov, R. D. R. Greenwood, M. R. Bryce, A. E. Smith and B. Tarbit, *J Org Chem*, 2008, **73**, 2176-2181.

113. R. W. Sinkeldam, P. Marcus, D. Uchenik and Y. Tor, *Chemphyschem*, 2011, **12**, 2260-2265.

114. J. Goodchild, *Bioconjugate Chem*, 1990, **1**, 165-187.

115. S. Verma and F. Eckstein, *Annu Rev Biochem*, 1998, **67**, 99-134.

116. E. T. Butler and M. J. Chamberlin, *J Biol Chem*, 1982, **257**, 5772-5778.

117. G. Krupp, *Gene*, 1988, **72**, 75-89.

118. P. Davanloo, A. H. Rosenberg, J. J. Dunn and F. W. Studier, *P Natl Acad Sci-Biol*, 1984, **81**, 2035-2039.

119. H. Kotani, Y. Ishizaki, N. Hiraoka and A. Obayashi, *Nucleic Acids Res*, 1987, **15**, 2653-2664.

120. S. G. Srivatsan and Y. Tor, *Tetrahedron*, 2007, **63**, 3601-3607.

121. Y. Tor and P. B. Dervan, *J Am Chem Soc*, 1993, **115**, 4461-4467.

122. S. G. Srivatsan and Y. Tor, *Nat Protoc*, 2007, **2**, 1547-1555.

123. S. G. Srivatsan and Y. Tor, *Chem-Asian J*, 2009, **4**, 419-427.

124. C. T. Martin and J. E. Coleman, *Biochemistry-Us*, 1987, **26**, 2690-2696.

125. R. Fiammengo, K. Musilek and A. Jaschke, *J Am Chem Soc*, 2005, **127**, 9271-9276.

126. F. Q. Huang, J. He, Y. L. Zhang and Y. L. Guo, *Nat Protoc*, 2008, **3**, 1848-1861.

127. F. Q. Huang, G. C. Wang, T. Coleman and N. Li, *RNA-Publ. RNA Soc.*, 2003, **9**, 1562-1570.

128. G. H. Lee, H. K. Lim, W. Jung and S. S. Hah, *B Korean Chem Soc*, 2012, **33**, 3861-3863.

129. E. Paredes and S. R. Das, *Chembiochem*, 2011, **12**, 125-131.

130. D. Williamson, M. J. Cann and D. R. W. Hodgson, *Chem Commun*, 2007, DOI: Doi 10.1039/B712066d, 5096-5098.

131. J. Wolf, V. Dombos, B. Appel and S. Muller, *Org Biomol Chem*, 2008, **6**, 899-907.

132. G. A. Diaz, M. Q. Rong, W. T. McAllister and R. K. Durbin, *Biochemistry-Us*, 1996, **35**, 10837-10843.

133. S. N. Kochetkov, E. E. Rusakova and V. L. Tunitskaya, *Febs Lett*, 1998, **440**, 264-267.

134. A. Fin, A. R. Rovira, P. A. Hopkins and Y. Tor, in *Modified Nucleic Acids*, eds. K. Nakatani and Y. Tor, Springer International Publishing, Cham, 2016, DOI: 10.1007/978-3-319-27111-8_1, pp. 1-26.

135. L. D. Lavis and R. T. Raines, *Acs Chem Biol*, 2008, **3**, 142-155.

136. R. W. Sinkeldam, P. A. Hopkins and Y. Tor, *Chemphyschem*, 2012, **13**, 3350-3356.

137. P. A. Hopkins, R. W. Sinkeldam and Y. Tor, *Org Lett*, 2014, **16**, 5290-5293.

138. P. A. Hopkins, L. S. McCoy and Y. Tor, *submitted*.

139. P. A. Hopkins, L. S. McCoy, A. Fin and Y. Tor, *in preparation*.

Chapter 2   Modified 6-Aza Uridines as Highly Emissive Fluorescent Nucleosides: Design, Synthesis and Applications

2.1 Design

To explore nucleic acids, the non-emissive canonical nucleosides need to be structurally modified to endow them with useful fluorescent properties. Such alterations cannot impede native W-C base-pairing and subsequent helix formation.[1-3] Successful nucleoside probes reconcile these structural and photophysical demands with an exclusively excitable wavelength, red-shifted from the ~260 nm absorption band of the native nucleobases. Responsive emissive nucleoside probes, unlike fluorescent tags or labels, are characterized by spectroscopic properties, which are sensitive to changes in the local environment. Several groups, including ours, have reported a rich variety of isomorphic fluorescent nucleoside analogs.[1-14] The basic design principle of our first generation emissive nucleobase analogs has relied on conjugating five membered aromatic heterocycles to the pyrimidine core at the 5-position.[2] The modification at this position appears to be structurally benign, with no impact on the *anti* orientation or sugar pucker of the nucleosides. The resulting nucleosides emit in the visible range (390–443 nm), have rather large Stokes shifts (8400–9700 cm^{-1}), but suffer from relatively low emission quantum efficiency (Φ=0.01–0.035).[15] These nucleosides behave like molecular rotors, with the 5-membered

heterocycles and the pyrimidine core acting as the "donor and acceptor", respectively.[16] Rotation around the biaryl linkage provides an effective channel for non-radiative torsional relaxation, thus leading to low emission quantum yield in non-viscous media.[16, 17] We hypothesized that altering the electronics by replacing the pyrimidine with the corresponding 1,2,4-triazine core (i.e., 6-aza-pyrimidine)[18] would enhance the charge transfer character, ideally yielding a bathochromic shift as well as a hyperchromic effect and higher brightness. Herein we report a simplified synthetic pathway to thiophene and furan-modified 6-azauridines, discuss their most relevant photophysical properties, and reveal their responsiveness to environmental polarity and pH.

As articulated above, replacing the uridine core with 6-azauracil, a more electron deficient heterocycle, while keeping the electron rich thiophene and furan moieties in the 5-position, was expected to maintain an extended $\pi-$system while augmenting the donor-acceptor nature of this motif (Scheme 2.1). This enhanced "push-pull" interaction is expected to affect a red-shift of the absorption maxima and likely a concomitant red-shifted emission maxima. Such a donor$-$acceptor interplay is also frequently associated with enhanced sensitivity toward environmental polarity.[19-22] The single bond linkage between the 5-membered heterocycle and the 6-membered nucleobase introduces a molecular rotor element likely to render **46** and **47** sensitive toward viscosity (molecular crowding) as demonstrated for other 5-aryl modified pyrimidines.[16, 17] Chromophores possessing such a feature typically reveal much higher emission intensity in viscous compared to non-viscous media. Viscosity (molecular

crowding) hampers formation of a twisted excited state thereby limiting the rotational relaxation, a non-emissive decay pathway, to the ground state.[23, 24]

Only a few fluorescent nucleosides have been shown to be pH sensitive.[17, 25, 26] 8-Aza-guanosine, an isomorphic fluorescent purine analog,[27] shows its pH-dependent emission characteristics as a nucleoside[28] and after enzymatic incorporation into RNA.[29] This feature was exploited in the investigation of the catalytic function of RNA,[30] An application reveals interesting opportunities for fluorescent nucleosides that display spectral sensitivity toward (de)protonation. No emissive pyrimidines with pKa values near physiological pHs have, however, been documented. Importantly, the presence of the nitrogen atom at position 6 has a profound influence on the acidity of the NH, lowering its pKa to 6.8 vs. a pKa value of 9.3 for the parent uridine.[31] Like uridine, 6-aza-uridine has no appreciable fluorescence properties hampering visualization of (de)protonation events with fluorescence spectroscopy. Assuming the emissive features discovered for the parent 5-modified uridines translate to (or are enhanced in) the 6-aza motif, deprotonation and protonation events are likely to be revealed by changes in the emission profile. This new motif would therefore constitute the first example of a pH responsive emissive pyrimidine with a pKa value near physiological pHs.

2.2 Synthesis

Condensation of commercially available thiophene glyoxylic acid (**56**) with semicarbazide (**58**) gave imine **59** which, in situ, cyclized upon heating under basic conditions to furnish **61** in good yields. The same approach, starting from **57**, provided **62**. Glycosylation of **61** and **62** was performed in a two-step one-pot

procedure. First, the nucleobase was activated by reaction with bis(trimethylsilyl)acetamide (BSA) in acetonitrile for ~30 minutes at rt, turning the initial suspension into a clear solution. Secondly, the reaction mixture was brought to 85°C after which β-d-ribofuranose 1-acetate 2,4,5-tribenzoate (**63**) was added, immediately followed by addition of trimethylsilyltriflate (TMSOTf). The reaction was allowed to run for 30 minutes after which TLC analysis indicated full conversion of the nucleobase to the desired glycosylated product. An aqueous work-up followed by purification with column chromatography yielded **64** and **65** in good yields. Deprotection with methanolic ammonia at 60°C, followed by column chromatographic purification gave desired **46** and **47**.
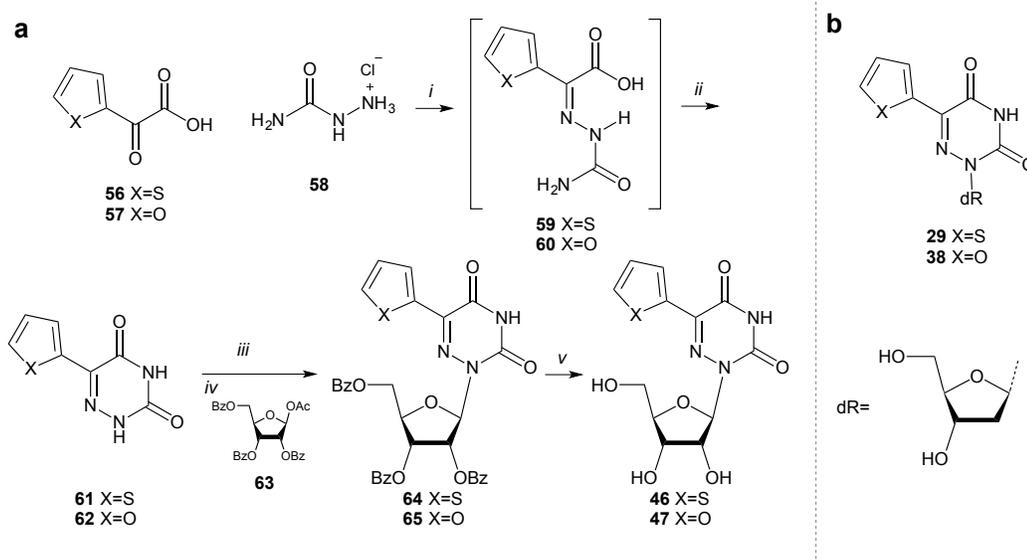
Scheme 2.1 a) Syntheses of nucleosides **46** and **47**, reagents and conditions: i) semicarbazide hydrochloride (**58**) (1.1 equiv.), H_2O, 60°C, 17h, ii) add. 1 M NaOH aq. (3 equiv.), 60°C, 4h, iii) BSA (3 equiv.), CH_3CN, rt, 30', iv) β-d-ribofuranose 1-acetate 2,4,5-tribenzoate (**6**) (1.1 equiv.), TMSOTf (1.1 equiv.), CH_3CN, 85°C, 30', v) NH_3/MeOH, 60°C, 24h; and b) structures of reported 5-(thiophene-2-yl)-2'-deoxyuridine (**29**) and 5-(furan-2-yl)-2'-deoxyuridine (**38**).

2.3 Photophysical properties

2.3.1 Basic photophysics in water

In unbuffered deionized water, **46** is characterized by a red-shifted absorption maximum at 332 nm, which upon excitation gives a visible fluorescence signal, peaking at 455 nm, with a respectable quantum yield (Φ = 0.20) and a fluorescence lifetime of 4.9 ns (Table 2.1). In contrast, a solution in dioxane yields an absorption maximum at 335 nm, an emission maximum at 415 nm (Φ = 0.80) and a slightly longer fluorescence lifetime of 5.4 ns. The small change in absorption maximum, but 40 nm difference between the emission maxima in the two solvents prompted us to study the sensitivity to polarity in greater detail (see below).

Table 2.1 Selected spectroscopic properties of **46** and **47**.[a]

	Solvent	Absorption		Fluorescence			Brightness	Stokes shift
		λ_{max}	ε	λ_{max}	Φ	τ	$\Phi \times \varepsilon$	$\nu_{abs}\text{-}\nu_{em}$
		nm	$\times 10^4\,M^{-1}cm^{-1}$	nm		ns	$\times 10^3$	cm^{-1}
46	Dioxane	335	1.3	415	0.80	5.4	10.4	6025
	Water	332	1.1	455	0.20	4.9	2.2	8492
	pH 2.55	336	1.2	462	0.13	3.0[b]	1.6	8485
	pH 10.13	325	1.2	426	0.39	6.6[b]	4.7	7588
	MeOH	334	1.1	433	0.50	6.6	5.5	7332
	Glycerol	339	1.0	444	0.66	7.9	6.6	7372
47	Dioxane	327	1.2	414	0.60	5.6	7.2	6824
	Water	320	1.0	443	0.05	3.1	0.5	9107
	pH 2.55	329	1.1	485	<0.01	<1[b]	<0.1	9861
	pH 10.13	319	1.1	446	0.11	2.8[b]	1.2	9273
	MeOH	326	1.1	442	0.04	1.6	0.4	8585
	Glycerol	331	1.1	434	0.46	2.7	5.1	7574

[a] The Stokes shifts are calculated with corrected emission maxima obtained using: Intensity $[\nu] = \lambda^2 \times$ Intensity $[\lambda]$.[19]   [b] Values for pH 2.55 and pH 10.13 (aq. 10 mM phosphate, 0.1 M NaCl buffers) represent the lower- and upper plateau values of the sigmoidal fit (Figure 2.2). All values represent averages, see references for error analysis.

2.3.1 Sensitivity to polarity

Influence of polarity on spectroscopic parameters of fluorescent probes was studied using samples in dioxane ($E_T(30)$ = 36.0 kcal/mol), water ($E_T(30)$ = 63.1 kcal/mol) and mixtures thereof.[32] Water–dioxane samples of **46** and **47** were prepared from a concentrated DMSO stock solution and subjected to an absorption and steady state emission spectroscopy study (Figure 2.1 and Table 2.1). The small amount of DMSO has been established to have negligible impact on polarity.[33]

The Stokes shifts ($\nu_{abs}-\nu_{em}$) were calculated for the samples in dioxane, water, and mixtures thereof and plotted, after conversion from cm^{-1} to kcal/mol, as a function of the experimentally-determined samples' $E_T(30)$ values (Figure 2.1b). The data points clearly line-up linearly as was established by a good linear fit with a slope value of 0.268, quantifying the sensitivity to polarity (Table 2.2). In addition to a poorer linear fit, the sensitivity to polarity for **47** was lower (Table 2.2 and references).

Table 2.2 Sensitivity to polarity[a]

	Slope cm^{-1}/kcal mol^{-1}	Y-int. cm^{-1}	Slope -	Y-int. kcal mol^{-1}	R^2
46	93.7	2548	0.268	7.29	0.99510
29	63.8	5188	0.182	14.83	0.86159
47	75.4	4290	0.216	12.27	0.84211
38	79.5	4437	0.227	12.69	0.96922

[a] Polarity sensitivity equals the slope of the linearized Stokes shift vs. $E_T(30)$ plot. Values for 5-(thiophen-2-yl)-2dU (**29**) and 5-(fur-2-yl)-2dU (**38**) in cm^{-1}/kcal mol^{-1} have been reported.[15] A unitless value for polarity sensitivity is obtained by conversion of the Stokes shift value from cm^{-1} to kcal mol and replotting it as a function of the sample $E_T(30)$ value (in kcal mol^{-1}). The adjusted R^2 represents the goodness of the linear fit.
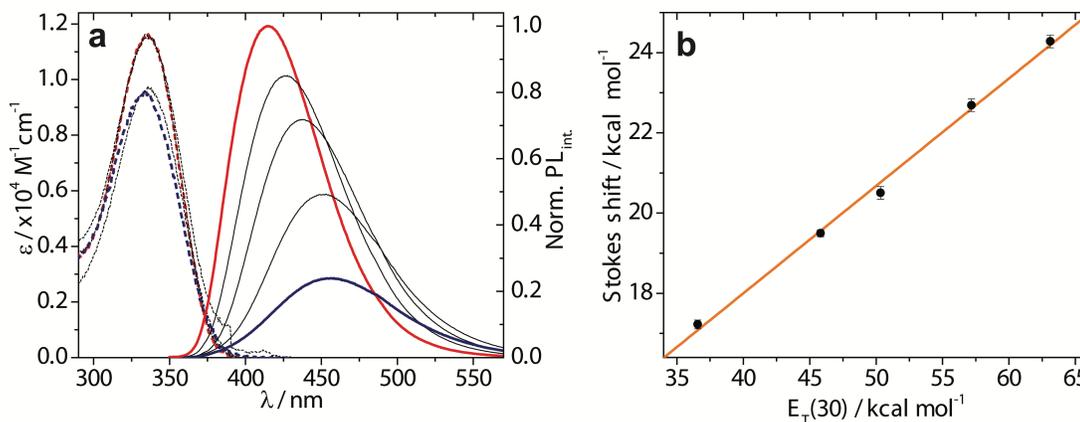
Figure 2.1 a) Absorption (dashed line) and emission (solid line) spectra of **46** in dioxane (red), water (blue), and binary mixtures of 10, 30, and 70 v% water in dioxane (red). Samples' concentration is 7.9×10^{-6} M, emission is recorded after excitation at 335 nm. b) A plot of the Stokes shift (in kcal/mol) vs. $E_T(30)$ values (in kcal/mol) obtained from dioxane–water mixtures (data points with error bars) and linearization (orange line). Absorption and emission spectra of each sample are recorded in duplicate; only one set is shown.

2.3.2 Sensitivity to pH

The sensitivity of the emissive 6-aza-U chromophores to pH was studied in aqueous phosphate buffers with pH values ranging from 0.5 to 12. Absorption spectra only showed small differences in maxima (λ_{max} = 326 nm), while the corresponding emission spectra revealed a strong influence of pH on the position as well as intensity of the emission maximum (Figure 2.2a, Table 2.1). The emission maximum shifted from ~426 nm (Φ = 0.39) to ~462 nm (Φ = 0.13), going from pH 10.13 to 2.55, respectively. Curve-fitting the normalized fluorescence intensity as a function of pH gave a sigmoidal curve from which a pKa, value of 6.7 was calculated (Figure 2.2b). Using the same approach a pKa value of 6.6 for probe **47** was determined. The difference between the quantum yield under acidic

and basic conditions was also reflected by the difference in fluorescence lifetimes, 3.0 and 6.6 ns, respectively (Table 2.1).
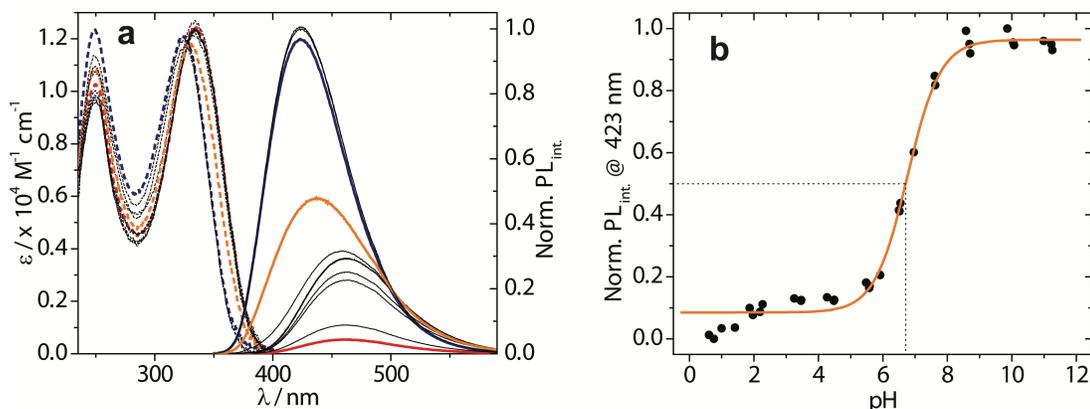
Figure 2.2 a) Absorption (dashed lines) and emission plots (solid lines), corrected for small differences in the O.D. at the λ_{ex}(335 nm) of **46** at pH 0.61 (red), pH 10.99 (blue), pH 6.55 (orange) and intermediate pH values (black). b) Plot of the normalized emission intensity as function of sample pH (solid circles) fitted to a sigmoidal curve (orange line). The dashed lines illustrate a graphical determination of the pK_a value.

2.3.3 Sensitivity to viscosity

Absorption and emission spectra of **46** in samples of methanol, glycerol and mixtures thereof reveal virtually no responsiveness toward viscosity for **46** (Figure 2.3), but indicate a significant sensitivity to viscosity for **47**. A double log plot of the fluorescence intensity as a function of sample viscosity for **47**, however, reveals poor linearity. The high fluorescence quantum yields for **46** in methanol (0.50) and glycerol (0.66), are accompanied by the longest lifetimes observed in this study, 6.6 and 7.9 ns, respectively.
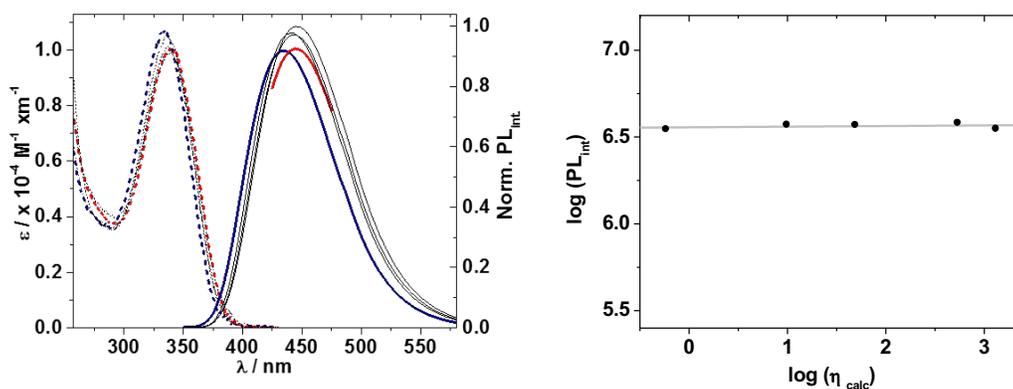
Figure 2.3 a) Absorption (dashed lines) and emission (solid lines) curves for **46**, corrected for small differences in the O.D. at the λ_{ex}(340 nm), in methanol (blue), glycerol (red) and mixtures thereof (black lines). b) correlation of the log fluorescence intensity as a function of the log viscosity (solid circles), with a linear fit (grey line) (the Förster–Hoffman plot).[34]

2.4 Discussion

Synthesis

The earliest scheme described for the synthesis of **62**, is a two-pot two-step procedure starting from furanglyoxylic acid (**57**), which was reacted with semicarbazide hydrochloride (**3**) to give the corresponding formic acid semicarbazone (**60**). This, in turn, yielded **62** upon refluxing in a mixture of propylene glycol–ethanol in the presence of sodium ethoxide.[35] Later, Walker et. al. reported a procedure for the synthesis of **56**, but used thio-semicarbazide instead of semicarbazide (**3**), necessitating a third step to convert the thiocarbonyl into a carbonyl moiety.[36] To shorten the synthesis of **61** and **62** we started from semicarbazide and convert the original two-pot two-step synthesis[35] to a simplified one-pot two-step procedure.[37] Importantly, this streamlined approach works well for the synthesis of **61** and, albeit with lower yield, also for **62**, while the three-step three-pot synthesis by Walker et. al. works for the synthesis of **61** it but proved

very challenging for the synthesis of **62**.[36, 38] Glycosylation of 5-substituted 6-aza uracils has been reported.[36, 37, 39] We used a slightly modified procedure involving activation of the nucleobase with BSA in CH_3CN instead of TMSCl and Et_3N in benzene. We performed the glycosylation directly in situ after BSA treatment, activating the ribose (**6**) with TMSOTf, instead of changing the solvent from benzene to dichloroethane and using stannic chloride to activate the ribose (**6**). Our procedure proved facile, quick, and high yielding for the synthesis of **64**, and **65**. Deprotection of the benzoyl groups to give **46** and **47** was performed using methanolic ammonia for 24 hours at 60°C in a pressure vessel. Deprotection also succeeded with aqueous ammonia in methanol in 17 hours at 40°C, albeit in lower yield. Both **46** and **47** were synthesized in 3 steps from commercially available starting materials in 41% and 20% overall yield, respectively.

Fluorescence Characteristics and their Sensitivity to Polarity

Diverse events can impact the structure of nucleic acids (e.g., (un)folding, binding of low MW ligand, lesions), which often result in changes in the local polarity. Nucleic acid probes able to detect such changes in environmental polarity have proven to be useful tools.[1, 33, 40-47] We therefore thoroughly evaluate the influence of polarity on a probe's photophysical properties. We regularly use water–dioxane mixtures, as they not only vary in polarity, but also present changes in hydrogen bonding donor and acceptor composition prone to exert specific solvent–solute interactions.[31] Moreover, the polarity window obtainable with such water–dioxane mixtures likely reflects most biological relevant environmental polarities. Using a microenvironmental polarity parameter such as the $E_T(30)$ scale

with values ranging from 36.0 to 63.1 kcal/mol for pure dioxane to water, respectively,[20] typically provides good correlation with spectral features like Stokes shifts.[32]

While the quantum yield for **46** in water is modest ($\Phi = 0.20$) and low for **47** ($\Phi = 0.05$), the quantum yield displayed by **46** in dioxane ($\Phi = 0.80$) is among the highest values reported for an isomorphic nucleoside (Table 2.1).[1]   Both the quantum yield in water and dioxane represent a huge improvement compared to the values reported for the corresponding 2'-deoxyuridine modified with either a thiophene (**29**) ($\Phi_{water} = $ 0.01) or furan (**38**) ($\Phi_{water} = 0.03$) in the five position (Scheme 2.1).[2, 15] Interestingly, uridine analogs **29** and **38** display the opposite behavior with even lower quantum yield in apolar organic media.[2, 15, 33] Incidentally, this remarkable difference in fluorescence quantum yield between these isoelectronic nucleoside analogs shows that photophysical properties are difficult to predict from the molecular structure alone.

It is worth noting that we previously reported the sensitivity to polarity for **29** and **38** in cm^{-1}/kcal mol^{-1} (Table 2.2). To obtain unitless values, the values for the Stokes shift in cm–1 were converted to kcal/mol, and replotted as a function of the samples' $E_T(30)$ value. The slope for the linear relationship, the polarity sensitivity, for modified uridines **29** and **38** was determined to be 0.182 and 0.227, respectively, both significantly less sensitive than **46** (Table 2.2).   These substantial differences illustrate how the enhanced push-pull motif present in **46** and **47**, as compared to **29** and **38** indeed results in the expected enhanced susceptibility to polarity.

Sensitivity to pH

Sensitivity to pH as a characteristic of fluorescent nucleosides is relatively uncommon, although a couple of examples have been reported.[17, 26, 28] In these cases the absorption spectra show a two-state transition with an isosbestic point. The influence of pH on the emission properties is reflected by a strong increase or decrease of the fluorescence intensity not necessarily accompanied by significant shifts of the emission maxima.

Nucleoside **46** reveals its pH sensitivity by small changes in absorption, but large changes in emission intensity and maxima (Figure 2.2a). Changing from basic to acidic conditions the fluorescence curve of **46** shifts from a maximum located at 426 nm to 462 nm with a concurrent drop in the fluorescence quantum yield from 0.39 to 0.13, respectively (Figure 2.2a, Table 2.1). Fluorescence decay analysis reveals that the fluorescence lifetime in basic conditions is more than doubled compared to the value in acidic conditions. The drop in normalized fluorescence intensity as a function of pH gives, after a Boltzmann fit of the data points, an S-curve indicative of a two-state (de)protonation process with a pKa of 6.7 and 6.6 for **46**, and **47** respectively (Figure 2.2b and references). These values are in agreement with a reported value for 6-aza-uridine (pKa = 6.8).[31] It further shows that the nature of the modification in the five position of the six membered heterocycle has little effect on the pKa of the NH as its electronic influence is limited. The observed pKa values significantly deviate from the reported values for the parent uridine (9.3–9.5),[31, 48, 49] demonstrating that the additional nitrogen renders the NH ~1000 fold more acidic. This finding is corroborated by others who

have reported that substitution of H-6 in native uracil with an electron withdrawing chloride has a spectacular influence on the pKa value causing a drop from 9.5 to 5.8.[50] Closer inspection of Figure 2.2a shows a downward trend of the fluorescent intensity at very low pHs. Although speculative at this stage, this might suggest that the 6-aza nitrogen gets protonated at pHs far below 2. A normalized plot excluding these low pH data points (pH < 2.5) gives a slightly higher pKa value of 6.85 for the heterocycle's NH (Figure 2.9). Regardless of this putative second transition, **46** shows fluorescence intensity as well as emission maximum sensitivity with a pKa close to physiological pH. This makes **46** an attractive probe for studies involving (de)protonation events on RNA, as there are no known isomorphic pyrimidine analogs with pKa values at this range.[30]

Sensitivity to Viscosity

The nucleobase moiety in **46** and **47** can be described as a molecular rotor, being comprised of a single bond linking two π-systems.[16] Such molecules can lose their excitation energy via a non-emissive rotational decay from a twisted excited state. Molecular crowding effects such as the one induced by viscous solvents can hinder formation of a twisted excited state, thereby limiting the non-emissive rotational decay process resulting in an increased emission intensity.[51] The contribution of this non-emissive decay pathway can be studied by plotting the emission intensity as a function of sample viscosity.[34] A practical approach to controlling sample viscosity is the use of binary mixtures of solvents with very low and very high viscosities that differ minimally in polarity.[16, 17, 52] Suitable non-viscous and highly viscous solvents are methanol ($\eta_{20°C}$ = 0.583 cp), and glycerol

($\eta_{20°C}$ = 1317 cp),[16, 53] respectively. They can be mixed in all ratios to control sample viscosity which can be calculated based on the weight fraction and viscosity values of its pure components.[54]

Clearly, **46** shows no fluorescence intensification upon increasing viscosity, suggesting that **46** does not adopt a twisted excited state upon excitation (Figure 2.3a). This contrasts the significant sensitivity for viscosity observed for **47**. The difference can perhaps be explained by their respective quantum yields in solvents of low viscosity (Table 2.1). Samples in non-viscous, protic solvents such as water and methanol display quantum yields of 0.20 and 0.50 for **46**, but 0.05 and 0.04 for **47**, respectively. The already much higher quantum yields for **46** suggest that rotational relaxation might not be a major non-radiative pathway for this chromophore's electronic decay. The very low initial quantum yields of **47**, however, benefit greatly from curtailment of this non-emissive decay pathway.
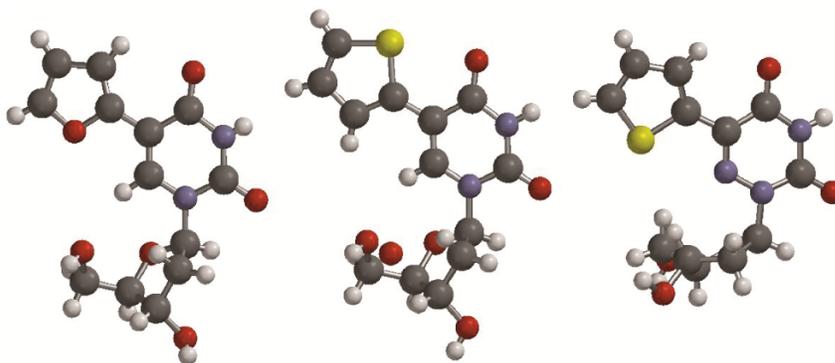
Figure 2.4 Crystal structures of **29** (left),[15] **38** (middle),[15] and the 2'-deoxy-ribose form of **46** (right).[55]

Explaining the distinct differences between the two related chromophores is challenging and must be sought in searching for differences in the relative ground state orientation of the thiophene (or furan) with respect to the nitrogenous

nucleobase. In the reported crystal structure of the deoxy-ribose form of 46, the sulfur atom of the thiophene moiety, which is coplanar with the nucleobase, points away ("*anti*") from the 4-carbonyl thereby facing ("*syn*") the 6-aza nitrogen (Figure 2.4).[55] This orientation is opposite to the one observed in our reported crystal structure of **29**, in which the thiophene moiety, also coplanar with the nucleobase, is oriented with its sulfur atom pointing toward the 4-carbonyl.[2, 15] The crystal structure of **38** shows again the opposite configuration, with the furan oxygen facing away from the 4-carbonyl (Figure 2.4).[2] In lieu of a crystal structure of **47**, we can only speculate on the orientation of its furan oxygen. While differences between solid and solution states could exist, the distinct orientational preferences of the thiophene and furan units in the crystal structures might also reflect their ground state orientation in solution. The electronic nature of the atom in the 6-position, CH (**29** and **38**) or N (**46** and **47**) seems, therefore, to influence the orientation of the 5-membered heterocycle. This, in turn, could have a profound influence on the excited state geometry and hence sensitivity toward viscosity.

2.5 Conclusion

Introducing 6aza into uridine core has proven to be a successful new design. The photophysical features of this novel class of fluorescent nucleosides has demonstrated that altering electronics by replacing pyrimidine with the

corresponding 1,2,4-triazine core enhanced the charge transfer character and yielded a bathochromic shift.

In summary, the 6-aza-uridine motif, decorated at the 5-position with either a thiophene (**46**) or furan (**47**), is a synthetically accessible pyrimidine analog with highly desirable photophysical properties. The Stokes shift of **46** is very sensitive to polarity. The pH sensitive fluorescence intensity starkly increases upon deprotonation with a pKa value of 6.7–6.9, close to physiological pHs. Most importantly, the fluorescence quantum yield remains sufficient (Φ_{water}(pH 2.55) = 0.13) to excellent ($\Phi_{dioxane}$ = 0.80) under all conditions and is thereby amongst the highest reported for isomorphic nucleoside probes.[1] The robust fluorescence quantum yield in neutral protic solvents (Φ_{water} = 0.20, Φ_{MeOH} = 0.50, and $\Phi_{glycerol}$ = 0.66) suggest minimal population of non-radiative decay pathways, which is arguably the cause for the lack of sensitivity to viscosity. The structurally related nucleoside **47**, displays relatively low quantum yields in protic solvents and sensitivity to viscosity, marked by a steep increasing fluorescence quantum yield with increasing medium viscosity. Interestingly, **47** shows almost identical pH sensitivity to **46**, but a weaker sensitivity to polarity. Despite their differences, both **46** and **47** exhibit more desirable photophysical properties than their reported 2′-deoxyuridine analogs (**29** and **38**), making the 6-aza motif one of the most attractive and promising emissive pyrimidines.

2.6 Experimental

Materials

Furanglyoxylic acid, *N,O*-bis(trimethylsilyl)acetamide (BSA) and dry acetonitrile were purchased from Sigma-Aldrich. Thiopheneglyoxylic acid was purchased from Alfa-Aesar. Trimethylsilyl-O-triflate was purchased from Acros. Semicarbazide hydrochloride and 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose were purchased from TCI. All reagents and solvents were used without further purification. Moister and oxygen sensitive reactions were performed in an inert argon atmosphere.
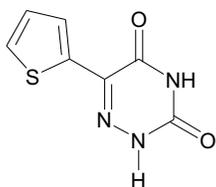
Synthesis

Synthesis of 5-(thiophen-2-yl)-6-aza uracil (61). A mixture of 2-thienylglyoxylic acid (**56**) (251 mg, 1.61 mmol) and semicarbazide hydrochloride (**58**) (210 mg, 1.88 mmol) in water (2.5 mL) was heated at 60°C until its off-white color changed to light green (~18 h). Keeping the reaction mixture at 60°C, a 1M NaOH aq. solution (6.7mL) was added turning the mixture into a light yellow solution. After 4 h, the solution was allowed to cool down to rt and acidified with 2M HCl aq. to reach a pH of 4. Recrystalization from water gave **61** as an off-white solid (232 mg, 1.19 mmol, 74%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.42 (br s, 1H),

12.17 (br, s, 1H), 7.93 (dd, J = 3.7 Hz, 1.1 Hz, 1H), 7.65 (dd, J = 5.1 Hz, 1.1 Hz, 1H), 7.13 (dd, J = 5.1 Hz, 3.7 Hz, 1H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 156.96, 149.01, 137.25, 135.04, 128.78, 128.29, 127.74; MS-ESI: m/z calcd. for C$_7$H$_5$N$_3$O$_2$S (M-H)$^-$: 194.00, found 194.13.
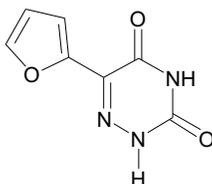
Synthesis of 5-(furan-2-yl)-6-aza uracil (62). A solution of (2-furyl)-glyoxylic acid (57) (251 mg, 1.88 mmol) in water (2 mL) was added dropwise to a solution of semicarbazide hydrochloride (**58**) (202 mg, 1.81 mmol) in water (0.5 mL) at rt. After 15 min. an off-white precipitate was formed and a 1M NaOH aq. (5.9 mL) was added. The reaction mixture was heated at 60°C for 42 h. The solution was allowed to cool down to rt and acidified with 2M HCl aq. to reach a pH of 4. After recrystalization from water, **62** was obtained as an off-white solid (98 mg, 0.55 mmol, 30%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.43 (br s, 1H), 12.13 (br s, 1H), 7.86 – 7.80 (m, 1H), 7.29-7.25 (m, 1H), 6.64-6.59 (m, 1H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 155.49, 149.03, 145.93, 144.64, 133.82, 114.25, 112.08; MS-ESI: m/z calcd. for C$_7$H$_5$N$_3$O$_3$ (M-H)$^-$: 178.03, found: 178.18.
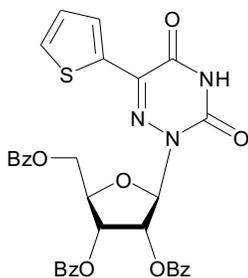
Synthesis of 5-(thiophene-2-yl)-6-aza-2',3',5'-tribenzoyl-uridine (64). Under argon, *N,O*-bis(trimethylsilyl)acetamide (0.76 mL, 3.0 mmol) was added to a stirring suspension of **61** (200 mg, 1.03 mmol) in dry acetonitrile (3.78 mL). The suspension turned into a clear solution after stirring at rt for 30 min.. The solution was heated to 85°C followed by addition of 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (569 mg, 1.1 mmol) and trimethylsilyl trifluoromethanesulfonate (0.2 mL, 1.1 mmol). The reaction mixture was stirred at 85°C under argon until monitoring by TLC (silica, 10 *v*% MeOH in DCM) indicated full conversion of starting material (~30 min.). The reaction mixture was cooled to rt diluted with EtOAc (15 mL) and washed with water (3 × 5 mL), and brine (5 mL). The combined organic layers were dried over Na_2SO_4, filtered over a glass frit and concentrated to dryness. Purification of the crude residue by column chromatography (silica, 1 *v*% MeOH in DCM) yielded **64** as an off-white foam (452 mg, 0.71 mmol, 69%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.56 (s, 1H), 8.10 – 7.76 (m, 6H), 7.72 – 7.18 (m, 12H), 6.50 (d, *J* = 1.9 Hz, 1H), 6.15 – 6.03 (m, 2H), 4.87 – 4.79 (m, 1H), 4.72 – 4.65 (m, 1H), 4.58 – 4.50 (m, 1H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 165.40, 164.66, 164.62, 155.41, 147.87, 138.43, 133.43, 129.50, 129.40, 129.31, 129.17, 129.06, 128.86, 128.68, 128.54, 128.51, 127.90, 87.85, 78.34, 74.21, 70.77,

63.41; MS-ESI: m/z calcd. for $C_{33}H_{25}N_3O_9S$ $(M+H)^+$: 640.14, and $C_{33}H_{25}N_3O_9S$ $(M+Na)^+$: 662.12; found: 640.09, and 662.18, respectively.
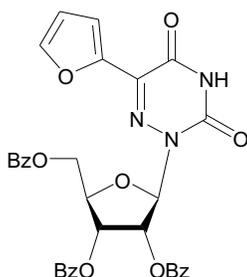
Synthesis of 5-(furan-2-yl)-6-aza-2',3',5'-tribenzoyl-uridine (65). Using the procedure described for **65**, but starting from **62** (0.101 g, 0.56 mmol), **65** was obtained as a white solid foam (300 mg, 0.48 mmol, 85%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.52 (s, 1H), 8.00 – 7.78 (m, 7H), 7.71 – 7.28 (m, 11H), 6.70 – 6.67 (m, 1H), 6.52 (d, J = 1.9 Hz, 1H), 6.11 – 6.04 (m, 2H), 4.85 – 4.77 (m, 1H), 4.74 – 4.67 (m, 1H), 4.61 – 4.54 (m, 1H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 165.42, 164.62, 164.55, 154.72, 147.79, 145.43, 145.22, 133.98, 133.82, 133.44, 129.49, 129.18, 128.69, 127.90, 115.80, 112.28, 88.28, 78.79, 74.27, 70.99, 63.68, 64.07; MS-ESI: m/z calcd. for $C_{33}H_{25}N_3O_{10}$ $(M+H)^+$: 624.16, and $C_{33}H_{25}N_3O_{10}$ $(M+Na)^+$: 646.14; found: $(M+Na)^+$ 646.14
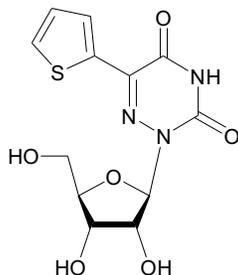
Synthesis of 5-(thiophen-2-yl)-6-aza-uridine (46). A solution of **64** (99 mg, 0.15 mmol) in saturated methanolic ammonia (10 mL) was heated at 60 °C in a pressure vessel for 24h. The mixture was cooled to rt, concentrated and the residue was triturated with DCM (4 × 5 mL) to give **46** as a white solid (41mg, 0.12 mmol, 80%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.41 (s, 1H), 7.98 – 7.93 (m, 1H), 7.74 – 7.69 (m, 1H), 7.19 – 7.11 (m, 1H), 5.98 (d, J = 2.6Hz, 1H), 5.33 (d, J = 4.8 Hz, 1H), 5.07 (d, J = 6.1 Hz, 1H), 4.66 (t, J = 5.8Hz, 1H), 4.30 – 4.12 (m, 2H), 3.87 – 3.78 (m, 1H), 3.64 – 3.54 (m, 1H), 3.50 – 3.39 (m, 1H); ^{13}C NMR (100 MHz, DMSO-d$_6$) δ 155.45, 148.36, 137.62, 134.69, 129.61, 129.11, 127.92, 89.65, 84.88, 73.26, 70.63, 62.32; MS-ESI: m/z calcd. for $C_{12}H_{13}N_3O_6S$ (M-H)$^-$ 326.04, found: 327.16.
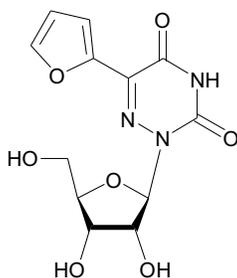
Synthesis of 5-(furan-2-yl)-6-aza-uridine (47). Using the procedure described for

46, but starting from **65** (88 mg, 0.14 mmol)**, 47** was obtained as a white solid (35

mg, 0.11 mmol, 79%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 7.90 – 7.87 (m, 1H), 7.36

– 7.33 (m, 1H), 6.66 – 6.63 (m, 1H), 6.00 (d, J = 3.2 Hz, 1H), 5.30 (d, J = 5.0 Hz,

1H), 5.03 (d, J = 6.3 Hz, 1H), 4.62 (t, J = 5.8 Hz, 1H), 4.30 – 4.25 (m, 1H), 4.19 –

4.12 (m, 1H), 3.86 – 3.79 (m, 1H), 3.65 – 3.51 (m, 1H), 3.52 – 3.40 (m, 1H); ^{13}C

NMR (100 MHz, DMSO-d$_6$) δ 154.74, 148.17, 145.46, 145.14, 133.85, 115.17,

112.16, 89.78, 84.90, 72.99, 70.58, 62.33; MS-ESI: m/z calcd. for $C_{12}H_{13}N_3O_7$ (M-

H)$^-$ 310.07, found 310.18

Photophysics

All solvents used were of spectroscopic grade and obtained from Sigma Aldrich. Spectroscopic grade glycerol was obtained from Acros. For all spectroscopic measurements a 1 cm four-sided Helma quartz cuvette was used. All spectroscopy samples were prepared from concentrated DMSO stock solutions, hence, all samples contain 0.4 v% DMSO.

Absorption spectra were measured on a Shimadzu UV-2450, UV-Vis spectrophotometer and corrected for the blank. The sample temperature was kept constant at 20 °C using a thermostat controlled ethylene glycol−water bath fitted to a specially designed cell holder.

Steady state emission and excitation spectra were taken on a PTI luminescence spectrometer using a 0.5 nm and 1 nm resolution, respectively. The sample temperature was kept constant at 20 °C with a Quantum Northwest TLC50 fluorescence cuvette holder in conjunction with a software controllable TC 125 temperature controller.

The polarity dependent steady state fluorescence studies were performed using an excitation wavelength of 335 nm (slit-widths: 0.35 mm = 1.4 nm), and 327 nm (slit-widths of 0.45 mm = 1.8 nm), for **46** and **47** respectively. Sample concentrations of 7.9 μM were used for **46** and **47**. Stokes shifts, in cm^{-1}, were calculated after correction of the emission intensity according the following equation: Intensity [ν] = λ^2 × Intensity [λ].[19] The cm^{-1} units were converted to kcal/mol by multiplication with 0.0028591 to plot the polarity sensitivity plot.

Sample $E_T(30)$ values (in kcal/mol) were determined by dissolving a small amount of Reichardt's dye in the solvent(mixture) used to prepare the sample.[20] The observed long–wavelength absorption maximum in nm (λ_{absmax}) was converted to the sample $E_T(30)$ value according the following equation:

Due to solubility limitations of Reichardt's dye, the literature $E_T(30)$ value for pure water was used.[20]

The pH dependent steady state fluorescence studies were performed using an excitation wavelength of 335 nm (slit-widths: 0.35 mm = 1.4 nm), and 327 nm (slit-widths of 0.8 mm = 3.2 nm), for **46** and **47** respectively. The pH studies were performed in aqueous 10 mM phosphate buffer containing 100 mM NaCl. The pH of each sample was determined prior to use. Sample concentrations of 7.9 μM were used for **46** and **47**. The normalized PL intensity vs. pH was fitted using a Boltzmann sigmoidal curve using OriginPro 8. The pK_a value was determined by interpolation of the fitted curve.

$$E_T(30) = \frac{28591}{\lambda_{abs\,max}}$$

The viscosity dependent steady state fluorescence studies were performed using an excitation wavelength of 340 nm (slit-widths: 0.5 mm = 2.0 nm), and 330 nm (slit-widths of 0.55 mm = 2.2 nm), for **46** and **47** respectively. To prevent photobleaching by the extended irradiation at high viscosities, a smaller window around the emission maximum was recorded. Sample concentrations of 7.9 μM were used for **46** and **47**. Samples of different viscosity were prepared by mixing

a methanol stock solution with a glycerol stock solution, both prepared from DMSO stock solutions, in different ratios. The highly viscous glycerol stock solution was prepared by first filling the volumetric flask halfway with glycerol followed by addition of a calculated amount of the DMSO sample stock solution using a syringe upon which the flask was properly filled with the viscous glycerol and capped. Subsequently, the volumetric flask, was partly submerged in an oil bath and warmed to 60–70 °C for ~30' occasionally rotating the flask to ensure mixing. A sufficient volume was transferred to the cuvette using a glass pipette. The cuvette was kept at ~60 °C for ~10' to obtain a perfectly transparent homogeneous sample solution followed by cooling to 20 °C prior to the measurements. The sample viscosity of mixtures was calculated based on the literature viscosity for the pure glycerol and methanol and their mixing ratio using the following equation:[54]

$$\ln h_{mix} = \sum_{i=1}^{2} w_i \cdot \ln h_i$$

Here η_{mix} and η_i stand for the viscosity of the mixture and the viscosity of component i respectively. Factor w_i stands for the weight fraction of component i. Solvent weight fractions can be calculated from the density of the solvent and the volume used. Ratios of methanol-glycerol were chosen to give an even distribution of data points over the entire viscosity window used. The viscosities of the binary methanol-glycerol mixtures used were calculated using the equation above (Table 2.3).

Table 2.3 Calculated viscosity of methanol-glycerol mixtures

Volume ratio methanol : glycerol	Viscosity [cp]
1 : 0	0.583
0.7 : 0.3	9.849
0.5 : 0.5	49.055
0.15 : 0.85	538.414
0 : 1	1317

$$\Phi_s = \frac{I_s}{I_{ref}} \cdot \frac{O.D._{ref}}{O.D._s} \cdot \left(\frac{n_s}{n_{ref}}\right)^2 \cdot \Phi_{ref}$$

Quantum yields were determined using Coumarin-1 in ethanol (Φ = 0.73) as a standard, using the following equation:

Here Φ, I, O.D. and n stand for quantum yield, integrated emission intensity, optical density at λ_{ex} and refractive index (n_{water} = 1.333, $n_{dioxane}$ = 1.422, n_{EtOH} = 1.36), respectively. Sample and reference are denoted by s and ref, respectively. The O.D. at λ_{ex} is kept below 0.05 to avoid inner filter effects. The λ_{ex} is the same for the sample and reference solutions to circumvent correction of the difference in excitation energy at different wavelengths. Since in glycerol only the top of the emission curve is measured, the quantum yield for the samples in pure glycerol were calculated based on the O.D. corrected fluorescence intensity observed in glycerol ($I_{glycerol}$) relative to the fluorescence intensity in methanol ($I_{methanol}$), and the averaged determined quantum yield ($\Phi_{methanol}$) therein, taking the difference of the refractive index of methanol ($n_{methanol}$ = 1.329) and glycerol ($n_{glycerol}$ = 1.4746) into account according the following equation:

$$\Phi_{glycerol} = \frac{I_{glycerol}}{I_{methanol}} \cdot \left(\frac{n_{glycerol}}{n_{methanol}}\right)^2 \cdot \Phi_{methanol}$$

Fluorescence decay curves were recorded on a PTI luminescence spectrometer after excitation at 340 nm using an LED laser. The sample temperature was kept constant at 20 °C with a Quantum Northwest TLC50 fluorescence cuvette holder in conjunction with a software controllable TC 125 temperature controller. Fluorescence lifetime calculations were performed using Felix32 corrected and corrected for the instrument response factor (IRF).'

Sensitivity to polarity

46: water-dioxane absorption and steady-state emission spectroscopy
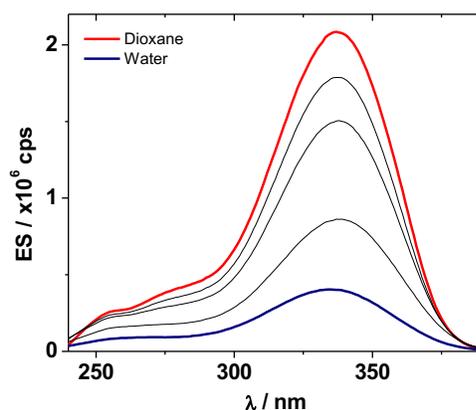
Figure 2.5 Excitation spectra of **46** in dioxane (red line), water (blue line), and mixtures thereof (black lines).

Table 2.4 $E_T(30)$ data for water-dioxane mixtures[a]

Water content	Abs-1	Abs-2	$E_T(30)$-1	$E_T(30)$-2	$E_T(30)$Avg	$E_T(30)$SEM	$E_T(30)$err
v%	Nm	nm	kcal/mol	kcal/mol	kcal/mol	kcal/mol	%
0	779	785	36.7	36.4	36.6	0.14	0.38
10	629	619	45.5	46.2	45.8	0.37	0.80
30	568	568	50.3	50.3	50.3	0	0
70	500	500	57.2	57.2	57.2	0	0
100	453	453	63.1	63.1	63.1	0	0

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro,
err=SEM/Avg*100.

Table 2.5 Absorption and emission (PL) maxima of **46** in water–dioxane mixtures[a]

Water Content	Abs-1	Abs-2	AbsAvg	AbsSEM	Abserr
v%	nm	Nm	nm	nm	%
0	336.0	334.5	335.25	0.75	0.22
10	334.0	335	334.5	0.5	0.15
30	336.5	336	336.25	0.25	0.07
70	335.5	334	334.75	0.75	0.22
100	333.0	331	332	1	0.30
Water Content	PL-1	PL-2	PLAvg	PLSEM	PLerr
v%	nm	nm	nm	nm	%
0	415.1	414.1	414.6	0.5	0.12
10	427.6	425.1	426.35	1.25	0.29
30	437.6	436.6	437.1	0.5	0.11
70	453.1	451.1	452.1	1	0.22
100	458.1	452.1	455.1	3	0.66

[a] Avg=average, SEM= standard error of mean as calculated
by OriginPro, err=SEM/Avg*100.

Table 2.6 Fluorescence quantum yield (QY) data for **46** in water and dioxane[a]

Solvent	QY-1	QY-2	QY-3	QY-4	QY-5	QY-6	QYAvg	QYSEM	QYerr
									%
Dioxane	0.77	0.78	0.83	0.79	0.79	0.85	0.80	0.013	1.6
Water	0.20	0.19	0.21	0.20	0.20	0.21	0.20	0.003	1.2

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro,
err=SEM/Avg*100.

Table 2.7 Stokes shift (SS) data for samples of **46** in water–dioxane mixtures[a]

Water Content	SS-1	SS-2	SSAvg	SSSEM	SSerr
v%	cm^{-1}	cm^{-1}	cm^{-1}	cm^{-1}	%
0	5986	6063	6025	38.40	0.64
10	6851	6788	6819	31.37	0.46
30	7226	7117	7171	54.29	0.76
70	7881	7991	7936	54.90	0.69
100	84364	8548	8492	55.64	0.66
Water Content	**SS-1**	**SS-2**	**SSAvg**	**SSSEM**	**SSerr**
v%	kcal/mol	kcal/mol	kcal/mol	kcal/mol	%
0	17.12	17.34	17.23	0.11	0.64
10	19.59	19.41	19.50	0.09	0.46
30	20.66	20.35	20.50	0.16	0.76
70	22.53	22.85	22.69	0.16	0.69
100	24.12	24.44	24.28	0.16	0.66

[a] Avg=average, SEM= standard error of mean as calculated by
OriginPro, err=SEM/Avg*100. N.B. Stokes shift are calculated
using corrected emission maxima

47: water-dioxane absorption and steady-state emission spectroscopy
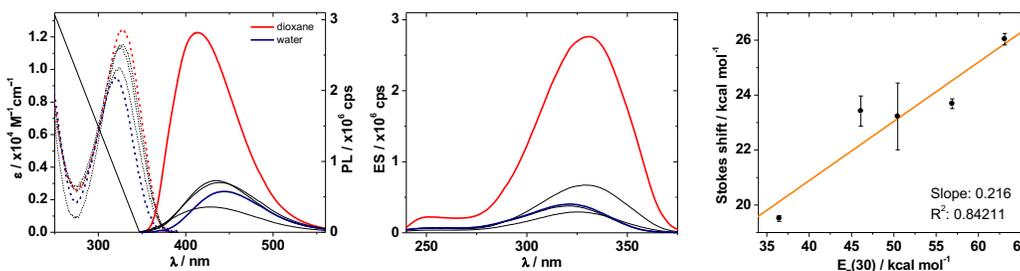
Figure 2.6 *Left:* Absorption (dashed line) and emission (solid line) of **47** in dioxane (red), water (blue) and 10v%, 30v%, and 70v% water in dioxane mixtures (black lines). *Middle*: Excitation spectra of **47** in the same mixtures. *Right*: Stokes shift (in kcal/mol) as a function of the sample $E_T(30)$ value (solid circles), with a linear fit (orange line).

Table 2.8 Absorption and emission (PL) maxima for samples of **47** in water–dioxane mixtures[a]

Water Content	Abs-1	Abs-2	AbsAvg	AbsSEM	Abserr
v%	nm	Nm	nm	nm	%
0	327.5	327.0	327.3	0.25	0.08
10	328.5	326.0	327.3	1.25	0.38
30	326.5	324.0	325.3	1.25	0.38
70	324.5	324.0	324.3	0.25	0.08
100	318.5	321.0	319.8	1.25	0.39
Water Content	PL-1	PL-2	PLAvg	PLSEM	PLerr
v%	nm	nm	nm	nm	%
0	413.6	415.0	414.3	0.70	0.17
10	438.1	437.0	437.6	0.55	0.13
30	427.1	438.0	432.6	5.45	1.26
70	433.1	436.0	434.6	1.45	0.33
100	443.6	443.0	443.3	0.30	0.07

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100.

Table 2.9 Fluorescence quantum yields (QY) for **47** in water and dioxane[a]

Solvent	QY-1	QY-2	QY-3	QY-4	QY-5	QYAvg	QYSEM	QYerr
								%
Dioxane	0.66	0.69	0.70	0.63	0.54	0.64	0.029	4.50
Water	0.00	0.01	0.01	0.01	0.01	0.01	0.000	5.52

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100.

Table 2.10 Stokes shift (SS) data for samples of **47** in water–dioxane mixtures[a]

Water Content	SS-1	SS-2	SSAvg	SSSEM	SSerr
v%	cm^{-1}	cm^{-1}	cm^{-1}	cm^{-1}	%
0	6787	6862	6824	37	0.55
10	8000	8383	8192	192	2.34
30	7697	8548	8123	425	5.23
70	8223	8347	8285	62	0.75
100	9180	9034	9107	73	0.80
Water Content	**SS-1**	**SS-2**	**SSAvg**	**SSSEM**	**SSerr**
v%	kcal/mol	kcal/mol	kcal/mol	kcal/mol	%
0	19.40	19.62	19.51	0.11	0.55
10	22.87	23.97	23.42	0.55	2.34
30	22.01	24.44	23.22	1.22	5.23
70	23.51	23.86	23.69	0.18	0.75
100	26.25	25.83	26.04	0.21	0.80

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100. N.B. Stokes shift are calculated using corrected emission maxima

Correlations for 5-(thiophen-2-yl)-2dU and 5-(furan-2-y)-2dU

Sensitivity toward polarity for 5-(thiophen-2-yl)-2dU and 5-(furan-2-yl)-2dU plotted with Stokes shifts in cm^{-1} has been reported.[16] The Stokes shifts in cm^{-1} were converted to kcal/mol values and plotted as a function of the sample $E_T(30)$ value. Now, after linearization, the slope, representing the polarity sensitivity is a unitless value.
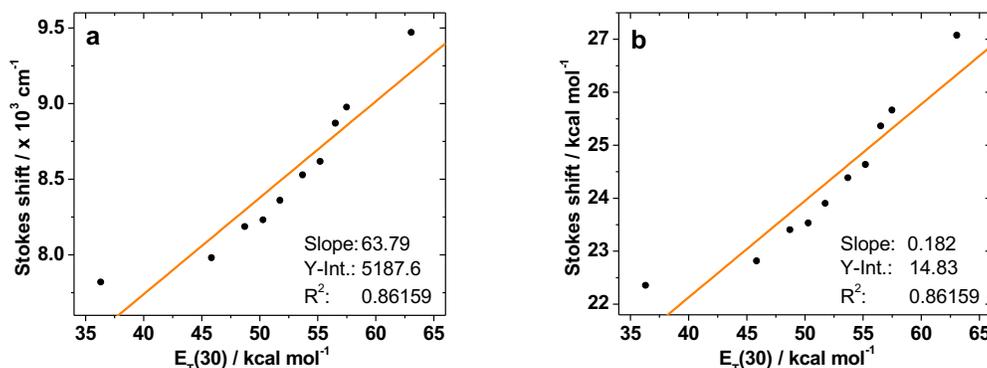
Figure 2.7 A plot of the Stokes shift of 5-(thiophen-2-yl)-2dU as a function of sample $E_T(30)$ value with a) Stokes shift in cm^{-1} and b) Stokes shift in kcal/mol. Orange lines represent linearization of the data points (solid circles).
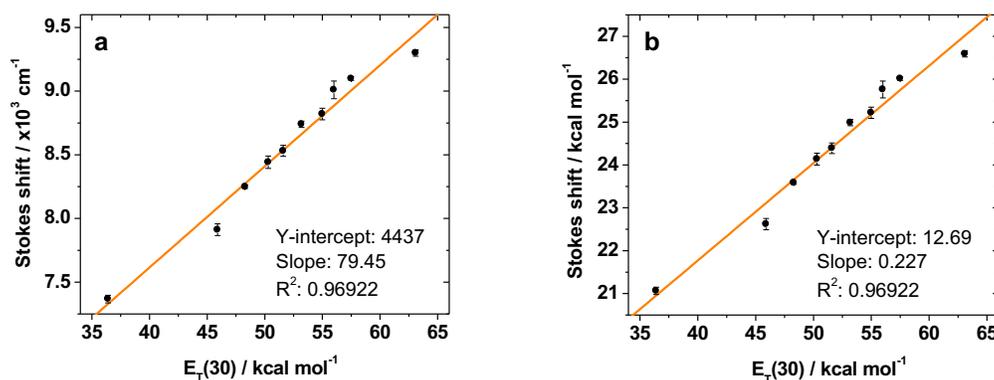
Figure 2.8 A plot of the Stokes shift of 5-(furan-2-yl)-2dU as a function of sample $E_T(30)$ value with a) Stokes shift in cm^{-1} and b) Stokes shift in kcal/mol. Orange lines represent linearization of the data points (solid circles). Note that the value for the slope in a) is slightly lower than the reported 81.85 cm^{-1}/kcal mol^{-1} due to a different error treatment used for the linear fit.[16]

Sensitivity to pH

46: absorption and steady-state emission spectroscopy
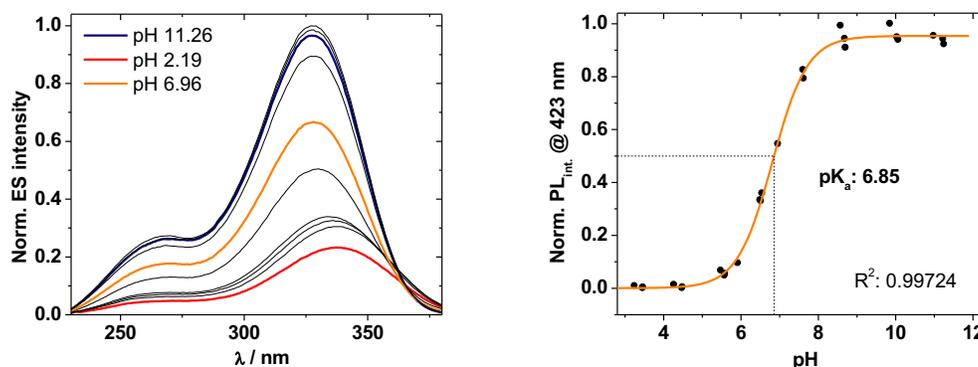
Figure 2.9 *Left:* Selected excitation curves for aqueous buffered samples of **46** at pH 2.19 (red line), 6.96 (orange line), pH 11.26 and intermediate pH values (black lines). Excitation spectra were recorded monitoring at the sample's emission maximum. Hence the intensity reflects the emission intensity for the corresponding samples. *Right:* a plot of the normalized fluorescent intensity of **46** (solid circles), excluding the samples with a pH<2.5, and a sigmoidal fit (orange line).

47: absorption and steady-state emission spectroscopy
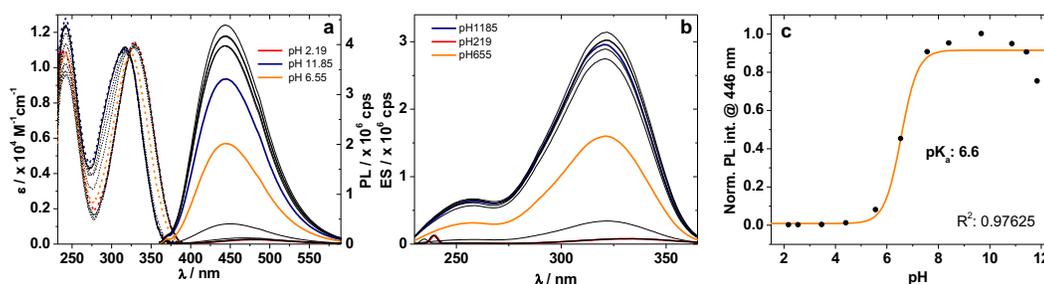
Figure 2.10 a) Absorption (dashed lines) and emission (solid lines) spectra and b) excitation curves of **47** in aqueous buffers of pH 2.19 (red line), pH 6.55 (orange line), pH 11.85 (blue line), and intermediate pH values (black lines) Excitation spectra were recorded monitoring at the sample's emission maximum. Hence, the intensity reflects the emission intensity for the corresponding samples. c) A plot the normalized fluorescence intensity vs the sample pH value (solid circles), with a sigmoidal fit (orange line) using OriginPro. Interpolation gives a pK$_a$ value (pH at y=0.5) for **47** of 6.6. The R^2 value describes the goodness of the fit.

Sensitivity to viscosity

46: absorption and steady-state emission spectra in methanol–glycerol
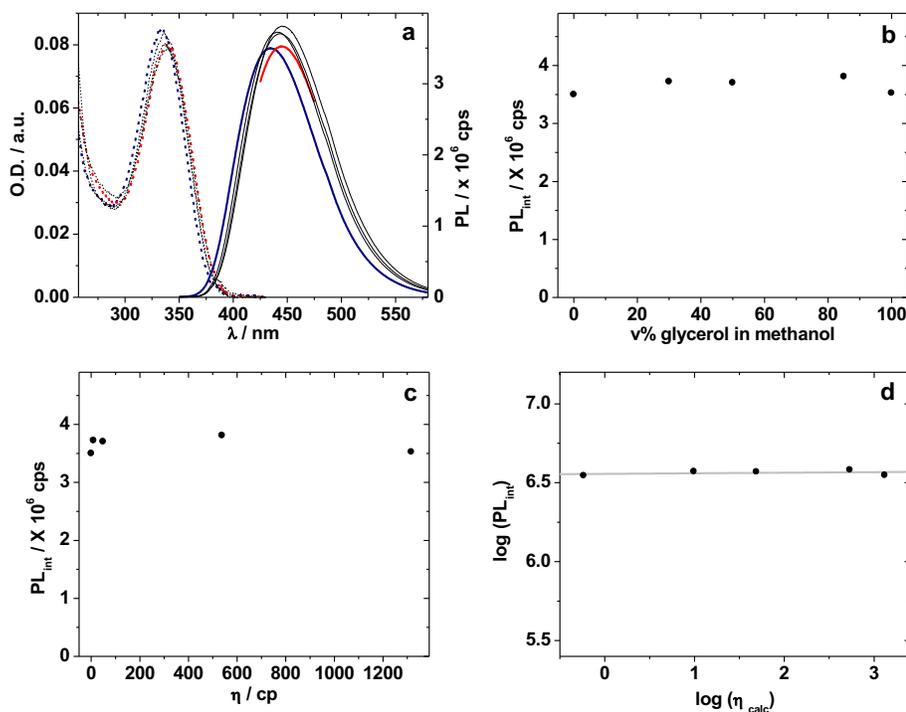
Figure 2.11 a) Absorption (dashed lines) and emission (solid lines) spectra of **46** in methanol (blue line), glycerol (red line), and binary mixtures thereof (black lines); b) correlation of fluorescent intensity vs. *v*% glycerol in methanol; c) plot of fluorescent intensity vs. sample viscosity, and d) correlation of the log fluorescence intensity as a function of the log viscosity (solid circles), with a linear fit (grey line) (the Förster–Hoffman plot).[34] This experiment, once performed, indicated that **46** is not responsive to changes in viscosity.

Table 2.11 Fluorescence quantum yields (QY) for **46** in methanol and glycerol[a]

Solvent	QY-1	QY-2	QYAvg	QYSEM	QYerr
					%
Methanol	0.49	0.52	0.50	0.016	3.1
Glycerol	0.62	0.69	0.66	0.034	5.2

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100.

47: absorption and steady-state spectra in methanol–glycerol
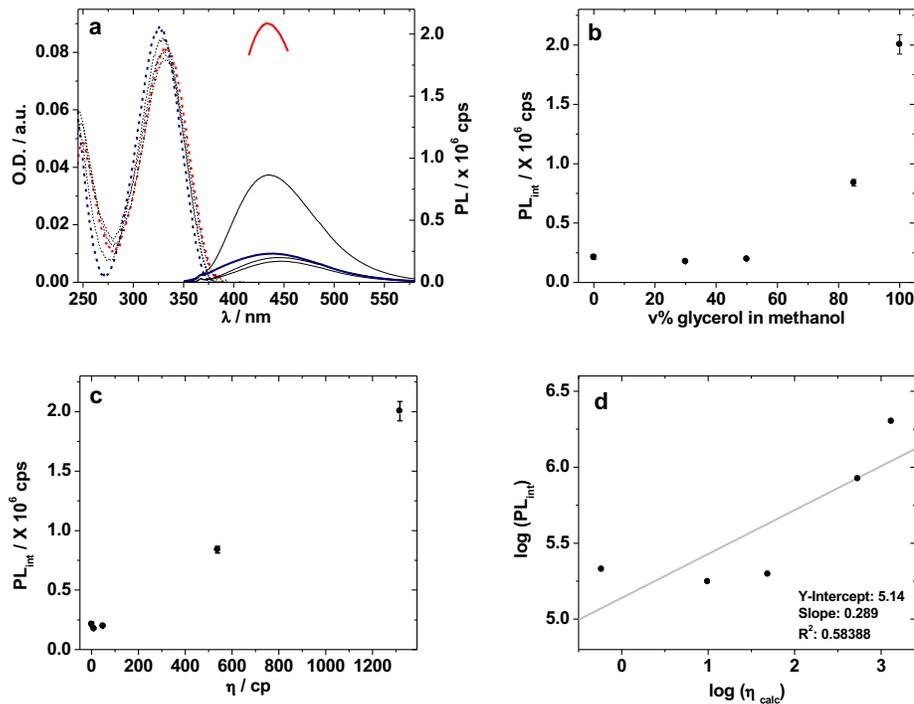
Figure 2.12 a) Absorption (dashed lines) and emission (solid lines) spectra of **47** in methanol (blue line), glycerol (red line), and binary mixtures thereof (black lines); b) correlation of fluorescent intensity vs. *v*% glycerol in methanol; c) plot of fluorescent intensity vs. sample viscosity, and d) correlation of the log fluorescence intensity as a function of the log viscosity (solid circles), with a linear fit (grey line) (the Förster–Hoffman plot). This experiment, performed in duplicate (see error bars in graphs b and c) shows that **47** has a strong but non-linear sensitivity toward viscosity, indicating that, besides viscosity other parameters (e.g. polarity) distort the viscosity measurement.

Table 2.12 Absorption and emission (PL) maxima for samples of **47** in

methanol–glycerol mixtures[a]

Glycerol Content	Abs-1	Abs-2	AbsAvg	AbsSEM	Abserr
v%	nm	Nm	nm	nm	%
0	324.5	326.5	325.5	1.00	0.31
10	327.5	329.0	328.3	0.75	0.23
30	330.5	331.5	331.0	0.50	0.15
70	332.0	330.0	331.0	1.00	0.30
100	331.5	330.5	331.0	0.50	0.15
Glycerol Content	PL-1	PL-2	PLAvg	PLSEM	PLerr
v%	nm	nm	nm	nm	%
0	446.0	438.0	442.0	4.00	0.90
10	441.0	446.0	443.5	2.50	0.56
30	443.5	444.5	444.0	0.50	0.11
70	435.0	434.0	434.5	0.50	0.12
100	435.0	433.0	434.0	1.00	0.23

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100.

Table 2.13 Fluorescence quantum yields (QY) for **47** in methanol and glycerol[a]

Solvent	QY-1	QY-2	QYAvg	QYSEM	QYerr
					%
Methanol	0.03	0.04	0.04	0.004	10.17
Glycerol	0.49	0.44	0.46	0.025	5.38

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100.

Fluorescence lifetime studies

46: Time resolved measurements
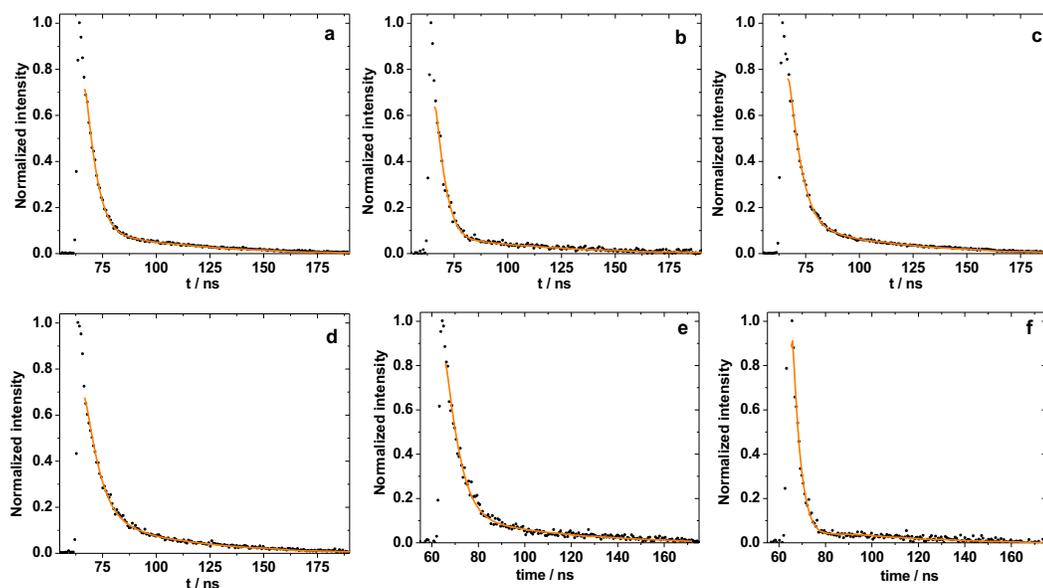
Figure 2.13 Fluorescence lifetime decay curves (black circles) and a mono-exponential fit generated in Felix32 (orange line) of **46** in a) dioxane, b) water, c) methanol, d) glycerol, e) aqueous buffer at pH 10.13, and f) aqueous buffer at pH 2.55. Lifetimes in dioxane, water, methanol, and glycerol have been determined in triplicate. Lifetimes in buffers are measured in duplicate. One decay curve for each solvent is shown as an example.

Table 2.14 Fluorescence lifetimes (LT) for **46** in selected solvents[a]

Solvent	LT-1	LT-2	LT-3	LTAvg	LTSEM	LTerr
	ns	ns	ns	ns	ns	%
Dioxane	5.5	5.4	5.5	5.4	0.034	0.6
Water	4.7	4.8	5.1	4.9	0.124	2.6
Methanol	6.4	6.8	6.7	6.6	0.131	2.0
Glycerol	8.2	7.8	7.8	7.9	0.116	1.5
aq. pH 10.13	6.8	6.5	-	6.6	0.126	1.9
aq. pH 2.55	3.0	3.0	-	3.0	0.021	0.7

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro, err=SEM/Avg*100.

47: Time resolved measurements
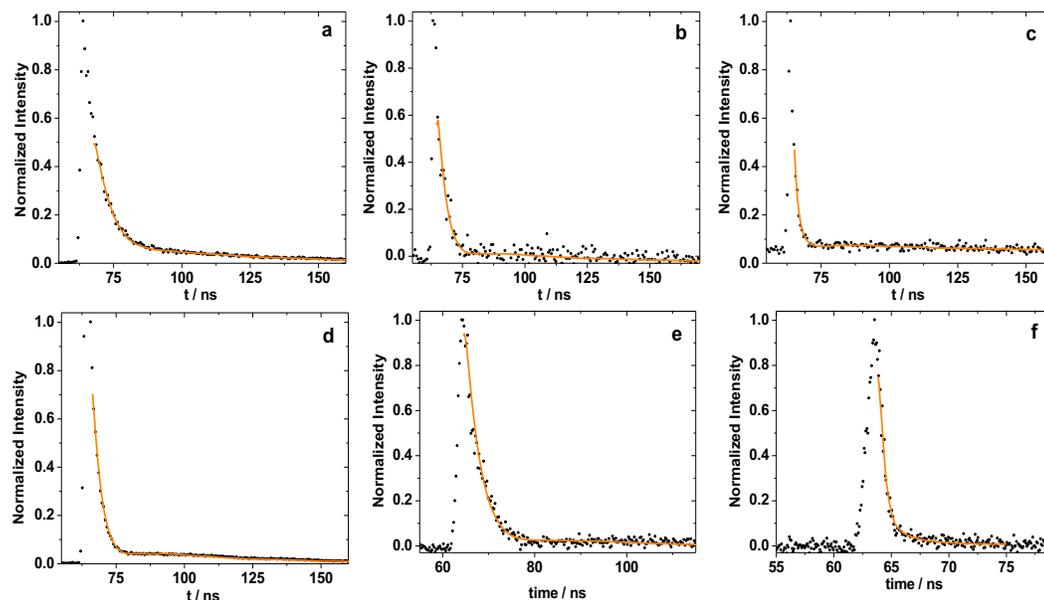
Figure 2.14 Fluorescence lifetime decay curves (black circles) and a mono-exponential fit generated in Felix32 (orange line) of **47** in a) dioxane, b) water, c) methanol, d) glycerol, e) aqueous buffer at pH 10.13, and f) aqueous buffer at pH 2.55. Lifetimes in dioxane, water, methanol, and glycerol have been determined in triplicate. Lifetimes in buffers are measured in duplicate. One decay curve for each solvent is shown as an example.

Table 2.15 Fluorescence lifetimes (LT) for **47** in selected solvents.[a]

Solvent	LT-1	LT-2	LT-3	LTAvg	LTSEM	LTerr
	ns	ns	ns	ns	ns	%
Dioxane	5.7	5.5	5.7	5.6	0.061	1.1
Water	2.8	2.6	4.0	3.1	0.431	13.7
Methanol	1.3	1.9	1.7	1.6	0.194	11.9
Glycerol	2.8	2.6	2.7	2.7	0.043	1.6018
aq. pH 10.13	2.841	2.702	-	2.7715	0.0695	2.50767
aq. pH 2.55	0.1278	0.1092	-	0.1185	0.0093	7.8481

[a] Avg=average, SEM= standard error of mean as calculated by OriginPro,

err=SEM/Avg*100.

Acknowledgment

Chapter 2 has been previously published in the following article: Sinkeldam, R. W.; Hopkins, P. A.; Tor, Y. "Modified 6-Aza Uridines: Highly Emissive pH-Sensitive Fluorescent Nucleosides" *ChemPhysChem* 2012, *13*, 3350. The dissertation author is one of the primary investigators and authors of this article.

2.7 References

1.　　R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

2.　　N. J. Greco and Y. Tor, *J Am Chem Soc*, 2005, **127**, 10784–10785.

3.　　D. Shin, R. W. Sinkeldam and Y. Tor, *J Am Chem Soc*, 2011, **133**, 14912-14915.

4.　　M. E. Hawkins, *Fluorescence Spectroscopy*, 2008, **450**, 201-231.

5.　　D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378-391.

6.　　Y. Tor, *Pure Appl Chem*, 2009, **81**, 263-272.

7.　　L. M. Wilhelmsson, *Quarterly reviews of biophysics*, 2010, **43**, 159-183.

8.　　M. Kimoto, R. S. I. Cox and I. Hirao, *Expert Review of Molecular Diagnostics*, 2011, **11**, 321-331.

9.　　D. C. Ward, E. Reich and L. Stryer, *J Biol Chem*, 1969, **244**, 1228-1237.

10.　　C. H. Liu and C. T. Martin, *J Mol Biol*, 2001, **308**, 465-475.

11.　　Y. Tor, S. Del Valle, D. Jaramillo, S. G. Srivatsan, A. Rios and H. Weizman, *Tetrahedron*, 2007, **63**, 3608-3614.

12.　　S. G. Srivatsan, N. J. Greco and Y. Tor, *Angew Chem Int Edit*, 2008, **47**, 6661-6665.

13.　　Y. Xie, A. V. Dix and Y. Tor, *J Am Chem Soc*, 2009, **131**, 17605-17614.

14. Y. Xie, T. Maxson and Y. Tor, *J Am Chem Soc*, 2010, **132**, 11896-11897.

15. N. J. Greco and Y. Tor, *Tetrahedron*, 2007, **63**, 3515-3527.

16. R. W. Sinkeldam, A. J. Wheat, H. Boyaci and Y. Tor, *Chemphyschem*, 2011, **12**, 567-570.

17. R. W. Sinkeldam, P. Marcus, D. Uchenik and Y. Tor, *Chemphyschem*, 2011, **12**, 2260-2265.

18. TrivialName, Although a trivial name, we refer to the new nucleosides as 6-aza uridines since their features are compared to their corresponding modified pyrimidines.

19. J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer, New York, 3rd edn., 2006.

20. C. Reichardt, *Chem Rev*, 1994, **94**, 2319–2358.

21. M. Y. Berezin and S. Achilefu, *Chem Rev*, 2010, **110**, 2641-2684.

22. B. C. Lobo and C. J. Abelt, *J Phys Chem A*, 2003, **107**, 10938-10943.

23. G. Oster and Y. Nishijima, *J Am Chem Soc*, 1956, **78**, 1581-1584.

24. R. O. Loutfy and B. A. Arnold, *J Phys Chem-Us*, 1982, **86**, 4205-4211.

25. L. M. Wilhelmsson, P. Sandin, A. Holmen, B. Albinsson, P. Lincoln and B. Norden, *J Phys Chem B*, 2003, **107**, 9094-9101.

26. K. M. Sun, C. K. McLaughlin, D. R. Lantero and R. A. Manderville, *J Am Chem Soc*, 2007, **129**, 1894-1895.

27. R. O. Roblin, J. O. Lampen, J. P. English, Q. P. Cole and J. R. Vaughan, *J Am Chem Soc*, 1945, **67**, 290-294.

28. J. Wierzchowski, B. Wielgus-Kutrowska and D. Shugar, *Biochimica Et Biophysica Acta-General Subjects*, 1996, **1290**, 9-17.

29. C. P. Da Costa, M. J. Fedor and L. G. Scott, *J Am Chem Soc*, 2007, **129**, 3426-3432.

30. L. Liu, J. W. Cottrell, L. G. Scott and M. J. Fedor, *Nat Chem Biol*, 2009, **5**, 351-357.

31. F. Seela and P. Chittepu, *J Org Chem*, 2007, **72**, 4358-4366.

32. R. W. Sinkeldam and Y. Tor, *Org Biomol Chem*, 2007, **5**, 2523–2528.

33. R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chembiochem*, 2008, **9**, 706-709.

34. T. Förster and G. Hoffmann, *Zeitschrift Fur Physikalische Chemie*, 1971, **75**, 63-76.

35. K. Hayes, *J Med Chem*, 1964, **7**, 819-820.

36. I. Basnak, P. L. Coe and R. T. Walker, *Nucleosides and Nucleotides*, 1994, **13**, 163-175.

37. W. L. Mitchell, P. Ravenscroft, M. L. Hill, L. J. S. Knutsen, B. D. Judkins, R. F. Newton and D. I. C. Scopes, *J Med Chem*, 1986, **29**, 809-816.

38. H. A. Burch, *J Med Chem*, 1970, **13**, 288-291.

39. C. Cristescu, *Revue Roumaine De Chimie*, 1975, **20**, 1287-1293.

40. R. Jin and K. J. Breslauer, *P Natl Acad Sci USA*, 1988, **85**, 8939–8942.

41. D. A. Barawkar and K. N. Ganesh, *Biochem Bioph Res Co*, 1994, **203**, 53–58.

42. D. A. Barawkar and K. N. Ganesh, *Nucleic Acids Res*, 1995, **23**, 159–164.

43. T. Kimura, K. Kawai and T. Majima, *Org Lett*, 2005, **7**, 5829–5832.

44. T. Kimura, K. Kawai and T. Majima, *Chem Commun (Camb)*, 2006, 1542–1544.

45. V. R. Jadhav, D. A. Barawkar and K. N. Ganesh, *J Phys Chem B*, 1999, **103**, 7383–7385.

46. A. Okamoto, K. Tainaka and I. Saito, *Bioconjug Chem*, 2005, **16**, 1105–1111.

47. K. Tainaka, K. Tanaka, S. Ikeda, K. Nishiza, T. Unzai, Y. Fujiwara, I. Saito and A. Okamoto, *J Am Chem Soc*, 2007, **129**, 4776-4784.

48. D. Shugar and J. J. Fox, *Biochim Biophys Acta*, 1952, **9**, 199-218.

49. J. J. Fox and D. Shugar, *Biochim Biophys Acta*, 1952, **9**, 369-384.

50. Pfleider.W and H. Deiss, *Isr J Chem*, 1968, **6**, 603-614.

51. M. S. A. Abdelmottaleb, R. O. Loutfy and R. Lapouyade, *J Photoch Photobio A*, 1989, **48**, 87-93.

52. M. A. Haidekker and E. A. Theodorakis, *Org Biomol Chem*, 2007, **5**, 1669-1678.

53. *DIPPR Project 801*, Design Institute for Physical Properties, AIChE, 2009.

54. M. A. Haidekker, T. P. Brady, D. Lichlyter and E. A. Theodorakis, *Bioorg Chem*, 2005, **33**, 415-425.

55. I. Bansak, M. Sun, T. A. Hamor, F. Focher, A. Verri, S. Spadari, B. Wroblowski, P. Herdewijn and R. T. Walker, *Nucleosides and Nucleotides*, 1998, **17**, 187-206.

Chapter 3   Visibly Emissive and Responsive

Extended 6-Aza-Uridines

3.1 Introduction and design

Diverse approaches have been devised for modifying the non-emissive pyrimidine and purine nucleobases in DNA and RNA into fluorescent surrogates.[1-4] The fundamental challenges result from both structural and electronic dilemmas, where any modification aimed at enhancing the electronic features favoring fluorescence can hamper the WC face and its tautomeric preferences, as well as the hybridization and folding features of the resulting oligomers. This issue is particularly challenging when one aims at shifting the emission bands further into the visible and red spectral domains. Such low energy emission is frequently associated with relatively large chromophores with physical footprints, which are much larger than the native nucleobases.[5]

Nature exploits various mechanisms to tune the photophysics of its small and environmentally-sensitive visibly emitting chromophores, such as oxyluciferin[6-8] and arylideneimidazolidones (in fluorescent proteins).[9-11] Strong charge transfer transitions and proton transfer processes in the excited state typically yield low energy emission, which is dependent on the compactness and polarity of the chromophore's environment.[12-15] Applying these motifs to visibly emitting nucleosides presents additional challenges. In particular, the electron withdrawing ability of the native nucleobases needs to be augmented to promote

 effective CT bands.   Diverse efforts have resulted in numerous motifs, most

however, either electronically decouple the native pyrimidine or purine from the

actual chromophore,[16] or significantly alter the native structure resulting in less

than favorable hybridization features.[17-19]
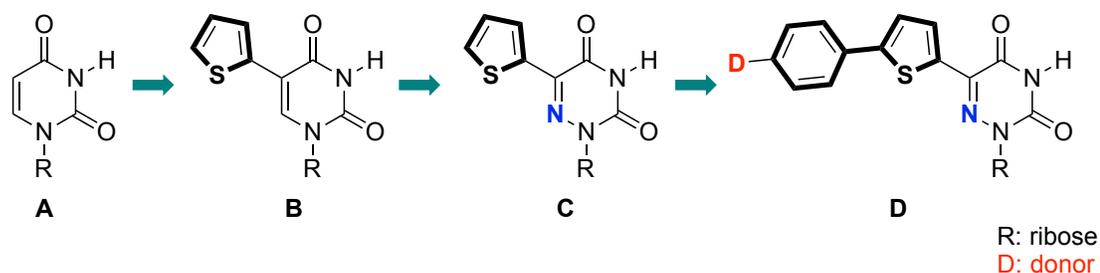
Figure 3.1 Evolution of the design elements leading to visibly emitting nucleosides.

        Here we combine several key features to advance a family of visibly emitting

pyrimidine analogs.   Figure 3.1 depicts the evolution of our design principles,

illustrating the transformation of uridine into visibly emissive nucleosides.

Conjugation of a 5-membered heterocycle, such as thiophene, at the 5 position of

uridine, a "dark" non-emissive native nucleoside (Φ ~10–4), yielded our first-

generation fluorescent nucleosides (**A**, Figure 3.1). Depending on the conjugated

heterocycle, such 5-modified pyrimidines emit in the visible range (390–443 nm),

have very large Stokes shifts (8400–9700 cm^{-1}), while their quantum efficiency is

relatively low (Φ=0.01–0.035).[19-21] Further enhancing the polarization of this

conjugated electron-poor/electron-rich biaryl system by introducing the

electronegative nitrogen at the pyrimidine's 6 position (**B**, Figure 3.1), resulted in

red-shifted absorption and emission maxima and substantially augmented

quantum yields (Φ =0.2–0.8).[22] To further shift the emission into the red region of

the spectrum, we have advanced the family shown here by directly conjugating a donor group through an extended aromatic system to the electron deficient 6-aza U (**C**, Figure 3.1). Here we disclose the synthesis, as well as structural and photophysical features of this advanced visibly emitting motif.

3.2 Synthesis

Although multiple synthetic approaches are conceivable, we have selected Suzuki coupling reactions as the key step in constructing all derivatives (**48**–**53**) from one common precursor (Scheme 3.1). Bromination of the protected nucleoside **64**, which was synthesized using a previously published procedure,[22] gave **66** in very good yields (Scheme 3.1, Figure 3.2a). Deprotection with methanolic ammonia at 60 °C, followed by recrystallization, yielded **67**. All extended nucleosides **48**–**53**, were obtained via a Suzuki coupling reaction between **67** and boronic acids **68**–**73**,[23, 24] using a combination of a water-soluble ligand, tris(3-sulfophenyl)phosphine trisodium salt (TPPTS), and palladium acetate as a catalyst.[25] Trituration with water and recrystallization from methanol afforded pure **48**–**53** (49–74%). All nucleosides were thoroughly characterized by ^{1}H- and ^{13}C-NMR spectroscopy as well as by HR MS and crystallography.[25]
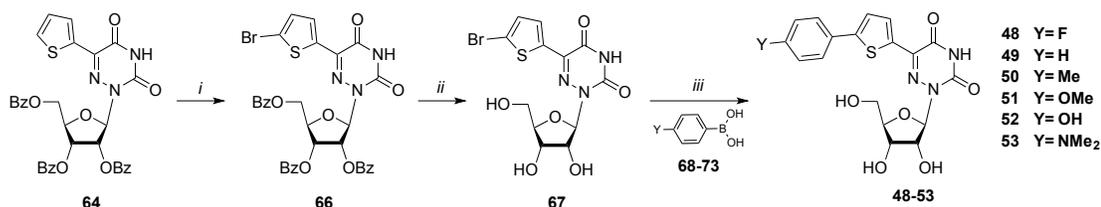
Scheme 3.1 Syntheses of nucleosides **48**–**53**. See experimental for synthetic procedures and analytical data.

3.3 X-ray crystal structures

The crystal structure of **51** (Figure 3.2b) illustrates the common structural features of these extended nucleosides (see also Figure 3.6-3.10). Overlaying the structures of uridine and **51** (Figure 3.2c) shows the minimal impact on the sugar pucker, although a noticeable difference in the dihedral angle χ (−164.41º and −89.0º for uridine and **51**, respectively; Figure 3.2d). Such differences are likely a result of crystal packing forces, as the extended derivatives frequently show extensive aromatic–aromatic interactions in the solid state (see Figure 3.11).
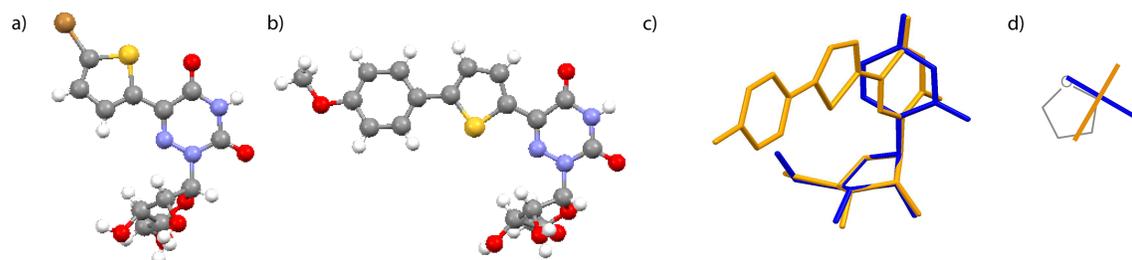
Figure 3.2 (a, b) X-ray crystal structures of **67** and **51**, respectively; (c) Overlay of X-ray the crystal structure of **51** (orange) with uridine (blue); overlaying the ribose rings shows minimal impact on the sugar pucker (rmsd=0.04 Å); (d) Schematic top view illustrating the relative conformation of the nucleobases in uridine (blue) and **51** (orange).

3.4 Photophysical Properties

To evaluate their basic features and assess the influence of the remote substituents on the photophysical properties of **48**–**53**, absorption and emission spectra were recorded in dioxane and water (Figure 3.3, Table 3.1). The absorption maxima of all nucleosides are redshifted compared to the parent conjugated aza-uridine,[22] and found in the low energy range of the UV spectrum (350–400 nm). All nucleosides **48**–**53** are visibly fluorescent, covering a wide window of emission

energies ranging from ca. 450 to 600 nm. All display rather large Stokes shifts (> 5,000 cm^{-1}), which for all derivatives but one (53) become even more pronounced as polarity increases (> 7,000 cm^{-1}). Their emission quantum yields in dioxane are moderate to high (0.2–0.75), but drop in a substitution-dependent manner, as discussed below, when taken in more polar solvents.

Increasing the electron-rich character of the substituent on the phenyl ring results in a bathochromic shift for both absorption and emission maxima in the following general order: **48≈49<50<51<52<53**.[26] This illustrates the impact of the substituent on both the ground and excited state. In apolar media the highest quantum yield is observed for derivatives with oxygen-containing substituents **51** and **52** (0.74 and 0.71, respectively). The opposite is observed in polar protic media where the less electron-rich substitutions show a higher fluorescent intensity while strong fluorescent quenching is observed for the most electron rich derivatives **52** and **53** compared to **48**–**51**. Such facilitation of non-radiative decay pathways for fluorophores capable of H-bonding is not uncommon.[13,27,28] Correlating the calculated Stokes shifts observed in dioxane against Hammett σ_{para} and σ^{+}_{para} parameters shows a reliable trend (Figure 3.3c,d). This provides a useful design tool enabling the use of established linear free energy parameters to confidentially anticipate select photophysical properties.[29-30]
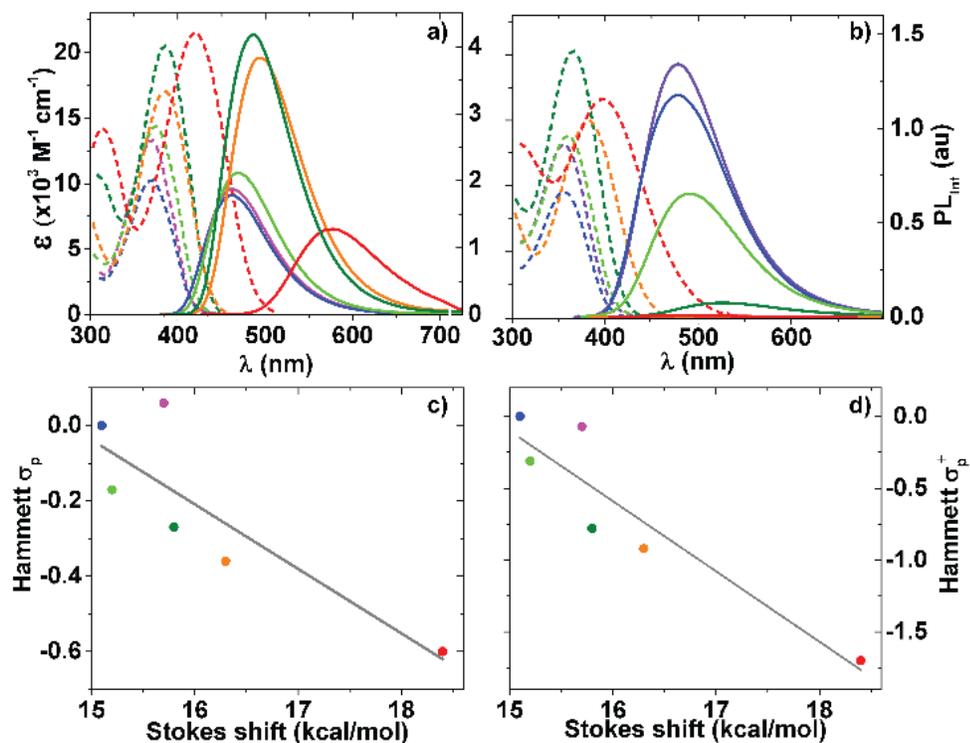
Figure 3.3 Molar absorptivity (dotted line) and emission (solid line) spectra for **48** (purple), **49** (blue), **50** (green), **51** (dark green), **52** (orange), **53** (red) in dioxane (a) and water (b). Emission was recorded after excitation at $\lambda_{abs\ max}$ for each derivative (Table 3.1). Calculated Stokes shift in kcal/mol for spectra taken in dioxane are correlated with Hammett σ_{para} (c) and Hammett σ^+_{para} (d).

Table 3.1 Photophysical properties of nucleosides **48**–**53**.[a]

compound	substituent Y	solvent	absorption λ_{max} [nm]	absorption ε [b]	emission λ_{max} [nm]	emission Φ [c]	brightness $\Phi \times \varepsilon \times 10^3$	Stokes shift $\nu_{abs} - \nu_{em}$ [kcal/mol]	Stokes shift $\nu_{abs} - \nu_{em}$ [cm^{-1}]
B^{22} (R=ribose) (X=S)		Dioxane	335	13	415	0.8	10.4	6025	16.4
		Methanol	334	11	433	0.2	5.5	7332	19.6
		Water	332	11	455	0.5	2.2	8492	23.3
48	F	Dioxane	368	15	462	0.32	4.8	5504	15.7
		Methanol	365	14	488	0.11	1.5	6867	19.6
		Water	357	12.6	478	0.24	3	7140	20.4
49	H	Dioxane	371	9.7	461	0.3	2.9	5286	15.1
		Methanol	367	8.9	488	0.16	1.4	6767	19.3
		Water	356	9.6	478	0.21	2	7184	20.5
50	Me	Dioxane	375	14.6	468	0.38	5.6	5308	15.2
		Methanol	372	14	499	0.07	1	6855	19.6
		Water	360	13.9	490	0.12	1.7	7370	21.1
51	MeO	Dioxane	383	20.8	486	0.74	15.4	5534	15.8
		Water	366	19.7	525	0.02	0.4	8275	23.7
52	HO	Dioxane	385	16.9	494	0.71	12	5695	16.3
		Water	381	15	517	<0.01	<0.1	6944	19.8
53	Me$_2$N	Dioxane	420	21.4	575	0.2	4.3	6428	18.4
		Water	398	16.8	484	<0.01	<0.1	4464	12.8

[a] Absorption and steady-state emission spectroscopy studies were performed using samples prepared from concentrated DMSO stock solution.[25] [b] ε in [$\times$ 10^3 M^{-1} cm^{-1}]. [c] Fluorescence standards: Coumarin 102 was used for **48**–**52**, and Coumarin 153 was used for **53**.

3.5 Sensitivity to polarity and pH

To evaluate the influence of polarity on the nucleosides' photophysical properties and hence their responsiveness, spectra were measured in dioxane [$E_T(30)$=36.0 kcal/mol], methanol [$E_T(30)$=55.4 kcal/mol], and mixtures thereof.[31] For each solution the $E_T(30)$ value was experimentally determined using Reichardt's dye.[32] While absorption spectra show little to no variation as polarity is systematically varied[33], significant changes are seen in both emission wavelengths and intensity (Figure 3.4a,b).[34] With increasing polarity a bathochromic shift of $\lambda_{em\ max}$ was seen for all extended nucleosides **48**–**53**. For example, the emission maximum of **51** is 468 in dioxane and 499 nm in methanol. Similar trends were reported for the parent 5-thiopheno-6-aza-uridine,[22] and 5-thiopheno-uridine.[35] To better quantify this effect, Stokes shifts were calculated for each sample containing dioxane/methanol mixture and plotted against the experimentally determined $E_T(30)$ values. As with related emissive nucleosides, a linear fit is observed (Figure 4c,d). Additionally, a steady decrease in the integrated emission was observed for **51** and **52** (as well as the parent 5-thiopheno-6-aza-uridine,[22] and 5-thiophene-uridine[35]) as the content of the the protic solvent (e.g., H_2O, MeOH) increases. This phenomenon is rather common in fluorophores that display significant charge transfer character in their excited state.[13, 32]
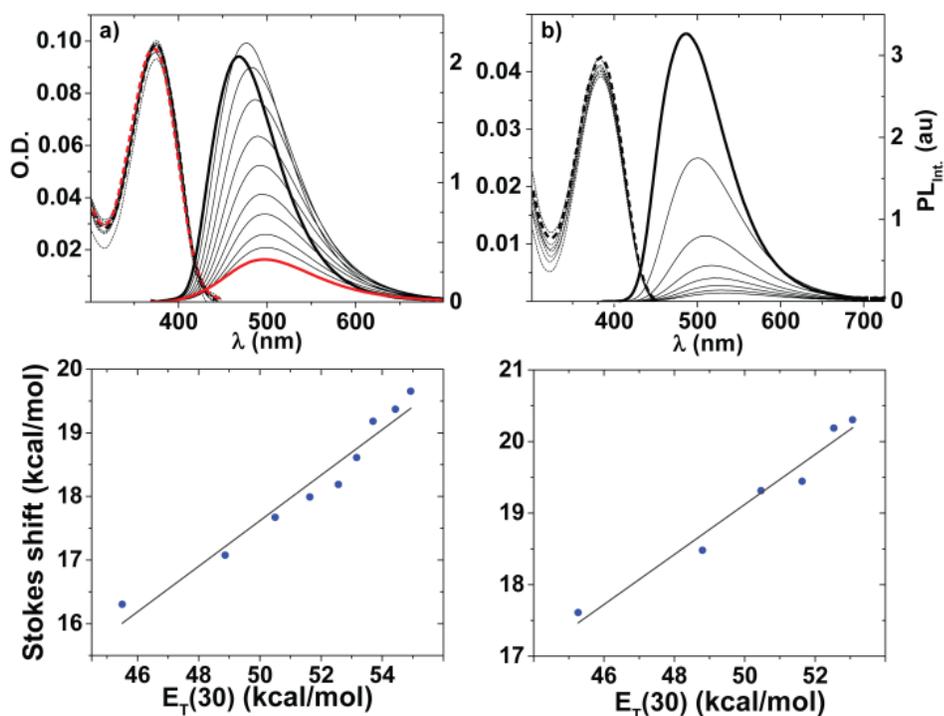
Figure 3.4 (a,b) Assessing the effect of solvent polarity on absorption (dotted line) and emission (solid line), in dioxane (bold black line) methanol (bold red line) and their mixtures (black lines) for **51** (6.7×10^{-6} M) and **52** (2.0×10^{-6} M); respectively.[25] (c, d) Correlating Stokes shift vs. $E_T(30)$ values obtained from dioxane–methanol mixtures for **51** (90%:10% → 10%:90%; slope: 0.36 and R^2=0.95) and **52** (90%:10% → 40%:60%; slope: 0.35 and R^2=0.98).[25] Experimental errors are smaller than the data symbols; see experimental for enlarged correlations (Figures 3.27 and 3.28).

3.6 Conclusion

In summary, visibly emitting, bright and responsive nucleosides have been obtained by implementing an enhanced charge transfer character in 5-substituted 6-aza-uridines. The photophysical features of these synthetically accessible analogs can be tuned by judiciously introducing substituents of distinct electronic character at a remote but conjugated position. In general, the extended analogs reported here display higher emission quantum yields in apolar solvents but remain sufficiently bright in polar media. This trend is, however, dependent on the nature

of the substituent with highly electron rich derivatives suffering the highest loss in emission quantum yield. Nevertheless, the highly desirable and tunable photophysical properties, including pronounced solvatochromism, make this 6-aza-uridine motif a very attractive scaffold for the design and development of useful biophysical probes.

3.7 Experimental

Materials

Palladium acetate, (4-(dimethylamino)phenyl)boronic acid, 1,2-dichloroethane and dry acetonitrile were purchased from Sigma-Aldrich. Sodium thiosulphate was purchased from Amresco. Ammonium chloride was purchased from Mallinckrodt. Sodium sulfate, sodium carbonate and all solvents were purchased from Fisher. Tris(3-sulfophenyl)phosphine trisodium salt was purchased from Alfa-Aesar. (4-Methoxyphenyl)boronic acid was purchased from Frontier Scientific. (4-hydroxyphenyl)boronic, (4-methylphenyl)boronic, phenylboronic and (4-fluorophenyl)boronic were purchased from Combi-Blocks. All reagents and solvents were used without further purification. Moister and oxygen sensitive reactions were performed in an inert argon atmosphere. 5-(thiophene-2-yl)-6-aza-2',3',5'-tribenzoyl-uridine **2** synthesized by previously published method in our laboratory.[22]

Synthesis
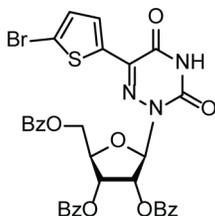
Synthesis of 5-(5-bromothiophene-2-yl)-6-aza-2',3',5'-tribenzoyl-uridine (66).

5-(thiophene-2-yl)-6-aza-2',3',5'-tribenzoyl-uridine **64** (1.1 g, 1.72 mmol) was dissolved in 1,2-dichloroethane (22 mL) at rt. Bromine (0.18 mL, 3.51 mmol) was then added and the reaction mixture left to stir at rt. After 1h, saturated aqueous solution of sodium thiosulphate (0.5 mL) was added and mixed until the red color changed to yellow. The reaction mixture was washed with water (3 × 100 mL). The combined organic layers were dried over Na_2SO_4 filtered over a glass frit and concentrated to dryness. Purification of the crude residue by column chromatography (silica, 7 *v*% DCM in EtOAc) yielded **66** as an almost white foam (1.18 g, 1.64 mmol, 95%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.64 (br s, 1H), 8.05 – 7.73 (m, 7H), 7.72 – 7.20 (m, 10H) 6.49 (s, 1H), 6.10 – 5.95 (m, 2H), 4.90 – 4.75 (m, 1H), 4.68 (dd, *J* = 12.2, 3.4 Hz, 1H), 4.56 (dd, *J* = 12.2, 4.6 Hz, 1H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 165.85, 165.14, 165.01, 155.85, 148.21, 137.93, 135.56, 134.44, 134.29, 133.93, 131.46, 129.92, 129.82, 129.76, 129.60, 129.46, 129.30, 129.15, 128.97, 128.92, 117.13, 88.11, 78.79, 74.36, 71.05, 63.67. HR-MS-ESI: m/z calcd. for $C_{33}H_{24}BrN_3O_9S$ (M+Na)$^+$: 740.0309, found 740.0307.
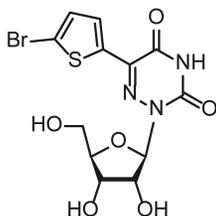
5-(5-bromothiophen-2-yl)-6-aza-uridine (67). A solution of **66** (1.14 g, 1.58 mmol) in saturated methanolic ammonia (10 mL) was heated at 60 °C in a pressure vessel for 24 h. The mixture was cooled to rt, concentrated and the residue was triturated with DCM (3 × 60 mL). Purification of the crude residue by recrystallization from water/methanol (60 mL/20 mL) yielded **67** as a white solid (495.5 mg, 1.22 mmol, 77%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.49 (br s, 1H), 7.75-7.67 (m, 1H), 7.35-7.25 (m, 1H), 5.97 (d, J = 2.0 Hz, 1H), 5.31 (d, J = 4.7 Hz, 1H), 5.09 (d, J = 6.1 Hz, 1H), 4.64 (t, J = 5.6 Hz, 1H), 4.28 – 4.12 (m, 2H), 3.88 – 3.78 (m, 1H), 3.64 – 3.54 (m, 1H), 3.49 – 3.38 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 155.69, 148.53, 137.10, 136.47, 131.56, 129.71, 116.23, 89.99, 85.14, 73.72, 70.80, 62.36; HR-MS-ESI: m/z calcd. for C$_{12}$H$_{12}$BrN$_3$O$_6$S (M-H)$^-$: 403.9557, found: 403.9558.
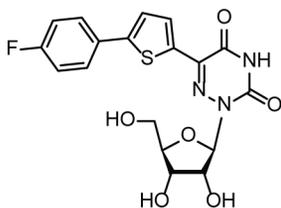
Synthesis of 5-(5-(4-fluorophenyl)thiophen-2-yl)-6-aza-uridine (48).

All solvents were purged with argon. Under argon, sodium carbonate (400 mg, 3.77 mmol) was dissolved in water (6 mL). Then **67** (500 mg, 1.23 mmol), (4-fluorophenyl)boronic acid **68** (258.3 mg, 1.85 mmol) and dry acetonitrile were added to the solution and additionally purged with argon for 15 minutes. Tris(3-sulfophenyl)phosphine trisodium salt (175 mg, 0.31 mmol) and palladium acetate (13.8 mg, 0.06 mmol) were added and the reaction mixture was heated and kept at 60 °C for 18 h. The mixture was cooled to rt, neutralized with saturated aqueous solution of ammonium chloride and diluted with methanol (100 mL). A black precipitate was filtrate off, and remaining filtrate was concentrated to dryness. The residue was triturated with water (3 × 60 mL). Purification of the crude solid by recrystallization from methanol (50 mL) yielded **48** as a bright yellow solid (337.0 mg, 0.80 mmol, 65%).^{1}H NMR (500 MHz, DMSO-d_6) δ 12.47 (br s, 1H), 7.96 (d, J = 4.0 Hz, 1H), 7.80 – 7.72 (m, 2H), 7.54 (d, J = 4.0 Hz, 1H), 7.34 – 7.26 (m, 2H), 5.99 (d, J = 3.0 Hz, 1H), 5.36 (d, J = 4.5 Hz, 1H), 5.12 (d, J = 6.0 Hz, 1H), 4.70 (t, J = 6.0 Hz, 1H), 4.32 – 4.19 (m, 2H), 3.87 –.3.81 (m, 1H), 3.66 – 3.59 (m, 1H), 3.51 – 3.43 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d_6) δ 163.47, 161.51, 155.72, 148.61, 145.29, 137.72, 134.32, 130.79, 130.20, 130.18, 128.18, 128.12, 124.97,

116.77, 116.60, 90.03, 85.24, 73.63, 70.95, 62.56; ^{19}F NMR (280 MHz, DMSO-d$_6$)

δ 116.78; HR-MS-ESI: m/z calcd. for C$_{18}$H$_{16}$FN$_3$O$_6$S (M-H)$^-$: 420.0671, found:

420.0673
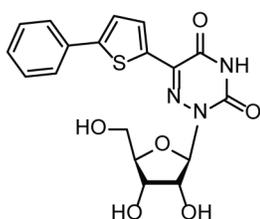
Synthesis of 5-(5-phenylthiophen-2-yl)-6-aza-uridine (49).

Using the procedure described for **48**, but starting from **69, 49** was obtained as a

bright yellow solid (335.2 mg, 0.83 mmol, 68%).^{1}H NMR (400 MHz, DMSO-d$_6$) δ

12.45 (br s, 1H), 7.97 (d, *J* = 4.0 Hz, 1H), 7.80 – 7.64 (m, 2H), 7.57 (d, *J* = 4.0 Hz,

1H), 7.51 – 7.40 (m, 2H), 7.40 – 7.27 (m, 1H), 6.00 (d, *J* = 3.2 Hz, 1H), 5.33 (d, *J*

= 5.2 Hz, 1H), 5.11 (d, *J* = 6.0 Hz, 1H), 4.68 (t, *J* = 6.0 Hz, 1H), 4.36 – 4.16 (m,

2H), 3.90 –.3.78 (m, 1H), 3.70 – 3.56 (m, 1H), 3.54 – 3.42 (m, 1H); ^{13}C NMR (125

MHz, DMSO-d$_6$) δ 155.73, 148.61, 146.45, 137.77, 134.32, 133.51, 130.81,

129.73, 128.90, 126.01, 124.85, 90.06, 85.24, 73.65, 70.97, 62.58; HR-MS-ESI:

m/z calcd. for C$_{18}$H$_{17}$N$_3$O$_6$S (M-H)$^-$: 402.0765, found: 402.0767.
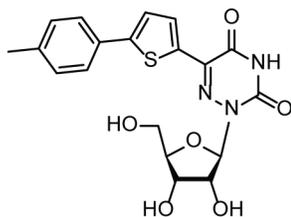
Synthesis of 5-(5-(4-methyl)phenyl)thiophen-2-yl)-6-aza-uridine (50).

Using the procedure described for **48**, but starting from **70**, **50** was obtained as a a bright yellow solid (354.5 mg, 0.85 mmol, 69%). ^{1}H NMR (500 MHz, DMSO-d$_6$) δ 12.45 (br s, 1H), 7.95 (d, J = 4.0 Hz, 1H), 7.60 (d, J = 8.0 Hz, 2H), 7.51 (d, J = 4.0 Hz, 1H), 7.25 (d, J = 8.0 Hz, 2H), 5.99 (d, J = 3.0 Hz, 1H), 5.35 (d, J = 4.5 Hz, 1H), 5.13 (d, J = 6.0 Hz, 1H), 4.70 (t, J = 6.0 Hz, 1H), 4.34 − 4.18 (m, 2H), 3.90 −.3.80 (m, 1H), 3.68 − 3.57 (m, 1H), 3.53 − 3.42 (m, 1H) 2.32 (s, 3H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 155.73, 148.61, 146.67, 138.48, 137.80, 133.74, 130.85, 130.78, 130.24, 125.90, 124.25, 90.01, 85.24, 73.62, 70.97, 62.59, 21.27; HR-MS-ESI: m/z calcd. for C$_{19}$H$_{19}$N$_3$O$_6$S (M-H)$^-$ 416.0922, found: 416.0923.
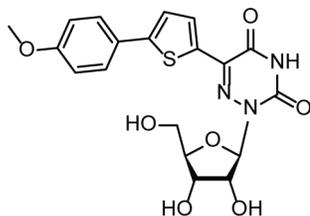
Synthesis of 5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (51).

Using the procedure described for **48**, but starting from **71, 51** was obtained as a yellow solid (351.9 mg, 0.81 mmol, 66%). ^{1}H NMR (500 MHz, DMSO-d$_{6}$) δ 12.45 (br s, 1H), 7.95 (d, J = 4.0 Hz, 1H), 7.70 – 7.60 (m, 2H), 7.44 (d, J = 4.0 Hz, 1H), 7.05 – 6.97 (m, 2H), 6.01 (d, J = 3.0 Hz, 1H), 5.37 (d, J = 5.0 Hz, 1H), 5.14 (d, J = 6.0 Hz, 1H), 4.71 (t, J = 5.5 Hz, 1H), 4.32 – 4.18 (m, 2H), 3.87 –.3.82 (m, 1H), 3.79 (s, 3H), 3.69 – 3.59 (m, 1H), 3.53 – 3.43 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d$_{6}$) δ 159.94, 155.74, 148.62, 146.66, 137.85, 133.22, 130.96, 127.44, 126.19, 123.61, 115.10, 90.02, 85.26, 73.63, 71.01, 62.64, 55.74; HR-MS-ESI: m/z calcd. for C$_{19}$H$_{19}$N$_{3}$O$_{7}$S (M-H)$^{-}$ 432.0874, found 432.0874
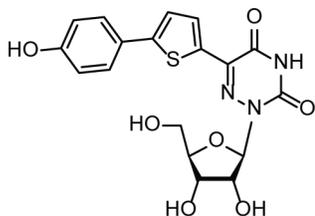
Synthesis of 5-(5-(4-hydroxyphenyl)thiophen-2-yl)-6-aza-uridine (52).

Using the procedure described for **48**, but starting from **72, 52** was obtained as a off-yellow solid (382.5 mg, 0.91 mmol, 74%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.41 (br s, 1H), 9.79 (br s, 1H), 7.92 (d, J = 4.0 Hz, 1H), 7.53 (d, J = 8.4 Hz, 2H), 7.35 (d, J = 4.0 Hz, 1H), 6.82 (d, J = 8.4 Hz, 2H), 5.99 (d, J = 2.8 Hz, 1H), 5.32 (d, J = 4.8 Hz, 1H), 5.09 (d, J = 6.0 Hz, 1H), 4.68 (t, J = 5.6 Hz, 1H), 4.34 − 4.15 (m, 2H), 3.90 −.3.77 (m, 1H), 3.68 − 3.56 (m, 1H), 3.53 − 3.40 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 158.41, 155.87, 148.73, 147.23, 137.89, 132.78, 130.99, 127.53, 124.66, 122.98, 116.42, 90.01, 85.23, 73.62, 71.03, 62.68; HR-MS-ESI: m/z calcd. for $C_{18}H_{17}N_3O_7S$ (M-H)$^-$ 418.0714, found 418.0715.
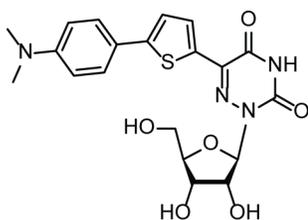
Synthesis of 5-(5-(4-(dimethylamino)phenyl)thiophen-2-yl)-6-aza-uridine (53).

Using the procedure described for **48**, but starting from **73, 53** was obtained as a red solid (269.3 mg, 0.60 mmol, 49%). ^{1}H NMR (500 MHz, DMSO-d$_6$) δ 12.39 (br s, 1H), 7.93 (d, J = 4.0 Hz, 1H), 7.57 – 7.49 (m, 2H), 7.33 (d, J = 4.0 Hz, 1H), 6.80 – 6.73 (m, 2H), 5.99 (d, J = 3.0 Hz, 1H), 5.32 (d, J = 5.0 Hz, 1H), 5.10 (d, J = 6.5 Hz, 1H), 4.68 (t, J = 5.5 Hz, 1H), 4.32 – 4.16 (m, 2H), 3.88 –.3.80 (m, 1H), 3.67 – 3.57 (m, 1H), 3.52 – 3.43 (m, 1H), 2.95 (s, 6H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 156.07, 150.76, 148.93, 147.86, 137.93, 131.91, 131.08, 126.93, 121.92, 121.24, 112.73, 90.01, 85.23, 73.62, 71.07, 62.76, 40.33; HR-MS-ESI: m/z calcd. for $C_{20}H_{22}N_4O_6S$ (M-H)$^-$ 445.1187, found 445.1190

^{1}H-NMR and ^{13}C-NMR spectra of 48–53
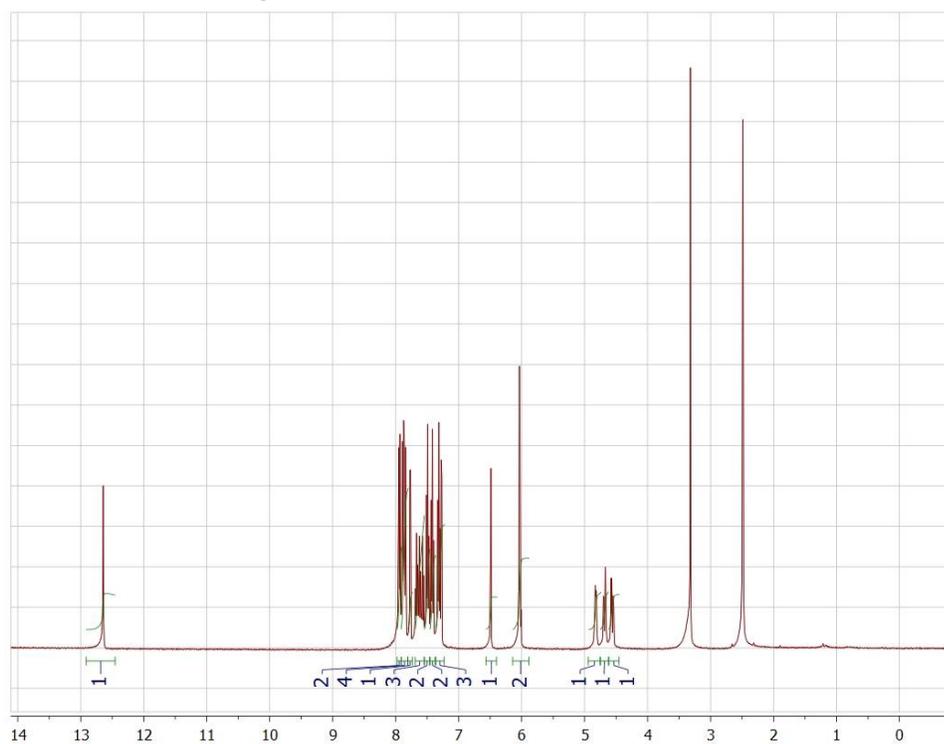
Spectrum 3.1 ^{1}H-NMR of 5-(5-bromothiophene-2-yl)-6-aza-2,3,5-tribenzoyl-uridine (**66**).
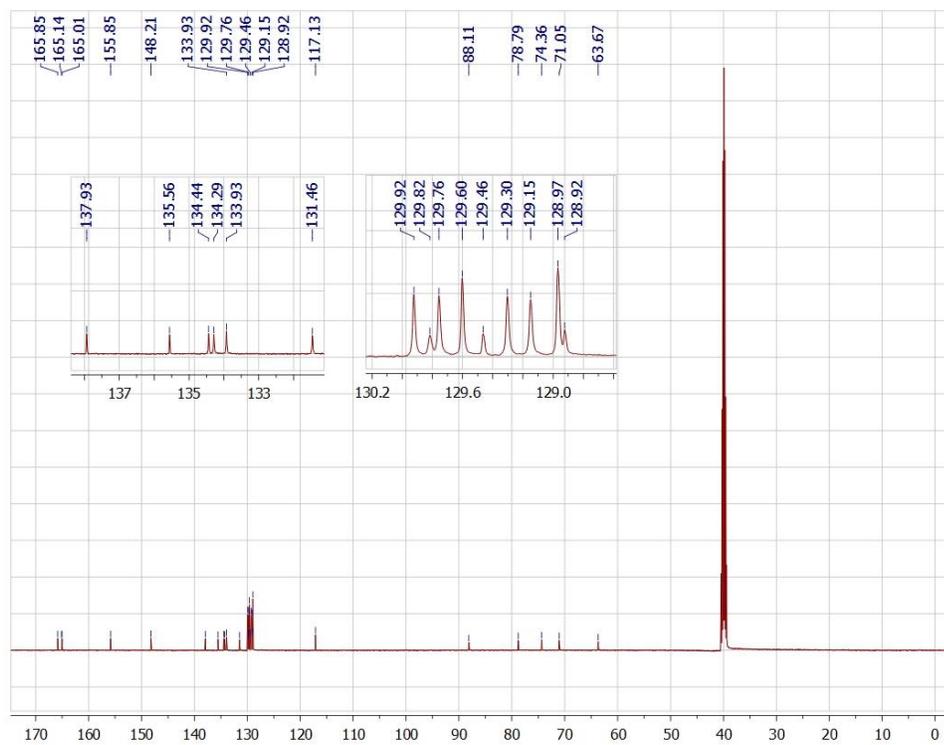
Spectrum 3.2 ^{13}C-NMR of 5-(5-bromothiophene-2-yl)-6-aza-2,3,5-tribenzoyl-uridine (**66**).
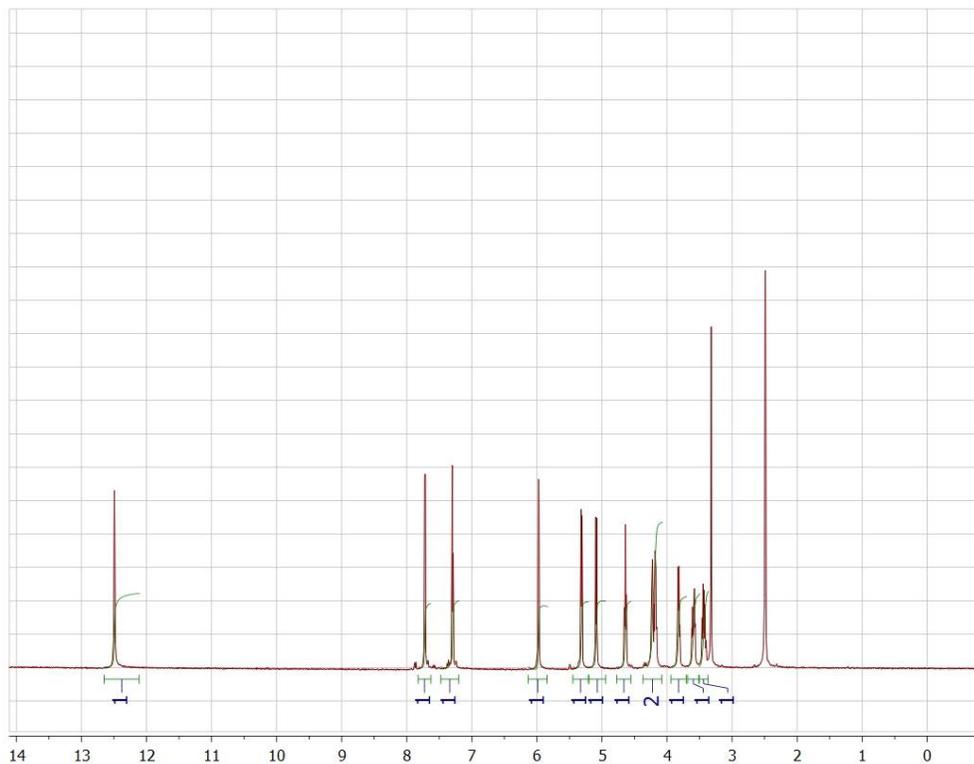
Spectrum 3.3 ^{1}H-NMR of 5-(5-bromothiophen-2-yl)-6-aza-uridine (**67**).
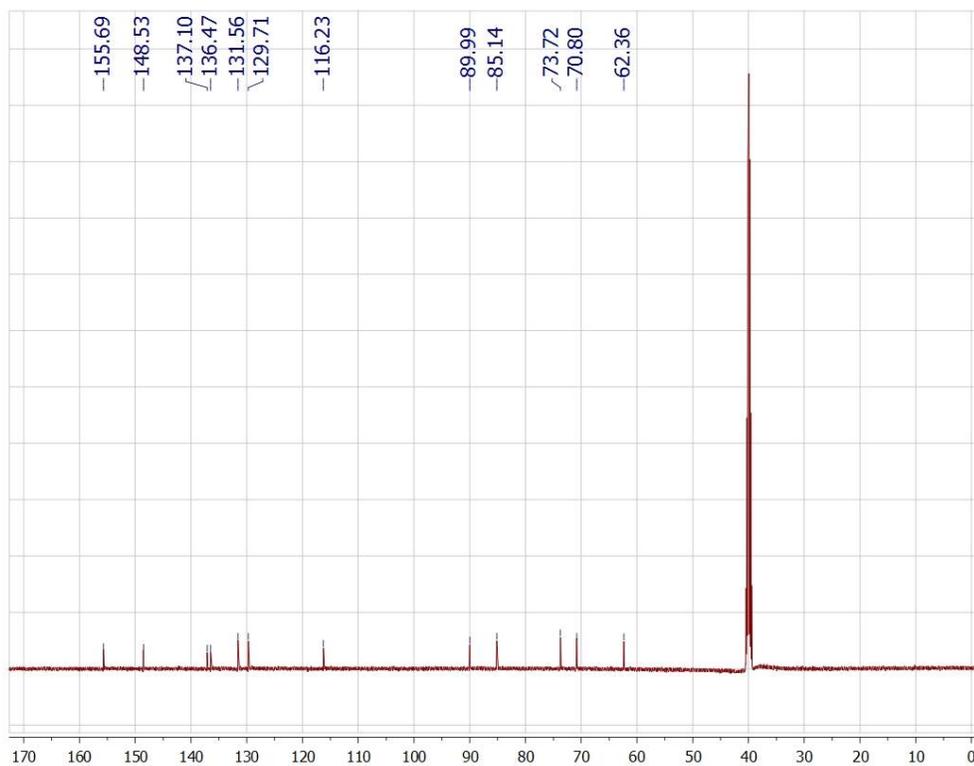
Spectrum 3.4 ^{13}C-NMR of 5-(5-bromothiophen-2-yl)-6-aza-uridine (**67**).
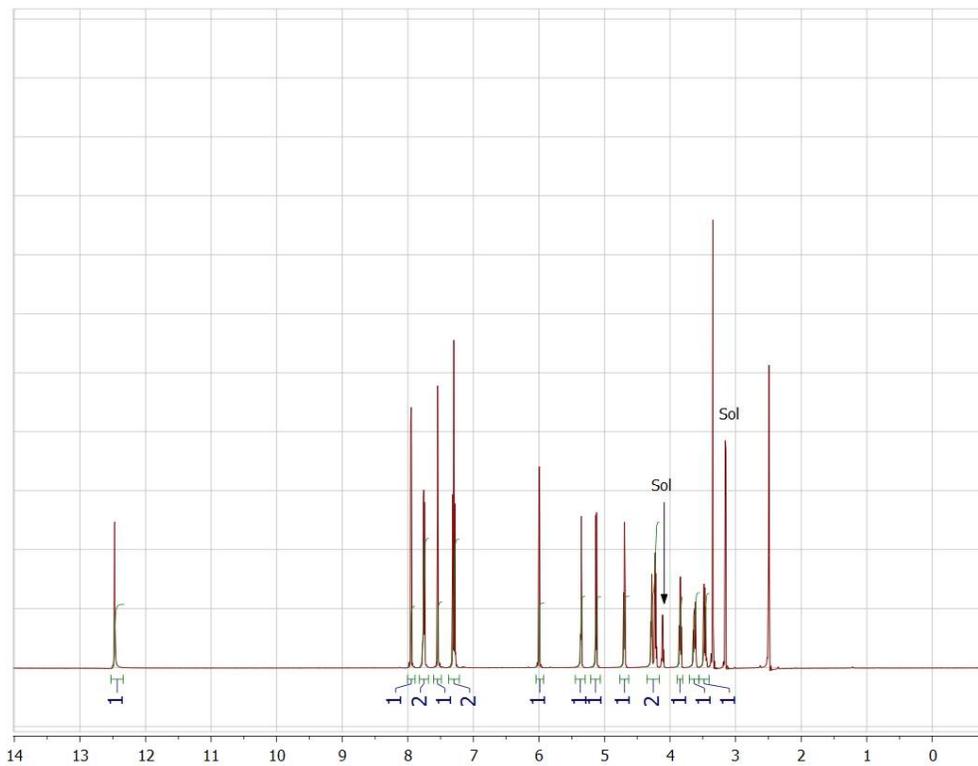
Spectrum 3.5 ^{1}H-NMR of 5-(5-(4-fluorophenyl)thiophen-2-yl)-6-aza-uridine (**48**).
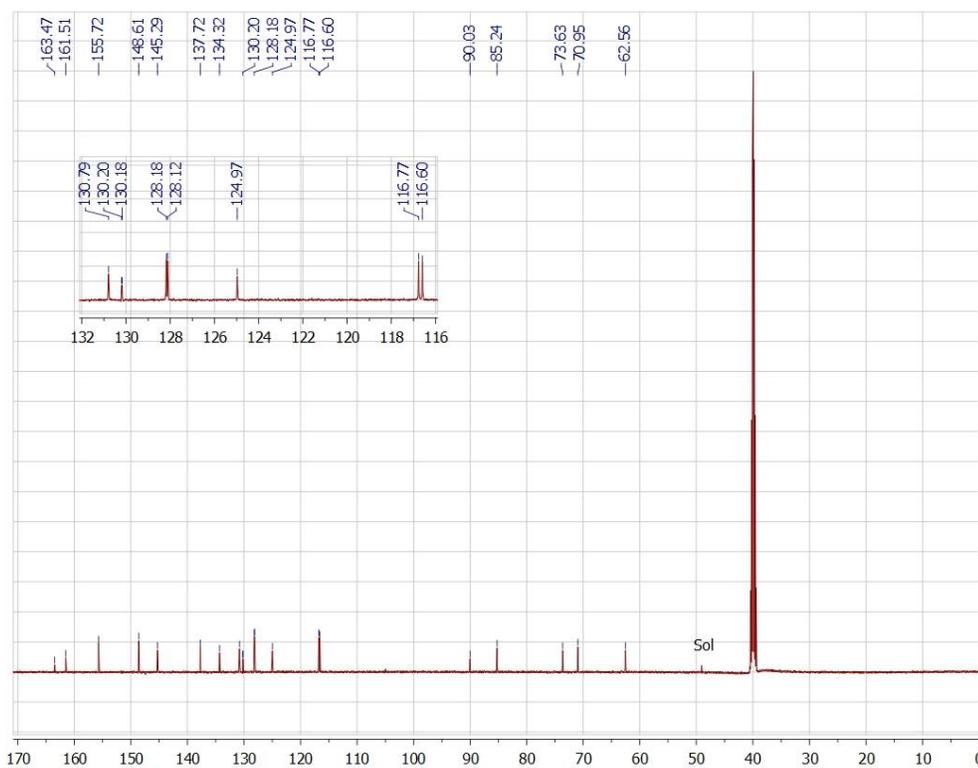
Spectrum 3.6 ^{13}C-NMR of 5-(5-(4-fluorophenyl)thiophen-2-yl)-6-aza-uridine (**48**).
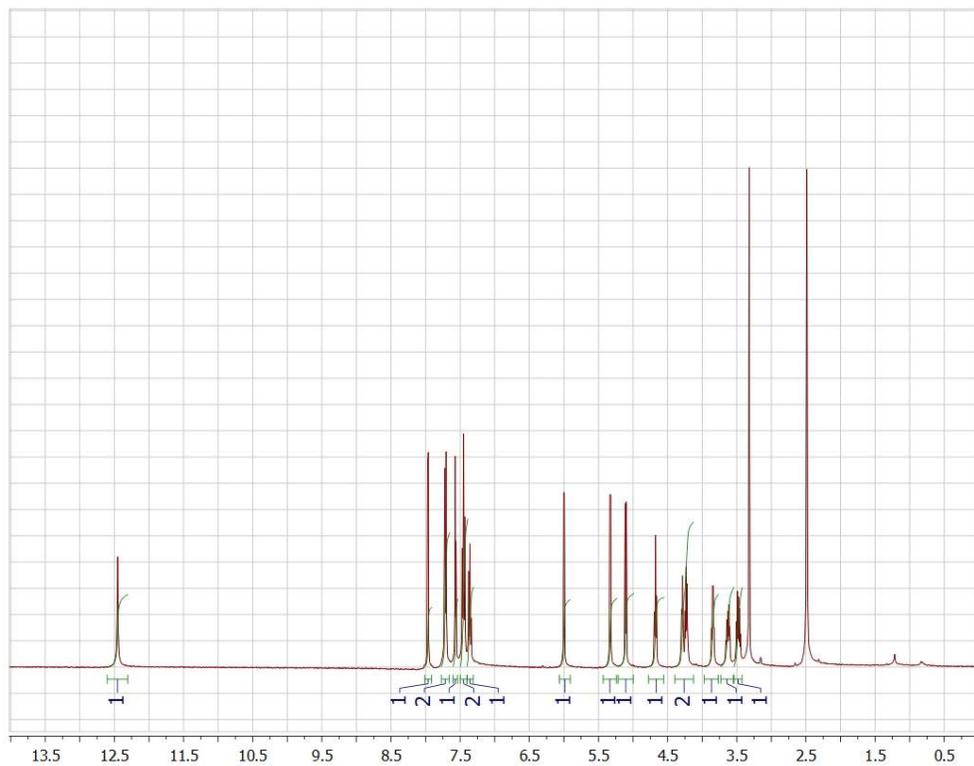
Spectrum 3.7 ^{1}H-NMR of 5-(5-phenylthiophen-2-yl)-6-aza-uridine (**49**).
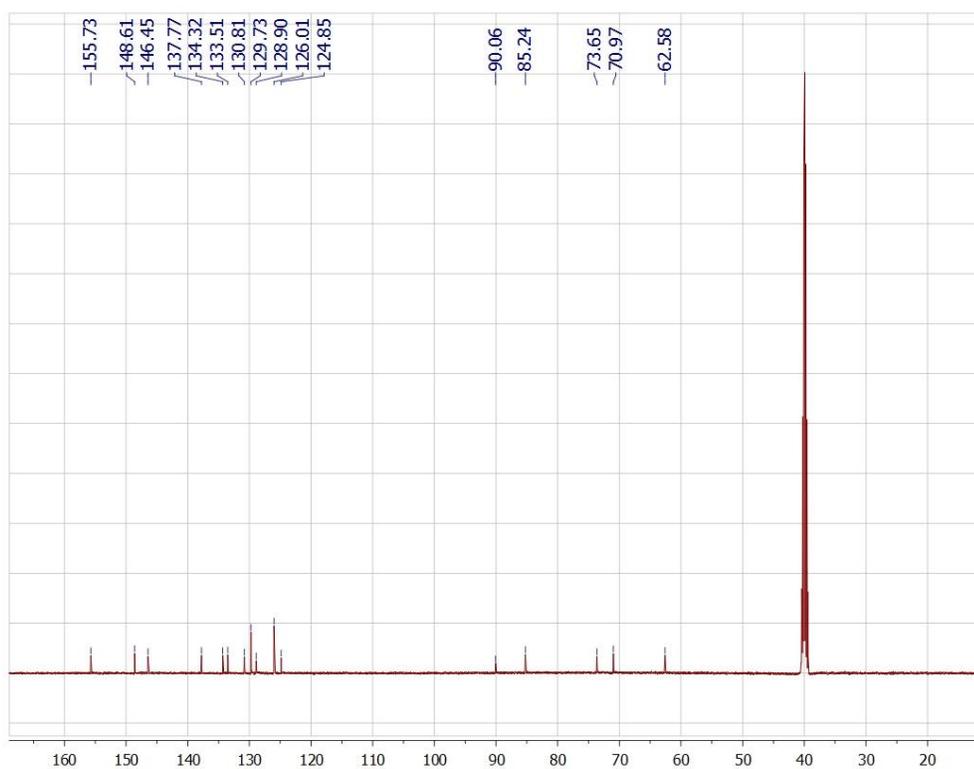
Spectrum 3.8 ^{13}C-NMR of 5-(5-phenylthiophen-2-yl)-6-aza-uridine (**49**).
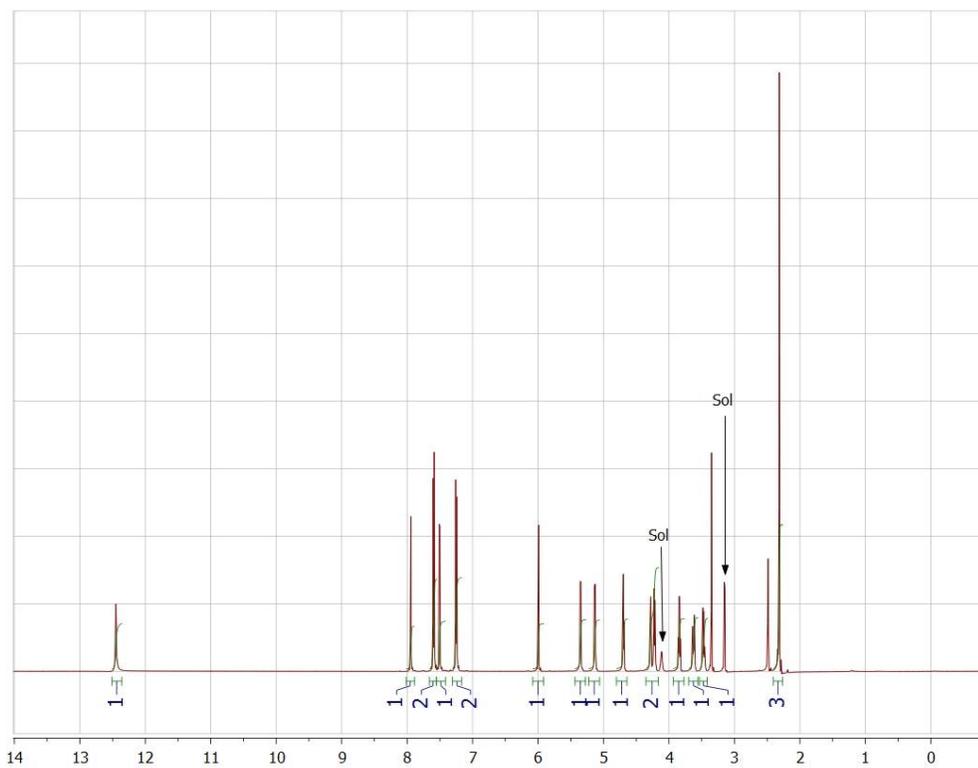
Spectrum 3.9 ¹H-NMR of 5-(5-(4-methyl)phenyl)thiophen-2-yl)-6-aza-uridine (**50**).
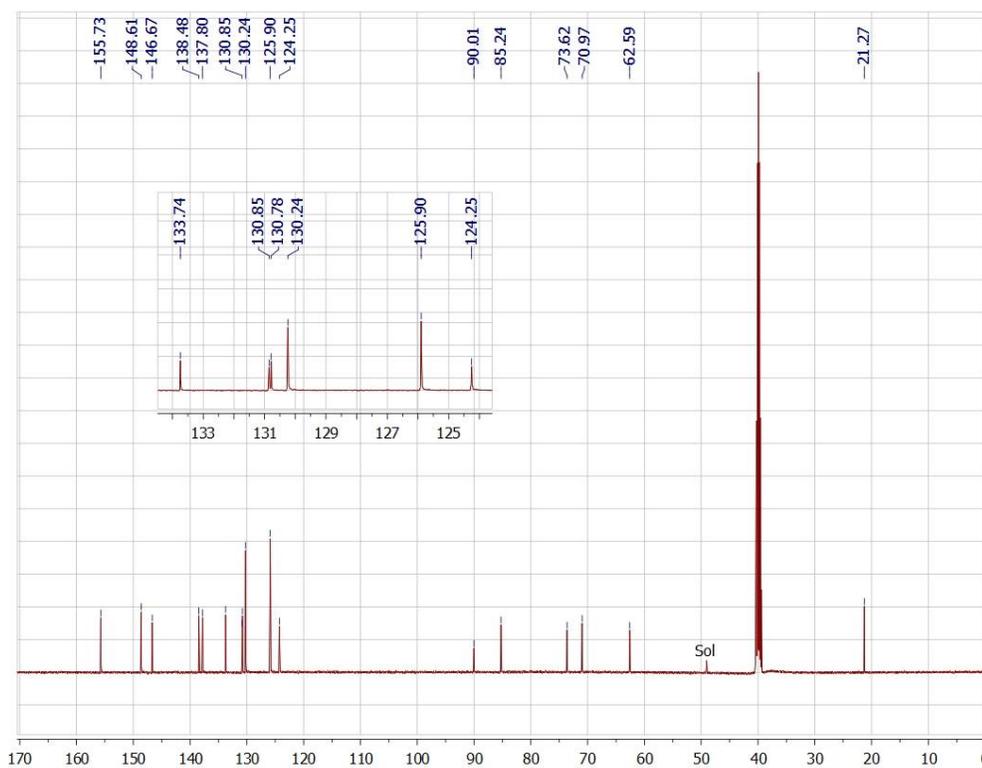
Spectrum 3.10 ¹³C-NMR of 5-(5-(4-methyl)phenyl)thiophen-2-yl)-6-aza-uridine (**50**).
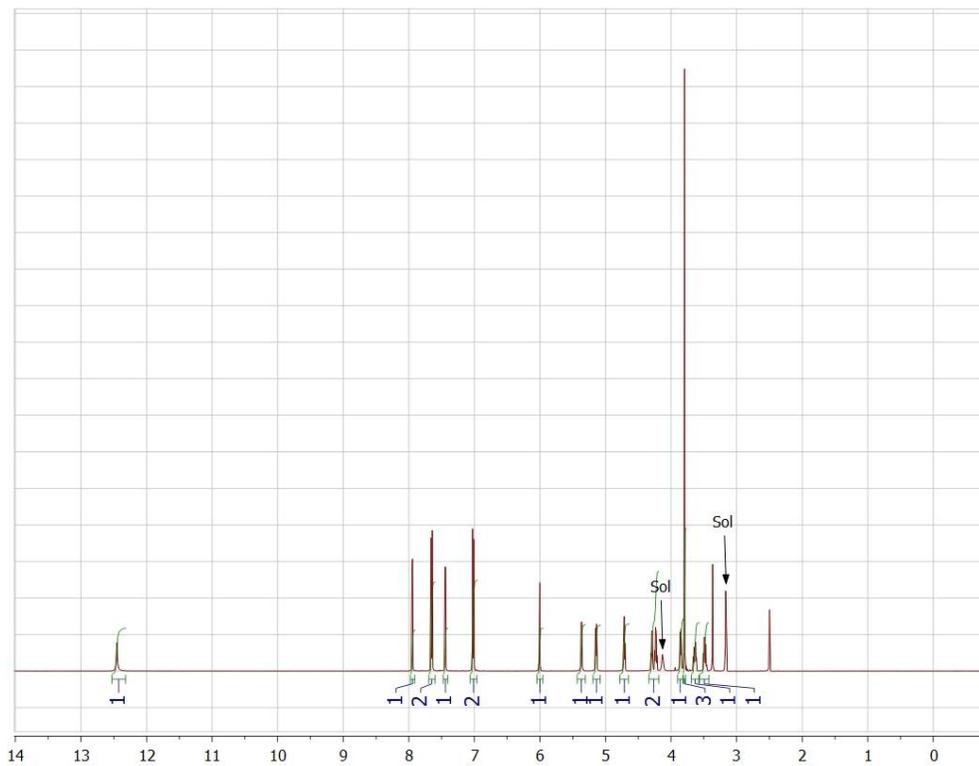
Spectrum 3.11 ^{1}H-NMR of 5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (**51**).
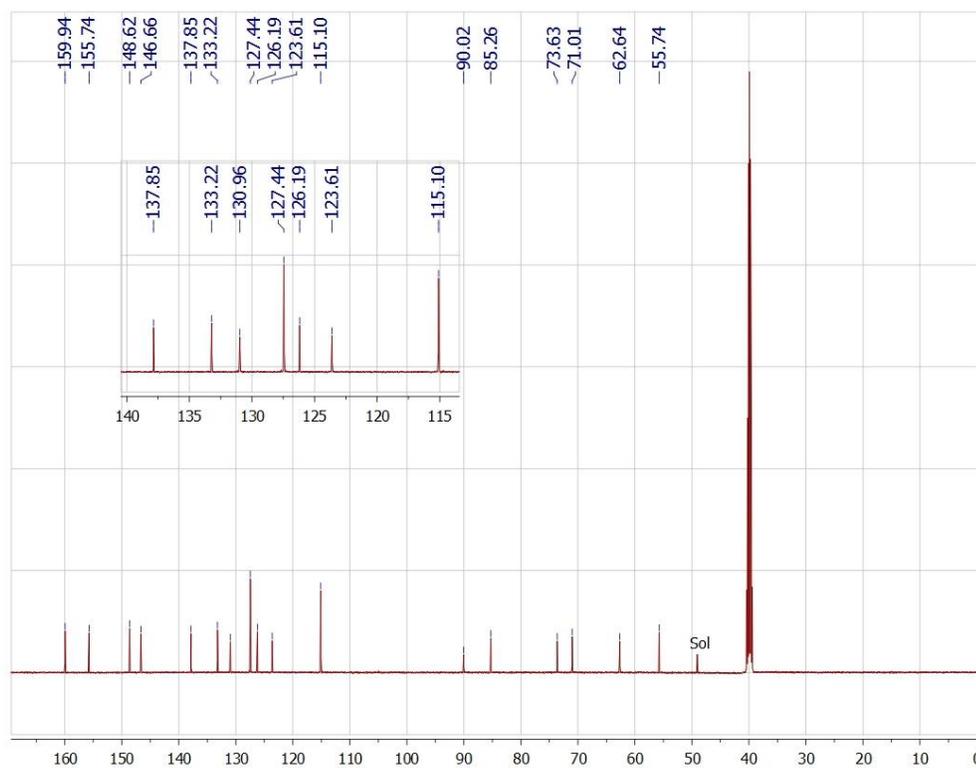
Spectrum 3.12 ^{13}C-NMR of 5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (**51**).
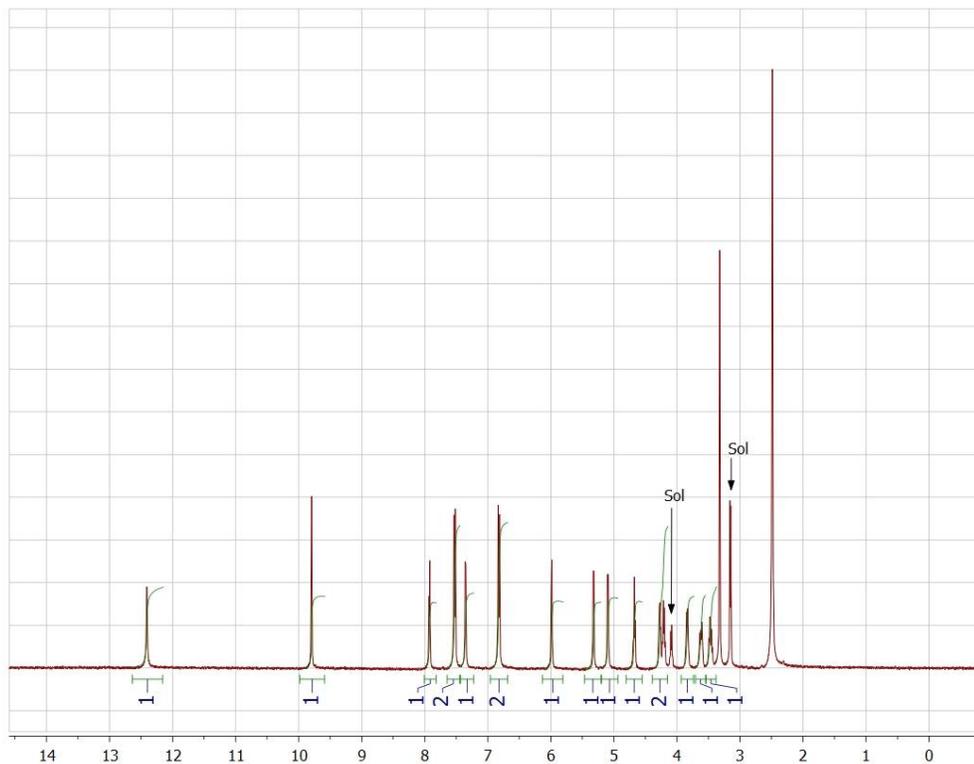
Spectrum 3.13 ^{1}H-NMR of 5-(5-(4-hydroxyphenyl)thiophen-2-yl)-6-aza-uridine (**52**).
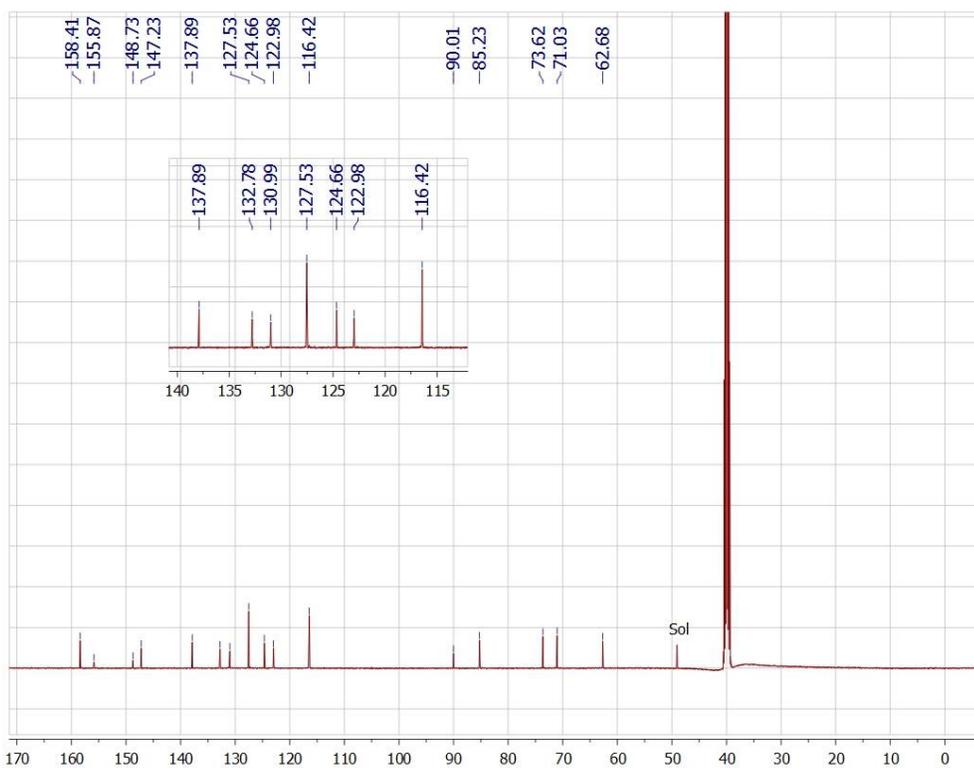
Spectrum 3.14 ^{13}C-NMR of 5-(5-(4-hydroxyphenyl)thiophen-2-yl)-6-aza-uridine (**52**).
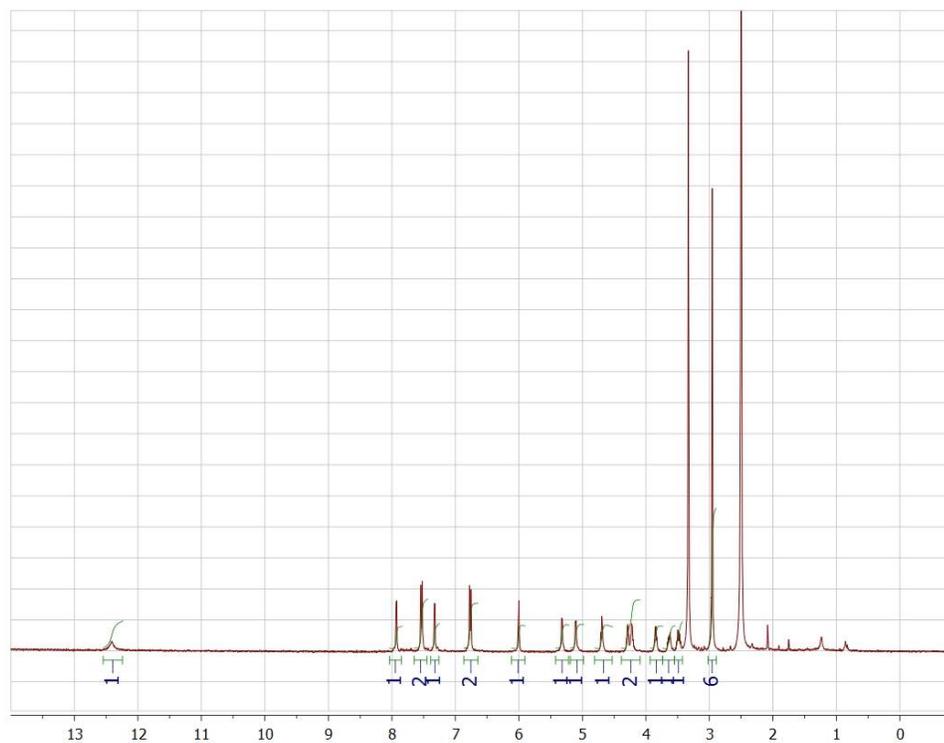
Spectrum 3.15 ^{1}H-NMR of 5-(5-(4-(dimethylamino)phenyl)thiophen-2-yl)-6-aza-uridine (**53**).
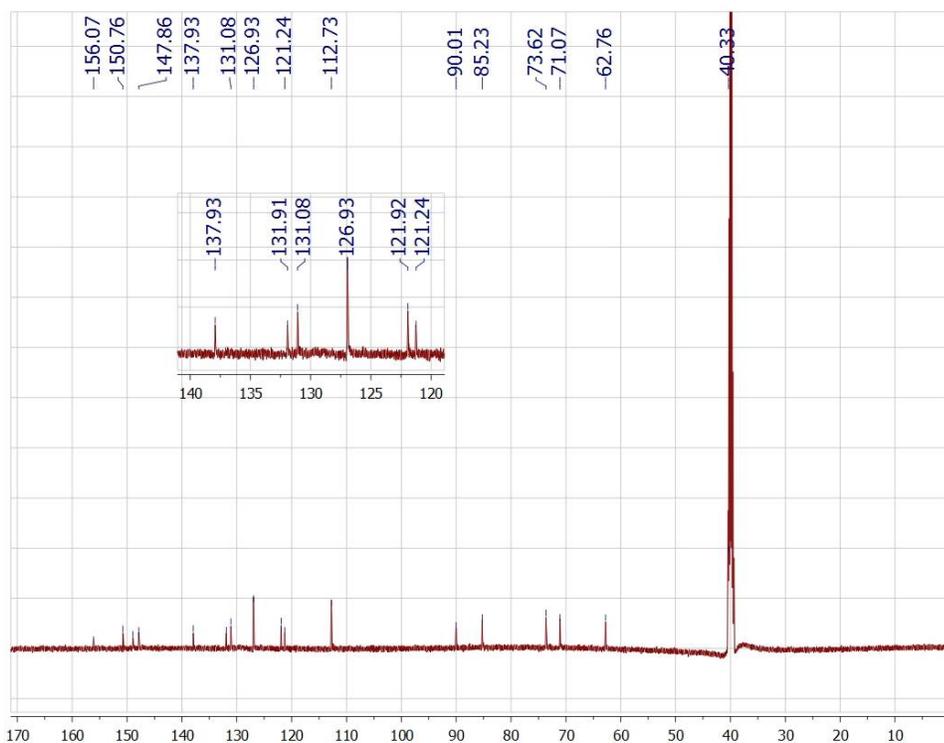
Spectrum 3.16 ^{13}C-NMR of 5-(5-(4-(dimethylamino)phenyl)thiophen-2-yl)-6-aza-uridine (**53**).

Crystal structures

Experimental Summary

The single crystal X-ray diffraction studies were carried out on a Bruker Kappa APEX-II CCD diffractometer equipped with Mo K$_\alpha$ radiation (λ = 0.71073 Å). A 0.103 x 0.011 x 0.005 mm colorless needle was mounted on a Cryoloop with Paratone oil. Data were collected in a nitrogen gas stream at 100(2) K using ϕ and ϖ scans. Crystal-to-detector distance was 30 mm and exposure time was 40 seconds per frame using a scan width of 1.0°. Data collection was 99.6% complete to 25.00° in θ. A total of 27161 reflections were collected covering the indices, -26<=h<=26, -7<=k<=7, -15<=l<=15. 3051 reflections were found to be symmetry independent, with a R$_{int}$ of 0.0621. Indexing and unit cell refinement indicated a *C*-centered, monoclinic lattice. The space group was found to be *C*2. The data were integrated using the Bruker SAINT software program and scaled using the SADABS software program. Solution by direct methods (SHELXT) produced a complete phasing model consistent with the proposed structure. All nonhydrogen atoms were refined anisotropically by full-matrix least-squares (SHELXL-2013). All hydrogen atoms were placed using a riding model. Their positions were constrained relative to their parent atom using the appropriate HFIX command in SHELXL-2013. Crystallographic data are summarized in Table 3.2-3.6.

Crystal structure(s) were deposited at the Cambridge Crystallographic Data Centre.

The data have been assigned to the following deposition numbers.

Summary of Data CCDC 1018947

Compound Name: **5-(5-bromothiophen-2-yl)-6-aza-uridine (67)**

Formula: C12 H12 Br1 N3 O6 S1

Unit Cell Parameters: a 21.582(3) b 5.6109(7) c 12.5574(14) C2

Summary of Data CCDC 1018950

Compound Name: **5-(5-(4-fluorophenyl)thiophen-2-yl)-6-aza-uridine (48)**

Formula: 2(C18 H16 F1 N3 O6 S1),C1 H4 O1

Unit Cell Parameters: a 9.4297(5) b 10.0067(4) c 10.3843(4) P1

Summary of Data CCDC 1025557

Compound Name: **5-(5-phenylthiophen-2-yl)-6-aza-uridine (49)**

Formula: C18 H17 N3 O6 S1,0.5(C1 H4 O1)

Unit Cell Parameters: a 20.4544(12) b 5.5733(3) c 16.1707(10) C2

Summary of Data CCDC 1018949

Compound Name: **5-(5-(4-methyl)phenyl)thiophen-2-yl)-6-aza-uridine (50)**

Formula: C19 H19 N3 O6 S1,C1 H4 O1

Unit Cell Parameters: a 9.4923(7) b 19.9718(15) c 10.7103(8) P21

Summary of Data CCDC 1018948

Compound Name: **5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (51)**

Formula: C19 H19 N3 O7 S1,C1 H4 O1

Unit Cell Parameters: a 9.0722(10) b 10.3704(11) c 12.3104(14) P1

Summary of Data CCDC 1025556

Compound Name: **5-(5-(4-(dimethylamino)phenyl)thiophen-2-yl)-6-aza-uridine (53)**

Formula: 2(C20 H22 N4 O6 S1),C2 H6 O1 S1

Unit Cell Parameters: a 8.9567(7) b 19.8870(16) c 24.7770(19) P212121
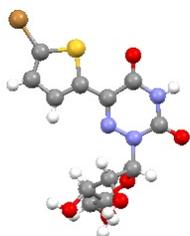
Figure 3.5 X-ray crystal structure of 5-(5-bromothiophen-2-yl)-6-aza-uridine (**67**).

Table 3.2 Crystal data and structure refinement for Tor75 5-(5-bromothiophen-2-yl)-6-aza-uridine (**67**).

Identification code	PH140
Empirical formula	C12 H12 Br N3 O6 S
Molecular formula	C12 H12 Br N3 O6 S
Formula weight	406.22
Temperature	100.0 K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	C 1 2 1
Unit cell dimensions	a = 21.582(3) Å $\quad\quad$ α= 90°.
	b = 5.6109(7) Å $\quad\quad$ β= 98.665(4)°.
	c = 12.5574(14) Å $\quad\quad$ γ = 90°.
Volume	1503.3(3) Å^3
Z	4
Density (calculated)	1.795 Mg/m^3
Absorption coefficient	2.909 mm^{-1}
F(000)	816
Crystal size	0.103 x 0.011 x 0.005 mm^3
Crystal color, habit	Colorless Needle
Theta range for data collection	1.640 to 26.440°.
Index ranges	-26<=h<=26, -7<=k<=7, -15<=l<=15
Reflections collected	27161
Independent reflections	3051 [R(int) = 0.0621]
Completeness to theta = 25.000°	99.6 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.0926 and 0.0654
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	3051 / 5 / 224
Goodness-of-fit on F^2	1.032
Final R indices [I>2sigma(I)]	R1 = 0.0251, wR2 = 0.0457
R indices (all data)	R1 = 0.0321, wR2 = 0.0469
Absolute structure parameter	0.039(4)
Extinction coefficient	n/a
Largest diff. peak and hole	0.417 and -0.408 e.Å^{-3}
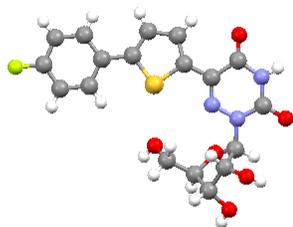
Figure 3.6 X-ray crystal structure of 5-(5-(4-fluorophenyl)thiophen-2-yl)-6-aza-uridine (**48**).

Table 3.3 Crystal data and structure refinement for tor85 5-(5-(4-fluorophenyl)thiophen-2-yl)-6-aza-uridine (**48**).

Identification code	PH188
Empirical formula	C37 H36 F2 N6 O13 S2
Formula weight	874.84
Temperature	100(2) K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	P 1
Unit cell dimensions	a = 9.4297(5) Å α= 74.895(2)°.
	b = 10.0067(4) Å β= 77.875(2)°.
	c = 10.3843(4) Å γ= 76.982(2)°.
Volume	909.69(7) Å^3
Z	1
Density (calculated)	1.597 Mg/m^3
Absorption coefficient	0.237 mm^{-1}
F(000)	454
Crystal size	0.290 x 0.080 x 0.040 mm^3
Theta range for data collection	2.058 to 26.443°.
Index ranges	-11<=h<=11, -12<=k<=12, -12<=l<=13
Reflections collected	18448
Independent reflections	7444 [R(int) = 0.0511]
Completeness to theta = 25.000°	100.0 %
Absorption correction	Multi-scan
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	7444 / 4 / 535
Goodness-of-fit on F^2	1.019
Final R indices [I>2sigma(I)]	R1 = 0.0401, wR2 = 0.0910
R indices (all data)	R1 = 0.0453, wR2 = 0.0944
Absolute structure parameter	0.03(4) [stereochem. confirmed]
Extinction coefficient	n/a
Largest diff. peak and hole	0.501 and -0.459 e.Å^{-3}
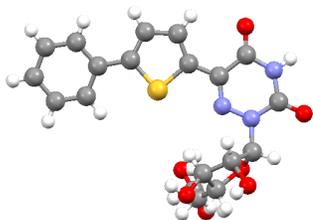
Figure 3.7 X-ray crystal structure of 5-(5-phenylthiophen-2-yl)-6-aza-uridine (**49**).

Table 3.4 Crystal data and structure refinement for tor93 5-(5-phenylthiophen-2-yl)-6-aza-uridine (**49**).

Identification code	PH189
Empirical formula	C18.50 H19 N3 O6.50 S
Molecular formula	C18 H17 N3 O6 S, 0.5(C H4 O)
Formula weight	419.43
Temperature	100.0 K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	C 1 2 1
Unit cell dimensions	a = 20.4544(12) Å α= 90°.
	b = 5.5733(3) Å β= 93.062(4)°.
	c = 16.1707(10) Å γ = 90°.
Volume	1840.80(19) Å^3
Z	4
Density (calculated)	1.513 Mg/m^3
Absorption coefficient	0.223 mm^{-1}
F(000)	876
Crystal size	0.053 x 0.005 x 0.005 mm^3
Crystal color, habit	Orange Needle
Theta range for data collection	1.994 to 25.732°.
Index ranges	-24<=h<=24, -6<=k<=6, -19<=l<=19
Reflections collected	18065
Independent reflections	3409 [R(int) = 0.1124]
Completeness to theta = 25.000°	99.9 %
Absorption correction	Semi-empirical from equivalents
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	3409 / 23 / 284
Goodness-of-fit on F^2	1.052
Final R indices [I>2sigma(I)]	R1 = 0.0708, wR2 = 0.1881
R indices (all data)	R1 = 0.0915, wR2 = 0.2038
Absolute structure parameter	0.03(19)
Extinction coefficient	n/a
Largest diff. peak and hole	0.550 and -0.493 e.Å^{-3}
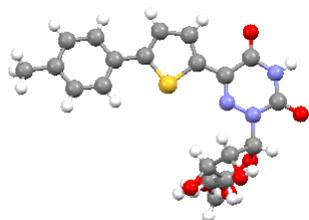
Figure 3.8 X-ray crystal structure of 5-(5-(4-methyl)phenyl)thiophen-2-yl)-6-aza-uridine (**50**).

Table 3.5 Crystal data and structure refinement for tor81 5-(5-(4-methyl)phenyl)thiophen-2-yl)-6-aza-uridine (**50**).

Identification code	PH187
Empirical formula	C20 H23 N3 O7 S
Formula weight	449.47
Temperature	100(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	P 21
Unit cell dimensions	a = 9.4923(7) Å α= 90°.
	b = 19.9718(15) Å β= 95.314(3)°.
	c = 10.7103(8) Å γ = 90°.
Volume	2021.7(3) Å^3
Z	4
Density (calculated)	1.477 Mg/m^3
Absorption coefficient	0.210 mm^{-1}
F(000)	944
Crystal size	0.300 x 0.100 x 0.050 mm^3
Theta range for data collection	2.039 to 28.349°.
Index ranges	-12<=h<=12, -26<=k<=26, -14<=l<=12
Reflections collected	31888
Independent reflections	10067 [R(int) = 0.0496]
Completeness to theta = 25.000°	99.9 %
Absorption correction	Multi-scan
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	10067 / 1 / 571
Goodness-of-fit on F^2	1.012
Final R indices [I>2sigma(I)]	R1 = 0.0444, wR2 = 0.1009
R indices (all data)	R1 = 0.0562, wR2 = 0.1077
Absolute structure parameter	0.00(3)
Extinction coefficient	n/a
Largest diff. peak and hole	0.771 and -0.401 e.Å^{-3}
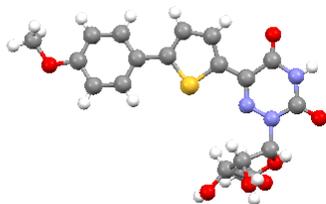
Figure 3.9 X-ray crystal structure of 5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (**51**).

Table 3.6 Crystal data and structure refinement for tor76 5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (**51**).

Identification code	PH176
Empirical formula	C20 H23 N3 O8 S
Formula weight	465.47
Temperature	100(2) K
Wavelength	1.54178 Å
Crystal system	Triclinic
Space group	P 1
Unit cell dimensions	a = 9.0722(10) Å α= 100.154(4)°.
	b = 10.3704(11) Å β= 103.516(4)°.
	c = 12.3104(14) Å γ = 105.629(4)°.
Volume	1048.5(2) Å^3
Z	2
Density (calculated)	1.474 Mg/m^3
Absorption coefficient	1.857 mm^{-1}
F(000)	488
Crystal size	0.340 x 0.080 x 0.060 mm^3
Theta range for data collection	3.820 to 68.341°.
Index ranges	-10<=h<=10, -10<=k<=12, -14<=l<=14
Reflections collected	8418
Independent reflections	5428 [R(int) = 0.0362]
Completeness to theta = 66.000°	95.7 %
Absorption correction	Multi-scan
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5428 / 3 / 582
Goodness-of-fit on F^2	1.042
Final R indices [I>2sigma(I)]	R1 = 0.0466, wR2 = 0.1257
R indices (all data)	R1 = 0.0475, wR2 = 0.1270
Absolute structure parameter	0.056(12)
Extinction coefficient	n/a
Largest diff. peak and hole	0.503 and -0.443 e.Å^{-3}
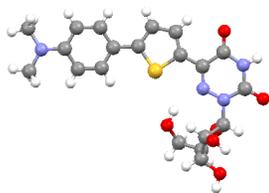
Figure 3.10 X-ray crystal structure of 5-(5-(4-(dimethylamino)phenyl)thiophen-2-yl)-6-aza-uridine (**53**).

Table 3.7 Crystal data and structure refinement for Tor91 5-(5-(4-(dimethylamino)phenyl)thiophen-2-yl)-6-aza-uridine (**53**).

Report date	2014-08-14
Identification code	Tor91
Empirical formula	C21 H25 N4 O6.50 S1.50
Molecular formula	C20 H22 N4 O6 S, 0.5(C2 H6 O S)
Formula weight	485.54
Temperature	100.0 K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	P 21 21 21
Unit cell dimensions	a = 8.9567(7) Å $\quad\quad$ α= 90°.
	b = 19.8870(16) Å $\quad\quad$ β= 90°.
	c = 24.7770(19) Å $\quad\quad$ γ = 90°.
Volume	4413.3(6) Å^3
Z	8
Density (calculated)	1.461 Mg/m^3
Absorption coefficient	0.244 mm^{-1}
F(000)	2040
Crystal size	0.377 x 0.035 x 0.031 mm^3
Crystal color, habit	Orange Needle
Theta range for data collection	2.048 to 26.413°.
Index ranges	-10<=h<=11, -24<=k<=24, -30<=l<=30
Reflections collected	29602
Independent reflections	9013 [R(int) = 0.0799]
Completeness to theta = 25.000°	99.9 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.0932 and 0.0655
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	9013 / 8 / 633
Goodness-of-fit on F^2	1.013
Final R indices [I>2sigma(I)]	R1 = 0.0530, wR2 = 0.0978
R indices (all data)	R1 = 0.0948, wR2 = 0.1119
Absolute structure parameter	-0.04(5)
Extinction coefficient	n/a
Largest diff. peak and hole	0.745 and -0.320 e.Å^{-3}

X-ray crystal structure: packing
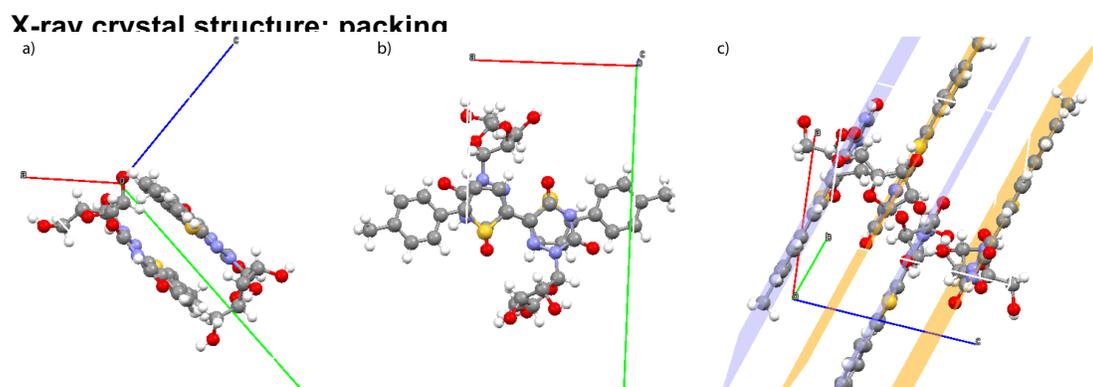
Figure 3.11 Crystal packing of derivative 5-(5-(4-methoxyphenyl)thiophen-2-yl)-6-aza-uridine (**51**) in three different views (a) side view (b) highlighting he overlap of the thiophene and the 6-aza-uridine rings (c) highlighting the aromatic layers.

Photophysics

Spectroscopic grade methanol and ethanol were obtained from Sigma Aldrich. Spectroscopic grade dioxane was obtained from Acros. Aqueous samples were prepared with de-ionized water. For all spectroscopic measurements a 1 cm four-sided Helma quartz cuvette was used. All spectroscopy samples were prepared from concentrated DMSO stock solutions, hence, all samples contain 0.4 v% or 0.2 v% DMSO.

Absorption spectra were measured on a Shimadzu UV-2450 UV-Vis spectrophotometer with 1 nm resolution and corrected for the blank. The sample temperature was kept constant at 20 °C using a thermostat.

Steady state emission and excitation spectra were taken on a PTI luminescence spectrometer with a 1 nm resolution. The sample temperature was kept constant at 20 °C with a Quantum Northwest TLC50 fluorescence cuvette holder in conjunction with a software controllable TC 125 temperature controller.

The polarity dependent steady state fluorescence studies were performed using an excitation wavelength (λ_{ex}) of:

	Dioxane-methanol		Water	
Sample	λ_{ex} (nm)	Concentration (µM)	λ_{ex} (nm)	Concentration (µM)
48	365	4.25	335	4.25
49	370	3.75	365	3.75
50	365	6.71	360	3.35
51	375	2.00	370	2.00
52	385	1.60	370	1.60
53	410	2.13	390	2.13
slit-widths	0.50 mm = 2.0 nm		1.00 mm = 4.0 nm	

Stokes shifts were calculated in cm^{-1} then converted to kcal/mol by multiplication with 0.0028591 to plot the polarity sensitivity correlations.

Samples' E$_T$(30) values (in kcal/mol) were determined by dissolving a small amount of Reichardt's dye in the solvent (mixture) used to prepare the sample.[13] The observed long wavelength absorption maximum in nm ($\lambda_{abs\,max}$) was converted to the sample E$_T$(30) value according the following equation:

$$E_T(30) = \frac{28591}{\lambda_{abs\,max}}$$

Quantum yields were determined using Coumarin-102 in ethanol (Φ = 0.80) as a standard for **48-52**, and Coumarin-153 in ethanol (Φ = 0.38) for **53** using dilute sample solutions with an O.D. < 0.05 at the λ_{ex}, using the following equation:

$$\Phi_s = \frac{I_s}{I_{ref}} \cdot \frac{O.D._{ref}}{O.D._s} \cdot \left(\frac{n_s}{n_{ref}}\right)^2 \cdot \Phi_{ref}$$

Here Φ, I, O.D. and *n* stand for quantum yield, integrated emission intensity, optical density at λ_{ex} and refractive index (n_{water} = 1.333, $n_{dioxane}$ = 1.42, $n_{methanol}$ = 1.326, $n_{ethanol}$ = 1.361), respectively. Sample and reference are denoted by *s* and *ref*, respectively. The λ_{ex} is in a very close proximity for the sample and reference solutions to circumvent correction of the difference in excitation energy at different wavelengths.

For dioxane solutions:

Sample	λ_{ex} (nm)
48	368
49	371
50	375
51	383
52	385
53	421

The excitation wavelengths for methanol and water solutions were kept the same as reported above for polarity dependent steady state fluorescence studies.

Sensitivity to polarity

Sensitivity to solvent polarity was studied in methanol, dioxane, and their mixtures (10, 20, 30, 40, 50, 60, 70, 80, and 90 v/v % of methanol in dioxane) at 20 °C. Emission spectra were recorded after excitation at the long wavelength emission maximum (O.D.'s < 0.1). The same samples were used to obtain excitation spectra probing at:

Sample	λ_{em} (nm)
48	470
49	480
50	475
51	480
52	500
53	560

All experiments were performed in triplicate with negligible differences; hence only one series is shown.

Absorption and steady-state emission spectroscopy for **48** in methanol/dioxane
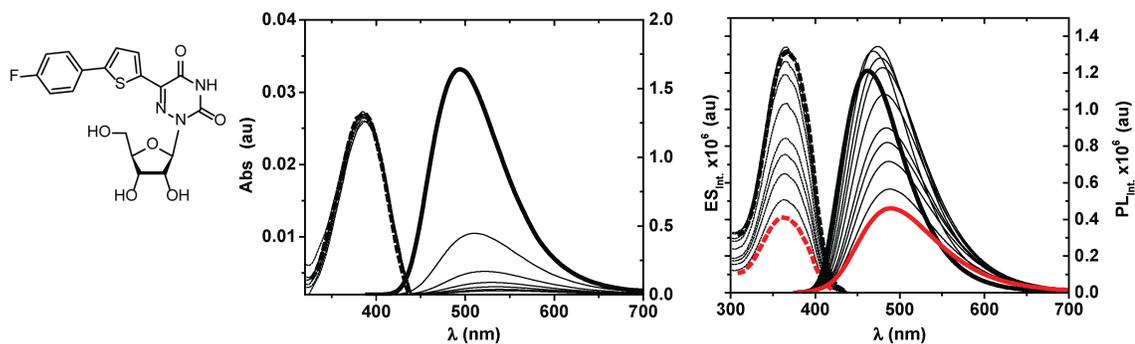
Figure 3.12 Assessing the effect of solvent polarity on absorption and excitation (dotted line), and emission (solid line), in dioxane (bold black line) and methanol (bold red line) and their mixtures (black lines).
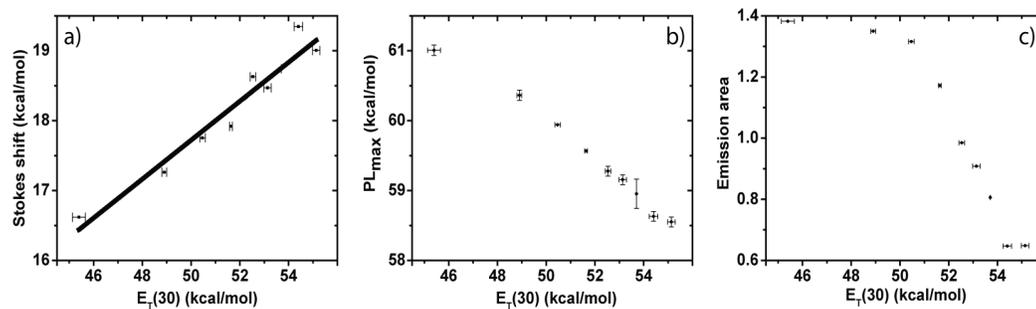
Figure 3.13 Correlating $E_T(30)$ vs. Stokes shift (slope:0.28 and R^2:0.97) (a), PL_{max} (b), emission area (c) values obtained from dioxane–methanol mixtures 90%:10% → 90%:10%.

Absorption and steady-state emission spectroscopy for **49** in methanol/dioxane
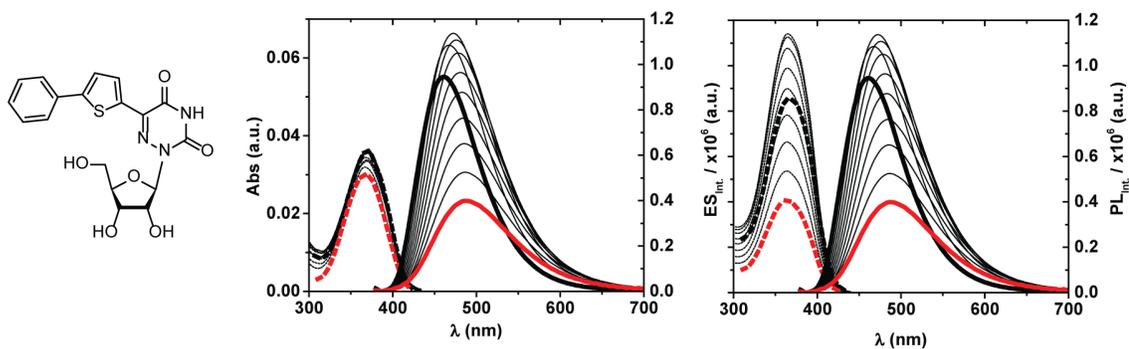
Figure 3.14 Assessing the effect of solvent polarity on absorption and excitation (dotted line), and emission (solid line), in dioxane (bold black line) and methanol (bold red line) and their mixtures (black lines).
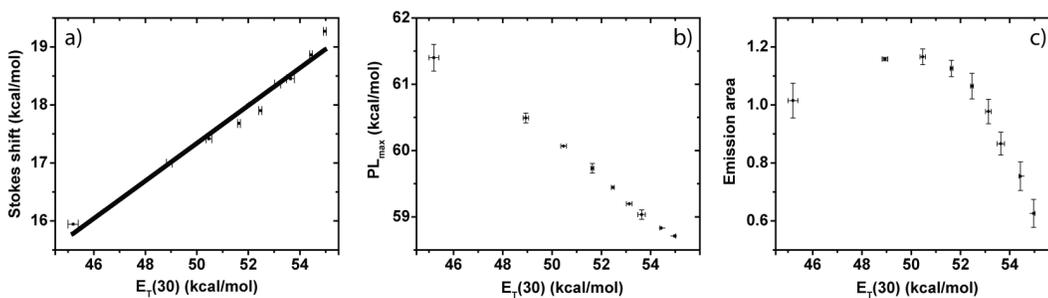
Figure 3.15 Correlating $E_T(30)$ vs. Stokes shift (slope:0.33 and R^2:0.99) (a), PL_{max} (b), emission area (c) values obtained from dioxane–methanol mixtures 90%:10% → 90%:10%.

Absorption and steady-state emission spectroscopy for **50** in methanol/dioxane
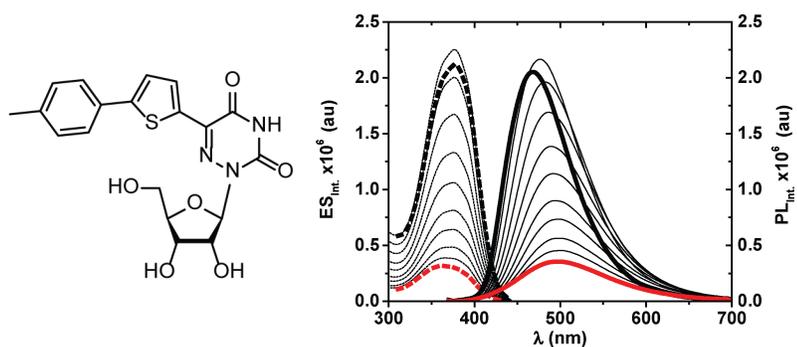
Figure 3.16 Assessing the effect of solvent polarity on excitation (dotted line), and emission (solid line), in dioxane (bold black line) and methanol (bold red line) and their mixtures (black lines).
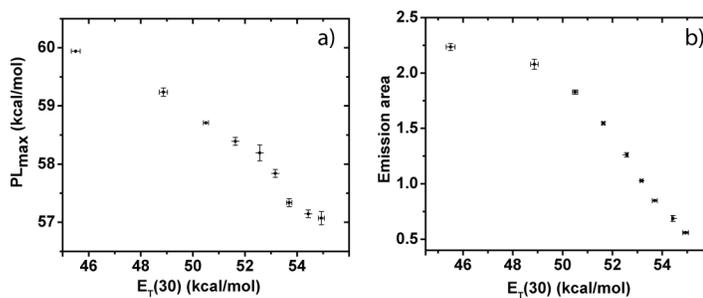
Figure 3.17 Correlating $E_T(30)$ vs. PL_{max} (a), emission area (b) values obtained from dioxane–methanol mixtures 90%:10% → 40%:60%.

Absorption and steady-state emission spectroscopy for **51** in methanol/dioxane
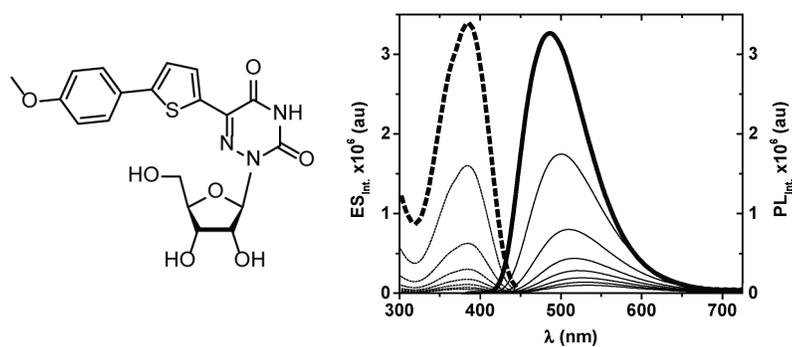
Figure 3.18 Assessing the effect of solvent polarity on excitation (dotted line), and emission (solid line), in dioxane (bold black line) and dioxane-methanol mixtures (black lines).
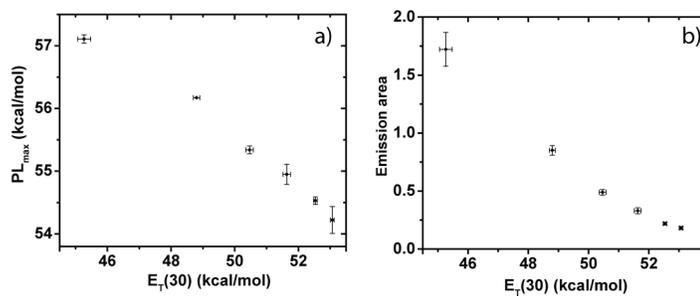
Figure 3.19 Correlating $E_T(30)$ vs. PL_{max} (a), emission area (b) values obtained from dioxane–methanol mixtures 90%:10% → 40%:60%.

Absorption and steady-state emission spectroscopy for **52** in methanol/dioxane
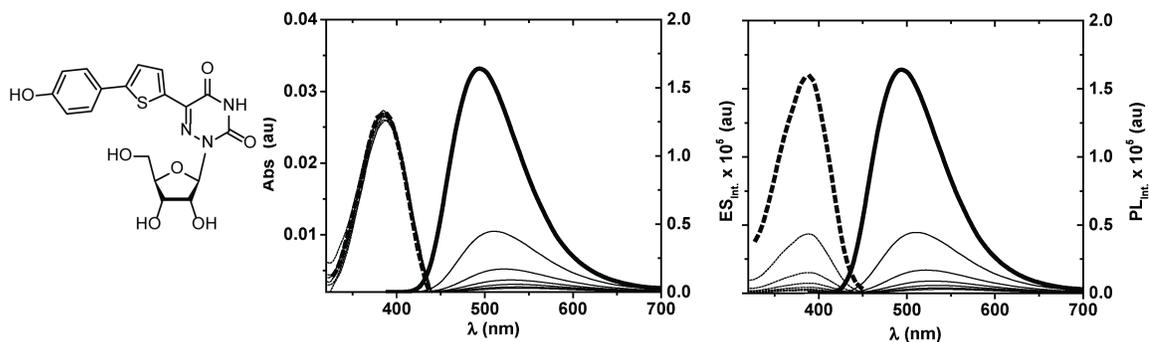
Figure 3.20 Assessing the effect of solvent polarity on absorption and excitation (dotted line), and emission (solid line), in dioxane (bold black line) and dioxane-methanol mixtures (black lines).
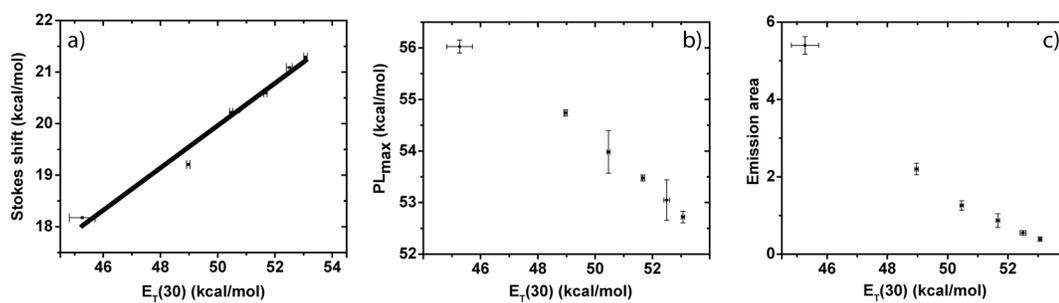
Figure 3.21 Correlating $E_T(30)$ vs. Stokes shift (slope:0.41 and R^2:0.99) (a), PL_{max} (b), emission area (c) values obtained from dioxane–methanol mixtures 90%:10% → 40%:60%.

Absorption and steady-state emission spectroscopy for **53** in methanol/dioxane
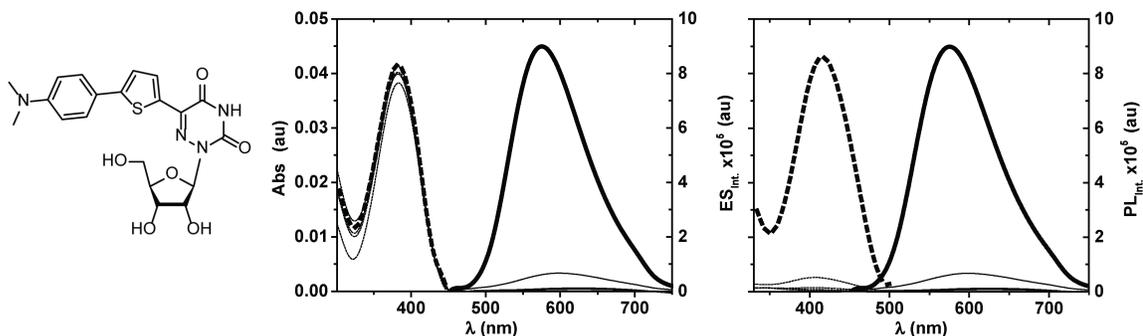
Figure 3.22 Assessing the effect of solvent polarity on absorption and excitation (dotted line), and emission (solid line), in dioxane (bold black line) and dioxane-methanol mixtures (black lines).
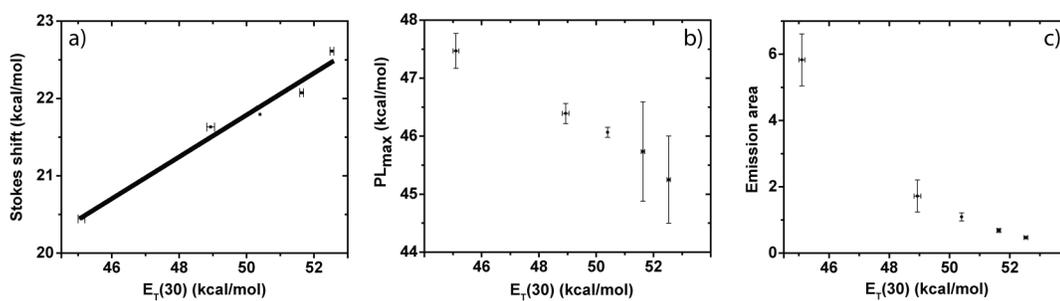
Figure 3.23 Correlating $E_T(30)$ vs. Stokes shift (slope:0.27 and R^2:0.99) (a), PL_{max} (b), emission area (c) values obtained from dioxane–methanol mixtures 90%:10% $\rightarrow$ 50%:50%.

Absorption and steady-state emission spectroscopy for **48**-**53** in water

The water steady state fluorescence studies were performed using an excitation wavelength (λ_{ex}) of:

Sample	λ_{ex} (nm)
48	355
49	365
50	360
51	370
52	370
53	390

The same samples were used to obtain excitation spectra probing at:

Sample	λ_{em} (nm)
48	450
49	450
50	475
51	490
52	475
53	475

All experiments were performed in triplicate with negligible differences; hence only one series is shown.
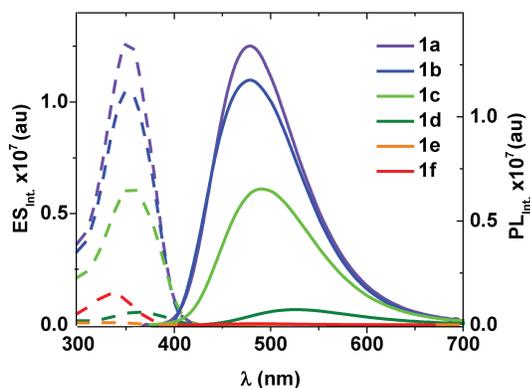
Figure 3.24 Excitation (dotted line) and emission (solid line) spectra for **48** (purple), **49** (blue), **50** (green), **51** (dark green), **52** (orange), **53** (red) in dioxane (a) and water (b) Emission was recorded after excitation at $\lambda_{abs\ max}$ for each derivatives (values in Table above). Slits width 1.00 mm.

Sensitivity to pH

Absorption and steady-state emission spectroscopy for **50**
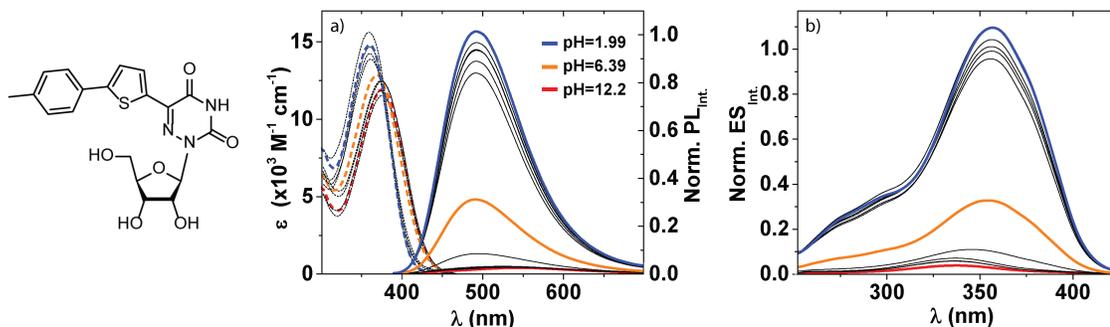
Figure 3.25 Assessing the effect of pH in aqueous buffers of pH 1.99 (red line), pH 6.39 (orange line) and pH 12.2 (blue line) and intermediate pH values (black lines) on (a) absorption (dotted line) and emission (solid line) (emission spectra were recorded after excitation at 370nm) and on (b) excitation (solid line) (spectra were recorded probing at 500nm).
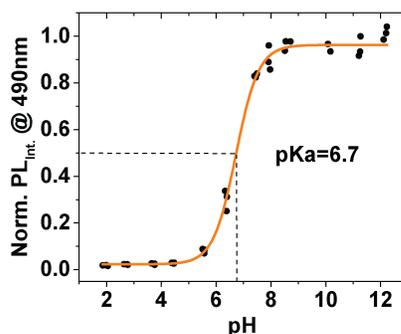
Figure 3.26 A plot of the normalized emission intensity as function of sample pH values (solid circles), with a sigmoidal fit (orange line) using OriginPro. The dashed lines illustrate a graphical determination of the pKa value (R^2: 0.99).

Enlarged Figure 3.3c and Figure 3.3d.
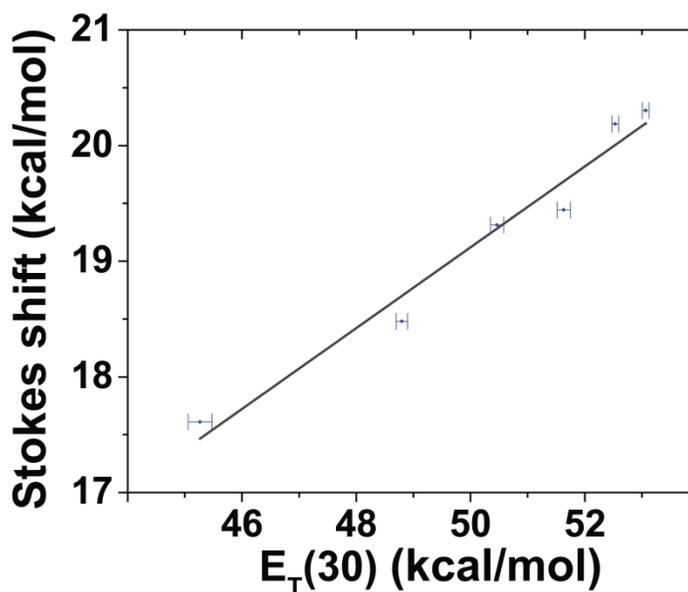
Figure 3.27 Correlating $E_T(30)$ vs. Stokes shift (slope:0.36 and R^2=0.95) values obtained from dioxane–methanol mixtures 90%:10% $\rightarrow$ 40%:60%.
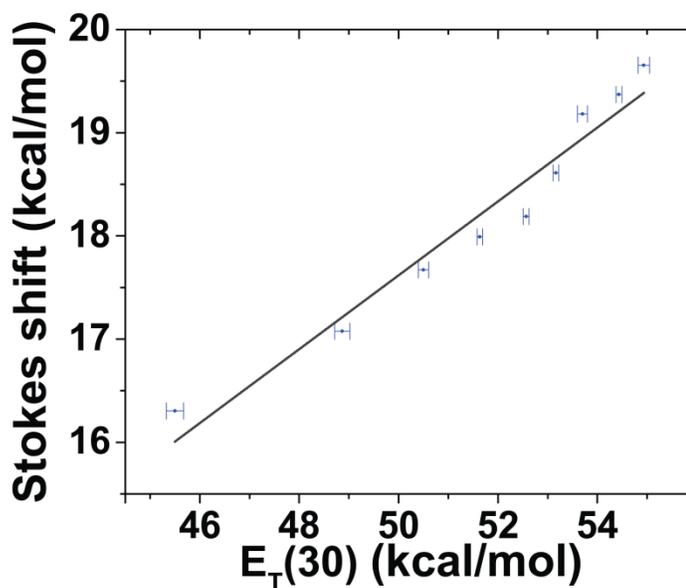
Figure 3.28 Correlating $E_T(30)$ vs. Stokes shift (slope:0.35 and R^2=0.98) values obtained from dioxane–methanol mixtures 90%:10% $\rightarrow$ 10%:90%.

Acknowledgment

We are grateful to Drs. Curtis Moore and Arnold Rheingold (UCSD Chemistry and Biochemistry X-ray Facility) for their help.

Chapter 3 has been previously published in the following article: Hopkins, P. A.; Sinkeldam, R. W.; Tor, Y. " Visibly Emissive and Responsive Extended 6‑Aza-Uridines " Org. Lett., 2014, 16, 5290. The dissertation author is the primary investigator and author of this article.

3.8 References

1. R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

2. L. M. Wilhelmsson, *Quarterly reviews of biophysics*, 2010, **43**, 159-183.

3. D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378-391.

4. M. J. Rist and J. P. Marino, *Current Organic Chemistry*, 2002, **6**, 775-793.

5. L. D. Lavis and R. T. Raines, *Acs Chem Biol*, 2008, **3**, 142-155.

6. K. V. Wood, Y. A. Lam, H. H. Seliger and W. D. Mcelroy, *Science*, 1989, **244**, 700-702.

7. B. R. Branchini, R. A. Magyar, M. H. Murtiashaw, S. M. Anderson, L. C. Helgerson and M. Zimmer, *Biochemistry-Us*, 1999, **38**, 13223-13230.

8. B. R. Branchini, T. L. Southworth, M. H. Murtiashaw, R. A. Magyar, S. A. Gonzalez, M. C. Ruggiero and J. G. Stroh, *Biochemistry-Us*, 2004, **43**, 7255-7262.

9. R. Y. Tsien, *Annu Rev Biochem*, 1998, **67**, 509-544.

10. R. M. Wachter, *Accounts Chem Res*, 2007, **40**, 120-127.

11. R. M. Wachter, *Photochem Photobiol*, 2006, **82**, 339-344.

12. T. Nakatsu, S. Ichiyama, J. Hiratake, A. Saldanha, N. Kobashi, K. Sakata and H. Kato, *Nature*, 2006, **440**, 372-376.

13. J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer, New York, 3rd edn., 2006.

14. S. B. Raymond, J. Skoch, I. D. Hills, E. E. Nesterov, T. M. Swager and B. J. Bacskai, *Eur J Nucl Med Mol Imaging*, 2008, **35 Suppl 1**, S93-S98.

15. S. T. Meek, E. E. Nesterov and T. M. Swager, *Org Lett*, 2008, **10**, 2991-2993.

16. J. Riedl, P. Menova, R. Pohl, P. Orsag, M. Fojta and M. Hocek, *J Org Chem*, 2012, **77**, 8287-8293.

17. S. T. Gaballah, G. Collier and T. L. Netzel, *J Phys Chem B*, 2005, **109**, 12175-12181.

18. C. Grunwald, T. Kwon, N. Piton, U. Forster, J. Wachtveitl and J. W. Engels, *Bioorgan Med Chem*, 2008, **16**, 19-26.

19. A. Okamoto, K. Tainaka and Y. Fujiwara, *J Org Chem*, 2006, **71**, 3592–3598.

20. N. J. Greco and Y. Tor, *J Am Chem Soc*, 2005, **127**, 10784–10785.

21. N. J. Greco and Y. Tor, *Tetrahedron*, 2007, **63**, 3515-3527.

22. R. W. Sinkeldam, P. A. Hopkins and Y. Tor, *Chemphyschem*, 2012, **13**, 3350-3356.

23. M. Kovaliov, M. Segal and B. Fischer, *Tetrahedron*, 2013, **69**, 3698-3705.

24. P. Capek, R. Pohl and M. Hocek, *Org Biomol Chem*, 2006, **4**, 2278-2284.

25. , See Supporting Information for additional details.

26. G. J. Zhao and K. L. Han, *Accounts Chem Res*, 2012, **45**, 404-413.

27. Y. H. Liu, G. J. Zhao, G. Y. Li and K. L. Han, *J Photoch Photobio A*, 2010, **209**, 181-185.

28. H. S. Joshi, R. Jamshidi and Y. Tor, *Angew Chem Int Edit*, 1999, **38**, 2722-2725.

29. M. W. Marina Kovaliov, Dan Thomas Major, Bilha Fischer, *J Org Chem*, 2014, **79**, 7051–7062.

30. C. Reichardt, *Chem Rev*, 1994, **94**, 2319–2358.

31. S. G. Srivatsan and Y. Tor, *Chem-Asian J*, 2009, **4**, 419-427.

32. J. Dhuguru, W. J. Liu, W. G. Gonzalez, W. M. Babinchak, J. Miksovska, R. Landgraf and J. N. Wilson, *J Org Chem*, 2014, **79**, 4940-4947.

Chapter 4   Development   of   the   Fluorescent 6-AzaUridine with Hydrogen Transfer in the Excited state

4.1 Introduction

Shifting the emission bands of fluorescent probes to the visible and red spectral domains, while maintaining isomorphicity, can be exceptionally challenging.[1, 2] Such low energy emission is frequently associated with extended chromophores, which are larger than the native nucleobases.[3] This chapter discusses the design of a synthetic uridine surrogate, which can undergo a proton transfer in the excited state, thus facilitating relatively large bathochromic shift while fulfilling the requirements of isomorphic nucleosides.

Augmentation of the acidity or basicity of a molecule upon photoexcitation has been documented by several research groups.[4-13] To describe compounds with enhanced acidity or basicity in the excited state the term "photo-acid" and "photo-base" could be used. These molecules either lose ("photo-acid") or gain ("photo-base") a proton upon absorption of a photon.[14, 15] Consequently, if a compound contains both acidic and basic groups in close proximity, a proton can be transferred in the excited state.[10] This process is referred to as Excited State Intramolecular Proton Transfer (ESIPT). Most commonly the acidic group consists of a hydroxyl group and the basic group can be a heterocyclic nitrogen, which are frequently H-bonded in the ground state. Therefore, proton transfer is very rapid since it involves little or no movement of the proton.
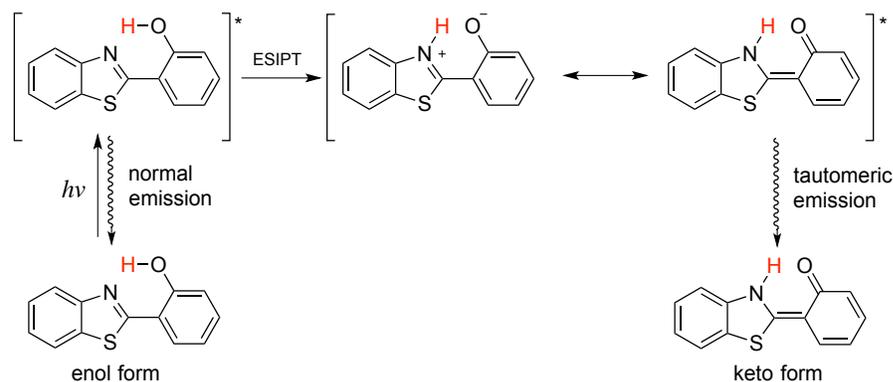
Figure 4.1 Origin of the dual emission. Excited State Intramolecular Proton Transfer (ESIPT).

ESIPT chromophores typically exist in the enol form at the ground state. Upon photoexcitation, they exhibit a normal excited state (enol form), which then undergoes ultrafast ($k_{ESIPT} > 10^{12}$ s^{-1}) ESIPT to produce a tautomeric excited state (keto form). Therefore, two distinct emission bands can be observed, namely the tautomeric band being typically 80–100 nm red shifted compared to the normal band. The large Stokes shift is the most remarkable photophysical property of ESIPT fluorophores, which is rather difficult to increase in conventional fluorophores.

4.2 Design

Notably, most of the ESIPT chromophores are related to structures such as HBT (**84**)[16], and very few new designs have been reported (**87**–**89**).[17-19] Here, we introduce a 5-substituted phenyl along with a judiciously placed hydroxyl group, which can H-bond to the basic site of the 6-aza-uridine core, which should facilitate the ESIPT process (Figure 4.1b). We have hypothesized that the N6 of the 6-aza-uridine core can act as a "photo-base" and the hydroxyl group being in a close

proximity can act as a "photo-acid" upon absorption of a photon of light. Therefore, ESIPT can occur upon excitation and a desired emission and large Stokes shift should be observed. Here, we reveal the synthesis of the ESIPT probe (**54**) and its parent nucleoside lacking the hydroxyl group (**55**), as a reference compound.
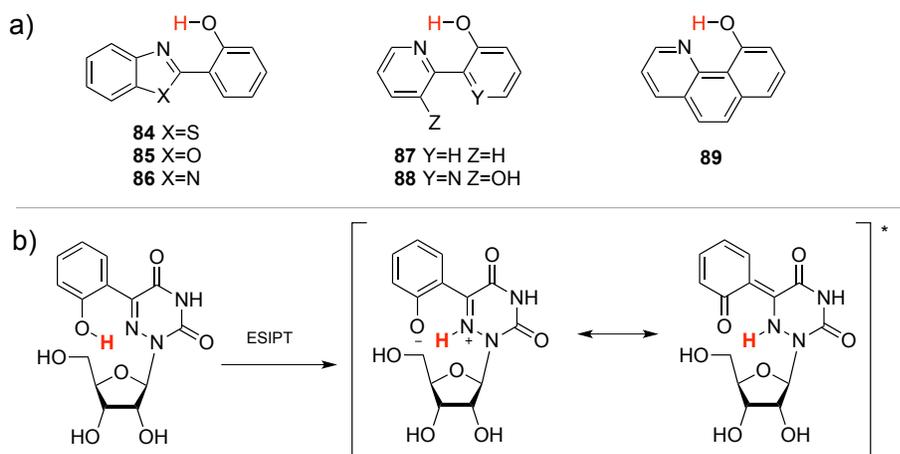
Figure 4.2 a) ESIPT based chromophores; b) Design elements leading to ESIPT 6-aza-uridines.

4.3 Synthesis

Glycosylation of commercially available 5-bromo-6-azauracil (**90**) was performed using a known protocol.[20] The nucleobase was activated with BSA for 30 min at room temperature. Subsequently, the reaction mixtures was brought to 85 °C and β-D-ribofuranose 1-acetate 2,4,5-tribenzoate (**63**) was added, immediately followed by addition of TMSOTf to yield glycosylated product **91** in good yield (80%). Deprotection with methanolic ammonia at 60 °C yielded substrate **92**. Both nucleosides **54** and **55** were obtained via Suzuki coupling reactions between **92** and boronic acids **93** and **94**, using a combination of the water soluble ligand TPPTS and palladium acetate as a catalyst.[21-23] All products

were thoroughly characterized by mass spectroscopy and ^{1}H, ^{13}C NMR spectroscopy (see experimental).
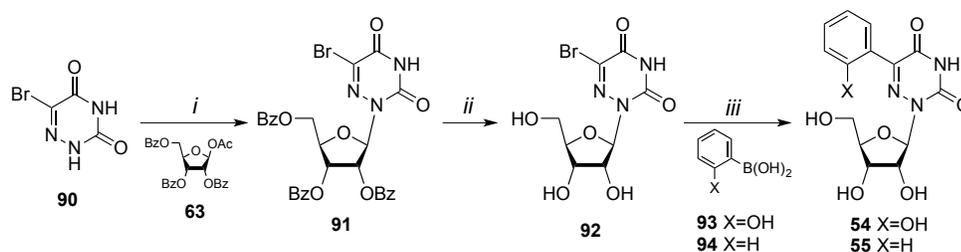
Scheme 4.1 Synthesis of nucleosides **54** and **55**, reagents and conditions: *i*) BSA (3 equiv), acetonitrile, rt, 30 min, 63 (1.1 equiv), TMSOTf (1.1 equiv) 85 °C 30min; *ii*) NH$_3$/MeOH 60 °C 18 h; *iii*) boronic acid 93-94 (1.2 equiv), Pd(OAc)$_2$ (5% mol), TPPTS (0.25 equiv), Na$_2$CO$_3$ (3 equiv), H$_2$O:MeCN (3:2) 60 °C, 6 h.

4.4 Photophysical properties

To evaluate the influence of the hydroxyl group near N6 basic site on the photopysical properties of **54** excitation and emission spectra were recorded in dioxane, ethyl acetate and acetonitrile (Figure 4.3, Table 4.1). Consequently, **55** was used as a reference compound. The absorption maxima are relatively similar in aprotic solvents for both probes **54** and **55** and vary between 300–310 nm. The emission maxima of **54** are significantly redshifted compared to the parent 6-aza-uridine **55**. The large Stokes shifts (10,000–12,000 cm^{-1}) observed only for **54** support our hypothesis that upon excitation proton transfer can occur when hydroxyl group is in a close proximity to 6-aza. Preliminary results showed that emission of **54** is drastically quenched in protic solvents compared to **55**.
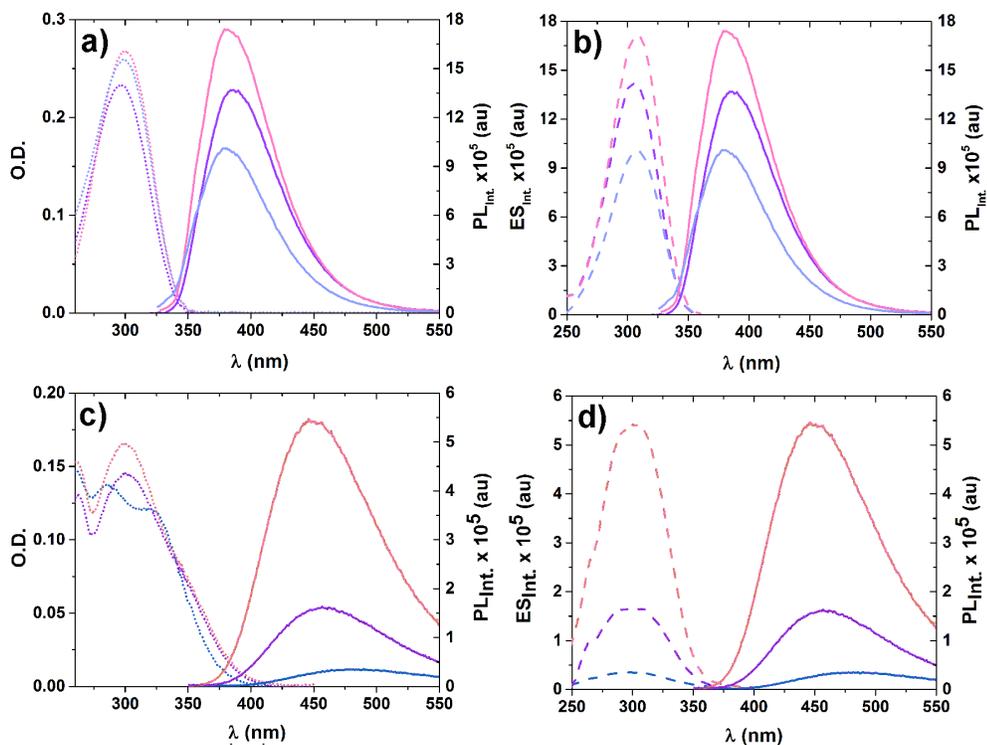
Figure 4.3 Absorption (dotted line), excitation (dash line) and emission (solid line) spectra for **55** (a-b) for **54** (c-d); dioxane (pink), acetonitrile (blue) and ethyl acetate (purple). Emission was recorded after excitation at $\lambda_{abs\,max}$ for each derivative (Table 4.1).

Table 4.1 Selected spectroscopic properties of **54** and **55**.

	Solvent	Excitation λ_{max} (nm)	Fluorescence λ_{max} (nm)	Stokes shift (nm)	x 10^3(cm^{-1})
54	Dioxane	301	449	148	10.9
	Acetonitrile	300	457	157	11.4
	Ethyl acetate	300	482	182	12.6
55	Dioxane	308	379	71	6.1
	Acetonitirle	306	386	80	6.8
	Ethyl acetate	309	378	69	5.9

4.5 Conclusion and future direction

The 6-aza-uridine motif, decorated at the 5-position with a phenol is a synthetically accessible pyrimidine analog. Preliminary spectroscopic studies suggest that **54** can undergo excited state intramolecular proton transfer upon excitation. Although relatively large Stokes shifts were observed a further investigation is necessary. Particularly time resolved studies, where fluorescence of **54** is monitored as a function of time after excitation, and lifetime can be calucalted from a fluorescence decay. This can indicate if the lifetime of the tautomeric keto form is longer than of the enol form, supporting presence of the ESIPT process. Furthermore, another isomer of 6aza-uridine substituted at the 5-position with the 4-phenol ring needs to be synthesized. Studying the photophysical properties of this isomer would ultimately allowed us to compare the electronic effect of the hydroxyl group on the emission maxima opposed to the ESIPT mechanism.

If proven ESIPT mechanism for **54**, this probe would be the first isomorphic nucleoside, which undergoes excited state intramolecular proton transfer, making **54** one of the most attractive and promising emissive uridines.

4.6 Experimental

General

N,*O*-bis(trimethylsilyl)acetamide (BSA) and dry acetonitrile were purchased from Sigma-Aldrich. Trimethylsilyl-O-triflate was purchased from Acros. Semicarbazide

hydrochloride and 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose were purchased from TCI. 5-bromo-6-azauracil and boronic acids were purchased from Combi-Blocks. TPPTS was purchased from Matrix Scientific, palladium acetate was purchased from Sigma Aldrich. All reagents and solvents were used without further purification. Moister and oxygen sensitive reactions were performed in an inert argon atmosphere.
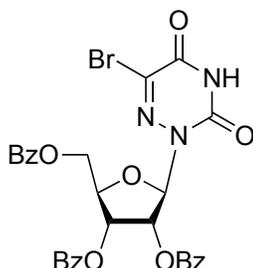
Synthesis

Synthesis of 5-bromo-6-aza-2',3',5'-tribenzoyl-uridine (91). Under argon, *N*,*O*-bis(trimethylsilyl)acetamide (15.3 mL, 60.2 mmol) was added to a stirring suspension of 5-bromo-6-azauracil (4.00 g, 20.8 mmol) in dry acetonitrile (40.0 mL). The suspension turns into a clear solution after stirring at rt for 30 min followed by heating to 85°C and addition of 1-*O*-acetyl-2,3,5-tri-*O*-benzoyl-β-D-ribofuranose (11.6 g, 22.9 mmol) and trimethylsilyl trifluoromethanesulfonate (4.1 mL, 22.9 mmol). The reaction mixture was stirred at 85°C under argon until monitoring by TLC indicated full conversion of starting material (~30 min.). The reaction mixture is cooled to rt diluted with DCM (1.5 L) and washed with water (3 × 0.5 L). The combined organic layers are dried over Na_2SO_4 filtered over a glass frit and concentrated to dryness. Purification of the crude residue by column

chromatography (silica, 20 *v*% EtOAc in DCM) yielded **91** as a white foam (11.8

g, 18.5 mmol, 80%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.69 (s, 1H), 8.00 – 7.82

(m, 6H), 7.63 – 7.33 (m, 9H), 6.41 (d, *J* = 3.3 Hz, 1H), 6.01 – 5.81 (m, 2H), 4.83 –

4.72 (m, 1H), 4.66–4.50 (m, 2H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 165.88, 165.00,

164.97, 153.63, 148.55, 134.44, 134.31, 134.01, 131.51, 129.91, 129.76, 129.72,

129.60, 129.30, 129.28, 129.16, 128.93, 128.87, 88.04, 79.09, 73.72, 70.98,

63.93; HR-MS-ESI: m/z calcd. for $C_{29}H_{22}BrN_3O_9$ (M–H$^-$): 635.0435, found: (M–H$^-$)

635.0432, (M–Na$^+$): 658.0435
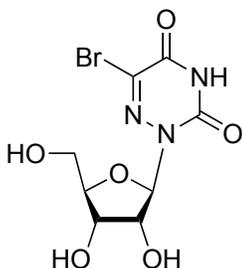
Synthesis of 5-(thiophen-2-yl)-6-aza-uridine (92). A solution of **91** (4.9 g, 7.7

mmol) in methanol (70 mL) and 27-30% ammonia (30mL) is heated at 80 °C for

44h. The solution was concentrated *in vacuo*. The precipitated is dissolved in water

(500 mL) and then water layer is washed with DCM (3 × 100 mL) to give **92** (1.3 g,

4.0 mmol, 52%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 12.58 (s, 1H), 5.83 (d, *J* = 3.4

Hz, 1H), 5.30 (d, *J* = 5.1 Hz, 1H), 5.07 (d, *J* = 6.0 Hz, 1H), 4.68 – 4.56 (m, 1H),

4.19 – 4.11 (m, 1H), 4.06 – 3.96 (m, 1H), 3.84 – 3.72 (m, 1H), 3.55 – 3.45 (m, 1H),

3.45 – 3.35 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 153.61, 148.78, 129.93,

90.00, 85.10, 73.15, 70.56, 62.30; HR-MS-ESI: m/z calcd. for $C_8H_{10}BrN_3O_6$ (M–H$^-$) 321.9682, found: 321.9680.
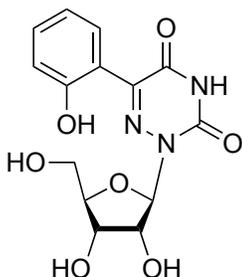
Synthesis of 5-(2hydroxyphenyl)-6-aza-uridine (54).

All solvents were purged with argon. Under argon, sodium carbonate (490.6 mg, 4.6 mmol) was dissolved in water (6 mL). Then **92** (500.0 mg, 1.5 mmol), (2-hydroxyphenyl)boronic acid **93** (319.2 mg, 2.3 mmol) and dry acetonitrile (4 mL) were added to the solution and additionally purged with argon for 15 minutes. Tris(3-sulfophenyl)phosphine trisodium salt (219.2 mg, 0.39 mmol) and palladium acetate (17.3 mg, 0.08 mmol) were added and the reaction mixture was heated and kept at 70 °C for 18h. The mixture was cooled to rt, neutralized with saturated aqueous solution of ammonium chloride and diluted with methanol (50 mL). A black precipitate was filtrate off, and remaining filtrate was concentrated to dryness. Purification of the crude residue by column chromatography (silica, 20 v% MeOH in DCM) yielded **54** as a white foam (181.3 mg, 0.05 mmol, 38%). ^{1}H NMR (500 MHz, DMSO-d$_6$) δ 12.24 (s, 1H), 9.62 (s, 1H), 7.35 – 7.15 (m, 2H), 7.00 – 6.76 (m, 2H), 5.95 (d, J = 3.2 Hz, 1H), 5.27 (d, J = 5.1 Hz, 1H), 4.99 (d, J = 6.1 Hz, 1H), 4.70 – 4.56 (m, 1H), 4.31 – 4.17 (m, 1H), 4.08 – 3.94 (m, 1H), 3.86 –

3.74 (m, 1H), 3.61 – 3.44 (m, 1H), 3.40 – 3.34 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d_6) δ 156.27, 156.05, 149.26, 144.25, 131.29, 131.06, 120.15, 119.12, 116.38, 89.93, 84.94, 73.15, 70.71, 62.33; HR-MS-ESI: m/z calcd. for $C_{14}H_{15}N_3O_7$ (M-H)$^-$: 336.0837, found: 336.0837
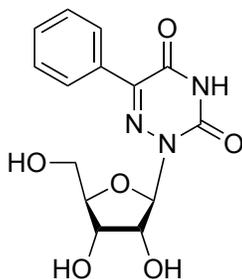
Synthesis of 5-phenyl-6-aza-uridine (55). Using the procedure described for **54**, but starting from **92** and phenyl boronic acid **94**; **55** was obtained as a white solid (245.2 mg, 0.76 mmol, 49%). ^{1}H NMR (400 MHz, DMSO-d_6) δ 12.32 (s, 1H), 7.96 – 7.82 (m, 2H), 7.50 – 7.40 (m, 3H), 5.99 (d, J = 3.2 Hz, 1H), 5.29 (d, J = 5.0 Hz, 1H), 5.03 (d, J = 6.2 Hz, 1H), 4.70 – 4.60 (m, 1H), 4.35 – 4.25 (m, 1H), 4.20 – 4.12 (m, 1H), 3.86 – 3.76 (m, 1H), 4.62 – 4.48 (m, 1H), 4.46 – 4.36 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d_6) δ 156.54, 149.04, 142.14, 132.56, 130.24, 128.71, 128.61, 89.88, 84.89, 73.35, 70.61, 62.00; MS-ESI: m/z calcd. for $C_{14}H_{15}N_3O_6$ (M-H$^-$) 320.0887, found 320.0888

Crystal structures

Experimental Summary

The single crystal X-ray diffraction studies were carried out on a Bruker Kappa APEX-II CCD diffractometer equipped with Cu K$_\alpha$ radiation (λ = 1.5478). A 0.213

x 0.053 x 0.031 mm piece of a colorless needle was mounted on a Cryoloop with Paratone oil. Data were collected in a nitrogen gas stream at 100(2) K using ϕ and ϖ scans. Crystal-to-detector distance was 50 mm using variable exposure time (1s-5s) depending on θ with a scan width of 1.0°. Data collection was 99.8% complete to 68.00° in θ. A total of 10353 reflections were collected covering the indices, -8<=h<=8, -14<=k<=16, -17<=l<=18. 2766 reflections were found to be symmetry independent, with a R_{int} of 0.0555. Indexing and unit cell refinement indicated a primitive, orthorhombic lattice. The space group was found to be $P2_12_12_1$. The data were integrated using the Bruker SAINT software program and scaled using the SADABS software program. Solution by direct methods (SHELXT) produced a complete phasing model consistent with the proposed structure.

All nonhydrogen atoms were refined anisotropically by full-matrix least-squares (SHELXL-2014). All carbon and oxygen bonded hydrogen atoms were placed using a riding model. Their positions were constrained relative to their parent atom using the appropriate HFIX command in SHELXL-2014. All nitrogen bonded hydrogen atoms were located in the difference map. There relative positions were restrained using DFIX commands and their thermals freely refined. Crystallographic data are summarized in Table 4.2.
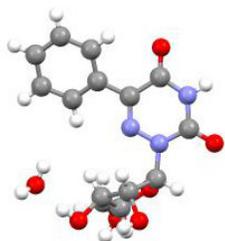
Figure 4.4 X-ray crystal structure of 5-phenyl-6-aza-uridine (**55**).

Table 4.2 Crystal data and structure refinement for Tor97 5-phenyl-6-aza-uridine (**55**).

Report date	2014-10-23
Identification code	PH207
Empirical formula	C14 H17 N3 O7
Molecular formula	C14 H15 N3 O6, H2 O
Formula weight	339.30
Temperature	100.0 K
Wavelength	1.54178 Å
Crystal system	Orthorhombic
Space group	P 21 21 21
Unit cell dimensions	a = 7.1137(5) Å α= 90°.
	b = 14.0400(9) Å β= 90°.
	c = 15.1922(11) Å γ = 90°.
Volume	1517.34(18) Å^3
Z	4
Density (calculated)	1.485 Mg/m^3
Absorption coefficient	1.033 mm^{-1}
F(000)	712
Crystal size	0.213 x 0.053 x 0.031 mm^3
Crystal color, habit	Colorless Needle
Theta range for data collection	4.288 to 68.182°.
Index ranges	-8<=h<=8, -14<=k<=16, -17<=l<=18
Reflections collected	10353
Independent reflections	2766 [R(int) = 0.0555]
Completeness to theta = 68.000°	99.8 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.3201 and 0.2247
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	2766 / 5 / 241
Goodness-of-fit on F^2	1.063
Final R indices [I>2sigma(I)]	R1 = 0.0322, wR2 = 0.0721
R indices (all data)	R1 = 0.0396, wR2 = 0.0751
Absolute structure parameter	-0.05(12)
Extinction coefficient	n/a
Largest diff. peak and hole	0.149 and -0.185 e.Å^{-3}

Photophysics

Spectroscopic grade ethyla acetate and acetonitrile were obtained from Sigma Aldrich. Spectroscopic grade dioxane was obtained from Acros. For all spectroscopic measurements a 1 cm four-sided Helma quartz cuvette was used. All spectroscopy samples were prepared from concentrated acetonitrile stock solutions.

Absorption spectra were measured on a Shimadzu UV-2450 UV-Vis spectrophotometer with 1 nm resolution and corrected for the blank. The sample temperature was kept constant at 20 °C using a thermostat.

Steady state emission and excitation spectra were taken on a PTI luminescence spectrometer with a 1 nm resolution. The sample temperature was kept constant at 20 °C with a Quantum Northwest TLC50 fluorescence cuvette holder in conjunction with a software controllable TC 125 temperature controller.

The steady state fluorescence studies were performed using an excitation wavelength (λ_{ex}) of:

a) for sample **54**:

Solvent	λ_{ex} (nm)
Dioxane	300
Acetonitrile	300
Ethyl acetate	288
slit-widths	0.75 mm = 3.0 nm

b) for sample **55**:

Solvent	λ_{ex} (nm)
Dioxane	308
Acetonitrile	303
Ethyl acetate	306
slit-widths	0.75 mm = 3.0 nm

4.7 References

1.   R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

2.   A. Fin, A. R. Rovira, P. A. Hopkins and Y. Tor, in *Modified Nucleic Acids*, eds. K. Nakatani and Y. Tor, Springer International Publishing, Cham, 2016, DOI: 10.1007/978-3-319-27111-8_1, pp. 1-26.

3.   L. D. Lavis and R. T. Raines, *Acs Chem Biol*, 2008, **3**, 142-155.

4.   A. Weller, *Prog React Kinet Mec*, 1961, **1**, 187-214.

5.   I. J. F. and W. P. A. H., *Advances in Physical Organic Chemistry* 1976, **12**, 131-221.

6.   J. Goodman and L. E. Brus, *J Am Chem Soc*, 1978, **100**, 7472-7474.

7.   P. F. Barbara, P. K. Walsh and L. E. Brus, *J Phys Chem-Us*, 1989, **93**, 29-34.

8.   A. Douhal, F. Lahmani and A. H. Zewail, *Chem Phys*, 1996, **207**, 477-498.

9.   J. S. Wu, W. M. Liu, J. C. Ge, H. Y. Zhang and P. F. Wang, *Chem Soc Rev*, 2011, **40**, 3483-3495.

10.   J. Z. Zhao, S. M. Ji, Y. H. Chen, H. M. Guo and P. Yang, *Physical Chemistry Chemical Physics*, 2012, **14**, 8803-8817.

11.   A. P. Demchenko, K. C. Tang and P. T. Chou, *Chem Soc Rev*, 2013, **42**, 1379-1408.

12.   P. T. Edlefsen, M. Rolland, T. Hertz, S. Tovanabutra, A. J. Gartland, A. C. deCamp, C. A. Magaret, H. Ahmed, R. Gottardo, M. Juraska, C. McCoy, B. B. Larsen, E. Sanders-Buell, C. Carrico, S. Menis, M. Bose, M. A. Arroyo, R. J. O'Connell, M. S. deSouza, S. Nitayaphan, P. Pitisuttithum, J. Kaewkungwal, S. Rerks-Ngarm, M. L. Robb, J. S. McLellan, I. S. Georgiev, T. Kirys, P. D. Kwong, J. M. Carlson, N. L. Michael, W. R. Schief, J. I. Mullins, J. H. Kim, P. B. Gilbert and R. S. Team, *Aids Res Hum Retrov*, 2014, **30**, A25-A26.

13.   V. S. Padalkar and S. Seki, *Chem Soc Rev*, 2016, **45**, 169-202.

14.   R. G. B. S. M. Ormson *Prog React Kinet*, 1994, **19**, 45-91.

15. D. Legourrierec, S. M. Ormson and R. G. Brown, *Prog React Kinet*, 1994, **19**, 211-275.

16. K. Sakai, T. Ishikawa and T. Akutagawa, *J Mater Chem C*, 2013, **1**, 7866-7871.

17. W. T. Chuang, C. C. Hsieh, C. H. Lai, C. H. Lai, C. W. Shih, K. Y. Chen, W. Y. Hung, Y. H. Hsu and P. T. Chou, *J Org Chem*, 2011, **76**, 8189-8202.

18. D. Y. Han, J. M. Kim, J. Kim, H. S. Jung, Y. H. Lee, J. F. Zhang and J. S. Kim, *Tetrahedron Lett*, 2010, **51**, 1947-1951.

19. K. C. Wu, Y. S. Lin, Y. S. Yeh, C. Y. Chen, M. O. Ahmed, P. T. Chou and Y. S. Hon, *Tetrahedron*, 2004, **60**, 11861-11868.

20. R. W. Sinkeldam, P. A. Hopkins and Y. Tor, *Chemphyschem*, 2012, **13**, 3350-3356.

21. P. A. Hopkins, R. W. Sinkeldam and Y. Tor, *Org Lett*, 2014, **16**, 5290-5293.

22. M. Kovaliov, M. Segal and B. Fischer, *Tetrahedron*, 2013, **69**, 3698-3705.

23. P. Capek, R. Pohl and M. Hocek, *Org Biomol Chem*, 2006, **4**, 2278-2284.

Chapter 5   Enzymatic Incorporation and Utilization of an Emissive 6-Aza Uridine

5.1 Introduction

The sensitivity and simplicity of fluorescence spectroscopy has made it one of the most invaluable techniques for exploring biomolecules, their structure and dynamics as well as their interactions with diverse ligands.[1, 2]   Unlike many proteins, nucleic acids are not intrinsically emissive. The non-emissive pyrimidine and purine nucleobases need to be modified to display usable fluorescent properties.[1] Diverse approaches for emissive nucleosides have indeed been devised over the years,[1, 3-8] but only a small fraction of the analogues made satisfies the criterion of isomorphicity.[9-13] Isomorphic nucleobases need to closely resemble their native counterparts with respect to their size, shape and hydrogen bonding pairing abilities. While only few useful purine analogues have been reported,[1, 10, 13-18] the largest number of chemical variations reported are for the pyrimidines, particularly 5-modified uridine analogues.[19] Several groups, including ours, have investigated a variety of isomorphic nucleosides whose design relied on conjugating five-membered rings such as thiophene (**29**),[11, 20, 21] furan (**38**),[9, 21-27] thiazole (**40**)[21, 27] and oxazole (**39**)[9, 24] to the pyrimidine core. The 5-subsituted uridine cores extend the chromophore, frequently augmenting the charge separation in the excited state, yielding useful fluorophores (Figure 5.1a). Interestingly, markedly fewer 6-modified uridines have been utilized in a similar manner.

Modifying the pyrimidine's 5 and 6 positions by judiciously placing conjugated electron poor/electron rich substituents can further enhance the chromophore's "push-pull" interactions, allowing one to potentially tune their photophysical features (Figure 5.1a). In our primary design, we have replaced the uracil core with the more electron-deficient 6-azauracil, while keeping the electron-rich thiophene ring expecting red-shifted absorption and emission maxima (**46,** Figure 5.1b). This indeed has been confirmed to be the case.[12] To explore the utility of such a modification within oligonucleotides we have sought to explore the enzymatic incorporation of the corresponding triphosphates. However, enzymatic incorporation of 6-azauridine triphosphate has been reported to be extremely inefficient[28-30] and has been proposed to potentially be due to the nucleoside's favorable *syn* conformation.[31-33]

We have previously observed enhanced T7 RNA polymerase-mediated incorporation of 5-thiopheneUTP (**98**).[11] We therefore hypothesized that conjugating a thiophene ring to 6-azaUTP at position 5 may partially overcome the detrimental effect of the N substituent at the 6 position, potentially providing "polymerase-friendly" substrate (**96**), which will allow the enzymatic fabrication of fluorescent RNAs. Here, we report the synthesis of this triphosphate and its enzymatic incorporation into oligonucleotides (Figure 5.1b). The impact of the "6-aza modification" on *in vitro* transcription reactions was investigated by comparing the performance of native UTP, 6-azaUTP (**97**) and 5-thiopheneUTP (**98**). We illustrate that the modification at the 5 position indeed significantly improves the enzymatic incorporation with essentially doubled transcription efficiency for 5-

thiophene-6azaUTP (**96**) compared to 6-azaUTP (**97**). We further demonstrate

that the resulting modified transcript can be ligated to unmodified oligonucleotides

to provide longer, fluorescent RNA constructs. As a proof of concept, a singly

modified A-site model RNA hairpin is assembled and used to monitor the binding

of aminoglycoside antibiotics. The emissive 5-thiophene-6-azauridine can probe

antibiotics binding either directly or via FRET-monitored displacement titrations.

We conclude that the efficient enzymatic incorporation of this 5-modified 6-azaUTP

and its responsiveness to environmental changes make it an attractive reporter

nucleoside for exploring RNA and its interactions.
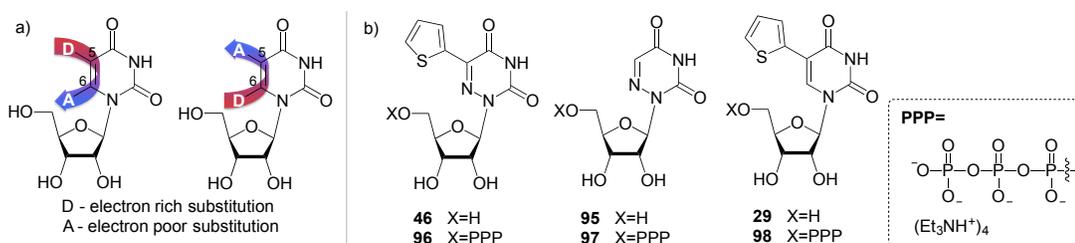
Figure 5.1 a) Nucleoside design: enhancing the polarization of the conjugated electron poor/electron rich biaryl system. b) Modified uridines: 5-(thiophen-2-yl)-6-azauridine (**46**), 6-azauridine (**95**), 5-(thiophen-2-yl)uridine (**29**) and their corresponding triphosphates **96**, **97**, and **98**.

5.2 Triphosphate synthesis

The parent nucleoside 5-(thiophen-2-yl)-6-azauridine **46** was synthesized

as reported.[12] It was then converted to the 5'-triphosphate **96** using $POCl_3$ and

tributylammonium pyrophosphate (Scheme 5.1), following standard procedures.[11, 34, 35] The analytically pure triphosphate **96** was obtained by ion-exchange

chromatography and reverse-phase HPLC. After precipitation with 3% sodium

perchlorate in acetone and treatment with Chelex 100, nucleotide **96** was

characterized by mass spectrometry and ^{1}H, ^{13}C and ^{31}P NMR spectroscopy (See

5.9 Experimental).
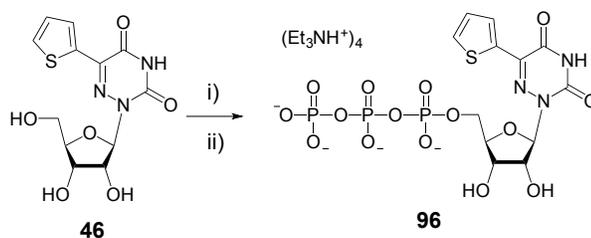
Scheme 5.1 Synthesis of modified nucleoside triphosphate **96**. Reagents and conditions: i) POCl$_3$, (MeO)$_3$PO; ii) tributylammonium pyrophosphate, proton sponge, 0–4°C. See experimental for details.

5.3 T7 RNA polymerase-mediated *in vitro* transcription reactions

To evaluate the enzymatic incorporation of the 6-aza-modified nucleosides into short RNA oligonucleotides, transcription reactions with the analytically pure triphosphates **96**–**98** and T7 RNA polymerase were performed (Figure 5.2a). A short DNA promoter-template duplex was assembled by annealing the 18-mer consensus promoter (**ODN1**) with commercial DNA template **ODN2**. All full-length RNA transcripts **ON1**–**ON4** were obtained under standard *in vitro* transcription conditions and purified by denaturing polyacrylamide gel electrophoresis.[11, 14, 23, 36, 37] UV shadowing (254 nm) was used to visualize the transcription reactions. Notably, under long UV illumination (365 nm) transcripts **OD2** and **OD4**, which contain fluorescent modifications (lanes 3, 4, 7 and 8), are highly emissive (Figure 5.3b).

a)

```
T7 promoter ODN1   5'-d-TAA TAC GAC TCA CTA TAG-3'
      Template ODN2   3'-d-ATT ATG CTG AGT GAT AT C GCG GCA CGT-5'
      Template ODN3   3'-d-ATT ATG CTG AGT GAT AT C GCC GCA GTG T-5'
```

T7 RNA polymerase
Natural NTPs/
Modified UTPs

Transcript **ON1** **X**=U
Transcript **ON2** **X**=46
Transcript **ON3** **X**=95
Transcript **ON4** **X**=29
Transcript **ON5** **X**=U
Transcript **ON6** **X**=46

```
ON1-ON4  5'-r-pppGCG CCG XGC A-3'
ON5-ON6  5'-r-pppGCG GCG XCA CA-3'
```

b)

"Nick"

Acceptor (**ON5-ON6**)   Donor (**ON8**)

```
5'-r-ppp GCG GCG XCA CA CCU UCG GGU GAA GUC GCC GC-3'
             |||  |||  |||  ||  |||  |||  |||  |||
Splint  3'-d-CGC CGC AGT GT GGA AGC CCA CTT-5'
```

T4 DNA ligase
ATP

Ligation products **ON8** (**X**=U) and **ON9** (**X**=46)

```
5'-r-ppp GCG GCG XCA CAC CUU CGG GUG AAG UCG CCG C-3'
```

Dissociated splint

```
3'-d-CGC CGC AGT GT  GGA AGC CCA CTT-5'
```

Figure 5.2 (a) Transcription reaction with T7 promoter (**ODN1**) and template **ODN2** in the presence of natural NTPs and UTP or modified UTPs **96**–**98**; resulting in transcripts **ON1**–**ON4** respectively. Transcription reaction with T7 promoter and template **ODN3** in the presence of natural NTPs and UTP or modified **96**, resulting in transcripts **ON5**, **ON6** respectively. (b) Ligation reaction with T4 DNA ligase, modified acceptor, donor and splint.

Following extraction and isolation, the incorporation efficiency of triphosphates **96**, **97** and **98** compared to native UTP (lane 1) in the full-length product was determined to be 77±2%, 37±3% and 97±2%, respectively (lanes 3, 6 and 8). Additionally, the aborted transcripts were extracted and characterized by MALDI-TOF MS. Nucleotide **46** produced two main truncated products, where only one is highly fluorescent (lane 3). These were assigned as the emissive 7-mer (*5'–pppGCG CCG X–3'*, *X*=**46**) and the shorter, non-emissive 6-mer (*5'–pppGCG CCG–3'*) (lane 3). The major abortive transcript for **95** was also found to correlate to a 6-mer (lane 6). Interestingly, when only native NTPs were used

(lane 1) the truncated construct was characterized as a 5-mer, where the last nucleotide is A instead of C (*5′– pppGCG CA–3′* instead of *5′–pppGCG CC–3′*).
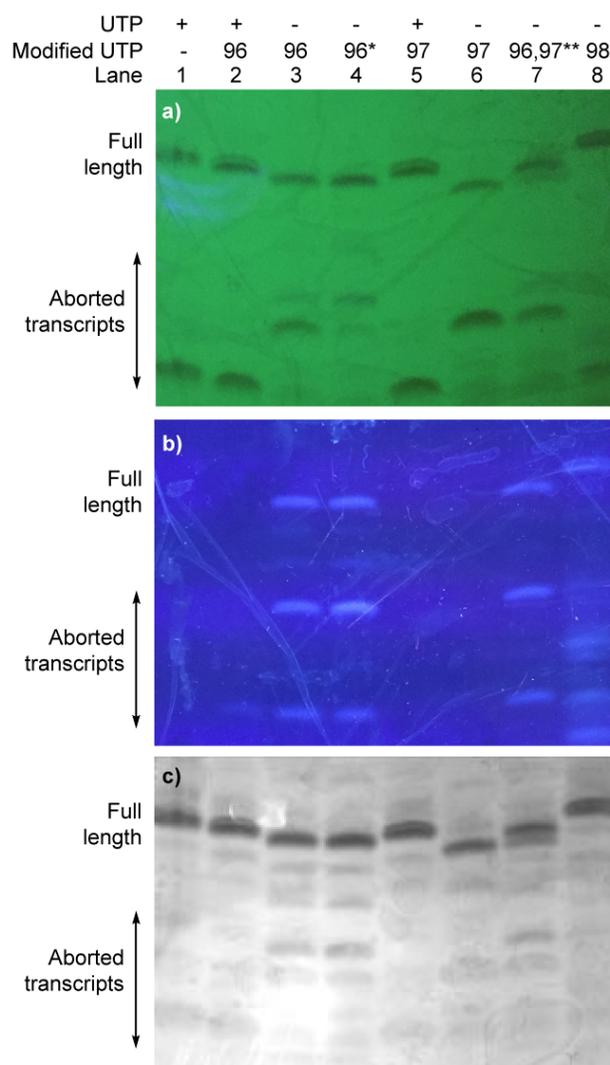
Figure 5.3 Transcription reactions were carried out with T7 promoter (**ODN1**) and template **ODN2** in the presence of natural NTPs and UTP or modified UTPs **96**, **97**, **98**. Lane 1: control reaction with all natural NTPs; lane 2: competition reaction in the presence of equimolar concentration of UTP and **96**, lane 3: reaction in the presence of **96**, lane 4.: reaction in presence of **96** (*5mM instead of 1mM, as for all other reactions), lane 5: reaction in the presence of equimolar concentration of UTP and **97**, lane 6: action in the presence of **97**, lane 7**: reaction in the presence of equimolar concentration of **96** and **97**, lane 8: reaction in the presence of **98**. The reaction was resolved by gel electrophoresis on a denaturing 20% polyacrylamide gel; under UV light at a) 254 nm (on TLC plate); b) 365 nm; c) black and white picture after Stains-All was used to visualize the oligonucleotides.

Additionally, competitive transcription reactions were performed in the presence of equimolar concentrations of UTP and the modified triphosphates **96** and **97**. The 6-azauridine-containing triphosphates (**96** and **97**) cannot compete with UTP and only the native full length 10-mer is produced. In contrast, 5-thiopheneuridine triphosphate **98**, can favorably compete with UTP.[11] T7 polymerase exhibits approximately 1.5x fold preference for the modified triphosphate (**96**), in which the thiophene ring is conjugated at position 5, over 6azaUTP (**97**) (lane 7). ‡

5.4 T4 DNA ligase-mediated *in vitro* ligation reaction

A singly modified A-site hairpin was designed for monitoring small molecule binding based on previously designed short RNA constructs (Figure 5.2b and 5.5).[23, 38] A singly modified 11-nt RNA construct with **46** at position 7 was transcribed to give transcript **ON6** using previously described conditions (Figures 5.2a and 5.8). The modified transcript **ON6**, serving as an acceptor, was ligated to a 5'-phosphorylated 20-nt donor oligonucleotide (**ON7**) containing only native nucleotides (Figure 5.2b) under standard *in vitro* ligation condition, including T4 DNA ligase, ATP and a bridging DNA splint, and resolved by 20% PAGE (Figure 5.9).[39] The ligation product **ON9** was extracted and analyzed for its composition by enzymatic digestion in the presence of S1 nuclease and alkaline phosphatase. Integrated reversed-phase HPLC gave the following nucleoside ratio: C/U/A/G/**46** = 10.7/4.0/4.0/12.6/1.1 (Figure 5.4). Fractions corresponding to each peak were collected and subjected to mass spectrometric analysis, which confirmed the authenticity of the native and modified nucleoside **46**. Thermal melting studies

showed that the incorporation of **46** had a small effect on the stability of the folded RNA. The unmodified control RNA ligation product **ON8** displayed a T_m of 82.8 (±0.1) °C, while the T_m of modified RNA ligation product **ON9** was 79.3 (±0.3) °C (Figure 5.17).
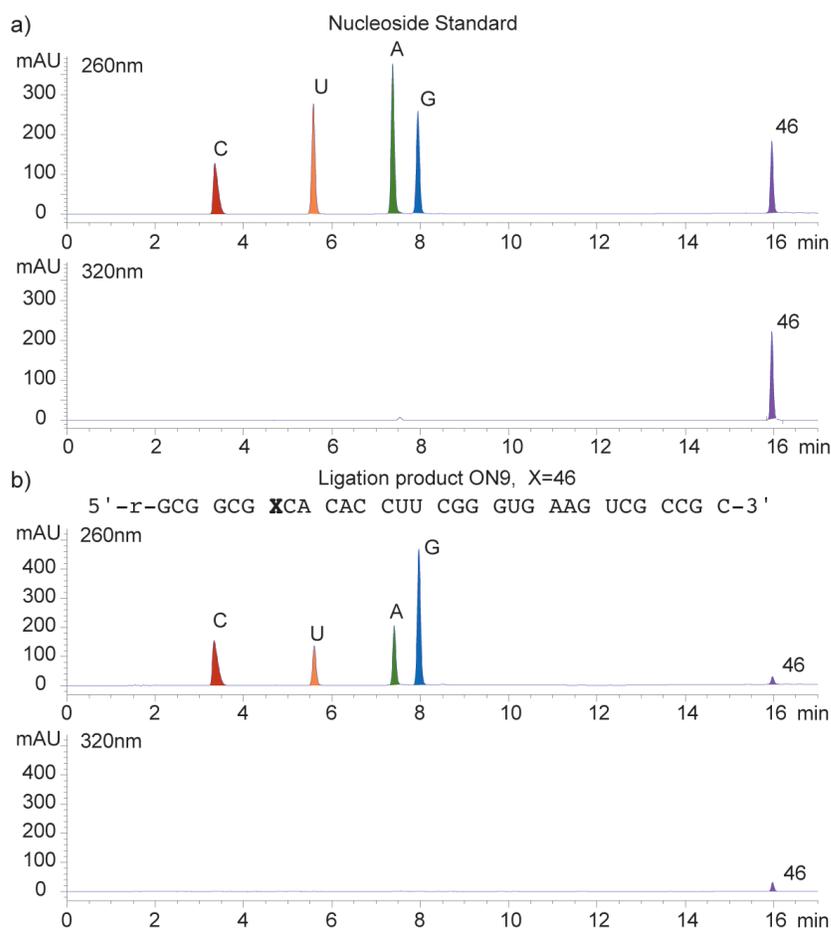
Figure 5.4 HPLC profile of enzymatic digestion reactions. (a) A mixture of nucleosides used as a standard and (b) digestion results of ligation product **ON9**. Digestion of 1−2 nmol of transcript was carried out using S1 nuclease for 2 h at 37 °C and followed by dephosphorylation with alkaline phosphatase for 2 h at 37 °C. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC with Agilent column eclipse XDB-C18 (5μm, 4.6 x 150mm). Mobile phase: 0−5% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 10 min and then increase 5−35% acetonitrile (0.1% formic acid) in water (0.1% formic acid) 10−17 min; flow rate 1 mL/min.

5.5 Binding of aminoglycosides to a fluorescently modified A-site

To assess the potential of the 6-aza modified uridine **46** as a reporter nucleoside for small molecule binding assays, we chose the A-site model hairpin **ON9** as a test system. Note that the modified residue **46** is in close proximity to the binding site but not part of it (Figure 5.5). A-site hairpin **ON9** was therefore excited at 320 nm and changes in emission at 420 nm were monitored upon addition of aminoglycosides. Titration curves were generated by plotting fractional fluorescence saturation versus aminoglycoside concentration. Titration of neomycin (**99**) resulted first in slight decrease and then more significant increase of the fluorescence intensity (Figure 5.6a). The first binding event, corresponding to EC_{50} value of $0.9 \pm 0.2 \times 10^{-3}$ mM, in close agreement with literature values, reflects the A-site binding of neomycin.[38] The second binding event, manifested by an increase of emission intensity, corresponding to EC_{50} value of $2 \pm 1 \times 10^{-2}$ mM, likely reflects remote neomycin–RNA association. Similarly, when tobramycin (**100**) was titrated into the labelled construct **ON9**, the fluorescent intensity first slightly increased (EC_{50} $4 \pm 1 \times 10^{-3}$ mM) and was then followed by more significant emission increase (EC_{50} 0.24 ± 0.03 mM) (Figure 5.18).
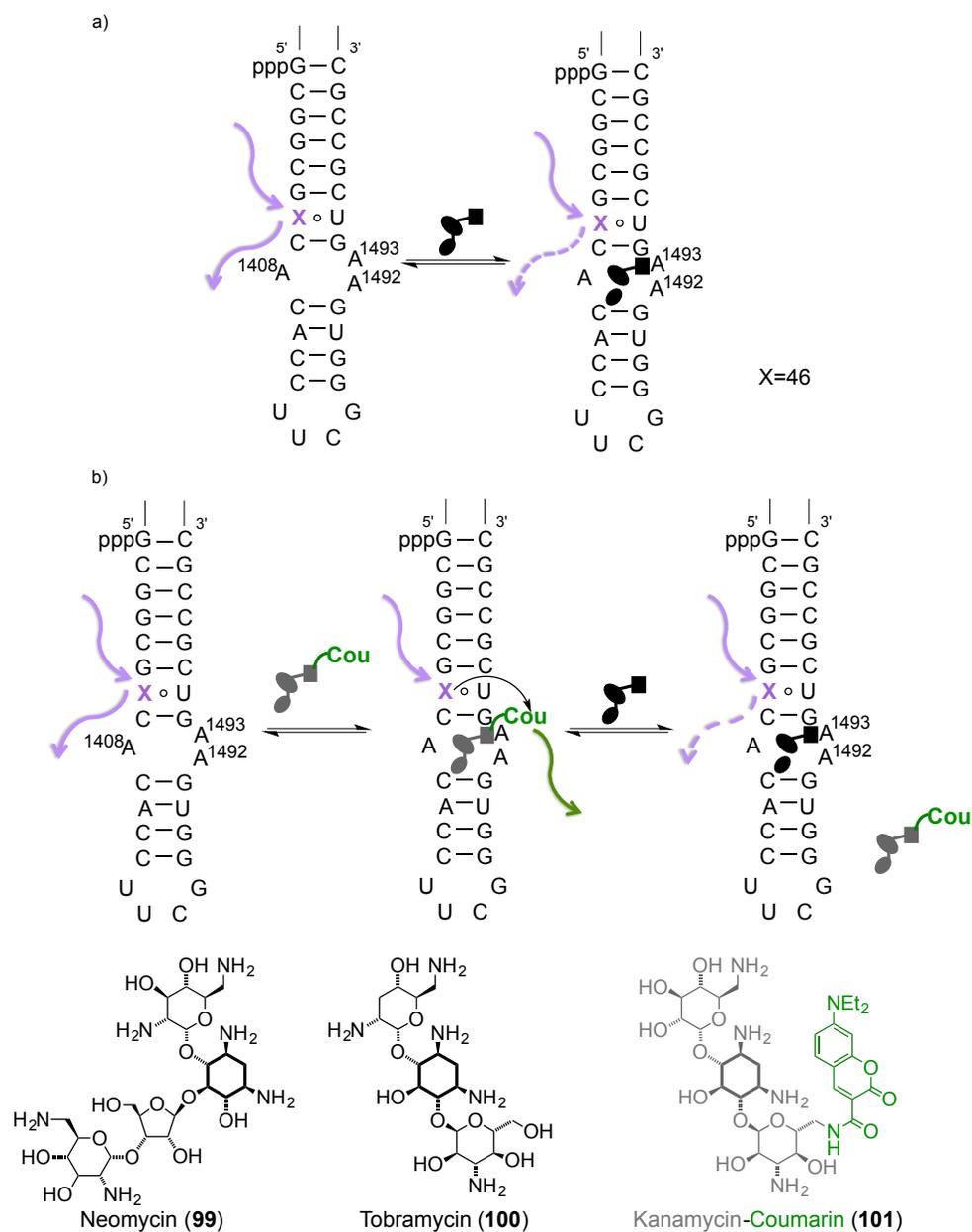
Figure 5.5 (a) Binding of aminoglycosides to a fluorescently modified A-site hairpin **ON9**. (b) Binding and displacement events of A-site bound coumarin-aminoglycoside with unlabelled RNA binders (where 6-aza modified uridine **46** is a FRET donor and coumarin is a FRET acceptor).
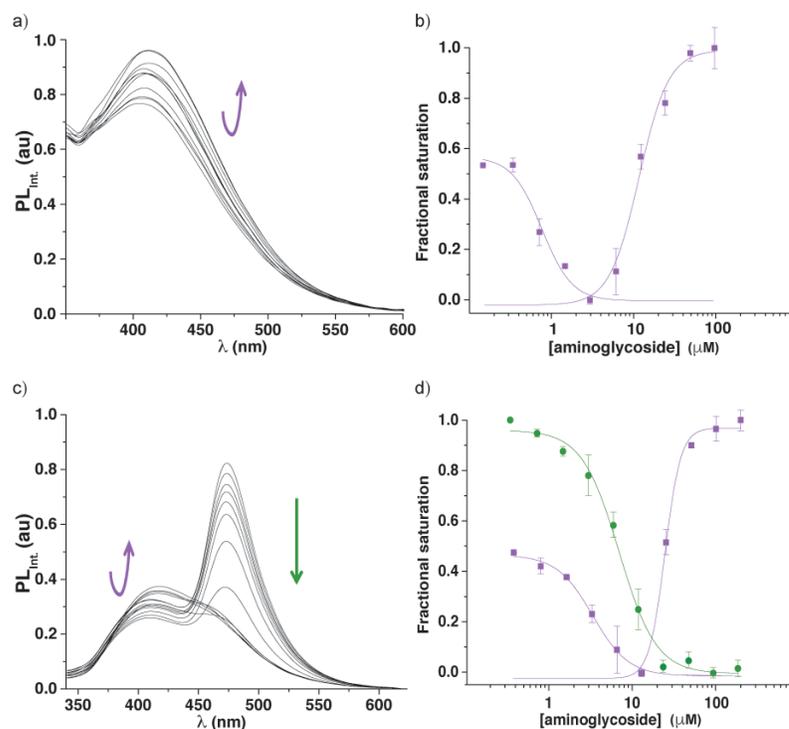
Figure 5.6 Fluorescence titrations of the A-site construct **ON9** with neomycin (**99**). (a) Representative emission spectra for titrations of A-site hairpin **ON9** with neomycin. (b) Curve fit for the titrations of A-site hairpin **ON9** with neomycin (**99**) with respective values EC_{50} 0.9 ±0.2 × 10^{-3} mM and 2 ±1 × 10^{-2} mM (fractional saturation vs. antibiotic concentration). (c) Representative emission spectra for displacement studies: titrations of A-site hairpin **ON9** with neomycin (**99**). (d) Curve fit for the displacement studies: titrations of A-site hairpin **ON9** with neomycin (**99**) with respective values EC_{50} 3.0 ±0.1 × 10^{-3} mM and 3 ±1 × 10^{-2} mM [fractional saturation vs antibiotic concentration; **46**, FRET donor (purple), 6"-amino-6"-deoxykanamycin-coumarin conjugate, FRET acceptor (green)].
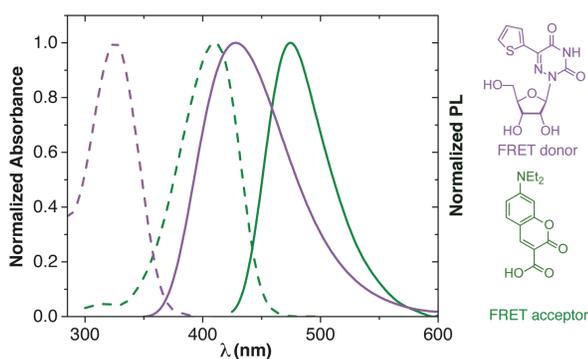
Figure 5.7 Normalized absorption (---) and emission (—) spectra of 5-thiophene-6-azauridine in purple (**46**; FRET donor) and coumarin in green (FRET acceptor).

5.6 Displacement of A-site bound coumarin-aminoglycoside with unlabelled RNA binders

In our displacement assay format, detection of binding events is dependent on the interactions between two fluorophores, where a labelled RNA serves as a FRET donor and an appropriately labelled aminoglycoside serves as a FRET acceptor.[38] The emissive 5-thiophene-6-azauridine was found to be an excellent FRET donor for 7-diethylaminocoumarin-3-carboxylic acid (Figure 5.7).

Assembly of the FRET complex was accomplished by adding the coumarin conjugate of 6″-amino-6″-deoxykanamycin **101** (0.53 μM) to the modified A-site hairpin **ON9** (1 μM) prior to the addition of an unlabelled aminoglycoside. Titration of unlabelled neomycin (**99**) into a pre-formed FRET complex excited at 320 nm resulted in a slight decrease followed by an increase of emission intensity of the FRET donor at 420 nm accompanied by a concurrent decrease of the acceptor emission at 470 nm. The corresponding EC_{50} values from the displacement assay for each event (3.0 $\pm$0.1 $\times$ 10^{-3} mM and 3 $\pm$1 $\times$ 10^{-2} mM, respectively are similar to the data obtained by the direct binding assay. Note that the first binding event is reflected by changes in the donor's and sensitized acceptor's emission, while the second event is manifested by changes in the donor's emission intensity only, as expected (Figure 5.6).

5.7 Discussion

Tuning the photophysical properties of emissive nucleoside analogues can be a challenging task. We envision that introducing an electron-withdrawing

modification at position 6 can enhance the polarization of 5-modified conjugated biaryl uridines, ideally yielding red-shifted emission maxima and increased fluorescence intensity (Figure 5.1a). While indeed realized with the 6-aza-uracil motif, the major challenge remaining was the sluggish recognition of such modified pyrimidines by common RNA polymerases.[28-30] To explore the utility of such a modification within oligonucleotides we have sought to explore the enzymatic incorporation of the corresponding triphosphates, hypothesizing that the conjugated aromatic ring at the 5-position may enhance enzymatic incorporation when compared to 6-azaUTP, a rather poorly incorporating parent nucleotide.

T7 RNA polymerase-mediated *in vitro* transcription reactions

Although in vitro transcription reactions can be very effective, the "pickiness" of RNA polymerase can be challenging. To investigate the "6-aza" effect on in vitro transcription we compared the incorporation efficiency of 5-thiophene-6-azauridine (**46**), 6-azauridine (**95**) and 5-thiophene-uridine (**29**).

T7 RNA polymerase accepts the triphosphates **96**–**98** as substrates and effectively incorporates all three modified nucleosides **46**, **95** and **29**. Triphosphate **98**, which is lacking the nitrogen at position 6 in the uridine core is nearly as effective as the native UTP. 6-substituted uridines are frequently forced into a less favourable *syn* orientation and perform less effectively in transcription reactions.[31-33] Indeed, analyses of the abortive transcripts suggest the enzyme aborts transcription when in a close proximity to a 6-aza-U (**95**) modification, in addition to a lower overall transcription yield. Conjugation of a thiophene ring to the 6-azauridine core, restores "enzyme friendly" behaviour, facilitating the enzymatic

generation of full length emissive transcripts, with 2-fold increase of incorporation efficiency compared to 6-azauridine (**95**). Additionally, competitive transcription reactions show, as hypothesized, that conjugating a thiophene ring to 6-aza-UTP at position 5 partially overcomes the detrimental effect of the CH/N modification at the 6 position.

T4 DNA ligase-mediated *in vitro* ligation reaction

A singly modified A-site hairpin was designed for a potential aminoglycoside binding assay (Figure 5.5). Considering that the construct contains multiple uridine residues and only one modification is desired, its preparation by an enzymatic transcription reaction is not plausible. A potential approach involves transcribing a smaller 11-nt construct with a single incorporation of **46** (transcript **ON6**, Figure 5.2a) and ligating it to a 20-nt oligonucleotide containing native uridines only (donor, Figure 5.2b).

Using T4 DNA ligase, an ATP-dependent enzyme, which catalyzes the joining of two duplexed oligonucleotides, an efficient ligation reaction between acceptor **ON6** and native donor yielded the modified A-site construct **ON9**. Enzymatic digestion followed by reversed-phase HPLC profiling clearly shows that the nucleoside composition matches with the expected 31-nt ligation product **ON9**'s sequence (Figure 5.4).

Binding of aminoglycosides to a fluorescently modified A-site

The bacterial decoding site within the 16S ribosomal RNA, also referred to as the A-site, possesses a small domain comprised of unpaired adenosines, targeted by aminoglycoside antibiotics.[40-42] For biophysical studies, a short RNA

construct representing a functional decoding site is commonly used.[43, 44] The dynamics and remodelling of this domain upon aminoglycoside binding can be exploited for fluorescence-based assays as structural changes are likely to translate into changes in the microenvironment of the fluorescent probe.[45-54] Both 2-aminopurine (2-AP) and pyrroloC (pC) have shown promise as such,[55-59] however, both suffer from significantly quenched emission upon incorporation into oligonucleotides.[1, 23] Additionally, the responsiveness of 2-aminopurine is antibiotic-dependent and appears to fail for certain binders, such as neomycin.[17]

We hypothesized that upon titration of aminoglycoside antibiotics (e.g., neomycin, tobramycin) into the labelled RNA construct **ON9**, the induced conformational and environmental changes would cause changes in emission intensity of the responsive fluorescent nucleoside **46** (Figure 5.6a), which is sensitive to changes in polarity, viscosity, and pH.[12, 60-62] Indeed, upon titration of neomycin (**99**) and tobramycin (**100**) into the fluorescent A-site hairpin **ON9** the emission intensity has changed with increasing concentrations of the antibiotic. For neomycin, the fluorescent intensity slightly decreased initially, which was then followed by a significant increase likely reflecting two discrete binding events. Similar behaviour was observed when tobramycin was titrated. The EC_{50} values extracted for the second event closely corresponded to previously reported literature values for A-site–aminoglycoside binding.[38] We speculate that the second event potentially reflects remote binding (such as to the hairpin stem or loop present in the RNA construct **ON9**), which propagates to trigger an

environmental change and photophysical response of the responsive nucleoside.[53, 63, 64]

In the second approach to detecting aminoglycoside–RNA binding, we envisioned an assay where detection is dependent on the interaction between two fluorophores acting as a FRET pair, with 5-thiophene-6-azauridine as a donor and 7-diethylaminocoumarin-3-carboxyamide as the acceptor. Following the binding of the singly labelled RNA **ON9** to a coumarin conjugate of 6″-amino-6″-deoxykanamycin **101** (Figure 5.5b), which exhibit a high FRET signal, a competitor (e.g., unlabelled neomycin **99**) is titrated in, displacing the fluorescent placeholder (**101**). The intensity of the FRET-acceptor's sensitized emission drops upon titrating neomycin, reflecting its displacement from the A-site. Further titration of neomycin, shows, as before, enhanced emission of the donor aza-U nucleotide. This event, likely reflecting neomycin binding to remote RNA sites (e.g., the tetraloop) nevertheless propagates through the folded structure and alters the probe's immediate environment, yielding enhanced emission.

The A-site and its interactions with aminoglycoside antibiotics have been thoroughly studied.[48–52] Models for specific as well as non-specific binding events have been discussed, including crystal structures showing specifically- and nonspecifically-bound neomycin.[53] In this context, such fluorescence based assays are frequently used for inhibitor discovery and not necessarily for thorough biophysical analyses of RNA–Ligand interactions,[23,38] as the photophysical changes and their physical triggers are rarely fully interpretable, even for established probes.[1]

5.8 Conclusions

The design of emissive nucleosides and their implementation for biochemical and biophysical applications remain an empirical exercise. The relationship between molecular structure and photophysical properties (particularly when it comes to excited state dynamics) is hard to predict and can only be experimentally validated. This issue is particularly challenging for emissive nucleosides since their photophysical features are strongly impacted upon incorporation into oligonucleotides. Here we illustrate that 6-modifed uridine analogues represent an intriguing family with potentially valuable applications. Particularly rewarding is the observation that upon modification of the 5-position, such nucleosides can be incorporated into RNA constructs by T7 RNA polymerase-mediated *in vitro* transcription reactions. 5-thiphene-6-azauridine triphosphate (**96**) is a "polymerase-friendly" substrate, and is incorporated twice as efficiently as 6-azauridine triphosphate, the parent analogue (77±2% vs 37±3%, respectively), which can then be utilized in RNA–ligand binding assays. We illustrate that this analogue can be used as the sole probe or in conjunction with a FRET acceptor for monitoring RNA binding by aminoglycoside antibiotics, either directly or via displacement assays, respectively. Our design is not necessarily limited to the A-site and its binders, but can possibly be used for the analysis of other RNA targets with biochemical interest or therapeutic potential.

5.9 Experimental

General

Tris(tetrabutylammonium)hydrogen pyrophosphate, phosphoryl chloride, trimethyl phosphate, 1,8-bis(dimethylamino)naphthalene (proton sponge), and anhydrous DMF were purchased from Sigma-Aldrich. All other solvents were purchased from Fisher Scientific. All reagents and solvents were used without further purification.

Synthesis of triphosphate 96

Previously published procedure was used with minor changes, as described below.[11, 14]

Tris(tetrabutylammonium) hydrogen pyrophosphate (0.91 g, 1.0 mmol) in a 10 mL round-bottom flask, and **46** (66 mg, 0.20 mmol) in a 25 mL round-bottom flask, were separately coevaporated with anhydrous pyridine and dried (3 × 1mL). Trimethyl phosphate (2 mL) was added to **46** and cooled in an ice bath to 0 °C. Phosphoryl chloride (46 µL, 0.5 mmol) was added slowly, and the reaction was stirred for 2 h at 0 °C. The coevaporated tris(tetrabutylammonium) hydrogen pyrophosphate was dissolved in anhydrous DMF (2 mL) and added to the reaction mixture with **46**. Then proton sponge (216 mg, 0.1 mmol) was added and the reaction was kept stirring at 0 °C for 40 min. To the reaction mixture was added 1 M triethylammonium bicarbonate buffer (TEAB) (6 mL), and the mixture was stirred briefly. The mixture was then transferred to a separatory funnel and washed with ethyl acetate (10 mL). The organic layer was then back-extracted with 1 M TEAB (5 mL). The aqueous layers were combined and concentrated under reduced pressure at room temperature to afford an oily yellow residue. The residue was

dissolved in 0.05 M ammonium bicarbonate buffer (10 mL) and loaded onto a DEAE Sephadex A25 anion-exchange column kept in a cold room at 4 °C. The column was eluted using a gradient mixer with 0.01−1.0 M of ammonium bicarbonate buffer. A fraction collector was used to collect 260 fractions that were about 8 mL (220 drops). The fractions containing the triphosphate were evaporated under reduced pressure at 10 °C, and the residue was then lyophilized. The triphosphate was further purified by HPLC (Phenomenex Synergi Fusion-RP 80A C18 column, 4 μm, 250 × 10 nm, 5−20% acetonitrile in 50 mM TEAA buffer, pH 6.0, 30 min). Appropriate fractions were lyophilized and treated with 3% sodium perchlorate in acetone and then with Chelex 100 for 30 minutes with shaking, then filtrated and dried to afford nucleotide **96** (29 mg, 20%). ^{1}H NMR (400 MHz, D_2O) δ 8.08 – 7.94 (m, 1H), 7.69 – 7.57 (m, 1H), 7.26 – 7.14 (m, 1H), 6.25 – 6.15 (m, 1H), 4.71−4.63 (m, 1H), 4.69 – 4.52 (m, 1H), 4.38 – 4.20 (m, 2H), 4.18 – 4.02 (m, 1H) and triethylammonium $(CH_3CH_2)_3N^+$ counter ion (1equiv.) 3.17 (q, J = 8.0 Hz, 6H), 1.25 (t, J = 6.0 Hz, 9H); ^{13}C NMR (125 MHz, D_2O) δ 159.18, 151.91, 139.35, 133.88, 129.84, 129.66, 127.91, 89.49, 82.64, 72.95, 70.47, 66.17 and triethylammonium $(CH_3CH_2)_3N^+$ counter ion (1equiv.) 46.44, 8.08; ^{31}P NMR (202 MHz, D_2O) δ −8.26 (d, J = 19.9 Hz, Pγ), −11.06 (d, J = 19.2 Hz, Pα), −22.34 (t, J = 19.5 Hz, Pβ); HR ESI-MS (negative ion mode)$[C_{12}H_{16}N_3O_{15}P_3S]^-$ calculated 565.9442, found 565.9438.

^{1}H-NMR, ^{13}C-NMR and ^{31}P-NMR spectra of triphosphate 96
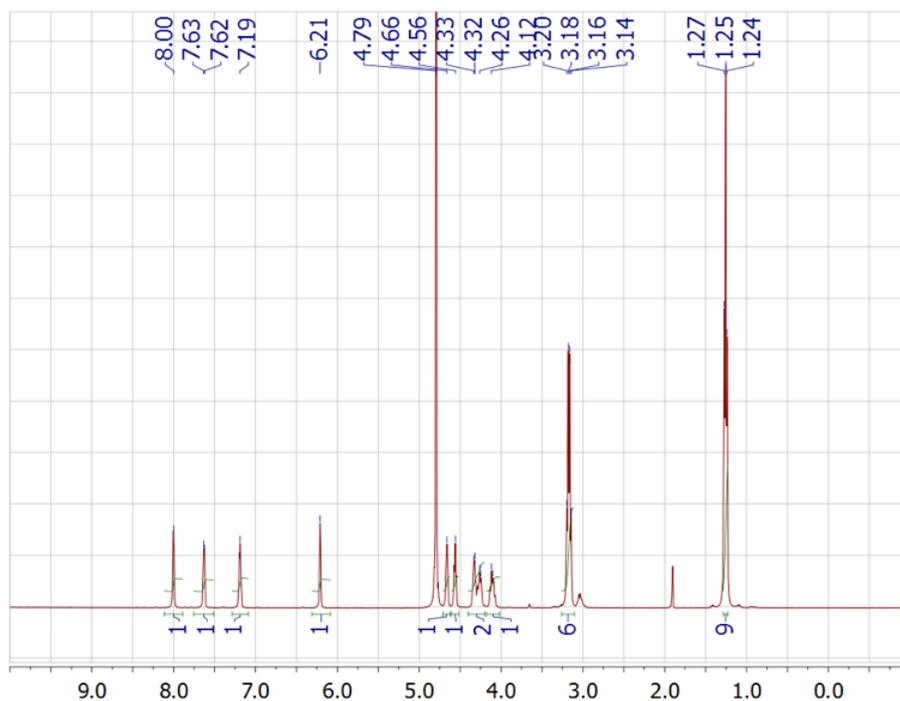
Spectrum 5.1 ^{1}H-NMR of 5-(thiophen2-yl)-6-azauridine triphosphate (**96**).
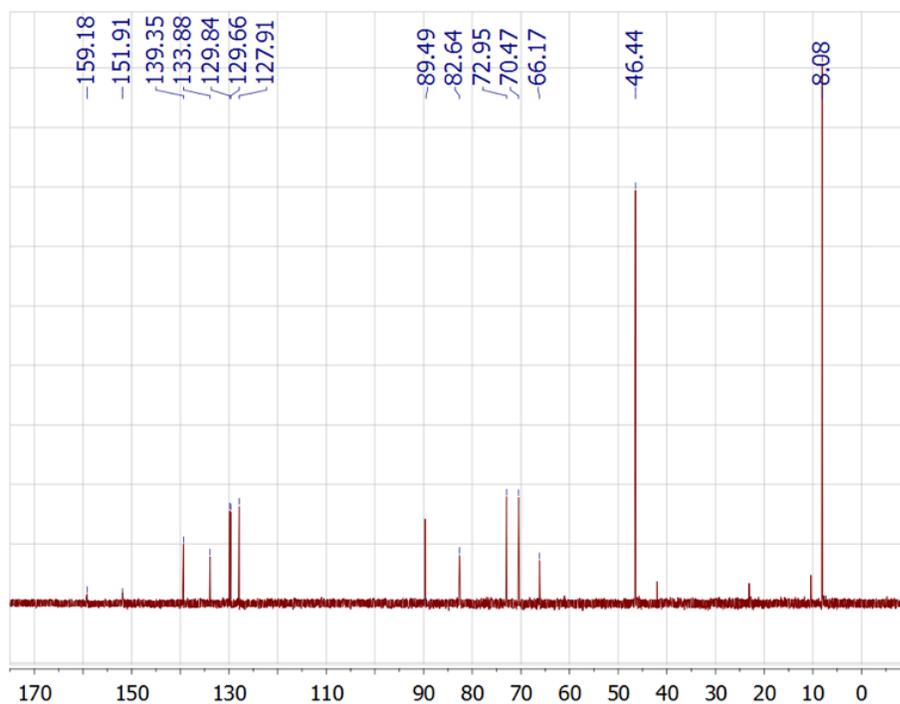
Spectrum 5.2 ^{13}C-NMR of 5-(thiophen2-yl)-6-azauridine triphosphate (**96**).
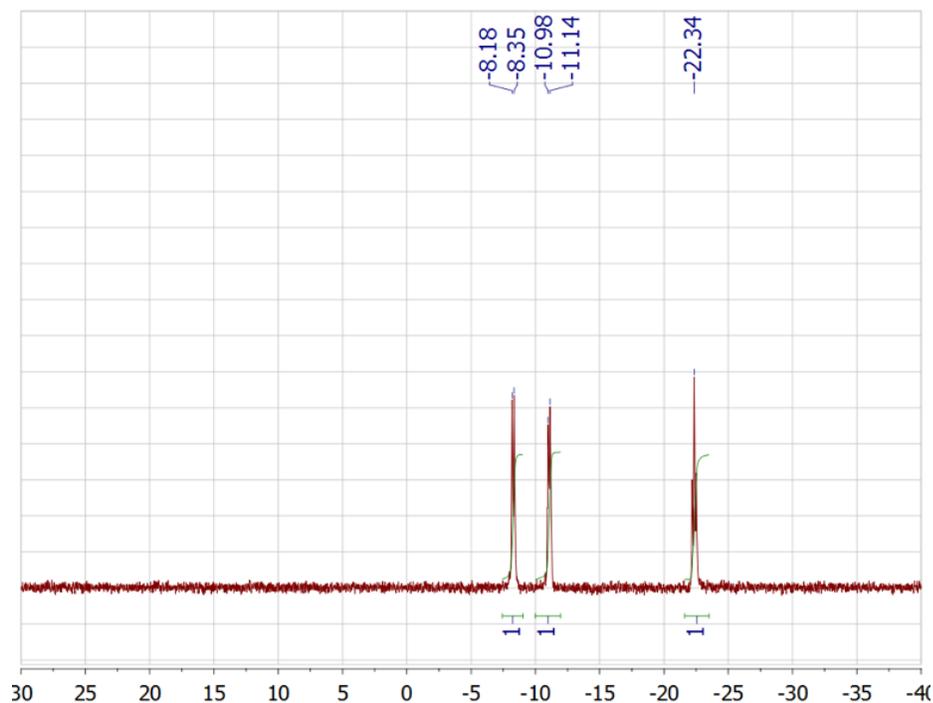
Spectrum 5.3 ^{31}P-NMR of 5-(thiophen2-yl)-6-azauridine triphosphate (**96**).

Oligonucleotides synthesis and purification

General

All unmodified DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. Native NTPs were purchased from Thermo Scientific Dharmacon. Triphosphate **97** (6AzaUTP) was purchased from TriLink BioTechnologies. T7 RNA polymerase (5000 U), T4 Polynucleotide Kinase (PNK) and kinase buffer (2500 U) were purchased from New England BioLabs (NEB). T4 DNA ligase, 50% PEG 4000 and RiboLock (40 U/µl) were purchased from Fermentas Inc. All other solutions and buffers were prepared in our laboratory.

T7 RNA polymerase-mediated in vitro transcription reactions

Transcription reactions with T7 RNA polymerase and the analytically pure triphosphates **96**, **97** and **98** were performed to analyze their enzymatic incorporation into short RNA transcripts **ON1–ON4**. The T7 promoter (**ODN1**) and template **ODN2** were annealed, and transcribed in the presence of natural NTPs or with modified UTPs **96**, **97**, **98** replacing native UTP. An agarose gel electrophoresis revealed full-length 10-mer products (transcripts **ON1–ON4**). Transcription reaction cocktail contains: annealed template/promoter (500 nM), 1X transcription buffer (pH 9.0), $MgCl_2$ (16 mM), dithiothreitol DTT (10 mM), RiboLock (1 U/µL), T7 RNA polymerase (0.15 µg/µL), native and modified triphosphates (1 mM). Next, transcription reaction was incubated for 4 h at 37 °C. The precipitated magnesium pyrophosphate was removed by centrifugation. The reaction was concentrated to half of the volume. Then 125 µL of loading buffer was added. The mixture was heated at 75 °C for 3 min, and loaded onto a preparative 20% denaturing polyacrylamide gel. The gel was UV shadowed; appropriate bands were excised, extracted with 0.5 M ammonium acetate, and desalted on a Sep-Pak column. Importantly, when visualizing the gel under UV illumination (365 nm), the product and initiation phase truncated transcripts are highly fluorescent.

T4 Polynucleotide kinase (PNK)-mediated phosphorylation

Donor RNA construct for the ligation reaction was purchased from IDT Integrated DNA Technologies and purified before phosphorylation. Phosphorylation reaction cocktail contains: oligonucleotide (0.006 mM), ATP (1 mM), dithiothreitol DTT (5 mM), 10X kinase buffer (0.1 v/v), PNK (0.2 U/µL). After reaction mixture was

incubated for 2 h at 37 °C and precipitated from ethanol, an agarose gel electrophoresis revealed quantitative conversion. [65]

T4 DNA ligase-mediated in vitro ligation reaction

Ligation reaction with T4 DNA ligase and analytically pure donor (**ON7**) and acceptor constructs (**ON5–ON6**) were performed with presence of splint DNA oligonucleotide for better selectivity. The reaction consists of two steps. First, acceptor (10 µM, 1 equiv.), donor (10 µM, 1.5 equiv.) and splint (1.5 equiv.) and Tris-HCl buffer (40 mM, pH 7.8) were mixed and heated at 90 °C for 3 minutes and cooled down slowly until 40 – 50 °C. Second, $MgCl_2$ (10 mM), dithiothreitol DTT (10 mM), ATP (0.5 mM) 50% PEG 4000 (0.1 V) and T4 DNA ligase (100 U/nmol) were added to the previous reaction mixture and incubated for 2 h at 37 °C. Then ethanol precipitation was performed and final product was purified by agarose gel electrophoresis.[39]

T7 RNA polymerase-mediated *in vitro* transcription reaction
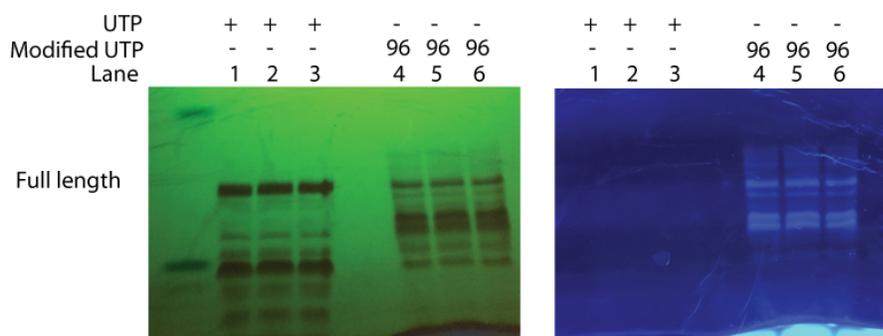
Figure 5.8 Transcription reaction was carried out with T7 promoter **ODN1** and template **ODN3** in the presence of natural NTPs and UTP or modified **96**. Lanes 1–3: control reaction with all natural NTPs; lanes 4–6: reaction in presence of **96**. The reaction was resolved by agarose gel electrophoresis on a denaturing 20% polyacrylamide gel; with UV light at 254 nm (on TLC plate) on the left, and 365 nm on the right.

T4 DNA ligase-mediated *in vitro* ligation reaction
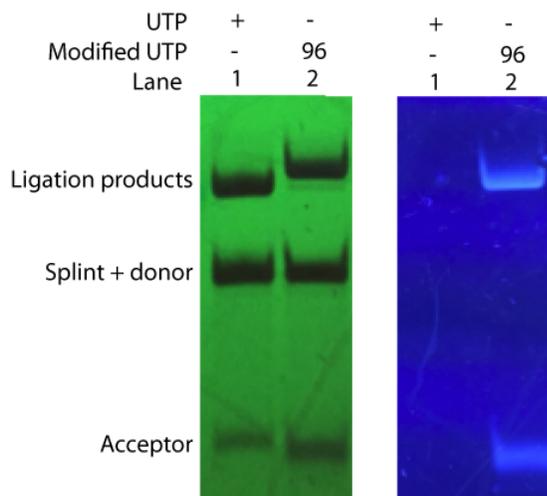
Figure 5.9 Ligation reaction was carried out with T4 DNA ligase in the presence of donor and native acceptor (**ON5**) or modified acceptor (**ON6**). Lane 1: control reaction with native acceptor (transcript **ON5**); lane 2: reaction with modified acceptor (transcript **ON6**). The reaction was resolved by agarose gel electrophoresis on a denaturing 20% polyacrylamide gel; with UV light at 254 nm (on TLC plate) on the left, and 365 nm on the right.

MALDI -TOF MS measurements

Transcripts ON1–ON6
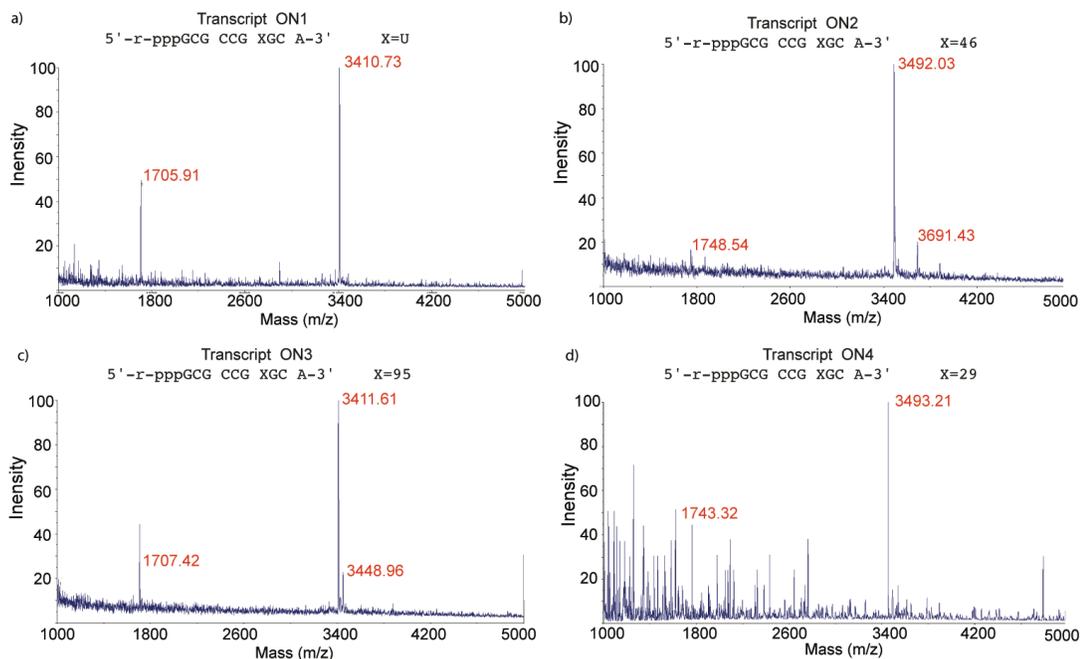
Figure 5.10 MALDI-TOF MS results for (a) transcript **ON1** calculated [M] 3413.34 found 3410.73, (b) transcript **ON2** calculated [M] 3496.58 found 3492.03, (c) transcript **ON3** calculated [M] 3414.46 found 3411.61, (d) transcript **ON4** calculated [M] 3495.59 found 3493.21.
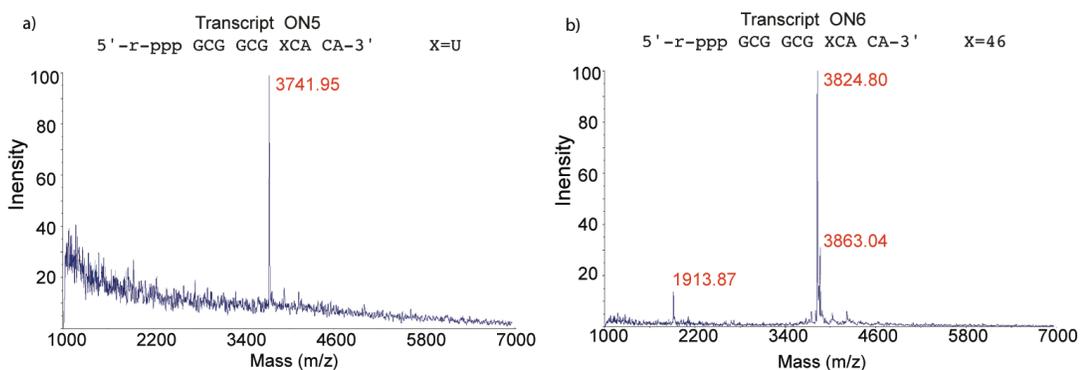
Figure 5.11 MALDI-TOF MS results for 11-mers - ligation acceptor RNA oligonucleotides (a) transcript **ON5** calculated [M+K] 3741.39 found 3741.95, (b) transcript **ON6** calculated [M+K] 3824.63 found 3824.80.

Ligation constructs ON8–ON9
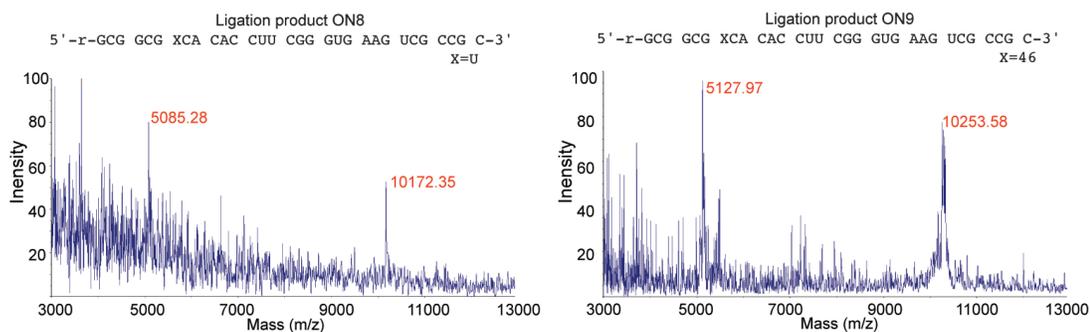
Figure 5.12 MALDI-TOF MS results for (a) ligation products **ON8** calculated [M] 10175.13 found 10172.35, (b) ligation products **ON9** calculated [M] 10258.37 found 10253.58.

Digestion of oligonucleotides

All transcripts **ON1**–**ON6** and ligation products **ON8**–**ON9** (1−2 nmol,) were incubated with S1 nuclease in a reaction buffer (Promega) for 2 h at 37 °C. The reaction was further treated with alkaline phosphatase and dephosphorylation buffer (Promega) for 2 h at 37 °C. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC with an Agilent column eclipse XDB-C18 (5 µm, 4.6 × 150 mm). Mobile phase: 0−5% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 10 min and then increase 5−35% acetonitrile (0.1% formic acid) in water (0.1% formic acid) 10−17 min; flow rate 1 mL/min. Only for transcript **ON3** and corresponding standard mixture the conditions were changed. Mobile phase: 0% acetonitrile (0.1% TEAA buffer) in water (0.1% TEAA buffer) over 5 min and then 0−7% acetonitrile (0.1% TEAA buffer) in water (0.1% TEAA buffer) 5−17 min; flow rate 1 mL/min.

Transcripts ON1–ON6
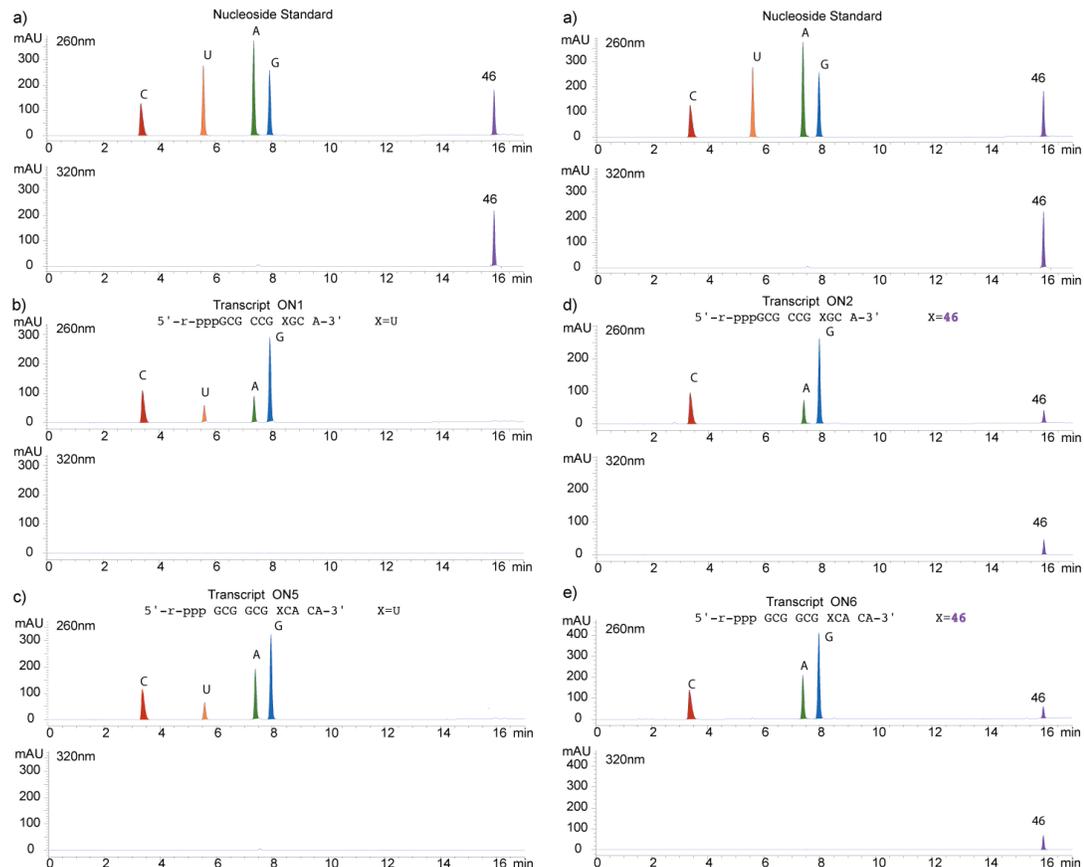
Figure 5.13 HPLC profile of enzymatic digestion (a) mixture of nucleosides used as a standard and digestion results of (b) transcript **ON1**, (c) transcript **ON5**, (d) transcript **ON2** (e) transcript **ON6.** The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC, using a mobile phase of 0−5% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 10 min and then increase 5−35% acetonitrile (0.1% formic acid) in water (0.1% formic acid) 10−17 min; flow rate 1 mL/min.
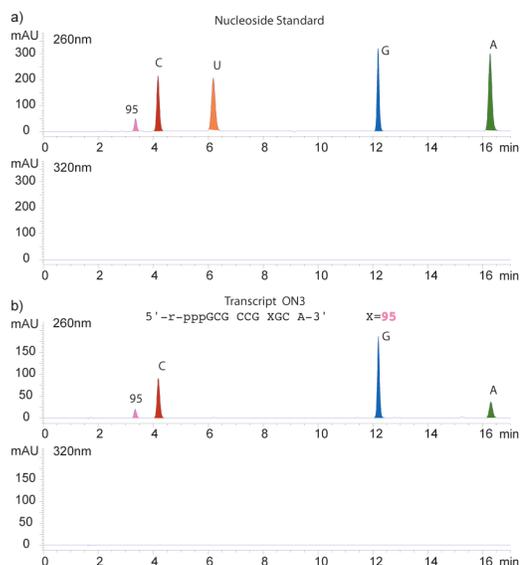
Figure 5.14 HPLC profile of enzymatic digestion (a) mixture of nucleosides used as a standard and digestion results of (b) transcript **ON3**. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC, using a mobile phase: 0% acetonitrile (0.1% TEAA buffer) in water (0.1% TEAA buffer) over 5 min and then 0−7% acetonitrile (0.1% TEAA buffer) in water (0.1% TEAA buffer) 5−17 min; flow rate 1 mL/min.
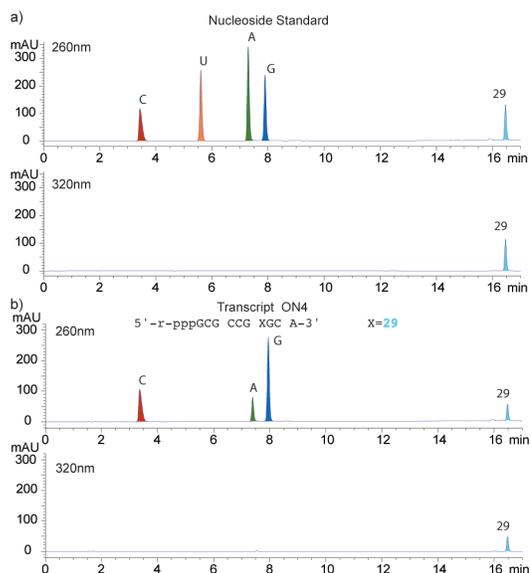
Figure 5.15 HPLC profile of enzymatic digestion (a) mixture of nucleosides used as a standard and digestion results of (b) transcript **ON4**. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC, using a mobile phase of 0−5% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 10 min and then increase 5−35% acetonitrile (0.1% formic acid) in water (0.1% formic acid) 10−17 min; flow rate 1 mL/min.

Ligation construct ON8
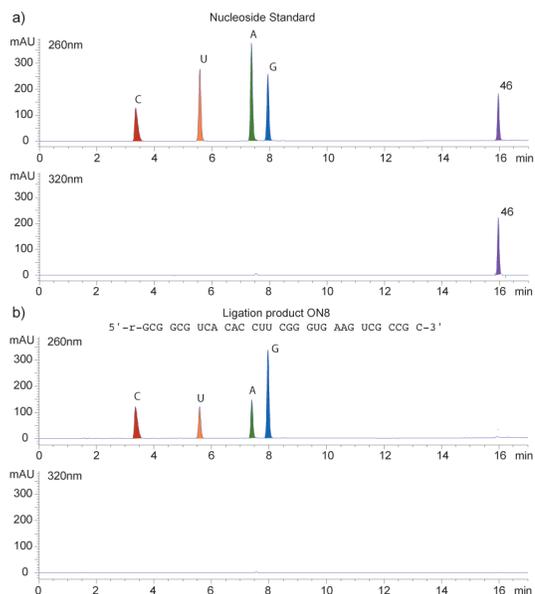
Figure 5.16 HPLC profile of enzymatic digestion (a) mixture of nucleosides used as a standard and (b) digestion results of ligation product **ON8**. Digestion of 1−2 nmol of transcript was carried out using S1 nuclease for 2 h at 37 °C and followed by dephosphorylation with alkaline phosphatase for 2 h at 37 °C. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC, using a mobile phase of 0−5% acetonitrile (0.1% formic acid) in water (0.1% formic acid) over 10 min and then increase 5−35% acetonitrile (0.1% formic acid) in water (0.1% formic acid) 10−17 min; flow rate 1 mL/min.

Thermal denaturation curves for control and modified RNA ligation constructs
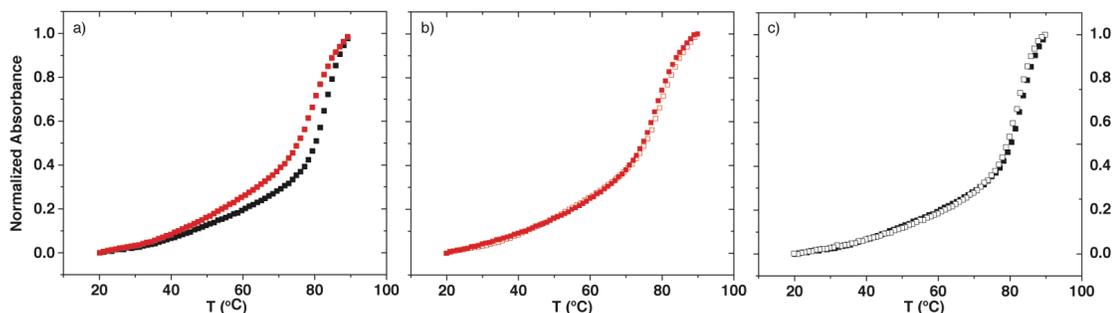
Figure 5.17 Thermal denaturation curves: (a) forward curve (black solid square) for control RNA ligation product **ON8** and forward curve (red solid square) for modified RNA ligation product **ON9**; (b) forward (red solid square) and reverse (red empty square) curve for modified RNA ligation product **ON9** (c) forward (black solid square) and reverse (black empty square) curve for modified RNA ligation product **ON8**. Tm of control RNA ligation product **ON8** is 82.8 ±0.1 °C, Tm of modified RNA ligation product **ON9** is 79.1 ±0.3 °C.

Photophysical studies of modified A-site construct

General

Aqueous samples were prepared with de-ionized water. For all spectroscopic measurements a 1 cm Hellma Analytics quartz cuvette was used. Steady state emission spectra were taken on a PTI luminescence spectrometer with a 1 nm resolution. The sample temperature was kept constant at 20 °C with a Quantum Northwest TLC50 fluorescence cuvette holder in conjunction with a software controllable TC 125 temperature controller.

EC_{50} values were calculated using OriginPro 8.5 software by fitting a dose response curve (eq 1) to the fractional fluorescence saturation (Fs) plotted against the log of antibiotic (A) concentration.

$$F_s = F_0 + (F_\infty [A]^n)/([IC_{50}]^n + [A]^n) \qquad (1)$$

F_s is the fluorescence intensity at each titration point. F_0 and F_∞ are the fluorescence intensity in the absence of aminoglycoside or at saturation, respectively, and n is the Hill coefficient or degree of cooperativity associated with binding. All reported standard deviations were calculated using STDEVP in Microsoft Excel.

Binding of aminoglycosides to a fluorescently modified A-site

All titrations were performed with working solutions of 1 µM labeled A-site **ON9** in 20 mM cacodylate buffer (pH 7.0, 100 mM NaCl, 0.5 mM EDTA). The solutions were heated to 75 °C for 5 min and cooled to room temperature over 2 h. Aliquots (1 µL) of increasing concentrations of an aminoglycoside (**99−100**) were added. Fluorescence spectrum was recorded after each addition until saturation was achieved. Modified A-site **ON9** was excited at 320 nm (slit width 8 nm), and changes in emission were monitored at 420 nm. For figure 5.6, each event was fitted separately for the binding of neomycin to the A-site; critical points were fixed at y=0, x=0.1 mM (for the first event) and y=0, x=1 $\times$ 10^{-3} mM (for the second event). Errors were calculated as standard deviation from two measurements. [23]

Displacement of A-site bound 6″-amino-6″-deoxykanamycin-coumarin with unlabelled RNA binders

All titrations were performed with working solutions of 1 µM labeled A-site **ON9** in 20 mM cacodylate buffer (pH 7.0, 100 mM NaCl, 0.5 mM EDTA). The solutions were heated to 75 °C for 5 min and cooled to room temperature over 2 h 6″-amino-6″-deoxykanamycin-courmarin was added, to give a working

concentration of 0.53 µM, just prior to aminoglycoside titrations (just as described above). For biding studies, **ON9** was excited 320 nm (slit width 8 nm), and changes is emission upon titration with aminoglycosides were monitored at 420 nm and at 470 nm. For figure 5.6, each event was fitted separately for the binding of neomycin to the A-site; critical points were fixed at y=0, x=0.1 mM (for the first event) and y=0, x=1 × 10^{-3} mM (for the second event). Errors were calculated as standard deviation from two measurements.[38]

Binding of Aminoglycosides to a Fluorescently Modified A-site

Tobramycin titration
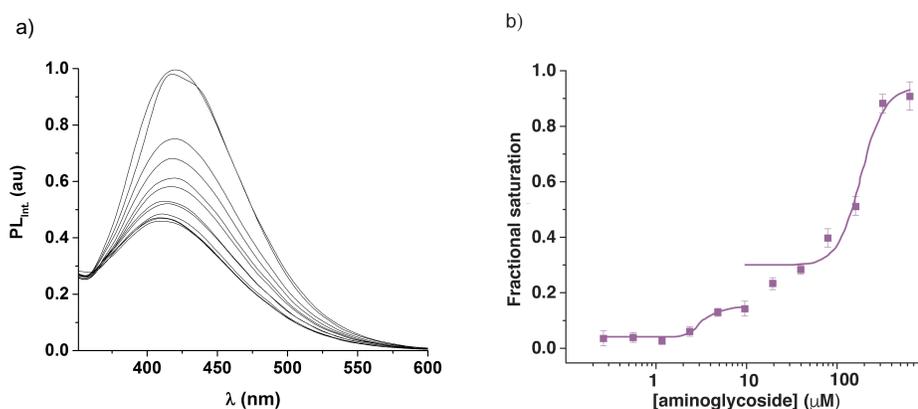
Figure 5.18 Fluorescence titrations of the A-site construct **ON9** with tobramycin (**100**). (a) Emission spectra for titrations of A-site hairpin **ON9** with tobramycin (**100**). (b) Curve fit for the titrations of A-site hairpin **ON9** with tobramycin (**100**) (fractional saturation vs. antibiotic concentration). Two slopes were separated and fitted individually.

Acknowledgements

5.10 References

‡   For full characterization of transcripts **ON1**–**ON4** by MALDI-TOF MS and digestions profiles see experimental Figure 5.10.

1. R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

2. M. Sameiro and T. Goncalves, *Chem Rev*, 2009, **109**, 190-212.

3. L. M. Wilhelmsson, *Quarterly reviews of biophysics*, 2010, **43**, 159-183.

4. Y. Tor, *Pure Appl Chem*, 2009, **81**, 263-272.

5. D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378-391.

6. J. N. Wilson and E. T. Kool, *Org Biomol Chem*, 2006, **4**, 4265-4274.

7. A. Okamoto, Y. Saito and I. Saito, *Journal of Photochemistry and Photobiology C*, 2005, **6**, 108-122.

8. M. J. Rist and J. P. Marino, *Current Organic Chemistry*, 2002, **6**, 775-793.

9. N. J. Greco and Y. Tor, *J Am Chem Soc*, 2005, **127**, 10784–10785.

10. D. Shin, R. W. Sinkeldam and Y. Tor, *J Am Chem Soc*, 2011, **133**, 14912-14915.

11. S. G. Srivatsan and Y. Tor, *Chem-Asian J*, 2009, **4**, 419-427.

12. R. W. Sinkeldam, P. A. Hopkins and Y. Tor, *Chemphyschem*, 2012, **13**, 3350-3356.

13. A. R. Rovira, A. Fin and Y. Tor, *J Am Chem Soc*, 2015, **137**, 14602-14605.

14. L. S. McCoy, D. Shin and Y. Tor, *J Am Chem Soc*, 2014, **136**, 15176-15184.

15. H. Liu, J. Gao, L. Maynard, Y. D. Saito and E. T. Kool, *J Am Chem Soc*, 2004, **126**, 1102-1109.

16. J. Gao, H. Liu and E. T. Kool, *J Am Chem Soc*, 2004, **126**, 11826-11831.

17. F. Seela and G. Becher, *Helv Chim Acta*, 2000, **83**, 928-942.

18. D. C. Ward, E. Reich and L. Stryer, *J Biol Chem*, 1969, **244**, 1228-1237.

19. A. Fin, A. R. Rovira, P. A. Hopkins and Y. Tor, in *Modified Nucleic Acids*, eds. K. Nakatani and Y. Tor, Springer International Publishing, Cham, 2016, DOI: 10.1007/978-3-319-27111-8_1, pp. 1-26.

20. R. S. K. Lane, R. Jones, R. W. Sinkeldam, Y. Tor and S. W. Magennis, *Chemphyschem*, 2014, **15**, 867-871.

21. P. Wigerinck, C. Pannecouque, R. Snoeck, P. Claes, E. De Clercq and P. Herdewijn, *J Med Chem*, 1991, **34**, 2383-2389.

22. N. J. Greco and Y. Tor, *Tetrahedron*, 2007, **63**, 3515-3527.

23. S. G. Srivatsan and Y. Tor, *J Am Chem Soc*, 2007, **129**, 2044-2053.

24. S. G. Srivatsan and Y. Tor, *Tetrahedron*, 2007, **63**, 3601-3607.

25. R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chembiochem*, 2008, **9**, 706-709.

26. N. J. Greco, R. W. Sinkeldam and Y. Tor, *Org Lett*, 2009, **11**, 1115-1118.

27. A. J. Gutierrez, T. J. Terhorst, M. D. Matteucci and B. C. Froehler, *J Am Chem Soc*, 1994, **116**, 5540-5544.

28. F. M. Kahan and J. Hurwitz, *J Biol Chem*, 1962, **237**, 3778-3785.

29. L. Shapiro and J. T. August, *J Mol Biol*, 1965, **14**, 214-220.

30. S. Asano, Kurashin.Y, Y. Anraku and D. I. Mizuno, *J Biochem-Tokyo*, 1971, **70**, 9-20.

31. V. Cody and T. I. Kalman, *Nucleos Nucleot*, 1985, **4**, 587-594.

32. K. Felczak, A. K. Drabikowska, J. A. Vilpo, T. Kulikowski and D. Shugar, *J Med Chem*, 1996, **39**, 1720-1728.

33. E. Sochacka, G. Czerwinska, R. Guenther, R. Cain, P. F. Agris and A. Malkiewicz, *Nucleos Nucleot Nucl*, 2000, **19**, 515-531.

34. S. G. Srivatsan, H. Weizman and Y. Tor, *Org Biomol Chem*, 2008, **6**, 1334-1338.

35. J. Ludwig, *Acta Biochim Biophys*, 1981, **16**, 131-133.

36. J. F. Milligan and O. C. Uhlenbeck, *Method Enzymol*, 1989, **180**, 51-62.

37. J. F. Milligan, D. R. Groebe, G. W. Witherell and O. C. Uhlenbeck, *Nucleic Acids Res*, 1987, **15**, 8783-8798.

38. Y. Xie, A. V. Dix and Y. Tor, *J Am Chem Soc*, 2009, **131**, 17605-17614.

39. K. Lang and R. Micura, *Nat Protoc*, 2008, **3**, 1457-1466.

40. D. Moazed and H. F. Noller, *Nature*, 1987, **327**, 389-394.

41. Y. Tor, *Angew Chem Int Edit*, 1999, **38**, 1579-1582.

42. L. S. McCoy, Y. Xie and Y. Tor, *Wires Rna*, 2011, **2**, 209-232.

43. P. Purohit and S. Stern, *Nature*, 1994, **370**, 659-662.

44. M. I. Recht, D. Fourmy, S. C. Blanchard, K. D. Dahlquist and J. D. Puglisi, *J Mol Biol*, 1996, **262**, 421-436.

45. M. L. Zapp, S. Stern and M. R. Green, *Cell*, 1993, **74**, 969-978.

46. H. Y. Mei, D. P. Mack, A. A. Galan, N. S. Halim, A. Heldsinger, J. A. Loo, D. W. Moreland, K. A. SannesLowery, L. Sharmeen, H. N. Truong and A. W. Czarnik, *Bioorgan Med Chem*, 1997, **5**, 1173-1184.

47. T. Hermann and E. Westhof, *Biopolymers*, 1998, **48**, 155-165.

48. Y. Tor, T. Hermann and E. Westhof, *Chem Biol*, 1998, **5**, R277-R283.

49. T. Hermann and E. Westhof, *J Mol Biol*, 1998, **276**, 903-912.

50. Q. Vicens and E. Westhof, *Biopolymers*, 2003, **70**, 42-57.

51. Q. Vicens and E. Westhof, *Chembiochem*, 2003, **4**, 1018-1023.

52. Y. Tor, *Chembiochem*, 2003, **4**, 998-1007.

53. F. Zhao, Q. Zhao, K. F. Blount, Q. Han, Y. Tor and T. Hermann, *Angew Chem Int Edit*, 2005, **44**, 5329-5334.

54. J. R. Thomas and P. J. Hergenrother, *Chem Rev*, 2008, **108**, 1171-1224.

55. M. Kaul, C. M. Barbieri and D. S. Pilch, *J Am Chem Soc*, 2004, **126**, 3447-3453.

56. S. Shandrick, Q. Zhao, Q. Han, B. K. Ayida, M. Takahashi, G. C. Winters, K. B. Simonsen, D. Vourloumis and T. Hermann, *Angew Chem Int Edit*, 2004, **43**, 3177-3182.

57. M. Kaul, C. M. Barbieri and D. S. Pilch, *J Am Chem Soc*, 2006, **128**, 1261-1271.

58. C. M. Barbieri, M. Kaul and D. S. Pilch, *Tetrahedron*, 2007, **63**, 3567-3574.

59. A. A. Marti, S. Jockusch, Z. M. Li, J. Y. Ju and N. J. Turro, *Nucleic Acids Res*, 2006, **34**, e50.

60. S. K. Pal and A. H. Zewail, *Chem Rev*, 2004, **104**, 2099-2123.

61. B. Nguyen, S. Neidle and W. D. Wilson, *Accounts Chem Res*, 2009, **42**, 11-21.

62. E. Duboue-Dijon, A. C. Fogarty, J. T. Hynes and D. Laage, *J Am Chem Soc*, 2016, **138**, 7610-7620.

63. J. H. Cho and R. R. Rando, *Biochemistry-Us*, 1999, **38**, 8548-8554.

64. K. F. Blount, F. Zhao, T. Hermann and Y. Tor, *J Am Chem Soc*, 2005, **127**, 9818-9829.

65. M. R. Green, J. Sambrook and J. Sambrook, *Molecular cloning : a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 4th edn., 2012.

Chapter 6   Enzymatic Synthesis of Partially and Fully

Modified RNAs

6.1 Introduction

      Nucleic acids and their canonical building blocks have been chemically modified, mainly for the development of biophysical and discovery assays. Although there are various designs available, we focus on modified surrogates, which closely resemble the size and shape of native nucleosides. Due to the structural similarity, our isomorphic mimics should fabricate functions and bio-compatibility of native building blocks.[1-4] Many probes have been designed based on the requirements of the isomorphicity[5-11], the most notable to be developed in our laboratory was the complete RNA alphabet consisting of purine (thG, thA) and pyrimidine (thC, thU) mimics published in 2011.[12]

      Previous incorporation of thG, thA, thC and thU (**30–33**) (into oligonucleotides was studied either via solid phase synthesis[13] or enzymatic transcription reactions[9, 14]. However, in both cases only one kind of the modified surrogate was incorporated into each oligonucleotide. There is a need to conduct a systematic study about incorporating multiple mimics into one RNA construct. Furthermore, affinity of T7 RNA polymerase for different modifications should be compared and closely evaluated. We choose T7 RNA polymerase-mediated transcription with consideration that, in many cases, it is more advantageous for the synthesis of longer biologically relevant constructs. Additionally, the

promiscuity of the enzyme has been previously studied by comparing three related pyrimidine mimics. Specifically, an analog where thiophene was conjugated at the C-5 position (^{5th}U) to a uridine core was enzymatically incorporated with increased yields relative to thiophene-fused analogues (e.g. thU).[9] We therefore hypothesize that conjugating a heterocyclic ring to the cytidine core may facilitate the enzymatic fabrication of modified oligonucleotides as well. Consequently, we selected two pyrimidine surrogates: ^{5fu}C (**28**) and ^{5th}U (**29**) (Figure 6.1). Furthermore, it has been previously demonstrated that guanosine analog thG (**30**) was efficiently incorporated into RNA constructs via enzymatic synthesis,[14] prompting our selection of thiophene-fused purines thG (**30**) and thA (**31**).

These findings have motivated the study reported here, where we explore affinity of T7 RNA polymerase by challenging it with incorporation of fused purines (thG, thA) and conjugated pyrimidines (^{5fu}C, ^{5th}U) in systematic way (Figure 6.1). Transcription efficiency of single/double/triple/quadruple incorporations of modified surrogates, allowing for multiple modifications in the same oligonucleotide is compared. Subsequently, the partially and fully modified transcripts were utilized in the ligation reaction mediated by 9DB1 deoxyribozyme. Both the transcription and ligation reaction are highly dependent on nucleobase recognition of the complementary strand, therefore comparing fully modified and native RNA constructs can improve our understanding of folding patterns for synthetic RNAs. Finally, we conclude that partially and fully synthetic oligonucleotides can moderately mimic functions and bio-compatibility of the native strands.
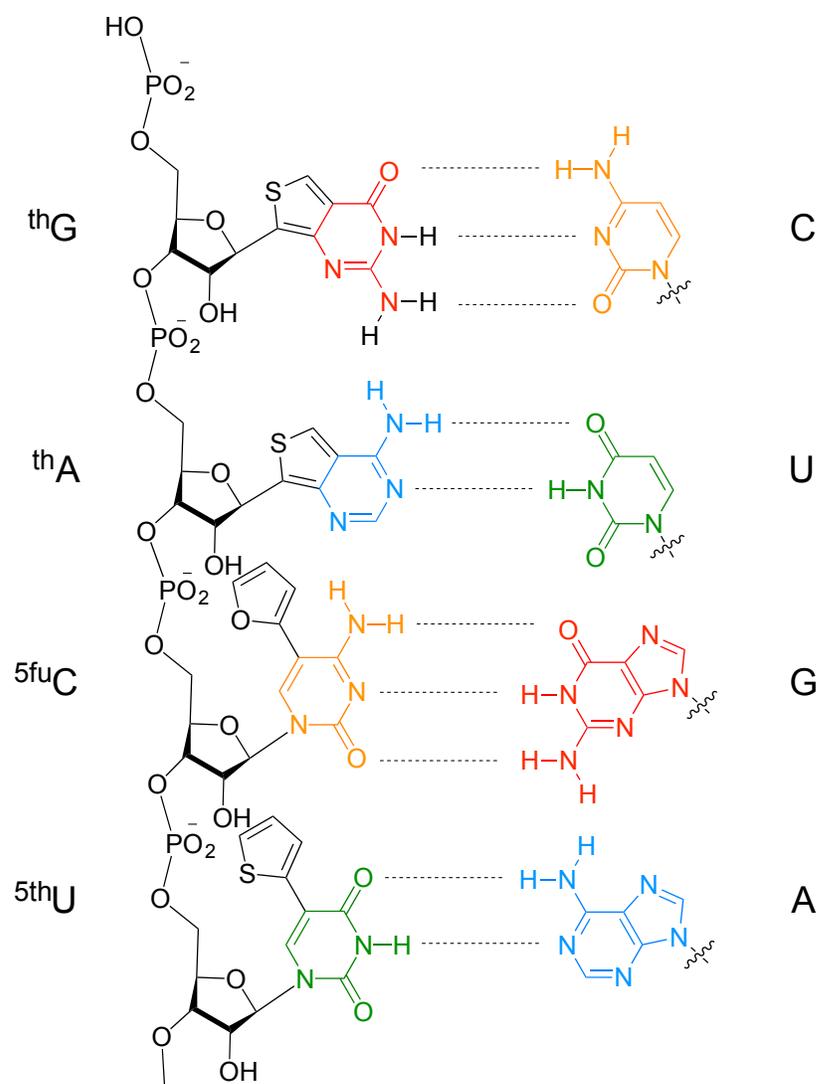
Figure 6.1 Base paring between isomorphic surrogates thG (**30**), thA (**31**), ^{5fu}C (**28**), ^{5th}U (**29**) and native nucleosides.

6.2 T7 RNA polymerase mediated transcription reactions initiated

with guanosine

To analyze enzymatic synthesis of partially and fully modified RNAs, the T7 promoter **ODN1** and template **ODN2** or **ODN4** were annealed and transcribed in the presence of native guanosine as well as native and synthetic triphosphates (Figure 6.2). A small aliquot was withdrawn from each reaction and G-terminated

transcripts were 5' labeled with T4 polynucleotide kinase in a presence of $\gamma\text{-}^{32}P$

ATP, according to standard protocols[15].
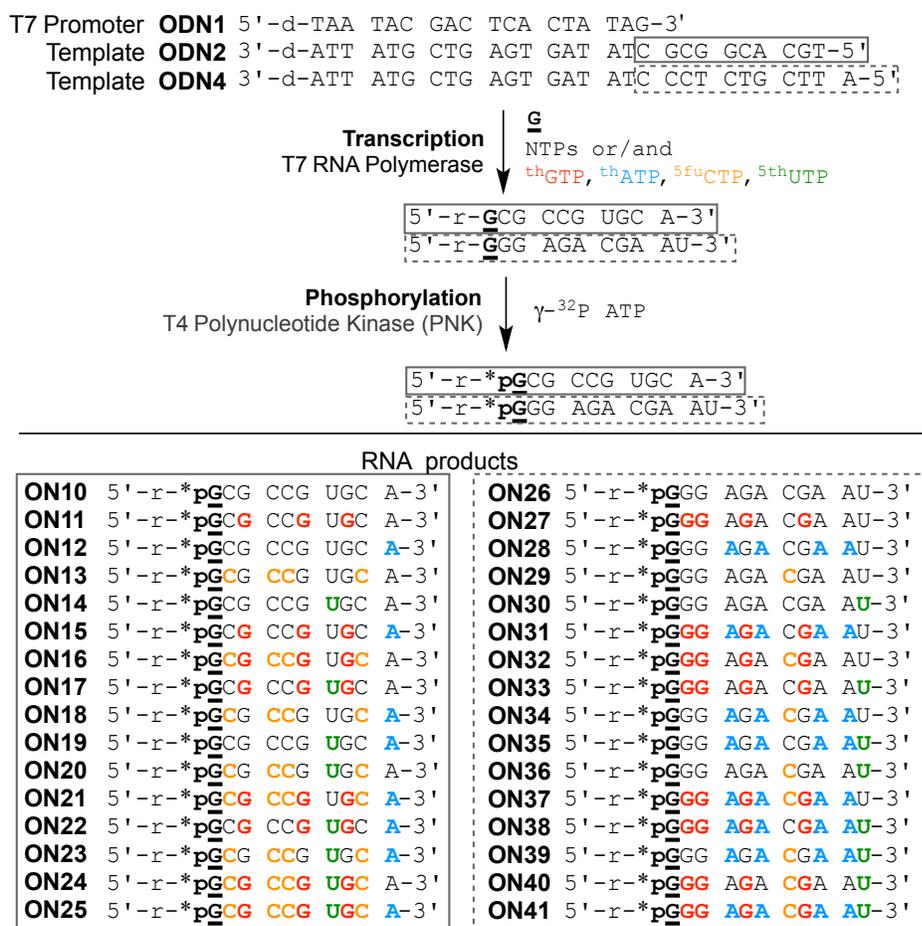
Figure 6.2 Transcription reactions with T7 promoter **ODN1** and templates **ODN2** or **ODN4** initiated with guanosine in the presence of natural NTPs and modified [th]GTP (**102**), [th]ATP (**103**), [5fu]CTP (**104**) and [5th]UTP (**105**); followed by T4 polynucleotide kinase mediated phosphorylation reactions with radioactive $\gamma\text{-}^{32}P$ ATP resulting in transcripts **ON10–ON25** and **ON26–ON41**.

A phosphorimager revealed transcription profiles with full-length products

as well as short, failed transcripts (Figure 6.3). Intensity of full-length RNA

constructs **ON10–ON25** and **ON26–ON41** was quantified and normalized to the

most intense signal corresponding to the most efficient transcription reaction

(Figure 6.3b and 6.3d, respectively).

The highest yielding transcriptions were observed when template **ODN2** was used with ^{5fu}CTP (**104**) and ^{5th}UTP (**105**) for single and double modification of oligonucleotides (**ON13**, **ON14** and **ON20**). However, when purines were incorporated less amount of final product was obtained (**ON11**, **ON12**). Additionally, when thG (**30**) and thA (**31**) were simultaneously used, oligonucleotides with one less residue (5'-r-pGCG CCG UGC-3') were transcribed as the longest constructs (**ON15**, **ON16** and **ON21**). Subsequently, template **ODN4** was utilized in a similar manner but providing different pattern, where thA (**31**) and ^{5th}U (**29**) got incorporated almost as well as the native nucleosides (**ON28**, **ON30**). Furthermore, the best yielding transcription was observed when both residues were used concurrently (**ON35, ON39**). The lowest intensity of the final product was shown whenever thG (**30**) was incorporated alone or in the presence of other surrogates.

Subsequently, templates **ODN5**–**ODN8** were selected for the incorporation of modified purines and pyrimidines into longer transcripts (Figure 6.7). Double modification of synthetic purines yielded the least amount of full-length product for 3 out of 4 templates (Figure 6.8b). Remarkably, for investigation of modified pyrimidines, ^{5th}U (**29**) surrogate was incorporated either better or as well as native uridine depending on the template (Figure 6.8d).
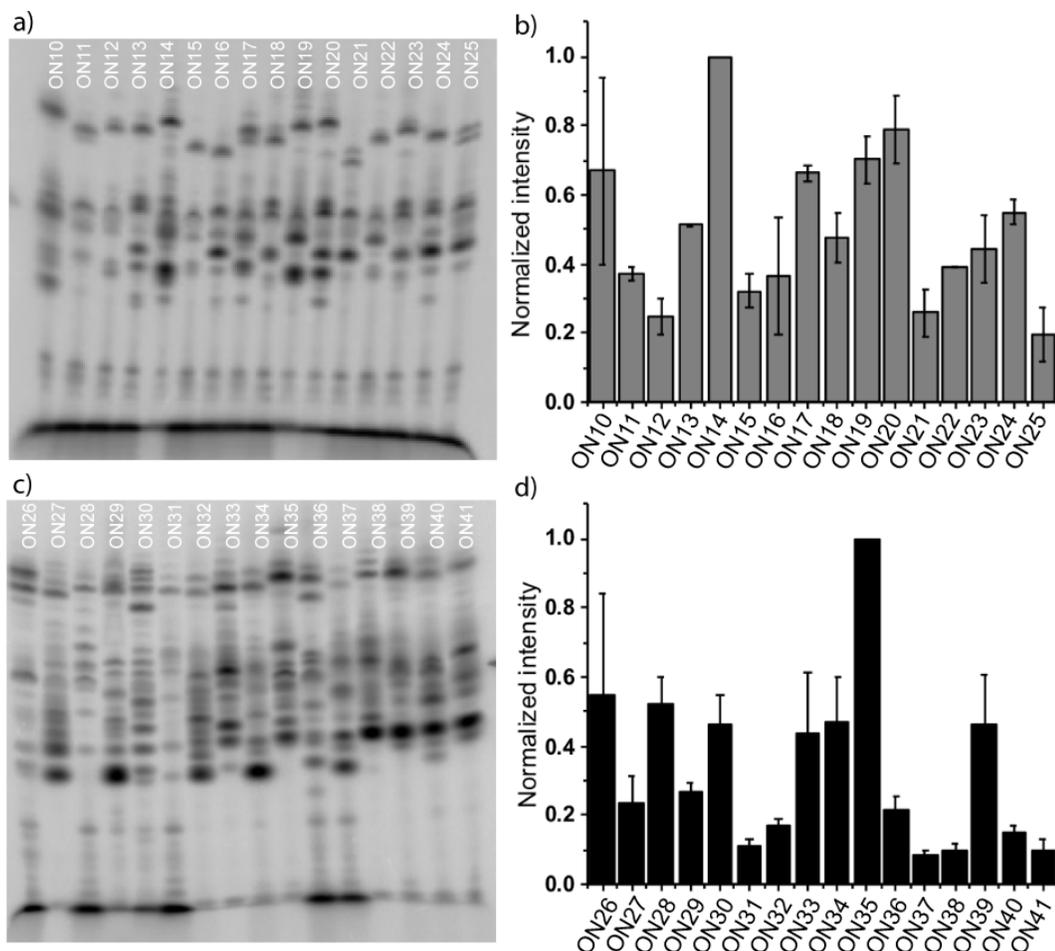
Figure 6.3 Labeled transcripts obtained after transcription reaction using a) template **ODN2**, b) template **ODN4**. Normalized intensity vs reaction number c) full length transcripts **ON10–ON25** d) full length transcripts **ON26–ON41**.

Based on these observations we hypothesized that T7 RNA polymerase frequently pauses in a close proximity to the modified mimics, more precisely directly before or after the incorporation.[9] It has been observed previously that this enzyme has a lower ability to accept unnatural substrates compared to native nucleosides.[9, 14] Although it appears that more modifications introduced at the same time will provide lower yields for the final products, we also observed that the positioning of the synthetic residues may have equally important impact on

enzyme's affinity. Moreover, T7 RNA polymerase is found to be error-prone, especially at the beginning of the initiation phase or just before the dissociation of the full-length transcript. As more residues are modified at the 5' or 3' end of the oligonucleotide, the yields are found to dramatically decrease (**ON21**, **ON37**, **ON38**) and more transcription errors are observed (**ON15**, **ON16**, **ON21** and **ON37**). We speculate that when modified surrogates are used, especially in case of ^{5fu}C and ^{5th}U, the base stacking is more intense compared to the native C or U residues. Presence of the furan or thiophene ring allows for more base stacking interactions due to larger surface. It is likely that the stability of the DNA-RNA hybrid is enhanced, when modified surrogates are present, which can perturb accurate fabrication of the native residues' functions. Hence, T7 RNA polymerase's ability to recognize when to initiate and abort transcription reaction may decrease.

6.3 9DB1 Deoxyribozyme-mediated ligation reaction: utilizing partially modified RNAs

Another known reaction depended on the formation of the DNA-RNA hybrid is ligation catalyzed by 9DB1 deoxyriboxyme (DNA enzyme, DNA catalyst or DNAzyme).[16-20] We hypothesized that positioning of modified surrogates and their recognition by canonical nucleobases can influence annealing of DNA enzyme and RNA constructs thus extending the scope on bio-compatibility of partially and fully modified oligonucleotides. Additionally, synthesis of longer partially modified oligonucleotides was introduced.

a) T7 promoter **ODN9** `5'-d-CTG TAA TAC GAC TCA CTA TAG-3'`
  Template **ODN10** `3'-d-ATT ATG CTG AGT GAT ATC TAC AAG ATC GCG GCT T-5'`

Transcription | NTPs or/and
T7 RNA Polymerase ↓ | thGTP, thATP, ^{5fu}CTP, ^{5th}UTP

ON70 `5'-r-pppG AUG UUC UAG CGC CGA A-3'`
ON71 `5'-r-pppG AUG UUC UAG CGC CGA A-3'`
ON72 `5'-r-pppA UG UUC UAG CGC CGA AU-3'`
ON73 `5'-r-pppG AUG UUC UAG CGC CGA A-3'`
ON74 `5'-r-pppG AUG UUC UAG CGC CGA A-3'`
ON75 `5'-r-pppG AUG UUC UAG CGC CGA A-3'`
ON76 `5'-r-pppG AUG UUC UAG CGC CGA A-3'`

b)

Acceptor RNA (**ON77**)   Donor RNA (**ON70–ON76**)

Figure 6.4 a) Transcription reactions with T7 promoter (**ODN9**) and template **ODN10** in the presence of natural NTPs and modified thGTP (**102**), thATP (**103**), ^{5fu}CTP (**104**) and ^{5th}UTP (**105**) resulting in transcripts **ON70–ON76**. b) Ligation reaction mediated by 9DB1 deoxyribozyme with 5'-³²P-labeled acceptor (**ON77**) and modified donors (**ON70–ON76**), resulting in ligation products (**ON78–ON84**).

Frist, 5'-triphosphorylated donors (**ON70–ON76**) were transcribed in a presence of native and modified triphosphates following previously described protocol (nucleoside guanosine was not used, Figure 6.4a). Subsequently, partially and fully synthetic donors were subjected to mild ligation conditions with 5'-³²P-labeled acceptor **ON77** (Figure 6.4b). Small aliquots were withdrawn from reaction mixtures at selected time points, and resolved by 20% PAGE (Figure 6.5). Band intensities were quantified, thus providing data of pseudo-first order kinetic profiles for ligations reaction of synthetic donors **ON70–ON76** (Figure 6.5c).

The rate constants obtained for single and double modified oligonucleotides with thG and ^{5fu}C (**ON71**, **ON73**, **ON75**) were 8.4 $\pm$ 0.1 min^{-1}, 39 $\pm$ 1 min^{-1} and 7.0 $\pm$ 0.6 min$^{-1,}$ respectively (Figure 6.5c). When thA or ^{5th}U modified donors (**ON72**, **ON74**) were used ligations was extremely slow, and when fully modified donor (**ON76**) was used reaction didn't occur.

Donor for the 9DB1 deoxyribozyme-mediated ligation was purposely selected because the ratio of nucleotides in the sequence is 5G:4A:4C:4U. Therefore, allowing to investigate if the positioning of the synthetic residues has more impact than the amount of modifications on rates of the reactions. The rate of ligation for single modified donors decreases in the general order **ON73**>**ON71**>**ON72**≈**ON74**. ^{5fu}C modified donor (**ON73**) was able to ligated almost as well as the native donor **ON70**, where thA and ^{5th}U modified donors (**ON72** and **ON74**, respectively) ligated with ~100-fold slower rate. We hypostasized that reaction's efficiency is depended on the distance of the synthetic residues from the ligation's active center. thrA2 and ^{5th}rU3 are the first two nucleotides involved in Watson-Crick base paring with DNA enzyme, therefore, formation of DNA-RNA hybrid essential for the reaction is likely disturbed.

Although thG is a first residue in donor's sequence (**ON71**), the rate of this reaction was much higher than for **ON72** or **ON74**. This is due to the fact that two reacting residues rA-1 (acceptor) and rG1 (donor) are embracing the ligation junction unpaired, before the phosphodiester bond is formed between them. Therefore, thrG1 is not hybridized to the DNA catalyst oppose to thrA2 and ^{5th}rU3.
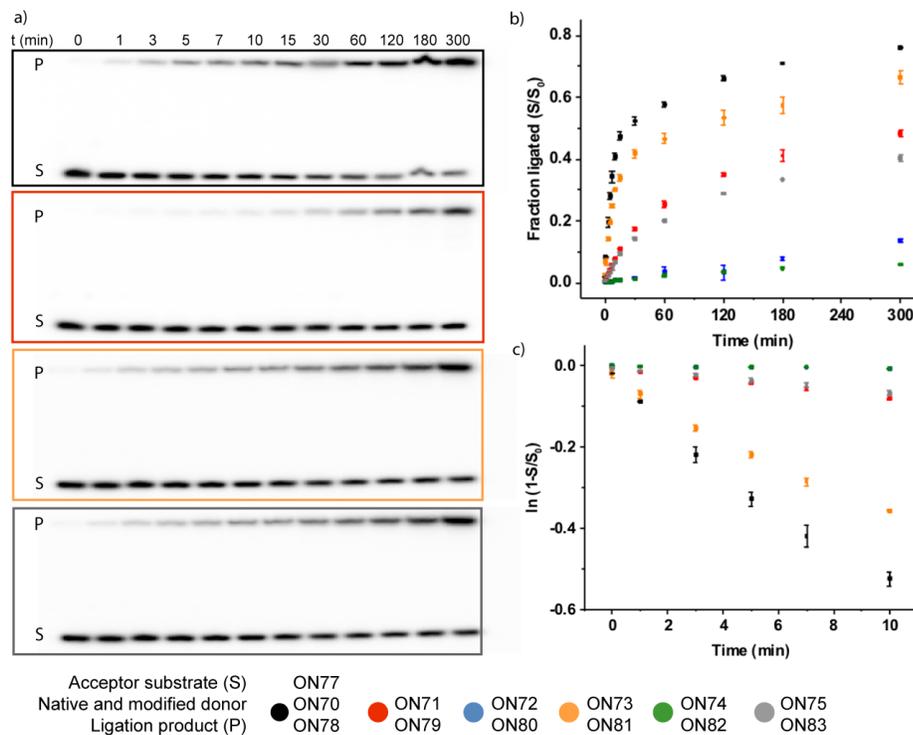
Figure 6.5 a) 9DB1 deoxyribozyme mediated ligation reaction results were followed by ^{32}P radioactive labeling of substrate strand (**ON77**) and product strands (**ON78**–**ON84**). The reactions were quenched at given time points and resolved by gel electrophoresis on a denaturation 20% gel. b) Ligation curves as determined by ^{32}P data for substrate (**ON77**) and products (**ON78**–**ON84**). Fraction cleaved (S/S_0) was determined by dividing the amount of substrate by the sum of the full length product and substrate. c) Initial kinetics of substrate (**ON77**) and products (**ON78**–**ON84**), pseudo first order rate constants (k_2) of the ligation reactions are determined as the slopes of ln(fraction ligated) versus time.

Consequently, double and fully modified donors (**ON75**–**ON76**) were investigated. **ON75** construct was able to ligate almost as well as **ON71**. Thus supporting the hypothesis that the ligation efficiency is more depended on the positioning of the synthetic nucleotides then on the amount of modifications. Notably, ligation didn't occur for fully modified donor (**ON76**).

6.4. Conclusions

In summary, we introduce synthesis of partially and fully modified RNAs, which to our knowledge have not yet been reported. These highly isomorphic surrogates (thG, thA, ^{5fu}C and ^{5th}U) are readily incorporated by T7 RNA polymerase into various length oligonucleotides. Although the affinity of the enzyme lowers when modifications are in a close proximity to initiation and dissociation phase of transcription, T7 RNA polymerase can still incorporate four mimics simultaneously. To further investigated the impact of synthetic modification on RNA function, we tested capability of efficient phosphodiester bond formation in 9DB1 DNA enzyme-mediated ligation. Both single and double modified donors **ON71**, **ON73**, **ON75** ligated to native acceptor providing an efficient way to the longer partially modified oligonucleotides.

Our results point to intriguing future applications of thG, thA, ^{5fu}C and ^{5th}U as isomorphic surrogates. Either site specific modification can be envisioned or partially and fully modified oligonucleotides can be utilized for biochemical assays.

6.5 Experimental

Synthesis of modified nucleosides (28–31) and triphosphates (102–105).

General

Tris(tetrabutylammonium) hydrogen pyrophosphate, phosphoryl chloride, trimethyl phosphate, 1,8-bis(dimethylamino)naphthalene (proton sponge), anhydrous DMF were purchased from Sigma-Aldrich. All other solvents were purchased from Fisher Scientific. All reagents and solvents were used without further purification.
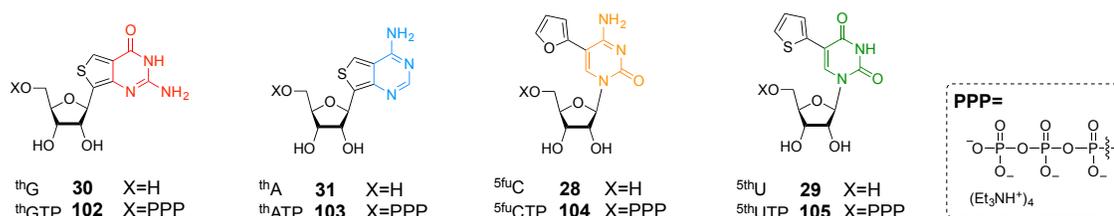
Figure 6.6 Modified nucleosides (**28–31**) and triphosphates (102–105).

a) Purine surrogates

The modified purine nucleosides [th]G (**30**) and [th]A (**31**) were previously synthesized and reported.[12] Furthermore the procedure for the synthesis of [th]GTP (**102**) was published as well.[14]
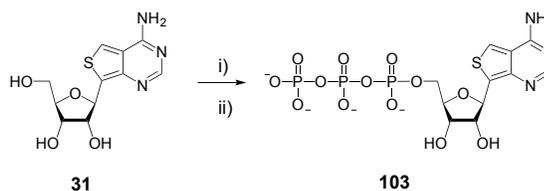
Scheme 6.1 Synthesis of modified nucleoside triphosphate **103**. Reagents and conditions: i) POCl$_3$, (MeO)$_3$PO; ii) tributylammonium pyrophosphate, proton sponge, 0–4°C.

Here we report synthesis of thATP (**103**) following the standard procedures.[9, 21-23]

Tris(tetrabutylammonium) hydrogen pyrophosphate (0.91 g, 1.0 mmol) in a 10 mL round-bottom flask, and **31** (66 mg, 0.20 mmol) in a 25 mL round-bottom flask, were separately coevaporated with anhydrous pyridine and dried (3 x 1mL). Trimethyl phosphate (2 mL) was added to **31** and cooled in an ice bath to 0 °C. Phosphoryl chloride (46 µL, 0.5 mmol) was added slowly, and the reaction was stirred for 2 h at 0 °C. The coevaporated tris(tetrabutylammonium) hydrogen pyrophosphate was dissolved in 2 mL of anhydrous DMF and added to the reaction mixture with **31**. Then proton sponge (216 mg, 0.1 mmol) was added and the reaction was kept stirring at 0 °C for 40 min. To the reaction mixture was added 6 mL of 1 M triethylammonium bicarbonate buffer (TEAB), and the mixture was stirred briefly. The mixture was then transferred to a separatory funnel and washed with 10 mL of ethyl acetate. The organic layer was then back-extracted with 5 mL of 1 M TEAB. The aqueous layers were combined and concentrated under reduced pressure at room temperature to afford an oily yellow residue. The residue was dissolved in 10 mL of 0.05 M ammonium bicarbonate buffer and loaded onto a DEAE Sephadex A25 anion-exchange column kept in a cold room at 4 °C. The column was eluted using a gradient mixer with 0.01−1.0 M of ammonium bicarbonate buffer. A fraction collector was used to collect 260 fractions that were about 8 mL (220 drops). The fractions containing the triphosphate were evaporated under reduced pressure at 10 °C, and then the residue was lyophilized. The

triphosphate was further purified by HPLC (Phenomenex Synergi Fusion-RP 80A C18 column, 4 μm, 250 × 10 nm, 5−20% acetonitrile in 50 mM TEAA buffer, pH 6.0, 30 min). Appropriate fractions were lyophilized and treated with 3% sodium perchlorate in acetone and then with Chelex 100 for 30 minutes with shaking, filtrated and dried to afford nucleotide **103** (29 mg, 20%). ^{1}H NMR (500 MHz, D$_2$O) δ 8.19 (s, 1H), 7.89 (s, 1H), 5.44 − 5.37 (m, 1H), 4.35 − 4.30 (m, 1H), 4.30−4.23 (m, 1H), 4.21 − 4.15 (m, 1H), 4.15 − 4.04 (m, 2H) and triethylammonium (CH$_3$CH$_2$)$_3$N$^+$ counter ion (0.33 equiv.) 3.04 (q, J = 7.3 Hz, 2H), 1.13 (t, J = 7.3 Hz, 3H); ^{13}C NMR (125 MHz, D$_2$O) δ 173.49, 157.57, 151.44, 130.83, 124.53, 118.77, 83.53, 77.27, 76.71, 71.39, 65.64 and triethylammonium (CH$_3$CH$_2$)$_3$N$^+$ counter ion (0.33 equiv.) 46.63, 8.25; ^{31}P NMR (202 MHz, D$_2$O) δ −7.58 (d, J = 19.2 Hz, Pγ), −11.16 (d, J = 19.2 Hz, Pα),−22.13 (t, J = 19.2 Hz, Pβ); ESI-MS (negative ion mode) [C$_{11}$H$_{16}$N$_3$O$_{13}$P$_3$S]− calculated 522.96, found 522.09.

b) Pyrimidine surrogates

Both the parent nucleoside [5th]U (**29**) and triphosphate [5th]UTP (**105**) were previously synthesized and reported.[9]
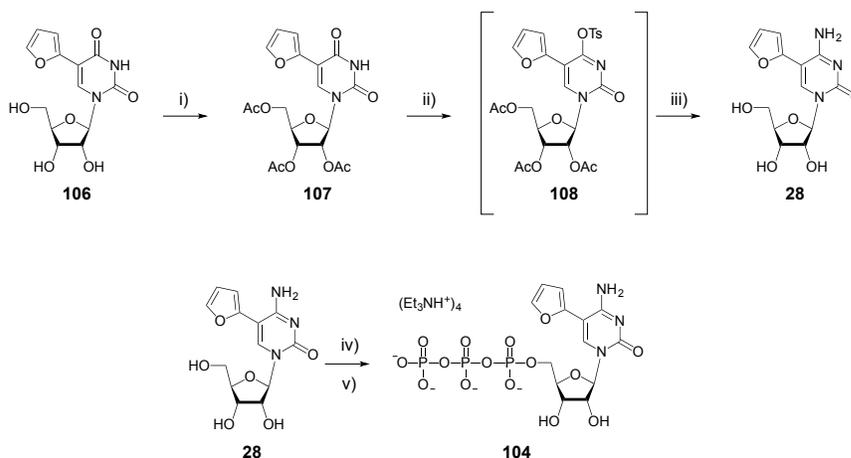
Scheme 6.2 Synthesis of modified nucleoside **28** and its triphosphate **104**. Reagents and conditions: i) Ac_2O, pyridine ii) TsCl, Et_3N, acetonitrile iii) 37% NH_4OH, acetonitrile iv) $POCl_3$, $(MeO)_3PO$; v) tributylammonium pyrophosphate, proton sponge, 0–4°C.

Here we report synthesis of cytidine surrogate [5fu]C (**28**) and [5fu]CTP (**104**).

Both [5fu]U and acetyl protected [5fu]U were previously synthesized and reported.[6, 24] Acetyl protected [5fu]U (**107**) (2.00 g, 4.58 mmol) was dissolved in dry acetonitrile (50 mL) and treated with tosyl chloride (1.75 g, 9.16 mmol, 2 equiv). After the mixture was stirred for 2 minutes, triethyl amine (1.41 mL, 10.08 mmol, 2.2 equiv) was added and left for 16 h at room temperature. The reaction mixture was diluted with EtOAc (100 mL) and washed with water (3 × 100 mL), and brine (5 mL). The combined organic layers were dried over Na_2SO_4, filtered and concentrated to dryness. Obtained solid was further washed with hexanes (3 × 20 mL). Crude residue was dissolved in dry acetonitrile (40 mL), aqueous ammonia

(37%, 20 mL) was added and heated at 80 °C in a pressure vessel for 24 h. The mixture was cooled to rt, concentrated to dryness, dissolved in methanol (40 mL) and treated with Dowex Monosphere 500A (OH) anion exchange resin for 16 h. The mixture was filtrated and concentrated to dryness. Purification of the crude residue by column chromatography (silica, 10 v% MeOH in DCM) yielded **28** as an off-white solid (487.2 mg, 1.58 mmol, 35%). ^{1}H NMR (400 MHz, DMSO-d$_6$) δ 8.37 (s, 1H), 7.86 – 7.56 (m, 2H), 6.88 – 6.38 (m, 3H), 5.88 – 5.70 (m, 1H), 5.50−5.26 (m, 1H), 5.25 – 5.13 (m, 1H), 5.02 – 4.92 (m, 1H), 4.03 – 3.91 (m, 2H), 3.91−3.80 (m, 1H), 3.76 – 3.64 (m, 1H), 3.62 – 3.50 (m, 1H); ^{13}C NMR (125 MHz, DMSO-d$_6$) δ 162.24, 154.42, 147.90, 142.91, 140.71, 111.83, 106.98, 98.49, 89.98, 84.34, 74.86, 69.22, 60.29; ESI-MS (positive ion mode) [C$_{13}$H$_{15}$N$_3$O$_6$]$^-$ calculated 309.10, found 310.22 and 332.31 as [M − 2H + Na]$^-$.

Using the procedure described for **103**, but starting from **28** (0.080 g, 0.26 mmol), **104** was obtained as a white solid (61.2 mg, 0.09 mmol, 37%). ^{1}H NMR (400 MHz, D$_2$O) δ 7.94 (s, 1H), 7.55 – 7.41 (m, 1H), 6.68 – 6.58 (m, 1H), 6.47 – 6.41 (m, 1H), 5.89−5.85 (m, 1H), 4.34 – 4.22 (m, 2H), 4.17 – 4.05 (m, 3H) and triethylammonium (CH$_3$CH$_2$)$_3$N$^+$ counter ion (0.66 equiv.) 3.03 (q, J = 7.3 Hz, 3H), 1.11 (t, J = 7.3 Hz, 6H); ^{13}C NMR (125 MHz, D$_2$O) δ 156.54, 145.93, 143.00, 139.32, 111.56, 108.45, 89.14, 82.93, 82.86, 74.07, 68.99, 64.58, 64.54 and triethylammonium (CH$_3$CH$_2$)$_3$N$^+$ counter ion (0.66 equiv.) 46.43, 8.06; ^{31}P NMR (202 MHz, D$_2$O) δ −6.65 (d, J = 20.1 Hz, Pγ), −11.33 (d, J = 19.5 Hz, Pα),−21.96

(t, J = 19.9 Hz, Pβ); ESI-MS (negative ion mode) $[C_{15}H_{18}N_3O_{15}P_3]^-$ calculated 549.00, found 548.06.

Oligonucleotides synthesis and purification

General

All unmodified DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies, Inc. Native NTPs were purchased from Thermo Scientific Dharmacon and ATP–[γ–32P] (6000 Ci/mmol; 10 mCi/ml) was purchased from Perkin Elmer. T4 Polynucleotide Kinase (PNK) and kinase buffer (2500 U) were purchased from New England BioLabs (NEB). 50% PEG 4000 and RiboLock (40 U/μl) were purchased from Fermentas Inc. All other solutions and buffers were prepared in our laboratory.

T7 RNA polymerase-mediated *in vitro* transcription reaction

Small scale of short and longer oligonucleotides

Transcription reactions with T7 RNA polymerase and the analytically pure native and modified triphosphates (thGTP, thATP, ^{5fu}CTP, ^{5th}UTP) were performed to analyze their enzymatic incorporation into short (**ON10**–**ON25, ON26**–**ON41)** and longer (**ON42**–**ON48, ON49**–**ON55, ON56**–**ON62, ON63**–**ON69**) RNA oligonucleotides (Figure 6.7). The T7 promoter **ODN1** and templates **ODN2**, **ODN4**–**ODN8** were annealed, and transcribed in the presence of natural NTPs or

with thGTP replacing GTP, with thATP replacing ATP, with ^{5fu}CTP replacing CTP, with ^{5th}UTP replacing UTP. Transcription reaction (10 μL) cocktail contains: annealed template/promoter (500nM), 1X transcription buffer (pH 9.0), MgCl$_2$ (16 mM), dithiothreitol DTT (10 mM), RiboLock (1 U/μL), T7 RNA polymerase (0.15 μg/μL), native and modified triphosphates (1 mM), guanosine nucleoside (5 mM). Next, transcription reaction was incubated for 4 h at 37 °C. A small aliquot was withdrawn from each reaction and G-terminated transcripts were 5' labeled with T4 polynucleotide kinase in a presence of γ–^{32}P ATP, according to standard protocols[15]. Phosphorylation reaction cocktail contains: oligonucleotide (3 μL), γ–^{32}P ATP (1 μL of 6000 Ci/mmol; 10 mCi/ml), dithiothreitol DTT (5 mM), 10X kinase buffer (0.1 v/v), PNK (16.7 U/μL). Reaction mixture was inculcated for 2 h at 37 °C and loaded onto a preparative 20% denaturing polyacrylamide gel. The products on the gel were analyzed using phosphorimager. Transcription efficiencies are reported after normalization to the most efficient reaction.

```
T7 Promoter ODN1  5'-d-TAA TAC GAC TCA CTA TAG-3'
     Template ODN2  3'-d-ATT ATG CTG AGT GAT ATC GCG GCA CGT-5'
     Template ODN4  3'-d-ATT ATG CTG AGT GAT ATC CCT CTG CTT A-5'
     Template ODN5  3'-d-ATT ATG CTG AGT GAT ATC GCC ACT TCA GCG GCG-5'
     Template ODN6  3'-d-ATT ATG CTG AGT GAT ATC CTG TAC TCC TAA TGG GTA CA-5'
     Template ODN7  3'-d-ATT ATG CTG AGT GAT ATC CCT CTG CTT ATC CGG CTT TCC CGG CTT TGC AAG CG-5'
     Template ODN8  3'-d-ATT ATG CTG AGT GAT ATC GCT ACT GGA CTA CTC CGG CTT TCC GGC TTT GCA AGC G-5'
```

Transcription
T7 RNA Polymerase | G / NTPs or/and / thGTP, thATP, 5fuCTP, 5thUTP

```
5'-r-GCG CCG UGC A-3'
5'-r-GGG AGA CGA AU-3'
5'-r-GCG GUG AAG UCG CCG C-3'
5'-r-GGA CAU GAG GAU UAC CCA UGU-3'
5'-r-GGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
5'-r-GCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
```

Phosphorylation
T4 Polynucleotide Kinase (PNK) | γ-32P ATP

```
5'-r-*pGCG CCG UGC A-3'
5'-r-*pGGG AGA CGA AU-3'
5'-r-*pGCG GUG AAG UCG CCG C-3'
5'-r-*pGGA CAU GAG GAU UAC CCA UGU-3'
5'-r-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
5'-r-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
```

NTPs, modified triphosphates | **RNA products**

GTP column	ATP column	CTP column	UTP column	ON	Sequence
GTP	ATP	CTP	UTP	ON10	5'-r-*pGCG CCG UGC A-3'
thGTP	ATP	CTP	UTP	ON11	5'-r-*pGCG CCG UGC A-3'
GTP	thATP	CTP	UTP	ON12	5'-r-*pGCG CCG UGC A-3'
GTP	ATP	5fuCTP	UTP	ON13	5'-r-*pGCG CCG UGC A-3'
GTP	ATP	CTP	5thUTP	ON14	5'-r-*pGCG CCG UGC A-3'
thGTP	thATP	CTP	UTP	ON15	5'-r-*pGCG CCG UGC A-3'
thGTP	ATP	5fuCTP	UTP	ON16	5'-r-*pGCG CCG UGC A-3'
thGTP	ATP	CTP	5thUTP	ON17	5'-r-*pGCG CCG UGC A-3'
GTP	thATP	5fuCTP	UTP	ON18	5'-r-*pGCG CCG UGC A-3'
GTP	thATP	CTP	5thUTP	ON19	5'-r-*pGCG CCG UGC A-3'
GTP	ATP	5fuCTP	5thUTP	ON20	5'-r-*pGCG CCG UGC A-3'
thGTP	thATP	5fuCTP	UTP	ON21	5'-r-*pGCG CCG UGC A-3'
thGTP	thATP	CTP	5thUTP	ON22	5'-r-*pGCG CCG UGC A-3'
thGTP	thATP	5fuCTP	5thUTP	ON23	5'-r-*pGCG CCG UGC A-3'
thGTP	ATP	5fuCTP	5thUTP	ON24	5'-r-*pGCG CCG UGC A-3'
thGTP	thATP	5fuCTP	5thUTP	ON25	5'-r-*pGCG CCG UGC A-3'

				ON	Sequence
				ON26	5'-r-*pGGG AGA CGA AU-3'
				ON27	5'-r-*pGGG AGA CGA AU-3'
				ON28	5'-r-*pGGG AGA CGA AU-3'
				ON29	5'-r-*pGGG AGA CGA AU-3'
				ON30	5'-r-*pGGG AGA CGA AU-3'
				ON31	5'-r-*pGGG AGA CGA AU-3'
				ON32	5'-r-*pGGG AGA CGA AU-3'
				ON33	5'-r-*pGGG AGA CGA AU-3'
				ON34	5'-r-*pGGG AGA CGA AU-3'
				ON35	5'-r-*pGGG AGA CGA AU-3'
				ON36	5'-r-*pGGG AGA CGA AU-3'
				ON37	5'-r-*pGGG AGA CGA AU-3'
				ON38	5'-r-*pGGG AGA CGA AU-3'
				ON39	5'-r-*pGGG AGA CGA AU-3'
				ON40	5'-r-*pGGG AGA CGA AU-3'
				ON41	5'-r-*pGGG AGA CGA AU-3'

GTP col	ATP col	CTP col	UTP col	ON	Sequence
GTP	ATP	CTP	UTP	ON42	5'-*pGCG GUG AAG UCG CCG C-3'
thGTP	ATP	CTP	UTP	ON43	5'-*pGCG GUG AAG UCG CCG C-3'
GTP	thATP	CTP	UTP	ON44	5'-*pGCG GUG AAG UCG CCG C-3'
thGTP	thATP	CTP	UTP	ON45	5'-*pGCG GUG AAG UCG CCG C-3'
GTP	ATP	5fuCTP	UTP	ON46	5'-*pGCG GUG AAG UCG CCG C-3'
GTP	ATP	CTP	5thUTP	ON47	5'-*pGCG GUG AAG UCG CCG C-3'
GTP	ATP	5fuCTP	5thUTP	ON48	5'-*pGCG GUG AAG UCG CCG C-3'

GTP col	ATP col	CTP col	UTP col	ON	Sequence
GTP	ATP	CTP	UTP	ON49	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'
thGTP	ATP	CTP	UTP	ON50	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'
GTP	thATP	CTP	UTP	ON51	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'
thGTP	thATP	CTP	UTP	ON52	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'
GTP	ATP	5fuCTP	UTP	ON53	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'
GTP	ATP	CTP	5thUTP	ON54	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'
GTP	ATP	5fuCTP	5thUTP	ON55	5'-*pGGA CAU GAG GAU UAC CCA UGU-3'

GTP col	ATP col	CTP col	UTP col	ON	Sequence
GTP	ATP	CTP	UTP	ON56	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
thGTP	ATP	CTP	UTP	ON57	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
GTP	thATP	CTP	UTP	ON58	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
thGTP	thATP	CTP	UTP	ON59	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
GTP	ATP	5fuCTP	UTP	ON60	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
GTP	ATP	CTP	5thUTP	ON61	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'
GTP	ATP	5fuCTP	5thUTP	ON62	5'-*pGGG AGA CGA AUA GGC CGA AAG GGC CGA AAC GUU CGC-3'

GTP col	ATP col	CTP col	UTP col	ON	Sequence
GTP	ATP	CTP	UTP	ON63	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
thGTP	ATP	CTP	UTP	ON64	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
GTP	thATP	CTP	UTP	ON65	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
thGTP	thATP	CTP	UTP	ON66	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
GTP	ATP	5fuCTP	UTP	ON67	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
GTP	ATP	CTP	5thUTP	ON68	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'
GTP	ATP	5fuCTP	5thUTP	ON69	5'-*pGCG AUG ACC UGA UGA GGC CGA AAG GCC GAA ACG UUC GC-3'

Figure 6.7 Transcription reactions with T7 promoter **ODN1** and templates **ODN2** or **ODN4**–**ODN8** initiated with guanosine in the presence of natural NTPs and modified thGTP (**102**), thATP (**103**), 5fuCTP (**104**) and 5thUTP (**105**); followed by T4 polynucleotide kinase mediated phosphorylation reactions with radioactive γ-32P ATP resulting in transcripts **ON10**–**ON69**.

Big scale of transcripts used as donors for the ligation reaction

Transcription reactions with T7 RNA polymerase and the analytically pure native and modified triphosphates (thGTP, thATP, ^{5fu}CTP, ^{5th}UTP) were performed to obtain partially and fully modified donors (**ON70–ON76)** for 9DB1 deozyriboxyme-madiated ligation. The T7 promoter **ODN9** and template **ODN10** were annealed, and transcribed in the presence of natural NTPs or with thGTP replacing GTP, with thATP replacing ATP, with ^{5fu}CTP replacing CTP, with ^{5th}UTP replacing UTP. An agarose gel electrophoresis revealed full-length products (transcript **ON70–ON76**). Transcription reaction cocktail contains: annealed template/promoter (500 nM), 1X transcription buffer (pH 9.0), MgCl$_2$ (30 mM), dithiothreitol DTT (10 mM), RiboLock (1 U/µL), T7 RNA polymerase (0.15 µg/µL), native and modified triphosphates (3 mM). Next, transcription reaction was incubated for 4 h at 37 °C. The precipitated magnesium pyrophosphate was removed by centrifugation. The reaction was concentrated to half of the volume, subsequently 125 µL of loading buffer was added. The reaction mixture was heated at 75 °C for 3 min, and loaded onto a preparative 20% denaturing polyacrylamide gel. The gel was UV shadowed; appropriate bands were excised, extracted with 0.5 M ammonium acetate, and desalted on a Sep-Pak column.
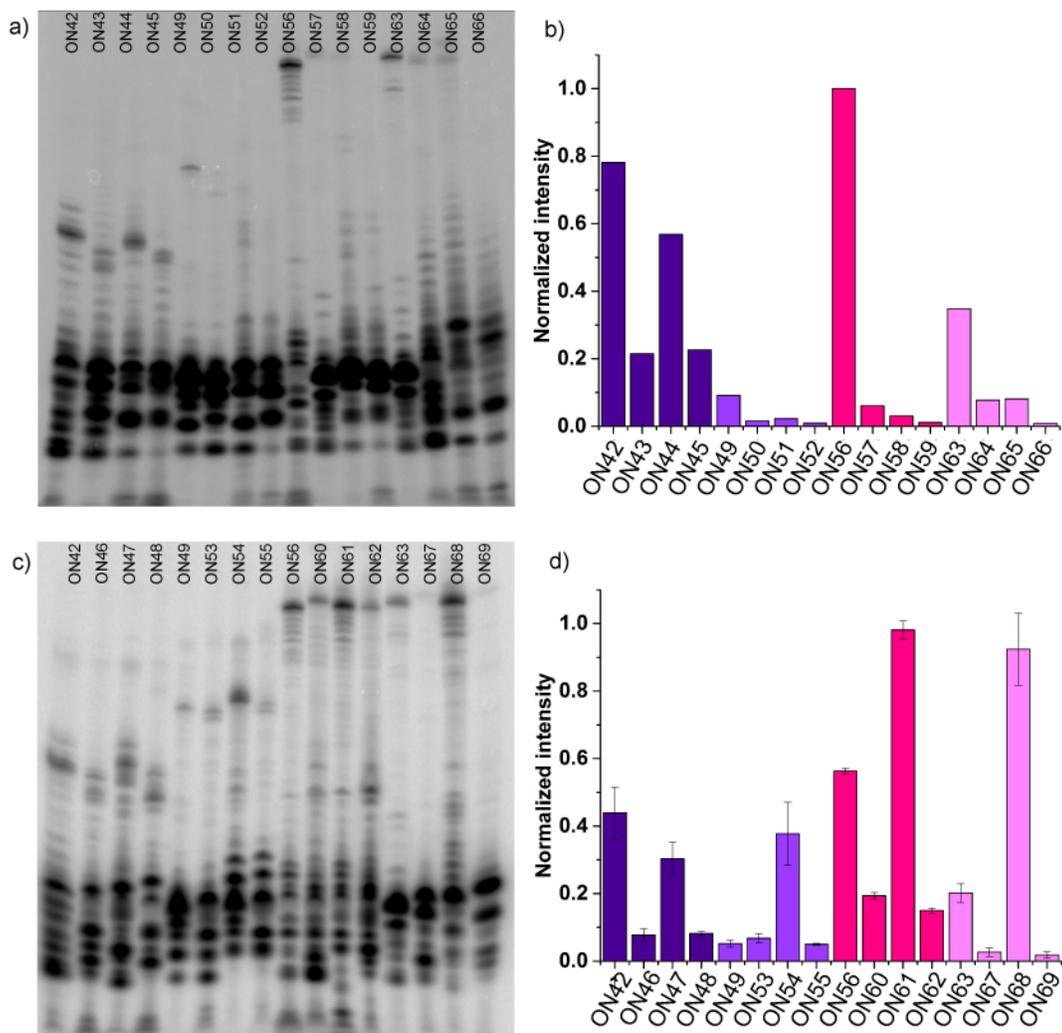
Figure 6.8 Labeled transcripts obtained after transcription reaction using templates **ODN5-ODN8**, Transcripts with single and double modification of synthetic purines a) and pyrimidines b). Normalized intensity vs reaction number c) full length transcripts **ON42**–**ON44**, **ON4**9–**ON52**, **ON56**–**ON59** and **ON63**–**ON66** d) full length transcripts **ON42**, **ON46**–**ON48**; **ON49**, **ON53**–**ON55**; **ON56**, **ON60**–**ON62** and **ON63**, **ON67**–**ON69**.

9DB1 deoxyribozyme-mediated *in vitro* ligation reaction

T4 Polynucleotide Kinase (PNK)-Mediated Phosphorylation

Acceptor RNA construct for ligation reaction was purchased from IDT Integrated DNA Technologies and purified before phosphorylation. Phosphorylation reaction cocktail contains: oligonucleotide (0.005mM), $\gamma–^{32}$P ATP (1.0 μL [6000 Ci/mmol; 10 mCi/ml]), dithiothreitol DTT (5 mM), 10X kinase buffer (0.1 v/v), PNK (1 U/ μL). Reaction mixture was inculcated for 2 h at 37 °C and used in the next step. [15]

9DB1 deoxyribozyme-mediated *in vitro* ligation reaction

Ligation reactions with 9DB1 deoxyribozyme and analytically pure donors (transcript **ON70**–**ON76**) and 5'-^{32}P-labeled acceptor were performed. The reaction consists of two steps. First, acceptor (2 pmol, 1 equiv.), donor (20 pmol, 10 equiv.) and 9DB1 deoxyribozyme (10 pmol, 5 equiv.) and HEPES annealing buffer (5 mM, pH 7.5) were mixed and heat at 95 °C for 2 minutes and cooled down slowly (15 min). Second, HEPES ligation buffer (50 mM) and $MnCl_2$ (20 mM) were added to the previous reaction mixture and incubated for 5 h at 37 °C. A small aliquot (2 μL) was withdrawn and quenched in 8 μL of quenching buffer at selected time points (0, 1, 3, 5, 7, 10, 15, 30, 60, 120, 180, 300 [min]). Samples were resolved by 20% denaturing polyacrylamide gel. The gel was exposed to Phosphor storage screen and the scanned image is analyzed by QuantityOne.[19, 20] Rate constants (*k*) were calculated as the slope of *ln*(1-(P/(P+S)) vs. time, where S is a substrate (5'-^{32}P-labeled acceptor) and P is a product of ligation reaction between acceptor and donor.

HEPES annealing buffer (pH 7.5, 50 mM HEPES, 1 mM EDTA, 150 mM NaCl).

HEPES ligation buffer (pH 7.5, 250 mM HEPES, 10 mM KCl, 750mM NaCl).

```
ON78   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA A-3'
ON79   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA A-3'
ON80   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA AU-3'
ON81   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA A-3'
ON82   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA A-3'
ON83   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA A-3'
ON84   5'-r-pGGA AGU CUC AUG UAC UAG AUG UUC UAG CGC CGA A-3'
```
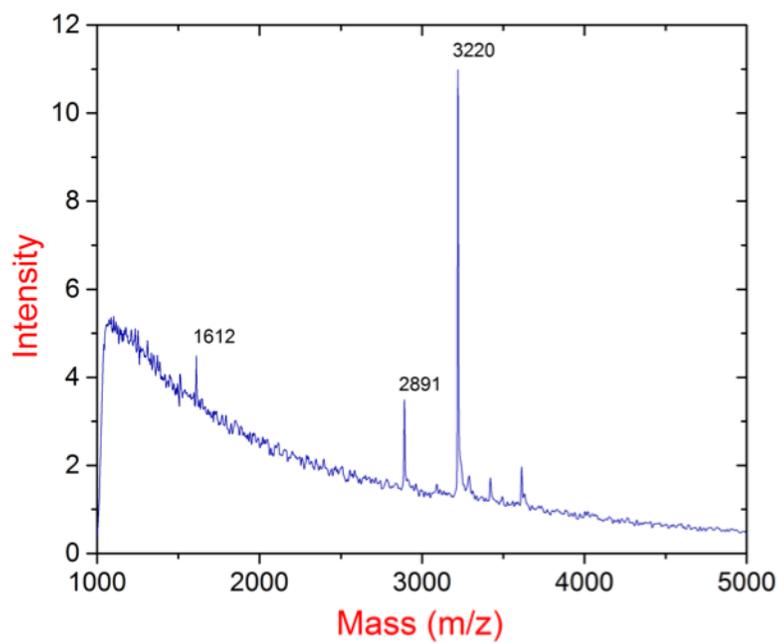
Figure 6.9 9DB1 deoxyribozyme mediated ligation reaction resulted in 5'-^{32}P-labeled radioactive product strands (**ON78**–**ON84**).

MALDI -TOF MS measurements
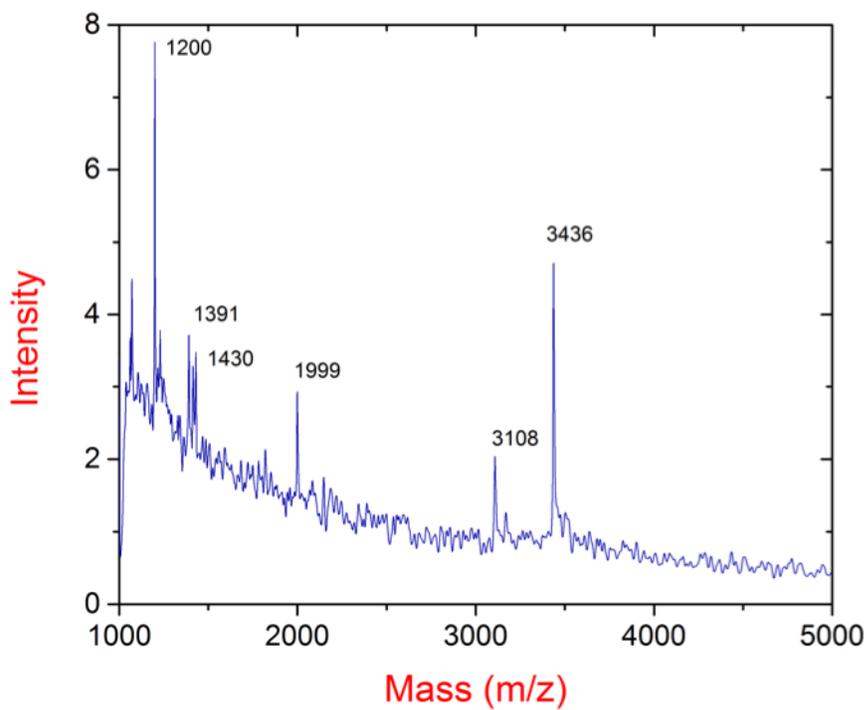
Transcripts **ON10**–**ON25**
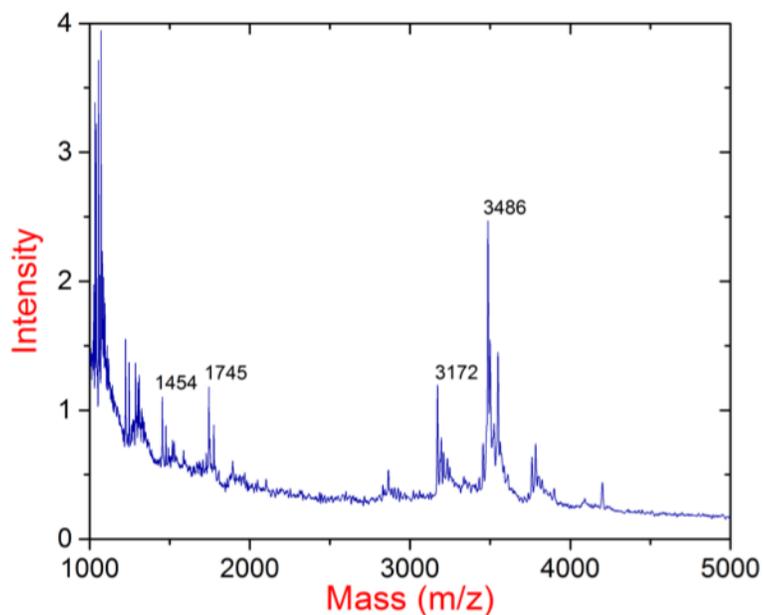
Spectrum 6.1. MALDI-TOF MS results for transcript **ON10** calculated [M] 3173 found 3172. 5'-r-pppGCG CCG UGC A-3' calculated [M+2K] 3491 found 3486.
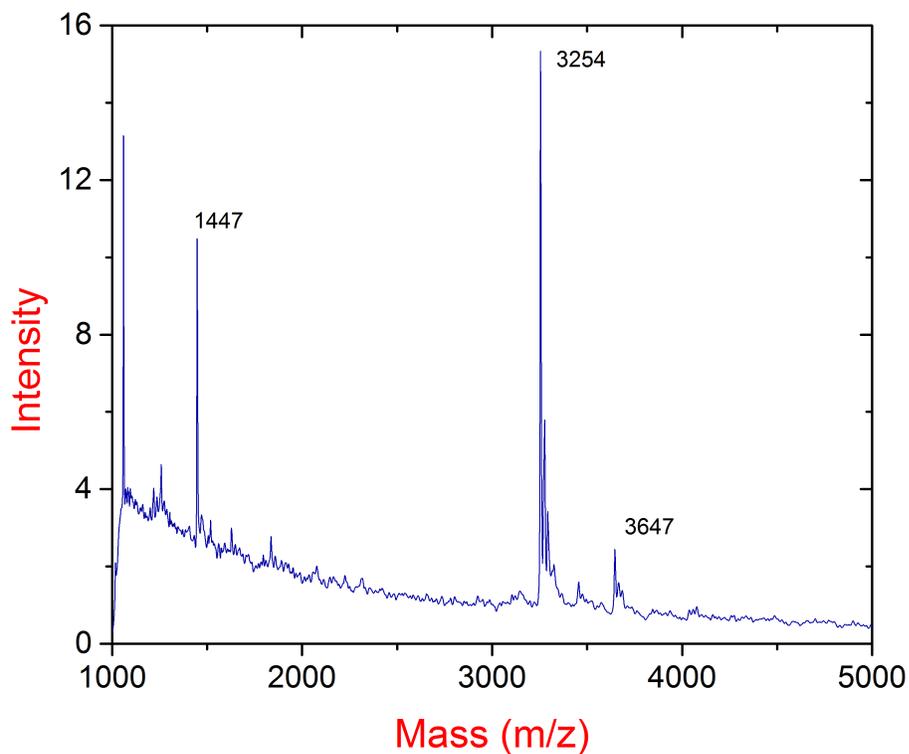

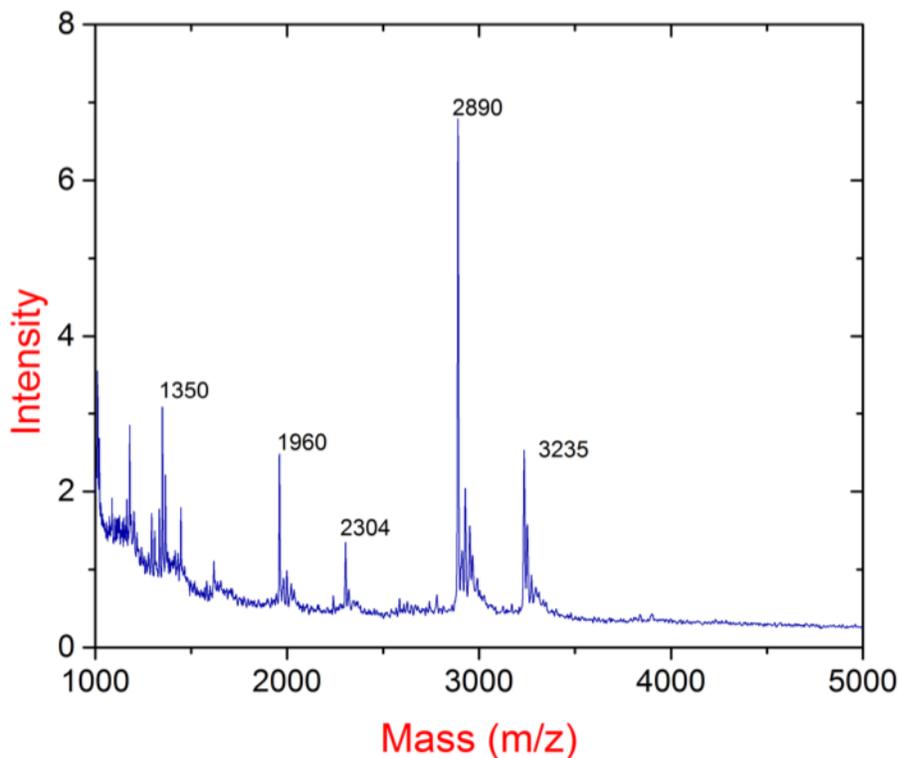Spectrum 6.2 MALDI-TOF MS results for transcript **ON11** calculated [M] 3221 found 3220.
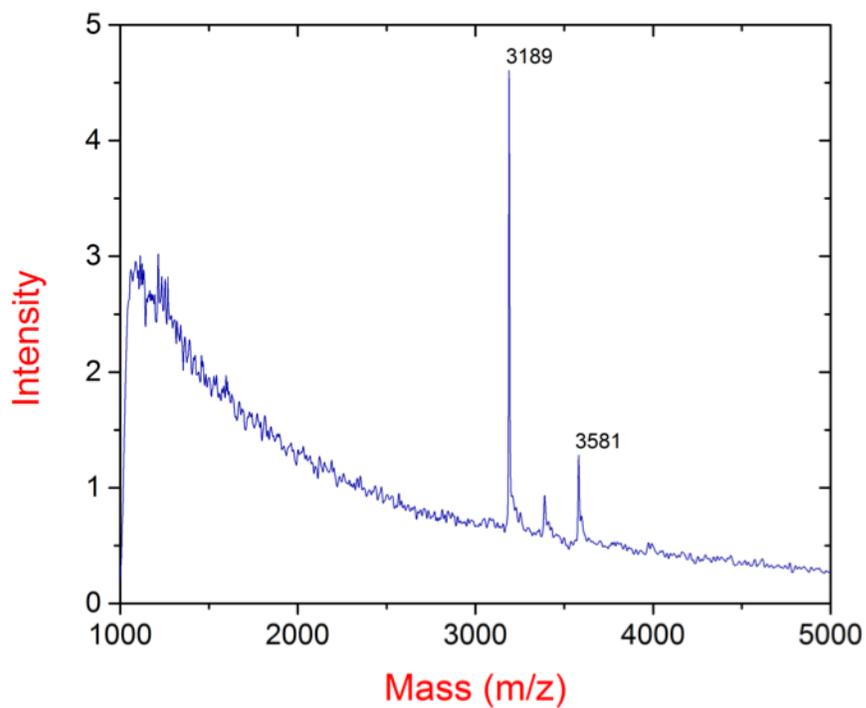
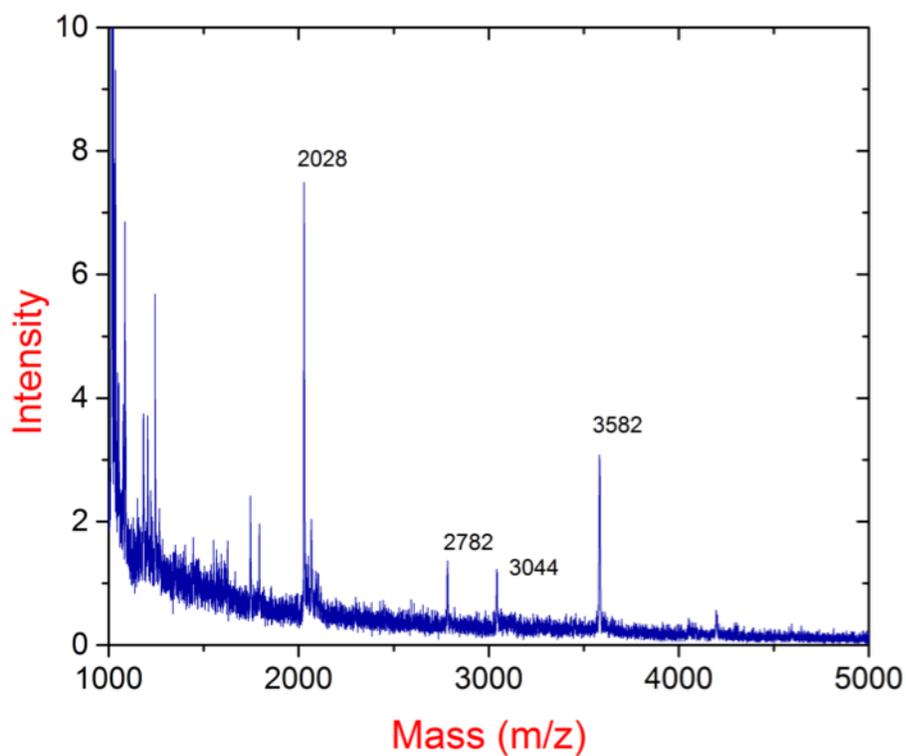Spectrum 6.3 MALDI-TOF MS results for transcript **ON12** calculated [M] 3189 found 3189.

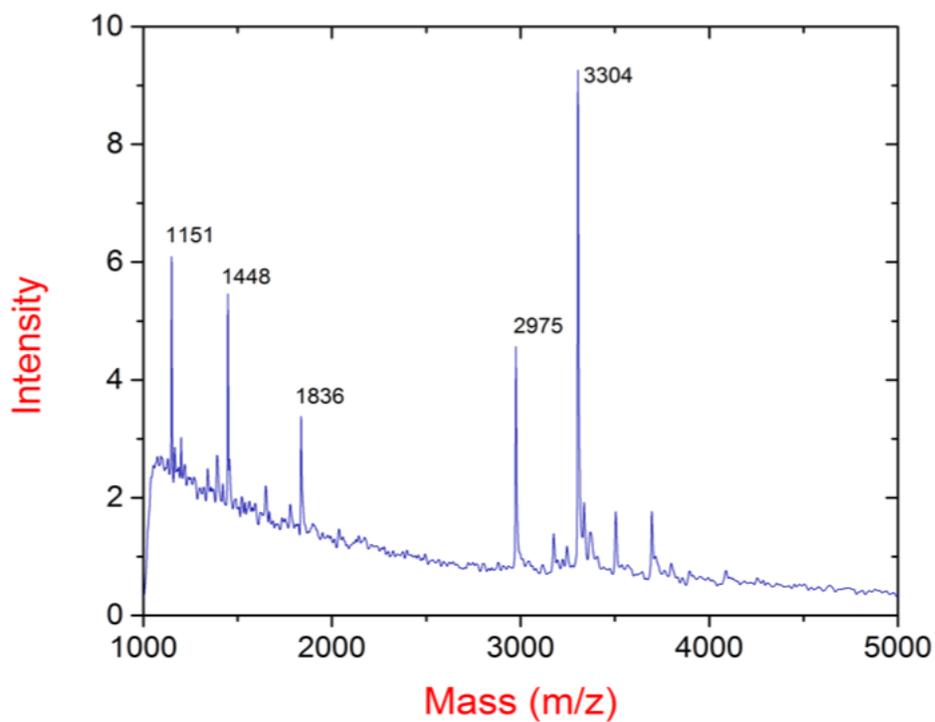Spectrum 6.4 MALDI-TOF MS results for transcript **ON13** calculated [M] 3437 found 3436.

Spectrum 6.5 MALDI-TOF MS results for transcript **ON14** calculated [M] 3255 found 3254.
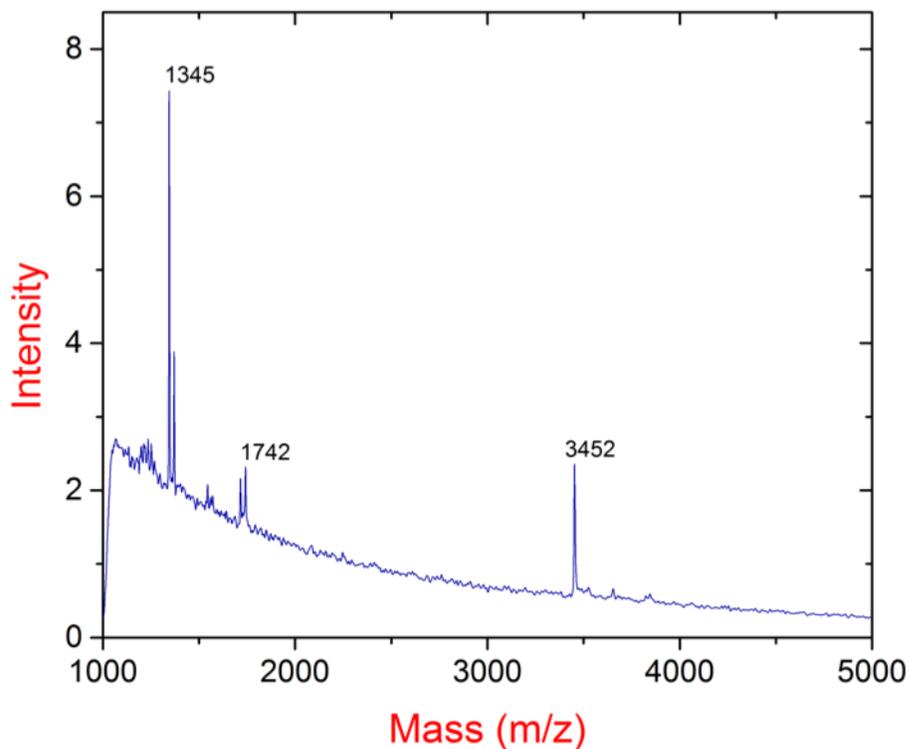

Spectrum 6.6 MALDI-TOF MS results for transcript **ON15** calculated [M] 3237 found 3235.
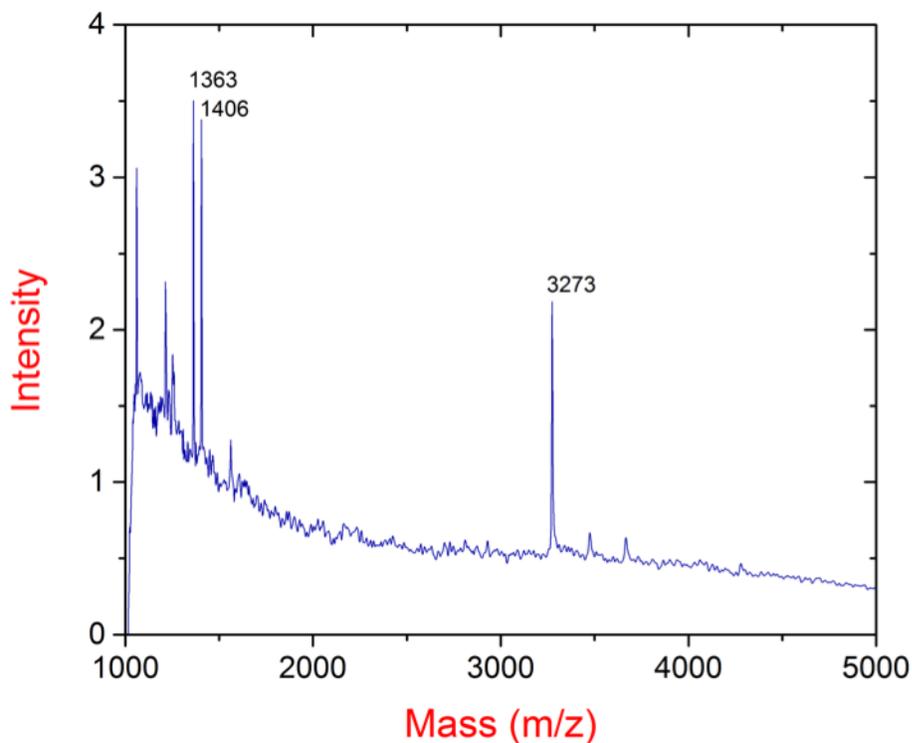5'-r-GCG CCG UGC -3' calculated [M] 2892 found 2890.

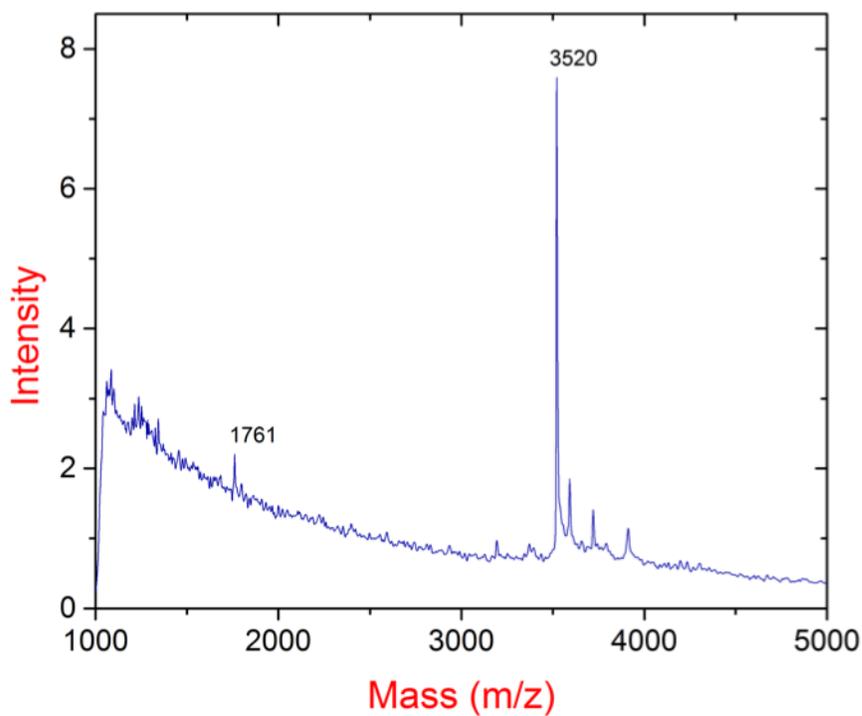Spectrum 6.7 MALDI-TOF MS results for (a) transcript **ON16** calculated [M+2K+Na] 3587 found 3582.

Spectrum 6.8 MALDI-TOF MS results for (a) transcript **ON17** calculated [M] 3303 found 3304. 5'-r-GCG CCG UGC -3' calculated [M] 2974 found 2975.
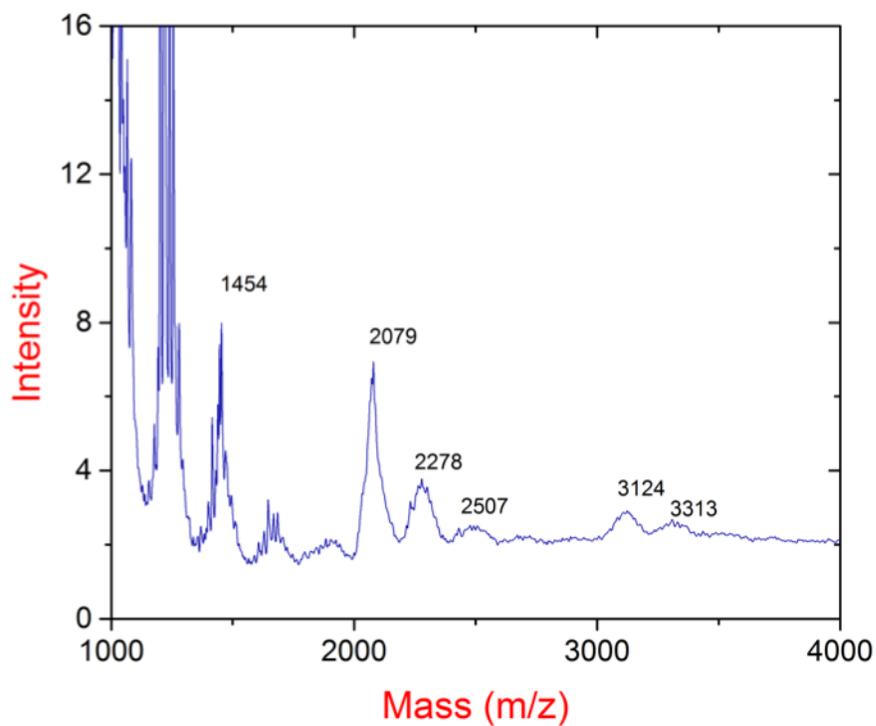
Spectrum 6.9 MALDI-TOF MS results for (a) transcript **ON18** calculated [M] 3453 found 3452.
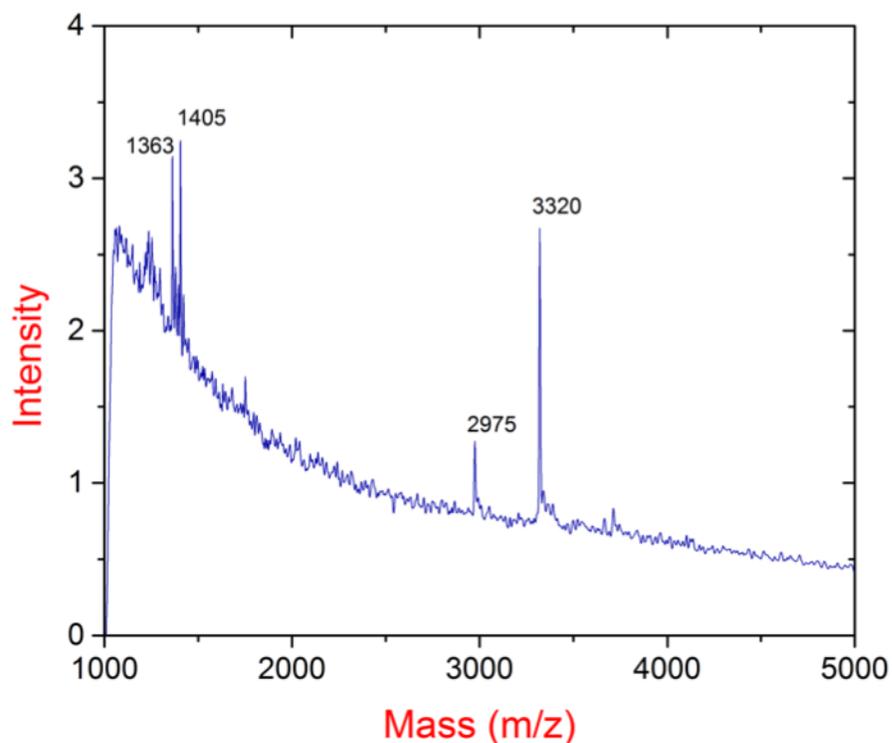

Spectrum 6.10 MALDI-TOF MS results for (a) transcript **ON19** calculated [M+K] 3271 found 3273.

Spectrum 6.11 MALDI-TOF MS results for (a) transcript **ON20** calculated [M] 3519 found 3520.

Spectrum 6.12 MALDI-TOF MS results for (a) transcript **ON21** (9mer) 5'-r-GCG CCG UGC -3' calculated [M+4Na] 3310 found 3313.

Spectrum 6.13 MALDI-TOF MS results for (a) transcript **ON22** calculated [M] 3319 found 3320. 5'-r-<u>G</u>CG CCG UGC -3' calculated [M] 2974 found 2975.

Spectrum 6.14 MALDI-TOF MS results for (a) transcript **ON23** calculated [M] 3537 found 3538.
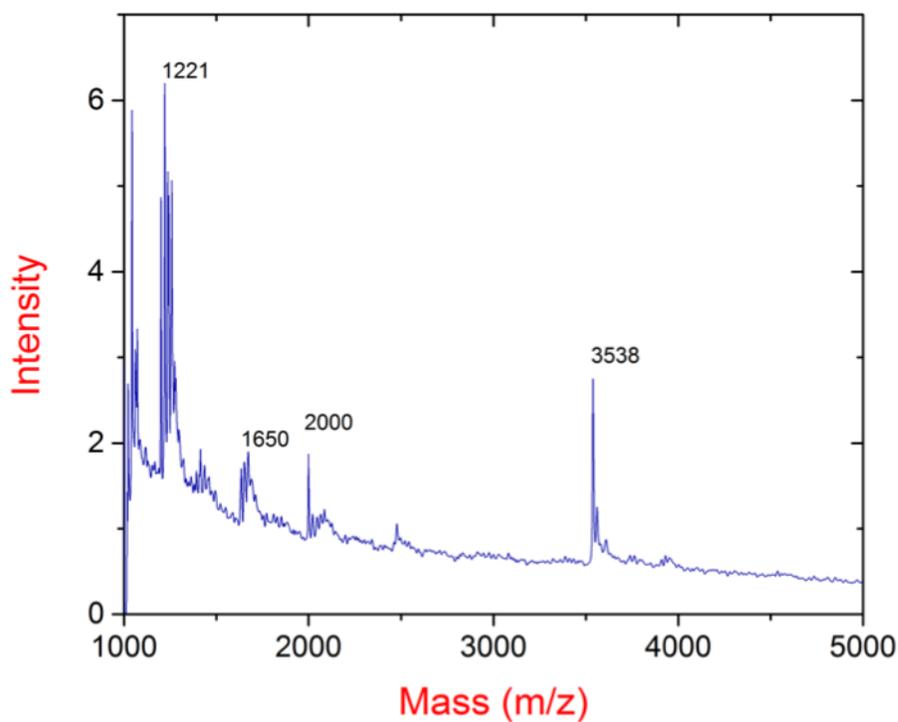
Transcripts **ON26–ON41**

Spectrum 6.15 MALDI-TOF MS results for (a) transcript **ON26** calculated [M] 3590 found 3588.
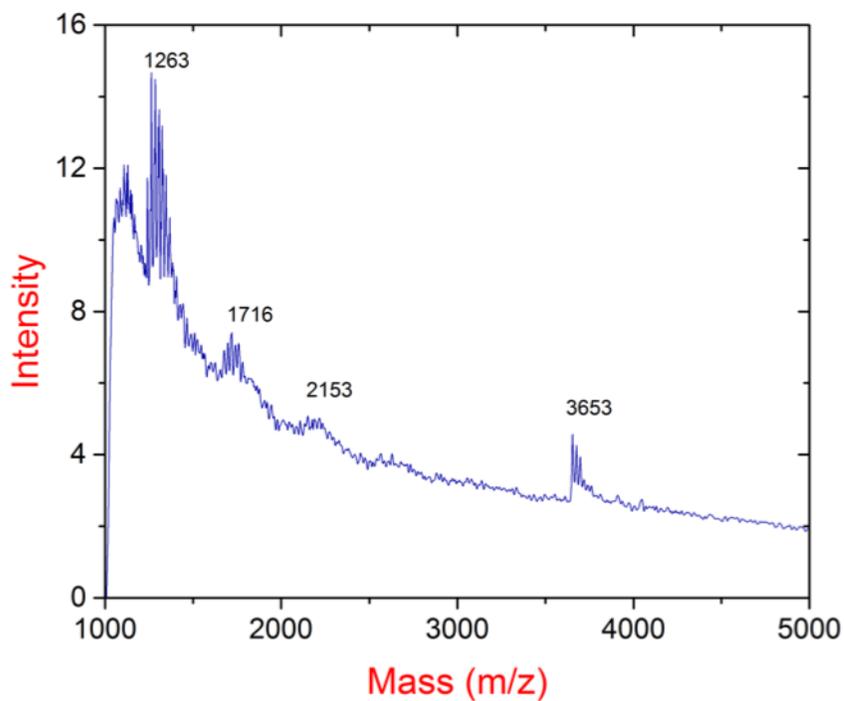
Spectrum 6.16 MALDI-TOF MS results for (a) transcript **ON27** calculated [M] 3654 found 3653.

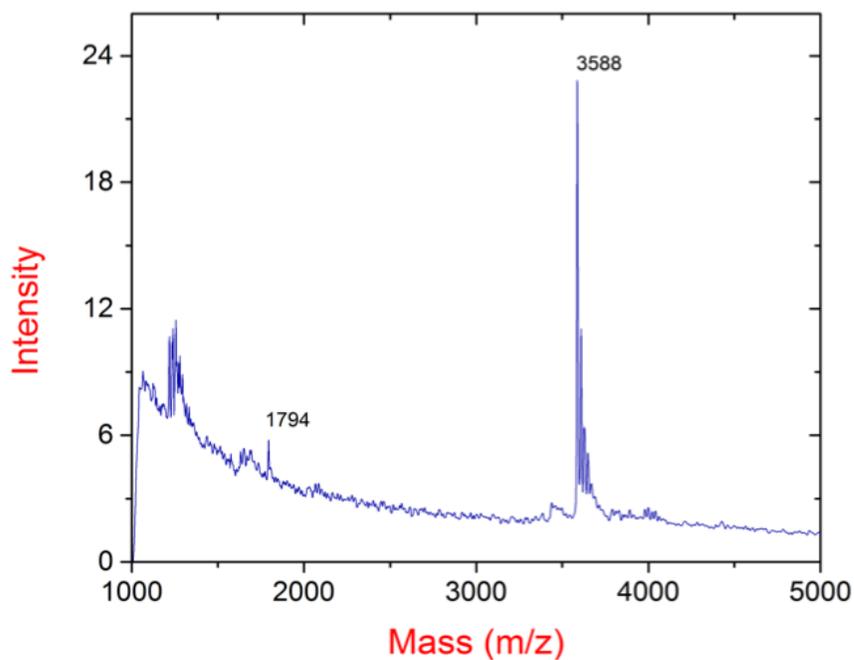Spectrum 6.17 MALDI-TOF MS results for (a) transcript **ON28** calculated [M] 3654 found 3655.

Spectrum 6.18 MALDI-TOF MS results for (a) transcript **ON29** calculated [M] 3656 found 3655.

Spectrum 6.19 MALDI-TOF MS results for (a) transcript **ON30** calculated [M] 3672 found 3669.

Spectrum 6.20 MALDI-TOF MS results for (a) transcript **ON31** calculated [M] 3721 found 3722.
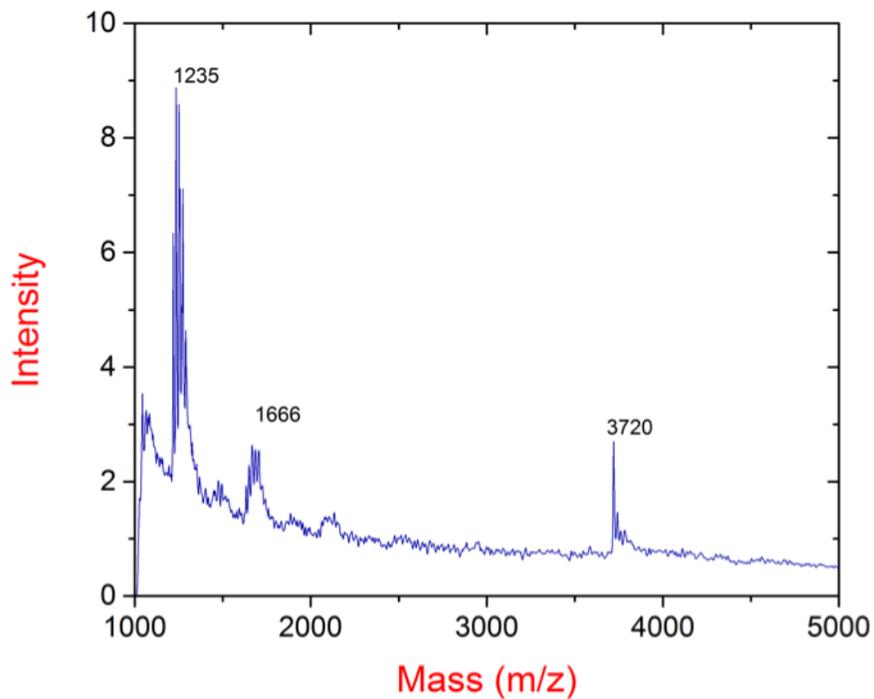
Spectrum 6.21 MALDI-TOF MS results for (a) transcript **ON32** calculated [M] 3720 found 3720.

Spectrum 6.22 MALDI-TOF MS results for (a) transcript **ON33** calculated [M] 3736 found 3737.

Spectrum 6.23 MALDI-TOF MS results for (a) transcript **ON34** calculated [M] 3720 found 3721.
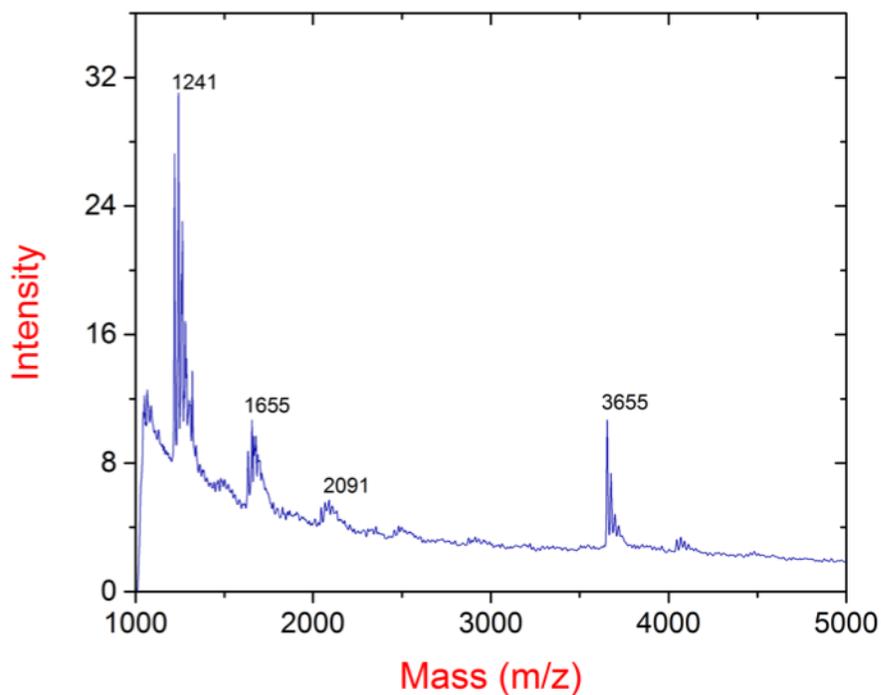

Spectrum 6.24 MALDI-TOF MS results for (a) transcript **ON35** calculated [M] 3736 found 3739.
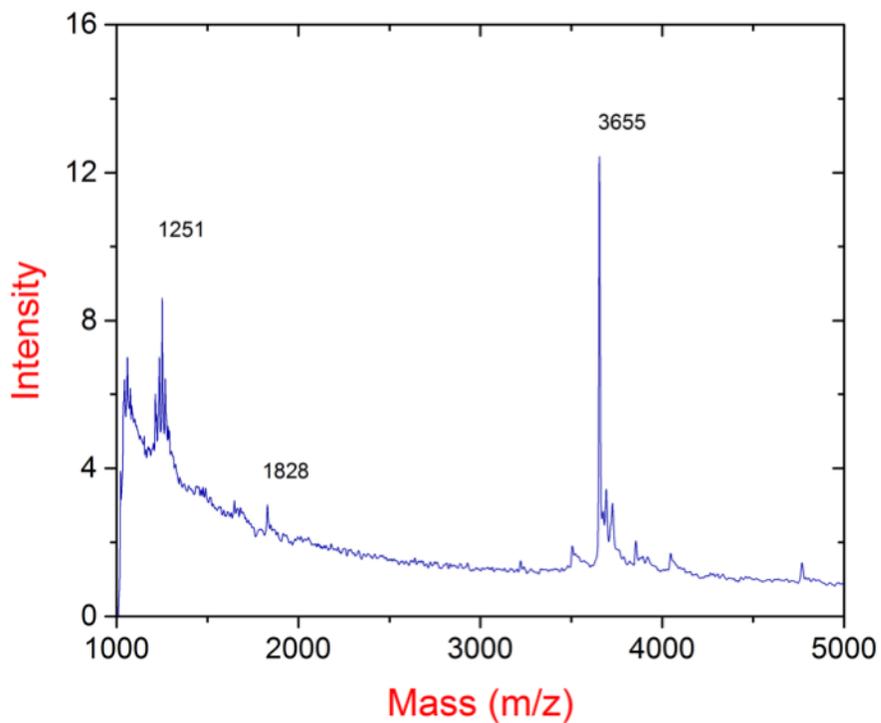
Spectrum 6.25 MALDI-TOF MS results for (a) transcript **ON36** calculated [M] 3738 found 3739.

Spectrum 6.26 MALDI-TOF MS results for (a) transcript **ON37** calculated [M+Na] 3807 found 3805. 5'-r-GGG AGA CGA AUA -3' calculated [M+Na] 4152\ found 4150,

Spectrum 6.27 MALDI-TOF MS results for (a) transcript **ON38** calculated [M] 3803 found 3805.
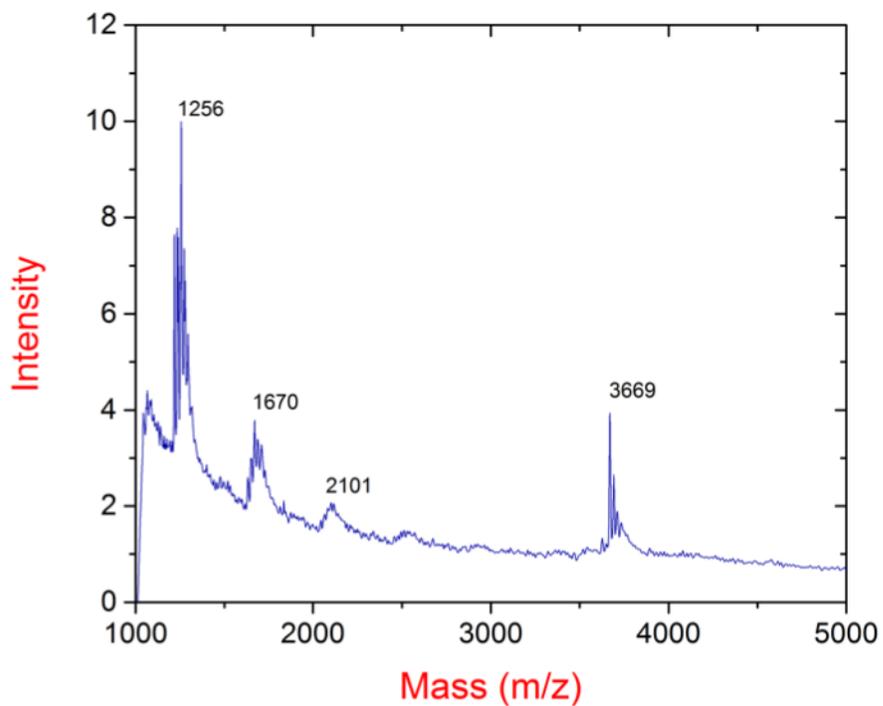

Spectrum 6.28 MALDI-TOF MS results for (a) transcript **ON39** calculated [M] 3805 found 3805.

Spectrum 6.29 MALDI-TOF MS results for (a) transcript **ON40** calculated [M] 3805 found 3806.

Spectrum 6.30 MALDI-TOF MS results for (a) transcript **ON41** calculated [M] 3869 found 3869.
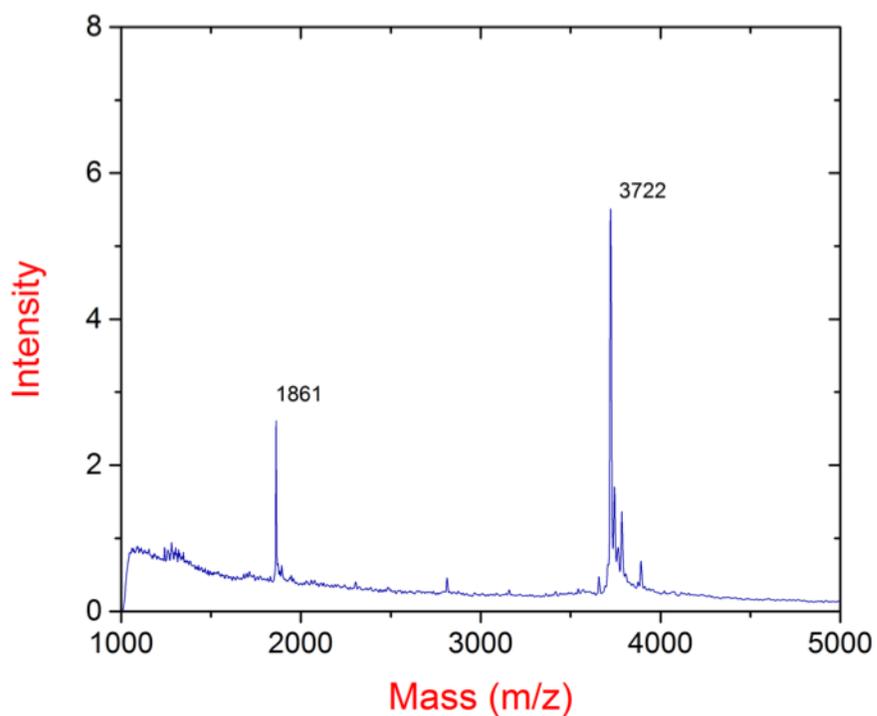
Transcripts **ON70**–**ON76**
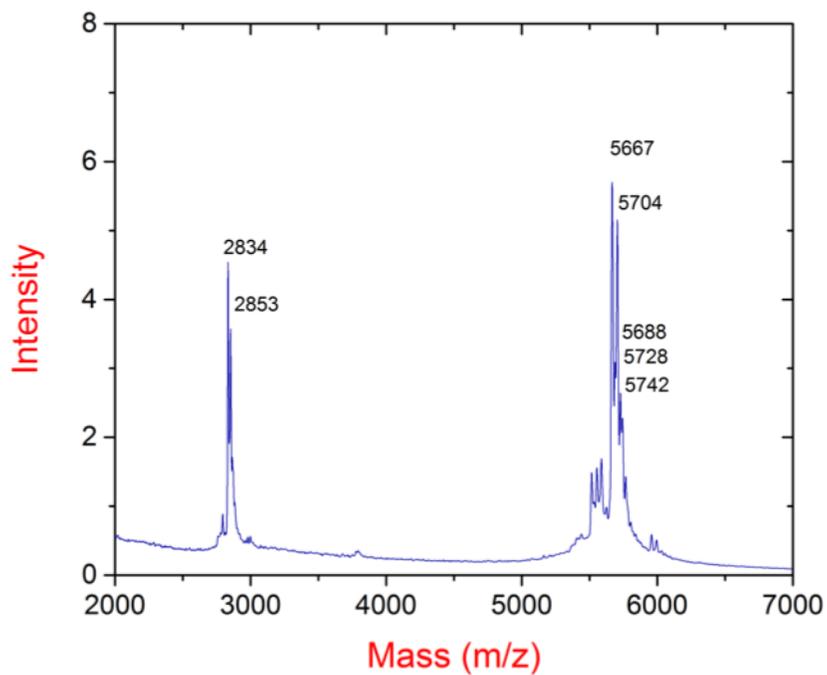
Spectrum 6.31 MALDI-TOF MS results for (a) transcript **ON70** calculated [M] 5664 found 5667.
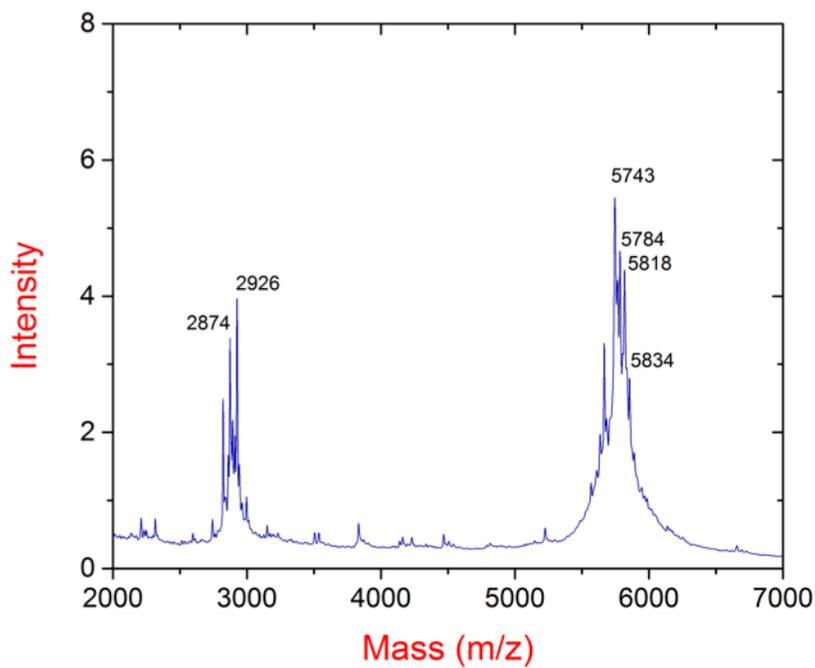
Spectrum 6.32 MALDI-TOF MS results for (a) transcript **ON71** calculated [M] 5743 found 5743.
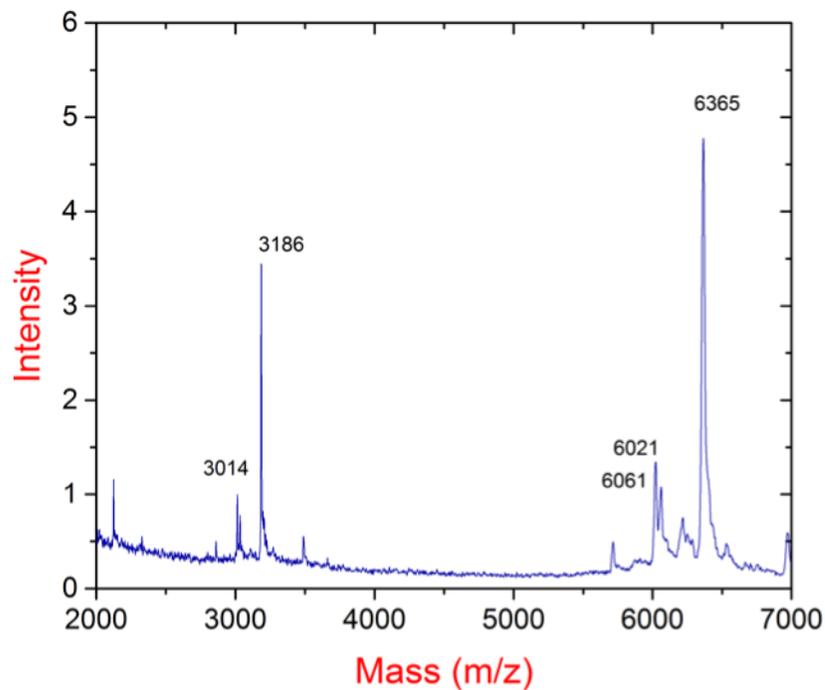
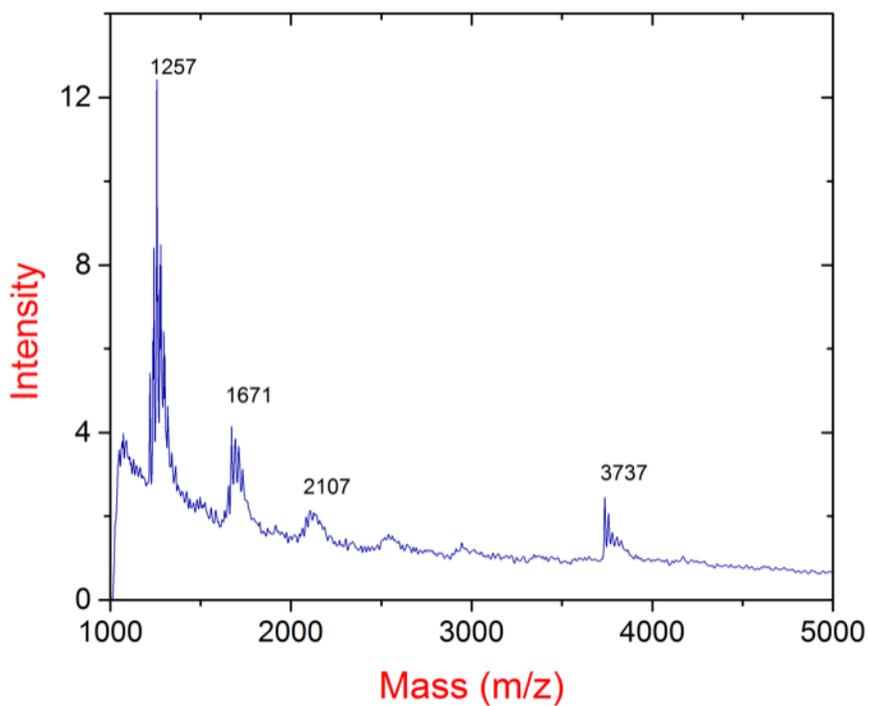Spectrum 6.33 MALDI-TOF MS results for (a) transcript **ON72** (19mer) calculated [M] 6378 found 6365.
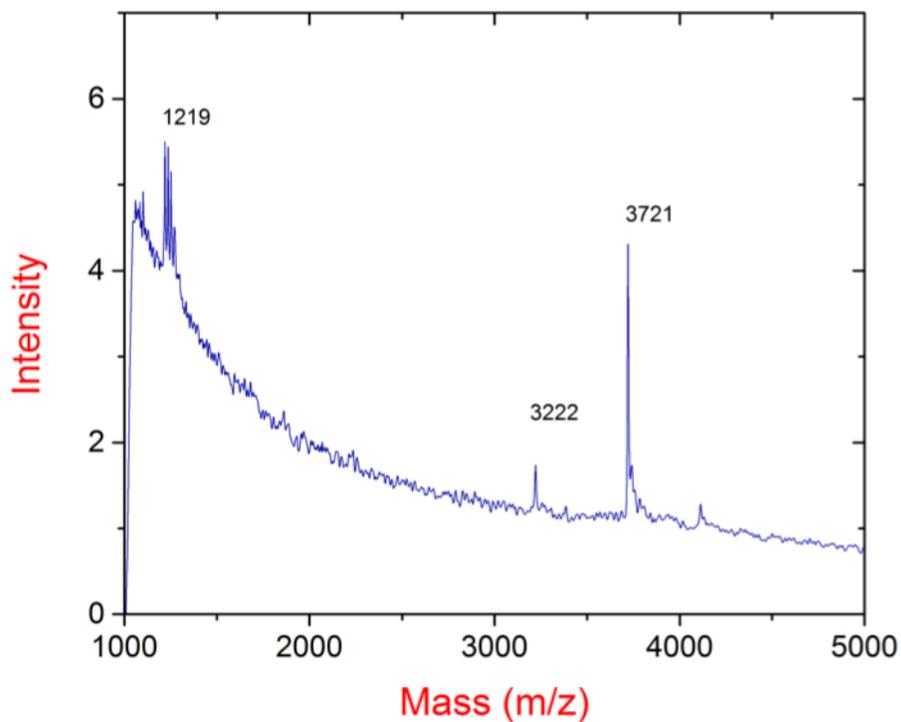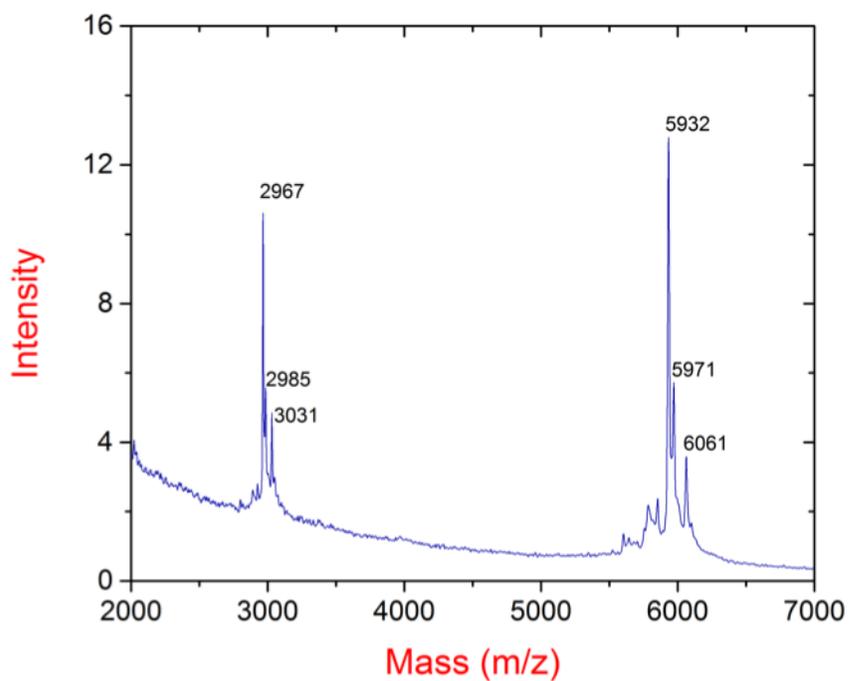

Spectrum 6.34 MALDI-TOF MS results for (a) transcript **ON73** calculated [M] 5928 found 5932.
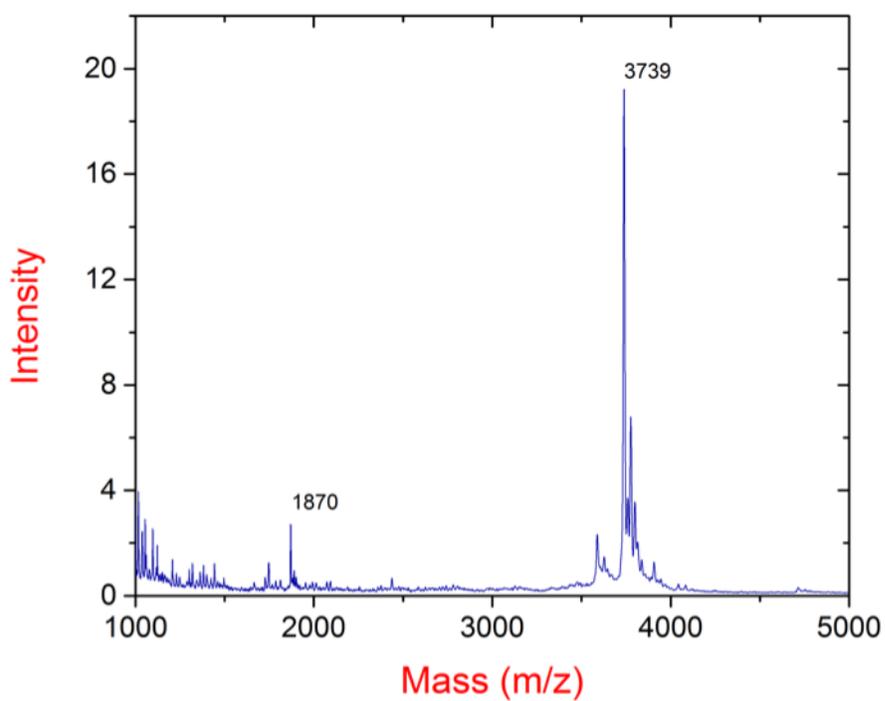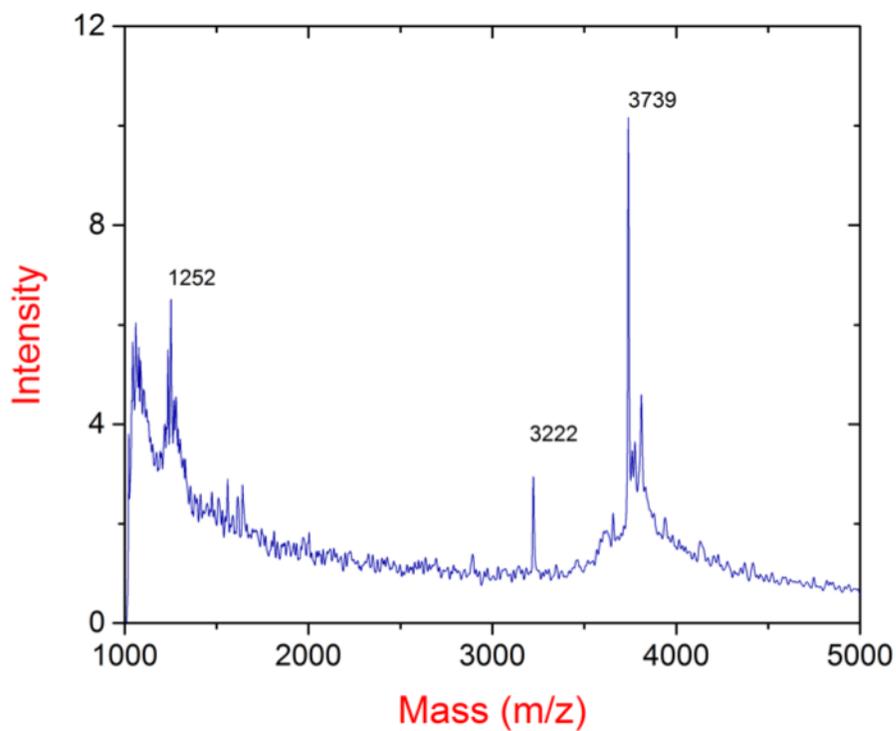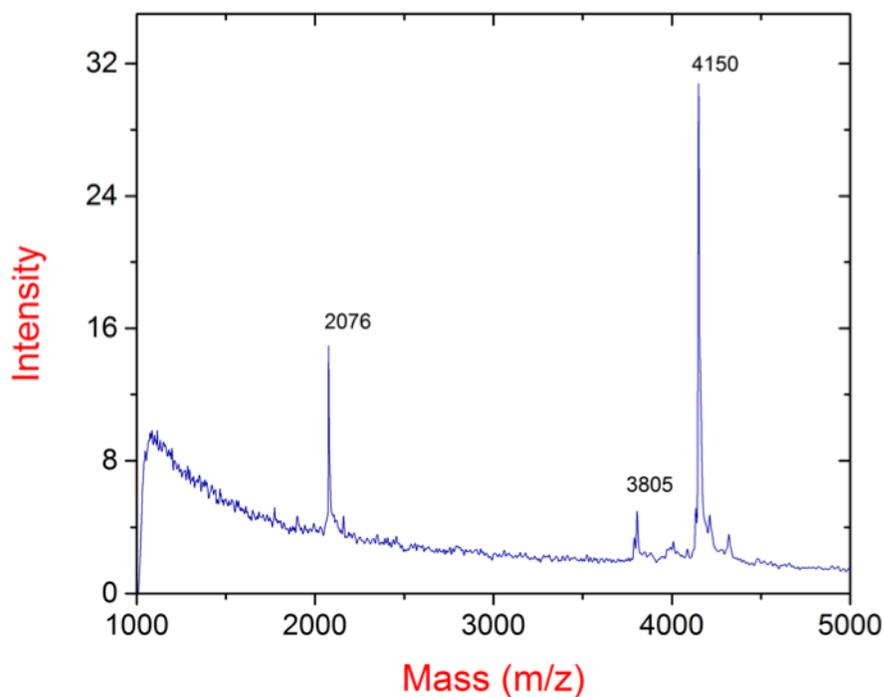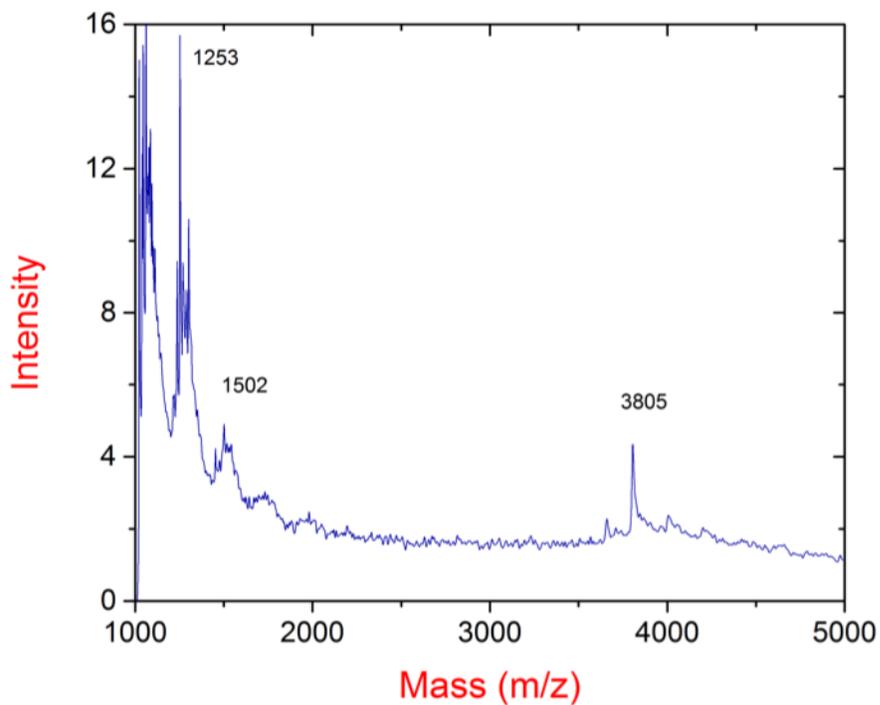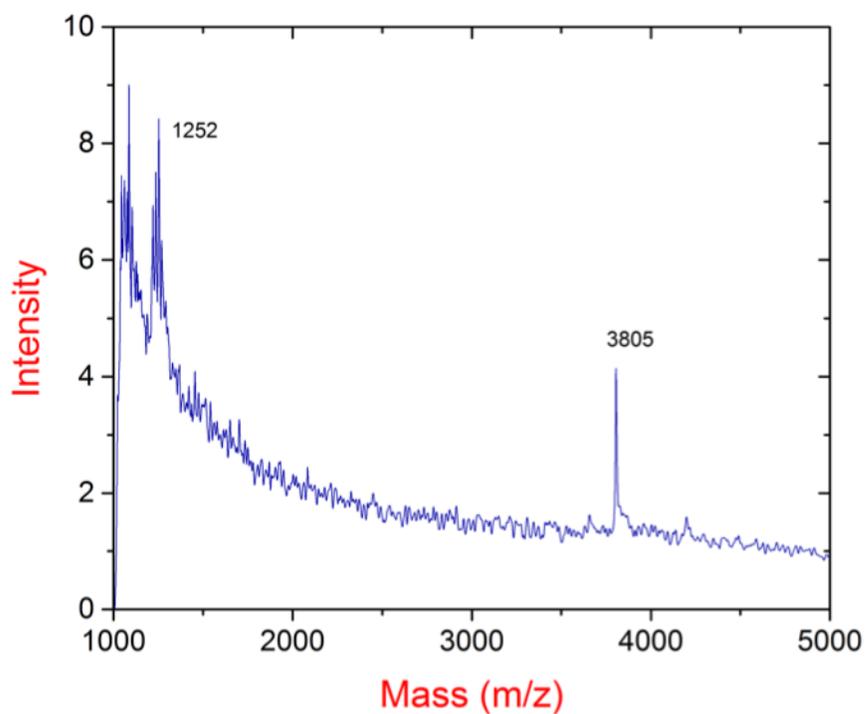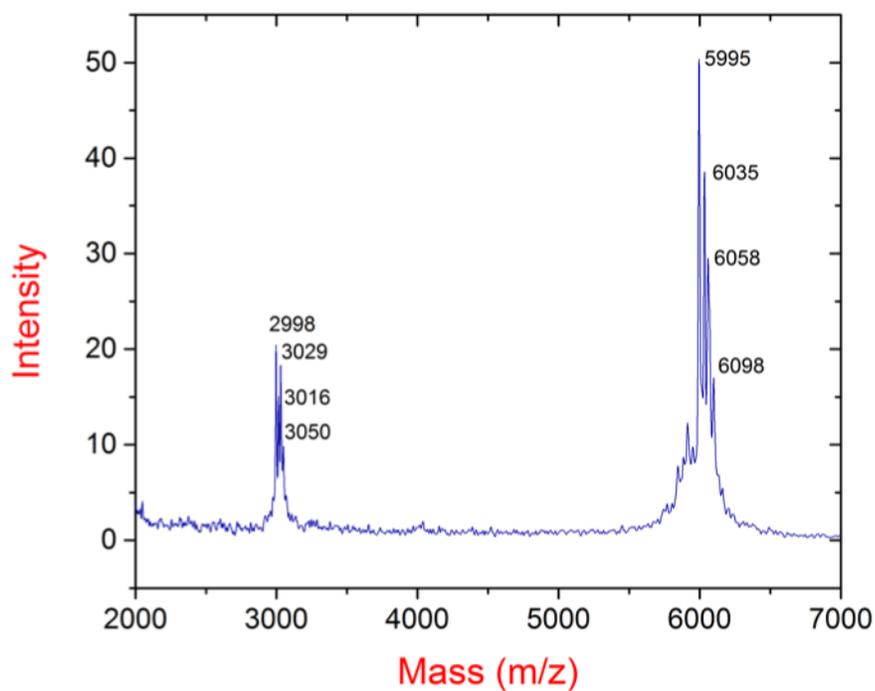
Spectrum 6.35 MALDI-TOF MS results for (a) transcript **ON74** calculated [M] 5992 found 5995.
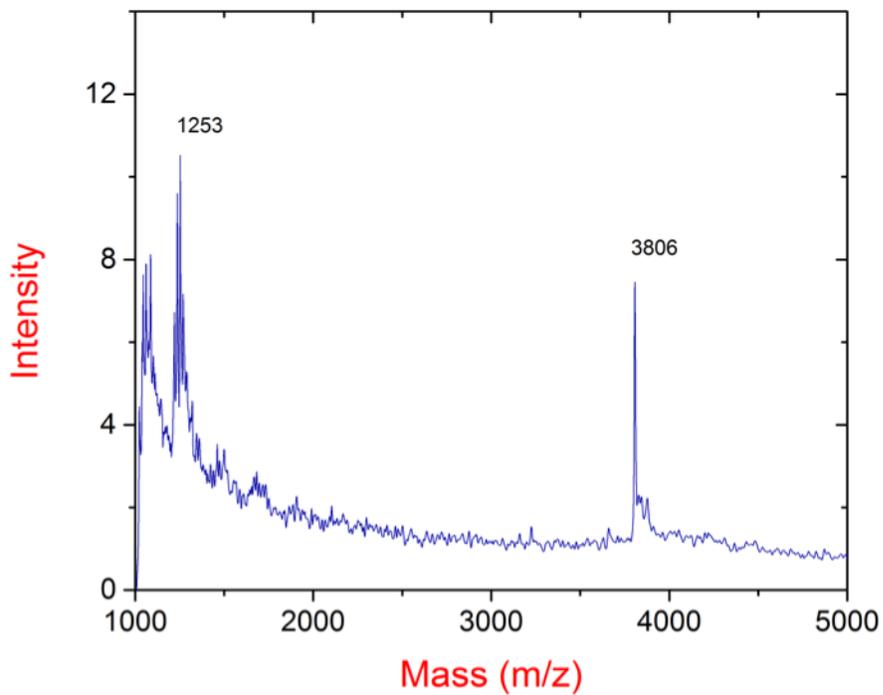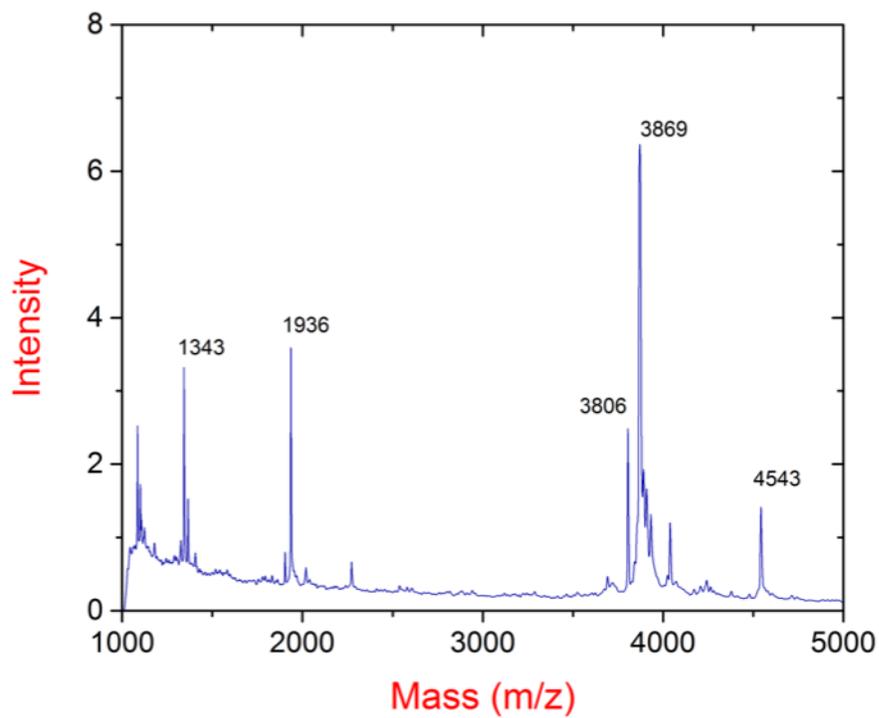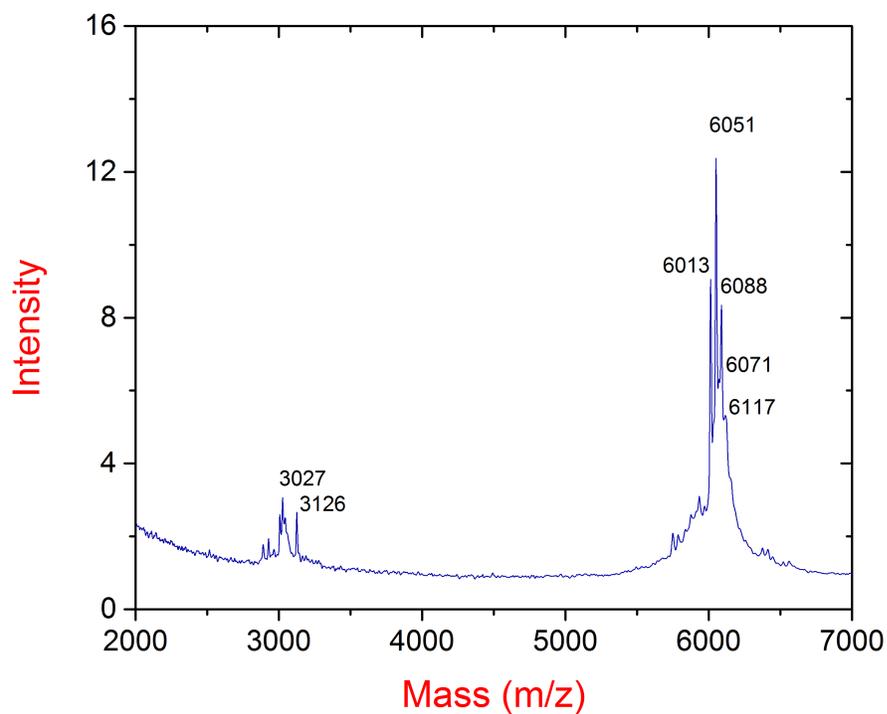

Spectrum 6.36 MALDI-TOF MS results for (a) transcript **ON75** calculated [M] 6011 found 6013 and [M+K] 6050 found 6051.
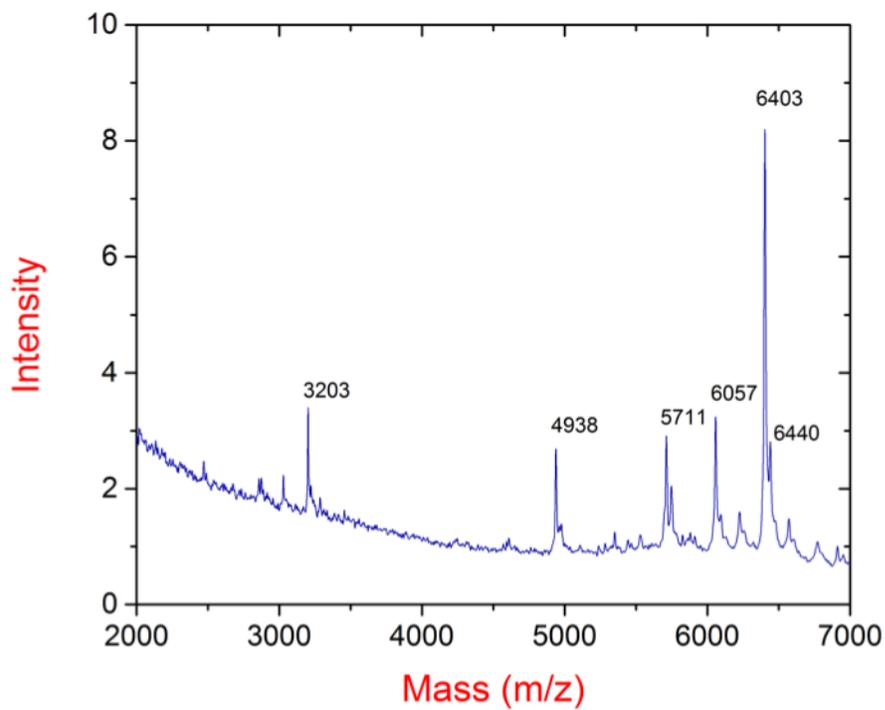
Spectrum 6.37 MALDI-TOF MS results for (a) transcript **ON76** calculated [M] 6403.39 found 6402.87.

Acknowledgements

Chapter 6 is a full reprint from: Hopkins, P. A.; McCoy, L.; Fin, F.; Tor, Y. "Enzymatic synthesis of partially and fully modified RNAs." *-in preparation*. The dissertation author is the primary investigator and author of this article.

6.6 References

1.   R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

2.   L. M. Wilhelmsson, *Quarterly reviews of biophysics*, 2010, **43**, 159-183.

3.   D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378-391.

4.   M. J. Rist and J. P. Marino, *Current Organic Chemistry*, 2002, **6**, 775-793.

5.   N. J. Greco and Y. Tor, *J Am Chem Soc*, 2005, **127**, 10784-10785.

6.   S. G. Srivatsan and Y. Tor, *J Am Chem Soc*, 2007, **129**, 2044-2053.

7.   S. G. Srivatsan, N. J. Greco and Y. Tor, *Angew Chem Int Edit*, 2008, **47**, 6661-6665.

8.   R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chembiochem*, 2008, **9**, 706-709.

9.   S. G. Srivatsan and Y. Tor, *Chem-Asian J*, 2009, **4**, 419-427.

10.   N. J. Greco, R. W. Sinkeldam and Y. Tor, *Org Lett*, 2009, **11**, 1115-1118.

11.   A. R. Rovira, A. Fin and Y. Tor, *J Am Chem Soc*, 2015, **137**, 14602-14605.

12.   D. Shin, R. W. Sinkeldam and Y. Tor, *J Am Chem Soc*, 2011, **133**, 14912-14915.

13.   D. Shin, P. Lonn, S. F. Dowdy and Y. Tor, *Chem Commun*, 2015, **51**, 1662-1665.

14.   L. S. McCoy, D. Shin and Y. Tor, *J Am Chem Soc*, 2014, **136**, 15176-15184.

15. M. R. Green, J. Sambrook and J. Sambrook, *Molecular cloning : a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 4th edn., 2012.

16. W. E. Purtha, R. L. Coppins, M. K. Smalley and S. K. Silverman, *J Am Chem Soc*, 2005, **127**, 13124-13125.

17. F. Wachowius, F. Javadi-Zarnaghi and C. Hobartner, *Angew Chem Int Edit*, 2010, **49**, 8504-8508.

18. F. Wachowius and C. Hobartner, *J Am Chem Soc*, 2011, **133**, 14888-14891.

19. L. Buttner, J. Seikowski, K. Wawrzyniak, A. Ochmann and C. Hobartner, *Bioorgan Med Chem*, 2013, **21**, 6171-6180.

20. A. Ponce-Salvatierra, K. Wawrzyniak-Turek, U. Steuerwald, C. Hobartner and V. Pena, *Nature*, 2016, **529**, 231-U272.

21. S. G. Srivatsan, H. Weizman and Y. Tor, *Org Biomol Chem*, 2008, **6**, 1334-1338.

22. Y. Tor, S. Del Valle, D. Jaramillo, S. G. Srivatsan, A. Rios and H. Weizman, *Tetrahedron*, 2007, **63**, 3608-3614.

23. J. Ludwig, *Acta Biochim Biophys*, 1981, **16**, 131-133.

24. Y. Liang, J. Gloudeman and S. F. Wnuk, *J Org Chem*, 2014, **79**, 4094-4103.

Chapter 7   Conclusions and Future Directions.

It has become apparent to the scientific community that exploring real-time dynamics of RNA and another biomolecule is very important. Carrying out experiments under biologically relevant conditions can ensure meaningful information about the mobility and flexibility of nucleic acids.[1, 2] Due to its sensitivity and time scale, fluorescent spectroscopy plays a predominant role among non-invasive techniques. Consequently, the utility of synthetic fluorescent nucleosides to investigate RNA structure and function as well as the interplay between the two has been rapidly growing for several decades.[2-10]

Among biomolecules and their building blocks, nucleosides and oligonucleotides are practically non emissive.[11, 12] Nevertheless, chemical biology and medicinal chemistry applications require new nucleoside analogues with well-defined photophysical properties. To confer favorable fluorescence features upon the native nucleosides, altering the electronic character of chromophore is essential, thus synthetic modification are often required. The challenge in the field is to develop such synthetic nucleosides while maintaining their high structural resemblance to the native counterparts.

Accordingly, the design of the first generation of isomorphic emissive nucleosides (**29**, **38**–**41**) was based on substituting the uridine core with small heterocyclic moieties at the nucleobase's 5 position (Figure 1.6).[13-18] This simple and small modification generally introduces unique properties upon the new

nucleoside analogues, such as exclusive absorption bands (≥ 300 nm) and red-shifted emission bands beyond 400 nm (Table 1.2). [13, 19-23]

Sensitivity to changes in polarity of nucleoside **38** (Figure 1.6), coupled with the proximity of the furan moiety to the nucleobase and its placement in the major groove, prompted its successful use as a probe for the groove's micropolarity.[24] The advantages of nucleoside **38** were not limited to fluorescent properties but also included enzymatic compatibility. The uridine derivative was converted to the corresponding 5'-triphosphate and subsequently enzymatically incorporated into oligonucleotides by T7 RNA polymerase with an overall efficiency of 78% compared to the natural UTP. This finding initiated further studies, wherein a bacterial A-site labeled with nucleoside **38** in close proximity of the binding site was prepared to monitor aminoglycoside antibiotic binding by monitoring fluorescence changes.[18] A similar design was used to monitor the HIV-1 TAR–Tat interaction.[25]

Although the first generation of fluorescent uridines have proven useful, their emission quantum efficiencies were relatively low (0.01–0.035).[13, 16, 26] This limitation justifies the development of new isomorphic uridine probes. The major obstacle in this field is the inability to predict the emissive features of small chromophores based on their structure. Therefore, designing such analogues remains an empirical exercise. Nonetheless, we wondered what kind of electronic alterations to small fluorophores would let us tune their photophysical properties; thus yielding "ideal" uridine analogues with relatively high quantum yields and emission bands shifted into the red regions of the spectrum.[27, 28]

Consequently, the polarization of the conjugated electron-poor/electron-rich biaryl system of a parent analogue (**46**) was enhanced by introducing an electronegative nitrogen at the pyrimidine's 6 position. This resulted in substantially augmented quantum yields (0.2–0.8, Table 2.1) and red-shifted emission maxima (Figure 3.1, **C**).[27] These findings have prompted the development of a family of visibly emissive extended 6-azauridines (**48**–**53**) by directly conjugating a donor group through an extended aromatic system to the electron deficient 6-azaurdine (Figure 3.1, **D**).[28] Generally, these uridine analogues display high emission quantum yields in apolar solvents and red-shifted emission bands compared to the parent modified uridines (Table 3.1). More importantly, both designs have shown that the photophysical properties of synthetically accessible analogs may be tuned by judiciously introducing substituents of distinct electronic character at a remote position.

While successful alterations of the emissive features of synthetic nucleosides have been shown,[27, 28] it is extremely important for such probes to sustain, at the same time, high bio-compatibility and bio-functionality. Tuning the interplay between these requirements within the design of fluorescent probes may often be a challenging task. To directly test these criteria, we often convert modified nucleosides to the corresponding 5'-triphosphates and subsequently test their enzymatic incorporation into RNA constructs by T7 RNA polymerase.

While previous reports have indicated that enzymatic incorporation of 6-azauridine triphosphate **97** was extremely inefficient,[29-31] we hypothesized that conjugating a thiophene ring to a 6-azauridine at position 5 may partially overcome

the unfavorable effect of the nitrogen substituent at the 6 position (Figure 5.1).[32] We have therefore synthesized the corresponding triphosphate **96** and subjected it to T7-RNA polymerase-mediated transcription reactions. Our results have illustrated that 5-thiophene-6-azauridine triphosphate **96** is a "polymerase-friendly" substrate and its incorporation is twice as efficient as 6-azauridine triphosphate **97** (Figure 5.6). Thus, we have demonstrated our approach's utility in restoring the enzymatic incorporation of 6-aza modified nucleotides. Furthermore, we utilized **46**, either as the sole probe or in conjugation with a FRET acceptor, for monitoring RNA–ligand binding assays. This probe can possibly be used for exploring other potentially therapeutic RNA targets. Particularly, its sensitivity to changes in pH (pK_a=6.7) makes it an attractive probe for studies involving (de)protonation events on RNA, as there are no known isomorphic pyrimidines analogs with pK_a values within this physiological range.[27, 33]

One of the most important developments in our laboratory was the RNA alphabet consisting of purine (thG, thA) and pyrimidine (thC, thU) mimics (Figure 1.5 **30**–**33**).[34] Subsequently, the emissive guanosine triphosphate surrogate (thGTP, **102**) was found to initiate enzymatic *in vitro* transcription reactions, as well as to be incorporated during the elongation phase in short and long oligonucleotides.[35] To investigate the impact of thG on RNA function the Hammerhead ribozyme's (HH ribozyme) phosphodiester bond cleavage catalysis was monitored. The modified substrate was effectively cleaved, demonstrating the ability of thG in replacing G residues while maintaining proper RNA folding. In contrast, the activity of the modified enzyme was severely diminished, suggesting that thG modification

possibly disrupted the catalytic center.[35] This work triggered further studies into probing the mechanism of HH ribozyme's cleavage, where several site-specifically modified enzymes were used to demonstrate the impact of each substitution on the cleavage reaction.[36] Both contributions support the previously proposed mechanism[37] and the importance of N7 in guanosine 10.1 coordination to metal ions. These findings prompted the development of the second generation emissive RNA alphabet where the nitrogen in a position equivalent to the purines' N7 was restored.[38]

Furthermore, each mimic ([th]G, [th]A, [th]U and [th]C) was incorporated into siRNAs to evaluate their cellular activity.[39] In general, all modified siRNA duplexes displayed potent cellular interference activity. However, the modified pyrimidines seemed to be more disruptive compared to purine analogues. The presence of the thiophene ring fused with cytidine (**32**) or uridine (**33**) core introduces additional heterocyclic moiety. Therefore, these modified nucleosides are more sterically demanding than their natural counterparts (Figure 1.5).

These successful applications of emissive RNA alphabet mimics **30**–**33** have inspired our research, where the impact of synthetic modifications on RNA's function was further investigated.[40] Moreover, there was a need to conduct a systematic study exploring the incorporation of multiple mimics (**28**–**31**) into the same RNA construct. We were interested in the limitation of enzymatic incorporation of multiple isomorphic probes and, more importantly, in the ability of modified transcripts to mimic function and folding of the native strands.

After testing several different DNA templates (**ODN2–8**), we obtained a vast variety of partially and fully modified transcripts (**ON10–ON69**, Figure 6.7). It became apparent that more modifications introduced at the same time decreased the transcription efficiency. We also observed that the positioning of the synthetic residues might have a significant impact on T7 RNA polymerase's ability to effectively incorporate modified nucleosides. Additionally, the enzyme is found to be error-prone, especially at the beginning of the initiation phase or just before the dissociation of the full-length product. T7 RNA polymerase was found to frequently pause in a close proximity to the modified mimics, more precisely directly before or after the incorporation.[14]

Countless biochemical processes depend on nucleobase recognition of the complementary strand. Consequently, to advance our knowledge regarding the recognition patterns of modified RNA constructs, we chose to evaluate the ability of partially and fully modified RNA transcripts to be ligated in the 9DB1 deoxyribozyme-mediated reaction (Figure 6.4).[41-45] We realized that the positioning of the emissive mimics may have an equally significant influence, as much as the quantity of modified residues, on the ability to form stable DNA–RNA hybrids in the ligation reactions. In order to test this, we purposefully chose an RNA sequence with a ratio of nucleotides of 5G:4A:4C:4U, to serve as a donor in the ligation. The ligation rate of the doubly modified donor was faster than the two singly modified donors. We speculate that when emissive nucleosides are embedded within oligonucleotides, especially in case of ^{5fu}C and ^{5th}U, the base stacking is more intense compared to the native C or U residues. The presence of

additional heterocycles, such as furan or thiophene, may enhance the stability of the DNA–RNA hybrid due to a larger structural surface.[18, 32, 46] Finally, we concluded that partially and fully synthetic oligonucleotides can moderately mimic the ability of the native strands to form phosphodiester bond.

Furthermore, we speculate that multiple modifications within RNA constructs may allow for highly efficient electronic communication of excitation energy.[47, 48] Moreover, several modifications may enhance the relative brightness of oligonucleotides compare to the singly modified strands.[49 50] This could address one of the limitations in the field, which is commonly reduced fluorescent intensity of emissive nucleosides when incorporated into oligonucleotides. Subsequently, these emissive isomorphic probes can theoretically advance the imaging of the transport and localization of potential therapeutic agents.

Notably, many self-cleaving ribozymes were discovered in late 80's and early 90' such as the Hammerhead,[51] hairpin,[52] HDV,[53, 54] and Varkud satellite (VS)[55] furthermore, their function and mechanism has been thoroughly studied.[35, 37, 56-60] However, the exact function and significance of the more recent and diverse CPEB3,[61] CoTC,[62, 63] and glmS[64] ribozymes is still vague, thus further investigation into their mechanism of action is needed.

A recent NMR solution structure of CPEB3 has been published, however its exact function is unclear.[65] The rate of self-cleaving can be tuned by single nucleotide polymorphism (SNP) at a specific position (36) in the ribozyme and thus the formation of the translated CPEB3 protein can be regulated.[66] It can be envisioned that following diminishing fluorescent signal of [th]G (**30**) at position 36 in

the CPEB3 ribozyme during translation process could allow us to study correlation between the self-cleavage rate of the ribozyme and the formation of the protein. Other two ribozymes (glmS and CoTC) are known to require a specific cofactor for self-cleavage. For example, gluocosamine-6-phosphate (GlcN6P) is thought to be used as a coenzyme in the self-cleaving reaction of the glmS ribozyme[64] and GTP is essential for auto-cleaving of CoTC.[62]

Fluorescent nucleosides, supplementing existing probes, provide a unique opportunity to identify new roles for RNAs that function independently of proteins, such as ribozymes and riboswitches. These may include mechanistic studies of strand cleavage and ligation reactions. Furthermore, emissive probes can be utilized in assays of therapeutic potentials. For example, in RNA-ligand interaction assays, where small molecules by binding to specific RNA sites (such as A-site) can regulate cell function. The careful evaluation of photophysical features of fluorescent probes is often limited in the field, therefore we hope that our research will help to accelerate the development of new discovery assays.

1.   J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer, New York, 3rd edn., 2006.

2.   R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chem Rev*, 2010, **110**, 2579–2619.

3.   M. J. Rist and J. P. Marino, *Current Organic Chemistry*, 2002, **6**, 775-793.

4.   A. Okamoto, Y. Saito and I. Saito, *Journal of Photochemistry and Photobiology C*, 2005, **6**, 108-122.

5.   J. N. Wilson and E. T. Kool, *Org Biomol Chem*, 2006, **4**, 4265-4274.

6.   M. Sameiro and T. Goncalves, *Chem Rev*, 2009, **109**, 190-212.

7.   Y. Tor, *Pure Appl Chem*, 2009, **81**, 263-272.

8.   D. W. Dodd and R. H. E. Hudson, *Mini-Reviews in Organic Chemistry*, 2009, **6**, 378-391.

9.   L. M. Wilhelmsson, *Quarterly reviews of biophysics*, 2010, **43**, 159-183.

10.   A. Fin, A. R. Rovira, P. A. Hopkins and Y. Tor, in *Modified Nucleic Acids*, eds. K. Nakatani and Y. Tor, Springer International Publishing, Cham, 2016, DOI: 10.1007/978-3-319-27111-8_1, pp. 1-26.

11.   C. A. Sprecher and W. C. Johnson, *Biopolymers*, 1977, **16**, 2243-2264.

12.   P. R. Callis, *Annual Review of Physical Chemistry*, 1983, **34**, 329–357.

13.   N. J. Greco and Y. Tor, *J Am Chem Soc*, 2005, **127**, 10784–10785.

14.   S. G. Srivatsan and Y. Tor, *Chem-Asian J*, 2009, **4**, 419-427.

15.   P. Wigerinck, C. Pannecouque, R. Snoeck, P. Claes, E. De Clercq and P. Herdewijn, *J Med Chem*, 1991, **34**, 2383-2389.

16.   N. J. Greco and Y. Tor, *Tetrahedron*, 2007, **63**, 3515-3527.

17.   A. J. Gutierrez, T. J. Terhorst, M. D. Matteucci and B. C. Froehler, *J Am Chem Soc*, 1994, **116**, 5540-5544.

18.   S. G. Srivatsan and Y. Tor, *J Am Chem Soc*, 2007, **129**, 2044-2053.

19.   M. G. Pawar, A. Nuthanakanti and S. G. Srivatsan, *Bioconjugate Chem*, 2013, **24**, 1367-1377.

20.   A. A. Tanpure and S. G. Srivatsan, *Chembiochem*, 2012, **13**, 2392-2399.

21.   T. Kanamori, H. Ohzeki, Y. Masaki, A. Ohkubo, M. Takahashi, K. Tsuda, T. Ito, M. Shirouzu, K. Kuwasako, Y. Muto, M. Sekine and K. Seio, *Chembiochem*, 2015, **16**, 167-176.

22.   M. G. Pawar and S. G. Srivatsan, *Org Lett*, 2011, **13**, 1114-1117.

23.   M. G. Pawar and S. G. Srivatsan, *J Phys Chem B*, 2013, **117**, 14273-14282.

24.   R. W. Sinkeldam, N. J. Greco and Y. Tor, *Chembiochem*, 2008, **9**, 706-709.

25.   S. G. Srivatsan and Y. Tor, *Tetrahedron*, 2007, **63**, 3601-3607.

26. R. W. Sinkeldam, A. J. Wheat, H. Boyaci and Y. Tor, *Chemphyschem*, 2011, **12**, 567-570.

27. R. W. Sinkeldam, P. A. Hopkins and Y. Tor, *Chemphyschem*, 2012, **13**, 3350-3356.

28. P. A. Hopkins, R. W. Sinkeldam and Y. Tor, *Org Lett*, 2014, **16**, 5290-5293.

29. F. M. Kahan and J. Hurwitz, *J Biol Chem*, 1962, **237**, 3778-3785.

30. L. Shapiro and J. T. August, *J Mol Biol*, 1965, **14**, 214-+.

31. S. Asano, Kurashin.Y, Y. Anraku and D. I. Mizuno, *J Biochem-Tokyo*, 1971, **70**, 9-&.

32. P. A. Hopkins, L. S. McCoy and Y. Tor, *submitted*.

33. L. Liu, J. W. Cottrell, L. G. Scott and M. J. Fedor, *Nat Chem Biol*, 2009, **5**, 351-357.

34. D. Shin, R. W. Sinkeldam and Y. Tor, *J Am Chem Soc*, 2011, **133**, 14912-14915.

35. L. S. McCoy, D. Shin and Y. Tor, *J Am Chem Soc*, 2014, **136**, 15176-15184.

36. Y. Li, A. Fin, L. S. McCoy and Y. Tor, *submitted*.

37. M. Martick, T. S. Lee, D. M. York and W. G. Scott, *Chem Biol*, 2008, **15**, 332-342.

38. A. R. Rovira, A. Fin and Y. Tor, *J Am Chem Soc*, 2015, **137**, 14602-14605.

39. D. Shin, P. Lonn, S. F. Dowdy and Y. Tor, *Chem Commun*, 2015, **51**, 1662-1665.

40. P. A. Hopkins, L. S. McCoy, A. Fin and Y. Tor, *in preparation*.

41. W. E. Purtha, R. L. Coppins, M. K. Smalley and S. K. Silverman, *J Am Chem Soc*, 2005, **127**, 13124-13125.

42. F. Wachowius, F. Javadi-Zarnaghi and C. Hobartner, *Angew Chem Int Edit*, 2010, **49**, 8504-8508.

43. F. Wachowius and C. Hobartner, *J Am Chem Soc*, 2011, **133**, 14888-14891.

44. L. Buttner, J. Seikowski, K. Wawrzyniak, A. Ochmann and C. Hobartner, *Bioorgan Med Chem*, 2013, **21**, 6171-6180.

45. A. Ponce-Salvatierra, K. Wawrzyniak-Turek, U. Steuerwald, C. Hobartner and V. Pena, *Nature*, 2016, **529**, 231-U272.

46. S. G. Srivatsan and Y. Tor, *Nat Protoc*, 2007, **2**, 1547-1555.

47. J. N. Wilson, Y. J. Cho, S. Tan, A. Cuppoletti and E. T. Kool, *Chembiochem*, 2008, **9**, 279-285.

48. J. N. Wilson, J. M. Gao and E. T. Kool, *Tetrahedron*, 2007, **63**, 3427-3433.

49. Y. N. Teo, J. N. Wilson and E. T. Kool, *J Am Chem Soc*, 2009, **131**, 3923-3933.

50. S. L. Wang, J. Guo, T. Ono and E. T. Kool, *Angew Chem Int Edit*, 2012, **51**, 7176-7180.

51. A. C. Forster and R. H. Symons, *Cell*, 1987, **49**, 211-220.

52. J. M. Buzayan, W. L. Gerlach and G. Bruening, *Nature*, 1986, **323**, 349-353.

53. L. Sharmeen, M. Y. P. Kuo, G. Dintergottlieb and J. Taylor, *Journal of virology*, 1988, **62**, 2674-2679.

54. H. N. Wu, Y. J. Lin, F. P. Lin, S. Makino, M. F. Chang and M. M. C. Lai, *P Natl Acad Sci USA*, 1989, **86**, 1831-1835.

55. B. J. Saville and R. A. Collins, *Cell*, 1990, **61**, 685-696.

56. A. B. Burgin, C. Gonzalez, J. MatulicAdamic, A. M. Karpeisky, N. Usman, J. A. McSwiggen and L. Beigelman, *Biochemistry-Us*, 1996, **35**, 14090-14097.

57. J. Han and J. M. Burke, *Biochemistry-Us*, 2005, **44**, 7864-7870.

58. S. R. Kirk, N. W. Luedtke and Y. Tor, *Bioorgan Med Chem*, 2001, **9**, 2295-2301.

59. M. Y. P. Kuo, L. Sharmeen, G. Dintergottlieb and J. Taylor, *Journal of virology*, 1988, **62**, 4439-4444.

60. S. Nakano, D. J. Proctor and P. C. Bevilacqua, *Biochemistry-Us*, 2001, **40**, 12022-12038.

61. K. Salehi-Ashtiani, A. Luptak, A. Litovchick and J. W. Szostak, *Science*, 2006, **313**, 1788-1792.

62. A. Teixeira, A. Tahiri-Alaoui, S. West, B. Thomas, A. Ramadass, I. Martianov, M. Dye, W. James, N. J. Proudfoot and A. Akoulitchev, *Nature*, 2004, **432**, 526-530.

63. S. West, N. Gromak and N. J. Proudfoot, *Nature*, 2004, **432**, 522-525.

64. W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins and R. R. Breaker, *Nature*, 2004, **428**, 281-286.

65. M. Skilandat, M. Rowinska-Zyrek and R. K. O. Sigel, *J Biol Inorg Chem*, 2014, **19**, 903-912.

66. C. Vogler, K. Spalek, A. Aerni, P. Demougin, A. Muller, K. D. Huynh, A. Papassotiropoulos and D. J. F. de Quervain, *Front Behav Neurosci*, 2009, **3**.