Functionalized Nucleoside Analogs as Probes for Nucleic Acid Recognition and Synthons for Supramolecular Assemblies

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

By

Noothanaganti Ashok

ID: 20123162

Indian Institute of Science Education and Research, Pune

IISER PUNE

Research Supervisor

Dr. Seergazhi G. Srivatsan

Associate Professor

Indian Institute of Science Education and Research, Pune

This Dissertation is Dedicated to My Parents

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER), PUNE
(An Autonomous Institution, Ministry of Human Resource Development, Govt. of India)
Dr. Homi Bhabha Road, Pashan, Pune–411 008

Dr. Seergazhi G. Srivatsan
Associate Professor, Chemistry

Dr. Seergazhi G. Srivatsan
Associate Professor
Department of Chemistry
IISER, Pune

CERTIFICATE

Certified that the work incorporated in the thesis entitled *"**Functionalized Nucleoside Analogs as Probes for Nucleic Acid Recognition and Synthons for Supramolecular Assemblies**"* submitted by Mr. Noothanaganti Ashok was carried out by the candidate under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or Institution.

S.G.Srivatsan

Date: 15/01/2018
Place: Pune

Dr. Seergazhi G. Srivatsan

Declaration

I hereby declare that the thesis entitled **"Functionalized Nucleoside Analogs as Probes for Nucleic Acid Recognition and Synthons for Supramolecular Assemblies"** submitted for Doctor of Philosophy in chemistry at Indian Institute of Science Education and Research (IISER), Pune, has not been submitted by me to any other university or institution. I declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or falsified any idea/data/fact/source in my submission. This work presented here was carried out at the, Indian Institute of Science Education and Research, Pune, India under the supervision of Dr. Seergazhi G. Srivatsan.

Date: 15/01/2018

Place: Pune

Noothanaganti Ashok
ID-20123162
Senior Research Fellow
Department of Chemistry
IISER, Pune-411008
India

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my thesis advisor, Dr. S. G. Srivatsan for his constant encouragement, support, and guidance throughout my doctoral research. His expertise in nucleic acids chemistry improved my research skills and prepared me for future challenges. During my tenure, he has given me intellectual freedom in my work, supporting my attendance at various conferences, engaging me in new ideas, and demanding a high quality of work in all my endeavors. His diligent effort and training was the reason behind every success achieved by me and undoubtedly is the asset for my future research. This thesis would not have been possible without his valuable support. I acknowledge Indian Institute of Science Education and Research (IISER), Pune for providing excellent research facilities and an outstanding research environment.

I thank our current and former Directors, Prof. J. B. Udgaonkar and Prof. K. N. Ganesh for giving excellent research platform, financial support and facilities that I have been fortunate with during my research here Indian Institute of Science Education and Research, Pune. I am grateful to the Research Advisory Committee (RAC) members, chair chemistry (IISER Pune) Prof. M. Jayakannan and Dr. M. V. Badiger (NCL Pune) for their suggestions and comments during RAC meetings. The critical examination of my research work and valuable comments by all of the RAC members were always very useful. Because of their guidance and suggestions I gained experience to work in diverse fields of research which was indeed very helpful. I am thankful to all the chemistry faculties in IISER Pune for their support. I am grateful to our collaborators, Dr. K. Saikrishnan (IISER Pune) and Prof. Thomas Hermann (UCSD) for giving me the opportunity to work with them in different collaborative research projects.

I thank IISER Pune librarians, IT staff and administrative staff especially Mayuresh, Tushar, Praveen and Yathish for their kind support. I would like to thank to DST-SERB and Infosys foundation for financial support to attend an international conference. It's my pleasure to thank all the members of the team, I belong to; Dr. Maroti , Dr. Arun, Dr. Anupam, Dr. Cornelia, Dr. Pramod, Dr. Vyankat, Sudeshna, Anurag, Jerrin, Manisha, Pankaj, Saddam, Uddhav and Sangamesh as they always maintained a very lively environment in lab. An all time running discussion and sharing of experiences helped me to expand the horizon of knowledge. Special thanks are due to Ms. Sudeshna, Ms. Shalini, Dr. koti, Dr. Hari, Dr. Narshimma, Dr. Rajandra, Dr. Rajesh, Dr. Rajkumar and Mr. Dinesh for *the many insightful discussions and critical assessment of my work and during my thesis writing.*

Special thank all my friends, seniors and juniors for their help and support during my research tenure. I can not find any word to express my gratitude to my parents. My parents are the key players who boost me in each and every step of life with their kind guidance, timely encouragement love and affection. I am able to reach this position. At last but not least I want to express my gratitude to my sisters Ms. Jyothi and Ms. Sri Likitha who supports me every time to do something constructive in life and encourages to dedicate myself increase a curiosity in my research. I dedicate this thesis to my parents who endless supported me during my years of study.

Noothanaganti Ashok

Chapter 2 is a reprint of: Nuthanakanti, A.; Boerneke, M. A.; Hermann, T.; Srivatsan, S. G. Structure of the ribosomal RNA decoding site containing a selenium-modified responsive fluorescent ribonucleoside Probe. *Angew. Chem., Int. Ed.* **2017**, *56*, 2640–2644.

The dissertation author is the main author and researcher for this work.

Chapter 4A is a reprint of: Nuthanakanti, A.; Srivatsan, S. G. Surface-tuned and metal-ion responsive supramolecular gels based on thymidine nucleolipids. *ACS Appl. Mater. Interfaces* **2017**, *9*, 22864–22874.

The dissertation author is the main author and researcher for this work.

Chapter 5 is a reprint of: Nuthanakanti, A.; Srivatsan, S. G. Hierarchical self-assembly of switchable nucleolipid supramolecular gels based on environmentally-sensitive fluorescent nucleoside analogs. *Nanoscale* **2016**, *8*, 3607–3619.

The dissertation author is the main author and researcher for this work.

Table of Contents

Chapter 2: Probing bacterial ribosomal decoding site RNA structure and antibiotic binding using selenophene-modified dual-purpose nucleoside probe.

Chapter 3: Probing G-quadruplex structures and ligand binding using selenophene-modified 2′-deoxyuridine.

Chapter 4A: Surface-tuned and metal ion-responsive supramolecular gels based on thymidine nucleolipids.

Chapter 4B: Microscopic assemblies of nucleolipid supramolecular synthons show self-sorting and cooperative self-assembling process

Chapter 5: Switchable nucleolipid supramolecular gels based on environmentally-sensitive fluorescent nucleoside analogs.

Abbreviations

ε	molar extinction coefficient
μL	microliter
μM	micro molar
2-AP	2-aminopurine
A	adenosine
ACN	acetonitrile
AMP	adenosine monophosphate
ANL	aminoacid-nucleotide-lipid
ATP	adenosine triphosphate
BrU	5-bromouridine
C	cytosine
CD	circular dichroism
CDP	cytidine diphosphate diacylglycerol
CGC	critical gelation concentration
CPL	circular polarized fluorescence
CW-EPR	continuous wave electron paramagnetic resonance
3D	three dimensional
DMAP	4-dimethylaminopyridine
DMSO	N, N-dimethyl sulfoxide
DMF	dimethylformamide
DMT	dimethoxytrityl
DNA	deoxyribonucleic acid
2-DOS	2-deoxystreptamine
DQC	double-quantum coherence
ds	double stranded
EDC	N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide
EDAX	energy dispersive X-ray spectroscopy

em	emission
EPR	electron paramagnetic resonance
FESEM	field emission scanning electron microscopy
FRET	förster resonance energy transfer
G	guanosine
GB	guanosine-borate
GNBA	glycosyl-nucleoside bola-amphiphiles
GQ	G-quadruplex
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPA	hydroxylpiccolinic acid
HPLC	high performance liquid chromatography
H-Telo	human telomeric
in vitro	outside living organism
in vivo	inside living organism
ITC	isothermal titration calorimetry
IU	5-iodouridine
K_d	dissociation constant
LNA	locked nucleic acid
LMWG	low molecular weight gel
MAD	multiwavelength anomalous diffraction
MALDI-TOF	matrix assisted laser desorption ionisation-time of flight
max	maximum
MB	methylene blue
MeOH	methanol
mg	milligram
MHz	megahertz
6-MI	6-methyl isoxanthopterin
MIR	multiple isomorphous replacement
MIRAS	multiple isomorphous replacement with anomalous scattering

µL	microliter
mM	milimolar
µM	micrometer
MOE	methoxyethyl
MR	molecular replacement
NAIM	nucleotide analog interference mapping
nm	nanometer
nmol	nanomolar
NMR	nuclear magnetic resonance
ON	oligonucleotide
Pa	pascal
PAGE	polyacrylamide gel electrophoresis
PAH	polycyclic aromatic hydrocarbon
PDB	protein data bank
PEG400	polyethylene glycol400
PELDOR	pulsed electron double resonance
Pd	palladium
PDS	pyridostatin
PNA	peptide nucleic acids
ppm	parts per million
PXRD	powder X-ray diffraction
RB	rose bengal
Rf	retention factor
RIPs	ribosome inactivating proteins
RMSD	root-mean-square deviation
RNA	ribonucleic acid
RNase H	ribonuclease H
SAD	single-wavelength anomalous dispersion
SIR	single isomorphous replacement

SIRAS	single isomorphous replacement with anomalous scattering
SNP	single nucleotide polymorphism
T	thymine
TBAF	tetrabutylammonium fluoride
TBDMS	tert-butyldimethylsilyl
TCA	trichloroacetic acid
TCSPC	time correlated single photon counting
TEAA	triethylammonium acetate
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TERRA	telomeric repeat-containing RNA
THF	tetrahydrofuron
ThT	thioflavin T
TLC	thin layer chromatography
Tm	thermal melting
TOM	triisopropylsilyloxymethy
Tris	tris (hydroxymethyl) amino methane
U	uridine
UV	ultraviolet
WC	Watson-Crick

Synopsis

Functionalized nucleoside analogs as probes for nucleic acid recognition and synthons for supramolecular assemblies

Background and Aim:

Nucleic acids, DNA and RNA, as we know perform their cellular functions by adopting various complex structures and by undergoing conformation changes upon protein or small molecule metabolite binding.[1] Several biophysical tools based on fluorescence, NMR, EPR, X-ray techniques, circular dichroism (CD), to name a few, have been developed to understand the structure and ensuing functions of nucleic acids.[2] Invariably, these investigations greatly rely on custom-labeled nucleic acid sequences as the components of nucleic acids do not contain intrinsic labels for effective analysis by fluorescence, NMR, X-ray, etc., techniques. In this context, modified nucleosides, which retain the Watson-Crick base paring and metal ion binding properties, have served as very useful biophysical tools in nucleic acid analysis, diagnosis and in setting up screening platforms for identifying small molecule ligands that bind and modulate the function of nucleic acids.[3]

Functionalized nucleoside analogs are usually prepared by attaching reporters or labels on to the nucleobase, sugar or phosphate backbone. The labeled oligonucleotides (ONs) are then prepared by incorporating the phosphoramidite or triphosphate substrate of the modified nucleoside using solid-phase ON synthesis or enzymatic methods. For example, fluorescent nucleoside analogs are typically prepared by attaching known fluorophores to purine and pyrimidine rings or sugar residue. Alternatively, direct conjugation of heterocycles onto nucleobases can be used to develop environment-sensitive fluorescent nucleoside analogs.[4] Similarly, isotope- (^{13}C, ^{15}N and ^{19}F) and spin-labeled (nitroxide radicals) nucleoside analogs incorporated into ONs serve as excellent tools for nucleic acid conformational analysis by NMR and EPR spectroscopy techniques.[5] ONs labeled with halogenated (Br or I) nucleosides are usually used in X-ray crystallography. More recently, the usefulness of Se atom as a very good anomalous scattering label has been also applied in nucleic acid structural analysis.[6] While these tools individually have provided valuable information on nucleic acids, correlating the information obtained from different techniques, working under different assay conditions, is difficult and prone to obscure correlation. This we strongly believe is because the scientific community has greatly relied on the traditional one-label one-technique approach due to paucity of tools that can be efficiently implemented in more than one biophysical technique.

The ability to program nucleic acid structure by using their well defined recognition properties has rendered nucleic acid as a very useful supramolecular synthon for the construction of nano-architectures.[7] Many such architectures have been used as nano machines,[8] molecular containers for transporting drugs and in diagnosis.[9] However, challenges in large-scale fabrication and scalability are the major concerns of this field. Alternatively, nucleobases, nucleosides, and their derivatives, serve as very useful supramolecular synthons for the construction of various architectures.[10] The advantages are that along with their inherent recognition, stacking and metal ion binding properties, the nucleosides can be easily modified to install additional functionalities by using established synthetic approaches. Self-assemblies like vesicles, fibres, gels and synthetic ion channels have been constructed by using tailor-made nucleobase and nucleoside derivatives.[11] Many such assemblies have been used as gene and drug delivery vehicles.[12] However, the potential of nucleoside-based synthons in designing smart self-assembled materials, which are tunable and can show interesting optical and sensing properties, is not well explored.

It is apparent from these studies that when carefully modified nucleoside analogs can be used as probes for studying nucleic acid structure and function or as therapeutic tool or can be used to build self-assembling systems (Figure 1). Limitations of traditional nucleic acid probe design strategy and the potential of nucleoside supramolecular synthons in materials applications inspired us to design new types of probes and materials, which would complement existing tools and resources to (i) study nucleic acids effectively and (ii) construct architectures with tunable and sensing properties.

The thesis is divided into two parts. The first part illustrates an innovative approach to investigate the nucleic acid structure and function in real-time in solution and in solid-state by using dual-purpose nucleoside probes. The dual-purpose nucleoside analogs, containing a fluorophore and X-ray anomalous scattering label (Se atom), are derived by attaching selenophene at the 5-position of 2′-deoxyuridine (^{Se}dU) and uridine (^{Se}U).[13] The nucleoside analogs incorporated into DNA and RNA ONs are minimally perturbing and were found to be sensitive to conformational changes. These features of the nucleoside probe enabled a comprehensive understanding of the structure and recognition properties of the bacterial ribosomal decoding site RNA and G-quadruplex-forming human telomeric repeats under equilibrium condition by fluorescence, and in 3D by X-ray crystallography.[14,15] The selenophene-modified nucleoside analogs represent the first example of a dual purpose probe, which provides this level of understanding.

The second part of the thesis illustrates a novel design strategy to construct supramolecular nucleolipid synthons by using environment-sensitivity fluorescent nucleoside analogs, based on 5-(benzofuran-2-yl)uracil and 5-(benzo[*b*]thiophen-2-yl)uracil cores, as the head group and fatty acids, attached to the ribose sugar, as the lipophilic group. Alternatively, nucleolipid synthons are simply prepared by attaching fatty acids to the sugar hydroxyl groups of native nucleosides. Depending on the modification on the nucleobase and sugar residue, the nucleolipids form gels, which are highly responsive to external stimuli (physical and chemical) and show aggregation-induced enhanced emission, or form materials whose surface can be switched from highly hydrophobic to hydrophilic and vice versa. We expect that this simple approach of probe and supramolecular synthon design could lead to the emergence of a new family of smart materials and probes.

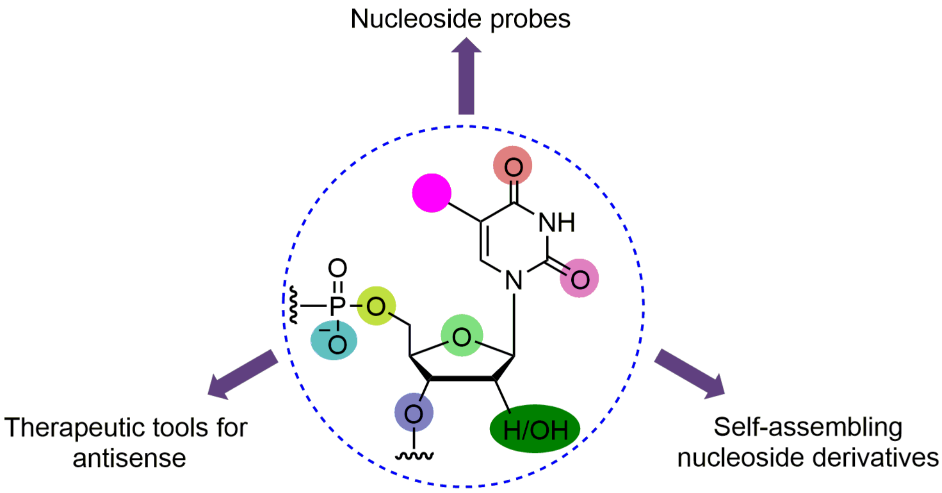


Figure 1. Schematic representation sites for chemical modifications of uridine nucleotide. Selectively derivatized nucleic acid components (nucleobase, nucleoside and nucleotide) act as functional nucleoside probes, self-assembling building blocks and therapeutic tools for antisense applications. Coloured labels indicate possible modification sites.

The thesis is organized as follows:

Chapter 1: Nucleic acid probes and supramolecular assemblies based on functionalized nucleosides

A brief overview of various nucleic acid structures including non-canonical structures and their cellular functions is provided. A detailed discussion on various biophysical tools that have been developed to investigate the structure and function of nucleic acid is presented.

Particular emphasis is laid on the recent developments in the design and applications of base-modified nucleoside analogs. Further, general approaches to design nucleoside-based supramolecular synthons, and their self-assembling behaviour and biomedical applications are discussed. The limitations of currently available nucleoside probes and the envisioned potential of nucleoside supramolecular synthons in materials application, which motivated the present work, are described in the research statement section.

Chapter 2: Probing bacterial ribosomal decoding site RNA structure and antibiotic binding using selenophene-modified dual-purpose nucleoside probe

The utility of selenophene-modified uridine analog (^{Se}U) in understanding the structure and antibiotic binding of bacterial ribosomal decoding site RNA motif (A-site) is described in this chapter. ^{Se}U represents a dual-purpose probe containing a fluorophore and an X-ray label. A-site RNA was chosen as the first test system as it is an important and well studied RNA target, and hence, the information on structure and antibiotic binding obtained using ^{Se}U-labeled A-site can be directly compared with the literature data. Bacterial rRNA, which plays important roles in protein synthesis process, is an attractive target for antibiotics such as aminoglycosides, macrolides, tetracyclines and synthetic oxazolidinoes.[16] A-site RNA motif is of particular significance as it monitors the cognate mRNA codon and tRNA anticodon interaction, and maintains the high fidelity of protein translation process.[17] The internal loop of this site is also a target for aminoglycoside antibiotics. In order to correlate the RNA conformation and recognition under equilibrium and in 3D, U1406, present in the internal loop and known to stabilize the interaction between the A-site and aminoglycosides, was replaced with ^{Se}U. The probe incorporated into the bacterial ribosomal decoding site RNA, fluorescently reported the antibiotic binding and provided diffraction information suitable for determining the structure (Figure 2). Comparison of CD, T_m and X-ray structures of ^{Se}U-labeled A-site with that of native A-site RNA indicated that ^{Se}U had very minimum impact on the native fold. Importantly, by comparing solution binding data and crystal structure, we gained insight on how the probe senses ligand-induced conformational change in RNA.[14] Taken together, our nucleoside probe represents a new class of biophysical tool that would compliment available tools for functional RNA investigations.

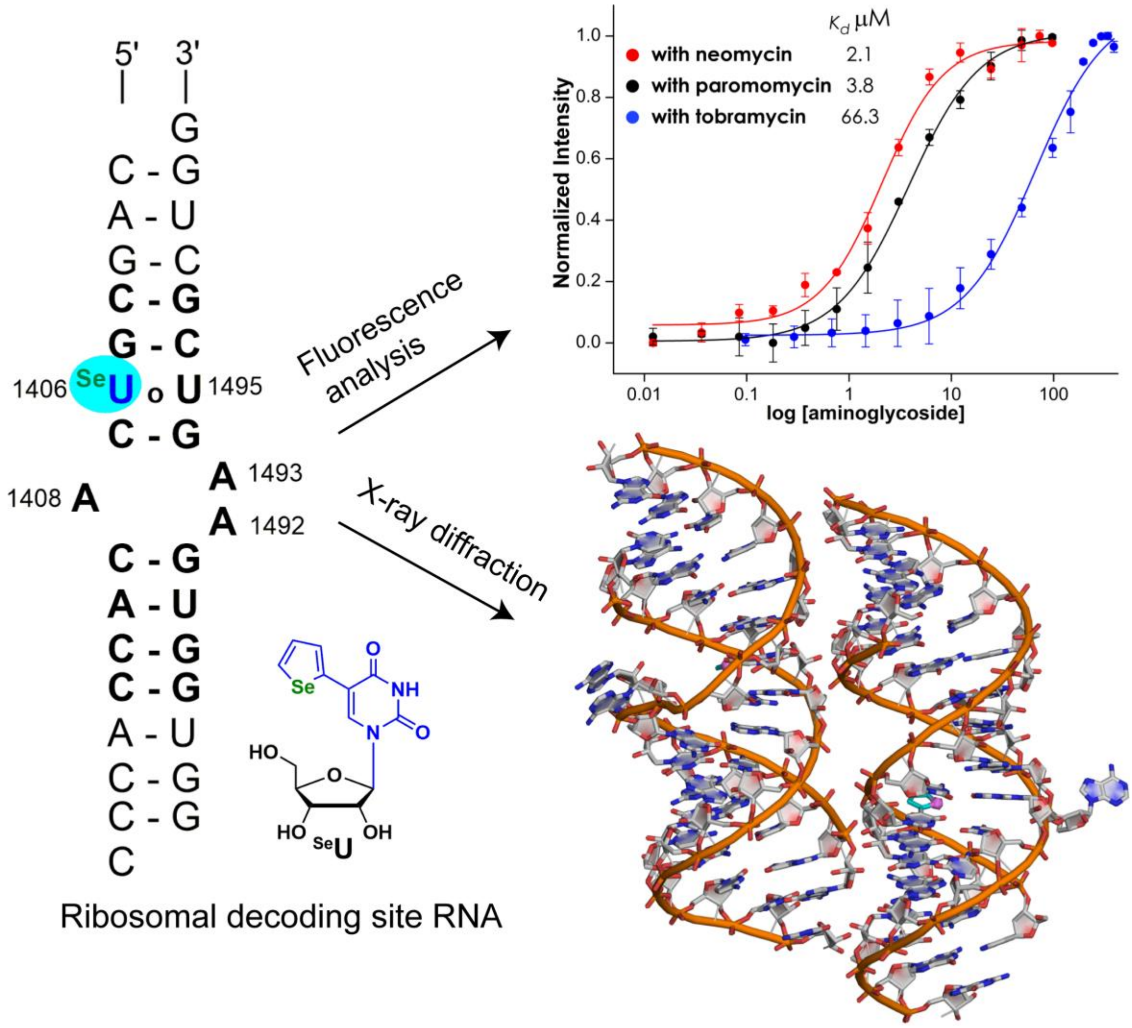


Figure 2. Two-in-one ribonucleoside probe (^{Se}U), when incorporated into the bacterial ribosomal decoding site RNA, enables the direct correlation of antibiotic-induced conformational change in real time and at atomic level without disrupting the native RNA folding.

Chapter 3: Probing G-quadruplex structures and ligand binding using selenophene-modified 2′-deoxyuridine

Biochemical and biophysical investigations strongly suggest that G-quadruplex (GQ) forming motifs play important roles in the maintenance of chromosomes and in the regulation of proliferation associated genes.[18] Several biophysical tools have been developed to study the structure, dynamics, small-molecule recognition properties and functions of GQ-forming sequences. Nevertheless, chemical probes that can be simultaneously used to understand the structure and recognition properties of GQs in real time and 3D will be highly useful in advancing our understanding of GQs, which in turn could aid GQ-based therapeutic strategies.

In order to expand the utility of selenophene-modified nucleoside analog, human telomeric DNA repeat was chosen. The human telomeric (H-Telo) DNA repeat $(TTAGGG)_n$, which end-caps the chromosomes, maintains the chromosomes from end-to-end fusion and degradation. The overhang of this sequence is known to adopt four-stranded structure called G-quadruplex. In terms of structure, H-Telo DNA repeat forms different GQ topologies (e.g., parallel, antiparallel, hybrid) *in vitro*, which depends on ionic conditions. Importantly, the conformation of loop residues in different GQ topologies is distinct. Based on this information, ^{Se}dU-labeled H-Telo DNA ONs were chemically synthesized wherein one of the loop dT/dU residues was replaced with the nucleoside analog. The nucleoside analog incorporated into H-Telo DNA ONs fluorescently reported the formation of GQ structures with significant enhancement in fluorescence intensity as compared to the duplex (Figure 3A). Further, we could successfully establish an assay to monitor and estimate the ligand binding to the GQ by fluorescence technique (Figure 3B). The H-Telo DNA ONs containing the ^{Se}dU modification at different positions within loop-2 was crystallized (Figure 3C). 3D structures determined by X-ray diffraction at 1.54 Å and 2.3 Å resolution indicated that the ON adopted a parallel GQ conformation. Superimposition of the modified GQ and native GQ indicated that the modification had only minor impact on the structure. Further analysis of the structure provided atomic level insights into the structural basis of the GQ sensing ability of the fluorescent nucleoside probe. This chapter describes the synthesis of ^{Se}dU-labeled H-Telo DNA ONs, fluorescence detection of GQ structure and ligand binding and X-ray structural analysis.[15]

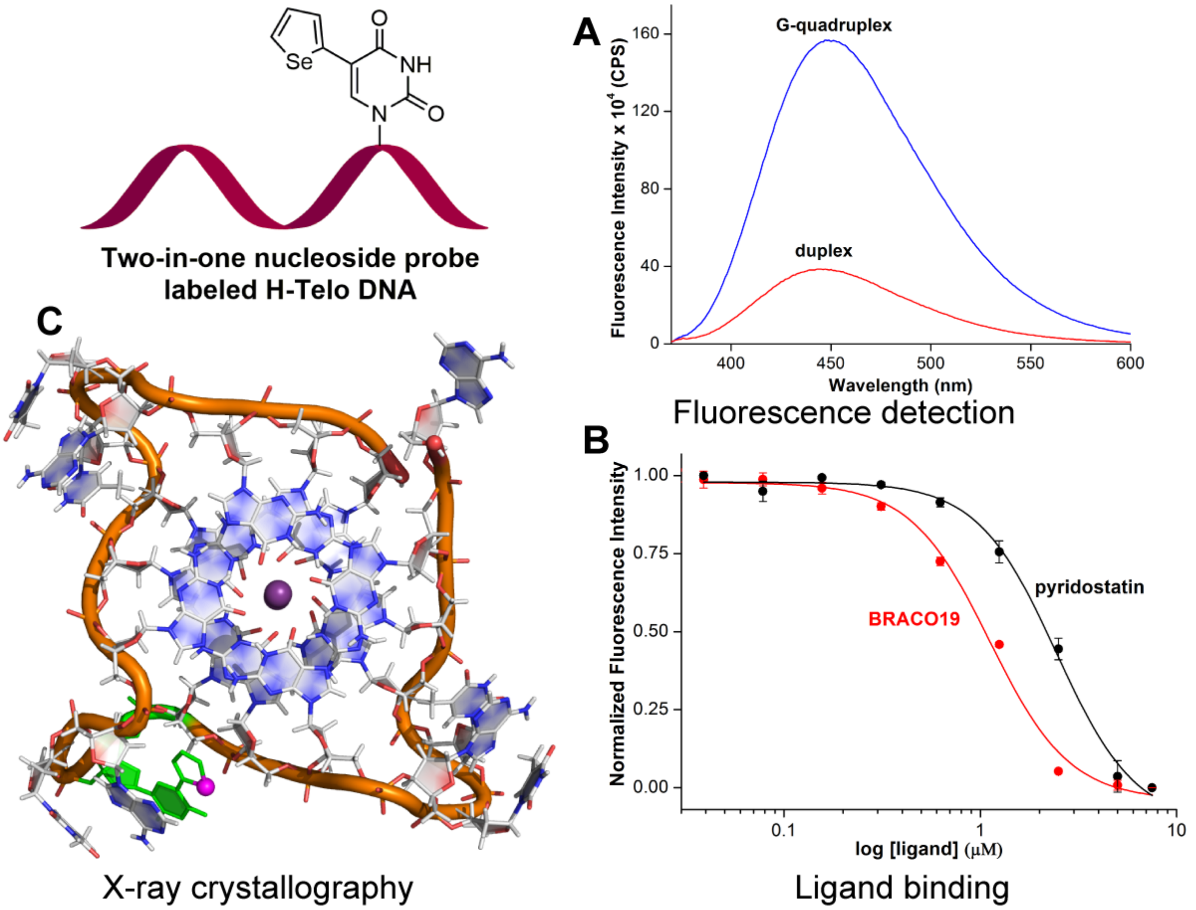


Figure 3. Two-in-one nucleoside probe, based on a 5-selenophene uracil core, for probing human telomeric (H-Telo) DNA repeats by fluorescence and X-ray crystallography techniques.

Chapter 4A: Surface-tuned and metal-ion-responsive supramolecular gels based on thymidine nucleolipids

The self-assembling properties of simple nucleoside-fatty acid nucleolipids was first investigated, wherein the nucleoside serves as the head group and long chain fatty acids as the lipophilic group. Depending on the nucleoside (2′-deoxy or ribonucleoside), length of fatty acid chain and position of fatty acid attachment to the sugar residue, the nucleolipids showed distinct self-assembling behaviour, morphology and recognition properties. This chapter is subdivided into two sections (4A and 4B) and describes two different potential applications of the nucleolipids.

In chapter 4A, the development of novel functional materials with surface tunability and metal ion responsiveness by using simple nucleolipid supramolecular synthons derived by attaching various fatty acids to the 3′-*O* or 3′,5′-*O* positions of the sugar residue of thymidine nucleoside is described. 3′,5′-*O*-di-fatty acid-substituted thymidines formed typical organogels in pure organic solvents, whereas, 3′-*O*-mono-fatty acid-substituted thymidine nucleolipids formed water-induced gels (Figure 3). A detailed morphological and structural analysis using microscopy, single crystal and powder X-ray diffraction and ^{1}H NMR techniques clearly revealed the molecular interactions invoked by nucleobase, sugar, fatty acid chain and water in setting up the path for hierarchical self-assembly and gelation of thymidine nucleolipids. Interestingly, the surface property of the xerogel film fabricated using 3′-*O*-monosubstituted nucleolipid gels could be switched from highly hydrophobic to hydrophilic and *vice versa* depending on the nature of the organic solvent-water mixture used in the gelation process. On the contrary, the gelation process of disubstituted thymidine nucleolipids was highly sensitive to the presence of Hg^{2+} ions as the metal ion formed T-Hg-T base pair, thereby disrupting the H-bonding interactions that favoured the gelation. Taken together, straightforward synthesis and modification-dependent gelation behaviour, surface tunability and metal ion responsiveness underscore the potential of these supramolecular nucleolipid synthons in constructing novel functional materials.[19]

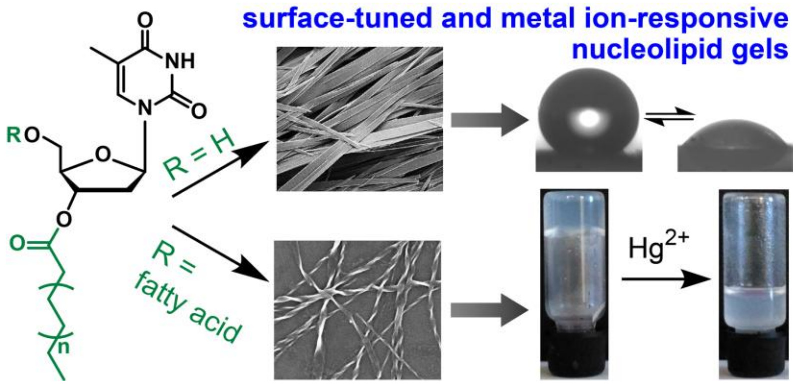


Figure 3. Design of self-assembling thymidine nucleolipids, Water induces the supramolecular gelation of 3′-*O*-mono-fatty acid-substituted nucleolipids dispersed in organic solvents. The surface of the xerogel films could be tuned between highly hydrophobic and hydrophilic by using an appropriate organic solvent-water mixture. 3′,5′-*O*-di-fatty acid-substituted nucleolipids form organogels, which is highly responsive to the presence of Hg^{2+} ions.

Chapter 4B: Microscopic assemblies of nucleolipid supramolecular synthons show self-sorting and cooperative self-assembling process

This section describes the self-assembling and recogntion properties of uridine and adenosine ribonucleolipids. The nucleolipids were sysnthesized by coupling fatty acids of different chain lengths to the 2′- and 3′-*O*-positions of ribose sugar through an ester linkage. The uridine nucleolipids formed stable opaque organogel in various organic solvents, whereas adenosine nucleolipids did not support organogelation (Figure 4A). The driving force for the hierarchical self-assembly leading to gelation was thoroughly characterized by microscopic, single crystal and powder X-ray diffraction, ^{1}H NMR and rheological analyses. In the packing structure of uridine nucleolipids, one uracil ring base pairs with two adjacent uridine nucleolipid residues via multiple H-bonding interactions. N3H and O4 atoms from one face and O4 and C5H atoms from the other face of the uracil ring participates in H-bonding interactions to form a 1D sheet structure. The formation of the sheet is also facilitated by strong H-bonding interaction between O2 and 5′-OH atoms of adjacent nucleolipids.

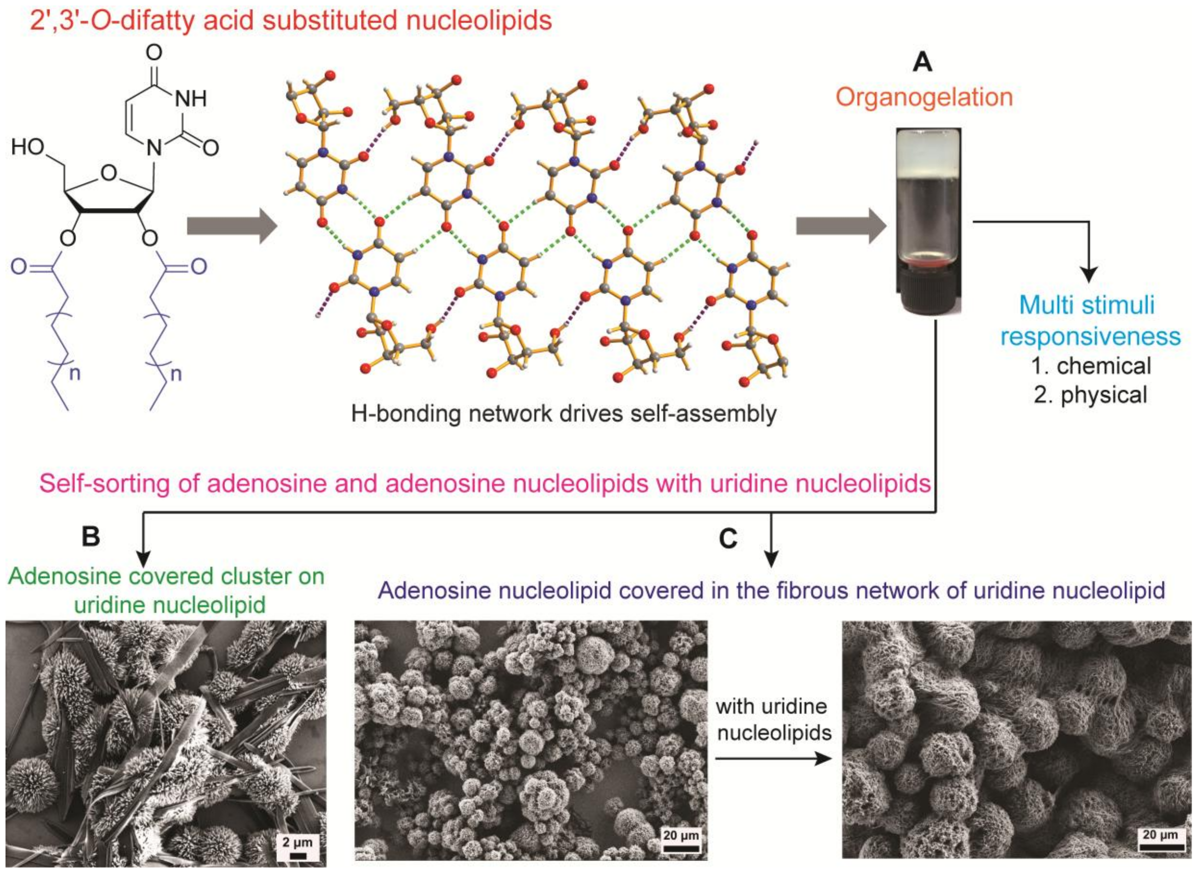


Figure 4. Structure of 2′-*O*- and 3′-*O*-dialkylated uridine based nucleolipids and their organogelation. These supramolecular assemblies are stabilized by multiple H-bonding and hydrophobic effect. Interestingly, adenosine or adenosine nucleolipid are self-sorting in the presence of complementary uridine nucleolipid. Furthermore, uridine nucleolipids exhibits multi stimuli responsiveness.

Importantly, the interactions in the crystal packing and in the gel states were similar as confirmed by ^{1}H NMR and PXRD measurements. Interestingly, SEM images of xerogels of uridine nucleolipids containing adenosine showed adenosine aggregating on the sheets formed by the nucleolipid (Figure 4B). However, such an interaction was not observed when other nucleosides were added (guanosine, cytidine and uridine). Addition of adenosine nucleolipid affected the gelation process of uridine nucleolipids. SEM images indicate that the porous microspheres of adenosine nucleolipid are covered in the fibrous network of uridine nucleolipid (Figure 4C). These morphological analyses along with ^{1}H NMR and rheological studies reveal that self-sorting of assemblies happen due to interactions at the microscopic level rather at the molecular level. On the other hand, a gel made of a combination of uridine nucleolipids containing fatty acids of different chain lengths displayed significantly higher storage modulus (higher gel strength) as compared to gels made of individual nucleolipids. SEM, PXRD and rheological studies indicate a cooperative self-assembling process leading to the formation of a stronger gel. Apart from these features, gels of uridine nucleolipids were found to be responsive to multiple stimuli (chemical and physical).[20]

Chapter 5: Switchable nucleolipid supramolecular gels based on environmentally-sensitive fluorescent nucleoside analogs

In this chapter, we describe a novel design strategy to construct supramolecular nucleolipid synthons by using environment-sensitivity fluorescent nucleoside analogs, based on 5-(benzofuran-2-yl)uracil and 5-(benzo[*b*]thiophen-2-yl)uracil cores, as the head group and fatty acids, attached to the ribose sugar, as the lipophilic group (Figure 5). Photophysical analysis in different solvents indicated that the fluorescence properties of nucleolipids are sensitive changes in solvent polarity and viscosity. Notably these studies revealed the presence of molecular rotor element in the nucleolipid, namely the heterobicycle attached to the nucleobase via an aryl-aryl single bond. The nucleolipids formed fluorescent organogels driven by hierarchical structures such as fiber, twisted ribbon, helical ribbon and nanotube, which depended on the nature of fatty acid chain and nucleobase modification. Thorough microscopy, ^{1}H NMR and X-ray structural analysis revealed the coordinated interplay of various non-covalent interactions in the gelation process.

Importantly, these nucleolipid gels retained or displayed aggregation-induced enhanced emission and their gelation behavior and photophysical properties could be reversibly switched by external stimuli such as temperature, ultrasound and chemicals. The selectively deprotonation of hydrogen atoms of nucleolipid gel involved in the H-bonding interaction by fluoride and Hg^{2+} ions enabled the two channel detection of these ions through visual phase transition and fluorescence change. Fluorescent organogels exhibiting such a combination of useful features is rare, and hence, we expect that this innovative design of fluorescent nucleolipid supramolecular synthons could lead to the emergence of a new family of smart optical materials and probes.[21]

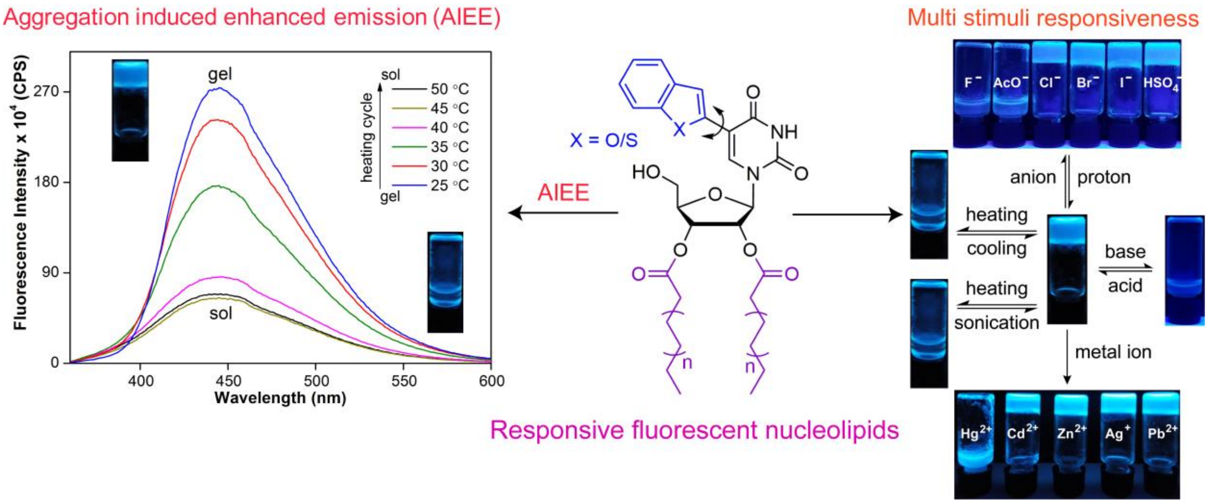


Figure 5. Structure of self-assembling and responsive fluorescent uridine based nucleolipids. These nucleolipids show aggreagation induced enhanced emission and multi stimuli responsiveness.

References:

1 David-Eden, H.; Mankin, A. S.; Mandel-Gutfreund, Y. *Nucleic Acids Res*. **2010**, *38*, 5982–5994.

2 (a) Olieric, V.; Rieder, U.; Lang, K.; Serganov, A.; Schulze-Briese, C.; Micura, R.; Dumas, P.; Ennifar, E. *RNA* **2009**, *15*, 707–715. (b) Juskowiak, B. *Anal. Bioanal. Chem*. **2011**, *399*, 3157–3176. (c) Bardaro Jr, M. F.; Varani, G. *WIREs RNA* **2012**, *3*, 122–132. (d) Shelke, S. A.; Sigurdsson, S. T. *Eur. J. Org. Chem*. **2012**, 2291–2301.

3 (a) Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z. *Nucleic Acids Res*. **2007**, *35*, 477–485. (b) Nelissen, F. H.; van Gammeren, A. J.; Tessari, M.; Girard, F. C.; Heus, H. A.; Wijmenga, S. S. *Nucleic Acids Res*. **2008**, *36*, e89. (c) Sigurdsson, S. T. *Pure Appl. Chem*. **2011**, *83*, 677–686. (d) Phelps, K.; Morris, A.; Beal, P. A. *ACS Chem Biol*. **2012**, *7*, 100–109.

4 (a) Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem*. **2013**, *53*, 366–378. (b) Xu, W.; Chan, K. M.; Kool, E. T. *Nat. Chem*. **2017**, *9*, 1043–1055.

5 Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.

6 Lin, L.; Sheng, J.; Huang, Z. *Chem. Soc. Rev*. **2011**, *40*, 4591–4602.

7 (a) Krishnan, Y.; Simmel, F. C. *Angew. Chem., Int. Ed*. **2011**, *50*, 3124–3156. (b) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. *Science* **2015**, *347*, 1260901.

8 Ackermann, D.; Schmidt, T. L.; Hannam, J. S.; Purohit, C. S.; Heckel, A.; Famulok, M. *Nat. Nanotechnol*. **2010**, *5*, 436–442.

9 (a) Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335*, 831–834. (b) Bhatia, D.; Chakraborty, S.; Mehtab, S.; Krishnan, Y. *Methods Mol. Biol*. **2013**, *99*1, 65–80.

10 Peters, G. M.; Davis, J. T. *Chem. Soc. Rev*. **2016**, *45*, 3188–3206.

11 (a) Berti, D.; Montis, C.; Baglioni, P. *Soft Matter* **2011**, *7*, 7150–7158. (b) Allain, V.; Bourgaux, C.; Couvreur, P. *Nucleic Acid Res*. **2012**, *40*, 1891–1903.

12 (a) Gao, Y.; Kuang, Y.; Guo, Z.-F.; Guo, Z.; Krauss, I. J.; Xu, B. *J. Am. Chem. Soc*. **2009**, *131*, 13576–13577. (b) Lanna, C. M.; Lusic, H.; Complo, M.; McIntosh, T. J.; Barthélémy, P.; Grinstaff, M. W. *Acc. Chem. Res*. **2012**, *45*, 1026–1038.

13 Pawar, M. G.; Nuthanakanti, A.; Srivatsan, S. G. *Bioconjugate Chem*. **2013**, *24*, 1367–1377.

14 Nuthanakanti, A.; Boerneke, M. A.; Hermann, T.; Srivatsan, S. G. *Angew. Chem., Int. Ed*. **2017**, *56*, 2640–2644.

15 Nuthanakanti, A.; Khan, I. A.; Kayarat, S.; Srivatsan, S. G. (*Manuscript under preparation)*.

16 Hermann, T. *Biopolymers* **2003**, *70*, 4–18.

17 Ogle, J. M.; Ramakrishnan, V. *Annu. Rev. Biochem*. **2005**, *74*, 129–177.

18 (a) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res*. **2007**, *35*, 7429–7455. (b) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261–275.

19 Nuthanakanti, A.; Srivatsan, S. G. *ACS Appl. Mater. Interfaces* **2017**, *9*, 22864–22874.

20 Nuthanakanti, A.; Srivatsan, S. G. (*Manuscript under preparation)*.

21 Nuthanakanti, A.; Srivatsan, S. G. *Nanoscale* **2016**, *8*, 3607–3619.

List of Publications

1. **Nuthanakanti, A**.; Boerneke, M. A.; Hermann, T.; Srivatsan, S. G. Structure of the ribosomal RNA decoding site containing a selenium-modified responsive fluorescent ribonucleoside Probe. *Angew. Chem., Int. Ed.* **2017**, *56,* 2640–2644.
2. **Nuthanakanti, A**.; Srivatsan, S. G. Surface-tuned and metal ion-responsive supramolecular gels based on thymidine nucleolipids *ACS Appl. Mater. Interfaces* **2017**, *9,* 22864–22874.
3. **Nuthanakanti, A**.; Srivatsan, S. G. Hierarchical self-assembly of switchable nucleolipid supramolecular gels based on environmentally-sensitive fluorescent nucleoside analogs. *Nanoscale* **2016**, *8,* 3607–3619.
4. Pawar, M. G.; **Nuthanakanti, A**.; Srivatsan, S. G. Heavy atom containing fluorescent ribonucleoside analog probe for the fluorescence detection of RNA-ligand binding. *Bioconjugate Chem*. **2013**, *24,* 1367–1377.
5. Sabale, P. M.; **Nuthanakanti, A**.; Srivatsan, S. G. Synthesis and fluorescence properties of a full set of extended RNA base analogues. *Ind. J. Chem.* **2013**, *52A,* 1004–1013.
6. **Nuthanakanti, A**.; Khan, I. A.; Kayarat, S.; Srivatsan, S. G. Probing human telomeric G-quadruplex structure and ligand binding in real time and 3D by using a dual-purpose nucleoside probe *(manuscript under preparation)*.
7. **Nuthanakanti, A**.; Srivatsan, S. G. Microscopic assemblies of nucleolipid supramolecular synthons show self-sorting and cooperative self-assembling process *(manuscript under preparation)*.
8. **Nuthanakanti, A**.; Srivatsan, S. G. Heterotypic self-assembly of 5′-*O*-fatty acid conjugated nucleolipid hydrogels selectively adsorbs cationic dyes *(manuscript under preparation)*.
9. **Nuthanakanti, A**.; Srivatsan, S. G. Multi-stimuli responsive ribothymidine nucleolipids synthesis and self-assembling properties *(manuscript under preparation)*.

Chapter 1

Nucleic acid probes and supramolecular assemblies based on functionalized nucleosides.

1.1 Introduction

Nucleic acids, DNA and RNA, as we know perform several essential cellular processes, which include transfer and storage of genetic information, catalysis, and regulation of gene expression.[1] Despite being chemically less diverse as compared to proteins and oligosaccharides, nucleic acids perform these functions by utilizing their in-built conformational dynamics and by adopting complex secondary and tertiary structures.[2] DNA exist in three helical forms like A-DNA, B-DNA and Z-DNA, however the most common conformation B-DNA forms a right-handed double helix. Major groove of B-DNA is known to be deep and wide and its minor groove is narrow and shallow.[3] A-DNA also adopts a right-handed helix however the major groove is deep and narrow and not accessible to proteins. The minor groove is wide and shallow and accessible to proteins.[4] The Z-DNA is left-handed double helix, with only one groove, which is deep and narrow.[5]

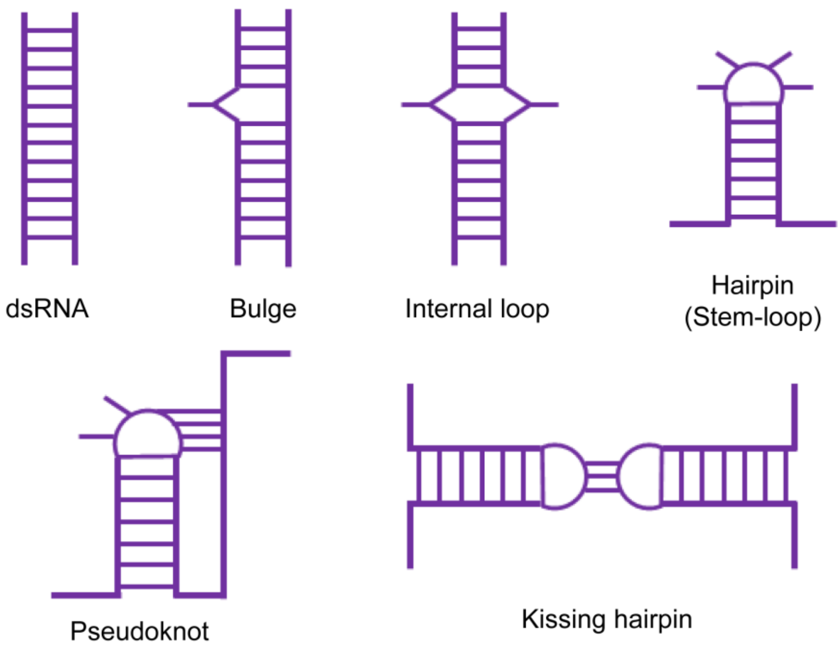


Figure 1. Representive examples of RNA secondary and tertiary structures.

Unlike DNA, RNA double helix exhibit less polymorphism, and predominantly adopts the A-form with deep and narrow major groove, and shallow and wide minor groove. Therefore, both major and minor grooves of the duplex region of RNA are not very conducible to binding to small and large proteins. In order to perform the essential biological functions, RNA adopts a wide variety of secondary and tertiary structures. A RNA molecule can have several secondary structural motifs assisted by a combination of paired and unpaired bases and stacking interaction between neighbouring bases. For example, RNA can fold to form duplex, duplex with a bulge, hair-pin loops and internal loops (Figure 1).[6] These structures can further interact to form tertiary structures like kissing loops and pseudoknots (Figure 1).[7]

Evidently, many of the nucleic acid structural motifs also provide binding sites for antibiotics and small molecule drugs.[8] Several biophysical tools based on fluorescence, NMR, EPR, X-ray techniques, circular dichroism (CD), to name a few, have been developed to advance the understanding of the structure-function relationship of nucleic acids.[9] Many of these biophysical investigations require appropriately labeled DNA and RNA oligonucleotides (ONs). For fluorescence analysis, it becomes imperative to use fluorescently modified ONs as the native nucleosides are practically non-emissive. Similarly, NMR, EPR and X-ray techniques require isotope-, radical- and heavy atom-labeled ONs, respectively.[10] In this context, modified nucleoside analogs, incorporated into target ON sequences, have provided efficient systems not only to study nucleic acid structure, dynamics and function but also to setup discovery platforms to identify efficient binders.

Functionalized nucleoside analogs are usually prepared by attaching reporters or labels on to the nucleobase, sugar or phosphate backbone. Fluorescent nucleoside analogs are typically prepared by attaching known fluorophores to purine and pyrimidine rings or 2′-position of sugar residue.[11] The corresponding phosphoramidite or triphosphate is then prepared and incorporated into ONs by solid-phase or enzymatic methods. Alternatively, attaching or fusing heterocycles onto nucleobases have generated environment-sensitive fluorescent nucleoside analogs.[12] Many such analogs are minimally perturbing, and importantly, are highly useful in surveying the local conformational changes that occur during the folding event or upon ligand binding. ONs suitable for NMR analysis are prepared by incorporating isotope (^{13}C, ^{15}N etc) enrich nucleosides by chemical or enzymatic methods. In case of X-ray analysis, ONs labeled with halogenated (Br or I) nucleosides are usually used. More recently, excellent anomalous scattering property of Se atom, which is widely used in protein X-ray crystallography, is also being applied in nucleic acid structural analysis. The Se-labeled ONs are prepared by incorporating nucleoside analogs in which the carbonyl oxygen of nucleobase or sugar oxygen is replaced with Se atom. While the benefits of such custom labeled ONs is arguably undeniable, correlating the information obtained from different techniques, which work under different conditions, is difficult and liable to give obscure correlation. This we strongly believe is because the scientific community has greatly relied on the traditional one-label one-technique due to paucity of tools that can be efficiently used in more than one biophysical technique.

Apart from the biological role and therapeutic potential, the ability to program nucleic acid structures by using their characteristic recognition properties has enabled the construction of several nano-architectures.[13] Many of these self-assembled systems have been used in devising nano machines[14] and containers,[15] and therapeutic and diagnostic tools.[16] However, challenges in large-scale fabrication and scalability problem have hampered their wide application. While efforts are being made to address these shortcomings, the basic constitute of nucleic acids, nucleobases and nucleosides, and their derivatives, are proving to be very useful supramolecular synthons for the fabrication of various architectures. The advantages are–(i) nucleosides can be easily modified to install functionalities without affecting the native H-binding interaction by using established synthetic approaches, (ii) scalability is not a major issue, and (iii) the architecture and properties can be tuned by appropriate chemical modification. Many tailor-made nucleobase and nucleoside derivatives self-assemble to form vesicles, fibres, gels and synthetic ion channels, which are useful as gene and drug delivery vehicles. However, the potential of nucleoside-based synthons in designing smart self-assembled materials, which are tunable and can show interesting optical and sensing properties, is not well explored.

In this chapter, a brief discussion on the different structures adopted by nucleic acid sequences and their role in cellular process is provided. Following this, an overview of various biophysical tools that have been developed to study the structure and function of nucleic acid is presented. In particular, a detailed discussion on recent developments in the design and applications of base-modified nucleoside analogs is presented. Further, design of nucleoside-based supramolecular synthons, and their self-assembling behaviour and biomedical applications are discussed. The limitations of currently available nucleoside probes and the envisioned potential of nucleoside supramolecular synthons in materials application, which laid the foundation for the present work, are described.

1.2 Non-canonical structures of DNA

Nucleic acid has the ability to acquire several non-canonical secondary structures such as triplexes, i-motif, G-quadruplex, hairpin, cruciform, etc., which depend on the sequence and its interaction with proteins and metabolites. Studies suggest that these non-canonical structures may influence the gene regulation process and also involve in several other biological processes.[17]

1.2.1 Triple helix structure

DNA triple helix, unlike DNA double-helix contain both of the Watson-Crick and Hoogsteen base pairing. Triplex is more prevalent in double-stranded DNA consisting of purine-rich on one strand and pyrimidine on the other (Figure 2). The triple helical structure forms when single-stranded pyrimidine-rich DNA interacts with the major groove of duplex DNA through Hoogsteen hydrogen bonds.[18] The third strand in triple helix is aligned antiparallel to the duplex. The triplexes are involved in the inhibition of DNA replication and interference of transcription processes.[19]

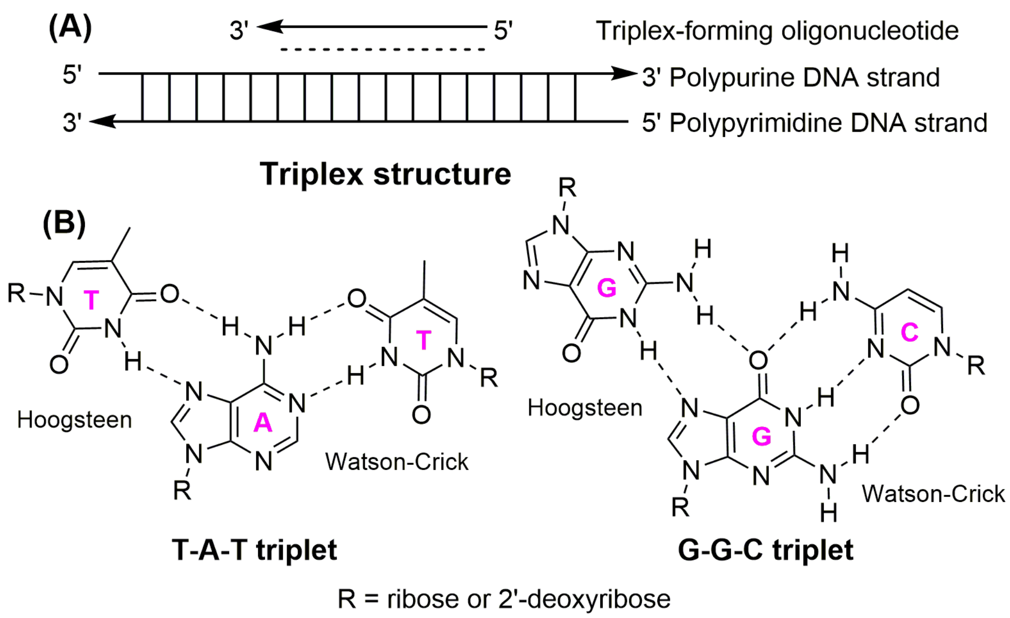


Figure 2. Triple helix formation (**A**) the orientation of the third strand in the pyrimidine triple-helix motif. (**B**) Base triplet structure of purine formed by using both Watson-Crick and Hoogsteen hydrogen bonds that stabilize triple-helix formation.[20]

1.2.2 G-quadruplex

Guanine-rich DNA and RNA sequences can form stacked nucleic acid structure. In 1962 Gellert *et al*. reported that guanylic acids can assemble into tetrameric structures.[21] In these motifs, four guanine molecules are held together to form a planar G-quartet (Figure 3A), in which each guanine is involved in four hydrogen bonds with two adjacent guanines. Two or more G-quartets stack on top of each other to form a G-quadruplex structure (Figure 3). The

G-quadruplex forming sequences are found in the genome, at promoters, telomeric regions and 5′-UTRs of mRNA.[22] The G-quadruplex core possess two or more stacked G-tetrad arranged in a right-handed helical twist. These motifs were stabilized by three main aspects: hydrogen bonding between the guanines within a G-tetrad, π–π stacking interactions between the adjacent G-tetrad and electrostatic interactions between the quartet core and monovalent cation. Replacing any of the guanine (G) to non-G bases adversely affects the stability, and such mutant sequences do not form G-quadruplexes.[23]

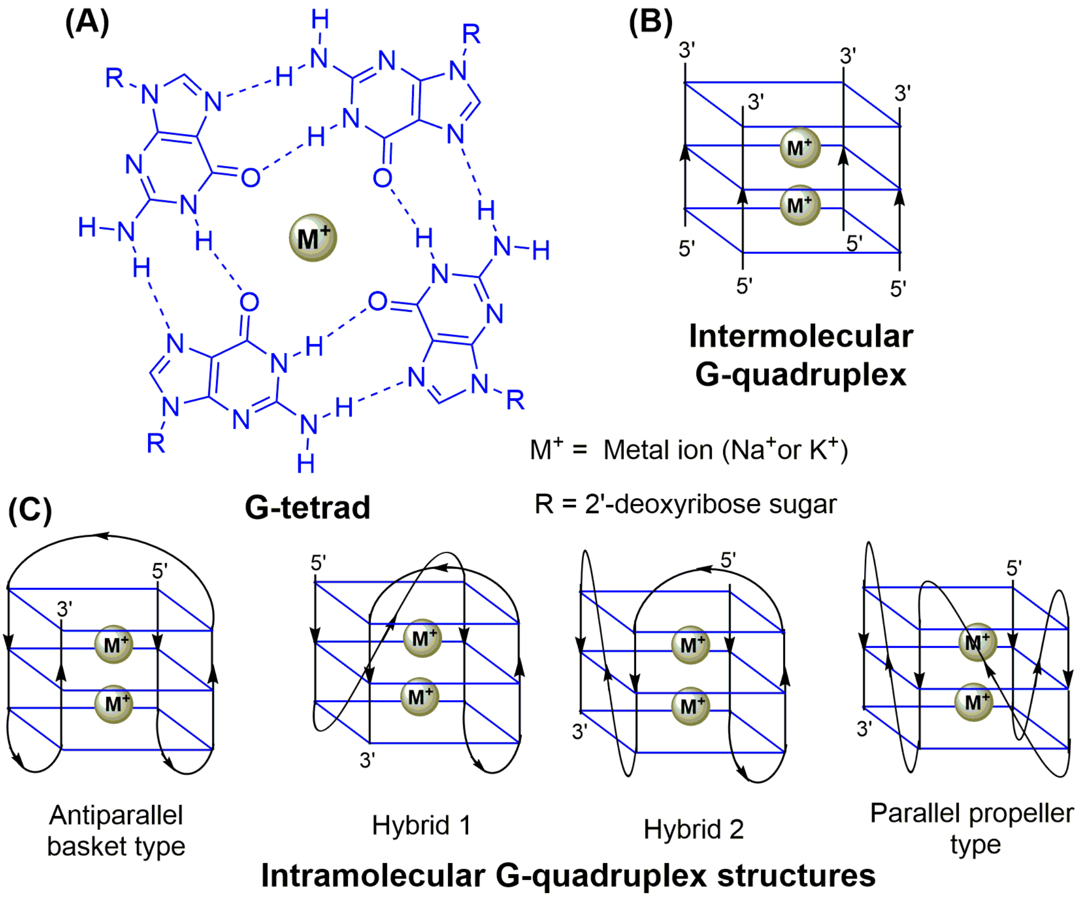


Figure 3. (A) Structure of the G-tetrad with central monovalent ion (Na^+ or K^+). (B) Schematic illustration of intermolecular telomeric G-quadruplex structure is shown in diagram. (C) Four distinct folding topologies of intramolecular telomeric G-quadruplex structures and the arrows indicate strand polarity.

These are connected by single-stranded flexible loops, which extrude from G-tetrad core. The quadruplex structures vary depending on the guanine content and number of nucleotides in the loop. Although, the loop length is typically 1–7 nucleotides, smaller loops stabilize effectively and facilitates the formation of higher order structures,[24] whereas longer loops promote more intramolecular and antiparallel conformation.[25] The negatively charged

G-quartet core is stabilized by monovalent cations, such as sodium or potassium ions, sandwiched between the stacks, and they co-ordinate with eight guanine oxygens.[26] G4 structures can adopt various topologies, depending on the orientation of the DNA strand. They can be classified under four major topologies such as antiparallel, hybrid-1, hybrid-2 and parallel structures. GQ formed within one strand are called intramolecular (Figure 3C) and different strand named as intermolecular with variety of loop structures (Figure 3B).[27] The stability and topology of the quadruplex mainly depends on the length of the sequence and number of G-tetrads, loop length[28] and nature of metal ion.[29]
G-quadruplex formation can play important regulatory role in cells such as telomeres capping, where in it protect the chromosome from telomerase, end to end fusion and exonuclease degradation.[30] Also, in the promoter region G-quadruplex forming sequences control the gene expression during the transcription.[31] The 5′-untranslated regions (5′-UTR) of mRNA having G-quadruplexes were shown to inhibit of translation.[32]

1.2.3 i-motif DNA

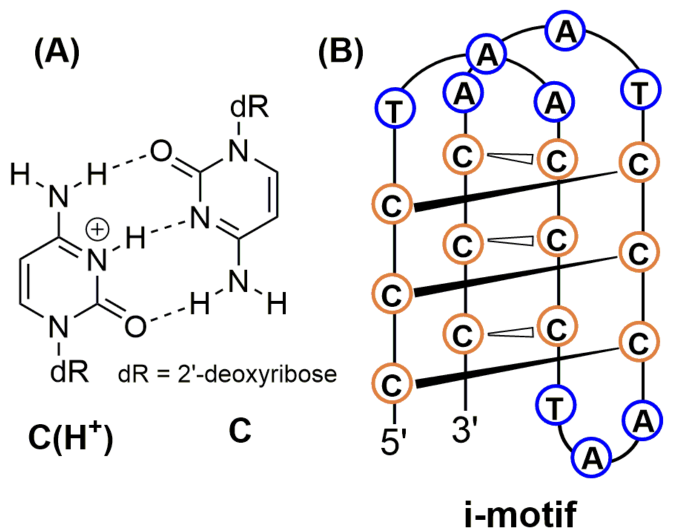


Figure 3. (**A**) Protonated-cytosine: cytosine base pair. (**B**) Schematic diagram of i-motif structure formation was shown above.

Analogous to G-rich sequences, cytosine-rich (C-rich) sequences are known to form i-motif structure at mild acidic pH (~5). Such sequences are present in the promoter region of certain oncogenes and human telomeric DNA repeat region.[33] The i-motif structure is comprised of two parallel (C:C+) hemiprotonated duplexes, which are interconnected by intercalated interactions in an antiparallel manner (Figure 4). Depending on the number of cytosine bases, loop length and solution pH, the C-rich sequence exhibit structural polymorphism.[34] The pH dependent folding and unfolding of i-motif structures have been used in devising pH switches for nanotechnology applications.[35] However, the importance of the i-motif in biological context in unclear. Recently, Hurley *et al*. have demonstrated that i-motif forming sequences regulates transcription of pro-survival oncoprotein.[36]

1.3 Functionalized nucleotide analogs

Nucleobase, sugar and phosphate groups contain important sites for incorporation of various functional moieties into nucleic acids. These modified ONs were used for elucidating nucleic acids structure and dynamics and also in devising various supramolecular assemblies and in therapeutic tools. Appropriate functional groups were inserted into the pyrimidine bases at C2, C4 and C5 positions and purine nucleobases at C2, C6 and C7 positions without affecting the H-bonding face. Sugar moiety at C2′ and O4′ are readily accessible sites for internal-labeling of nucleic acids. However, labeling at O5′ and O3′-positions limits the modification to the ONs terminals. The phosphate backbone of ONs can also be used for further derivatization, wherein the phosphorous non-bridging oxygen is replaced with suitable labels. These modified nucleic acid components can serve as nucleoside probes, building blocks for supramolecular self-assemblies and therapeutic tools for antisense technology (Figure 5)

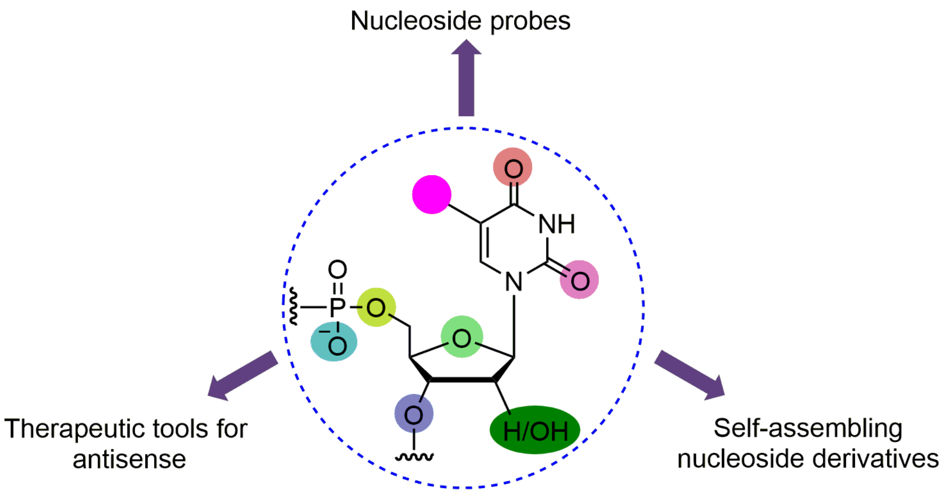


Figure 5. Schematic illustration of sites for chemical modifications of uridine nucleotide. Appropriately derivatized nucleic acid components (nucleobase, nucleoside and nucleotide) can serve as nucleoside probes, self-assembling building block and therapeutic tools for antisense therapy. Colour labels indicate possible modification sites.

1.3.1 Biophysical techniques and nucleoside probes for studying structure and dynamics of nucleic acids

To study nucleic acid structure, dynamics and function, several biochemical and biophysical methods such as fluorescence, NMR, EPR, CD, FRET, ITC and X-ray crystallography have been employed. Among them X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy provide detailed 3D information at atomic level. Obtaining diffraction quality crystals of nucleic acids is a challenge and sometimes the acquired crystal structure may not be of biologically active conformation.[37] NMR spectroscopy provides high-resolution solution structural information under near physiological conditions but becomes more challenging when the size of the nucleic acids is more than 30 kDa.[38] FRET and EPR have been utilized to study nucleic acids in biological systems. Using these two techniques one can evaluate long-range distances and hence the conformational changes in folding by monitoring the probes response.[39]

1.3.1.1 Nuclear Magnetic Resonance (NMR) spectroscopy

NMR spectroscopy being a powerful biophysical technique, and it provides 3D structure of nucleic acids in the solution state. Determination of nucleic acid structure by NMR is not straightforward as it requires nucleic acid selectively labeled with appropriate stable isotopes (Figure 6).[40] The isotope-labeled nucleic acids (^{1}H, ^{13}C, ^{15}N, ^{31}P) could provide the information about the base-pairing pattern, conformational equilibria, metal ion binding, formation of secondary structure motifs, local and global conformational dynamics and interactions with small ligands and proteins.[41] Williamson and co-worker demonstrated *de novo* purine biosynthesis of isotopically labeled purine nucleotides (**1**-2).[42] Pitsch and co-worker prepared phosphoramidite of 2′-TOM-protected ^{15}N-labeled ^{15}N(3)-pyrimidine (**3**) and ^{15}N(1) purine (**4**) for site-specific labelling of RNA, which allowed the study of RNA folding kinetics (Figure 6).[43] Micura and co-workers incorporated ^{15}N(4) cytidine (**5**) into preQ1 riboswitch class I aptamer for probing pseudoknot formation of the aptamer by NMR.[44] Likewise, site-specific incorporation of fluorine atom into nucleic acids enabled the validation of different conformations by ^{19}F NMR spectroscopy. Sever resonance degeneracy was observed in ^{1}H NMR imino proton, but in ^{19}F spectra, because of 100% natural abundance and larger chemical shift distribution of fluorine as compared to ^{1}H, it was found to be strongly diminished. Williamson *et al*. incorporated 5-fluorouridine-5′-triphosphate (**6**) into RNA by using an enzyme, and this non-perturbing probe was used to study the RNA structure and dynamics by ^{19}F NMR (Figure 6).[45] Fluorinated adenosine analogs (**7-8**) were used for studying nucleobase protonation in ribosomes by nucleotide analog interference mapping (NAIM).[46]

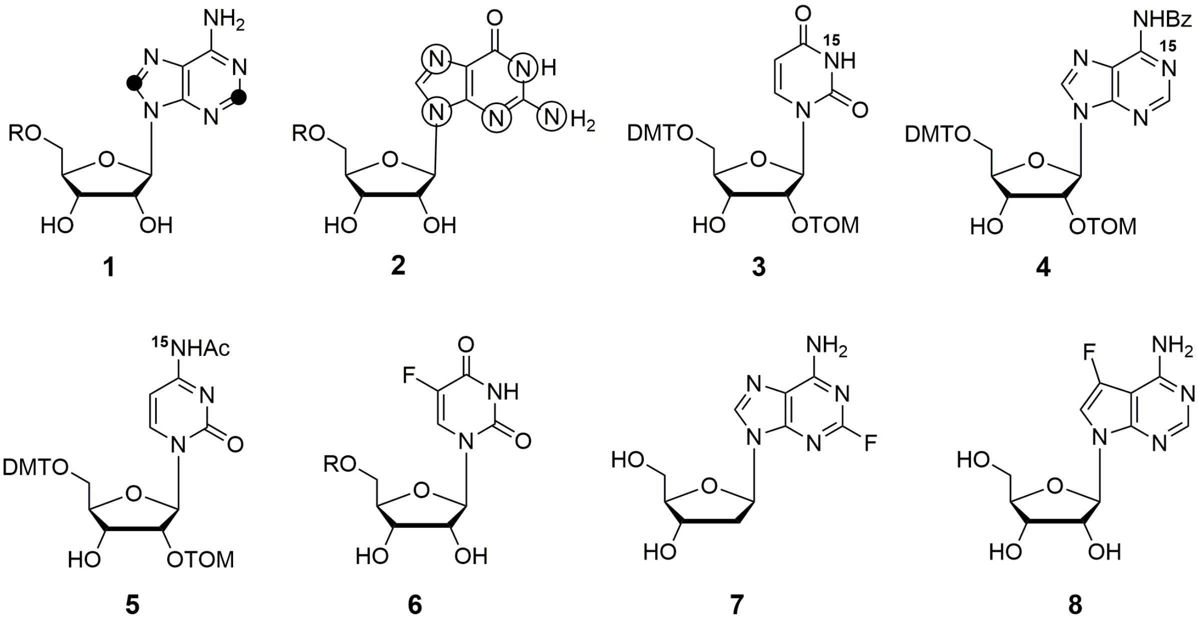


Figure 6. Representative examples of isotope-labeled nucleoside probes used for NMR spectroscopy. Carbon isotopes (^{13}C) are shown in block solid dots and nitrogen (^{15}N) are indicated by open circles. (R = triphosphate, DMT = 4,4′-dimethoxytrityl, TOM = tri-iso-propylsilyloxymethyl).

1.3.1.2 Electron Paramagnetic Resonance (EPR) spectroscopy

EPR spectroscopy can be applied to study the structure of different nucleic acids and determine local environment of paramagnetic EPR centers in DNA and RNA. To do so, incorporation of stable free radical into the oligonucleotide is required. This enables structural characterization and elucidation of dynamic structural changes in nucleic acids.[47] Bobst *et al.* described the synthesis of conformation-sensitive, sequence-specific, and spin-labeled RNA ON (**9**), and used it as nucleic acid hybridization probe. They have shown that this probe exhibits a characteristic EPR signature in both single-strand and double-helix states (Figure 7).[48] Gannett and co-workers prepared spin-labeled (**10**), nitroxide functional group (N-O) bond aligned with respect to the nucleobase.[49] This feature proved to be highly useful to evaluate accurate distances in **10-**labelled nucleic acids by EPR spectroscopy (Figure 7). Therefore, this has been exploited as a probe to distinguish between the formation of DNA duplex and triplex structures by continuous wave electron paramagnetic resonance CW-

EPR,[50] and for examining the G-quadruplex formation and topologies in human telomeric DNA by pulsed electron double resonance (PELDOR).[51] Sigurdsson introduced a TEMPO spin-label on exocyclic amino groups of 2′-deoxycytidine (**11**), which was involved in hydrogen bonding with its complementary nucleobase. Furthermore, they used this feature of the probe to determine single base mismatches. It has also been shown to discriminate between perfect and mismatched base-pairing partner in duplex DNA by using a CW-EPR spectroscopy.[52]

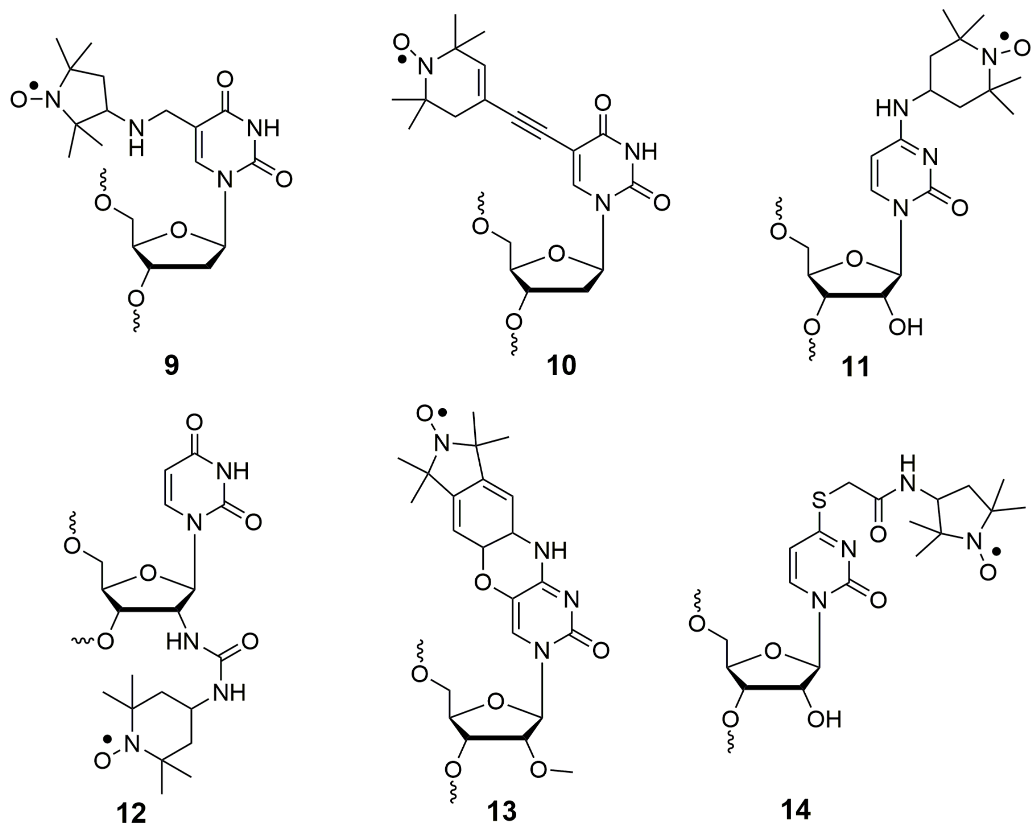


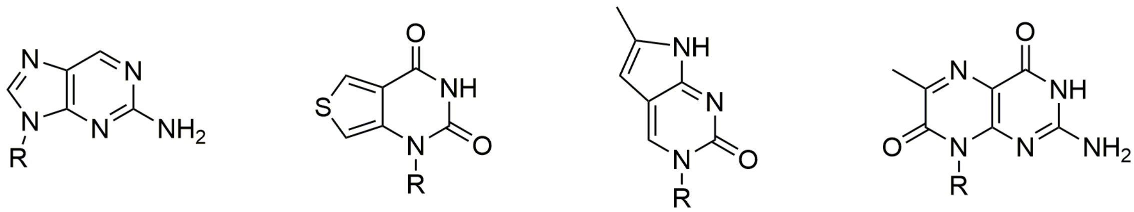
Figure 7. Representative examples of spin labeled nucleic acid probes for EPR spectroscopy.

Sigurdsson and co-workers developed a urea-linked 2′-spin-labeled (**12**). This label was allowed to monitor the structure-dependent conformations of the trans-activation-responsive (TAR) RNA.[53] Further, it was used for investigating the ligand-induced conformational changes of the tetracycline aptamer[54] and nucleic acids distance measurements by pulsed EPR.[55] They have also implemented bifunctional cytosine based rigid spin label (**13**), which possess a methoxy group at 2′-position. It has been chemically incorporated into different RNA ONs[56] and exploited for distance determinations and orientation selection in RNA ONs by PELDOR.[57] Recently, Famulok group postsynthetically incorporated spin label (**13**) into RNA ONs through click-chemistry for PELDOR measurements.[58] Varani and co-workers site-specifically incorporated 4-thiouridine into RNA by solid phase synthesis, followed by post-synthetic base labeling with iodo-derivatized spin-label to obtain the EPR labelled (**14**) RNA.[59] This probe enabled NMR studies of RNA-protein complexes and also for discrimination of long-range distances in RNA by double-quantum coherence (DQC) EPR spectroscopy.[60]

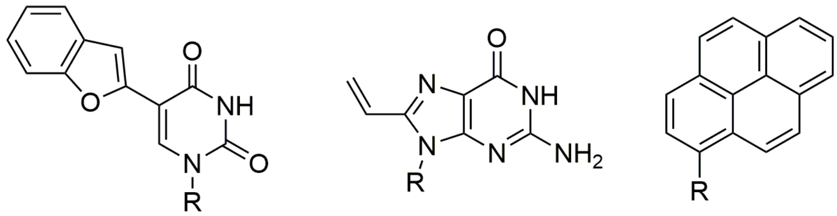
1.3.1.3 Fluorescence spectroscopy

Fluorescence spectroscopy is one of the most powerful biophysical techniques for analysing biomolecules–their structure, dynamics and interactions with various ligands with high sensitivity.[61] Unlike, some proteins, nucleic acids are non-emissive therefore purine and pyrimidine nucleobases need to be modified to show useful fluorescence properties (Figure 8). Different synthetic organic approaches have been developed for synthesizing emissive nucleosides analogs but only a small number of analogs fulfil the criterion of isomorphicity. Isomorphic nucleobases mimic the native structure with respect to their shape, size and hydrogen bonding face.[62] Base-modified nucleobase analogs are classified into five major categories, (1) size-expanded base, (2) pteridines, (3) extended bases, (4) isomorphic bases and (5) polycyclic aromatic hydrocarbon (PAH) base analogs. Size-expanded fluorescent nucleosides are generated by fusing additional aromatic rings onto purine and pyrimidine bases.[63] Pteridines are natural bicyclic planar compounds, which are emissive and structurally similar to purines.[64] Extended nucleobases are produced by attaching known fluorophores such as pyrene,[65] perylene,[66] bipyridine, thiazole orange and anthracene to the nucleobases via flexible or rigid linkers. PAH bases are formulated by natural nucleobase replaced with polycyclic aromatic hydrocarbons fluorophores, which comprises nonpolar aromatic ring and there is no Watson-Crick hydrogen bonding face.[63]

2-Aminopurine (**15**), an adenosine analog, is the most commonly used fluorescent nucleoside probe, which can form stable base pairs with thymine or uridine as it is equivalent to A-T/U base pair (Figure 8). It exhibits lower fluorescence intensity while stacked in-between adjacent bases, however the fluorescent intensity substantially increases upon unstacking and exposure to aqueous media.[67] However, its excitation and emission maxima lie in the UV region, which limits its applications in cell-based assays.[68] 2-Aminopurine (**15**) permits investigation of the RNA folding and cleavage mechanism of ribozymes,[69] and it also allows the study of RNA-protein interactions and ligand binding of RNA.[70] Tor and co-worker prepared a fused thiophene derivative of uridine analogs (**16**), which had shown higher quantum yield and better solvatochromism. This probe served as a nucleobase discriminating probe[71] and was used for detection of abasic sites in RNA caused by ribosome inactivating protein (RIP) toxin.[72]



15 **16** **17** **18**



19 **20** **21** **22**

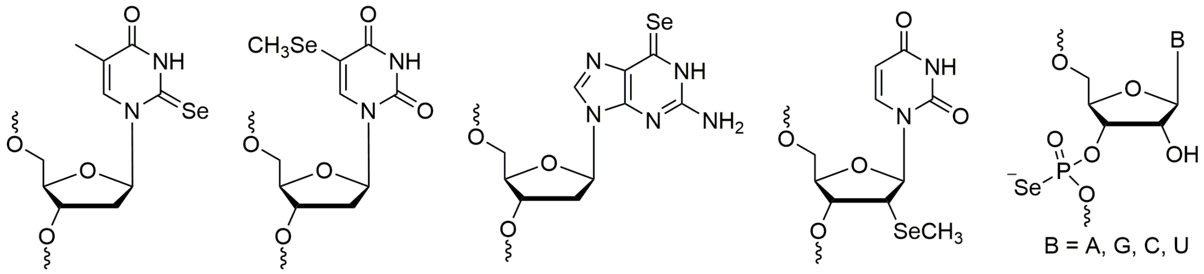
Figure 8. A collection of fluorescent nucleoside analogs, R = ribose or 2′-deoxyribose.

Pyrrolocytidine (**17**), an analog of cytidine has been exerted for examining RNA secondary structures and forms stable base-pairing to guanosine, like cytidine (Figure 8). This analog is highly fluorescent as a free nucleoside, but upon incorporation into ON duplexes, fluorescence is drastically quenched. The responsiveness of this probe has been used for monitoring the association and dissociation kinetics of RNA-DNA complexes.[73] The emissive guanosine analog 6-methylisoxanthopterin (**18**) was exploited in a fluorescence polarization anisotropy study to investigate the structure and dynamics of a single RNA helix namely tetrahymena group I intron ribozyme within nanosecond time scale.[74] The furan-conjugated uridine analog (**19**) is highly microenvironment sensitive and it can be effectively incorporated into nucleic acid by using either chemical or enzymatic methods. For instance, probe (**19**) was enzymatically incorporated into shorter construct of bacterial ribosomal decoding site (A-site) rRNA which was used to monitor the aminoglycoside antibiotic binding studies and also correlation with 2-aminopurine-modified A-site constructs.[75] Benzofuran was tethered at 5-position of uridine to generate highly emissive fluorescence nucleoside analog (**20**). This probe, when selectively incorporated into one of the loop residues of human telomeric DNA and TERRA ONs, not only distinguished between the quadruplex versus duplex, but also discriminated various topologies by changes in the fluorescence.[76] Diederichsen *et al*. reported an isomorphic guanosine analog (**21**) by attaching a vinyl group at the 8-position of guanine. This fluorescent nucleoside analog has propensity to acquire both *syn* and *anti* conformation, which makes it amenable to study the distinct topologies of G-quadruplexes.[77] Recently, Kool and co-workers delineated an oligodeoxyfluoroside (**22**), where base was replaced with a chromophore. This oligodeoxyfluorosides oligomer showed large Stoke shifts, high quantum yield and emission wavelength near to visible region. When pyrene 2′-deoxyribonucleoside was paired against abasic site, the duplex stability was enhanced in comparison to placing it against a natural base. This assumption was attributed to pyrene being accommodate within the double helix and intercalated between the two adjoining base pairs.[78]

1.3.1.4 X-ray crystallography

X-ray crystallography is a valuable technique for 3D-structure determination of nucleic acids and nucleic acid-protein/small molecule complexes. Crystallographic investigation allows structural insights into nucleic acids hydration, conformation, and cation interactions. The trial and error method has been the only way to achieve diffraction quality crystallization. To overcome the phase problem in nucleic acid crystallography, several strategies have been employed, including heavy-atom soaking, halogen derivatization, co-crystallization and indirect Se-protein derivatization.[79] In nucleic acid crystallography, there are three steps involved: crystallization, diffraction data collection and structural calculation from the built 3D-structure. There are common phasing methods available for macromolecular crystallography such as molecular replacement, multiple and single isomorphous replacement (MIR and SIR), multi-wavelength anomalous diffraction (MAD) and single-wavelength anomalous dispersion (SAD), and multiple and single isomorphous replacement with anomalous scattering (MIRAS and SIRAS). To solve the phase problem, MAD and SAD methods have been employed, which use the anomalous scattering property of heavy atoms (e.g., selenium and bromine).[80]

In protein X-ray crystallography, selenomethionine is extensively used in phase and structure determination of various proteins by using multi-wavelength anomalous diffraction (MAD).[81] Investigation of 3D-structure of nucleic acids by indirect protein derivatization (selenomethionine-labeled RNA-binding protein) is seemingly labor-intensive. Therefore, there is a need to develop methods to directly incorporate selenium covalently into nucleic acids for phasing and structural studies. Bromine derivatization (Br K edge = 0.9202 Å) of nucleic acids was used to determine the phase problem,[82] Additionally, halogens have limited derivatization sites on nucleic acids, commonly at 8-position of purines and the 5-position of pyrimidines. It was found that halogenated oligonucleotides are photo-labile and X-ray induced dehalogenation happens when exposed to X-ray radiation.[83] Selenium modified ONs are stable under intense X-ray irradiation as compared to bromine derivatized ONs. Interestingly, the anomalous diffraction signals from selenium and bromine are relatively closer, thus the phasing power of both selenium and bromine are comparable.[84] Also selenium substitution does not cause structural perturbations (Figure 9).[85] Since selenium atom enables phase determination for nucleic acids of up to 30 nucleotides length, it has been used extensively for MAD experiments.[86]



23 24 25 26 27

Figure 9. Nucleic acids with selenium-modification at the nucleobase, ribose and the phosphate backbone.

Huang and co-workers have replaced the 2-position oxygen atom of thymidine by selenium to generate a probe (**23**). Insertion of selenium atom at the 2-position can cause enhanced electronic and steric effects, which promotes its ability to distinguish between natural T-A pairing and T-G wobble pairing. The crystallographic results imply 2-Se-T substitution strengthens the T-A base pair fidelity by decreasing the T-G and T-C mispairs. In addition to that, selenium substitution did not change the complementary T-A base pairing and the structure.[87] Huang *et al.* inserted a selenium atom between the 5-methyl group and C5 carbon of thymidine to give compound (**24**). The biophysical analysis revealed 5-Se-thymidine, methyl and phosphate groups are involved in hydrogen bonds (CH---O---PO_3) and Se-modification did not alter the native DNA structure. This distinct methyl and phosphate interactions can be used in developing inhibitors for the backbone digestion.[88] Huang and co-workers reported selenium containing probes (6-Se-2′-deoxyguanosine (**25**) and Se-G DNAs) that are yellow in colour with UV absorption at 360 nm. They crystallized a ternary complex comprising Se-G-DNA, RNA and RNase H, and have shown that the structure of native and Se-modified duplexes is nearly identical. Interestingly, base pairing between Se-G and C is resemble to native G-C pair, and the selenium atom involves in a Se-mediated hydrogen bonding.[89] Egli, Huang and co-workers inserted selenium to the 2′-position of uridine (**26**) as methylseleno functionality. The sugar puckering was locked into the 2′-*exo* conformation by bulky 2′-Se modification which promoted structural rigidity of the Se-derivatized DNAs and thereby help better molecular packing and crystal growth than native DNAs.[84]

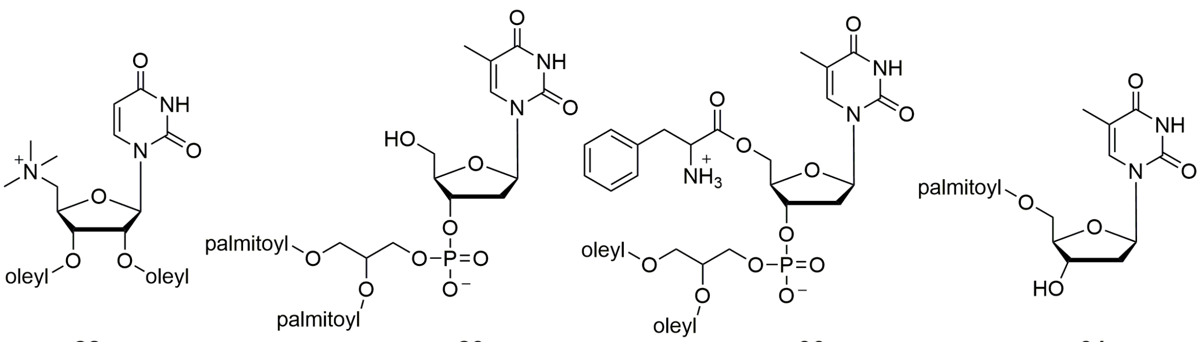
Egli's group synthesized a DNA ON with selenium at the non-bridging phosphate backbone (Figure 9). The phosphoroselenoate (**27**) modified DNA ON was found to be stable under the crystallographic conditions and was further used for MAD phasing for hexameric DNA duplex.[90]

1.4 Nucleoside derivatives as supramolecular synthons

Nucleic acids have received much attention as unique materials for constructing nanoscale structures.[91] Recently, DNA nanostructures have been extensively utilized as functional biomaterial for various applications,[92] such as multiplexed diagnosis,[93] gene and drug delivery,[94] and molecular patterning[95]. Polymers or long-chain alkyl groups conjugated to oligonucleotide scaffolds are used for gene and drug delivery applications.[96] However, the major impediment of nucleic acid based nanotechnology is scalability and high error rate in the self-assembly. In order to overcome above limitations, nucleic acid components (base, sugar, phosphate) are generally derivatized with hydrophobic moieties or tails, so as to introduce a subtle balance of hydrophobicity for constructing various hierarchical supramolecular assemblies. The nucleobase specific base recognition property has been exploited to design certain supramolecular structures at monomeric level, such as G-quartets,[97] columnar nanotubes,[98] nanoparticles[99] and complementary oligonucleotide-template.[100] Guanosine-borate hydrogel is able to absorb a cationic dye and nucleosides via hydrogen bonding and electrostatic interactions.[101]

Naturally occurring nucleolipids have therapeutic applications because of their antimicrobial, antifungal, antiviral and antitumor properties.[102] Tunicamycins, the first nucleolipid antibiotic isolated from the fermentation broths of *Streptomyces lysosuperificus,* is a potent inhibitor of biosynthesis of oligosaccharide. In addition to therapeutic application, nucleolipids are also involved in the biosynthesis of membrane components like glyceophospholipids, phosphatidylinosides as well as cardiolipins, present both in mammalian and prokaryotic cells. Cytidine diphosphate diacylglycerol (CDP-diacylglycerol), a natural nucleolipid, is important in the biosynthesis of membrane phospholipids.[103] In eukaryotic cells, CDP-diacylglycerol is synthesized from cytidine-5-triphosphate and phosphatidic acid and it acts as an activated energy-rich intermediate.

Amphiphilic compounds containing natural building blocks like nucleosides, peptides, sugars and lipid moieties display interesting self-assembling properties.[104] Nucleolipids are biomolecules, where nucleic acid components are covalently attached to long-chain fatty acids. Most of the nucleolipids are 5′-OH group functionalized with polar head group and hydrophobic long-chain fatty acid attached at 2′-*O*- and 3′-*O*-positions of sugar moiety, thereby mimicking the natural glycerophospholipids, which are important components in cell membrane. The nucleoside based nucleolipids were classified into four different categories depending on the charge: cationic, anionic, zwiterrionic and neutral in nature (Figure 10). They self-assemble to form various supramolecular structures like micelles, bilayers, cylindrical and helical structures.[105] Most of the synthetic nucleolipids were designed to interact with nucleic acids through hydrogen bonding and electrostatic interactions.[106] Nucleolipids, which integrate molecular recognition of the base and self-assembling feature of the lipid chain, produced self-assembled systems for gene delivery. For example, cationic nucleolipid vector that binds DNA through Watson-Crick or Hoogsteeen interactions, performed better at delivering nucleic acids to cells as compared to conventional cationic groups (e.g., choline).



28 **29** **30** **31**

Figure 10. Chemical structures of different synthetic nucleolipids cationic (**28**)[107] anionic (**29**)[108] zwiterrionic (**30**)[109] and non-ionic (**31**)[110] respectively.

Grinstaff *et al.* developed a novel nucleic acid vector based on a cationic uridine nucleolipid (**28**), which interacted with both DNA double helix and single-stranded polyadenylic acid. In the presence of nucleic acid, the cationic nucleolipids formed multi-lamellar structures that resembled lipoplexes (cationic liposome-DNA complex, Figure 10). *In vitro* DNA transfection assay employing a green fluorescent protein-encoded plasmid (pEGFP) revealed the efficacy of the cationic nucleolipid as a good tranfecting agent. The nucleoside moiety on the cationic lipid has profound effect on the lipoplex efficacy and cell toxicity. Collectively, these cationic nucleolipids are potential candidates as transfection reagents for mammalian cells.[107] Barthélémy *et al.* investigated anionic nucleolipids (**29**) made of thymidine 3′-monophosphate and 1,2-diacyl-sn-glycerol moieties for the transport of nucleic acids in the presence of Ca^{2+} ion trigger. At lower concentrations of Ca^{2+}, they could not support the formation of stable nucleolipid/DNA complex and exist only at a concentration of 1 mM Ca^{2+} or higher. However, this concentration is much higher than intracellular calcium (1μM). Therefore, by using this Ca^{2+} gradient variations *in vitro* and *in vivo* provide viable mechanism to release the DNA from the lipoplex inside the cells. Subsequently, the transfection experiment was performed on human embryo kidney (HEK 293) cells which demonstrated that the nucleotide moiety increases the transfection efficacy as compared to the anionic phospholipid.[108] Barthélémy *et al.* reported a new type of zwitterionic nucleolipids (**30**), which was constructed by integrating three biomolecules such as amino acid-nucleotide-lipid (ANLs). The zwitterionic nucleolipids made up of phenylalanine and thymidine and unsaturated dioleyl glycerol lipid moieties. Placing an amino acid into the nucleolipid structure modulated the self-assembly properties and also assisted the formation of multilamellar systems. These multilamellar vesicles enabled not only the encapsulation of hydrophobic drug molecules but also hydrophilic therapeutic molecules (Figure 10).[109] Shinkai *et al.* have reported nonionic thymidine-based nucleolipids, which could form organogels in variety of non-polar solvents. 5′-Palmitoyl thymidine self-assembled through non-covalent interactions and forms organogels. These organogelators (**31**) interacts with complementary oligonucleotide poly(A)/lipid in the gel phase, and causes the opaque gel to turn into transparent gel along with substantial changes in the morphologies (Figure 10).[110]

1.5 Applications of nucleobase/nucleoside-based supramolecular synthons

1.5.1 Biomedical applications

The glycosyl-nucleoside bola-amphiphiles (GNBA) are composed of one hydrophobic chain that is covalently attached at both ends to two hydrophilic moieties.[111] These nucleolipid derivatives have the tendency to form hydrogels, which are amenable to the culture of stem cells (Figure 11). GNBAs contain triazole ring and nucleosides, which are extensively involved in π–π stacking interactions in the supramolecular hydrogelation. The therapeutic utility of the thixotropic GNBA-hydrogel was evaluated using a mouse model injection. Thixotropic behaviour, which indicates that the gel convert into sol under shear stress and then rapidly recovers the gel state on resting. GNBA solution was allowed to form gel in syringe and which when pushed with the piston caused a gel-sol transition that enabled its injection into mouse. Few hours after injection, a skin biopsy was examined, which revealed the presence of gels in mouse tissue. These experiments underscore the injectability of the GNBA gel by exploiting its thixotropic property. To investigate whether GNBA gels could be used as matrices for cell adhesion and proliferation, different cell lines were incubated with GNBA solution at RT under gentle stirring. In case of human mesenchymal stem cells, one week after seeding in GNBA hydrogel, most of the cells appeared on the gel fibrous network with cell-divisions and fibro-blastoid morphology were observed.[112]

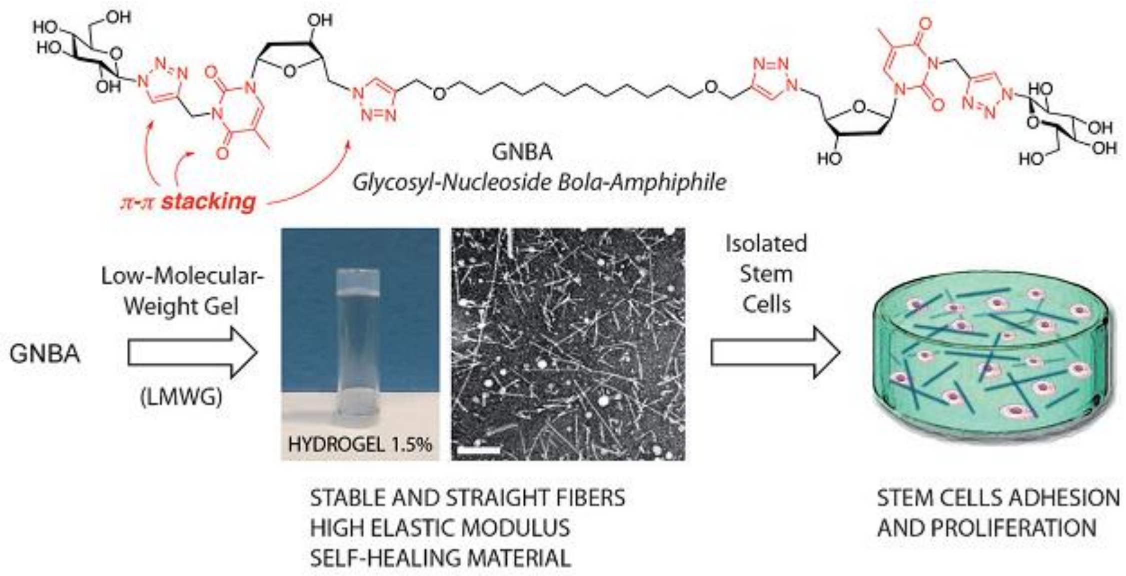


Figure 11. Formation of the supramolecular hydrogel, amenable for the growth and culture of stem cells. Nucleolipids (GNBA) self-assemble through π-π stacking interactions to form stable fibers. The resulting self-healing hydrogels promote the adhesion and proliferation of stem cells. Above figure is adopted from ref. 112.

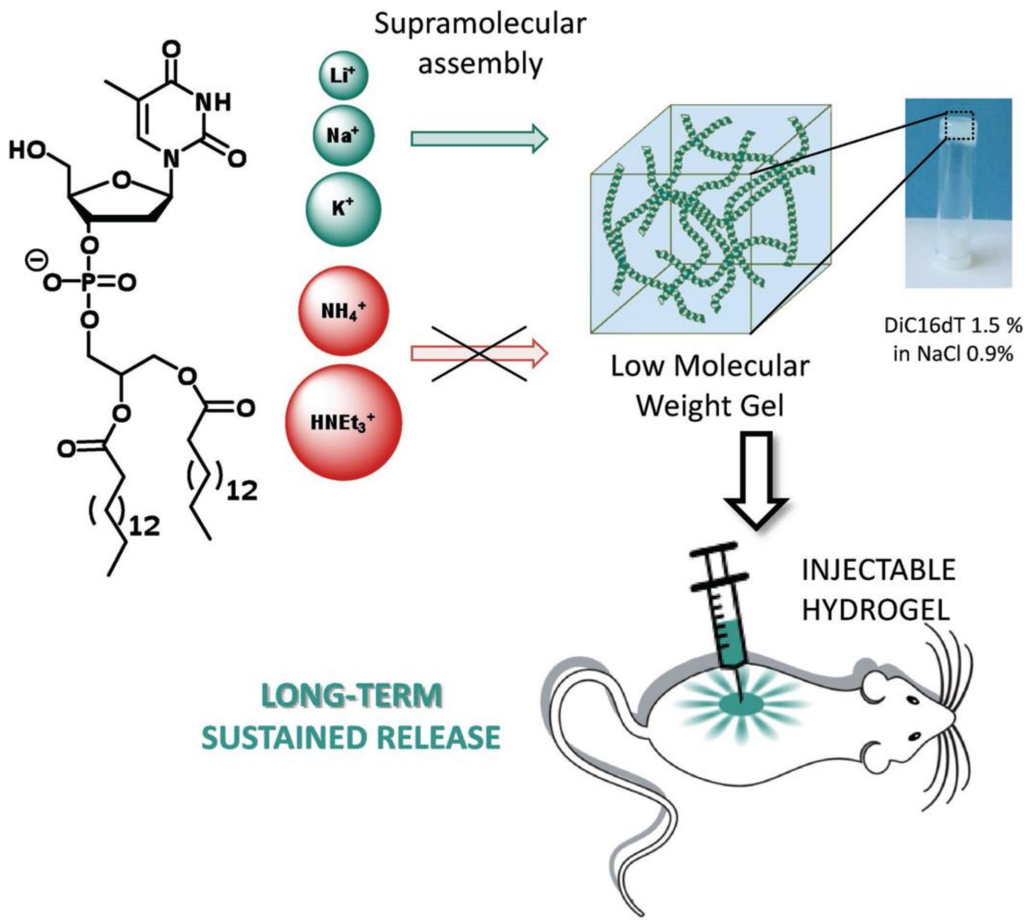


Figure 12. Structure of the nucleolipid and different cations. Schematic illustration of the supramolecular fiber and their gel network observed in the presence of monovalent cations. These hydrogels are suitable for *in vivo* injection and implantation (biocompatibility, biostability adequate mechanical features, etc.). Above schematic illustration is adopted from ref. 113.

The thymidine based nucleotide lipid hybrid possessing a thymidine head group and 1,2-dipalmitoyl-glycerol phosphate (diC16dT) moiety paired with monovalent cations (Li^+, Na^+, K^+, NH_4^+ and $HNEt_3^+$) spontaneously assembles into supramolecular hydrogel (Figure 12). However, in the presence of larger cations such as ammonium and triethyl ammonium, it could not support hydrogelation. The self-assembling process not only depended on nucleolipid hydrophobicity and structure, but also the counterion involved. Hydrogel of diC16T/Na^+ are stable for several weeks upon subcutaneous injection into mice, and remained at the specified location of injection without any adverse immune response. Hydrogel loaded with different model molecules of different molecular weights show long-term sustained release *in vitro*. They also have capability to deliver proteins into the bloodstream *in vivo*. The nucleolipid hydrogels are useful in controlling the *in vivo* release kinetics of small molecules and proteins, and hence, could potentially complement polymer-based drug release systems.[113]

Conventional phospholipids have a covalent linkage between the phospholipid head group and tail. Wang *et al*. established a new type of phospholipids by exploiting non-covalent interactions such as hydrogen bonding and hydrophobic effect to connect the uridine-functionalized head group and adenosine-functionalized tails (Figure 13).[114] As compared to regular phospholipid system, these supramolecular phospholipids are more advantageous due to their good responsiveness and facile preparation. Upon mixing of the two nucleoside phospholipids, it self-assembled into liposome-like bilayer nano-vesicles in water.[115] It has been known that complementary hydrogen bonding interactions among the nucleobases are strong, directional and responsive to acidic pH.[116] Therefore, the liposomes composed from uridine phospholipid and adenosine nucleolipid, shows sensitivity towards acidic stimuli. This property is highly useful in drug delivery vehicles for treatment of various types of diseases. In addition to this, the doxorubicin-loaded nucleolipid based liposomes displayed higher anticancer efficacy over conventional liposome.

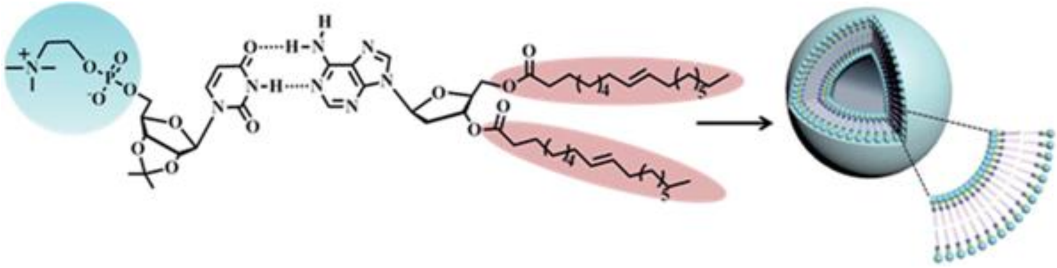


Figure 13. Schematic illustration of a supramolecular liposome-like bilayer formation through self-assembled from the adenosine nucleolipid:uridine phospholipids in aqueous media. This figure is adopted from ref. 114.

1.5.2 Materials applications

Gazit and co-workers developed the dimers of peptide nucleic acid (PNA), which can self-assemble to form ordered nanostructures with unusual physical properties.[117] They have synthesised 16 possible di- PNA building blocks with combination of the adenine (A), cytosine (C), guanine (G) and thymine (T) bases, and tested their self-assembling processes. Among them only three di-PNA (CG, GC and GG) could support the formation of one dimensional nanofiber and spheroids (Figure 14). These observations suggested that GC base pairing was required to form self-assembled structures. The crystal structure of GC-PNA, shows guanine and cytosine in each molecule involve in π–π stacking interactions. Furthermore, each molecule interacts with adjacent cytosine and guanine through hydrogen bonding, which is resembled to regular Watson-Crick duplexes. These assemblies also exhibit wide range of excitation-dependent emission in the visible region and optical properties like voltage-dependent electroluminescence.

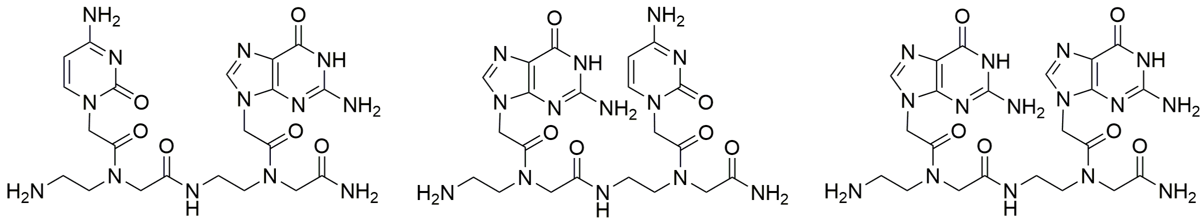


Figure 14. Chemical structures of the three nanostructures forming di-PNAs (CG, GC and GG).

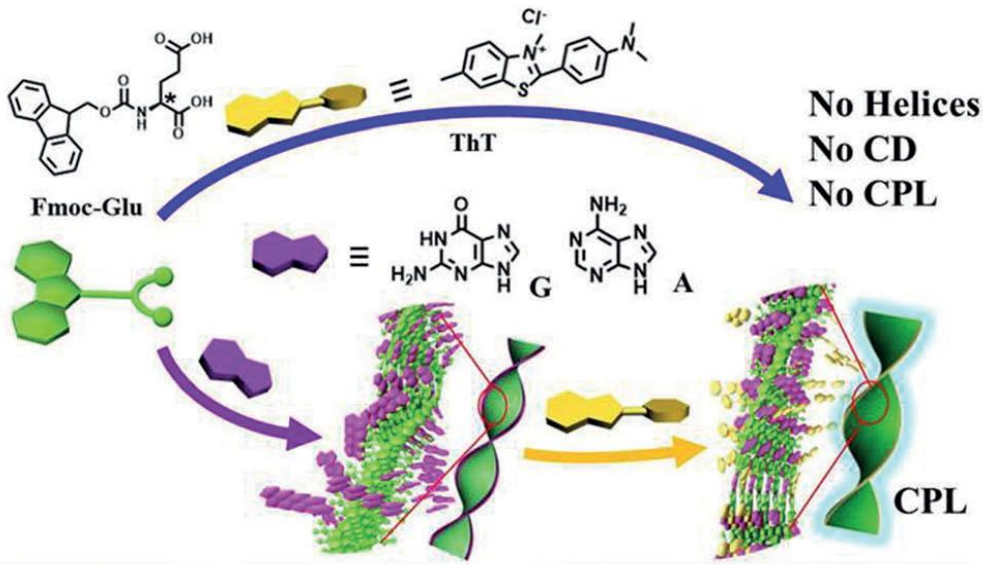


Figure 15. Schematic representation and chemical structures of Fmoc-Glu, purine bases (A and G) and achiral thioflavin T (ThT). Here, achiral purine nucleobases transfer the chirality from Fmoc-Glu to achiral ThT. Fmoc-Glu self-assemble in the presence of purine base to promote the helical structure, upon addition of ThT, they are further converted into helices. But in the absence of A or G no chiral assembly are obtained. (Fmoc-Glu = N-(9-fluorenylmethoxycarbonyl) (Fmoc)-protected glutamic acid, CPL = circular-polarized fluorescence). This schematic illustration is adopted form ref. 118.

Liu and co-workers have investigated chiral amino acid (Fmoc-Glu), achiral purine nucleobases (A, G) and dye molecule (thioflavin T), their roles in multicomponent chiral self-assembly processes (Figure 15). The Fmoc-Glu did not support helical-assembly and hydrogelation, but upon addition of purine bases promote the formation of hydrogel. In contrast, pyrimidine bases (C and T) cannot induce Fmoc-Glu either into chiral-assembly or hydrogelation. It implies that purine bases are effectively involved in π–π stacking interactions with Fmoc-Glu originating the chiral structures as compared to pyrimidine bases. Furthermore, when fluorescent achiral dye (ThT) molecule mixed with two-component assemblies, it also exhibit helicity and subsequently transfers the chirality to thioflavin T (ThT), thus causing a prominent circular-polarized fluorescence (CPL). These results indicating that the helical structures served as a matrix and purine bases acted as bridge to transfer the chirality from Fmoc-Glu to ThT.[118] Therefore purine nucleobases plays an important role in chiral self-assembly and emanating the emissive hydrogel with CPL property.

The anionic guanosine-borate (GB) hydrogel is used for separation of two different dyes from bulk solution. Since the GB hydrogel has anionic borate ester and stacking faces of the G4-tetrad, it would facilitate the non-covalent binding.[101] Here, the cationic methylene blue serves as G-quadruplex ligand[119] and is selectively absorbed into the gel, unlike the non-planar, anionic dye rose bengal. As shown the original colourless gel turned bluish around the edges (Figure 16), and subsequently the dye diffused all over the gel. After 24 h, the gel was blue, whereas the solution remained pink in colour, which implies that GB hydrogel is selectively absorbing cationic dye (MB) as compared to anionic dye (RB). It was suggested that the selectivity for MB could mainly be due to favorable electrostatic interactions between MB and anionic borate, and π–π stacking interactions with the G4-tetrad.

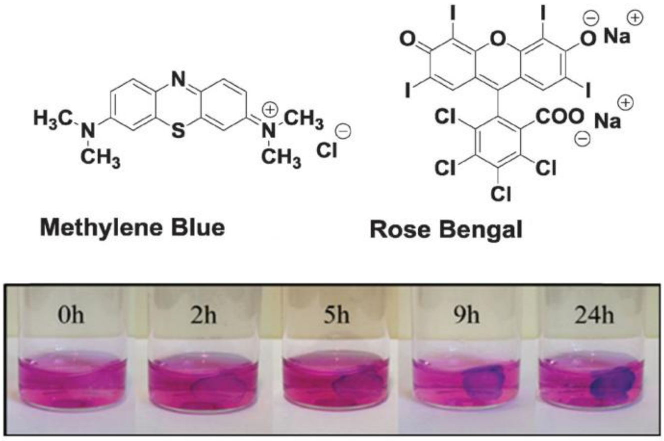


Figure 16. The anionic guanosine-borate (GB) hydrogel selectively absorbs and separates the cationic dye methylene blue (MB) from mixture of anionic rose bengal (RB). This figure is taken from ref. 101.

1.6 Research statement

It is apparent from these studies that when carefully modified, nucleosides can be used as probes for investigating nucleic acid structure and function or as therapeutic tool or can be used to build self-assembling systems or biomedical applications. The limitations of traditional nucleic acid probe design strategy and unexplored potential of nucleoside supramolecular synthons in materials applications inspired us to design new types of probes and materials, which would complement existing tools and resources to (i) study nucleic acids effectively and (ii) construct architectures with tunable and sensing properties.

Objectives of the thesis: The first part of the thesis describes the use of an innovative approach to study nucleic acids by using novel conformation-sensitive dual-purpose nucleoside analogs equipped with a fluorophore and X-ray crystallography label. Such analogs incorporated into target ON sequences enabled the probing of the structure and recognition properties concurrently in real time by fluorescence and in 3D by X-ray crystallography. The second part of the thesis describes the development of nucleoside-fatty acid hybrids (nucleolipids), which show interesting self-assembling behaviour. These synthons show hierarchical self-assembling process with different morphologies, surface property and stimuli responsiveness. We expect that this simple approach of probe and supramolecular synthon design could lead to fabrication of smart materials and probes.

a) Dual-purpose selenophene-modified nucleoside analog probes: Simple physical organic concepts and literature precedents was used to assemble the dual-purpose nucleoside probes. It is known from others and our studies that attaching heterocyclic rings on to purine and pyrimidine bases can generate environment-sensitive nucleosides. In addition, we took advantage of the excellent anomalous X-ray dispersion property of Se atom in assembling the probes. The dual-purpose nucleoside analogs, containing a fluorophore and X-ray anomalous scattering label (Se atom), are derived by attaching selenophene at the 5-position of 2′-deoxyuridine (^{Se}dU) and uridine (^{Se}U, Figure 17). An important aspect of this probe design is that the fluorophore (heterocycle core), and heavy atom (selenium) are part of the same electronic system. Hence, when the nucleoside is incorporated into target nucleic acid, the individual labels will experience similar environment and the outcome from individual spectroscopy technique could be effectively correlated with each other. The nucleoside analogs incorporated into DNA and RNA ONs are minimally perturbing and were found to be sensitive to conformational changes. These features of the nucleoside probe enabled a comprehensive understanding of the structure and recognition properties of the bacterial ribosomal decoding site RNA and G-quadruplex-forming human telomeric repeats under equilibrium condition by fluorescence and in 3D by X-ray crystallography. The selenophene-modified nucleoside analogs represent the first example of a dual purpose probe, which provides this level of understanding.

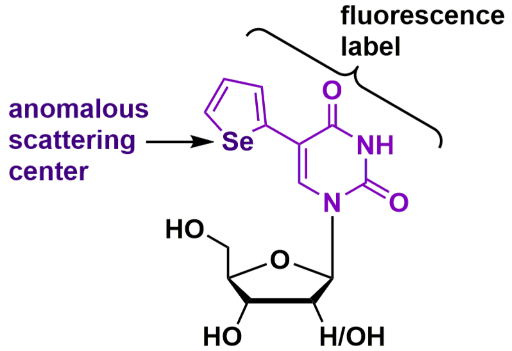


Figure 17. Design of dual-purpose nucleoside probe for studying nucleic acids by using fluorescence and X-ray crystallography.

b) Functionalized nucleolipid supramolecular synthons: Responsive nucleolipid synthons were constructed by using environment-sensitivity fluorescent nucleoside analogs, based on 5-(benzofuran-2-yl)uracil and 5-(benzo[*b*]thiophen-2-yl)uracil cores, as the head group and fatty acids, attached to the ribose sugar, as the lipophilic group (Figure 18). Depending on the modification on the nucleobase and sugar residue, the nucleolipids form gels, which are highly responsive to external stimuli (physical and chemical) and show signs of aggregation-induced enhanced emission. Further, simple nucleolipid synthons prepared by attaching fatty acids to the sugar hydroxyl groups of native nucleosides also displayed interesting self-assembling behaviour. The gel thus formed showed self-sorting property and surface tunability.

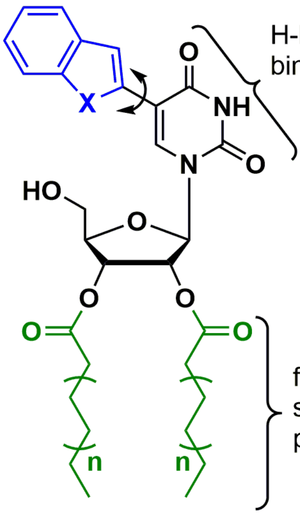


Figure 18. Design of self-assembling and responsive nucleoside-fatty acid hybrids (nucleolipids).

1.7 References:

1 (a) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737–738. (b) Crick, F. *Nature* **1970**, 272, 561–563. (c) Dickerson, R. E.; Drew, H. R.; Conner, B. N.; Wing, R. M.; Fratini, A. V.; Mary L. Kopka, M. L. *Scienec* **1982**, *216*, 475–485. (d) Altman, S.; Baer, M.; Guerrier-Takada, C.; Vioque, A. *Trends Biochem. Sci*. **1986**, *11*, 515–518. (e) Sharp, P. A. *Cell* **2009**, *136*, 577–580.

2 (a) Hermann, T.; Patel, D. J. *Structure* **2000**, *8*, R47–R54. (b) Leulliot, N.; Varani, G. *Biochemistry* **2001**, *40*, 7947–7956. (c) Mandal, M.; Breaker, R. R. *Nat. Rev. Mol. Cell Biol*. **2004**, *5*, 451–463

3 Langridge, R.; Wilson, H. R.; Hooper, C. W.; Wilkins, M. H. F.; Hamilton, L. D. *J. Mol. Biol.* **1960**, *2*, 19–37.

4 Fuller, W.; Wilkins, M. H. F.; Wilson, H. R.; Hamilton, L. D.; Arnott, S. *J. Mol. Biol.* **1965**, *12*, 60–76.

5 Wang, A. H.; Quigley, G. J.; Kolpak, F. J.; Crawford, J. L.; van Boom, J. H.; van der Marel, G.; Rich, A. *Nature* **1979**, *282*, 680–686.

6 (a) Turner, D. H. *Curr. Opin. Struct. Biol*. **1996**, *6*, 299–304. (b) Tian, B.; Bevilacqua, P. C.; Diegelman-Parente, A.; Mathews, M. B. *Nat. Rev. Mol. Cell Biol*. **2004**, *5*, 1013–1023.

7 (a) Brion, P.; Westhof, E. *Annu. Rev. Biophys. Biomol. Struct*. **1997**, *26*, 113–137. (b) Tinoco, I. Jr.; Bustamante, C. *J. Mol. Biol*. **1999**, *293*, 271–281.

8 David-Eden, H.; Mankin, A. S.; Mandel-Gutfreund, Y. *Nucleic Acids Res*. **2010**, *38*, 5982–5994.

9 (a) DeRose, V. J. *Curr. Opin.Struct. Biol*. **2003**, *13*, 317–324. (b) Asseline, U. *Curr. Org. Chem*. **2006**, *10*, 491–518. (c) Piton, N.; Mu, Y.; Stock, G.; Prisner, T. F.; Schiemann, O.; Engels, J. W. *Nucleic Acids Res*. **2007**, *35*, 3128–3143. (d) Olieric, V.; Rieder, U.; Lang, K.; Serganov, A.; Schulze-Briese, C.; Micura, R.; Dumas, P.; Ennifar, E. *RNA* **2009**, *15*, 707–715. (e) Juskowiak, B. *Anal. Bioanal. Chem*. **2011**, *399*, 3157–3176. (f) Bardaro Jr, M. F.; Varani, G. *WIREs RNA* **2012**, *3*, 122–132. (g) Serganov, A.; Patel, D. J. *Curr. Opin. Struct. Biol*. **2012**, *22*, 279–286.

10 Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469–480.

11 Phelps, K.; Morris, A.; Beal, P. A. *ACS Chem Biol*. **2012**, *7*, 100–109.

12 Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem*. **2013**, *53*, 366–378.

13 (a) Krishnan, Y.; Simmel, F. C. *Angew. Chem., Int. Ed.* **2011**, *50*, 3124–3156. (b) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. *Science* **2015**, *347*, 1260901.

14 (a) Tian, Y.; Mao, C. *J. Am. Chem. Soc*. **2004**, *126*, 11410–11411. (b) Ackermann, D.; Schmidt, T. L.; Hannam, J. S.; Purohit, C. S.; Heckel, A.; Famulok, M. *Nat. Nanotechnol*. **2010**, *5*, 436–442.

15 (a) Erben, C. M.; Goodman, R. P.; Turberfield, A. J. *Angew. Chem., Int. Ed*. **2006**, *45*, 7414–7417.

16 (a) Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335*, 831–834. (b) Guo, P.; Haque, F. *RNA Nanotechnology and Therapeutics*, CRS Press, Taylor and Francis Group, **2014**.

17 (a) Wells, R. D.; Dere, R.; Hebert, M.L.; Napierala, M.; Son, L. S. *Nucleic Acids Res*. **2005**, *33*, 3785–3798. (b) Choi, J.; Majima, T. *Chem. Soc. Rev*. **2011**, *40*, 5893–5909.

18 Jain, A.; Wang, G.; Vasquez, K. M. *Biochimie* **2008**, *90*, 1117–1130.

19 Bacolla, A.; Collins, J. R.; Gold, B.; Chuzhanova, N.; Yi, M.; Stephens, R. M.; Stefanov, S.; Olsh, A.; Jakupciak, J. P.; Dean, M.; Lempicki, R. A.; Cooper, D. N.; Wells, R. D. *Nucleic Acids Res*. **2006**, *34*, 2663–2675.

20 Seidman, M. M.; Glazer, P. M. *J. Clin. Invest*. **2003**, *112*, 487–494.

21 Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U. S. A.* **1962**, *48*, 2013–2018.

22 Lipps, H. J.; Rhodes, D. *Trends Cell Biol*. **2009**, 19, 414–422.

23 Gros, J.; Rosu, F.; Amrane, S.; De Cian, A.; Gabelica, V.; Lacroix, L.; Mergny, J.-L. *Nucleic Acids Res*. **2007**, *35*, 3064–3075.

24 Huppert, J. L. *FEBS J*. **2010**, *277*, 3452–3458.

25 Smargiasso, N,; Rosu, F.; Hsia, W.; Colson, P.; Baker, E. S.; Bowers, M.T.; De Pauw, E.; Gabelica, V. *J. Am. Chem. Soc*. **2008**, *130*, 10208–10216.

26 (a) Williamson, J. R.; Raghuraman, M. K.; Cech, T. R. *Cell* **1989**, *59*, 871–880. (b) Wong, H. M.; Payet, L.; Huppert, J. L. *Curr. Opin. Mol. Ther*. **2009**, *11*, 146–155.

27 (a) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res*. **2006**, *34*, 5402–5415. (b) Hazel, P.; Parkinson, G. N.; Neidle, S. *Nucleic Acids Res*. **2006**, *34*, 2117–2127.

28 (a) Hardin, C. C.; Perry, A. G.; White, K. *Biopolymers* **2000**, *56*, 147–194. (b) Bugaut, A.; Balasubramanian, S. *Biochemistry* **2008**, *47*, 689–697. (c) Guédin, A.; Gros, J,; Alberti, P.; Mergny, J.-L. *Nucleic Acids Res*. **2010**, *38*, 7858–7868.

29 Patel, D. J.; Phan, A. T.; Kuryavy, V. *Nucleic Acids Res*. **2007**, *35*, 7429–7455.

30 Neidle, S.; Parkinson, G. H. *Curr. Opin. Struct. Biol*. **2003**, *13*, 275–283.

31 Dexheimer, T. S.; Fry, M.; Hurley, L. H. *Quadruplex Nucleic Acids*, RSC Publishing, Cambridge, UK, **2006**, 180–207.

32 Rhodes, D.; Lipps, H. J. *Nucleic Acids Res*. **2015**, *43*, 8627–8637.

33 Brooks, T. A.; Kendrick, S.; Hurley, L. *FEBS J*. **2010**, *277*, 3459–3469.

34 (a) Miyoshi, D.; Karimata, H.; Sugimoto, N. *J. Am. Chem. Soc*. **2006**, *128*, 7957–7963. (b) Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Mol. BioSyst*. **2010**, *6*, 580–586.

35 Dong, Y.; Yang, Z.; Liu, D. *Acc. Chem. Res*. **2014**, *47*, 1853–1860.

36 Kendrick, S.; Kang, H.-J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H. *J. Am. Chem. Soc*. **2014**, *136*, 4161–4171.

37 Holbrook, S. R. *Annu. Rev. Biophys*. **2008**, *37*, 445–464.

38 Davis, J. H.; Tonelli, M.; Scott, L. G.; Jaeger, L.; Williamson, J. R.; Butcher, S. E. *J. Mol. Biol*. **2005**, *351*, 371–382.

39 (a) Blanchard, S. C. *Curr. Opin. Struct. Biol*. **2009**, *19*, 103–109. (b) Sigurdsson, S. T. *Pure Appl. Chem*. **2011**, *83*, 677–686.

40 (a) Latham, M. P.; Brown, D. J.; McCallum, S. A.; Pardi, A. *ChemBioChem* **2005**, *6*, 1492–1505. (b) Nelissen, F. H.; van Gammeren, A. J.; Tessari, M.; Girard, F. C.; Heus, H. A.; Wijmenga, S. S. *Nucleic Acids Res*. **2008**, *36*, e89.

41 (a) Fürtig, B.; Richter, C.; Wöhnert, J.; Schwalbe, H. *ChemBioChem* **2003**, *4*, 936–962. (b) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle,S.; Patel, D. J. *J. Am. Chem. Soc*. **2007**, *129*, 4386–4392. (c) Rinnenthal, J.; Buck, J.; Ferner, J.; Wacker, A.; Fürtig, B.; Schwalbe, H. *Acc. Chem. Res*. **2011**, *44*, 1292–1301. (d) Bothe, J. R.; Nikolova, E. N.; Eichhorn, C. D.; Jeetender Chugh, J.; Hansen, A. L.; Al-Hashimi, H. M. *Nat. Methods* **2011**, *8*, 919–931.

42 Schultheisz, H. L.; Szymczyna, B. R.; Scott, L. G.; Williamson, J. R. *ACS Chem. Biol*. **2008**, *3*, 499–511.

43 Wenter, P.; Bodenhausen, G.; Dittmer, J.; Pitsch, S. *J. Am. Chem. Soc*. **2006**, 128, 7579–7587.

44 Rieder, U.; Lang, K.; Kreutz, C.; Polacek, N.; Micura, R. *ChemBioChem* **2009**, *10*, 1141–1144.

45 Hennig, M.; Scott, L. G.; Sperling, E.; Bermel, W.; Williamson, J. R. *J. Am. Chem. Soc*. **2007**, *129*, 14911–14921.

46 Suydam, I. T.; Strobel, S. A. *J. Am. Chem. Soc*. **2008**, *130*, 13639–13648.

47 (a) Qin, P. Z.; Dieckmann, T. *Curr. Opin. Struc. Biol*. **2004**, *14*, 350–359. (b) Schiemann, O.; Prisner, T. F. *Q. Rev. Biophys*. **2007**, *40*, 1–53. (c) Sowa, G. Z.; Qin, P. Z. *Prog. Nucleic Acid Res*. **2008**, *82*, 147–197. (d) Zhang, X.; Cekan, P.; Sigurdsson, S. T.; Qin, P. *Methods Enzymol*. **2009**, *469*, 303–328. (e) Shelke, S. A.; Sigurdsson, S. T. *Eur. J. Org. Chem*. **2012**, 2291–2301.

48 Strobel, O. K.; Kryak, D. D.; Bobst, E. V.; Bobst, A. M. *Bioconjugate Chem.* **1991**, *2*, 89–95.

49 Gannett, P. M.; Darian, E.; Powell, J. H.; Johnson, E.M. *Syn. Commun*. 2001, *31*, 2137–2141.

50 Gannett, P. M.; Darian, E.; Powell, J.; E. M. Johnson, E. M. 2nd.; Mundoma, C.; Greenbaum, N. L.; Ramsey, C. M.; Dalal, N. S.; Budil, D. E. *Nucleic Acids Res.* **2002**, *30*, 5328–5337.

51 Singh, V.; Azarkh, M.; Exner, T. E.; Hartig, J. S.; Drescher, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 9728–9730.

52 Cekan, P.; Sigurdsson, S. T. *J. Am. Chem. Soc.* **2009**, *131*, 18054–18056.

53 Edwards, T. E.; Robinson, B. H.; Sigurdsson, S. T. *Chem. Biol*. **2005**, *12*, 329–337.

54 Wunnicke, D.; Strohbach, D.; Weigand, J. E.; Appel, B.; Feresin, E.; Suess, B.; Muller, S.; Steinhoff, H. J. *RNA* **2011**, *17*, 182–188.

55 Schiemann, O.; Weber, A.; Edwards, T. E.; Prisner, T. F.; Sigurdsson, S. T. *J. Am. Chem. Soc*. **2003**, *125*, 3434–3435.

56 Höbartner, C.; Sicoli, G.; Wachowius, F.; Gophane, D. B.; Sigurdsson, S. T. *J. Org. Chem*. **2012**, *77*, 7749–7754.

57 Tkach, I.; Pornsuwan, S.; Höbartner, C.; Wachowius, F.; Sigurdsson, S. T.; Baranova, T. Y.; Diederichsen, U.; Sicoli, G.; Bennati, M. *Phys Chem Chem Phys*. **2013**, *15*, 3433–3437.

58 Kerzhner, M.; Abdullin, D.; Więcek, J.; Matsuoka, H.; Hagelueken, G.; Schiemann, O.; Famulok, M. *Chem. Eur. J.* **2016**, *22*, 12113–12121.

59 Ramos, A.; Varani, G. *J. Am. Chem. Soc.* **1998**, *120*, 10992–10993.

60 Borbat, P. P.; Davis, J. H.; Butcher, S. E,; Freed, J. H. *J. Am. Chem. Soc.* **2004**, *126*, 7746–7747.

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

62 (a) Sinkeldam, R. W.; Hopkins, P. A.; Tor, Y. *ChemPhysChem* **2012**, *13*, 3350–3356. (b) Rovira, A. R.; Fin, A.; Tor, Y. *J. Am. Chem. Soc*. **2015**, *137*, 14602–14605.

63 Wilson, J. N.; Kool. E. T. *Org. Biomol. Chem.* **2006**, *4*, 4265–4274.

64 Hawkins, M. E. *Cell. Biochem. Biophys.* **2001**, *34*, 257–281.

65 Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. *Chem. Sci*. **2015**, *6*, 3314–3320.

66 Cobb, S. L.; Murphy, C. D. *J. Fluorine Chem*. **2009**, *130*, 132–143.

67 Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem*. **1969**, *244*, 1228–1237.

68 Rachofsky, E. L.; Osman, R.; Alexander Ross, J. B. *Biochemistry* **2001**, *40*, 946–956.

69 Menger, M.; Eckstein, F.; Pörschke, D. *Nucleic Acids Res*. **2000**, *28*, 4428–4434.

70 Bradrick, T. D.; Marino, J. P. *RNA* **2004**, *10*, 1459–1468.

71 Srivatsan, S. G.; Weizman, H.; Tor, Y. *Org. Biomol. Chem*. **2008**, *6*, 1334–1338.

72 Srivatsan, S. G.; Greco, N. J.; Tor, Y. *Angew. Chem., Int. Ed.* **2008**, *47*, 6661–6665.

73 Tinsley, R. A.; Walter, N. G. *RNA* **2006**, *12*, 522–529.

74 Shi, X.; Mollova, E. T.; Pljevaljcic, G.; Millar, D. P.; Herschlag, D. *J. Am. Chem. Soc*. **2009**, *131*, 9571–9578.

75 Srivatsan, S.G.; Tor, Y. *J. Am. Chem. Soc*. **2007**, *129*, 2044–2053.

76 Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res*. **2015**, *43*, e149.

77 Nadler, A.; Strohmeier, J.; Diederichsen, U. *Angew. Chem., Int. Ed.* **2011**, *50*, 5392–5396.

78 Teo, Y. N.; Kool, E. T. *Chem. Rev.* **2012**, *112*, 4221–4245.

79 (a) Ferré-D'Amaré, A. R.; Zhou, K.; Doudna, J. A. *Nature* **1998**, *395*, 567–574. (b) Xiong, Y.; Sundaralingam, M. *Nucleic Acids Res*. **2000**, *28*, 2171–2176. (c) Xiao, H.; Murakami, H.; Suga, H.; Ferré-D'Amaré, A. R. *Nature* **2008**, *454,* 358–361.

80 Boggon, T. J.; Shapiro, L. *Structure* **2000**, *8,* R143–149.

81 Hendrickson, W. A. *Science* **1991**, *254,* 51–58.

82 (a) Egli, M. *Curr. Opin. Chem. Biol*. **2004**, *8*, 580–591. (b) Jiang, J.; Sheng, J.; Carrasco, N.; Huang, Z. *Nucleic Acids Res*. **2007**, *35*, 477–485.

83 Ennifar, E.; Carpentier, P.; Ferrer, J. L.; Walter, P.; Dumas, P. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2002**, *58*, 1262–1268.

84 Teplova, M.; Wilds, C. J.; Wawrzak, Z.; Tereshko, V.; Du, Q.; Carrasco, N.; Huang, Z.; Egli, M. *Biochimie* **2002**, *84*, 849–858.

85 Salon, J.; Sheng, J.; Gan, J.; Huang, Z. *J. Org. Chem*. **2010**, *75*, 637–641.

86 (a) Hendrickson, W. A. *Trends Biochem. Sci*. **2001**, *25*, 637–643. (b) Willis, M. C.; Hicke, B. J.; Uhlenbeck, O. C.; Cech, T. R.; Koch, T. H. *Science* **1993**, *262*, 1255–1257.

87 Hassan, A. E.; Sheng, J.; Zhang, W.; Huang, Z. *J. Am. Chem. Soc*. **2010**, *132*, 2120–2121.

88 Hassan, A. E.; Sheng, J.; Jiang, J.; Zhang, W.; Huang, Z. *Org. Lett*. **2009**, *11*, 2503–2506.

89 Salon, J.; Jiang, J.; Sheng, J.; Gerlits, O.O.; Huang, Z. *Nucleic Acids Res*. **2008**, *36*, 7009–7018.

90 Wilds, C. J.; Pattanayek, R.; Pan, C.; Wawrzak, Z.; Egli, M. *J. Am. Chem. Soc*. **2002**, *124*, 14910–14916.

91 (a) Semeen, N. C. *Nature* **2003**, *421*, 427–431. (b) Yan, H. *Science* **2004**, *306*, 2048–2049.

92 Wang, Z.-G.; Ding, B. *Adv. Mater.* **2013**, *25*, 3905–3914.

93 Lee, J. B.; Roh, Y. H.; Um, S. H.; Funabashi, H.; Cheng, W.; Cha, J. J.; Kiatwuthinon. P.; Muller, D. A.; Luo, D. *Nat. Nanotechnol*. **2009**, *4*, 430–436.

94 (a) Cheng, E.; Xing, Y.; Chen, P.; Yang, Y.; Sun, Y.; Zhou, D.; Xu, L.; Fan, Q.; Liu, D. *Angew. Chem., Int. Ed.* **2009**, *48,* 7660–7663. (b) Lee, H. *et al*. *Nat. Nanotechnol*. **2012**, *7*, 389–393.

95 Han, D.; Pal, S.; Nangreave, J.; Deng, Z.; Liu, Y.; Yan, H. *Science* **2011**, *332*, 342–346.

96 (a) Patwa, A.; Gissot, A.; Bestel, I.; Barthélémy, P. *Chem. Soc. Rev*. **2011**, *40*, 5844–5854. (b) Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I. *J. Am. Chem. Soc*. **2013**, *135*, 6211–6219. (c) Banga, R. J.; Chernyak, N.; Narayan, S. P.; Nguyen, S. T.; Mirkin, C. A. *J. Am. Chem. Soc*. **2014**, *136*, 9866–9869.

97 Rivera, J. M.; Silva-Brenes, D. *Org. Lett*. **2013**, *15*, 2350–2353.

98 Seela, F.; Wiglenda, T.; Rosemeyer, H.; Eickmeier, H.; Reuter, H. *Angew. Chem., Int. Ed*. **2002**, *41*, 603–605.

99 Lanna, C. M.; Lusic, H.; Complo, M.; McIntosh, T. J.; Barthélémy, P.; Grinstaff, M. W. *Acc. Chem. Res*. **2012**, *45*, 1026–1038.

100 Iwaura, R.; Yoshida, K.; Masuda, M.; Ohnishi-Kameyama, M.; Yoshida, M.; Shimizu, T. *Angew. Chem., Int. Ed*. **2003**, *115*, 1039–1042.

101 Peters, G. M.; Skala, L. P.; Plank, T. N.; Hyman, B. J.; Reddy, G. N. M.; Marsh, A.; Brown, S. P.; Davis, J. T. *J. Am. Chem. Soc*. **2014**, *136*, 12596–12599.

102 Rosemeyer, H. *Chem. Biodiversity* **2005**, *2*, 977–1062.

103 Allain, A.; Bourgaux, C.; Couvreur, P. *Nucleic Acids Res*. **2012**, *40*, 1891–1903.

104 (a) Araki, K.; Yoshikawa, I. *Top. Curr. Chem*. **2005**, *256*, 133–165. (b) Peters, G. M.; Davis, J. T. *Chem. Soc. Rev*. **2016**, *45*, 3188–3206.

105 (*a*) Itojima, Y.; Ogawa, Y.; Tsuno, K.; Hands, N.; Yanagawa, H. *Biochemistry* **1992**, *31*, 4757–4765. (*b*) Milani, S.; Bombelli, F. B.; Berti, D.; Baglioni. P. *J. Am. Chem. Soc*. **2007**, *129*, 11664–11665.

106 Berti, D.; Montis, C.; Baqlioni, P. *Soft Matter* **2011**, *7*, 7150–7158.

107 Chabaud, P.; Camplo, M.; Payet, D.; Serin, G.; Moreau, L.; Barthélémy, P.; Grinstaff, M. W. *Bioconjugate Chem*. **2006**, *17*, 466−472.

108 Khiati, S.; Pierre, N.; Andriamanarivo, S.; Grinstaff, M. W.; Arazam, N.; Nallet, F.; Navailles, L.; Barthélémy, P. *Bioconjugate Chem*. **2009**, *20*, 1765–1772.

109 Tonelli, G.; Oumzil, K.; Nallet, F.; Gaillard, C.; Navailles, L.; Barthélémy, P. *Langmuir* **2013**, *29*, 5547−5555.

110 Sugiyasu, K.; Numata, M.; Fujita, N.; Park, S. M.; Yun, Y. J.; Kim, B. H.; Shinkai, S. *Chem. Commun*. **2004**, 1996–1997.

111 Estroff, L. A.; Hamilton, A. D. *Chem. Rev*. **2004**, *104*, 1201–1218.

112 Latxague, L.; Ramin, M. A.; Appavoo, A.; Berto, P.; Maisani, M.; Ehret, C.; Chassande, O.; Barthélémy, P. *Angew. Chem., Int. Ed*. **2015**, *54*, 4517–4521.

113 Ramin, M. A.; Sindhu, K. R.; Appavoo, A.; Oumzil, K.; Grinstaff, M. W.; Chassande, O.; Barthélémy, P. *Adv. Mater.* **2017**, 1605227.

114 Wang, D.; Tu, C.; Su, Y.; Zhang, C.; Greiser, U.; Zhu, X.; Yana, D.; Wang, W. *Chem. Sci*. **2015**, *6*, 3775–3787.

115 (a) Gissot, A.; Camplo, M.; Grinstaff, M. W.; Barthélémy, P. *Org. Biomol. Chem*. **2008**, *6*, 1324–1333. (b) Wang, C.; Wang, Z.; and Zhang, X. *Acc. Chem. Res*. **2012**, *45*,608–618.

116 Li, Y.; Park, T.; Quansah, J. K.; Zimmerman, S. C. *J. Am. Chem. Soc*. **2011**, *133*, 17118–17121.

117 Berger, O.; Adler-Abramovich, L.; Levy-Sakin, M.; Grunwald, A.; Liebes-Peer, Y.; Bachar, M.; Buzhansky, L.; Mossou, E.; Forsyth, V. T.; Schwartz, T.; Ebenstein, Y.; Frolow, F.; Shimon, L. J. W.; Patolsky, F.; Gazit, E. *Nat. Nanotechnol*. **2015**, *10*, 353–360.

118 Deng, M.; Zhang, L.; Jiang, Y.; Liu, M. *Angew. Chem., Int. Ed*. **2016**, *55*, 15062–15066.

119 Rodríguez-Llansola, F.; Miravet, J. F.; Escuder, B. *Chem. Commun*. **2011**, *47*, 4706–4708.

Chapter 2

Probing bacterial ribosomal decoding site RNA structure and antibiotic binding using selenophene-modified dual-purpose nucleoside probe

2.1 Introduction

Biophysical techniques including fluorescence labeling, NMR, EPR and X-ray crystallography have provided valuable information on RNA folding, molecular recognition and function.[1–4] Such investigations mostly involve RNA sequences labeled with appropriate reporters as natural nucleosides are practically non-fluorescent and do not provide intrinsic labels that are suitable for analysis by spectroscopic and diffraction techniques.[5] Minimally perturbing fluorescent nucleoside analogs,[6] isotope- (e.g., ^{13}C, ^{15}N),[7] spin- (e.g., nitroxide)[8] and heavy-atom-labeled (e.g., Br, Se)[9] nucleosides incorporated into oligonucleotides (ONs) have been applied in assays to study structure, dynamics and function of RNA. However, given the multitude and diversity of RNA transcripts, probing structure-function relationships of RNA sequences using established biochemical and biophysical methods remains a major challenge.[10] This is largely due to the lack of chemical probes that can be efficiently deployed in multiple biophysical techniques as almost all studies use the traditional approach of "one label-one technique", wherein a custom labeled RNA suitable for a particular technique is used. Therefore, development of smart chemical tools that can (i) be readily incorporated into RNA ONs and (ii) enable the direct correlation of RNA structure and function in real time and in 3D is highly desired. We concluded that ribonucleoside probes containing multiple labels (e.g., a fluorophore and an anomalous X-ray scattering label) would be advantageous as they would serve as common probes to analyze RNA conformation and function in solution in real time by fluorescence and, concurrently, RNA structure in the solid-state by X-ray crystallography (Figure 1).

We recently introduced a 5-selenophene-modified uridine analog **1** (^{Se}U), which is composed of an environment-sensitive base-modified fluorescent nucleobase and an X-ray crystallography phasing agent (Se atom, Figure 1).[11] Selenium provides key advantages over halogen modification. First, Se gives rise to a strong anomalous scattering signal, which has been widely used in protein crystallography and more recently in nucleic acid analysis.[12] Second, halogen-modified ONs are known to undergo dehalogenation when exposed to X-ray radiation causing failures in phasing.[13] Selenium is the most preferred anomalous scattering atom form phase determination using either multi-wavelength anomalous diffraction (MAD) or single-wavelength anomalous diffraction (SAD). The selenium K absorption edge, which is at 12.658 kev, is well within the energy range of both tunable and fixed wavelength synchrotron radiation source. The anomalous scattering power of selenium is sufficiently good. If data collection is carefully performed, one anomalous signal from one selenium can phase about 1000 atoms. The popularity of selenium as an anomalous scattered has influenced many synchrotron beam lines with fixed wavelength to tune the energy to selenium K absorption edge. This has facilitated SAD experiments at these beam lines.[14]

Ribonucleoside **1** has an emission maximum in the visible region (λ_{em} = 452 nm in water), and exhibits probe-like properties (e.g., solvatochromism and viscochromism). The corresponding ribonucleotide can be effectively incorporated into ONs by enzymatic in vitro transcription[11] and used to monitor RNA-ligand binding by fluorescence. Based on these observations we sought to explore the fluorescent and anomalous X-ray scattering properties of the uridine analog **1** as a common RNA probe in complementing solution and solid-state techniques, namely fluorescence and X-ray crystallography. We chose bacterial A-site RNA as the test system, which is a well-studied target for aminoglycoside antibiotics.[15] In this system, the effect of selenophene modification on the RNA structure and its ability to faithfully report drug binding to A-site in solution and solid state can be validated by direct comparison with previously reported data on A-site aminoglycoside complexes.[16]

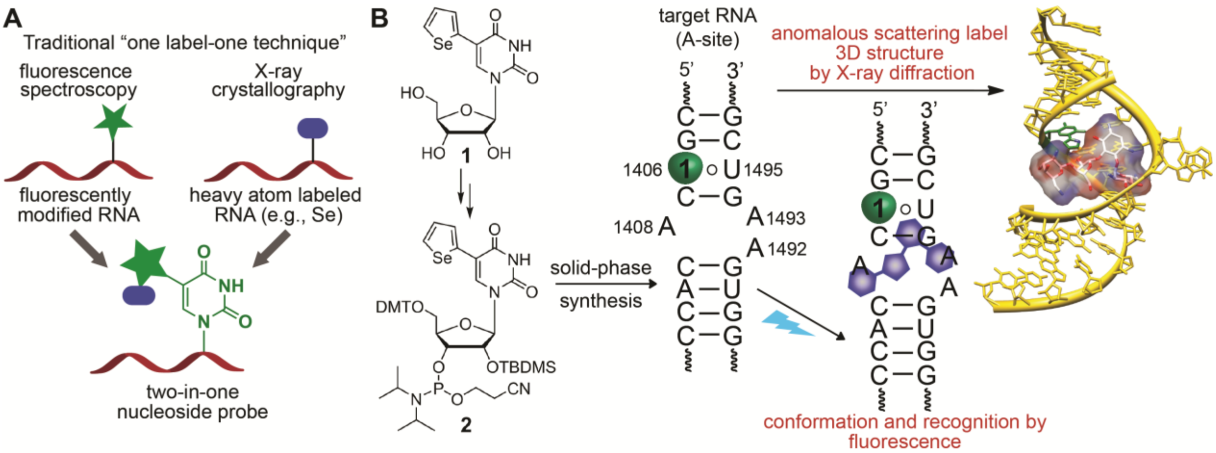


Figure 1. (**A**) Design of "two-in-one" ribonucleoside probe for RNA analysis in solution and solid-state. In this design the fluorophore core and anomalous X-ray scattering Se atom are placed in the same electronic system so that they sense similar environment when incorporated into RNA ONs. (**B**) Chemical structure of 5-selenophene-modified nucleoside probe (**1** or ^{Se}U) and corresponding phosphoramidite substrate (**2**). Incorporation of ^{Se}U into the RNA target (e.g., A-site) would enable simultaneous biophysical investigation of the same RNA construct in solution by fluorescence techniques and in the solid-state by X-ray crystallography.

In this chapter, we describe the first example of a dual-purpose ribonucleoside probe (^{Se}U), based on the 5-(selenophen-2-yl)uracil scaffold, which provides a tool for comprehensive investigations of RNA-drug interaction by fluorescence and X-ray crystallography techniques (Figure 1B). The nucleoside probe was chemically incorporated into the bacterial ribosomal decoding site (A-site), and the environmental sensitivity of the fluorophore was used for monitoring the binding of aminoglycoside antibiotics. X-ray diffraction analysis of A-site RNA crystals containing the modified uridine revealed a prominent diffraction signal from the selenium atom and also provided insight into the structural basis on how the nucleoside probe senses conformational changes in RNA during ligand binding.

2.2 Results and Discussion

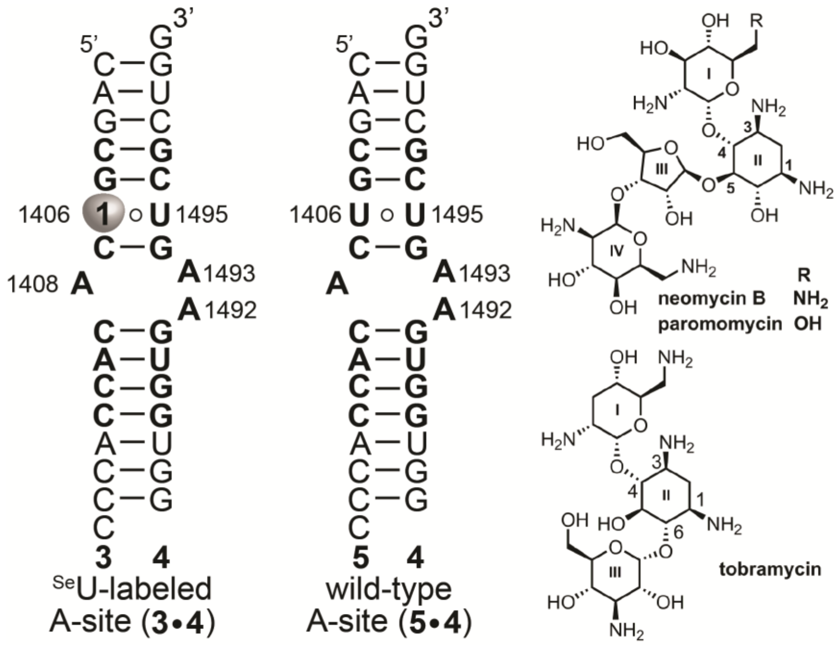
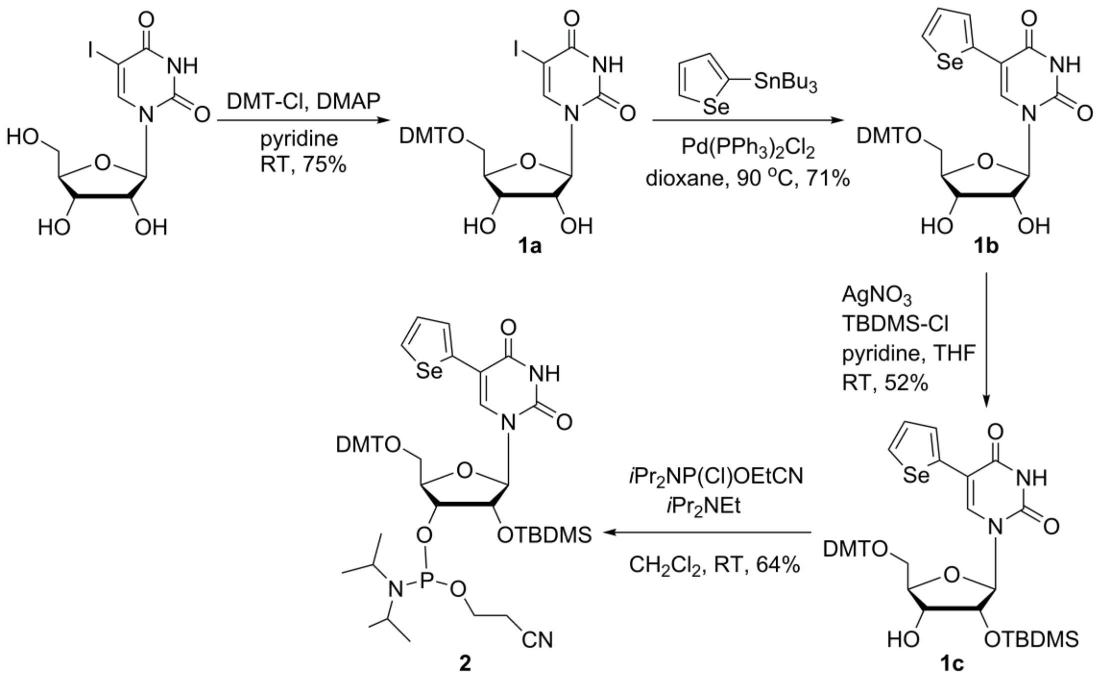


Figure 2. Secondary structure of selenophene-modified (**3·4**) and native (**5·4**) bacterial A-site model RNA. The decoding site internal loop that binds aminoglycosides is highlighted in bold letters. On the right, aminoglycosides used in this study are shown.

The A-site motif in 16S rRNA serves as the decoding site for protein translation by screening cognate paring between the mRNA codon and tRNA anticodon.[17] Biochemical and structural studies revealed that natural aminoglycoside antibiotics bind to the RNA decoding site at an internal loop which contains two conformationally flexible adenosine residues (A1492 and A1493, Figure 2).[18] A-site-aminoglycoside complexes are further stabilized by direct and water bridged H-bonding interactions with the U1406-U1495 noncanonical base pair. Aminoglycoside binding induces a conformational change analogous to cognate tRNA bound to mRNA in the A-site wherein A1492 and A1493 are pushed out of the internal loop. This conformational change reduces the ability of the decoding site to discriminate between cognate and near-cognate tRNAs and ultimately leads to mistranslation. Short duplexes and stem-loop RNA ONs faithfully representing the wild-type decoding site of rRNA have been used as models to investigate the structure and recognition properties of the A-site.[18a,b] A-site RNA models containing fluorescent nucleoside analogs (e.g., 2-aminopurine; 2-AP) at positions A1492/A1493/U1406 have been used in assays to monitor aminoglycoside binding.[19] Based on these studies, we devised a labeled A-site model suitable for both fluorescence and X-ray studies by replacing U1406 with the nucleoside ^{Se}U.

2.2.1 Synthesis of selenophene-modified bacterial ribosomal decoding site (A-site)

We have used bipartite ONs **3** and **4** comprising the bacterial decoding site sequence to assemble the A-site RNA motif **3·4** (Figure 2). Preparation of the phosphoramidite building block **2** required for the synthesis of ^{Se}U-modified RNA ON **3** is outlined in Scheme 1. The phosphoramidite **2** was site-specifically incorporated by solid-phase ON synthesis. Fully deprotected ON **3** was purified by polyacrylamide gel electrophoresis (PAGE, Figure 3A). The integrity of ^{Se}U-modified ON **3** was confirmed by MALDI-TOF mass analysis (Figure 3B). The labeled A-site RNA construct **3·4** was assembled by hybridizing equimolar amounts of ON **3** and **4** in cacodylate buffer. The impact of the selenophene modification on the A-site structure and stability was studied by circular dichroism (CD) and thermal melting experiments. Both, control unmodified **5·4** and modified A-site **3·4** RNAs displayed similar CD spectra and T_m values indicating that the selenophene modification had only a minor effect on the native structure and stability of the A-site RNA (Figure 4).



Scheme 1. Synthesis of 5-selenophene-modified uridine phosphoramidite substrate **2**.

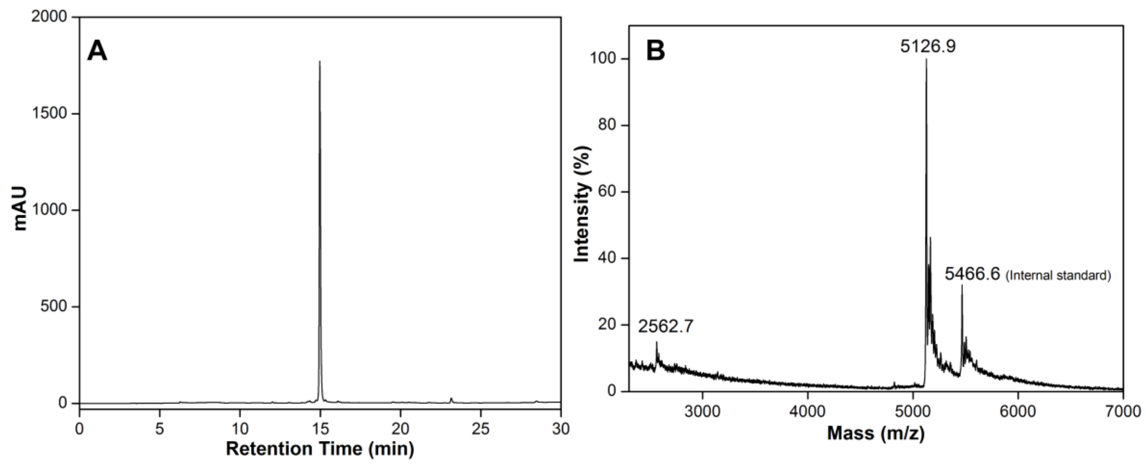


Figure 3. (**A**) HPLC chromatograms of PAGE purified ^{Se}U-modified ON **3** at 260 nm. Mobile phase A = 100 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0–10% B in 10 min and 10–100% B in 20 min. HPLC analysis was performed using Luna C18 column (250 x 4.6 mm, 5 micron). (**B**) MALDI-TOF mass spectrum of RNA ON **3**. The spectrum was calibrated relative to the +1 ion (m/z = 5466.6) of an internal DNA ON standard. Calcd. mass for ON **3**: $[M]^+$ 5127.1; found: $[M]^+$ 5126.9 and 2562.7 $[M]^{2+}$. For more details see experimental section.

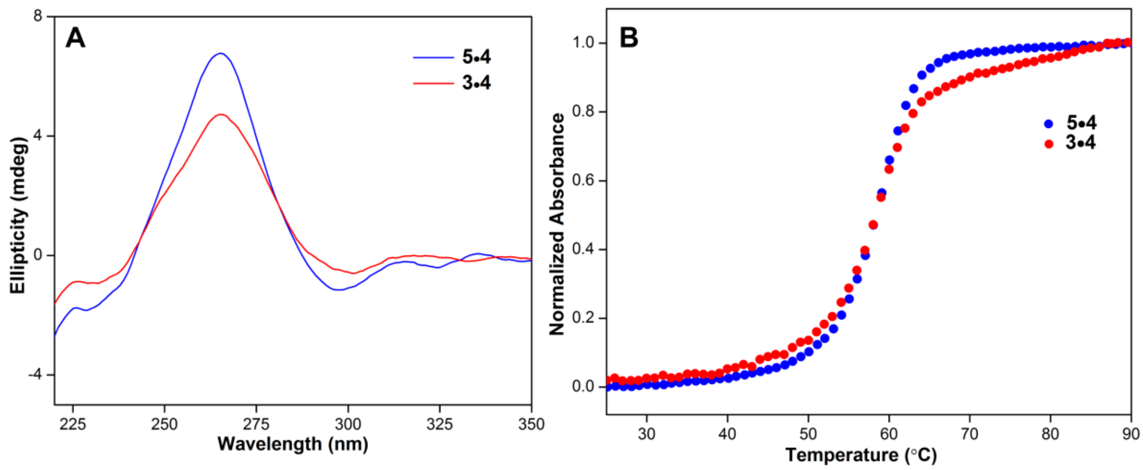


Figure 4. (**A**) CD spectra (5 µM) of control unmodified A-site RNA **5•4** (blue) and ^{Se}U-modified A-site RNA **3•4** (red). Both unmodified and modified ONs showed similar CD profiles, which were in consensus with the literature report.[29] (**B**) UV-thermal melting profile of control A-site RNA **5•4** (blue) and ^{Se}U-modified A-site RNA **3•4** (red) at 260 nm. T_m values for **5•4** and **3•4** were found to be 59.0 ± 0.7 °C and 58.8 ± 0.3 °C, respectively. For details see experimental section.

2.2.2 Aminoglycoside binding to A-site by fluorescence spectroscopy

The ability of the nucleoside probe to photophysically report a ligand-induced conformational change in the A-site was evaluated by titrating RNA **3·4** with aminoglycoside antibiotics derived from 2-deoxystreptamine (2-DOS). We chose 4,5-disubstituted 2-DOS (paromomycin and neomycin B) and 4,6-disubstituted 2-DOS (tobramycin) aminoglycosides as they are known to reduce translation fidelity in bacteria by interacting with residues present in the internal loop of the decoding site (Figure 2).[17,18] A-site model RNA was excited at 330 nm, and changes in emission intensity at λ_{em} =450 nm were monitored as a function of increasing concentration of aminoglycosides. Titration with paromomycin and neomycin resulted in an increase in fluorescence intensity, corresponding to an apparent K_d of 3.80 ± 0.20 µM and 2.10 ± 0.22 µM, respectively (Figure 5 and Figure 6). The binding constant and higher affinity exhibited by neomycin compared to paromomycin are consistent with literature reports ($K_d \approx 10^{-6}$ M).[19a–c]

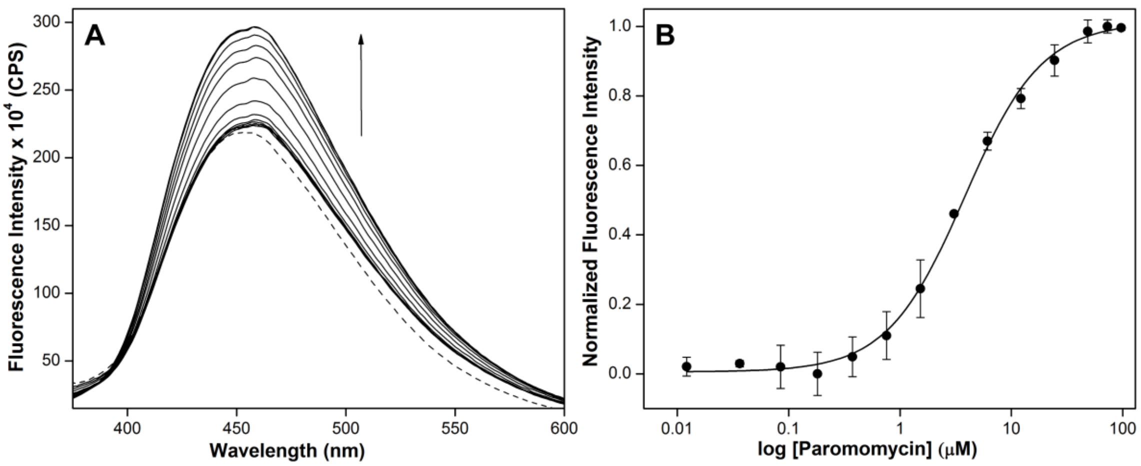


Figure 5. (**A**) Emission spectra (solid lines) of A-site RNA **3•4** as a function of increasing concentration of paromomycin. Dashed line represents emission profile in the absence of aminoglycoside. (**B**) Curve fit for the titration of A-site with paromomycin. Normalized fluorescence intensity at λ_{em} = 450 nm is plotted against log [aminoglycoside].

It is important to note that 2-AP, the most widely used conformation-sensitive nucleoside probe, fails to report the binding of neomycin to A-site RNA despite that both paromomycin and neomycin are known to bind the RNA target in a similar fashion and result in ribosomal miscoding during translation.[16, 19] Next, we studied the binding of tobramycin, which binds A-site RNA with lower affinity compared to 4,5-disubstituted 2-DOS aminoglycosides.[18b] ^{Se}U reported the binding of tobramycin with discernible enhancement in fluorescence intensity, albeit with reduced affinity compared to the native decoding site (66.3 ± 5.9 μM and 9.5 ± 2.0 μM, respectively, Figure 7).[19a]

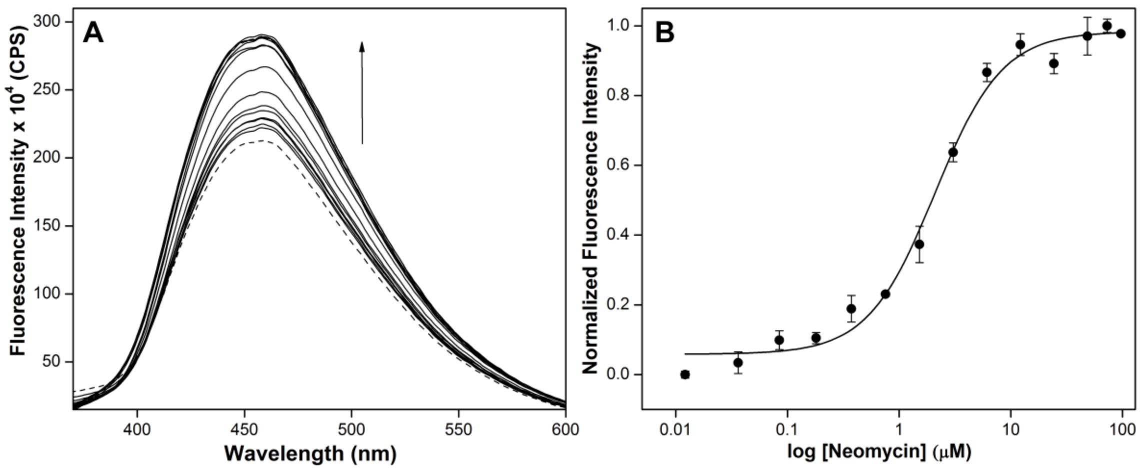


Figure 6. (**A**) Emission spectra (solid lines) of A-site **3•4** as a function of increasing concentration of neomycin B. Dashed line represents emission profile in the absence of aminoglycoside. (**B**) Curve fit for the titration of A-site with neomycin B. Normalized fluorescence intensity at λ_{em} = 450 nm is plotted against log [aminoglycoside].

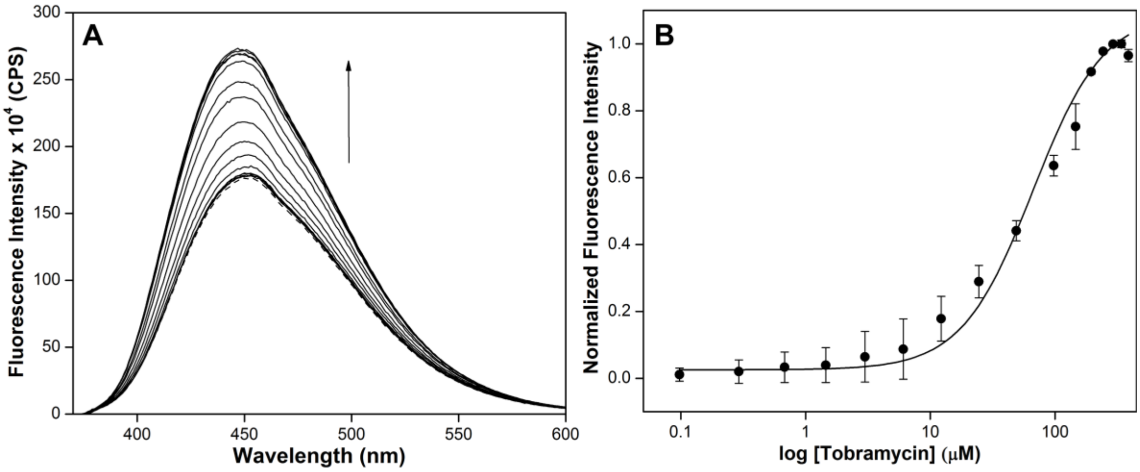


Figure 7. (**A**) Emission spectra (solid lines) of A-site **3•4** as a function of increasing concentration of tobramycin. Dashed line represents emission profile in the absence of aminoglycoside. (**B**) Curve fit for the titration of A-site with tobramycin. Normalized fluorescence intensity at λ_{em} = 450 nm is plotted against log [aminoglycoside].

2.2.3 Crystal structure of ^{Se}U-labeled A-site RNA

In order to elucidate the structural basis of conformation sensitivity of the nucleoside analog and allow a direct comparison between solution binding data and crystallographic structure, selenophene-modified A-site model RNA **3·4** was crystallized and its 3D structure determined by X-ray diffraction at 2.14 Å resolution (PDB: 5T3K). Hanging-drop vapor diffusion resulted in well-diffracting crystals, which contained two unique A-site RNA molecules with 50% occupancy in the unit cell (Figure 8, Table 1). In one of the RNA copies, both A1492 and A1493 residues were flipped out of the loop, closely resembling the structure of the decoding site bound to aminoglycosides (Figure 9A). In the second RNA copy, only A1493 was flipped-out of the loop, while A1492 was found to be inside, base pairing and stacking with A1408 and G1491, respectively (Figure 9B and Table 2). Other than these differences the two A-site copies were nearly identical. Importantly, the electron density of the heavy Se atom was clearly visible, establishing the orientation of the 5-membered ring as coplanar with the pyrimidine and Se facing the carbonyl O4. Like U1406 in the native decoding site, ^{Se}U1406 also formed a noncanonical Watson-Crick pair with U1495 (Figure 10, Table 2). The characteristically strong diffraction signal of the Se atom in the ^{Se}U analog suggests that anomalous dispersion from Se can be used in direct RNA structure determination by SAD or MAD phasing techniques.[9, 12]

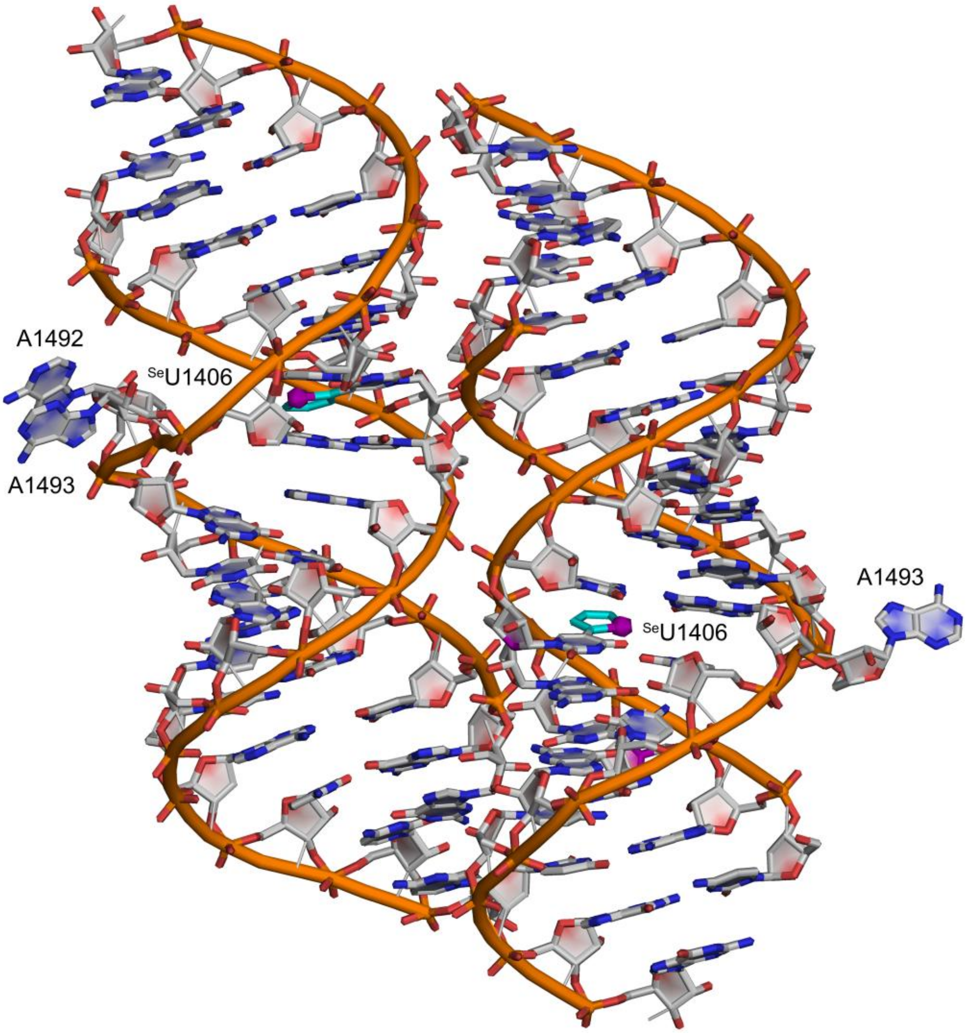


Figure 8. Crystal structure of ^{Se}U-modified A-site RNA **3•4** containing two unique A-sites with 50% occupancy in the unit cell. (left) RNA copy in which A1492 and A1493 residues are flipped-out of the loop. This structure closely resembles the structure of the decoding site bound to the aminoglycosides. (right) RNA copy in which only A1493 is flipped-out of the loop. ^{Se}U is shown in cyan color.

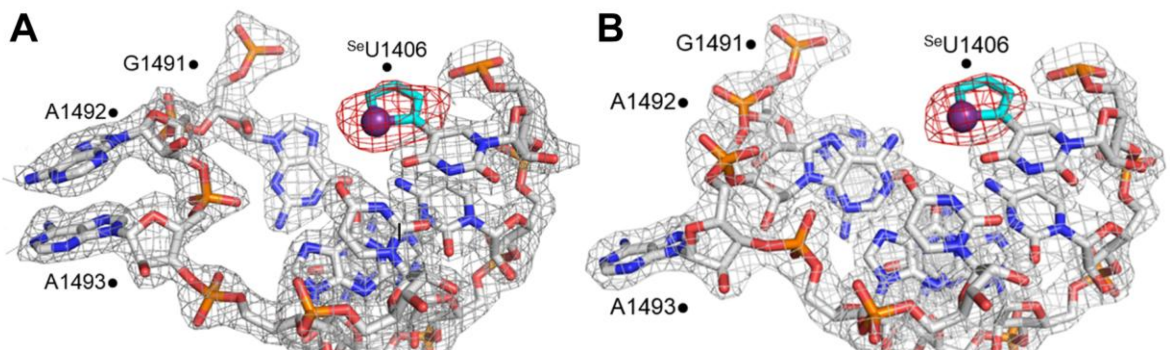


Figure 9. Electron density map around the ^{Se}U-modified A-site RNA **3·4** with A1492 flipped out (left) or in (right) the RNA helix, as is seen in a previous A-site model RNA structure.[19a] The selenophene ring is shown in cyan with the selenium atom shown as an indigo sphere. The $2F_o$-F_c electron density map, contoured at 1.0σ, is shown in gray, and the $2F_o$-F_c electron density map, contoured at 3.0σ, is shown in red (PDB ID: 5T3K).

Table 1. Crystallographic data collection and refinement statistics.

Data Collection	
Wavelength (Å)	1.54
High-resolution limit (Å)	2.14
Low-resolution limit (Å)	26.02
Redundancy[a]	2.9 (1.7)
Completeness (%)[a]	77.6 (46.7)
$I/\sigma\ (I)$[a]	11.32 (4.25)
Total reflections	103258
Unique reflections	7483
Refinement	
Space group	$P2_1$
Cell dimensions (Å)	
a	31.55
b	86.96
c	32.56
α	90.00
β	94.14
γ	90.00
R_{work}/R_{free}	0.197 / 0.247
No. atoms	
RNA atoms	1408
Solvent atoms	60
Metal ions	10 Mg^{2+}
Mean B factors ($Å^2$)	
RNA	32.6
Solvent	32.2
Metal	37.5
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.366
Dihedral angles (°)	20.892

[a]Numbers in parentheses are for the highest-resolution shell.

Table 2. Important H-bonding and stacking interactions present in control, ^{Se}U-modified A-site RNA constructs, and unmodified A-site RNA bound to aminoglycoside antibiotics.

A-site RNA construct	Hydrogen bond	distance (Å)	π-π distance (Å)[a]
Control A-site (PDB: 1T0D)			
G1405-C1496	N2H---O2	2.827	G1405 (Py)-U1406
	N1H---N3	2.857	3.769
	O6---HN	2.857	
U1406-U1495	N3H---O2	3.015	U1406-C1407
	O4---HN3	2.892	3.888
C1407-G1494	O2---HN2	2.741	
	N---HN1	2.791	
	NH---O6	2.818	
^{Se}U-modified **3•4** A-site (PDB: 5T3K)			
G1405-C1496	N2H---O2	2.588	G1405 (Py)-^{Se}U1406 (uracil)
	N1H---N3	2.857	3.701
	O6---HN	2.985	U1406-C1407
U1406-U1495	N3H---O2	3.166	3.713
	O4---HN3	2.913	G1405 (Im)-U1406 (Se)
C1407-G1494	O2---HN2	2.456	3.720
	N---HN1	2.716	
	NH---O6	2.823	
Paromomycin bound to A-site RNA (PDB: 1J7T)			
G1405-C1496	N2H---O2	2.206	G1405 (Py)-U1406
	N1H---N3	2.564	3.776
	O6---HN	2.776	U1406-C1407
U1406-U1495	N3H---O4	2.808	4.205
C1407-G1494	O2---HN2	2.429	
	N---HN1	2.873	
	NH---O6	3.220	
Neomycin bound to A-site RNA (PDB: 2A04)			
G1405-C1496	N2H---O2	2.530	G1405 (Py)-U1406
	N1H---N3	2.847	3.860
	O6---HN	3.035	U1406-C1407
U1406-U1495	O4---HN3	3.203	3.958
C1407-G1494	O2---HN2	2.891	
	N---HN1	2.892	
	NH---O6	2.767	
Tobramycin bound to A-site RNA (PDB: 1LC4)			
G1405-C1496	N2H---O2	2.711	G1405 (Py)-U1406
	N1H---N3	2.692	3.715
	O6---HN	2.564	U1406-C1407
U1406-U1495	N3H---O4	2.365	4.273
C1407-G1494	O2---HN2	2.444	
	N---HN1	2.577	
	NH---O6	2.583	

[a]Centroid to centroid distance has been reported.

Py = pyrimidine ring; Im = imidazole ring of guanine; Se = selenophene ring of ^{Se}U ribonucleoside.

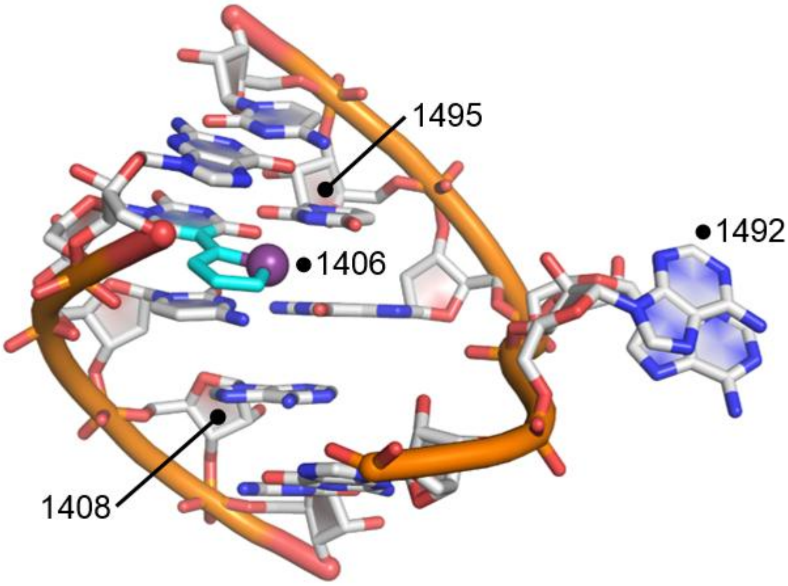


Figure 10. Structure showing noncanonical Watson-Crick pairing between ^{Se}U1406 and U1495 as in the native decoding site RNA.

The selenophene ring of the base-paired fluorophore (^{Se}U1406) is projected into the major groove and stacked with the imidazole ring of G1405 (3.72 Å, Figure 11). The uracil ring is stacked with the pyrimidine rings of G1405 (3.70 Å) and C1407 (3.71 Å). It is well documented that stacking interaction with neighboring bases and proximity to guanine residue can reduce the fluorescence efficiency of fluorophores incorporated into ONs.[20] Hence, the stacked conformation of ^{Se}U in the unbound state of the A-site RNA exhibits lower fluorescence intensity (Figure 5). However, upon titrating A-site with aminoglycosides, the nucleoside probe reported the binding event by enhancement of fluorescence intensity. To gain further insight on the effect of modification on the A-site structure and responsiveness of the label, the ^{Se}U-A-site crystal structure was superimposed onto structures of native A-site RNA and aminoglycosides bound to the native A-site RNA.

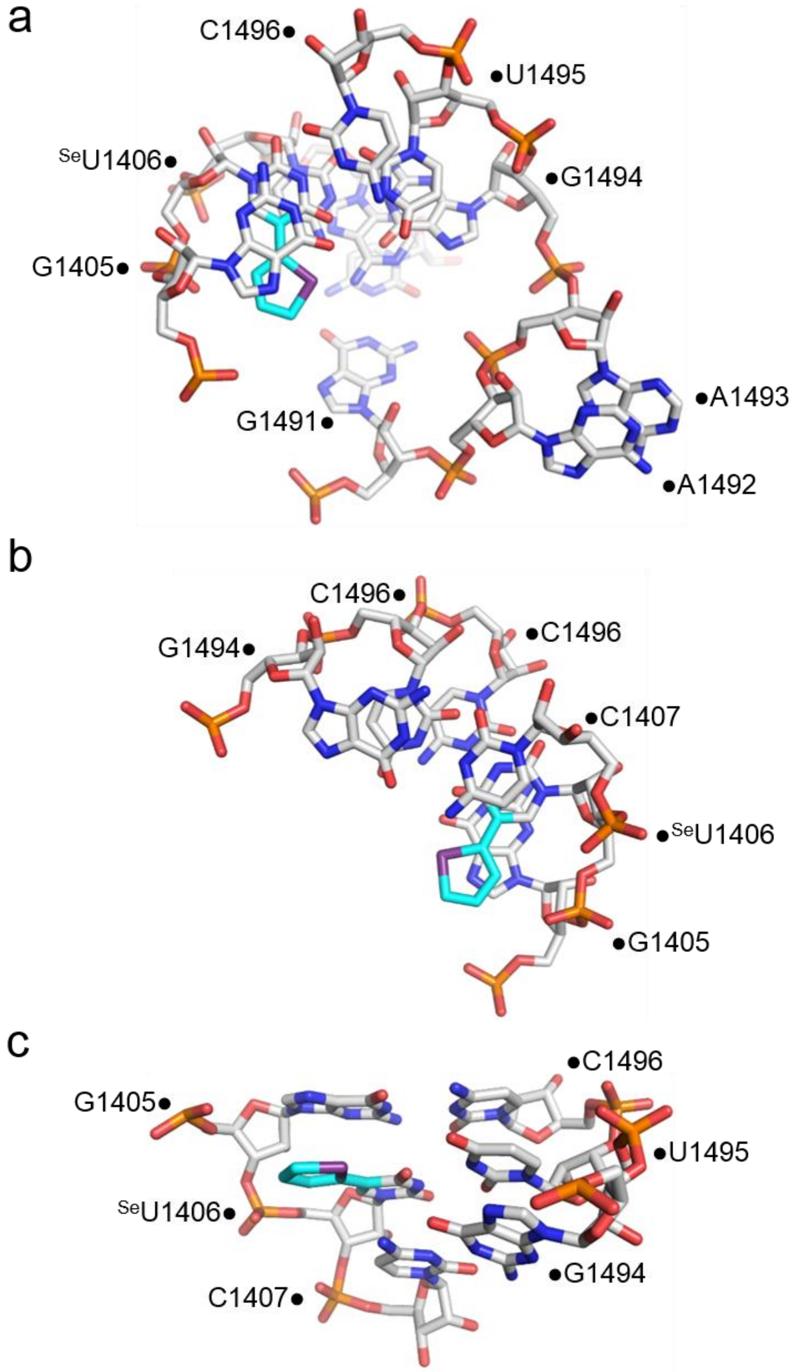


Figure 11. Stacking of ^{Se}U1406 with neighboring residues G1405 and C1407. (a) Top-down view with G1405 in the foreground in which the imidazole ring of G1405 is stacked with the selenophene ring of ^{Se}U. (b) Bottom-up view with C1407 in the foreground in which the uracil ring of ^{Se}U is stacked with the C1407 base. (c) Side view with G1405 above and C1407 below the ^{Se}U1406.

2.2.4 Superimposition of ^{Se}U A-site RNA with native and antibiotic bound A-site RNA

Superimposition of the selenophene-modified and native decoding site structures reveal that the structures are almost identical with a RMS distance of 0.367 Å for 1305 identical atoms out of 1362 total atoms (Figure 13). The 2′-OH group of the G1491 sugar is slightly drawn into the interior of the decoding site in the A1492-out conformation of the modified RNA, whereas this is not seen in the structure of the native A-site model. This 2′-OH group interacts with a bridging water molecule to A1408 and with a chain of water molecules beneath the selenophene ring and above a Mg^{2+} ion (Figure 12). In addition to thermal melting and CD studies, the crystal structure confirms that selenophene-modification has only a minor impact on the native A-site RNA structure.

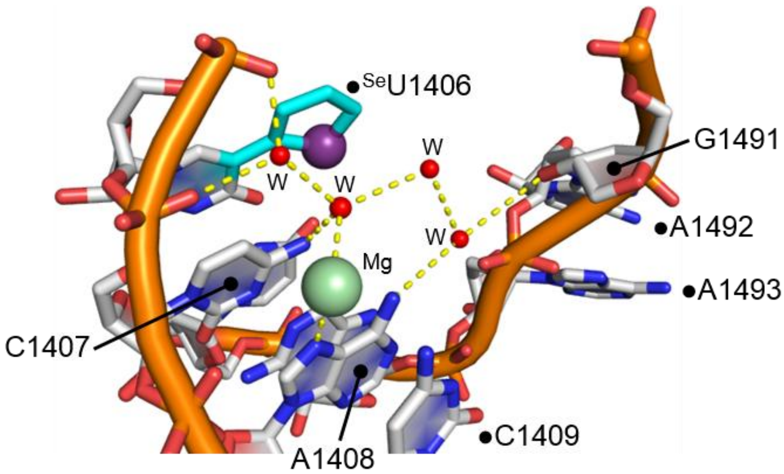


Figure 12. Structure showing the interaction between the 2′-OH group of G1491 and the N6 of A1408 through a bridging water molecule.

In the superimposed structures of paromomycin and neomycin there was no apparent steric clash between the aminoglycosides and selenophene modification (Figure 13B and Figure 13C).[21, 22] Other than the conserved interactions between the ligands and decoding site, 6- and 2′-hydroxyl groups of aminoglycoside rings II and III, respectively, were found to be in close proximity with the Se atom (2.2/2.4 Å and 3.4/2.6 Å, respectively). These superimposed models indicate that the ^{Se}U modification should not affect binding of paromomycin and neomycin to the decoding site. This notion is in consensus with the fluorescence binding experiments wherein ^{Se}U-labeled A-site exhibited binding affinity comparable to literature reports.[19] In case of tobramycin, which binds differently to the decoding site as compared to 4,5-disubstituted 2-DOS aminoglycosides,[23] the superimposed structure revealed a steric clash between ring III and the selenophene moiety (Figure 13D).

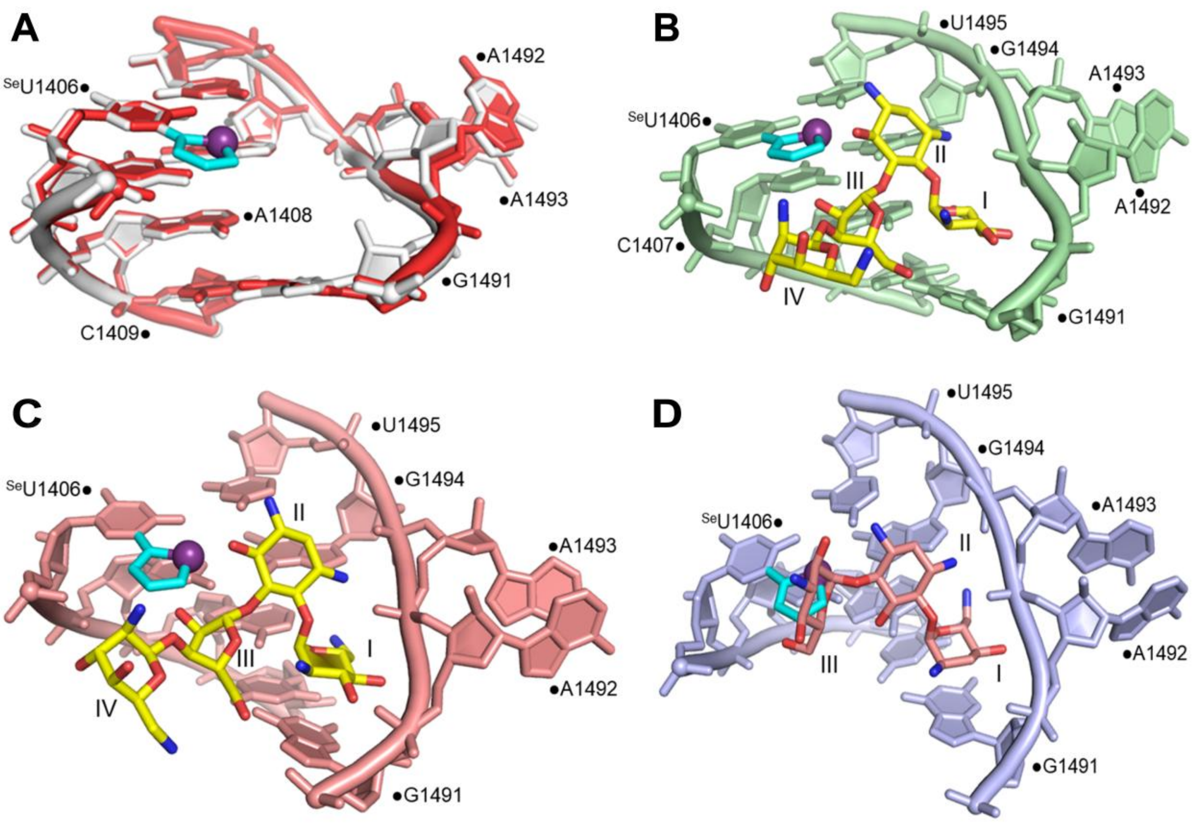


Figure 13. (**A**) Superimposition of the ^{Se}U-modified (white) and native (red; PDB: 1T0D)[19a] A-site RNA structures. Superimposition of the ^{Se}U-modified A-site RNA and native A-site RNA bound to (**B**) paromomycin (yellow; PDB: 1J7T).[21] Superimposition of the ^{Se}U-modified A-site RNA and native A-site RNA bound to (**C**) neomycin B (yellow; PDB: 2A04).[22] (**D**) tobramycin (salmon; PDB: 1LC4).[23] For clarity, superimposed selenophene-modified uridine (cyan) is only shown.

This would explain the observed lower affinity of tobramycin for the ^{Se}U-labeled A-site. Not surprisingly, mutations of U1406, predicted to confer antibiotic resistance, are known to reduce the binding of tobramycin due to a comparable steric clash.[24] For example, replacing U1406 with adenine residue results in reduction in affinity of tobramycin (EC_{50} values, wt: 9.5 μM; mut: 110 μM).[19a] However, mutation of U1406 does not affect the binding of 4,5-disubstituted 2-DOS aminoglycosides to A-site. In this context, ^{Se}U-labeled A-site can potentially serve as a mutant model because the nucleoside probe faithfully distinguishes the binding of aminoglycosides based on affinity and 2-DOS class. Such labeled RNA constructs could be used for screening of small molecule ligands against resistant bacterial strains.[25]

2.2.5 Fluorescence outcome based on crystal structure

Further, we compared the conformation of U1406 and its neighboring bases in the structures of native A-site RNA in the absence and presence of aminoglycosides and ^{Se}U-labeled A-site to understand the photophysical behavior of ^{Se}U in solution binding experiments. The stacking interaction between U1406 and G1405 in the unbound state (3.77 Å) and aminoglycoside-bound state (3.72–3.86 Å) is similar. However, U1406, which is completely stacked with C1407 (3.89 Å) in the free A-site structure, is found to be only partially stacked with C1407 (3.96–4.27 Å) in all three aminoglycoside-A-site complexes (Figure 14, Table 2). Based on these structural observations, it is suggested that ^{Se}U placed at the 1406 position is likely to sense a similar environment as U1406 upon ligand binding. Hence, enhanced fluorescence intensity displayed by ^{Se}U in ligand binding experiments can be attributed to a reduced stacking interaction between the fluorophore and neighboring bases.

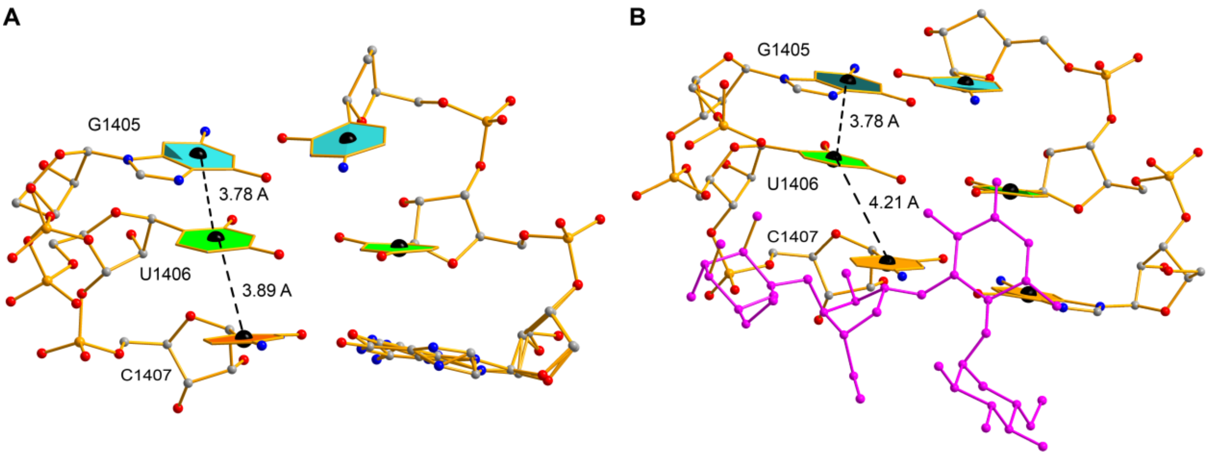


Figure S14. Comparison of stacking interactions between U1406 and its neighboring bases (G1405 and C1407) in the structure of the native A-site in the absence (**A**), and presence (**B**) of an aminoglycoside. A representative structure of aminoglycoside (paromomycin in magenta; PDB: 1J7T) bound to A-site RNA is shown here.

2.3 Conclusions

In summary, we have introduced a multipurpose nucleoside probe functioning both as an environment-sensitive fluorophore and an anomalous X-ray diffraction label, which facilitated the direct correlation of RNA structure and ligand recognition under equilibrium conditions and in 3D. The new bifunctional nucleoside probe will be a powerful tool for combined fluorescence and crystallography studies, including time-resolved experiments, for example with femtosecond X-ray laser pulses, on dynamic RNA systems such as riboswitches.[26] The responsiveness of the fluorophore to subtle conformational changes and the strong diffraction signal from the Se atom as demonstrated by using decoding site RNA suggest that such a probe, when judiciously placed, would not only allow the concurrent investigation of RNA structure and function but also could support integrated approaches to antibiotics discovery against resistant bacterial strains.

2.4 Experimental section

2.4.1 Materials

5-iodouridine, selenophene, *n*-butyllithium, tributyltin chloride, *bis*(triphenylphosphine)-palladium(II) chloride, 4,4′-dimethoxytrityl chloride, paromomycin, tobramycin and hygromycin were obtained from Sigma-Aldrich. Neomycin was procured from Fluka. 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite was purchased from Alfa Aesar. TBDMS-protected ribonucleoside phosphoramidite substrates and solid supports required for RNA synthesis were purchased from Glen Research. All other reagents for solid-phase oligonucleotide (ON) synthesis were obtained from ChemGenes Corporation. Custom synthesized RNA ONs were purchased from Dharmacon RNAi Technologies, deprotected according to the supplier's protocol, PAGE-purified and desalted using Sep-Pak Classic C18 cartridge. Chemicals (BioUltra grade) for preparing buffer solutions were purchased from Sigma-Aldrich. Autoclaved water was used for preparation of all buffer solutions and in biophysical analysis.

2.4.2 Instrumentation

NMR spectra were recorded on a 400 MHz Jeol ECS-400 spectrometer. Mass measurements were recorded on the Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer and Water Synapt G2 High Definition mass spectrometers. Modified RNA ONs were synthesized on Applied Biosystems RNA/DNA synthesizer (ABI-394). Absorption spectra were recorded on Shimadzu UV-2600 spectrophotometer. HPLC analysis was performed using Agilent Technologies 1260 Infinity. UV-thermal melting studies of ONs were performed on Cary 300Bio UV-Vis spectrophotometer and CD analysis was performed on JASCO J-815 CD spectrometer. Steady-state and time-resolved fluorescence experiments were carried out in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on Horiba Jobin Yvon, Fluorolog-3.

Crystal screening and X-ray diffraction data were collected at 110 K on a Rigaku rotating anode X-ray generator (λ = 1.54 Å) equipped with a MAR345 imaging plate detector system.

2.4.3 Synthesis of 5-selenophene-modified uridine phosphoramidite substrate 2

2.4.3.1 5-Selenophene-modified 5′-*O*-DMT-protected uridine 1b: A mixture of 5′-*O*-DMT-protected 5-iodouridine **1a** [27] (1.0 g, 1.49 mmol, 1.0 equiv), 2-(tri-*n*-butylstannyl)selenophene [28] (1.37 g, 3.27 mmol, 2.2 equiv) and *bis*(triphenylphosphine)-palladium(II) chloride (0.084 g, 0.12 mmol, 0.08 equiv) was dissolved in anhydrous dioxane (25 ml). The reaction mixture was heated at 90 °C for 2 h and filtered through celite pad. The celite pad was washed with dichloromethane (2 × 20 ml), filtrate was evaporated and the resulting residue was purified by silica gel column chromatography to afford compound **1b** as an off white foam (0.71 g, 71% yield). TLC (CH_2Cl_2:MeOH = 95:5 with few drops of Et_3N); R_f = 0.36; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.02 (s, 1H), 7.85 (dd, J_1 = 5.0 Hz, J_2 = 1.4 Hz, 1H), 7.40–7.38 (m, 2H), 7.28–7.26 (m, 4H), 7.23–7.15 (m, 3H), 6.85–6.83 (m, 2H), 6.77–6.75 (m, 4H), 5.99 (d, J = 4.8 Hz, 1H), 4.49 (t, J = 5.0 Hz, 1H), 4.35 (t, J = 4.6 Hz, 1H), 4.32–4.31 (m, 1H), 3.73 (s, 6H), 3.49–3.47 (m, 1H), 3.35 (dd, J_1 = 10.6 Hz, J_2 = 3.8 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 162.2, 158.7, 151.0, 144.4, 136.3, 135.6, 135.5, 132.0, 130.1, 130.1, 128.9, 128.2, 128.1, 127.2, 124.7, 113.4, 112.2, 90.7, 86.9, 84.8, 75.6, 71.2, 63.4, 55.3; HRMS: m/z Calcd. for $C_{34}H_{33}N_2O_8Se$ $[M+H]^+$ = 677.1402, found = 677.1394.

2.4.3.2 5-Selenophene-modified 2′-*O*-TBDMS protected uridine 1c: Compound **1b** (0.95 g, 1.41 mmol, 1.0 equiv) and silver nitrate (0.62 g, 3.65 mmol, 2.6 equiv) were dissolved in anhydrous pyridine (4.8 ml). To the above solution was added anhydrous THF (14 ml). After stirring for 10 min *tert*-butyldimethylsilyl chloride (TBDMS-Cl, 0.55 g, 3.65 mmol, 2.6 equiv) was added. Reaction mixture was stirred for additional 0.5 h and filtered through celite pad. Celite pad was washed with ethyl acetate (2 × 15 ml) and the combined organic solution was then washed with 5 % sodium bicarbonate (20 ml) followed by brine solution (20 ml). The organic extract was dried over sodium sulphate and evaporated. The resulting crude residue was purified by silica gel column chromatography to afford the product **1c** as a white solid (0.57 g, 52% yield). TLC (petroleum ether:EtOAc = 60:40 with few drops of Et_3N); R_f = 0.42; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.62 (br, 1H), 7.98 (s, 1H), 7.88 (d, J = 5.6 Hz, 1H), 7.42–7.40 (m, 2H), 7.30–7.28 (m, 5H), 7.24–7.19 (m, 2H), 6.78–6.67 (m, 6H), 6.12 (d, J = 5.6 Hz, 1H), 4.51 (t, J = 5.2 Hz, 1H), 4.26–4.22 (m, 2H), 3.74 (s, 6H), 3.54–3.52 (m, 1H), 3.40 (dd, J_1 = 11.0 Hz, J_2 = 3.0 Hz, 1H), 0.93–0.91 (s, 9H), 0.15 (s, 3H), 0.14 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 161.3, 158.8, 149.4, 144.4, 136.1, 135.4, 135.3, 132.6, 132.2, 130.2, 130.1, 128.9, 128.2, 128.1, 127.3, 124.9, 123.9, 113.4, 88.0, 87.0, 84.0, 75.8, 71.3, 63.4, 55.4, 25.8, 18.1, -4.5, -4.9; HRMS: m/z Calcd. for $C_{40}H_{46}N_2O_8SeSiNa$ $[M+Na]^+$ = 813.2086, found = 813.2123.

2.4.3.3 5-Selenophene-modified uridine phosphoramidite substrate 2: To a solution of compound **1c** (0.41 g, 0.52 mmol, 1.0 equiv) in anhydrous dichloromethane (4.1 ml) was added *N*,*N*-diisopropylethylamine (0.226 mL, 1.30 mmol, 2.5 equiv) and stirred for 10 min. To this solution was slowly added 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (0.14 mL, 0.63 mmol, 1.2 equiv) and the reaction mixture was stirred for 12 h. The solvent was evaporated to dryness and the residue was redissolved in ethyl acetate (20 ml), which was washed with 5% sodium bicarbonate solution (15 ml) followed by brine solution (15 ml). The organic extract was dried over sodium sulphate, evaporated and the crude solid residue was purified by column chromatography to afford the product **2** as a white solid (0.33 g, 64%). TLC (petroleum ether:EtOAc = 60:40 with few drops of Et_3N); R_f = 0.40; Analytical data was obtained using one of the diastereomers, which eluted first. ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.03 (s, 1H), 7.86–7.84 (m, 1H), 7.44–7.42 (m, 2H), 7.32–7.28 (m, 5H), 7.25–7.20 (m, 2H), 6.78–6.70 (m, 6H), 6.12 (d, J = 6.4 Hz, 1H), 4.49 (t, J = 5.8 Hz, 1H), 4.37 (br, 1H), 4.24–4.20 (m, 1H), 3.74 (s, 6H), 3.61–3.51 (m, 5H), 3.30 (dd, J_1 = 10.6 Hz, J_2 = 2.6 Hz, 1H), 2.28–2.24 (m, 2H), 1.18–1.13 (m, 12H), 0.88 (s, 9H), 0.11 (s, 3H), 0.07 (s, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 161.3, 158.8, 149.4, 144.4, 135.4, 132.8, 132.0, 130.3, 130.3, 130.2, 128.9, 128.4, 128.2, 128.1, 127.3, 124.8, 113.4, 87.6, 87.0, 84.2, 75.8, 74.9, 63.4, 55.4, 43.6, 43.5, 29.8, 25.8, 24.8, 22.7, 20.3, 18.1, -4.4, -4.7; ^{31}P NMR (162 MHz, $CDCl_3$): δ (ppm) 151.6; HRMS: m/z Calcd. for $C_{49}H_{63}N_4O_9PSeSiNa$ $[M+Na]^+$ = 1013.3165, found = 1013.3191.

2.4.4 Solid-phase synthesis of selenophene-modified RNA ON 3:

RNA ON **3** was synthesized on a 1 μmole scale using CPG solid support (1000 Å). Standard synthesis cycles employed for 2′-*O*-TBDMS-protected phosphoramidite substrates were used. While incorporation of native 2′-*O*-TBDMS-protected phosphoramidites was performed with a coupling time of 10 min (two times), incorporation of 5-selenophene-modified-2′-*O*-TBDMS-protected phosphoramidite substrate **2** was performed with a coupling time of 15 min (two times). The coupling efficiency for phosphoramidite substrate **2** based on trityl monitor-assay was found to be ~40%. After deprotection of trityl group on the synthesizer, the solid support was treated with a 1:1 solution of 10 M methylamine in ethanol and water (1.5 mL) for 12 h at RT. The mixture was centrifuged and the supernatant was evaporated to dryness on a Speed Vac. The residue was dissolved in DMSO (100 μL) and was added TEA.3HF (150 μL). The resulting solution was heated at 65 °C for 2.5 h and was brought to RT. The completely deprotected ON solution was lyophilized and was purified by PAGE (20% gel) under denaturing conditions. The band corresponding to the full-length product was identified by UV shadowing. The ON was extracted with ammonium acetate buffer (0.5 M, 3 ml) and desalted using a Sep-Pak classic C18 cartridge. A 1 μmol scale synthesis gave ~100 nmol of the labeled RNA after PAGE purification. See Figure 3 for HPLC chromatogram and mass spectrum of purified ON **3**.

2.4.5 MALDI-TOF mass analysis of selenophene-modified RNA ON 3: 2 µL of the modified ON **3** (~200 µM) was combined with 1 µL of ammonium citrate buffer (100 mM, pH 9), 2 µL of a DNA internal standard (200 µM) and 4 µL of saturated 3-hydroxypicolinic acid solution. The sample was desalted using an ion-exchange resin (Dowex 50W-X8, 100-200 mesh, ammonium form), spotted on the MALDI plate and air dried.

2.4.6 Reconstitution of bacterial A-site RNA models: A-site containing selenophene-modified uridine at position 1406 (**3•4**) and control native (**5•4**) A-site models were assembled by heating 1:1 mixture of respective ONs (40 µM) in buffer (30 mM sodium cacodylate, 100 mM NaCl, pH 6.8) at 70 °C for 3 min. Samples were slowly cooled to RT and placed on ice for 30 min before fluorescence, CD and thermal melting analysis were performed. For fluorescence and thermal melting analysis, samples of A-site constructs were diluted with cacodylate buffer to a final RNA concentration of 2 µM. For CD analysis, 5 µM concentration of A-site constructs was used.

2.4.7 CD analysis: 5 µM of the control unmodified and selenophene modified A-site constructs was used in the CD experiments. CD spectra were collected from 350 to 220 nm using a quartz cuvette (Sterna Scientific, path length 2 mm) on a Jasco J-815 CD spectrometer using 1 nm bandwidth at 20 °C. Experiments were performed in duplicate wherein each spectrum was an average of three scans. The spectrum of buffer was subtracted from all sample spectra.

2.4.8 Fluorescence monitoring of aminoglycosides binding to selenophene modified A-site construct 3•4: 400 μL of A-site construct **3•4** (2 μM) was taken in a micro fluorescence cuvette (Hellma, path length 1.0 cm) and titrated by adding 1 μL of stock solutions of increasing concentration of aminoglycosides. After each addition of aminoglycoside the sample was excited at 330 nm with excitation and emission slit widths of 4 and 6 nm, respectively, and changes in fluorescence intensity at emission maximum (450 nm) were measured. The aminoglycoside was added till the fluorescence saturation point. The total volume change upon addition of aminoglycosides throughout the titration was ≤ 4%. A spectral blank of cacodylate buffer in the absence of RNA and aminoglycoside was subtracted from all titrations. A control titration experiment performed as above by adding buffer (instead of aminoglycosides) to A-site **3•4** did not result in detectable changes in fluorescence indicating that repeated excitation of the sample did not cause photodegradation or photobleaching of the labeled A-site. Therefore, the increase in fluorescence intensity observed upon addition of aminoglycoside antibiotics to A-site reflects the ability of selenophene-modified uridine to report RNA-ligand binding. All fluorescence experiments were performed in triplicate at 20 °C. Normalized fluorescence intensity (F_N) versus log of aminoglycoside (AG) concentration plots were fitted using Hill equation (OriginPro 8.5.1) to determine the apparent dissociation constant (K_d) values for the binding of aminoglycosides to A-site construct.[30,19a] One-site binding model was used. χ^2 values for the curve fits were found to be very close to unity. Hill coefficient (n) was in the range of 1.2–1.5.

$$F_N = \frac{F_i - F_0}{F_s - F_0}$$

F_i is the fluorescence intensity at each titration point. F_0 and F_s are the fluorescence intensity in the absence of aminoglycoside and at saturation, respectively. n is the Hill coefficient or degree of cooperativity associated with the binding.

$$F_N = F_0 + (F_S - F_0)\left(\frac{[\mathrm{AG}]^n}{[KD]^n + [\mathrm{AG}]^n}\right)$$

2.4.9 Crystallization and Data Collection: The ^{Se}U-modified A-site model RNA **3•4** was assembled by annealing stoichiometric amounts of the ON **3** and **4** in 10 mM sodium cacodylate, pH 6.5 by heating to 65 °C for 4 min followed by snap cooling on ice for 10 min followed by slow warming to room temperature. After annealing, the RNA was crystallized at 22 °C by hanging drop vapor diffusion. For crystallization, 1 μL of 0.2 mM RNA was mixed with an equal volume of crystallizing solution containing 2.40–2.55 M ammonium sulfate, 10 mM magnesium acetate, and 40 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer (pH 5.6). Crystals grew over the course of four to seven months after equilibration against 700 μL of well solution containing crystallizing solution. Crystals were flash-cooled in liquid nitrogen. Crystal screening and X-ray diffraction data were collected at 110 K on a Rigaku rotating anode X-ray generator (λ = 1.54 Å) equipped with a MAR345 imaging plate detector system. Datasets were processed, integrated, and scaled with the HKL2000 package.[31]

Structure solution and refinement: The three-dimensional structures of the ^{Se}U-modified A-site RNA construct was solved by molecular replacement with the program Phaser [32] using a previously determined structure,[19a] as the search model and refined by the program Refmac [33] both within the CCP4 package.[34] Subsequent iterative rounds of manual building and refinement, alternating between Refmac and manual rebuilding in Coot,[35] were based on the obtained 2Fo-Fc and Fo-Fc density maps. Final refinement was carried out in PHENIX [36] with individual isotropic atomic displacement parameters and water picking. Crystallographic data collection and refinement statistics are provided in Table 1. Atomic coordinates and structure factors of the ^{Se}U-modified A-site RNA have been deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank, www.rcsb.org (PDB ID code 5T3K).

2.5 References:

1 (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev*. **2010**, *110*, 2579−2619. (b) Phelps, K.; Morris, A.; Beal, P. A. *ACS Chem. Biol*. **2012**, *7*, 100–109.

2 (a) Dethoff, E. A.; Chugh, J.; Mustoe, A. M.; Al-Hashimi, H. M. *Nature* **2012**, *482*, 322−330. (b) Bardaro Jr, M. F.; Varani, G. *WIREs RNA* **2012**, *3*, 122–132.

3 Nguyen, P.; Qin, P. Z. *WIREs RNA* **2012**, *3*, 62−72.

4 Serganov, A.; Patel, D. J. *Curr. Opin. Struct. Biol*. **2012**, *22*, 279–286.

5 (a) Wachowius, F.; Höbartner, C. *ChemBioChem* **2010**, *11*, 469−480. (b) Khakshoor, O.; Kool, E. T. *Chem. Commun*. **2011**, *47*, 7018–7024.

6 Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem*. **2013**, *53*, 366−378.

7 (a) Latham, M. P.; Brown, D. J.; McCallum, S. A.; Pardi, A. *ChemBioChem* **2005**, *6*, 1492−1505.(b) Hänsel, R.; Luh, L. M.; Corbeski, I.; Trantirek, L.; Dötsch, V. *Angew. Chem., Int. Ed*. **2014**, *53*, 10300–10314. (c) Salgado, G. F.; Cazenave, C.; Kerkour, A.; Mergny, J.-L. *Chem. Sci*. **2015**, *6*, 3314–3320.

8 (a) Shelke, S. A.; Sigurdsson, S. T. *Angew. Chem., Int. Ed*. **2010**, *49*, 7984−7986. (b) Helmling, C.; Bessi, I.; Wacker, A.; Schnorr, K. A.; Jonker, H. R. A.; Richter, C.; Wagner, D.; Kreibich, M.; Schwalbe, H. *ACS Chem. Biol*. **2014**, *9*, 1330−1339.

9 (a) Zhang, W.; Szostak, J. W.; Huang, Z. *Front. Chem. Sci. Eng*. **2016**, *10*, 196−202. (b) Westhof, E. *RNA* **2015**, *21*, 486–487. (c) Egli, M.; Pallan, P. S. *Annu. Rev. Biophys. Biomol. Struct*. **2007**, *36*, 281–305.

10 Mortimer, S. A.; Kidwell, M. A.; Doudna, J. A. *Nature Rev. Genet*. **2014**, *14*, 469–479.

11 Pawar, M. G.; Nuthanakanti, A.; Srivatsan, S. G. *Bioconjugate Chem*. **2013**, *24*, 1367–1377.

12 (a) Du, Q.; Carrasco, N.; Teplova, M.; Wilds, C. J.; Egli, M.; Huang, Z. *J. Am. Chem. Soc*. **2002**, *124*, 24–25. (b) Serganov, A.; Keiper, S.; Malinina, L.; Tereshko, V.; Skripkin, E.; Höbartner, C.; Polonskaia, A.; Phan, A. T.; Wombacher, R.; Micura, R.; Dauter, Z.; Jäschke, A.; Patel, D. J. *Nat. Struct. Mol. Biol*. **2005**, *12*, 218–224. (c) Höbartner, C.; Rieder, R.; Kreutz, C.; Puffer, B.; Lang, K.; Polonskaia, A.; Serganov, A.; Micura, R. *J. Am. Chem. Soc*. **2005**, *127*, 12035−12045. (d) Freisz, S.; Lang, K.; Micura, R.; Dumas, P.; Ennifar, E. *Angew. Chem., Int. Ed*. **2008**, *47*, 4110–4113. (e) Sheng, J.; Gan, J.; Soares, A. S.; Salon, J.; Huang, Z. *Nucleic Acids Res*. **2013**, *41*, 10476–10487.

13 Ennifar, E.; Carpentier, P.; Ferrer, J.-L.; Walter, P.; Dumas, P. *Acta Crystallogr., Sect D: Biol. Crystallogr*. **2002**, *58*, 1262–1268.

14 Deacon, A. M.; Ealick, S. E. *Structure* **1999**, *7*, R161–R166.

15 Moazed, D.; Noller, H. F. *Nature* **1987**, *327*, 389–394.

16 François, B.; Russell, R. J. M.; Murray, J. B.; Aboul-ela, F.; Masquida, B.; Vicens, Q.; Westhof, E. *Nucleic Acids Res*. **2005**, *33*, 5677–5690.

17 Ogle, J. M.; Carter, A. P.; Ramakrishnan, V. *Trends Biochem. Sci*. **2003**, *28*, 259–266.

18 (a) Fourmy, D.; Recht, M. I.; Blanchard, S. C.; Puglisi, J. D. *Science* **1996**, *274*, 1367–1371. (b) Wong, C.-H.; Hendrix, M.; Priestley, E. S.; Greenberg, W. A. *Chem.*

Biol. **1998**, *5*, 397–406. (c) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. *Nature* **2000**, *407*, 340–348. (d) Demeshkina, N.; Jenner, L.; Westhof, E.; Yusupov, M.; Yusupova, G. *Nature* **2012**, *484*, 256–259.

19 (a) Shandrick, S.; Zhao, Q.; Han, Q.; Ayida, B. K.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Vourloumis, D.; Hermann, T. *Angew. Chem., Int. Ed*. **2004**, *43*, 3177–3182. (b) Kaul, M. Barbieri, C. M.; Pilch, D. S. *J. Am. Chem. Soc*. **2004**, *126*, 3447–3453. (c) Srivatsan, S. G.; Tor, Y. *J. Am. Chem. Soc*. **2007**, *129*, 2044–2053. (d) Parsons, J.; Hermann, T. *Tetrahedron* **2007**, *63*, 3548–3552. (e) Chao, P.-W.; Chow, C. S. *Bioorg. Med. Chem*. **2007**, *15*, 3825–3831.

20 (a) Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 37−41. (b) Doose, S.; Neuweiler, H.; Sauer, M. *ChemPhysChem* **2009**, *10*, 1389–1398.

21 Vicens, Q.; Westhof, E. *Structure* **2001**, *9*, 647–658.

22 Zhao, F.; Zhao, Q.; Blount, K. F.; Han, Q.; Tor, Y.; Hermann, T. *Angew. Chem., Int. Ed*. **2005**, *44*, 5329–5334.

23 Vicens, Q.; Westhof, E. *Chem. Biol*. **2002**, *9*, 747–755.

24 Pfister, P.; Hobbie, S.; Vicens, Q.; Böttger, E. C.; Westhof, E. *ChemBioChem* **2003**, *4*, 1078–1088.

25 (a) Perez-Fernandez1, D.; Shcherbakov, D.; Matt, T.; Leong, N. C.; Kudyba, I.; Duscha, S.; Boukari, H.; Patak, R.; Dubbaka, S. R.; Lang, K.; Meyer, M.; Akbergenov, R.; Freihofer, P.; Vaddi, S.; Thommes, P.; Ramakrishnan, V.; Vasella, A.; Böttger, E. C. *Nat. Commun*. **2014**, *5*, 3112. (b) Kondo, J. *Angew. Chem., Int. Ed*. **2012**, *51*, 465–468.

26 Stagno, J. R.; Liu, Y.; Bhandari, Y. R.; Conrad, C. E.; Panja, S.; Swain, M.; Fan, L.; Nelson, G.; Li, C.; Wendel, D. R.; White, T. A.; Coe, J. D.; Wiedorn, M. O.; Knoska, J.; Oberthuer, D.; Tuckey, R. A.; Yu, P.; Dyba, M.; Tarasov, S. G.; Weierstall, U.; Grant, T. D.; Schwieters, C. D.; Zhang, J.; Ferré-D'Amaré, A. R.; Fromme, P.; Draper, D. E.; Liang, M.; Hunter, M. S.; Boutet, S.; Tan, K.; Zuo, X.; Ji, X.; Barty, A.; Zatsepin, N. A.; Chapman, H. N.; Spence, J. C. H.; Woodson, S. A.; Wang, Y.-X. *Nature* **2017**, *541*, 242–246.

27 Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem*. **1994**, *5*, 508–512.

28 Liebeskind, L. S.; Wang, J. *J. Org. Chem*. **1993**, *58*, 3550–3556.

29 Barbieri, C. M.; Kaul, M.; Pilch, D. S. *Tetrahedron* **2007**, *63*, 3567–3574.

30 (a) Tam, V. K.; Kwong, D.; Tor, Y. *J. Am. Chem. Soc.* **2007**, *129*, 3257–3266. (b) Tanpure, A. A.; Srivatsan. S. G. *Nucleic Acids Res*. **2015**, *43*, e16.

31 Otwinowski, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307–326.

32 McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007**, *40*, 658–674.

33 Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect D: Biol. Crystallogr.***1997**, *53*, 240–255.

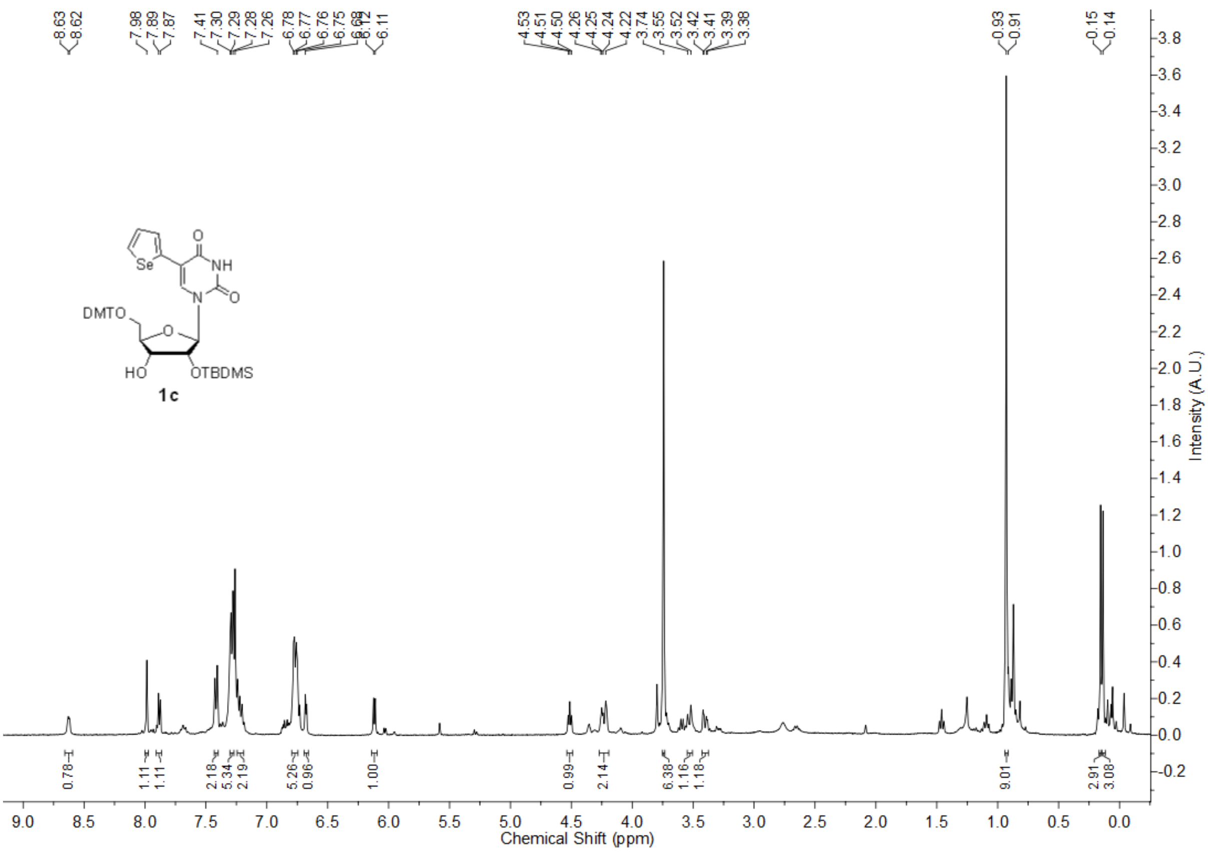
34 *Acta Crystallogr., Sect D: Biol. Crystallogr.***1994**, *50*, 760–763.

35 Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.

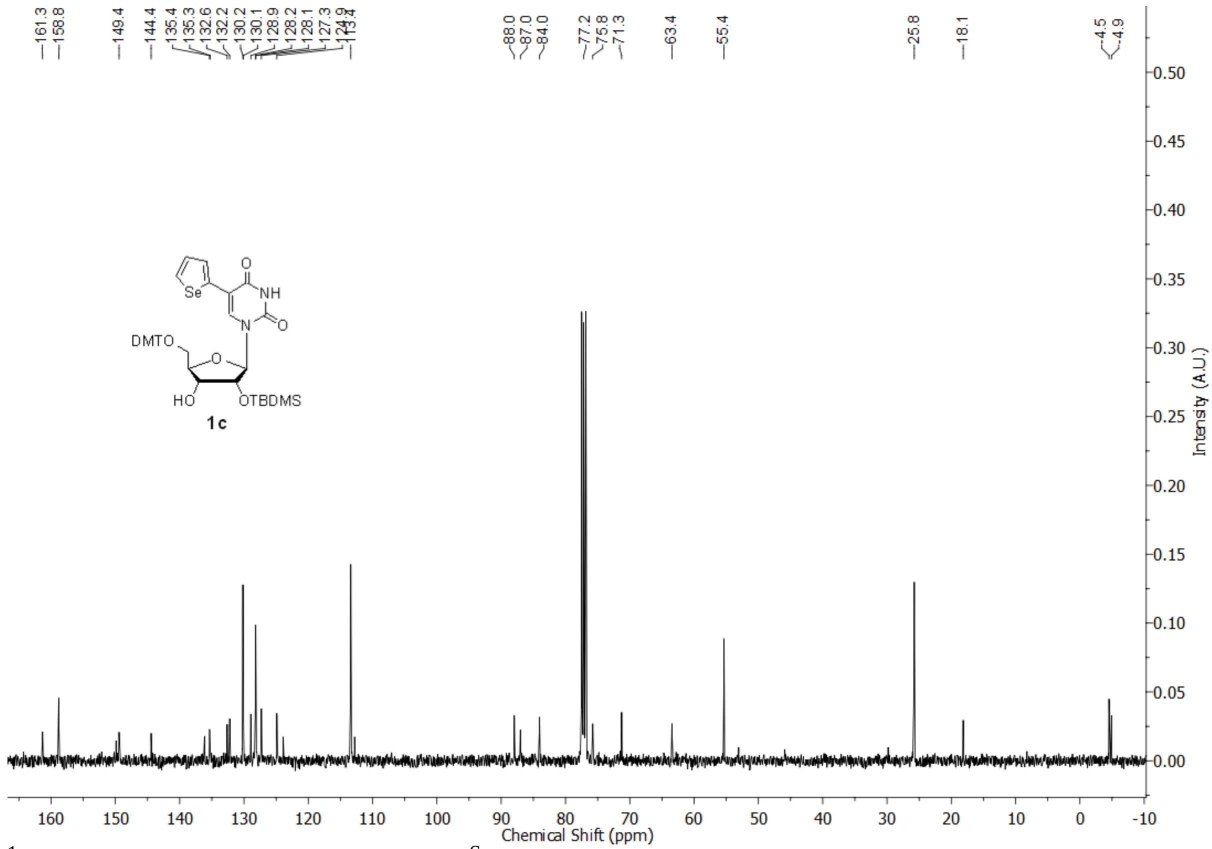
36 Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.;. Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2002**, *58*, 1948–1954.

2.6 Appendix-I: Characterization data of synthesized compounds

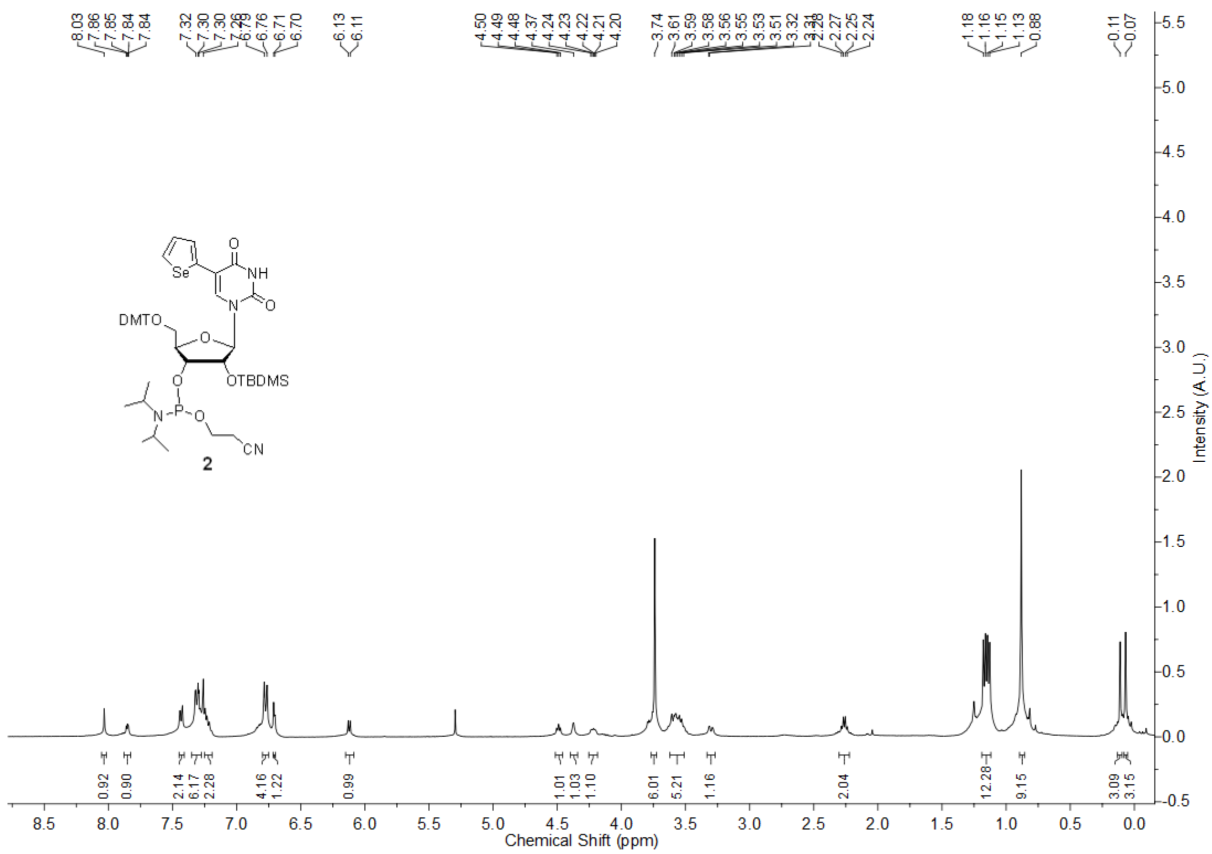
^{1}H NMR of the phosphoramidite of ^{Se}U **1c** in $CDCl_3$ (400 MHz)



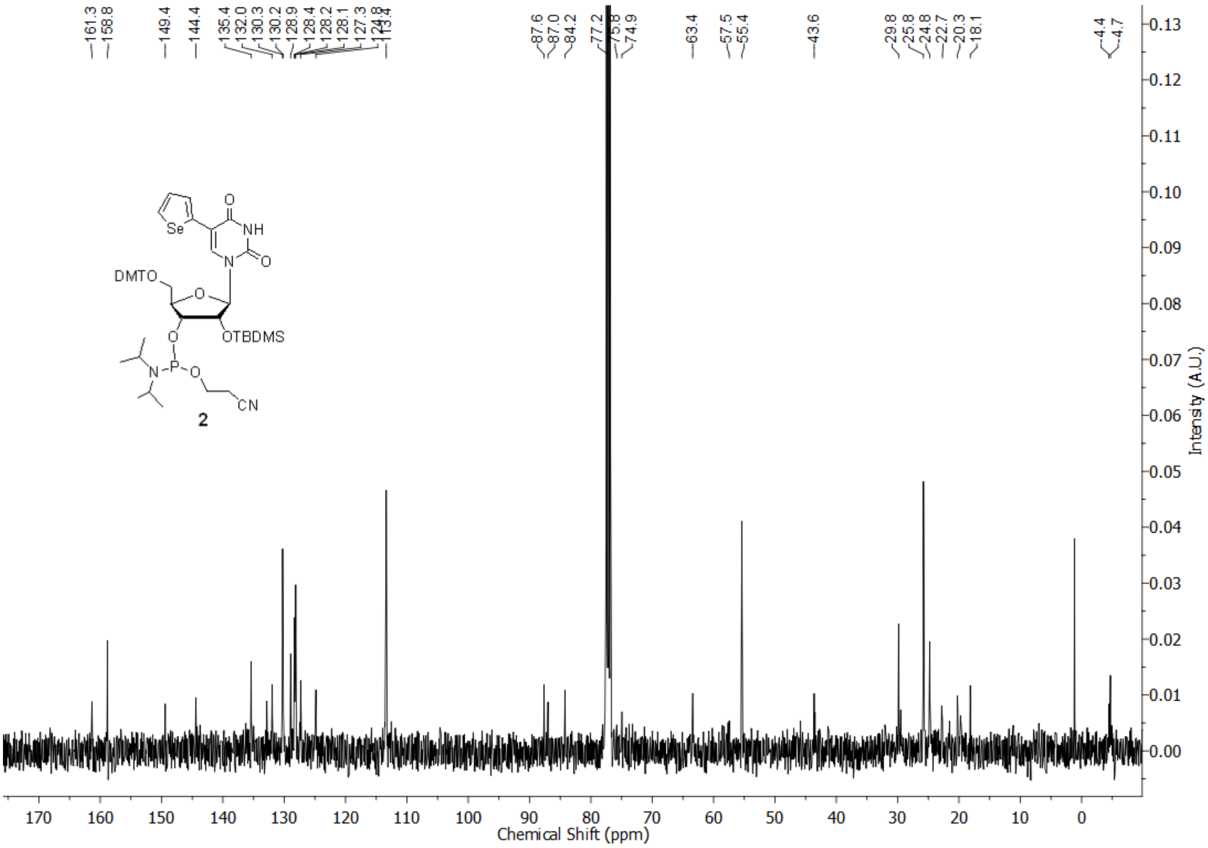
^{13}C NMR of the phosphoramidite of ^{Se}U **1c** in $CDCl_3$ (100 MHz)



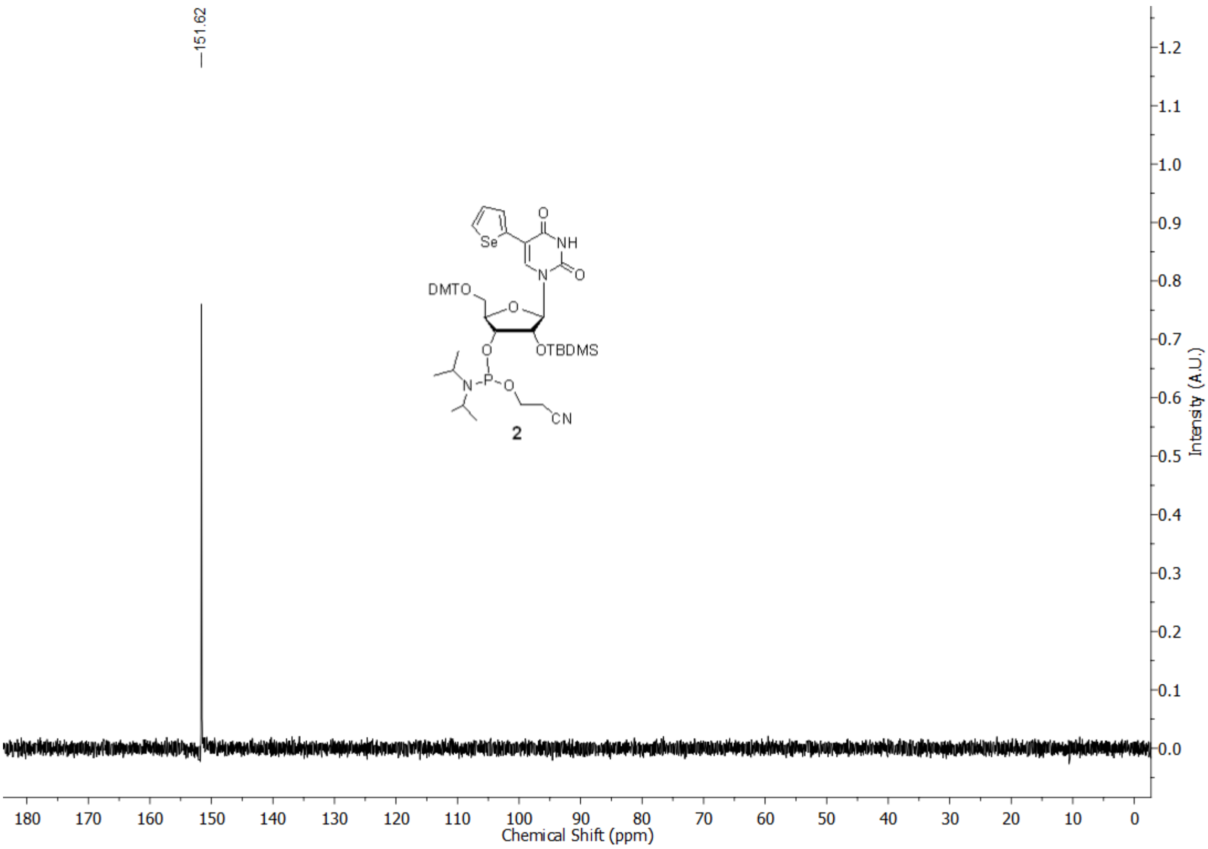
^{1}H NMR of the phosphoramidite of ^{Se}U **2** in $CDCl_3$ (400 MHz)



^{13}C NMR of the phosphoramidite of ^{Se}U **2** in $CDCl_3$ (100 MHz)



^{31}P NMR of the phosphoramidite of ^{Se}U **2** in $CDCl_3$ (162 MHz)



Chapter 3

Probing G-quadruplex structures and ligand binding using selenophene-modified 2′-deoxyuridine

3.1 Introduction

DNA and RNA sequences containing closely placed guanine residues are known to adopt non-canonical four-stranded structures called G-quadruplexes (GQs).[1] Bioinformatics studies indicate that such putative GQ forming sequences are abundantly found in the telomeres, DNA promoter regions and untranslated regions of RNA.[2] Further, conservation of their location in the genome across most eukaryotes suggested that these structural motifs could play important gene regulatory roles.[3] Recent biochemical and biophysical investigations strongly support this notion as GQ-forming motifs have been shown to play important roles in the maintenance of chromosomes and in the regulation of proliferation associated genes.[4] For this reason, GQ is pursued as an important nucleic acid target for the development of therapeutic agents.[5]

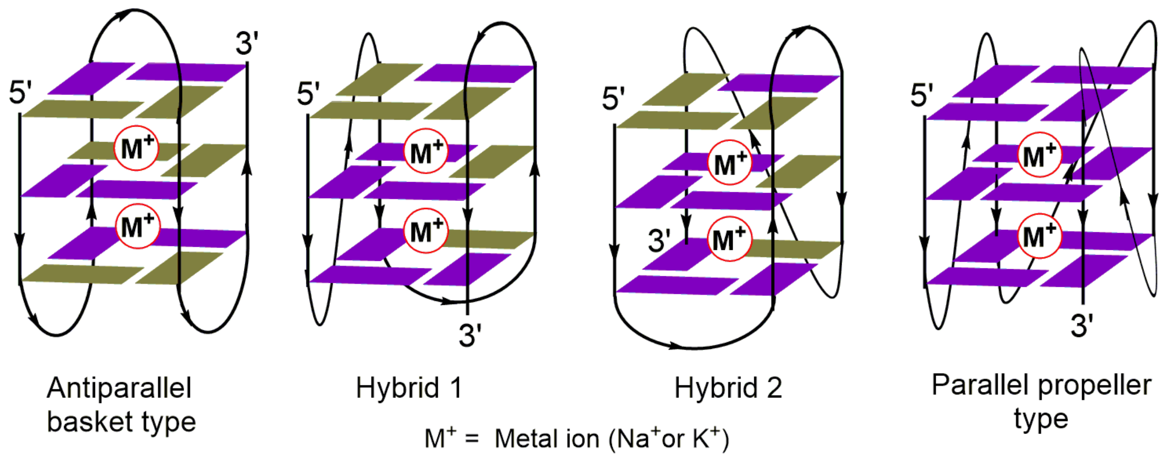


Figure 1. Polymorphism of intramolecular H-Telo DNA quadruplexs formed by four-repeat human telomeric sequences, olive green and purple boxes indicate guanine bases in the *syn* and *anti* conformations, respectively.

Several biophysical tools have been developed to study the structure, dynamics, small-molecule recognition properties and functions of GQ-forming sequences.[6] GQ exhibits a variety of folding topologies *in vitro*, which depend on the sequence,[7] monovalent cations[8] and molecular crowding.[9] Depending on the number of contiguous G residues and how they are spaced, a sequence can adopt several of the GQ structures, which could be made of two or three tetrads.[10] For example, the overhang of human telomeric (H-Telo) DNA repeat $(TTAGGG)_n$ forms parallel, anti-parallel and hybrid type parallel-antiparallel stranded structures (Figure 1).[11] In presence of Na^+ ions, it adopts predominantly an antiparallel structure,[12] whereas in the presence of K^+ ions, it adopts mixture of structures with major being hybrid type structures.[13] In synthetic crowded environment like PEG[14] and in solid state, the H-Telo DNA forms a parallel structure.[15] Interestingly, equivalent RNA repeat (TERRA) forms predominantly a parallel structure irrespective of the ionic conditions.[16]

GQ structure, stability and its interaction with small molecules are usually studied by using FRET systems,[17] and fluorescent ligands[18] and metal complexes[19] that bind and stabilize GQs. Alternatively, fluorescent nucleoside probes incorporated into tetrad or loop positions also enable the study of GQs.[20] Diederichsen and co-workers have reported fluorescent guanine analog 8-vinyl-2′-deoxyguanosine, by attaching a vinyl group at the 8-position of guanine. This analog capable of acquire both the *syn* and *anti* conformation, which makes it amenable to study the different topologies of G-quadruplexes.[21] Manderville and co-workers, developed emissive guanosine derivatives which they employed to investigate GQ folding by using C8-aryl-dG or 8-(2″-furyl)-2′-dG nucleoside analog.[22] Luedtke *et al*. have demonstrated that 8-(2-pyridyl)-2′-deoxyguanosine, serves as turn-on fluorescent probe which enable the monitoring of G-quadruplex folding and intramolecular energy transfer process, in the presence of a monovalent ion-containing.[23] Recently, our group developed a conformational sensitive fluorescence nucleoside analog (5-benzofuran-2′-deoxyuridine) was selectively incorporated into one of the loop residues of human telomeric DNA ONs, which is not only distinguished between the quadruplex versus duplex, but also discriminated various topologies by changes in the fluorescence.[24]

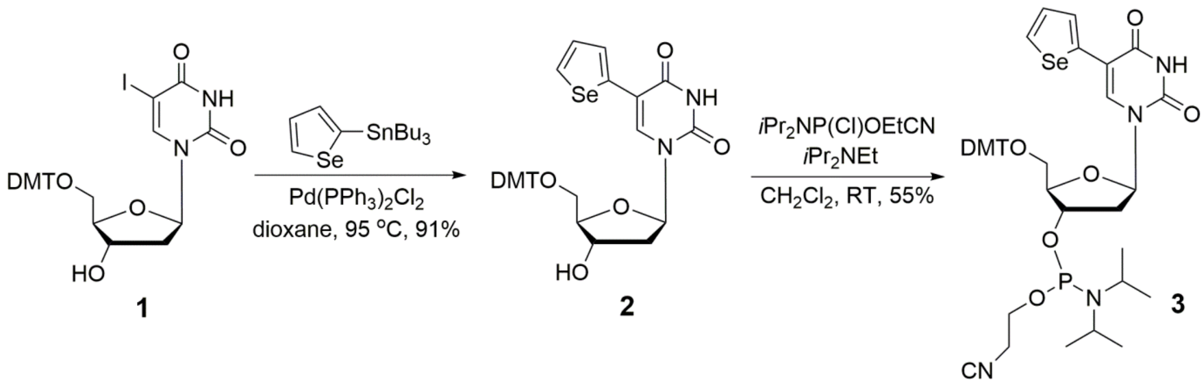
Further, recent studies using structure-specific antibodies[25] and fluorescent probes,[20] for sure, indicate the formation of GQ structures in cellular environment. Several small molecule ligands, which bind and stabilize GQs or induce the formation GQ structures, have been developed. While many of these ligands have been useful in understand the biological role of GQs, a very few have shown promise as clinically potential candidates for chemotherapy.[26] This we believe is because, most of the currently available tools cannot efficiently discriminate different GQ structures based on topology and nucleic acid type, and hence cannot be efficiently used in estimating and comparing the binding of ligands to different GQ structures. Further, we hypothesized that the development of a dual-purpose probe, which can be simultaneously used to understand the structure and ligand binding of GQs in real time and 3D, will not only be useful in advancing our understanding of the GQ structure but also could aid GQ-based therapeutic strategies.

In this context, the utility of selenophene-modified nucleoside analog in investigating the structure and recognition properties of GQ-forming sequences by fluorescence and X-ray crystallography was explored. The H-Telo DNA repeat $(TTAGGG)_n$, which end-caps the chromosomes, maintains the chromosomes from end-to-end fusion and degradation, was chosen as the test system. The overhang of this sequence is known to adopt GQ structures, and several biophysical studies have been conducted on this repeat sequence. ^{Se}dU incorporated into the loop region of the H-Telo DNA ONs fluorescently reported the formation of GQ structures with significant enhancement in fluorescence intensity as compared to the duplex. Further, we could successfully establish an assay to monitor and estimate the ligand binding to the GQ by fluorescence technique. The H-Telo DNA ONs containing the ^{Se}dU modification at different positions within a second loop were crystallized. 3D structures determined by X-ray diffraction indicated that the ON adopted a parallel GQ conformation. Superimposition of the modified GQ and native GQ indicated that the modification had only minor impact on the structure. Further analysis of the structure provided atomic level insights into the structural basis of the GQ sensing ability of the fluorescent nucleoside probe.

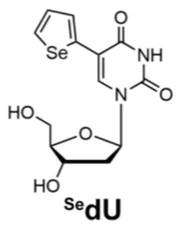
3.2 Results and discussion

3.2.1 Synthesis of 5-selenophene 2′-deoxyuridine-modified H-Telo DNA G-quadruplexes

The loop residues (TTA) of H-Telo DNA show considerable differences in their conformation in different topologies of GQ and upon binding of planar heteroaromatic ligand.[27] Based on this understanding, we decided to incorporate ^{Se}dU in different loop regions of H-Telo DNA repeat (ONs **4–7**, **10**, Scheme 1). The phosphoramidite **3** was synthesized as shown in Scheme 1. 5′-*O*-DMT protected 5-iodouridine was reacted with 2-(tri-*n*-butyl stannyl) selenophene in the presence of a palladium catalyst to give compound **2**. A reaction in the presence of 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite gave the desired phosphoramidite substrate required for the incorporation of ^{Se}dU into DNA ONs by solid-phase ON synthesis. ONs **4–7**, **10** were synthesized by standard solid-phase ON synthesis cycle, and were purified by PAGE. The purity and integrity of the ^{Se}dU-labeled ONs were confirmed by HPLC and mass analysis (Figure 2 and Table 1).



4 5' d(AGGG^{Se}**U**TAGGGTTAGGGTTAGGG) 3'
5 5' d(AGGGTTAGGG^{Se}**U**TAGGGTTAGGG) 3'
6 5' d(AGGGTTAGGGT^{Se}**U**AGGGTTAGGG) 3'
7 5' d(AGGGTTAGGGTTAGGG^{Se}**U**TAGGG) 3'
8 5' d(AGGGTTAGGGTTAGGGTTAGGG) 3'
9 5' d(CCCTAACCCTAACCCTAACCCT) 3'
10 5' d((AGGGTTAGGG^{Se}**U**TAGGGTT)$_2$AGGG) 3'



^{Se}**dU**

Scheme 1. Synthesis of 5-selenophene-modified 2′-deoxyuridine phosphoramidite substrate **3** for the solid-phase synthesis of H-Telo DNA repeats. DMT = 4,4′-dimethoxytrityl, *i*Pr_2NP(Cl)OEtCN = 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite. The ^{Se}dU labeled H-Telo DNA repeats ONs containing the modification at position 5 (**4**), 11 (**5**), 12 (**6**) and 17 (**7**), respectively. ON **8** is a control unmodified H-Telo DNA and **9** is complementary to H-Telo DNA **4–8**. Longer sequence of H-Telo DNA repeats with double modification used in this study **10**.

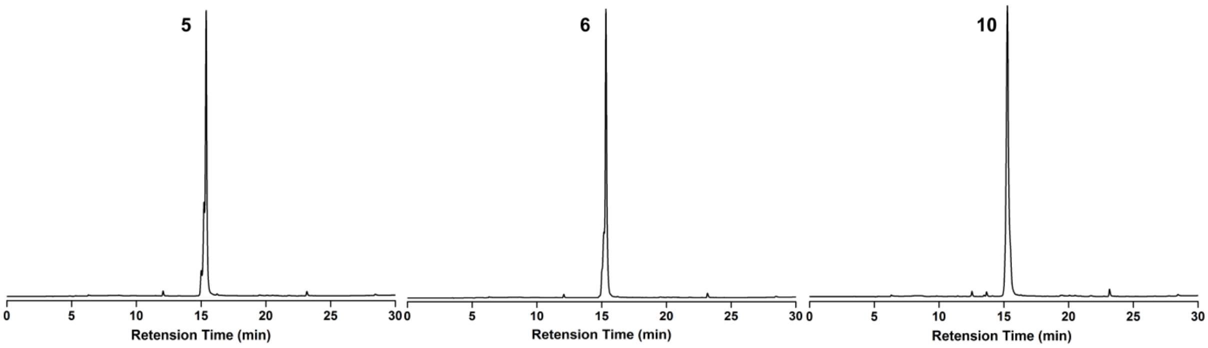


Figure 2. HPLC chromatograms of PAGE purified fluorescent ONs **5**, **6** and **10** at 260 nm. Mobile phase A = 100 mM triethylammonium acetate buffer (pH 7.5), mobile phase B = acetonitrile. Flow rate = 1 mL/min. Gradient = 0–10% B in 10 min and 10–100% B in 20 min. HPLC analysis was performed using Luna C18 column (250 x 4.6 mm, 5 micron).

Table 1. Extinction coefficient and mass of modified H-Telo DNA ONs **4**–**7** and **10**

Modified ON	ε_{260} ($M^{-1}cm^{-1}$)[a]	Calcd. mass	Observed mass
4	228676	7081.5	7082.4
5	228676	7081.5	7080.2
6	228676	7081.5	7081.6
7	228676	7081.5	7081.5
10	412152	12924.2	12924.5

[a]Molar absorption coefficient of modified ONs was determined by using OligoAnalyzer 3.1, which was used for the determination of concentration of modified ONs. The extinction coefficient of nucleoside (ε_{260} = 8576 $M^{-1}cm^{-1}$) was used in place of thymidine, respectively.

3.2.2 Circular dichroism and thermal melting studies

The formation of GQ structure and the effect of selenophene modification on the GQ DNA structure and stability were studied by employing circular dichroism (CD) and thermal melting experiments. H-Telo DNA in the presence of Na^+ ions forms an antiparallel basket-type topology, whereas in the presence of K^+ ions forms mixed parallel-antiparallel stranded hybrid type structures.[28] The CD profiles of modified and control unmodified H-Telo DNA ONs in Na^+ ionic condition were found to be similar and characteristic of an antiparallel GQ structure (positive peaks at 290 nm and 240 nm and a negative peak at 260 nm, Figure 3A and 3B). Similarly, characteristic CD pattern for hybrid type structures was displayed by the ONs in the presence of K^+ ions. The thermal melting analysis of unmodified and modified GQ structures of ONs in different salt (NaCl/KCl) conditions gave similar T_m values, which matched well with the literature reports (Figure 3B and Table 2).[29] The above CD and thermal melting measurements clearly indicate that ^{Se}dU modified H-Telo DNA ONs form respective GQ structures and the modification is only minimally perturbing in terms of structure and stability (Figure 3 and Table 2).

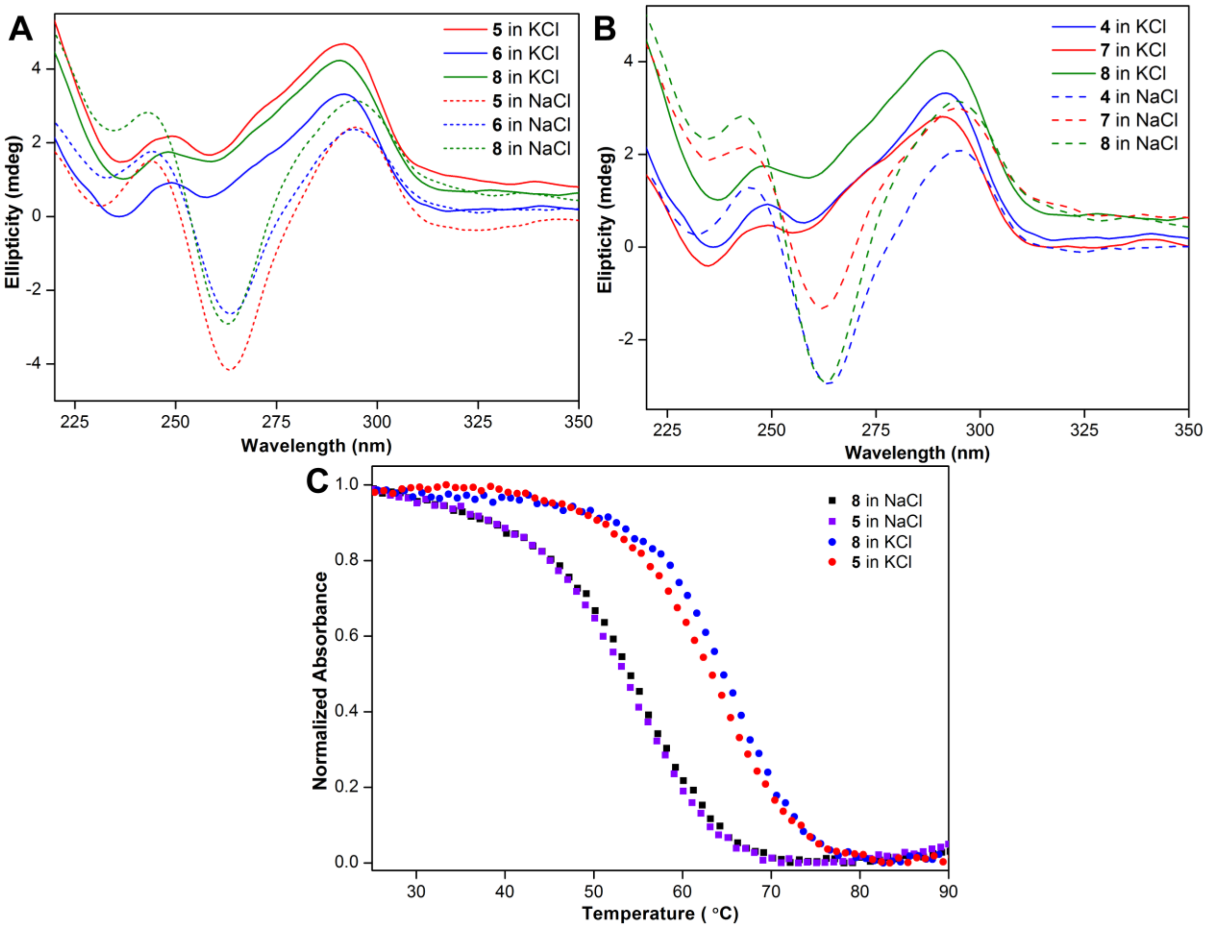


Figure 3. (**A**) CD spectra of fluorescently modified ON **5** (6 μM), **6** (6 μM) and control unmodified **8** (6 μM) in 10 mM Tris-HCl buffer (pH 7.5) containing either 100 mM KCl or 100 mM NaCl. (**B**) CD spectra of fluorescently modified ON **4** (6 μM), **7** (6 μM) and control unmodified **8** (6 μM) in 10 mM Tris-HCl buffer (pH 7.5). Both unmodified and modified ONs show similar CD profiles in respective ionic conditions. (**C**) UV-thermal melting profiles of fluorescently modified ON **5** and control unmodified **8** in 10 mM Tris-HCl buffer (pH = 7.5) containing 100 mM NaCl or KCl at 295 nm.

Table 2. T_m values of fluorescently modified H-Telo (**4–7**) and control unmodified H-Telo GQs (**8**) in the presence of KCl and NaCl.

H-Telo GQ	T_m (°C)
4 in NaCl	52 ± 0.8
4 in KCl	65 ± 0.9
5 in NaCl	54 ± 0.9
5 in KCl	65 ± 0.7
6 in NaCl	53 ± 0.8
6 in KCl	67 ± 0.9
7 in NaCl	52 ± 0.8
7 in KCl	65 ± 0.8
8 in NaCl	56 ± 0.8
8 in KCl	67 ± 0.7

3.2.3 Fluorescence detection of GQ structure of H-Telo DNA ONs

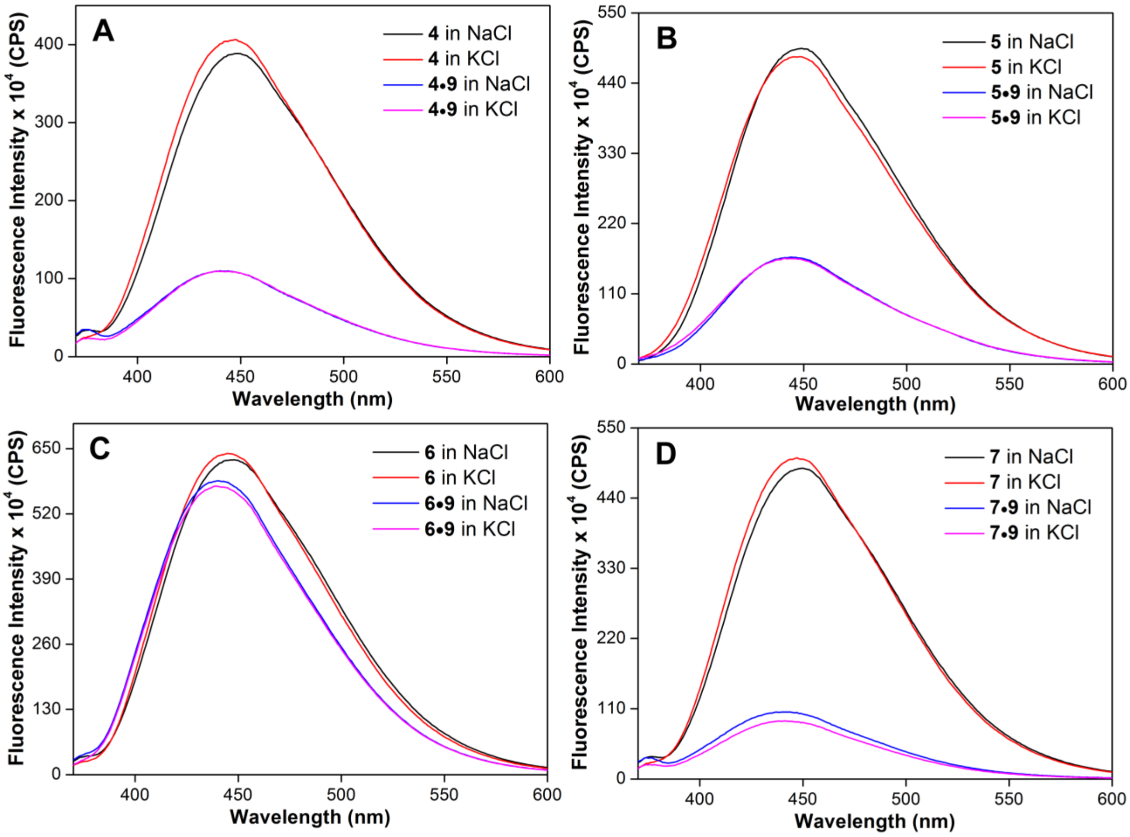


Figure 4. Steady-state fluorescence spectra of H-Telo DNA ON (**A**) **4** and corresponding duplex **4•9**, (**B**) **5** and corresponding duplex **5•9**, (**C**) **6** and corresponding duplex **6•9** and (**D**) **7** and corresponding duplex **7•9** in 10 mM Tris-HCl buffer (pH 7.5) containing 100 mM NaCl or 100 mM KCl. DNA ON samples (1 µM) were excited at 330 nm with an excitation and emission slit width of 5 nm and 9 nm, respectively.

Fluorescence of annealed H-Telo DNA ONs **4**–**7** was recorded and compared with that of the corresponding duplexes in different ionic conditions. G-quadruplex of ONs **4**, **5** and **7** in which the emissive analog is placed in the first (dT5), second (dT11) and third (dT17) loops, respectively, displayed significant enhancement in fluorescence intensity (3–5 fold) as compared to the corresponding duplexes (Figure 4). Notably, the fluorescence profile of GQs in NaCl and KCl was found to be similar. While this observation indicates that ^{Se}dU in these positions distinguishing between quadruplex and duplex structures, it poorly distinguished individual topologies of GQ formed in different ionic conditions. The formation of G-quadruplex structures can be predicted by changes in emission maximum and fluorescence intensity. ON **4**, **5** and **7** reported the formation of GQs and distinguished them from duplex via changes in intensity and emission maximum. However, ON **6** containing the selenophene label in the second thymidine residue of the loop distinguished the GQ from duplex structure by a discernible change in emission maximum (Figure 4C). These results indicate that ^{Se}dU, depending on the position of modification, can report the formation of GQ structure.

3.2.4 Probing ligand binding in aqueous buffers

Based on biochemical and structural analysis, H-Telo DNA GQs trinucleotide loops are highly flexible, and undergo substantial structural reorganisation during ligand binding event.[27] In order to probe ligand binding, we performed fluorescence binding assay in 10 mM Tris-HCl buffer (pH 7.5) containing either 100 mM NaCl or 100 mM KCl to assess the probe response to ligand-induced conformational change in H-Telo structures. In this regard, we have chosen two well known GQ binders pyridostation (PDS) and BRACO19, which are used in biophysical and therapeutic analysis of different GQ structures (Figure 5).[30]

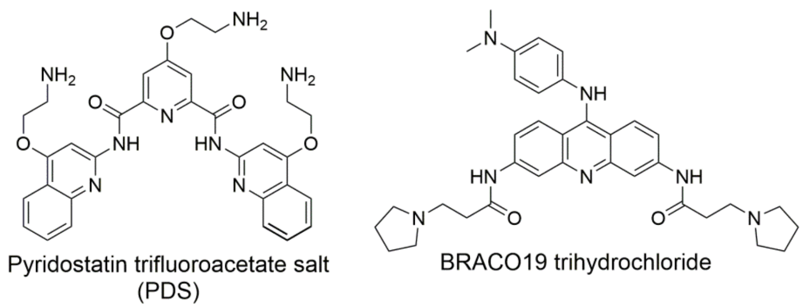


Figure 5. Chemical structure of GQ binders, pyridostatin (PDS) and BRACO19.

Fluorescent H-Telo DNA **5** was annealed into antiparallel and hybrid-type GQs in buffer having NaCl and KCl, respectively. Upon increasing the concentration of PDS in the presence of an antiparallel GQ of ON **5**, a near 2-fold reduction in fluorescence intensity was observed at the saturation point of PDS (10 μM, Figure 6A). An apparent K_d of 1.77 ± 0.16 μM was obtained (Figure 6B). Also, binding of PDS to hybrid-type GQ in K^+ promoted dose-dependent decrease in fluorescence intensity of nearly 2-fold (10 μM, Figure 6C). The observed K_d values of 2.34 ± 0.13 μM (Figure 6D) reveals that PDS has higher binding affinity for antiparallel H-Telo DNA GQ **5** as compared to hybrid-type H-Telo GQ **5** structure (Table 3). Furthermore, the fluorescence response of ^{Se}dU probe with BRACO19 binding was studied in both antiparallel and hybrid-type GQ forms. Upon addition of BRACO19 to ON **5**, a progressive quenching in fluorescence intensity was noticed (Figure 7). A higher binding affinity for hybrid-type GQ topology was observed as compared to antiparallel GQ form of H-Telo DNA **5** (Table 3). In comparison to BRACO19, PDS has the ability to acquire a planar and flexible conformation in the presence of GQs which makes PDS more selective for binding to anti-parallel strucrture.[31] BRACO19 ligand being a planar molecule, stacks G-tetrad as it binds more effectively to hybrid-type GQ topology. This notion is consistent with the binding data obtained from fluorescence experiment.

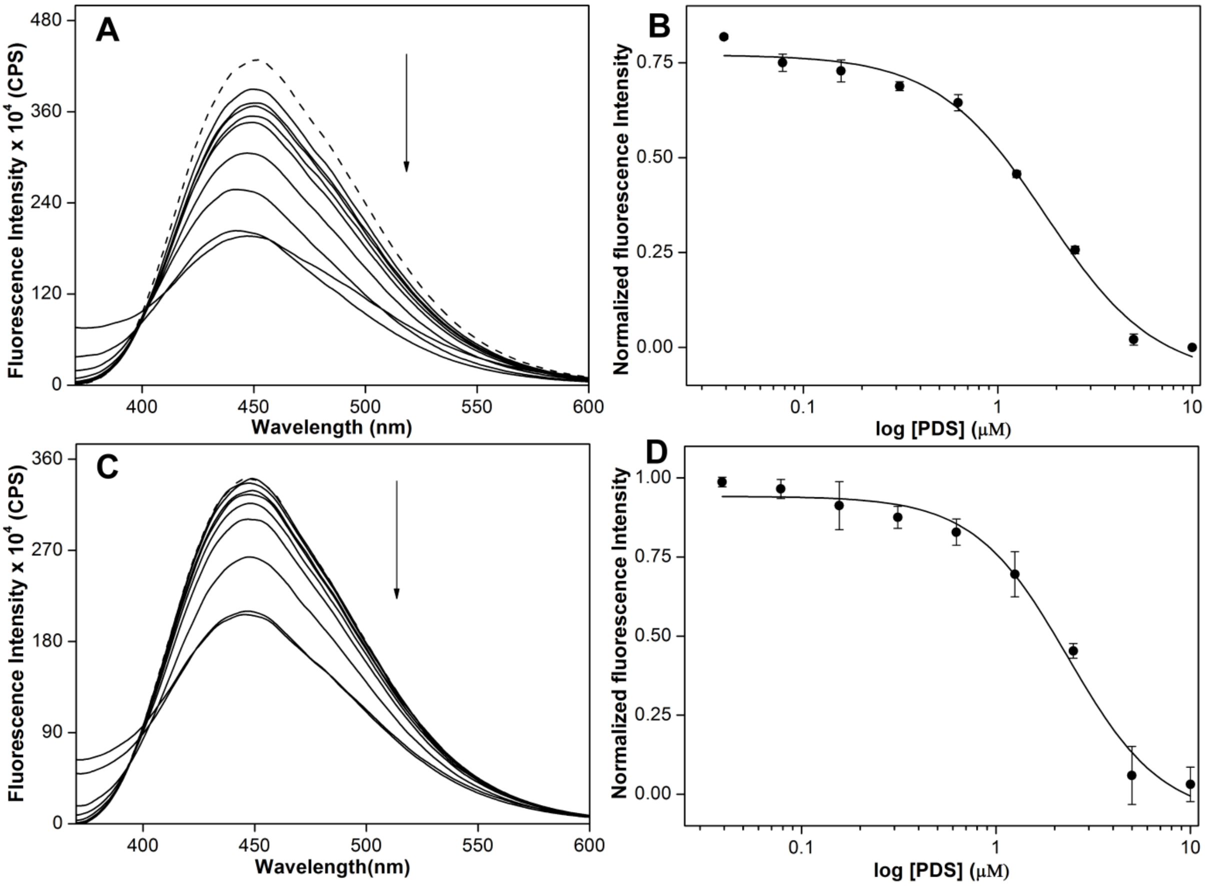


Figure 6. Emission spectra of H-Telo DNA ONs **5** antiparallel (**A**) and hybrid-type (**C**) in aqueous buffer (pH = 7.5) in the presence of Na^+ or K^+ ions as a function of increasing concentration of PDS. Dashed line indicates fluorescence spectrum of GQs in the absence of PDS. The ONs (1 μM) sample were excited at 330 nm with excitation and emission slit width of 5 nm and 9 nm, respectively. Curve fit for the binding of PDS to H-Telo DNA **5** in the presence of NaCl (**B**) and KCl (**D**). Normalized fluorescence intensity at λ_{em} = 450 nm is plotted against log [PDS].

Table 3. Binding constant (K_d) for PDS and BRACO19 binding to H-Telo DNA **5** and **10**

ON	PDS (μM)		BRACO19 (μM)	
	in NaCl	in KCl	in NaCl	in KCl
5	1.77 ± 0.16	2.34 ± 0.13	1.53 ± 0.07	1.21 ± 0.09
10	2.86 ± 0.29	3.67 ± 0.43	0.65 ± 0.03	0.64 ± 0.02

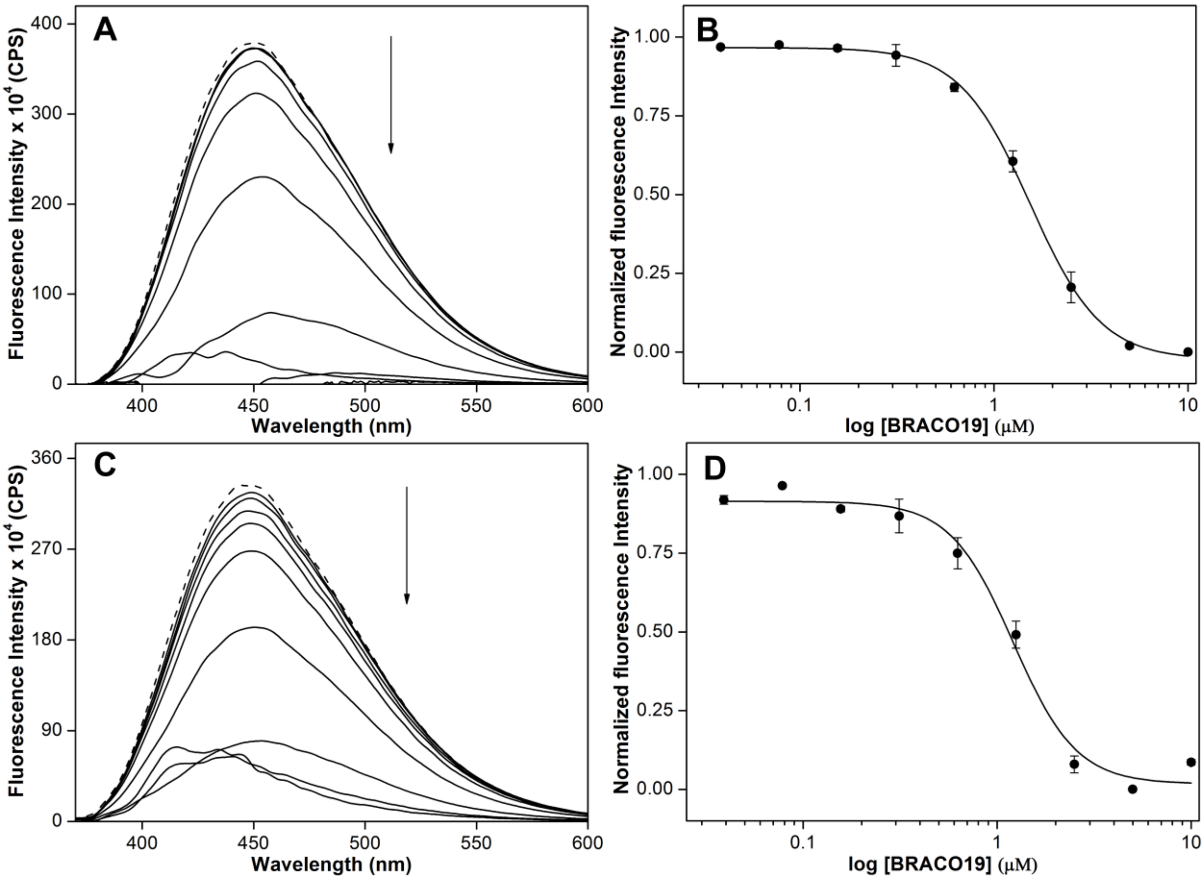


Figure 7. Emission spectra of H-Telo DNA ONs **5** antiparallel (**A**) and hybrid-type (**C**) in aqueous buffer (pH = 7.5) in the presence of Na^+ or K^+ ions as a function of increasing concentration of BRACO19. Dashed line indicates fluorescence spectrum of GQs in the absence of BRACO19. The ONs (1 μM) sample were excited at 330 nm with excitation and emission slit width of 5 nm and 9 nm, respectively. Curve fit for the binding of BRACO19 to H-Telo DNA **5** in the presence of NaCl (**B**) and KCl (**D**). Normalized fluorescence intensity at λ_{em} = 450 nm is plotted against log [BRACO19].

3.2.5 Higher-order H-Telo DNA GQ structures and ligand binding

The human telomeric overhang composed of a G-rich hexameric repeat (T_2AG_3) has the ability to form higher-order GQ structure (e.g., intra or intermolecular). Such structures are attractive targets for therapeutic intervention and also they play key roles in recognition and telomeric DNA length maintenance.[32] Based on this information, a longer ^{Se}dU-labeled H-Telo DNA repeat ON **10**, which can form two consecutive GQ units, was synthesized. The higher order of H-Telo DNA **10** carrying two modifications potentially forms two successive GQ units. The CD analysis of ON **10** in the presence of NaCl indicates the formation of an antiparallel GQ topology (positive peak at 295 nm and 245 nm and negative peak at 265 nm). In the presence of KCl, CD signatures for hybrid type and parallel GQ structures were observed (Figure 8A).[28]

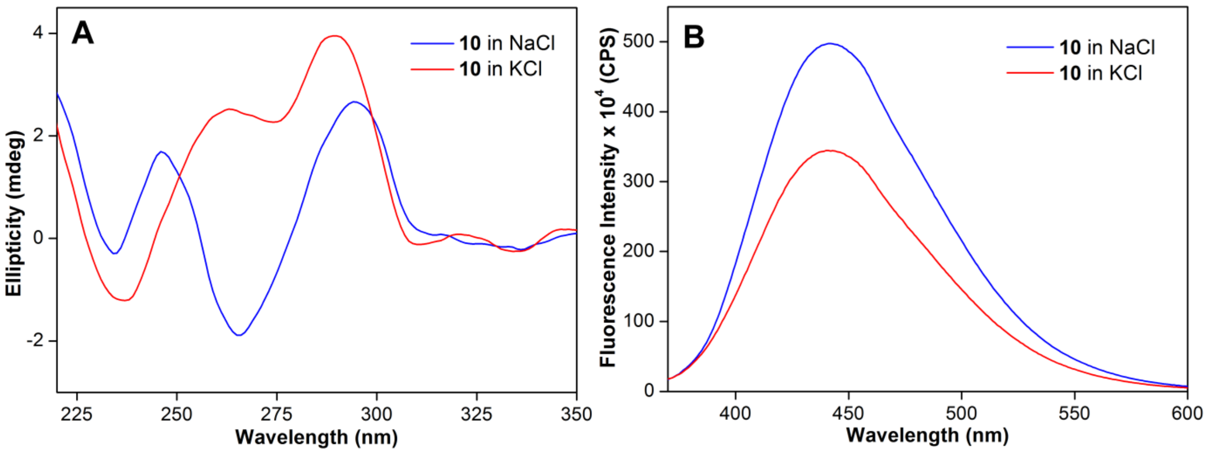


Figure 8. (**A**) CD and (**B**) fluorescence spectra of doubly modified longer H-Telo DNA repeat **10** in 10 mM Tris-HCl buffer (pH 7.5) containing either 100 mM KCl or 100 mM NaCl. CD and fluorescence profiles of ON **10** in the presence of KCl and NaCl were found to be different to H-Telo DNA **5**.

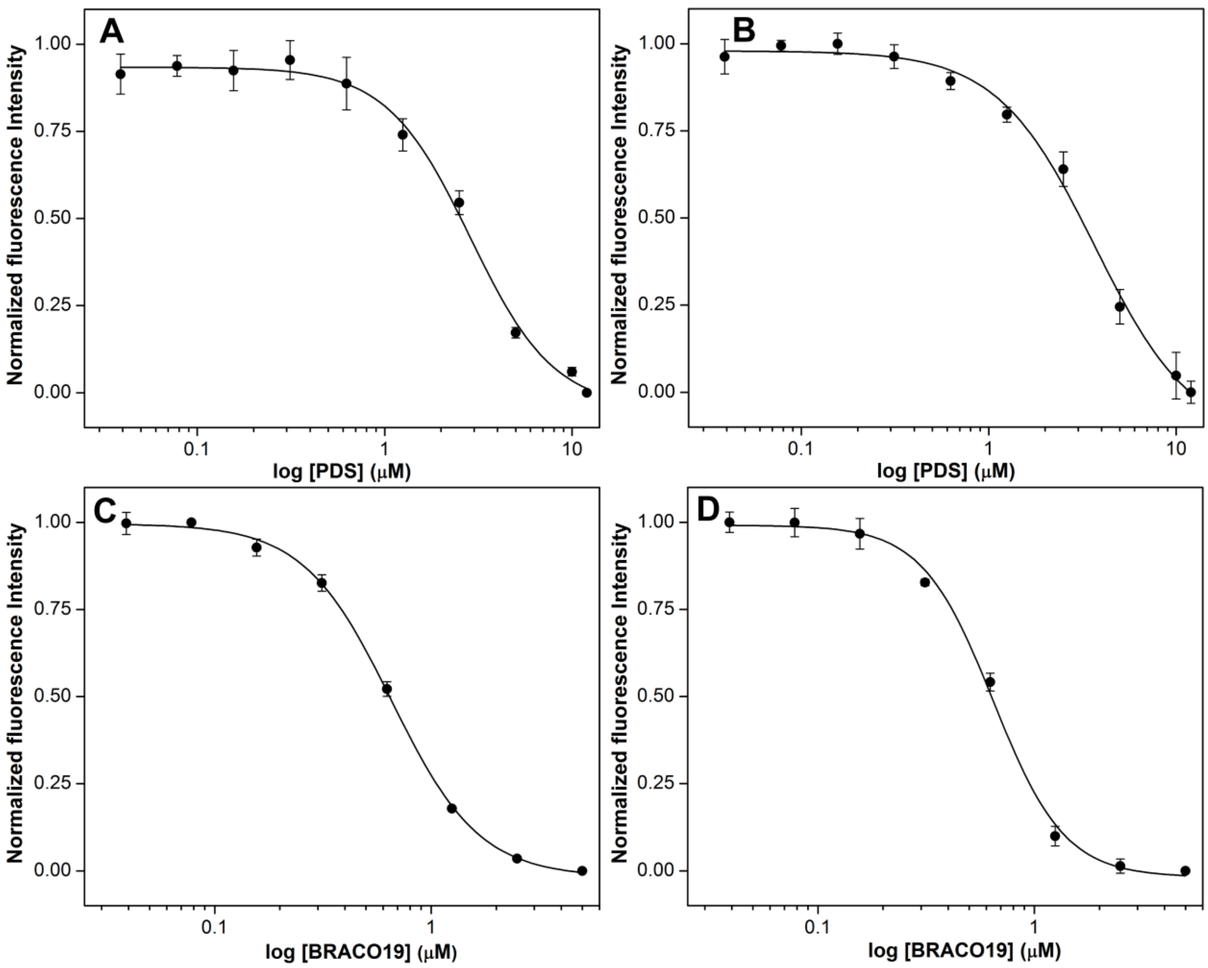


Figure 9. Curve fit for the binding of PDS to H-Telo DNA **10** in the presence of NaCl (**A**) and KCl (**B**) and BRACO19 in the presence of NaCl (**C**) and KCl (**D**). Normalized fluorescence intensity at λ_{em} = 442 nm is plotted against log of [ligand]. The samples were excited at 330 nm with an excitation and emission slit width of 5 nm and 6 nm, respectively.

Unlike 22-mer H-Telo DNA **5**, longer ON **10** displayed different fluorescence intensity in NaCl and KCl conditions (Figure 8B). In ligand binding studies, when the concentration of PDS or BRACO19 was increased in presence of ON **10**, a dose-dependent decrease in the fluorescence intensity was noticed. The resultant binding constant in the presence of PDS was distinct in Na^+ and K^+ conditions but in case of BRACO19 nearly identical binding constant in both Na^+ and K^+ ionic conditions was observed (Figure 9 and Table 3).

3.2.6 ^{Se}dU reports the formation GQ in intraoocyte conditions

3.2.6.1 Circular dichroism and thermal melting studies

CD and thermal melting studies were first performed to investigate the effect of modification on the structure and stability of GQs of ONs **4–8**. The CD spectrum of control (**8**) in intraoocyte buffer solution[33] (25 mM HEPES buffer (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 0.13 µM $CaCl_2$ and 1 mM $MgCl_2$) shows a distinctly broad positive band peaking at 290 nm and a small negative peak at 235 nm. This spectrum is considered a summation of multiple GQ structures composing mainly of mixed hybrid topologies (black solid line). Similar kind of spectrum is displayed in case of modified ONs **4**, **5**, **6** and **7** (Figure 10A). In case of higher order structure forming sequence (**10**) the CD profile in intraoocyte conditions favour the formation of mixed-parallel structure. A prominent peak at 265 nm indicated the presence of parallel GQ structure and 290 nm indicated hybrid structures (Figure 10B).[28] The CD measurements clearly indicates that the selenophene modification does not perturb the native H-Telo topology.

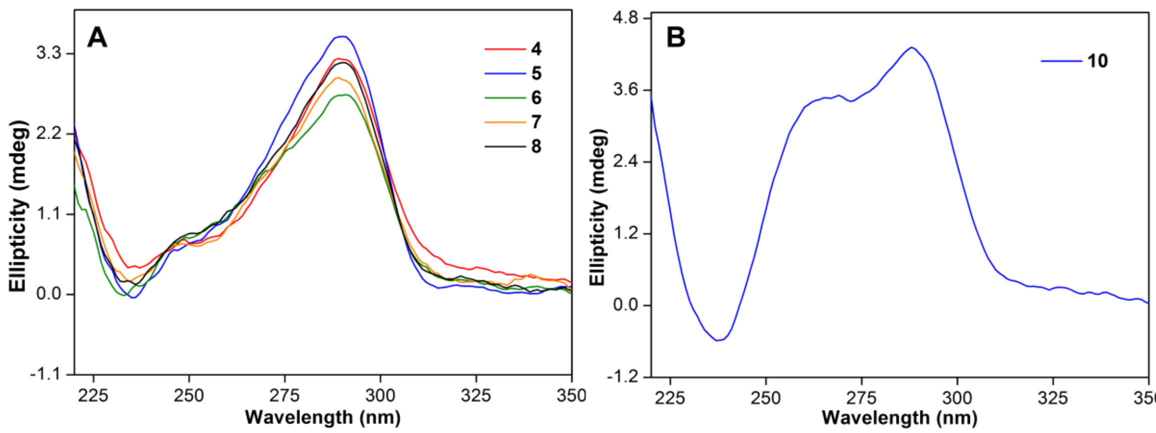


Figure 10. (**A**) CD spectra of fluorescently modified H-Telo DNA **4**, **5**, **6**, **7** (6 µM) and control unmodified H-Telo DNA **8** (6 µM) in intraoocyte buffer (pH 7.5). Both unmodified and modified ONs show similar CD profiles in respective ionic conditions. (**B**) CD spectra of doubly modified longer H-Telo DNA repeat **10** in intraoocyte buffer (pH 7.5).

UV-thermal melting values of unmodified and modified DNA GQs in intraoocyte buffer were found to be similar (Figure 11 and Table 4).[29] These results indicate that the selenophene modification does not affect on GQ stability.

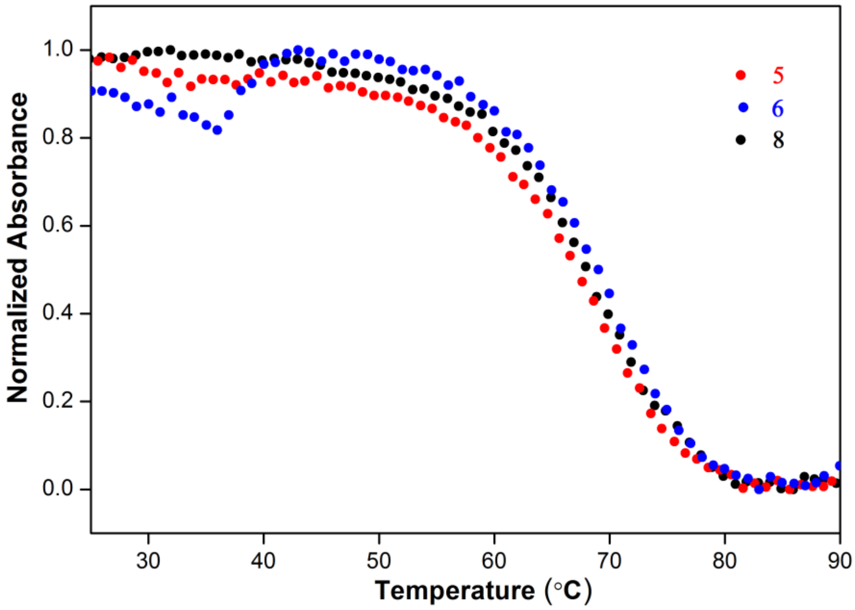


Figure 11. UV-thermal melting profile of fluorescent H-Telo DNA ONs **5** and **6** (1 μM) and control unmodified DNA ON **8** (1 μM) in intraoocyte buffer (pH 7.5).

Table 4. T_m values of fluorescently modified and control unmodified GQs in the intraoocyte conditions.

H-Telo GQ	T_m (°C)
4	69 ± 0.4
5	70 ± 0.8
6	70 ± 1.0
7	69 ± 0.8
8	70 ± 0.2

3.2.6.2 Fluorescence detection of H-Telo DNA

Fluorescence response of modified ONs **4**, **5**, **6** and **7** in intraoocyte buffer showed a significant difference in emission intensity between duplex and GQ structure. Steady-state fluorescence analysis of ONs **4–7** in intraoocyte buffer was carried out by exciting the samples at 330 nm. H-Telo DNA **4**, **5** and **7**, which predominantly formed hybrid-type mixed parallel-antiparallel strand GQs in intraoocyte buffer, displayed noticeably higher fluorescence intensity (~4 to 6 fold at λ_{em} = 450 nm) as compared to respective duplexes **4•9**,

5•9 and **7•9** (Figure 12A and B). However, in case of **6**, no discernible change in emission spectra between duplex and quadruplex structures was observed. This showed that ON **6** could not discriminate between the duplex versus quadruplex structures.

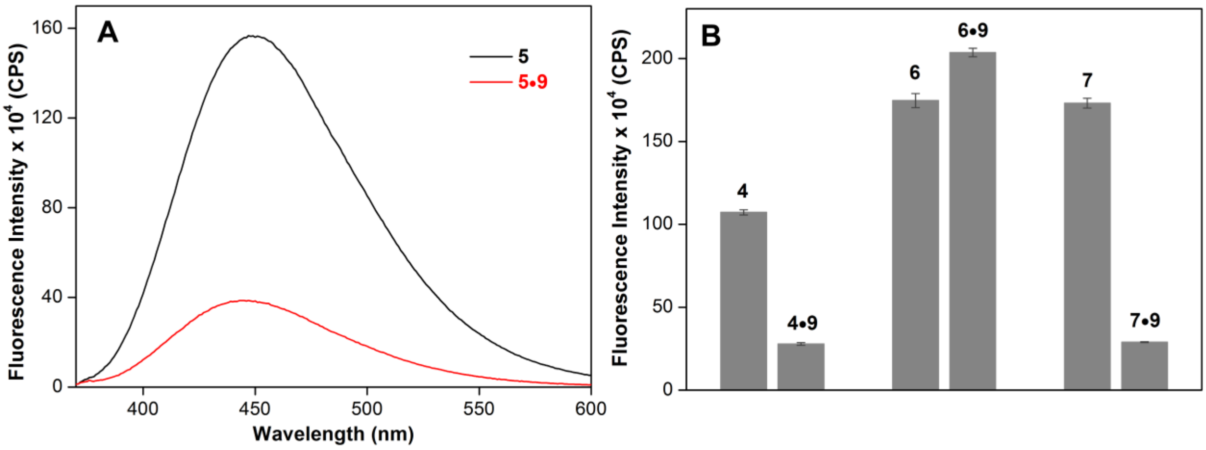


Figure 12. Steady-state fluorescence spectra of H-Telo DNA ON (**A**) **5** and corresponding duplex **5•9**, (**B**) **4**, **6** and **7** and their corresponding duplexs **4•9**, **6•9** and **7•9** in intraoocyte buffer containing 25 mM HEPES buffer (pH = 7.5) containing 10.5 mM NaCl, 110 mM KCl, 0.13 µM $CaCl_2$ and 1 mM $MgCl_2$. DNA ON samples (1 µM) were excited at 330 nm with an excitation and emission slit width of 5 nm and 9 nm, respectively.

3.2.6.3 Probing ligand binding in intraoocyte buffer

Fluorescent H-Telo DNA ONs were annealed in intraoocyte buffer (pH = 7.5). GQ of **5** was excited at 330 nm and changes in emission intensity upon increasing the concentration of PDS were monitored. A dose dependent quenching in fluorescence intensity of ON **5** (~3-fold) corresponding to an apparent K_d of 2.44 ± 0.2 µM was obtained (Figure 13A, B and Table 5). The ^{Se}dU probe also reported the binding of BRACO19 to hybrid GQs structures of **5** with significant quenching in fluorescence intensity upon addition of BRACO19 (Figure 13C). The apparent K_d values implied that BRACO19 has higher binding affinity for hybrid H-Telo DNA GQ ONs as compared to PDS (Figure 13D and Table 5).

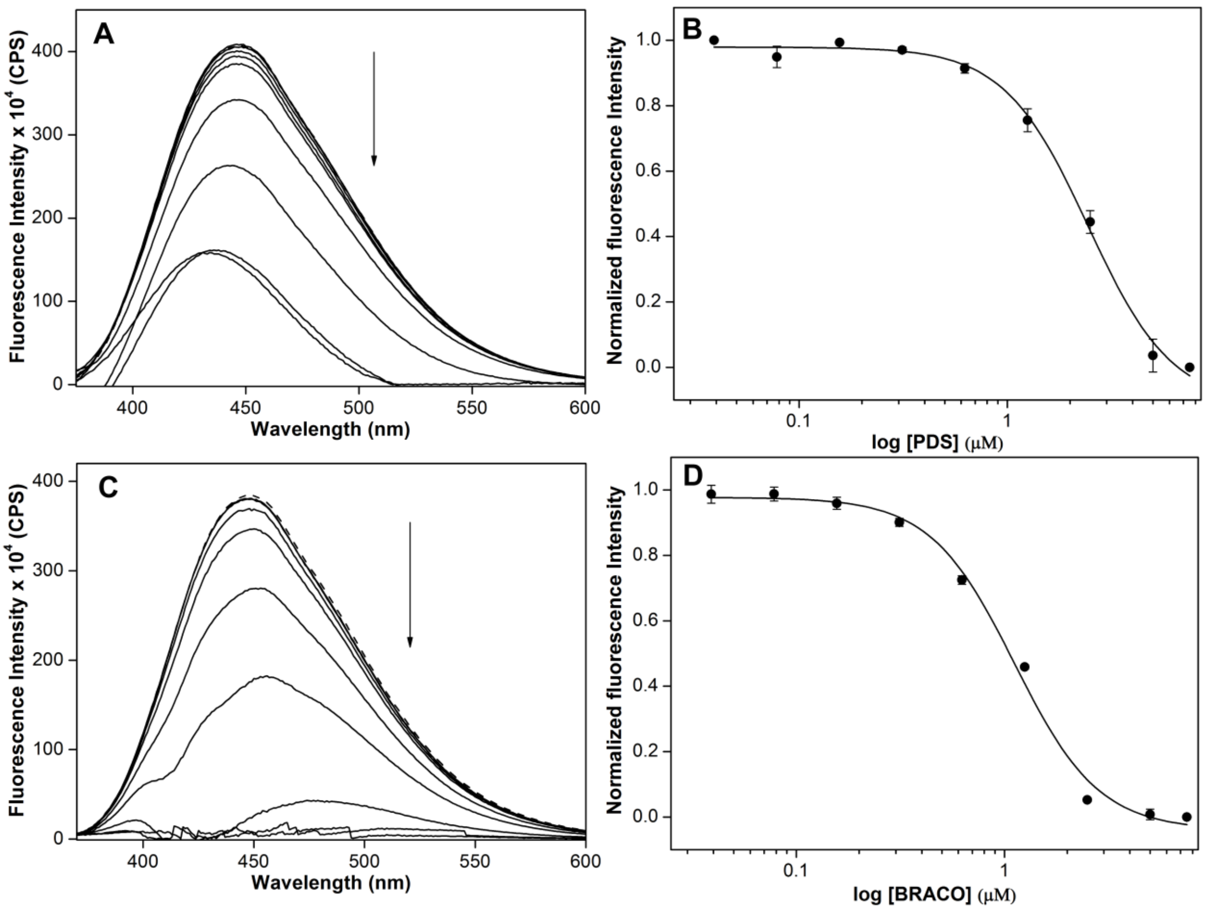


Figure 13. Emission spectra of hybride (**A**) and (**C**) GQs of H-Telo DNA **5** (1 µM) in the presence of intraoocyte buffer (pH = 7.5), respectively, with increasing concentrations of pyridostatin (PDS) or BRACO19. Dashed lines represent emission profile of GQs in the absence of PDS or BRACO19. Curve fit for the binding of PDS to H-Telo DNA **5** (**B**) and (**D**) curve fit for the binding of BRACO19 to H-Telo DNA **5**. Normalized fluorescence intensity at λ_{em} = 450 nm is plotted against log of [ligand].

Table 5. Binding constant (K_d) for PDS and BRACO19 binding to H-Telo DNA **5**

ON	PDS (µM)	BRACO19 (µM)
5	2.44 ± 0.2	1.12 ± 0.08

3.2.7 Fluorescence detection of parallel topology of the H-Telo DNA ONs

We next sought to explore the dual purpose of ^{Se}dU in correlating the structure and recognition property of H-Telo DNA GQ in solution and solid state. The crystal structure of H-Telo DNA ON repeat indicates the formation of parallel stranded GQ structure. In order to directly compare the solution data with that of X-ray structure, it is important to indentify if the ^{Se}dU can report the formation of the parallel GQ structure. Dehydrating agents like polyethylene glycol and ions like strontium are known to favour the formation of parallel GQ topology.[9,34] Hence, H-Telo DNA ONs **4–8** were annealed in the presence of $SrCl_2$ to favour the formation of the parallel structure. The CD spectrum of **4–8** exhibited a strong positive peak at ~260 nm and negative peak at ~240 nm, which is characteristic signature of parallel topology (Figure 14A). Thermal melting studies of unmodified and modified H-Telo GQs in Tris-HCl buffer containing Sr^{2+} gave nearly identical T_m values (Figure 14B and Table 6).[35] It revealed that the selenophene modification does not hamper the GQ structure formation (parallel-type) and stability.

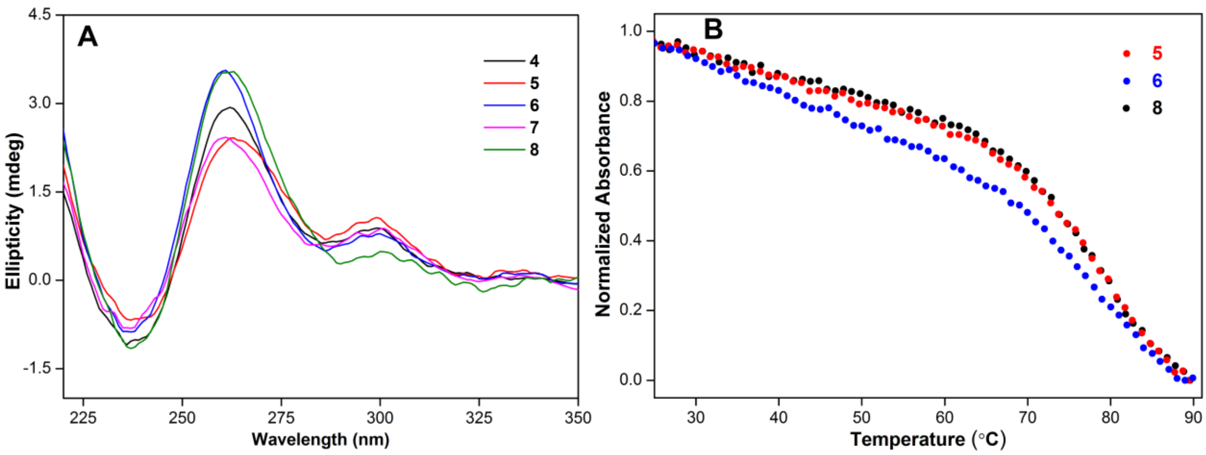


Figure 14. (**A**) CD spectra of fluorescently modified H-Telo DNA **4**, **5**, **6**, **7** (6 μM) and control unmodified H-Telo DNA **8** (6 μM) in Tris-HCl buffer (pH 7.4). Both unmodified and modified ONs show similar CD profiles in respective ionic conditions. (**B**) UV-thermal melting profile of fluorescent H-Telo DNA ONs **5** and **6** (1 μM) and control unmodified DNA ON **8** (1 μM) in Tris-HCl buffer (pH 7.4).

Table 6. T_m values of fluorescently modified (**4**–**7**) and control unmodified GQs (**8**) in the Tris HCl (pH = 7.4) conditions.

H-Telo GQ	T_m (°C)
4	79 ± 0.5
5	79 ± 0.6
6	78 ± 0.3
7	79 ± 0.5
8	80 ± 0.8

In the presence of Sr^{2+} ions, ^{Se}dU modified H-Telo GQs was able to distinguish parallel GQ structure from the duplex structure by the change in the fluorescence intensity. Parallel GQ structures of DNA ONs **4**, **5** and **7** showed enhancement in fluorescence intensity (nearly two fold at λ_{em} = 445 nm) as compared to respective duplexes **4•9**, **5•9** and **7•9**

(Figure 15A and B). This indicates that ^{Se}dU is free from stacking interaction in parallel GQ topology. However, in case of duplex, ^{Se}dU is involved in stacking interactions with adjacent guanine residue. Guanine is known to quench the fluorescence intensity by electron transfer and stacking interactions.[36] In order to assess the positional impact of ^{Se}dU probe, the fluorescence of ONs **6** was recorded under Sr^{2+} condition. Both parallel and its respective duplex **6•9** exhibited similar emission (Figure 15B). As before, ^{Se}dU placed at position 12 is not a good GQ sensor.

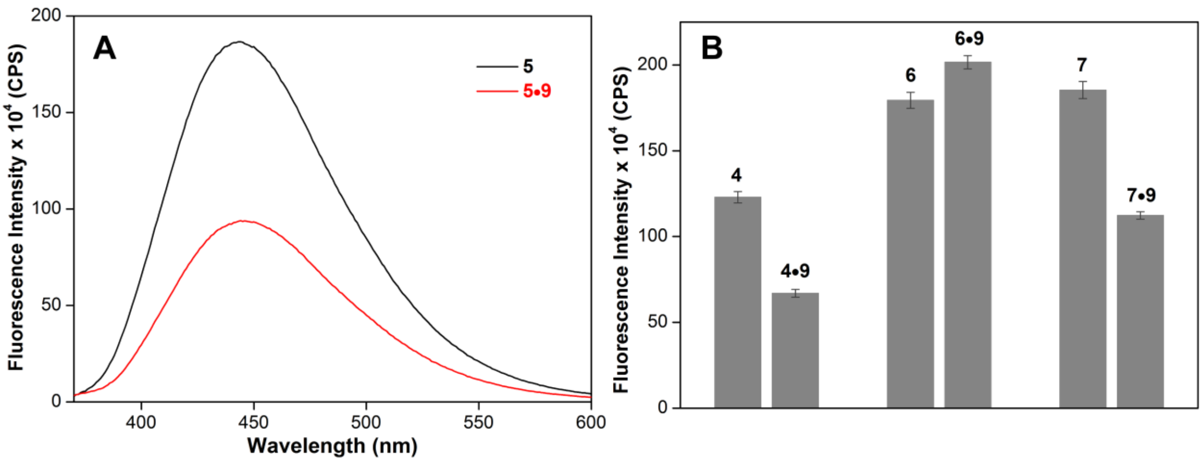


Figure 15. Steady-state fluorescence spectra of H-Telo DNA ON (**A**) **5** and corresponding duplex **5•9**, (**B**) **4**, **6** and **7** and their corresponding duplexes **4•9**, **6•9** and **7•9** in Tris-HCl buffer containing 50 mM Tris-HCl buffer (pH = 7.4) containing 150 mM $SrCl_2$. DNA ON samples (1 µM) were excited at 330 nm with an excitation and emission slit width of 5 nm and 9 nm, respectively.

The ON **5** annealed in the presence of Sr^{2+} ions was subjected to binding studies. Addition of increasing concentrations of PDS resulted in a dose-dependent quenching in fluorescence intensity, which gave an apparent K_d of 1.19 ± 0.1 µM (Figure 16A and B). BRACO19 exhibited a lower binding affinity for parallel GQs as compared to PDS (Figure 16B, D and Table 7).

Table 7. Binding constant (K_d) for PDS and BRACO19 binding to H-Telo DNA **5**, **6** and **10**

ON	PDS (µM)	BRACO19 (µM)
5	1.19 ± 0.1	2.39 ± 0.7

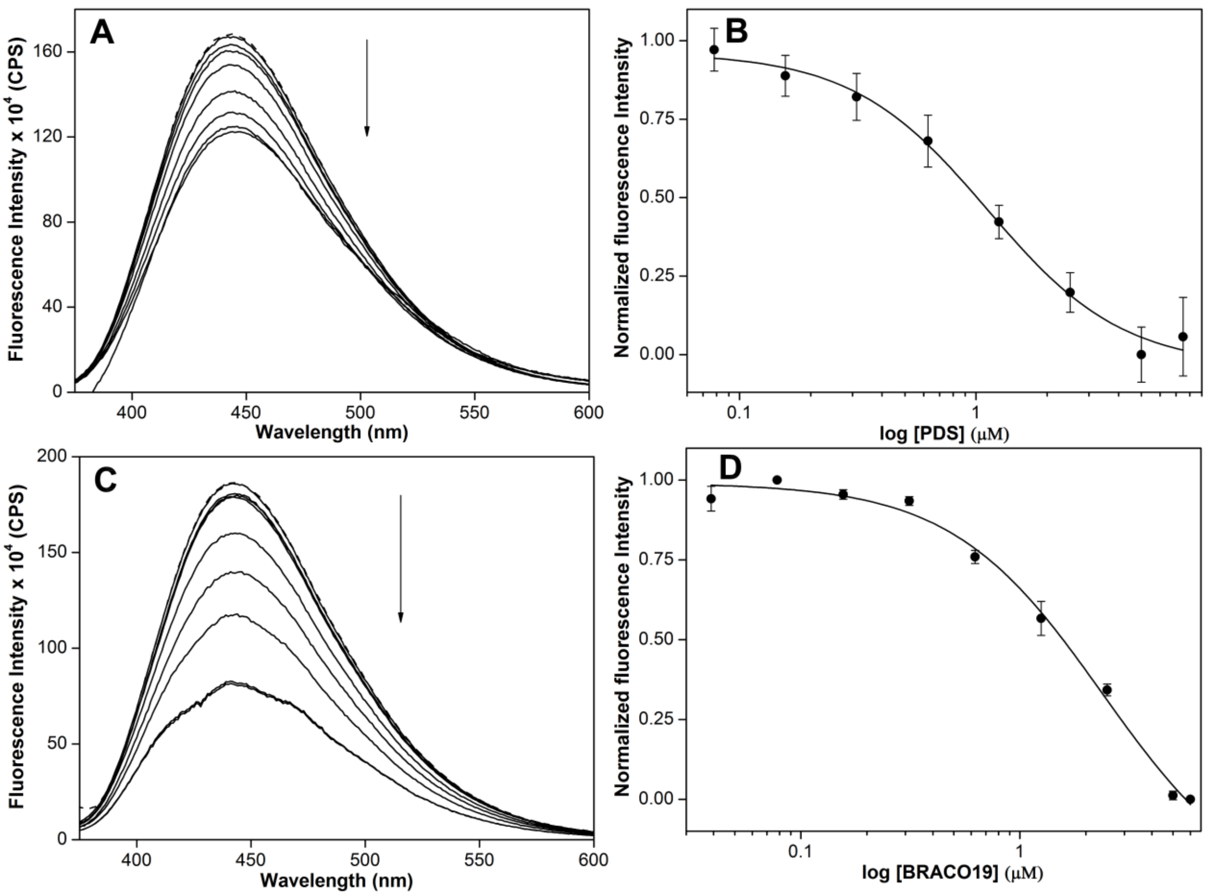


Figure 16. Emission spectra of hybride (**A**) and (**C**) GQs of H-Telo DNA **5** (1 µM) in the presence of Tris-HCl buffer (pH = 7.4), respectively, with increasing concentrations of pyridostatin (PDS) or BRACO19. Dashed lines represent emission profile of GQs in the absence of PDS or BRACO19. Curve fit for the binding of PDS to H-Telo DNA **5** (**B**) and (**D**) curve fit for the binding of BRACO19 to H-Telo DNA **5**. Normalized fluorescence intensity at λ_{em} = 445 nm is plotted against log of [ligand].

3.3 Crystal structures of ^{Se}dU-modified H-Telo DNA

Native (**8**) and selenophene-modifed H-Telo DNA ONs (**5** and **6**) were crystallized and their 3D structure was obtained at 1.40 Å, 1.55 Å and 2.3 Å resolutions, respectively. The strong anomalous signal of Se atom was used for phase determination of modified H-Telo structure. The structure of the native H-Telo DNA ON **8** and selenophene modified H-Telo DNA ON **5**, wherein the modified nucleoside was placed in position 11, was found to confer with the P6 space group. In case of selenophene modified H-Telo DNA ON **6**, wherein the modified nucleoside was placed in position 12, the structure was found to be in P21221 space group (Table 8). In the X-ray structure of ^{Se}dU modified ONs, the asymmetric unit was composed of one intramolecular parallel G-quadruplex unit, which was consistent with the native structure of ON **8** (Figure 17, 18, 19 and Table 8). In the GQ structure of **5** and **6**, the selenophene ring was found to be nearly coplanar with the uracil ring (-3.9° and 0.7°). In both the structures Se was facing the C4 carbonyl of uracil base (Figure 20). Native as well as the modified GQs adopted a parallel topology in which the tetrads were formed by guanines in *anti* conformation (Table 9).[15,37] The distance between the tetrads was 3.3 Å. Further, the GQ structure was stabilized by three K^+ ions. Each K^+ ion was located in-between the tetrads and was coordinated to eight C6 carbonyl oxygen atoms. It is important to mention here that the GQ structure of unmodified ON **8** is very similar to the GQ structure of the same sequence (1KF1) reported by Neidle.[15]

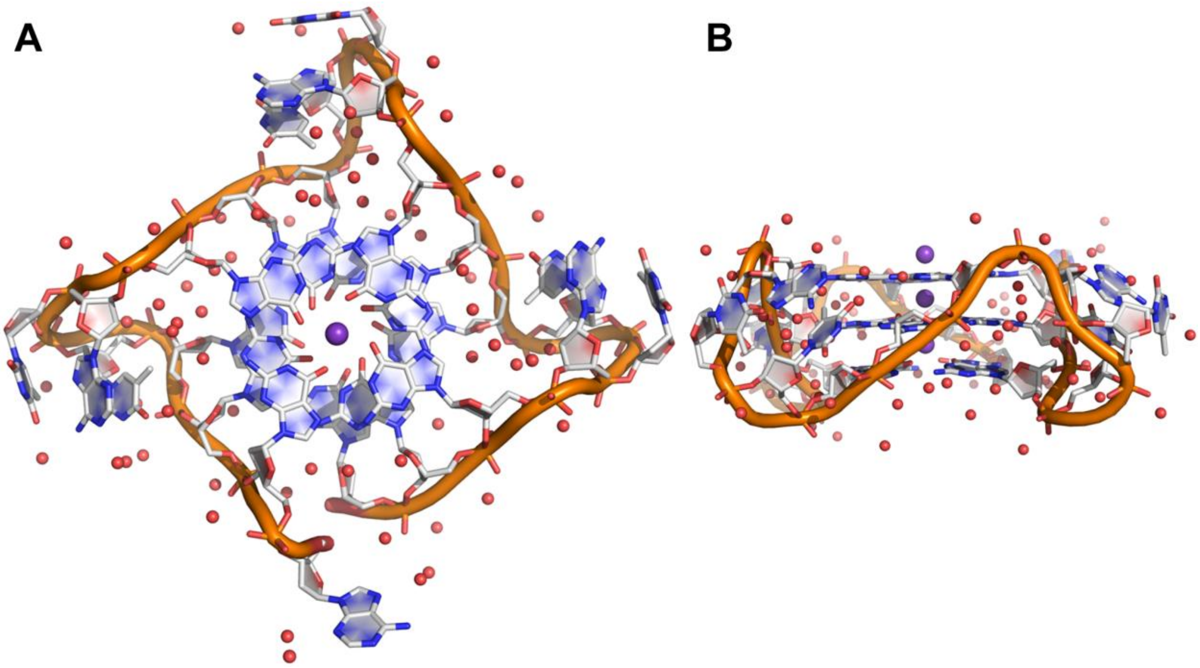


Figure 17. Crystal structure of native human telomeric DNA G-quadruplex (H-Telo) ON **8**, top view (**A**) and side view (**B**), indigo spheres implies potassium ions, red spheres indicates water molecules and resolution at 1.40 Å.

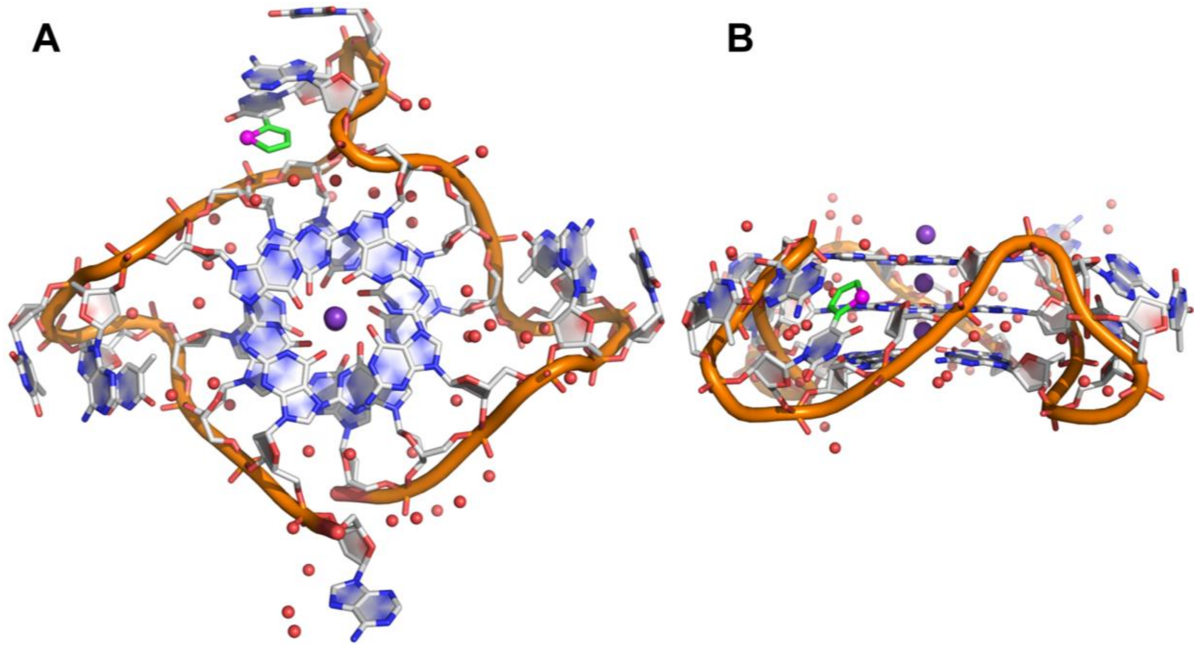


Figure 18. Crystal structure of selenophene modified H-Telo ON **5**, top view (**A**) and side view (**B**). Selenophene ring placed at 11-position of second loop denotes as green colour and selenium shown as magenta colour, indigo spheres implies potassium ions, red spheres indicates water molecules and resolution at 1.55 Å.

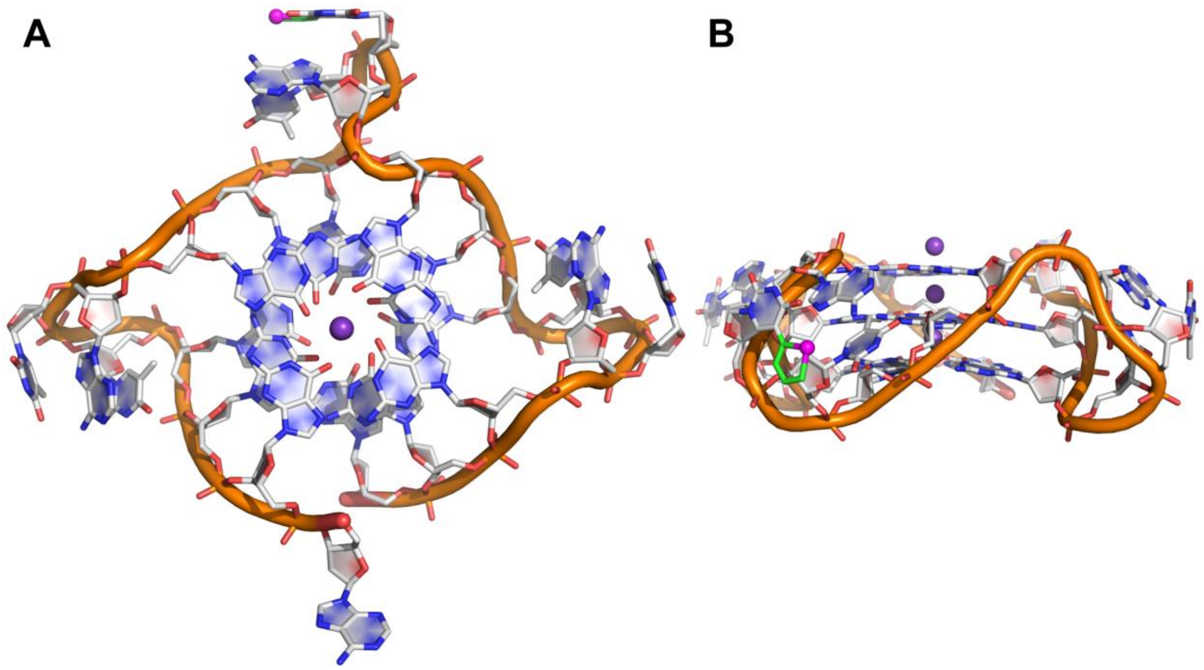


Figure 19. Crystal structure of selenophene modified H-Telo ON **6**, top view (**A**) and side view (**B**). Selenophene ring placed at 12-position of second loop denotes as green colour and selenium shown as magenta colour, indigo spheres implies potassium ions and resolution at 2.30 Å.

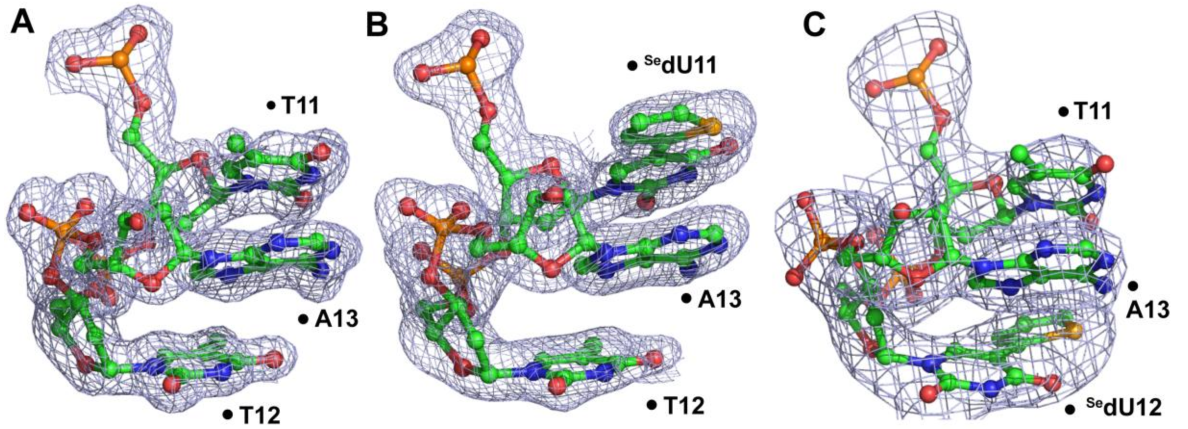


Figure 20. The $2F_O$-F_C electron density maps and contoured at 1.0σ level, is shown in gray. (**A**) The structure of the native H-Telo (**8**) at 1.40 Å resolution. (**B**) The structure of the ^{Se}dU-modified H-Telo (**5**) at 1.55 Å resolution, the selenium atom is shown as yellow sphere. (**C**) The structure of the ^{Se}dU-modified H-Telo (**6**) at 2.30 Å resolution, the selenium atom is shown as yellow sphere.

Table 8. Crystallographic data collection and refinement statistics

Structure	ON **8**	ON **5**	ON **6**
Space group	P6	P6	$P2_12 2_1$
Cell dimensions	56.839, 56.839,	56.483 56.483	35.204, 42.280,
a, b, c (Å)	42.411	42.415	49.924
α, β, γ (deg)	90, 90, 120	90, 90, 120	90, 90, 90
Wavelength (Å)	0.9795	0.9795	0.9792
Resolution (Å)	28.4-1.4 (1.42-1.4)	50-1.55 (1.58-1.55)	42.3-2.3 (2.38-2.3)
(Highest resolution shell)			
R_{merge} (%) overall	0.10 (0.88)	0.058 (0.90)	0.119 (0.744)
I/σ	16.4(3.5)	23.6 (3.0)	13.9 (3.7)
Completeness (%)	100 (100)	100 (100)	99.7 (99.6)
Redundancy	16.3 (14.7)	12.5 (12.7)	12.7 (12.3)
Refinement			
Resolution (Å)	28.4-1.4	50-1.55	42.3-2.3
No. of reflections	30199	21923	6296
R_{work}/R_{free} (%)	16.0/18.4	18.0/25.0	25.0/28.9
No. of atoms	550	522	472
No. of ions	3	3	3
No. of water molecules	78	50	0
RMS deviations in			
Bond lengths (Å)	0.012	0.019	0.008
Bond angles (deg)	1.187	2.141	2.233

Tetrad conformation: The GQ structure of unmodified and modified ONs are made of three tetrads, which stack one up the other. The tetrads are formed by G2-G8-G14-G20, G3-G9-G15-G21 and G4-G10-G16-G22 residues, which are *anti* in conformation (Table 9). The tetrads, apart from K^+ ions, are also stabilized by several water-mediated H-bonding interactions with N3 and N2 atoms of guanine residues (Figure 17B and 18B). The sugar conformation of the G9 residue of the middle tetrad of ON **5**, containing the modification at position 11 (second loop), adopted a C2′-*exo* conformation as opposed to C2′-*endo* in the native structure (Table 9). The G9 residue was found to display higher degree of *anti* conformation about the glycosidic bond (-153.6°) as compared to the native structure (111.8°). A torsion angle in the range of 180 ± 90° denotes *anti* conformation and in the range of 0°± 90° denotes *syn* conformation.[38] This is potentially due to the presence of the selenophene ring in the groove, which projects towards the G9 residue of the middle tetrad. However, the modification placed at the 12th position (ON **6**) did not affect the tetrad conformation as it was project away from the tetrad core. Other than these minor differences the tetrad conformation of native and modified ONs was similar.

Loop conformation: The conformation of three loops formed by TTA residues were found to be similar in native and modified GQ structures (Figure 21, Table 10 and 11). In the GQ structure of ON **5** in which the modification is in the T11 position (second loop), the loop residue A13 was stacked in-between ^{Se}dU11 and T12 residues (stacking distance 3.94 Å, Figure 21B). In case of native GQ structure the stacking distance is 3.16 Å (Figure 21A). The slight deviation in stacking interaction is due to the differences in the torsion angle (O4′-C1′-N1-C2) about the glycosidic bond. In ON **5**, ^{Se}dU11 is less *anti* (–92.0°) as compared to the T11 residue (–111.6°) in the native structure (Table 10). Due to this, ^{Se}dU11 is slightly pushed away from the loop, thereby experiences less stacking interaction. In the GQ structure of ON **6**, ^{Se}dU12 does not show this deviation and exhibits similar stacking interaction as that of the native form (Figure 21C).

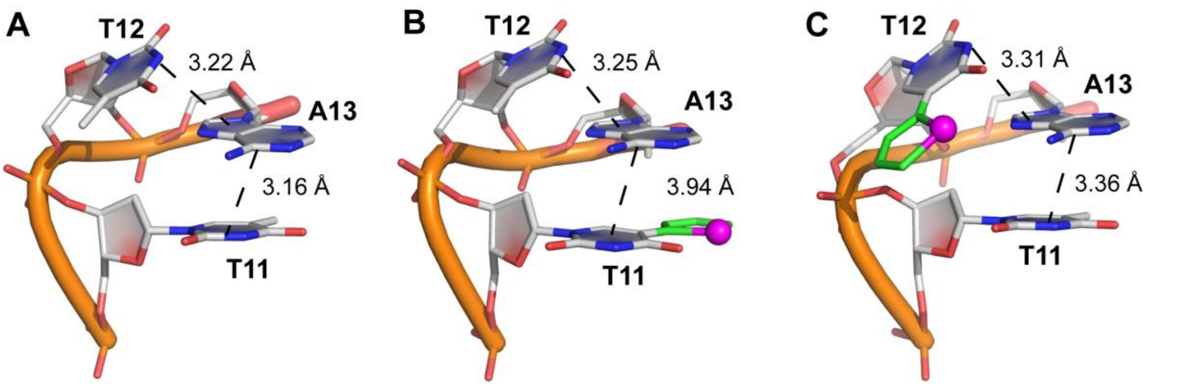


Figure 21. The trinucleotide(TTA) second loop geometry of native H-TeloDNA (**A**), selenophene modified H-Telo (**5**) (**B**) and **6** (**C**). This loop orientation similar to type-1 loop (1KF1) arrangement,[15] where adenine (A13) involves π–π stacking interactions with adjacent thymine moieties (T11 and T12). Here selenophene ring and selenium atom are denote as green and magenta colour, respectively.

Packing structure: The packing structure of unmodified and modified H-Telo DNA ONs show two parallel GQs interacting from 5′–5′ end via stacking interaction. The dimeric structure is stabilized by a K^+ ion, which coordinates with the C6 carbonyl oxygen atoms of tetrads of adjacent GQs in an sandwiched fashion (Figure 22). The distance between the interacting tetrads is 3.3 ± 0.1 Å. In the native structure, A1 residue of one of the GQs forms reverse Watson-Crick hydrogen bond interaction with T12 and also stacks with A13 residue of adjacent GQ (Figure 22A). However, when the modification is placed in the T11 position, the dimer structure is stabilized by canonical Watson-Crick hydrogen bonding between A1 and T12 residues of neighboring GQ (Figure 23A). Modification present in the T12 position also forms similar interactions in the packed structure (Figure 24A). Two adjacent dimers interact with each other through strong π–π stacking interactions via two pairs of stacked T6′–T18 bases (Figure 22B, 23B and 24B), which are separated by 3.3 ± 0.1 Å. These bases are integral part in the parallel TTA loops (1 and 3).

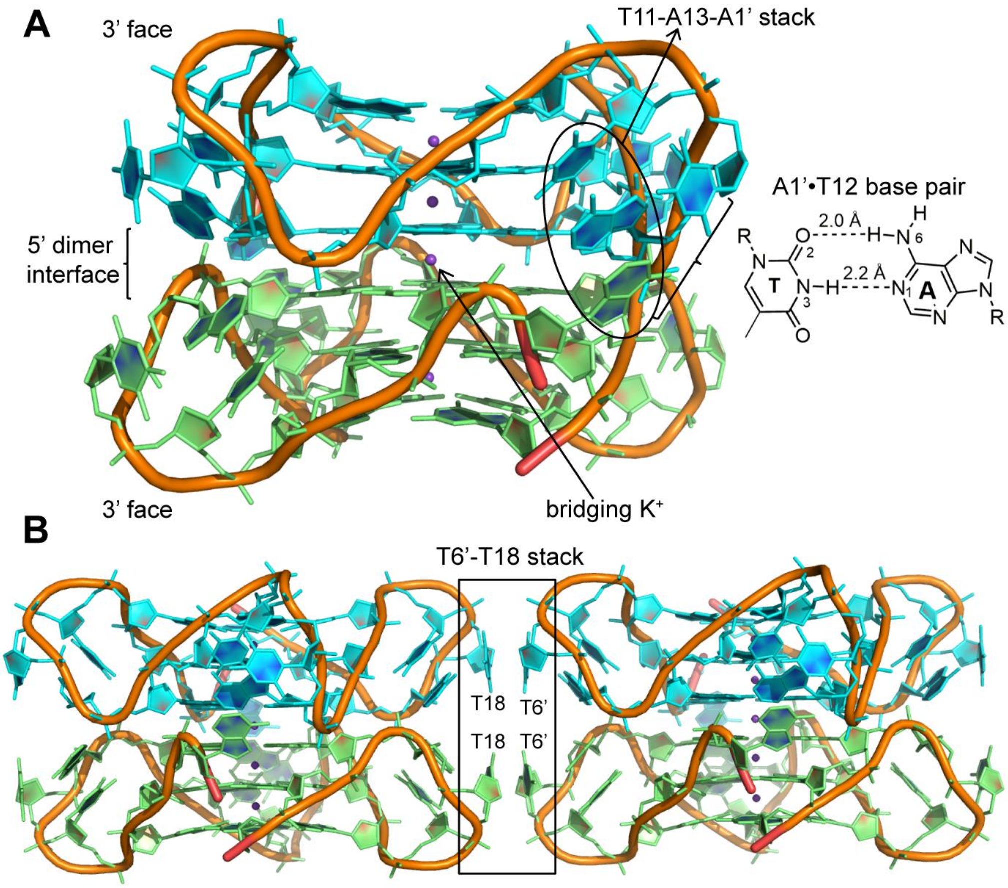


Figure 22. Intermolecular interactions in native H-Tel22 DNA ON **8** crystal. (**A**) Dimer formation is guided by the head-to-head stacking of 5′ G-tetrads from two intramolecular parallel GQs (cyan and lime), a bridging K^+ ion, an A1′-T12 base pair, and π–π stacking of A1′-A13-T11 involving loops from each monomer. Interactions between dimers are mediated by (**B**) π–π stacking of two sets of T bases.

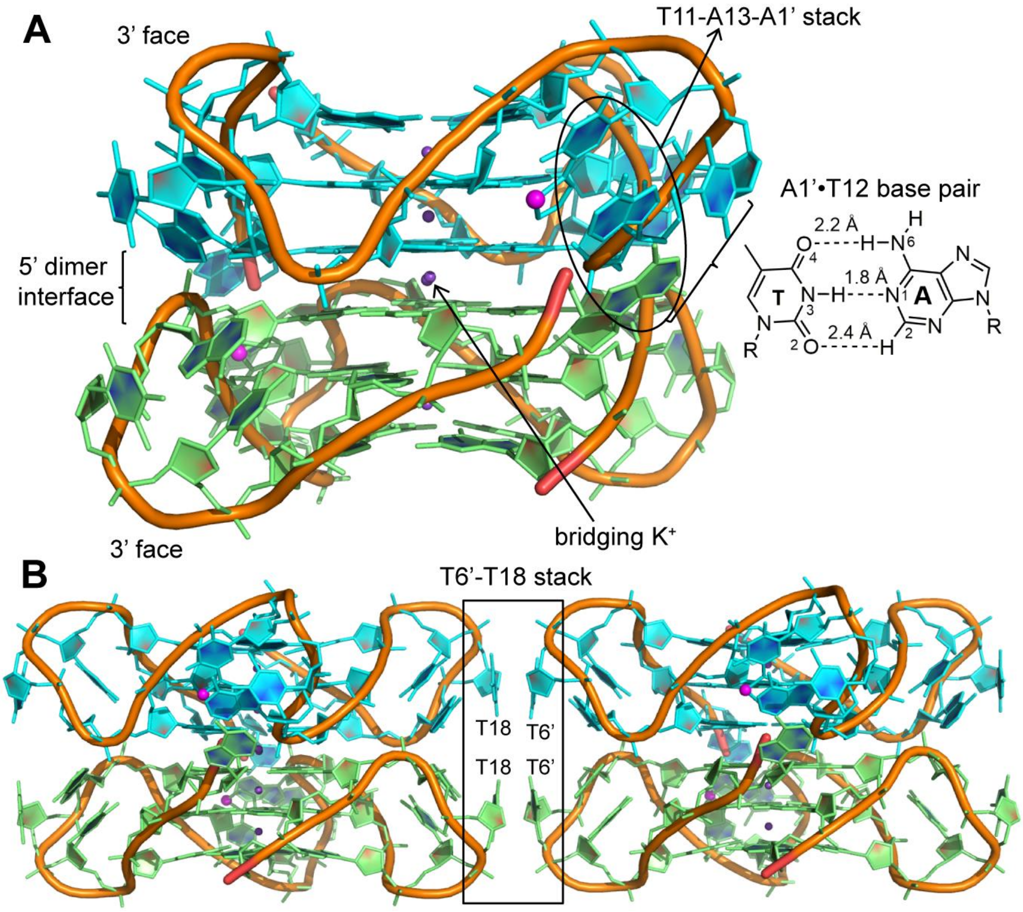


Figure 23. Intermolecular interactions in selenophene modified H-Tel22 DNA ON **5** crystal. (**A**) Dimer formation is guided by the head-to-head stacking of 5′ G-tetrads from two intramolecular parallel GQs (cyan and lime), a bridging K^+ ion, an A1′**-**T12 base pair, and π–π stacking of A1′-A13-T11 involving loops from each monomer. Interactions between dimers are mediated by (**B**) π–π stacking of two sets of T bases.

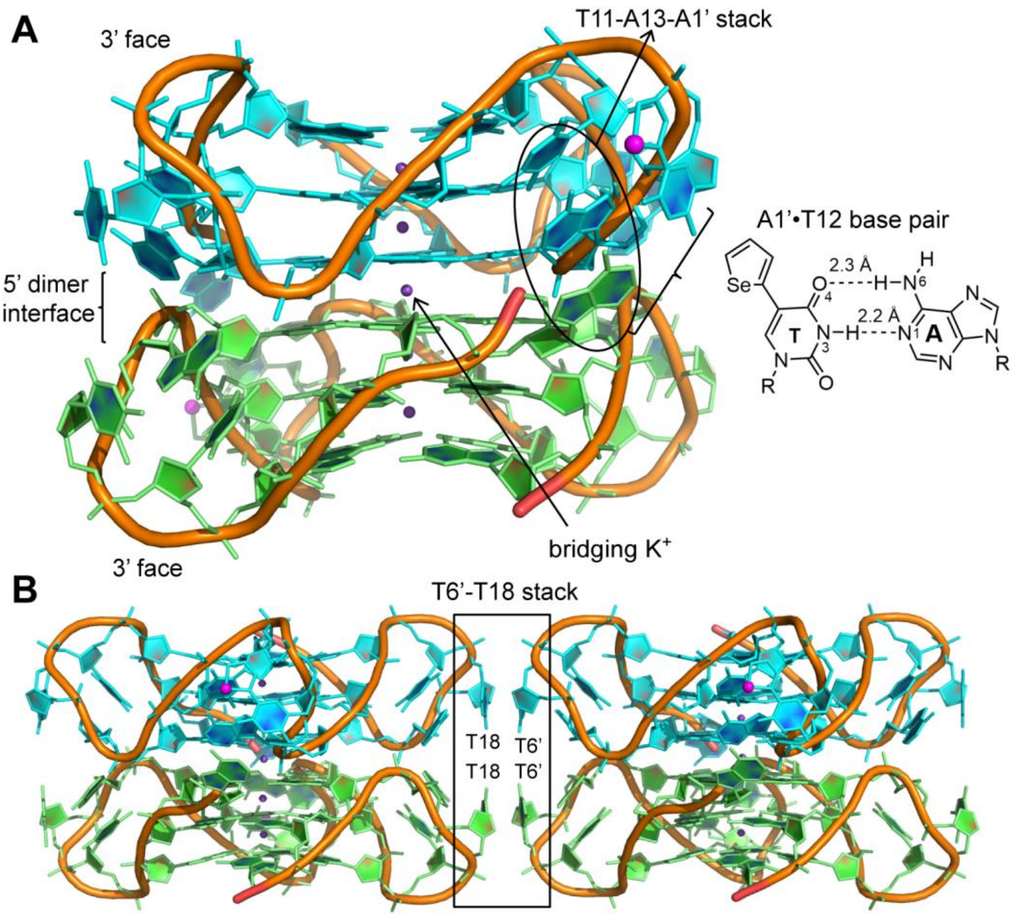


Figure 24. Intermolecular interactions in selenophene modified H-Telo DNA ON **6** crystal. (**A**) Dimer formation is guided by the head-to-head stacking of 5′ G-tetrads from two intramolecular parallel GQs (cyan and lime), a bridging K^+ ion, an A1′•T12 base pair, and π–π stacking of A1′-A13-T11 involving loops from each monomer. Interactions between dimers are mediated by (**B**) π–π stacking of two sets of T bases.

3.3.1 Superimposed structure of native and ^{Se}dU modified GQs

In order to further understand the effect modification on the native fold of the GQ structure, the GQ structure of unmodified ON **8** and modified ON **5**/**6** was superimposed. First, the GQ structure of ON **8** crystallized in the present study was superimposed on to the structure reported by Neidle group for the same sequence.[15] A RMSD (root mean square deviation) of 0.36 Å was observed for 465 atoms out of 465 atoms, which reveals that the structures are nearly identical (Figure 25). Superimposition of GQs of ON **8** and **5/6**, revealed a RMSD of 0.51/0.60 Å (Figure 26 and 27). Taken together, the crystal structure data indicates that the modification has only a minor impact on the native fold of the GQ structure.

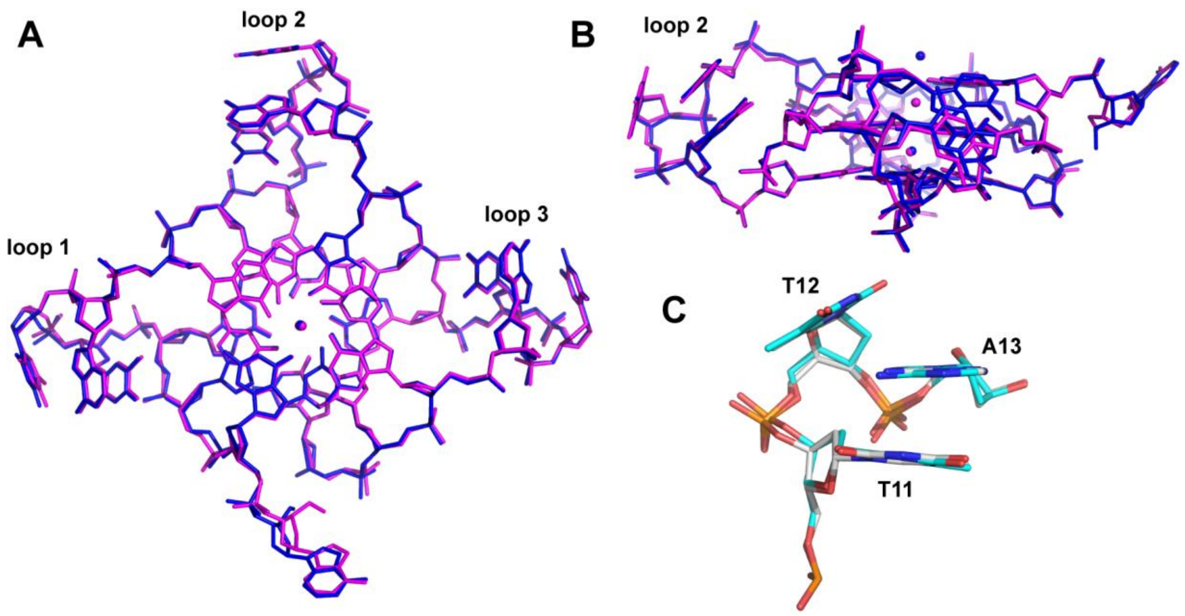


Figure 25. (**A**) Structural alignment of the both native H-Telo ON **8** (blue) and 1KF1 (magenta). RMSD (root mean square deviation) = 0.36 Å (465 to 465 atoms). (**B**) Side view of loop 2 region. (**C**) Superimposed trinucleotide loop 2 orientation. All the residues align similar crystal form.

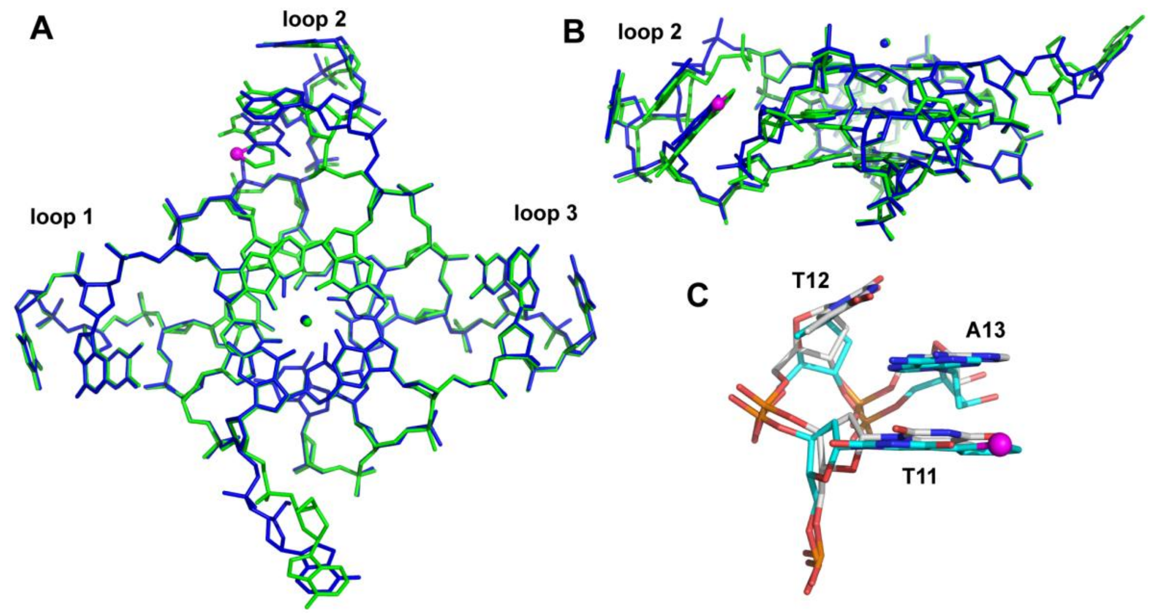


Figure 26. (**A**) Structural alignment of the native H-Telo ON **8** (blue) and the modified ON **5** (green). RMSD (root mean square deviation) = 0.51 Å (360 to 360 atoms). (**B**) Side view of loop 2 region where little structural deviation was observed. (**C**) Superimposed trinucleotide loop 2 orientation and selenium denotes as (magenta).

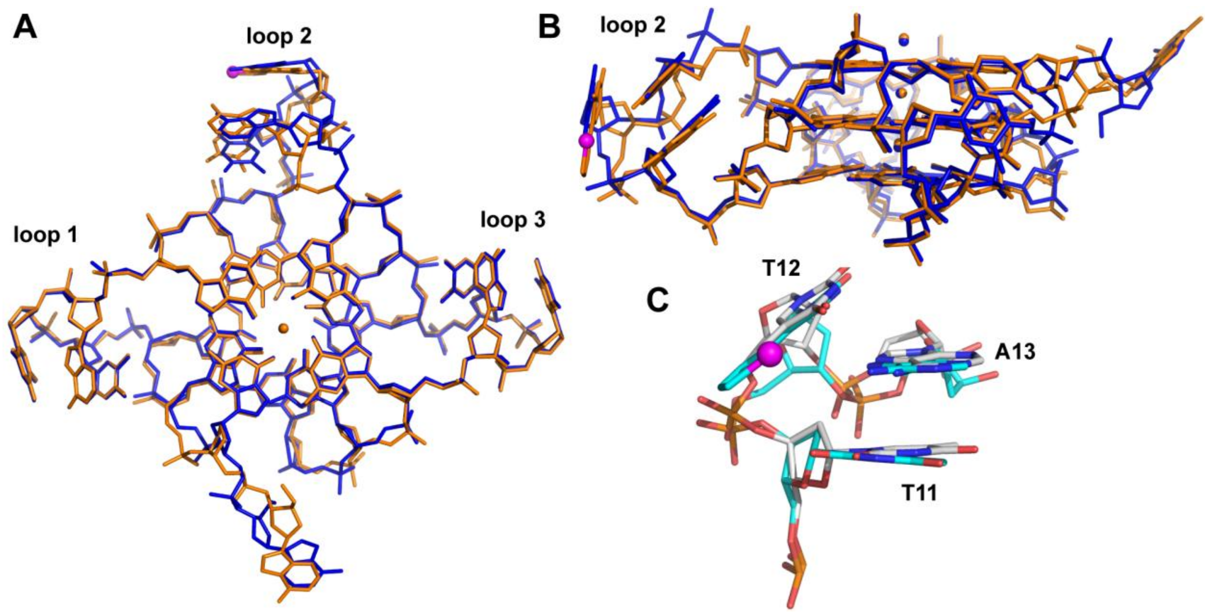


Figure 27. (**A**) Structural alignment of the native H-Telo ON **8** (blue) and the modified ON **6** (orange). RMSD (root mean square deviation) = 0.60 Å (360 to 360 atoms). (**B**) Side view of loop 2 region, where little structural deviation was observed. (**C**) Superimposed comparison of trinucleotide loop 2 orientation and selenium denotes as (magenta).

3.3.2 Structural basis for the fluorescence sensing of the GQ structure by ^{Se}dU

Careful analysis of the structure, particularly the interaction between ^{Se}dU and neighboring bases provided information on how the nucleoside probe senses the formation of GQ structure. Since in the solid-state, the H-Telo DNA adopts a parallel GQ topology, the fluorescence data obtained for the parallel GQ topology (obtained in presence of strontium ion) was compared to the structure. ^{Se}dU placed in the second loop (position T11) of the ON **5** reports the formation of GQ structure with significant enhancement in fluorescence intensity as compared to corresponding duplex. In the crystal structure of **5**, ^{Se}dU was involved in partial stacking interaction with adjacent A13 residue. But when it is in duplex form ^{Se}dU11 could strongly stack with neighboring T12 and G10 residues, leading to quenching in fluorescence intensity. Likewise, in ON **6** ^{Se}dU located at second loop (position T12), both GQ and duplex structures were shown to have similar fluorescence intensity. The ON **6** in crystal form ^{Se}dU12 participated in partial stacking interactions with A13 base. In duplex structure of ON **6**, ^{Se}dU12 was flank between the T11 and A13 bases. These results clearly substantiate that fluorescent ^{Se}dU probe in close proximity to guanine is drastically quenched,[36] unlike in the presence of adenine. The reduction or enhancement in fluorescence intensity depends on extent of stacking interactions of the emissive probe with flanking bases.

3.4 Conclusions

The dual-purpose ^{Se}dU probe was site-specifically incorporated into three loops of H-Telo DNA GQs. The modification minimally perturbs the native structure of H-Telo DNA. This probe was able to distinguish GQ structure from duplex by changing its fluorescence intensity in different salt containing buffers. Further, the probe has been utilized to investigate GQ formation and its ligand binding ability in intraoocyte buffer, which mimics the physiological salt conditions. To compare the fluorescence response of ^{Se}dU modified H-Telo in solution state to 3D structure, the ^{Se}dU modified H-Telo is induced into the parallel topology upon addition of Sr^{2+} ion. Collectively this ^{Se}dU probe faithfully reports the ligand-binding to different GQ topologies, such as antiparallel, mixed-hybrid and parallel structures. In addition the selenium-derivatization is highly useful in the determination of non-canonical nucleic acids crystal structures via MAD or SAD phasing. We have reported the first X-ray crystal structures of ^{Se}dU-modified H-Telo DNA at highest resolution reported to date. The native and ^{Se}dU-derivatized H-Telo structures are solved at 1.40 Å and 1.54 Å resolutions, respectively. The native and modified H-Telo structures are nearly identical crystal form (parallel topology). Thereby the ^{Se}dU probe was permitted the direct correlation of the information obtained from two independent techniques, namely fluorescence and X-ray crystallography.

3.5.Experimental section

3.5.1 Materials: 5-iodo deoxyuridine, selenophene, butyllithium, tributyltin chloride, *bis*(triphenylphosphine)-palladium(II) chloride, 4,4′-dimethoxytrityl chloride, Pyridostatin (PDS), BRACO19 were obtained from Sigma-Aldrich. 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite was purchased from Alfa Aesar. *N*-benzoyl-protected dA, dT, *N*,*N*-dimethylformamidine-protected dG and *N*-acetyl-protected dC phosphoramidite substrates for solid-phase DNA synthesis were purchased from ChemGenes Corporation. Solid supports required for DNA synthesis were purchased from ChemGenes Corporation. All other reagents for solid-phase oligonucleotide (ON) synthesis were obtained from either ChemGenes Corporation or Sigma Aldrich. Human telomeric DNA ONs **4**, **5**, **6**, **7** and **10** were synthesized by following earlier reported procedure using phosphoramidite **3**.[39] Custom synthesized DNA ONs **8** and **9** were purchased from Integrated DNA Technologies, Inc. and purified by polyacrylamide gel electrophoresis (PAGE) under denaturing condition, and desalted using Sep-Pak Classic C18 cartridge (Waters Corporation). Chemicals (BioUltra grade) for preparing buffer solutions were purchased from Sigma-Aldrich. Autoclaved water was used for preparation of all buffer solutions and in biophysical analysis.

3.5.2 Instrumentation: NMR spectra were recorded on 400 MHz Jeol ECS-400 spectrometer. Mass measurements were done on Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer and Water Synapt G2 High Definition mass spectrometers. Modified RNA ONs were synthesized on Applied Biosystems RNA/DNA synthesizer (ABI-394). Absorption spectra were recorded on Shimadzu UV-2600 spectrophotometer. HPLC analysis was performed using Agilent Technologies 1260 Infinity. UV-thermal melting studies of oligonucleotides were performed on Cary 300Bio UV-Vis spectrophotometer and CD analysis was carried out on JASCO J-815 CD spectrometer. Steady-state and time-resolved fluorescence experiments were carried out in a micro fluorescence cuvette (Hellma, path length 1.0 cm) on a TCSPC instrument (Horiba Jobin Yvon, Fluorolog-3), respectively. X-ray diffraction data were collected at UK Diamond Light Source Synchrotron and European Synchrotron Radition Facility (ESRF) at wavelength 0.9793 Å, beamline IDs I02, I04 and BM14, respectively.

3.5.3 Synthesis of 5-selenophene-modified 2′-deoxyuridine phosphoramidite substrate 3

3.5.3.1 5-Selenophene-modified 5′-*O*-DMT-protected deoxyuridine 2: A mixture of 5′-*O*-DMT-protected 5-iodo deoxyuridine **1** [39] (0.71 g, 1.04 mmol, 1.0 equiv), 2-(tri-*n*-butylstannyl) selenophene [40] (1.0 g, 2.37 mmol, 2.2 equiv), and *bis*(triphenylphosphine)-palladium(II) chloride (0.05 g, 0.06 mmol, 0.06 equiv) was dissolved in anhydrous dioxane (25 mL). The reaction mixture was heated at 90 °C for 2.5 h and filtered through celite pad. The celite pad was washed with dichloromethane (2 × 20 ml), filtrate was evaporated and the resulting solid residue was purified by silica gel column chromatography to afford the compound **3** as a colourless solid foam, 91% yield. TLC (CH_2Cl_2:MeOH = 95:5, few drops of triethylamine); R_f = 0.62; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 9.39 (s, 1H), 7.99 (s, 1H), 7.91–7.90 (m, 1H), 7.41–7.39 (m, 2H), 7.29–7.27 (m, 4H), 7.25–7.17 (m, 3H), 6.90–6.88 (m, 2H), 6.78–6.76 (m, 4H), 6.39 (dd, J_1 = 7.6 Hz, J_2 = 6.0 Hz, 1H), 4.53–4.51 (m, 1H), 4.17–4.14 (m, 1H), 3.74 (s, 6H), 3.46–3.42 (m, 1H), 3.38 (dd, J_1 = 10.4 Hz, J_2 = 4.0 Hz, 1H), 2.57–2.51 (m, 1H), 2.33–2.26 (m, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 161.5, 158.8, 149.5, 144.4, 136.8, 135.6, 135.5, 133.0, 132.1, 130.1, 130.1, 128.9, 128.2, 128.1, 127.2, 124.7, 113.4, 112.3, 87.0, 86.4, 85.7, 72.5, 63.7, 55.4, 40.9; HRMS: m/z Calcd. for $C_{34}H_{32}N_2O_7Se$ $[M-H]^+$ = 659.1297, found = 659.1290.

3.5.3.2 5-Selenophene-modified 2′-deoxyuridine phosphoramidite substrate 3: To a solution of compound **2** (0.33 g, 0.55 mmol, 1.0 equiv) in anhydrous dichloromethane (2.5 ml) was added *N*,*N*-diisopropylethylamine (0.44 mL, 2.53 mmol, 2.5 equiv) and stirred for 10 min under ice cold conditions. To this solution was slowly added 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (0.17 mL, 0.75 mmol, 1.5 equiv) and the reaction mixture was allowed RT and stirred for 1.5 h. The solvent was evaporated to dryness and the residue was redissolved in ethyl acetate (20 ml). It was washed with 5% sodium bicarbonate solution (15 ml) followed by brine solution (15 ml). The organic extract was dried over sodium sulphate, evaporated and the crude solid residue was purified by column chromatography to afford the product **3** as a colourless solid, 55% yield. TLC (CH_2Cl_2:MeOH = 95:5 few drops of triethyl amine); R_f = 0.58; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.02 (br, 1H), 7.88-7.87 (m, 1H), 7.40–7.38 (m, 2H), 7.28–7.17 (m, 8H), 6.82–6.75 (m, 6H), 6.82–6.75 (m, 6H), 6.39–6.35 (m, 1H), 4.61–4.58 (m, 1H), 4.24 (br, 1H), 3.73 (s, 6H), 3.66–3.54 (m, 4H), 3.50–3.47 (m, 1H), 3.33–3.30 (m, 1H), 2.57–2.54 (m, 1H), 2.40 (t, *J* = 6.4 Hz, 2H), 2.33–2.26 (m, 1H), 1.16–1.15 (m, 12H), ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 161.4, 158.8, 149.2, 144.4, 136.6, 135.5, 135.5, 132.9, 132.0, 130.2, 130.2, 128.9, 128.3, 128.1, 127.2, 124.7, 117.5, 113.3, 112.3, 86.8, 86.0, 85.5, 73.7, 73.5, 63.2, 58.4, 58.2, 55.4, 43.5, 43.3, 40.4, 40.4, 24.8, 24.7, 24.6, 20.3, 20.3; ^{31}P NMR (162 MHz, $CDCl_3$): δ (ppm) 149.7; HRMS: m/z Calcd. for $C_{43}H_{49}N_4O_8PSe$ $[M+H]^+$ = 899.2092, found = 899.2299.

3.5.4 Solid-phase synthesis of selenophene modified H-Telo DNA ONs: Selenophene-modified H-Telo DNA ONs **4–7**, **10** were synthesized on a 1.0 µmol scale (1000 CPG solid supports) by a standard DNA ON synthesis protocol using phosphoramidite **3**.[40] Selenophene phosphoramidite (**3**) substrate was site specifically incorporated into therapeutically related human telomeric DNA G-quadruplexs ONs (coupling efficiencies were 55–70%). The solid support was treated with 30% aqueous ammonium hydroxide for 15 h at ~55 °C. The aqueous ammonium hydroxide solution was evaporated to dryness on speedVac and deprotected ON products were purified by 20% polyacrylamide gel electrophoresis (PAGE) under the denaturing conditions. The band corresponding to the full-length product was identified by

UV shadowing, desire bands cut and transferred to a Poly-Prep column (Bio-Rad). The gel pieces were crushed with sterile glass rod and ONs extracted in ammonium acetate buffer (0.5 M, 3 mL) for 12 h and desalted using a Sep-Pak classic C18 cartridge (Waters). The purity of ONs were investigated by RP-HPLC and characterized by MALDI TOF mass analysis (see Table 1 for ε_{260} and mass analysis data of modified DNA).

3.5.5 MALDI-TOF mass analysis of modified DNA ONs: 2 µL of the modified ON (~200 µM) was combined with 1 µL of ammonium citrate buffer (100 mM, pH 9), 2 µL of a DNA internal standard (200 µM) and 4 µL of saturated 3-hydroxypicolinic acid solution. The sample was desalted using an ion-exchange resin (Dowex 50W-X8, 100-200 mesh, ammonium form), spotted on the MALDI plate and was air dried. The resulting spectrum was calibrated relative to an internal DNA ON standard.

3.5.6 CD measurements:

Sample preparation in aqueous buffer: Modified and unmodified H-Telo DNA ONs **4**–**8**, **10** (12 µM) were annealed at 90 °C for 5 min in 10 mM Tris-HCl buffer (pH 7.5) containing either 100 mM NaCl or 100 mM KCl. Samples were then allowed cool RT and kept at ~4 °C for 1 h. The final concentrations of samples in CD analysis of modified H-Telo DNA ONs **4**–**7**, **10** (6 µM) and control unmodified H-Telo DNA ON **8** (6 µM), respectively. The CD spectrum were collected from 350 nm to 200 nm using a quartz cuvette (Sterna Scientific, path length 2 mm) on a Jasco J-815 CD spectrometer using 1 nm bandwidth at 20 °C. CD measurements were performed in duplicate and the spectrum of buffer was subtracted from all sample spectra.

Sample preparation in intraoocyte buffer: Modified and unmodified H-Telo DNA ONs **4**–**8**, **10** (12 µM) were annealed at 90 °C for 5 min in intraoocyte buffer (25 mM HEPES buffer (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 0.13 µM $CaCl_2$ and 1 mM $MgCl_2$). The final concentrations of samples in CD analysis of modified H-Telo DNA ONs **4**–**7**, **10** (6 µM) and control unmodified H-Telo DNA ON **8** (6 µM), respectively. The CD spectrum were collected from 350 nm to 200 nm using a quartz cuvette (Sterna Scientific, path length 2 mm) on a Jasco J-815 CD spectrometer using 1 nm bandwidth at 20 °C. CD measurements were performed in duplicate and the spectrum of buffer was subtracted from all sample spectra.

Sample preparation in strontium (Sr^{2+}) containing Tris-HCl buffer: Modified and unmodified H-Telo DNA ONs **4**–**8** (12 µM) were annealed at 90 °C for 5 min in 50 mM Tris-HCl buffer (pH = 7.4) containing 150 mM $SrCl_2$. The final concentrations of samples in CD analysis of modified H-Telo DNA ONs **4**–**7** (6 µM) and control unmodified H-Telo DNA ON **8** (6 µM), respectively. The CD spectrum were collected from 350 nm to 200 nm using a quartz cuvette (Sterna Scientific, path length 2 mm) on a Jasco J-815 CD spectrometer using 1 nm bandwidth at 20 °C. CD measurements were performed in duplicate and the spectrum of buffer was subtracted from all sample spectra.

3.5.7 Thermal melting analysis:

Sample preparation in aqueous buffer: Modified H-Telo DNA ONs **4**–**7** (10 µM), and control unmodified H-Telo DNA ON **8** (10 µM) were annealed by heating at 90 °C for 5 min in 10 mM Tris-HCl buffer (pH 7.5) containing either 100 mM NaCl or 100 mM KCl. The samples were slowly cooled to RT and were kept at ~4 °C for 1 h. Thermal melting analysis of the ONs (1 µM) was performed in triplicate by using Cary 300Bio UV-Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 295 nm. Forward and reverse cycles were used to determine the T_m values.

Sample preparation in intraoocyte buffer: Modified H-Telo DNA ONs **4**–**7** (10 µM), and control unmodified H-Telo DNA ON **8** (10 µM) were annealed by heating at 90 °C for 5 min in intraoocyte buffer(25 mM HEPES buffer (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 0.13 µM $CaCl_2$ and 1 mM $MgCl_2$ and cooled to RT overnight). Thermal melting analysis of the ONs (1 µM) was performed in triplicate by using Cary 300Bio UV-Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 295 nm. Forward and reverse cycles were used to determine the T_m values.

Sample preparation in strontium (Sr^{2+}) containing Tris-HCl buffer: Modified H-Telo DNA ONs **4**–**7** (10 µM), and control unmodified H-Telo DNA ON **8** (10 µM) were annealed by heating at 90 °C for 5 min in 50 mM Tris-HCl buffer (pH = 7.4) containing 150 mM $SrCl_2$ and cooled to RT overnight. Thermal melting analysis of the ONs (1 µM) was performed in triplicate by using Cary 300Bio UV-Vis spectrophotometer. The temperature was increased from 20 °C to 90 °C at 1 °C/min and the absorbance was measured every 1 °C interval at 295 nm. Forward and reverse cycles were used to determine the T_m values.

3.5.8 Fluorescence analysis: Steady-state fluorescence measurements of DNA ONs samples in aqueous buffer or in intraoocyte buffer or in strontium (Sr^{2+}) containing Tris-HCl buffer were carried out by exciting the sample at 330 nm with an excitation and emission slit width of 5 nm and 9 nm and in case of **10** ON excitation and emission slit width of 5 nm and 6 nm, respectively.

3.5.9 Photophysical analysis of fluorescently modified DNA and corresponding duplexes

Sample preparation in aqueous buffer: The ONs were annealed similarly like thermal melting analysis. Duplexes of ONs **4–7** (10 µM) were made by heating a 1:1 mixture of H-Telo DNA and complementary ON **9** at 90 °C for 5 min in 10 mM Tris-HCl buffer (pH 7.5) containing either 100 mM NaCl or 100 mM KCl. The duplex (**4•8**, **5•8, 6•8** and **7•8)** solutions were slowly cooled to RT and were kept at ~4 °C for 1 h. The fluorescence measurements of the ONs (1 µM) were performed in triplicate.

Sample preparation in intraoocyte buffer: The ONs were annealed similarly like thermal melting analysis. Duplexes of ONs **4–7** (10 µM) were made by heating a 1:1 mixture of H-Telo DNA and complementary ON **9** at 90 °C for 5 min in intraoocyte buffer (25 mM HEPES buffer (pH = 7.5), 10.5 mM NaCl, 110 mM KCl, 0.13 µM $CaCl_2$ and 1 mM $MgCl_2$). The duplex (**4•8**, **5•8, 6•8** and **7•8)** solutions were cooled to RT overnight. The fluorescence measurements of the ONs (1 µM) were performed in triplicate.

Sample preparation in strontium (Sr^{2+}) containing Tris-HCl buffer: The ONs were annealed similarly like thermal melting analysis. Duplexes ONs **4–7** (10 µM) were made by heating a 1:1 mixture of H-Telo DNA and complementary ON **9** at 90 °C for 5 min in 50 mM Tris-HCl buffer (pH = 7.4) containing 150 mM $SrCl_2$. The duplex (**4•8**, **5•8, 6•8** and **7•8)** solutions were cooled to RT overnight. The fluorescence measurements of the ONs (1 µM) were performed in triplicate.

3.5.10 Fluorescence binding assay: PDS and BRACO19 binding to H-Telo DNA ONs 5 and 10

The GQ structure of H-Telo DNA ONs **5** and **10** was formed by annealing the ONs with respective conditions (aqueous buffer or intraoocyte buffer or strontium (Sr^{2+})containing Tris-HCl buffer), respectively.

3.5.11 Binding studies in three buffer conditions: A series of DNA ON samples (1 µM) in either aqueous buffer or intraoocyte buffer or strontium (Sr^{2+}) containing Tris-HCl buffer, increasing concentration of pyridostatin (PDS, 4 nM to 7.5 µM) or BRACO19 (BC, 4 nM to 5 µM) were prepared and incubated at RT for 30 min. Samples were excited at 330 nm with an excitation and emission slit width of 5 nm and 9 nm and in case of ON **10** slit width 5 nm and 6 nm, respectively. Fluorescence experiments were performed in triplicate in a micro fluorescence cuvette (Hellma, path length 1.0 cm) at 20 °C. Appropriate blank in absence of ONs, but containing respective concentration of the ligand, was subtracted from the individual spectrum.

From the dose-dependent quenching curves the apparent dissociation constants (K_d) for the binding of PDS or BRACO19 to H-Telo DNA ONs **5** and **10** in aqueous buffer, intraoocyte buffer and strontium(Sr^{2+}) containing Tris-HCl buffer were determined by fitting normalized fluorescence intensity (F_N) versus log of PDS or BRACO19 concentration plot to Hill equation (Origin 8.5).[24,41]

$$F_N = \frac{F_i - F_s}{F_0 - F_s}$$

F_i is the fluorescence intensity at each titration point. F_0 and F_s are the fluorescence intensity in the absence of ligand (L) and at saturation, respectively. n is the Hill coefficient or degree of cooperativity associated with the binding.

$$F_N = F_0 + (F_s - F_0)\left(\frac{[L]^n}{[K_d]^n + [L]^n}\right)$$

3.5.12 Crystallization and Data Collection:

The control unmodified DNA G-Quadruplex ONs **8** was prepared 3 mM solution in 20 mM potassium cacodylate (pH = 6.5) and 50 mM KCl. The buffered DNA annealed at 363 K for 5 min, slowly cooled to room temperature at 298 K overnight. Crystals have grown using hanging drop vapour diffusion methods at 4 °C. The DNA final concentration 1.2 mM in native GQ crystals were obtained using the well solution of 0.05 M sodium cacodylate (pH = 7.2), 0.4 M ammonium sulfate, 0.05 M KCl, 0.01 M $CaCl_2$, 15% PEG400. Gradients 1 µL of DNA sample was mixed with 0.5 µL of well solution for crystallization. Crystals grew in three months as hexagonal rods of dimensions (0.26 x 0.10 x 0.08 mm^3), the diffraction quality crystals are harvested and cryoprotected in a solution of the mother liquor and bringing PEG400 concentration to 30%.

The modified H-Telo ON **5** was first dissolved at 3 mM solution containing 20 mM potassium cacodylate buffer (pH = 7.2) and 50 mM KCl. The buffered DNA was annealed at 363 K for 5 min, followed by slow cooling to room temperature overnight. Crystals were grown using hanging drop vapour diffusion methods at 4 °C. The DNA final concentration 1.2 mM, crystals of **5** was achieved 0.05 M potassium cacodylate (pH = 7.2), 0.625 M ammonium acetate, 0.2 M KCl, 15% PEG400. One microliter of sample was mixed with 0.5 µL of well solution for crystallization respectively. Crystal grew in two months hexagonal rods (0.16 x 0.16 x 0.15 mm^3) they were harvested and cryoprotected in a solution of the mother liquor with an additional 15% PEG400, bringing total PEG 400 concentration to 30%.

After annealing a premixed solution of ON **6** and BRACO19 incubated at 25 °C for 1 hour for crystallization of H-Telo DNA **6**. The 1 µL drop was contained 1.3 mM H-Telo DNA **6**, 1.56 mM BRACO19, which is mixed with 0.5 µL of well solution 0.05 M potassium cacodylate (pH = 7.2), 0.7 M ammonium sulfate, 0.05 M KCl, 0.01 M $CaCl_2$ and 12.5% PEG400, to obtained crystal of **6**. The crystals appeared after four months as rhombic crystals (0.13 x 0.08 x 0.08 mm^3). They were procured and cryoprotected in a solution of the mother liquor in addition to PEG400 concentration bringing to 30% prior to X-ray diffraction experiments. Crystals were flash-frozen in liquid nitrogen and data collected at the UK Diamond Light Source Synchrotron and European Synchrotron Radition Facility (ESRF) at wavelength 0.9793 Å, beamline IDs I02, I04 and BM14, respectively.

***Structure solution and refinement**:* Data sets were collected at Diamond Light Source, UK, and European Synchrotron Radiation Facility, France. Data were processed using either imosflm[42] or XDS[43] and scaled using AIMLESS[44]. Structures were determined by molecular replacement method using (1KF1) as model.[15] The structures were refined using PHENIX[45] or REFMAC[46].

3.6 References:

1 Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. *Genes Dev.***1997**, *11*, 2801–2809.

2 (a) Collie, G.W.; Parkinson, G. N. *Chem. Soc. Rev*. **2011**, *40*, 5867–5892. (b) Bugaut, A.; Balasbramanian, S. *Nucleic Acids Res.* **2012**, *40*, 4727–4741.

3 Huppert, J. L.; Balasubramanian, S. *Nucleic Acids Res.* **2005**, *33*, 2908–2916.

4 Rhodes, D.; Lipps, H. J. *Nucleic Acids. Res.* **2015**, *43*, 8627–8637.

5 (a) Monchaud, D.; Teulade-Fichou, M. P. *Org. Biomol. Chem*. **2008**, *6*, 627–636. (b) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discov*. **2011**, *10*, 261–275. (c) Ohnmacht, S. A.; Neidle, S. *Bioorg. Med. Chem. Lett*. **2014**, *24*, 2602–2612.

6 (a) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. *Methods* **2012**, *57*, 64–75. (b) Adrian, M.; Heddi, B.; Phan, A. T. *Methods* **2012**, *57*, 11–24.

7 (a) Hazel, P.; Huppert, J. L.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc*. **2004**, *126*, 16405–16415. (b) Guédin, A.; Gros, J,; Alberti, P.; Mergny, J.-L. *Nucleic Acids Res*. **2010**, *38*, 7858–7868.

8 (a) Yaku, H.; Fujimoto, T.; Murashima, T.; Miyoshi, D.; Sugimoto, N. *Chem. Commun.* **2012**, *48*, 6203–6216. (b) Zhang, S.; Wu, Y.; Zhang, W. *ChemMedChem* **2014**, *9*, 899–911.

9 Heddi, B.; Phan, A. T. *J. Am. Chem. Soc.* **2011**, *133*, 9824–9833.

10 (a) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res*. **2006**, *34*, 5402–5415. (b) Murat, P.; Balasubramanian, S. *Curr. Opin. Genet. Dev*. **2014**, *25*, 22–29.

11 Dai, J.; Carver, M.; Yang, D. *Biochimie* **2008**, *90*, 1172–1183.

12 Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.

13 Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *J. Am. Chem. Soc*. **2006**, *128*, 9963–9970.

14 Xue, Y.; Kan, Z.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y.; Tan, Z. *J. Am. Chem. Soc.* **2007**, *129*, 11185–11191.

15 Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876–880.

16 (a) Xu, Y.; Kaminaga, K.; Komiyama, M. *J .Am. Chem. Soc*. **2008**, *130*, 11179–11184. (b) Martadinata, H.; Phan, A. T. *J .Am. Chem. Soc*. **2009**, *131*, 2570–2578.

17 Rache, A. D.; Mergny, J.-L. *Biochimie* **2015**, *115*, 194–202.

18 (a) Faverie, A. R.; Guédin, A.; Bedrat, A.; Yatsunyk, L. A.; Mergny, J.-L. *Nucleic Acids Res*. **2014**, *42*, *e*65. (b) Laguerre, A.; Stefan, L.; Larrouy, M.; Genest, D.; Novotna, J.; Pirrotta, M.; Monchaud, D. *J. Am. Chem. Soc.* **2014**, *136*, 12406–12414. (c) Bhasikuttan, A. C.; Mohanty, J. *Chem. Commun.* **2015**, *51*, 7581–7597.

19 Vummidi, B. R.; Alzeer, J.; Luedtke, N. W. *ChemBioChem* **2013**, *14*, 540–558.

20 (a) Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Chem. Commun.* **2006**, 401–402. (b) Gray, R. D.; Petraccone, L.; Trent, J. O.; Chaires, J. B.; *Biochemistry* **2010**, *49*, 179–194. (c) Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A. *ACS Chem. Biol*. **2015**, *10*, 1311–1318.

21 Nadler, A.; Strohmeier, J.; Diederichsen, U. *Angew. Chem., Int. Ed.* **2011**, *50*, 5392–5396.

22 Sproviero, M.; Fadock, K. L.; Witham, A. A.; Manderville, R. A.; Sharma, P.; Wetmore, S. D. *Chem. Sci*. **2014**, *5*, 788–796.

23 Dumas, A.; Luedtke, N. W. *J. Am. Chem. Soc.* **2010**, *132*, 18004–18007.

24 Tanpure, A. A.; Srivatsan, S. G. *Nucleic Acids Res*. **2015**, *43*, e149.

25 (a) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem*. **2013**, *5*, 182–186. (b) Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. *Nat. Chem*. **2014**, *6*, 75–80.

26 Neidle, S. *J. Med. Chem.* **2016**, *59*, 5987–6011.

27 Collie, G. W.; Campbell, N. H.; Neidle, S. *Nucleic Acids Res.* **2015**, *43*, 4785–4799.

28 (a) Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res*. **2006**, *34*, 2723–2735. (b) Petraccone, L.; Spink, C.; Trent, J. O.; Garbett, N. C.; Mekmaysy, C. S.; Giancola, C.; Chaires, J. B. *J. Am. Chem. Soc*. **2011**, *133*, 20951–20961.

29 (a) Rachwal, P. A.; Fox, K. R. *Methods* **2007**, *43*, 291–301. (b) Tran, P. L. T.; Mergny, J.-L.; Alberti, P. *Nucleic Acids Res.* **2011**, *39*, 3282–3294.

30 (a) Moore, M. J. B.; Schultes, C. M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F. A.; Wilson, W. D.; Neidle, S. *J. Med. Chem*. **2006**, *49*, 582–599. (b) Rodriguez, R.; Müller, S.; Yeoman, J. A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. *J. Am. Chem. Soc*. **2008**, *130*, 15758–15759.

31 Müller, S.; Pantoş, G. D.; Rodriguez, R.; Balasubramanian, S. *Chem. Commun*. **2009**, 80–82.

32 Zhao, C.; Wu, L.; Ren, J.; Xu, Y.; Qu, X. *J. Am. Chem. Soc*. **2013**, *135*, 18786−18789.

33 Hänsel, R.; Foldynová-Trantírková, S.; Löhr, F.; Buck, J.; Bongartz, E.; Bamberg, E.; Schwalbe, H.; Dötsch, V.; Trantírek, L. *J. Am. Chem. Soc*. **2009**, *131*, 15761–15768.

34 (a) Pedroso, I. M.; Duarte, L. F.; Yanez, G.; Baker, A. M.; Fletcher, T. M. *Biochem. Biophys. Res. Commun*. **2007**, *358*, 298–303. (b) Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J,-H.; Lv, P.; Ou, T.-M.; Tan, J.-H.; Li, D.; Gu, L.-Q.; Ren, J.; Zhao, Y.; Huang, Z.-S. *Cell Chem. Biol.* **2016**, *23*, 1261–1270.

35 Chen, F.-M. *Biochemistry* **1992**, *31*, 3769-3776.

36 (a) Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 37–41. (b) Doose, S.; Neuweiler, H.; Sauer, M. *ChemPhysChem* **2009**,*10*, 1389–1398.

37 Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. *J. Am. Chem.Soc*. **2012**, *134*, 2723−2731.

38 IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. *Pure Appl. Chem*. **1983**, *55*, 1273–1280.

39 Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem.* **1994**, *5*, 508–512.

40 Tanpure, A. A.; Srivatsan, S. G. *ChemBioChem* **2012**, *13*, 2392–2399.

41 (a) Shandrick, S.; Zhao, Q.; Han, Q.; Ayida, B. K.; Takahashi, M.; Winters, G. C.; Simonsen, K. B.; Vourloumis, D.; Hermann, T. *Angew. Chem., Int. Ed.* **2004**, *43*, 3177–3182. (b) (a) Tam, V. K.; Kwong, D.; Tor, Y. *J. Am. Chem. Soc.* **2007**, *129*, 3257–3266.

42 Leslie, A. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2006**, *62*, 48–57.

43 Kabsch W. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010** *D66* 125–132.

44 Evans, P. R. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *D69*, 1204–1214.

45 Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr*. **2002**, *58*, 1948–1954.

46 Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D* **2004**, *60*, 2126–2132.

Appendix-II: Characterization data of synthesized compounds

Table 9. Analysis of sugar puckering and glycosidic torsional angles of the quadruplex core.

ON **8**	χ	puckering	**1KF1**	χ	puckering
G2-G4			G2-G4		
G2	-159.4	C2′-exo	G2	-160.3	C2′-exo
G3	-122.3	C2′-endo	G3	-115.0	C2′-endo
G4	-109.4	C2′-endo	G4	-107.2	C2′-endo
G8-G10			G8-G10		
G8	-122.4	C2′-endo	G8	-135.5	C2′-endo
G9	-111.8	C2′-endo	G9	-117.7	C2′-endo
G10	-109.4	C2′-endo	G10	-109.6	C2′-endo
G14-G16			G14-G16		
G14	-125.4	C2′-endo	G14	-119.3	C2′-endo
G15	-109.7	C2′-endo	G15	-107.5	C2′-endo
G16	-110.0	C2′-endo	G16	-106.3	C2′-endo
G20-G22			G20-G22		
G20	-132.3	C2′-endo	G20	-129.1	C2′-endo
G21	-108.2	C2′-endo	G21	-117.0	C2′-endo
G22	-102.9	C2′-endo	G22	-115.8	C2′-endo

ON **5**	χ	puckering	ON **6**	χ	puckering
G2-G4			G2-G4		
G2	-159.4	C2′-exo	G2	-132.6	C2′-endo
G3	-127.8	C2′-endo	G3	-106.6	C2′-endo
G4	-110.0	C2′-endo	G4	-117.4	C2′-endo
G8-G10			G8-G10		
G8	-122.7	C2′-endo	G8	-122.2	C2′-endo
G9	-153.6	C2′-exo	G9	-112.4	C2′-endo
G10	-114.1	C2′-endo	G10	-112.5	C2′-endo
G14-G16			G14-G16		
G14	-125.7	C2′-endo	G14	-118.6	C2′-endo
G15	-110.5	C2′-endo	G15	-117.0	C2′-endo
G16	-106.6	C2′-endo	G16	-117.7	C2′-endo
G20-G22			G20-G22		
G20	-131.5	C2′-endo	G20	-126.0	C2′-endo
G21	-101.2	C2′-endo	G21	-93.7	C2′-endo
G22	-98.9	C2′-endo	G22	-113.4	C2′-endo

Table 10. Analysis of backbone and glycosidic torsional angles of the quadruplex loops.

KF1	α	β	γ	δ	ε	ζ	χ
loop 1							
T5	78.6	-169.7	-177.5	116.8	-117.8	-55.6	-115.6
T6	-153.9	144.1	32.5	84.1	-139.7	-77.6	-169.2
A7	-86.2	-152.0	62.7	144.3	-69.0	-121.5	-125.9

loop 2							
T11	77.0	-171.0	176.3	123.3	-110.9	-62.5	-99.3
T12	-142.5	151.5	18.2	91.4	-150.7	-67.1	-164.8
A13	-75.1	-176.8	67.3	132.5	-75.9	-67.0	-155.0
loop 3							
T17	63.1	-176.2	-165.1	138.7	-124.4	-84.3	-100.0
T18	148.3	-160.4	57.5	85.5	-168.2	-40.5	-160.4
A19	-91.8	-133.7	45.5	147.7	-91.3	-93.2	-117.0

ON **8**	α	β	γ	δ	ε	ζ	χ
loop 1							
T5	72.1	-171.6	-168.4	122.1	-127.8	-65.6	-112.2
T6	161.1	179.4	54.1	83.8	-142.3	-65.7	-168.2
A7	-68.3	-161.8	54.4	142.3	-88.6	-83.5	-126.4
loop 2							
T11	67.0	-167.9	-166.0	117.5	-119.9	-58.8	-111.6
T12	168.9	175.6	47.0	85.2	-144.5	-65.4	-172.0
A13	-60.0	-170.4	57.3	148.1	-83.0	-70.4	-142.2
loop 3							
T17	78.5	-177.1	-171.2	125.5	-126.8	-66.8	-111.3
T18	158.8	-179.3	55.9	84.8	-140.9	-63.6	-168.1
A19	-64.9	-156.5	46.7	145.1	-90.3	-88.5	-124.9

ON **5**	α	β	γ	δ	ε	ζ	χ
loop 1							
T5	65.1	-170.1	-162.1	128.0	-127.1	-69.1	-105.8
T6	148.5	-175.4	64.6	85.5	-138.6	-67.2	-163.0
A7	-65.5	-163.3	50.1	142.9	-90.7	-83.7	-126.3
loop 2							
T11	59.8	-173.0	-174.6	123.4	-114.8	-52.1	-92.2
T12	155.3	-176.5	60.5	79.5	-137.9	-65.0	-175.9
A13	-66.2	-174.0	58.0	146.6	-155.5	63.8	-154.0
loop 3							
T17	73.3	-170.8	-167.6	118.2	-126.8	-67.5	-116.6
T18	148.6	-176.8	68.1	83.3	-135.3	-66.2	-167.9
A19	-64.2	-160.2	49.9	145.2	-92.9	-80.3	-124.3

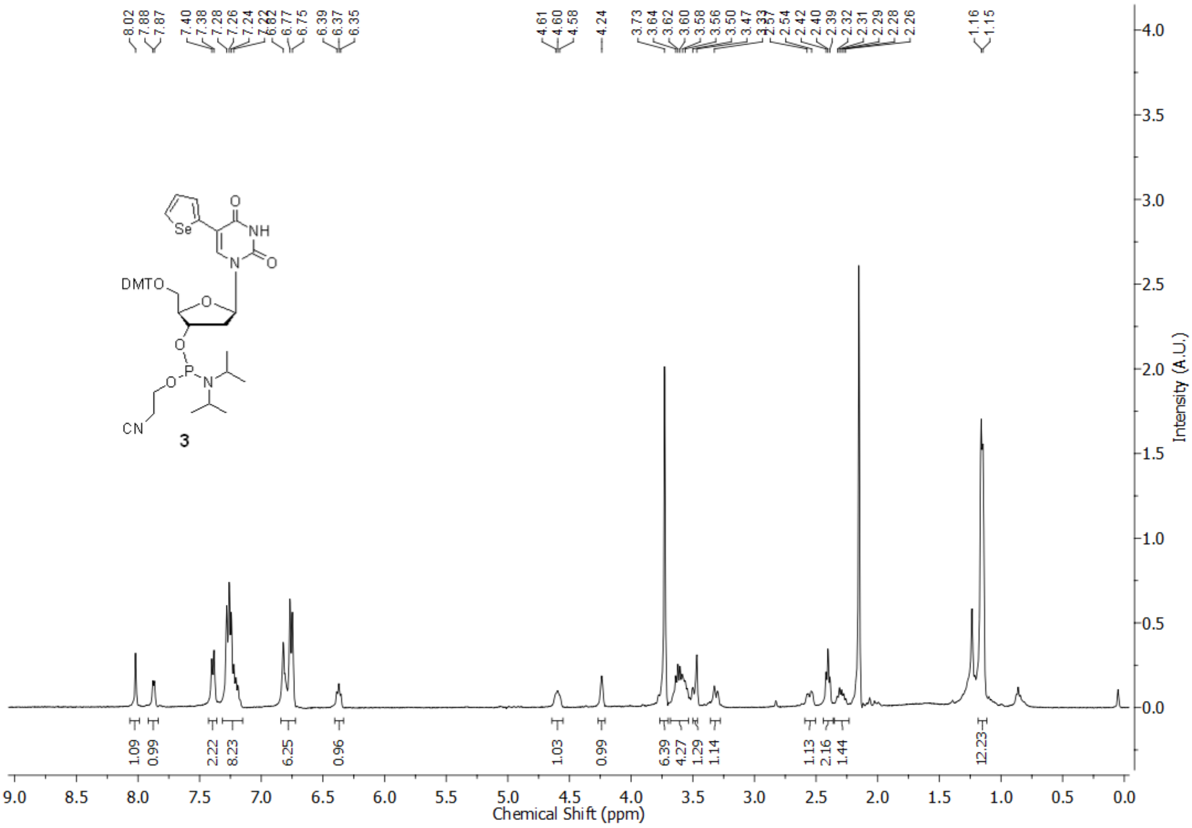
ON **6**	α	β	γ	δ	ε	ζ	χ
loop 1							
T5	65.8	-176.9	-159.2	126.1	-133.9	-65.7	-111.1
T6	170.3	-179.0	47.8	85.6	-139.3	-66.5	-163.4
A7	-63.8	-168.2	48.8	140.1	-78.5	-134.1	-136.0
loop 2							
T11	71.1	179.1	173.0	124.4	-70.7	-104.5	-101.8
T12	55.8	-78.4	63.3	61.5	59.9	110.3	169.9
A13	-86.8	149.9	28.7	137.5	-160.7	62.7	-148.3
loop 3							

T17	68.2	176.2	-162.8	131.8	-130.7	-67.6	-105.2
T18	170.8	179.3	48.0	87.0	-140.6	-62.1	-166.7
A19	-65.2	-163.5	48.3	139.8	-85.0	-83.8	-130.8

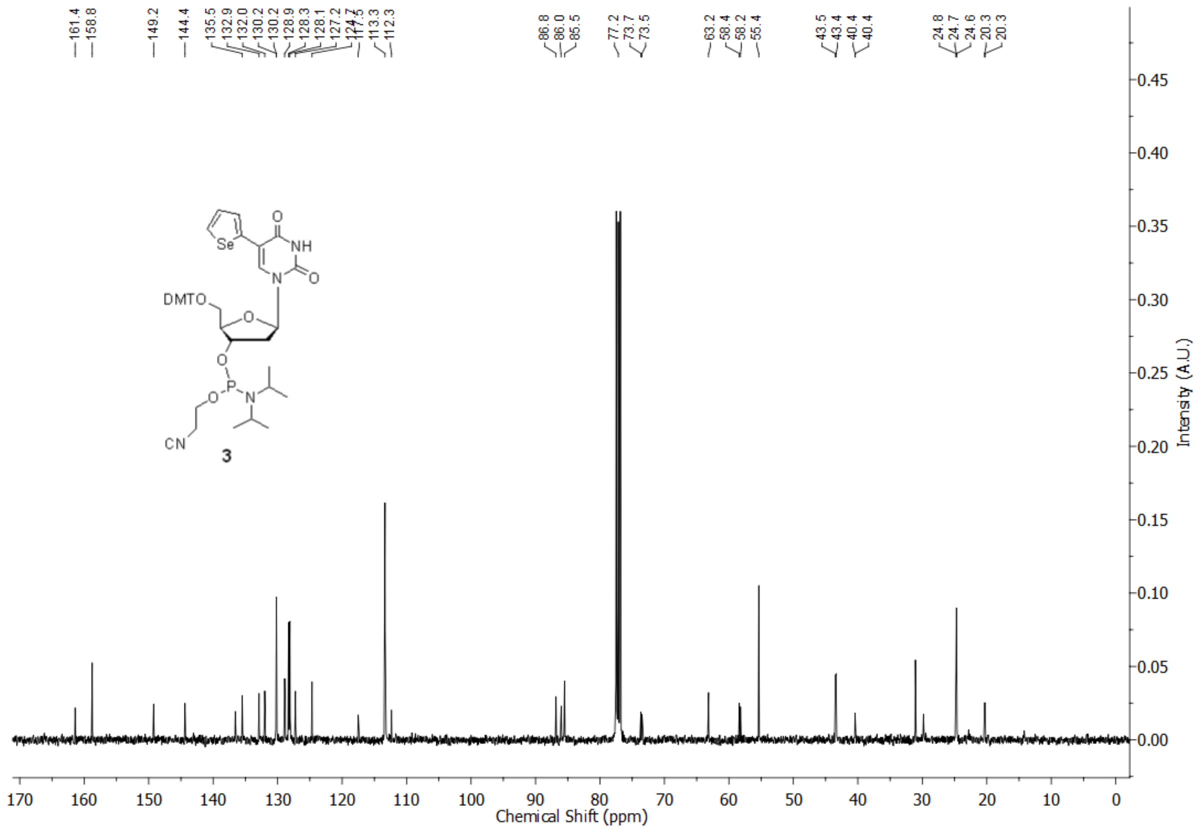
Table 11. Analysis of sugar puckering of control and modified nucleotides of the loops.

loop	Puckering (KF1)	SG-native (**8**)	Puckering (**5**)	Puckering (**6**)
loop 1				
T5	C2′-endo	C2′-endo	C2′-endo	C2′-endo
T6	C2′-exo	C2′-exo	C2′-exo	C2′-exo
A7	C2′-endo	C2′-endo	C2′-endo	C2′-endo
loop 2				
T11	C2′-endo	C2′-endo	C2′-endo	C2′-endo
T12	C2′-exo	C2′-exo	C2′-exo	C2′-exo
A13	C2′-endo	C2′-endo	C2′-endo	C2′-endo
loop 3				
T17	C2′-endo	C2′-endo	C2′-endo	C2′-endo
T18	C2′-exo	C2′-exo	C2′-exo	C2′-exo
A19	C2′-endo	C2′-endo	C2′-endo	C2′-endo

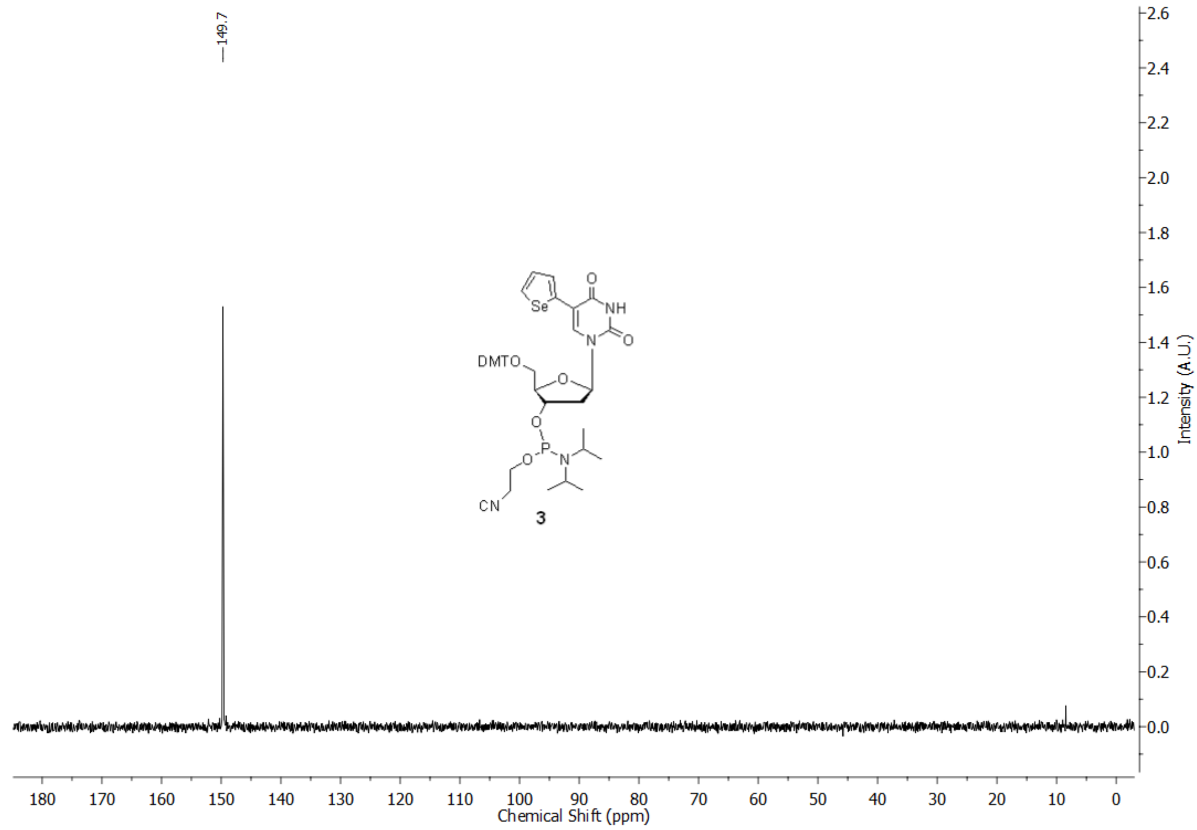
^{1}H NMR of 5-Selenophene phosphoramidite substrate **3** in $CDCl_3$



^{13}C NMR of 5-Selenophene phosphoramidite substrate **3** in $CDCl_3$



^{31}P NMR of the 5-Selenophene phosphoramidite substrate **3** in $CDCl_3$ (162 MHz)



Chapter 4A

Surface-tuned and metal-ion-responsive supramolecular gels based on thymidine nucleolipids

4A.1 Introduction

Exquisite recognition properties and the ability to adopt as well as switch between complex structures depending on the sequence, physical conditions, and ligands have rendered nucleic acids as powerful supramolecular synthons for assembling functional nanoarchitectures.[1–4] Many design strategies using DNA and RNA oligonucleotides have enabled the development of nanomachines and containers,[5–7] synthetic membranes and channels,[8–10] and therapeutic and diagnostic tools.[11,12] However, large-scale fabrication of such nucleic acid nanostructures has remained a major challenge. While constant effort is being put to address this concern, the basic components of nucleic acids, nucleobases, and nucleosides and their derivatives are proving to be highly useful synthons for constructing a variety of supramolecular architectures.[13–15] The advantages of nucleoside-based supramolecular synthons are that along with their inherent H-bonding, π-stacking, and metal-ion-binding abilities additional functionalities can be conveniently imparted to the base and or sugar residues by established synthetic routes. Such tailor-made nucleobase- and sugar-modified nucleoside derivatives have enabled the construction of a diverse variety of self-assemblies like vesicles, fibers, gels, and synthetic ion channels.[16–19] For example, nucleoside-lipid hybrids, also called nucleolipids, which combine the self-assembling character of amphiphilic lipids and base pairing property of nucleobases, have been shown to form vesicles, microspheres, and hydrogels.[20–26] Multicomponent systems like glycosyl/nucleoside/lipid/peptide hybrids have also been introduced as efficient hydrogelators.[27,28] Many of these molecular containers have been successfully used as drug and gene delivery systems. Interestingly, the ability of G-rich sequences to form tetrads has also been invoked at the nucleoside level to assemble ion-selective channels and nanowires by using guanosine and its derivatives.[29–32]

Most of the design strategies have been dedicated toward developing biocompatible nucleolipids for biotechnological applications. However, the potential of nucleoside hybrids in constructing smart self-assembled materials, which are tunable and stimuli-responsive, and showing interesting optical and sensing properties has recently gained significant attention. For example, the ability of achiral nucleobases to trigger the helical assembly and transfer the chiral sense of *N*-(9-fluorenylmethoxycarbonyl)-(Fmoc)-protected glutamic acid to an achiral cationic dye has been elegantly demonstrated.[33] Using this combination of chirality transfer and fluorescence, a hydrogel emitting circularly polarized light was developed. Gazit and co-workers used simple guanine-containing peptide nucleic acid dimers and monomers to construct rods and spheroids, which served as good organic light-emitting materials and photonic crystals.[34,35] Recently, we have developed a new family of environment-sensitive fluorescent uridine nucleolipids by conjugating heterobicycle aryl moieties to the uracil base and long-chain fatty acids to the 2′-*O*- and 3′-*O*-positions of the ribose sugar. The fluorescent nucleolipids exhibited good solvatochromism and viscochromism and formed supramolecular organogels driven by hierarchical structures such as fibers, helical ribbons, and nanotubes.[36] In addition, they displayed aggregation-induced enhanced emission and were found to be responsive to various external stimuli.

In the examples discussed above, noncovalent interactions like H-bonding, π–π stacking, and van der Waals and hydrophobic effect drive the formation of hierarchical supramolecular assemblies, which sometimes lead to the formation of gels. Typically, in these systems the gelator is dissolved in a pure or combination of organic solvents or in water by heating, and then the respective gels are formed upon cooling or sonication. However, certain gelators that do not form gels in organic solvents or water have been induced to form organogels by adding small amounts of water to a solution of a gelator in an organic solvent.[37,38] The mechanism of water-induced gelation is intriguing as water is known to break the H-bonding between the gelator molecules, often leading to precipitation. Nevertheless, this water-induced gelation strategy has been used to tune the supramolecular chirality in organogels, promote organogelation of an oleanlic acid-adenine conjugate, and in understanding the subtle balance between dissolution and aggregation of small molecular gelators.[39–41] However, the utility of this supramolecular assembling strategy has not been well explored in fabricating nucleoside-lipid hybrid materials with tunable properties.

4A.2 Design of thymidine nucleolipids

During the course of our continuous effort to develop smart nucleolipid materials, we stumbled upon simple thymidine-fatty acid hybrids exhibiting interesting gelation behaviour, which was found to influence the surface hydrophobicity-hydrophilicity and metal ion sensing ability of the gels (Figure 1). Here, we describe the development of nucleolipids with different hydrophilic-hydrophobic quotients, prepared by attaching different fatty acid hydrocarbon chains to the 3′-*O* or 3′,5′-*O* positions of the sugar residue of thymidine. While 3′,5′-*O*-attached and 5′-*O*-attached[42] fatty acid nucleolipids formed organogels in various pure organic solvents by conventional heating-cooling steps, 3′-*O*-attached fatty acid nucleolipids exhibited water-induced organogelation. The morphology and non-covalent interactions that drive the hierarchical self-assembling process were investigated by SEM, single crystal and powder X-ray diffraction analyses. Further, ^{1}H NMR analysis provided insights on the molecular interactions invoked by nucleobase, sugar, fatty acid chains and water in the water-induced gelation of 3′-*O*-mono-fatty acid-substituted nucleolipids. Interestingly, the surface of the xerogel film fabricated using monosubstituted thymidine gels formed in toluene-water mixture was highly hydrophobic, whereas a surface fabricated using the gel formed in a polar organic solvent-water mixture (methanol-water) was hydrophilic. On the other hand, the gelation process of 3′,5′-*O*-di-fatty acid-substituted nucleolipids was found to be highly sensitive to the presence of Hg^{2+} ions as a result of a specific interaction between Hg^{2+} ion and N3 nitrogen of thymine nucleobase. Collectively, this simple design of self-assembling supramolecular nucleolipid synthons, which has enabled the construction of gels with tunable and metal ion-responsive properties, could be also applied to other nucleosides to develop novel functional materials.

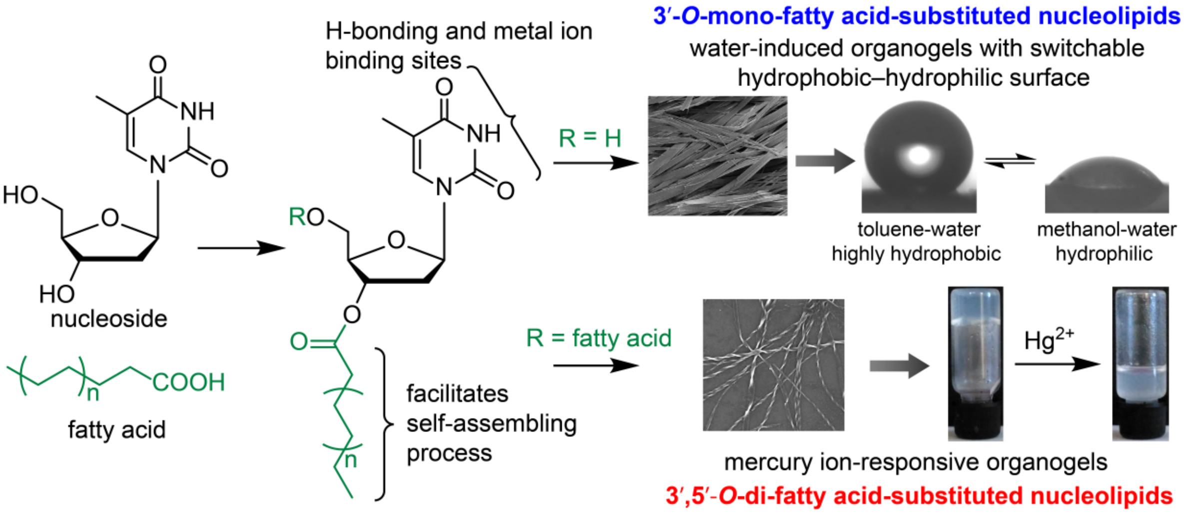
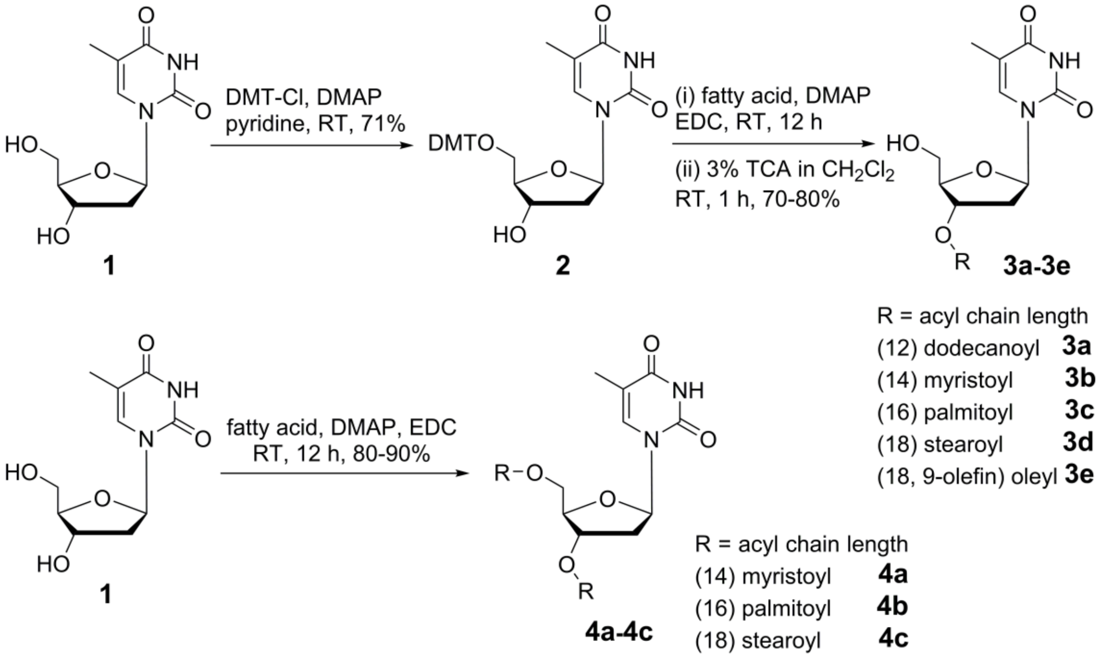


Figure 1. Design of self-assembling thymidine nucleolipids, which show different gelation behaviour, morphology, surface tunability and metal ion responsiveness depending on the site of attachment of fatty acid acyl chain onto the sugar residue. Water induces the supramolecular gelation of 3′-*O*-mono-fatty acid-substituted nucleolipids dispersed in organic solvents. The surface of the xerogel films could be tuned between highly hydrophobic and hydrophilic by using an appropriate organic solvent-water mixture. 3′,5′-*O*-di-fatty acid-substituted nucleolipids form organogels, which is highly responsive to the presence of Hg^{2+} ions.

4A.3 Results and Discussion

4A.3.1 Synthesis of thymidine nucleolipids

The nucleoside-fatty acid hybrids of different amphiphilicity were synthesized by attaching fatty acid chains of different lengths at the 3′-*O*- and 3′,5′-*O*-positions of thymidine via an ester linkage (Scheme 1). In this design, the nucleoside serves as a head group and long chain fatty acid attached to the sugar serves as a lipophilic group. The 3′-*O*-monosubstituted nucleolipids **3a**–**3e** were synthesized by first protecting the 5′-OH group with acid-labile dimethoxytrityl (DMT) moiety followed by a coupling reaction with various saturated and unsaturated fatty acids in the presence of *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide (EDC) hydrochloride. Deprotection of the DMT group in the presence of trichloroacetic acid (TCA) gave the desired monosubstituted nucleolipids **3a**–**3e** in good yields. 3′,5′-*O*-di-fatty acid-substituted thymidine nucleolipids **4a**–**4c** were synthesized in good yields by coupling thymidine with fatty acids in the presence of EDC.



Scheme 1. Synthesis of 3′-*O*-fatty acid-substituted (**3a-3e**) and 3′,5′-*O*-di-fatty acid-substituted (**4a**–**4c**) thymidine nucleolipids. DMT-Cl = 4,4′-dimethoxytrityl chloride; DMAP = 4-dimethylaminopyridine; EDC = *N*-(3-dimethylamino)propyl)-*N*′-ethylcarbodiimide hydrochloride.

4A.3.2 Gelation behavior of nucleolipids

The ability of mono- and disubstituted nucleolipids to form gels was studied in various solvents by an inverted vial method. Both types of nucleolipids were practically insoluble in water. Monosubstituted nucleolipids **3a**–**3e** were dissolved in a range of polar and nonpolar organic solvents by heating. Upon cooling or ultrasonication, which is known to induce gelation by facilitating intermolecular interactions, these nucleolipids did not form gels. Rewardingly, addition of a small amount of water into the organic dispersion of the nucleolipids (**3a**–**3d**) containing saturated fatty acid chains resulted in the spontaneous formation of opaque gels irrespective of the miscibility of water in the organic solvent (Figure 2A, 3 and Table 1). The amount of water required to obtain a stable gel was determined by adding increasing percentage of water into samples of a particular concentration of the gelator in organic solvent. For example, 20% of water was required to induce the formation of stable gels in nucleolipids **3a**–**3d** dispersed in DMSO at respective critical gelation concentration (CGC). Addition of water beyond 20% led to precipitation or formation of a suspension (Figure 2B).

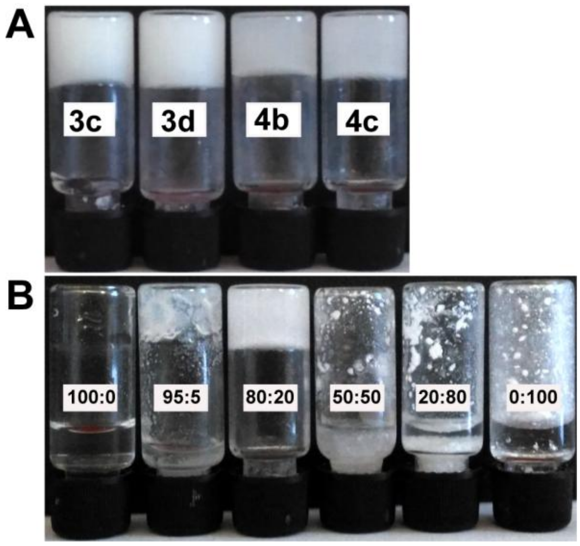


Figure 2. (**A**) A representative photograph of water-induced gels of monosubstituted nucleolipids **3c** and **3d** obtained in a mixture of DMSO-water (80:20) and organogels of disubstituted nucleolipids **4b** and **4c** in DMSO at respective CGC (Table 1). (**B**) Photograph showing the effect of added water on the gelation of **3c**. Left to right: amount of water (v %) was increased from 0−100%.

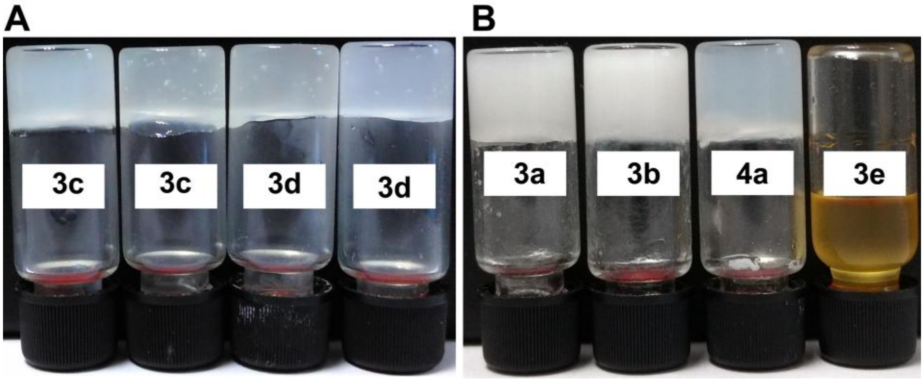


Figure 3. (**A**) Photograph showing the water-induced gelation of thymidine nucleolipids at respective CGC values. **3c** (methanol:H_2O = 80:20), **3c** (toluene:H_2O = 80:20), **3d** (DMF:H_2O = 86:14) and **3d** (methanol:H_2O = 80:20). (**B**) Photograph of gels of **3a** (DMSO:H_2O = 80:20), **3b** (DMSO:H_2O = 80:20), **4a** (DMSO) and **3e** (DMSO:H_2O = 80:20) light-yellow in colour.

The gelation capacity of 3′-*O*-monosubstituted nucleolipids was found to depend on the nature and length of the fatty acid alkyl chain. As the fatty acid alkyl chain length of nucleolipids was increased from dodecanoyl to stearoyl (**3a–3d**), the gelation capacity in terms of CGC increased in all solvents that supported a water-induced organogelation process (Table 1). Notably, this water-induced gelation process produced nucleolipid gels, which were found to be thermoreversible and stable at room temperature for months. Expectedly, nucleolipid **3e** made of an unsaturated oleyl chain did not support the gelation process in any of the solvents tested in this study. These observations reveal that a nucleolipid containing a longer and saturated fatty acid chain requires a smaller amount of the gelator to form a stable gel, possibly due to the enhanced hydrophobicity and van der Waals interaction resulting in a better packing as compared to nucleolipids containing shorter and unsaturated chains.[43–45]

In contrast, 3′,5′-*O*-disubstituted nucleolipids **4a**–**4c** dissolved in DMSO formed stable opaque organogels upon heating-cooling/sonication steps (Figure 2A). In acetonitrile, **4b** and **4c** formed stable gels. However, other organic solvents or addition of water did not facilitate the gelation of the disubstituted nucleolipids. Again, a nucleolipid containing longer fatty acid acyl chains exhibited lower CGC values as compared to systems containing shorter hydrocarbon chains. It is important to mention here that 5′-*O*-mono-fatty acid-substituted thymidine reported by Kim and co-workers also formed organogels in several pure nonpolar organic solvents.[42] Taken together, these results indicate that depending on the site and number of fatty acid chains on the nucleoside either conventional heating-cooling steps or water-induced gelation procedure could be adopted to obtain supramolecular gels.

Table 1. Gelation properties of thymidine nucleolipids **3a–3e** and **4a–4c**

Solvent	**3a**	**3b**	**3c**	**3d**	**3e**	**4a**	**4b**	**4c**
DMSO	G (2.2, 20)	G (1.8, 20)	G (1.4, 20)	G (1.2, 18)	S	G (3.0)	G (2.1)	G (1.1)
DMF	G (3.0, 15)	G (2.2, 15)	G (1.8, 13)	G (1.4, 14)	S	S	P	P
ACN	G (2.0, 13)	G (1.5, 12)	G (1.5, 9)	G (1.4, 8)	S	P	G (1.5)	G (1.2)
methanol	G (2.4, 19)	G (1.5, 18)	G (1.3, 20)	G (1.1, 20)	S	P	PG	PG
CCl_4	G (3.2, 16)	G (2.3, 15)	G (2.0, 12)	G (1.8, 10)	S	S	PG	PG
dioxane	G (3.0, 20)	G (2.8, 20)	G (2.5, 20)	G (2.0, 18)	S	S	S	S
toluene	G (3.1, 20)	G (3.0, 20)	G (2.1, 20)	G (2.0, 18)	S	S	S	S
heptane	I	P	G (2.1, 20)	P	S	P	E	E

I: insoluble, P: precipitate, E: emulsion, G: stable gel, PG: partial gel, S: sol. In parenthesis, critical gelation concentration (CGC, w/v %) and amount of water (v %) required to form stable water-induced gels, respectively, are given.

4A.3.3 Rheological analysis

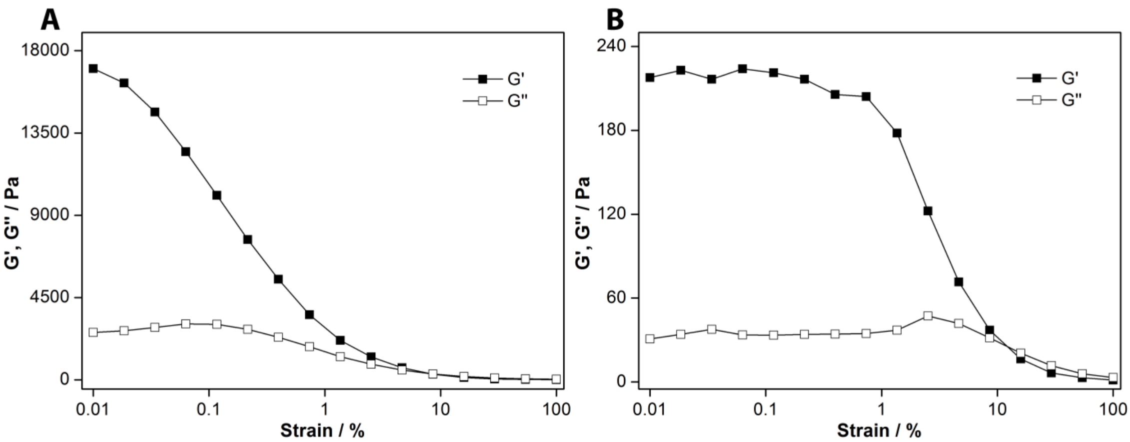


Figure 4. Strain-sweep rheological measurements of nucleolipid gel **3c** formed in DMSO-H_2O (80:20) (**A**) and **4c** in DMSO (**B**) at respective CGC at constant oscillating frequency.

The mechanical properties of gels of palmitoyl-containing nucleolipids **3c** and **4c** were evaluated by rheological analysis. The storage modulus (G′) and loss modulus (G″) of the gels were measured as a function of increasing shear strain at constant oscillating frequency. At a low strain value, gels of **3c** and **4c** exhibited a G′ value of 17020 and 217 Pa, respectively, which was nearly 1 order of magnitude greater than corresponding G″ values (2593 and 30 Pa, Figure 4). The crossover point of G′ and G″, which indicates the gel to sol transformation point, was observed at ~8% and 12% strain for the gels of **3c** and **4c**. These results reveal the elastic character of the gels. In particular, very high storage modulus displayed by the gel of **3c** as compared to **4c** indicates that the gel of monofatty acid-substituted thymidine, triggered by water, has significantly higher viscoelastic character.

4A.3.4 Morphology of nucleolipid gels

The morphology of self-assembled nucleolipid gels was analyzed by field emission scanning electron microscopy (FESEM). The SEM images of xerogel of 3′-*O*-monosubstituted and 3′,5′-*O*-disubstituted nucleolipids exhibited distinctly different morphologies (Figure 5). To a hot dispersion of 3′-*O*-palmitoyl- and 3′-*O*-stearoyl-tagged thymidines (**3c** and **3d**, respectively) in DMSO was added water (20%), and the clear dispersion was quickly drop casted on a silicon wafer substrate. The resulting gel was dried under vacuum and was subjected to SEM analysis, which revealed the formation of long-range sheets of 2–8 μm in width and several micrometers in length (Figure 5A and 5B). Similarly, in other polar as well as nonpolar organic solvents, images of water-triggered gels of monosubstituted thymidines revealed the formation of highly entangled long-range sheets (Figure 6). In order to further understand the role of water in the gelation process, SEM images of xerogels of **3c** at different DMSO to water content were compared with the SEM image of the aggregate in neat DMSO. In DMSO, **3c** formed a fibrous network, which, however, did not support the gelation process (Figure 2B, Figure 7). Addition of a small amount of water (5%) resulted in the formation of a partial gel, which exhibited flaky morphology. Upon increasing the water content to 20%, wherein a stable gel was formed, the xerogel of **3c** displayed long-range sheets, which were highly entangled (Figure 7C).

This morphological transition from fiber to flake to sheet as a function of increasing water content points out that a hierarchical self-assembling process is responsible for the formation of nucleolipid gels.[46] Further addition of water resulted in the formation of a suspension or precipitate, which was due to the breakdown of an entangled network structure as evident from the SEM images (Figure 2B, Figure 7). Addition of an aqueous solution of other nucleosides (A/T/G/C, 1–2 equiv) into the organic dispersion of **3c** at its CGC did not affect the gelation process or morphology (data not shown). On the other hand, the xerogels of 3′,5′-*O*-disubstituted nucleolipids **4b** and **4c**, containing palmitoyl and stearoyl chains, formed long-range twisted tapes of width nearly 110 and 150 nm, respectively (Figure 5C and 5D). Morphological analysis clearly indicates that the gelation mechanism differs depending on the site as well as number of fatty acid chain attached to the nucleoside.

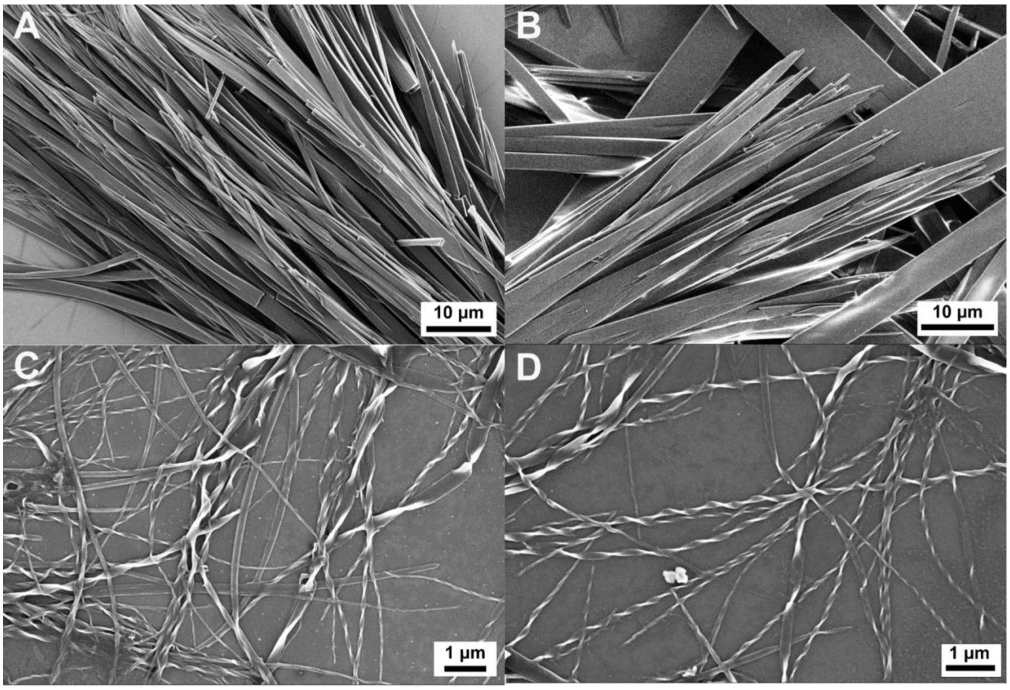


Figure 5. FESEM images showing the morphology of xerogel of nucleolipids containing palmitoyl and stearoyl chains. Xerogel of **3c** (**A**) and **3d** (**B**) obtained from the DMSO-water mixture (80:20). Xerogel of **4b** (**C**) and **4c** (**D**) obtained from DMSO.

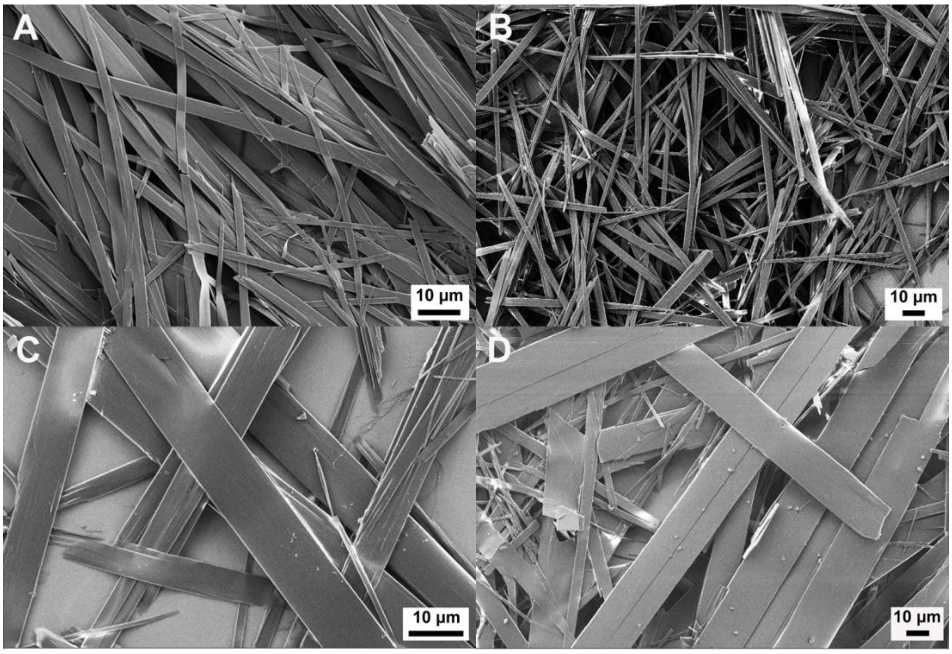


Figure 6. FESEM images of xerogel of nucleolipids. (**A**) **3c** in DMF, (**B**) **3c** in dioxane, (**C**) **3d** in DMF and (**D**) **3d** in methanol.

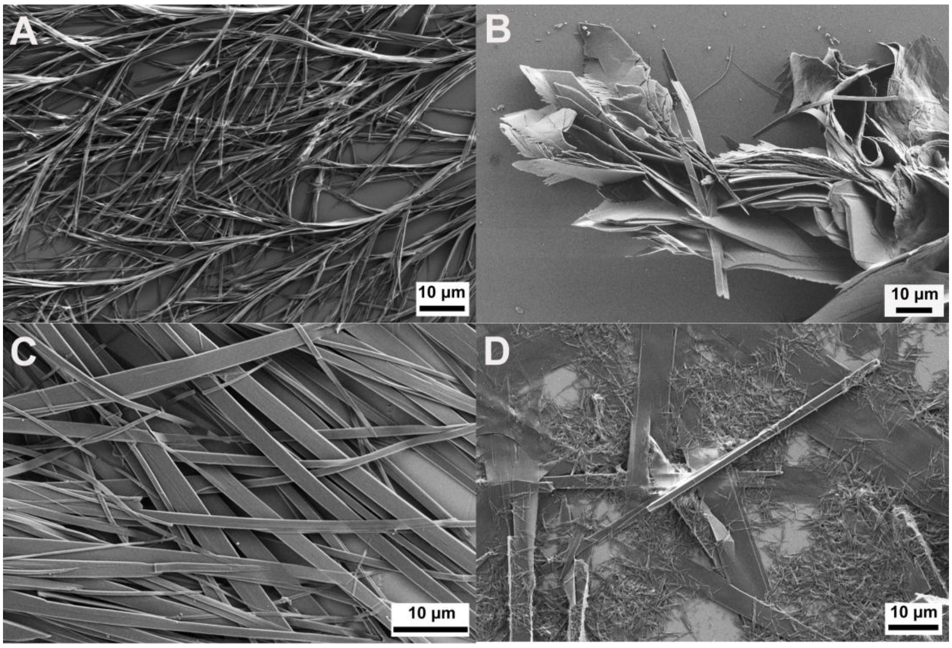
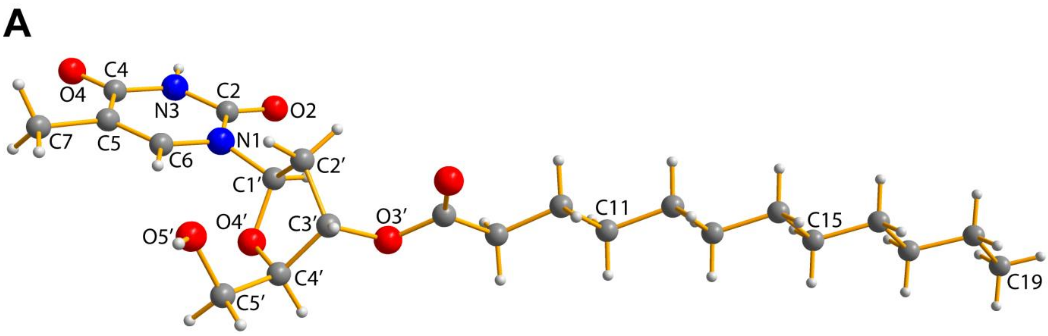


Figure 7. (**A**) FESEM image of nucleolipid **3c** dispersed in neat DMSO. (**B-D**) FESEM images of nucleolipid **3c** in different ratios of DMSO-water. (**B**) Partial gel, DMSO-H_2O (95:5). (**C**) Gel, DMSO-H_2O (80:20). (**D**) Precipitate, DMSO-H_2O (50:50).

4A.3.5 Driving force for the water-induced gelation

To understand the role played by water in the gelation process and also determine the interactions that drive the hierarchical assembly, we relied on single-crystal and powder X-ray diffraction (PXRD) and ^{1}H NMR spectroscopy techniques.

4A.3.5.1 Single crystal X-ray diffraction analysis



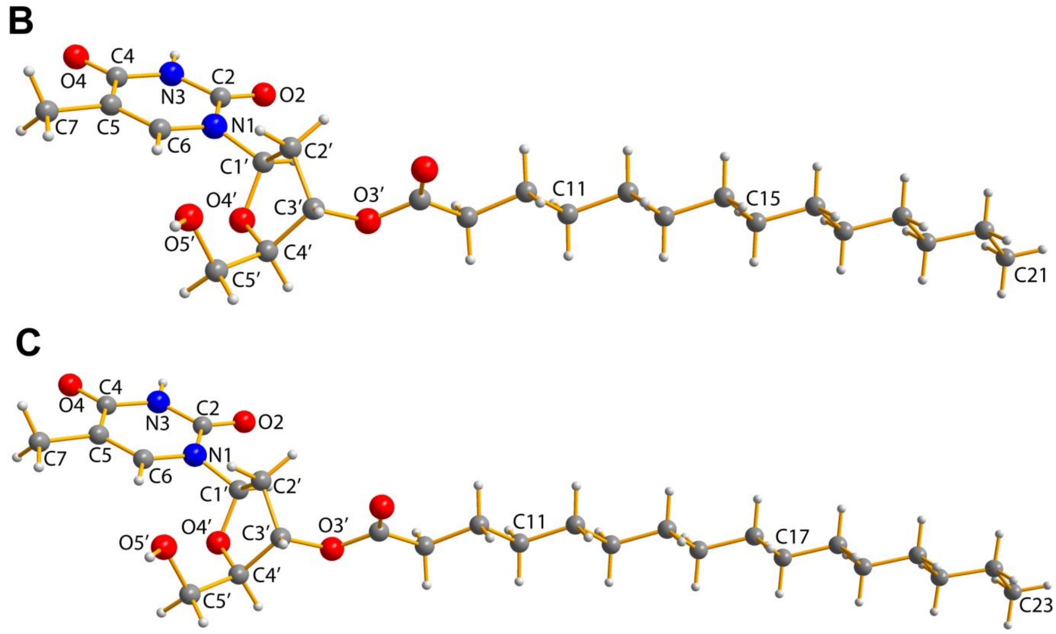


Figure 8. Single-crystal X-ray structure of nucleolipids (**A**) **3a**, (**B**) **3b** and (**C**) **3c** showing one molecule in the unit cell. Thymidine in all the nucleolipids adopts a C2′-endo *anti* conformation. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

Irrespective of the fatty acid chain length, all three nucleolipids (**3a–3c**) crystallized in orthorhombic system with the space group $P2_12_12_1$ (see experimental section). One molecule of the nucleolipid was present in the asymmetric unit (Figure 8). The conformation of the thymine base relative to the deoxy ribose sugar ring (*syn/anti*) in the nucleolipids was determined by measuring the torsion angle between C2-N1-C1′-O4′ atoms. Consistent with the recommendations of IUPAC,[47] a torsion angle of ~239° indicated that thymidine adopted a C2′-endo *anti* conformation in **3a–3c** (Table 2). Crystal packing of 3′-*O*-palmitoyl-substituted thymidine **3c** showed a 2D layered sheet structure formed by interdigitation of fatty acid acyl chains and multiple H-bonding interactions involving nucleosides. Each nucleoside interacted with four other nucleosides in the layered network (Figure 9A). Unlike in native Watson-Crick base pairing, the N3-H and C4-O atoms of one thymine residue were found to H-bond with N3-H and C4-O atoms of two adjacent thymine residues, respectively (Table 2). Further, O4′ and O5′-H atoms of the sugar residue interacted via H-bonds with O5′-H and O4′ atoms, respectively, of two other thymidine residues. These interactions essentially resulted in the formation of a 1D layered structure. The layers thus formed were further extended into 2D sheets through strong hydrophobic effect facilitated by the interdigitation of fatty acid acyl chains (Figure 9B). Crystal structure of nucleolipids **3a** and

3b containing a dodecyl and myristoyl chain, respectively, also exhibited similar H-bonding interactions and packing structure (Figures 10 and 11 and Table 2).

Table 2. H-bonding distances and angles, torsional angles measured from the crystal structure of thymidine nucleolipids **3a**, **3b** and **3c**.

nucleolipid	hydrogen bond	distance (Å)	angle (°)	torsion angle (χ) (°)
3a	N3H---04	2.005(4)	165.9(1)	C2-N1-C1′-O4′ 239.1(3)
	O4---HN3	2.005(4)	165.9(1)	
	O5′H---O4′	2.028(6)	150.6(2)	
	O4′---HO5′	2.028(6)	150.6(2)	
3b	N3H---04	1.975(1)	166.1(2)	C2-N1-C1′-O4′ 237.8(3)
	O4---HN3	1.975(1)	166.1(2)	
	O5′H---O4′	2.020(6)	146.6(3)	
	O4′---HO5′	2.020(6)	146.6(3)	
3c	N3H---04	1.970(2)	165.9(8)	C2-N1-C1′-O4′ 239.1(1)
	O4---HN3	1.970(2)	165.9(8)	
	O5′H---O4′	2.007(3)	147.4(3)	
	O4′---HO5′	2.020(6)	146.6(3)	

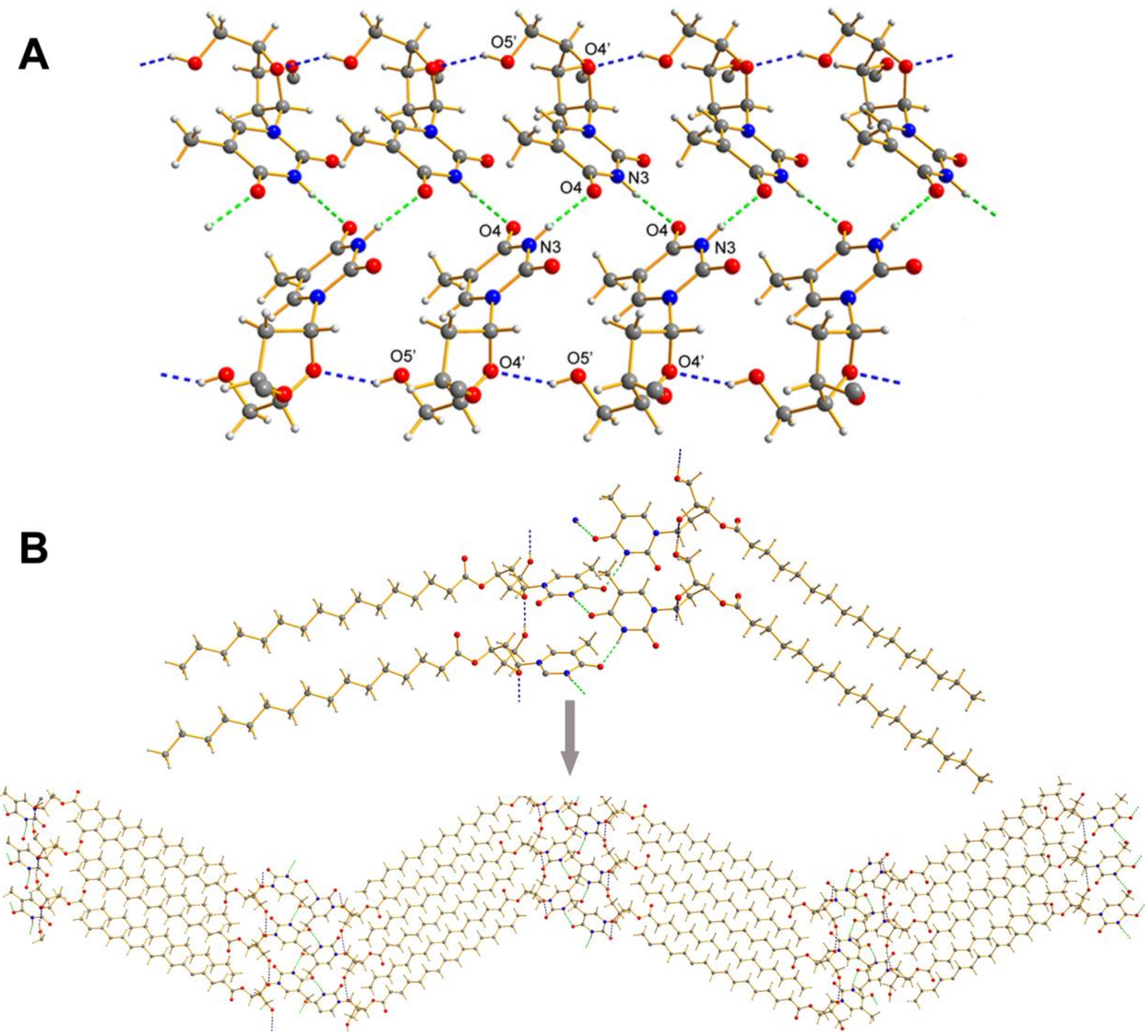


Figure 9. (**A**) X-ray crystal structure of **3c** along the crystallographic c-axis showing a detailed view of the H-bonding interactions. Intermolecular H-bonding interactions between the N3-H and C4-O atoms of thymine bases are shown in green dashed lines. H-bonding interactions between the O4′ and O5′-H atoms of sugar residues are shown in blue dashed lines. Fatty acid chains are not shown for clarity. (**B**) Packing diagram of nucleolipid **3c** along the crystallographic b-axis. 1D layer formed by H-bonding interactions between the nucleobases and sugars is further extended into a 2D sheet by the interdigitation of the fatty acid acyl chains. Head to head and tail to tail interaction is clearly seen. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

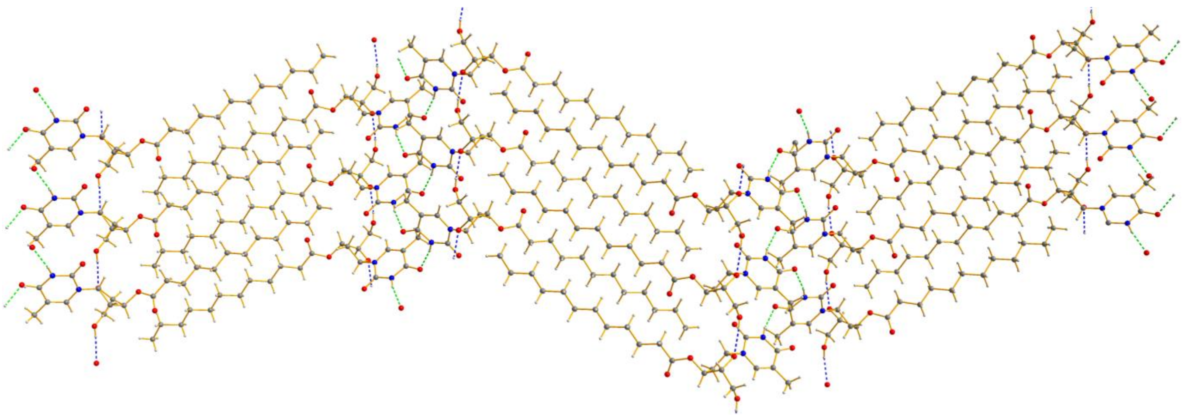


Figure 10. X-ray crystal packing structure of **3a** showing a view of the fully interdigitated alkyl chains and a detailed view of the H-bonding interactions along the crystallographic b-axis. Intermolecular H-bonding interactions between the N3-H and C4-O atoms of thymine bases are shown in green dashed lines. H-bonding interactions between the O4′ and O5′-H atoms of sugar residues are shown in blue dashed lines. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

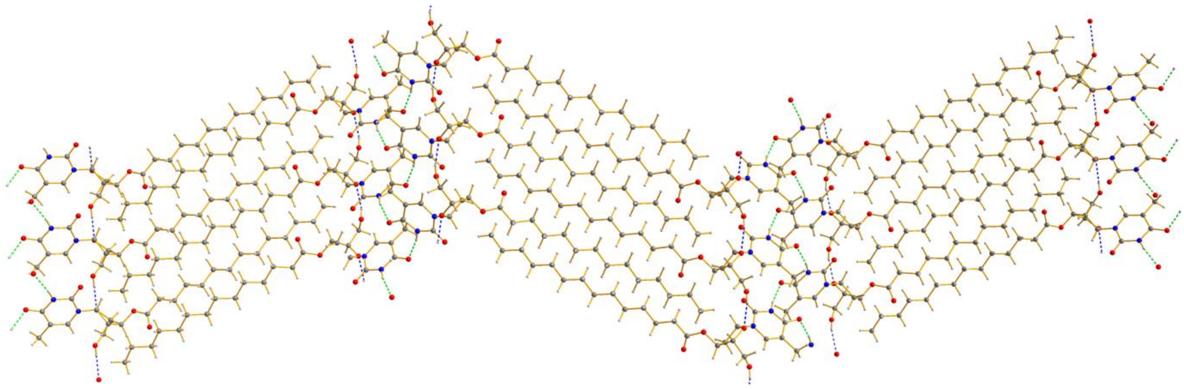


Figure 11. X-ray crystal packing structure of **3b** showing a view of the fully interdigitated alkyl chains and a detailed view of the H-bonding interactions along the crystallographic b-axis. Intermolecular H-bonding interactions between the N3-H and C4-O atoms of thymine bases are shown in green dashed lines. H-bonding interactions between the O4′ and O5′-H atoms of sugar residues are shown in blue dashed lines. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

4A.3.5.2 Powder X-ray diffraction (PXRD) analysis

We resorted to PXRD analysis to further understand the molecular packing of the monosubstituted nucleolipids in the gel state. For this purpose, **3c** was chosen as the test system. The PXRD spectrum of the xerogel of **3c** obtained from a mixture of DMSO-water (80:20) gave a prominent peak corresponding to an interplanar distance of 3.17 nm (001 plane, Figure 12). This distance is shorter than the distance of two interdigitated molecules (3.47 nm) and longer than one molecule length (2.85 nm) as obtained from single-crystal structure (Figure 13). These measurements suggest that the basic unit of the xerogel of **3c** is a bilayer structure composed of interdigitated nucleolipid molecules (Figure 14).[48,49]

Furthermore, shorter interplanar distance exhibited by the gel indicates that the extent of interdigitation is more in the gel state as compared to in the crystal structure. This is likely due to the added water, which augments the hydrophobic effect between the fatty acid acyl chains leading to effective packing. The PXRD data also revealed diffraction peaks corresponding to layer spacings of 3.17, 1.59, 0.79, 0.63, 0.53, 0.45, and 0.39 nm, which were in the ratio of 1:1/2:1/4:1/5:1/6:1/7:1/8, respectively (Figure 12). This ratio of *d*-spacing clearly substantiates the existence of an ordered lamellar structure originating from multiple H-bonding interactions between the nucleosides of the basic bilayer unit (Figure 14).[50–52] The PXRD pattern and *d*-spacing of the xerogel of **3c** obtained from methanol-water and toluene-water mixtures were found to be similar (Figure 15). Furthermore, a PXRD spectrum of xerogel and a spectrum simulated using the single-crystal data matched reasonably well, suggesting that the packing pattern of **3c** in the gel state and crystal structure could be similar (Figure 16).

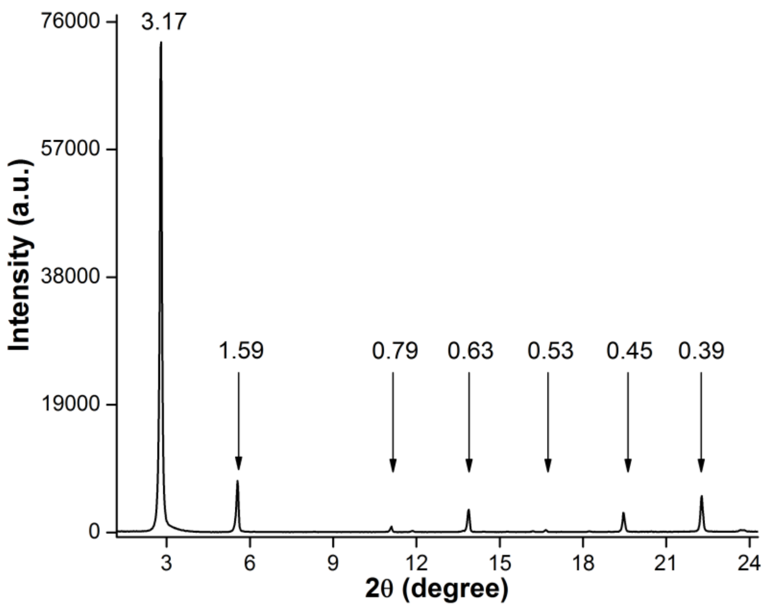


Figure 12. PXRD spectrum of xerogel of **3c** obtained from a mixture of DMSO-water (80:20). The layer spacing (nm) for prominent diffraction peaks is also given.

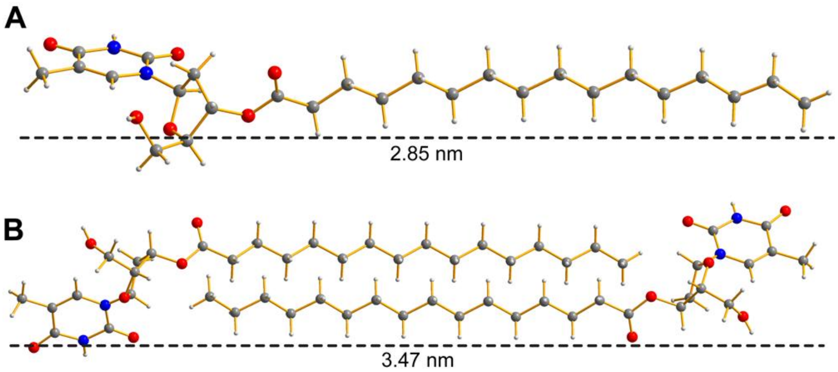


Figure 13. X-ray crystal structure of **3c** (**A**) single molecule and (**B**) fully interdigitated molecules as viewed using Diamond 3.0. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen. The length was measured as shown above (dashed line represent the length of the single molecule and interdigitated molecules in nm).

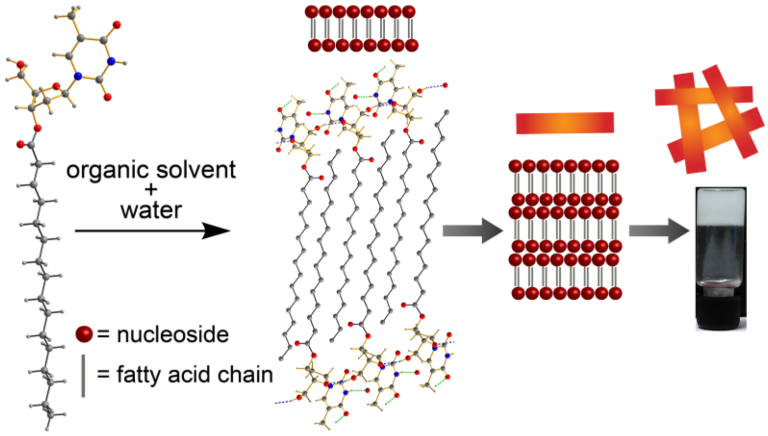


Figure 14. Pictorial representation of the possible mechanism for the hierarchical self-assembly of the monosubstituted nucleolipids (e.g., **3c**) leading to the formation of gel. The basic bilayer unit formed by the interdigitation of fatty acid acyl chains of nucleolipids serves as the precursor for the ordered lamellar structure through multiple H-bonds between nucleosides. These interactions eventually lead to the formation of entangled sheets resulting in the immobilization of the solvent.

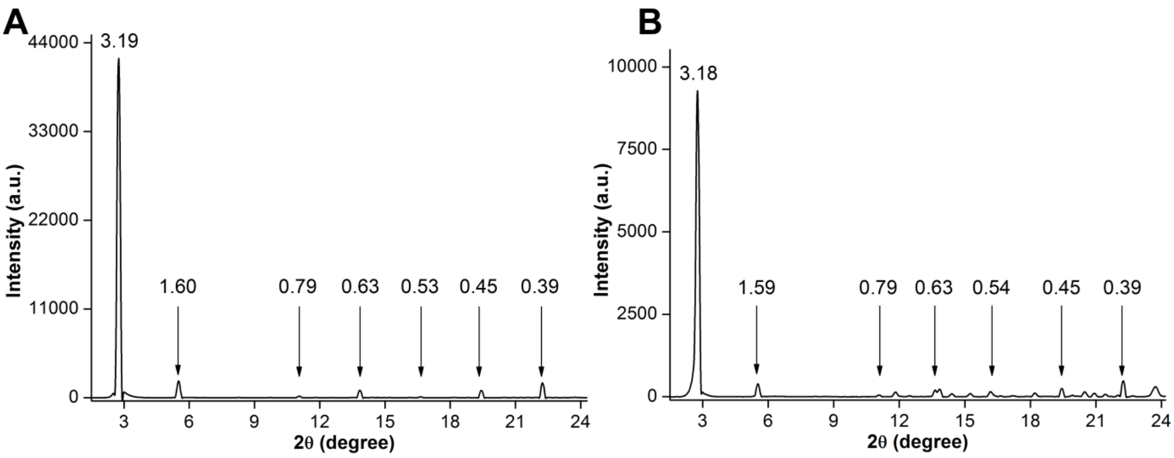


Figure 15. PXRD spectra of xerogel of **3c** obtained from a mixture of (**A**) methanol-water (80:20) and (**B**) toluene-water (80:20). Layer spacing (nm) for prominent diffraction peaks is given.

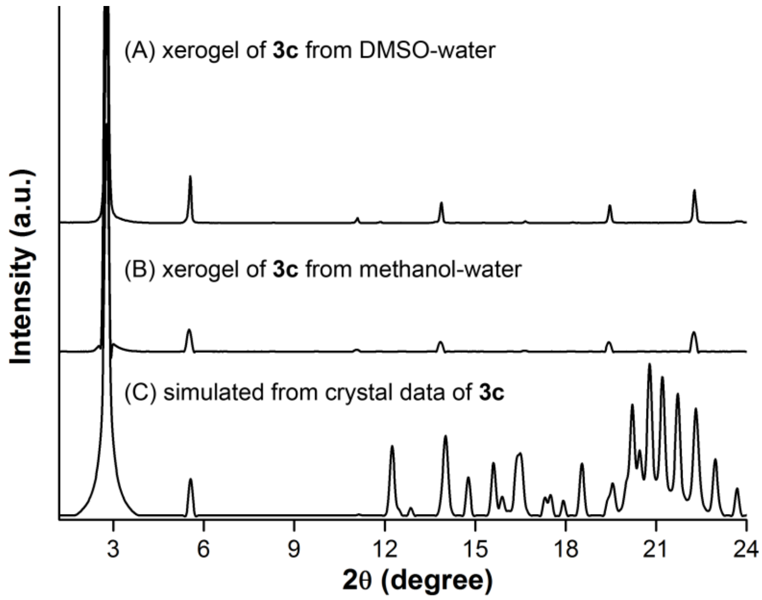


Figure 16. Comparison of PXRD spectra of xerogels of **3c** obtained from (**A**) DMSO-water and (**B**) methanol-water mixtures with that of the (**C**) PXRD pattern simulated from X-ray crystal data of **3c**.

4A.3.5.3 Variable temperature ^{1}H NMR experiments

In order to ascertain the H-bonding interaction partners in the gel state, we first performed variable temperature ^{1}H NMR experiment using nucleolipid **3c** at its CGC. As the temperature of the partial gel of **3c**, prepared in d_6-DMSO-water mixture (95:5), was increased from 25 °C to 55 °C, the N3-H and 5′-OH proton signals gradually shifted upfield ($\Delta\delta$ = 0.23 and 0.21 ppm, respectively, Figure 17). Partial gel was used in this experiment as it gave detectable NMR signals for N3-H and 5′-OH atoms. This shift in the proton signal as a function of increasing temperature is due to the breaking of H-bonding interactions during gel to sol transition. Apart for these hydrogens, chemical shift of other hydrogen atoms were found to be unaltered in both the states. Similar results were obtained for nucleolipid **3b** containing a shorter myristoyl chain (Figure 18). These observations suggest that the H-bonded atoms involved in the crystal packing and supramolecular gels are same.

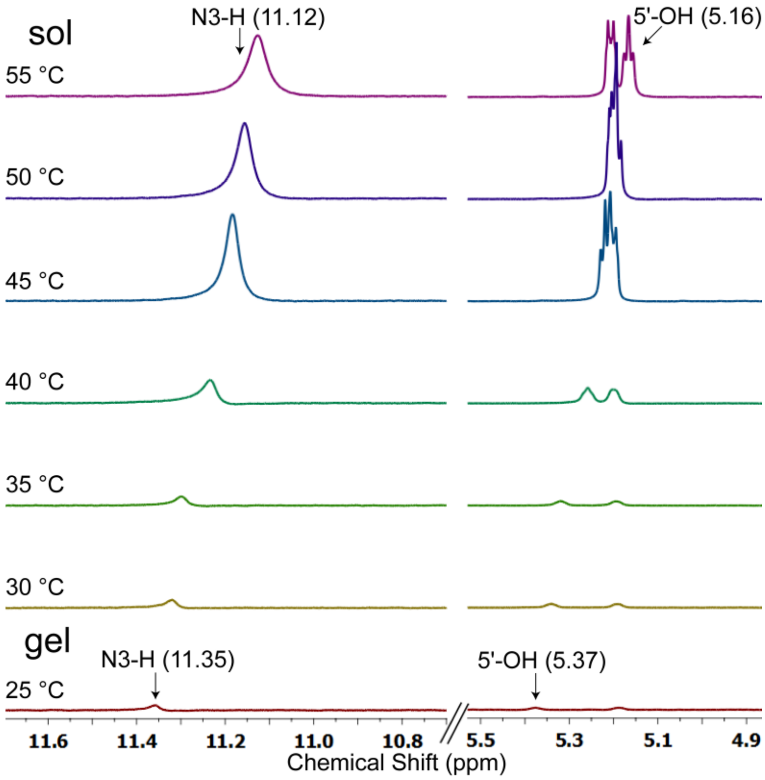


Figure 17. [1]H NMR spectra of **3c** gel (partial gel, d_6-DMSO-water = 95:5) at its CGC as a function of increasing temperature. N3-H and 5′-OH atoms exhibited discernible upfield shift in their proton signals during gel to sol transition.

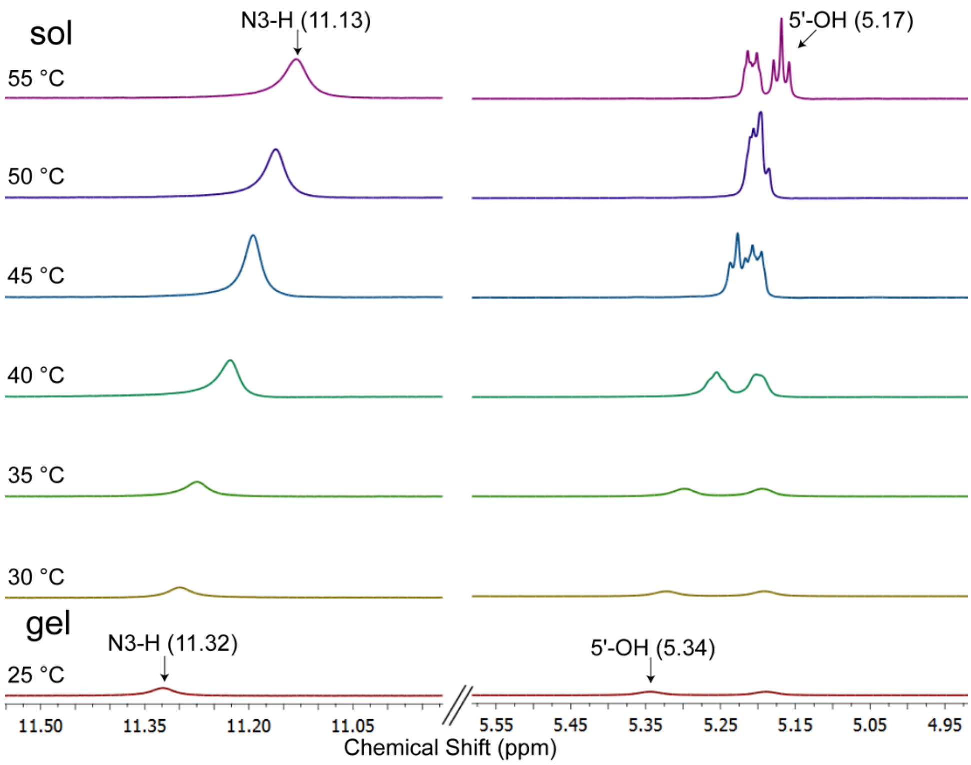


Figure 18. [1]H NMR spectra of **3b** gel (partial gel, d_6-DMSO-water = 95:5) at its CGC as a function of increasing temperature. N3-H of nucleobase and 5′-OH of sugar exhibited discernible upfield shift in their proton signals during gel to sol transition.

The role of water in the gelation of nucleolipids was investigated by performing ^1H NMR analysis of **3c** in d_6-DMSO containing different amounts of D_2O. Typically, HDO present in trace amounts as an impurity in D2O shows a peak at ~3.3 ppm.[39] The proton NMR of the nucleolipid in neat d_6-DMSO showed a peak corresponding to HDO at 3.32 ppm. As the percentage of D_2O was increased from 5 to 20, transformation from a sol to partial gel to stable gel was observed much like the one discussed above (Figure 2B). Interestingly, during this sol-gel transformation process, the signal associated with HDO also shifted downfield from 3.32 to 3.87 ppm (Figure 19). This deshielding effect of proton signal in the gel state points out that the hydrogen atoms of water could be engaged in H-bonding interaction with nucleoside H-bonding sites, thereby facilitating the self-assembling process. Taken together, combining the information obtained from crystal structure, PXRD, and ^1H NMR analyses, it can be inferred that the tail-to-tail interaction (interdigitation) between the hydrocarbon chains and head-to-head interaction between the nucleosides assisted by multiple H-bonding partners including water could have set the path for the hierarchical self-assembly leading to the gel formation (Figure 14).

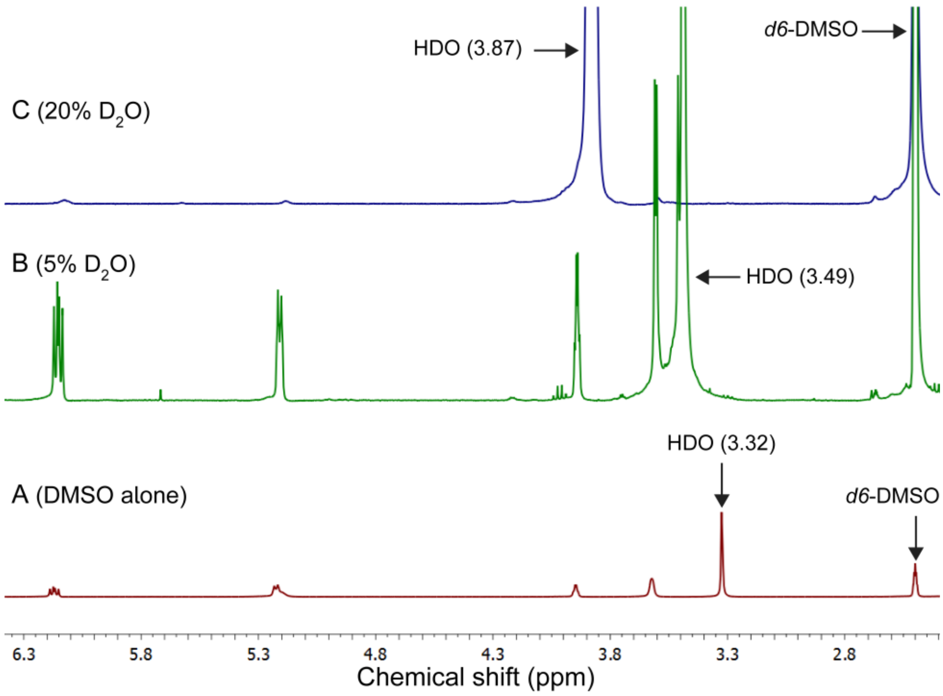


Figure 19. (**A**) ^1H NMR spectrum of nucleolipid **3c** in d_6-DMSO (sol state). (**B**) ^1H NMR spectrum of a partial gel of **3c** in d_6-DMSO-water mixture (95:5). (**C**) ^1H NMR spectrum of a stable gel of **3c** in d_6-DMSO-water (80:20). During sol-gel transformation process, the signal associated with HDO shifted downfield from 3.32 ppm to 3.87 ppm. This observation suggests that apart from enhancing the hydrophobic effect and van der Waals interaction between the fatty acid chains, water could be also involved in H-bonding interaction with thymidine nucleolipids, thereby setting up the pathway for gelation process.[39]

4A.3.6 Surface character of films made of water-induced nucleolipid gels

It is well documented that surface texture and chemistry can be manipulated to generate surfaces, which are highly (super)hydrophobic.[53,54] Though not straightforward, material composites with such surface properties have been developed by coating surfaces with highly hydrophobic organosilanes, perfluoroalkanes, carboranes, and, recently, metal/covalent organic frameworks.[55–60] Some of these systems have been used in designing self-cleaning and corrosion-resistant materials and in biomedical research.[61–66] As the self-organization in monosubstituted thymine nucleolipid gels (e.g., **3c**) is driven by tail-to-tail and head-to-head interactions between fatty acid acyl chains and nucleosides, respectively, we hypothesized that the surface hydrophobicity/hydrophilicity of the film made of thymidine nucleolipid xerogels could be tuned by using an appropriate organic solvent-water mixture during the gelation process.[39–41] For example, a film made of a xerogel of **3c** in a toluene-water mixture could produce a surface ending with fatty acid acyl chains due to the favored hydrophobic-hydrophobic interaction between the hydrocarbon chains and solvent, toluene (Figure 20, inset). Similarly, a film made of a xerogel of **3c** in a polar organic solvent-water mixture (e.g., methanol-water) could produce a surface in which the nucleosides could be projected outside in the polar solvent. In the former scenario, the surface could be hydrophobic, and in the latter, it could be hydrophilic.[41,67]

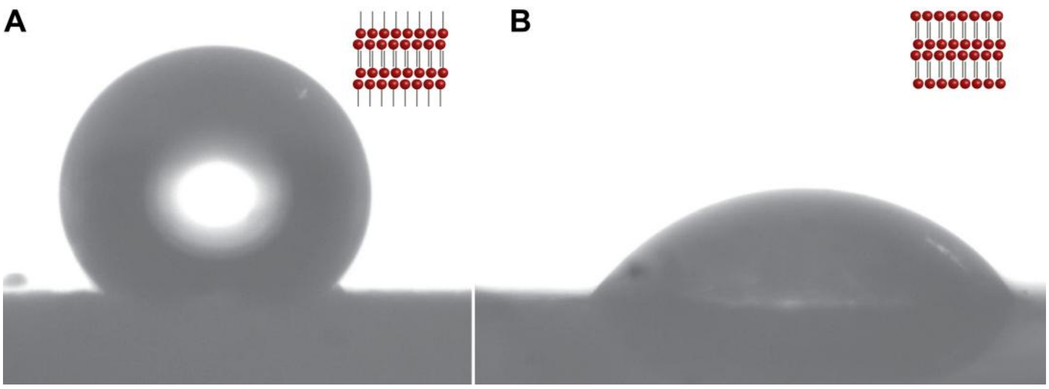


Figure 20. Images showing the contact angle of water droplets on the surface of xerogel films of **3c** fabricated from (**A**) toluene-water mixture and (**B**) methanol-water mixture. The possible orientation of the fatty acid tail and nucleoside head group as predicted from the contact angle measurements in shown in the respective inset.

In order to test our hypothesis, we made films of xerogel of **3c** gelated in toluene-water or methanol-water mixture (organic solvent:water = 80:20) and measured the contact angle (θ_{CA}) of sessile water drops on the surface. In general, a surface is regarded as hydrophobic and hydrophilic if θ_{CA} is > 90° and < 90°, respectively.[53,54] The θ_{CA} of water droplets on the surface of the xerogel film fabricated from toluene-water mixture was found to be 151 ± 3°, which indicates that the surface of the film is highly hydrophobic (Figure 20). Interestingly, water droplet on this surface rolled off upon slightly tilting the surface, similar to the Cassie state, which exhibits low adhesion.[68] A tilt angle as low as 4° resulted in the rapid rolling off of the water droplet, which further indicates that the surface is in the realm of superhydrophobicity.[54] On the other hand, xerogel film fabricated from methanol-water mixture gave a θ_{CA} of 62 ± 2° indicating that the surface is hydrophilic. The θ_{CA} was also measured for xerogel films of **3c** made from water-miscible organic solvent mixtures such as DMSO-water and dioxane-water. The θ_{CA} was found to be 105 ± 3° and 74 ± 4°, respectively, which further indicates that the choice of solvent mixture in the gelation process is important in tuning the surface property.

Next, we wanted to evaluate if the surface of the gel could be switched from highly hydrophobic to hydrophilic and vice versa. Highly hydrophobic xerogel film fabricated from toluene-water mixture was recovered and redispersed in methanol-water mixture, which upon drop-casting and drying was subjected to θ_{CA} measurements. The surface of the xerogel film thus obtained was found be hydrophilic (θ_{CA} = 61 ± 2°). Similarly, film fabricated from methanol-water mixture was recovered and redispersed in toluene-water mixture, which was then used in forming the xerogel film. The θ_{CA} of sessile water drops on the film was found to be 152 ± 2°, which indicated that the surface property was switched from hydrophilic to highly hydrophobic. Collectively, these results demonstrate that water-induced gelation of nucleoside-fatty acid hybrids in an appropriate organic solvent could provide an alternative method for developing functional gels, in this case, gel surface with switchable wettability property.

4A.3.7 Metal ion responsiveness of 3′,5′-*O*-disubstituted nucleolipid gels

Apart from their biological and therapeutic roles, nucleoside-metal ion interactions have been utilized in designing sensors and nucleic acid-based molecular machines.[69-71] In particular, the thymine-Hg^{2+}-thymine non-canonical base pair formed by the specific interaction of Hg^{2+} ion with N3 atom of thymine residues has received significant attention.[72,73] Since N3-H serves as an important H-bonding site for the supramolecular gelation process of thymine nucleolipids, we became interested in studying their gelation behaviour in response to added metal ions. For this purpose, increasing amount of a small series of metal ions Hg^{2+}, Cd^{2+}, Zn^{2+}, Ag^{+} and Pb^{2+}, which are known to coordinate to nucleosides, was added to a hot solution of **4c** (distearoyl-attached thymidine) in DMSO at its CGC. The samples were left to come to room temperature and the gelling ability was studied by inverted vial method. Addition of as low as 0.02 equiv of Hg^{2+} ion (0.29 mM) did not the support the gelation process (Figure 21). However, even 5 equiv of other metal ions listed above did not affect the gelation ability of **4c**. These results indicate that the self-assembling process of **4c** is selectively responsive to Hg^{2+} ions among the metal ions tested. Further, addition of aqueous stock solutions of the metal ions to the monosubstituted nucleolipid **3c** did not support water-induced gelation.

Figure 21. Photograph of 3′,5′-*O*-distearoyl-substituted nucleolipid gel **4c** at its CGC in the presence of various metal ions.

In order to understand the mechanism by which Hg^{2+} ions disrupt the gelation of **4c**, SEM, NMR, and mass analyses were performed. The long-range twisted and entangled tapes of **4c**, which were responsible for immobilizing the solvents, disintegrated into smaller fragments upon addition of Hg^{2+} ions (Figure 22). The ^{1}H NMR spectrum of **4c** in the presence of Hg^{2+} ions revealed the disappearance of the N3-H signal of the thymine base. It is likely that the Hg^{2+} ion could have displaced the N3-H atom and formed a **4c**-Hg-**4c** base pair by coordinating with N3 atoms.[72] MALDI-TOF mass analysis also supported the notion of formation of the dimer complex (Figure 23). All these results point out that the dimer formation in the presence of Hg^{2+} ions would have disrupted the H-bonding interactions involved in the gelation process, thereby resulting in the disassembly of the nucleolipid gel.

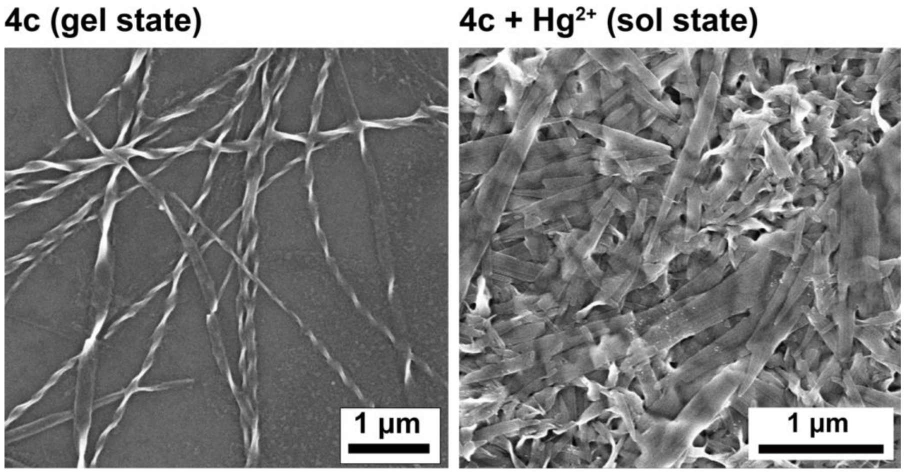


Figure 22. FESEM image of nucleolipid **4c** gel (**A**) in the absence and (**B**) in the presence of Hg^{+2} ions.

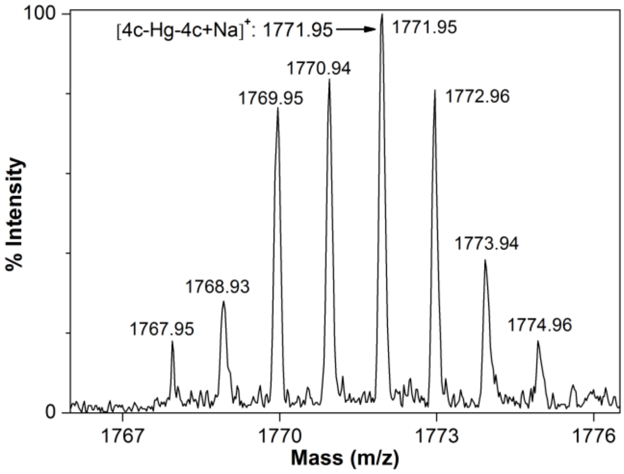


Figure 23. Mass spectrum showing the molecular ion peak of **4c**-Hg-**4c** base pair complex. Characteristic isotopic pattern supports the presence of Hg^{2+} ion.

4A.4 Conclusions

We have developed nucleoside-fatty acid hybrids, which serve as efficient supramolecular synthons for gelling solvents, by attaching different fatty acid chains to the 3′-*O* or 3′,5′-*O* positions of thymidine. Depending on the site, nature (saturated/unsaturated), and number of fatty acid chains attached to the nucleoside, the nucleolipids formed typical organogels in pure organic solvents or water-induced organogels with distinct morphologies. Thorough investigation of the morphology and structure of the nucleolipids and their gels provided insight into the mechanism of the hierarchical supramolecular gelation process, which was found to be facilitated by a coordinated interplay of H-bonding interactions and hydrophobic effect. Importantly, the surface of the film fabricated using monofatty acid-substituted thymidine nucleolipid gels could be tuned between highly hydrophobic and hydrophilic surface depending on the nature of the organic solvent-water mixture used in the gelation process. On the other hand, the gelation ability of difatty acid-substituted thymidine nucleolipids could be selectively modulated by adding trace amounts of Hg^{2+} ions as it formed a T-Hg-T base pair by coordinating with the N3 atom of thymidine. Taken together, our results demonstrate that supramolecular nucleolipid gels with surface-tuned and metal-ion-sensing properties can be fabricated by design. This approach could be extended to other purine and pyrimidine nucleoside derivatives, thereby providing new opportunities to construct novel nucleolipid-based function materials.

4A.5 Experimental section

4A.5.1 Materials

Thymidine, 4,4′-dimethoxytrityl chloride (DMT-Cl), dodecanoic acid, palmatic acid, stearic acid, oleic acid, mercury(II) perchlorate hexahydrate, dry pyridine were purchased from Sigma-Aldrich. EDC (1-(3-dimethyl aminopropyl)-3-ethyl carbodiimide hydrochloride) and 4-dimethylaminopyridine were obtained from Avra Synthesis. Myristic acid was procured from Fluka. Silicon wafers (N-type without dopant) were purchased from Sigma-Aldrich. 5′-*O*-DMT-protected thymidine **2** was prepared by following a reported procedures.[74]

4A.5.2 Instrumentation

NMR spectra were recorded on 400 MHz Jeol ECS-400 and Bruker 500 MHz spectrometers. The morphology of gels was analyzed using Zeiss Ultra Plus field-emission scanning electron microscope (FESEM). Powder X-ray diffraction (PXRD) spectra were obtained at room temperature using Bruker D8 Advance diffractometer (Cu Kα radiation, λ = 1.5406 Å). Contact angle measurements were performed on the Contact Angle Instrument model (HO-IAD-CAM-01). Rheology measurements were carried out in Anton paar MCR 302 instrument. Single crystal X-ray data for structure determination were collected from Bruker APEX II DUO diffractometer using MoKα (λ= 0.71073 Å) and Cu Kα (λ = 1.54178 Å) graphite monochromated radiation. Mass measurements were performed using Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer and Water Synapt G2 High Definition mass spectrometers.

4A.5.3 Characterization data for compounds 3a-3e and 4a-4c

4A.5.3.1 General procedure for the synthesis of mono-fatty acid-substituted nucleolipids (3a–3e): 5′-*O*-DMT-protected thymidine **2** (1.0 equiv), fatty acid (dodecanoic acid, myristic acid, palmitic acid, stearic acid, oleic acid, 1.2 equiv), EDC (1.2 equiv) and DMAP (0.3 equiv) were dissolved in anhydrous dichloromethane (10 mL/g of **2**). The reaction mixture was stirred for 12 h at room temperature under nitrogen atmosphere. After completion of the reaction, the crude product was diluted with dichloromethane and extracted with saturated NH_4Cl solution and dried over sodium sulfate. The residue was treated with 3% trichloroacetic acid (TCA) in dichloromethane for 30 min at RT to remove the DMT group. The volatile solvents were evaporated and the crude residue was purified by silica gel column chromatography (40% ethyl acetate in hexane) to afford the products.

Compound 3a: Off white solid, 68 % yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5 v/v); R_f = 0.36; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 7.51 (br, 1H), 6.25 (dd, J_1 = 8.2 Hz, J_2 = 6.2 Hz, 1H), 5.36–5.33 (m, 1H), 4.08–4.06 (m, 1H), 3.96–3.88 (m, 2H), 2.45–2.37 (m, 2H), 2.33 (t, J = 7.6 Hz, 2H), 1.92 (br, 3H), 1.65–1.58 (m, 2H), 1.29–1.25 (m, 16 H), 0.87 (t, J = 6.8 Hz, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.8, 163.8, 150.6, 136.4, 111.5, 86.1, 85.3, 74.6, 62.7, 37.3, 34.3, 32.0, 29.7, 29.6, 29.5, 29.4, 29.2, 24.9, 22.8, 14.3, 12.7; HRMS: (m/z) Calculated for $C_{22}H_{37}N_2O_6$ $[M+H]^+$ = 425.2652, found = 425.2657.

Compound 3b: Off white solid, 76 % yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5 v/v); R_f = 0.40; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 9.02 (br, 1H), 7.51 (d, J = 1.2 Hz, 1H), 6.25 (dd, J_1 = 8.0 Hz, J_2 = 6.4 Hz, 1H), 5.36–5.33 (m, 1H), 4.07–4.05 (m, 1H), 3.96–3.87 (m, 2H), 2.45–2.36 (m, 2H), 2.33 (t, J = 7.4 Hz, 2H), 1.91 (d, J = 0.8 Hz, 3H), 1.65–1.58 (m, 2H), 1.29–1.25 (m, 20H), 0.87 (t, J = 6.8 Hz, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.8, 163.8, 150.6, 136.4, 111.5, 86.1, 85.3, 74.5, 62.7, 37.4, 34.3, 32.1, 29.8, 29.8, 29.7, 29.6, 29.5, 29.4, 29.2, 25.0, 22.8, 14.2, 12.7; HRMS: (m/z) Calculated for $C_{24}H_{40}N_2O_6Na$ $[M+Na]^+$ = 475.2784, found = 475.2785.

Compound 3c: Off white solid, 79 % yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5 v/v); R_f = 0.41; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.84 (br, 1H), 7.51 (d, J = 1.2 Hz, 1H), 6.25 (dd, J_1 = 8.0 Hz, J_2 = 6.0 Hz, 1H), 5.36–5.33 (m, 1H), 4.07–4.05 (m, 1H), 3.96–3.88 (m, 2H), 2.45–2.37 (m, 2H), 2.33 (t, J = 7.6 Hz, 2H), 1.91 (d, J = 0.8 Hz, 3H), 1.65–1.58 (m, 2H), 1.29–1.25 (m, 24 H), 0.87 (t, J = 7.0 Hz, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.8, 163.7, 150.5, 136.4, 111.5, 86.1, 85.3, 74.5, 62.7, 37.3, 34.3, 32.1, 29.8, 29.7, 29.7, 29.6, 29.5, 29.4, 29.3, 25.0, 22.8, 14.3, 12.7; HRMS: (m/z) Calculated for $C_{26}H_{44}N_2O_6Na$ $[M+Na]^+$ = 503.3097, found = 503.3098.

Compound 3d: Off white solid, 73 % yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5 v/v); R_f = 0.51; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.78 (s, 1H), 7.50 (d, J = 0.8 Hz, 1H), 6.24 (dd, J_1 = 8.0 Hz, J_2 = 6.4 Hz, 1H), 5.36–5.33 (m, 1H), 4.07–4.06 (m, 1H), 3.96–3.89 (m, 2H), 2.46–2.37 (m, 2H), 2.33 (t, J = 7.6 Hz, 2H), 1.92 (d, J = 0.8 Hz, 3H), 1.66–1.58 (m, 2H), 1.29–1.25 (m, 28 H), 0.87 (t, J = 6.8 Hz, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.8, 163.7, 150.5, 136.4, 111.5, 86.1, 85.3, 74.6, 62.8, 37.4, 34.3, 32.1, 29.8, 29.8, 29.7, 29.6, 29.5, 29.4, 29.3, 25.0, 22.8, 14.3, 12.8; HRMS: (m/z) Calculated for $C_{28}H_{48}N_2O_6Na$ $[M+Na]^+$ = 531.3410, found = 531.3405.

Compound 3e: Viscous oil, 78 % yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5 v/v); R_f = 0.52; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.98 (br, 1H), 7.52 (s, 1H), 6.25–6.22 (m, 1H), 5.38–5.30 (m, 3H), 4.09–4.05 (m, 1H), 3.96–3.89 (m, 2H), 2.83–2.75 (m, 2H), 2.46–2.38 (m, 2H), 2.34 (t, J = 7.6 Hz, 2H), 2.03–1.98 (m, 2H), 1.92 (s, 3H), 1.66–1.57 (m, 2H), 1.30–1.21 (m, 20 H), 0.87 (t, J = 6.8 Hz, 3H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.8, 164.0, 150.5, 136.7, 130.2, 129.8, 111.6, 86.3, 85.3, 74.5, 62.7, 37.3, 34.3, 32.0, 29.9, 29.8, 29.7, 29.5, 29.3, 29.3, 29.2, 27.4, 27.3, 24.9, 22.8, 14.3, 12.7; HRMS: (m/z) Calculated for $C_{28}H_{46}N_2O_6Na$ $[M+Na]^+$ = 529.3254, found = 529.3253.

4A.5.3.2 General procedure for synthesis of di-fatty acid-substituted nucleolipids (4a–4c)

Thymidine (1.0 equiv), fatty acid (myristic acid, palmitic acid, stearic acid, 2.1 equiv), EDC (2.1 equiv) and DMAP (0.25 equiv) were dissolved in anhydrous dichloromethane. The reaction mixture was stirred for 12 h at RT under nitrogen atmosphere. The reaction mixture was diluted with dichloromethane and extracted with saturated NH_4Cl solution. The organic extract was dried over sodium sulphate, filtered and the filtrate was evaporated. The residue was purified by silica gel column chromatography (40% ethyl acetate in hexane) to afford the products.

Compound 4a: Off white solid, 79 % yield. TLC (hexane:EtOAc = 70:30 v/v); R_f = 0.42; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.96–8.69 (br, 1H), 7.29 (s, 1H), 6.31 (dd, J_1 = 8.4 Hz, J_2 = 5.6 Hz, 1H), 5.20 (d, J = 6.4 Hz, 1H), 4.40 (dd, J_1 = 12.2 Hz, J_2 = 4.2 Hz, 1H), 4.31 (dd, J_1 = 12.4 Hz, J_2 = 2.8 Hz, 1H), 4.23 (br, 1H), 2.46 (dd, J_1= 14.0 Hz, J_2 = 5.2 Hz, 1H), 2.34 (t, J = 7.4 Hz, 4H), 2.16–2.09 (m, 1H), 1.93 (s, 3H), 1.67–1.60 (m, 4H), 1.29–1.25 (m, 40H), 0.87(t, J = 6.6 Hz, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.4, 173.2, 163.4, 150.2, 134.8, 111.8, 85.0, 82.5, 74.1, 63.8, 37.9, 34.3, 34.2, 32.1, 29.8, 29.8, 29.7, 29.6, 29.5, 29.4, 29.4, 29.3, 29.2, 25.0, 24.9, 24.9, 22.8, 14.3, 12.5; HRMS: (m/z) Calculated for $C_{38}H_{67}N_2O_7$ $[M+H]^+$ = 663.4948, found = 663.4942.

Compound 4b: Off white solid, 87 % yield. TLC (hexane:EtOAc = 70:30 v/v); R_f = 0.43; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 9.04–8.81 (br, 1H), 7.29 (s, 1H), 6.31 (dd, J_1 = 8.4 Hz, J_2 = 5.6 Hz, 1H), 5.20 (d, J = 6.4 Hz, 1H), 4.41 (dd, J_1 = 12.2 Hz, J_2 = 4.2 Hz, 1H), 4.31 (dd, J_1 = 12.0 Hz, J_2 = 2.8 Hz, 1H), 4.23 (br, 1H), 2.46 (dd, J_1 = 14.0 Hz, J_2 = 5.2 Hz, 1H), 2.34 (t, J = 7.4 Hz, 4H), 2.16–2.10 (m, 1H), 1.93 (s, 3H), 1.65–1.60 (m, 4H), 1.29–1.25 (m, 48H), 0.87(t, J = 6.6 Hz, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.4, 173.2, 163.7, 150.3, 134.7, 111.6, 85.0, 82.4, 74.0, 63.8, 37.8, 34.3, 34.2, 32.1, 29.8, 29.8, 29.8, 29.7, 29.6, 29.5, 29.4, 29.4, 29.3, 29.2, 25.0, 24.9, 22.8, 14.3, 12.8; HRMS: (m/z) Calculated for $C_{42}H_{75}N_2O_7$ $[M+H]^+$ = 719.5574, found = 719.5574.

Compound 4c: Colourless solid, 90 % yield. TLC (hexane:EtOAc = 70:30 v/v); R_f = 0.44; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.88–8.54 (br, 1H), 7.29 (s, 1H), 6.31 (dd, J_1 = 8.4 Hz, J_2 = 5.6 Hz, 1H), 5.20 (d, J = 6.4 Hz, 1H), 4.40 (dd, J_1 = 12.2 Hz, J_2 = 4.2 Hz, 1H), 4.31 (dd, J_1 = 12.0 Hz, J_2 = 2.8 Hz, 1H), 4.23 (br, 1H), 2.46 (dd, J_1 = 14.2 Hz, J_2 = 5.4 Hz, 1H), 2.34 (t, J = 7.4 Hz, 4H), 2.16–2.09 (m, 1H), 1.93 (s, 3H), 1.67–1.59 (m, 4H), 1.29–1.25 (m, 56H), 0.87(t, J = 6.6 Hz, 6H). ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.4, 173.2, 163.7, 150.3, 134.7, 111.5, 85.0, 82.5, 74.1, 63.8, 37.8, 34.3, 34.2, 32.1, 29.8, 29.8, 29.8, 29.8, 29.7, 29.6, 29.5, 29.4, 29.4, 29.3, 25.0, 24.9, 22.8, 14.3, 12.8; HRMS: (m/z) Calculated for $C_{46}H_{83}N_2O_7$ $[M+H]^+$ = 775.6200, found = 775.6193.

4A.5.4 Gelation test by the inverted vial method

4A.5.4.1 Water-induced gelation of nucleolipids (3a–3e): A weighed amount of the mono-fatty acid-substituted nucleolipids in different sample vials was dissolved in an organic solvent by heating. Different amounts of water (v %) was added to each vial and was left to form gels. The formation of a stable gel was monitored by inverted vial method. Except in dioxane and heptane, in other solvents tested, the nucleolipids formed stable gels within a min (Table 1). In dioxane and heptane, the gelation process was slow and it took nearly ~3 h to for a stable gel. The CGC and amount of water required to promote gelation was determined by preparing samples containing different amounts of the nucleolipid and organic solvent-water ratios. Repeated heating and cooling steps confirmed the thermo-reversibility of the gelation process. All experiments were performed at least in duplicate.

4A.5.4.2 Organogels of nucleolipids (4a–4c): A weighed amount of the di-fatty acid-substituted nucleolipids was dissolved in various organic solvents by heating. The samples were allowed to come to RT and the formation of gels was confirmed by inverted vial method. In DMSO and acetonitrile, **4a–4c** formed stable gels within few min. **4a–4c** did not for a stable gel in other solvents tested in this study. Repeated heating and cooling steps confirmed the thermo-reversibility of the gelation process. All experiments were performed at least in duplicate.

4A.5.5 FESEM analysis: The morphology of the nucleolipid gels was analyzed by FESEM. Diluted gel samples in respective solvents were drop-casted on silicon wafers and dried in a vacuum desiccator for ~12 h. A thin layer of Au was deposited on the samples to minimize sample charging before SEM analysis. The SEM images were analyzed by using ImageJ 1.46r software to determine the dimensions of the sheets and twisted tapes.

4A.5.6 Crystallography: Single crystals of **3a** and **3b** were obtained from methylene chloride and dimethyl formamide, respectively, by vapour diffusion method. Crystals of **3c** were obtained from isopropanol by slow evaporation method. The diffraction data of **3a**, **3b** and **3c** were collected using Bruker KAPPA APEX II CCD Duo diffractometer (operated at 1500 W, 50 kV and 30 mA) equipped with a graphite-monochromated MoKα (λ = 0.71073 Å) or Cu Kα (λ = 1.54178 Å) radiation source. The structure of nucleolipids was refined on F^2 by full-matrix least-squares technique using SHELX program.[75] All non-hydrogen atoms were refined anisotropically unless otherwise mentioned. Hydrogen atoms were constrained in geometric positions with respect to their parent atoms. Crystallographic data for the nucleolipids are listed in appendix-IVA. The X-ray crystallographic coordinates for the structures reported in this article have been deposited at the Cambridge Crystallographic Data

Centre (CCDC), under deposition number CCDC: 1479220 (**3a**), 1479224 (**3b**) and 1479225 (**3c**). These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

4A.5.7 Powder X-ray diffraction (PXRD) analysis of nucleolipid 3c: Xerogels of **3c** were formed by drop-casting a dispersion of the nucleolipid in different organic solvent-water mixtures (DMSO:water, 80:20; methonal:water, 80:20; toluene:water, 80:20) at its respective CGC. The samples were allowed to come to RT and were placed in a vacuum desiccator and dried under vacuum for nearly 15 h to obtain the xerogels. PXRD data was collected using Bruker D8 Advance diffractometer with CuKα source (1.5406 Å). Diffraction data were collected at 2θ angle from 1° to 30° using a 0.01° step size and 0.5 s per step. Low angle diffraction data was collected by keeping the motorized divergence slit in automatic mode so as to maintain the X-ray beam footprint on the sample to 12 x 12 mm. Further, the position sensitive detector (Lynxeye) channels were reduced to minimize the background X-ray scattering entering the detector.

4A.5.8 Variable temperature ^{1}H NMR experiments: Partial gels of **3b** (1.8 w/v %) and **3c** (1.4 w/v %) in *d6*-DMSO:water (95:5) were formed in individual NMR tubes by heating and cooling process. ^{1}H NMR was recorded on a 500 MHz spectrometer as a function of increasing temperature. The temperature of the sample was increased from 25 °C to 55 °C with an increment of 5 °C and equilibration time of 10 min. The spectrum was recorded at every 5 °C interval.

4A.5.9 Contact Angle (θ_{CA}) measurements: Contact angle measurements were performed using the sessile drop method on Contact Angle Instrument model (HO-IAD-CAM-01). A hot dispersion of nucleolipid **3c** in an organic solvent-water mixture (80:20, v %) at its CGC was drop-casted on a glass slide and allowed to form gel by cooling. The glass slide was dried under vacuum in a desiccator for ~6 h to form a xerogel film. Sessile water droplets were placed on the xerogel film at different places and the static contact angle was determined by processing the images obtained using the CCD camera and software provided with the instrument. θ_{CA} reported in this study is an average of at least ten different measurements.

4A.5.10 Slide angle measurements: The angle at which the water droplets rolled off the surface of the xerogel film of **3c** was determined by placing the film on glass slides of know thickness in an inclined position. A drop of water was dropped on the film and the rolling off of the droplet was monitored using a CCD camera provided with the instrument.

4A.6 References:

1 Guo, P. *Nat. Nanotech*. **2010**, *5*, 833–842.

2 Pinheiro, A.V.; Han, D.; Shih, W. M.; Yan, H. *Nat. Nanotech*. **2011**, *6*, 863–872.

3 Krishnan, Y.; Simmel, F. C. *Angew. Chem., Int. Ed*. **2011**, *50*, 3124–3156.

4 Keller, S.; Marx, A. *Chem. Soc. Rev*. **2011**, *40*, 5690–5697.

5 Bath, J.; Turberfield, A. J. *Nat. Nanotech*. **2007**, *2*, 275–284.

6 Ackermann, D.; Schmidt, T. L.; Hannam, J. S.; Purohit, C.S.; Heckel, A.; Famulok, M. *Nat. Nanotech*. **2010**, *5*, 436–442.

7 Saha, S.; Prakash, V.; Halder, S.; Chakraborty, K.; Krishnan, Y. *Nat. Nanotech*. **2015**, *10*, 645–651.

8 Harrell, C. C.; Kohli, P.; Siwy, Z.; Martin, C. R. *J. Am. Chem. Soc*. **2004**, *126*, 15646–15647.

9 Langecker, M.; Arnaut, V.; Martin, T. G.; List, J.; Renner, S.; Mayer, M.; Dietz, H.; Simmel, F. C. *Science* **2012**, *338*, 932–936.

10 Han, D.; Park, Y.; Kim, H.; Lee, J. B. *Nat. Commun*. **2014**, *5*, 4367.

11 Guo, P.; Haque, F. RNA Nanotechnology and Therapeutics, CRS press, *Taylor and Francis Group*, **2014**.

12 Angell, C.; Xie, S.; Zhang, L.; Chen, Y. *Small* **2016**, *12*, 1117–1132.

13 Sivakova, S.; Rowan, S. J. *Chem. Soc. Rev*. **2005**, *34*, 9–21.

14 Kwak, M.; Herrmann, A. *Chem. Soc. Rev*. **2011**, *40*, 5745–5755.

15 Du, X.; Zhou, J.; Shi, J.; Xu, B. *Chem. Rev*. **2015**, *115*, 13165–13307.

16 Patwa, A.; Gissot, A.; Bestel, I.; Barthélémy, P. *Chem. Soc. Rev*. **2011**, *40*, 5844–5854.

17 Berti, D.; Montis, C.; Baqlioni, P. *Soft Matter* **2011**, *7*, 7150–7158.

18 Allain, V.; Bourgaux, C.; Couvreur, P. *Nucleic Acids Res*. **2012**, *40*, 1891–1903.

19 Peters, G. M.; Davis, J. T. *Chem. Soc. Rev*. **2016**, *45*, 3188–3206.

20 Khiati, S.; Luvino, D.; Oumzil, K.; Chauffert, B.; Camplo, M.; Barthélémy, P. *ACS Nano* **2011**, *5*, 8649–8655.

21 Li, X.; Kuang, Y.; Lin, H.-C.; Gao, Y.; Shi, J.; Xu, B. *Angew. Chem., Int. Ed*. **2011**, *50*, 9365–9369.

22 Peters, G. M.; Skala, L. P.; Plank, T. N.; Hyman, B. J.; Reddy, G. N. M.; Marsh, A.; Brown, S. P.; Davis, J. T. *J. Am. Chem. Soc*. **2014**, *136*, 12596–12599.

23 Wang, D.; Tu, C.; Su, Y.; Zhang, C.; Greiser, U.; Zhu, X.; Yana, D.; Wang, W. *Chem. Sci*. **2015**, *6*, 3775–3787.

24 Li,Y.; Liu, Y.; Ma, R.; Xu, Y.; Zhang, Y.; Li, B.; An, Y.; Shi, L. *ACS Appl. Mater. Interfaces* **2017**, *9*, 13056−13067.

25 Ramin, M. A.; Sindhu, K. R.; Appavoo, A.; Oumzil, K.; Grinstaff, M. W.; Chassande, O.; Barthélémy, P. *Adv. Mater*. **2017**, *29*, 1605227.

26 Jain, D.; Karajic, A.; Murawska, M.; Goudeau, B.; Bichon, S.; Gounel, S.; Mano, N.; Kuhn, A.; Barthélémy, P. *ACS Appl. Mater. Interfaces* **2017**, *9*, 1093–1098.

27 Latxague, L.; Ramin, M. A.; Appavoo, A.; Berto, P.; Maisani, M.; Ehret, C.; Chassande, O.; Barthélémy, P. *Angew. Chem., Int. Ed*. **2015**, *54*, 4517–4521.

28 Yuan, D.; Du, X.; Shi, J.; Zhou, N.; Zhou, J.; Xu, B. *Angew. Chem., Int. Ed*. **2015**, *54*, 5705–5708.

29 Davis, J. T.; Spada, G.P. *Chem. Soc. Rev*. **2007**, *36*, 296–313.

30 Davis, J. T. *Angew. Chem., Int. Ed*. **2004**, *43*, 668–698.

31 Das, R. N.; Kumar, Y. P.; Schütte, O. M.; Steinem, C.; Dash, J. *J. Am. Chem. Soc*. **2015**, *137*, 34–37.

32 Reddy, G. N. M.; Marsh, A.; Davis, J. T.; Masiero, S.; Brown, S. P. *Cryst. Growth Des*. **2015**, *15*, 5945–5954.

33 Deng, M.; Zhang, L.; Jiang, Y.; Liu, M. *Angew. Chem., Int. Ed.* **2016**, *55*, 15062–15066.

34 Berger, O.; Adler-Abramovich, L.; Levy-Sakin, M.; Grunwald, A.; Liebes-Peer, Y.; Bachar, M.; Buzhansky, L.; Mossou, E.; Forsyth, V. T.; Schwartz, T.; Ebenstein, Y.; Frolow, F.; Shimon, L. J. W.; Patolsky, F.; Gazit, E. *Nat. Nanotech.* **2015**, *10*, 353–360.

35 Berger, O.; Yoskovitz, E.; Adler-Abramovich, L.; Gazit, E. *Adv. Mater*. **2016**, *28*, 2195–2200.

36 Nuthanakanti, A.; Srivatsan, S. G. *Nanoscale* **2016**, *8*, 3607–3619.

37 Pal, A.; Dey, J. *Langmuir* **2011**, *27*, 3401–3408.

38 Gao, Y.; Hao, J.; Wu, J.; Zhang, X.; Hu, J.; Ju, Y. *Langmuir* **2016**, *32*, 1685–1692.

39 Lu, J.; Hu, J.; Liu, C.; Gao, H.; Ju, Y. *Soft Matter* **2012**, *8*, 9576–9580.

40 Yan, N.; Xu, Z.; Diehn, K. K.; Raghavan, S. R.; Fang, Y.; Weiss, R. G. *J. Am. Chem. Soc*. **2013**, *135*, 8989–8999.

41 Liu, C.; Jin, Q.; Lv, K.; Zhang, L.; Liu, M. *Chem. Commun*. **2014**, *50*, 3702–3705.

42 Yun, Y. J.; Park, S. M.; Kim, B. H. *Chem. Commun*. **2003**, 254–255.

43 Bhattacharya, S.; Acharya, S. N. G. *Chem. Mater*.**1999**, *11*, 3121–3132.

44 Gao, Y.; Hao, J.; Wu, J.; Zhang, X.; Hu, J.; Ju, Y. *Nanoscale* **2015**, *7*, 13568–13575.

45 Jung, J. H.; Do, Y.; Lee, Y-A.; Shimizu, T. *Chem. Eur. J*. **2005**, *11*, 5538–5544.

46 Zhang, L.; Wang, X.; Wang, T.; Liu, M. *Small* **2015**, *11*, 1025–1038.

47 IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and symbols for the description of conformations of polynucleotide chains. *Pure Appl. Chem*.**1983**, *55*, 1273–1280.

48 John, G.; Jung, J. H.; Masuda, M.; Shimizu, T. *Langmuir*, **2004**, *20*, 2060–2065.

49 Wang, X.; Xie, F.; Duan, P.; Liu, M. *Langmuir* **2016**, *32*, 12534–12541.

50 Ajayaghosh, A.; Praveen, V. K. *Acc. Chem. Res*. **2010**, *40*, 644–656.

51 Basak, S.; Nanda, J.; Banerjee, A. *J. Mater. Chem*. **2012**, *22*, 11658–11664.

52 Bhattacharjee, S.; Maiti, B.; Bhattacharya, S. *Nanoscale* **2016**, *8*, 11224–11233.

53 Liu, M.; Zheng, Y.; Zhai, J.; Jiang, L. *Acc. Chem. Res*. **2010**, *43*, 368–377.

54 Chu, Z.; Seeger, S. *Chem. Soc. Rev*. 2014, **43**, 2784–2798.

55 Zhang, J.; Seeger, S. *Angew. Chem., Int. Ed*. **2011**, *50*, 6652–6656.

56 Li, Y.; Li, L.; Sun, J. *Angew. Chem., Int. Ed*. **2010**, *49*, 6129–6133.

57 Yao, X.; Gao, J.; Song, Y.; Jiang, L. *Adv. Funct. Mater*. **2011**, *21*, 4270–4276.

58 Neirynck, P.; Schimer, J.; Jonkheijm, P.; Milroy, L.-G.; Cigler, P.; Brunsveld, L. *J. Mater. Chem. B* **2015**, *3*, 539–545.

59 Rao, K. P.; Higuchi, M.; Sumida, K.; Furukawa, S.; Duan, J.; Kitagawa, S. *Angew. Chem., Int. Ed*. **2014**, *53*, 8225–8230.

60 Mullangi, D.; Shalini, S.; Nandi, S.; Choksi, B.; Vaidhyanathan, R. *J. Mater. Chem. A* **2017**, *5*, 8376–8384.

61 Zhang, X.; Shi, F.; Niu, J.; Jiang, Y.; Wang, Z. *J. Mater. Chem*. **2008**, *18*, 621–633.

62 Liu, H.; Szunerits, S.; Xu, W.; Boukherroub, R. *ACS Appl. Mater. Interfaces* **2009**, *1*, 1150–1153.

63 Ishizaki, T.; Saito, N. *Langmuir* **2010**, *26*, 9749–9755.

64 Deng, X.; Mammen, L.; Butt, H.-J.; Vollmer, D. *Science* **2012**, *335*, 67–70.

65 Valamehr, B.; Jonas, S. J.; Polleux, J.; Qiao, R.; Guo, S.; Gschweng, E. H.; Stiles, B.; Kam, K.; Luo, T.-J. M.; Witte, O. N.; Liu, X.; Dunn, B.; Wu, H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 14459–14464.

66 Seo, J.; Lee, J. S.; Lee, K.; Kim, D.; Yang, K.; Shin, S.; Mahata, C.; Jung, H. B.; Lee, W.;. Cho, S.-W.; Lee, T. *Adv. Mater*. **2014**, *26*, 7043–7050.

67 Jin, Q.; Zhang, L.; Liu, M. *Chem. Eur. J*. **2013**, *19*, 9234–9241.

68 Cassie, A. B. D.; Baxter, S. *Trans. Faraday Soc*. **1944**, *40*, 546–550.

69 Clever, G. H.; Kaul, C.; Carell, T. *Angew. Chem., Int. Ed*. **2007**, *46*, 6226–6236.

70 Ono, A.; Torigoe, H.; Tanaka, Y.; Okamoto, I. *Chem. Soc. Rev*. **2011**, *40*, 5855–5866.

71 Scharf, P.; Müller, J. *ChemPlusChem* **2013**, *78*, 20–34.

72 Miyake, Y.; Togashi, H.; Tashiro, M.; Yamaguchi,H.; Oda, S.; Kudo, M.; Tanaka, Y.; Kondo, Y.; Sawa,R.; Fujimoto, T.; Machinami, T.; Ono, A. *J. Am. Chem. Soc*. **2006**, *128*, 2172–2173.

73 Urata, H.; Yamaguchi, E.; Funai, T.; Matsumura,Y.; Wada, S. *Angew. Chem., Int. Ed.* **2010**, *49*, 6516–6519.

74 Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem*. **1994**, *5*, 508–512.

75 G. M. Sheldrick, A Short History of SHELX. *Acta Crystallogr. Sect.* **2008**, *A64*, 112–122.

4A.7 Appendix-III: Characterization data of synthesized compounds

4A.7.1 Crystallographic Information

4A.7.1.1 Crystallographic data for nucleolipid **3a**

Compound identity	**3a**
Empirical formula	C_{22} H_{36} N_2 O_6
Formula weight	424.53
Temperature	296(2) K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	P 21 21 21

Unit cell dimensions	a = 5.6365(19) α = 90
	b = 7.316(2) β = 90
	c = 54.623(18) γ = 90
Volume	2252.6(13) $Å^3$
Z	4
Density (calculated)	1.252 Mg/cm^3
Absorption coefficient (μ)	0.091 mm^{-1}
F(000)	920
Crystal size	0.38 X 0.18 X 0.11 mm^3
Theta range for data collection	0.74 to 28.39
Index ranges	-7<=h<=7, -9<=k<=9, -71<=l<=72
Reflections collected	39741
Independent reflections	5625 [R(int) = 0.0767]
Completeness to theta = 28.39°	99.0 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.990 and 0.981
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5625 / 0 / 274
Goodness-of-fit on F^2	0.702
Final R indices [I>2sigma(I)]	R1 = 0.0351, wR2 = 0.0938
R indices (all data)	R1 = 0.0414, wR2 = 0.1018
Largest diff. Peak and hole	0.331 and -0.352 e.$Å^{-3}$
CCDC	1479220

4A.7.1.2 Crystallographic data for nucleolipid **3b**

Compound identity	**3b**
Empirical formula	$C_{24}H_{40}N_2O_6$
Formula weight	452.58
Temperature	100(2) K
Wavelength	1.54178 Å
Crystal system	Orthorhombic
Space group	P 21 21 21
Unit cell dimensions	a = 5.6059(2) α = 90

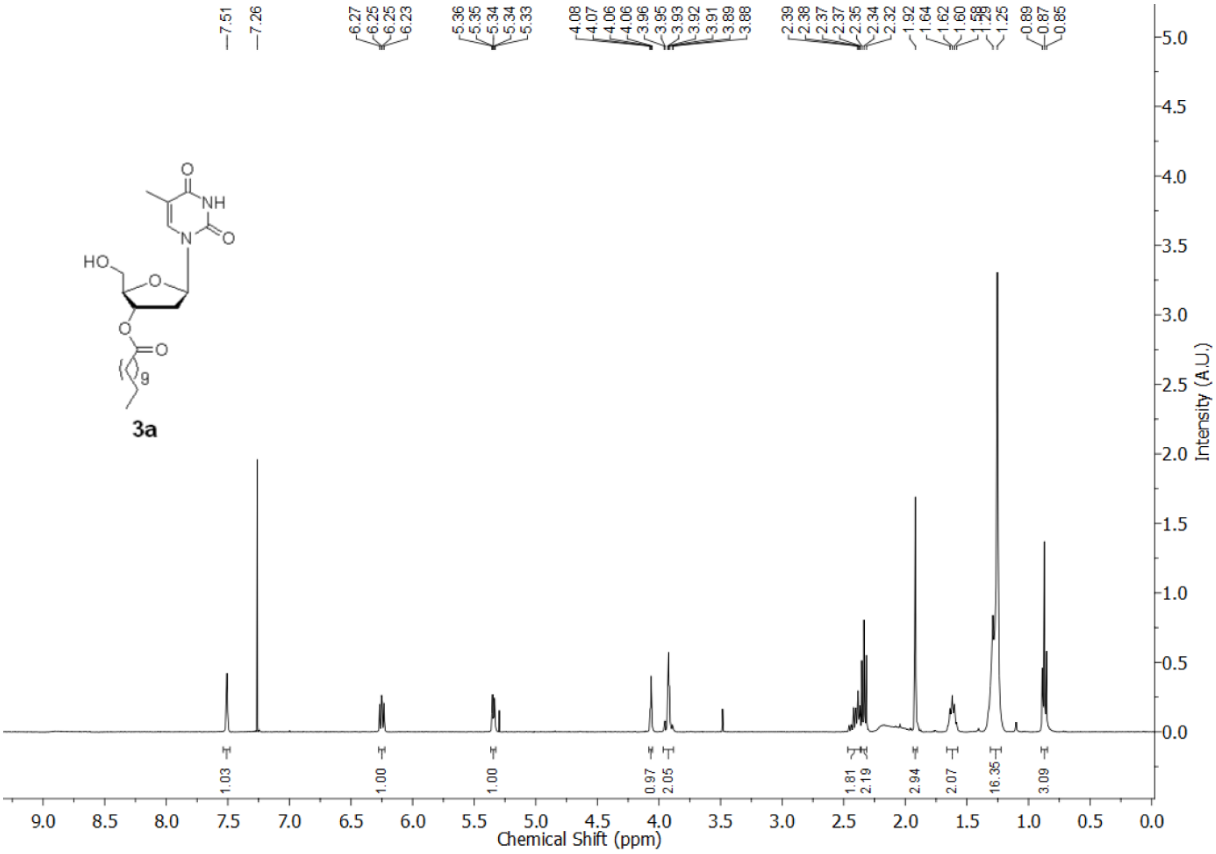
	b = 7.3036(2) β = 90
	c = 59.161(2) γ = 90
Volume	2422.24(16) $Å^3$
Z	4
Density (calculated)	1.241 Mg/cm^3
Absorption coefficient (μ)	0.719 mm^{-1}
F(000)	984
Crystal size	0.32 X 0.20 X 0.15 mm^3
Theta range for data collection	2.98 to 74.16
Index ranges	-6<=h<=6, -8<=k<=8, -72<=l<=73
Reflections collected	22088
Independent reflections	4664 [R(int) = 0.0704]
Completeness to theta = 74.16°	96.0 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.898 and 0.842
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	4664 / 0 / 292
Goodness-of-fit on F^2	1.481
Final R indices [I>2sigma(I)]	R1 = 0.0872, wR2 = 0.2235
R indices (all data)	R1 = 0.1028, wR2 = 0.2306
Largest diff. Peak and hole	0.340 and -0.322 e.$Å^{-3}$
CCDC	1479224

4A.7.1.3 Crystallographic data for nucleolipid **3c**

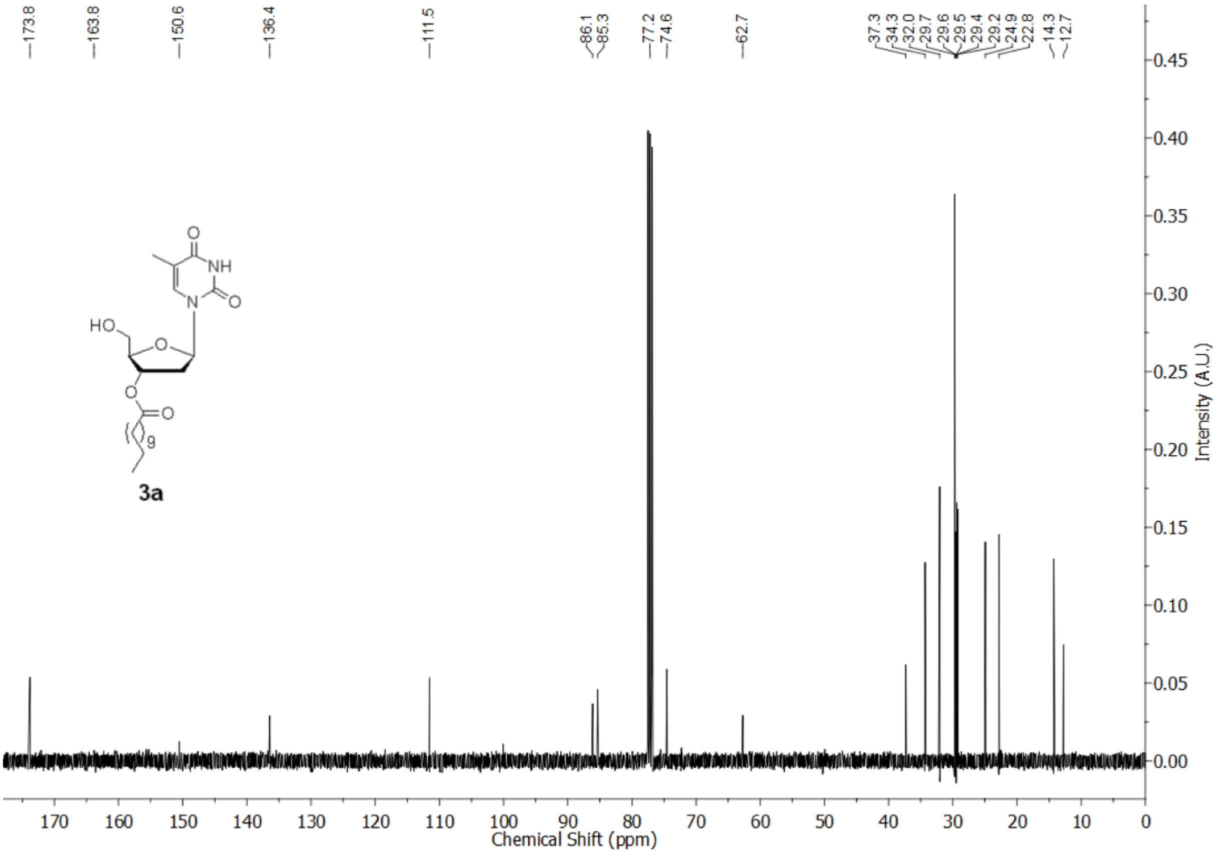
Compound identity	**3c**
Empirical formula	$C_{26}H_{44}N_2O_6$
Formula weight	480.63
Temperature	98(2) K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	P 21 21 21
Unit cell dimensions	a = 5.5928(19) α = 90
	b = 7.2712(2) β = 90

	c = 63.511(18) γ = 90
Volume	2582.8(7) $Å^3$
Z	4
Density (calculated)	1.236 Mg/cm^3
Absorption coefficient (μ)	0.087 mm^{-1}
F(000)	1048
Crystal size	0.80 X 0.45 X 0.10 mm^3
Theta range for data collection	0.64 to 28.44
Index ranges	-7<=h<=7, -9<=k<=9, -82<=l<=85
Reflections collected	40581
Independent reflections	6475 [R(int) = 0.0849]
Completeness to theta = 28.44°	99.5 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.991 and 0.954
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	6475 / 0 / 310
Goodness-of-fit on F^2	0.929
Final R indices [I>2sigma(I)]	R1 = 0.0519, wR2 = 0.1245
R indices (all data)	R1 = 0.0826, wR2 = 0.1503
Largest diff. Peak and hole	0.297 and -0.448 e.$Å^{-3}$
CCDC	1479225

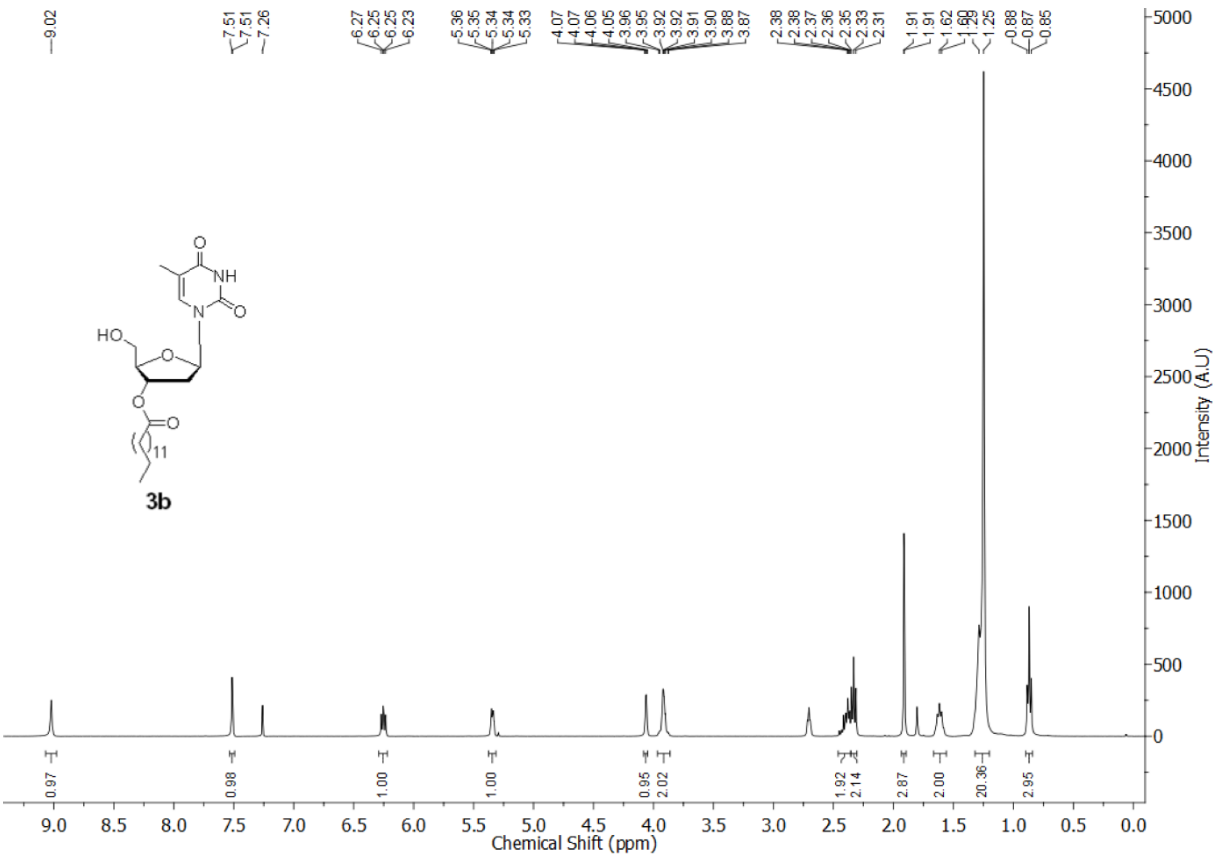
^1H NMR of nucleolipid **3a** in $CDCl_3$



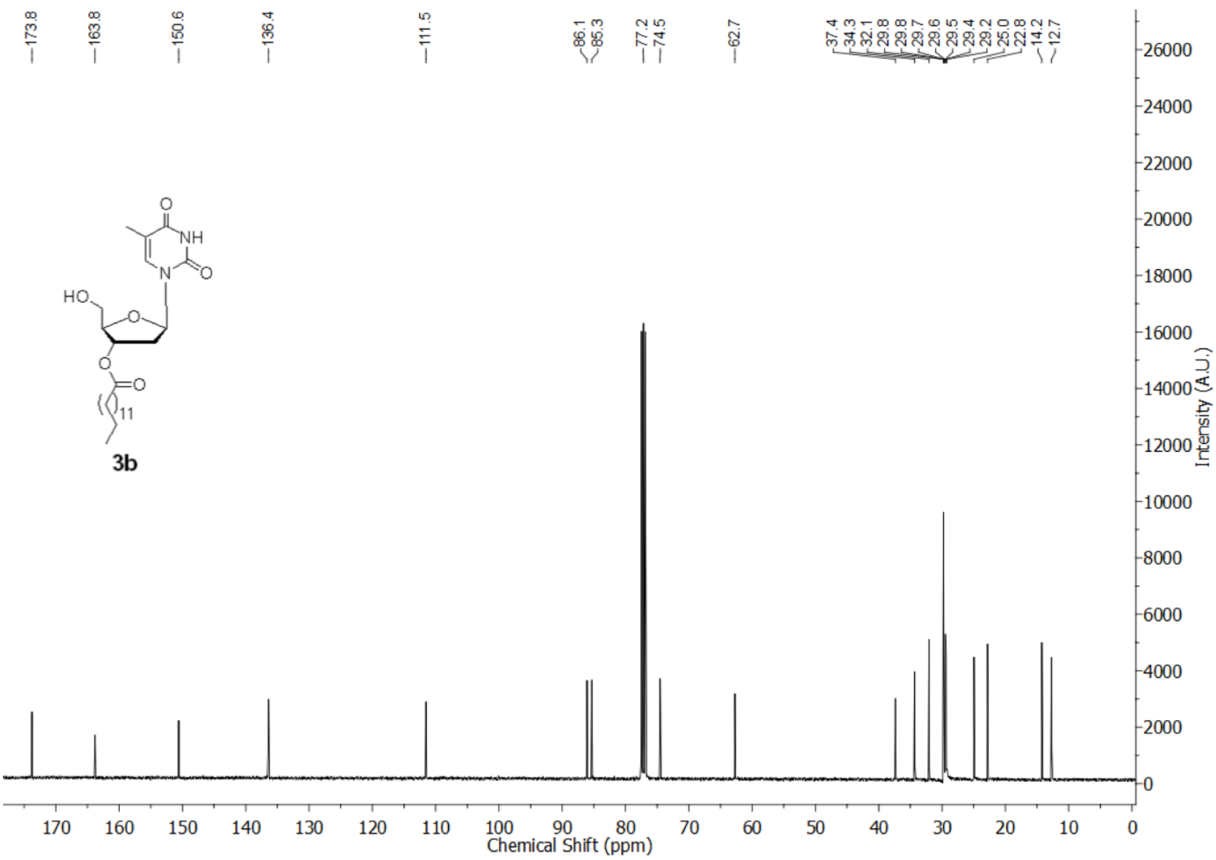
^{13}C NMR of nucleolipid **3a** in $CDCl_3$



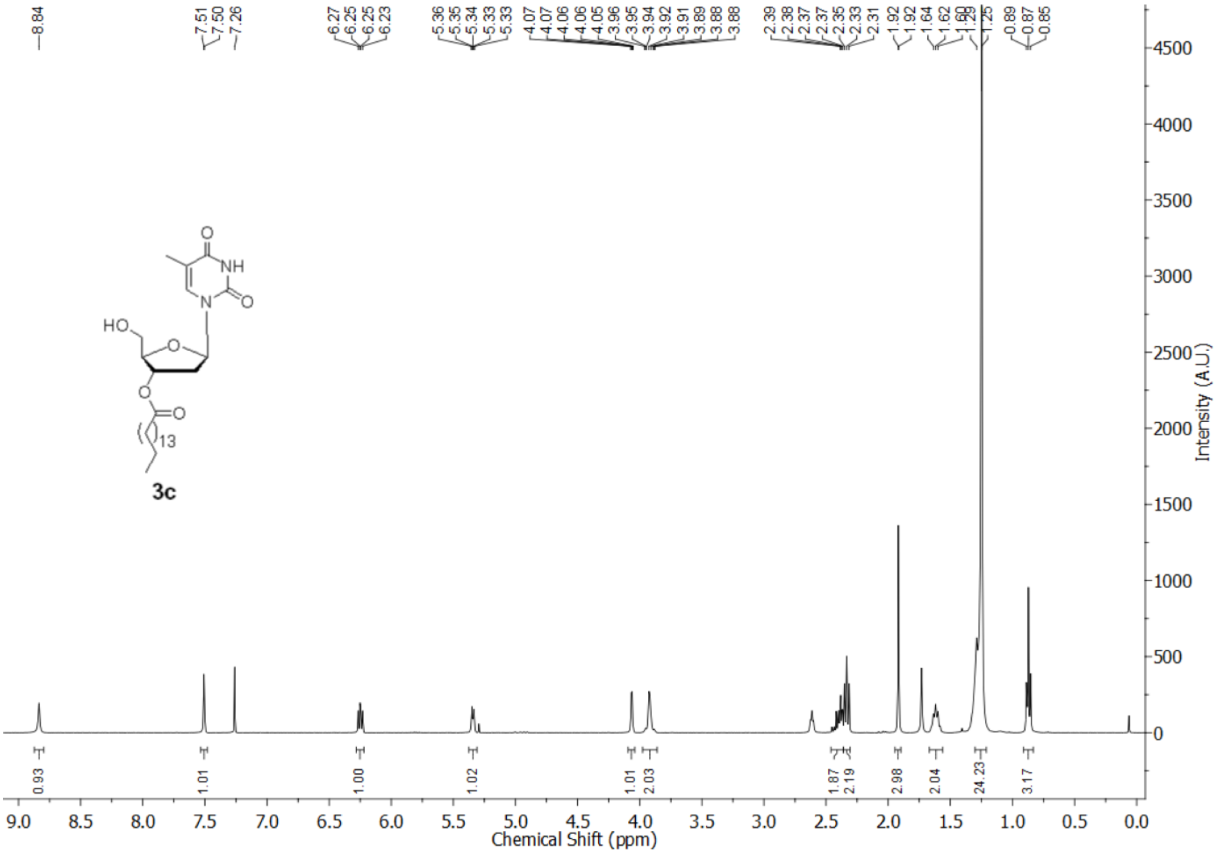
[1]H NMR of nucleolipid **3b** in $CDCl_3$



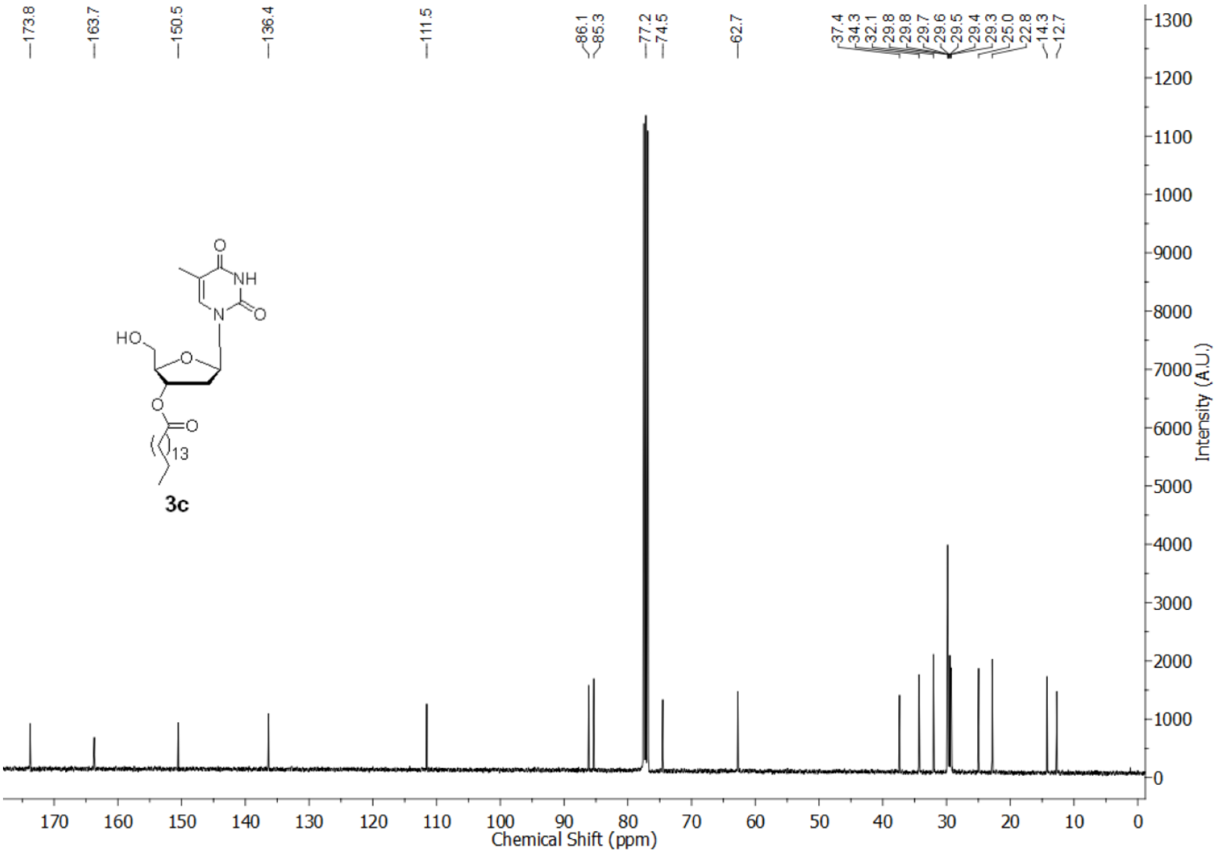
[13]C NMR of nucleolipid **3b** in $CDCl_3$



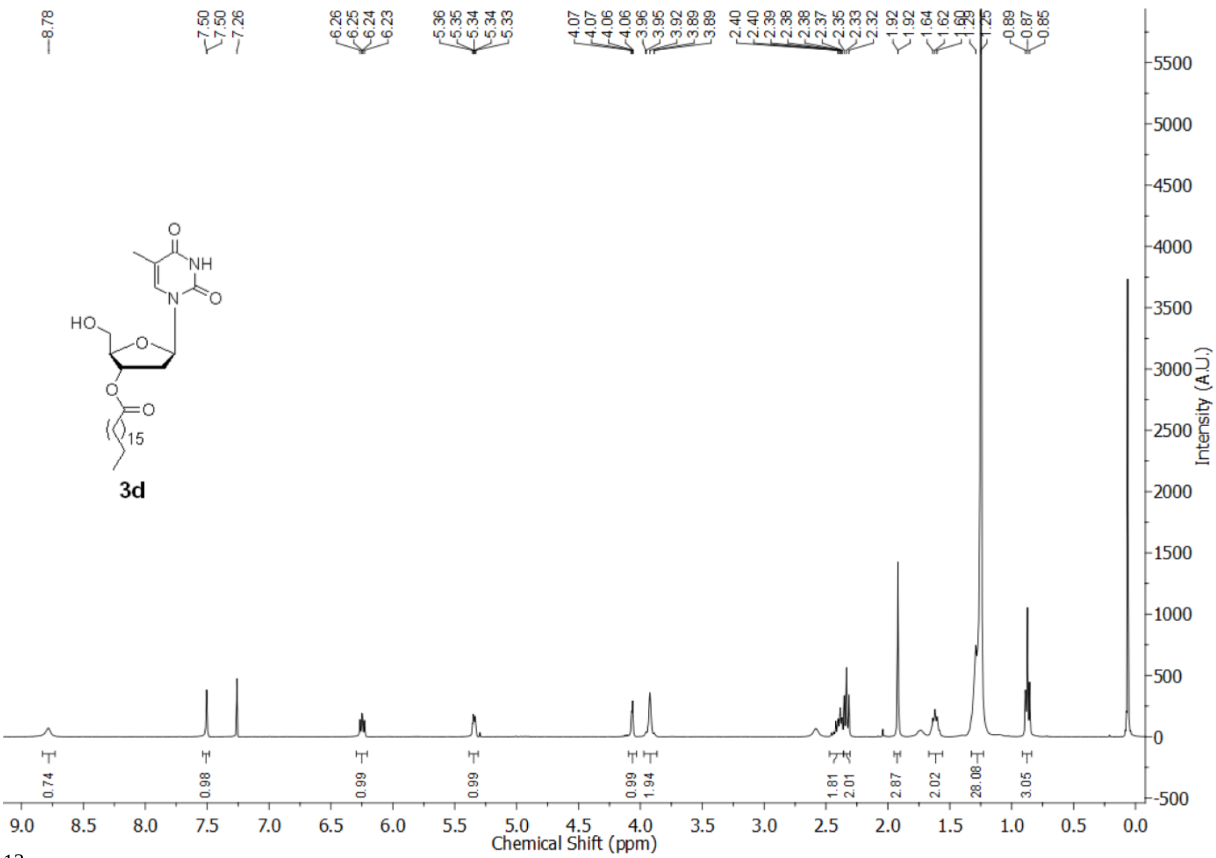
^{1}H NMR of nucleolipid **3c** in $CDCl_3$



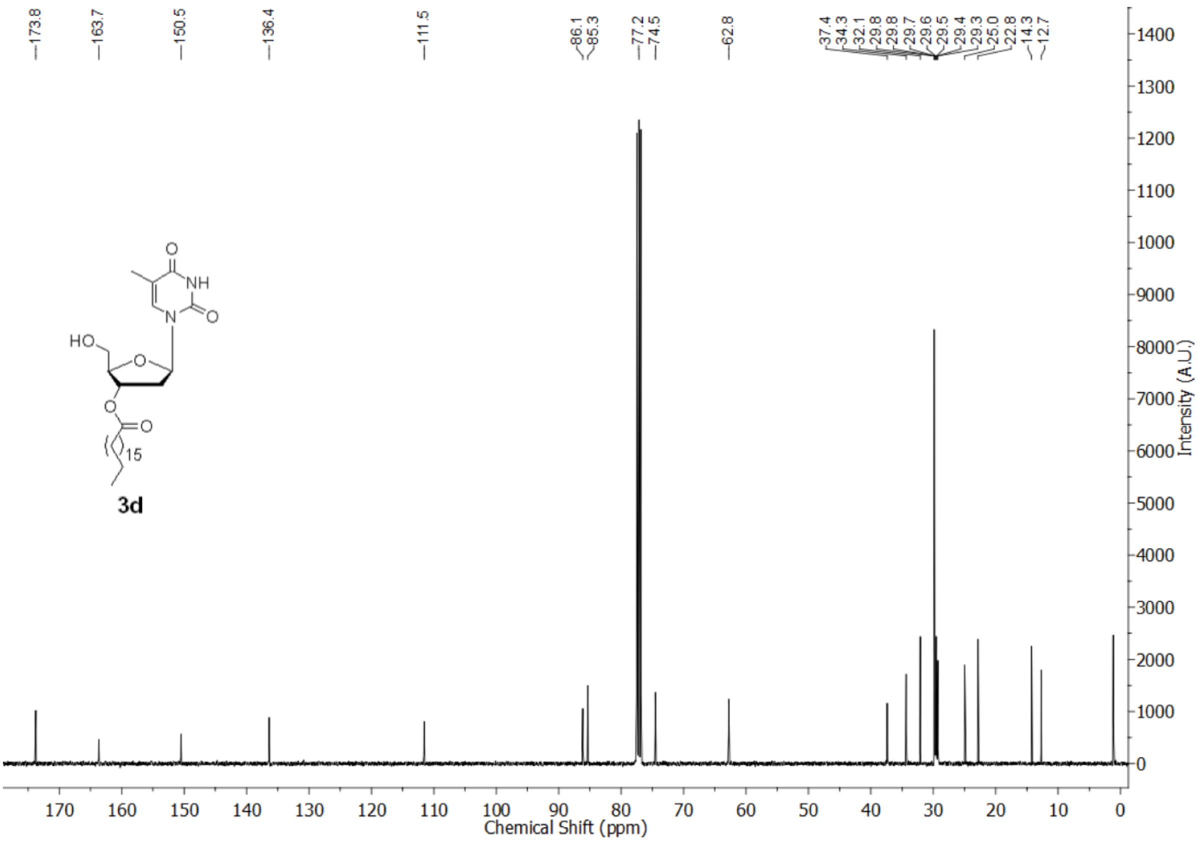
^{13}C NMR of nucleolipid **3c** in $CDCl_3$



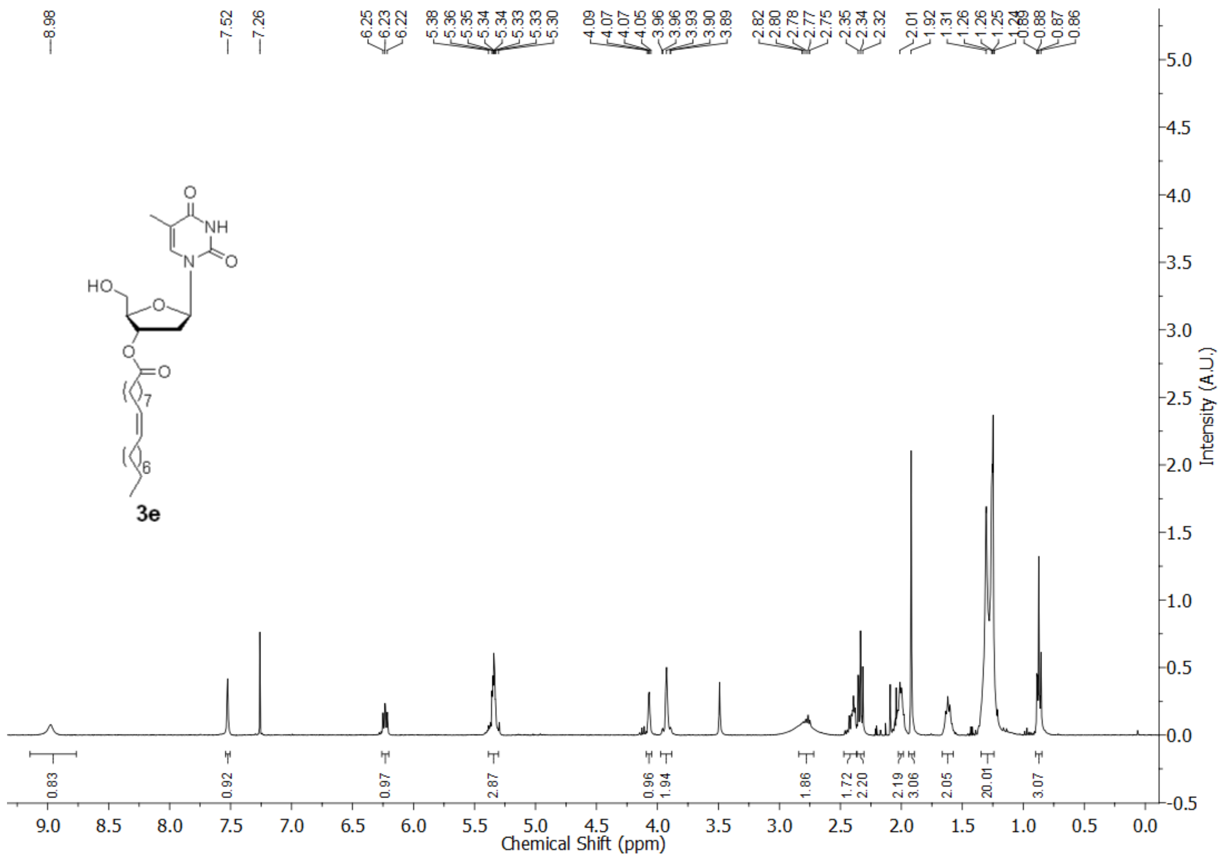
^{1}H NMR of nucleolipid **3d** in $CDCl_3$



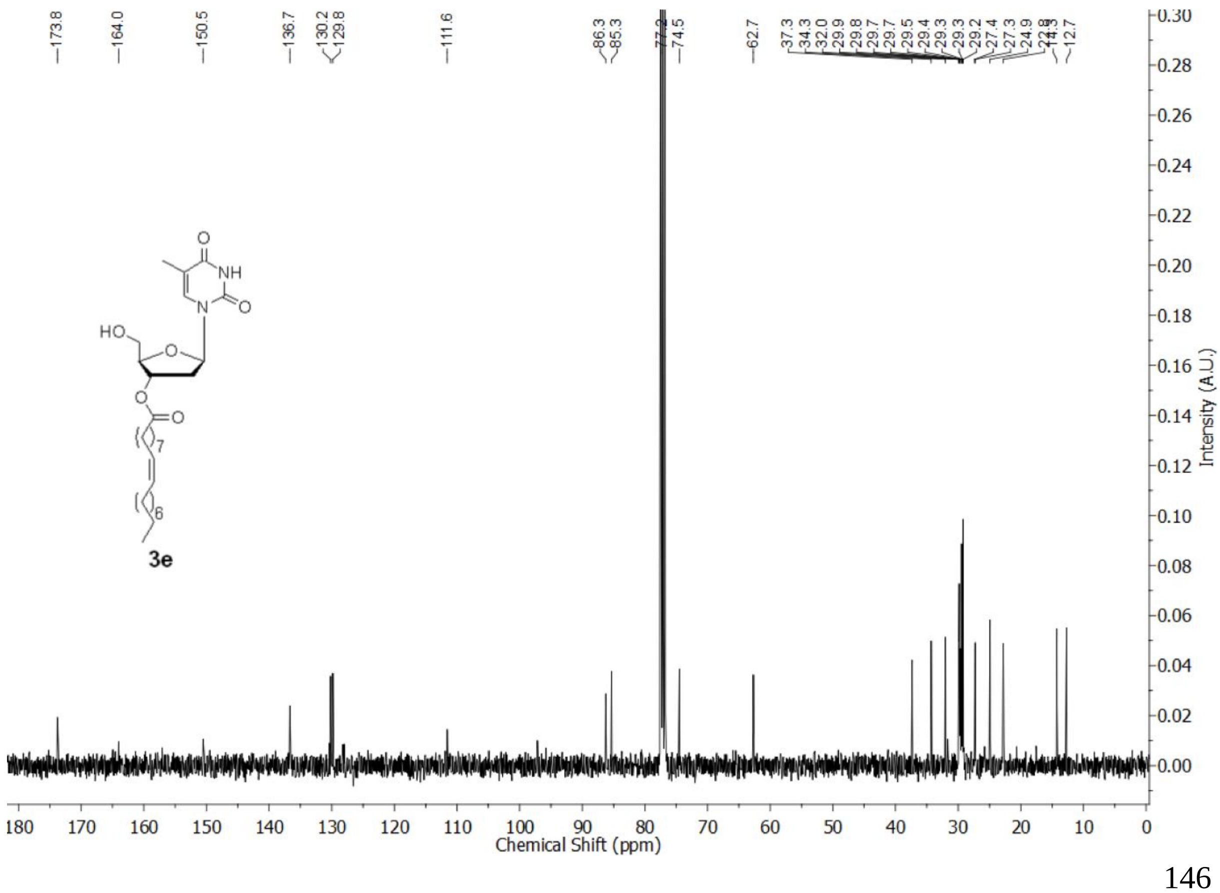
^{13}C NMR of nucleolipid **3d** in $CDCl_3$



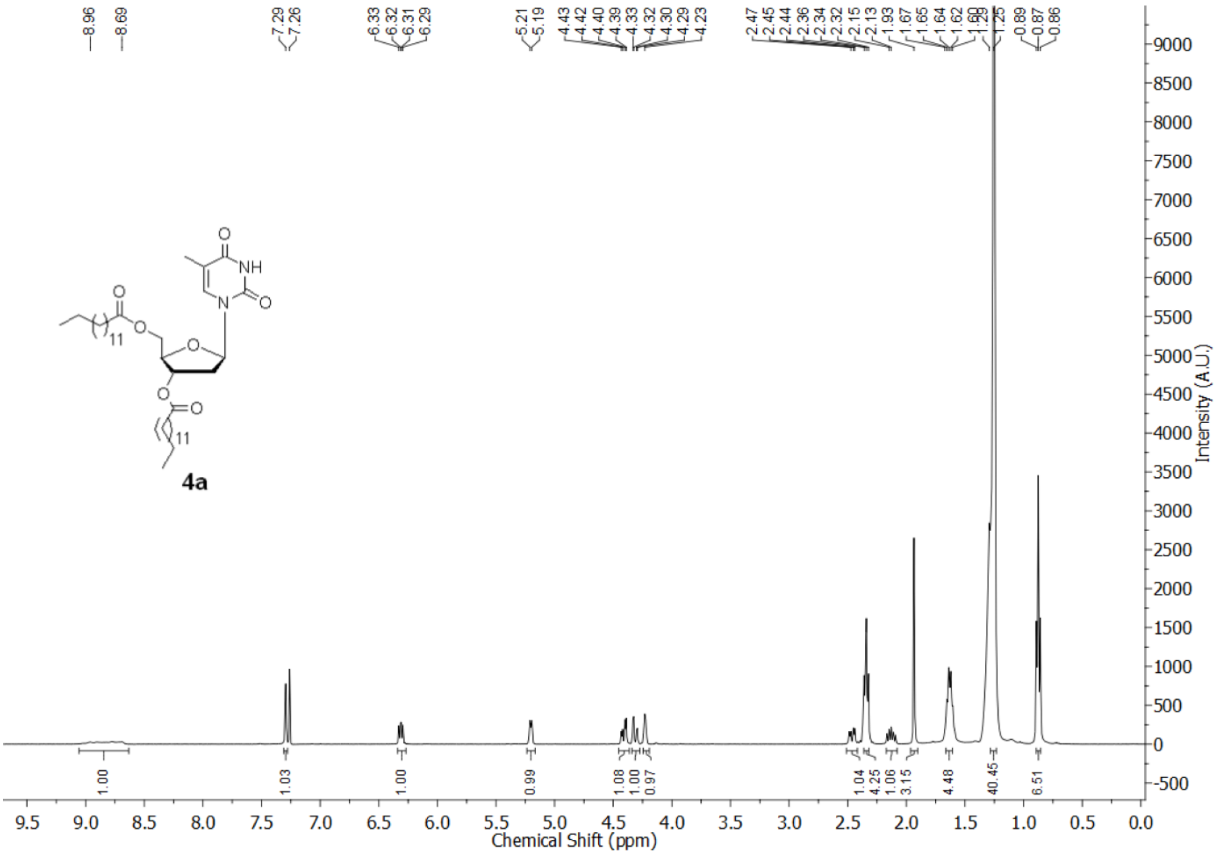
[1]H NMR of nucleolipid **3e** in $CDCl_3$



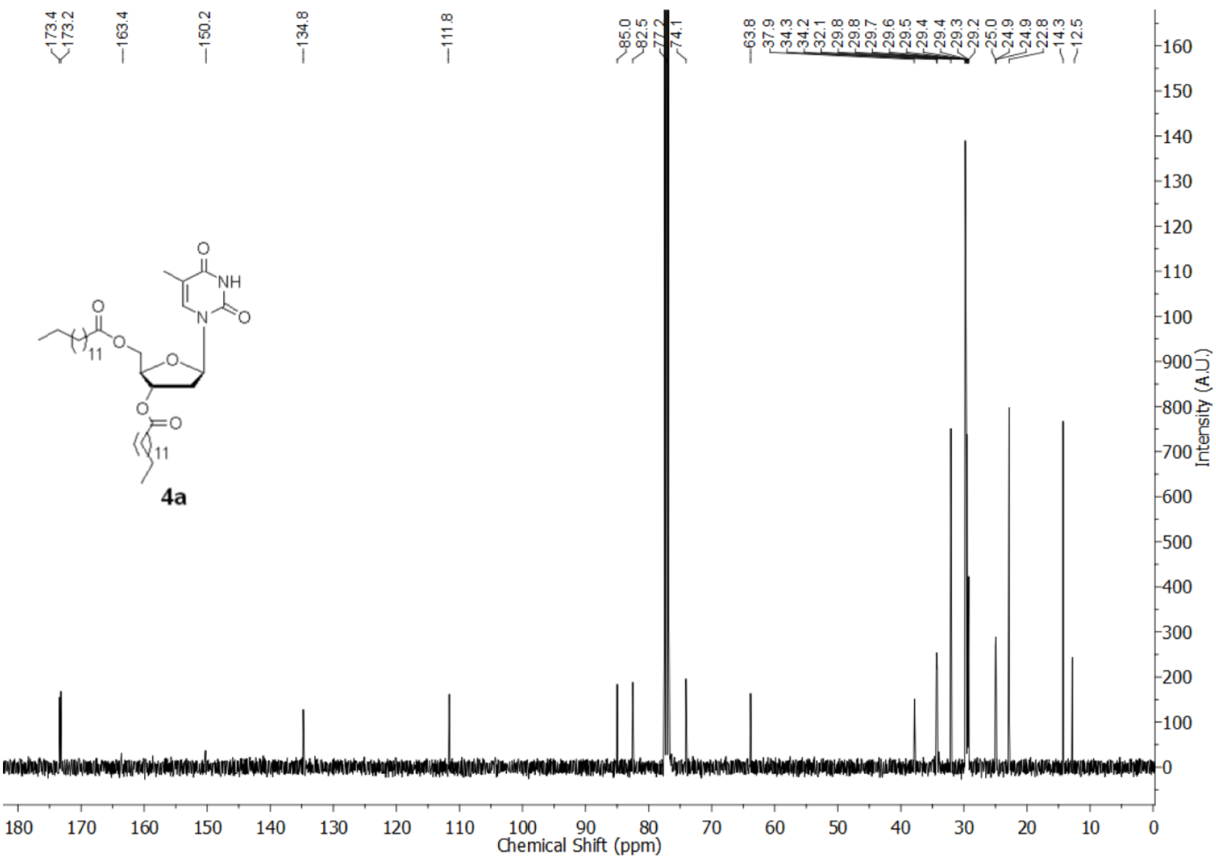
[13]C NMR of nucleolipid **3e** in $CDCl_3$



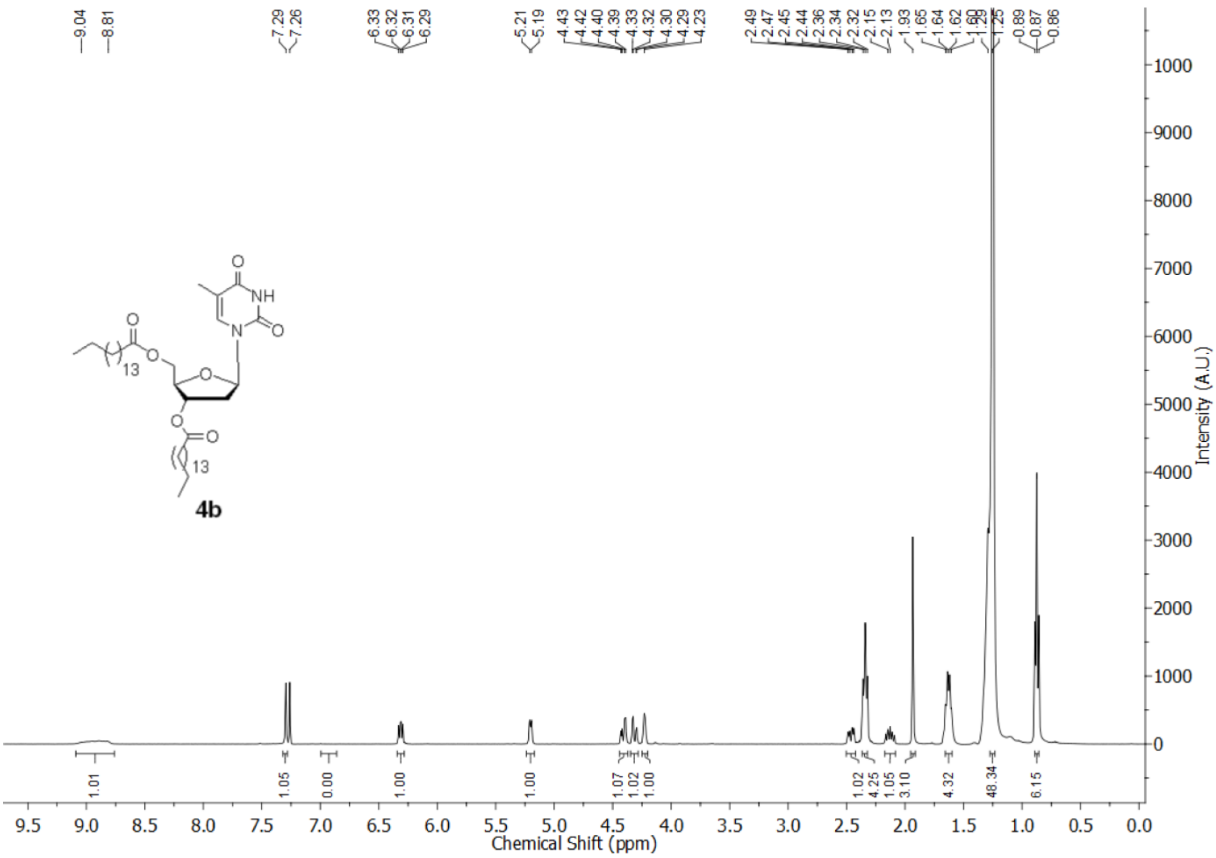
^{1}H NMR of nucleolipid **4a** in $CDCl_3$



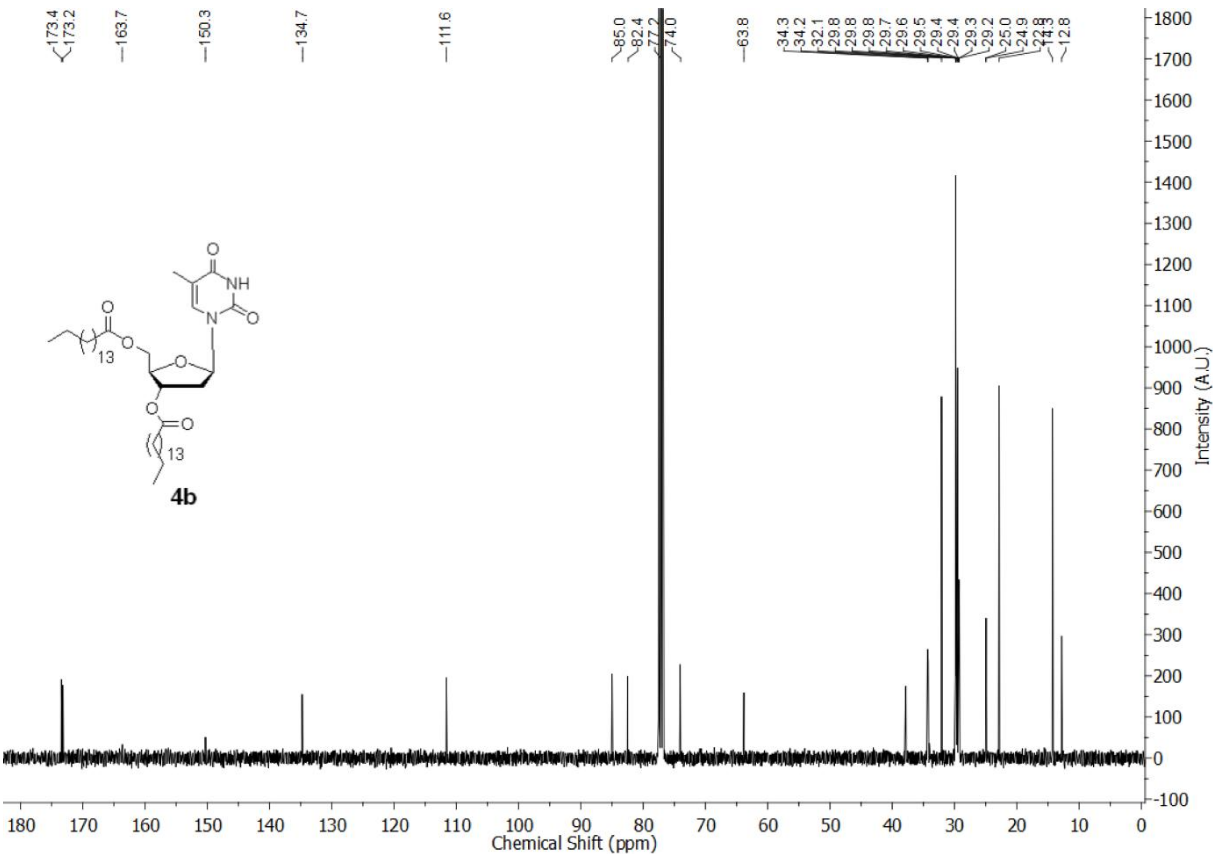
^{13}C NMR of nucleolipid **4a** in $CDCl_3$



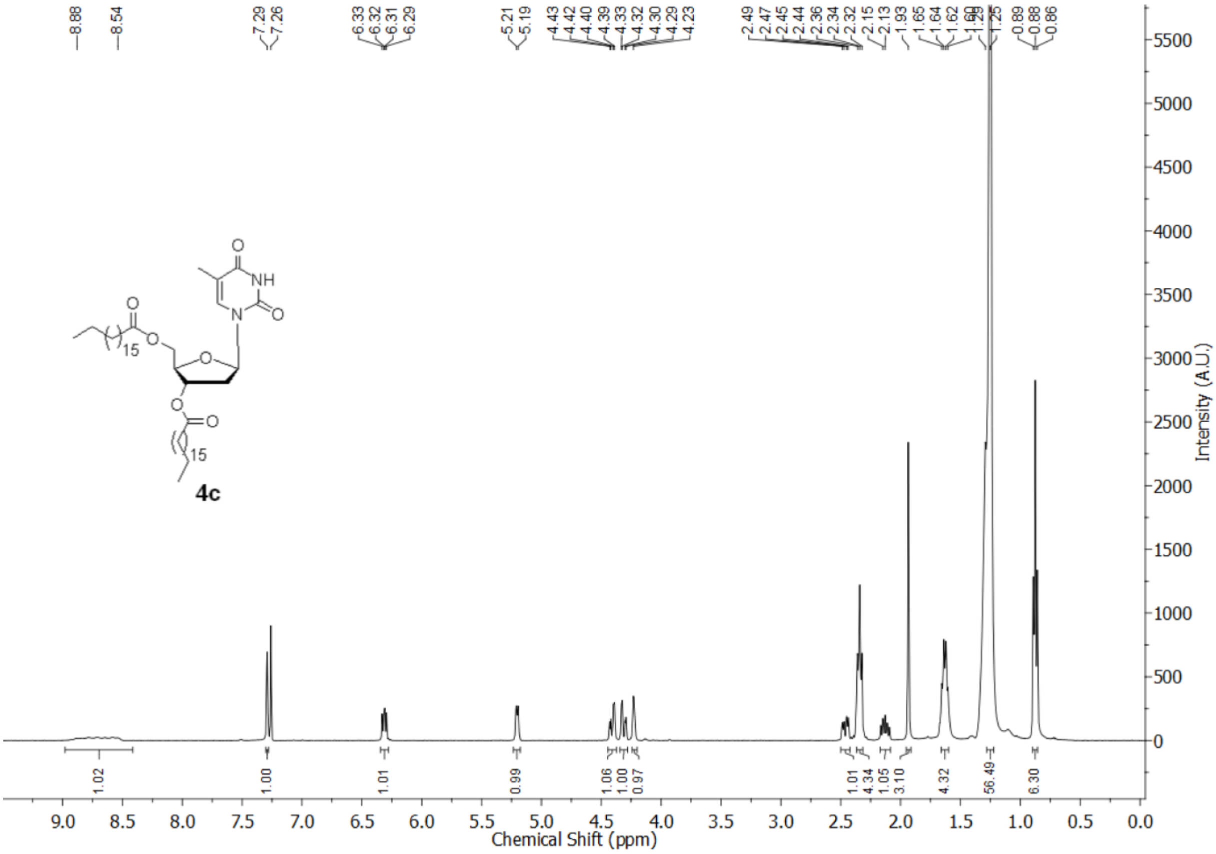
^{1}H NMR of nucleolipid **4b** in $CDCl_3$



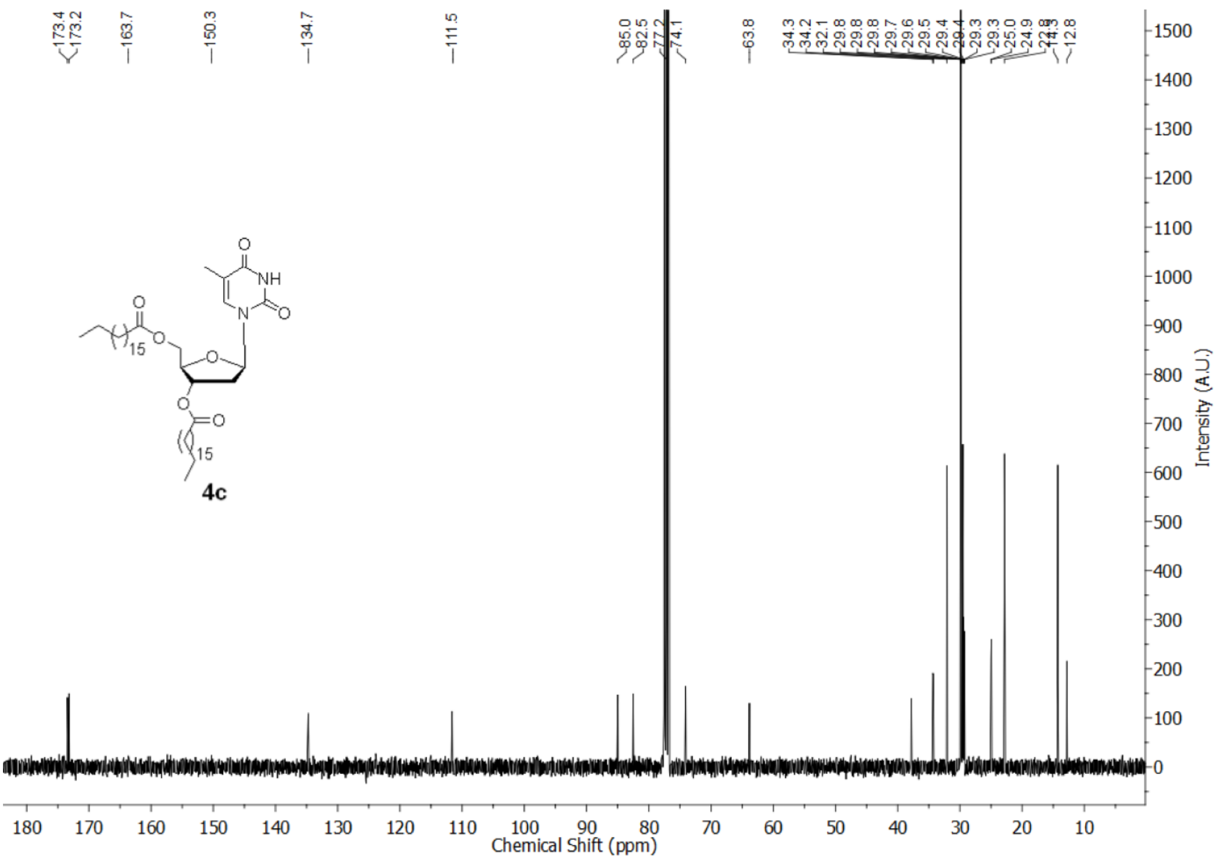
^{13}C NMR of nucleolipid **4b** in $CDCl_3$



^{1}H NMR of nucleolipid **4c** in $CDCl_3$



^{13}C NMR of nucleolipid **4c** in $CDCl_3$



Chapter 4B

Microscopic assemblies of nucleolipid supramolecular synthons show self-sorting and cooperative self-assembling process

4B.1 Introduction

Cocktails of two or more self-assembling species can self-sort or co-assemble to from complex but order systems.[1] These processes are common in biological systems (e.g., living cells), wherein dynamic assembling/disassembling events happening in a cooperative and or orthogonal manner maintain the overall function of the system.[2] Individual self-assembling species can be chemically similar or diverse. For example, DNA and RNA are made of largely similar chemical components and have similar recognition features, yet they have a definite role in certain cellular processes (e.g., DNA replication and RNA interference) and a cooperative role in certain other cellular processes (e.g., transcription). Ribosome, the protein translation machinery, when functioning, is composed of rRNA, mRNA, tRNA and several protein factors. This machinery is essentially an assembly of self-sorted as well as co-assembled RNAs and proteins. This level of organization originates from the inherent ability of different nucleic acid and protein sequences to adopt complex but defined structures either orthogonally or cooperatively. However, controlling the self-assembling process of a multicomponent system composed of synthetic supramolecular synthons is extremely challenging.[3]

Two or more supramolecular synthons, when mixed, can form self-sorted or co-assemblies. Formation of such assemblies will depend on the subtle differences in their interactions with one another and the conditions under which they form.[4] Further, the interactions that drive such processes are observed at the molecular to nanometer to micrometer levels. While predicting the self-sorting and co-assembling behavior of a given multi-component system is difficult, researcher have identified that such processes can lead to the developed of new functional materials.[5] Adams and coworker have shown that the mechanical properties and patterns of low molecular weight hydrogels can be modulated by using pH/light as a trigger to induce self-sorting, disruptive or co-assemblies.[6] Self-sorting supramolecular moieties have been used to tailor catalytic boxes for selective oxidation.[7] Using orthogonal self-assemblies of fibers of electron rich and electron poor gelators p-n bulk heterojunctions have been created.[8] Likewise, Smith and coworkers have used multi-component hybrid gels made of sorbitol, agarose and cationic micelles to deliver heparin in a controlled fashion.[9]

Combinations of synthetic oligonucleotides have been used to program co-assembling process to develop nano-architectures.[10] Many such assemblies have been used as machines,[11] ion channels,[12] membranes[13] and as containers for drug delivery.[14] Alternative, the basic components of nucleic acids, proteins and oligosaccharides namely, nucleosides, amino acids and glycosides, respectively, also co-assemble to form several architectures. Xu and coworkers have extensively used nucleobase, amino acids and glycosides to prepare biocompatible and multifunctional hydrogels.[15] Barthélémy's group has used glycosyl–nucleoside–lipid self-assembles to produce highly organized hydrogels for the delivery of nucleic acids into human cells.[16] While such material fabrications have used multi-component systems composed of chemically different biomolecules or synthetic molecules, use of chemically similar molecules has not been that fruitful.[17]

In this context, we describe the self-sorting and co-assembling properties of combinations of uridine nucleolipid, adenosine nucleolipid and nucleosides (Figure 1). Uridine nucleolipids containing fatty acid acyl chain of different lengths attached at 2′-*O*- and 3′-*O*-positions formed stable organogels. Depending on the solvent and alkyl chain length they formed twisted ribbon, long-range sheets and fibrous network. Whereas 2′,3′-*O*-difatty acid substituted adenosine nucleolipids formed porous microspheres and did not support gelation. When uridine nucleolipids were combined with adenosine nucleolipids of similar alkyl chain lengths, the gelation ability of uridine nucleolipids was completely arrested. On the other hand, addition of adenosine did not degelate the uridine nucleolipids, but reduced the gel strength. Morphological, ^{1}H NMR and rheological studies revealed self-sorting of assemblies due to interactions at the microscopic level. On the other hand, a gel made of a combination of uridine nucleolipids containing fatty acids of different chain lengths displayed significantly higher storage modulus (higher gel strength) as compared to gels made of individual nucleolipids. SEM, PXRD and rheological studies indicated a cooperative self-assembling process leading to the formation of a stronger gel. Apart from these features, gels of uridine nucleolipids were found to be responsive to multiple stimuli (chemical and physical).

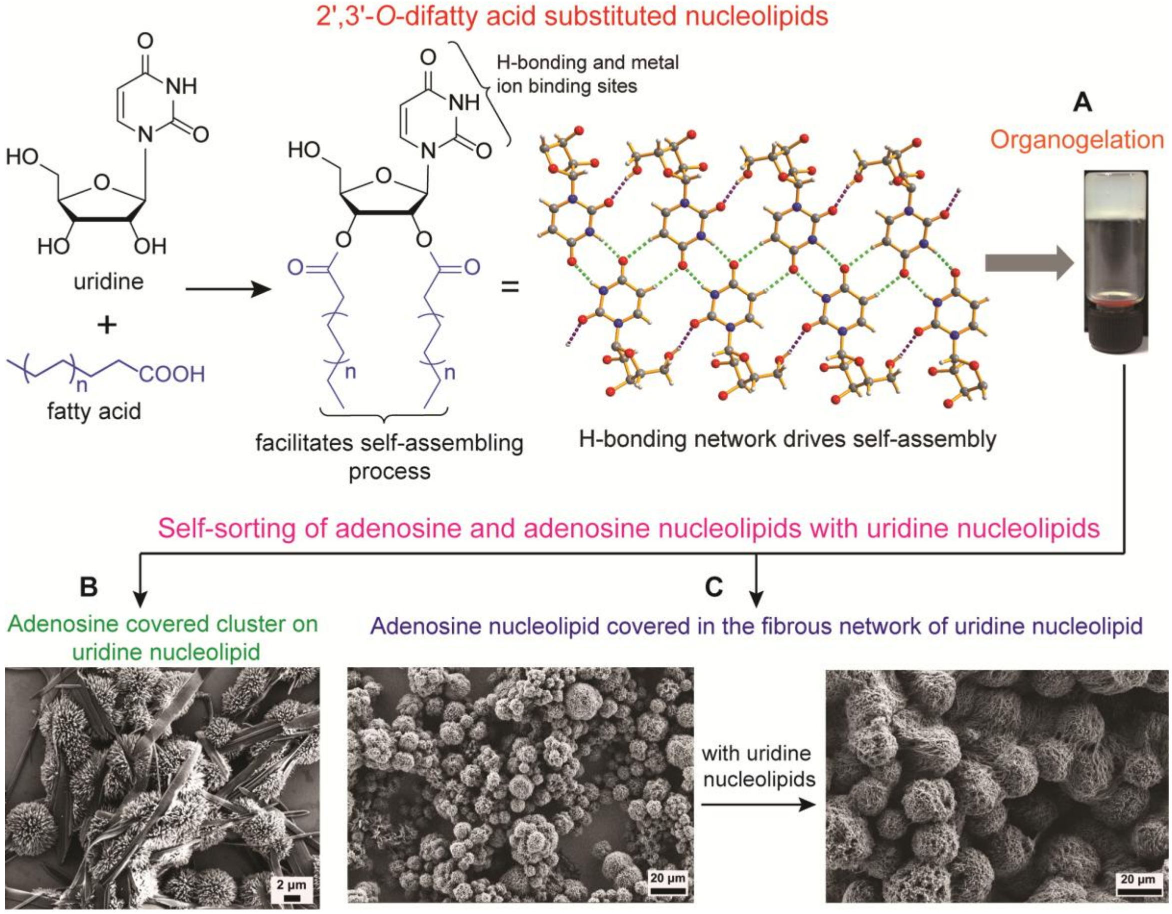
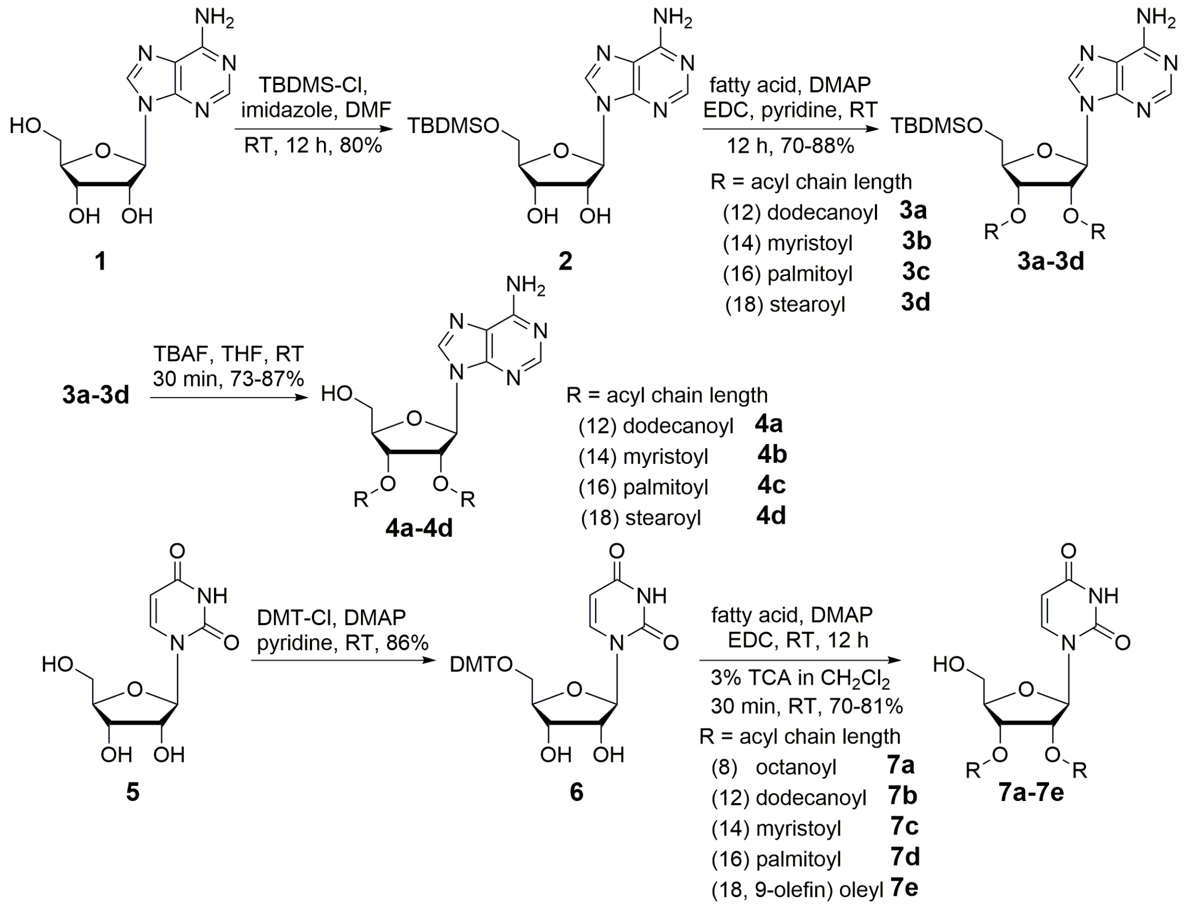


Figure 1. Design, hierarchical self-assembly and organogelation behavior of uridine nucleolipids. The coupling of long-chain fatty acids at the 2′,3′-*O*-position of uridine to facilitate the hydrophobic effect and without affecting the base H-bonding and metal binding sites of uridine nucleolipids. Interestingly, adenosine or adenosine nucleolipid are self-sorting in the presence of complementary uridine nucleolipid. Furthermore, uridine nucleolipids also exhibits multi stimuli responsiveness.

4B.2 Results and discussion

4B.2.1 Synthesis of adenosine (4a–4d) and uridine (7a–7e) nucleolipids

2′,3′-*O*-difatty acid-substituted adenosine (**4a**–**4d**) and uridine nucleolipids (**7a**–**7e**) were synthesized as illustrated in Scheme 1. Adenosine was 5′-*O*-protected with TBDMS group and was then coupled with different fatty acids using EDC. TBDMS group was deprotected in the presence of TBAF to give adenosine nucleolipids **4a**–**4d**. Uridine was first DMT-protected at the 5′-*O*-position and then coupled with fatty acids using EDC. Further, deprotection under acidic conditions produced uridine nucleolipids (**7a**–**7e**).



Scheme 1. Synthesis of 2′,3′-diacylated adenosine (**4a–4d**) and uridine (**7a**–**7e**) nucleolipids. DMT-Cl = 4,4′-dimethoxytrityl chloride; TCA = Trichloroacetic acid; DMAP = 4-Dimethylaminopyridine; EDC = *N*-(3-(dimethylamino)propyl)-*N*′-ethylcarbodiimide hydrochloride; TBDMS-Cl = tertiarybutyl dimethylsilyl chloride; TBAF = tetrabutylammonium fluoride.

4B.2.2 Gelation behaviour of nucleolipids

Attempts to form gels using adenosine nucleolipids failed in all tested pure solvents and solvent mixtures. Rewardingly, uridine nucleolipids (**7b**–**7d**) formed stable opaque organogels in various organic solvents ranging from polar to non polar (Figure 2). The gelation ability of nucleolipids was found to depend on alkyl chain length and nature of fatty acid (saturated or unsaturated). A nucleolipid of shorter chain length (**7a**) did not support gelation. As the fatty acid acyl chain length was increased from dodecanoyl to palmitoyl, the CGC was found to decrease (Table 1). Nucleolipid **7e** made of oleyl chain containing internal *cis* double bond did not support gelation. These observations indicate a direct correlation between the chain length and unsaturation in the self-assembling process leading to gelling of the solvent.[18] It was observed that the gel-sol transition was thermally reversible and the gels were stable for several months at room temperature.

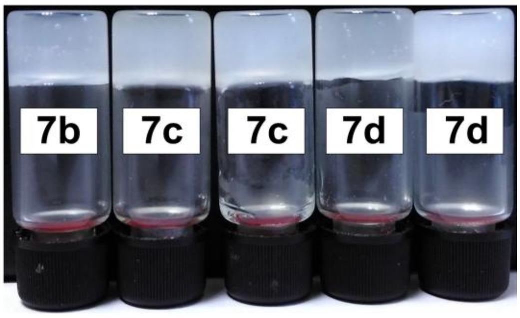


Figure 2. Photographs of uridine based nucleolipids organogels from left to right **7b** (2.4 w/v %) in MeOH, **7c** (1.2 w/v %) in DMSO, **7c** (1.0 w/v %) in MeOH and **7d** (0.7 w/v %) in DMSO, **7d** (0.8 w/v %) in MeOH, at respective CGC are given in bracket.

Table 1. Gelation behaviors of uridine nucleolipids in various organic solvents (**7a**–**7e**).

Solvent	**7a** (w/v %)	**7b** (w/v %)	**7c** (w/v %)	**7d** (w/v %)	**3e** (w/v %)
DMSO	S	G (5.0)	G (1.2)	G (0.7)	S
DMF	S	G (4.2)	G (4.0)	G (3.0)	S
acetonitrile	S	G (2.3)	G (1.4)	G (1.2)	S
methanol	S	G (2.4)	G (1.0)	G (0.8)	S
cyclohexanone	S	PG	G (4.0)	G (2.2)	S
toluene	S	S	G (4.0)	G (1.8)	S

G: gel; S: sol; PG: partial gel and parenthesis critical gelation concentration (CGC)

4B.2.3 Rheological analysis

The mechanical property of uridine nucleolipid organogels **7c** and **7d** were investigated by rheological measurements. The storage modulus (G′) and loss modulus (G″) of organogel **7c** and **7d** were examined as a function of shear strain at constant oscillating frequency. At lower strain values, organogels of **7c** and **7d** exhibited a G′ of 1721 and 2483 Pa, respectively, which are nearly one order of magnitude greater than its G″ of 395 and 360 Pa. This observation implies dominant elastic character of the gel. The crossover point of G′ and G″ for **7c** and **7d,** where gel transform into sol, was observed at 2.5% and 21.6% of strain (Figure 3). Therefore, the high storage modulus shown by the gel of **7d** as compared to **7c** reveals that the fatty acid chain length has profound effect on viscoelastic character of the gel.

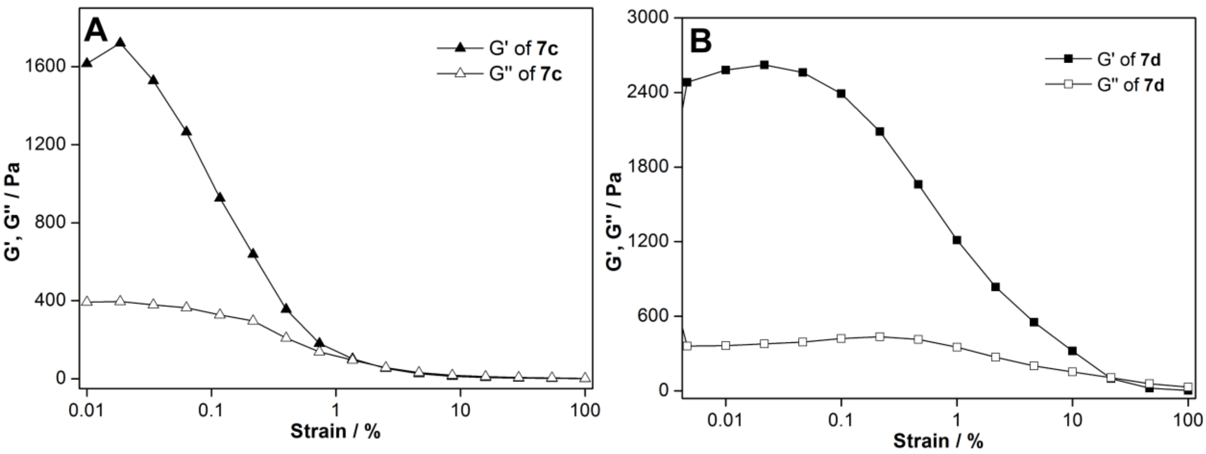


Figure 3. Strain sweep rheological measurements of nucleolipid organogels of **7c** (1.4 w/v %) (**A**) and **7d** (0.7 w/v %) (**B**) were taken at respective CGC in DMSO at constant oscillating frequency.

4B.2.4 Morphology analysis

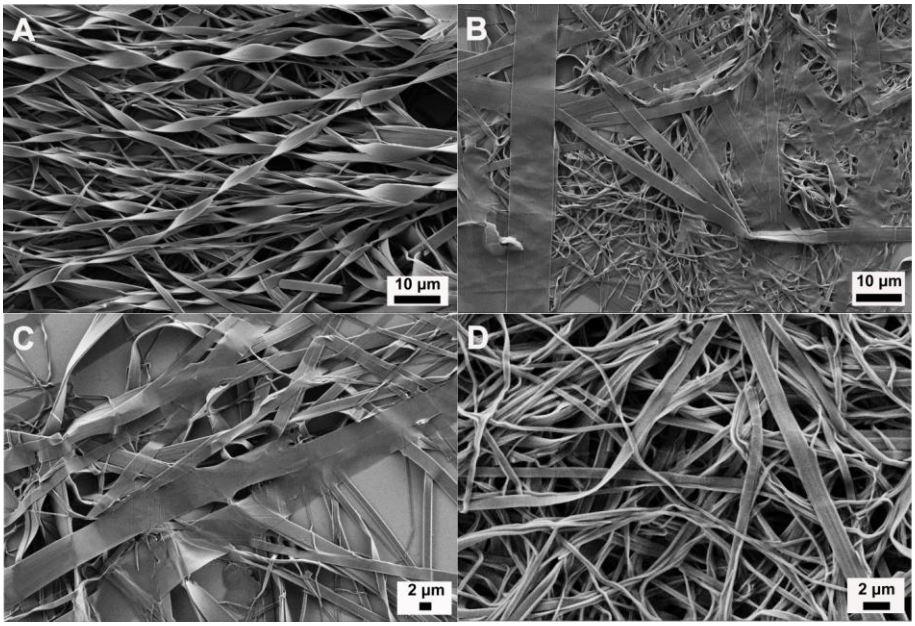


Figure 4. FESEM images of xerogels of nucleolipids. (**A**) Compound **7c** in DMF, (**B**) **7c** in DMSO, (**C**) **7d** in DMF and (**D**) **7d** in DMSO.

Morphology of xerogel of uridine nuclelipid gels was characterized by field emission scanning electron microscopy (FESEM). Morphology of xerogel was found to differ with fatty acid chain length and solvents. For instance, **7c** containing myristoyl chains in DMF formed twisted ribbon, whereas in DMSO formed long-range sheets (Figure 4A and 4B). SEM image of **7c** in DMSO also showed the presence of fibrous network, possibly the precursor for the sheets. **7d** containing palmitoyl chains in DMF and DMSO formed highly entangled sheets and fibrous network (Figure 4C and 4D), respectively. Here also we observed the presence of precursors namely, fibers and tapes formed by joining of fibers, which subsequently assembles into long-range sheets.

4B.2.5 Driving force for organogelation

The molecular interactions involved in the supramolecular assembly process were thoroughly studied by using single crystal analysis, powder X-ray diffraction, ^{1}H NMR techniques, and rheological analysis.

4B.2.5.1 Single crystal X-ray diffraction analysis

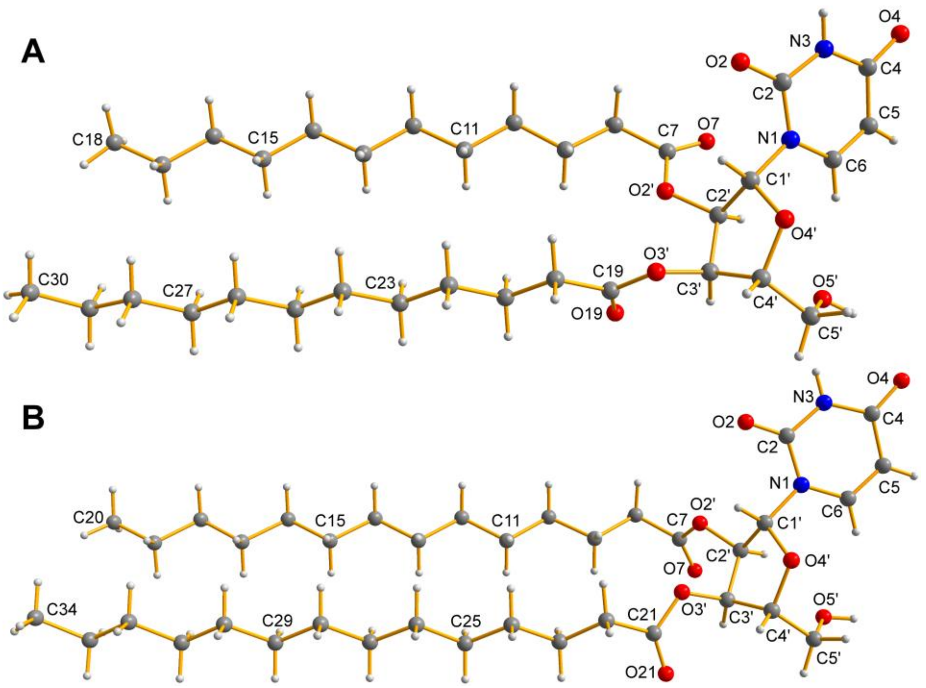


Figure 5. X-ray crystal structures of nucleolipid **7b** (**A**) and **7c** (**B**) showing one molecule in the unit cell. Nucleobase adopts an *anti* conformation relative to the sugar ring with C2′-endo sugar puckering. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

Crystal structure of nucleolipids **7b** and **7c** was found to be monoclinic with a space group C2 (Figure 5). According to the IUPAC recommendations, torsion angle (C2-N1-C1′-O4′) in the range of 180 ± 90° indicates an *anti* conformation.[19] Both the compounds displayed a torsion angle of **7b** and **7c** –124.3(4)° and –138.21(4)° indicating that the uracil ring is *anti* to sugar (see experimental section). Interestingly, the crystal structure of nucleolipid **7b** containing dodecyl chains did not show bifurcated and C6H---O H-bonding interaction as in **7c**. However, O2, N3H, O4 and 5′-OH atoms of the uridine were involved in H-bonding interactions with adjacent four independent molecules. The crystal packing of nucleolipid **7b** indicated that one uracil ring interacts with two adjacent uracil rings via a single H-bonding contact. An O2 and 5′-OH atoms of the same nucleolipid is H-bonds with 5′-OH and O2 atoms of adjacent nucleolipids (Figure 6 and Table 2).

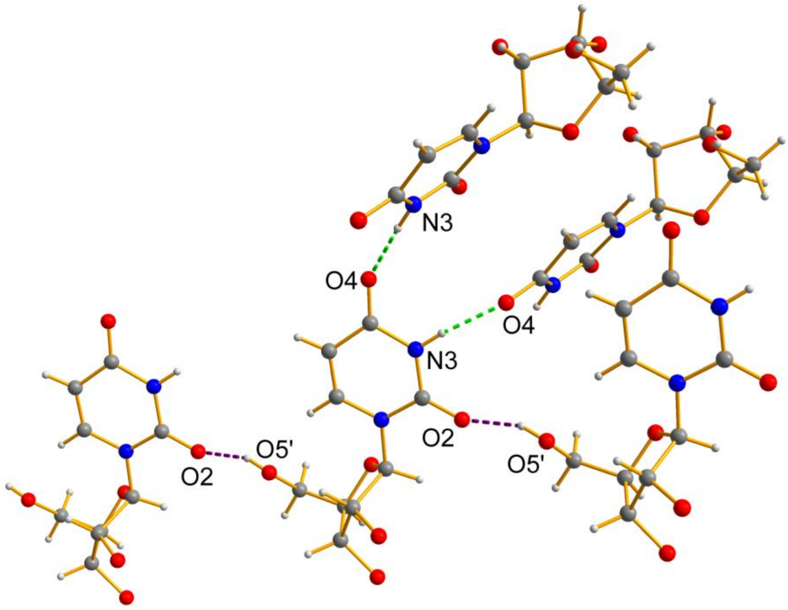


Figure 6. X-ray crystal structure of **7b** showing a detailed view of the H-bonding interactions along the crystallographic b-axes. The base recognition site is denoted green dashed lines and other interactions shown as violet dashed lines. The alkyl chains and hydrogen atoms other than the nucleosides are not shown for clarity. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

The packing structure of **7c** showed complex H-bonding interactions. One uracil ring was found to base pair with two adjacent uridine nucleolipid residues via multiple H-bonding interactions. N3H and O4 atoms from one face and O4 and C5H atoms from the other face of the uracil ring participates in H-bonding interactions to form a 1D sheet-like structure. The sheet is also stabilized by a strong H-bonding interaction between O2 (uracil) and 5′-OH (sugar) atoms of adjacent nucleolipids (Figure 7 and Table 2). Each 1D sheet further stabilized by hydrophobic interactions (Figure 8). The packing structure revealed head to head and tail to tail interaction between nucleolipids (Figure 9–11).

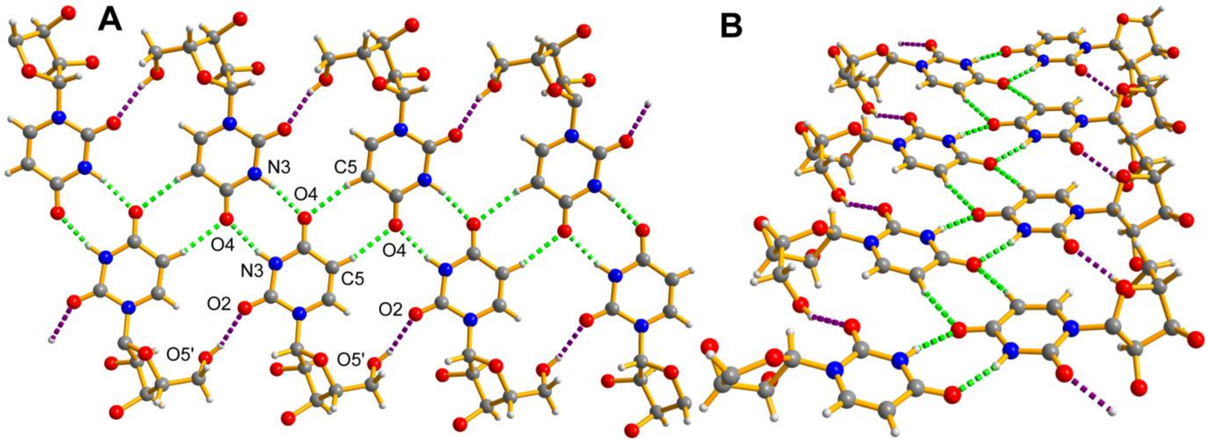


Figure 7. X-ray crystal structure of **7c** is showing a detailed view of the H-bonding interactions (**A**) along b-axes and (**B**) along c-axes. The base recognition part are shown in green dashed lines, other non-canonical H-bonds are denoted in violet dashed lines. Alkyl chains and hydrogen atoms other than the nucleosides are not shown for clarity. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

Table 2. H-bonding distances and angles, torsional angles measured from the crystal structure of uridine nucleolipids **7b** and **7c**.

nucleolipid	hydrogen bond	distance (Å)	Angle (°)	torsion angle (χ) (°)
7b	N3H---O4	1.949(4)	173.6(3)	C2-N1-C1′-O4′, –124.3(4)
	O2---HO5′	2.010(4)	151.9(3)	
	O4---HN3	1.949(5)	173.6(3)	
	O5′H---O2	2.010(5)	151.9(3)	
7c	N3H---O4	1.949(4)	167.7(3)	C2-N1-C1′-O4′, –138.21(4)
	O2---HO5′	2.020(4)	158.1(3)	
	O4---HN3	1.949(4)	167.7(3)	
	O4---HC5	2.278(4)	166.3(3)	
	C5H---O4	2.278(4)	166.3(4)	
	O5′H---O2	2.020(4)	158.1(3)	

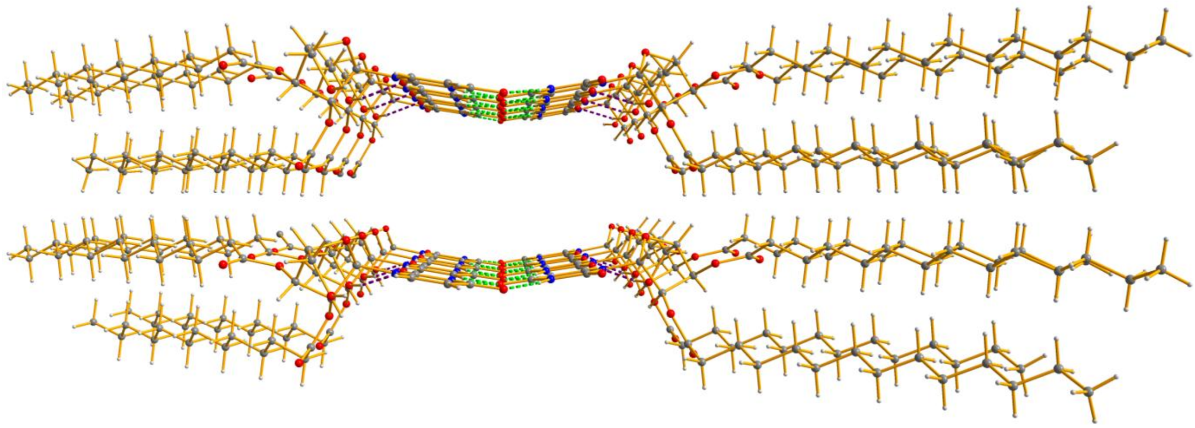


Figure 8. X-ray crystal structure showing a complete view of the non-covalent interactions between nucleolipid **7c** along the crystallographic c-axes. Here two 1D sheets join together through strong interlayer hydrophobic interactions, for sake of brevity base recognition parts are shown in green dashed lines, other non-canonical H-bonds are denoted in violet dashed lines. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

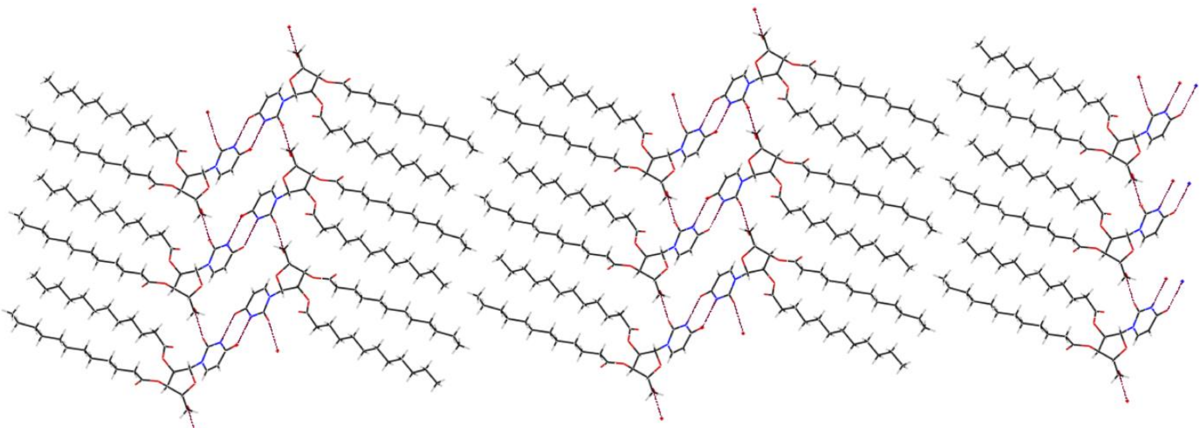


Figure 9. Complete packing diagram of compound **7b** along the crystallographic b-axes.

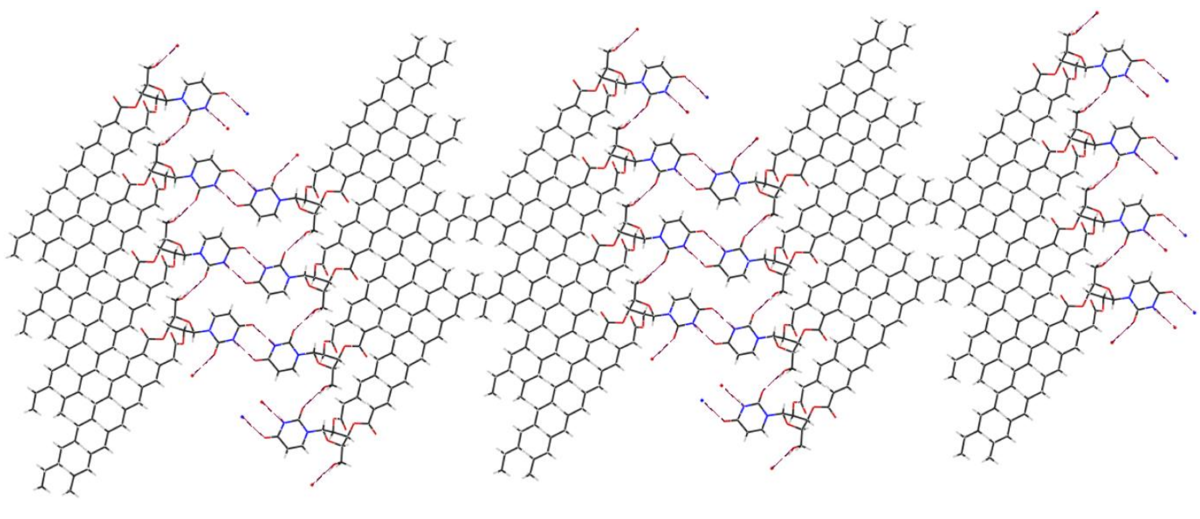


Figure 10. Complete packing diagram of compound **7c** along the crystallographic b-axes.

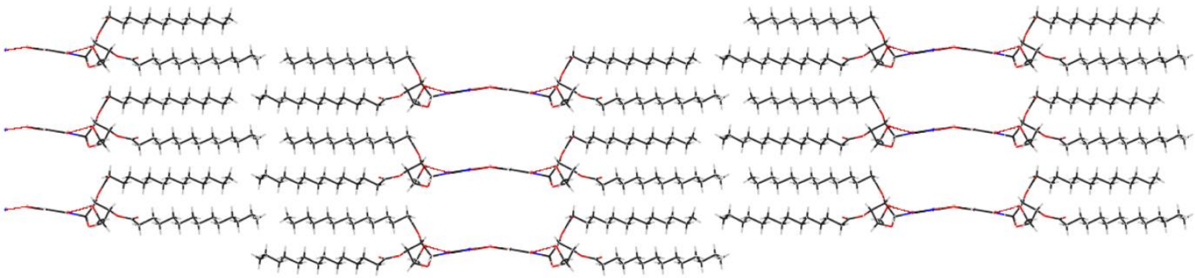


Figure 11. Complete packing diagram of compound **7c** along the crystallographic c-axes.

4B.2.5.2 Powder X-ray diffraction (PXRD) analysis

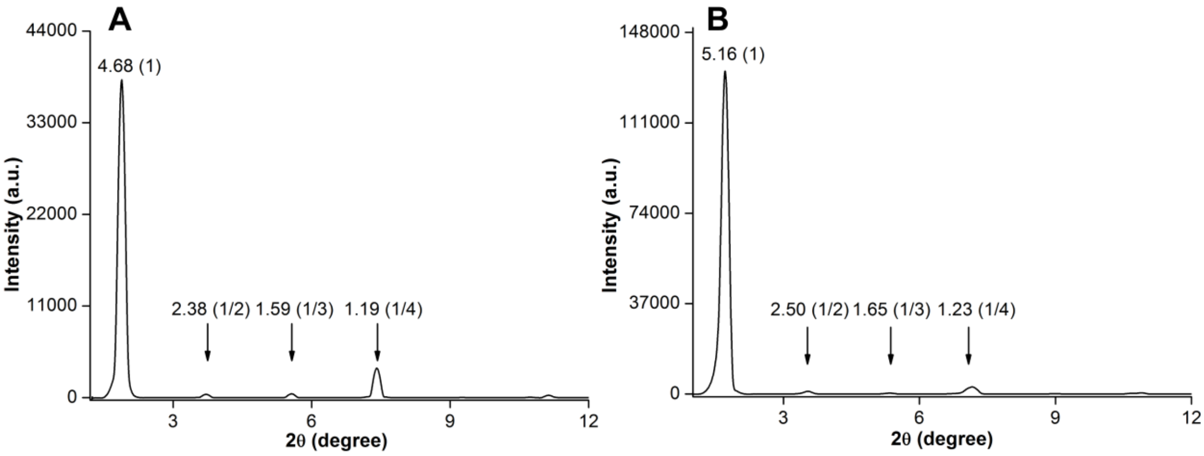


Figure 12. PXRD spectra of xerogels of (**A**) **7c** from DMSO, (**B**) **7d** from DMSO. Layer spacing (nm) for prominent diffraction peaks and relative ratio also given in bracket.

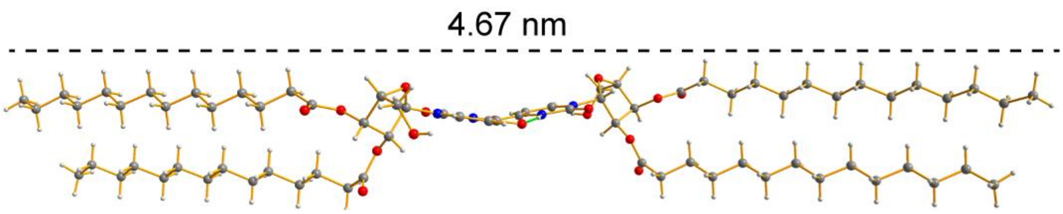
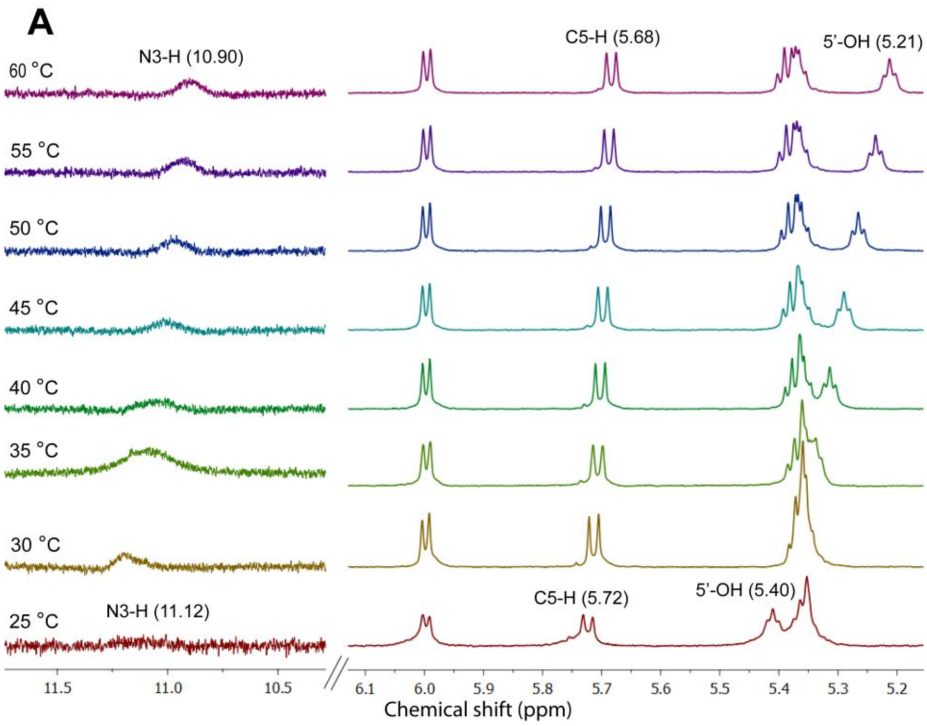


Figure 13. X-ray crystal structure of **7c** two molecule length as viewed using Diamond 3.0. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen. The length was measured as shown above (dashed line represent the length of the two molecules in nm).

In order to gain further insights into the molecular interactions and arrangement of the gelator molecules in the gel state, powder X-ray diffraction technique was employed. The dried xerogel of nucleolipid **7c** obtained from DMSO showed a well defined diffraction pattern. A low angle diffraction peak corresponding to an interplanar distance of 4.68 nm was observed. This distance matched well with the distance obtained from the crystal structure for two molecules of **7c** arranged in a head to head fashion (Figure 13). Further, diffraction signals corresponding to 4.68, 2.38, 1.59 and 1.19 nm are in the ratio of 1:1/2:1/3:1/4 (Figure 12A). This pattern is typical of a lamellar organization with bilayer formed by two molecules of nucleolipid as the basic bilayer unit. Similarly, xerogel of **7d** composed of palmitoyl chains also showed periodic lamellar organization (5.16, 2.50, 1.65, 1.23 nm, Figure 12B).[20] These observations indicate that the gels are formed by a hierarchical self-assembling process starting with basic bilayer unit, which forms lamellar structures. The lamellar structures assemble to form fibrils, which further assemble to form twisted ribbons, entangled sheets or tapes.[21]

4B.2.5.3 Variable temperature ^{1}H NMR experiments

^{1}H NMR experiment was carried out by using nucleolipids **7c** and **7d** at respective CGC in *d*6-DMSO. As the temperature of the gel was elevated from 25 °C to 60 °C, the N3-H, C5-H and 5′-OH proton signals of nucleolipids progressively shifted to upfield ($\Delta\delta$ = 0.22, 0.04, 0.19 and 0.18, 0.03, 0.17 ppm, respectively, Figure 14). This shift is due to the breaking of H-bonding interactions during the gel-sol transition. Apart from these hydrogens, chemical shifts of other hydrogen atoms remained unchanged. These results confirm that the H-bonding interactions in the gel state and crystal structure are the same.



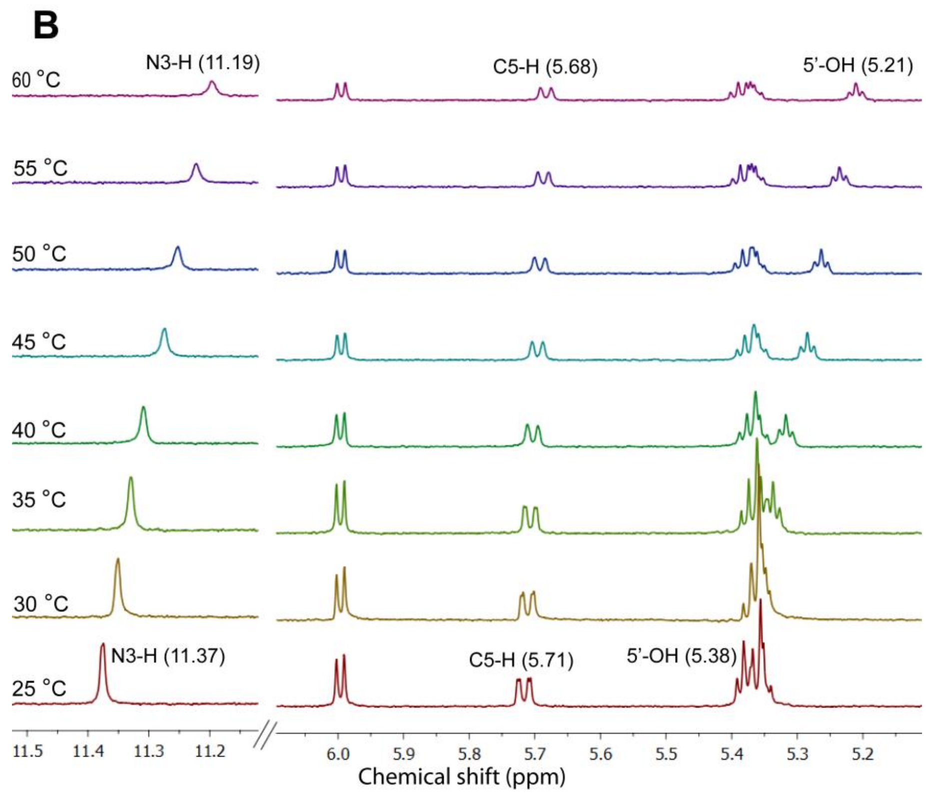


Figure 14. Partial ^{1}H NMR spectra of nucleolipid gels of **7c** (1.2 w/v %) (**A**) and **7d** (0.7 w/v %) (**B**) in *d6*-DMSO as a function of increasing temperature. N3-H, C5-H and 5′-OH, which are involved in strong intermolecular H-bonding in crystalline state, shown perceptible upfield shift in their proton signals during gel-sol transition. These results reveal that the H-bonding interactions in crystal structure and in gels are similar. For more details see experimental section.

4B.2.6 Self-sorted assemblies and co-assemblies of nucleolipids

An important aspect in our supramolecular synthon design is to retain the intrinsic H-bonding capability of the nucleobase so that it could base pair with itself or complementary nucleoside/nucleolipid. In this way, individual components and a mixture of components could exhibit distinct self-assembling behavior, thereby enabling the development of materials with different properties. A single component system of uridine nucleolipids as discussed above shows very good gelation ability in organic solvents. The gelation process is driven by hierarchical self-assembly with the basic bilayer unit made of two nucleolipids arranged in a head to head fashion. The head to head interaction is created by multiple H-bonding interactions between uracil rings. Further, crystal packing indicates that one uracil pairs two adjacent uracil rings. These observations prompted us to investigate the effect of addition of nucleolipids and nucleosides on the gelation as well as self-assembling process of uridine nucleolipid gelators (Figure 15).[22]

Figure 15. Photographs of nucleolipids gels **7d** DMSO at CGC with various additives from left to right (1–7 respectively, 7d, 7d-4c, 7d-7c, 7d-adenosine, 7d-guanosine, 7d-cytidine and 7d-uridine).

4B.2.6.1 Uridine and adenosine nucleolipids form self-sorted assemblies: Addition of 1 equivalent of adenosine nucleolipid to uridine nucleolipid (of same fatty acid chain length) completely abrogated the gelation ability of the uridine nucleolipid (Figure 15). Rheological measurement was carried out using a partial gel obtained using 1 equivalent of uridine nucleolipid **7d** and 0.5 equivalent of adenosine nucleolipid **4c**. The storage modulus (G′) of organogel of **7d** dramatically decreased from 2483 Pa to 14 Pa for the partial gel, which suggests that adenosine nucleolipid destabilizes the gel of uridine nucleolipid (Figure 16).

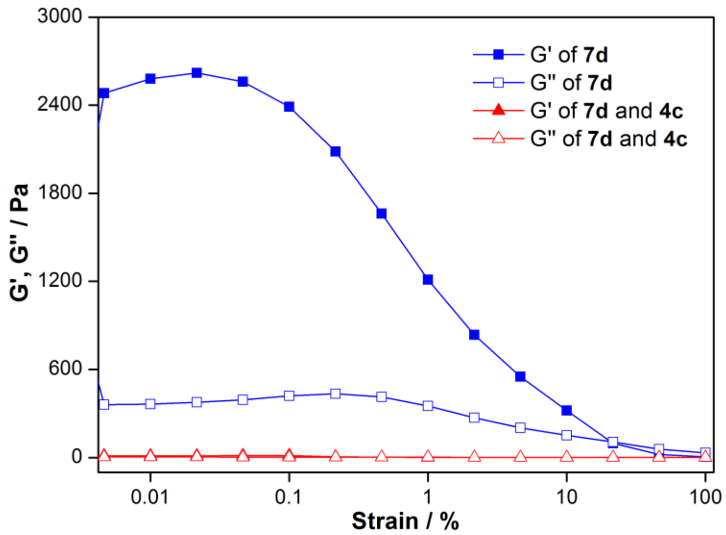


Figure 16. Comparison of strain sweep rheological measurements of nucleolipid organogel **7d** (0.7 w/v %) with 0.5 equiv. of **4c** in DMSO at constant oscillating frequency.

Comparison of morphologies of individual nucleolipids and their combination revealed interesting self-assembling phenomenon. While adenosine nucleolipids alone (does not form gels by itself) formed clusters of porous microspheres (Figure 17A and B), uridine nucleolipids largely formed highly entangled fibrous network (Figure 17C and D). SEM images of a mixture of uridine and adenosine nucleolipids revealed self-sorted assemblies,[4,6] wherein the porous microspheres of adenosine nucleolipid were covered in the fibrous network of uridine nucleolipid (Figure 17E and F). Despite the possibility of uridine base pairing with adenine, which could potentially lead to co-assembly, the mixture self-sorts. A closer look at the images indicated that the morphology of individual nucleolipids was not affected by the mixing of the two nucleolipids. Importantly, the images suggest that the driving force for the self-sorting is relative differences in the kinetics of formation of individual assemblies. The porous microspheres of adenosine nucleolipids are formed first and then the fibers of uridine nucleolipids grow on these spheres. This order of self-sorting is the likely reason for the inability of uridine nucleolipids to form gels in the presence of adenosine nucleolipid. ^{1}H NMR experiment with different mixtures of the nucleolipids did not show any changes in the ^{1}H NMR signal of individual nucleolipids (Figure 18). This indicates that there is no interaction between the nucleolipids at the molecular level.

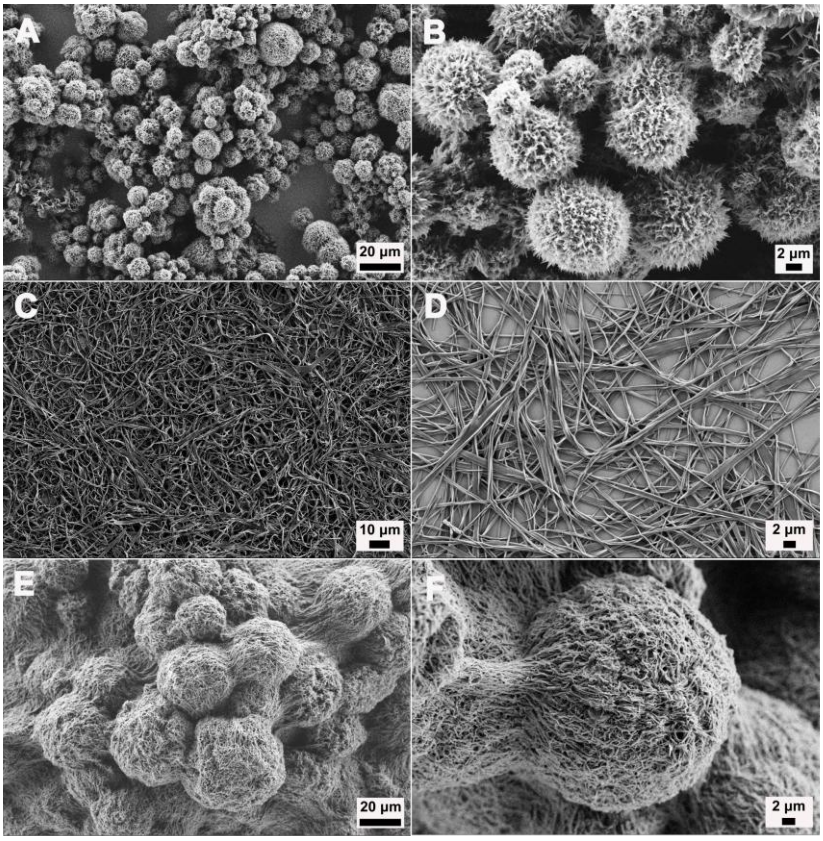


Figure 17. FESEM images of adenosine and uridine based nucleolipids. (**A** and **B**) Compound **4c** in DMSO (0.7 w/v %), (**C** and **D**) xerogel of **7d** (0.7 w/v %) and (**E** and **F**) **7d** (0.7 w/v %) with **4c** (0.7 w/v %) 1:1 equiv. in DMSO.

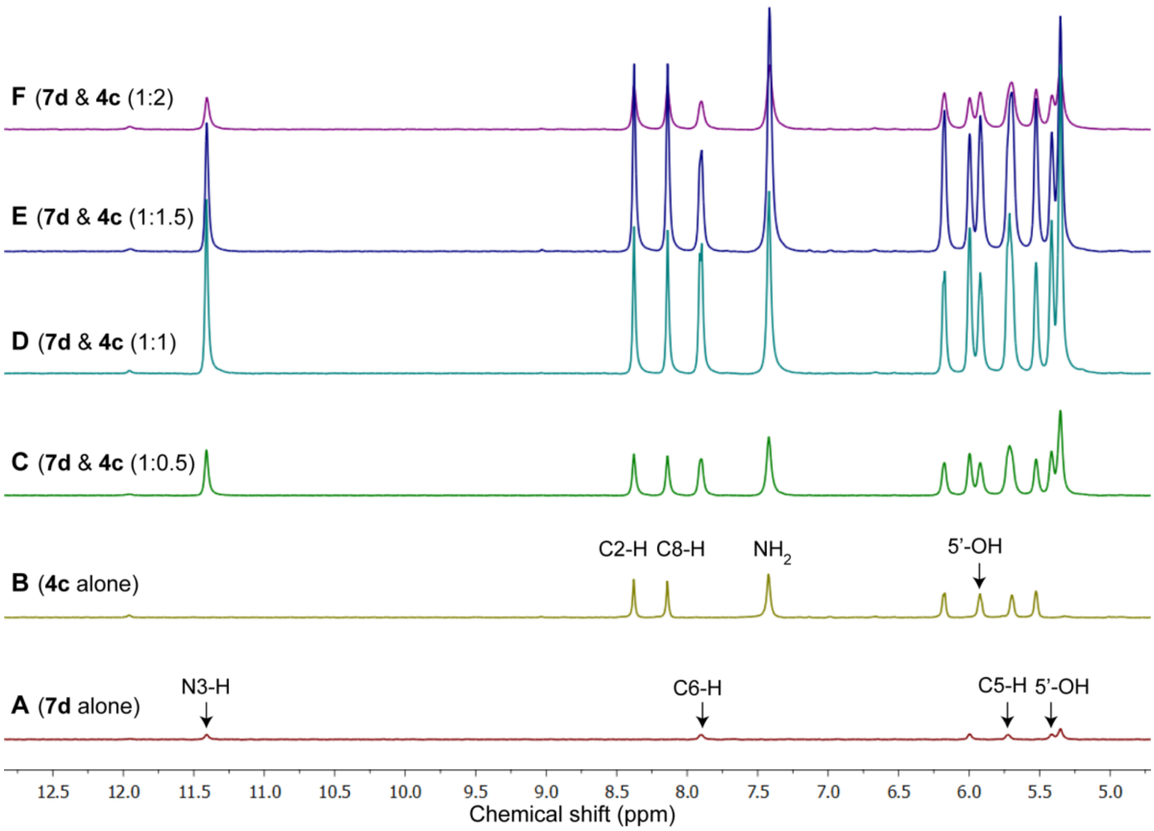


Figure 18. Partial ^{1}H NMR spectra of nucleolipid gel of **7d** (0.7 w/v %) (**A**) and **4c** (0.7 w/v %) (**B**) in *d6*-DMSO an increase stoichiometric ratios of **4c**. The nucleolipids **7d** and **4c** ratios are 1:0.5 (**C**), 1:1 (**D**), 1:1.5 (**E**) and 1:2 (**F**). Here, H-bonding involving and aromatic protons chemical shift values there are no significant changes were observed. It implies **7d** gel not interacts with **4c** because it was already self-sorted system rather than co-assembly.

4B.2.6.2 Uridine nucleolipid and adenosine mixtures form self-sorted assemblies

Uridine nucleolipids formed organogels in the presence of adenosine, albeit with reduced strength.[23] Rheological measurements revealed a significant decrease in storage modulus (G′) of **7d** gel in the presence of 1 equivalent of adenosine (2483 Pa to 1606 Pa). Notably, addition of other nucleosides (guanosine, cytidine and uridine) had only minor effect on the gelation process and strength (Figure 19).[24]

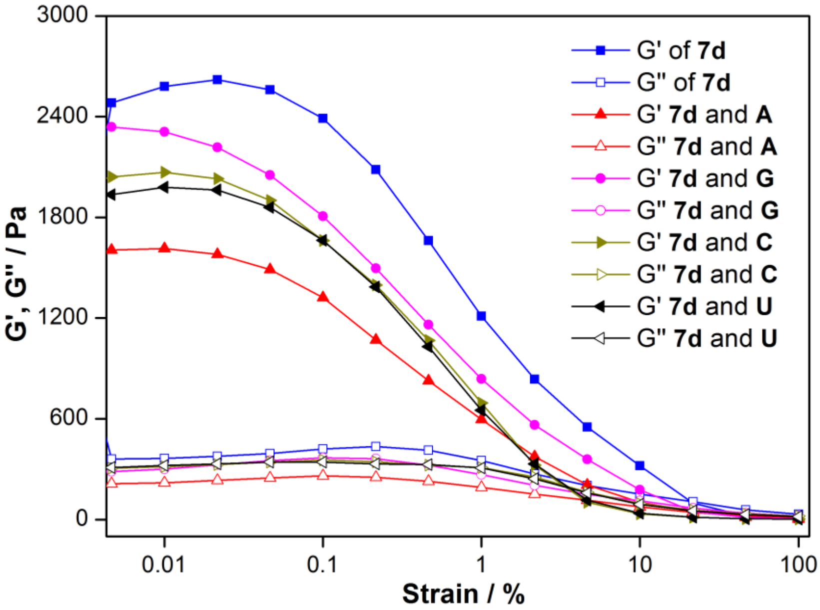


Figure 19. Comparison of strain sweep rheological measurements of nucleolipid organogel **7d** (0.7 w/v %) with nucleosides 1:1 equiv. (A = adenosine, G = guanosine, C = cytidine and U = uridine) at respective CGC in DMSO at constant oscillating frequency.

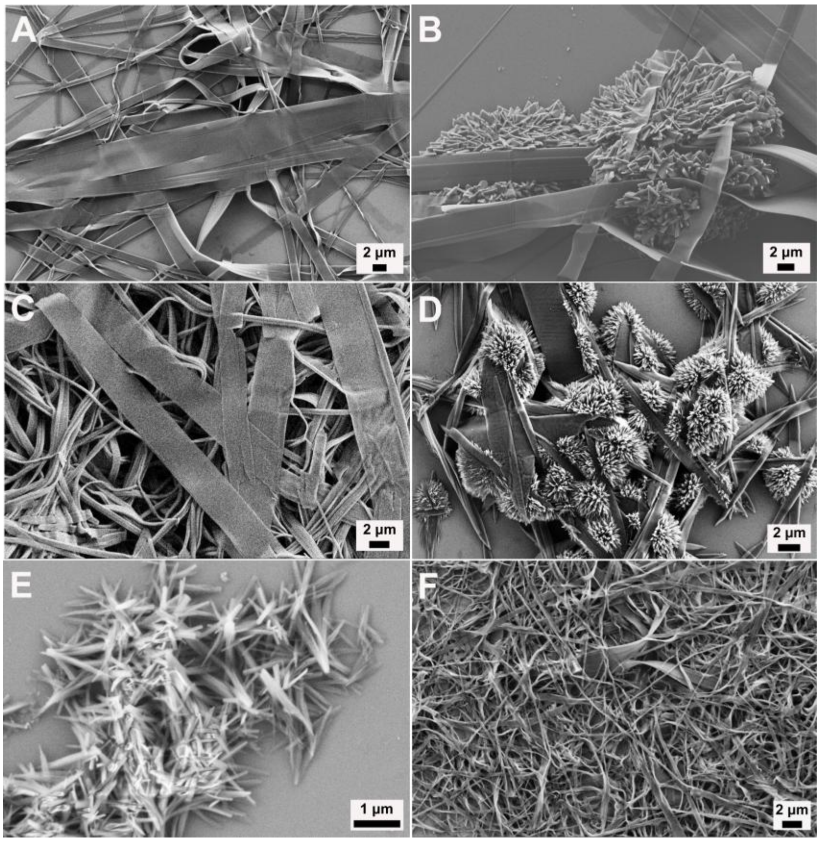
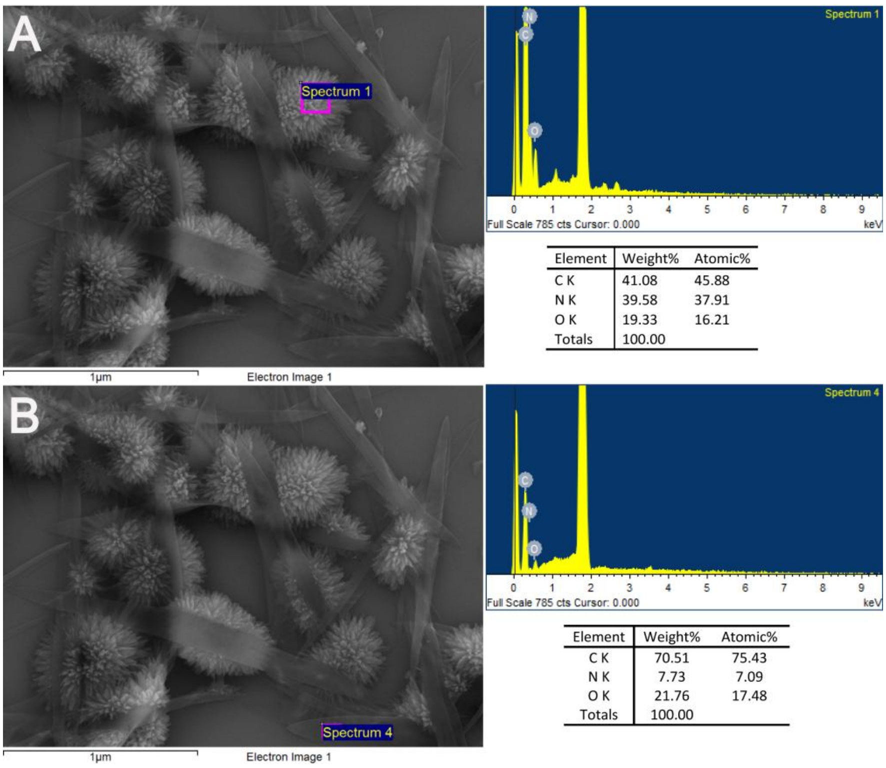


Figure 20. FESEM images of xerogels of nucleolipids and adenosine. (**A**) Compound **7c** (1.2 w/v %, 18 mM), (**B**) **7c** with adenosine (1:2 equiv. 36 mM), (**C**) **7d** (0.7 w/v %, 9.7 mM), (**D**) **7d** with adenosine (1:2 equiv. 19.4 mM), (**E**) adenosine (19.4 mM) and (**F**) **7d** with **7c** (1:1 equiv. 9.7 mM) in DMSO.

Interestingly, SEM images of xerogels of uridine nucleolipids containing adenosine showed aggregation of adenosine in the form nanoclusters on the sheets formed by the nucleolipid (Figure 20). EDAX analysis confirmed that the clusters are made of adenosine (Figure 21). SEM images also indicated that entangled sheets and fibres of uridine nucleolipids, which are responsible for gelling solvents, formed first. Subsequently, these assemblies *in situ* served as a microscopic scaffold for the assembling of adenosine clusters. This observation is intriguing given that there is no interaction between the uridine nucleolipid and adenosine at the molecular level as confirmed by ^{1}H NMR analysis. ^{1}H NMR of a gel made of uridine nucleolipid and adenosine indicate no change in the proton signal of the individual components (Figure 22). Collectively, these observations indicate that self-sorting of assemblies happened due to interactions at the microscopic level rather than at the molecular level. In control experiments, addition of guanosine, cytidine and uridine did not change the morphology of **7d**.



Element	Weight%	Atomic%
C K	41.08	45.88
N K	39.58	37.91
O K	19.33	16.21
Totals	100.00	

Element	Weight%	Atomic%
C K	70.51	75.43
N K	7.73	7.09
O K	21.76	17.48
Totals	100.00	

Figure 21. EDAX analysis of supramolecular organogel of **7d** and adenosine (1:2 equiv) hybride, adenosine aggregates are clustered on top of the uridine nucleolipid sheet. The clustered elemental composition matches to adenosine moiety (**A**) and sheet elemental composition matches the nucleolipid **7d** (**B**), respectively. It implies that nucleolipids (**7d**) and adenosine self-sorting rather than co-assembly.

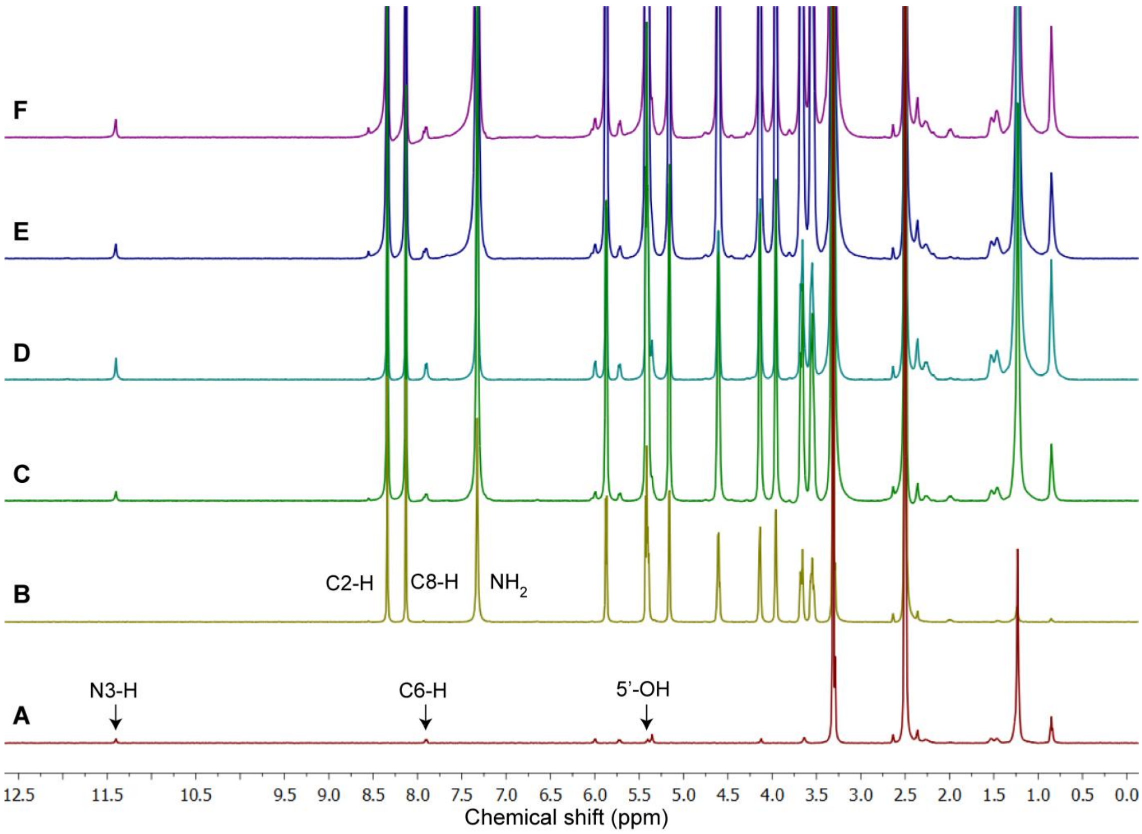


Figure 22. ^{1}H NMR spectra of nucleolipid gels of **7d** (0.7 w/v %) (**A**) and adenosine (0.7 w/v %) (**B**) in *d6*-DMSO an increase stoichiometric ratios of adenosine. The nucleolipids **7d** and adenosine ratios are 1:0.5 (**C**), 1:1 (**D**), 1:1.5 (**E**) and 1:2 (**F**). Here, H-bonding involving and aromatic protons chemical shift, no appreciable changes were observed. It implies **7d** gel not interacts with adenosine in gel state because it was already self-sorted system.

4B.2.6.3 Order/kinetics of self-sorting influences gelation ability: An important aspect of these self-sorted assemblies is the difference in the order/kinetics of formation of individual assemblies. In a mixture of uridine and adenosine nucleolipids, microspheres of adenosine nucleolipid, which does not support gelation, formed first and abrogated the gelling ability of uridine nucleolipid. In case of a mixture of uridine nucleolipid and adenosine, a highly entangled network of uridine nucleolipid, which supports gelation, formed first and then facilitated the growth of adenosine. This order of self-sorting did not adversely affect the gelation ability of the nucleolipid.

4B.2.6.4 Co-assembly of uridine nucleolipids of different fatty acid acyl chain length enhances the gel strength: Having studied the self-sorting behavior of uridine nucleolipids in the presence of complementary nucleoside and nucleolipid, we next sought to study the self-assembling process of a two component system composed of uridine nucleolipids of different fatty acid acyl chain lengths. A 1:1 mixture of myristoyl-containing nucleolipid **7c** and palmytoyl-containing nucleolipid **7d** (9.7 mM, at CGC of **7d**) formed a very strong gel with a remarkably higher storage modulus (G′, 14584 Pa) as compared to individual gels of **7c** (1721 Pa) and **7d** (2483 Pa) (Figure 23).[6b]

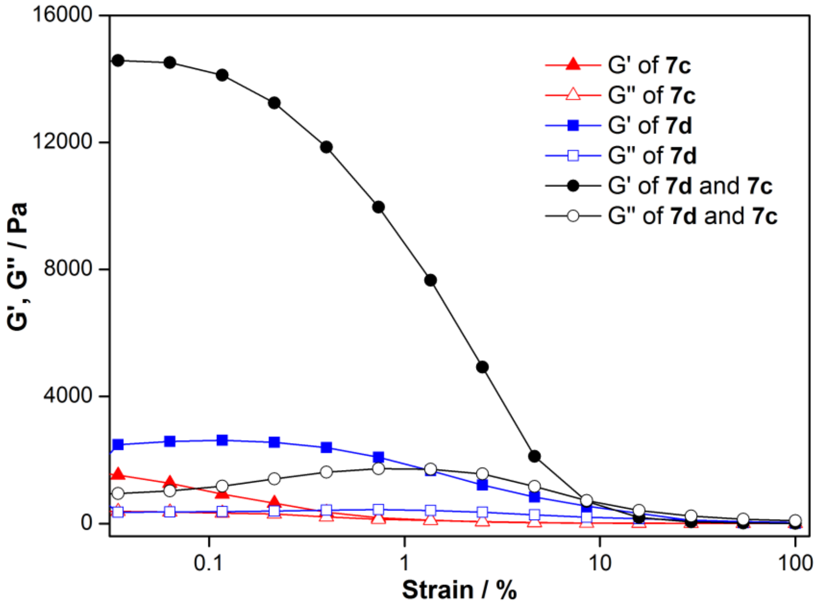


Figure 23. Comparison of strain sweep rheological measurements of nucleolipid organogel of **7c** (1.2 w/v %), **7d** (0.7 w/v %) and **7d** with 1:1 equiv of **7c** were performed at respective CGC in DMSO at constant oscillating frequency.

SEM analysis of this mixture indicated the formation of a highly entangled fibrous network (Figure 20F). A comparison of PXRD spectrum of **7c**, **7d** and a mixture of **7c–7d** suggested the existence of a co-assembly. PXRD analysis of gels of **7c** and **7d** showed that the basic unit of the self-assembly is a bilayer made of two molecules of nucleolipid arranged in a head to head fashion (vide supra, Figure 12 and 13). The interplanar distance corresponding to the basic unit was found to be 4.68 nm and 5.16 nm, respectively. However, the xerogel of a mixture of **7c–7d** gave an interplanar distance of 4.85 nm, which did not match the length of the basic bilayer unit of pure gels. An interplanar distance of 4.85 nm is more close to the bilayer unit formed by a co-assembly **7c** and **7d** (Figure 24). Collectively, these results underscore the potential of a co-assembly of nucleolipids composed of different alkyl chain lengths in producing gels of very high strength as compared to gels formed by using individual gelators.

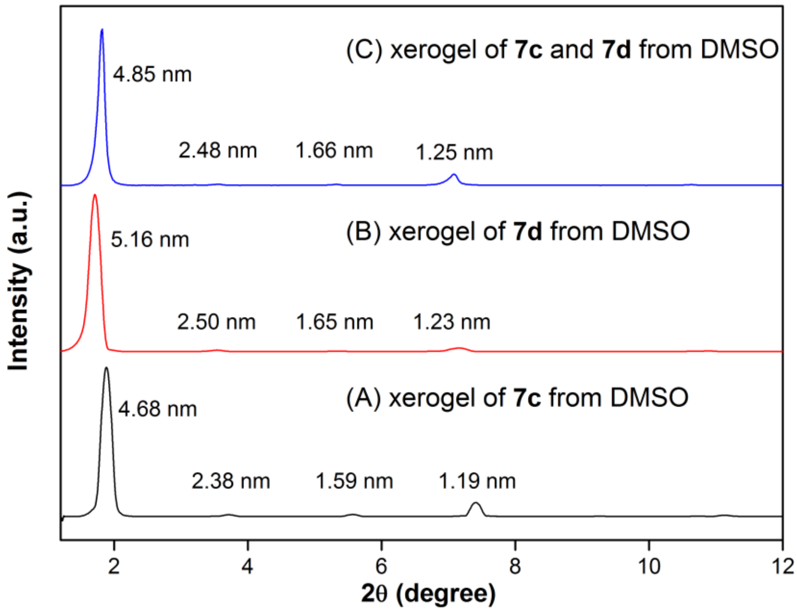


Figure 24. Comparison of PXRD spectra of xerogels of **7c** (A) **7d** (B) and **7c**–**7d** with 1:1 equiv. (C) obtained from DMSO at CGC. Layer spacing (nm) for prominent diffraction peaks.

4B.2.7 Multi stimuli-responsiveness

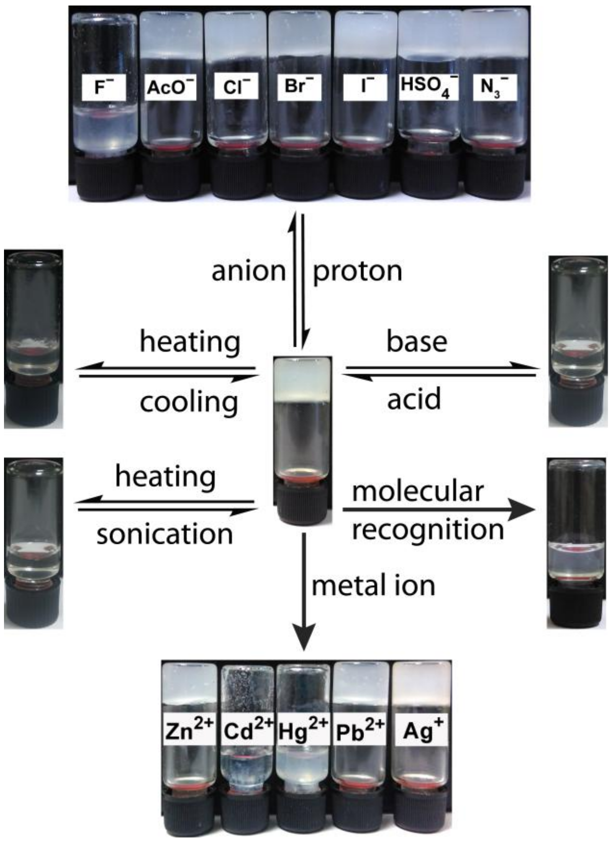


Figure 25. Photograph of multi stimuli responsiveness for uridine based nucleolipid **7d** (0.7 w/v %) gel to physical (temperature and sonication) and chemical stimuli (anion, acid-base, molecular recognition and metal ion).

As evident from diffraction data and ^{1}H NMR analysis, the self-assembling process in uridine nucleolipids is mainly driven by a cascade of H-bonding. Hence, chemical or physical stimuli that can disrupt these H-bonding interactions can affect the self-assembling as well as the gelation behavior. In order to study the responsiveness of nucleolipids to various chemical and physical stimuli, we chose nucleolipid **7d**, which has low CGC values (Figure 25). The responsiveness of gel of **7d** formed in DMSO to anions was studied by using tetrabutylammonium salts of F^-, AcO^-, Cl^-, Br^-, I^-, HSO_4^- and N_3^-. When 0.1 equiv. of fluoride salt was added to preformed gel of **7d**, gel to sol transition occurred. However, addition of same amount or higher equivalents (5 equiv.) of other anions, the gelation ability of **7d** was preserved (see the top panel Figure 25). These results clearly demonstrated that the gelation of **7d** is specifically affected by F^- anion. The addition of 2 equiv. of F^- to **7d** gel resulted in the disappearance of N3H proton and emergence of a new triplet at 16.16 ppm corresponding to HF_2^-.[25] This result indicates the effective deprotonation of N3H by fluoride ion, thereby disrupting the key H-bonding partner required for the self-assembling and gelation process (Figure 26). Interestingly, the gelation ability of **7d** was restored upon addition of proton source (1 equiv. acetic acid). Similarly, gel-sol transition was reversible upon addition of a base and acid.[26] It was found that 0.5 equiv of DBU disassembled the gel, and subsequent addition of 0.5 equiv of trichloroacetic acid restored the **7d** gel. DBU induced phase transition mainly due to deprotonation of N3H and 5′-OH of nucleolipid. The addition of 1 equivalent of adenosine nucleolipid (**4c**) into uridine nucleolipid (**7d**) gel to sol transition was observed. This is because of molecular recognition between the microscopic objects of two complement nucleolipids.

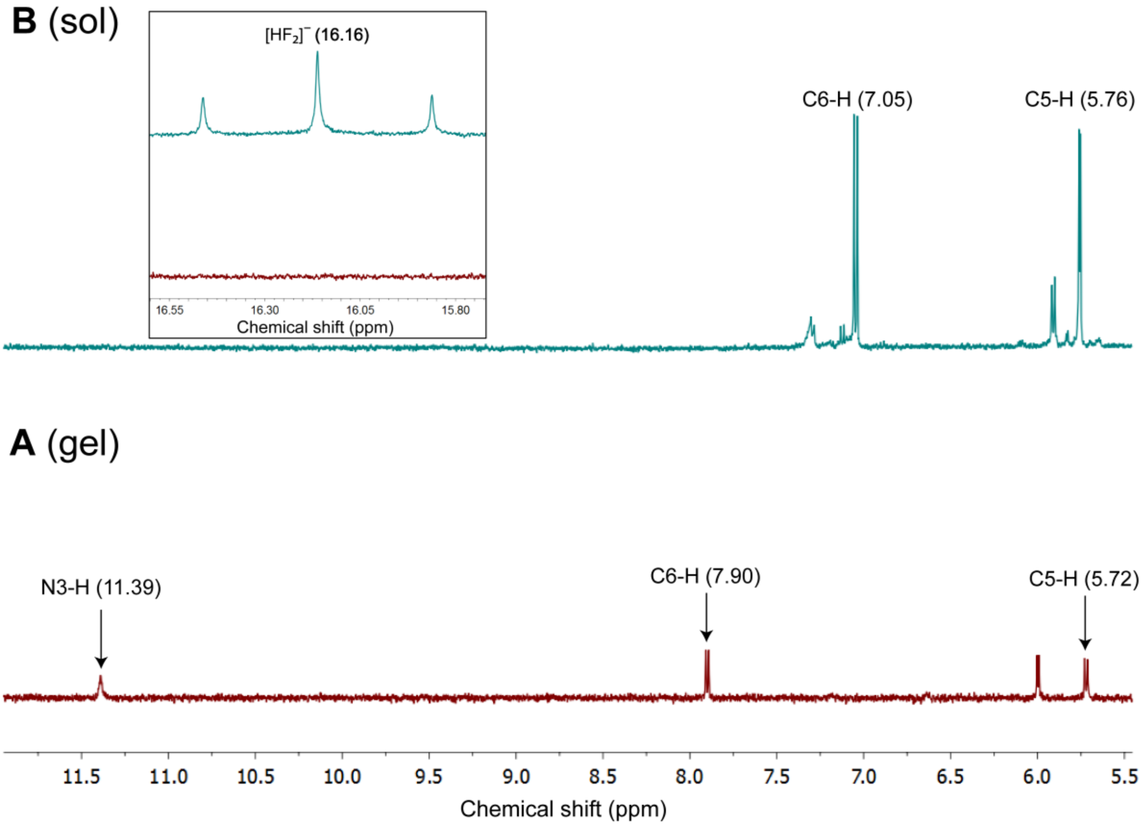


Figure 26. Partial ^{1}H NMR spectrum of nucleolipid **7d** in *d6*-DMSO (0.7 w/v %) in the absence (**A**, gel state) and presence of 2 equiv. of fluoride ion (**B**, sol state). The disappearance of N3-H (11.39 ppm) signal and appearance of a new triplet signal at 16.16 ppm corresponding to HF_2^- confirms the deprotonation of imino hydrogen atoms by fluoride anion. Similarly the C6-H (7.90 ppm) signal shifted towards upfield C6-H (7.05 ppm) when the fluoride ion deprotonate N3-H. ^{1}H NMR spectra was obtained using Bruker 500 MHz spectrometer.

Effect of added metal ions: The effect of added metal ions (Zn^{+2}, Cd^{+2}, Hg^{+2}, Pb^{+2} and Ag^{+}) on the gelation ability of uridine nucleolipids was tested using **7d** as the test example. Among the metal ions tested, Cd^{+2} (0.5 equiv.) and Hg^{+2} (0.1 equiv.) completely disassembled the gel. However, higher amounts (5 equiv.) of the other metals did not affect the gelation ability of **7d** (Figure 25). ^{1}H NMR experiments were carried out to understand the specific role of metal ions in the degelation process. Addition of 1 equivalent of Cd^{+2} ions resulted in the disappearance of N3H and 5′-OH proton signals (Figure 27). Similarly, in the presence of 1 equivalent of a Hg^{+2} ions, 5′-OH proton signal vanished and there was noticeable change in N3H signal (Figure 28). Uridine-Hg-uridine complex formation was further confirmed by MALDI-TOF mass analysis. The mass corresponding to the dimer along with characteristic isotopic pattern of a Hg^{+2} was observed (Figure 29). These observations collectively indicate that the metal ions interact with the key H-bonding sites, which are involved in the gelation process, thereby disrupting the gelation ability of the uridine nucleolipids.[27]

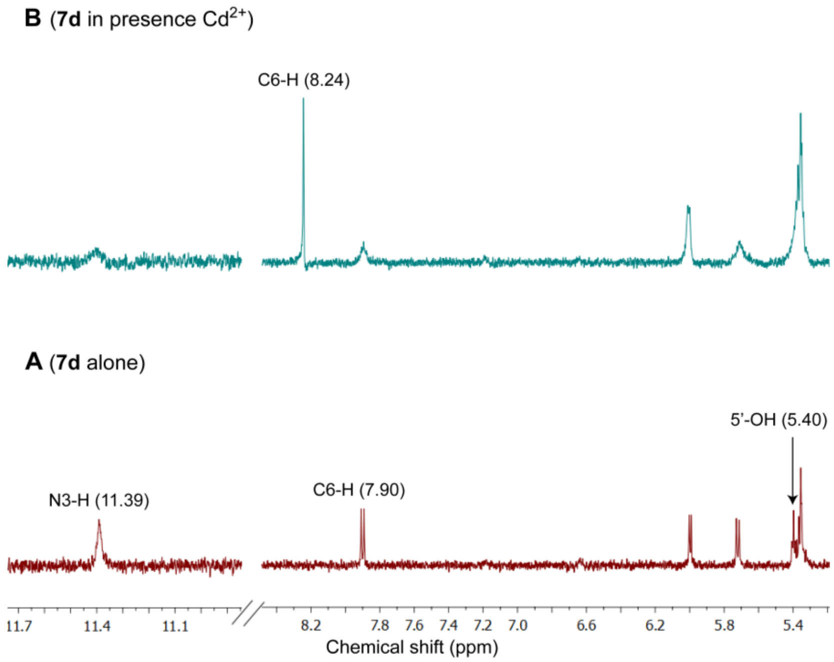


Figure 27. Partial ^{1}H NMR spectrum of nucleolipid gel **7d** in *d6*-DMSO (0.7 w/v %) in the absence (**A**) and presence of 1 equiv. of Cd^{2+} ion (**B**). The disappearance of N3-H (11.39 ppm) and 5′-OH signals (5.40 ppm) and C6-H (7.90 ppm) signal shifted to downfield (8.24 ppm). It indicates that the deprotonation of N3-H and 5′-OH and C6-H downfield due to binding of cadmium to nucleolipid, leads to gel-sol transition. ^{1}H NMR spectra was obtained using Bruker 500 MHz spectrometer.

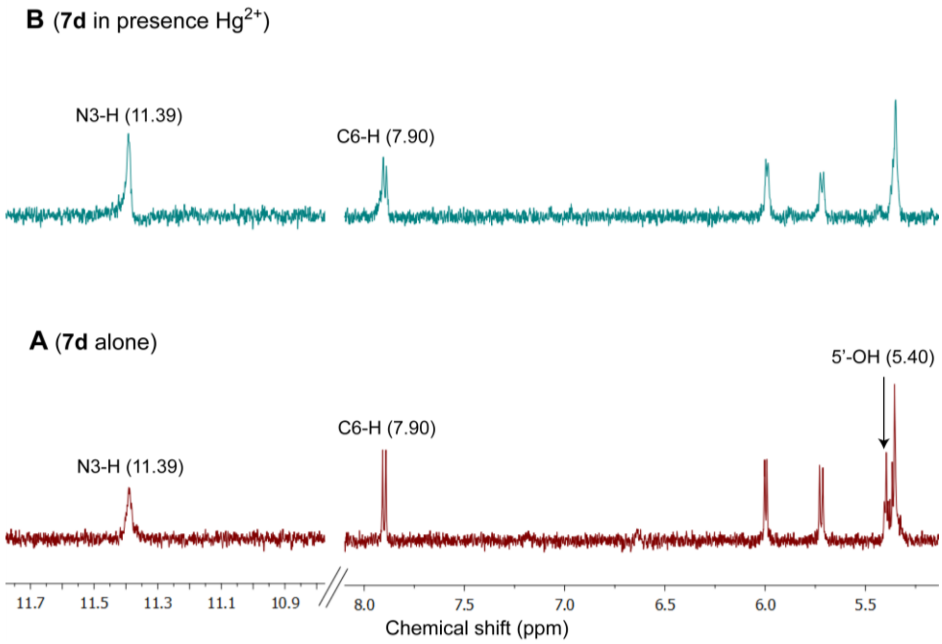


Figure 28. Partial ^{1}H NMR spectrum of nucleolipid gel **7d** in *d6*-DMSO (0.7 w/v %) in the absence (**A**) and presence of 1 equiv. of Hg^{2+} ion (**B**). The disappearance of 5′-OH signal (5.40 ppm) but N3-H (11.39 ppm) intact in both phases. It indicates the deprotonation of 5′-OH due to binding of mercury to nucleolipid, leads to gel-sol transition. The mode of binding of Hg^{2+} is 5′-*O*-Hg-*O*-5′ with respective sugar moiety of nucleolipid **7d**. ^{1}H NMR spectra was obtained using Bruker 500 MHz spectrometer.

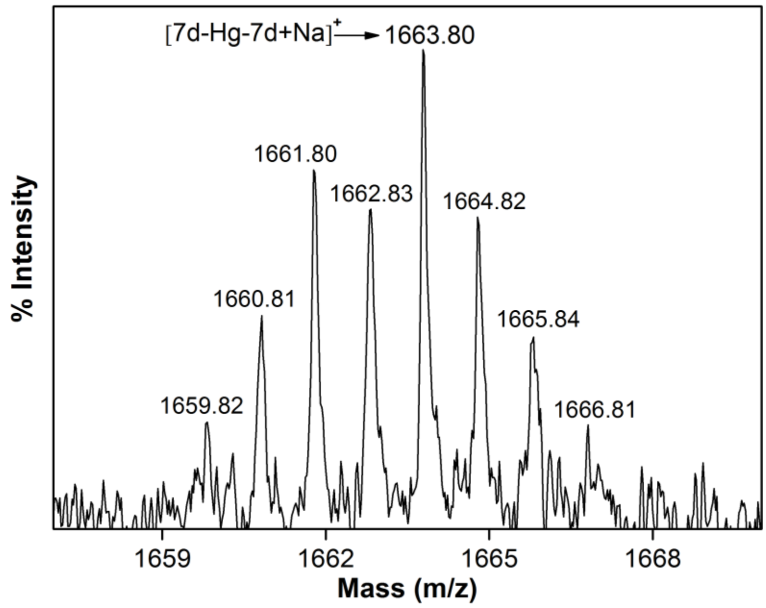


Figure 29. Mass spectrum showing the molecular ion peak of the 7d-Hg-7d base pair complex **(7d–Hg–7d+Na$^+$)** calculated mass 1663.63, found mass 1663.80. Characteristic isotopic pattern supports the presence of a Hg^{2+} ion.

4B.3 Conclusions

Ribonucleoside-fatty acids conjugates made of uridine and adenosine serve as supramolecular synthons with distinct self-assembling behavior. Uridine nucleolipids form gel in various organic solvents, whereas adenosine nucleolipids do not support gelation process. The ability to support and not support gelation is also evident from the respective self-assembling process. Uridine nucleolipids form highly entangled sheet or fibrous network, which is an essential feature of a gelator assembly. Adenosine nucleolipids form microspheres, which is not support an assembly to gel solvents. However, when uridine nucleolipids where combined with complementary adenosine nucleolipids or adenosine or uridine nucleolipids interesting self-assembling phenomena like self-sorting and co-assembly were observed, which were not governed by complementary H-bonding interactions. While addition of adenosine nucleolipids to uridine nucleolipids resulted in degelation, addition of adenosine resulted in the weakening of the uridine nucleolipid gel strength. SEM images, and rheological and ^{1}H NMR analyses indicate that the order of formation of individual assemblies and interactions at the microscopic level drive the assemblies to self-sort. Interestingly, a two component system composed of uridine nucleolipids of different fatty acid acyl chain lengths formed a co-assembly, which exhibited a remarkable increase in the gel strength as compared to the gels formed by the individual components. Collectively, the results presented in this chapter underscore the potential of these simple ribonucleolipid synthons, individually or in combination, in programming the self-assembling process to generate materials with different or improved properties.

4B.4 Experimental section

4B.4.1 Materials

tert-Butyl(chloro)dimethylsilane (TBDMS-Cl) 4,4′-dimethoxytrityl chloride (DMT-Cl), octanoic acid, dodecanoic acid, palmatic acid, stearic acid, oleic acid, tetrabutylammonium salts (fluoride, chloride, bromide, iodide, acetate, sulfate and azide), silver nitrate, zinc nitrate hexahydrate, lead(II) nitrate, cadmium acetate dihydrate, mercury(II) perchlorate hexahydrate, cytidine and dry pyridine were purchased from Sigma-Aldrich. EDC (1-(3-dimethyl aminopropyl)-3-ethyl carbodiimide hydrochloride), 4-dimethylaminopyridine were obtained from Avra synthesis. Myristic acid was procured from Fluka. Adenosine and guanosine were obtained from Sisco Research Laboratories and uridine from Molychem. Silicon wafers (N-type without dopant) were purchased from Sigma-Aldrich. 5′-*O*-DMT-protected uridine **6**[28] was prepared by following the reported procedures.

4B.4.2 Instrumentation

NMR spectra were recorded on 400 MHz Jeol ECS-400 spectrometer and Bruker 500 MHz spectrometer. The morphology of gels was analyzed using Zeiss Ultra Plus field-emission scanning electron microscope (FESEM). Powder X-ray diffraction (PXRD) spectra were obtained at room temperature using Bruker D8 Advance diffractometer (Cu Kα radiation, λ = 1.5406 Å). Rheology measurements were carried out in Anton paar MCR 302 instrument. Single crystal X-ray data for structure determination were collected from Bruker APEX II DUO diffractometer using MoKα (λ= 0.71073 Å) and CuKα (λ = 1.54178 Å) graphite monochromated radiation. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometric analysis was used for a U–Hg–U dimer formation. Mass measurements were recorded on Applied Biosystems 4800 Plus MALDI TOF/TOF analyzer and Water Synapt G2 High Definition mass spectrometers.

4B.4.3 Characterization data for compounds 4a-4d and 7a-7e

4B.4.3.1 Synthesis of 5′-*O*-TBDMS-adenosine (2): The reaction mixture of adenosine and anhydrous DMF (20 mL/g) was elevated the temperature 80 °C, in order to solubilize adenosine. Then TBDMS-Cl and imidazole were added to adenosine suspension, and allowed to stir at RT for 12 h under nitrogen atmosphere. The resultant residue was diluted with dichloromethane and extracted with $NaHCO_3$ and dried over sodium sulfate. The residue was purified by silica gel column chromatography (7% methanol in dichloromethane) to afford the products.

Compound 2: Colourless solid, 80% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.40; ^1H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.29 (s, 1H), 8.14 (s, 1H), 7.31 (br, 2H), 5.91 (d, J = 5.2 Hz, 1H), 5.55 (d, J = 6.0 Hz,1H), 5.21 (d, J = 5.6 Hz, 1H), 4.55–4.51 (m, 1H), 4.19–4.16 (m, 1H), 3.97–3.94 (m, 1H), 3.88 (dd, J_1 = 11.4 Hz, J_2 = 3.8 Hz, 1H), 3.75 (dd, J_1 = 11.2 Hz, J_2 = 4.0 Hz, 1H), 0.88–0.86 (m, 9H), 0.05–0.04 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 156.0, 152.7, 149.4, 138.9, 119.0, 87.4, 84.4, 73.7, 69.9, 62.9, 25.8, 18.1, -5.4, -5.4; HRMS: m/z Calcd. for $C_{16}H_{28}N_5O_4Si$ $[M+H]^+$ = 382.1910, found = 382.1910.

4B.4.3.2 General procedure for synthesis of 2′,3′-*O*-disubstituted adenosine nucleolipids (4a–4d).

5′-*O*-TBDMS-adenosine **2** (1.0 equiv.), fatty acid (dodecanoic acid, myristic acid, palmitic acid and stearic acid, 2.0 equiv.), EDC (2.0 equiv.), and DMAP (0.2 equiv.) were dissolved in anhydrous pyridine (10 mL/g) under nitrogen atmosphere at RT, allowed to stir overnight. The crude residue was diluted with dichloromethane and extracted with water and dried over sodium sulfate. The resultant product was purified by silica gel column chromatography (2% methanol in dichloromethane) to afford the product.

The 5′-*O*-TBDMS-adenosine nucleolipid (**3a–3d**) was dissolved in anhydrous THF and TBAF (1M, THF) added drop wise manner under nitrogen atmosphere and stirred at RT for 30 min. The crude residue was diluted with dichloromethane and extracted with water and dried over sodium sulfate. The crude product was purified by silica gel column chromatography (2.5% methanol in dichloromethane) to obtained desire product.

Compound 3a: Colourless solid, 82% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.48; ^1H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.31 (br, 1H), 8.24 (br, 1H), 6.34 (d, J = 6.8 Hz, 1H), 6.28–6.26 (br, 2H), 5.74–5.71 (m, 1H), 5.54–5.52 (m, 1H), 4.31–4.26 (m, 1H), 3.95 (dd, J_1 = 11.4 Hz, J_2 = 2.2 Hz, 1H), 3.89 (dd, J_1 = 11.2 Hz, J_2 = 2.0 Hz, 1H), 2.37 (t, J = 7.4 Hz, 2H), 2.26 (t, J = 7.6 Hz, 2H), 1.69–1.62 (m, 2H), 1.56–1.49 (m, 2H), 1.26–1.20 (m, 32H), 0.96 (s, 9H), 0.89–0.85 (m, 6H), 0.15 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.8, 172.3, 155.7, 153.1, 150.2, 138.6, 119.5, 85.1, 84.5, 74.3, 71.8, 63.3, 34.2, 33.9, 32.1, 29.8, 29.8, 29.6, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 26.1, 25.1, 25.0, 24.8, 22.8, 18.6, 14.2, -5.3; HRMS: m/z Calcd. for $C_{40}H_{72}N_5O_6Si$ $[M+H]^+$ = 746.5252, found = 746.5242.

Compound 3b: Colourless solid, 88% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.32; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.34 (s, 1H), 8.22 (s, 1H), 6.35 (d, J = 7.2 Hz, 1H), 5.92 (br, 2H), 5.73 (dd, J_1 = 7.0 Hz, J_2 = 5.4 Hz, 1H), 5.54 (dd, J_1 = 5.2 Hz, J_2 = 2.4 Hz, 1H), 4.29–4.27 (m, 1H), 3.95 (dd, J_1 = 11.4 Hz, J_2 = 2.2 Hz, 1H), 3.89 (dd, J_1 = 11.6 Hz, J_2 = 2.2 Hz, 1H), 2.37 (t, J = 7.4 Hz, 2H), 2.26 (t, J = 7.6 Hz, 2H), 1.68–1.60 (m, 2H), 1.55–1.47 (m, 2H), 1.29–1.20 (m, 40H), 0.96 (s, 9H), 0.89–0.85 (m, 6H), 0.15 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.8, 172.3, 155.6, 153.3, 150.3, 138.6, 119.7, 85.0, 84.4, 74.2, 71.8, 63.3, 34.2, 33.9, 32.1, 29.8, 29.8, 29.8, 29.6, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 26.1, 25.0, 24.8, 22.8, 18.6, 14.3, -5.3, -5.3; HRMS: m/z Calcd. for $C_{44}H_{80}N_5O_6Si$ $[M+H]^+$ = 802.5878, found = 802.5880.

Compound 3c: Colourless solid, 70% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.41; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.32 (s, 1H), 8.24 (s, 1H), 6.35 (d, J = 6.8 Hz, 1H), 6.16 (br, 2H), 5.73 (dd, J_1 = 6.8 Hz, J_2 = 5.2 Hz, 1H), 5.53 (dd, J_1 = 5.4 Hz, J_2 = 2.2 Hz, 1H), 4.29–4.28 (m, 1H), 3.95 (dd, J_1 = 11.4 Hz, J_2 = 2.2 Hz, 1H), 3.89 (dd, J_1 = 11.6 Hz, J_2 = 2.0 Hz, 1H), 2.37 (t, J = 7.4 Hz, 2H), 2.26 (t, J = 7.6 Hz, 2H), 1.67–1.60 (m, 2H), 1.55–1.47 (m, 2H), 1.29–1.20 (m, 48H), 0.96 (s, 9H), 0.89–0.86 (m, 6H), 0.15 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.9, 172.3, 155.7, 153.2, 150.2, 138.6, 119.5, 85.0, 84.4, 74.3, 71.8, 63.3, 34.2, 33.9, 32.1, 29.9, 29.8, 29.8, 29.7, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 26.1, 25.1, 25.0, 24.8, 22.8, 18.6, 14.3, -5.3; HRMS: m/z Calcd. for $C_{48}H_{88}N_5O_6Si$ $[M+H]^+$ = 858.6505, found = 858.6500.

Compound 3d: Colourless solid, 72% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.42; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.35 (s, 1H), 8.22 (s, 1H), 6.35 (d, J = 6.8 Hz, 1H), 5.83 (br, 2H), 5.73 (dd, J_1 = 6.8 Hz, J_2 = 5.2 Hz, 1H), 5.54 (dd, J_1 = 5.2 Hz, J_2 = 2.4 Hz, 1H), 4.29–4.27 (m, 1H), 3.95 (dd, J_1 = 11.2 Hz, J_2 = 2.4 Hz, 1H), 3.89 (dd, J_1 = 11.6 Hz, J_2 = 2.4 Hz, 1H), 2.37 (t, J = 7.4 Hz, 2H), 2.26 (t, J = 7.6 Hz, 2H), 1.68–1.60 (m, 2H), 1.55–1.49 (m, 2H), 1.31–1.20 (m, 56H), 0.96 (s, 9H), 0.89–0.86 (m, 6H), 0.15 (s, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.8, 172.3, 155.6, 153.4, 150.4, 138.7, 119.8, 85.0, 84.4, 74.2, 71.8, 63.3, 34.2, 33.9, 32.1, 29.9, 29.8, 29.8, 29.7, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 26.1, 25.0, 24.8, 22.8, 18.6, 14.3, -5.3, -5.3; HRMS: m/z Calcd. for $C_{52}H_{96}N_5O_6$ Si $[M+H]^+$ = 914.7130, found = 914.7141.

Compound 4a: Colourless solid, 73% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.48; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.33 (s, 1H), 7.81 (s, 1H), 6.62 (br, 1H), 6.05–6.03 (br, 1H), 6.01–6.00 (m, 2H), 5.70–5.69 (m, 1H), 4.34 (br, 1H), 3.98 (dd, J_1 = 13.0 Hz, J_2 = 1.4 Hz, 1H), 3.87–3.84 (m, 1H), 2.39 (t, J = 7.6 Hz, 2H), 2.23 (t, J = 7.6 Hz, 2H), 1.69–1.62 (m, 2H), 1.56–1.48 (m, 2H), 1.26–1.22 (m, 32H), 0.89–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.6, 171.9, 156.2, 152.9, 148.8, 140.3, 121.3, 88.9, 86.8, 72.8, 72.8, 62.8, 34.3, 33.8, 32.1, 29.8, 29.8, 29.7, 29.7, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 25.1, 24.8, 22.8, 14.2; HRMS: m/z Calcd. for $C_{34}H_{58}N_5O_6$ $[M+H]^+$ = 632.4387, found = 632.4377.

Compound 4b: Colourless solid, 83% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.36; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.34 (br, 1H), 7.81 (s, 1H), 6.62 (d, J = 10.8 Hz, 1H), 6.01–6.00 (m, 1H), 5.95–5.80 (br, 2H), 5.70 (dd, J_1 = 4.4 Hz, J_2 = 0.8 Hz, 1H), 4.34 (br, 1H), 3.99–3.96 (m, 1H), 3.89–3.83 (m, 1H), 2.39 (t, J = 7.4 Hz, 2 H), 2.23 (t, J = 7.6 Hz, 2H), 1.70–1.62 (m, 2H), 1.58–1.47 (m, 2H), 1.32–1.22 (m, 40H), 0.89–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.6, 171.9, 156.1, 152.9, 148.9, 140.3, 121.4, 100.1, 88.9, 86.8, 72.8, 72.8, 62.8, 34.3, 33.8, 32.1, 29.8, 29.8, 29.8, 29.8, 29.7, 29.7, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 25.1, 24.8, 22.8, 14.2; HRMS: m/z Calcd. for $C_{38}H_{66}N_5O_6$ $[M+H]^+$ = 688.5013, found = 688.5002.

Compound 4c: Colourless solid, 75% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.37; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.35 (br, 1H), 7.81 (s, 1H), 6.59 (d, J = 11.2 Hz, 1H), 6.01–6.00 (m, 1H), 5.85–5.73 (br, 2H), 5.70 (dd, J_1 = 4.6 Hz, J_2 = 0.6 Hz, 1H), 4.34 (br, 1H), 4.00–3.97 (m, 1H), 3.89–3.83 (m, 1H), 2.39 (t, J = 7.6 Hz, 2H), 2.23 (t, J = 7.6 Hz, 2H), 1.68–1.63 (m, 2H), 1.56–1.47 (m, 2H), 1.31–1.22 (m, 48H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.6, 171.9, 156.1, 152.9, 148.9, 140.3, 121.4, 88.9, 86.9, 72.8, 72.8, 62.8, 34.3, 33.8, 32.1, 29.9, 29.8, 29.8, 29.7, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 25.1, 24.8, 22.8, 14.3; HRMS: m/z Calcd. for $C_{42}H_{74}N_5O_6$ $[M+H]^+$ = 744.5639, found = 744.5622.

Compound 4d: Colourless solid, 87% yield. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.38; ^1H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.35 (s, 1H), 7.80(s, 1H), 6.61 (d, *J* = 11.2 Hz, 1H), 6.03–5.98 (m, 2H), 5,71–5.70 (m, 2H), 4.34 (br, 1H), 4.00–3.97 (m, 1H), 3.89–3.83 (m, 1H), 2.39 (t, *J* = 7.6 Hz, 2H), 2.23 (t, *J* = 7.4 Hz, 2H), 1.68–1.60 (m, 2H), 1.56–1.48 (m, 2H), 1.29–1.23 (m, 56H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.6, 171.9, 156.1, 153.0, 148.9, 140.3, 121.4, 100.2, 88.9, 86.9, 72.8, 62.8, 34.3, 33.8, 32.1, 29.9, 29.8, 29.8, 29.7, 29.6, 29.5, 29.5, 29.4, 29.4, 29.2, 25.1, 24.8, 22.8, 14.3; HRMS: m/z Calcd. for $C_{46}H_{82}N_5O_6$ $[M+H]^+$ = 800.6265, found = 800.6265.

4B.4.3.3 General procedure for synthesis of 2′,3′-*O*-disubstituted uridine nucleolipids (7a–7e).

5′-*O*-DMT–protected uridine **6**[28] (1.0 equiv), fatty acid (octanoic acid, dodecanoic acid, myristic acid, palmitic acid and oleic acid, 2.1 equiv), EDC (2.1 equiv), and DMAP (0.3 equiv) were dissolved in anhydrous dichloromethane (10 mL/g of **6**). The reaction mixture was stirred for 12 h at room temperature under nitrogen atmosphere. After completion of the reaction, the crude product was diluted with dichloromethane and extracted with saturated NH_4Cl solution and dried over sodium sulfate. The product was treated with 3% trichloroacetic acid (TCA) in dichloromethane for 30 min at RT to remove DMT group. The solvent was evaporated, and the resultant residue was purified by silica gel column chromatography (40% ethyl acetate in hexane) to get desired product.

Compound 7a: Colourless solid, 73% yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.29; ^1H NMR (400 MHz, $CDCl_3$): δ (ppm) 9.04 (br, 1H), 7.75 (d, *J* = 8.0 Hz, 1H), 6.06–6.03 (m, 1H), 5.78 (d, *J* = 8.0 Hz, 1H), 5.48–5.45 (m, 2H), 4.19–4.18 (m, 1H), 3.92 (dd, J_1 = 12.2 Hz, J_2 = 2.2 Hz, 1H), 3.86 (dd, J_1 = 12.4 Hz, J_2 = 2.0 Hz, 1H), 2.38–2.29 (m, 4H), 1.66–1.54 (m, 4H), 1.30–1.25 (m, 16H), 0.90–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.1, 172.7, 163.2, 150.5, 140.9, 103.3, 87.9, 83.7, 73.0, 71.1, 62.0, 34.1, 33.9, 31.8, 29.2, 29.1, 29.1, 29.0, 25.0, 24.8, 22.7, 14.2; HRMS: m/z Calcd. for $C_{25}H_{41}N_2O_8$ $[M+H]^+$ = 497.2863, found = 497.2855.

Compound 7b: Colourless solid, 78% yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.42; ^1H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.78 (br, 1H), 7.72 (d, *J* = 8.4 Hz, 1H), 6.05–6.01 (m, 1H), 5.78 (dd, J_1 = 8.2 Hz, J_2 = 2.2 Hz, 1H), 5.49–5.45 (m, 2H), 4.20–4.18 (m, 1H), 3.95 (dd, J_1 = 12.0 Hz, J_2 = 2.0 Hz, 1H), 3.86 (dd, J_1 = 12.0 Hz, J_2 = 2.4 Hz, 1H), 2.38–2.29 (m, 4H), 1.66–1.54 (m, 4H), 1.29–1.25 (m, 32H), 0.89–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.1, 172.7, 163.0, 150.4, 140.9, 103.4, 88.0, 83.7, 73.0, 71.0, 62.0, 34.1, 33.9, 32.1, 29.8, 29.6, 29.5, 29.4, 29.3, 29.2, 25.0, 24.8, 22.8, 14.3; HRMS: m/z Calcd. for $C_{33}H_{57}N_2O_8$ $[M+H]^+$ = 609.4115, found = 609.4107.

Compound 7c: Colourless solid, 81% yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.23; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.82 (br, 1H), 7.72 (d, J = 8.0 Hz, 1H), 6.05–6.01 (m, 1H), 5.78 (d, J = 8.0 Hz, 1H), 5.48–5.45 (m, 2H), 4.19–4.18 (m, 1H), 3.95 (dd, J_1 = 11.8 Hz, J_2 = 1.8 Hz, 1H), 3.86 (dd, J_1 = 12.0 Hz, J_2 = 2.0 Hz, 1H), 2.38–2.29 (m, 4H), 1.66–1.54 (m, 4H), 1.29–1.25 (m, 40 H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.0, 172.6, 162.9, 150.4, 140.8, 103.4, 88.1, 83.7, 73.0, 71.0, 62.0, 34.2, 34.0, 32.1, 29.8, 29.8, 29.8, 29.6, 29.5, 29.4, 29.4, 29.3, 29.2, 25.0, 24.9, 22.8, 14.2; HRMS: m/z Calcd. for $C_{37}H_{65}N_2O_8$ $[M+H]^+$ = 665.4741, found = 665.4729.

Compound 7d: Colourless solid, 70% yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.40; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.41 (br, 1H), 7.71 (d, J = 8.0 Hz, 1H), 6.04–6.01 (m, 1H), 5.77 (dd, J = 8.0 Hz, 1H), 5.48–5.45 (m, 2H), 4.20–4.19 (m, 1H), 3.96 (dd, J_1 = 12.0 Hz, J_2 = 2.0 Hz, 1H), 3.86 (dd, J_1 = 12.2 Hz, J_2 = 2.2 Hz, 1H), 2.38–2.30 (m, 4H), 1.64–1.55 (m, 4H), 1.30–1.25 (m, 48H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.1, 172.6, 162.7, 150.3, 140.7, 103.4, 88.1, 83.6, 73.0, 71.0, 62.0, 34.2, 33.9, 32.1, 29.9, 29.8, 29.8, 29.6, 29.5, 29.4, 29.3, 29.2, 25.0, 24.9, 22.8, 14.3; HRMS: m/z Calcd. for $C_{41}H_{73}N_2O_8$ $[M+H]^+$ = 721.5367, found = 721.5358.

Compound 7e: Colourless solid, 74% yield over two steps. TLC (CH_2Cl_2:MeOH = 95:5); R_f = 0.38; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.79 (br, 1H), 7.73 (d, J = 8.4 Hz, 1H), 6.04–6.02 (m, 1H), 5.78 (d, J = 8.0 Hz, 1H), 5.47–5.46 (m, 2H), 5.38–5.29 (m, 4H), 4.19–4.18 (m, 1H), 3.95 (dd, J_1 = 12.0 Hz, J_2 = 2.2 Hz, 1H), 3.86 (dd, J_1 = 12.2 Hz, J_2 = 2.0 Hz, 1H),2.38–2.26 (m, 8H), 2.01–1.99 (m, 4H), 1.62–1.58 (m, 4H), 1.30–1.26 (m, 40 H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.0, 172.7, 163.1, 159.5, 150.4, 140.9, 130.2, 129.8, 103.3, 88.0, 83.7, 73.1, 71.0, 62.0, 34.1, 33.9, 32.1, 29.9, 29.9, 29.7, 29.5, 29.3, 29.3, 29.2, 27.4, 27.3, 25.0, 24.8, 22.8, 14.2; HRMS: m/z Calcd. for $C_{45}H_{76}N_2O_8Na$ $[M+Na]^+$ = 795.5499, found = 795.5486.

4B.4.4 Gelation test by the inverted vial method: Organogels of nucleolipids (**7a–7e**) a weighed amount of the 2′,3′-difatty acid substituted nucleolipids was dissolved in various organic solvents by elevating temperature. The sample was allowed to RT, where a sturdy opaque organogelation occurred. The gel formation was confirmed by the inverted vial method and repeating heating and cooling steps supported the thermoreversibility of the organogel. All experiments were performed at least in triplicate.

4B.4.5 Rheological studies: Rheology measurements was carried out in Anton paar MCR 302 instrument by using 15 mm diameter parallel plate, Measurement were done at room temperature 25 °C. A strain sweep experiment at a constant frequency (10 rad/s) was performed in the 0.01–100% range to determine the linear viscoelastic region of the gel sample.

4B.4.6 FESEM analysis: The morphology of the nucleolipid gels was characterized by FESEM. Diluted gel samples with respective solvent was drop-casted on silicon wafer and dried in a vacuum desiccators for ~12 h. Gold sputtering was done on the sample to minimize sample charging before SEM analysis. The SEM images were analyzed by using ImageJ 1.46r software to measure the dimensions of the sheets and twisted ribbons.

4B.4.7 Crystallography: Single crystals of **7b** and **7c** were obtained from dioxane by slow evaporation method. The diffraction data of **7b** and **7c** were collected using Bruker KAPPA APEX II CCD Duo diffractometer (operated at 1500 W, 50 kV and 30 mA) equipped with a graphite-monochromated MoKα (λ = 0.71073 Å) radiation source. The structure of nucleolipids was refined on F^2 by full-matrix least-squares technique using SHELX program.[29] All non-hydrogen atoms were refined anisotropically unless otherwise mentioned. Hydrogen atoms were constrained in geometric positions with respect to their parent atoms. The X-ray crystallographic coordinates for the structures reported in this article have been deposited at the Cambridge Crystallographic Data Centre (CCDC) under deposition number CCDC: 1831323 (**7b**) and 1831324 (**7c**). These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif. Crystallographic data for the nucleolipids are listed in appendix.

4B.4.8 Powder X-ray diffraction (PXRD) analysis of nucleolipid 7c and 7d: The PXRD data taken from the xerogels of **7c** in DMSO and **7d** in DMSO at respective CGC were drop-casted on glass slide and was allowed to from gel at RT. The glass slide was placed in a vacuum desiccator and was dried under vacuum for nearly 15 h to obtain the xerogel. PXRD data was collected using Bruker D8 Advance diffractometer with CuKα source (1.5406 Å). Diffraction data were collected at 2θ angle from 1° to 30° using a 0.01° step size and 0.5 s per step. Low angle diffraction data was collected by keeping the motorized divergence slit in automatic mode so as to maintain the X-ray beam footprint on the sample to12 x 12 mm. Further, the position sensitive detector (Lynxeye) channels were reduced to minimize the background X-ray scattering entering the detector.

4B.4.9 Variable temperature ^{1}H NMR experiments: Gels of **7c** (1.2 w/v %) and **7d** (0.7 w/v %) in *d6*-DMSO were formed in individual NMR tubes by heating and cooling process. ^{1}H NMR was recorded on a 500 MHz spectrometer as a function of increasing temperature. The temperature of the sample was elevated from 25 °C to 60 °C with an increment of 5 °C and equilibration time of 10 min. The spectrum was recorded at every 5 °C interval.

4B.5 References:

1 (a) Safont-Sempere, M. M.; Fernandez, G.; Wurthner, F. *Chem. Rev*. **2011**, *111*, 5784–5814. (b) Zhou, M.; Smith, A. M.; Das, A. K.; Hodson, N. W.; Collins, R. F.; Ulijn. R. V.; Gough, J. E. *Biomaterials* **2009**, *30*, 2523–2530.

2 Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry* 2007, 6th ed W. H. Freeman.

3 (a) Edwards, W.; Smith, D. K. *J. Am. Chem. Soc*. **2013**, *135*, 5911–5920. (b) Raeburn, J.; Adams, D. J. *Chem. Commun*. **2015**, *51*, 5170–5180.

4 Draper, E. R.; Eden, E. G. B.; McDonald, T. O.; Adams, D. J. *Nat. Chem*. **2015**, *7*, 848–852.

5 (a) Moffat, J. R.; Smith, D. K. *Chem. Commun*. **2009**, 316–318. (b) Raeburn, J.; Alston, B.; Kroeger, J.; McDonald, T. O.; Howse, J. R.; Cameron, P. J.; Adams, D. J. *Mater. Horiz*. **2014**, *1*, 241–246. (c) Singh, N.; Zhang, K.; Angulo-Pachón, C. A.; Mendes, E.; van Esch, J.; Escuder, B. *Chem. Sci*. **2016**, *7*, 5568–5572.

6 (a) Morris, K. L.; Chen, L.; Raeburn, J.; Sellick, O. R.; Cotanda, P.; Paul, A.; Griffiths, P. C.; King, S. M.; O′ Reilly, R. K.; Serpell, L. C.; Adams, D. J. *Nat. Commun*. **2014**, *4*, 1480. (b) Colquhoun, C.; Draper, E. R.; Eden, E. G. B.; Cattoz, B. N.; Morris, K. L.; Chen, L.; McDonald, T. O.; Terry, A. E.; Griffiths, P. C.; Serpell,

L. C.; Adams, D. J. *Nanoscale*, **2014**, *6*, 13719–13725. (c) Draper, E. R.; Lee, J. R.; Wallace, M.; Jäckel, F.; Cowan, A. J.; Adams, D. J. *Chem. Sci*. **2016**, *7*, 6499–6505.

7 Lee, S. J.; Cho, S.-H.; Mulfort, K. L.; Tiede, D. M.; Hupp, J. T.; Nguyen, S. B. T. *J. Am. Chem. Soc*. **2008**, *130*, 16828–16829.

8 Sugiyasu, K.; Kawano, S. I.; Fujita, N.; Shinkai, S. *Chem. Mater*. **2008**, *20*, 2863–2865.

9 Vieira, V. M. P.; Hay, L. L.; Smith, D. K. *Chem. Sci*. **2017**, *8*, 6981–6990.

10 (a) Semeen, N. C. *Nature* **2003**, *421*, 427–431. (b) Yan, H. *Science* **2004**, *306*, 2048–2049.

11 (a) Tian, Y.; Mao, C. *J. Am. Chem. Soc*. **2004**, *126*, 11410–11411. (b) Saha, S.; Prakash, V.; Halder, S.; Chakraborty, K.; Krishnan, Y. *Nat. Nanotechnol*. **2015**, *10*, 645–651. (c) Yang, X, Tang, Y.; Mason, S. D.; Chen, J.; Li, F. *ACS Nano* **2016**, *10*, 2324−2330.

12 Langecker, M.; Arnaut, V.; Martin, T. G.; List, J.; Renner, S.; Mayer, M.; Dietz, H.; Simmel, F. C. *Science* **2012**, *338*, 932–936.

13 Han, D.; Park, Y.; Kim, H.; Lee, J. B. *Nat. Commun*. **2014**, *5*, 4367.

14 Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335*, 831–834.

15 (a) Li, X.; Kuang, Y.; Shi, J.; Gao, Y.; Lin, H.-C.; Xu, B. *J. Am. Chem. Soc*. **2011**, *133*, 17513–17518. (b) Li, X.; Kuang, Y.; Xu, B *Soft Matter* **2012**, *8*, 2801–2806.

16 Godeau, G.; Bernard, J.; Staedel, C.; Barthélémy, P. *Chem. Commun*. **2009**, 5127–5129.

17 Yuan, D.; Xu, B. *J. Mater. Chem. B* **2016**, *4*, 5638–5649.

18 (a) Bhattacharya, S.; Acharya, S. N. G. *Chem. Mater*. **1999**, *11*, 3121–3132. (b) Jung, J. H.; Do, Y.; Lee, Y.-A.; Shimizu, T. *Chem. Eur. J*. **2005**, *11*, 5538–5544. (c) Yu, X.; Chen, L.; Zhang, M.; Yi, T. *Chem. Soc. Rev*. **2014**, *43*, 5346–5371.

19 IUPAC-IUB Commission on Biochemical Nomenclature. Abbreviations and Symbols for the Description of Conformations of Polynucleotide Chains. *Pure Appl. Chem.* **1983**, *55*, 1273– 1280.

20 Nuthanakanti, A.; Srivatsan, S. G. *Nanoscale* **2016**, *8*, 3607–3619.

21 (a) Basak, S.; Nanda, J.; Banerjee, A. *J. Mater. Chem*. **2012**, *22*, 11658−11664. (b) Bhattacharjee, S.; Maiti, B.; Bhattacharya, S. *Nanoscale* **2016**, *8*, 11224−11233.

22 (a) Lu, J.; Hu, J.; Liu, C.; Gao, H.; Ju, Y. *Soft Matter* **2012**, *8*, 9576−9580. (b) Deng, M.; Zhang, L.; Jiang, Y.; Liu, M. *Angew. Chem., Int. Ed*. **2016**, *55*, 15062–15066. (c)

Liu, Z.; Wang, D.; Cao, M.; Han, Y.; Xu, H.; Wang, Y. *ACS Appl. Mater. Interfaces* **2015**, *7*, 15078–15087.

23 (a) Banchelli, M.; Berti, D.; Baglioni, P. *Angew. Chem., Int. Ed.* **2007**, *46*, 3070–3073. (b) Moreau, L.; Camplo, M.; Wathier, M.; Taib, N.; Laguerre, M.; Bestel, I.; Grinstaff, M. W.; Barthélémy, P. *J. Am. Chem. Soc*. **2008**, *130*, 14454–14455.

24 (a) Li, X.; Kuang, Y.; Lin, H.-C.; Gao, Y.; Shi, J.; Xu, B. *Angew. Chem., Int. Ed.* **2011**, *50*, 9365–9369. (b) Li, D.; Shi, Y.; Wang, L. *Chin. J. Chem.* **2014**, *32*, 123–127.

25 Kang, S. O.; Powell, D.; Day, V. W.; Bowman-James, K. *Angew. Chem., Int. Ed.* **2006**, *45*, 1921–1925.

26 Chatterjee, S.; Pathmasiri, W.; Plashkevych, O.; Honcharenko, D.; Varghese, O. P.; Maiti, M.; Chattopadhyaya, J. *Org. Biomol. Chem*. **2006**, *4*, 1675–1686.

27 (a) Clever, G. H.; Kaul, C.; Carell, T. *Angew. Chem., Int. Ed.* **2007**, *46*, 6226−6236. (b) Ono, A.; Torigoe, H.; Tanaka, Y.; Okamoto, I. *Chem. Soc. Rev*. **2011**, *40*, 5855−5866. (c) Scharf, P.; Müller, J. *ChemPlusChem* **2013**, *78*, 20−34.

28 Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem*. **1994**, *5*, 508–512.

29 G. M. Sheldrick, A Short History of SHELX. *Acta Crystallogr. Sect.* **2008**, *A64*, 112–122.

4B.6 Appendix-IV: Characterization data of synthesized compounds

4B.6.1 Crystallographic Information

4B.6.1.1 Crystallographic data for nucleolipid **7b**

Compound identity	**7b**
Empirical formula	$C_{33} H_{56} N_2O_8$
Formula weight	608.80
Temperature	296(2) K
Wavelength	0.71073 Å
Crystal system	Monoclinic
Space group	C2
Unit cell dimensions	a = 16.404(8) α = 90
	b = 5.345(3) β = 96.154(10)

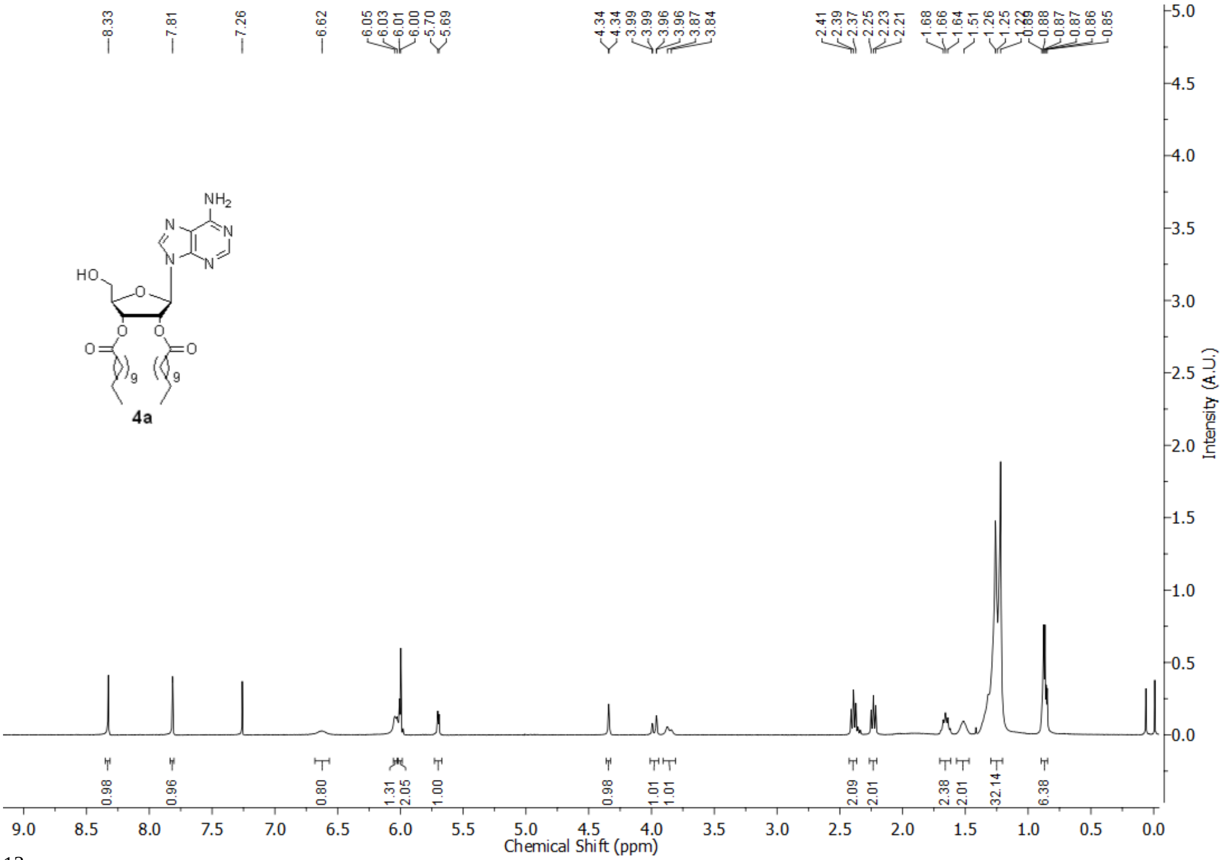
	c = 38.959(18) γ = 90
Volume	3396(3) $Å^3$
Z	4
Density (calculated)	1.191 Mg/cm^3
Absorption coefficient (μ)	0.085 mm^{-1}
F(000)	1328
Crystal size	0.30 X 0.16 X 0.06 mm^3
Theta range for data collection	2.63 to 19.00
Index ranges	-17<=h<=17, -5<=k<=5, -42<=l<=42
Reflections collected	14394
Independent reflections	4195 [R(int) = 0.0592]
Completeness to theta = 24.86°	95.4 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.984 and 0.995
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	4195 / 1 / 391
Goodness-of-fit on F^2	1.117
Final R indices [I>2sigma(I)]	R1 = 0.0619, wR2 = 0.1527
R indices (all data)	R1 = 0.0745, wR2 = 0.1630
Largest diff. Peak and hole	0.191 and -0.245 e.$Å^{-3}$
CCDC	1831323

4B.6.1.2 Crystallographic data for nucleolipid **7c**

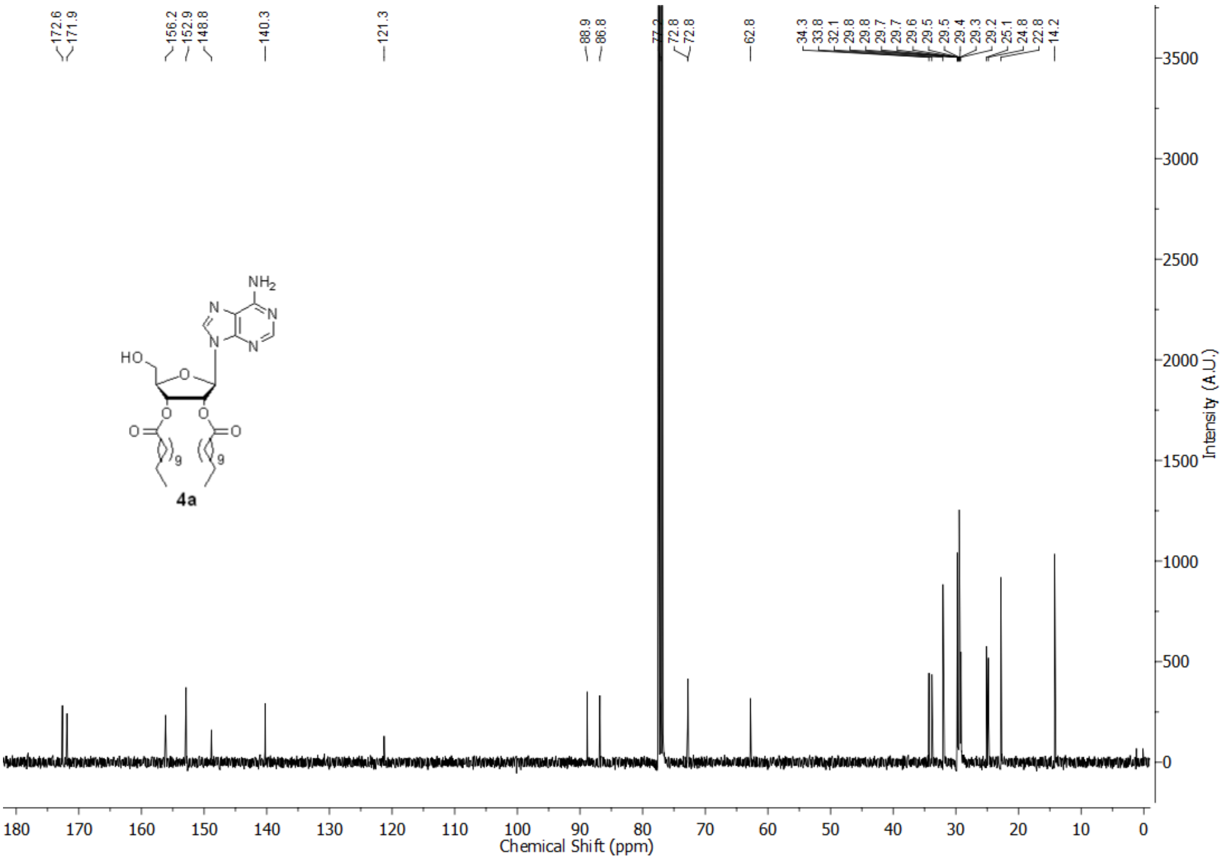
Compound identity	**7c**	
Empirical formula	$C_{37} H_{64} N_2 O_8$	
Formula weight	664.90	
Temperature	150(2) K	
Wavelength	0.71073 Å	
Crystal system	Monoclinic	
Space group	C2	
Unit cell dimensions	a = 83.98(4)	α = 90
	b = 7.173(4)	β = 94.66(2)
	c = 7.066(3)	γ = 90

Volume	4242(3) Å^3
Z	4
Density (calculated)	1.041 Mg/cm^3
Absorption coefficient (μ)	0.072 mm^{-1}
F(000)	1456
Crystal size	0.40 X 0.25 X 0.12 mm^3
Theta range for data collection	2.85 to 24.01
Index ranges	-110<=h<=110, -9<=k<=9, -6<=l<=9
Reflections collected	32581
Independent reflections	9804 [R(int) = 0.1700]
Completeness to theta = 25.242°	97.8 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.582 and 0.746
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	9804 / 13 / 427
Goodness-of-fit on F^2	0.972
Final R indices [I>2sigma(I)]	R1 = 0.0833, wR2 = 0.1736
R indices (all data)	R1 = 0.2197, wR2 = 0.2163
Largest diff. Peak and hole	0.325 and -0.321 e.Å^{-3}
CCDC	1831324

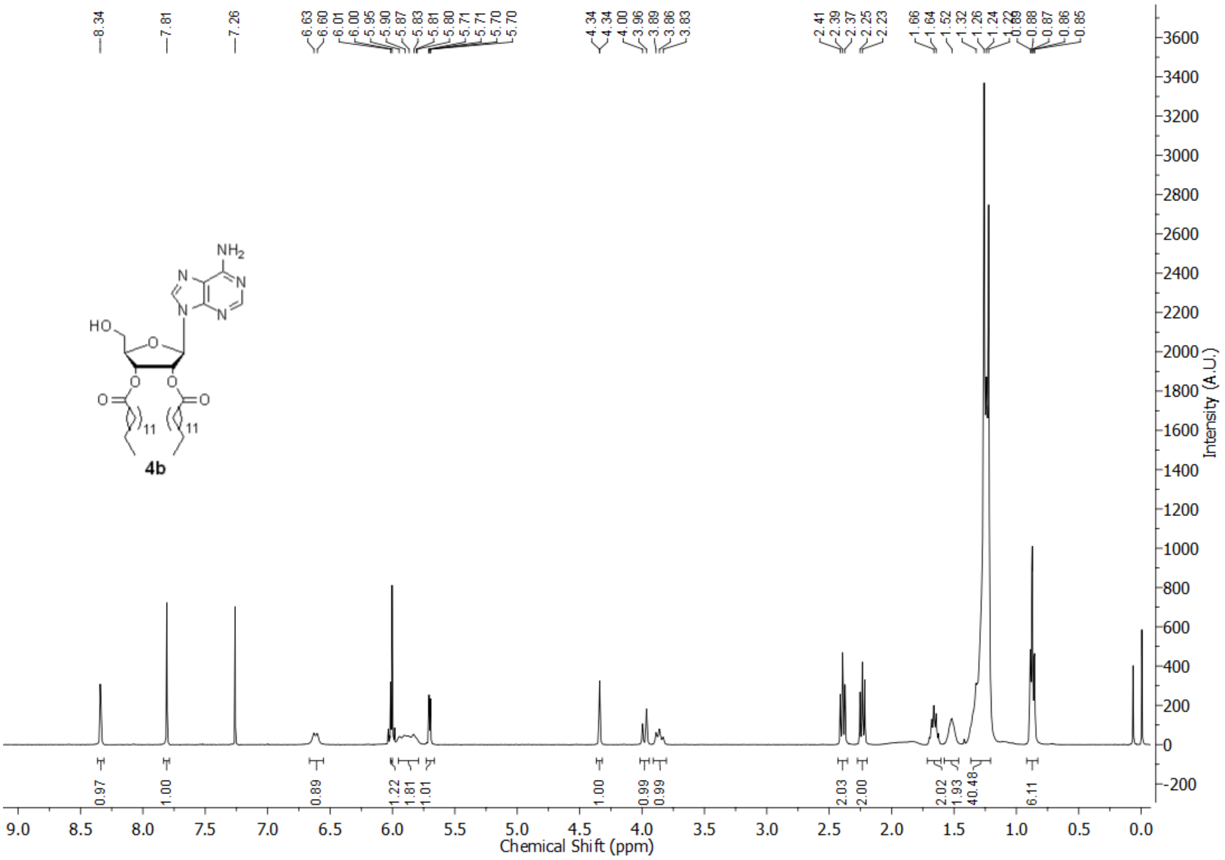
^{1}H NMR of nucleolipid **4a** in $CDCl_3$



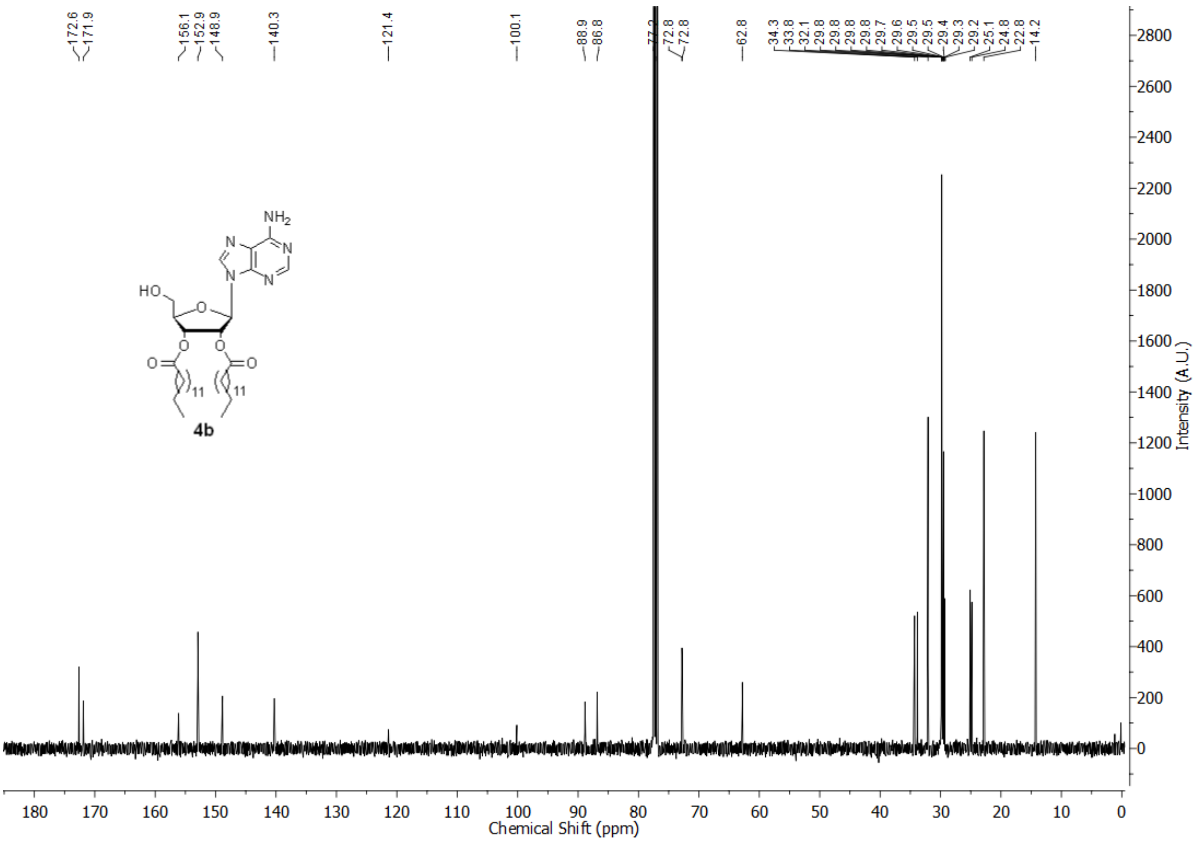
^{13}C NMR of nucleolipid **4a** in $CDCl_3$



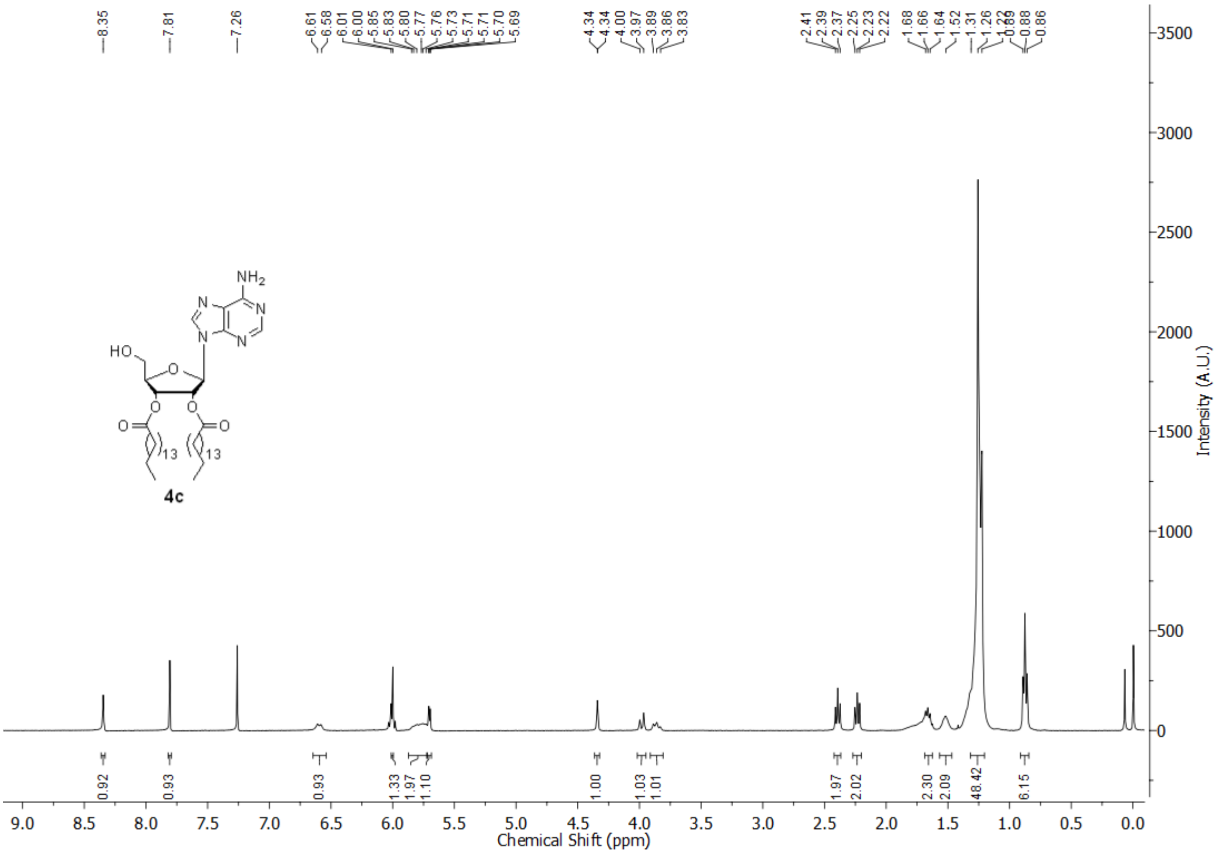
[1]H NMR of nucleolipid **4b** in $CDCl_3$



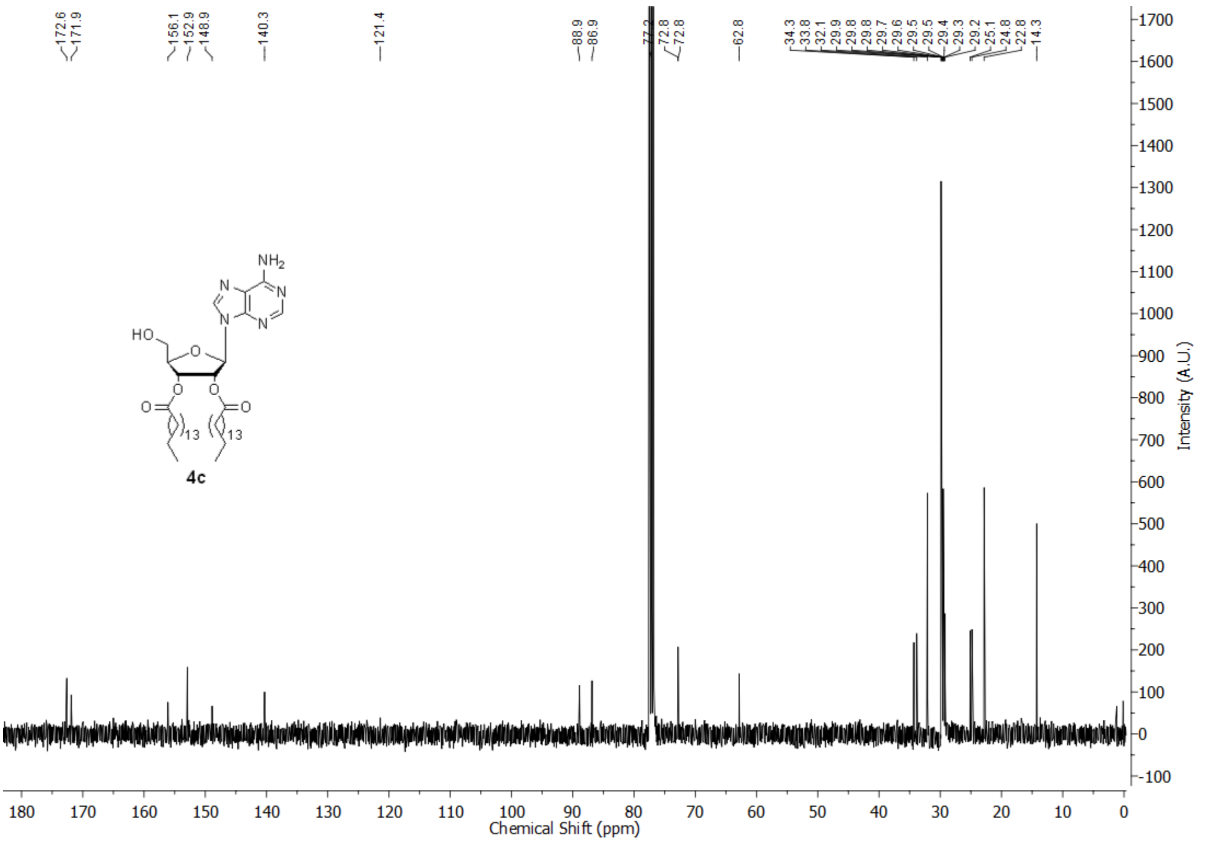
[13]C NMR of nucleolipid **4b** in $CDCl_3$



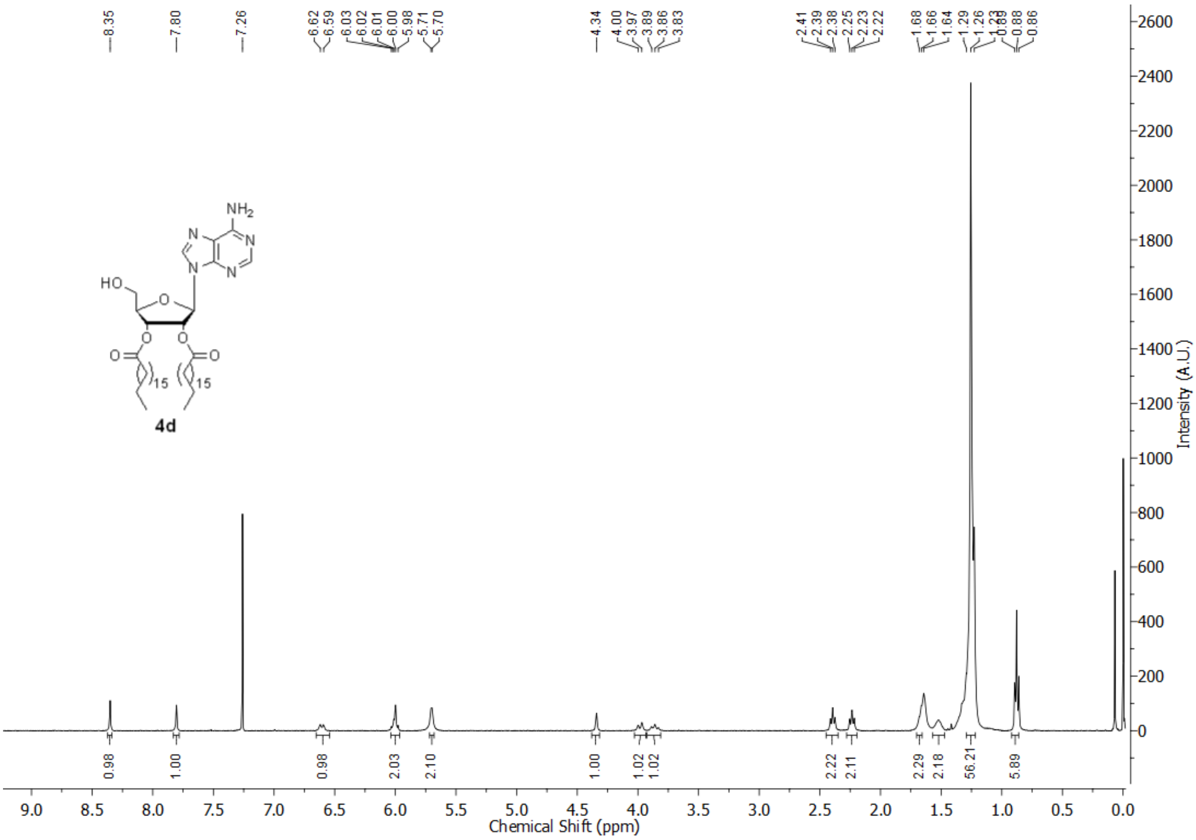
^{1}H NMR of nucleolipid **4c** in $CDCl_3$



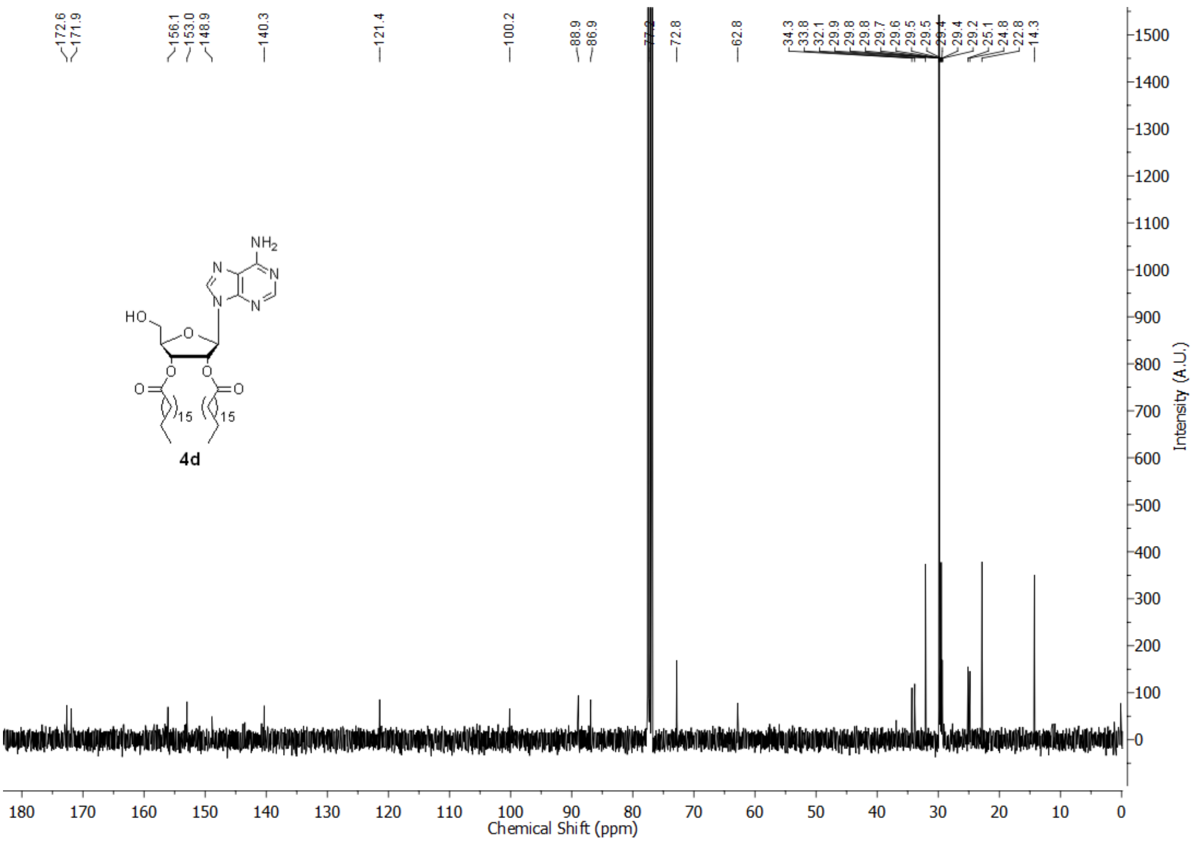
^{13}C NMR of nucleolipid **4c** in $CDCl_3$



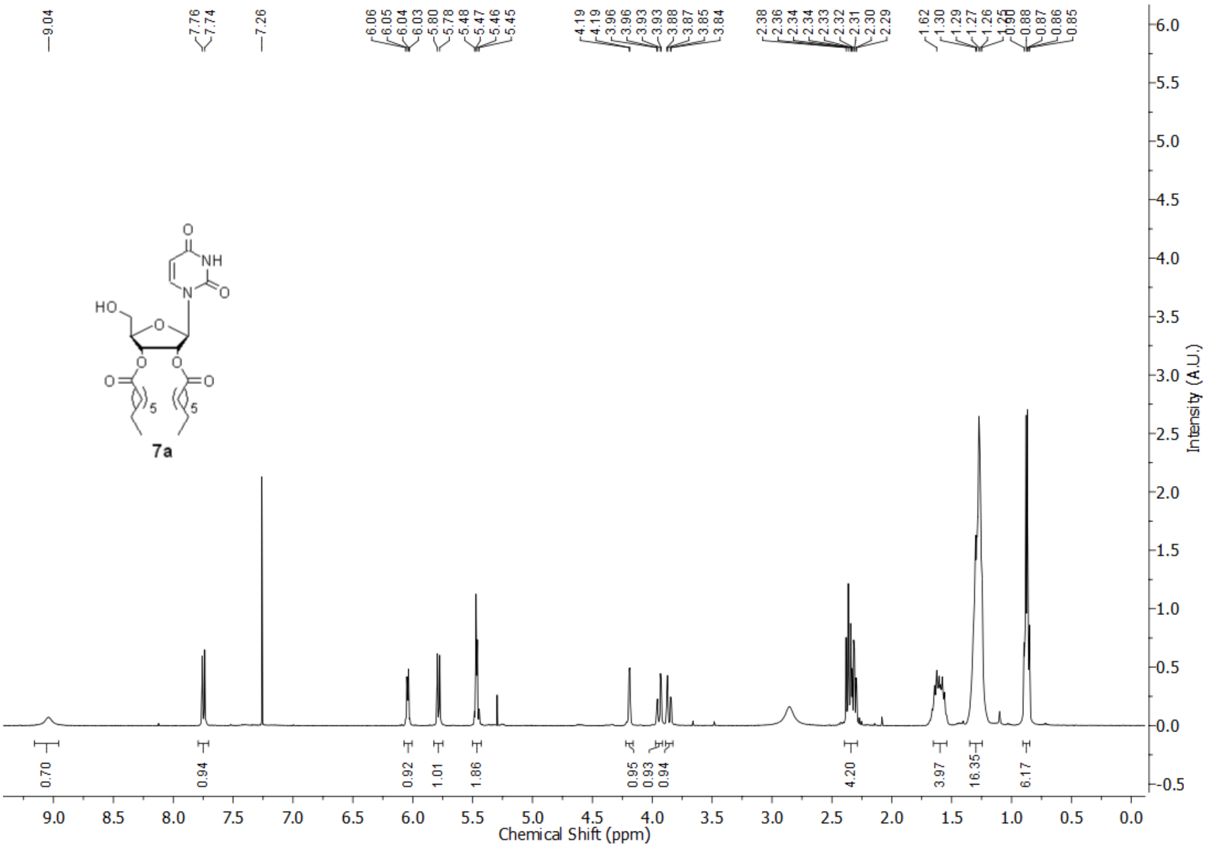
[1]H NMR of nucleolipid **4d** in $CDCl_3$



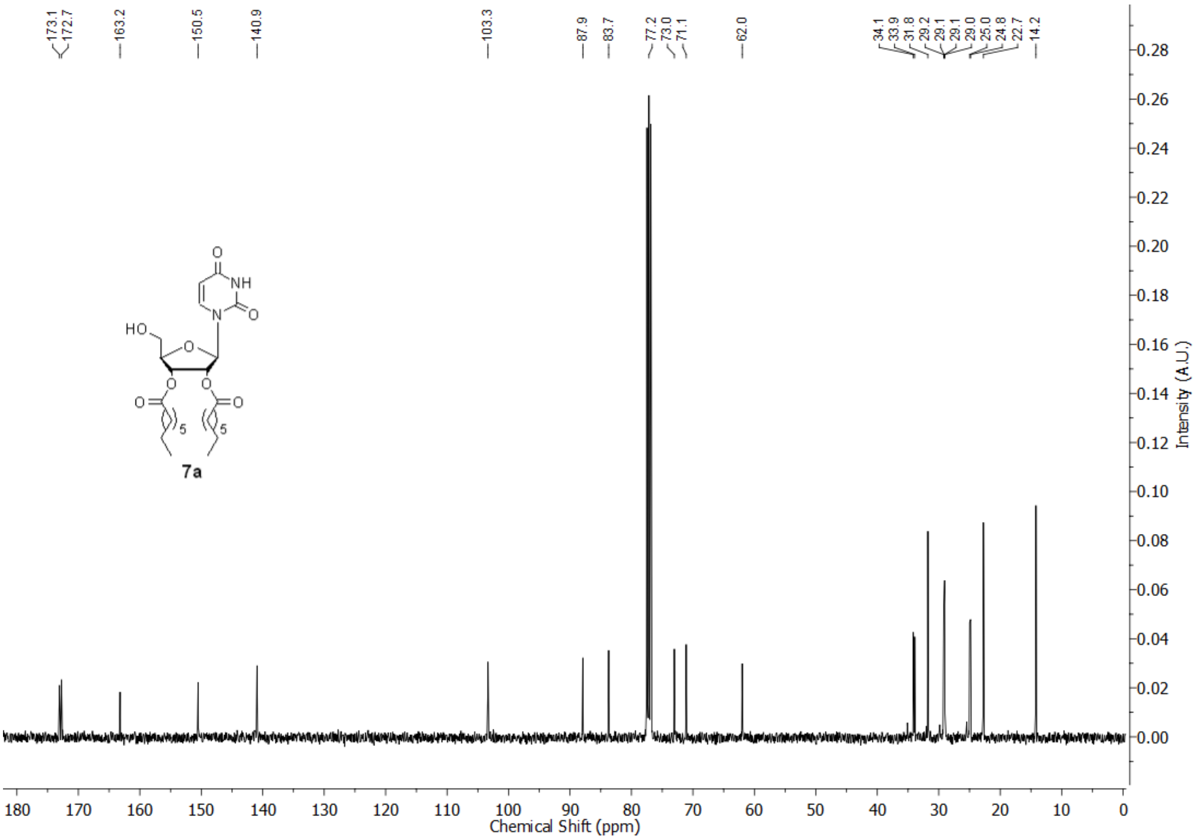
[13]C NMR of nucleolipid **4d** in $CDCl_3$



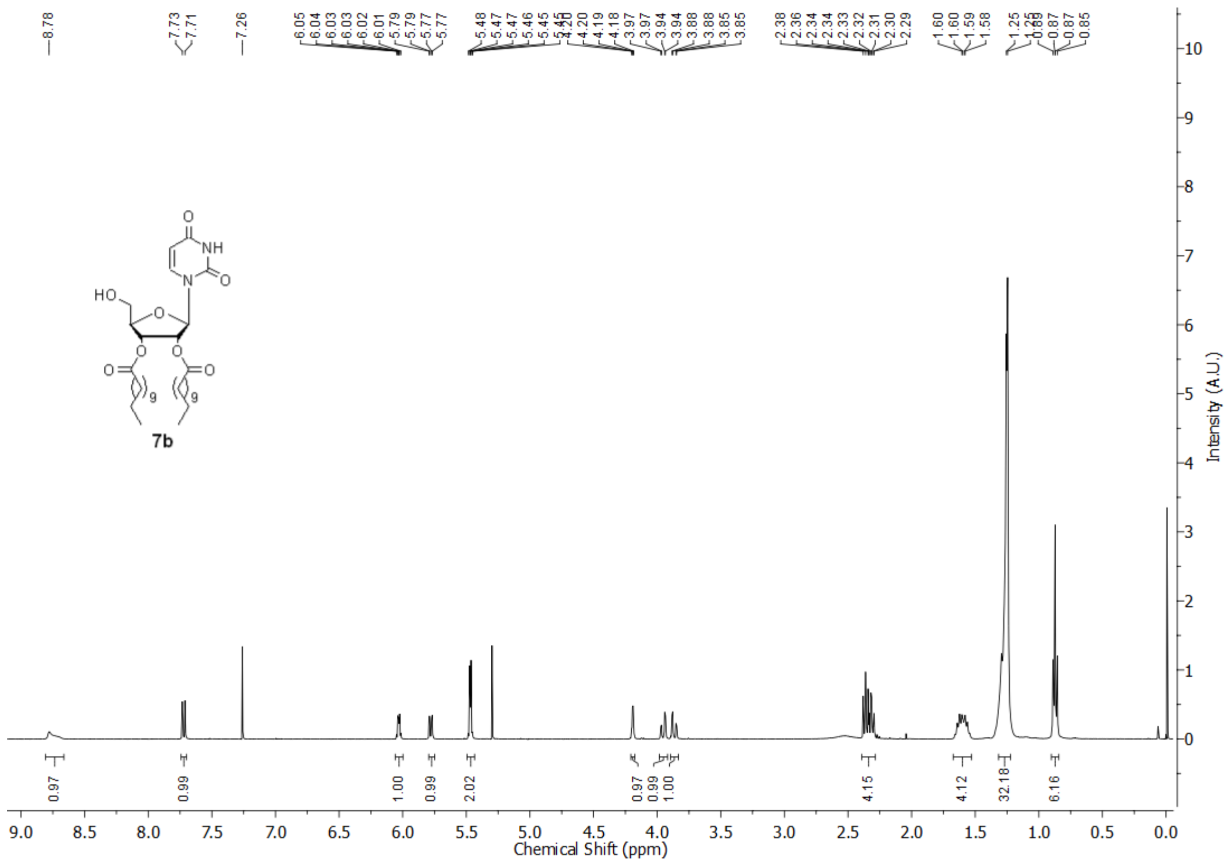
^{1}H NMR of nucleolipid **7a** in $CDCl_3$



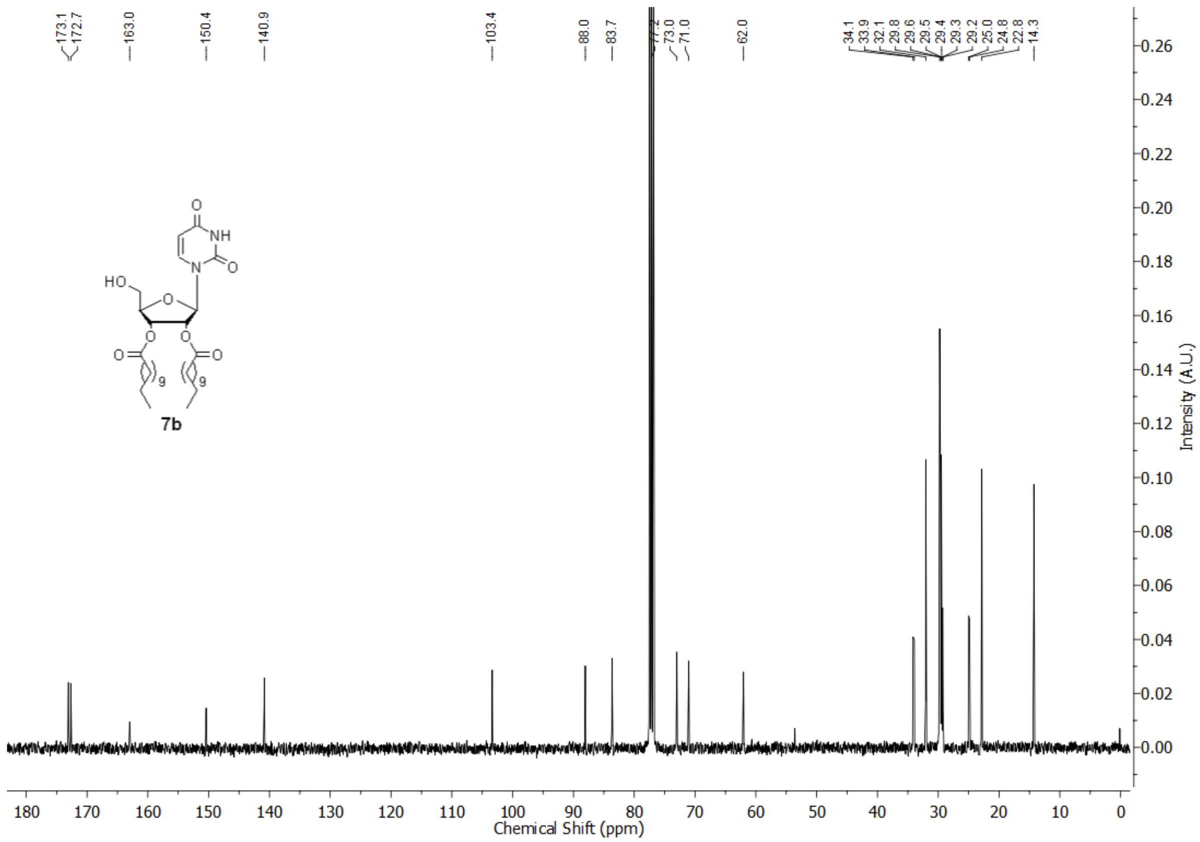
^{13}C NMR of nucleolipid **7a** in $CDCl_3$



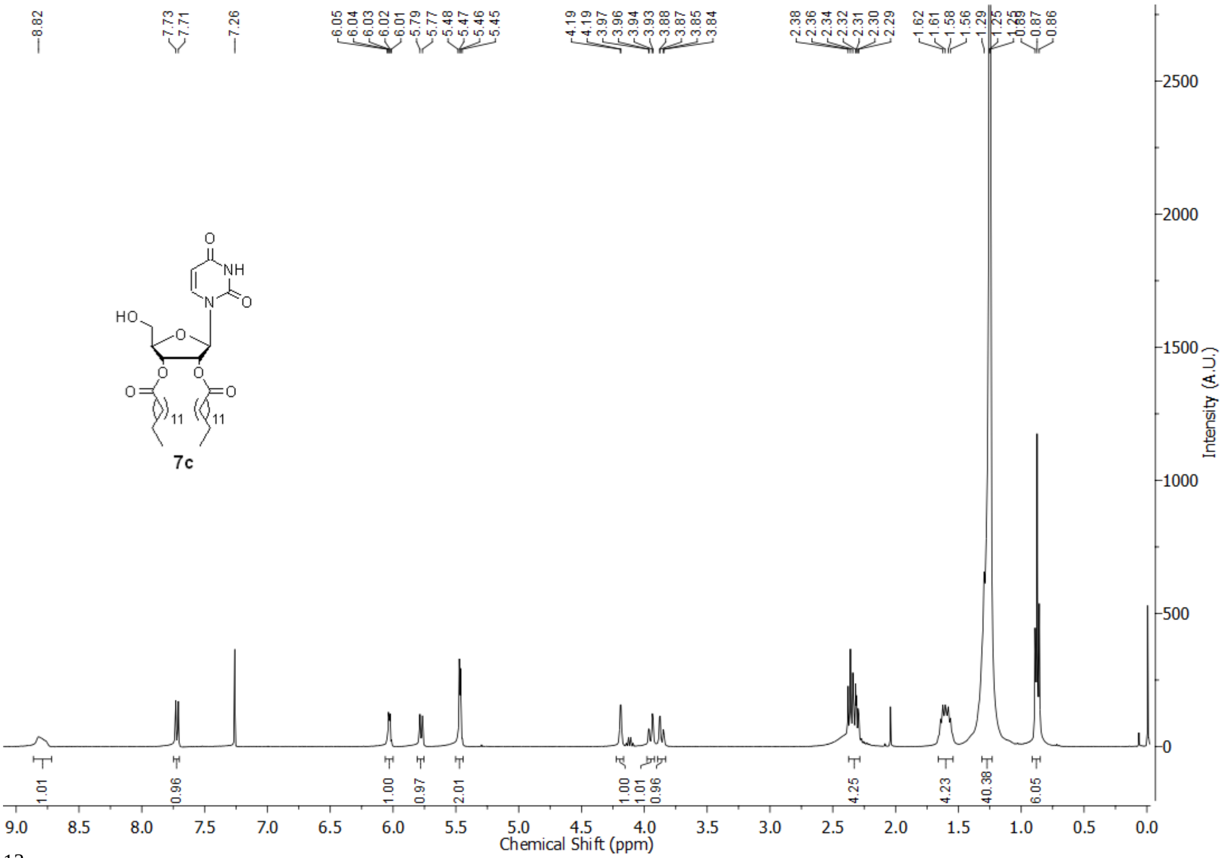
[1]H NMR of nucleolipid **7b** in $CDCl_3$



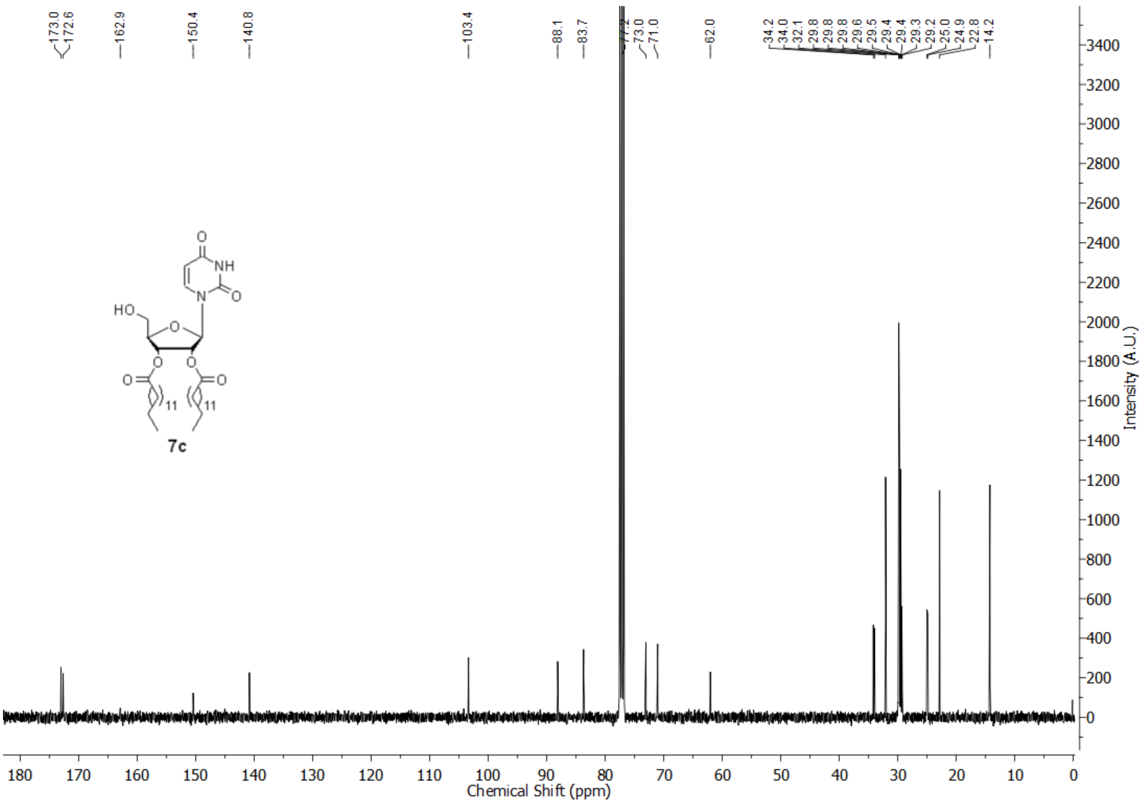
[13]C NMR of nucleolipid **7b** in $CDCl_3$



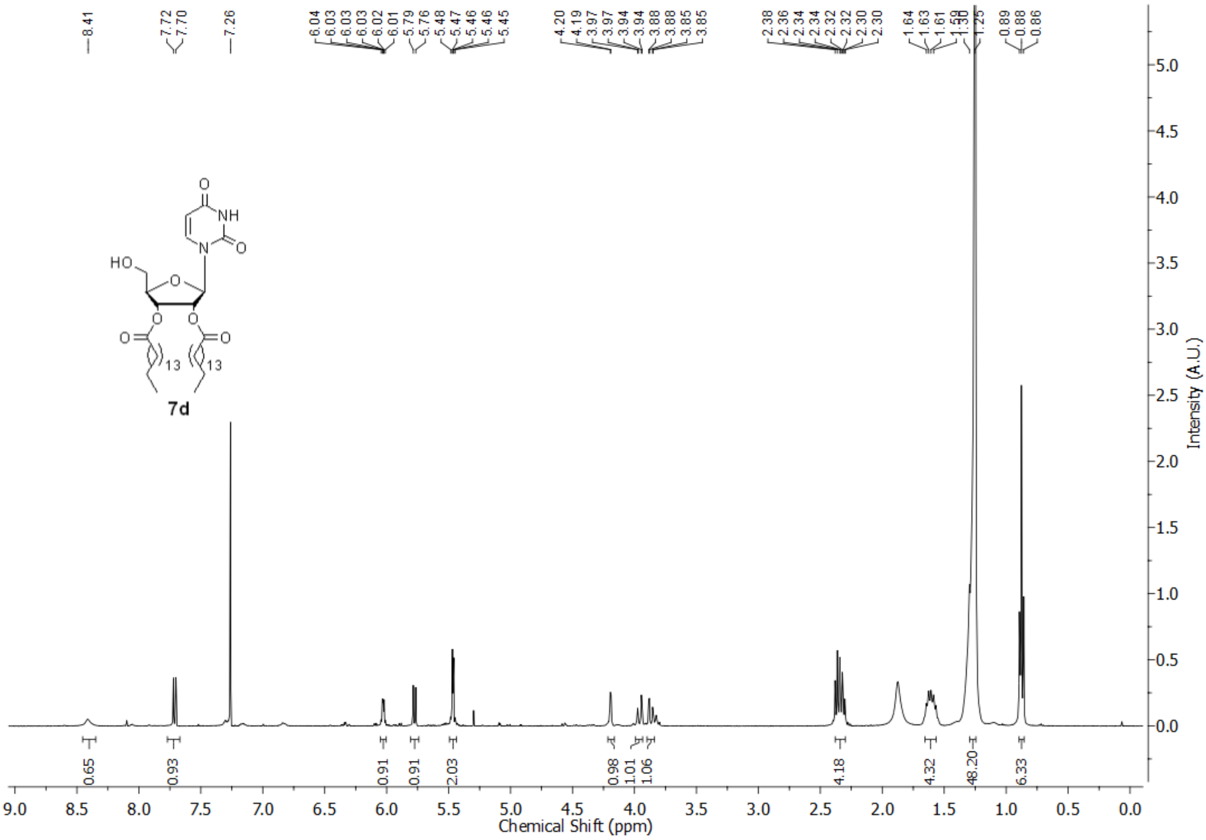
^1H NMR of nucleolipid **7c** in $CDCl_3$



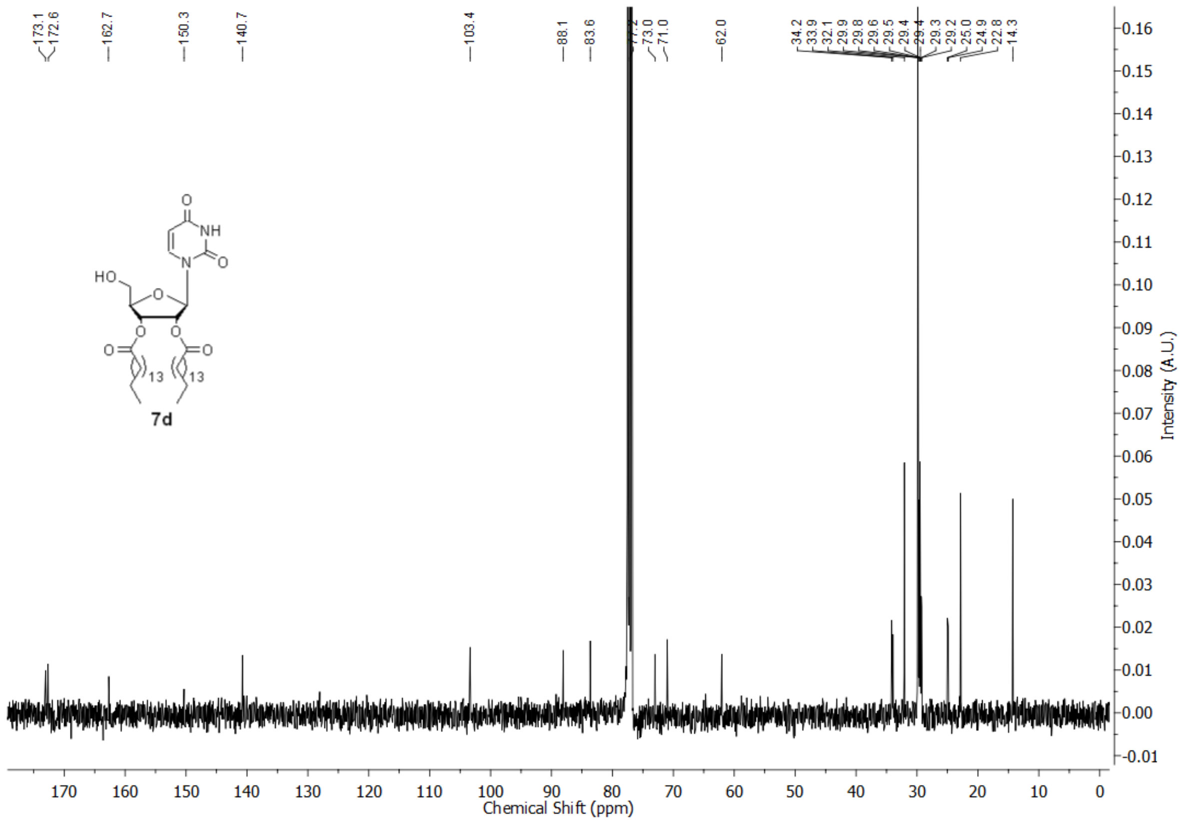
^{13}C NMR of nucleolipid **7c** in $CDCl_3$



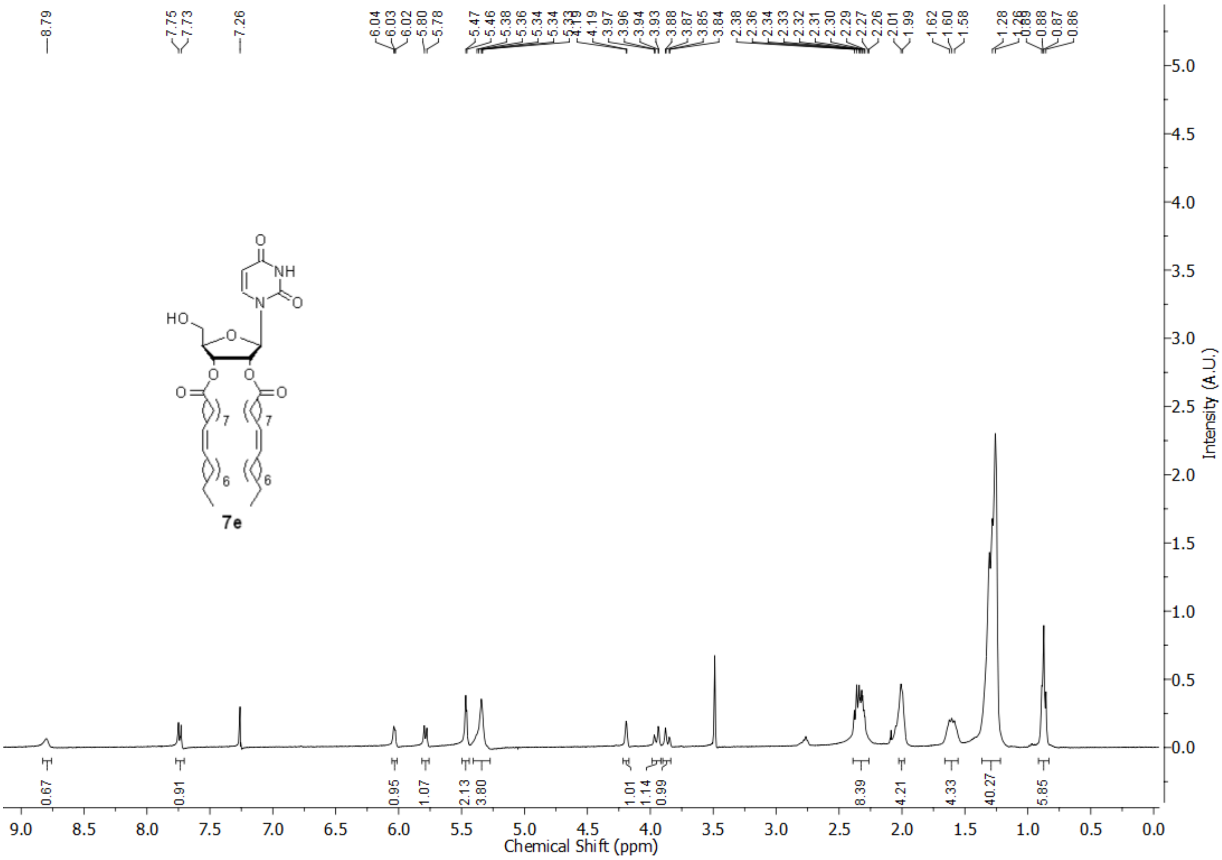
^{1}H NMR of nucleolipid **7d** in $CDCl_3$



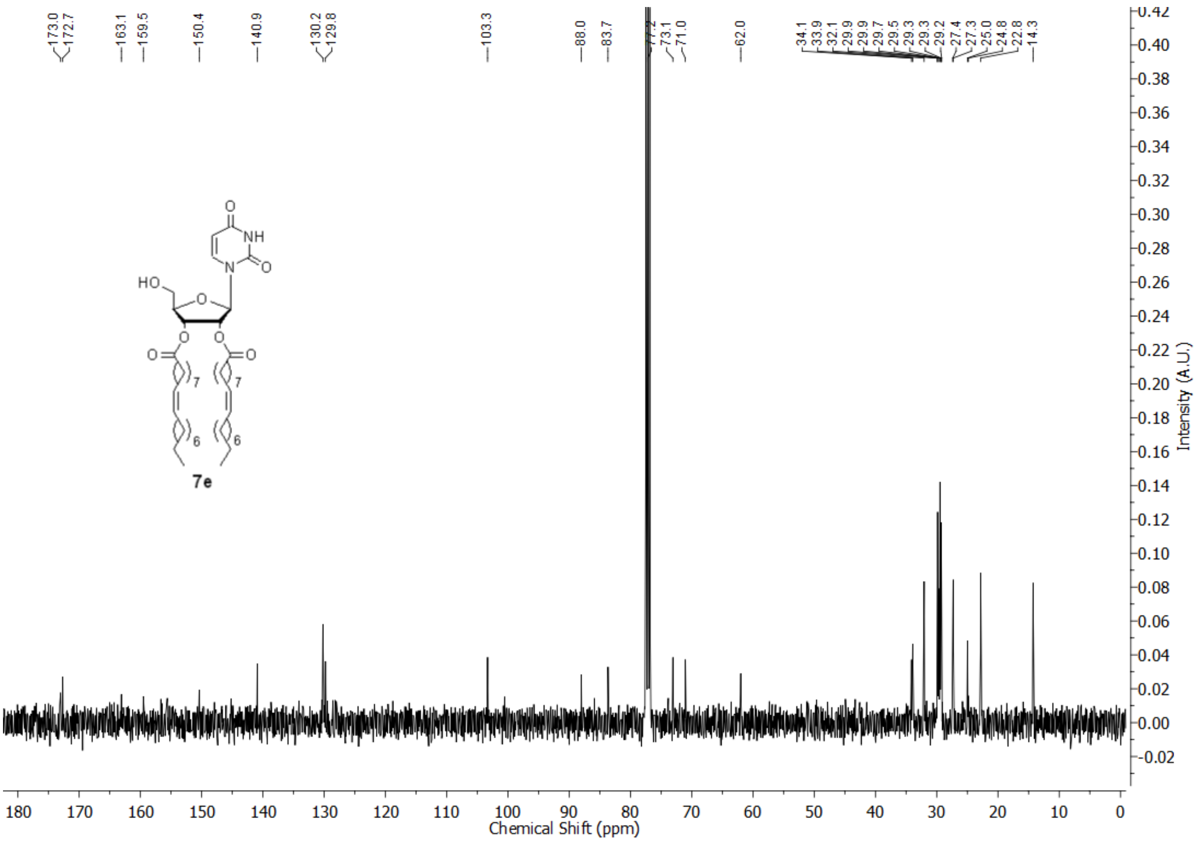
^{13}C NMR of nucleolipid **7d** in $CDCl_3$



[1]H NMR of nucleolipid **7e** in $CDCl_3$



[13]C NMR of nucleolipid **7e** in $CDCl_3$



Chapter 5

Switchable nucleolipid supramolecular gels based on environmentally-sensitive fluorescent nucleoside analogs

5.1 Introduction

Unique but predictable base pairing property of nucleic acids has been extensively utilized in the construction of programmable self-assembled architectures for applications in nanotechnology.[1,2] Many such nanoscale structures have been developed as molecular machines,[3] nanocapsules,[4] synthetic membranes and ion channels[5] and as therapeutic tools.[2,6] Hybrid nucleic acid conjugates made of polymers and lipids that form supramolecular assemblies have also attracted significant interest, namely in the development of scaffolds for gene and drug delivery.[7] While scalability and fabrication of such nucleic acid nanostructures still remain a challenge, nucleic acid components (e.g., nucleobase and nucleosides) and their derivatives provide an alternative route for the construction of a variety of supramolecular architectures.[8] For example, guanosine and its derivatives form ordered supramolecular assemblies in the presence of monovalent cations resulting in stable gels, ion channels, nano wires, etc.[9,10] Especially, tailor-made nucleoside-fatty acid hybrids, which combine the aggregation property of amphiphiles and exquisite recognition property of nucleobases, have gained significant attention as they form diverse supramolecular assemblies like vesicles, fibres, hydro- and organogels.[11] Many such nucleolipids have been used as efficient gene and drug delivery systems.

While most of the efforts have been focused towards developing efficient biocompatible nucleolipid transport systems,[11,12] the potential of nucleolipid self-assemblies in constructing responsive materials and sensors has not been well explored. Incidentally, the nucleosides can be conveniently modified to impart functionalities in addition to their inherent H-bonding, π-stacking and metal ion binding capabilities. Therefore, we envisioned that incorporating appropriate tunable features (e.g., multistimuli responsiveness, fluorescence, etc.) in nucleolipids would enable the development of new self-assembling smart materials and sensors.[7a,15] Here, we describe the development of a new family of nucleolipids containing environmentally-sensitive fluorescent pyrimidine nucleoside analogs, based on 5-(benzofuran-2-yl)uracil and 5-(benzo[b]thiophen-2-yl)uracil cores, as head groups and classical long chain fatty acids, attached at the 2′-*O*- and 3′-*O*-positions of the nucleoside, as lipophilic groups. The fluorescence properties of benzofuran- and benzothiophene-modified nucleolipids are highly sensitive to solvent polarity and viscosity changes. These fluorescent nucleolipids readily form supramolecular organogels driven by hierarchical structures (fibres, twisted ribbons, helical ribbons and nanotubes), which depend on the length and nature of fatty acid chain and nucleobase modification. Rewardingly, in addition to aggregation-induced enhanced emission (AIEE), the nucleolipid organogels also exhibit multistimuli responsiveness towards external stimuli including temperature, ultrasound and chemicals (anions, acid–base and metal ions).

5.2 Results and discussion

5.2.1 Design and synthesis of fluorescent nucleolipids

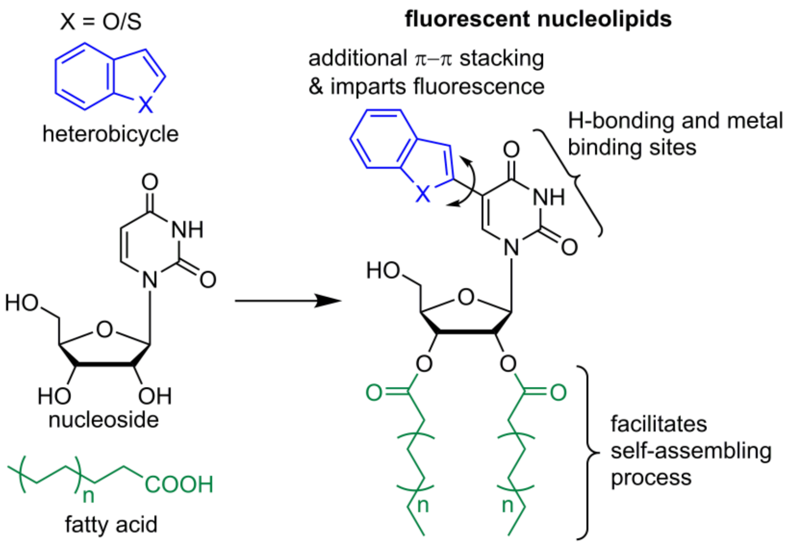
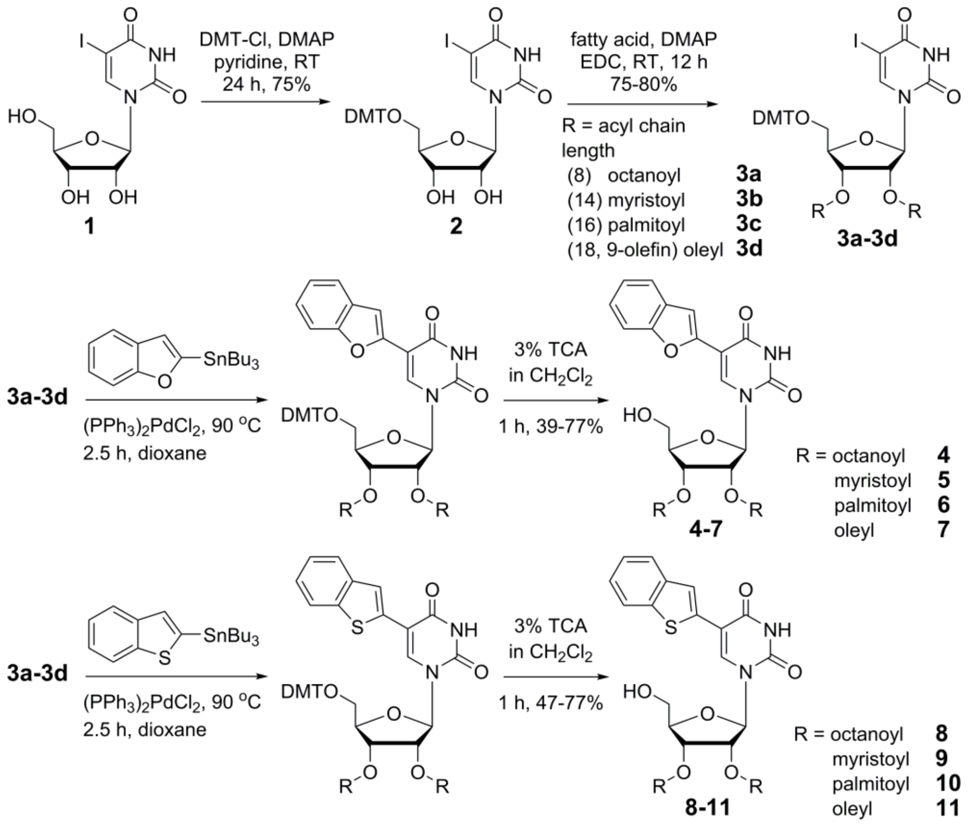


Figure 1. Design of self-assembling and responsive fluorescent nucleolipids. Conjugation of heterobicyclic moieties at the 5-position of uridine could impart fluorescence properties and also additional π–π stacking interaction without affecting the Watson-Crick H-bonding capabilities of the nucleoside. Coupling of fatty acids at the 2′-*O*- and 3′-*O*-positions of uridine could facilitate the self-assembling process. The fluorescence properties of the obtained nucleolipids could be sensitive to rigidification-derigidification process due to the presence of a rotatable aryl-aryl bond between the heterobicycle and uracil rings.

We took a cue from a fluorescent amino acid, tryptophan (indole derivative), which is reasonably fluorescent and highly sensitive to its microenvironment, to construct environmentally-sensitive fluorescent nucleoside–lipid hybrids. While conjugation of indole and N-methylindole heterobicycles at the 5-position of uracil base did not impart fluorescence, conjugation of benzofuran and benzothiophene moieties afforded fluorescent uridine analogs.[16] Therefore, we decided to synthesize a series of base-modified fluorescent nucleolipids by attaching (i) benzofuran and benzothiophene at the 5-position of uridine, which would serve as head groups and (ii) fatty acids at the 2′-*O*- and 3′-*O*-positions of uridine, which would serve as lipophilic groups (Figure 1). In this design, we have purposely chosen to attach the heterobicyclic moieties to the uracil ring via a rotatable aryl–aryl bond, which upon self-assembling could result in rigidification of the fluorophore and hence, exhibit aggregation-induced enhancement in fluorescence. Furthermore, heterocycles conjugated at the 5-position of uracil, should not only retain the Watson–Crick H-bonding ability of the base but also could assist a self assembling process by providing additional π–π stacking and H-bonding interactions.



Scheme 1. Synthesis of 5-benzofuran- (**4**–**7**) and 5-benzothiophene-modified (**8**–**11**) fluorescent nucleolipids. DMT-Cl = 4,4′-Dimethoxytrityl chloride; DMAP = 4-Dimethylaminopyridine; EDC = *N*-(3-Dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride.

5-Benzofuran- and 5-benzothiophene-modified fluorescent nucleolipids **4**–**11** containing varying lengths of long chain saturated as well as unsaturated hydrocarbons were synthesized by following the steps illustrated in Scheme 1. Precursors **3a**–**3d** required to synthesize fluorescent nucleolipids were prepared by first protecting the 5′-OH group of 5-iodouridine with the dimethoxytrityl group, followed by coupling hydrocarbon chains (octanoyl, myristoyl, palmitoyl and oleyl) to secondary hydroxyl groups (2′-*O*- and 3′-OH) using the corresponding acid in the presence of a coupling agent. 5-Benzofuran- (**4**–**7**) and 5-benzothiophene-modified (**8**–**11**) fluorescent nucleolipids were then synthesized by reacting **3a**–**3d** with 2-(tri-n-butylstannyl)benzofuran and 2-(tri-n-butylstannyl) benzothiophene, respectively, under Stille cross-coupling reaction conditions using a palladium catalyst, $Pd(PPh_3)_2Cl_2$. This series of fluorescent nucleolipids would enable the study of effects of nucleobase modification, hydro-carbon chain length and saturation/unsaturation on the physicochemical properties and self-organization capabilities of the nucleolipids.

5. 2.2 Sensitivity to solvent polarity and viscosity

We performed UV absorption and steady-state fluorescence spectroscopic analysis to assess the responsiveness of fluorescent nucleolipids to changes in solvent polarity and viscosity. 5-Benzofuran- and 5-benzothiophene-modified uridine nucleolipids containing myristoyl and palmitoyl acyl chains were used as test systems. The ground-state electronic spectrum of 5-benzofuran-modified nucleolipids **5** and **6** in methanol revealed absorption maxima at around 262 nm, 271 nm and 321 nm. When solvent polarity was changed from methanol to dioxane, the nucleolipids exhibited a marginal shift in absorption maxima accompanied by a small hypochromic effect (Figure 2A, B and Table 1). Similar absorption behaviour was also exhibited by 5-benzothiophene-modified nucleolipids **9** and **10** (Figure 2C, D and Table 1). Emission profile of nucleolipids **5** and **6** containing the benzofuran moiety revealed a slight reduction in quantum yield and reasonable blue shift (∼16 nm) in the emission maximum as the solvent polarity was decreased from methanol to dioxane (Figure 2A, B and Table 1). On the other hand, nucleolipids **9** and **10** containing the benzothiophene moiety exhibited an enhancement in quantum yield and blue-shifted (∼10 nm) emission maximum as the solvent polarity was decreased from methanol to dioxane (Figure 2C, D and Table 1).

The heterobicyclic and uracil rings in 5-benzofuran-modified (**5** and **6**) and 5-benzothiophene-modified uridine nucleolipids (**9** and **10**) are separated by a rotatable aryl–aryl bond. The relative conformation of these ring systems, which could be influenced by molecular crowding effects and viscosity, could affect the conjugation and hence, the fluorescence properties of the nucleolipids.[17] Interestingly, both benzofuran- and benzothiophene-modified fluorescent nucleolipids exhibited significant enhancement in quantum yield in a viscous solvent (ethylene glycol) as compared to in a less viscous solvent (methanol, Figure 2 and Table 1). Further, nucleolipids revealed higher fluorescence anisotropy values in ethylene glycol as compared to in methanol (Table 1). The observed enhancement in fluorescence quantum yield and high anisotropy is possibly due to rigidification of the fluorophores as a result of restricted rotation about the aryl–aryl bond between the heterocyclic moiety and uracil ring in a more viscous solvent.[17] Together, these results indicate that the photophysical properties of fluorescent nucleolipids are sensitive to their surrounding environment.

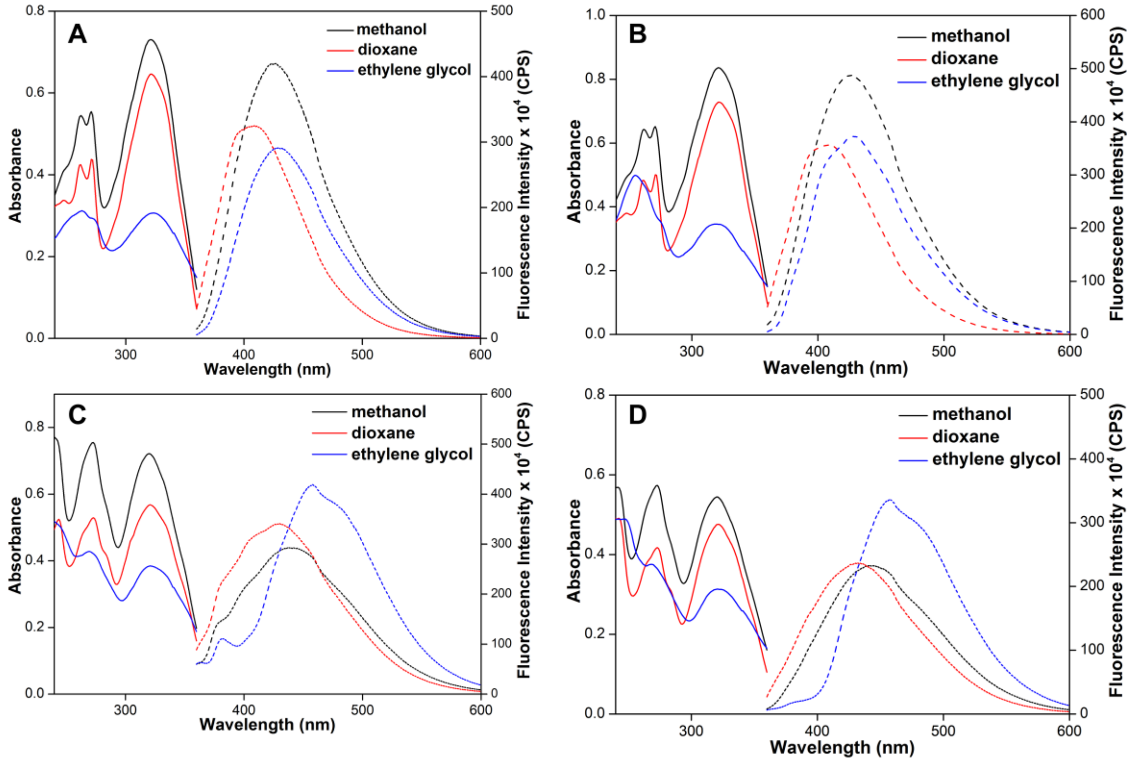


Figure 2. Absorption (solid line, 50 μM) and emission (dashed line, 5 μM) spectra of nucleolipids **5** (**A**), **6** (**B**), **9** (**C**) and **10** (**D**) in different solvents. Steady-state fluorescence experiments were performed by exciting the samples at 320 nm. The excitation and emission slit widths for **5** and **6** were 2 nm and 3 nm, and for **9** and **10** were 3 nm and 3 nm, respectively. All solutions for absorption and emission studies contained 5% and 0.5% dioxane, respectively.

The fluorescence properties of emissive nucleoside incorporated into oligonucleotides are known to be influenced by H-bonding and stacking interactions between neighbouring bases, solvation–desolvation effects and rigidification–derigidification of the fluorophore.[18,19] We envisioned that such factors could also influence the photophysical behaviour of fluorescent nucleolipids possessing chemical components that induce hierarchical self-organization. Hence, we expect that 5-benzofuran- and 5-benzothiophene-functionalized nucleolipids would not only support self-organization but also could be photophysically responsive to reversible self-assembling and disassembling processes.

Table 1. Photophysical properties of 5-benzofuran- and 5-benzothiophene-modified uridine nucleolipids (**5**, **6**, **9** and **10**).

nucleolipid	solvent	$\lambda_{max}{}^{a}$ (nm)	λ_{em} (nm)	Φ^{b}	r^{b}
5	methanol	321	425	0.13	0.037
	dioxane	321	409	0.12	-
	ethylene glycol	321	429	0.25	0.159
6	methanol	321	426	0.13	0.037
	dioxane	321	409	0.12	-
	ethylene glycol	320	427	0.28	0.109
9	methanol	320	439	0.07	0.058
	dioxane	321	430	0.12	-
	ethylene glycol	321	458	0.20	0.163
10	methanol	321	442	0.07	0.059
	dioxane	321	432	0.09	-
	ethylene glycol	324	458	0.18	0.098

[a]Lowest absorption energy maximum is given. [b]Standard deviations for quantum yield (Φ) and fluorescence anisotropy are ≤ 0.01 and ≤ 0.004, respectively. See experimental section.

5.2.3 Gelation behaviour of nucleolipids

The gelation ability of fluorescent nucleolipids was evaluated in various organic solvents by the inverted vial method. The nucleolipids were dissolved in a solvent or in a mixture of solvents by heating and were left to come to room temperature. Rewardingly, 5-benzofuran and 5-benzothiophene-modified uridine nucleolipids containing myristoyl (**5** and **9**) and palmitoyl (**6** and **10**) fatty acid acyl chains formed stable opaque fluorescent organogels in anhydrous DMSO at RT (~25 °C, Figure 3). The gel–sol interconversion was found to be thermoreversible over several cycles of heating and cooling steps, and the gels were found to be stable for months at RT. Other common organic solvents did not support gel formation even at very high concentration of the nucleolipids (∼100 mg mL^{-1}). Ultrasound waves as an external stimulus can promote rapid gelation by facilitating the conversion of intramolecular interactions into intermolecular interactions.[20] Hot nucleolipid solutions when subjected to ultrasonication (∼1 min) resulted in rapid formation of gels, while it took few minutes (palmitoyl-containing nucleolipids) to nearly two hours (myristoyl-containing nucleolipids) to form gels by the heating–cooling process. The gelation capacity was found to depend on the alkyl chain length and nature of fatty acid chains (saturated or unsaturated). Nucleolipids containing shorter hydrocarbon chains (octanoyl) did not yield gels in any of the solvents tested. Although nucleolipids **5** and **9** containing C14 myristoyl chains displayed the gelation property in DMSO, critical gelation concentration (CGC) was found to be significantly higher as compared to nucleolipids **6** and **10** containing C16 palmitoyl chains (Table 2). Furthermore, gel–sol phase transition temperatures (Tgel) increased with increasing chain length, indicating that the palmitoyl-tagged fluorescent nucleolipid gels with lower CGC are more stable than nucleolipids containing shorter myristoyl chains (Table 2). Next we studied the effect of unsaturation in the fatty acid acyl chain on the gelation ability of nucleolipids. Nucleolipids **7** and **11** made of oleyl chain containing an internal *cis* double bond failed to self-assemble into organogels. Collectively, these results indicate that the fluorescent nucleolipids containing longer and saturated fatty acid acyl chains promote the formation of stable thermoreversible gels possibly due to enhanced hydrophobic and van der Waals interactions as compared to in nucleolipids containing shorter or unsaturated fatty acid acyl chains.[21,22]

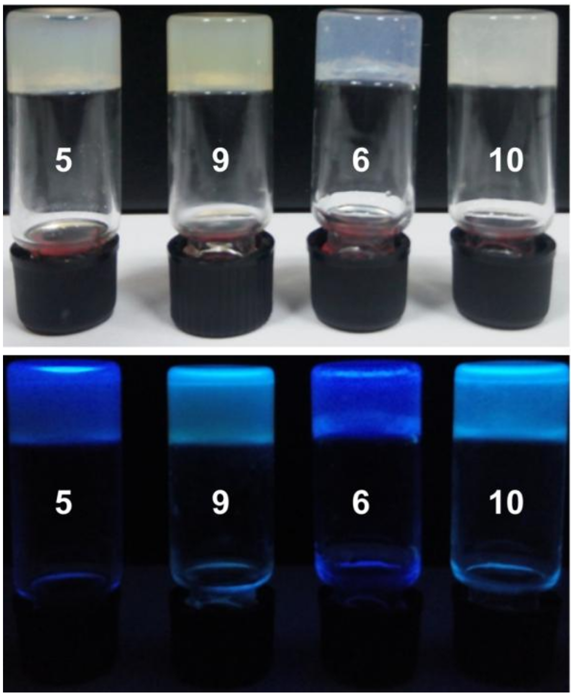


Figure 3. Top: Photograph of organogels of benzofuran-modified nucleolipids **5** and **6**, and benzothiophene-modified nucleolipids **9** and **10** in DMSO at the respective CGC values (Table 2). Bottom: Same samples under UV illumination (365 nm) indicate that these nucleolipids retain their fluorescence upon aggregation. This is in contrast to the majority of low molecular weight gelators, which lose their strong fluorescence upon self-assembling.

Table 2. Gelation properties of fluorescent nucleolipids **4–11**

nucleolipid in DMSO		state	CGC (w/v %)	T_{gel} (^oC)
benzofuran-modified nucleolipids	**4**	sol	-	-
	5	gel	4.2	32
	6	gel	0.9	46
	7	sol	-	-
benzothiophene-modified nucleolipids	**8**	sol	-	-
	9	gel	4.4	33
	10	gel	1.4	48
	11	sol	-	-

5.2.4 Rheological measurements of nucleolipid gels

The mechanical properties of nucleolipid organogels were studied by performing rheological measurements. The storage modulus (G′) and loss modulus (G″) of nucleolipid gels **6** and **10** in DMSO were monitored as a function of strain at constant oscillating frequency. Significantly higher G′ values of ~2480 Pa and ~4130 Pa over the G″ values of ~930 Pa and ~1780 Pa for nucleolipid gels **6** and **10**, respectively, at low strain values suggest the elastic behaviour of organogels (Figure 4). The crossover of G′ and G″ for **6** and **10** was observed at 1.8% and 1.3% of strain, respectively, where gel disintegrates into sol.

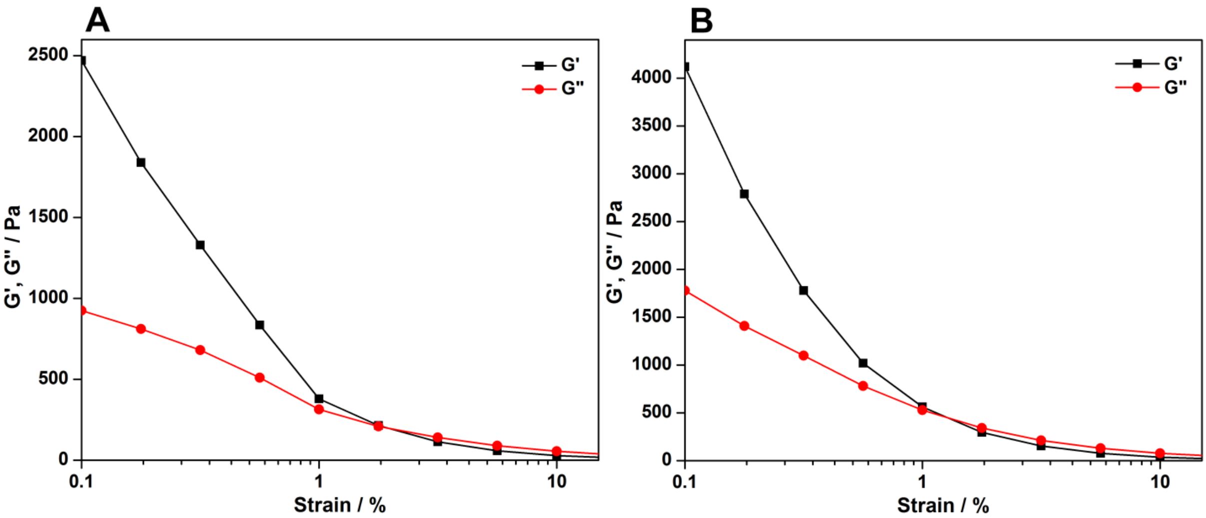


Figure 4. Strain sweep rheological measurements of nucleolipid gels **6** (**A**) and **10** (**B**) at respective CGC in DMSO at constant oscillating frequency.

5.2.5 Morphologies and characterization of organogels

Field emission scanning electron microscopy (FESEM) images of xerogels of nucleolipids revealed distinctly different morphologies and dimensions, which depended on the acyl chain length and type of modification on the nucleobase (Figure 5). Nucleolipids **5** and **9** containing myristoyl chains self-assembled into nanotubes, which were of several micrometers in length. Tubular structure of benzofuran-modified nucleolipid **5** had an outer diameter of approximately 360–420 nm and a wall thickness of nearly 70–90 nm (Figure 5A and B). The benzothiophene-modified nucleolipid **9** also formed nanotubes, however, with visible helical traces on the exterior (Figure 5C and D). The images of nanotubes of **9** revealed a similar outer diameter (380–410 nm) but a larger wall thickness (90–120 nm) as compared to **5**. Interestingly, the morphological transition from helical ribbons to nanotubes was occasionally observed in the SEM images of **9** (Figure 6). The width of helical traces on the exterior of the tubes were found to be reasonably larger than the intermediate helical ribbons, thereby indicating that the formation of tubular structures could be possibly due to widening of helical width resulting in the fusion of helical ribbons. Such morphological transformations due to the growth of helical width have been observed for low molecular weight amphiphiles.[22, 23] Surprisingly, nucleolipids **6** and **10** containing longer palmitoyl acyl chains formed predominantly entangled aggregates of twisted ribbons (width 130–160 nm) and fibres, respectively (Figure 5E and F). In general, gelators self-assembling into intertwined fibrous networks are known to have higher stability and gelation capacity as they can immobilize solvents effectively.[24] Hence, the fibrous morphology exhibited by the gels of **6** and **10** support the observed lower CGC and higher gel–sol transition temperature in comparison to the gels of nucleolipids **5** and **9** containing shorter myristoyl chains, which form higher ordered tubular structures.

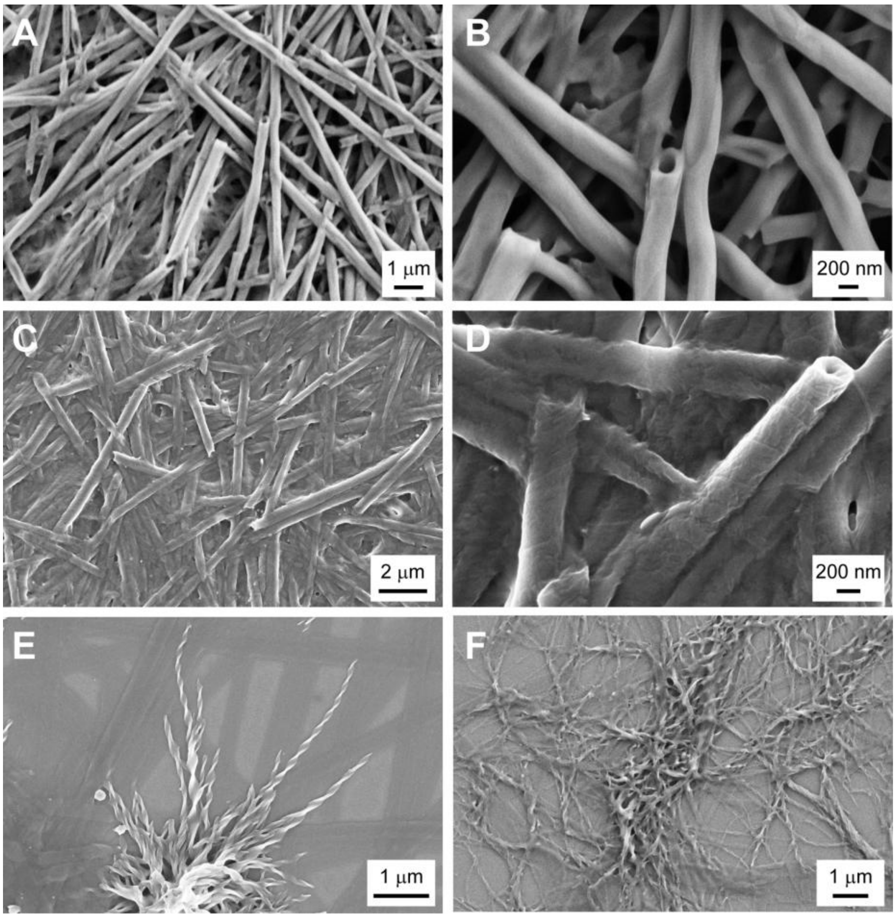


Figure 5. FESEM images of xerogels of nucleolipids. (**A** and **B**) Benzofuran-modified nucleolipid **5** containing myristoyl chains, (**C** and **D**) benzothiophene- modified nucleolipid **9** containing myristoyl chains, (**E** and **F**) benzofuran- and benzothiophene-modified nucleolipids **6** and **10**, respectively, containing palmitoyl chains. While myristoyl chain-containing nucleolipids **5** and **9** form tubular structures, palmitoyl chain-containing nucleolipids **6** and **10** form twisted ribbons and fibres, respectively.

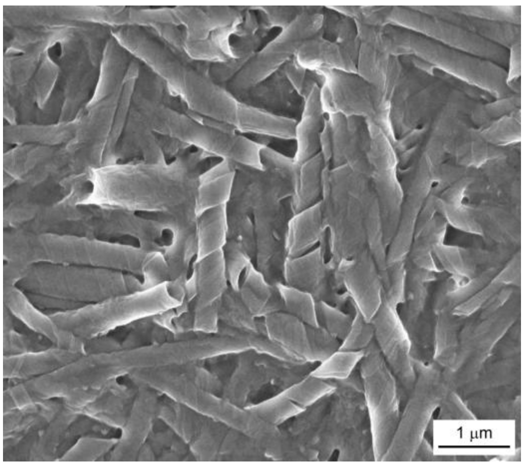


Figure 6. FESEM image of xerogel of benzothiophene-modified nucleolipid **9** showing helical ribbon intermediate structures, a possible precursor for nanotubes.

5.3 Driving force for hierarchical self-assembly

The chemical components of nucleolipids, namely the heterocyclic-conjugated uracil core, sugar residue and fatty acid acyl chains, could invoke various non-covalent interactions including H-bonding, π–π stacking, van der Waals and hydrophobic interactions to drive the self-assembly process. The FESEM images in part provide evidence for the formation of gels via the hierarchical self-assembly process driven by the cooperative action of the various non-covalent interactions mentioned above. To further understand the role played by individual components of nucleolipids in inducing gelation we performed single-crystal and powder X-ray diffraction (PXRD) and variable temperature ^{1}H NMR analyses.

5.3.1 Single-crystal X-ray diffraction analysis

Single crystals of 5-benzofuran-modified nucleolipids **5** and **6** were obtained from a chloroform–dimethylformamide mixture by the vapour diffusion method, while **5** benzothiophene- modified nucleolipids **9** and **10** were obtained from benzene by the slow evaporation method. The torsion angle (O4′-C1′-N1-C2) was measured to determine the orientation (*syn*/*anti*) of the nucleobase with respect to the sugar ring. As per IUPAC recommendations, torsion angle in the range of 180 ± 90° denotes anti conformation.[25] In all the structures the torsion angle was found to be in the range of 130°–237°, which confirmed that modified nucleobases adopted an *anti* conformation relative to the sugar ring (Table 3). The ribonucleosides usually adopt C3′-*endo* sugar pucker conformation. However, in the presence of the fatty acid acyl chain, the ribonucleolipids exhibited C3′-*exo* sugar puckering. Other than these common features the modification on the nucleobase (benzofuran and benzothiophene) had a significant impact on the structure and packing pattern of nucleolipids. The single crystal X-ray analysis of 5-benzofuran-modified nucleolipid **5** revealed the presence of two molecules of **5** in the asymmetric unit (Figure 7).

Table 3. H-bonding distances and angles, torsional angles and π–π interactions measured from the crystal structure of fluorescent nucleolipids **5**, **6**, **9** and **10**.

nucleolipid	hydrogen bond	distance (Å)	angle (°)	torsion angle (χ) (°)	π–π distance (Å)
5	N3AH---04B	2.011(5)	167.6(4)	C2A-N1A-C1′A-O4′A 149.7(4) C2B-N1B-C1′B-O4′B 130.0(4)	C11A----C12B 3.594(1) C12A----C13B 3.448(9)
	O2A---HN3B	1.990(5)	166.5(4)		
	O5′AH---O5′B	1.990(7)	154.2(5)		
	C6BH---O5′B	2.288(5)	157.6(4)		
6	N3AH---O4B	2.003(12)	166.9(1)	C2A-N1A-C1′A-O4′A 151.2(1) C2B-N1B-C1′B-O4′B 129.9(1)	C11A----C12B 3.583(2) C12A----C13B 3.468(2)
	O2A---HN3B	2.003(12)	168.3(1)		
	O5′AH---O5′B	1.945(16)	167.3(2)		
	C6BH---O5′B	2.296(13)	155.9(1)		
9	N3H---O29	2.034(1)	170.6(1)	C2-N1-C1′-O4′ 236.8(1)	C4----S1 3.387(2) C8----C13 3.390(2)
	O4′---HO5′	2.021(2)	162.3(1)		
	C6H---O5′	2.403(2)	157.6(1)		
10	N3H---O6	2.033(2)	168.5(1)	C2-N1-C1′-O4′ 237.2(2)	C4----S1 3.324(3) C8----C13 3.358(3)
	O4′---O5′[a]	2.792(3)	-		
	C6H---O5′	2.374(3)	159.4(1)		

[a]The position of hydrogen atom of 5′-OH could not be fixed satisfactorily and hence, O4′---O5′ distance is given.

The uracil and benzofuran rings were almost coplanar with only small differences in the mean-plane angle (~3°) between the planes of two rings. The Watson–Crick base pairing face of nucleolipids was disposed in such a fashion that each molecule in the unit cell base paired with an adjacent molecule of the immediate unit cell via strong H-bonds involving N3 imino hydrogens and C2 and C4 carbonyl oxygens (Figure 8 and Table 3). The adjacent base pairs were further connected to each other by a strong H-bond formed between the 5′-OH of one pair with the 5′-O of the other. Interestingly, the connection between the base pairs was also strengthened by an intramolecular non-canonical H-bonding between C6–H and 5′-O of the nucleoside and by partial π–π stacking interaction between the benzofuran rings [C11A C12B: 3.594(1) Å and C12A–C13B: 3.448(9) Å, Table 3]. These interactions essentially resulted in a one dimension (1D) layered structure with the base pairs arranged in a step-wise manner (Figure 9). Furthermore, a strong π–π stacking interaction between the uracil ring of one layer and the furan ring of the immediate layer, augmented by van der Waals interaction and hydrophobic effect between the myristoyl chains, generated a multilayered supramolecular architecture (Figure 10 and 11). The crystal structure of 5-benzofuranmodified nucleolipid **6** containing palmitoyl chains also exhibited a similar H-bonding pattern and structural organization (Figure 12–14 and Table 3).

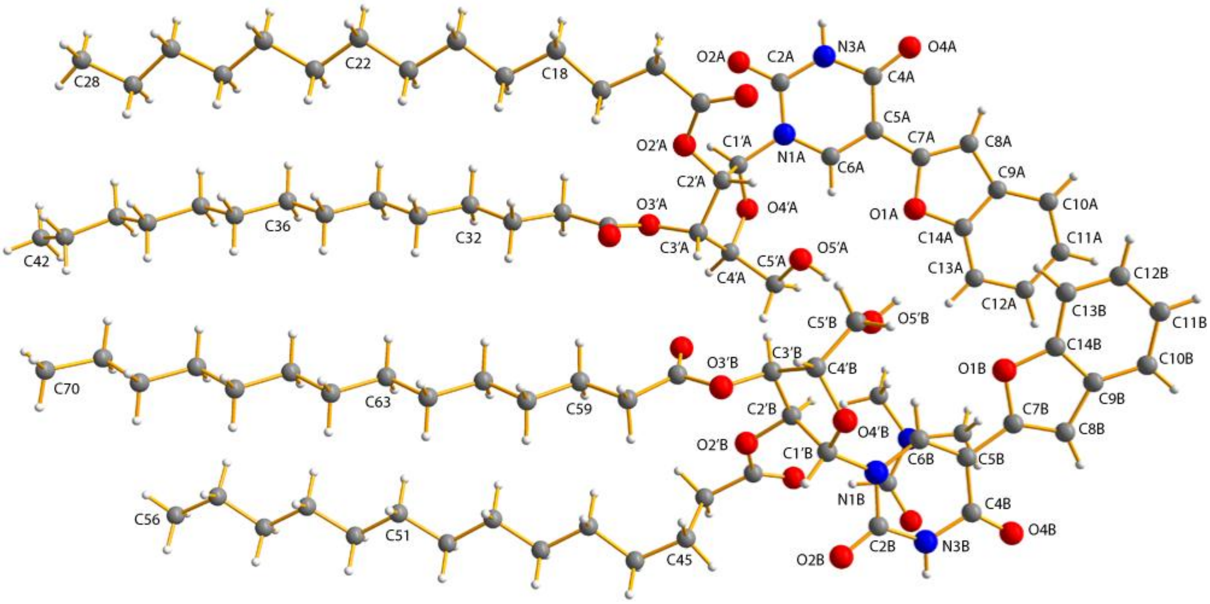


Figure 7. Crystal structure showing two molecules of 5-benzofuran-modified nucleolipid **5** in the asymmetric unit, which are designated as A and B. Solvent (DMF) has not been labeled for clarity. The uracil and benzofuran rings are almost coplanar. Modified nucleobase adopts an *anti* conformation relative to the sugar ring with C3′-*exo* sugar puckering. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

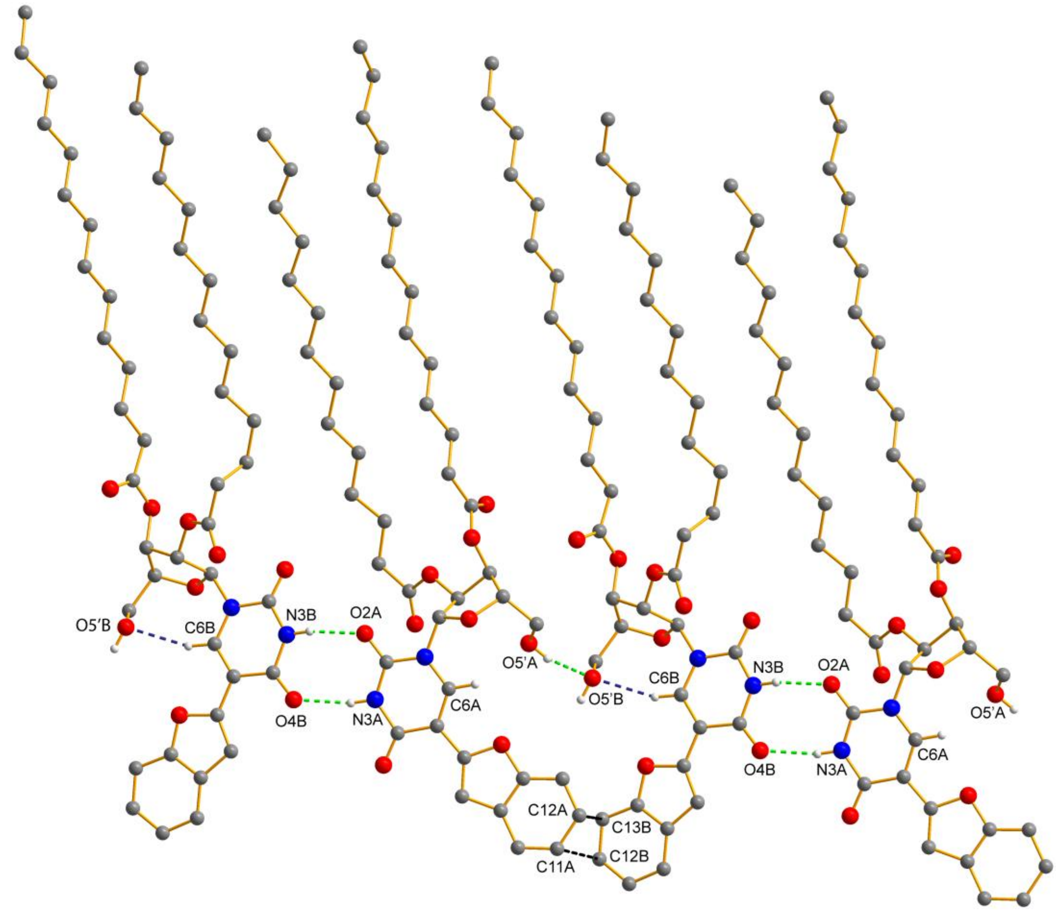


Figure 8. X-ray crystal structure showing a detailed view of the H-bonding and stacking interactions in 5-benzofuran-modified nucleolipid **5**. Canonical intermolecular H-bonding interactions (base pair and O5′AH---O5′B) are shown in green dashed lines. Intramolecular H-bond C6BH---O5′B is shown in dark blue dashed lines. Atoms involved in partial π–π stacking interaction are shown in black dashed lines. Hydrogen atoms other than the ones involved in H-bonding are not shown for clarity. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

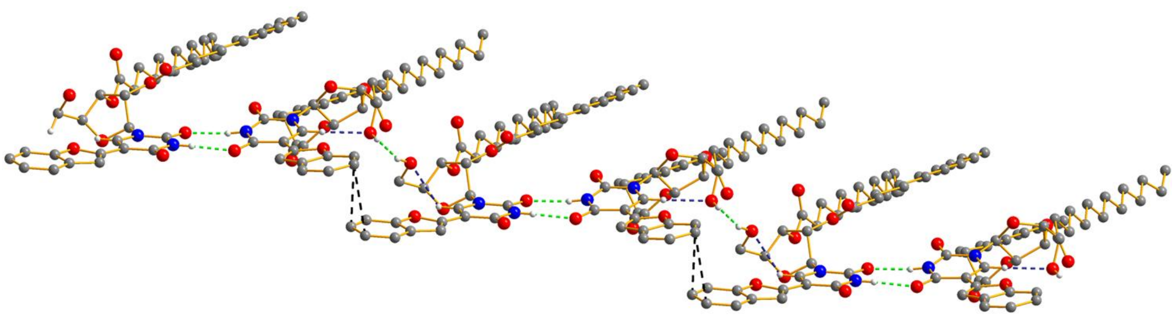


Figure 9. Structure of **5** showing 1D layered architecture with base pairs arranged in a step-wise manner. Hydrogen atoms are omitted for clarity and atom labels are same as in Figure 8.

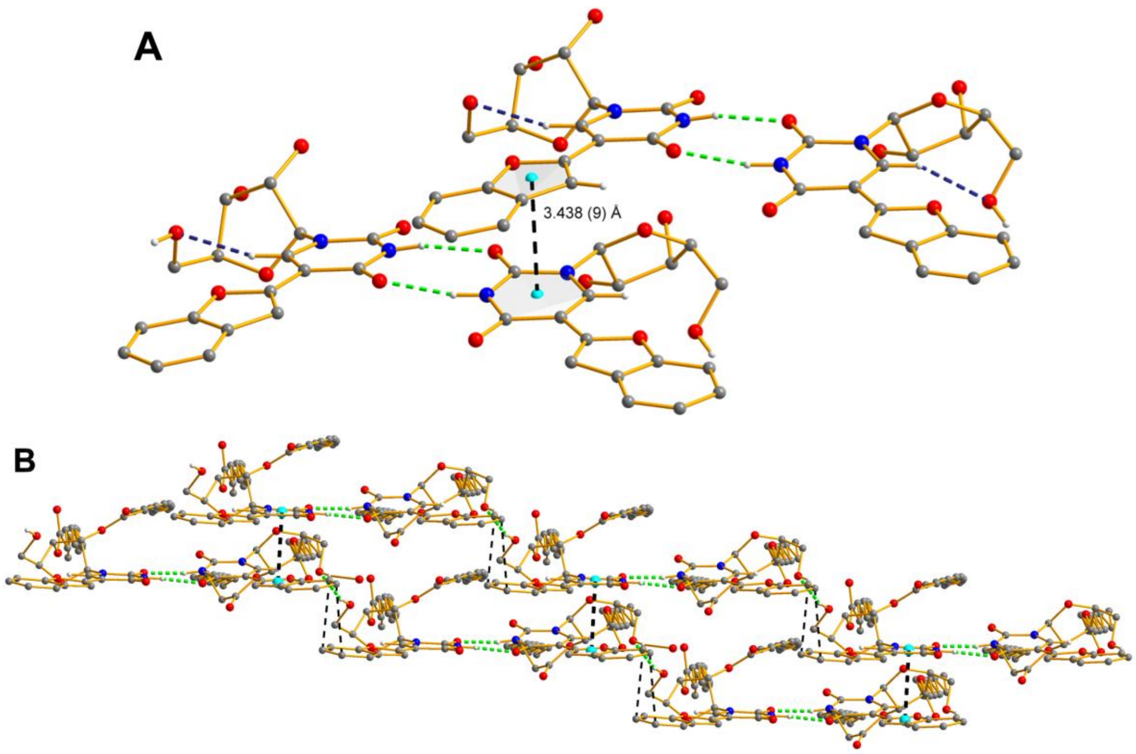


Figure 10. (**A**) X-ray crystal structure of nucleolipid **5** showing strong π–π stacking interaction between the uracil ring of one layer and furan ring of benzofuran moiety of the next layer. π–π stacking distance (3.438 Å) between the uracil and furan rings is shown in black dashed lines. (**B**) Layered supramolecular architecture formed by various non-covalent interactions. Hydrogen atoms are omitted for clarity and atom labels are same as in Figure 8.

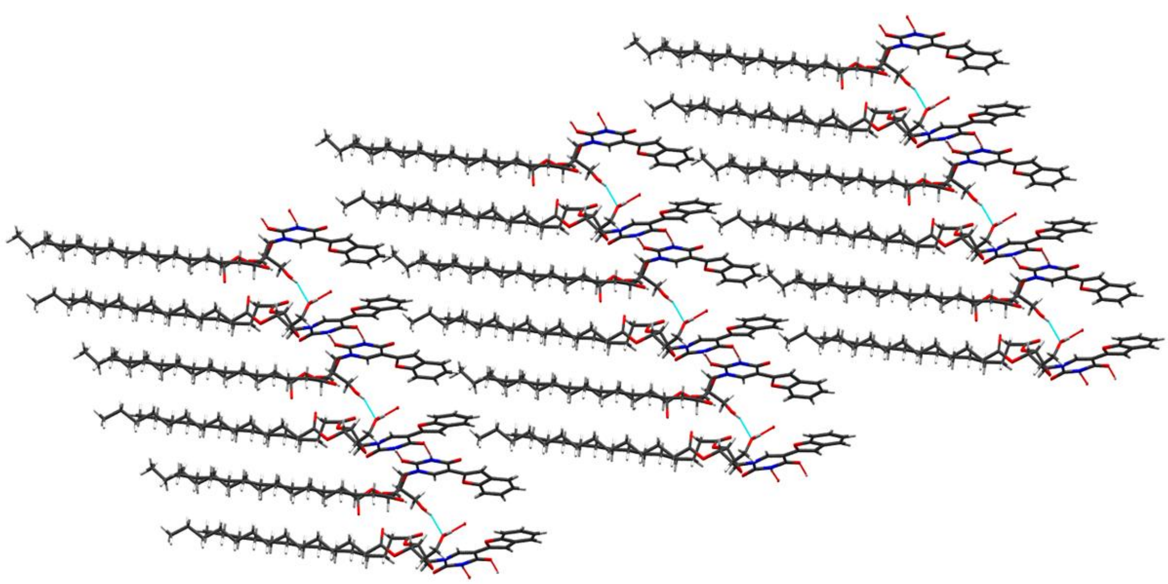


Figure 11. Packing diagram of nucleolipid **5** along the crystallographic a-axes.

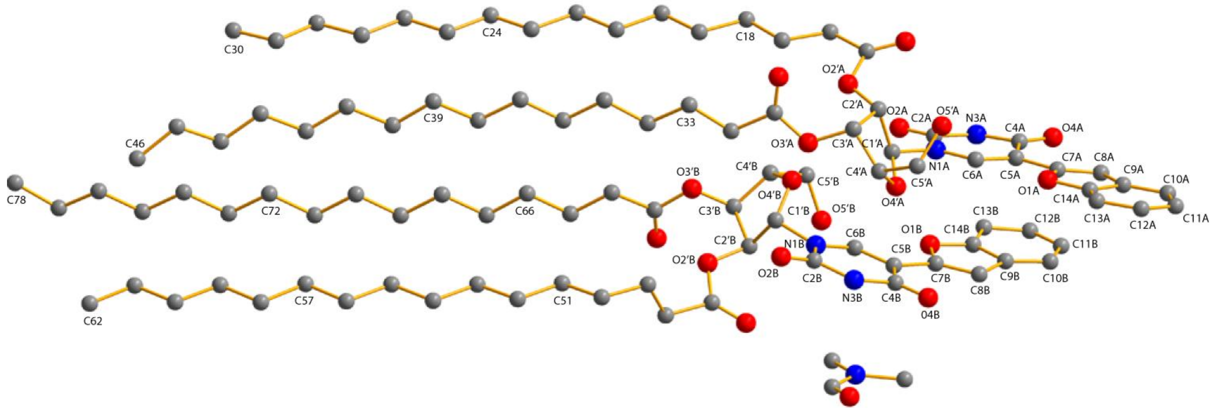


Figure 12. X-ray crystal structure of 5-benzofuran-modified nucleolipid **6** containing palmitoyl acyl chains. Crystal structure shows two molecules of **6** in the asymmetric unit, which are designated as A and B. Solvent (DMF) has not been labeled for clarity. Akin to **5**, the uracil and benzofuran rings in **6** are almost coplanar. Modified nucleobase adopts an *anti* conformation relative to the sugar ring with C3′-*exo* sugar puckering. Hydrogen atoms were omitted for clarity. Atoms are coded as follows: dark gray, carbon; blue, nitrogen; red, oxygen.

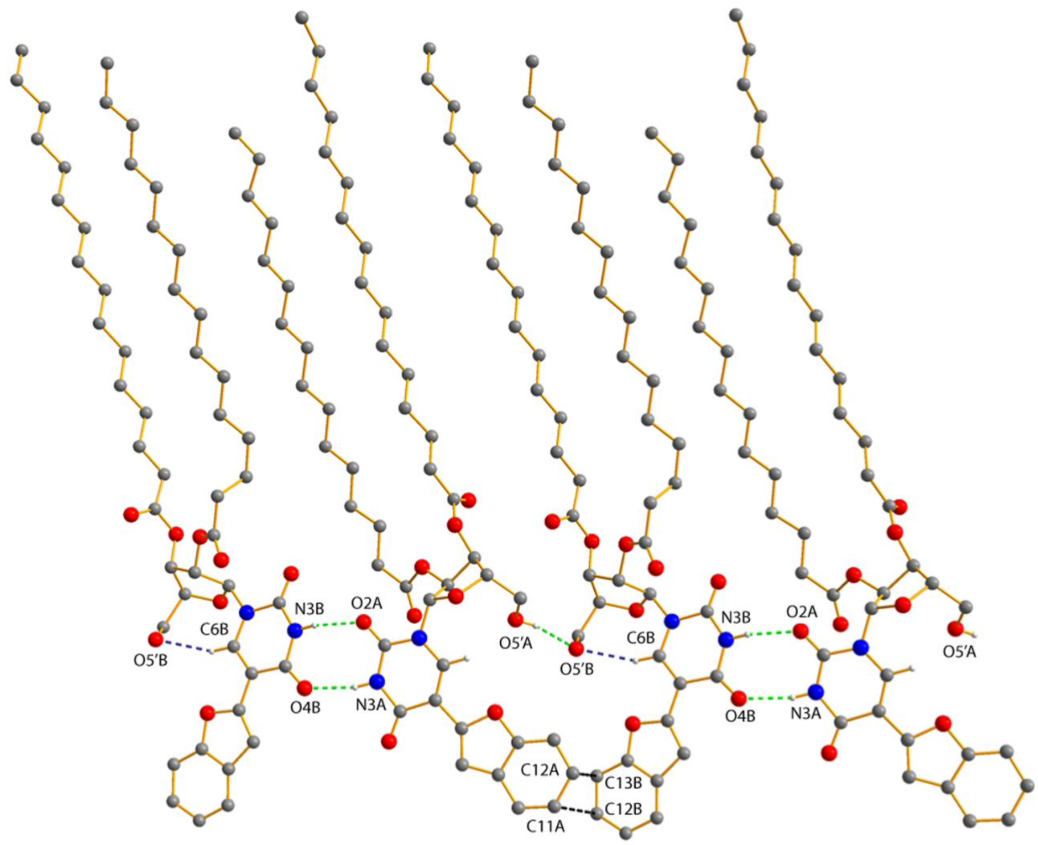


Figure 13. X-ray crystal structure of nucleolipid **6** showing a detailed view of the H-bonding and π–π interactions. Canonical intermolecular H-bonding interactions (base pair and O5′AH---O5′B) are shown in green dashed lines. Intramolecular H-bond C6BH---O5′B is shown in dark blue dashed lines. Atoms involved in partial π–π stacking interaction (C11A-C12B and C12A-C13B) are shown in black dashed lines. Hydrogen atoms not involved in H-bonding are omitted for clarity. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen.

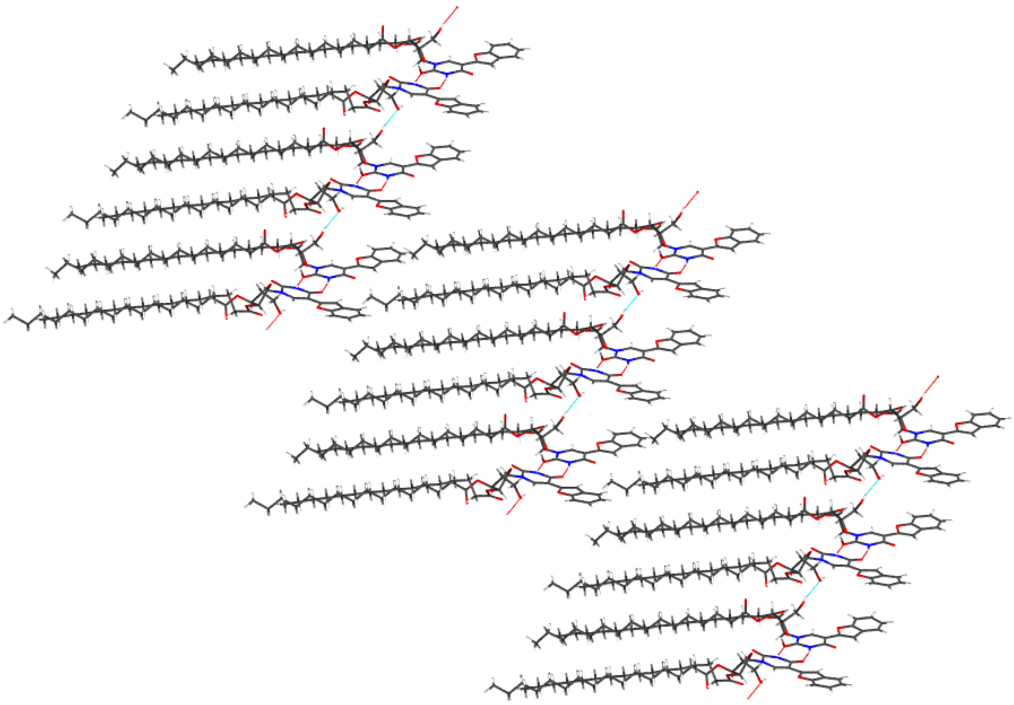


Figure 14. Packing diagram of nucleolipid **6** along the crystallographic a-axes.

The crystal structures of 5-benzothiophene-modified nucleolipids **9** and **10** revealed markedly different structural features as compared to benzofuran-modified nucleolipids. Notably, the asymmetric unit contained only one molecule, which was significantly bent, the Watson–Crick base pairing was absent, and the uracil and benzothiophene rings were not coplanar (Figure 15). The mean-plane angle between the planes of these two rings was found to be nearly ~18°. Molecules of **9** were connected in a series by unconventional intermolecular hydrogen bonds involving furanose sugar O4′ and 5′-OH atoms to form a 1D chain in which the modified nucleobases were arranged in a step-wise fashion (Figure 16 and Table 3). The chain was also stabilized by a partial π–π stacking interaction between the aromatic systems of adjacent molecules and intramolecular H-bonding between C6–H and 5′-O of the nucleoside. The individual chains were further extended in 2D by strong H-bonding between nucleobase N3H and C3′ ester carbonyl oxygen atom resulting in a multilayered structure (Figure 16 and 17). Expectedly, nucleolipid **10** containing longer fatty acid acyl chains displayed a similar H-bonding pattern and structural features as that of nucleolipid **9** (Figure 18 and 19).

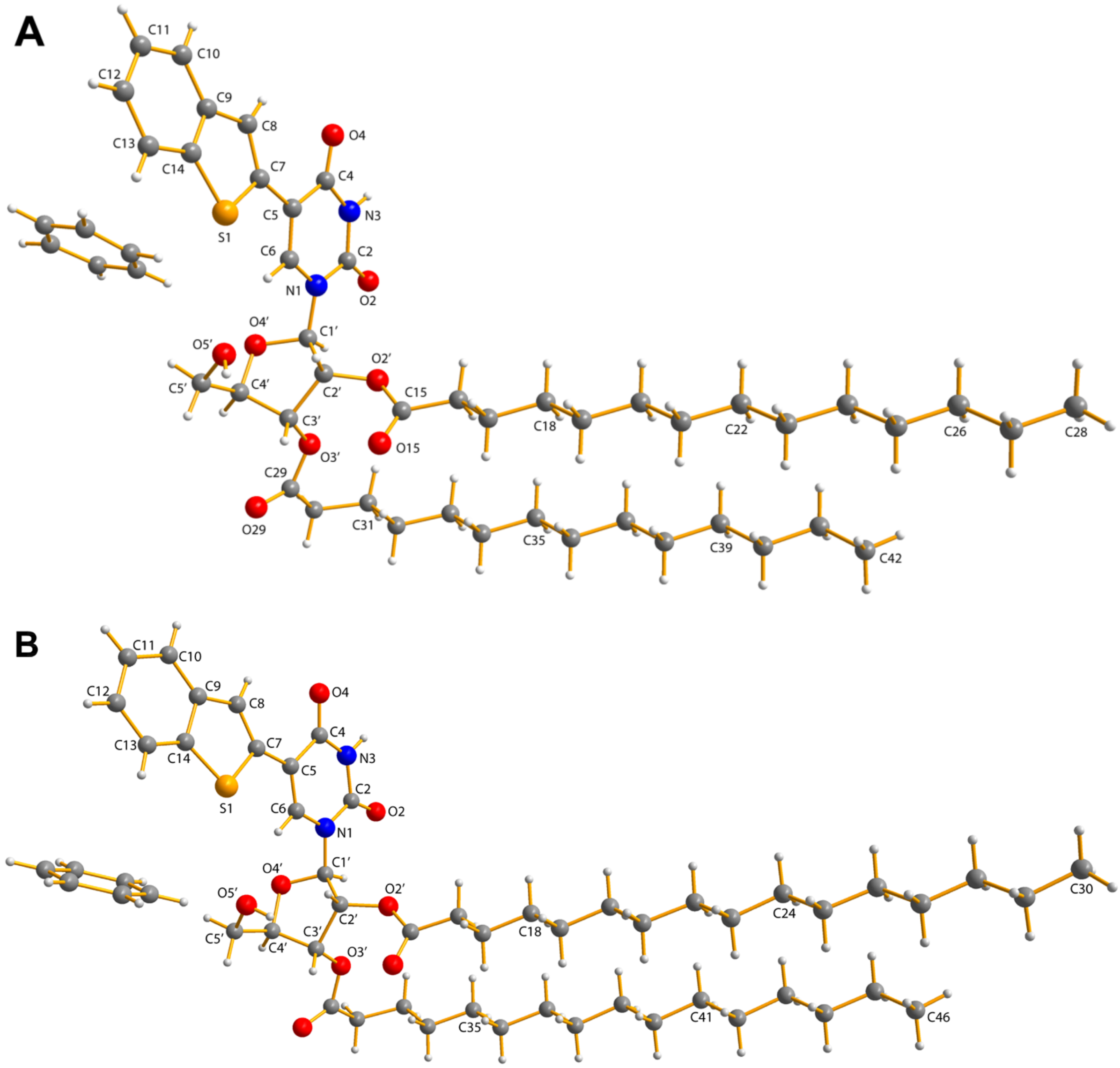


Figure 15. X-ray crystal structures of 5-benzothiophene-modified nucleolipid **9** (**A**) and **10** (**B**) showing one molecule in the unit cell. Uracil and benzothiophene rings are not coplanar. Modified nucleobase adopts an *anti* conformation relative to the sugar ring with C3′-*exo* sugar puckering. Atoms are coded as follows: off white, hydrogen, dark gray, carbon; blue, nitrogen; red, oxygen; golden yellow, sulfur.

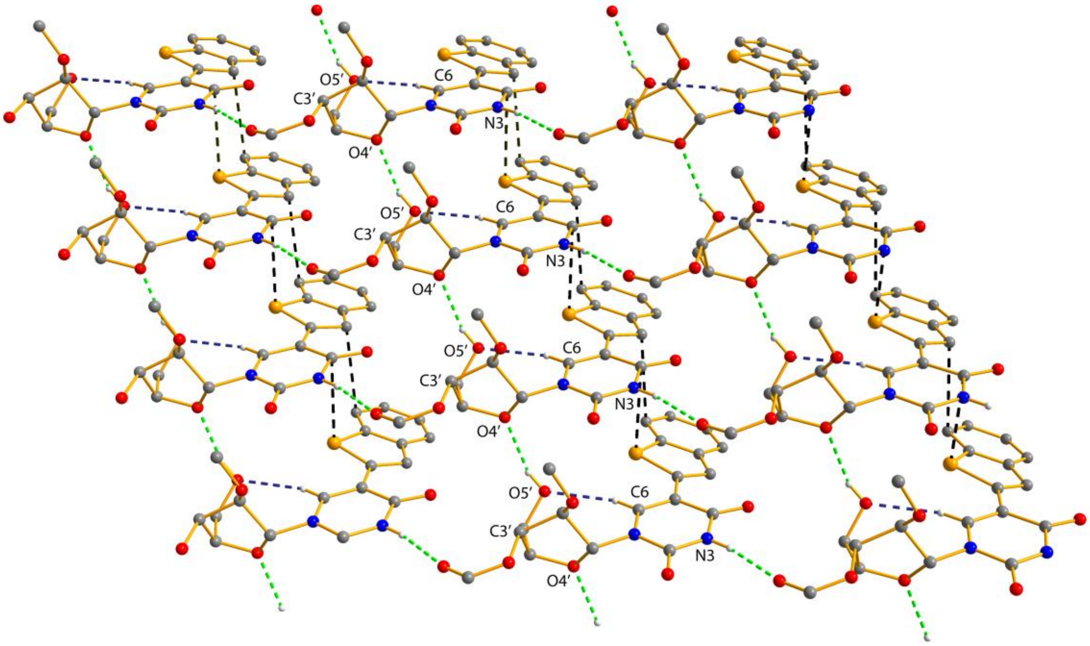


Figure 16. X-ray crystal structure showing a detailed view of the H-bonding and stacking interactions in 5-benzothiophene-modified nucleolipid **9**. Intermolecular H-bonding interactions are shown in green dashed lines. Intramolecular H-bond C6H---O5′ is shown in dark blue dashed lines. Atoms involved in partial π–π stacking interaction are shown in black dashed lines. See Table 3 for details. Hydrogen atoms not involved in H-bonding and part of the fatty acid acyl chains have been omitted for clarity. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen; golden yellow, sulfur.

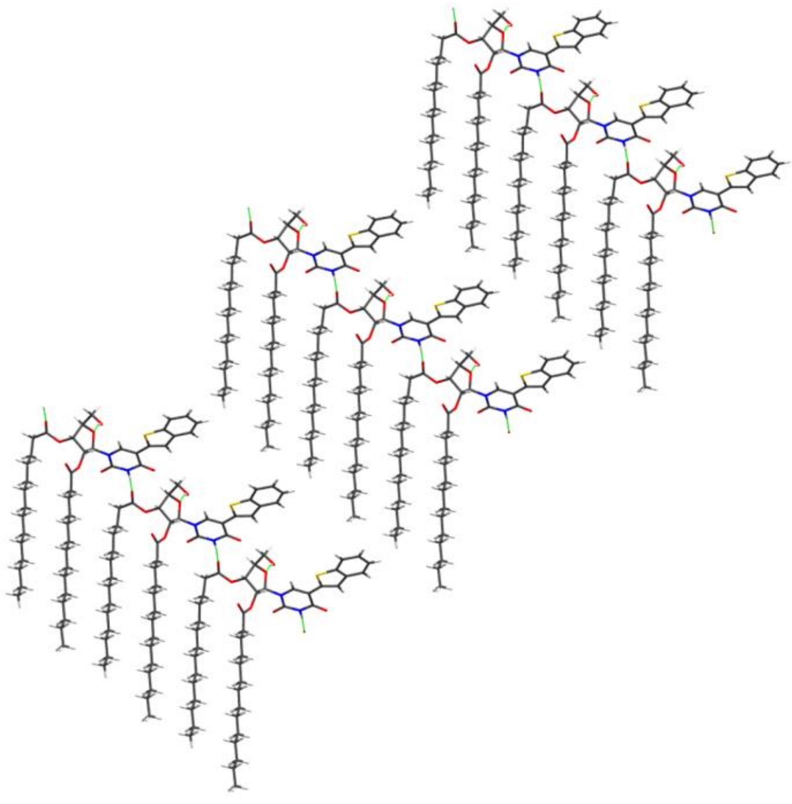


Figure 17. Packing diagram of nucleolipid **9** along the crystallographic a-axes.

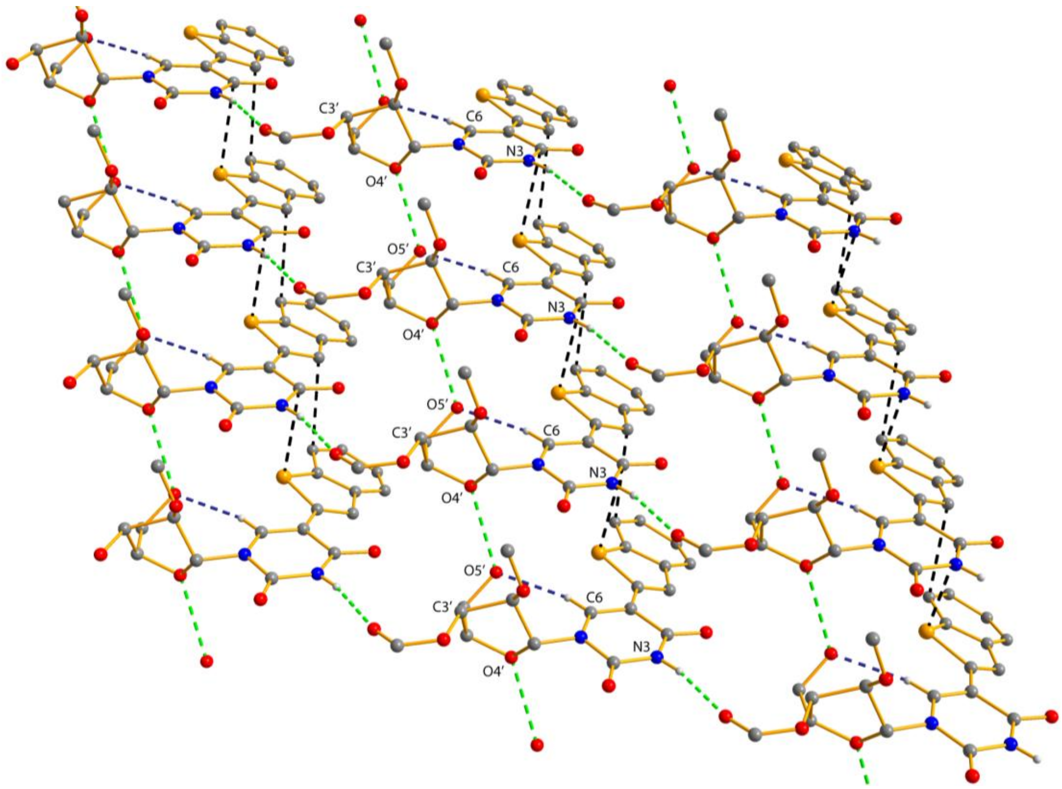


Figure 18. X-ray crystal structure showing a detailed view of the H-bonding and π–π interactions in 5- benzothiophene-modified nucleolipid **10**. Intermolecular H-bonding interactions are shown in green dashed lines. Intramolecular H-bond C6H---O5′ is shown in dark blue dashed lines. Atoms involved in partial π–π stacking interaction are shown in black dashed lines. See Table 3 for details. Hydrogen atoms not involved in H-bonding and part of the fatty acid acyl chains have been omitted for clarity. Atoms are coded as follows: off white, hydrogen; dark gray, carbon; blue, nitrogen; red, oxygen; golden yellow, sulfur.

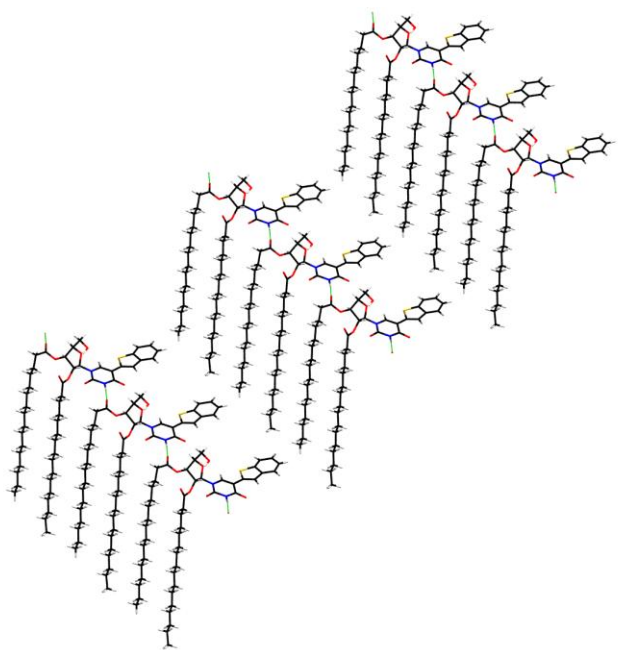


Figure 19. Packing diagram of nucleolipid **10** along the crystallographic a-axes.

The differences in crystal structures and packing of benzothiophene-and benzofuran-modified nucleolipids, especially the conformation around the uracil and heterobicycle rings could be due to the larger size of the sulfur atom. The steric hindrance between the sulfur atom and C6 hydrogen atom in **9** and **10** renders these rings in a nonplanar geometry as compared to nucleolipids **5** and **6** in which the benzofuran and uracil rings are almost coplanar. This nonplanar geometry possibly prevents base pairing in **9** and **10**, but facilitates partial π–π stacking interaction between the aromatic system of adjacent molecules involving sulfur and carbon atoms (Figure 16). The influence of sulfur atom on the planarity and packing in the crystal structures of heterocycle containing molecules has been reported earlier.[26]

5.3.2 Variable temperature ^{1}H NMR analysis

To ascertain if the H-bonding interactions, which were present in the crystal packing structures were also present in the gel state, we performed variable temperature ^{1}H NMR experiment using **6** and **10** having lower CGC values. As the temperature of gel was increased from 25 °C to 55 °C, the N3 imino H and 5′-OH exhibited significant upfield shift in their proton signals in the sol state ($\Delta\delta$ = 0.13 and 0.16 ppm, respectively, Figure 20). C6H of nucleobase, which formed non-canonical C-H-O H-bond with 5′-O atom of ribose sugar in the crystalline state, also showed small upfield shift during gel-sol transition ($\Delta\delta$ = 0.06 ppm). The progressive shift in NMR signal of the protons mentioned above is possibly due to loosening or breaking of the respective H-bonds as the gel is converted to solution. These observations indicate that the H-bonding interaction partners in crystal structures and in supramolecular gels are same.

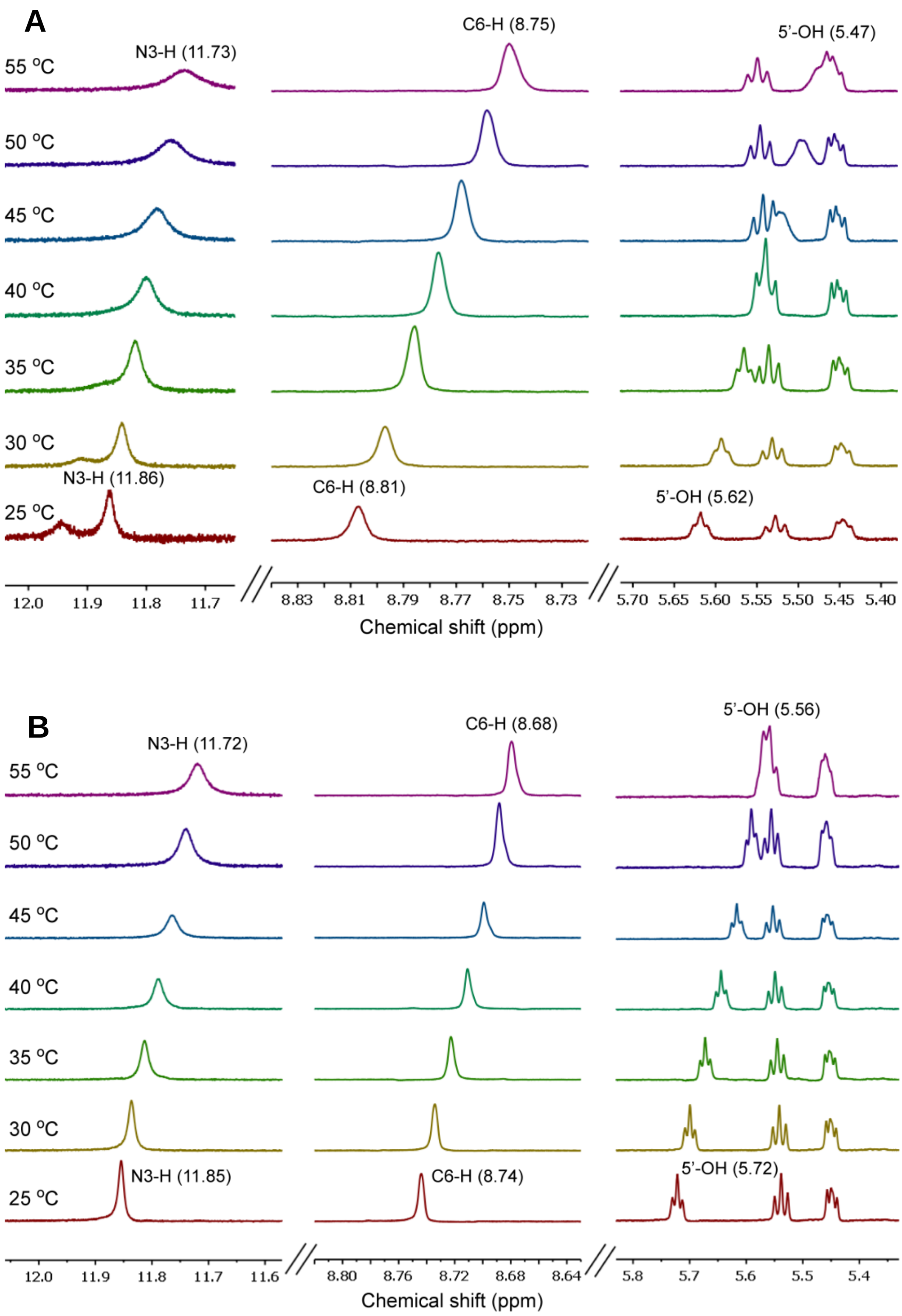


Figure 20. Partial ^{1}H NMR spectra of nucleolipid gels of **6 (A)** and **10 (B)** in *d6*-DMSO as a function of increasing temperature. N3-H and 5′-OH, which participated in strong intermolecular H-bonding in crystalline state, exhibited significant upfield shift in their proton signals during gel-sol transition. C6H of nucleobase, which participated in a weak intramolecular C-H-O H-bonding interaction, also showed small upfield shift during gel-sol transition. These results indicate that the H-bonding interactions in crystal structure and in supramolecular gel are similar. See the experimental section**.**

5.3.3 Powder X-ray diffraction (PXRD) analysis

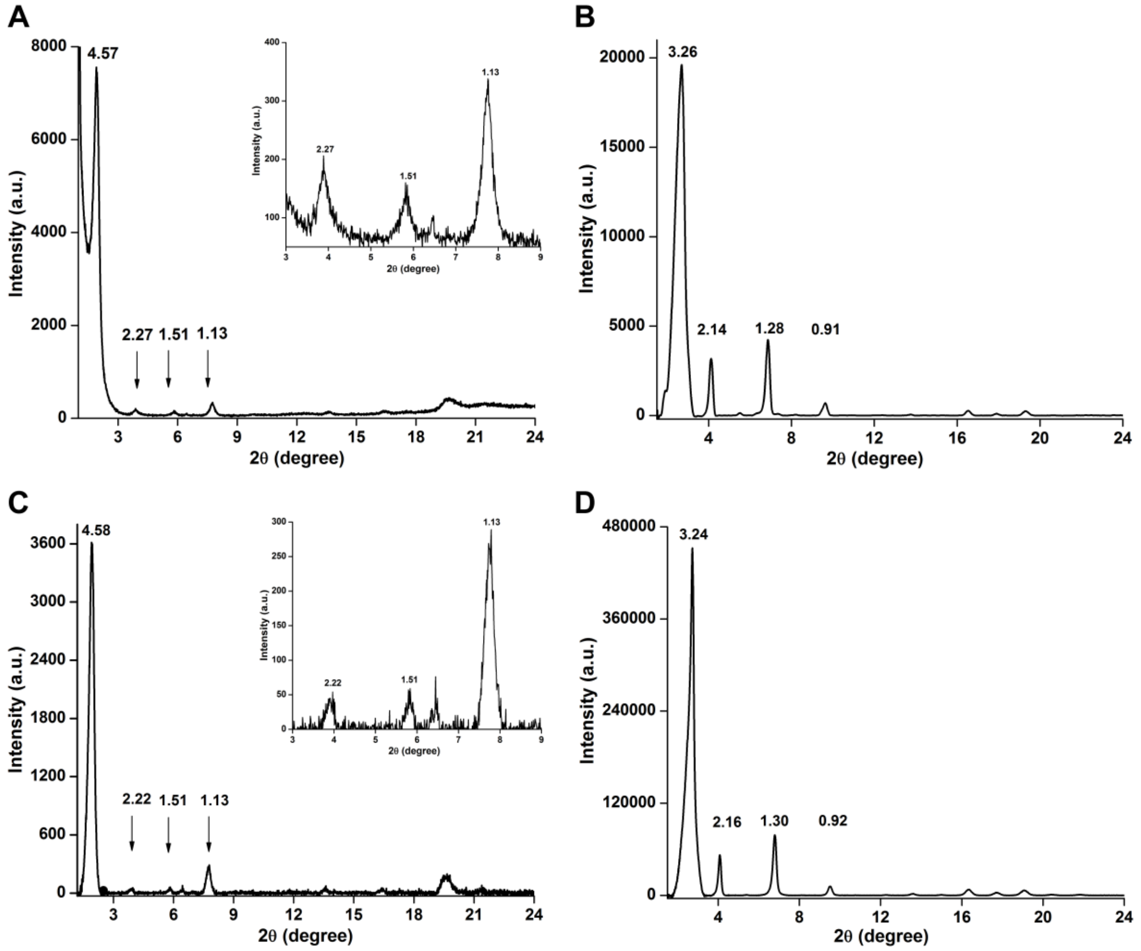


Figure 21. PXRD spectra of xerogels of (**A**) **5**, (**B**) **6**, (**C**) **9**, (**D**) **10**. Inset in (**A**) and (**C**): peaks in the 2θ value range of 3–9° have been magnified. Layer spacing (nm) for prominent diffraction peaks are also given. For details see experimental section.

The PXRD data obtained from xerogels of benzofuran- and benzothiophene-modified nucleolipids (**5** and **9**) containing myristoyl chains showed diffraction peaks corresponding to layer spacings in the ratio 1:1/2:1/3:1/4 (Figure 21). This ratio of d-spacing is indicative of an ordered lamellar structure, which could have further promoted the formation of hierarchical self-organized structures, namely ribbons, helical ribbons and finally nanotubes. Such a mechanism for the formation of nanotubes by low molecular weight self-assembling systems has been well documented.[22,27] In the case of xerogels of **6** and **10** containing longer palmitoyl chains, the PXRD data did not reveal a lamellar packing arrangement (Figure 21B and D). It is likely that the longer palmitoyl chains in **6** and **10** could have altered the balance between non-covalent interactions exerted by the modified nucleoside head group and fatty acid tail resulting in fibrous networks. Taken together, these analyses clearly revealed the synergistic role of various non-covalent interactions invoked by the modified nucleobase, and the sugar and fatty acid acyl chain in setting up the path for the self-assembly process.

5.4 AIEE behaviour of nucleolipids

Most often fluorescent organic molecules used as building blocks in constructing supramolecular structures lose their strong fluorescence upon self-assembly, which greatly restrict their practical applications.[28] Therefore, self-assembling organic fluorophores that retain or exhibit enhanced fluorescence upon aggregation are promising candidates for fabricating optical materials and sensors.[29] In the chromophore system of our nucleolipids the heterobicyclic moiety and uracil rings are attached via a rotatable aryl-aryl bond, which upon rigidification could show enhancement in fluorescence. To test this behaviour, the fluorescence of nucleolipids at respective CGC was recorded in the assembled and disassembled states. Gels of benzofuran-modified nucleolipids **5** and **6** retained their fluorescence efficiency in gel as well as in the sol state (Figure 22). In **5** and **6** the strong base pairing interaction between adjacent nucleobases, which was observed in the crystal structure (Figure 8 and 13) could have rigidified the fundamental unit responsible for hierarchical self-assembly, thereby resulting in only minor changes in the fluorescence intensity in gel and sol states.

Interestingly, as the temperature of gels of benzothiophenemodified nucleolipids **9** and **10** was increased, a progressive decrease in the fluorescence intensity was observed (Figure 23). These nucleolipids in the gel state displayed 3- to 4-fold higher fluorescence intensity as compared to in the sol state. This effect was found to be significantly reversed as hot nucleolipid solutions were gradually cooled to RT to form gels (Figure 24). Importantly, in a control experiment the fluorescence of nucleolipids at well below the CGC was not affected by changes in temperature (Figure 25). In the crystal packing of **9** and **10** the benzothiophene and uracil rings are not coplanar and the nucleobases are not base paired (Figure 16 and 18). Hence, this enhancement in emission in the gel state could be related to AIEE, which is more likely due to the rigidification of the fluorophore, i.e., planarization of benzothiophene and uracil rings during the self-assembly process.

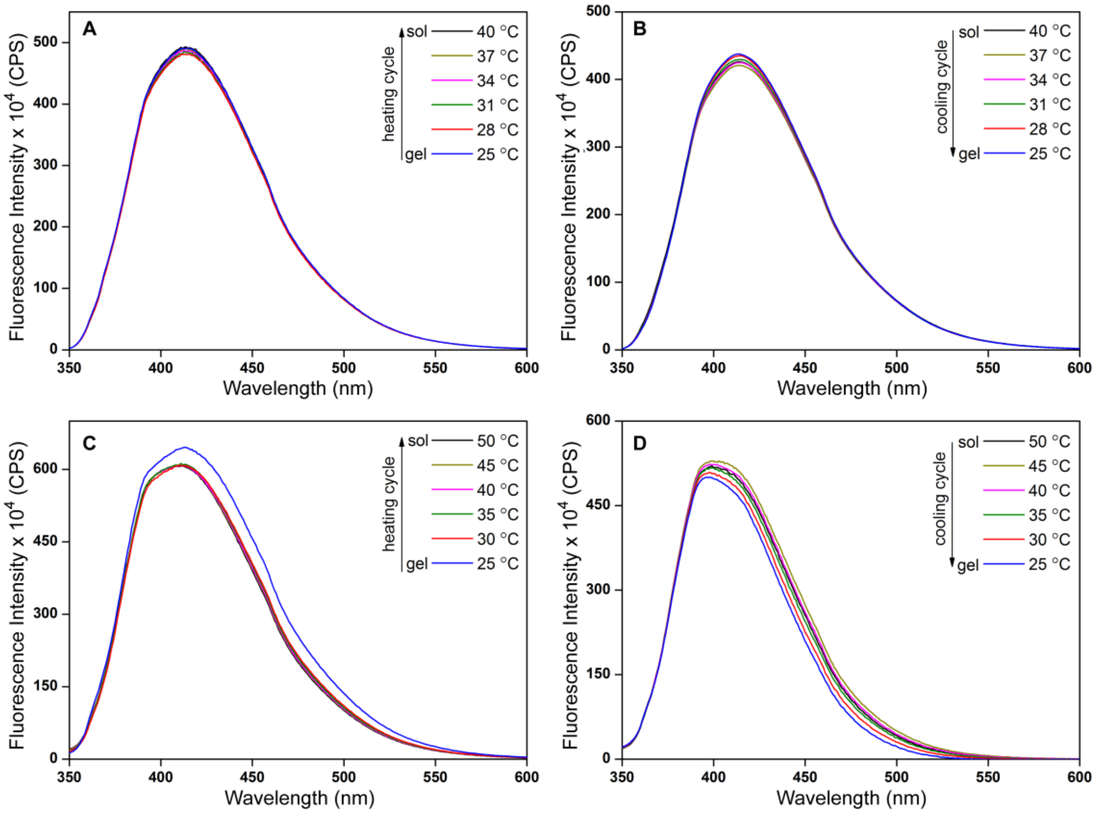


Figure 22. Fluorescence spectra of benzofuran-modified nucleolipid gels **5** and **6** at respective CGC as a function of temperature. Fluorescence spectra of **5** during heating (**A**) and cooling (**B**) cycles. Fluorescence spectra of **6** during heating (**C**) and cooling (**D**) cycles. Gels of **5** and **6** retained their fluorescence intensity in gel and solution states during heating and cooling cycles. Samples were excited at 320 nm with excitation and emission slit widths of 2 nm.

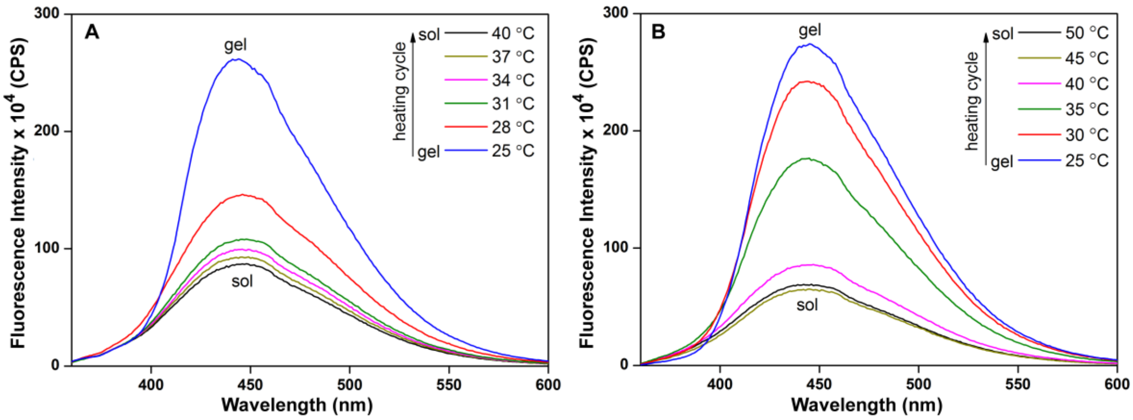


Figure 23. Fluorescence spectra of benzothiophene-modified nucleolipid gel **9** (**A**) and **10** (**B**) at respective CGC as a function of increasing temperature. Samples were excited at 320 nm with an excitation slit width of 2 nm and emission slit width of 2 nm. As the temperature was increased a significant decrease in the fluorescence intensity was observed.

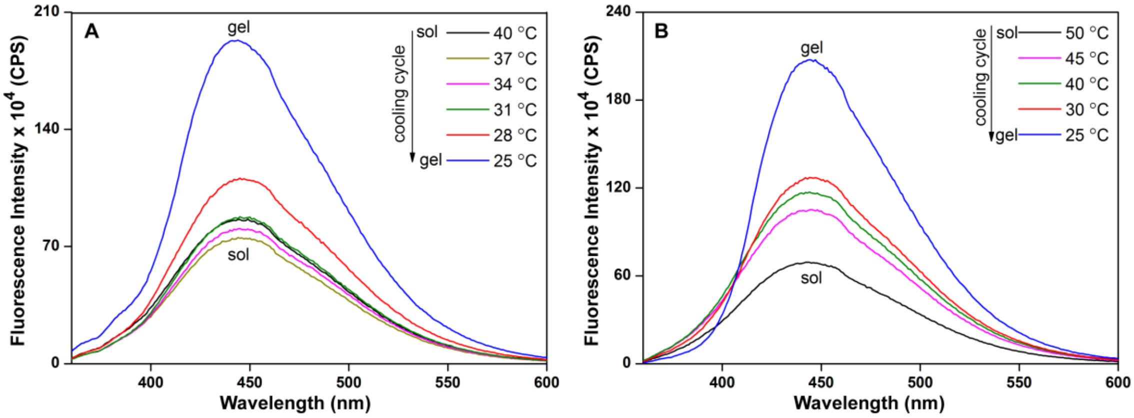


Figure 24. Fluorescence spectra of benzothiophene-modified nucleolipid gel **9** and **10** at respective CGC as a function of temperature. Fluorescence spectra of **9** (**A**) and **10** (**B**) during cooling cycle. As the nucleolipid solutions were gradually cooled to form the gels, a significant increase in fluorescence intensity was observed. Samples were excited at 320 nm with excitation and emission slit widths of 2 nm.

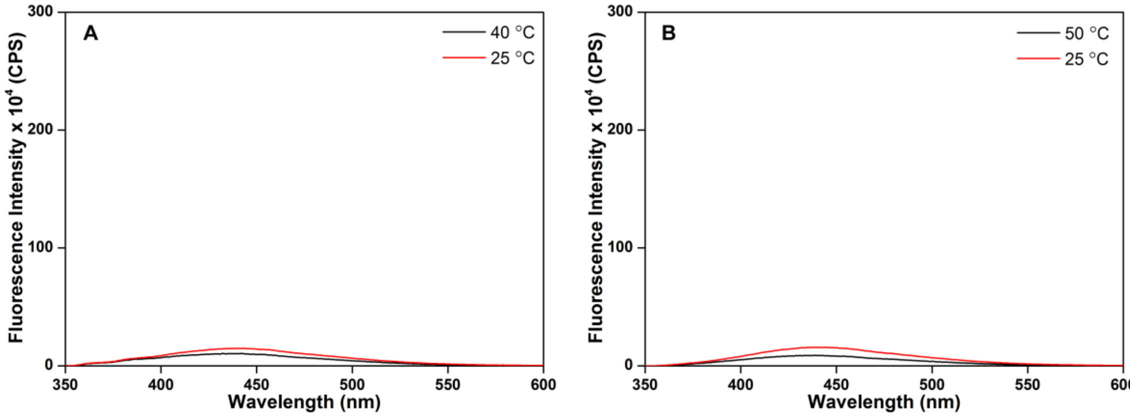


Figure 25. Fluorescence spectra of nucleolipid **9** (**A**) and **10** (**B**) at different temperatures. The fluorescence of nucleolipids (5 μM) at well below the CGC was not affected by changes in temperature. Samples were excited at 320 nm with excitation and emission slit widths of 2 nm. The Y-axis scale has been kept similar to that of Figure 23 for better comparison.

5.5 Multistimuli-responsiveness of nucleolipid gels

Altering the physicochemical properties of atoms involved in the non-covalent interactions by using external stimuli (e.g., physical and chemical) has enabled the tuning of gelation process.[20,30,31] In this context, multistimuli-responsive supramolecular gels whose fluorescence properties are sensitivity to the assembling and disassembling process are highly suitable for developing smart materials and sensors.[32] It is evident that the gelation process of our fluorescent nucleolipids is assisted by multiple H-bonds, in particular involving ionisable N3H and 5′-OH atoms. Hence, we decided to study the effect of multiple external stimuli namely, sonication, temperature, anions, acid/base and metal ions, on the gelation behaviour of nucleolipids. Moreover, metal ions could also be used to modulate the morphology of supramolecular self-assemblies. For example, uridine phosphocholine amphiphiles, which form fibers in water, can be converted into hollow microspheres in the presence of actinides and lanthanides.[33]

The sensitivity of nucleolipid gels to physical stress, sonication and temperature, has already been described in the above sections (vide supra). Briefly, while sonication of a hot solution of nucleolipids promoted rapid formation of gels, heating–cooling cycles resulted in a reversible gel–sol phase transition (Figure 23 and 26). To study the responsiveness of fluorescent nucleolipids to chemical stimuli, we chose the benzothiophene- modified nucleolipid **10**, which has low CGC value and shows AIEE, as the test system.

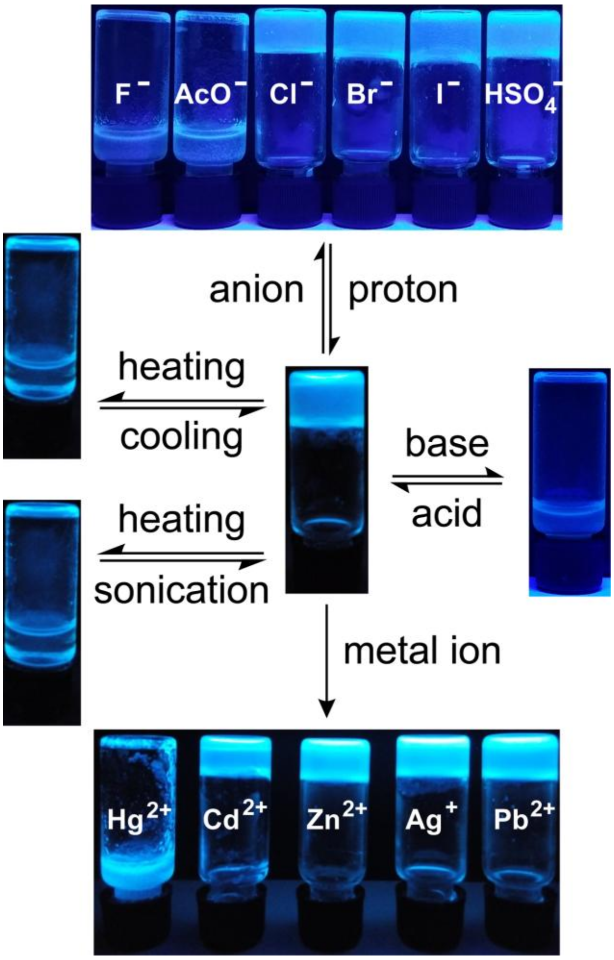


Figure 26. Responsiveness of benzothiophene-modified nucleolipid gel **10** to physical (temperature and sonication) and chemical stimuli (anion, acid-base, metal ion). Photographs have been taken under UV illumination (365 nm), which clearly show the changes in fluorescence upon application of external stimuli. See details experimental section.

The responsiveness of gelator **10** to various anions was evaluated by using tetrabutylammonium salts of F^-, AcO^-, Cl^-, Br^-, I^- and HSO_4^-. Addition of F^- or AcO^- (0.5

equiv.) to a hot solution of nucleolipid **10** completely arrested the formation of gel (Figure 26). Furthermore, addition of F^- (1 equiv.) to the pre-formed gel of **10** at RT (concentration of gelator was maintained above CGC) slowly disintegrated the gel to solution. However, other anions did not affect the gelation process even at a high concentration (5 equiv.) of the anions. It is likely that the deprotonation of N3H and 5′-OH by F^- and AcO^- anions, which are more basic than other anions tested, would have prevented the H-bond formation thereby arresting the gelation process. This assumption was confirmed by adding proton source (0.5 equiv. acetic acid) to a degelated sample containing F^- (0.5 equiv.), which reversed the effect of F^- and promoted gelation. Moreover, addition of F^- to gel resulted in the disappearance of N3H and 5′-OH proton NMR signals and appearance of a new broad signal at 15.9 ppm corresponding to HF_2^-, which also confirmed the deprotonation by F^- (Figure 27).[34] Interestingly, gel-sol transition in the presence of F^- and AcO^- could be easily visualized by changes in fluorescence (Figure 26 and 28A).

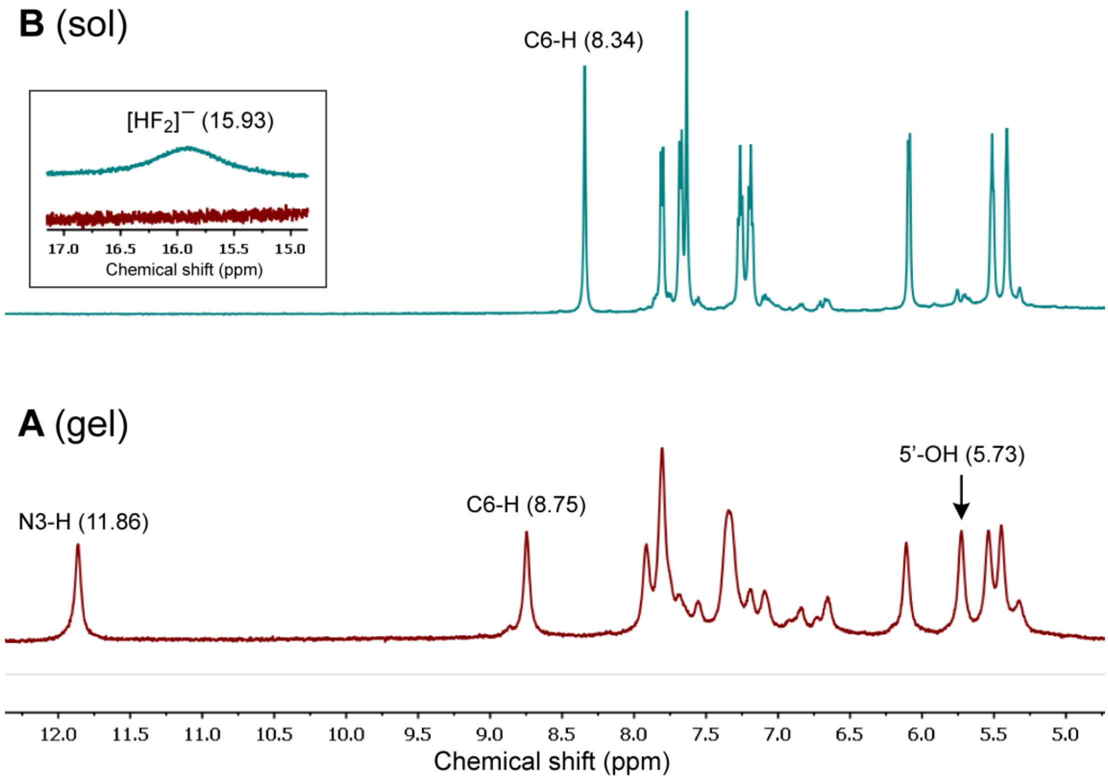


Figure 27. Partial ^{1}H NMR spectrum of nucleolipid **10** in *d6*-DMSO (1.4 w/v %) in the absence (**A**, gel state) and presence of 3 equivalence of fluoride ion (**B**, sol state). The disappearance of N3H (11.86 ppm) and 5′-OH (5.73 ppm) signals and appearance of a new broad signal at 15.93 ppm corresponding to HF_2^- confirms the deprotonation of imino and 5′-O hydrogen atoms by fluoride anion. ^{1}H NMR spectra were obtained using Bruker 500 MHz spectrometer.

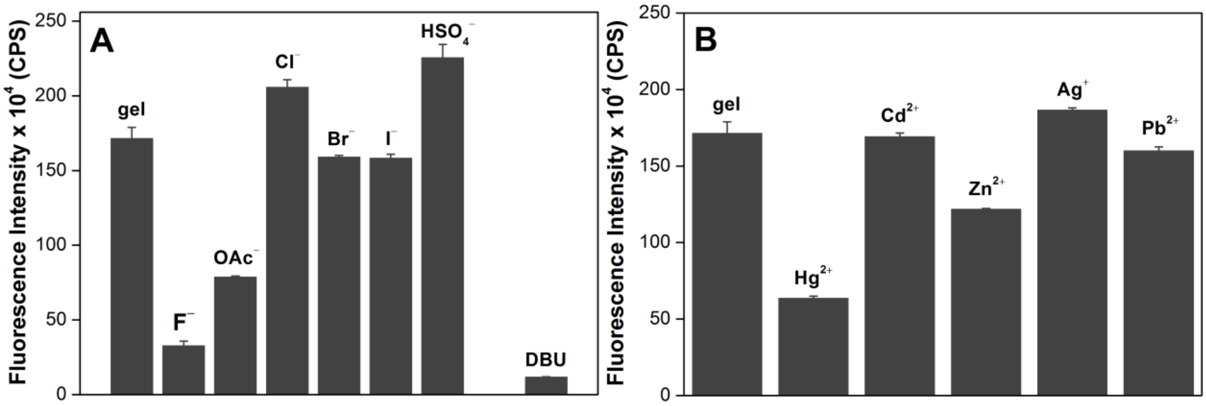


Figure 28. Plots showing the effect of addition of anions, a base and metal ions on the fluorescence intensity (λ_{em} = 445 nm) of benzothiophene-modified nucleolipid gel **10**. Fluorescence spectra were recorded by exciting the samples at 320 nm with excitation and emission slit widths of 2 nm. (**A**) Fluorescence intensity of **10** in the presence of various anions. Addition of F^- and AcO^- completely arrested the gel formation and displayed significant reduction in fluorescence intensity in sol state. Other anions tested did not affect the gelation of **10**. Addition of a strong organic base, 1,8-diazabicyclo(5.4.0) undec-7-ene (DBU), arrested the gelation process and resulted in significant reduction in fluorescence intensity in sol state. (**B**) Fluorescence intensity of **10** in the presence of various metal ions. Among the metal ions tested addition of Hg^{2+} alone lead to the precipitation of **10** and reduction (~3-fold) in fluorescence intensity.

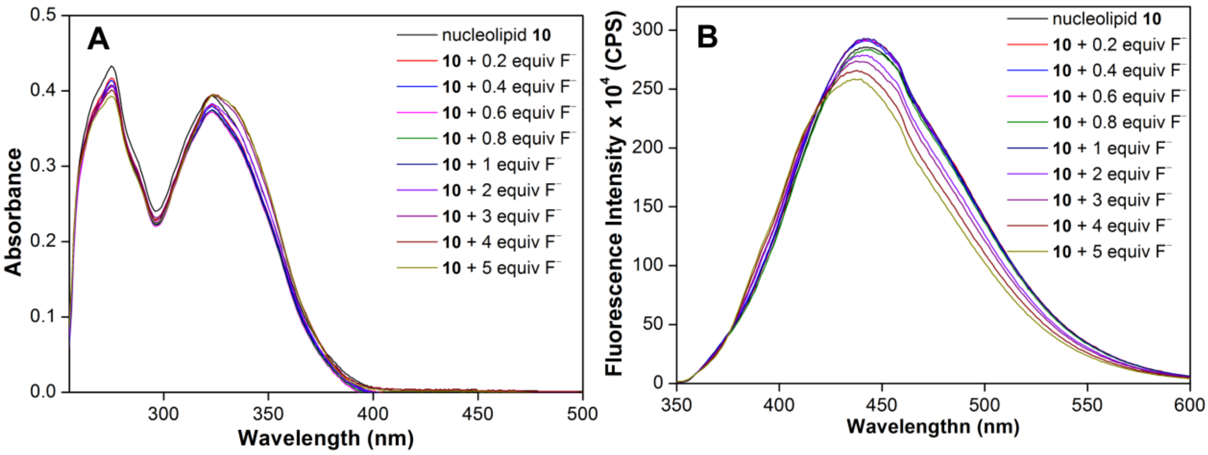


Figure 29. Effect of fluoride ion on UV absorption (**A**) and fluorescence (**B**) of nucleolipid **10** at well below its CGC. Concentration of **10** for absorption and fluoresence analysis was 25 µM and 10 µM, respectively. Samples were excited at 320 nm with an excitation slit width of 2 nm and emission slit width of 2 nm. Only small changes in absorption and fluorescence profile upon addition of fluoride ion were observed.

We measured UV absorption and fluorescence of nucleolipid **10** at well below its CGC as a function of the increasing concentration of fluoride ion. We observed only small changes in the absorption and fluorescence profile upon the addition of fluoride ion (Figure 29). These control experiments indicate that the decrease in the fluorescence intensity is most likely due to the disintegration of the gel to solution by fluoride ions as a result of the disruption of H-bonding interactions.

Similarly, reversible gel–sol transition induced by deprotonation and protonation steps by the addition of a base and an acid, respectively, was confirmed by the inverted vial method and fluorescence analysis (Figure 26). Addition of DBU (0.5 equiv.) arrested the gelation process, and subsequent addition of trichloroacetic acid (0.5 equiv.) restored the gelling ability of **10**. Attempts to determine the pK_a of modified nucleolipids failed as they were insoluble in aqueous buffer. So we determined the pK_a of N3–H of benzothiophene-modified uridine nucleoside in aqueous buffer by monitoring the changes in emission maximum as a function of pH (Figure 30). The p*K*a of modified nucleoside was found to be ~9.0, which is slightly lower than the pK_a of N3–H of uridine (~9.3).[35] Although the pH value was not reproducible in DMSO, DBU (0.5 equiv. with respect to nucleolipid) in DMSO gave a pH value more than 14. Hence, DBU should essential deprotonate the N3–H involved in the H-bonding network and disintegrate the gel. Upon providing the proton source we could restore the gelling ability of **10**.

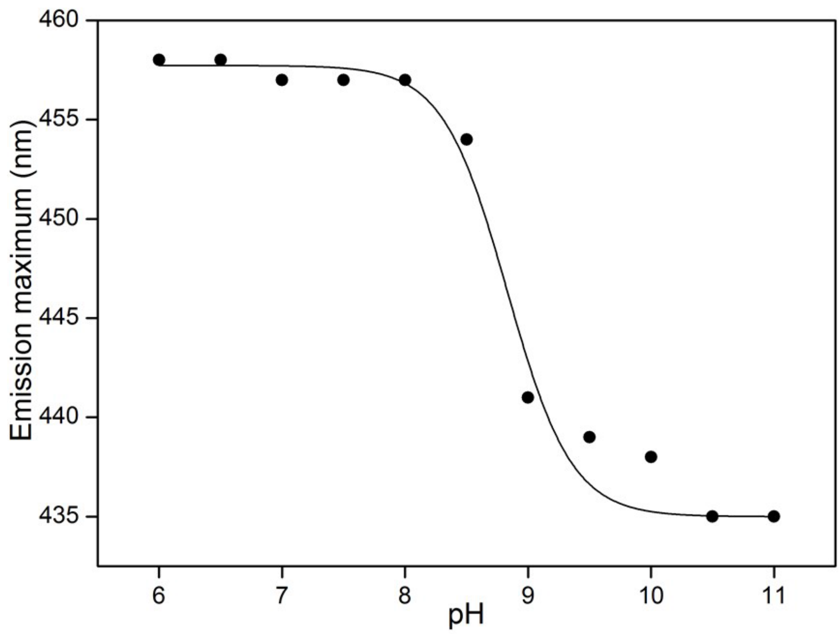


Figure 30. A plot of emission maximum of benzothiophene-modified uridine nucleoside versus pH. The pK_a of N3-H of benzothiophene-modified uridine nucleoside (10 μM) in aqueous buffer was determinmed by fitting the changes in emission maximum as a function of pH. The curve fit gave a pK_a of 9.0 ± 0.1. Under these conditions there were no significant changes in fluorescence intensity.

Metal-mediated base pairing in oligonucleotides have been elegantly utilized in developing metal ion sensors and nucleic acid-based molecular machines.[36] Notably, the thymine-Hg^{2+}-thymine base pair in which Hg^{2+} binds directly to N3 atoms of thymine residues has received much of the attention owing to its cytotoxic and mutagenic effects.[37] Encouraged by these reports, we wanted to evaluate if the metal ion binding to nucleolipids would affect the H-bonding interactions and hence, the self-assembly process. Increasing concentration of a small series of metal ions (Hg^{2+}, Cd^{2+}, Zn^{2+}, Ag^{+} and Pb^{2+}) that are known to interact with nucleosides was added to a hot solution of nucleolipid **10**. As low as 0.2 equiv. of Hg^{2+} lead to the precipitation of the nucleolipid with a nearly 3-fold reduction in the fluorescence intensity (Figure 26 and 28B). Interestingly, in the ^{1}H NMR spectrum of **10** containing the Hg^{2+} ion, the proton signal from N3H was intact, whereas the signal from 5′-OH was absent (Figure 31).

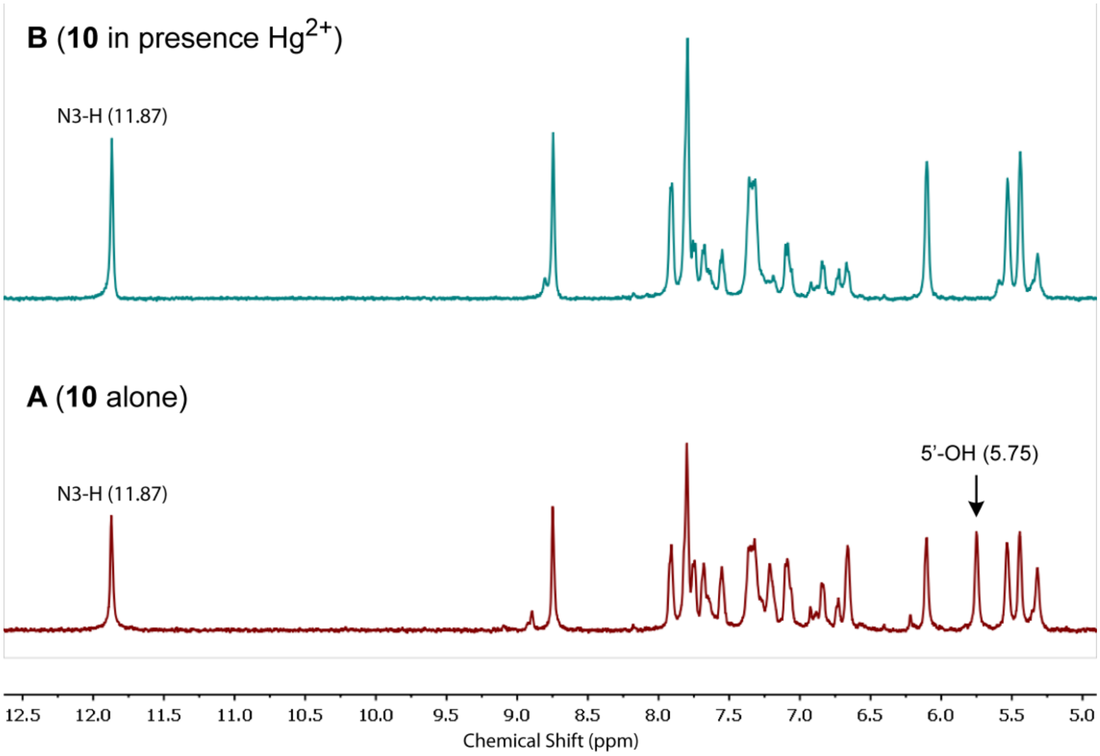


Figure 31. Partial 1H NMR spectrum of nucleolipid gel **10** in *d6*-DMSO (1.4 w/v %) in the absence (**A**) and presence of 1 equivalence of Hg^{2+}ion (**B**). Addition of Hg^{2+} ion did not affect N3-H signal (11.87 ppm), but resulted in the disappearance of 5′-OH signal (5.75 ppm). The deprotonation of 5′-OH due to binding of mercury to nucleolipid is the likely reason for the disruption of H-bonding interaction, and hence, the gelating ability of the nucleolipid.[39]

It is likely that the binding of mercury to nucleolipid would have resulted in the deprotonation of 5′-OH, thereby disrupting the H-bonding interaction and gelating ability of the nucleolipid.[38] However, other metal ions, even at a high concentration (5 equiv.), did not affect the gelation of **10**. Collectively, these results underscore the potential of this switchable nucleolipid gel as a useful two channel probe for fluoride and Hg^{2+} ions through visual phase transition and fluorescence change.[39,40] It is worth mentioning here that addition of adenosine, guanosine, cytidine and uridine nucleosides did not affect the gelation ability of fluorescent nucleolipids. In addition to these useful features, reducing the CGC and enhancing the mechanical strength of these fluorescent nucleolipid gels would augment their practical utility. In this regard, mono-, di-, and tri-substituted derivatives and also introduction of amide linkage between the nucleoside and fatty acids could potentially reduce CGC, improve mechanical strength, and show different supramolecular structures and fluorescence sensitivity.

5.6 Conclusions

In summary, we have introduced a new family of supramolecular nucleolipid synthons by combining the properties of (i) environmentally-sensitive fluorescent nucleoside analogs, which invoke multiple non-covalent interactions and (ii) fatty acids, which support aggregation by hydrophobic effect. These fluorescent nucleolipids readily formed organogels via a hierarchical self-organization process, which was found to be driven by a systematic interplay of a variety of non-covalent interactions including H-bonding, π–π interactions and hydrophobic effects. Importantly, the nucleolipid organogels not only exhibited enhanced emission upon aggregation, their gelation behaviour could be reversibly switched by the application of multiple external stimuli such as temperature, ultrasound and chemicals. Together, straight-forward synthesis, useful fluorescence properties and multistimuli responsiveness highlight the potential of these novel nucleolipid gels in developing new smart optical materials and probes. Moreover, this approach of building self-assembling nucleolipids could also be extended to other structurally and photophysically diverse nucleoside analogs, which will expand the functional repertoire of nucleoside–lipid hybrid materials.

5.7 Experimental section

5.7.1 Materials

5-iodouridine, benzofuran, benzo[*b*]thiophene, butyllithium, tributyltin chloride, bis(triphenylphosphine)palladium(II)dichloride, 4,4′-dimethoxytrityl chloride (DMT-Cl), octanoic acid, oleic acid, palmatic acid, tetrabutylammonium salts, silver nitrate, zinc nitrate hexahydrate, lead(II) nitrate, cadmium acetate dihydrate, mercury(II) perchlorate hexahydrate, DBU (1,8-diazabicyclo(5.4.0) undec-7-ene), dry pyridine were purchased from Sigma-Aldrich. EDC (1-(3-dimethyl aminopropyl)-3-ethyl carbodiimide hydrochloride) and 4-dimethylaminopyridine were obtained from Avra Synthesis. Myristic acid was procured from Fluka. Silicon wafers (N-type without dopant) were purchased from Sigma-Aldrich. 5-Iodo-5′-*O*-DMT-protected uridine **2**,[41] 2-(tri-n-butylstannyl)benzofuran[42] and 2-(tri-nbutylstannyl) benzothiophene[42] were prepared by following the reported procedures.

5.7.2 Instrumentation

NMR spectra were recorded on 400 MHz Jeol ECS-400 spectrometer and Bruker 500 MHz spectrometer. Absorption spectra were recorded on a Shimadzu UV-2600 spectrophotometer. Steady-state and time-resolved fluorescence experiments were carried out in a micro fluorescence cuvette (Hellma, path length 1.0 cm) either on a Fluoromax-4 spectrophotometer (Horiba Jobin Yvon) or TCSPC instrument (Horiba Jobin Yvon, Fluorolog-3). The morphology of gels was analyzed using Zeiss Ultra Plus field-emission scanning electron microscope (FESEM). Powder X-ray diffraction (PXRD) spectra were obtained at room temperature using Bruker D8 Advance diffractometer (Cu Kα radiation, λ = 1.5406 Å). Single crystal X-ray data for structure determination were collected from Bruker APEX II DUO diffractometer using MoKα (λ= 0.71073 Å) graphite monochromated radiation. Gel-sol transition temperature was determined using Equitron round bath clear bottom series medica instrument. MCR-301 (Anton-Paar) rheometer was used for Rheological studies.

5.7.3 Characterization data for compounds 3a–3d

5.7.3.1 General procedure for the synthesis of 5-Iodo-5′-*O*-DMT-protected nucleolipids (3a–3d): 5-Iodo-5′-*O*-DMT-protected uridine **2** (1.0 equiv), fatty acid (octanoic acid/myristic acid/palmitic acid/oleic acid, 2.4 equiv), EDC (2.4 equiv) and DMAP (2.4 equiv) were dissolved in dry dichloromethane (15 mL/g of **2**). The reaction mixture was stirred for 12 h at room temperature (RT) under nitrogen atmosphere. The reaction mixture was diluted with dichloromethane and was washed with saturated solution of NH_4Cl. The organic extract was dried over sodium sulphate and the solvent was evaporated in a rotary evaporator. The residue was purified by silica gel column chromatography (30% ethyl acetate in hexane containing 1% triethylamine) to afford the desired products (**3a–3d**).

Compound 3a: off white solid, 81% yield. TLC (petroleum ether:EtOAc = 60:40 v/v containing few drops of Et_3N); R_f = 0.48; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.11 (s, 1H), 7.46–7.44 (m, 2H), 7.37–7.30 (m, 6H), 7.26–7.22 (m, 1H), 6.87–6.85 (m, 4H), 6.24 (d, J = 7.2 Hz, 1H), 5.67 (dd, J_1 = 7.2 Hz, J_2 = 5.6 Hz, 1H), 5.59 (dd, J_1 = 5.6 Hz, J_2 = 2.0 Hz, 1H), 4.21–4.20 (m, 1H), 3.80 (s, 6H), 3.46 (dd, J_1 = 10.8 Hz, J_2 = 2.0 Hz, 1H), 3.41 (dd, J_1 = 11.0 Hz, J_2 = 2.2 Hz, 1H), 2.39–2.32 (m, 4H), 1.66–1.59 (m, 4H), 1.33–1.24 (m, 16H), 0.90–0.86 (m. 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.8, 172.6, 159.6, 158.9, 150.2, 144.0, 143.9, 135.3, 135.0, 130.3, 130.3, 128.4, 128.3, 127.4, 113.6, 87.8, 85.4, 82.9, 72.8, 71.6, 69.8, 63.1, 55.4, 34.2, 33.9, 31.8, 29.2, 29.2, 29.1, 29.0, 25.0, 24.8, 22.7, 14.2; HRMS: (m/z) calculated for $C_{46}H_{57}IN_2O_{10}$ $[M+Na]^+$ = 947.2956, found = 947.2946.

Compound 3b: Viscous oil, 77% yield. TLC (petroleum ether:EtOAc = 60:40 v/v containing few drops of Et_3N); R_f = 0.53; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.38 (br, 1H), 8.11 (s, 1H), 7.45–7.43 (m, 2H), 7.37–7.30 (m, 6H), 7.25–7.22 (m, 1H), 6.87–6.84 (m, 4H), 6.23 (d, J = 7.2 Hz , 1H), 5.67 (dd, J_1 = 7.2 Hz, J_2 = 5.6 Hz, 1H), 5.59 (dd, J_1 = 5.6 Hz, J_2 = 2.0 Hz, 1H), 4.21–4.20 (m, 1H), 3.79 (s, 6H), 3.45 (dd, J_1 = 11.0 Hz, J_2 = 2.2 Hz, 1H), 3.40 (dd, J_1 = 11.0 Hz, J_2 = 2.2 Hz, 1H), 2.38–2.31 (m, 4H), 1.63 (br, 4H), 1.28–1.25 (m, 40H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.8, 172.7, 158.9, 158.9, 144.0, 136.2, 135.0, 130.3, 130.3, 128.4, 128.3, 127.4, 127.4, 124.9, 113.6, 100.4, 87.8, 85.5, 82.9, 72.8, 71.6, 63.1, 55.4, 34.2, 33.9, 32.1, 29.8, 29.6, 29.5, 29.4, 29.3, 29.3, 25.0, 24.8, 22.8, 14.3; HRMS: (m/z) calculated for $C_{58}H_{81}IN_2O_{10}$ $[M+Na]^+$ = 1115.4834, found = 1115.4792.

Compound 3c: Viscous oil, 79% yield. TLC (petroleum ether:EtOAc = 60:40 v/v containing few drops of Et_3N); R_f = 0.54; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.47 (s, 1H), 8.27 (s, 1H), 7.37–7.27 (m, 5H), 7.19–7.17 (m, 2H), 7.16–7.15 (m, 2H), 6.85–6.81 (m, 4H), 6.05 (d, J = 6.0 Hz, 1H), 5.49–5.42 (m, 2H), 4.23–4.20 (m, 1H), 3.99 (dd, J_1 = 12.0 Hz, J_2 = 2.0 Hz, 1H), 3.88 (dd, J_1 = 11.4 Hz, J_2 = 1.0 Hz, 1H), 3.80 (s, 6H), 2.38–2.30 (m, 4H), 1.64–1.59 (m, 4H), 1.28–1.25 (m, 48 H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.0, 172.6, 159.6, 158.7, 150.0, 147.4, 145.3, 144.0, 139.6, 129.3, 128.0, 127.9, 127.2, 113.3, 87.8, 83.9, 73.4, 71.2, 69.2, 62.0, 55.4, 34.1, 33.9, 32.1, 29.9, 29.8, 29.6, 29.5, 29.4, 29.3, 29.2, 25.0, 24.8, 22.8, 14.3; HRMS: (m/z) calculated for $C_{62}H_{89}IN_2O_{10}$ $[M+Na]^+$ = 1171.5460, found = 1171.5465.

Compound 3d: Viscous oil, 88% yield. TLC (petroleum ether:EtOAc = 60:40 v/v containing few drops of Et_3N); R_f = 0.57; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.11 (s, 1H), 7.46–7.44 (m, 2H), 7.37–7.30 (m, 6H), 7.25–7.22 (m, 1H), 6.87–6.85 (m, 4H), 6.23 (d, J = 7.2 Hz, 1H), 5.69–5.65 (m, 1H), 5.59 (dd, J_1 = 5.4 Hz, J_2 = 2.0 Hz 1H), 5.38–5.31 (m, 4H), 4.21–4.20 (m, 1H), 3.79 (s, 6H), 3.45 (dd, J_1 = 11.0 Hz, J_2 = 2.2 Hz, 1H), 3.41 (dd, J_1 = 10.8 Hz, J_2 = 2.0 Hz, 1H), 2.39–2.30 (m, 4H), 2.07–1.96 (m, 8H), 1.66–1.60 (m, 4H), 1.30–1.25 (m, 40H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.7, 172.6, 159.7, 158.9, 150.2, 144.0, 135.3, 135.0, 130.3, 130.3, 130.2, 129.8, 129.8, 128.4, 128.3, 128.2, 127.4, 113.6, 87.8, 85.4, 82.9, 72.8, 71.6, 69.7, 63.1, 55.4, 34.1, 33.9, 32.0, 29.9, 29.9, 29.7, 29.5, 29.3, 29.3, 27.4, 27.3, 25.0, 24.8, 22.8, 14.3; HRMS: (m/z) calculated for $C_{66}H_{93}IN_2O_{10}$ $[M+Na]^+$ = 1223.5773, found = 1223.5764.

5.7.3.2 General procedure for the synthesis of fluorescent nucleolipids (4–11)

5-Iodo-5′-*O*-DMT-protected nucleolipid (**3a**/**3b**/**3c**/**3d**, 1.0 equiv), 2-(tri-n-butylstannyl)benzofuran or 2-(tri-n-butylstannyl)benzothiophene (2.0 equiv) and bis(triphenylphosphine)-palladium(II) chloride (0.05 equiv) were dissolved in anhydrous dioxane (50 mL/g of **3**). The reaction mixture was heated at 95 °C for 2.5 h and filtered through Celite pad. The Celite pad was washed with dioxane (2 x 15 mL) and the filtrate was evaporated. The resulting DMT-protected fluorescent nucleolipid was used in the next step without further purification. The deprotection was performed using 3% trichloroacetic acid in methylene chloride (20 mL) for 30 min. The reaction mixture was evaporated and the residue was purified by silica gel column chromatography (40% ethyl acetate in hexane) to afford the product (**4–11**).

Compound 4: Off white solid, 77% yield over two steps. TLC (petroleum ether:EtOAc = 70:30); R_f = 0.54; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 9.13 (s, 1H), 8.40 (s, 1H), 7.56–7.53 (m, 1H), 7.46–7.45 (m, 1H), 7.40–7.38 (m, 1H), 7.27–7.17 (m, 2H), 6.09 (d, J = 6.4 Hz, 1H), 5.66 (t, J = 6.0 Hz, 1H), 5.57 (dd, J_1 = 6.0 Hz, J_2 = 3.2 Hz, 1H), 4.26 (dd, J_1 = 5.2 Hz, J_2 = 2.4 Hz, 1H), 4.05 (dd, J_1 = 12.0 Hz, J_2 = 2.4 Hz, 1H), 3.94 (dd, J_1 = 12.4 Hz, J_2 = 2.4 Hz, 1H), 2.41–2.31 (m, 4H), 1.69–1.55 (m, 4H), 1.36–1.20 (m, 16 Hz), 0.91–0.83 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.0, 172.7, 160.0, 153.8, 149.5, 147.5, 136.7, 129.2, 124.8, 123.2, 121.6, 110.7, 107.4, 106.4, 89.4, 84.1, 73.0, 71.2, 62.2, 34.2, 33.9, 31.8, 31.7, 29.2, 29.2, 29.1, 29.0, 25.0, 24.8, 22.7, 22.7, 14.2; HRMS: (m/z) calculated for $C_{33}H_{44}N_2O_9$ $[M+Na]^+$ = 635.2945, found = 635.2934.

Compound 5: Off white solid, 66% yield over two steps. TLC (petroleum ether:EtOAc = 70:30); R_f = 0.42; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.80 (br, 1H), 8.40 (s, 1H), 7.57–7.54 (m, 1H), 7.49–7.47 (m, 1H), 7.41–7.39 (m, 1H), 7.32–7.28 (m, 1H), 7.24–7.19 (m, 1H), 6.09 (d, *J* = 6.4 Hz, 1H), 5.65 (t, *J* = 6.0 Hz, 1H), 5.57 (dd, J_1 = 5.6 Hz, J_2 = 3.2 Hz, 1H), 4.26 (dd, J_1 = 5.2 Hz, J_2 = 2.0 Hz, 1H), 4.05 (dd, J_1 = 12.2 Hz, J_2 = 2.2 Hz, 1H), 3.94 (dd, J_1 = 12.4 Hz, J_2 = 2.2 Hz, 1H), 2.40–2.31 (m, 4H), 1.68–1.57 (m, 4H), 1.32–1.23 (m, 40), 0.90–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.9, 172.6, 159.7, 153.8, 149.4, 147.5, 136.6, 129.2, 124.8, 123.2, 121.6, 110.7, 107.4, 106.4, 89.4, 84.1, 73.0, 71.2, 62.2, 34.2, 33.9, 32.1, 29.8, 29.6, 29.5, 29.5, 29.4, 29.3, 29.2, 28.0, 27.0, 25.0, 24.9, 22.8, 14.3; HRMS: (m/z) calculated for $C_{45}H_{68}N_2O_8$ $[M+Na]^+$ = 803.4822, found = 803.4835.

Compound 6: Off white solid, 69% yield over two steps. TLC (petroleum ether:EtOAc = 70:30); R_f = 0.56; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.40 (br, 1H), 7.57–7.55 (m, 1H), 7.47 (br, 1H), 7.41–7.39 (m, 1H), 7.28–7.26 (m, 1H), 7.24–7.19 (m, 1H), 6.09 (d, *J* = 6.4 Hz, 1H), 5.65 (t, *J* = 6.0 Hz, 1H), 5.57 (dd, J_1 = 5.6 Hz, J_2 = 3.2 Hz 1H), 4.26 (dd, J_1 = 5.2 Hz, J_2 = 2.4 Hz, 1H), 4.05 (dd, J_1= 12.0 Hz, J_2 = 2.4 Hz, 1H), 3.94 (dd, J_1 = 12.0 Hz, J_2 = 2.4 Hz, 1H), 2.38–2.31 (m, 4H), 1.68–1.55 (m, 4H), 1.29–1.22 (m, 48H), 0.90–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.4, 173.4, 160.3, 153.7, 150.2, 147.4, 134.5, 129.1, 124.9, 123.3, 121.5, 110.6, 107.3, 106.4, 89.7, 80.9, 74.1, 72.1, 63.3, 34.1, 34.0, 32.1, 29.8, 29.8, 29.7, 29.6, 29.5, 29.5, 29.4, 29.4, 29.3, 29.2, 24.9, 22.8, 14.3; HRMS: (m/z) calculated for $C_{49}H_{76}N_2O_9$ $[M+Na]^+$ = 859.5449, found = 859.5482.

Compound 7: Viscous oil, 39% yield over two steps. TLC (petroleum ether:EtOAc = 70:30); R_f = 0.49; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.99 (s,1H), 8.41 (s, 1H), 7.57–7.56 (m, 1H), 7.46 (br, 1H), 7.41–7.39 (m, 1H), 7.28–7.26 (m, 1H), 7.24–7.18 (m, 1H), 6.09 (d, *J* = 6.0 Hz, 1H), 5.65 (t, *J* = 6.0 Hz, 1H), 5.57 (dd, J_1 = 5.6 Hz, J_2 = 3.2 Hz, 1H), 5.39–5.28 (m, 4H), 4.29–4.24 (m, 1H), 4.05 (dd, J_1 = 12.2 Hz, J_2 = 2.2 Hz, 1H), 3.94 (dd, J_1 = 12.0 Hz, J_2 = 2.0 Hz, 1H), 2.40–2.31 (m, 4H), 2.07–1.91 (m, 8H), 1.66–1.55 (m. 4H), 1.32–1.22 (m, 40H), 0.89–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 172.9, 172.6, 159.9, 153.8, 149.4, 147.5, 136.6, 130.2, 130.2 129.8, 129.2, 124.8, 123.2, 121.6, 110.7, 107.4, 106.4, 89.3, 84.1, 73.0, 71.2, 62.2, 34.1, 33.9, 32.0, 29.9, 29.8, 29.8, 29.7, 29.5, 29.4, 29.3, 29.2, 27.4, 27.3, 25.0, 24.8, 22.8, 14.3; HRMS: (m/z) calculated for $C_{53}H_{80}N_2O_9$ $[M+Na]^+$ = 911.5762, found = 911.5715.

Compound 8: Off white solid, 67% yield over two steps. TLC (EtOAc:hexane = 30:70); R_f = 0.29; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.81 (s, 1H), 8.36 (s, 1H), 7.89 (s, 1H), 7.76–7.71 (m, 2H), 7.34–7.27 (m, 2H), 6.20 (d, J = 5.2 Hz, 1H), 5.59–5.52 (m, 2H), 4.28–4.27 (m, 1H), 4.07 (dd, J_1 = 12.0 Hz, J_2 = 1.6 Hz, 1H), 3.96 (dd, J_1 = 11.6 Hz, J_2 = 1.2 Hz, 1H), 2.41–2.32 (m, 4H), 1.68–1.56 (m, 4H), 1.37–1.21 (m, 16H), 0.91–0.84 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.1, 172.7, 160.7, 149.4, 140.1, 138.7, 136.9, 133.8, 124.8, 124.8, 123.9, 122.8, 121.9, 110.7, 87.9, 84.0, 73.6, 71.4, 62.2, 34.2, 33.9, 31.8, 31.8, 29.2, 29.2, 29.1, 29.0, 25.0, 24.8, 22.7, 22.7, 14.2; HRMS: (m/z) calculated for $C_{33}H_{44}N_2O_8S$ $[M+Na]^+$ = 651.2715, found = 651.2712.

Compound 9: Off white solid, 70% yield over two steps. TLC (EtOAc:hexane = 30:70); R_f = 0.41; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.68–8.62 (br, 1H), 8.37 (s, 1H), 7.89 (s, 1H), 7.77–7.72 (m, 2H), 7.34–7.27 (m, 2H), 6.2 (d, J = 5.6 Hz, 1H), 5.57–5.53 (m, 2H), 4.27–4.26 (m, 1H), 4.08–4.05 (m, 1H), 3.97–3.94 (m, 1H), 2.40–2.32 (m, 4H) 1.67–1.64 (m, 4H), 1.31–1.24 (m, 40H), 0.89–0.85 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.1, 172.7, 160.6, 149.4, 140.1, 138.7, 136.9, 133.9, 124.8, 124.6, 123.9, 122.8, 121.9, 110.7, 87.8, 84.0, 73.6, 71.4, 62.2, 34.2, 34.0, 32.1, 29.8, 29.8, 29.6, 29.5, 29.4, 29.3, 29.2, 25.0, 24.9, 22.8, 14.3; HRMS: (m/z) calculated for $C_{45}H_{68}N_2O_8S$ $[M+Na]^+$ = 819.4594, found = 819.4597.

Compound 10: Off white solid, 77% yield over two steps. TLC (EtOAc:hexane = 40:60); R_f = 0.72; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.37 (s, 1H), 7.90–7.89 (s, 1H), 7.76–7.72 (m, 2H), 7.34–7.29 (m, 2H), 6.20 (d, J = 4.8 Hz, 1H), 5.58–5.55 (m, 2H), 4.27 (br, 1H), 4.08–4.05 (m, 1H), 3.97–3.94 (m, 1H), 2.40–2.32 (m, 4H), 1.68–1.56 (m, 4H), 1.36–1.23 (m, 48H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.1, 172.7, 160.6, 149.4, 140.1, 138.7, 136.9, 133.8, 124.8, 124.6, 123.9, 122.8, 121.9, 121.1, 110.7, 87.9, 84.0, 73.6, 71.4, 62.2, 34.2, 34.0, 32.1, 29.9, 29.8, 29.6, 29.5, 29.4, 29.4, 29.3, 29.2, 25.0, 24.9, 22.8, 14.3; HRMS: (m/z) calculated for $C_{49}H_{76}N_2O_8S$ $[M+Na]^+$ = 875.5220, found = 875.5190.

Compound 11: Viscous oil, 47% yield over two steps. TLC (EtOAc:hexane = 30:70); R_f = 0.40; ^{1}H NMR (400 MHz, $CDCl_3$): δ (ppm) 8.68 (s, 1H), 8.37 (s, 1H), 7.89 (s, 1H), 7.76–7.72

(m, 2H), 7.34–7.27 (m, 2H), 6.20 (d, J = 5.2 Hz, 1H), 5.58–5.53 (m, 2H), 5.38–5.28 (m, 4H), 4.27–4.26 (m, 1H), 4.07 (dd, J_1 = 12.0 Hz, J_2 = 2.0 Hz, 1H), 3.96 (dd, J_1 = 11.8 Hz, J_2 = 1.8 Hz, 1H), 2.40–2.32 (m, 4H), 2.07–1.91 (m, 8H), 1.68–1.56 (m, 4H), 1.32–1.23 (m, 40H), 0.89–0.86 (m, 6H); ^{13}C NMR (100 MHz, $CDCl_3$): δ (ppm) 173.0, 172.7, 160.6, 149.4, 140.1, 138.7, 136.9, 133.9, 130.2, 130.2, 129.8, 129.8, 124.8, 124.6, 123.9, 122.8, 121.9, 110.7, 87.8, 84.0, 73.6, 71.4, 62.2, 34.1, 33.9, 32.0, 29.9, 29.8, 29.7, 29.5, 29.4, 29.3, 29.2, 27.4, 27.3, 25.0, 24.8, 22.8, 14.3; HRMS: (m/z) calculated for $C_{53}H_{80}N_2O_8S$ $[M+Na]^+$ = 927.5533, found = 927.5550.

5.7.4 Fluorescence properties of nucleolipids in different solvents:

Quantum yield of nucleolipids in various solvents: Quantum yield of 5-benzofuran-modified (**5** and **6**) and 5-benzothiophene-modified (**9** and **10**) nucleolipids in different solvents relative to 2-aminopurine as the standard was determined using the following equation.[43]

$$\Phi_{F(x)} = \left(A_s/A_x\right)\left(F_x/F_s\right)\left(n_x/n_s\right)^2 \Phi_{F(s)}$$

Where s is the standard, x is the nucleolipid, A is the absorbance at excitation wavelength, F is the area under the emission curve, n is the refractive index of the solvent, and Φ_F is the quantum yield. Quantum yield of 2-aminopurine in water is 0.68.[44]

Steady-state fluorescence anisotropy measurements: Steady-state fluorescence anisotropy measurements in different solvents were performed by exciting the samples (5 μM) at 320 nm. The anisotropy value (r) was determined by analyzing the data using software provided with the instrument. Anisotropy measurements were performed in duplicate and the values reported in this study are an average of ten successive measurements for each sample.

5.7.5 Gelation test and gel–sol transition temperature (T_{gel}):

A weighed amount of the nucleolipid gelator was mixed with an organic solvent (e.g., DMSO, 0.5 mL) in a septum-capped vial and heated until a clear solution was obtained. The sample vial was cooled and left to stand at RT for at least 2 h. The formation of the gel was tested by inverted vial method. If no flow was observed upon inverting the vial, then it was noted as gel. If there was flow and the gelator solution was clear then it was noted as sol. The minimum amount of the gelator required for the formation of a stable gel at RT was determined by performing this test at different concentrations of the gelator. The thermo-reversibility of gel-sol transition was confirmed by repeated heating–cooling steps. Sonication-induced gelation process was tested by subjecting a hot solution of the gelator at respective CGC to ultrasound (Bransonic 2510 EDTH, 40 kHz) for ~1 min at the beginning of the cooling process. Photographs of organogels of nucleolipids **5**, **6**, **9** and **10** in DMSO were taken under white light and under UV illumination (365 nm) to observe the fluorescence from the gel state. All experiments were performed at least in duplicate.

A sealed vial containing the nucleolipid gel at CGC was immersed upside-down in a thermostated water bath. The temperature of the bath was raised at 2 °C/min. The temperature at which the gel started to flow was noted as T_{gel}. The experiments were performed in duplicate and error in T_{gel} was found to be ± 1 °C.

5.7.6 FESEM experiments: The morphology of nucleolipid organogels was characterized by FESEM. Diluted gel samples in DMSO were drop-casted on silicon wafers and dried under vacuum in a desiccator for nearly 15 h. To minimize sample charging, a thin layer of Au was deposited onto the samples before FESEM analysis. ImageJ 1.46r software was used to calculate the dimensions of the nucleolipid nanotubes, ribbons and fibres.

5.7.7 Crystallography: Single crystals of 5-benzofuran-modified nucleolipids **5** and **6** were obtained from their corresponding chloroform solution containing few drops of dimethylformamide by vapour diffusion method. Crystals of 5-benzothiophene-modified nucleolipids **9** and **10** were obtained from benzene by slow evaporation method. The diffraction data of **5**·DMF, **6**·DMF, **9**·C_6H_6 and **10**·C_6H_6 were collected using Bruker KAPPA APEX II CCD Duo diffractometer (operated at 1500 W, 50 kV and 30 mA) equipped with a graphite-monochromated MoKα (λ = 0.71073 Å) radiation. All the structures were refined on F^2 by full-matrix least-squares technique using SHELX program.[45] All non-hydrogen atoms were refined aniostropically unless otherwise stated. Hydrogen atoms were constrained in geometric positions with respect to their parent atoms. Crystallographic data for the nucleolipids see appendix-V. Crystals of **6** and **10** were weakly diffracting at higher angles, and hence, a 2θ = 50° cut-off was applied. The solvated benzene carbon atoms in **10** were disordered and thus were freely refined isotropically over two positions using similar distances and similar U-restraints. A few C, N, and O atoms in the nucleolipid structures exhibited slightly bad ellipsoids and were refined with partial isotropic parameters. The X-ray crystallographic coordinates for structures reported in this article have been deposited at the

Cambridge Crystallographic Data Centre (CCDC), under deposition number CCDC: 1423164 (**5**), 1423177 (**6**), 1423141 (**9**) and 1423142 (**10**). These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif.

5.7.8 Variable temperature ^{1}H NMR: Gels of **6** (0.9 w/v %) and **10** (1.4 w/v %) in *d6*-DMSO were formed in individual NMR tubes by heating and cooling process. ^{1}H NMR was recorded on a 500 MHz spectrometer as a function of increasing temperature. The temperature of the sample was increased from 25 °C to 55 °C with an increment of 5 °C and equilibration time of 10 min. The spectrum was recorded at every 5 °C interval.

5.7.9 Powder X-ray diffraction (PXRD) analysis of nucleolipids 5, 6, 9 and 10: A hot solution of nucleolipid in DMSO at respective CGC was drop-casted on glass slide and was allowed to from gel at RT. The glass slide was placed in a vacuum desiccator and was dried under vacuum for nearly 15 h to obtain the xerogel. PXRD data was collected using Bruker D8 Advance diffractometer with CuKα source (1.5406 Å). Diffraction data were collected at 2θ angle from 1° to 40° using a 0.01° step size and 0.5 s per step. Low angle diffraction data was collected by keeping the motorized divergence slit in automatic mode so as to maintain the X-ray beam footprint on the sample to12 x 12 mm. Further, the position sensitive detector (Lynxeye) channels were reduced to minimize the background X-ray scattering entering the detector.

5.7.10 Fluorescence of nucleolipid gels 5, 6, 9 and 10. A hot solution of the nucleolipid at respective CGC was taken in a micro fluorescence cuvette (Hellma, path length 1.0 cm) and was allowed to stand at RT till stable gel was formed (1–2 h, cuvette was inverted to test the gel formation). Fluorescence spectrum of gel samples was recorded as a function of temperature by exciting the samples at 320 nm with an excitation slit width of 2 nm and emission slit width of 2 nm. Nucleolipid gels **5** and **9** containing myristoyl chains were subjected to heating–cooling process as the following: 25 °C to 40 °C with an increment of 3 °C in temperature and an equilibration time of 10 min. Fluorescence spectrum was recorded at every 3 °C interval. Same steps were followed during the cooling cycle. Similarly, fluorescence spectrum of nucleolipid gels **6** and **10** containing palmitoyl chains was recorded at every 5 °C interval (heating–cooling cycle: 25 °C to 50 °C and then to 25 °C, 5 °C interval and 10 min equilibration time). The experiments were performed in duplicate.

5.7.11 Responsiveness of nucleolipid gel 10 to chemical stimuli.

Effect of anions. To a hot solution of nucleolipid **10** (1.4 w/v %) in DMSO was added increasing concentration of tetrabutylammonium salts of F^-, AcO^-, Cl^-, Br^-, I^- and HSO_4^- without altering the CGC of the gelator. The samples were allowed to stand at RT temperature for 1 h and the gelling ability was tested by inverted vial method. In the presence of F^- and AcO^- (0.5 equiv), nucleolipid **10** did not show any signs of gelation even after a month. The gelation was restored by addition of acetic acid (0.5 equiv). However, Cl^-, Br^-, I^- and HSO_4^- (5 equiv) did not affect the gelation process and formed stable gels within 20 min, which did not disintegrate for months at RT. The fluorescence of gel samples in the presence and absence of anions was recorded by exciting the samples at 320 nm with an excitation slit width of 2 nm and an emission slit width of 2 nm.

Effect of base. To a hot solution of nucleolipid **10** (1.4 w/v %) in DMSO was added increasing concentration a strong organic base, 1,8-diazabicyclo(5.4.0) undec-7-ene (DBU). The sample was allowed to stand at RT temperature for 1 h. The gelation test and fluorescence was recorded as above. DBU (0.5 equiv) arrested the gelation process, and subsequent addition of trichloroacetic acid (0.5 equiv) restored the gelling ability of **10**.

Effect of metal ions. To a hot solution of nucleolipid **10** (1.4 w/v %) in DMSO was added increasing concentration of $Hg(ClO_4)\cdot6H_2O$, $Hg(NO_3)_2\cdot H_2O$, $Cd(OAc)_2\cdot2H_2O$, $Zn(NO_3)_2\cdot6H_2O$, $AgNO_3$ and $Pb(NO_3)_2$ without altering the CGC of gelator. The samples were allowed to stand at RT temperature for 1 h and the fluorescence was recorded as above. While Hg^{2+} lead to the precipitation, other metal ions tested did not affect the gelling ability of **10**.

5.8 References:

1 (a) Krishnan, Y.; Simmel, F. C. *Angew. Chem., Int. Ed.* **2011**, 50, 3124–3156. (b) Jones, M. R.; Seeman, N. C.; Mirkin, C. A. *Science* **2015**, *347*, 1260901.

2 Guo, P.; Haque, F. *RNA Nanotechnology and Therapeutics*, CRS Press, Taylor and Francis Group, **2014**.

3 (a) Tian, Y.; Mao, C. *J. Am. Chem. Soc.* **2004**, *126*, 11410–11411. (b) Ackermann, D.; Schmidt, T. L.; Hannam, J. S.; Purohit, C. S.; Heckel, A.; Famulok, M. *Nat. Nanotechnol.* **2010**, *5*, 436–442. (c) Saha, S.; Prakash, V.; Halder, S.; Chakraborty, K.; Krishnan, Y. *Nat. Nanotechnol.* **2015**, *10*, 645–651.

4 (a) Erben, C. M.; Goodman, R. P.; Turberfield, A. J. *Angew. Chem., Int. Ed.* **2006**, *45*, 7414–7417. (b) Bhatia, D.; Chakraborty, S.; Mehtab, S.; Krishnan, Y. *Methods Mol. Biol.* **2013**, *991*, 65–80.

5 (a) Han, D.; Park, Y.; Kim, H.; Lee, J. B. *Nat. Commun.* **2014**, *5*, 4367. (b) Langecker, M.; Arnaut, V.; Martin, T. G.; List, J.; Renner, S.; Mayer, M.; Dietz, H. Simmel, F. C. *Science* **2012**, *338*, 932–936.

6 (a) Douglas, S. M.; Bachelet, I.; Church, G. M. *Science* **2012**, *335*, 831–834. (b) Meyer, R.; Faesen, A.; Vogel, K.; Jeganathan, S.; Musacchio, A.; Niemeyer, C. M.; *Small* **2015**, *11*, 2669–2674.

7 (a) Cobo, I.; Li, M.; Sumerlin, B. S.; Perrier, S. *Nat. Mater.* **2015**, *14*, 143–159. (b) Beales, P. A.; Vanderlick, T. K. *Adv. Colloid Interface Sci.* **2014**, *207*, 290–305. (c) Patwa, A.; Gissot, A.; Bestel, I.; Barthélémy, P. *Chem. Soc. Rev.* **2011**, *40*, 5844–5854. (d) Ruff, Y.; Moyer, T.; Newcomb, C. J.; Demeler, B.; Stupp, S. I. *J. Am. Chem. Soc.* **2013**, *135*, 6211–6219. (e) Banga, R. J.; Chernyak, N.; Narayan, S. P.; Nguyen, S. T.; Mirkin, C. A. *J. Am. Chem. Soc.* **2014**, *136*, 9866–9869.

8 (a) Sivakova, S.; Rowan, S. J. *Chem. Soc. Rev.* **2005**, *34*, 9–21. (b) Llanes-Pallas, A.; Palma, C.-A.; Piot, L.; Belbakra, A.; Listorti, A.; Prato, M.; Samorì, P.; Armaroli, N.; Bonifazi, D. *J. Am. Chem. Soc.* **2009**, *131*, 509–520. (c) Zhao, H.; Guo, X.; He, S.; Zeng, X.; Zhou, X.; Zhang, C.; Hu, J.; Wu, X.; Xing, Z.; Chu, L.; He, Y.; Chen, Q. *Nat. Commun.* **2014**, *5*, 3108. (d) Li, X.; Kuang, Y.; Lin, H.-C.; Gao, Y.; Shi, J.; Xu, B. *Angew. Chem., Int. Ed.* **2011**, *50*, 9365–9369. (e) Wu, D.; Zhou, J.; Shi, J.; Du, X.; Xu, B. *Chem. Commun.* **2014**, *50*, 1992–1994. (f) Berger, O.; Adler-Abramovich, L.; Levy-Sakin, M.; Grunwald, A.; Liebes-Peer, Y.; Bachar, M.; Buzhansky, L.; Mossou, E.; Forsyth, V. T.; Schwartz, T.; Ebenstein, Y.; Frolow, F.; Shimon, L. J. W.; Patolsky, F.; Gazit, E. *Nat. Nanotechnol.* **2015**, *10*, 353–360.

9 Lena, S.; Masiero, S.; Pieraccini, S.; Spada, G. P. *Chem. Eur. J.* **2009**, *15*, 7792–7806.

10 (a) Ma, L.; Melegari, M.; Colombini, M.; Davis, J. T. *J. Am. Chem. Soc.* **2008**, *130*, 2938–2939. (b) Simeone, L.; Milano, D.; Napoli, L. D.; Irace, C.; Pascale, A. D.; Boccalon, M.; Tecilla, P.; Montesarchio, D. *Chem. Eur. J.* **2011**, *17*, 13854–13865.

(c) Das, R. N.; Kumar, Y. P.; Schütte, O. M.; Steinem, C.; Dash, J. *J. Am. Chem. Soc*. **2015**, *137*, 34–37.

11 (a) Rosemeyer, H. *Chem. Biodiversity*, **2005**, *2*, 977–1062. (b) Sessler, J. L.; Jayawickramarajah, J. *Chem. Commun*. **2005**, 1939–1949. (c) Gissot, A.; Camplo, M.; Grinstaff, M. W.; Barthélémy, P. *Org. Biomol. Chem*. **2008**, *6*, 1324–1333. (d) Berti, D.; Montis, C.; Baglioni, P. *Soft Matter* **2011**, *7*, 7150–7158. (e) Allain, V.; Bourgaux, C.; Couvreur, P. *Nucleic Acids Res*. **2012**, *40*, 1891–1903.

12 (a) Khiati, S.; Luvino, D.; Oumzil, K.; Chauffert, B.; Michel Camplo, M.; Barthélémy, P. *ACS Nano* **2011**, *5*, 8649–8655. (b) Ceballos, C.; Khiati, S.; Prata, C. A. H.; Zhang, X.-X.; Giorgio, S.; Marsal, P.; Grinstaff, M. W.; Barthélémy, P.; Camplo, M. *Bioconjugate Chem*. **2010**, *21*, 1062–1069. (c) Wang, D.; Tu, C.; Su, Y.; Zhang, C.; Greiser, U.; Zhu, X.; Yana, D.; Wang, W. *Chem. Sci*. **2015**, *6*, 3775–3787.

13 Rosemeyer, H.; Stürenberg, E. M.; Herdewijn, P. *Nucleosides, Nucleotides Nucleic Acids* **2007**, *26*, 995–999.

14 Moreau, L.; Barthélémy, P.; El Maataoui, M.; Grinstaff, M. W. *J. Am. Chem. Soc*. **2004**, *126*, 7533–7539.

15 (a) Busseron, E.; Ruff, Y.; Moulin, E.; Giuseppone, N. *Nanoscale* **2013**, *5*, 7098–7140. (b) Babu, S. S.; Praveen, V. K.; Ajayaghosh, A. *Chem. Rev*. **2014**, *114*, 1973–2129. (c) Weiss, R. G. *J. Am. Chem. Soc*. **2014**, *136*, 7519–7530. (d) Datta, S.; Bhattacharya, S. *Chem. Soc. Rev*. **2015**, *44*, 5596–5637.

16 (a) Tanpure, A. A.; Srivatsan, S. G. *Chem. Eur. J*. **2011**, *17*, 12820–12827. (b) Pawar, M. G.; Srivatsan, S. G. *Org. Lett*. **2011**, *13*, 1114–1117. (c) Tanpure, A. A.; Srivatsan, S. G. *ChemBioChem* **2012**, *13*, 2392–2399.

17 Sinkeldam, R. W.; Wheat, A. J.; Boyaci, H.; Tor, Y. *ChemPhysChem* **2011**, *12*, 567–570.

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

19 Tanpure, A. A.; Pawar, M. G.; Srivatsan, S. G. *Isr. J. Chem*. **2013**, *53*, 366–378.

20 Yu, X.; Chen, L.; Zhanga, M.; Yi, T. *Chem. Soc. Rev*. **2014**, *43*, 5346–5371.

21 (a) Bhattacharya, S.; Acharya, S. N. G. *Chem. Mater*. **1999**, *11*, 3121–3132. (b) Gao, Y.; Hao, J.; Wu, J.; Zhang, X.; Hu, J.; Ju, Y. *Nanoscale*, **2015**, *7*, 13568–13575.

22 Jung, J. H.; Do, Y.; Lee, Y.-A.; Shimizu, T. *Chem. Eur. J*. **2005**, *11*, 5538–5544.

23 (a) Lin, S.-C.; Ho, R.-M.; Chang, C.-Y.; Hsu, C.-S. *Chem. Eur. J*. **2012**, *18*, 9091–9098. (b) Liu, C.; Jin, Q.; Lv, K.; Zhang, L.; Liu, M. *Chem. Commun*. **2014**, *50*,

3702–3705. (c) Malakar, P.; Prasad, E. *Chem. Eur. J*. **2015**, *21*, 5093–5100. (d) Song, B.; Liu, B.; Jin, Y.; He, X.; Tang, D.; Wu, G.; Yin, S. *Nanoscale* **2015**, 7, 930–935.

24 Dastidar, P. *Chem. Soc. Rev*. **2008**, *37*, 2699–2715.

25 Karlson, P.; Dixon, H. B. F.; Cornish-Bowden, A.; Liébecq, C.; Loening, K. L.; Moss, G. P.; Reedijk, J.; Velick, S. F.; Vliegenthart, J. F. G. Abbreviations and symbols for the description of conformations of polynucleotide chains, *Pure Appl. Chem*. **1983**, *55*, 1273–1280.

26 (a) Greco, N. J.; Tor, Y. *Tetrahedron* **2007**, *63*, 3515– 3527. (b) Busschaert, N.; Elmes, R. B. P.; Czech, D. D.; Wu, X.; Kirby, I. L.; Peck, E. M.; Hendzel, K. D.; Shaw, S. K.; Chan, B.; Smith, B. D.; Jolliffe, K. A.; Gale, P. A. *Chem. Sci*. **2014**, *5*, 3617–3626.

27 (a) Moralez, J. G.; Raez, J.; Yamazaki, T.; Motkuri, R. K.; Kovalenko, A.; Fenniri, H. *J. Am. Chem. Soc*. **2005**, *127*, 8307–8309. (b) Pantos, G. D.; Pengo, P.; Sanders, J. K. M.; *Angew. Chem., Int. Ed*. **2007**, *46*, 194–197. (c) Ziserman, L.; Lee, H.-Y.; Raghavan, S.R.; Mor, A.; Danino, D. *J. Am. Chem. Soc*. **2011**, *133*, 2511–2517. (d) Jin, Q.; Zhang, L.; Liu, M. *Chem. Eur. J*. **2013**, *19*, 9234–9241.

28 Birks, J. B. *Photophysics of Aromatic Molecules*, Wiley, London, **1970**.

29 (a) Pérez, A.; Serrano, J. L.; Sierra, T.; Ballesteros, A.; Saá, D.; Barluenga, J. *J. Am. Chem. Soc*. **2011**, *133*, 8110–8113. (b) Zhao, J.; Yang, D.; Zhao, Y.; Yang, X.-J.; Wang, Y.-Y.; Wu, B. *Angew. Chem., Int. Ed*. **2014**, *53*, 6632–6636. (c) Hu, R.; Leung, N. L. C.; Tang, B. Z. *Chem. Soc. Rev*. **2014**, *43*, 4494–4562. (d) Zhang, L.; He, N.; Lu, C. *Anal. Chem*. **2015**, *87*, 1351–1357.

30 (a) Yang, X.; Zhanga, G.; Zhang, D. *J. Mater. Chem*. **2012**, *22*, 38–50. (b) Segarra-Maset, M. D.; Nebot, V. J.; Miravet, J. F.; Escuder, B. *Chem. Soc. Rev*. **2013**, *42*, 7086–7098.

31 (a) Basak, S.; Nanda, J.; Banerjee, A. *J. Mater. Chem*. **2012**, *22*, 11658–11664. (b) Sun, Z.; Li, Z.; He, Y.; Shen, R.; Deng, L.; Yang, M.; Liang, Y.; Zhang, Y. *J. Am. Chem. Soc*. **2013**, *135*, 13379–13386. (c) Miao, W.; Yang, D.; Liu, M. *Chem. Eur. J*. **2015**, *21*, 7562–7570. (d) Vidyasagar, A.; Sureshan, K. M. *Angew. Chem., Int. Ed*. **2015**, *54*, 12078–12082. (e) Miao, Q.; Wu, Z.; Hai, Z.; Tao, C.; Yuan, Q.; Gong, Y.; Guan, Y.; Jiang, J.; Liang, G. *Nanoscale*, **2015**, *7*, 2797–2804.

32 (a) Hirst, A. R.; Escuder, B.; Miravet, J. F.; Smith, D. K. *Angew. Chem., Int. Ed*. **2008**, *47*, 8002–8018. (b) Piepenbrock, M.-O. M.; Lloyd, G. O.; Clarke, N.; Steed, J.

W. *Chem. Rev*. **2010**, *110*, 1960–2004. (c) Yan, N.; Xu, Z.; Diehn, K. K.; Raghavan, S. R.; Fang, Y.; Weiss, R. G. *J. Am. Chem. Soc*. **2013**, *135*, 8989–8999. (d) Chen, H.; Feng, Y.; Deng, G.-J.; Liu, Z.-X.; He, Y.-M.; Fan, Q.-H. *Chem. Eur. J*. **2015**, *21*, 11018–11028.

33 Moreau, L.; Ziarelli, F.; Grinstaff , M. W.; Barthélémy, P. *Chem. Commun*. **2006**, 1661–1663.

34 Kang, S. O.; Powell, D.; Day, V. W.; Bowman-James, K. *Angew. Chem., Int. Ed*. **2006**, *45*, 1921–1925.

35 Chatterjee, S.; Pathmasiri, W.; Plashkevych, O.; Honcharenko, D.; Varghese, O. P.; Maiti, M.; Chattopadhyaya, J. *Org. Biomol. Chem*. **2006**, *4*, 1675–1686.

36 (a) Clever, G. H.; Kaul, C.; Carell, T. *Angew. Chem., Int. Ed*. **2007**, *46*, 6226–6236. (b) Scharf, P.; Müller, J. *ChemPlusChem* **2013**, *78*, 20–34.

37 Miyake, Y.; Togashi, H.; Tashiro, M.; Yamaguchi, H.; Oda, S.; Kudo, M.; Tanaka, Y.; Kondo, Y.; Sawa, R.; Fujimoto, T.; Machinami, T.; Ono, A. *J. Am. Chem. Soc*. **2006**, *128*, 2172–2173.

38 (a) Nolan, E. M.; Lippard, S. J. *Chem. Rev*. **2008**, *108*, 3443–3480. (b) Kaur, N.; Dhaka, G.; Singh, J. *New J. Chem*. **2015**, *39*, 6125–6129.

39 Foster, J. A.; Piepenbrock, M.-O. M.; Lloyd, G. O.; Clarke, N.; Howard, J. A. K.; Steed, J. W. *Nat. Chem*. **2010**, *2*, 1037–1043.

40 Zhang, Y.-M.; Lin, Q.; Wei, T.-B.; Qin, X.-P.; Li, Y. *Chem. Commun*. **2009**, 6074–6076.

41 Shah, K.; Wu, H.; Rana, T. M. *Bioconjugate Chem*. **1994**, *5***,** 508–512.

42 Liebeskind, L. S.; Wang, J. *J. Org. Chem*. **1993**, *58***,** 3550–3556.

43 Lavabre, D.; Fery-Forgues, S. *J. Chem. Educ.* **1999**, *76*, 1260–1264.

44 Ward, D. C.; Reich, E.; Stryer, L. *J. Biol. Chem*. **1969**, *244*, 1228–1237.

45 Sheldrick, G. M. A short history of SHELX, *Acta Crystallogr., Sect. A: Fundam. Crystallogr*. **2008**, *64*, 112–122.

5.9 Appendix-V: Characterization data of synthesized compounds

5.9.1 Crystallographic Information

5.9.1.1 Crystallographic data for nucleolipid **5**

Compound identity	**5**
CCDC (deposition number)	1423164
Empirical formula	$C_{45} H_{68} N_2 O_9$
Formula weight	781.04
Temperature	100(2) K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	P 1
Unit cell dimensions	a = 8.615(3) α = 101.343(7)
	b = 11.124(4) β = 91.424(7)
	c = 25.315(9) γ = 109.601(6)
Volume	2229.9(14) $Å^3$
Z	2
Density (calculated)	1.218 Mg/cm^3
Absorption coefficient (μ)	0.084 mm^{-1}
F(000)	888
Crystal size	0.40 x 0.17 x 0.05 mm^3
Theta range for data collection	4.11 to 28.28
Index ranges	-11<=h<=11, -14<=k<=10, -31<=l<=33
Reflections collected	32572
Independent reflections	13635 [R(int) = 0.0985]
Completeness to theta = 28.28°	97.0 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.983 and 0.996
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	13635 / 9 / 1063
Goodness-of-fit on F^2	0.958
Final R indices [I>2sigma(I)]	R1 = 0.0652, wR2 = 0.1124
R indices (all data)	R1 = 0.1361, wR2 = 0.1372
Largest diff. Peak and hole	0.285 and -0.288 e.$Å^{-3}$
CCDC	1423164

5.9.1.2 Crystallographic data for nucleolipid **6**

Compound identity	**6**
CCDC (deposition number)	1423177
Empirical formula	$C_{49} H_{76} N_2 O_9$
Formula weight	836.55
Temperature	100(2) K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	P1
Unit cell dimensions	a = 8.671(7) α = 94.389(14)
	b = 11.183(9) β = 97.977(16)
	c = 27.36(2) γ = 109.553(14)
Volume	2455(4) $Å^3$
Z	2
Density (calculated)	1.182 Mg/cm^3
Absorption coefficient (μ)	0.081 mm^{-1}
F(000)	952
Crystal size	0.39 x 0.19 x 0.05 mm^3
Theta range for data collection	4.10 to 25.02
Index ranges	-10<=h<=9, -13<=k<=13, -32<=l<=32
Reflections collected	32275
Independent reflections	11984[R(int) = 0.3614]
Completeness to theta = 25.02°	99.4 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.982 and 0.996
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	11984 / 45 / 1134
Goodness-of-fit on F^2	0.643
Final R indices [I>2sigma(I)]	R1 = 0.0725, wR2 = 0.1467
R indices (all data)	R1 = 0.2949, wR2 = 0.2409
Largest diff. Peak and hole	0.239 and -0.262 e.$Å^{-3}$
CCDC	1423177

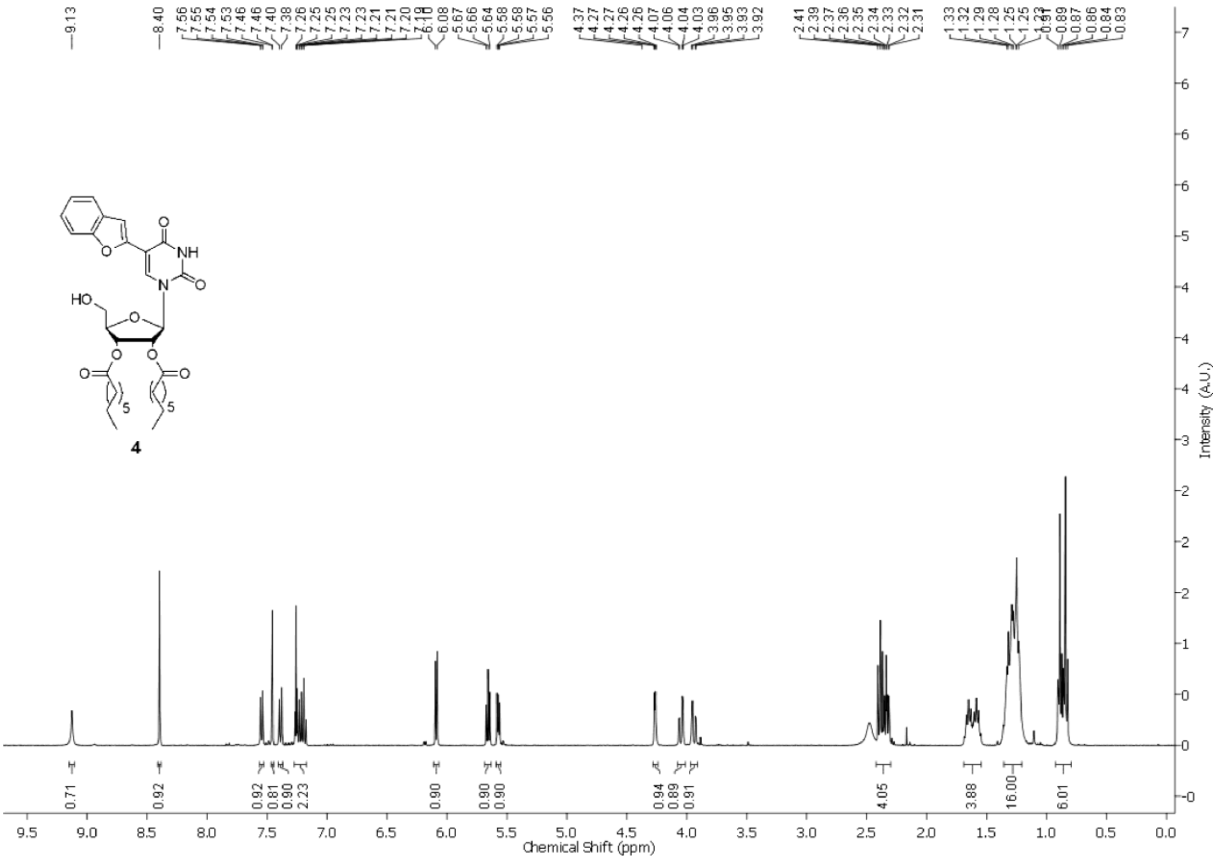
5.9.1.3 Crystallographic data for nucleolipid **9**

Compound identity	**9**
CCDC (deposition number)	1423141
Empirical formula	$C_{45}\,H_{68}\,N_2\,O_8\,S$
Formula weight	796.46
Temperature	296(2) K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	P1
Unit cell dimensions	a = 5.4174(5) α = 95.858(2)
	b = 9.3296(9) β = 92.994(2)
	c = 24.033(2) γ = 93.330(2)
Volume	1204.2(2)
Z	2
Density (calculated)	1.207 Mg/cm^3
Absorption coefficient (μ)	0.122 mm^{-1}
F(000)	474
Crystal size	0.35 x 0.15 x 0.04 mm^3
Theta range for data collection	4.23 to 27.59
Index ranges	-7<=h<=7, -12<=k<=9, -30<=l<=31
Reflections collected	40246
Independent reflections	8197[R(int) = 0.0612]
Completeness to theta = 27.59 °	99.2 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.978 and 0.995
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	8197 / 9 / 520
Goodness-of-fit on F^2	1.054
Final R indices [I>2sigma(I)]	R1 = 0.0676, wR2 = 0.1557
R indices (all data)	R1 = 0.0918, wR2 = 0.1700
Largest diff. Peak and hole	0.723 and -0.661 e.$Å^{-3}$
CCDC	1423141

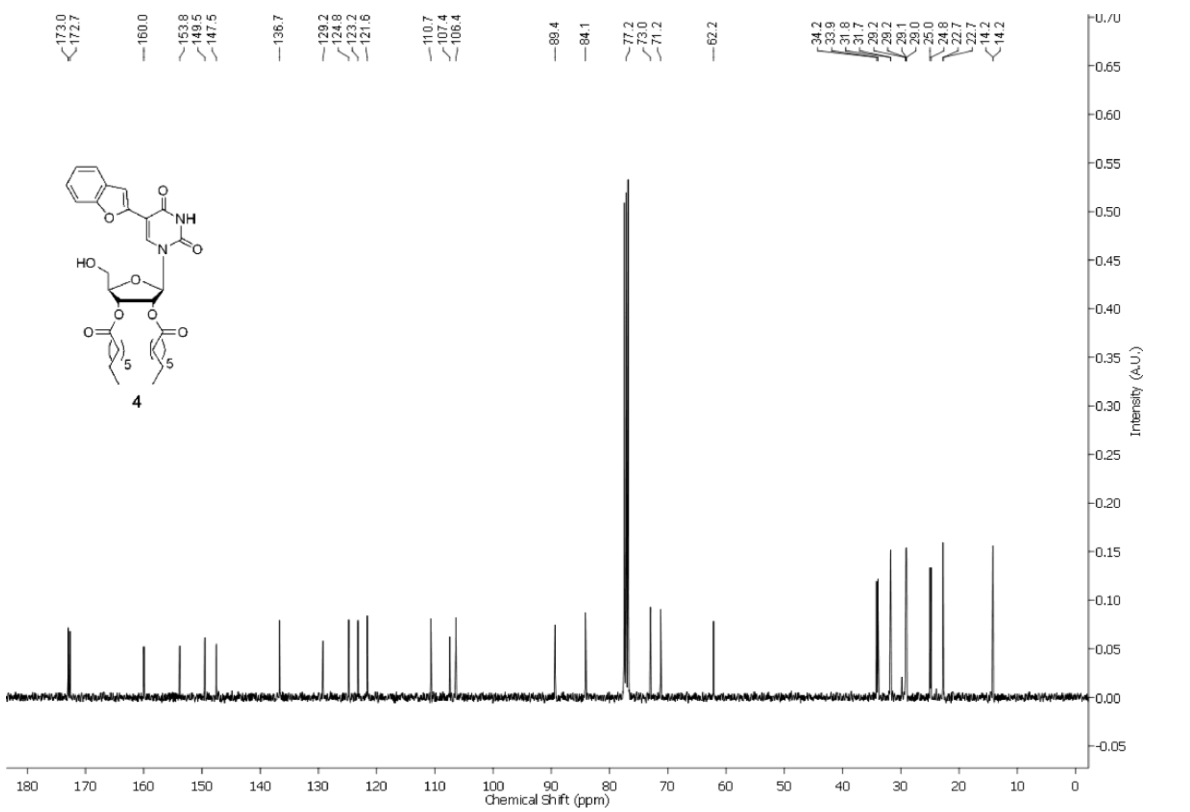
5.9.1.4 Crystallographic data for nucleolipid **10**

Compound identity	**10**
CCDC (deposition number)	1423142
Empirical formula	$C_{49} H_{76} N_2 O_8 S$
Formula weight	852.53
Temperature	296(2) K
Wavelength	0.71073 Å
Crystal system	Triclinic
Space group	P1
Unit cell dimensions	a = 5.385(8) α = 92.65(2)
	b = 9.274(13) β = 95.509(19)
	c = 25.69(4) γ = 93.348(20)
Volume	1273(3)
Z	1
Density (calculated)	1.215 Mg/cm^3
Absorption coefficient (μ)	0.119 mm^{-1}
F(000)	506
Crystal size	0.39 x 0.18 x 0.03 mm^3
Theta range for data collection	4.27 to 25.02
Index ranges	-6<=h<=6, -10<=k<=10, -30<=l<=30
Reflections collected	11701
Independent reflections	5427[R(int) = 0.1855]
Completeness to theta = 25.02°	92.2 %
Absorption correction	MULTI-SCAN
Max. and min. Transmission	0.975 and 0.996
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	5427 / 168 / 569
Goodness-of-fit on F^2	1.001
Final R indices [I>2sigma(I)]	R1 = 0.1082, wR2 = 0.2511
R indices (all data)	R1 = 0.2437, wR2 = 0.3229
Largest diff. Peak and hole	0.449 and -0.428 e.Å^{-3}
CCDC	1423142

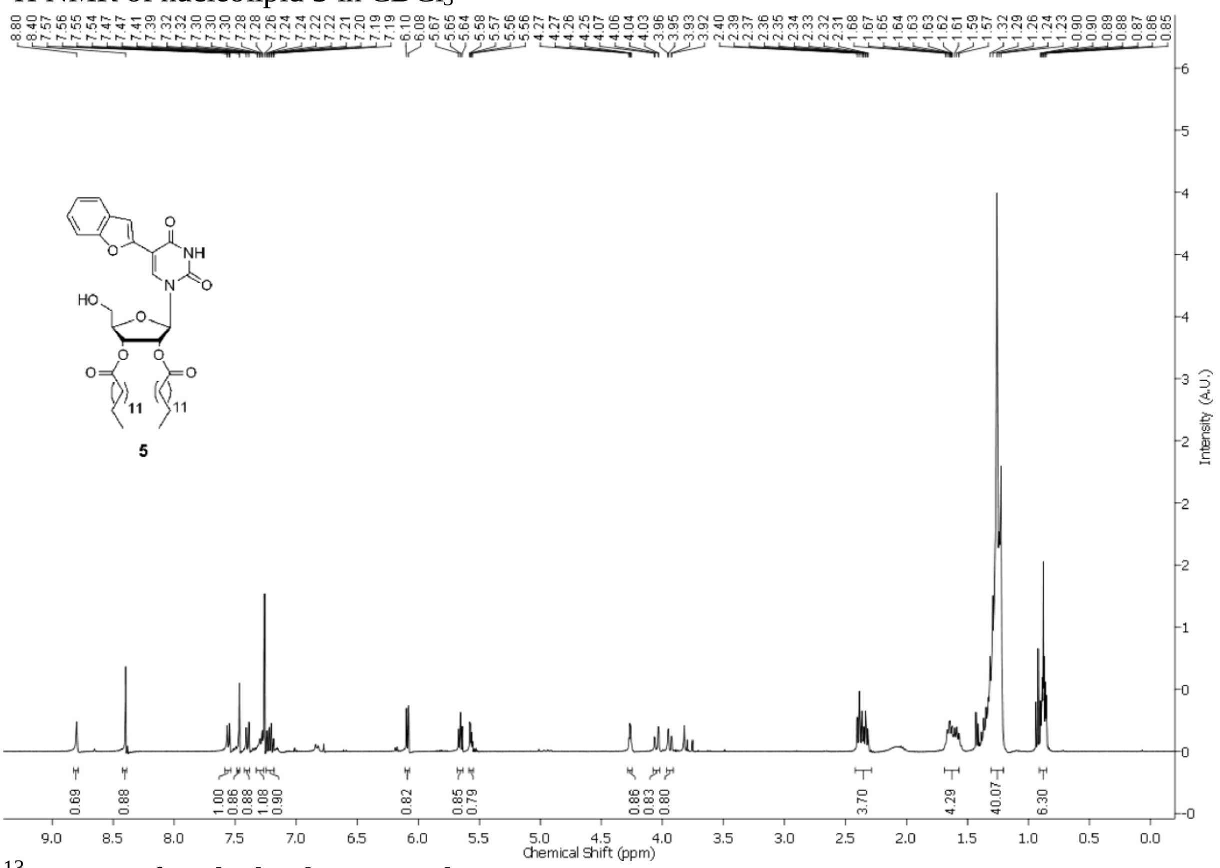
^{1}H NMR of nucleolipid **4** in $CDCl_3$



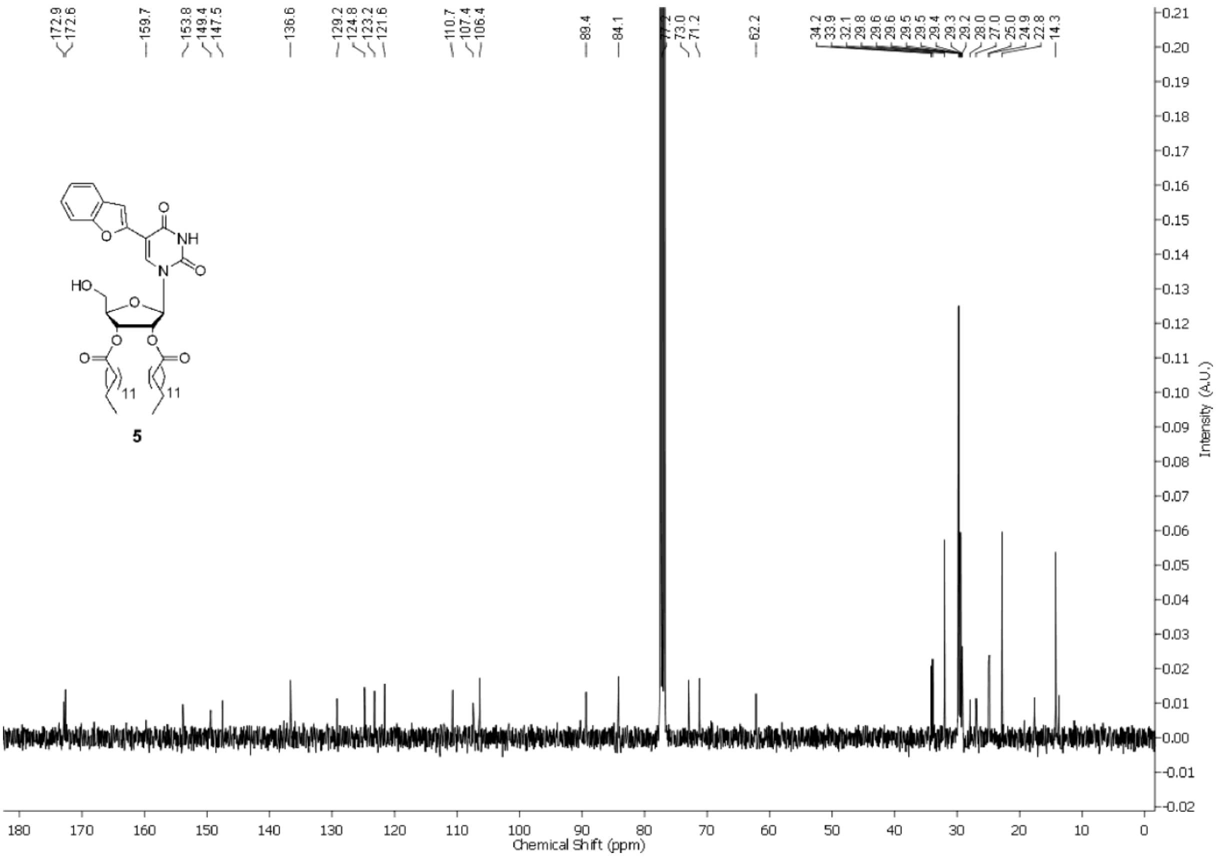
^{13}C NMR of nucleolipid **4** in $CDCl_3$



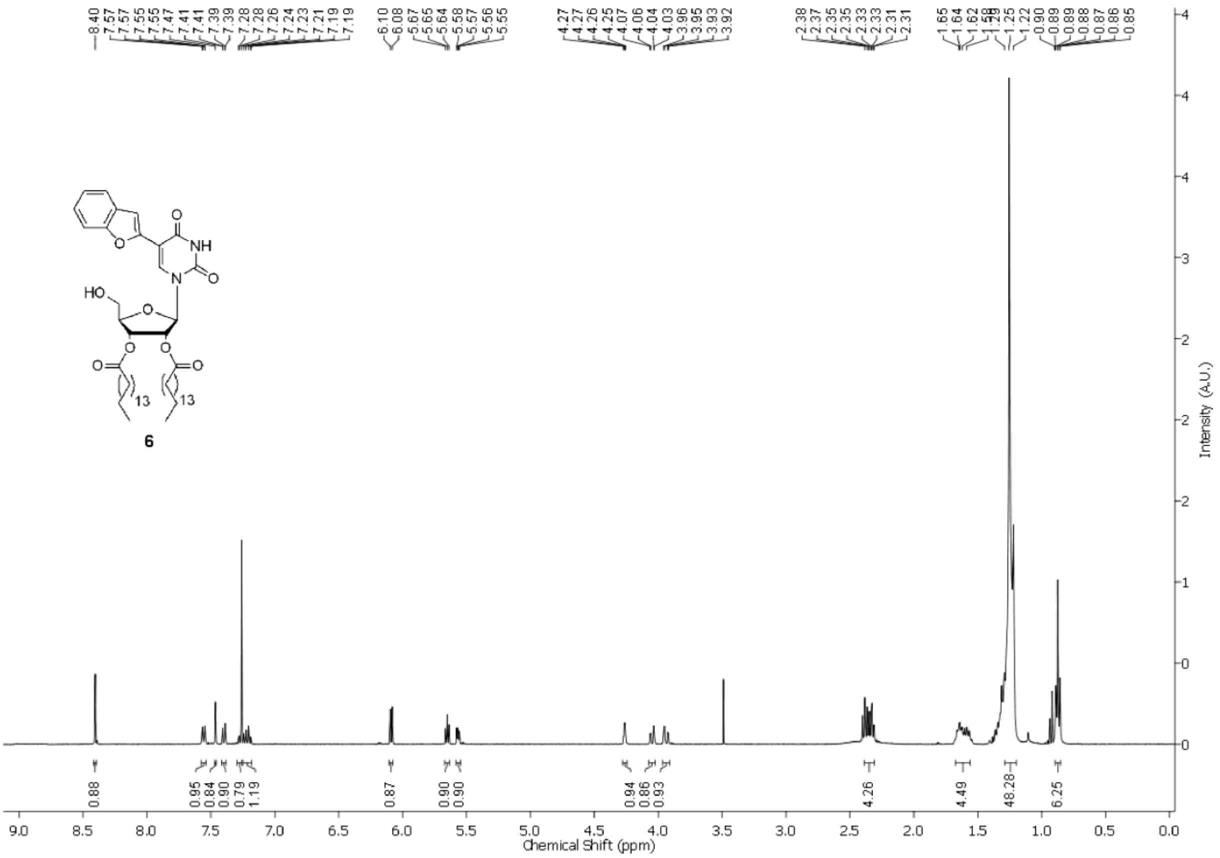
^{1}H NMR of nucleolipid **5** in $CDCl_3$



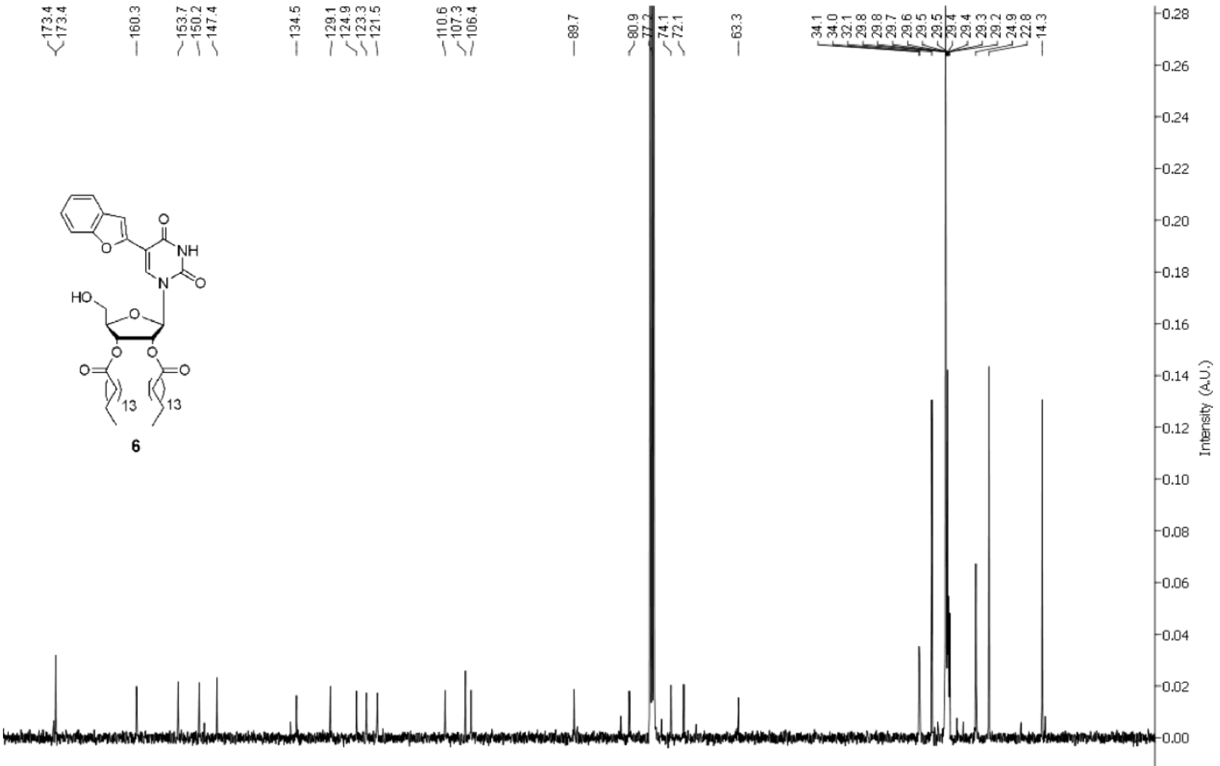
^{13}C NMR of nucleolipid **5** in $CDCl_3$



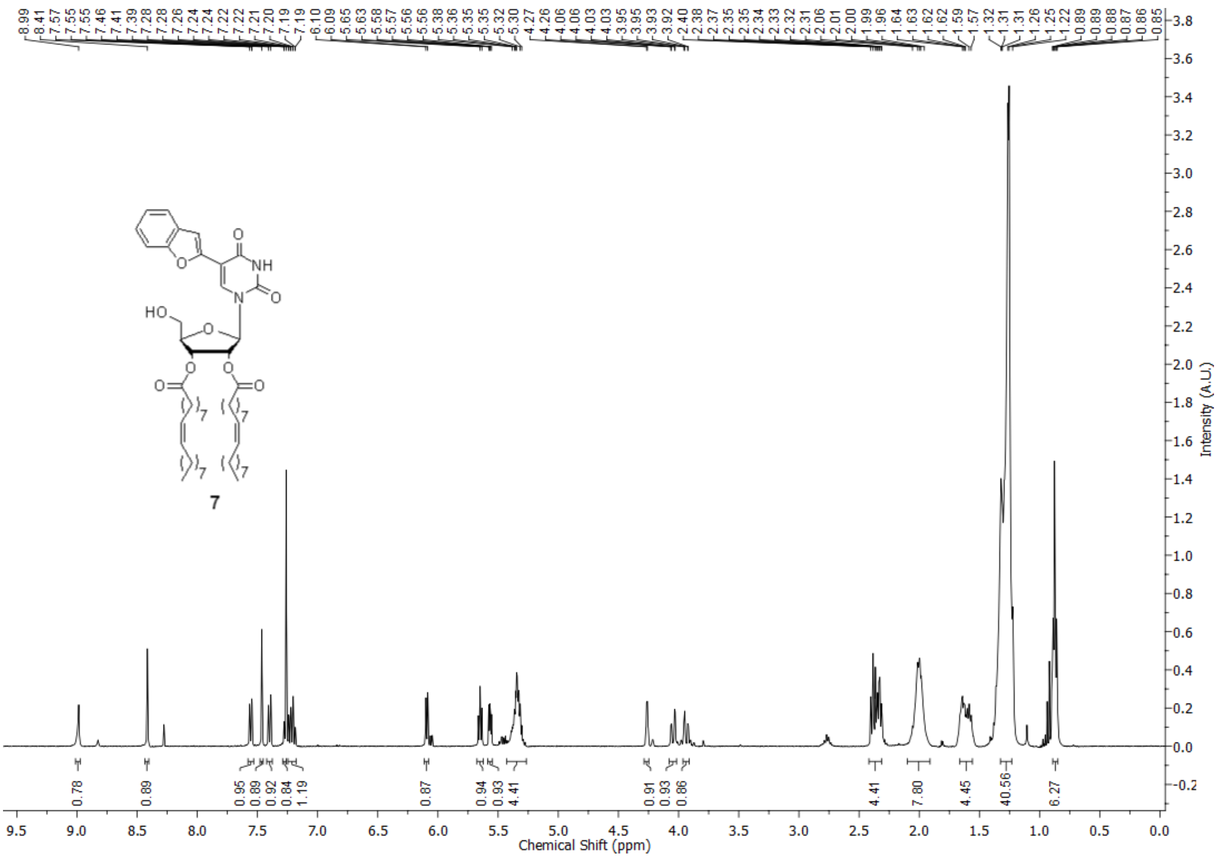
^{1}H NMR of nucleolipid **6** in $CDCl_3$



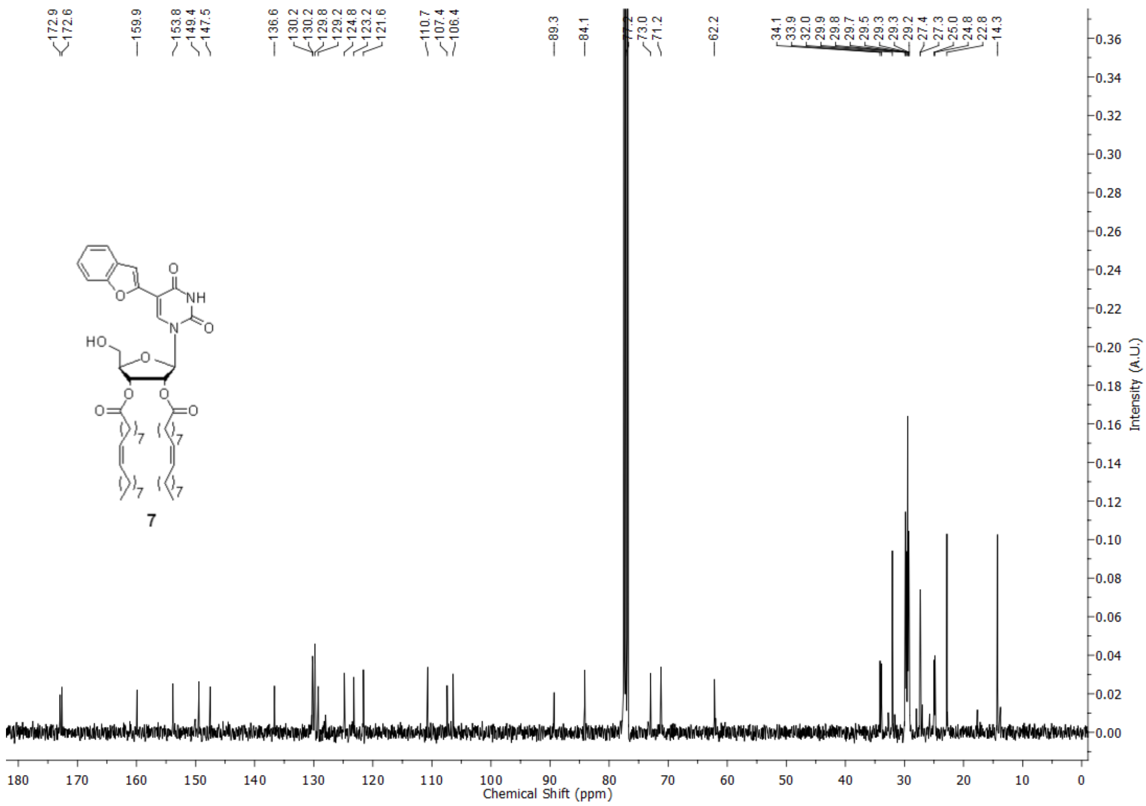
^{13}C NMR of nucleolipid **6** in $CDCl_3$



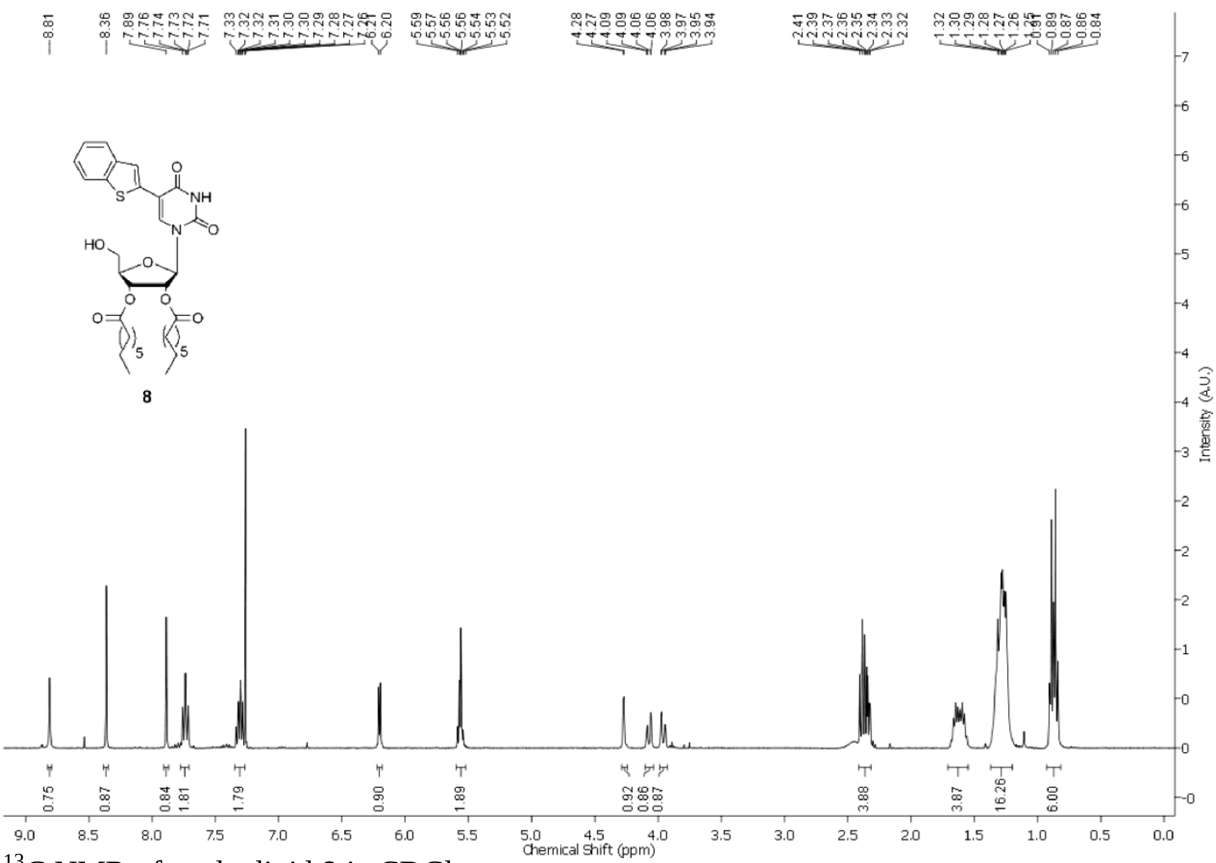
[1]H NMR of nucleolipid **7** in $CDCl_3$



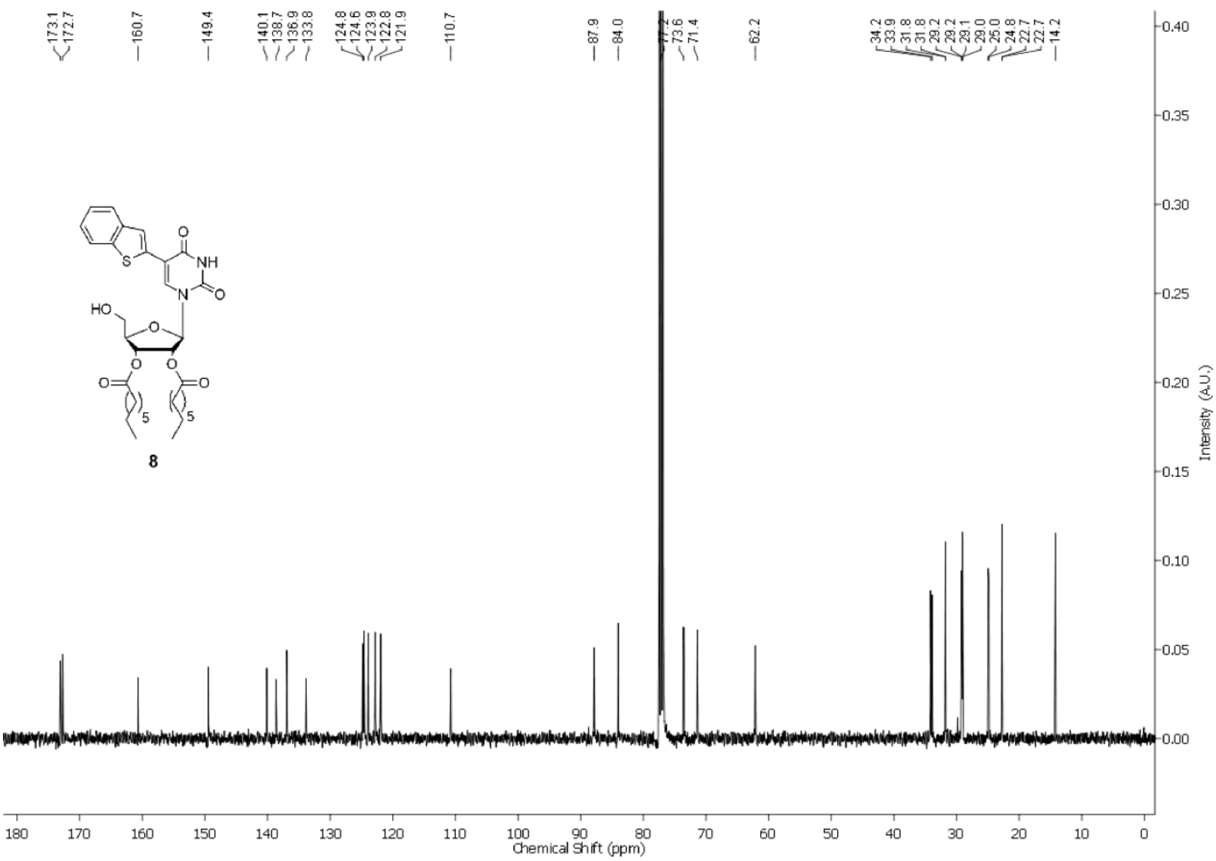
[13]C NMR of nucleolipid **7** in $CDCl_3$



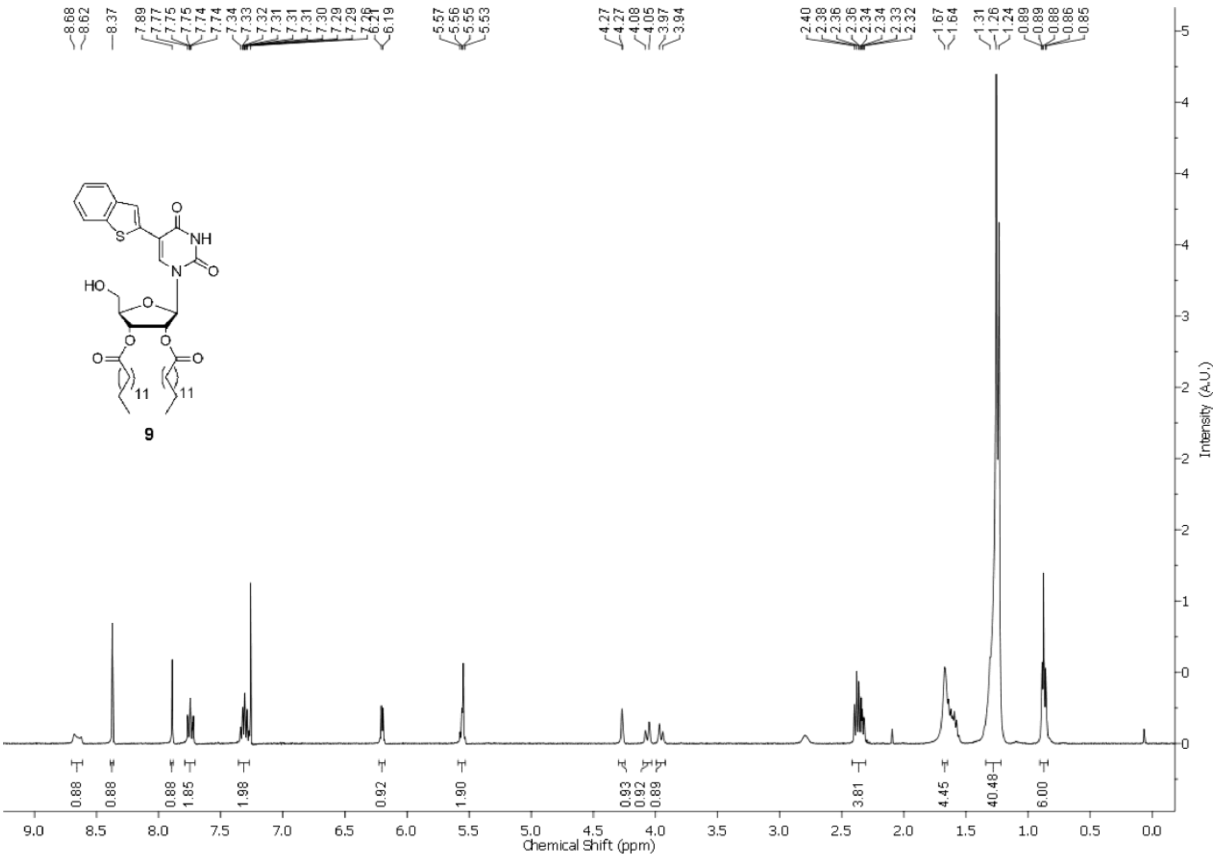
^{1}H NMR of nucleolipid **8** in $CDCl_3$



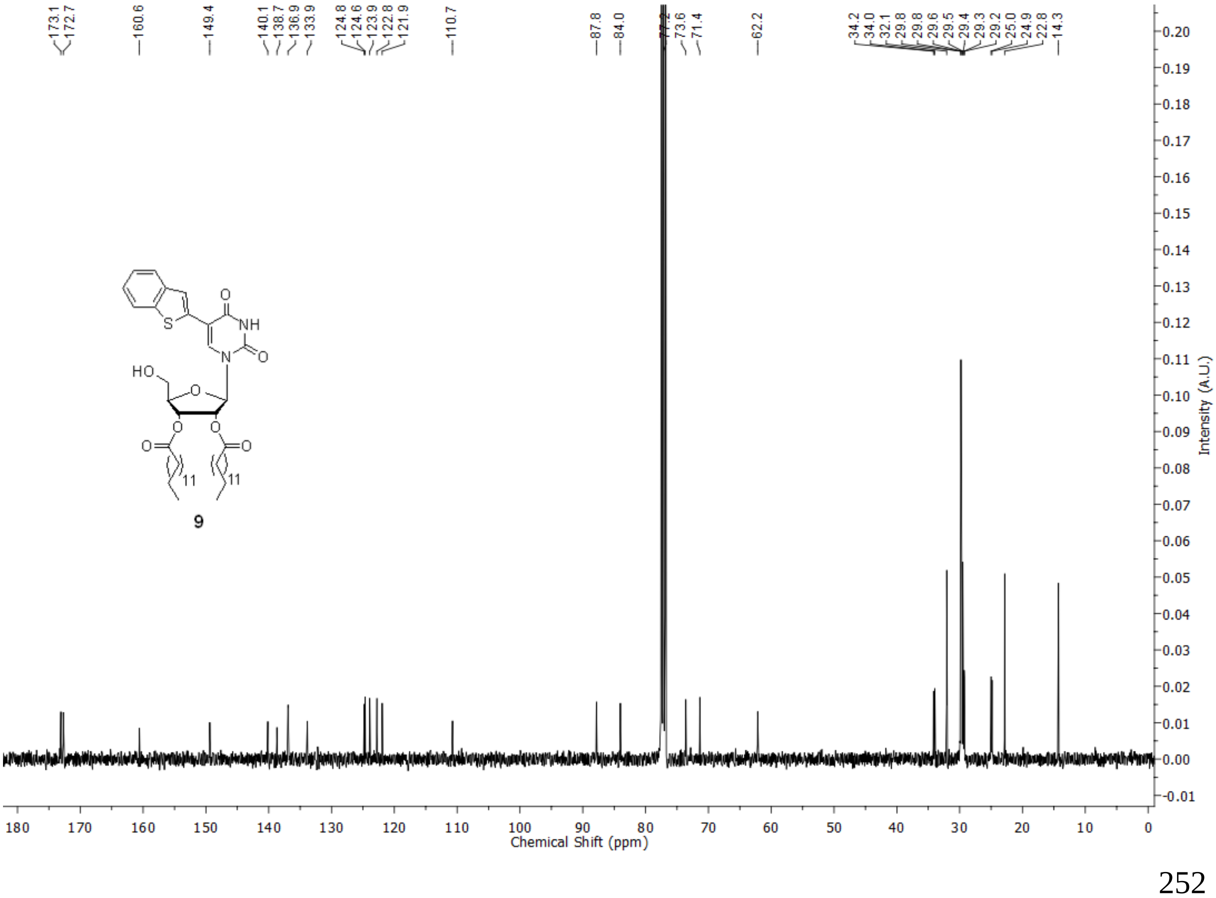
^{13}C NMR of nucleolipid **8** in $CDCl_3$



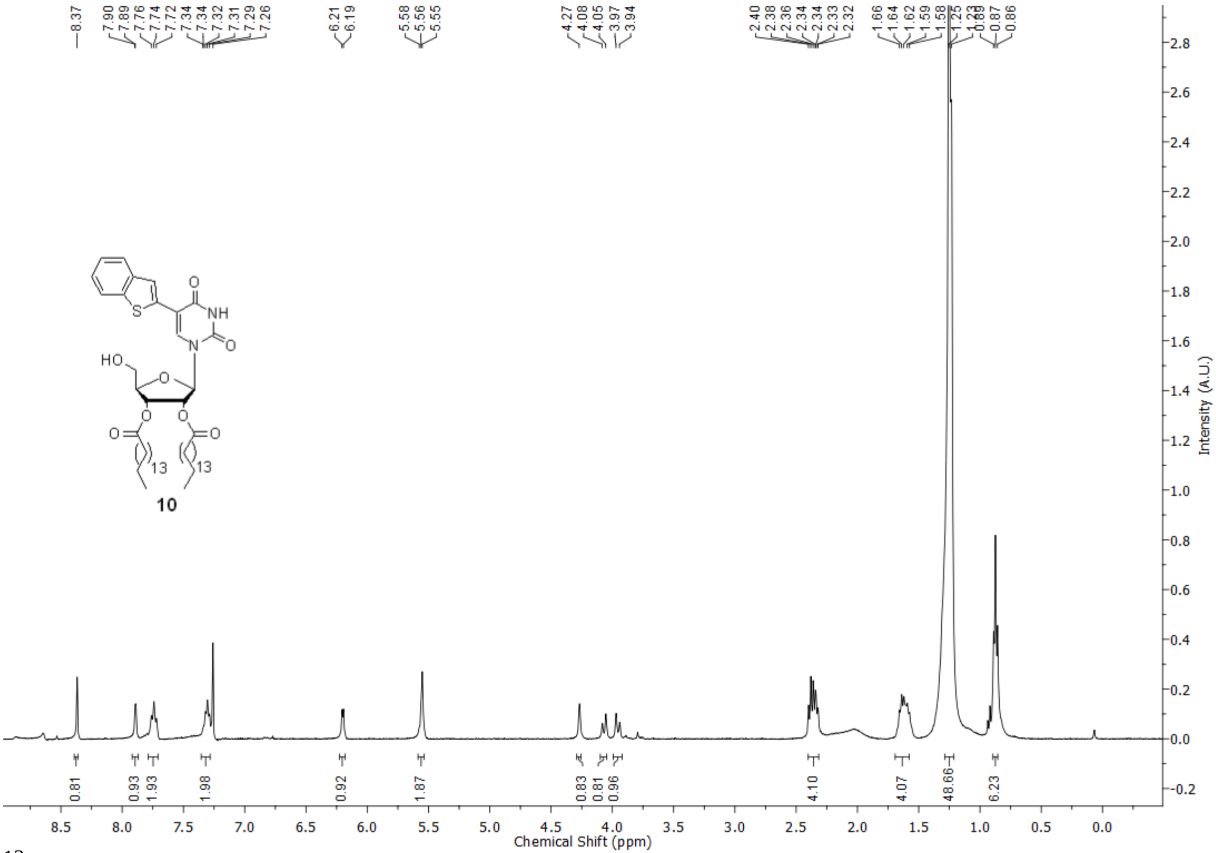
^{1}H NMR of nucleolipid **9** in $CDCl_3$



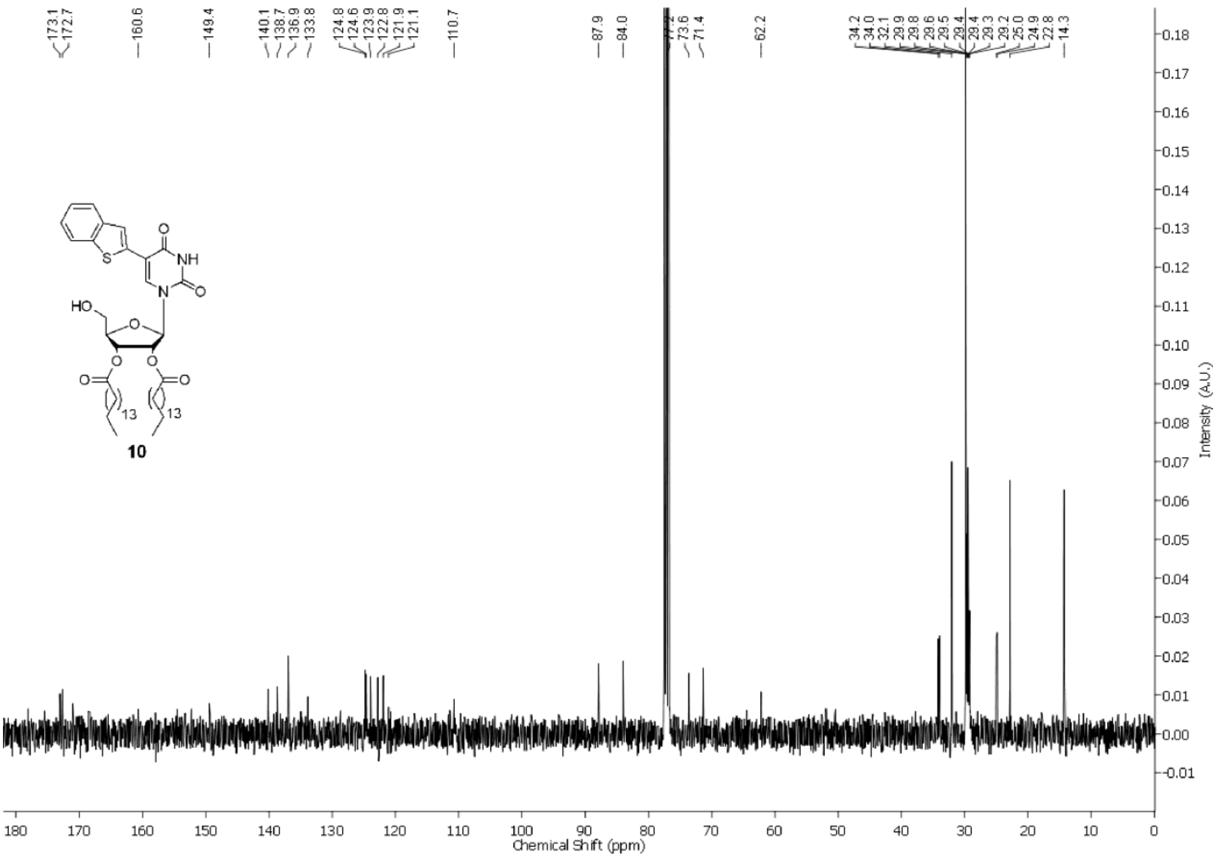
^{13}C NMR of nucleolipid **9** in $CDCl_3$



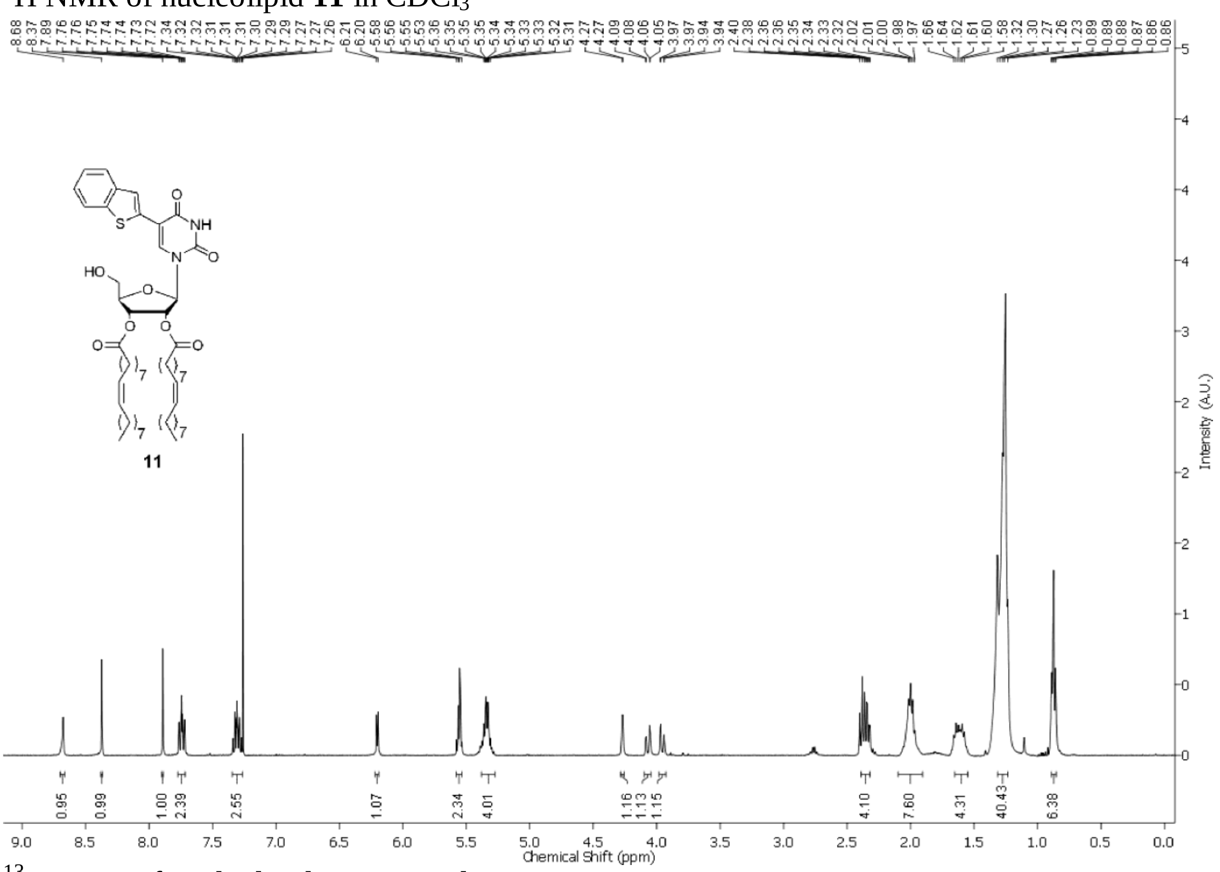
^{1}H NMR of nucleolipid **10** in $CDCl_3$



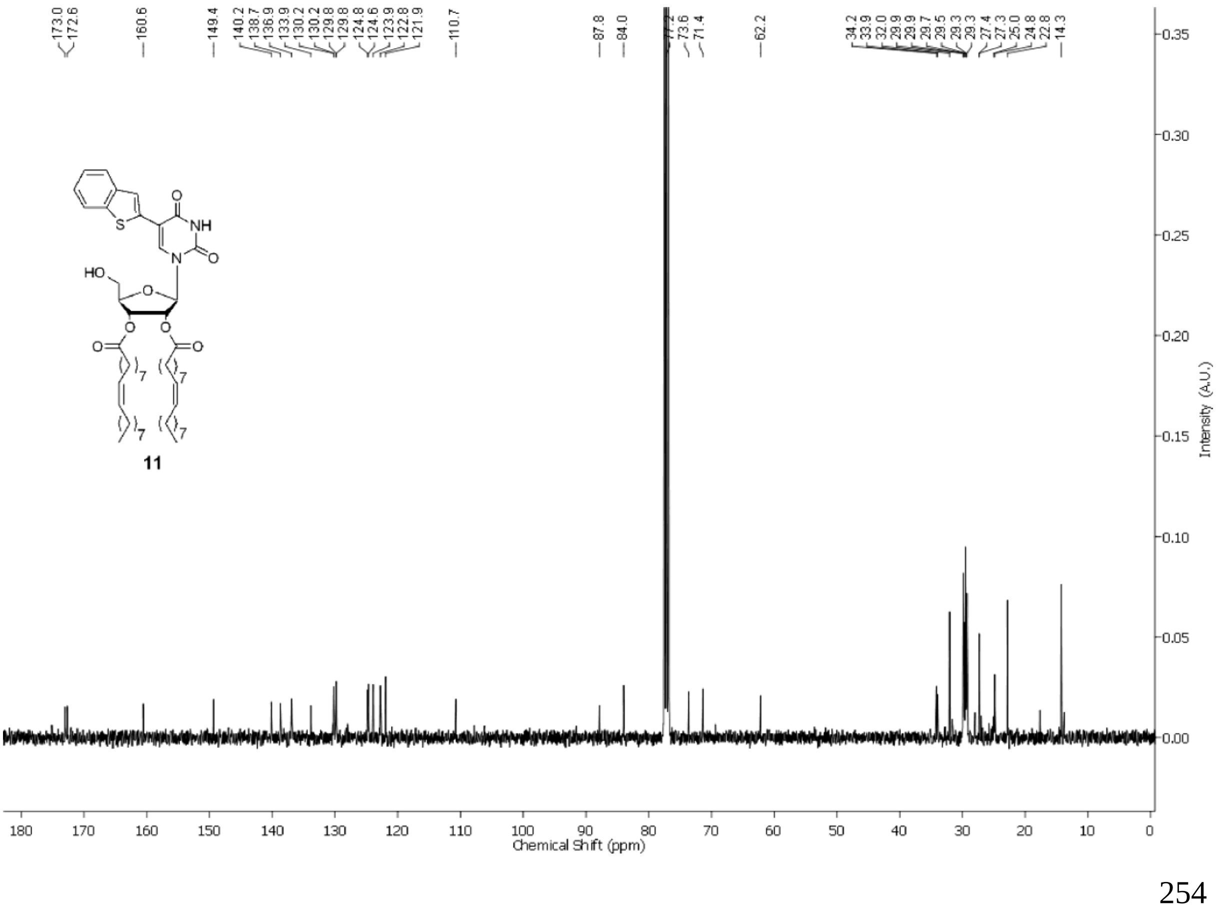
^{13}C NMR of nucleolipid **10** in $CDCl_3$



^{1}H NMR of nucleolipid **11** in $CDCl_3$



^{13}C NMR of nucleolipid **11** in $CDCl_3$



General conclusions and future perspectives

A novel dual-propose nucleoside probe system made of selenophene has been synthesized, which serves both as a conformation-sensitive fluorescent reporter and an anomalous X-ray diffraction label. The selenophene modified ribonucleoside incorporated into the bacterial ribosomal decoding site RNA enabled the monitoring of the RNA-antibiotic interaction by fluorescence and X-ray crystallography. Further, the utility of the dual-purpose probe was expanded by detecting the formation of non-canonical GQ structure of human telomeric DNA repeat and its ligand binding in real time by fluorescence and in solid-state by X-ray crystallography. Importantly, the diffraction signal from Se atom held in determining the solution of human telomeric GQ structure. X-ray, thermal melting and CD studies indicated that the selenophene modification had only a minor impact on the native fold of the target oligonucleotides. Further, fluorescence data obtained under equilibrium conditions and crystal structure obtained from X-ray analysis using our dual-purpose could be directly correlated with each other, which otherwise is not very easy with uniquely labeled probes.

Our probe represents a new class of biophysical tool that not only provides a means to concurrently use two complementing biophysical techniques to study nucleic acids but also could support discovery approaches to indentify efficient binders against a given target. In this context, future directions will involve implementation of these analogs in different targets like hepatitis C IRES RNA motif, i-motifs, etc. In parallel, colleagues in my group are working towards the development of next generation dual-purpose nucleoside probes with better photophysical properties (high quantum yield and excitation and emission the visible region).

We have also demonstrated novel nucleoside-fatty acid hybrids nucleolipids design and self-assembling process. Through this study, we have demonstrated how small chemical variations like site of modification (fluorophore/fatty acid), nature of fatty acid chain (saturated or unsaturated), and number of fatty acids acyl chain conjugated to nucleoside, can produce different self-organized materials via different hierarchical self-assembling process. Aggregation-induced fluorescence, surface-tunability, self-sorting, co-assembly behavior and multistimuli responsiveness are some of the features, which we could invoked by small variation in the chemical nature of the nucleolipid. Currently, we are in the process of expanding the utility of our supramolecular design to other fluorophores and nucleosides.