Université catholique de Louvain

Institute of Condensed Matter and Nanosciences

Bio- and Soft Matter (UCLouvain/IMCN/BSMA)

Chiral sequence-defined oligomers for self-assembly and catalysis

Qian Qin

Jury members:

Prof. Alain Jonas (UCLouvain, BE), supervisor

Prof. Karine Glinel (UCLouvain, BE), supervisor

Prof. Jean-François Gohy (UCLouvain, BE), president

Prof. Charles-André Fustin (UCLouvain, BE), member

Prof. Anja Palmans (Tu/e, NL), member

Dr. Antony Fernandes (Certech, BE), member

Dr. Jie Li (BUCT, CHN), member

Thesis printed in partial fulfilment of the degree of

Doctor in Engineering sciences and technology

Louvain-la-Neuve, March 2025

Acknowledgements

The writing comes to an end with myriad emotions stirring within me. First and foremost, I must express my deepest gratitude to my two supervisors, Prof. Alain Jonas and Prof. Karine Glinel, for providing me with this precious opportunity to pursue a doctoral degree. They are diligent, rigorous, and open-minded, always granting students ample freedom to pursue their ideas as well as being readily available for discussions. It's impossible for me to successfully complete my PhD thesis without their patient supervision and ongoing assistance. In daily life, they are very nice, modest, and eager to help whenever feasible. I appreciate their concern for me and my family, their encouragement when I was struggling, and willingness to share my happiness. How lucky I am to have two of the best professors in my life!

I am grateful to the EOS project and team. The EOS project not only provided me with financial support but also served as a platform for discussion and learning with other participants. Most importantly, within the framework of the EOS project, I met an amazing group. In addition to my two supervisors, the team includes several postdoctoral fellows: Dr. Jie Li, Dr. Antony Fernandes, Dr. Kseniia Grafskaia, and Dr. Imane Nekkaa. During my research, they were very supportive and gave me great suggestions. Especially Dr. Jie Li, who was not only my colleague but also my teacher and close friend. He taught me how to perform organic synthesis and encouraged and guided me when I encountered difficulties. Thanks to this superb team for making every meeting so fruitful.

I have to say thank you to my (ex)colleagues, because they have

been highly helpful in both my studies and my personal life. They taught me a lot, not just about the use of equipment, but also about Belgian culture. When I turn to them, I can receive a lot of guidance and inspiration. In addition, we had a great time doing a variety of interesting activities together. However, it is difficult to mention all of them due to the potential of missing someone.

Thank you, my Chinese friends, for brightening up my otherwise dreary life. I'll never forget the beautiful scenery we saw together, the games we played, and the excellent cuisine we ate together. Even though the wonderful times are transient, our friendship will last forever. It would be impossible to name all my dear friends here, but you guys are all sincerely remembered in my heart.

Finally, I'd like to express my endless appreciation to my family. My parents, despite your limited education, understood that knowledge was invaluable. You work the hardest but make the least amount of money, which you generously spend on supporting me achieve higher education. Although you can only say the most straightforward phrases to me and perform the most ordinary things for me, you are unaware that this is the most ordinary, unselfish, and greatest love in the world. Thank you for your unselfish devotion and unconditional affection. My husband has always respected, supported, encouraged, and tolerated me. You are not only my strongest supporter, but also a fantastic partner who helps and monitors each other as we strive to progress together. Thank you for your commitment and sacrifice for our family. My son, you truly are a little angel. Thank you for choosing us as your parents. You inspire us to work hard for a better life and give us the courage to overcome various challenges. I wish you healthy and happy growth. We will all love you forever! (最后，我想感

谢的我家人。我的父母，尽管学历有限，但深知知识无价。你们用最辛苦的方式挣着微薄的收入，却毫无保留的来支撑我完成学业。你们对我说着最朴素的话，为我做最平常的事，殊不知这正是这个世界上最平凡、最无私、最伟大的爱。谢谢你们的无私奉献和无条件的爱。我的丈夫，你一直尊重我、鼓励我、支持我和包容我。你不仅是我最坚实的后盾，也是我相互扶持相互监督共同进步的好伙伴。谢谢你为我们家庭的付出和坚守。我的儿子，你真是一个小天使。谢谢你选择我们做你的爸爸妈妈。你是我们奔赴美好生活的动力，你是我们直面重重困难的勇气。儿子，愿你健康快乐的成长，爸爸妈妈永远爱你!)

Table of Contents

List of figures

List of Abbreviations and Symbols

(c)SUMI	(cationic)Single-unit monomer insertions
Δ	Time between two pulses
ΔH	Standard binding enthalpy
ΔS	Standard binding entropy
ΔS_c°	Standard entropy penalty
μ_0	Viscosity
1D/2D/3D	One/two/three dimensional
4-NPC	4-Nitrophenyl chloroformate
A	Adenine
ACN	Acetonitrile
AzUTMS	11-Azidoundecyltrimethoxysilane
BnOH	Benzyl alcohol
Boc	*Tert*-butyloxycarbonyl
bpy	Bipyridine
BzH	Benzaldehyde
C	Cytosine
C_6	Hexane
CA	Carbamate
calcd.	Calculated
C_b	Betweenness centrality
CB[7]	Cucurbit[7]uril
C_c	Closeness centrality
CD	Circular dichroism

CD_3CN	Deuterated acetonitrile
CSP	Chiral stationary phase
CuAAC	Copper-catalyzed azide–alkyne cycloaddition
$CuSO_4$	Copper(II) sulfate
D	2,6-Diaminopyridine derivative
D	Diffusion coefficient
D-A	Donor-Accepter
DCM	Dichloromethane
DFT	Density functional theory
DMF	N,N-dimethylformamide
DMSO-d_6	Deuterated dimethyl sulfoxide
DOSY	Diffusion Ordered Spectroscopy
DTBN	Di-tert-butylnitroxyl
DTS	DNA-templated synthesis
E_a	Activation energy
ee	Enantiomeric excess
eq.	Equivalent
Et_3N	Triethylamine
EtOAc	Ethyl acetate
EtOH	Ethanol
G	Guanine
g	Gradient strength
GC	Gas chromatography
H-bonding	Hydrogen bonding
HPLC	High-performance liquid chromatography

Hz	Hertz
I	Imidazole
IEG	Iterative exponential growth
IEG+	Iterative exponential growth plus side-chain functionalization
ISG	Iterative stepwise growth
J	Coupling constant
K	Binding constant
LC-QTOF-MS/MS	Qualitative tandem liquid chromatography quadrupole time of flight mass spectrometry
LRP	Living radical polymerization
MD	Molecular dynamics
MOF	Metal organic framework
MPLA	Monodisperse poly(lactic acid)
MS	Mass spectrometry
n	Molar concentration
Na_2EDTA	Ethylenediaminetetraacetic acid disodium salt dihydrate
Na_2SO_4	Sodium sulfate anhydrous
NaClO	Oxidant sodium hypochlorite
$NaHCO_3$	Sodium bicarbonate
NMI	N-methylimidazole
NMR	Nuclear magnetic resonance spectroscopy
OAc	Acetyl
OBn	Benzyl
OEGs	Oligo(ethylene glycol)s
P	Pyridine

p	The probability of formation of a complete catalytic site
P-3CR	Passerini three-component reactions
PEG	Poly(ethylene glycol)
PINO	Phthalimide N-oxyl radical
ppm	Parts per million
pyta	Pyridine-triazole
q	The number of equivalent protons
r	Hydrodynamic radius
R	Ideal gas constant
R/S	Right/left absolute configuration
RCA	Rolling circle amplification
R_g	Radius of gyration
RT	Room temperature
s/d/t/m	Singlet/doublet/triplet/multiplet
SCNP	Single-chain polymeric nanoparticle
SDP	Sequence-defined polymer
SFC	Supercritical fluid chromatography
ssDNA	Single-stranded DNA
SuFEx	Sulfur-fluoride exchange reaction
T	Thymine
T	Temperature
TBAF	Tetrabutylammonium fluoride
TBS	Tert-butyl(dimethyl)silyl
Tempo（M）	2,2,6,6-Tetramethylpiperidinyl-1-oxy

TFA	Trifluoroacetic acid
TIPS	Triisopropylsilyl
TOF or *f*	Turnover frequency
Ugi-4CR	Ugi four-component reaction
γ	Proton magnetogyric ratio
δ	Duration of a pulse
δ or *I*	Chemical shift
ε	Dielectric constant

Chapter 1 Motivation and Objectives

Catalysis is a crucial method for accelerating chemical reactions. In enzyme catalysis, the various chemical groups involved in the catalytic site assemble in spatially close positions, leading to optimal availability and often adaptive activity. This optimized self-assembly of active groups is intrinsically dynamic, allowing accessibility to and from the reactive site, and the realization of the reaction transition state.[1,2] Although biomolecule-based catalysis is efficient and fascinating, biomolecular catalysts are hindered by limitations such as susceptibility to the reaction environment and the high costs associated with their production and storage, which restrict their application.

To overcome these shortcomings of biomolecular catalysts, chemists have attempted, among other strategies, to utilize synthetic molecules with catalytically reactive units for the purpose of creating catalytic structures by self-assembly. This method allows for the construction of dynamic active sites, which are essential for the development of highly efficient catalysts reminiscent of biomolecular ones. A series of synthetic catalysts relying on some form of self-assembly to create catalytic sites have been developed, comprising peptides,[3–10] DNA scaffolds,[11–17] single-chain polymeric nanoparticles (SCNPs),[18–25] peptoids,[26,27] metal organic frameworks (MOF),[28–30] supramolecular polymers,[31] foldamers,[32,33] and helical (supramolecular) polymers.[34–37] However, high catalytic efficiency via the self-assembly of active groups similar as what happens in enzymes, remains challenging in synthetic molecular catalysts. Nevertheless, the pursuit of improving artificial enzymes has remained uninterrupted. For instance, it was demonstrated that supramolecular and covalent macrocycles can be

synthesized to create catalytic sites involving the concerted action of different functional groups,[38–42] because cyclic configurations decrease the accessible molecular conformational space and thereby increase the probability to bring catalytically-active groups in spatially-close positions. These inspiring examples motivated us to try and synthesize molecules capable to self-assemble in dynamic catalytic supramolecular macrocycle structures, simultaneously allowing accessibility to and from the reactive site, while containing catalytic groups in close position.

More precisely, the main goal of this thesis is to design precision oligomers and use them to construct self-assembled supramolecules with synergistic catalytic groups positioned in reactive pockets for the aerobic oxidation of alcohols, as shown in Figure 1.1. To create the catalytic pocket, two base pairs were selected as recognition units: guanine (G)/cytosine (C) and 2,6-diaminopyridine derivative (D)/thymine (T). The complementary bases are individually inserted at the ends of two strands to realize the self-assembly of the oligomers. The cooperative catalytic groups involved in the copper-catalyzed aerobic oxidation of alcohols, which are distributed between the two oligomers, include Tempo (2,2,6,6-tetramethylpiperidinyl-1-oxy, M), pyridine (P) and imidazole (I). By properly designing the sequence order of the functional moieties to form supramolecular macrocycles, we wanted to demonstrate that it is possible to reach an efficient spatial positioning of the cooperative catalytic residues within a confined space.

To achieve this, two sub-objectives are proposed in this text. In chapter 3, we describe the methodology to synthesize the sequence-defined oligomers, including a description of fluoride-mediated desilylation of alkynes, and the application of copper-catalyzed azide–

alkyne cycloadditions (CuAAC), as well as the verification of their structures and sequence orders, including by nuclear magnetic resonance spectroscopy (NMR), mass spectrometry (MS), and qualitative tandem liquid chromatography quadrupole time-of-flight mass spectrometry (LC-QTOF-MS/MS). We then use this methodology to synthesise a series of short oligomers substituted by nucleic acid bases to demonstrate the versatility of our synthetic route. Moreover, these oligomers were used in an independent study to investigate recognition processes between flexible short oligomers equipped with complementary and non-complementary nucleobase analogues.

In chapter 4, we switch to our main goal which is to design precision oligomers with complementary hydrogen-binding recognition units and cooperative catalytic groups involved in the copper-catalyzed aerobic oxidation of alcohols, investigate their self-assembly behaviour including with molecular dynamics (MD) simulations performed in collaboration with the group of Mathieu Surin from University of Mons, and study and model their catalytic activity.

Before detailing this work, Chapter 2 briefly reviews the background knowledge related to self-assembly, notably based on H-bonding, synthetic methods and characterization of precision macromolecules as well as their applications including the aerobic oxidation of alcohols. Chapter 5 finally summarizes the main results and discussions derived from our work and presents outlooks for future work.

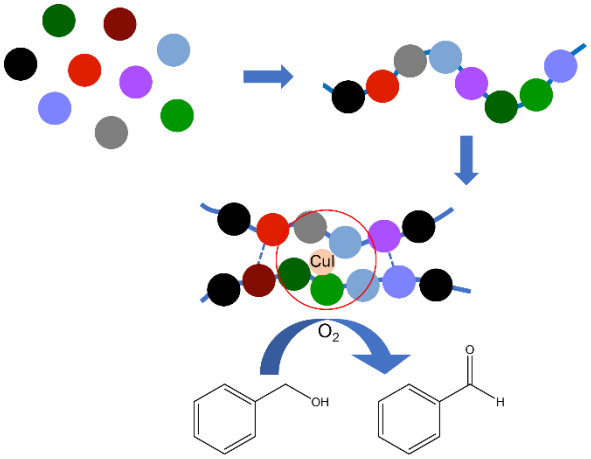


Figure 1.1 Schematic representation of the objective of thesis.

Chapter 2 General Introduction and State-of-the-Art

Abstract: In this chapter, the background and methodology related to this thesis are presented. Since this thesis focuses on designing a self-assembled sequence-defined catalytic pocket using precision oligomers ended with DNA-like bases, a general introduction to self-assembly and hydrogen based self-assembly is first given in Section 2.1. Then, the importance to control sequence and chirality, and synthetic strategies to reach this aim, are discussed, before methods of characterization and application of precision polymers are briefly presented. Finally, systems mimicking enzymes involved in oxidation of alcohols are discussed, and the copper/TEMPO-mediated aerobic oxidation of alcohols, which is the specific catalytic reaction used to demonstrate our concept, is introduced.

“In design, Mother Nature is our best teacher.”

— Van Day Truex

Nature has designed biomaterials using evolution for wide purposes, such as catalysis, molecular recognition, and data storage. Those sophisticated biomaterials are inspiring for scientists but are still too complex to be synthesized from scratch. Among the lectures that nature has given us, one of the most important parts is self-assembly.

2.1 Self-assembly

2.1.1 Introduction to self-assembly

The main idea of self-assembly is that disordered building blocks/subunits spontaneously may form ordered structures owing to local interactions among the components, without any external work.[43,44] From the concept, we can easily find there are three critical features for self-assembly. First, the free energy of the mixture of building blocks and self-assembled system should reach a minimum for a high concentration of self-assembled system (excluding kinetic trapping and dynamic assembly from our discussion). Therefore, assembly will be thermodynamically favored, and the lifetime of the self-assembled system will be significant. Second, the building blocks, which are very diverse, must comprise multiple chemical functions and have complementary "shapes" that allow the self-assembled system to be of low(er) energy, thereby favoring assembly despite it is entropically disfavored. Third, self-assembly needs sufficiently weak interactions compared to thermal energy, which allows the system to remain dynamic and form reversible adjustable

associations before reaching the final state.

Self-assembly thus involves a balance of multiple intermolecular forces, those forces being attractive (*e.g.*, for stabilization of the assembly) or repulsive (*e.g.*, for driving assembly to specific configurations). When the system has net interaction with an equilibrium separation between components, aggregation take places, as shown in Figure 2.1.[45] During self-assembly, normally the interactions are relatively weak interactions or weak covalent bonds, such as van der Waals, hydrogen bonds, coordination bonds, etc. (Table 2.1).[45] Self-assembly processes are thus determined by a delicate balance of attractive and repulsive opposition forces. The attractive force brings self-assembly building units together; the repulsive force balances the attractive force to tune the dynamics and configurations of self-assembled structures.

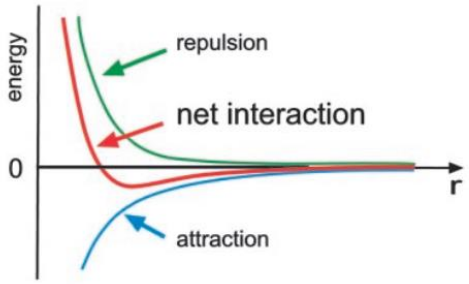


Figure 2.1 Typical scheme of the intermolecular or colloidal potential energies depending on the distance between two components. Adapted from[45]

Because self-assembly processes are determined by multiple intermolecular forces, the external environment can also influence strongly self-assembly.[46–48] For example, the polarity of the solvent can affect the strength of hydrogen bonding interactions and of

hydrophobic interactions. A polar solvent can disrupt hydrogen bonding interactions because it can solvate the hydrogen bond donors and acceptors, weakening the interaction. In contrast, a non-polar solvent, such as benzene, can enhance hydrogen bonding interactions because it cannot solvate the hydrogen bond donors and acceptors, leading to a stronger interaction. Hydrophobic interactions are a type of non-polar interaction between non-polar molecules or parts of molecules. A polar solvent can increase hydrophobic interactions because it cannot solvate the non-polar molecules, increasing their tendency to interact with each other. In contrast, a non-polar solvent can decrease hydrophobic interactions because it solvates the non-polar molecules, leading to a weaker interaction. pH is also an external variable that can impact self-assembly through its effect on the ionization state of functional groups on the molecules. For example, in a system with molecules containing carboxylic acid groups, as the pH is increased above the pK_a of the carboxylic acid group, the group becomes more deprotonated and negatively charged. This can lead to repulsion between molecules containing these groups and can disrupt a self-assembly process. Similarly, in a system containing amine groups, as the pH is decreased below the pK_a of the amine group, the group becomes more protonated and positively charged. This can also lead to repulsion between molecules and disrupt self-assembly. Besides, ionic strength can also impact self-assembly. The presence of ions can affect the solubility of the molecules and their ability to interact with each other. For example, increasing the concentration of salt in a solution leads to the screening of electrostatic interactions between charged groups on the molecules, reducing the strength of these interactions and potentially disrupting self-assembly. On the other hand, in some cases, adding salt can promote self-assembly by reducing the solubility of the

molecules and promoting their aggregation (salting out effect). In addition, temperature can affect self-assembly by influencing the rate and extent of molecular motion. As temperature is increased, the molecules in solution, on surface or in gas phase become more energetic and can move more freely. This increased mobility can lead to an increased rate of self-assembly and to the formation of larger, more complex structures. However, if the temperature becomes too high, the assemblies may be destabilized, due to the dominance of entropy over energy at higher temperature.

Table 2.1 Typical weak interactions or weak covalent bonds and force types in self-assembly. Adapted from[49]

Attractive Force	Repulsive Force
van der waals[a]	Electric double-layer[b]
Solvation	Solvation
Depletion	Hydration
Bridging	Steric
Hydrophobic	
π–π stacking	
Hydrogen bond	
Coordination bond[c]	

[a] Some cases of interaction between dissimilar colloidal objects can be repulsive.
[b] This force sometimes can be attractive when (1) interaction occurs between molecules or colloids with different charges, (2) with the same charge but at very small separation, and (3) between zwitterionic molecules and colloids.
[c] Coordination bond is a strong chemical bond compared with the rest of the forces, but serves as a unique attractive force for some of the supramolecular self-assembly systems.

Self-assembly is fascinating in nanotechnology because it allows for the creation of complex structures with minimal human intervention by controlling the molecular interactions and external environment, which makes it possible to ideally design and create complex

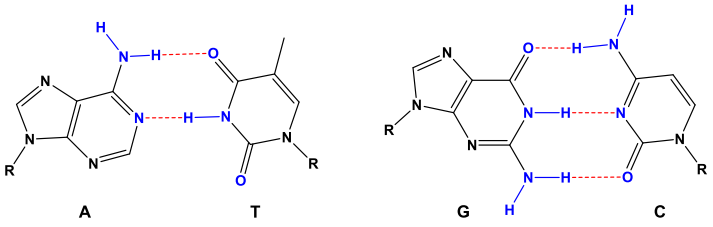
nanostructures with precise control over their size, shape, and properties.

2.1.2 Self-assembly driven by hydrogen bonds

Hydrogen bonding (H-bonding) is one of the most commonly used weak interactions to drive self-assembly and form supramolecular architectures; it has various advantages such as synthetic accessibility, high directionality, and tunable strength.[50–52] Theoretically, any atom or group with a higher electronegativity than its bonded hydrogen atom can create a dipole moment, allowing the hydrogen to act as a donor (D) and form a hydrogen bond with another more electronegative atom; however, such a weak bonding is usually considered as a van der Waals interaction rather than as H-bonding when the electronegativity difference is not large enough.[53] Therefore, standard hydrogen bonds are usually restricted to highly electronegative atoms such as oxygen (O), nitrogen (N) and fluorine (F) atoms.[53]

A single hydrogen bond is weak, with a typical energy of 5-30 kJ/mol, which is not strong enough to fabricate stable assemblies. A commonly used method to obtain more stable structures is to increase bonding strength and directionality by forming hydrogen-bonding arrays comprising multiple hydrogen bonds.[54] A well-known example is nucleobases of DNA (or of RNA), which are the basis of its structure and functions. The base pairs in DNA include adenine(A)/thymine(T) and guanine(G)/cytosine(G). They can form double or triple hydrogen arrays, as shown in Scheme 2.1; the binding constant of A/T and G/C are $\sim 10^2$ and $\sim 10^4$ in chloroform, respectively.[55] In fact, these natural nucleobases have already been used to functionalize molecules and macromolecules and direct

supramolecular self-assembly owing to their unique properties like reactivity, responsiveness, biocompatibility, etc.[50,56–58] Although there are other 28 possibilities for the binding of the four bases out of the Watson-Crick base pairing with at least two hydrogen bonds,[59] the binding force of double hydrogen bonds is relatively weak.[50] Therefore, much effort has been invested into developing multiple hydrogen bond arrays with higher stability and orthogonality by modifying existing bases or developing new heterocyclic subunits mimicking the pairing of natural bases.[60–69] As non-exhaustive example, Blight *et al*. reported a cationic AAAA–DDDD complex with a binding constant up to 10^{12} M^{-1} in dichloromethane.[70] The multiple-hydrogen-bonding arrays commonly used for self-assembly, especially for supramolecular chemistry, include triple, quadruple and sextuple hydrogen bonds; some typical examples are shown in Table 2.2. The binding constant of these multiple-hydrogen-bonding arrays is related to the number of H-bonds and secondary electrostatic interactions between adjacent hydrogen bonds;[71] hence, there exists various hydrogen-bonding motifs with appropriate binding strength allowing to design a self-assembled structure according to the needs.

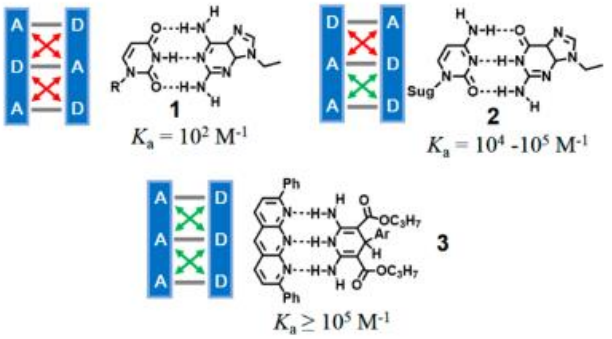
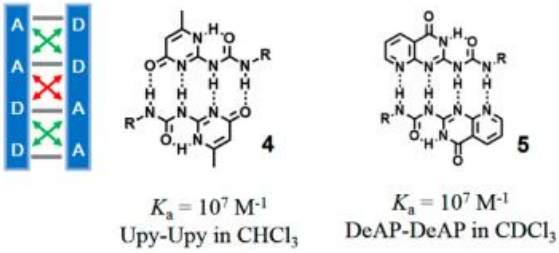
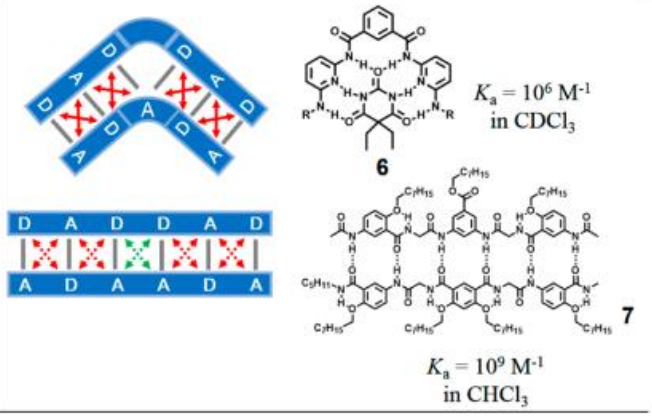


Scheme 2.1 Chemical structure of DNA base pairs

As written before, H-bonds have many interesting properties, such as being reversible, directional and strength-tunable, making them ideal for creating dynamic, functional, and responsive materials. Self-

assembly based on H-bonding is thus a versatile and powerful approach that has already been widely used in fields ranging from nanotechnology to biomedicine.[72–77]

Table 2.2 Typical examples of multiple-hydrogen-bonding arrays and their binding constants. Adapted from[54]

Type	Example
Triple	
Quadruple	
Sextuple	

2.2 Precision macromolecules

Natural molecules like DNA and proteins typically exhibit precise molecular lengths, perfect sequence order and topological structures, thereby acquiring specific biological functions.[78,79] Such precision and regularity of natural molecules is essentially dependent on their primary structure.[80] Inspired by these natural polymers, sequence-defined polymers (SDPs) were proposed to achieve similar precision. Compared with natural macromolecules, synthetic SDPs have more diverse backbones and side groups.[81] By controlling the primary structure of SDPs, the intramolecular conformations and intermolecular interactions could potentially be further tuned to achieve complex and predictable functional properties,[81,82] such as data storage, catalysis, etc. However, to get these properties, we need to be able to predict function from primary structure, which requires understanding higher order structures, which remains very difficult.

2.2.1 Effects of sequence and chirality on molecules

Sequence definition and stereochemistry are critical for precision macromolecules. For biomolecules as well as for synthetic polymers and small molecules, their functions, such as enzymatic activity, binding affinity, or mechanical strength, are intimately tied to their precise three-dimensional shape. Thus, controlling sequence and chirality is essential for designing molecules with tailored properties and functionalities for various applications.

Indeed, the sequence order of the monomer units (or primary structure) directly decides the physicochemical properties of the macromolecule, and influences the folding of the molecules into specific secondary and tertiary structures, ultimately determining its

three-dimensional conformation and controlling functionality.[83] For example, a previous work in our group by Li *et al*. focused on the preparation of a series of sequence-defined catalysts based on TEMPO, imidazole, and pyridine-triazole (pyta)-Cu as catalytic units (but of non-controlled chirality). Among these, the catalytic efficiency of two catalysts having their two active sites switched in position, differed by a factor of up to ten for the aerobic oxidation of benzyl alcohol (BnOH) at a catalyst concentration of 5 mol%.[84]

Chirality can also significantly affect the shape and conformation of the molecule by governing the spatial arrangement of atoms within the molecule, resulting in enantiomers or diastereomers that can exhibit vastly different biological activities, chemical reactivities, and physical properties.[83,85] For instance, Clover *et al*.[86] prepared homochiral counterparts LL(FKFEFKFE), DD(fkfefkfe) and heterochiral counterparts LD(FKFEfkfe) and DL(fkfeFKFE) by replacing one or two L-FKFE repeats in the amphiphilic peptide KFE8 (of sequence Ac-FKFEFKFE-NH_2) by D-amino acids ("fkfe") of opposite chirality. Although the homochiral counterparts formed easily nanofibers, the heterochiral counterparts assembled into larger helical tapes of large dimension and pitch. Through energy minimization models, the authors speculated that the difference in behavior originated from the appearance of kinks at the interface between chirally-different blocks, and from the introduction of internal strains which offset the natural distortion of the β-sheet structure and flatten it out.

2.2.2 Synthesis methodologies

As was covered in the previous section, the synthesis of monodisperse sequence-defined and stereo-controlled polymers is

crucial to control their properties and to custom their structure and functionality. Therefore, developing methods to efficiently synthesize functional polymers while simultaneously controlling sequence and chirality should be of high priority. A timesaving strategy for synthesizing SDPs in a few steps is controlled polymerization, but the polymers and sequences obtained by this method are usually not monodisperse, and thus separation techniques like automated flash chromatography systems are necessary to purify the products.[79,82,83] Therefore, although the synthetic protocol is simple and scalable, controlled polymerization is usually only suitable for the synthesis of sequence-defined oligomers, because the purification step is more difficult for longer chains.[83] To synthesize SDPs with a control of chirality, living radical polymerization (LRP) might be used. For example, vinyl monomers can be used to synthesize syndiotactic or heterotactic polymers by adding different catalysts or tuning the activity of the catalyst.[87,88] In addition, stereo-regular and sequence-controlled polymers can also be prepared by stereoselective ring-opening polymerization of racemic monomers containing chiral side chains or ring-opening polymerization of chiral monomers.[89,90] The isolation of monodisperse stereo-controlled polymers can be realized by supercritical fluid chromatographic separation or column chromatography, but is usually not applicable for copolymers with different block lengths.[83]

Another more precise way to synthesize SDPs is more commonly used, namely multistep synthesis. As the name implies, multistep synthesis achieves precise positioning of functional groups by coupling monomers or oligomers step by step. Due to the multistep reaction, it is possible to synthesize SDPs with high sequence complexity, comprising various monomers (functional groups), with stereo-control, leading to diversified backbone structures.[82,83] On the

other hand, multi-step reactions will result in a low total yield of the final product due to the finite yield of each step and losses during purification processes performed between each step. Because the yield of each step strongly matters, highly efficient reactions like click chemistry, thioacetone chemistry, Michael addition, etc. are usually selected.[79,83] The strategies to synthesize SDPs by multistep synthesis can be broadly categorized into support-free liquid-phase synthesis and supported synthesis, including solid-supported and soluble-supported synthesis.[79,81] With these synthetic pathways, various structural levels including backbones, side groups, and stereochemistries can be controlled.[82]

Support-free liquid-phase synthesis typically involves the use of low-molecular-weight organic molecules as starting materials, using some efficient chemical reaction such as CuAAC, where monomers or oligomers are sequentially incorporated into the main chain. For each elongation step, the intermediate product needs purification, often through classical column chromatography. Therefore, this method is intricate and time-consuming. However, the products synthesized by this method have well-defined structures, facilitating relatively straightforward analyses. Moreover, this approach is applicable to large-scale synthesis. SDPs via support-free liquid-phase synthesis methods include iterative stepwise growth (ISG), iterative exponential growth (IEG), bidirectional growth or single-unit monomer insertions (SUMI).[79,82] As Figure 2.2 shows, the ISG method starts with a monofunctional molecule and is followed by the progressive coupling of individual monomers. Although more iterative steps are required for SDPs via ISG, the ability to independently control the monomer structure and function of each step allows for highly complicated sequences.[79,82] ISG is widely used in multicomponent reactions. For example, Meier's group reported a series of multi-component

sequence-defined oligomers using Passerini three-component reactions (P-3CR) and Ugi four-component reactions (Ugi-4CR); the strategy is free of protecting groups or activating agents, and is efficient and scalable.[91,92] The IEG strategy, for its part, is usually based on monomers protected orthogonally which, after deprotection and activation, are coupled to elongate the main chain by a factor of two. Compared with ISG, IEG needs less iterative steps for a similar chain length, but the prepared sequences have higher repetitiveness. Besides, group protection is necessary during the synthesis. Jamison and Johnson *et al*. proposed a semi-automated method for the synthesis by iterative exponential growth by combining continuous flow chemistry with IEG.[93] They employed an alkyl bromide and a triisopropylsilyl (TIPS)-protected alkyne to accomplish CuAAC coupling of monomers or oligomers via azide substitution and excess fluoride-mediated desilylation. Additionally, they reported an iterative exponential growth plus side-chain functionalization (IEG+) strategy with *tert*-butyldimethylsilyl (TBS) protected chiral monomers synthesized by enantiomerically pure epichlorohydrin and propargyl alcohol.[94] Similar to the previous method, the monomer was subjected to azide substitution/desilylation followed by a click reaction to give sequence-defined and chiral-controlled macromolecules. Meanwhile, this IEG+ allows the introduction of different functional groups via acetyl (OAc) and benzyl (OBn) groups. However, Zhang *et al*.[95] have proposed a novel IEG strategy without protecting groups. This approach employs three monomers with orthogonal reactive end groups, including aldehyde-X-alkyne, azide-X-fluorosulfate, and *tert*-butyldimethylsilyl ether-X-aniline. Through the parallel orthogonal click reactions of CuAAC, sulfur-fluoride exchange reaction (SuFEx), and Ugi-4CR, three SDPs with different repetitive main-chain sequences were synthesized. For bidirectional growth synthesis, the

starter is a bifunctionalized molecule, and the main chain is elongated bidirectionally by the stepwise addition of two equivalents of monomers. Similar to the IEG strategy, bidirectional growth can achieve rapid chain growth with fewer steps compared to ISG. However, the SDPs synthesized by bidirectional growth are symmetric. The common strategy includes the use of oligo(ethylene glycol)s (OEGs) coupling with monoprotected and mesylated or tosylated OEGs, followed by deprotection and iteration;[96–98] or photoligation, such as the bidirectional stepwise addition with a molecule containing two maleimides at the end as a core.[99–101] Zhao *et al*.[102] proposed a novel bidirectional synthetic strategy. They used N,N,N',N'-tetramethylethylenediamine as the starting unit, which reacted with propargyl bromide to give an oligomer with azide groups at both ends; this oligomer was then coupled with functional monomers with an azide group and a N,N-dimethylamine group at the chain ends. Through alternating Menschutkin reaction and CuAAC, a series of sequence-defined multifunctional polymers with positively charged main chains were obtained through precipitation in weak polar solvents, without any protecting groups and solid support.

SUMI is typically based on a free radical addition mechanism, in which individual vinyl monomers are sequentially incorporated into the main chain by reducing the reactivity of chain ends. For example, Huang *et al*.[103] reported a SUMI methodology to synthesize sequence-defined and stereo-controlled oligomers by sequential and alternating insertion of stereospecific indene and N-substituted maleimide derivatives. Recently, a novel SUMI based on a cationic mechanism (cSUMI) by Tao *et al*.[104] was introduced, enabling the synthesis of vinyl ethers and styrene oligomers that are partially inaccessible through conventional free radical SUMI methods.

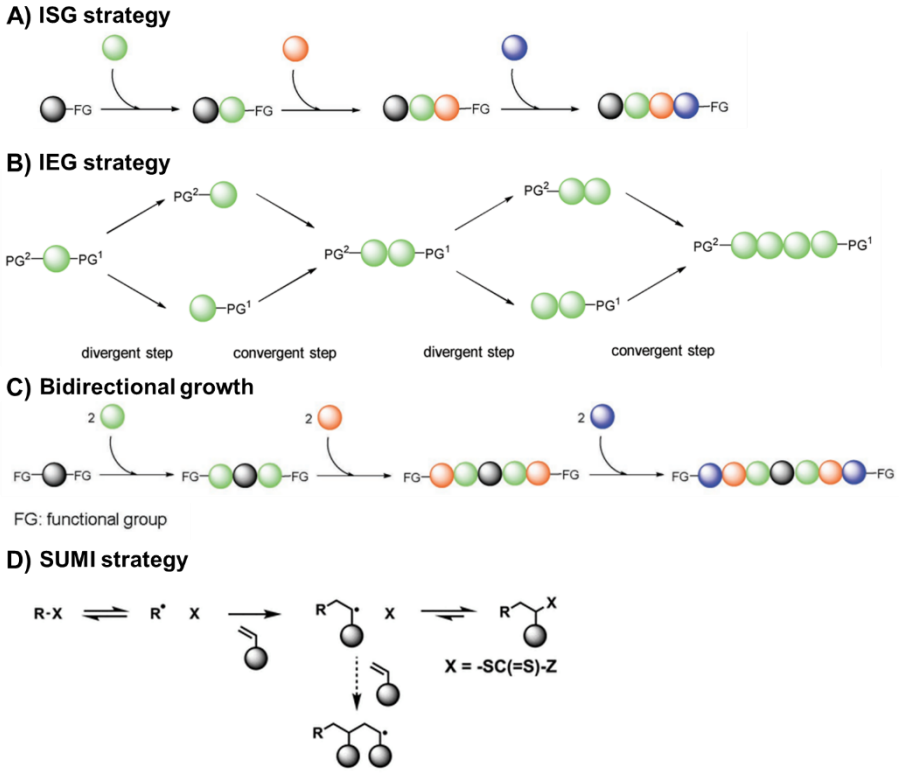


Figure 2.2 Synthesis of sequence-defined polymers by support-free liquid-phase synthesis. A~ C, adapted from[82]; D, adapted from[104]

Solid-supported synthesis (Figure 2.3) refers to the initial anchoring of the reactant onto a solid support, which is followed by stepwise coupling of the protected monomer units and deprotection, or coupling of subunits (monomers with reactive terminal groups or two species with different end groups amenable to orthogonal coupling reactions).[105] When the synthetic steps are complete, SDP's can be released from the support by cleavage. Due to the heterogeneous nature of the synthesis, side products and reagents can be easily removed by simple filtration or washing. However, due to the limitations in the loading capacity of the support or coupling efficiency, solid-supported synthesis faces challenges in preparing high molar

mass SDPs, and the mass scale of the product is relatively small.[82,105] Thus, a soluble-supported synthesis strategy was proposed by combining the advantages of liquid-phase synthesis, for instance, low reagent dosage, characterizability of intermediates, scalability, as well as simple post-treatment with the benefits of solid-supported synthesis. This approach utilizes a soluble carrier for the stepwise coupling of monomeric units in a homogeneous system. Commonly used soluble carriers include soluble fluorous tags and polymers.[82] However, because the system is homogeneous, the separation of the resulting SDP and carrier is challenging after cleavage.[82]

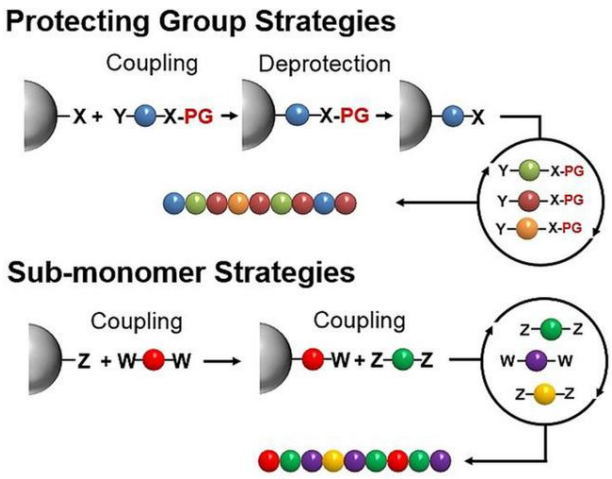


Figure 2.3 Synthesis of sequence-defined polymers by solid-supported synthesis. Adapted from[105]

There are also other methods to synthesize SDPs. For instance, template synthesis, especially DNA-templated synthesis (DTS), in which the template is pre-designed to interact with the introduced monomers, leading to the same degree of ordered sequence in the synthetic polymer.[79,106] However, DTS is only suitable for systems compatible with DNA.[79]

2.2.3 Characterization of sequenced defined polymers

Characterization techniques are crucial for exploring molecular structure and physicochemical properties. Methods for characterizing SDPs encompass mass spectroscopy (MS), nuclear magnetic resonance (NMR), nanopore sequencing, simulation, etc. These techniques provide diverse information about the structure, sequence order, morphology, and other aspects of SDPs. Detailed presentations of these techniques can be found in some reviews;[78,107] therefore, only a brief introduction will be given to those commonly used in this work.

NMR is one of the most used and valuable techniques for structural characterization. It relies on interactions between atomic nuclei with non-zero magnetic moments, such as those with odd numbers of protons or neutrons, like the common ^{1}H, ^{13}C, and ^{19}F nuclei, and an applied magnetic field; these nuclei absorb energy corresponding to the energy difference between two levels in the presence of a strong magnetic field, leading to the generation of a resonance spectrum. Chemical shifts, scalar coupling, and spectral intensities in the resonance spectrum can be used to identify chemical groups and their bonding environments, thereby determining the molecular structures.[108,109] Additionally, in this thesis, diffusion ordered spectroscopy (DOSY) has been used to analyze self-assembly behavior. DOSY uses pulsed gradient field NMR to virtually separate components with different diffusion coefficients (D), which depend on molecule properties like size and shape, thereby providing information such as molar mass or the formation of assemblies.

Tandem mass spectrometry (MS/MS) is another commonly used technique for sequencing SDPs. Ions with specific mass-to-charge ratios (m/z) are fragmented after the ionization of the sample

molecule, and the resulting fragments are detected by the mass spectrometer to generate MS/MS spectra. By analyzing the mass differences between these fragments, sequence information for the polymer can be inferred.[78,108] Additionally, the combination of standard MS analysis with other techniques can also be utilized for sequencing. For instance, mass spectrometry can be coupled with chromatography, or backbone cleavage before standard MS detection.[78]

Computational modeling and theoretical simulations are also used in SDP research. They can predict the differences between various SDPs to optimize experimental design. Simultaneously, modeling provides detailed insights into the relationships between structure, property and reactivity.[78,110] Additionally, for self-assembled systems, thermodynamic and dynamic information is challenging to directly infer from experimental measurements; however, theoretical simulations can offer these details.[111]

Apart from structure and sequence order, the chiral purity is also an important parameter for stereo-controlled SDPs, particularly the enantiomeric excess (ee). There are many techniques used for chiral detection, including chiral chromatography, such as chiral high-performance liquid chromatography (chiral HPLC) and supercritical fluid chromatography (SFC), spectroscopic methods such as circular dichroism (CD) spectroscopy and NMR with chiral solvents or shift reagents, electrochemical methods like electrochemical sensors, and so on.[112–115] Among these, chiral HPLC is one of the most widely used technology. Typically, chiral HPLC makes use of a chiral stationary phase (CSP) or chiral mobile phase additive that interacts differently with the enantiomers due to their chirality; these differences in binding affinity allow the enantiomers to separate and

elute from the column at different rates; a detector is used to identify and quantify the enantiomers based on their retention times.[116,117] Chiral HPLC is widely used for chiral detection due to its high speed, sensitivity, and reproducibility.[114]

2.2.4 Applications of sequence-defined polymers

SDPs have diverse backbones and complex functional groups ideally with control chirality, providing a broad scope for customization, which makes them useful for various applications in biomedicine, nanotechnology, data storage, catalysis, etc. Since SDPs have a wide range of applications and some reviews[79,83,107,118–123] have already well summarized these applications, only a few examples will be discussed here.

SPDs provide a platform to create highly ordered structures for complex functions like drug delivery and nanoreactors by precisely controlling the interactions within and between SDPs by tuning the backbone, side groups, stereochemistry, etc. For example, Kwon *et al*.[124] reported a crystallization-driven self-assembly of block copolymers composed of monodisperse poly(lactic acid) (MPLA) blocks with enantiomeric sequences and poly(ethylene glycol) (PEG). The morphology of the self-assembled nanostructures varied from nanoparticles to 2D triangles and vesicles, depending on the number and sequence of enantiomeric lactic acid units.

Despite containing only four nucleotides, DNA is a natural memory device that can store massive amounts of information about the entire genome due to its various nucleotide sequences and chain lengths. Compared with DNA, SDPs have more abundant monomers and better stability, which enables them to store information at the molecular level.[125] For example, Du Prez *et al*.[126] reported a strategy

of chemical data storage by multifunctional SDPs synthesized with an adapted peptide synthesizer. Fifteen acrylate monomers allowed to generate different sequence-defined structures, which could be converted into a machine-readable QR code with a data capacity of 1089 bits per a 33 × 33 QR code.

Self-assembly of SPDs mimicking enzyme structure is another example of application. Enzymes can catalyze different reactions with high specificity thanks to their stereo-regularity and amino acid sequence, which allow the precise folding of proteins and the creation of highly specific catalytic pockets.[83] Similar to enzymes, polymers can theoretically fold into multiple shapes through intramolecular interactions to form catalytic pockets suitable for specific substrates. However, only a fraction of macromolecules can form the expected folding due to the polydispersity of polymers.[79,107,127] Compared with traditional polymers, SDPs can be tailored to precisely control the position of catalytic units and control the folding of SDPs, thereby forming more efficient enzyme mimics.[79,107] In our group, the work of Chandra *et al*. about the aerobic oxidation of alcohols with the solid-supported precision oligomers catalysts containing the synergistic catalytic units of TEMPO, NMI and pyridyl triazole (pyta) showed that the restricted conformational space in grafted layers was beneficial to enhance the interaction within oligomer chains,[128,129] and that the sequence of the catalytic units was very important for catalytic activity.[130] When the NMI was close to pyta, the catalytic activity was the highest.[130] Another work in our group by Li *et al*.[84] already mentioned in section 2.2.1, reached a similar conclusion for homogeneous catalytic systems based on precise oligomers having the same catalytic units. Stereocontrol is also important in catalysis, especially for enantioselective transformations, because the relative orientation of the catalyst and substrate may largely control the

catalytic process.[118] Maayan *et al*.[131] prepared a series of a phenylethylamine peptoid catalysts with a TEMPO catalytic center and investigated the role of the nano-environment of the catalyst on the oxidative kinetic resolution of racemic mixtures of chiral secondary alcohols when tuning the sequential position of TEMPO and the chirality and helicity of the peptoid. They found that the catalyst activity is sequence-dependent, with the catalyst having the highest catalytic activity when TEMPO was at the end of the chain. The chirality of the asymmetric environment generated by the helical backbone was more selective for enantiomers of matched chirality.

2.3 Aerobic oxidation of alcohol

The selective oxidation of alcohols is a crucial class of oxidation reactions in organic synthesis. The oxidation products, aldehydes or ketones, can serve as intermediates for the preparation of various organic compounds or industrial products; therefore, the alcohol oxidation reaction has drawn significant attention.[132,133] Initially, the oxidation of alcohols was performed with inorganic oxidants such as hexavalent chromium and manganese dioxide. But those inorganic oxidants themselves are toxic. Besides, there are also other problems like harsh reaction conditions, a lot of side products and waste.[132,134] With the progress of research, the conditions for alcohol oxidation have been continually optimized, looking for milder and effective catalysts.

2.3.1 Enzyme mimics for the aerobic oxidation of alcohols

Enzymes are effective natural catalysts; the enzyme residues with or without cofactors create cooperative catalytic sites within a three-dimensional cleft formed by assembly and folding. The cavity is

designed for substrate binding and transition-state stabilization, resulting in accelerated reactions and quickly adapting to changes.[6,135] Therefore, enzymes are widely investigated and mimicked.

Inspired by enzyme catalytic sites, the construction of dynamic active sites within catalytic pockets by self-assembly or folding has been investigated, as shown in Figure 2.4. For example, peptides are promising candidates to produce self-assembled supramolecular catalysts due to their strong biocompatibility and programmability.[3,5] Li *et al*.[3] used the co-assembly of peptides and Cu^{2+} ions to produce nanofibers mimicking peptidyl catecholase. The nanofibers have switchable catalytic activity through reversible covalent modification with kinases/phosphatases and high selectivity. Similarly, DNA molecules are used as molecular scaffolds to construct dynamic active sites thanks to their properties such as programmability and ease of modification. Pimentel *et al.*[11] prepared a DNA scaffold catalyst by anchoring the catalyst TEMPO and co-catalyst Cu at the two close ends of a double-stranded DNA chain, which maintained a high turnover number after dilution; in addition, the introduction of competing single-stranded DNA (ssDNA) could reversibly tune the DNA assembly, thereby controlling the formation and closure of cooperative catalytic sites and further controlling the catalytic reaction rate.

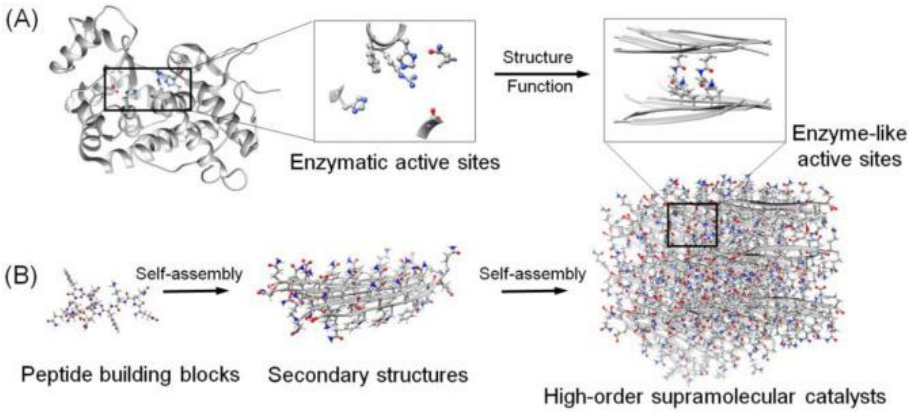


Figure 2.4 Design of peptide-based self-assembled supramolecular catalysts with built-in enzyme-like active sites. (A) Formation of enzyme catalytic sites and realization of catalytic function. (B) Construction of supramolecular catalysts based on peptide self-assembly. Adapted from [6]

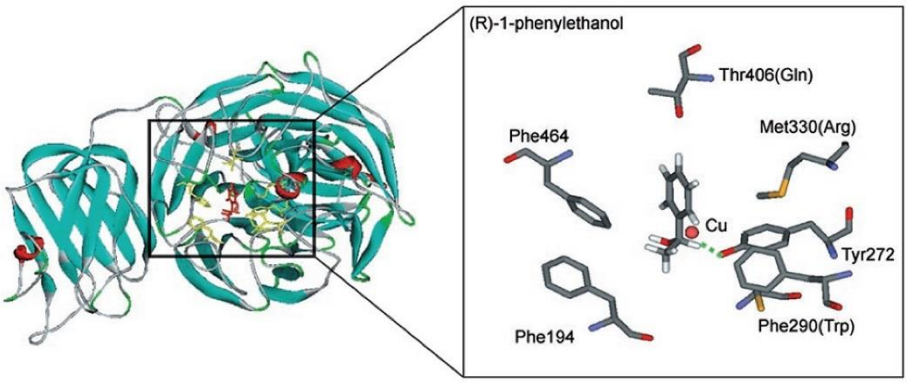
Without doubt, the utility of natural molecules such as DNA and peptides to construct dynamic catalytic active sites is interesting, but the application of natural molecules is still limited due to cost and environmental constraints. To overcome the above limitations, an alternative approach is to modify these natural molecules. For example, Prathap *et al*. selected a peptoid having catalytic units (ligand phenanthroline and TEMPO) as side groups; thereby, they developed a highly active Cu-TEMPO-based catalyst that can achieve a low catalyst loading (0.1 mol%) and a high turnover number in the catalytic oxidation of primary alcohols by tuning the spacing and order of the catalytic groups.[27]

In order to explore more possible catalytic conditions, scientists have started to make use of synthetic macromolecular architectures to construct active catalytic sites. Metal-organic frameworks (MOFs) are one of the structures inspired by enzyme catalysis, in which a confined environment similar to enzyme is provided by an ordered

arrangement of metal active centers and multichelating organic ligands in tunable cavities or channels.[28,136] Taher *et al*.[137] developed an efficient MOF catalytic system with $Cu^{(II)}$ by anchoring copper complexes through amine-functionalized MOF layers which, together with the co-catalyst TEMPO, successfully achieved aerobic oxidation of various alcohols even containing inactive heteroaryl and long-chain alkyl units. Supramolecular catalysts were also developed to create a stable microenvironment for the active site. Jiao *et al*.[138] fabricated a host-guest complexed supramolecular catalyst incorporating TEMPO and cucurbit[7]uril (CB[7]), which together with an oxidant sodium hypochlorite (NaClO) was used in the biphasic oxidation of alcohols. During the catalysis process, the electrostatic interaction between the carbonyl portal of CB[7] and $TEMPO^+$ favors the stabilization of $TEMPO^+$ in the aqueous phase, helping the oxidation of TEMPO by NaClO into $TEMPO^+$ and inhibiting side reactions. When migrating into the organic phase, TEMPO+ was separated from CB[7] due to the insolubility of CB[7] in the organic solvent, and its high activity is recovered to drive the fast oxidation of substrate. The adaptive reactivity of the intermediate $TEMPO^+$ significantly improved the turnover rate for various alcohols. There are also other reported polymeric architectures that can self-assemble or fold to create a catalytic pocket that resembles an enzyme, such as single-chain polymeric nanoparticles (SCNPs),[18–25] foldamer,[32,33] helical (supramolecular) polymers, etc.[34–37] No matter what kind of macromolecular architecture is selected, the choice of polymeric backbone, the selection of side chains, and the design of self-assembly and folding are all important to the construction of the active site.

In addition to structural mimicry, there are also various studies based on the mimicry of enzyme functions; one of the most famous

examples is galactosidase, whose active site is highly dependent of the catalytic synergy between copper and the phenoxyl radical of the cysteine-modified Tyrosine 272 residue,[139] as shown in Scheme 2.2. This synergistic catalytic mechanism between metals and radicals has been the inspiration for numerous systems for the aerobic oxidation of alcohols. Cu/TEMPO is one of the most well-known catalytic systems. This system can oxidize a variety of substrates, but it usually works better for the oxidation of primary alcohols than secondary alcohols.[140–142] Normally, the catalytic activity of the system can be properly tuned by the source of copper, ligands, solvent and base.[143] The Fe/TEMPO system is another metal-radical synergistic catalytic system, which is usually used together with nitrate or nitrite additives, but without base or organic ligand which are not needed.[142,144] Compared with Cu/TEMPO, the Fe/TEMPO system oxidizes more readily secondary alcohols, which may be because the nitrate or nitrite additives can oxidize TEMPO to oxyammonium salts; there would be no need for the ligand, and the system would thus have less steric hindrance.[142,145,146] There also other transition metal/TEMPO systems which have been reported, such as cobalt or manganese and their co-catalysts with TEMPO,[147–152] ruthenium/TEMPO,[153–155] etc. Although the choice of metal types is diverse, copper and iron are among the most widely used metals considering the catalytic effect, source, cost, safety and other aspects of the metal.[142]



Scheme 2.1 Active site of Galactose oxidase. Adapted from[139]

Apart from the type of metal, free radicals are also an important part of biomimetic catalysis. Among these, the most widely used are the nitroxyl radicals due to their low sensitivity to air or water.[156] The nitroxyl radicals are a free radicals containing a N,N-disubstituted -NO group, which bears one unpaired electron. The nitroxyl radicals can be divided into two categories: stable free radicals and reactive free radicals, as shown in Figure 2.5.[157] Stable free radicals include conjugated radicals like diphenyl-nitroxyl radicals or non-conjugated radicals, such as di-tert-butylnitroxyl (DTBN) and TEMPO. For the reactive nitroxyl radicals, the phthalimide N-oxyl radical (PINO) obtained from N-hydroxyphthalimide (NHPI) is the most well-known example. Although PINO·is more reactive, the relative stability of stable radicals makes them more widely used.[157]

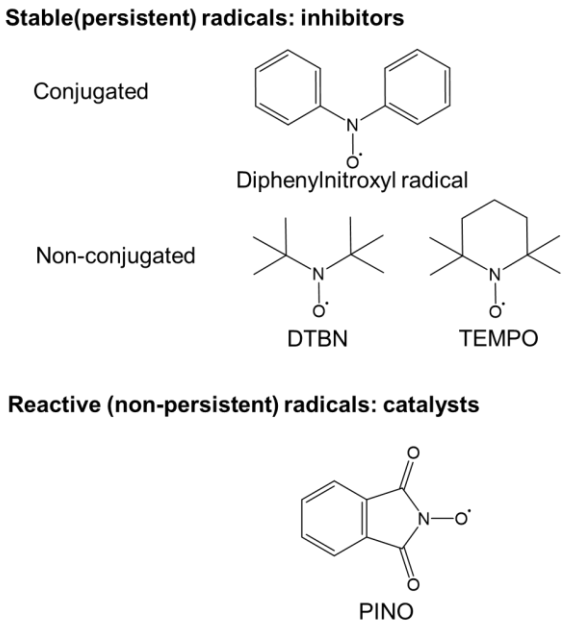


Figure 2.5 Typical examples of organic nitroxyl radical. Adapted from[157]

2.3.2 Cu(I)/TEMPO-catalyzed aerobic alcohol oxidation

As mentioned before, galactose oxidase has already inspired the free radical-coupled copper catalytic oxidation,[158] especially, the Cu/TEMPO catalytic system has attracted much attention. On the one hand, copper has wide availability, low cost and already exists in the active sites of many enzymes.[132] On the other hand, TEMPO· is a stable radical, that can act as an oxidant itself capable of oxidizing various functional groups.[156] Notably, TEMPO· exhibits high selectivity in the oxidation of alcohols, typically converting alcohols solely into the corresponding aldehydes or ketones, and it has high coordination ability, allowing synergistic catalysis.[159]

The Cu/TEMPO catalytic system was first reported in 1984 by Semmelhack *et al*.[160] They used $Cu^{(I)}Cl$/TEMPO to oxidize benzyl

alcohol and propargyl alcohol into the corresponding aldehydes in N,N-dimethylformamide (DMF) solution under oxygen conditions at room temperature (RT). Subsequently, Sheldon *et al*.[161] developed a catalytic system utilizing $CuBr_2$/TEMPO/KOtBu equipped with bipyridine (bpy) as a ligand. This system was employed in a mixed solvent of acetonitrile (ACN) and H_2O (2:1), and it can efficiently oxidize various primary alcohols to aldehydes. In comparison to previously reported systems, their catalytic system introduced a base, which can facilitate the deprotonation of alcohol substrates to provide alcoholates capable of coordinating with the $Cu^{(II)}$ center.[156] The ligand is necessary for their system, because the ligand contributes to enhanced solubility of Cu and provides beneficial electronic effects.[161] Based on Sheldon's catalytic system, Koskinen *et al*.[162] investigated the catalytic conditions. To enhance the oxidation of hydrophobic substrate alcohols, they replaced the catalytic solvent with ACN. Additionally, they evaluated the type and amount of base, observing that the base can significantly impact reactivity. Later, Stahl *et al*. found that $Cu^{(I)}$ salts exhibit higher reaction rates than $Cu^{(II)}$, and they designed a (bpy)$Cu^{(I)}$/TEMPO/N-methylimidazole (NMI) catalytic system, demonstrating effective oxidation of various primary alcohols.[133,163] This ternary system has been applied in many cases.[84,128,130,164–169,174] It is not only applicable to homogeneous catalysis but also exhibits high efficiency in heterogeneous catalysis.[128,164] Furthermore, it has evolved from an initial discrete cooperative catalysis system into a multifunctional catalytic system.[84,130]

2.3.2 Mechanism

In the past decades, the oxidation mechanism of Cu/TEMPO has

been discussed. It was described by different mechanistic pathways:[156] oxoammonium,[160] $Cu^{(II)}$-TEMPO-H,[161] $Cu^{(II)}$-binuclear[163] and $Cu^{(II)}$-monomer.[170] Here the $Cu^{(II)}$-binuclear mechanism is presented to introduce the (bpy)$Cu^{(I)}$/TEMPO/NMI ternary system (Figure 2.6), because our results are compatible with this mechanism. Firstly, $Cu^{(I)}$ would form a coordination compound with bpy and NMI, namely (bpy)(NMI)$Cu^{(I)}$. This coordination compound reacts with O_2 to generate a $Cu^{(II)}$ superoxide ((bpy)(NMI)$Cu^{(II)}O_2{\cdot}^+$). Subsequently, (bpy)(NMI)$Cu^{(II)}O_2{\cdot}^+$ reacts with a second (bpy)(NMI)$Cu^{(I)}$ to form a peroxo-bridged binuclear $Cu^{(II)}$ dimer. Next, TEMPOH is oxidized by the binuclear copper complex to give TEMPO·, and the hydrogen atom is transferred to form a Cu^{II} peroxo-hydroxide intermediate ($Cu^{(II)}$-OOH). This process is accompanied by the generation of $Cu^{(I)}$ as a byproduct. The intermediate $Cu^{(II)}$-OOH reacts with water (or alcohol substrate) to produce $Cu^{(II)}$-OH (or $Cu^{(II)}$-OCH_2R) species and H_2O_2, which will rapidly disproportionate. After that, the $Cu^{(II)}$-OH intermediate oxidizes the alcohol substrate, which creates a $Cu^{(II)}$-alkoxide through a step called equilibration. Upon removal of the hydrogen atom by TEMPO·, $Cu^{(II)}$-OH yields an aldehyde and TEMPOH.

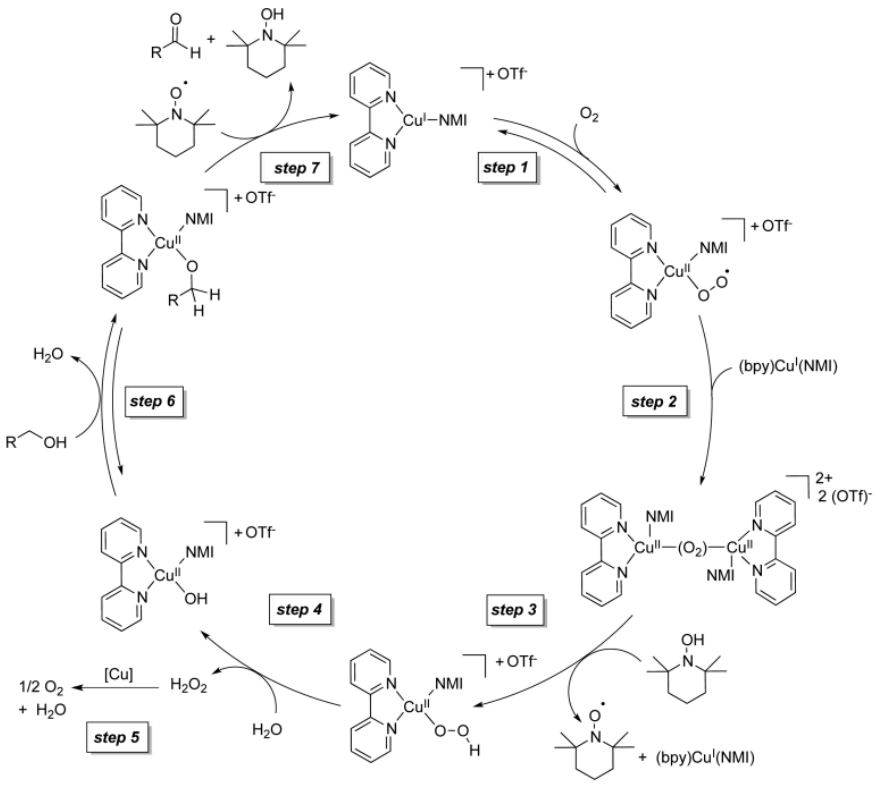


Figure 2.6 The proposed catalytic cycle for the (bpy)Cu^{I}/TEMPO·/NMI catalyst system. Adapted from[163]

Chapter 3 Methodology to synthesize stereo-controlled and sequence-defined oligomers

Abstract: In this chapter, we focus on the synthesis of stereo-controlled and sequence-defined oligomers. We have investigated ways to synthesize chains with a stable backbone, functionalized with a controlled sequence of functional units. These functional units include base pairs for self-assembly and catalytic units for aerobic alcohol oxidation. To control their sequence and chirality, a chiral precursor is used; the chirality of the oligomers is determined by this precursor. Then, the monomers are functionalized with an alkyne group for incorporation in oligomer backbones to give chiral sequence-defined oligomers. The reactions involved in this synthesis procedure include opening epoxides with azides, fluoride-mediated desilylation of alkynes, and CuAAC. The general information on the synthetic procedure is in Section 3.2. Before we start to synthesize the oligomers, we prepared the functional units (Section 3.3). In Section 3.4, the final oligomers we used are described and their chemical structures and sequences are confirmed by NMR and TOF-MS/MS, respectively. In section 3.5, we finally use this methodology to prepare a series of oligomers substituted by nuclei acid bases. Thanks to these highly efficient and orthogonal reactions, we produced stereo-controlled and sequence-defined oligomers with excellent yields together with good atom economy.

3.1 Introduction

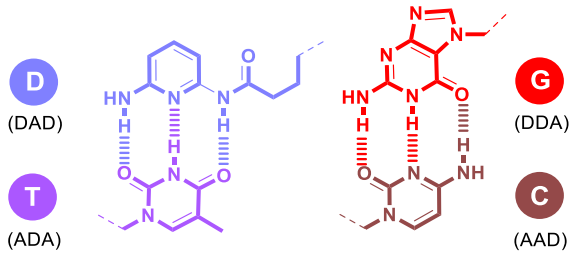
A The chirality and primary sequence of proteins have a significant impact on their structure, reactivity, and functionality; they affect everything from secondary, tertiary and even quaternary structures to folding and assembly, which in turn control their function and chemical interactions. Therefore, when synthesizing defined polymers, it is important to not only control the sequence order but also the chirality of the molecules to possibly mimic the assembly of natural macromolecules and to replicate some of their functions such as catalysis, as we want to do in this thesis.

There are several methodologies to synthesize SDPs as discussed in Section 2.2.2. Among these, an iterative exponential growth plus side-chain functionalization (IEG+) route developed by Johnson *et al*.[94] works well for controlling the sequence and stereochemistry of a macromolecular chain, which should be important parameters for assembly and catalysis. However, in the original scheme, an ester function is created, which does not resist some deprotection steps. As already discussed in previous work by our group, ester groups are not stable under hydrolytic conditions and in the presence of TBAF (tetrabutylammonium fluoride), which are used in deprotection steps.[128,130,171] Therefore, in our lab, an improved route has been developed, in which the ester is replaced by a more stable urethane (carbamate) bond, and the sequence of chain elongation and side chain installation are reversed compared to the original work of Johnson et al. This development, presented in our published articles, allowed us also to extend the range of functionalities of the produced polymers.[84,171] Additionally, the carbamate may contribute to the formation of H-bonded secondary structures in our precision macromolecules, as peptide bonds in proteins.[172] These H-bond

interactions may actually compete with the base pairing which we plan to use to favor the assembly of different chains in a catalytically-active center (see below); however, since base pairing is more stable and specific due to existence of an array of hydrogen bonds, we hypothesized that the addition of carbamate groups in the chain backbones would not be an issue. The reactions involved in this synthesis procedure include opening epoxides with azides, fluoride-mediated desilylation of alkynes, and CuAAC. Thanks to these highly efficient and orthogonal reactions, we produced macromolecules with excellent yields together with good atom economy.[173]

In this work, we used this modified IEG+ strategy to synthesize both stereo-controlled and sequence-defined chains, which can self-assemble and form a multifunctional catalysis structure for the aerobic oxidation of alcohols. For this, as mentioned above, we chose hydrogen bond arrays (base pairs) to control the assembly of the strands, because hydrogen bonds exhibit remarkable dynamics, directionality, and selectivity.[51] By playing with the number, direction, and external environment, such as temperature and solvent, of the hydrogen bonds, molecular assembly can be tuned.[111,174] However, single hydrogen bonds are usually weak; therefore, hydrogen bonds should preferably be used in the form of hydrogen arrays. For example, in DNA, hydrogen bonds exist in base pairs. Using DNA as a source of inspiration, we thus selected two base pairs to be incorporated in the oligomer sequences for recognition and assembly (Scheme 3.3.2). Within the constraints of the Watson-Crick rules, the number of possible nucleobases is only twelve;[175] however, these rules can be relaxed to extend the pairing possibilities. According to this, a 2,6-diaminopyridine derivative (D) was selected instead of adenine (A) to form the second base pair with thymine (T) (Scheme 3.1). Indeed, in contrast with the A/T pair which leads to divalent

hydrogen bonds, the D/A pairing results in the formation of trivalent hydrogen bonds (ADA-DAD) that should reinforce the specificity of the recognition process between the two oligomers.[176] The other base pair is guanine (G)-cytosine (C) (ADD-DAA). We expect that these base pairs will form H-bonds between complementary bases (scheme 3.1) and therefore control the assembly of strands equipped with these complementary bases. To construct the catalytic site, we selected a group of catalytic units as our synergistic catalysis center, namely TEMPO (M), imidazole (I) and Cu-bipyridine (P), which cooperate to realize the oxidation of alcohols. These three catalytic units have been widely used for the aerobic oxidation of alcohols, aiming to mimic the galactose oxidase free radical-coupled copper catalytic oxidation.[156] Additionally, this catalytic system has been proved to be highly efficient for the selective oxidation of alcohols.[156] Finally, alkyls chains (C_6) were also introduced on the chains to help balance the hydrophobicity of the catalytic system and improve solubility.



Scheme 3.1 Chemical structure of selected base pairs.

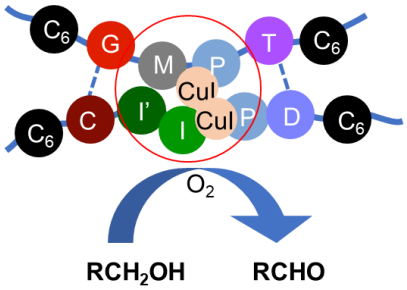
As mentioned before, the catalytic site will be formed by self-assembly of complementary chains, which together should bring all the needed catalytic groups in close spatial proximity. This requires to distribute all the needed functional moieties between the two chains.

First of all, the two base pairs need to be well placed in the two strands to allow the self-assembly of the complementary chains; we thus decided to place the bases at the two ends of the chains. For the catalytic center, it is critical for all catalytic units to be as close as possible to form a complete active site. Therefore, the catalytic groups were placed at the center of the chains. Then, the alkyl chains were finally placed at the very ends of the chains, in order to avoid disturbing the catalytic site and possibly contribute to reinforce chain end assembly.

The mechanism of $Cu^{(I)}$/TEMPO-catalyzed aerobic oxidation of alcohols proposed by Stahl *et al*.[163] involves a binuclear copper complex comprising two bidentate nitrogen ligands (such as bipyridine) and two auxiliary ligands (such as NMI). The catalytic cycle also requires a TEMPO moiety (the simplified catalytic cycle is shown in Figure 4.1.A). To confirm this mechanism, we designed three different catalytic systems made of two chains, a system including all the units needed to form a complete catalytic center including two P, two I and one TEMPO, and two systems with incomplete catalytic centers in which one P or one I is missing. The catalytic units were then split and distributed between the two oligomeric chains, making sure that no single chain would be capable to catalyze significantly the oxidation reaction. In order to reduce variables as well as have less synthetic work, we fixed the primary structure of one chain, whose sequence is $C_6G\underline{MP}TC_6$; the complementary chains were then built with only the catalytic groups in the center differing from system to system, leading to $C_6C\underline{I'IP}DC_6$, $C_6C\underline{IP}DC_6$ and $C_6C\underline{II}DC_6$ (the catalytic units are underlined) primary structures.

Having selected how to split between the chains all the functional

moieties and decided their relative position, we just needed to incorporate them step by step to form catalytic precision poly(triazole-carbamate) chains; as an example, the structure of the complete catalytic system is shown in Scheme 3.2.



Scheme 3.2 Structure of designed catalytic system.

3.2 General synthetic scheme of chiral precision oligomers

As said before, we selected a modified IEG+ strategy to synthesize both stereo- and sequence-controlled chains. During this synthesis, the control of chirality and sequence is carried out by using chiral key monomers. As shown in Scheme 3.3, (*R*)-3/(*S*)-3 (99.2% ee, details are available in Annex 3.1) are the chiral precursors, whose chirality is determined by the starting material epichlorohydrin. One can estimate the chirality of the oligomers containing *n* such repeat units (it was not possible to measure it directly). Suppose the enantiomeric excess of each monomer unit is e_1; if we denote by p_R and p_S=1-p_R the probability of *R* and *S* enantiomers (identical to their molar fractions) respectively, we can write:

$$e_1 = p_R - p_S = 2\,p_R - 1 \quad ,$$

leading to:

$$p_R = (1 + e_1)/2 \ .$$

The coupling of *n* monomer units in a single chain will lead to a collection of chains of different chirality. If we define the enantiomeric excess of the *R* enantiomer in this collection as:

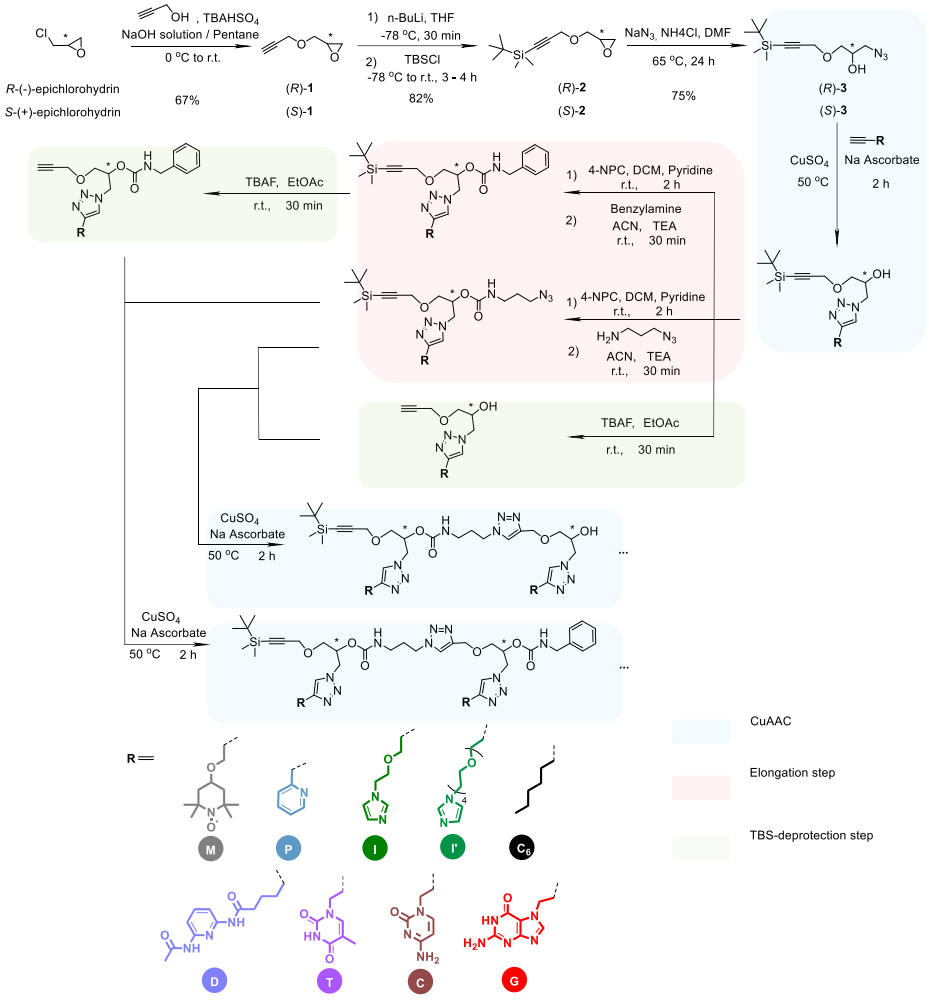
$$e_n = p_{(nR)} - p_{(\text{all other enantiomers})} \quad ,$$

we can write:

$$e_n = p_{(nR)} - (1- p_{(nR)}) = 2\, p_{(nR)} - 1 = 2\, (p_R)^n - 1 = 2^{1-n}\, (1 + e_1)^n - 1 \quad .$$

Therefore, with e_1 = 0.992, one finds e_4 = 0.97, e_6 = 0.95 and e_7 = 0.94 for the lengths of the oligomers which will be synthesized in this thesis. These values indicate highly chiral oligomers.

These chiral monomers are then orthogonally functionalized[177,178]/deprotected, followed by fluoride-mediated desilylation of alkynes and CuAAC of the intermediates. This process results in the synthesis of polymers with precise sequences, uniform masses, and defined stereo-configurations after elongation and deprotection cycles. Here, we wanted to adapt this route for the synthesis of a variety of precision oligomers capable to bind and generate a catalytic site based on TEMPO (M), imidazole (I) and Cu-bipyridine (P) as required for the aerobic oxidation of alcohols. This requires to synthesize alkyne-substituted functional units M, P, I and I' (Scheme 3.4), which are then linked on the monomers by CuAAC (Scheme 3.3). For assembly, we select complementary base units D/T and C/G (scheme 3.1); additionally, to favor solubility, we also introduced hexyl side chains C_6 in the SDPs. The general procedures used for the synthesis of SDPs are listed below.



Scheme 3.3 General synthetic route of the stereo-controlled and sequence-defined oligomers.

Synthesis of (*R*)/(*S*)-1, (*R*)/(*S*)-2 and (*R*)/(*S*)-3 was performed according to the literature done in our group[171,179] and our previous work,[180] and during my PhD;[84] the details of synthesis and characteristics are in ESI section 3.1, 4 and 5.1 of our co-authored article titled “Sequence Rules the Functional Connections and Efficiency of Catalytic Precision Oligomers”.[84]

The synthesis of monomers and other oligomers, as shown in blue in scheme 3.3, relies on the coupling of azide-ended molecules and alkyne-ended molecules in solution. This CuAAC reaction is highly efficient, and the solvent used is a mixture of ethanol (EtOH) and water, which is non-toxic and easily removed. Generally, azide oligomers (1 mmol or depending on the needs) and alkyl-end oligomers (1 equiv.) are added into a flask followed by adding EtOH (7 mL) with stirring under a nitrogen atmosphere at 50 °C. Sodium ascorbate solution (0.2 equiv.) and copper (II) sulfate ($CuSO_4$) solution (0.1 equiv.) dissolved in MiliQ-water (3 mL) are successively added to the reaction mixture. The mixture is stirred for 2 h followed by cooling to room temperature. Ethyl acetate (EtOAc) or dichloromethane (DCM) (30 mL) is added into the flask to extract the organic products, which is followed by pouring ethylenediaminetetraacetic acid disodium salt dihydrate solution (Na_2EDTA) (0.05 M, 30 mL) into the mixture to extract the copper ions dissolved in aqueous phase; this step is done while bubbling air for 30 min to fully oxidize Cu(I) into Cu(II) and quench the reaction. Then, the mixture is transferred into a separatory funnel, the organic phase is collected, followed by extraction (twice) with EtOAc, DCM or chloroform ($CHCl_3$) (depending on the solubility). The organic phases are combined before being dried with sodium sulfate anhydrous (Na_2SO_4) and concentrated under vacuum. The residue is purified by column chromatography to give the final product. The yield of CuAAC reactions during the synthesis varies from 57% to quantitative. Some click reactions have low yields because one of the amine protections, the *tert*-butoxycarbonyl protecting (BOC) group, is partially deprotected or the TEMPO radical is partially reduced during the reactions.

For the elongation steps, as underlined in red in scheme 3.3, the

hydroxyl group of the monomer or other oligomers is elongated with an azide end or a benzyl, which is realized by the activation of alcohol followed by an aminolysis reaction. During the elongation process, the hydroxyl group is first reacted with 4-nitrophenyl chloroformate (4-NPC) to give an activated carbonate intermediate, which is subsequently attacked by the amine to form a carbamate bond. 4-nitrophenyl group is a good leaving group that can be easily cleaved in mild basic conditions, and the hydrolysis product 4-nitrophenol can be removed by washing with water.[180] Therefore, these two steps reaction is usually with high yield.[181] The only significant issue is the toxicity of the chloroformate, which should preferably be banned in future work.

In this thesis, oligomers with a hydroxyl end group (1 mmol or depending on the needs) are dissolved in EtOAc or DCM (5 mL, depending on solubility) in a flask followed by adding 4-nitrophenyl chloroformate (1.5 monomer equiv. or 2 dimer and trimer equiv.) and pyridine (1.5 monomer equiv. or 2 dimer and trimer equiv.). The mixture is stirred under a nitrogen atmosphere at room temperature for several hours (1 h and 2-4h for monomer and dimer or trimer, respectively) followed by solvent evaporation. The residue is dissolved in ACN (5 mL) again, followed by adding benzylamine (2 monomer equiv. or 3 dimer and trimer equiv.) or 3-azido-1-propylamine (2 monomer equiv. or 3 dimer and trimer equiv.) according to the aims, and triethylamine (Et_3N) (3 monomer equiv. or 4 dimer and trimer equiv.). The mixture is stirred under a nitrogen atmosphere at room temperature for 1or 2 h followed by solvent evaporation. The residue is dissolved in EtOAc or DCM (50 mL) and washed with Mili-Q water (50 mL) three times to remove the 4-nitrophenol; it is washed twice with brine (50 mL) before being dried with Na_2SO_4 and concentrated under vacuum. The final product is

obtained after passing the residue through a chromatography column. The yield of most elongation reactions is or is close to quantitative. However, during some elongation process, one of the BOC groups was deprotected; as a result, the yield of some reactions decreased to 70-90%.

As underlined in green in Scheme 3.3, oligomers with *tert*-butyl(dimethyl)silyl (TBS) protection need to be deprotected to free the alkyne group for the following click reaction. A widely used method is to use fluoride-mediated desilylation, typically adding TBAF. Fluoride ions have a strong affinity for silicon atoms, allowing the desilylation reaction to be very efficient.[182] During the synthesis, TBS-protected molecules (1 mmol or depending on the needs) are dissolved in EtOAc or DCM (5 mL) followed by adding TBAF (1.5 monomer equiv., 3 dimer equiv., 3 or 5 trimer equiv.). The mixture is stirred for 30 min or 1 h at room temperature under a nitrogen atmosphere, then MeOH (5 mL) is added into the flask to quench the reaction before solvent evaporation. The residue is dissolved in DCM (50 mL) and washed thrice with brine (50 mL) to remove the side products, which is followed by drying with Na_2SO_4 and concentration under vacuum. The crude product is obtained without any further purification and can directly be used for the click reaction, since we checked that the following click reactions behave identically irrespective of the fact that the product is or not purified.

In addition to the above-mentioned main reactions, the protection and deprotection of amino groups and oxidation of chloropurine are also involved during the synthesis process. The amino groups are protected by a BOC group which is easily removed by trifluoroacetic acid (TFA) treatment (if not properly discarded, TFA may pose environmental problems, as it belongs to the category of forever

chemicals). The deprotection of amino groups and oxidation of chloropurine are normally performed once the desired sequences are obtained. For example, oligomers with double BOC-protection (1mmol or depending on the needs) are dissolved in DCM (3 mL) followed by the addition of TFA (1.5 mL) into the flask. The mixture is stirred for 2 h at room temperature under a nitrogen atmosphere before solvent evaporation. Then, the mixture is neutralized with saturated sodium bicarbonate ($NaHCO_3$) solution to fully free the amino groups and extracted thrice with DCM (50 mL). The organic phase is washed twice with brine (50 mL) before being dried with Na_2SO_4 and concentrated under vacuum. The final product is obtained after passing the residue through a chromatography column or precipitation. The yield of this reaction varies from 44 to 98%.

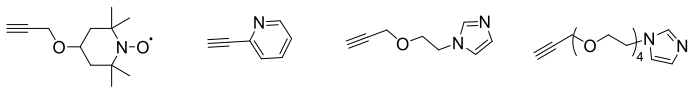
For the oxidation of chloropurine, TFA was also used. Oligomers containing chloropurine (1mmol or depending on the needs) are added in a flask followed by adding TFA (3 mL) and Milli-Q water (1mL). The mixture is stirred for 48 h at room temperature. Then, the mixture is neutralized with saturated $NaHCO_3$ solution to fully free the amino groups and extracted thrice with DCM (50 mL). The organic phase is washed twice with brine (50 mL) before being dried with Na_2SO_4 and concentrated under vacuum. The final product is obtained after passing the residue through a chromatography column or purified through precipitation. The yield of this reaction varies from 51 to 92%.

3.3 Synthesis of precursors

3.3.1 Synthesis of catalytic side groups

As mentioned before, the(bpy)Cu^{I}/TEMPO/NMI catalytic system is

suitable for the aerobic oxidation of alcohols. In our catalytic system, we thus also selected TEMPO, pyridine and imidazole as catalytic units. To incorporate the catalytic units in the oligomer backbone, we functionalized these catalytic units with an alkyne group; the chemical structure of these compounds is shown in Scheme 3.4, as reported in my co-authored article[84] and in previous work done in our group.[128,164,171,179] The detailed synthesis and characteristics are reported in ESI section 3.1, 4 and 5.1 of my co-authored article "Sequence Rules the Functional Connections and Efficiency of Catalytic Precision Oligomers"[84] and in section 3 of the article "Molecular Engineering of Trifunctional Supported Catalysts for the Aerobic Oxidation of Alcohols" published by our group.[164] Note that we explored two possible structures for the N-imidazole group (**6** and **7**), in order to permit the exploration of the different volumes of space by the imidazole unit, which needs to complex with copper as shown in Figure 2.5.1.

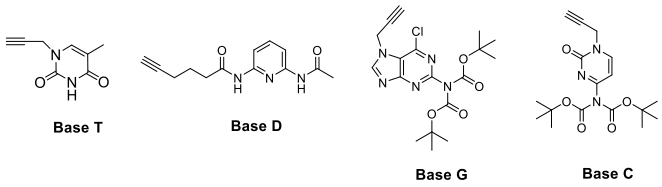


Scheme 3.4 Chemical structure of catalytic side groups.

3.3.2 Synthesis of base side groups

As already presented in section 3.1, the selected units for self-assembly include 2,6-diaminopyridine derivative (D)/thymine (T) and guanine (G)/cytosine (C) complementary pairs. As for the catalytic units, these binding units need to be alkynylated to allow their installation onto the backbone of the oligomers. Besides, the amino groups of base G and C must be protected to avoid side reactions; the deprotection is performed only after incorporation in the SDPs.

More specifically, G and C need BOC-deprotection. Additionally, base G is provided as a chloropurine which is later on oxidized into purine. The structure of the alkynylated bases as shown in Scheme 3.5. The synthetic details are available in Annex 3.2.



Scheme 3.5 Chemical structure of alkynylated bases.

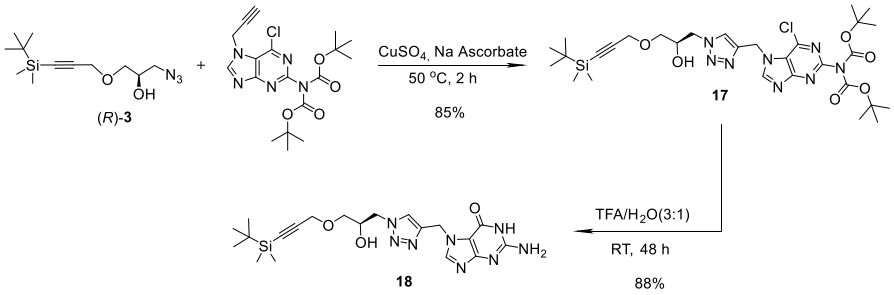
3.3.3 Synthesis of monomer units

The alkyne-modified side groups are then introduced on **(*R*)-3** or **(*S*)-3** by CuAAC, using the functional procedure described in section 3.2. The deprotection of precursors to bases G and C can be performed at this step. As a typical, we provide here the synthesis of monomer **18** comprising base G. The synthesis of other monomer units (compounds **11** to **20**) is described in the Annex 3.1 and 3.2 at the end of this chapter.

Compound **(*R*)-3** (0.808 g, 3 mmol) and **base G** (1.224 g, 1 equiv.) were added into a flask followed by EtOH (21 mL) with stirring. Water (9 mL), sodium ascorbate solution (120 mg in 4 mL water, 0.2 equiv.) and $CuSO_4$ solution (48 mg in 3 mL water, 0.1 equiv.) were successively added to the reaction mixture. The reaction was stirred for 2 h at 50 °C followed by cooling to room temperature. Then, EtOAc (150 mL) and Na_2EDTA (0.05 M, 200 mL) were added into the mixture followed by air bubbling for 30 min. After that, the organic phase was collected, and the water phase was extracted with DCM (150 × 3 mL). The organic phases were combined, dried with Na_2SO_4 and concentrated under vacuum. The residue was purified by column

chromatography (EtOAc /n-hexane = 5/5 to 0/10) to give the final product **17** as foamy white solid (1.727 g, 85%). (*R*)-**17**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 8.32 (s, 1H), 7.92 (s, 1H), 5.53 (s, 2H), 4.53 (dd, *J* = 14.0, 3.3 Hz, 1H), 4.34 (dd, *J* = 14.1, 7.7 Hz, 1H), 4.19 (s, 3H), 3.60 (dd, *J* = 9.7, 4.5 Hz, 1H), 3.49 (dd, *J* = 9.7, 5.6 Hz, 1H), 1.48 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 151.98, 151.48, 151.00, 145.96, 125.12, 101.26, 84.04, 70.69, 69.17, 59.63, 53.36, 39.52, 28.05, 26.14, 16.55, -4.59. **HRMS** *m/z*= 677.2993 (calcd. for $C_{30}H_{45}ClN_8O_6Si$ 677.2993 $[M+H]^+$).

Compound (*R*)-**17** (0.211 g, 0.3 mmol) was added in a flask followed by TFA (3 mL) and Milli-Q water (1mL). The mixture was stirred for 48 h at room temperature. Then, the mixture was neutralized with saturated $NaHCO_3$ and extracted with DCM (3 × 50 mL). The organic phase was washed with brine (2 × 50 mL), dried with Na_2SO_4 and concentrated under vacuum. The final product **18** was obtained after passing the residue through a chromatography column (MeOH/DCM = 0/100 to 15/85) as a white solid (0.126 g, 88%). **18**, **^{1}H NMR (300 MHz, DMSO)** *δ* 10.57 (s, 1H), 8.01 (s, 1H), 7.77 (t, *J* = 6.2 Hz, 1H), 7.64 (s, 1H), 7.39 – 7.09 (m, 5H), 6.45 (s, 2H), 5.20 (s, 3H), 4.57 (t, *J* = 6.3 Hz, 2H), 4.30 – 3.97 (m, 4H), 3.56 (td, *J* = 10.5, 9.9, 4.6 Hz, 2H), 0.90 (s, 9H), 0.08 (s, 6H). **^{13}C NMR (75 MHz, DMSO)** *δ* 156.76, 153.73, 151.02, 142.34, 137.09, 124.34, 116.37, 103.17, 88.84, 71.27, 68.04, 58.58, 52.94, 37.82, 25.86, 16.12, -4.76. **HRMS** *m/z*= 459.2280 (calcd. for $C_{20}H_{30}N_8O_3Si$ 459.2283 $[M+H]^+$).



Scheme 3.6 Synthesis and deprotection of monomer 17 into monomer 18.

3.3.4 Termination of monomer units

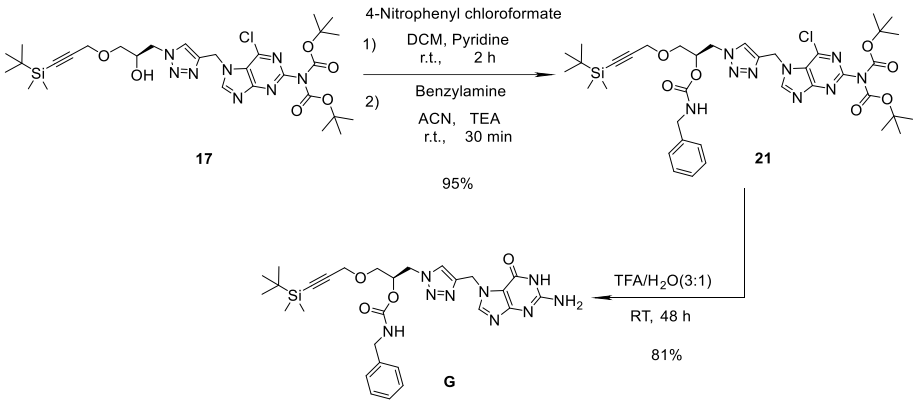
The synthesized monomer units contain a hydroxyl group, which is used for elongation of oligomer chains (Scheme 3.3). In some circumstances, this -OH group must be transformed to avoid reacting with the catalytic system (which oxidizes alcohols).

In this case, the hydroxyl group is reacted to form a benzyl carbamate group. This is schematized below (Scheme 3.10) for the monomer unit comprising base G; the case of other units (compounds **21, 22, M, P, I, T, D, C** and **C_6**) is detailed at the end of this chapter, in Annexes 3.1 and 3.2.

Compound (*R*)-**17** (260 mg, 0.38 mmol, 1 equiv.) was added in a flask followed by DCM (5 mL), 4-nitrophenyl chloroformate (160 mg, 1.5 equiv.) and pyridine (64 µL, 1.5 equiv.). The mixture was stirred for 2 h at room temperature followed by solvent evaporation. The residue was dissolved in ACN (5 mL) and added to a flask followed by benzylamine (114 mg, 2 equiv.) and Et_3N (221 µL, 3 equiv.). The mixture was stirred for 1 h at room temperature, then EtOAc (50 mL) and water (50 mL) were added to the mixture. The organic phase was washed with water (2 × 50 mL) and brine (2 × 50 mL), dried with Na_2SO_4 and concentrated under vacuum. The final product **21** was obtained after passing the residue through a chromatography column

(EtOAc/*n*-hexane = 2/8 to 6/4) as foamy white solid (306 mg, 95%). **21**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.27 (s, 1H), 7.80 (s, 1H), 7.32 - 7.24 (m, 5H), 7.18 – 7.11 (m, 2H), 5.47 (dd, *J* = 15.5, 4.3 Hz, 3H), 5.16 (quint, *J* = 4.1 Hz, 1H), 4.71 – 4.46 (m, 2H), 4.28 – 4.01 (m, 4H), 3.72 – 3.54 (m, 2H), 1.48 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 155.12, 152.55, 151.53, 151.21, 145.97, 141.35, 138.12, 128.80, 127.67, 127.52, 124.90, 101.23, 90.89, 84.19, 71.25, 68.18, 59.65, 51.06, 45.03, 39.67, 28.06, 26.16, 16.56, -4.57. **HRMS** *m/z* = 832.3337 (calcd. for $C_{38}H_{52}ClN_9O_7Si$ 832.3340 $[M+Na]^+$).

Compound **21** (270 mg, 0.34 mmol, 1 equiv.) was added in a flask followed by TFA (3 mL) and Milli-Q water (1mL). The mixture was stirred for 48 h at room temperature. Then, the mixture was neutralized with saturated $NaHCO_3$ and extracted with DCM (3 × 50 mL). The organic phase was washed brine (2 × 50 mL), dried with Na_2SO_4 and concentrated under vacuum. The final product **G** was obtained after passing the residue through a chromatography column (MeOH/DCM = 3/97 to 8/92) as a white solid (0.16 g, 81%). **G**, **^{1}H NMR (300 MHz, DMSO-d_6)** δ 10.57 (s, 1H), 8.01 (s, 1H), 7.77 (t, *J* = 6.2 Hz, 1H), 7.64 (s, 1H), 7.39 – 7.09 (m, 5H), 6.45 (s, 2H, H_{11}), 5.20 (s, 3H), 4.57 (t, *J* = 6.3 Hz, 2H), 4.30 – 3.97 (m, 4H), 3.56 (td, *J* = 10.5, 9.9, 4.6 Hz, 2H), 0.90 (s, 9H), 0.08 (s, 6H). **^{13}C NMR (75 MHz, DMSO-d_6)** δ 156.75, 155.34, 153.63, 151.00, 142.51, 139.41, 136.93, 128.19, 126.83, 126.72, 124.23, 116.38, 102.74, 89.22, 70.49, 68.24, 58.59, 50.14, 43.66, 37.75, 25.81, 16.06, -4.81. **HRMS** *m/z* = 592.2811 (calcd. for $C_{28}H_{37}N_9O_4Si$ 592.2811 $[M+H]^+$).



Scheme 3.7 Synthesis and deprotection of monomer 21.

3.4 Synthesis of precision oligomers

3.4.1 Precision oligomers decorated with nucleobases and analogs

The specificity of DNA duplex formation has been used to develop applications such as DNA origami and nanoparticle assembly.[183–185] These rely on the formation of double helices, in which hydrogen bonds between nucleobases on complementary strands and base-stacking interactions, provide the stabilization mechanism for duplex formation.[186] The persistence length of single- and double-strand DNA and the short distance between successive nucleobases along the DNA backbone are important factors in this stabilization. Synthetic oligomers bearing complementary recognition units can also form sequence-selective duplexes.[187,188] Among those, a number of synthetic sequence-defined oligomers bearing nucleobases or analogous side groups have been demonstrated to form H-bonded duplexes,[189,190] sometimes including with complementary oligonucleotides; peptide nucleic acids are a well-known example.

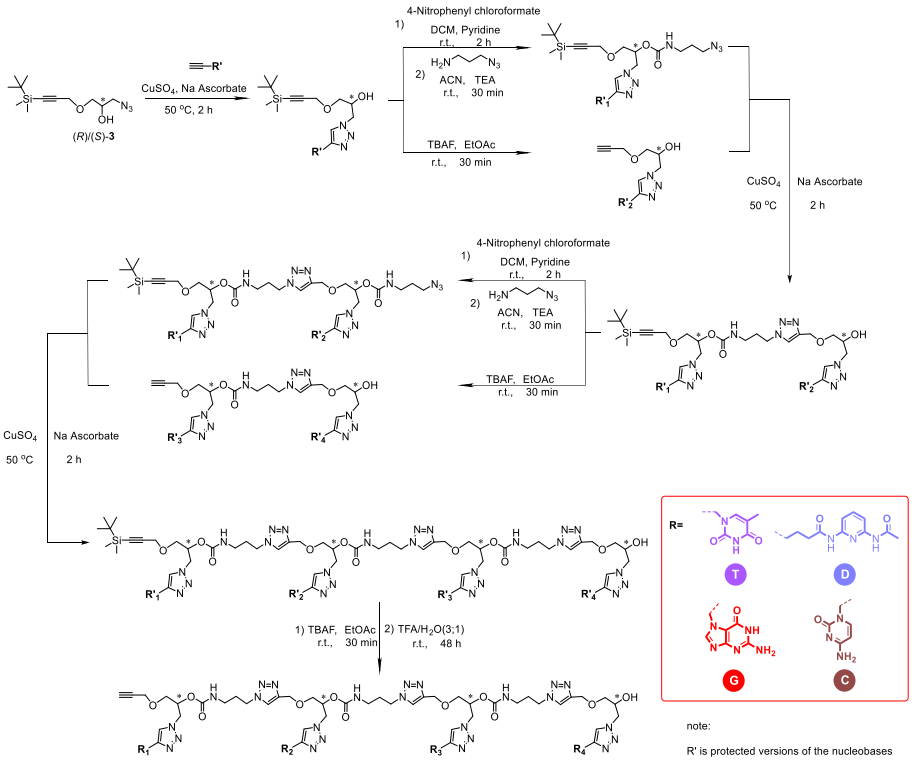
Most nucleobase-bearing synthetic systems have relatively rigid backbones and/or moderate distances between successive bases, leading to some preservation of the chemical (primary) sequence of bases in the average three-dimensional conformation of the chains. This partial spatial preservation of the chemically-encoded information stored in the primary structure of the chain presumably helps in the recognition of a complementary sequence. In the case of our oligomers, however, the distances between bases and the flexibility of the backbone may fully scramble the information contained in the primary sequence of the chains. It is thus not clear whether the blurring resulting from conformational fluctuations would completely prevent recognition between complementary chains.

To answer this question, we synthesized flexible precise oligomers with a well-defined sequence of nucleobases using the monomers developed in the previous part of this chapter (section 3.3.3 and 3.3.4). To avoid other scrambling mechanisms than conformational blurring, a complete control over molar mass, monomer sequence, and chirality of the chains is needed. Our route provides such a control. Therefore, we synthesized a series of complementary and non-complementary nucleobase tetramers, with the aim of them being grafted on surfaces to study recognition by ellipsometry. This recognition part of the work was performed by another member of the group and is thus not discussed in detail in this thesis. Here we focus on the synthesis of the oligomers.

Synthesis of precision oligomers decorated with nucleobases or analogues

The synthesis of sequence-defined oligo(urethane-triazole) tetramers of controlled chirality was performed as shown in Scheme 3.3. The monomers are equipped with guanine (G), cytosine (C), a 2,6-

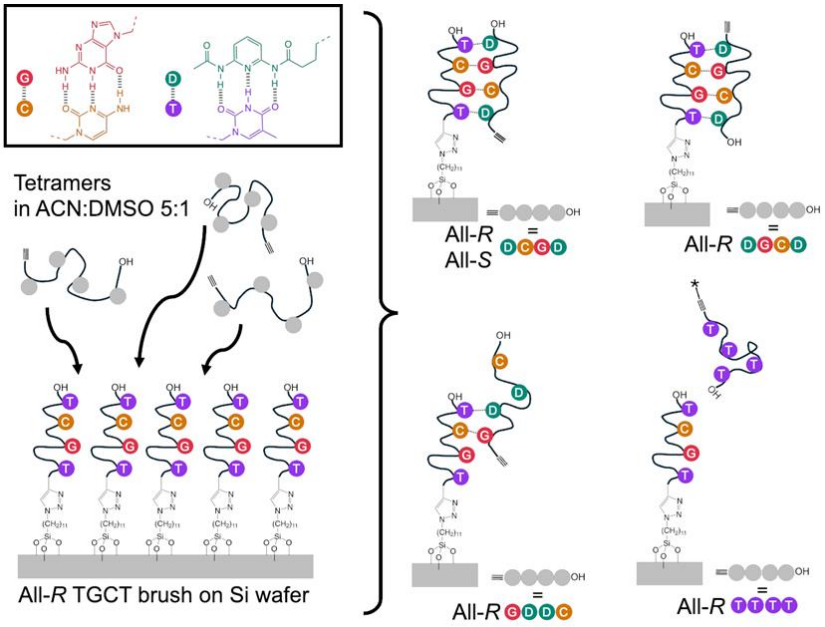
diaminopyridine derivative (D) and thymine (T) side groups (Scheme 3.8); trivalent H-bonded duplexes can be formed between the complementary base pair analogues C and G, or D and T (Scheme 3.1).



Scheme 3.8 General synthetic route of precision oligomers decorated with nucleobases or analogues.

Six different tetramers were obtained from these monomer units; they are designed by a succession of letters corresponding to the successive nucleobases, starting from the alkyne end group and ending in the hydroxyl terminal group. An all-*R* TGCT tetramer was selected as target chain and end-grafted on a silicon surface (Scheme 3.9), using its alkyne terminal group to click it onto a layer of

11-azidoundecyltrimethoxysilane (AzUTMS). The four units and sequence of this target chain were selected having in mind the following considerations: 1. maximal sequence specificity but minimal possible self-binding for the target and complementary probe chain; therefore, there are only two complementary bases in the sequence, and one repeated base; 2. no sequence symmetry to discriminate the chain direction; 3. good solubility in the conditions of grafting. Four different tetramers having an identical composition but different sequences or chirality were synthesized as probe chains for recognition (Scheme 3.9) to investigate the influence of sequence and chirality: all-*R* DCGD, all-*R* DGCD (flipped version of the previous one) and all-*S* DCGD chains that have a sequence fully complementary to the target chain when in extended conformation, and a permuted all-*R* GDDC sequence which can only form a partial complementary duplex with the TGCT target chain unless a strongly folded conformation is taken (Scheme 3.9). Additionally, a non-complementary all-*R* TTTT chain which cannot form a duplex with the probe chain was also synthesized.



Scheme 3.9 Different sequences of synthesized tetramers and schematic possible binding pattern between the target grafted chains and the different probe chains. The pattern of hydrogen bonding between the nucleobases or analogues installed on the tetramers is displayed in the top left box. The alkyne end-group of the TTTT oligomer was not deprotected (as indicated by the '*') to improve solubility; it is terminated by a *tert*-butyl-silyl group.

The chain structures were checked by NMR and tandem mass spectrometry, as fully described in the Annex 3.3 (it is the ESI of my co-authored article titled "Chain stretching in brushes favors sequence recognition for nucleobase-functionalized flexible precise oligomers"), in which the reader will have access to full synthetic details. As a brief summary, the all-over yields of these tetramers, starting from monomer units, are as follows: all-*R* TGCT, all-*R* DCGD, all-*S* DCGD, all-*R* DGCD, all-*R* GDDC and all-*R* TTTT are 31%, 31%, 36%, 37%, 53% and 66%, respectively.

The work on recognition demonstrated the possibility to graft the

target chains onto azide-modified silicon wafers in mushroom or brush layer configurations. In the mushroom regime, the chains lack recognition specificity due to competing hydrogen bonds with the triazole-urethane backbone and spatial scrambling of nucleobases. In the brush regime, recognition improves, with complementary probe chains showing twice the binding probability of non-complementary ones irrespective of their chirality or directionality, driven by partial chain stretching and reduced scrambling. However, specificity remains lower than DNA, which benefits from a rigid, charged backbone and closer nucleobase spacing. While DNA-like applications are unlikely, the observed irreversible binding suggests potential for simpler uses, such as nanoparticle superlattices, warranting further exploration of flexible precision oligomers as DNA alternatives. These results are given in Annex 3.6, which is the preprint of my co-authored article titled “Chain stretching in brushes favors sequence recognition for nucleobase-functionalized flexible precise oligomers”.

3.4.2 Synthesis of oligomers for catalysis

3.4.2.1 Synthesis strategy

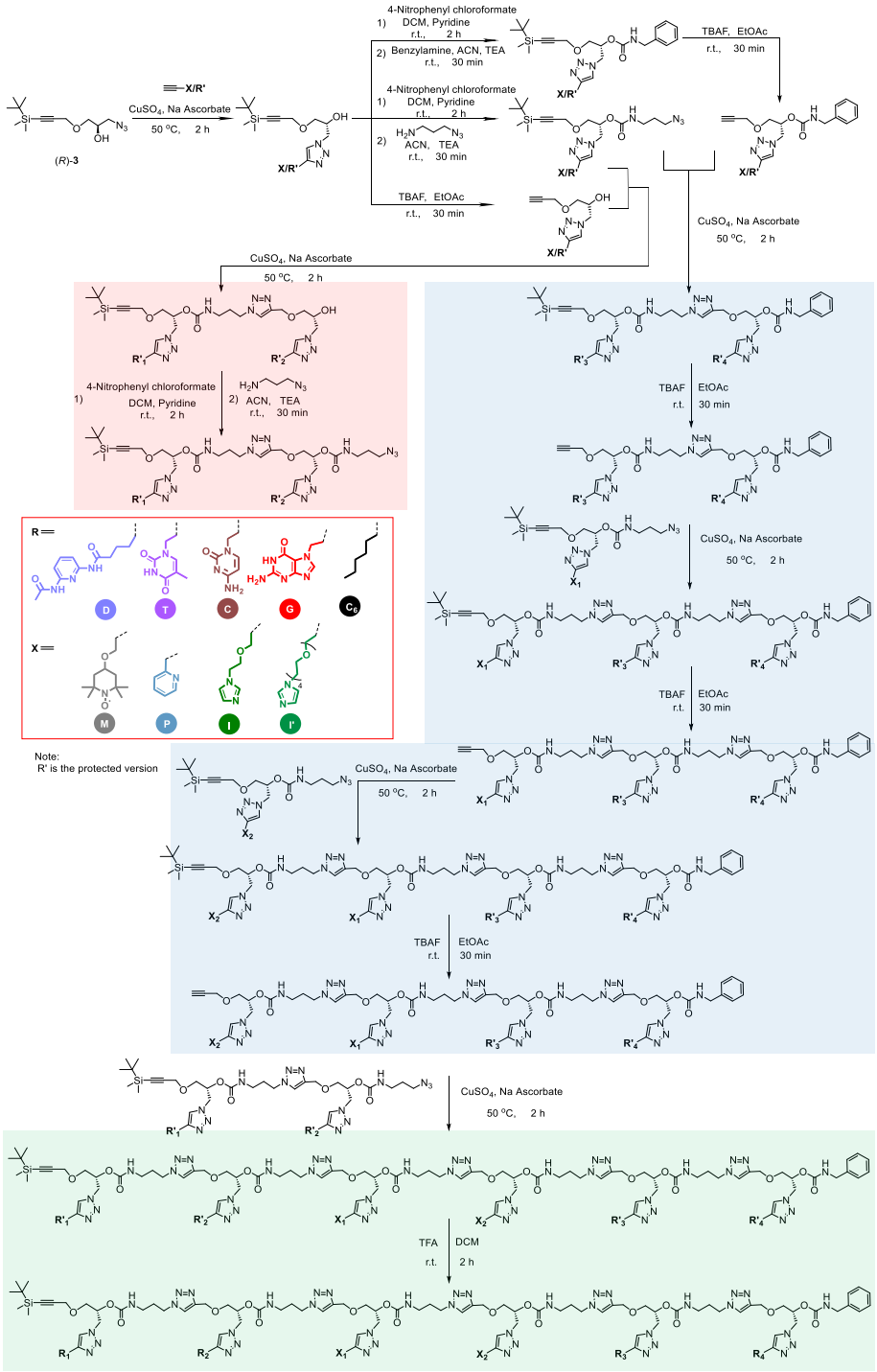
We then concentrated on the synthesis of oligomers capable to assemble to form a catalytically-active pocket involving three functional groups (I, M and P as described previously – chapter 2, section 2.3.2) for the aerobic oxidation of alcohols. The rationale behind the selection of specific sequences to favor catalytic activity will be discussed in the next chapter. Here, we only focus on the synthetic pathway of the catalytic oligomers. We also need to point out that all the sequences are in all-*R* configuration. This was selected because homochiral polymers are more conducive to form

stable secondary structures, which in turn may promote the formation of stable functional entities.[191,192]

The general synthetic route is shown in Scheme 3.10; all the reaction conditions have already been described in section 3.2. To make the thesis more readable, here we will only briefly present how we synthesized all the oligomers step by step; the rest of synthetic details for all sequences are available in Annexes 3.3 and 3.4 at the end of this chapter.

As mentioned in Section 3.1, the oligomers for self-assembled catalytic pockets include $C_6G\underline{MP}TC_6$, $C_6C\underline{I'IP}DC_6$, $C_6C\underline{IP}DC_6$ and $C_6C\underline{II}DC_6$ (the catalytic units are underlined). The synthetic methods are essentially the same for these four sequences. The synthesis was divided into the synthesis of a dimer or trimer (in red in Scheme 3.10) and a tetramer (in blue in Scheme 3.10). The dimer or trimer was synthesized sequentially, whereas the corresponding tetramer was synthesized in reverse order. Then, the final sequenced oligomer was obtained by the coupling of an elongated dimer or trimer with an alkyne-ended tetramer. There are several reasons for this procedure: first, both ends of each sequence are C_6; therefore, starting with two chain ends is time saving. Besides, $C_6C\underline{I'IP}DC_6$, $C_6C\underline{IP}DC_6$ and $C_6C\underline{II}DC_6$ have the same first two functional groups and last two functional groups in the sequence; therefore, our procedure allowed us to spare a lot of synthesis work. Additionally, we found it easier to perform the elongation of long chains in reverse order, adding the TBS-protected monomer to the growing chains - which we thus did for the tetramer. Once the desired sequence is ready, the BOC-deprotection (in green in Scheme 3.10) is performed to get the final products, except for $C_6G\underline{MP}TC_6$, for which the oxidation of chloropurine needs to be done before coupling with the tetramer to

avoid oxidizing TEMPO. All the chain structures were checked by NMR and tandem mass spectrometry, as fully described in Annex 3.3 and SI of our published paper from section 2 to 4. In short, the oligomers for the catalytic pockets were synthesized with reasonable yields (21%, 39%, 25% and 26%, for C_6GMPTC_6, C_6CI'IPDC_6, C_6CIPDC_6 and C_6CIIDC_6, respectively), starting from monomer units. The synthesis was usually without problem, except longer chains for which solubility issues appeared and for which purification became increasingly tedious.



Scheme 3.10 General synthetic route of precision oligomers for catalysis.

The synthesis details of stereo-controlled and sequence-defined oligomers for catalysis and all the intermediates are available in Annexes 3.3, 3.4 and 3.5.

3.4.2.2 Sequence confirmation by TOF-MS/MS

As said in Section 2.2.3, MS/MS is a useful technique to analyse the sequences by ionizing the given molecules and detecting the corresponding fragments. Therefore, TOF-MS/MS was performed to get sequence information for each oligomer chain. During the ionization process, cleavage mainly happened at the ether bond of the main chain, resulting in the molecule being split into two complementary fragments. Since the synthesized molecules contain several cations, multiple charging is also possible. Here we only take C_6GMPTC_6 as example, as shown in Figure 3.1. The sequence of C_6GMPTC_6 reading from left to right was detected at *m/z* values 544.36, 1014.56, 1506.85, 1889.98 and 2355.15, respectively, corresponding to the sequences C_6, C_6G, C_6GM, C_6GMP and C_6GMPT. Reading from the opposite direction, there is a fragment at *m/z*=343.23, which corresponds to C_6. Thus, the whole sequence of C_6GMPTC_6 is found in the spectrum. When reading from right to left, a similar conclusion can be obtained. Note that, during this process, the fragment $GMPTC_6$ was doubly charged. In conclusion, these fragments confirm that the desired sequence is obtained. Similarly, we confirmed that the sequences of C_6CIIDC_6, C_6CIPDC_6 and $C_6CI'IPDC_6$ were well defined according to their TOF-MS/MS spectra. The details of these oligomers are available in Annex 3.5.

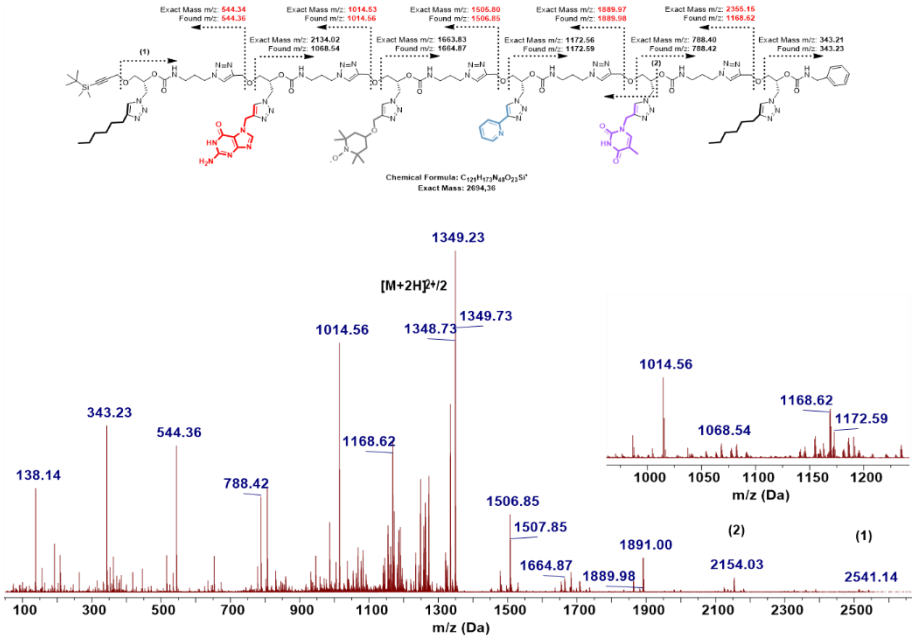


Figure 3.1 TOF-MS/MS of C_6GMPTC_6.

3.5 Conclusions

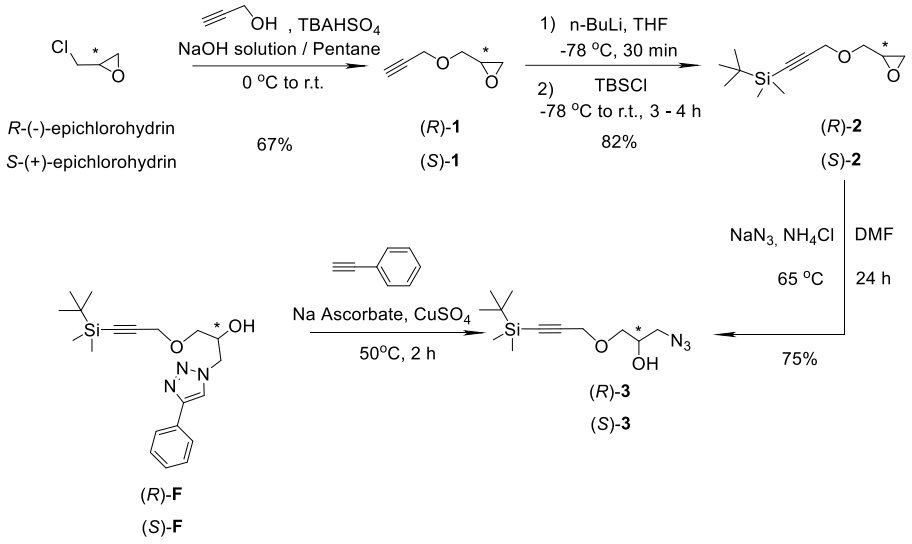
We synthesized nucleobase-substituted and multifunctional precision poly(triazole-carbamate) chains with precise sequences, uniform masses, and defined stereo-configurations. Those chains contain self-assembly units or catalytic units. The NMR results confirm the structure of the oligomers; stereochemistry was established in previous work of Li *et al*.[171] and their mass is confirmed by MS. MS/MS results also show that the synthesized oligomers have the proper sequence as designed. Those synthesized oligomers were obtained with excellent yields together with good atom economy, even though the purification was not always easy.

Our work thus demonstrates the feasibility of our route to provide access to precision oligomers with a wide variety of functional groups. Its main drawback is, however, the limited chain lengths that can be

achieved, due to the solution synthesis and purification work needed.

Annex 3.1 Chiral purity

The chirality of the chiral precursor is determined by the starting material, epichlorohydrin, which was purchased from Sigma-Aldrich. The chiral epichlorohydrin was modified and ring-opened to give the chiral precursors (*R*)-3 and (*S*)-3, used in this thesis. The chiral purity of the chiral precursors (*R*)-3 and (*S*)-3 were tested by coupling with phenylacetylene to give chiral (*R*)-F and (*S*)-F, which were already reported previous work done in our group.[171]



Scheme A.3.1 Synthesis of (*R*)-F and (*S*)-F.

Chiral high-performance liquid chromatography (Chiral HPLC) was performed on a Waters 600 Controller with a Waters 996 photodiode array detector and a Waters 717 plus autosampler. A CHIRALPAK IA column (4.6 × 250 mm) was used to separate samples under a mobile phase of iso-hexane/EtOH mixture (v/v = 9/1) at a flow rate of 1 mL/min. The detection wavelength was set to 245 nm. As Figure A.3.1 shows, the ee of both isomers (*R*)-F and (*S*)-F is 99.2%,

showing that the chirality was well controlled under the synthetic conditions. Through calculation, it can be inferred that the ee of the hexamer and heptamer used for catalysis is 95.2% and 94.5%, respectively (see main text).

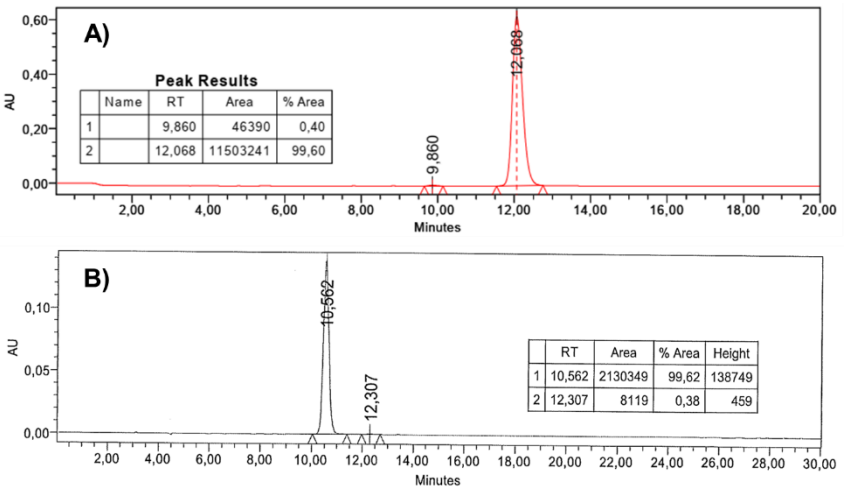
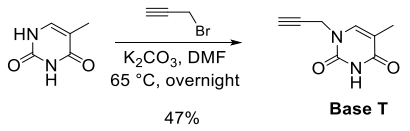


Figure A.3.1 Chiral HPLC traces of monomer (*R*)-F (A) and (*S*)-F (B) in isohexane/EtOH (v/v = 90/10).

Annex 3.2 Alkynylation of the bases

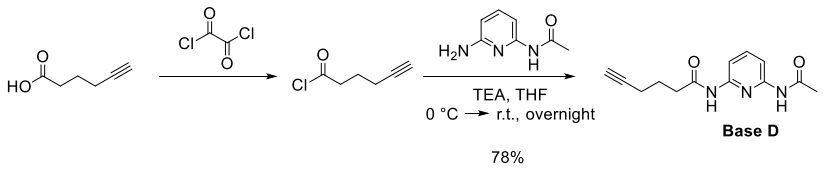
Thymine (5 g, 39.7 mmol) and K_2O_3 (2.743 g, 19.8 mmol, 0.5 equiv.) a added into a flask followed by adding DMF (200 mL) and propargyl bromide (W% = 80% in toluene) (4.42 mL, 39.7 mmol, 1 equiv.), and the mixture was stirred at 60 °C overnight.[193,194] After that, the solvent was removed under reduced pressure. Milli-Q water (250 mL) was added into the residue and extracted with EtOAc (150×3 mL). The organic phase was collected, and dried with Na_2SO_4 before being concentrated under reduced pressure. The crude product was purified by flash chromatography on silica gel column (EtOAc /n-hexane = 3/7 to 6/4) to give the final product base T as a white solid (3.055 g, 47%). **Base T**, **^{1}H NMR (300 MHz, DMSO)** δ 11.36 (s, 1H), 7.56 (d, *J* = 1.3 Hz, 1H), 4.47 (d, *J* = 2.5 Hz, 2H), 3.39 (t, *J* = 2.5 Hz, 1H), 1.77 (d, J = 1.2 Hz, 3H). **^{13}C NMR (75 MHz, DMSO)** δ 164.12, 150.35, 140.10, 109.37, 78.65, 75.63, 36.29, 11.90. **HRMS** *m/z*= 165.06593 (calcd. for $C_8H_8N_2O_2$ 165.06585 $[M+H]^+$)



Scheme A.3.2 Synthesis of base T.

5-hexynoic acid (3.24 g, 28.9 mmol) was added into anhydrous THF (30 mL) followed by adding oxalyl chloride (5 g, 39.4 mmol, 1.4 equiv.) dropwise at 0 ℃. Then, a few drops of DMF were added into the mixture, and the mixture was stirring at room temperature for 2 h. After that, the mixture was concentrated via distillation in vacuo to give the product 5-hexynoic chloride without further purification.[195] 2,6-Diaminopyridine (3.976 g, 26.3 mmol) and triethylamine (3.992 g, 39.4 mmol, 1.5 equiv.) were dissolved in dry THF (80 mL), then the

mixture was cooled to 0 °C in an ice bath. 5-hexynoyl chloride in anhydrous THF (20 mL) was added dropwise within 1 h. Then the mixture was stirred at 0 °C for 1 h followed by stirring at room temperature overnight.[196] The solvent was removed under reduced pressure, and then the residue was dissolved in DCM (100 mL) and washed with 0.5 mol/L HCl (50×3 mL). NaOH aq (1mol/L) was used to neutralize excess acid before washing with saturated NaCl. The organic layer was collected, dried with Na_2SO_4 and concentrated under reduced pressure. The final product is a light-yellow solid (5.03 g, 78%) after recrystallization in EtOAc. **Base D**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.88 (d, *J* = 8.0 Hz, 2H), 7.69 (t, *J* = 8.1 Hz, 3H), 2.52 (d, *J* = 7.3 Hz, 2H), 2.32 (d, *J* = 2.6 Hz, 2H), 2.19 (s, 3H), 2.00 (t, *J* = 2.6 Hz, 1H), 1.97 – 1.88 (m, 2H).**^{13}C NMR (75 MHz, $CDCl_3$)** δ 170.88, 168.66, 149.53, 149.43, 141.06, 109.59, 83.36, 69.64, 36.09, 24.88, 23.78, 17.89. **HRMS** *m/z*= 246.12346 (calcd. for $C_{13}H_{15}N_3O_2$ 246.12370 $[M+H]^+$)

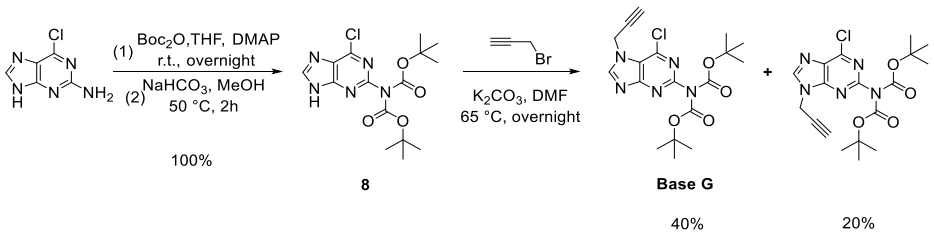


Scheme A.3.3 Synthesis of base D.

Compound **8** was synthesized according to the reported literature.[197] **8**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.39 (s, 1H), 6.46 (s, 1H), 1.49 (s, 18H).**^{13}C NMR (75 MHz, $CDCl_3$)** δ 154.09, 151.45, 150.82, 150.70, 145.97, 129.33, 84.51, 28.01. **HRMS** *m/z*= 370.1275 (calcd. for $C_{15}H_{20}ClN_5O_4$ 370.1277 $[M+H]^+$)

Compound 8 (4.24 g, 11.5 mmol) and K_2O_3 (0.794 g, 5.75 mmol, 0.5 equiv.) were added into a flask followed by adding DMF (60 mL) and

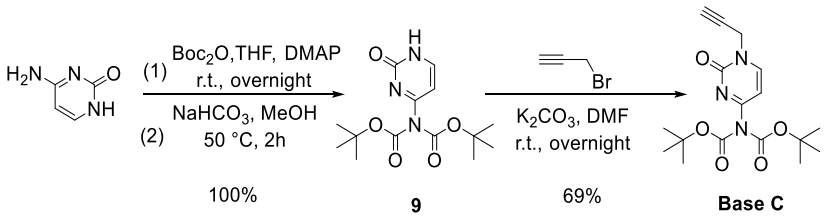
propargyl bromide (W% = 80% in toluene) (1.28 mL, 11.5 mmol, 1 equiv.) with stirring at 60 °C overnight. The solvent was removed under reduced pressure. Milli-Q water (100 mL) was added into the residue and extracted with EtOAc (100×3 mL). The organic phase was collected and dried with Na_2SO_4 before being concentrated under reduced pressure. The crude product was purified by flash chromatography on silica gel column (EtOAc /n-hexane = 1/9 to 5/5) to give a homolog white solid, namely the substitution by propargyl occurred at both N and NH, with the product substituted by propargyl at the N atom being dominant (2:1). Then we selected tert-butyl (tert-butoxycarbonyl) (6-chloro-7-(prop-2-yn-1-yl)-7H-purin-2-yl) carbamate as base G (1.88 g, 40%). **Base G**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.35 (s, 1H), 5.03 (d, *J* = 2.6 Hz, 2H), 2.58 (t, *J* = 2.6 Hz, 1H), 1.45 (s, 18H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 152.27, 151.61, 150.75, 145.18, 130.16, 83.96, 76.17, 33.91, 28.04. **HRMS** *m/z*= 408.14340 (calcd. for $C_{18}H_{22}ClN_5O_4$ 408.14331 $[M+H]^+$).



Scheme A.3.4 Synthesis of base G.

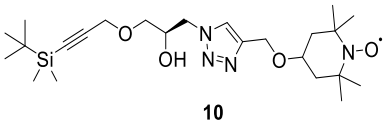
Compound 9 was synthesized according to the reported literature.[197] **^{1}H NMR (300 MHz, $CDCl_3$)** δ 13.13 (s, 1H), 7.74 (d, *J* = 7.2 Hz, 1H), 7.12 (d, *J* = 7.1 Hz, 1H), 1.56 (s, 18H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 163.80, 158.67, 149.59, 145.92, 96.90, 85.13, 27.84. **HRMS** *m/z*= 312.1553 (calcd. for $C_{14}H_{21}N_3O_5$ 312.1554 $[M+H]^+$)

The synthetic route of Base **C** was the same as base **G.** Compound 9 (4.98 g, 16mmol) was used. After the reaction, **base C** was obtained as a light-yellow solid (3.86 g, 69%) purified by flash chromatography on silica gel column (EtOAc /n-hexane = 1/9 to 5/5). **Base C**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.92 (d, *J* = 7.5 Hz, 1H), 7.10 (d, *J* = 7.5 Hz, 1H), 4.67 (d, *J* = 2.6 Hz, 2H), 2.56 (t, *J* = 2.6 Hz, 1H), 1.55 (s, 18H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 162.66, 154.54, 149.63, 145.89, 96.74, 85.15, 38.98, 27.83. **HRMS** *m/z*= 350.1712 (calcd. for $C_{17}H_{23}N_3O_5$ 350.1711 $[M+H]^+$)



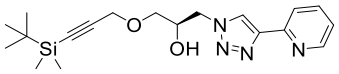
Scheme A.3.5 Synthesis of base C.

Annex 3.3 Report of synthesized intermediates



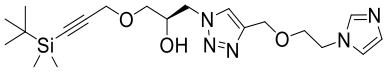
10

10, HRMS *m/z*= 479.3050 (calcd. for $C_{24}H_{43}N_4O_4Si$• 479.3048 $[M+H]^+$)



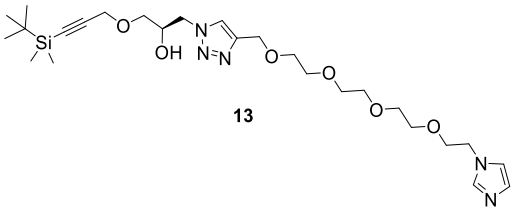
11

The compound **11** were reported in in previous works of Li *et al*.[171]



12

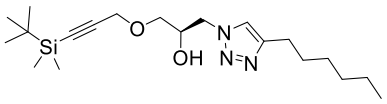
12, ^{1}H NMR (300 MHz, $CDCl_3$) *δ* 7.47 (s, 1H), 7.24 (s, 1H), 6.87 (d, *J* = 21.4 Hz, 2H), 4.58 (m, 3H), 4.25 (m, 4H), 4.03 (m, 2H), 3.83 (m, 2H), 3.63 (dd, *J* = 5.2, 3.7 Hz, 2H, H_4), 0.94 (s, 9H, H_1), 0.12 (s, 6H, H_2). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 144.14, 137.30, 128.42, 124.42, 119.74, 101.69, 90.49, 71.29, 68.92, 68.72, 64.09, 59.61, 54.01, 47.50, 26.16, 16.57, -4.57. **HRMS** *m/z*= 420.2427 (calcd. for $C_{20}H_{33}N_5O_3Si$ 420.2425 $[M+H]^+$)



13

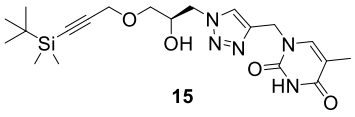
13, **^{1}H NMR (300 MHz, $CDCl_3$** *δ* 7.77 (s, 1H), 7.47 (s, 1H), 6.96 (d, *J* = 19.6 Hz, 2H), 4.67 (d, *J* = 0.6 Hz, 2H), 4.61 (dd, *J* = 13.9, 3.0 Hz, 1H), 4.34 (dd, *J* = 13.9, 8.0 Hz, 1H), 4.25 – 4.14 (m, 3H), 4.09 (dd, *J* = 5.6, 4.4 Hz, 2H), 3.76 – 3.70 (m, 2H), 3.68 – 3.55 (m, 14H), 0.93 (s,

9H), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 144.74, 137.26, 128.07, 124.29, 119.63, 101.56, 90.35, 71.11, 70.72, 70.70, 70.55, 70.28, 69.73, 68.88, 64.67, 59.46, 53.46, 47.30, 26.03, 16.43, -4.70. **HRMS** *m/z*= 552.3214 (calcd. for $C_{26}H_{45}N_5O_6Si$ 552.3212 $[M+H]^+$).



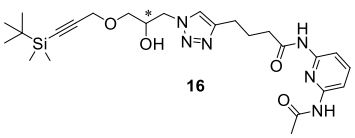
14

14, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.39 (s, 1H), 4.50 (dd, *J* = 13.9, 3.5 Hz, 1H), 4.37 (d, *J* = 7.1 Hz, 1H), 4.20 (s, 3H), 3.64 – 3.44 (m, 2H), 3.35 – 2.84 (m, 1H), 2.76 – 2.61 (m, 2H), 1.63 (q, *J* = 7.7 Hz, 2H), 1.42 – 1.22 (m, 6H), 0.92 (s, 12H), 0.10 (s, 6H).**^{13}C NMR (75 MHz, $CDCl_3$)** δ 148.35, 122.28, 101.48, 90.72, 70.83, 69.30, 59.63, 52.95, 31.69, 29.49, 29.07, 26.15, 25.74, 22.69, 16.55, 14.19, -4.59. **HRMS** *m/z*= 380.2728 (calcd. for $C_{20}H_{37}N_3O_2Si$ 380.2728 $[M+H]^+$).



15

15, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 9.05 (s, 1H), 7.90 (s, 1H), 7.35 (d, *J* = 1.4 Hz, 1H), 5.01 – 4.88 (m, 2H), 4.56 (dd, *J* = 14.1, 3.2 Hz, 1H), 4.38 (dd, *J* = 14.0, 7.7 Hz, 1H), 4.22 (s, 3H), 3.67 – 3.47 (m, 2H), 1.89 (d, *J* = 1.1 Hz, 3H), 0.92 (s, 9H), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 164.19, 151.13, 141.94, 140.43, 125.30, 111.46, 101.31, 90.94, 70.84, 69.27, 59.64, 53.35, 43.15, 26.15, 16.56, 12.43, -4.58. **HRMS** *m/z*= 434.2218 (calcd. for $C_{20}H_{31}N_5O_4Si$ 434.2218 $[M+H]^+$).



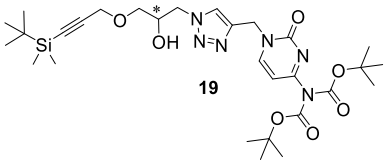
16

(*R*)-**16**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.39 (s, 1H), 8.17 (s, 1H), 7.85 (t,

J = 7.3 Hz, 2H), 7.66 (t, *J* = 8.0 Hz, 1H), 7.48 (s, 1H), 4.53 (dd, *J* = 14.0, 3.4 Hz, 1H), 4.39 (dd, *J* = 14.0, 7.5 Hz, 1H), 4.21 (s, 3H), 3.68 – 3.50 (m, 2H), 2.79 (t, *J* = 6.9 Hz, 2H), 2.41 (t, *J* = 7.1 Hz, 2H), 2.20 (s, 3H), 2.14 – 2.01 (m, 2H), 0.92 (s, 9H), 0.10 (s, 6H).**^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 171.97, 169.28, 149.76, 149.57, 146.88, 140.86, 123.08, 109.55, 101.31, 90.90, 70.96, 69.25, 59.65, 53.24, 36.37, 26.12, 25.18, 24.78, 24.25, 16.54, -4.60. **HRMS** *m/z*= 515.2800 (calcd. for $C_{25}H_{38}N_6O_4Si$ 515.2797 $[M+H]^+$).

(*S*)-**16**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 8.55 (s, 1H), 8.34 (s, 1H), 7.82 (t, *J* = 8.4 Hz, 2H), 7.67 (t, *J* = 8.0 Hz, 1H), 7.49 (s, 1H), 4.54 (dd, *J* = 14.0, 3.3 Hz, 1H), 4.39 (dd, *J* = 14.0, 7.5 Hz, 1H), 4.21 (s, 3H), 3.67 – 3.52 (m, 2H), 2.80 (t, *J* = 6.9 Hz, 2H), 2.43 (t, *J* = 7.1 Hz, 2H), 2.21 (s, 3H), 2.08 (t, *J* = 7.0 Hz, 2H), 0.92 (s, 10H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 171.96, 169.20, 149.40, 146.89, 142.65, 141.12, 123.10, 109.51, 101.32, 90.98, 70.95, 69.33, 59.70, 53.22, 36.39, 26.15, 25.19, 24.84, 24.23, -4.57. **HRMS** *m/z*= 537.2637 (calcd. for $C_{25}H_{38}N_6O_4Si$ 537.2640 $[M+Na]^+$).

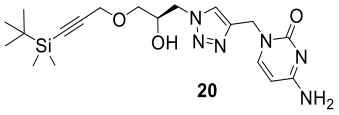
The products (*S*)-**17** was prepared following the same protocol as (*R*)-**17**. (*S*)-**17, ^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 8.32 (s, 1H), 7.92 (s, 1H), 5.53 (s, 2H), 4.53 (dd, *J* = 14.1, 3.3 Hz, 1H), 4.34 (dd, *J* = 14.1, 7.7 Hz, 1H), 4.19 (s, 3H), 3.61 (dd, *J* = 9.7, 4.5 Hz, 1H), 3.49 (dd, *J* = 9.7, 5.6 Hz, 1H), 1.48 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 152.47, 152.01, 151.49, 151.01, 145.94, 141.18, 130.15, 125.11, 101.27, 84.03, 70.70, 69.19, 59.64, 53.36, 39.54, 28.07, 26.15, 16.56, -4.58. **HRMS** *m/z*= 677.2990 (calcd. for $C_{30}H_{45}ClN_8O_6Si$ 677.2993 $[M+H]^+$).



19

(*R*)-**19**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 7.94 (s, 1H), 7.86 (d, *J* = 7.4 Hz, 1H), 7.04 (d, *J* = 7.3 Hz, 1H), 5.08 (s, 2H), 4.52 (dd, *J* = 14.0, 3.5 Hz, 1H), 4.36 (dd, *J* = 14.0, 7.4 Hz, 1H), 4.20 (s, 3H), 3.63 – 3.45 (m, 2H), 1.54 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 162.71, 155.08, 149.62, 148.01, 101.43, 96.84, 90.76, 85.12, 70.69, 69.20, 59.61, 53.31, 45.53, 27.82, 26.16, 16.56, -4.58. **HRMS** *m/z*= 619.3270 (calcd. for $C_{29}H_{46}N_6O_7Si$ 619.3270 $[M+H]^+$).

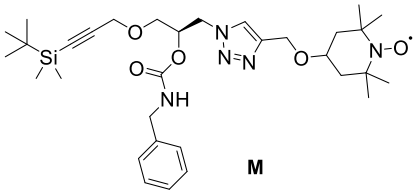
(*S*)-**19**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 7.93 (s, 1H), 7.85 (d, *J* = 7.4 Hz, 1H), 7.03 (d, *J* = 7.3 Hz, 1H), 5.09 (s, 2H), 4.51 (dd, *J* = 14.0, 3.5 Hz, 1H), 4.36 (dd, *J* = 14.0, 7.4 Hz, 1H), 4.20 (s, 3H), 3.66 – 3.43 (m, 2H), 1.54 (s, 18H), 0.93 (s, 9H), 0.11 (s, 6H**). ^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 162.74, 155.05, 149.64, 147.96, 141.75, 125.84, 101.42, 96.83, 90.83, 85.11, 70.66, 69.24, 59.63, 53.28, 45.51, 27.84, 26.17, 16.57, -4.57. **HRMS** *m/z*= 619.3270 (calcd. for $C_{29}H_{46}N_6O_7Si$ 619.3270 $[M+H]^+$).



20

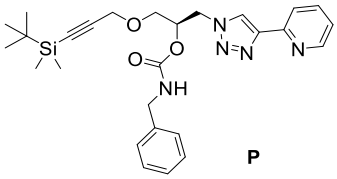
20, **^{1}H NMR (300 MHz, DMSO)** *δ* 7.92 (s, 1H), 7.66 (d, J = 7.2 Hz, 1H), 7.04 (d, *J* = 27.7 Hz, 2H), 5.66 (d, *J* = 7.2 Hz, 1H) , 5.36 (d, *J* = 5.7 Hz, 1H), 4.88 (s, 2H), 4.41 (dd, *J* = 13.9, 3.6 Hz, 1H), 4.23 (d, *J* = 11.9 Hz, 3H), 3.97 (ddd, J = 7.8, 5.6, 3.6 Hz, 1H), 3.41 (td, J = 9.7, 5.6 Hz, 2H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (75 MHz, DMSO)** *δ*

166.00, 155.53, 145.77, 142.81, 124.51, 103.19, 93.57, 88.84, 71.28, 68.06, 58.58, 52.85, 43.43, 25.88, 16.14, -4.75. **HRMS** *m/z*= 419.2219 (calcd. for $C_{19}H_{30}N_6O_3Si$ 419.2221 $[M+H]^+$).



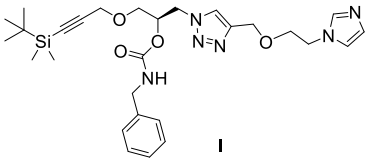
M

M, HRMS *m/z*= 612.3579 (calcd. for $C_{32}H_{51}N_5O_5Si$• 612.3576 $[M+H]^+$).



P

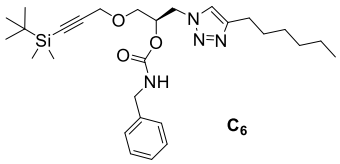
P, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.56 (m, 1H), 8.18 (m, 2H), 7.78 (td, *J* = 7.7, 1.8 Hz, 1H), 7.25 (m, 6H), 5.30 (m, 2H), 4.73 (dd, *J* = 5.3, 2.3 Hz, 2H), 4.28 (dd, *J* = 31.8, 4.1 Hz, 4H), 3.66 (d, *J* = 4.9 Hz, 2H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 155.17, 150.38, 149.50, 148.57, 138.01, 137.06, 128.83, 127.72, 127.63, 123.26, 123.01, 120.46, 101.23, 90.92, 71.26, 67.97, 59.66, 50.64, 45.29, 26.16, 16.56, -4.60. **HRMS** *m/z*= 506.2584 (calcd. for $C_{27}H_{35}N_5O_3Si$ 506.2582 $[M+H]^+$).



I

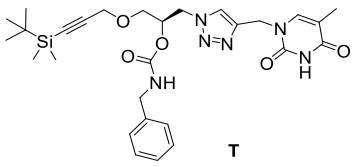
I, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.23 (m, 8H), 7.06 (s, 1H), 6.99 (t, *J* = 6.0 Hz, 1H), 6.79 (d, *J* = 18.0 Hz, 2H), 5.12 (t, *J* = 4.9 Hz, 1H), 4.51 (m, 4H), 4.27 (t, *J* = 5.8 Hz, 2H), 4.07 (m, 2H), 3.96 (t, *J* = 4.9 Hz, 2H), 3.60 (m, 2H), 3.46 (m, 2H), 0.83 (s, 9H), 0.01 (s, 6H).**^{13}C NMR (75**

MHz, $CDCl_3$) δ 155.59, 145.04, 138.76, 137.93, 128.97, 128.74, 127.75, 127.50, 123.48, 119.39, 101.41, 90.68, 70.74, 68.99, 67.90, 64.83, 59.60, 50.75, 47.28, 45.15, 26.17, 16.57, -4.57. **HRMS** *m/z*= 553.2953 (calcd. for $C_{28}H_40N_6O_4Si$ 553.2953 $[M+H]^+$).



C_6

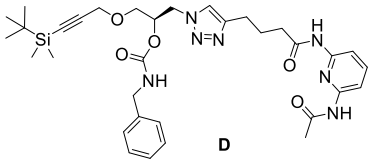
C_6, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.43 – 7.13 (m, 6H), 5.24 (dd, *J* = 6.0, 4.6 Hz, 1H), 5.10 (d, *J* = 6.2 Hz, 1H), 4.61 (t, *J* = 5.9 Hz, 2H4.34 (d, *J* = 6.0 Hz, 2H), 4.21 (s, 2H), 3.63 (dd, J = 4.9, 2.9 Hz, 2H), 2.69 (t, *J* = 7.7 Hz, 2H), 1.63 (d, *J* = 21.7 Hz, 2H), 1.39 – 1.24 (m, 6H), 0.93 (d, *J* = 1.2 Hz, 9H), 0.90 – 0.84 (m, 3H), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 155.27, 148.74, 128.89, 127.80, 127.62, 121.55, 101.30, 71.44, 68.20, 59.64, 50.21, 45.30, 31.72, 29.57, 29.08, 26.17, 25.82, 22.72, 16.57, 14.21, -4.56. **HRMS** *m/z*= 513.3260 (calcd. for $C_{28}H_{44}N_4O_3Si$ 513.3255 $[M+H]^+$).



T

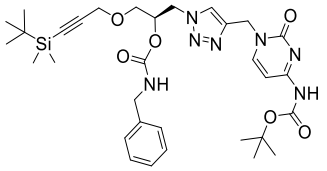
T, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 9.11 (s, 1H), 7.76 (s, 1H), 7.38 – 7.19 (m, 6H), 5.45 (t, *J* = 6.0 Hz, 1H), 5.22 (dd, *J* = 6.5, 4.2 Hz, 1H), 4.90 (d, *J* = 3.1 Hz, 2H), 4.63 (dd, *J* = 7.7, 5.4 Hz, 2H), 4.30 (dd, *J* = 6.1, 2.6 Hz, 2H), 4.20 (d, *J* = 2.5 Hz, 2H), 3.64 (dd, *J* = 4.9, 3.3 Hz, 2H), 1.86 (d, *J* = 1.2 Hz, 3H), 0.92 (s, 9H), 0.11 (s, 6H).**^{13}C NMR (75 MHz, $CDCl_3$)** δ 164.15, 155.18, 142.14, 140.31, 128.85, 127.74, 127.59, 124.97, 111.43, 101.22, 90.93, 71.27, 68.10, 59.63, 50.81, 45.23, 43.16, 26.17, 16.56, 12.39, -4.57. **HRMS** *m/z*= 589.25680 (calcd. for

$C_{28}H_{38}N_6O_5Si$ 589.25652 $[M+H]^+$).



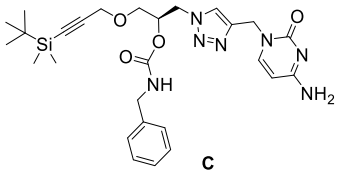
D

D, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.47 (s, 1H), 8.01 – 7.74 (m, 3H), 7.66 (t, *J* = 8.1 Hz, 1H), 7.41 – 7.14 (m, 6H), 5.44 – 5.23 (m, 2H), 4.61 (d, *J* = 5.5 Hz, 2H), 4.31 (d, *J* = 6.0 Hz, 2H), 4.23 (s, 2H), 3.68 (d, *J* = 5.0 Hz, 2H), 2.77 (d, *J* = 7.8 Hz, 2H), 2.32 (t, J = 7.1 Hz, 2H), 2.21 – 1.90 (m, 5H), 0.93 (s, 9H), 0.11 (s, 6H).**^{13}C NMR (75 MHz, $CDCl_3$)** δ 171.79, 168.83, 155.39, 149.77, 147.21, 140.75, 137.94, 128.90, 127.81, 127.49, 122.42, 109.45, 101.14, 91.10, 71.56, 68.43, 59.74, 50.76, 45.22, 36.03, 26.16, 25.22, 24.80, 24.05, 16.57, -4.57. **HRMS** *m/z*= 648.3327 (calcd. for $C_{33}H_{45}N_7O_5Si$ 648.3324 $[M+H]^+$).



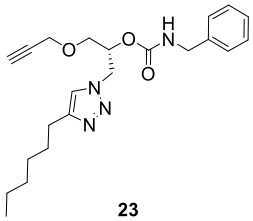
22

22, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.83 (d, *J* = 7.1 Hz, 2H), 7.36 – 7.21 (m, 6H), 7.16 (d, *J* = 7.3 Hz, 1H, H_{10}), 5.34 (t, *J* = 6.0 Hz, 1H), 5.19 (q, *J* = 5.1 Hz, 1H), 5.07 (s, 2H), 4.62 (t, *J* = 5.1 Hz, 2H), 4.31 (t, *J* = 5.3 Hz, 2H), 4.20 (d, *J* = 3.2 Hz, 2H), 3.61 (dd, *J* = 4.9, 3.6 Hz, 2H), 1.49 (s, 9H), 0.92 (s, 9H), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 162.88, 155.16, 151.17, 148.33, 141.88, 128.83, 127.67, 127.64, 125.66, 101.28, 95.44, 90.87, 83.02, 71.17, 68.00, 59.62, 50.67, 45.42, 45.23, 28.16, 26.17, 16.56, -4.57. **HRMS** *m/z*= 652.3272 (calcd. for $C_{32}H_{45}N_7O_6Si$ 652.3273 $[M+H]^+$).



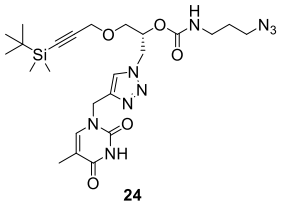
C

C, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.82 (s, 1H), 7.50 (d, *J* = 7.2 Hz, 1H), 7.36 – 7.17 (m, 5H), 6.06 (s, 1H), 5.66 (dd, *J* = 11.7, 6.7 Hz, 2H), 5.20 (t, *J* = 5.6 Hz, 1H), 4.94 (s, 2H), 4.60 (d, *J* = 4.3 Hz, 2H), 4.37 – 4.05 (m, 4H), 3.60 (dd, *J* = 4.9, 2.7 Hz, 2H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 166.00, 156.45, 155.29, 145.91, 142.77, 138.36, 128.78, 127.60, 125.41, 101.32, 94.62, 71.17, 68.11, 59.60, 50.70, 45.14, 44.80, 26.17, 16.56, -4.57. **HRMS** *m/z*= 552.2752 (calcd. for $C_{27}H_{37}N_7O_4Si$ 552.2749 $[M+H]^+$).



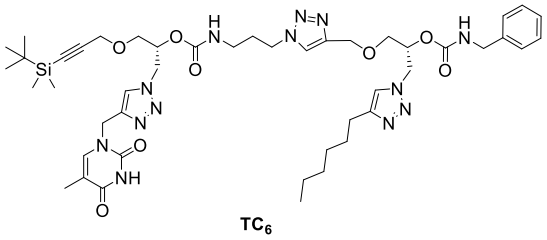
23

23, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.42 – 7.14 (m, 6H, H_7, H_8, H_9 and H_{10}), 5.30 – 5.06 (m, 2H, H_4 and H_5), 4.61 (dd, *J* = 6.8, 5.5 Hz, 2H, H_{10}), 4.42 – 4.30 (m, 2H, H_6), 4.20 (d, *J* = 2.4 Hz, 2H, H_2), 3.62 (t, *J* = 4.7 Hz, 2H, H_3), 2.69 (t, *J* = 7.7 Hz, 2H, H_{12}), 2.44 (t, *J* = 2.4 Hz, 1H, H_1), 1.73 – 1.49 (m, 2H, H_{13}), 1.44 – 1.20 (m, 6H, H_{14}, H_{15} and H_{16}), 1.04 – 0.79 (m, 3H, H_{17}). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 148.72, 128.89, 128.79, 127.81, 127.62, 121.70, 79.00,75.41, 71.34, 68.18, 58.85, 50.07, 45.29, 31.71, 29.53, 29.06, 25.78, 22.70, 14.20. **HRMS** *m/z* 399.2389 (calcd. for $C_{32}H_{51}N_5O_5Si$• 399.2391 $[M+H]^+$).



24

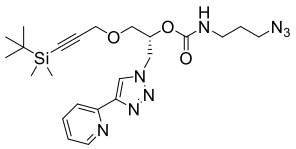
24, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 9.50 (s, 1H), 7.78 (s, 1H), 7.34 (q, J = 1.2 Hz, 1H), 5.35 (t, J = 6.0 Hz, 1H), 5.17 (h, J = 4.5 Hz, 1H), 4.94 (d, J = 3.9 Hz, 2H), 4.71 – 4.53 (m, 2H), 4.20 (d, J = 2.5 Hz, 2H), 3.63 (dd, J = 4.8, 2.8 Hz, 2H), 3.43 – 3.13 (m, 6H), 1.89 (d, J = 1.2 Hz, 3H), 1.77 (dt, J = 9.9, 6.6 Hz, 2H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 164.32, 155.19, 151.32, 142.17, 140.41, 124.98, 111.45, 90.91, 71.16, 68.11, 59.61, 50.83, 49.07, 43.25, 38.66, 38.11, 29.02, 26.14, 16.55, 12.43, -4.60. **HRMS** *m/z*= 560.2760 (calcd. for $C_{24}H_{37}N_9O_5Si$ 560.2760 $[M+H]^+$).



TC_6

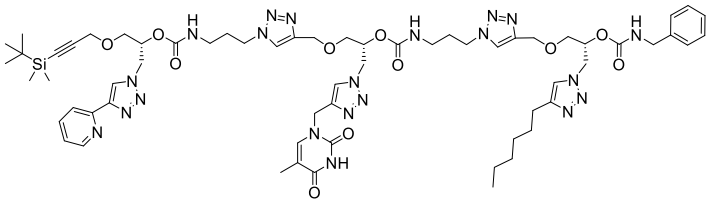
TC_6, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.51 (s, 1H), 7.77 (d, *J* = 20.6 Hz, 2H), 7.37 – 7.15 (m, 7H), 5.81 (t, *J* = 6.1 Hz, 1H), 5.64 (t, *J* = 6.0 Hz, 1H), 5.18 (dd, *J* = 11.4, 6.1 Hz, 2H), 4.88 (q, *J* = 15.0 Hz, 2H), 4.72 – 4.51 (m, 6H), 4.43 – 4.28 (m, 4H), 4.25 – 4.13 (m, 2H), 3.70 – 3.50 (m, 4H), 3.09 (p, *J* = 7.0 Hz, 2H), 2.65 (t, *J* = 7.7 Hz, 2H), 2.05 (d, *J* = 8.3 Hz, 2H), 1.84 (d, *J* = 1.2 Hz, 3H), 1.60 (q, *J* = 7.5 Hz, 2H), 1.28 (tt, *J* = 9.7, 5.2 Hz, 6H), 0.92 (s, 9H), 0.89 – 0.83 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 164.44, 155.61, 155.47, 151.21, 150.13, 149.45, 148.69, 148.35, 144.54, 144.24, 142.26, 140.38, 138.25, 137.18, 128.80, 127.66, 127.55, 125.27, 123.61, 123.58, 123.12,

121.80, 120.32, 111.27, 101.20, 90.95, 71.40, 71.15, 70.89, 68.56, 68.38, 68.27, 64.93, 64.80, 59.62, 50.94, 50.41, 50.05, 47.45, 47.26, 45.14, 37.80, 37.72, 31.66, 30.28, 30.16, 29.48, 29.02, 26.13, 25.70, 22.67, 16.53, 14.19, 12.39, -4.62. **HRMS** *m/z*= 958.5074 (calcd. for $C_{46}H_{67}N_{13}O_8Si$ 958.5078 $[M+H]^+$).



25

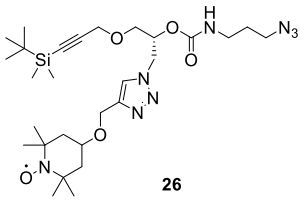
The synthetic details of **25** are available in previous work done by our group.[303]



PTC_6

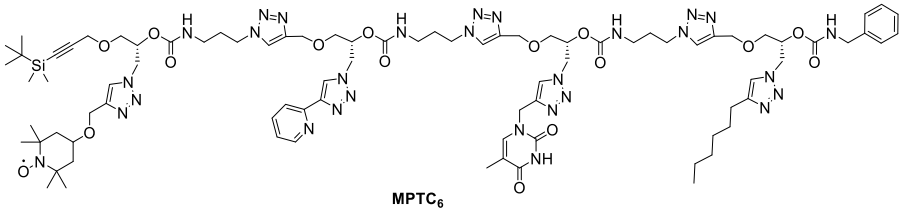
PTC_6, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.79 (s, 1H), 8.58 – 8.47 (m, 1H), 8.33 (s, 1H), 8.08 (dt, *J* = 7.9, 1.1 Hz, 1H), 7.85 – 7.59 (m, 4H), 7.35 – 7.15 (m, 8H), 6.01 (t, *J* = 6.1 Hz, 1H), 5.89 (t, *J* = 6.2 Hz, 1H), 5.69 (t, *J* = 6.1 Hz, 1H), 5.38 – 5.25 (m, 1H), 5.22 – 5.06 (m, 2H), 4.88 (s, 2H), 4.77 – 4.50 (m, $10H_8$), 4.44 – 4.15 (m, 8H), 3.68 (dd, *J* = 6.8, 4.8 Hz, 2H), 3.54 (dd, *J* = 25.0, 6.3 Hz, 4H), 3.08 (ddd, *J* = 34.0, 26.4, 6.5 Hz, 4H), 2.63 (t, *J* = 7.7 Hz, 2H), 2.04 (dt, *J* = 13.2, 7.0 Hz, 4H), 1.89 – 1.78 (m, 3H), 1.58 (q, J = 7.5 Hz, 2H), 1.34 – 1.20 (m, 6H), 0.90 (s, 9H), 0.87 – 0.82 (m, 3H), 0.08 (s, 6H).**^{13}C NMR (126 MHz, $CDCl_3$)** δ 164.44, 155.61, 155.47, 151.21, 150.13, 149.45, 148.69, 148.35, 144.54, 144.24, 142.26, 140.38, 138.25, 137.18, 128.80, 127.66, 127.55, 125.27, 123.61, 123.58, 123.12, 121.80, 120.32, 111.27, 101.20, 90.95, 71.40, 71.15, 70.89, 68.56, 68.38, 68.27, 64.93, 64.80,

59.62, 50.94, 50.41, 50.05, 47.45, 47.26, 45.14, 43.26, 37.80, 37.72, 31.66, 30.28, 30.16, 29.48, 29.02, 26.13, 25.70, 22.67, 16.53, 14.19, 12.39, -4.62. **HRMS** *m/z*= 1342.6734 (calcd. for $C_{63}H_{87}N_{21}O_{11}Si$ 1342.6736 $[M+H]^+$).



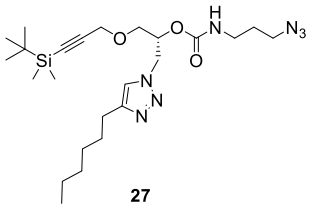
26

26, **HRMS** *m/z* = 605.3591 (calcd. for $C_{28}H_{49}N_8O_5Si$• 605.3590 $[M+H]^+$).



MPTC$_6$

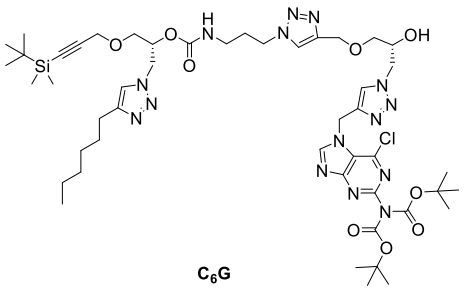
MPTC$_6$, **HRMS** *m/z* = 1833.9468 (calcd. for $C_{85}H_{122}N_{29}O_{16}Si$• 1833.9466 $[M+H]^+$).



27

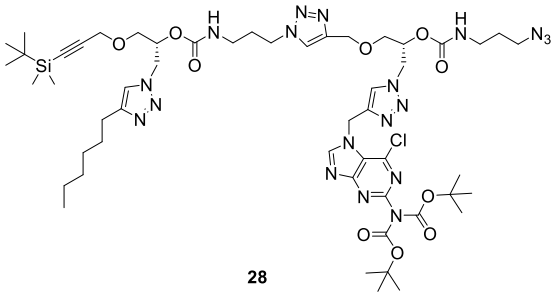
27, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.31 (s, 1H), 5.20 (t, *J* = 5.5 Hz, 1H), 4.94 (d, *J* = 6.4 Hz, 1H), 4.69 – 4.48 (m, 2H), 4.20 (s, 2H), 3.61 (dd, J = 5.0, 1.9 Hz, 2H), 3.35 (t, *J* = 6.5 Hz, 2H), 3.24 (d, *J* = 6.5 Hz, 2H), 2.74 – 2.59 (m, 2H), 1.70 (dt, *J* = 32.5, 7.1 Hz, 4H), 1.41 – 1.22 (m, 6H), 0.93 (s, 12H), 0.11 (d, *J* = 1.0 Hz, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 155.27, 148.74, 121.54, 101.25, 90.87, 71.25, 68.18, 59.63, 50.23, 49.13, 38.71, 31.71, 29.58, 29.07, 26.42, 26.16, 25.81, 22.71, 16.57,

14.20, -4.57. **HRMS** *m/z*= 506.3269 (calcd. for $C_{24}H_{43}N_7O_3Si$ 506.3269 $[M+H]^+$).



C_6G

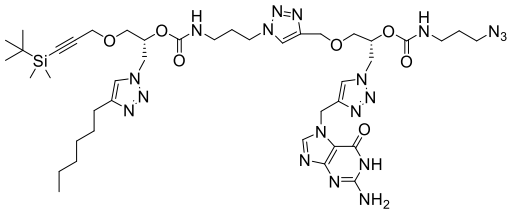
C_6G, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.35 (s, 1H), 7.93 (s, 1H, H2), 7.70 (s, 1H), 7.36 (s, 1H), 5.53 (s, 2H), 5.34 (s, 1H), 5.26 – 5.19 (m, 1H), 4.67 – 4.48 (m, 5H), 4.42 – 4.32 (m, 3H), 4.21 (d, J = 1.5 Hz, 2H), 4.17 – 4.11 (m, 1H), 3.69 – 3.53 (m, 3H), 3.47 (dd, *J* = 10.0, 5.5 Hz, 1H), 3.23 – 2.98 (m, 2H), 2.65 (t, *J* = 7.8 Hz, 2H), 2.07 (ddq, J = 27.5, 13.9, 6.9 Hz, 2H), 1.61 (p, *J* = 7.3 Hz, 2H), 1.46 (s, 18H), 1.35 – 1.22 (m, 6H), 0.92 (s, 9H), 0.88 – 0.83 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 155.51, 152.54, 151.99, 151.34, 150.95, 148.55, 146.18, 144.51, 141.04, 130.10, 125.16, 123.47, 121.93, 101.17, 90.99, 83.99, 71.48, 71.06, 69.20, 68.20, 64.78, 59.65, 53.21, 50.51, 47.42, 39.46, 37.80, 31.67, 30.41, 29.52, 29.02, 28.04, 26.15, 25.66, 22.68, 16.56, 14.19, -4.57. **HRMS** *m/z*= 1068.5326 (calcd. for $C_{63}H_{87}N_{21}O_{11}Si$ 1068.5325 $[M+H]^+$).



28

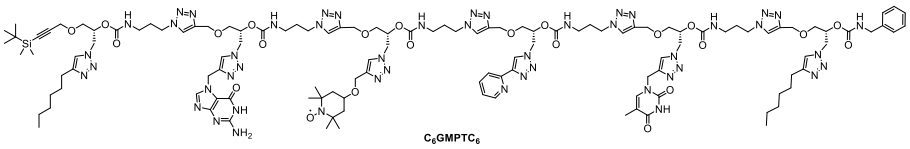
28, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.35 (s, 1H), 7.83 (s, 1H), 7.70 (s,

1H), 7.35 (s, 1H), 5.58 – 5.49 (m, 2H), 5.42 (t, *J* = 6.1 Hz, 2H), 5.22 (dd, *J* = 6.6, 4.5 Hz, 1H), 5.04 (d, *J* = 7.6 Hz, 1H), 4.68 – 4.46 (m, 6H), 4.38 (t, *J* = 6.6 Hz, 2H), 4.20 (d, *J* = 1.7 Hz, 2H), 3.60 (dd, *J* = 15.3, 4.8 Hz, 3H), 3.48 (dd, *J* = 10.3, 6.0 Hz, 1H), 3.29 – 2.97 (m, 6H), 2.66 (t, *J* = 7.8 Hz, 2H), 2.08 (tt, *J* = 13.8, 7.1 Hz, 4H), 1.62 (tt, *J* = 9.7, 6.9 Hz, 2H), 1.48 (s, 18H), 1.36 – 1.22 (m, 6H), 0.92 (s, 9H), 0.88 – 0.83 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 155.51, 155.16, 152.57, 151.91, 151.42, 151.23, 148.64, 146.24, 144.35, 141.40, 130.23, 125.06, 123.48, 121.80, 101.19, 90.93, 84.25, 71.10, 70.98, 68.38, 68.19, 64.84, 59.63, 50.81, 50.39, 49.03, 47.40, 39.65, 38.51, 37.83, 31.67, 29.59, 29.54, 29.02, 28.96, 28.03, 26.14, 25.72, 22.67, 16.55, 14.18, -4.59. **HRMS** *m/z*= 1216.5685 (calcd. for $C_{52}H_{80}ClN_{19}O_{10}Si$ 1216.5686 $[M+Na]^+$).



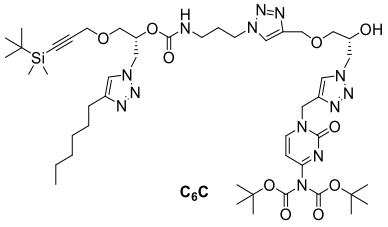
29

29, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 8.67 – 7.50 (m, 3H), 7.39 (s, 1H), 6.33 (d, *J* = 240.3 Hz, 2H), 5.15 (d, *J* = 40.1 Hz, 4H), 4.90 – 4.10 (m, 10H), 3.78 – 2.89 (m, 10H), 2.64 (s, 2H, H_{20}), 2.04 (s, 2H), 1.76 – 1.50 (m, 4H), 1.38 – 1.14 (m, 6H), 0.97 – 0.75 (m, 12H), 0.09 (d, *J* = 2.5 Hz, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 155.67, 155.58, 148.66, 144.23, 121.98, 121.92, 101.34, 90.83, 71.05, 70.99, 68.22, 64.65, 59.61, 50.41, 49.01, 47.63, 38.48, 37.97, 31.68, 30.36, 29.54, 29.03, 26.15, 25.74, 22.68, 16.55, 14.20, -4.57. **HRMS** *m/z*= 976.5155 (calcd. for $C_{42}H_{65}N_{19}O_7Si$ 976.5156 $[M+H]^+$).



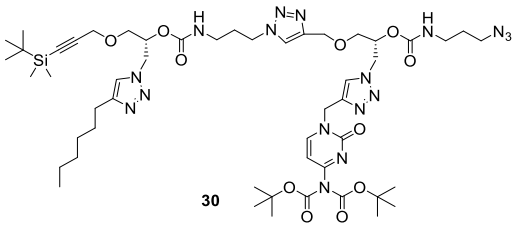
C_6GMPTC_6

C_6GMPTC_6, **TOF MS ES +** *m/z* = 2695,3744 (calcd. for $C_{121}H_{173}N_{48}O_{23}Si$• 2695,3685 $[M+H]^+$).



C_6C

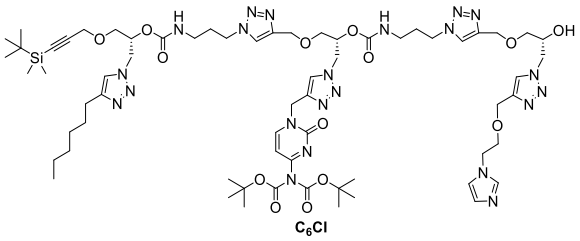
C_6C, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 7.94 – 7.84 (m, 2H), 7.74 (s, 1H), 7.35 (s, 1H), 7.03 (d, *J* = 7.4 Hz, 1H), 5.53 (t, *J* = 6.2 Hz, 1H), 5.21 (dd, *J* = 6.9, 4.5 Hz, 1H), 5.08 (s, 2H), 4.67 – 4.46 (m, 5H), 4.45 – 4.31 (m, 3H), 4.20 (d, *J* = 1.5 Hz, 2H), 4.15 (q, *J* = 6.1, 5.7 Hz, 1H), 3.85 (d, *J* = 5.8 Hz, 1H), 3.72 – 3.43 (m, 4H), 3.10 (dt, *J* = 36.1, 7.2 Hz, 2H), 2.64 (t, *J* = 7.8 Hz, 2H), 2.07 (dt, *J* = 13.5, 6.8 Hz, 2H), 1.61 (dd, *J* = 9.2, 5.8 Hz, 2H), 1.52 (s, 18H), 1.41 – 1.22 (m, 6H), 0.92 (s, 9H), 0.88 – 0.83 (m, 3H), 0.10 (s, 6H).**^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 162.69, 155.57, 155.06, 149.63, 148.64, 148.17, 144.69, 141.66, 125.79, 123.54, 121.79, 101.28, 96.80, 90.83, 85.10, 71.35, 71.16, 69.19, 68.27, 64.88, 59.62, 53.05, 50.38, 47.34, 45.53, 37.75, 31.68, 30.31, 29.55, 29.04, 27.81, 26.15, 25.75, 22.68, 16.56, 14.19, -4.57. **HRMS** *m/z*= 1010.5606 (calcd. for $C_{47}H_{75}N_{13}O_{10}Si$ 1010.5602 $[M+H]^+$).



30

30, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 7.92 – 7.76 (m, 3H), 7.35 (s, 1H),

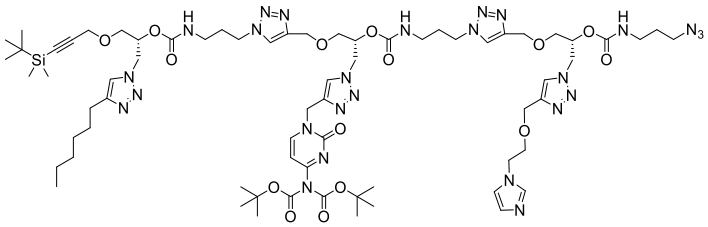
7.04 (d, *J* = 7.3 Hz, 1HHz, 1H), 5.71 (t, *J* = 6.2 Hz, 1H), 5.38 (t, *J* = 6.0 Hz, 1H), 5.22 (t, *J* = 5.7 Hz, 1H), 5.15 – 5.00 (m, 3H), 4.62 (dd, *J* = 34.8, 6.1 Hz, 6H), 4.41 (t, *J* = 6.5 Hz, 2H), 4.19 (d, *J* = 1.8 Hz, 2H), 3.70 – 3.01 (m, 10H), 2.66 (t, *J* = 7.8 Hz, 2H), 2.10 (dd, *J* = 10.5, 3.7 Hz, 2H), 1.82 – 1.69 (m, 2H), 1.61 (d, *J* = 7.3 Hz, 2H), 1.52 (s, 18H), 1.36 – 1.23 (m, 6H), 0.92 (s, 9H), 0.89 – 0.83 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 162.74, 155.54, 155.16, 149.60, 148.68, 148.21, 144.55, 141.66, 126.15, 123.56, 121.72, 101.24, 96.84, 90.86, 85.11, 71.16, 70.60, 68.27, 68.07, 65.01, 59.61, 50.32, 50.22, 47.35, 45.79, 38.63, 37.99, 37.79, 31.68, 30.23, 29.57, 29.10, 29.05, 27.80, 26.15, 25.78, 22.68, 16.55, 14.19, -4.58. **HRMS** *m/z*= 1136.6147 (calcd. for $C_{51}H_{81}N_{17}O_{11}Si$ 1136.6144 $[M+H]^+$).



C$_6$Cl

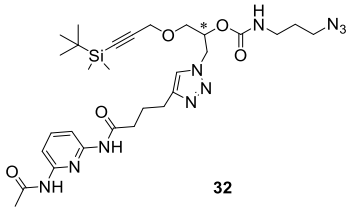
C$_6$Cl, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 7.91 – 7.84 (m, 2H), 7.80 (d, *J* = 5.2 Hz, 2H), 7.52 (s, 1H), 7.36 (s, 2H), 7.04 (d, *J* = 7.4 Hz, 1H), 6.89 (d, J = 21.4 Hz, 2H), 5.94 (dt, *J* = 12.2, 6.0 Hz, 2H), 5.22 (t, *J* = 5.6 Hz, 1H), 5.13 – 4.98 (m, 3H), 4.73 – 4.49 (m, 11H), 4.45 – 4.28 (m, 5H), 4.24 – 4.10 (m, 3H), 4.05 (dd, *J* = 6.1, 3.9 Hz, 2H), 3.84 – 3.73 (m, 2H), 3.67 – 3.44 (m, 6H), 3.10 (dq, *J* = 18.5, 6.9, 6.1 Hz, 4H), 2.65 (t, *J* = 7.8 Hz, 2H), 2.13 – 2.01 (m, 4H), 1.61 (p, *J* = 7.8, 7.1 Hz, 2H), 1.51 (s, 18H), 1.35 – 1.22 (m, 6H), 0.92 (s, 9H), 0.88 – 0.82 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 162.72, 155.59, 155.48, 155.18, 149.59, 148.66, 148.37, 144.77, 144.47, 144.16, 124.50, 123.69, 123.63, 121.79, 101.25, 96.90, 90.86, 85.20, 71.84, 71.03, 70.73, 69.08, 68.90, 68.26, 64.92, 64.20, 59.61, 53.63, 50.37, 47.53,

47.42, 45.77, 37.80, 31.68, 30.34, 30.27, 29.56, 29.04, 27.79, 26.15, 25.76, 22.69, 16.56, 14.21, -4.58. **HRMS** *m/z*= 1441.7629 (calcd. for $C_{65}H_{100}N_{22}O_{14}Si$ 1441.7631 $[M+H]^+$).



31

31, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 7.92 – 7.73 (m, 4H), 7.54 – 7.39 (m, 2H), 7.35 (s, 1H), 7.19 (s, 1H), 7.07 – 6.86 (m, 3H), 5.90 (dt, *J* = 71.8, 6.3 Hz), 5.33 – 4.89 (m, 5H), 4.79 – 4.32 (m, 16H), 4.30 – 4.02 (m, 4H, H_3 and H_{44}), 3.78 – 2.97 (m, 14H), 2.65 (t, *J* = 7.8 Hz, 2H), 2.20 – 1.93 (m, 7H), 1.66 – 1.57 (m, 2H), 1.51 (s, 18H), 1.36 – 1.19 (m, 8H), 0.92 (s, 9H), 0.88 – 0.82 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 162.72, 155.67, 155.57, 155.46, 155.09, 149.60, 148.68, 148.29, 145.08, 144.50, 141.65, 126.15, 123.71, 123.61, 121.74, 101.25, 96.82, 90.85, 85.14, 71.10, 70.28, 69.01, 68.27, 67.91, 64.95, 64.84, 59.60, 50.66, 50.33, 49.14, 47.47, 47.39, 45.74, 38.46, 37.82, 31.68, 30.27, 29.57, 29.10, 29.05, 27.79, 26.15, 25.77, 22.69, 16.56, 14.21, -4.58. **HRMS** *m/z*= 1567.8190 (calcd. for $C_{69}H_{106}N_{26}O_{15}Si$ 1567.8173 $[M+H]^+$).

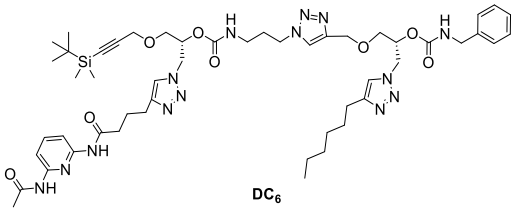


32

(*R*)-**32**, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.44 (br s, 1H), 8.10 – 7.75 (m, 3H), 7.67 (t, *J* = 8.1 Hz, 1H), 7.38 (s, 1H), 5.28 (quint, *J* = 5.2 Hz, 1H), 5.13 (t, *J* = 6.1 Hz, 1H), 4.59 (d, *J* = 5.6 Hz, 2H), 4.23 (s, 2H), 3.66 (d,

J = 4.9 Hz, 2H), 3.43 – 3.13 (m, 5H), 2.81 (t, *J* = 6.9 Hz, 2H), 2.51 – 2.27 (m, 2H), 2.19 (s, 3H), 2.08 (td, *J* = 6.9, 2.5 Hz, 2H), 1.81 – 1.66 (m, 2H), 0.93 (s, 9H), 0.11 (s, 6H).**^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 171.77, 168.85, 158.29, 155.32, 149.74, 147.21, 140.78, 122.41, 109.60, 109.48, 101.07, 91.09, 71.41, 68.37, 59.73, 50.76, 49.10, 38.73, 36.07, 28.99, 26.15, 25.22, 24.82, 24.09, 16.57, -4.58. **HRMS** *m/z* = 641.3337 (calcd. for $C_{29}H_{44}N_{10}O_5Si$ 641.3338 $[M+H]^+$).

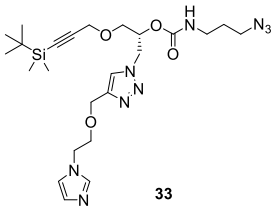
(*S*)-**32**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 8.60 (s, 1H), 8.17 (s, $1H_9$), 7.91 – 7.77 (m, 2H), 7.68 (t, *J* = 8.0 Hz, 1H), 7.39 (s, 1H), 5.37 – 5.12 (m, 2H), 4.59 (d, *J* = 5.6 Hz, 2H), 4.22 (s, 2H), 3.65 (d, *J* = 4.9 Hz, 2H), 3.33 (t, *J* = 6.5 Hz, 2H), 3.21 (q, *J* = 6.5 Hz, 2H), 2.90 – 2.71 (m, 2H), 2.39 (t, *J* = 8.0 Hz, 2H), 2.19 (s, 3H), 2.09 (q, *J* = 7.0 Hz, 2H), 1.73 (p, *J* = 6.5 Hz, 2H), 0.92 (s, 9H, H_1), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 171.96, 169.08, 161.28, 155.34, 149.62, 147.20, 140.93, 122.46, 109.47, 101.07, 91.08, 71.38, 68.35, 59.72, 50.75, 49.09, 38.72, 36.06, 28.98, 26.13, 25.19, 24.79, 24.09, 16.56, -4.59. **HRMS** *m/z*= 641.3338 (calcd. for $C_{29}H_{44}N_{10}O_5Si$ 641.3316 $[M+H]^+$).



DC$_6$

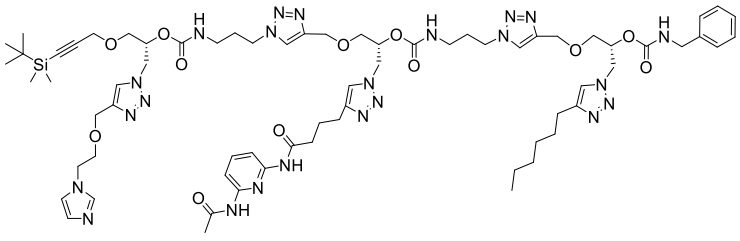
DC$_6$, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 8.64 (s, 1H), 8.37 (s, 1H), 7.84 (s, 2H), 7.73 (s, 1H), 7.64 (s, 1H), 7.40 (s, 1H), 7.34 – 7.19 (m, 6H), 5.63 (q, *J* = 6.2 Hz, 2H), 5.23 (dt, *J* = 27.2, 6.2 Hz, 2H), 4.68 – 4.50 (m, 6H), 4.42 – 4.17 (m, 6H), 3.78 – 3.47 (m, 4H), 3.26 – 2.90 (m, 2H), 2.87 – 2.54 (m, 4H), 2.34 (s, 2H), 2.16 (s, 3H), 2.00 (d, J = 7.4 Hz, 2H), 1.59 (d, J = 7.7 Hz, 2H), 1.37 – 1.21 (m, 6H), 0.92 (s, 9H, H_1), 0.89 – 0.81 (m, 3H), 0.11 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ*

171.93, 169.21, 155.52, 149.95, 149.80, 148.72, 147.26, 144.46, 140.64, 138.17, 128.82, 127.72, 127.56, 123.67, 122.54, 121.84, 109.57, 101.13, 91.05, 71.46, 71.30, 68.48, 64.90, 59.71, 50.89, 50.05, 47.39, 45.18, 37.83, 36.06, 31.67, 30.18, 29.49, 29.04, 26.14, 25.71, 25.21, 24.71, 24.23, 22.68, 16.55, 14.19, -4.58. **HRMS** *m/z*= 1039.5657 (calcd. for $C_{51}H_{74}N_{14}O_8Si$ 1039.5657 $[M+H]^+$).



33

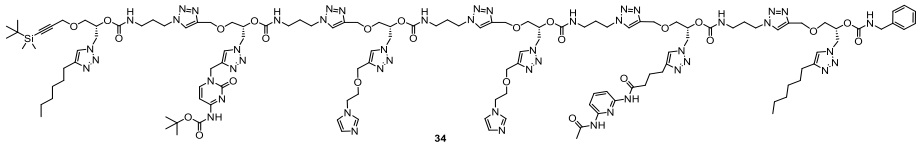
33, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 7.56 (s, 1H), 7.22 (s, 1H), 7.04 (s, 1H), 6.92 (dd, *J* = 6.7, 2.6 Hz, 1H), 6.77 (t, *J* = 5.9 Hz, 1H), 5.32 – 5.05 (m, 1H), 4.71 – 4.50 (m, 4H), 4.26 – 4.04 (m, 4H), 3.75 (ddd, *J* = 6.5, 4.2, 2.3 Hz, 2H), 3.62 – 3.48 (m, 2H), 3.41 – 3.21 (m, 4H), 1.81 (t, *J* = 6.7 Hz, 2H), 0.92 (s, 9H), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 155.55, 144.97, 128.43, 126.34, 123.51, 115.93, 101.35, 90.70, 70.61, 68.94, 67.87, 64.81, 59.59, 50.91, 50.80, 49.14, 47.48, 38.53, 29.10, 26.15, 16.56, -4.59. **HRMS** *m/z*= 546,2971 (calcd. for $C_{24}H_{39}N_9O_4Si$ 546,2967 $[M+H]^+$).



IDC_6

IDC_6, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.83 (d, *J* = 62.4 Hz, 2H), 8.08 (s, 1H), 7.79 (d, *J* = 35.8 Hz, 4H), 7.62 (t, *J* = 8.1 Hz, 1H), 7.48 (d, *J* = 7.0 Hz, 2H), 7.35 – 7.19 (m, 6H), 7.07 (d, *J* = 33.5 Hz, 2H), 6.93 (d, *J*

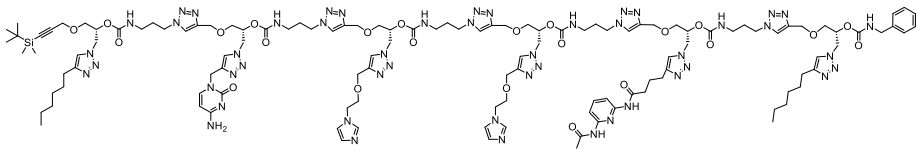
= 6.5 Hz, 1H), 6.30 (s, 1H), 5.80 (t, *J* = 6.0 Hz, 1H), 5.19 (d, *J* = 6.9 Hz, 3H), 4.70 – 4.49 (m, 12H), 4.44 – 4.11 (m, 10H), 3.82 – 3.44 (m, 8H), 3.27 – 2.95 (m, 4H), 2.85 – 2.53 (m, 4H), 2.48 – 2.26 (m, 2H), 2.16 (s, 3H), 2.04 (dtd, *J* = 32.1, 13.1, 12.6, 6.5 Hz, 6H), 1.74 – 1.48 (m, 2H), 1.37 – 1.18 (m, 6H), 0.91 (s, 9H), 0.88 – 0.81 (m, 3H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.19, 169.62,155.77, 155.56, 150.13, 149.95, 148.76,147.17, 144.51, 144.36, 144.20, 140.55, 128.80, 127.67, 127.54, 124.14, 123.94, 123.78, 121.93, 120.69, 109.58, 109.51, 90.82, 71.44, 71.14, 70.83, 68.63, 68.39, 68.17, 64.84, 64.71, 64.46, 59.60, 50.95, 50.12, 47.65, 45.10, 37.83, 36.06, 31.67, 30.34, 30.18, 29.50, 29.04, 26.15, 25.71, 25.06, 24.65, 22.68, 16.55, 14.21, -4.58. **HRMS** *m/z*= 1470.7658 (calcd. for $C_{69}H_{99}N_{23}O_{12}Si$ 1470.7686 $[M+H]^+$).



34

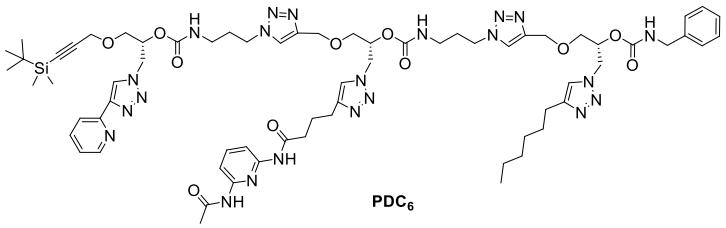
34, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 9.10 (d, *J* = 33.7 Hz, 2H), 7.92 – 7.69 (m, 9H), 7.61 (t, *J* = 7.3 Hz, 3H), 7.42 (d, *J* = 11.6 Hz, 3H), 7.34 (d, *J* = 8.1 Hz, 3H), 7.30 – 7.13 (m, 6H), 6.95 (d, *J* = 21.3 Hz, 4H), 6.37 (d, *J* = 25.7 Hz, 2H), 6.11 – 5.87 (m, 2H), 5.26 – 4.97 (m, 8H), 4.74 – 4.15 (m, 42H), 4.14 – 4.03 (m, 4H), 3.77 – 3.34 (m, 16H), 3.21 – 2.87 (m, 12H), 2.64 (dt, *J* = 15.3, 9.0 Hz,,4H), 2.34 (s, 2H), 2.14 (s, 3H), 2.12 – 1.88 (m, 12H), 1.59 (h, *J* = 8.0 Hz, 4H), 1.46 (d, J = 8.4 Hz, 9H), 1.35 – 1.18 (m, 12H), 0.90 (s, 9H), 0.87 – 0.81 (m, 6H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 169.57, 163.33, 155.88, 155.88, 155.86, 155.71, 155.62, 155.56, 155.56, 148.63, 147.17, 144.84, 144.23, 140.55, 138.34, 137.79, 128.75, 128.10, 127.60, 127.51, 125.86, 123.99, 123.86, 123.71, 122.82, 121.87, 121.74, 119.83, 109.56, 101.22, 90.85, 82.87, 71.43, 70.98, 70.62, 69.00, 68.36, 64.71, 64.60, 59.57, 50.65, 50.35, 50.10, 47.66, 47.53, 47.37, 45.89,

45.07, 37.83, 36.03, 31.65, 30.31, 30.20, 29.54, 29.48, 29.02, 28.13, 26.13, 25.74, 25.09, 24.60, 22.66, 16.54, 14.19, -4.59. **TOF MS ES +** *m/z*= 2822.6011 (calcd. for $C_{128}H_{184}N_{48}O_{25}Si$ 2822.4444 $[M+H]^+$).



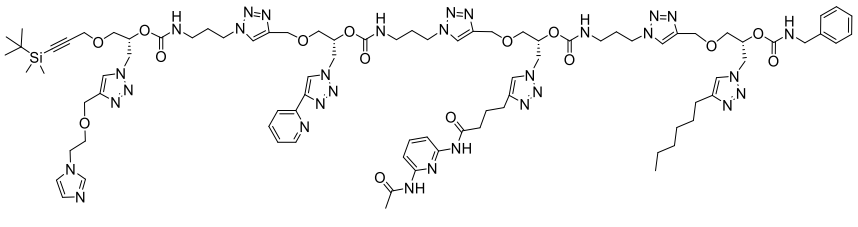
C_6CIIDC_6

C_6CIIDC_6, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.09 (s, 2H), 7.90 – 7.68 (m, 8H), 7.61 (dt, *J* = 16.1, 9.4 Hz, 3H), 7.52 (d, *J* = 6.9 Hz, 1H), 7.48 – 7.31 (m, 5H), 7.31 – 7.16 (m, 5H), 6.96 (d, *J* = 15.8 Hz, 6H) 6.54 (d, *J* = 31.6 Hz, 2H, H_{22} and H_{30}), 6.20 (s, 1H, H_{63}), 6.05 (s, 1H, H_6), 5.79 (d, *J* = 6.8 Hz, 1H), 5.33 – 4.79 (m, 8H),4.76 – 4.16 (m, 40H), 4.14 – 4.03 (m, 4H), 3.75 – 3.39 (m, 16H), 3.07 (d, *J* = 29.6 Hz, 10H), 2.63 (q, *J* = 8.2, 7.7 Hz, 4H), 2.33 (s, 2H), 2.13 (s, 3H), 2.11 – 1.87 (m, 12H), 1.58 (dt, *J* = 16.3, 8.0 Hz, 4H), 1.35 – 1.18 (m, 12H), 0.91 (s, 9H), 0.86 – 0.81 (m, 6H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.24, 171.33, 165.68, 156.15, 155.90, 155.78, 155.66, 150.02, 148.65, 147.18, 144.79, 144.21, 140.97, 140.47, 138.40, 137.75, 128.76, 127.92, 127.60, 127.51, 124.13, 123.94, 122.90, 121.94, 121.86, 119.93, 109.61, 101.26, 95.08, 90.89, 71.46, 71.00, 70.72, 69.04, 68.64, 68.25, 64.67, 64.52, 59.60, 50.63, 50.47, 50.33, 50.14, 47.68, 47.60, 47.40, 45.05, 38.61, 37.86, 31.67, 30.30, 29.78, 29.54, 29.04, 26.15, 25.74, 24.59, 22.68, 16.55, 14.21, -4.57. **TOF MS ES +** *m/z*= 2723.4556 (calcd. for $C_{122}H_{175}N_{49}O_{23}Si$ 2723,3873 $[M+H]^+$).



PDC$_6$

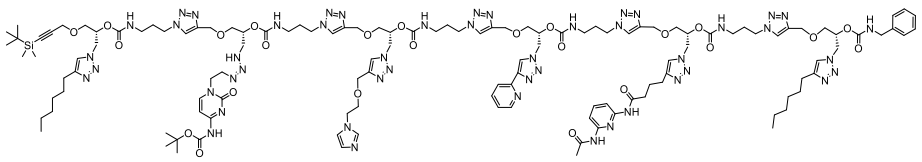
PDC$_6$, **^{1}H NMR (500 MHz, CDCl$_3$)** δ 8.83 (s, 1H), 8.67 (s, 1H), 8.49 (d, *J* = 4.9 Hz, 1H), 8.27 (s, 1H), 8.07 (d, *J* = 7.9 Hz, 1H), 7.88 – 7.69 (m, 5H), 7.61 (s, 1H, H_{40}), 7.49 (s, 1H), 7.33 (s, 1H), 7.31 – 7.16 (m, 6H), 6.29 (d, *J* = 6.4 Hz, 1H), 6.05 – 5.73 (m, 2H), 5.32 (dd, *J* = 7.6, 4.0 Hz, 1H), 5.18 (q, *J* = 6.9, 6.1 Hz, 2H), 4.79 – 4.47 (m, 10H), 4.43 – 4.13 (m, 8H), 3.73 – 3.39 (m, 6H), 3.20 – 2.94 (m, 4H), 2.79 – 2.53 (m, 4H), 2.35 (s, 2H), 2.15 (s, 3H), 2.03 (ddd, *J* = 29.4, 13.9, 7.0 Hz, 6H), 1.70 – 1.48 (m, 2H), 1.34 – 1.18 (m, 6H), 0.89 (s, 9H), 0.87 – 0.82 (m, 3H), 0.08 (s, 6H). **^{13}C NMR (126 MHz, CDCl$_3$)** δ 172.04, 169.45, 155.71, 155.54, 150.08, 149.43, 148.66, 148.41, 147.12, 144.37, 144.22, 140.47, 138.28, 137.12, 128.78, 127.65, 127.53, 123.89, 123.76, 123.12, 121.89, 120.33, 109.48, 101.16, 90.95, 71.47, 71.09, 68.58, 68.23, 64.88, 59.62, 59.08, 50.94, 50.07, 47.26, 45.11, 37.86, 37.71, 36.11, 31.66, 30.21, 29.49, 29.03, 26.13, 25.71, 24.66, 24.11, 22.67, 19.85, 14.19, -4.62. **HRMS** *m/z*= 1423.7313 (calcd. for $C_{68}H_{94}N_{22}O_{11}Si$ 1423.7314 $[M+H]^+$).



IPDC$_6$

I**PDC$_6$**, **^{1}H NMR (500 MHz, CDCl$_3$)** δ 8.75 (d, *J* = 48.8 Hz, 2H), 8.50 (d, *J* = 4.9 Hz, 1H), 8.24 (s, 1H), 8.04 (d, *J* = 7.9 Hz, 1H), 7.92 – 7.66 (m, 6H), 7.62 (t, *J* = 8.0 Hz, 1H), 7.55 – 7.36 (m, 2H), 7.34 – 7.12 (m,

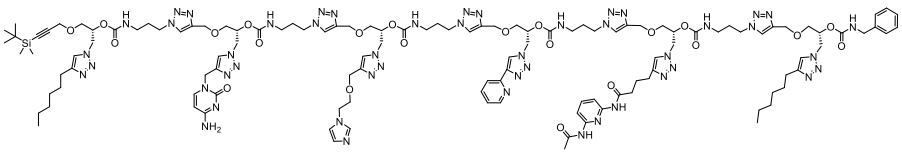
8H), 6.94 (d, *J* = 40.8 Hz, 2H), 6.23 (t, *J* = 7.0 Hz, 2H), 5.80 (t, *J* = 6.0 Hz, 1H), 5.32 – 5.06 (m, 4H), 4.75 – 4.47 (m, 15H), 4.41 – 4.02 (m, 12H), 3.76 – 3.44 (m, 10H), 3.27 – 2.92 (m, 6H), 2.79 – 2.54 (m, 4H), 2.33 (s, 3H), 2.15 (s, 3H), 2.13 – 1.90 (m, 8H), 1.58 (q, *J* = 7.5 Hz, 2H), 1.34 – 1.18 (m, 6H), 0.91 (s, 9H), 0.87 – 0.81 (m, 4H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.15, 169.52, 155.79, 155.57, 150.14, 149.99, 149.92, 149.34, 148.71, 147.16, 144.92, 144.39, 144.21, 144.17, 140.54, 138.27, 137.36, 128.79, 127.66, 127.53, 123.92, 123.77, 123.71, 123.23, 122.91, 121.92, 120.42, 109.58, 109.48, 101.31, 90.78, 71.41, 71.11, 70.98, 70.69, 69.04, 68.60, 68.39, 68.05, 64.83, 64.75, 64.64, 59.59, 50.84, 50.42, 50.09, 47.63, 47.38, 45.11, 37.83, 36.08, 31.66, 30.37, 30.17, 29.48, 29.03, 26.15, 25.70, 25.06, 24.62, 24.34, 22.68, 14.20, -4.58. **TOF MS ES +** *m/z*= 1855,0697 (calcd. for $C_{86}H_{119}N_{31}O_{15}Si$ 1854,9344 $[M+H]^+$).



35

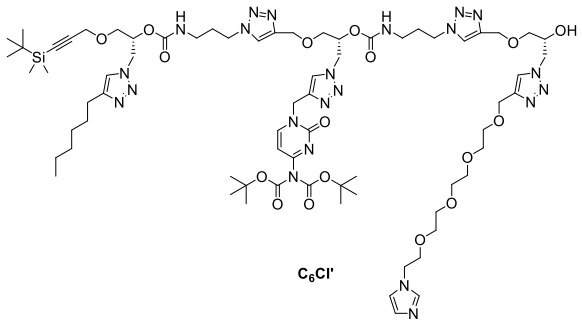
35, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 9.03 (dd, *J* = 59.8, 34.3 Hz, 2H), 8.47 (s, 1H), 8.30 (s, 1H), 8.13 – 7.42 (m, 15H), 7.38 – 7.13 (m, 10H), 7.11 – 6.98 (m, 2H), 6.67 – 6.16 (m, 3H), 6.12 – 5.83 (m, 2H), 5.37 – 4.92 (m, 8H), 4.80 – 4.12 (m, 42H), 3.78 – 3.37 (m, 14H, H_4, H_{12}, H_{20}, H_{28}, H_{36}, H_{44} and H_{69}), 3.24 – 2.95 (m, 10H), 2.80 – 2.48 (m, 6H), 2.34 (s, 2H), 2.08 (d, *J* = 51.0 Hz, 15H), 1.59 (q, *J* = 8.2 Hz, 4H), 1.49 (s, 9H), 1.35 – 1.18 (m, 12H), 0.90 (s, 9H, H_1), 0.84 (t, *J* = 3.3 Hz, 6H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.31, 169.33, 155.86, 155.83, 155.66, 155.64, 150.22, 149.99, 149.60, 149.36, 148.65, 148.62, 147.16, 144.28, 144.20, 144.19, 144.07, 144.01, 140.52, 140.48, 138.39, 137.41, 137.39, 128.73, 127.56, 127.52, 127.46, 124.05, 123.82, 123.30, 123.22, 123.11, 123.02, 121.97, 121.94,

121.85, 109.52, 109.43, 101.23, 90.86, 85.30, 85.27, 77.39, 77.14, 76.88, 71.42, 71.12, 70.93, 68.74, 68.52, 68.19, 64.68, 59.57, 50.46, 50.35, 50.14, 47.59, 47.45, 45.96, 45.94, 45.02, 37.80, 36.08, 31.64, 30.24, 29.80, 29.51, 29.01, 28.10, 27.75, 26.12, 25.71, 25.02, 24.55, 22.65, 14.19, -4.58, -4.59. **TOF MS ES +** *m/z*= 2776,6165 (calcd. for $C_{126}H_{178}N_{48}O_{24}Si$ 2776,4026 $[M+H]^+$).



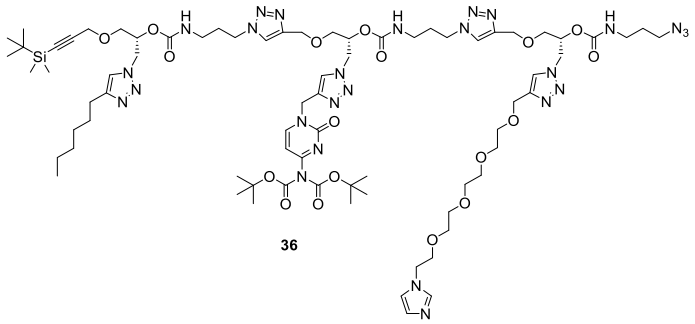
C_6ClPDC_6

C_6ClPDC_6, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.02 (d, *J* = 18.7 Hz, 2H), 8.48 (d, *J* = 4.9 Hz, 1H), 8.29 (s, 1H), 8.02 (d, *J* = 8.0 Hz, 1H), 7.89 – 7.57 (m, 11H), 7.55– 7.32 (m, 6H), 7.31 – 7.12 (m, 6H), 7.04 – 6.35 (m, 6H), 6.16 (s, 1H), 5.98 (s, 1H), 5.77 (d, *J* = 7.6 Hz, 1H), 5.29 – 4.84 (m, 8H), 4.77 – 4.02 (m, 42H), 3.73 – 3.38 (m, 14H), 3.06 (d, *J* = 34.4 Hz, 10H), 2.73 – 2.58 (m, 6H), 2.33 (s, 2H), 2.18 – 1.87 (m, 12H), 1.60 (q, *J* = 7.9 Hz, 4H), 1.35 – 1.18 (m, 15H), 0.91 (s, 9H), 0.87 – 0.81 (m, 6H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.25, 169.76, 155.89, 155.85, 155.82, 155.66, 155.62, 150.22, 149.98, 149.42, 148.64, 148.15, 147.17, 146.13, 144.31, 144.19, 144.09, 142.66, 140.48, 138.40, 137.28, 128.75, 127.73, 127.59, 127.49, 124.21, 124.08, 123.86, 123.79, 123.21, 122.96, 121.96, 121.88, 120.37, 119.96, 109.59, 109.50, 101.26, 90.89, 71.45, 71.12, 70.99, 70.78, 69.02, 68.74, 68.58, 68.23, 64.79, 64.70, 64.48, 59.60, 53.57, 50.84, 50.32, 50.15, 47.61, 47.42, 45.04, 37.88, 36.09, 31.66, 30.26, 29.82, 29.54, 29.49, 29.03, 26.14, 25.73, 25.10, 24.58, 22.67, 14.21, -4.57. **TOF MS ES +** *m/z*= 2676,3304 (calcd. for $C_{121}H_{170}N_{48}O_{22}Si$ 2676,3501 $[M+H]^+$).



C_6Cl'

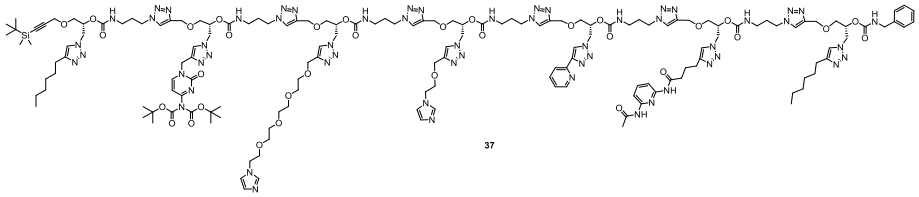
C_6Cl' , **^{1}H NMR (500 MHz, $CDCl_3$)** δ 7.92 – 7.74 (m, 5H), 7.51 (s, 1H), 7.36 (s, 1H), 7.06 – 6.88 (m, 3H), 5.94 (dd, *J* = 32.4, 6.2 Hz, 2H), 5.21 (d, *J* = 5.4 Hz, 1H), 5.13 – 4.97 (m, 3H), 4.71 – 4.51 (m, 11H), 4.39 (q, *J* = 6.9 Hz, 5H), 4.22 – 4.06 (m, 4H), 3.75 – 3.46 (m, 21H), 3.11 (ddq, J = 21.8, 14.4, 6.7 Hz, 4H), 2.65 (t, J = 7.8 Hz, 2H), 2.13 – 2.00 (m, 4H), 1.66 – 1.56 (m, 2H), 1.51 (s, 18H), 1.35 – 1.21 (m, 6H), 0.92 (s, 9H), 0.88 – 0.83 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 162.72, 155.59, 155.49, 155.20, 149.60, 148.66, 148.40, 144.79, 144.76, 144.49, 141.60, 128.67, 126.22, 124.55, 123.69, 123.63, 121.79, 119.77, 101.26, 96.91, 90.85, 85.20, 71.79, 71.04, 70.78, 70.66, 70.61, 70.49, 69.82, 69.10, 68.25, 64.92, 64.71, 59.60, 53.38, 50.36, 47.51, 47.42, 47.28, 45.77, 37.79, 31.68, 30.32, 29.56, 29.04, 27.79, 26.15, 25.75, 22.69, 16.56, 14.21, -4.58. **HRMS** *m/z*= 1573.8422 (calcd. for $C_{71}H_{112}N_{22}O_{17}Si$ 1573.8418 $[M+H]^+$).



36

36, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 7.94 – 7.74 (m, 4H), 7.66 (s, 1H), 7.56 (s, 1H), 7.36 (s, 1H), 7.07 – 6.96 (m, 3H), 6.04 (dt, *J* = 47.0, 6.0

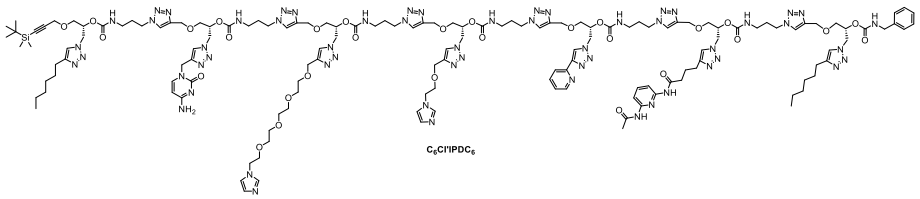
Hz, 2H), 5.86 (t, *J* = 6.1 Hz, 1H), 5.26 – 4.94 (m, 5H), 4.72 – 4.51 (m, 12H), 4.46 – 4.33 (m, 4H), 4.20 (d, J = 2.2 Hz, 2H), 4.10 (t, J = 5.1 Hz, 2H), 3.77 – 3.42 (m, 20H), 3.32 (t, J = 6.6 Hz, 2H), 3.23 – 2.99 (m, 6H), 2.65 (t, J = 7.8 Hz, 2H), 2.17 – 2.00 (m, 4H), 1.68 (dt, J = 62.8, 7.1 Hz, 4H), 1.51 (s, 18H), 1.35 – 1.23 (m, 6H), 0.92 (s, 9H), 0.88 – 0.83 (m, 3H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 162.72, 155.58, 155.50, 149.59, 148.67, 148.36, 145.17, 144.50, 144.38, 141.63, 137.67, 128.97, 126.20, 124.10, 123.78, 123.61, 121.76, 119.75, 101.25, 96.87, 90.85, 85.17, 71.08, 70.98, 70.75, 70.69, 70.67, 70.62, 70.58, 69.83, 68.47, 68.26, 64.94, 64.65, 59.60, 50.36, 49.05, 47.46, 47.40, 47.20, 45.77, 38.50, 37.79, 31.68, 30.27, 29.57, 29.04, 27.79, 26.14, 25.77, 22.68, 16.55, 14.20, -4.58. **HRMS** *m/z*= 1699.8971 (calcd. for $C_{75}H_{118}N_{26}O_{18}Si$ 1699.8959 $[M+H]^+$).



37

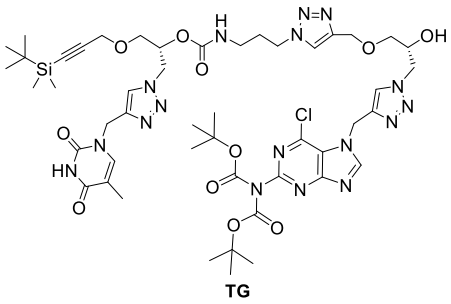
37, **^{1}H NMR (300 MHz, $CDCl_3$)** δ 8.98 (t, *J* = 42.6 Hz, 2H), 8.48 (d, *J* = 5.0 Hz, 1H), 8.27 (s, 1H), 8.03 (d, *J* = 7.9 Hz, 1H), 7.95 – 7.50 (m, 15H), 7.48 – 7.10 (m, 11H), 7.08 – 6.85 (m, 5H), 6.71 – 5.75 (m, 5H), 5.31 – 4.94 (m, 9H), 4.79 – 4.46 (m, 31H), 4.43 – 4.15 (m, 16H, H_3, H_9, H_{17}, H_{25}, H_{33}, H_{42}, H_{47} and H_{55}), 4.43 – 4.01 (m, 20H), 3.89 – 3.31 (m, 30H), 3.29 – 2.95 (m, 12H, H_7, H_{15}, H_{23}, H_{31}, H_{39} and H_{47}), 2.82 – 2.25 (m, 13H), 2.09 (d, *J* = 54.3 Hz, 16H), 1.58 (p, *J* = 8.3 Hz, 4H), 1.50 (s, 16H), 1.35 – 1.19 (m, 12H), 0.91 (s, 9H), 0.87 – 0.79 (m, 6H), 0.09 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** δ 169.59, 162.68, 155.88, 155.78, 155.60, 149.61, 149.43, 148.67, 148.31, 147.16, 145.02, 144.89, 144.37, 144.27, 144.20, 144.16, 141.68, 140.53, 140.49, 138.20, 137.61, 137.23, 128.77, 128.43, 127.62, 127.52, 124.36, 123.99, 123.96, 123.92, 123.86, 123.72, 123.18, 122.89, 121.90,

121.80, 120.32, 119.91, 119.78, 109.58, 109.48, 101.26, 96.86, 90.84, 85.20, 71.43, 71.04, 70.65, 70.62, 70.54, 70.43, 69.79, 69.08, 68.66, 68.24, 64.75, 64.60, 64.50, 59.58, 50.81, 50.67, 50.43, 50.32, 50.10, 47.66, 47.54, 47.45, 47.37, 47.27, 45.07, 37.82, 36.07, 31.66, 30.26, 29.81, 29.55, 29.48, 29.02, 28.13, 27.77, 26.13, 25.74, 25.71, 25.08, 24.61, 22.67, 16.54, 14.20, -4.59. **TOF MS ES +** *m/z*= 3439,0331 (calcd. for $C_{155}H_{223}N_{57}O_{33}Si$ 3439,7366 $[M+H]^+$).



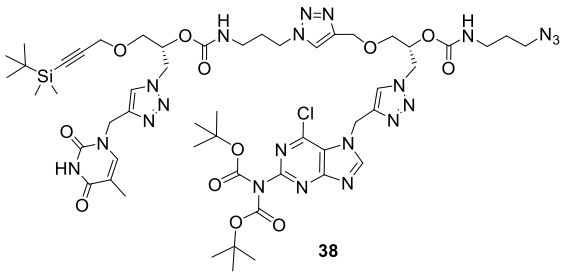
C_6Cl'IPDC$_6$, **^{1}H NMR (500 MHz, $CDCl_3$)** *δ* 9.11 (d, *J* = 20.3 Hz, 2H), 8.47 (d, *J* = 4.9 Hz, 1H), 8.30 (s, 1H), 8.02 (d, *J* = 8.0 Hz, 1H), 7.92 – 7.64 (m, 11H), 7.62 – 7.12 (m, 14H), 7.03 – 6.79 (m, 5H), 6.61 (d, *J* = 45.4 Hz, 4H), 6.13 (d, *J* = 68.9 Hz, 2H), 5.71 (d, *J* = 7.1 Hz, 1H), 5.29 – 4.78 (m, 7H), 4.90 (d, *J* = 35.3 Hz, 2H), , 4.76 – 4.43 (m, 32H), 4.42 – 4.16 (m, 16H), 4.07 (d, J = 6.1 Hz, 4H), 3.73 – 3.38 (m, 30H), 3.07 (dd, J = 13.6, 6.4 Hz, 12H), 2.63 (q, *J* = 11.2, 9.7 Hz, 8H), 2.06 (d, *J* = 58.1 Hz, 16H), 1.59 (d, *J* = 6.0 Hz, 4H), 1.35 – 1.19 (m, 12H), 0.91 (s, 9H), 0.84 (t, *J* = 6.5 Hz, 7H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 13C NMR (126 MHz, CDCl3) δ 155.79, 155.60, 150.05, 149.61, 149.43, 148.66, 145.02, 144.89, 144.37, 144.27, 137.23, 128.77, 128.43, 127.62, 127.52, 123.86, 123.74, 123.18, 121.90, 121.80, 120.32, 85.20, 71.43, 71.07, 70.65, 70.62, 70.60, 70.54, 70.43, 69.79, 69.08, 68.66, 68.24, 64.76, 64.50, 59.59, 50.67, 50.43, 50.10, 47.66, 47.54, 47.28, 45.07, 37.82, 36.07, 31.66, 30.25, 29.81, 29.55, 29.48, 29.02, 28.13, 27.77, 26.13, 25.74, 25.71, 25.08, 24.61, 22.67, 16.54, 14.20, -4.59. **TOF MS ES +** *m/z*= 3239.6485 (calcd. for

$C_{145}H_{207}N_{57}O_{29}Si$ 3239.6317 $[M+H]^+$).



TG

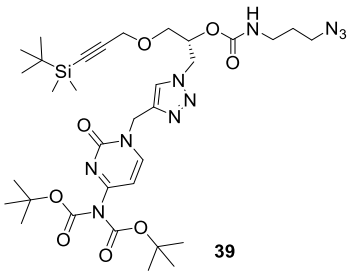
TG, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.69 (s, 1H), 8.38 (s, 1H), 7.95 (s, 1H), 7.78 (s, 1H), 7.69 (s, 1H), 7.30 (d, *J* = 1.5 Hz, 1H), 5.62 (t, *J* = 6.2 Hz, 1H), 5.55 (s, 2H), 5.18 (dd, *J* = 7.2, 4.0 Hz, 1H), 4.94 – 4.82 (m, 2H), 4.69 – 4.59 (m, 4H), 4.51 (dd, *J* = 14.0, 3.7 Hz, 1H), 4.32–4.41 (m, 3H), 4.25 – 4.16 (m, 2H), 3.70 – 3.57 (m, 3H), 3.51 – 3.46 (m, 1H), 3.08 (q, *J* = 6.3 Hz, 2H), 2.08 – 1.99 (m, 2H), 1.85 (d, *J* = 1.2 Hz, 3H), 1.46 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 164.49, 155.41, 152.55, 151.96, 151.31, 151.15, 150.98, 146.32, 144.54, 142.26, 141.09, 140.56, 130.09, 125.26, 125.10, 123.53, 111.28, 101.17, 90.97, 84.03, 71.60, 71.19, 69.15, 68.22, 64.64, 59.64, 53.19, 50.90, 47.39, 43.35, 39.41, 37.70, 28.48, 28.05, 26.16, 16.57, -4.57. **HRMS** *m/z*= 1122.4815 (calcd. for $C_{48}H_{68}ClN_{17}O_{11}Si$ 1122.4815 $[M+H]^+$).



38

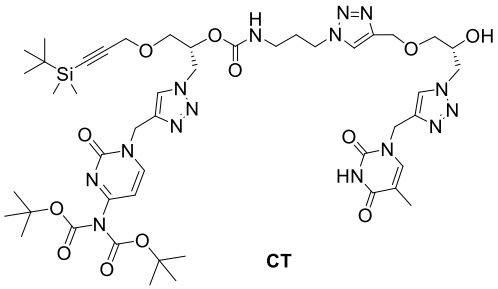
38, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.55 (s, 1H), 8.40 (s, 1H), 7.86 (s, 1H), 7.78 (s, 1H), 7.72 (s, 1H), 7.35 – 7.28 (m, 1H), 5.68 – 5.42 (m, 4H, H_6), 5.19 (s, 1H), 5.07 (s, 1H), 4.98 – 4.82 (m, 2H), 4.72 – 4.51

(m, 6H), 4.40 (t, *J* = 6.7 Hz, 2H), 4.27 – 4.14 (m, 2H), 3.73 – 3.42 (m, 4H), 3.31 – 3.22 (m, 2H), 3.18 – 2.98 (m, 4H), 2.06 – 1.96 (m, 2H), 1.87 (d, *J* = 1.1 Hz), 1.66 – 1.59 (m, 2H), 1.49 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 164.43, 155.38, 155.20, 152.59, 151.87, 151.39, 151.26, 151.14, 146.43, 144.45, 142.30, 140.46, 130.22, 125.21, 125.08, 123.56, 111.28, 101.19, 90.93, 84.28, 71.13, 70.95, 68.43, 68.18, 64.84, 59.62, 50.88, 49.01, 47.39, 43.30, 39.54, 38.49, 37.78, 30.27, 28.94, 28.47, 28.04, 26.15, 16.56, 12.43, -4.58. **HRMS** *m/z*= 1248.5367 (calcd. for $C_{57}H_{83}ClN_{19}O_{13}Si$ 1248.5356 $[M+Na]^+$).



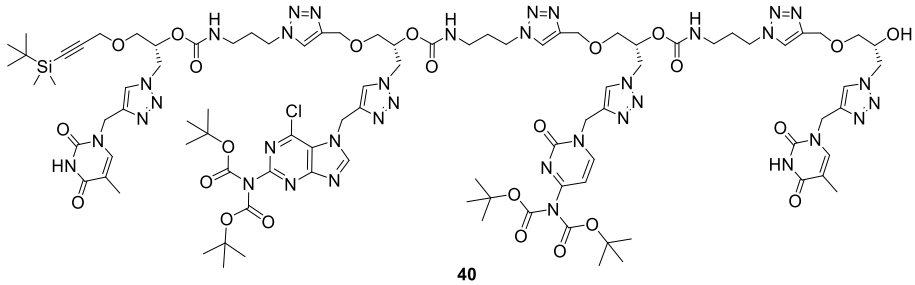
39

39, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 7.84 (t, *J* = 3.7 Hz, 2H), 7.05 (d, *J* = 7.4 Hz, 1H), 5.25 – 5.10 (m, 2H), 5.07 (d, *J* = 5.2 Hz, 2H), 4.71 – 4.55 (m, 2H), 4.20 (d, *J* = 4.0 Hz, 2H), 3.60 (d, *J* = 4.2 Hz, 2H), 3.43 – 3.32 (m, 4H), 1.83 – 1.73 (m, 2H), 1.54 (s, 18H), 0.92 (s, 9H), 0.11 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 162.73, 158.39, 155.11, 149.58, 148.01, 141.70, 125.76, 101.23, 96.89, 90.83, 85.15, 71.00, 68.00, 59.59, 50.66, 49.08, 38.65, 37.98, 29.08, 27.80, 26.15, 16.55, -4.59. **HRMS** *m/z*= 745.3810 (calcd. for $C_{33}H_{52}N_{10}O_8Si$ 745.3812 $[M+H]^+$).



CT

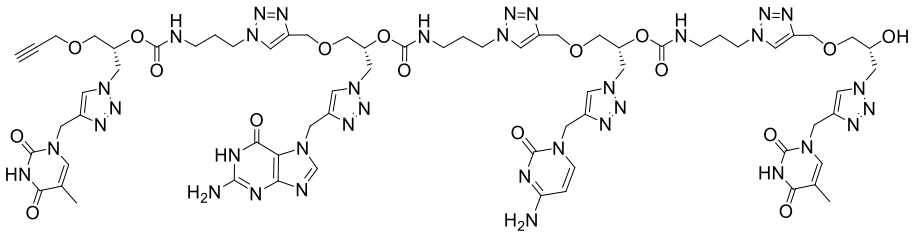
CT, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.36 (s, 1H), 7.94 – 7.84 (m, 3H), 7.79 (s, 1H), 7.37 – 7.31 (m, 1H), 7.04 (d, *J* = 7.4 Hz, 1H), 5.59 (t, *J* = 6.1 Hz, 1H), 5.20 – 5.01 (m, 3H), 4.93 (s, 2H), 4.70 – 4.49 (m, 5H), 4.47 – 4.34 (m, 3H), 4.26 – 4.13 (m, 3H), 3.71 – 3.60 (m, 2H), 3.58 – 3.45 (m, 2H), 3.21 – 2.92 (m, 2H), 2.05 (m, 2H), 1.88 (d, *J* = 1.2 Hz, 3H), 1.51 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 164.32, 162.74, 155.51, 155.27, 151.06, 149.60, 148.39, 144.50, 142.04, 141.68, 140.43, 126.10, 125.26, 123.72, 111.28, 101.25, 97.00, 90.88, 85.26, 71.49, 71.12, 69.25, 68.20, 64.78, 59.62, 53.11, 50.76, 47.45, 45.75, 43.15, 37.76, 30.33, 27.80, 26.16, 16.56, -4.57. **HRMS** *m/z*= 1064.5099 (calcd. for $C_{47}H_{69}N_{15}O_{12}Si$ 1064.5092 $[M+H]^+$).



40

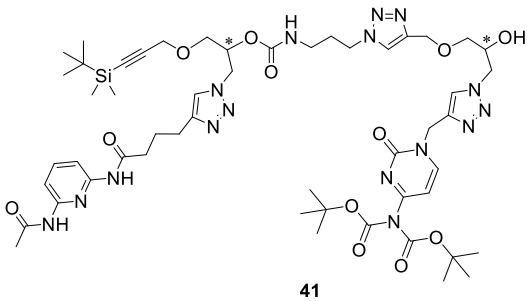
40, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 10.27 – 9.98 (m, 2H), 8.43 (s, 1H), 7.98 – 7.68 (m, 8H), 7.36 – 7.28 (m, 2H), 7.02 (d, *J* = 7.3 Hz, 1H), 6.31 – 5.94 (b, 3H), 5.61 – 5.34 (s, 2H), 5.23 – 4.80 (m, 9H), 4.58 (m, 13H), 4.44 – 4.27 (m, 7H), 4.25 – 4.09 (m, 3H), 3.74 – 3.41 (m, 8H),

3.16– 2.90 (m, 5H), 2.09 – 1.89 (m, 6H), 1.86 – 1.77 (m, 6H), 1.56 – 1.40 (m, 36H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 164.68, 162.68, 155.61, 155.49, 155.22, 152.64, 151.27, 151.14, 149.60, 148.60, 146.72, 144.53, 144.27, 144.18, 143.31, 142.30, 142.03, 141.69, 141.38, 140.64, 140.53, 126.03, 125.27, 125.14, 123.85, 111.09, 101.26, 96.90, 90.84, 85.21, 84.17, 71.57, 71.14, 69.17, 68.74, 68.33, 64.73, 59.59, 53.25, 50.97, 50.52, 47.49, 45.60, 43.16, 39.46, 37.83, 30.23, 29.82, 28.47, 28.03, 27.78, 26.15, 16.55, 12.40, -4.57. **TOF MS ES +** *m/z*= 2199.0055 (calcd. for $C_{93}H_{129}ClN_{36}O_{24}Si$ 2198.9545 $[M+H]^+$).



TGCT

TGCT , **^{1}H NMR (500 MHz, DMSO)** δ 11.30 (s, 2H), 11.06 (s, 1H), 8.19 – 7.85 (m, 7H), 7.78 – 7.48 (m, 3H, H_{34}, H_{43} and H_{52}), 7.46 – 7.32 (m, 2H), 7.11 (d, *J* = 124.8 Hz, 1H), 6.68 (s, 2H), 5.31 (d, *J* = 5.6 Hz, 1H), 5.21 (s, 2H), 5.09 (s, 3H), 4.88 (d, *J* = 4.2 Hz, 6H), 4.64 – 4.46 (m, 12H), 4.42 (dd, *J* = 13.9, 3.5 Hz, 1H), 4.36 – 4.20 (m, 7H), 4.17 (d, *J* = 2.4 Hz, 2H), 3.96 (s, 1H), 3.62 – 3.39 (m, 8H), 2.98 – 2.79 (m, 6H), 2.46 (t, *J* = 4.0, 2.0 Hz, 1H), 1.95 – 1.80 (m, 6H), 1.78 – 1.68 (m, 6H). **^{13}C NMR (126 MHz, DMSO)** δ 166.10, 166.01, 164.33, 155.60, 155.20, 151.23, 150.77, 145.90, 145.74, 143.67, 143.40, 142.96, 142.41, 142.13, 141.15, 141.01, 124.53, 124.45, 124.28, 124.13, 108.84, 93.62, 79.82, 77.68, 71.62, 70.43, 68.67, 68.21, 63.87, 57.92, 54.93, 52.90, 50.06, 48.61, 46.96, 42.15, 37.41, 29.95, 26.24, 11.98. **HRMS** *m/z*= 1665,6914 (calcd. for $C_{67}H_{84}N_{36}O_{17}$ 1665,6888 $[M+H]^+$).

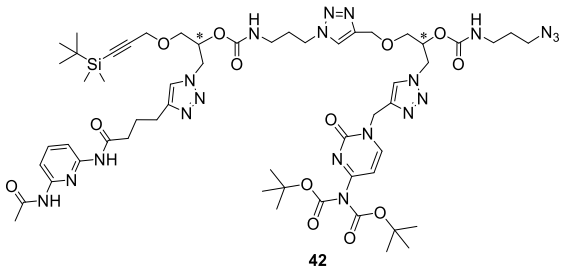


41

(*R*)-**41**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.81 (s, 1H), 8.55 (s, 1H), 8.00 – 7.58 (m, 6H), 7.46 (d, *J* = 7.0 Hz, 1H), 7.04 (d, *J* = 7.3 Hz, 1H), 5.64 (b, 1H), 5.26 (d, *J* = 5.9 Hz, 1H), 5.10 (s, 2H), 4.70 – 4.31 (m, 8H), 4.29 – 4.08 (m, 3H), 3.77 – 3.60 (m, 2H), 3.57 – 3.42 (m, 2H), 3.22 – 2.94 (m, 2H), 2.80 – 2.64 (m, 2H), 2.39 (s, 2H), 2.20 (s, 3H), 2.09 – 1.93 (m, 4H), 1.51 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.08, 169.25, 162.70, 155.59, 155.19, 149.67, 148.28, 147.31, 144.91, 144.67, 144.18, 141.69, 140.75, 125.73, 123.69, 122.63, 109.53, 101.19, 96.85, 90.98, 85.28, 71.43, 71.24, 69.17, 68.50, 64.81, 59.70, 53.10, 50.90, 47.45, 45.68, 37.83, 36.17, 30.27, 27.80, 26.15, 25.15, 24.72, 24.41, 16.57, -4.56. **HRMS** *m/z*= 1145.5678 (calcd. for $C_{52}H_{76}N_{16}O_{12}Si$ 1145.5671 $[M+H]^+$).

(*S*)-**41**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.07 – 8.43 (m, 2H), 7.94 – 7.59 (m, 6H), 7.52 – 7.40 (m, 1H), 7.05 (d, *J* = 7.4 Hz), 5.77 (b, 1H), 5.37 – 5.21 (m, 1H), 5.10 (s, 2H), 4.67 – 4.29 (m, 8H), 4.28 – 4.10 (m, 3H), 3.77 – 3.58 (m, 2H), 3.57 – 3.40 (m, 2H), 3.25 – 2.93 (m, 2H), 2.85 – 2.66 (m, 2H), 2.40 (s, 2H), 2.20 (s, 3H), 2.10 – 1.89 (m, 4H), 1.51 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 171.99, 169.21, 162.71, 155.61, 155.25, 149.66, 148.33, 147.29, 144.82, 144.64, 144.29, 141.64, 140.79, 125.72, 123.69, 122.67, 109.49, 101.21, 96.90, 90.95, 85.29, 71.47, 71.21, 69.15, 68.48, 64.78, 59.68, 53.12, 50.90, 47.47, 45.68, 37.83, 36.17, 30.26, 27.79, 26.15, 25.10, 24.69, 24.43, 16.57, -4.57. **HRMS** *m/z*= 1145.5672 (calcd. for

$C_{52}H_{76}N_{16}O_{12}Si$ 1145.5671 $[M+H]^+$).

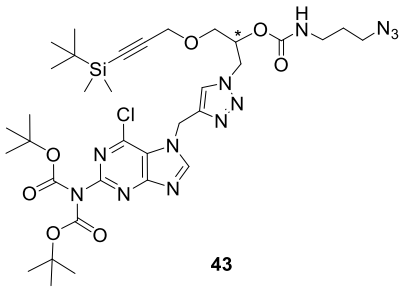


42

(*R*)-**42**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.92 – 8.38 (m, 2H), 7.97 – 7.64 (m, 6H), 7.44 (s, 1H, H_{19}), 7.07 (d, *J* = 7.3 Hz, 1H), 5.92 – 5.54 (b, 2H), 5.26 (d, *J* = 6.2 Hz, 1H), 5.17 – 5.05 (m, 3H), 4.68 – 4.51 (m, 6H), 4.39 (t, *J* = 6.6 Hz, 2H), 4.22 (d, *J* = 2.3 Hz, 2H), 3.80 – 3.58 (m, 2H), 3.54 – 3.39 (m, 2H), 3.33 (t, *J* = 6.6 Hz, 2H), 3.25 – 3.10 (m, 3H), 3.08 – 2.96 (m, 1H), 2.88 – 2.67 (m, 2H), 2.38 (s, 2H), 2.20 (s, 3H), 2.13 – 1.96 (m, 6H), 1.78 – 1.70 (m, 2H), 1.52 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.27, 169.45, 162.75, 155.56, 155.33, 155.25, 149.92, 149.66, 148.34, 148.17, 147.28, 144.44, 141.73, 140.72, 125.92, 123.74, 122.61, 109.51, 101.17, 96.93, 91.00, 85.28, 71.23, 70.64, 68.47, 68.13, 64.88, 59.69, 50.86, 49.05, 47.42, 45.77, 38.57, 37.84, 36.08, 30.20, 29.05, 27.79, 26.14, 25.23, 24.72, 24.31, 16.56, -4.58. **HRMS** *m/z*= 1271.6209 (calcd. for $C_{56}H_{82}N_{20}O_{13}Si$ 1271.6212 $[M+H]^+$).

(*S*)-**42**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.13 – 8.20 (m, 2H), 7.97 – 7.59 (m, 6H), 7.46 (d, *J* = 6.8 Hz, 1H), 7.07 (t, *J* = 8.3 Hz, 1H), 5.93 – 5.55 (b, 2H), 5.26 (d, *J* = 5.6 Hz, 1H), 5.18 – 4.99 (m, 3H), 4.68 – 4.50 (m, 6H), 4.40 (t, *J* = 6.6 Hz, 2H), 4.30 – 4.16 (m, 2H), 3.80 – 3.60 (m, 2H), 3.56 – 3.39 (m, 2H), 3.33 (t, *J* = 6.6 Hz, 2H), 3.25 – 3.11 (m, 3H), 3.02 (dd, *J* = 14.0, 6.5 Hz, 1H), 2.85– 2.67 (m, 2H), 2.40 (s, 2H), 2.21 (s, 3H), 2.14 – 1.94 (m, 4H), 1.74 (p, *J* = 6.6 Hz, 2H), 1.52 (s, 18H), 0.92 (s, 9H), 0.11 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.29, 169.78,

162.54, 155.67, 155.63, 155.19, 150.28, 150.11, 149.64, 148.63, 146.48, 144.57, 141.82, 140.20, 126.32, 124.18, 123.55, 109.30, 101.32, 96.85, 90.62, 85.08, 71.42, 70.84, 69.04, 68.08, 64.76, 59.54, 53.00, 50.76, 47.43, 45.01, 37.65, 36.11, 35.97, 30.17, 28.08, 27.74, 26.11, 25.09, 24.67, 24.38, 16.50, -4.62. **HRMS** *m/z*= 1271.6187 (calcd. for $C_{56}H_{82}N_{20}O_{13}Si$ 1271.6212 $[M+H]^+$).

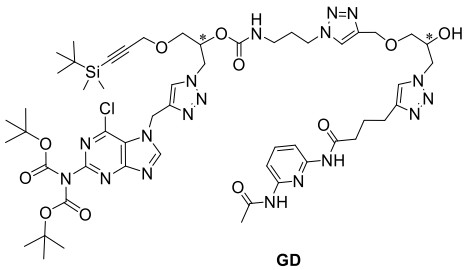


43

(*R*)-**43**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* δ 8.30 (s, 1H), 7.81 (s, 1H), 5.53 (s, 2H), 5.28 (t, *J* = 6.0 Hz, 1H), 5.09 (dd, *J* = 8.4, 4.2 Hz, 1H), 4.69 – 4.41 (m, 2H), 4.18 (d, *J* = 3.6 Hz, 2H), 3.69 – 3.54 (m, 2H), 3.23 (t, *J* = 6.5 Hz, 2H), 3.13 – 2.89 (m, 2H), 1.62 – 1.54 (m, 2H), 1.51 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 155.09, 152.57, 151.93, 151.56, 151.31, 146.06, 141.39, 130.31, 125.02, 101.20, 90.84, 84.35, 71.18, 68.22, 59.63, 51.10, 49.04, 39.79, 38.49, 28.95, 28.07, 26.15, 16.55, -4.57. **HRMS** *m/z*= 803.3535 (calcd. for $C_{34}H_{51}ClN_{12}O_7Si$ 803.3534 $[M+H]^+$)

(*S*)-**43**, **^{1}H NMR (300 MHz, $CDCl_3$)** *δ* 8.30 (s, 1H), 7.81 (s, 1H), 5.53 (s, 2H), 5.27 (t, *J* = 6.0 Hz, 1H), 5.09 (dd, *J* = 8.1, 4.0 Hz, 1H), 4.72 – 4.45 (m, 2H), 4.18 (d, *J* = 3.6 Hz, 2H), 3.63 (dd, *J* = 9.0, 4.6 Hz, 2H), 3.30 – 3.19 (m, 2H), 3.12 – 2.90 (m, 2H), 1.83 – 1.68 (m, 2H), 1.51 (s, 18H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (75 MHz, $CDCl_3$)** *δ* 155.09, 152.57, 151.92, 151.56, 151.31, 146.06, 141.38, 130.32, 125.02, 101.19, 90.84, 84.35, 71.18, 68.22, 59.63, 51.10, 49.04, 39.79, 38.49, 28.95, 28.07, 26.15, 16.56, -4.57. **HRMS** *m/z*= 803.3534 (calcd. for

$C_{34}H_{51}ClN_{12}O_7Si$ 803.3534 $[M+H]^+$)

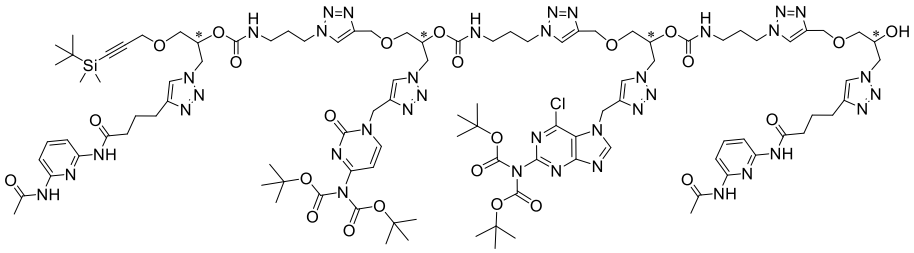


GD

(*R*, *R*)-**GD**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.95 – 8.42 (m, 2H), 8.38 (s, 1H), 7.90 – 7.74 (m, 3H), 7.70 – 7.62 (m, 2H), 7.50 (s, 1H), 5.76 (t, *J* = 6.0 Hz, 1H), 5.52 (d, *J* = 4.5 Hz, 2H), 5.15 (dd, *J* = 8.3, 4.1 Hz, 1H), 4.70 – 4.09 (m, 11H), 3.72 – 3.41 (m, 4H), 2.97 – 2.86 (m, 2H), 2.74 (t, *J* = 7.0 Hz, 2H), 2.40 (t, *J* = 7.2 Hz, 2H), 2.20 (s, 3H), 2.04 (p, *J* = 7.0 Hz, 2H), 1.97 – 1.82 (m, 2H), 1.49 (s, 18H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.23, 169.35, 155.45, 152.59, 151.90, 151.34, 151.31, 149.77, 149.58, 146.82, 146.55, 144.53, 141.31, 140.90, 130.16, 125.19, 123.44, 123.18, 109.51, 109.48, 101.12, 90.95, 84.45, 71.37, 71.17, 69.27, 68.30, 64.70, 59.63, 52.96, 51.18, 47.45, 39.69, 37.76, 36.30, 30.17, 28.06, 26.14, 25.11, 24.77, 24.28, 16.55, -4.57. **HRMS** *m/z*= 1225.5204 (calcd. for $C_{53}H_{75}ClN_{18}O_{11}Si$ 1225.5213 $[M+Na]^+$)

(*S, S*)-**GD**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.13 – 8.47 (m, 2H), 8.39 (s, 1H), 7.94 – 7.73 (m, 3H), 7.69 – 7.62 (m, 2H), 7.51 (s, 1H), 5.85 – 5.74 (t, 1H), 5.53 (d, *J* = 3.4 Hz, 2H), 5.15 (dd, *J* = 7.9, 3.9 Hz, 1H), 4.74 – 4.06 (m, 11H), 3.71 – 3.43 (m, 4H), 2.99 – 2.86 (m, 2H), 2.74 (t, *J* = 6.9 Hz, 2H),), 2.40 (t, *J* = 7.2 Hz, 2H), 2.20 (s, 3H), 2.04 (q, *J* = 7.2 Hz, 2H), 1.89 (p, *J* = 27.5, 6.8 Hz, 2H), 1.49 (s, 18H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.14, 169.55, 155.45, 152.58, 151.87, 151.33, 151.29, 149.69, 149.48, 146.79, 146.59, 144.50, 141.30, 141.01, 130.14, 125.21, 123.47, 123.22, 109.48,

109.44, 101.11, 90.93, 84.46, 71.40, 71.15, 69.24, 68.28, 64.67, 59.63, 52.98, 51.17, 47.45, 39.67, 37.74, 36.26, 30.16, 28.05, 26.13, 25.07, 24.75, 24.27, 16.54, -4.58. **HRMS** *m/z*= 1203.5392 (calcd. for $C_{53}H_{75}ClN_{18}O_{11}Si$ 1203.5393 $[M+H]^+$)

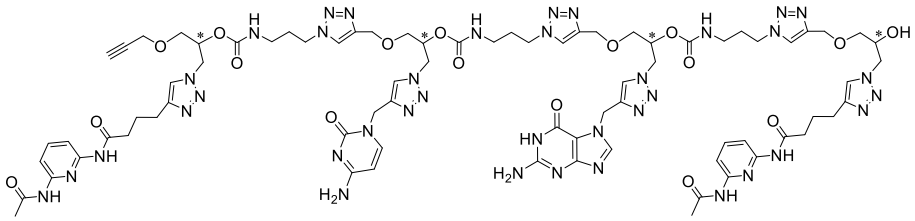


44

(*R*)-**44**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.48 – 8.48 (m, 4H), 8.41 (s, 1H, H_{51}), 8.01 – 7.57 (m, 12H), 7.51 (s, 1H, H_{54}), 7.44 (s, 1H, H_{32}), 7.08 – 6.99 (m, 1H), 6.38 – 5.91 (b, 3H), 5.61 – 4.91 (m, 7H), 4.87 – 4.02 (m, 25H), 3.77 – 3.41 (m, 8H), 3.30 – 2.86 (m, 6H), 2.76 – 2.64 (m, 4H), 2.43 – 1.87 (m, 20H), 1.59 – 1.38 (m, 36H), 0.91 (s, 9H, H_1), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.31, 169.26, 161.91, 155.85, 155.76, 155.73, 155.11, 152.56, 152.00, 151.12, 150.94, 149.69, 148.80, 147.28, 147.09, 146.71, 144.54, 144.17, 144.12, 141.90, 141.37, 125.68, 125.19, 123.53, 123.44, 123.32, 122.71, 109.54, 109.49, 101.30, 96.84, 90.86, 84.16, 83.25, 77.00, 71.12, 71.06, 71.01, 70.91, 70.88, 69.11, 68.69, 68.49, 64.85, 64.78, 64.64, 59.65, 53.07, 50.94, 50.89, 50.66, 50.64, 47.59, 45.94, 37.83, 37.79, 36.15, 36.08, 30.20, 30.15, 29.82, 28.13, 28.02, 26.15, 25.17, 25.01, 24.64, 24.51, 16.55, -4.56. **TOF MS ES +** *m/z*= 2360.1295 (calcd. for $C_{103}H_{143}ClN_{38}O_{24}Si$ 2360.0668 $[M+H]^+$).

(*S*)-**44**, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.90 9.42 – 8.42 (m, 4H), 8.41 (s, 1H, H_{51}), 7.98 – 7.39 (m, 14H), 7.04 (d, *J* = 7.4 Hz, 1H), 6.41– 5.90 (b, 3H), 5.52 (s, 2H), 5.30 – 5.21 (m, 1H), 5.16 – 4.93 (m, 4H), 4.77 –

4.11 (m, 25H), 3.80 – 3.36 (m, 8H), 3.22 – 2.87 (m, 6H), 2.77 – 2.63 (m, 4H), 2.48 – 1.86 (m, 20H), 1.52 – 1.42 (m, 36H), 0.91 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.25, 169.71, 162.73, 155.71, 155.64, 155.59, 155.30, 152.63, 151.86, 151.23, 151.15, 149.67, 148.60, 147.28, 146.82, 146.71, 144.50, 144.21, 144.12, 141.72, 141.28, 125.87, 125.27, 123.89, 123.69, 123.21, 122.66, 109.50, 109.46, 101.19, 96.99, 90.97, 85.37, 84.31, 71.54, 71.44, 71.15, 70.88, 70.86, 69.17, 68.72, 68.49, 64.73, 64.63, 64.53, 59.67, 53.03, 50.96, 50.89, 50.55, 50.52, 47.51, 45.67, 39.52, 37.88, 37.77, 36.25, 36.12, 30.24, 30.10, 29.82, 28.03, 27.77, 26.14, 25.16, 25.06, 24.66, 24.43, 16.55, -4.57. **TOF MS ES +** *m/z*= 2360.1712 (calcd. for $C_{103}H_{143}ClN_{38}O_{24}Si$ 2360.0668 $[M+H]^+$).

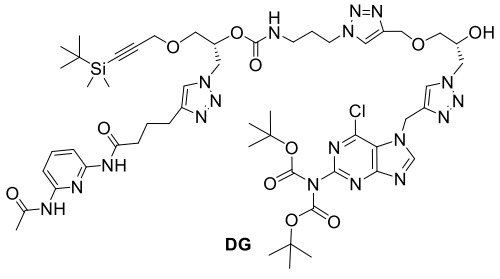


DCGD

(*R, R, R, R*)-**DCGD**, **^{1}H NMR (500 MHz, DMSO)** *δ* 11.10 (s, 1H), 10.03 (dd, *J* = 24.4, 7.6 Hz, 4H), 8.13 – 8.06 (m, 3H), 7.95 (d, *J* = 14.5 Hz, 2H), 7.78 (s, 2H), 7.74 – 7.64 (m, 8H), 7.40 (dq, *J* = 15.9, 5.6 Hz, 3H), 7.22 (s, 2H), 6.69 (s, 2H), 5.68 (d, *J* = 7.2 Hz, 1H), 5.21 (s, 2H), 5.09 (tt, *J* = 8.4, 4.9 Hz, 3H), 4.93 – 4.83 (m, 2H), 4.62 – 4.47 (m, 14H), 4.39 (dd, *J* = 13.9, 3.7 Hz, 1H), 4.31 (dd, *J* = 7.2, 5.0 Hz, 6H), 4.18 (d, *J* = 2.5 Hz, 2H), 4.10 (d, *J* = 5.5 Hz, 1H), 3.60 – 3.40 (m, 8H), 2.97 – 2.84 (m, 6H), 2.63 (td, *J* = 7.7, 5.2 Hz, 4H), 2.45 (q, *J* = 7.9 Hz, 5H), 2.09 (s, 6H), 1.88 (q, *J* = 7.5, 6.8 Hz, 10H). **^{13}C NMR (126 MHz, DMSO)** *δ* 172.17, 169.67, 166.31, 157.51, 155.49, 153.92, 151.43, 150.51, 150.44, 146.66, 146.31, 146.15, 143.96, 143.65, 143.61,

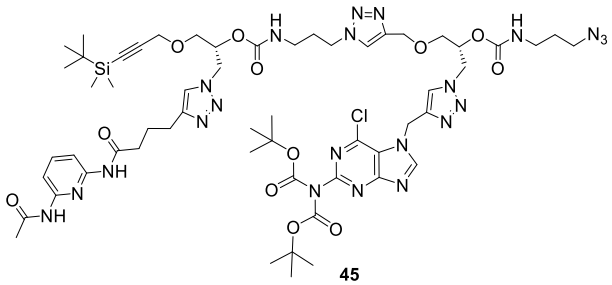
143.12, 142.80, 140.22, 137.53, 124.78, 124.64, 124.38, 123.12, 123.07, 109.19, 109.08, 94.22, 80.09, 77.80, 71.81, 70.77, 68.84, 68.55, 68.47, 64.01, 58.14, 52.89, 50.28, 50.06, 48.85, 47.21, 43.65, 37.61, 35.83, 35.73, 30.14, 25.04, 24.69, 24.65, 24.20.

(*S, S, S, S*)-**DCGD**, **^{1}H NMR (500 MHz, DMSO)** δ 11.03 (s, 1H), 10.03 (dd, *J* = 24.2, 7.5 Hz, 4H), 8.14 – 8.05 (m, 3H), 7.95 (d, *J* = 14.7 Hz, 2H), 7.80 (d, *J* = 12.8 Hz, 2H), 7.75 – 7.60 (m, 8H), 7.41 (dt, *J* = 16.9, 5.2 Hz, 3H), 7.22 (s, 2H), 6.66 (s, 2H), 5.76 (s, 1H), 5.68 (d, *J* = 7.2 Hz, 1H), 5.29 (d, *J* = 5.5 Hz, 1H), 5.21 (s, 2H), 5.09 (tt, *J* = 8.7, 4.7 Hz, 3H), 4.95 – 4.81 (m, 2H), 4.67 – 4.44 (m, 12H), 4.39 (dd, *J* = 13.9, 3.7 Hz, 1H), 4.31 (q, *J* = 6.9 Hz, 6H), 4.24 – 4.14 (m, 3H), 4.10 (d, *J* = 5.1 Hz, 2H), 3.96 (s, 1H), 3.64 – 3.44 (m, 6H), 3.40 (td, *J* = 11.9, 10.9, 5.5 Hz, 2H), 2.91 (td, *J* = 14.2, 12.2, 6.5 Hz, 6H), 2.63 (td, *J* = 7.7, 5.4 Hz, 5H), 2.44 (t, *J* = 8.0 Hz, 4H), 2.09 (s, 6H), 1.94 – 1.80 (m, 10H). **^{13}C NMR (126 MHz, DMSO)** δ 171.87, 169.28, 166.08, 157.19, 155.20, 153.71, 151.19, 150.36, 150.27, 146.38, 146.02, 146.00, 145.87, 143.70, 143.70, 143.39, 142.97, 142.66, 139.86, 137.14, 124.47, 124.31, 124.12, 122.80, 122.76, 109.00, 108.87, 93.82, 79.84, 77.67, 71.66, 70.52, 68.64, 68.33, 68.23, 63.85, 57.91, 52.69, 49.99, 49.78, 48.61, 46.95, 43.39, 37.41, 35.57, 35.48, 29.96, 24.86, 24.54, 24.50, 24.00.



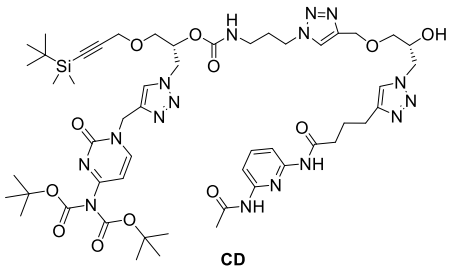
DG, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.14 (d, *J* = 105.7 Hz, 2H), 8.43 (s, 1H), 7.99 (s, 1H), 7.80 (d, *J* = 9.4 Hz, 3H), 7.65 – 7.51 (m, 2H), 6.14 (t,

J = 6.1 Hz, 1H), 5.51 (s, 2H), 5.27 – 5.14 (m, 1H), 4.84 (d, *J* = 5.8 Hz, 1H), 4.66 – 4.07 (m, 11H), 3.72 – 3.40 (m, 4H), 3.23 – 3.16 (m, 3H), 2.75 – 2.62 (m, 2H), 2.37 (q, *J* = 7.1 Hz, 2H), 2.16 (s, 3H), 2.07 – 1.91 (m, 4H), 1.43 (s, 19H), 0.90 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.23, 169.58, 155.64, 152.58, 151.93, 151.12, 150.93, 150.07, 147.23, 146.57, 144.44, 140.82, 140.50, 129.97, 125.23, 123.82, 122.88, 109.46, 101.20, 90.91, 84.00, 71.53, 71.20, 69.09, 68.45, 64.67, 59.02, 53.26, 50.89, 47.47, 39.32, 37.77, 36.20, 30.27, 28.00, 26.13, 25.12, 24.70, 24.49, 16.54, -4.59. **HRMS** *m/z*= 1203.5397 (calcd. for $C_{53}H_{75}ClN_{18}O_{11}Si$ 1203.5393 $[M+H]^+$)



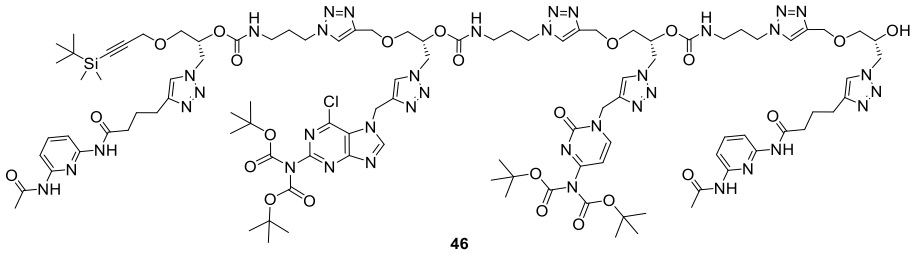
45

45, ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.61 (s, 1H), 8.40 (d, *J* = 7.8 Hz, 2H), 7.84 (s, 3H), 7.71 (s, 1H), 7.65 (s, 1H), 7.42 (s, 1H), 5.61 (t, *J* = 6.0 Hz, 1H), 5.55 (dd, *J* = 10.1, 4.6 Hz, 2H), 5.28 (t, *J* = 5.8 Hz, 1H), 5.10 (s, 1H), 4.74 – 4.11 (m, 10H), 3.86 – 3.37 (m, 4H), 3.32 – 2.87 (m, 6H), 2.74 (d, *J* = 9.2 Hz, 2H), 2.36 (d, *J* = 4.7 Hz, 2H), 2.20 (s, 3H), 2.02 (t, *J* = 7.7 Hz, 4H), 1.62 (d, *J* = 1.4 Hz, 2H), 1.48 (s, 19H), 0.92 (s, 9H), 0.10 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 171.92, 169.21, 155.52, 155.27, 152.59, 151.91, 151.41, 151.29, 149.96, 147.29, 146.41, 144.30, 141.36, 140.67, 130.19, 125.12, 123.62, 122.57, 109.61, 101.10, 91.10, 84.33, 71.31, 70.97, 68.45, 68.38, 64.79, 59.74, 50.87, 49.02, 47.42, 39.59, 38.50, 37.85, 36.07, 30.30, 28.94, 28.08, 28.04, 26.15, 25.22, 24.77, 24.27, 16.57, -4.57. **HRMS** *m/z*= 1329.5743 (calcd. for $C_{57}H_{81}ClN_{22}O_{12}Si$ 1329.3935 $[M+H]^+$)



CD

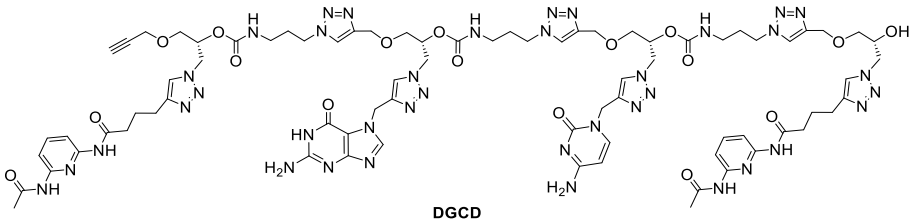
CD, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.26 (d, *J* = 18.1 Hz, 2H), 8.12 (s, 1H), 8.01 (t, *J* = 3.7 Hz, 2H), 7.80 (t, *J* = 7.0 Hz, 2H), 7.70 – 7.51 (m, 2H), 6.98 (d, *J* = 7.4 Hz, 1H), 6.57 (s, 1H), 5.12 (s, 4H), 4.71 – 4.32 (m, 8H), 4.16 (d, *J* = 13.5 Hz, 3H), 3.68 – 3.45 (m, 4H), 3.06 (dd, *J* = 14.4, 6.6 Hz, 2H), 2.73 (t, *J* = 6.8 Hz, 2H), 2.38 (dt, *J* = 7.2, 3.2 Hz, 2H), 2.18 (s, 3H), 2.04 (dt, *J* = 21.5, 7.3 Hz, 4H), 1.49 (s, 18H), 0.89 (s, 9H), 0.08 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.22, 169.68, 162.72, 155.66, 155.24, 150.00, 149.70, 149.66, 148.36, 147.32, 144.39, 143.61, 141.74, 140.65, 125.67, 123.74, 122.60, 109.54, 101.17, 96.92, 85.28, 68.48, 64.88, 61.72, 59.69, 50.86, 50.47, 49.05, 47.64, 45.68, 38.57, 37.84, 36.10, 28.13, 27.79, 26.40, 26.15, 25.21, 24.64, 24.30, 16.56, -4.57. **HRMS** *m/z*= 1145.5679 (calcd. for $C_{52}H_{76}N_{16}O_{12}Si$ 1145.5671 $[M+H]^+$).



46

46, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.00 – 8.67 (m, 4H), 8.43 (s, 1H), 7.94 – 7.38 (m, 14H), 7.02 (d, *J* = 7.3 Hz, 1H), 6.31 – 6.06 (m, 2H), 5.93 (s, 1H), 5.52 (s, 2H), 5.38 – 4.95 (m, 5H), 4.77 – 4.08 (m, 24H),

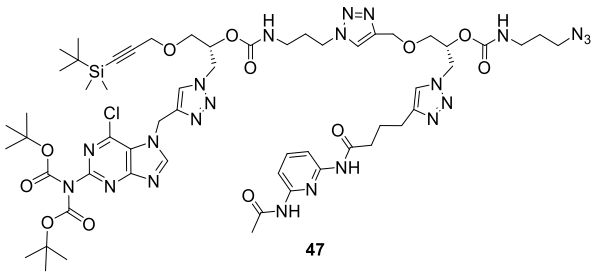
3.78 – 3.41 (m, 8H), 3.25 – 2.87 (m, 6H), 2.69 (d, *J* = 7.8 Hz, 4H), 2.44 – 1.83 (m, 20H), 1.48 (d, *J* = 17.5 Hz, 36H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.15, 169.58, 162.69, 155.66, 155.27, 152.63, 151.87, 151.21, 151.15, 150.10, 149.91, 149.68, 148.58, 147.29, 146.83, 146.74, 144.48, 144.24, 144.09, 141.70, 141.29, 140.57, 130.06, 125.86, 125.24, 123.79, 123.19, 122.65, 109.54, 101.17, 96.93, 91.00, 85.36, 84.29, 71.42, 71.16, 70.93, 69.20, 68.62, 68.46, 64.65, 59.68, 53.00, 50.87, 50.56, 47.54, 45.60, 39.47, 37.82, 36.24, 36.11, 30.28, 30.17, 29.81, 28.02, 27.77, 26.13, 25.17, 25.09, 24.65, 24.42, 16.55, -4.58. **TOF MS ES +** *m/z*= 2360.1666 (calcd. for $C_{103}H_{143}ClN_{38}O_{24}Si$ 2360.0668 $[M+H]^+$).



DGCD

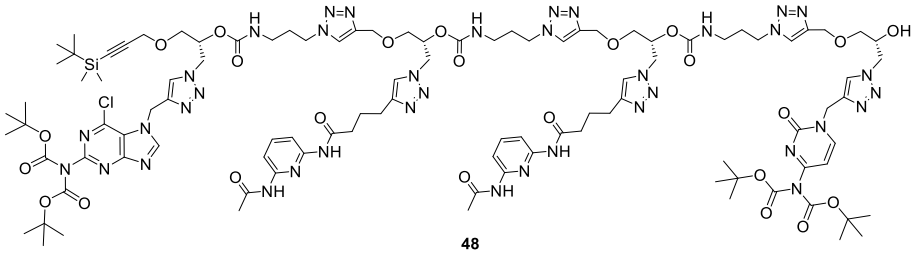
DGCD, **^{1}H NMR (500 MHz, DMSO)** *δ* 11.17 (s, 1H), 10.03 (dd, *J* = 24.2, 8.0 Hz, 4H), 8.18 – 8.02 (m, 3H), 7.95 (d, *J* = 16.5 Hz, 2H), 7.86 – 7.59 (m, 10H), 7.48 – 7.32 (m, 3H), 7.22 (s, 1H), 7.05 (d, *J* = 49.0 Hz, 1H), 6.72 (s, 1H), 5.68 (d, *J* = 7.2 Hz, 1H), 5.37 – 5.02 (m, 6H), 4.87 (dd, *J* = 8.7, 6.1 Hz, 2H), 4.70 – 4.43 (m, 13H), 4.39 (dd, *J* = 13.9, 3.7 Hz, 1H), 4.31 (p, *J* = 8.2, 7.3 Hz, 6H), 4.24 – 4.15 (m, 3H), 4.11 (d, *J* = 5.2 Hz, 1H), 3.95 (d, *J* = 15.6 Hz, 1H), 3.61 – 3.39 (m, 8H), 3.03 – 2.78 (m, 6H), 2.70 – 2.56 (m, 4H), 2.45 (d, *J* = 8.3 Hz, 4H), 2.09 (s, 6H), 1.97 – 1.75 (m, 10H). **^{13}C NMR (126 MHz, DMSO)** *δ* 171.88, 169.29, 166.02, 155.25, 155.22, 150.37, 150.29, 146.39, 146.02, 145.92, 145.75, 143.68, 143.39, 139.88, 137.12, 124.54, 124.13, 122.82, 122.77, 116.28, 109.02, 108.88, 93.85, 79.85, 79.56, 79.27, 77.69, 71.65, 70.48, 68.34, 68.24, 63.85, 57.92, 54.94, 52.70,

48.62, 46.96, 37.42, 35.58, 35.49, 29.99, 24.88, 24.55, 24.50, 24.01.



47

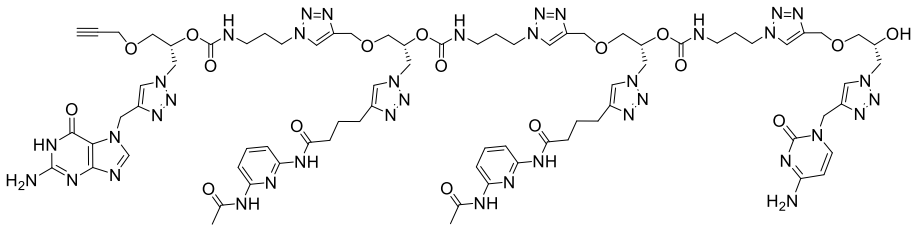
47, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 8.42 (s, 3H), 7.99 – 7.55 (m, 5H), 7.41 (s, 1H), 5.78 (d, *J* = 77.3 Hz, 2H), 5.53 (s, 2H), 5.16 (d, *J* = 4.8 Hz, 2H), 4.78 – 4.44 (m, 6H), 4.38 – 4.02 (m, 4H), 3.69 – 3.57 (m, 2H), 3.33 (t, *J* = 6.5 Hz, 2H), 3.21 (q, *J* = 6.5 Hz, 2H), 3.09 (qd, *J* = 7.3, 4.7 Hz, 4H), 2.74 (d, *J* = 7.2 Hz, 2H), 2.38 (t, *J* = 7.4 Hz, 2H), 2.19 (s, 3H), 2.08 – 1.89 (m, 4H), 1.73 (q, *J* = 6.6 Hz, 2H), 1.48 (s, 18H), 0.90 (s, 9H), 0.08 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 172.05, 169.39, 155.44, 152.60, 151.87, 151.26, 151.26, 149.88, 149.69, 147.09, 146.58, 144.27, 141.31, 140.72, 130.14, 125.15, 123.65, 122.74, 109.50, 101.11, 90.90, 84.37, 71.06, 68.25, 68.14, 64.69, 59.61, 51.13, 50.22, 49.07, 47.44, 39.63, 38.62, 37.76, 36.13, 30.11, 29.03, 28.08, 28.03, 26.11, 25.04, 24.71, 24.30, 16.52, -4.60. **HRMS** *m/z*= 1329.5933 (calcd. for $C_{57}H_{81}ClN_{22}O_{12}Si$ 1329.5935 $[M+H]^+$).



48

48, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 9.02 (s, 4H), 8.44 (s, 1H), 8.02 – 7.54 (m, 12H), 7.46 (d, *J* = 7.4 Hz, 2H), 7.03 (t, *J* = 7.2 Hz, 1H), 6.36 (s, 2H), 5.86 (s, 1H), 5.54 (s, 2H), 5.12 (d, *J* = 46.0 Hz, 5H), 4.70 –

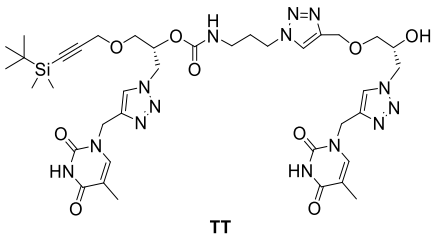
4.09 (m, 23H), 3.79 – 3.42 (m, 4H), 3.17 – 2.83 (m, 6H), 2.67 (s, 4H), 2.26 (d, *J* = 96.5 Hz, 10H), 1.98 (s, 10H), 1.49 (d, *J* = 5.7 Hz, 36H), 0.90 (s, 9H), 0.08 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** *δ* 172.35, 169.72, 155.80, 155.46, 155.24, 152.63, 151.86, 151.30, 151.18, 150.24, 149.69, 148.53, 147.19, 146.77, 144.60, 144.12, 141.63, 141.36, 140.64, 125.66, 125.26, 123.94, 123.78, 122.91, 109.64, 101.16, 96.88, 90.90, 85.33, 84.41, 71.55, 71.06, 69.08, 68.67, 68.31, 64.71, 64.53, 59.61, 53.19, 51.17, 50.52, 47.58, 45.56, 39.62, 37.85, 37.78, 36.06, 28.35, 28.05, 27.78, 25.05, 24.61, 24.46, 16.54, -4.58. **TOF MS ES +** *m/z*= 2360.2353 (calcd. for $C_{103}H_{143}ClN_{38}O_{24}Si$ 2360.0668 $[M+H]^+$).



GDDC

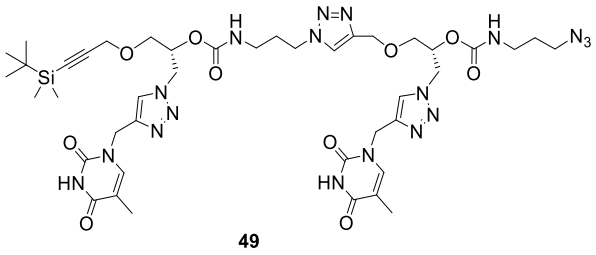
GDDC, **^{1}H NMR (500 MHz, DMSO)** *δ* 10.97 (s, 1H), 10.04 (d, *J* = 29.5 Hz, 4H), 8.16 – 8.05 (m, 3H), 8.02 (s, 1H), 7.91 (s, 1H), 7.84 – 7.57 (m, 9H), 7.44 (t, *J* = 6.0 Hz, 3H), 7.13 (d, *J* = 79.0 Hz, 2H), 6.95 – 6.65 (m, 2H), 5.69 (d, *J* = 7.2 Hz, 1H), 5.20 (s, 2H), 5.08 (dq, *J* = 9.1, 4.7 Hz, 3H), 4.88 (d, *J* = 2.5 Hz, 2H), 4.73 – 4.04 (m, 22H), 4.00 – 3.89 (m, 1H), 3.60 – 3.42 (m, 8H), 2.91 (q, *J* = 6.4 Hz, 6H), 2.61 (t, *J* = 7.7 Hz, 5H), 2.44 (t, *J* = 7.4 Hz, 4H), 2.08 (d, *J* = 3.4 Hz, 6H), 1.87 (dt, *J* = 13.9, 7.0 Hz, 10H). **^{13}C NMR (126 MHz, DMSO)** *δ* 171.92, 169.36, 166.08, 155.30, 155.22, 153.93, 150.39, 150.30, 146.40, 145.87, 143.71, 143.42, 142.80, 139.90, 137.06, 124.53, 124.45, 124.21, 124.12, 122.81, 109.03, 108.90, 93.78, 79.84, 79.28, 77.71, 71.62, 70.64, 68.66, 68.23, 63.85, 57.94, 52.83, 50.03, 49.98, 49.87,

48.62, 46.99, 43.50, 37.43, 35.53, 30.03, 24.89, 24.52, 24.04.



TT

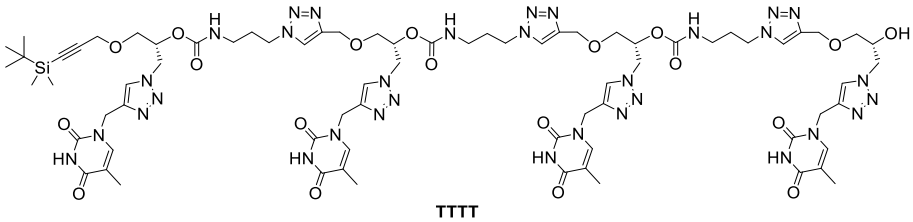
TT, **^{1}H NMR (500 MHz, $CDCl_3$)** δ 10.26 (s, 1H), 10.05 (s, 1H), 7.98 (d, J = 5.2 Hz, 2H), 7.84 (s, 1H), 7.39 – 7.31 (m, 2H), 6.18 (t, J = 6.0 Hz, 1H), 5.16 (d, J = 3.6 Hz, 1H), 4.91 (d, J = 12.8 Hz, 4H), 4.72 – 4.35 (m, 9H), 4.25 – 4.13 (m, 3H), 3.72 – 3.47 (m, 4H), 3.08 (dd, J = 16.5, 6.4 Hz, 2H), 2.12 – 1.97 (m, 2H), 1.84 (dd, J = 7.8, 1.1 Hz, 6H), 0.91 (s, 9H), 0.09 (s, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ 164.82, 164.76, 155.60, 151.57, 151.49, 144.53, 142.26, 141.91, 140.67, 125.56, 123.88, 111.33, 111.27, 101.31, 90.78, 71.59, 71.13, 69.09, 68.27, 64.66, 59.59, 59.12, 53.15, 50.99, 47.49, 43.19, 37.74, 30.21, 26.15, 19.88, 16.54, 12.39, -4.58. **HRMS** *m/z*= 879.4043 (calcd. for $C_{38}H_{54}N_{14}O_9Si$ 879.4040 $[M+H]^+$).



49

49, ^{1}H NMR (500 MHz, DMSO) δ 11.30 (s, 2H), 8.12 – 7.96 (m, 3H), 7.56 (dd, J = 2.3, 1.3 Hz, 2H), 7.37 (d, J = 41.9 Hz, 2H), 5.24 – 4.99 (m, 2H), 4.88 (t, J = 1.7 Hz, 4H), 4.70 – 4.46 (m, 6H), 4.40 – 4.16 (m, 4H), 3.70 – 3.46 (m, 4H), 3.29 (t, J = 6.7 Hz, 2H), 3.02 – 2.81 (m, 4H), 1.93 – 1.83 (m, 2H), 1.74 (dd, J = 5.6, 1.2 Hz, 6H), 1.56 (t, J = 6.8 Hz, 2H), 0.89 (s, 9H), 0.07 (s, 6H**). ^{13}C NMR (126 MHz, DMSO)** δ 164.28,

155.16, 150.72, 143.42, 142.39, 142.34, 141.03, 124.41, 124.12, 108.85, 102.78, 89.27, 70.49, 70.44, 68.70, 68.27, 63.87, 58.61, 50.12, 48.18, 46.97, 42.01, 37.52, 37.42, 29.99, 28.49, 25.85, 16.11, 11.96, -4.79. **HRMS** *m/z*= 1005.4585 (calcd. for $C_{42}H_{60}N_{18}O_{10}Si$ 1005.4582 $[M+H]^+$).



TTTT

TTTT, **^{1}H NMR (500 MHz, DMSO)** δ 11.30 (d, *J* = 2.1 Hz, 4H), 8.10 (d, *J* = 3.3 Hz, 3H), 8.05 – 7.94 (m, 4H), 7.63 – 7.53 (m, 4H), 7.41 (d, *J* = 4.3 Hz, 3H), 5.17 – 5.06 (m, 3H), 4.88 (d, *J* = 6.1 Hz, 8H), 4.65 – 4.46 (m, 13H), 4.45 – 4.20 (m, 10H), 3.96 (ddt, *J* = 7.1, 4.7, 2.4 Hz, 1H), 3.56 (dtd, *J* = 24.1, 10.6, 4.7 Hz, 6H), 3.44 – 3.36 (m, 2H), 2.90 (d, *J* = 7.4 Hz, 6H), 1.94 – 1.82 (m, 6H), 1.78 – 1.67 (m, 12H), 0.88 (s, 10H), 0.07 (s, 6H). **^{13}C NMR (126 MHz, DMSO)** δ 164.28, 155.21, 150.72, 143.70, 143.41, 142.38, 142.13, 141.04, 124.47, 124.43, 124.14, 108.86, 102.78, 89.27, 71.64, 70.56, 68.69, 68.21, 63.87, 58.61, 54.94, 52.90, 50.11, 46.96, 42.15, 42.03, 37.41, 29.97, 25.84, 16.10, 11.95, -4.79.

Annex 3.4 Characteristics of synthesized intermediates

Oligomer	Purification	Yield (%)	Product status
10	EtOAc/*n*-hexane = 1/9 to 5/5	77	Brown oil
12	MeOH/DCM = 0/100 to 5/95	100	Yellow oil
13	MeOH/DCM = 2/98 to 8/92	93	Yellow oil
14	EtOAc/*n*-hexane = 0/10 to 4/6	92	Colorless oil
15	Washed with *n*-hexane	100	White solid
(*R*)-16	Crystallized from cold EtOAc	98	Yellow solid
(*S*)-16	Crystallized from cold EtOAc	100	Yellow solid
(*S*)-17	EtOAc /*n*-hexane = 5/5 to 0/10	93	Foamy-white solid
(*R*)-19	EtOAc /*n*-hexane = 5/5 to 0/10	75	Foamy-yellow solid
(*S*)-19	EtOAc /*n*-hexane = 5/5 to 0/10	81	Foamy-yellow solid
20	MeOH/DCM = 2/98 to 7/93	100	Foamy-yellow solid
M	EtOAc/*n*-hexane = 1/9 to 5/5	97	Brown oil
P	EtOAc/*n*-hexane = 1/9 to 5/5	91	White solid
I	MeOH/DCM = 0/100 to 5/95	100	Yellow oil
C_6	EtOAc/*n*-hexane = 1/9 to 7/3	99	Colorless oil
T	MeOH/DCM = 0/100 to 3/97	100	White solid
D	EtOAc/*n*-hexane = 25/75 to 100/0	100	White solid
22	EtOAc/*n*-hexane = 5/5 to 10/0	95	Foamy-white solid

C	MeOH/DCM = 2/98 to 7/93	98	Foamy-yellow solid
23	EtOAc/*n*-hexane = 1/9 to 4/6	100	Brown oil
24	MeOH/DCM = 0/100 to 3/100	100	Foamy-white solid
TC_6	MeOH/DCM = 1/99 to 5/95	92	Foamy-white solid
PTC_6	MeOH/DCM = 2/98 to 6/94	95	Foamy-white solid
26	EtOAc/*n*-hexane = 2/8 to 7/3	98	Pink oil
$MPTC_6$	MeOH/DCM = 2/98 to 6/94	57	Foamy-white solid
27	EtOAc/*n*-hexane = 0/10 to 5/5	100	Colorless oil
C_6G	MeOH/DCM = 0/100 to 4/96	86	Foamy-white solid
28	MeOH/DCM= 0/100 to 5/95	100	Foamy-yellow solid
29	MeOH/DCM = 0/100 to 12/88	70	Yellow solid
C_6GMPTC_6	MeOH/DCM = 0/100 to 5/95	49	Foamy-white solid
C_6C	MeOH/DCM = 0/100 to 5/95	76	Foamy-white solid
30	MeOH/DCM = 0/100 to 12/88	100	Yellow solid
C_6CI	MeOH /DCM = 2/98 to 10/90	82	Foamy-white solid
31	MeOH /DCM = 2/98 to 10/90	83	Foamy-yellow solid
(*R*)-32	MeOH/DCM= 0/100 to 7/93	100	Foamy-white solid
(*S*)-32	MeOH/DCM= 0/100 to 5/95	100	Foamy-white solid
DC_6	MeOH/DCM = 0/100 to 5/95	83	Foamy-white solid
33	MeOH/DCM= 0/100 to 7/93	86	Foamy-white solid
IDC_6	MeOH/DCM = 2/98 to 10/90	63	Foamy-white solid
34	MeOH/DCM = 2/98 to 10/90	70	Foamy-white solid
C_6CIIDC_6	Washed with Et_2O	83	Foamy-white solid

PDC_6	MeOH/DCM = 2/98 to 6/94	100	Foamy-white solid
$IPDC_6$	MeOH/DCM = 3/97 to 10/90	84	Foamy-white solid
35	MeOH/DCM = 3/97 to 10/90	84	Foamy-white solid
C_6ClPDC_6	Washed with Et_2O	44	Foamy-white solid
C_6Cl'	MeOH /DCM = 0/100 to 9/91	85	Foamy-white solid
36	MeOH /DCM = 0/100 to 10/90	80	Foamy-white solid
37	MeOH /DCM = 3/97 to 15/85	84	Foamy-white solid
$C_6Cl'IPDC_6$	Washed with Et_2O	90	Foamy-white solid
TG	MeOH /DCM = 2/98 to 6/94	90	Foamy-white solid
38	MeOH/DCM= 0/100 to 6/94	75	Foamy-yellow solid
39	EtOAc/*n*-hexane = 5/5 to 10/0	100	Foamy-fellow solid
CT	MeOH /DCM = 2/98 to 6/94	96	Foamy-white solid
40	MeOH /DCM = 2/98 to 12/88	65	Foamy-white solid
TGCT	Neutralization by saturated $NaHCO_3$ before washing with brine, water, acetone and Et_2O	71	Foamy-yellow solid
(*R*)-41	MeOH /DCM = 2/98 to 6/94	70	Foamy-white solid
(*S*)-41	MeOH /DCM = 2/98 to 6/94	72	Foamy-white solid
(*R*)-42	MeOH/DCM= 3/97 to 8/92	100	Foamy-white solid
(*S*)-42	MeOH/DCM= 3/97 to 8/92	88	Foamy-white solid
(*R*)-43	MeOH/DCM= 3/97 to 8/92	100	Foamy-white solid
(*S*)-43	MeOH/DCM= 3/97 to 8/92	88	Foamy-white solid
(*R, R*)-GD	MeOH/DCM= 3/97 to 8/92	90	Foamy-yellow solid
(*S, S*)-GD	MeOH/DCM= 3/97 to 8/92	79	Foamy-yellow solid

All-*R* 44	MeOH/DCM= 3/97 to 8/92	87	Foamy-yellow solid
All-*S* 44	MeOH/DCM= 3/97 to 8/92	88	Foamy-yellow solid
All-*R* DCGD	Neutralization by saturated $NaHCO_3$ before washing with brine, water, acetone and Et_2O	51	Foamy-yellow solid
All-*S* DCGD	Neutralization by saturated $NaHCO_3$ before washing with brine, water, acetone and Et_2O	92	Foamy-yellow solid
DG	MeOH /DCM = 2/98 to 8/92	81	Foamy-yellow solid
45	MeOH/DCM= 3/97 to 8/92	100	Foamy-yellow solid
CD	MeOH/DCM= 3/97 to 8/92	60	Foamy-white solid
46	MeOH /DCM = 3/97 to 12/88	75	Foamy-yellow solid
All-*R* DGCD	Neutralization by saturated $NaHCO_3$ before washing with brine, water, acetone and Et_2O	83	Foamy-yellow solid
47	MeOH/DCM= 2/98 to 6/94	100	Foamy-white solid
48	MeOH /DCM = 3/97 to 14/86	87	Foamy-yellow solid
All-*R* GDDC	Neutralization by saturated $NaHCO_3$ before washing with brine, water, acetone and Et_2O	87	Foamy-yellow solid
TT	MeOH /DCM = 2/98 to 10/90	87	Foamy-white solid
49	MeOH /DCM = 2/98 to 10/90	87	Foamy-white solid
All-*R* TTTT	MeOH /DCM = 3/97 to 14/86	87	Foamy-white solid

Annex 3.5 TOF-MS/MS

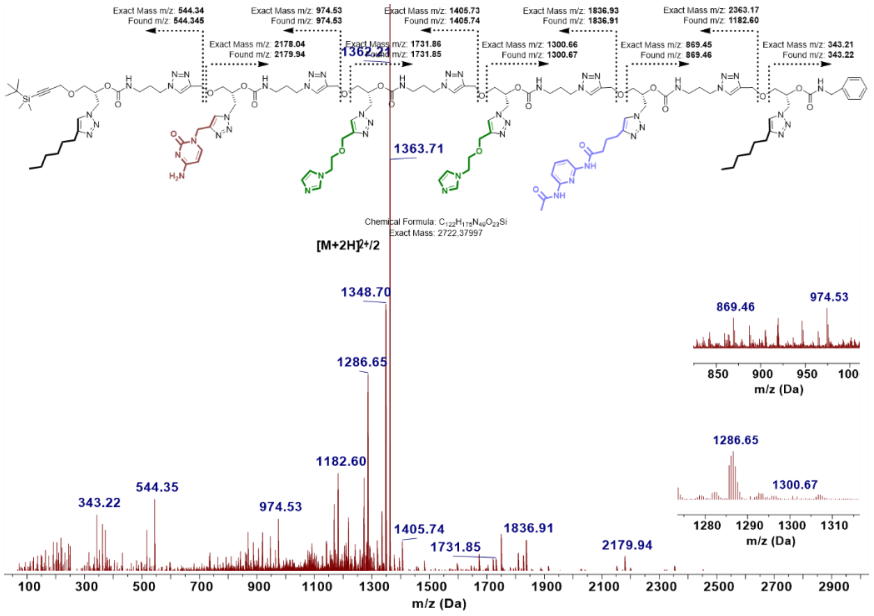


Figure A.3.2 TOF-MS/MS of C_6CIIDC_6.

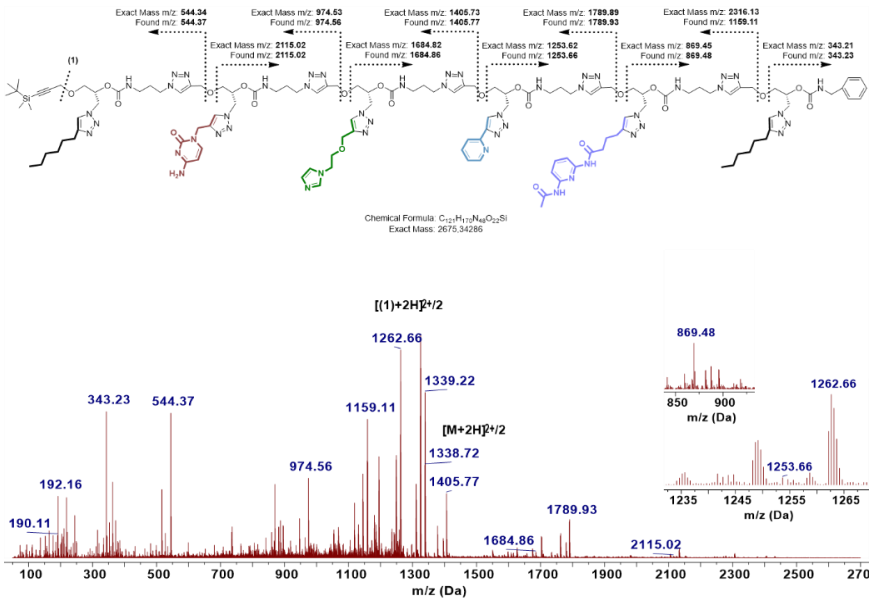


Figure A.3.3 TOF-MS/MS of C_6CIPDC_6.

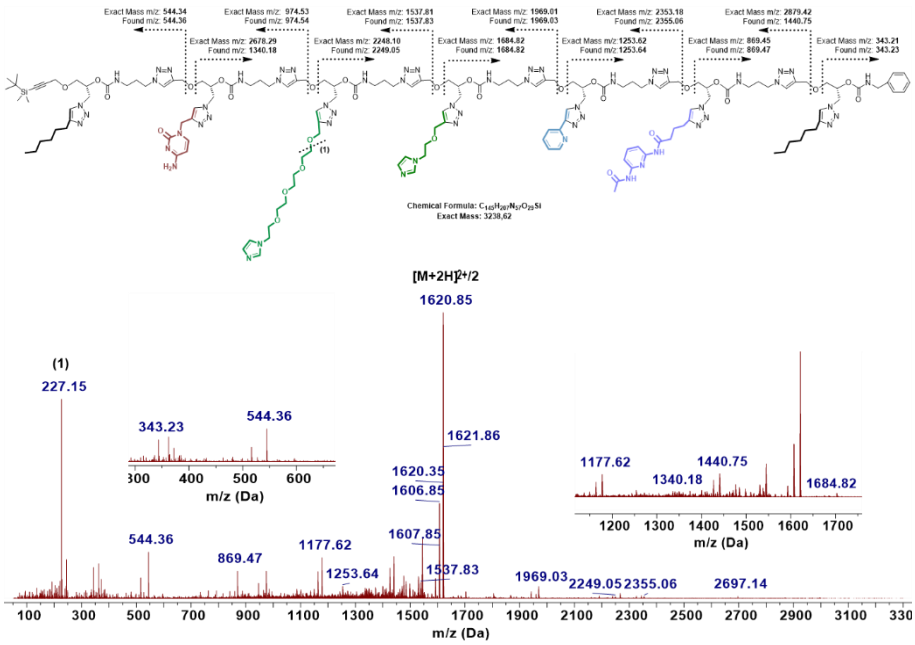


Figure A.3.4 TOF-MS/MS of $C_6Cl'IPDC_6$.

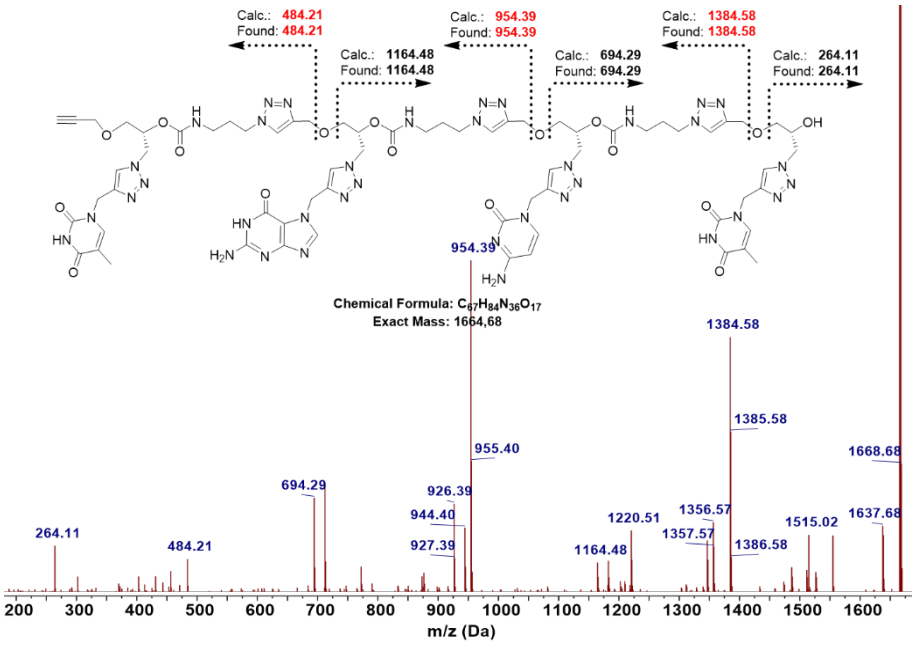


Figure A.3.5 TOF-MS/MS of TGCT.

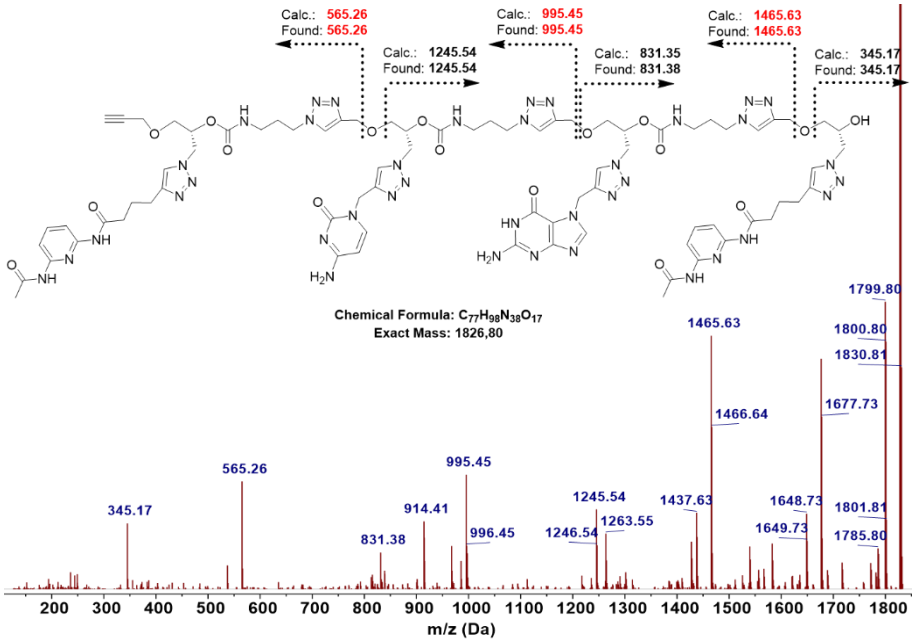


Figure A.3.6 TOF-MS/MS of all-*R* DCGD.

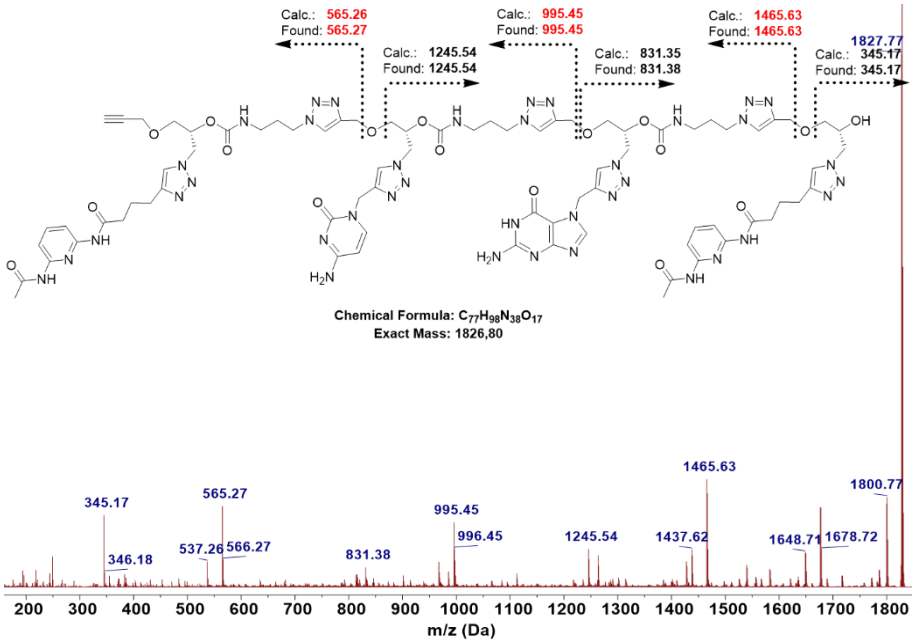


Figure A.3.7 TOF-MS/MS of all-*S* DCGD.

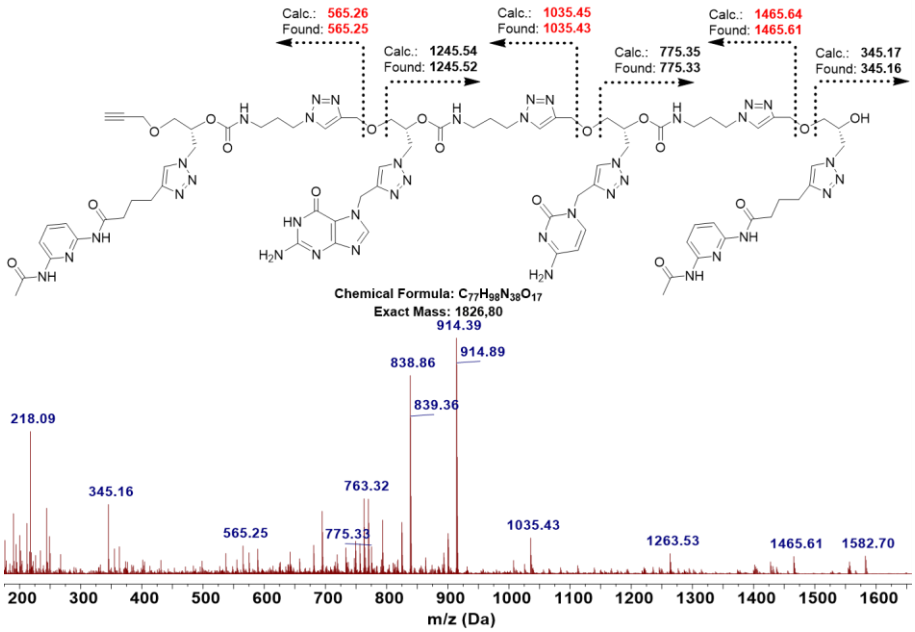


Figure A.3.8 TOF-MS/MS of DGCD.

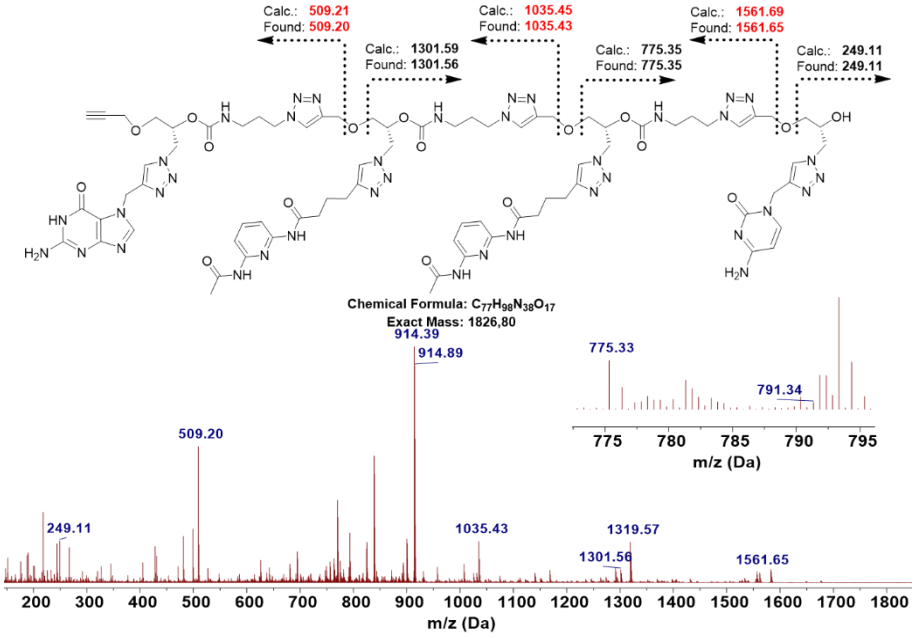


Figure A.3.9 TOF-MS/MS of GDDC.

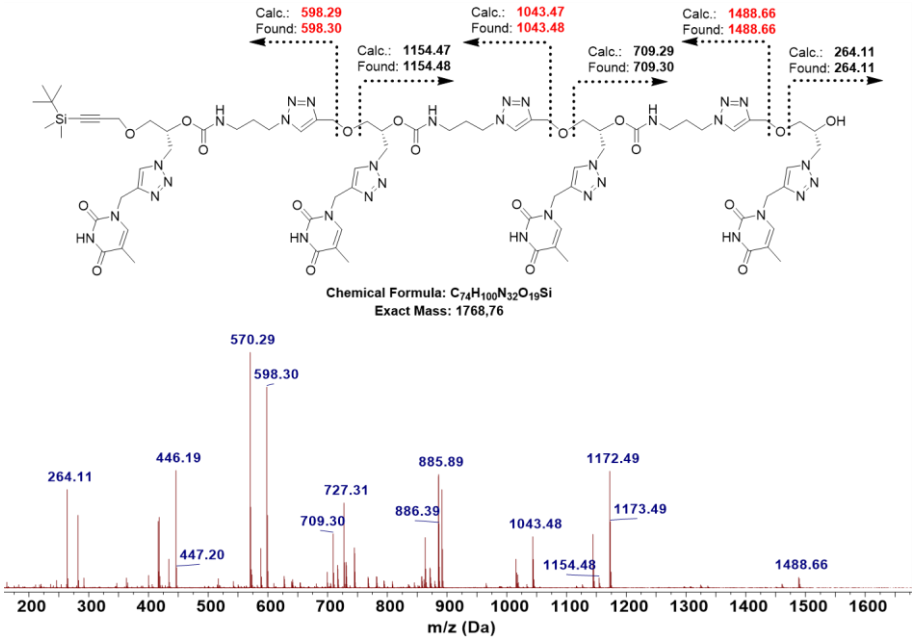


Figure A.3.10 TOF-MS/MS of TTTT.

Annex 3.6 Preprint of article on recognition experiments with nucleobase-decorated oligomers

Soft Matter **2024**

Chain stretching in brushes favors sequence recognition for nucleobase-functionalized flexible precise oligomers

Kseniia Grafskaia,[(1,††)] Qian Qin,[(1,††)] Jie Li,[(1,†)] Delphine Magnin,[(1)] David Dellemme,[(2)] Mathieu Surin,[(2)] Karine Glinel,[(1,*)] and Alain M. Jonas[(1,*)]

(1) Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, Croix du Sud 1 L7.04.02, Louvain-la-Neuve, Belgium

(2) Laboratory for Chemistry of Novel Materials, Université de Mons – UMONS, Avenue Maistriau, 17, B-7000 Mons, Belgium

(†) Current address: State Key Laboratory of Chemical Resource Engineering, Beijing Laboratory of Biomedical Materials, Beijing University of Chemical Technology, Beijing 100029, P. R. China

(*) Corresponding authors

(††) These authors contributed equally to this work

Abstract

Six different flexible stereocontrolled oligo(triazole-urethane)s substituted by precise sequences of nucleobases or analogs are synthesized. Molecular dynamics simulations indicate that the flexibility of the backbone leads to unspecific complexation of pairs of oligomers, irrespective of the complementarity of their sequences. This is ascribed to the existence of other interactions between pairs of

oligomers, as well as to the spatial blurring of the sequence order encoded in the chemical structure of the chain due to its flexibility. The same conclusions are drawn when investigating the irreversible adsorption of different probe oligomers onto a layer of target oligomers grafted by click chemistry in a mushroom configuration on a silicon substrate. In contrast, when the target oligomers are grafted in denser brush configurations, irreversible adsorption becomes more specific, with probe chains of complementary sequence being twice more probable to be irreversibly-bound to the layer of target chains than those of non-complementary sequence. This is ascribed to lateral excluded volume interactions between chains in the brush, leading to partial chain stretching and increased spatial preservation of the information contained in the monomer sequence of the chains. At even higher grafting densities, however, the penetration of the probe chains in the brush becomes increasingly difficult, resulting in a loss of binding efficiency. Our work thus demonstrates the adverse role of chain flexibility in the specificity of complexation between nucleobase-functionalized oligomers and provides directions for an improvement of specificity by tuning the grafting density of target chains on a substrate.

Introduction

The specificity of DNA duplex formation has been used to develop applications such as DNA origami,[1–3] programmed nanoparticle assembly[4–14] or molecular algorithms.[15] These rely on the formation of double helices, in which hydrogen bonds between nucleobases on complementary strands and base-stacking interactions provide the stabilization mechanism for duplex formation.[16] Importantly, the large electrostatic persistence length of single-stranded DNA[17] compared to the distance between successive nucleobases along the DNA sugar-

phosphate backbone results in substantial chain rigidity, which contributes to the local spatial preservation of the chemical information stored in the sequence of nucleobases along the chains.

Synthetic oligomers bearing complementary recognition units can also form sequence-selective duplexes.[18–24] Among those, a number of synthetic sequence-defined oligomers bearing nucleobases or analogous side groups have been demonstrated to form H-bonded duplexes, sometimes including with complementary oligonucleotides;[25–29] peptide nucleic acids are a well-known example.[30] Most nucleobase-functionalized synthetic systems have relatively rigid backbones and/or moderate distances between successive bases, leading to the short-range spatial preservation of the chemically-encoded sequence information contained in the primary structure of the chains.[29] This partial spatial preservation presumably helps in the recognition of a complementary sequence. Here, we want to address the issue of recognition when the distances between bases and the flexibility of the backbone scramble more the information contained in the primary sequence of the chains. Would the blurring resulting from conformational fluctuations completely prevent recognition between complementary chains and if this is the case, what would be possible solutions to this issue?

To answer this question, we synthesized flexible precise oligomers with a well-defined sequence of nucleobases. A variety of synthetic routes towards precision oligomers are available, as reviewed in recent articles;[31–35] they differ in their scalability, tolerance towards different functional groups, chemical stability of the products, and ability to control chirality. Considering our aims, we selected a flexible and tolerant synthetic route which offers a complete control over molar mass, monomer sequence, and chirality, as demonstrated in

our previous work (Scheme A.3.6).[36,37] Molecular dynamics (MD) simulations of short oligomers prepared by this route have shown that they tend to form random coils of small radius of gyration (in the range of 1 nm and below) owing to the flexibility of their urethane-triazole backbone.[36–39] Therefore, owing to our precise control over sequence and configuration, conformational blurring will be the only scrambling mechanism that contributes to complicate sequence recognition in this system.

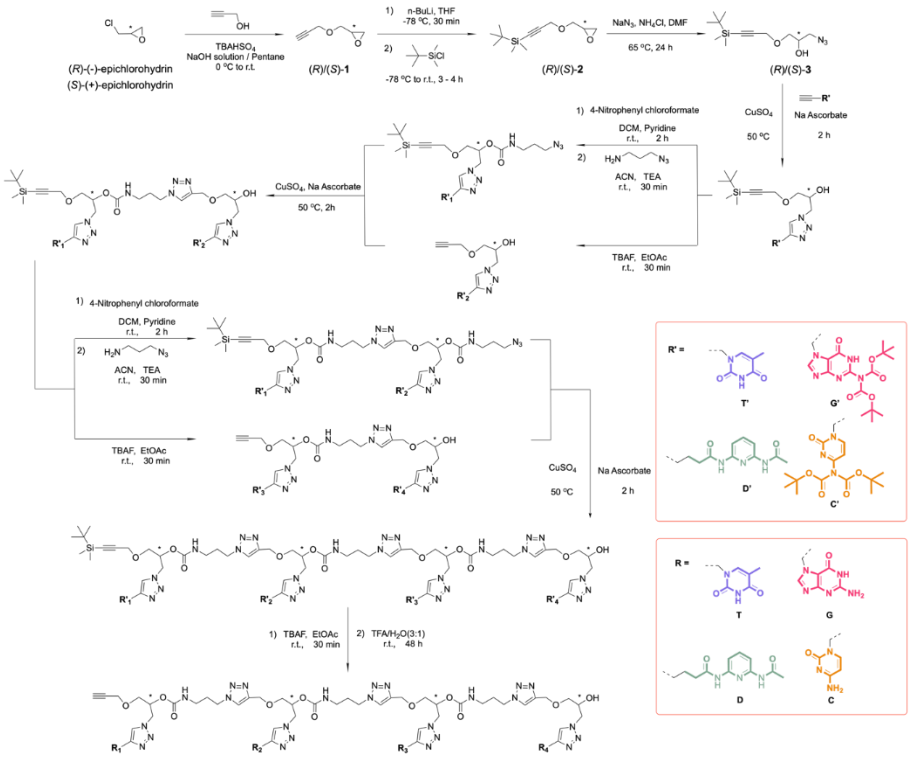
Here, we use this route to synthesize a series of complementary and non-complementary nucleobase tetramers, and we show by MD simulations that chain pairing occurs in solution irrespective of the sequence-complementarity of the chains. We confirm this conclusion by studying with *in situ* ellipsometry the binding of complementary and non-complementary probe chains in solution to a layer of target chains grafted in the mushroom regime on a silicon surface. We then show how the specificity (or selectivity) of the recognition process can be substantially improved by playing with excluded volume interactions in the grafted layer using the grafting density of the chains as controlling parameter.

Results and Discussion

Synthesis of precision oligomers decorated with nucleobases or analogues

The synthesis of sequence-defined oligo(urethane-triazole) tetramers of controlled chirality was adapted from our previous work (Scheme 3.8 and ESI section S2).[36,37,39] The monomers are equipped with guanine (G), cytosine (C), a 2,6-diaminopyridine xenobase (D) or thymine (T) side groups (Scheme A.3.6); trivalent H-bonded duplexes can be formed between the complementary base pair analogues C

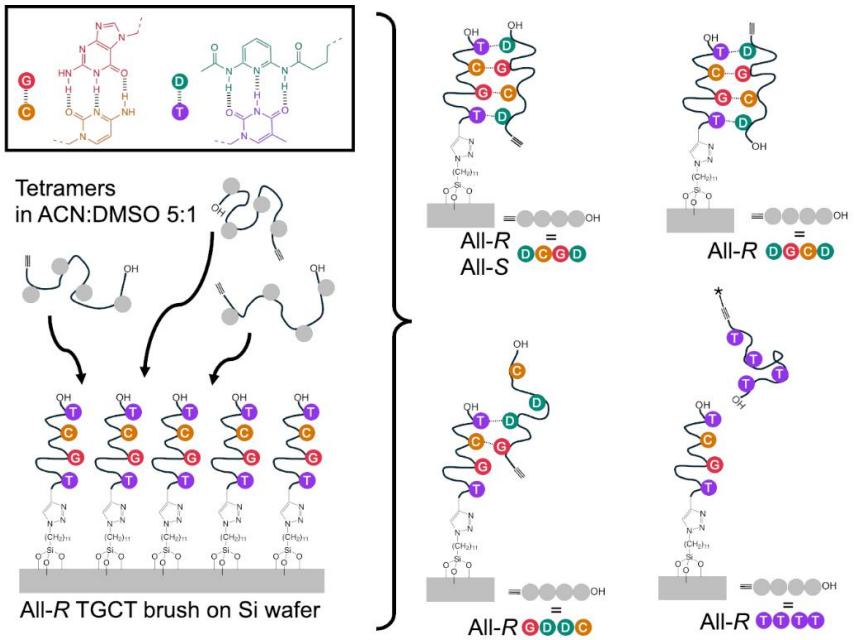
and G, or D and T (Scheme A.3.7).



Scheme A.3.6 General synthetic route of the nucleobase-functionalized oligo(triazole-urethane) tetramers.

Six different tetramers were obtained from these monomer units; they are designed by a succession of letters corresponding to the successive nucleobases, starting from the alkyne end group and ending in the hydroxyl terminal group (which defines the direction of the chain). An all-*R* TGCT tetramer was selected as target chain to be end-grafted on a silicon surface (Scheme A.3.7). Then, we prepared three probe oligomers that have a sequence fully complementary to the target chain when in extended conformation but differ in their chirality or chain direction: all-*R* DCGD, all-*R* DGCD (flipped version of the previous one) and all-*S* DCGD tetramers. Additionally, we

prepared a permuted all-*R* GDDC sequence which can only form a partial complementary duplex with the TGCT target chain unless a strongly folded conformation is taken (Scheme A3.4.2). Finally, a non-complementary all-*R* TTTT chain which cannot form a duplex with the target chain was also synthesized as a negative control probe chain. The chain structures were checked by nuclear magnetic resonance (NMR) and tandem mass spectrometry as fully detailed in ESI sections S2 and S3.



Scheme A.3.7 Different sequences of synthesized tetramers and schematics of possible binding patterns between the target grafted chains and the different probe chains. The pattern of hydrogen bonding between the nucleobases or analogues installed on the tetramers is displayed in the top left box. The alkyne end-group of the TTTT oligomer was not deprotected (as indicated by the '*') to improve solubility; it is terminated by a *tert*-butyl-silyl group.

Binding constants of the monomer units

NMR titration and dilution experiments at different temperatures were used to investigate H-binding between the monomers in a 5:1 v:v mixture of acetonitrile (ACN) and dimethyl sulfoxide (DMSO) (ESI, section S4). DMSO was added to ensure complete solubility of all monomer units. The binding constants and standard binding enthalpies and entropies of the different monomer units at 295K are given in Table A.3.1, from which the binding constant between the probe and target chains when four nucleobases are paired can be estimated to be *ca.* 11000 L/mol at 295K. This estimation is very imprecise, since it ignores other interactions between chains, the decrease of the number of degrees of freedom resulting from the restriction of the conformational space of the chains upon complexation, the effects of chirality and chain direction, and synergetic binding effects.

Table A.3.1 Binding constants (Ka) and standard binding enthalpies and entropies of the different monomer units, obtained from NMR measurements performed in deuterated ACN:DMSO 5:1 v:v at 295K. "-" indicates that no measurement was performed but that the parameter is likely to be negligeable; “0” indicates that a measurement was performed but that the variations of NMR shift were too small to extract a significant value for the parameter. The selected standard concentration is 1 M.

Complex	Ka (L/mol)	ΔH° (kJ/mol)	ΔS° (J/K·mol)
G-G	1.4±0.2	-	-
C-C	0.1±0.1	-	-
T-T	0	-	-
D-D	0	-	-
D-T	0.99±0.06	-24±4	-43±14
G-C	104±15	-5.7±2.3	-19±8
D-G	0	-	-

Molecular dynamics (MD) simulations of the interaction between probe and target chains in solution

MD simulations of assemblies comprising the target chain in the presence of each of the probe chains were performed over 1 μs time scale with a 1 fs time step, using the AMBER package (details in ESI, section 5).[40] The simulations were carried out starting from the target chain and the probe chain relatively distant (around 25 Å), to avoid any bias in the initial intermolecular contacts. All the simulations were carried implicitly in acetonitrile, using a Generalized Born solvation model considering the acetonitrile dielectric constant (37.5 at room temperature).[41–43] This implicit solvent model is accurate to model aprotic solvents and permits a large conformational sampling at reasonable computational cost. At the early stages of the MD simulations, the probe and target oligomers were found to rapidly interact (in less than 1 ns) to form heteromolecular assemblies. In each assembly, both chains were found to be very flexible and fold into compact globular shapes, with an average radius of gyration R_g ~0.85 nm (ESI Fig.S5.1). The flexibility of the oligomeric chains is also evident from the large variations of the chain end-to-end distances with simulation time, ranging from 0.3 nm to *ca.* 3.9 nm (ESI Fig.S5.2-3). Additionally, the chain-ends that are far apart in the primary sequence remain, on average, close in the 3D structure (average chain end-to-end distances of *ca.* 1.3 and 1.4 nm for target and probe chains, respectively) (ESI Table S5.1). On average, the assembly of the probe and target chains form compact assemblies in which the two oligomers remain close together but keep a high flexibility; a snapshot is shown in Fig.A.3.11.

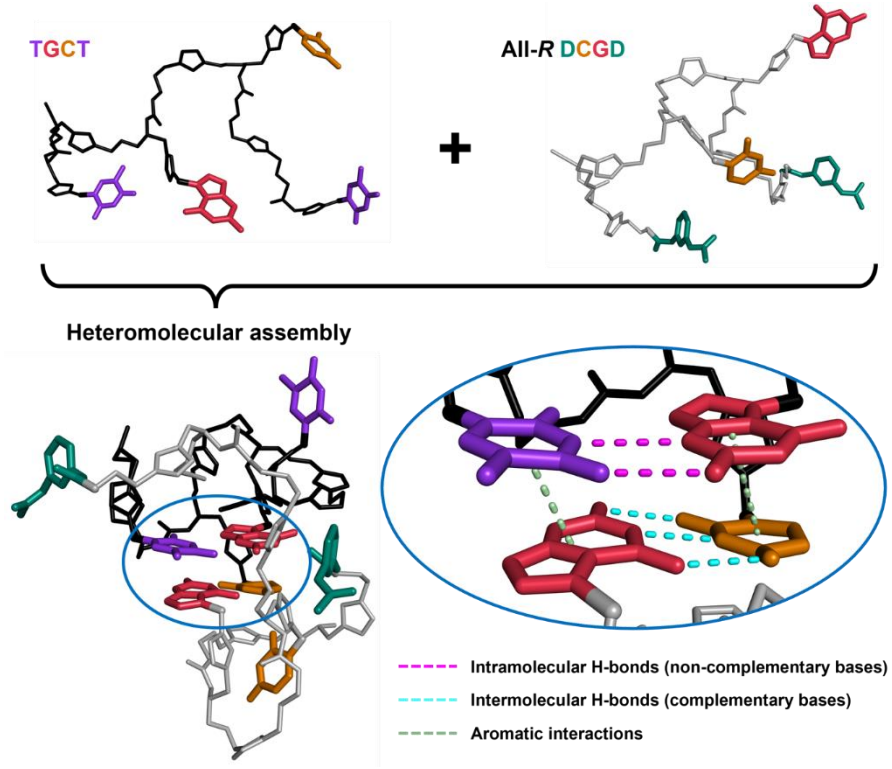


Figure A.3.11 MD snapshots of the heteromolecular assembly of the target chain, TGCT, and one of the probe chains, all-*R* DCGD. The backbones of TGCT and all-R DCGD are coloured in black and grey, respectively. The nucleobases are colored as follows: thymine (T) in purple, guanine (G) in red, cytosine (C) in orange and the xenobase 2,6-diaminopyridine (D) in blue-green. Top: starting structure of the target and probe chains, before their assembly and folding. Bottom: one conformation of the heteromolecular assembly, showing its globular shape. A zoom shows the four bases within the assembly and the coloured dots highlight different types of interactions.

The occurrence of H-bonds and parallel stacking of aromatic cycles was computed for each MD snapshot (total of 4000 conformations for each assembly), with on average 4.5-4.7 aromatic interactions and *ca.* 4.5 H-bonds per configuration. Less than 20 % of these H-bonds occur between complementary base pairs (ESI Fig.5.6). Hence, in

acetonitrile, the oligomers do not form well-organized duplexes with a regular pairing, but rather form disordered assemblies, regardless of the complementarity of the monomer sequences. This is due to the high intrinsic flexibility of the chain backbones, the formation of a competitive network of H-bonds (involving H-bonding units such as triazole, carbamate and non-complementary bases), and significant folding promoted by the stacking of aromatic cycles (Figure A.3.1 bottom). This folding also brings numerous intramolecular interactions (around 40 % of the total H-bonds and around 46 % of the total aromatic interactions). These results indicate that a specific recognition of complementary tetramers is prevented by the flexibility of the backbone of the oligomers which spatially blurs their sequence.

We surmised that a better recognition could be achieved by grafting the target chains on a surface, which should restrict the available conformational space of the chains provided the grafting density be large enough. Indeed, in the brush regime in which the average distance between the chains becomes lower than their unperturbed radius of gyration, the chains should on average stretch vertically, resulting in a vertical spatial distribution of the nucleobases on average closer to the chemical sequence of the chains.

Grafting of target oligomers onto azido-silanized wafers

Therefore, we grafted the all-*R* TGCT target chains onto azido-silanized silicon wafers, using their alkyne terminal groups to click them unto a monolayer of 11-azidoundecyltrimethoxysilane (AzUTMS**)**. A reproducible AzUMTS monolayer could be obtained over the native oxide of Si wafers by overnight gas phase silanation at 100°C, using a well-established silanation protocol.[44–49] XPS confirmed the grafting of the silane onto the silicon wafer (ESI section 7). Measurements of the layer thickness by spectroscopic

ellipsometry provided an average thickness of 1.1±0.03 nm (standard deviation 0.32 nm for 124 different samples prepared at different times, each measured in triplicate). Additionally, x-ray reflectivity (XRR) was performed on 35 different samples (ESI Fig.S6.1); from the first minimum arising from destructive interference in the reflectivity curves, an average thickness of 0.9±0.02 nm (standard deviation of 0.12 nm) was found. The XRR data was fit to extract electron density profiles (ESI Fig.S6.1) as described in previous work,[47] from which an average electron density profile was computed (Fig.A.3.12). An average grafting density of 2.9±0.1 molecules/nm^2 (0.32 molecules/nm^2 standard deviation) was obtained by integration of the profiles after subtracting the one of a clean Si wafer, considering that each grafted AzUMTS molecule contains 147 electrons. This is a typical grafting density for silanation in the gas phase.[44] As a comparison, the chain areal density of a polyethylene crystal over its *ab* plane (perpendicular to the chain axes) is *ca.* 5.5 molecules/nm^2; hence, the layer is made of chains in a partially coiled conformation. The average water contact angle of 82.8±0.2° (standard deviation 2.1° for five measurements on 46 samples) is in good agreement with what was reported before by Vos *et al*.[50]

The alkyne-ended all-*R* TGCT target chains were then grafted by Copper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC), coupling the terminal alkyne group of the all-*R* TGCT oligomers with the azide-terminated AzUTMS layer. The grafting was performed overnight at 80°C in four solvents, DMF, DMF:DMSO 75:25, DMF:DMSO 50:50 v:v and DMSO. XPS again confirmed the successful grafting of the TGCT chains (ESI section 7). Spectroscopic ellipsometry indicated an average increase of the layer thickness upon reaction (Table A.3.2) from 0.59±0.06 nm in DMF to 1.55±0.06 nm in DMSO, revealing a higher number of grafted chains when the click reaction is performed

in the presence of increasing DMSO content. Consistently, the water contact angle (Table A.3.2) decreased from 60.5±1.1° in DMF to 40.5±0.4° in DMSO, due to the higher number of grafted chains whose terminal ends comprise a hydrophilic hydroxyl group and a polar triazole-thymine side chain. The differences in the density of the brush depending on the solvent used for the grafting process result from differences in the ratio between intramolecular segmental interactions and segment/solvent interactions depending on solvent quality. DMSO being a better solvent, it favours segment/solvent interactions, brush swelling and penetration of oligomers during the grafting-to process, resulting in a higher final grafting density.

Table A.3.2 Morphological characterization of the target layer composed of grafted all-R TGCT molecules, depending on the solvent used for the click reaction.

DMF:DMSO vol:vol	*d* (nm)[a]	error[b]	*N*[c]	*s* (molec/nm^2)[d]	error[b]	*N*[c]	*x* (nm)[e]	*q* (°)[f]	error[b]	N[c]
100:0	0.6	0.06	15	0.30	0.11	2	1.8	60.5	1.1	25
75:25	1.1	0.12	12	0.60[g]	0.06[g]	4[g]	1.3[g]	51.9	2.1	19
50:50	1.3	0.10	15					47.1	1.1	20
0:100	1.6	0.06	63	0.74	0.03	8	1.2	40.5	0.4	70

(a) Average ellipsometric thickness of the grafted all-*R* TGCT layer; (b) standard error on the measurement (same units); (c) number of measurements; (d) XRR-determined average number of grafted all-*R* TGCT molecules per nm^2; (e) average distance between grafted all-*R* TGCT molecules, from XRR; (f) water contact angle; (g) the density profiles of samples clicked in solvents containing from 25 to 50 vol% of DMSO were averaged together.

The electron density profiles of all-*R* TGCT grafted layers prepared in different solvents were obtained by XRR (Fig.A.3.12), confirming the increasing grafting density when increasing the proportion of DMSO in the solvent. Owing to the limited precision of XRR, and the

proximity of the ellipsometry-derived thickness and contact angles of samples clicked in solvents having from 25 to 50% DMSO, the electron density profiles of these samples were averaged together. For pure DMF and DMSO, the profiles of different samples were also averaged to increase precision.

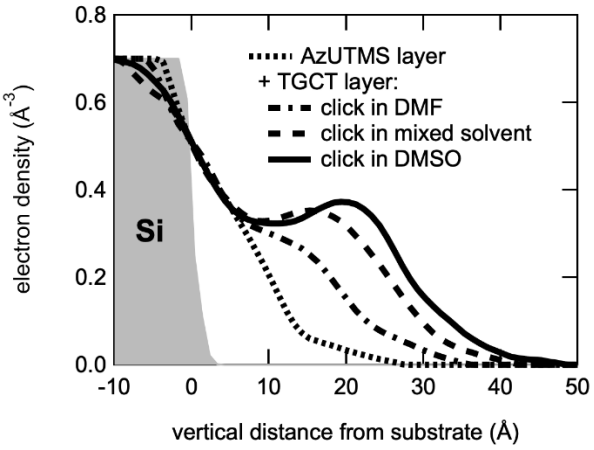


Figure A.3.12 Average electron density profiles of a bare Si wafer (gray), a wafer silanized by AzUTMS, and a silanized wafer reacted with the all-*R* TGCT oligomer by CuAAC in DMF, DMSO, or mixtures of DMF and DMSO containing from 25 to 50% DMSO. All XRR data and fits are shown in Fig.S6.2 of the ESI.

The average grafting densities were then obtained by subtracting the average electron density profile of the AzUTMS-silanized wafer, integrating the subtracted profiles and dividing by the number of electrons per TGCT molecule (874). This provided an average grafted density *s* increasing from 0.30±0.11 molecule per nm^2 when the click reaction is performed in pure DMF to 0.74±0.03 molecule per nm^2 in DMSO (Table A.3.2). This corresponds to average distances $x = s^{-1/2}$ between all-*R* TGCT grafted chains of *ca.* 1.2 and 1.8 nm for DMSO and DMF, respectively, with an intermediate value for the

mixed solvents. Given a radius of gyration R_g of 0.85 nm from the MD simulations, the oligomer layers are in the brush regime ($d < 2R_g$) when the click reaction is performed in pure DMSO, and in the mushroom regime ($d > 2R_g$) when performed in pure DMF.[51] Hence, by selecting the click conditions, the TGCT chains can be placed under different lateral constraints, corresponding to more or less stretched chain conformations, with consequently different degrees of spatial memory loss of the primary sequence.

Recognition between sequence-defined oligomers and the grafted all-R TGCT layers

The recognition experiments were performed at 25°C in a temperature-controlled ellipsometry flow cell (Fig.A.3.13). Importantly, all experiments were performed in ACN with 16.6 vol% of DMSO added to ensure full solubility of the tetramers, irrespective of the solvent previously used for the grafting of the target layer. The experimental procedure involved flowing pure degassed solvent in the ellipsometer cell, measuring a complete spectroscopic ellipsometry spectrum of the grafted target layer (A in Fig.A.3.13), switching to a 0.4 mM solution of a probe tetramer in the solvent while monitoring the evolution of the ellipsometric angle D at 400 nm with time until a plateau was reached, switching to the pure solvent while monitoring D at 400 nm until a plateau was reached, and finally measuring the complete ellipsometry spectrum of the target layer plus irreversibly-adsorbed probe layer in the pure solvent (B in Fig.A.3.13).

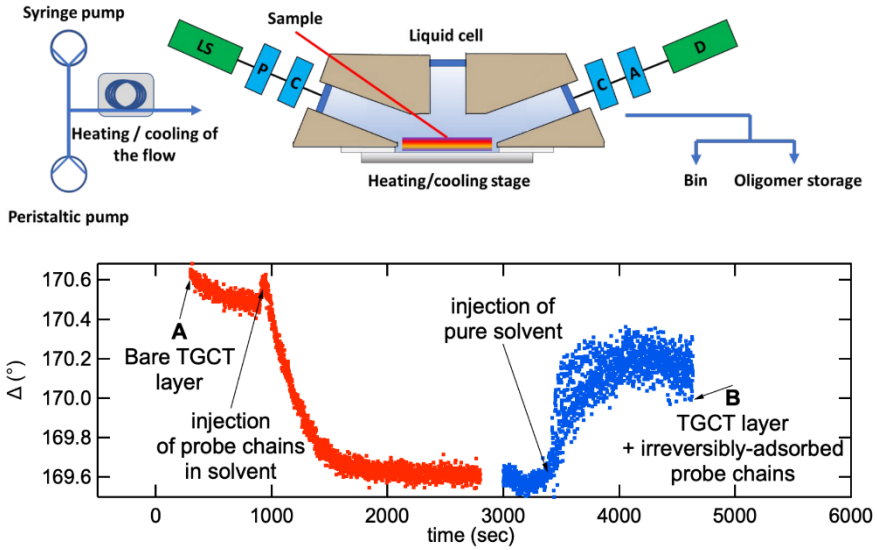


Figure A.3.13 Top. Scheme of the ellipsometry flow cell used for the measurement of adsorption (LS: light source, P: polarizer, C: (rotating) compensator, A: analyzer, D: detector). Bottom. Typical variation of the ellipsometric angle D measured at 400 nm during adsorption at 25°C of a probe tetramer (here, all-*R* DCGD) on a substrate (here, a grafted target layer of all-*R* TGCT of 0.74 molecule/nm^2 grafting density) from a 0.4 mM solution in ACN:DMSO 5:1. A and B indicate the times at which complete ellipsometric spectra were measured (two zone averaging).

The volume of irreversibly-adsorbed probe tetramer per surface area of target layer (nm^3/nm^2) was computed as $V_{irr} = c(\overline{\Delta_B} - \overline{\Delta_A})$, in which $\overline{\Delta_A}$ and $\overline{\Delta_B}$ are the ellipsometric angles measured at time *A* and *B*, respectively, averaged over the 400-700 nm wavelength range. This corresponds to a Taylor expansion limited to first order of the ellipsometric angle versus surfacic adsorbed volume, which is valid for limited thicknesses of the adsorbed layer as in our case. The calibration constant *c*=-1/1.2615 nm/° was obtained from simulations

of ellipsometric angles depending on thickness and solvent content of the adsorbed layer (ESI, section 8).

The results gathered in Fig.A.3.4 (left) show that substantial irreversible binding occurs, confirming that the equilibrium is strongly displaced towards complex formation despite the relatively moderate binding constant estimated above. This is due to the locally very high local concentration of nucleobases in the brush, to the existence of other types of H-bonds in the system as shown by MD simulations, as well as to the low probability to have a simultaneous release of the four pairs of nucleobases.

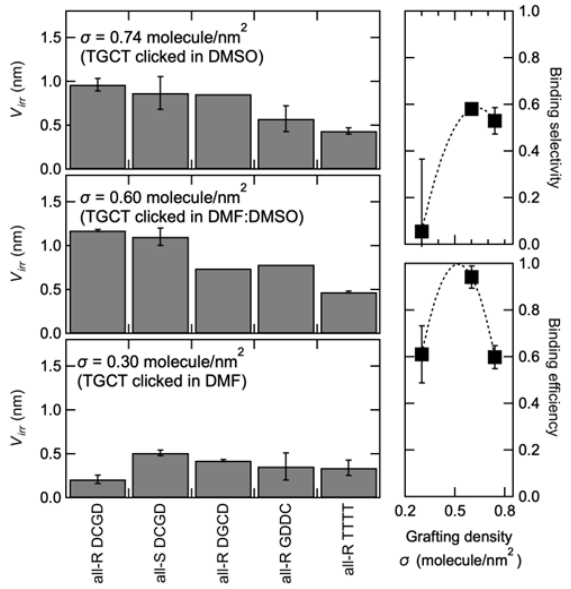


Figure A.3.14 (left) Volume V_{irr} of probe chains irreversibly-adsorbed per area of target chain layer, for different grafting densities σ of the grafted all-*R* TGCT target chains, depending on the sequence of the probe chains. The experiments were performed at 25°C in ACN:DMSO 5:1 at a 0.4 mM concentration of probe chains. (right) Binding selectivity and binding efficiency of the target layer depending on its grafting density; the dashed lines are parabolic fits drawn to guide the eye.

We computed two indicators that quantify the recognition of the target layer by probe chains. First, we defined the *binding efficiency* E as the number of irreversibly-bound probe chains per target chain in the layer, for fully complementary target and probe chains. It can be estimated by the ratio of the average surfacic volume V_{irr} of irreversibly adsorbed DCGD oligomers irrespective of chirality (hence probe chains having strictly a complementary sequence and same direction as the target chains), to the dry thickness d of the target layer:

$$E = \frac{V_{irr(R-DCGD)} + V_{irr(S-DCGD)}}{2d_{R-TGCT}}$$

An efficiency E=1 is obtained when each TGCT target chain is complexed by a complementary DGCD probe chain, whatever their chirality; in contrast, when no complementary probe chain irreversibly adsorbs in the target layer, E=0.

The second indicator is the *binding selectivity* S, defined as one minus the ratio of the average irreversibly adsorbed surfacic volume of fully non-complementary TTTT oligomers to the one of fully complementary DCGD oligomers (whatever their chirality):

$$S = 1 - \frac{2V_{irr(R-TTTT)}}{V_{irr(R-DCGD)} + V_{irr(S-DGCD)}}$$

A selectivity S=0 indicates that fully non-complementary TTTT chains adsorb identically to fully complementary DCGD chains, whereas S=1 when non-complementary chains do not adsorb at all. The binding efficiency and selectivity are displayed in Figure A.3.4 (right) versus the grafting density of the target layer.

Regarding the selectivity of the irreversible adsorption, when the all-*R*

TGCT target layer is in the mushroom regime after having been clicked in DMF (bottom panel in Fig.A.3.14 left), the adsorption is not sequence-specific with, *e.g.*, complementary all-*R* DCGD probe chains being even less irreversibly-adsorbed than non-complementary all-*R* TTTT probe chains. This is in good agreement with expectations from the MD simulations. In contrast, for all-*R* TGCT target chains in the brush regime (grafted in DMSO, top panel in Fig.A.3.14 left), complementary probe chains (all-*R* DCGD, all-*S* DCGD, and even all-*R* DGCD) are more bound to the layer, irrespective of their chirality or directionality, compared to all-*R* GDDC probe chains with a permuted sequence or completely non-complementary all-*R* TTTT probe chains. The indicator of binding selectivity in Figure A.3.14 (top right) confirms this conclusion, with almost no selectivity at the lower grafting density whereas values of *ca.* 0.6 are found for denser brushes.

Hence, these results indicate some sequence selectivity in the recognition process when the target chains adopt a more stretched average conformation in denser brushes, although it is not perfect since fully non-complementary all-*R* TTTT probe chains still bind to the all-*R* TGCT target layer (*ca*. 50% less than the complementary probe chains). Clearly, the lateral constraints due to volume exclusion in the brush, which result in thicker layers corresponding to more extended average TGCT chain conformations, favour the sequence-specificity of the recognition process. However, this effect remains moderate, with probe chains having the proper sequence only adsorbing approximately twice as much as chains of improper sequence.

It would thus be tempting to try to increase further the grafting density of the target chains to stretch them even more. This would not be

easy to perform in a grafting-to process and might require grafting-from strategies which are known to generally result in higher grafting densities. However, this is not likely to be of much interest, considering the way binding efficiency varies with grafting density (Figure A.3.14, bottom right): when increasing too much the grafting density of the target chains, the binding efficiency decreases due to increased excluded volume interactions within the brush, which limits the penetration of probe chains in the brush. As a result, even though the binding selectivity might be better, the number of probe chains bound per target chain will decrease, which will impair applications requiring the binding of a substantial number of chains.

Conclusions

In this work, we have shown that flexible precision tetramers substituted by nucleobases or analogs can be synthesized by a versatile route leading to alkyne-ended chains that can be grafted onto azide-modified silicon wafers in configurations ranging from mushroom to brush layers. When in the mushroom regime, the chains do not exhibit any recognition specificity towards probe chains in solution. This is also the conclusion drawn from molecular modelling simulations of the recognition process in solution, the reason being the existence of numerous competing H-bonding interactions with the triazole-urethane backbone of the oligomers as well as the complete scrambling of the spatial positioning of the nucleobases due to the flexibility of the backbone.

In contrast, when the target chains are grafted in the brush regime, the recognition process becomes more specific, with probe chains of complementary sequence being twice more probable to be irreversibly-adsorbed than non-complementary sequences. This undoubtedly results from lateral volume interactions between chains

leading to partial chain stretching and a decreased scrambling of the information contained in the chemical sequence of the chains. However, although encouraging, this remains much less specific than DNA, in which the charged chain backbone is of higher rigidity due to electrostatic repulsion, and for which the distance between successive nucleobases is smaller than in our synthetic system by a factor of *ca*. 2. Simulations could be extended to other types of synthetic backbones and recognition units to improve selectivity in the brush regime.

It is thus very improbable that applications similar to DNA origami, which need a high binding selectivity of specific nucleobase sequences in the presence of many competitors, could be transposed to more flexible precision oligomers. Nevertheless, the irreversible binding of probe and target chains we have shown is probably strong enough for simpler applications such as the formation of superlattices of nanoparticles, because such applications do not require to single out a specific sequence among many. Further work in these directions is thus of interest to explore more completely the limits and advantages of flexible precision oligomers as alternatives to DNA strands.

Author Contributions (CRediT)

Investigation and Formal analysis: K.Gr. and A.M.J. (XRR and ellipsometry experiments), Q.Q. and L.J. (synthesis and characterization of the molecules), D.D. and M.S. (molecular modeling simulations); Methodology: K.Gr. and D.M. (*in-situ* ellipsometry); Supervision: K.G., M.S. and A.M.J.; Writing – original draft: K.Gr., D.D., A.M.J.; Writing – review & editing: K.G., M.S., A.M.J.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

We acknowledge the Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek under EOS project n° 30650939 for financial support. We also acknowledge support from the ASM and SUCH platforms of UCLouvain for MS, NMR and XPS analyses. Computational resources are provided by the Consortium des Equipements de Calcul Intensif (CECI), funded by FNRS (Grant No. U.G.018.18) and Wallonia Region. K.G. and M.S. are Research Directors of the F.R.S.−FNRS. D.D. is supported by a FRIA grant funded by F.R.S.- FNRS.

References

1 G. Tikhomirov, P. Petersen and L. Qian, *Nature*, 2017, 552, 67–71.

2 K. F. Wagenbauer, C. Sigl and H. Dietz, *Nature*, 2017, 552, 78–83.

3 P. W. K. Rothemund, *Nature*, 2006, 440, 297–302.

4 G. Chen, K. J. Gibson, D. Liu, H. C. Rees, J.-H. Lee, W. Xia, R. Lin, H. L. Xin, O. Gang and Y. Weizmann, *Nat. Mater.*, 2018, 18, 1–8.

5 G. Yao, J. Li, Q. Li, X. Chen, X. Liu, F. Wang, Z. Qu, Z. Ge, R. P. Narayanan, D. Williams, H. Pei, X. Zuo, L. Wang, H. Yan, B. L. Feringa and C. Fan, *Nat. Mater.*, 2020, 1–10.

6 E. Auyeung, T. I. N. G. Li, A. J. Senesi, A. L. Schmucker, B. C. Pals, M. O. de la Cruz and C. A. Mirkin, *Nature*, 2014, 505, 73–77.

7 M. R. Jones, R. J. Macfarlane, B. Lee, J. Zhang, K. L. Young, A. J. Senesi and C. A. Mirkin, *Nature*, 2010, 9, 913–917.

8 R. J. Macfarlane, B. Lee, M. R. Jones, N. Harris, G. C. Schatz and

C. A. Mirkin, *Science*, 2011, 334, 204–208.

9 P. J. Santos, P. A. Gabrys, L. Z. Zornberg, M. S. Lee and R. J. Macfarlane, *Nature*, 2021, 591, 586–591.

10 K. J. Si, Y. Chen, Q. Shi and W. Cheng, *Adv. Sci.*, 2018, 5, 1700179.

11 M. Xie, J. Jiang and J. Chao, *Sensors*, 2023, 23, 9229.

12 S. Julin, S. Nummelin, M. A. Kostiainen and V. Linko, *J. Nanoparticle Res.*, 2018, 20, 119.

13 D. J. Lewis, D. J. D. Carter and R. J. Macfarlane, *J. Am. Chem. Soc.*, 2020, 142, 19181–19188.

14 S. Wang, X. Xie, Z. Chen, N. Ma, X. Zhang, K. Li, C. Teng, Y. Ke and Y. Tian, *Int. J. Mol. Sci.*, 2021, 22, 7558.

15 D. Woods, D. Doty, C. Myhrvold, J. Hui, F. Zhou, P. Yin and E. Winfree, *Nature*, 2019, 567, 366–372.

16 P. Yakovchuk, *Nucleic Acids Res.*, 2006, 34, 564–574.

17 B. Tinland, A. Pluen, J. Sturm and G. Weill, *Macromolecules*, 1997, 30, 5763–5765.

18 H. Zeng, X. Yang, A. L. Brown, S. Martinovic, R. D. Smith and B. Gong, *Chem. Commun.*, 2003, 1556–1557.

19 A. Marquis, V. Smith, J. Harrowfield, J. Lehn, H. Herschbach, R. Sanvito, E. Leize-Wagner and A. V. Dorsselaer, *Chem. Eur. J.*, 2006, 12, 5632–5641.

20 B. Gong, *Acc. Chem. Res.*, 2012, 45, 2077–2087.

21 D. Núñez-Villanueva, G. Iadevaia, A. E. Stross, M. A. Jinks, J. A. Swain and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, 139, 6654–6662.

22 Y. Zhang, R. Cao, J. Shen, C. S. F. Detchou, Y. Zhong, H. Wang, S. Zou, Q. Huang, C. Lian, Q. Wang, J. Zhu and B. Gong, *Org. Lett.*, 2018, 20, 1555–1558.

23 J. A. Swain, G. Iadevaia and C. A. Hunter, *J. Am. Chem. Soc.*, 2018, 140, 11526–11536.

24 D. Núñez-Villanueva and C. A. Hunter, *J. Am. Chem. Soc.*, 2022, 144, 17307–17316.

25 D. H. Appella, *Curr. Opin. Chem. Biol.*, 2009, 13, 687–696.

26 H. V. Nguyen, Z. Zhao, A. Sallustrau, S. L. Horswell, L. Male, A. Mulas and J. H. R. Tucker, *Chem. Commun.*, 2012, 48, 12165–12167.

27 P. Karri, V. Punna, K. Kim and R. Krishnamurthy, *Angew. Chem. Int. Ed.*, 2013, 52, 5840–5844.

28 L. Maes, D. M. Roqeuro, L. M. Pitet, P. Adriaensens and T. Junkers, *Polym. Chem.*, 2020, 11, 2027–2033.

29 J. Li, Z. Wang, Z. Hua and C. Tang, *J. Mater. Chem. B*, 2020, 8, 1576–1588.

30 P. E. Nielsen, *Chem. Biodivers.*, 2010, 7, 786–804.

31 M. A. R. Meier and C. Barner-Kowollik, *Adv. Mater.*, 2019, 31, 1806027.

32 R. Aksakal, C. Mertens, M. Soete, N. Badi and F. D. Prez, *Adv. Sci.*, 2021, 65, 2004038.

33 B. van Genabeek, B. A. G. Lamers, C. J. Hawker, E. W. Meijer, W. R. Gutekunst and B. V. K. J. Schmidt, *J. Polym. Sci.*, 2021, 59, 373–403.

34 Z. Deng, Q. Shi, J. Tan, J. Hu and S. Liu, *ACS Mater. Lett.*, 2021, 3, 1339–1356.

35 J.-F. Lutz, *Eur. Polym. J.*, 2023, 199, 112465.

36 J. Li, M. Leclercq, M. Fossepré, M. Surin, K. Glinel, A. M. Jonas and A. E. Fernandes, *Polym. Chem.*, 2020, 11, 4040–4046.

37 Q. Qin, J. Li, D. Dellemme, M. Fosséprés, G. Barozzino-Consiglio, I. Nekkaa, A. Boborodea, A. E. Fernandes, K. Glinel, M. Surin and A. Jonas, *Chem. Sci.*, 2023, 14, 9283–9292.

38 S. Kardas, M. Fosséprés, V. Lemaur, A. E. Fernandes, K. Glinel, A. M. Jonas and M. Surin, *J. Chem. Inf. Model.*, 2022, 62, 2761–2770.

39 J. Li, Q. Qin, S. Kardas, M. Fosséprés, M. Surin, A. E. Fernandes, K. Glinel and A. M. Jonas, *ACS Catal.*, 2022, 12, 2126–2131.

40 D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, 26, 1668–1688.

41 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *Chem. Phys. Lett.*, 1995, 246, 122–129.

42 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.*, 1996, 100, 19824–19839.

43 V. Tsui and D. A. Case, *Biopolymers*, 2000, 56, 275–291.

44 A. Pallandre, K. Glinel, A. M. Jonas and B. Nysten, *Nano Lett.*, 2004, 4, 365–371.

45 A. M. Jonas, Z. Hu, K. Glinel and W. T. S. Huck, *Macromolecules*, 2008, 41, 6859–6863.

46 A. M. Jonas, Z. Hu, K. Glinel and W. T. S. Huck, *Nano Lett.*, 2008, 8, 3819–3824.

47 X. Pei, A. Fernandes, B. Mathy, X. Laloyaux, B. Nysten, O. Riant and A. M. Jonas, *Langmuir*, 2011, 27, 9403–9412.

48 O. Deschaume, D. Magnin, Z. A. Cheng, C. Douchamps, P. Labbé, S. Yunus, M.-C. Durrieu, B. Nysten, K. Glinel, S. Demoustier-Champagne and A. M. Jonas, *Biomacromolecules*, 2014, 15, 3706–3716.

49 A. Dirani, A. E. Fernandes, D. R. Wong, P. Lipnik, C. Poleunis, B. Nysten, K. Glinel and A. M. Jonas, *Langmuir*, 2014, 30, 10057–10065.

50 R. Vos, T. Steylaerts, A. Franquet, A. Moussa, T. Stakenborg and K. Jans, *Sol. St. Phen.*, 2018, 282, 31–36.

51 R. R. Netz and D. Andelman, *Phys. Rep.*, 2003, 380, 1–95.

Chapter 4 Self-assembling chiral sequence-defined oligomers for homogeneous catalysis

This chapter is adapted from the article: Qin, Q.; Li, J.; Dellemme, D.; Fosséprés, M.; Barozzino-Consiglio, G.; Imane Nekkaa; Boborodea, A.; Fernandes, A. E.; Glinel, K.; Surin, M.; Jonas, A. M. Dynamic Self-Assembly of Supramolecular Catalysts from Precision Macromolecules. *Chem. Sci.* **2023**, *14* (35), 9283–9292.

Abstract: In this chapter, the catalysts generated by the self-assembly of complementary precision oligomers were tested in the aerobic oxidation of alcohols. The catalytic results show that only the two self-assembled chains containing all the needed components for catalysis can exhibit remarkable catalytic activity. In contrast, the low efficiency of the systems in which the two chains lack one component can be improved by introducing the missing unit as a free component in the solution. These results support the formation of a binuclear copper complex in the catalytic cycle. In section 4.3.2, molecular dynamics (MD) simulations suggest that the complementary chains can rapidly self-assemble into globules, in which different interactions contribute to stabilization. Besides, network analysis indicates that the base analogs at the ends of the strands provide a continuous stabilization mechanism and connectivity between the catalytic groups. Additionally, a mathematical model is developed to quantitatively predict the self-assembly of catalytic species and the catalytic properties of the resulting constitutional library in section

4.3.4, thereby providing design rules for self-assembled macromolecular catalytic systems.

* Binding affinity of recognition pairs, Cu/base interactions, and DOSY experiments were done with the help of Dr. Jie Li; the corresponding data analysis and mathematical models were performed by Prof. Alain Jonas.

* Molecular dynamics simulations data are from Prof. Mathieu Surin and Mathieu Fossépré in UMONS.

4.1 Introduction

In biomolecular catalysts, the different chemical groups involved in the catalytic site self-assemble in spatially close positions, resulting in optimal availability and often adaptive activity. This optimized self-assembly of active groups is intrinsically dynamic, enabling easy access to and from the reactive site and facilitating the transition state of the reaction.[1,2] Achieving such efficient and dynamic self-assembly of active groups, as well as the corresponding catalytic activity, is still a significant challenge for synthetic molecular catalysts.

Here, we will demonstrate by molecular dynamics simulations and experiments that, by designing sequence-defined oligomeric chains equipped with functional catalytic groups and complementary recognition units, it is possible to induce the dynamic self-assembly of supramolecular cyclic structures displaying unusually high catalytic efficiency at low concentrations. Importantly, we also show that the catalytic properties of such oligomers can be accurately predicted by a compact mathematical model, despite the multiplicity of their possible assembly configurations.

As mentioned before, attempts to form catalytic sites by self-assembly of synthetic macromolecules have been reported using peptides, DNA scaffolds, supramolecular polymers,[4-37] etc. Especially, supramolecular and covalent macrocycles have been successfully used to creat catalytic sites involving the concerted action of different functional groups[38–42] thanks to their cyclic configurations.

In this chapter, we combine these ideas by designing two precision oligomers (O_a and O_{b1} in Figure 4.1D) that can self-assemble in a variety of configurations, including di(oligomeric) macrocycles (Figure 4.1E). All these supramolecular species possess the five cooperative

catalytic groups involved in the copper-catalyzed aerobic oxidation of alcohols (Figure 2.6 in chapter 2), which are distributed between the two oligomers. We want to demonstrate the emergence of catalytic activity from the formed dynamic constitutional library at low concentrations, and that this activity mainly results from supramolecular di(oligomeric) macrocycles that enable the ideal positioning and connection of the cooperative catalytic residues within a confined space. Additionally, a mathematical model will be developed to quantitatively predict the self-assembly of catalytic species and the catalytic properties of the constitutional library, thereby providing designing rules for self-assembled macromolecular catalytic systems.

4.2. Strategy

To demonstrate the possibility to self-assemble two properly-designed precision oligomers into a catalytically-active superstructure, we use the system developed for the sustainable aerobic oxidation of alcohols based on TEMPO and $Cu^{(I)}$.[133,143,163,170,198] Although the catalytic mechanism is still discussed,[170] it was proposed[163,198] that the catalytic cycle requires the formation of a binuclear copper complex involving two bidentate nitrogen ligands (such as bipyridine) and two auxiliary ligands (such as NMI) (Figure 4.1A). In support of this hypothesis, recent findings have shown the effectiveness of dinuclear copper complexes in the TEMPO-catalyzed aerobic oxidation of alcohols.[199,200]

These five needed functional groups were precisely distributed between two strands O_a and O_b, each strand having intrinsically a very low catalytic activity, if any. Additionally, the two strands were equipped with complementary hydrogen-binding recognition units at

their ends to favor their assembly into a variety of configurations, including supramolecular di(oligomeric) cycles, while avoiding the formation of homo-oligomeric supramolecular structures (Figure 4.1B-E). The first strand, O_a, consists of a $C_6G\underline{MP}TC_6$ sequence (catalytic groups underlined), complementary to the second strand, O_{b1}, with a $C_6C\underline{I'IP}DC_6$ sequence (Figure 4.1D). When self-assembled in a di(oligomeric) cycle in the presence of $Cu^{(I)}$, the two strands should provide together the five groups needed for catalysis (Figure 4.1A). To demonstrate the completeness of this self-assembled active site, two other O_b strands were also synthesized, both of which can only form an incomplete catalytic site in combination with O_a: a $C_6C\underline{II}DC_6$ (O_{b2}) sequence that lacks a P group, and a $C_6C\underline{IP}DC_6$ (O_{b3}) sequence that lacks a I group (Figure 4.1D). Due to the removal of one unit in one strand of each of these two systems, the O_{b2} and O_{b3} chains are one unit shorter than the O_{b1} chain of the complete system. The libraries of self-assembled configurations that form from these three systems are shown in Figure 4.1E.

The functional groups are attached as side chains to the oligo(urethane triazole) backbone we described in chapter 3 (Figure 4.1B); the precise sequence of these complementary oligomers is displayed in Figure 4.1D. The oligomers comprise complementary nucleobase analogs (cytosine C, guanine G, thymine T and 2,6-diamidopyridine D) acting as recognition units through G---C and D---T pairing (Figure 4.1C). The catalytic groups comprise two pyridyltriazole (P)-copper complexes, two imidazole co-ligands (I and I' having different spacer lengths for optimal accessibility), and one TEMPO radical (M). Additionally, hexyl side chains (C_6) were added at both ends of the chains to favor solubility in the acetonitrile (ACN):dimethylsulfoxide(DMSO) 95:5 v:v solvent system used for the experiments and possibly contribute to stabilization of the self-

assembled structures. The synthetic procedures and characterization of the intermediary and final products by NMR and mass spectrometry are described in chapter 3. Experimental conditions are in Annex 4.1.

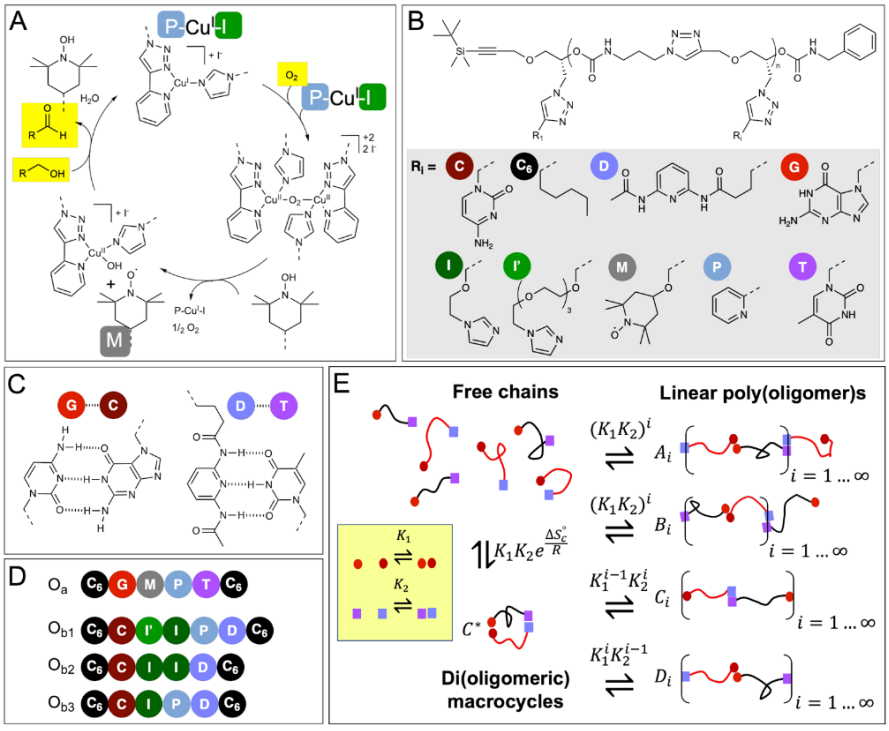


Figure 4.1 Postulated mechanism of the Cu^{I}/TEMPO-catalyzed aerobic oxidation of alcohols, and chemical structure of the oligomers and constitutional libraries. **A.** Simplified catalytic cycle, adapted from reference;[163] the mechanism requires the cooperative action of 2 $Cu^{(I)}$ atoms, 2 pyridyltriazole ligands (P), 2 imidazole co-ligands (I), and one TEMPO radical (M). **B.** Chemical structure of the oligomers (n= 5 or 6). **C.** Hydrogen-bond pairing of the recognition units. **D.** Sequence of the oligomers synthesized in this study; two complementary chains of the O_a/O_{b1} system contain all the needed catalytic groups when paired (2 × P, 2 × I, 1 × M), whereas O_a/O_{b2} or O_a/O_{b3} paired chains lack one functional group, *i.e.*, 1 P or 1 I, respectively. **E.** Equilibrium constitutional library of co-existing species in solutions.

4.3 Results and Discussion

4.3.1 Binding constants of the recognition units

The two monomeric pairs can form a trivalent hydrogen-bonded duplex, with two donor sites coming from one monomer and one donor site from the other monomer, as illustrated in Figure 4.2.

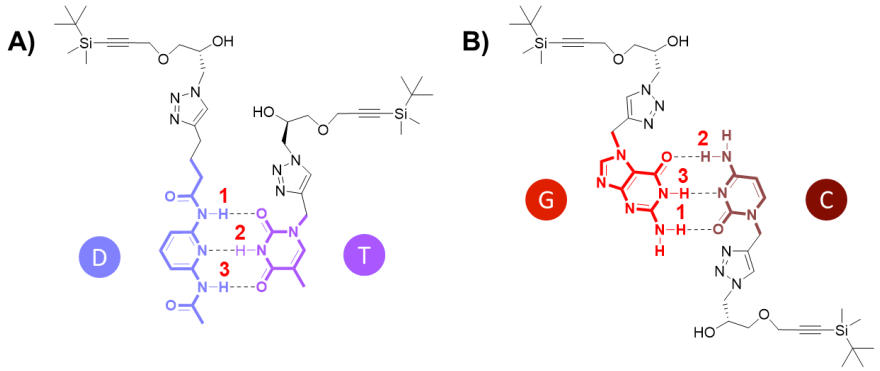


Figure 4.2 Chemical structure of the trivalent hydrogen-bonded duplex of the monomeric pairs.

In addition to the desired trivalent duplex, monomers can form a series of other hydrogen-bonded complexes. Here, we restrict ourselves to measurements performed in hydrogen bond-accepting solvents; therefore, the generic acceptor sites of the monomers are completely negligible compared to the vast amount of accepting sites of the solvent.

NMR titration and dilution experiments at different temperatures were used to investigate the H-binding properties of the monomers in mixtures of acetonitrile (ACN) and dimethyl sulfoxide (DMSO). The detailed analysis is available in Annex 4.2. The binding constants and standard binding enthalpies and entropies of the different monomer

units at 295 K are given in Table 4.1.

Table 4.1 Binding constants and standard binding enthalpies and entropies of monomeric pairs in deuterated ACN/DMSO at 295 k.

System	ACN/DMSO (v:v)	*K* (L/mol)	ΔH^o (kJ/mol)	ΔS^o (J/K·mol)
D/T	100:0	16 ± 1.7	-21 ± 5.7	-48 ± 19
	95:5	3.3 ± 0.2	-6.1 ± 1.8	-10.8 ± 6
	5:1	0.99 ± 0.06	-5.7 ± 2.3	-19.4 ± 8
G/C	5:1	104 ± 15	-24 ± 4	-42.7 ± 14

* The errors on the fit parameters were estimated from the diagonal terms of the covariance matrix, assuming a standard error on NMR shifts of 0.01 ppm.

Due to the limited solubility of the G/C monomeric pair, titration experiments could not be performed in deuterated ACN/DMSO (v:v, 95:5), which is the solvent (in non- deuterated form) used for catalytic experiments. Therefore, to predict the binding constant of the G/C monomeric pair in deuterated ACN/DMSO (v:v, 95:5), the effect of DMSO content should be understood. For this, the effect of the DMSO content on the formation of the D/T duplex was measured, as summarized in Figure 4.3. It is obvious that lower DMSO contents lead to stronger binding as expected. For the D/T duplex, the reduction of concentration in DMSO from 16 to 5% results in a multiplication of the binding constant by a factor of 3.3/0.99≈3.3. If this correction factor is applied to the G/C pair as well, the predicted binding constant of the G/C pair in ACN/DMSO (v:v, 95:5) should be *ca*. 343 L/mol. Since the binding constant of D/T at 295 K in ACN/DMSO (v:v 95:5) obtained from the dilution experiment is 3.3 L/mol (Table 4.1), the binding constant of the O_a/O_b pocket due only to

independent D/T and G/C base pairing is obtained by the product of the binding constants of the two pairs, *ca*. 1120 L/mol, corresponding to a free energy of binding of *ca*. -17.2 kJ/mol at 295 K. These parameters are large enough to create a relatively stable pocket, compared to thermal energy *RT* = 2.5-2.8 kJ/mol between 30 and 60°C.

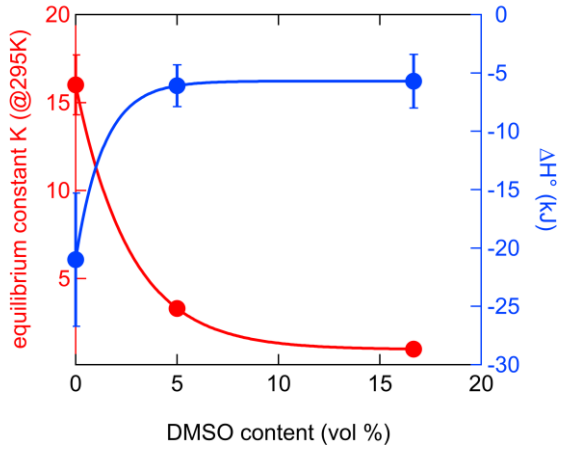


Figure 4.3 The effect of DMSO content on D/T duplex formation at 295 K.

4.3.2 Molecular dynamics simulations

MD simulations of one O_a and one O_{b1} chains placed in an effective solvent were performed as well as molecular network analysis to predict conformation and interaction of chains. This was done in the group of Mathieu Surin in UMons; details are in Annex 4.3.

MD simulations of one O_a and one O_{b1} chains placed in an effective solvent indicate the rapid assembly of the oligomers in a stable heteromolecular complex of *ca.* 1 nm radius of gyration as Figure 4.4.A shows (details in Annex 4.3 as well as ESI Section 10 of our paper).[201] Inside this compact globule, the two strands remain flexible and highly dynamic, allowing the different units to rearrange in space

due to a vast network of interactions involving functional moieties (catalytic groups and nucleobase analogs) as well as the carbamate and triazole moieties of the backbone. Interestingly, the catalytic units are among the less buried residues and are accessible at the periphery of the globule (Figure A.4.6).

Network theory applied to the collection of conformations generated over a 10 μs timescale indicates that interchain connectivity arises dominantly through interactions between complementary bases, whereas intrachain connectivity replicates the primary structure of the chains as shown in Figure 4.4.B. Other parameters of the network such as betweenness and average shortest path length provide further support to the formation of a globule stabilized by interactions between nucleobases (Figure A.4.8). A modular representation of the network shows that the catalytic modules (M, I', I, P in Figure 4.4.C) are connected *via* the connectivity between the bases. Analysis of MD simulations indicates that the formation of this supramolecular complex is driven both by hydrogen bonding and π-stacking interactions (Figure 4.4.D).

Heatmaps enumerating the frequency of hydrogen bonds (Figure 4.4.E) show that C---G and D---T pairings provide an important but not exclusive contribution to the persistence of the globular assembly, with D---G pairing providing a significant but less frequent mechanism of stabilization. Stacking interactions are far less specific, involving numerous residues including triazole groups from the backbone and aromatic catalytic units (I', I, P) (Figure A.4.5).

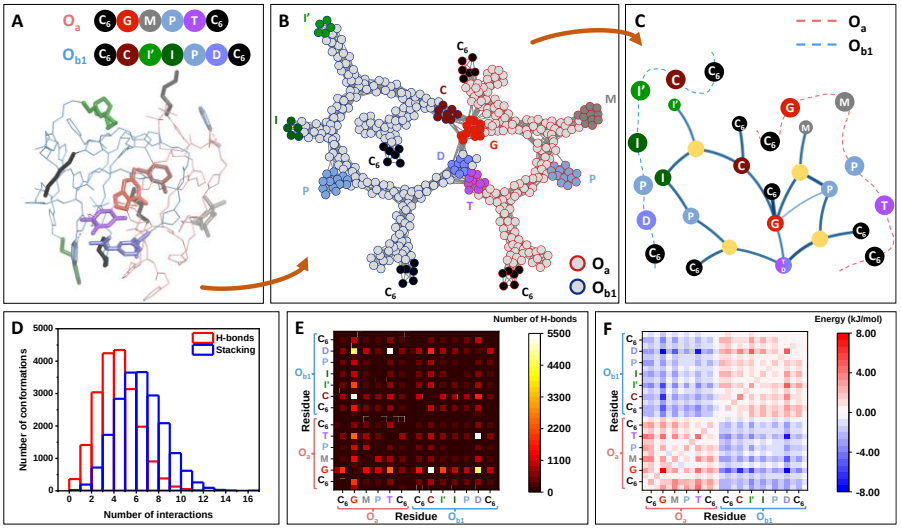


Figure 4.4 Molecular simulations of the self-assembly of O_a and O_{b1} oligomers. A. Molecular dynamics snapshot of a 3D conformation of the complex, representative of the globular conformation. The two strands, O_a and O_{b1}, are depicted in pink and blue lines, respectively, and the functional units are depicted in sticks and coloured according to the ball representation as sketched above. B. Network representation of the system, highlighting the persistent contacts observed during the MD simulation. Atoms (except H) constitute the nodes of the network. The nodes belonging to the strand O_a and O_{b1} are circled in red and blue, respectively, with the same color code for the functional groups. C. Modular representation of the network. The modules in yellow contain backbone or chain-ends nodes. For clarity, the primary structures of the two strands are sketched in dashed lines, showing that the sequence order is preserved in this representation, except for the complementary nucleobase analogs T and D, which are merged in the same module. D. Distribution of the number of hydrogen-bonding and stacking interactions for the 20,000 conformations from MD, showing that both H-bonds and aromatic interactions strongly participate to the stabilization of the system. E. Heatmaps showing the decomposition of the hydrogen-bonds by pair of residues. Each square, localized at the crossing of a pair of residues, indicates the number of interactions detected between the two residues over the 20,000 conformations. White squares indicating the highest number of

interactions are localized for the pairs G---C and T---D. Interactions are also found for the pair G---D. F. Heatmaps showing the decomposition of the enthalpy of binding by pair of residues, similarly as for Figure 2E. The most stabilizing residue-residue interactions (dark blue squares) are localized for the pairs G---D, T---D and G---C.

Heatmaps of the enthalpy of binding per residue provide a detailed view on the stabilization per residue pair, globally showing the stabilization of the supramolecular complex by interchain interactions between the different bases. This indicates that when forming the complex, intermolecular interactions outperform the intramolecular interactions of each strand when alone. The same analysis was performed on the O_a/O_{b2} system and provided similar results, with the same type of interactions stabilizing the formation of a globular complex (Figure A.4.8).

4.3.3 Aerobic oxidation of benzyl alcohol by the self-assembled oligomers

The catalytic systems were tested in the aerobic oxidation of benzyl alcohol (BnOH) into benzaldehyde (BzH), for different temperatures between 30 and 60°C and at different molar concentrations of catalyst. $Cu^{(I)}$ was introduced in stoichiometric amount with respect to P groups, and catalyst concentration is expressed as the content in M units relative to the molar concentration of introduced alcohol (0.2 mol/L). Gas chromatography was used to monitor the appearance of BzH; no other oxidation product could be found. Previous studies indicated that the competition of urethane and triazole motifs for Cu complexation is limited compared to P; additionally, competition of base analogs for copper binding is also limited according to NMR (Annex 4.4). Results from a typical catalytic experiment are displayed

in Figure 4.5.A; similar graphs for all other conditions are in Annex 4.5. From these kinetic data, the turnover frequencies (TOFs) were obtained at different temperatures for the different oligomeric systems at different concentrations. The TOF is defined as the slope versus time of the BzH molar concentration at moderate conversion rates, divided by the molar concentration in catalyst; it represents the catalytic efficiency per mole of catalyst at low to moderate conversion.

The catalytic activity of a single oligomeric chain (O_a, O_{b1}, O_{b2} alone) was first tested for the aerobic oxidation of BnOH. The limited catalytic activity of a single strand is evident in Figure 4.5.D, particularly at low temperatures. High temperatures accelerate the chain dynamics and facilitate the crossing of energy barriers, thereby generally increasing activity. At high temperatures, the presence in the chains of TEMPO· which acts as a cooperating oxidizing agent makes the catalytic activity of the TEMPO-containing single O_a chain relatively higher compared to chains without TEMPO. However, in any case, the oligomer alone lacks some of the necessary catalytic groups, which explains the poor catalytic activity.

Based on the above results, the complementary oligomer was added to the corresponding single oligomer chain. The catalytic activities of these systems with complementary oligomers are very different depending on whether the two chains together make up a full catalytic system or not, as shown in Figure 4.5. Indeed, the collection of poly(oligomers) of the incomplete systems in O_a/O_{b2} and O_a/O_{b3} solutions exhibit a very low TOF despite collectively comprising all the groups needed for catalysis, as Figure 4.5.C shows. In contrast, the library of poly(oligomers) of the complete O_a/O_{b1} system exhibits a much higher TOF at low concentration displayed in Figure 4.5.B. In addition, the catalytic rate and catalytic efficiency per mole of catalyst

improve as the temperature increases for the different self-assembled systems, as expected. However, the performance within the concentration range varies among these systems. The optimized catalytic concentration for the complete self-assembled catalytic system O_a/O_{b1} falls within the range of 1 mol%. Conversely, in the case of the incomplete catalytic systems O_a/O_{b2}, the efficiency of the catalyst per unit of catalyst improves continuously as the concentration increases. According to the simulation, O_a and O_b can form a self-assembled pocket no matter what the catalytic center is. In this case, the low catalytic activity of O_a/O_{b2} and O_a/O_{b3} system is thus due to incompleteness of the catalytic center (missing one catalytic unit).

Additionally, to investigate the influence of covalent bonding and self-assembly, a catalytic system made of monomer units with the same constituent units as the complete O_a/O_{b1} system was introduced. The catalytic activity of the separated monomeric catalytic system is shown as crosses in Figure 4.5.D. Its performance rapidly decreases upon dilution, owing to the decrease of the probability of chance encounter of the catalytic groups. This is in marked contrast with the complete O_a/O_{b1} catalytic system that resists dilution and intriguingly exhibits an optimum close to 1 mol% (Figure 4.5.B). Meanwhile, a catalytic system composed of dimer MP and trimer I'IP, which possesses the same catalytic center structure as the complete O_a/O_{b1} catalytic system but lacks the recognition units, was also tested to make a comparison. Although the system consisting of the O_a and O_{b1} oligomers devoid of H-binding G, T, D and C units and of hydrophobic C_6 units (pentagons in Figure 4.5.D) is very effective at high concentrations due to its limited steric crowding, the TOF rapidly decreases with concentration and becomes less efficient than the hydrogen-bonded system below 1 mol% of catalyst. This

phenomenon is caused by the lesser probability for the catalytic groups to meet each other when diluting in the absence of assembling units. In contrast, the assembling units serve to keep the probability of meeting high, even at low dilutions.

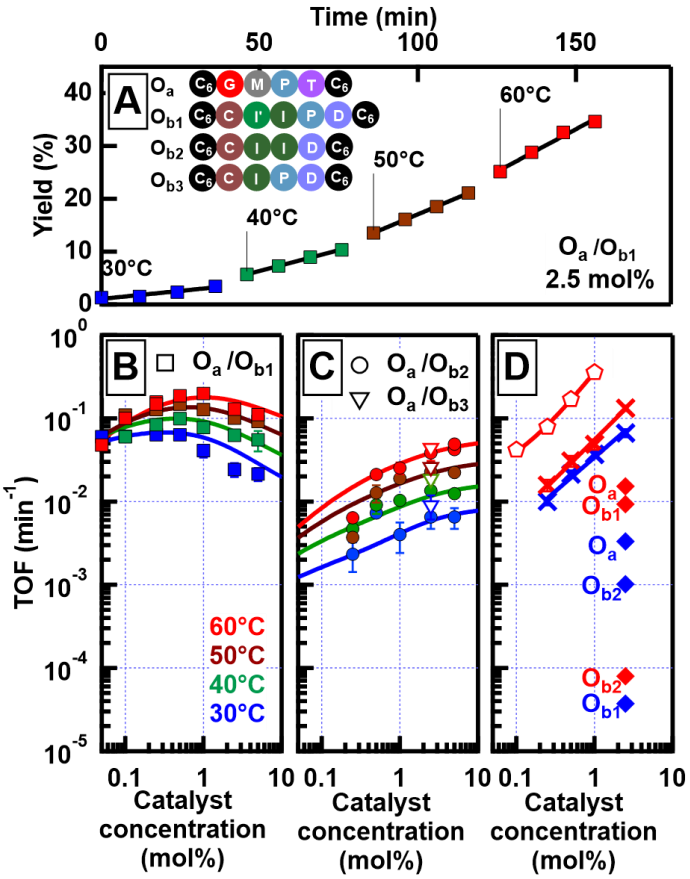


Figure 4.5 Catalytic properties of the libraries of self-assembled precise oligomers. The catalyst concentration is expressed in mol% of M units relative to a reference concentration of 0.2 M BnOH in ACN:DMSO 95:5. A. Typical catalytic experiment: BzH yield versus reaction time for the conversion of BnOH (0.2 M) in the presence of oxygen, 2.5 mol% of the complete O_a/O_{b1} catalytic system (with CuI), at four different temperatures. The slopes give the catalytic rates which, once divided by the concentration in catalyst, provide the TOF. B-D. TOF versus catalyst concentration at four

temperatures (colors as indicated), for (B) the complete O_a/O_{b1} system, (C) the incomplete O_a/O_{b2} (circles) and O_a/O_{b3} (open triangles) systems, (D) a mixture of monomer units with the same composition as the complete oligomeric system (crosses), an equimolar mixture of MP dimers and II'P trimers lacking the hydrogen-binding groups and hydrophobic C_6 units of O_a/O_{b1} (pentagons), and single oligomer chains at 2.5 mol% in the presence of an equimolar amount of CuI (diamonds – blue 30°C, red 60°C). The continuous curves fit the model of Figure 4.1E to the data, except in panel D where they are power law fits.

The complete O_a/O_{b1} system is thus peculiar in its specific ability to maintain catalytic activity at low dilutions. Whereas higher catalytic activities can certainly be obtained for the other systems simply by introducing larger amounts of catalytic components (such as for the system made of separated monomer units or the MP/I'IP system), the complete system is unique in its preservation of catalytic efficiency when diluted.

4.3.4 Mathematical models of the composition and properties of the constitutional libraries

To explain why the complete O_a/O_{b1} system has such special performance, we tried to analyze the composition of the system. The oligomer chains dynamically bind in a variety of configurations shown in Figure 4.1E, resulting in the coexistence of different species including free oligomeric strands, linear poly(oligomer)s of varying length and chain ends, and di(oligomeric) cycles. In the incomplete O_a/O_{b2} and O_a/O_{b3} systems, self-assembly configurations comprising at least three oligomers generally contain all groups needed for catalysis, whereas di(oligomeric) configurations do not; as shown in Figure 4.5C, these self-assembly configurations only provide

mediocre catalytic efficiency at low catalytic concentration. The fully different behavior of the complete O_a/O_{b1} system thus necessarily arises from di(oligomer) assemblies, which now comprise all needed catalytic groups. Since these di(oligomeric) assemblies tend to easily form self-assembled macrocycles as shown by molecular dynamics simulations, it follows that the high catalytic efficiency at low concentrations of the O_a/O_{b1} system arises from the formation of di(oligomeric) cycles.

Despite the virtually-infinite number of species in each constitutional library, their composition and catalytic activity can be modeled in a compact form by a mathematical model. Besides, the NMR data in Figure A.4.10 (Annex 4.4)**.** indicates that competition of base analogs for copper binding is also limited. However, this limited competition probably contributes to a decreased activity of catalytic sites, therefore, it needs to be taken into consideration in setting up the mathematical model. In brief, the mathematical model comprises a small set of physically-meaningful parameters including two binding constants, K_1 and K_2 associated to chain end pairing and a standard entropy penalty ΔS_c° for cycle formation, common to all libraries, temperatures and concentrations; the details are provided in Annex 4.6. The concentration in each species can be computed numerically from this model, as well as their contribution to the turnover frequencies at each temperature and concentration. This compact mathematical model was simultaneously fitted to all experimental catalytic turnover frequencies measured at different temperatures and concentrations, for the two O_a/O_{b1} and O_a/O_{b2} systems taken together, leading to a quantitative reproduction of the experimental data (continuous line in Figure 4.5 B and C).

However, the model needs to be validated by an independent measurement of the concentration of each species in the solution. These species are in constant dynamic equilibrium and cannot be isolated; furthermore, the oligomers they contain are involved in similar interactions in all species. Their concentrations are thus very difficult to measure by any available technique. Nevertheless, each species possesses its own molar mass and hydrodynamic volume. The diffusion coefficient of a proton attached to an oligomer included in a species of the constitutional libraries depends directly on the hydrodynamic volume of this species, which can be computed from its composition in each oligomer and from the hydrodynamic volume of the oligomers. Additionally, the diffusion coefficient averaged over all co-existing species only depends on the hydrodynamic radii and concentrations in solution of these species. Hence, measuring the average diffusion coefficient of the protons provides an elegant way to check *in situ* the compositions of the constitutional libraries predicted by our model.

The measured and predicted diffusion coefficients of the complete O_a/O_{b1} system were compared, using as parameters for the prediction the fit parameters obtained from the analysis of the catalytic data in Figure 4.5. The DOSY data is available in Annex 4.7. As shown in Figure 4.6 (continuous line), the experimental data agree very well with the predicted values assuming good solvent conditions. This model was thus used to compute the distribution of O_a oligomers between the different species for the O_a/O_b systems, depending on temperature and concentration (Figure 4.7 top row). Oligomers are essentially free at very low concentrations, whereas they dominantly belong to linear poly(oligomeric) chains above 1 mol% of catalyst.

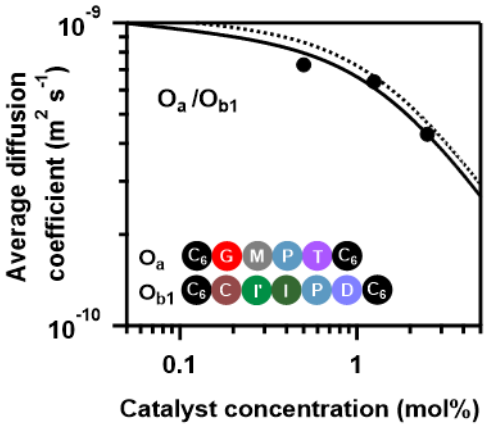


Figure 4.6 DOSY-measured average diffusion coefficient of the protons of the oligomers in the complete O_a/O_{b1} system, depending on the concentration in M units (circles). The standard error is smaller than the point size. The curves are predictions based on the parameters from the fit of the catalytic activity, assuming either good (continuous curve) or Θ (dotted curve) solvent conditions.

The decrease of the binding constants K_1 and K_2 with temperature explains why the average degree of poly(oligomerization) of the linear chains slightly decreases with temperature (inset in top row of Figure 4.7); however, the total amount of linear chains is very insensitive to temperature (Figure 4.7, top right). Di(oligomeric) cycles slightly dominate the constitutional libraries at intermediate concentrations and lower temperatures, but this dominance is progressively lost as temperature increases. The decreased amount of cycles as temperature increases essentially results from the increased free energy cost associated to the entropy penalty for cycle formation (ΔS_c°).

The contributions of each species to the total catalytic turnover

frequency of the complete O_a/O_{b1} system, which is the product between the amount of O_a in each species and the turnover frequency per O_a, are collected in the bottom row of Figure 4.7. The contribution of supramolecular di(oligomeric) cycles to the total turnover frequency is the dominant factor that explains the high catalytic activity of the complete system, even when the supramolecular cycles do not dominate the composition (*e.g.*, at 60°C). At this higher temperature, the relative proportion of cycles decreases but their catalytic activity increases due to thermal activation, resulting in an overall dominance of the cycles in the catalytic performance.

Cycles are thus much better than linear chains to reach a high catalytic activity, despite both configurations having all required active groups. At the concentration maximizing di(oligomeric) cycles, the ratio between the contributions to the catalytic activity of cycles versus all linear chains is *ca.* 30 and 10, at 30 and 60°C, respectively. This is due to the much higher probability for the catalytic groups to be in an active site when assembled in a supramolecular cycle, compared to when in linear chains. This demonstrates that, by assembling the complementary strands in a supramolecular cycle, and thereby maximizing the probability that the five required groups meet in space, dramatic improvements of the catalytic efficiency by more than one order of magnitude can be obtained compared to linear configurations. In contrast, for the incomplete O_a/O_{b2} system, supramolecular cycles cannot form complete catalytic sites and therefore do not contribute to the catalytic activity; consequently, this system only exhibits low activity associated to the low probability of formation of complete catalytic sites in linear poly(oligomeric) chains.

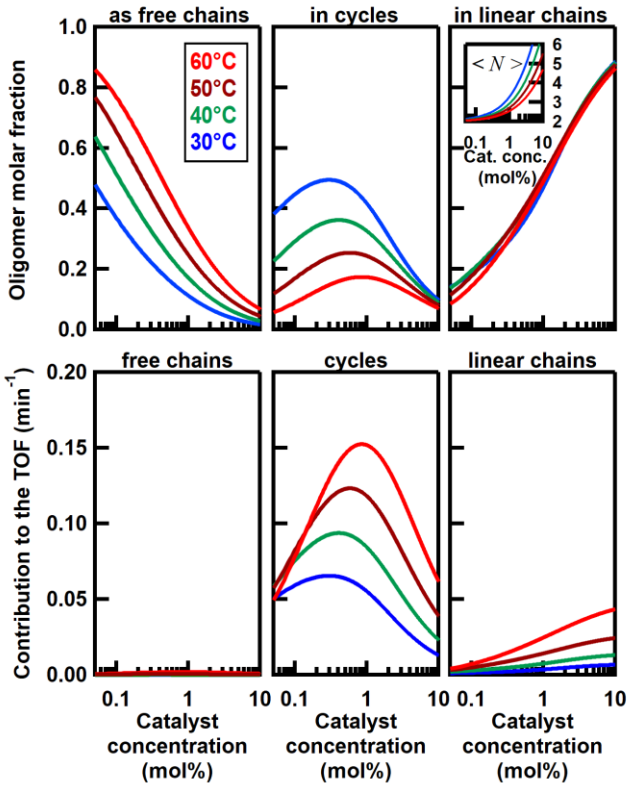


Figure 4.7 Top row. Distribution of O_a oligomers between the different co-existing species in the O_a/O_{bi} systems (i=1 or 2), depending on temperature and concentration, computed from the model. The inset in the right graph is the average degree of poly(oligomerization) by number of the linear chains *vs* catalyst concentration. Bottom row. Contribution of the different co-existing species to the catalytic turnover frequency, in the complete O_a/O_{b1} system.

4.3.5 Verification of the need to form binuclear copper complexes for catalysis

The di(oligomeric) cycles formed in the O_a/O_{b1} system comprise two P bidentate ligands, two I auxiliary ligands, and one TEMPO radical. In contrast, the supramolecular cycles of the O_a/O_{b2} system lack one P-

Cu, whereas the supramolecular cycles of the O_a/O_{b3} system lack one I, as shown in Figure 4.1. However, the O_a/O_{b1} system is much more efficient compared with systems where one unit is missing. In agreement with the catalytic models in section 4.3.4, a possible reason is that a dinuclear copper complex needs to be formed in the catalytic cycle. Therefore, experiments were performed in which one equivalent of monomer P and copper was added to the O_a/O_{b2} system, and one equivalent of *N*-methyl imidazole (analogue to I) was added to the O_a/O_{b3} system for a concentration in M units of 2.5 mol%. An experiment in which one equivalent of *N*-methyl imidazole was added to the O_a/O_{b2} system was also performed.

The TOFs of the different systems were compared before and after addition of missing units in Figure 4.8. The incomplete O_a/O_{b2} and O_a/O_{b3} systems have a much lower catalytic efficiency than the complete O_a/O_{b1} system. Adding one P-Cu equivalent to the O_a/O_{b2} system brings the TOF close to the one of the complete system, whereas adding one NMI equivalent (light blue) does not change significantly the TOF. Two ligands are needed in the catalytic cycle, to form the binuclear copper complexes.[143] Therefore, for the O_a/O_{b2} system containing only one P-Cu unit, the addition of free P-Cu contributes to the forming of the binuclear copper complex which significantly improves the TOF of O_a/O_{b2} system. However, the addition of NMI to O_a/O_{b2} system does not change significantly the TOF, which is expected since the two needed I units are already available in the self-assembled di(oligomeric) cycles. In contrast, adding one NMI equivalent to the O_a/O_{b3} system which lacks one I unit (light green) brings the TOF close to the one of the complete system. The improvement of the TOF following the addition of the NMI base results from the stabilization of the binuclear copper complex.[143,156] In each case, the addition of the missing unit results in

an efficiency similar to the one of the complete system, which strongly supports the need to form a binuclear copper complex involving two bidentate P ligands and two auxiliary I (or NMI) ligands.

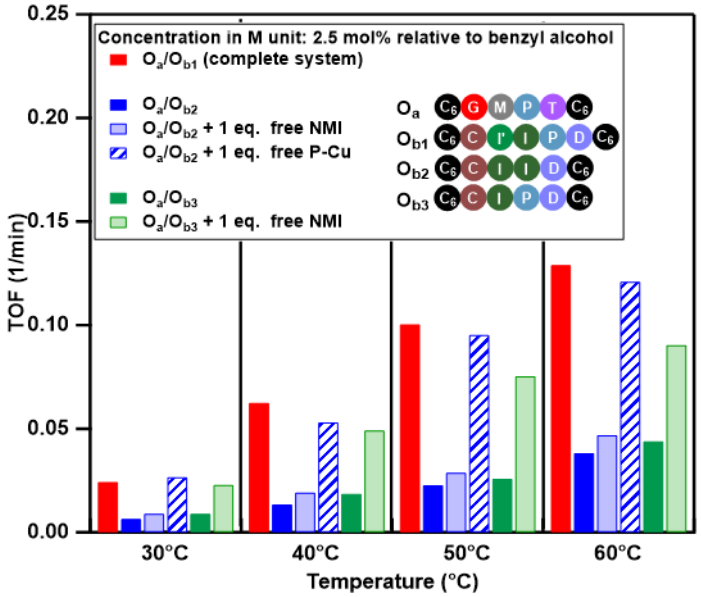


Figure 4.8 TOF of the three systems measured at different temperatures and 2.5 mol% concentration in M units. For the O_a/O_{b2} and O_a/O_{b3} systems which form di(oligomeric) supramolecular cycles that lack either one P-Cu or one I unit, one equivalent of free P-Cu and/or free NMI (analogue to I) was added as indicated.

4.3.6 Aerobic oxidation of other alcohols

In addition to the aerobic oxidation of BnOH, the complete O_a/O_{b1} system was also utilized in the oxidation of other alcohols. The TOFs of O_a/O_{b1} system for different alcohols are shown in Figure 4.9. The catalytic results show that the complete O_a/O_{b1} system is applicable for different primary alcohols, even though the catalysts are more efficient with the activated alcohols like benzylic and allylic alcohols. It may be due to the adjacent aromatic ring or double bond which

stabilizes the transition state through conjugation to help the formation of intermediates during oxidation.[156,160] This catalytic behavior is similar to the results of other reported metal/TEMPO systems.[155,202,203] Though the catalytic efficiency on these alcohols is different, its temperature-dependence is preserved. Further mechanistic studies are needed to explain these differences. Since our catalyst is chiral, the aerobic oxidation of chiral alcohols should be interesting. However, the oxidation of secondary alcohols is usually much slower than primary alcohols; therefore, in this work, we only focused on the oxidation of primary alcohols.

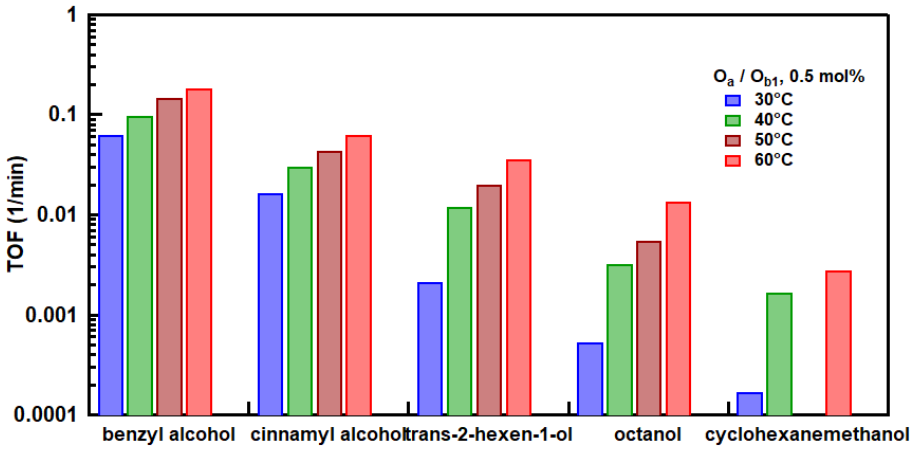


Figure 4.9 TOF of O_a/O_{b1} system catalysing the oxidation of different alcohols, measured at four different temperatures and 0.5 mol% concentration in M units.

4.3.7. A word on copper

Copper is a crucial component of the catalyst. However, as we already discussed in sections 4.3.3 and 4.3.4, the bases may compete with P as a ligand of copper, which would limit its availability in the catalytic cycle. On the other hand, the complexation of copper

with bases might favor the assembly of the complementary chains. It would thus have been interesting to include copper in the molecular dynamics simulations of the assembly of the oligomers. However, the numerous oxidation states of copper made this simulation too complex at this stage; this is clearly a limit of our study.

4.3.8 Comparison of the self-assembled catalysts with other systems

It is tempting to compare the efficiency of our supramolecular catalyst to other systems based on the same composition, in which the different catalytic groups are positioned differently in space.[204] Depending on the sequence of units along chains[84,130,205] and on their grafting configurations when heterogeneized,[128–130,164] catalysts based on M, I and P units have been shown to exhibit very different catalytic efficiencies, as judged from their initial TOF. Compared to systems grafted on porous silica beads, the current supramolecular cycles are generally slightly inferior, probably because of a smaller accessibility of the substrates in the crowded globules of the cycles. In contrast, at 60°C and below 1 mol%, the current supramolecular catalysts are better than any other homogeneous system tested so far in our group; this advantage is rapidly lost as concentration increases, because supramolecular cycles are disfavoured at higher concentrations while the probability of encounter is increased for other systems.

4.3.9. Predictions for improved activity

As already discussed in section 4.3.4, there are three different species co-existing in a solution, but the catalytic activity is dominated by the di(oligomeric) cycles. To obtain better catalytic activity for the system, strategies should thus be done to form more di(oligomeric)

cycles. The mathematical model proposed in this thesis can be used to predict directions for a larger range of conditions favouring the formation of the catalytically-active self-assembled di(oligomeric) cycles. For instance, increasing the chain end pairing constants by a factor of *ca.* 10 compared to their current geometric average value at 30°C would result in a solution consisting almost exclusively of supramolecular cycles below 1 mol% catalyst. To favor cycles at higher concentrations, which would result in higher absolute catalytic rates, a supplementary stabilization mechanism must be given to cycles compared to linear chains. According to the mathematical model, the number of the di(oligomeric) cycles in co-existing species is related to the binding constants of the base pairs and the entropy cost of cycle formation (ΔS_c°). Therefore, one way to achieve this is to reduce the entropy cost of cycle formation (ΔS_c°), which might be attained by increasing the rigidity of the oligomer backbones, for instance by introducing more aromatic or rigid cyclic groups in the backbones. However, this might be counter-productive with respect to the probability to have the catalytic groups meet in space. Another way would be to provide a supplementary enthalpy of stabilization specific to macrocycles. One could think of using higher binding affinity pairs like strong DDDD-AAAA quadruple hydrogen-bonding arrays; however, this would simultaneously contribute to stabilize both di(oligomeric) cycles and linear poly(oligomeric) chains. In order to favor cycles over linear chains, one might attempt to use different pairs of recognition units at each ends of the oligomers, which would create a directionality in the pairing of ends and might thereby favor cycles over linear chains.

4.4 Conclusions

Our study investigated the self-assembly of complementary oligomers into a dynamic library of poly(oligomeric) species, most of which contribute little to catalytic activity. In this dynamic library, di(oligomeric) cycles were demonstrated to be highly catalytically-active, provided the two chains together comprise all the needed elements for catalysis, as in the O_a/O_{b1} system. When only one element is missing as in O_a/O_{b2} and O_a/O_{b3}, the activity drops very significantly, despite all needed elements are present in the solution. The activity can be recovered by adding the missing unit as a free component in solution. These results support the assumption[35] that five functional groups, including a dinuclear copper complex, are involved in the catalytic cycle.

Molecular dynamics simulations demonstrate that these self-assembled cycles form globules in which different interactions contribute to stabilization, even though the base analogs at the ends of the strands provide a continuous stabilization mechanism and connectivity between the catalytic groups, as shown by network analysis. Measurement of hydrodynamic radius also provides experimental support for the formation of di(oligomeric) cycles.

Hence, the primary structure of the oligomers is important for the emergence of catalytic activity, because the presence or absence of the needed catalytic groups on the strands directly affects the efficiency of the di(oligomeric) cycles. However, it is likely that the sequence order is not a critical factor at this stage, considering the limited length of the investigated oligomers. Nevertheless, quaternary structures are crucial for the emergence of catalytic activity, which here arises dominantly from the self-assembly of the strands in di(oligomeric) cycles. Interestingly, the self-assembly of di(oligomeric)

cycles happens in a relatively dilute range of concentrations, in which cycles totally outperform simple mixtures of catalytic groups or any combination of the catalytic groups in linear chains. This demonstrates that an efficient catalytic activity can be obtained at rather high dilutions in the presence of properly self-assembled structures.

Our study also demonstrates that a small number of experimental parameters is sufficient to describe the mathematical complexity of the dynamic constitutional library of species resulting from the assembly of complementary oligomers. Indeed, the reproduction of catalytic activities and diffusion coefficients by a compact mathematical model is an important result, showing that the complexity of this system is still amenable to a deterministic description.

Annex 4.1 Experimental conditions

Reagents and solvents were obtained from commercial sources and used without further purification. All reactions were carried out under argon atmosphere. Flash column chromatography was carried out using silica gel 230-400 mesh (Sigma-Aldrich) as the stationary phase. Milli-Q water (resistivity 18.2 MΩ.cm) was obtained from a Merck Millipore system. NMR spectra were recorded on Bruker-300 and Bruker-500 spectrometers. Chemical shifts (δ) are reported in parts per million (ppm) from low to high field and referenced to residual solvent. Coupling constants (J) are reported in Hertz (Hz). Standard abbreviations indicating multiplicity are used as follows: br = broad, s = singlet, d = doublet, t = triplet, q = quartet, quint = quintet, m = multiplet. 2D NMR DOSY experiments were recorded on a Bruker Avance 500 MHz Spectrometer equipped with a 5 mm BBO probehead. All NMR samples were equilibrated at the measurement temperature of 303 K for 5 min before data collection. Spinning was deactivated to avoid convection. ^{1}H NMR spectra were acquired with the zg pulse program from the Bruker library using 32 scans and calibrated to the CD_3CN residual signal (δ = 1.94 ppm) as an internal reference.

2D ^{1}H DOSY experiments were performed using the standard Bruker ledbpgp2s pulse program with 16 t1 increments on 32 K data points. The acquisition time was 2.04 s and the relaxation delay was 10.0 s (D1). Diffusion time was set between 0.2 – 0.1 s (D20) and rectangular gradient pulse duration was between 1.0 – 0.8 ms (P30). Gradient recovery delays of 0.15 ms followed the application of each gradient pulse. Data was accumulated by linearly varying the diffusion encoding gradients over a range from 2 to 95 % for 16 gradient

increment values with 32 scans each. Individual rows of the quasi-2D diffusion databases were phased and baseline corrected. NMR data processing was performed using Topspin software, and the actual diffusion coefficients used for diffusion analysis were measured by T1/T2 relaxation module.

The diffusion coefficient at 30°C measured over different displacement ranges corresponding to protons of the oligomers were averaged to obtain the average diffusion coefficient $\overline{D}$. The hydrodynamic radii of the single oligomers were obtained from DOSY measurements performed at 30°C and 1 mM to avoid aggregation; the radii were computed from the Stokes-Einstein equation using $\mu_0 = 0.355$ mPa.s for the dynamic viscosity.[1] The average diffusion coefficient of equimolar mixtures of two oligomers were measured for a molar concentration in each oligomer of 1, 2.5 and 5 mM, corresponding to 0.5, 1.25 and 2.5 mol% relative to the (virtual) concentration of 0.2 M in alcohol used in the catalytic experiments.

High-resolution mass spectra (HRMS) were measured on a Q-Exactive (Orbitrap) from ThermoFisher using an atmospheric pressure chemical ionization (APCI) source. Electrospray ionisation mass spectrometry (ESI-MS) and ESI-MS/MS were performed on an SYNAPT G2-Si high definition mass spectrometer (Waters) equipped with a NanoLockSpray dual electrospray ion source (Waters). Precut fused silica PicoTipR Emitters (outer diameters: 360 µm; inner diameter: 20 µm; 10 µm tip; 2.5” length (Waters)) were used to carry samples for nanoelectrospray injecting the test solution.

LC-QTOF-MS/MS analysis were performed on an SYNAPT G2-Si high definition mass spectrometer (Waters) equipped with a NanoLockSpray dual electrospray ion source (Waters). Samples were

diluted 2 to 10 times in 50% (v/v) acetonitrile, 0.1% formic acid or in 50% (v/v) methanol, 0.1% formic depending on the quality and the intensity of the signal. Coated fused silica PicoTip™ Econo12 Emitters for nanoelectrospray, outer diameters: 1 µm Tip (New Objective, USA) were filled with 5 µL of samples and placed on the Universal NanoFlow™ Sprayer (Waters). The eluent was sprayed at a spray voltage of 2.8 k. The source temperature was set to 100 °C. The cone gas flow was 20 L/h with a nano flow gas pressure of 0.3 bar. MS spectra were acquired and processed with MassLynx software (Waters). Full scan MS and MS2 spectra (*m*/*z* 50 to 2000) were in resolution mode (20,000 resolution FWHM at *m*/*z* 400) with a scan time of 0.1 sec. Tandem mass spectra of the precursor were generated in the trapping region of the ion mobility cell by using a collision energy ramp from 10 V to 70 V. Charged ions are selected to be submitted to the MSMS fragmentation over the *m/z* range from 50 to 2000 with a scan time of 0.25 sec. For the post-acquisition lock mass correction of the data in the MS method, the doubly charged monoisotopic ion of [Glu^1]-fibrinopeptide B was used at 100 fmol/µL using the reference sprayer of the nanoESI source with a frequency of 30 s at 0.5 µL/min into the mass spectrometer.

Aerobic oxidation experiments were performed from 30 °C to 60 °C under O_2 bubbling (4.93 mL/min) using a SLA5850 thermal mass flow controller from Brooks, in a 0.2M solution of alcohol in the mixture solvent of acetonitrile (95%) and dimethyl sulfoxide (5%) (2 mmol of alcohol in 1 mL mixture solvent). Different loadings of catalytic oligomers were used, expressed in mol% relative to the amount of alcohol, keeping a 1:1 pyridyltriazole (P):Cu molar ratio (except for single oligomers missing P units, in which case Cu was introduced in equimolar amounts as the oligomer). Before the experiments, different catalytic oligomers were dissolved in acetonitrile or

acetonitrile and dimethyl sulfoxide mixture solvent, and the CuI was separately dissolved in acetonitrile at a suitable concentration (stock solutions). For catalytic experiments, alcohol and *p*-xylene (used as internal standard) was introduced into 2 mL vials, followed by catalytic oligomers and an appropriate amount of acetonitrile and dimethyl sulfoxide. The mixture was stirred to form a uniform solution, then a controlled amount of CuI solution was added into the mixture under stirring. The reaction time was counted when the mixture was at the set temperature and O_2 was introduced. Although the solution was turbid just after copper salt addition, it turned transparent and greenish during reaction (Figure A.4.1).

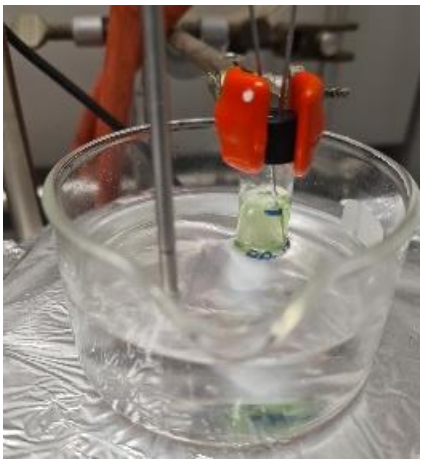


Figure A.4.1 Picture of the catalytic medium during reaction at 30°C for the O_a/O_{b1} system, with 2.5 mol% M units.

10 µL aliquots were then taken at selected time points and diluted by 990 µL acetonitrile to prepare a gas chromatograph (GC) sample. GC was performed on a Shimadzu GC-2010 equipped with an FID detector, with *p*-xylene (internal standard) only used for the detection of possible experimental problems. Benzaldehyde, benzyl alcohol,

cinnamaldehyde, cinnamyl alcohol, *trans* 2-hexen-1-al, *trans* 2-hexen-1-ol, octanal, octanol, cyclohexanecarboxaldehyde and cyclohexanemethanol were detected at 9.1, 11.3, 5.74, 5.34, 7.47, 8.0, 3.1, 3.5, 3.2- and 2.9-min elution time, respectively. The yields were computed as the ratios between the area below the aldehyde peak and the sum of the areas below the aldehyde and alcohol peaks. No other oxidation product than the aldehyde was detected in the chromatograms. The slope of the yield versus time was defined as the catalytic rate; the turnover frequency was obtained by dividing the rate by the molar content in M units in the catalyst, except for single oligomers with no M units, in which case the rate was divided by the molar content in oligomer.

Annex 4.2 NMR titration and dilution experiments

General formalism

Mass balance equations

Consider two molecules, A and B, which can form a trivalent hydrogen-bonded duplex $A\cdot\overset{3}{\cdot}\cdot B$ in which two donor sites of A (hereafter called Dn1 and Dn3) and one donor site of B (hereafter called Dn2) are involved. The protons associated to these sites are numbered 1 and 3 for A, and 2 for B. Reciprocally, A and B have one and two hydrogen bond-accepting sites, respectively, hereafter called Ac2, Ac1 and Ac3, respectively. This is for instance the case for the pairs of monomers D and T, and G and C, displayed in Figure 1.

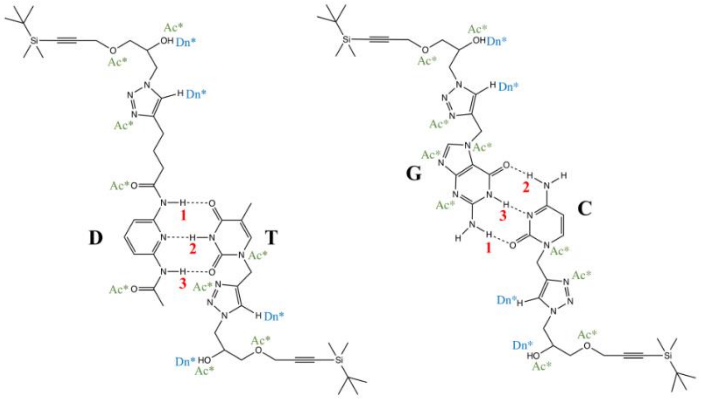


Figure 1 Chemical structure of the H-bonding monomers of this study, with the main trivalent duplexes shown.

Additionally, there are a_A and a_B generic accepting sites Ac* not involved in the main binding triad of A and B, respectively (here, we neglect the dual donating/accepting character of Dn1, Dn2 and Dn3). Likewise, there are d_A and d_B generic donor sites Dn* in A and B, respectively. For instance, for the D/T pair, $a_D = 5$, $a_T = 4$, and $d_D = d_T = 2$; for the G/C pair, $a_G = 6$, $a_C = 4$, and $d_G = d_B = 2$.

In addition to the desired trivalent duplex, A and B can form a series of other hydrogen-bonded complexes. Here, we will we restrict ourselves to measurements performed in hydrogen bond-accepting solvents ; therefore, the generic acceptor sites Ac* of A and B are completely negligible compared to the vast amount of accepting sites of the solvent, and may be ignored in the

equilibrium equations. The remaining possible hydrogen-bonded complexes involving the sites of the trivalent duplex are :

1. the trivalent duplex of interest, A$\cdot\overset{3}{\cdot}\cdot$B ;
2. possible trivalent self-duplexes A$\cdot\overset{3}{\cdot}\cdot$A or B$\cdot\overset{3}{\cdot}\cdot$B ;
3. possible divalent A$\cdot\overset{2}{\cdot}\cdot$B, A$\cdot\overset{2}{\cdot}\cdot$A and B$\cdot\overset{2}{\cdot}\cdot$B duplexes ;
4. a collection of simple complexes involving the accepting sites Ac1–3 and generic donating sites Dn*, such as A$\cdot\overset{1}{\cdot}\cdot$Dn*, B(Ac2)$\cdot\overset{1}{\cdot}\cdot$Dn*, B(Ac3)$\cdot\overset{1}{\cdot}\cdot$Dn* and B$\cdot\overset{2}{\cdot}\cdot$Dn$_2^*$.

Given the impossibility to obtain separate values for all equilibrium constants associated to all the possible complexes of the problem, we will combine equilibrium constants in the following way, defining :

1. $k_{AB} = K_{AB}/(VC^\circ)$, in which K_{AB} is the sum of all equilibrium constants for the formation of A$\cdots$B duplexes, be them trivalent or divalent ;
2. $k_{AA} = K_{AB}/(VC^\circ)$, in which K_{AA} is the sum of all equilibrium constants for the formation of A$\cdots$A duplexes, be them trivalent or divalent ;
3. $k_{BB} = K_{BB}/(VC^\circ)$, in which K_{BB} is the sum of all equilibrium constants for the formation of B$\cdots$B duplexes, be them trivalent or divalent ;
4. $k_A^* = K_A^*/(VC^\circ)$, in which K_A^* is the equilibrium constant for the formation of a simple A$\cdot\overset{1}{\cdot}\cdot$Dn* complex ;
5. $k_B^* = K_B^*/(VC^\circ)$, in which K_B^* is the equilibrium constant for the formation of a simple B$\cdot\overset{1}{\cdot}\cdot$Dn* complex ;
6. $k_{B_2}^* = K_{B_2}^*/(VC^\circ)^2$, in which $K_{B_2}^*$ is the equilibrium constant for the formation of a doubly bound B$\cdot\overset{2}{\cdot}\cdot$Dn$_2^*$ complex. For convenience, we define $r_2 = k_{B_2}^*/k_B^*$.

In these equations, V is the volume of solution, and $C^\circ = 1$ mol/L is the standard concentration.

Suppose now that we introduce $n_{0\mathrm{A}}$ and $n_{0\mathrm{B}}$ moles of A and B in the solvent. The introduced number of moles of generic acceptor and donor sites are thus $n_{0\mathrm{Ac}^*} = a_\mathrm{A} n_{0\mathrm{A}} + a_\mathrm{B} n_{0\mathrm{B}}$ and $n_0^* \triangleq n_{0\mathrm{Dn}^*} = d_\mathrm{A} n_{0\mathrm{A}} + d_\mathrm{B} n_{0\mathrm{B}}$, respectively. The mass balance equations are :

$$\frac{n_{0\mathrm{A}}}{n_\mathrm{A}} = 1 + k_{AB} n_\mathrm{B} + 2k_{AA} n_\mathrm{A} + k_A^* n^* \tag{1}$$

$$\frac{n_{0\mathrm{B}}}{n_\mathrm{B}} = 1 + k_{AB} n_\mathrm{A} + 2k_{BB} n_\mathrm{B} + k_B^* \big(1 + r_2 n^*\big) n^* \tag{2}$$

$$\frac{n_0^*}{n^*} = 1 + k_A^* n_\mathrm{A} + k_B^* \big(1 + 2r_2 n^*\big) n_\mathrm{B} \tag{3}$$

in which n_A, n_B and n^* are the number of moles of non-complexed A, B and

generic donor sites, respectively.

For a faster numerical solution of eqs.(1)–(3), we first solve them when the number of generating donating sites is vanishingly small ($n_0^* \approx n^* \to 0$), using :[1]

$$2k_{AA}n_{\mathsf{A}}^2 + (1 + k_{AB}n_{\mathsf{B}})n_{\mathsf{A}} - n_{0\mathsf{A}} = 0 \tag{4}$$

$$\text{with } n_{\mathsf{B}} = \frac{1 + k_{AB}n_{\mathsf{A}}}{4k_{BB}}\Big(\sqrt{1 + \frac{8k_{BB}n_{0\mathsf{B}}}{(1 + k_{AB}n_{\mathsf{A}})^2}} - 1\Big) \tag{5}$$

or, alternatively,

$$2k_{BB}n_{\mathsf{B}}^2 + (1 + k_{AB}n_{\mathsf{A}})n_{\mathsf{B}} - n_{0\mathsf{B}} = 0 \tag{6}$$

$$\text{with } n_{\mathsf{A}} = \frac{1 + k_{AB}n_{\mathsf{B}}}{4k_{AA}}\Big(\sqrt{1 + \frac{8k_{AA}n_{0\mathsf{A}}}{(1 + k_{AB}n_{\mathsf{B}})^2}} - 1\Big) \tag{7}$$

Once this approximative solution is found, the complete set of eqs.(1)–(2) is solved numerically, starting from the approximate solution obtained for $n_0^* \approx n^* \to 0$, and using eq.(3) to express n^* as a function of n_{A} and n_{B} :

$$n^* = \frac{1 + k_A^* n_{\mathsf{A}} + k_B^* n_{\mathsf{B}}}{4r_2 k_B^* n_{\mathsf{B}}}\Big(-1 + \sqrt{1 + \frac{8r_2 k_B^* n_{\mathsf{B}} n_0^*}{(1 + k_A^* n_{\mathsf{A}} + k_B^* n_{\mathsf{B}})^2}}\Big) \tag{8}$$

with the following limit when n_{B} tends to zero :

$$\lim_{n_{\mathsf{B}} \to 0} n^* = \frac{n_0^*}{1 + k_A^* n_{\mathsf{A}}} \tag{9}$$

NMR displacements of protons 1–3

When the rate of exchange between the complexes is much faster than the NMR characteristic time, the displacement δ of a proton involved in hydrogen bonding is given by the molar average of the displacements of this proton in the different coexisting species. The NMR displacements can then be expressed from the knowledge of the composition of the solution (remember that protons 1 and 3 belong to A, while proton 2 belongs to B) :

$$\delta_1 = \delta_{10} + k_{AB}\frac{n_{\mathsf{A}}n_{\mathsf{B}}}{n_{0\mathsf{A}}}\alpha_{1AB}\overline{\Delta_{1AB}} + k_{AA}\frac{n_{\mathsf{A}}^2}{n_{0\mathsf{A}}}\alpha_{1AA}\overline{\Delta_{1AA}} \tag{10}$$

$$\delta_2 = \delta_{20} + k_{AB}\frac{n_{\mathsf{A}}n_{\mathsf{B}}}{n_{0\mathsf{B}}}\alpha_{2AB}\overline{\Delta_{2AB}} + k_{BB}\frac{n_{\mathsf{B}}^2}{n_{0\mathsf{B}}}\alpha_{2BB}\overline{\Delta_{2BB}} \tag{11}$$

$$\delta_3 = \delta_{30} + k_{AB}\frac{n_{\mathsf{A}}n_{\mathsf{B}}}{n_{0\mathsf{A}}}\alpha_{3AB}\overline{\Delta_{3AB}} + k_{AA}\frac{n_{\mathsf{A}}^2}{n_{0\mathsf{A}}}\alpha_{3AA}\overline{\Delta_{3AA}} \tag{12}$$

1. When $n_{0\mathsf{A}} = 0$, $n_{\mathsf{A}} = 0$ and $n_{\mathsf{B}} = \frac{\sqrt{1+8k_{BB}n_{0\mathsf{B}}}-1}{4k_{BB}}$. Likewise, when $n_{0\mathsf{B}} = 0$, $n_{\mathsf{B}} = 0$ and $n_{\mathsf{A}} = \frac{\sqrt{1+8k_{AA}n_{0\mathsf{A}}}-1}{4k_{AA}}$.

In these equations, $\overline{\Delta_{iXY}} \triangleq \overline{\delta_{iXY}} - \delta_{i0}$ in which δ_{i0} is the NMR displacement of proton i in the solvent, and $\overline{\delta_{iXY}}$ is an average NMR displacement of proton i in the ensemble of XY duplexes, whose exact expression is complex and not especially useful. The α_{iXY} coefficients appearing in equations (10)–(12) refer to the ratio between the number of times proton i is involved in duplexes of type XY, and the number of such duplexes. These coefficients are introduced to have the $\overline{\Delta_{iXY}}$ shifts normalized per proton, but are not *per se* required. Examples will provided later on.

The previous equations are actually a simplification of a large ensemble of equations. Actually, the average chemical shifts $\overline{\Delta_{iXY}}$ involve a family of duplexes, whose composition might vary slightly when in competition with other species resulting in variations of $\overline{\Delta_{iXY}}$. However, without this simplification, the number of unknowns would be too large to extract reliable numbers from experimental data.

Complexation of monomer units D & T

NMR displacements were obtained for a range of concentrations in deuterated acetonitrile (CD_3CN) for temperatures ranging from room temperature to 333 K, in deuterated chloroform/deuterated dimethylsulfoxide (CD_3CN:DMSO-d_6 95:5 and 5:1 v :v) at room temperature, and in $CDCl_3$:DMSO-d_6 5:1 v:v at room temperature, 300 K and 333 K (Figure 2). Two types of experiments were performed : dilution experiments of T or D alone, to check the self-complexation of these molecules ; and titration experiments, in which the concentration of one molecule was kept constant while the other's was varied.

The possible bivalent and trivalent duplexes are shown in Figure 3. In the formalism of the previous section, A=D and B=T, and $a_{\mathsf{D}} = 5$, $a_{\mathsf{T}} = 4$, $d_{\mathsf{D}} = d_{\mathsf{T}} = 2$ as mentioned before. Proton 1 is involved twice in hydrogen bonds in the possible two D$\cdots$T duplexes ; hence, $\alpha_{1AB} = 2/2$. Likewise, it is involved four times in the four possible D$\cdots$D duplexes (once for c, once for d, and twice for f) ; hence, $\alpha_{1AB} = 4/4$. Similarly, it can easily be found that $\alpha_{2AB} = 2/2$, $\alpha_{2BB} = 8/4$, $\alpha_{3AB} = 2/2$, and $\alpha_{3AA} = 4/4$.

The number of unknowns being large, simplifications have to be introduced to ensure statistical significance. The symmetry of the H-binding part of monomer D results in very close NMR shifts for protons 1 and 3, with an identical variation with concentration or temperature. These displacements were therefore averaged and the fits performed on the average only. Additionally, K_D^* and K_T^* were invariably found to converge towards very small values, indicating that the generic donor sites of the molecules can be ignored in the equilibrium. Therefore, we set $K_D^* = K_T^* = 10^{-20}$ and $r_2 = 1$, with no dependence on temperature.

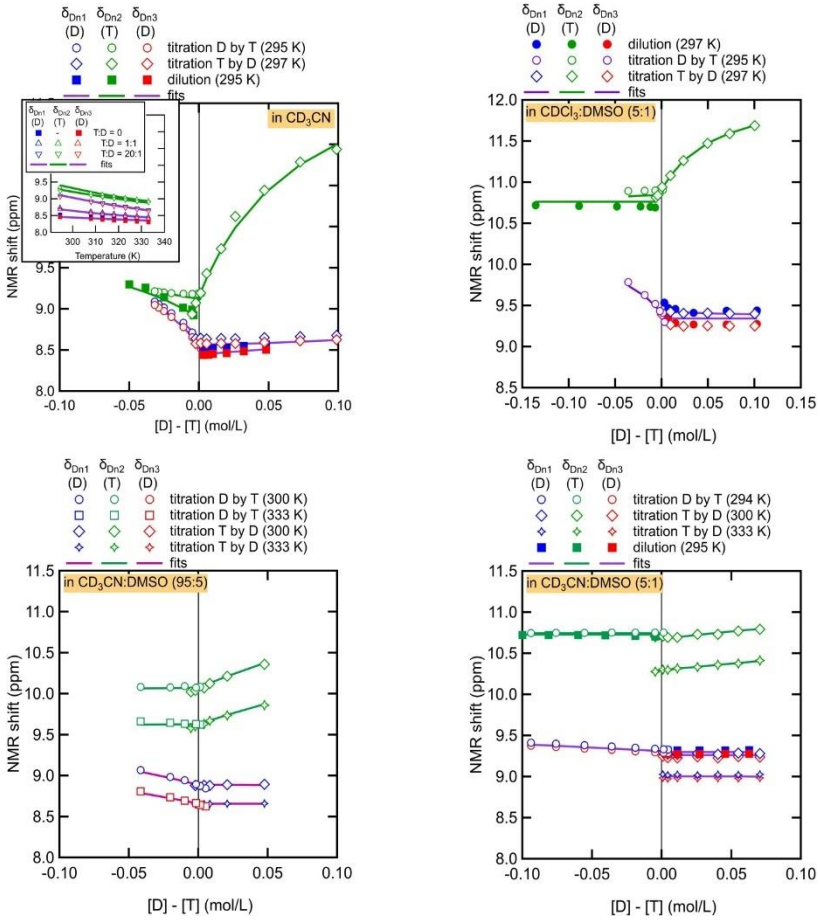


Figure 2 NMR shifts of protons 1 (blue symbols), 2 (green symbols) and 3 (red symbols) acquired at different temperatures for different concentrations of D and T in different solvents, including simple dilution experiments of D and T alone, and titration experiments. The shifts are plotted versus the difference of concentration of D and T in mol/L. The continuous lines are fits to the data using the model developed above, with fit parameters collected in Table 1. Because protons 1 and 3 have similar NMR shifts, the fits were performed on the average shift of these two protons (purple lines). The inset in the top left panel is the temperature variation of the NMR shifts in deuterated acetonitrile, together with the fits.

Parameter	in CD_3CN at 295 K	Error	in CD_3CN: DMSO-d_6 95:5 at 295 K	Error	in CD_3CN: DMSO-d_6 5:1 at 295 K	Error	in $CDCl_3$: DMSO-d_6 5:1 at 295 K	Error
K_{DT}	16.0	1.7	3.3	0.2	0.99	0.06	20.1	1.2
K_{DD}	0.4	0.1	10^{-20}	-	10^{-20}	-	0.09	0.06
K_{TT}	3.2	0.6	10^{-20}	-	10^{-20}	-	0.09	0.05
ΔH°_{DT} (kJ)	-21	5.7	-6.1	1.8	-5.7	2.3	-	-
ΔH°_{DD} (kJ)	24	50	0	-	0	-	-	-
ΔH°_{TT} (kJ)	-21	10	0	-	0	-	-	-
ΔS°_{DT} (J/K)	-48	19	-10.8	6	-19	8	-	-
$\frac{\delta_{10}+\delta_{30}}{2}$ (ppm)	8.44	0.01	8.898	0.006	9.297	0.002	9.346	0.004
δ_{20} (ppm)	8.88	0.02	10.097	0.006	10.729	0.002	10.771	0.004
$\frac{d(\delta_{10}+\delta_{30})/2}{dT}$ (ppm/K)	-0.0036	0.0022	-0.0068	0.0002	-0.0079	0.0001	0	-
$\frac{d\delta_{20}}{dT}$ (ppm/K)	-0.0048	0.002	-0.0132	0.0002	-0.0116	0.0002	0	-
Δ_{1DT} (ppm)	1.91	0.13	1.45	-	1.05	-	0.919	0.04
Δ_{2DT} (ppm)	3.64	0.17	2.42	-	1.79	-	1.375	0.03

Table 1 Fit para meters for D---T complexation in different solvents (Fig.2). The errors on the fit para meters were estimated from the diagonal terms of the covariance matrix, assuming a standard error on NM shifts of 0.01 ppm; ' -' in the error columns indicates that the corresponding parameter was not fitted. The standard entropies of complexation were not fitted and are computed from the equilibrium constants and the standard enthalpies of complexation.

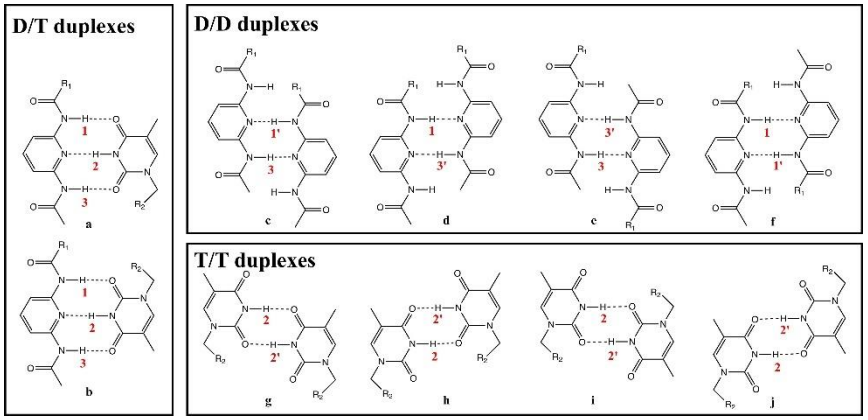


Figure 3 Possible hydrogen-bonded duplexes of monomers D and T.

Furthermore, an inspection of the pattern of hydrogen bonds in the trivalent D···T and divalent D···D and T···T duplexes (Figure) indicates that $\overline{\Delta_{1DD}} \simeq \overline{\Delta_{2DT}}$ and that $\overline{\Delta_{2TT}} \simeq \overline{\Delta_{1DT}}$. Therefore, these equalities were enforced during the fits.

Finally, a linear dependence of the δ_{i0}'s was assumed with temperature, with no temperature dependence of $\overline{\Delta_{iXY}}$'s ; the temperature dependence of the equilibrium constants was taken into account using :

$$K(T) = K(T_0) \exp\Big(- \frac{\Delta H^\circ}{R} (\frac{1}{T} - \frac{1}{T_0}) \Big) \tag{13}$$

in which $T_0 = 295$ K is an arbitrarily-selected reference temperature and ΔH° is the standard enthalpy of binding.

For experiments performed in the more hydrogen-binding CD_3CN:DMSO solvents, we also observed that K_{DD} and K_{TT} converged towards low values, presumably because divalent duplexes were essentially destroyed in this solvent ; they were therefore neglected as well. We also requested that $\delta_{i0} + \overline{\Delta_{iDT}}$ in these mixed solvents be equal at 295K to the value found in pure acetonitrile at this temperature. This corresponds to neglecting solvent effects on the NMR shifts of the D···T duplexes, which is reasonable since the protons in duplexes are less prone to interact with solvent molecules.

As for experiments performed in $CDCl_3$:DMSO, for which experiments were only performed close to room temperature, all thermal dependences were ignored.

The fits are displayed in Figure 2, and represent properly the experimental data, testifying for the plausibility of the model. The resulting fit parameters are collected in Table 1. The equilibrium constant K_{DT} for the formation of the trivalent duplex remains small in the selected solvents, with a decrease of a factor of *ca.* 5 and 16 observed when adding 5 and 16% of DMSO to acetonitrile,

respectively. The equilibrium constant in chloroform with 16% added DMSO is *ca.* 20 times higher than in acetonitrile added with the same amount of DMSO, which results from acetonitrile being a much stronger hydrogen bond acceptor than chloroform.

Complexation of monomer units G & C

NMR displacements were obtained for a range of concentrations in a deuterated chloroform/deuterated dimethylsulfoxide mixture ($CDCl_3$:DMSO-d_6 5:1 v:v) at room temperature, and in CD_3CN:DMSO-d_6 5:1 v:v at room temperature, 300 K and 333 K (monomer G is not soluble in pure acetonitrile, nor with only 5% of added DMSO). Again, two types of experiments were performed : dilution experiments of C or G alone, to check the self-complexation of these molecules ; and titration experiments, in which the concentration of one molecule is kept constant while the other's is varied (Figure 4).

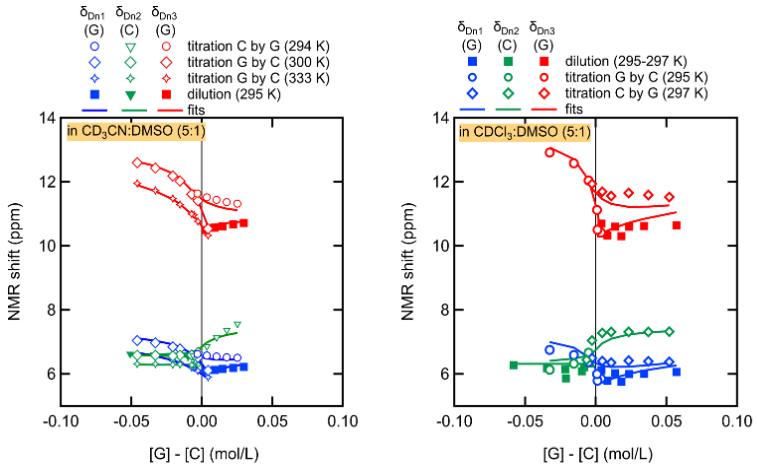


Figure 4 NMR shifts of protons 1 (blue symbols), 2 (green symbols) and 3 (red symbols) acquired at different temperatures for different concentrations of G and C in different solvents, including simple dilution experiments of G and C alone, and titration experiments. The shifts are plotted versus the deference of concentration of G and C in mol/L. The continuous lines are fits to the data using the model developed above, with fit parameters collected in Table 2. In some experiments, the NMR signal of the primary amine protons 2 of C could be differentiated (triangles), in which case the fit was performed on the average shift.

The possible bivalent and trivalent duplexes are shown in Figure 5. In the formalism of the previous section, A=G and B=C, and $a_{\mathsf{G}} = 6$, $a_{\mathsf{C}} = 4$, and $d_{\mathsf{G}} = d_{\mathsf{B}} = 2$ as mentioned before. Proton 1 is involved four times in hydrogen bonds in the possible four G$\cdots$C duplexes; hence, $\alpha_{1AB} = 4/4$. Likewise, it is involved six times in the five possible G$\cdots$G duplexes; hence, $\alpha_{1AA} = 6/5$.

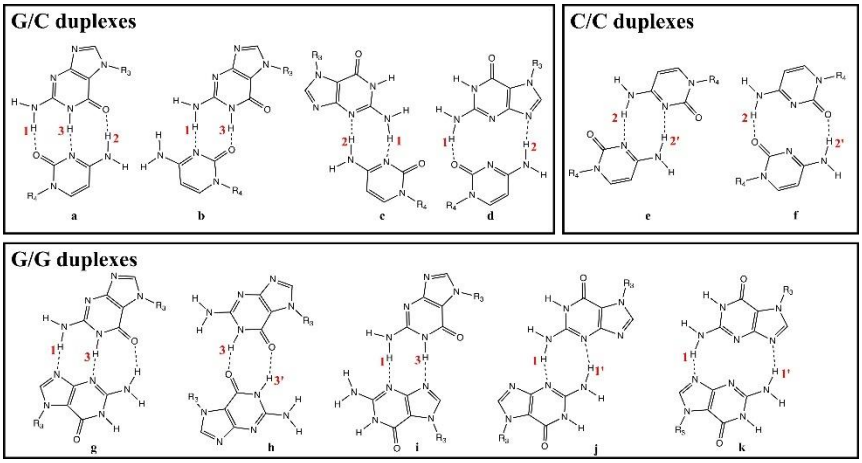


Figure 5 Possible hydrogen-bonded duplexes of monomers G and C.

Similarly, it can easily be found that $\alpha_{2AB} = 3/4$, $\alpha_{2BB} = 4/2$, $\alpha_{3AB} = 2/4$, and $\alpha_{3AA} = 4/5$.

As before, simplifications were performed to increase the significance of the fit parameters. Here again, K_G^* and K_C^* always converged towards very small values; hence, we set $K_G^* = K_C^* = 10^{-20}$ and $r_2 = 1$, with no dependence on temperature. Furthermore, we imposed $\overline{\Delta_{2GC}} = \overline{\Delta_{1GC}}$ and $\overline{\Delta_{2CC}} = \overline{\Delta_{1GG}}$, because protons 1 and 2 have similar NMR shifts and binding patterns, and $\overline{\Delta_{3GC}} = \overline{\Delta_{3GG}}$ again because the binding pattern of proton 3 in the C$\cdots$G and G$\cdots$G duplexes is similar. Finally, no temperature dependence was considered for experiments performed in $CDCl_3$:DMSO, which were only performed close to room temperature, whereas only the temperature dependence of K_{GC}, δ_{10}, δ_{20} and δ_{30} had to be taken into account in CD_3CN:DMSO.

The fits are displayed in Figure 4, and represent properly the experimental data, admittedly less so in $CDCl_3$:DMSO. The resulting fit parameters are collected in Table 2. The equilibrium constants K_{GC} for the formation of the trivalent duplex are similar for both solvents, and much larger than for the D$\cdots$T case, showing the greater stability of G$\cdots$C complexes compared to D$\cdots$T complexes. Additionally, the self-complexation of G is significant, with a K_{GG} constant only *ca.* 70 times smaller than K_{GC}. This comes as no surprise, as it is well-known that guanine has a strong tendency to form self-duplexes; this is most probably the cause for the insolubility of G in pure acetonitrile.

Table 2 – Fit para meters for G---C complexation in different solvents (Fig.4). The errors on the fit para meters were estimated from the diagonal terms of the covariance matrix, assuming a standard error on NM shifts of 0.01 ppm; '-' in the error columns indicates that the corresponding parameter was not fitted. The standard entropies of complexation were not fitted and are computed from the equilibrium constants and the standard enthalpies of complexation.

	in CD_3CN: DMSO-d_6 5:1		in $CDCl_3$: DMSO-d_6 5:1	
Parameter	at 295 K	Error	at 295 K	Error
K_{GC}	104	15	173	3
K_{GG}	1.4	0.2	5.1	0.1
K_{CC}	0.11	0.11	10^{-20}	-
ΔH°_{GC} (kJ)	-24	4	0	-
ΔS°_{GC} (J/K)	-43	14	-	-
δ_{10} (ppm)	6.05	0.05	5.661	0.004
δ_{20} (ppm)	6.53	0.03	6.315	0.003
δ_{30} (ppm)	10.62	0.03	10.293	0.006
$\frac{d\delta_{10}}{dT}$ (ppm/K)	-0.0061	0.0015	0	-
$\frac{d\delta_{20}}{dT}$ (ppm/K)	-0.0092	0.0011	0	-
$\frac{d\delta_{30}}{dT}$ (ppm/K)	-0.0077	0.002	0	-
$\overline{\Delta_{1GC}}$ (ppm)	1.36	0.09	1.572	0.009
$\overline{\Delta_{1GG}}$ (ppm)	5.89	1.91	3.48	0.06
$\overline{\Delta_{3GC}}$ (ppm)	5.10	0.25	6.54	0.03

Annex 4.3 All-atom molecular dynamics simulations and molecular network analysis (from UMons)

Simulation protocol

The oligomeric chains were built as a series of fragments, or residues, with the Avogadro software (Figure A.4.2).[206]

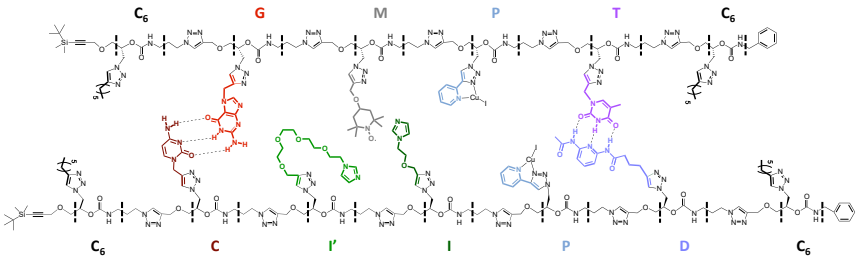


Figure A.4.2 Chemical structure of the complete O_a/O_{b1} catalytic system. The two strands, O_a and O_{b1}, are decomposed into a series of 28 residues, separated by black dots. The residues containing the functional side-chains are identified by letters.

These fragments were subsequently connected to one another to form the complete strands. The calculations were then performed with the AMBER16 simulation package, while free energy calculations were performed with the 2020 version of *ambertools*.[207] The atomic partial charges were assigned to the fragments using the *antechamber* module[208] of AMBER16 with the semi-empirical AM1-BCC method.[209] Structural parameters and partial charges of the 2,2,6,6-tetramethyl-1-piperidinyloxyl (TEMPO) radical were used as reported by Stendardo *et al.* after an ab-initio re-parameterization.[210] Structural parameters and partial charges of the pyridyltriazole copper ligands were obtained by calculations at the quantum-chemical level

using density functional theory (DFT) with the B3LYP/6-31G** model and extra basis sets LanL2DZ and MIDI! for the copper and iodine atoms, respectively. All other force-field parameters were given by the *General Amber Force-Field* (GAFF) 2.1.[211] The individual molecular fragments were then connected in the desired sequence with the *LEaP* module of AMBER16 to constitute the complete oligomeric chains. When building the complete O_a/O_b catalytic system, the second strand was translated by 40 Å in the *x*, *y* and *z* directions in order to avoid contacts between the two oligomers in the starting structure. A geometry optimization was then performed by molecular mechanics, with a total of 10,000 steps distributed in 1,000 steps of steepest descent and 9,000 steps of conjugated gradient, in order to get a stable starting point for the subsequent MD simulations. These were carried out with an implicit solvent model, the Generalized-Born (GB) model, to ensure a sufficient conformational sampling in a reasonable computational time.[212] The dielectric constant was set as the one of acetonitrile at 20 °C (ε = 37.5). For the catalytic system, constituted of the two strands O_a and O_b, two replicas of 10 µs were realized. Each oligomer was also simulated alone in two replicas of 5 µs. The timestep was fixed to 1 fs and the temperature was maintained at 300 K with a Langevin thermostat, with a collision frequency of 1 ps^{-1}. A bond restraint was applied in the simulations of the two strands to avoid that they translate in opposite directions and never meet each other, as we are not in periodic conditions: when the distance between the chains exceeds about 75 Å, a force constant of 10 $kcal.mol^{-1}.Å^{-2}$ is activated to prevent the chains from moving further away. An infinite cut-off was selected for the non-bonded interactions. A snapshot was saved each ns during the MD simulations and extracted for further analyses, giving a total of 20,000 conformations for the complete catalytic assembly and 10,000 for

each oligomeric chain alone. The GPU version of AMBER16 was used for all minimizations and MD simulations.[213] PyMOL 2.5.4 was used to visualize the snapshots and to extract images.[214]

Analyses of the trajectories

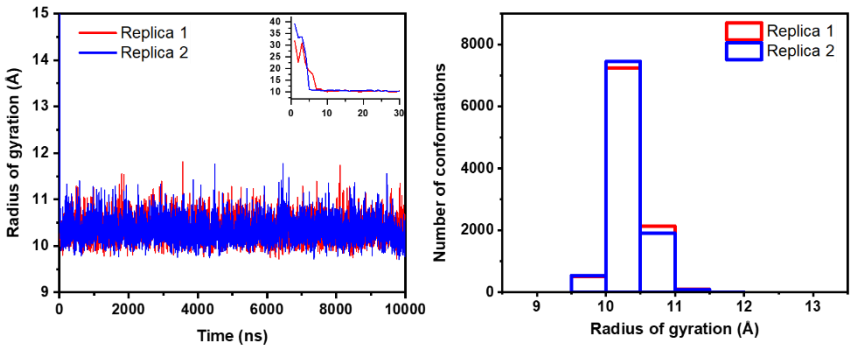
The analyses were performed using the *cpptraj* program of AMBER16. Mass-weighted radii of gyration (R_g) were computed with respect to heavy atoms (all atoms except hydrogens). End-to-end distances were calculated as the distance between the carbon directly linked to the silicon in the tert-butyl moiety at one end, and the carbon in para position of the phenyl ring at the other end. Hydrogen-bonds were detected with the default *cpptraj* parameters, i.e., a distance cutoff of 3.0 Å between the acceptor and the donor heavy atom and an angle cutoff of 135° between the donor, the hydrogen atom and the acceptor. Aromatic interactions (parallel stacking) were found with geometric criterion: two aromatic units are considered in interaction if the distance between their centres of mass is less than or equal to 5.0 Å and if the angle between the normal vectors of their planes is comprised between 45 and 135°. The heatmaps of H-bonds and stacking interactions were built with in-house scripts. The residue decomposition follows the sequence order and corresponds to the fragments used to build the chains, as is represented in Figure A.4.2.. Solvent-accessible surface-area (SASA) values were calculated with the LCPO model, as implemented within *cpptraj*.[215] The per-residue ΔSASA was simply computed as the difference between the SASA calculated for a residue in the complex of two chains and the SASA calculated for the same residue in the oligomer alone (the conformations of the individual oligomers were, in this case, obtained by removing the other chain in the trajectory of the complex). The ΔSASA can only be inferior or equal to 0: a residue that would be

located far from the interface of the two strands, without contact with the second chain, would have a ΔSASA of 0, meaning that it is as accessible with or without the second chain. In contrast, a residue that would be buried at the interface would have a highly negative ΔSASA, showing that it is much less accessible in the presence of the second strand.

Binding enthalpy calculations were performed with the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method, using the parallelized version of the Python program MMPBSA.py 14.0, implemented in Amber.[216] The binding enthalpy is given as the difference between the energy calculated for the complex (here, the complete catalytic system) and the sum of energies for the receptor (O_a) and ligand (O_b) alone. The "multiple-trajectory" approach was followed, which means that the conformations for the complex, receptor and ligand were obtained by independent simulations. MMPBSA.py post-processes the trajectories and calculates the energy of each frame with an implicit representation of the solvent. The energy is divided in two parts: an internal contribution and a solvation contribution. The internal contribution was given by the force-field and can be seen as the energy of the system in vacuum (bonds, angles, dihedrals, van der Waals and electrostatics). The solvation contribution is further divided into a polar and a non-polar part. The polar part represents the electrostatic interactions between the solute and the solvent and was obtained by solving the PB equation with a finite difference method. The non-polar part was calculated as the sum of a favourable "dispersion term" and an unfavourable "cavity term", representing the stabilizing solute-solvent dispersion interactions and the cost of creating a cavity in the solvent, respectively. These two terms are proportional to the SASA. Several calculations were carried out by varying the internal dielectric

constant, i.e., the dielectric constant of the solute, ranging from 1 to 4.

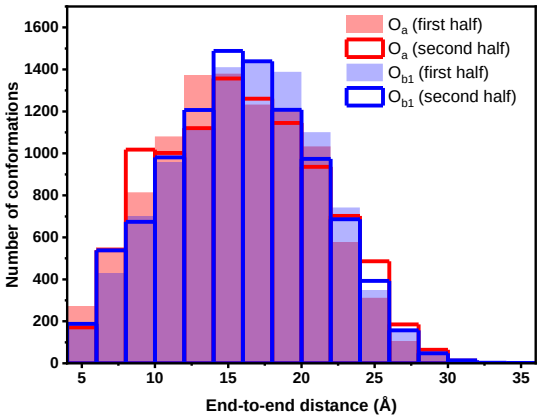
The default value is set to 1, but in some cases, a better agreement with experimental results was obtained with higher values of the internal dielectric constant.[217,218] In our case, a better agreement was reached with an internal dielectric value of 4. The first 100 ns of each replica were skipped for the calculations to let the systems reach equilibrium, giving a total of 19,800 conformations for the O_a/O_{b1} complex and 9,800 conformations for the oligomeric strands alone. The external dielectric constant was fixed at 37.5, the one of acetonitrile at room temperature, as was done during the MD simulations. MMPBSA.py offers the possibility to decompose the energy by residue and by pairs of residues (using the same residue division as the one shown in Figure A.4.3). The residue pairwise decomposition scheme was used to highlight pairs of residues playing an important part in the binding of the two oligomeric chains.



	Avg. (Å)	Std. Dev. (Å)	Initial (before assembly) (Å)	Final (Å)
Replica 1	10.3	0.4	32.0	10.3
Replica 2	10.3	0.5	39.2	10.1

Figure A.4.3 Top: radii of gyration of the di(oligomeric) complex for the two replicas as a function of the simulation time with a zoom on the first 30 ns on

the inset, showing the decrease of the R_g and its stabilization around 10 Å, related to the formation of the complex, in about 10 ns for each replica (left) and distributions of the radii for the 10,000 conformations sampled in each replica, with a bin size of 0.5 Å (right). Bottom: average, standard deviation, initial and final values of the radii of gyration for the two replicas.



	Avg (Å)	Std dev (Å)	Max (Å)	Min (Å)
O_a (first half)	15.5	5.2	34.9	4.3
O_a (second half)	15.8	5.4	32.4	4.4
O_{b1} (first half)	16.2	5.1	32.8	4.4
O_{b1} (second half)	15.9	5.2	34.9	4.2

Figure A.4.4 Top: distributions of the end-to-end distances for the 20,000 conformations obtained during the simulation of the di(oligomeric) cycle, for each oligomeric chain, highlighting the wide variety of conformations accessible to the chains inside the complex. The distributions are similar for the first and second halves of the simulation, showing that the strands remain strongly flexible during the 10 μs. The bin size is 2 Å. Bottom: average, standard deviation, maximum and minimum values of the end-to-end distance for each chain.

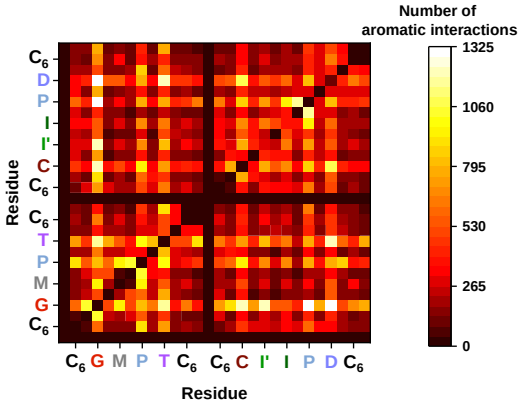


Figure A.4.5 Heatmap showing the decomposition of the aromatic interactions (parallel stacking) by pairs of residues. Each square, localized at the crossing of a pair of residues, indicate, by its colour, the number of interactions detected between two residues. The map is symmetric with respect to the diagonal. The map is rather homogeneous, showing that most aromatic units can be close to each other and stabilize the system, even though the most persistent interactions usually involve at least one nucleobase analogue.

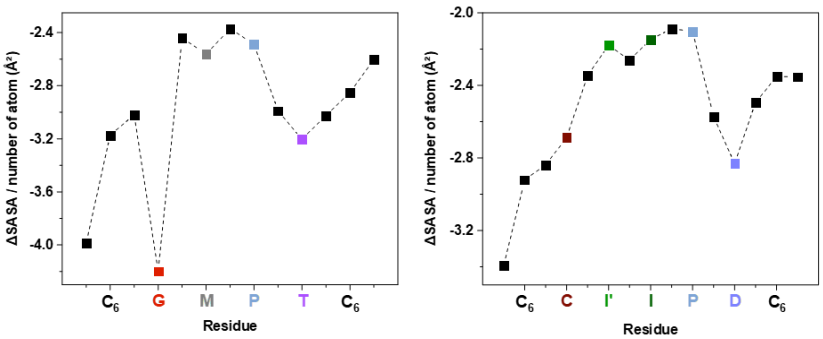


Figure A.4.6 Per-residue ΔSASA values normalized by the number of atoms in each residue and averaged over the whole simulation, for the oligomeric strands Oa (left) and Ob1 (right). The more negative the ΔSASA, the more the residue is hindered by the other strand in the di(oligomeric) complex, i.e., the more it is at the interface of the two strands. This graph shows that, for each chain, the catalytic units (M, P, I', I) tend to be among the less buried

residues. This makes them more accessible to the surrounding environment, which could be catalytic substrates, for example.

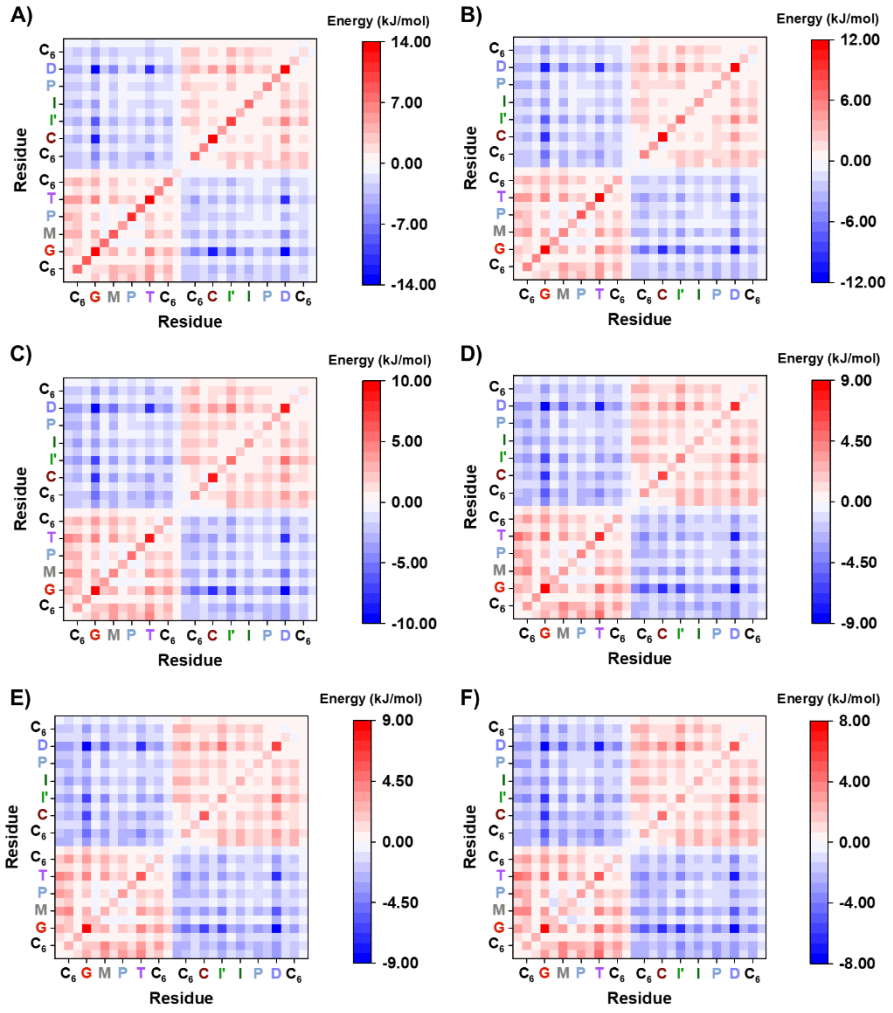


Figure A.4.7 Heatmaps showing the decomposition of the binding energy by pairs of residues. The most stabilizing residue-residue interactions are localized for the pairs G---D, T---D and G---C. This highlights again the role of the nucleobase analogues in the stabilization of the complex. A. Internal dielectric constant = 1. B. Internal dielectric constant = 1.5. C. Internal dielectric constant = 2. D. Internal dielectric constant = 2.5. E. Internal dielectric constant = 3. F. Internal dielectric constant = 3.5.

Molecular network analysis

The 3D conformations were converted into 2D networks. In this representation, all heavy atoms constitute *nodes*, and two nodes are connected by an *edge* if their distance is inferior or equal to 5 Å. This cutoff value allows to take into account hydrogen bonding interactions as well as aromatic stacking. In practice, one network file was created for each conformation of the catalytic system O_a/O_b from MD simulation, as soon as the duplex was formed (interchain distance < 14 Å) using in-house scripts, resulting in a total of 19,988 networks for O_a/O_{b1} and 19,953 for O_a/O_{b2}. These files, representing one conformation each, have been used to build one global network for each system, following this procedure: two nodes are considered connected by an edge only if they have been in contact during at least 10 % of the MD time, i.e., if their contact was detected in at least 10 % of the 19,998 conformations for O_a/O_{b1} and of the 19,953 conformations for O_a/O_{b2}. The resulting network is thus focusing on persistent contacts and neglecting more transient interactions. All edges are undirected and unweighted. To analyze and visualize the network, the *Cytoscape* 3.9.1 software[219] was used with its included analyzer *NetworkAnalyzer* 4.4.8.[220] Two descriptors were chosen to characterize the nodes inside the network. The betweenness centrality C_b for one node relies to the number of shortest paths connecting two other nodes passing through this node, i.e., the importance of the node to put the other ones into communication. The closeness centrality C_c for one node reflects the reciprocal of the average shortest paths length connecting this node to all the other nodes in the network: the higher the closeness, the more “central” is the node in the network, the more easily it communicates with the other nodes.

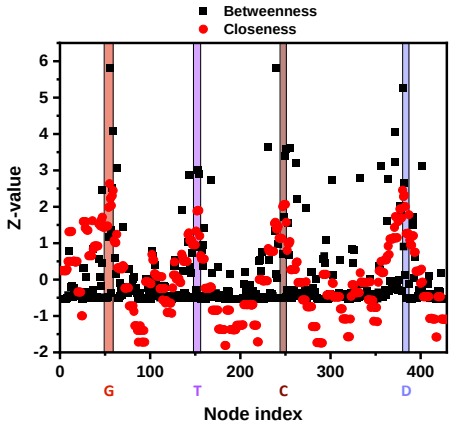


Figure A.4.8 Betweenness and closeness centralities for each node in the O_a/O_{b1} network, presented with their z-values. The nodes belonging to the four nucleobase analogs are highlighted. This graph shows that maxima of C_c (closeness centrality) and high values of C_b (betweenness centrality) tend to be located at or near the nodes of the bases, which demonstrates their role in the assembly of the two chains.

The network file was then submitted to the *Infomap* algorithm, in order to detect communities or modules, which are defined as highly connected groups of nodes.[221] The modular representation obtained can thus be thought as a coarse-grained view of the previous network. Various methods exist to decipher the modular topology of a network: the one that we used is called the *map equation*.[222,223] It is a flow-based method, which means that it focuses on "how is flowing, propagating the information from one node to another?". The propagation of information is materialized by a random walker that can move on the links between nodes while an information cost, in bits, can be associated with the movements of the walker. The main idea behind the *map equation* is that finding modules in the network can be seen as an encoding problem: to compress, reduce at best the information cost, it is necessary to efficiently partition, modularize

the network. The *map equation*, based on Shannon's source coding theorem, gives the theoretical lower limit of the information cost associated with one step of the random walker, on average, on the network.[224] *Infomap* is the algorithm used to minimize the map equation. The principle is as follows: each node begins in its own module. Then, the nodes were moved into their neighbouring module that reduces the most the *map equation*. This operation is repeated, the newly formed modules are merged with their neighbours, until no more minimization can be attained. The main goal of *Infomap* is thus to find the best partition of the network, i.e., the optimal organization of nodes inside the optimal number of modules, to reduce the most efficiently the information cost associated with the movements of a random walker. A two-level partition of the network was chosen, such as there is only one layer of modules containing the nodes (no possibility to have "module of modules").

Visualization and analysis of the network was done with the 2.6.0 version of the web server utility of *Infomap*.

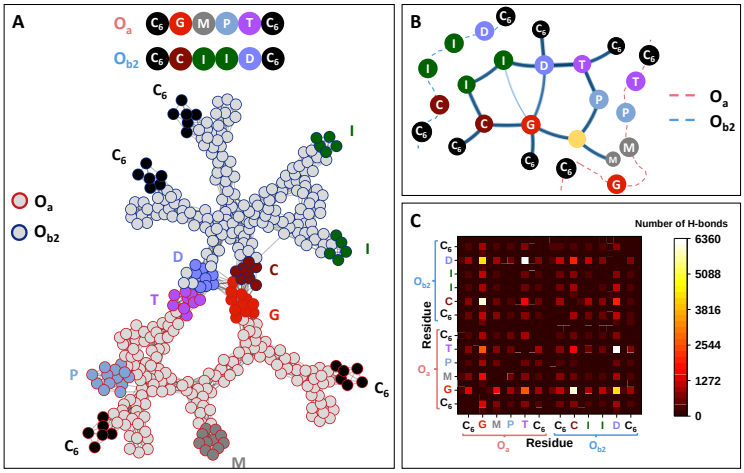


Figure A.4.9 Molecular simulations of the assembly of O_a and O_{b2} oligomers into catalytically-incomplete supramolecular cycles. A. Network representation of the system, highlighting the persistent contacts observed during the MD simulation. Heavy atoms constitute the nodes of the network. The nodes belonging to the strand O_a are circled in red while the nodes belonging to the strand O_{b2} are circled in blue. The nodes belonging to functional groups are identified by the corresponding code and colour. B. Modular representation of the network. Modules are named and coloured according to the functional unit included (but may contain other nodes). The module in yellow contains backbone nodes. For clarity, the primary structures of the two strands are sketched in dashed lines next to the modular network. C. Heatmap showing the decomposition of the hydrogen-bonds by pair of residues. Each square, localized at the crossing of a pair of residues, indicates by the scale bar the number of interactions detected between the two residues along the simulation. The residues from functional units are annotated while the backbone and chain-ends residues are hidden, for the sake of clarity. The map is symmetric with respect to the diagonal. White squares indicating the highest number of interactions are localized for the pairs G---C and T---D. Interactions are also found for the pair G---D. The map is very similar to the one of the O_a/O_{b1} system (Figure 4.4.E), showing that the composition of the catalytic centre does not strongly modify the interactions leading to the assembly of the chains.

Annex 4.4 NMR study of copper/base interactions

Although the pyridyltriazole group is a strong ligand for Cu, the base analogues may compete with it. Proton NMR experiments were thus performed with monomer P combined with either monomers G, T, D, or C (in 1:1 molar ratio) in the presence of increasing amounts of CuI. Limited competing complexation of $Cu^{(I)}$ happens with monomers G and to a lesser extent C, and negligible complexation of $Cu^{(I)}$ happens with D or T (Figure A.4.10). This limited competition probably contributes to a decreased activity of catalytic sites, which is taken into consideration in the factors p_c, p_L and p_f in the expressions of the turnover frequency of each species (Annex 4.6).

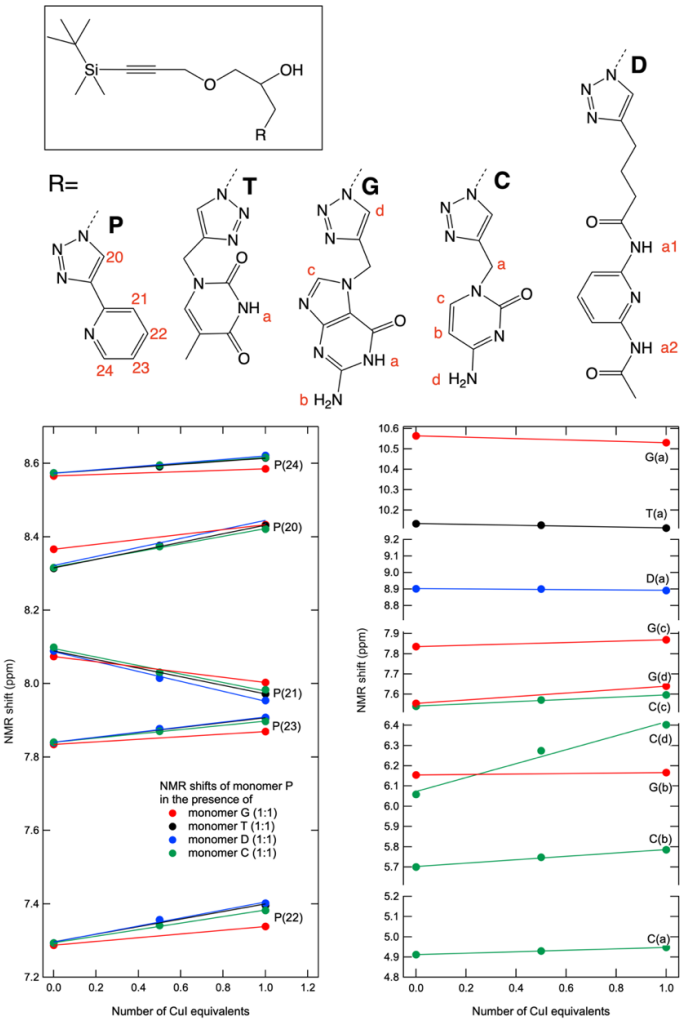
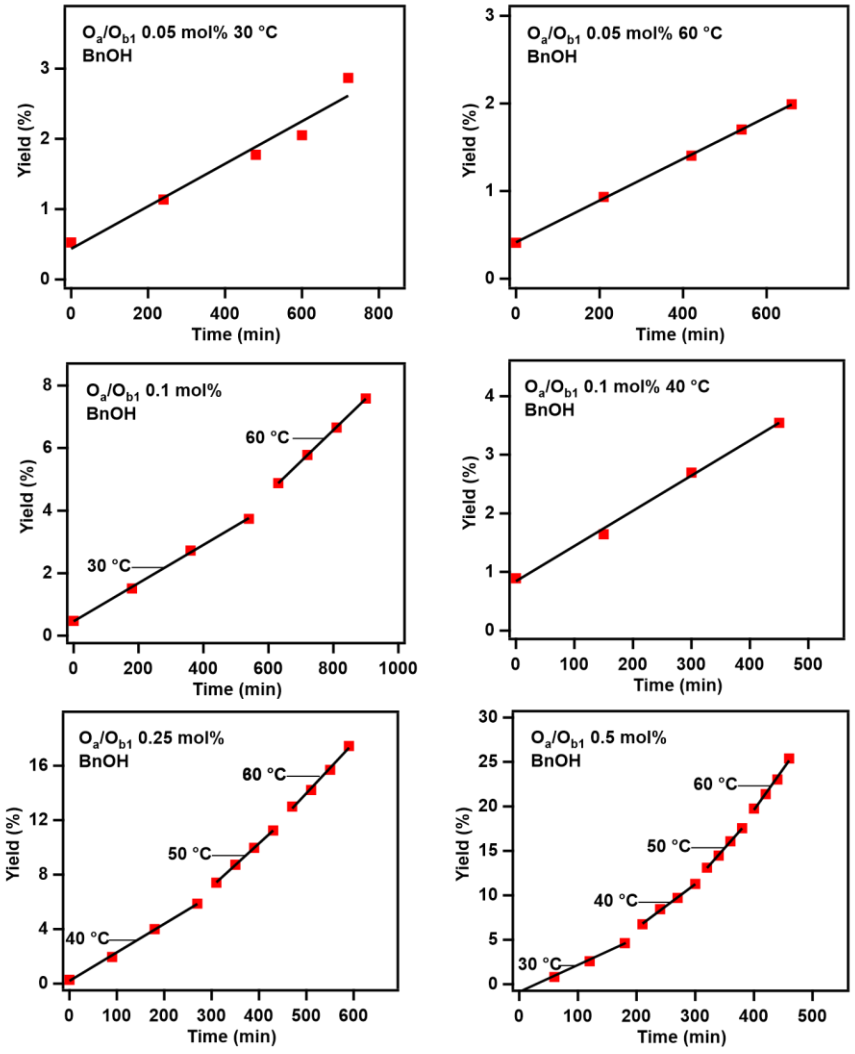
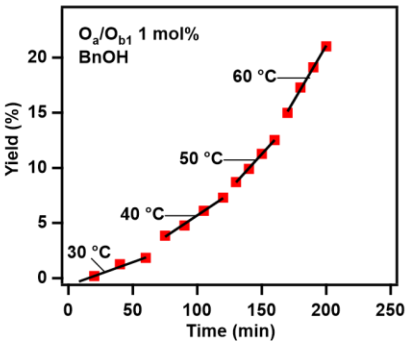
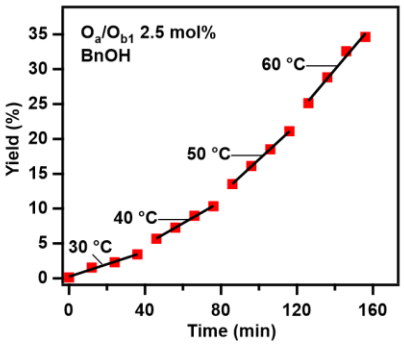


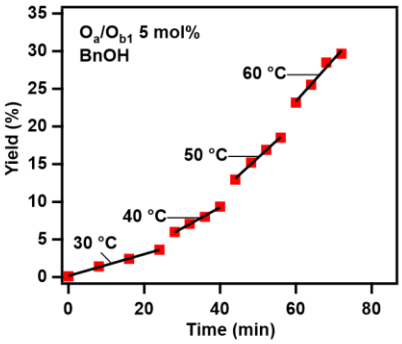
Figure A.4.10 Variation of the proton NMR displacements in P:X 1:1 monomer mixtures, with X=G, D, T or C, in the presence of increasing amounts of CuI. The left graph provides the NMR displacements of the protons of the pyridyltriazole ligand; the right graph is for the protons of the base analogues. The data indicate a small competition for copper by base G, as testified by the slightly reduced variation of the chemical displacements of the pyridyltriazole protons in the presence of Cu and G. The proton displacements of C in the presence of Cu and P are also modified, indicating also partial interaction with copper; however, this does not modify the effect of Cu on the displacements of the pyridyltriazole.

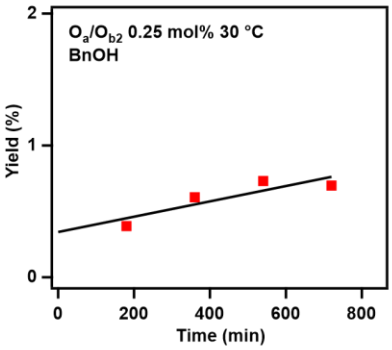
Annex 4.5 Graphs of yields versus time for the aerobic oxidation of alcohols

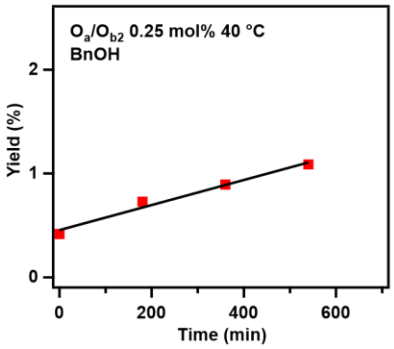


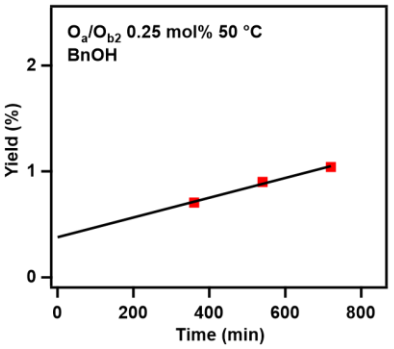

O_a/O_{b1} 1 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)

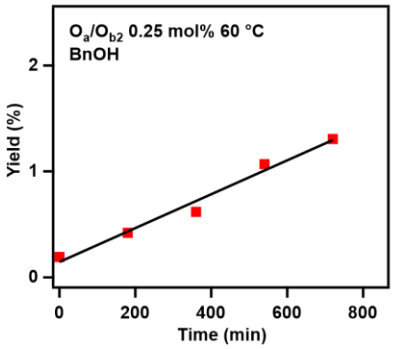

O_a/O_{b1} 2.5 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)

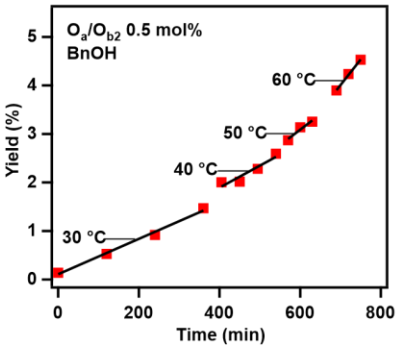

O_a/O_{b1} 5 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)

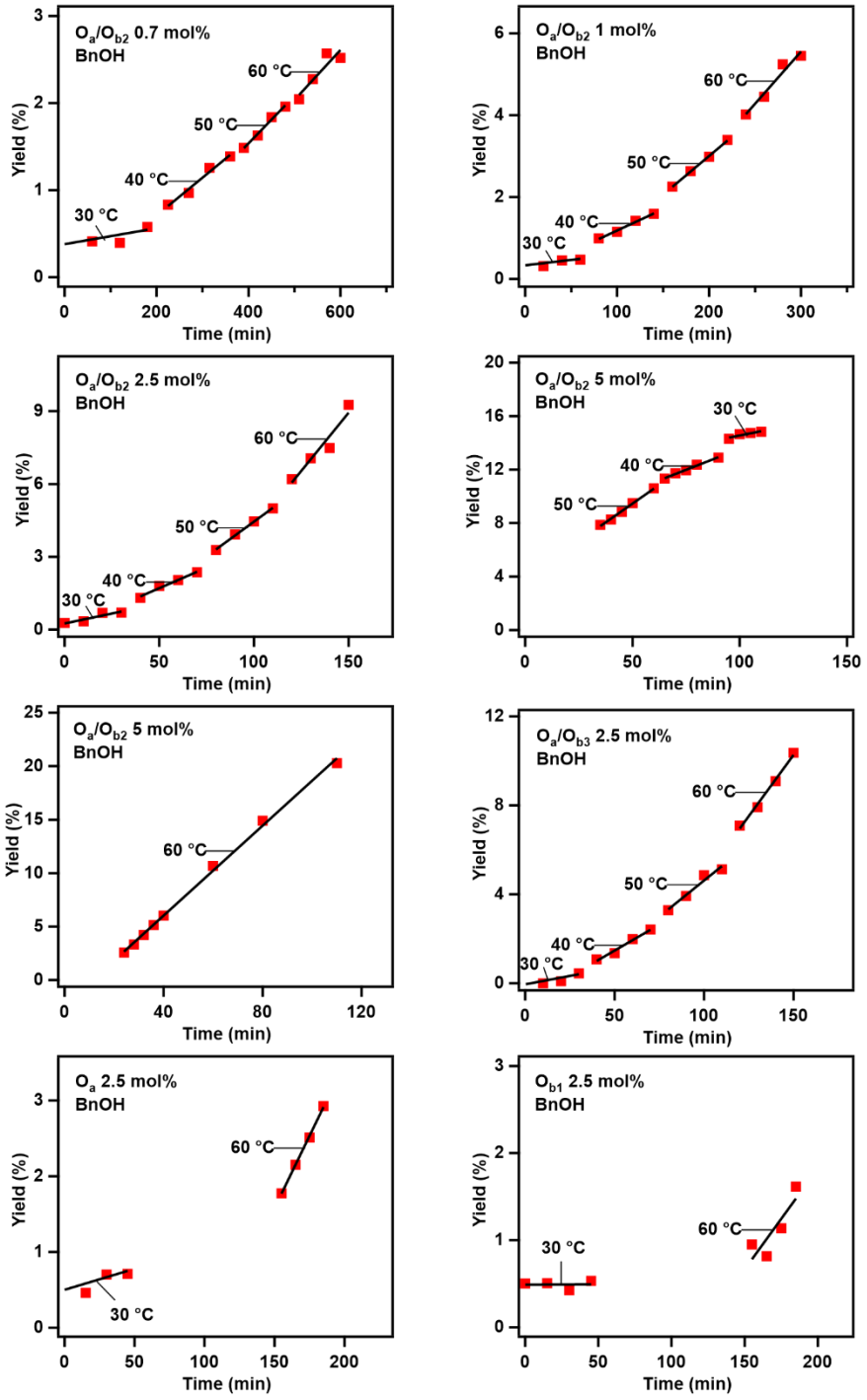

O_a/O_{b2} 0.25 mol% 30 °C
BnOH
Yield (%)
Time (min)

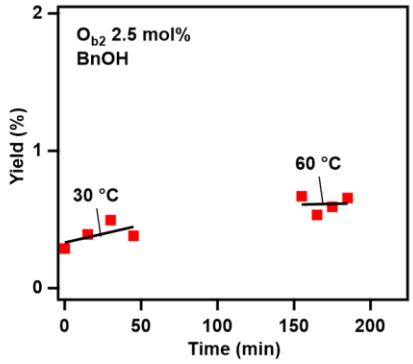

O_a/O_{b2} 0.25 mol% 40 °C
BnOH
Yield (%)
Time (min)

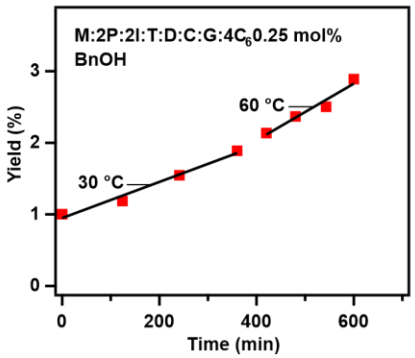

O_a/O_{b2} 0.25 mol% 50 °C
BnOH
Yield (%)
Time (min)

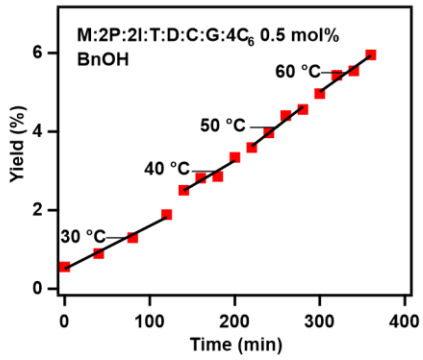

O_a/O_{b2} 0.25 mol% 60 °C
BnOH
Yield (%)
Time (min)

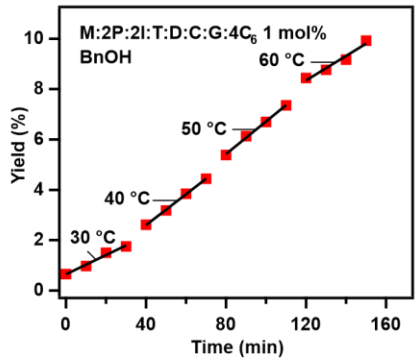

O_a/O_{b2} 0.5 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)

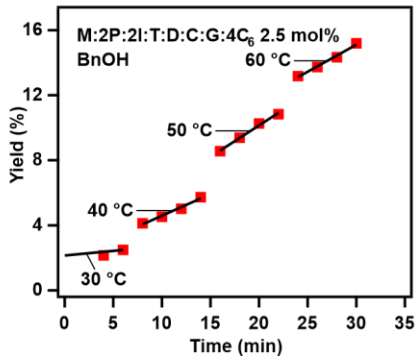

O_a/O_{b2} 0.7 mol% BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)
O_a/O_{b2} 1 mol% BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)
O_a/O_{b2} 2.5 mol% BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)
O_a/O_{b2} 5 mol% BnOH
30 °C
40 °C
50 °C
Yield (%)
Time (min)
O_a/O_{b2} 5 mol% BnOH
60 °C
Yield (%)
Time (min)
O_a/O_{b3} 2.5 mol% BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)
O_a 2.5 mol% BnOH
30 °C
60 °C
Yield (%)
Time (min)
O_{b1} 2.5 mol% BnOH
30 °C
60 °C
Yield (%)
Time (min)

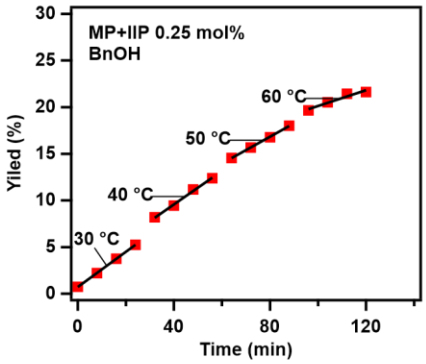

O_{b2} 2.5 mol%
BnOH
30 °C
60 °C
Yield (%)
Time (min)

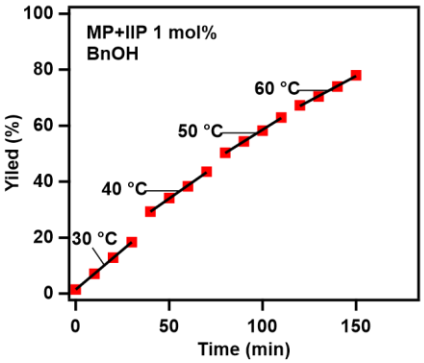

M:2P:2I:T:D:C:G:4$C_6$0.25 mol%
BnOH
30 °C
60 °C
Yield (%)
Time (min)

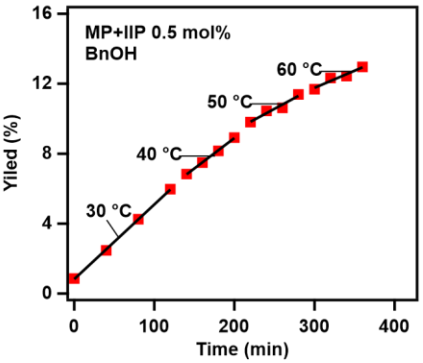

M:2P:2I:T:D:C:G:4C_6 0.5 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)


M:2P:2I:T:D:C:G:4C_6 1 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)


M:2P:2I:T:D:C:G:4C_6 2.5 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yield (%)
Time (min)


MP+IIP 0.25 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yiled (%)
Time (min)


MP+IIP 1 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yiled (%)
Time (min)


MP+IIP 0.5 mol%
BnOH
30 °C
40 °C
50 °C
60 °C
Yiled (%)
Time (min)

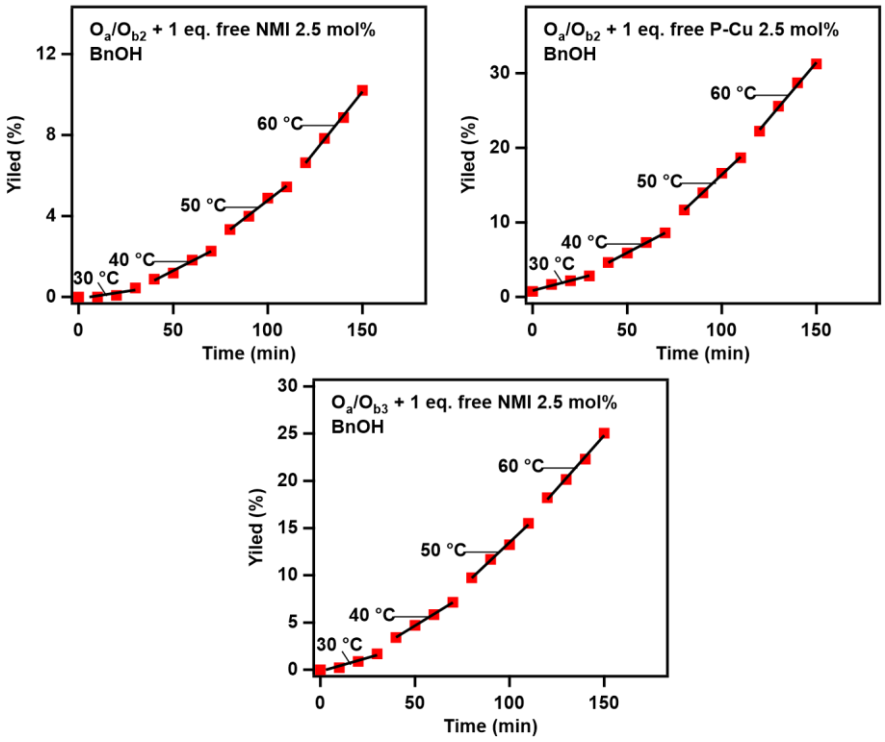


Figure A.4.11 Catalytic activity of the different systems with CuI in the aerobic oxidation of alcohols.

Annex 4.6 Mathematical model

Composition of the libraries

The different species co-existing in a solution of O_a and O_b oligomers, each introduced at the initial molar concentration *n*, are listed in Figure A.4.12.

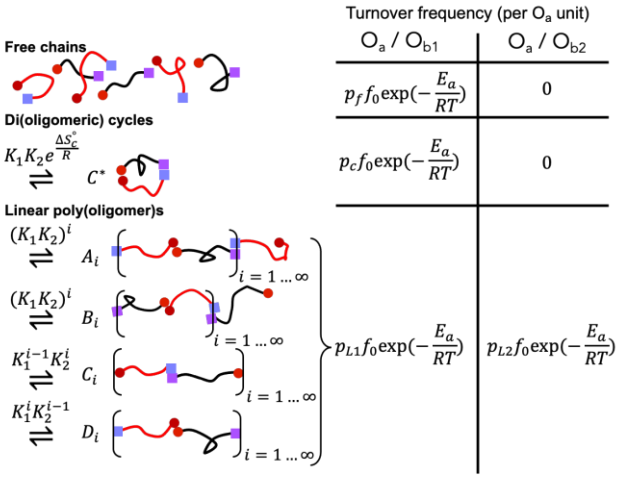


Figure A.4.12 Libraries of co-existing species in solutions of the complementary oligomers, and parameters used to model the catalytic activity of the system.

The binding constants for two complementary chain end segments are $K_m = \exp(-(\Delta H_m^\circ - T\Delta S_m^\circ)/RT)$, $m = 1,2$ with ΔH_m° and ΔS_m° the standard enthalpy and entropy of binding (*m*=1 for G- and C-containing end segments, and *m*=2 for D- and T-containing end segments). Therefore, the equilibrium constant for cycle formation is $K^* \triangleq K_1K_2 \exp(\Delta S_c^\circ)$ in which ΔS_c° is the standard entropy cost for cycle formation; likewise, the equilibrium constants for the different linear

poly(oligomers) are $(K_1K_2)^i$ for A_i and B_i, $K_1^{i-1}K_2^i$ for C_i and $K_1^iK_2^{i-1}$ for D_i (Figure A.4.12). Defining the molar concentration at equilibrium as *a* for free O_a, *b* for free O_b, n^* for di(oligomeric cycles), a_i, b_i, c_i and d_i for A_i, B_i, C_i and D_i, respectively, and neglecting supramolecular rings combining more than two strands as they are entropically much less likely to form and are not expected to result in catalytic activities significantly different from linear chains, the equilibrium equations are:

$$K^* = \frac{n^*}{ab}$$

$$(K_1K_2)^i = \frac{a_i}{a^i b^{i+1}}$$

$$(K_1K_2)^i = \frac{b_i}{a^{i+1} b^i}$$

$$K_1^{i-1}K_2^i = \frac{c_i}{a^i b^i}$$

$$K_1^iK_2^{i-1} = \frac{d_i}{a^i b^i} \text{ , } i = 1 \cdots \infty \text{ .}$$

The mass balance equation for O_a is:

$$n = n^* + a + \sum_{i=1}^{\infty}(i+1)b_i + \sum_{i=1}^{\infty} i\,(a_i + c_i + d_i)$$
$$= K^*ab + \frac{a(1 + K_1K_2b^2 + (K_1 + K_2)b)}{(1 - K_1K_2ab)^2}$$

for which we have used $\sum_{i=0}^{\infty} i\,x^{i-1} = (1-x)^{-2}$ if $|x| < 1$.

A similar equation can be obtained from the mass balance on O_b; however, this second equation is redundant since the two oligomers are introduced at the same concentration, and *a*=*b* at equilibrium. Therefore, the equilibrium concentration in free O_a oligomer, *a*, can be obtained by solving:

$$n = K^* a^2 + \frac{a(1 + (K_1 + K_2)a + K_1 K_2 a^2)}{(1 - K_1 K_2 a^2)^2}$$

which is equivalent to numerically finding the physically-compatible root of:

$$K_1^2 K_2^2 K^* a^6 - (K_1^2 K_2^2 n + 2K_1 K_2 K^*)a^4 + K_1 K_2 a^3 + (K^* + K_1 + K_2 + 2K_1 K_2 n)a^2 + a - n = 0$$

.

The fraction of O_a oligomers in free chains, di(oligomeric) cycles and linear chains is then a/n, $K^* a^2/n$ and $1 - (a + K^* a^2)/n$, respectively.

The average degree of poly(oligomerization) by number of the set of linear chains, excluding free oligomers and cycles, is:

$$< N >= \frac{\sum_{i=1}^{\infty} 2i(c_i + d_i) + (2i + 1)(a_i + b_i)}{\sum_{i=1}^{\infty} c_i + d_i + a_i + b_i} = \frac{2}{1 - K_1 K_2 ab} + \frac{K_1 K_2 (a + b)}{K_1 + K_2 + K_1 K_2 (a + b)}$$

in which *b*=*a* at equilibrium.

Catalytic activity of the constitutional libraries

The different co-existing species each contribute to the measured catalytic activity, here defined as the turnover frequency (*f*) measured

by dividing the initial rate of oxidation (moles of alcohol converted per unit time) by the number of moles of introduced catalyst (defined here as the number of introduced M units).

Each member of the constitutional libraries has a specific probability that its required catalytic groups meet in space, which depends on conformational sampling and on the probability that two atoms of copper be properly complexed in the site. Additionally, access of the substrate to the catalytic site, and diffusion of the products from it, may differ between species. Therefore, each species *j* existing into the solution is assigned a specific turnover frequency, *per* O_a *it contains*, $f_j = p_j f_0 \exp(-E_a/RT)$, in which f_0 and E_a are the intrinsic frequency and activation energy associated to a complete catalytic site, and p_j takes into account the probability of formation of a complete catalytic site for species *j*, as well as other factors such as the ease of diffusion of substrate and products to and from the site, etc.

For di(oligomeric) cycles, the contribution to the total turnover frequency is thus $f_c = p_c f_0 \exp(-E_a/RT)\, K^* a^2/n$, with $p_c = 0$ if the cycle does not contain the five groups required for the catalysis.

Linear chains collectively comprise all needed groups for catalysis. Folding of the chains may bring these groups in close proximity; additionally, the chains dynamically redistribute their component oligomers intra- and inter-molecularly, thereby adding supplementary degrees of freedom for the creation of a catalytically-active site. Therefore, they are assigned a collective turnover frequency $f_L = p_L f_0 \exp(-E_a/RT)\,(1 - (a + K^* a^2)/n)$, with p_L different for each

considered system.

Free oligomer chains only contribute to catalysis when they encounter by chance. For incomplete systems, this requires the random meeting of three chains, which is improbable and neglected. For the complete system, chance encounter of one O_a and one O_b is needed for catalysis. Therefore, in this case, $f_f = p_f f_0 \exp(-E_a/RT)\, a^2/n$. It should be noted that the units of p_f are different from the ones of p_c and p_L.

Combining these expressions leads to the expression of the turnover frequency of the catalytic solutions, $f = f_c + f_L + f_f$.

For the complete O_a/O_{b1} system:

$$f = p_c f_0 \exp(-E_a/RT)\, K^* a^2/n + p_{L1} f_0 \exp(-E_a/RT)\, (1 - (a + K^* a^2)/n) + p_f f_0 \exp(-E_a/RT)\, a^2/n$$

For the incomplete O_a/O_{b2} system:

$$f = p_{L2} f_0 \exp(-E_a/RT)\, (1 - (a + K^* a^2)/n)$$

with *n* the number of moles of introduced O_a, and *a* the number of free O_a chains at equilibrium, computed using the equations of the previous section.

Fits of the model to the experimental catalytic data

The theoretical expressions for the turnover frequencies were fitted simultaneously to all the experimental data acquired on both the O_a/O_{b1} and the O_a/O_{b2} systems, for the whole set of concentrations

and temperatures. The fit parameters and their optimal values obtained by the Marquardt-Levenberg fitting procedure are given in Table A.4.1.

Table A.4.1 Parameters obtained from the fit of the catalytic data.[a]

	O_a/O_{b1}	O_a/O_{b2}
ΔH_1° (kJ/mol, G---C pairing)	-45.0 ± 0.2	Same as for O_a/O_{b1}
ΔS_1° (J/mol.K, G---C pairing)	-82 ± 10	Same as for O_a/O_{b1}
ΔH_2° (kJ/mol, D---T pairing)	-41.2 ± 0.2	Same as for O_a/O_{b1}
ΔS_2° (J/mol.K, D---T pairing)	-72 ± 11	Same as for O_a/O_{b1}
ΔS_c° (J/mol.K)	-49 ± 0.7	Same as for O_a/O_{b1}
Ea (kJ/mol)	52.9 ± 1	Same as for O_a/O_{b1}
$p_c f_0$ (1/min)	$1.7\text{x}10^8 \pm 7.1\text{x}10^7$	0
$p_{Li} f_0$ (1/min), i=1,2	$1.0\text{x}10^7 \pm 0.4\text{x}10^7$	$1.2\text{x}10^7 \pm 0.4\text{x}10^7$
$p_f f_0$ (1/min.mol)	$1.7\text{x}10^9 \pm 1\text{x}10^9$	0

[a] The standard errors are the square roots of the corresponding diagonal elements of the covariance matrix.

Model of the diffusion coefficients measured by DOSY NMR

Diffusion coefficients *D* are obtained from DOSY proton NMR experiments, in which the signal intensity at a given chemical shift, *I*, is fitted as a function of gradient strength *g* according to the Stejskal–Tanner equation:[225]

$$I(g) = I(0)\exp\left(-D(\gamma g\delta)^2\left(\Delta - \frac{\delta}{3}\right)\right)$$

in which γ is the proton magnetogyric ratio, δ is the duration of a pulse and Δ the time between two pulses. For a system composed of different species in dynamic equilibrium at the timescale of the experiment (*ca.* 0.1 ms), the diffusion coefficient $\overline{D}$ is an average performed over the *N* different coexisting species:

$$I(g) = I(0)\exp\left(-\overline{D}(\gamma g\delta)^2\left(\Delta - \frac{\delta}{3}\right)\right)$$
$$= \sum_{j=1}^{N} I_j(0)\exp\left(-D_j(\gamma g\delta)^2\left(\Delta - \frac{\delta}{3}\right)\right)$$

in which $I_j(0)$ is the contribution of species *j* having diffusion coefficient D_j. Developing to first order leads to:

$$\overline{D} = \frac{\sum_{j=1}^{N} I_j(0)D_j}{I(0)} = \frac{\sum_{j=1}^{N} q_j n_j D_j}{\sum_{j=1}^{N} q_j n_j}$$

with n_j the molar concentration in species *j*, and q_j the number of equivalent protons of species *j* at the considered NMR displacement.

Suppose that there are q_0 equivalent protons belonging to oligomer O_a, and that we measure the diffusion coefficient at the NMR displacement of these protons. The number of equivalent protons of any species, q_j, at this specific NMR displacement is then equal to the product of q_0 and the number of O_a oligomers which this species contains, $N_{O_{a,j}}$:

$$q_j = q_0 N_{Oa,j}$$

Therefore, the average diffusion coefficient measured at this specific NMR displacement only depends on the number of oligomers O_a in each species, and on the concentration of each species:

$$\overline{D} = \frac{\sum_{j=1}^{N} N_{Oa,j} n_j D_j}{\sum_{j=1}^{N} N_{Oa,j} n_j} = \frac{\sum_{j=1}^{N} N_{Oa,j} n_j D_j}{n}$$

with *n* the total initial concentration of O_a, as before. At equilibrium, the catalyst solution comprises free O_a and O_b oligomers each at the molar concentration *a*, cyclic di(oligomers) at the molar concentration $n^* = K^* a^2$, and linear poly(oligomeric) chains of different types, A_i, B_i, C_i and D_i, $i = 1 \ldots \infty$, at the molar concentrations $a_i = b_i = (K_1 K_2)^i a^{2i+1}$, $c_i = K_1^{i-1} K_2^i a^{2i}$, and $d_i = c_i K_1 / K_2$, respectively (see companion article). Therefore,

$$\overline{D} = \frac{1}{n}\left(a D_{Oa} + n^* D^* + \sum_{i=1}^{\infty} (i a_i D_{Ai} + (i+1) b_i D_{Bi} + i c_i D_{Ci} + i d_i D_{Di}) \right)$$

$$\overline{D} = \frac{1}{n}\Bigg(a D_{Oa} + K^* a^2 D^* + \sum_{i=1}^{\infty} \Big(i (K_1 K_2)^i a^{2i+1} D_{Ai} + (i+1)(K_1 K_2)^i a^{2i+1} D_{Bi} + i K_1^{i-1} K_2^i a^{2i} (D_{Ci} + D_{Di} K_1 / K_2) \Big) \Bigg)$$

in which D_{Oa}, D^*, D_{Ai}, D_{Bi}, D_{Ci}, D_{Di} are the diffusion coefficients of the

oligomer O_a, di(oligomeric) cycles, and linear chains A_i, B_i, C_i and D_i, respectively.

Since species A_i and B_i contain the same total number of oligomers, $D_{Ai} \approx D_{Bi}$; the same holds true for C_i and D_i, leading to $D_{Ci} \approx D_{Di}$. Defining

$$D_{ABi} = (D_{Ai} + D_{Bi})/2 \text{ and}$$

$$D_{CDi} = (D_{Ci} + D_{Di})/2$$

leads to:

$$\overline{D} \approx \frac{1}{n}\left(aD_{Oa} + K^* a^2 D^* + \sum_{i=1}^{\infty} \left((2i+1)(K_1K_2)^i a^{2i+1} D_{ABi} + 2i(K_1K_2)^{i-1} a^{2i} K_{12} D_{CDi} \right) \right)$$

in which $K_{12} = (K_1 + K_2)/2$.

To progress further, expressions are needed for the diffusion coefficients. In the non-draining approximation which applies to our oligomers which form coiled globules (as indicated by molecular dynamics simulations), the diffusion coefficient of a chain may be estimated by the Stokes-Einstein equation:

$$D = \frac{kT}{6\pi\mu_0 r}$$

in which k is the Boltzmann constant, T is temperature, μ_0 is the viscosity of the medium (*i.e.*, the solvent in our very dilute conditions), and r is the hydrodynamic radius of the chain. The hydrodynamic radii of the free oligomers, r_a and r_b for O_a and O_b, respectively, can be obtained by measuring their diffusion coefficients D_{Oa} and D_{Ob} by DOSY NMR at low concentrations.

The diffusion coefficients of the other species can then be obtained from the hydrodynamic radii of the oligomers. In the non-draining limit, the hydrodynamic radius of a poly(oligomeric) chain containing m O_a and n O_b oligomers is:

$$r_{mn} = \left(m r_a^{1/\nu} + n r_b^{1/\nu}\right)^{\nu}$$

in which $\nu = 0.5$ in a theta solvent and ~0.57 in a good solvent with excluded volume interactions, according to numerical simulations.[13] If r_a is close to r_b, which is the case in our system, then:

$$r_{mn} \approx \left(\frac{r_a^{1/\nu} + r_b^{1/\nu}}{2}\right)^{\nu} (m+n)^{\nu} = \bar{r}(m+n)^{\nu}$$

in which $\bar{r}$ is the average hydrodynamic radius of the two oligomers.

Then, the diffusion coefficients of the poly(oligomeric) components of the catalytic medium are:

$$D_{Ai} = D_{Bi} = D_{ABi} \approx \frac{kT}{6\pi\mu_0 \bar{r}} \frac{1}{(2i+1)^{\nu}}$$

$$D_{Ci} = D_{Di} = D_{CDi} \approx \frac{kT}{6\pi\mu_0\bar{r}}\frac{1}{(2i)^{\nu}}$$

Likewise, the diffusion coefficient of the di(oligomeric) cycles is:

$$D^* \approx \frac{kT}{6\pi\mu_0\left(r_a^{1/\nu} + r_b^{1/\nu}\right)^{\nu} h^*} \approx \frac{kT}{6\pi\mu_0\bar{r}}\frac{1}{2^{\nu}}\frac{1}{h^*}$$

in which h^* is the ratio between the hydrodynamic radius of a cyclic di(oligomer) and a linear di(oligomer). According to numerical simulations performed on flexible chains,[13] $h^* \approx 0.82$ in a theta solvent and $h^* \approx 0.89$ in a good solvent with positive excluded volume.

Introducing these diffusion coefficients in the expression of the diffusion constant of the mixture, measured by DOSY on a proton belonging to oligomer O_a, gives the diffusion coefficient of the equilibrated mixture:

$$\overline{D} \approx \frac{kT}{6\pi\mu_0\bar{r}}\frac{1}{n}\left(a + \frac{K^* a^2}{2^{\nu} h^*} + \sum_{i=1}^{\infty}\left((2i+1)^{(1-\nu)}(K_1K_2)^i a^{2i+1} + (2i)^{(1-\nu)}(K_1K_2)^{i-1}a^{2i}K_{12}\right)\right)$$

.

In this expression, the terms of the infinite series successively correspond to dimers ($2i$=2), trimers ($2i$+1=3), tetramers ($2i$=4), etc. The terms of this series will become negligible from some value $i = i_{\infty}$, leading to:

$$\overline{D} \approx \frac{kT}{6\pi\mu_0\bar{r}}\frac{1}{n}\left(a + \frac{K^* a^2}{2^{\nu} h^*} + \sum_{i=1}^{i_{\infty}}\left((2i+1)^{(1-\nu)}(K_1K_2)^i a^{2i+1} + (2i)^{(1-\nu)}(K_1K_2)^{i-1}a^{2i}K_{12}\right)\right)$$

.

The same equation is obtained if one considers a proton belonging to the O_b oligomer; therefore, this is the final expression for the average diffusion coefficient of the mixture.

Annex 4.7 DOSY results

Preparation of the O_a (C_6GMPTC_6) NMR sample was carried out by filling a 5 mm NMR tube with 0.50 mL of a solution of O_a in a 95:5 (*v*/*v*) CD_3CN/DMSO-d_6 mixture (*conc*: 1mmol/L (=0.5 mol%)). The O_{b1} ($C_6CI'IPDC_6$) NMR samples were instead prepared using two equimolar amounts of phenylhydrazine. The two compounds were mixed and then dissolved in 0.50 mL of a 95:5 (*v*/*v*) CD_3CN/DMSO-d_6 mixture (*conc*: 1mmol/L (=0.5 mol%)). A series of NMR samples containing an equimolar amount of O_a and O_{b1} was obtained by filling 5 mm NMR tubes with 0.25 mL of a solution of O_a and 0.25 mL of a solution of O_{b1}, both prepared at the same concentration with a 95:5 (*v*/*v*) CD_3CN/DMSO-d_6 solvent mixture. The final NMR tube concentrations are as follows: 1 mmol/L each (=0.5 mol%), 2.5 mmol/L each (=1.25 mol%) and 5 mmol/L each (=5 mol%). The average diffusion coefficient of the pure monomers and their hydrodynamic radii obtained from the Stokes-Einstein equation are in Table S1. Tables of DOSY diffusion coefficients depending on NMR displacement for samples O_a, O_{b1} and O_a/O_{b1} at different concentrations are available at https://doi.org/10.14428/DVN/1XCNOK.

Table A.4.2 Average diffusion coefficient $\overline{D}$ and hydrodynamic radius r of oligomers O_a and O_{b1}, measured by DOSY NMR in deuterated acetonitrile:DMSO 95:5 vol:vol (1 mM).

	O_a[a]	O_{b1}[a,b]
$\overline{D}$ ± standard error (m^2/s)[c]	$1.26 \times 10^{-9} \pm 3.4 \times 10^{-11}$ (1.9×10^{-10})	Meas. 1: $1.15 \times 10^{-9} \pm 3.8 \times 10^{-11}$ (2.3×10^{-10}) Meas. 2: $1.02 \times 10^{-9} \pm 3.6 \times 10^{-11}$ (2.3×10^{-10})
Average $\overline{D}$ (m^2/s)	$1.26 \times 10^{-9} \pm 3.4 \times 10^{-11}$	$1.09 \times 10^{-9} \pm 2.6 \times 10^{-11}$
r ± standard error (nm)[d]	0.5 ± 0.01	Meas.1: 0.54 ± 0.02 Meas. 2: 0.61 ± 0.02
Average *r* (nm)	0.5 ± 0.01	0.57 ± 0.014

[a] Standard deviations are indicated between parentheses. [b] Two independent measurements were performed for O_{b1}. [c] DOSY-average diffusion coefficient; the averages and standard errors were computed from the set of diffusion coefficients obtained for all protons. [d] DOSY-average hydrodynamic radius computed from the average diffusion coefficients using the Stokes-Einstein relationship.

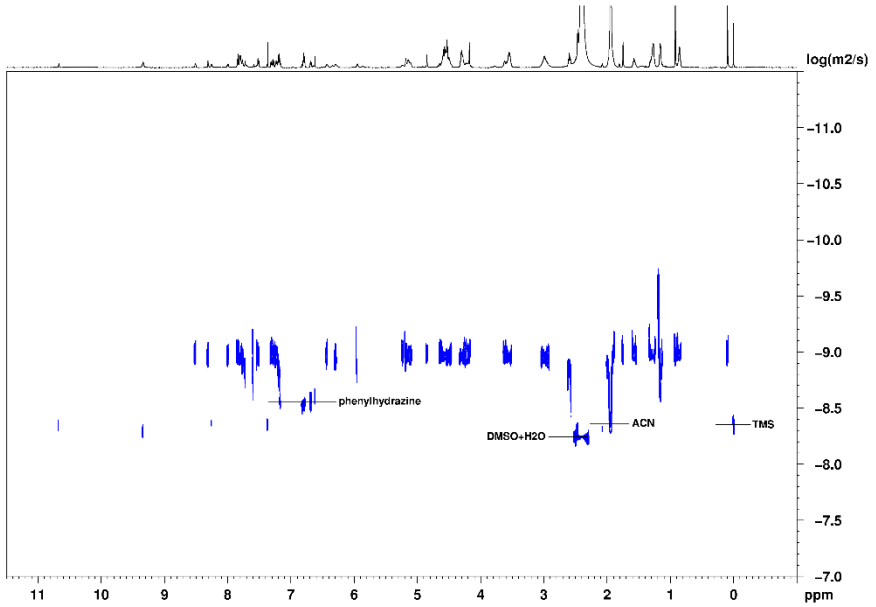


Figure A.4.13 ^{1}H DOSY NMR spectrum of O_a in 95:5 CD_3CN/DMSO-d_6 at 0.5 mol%.

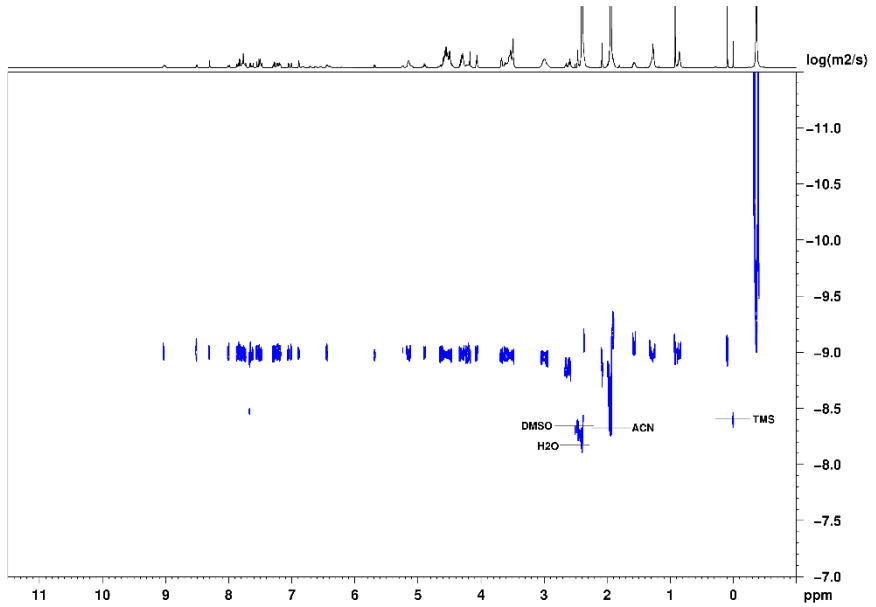


Figure A.4.14 ^{1}H DOSY NMR spectrum of O_{b1} in 95:5 CD_3CN/DMSO-d_6 at 0.5 mol%.

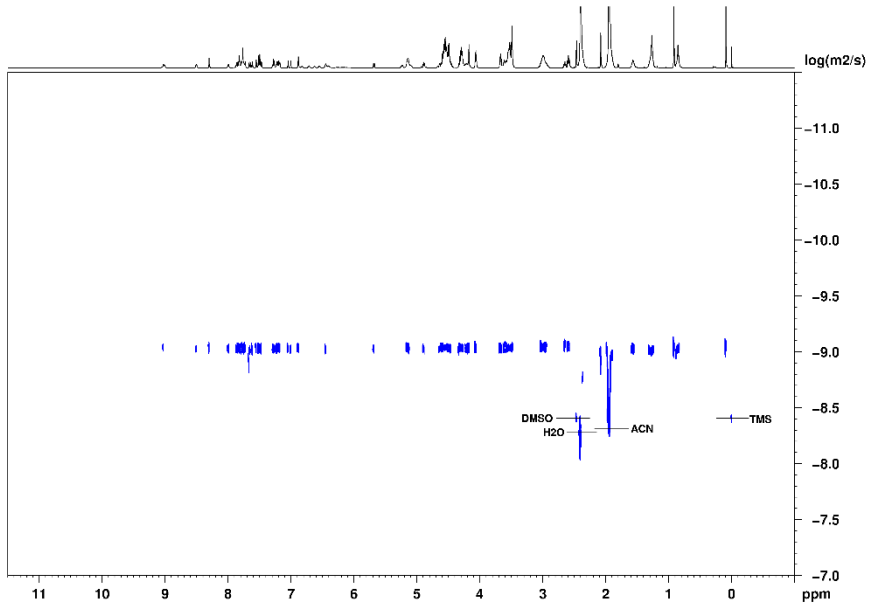


Figure A.4.15 ^{1}H DOSY NMR spectrum of O_{b1} in 95:5 CD_3CN/DMSO-d_6 at 0.5 mol%.

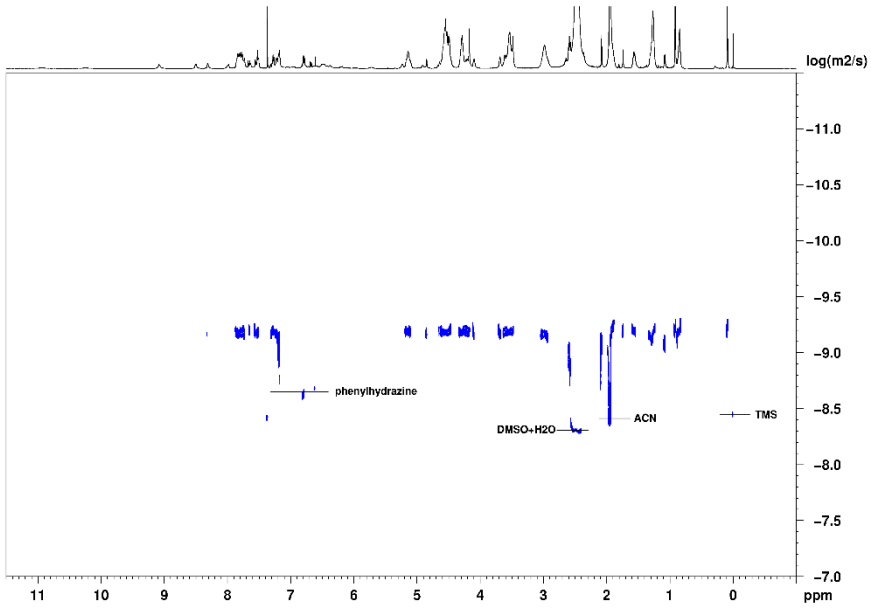


Figure A.4.16 ^{1}H DOSY NMR spectrum of O_a+O_{b1} in 95:5 CD_3CN/DMSO-d_6 at 0.5 mol%.

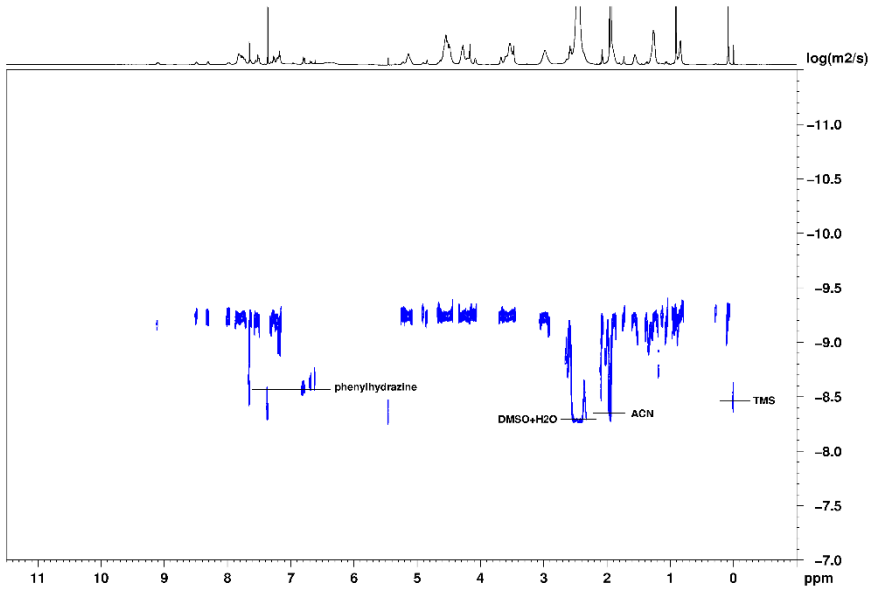


Figure A.4.17 ^{1}H DOSY NMR spectrum of O_a+O_{b1} in 95:5 CD_3CN/DMSO-d_6 at 1.25 mol%.

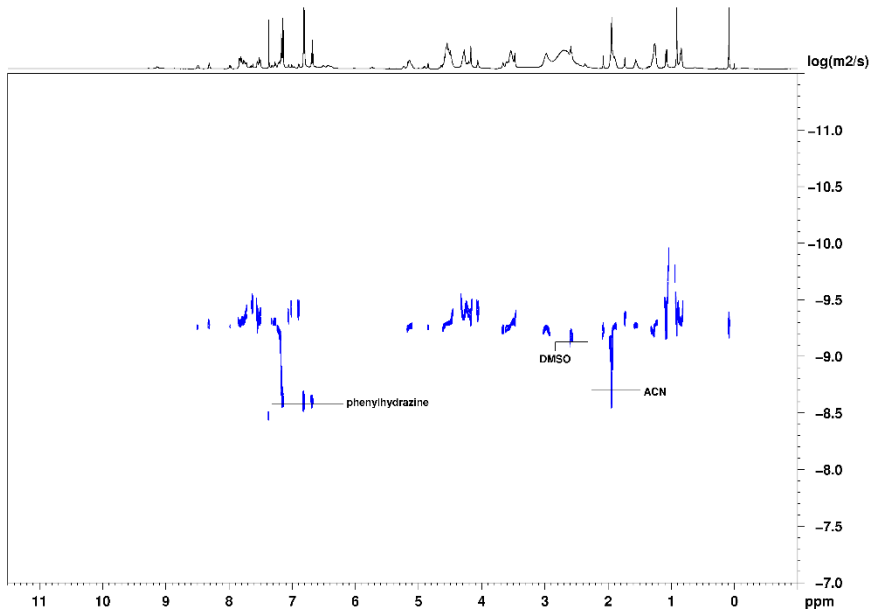


Figure A.4.18 ^{1}H DOSY NMR spectrum of O_a+O_{b1} in 95:5 CD_3CN/DMSO-d_6 at 5 mol%.

Chapter 5 Summary and outlook

The high catalytic effectiveness and remarkable substrate specificity of enzymes under mild conditions strongly depend on the chemical structure of the active site formed by the folding of residues from the primary sequence.[226] Inspired by the relationship between enzyme structure and catalytic activity, the mimicking of enzyme structure and function by synthetic molecules has been widely studied. Self-assembly of supramolecules and covalent macrocycles have proven to be effective ways to construct catalytic sites. These ideas were combined in this thesis to fabricate a Cu/TEMPO catalytic system that allows supramolecular cyclic structures to self-assemble in a dynamic way for the aerobic oxidation of alcohols.

For this, sequence-defined poly(triazole-carbamate) chains equipped with a variety of functional groups were synthesized. We showed that our route could provide precision tetramers with a specific sequence of nucleobases. These nucleobase-modified tetramers however displayed limited specific recognition of complementary strands, which was thus not discussed in this thesis. We then synthesized chains equipped with functional catalytic groups and complementary nucleobase recognition units. The catalytic moieties, which are involved in the copper-catalyzed aerobic oxidation of alcohols, and recognition pairs were distributed between two strands O_a and O_b for assembly driven by recognition units. O_a consists of a C_6GMPTC_6 sequence, complementary to O_{b1} ($C_6CI'IPDC_6$ sequence), O_{b2} (C_6CIIDC_6 sequence) or O_{b3} (C_6CIPDC_6 sequence).

In Chapter 4, the self-assembly of complementary oligomers into a dynamic library of poly(oligomers) and their catalytic activity were investigated. In the dynamic library, di(oligomeric) cycles exhibit high

catalytic, provided the two chains together comprise all the needed elements for catalysis, as is the case for the O_a/O_{b1} system. The lack of a required cooperative catalytic unit leads to a significant drop of catalytic activity, as shown for the O_a/O_{b2} and O_a/O_{b3} systems.

The formation of rather stable di(oligomeric) cycles was demonstrated by simulations, and we also showed how this complex catalytic system could be modeled by a limited set of equations, allowing us to predict properties and helping to suggest improvements. Although our system is essentially a proof-of-concept rather than the most efficient system one can devise, it is interesting to point that the system we created is very flexible, theoretically allowing to introduce a very wide variety of functional groups that could achieve an even more complex catalytic cycle. We also would like to point out that the use of precision oligomers for catalysis remains largely not explored, and that the formation of supramolecular structures based on such precision systems is a new field. This study has thus opened new avenues for research in both supramolecular chemistry and catalysis.

On the negative side, it is important to recognize that the catalytic cycle we selected, which involves the complexation of copper atoms, is especially complex. There are a variety of sites with which copper can interact in our systems; even though we checked that copper does not very strongly interact with sites other than the pyridyl-triazole group, it remains unclear whether the possible shuffling of copper atoms does not play a role in the catalytic properties. Further studies of copper complexation with our precision oligomers would thus be useful. Additionally, the efficiency of our system is optimal at relatively low concentrations, resulting in high TOFs but low rates, due to the inherent dynamic nature of cycle formation, which is disfavored at higher concentrations. As already discussed, there is little that can be

done to solve this issue, except playing with binding constants and entropy penalties. This is nevertheless easier said than done, and further work is clearly needed here. Playing with spatial anchoring of one end of the oligomers onto nanoparticles or introducing a mechanism that would covalently stabilize the cycles once formed, would be possible solutions.

More generally, one could wonder whether the synthetic effort needed to generate precision oligomers is really useful, compared to simpler methodologies that prove to be efficient, such as the grafting of each required unit on porous silica beads. At the end of this thesis, this remains an open question – although our work certainly provided interesting scientific insights, its application to real systems still remains elusive. Clearly, we are still far from reaching the sort of efficiency exhibited by natural systems such as enzymes.

Precision polymers are truly fantastic due to the various chemical synthesis methods utilized to strategically incorporate specific functional groups into molecules, allowing for tailoring of intramolecular conformation and intermolecular interactions.[227] This flexibility facilitates applications in catalysis, molecular recognition, data storage, and fabrication of complex self-assembled systems. However, despite their potential, synthetic precision polymers face challenges and complexities.

Efficient and simple synthesis methods need to be developed. The current synthetic strategies have their limitations; for instance, as discussed in chapter 2, solid-support synthesis is limited in scale, and support-free liquid-phase synthesis has complex and time-consuming purification procedures. Furthermore, even though backbone extension processes have high atom utilization, the overall yield is restricted after several chemical steps, consequently limiting the

chain length of the precision polymers.

In recent years, the exploration of precision polymer focused not only on sequence control but also on stereochemistry. Stereochemistry plays a crucial role in determining the physical and chemical properties of polymers, including thermal performance, folding, and assembly.[83] Stereocontrol is essential for creating macromolecules with complex structures and functions akin biological macromolecules. Methodologies to control stereochemistry often involve the utilization of chiral building blocks; while methods like IEG+ or automated synthesis can control chirality, challenges such as low yield and limited chain length of final products persist again. Additionally, the limited solubility of chiral-controlled polymers presents obstacles to the efficient synthesis of stereo and sequence-defined polymers.[83]

In addition to the challenges mentioned, the applications of precise polymers still need to be developed. Biomolecules have the capability to perform intricate tasks like replication and transcription, however, such applications currently are too far-reaching for synthetic precision polymers. Besides, industrial applications of precision polymer are limited due to elaborate production processes and high costs. Simplifying production and reducing costs are necessary for the industrial application of precision polymers. It is a rough road that leads to the heights of greatness. With more significant time and effort, we might finally achieve precise control over the sequence and stereochemistry of synthetic molecules and reach functions similar to biological macromolecules at affordable costs and efforts.

References

(1) Kohen, A. Role of Dynamics in Enzyme Catalysis: Substantial versus Semantic Controversies. *Acc. Chem. Res.* **2014**, *48* (2), 466–473.

(2) Agarwal, P. K.; Bernard, D. N.; Bafna, K.; Doucet, N. Enzyme Dynamics: Looking beyond a Single Structure. *ChemCatChem* **2020**, *12* (19), 4704–4720.

(3) Li, Q.; Min, J.; Zhang, J.; Meital Reches; Shen, Y.; Su, R.; Wang, Y.; Qi, W. Enzyme-Driven, Switchable Catalysis Based on Dynamic Self-Assembly of Peptides. *Angew. Chem. Int. Ed.* **2023**, *62* (41).

(4) Kurbasic, M.; Garcia, A. M.; Viada, S.; Marchesan, S. Heterochiral Tetrapeptide Self-Assembly into Hydrogel Biomaterials for Hydrolase Mimicry. *J. Pept. Sci.* **2022**, *28* (1), e3304.

(5) Liu, Q.; Akinori Kuzuya; Wang, Z.-G. Supramolecular Enzyme-Mimicking Catalysts Self-Assembled from Peptides. *iScience* **2023**, *26* (1), 105831–105831.

(6) Lou, Y.; Zhang, B.; Ye, X.; Wang, Z.-G. Self-Assembly of the de Novo Designed Peptides to Produce Supramolecular Catalysts with Built-in Enzyme-like Active Sites: A Review of Structure–Activity Relationship. *Mater. Today Nano* **2023**, *21* (100302), 100302.

(7) Zhang, C.; Xue, X.; Luo, Q.; Li, Y.; Yang, K.; Zhuang, X.; Jiang, Y.; Zhang, J.; Liu, J.; Zou, G.; Liang, X.-J. Self-Assembled Peptide Nanofibers Designed as Biological Enzymes for Catalyzing Ester Hydrolysis. *ACS Nano* **2014**, *8* (11), 11715–11723.

(8) Gulseren, G.; Khalily, M. A.; Tekinay, A. B.; Guler, M. O. Catalytic Supramolecular Self-Assembled Peptide Nanostructures for Ester Hydrolysis. *J. Mater. Chem. B* **2016**, *4* (26), 4605–4611.

(9) Tena-Solsona, M.; Nanda, J.; Díaz-Oltra, S.; Chotera, A.; Ashkenasy, G.; Escuder, B. Emergent Catalytic Behavior of Self-Assembled Low

Molecular Weight Peptide-Based Aggregates and Hydrogels. *Chem. Eur. J.* **2016**, *22* (19), 6687–6694.

(10) Zhang, C.; Shafi, R.; Lampel, A.; MacPherson, D.; Pappas, C. G.; Narang, V.; Wang, T.; Maldarelli, C.; Ulijn, R. V. Switchable Hydrolase Based on Reversible Formation of Supramolecular Catalytic Site Using a Self-Assembling Peptide. *Angew. Chem. Int. Ed.* **2017**, *129* (46), 14703–14707.

(11) Pimentel, E. B.; Peters-Clarke, T. M.; Coon, J. J.; Martell, J. D. DNA-Scaffolded Synergistic Catalysis. *J. Am. Chem. Soc.* **2021**, *143* (50), 21402–21409.

(12) Zheng, T.; Tang, Q.; Wan, L.; Zhao, Y.; Xu, R.; Xu, X.; Li, H.; Han, D. Controlled Self-Assembly of the Catalytic Core of Hydrolases Using DNA Scaffolds. *Nano Lett.* **2023**, *23* (6), 2081–2086.

(13) Cox, C. A.; Ogorek, A. N.; Jean Paul Habumugisha; Martell, J. D. Switchable DNA Photocatalysts for Radical Polymerization Controlled by Chemical Stimuli. *J. Am. Chem. Soc.* **2023**, *145* (3), 1818–1825.

(14) Drienovská, I.; Roelfes, G. Artificial Metalloenzymes for Asymmetric Catalysis by Creation of Novel Active Sites in Protein and DNA Scaffolds. *Isr. J. Chem.* **2014**, *55* (1), 21–31.

(15) Garcia, M. A. A.; Hu, Y.; Willner, I. Switchable Supramolecular Catalysis Using DNA-Templated Scaffolds. *Chem. Commun.* **2016**, *52* (10), 2153–2156.

(16) Marek, J. J.; Singh, R. P.; Heuer, A.; Hennecke, U. Enantioselective Catalysis by Using Short, Structurally Defined DNA Hairpins as Scaffold for Hybrid Catalysts. *Chem. Eur. J.* **2017**, *23* (25), 6004–6008.

(17) Green, S. A.; Montgomery, H. R.; Benton, T. R.; Chan, N. J.; Nelson, H. M. Regulating Transition-Metal Catalysis through Interference by Short RNAs. *Angew. Chem. Int. Ed.* **2019**, *58* (46), 16400–16404.

(18) Mundsinger, K.; Aiden Izuagbe; Tuten, B.; Roesky, P.; Barner-

Kowollik, C. Single Chain Nanoparticles in Catalysis. *Angew. Chem. Int. Ed.* **2023**, *63* (7), e202311734.

(19) Piane, J. J.; Chamberlain, L. E.; Huss, S.; Alameda, L. T.; Hoover, A. C.; Elacqua, E. Organic Photoredox-Catalyzed Cycloadditions under Single-Chain Polymer Confinement. *ACS Catal.* **2020**, *10* (22), 13251–13256.

(20) Liu, Y.; Pujals, S.; Stals, P. J. M.; Paulöhrl, T.; Presolski, S. I.; Meijer, E. W.; Albertazzi, L.; Palmans, A. R. A. Catalytically Active Single-Chain Polymeric Nanoparticles: Exploring Their Functions in Complex Biological Media. *J. Am. Chem. Soc.* **2018**, *140* (9), 3423–3433.

(21) Terashima, T.; Mes, T.; De Greef, T. F. A.; Gillissen, M. A. J.; Besenius, P.; Palmans, A. R. A.; Meijer, E. W. Single-Chain Folding of Polymers for Catalytic Systems in Water. *J. Am. Chem. Soc.* **2011**, *133* (13), 4742–4745.

(22) Rubio-Cervilla, J.; González, E.; Pomposo, J. Advances in Single-Chain Nanoparticles for Catalysis Applications. *Nanomaterials* **2017**, *7* (10), 341.

(23) Rothfuss, H.; Knöfel, N. D.; Roesky, P. W.; Barner-Kowollik, C. Single-Chain Nanoparticles as Catalytic Nanoreactors. *J. Am. Chem. Soc.* **2018**, *140* (18), 5875–5881.

(24) Nghiem, T. L.; Coban, D.; Tjaberings, S.; Gröschel, A. H. Recent Advances in the Synthesis and Application of Polymer Compartments for Catalysis. *Polymers* **2020**, *12* (10), 2190–2190.

(25) Artar, M.; Terashima, T.; Sawamoto, M.; Meijer, E. W.; Palmans, A. Understanding the Catalytic Activity of Single-Chain Polymeric Nanoparticles in Water. *J. Polym. Sci. Part A: Polym. Chem.* **2013**, *52* (1), 12–20.

(26) Kim, T.; Jian, T.; Jin, B.; Nguyen, D. T.; Zuckermann, R. N.; Chen, C. L. Designed Metal-Containing Peptoid Membranes as Enzyme

Mimetics for Catalytic Organophosphate Degradation. *ACS Appl. Mater. Interfaces.* **2023**, *15* (44), 51191–51203.

(27) Prathap, K. J.; Maayan, G. Metallopeptoids as Efficient Biomimetic Catalysts. *Chem. Commun.* **2015**, *51* (55), 11096–11099.

(28) Zhang, X.; Yang, C.; An, P.; Cui, C.; Ma, Y.; Liu, H.; Wang, H.; Yan, X.; Li, G.; Tang, Z. Creating Enzyme-Mimicking Nanopockets in Metal-Organic Frameworks for Catalysis. *Sci. Adv.* ***2022****, 8 (40),* eadd5678.

(29) Wang, D.; Wu, H.; Phua, S. Z. F.; Yang, G.; Qi Lim, W.; Gu, L.; Qian, C.; Wang, H.; Guo, Z.; Chen, H.; Zhao, Y. Self-Assembled Single-Atom Nanozyme for Enhanced Photodynamic Therapy Treatment of Tumor. *Nat. Commun.* **2020**, *11* (1), 357.

(30) Jiao, L.; Wang, Y.; Jiang, H. L.; Xu, Q. Metal-Organic Frameworks as Platforms for Catalytic Applications. *Adv. Mater.* **2017**, *30* (37), 1703663.

(31) Raynal, M.; Portier, F.; Leeuwen, van; Bouteiller, L. Tunable Asymmetric Catalysis through Ligand Stacking in Chiral Rigid Rods. *J. Am. Chem. Soc.* **2013**, *135* (47), 17687–17690.

(32) Girvin, Z. C.; Gellman, S. H. Exploration of Diverse Reactive Diad Geometries for Bifunctional Catalysis via Foldamer Backbone Variation. *J. Am. Chem. Soc.* **2018**, *140* (39), 12476–12483.

(33) Girvin, Z. C.; Gellman, S. H. Foldamer Catalysis. *J. Am. Chem. Soc.* **2020**, *142* (41), 17211–17223.

(34) Nagata, Y.; Takeda, R.; Suginome, M. Asymmetric Catalysis in Chiral Solvents: Chirality Transfer with Amplification of Homochirality through a Helical Macromolecular Scaffold. *ACS Cent. Sci.* **2019**, *5* (7), 1235–1240.

(35) Li, Y.; Caumes, X.; Raynal, M.; Bouteiller, L. Modulation of Catalyst Enantioselectivity through Reversible Assembly of Supramolecular Helices. *Chem. Commun.* **2019**, *55* (15), 2162–2165.

(36) Yamamoto, T.; Murakami, R.; Komatsu, S.; Michinori Suginome. Chirality-Amplifying, Dynamic Induction of Single-Handed Helix by Chiral Guests to Macromolecular Chiral Catalysts Bearing Boronyl Pendants as Receptor Sites. *J. Am. Chem. Soc.* **2018**, *140* (11), 3867–3870.

(37) Zimbron, J. M.; Caumes, X.; Li, Y.; Thomas, C. M.; Raynal, M.; Bouteiller, L. Real-Time Control of the Enantioselectivity of a Supramolecular Catalyst Allows Selecting the Configuration of Consecutively Formed Stereogenic Centers. *Angew. Chem.* **2017**, *129* (45), 14204–14207.

(38) Ready, J. M.; Jacobsen, E. N. Highly Active Oligomeric (Salen)Co Catalysts for Asymmetric Epoxide Ring-Opening Reactions. *J. Am. Chem. Soc.* **2001**, *123* (11), 2687–2688.

(39) Gianneschi, N. C.; Bertin, P. A.; Nguyen, S. T.; Mirkin, C. A.; Zakharov, L. N.; Rheingold, A. L. A Supramolecular Approach to an Allosteric Catalyst. *J. Am. Chem. Soc.* **2003**, *125* (35), 10508–10509.

(40) Zheng, X.; Jones, C. W.; Weck, M. Ring-Expanding Olefin Metathesis: A Route to Highly Active Unsymmetrical Macrocyclic Oligomeric Co-Salen Catalysts for the Hydrolytic Kinetic Resolution of Epoxides. *J. Am. Chem. Soc.* **2007**, *129* (5), 1105–1112.

(41) Liu, Y.; Rawlston, J.; Swann, A. T.; Takatani, T.; Sherrill, C. D.; Ludovice, P. J.; Weck, M. The Bigger, the Better: Ring-Size Effects of Macrocyclic Oligomeric Co(Iii)-Salen Catalysts. *Chem. Sci.* **2011**, *2* (3), 429–438.

(42) Bai, C.; Tian, B.; Zhao, T.; Huang, Q.; Wang, Z. Cyclodextrin-Catalyzed Organic Synthesis: Reactions, Mechanisms, and Applications. *Molecules* **2017**, *22* (9), 1475.

(43) Jørgen Kjems; Ferapontova, E.; Gothelf, K. V. *Nucleic Acid Nanotechnology*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2014; Vol. 29.

(44) Whitesides, G.; Mathias, J.; Seto, C. Molecular Self-Assembly and Nanochemistry: A Chemical Strategy for the Synthesis of Nanostructures. *Science* **1991**, *254* (5036), 1312–1319.

(45) Whitesides, G. M.; Boncheva, M. Beyond Molecules: Self-Assembly of Mesoscopic and Macroscopic Components. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 4769–4774.

(46) Wei, W.; Bai, F. Self-Assembly of Na Self-Assembly of Nanoparticles noparticles. *World Scientific series in nanoscience and nanotechnology/World scientific series in nanoscience and nanotechnology* **2017**, *14*, 1–56.

(47) Wang, L.; Gong, C.; Yuan, X.; Wei, G. Controlling the Self-Assembly of Biomolecules into Functional Nanomaterials through Internal Interactions and External Stimulations: A Review. *Nanomaterials* **2019**, *9* (2), 285.

(48) Gao, Y.; Wang, L.; Zhang, X.; Zhou, Z.; Shen, X.; Hu, H.; Sun, R.; Tang, J. Advances in Self-Assembled Peptides as Drug Carriers. *Pharmaceutics* **2023**, *15* (2), 482–482.

(49) Lee, Y. S. *Self-Assembly and Nanotechnology Systems: Design, Characterization, and Applications*; John Wiley & Sons, 2011.

(50) Yang, S. K.; Zimmerman, S. C. Hydrogen Bonding Modules for Use in Supramolecular Polymers. *Isr. J. Chem.* **2013**, *53* (8), 511–520.

(51) Liu, Y.; Wang, L.; Zhao, L.; Zhang, Y.; Li, Z.-T.; Huang, F. Multiple Hydrogen Bonding Driven Supramolecular Architectures and Their Biomedical Applications. *Chem. Soc. Rev.* **2024**, *53* (3), 1592–1623.

(52) Verjans, J.; Hoogenboom, R. Supramolecular Polymer Materials Based on Ureidopyrimidinone Quadruple Hydrogen Bonding Units. *Prog. Polym. Sci.* **2023**, *142*, 101689.

(53) Li, Z.; Wu, L. Z. *Lecture Notes in Chemistry 88 Hydrogen Bonded Supramolecular Materials*; Springer, 2015.

(54) Yang, L.; Tan, X.; Wang, Z.; Zhang, X. Supramolecular Polymers: Historical Development, Preparation, Characterization, and Functions. *Chem. Rev.* **2015**, *115* (15), 7196–7239.

(55) Kyogoku, Y.; Lord, R. C.; Rich, A. An Infrared Study of the Hydrogen-Bonding Specificity of Hypoxanthine and Other Nucleic Acid Derivatives. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. **1969**, *179* (1), 10–17.

(56) Sikder, A.; Esen, C.; O'Reilly, R. K. Nucleobase-Interaction-Directed Biomimetic Supramolecular Self-Assembly. *Acc. Chem. Res.* **2022**, *55* (12), 1609–1619.

(57) Sivakova, S.; Rowan, S. J. Nucleobases as Supramolecular Motifs. *Chem. Soc. Rev.* **2005**, *34* (1), 9–21.

(58) Li, J.; Wang, Z.; Hua, Z.; Tang, C. Supramolecular Nucleobase-Functionalized Polymers: Synthesis and Potential Biological Applications. *J. Mater. Chem. B*, **2020**, *8* (8), 1576–1588.

(59) Saenger, W. *Principles of Nucleic Acid Structure*; Springer Science & Business Media, 2013.

(60) Liang, Y.; Wnuk, S. Modification of Purine and Pyrimidine Nucleosides by Direct C-H Bond Activation. *Molecules* **2015**, *20* (3), 4874–4901.

(61) Sengupta, S.; Das, P. Application of Diazonium Chemistry in Purine Modifications: A Focused Review. *J. Heterocycl. Chem.* **2021**, *59* (1), 5–21.

(62) Berney, M.; McGouran, J. F. Methods for Detection of Cytosine and Thymine Modifications in DNA. *Nat. Rev. Chem.* **2018**, *2* (11), 332–348.

(63) Xu, H.; Chen, J.; Cheng, J.; Kong, L.; Chen, X.; Inoue, M.; Liu, Y.; Kriaucionis, S.; Zhao, M.; Song, C.-X. Modular Oxidation of Cytosine Modifications and Their Application in Direct and Quantitative Sequencing of 5-Hydroxymethylcytosine. *J. Am. Chem.*

Soc. **2023**, *145* (13), 7095–7100.

(64) Kozak, W.; Demkowicz, S.; Daśko, M.; Rachon, J.; Rak, J. Modifications at the C(5) Position of Pyrimidine Nucleosides. *Russ. Chem. Rev.* **2020**, *89* (3), 281–310.

(65) Krueger, A. T.; Lu, H.; Alex; Kool, E. T. Synthesis and Properties of Size-Expanded DNAs: Toward Designed, Functional Genetic Systems. *Acc. Chem. Res.* **2006**, *40* (2), 141–150.

(66) Malyshev, D. A.; Romesberg, F. E. The Expanded Genetic Alphabet. *Angew. Chem. Int. Ed.* **2015**, *54* (41), 11930–11944.

(67) Chen, J.; Liu, B.; Wang, X.; Liu, X.; Yang, D.; Yang, Y. Constructing Artificial Nucleobase Compilation to Enable Precise Molecular Medicine†. *Chin. J. Chem.* **2023**, *41* (11), 1385–1398.

(68) Piccirilli, J. A.; Benner, S. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature* **1990**, *343* (6253), 33–37.

(69) Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.-J.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J.; Georgiadis, M. M.; Benner, S. A. Hachimoji DNA and RNA: A Genetic System with Eight Building Blocks. *Science* **2019**, *363* (6429), 884–887.

(70) Blight, B. A.; Hunter, C. A.; Leigh, D. A.; McNab, H.; Thomson, P. I. T. An AAAA–DDDD Quadruple Hydrogen-Bond Array. *Nat. Chem.* **2011**, *3* (3), 244–248.

(71) Jorgensen, W. L.; Julianto Pranata. Importance of Secondary Interactions in Triply Hydrogen Bonded Complexes: Guanine-Cytosine vs Uracil-2,6-Diaminopyridine. *J. Am. Chem. Soc.* **1990**, *112* (5), 2008–2010.

(72) Del Prado, A.; González-Rodríguez, D.; Wu, Y. L. Functional Systems Derived from Nucleobase Self-Assembly. *ChemistryOpen*

2020, *9* (4), 409–430.

(73) Ciesielski, A.; El Garah, M.; Masiero, S.; Samorì, P. Self-Assembly of Natural and Unnatural Nucleobases at Surfaces and Interfaces. *Small* **2016**, *12* (1), 83–95.

(74) Lugger, S. J. D.; Houben, S. J. A.; Foelen, Y.; Debije, M. G.; Schenning, A. P. H. J.; Mulder, D. J. Hydrogen-Bonded Supramolecular Liquid Crystal Polymers: Smart Materials with Stimuli-Responsive, Self-Healing, and Recyclable Properties. *Chem. Res.* **2021**, *122* (5), 4946–4975.

(75) Sijbesma, R. P.; Meijer, E. Self-Assembly of Well-Defined Structures by Hydrogen Bonding. *Curr. Opin. Colloid Interface Sci.* **1999**, *4* (1), 24–32.

(76) Chen, S.; Zhang, K.; Li, Z.; Wu, Y.; Zhu, B.; Zhu, J. Hydrogen-Bonded Supramolecular Adhesives: Synthesis, Responsiveness, and Application. *Supramol. Mater.* **2023**, *2*, 100032.

(77) Liu, Y.; Wang, L.; Zhao, L.; Zhang, Y.; Li, Z. T.; Li, Z. T.; Huang, F.; Huang, F. Multiple Hydrogen Bonding Driven Supramolecular Architectures and Their Biomedical Applications. *Chem. Soc. Rev.* **2024**, *53* (3), 1592–1623.

(78) Hakobyan, K.; Noble, B. B.; Xu, J. The Current Science of Sequence-Defined Macromolecules. *Prog. Polym. Sci.* **2023**, *147*, 101754.

(79) Shi, Q.; Deng, Z.; Hou, M.; Hu, X.; Liu, S. Engineering Precise Sequence-Defined Polymers for Advanced Functions. *Prog. Polym. Sci.* **2023**, *141*, 101677.

(80) Neve, J. D.; Haven, J. J.; Maes, L.; Junkers, T. Sequence-Definition from Controlled Polymerization: The next Generation of Materials. *Polym. Chem.* **2018**, *9* (38), 4692–4705.

(81) Nanjan, P.; Porel, M. Sequence-Defined Non-Natural Polymers: Synthesis and Applications. *Polym. Chem.* **2019**, *10* (40), 5406–5424.

(82) Solleder, S. C.; Schneider, R. V.; Wetzel, K. S.; Boukis, A. C.; Meier, M. A. R. Recent Progress in the Design of Monodisperse, Sequence-Defined Macromolecules. *Macromol. Rapid Commun.* **2017**, *38* (9), 1600711.

(83) Szweda, R. Sequence-and Stereo-Defined Macromolecules: Properties and Emerging Functionalities. *Prog. Polym. Sci.* **2023**, *145*, 101737.

(84) Li, J.; Qin, Q.; Kardas, S.; Mathieu Fossépré; Surin, M.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. Sequence Rules the Functional Connections and Efficiency of Catalytic Precision Oligomers. *ACS Catal.* **2022**, *12* (3), 2126–2131.

(85) Goldup, S. M. The End of the Beginning of Mechanical Stereochemistry. *Acc. Chem. Res.* **2024**, *57* (12), 1696–1708.

(86) Clover, T. M.; O'Neill, C. L.; Appavu, R.; Lokhande, G.; Gaharwar, A. K.; Posey, A. E.; White, M. A.; Rudra, J. S. Self-Assembly of Block Heterochiral Peptides into Helical Tapes. *J. Am. Chem. Soc.* **2020**, *142* (47), 19809.

(87) Tariq, A.; Fayzan Shakir, H. M. Stereospecific Polymerization Techniques. In *Advanced Functional Polymers: Synthesis to Applications* (pp. 3-21). Singapore: Springer Nature Singapore.

(88) Satoh, K.; Kamigaito, M. Stereospecific Living Radical Polymerization: Dual Control of Chain Length and Tacticity for Precision Polymer Synthesis. *Chem. Rev.* **2009**, *109* (11), 5120–5156.

(89) Kramer, J. W.; Treitler, D. S.; Dunn, E. W.; Castro, P. M.; Thierry Roisnel; Thomas, C. M.; Coates, G. W. Polymerization of Enantiopure Monomers Using Syndiospecific Catalysts: A New Approach to Sequence Control in Polymer Synthesis. *J. Am. Chem. Soc.* **2009**, *131* (44), 16042–16044.

(90) Li, H.; Shakaroun, R. M.; Guillaume, M.; Carpentier, J.-F. Recent Advances in Metal-Mediated Stereoselective Ring-Opening

Polymerization of Functional Cyclic Esters towards Well-Defined Poly(Hydroxy Acid)s: From Stereoselectivity to Sequence-Control. *Chem. Eur.J*. **2020**,*26* (1), 128–138.

(91) Solleder, S. C.; Kam Chiu Tam. Sequence Control in Polymer Chemistry through the Passerini Three-Component Reaction. *Angew. Chem.* **2014**, *53* (3), 711–714.

(92) Solleder, S. C.; Wetzel, K. S.; Meier, M. A. R. Dual Side Chain Control in the Synthesis of Novel Sequence-Defined Oligomers through the Ugi Four-Component Reaction. *Polym. Chem.* **2015**, *6* (17), 3201–3204.

(93) Leibfarth, F. A.; Johnson, J. A.; Jamison, T. F. Scalable Synthesis of Sequence-Defined, Unimolecular Macromolecules by Flow-IEG. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (34), 10617–10622.

(94) Barnes, J. C.; Ehrlich, D. J. C.; Gao, A. X.; Leibfarth, F. A.; Jiang, Y.; Zhou, E.; Jamison, T. F.; Johnson, J. A. Iterative Exponential Growth of Stereo- and Sequence-Controlled Polymers. *Nat. Chem.* **2015**, *7* (10), 810–815.

(95) Li, Z.; Ren, X.; Sun, P.; Ding, H.; Li, S.; Zhao, Y.; Zhang, K. Protecting-Group-Free Iterative Exponential Growth Method for Synthesizing Sequence-Defined Polymers. *ACS Macro Lett.* **2021**, *10* (2), 223–230.

(96) Ahmed, S. A.; Tanaka, M. Synthesis of Oligo(Ethylene Glycol) toward 44-Mer. *J. Org. Chem.* **2006**, *71* (26), 9884–9886.

(97) Niculescu-Duvaz, D.; Getaz, J.; Springer, C. J. Long Functionalized Poly(Ethylene Glycol)S of Defined Molecular Weight: Synthesis and Application in Solid-Phase Synthesis of Conjugates. *Bioconjugate Chem.* **2008**, *19* (4), 973–981.

(98) Erik; Zwikker, J. W.; Roest, M. R.; Jenneskens, L. W. A Highly Selective Synthesis of Monodisperse Oligo(Ethylene Glycols). *J. Org. Chem.* **1992**, *57* (24), 6678–6680.

(99) Konrad, W.; Fengler, C.; Putwa, S.; Barner-Kowollik, C. Protection-Group-Free Synthesis of Sequence-Defined Macromolecules via Precision λ-Orthogonal Photochemistry. *Angew Chem. Int. Ed.* **2019**, *58* (21), 7133–7137.

(100) Zydziak, N.; Konrad, W.; Feist, F.; Afonin, S.; Weidner, S.; Barner-Kowollik, C. Coding and Decoding Libraries of Sequence-Defined Functional Copolymers Synthesized via Photoligation. *Nat. Commun.* **2016**, *7* (1).

(101) Zydziak, N.; Feist, F.; Huber, B.; Mueller, J. O.; Barner-Kowollik, C. Photo-Induced Sequence Defined Macromolecules via Hetero Bifunctional Synthons. *Chem*. *Commun.* **2014**, *51* (10), 1799–1802.

(102) Zhao, B.; Gao, Z.; Zheng, Y.; Gao, C. Scalable Synthesis of Positively Charged Sequence-Defined Functional Polymers. *J. Am. Chem. Soc.* **2019**, *141* (11), 4541–4546.

(103) Huang, Z.; Noble, B. B.; Corrigan, N.; Chu, Y.; Satoh, K.; Thomas, D. S.; Hawker, C. J.; Moad, G.; Kamigaito, M.; Coote, M. L.; Boyer, C.; Xu, J. Discrete and Stereospecific Oligomers Prepared by Sequential and Alternating Single Unit Monomer Insertion. *J. Am. Chem. Soc.* **2018**, *140* (41), 13392–13406.

(104) Tao, W.; He, W.; Feng, X.; Liu, G.; Shi, Q.; Tan, J.; Hu, J.; Yang, S.; Liu, G.; Yang, R. Cationic Single-Unit Monomer Insertion (CSUMI): From Discrete Oligomers to the α-/ω-End and In-Chain Sequence-Regulated Polymers. *J. Am. Chem. Soc.* **2023**, *145* (6), 3636–3646.

(105) Hill, S. A.; Gerke, C.; Hartmann, L. Recent Developments in Solid-Phase Strategies towards Synthetic, Sequence-Defined Macromolecules. *Chem. Asian J.* **2018**, *13* (23), 3611–3622.

(106) Kaligian, K. L.; Sprachman, M. M. Controlled Polymers: Accessing New Platforms for Material Synthesis. *Mol. Syst. Des. Eng.* **2019**, *4* (1), 144–161.

(107) Shi, Q.; Zhang, Z.; Liu, S. Precision Sequence-Defined Polymers:

From Sequencing to Biological Functions. *Angew Chem. Int. Ed.. **2023**, 136* (6), e202313370.

(108) Simpson, A. J.; Simpson, M. J.; Soong, R. Nuclear Magnetic Resonance Spectroscopy and Its Key Role in Environmental Research. *Environ. Sci. Technol.* **2012**, *46* (21), 11488–11496.

(109) Mlynárik, V. Introduction to Nuclear Magnetic Resonance. *Anal. chem.* **2017**, *529*, 4–9.

(110) Noble, B. B.; Coote, M. L. First Principles Modelling of Free-Radical Polymerisation Kinetics. *Int. Rev. Phys. Chem.* **2013**, *32* (3), 467–513.

(111) Xu, Z.; Lin, J.; Zhang, Q.; Wang, L.; Tian, X. Theoretical Simulations of Nanostructures Self-Assembled from Copolymer Systems. *Polym. Chem.* **2016**, *7* (23), 3783–3811.

(112) Labuta, J.; Ishihara, S.; Šikorský, T.; Futera, Z.; Shundo, A.; Hanyková, L.; Burda, J. V.; Ariga, K.; Hill, J. P. NMR Spectroscopic Detection of Chirality and Enantiopurity in Referenced Systems without Formation of Diastereomers. *Nat. Comm.* **2013**, *4*(1), 2188.

(113) Zou, J.; Zhao, G. Q.; Zhao, G. L.; Yu, J. G. Fast and Sensitive Recognition of Enantiomers by Electrochemical Chiral Analysis: Recent Advances and Future Perspectives. *Coord. Chem. Rev.* **2022**, *471*(51), 214732.

(114) Cavazzini, A.; Pasti, L.; Massi, A.; Marchetti, N.; Dondi, F. Recent Applications in Chiral High Performance Liquid Chromatography: A Review. *Anal. Chim. Acta* **2011**, *706* (2), 205–222.

(115) Yang, C.; Hu, W.; Guo, X. Precise Detection, Control and Synthesis of Chiral Compounds at Single-Molecule Resolution. *Nano-Micro Lett.* **2023**, *15* (1), 1–18.

(116) uHPLCs. *Chiral Column HPLC All You Should Know - uHPLCs*. Professional HPLC Column Hardware Consumables Supplier.

(117) Ward, K. D.; Bravenec, A. D.; Ward, T. J. Chiral Separations by

High-performance Liquid Chromatography. *Encycl. Anal. Chem. Appl. Theory Instrum.* **2006**, 1–28.

(118) van Genabeek, B.; Lamers, B. A. G.; Hawker, C. J.; Meijer, E. W.; Gutekunst, W. R.; Schmidt, B. V. K. J. Properties and Applications of Precision Oligomer Materials; Where Organic and Polymer Chemistry Join Forces. *J. Polym. Sci.* **2021**, *59* (5), 373–403.

(119) Deng, Z.; Shi, Q.; Tan, J.; Hu, J.; Liu, S. Sequence-Defined Synthetic Polymers for New-Generation Functional Biomaterials. ACS Mater. Lett. **2021**, *3* (9), 1339–1356.

(120) Nanjan, P.; Porel, M. Sequence-Defined Non-Natural Polymers: Synthesis and Applications. *Polym. Chem.* **2019**, *10* (40), 5406–5424.

(121) Meier, M. A. R.; Barner-Kowollik, C.. A New Class of Materials: Sequence-Defined Macromolecules and Their Emerging Applications. *Adv. Mater.* **2019**, *31* (26), 1806027.

(122) Destefano, A. J.; Segalman, R. A.; Davidson, E. C. Where Biology and Traditional Polymers Meet: The Potential of Associating Sequence-Defined Polymers for Materials Science. *JACS Au* **2021**, *1* (10), 1556–1571.

(123) Li, Z.; Cai, B.; Yang, W.; Chen, C. L. Hierarchical Nanomaterials Assembled from Peptoids and Other Sequence-Defined Synthetic Polymers. *Chem. Rev.* **2021**, *121* (22), 14031–14087.

(124) Kwon, Y.; Kim, K. T. Crystallization-Driven Self-Assembly of Block Copolymers Having Monodisperse Poly(Lactic Acid)s with Defined Stereochemical Sequences. *Macromolecules* **2021**, *54* (22), 10487–10498.

(125) Soete, M.; Mertens, C.; Badi, N.; Du Prez, F. E. Reading Information Stored in Synthetic Macromolecules. *J. Am. Chem. Soc* **2022**, *144*(49), 22378–22390.

(126) Martens, S.; Landuyt, A.; Espeel, P.; Devreese, B.; Dawyndt, P.; Du Prez, F. Multifunctional Sequence-Defined Macromolecules for

Chemical Data Storage. *Nat. Commun. 2018 91* **2018**, *9* (1), 1–8.

(127) Hartmann, L.; Börner, H. G. Precision Polymers: Monodisperse, Monomer-Sequence-Defined Segments to Target Future Demands of Polymers in Medicine. *Adv. Mater.* **2009**, *21* (32–33), 3425–3431.

(128) Chandra, P.; Jonas, A.; Fernandes, A. E. Synthesis of Discrete Catalytic Oligomers and Their Potential in Silica-Supported Cooperative Catalysis. *RSC Adv.* **2019**, *9* (25), 14194–14197.

(129) Chandra, P.; Jonas, A. M.; Fernandes, A. E. Spatial Coordination of Cooperativity in Silica-Supported Cu/TEMPO/Imidazole Catalytic Triad. *ACS Catal.* **2018**, *8* (7), 6006–6011.

(130) Chandra, P.; Jonas, A. M.; Fernandes, A. E. Sequence and Surface Confinement Direct Cooperativity in Catalytic Precision Oligomers. *J. Am. Chem. Soc.* **2018**, *140* (15), 5179–5184.

(131) Maayan, G.; Ward, M. D.; Kirshenbaum, K. Folded Biomimetic Oligomers for Enantioselective Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13679–13684.

(132) Silva, T. F. S.; Martins, L. M. D. R. S. Recent Advances in Copper Catalyzed Alcohol Oxidation in Homogeneous Medium. *Molecules* **2020**, *25* (3), 748.

(133) Hoover, J. M.; Stahl, S. S. Highly Practical Copper(I)/TEMPO Catalyst System for Chemoselective Aerobic Oxidation of Primary Alcohols. *J. Am. Chem. Soc.* **2011**, *133* (42), 16901–16910.

(134) Hoover, J. M.; Steves, J. E.; Stahl, S. S. Copper(I)/TEMPO Catalyzed Aerobic Oxidation of Primary Alcohols to Aldehydes with Ambient Air. *Nat. Protoc.* **2012**, *7* (6), 1161–1166.

(135) Wang, Q.-Q. Supramolecular Catalysis Using Organic Macrocycles. *In Handbook of Macrocyclic Supramolecular Assembly*. Springer, Singapore.

(136) Lee, J.; Hong, S.; Lee, J.; Kim, S.; Kim, J.; Kim, M. Strategies in

Metal-Organic Framework-Based Catalysts for the Aerobic Oxidation of Alcohols and Recent Progress. *Bull. Korean Chem. Soc.* **2021**, *42* (3), 359–368.

(137) Taher, A.; Kim, D. W.; Lee, I.-M. Highly Efficient Metal Organic Framework (MOF)-Based Copper Catalysts for the Base-Free Aerobic Oxidation of Various Alcohols †. *RSC Adv.* **2017**, *7* (29), 17806–17812.

(138) Jiao, Y.; Tang, B.; ucheng Zhang, Y.; iang-Fei Xu, J.; Wang, Z.; Xi Zhang, nd; Jiao, Y.; Tang, B.; Zhang, Y.; Xu, D.-F.; Wang, Z.; Zhang, X. Highly Efficient Supramolecular Catalysis by Endowing the Reaction Intermediate with Adaptive Reactivity. *Angew Chem. Int. Ed.* **2018**, *57* (21), 6077–6081.

(139) Safaei, E.; Bahrami, H.; Pevec, A.; Kozlevčar, B.; Jagličić, Z. Copper(II) Complex of New Non-Innocent O-Aminophenol-Based Ligand as Biomimetic Model for Galactose Oxidase Enzyme in Aerobic Oxidation of Alcohols. *J. Mol. Struct.* **2017**, *1133*, 526–533.

(140) Semmelhack, M. F.; Schmid, C. R.; Cortes, D. A.; Chou, C. S. Oxidation of Alcohols to Aldehydes with Oxygen and Cupric Ion, Mediated by Nitrosonium Ion. *J. Am. Chem. Soc.* **1984**, *106* (11), 3374–3376.

(141) Semmelhack, M. F.; Schmid, C. R.; Cortés, D. A. Mechanism of the Oxidation of Alcohols by 2,2,6,6-Tetramethylpiperidine Nitrosonium Cation. *Tetrahedron Lett.* **1986**, *27* (10), 1119–1122.

(142) Cao, Q.; Dornan, L. M.; Rogan, L.; Hughes, N. L.; Muldoon, M. J. Aerobic Oxidation Catalysis with Stable Radicals. *Chem. Commun.* **2014**, *50* (35), 4524–4543.

(143) Hoover, J. M.; Ryland, B. L.; Stahl, S. S. Copper/TEMPO-Catalyzed Aerobic Alcohol Oxidation: Mechanistic Assessment of Different Catalyst Systems. *ACS Catal.* **2013**, *3* (11), 2599–2605.

(144) Ma, S.; Liu, J.; Li, S.; Chen, B.; Cheng, J.; Kuang, J.; Liu, Y.; Wan, B.;

Wang, Y.; Ye, J.; Yu, Q.; Yuan, W.; Yu, S. Development of a General and Practical Iron Nitrate/TEMPO-Catalyzed Aerobic Oxidation of Alcohols to Aldehydes/Ketones: Catalysis with Table Salt. *Adv. Synth. Catal.* **2011**, *353* (6), 1005–1017.

(145) Chou, S.; Nelson, J. A.; Spencer, T. A. Oxidation and Mass Spectra of 4,4-Dimethyloxazolidine-N-Oxyl (Doxyl) Derivatives of Ketones. *J. Org. Chem.* **1974**, *39* (16), 2356–2361.

(146) Ma, Z.; Huang, Q.; Bobbitt, J. M. Oxoammonium Salts. 5. A New Synthesis of Hindered Piperidines Leading to Unsymmetrical TEMPO-Type Nitroxides. Synthesis and Enantioselective Oxidations with Chiral Nitroxides and Chiral Oxoammonium Salts. *J. Org. Chem.* **1993**, *58* (18), 4837–4843.

(147) Yang, G.; Zhu, W.; Zhang, P.; Xue, H.; Wang, W.; Tian, J.; Song, M. Recyclable Carbon Supported Copper-Manganese Oxide for Selective Aerobic Oxidation of Alcohols in Combination with 2,2,6,6-Tetramethylpiperidyl-1-Oxyl under Neutral Condition. *Adv. Synth. Catal.* **2008**, *350* (4), 542–546.

(148) Yang, G.; Ma, J.; Wang, W.; Zhao, J.; Lin, X.; Zhou, L.; Gao, X. Heterogeneous Cu–Mn Oxides Mediate Efficiently TEMPO-Catalyzed Aerobic Oxidation of Alcohols. *Catal. Lett.* **2006**, *112* (1-2), 83–87.

(149) Gilhespy, M.; Lok, M.; Baucherel, X. Polymer-Supported Nitroxyl Radical Catalyst for Selective Aerobic Oxidation of Primary Alcohols to Aldehydes. *Chem*. *Comm*. **2005**, No. 8, 1085–1086.

(150) Cecchetto, A.; Fontana, F.; Minisci, F.; Recupero, F. Efficient Mn–Cu and Mn–Co–TEMPO-Catalysed Oxidation of Alcohols into Aldehydes and Ketones by Oxygen under Mild Conditions. *Tetrahedron Lett.* **2001**, *42* (38), 6651–6653.

(151) Lagerblom, K.; Lagerspets, E.; Keskiväli, J.; Cook, C.; Ekholm, F.; Parviainen, A.; Repo, T. Practical Aerobic Oxidation of Alcohols: A Ligand-Enhanced 2,2,6,6-Tetramethylpiperidine-1-Oxy/Manganese

Nitrate Catalyst System. *ChemCatChem* **2017**, *9* (20), 3880–3887.

(152) Lagerblom, K.; Keskiväli, J.; Parviainen, A.; Mannisto, J.; Repo, T. Selective Aerobic Oxidation of Alcohols with NO_3^- Activated Nitroxyl Radical/Manganese Catalyst System. *ChemCatChem* **2018**, *10* (13), 2908–2914.

(153) Dijksman, A.; Arends, I. W. C. E.; Sheldon, R. A. The Ruthenium/TEMPO-Catalysed Aerobic Oxidation of Alcohols. *Platinum Met. Rev.* **2001**, *45* (1), 15–19.

(154) Sheldon, R. A.; Arends, I. W. C. E.; ten Brink, G.-J.; Dijksman, A. Green, Catalytic Oxidations of Alcohols. *Acc. Chem. Res.* **2002**, *35* (9), 774–781.

(155) Dijksman, A.; Marino-González, A.; Payeras, A. M.; Arends, I. W. C. E.; Sheldon, R. A. Efficient and Selective Aerobic Oxidation of Alcohols into Aldehydes and Ketones Using Ruthenium/TEMPO as the Catalytic System. *J. Am. Chem. Soc.* **2001**, *123* (28), 6826–6833.

(156) Marais, L.; Swarts, A. J. Biomimetic Cu/Nitroxyl Catalyst Systems for Selective Alcohol Oxidation. *Catalysts* **2019**, *9* (5), 395.

(157) Sheldon, R. A.; Arends, I. W. C. E. Catalytic Oxidations Mediated by Metal Ions and Nitroxyl Radicals. *J. Mol. Catal. A:Chem.* **2006**, *251* (1-2), 200–214.

(158) Whittaker, J. W. Free Radical Catalysis by Galactose Oxidase. *Chem. Rev.* **2003**, *103* (6), 2347–2364.

(159) Ma, Z.; Mahmudov, K. T.; Aliyeva, V. A.; Gurbanov, A. V.; Armando. TEMPO in Metal Complex Catalysis. *Coord. Chem. Rev.* **2020**, *423*, 213482.

(160) Semmelhack, M. F.; Schmid, C. R.; Cortes, D. A.; Chou, C. S. Oxidation of Alcohols to Aldehydes with Oxygen and Cupric Ion, Mediated by Nitrosonium Ion. *J. Am. Chem. Soc.* **1984**, *106* (11), 3374–3376.

(161) Gamez, P.; Arends, I. W. C. E.; Reedijk, J.; Sheldon, R. A. Copper(Ii)-Catalysed Aerobic Oxidation of Primary Alcohols to Aldehydes. *Chem. Commun.* **2003**, No. 19, 2414.

(162) Kumpulainen, E. T. T.; Koskinen, A. M. P. Catalytic Activity Dependency on Catalyst Components in Aerobic Copper-TEMPO Oxidation. *Chem. Eur. J.* **2009**, *15* (41), 10901–10911.

(163) Hoover, J. M.; Ryland, B. L.; Stahl, S. S. Mechanism of Copper(I)/TEMPO-Catalyzed Aerobic Alcohol Oxidation. *J. Am. Chem. Soc.* **2013**, *135* (6), 2357–2367.

(164) Fernandes, A. E.; Riant, O.; Jensen, K. F.; Jonas, A. M. Molecular Engineering of Trifunctional Supported Catalysts for the Aerobic Oxidation of Alcohols. *Angew. Chem.* **2016**, *128* (37), 11210–11214.

(165) Adomeit, S.; Rabeah, J.; Surkus, A. E.; Bentrup, U.; Brückner, A. Effects of Imidazole-Type Ligands in CuI/TEMPO-Mediated Aerobic Alcohol Oxidation. *Inorg. Chem.* **2016**, *56* (1), 684–691.

(166) Liu, Z.; Shen, Z.; Zhang, N.; Zhong, W.; Liu, X. Aerobic Oxidation of Alcohols Catalysed by Cu(I)/NMI/TEMPO System and Its Mechanistic Insights. *Catal. Lett.* **2018**, *148* (9), 2709–2718.

(167) Ma, R.; Xiao, Z.; Zhong, W.; Lü, C. X.; Shen, Z.; Zhao, D.; Liu, X. The Superiority of Cuprous Chloride to Iodide in the Selective Aerobic Oxidation of Benzylic Alcohols at Ambient Temperature. *Appl. Organomet. Chem.* **2021**, *35* (7), e6245.

(168) Bete, S. C.; Würtele, C.; Otte, M. A Bio-Inspired Imidazole-Functionalised Copper Cage Complex. *Chem. Commun.* **2019**, *55* (30), 4427–4430.

(169) Jain, R.; Gibson, T. J.; Mashuta, M. S.; Buchanan, R. M.; Grapperhaus, C. A. Copper Catalysed Aerobic Oxidation of Benzylic Alcohols in an Imidazole Containing N_4 Ligand Framework. *Dalt. Trans.* **2016**, *45* (45), 18356–18364.

(170) Rabeah, J.; Bentrup, U.; Stößer, R.; Brückner, A. Selective Alcohol

Oxidation by a Copper TEMPO Catalyst: Mechanistic Insights by Simultaneously Coupled Operando EPR/UV-Vis/ATR-IR Spectroscopy. *Angew. Chem. Int. Ed.* **2015**, *54* (40), 11791–11794.

(171) Li, J.; Leclercq, M.; Fossepré, M.; Surin, M.; Glinel, K.; Jonas, A. M.; Fernandes, A. E. Discrete Multifunctional Sequence-Defined Oligomers with Controlled Chirality. *Polym. Chem.* **2020**, *11* (24), 4040–4046.

(172) Mundlapati, V. R.; Imani, Z.; D'Mello, V. C.; Brenner, V.; Gloaguen, E.; Baltaze, J. P.; Robin, S.; Mons, M.; Aitken, D. J. N-H···X Interactions Stabilize Intra-Residue C5 Hydrogen Bonded Conformations in Heterocyclic α-Amino Acid Derivatives. *Chem. Sci.* **2021**, *12* (44), 14826–14832.

(173) Jiang, Y.; Golder, M. R.; Nguyen, H. V. T.; Wang, Y.; Zhong, M.; Barnes, J. C.; Ehrlich, D. J. C.; Johnson, J. A. Iterative Exponential Growth Synthesis and Assembly of Uniform Diblock Copolymers. *J. Am. Chem. Soc.* **2016**, *138* (30), 9369–9372.

(174) Jinks, M. A.; Howard, M.; Rizzi, F.; Goldup, S. M.; Burnett, A. D.; Wilson, A. J. Direct Detection of Hydrogen Bonds in Supramolecular Systems Using ^{1}H-^{15}N Heteronuclear Multiple Quantum Coherence Spectroscopy. *J. Am. Chem. Soc.* **2022**, *144* (50), 23127–23133.

(175) Piccirilli, J. A.; Benner, S. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. Enzymatic Incorporation of a New Base Pair into DNA and RNA Extends the Genetic Alphabet. *Nature* **1990**, *343* (6253), 33–37.

(176) Feibush, Binyamin.; Figueroa, Alvaro.; Charles, Rosita.; Onan, K. D.; Feibush, Pnina.; Karger, B. L. Chiral Separation of Heterocyclic Drugs by HPLC: Solute-Stationary Phase Base-Pair Interactions. *J. Am. Chem. Soc.* **1986**, *108* (12), 3310–3318.

(177) Gramlich, P. M. E.; Warncke, S.; Gierlich, J.; Carell, T. Click–Click–Click: Single to Triple Modification of DNA. *Angew. Chem. Int. Ed.* **2008**, *47* (18), 3442–3444.

(178) Fernandes, A.; Viterisi, A.; Coutrot, F.; Potok, S.; Leigh, David A.; Aucagne, V.; Papot, S. Rotaxane-Based Propeptides: Protection and Enzymatic Release of a Bioactive Pentapeptide. *Angew. Chem. Int. Ed.* **2009**, *48* (35), 6443–6447.

(179) Fernandes, A. E.; Olivier Riant; Jonas, A. M.; Jensen, K. F. One “Click” to Controlled Bifunctional Supported Catalysts for the Cu/TEMPO-Catalyzed Aerobic Oxidation of Alcohols. *RSC Adv.* **2016**, *6* (43), 36602–36605.

(180) Kocalar, S.; Zhou, A.; Salahifar, A.; Njoo, E. Spectrophotometric Comparison of 4-Nitrophenyl Carbonates & Carbamates as Base-Labile Protecting Groups. *J. Emerg. Investig.* **2022**, *5* (1), 1–6.

(181) Bodanszky, M.; du Vigneaud, V. A Method of Synthesis of Long Peptide Chains Using a Synthesis of Oxytocin as an Example. *J. Am. Chem. Soc.* **1959**, *81* (21), 5688–5691.

(182) Uneyama, K. Fluoride Ion-Catalyzed Desilylative-Defluorination for Synthetic Organic Chemistry. *J. Fluor. Chem.* **2007**, *128* (10), 1087–1090.

(183) Zhan, P.; Peil, A.; Jiang, Q.; Wang, D.; Mousavi, S.; Xiong, Q.; Shen, Q.; Shang, Y.; Ding, B.; Lin, C.; Ke, Y.; Liu, N. Recent Advances in DNA Origami-Engineered Nanomaterials and Applications. *Chem. Rev.* **2023**, *123* (7), 3976–4050.

(184) Jiang, W.; Li, J.; Lin, Z.; Guo, J.; Ma, J.; Wang, Z.; Zhang, M.; Wu, Y. Recent Advances of DNA Origami Technology and Its Application in Nanomaterial Preparation. *Small Struct.* **2023**, *4* (8), 2200376.

(185) Dey, S.; Fan, C.; Gothelf, K. V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M. A. D.; Saccà, B.; Simmel, F. C.; Yan, H.; Zhan, P. DNA Origami. *Nat. Rev. Methods Primers* **2021**, *1* (1), 1–24.

(186) Yakovchuk, P. Base-Stacking and Base-Pairing Contributions into Thermal Stability of the DNA Double Helix. *Nucleic Acids Res.* **2006**, *34* (2), 564–574.

(187) Iadevaia, G.; Hunter, C. A. Recognition-Encoded Synthetic Information Molecules. *Acc. Chem. Res.* **2023**, *56* (6), 712–727.

(188) Strom, K. R.; Szostak, J. W. Folding and Duplex Formation in Sequence-Defined Aniline Benzaldehyde Oligoarylacetylenes. *J. Am. Chem. Soc.* **2022**, *144* (40), 18350–18358.

(189) Lin, Y.; Thomas, M. R.; Gelmi, A.; Leonardo, V.; Pashuck, E. T.; Maynard, S. A.; Wang, Y.; Stevens, M. M. Self-Assembled 2D Free-Standing Janus Nanosheets with Single-Layer Thickness. *J. Am. Chem. Soc.* **2017**, *139* (39), 13592–13595.

(190) Chidchob, P.; Edwardson, T. G. W.; Serpell, C. J.; Sleiman, H. F. Synergy of Two Assembly Languages in DNA Nanostructures: Self-Assembly of Sequence-Defined Polymers on DNA Cages. *J. Am. Chem. Soc.* **2016**, *138* (13), 4416–4425.

(191) Weller, M. G. The Mystery of Homochirality on Earth. *Life* **2024**, *14* (3), 341–341.

(192) Baeriswyl, S.; Personne, H.; Di Bonaventura, I.; Köhler, T.; Van Delden, C.; Stocker, A.; Javor, S.; Reymond, J. L. A Mixed Chirality α-Helix in a Stapled Bicyclic and a Linear Antimicrobial Peptide Revealed by X-Ray Crystallography. *RSC Chem. Biol.* **2021**, *2* (6), 1608–1617.

(193) Kramer, R. A.; Bleicher, K. H.; Wennemers, H. Design and Synthesis of Nucleoproline Amino Acids for the Straightforward Preparation of Chiral and Conformationally Constrained Nucleopeptides. *Helv. Chim. Acta* **2012**, *95* (12), 2621–2634.

(194) Vo; D. D. Staedel, C.; Zehnacker, L.; Benhida, R.; Darfeuille, F.; Duca, M. Targeting the Production of Oncogenic MicroRNAs with Multimodal Synthetic Small Molecules. *ACS Chem. Biol.* **2014**, *9* (3), 711–721.

(195) Kolb, R.; Bach, N. C.; Sieber, S. A. β-Sultams Exhibit Discrete Binding Preferences for Diverse Bacterial Enzymes with Nucleophilic

Residues. *Chem. Commun.* **2014**, *50* (4), 427–429.

(196) Cheng, C. C.; Wang, Y. S.; Chang, F. C.; Lee, D. J.; Yang, L. C.; Chen, J. K. Supramolecular Assembly-Induced Enhanced Emission of Electrospun Nanofibers. *Chem. Commun.* **2015**, *51* (4), 672–675.

(197) Porcheddu, A.; Giacomelli, G.; Piredda, I.; Carta, M.; Nieddu, G. A Practical and Efficient Approach to PNA Monomers Compatible with Fmoc-Mediated Solid-Phase Synthesis Protocols. *Eur. J. Org. Chem.* **2008**, *2008* (34), 5786–5797.

(198) Ryland, B. L.; McCann, S. D.; Brunold, T. C.; Stahl, S. S. Mechanism of Alcohol Oxidation Mediated by Copper(II) and Nitroxyl Radicals. *J. Am. Chem. Soc.* **2014**, *136* (34), 12166–12173.

(199) Warm, K.; Tripodi, G.; Andris, E.; Mebs, S.; Kuhlmann, U.; Dau, H.; Hildebrandt, P.; Roithová, J.; Ray, K. Spectroscopic Characterization of a Reactive $[Cu_2(\mu\text{-}OH)_2]^{2+}$ Intermediate in Cu/TEMPO Catalyzed Aerobic Alcohol Oxidation Reaction. *Angew Chem. Int. Ed.* **2021**, *60* (42), 23018–23024. https://doi.org/10.1002/anie.202108442.

(200) Zhong, W.; Luo, J.; Liu, Z.; Zhan, G.; Zhu, L.; Lu, C.; Shen, Z.; Li, X.; Liu, X. The Mechanistic Diversity of the Selective Aerobic Oxidation of Alcohols Catalyzed by Systems Derived from CuI and a Diamine Ligand. *Inorg. Chem. Front.* **2023**, *10* (10), 3139–3150.

(201) Qin, Q.; Li, J.; Dellemme, D.; Fossépré, M.; Barozzino-Consiglio, G.; Nekkaa, I; Boborodea, A.; Fernandes, A. E.; Glinel, K.; Surin, M.; Jonas, A. M. Dynamic Self-Assembly of Supramolecular Catalysts from Precision Macromolecules. *Chem. Sci.* **2023**, *14* (35), 9283–9292.

(202) Mannam, S.; Alamsetti, S. K.; Sekar, G. Aerobic, Chemoselective Oxidation of Alcohols to Carbonyl Compounds Catalyzed by a DABCO-Copper Complex under Mild Conditions. *Adv. Synth. Catal.* **2007**, *349* (14-15), 2253–2258.

(203) Al-Hunaiti, A.; Abu-Radaha, B.; Wraith, D.; Repo, T. Catalytic

Behaviour of the Cu(I)/L/TEMPO System for Aerobic Oxidation of Alcohols - a Kinetic and Predictive Model. *RSC Adv.* **2022**, *12* (13), 7864–7871.

(204) Fernandes, A. E.; Jonas, A. M. Design and Engineering of Multifunctional Silica-Supported Cooperative Catalysts. *Catal. Today* **2019**, *334*, 173–186.

(205) Kardas, S.; Fossépré, M.; Lemaur, V.; Fernandes, A. E.; Glinel, K.; Jonas, A. M.; Surin, M. Revealing the Organization of Catalytic Sequence-Defined Oligomers via Combined Molecular Dynamics Simulations and Network Analysis. *J. Chem. Inf. Model.* **2022**, *62* (11), 2761–2770.

(206) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminf.* **2012**, *4* (1), 17.

(207) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(208) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260.

(209) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21* (2), 132–146.

(210) Stendardo, E.; Pedone, A.; Cimino, P.; Cristina Menziani, M.; Crescenzi, O.; Barone, V. Extension of the AMBER Force-Field for the Study of Large Nitroxides in Condensed Phases: An Ab Initio Parameterization. *Phys. Chem. Chem. Phys.* **2010**, *12* (37), 11697.

(211) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A.

Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174.

(212) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246* (1-2), 122–129.

(213) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542–1555.

(214) Schrodinger, L. L. C. The PyMOL Molecular Graphics System. *Version* **2015**, *1*, 8.

(215) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem.* **1999**, *20* (2), 217–230.

(216) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput*. **2012**, 8 (9), 3314–3321.

(217) Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 4. Accuracies of MM/PBSA and MM/GBSA Methodologies Evaluated by Various Simulation Protocols Using PDBbind Data Set. *Phys. Chem. Chem. Phys.* **2014**, *16* (31), 16719–16729.

(218) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10* (5), 449–461.

(219) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504.

(220) Assenov, Y.; Ramírez, F.; Schelhorn, S.-E.; Lengauer, T.; Albrecht, M. Computing Topological Parameters of Biological Networks. *Bioinformatics* **2007**, *24* (2), 282–284.

(221) Ding, Y.; Rousseau, R.; Wolfram, D. Measuring Scholarly Impact. *Methods Pract. Springer Int. Publ.* **2014**.

(222) Rosvall, M.; Bergstrom, C. T. Maps of Random Walks on Complex Networks Reveal Community Structure. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (4), 1118–1123.

(223) Rosvall, M.; Axelsson, D.; Bergstrom, C. THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS the Map Equation. *Eur. Phys. J. Spec. Top.* **2009**, *178* (1), 13–23.

(224) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27* (3), 379–423.

(225) Garcia, M. A. A.; Hu, Y.; Willner, I. Switchable Supramolecular Catalysis Using DNA-Templated Scaffolds. *Chem. Commun.* **2016**, *52* (10), 2153–2156.

(226) Liu, S.; Du, P.; Sun, H.; Yu, H.; Wang, Z. Bioinspired Supramolecular Catalysts from Designed Self-Assembly of DNA or Peptides. *ACS Catal.* **2020**, *10* (24), 14937–14958.

(227) Lutz, J. F. The Future of Sequence-Defined Polymers. *Eur. Polym. J.* **2023**, *199*, 112465–112465.

List of scientific communications

Articles

1. Grafskaia, K.; **Qin, Q.**; Li, J.; Magnin, D.; Dellemme, D.; Surin, M.; Glinel, K.; Jonas, A. Chain Stretching in Brushes Favors Sequence Recognition for Nucleobase-Functionalized Flexible Precise Oligomers. *Soft Matter* **2024**.

2. **Qin, Q.**; Li, J.; Dellemme, D.; Fossépré, M.; Barozzino-Consiglio, G.; Imane Nekkaa; Boborodea, A.; Fernandes, A. E.; Glinel, K.; Surin, M.; Jonas, A. M. Dynamic Self-Assembly of Supramolecular Catalysts from Precision Macromolecules. *Chem. Sci.* **2023**, *14* (35), 9283–9292.

3. Li, J.; **Qin, Q.;** Kardas, S.; Fossépré, M.; Surin, M.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. Sequence Rules the Functional Connections and Efficiency of Catalytic Precision Oligomers. *ACS Catal.* **2022**, *12* (3), 2126–2131.

4. Dellemme, D.; Leclercq, M.; Fossépré, M.; Li, J.; **Qin, Q**.; Jonas, A. M; Glinel, K.; Surin, M. Folding and self-assembly of sequence-defined oligomers containing nucleobases. *Chim. Nouv.* **2021**, No 138.

Conferences

1. Dellemme, D.; **Qin, Q.**; Fossépré, M.; Glinel, K.; Jonas, A. M.; Surin, M. Understanding the Self-Assembly of Precision Catalytic Oligomers through Molecular Dynamics Simulations. Montpellier, France, June 1-2, 2023.

2. Grafskaia, K.; **Qin, Q.**; Li, J.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. In-situ characterization by liquid-cell ellipsometry of interfacial

recognition of sequence-defined oligomers. Brussels, Belgium, July 13-14, 2023.

3. Dellemme, D.; Fosséprè, M.; **Qin, Q.**; Glinel, K.; Jonas, A. M.; Surin, M. Understanding the Self-Assembly of Precision Catalytic Oligomers through Molecular Dynamics Simulations. Biarritz, France, October 4-7, 2022.

4. Grafskaia, K.; **Qin, Q.**; Li, J.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. In-situ characterization of molecular recognition by sequence-defined oligomers via liquid-cell ellipsometry. UMONS, Belgium, September 12-14, 2022

5. **Qin, Q.**; Grafskaia, K.; Li, J.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. Chiral sequence-defined oligomers for molecular recognition and self-assembling at the solid-liquid interface. Bordeaux, France, June 13-16, 2022.

6. **Qin, Q.**; Li, J.; Fernandes, A. E.; Glinel, K.; Jonas, A. M. Recognition and self-assembly of stereo-controlled and sequence-defined oligomers at solid-liquid interfaces. Veldhoven, Netherlands, November 7-9, 2021.