Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Aromatic foldamers: mimicking the DNA surface to target DNA-binding proteins

Jiaojiao Wu

aus

Shanxi,Linfen, China

2025

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Ivan Huc betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 21.10.2025 _____

Jiaojiao Wu

Dissertation eingereicht am	21.10.2025
1. Gutachter:	Prof. Dr. Ivan Huc
2. Gutachter:	Prof. Dr. Franz Paintner
Mündliche Prüfung am	08.12.2025

Table of contents

1. List of Publications

Published:

V. Corvaglia, J. Wu, D. Deepak, M. Loos and I. Huc*, Enhancing the Features of DNA Mimic Foldamers for Structural Investigations, *Chem. Eur. J*. **2024**, 30, e202303650. (doi.org/10.1002/chem.202303650).

(See chapter 5)

D. Deepak, J. Wu[+], V. Corvaglia[+], L. Allmendinger, M. Streckenbach, P. Tinnefeld and Ivan Huc*, DNA Mimic Foldamer Recognition of a Chromosomal Protein, *Angew. Chem. Int. Ed*. **2025**, 64, e202422958. (doi.org/10.1002/anie.202422958).

(See chapter 6)

[+] These authors contributed equally

2. Abstract

Protein-nucleic acid interactions are fundamental in various biological processes and thus represent attractive targets for therapeutic intervention. Consequently, numerous molecules mimicking the feature of DNA have been developed and interfere with pharmacologically relevant protein-DNA interactions. Among these strategies, aromatic foldamers stand out as a new class of molecules that give access to functions beyond natural DNA. Specifically, aromatic oligoamides that mimic the charge surface of B-DNA have shown the potential in inhibiting the activity of several non-sequence-selective DNA-binding proteins. Despite these advances, a major challenge remains: designing DNA mimics that can outcompete DNA for binding to sequence-selective DNA-binding proteins.

In this thesis, we attempted to address this challenge by feature-driven design of DNA mimic foldamers. We introduced stereo genic centre, C_2-symmetry and sticky ends to enhance foldamer's similarity to double-stranded B-DNA. These new designs lead to the quantitative control of helical handedness, the construction of a foldamer mimicking palindromic DNA, and, significantly, the first crystal structure of a DNA mimic foldamer bearing anionic phosphonic side chains. Based on the optimization of DNA mimic foldamers, we attempted to investigate their interactions with the chromosomal protein Sac7d, a non-selective DNA-binding protein. Using a comprehensive set of techniques (including SPR, ITC, NMR, AFM and single-crystal X-ray crystallography), we demonstrated that DNA mimic foldamers bind to the DNA binding site of Sac7d but though a different mode, and that the binding affinity is orders of magnitude better than natural DNA. However the interactions between Sac7d and DNA mimic foldamers remain non sequence-selective. To advance toward sequence-selective protein recognition, we further modulated the foldamer's groove features by introducing new building blocks. Solution and solid-state studies confirm the new monomers' conformations and show their incorporation significantly changes helix flexibility and enhances groove characteristics, which is crucial to recognize specific DNA-binding proteins in the future.

Overall, the work presented here validates the concept of using DNA surface mimicry as a potent DNA competitor to target DNA-binding proteins. Importantly, we obtained the first crystal structure of a protein-foldamer complex, providing the solid foundation for structural-based design to improve the binding affinity and specificity. These results, combined with improvements to the groove features of our DNA mimics, represent a major step toward optimizing DNA mimic foldamers for highly specific protein recognition.

3. Introduction

3.1. Nucleic acids and their mimicry

3.1.1 Structure of nucleic acids

Nucleic acids are large biomolecules that create, encode, and store information in all living organisms on earth. Structurally, they are composed of nucleotide monomers, each consisting of a five-carbon sugar (deoxyribose in DNA and ribose in RNA), a nitrogenous base (adenine, cytosine, guanine, thymine in DNA; uracil in RNA), and a phosphate group (Figure 1a).[1] These nucleotides are covalently linked via phosphodiester bonds, giving nucleic acids directional polymer chains from the 5' to 3' end.

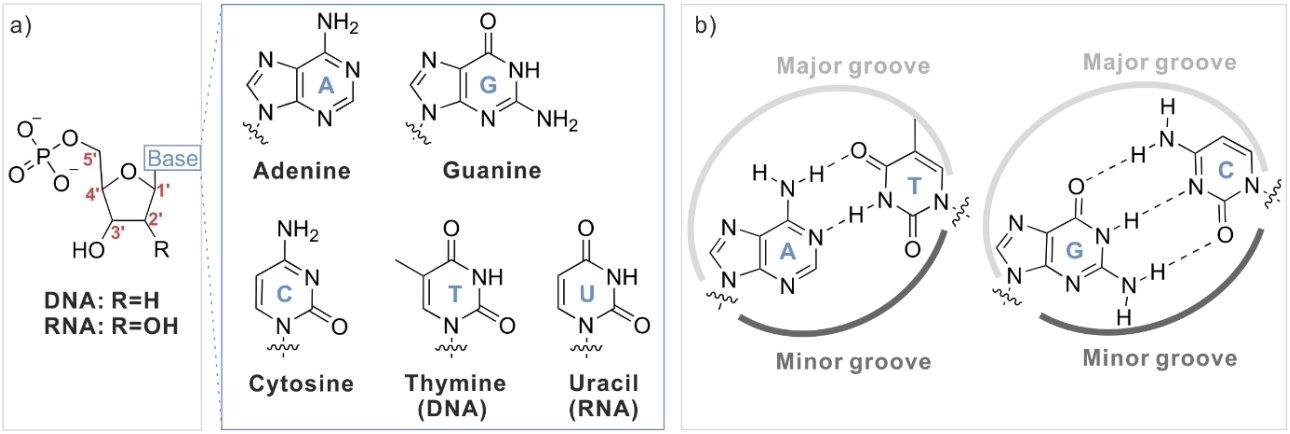


Figure 1. a) The monomer building blocks of nucleic acids. The site of the nitrogenous base attached to the sugar residue is shown in blue. b) Top view of an A-T base pair and a G-C base pair showing the formation of the major and minor groove sides of the DNA.

Nucleic acids fold under the following principles: the hydrophobic effect drives compactions, while buried donors and acceptors form hydrogen bonds (Figure 1b), van der Waals contacts are maximized, and the charged phosphate groups of the backbone are either solvated or neutralized. Based on the folding principle, the structure of nucleic acids can be divided into four different levels: primary, secondary, tertiary, and quaternary structure (Figure 2). Primary structure is the linear sequence of nucleotides; secondary structure involves interactions between bases; tertiary structure is the 3D folded shape of nucleic acid molecule; and quaternary structure refers to interactions with other molecules. DNA differs from RNA in that it mostly exists as fully base paired double helices. This helical structure creates specific features such as the major and minor grooves, which are key sites for protein recognition. In B-form duplexes, the major groove is wider than the minor groove and

both are readily accessible for protein recognition (Figure 1b, 2).[2] Moreover, DNA also exist in other conformations, such as A-DNA—a dehydrated, more compact form with a deeper and narrower major groove than B-DNA, and Z-DNA—a left-handed helical structure.[3]

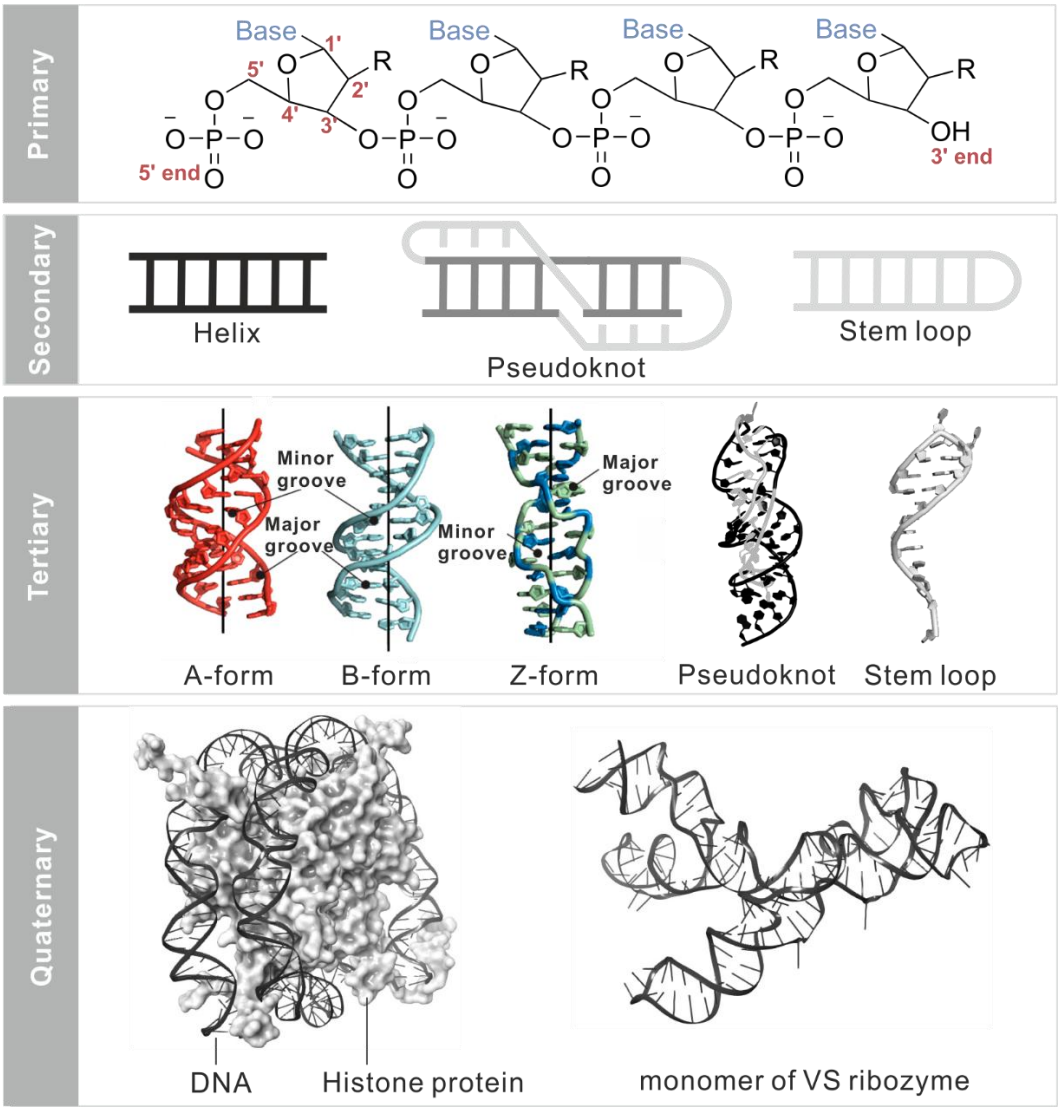


Figure 2. Nucleic acid structure at four levels: primary, secondary, tertiary,[4] and quaternary. Tertiary structure: A-form (PDB:413D), B-form (PDB:6CQ3), Z-form (PDB:4OCB), Pseudoknot (PDB: 1YMO), Stem loop (PDB: 4I8R); Quaternary structure: Nucleosome (PDB: 1EQZ), Varkud satellite (VS) ribozyme (PDB: 4R4V)

In contrast, RNA is often single-stranded but can form duplex, such as stem loop and pseudoknot.[5] These RNA duplex are largely limited to the A-form.[6] RNA also fold into complex tertiary structures, facilitated by hydrogen bonding involving its extra hydroxyl group on the ribose sugar, allowing it to serve not only as a messenger(mRNA) and but also as a catalyst (ribozyme) and regulator(siRNA). Thus their biological functions extend far beyond passive information carriers. They are also involved

in a wide range of cellular processes, including replication, transcription, translation, epigenetic regulation, and response to environmental stimuli.[7-9]

3.1.2 Synthetic nucleic acids analogues

Natural nucleic acids present several restrictions that limit their utility. For example, unmodified nucleic acids are rapidly degraded by nuclease before reaching their target receptors, making them less effective for long-term use in vivo.[10] Due to their reduced cell membrane permeability, they often require to integrate a delivery system that includes nanoparticles and liposomes.[11] A further drawback involves the lack of structural diversity, where the sugar-backbone constrains the functional derivatives for therapeutic purposes.

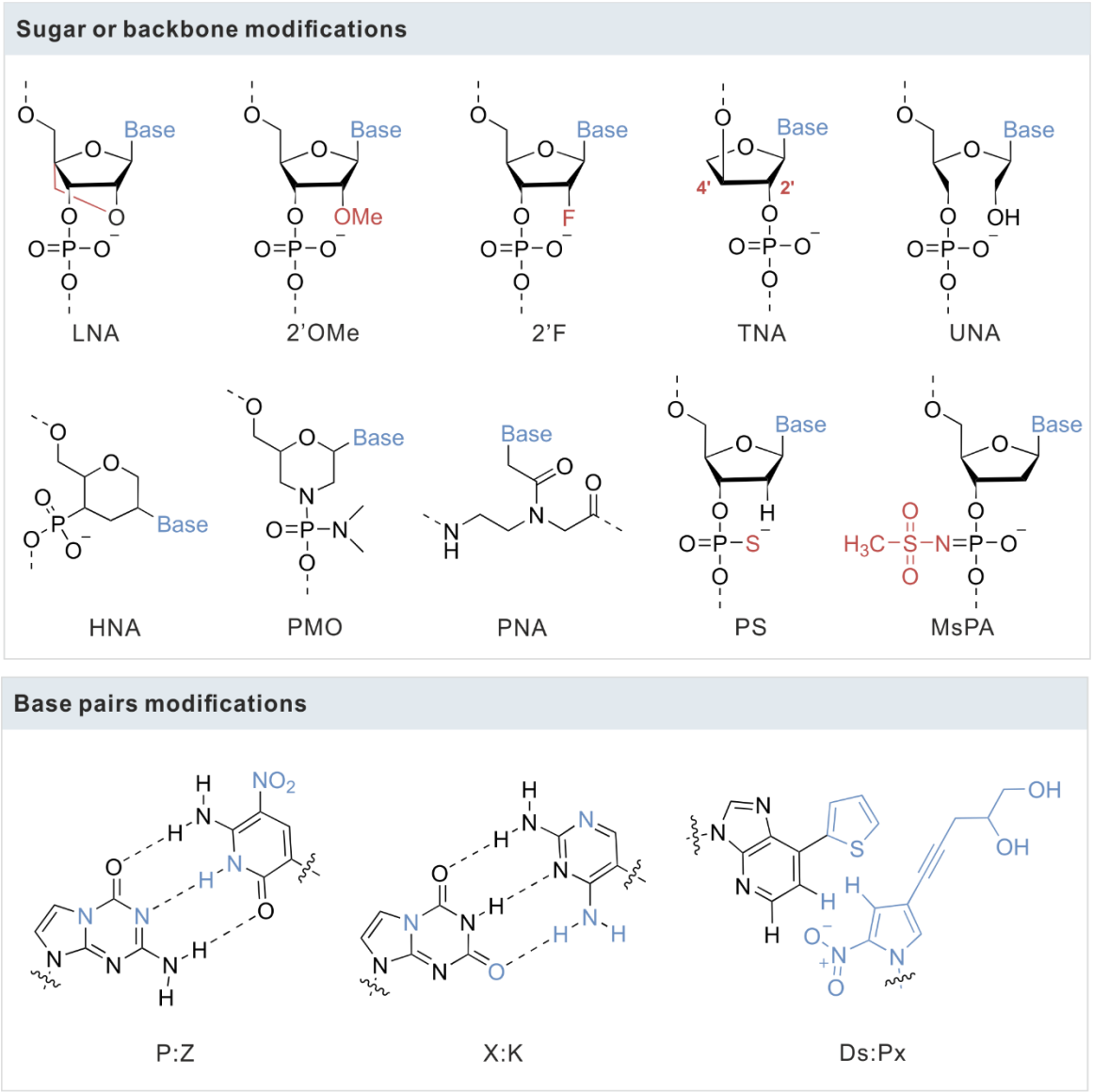


Figure 3. Nucleic acid modifications. LNA: locked nucleic acid; 2′OMe: 2′-O-methyl; 2′F: 2′-fluoro; TNA: threose nucleic acid; UNA: unlocked nucleic acid; HNA: hexitol nucleic acid; PMO:

morpholino phosphorodiamidate oligomer; PNA: peptide nucleic acid; PS: phosphorothioate; MsPA, mesyl phosphoramidate.

Scientists employ a series of design principle to create effective nucleic acid mimicry for medicinal applications. The central design principle involves structural modifications to backbone, sugar or nucleobases (Figure 3). Backbone modifications include functionalities and groups such as guanidinium,[12] amides, thioethers,[13] triazole,[14] boranophosphate,[15] methylphosphonate,[16] N-3′-phosphoramidate,[17] and S-methylthiourea.[18] These modifications are associated with enhanced physical and biological properties, significantly improving their performance in clinical applications. For instance, peptide nucleic acids (PNA) – neutral amide linkage confers resistance to nuclease degradation and strong binding affinity to DNA/RNA.[19-20] Another key modification replaces the ribose or deoxyribose sugar backbone with a six-membered morpholine ring, forming Phosphorodiamidate Morpholino Oligomers (PMOs), which have been widely used in antisense therapies with a proven safety profile.[21] In the case of Phosphorothioates (PS), one non-bridging oxygen is replaced with sulphur to improve nuclease resistance, resulting in better bioavailability of the oligonucleotides.[22]

Furthermore, sugar modification is another way to create nucleic acid mimicry. Locked nucleic acids (LNAs) retain the phosphodiester backbone but lock the furanose into a C3'-endo conformation; 2′-O-methyl (2′-OMe) and 2′-fluoro (2′-F) RNAs functionalize the sugar backbone, substantially enhancing rigidity and binding affinity.[23] Besides modifying either the sugar or backbone, nucleobase modifications can also be considered to increase the sequence diversity. Here are examples of artificial nucleobase with non-Watson-Crick base-pairing complementarity (Figure 3). Compared to the natural C:G base-pair, one hydrogen bond donor and acceptor are repositioned in xenobiotic P:Z and X:K nucleobase pairs. Also, spatial alignment of hydrophobic groups in Ds and Px bases induce complementary pairing.[24] Along with naturally occurring pyrimidine and purine scaffolds, other bases derived from pyrrole, imidazole, and indoles have also been developed.[25]

Based on these structural changes, synthetic nucleic acid mimics are designed to display improved hybridization behaviour, significantly increasing the binding affinity and mismatch discrimination of duplexes when binding to their complementary targets.[26-27] LNA, for example, show increased melting temperatures (T_m), making them ideal for applications requiring short, high-affinity probes.[28] PNAs also show strong hybridization to DNA and RNA due to their neutral backbone.[29] To note, nucleic acid mimics have distinct preference to A-form or B-form geometry. For example, Phosphorothioates DNA (PS-DNA) duplex retains B-form geometry and support RNase H activity,[30]

thereby have been widely used in therapeutic design; LNA strongly favour A-form conformation[28] while PNA duplex with DNA or RNA have intermediate geometry between A-form and B-form.[31]

3.1.3 Application of nucleic acids and their analogues

In the last three decades, oligonucleotides have been extensively investigated in processes such as polymerase chain reaction (PCR),[32] molecular probing,[33] artificial gene construction[34] and molecular cloning.[35] They have also been utilized to treat different diseases or are undergoing clinical trials (Figure 4).[36-39]

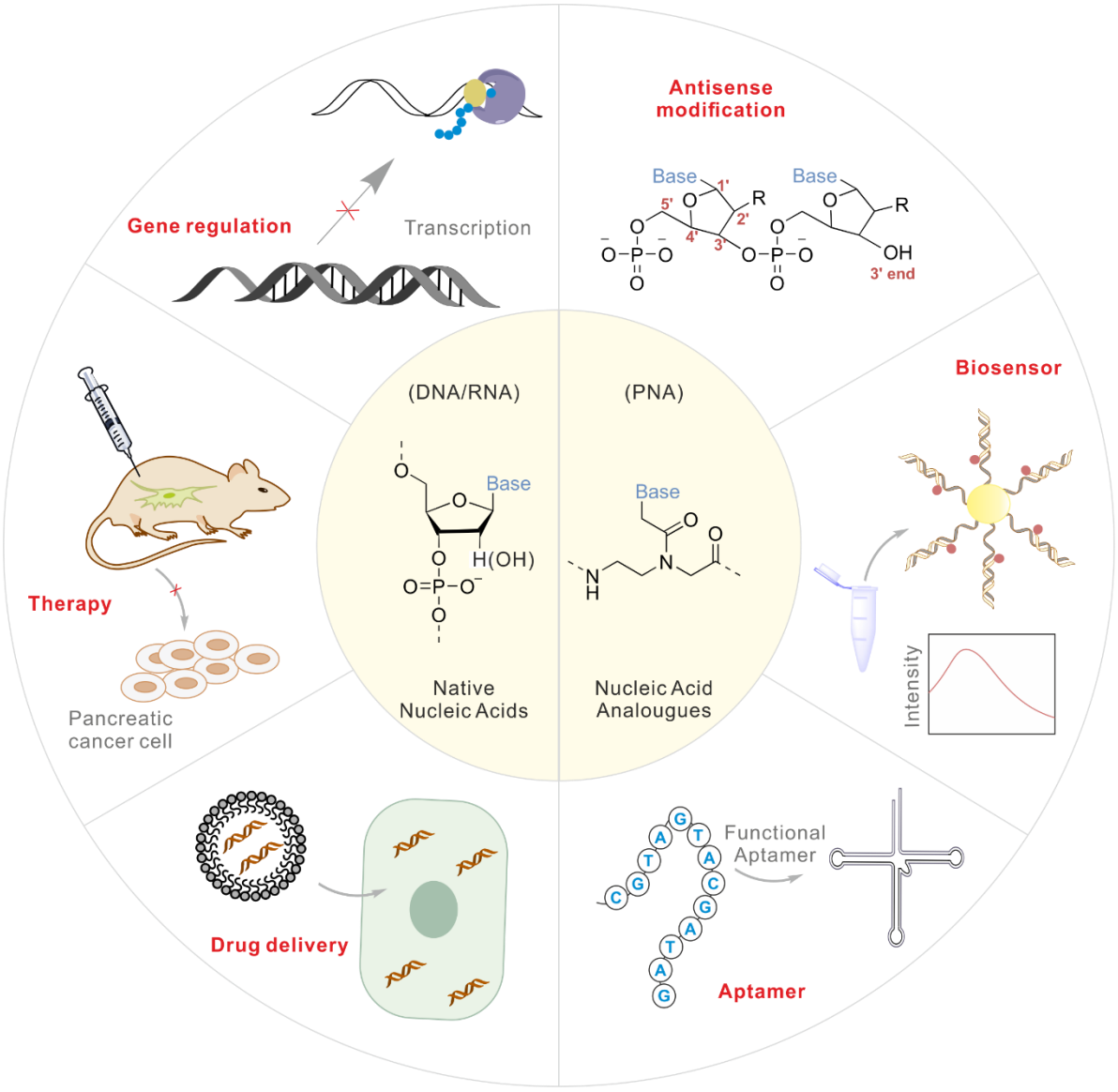


Figure 4. Schematic illustration of nucleic acid and their analogues for biomedical application.

Unlike natural nucleic acids, nucleic acid mimics are typically uncharged or structurally distinct and they often face significant challenges in cellular uptake due to their size, charges and hydrophilicity.

To overcome these barriers, multifunctional delivery strategies have been developed, including conjugation with cell-penetrating peptides (CPPs),[40] 2,4-diaminobutanoic acid (Dab) dendrons;[41] encapsulation into lipid nanoparticles (LNPs) or liposomes;[42-43] and the use of engineered or biomimetic nanocarriers.[44-45] In fact, these systems that enable nucleic acid mimics not only to cross the cellular membrane but also to manipulate intracellular trafficking, are now available.[46]

3.1.4 Natural DNA mimic proteins

Beyond synthetic nucleic acids mimicry that hybridize with DNA or RNA, a fundamentally different strategy consist of emerged-proteins that mimic DNA itself (Figure 5).[47] These proteins resemble negative surface charge distributions of B-DNA. Instead of binding to nucleic acids, DNA mimic proteins occupy the DNA-binding domains to regulate biological activities such as DNA repair, gene expression and DNA package in the living cells.[48] So far, only a small amount of DNA mimic proteins have been studied, and they have been found in prokaryotes and eukaryotes as well as in bacteriophages and in a eukaryotic virus. For example, TAFII230 (transcription initiation factors),[49] HI1450 (highly acidic protein from Haemophilus influenzae),[50] UGI (phage uracil DNA glycosylase inhibitor)[51] and Ocr (overcome classical restriction)[52] and p56.[53] Inspired by these natural systems, researchers have developed de novo designed DNA mimic proteins, engineered to mimic the charge pattern of DNA and to inhibit restriction activity of a type I restriction enzyme.[54] This work validates the use of charge patterning as a design principle to create protein mimics of DNA, and serves as a starting point to develop therapeutic peptide inhibitors against human pathogens.

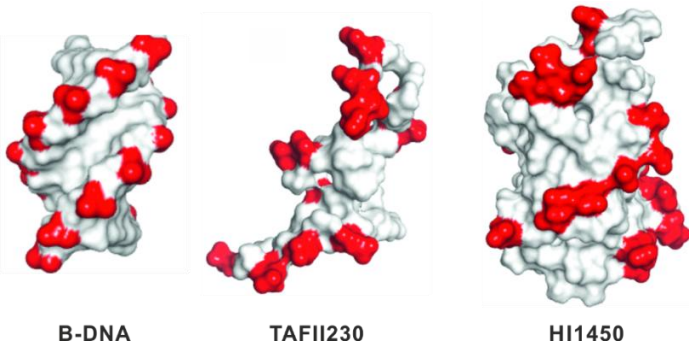


Figure 5. Negative charge distribution (in red) of B-DNA and DNA mimic proteins: TAFII230 (PDB: 1TBA); HI1450 (PDB: 1NNV).

3.2. Protein-nucleic acid interactions (PNIs)

3.2.1 Biological significance of protein-nucleic acid interactions

Proteins, found in every cell, are biological macromolecules made up of amino acids that are linked by peptide bonds. They serve as molecular "tools" and have very specific functions in the body. For

example: Collagens, as structural proteins, determine the nature of the skin, connective tissue and bones, making up to $^{1}/_{3}$ of the total body protein; in the muscles, myosins and actins change their shape and enable muscle contraction and thus movement; in the saliva, amylase start to digest the starchy food the moment they enter your mouth; as ion channels, proteins are present in neurons and heart cells, regulating heartbeat and nerve signal. Rather than functioning in isolation, these proteins operate in close coordination with nucleic acids, and mutations in a particular gene can potentially cause the dysfunction of the corresponding protein and thereby the development of diseases.

The principle behind is known as the central dogma of life,[55] which describes the generic flow within a biological system. It states that DNA is transcribed into RNA, which is then translated into proteins that carry out cellular functions. However, this process is not a simple and linear cascade. It is the result of the constant interaction between functional nucleic acids and specific proteins in life's process.

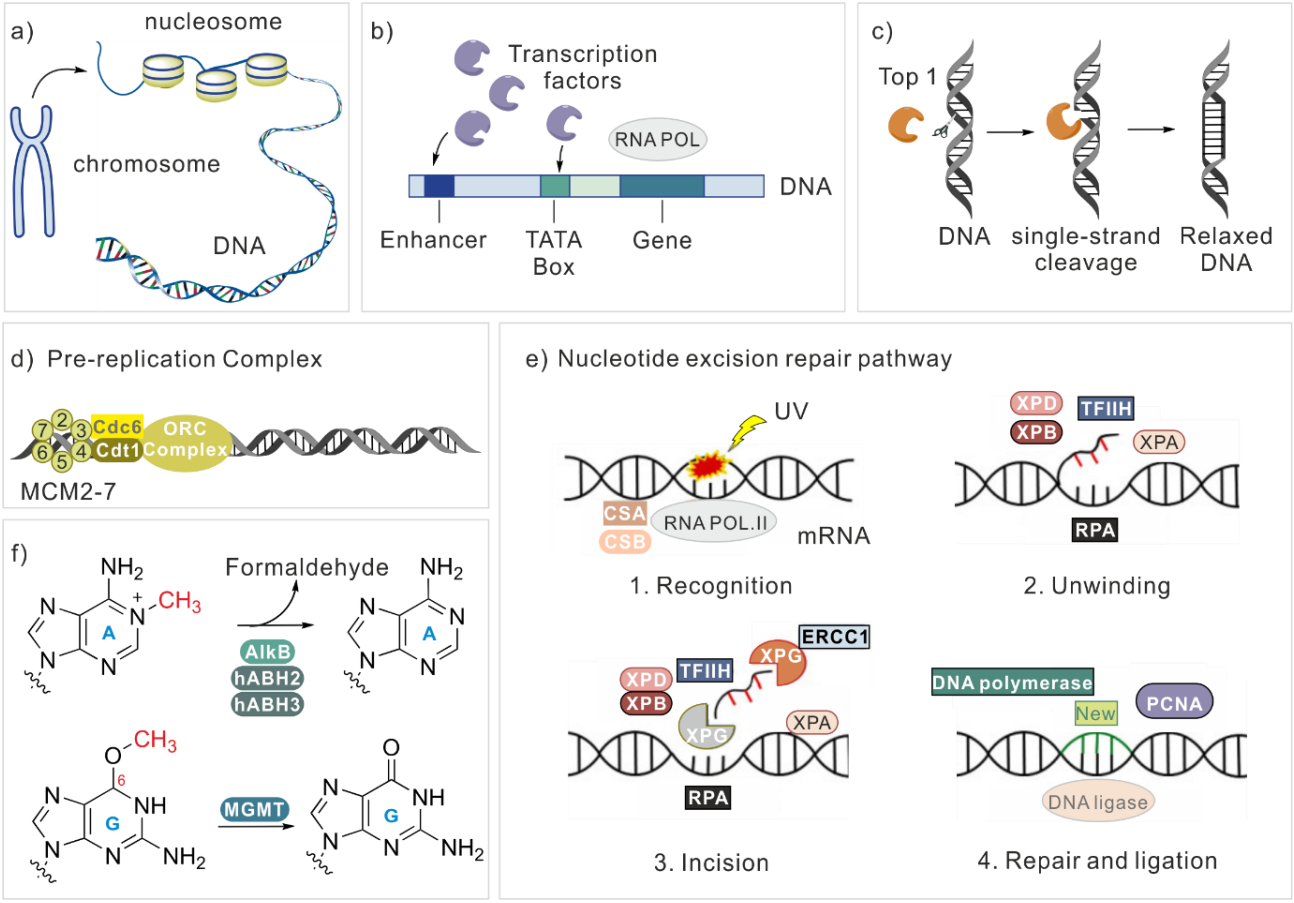


Figure 6. Examples of PNIs in biological processes a) A chromosome and its packaged long strand of DNA unravelled. b) Regulation of gene expression in eukaryotes. RNA POL: RNA polymerase. c) Function of DNA topoisomerase 1. d) Pre-replication complex in eukaryotes. MCM2-7: minichromosome maintenance protein 2-7; Cdc6: cell division cycle 6; Cdt1: chromatin licensing

and DNA replication factor 1. e) Nucleotide excision repair (NER) pathway. CSA/CSB: Cockayne syndrome A and B proteins, RPA: replication protein A; ERCC-1: excision repair cross-complementation group 1; TFIIH: transcription factor II H; PCNA: proliferating cell nuclear antigen. f) Repair of DNA alkylation lesions. AlkB: alkylation B; hABH: human alpha/beta hydrolase; MGMT: methylated-DNA-protein-cysteine methyltransferase.

During cell division, especially in eukaryotic cells, protein-DNA interactions play critical roles in ensuring that the DNA is accurately replicated, packaged, and distributed to daughter cells. A major function of protein-DNA interactions is chromosome packaging, where the extensive length of the DNA (approximately 2 meters in each human cell) wraps around histone proteins to form nucleosomes (Figure 6a), allowing for efficient organization of the genetic materials and access for other cellular machinery.[56] Once unravelled, genomic DNA can be transcribed; however, not all of the DNA sequence codes for proteins. Only genes are transcribed to produce RNA, and the sequences between the genes (and within), such as promoters, enhancers, insulators and spacers, are important for transcriptional control. Transcription factor proteins bind to specific DNA promoter sequences located immediately adjacent to the gene transcription start site (Figure 6b). As a result, a messenger RNA (mRNA) is produced by controlling the RNA polymerase recruitment. This precise regulation ensures the expression of genes at the correct time and location.

In addition, DNA replication and DNA repair mechanisms also involve proteins interacting with nucleic acids. In higher eukaryotes, the origin recognition complex (ORC) is present throughout the cell cycle bound to the origin of replication, but is active only in late mitosis and early G1 phase (Figure 6d). Beyond DNA replication, the ORC complex also take part in RNA export and in setting up a defined chromatin structure around the replication origins.[57-59] Another protein that have gained increasing attention in recent year are DNA topoisomerases. These enzymes can remove DNA supercoils during transcription and DNA replication, by catalysing transient single (type I topoisomerases) or double strand breaks (type II topoisomerases), crossing the strands through one another, then resealing the breaks without introducing mutations (Figure 6c). If the DNA supercoils are left unsolved, this torsion would eventually stop the DNA or RNA polymerases from continuing along the DNA helix.

Preserving genomic sequence information in living organisms is necessary for life and continuity of the species. At the same time, mutagenesis plays a role in its maintenance and evolution, while also impacting human health and contributing to aging. To our relief, cells have evolved with sophisticated systems, for example, DNA repair, damage tolerance, cell death pathways, that collectively function

to increase genomic stability, thereby reduce the deleterious implications of DNA damage. In our daily life, Sunlight exposure is usually the trigger for UV-induced DNA damage. Laboratory studies have shown that UV rays cause thymine dimers which are covalently linked between two adjacent pyrimidines.[60] Proteins involved in nucleotide excision repair (NER) pathway can recognise the bulky DNA segment and replace it with the correct sequences using DNA polymerase and ligase (Figure 6e). Deficiencies in certain proteins can lead to disease. For example, XPA, XPB, XPC, XPD, XPE, XPF, and XPG are associated with xeroderma pigmentosum (XP). On the other hand, cigarette smoke, industrial processing and certain drugs from chemotherapy can produce exogenous chemical agents leading to alkylated DNA damage. There are two different classes of enzymes in humans and mammals that repair the alkylated bases (Figure 6f). First, the O^6-alkylguanine-DNA methyltransferase (MGMT) reverses O-alkylated DNA lesions by transferring the alkyl group from the oxygen of the DNA base to the cysteine residue in its catalytic pocket;[61] The second class of direct reversal enzymes, the AlkB-related α-ketoglutarate-dependent dioxygenases (AlkB), reverse N-alkylated base adducts.[62] The oxidized alkyl group is released as formaldehyde, leaving behind the original base.

3.2.2. Molecular recognition in protein-nucleic acid complexes

To better understand the dynamic protein-DNA interactions in living cells, we now turn to the molecular mechanism of DNA recognition by proteins. In general, proteins interact with DNA through diverse physical forces. Among them, hydrogen-bonding interactions are a key component of protein-nucleic acid recognition, mediated by donor and acceptor groups from the bases of DNA and the side chains of most of the polar amino acids (Figure 7a-7b). Salt bridges are also commonly observed in protein-nucleic acid complexes, they are electrostatic interactions between groups of opposite charge (Figure 7c). Hydrophobic effect refers to the tendency of nonpolar molecules to aggregate in aqueous solutions, minimizing their contact with water. For example, the non-polar parts of nucleic acid, including the 5-methyl group of thymine, the heterocyclic carbon atoms within the purine and pyrimidine rings, and the ribose carbon atoms, are often contacted by aliphatic side chains of amino acid in protein-nucleic acid complexes (Figure 7d).[63] Additionally, dispersions attract can further stabilize protein-DNA complexes (Figure 7e). Stacking interactions can also be observed between aromatic residues in a protein and unstacked nucleotides (Figure 7f).

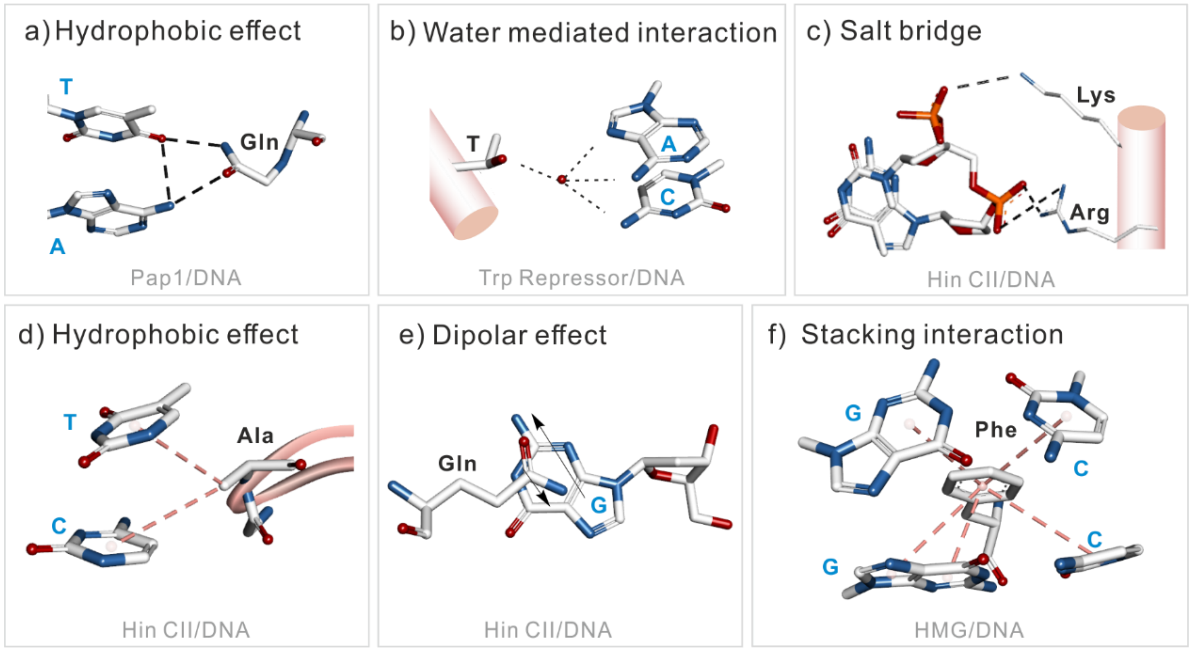


Figure 7. Physical forces between protein side chains and nucleic acids. a) Hydrogen-bonding interactions of glutamine with T and A, respectively (PDB: 1GD2). Hydrogen-bonding interactions are indicated by the dashed lines. b) Water molecules (red) provide a non-covalent extension of the surface of the DNA-binding protein (PDB: 1TRO). c) Salt bridge between positively charged protein side chains and negatively charged DNA phosphate groups. (PDB: 1OCT). d) Hydrophobic effect: recognition from contacts of non-polar atoms (PDB: 1KC6). e) Dipolar interaction (PDB: 1KC6). Dipoles are indicated by the black arrows. f) Stacking interactions between a Phe side chain and unstacked bases (PDB: 1CKT). Only the phenylalanine side chain of the protein is shown for clarity.

To elaborate further, DNA binds to proteins in a sequence-specific or non-sequence-specific manner. The specific protein-DNA interactions occur through direct contact, also known as base readout. EcoRI, a widely used type II restriction endonucleases in bacteria, can recognizes the palindromic sequence 5'-GAATTC-3', while prevent the cleavage of the host DNA.[64-65] Another restriction endonuclease BamHI recognizes its DNA binding sites, AGATCT and GGATCC, with an identical core region (underlined), but bending differentiates both binding sites.[65]

In addition to reading the base pairs, many proteins perform indirect readout by sensing local DNA shape features (eg: minor groove width, helical twist between base pairs and DNA flexibility…), or global shape features when the binding site is not in a classic B-form helix (Figure 8). To achieve high affinity and specificity, all DNA-binding protein, especially those that are sequence-specific, recognize DNA using a combination of base readout and shape readout.[66] For example, *E. coli* nucleoid protein Integration host factor (IHF) achieve DNA binding affinity by a combination of

bending, kinking and intercalation. Specifically, the IHF a/b heterodimer sharply bends DNA by 160° and also recognize three DNA sites: a TTG region at its flank, TATCAA in the central region of its binding site, and a 6-bp A tract.[65, 67] To note, there is no corresponding relationship between amino acids and DNA bases, but some particular amino acid-base pairings are enriched, such as arginine with guanine, and asparagine and glutamine with adenine.[68-70]

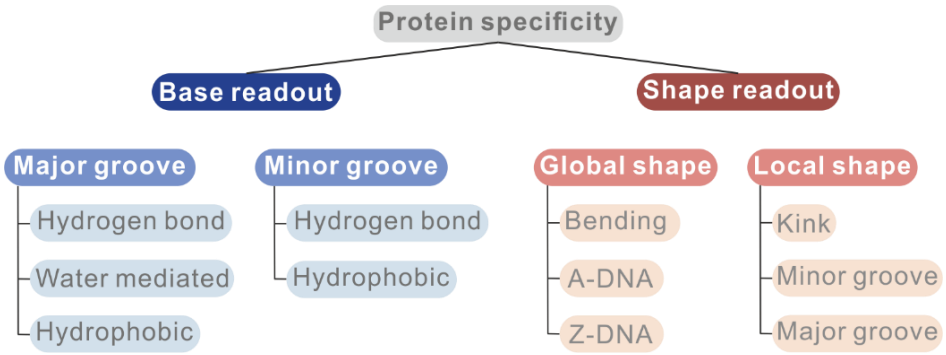


Figure 8. Types of protein-DNA recognition mechanisms used for specificity.[65]

In contrast, some DNA binding proteins bind to DNA non-specifically. Their binding is primarily driven by general physical and chemical properties of DNA rather than a specific nucleotide sequence. For example, histones bind to DNA in a non-specific manner through strong electrostatic interactions between their positively charged amino acids (lysine and arginine) and the negatively charged phosphate backbone of the DNA. Similarly, high mobility group (HMG) proteins, which are the most abundant non-histone chromosomal proteins, bind to DNA in a non-sequence-specific manner, promoting gene regulation and chromatin function.[71] Some structural aspects of non-sequence-specific DNA recognition have been examined: high mobility group (HMG)[72], Sac7d[73], a chromosomal protein from an archaeal hyperthermophile that kinks DNA by 61°, and the nucleosome itself.[74] Although these proteins do not ‘‘read’’ the DNA sequence, they do recognize other structural features of DNA, contributing to DNA architecture and essential cellular processes, such as compaction,[75] repair[76] and chromatin remodelling.[77]

Both specific and non-specific binding modes affect gene expression: the former provide specific gene regulations via direct base recognition; the latter enables global DNA organization and influences accessibility of transcriptional machinery[78-80]. Altogether, deciphering the protein-DNA recognition codes can not only help us better understand the mechanisms of these specific binding events, but is also crucial to reveal diseases caused by mutations that affect protein-DNA binding specificity, and to design therapeutic drugs.

3.2.3. Strategies to modulate protein-nucleic acid interactions

Transcription factors are found in all eukaryotes and regulate gene transcription in cell growth,[81] stress response,[82] immune responses.[83] Together with some other DNA-binding proteins, they were considered as undruggable due to their lack of well-defined ligand-binding pockets. Yet, recent successful examples have demonstrated that small molecules can indeed interfere with protein-DNA interactions. This chemical approach spans from groove binders to the inhibitors of the DNA-binding domain in target proteins. Compounds like netropsin[84], distamycin A[85], and microgonotropens (MGTs)[86] bind specifically in the A/T-rich minor groove (Figure 9a), blocking transcription factors from accessing their DNA sites. In addition to targeting DNA itself, small molecules also bind directly to the transcription factors. Key examples include factor quinolinone inhibitor 1 (FQI1)[87] and NSC194598 (Figure 9a) [88].

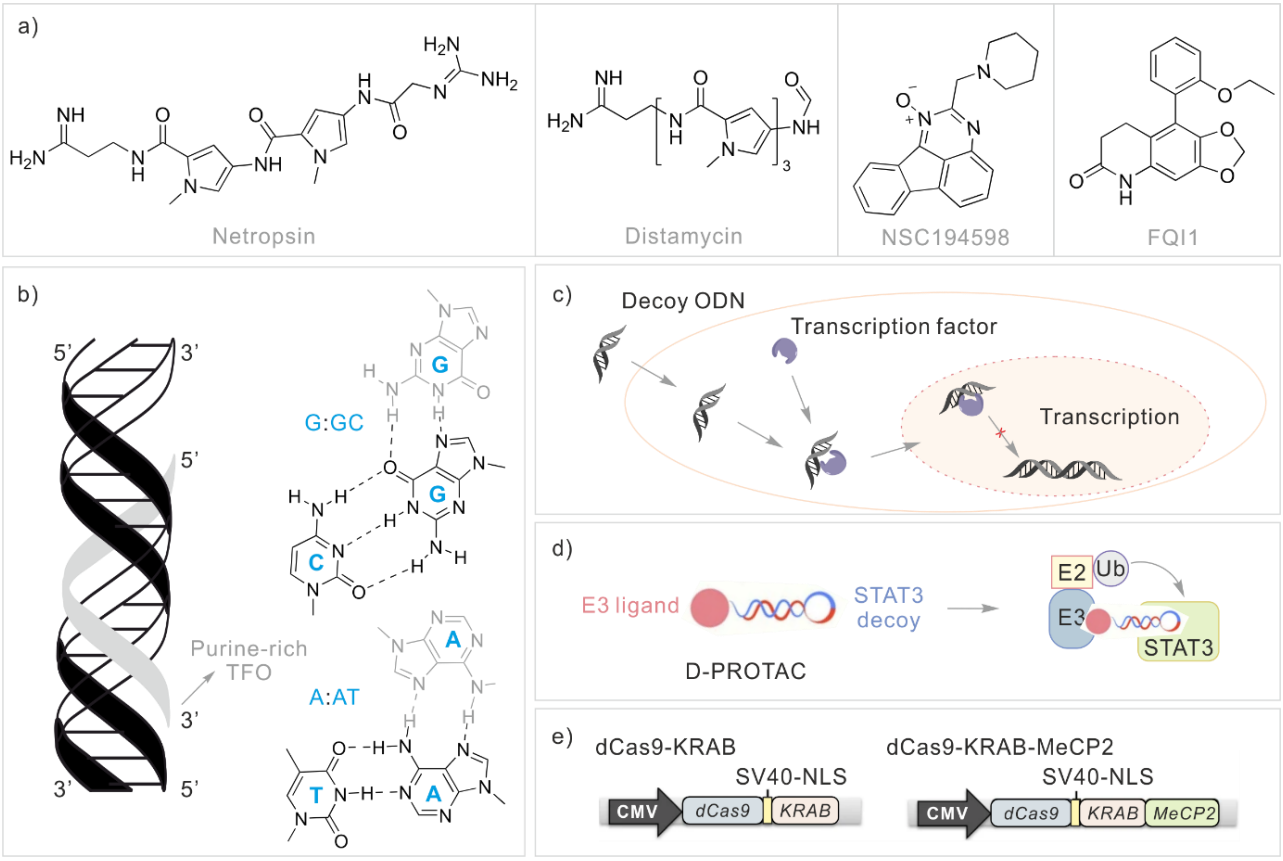


Figure 9. Strategies to interfere with protein-DNA interactions. a) Small molecules that bind to minor grooves of DNA.[84-85, 87-88] b) Triplex binding code in the purine-rich TFO motifs.[89] c) Schematic of gene expression inhibition by decoy oligodeoxynucleotides (decoy ODNs).[90] d) Decoy-PROTAC for specific degradation of "undruggable" STAT3 transcription factor.[91] e) Schematic of dCas9-KRAB and dCas9-KRAB-MeCP2 repressors. KRAB: a Krüppel-associated box transcriptional repression domain; MeCP2: methyl-CpG binding protein 2 NLS: nuclear localization signal.[92]

Beyond small molecules, triplex-forming oligonucleotides (TFOs) offer complementary strategy to disrupt protein-DNA interactions (Figure 9b) [89]. These agents bind specifically to genomic DNA through Hoogsteen hydrogen interactions, creating non-B-DNA structures, physically blocking transcription factor access[93-94]. Decoy oligodeoxynucleotides (ODNs) are short, double-stranded DNA sequences, and they can be transfected into the target cell, bind to the respective transcription factor and inhibit the transcription of target gene (Figure 9c).[90] Moving beyond mere inhibition, oligonucleotide-based PROTACs (proteolysis targeting chimeras) go a step further by inducing targeted degradation of DNA-binding proteins (Figure 9d). These bifunctional molecules link DNA/RNA binders with a ligand for ubiquitin ligase and drive the proteasomal degradation of transcription factors.[91] In addition to chemical and nucleic acid-based approaches, recent advances in synthetic biology have led to the development of engineered DNA-binding proteins—such as zinc finger repressors (ZFP),[95] transcription activator-like effector repressors (TALE-based repressors),[96] and CRISPR/dCas9-KRAB systems[97]. They illustrate how modular DNA-binding scaffolds (ZFP, TALE, dCas9) fused to repressor domains (KRAB, MeCP2, etc.)[92] can reprogram gene expression for both research and potential therapeutic applications (Figure 9e).

3.3. Aromatic foldamers as DNA surface mimicry

Folding is the process nature has selected for biopolymers, for example, polysaccharides, proteins, nucleic acids, lipids, etc, to carry out complex biological tasks such as enzyme catalysis, information storage and energy conversion. Inspired by the structures and functions of biopolymers, researchers have shown that folding is not exclusive to natural polymers, but can also occur in a wide range of non-natural backbones. These artificially folded molecular architectures are known as “foldamers.” This matters because backbones that differ from those of proteins and nucleic acids may give access to functions beyond the capabilities of natural biopolymers.

Synthetic foldamers are far too numerous to be simply mentioned here.[98] There are two general molecular classes of foldamers, determined by the presence or absence of aromatic units within the monomer unit (Figure 10). The design of “aliphatic” foldamers, also called biotic foldamers, has been inspired by their biopolymer counterparts, and they share similar folding principles. Typical examples are peptide nucleic acids (PNAs) and *N*-substituted oligoglycines (peptoids). This triggered a comprehensive investigations of β-, γ- and δ-peptides, oligoureas and azapeptides.[99] In contrast, “aromatic” foldamers have emerged with backbones and folding modes that differ significantly from those of biopolymers. The poly-pyrrole/imidazole DNA-binding oligomers provided early examples of heteroaryl oligomers that bind minor groove of DNA sequence-specifically.[100] To achieve more

distinct properties and therefore functions, the aromatic foldamers have been expanded into more diverse, abiotic backbone types. In particular, the linkage between units could be one such criteria: amide based on aza-heterocycles (pyridines, pyrimidines, pyridazines, etc.);[101] hydrazide;[102] urea[103] or alkyne groups.[104] Additionally, foldamers may be a composite of two or more different building blocks, which can be arranged in an alternating manner.[105-106]

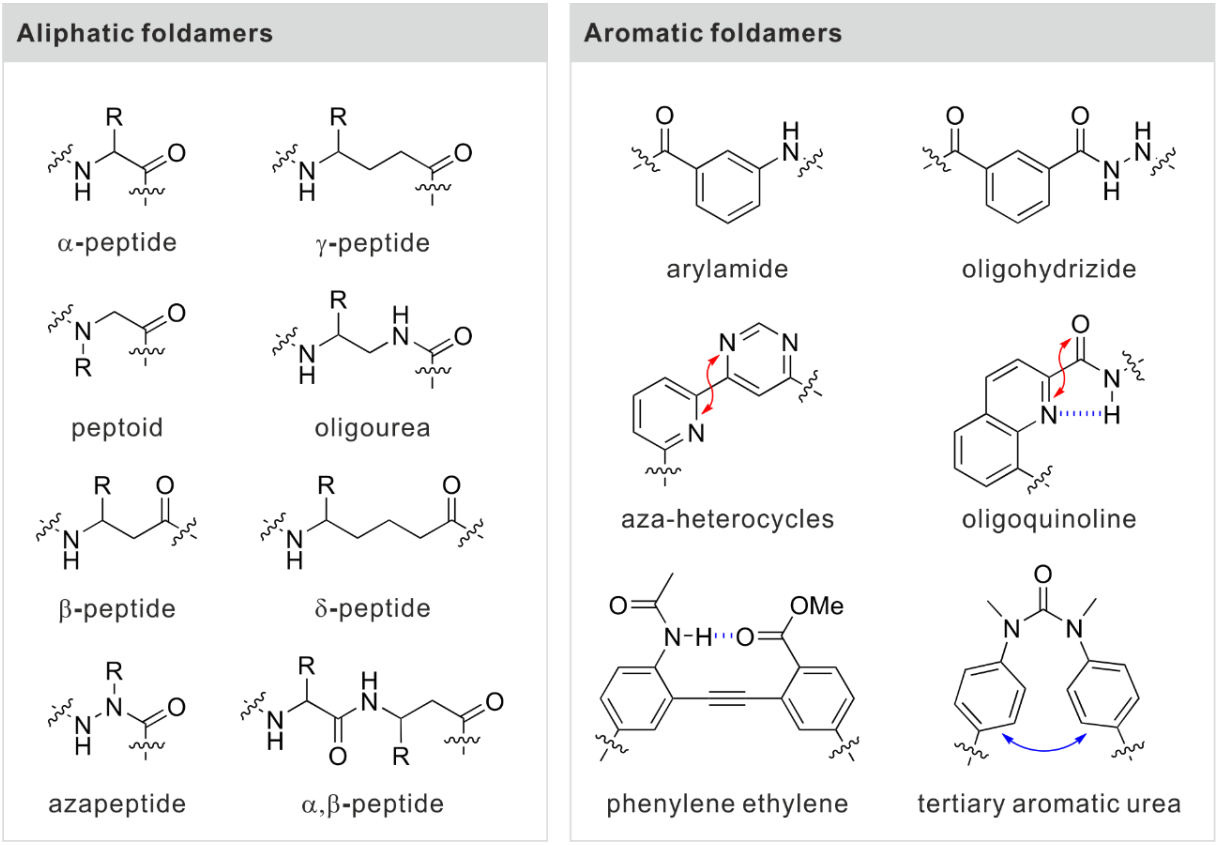


Figure 10. Examples of foldamer backbones. Red and blue arrows represent intramolecular repulsive and attractive interactions, respectively.

Despite the large number of building blocks reported so far, there are not a wide variety of different folding patterns. The common motif in biopolymers especially helices still prevail in non-natural oligomers (Figure 11a), even though more and more examples of β-sheets are reported in the literature. Additionally, Otto and coworkers showed folded oligomers containing only an aromatic disulfide monomers, which is neither α-helix nor β-sheets (Figure 11b).[107] Like the secondary structures of peptides, the construction of the artificial foldamers requires consideration of localized noncovalent interactions such as hydrogen bonding and electrostatic repulsions, but π-π stacking and specific geometric constrains also play significant roles in the case of aromatic foldamers (Figure 11c-11d).

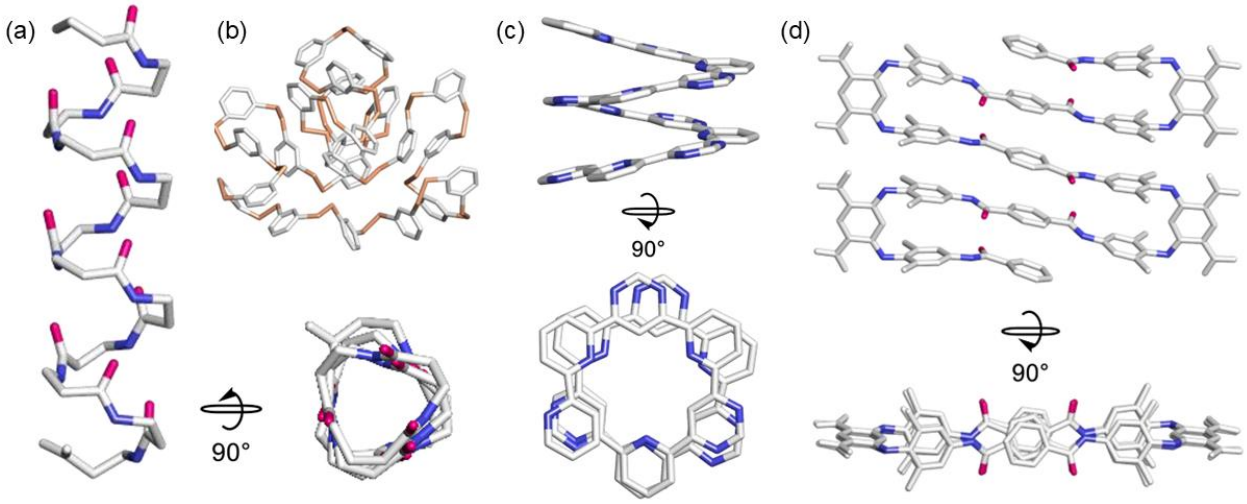


Figure 11. Folding patterns of different oligomers. a) β-peptide 12-mer;[108] b) aryl disulfide macrocycle;[109] c) helical aryl oligomer;[110] d) sheet-forming aromatic oligoamide.[111] Disulfide bonds are highlighted in orange. Oxygen and nitrogen atoms are highlighted in purple and blue, respectively. In all structures hydrogen atoms and side chains are omitted for clarity. Figure adapted from figure 2.1 from Bindl D.[112]

Among all the synthetic foldamers mentioned above, aromatic foldamers have been rapidly growing as an important class of foldamers due to a number of remarkable properties, including the stable conformation, predictable folding modes, their propensity to crystallize, and accessible synthesis. These advantages of aromatic foldamers have opened up avenue towards the design of α-mimicry,[113] coordination of metal ions,[114] molecular motions,[115] and self-assembly, etc.[116] Among the significant advances in foldamer science, α-helix mimicry stands out as a successful strategy for protein epitope mimicry, and may thus serve as an inspiration for nucleic acids surface mimicry.

Nucleic acid mimicry designed to reproduce their Watson-Crick base-pairing abilities, can result in improved behaviour (See section 3.1). Together with the DNA mimic protein mentioned above, this background exemplifies well the importance and potential for applications of nucleic acid mimicry. However, limitations remain. For example: DNA-binding proteins do not recognize well PNA-DNA duplexes.[117] DNA mimic proteins exist in nature and are difficult to predict due to their various sequence and structure.[47] Sulfonated polysaccharides, such as heparin, can inhibit multiple DNA-binding proteins due to their polyanionic nature,[118] but they do so without selectivity. Furthermore, their chemical nature makes the synthesis of variants difficult, and the conformation of heparin derivatives is not well-defined. Given all of these challenges associated with nucleic acid mimicry,

there is a growing interest in designing nucleic acids surface mimicry that go beyond the conventional backbone chemistry.

We recently found that alternation in the same sequence of δ-amino acid monomer **Q** with ε-amino acid monomer **^{m}Q**, which possesses an additional methylene group, can produce a single helix whose geometrical parameters match those of double stranded B-DNA. The reason for this match comes from the contributions to helix curvature of **Q** (0.4 turn/unit) and **^{m}Q** (0.5 turn per unit) which sum up to 0.9 turn per **^{m}QQ** dimer. Thus, in an (**^{m}QQ**)$_n$ helix, the angular shift between adjacent **^{m}QQ** blocks is predicted to be a tenth of a turn, which is equal to the angular shift between base pairs in double-stranded B-DNA (ten base pairs span a full double helix turn). An **^{m}QQ** dimer also contributes a vertical rise of 3.5 Å (the thickness of one aromatic ring) along the helix axis that matches the base-pair distances in B-DNA.[119]

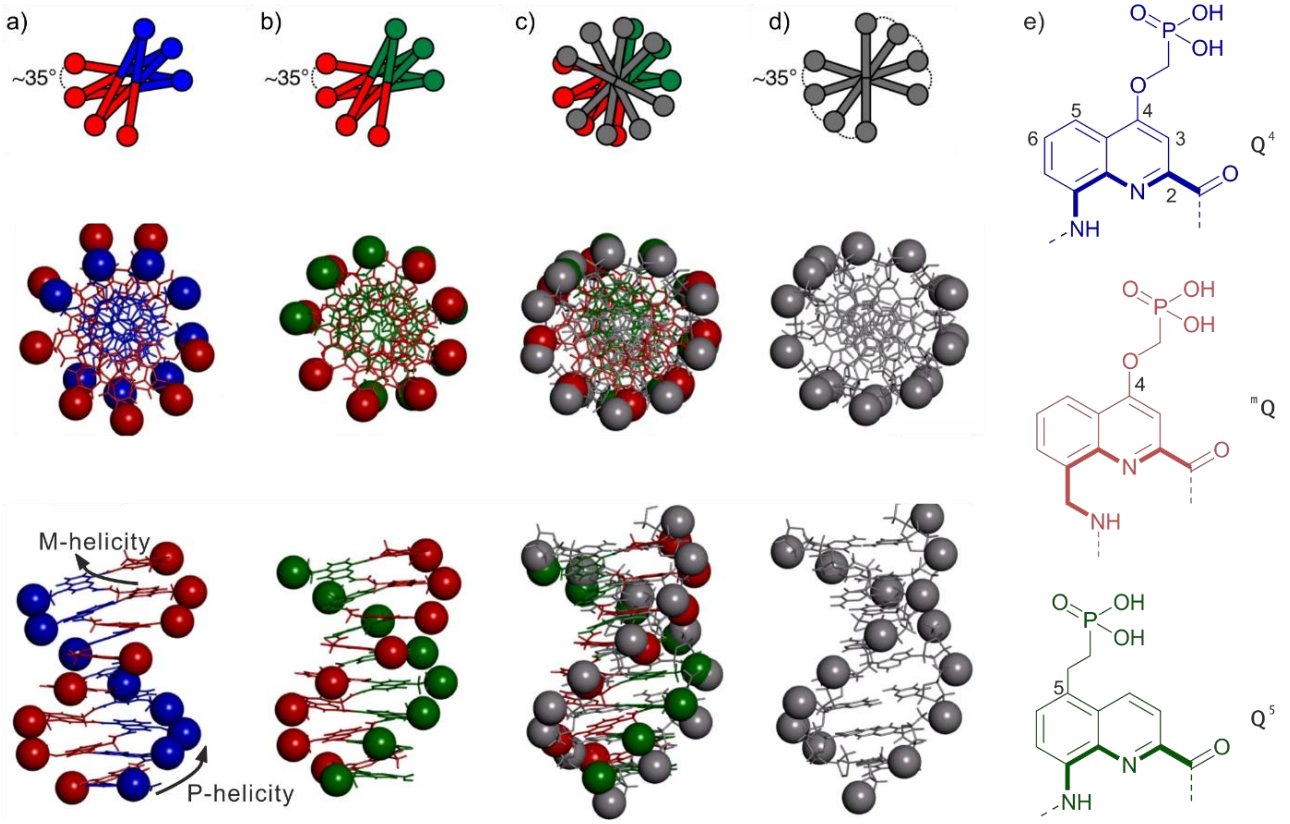


Figure 12. DNA mimic design.[119] Schematic representations, top views and side views of molecular models of a) (**mQQ^4**)$_8$; b) (**mQQ^5**)$_8$; and d) an eight-base-pair B-DNA. The structure in c) is an overlay of b) and c). (e) Formulae of amino-acid monomers **Q^4**, **^{m}Q** and **Q^5**. Models are shown at the same scale as the stick representations except that the phosphorus atoms are shown as spheres. The same blue, red and green colours are used in the models and schemes as in the formulae shown in e).

In order to mimic the B-DNA charge surface, phosphonate side chains were installed to ^{m}**Q** and **Q**. As a result, the negatively charged side chain at the surface of an $(^{m}\mathbf{QQ})_n$ single helix form a double helical array that matches the positions of phosphates in duplex B-DNA (Figure 12).[119] However. The grooves of the $(^{m}\mathbf{QQ^4})_n$ helix are very similar in width. To widen the major groove and narrow down the minor groove without changing the curvature contribution, the side chains of the Q monomers was installed at position 5 as in monomer $\mathbf{Q^5}$. Interestingly, charge distribution of the $(^{m}\mathbf{QQ^5})_n$ oligomer predicted by molecular model shows an excellent match with the positions of phosphorus atoms in duplex B-DNA.

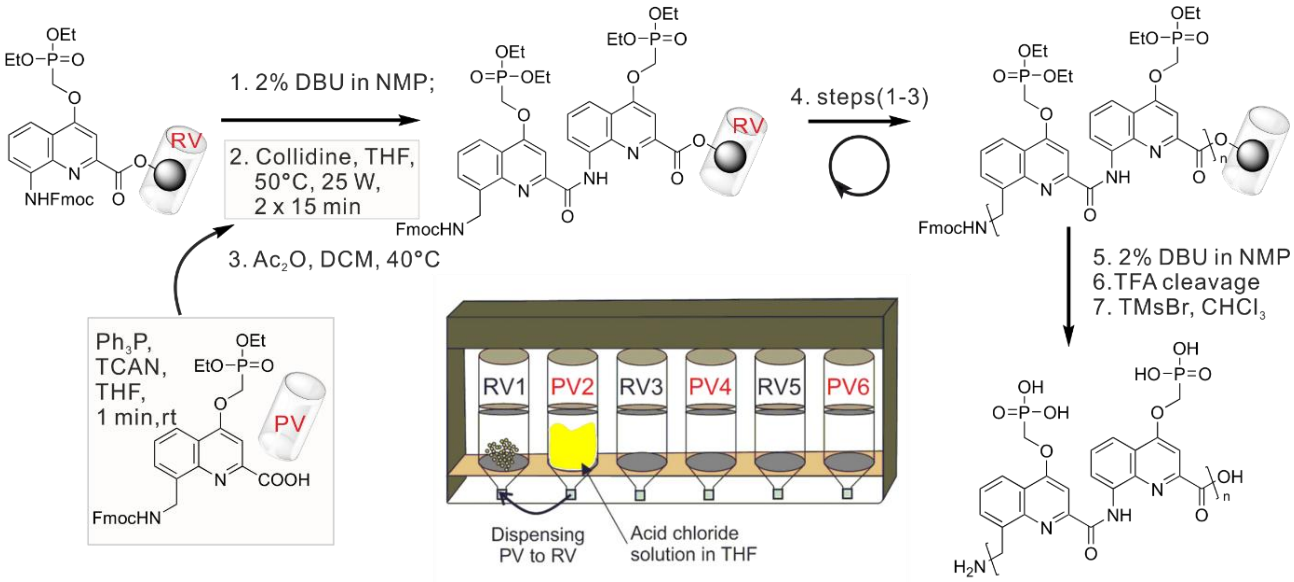


Figure 13. Schematic of the PurePep Chorus synthesizer with reaction vessels (RV) and pre-active vessels (PV) used during the coupling cycles.[120]

The synthesis of DNA-mimic foldamers was initially achieved in solution with low efficiency, which requires multiple purification after each coupling steps. Taking inspiration from the oligomer solid-phase synthetic methods, which has extended beyond peptides and nucleotides, we managed to implement the solid phase synthesis (SPS) protocols on a commercial peptide synthesizer, in particular using *in situ* acid chloride activation under Appel's conditions. The automated solid phase foldamer synthesis (SPFS) contains cycles iterations similar to those of manual SPFS: 1) Fmoc deprotection, 2) resin washings, 3) coupling, and 4) new round of washings (Figure 13). To note, all the building blocks needed by the DNA-mimic foldamers have been synthesized as Fmoc-protected amino acids. Starting from the Fmoc-monomer loaded resin, Fmoc group is removed using 2% DBU in NMP to obtain a free amine on the solid support. When coupling to the aromatic amine, the conversion of the acid into the acid chloride is implemented in the pre-active vessel (PV) *in situ*

condition. Then, shaking the acid chloride solution in PV for 1 min before dispensing to the reaction vessel (RV) containing the pre-swollen resin with a solution of 2,4,6-collidine. To note, aggregation on the resin is a common problem in peptide synthesis, which hampers further couplings. Moreover, automation of SPFS plus the possibility to synthesize several sequences in parallel significantly reduce the work time, offering a rapid access to new and longer sequences, and thus to quicker exploration of their properties.

These single stranded foldamers, featuring aryl rings in their main chain, differ remarkably from B-DNA and thus hold the potential to go beyond the function of natural nucleic acids. Alone with this line, we assessed the inhibition of DNA-mimic foldamers on several DNA-binding enzymes. To our surprise, two therapeutically important enzymes were remarkably inhibited by $(^{m}\mathbf{QQ^4})_n$. Specifically, topoisomerase 1 (Top1), which cleaves one strand of DNA duplexes to allow relaxation of supercoiling, was strongly inhibited by $(^{m}\mathbf{QQ^4})_n$ in a length dependent manner. Human immune deficiency virus 1 integrase (HIV-1 IN), which catalyses the insertion of the HIV-1 DNA into the host genome, was also found to be inhibited (Figure 14a). However, similar inhibition was not observed in the case of a random double-stranded DNA, which has same base pair length as $(^{m}\mathbf{QQ^4})_8$. That is, the aromatic oligoamide backbone that make DNA-mimic foldamers distinct from DNA eventually give rise to enhanced inhibition. Moreover, inhibition by $(^{m}\mathbf{QQ^4})_{16}$ occurred at sub-micromolar concentrations, matching or even exceeding the performance of the best inhibitors of these enzymes, camptothecin for Top1 and raltegravir for HIV-1 IN.[119]

Inspired by the initial design of DNA-mimic foldamers, we started to work on variants of anionic foldamers bearing carboxylate groups in place of phosphonates. We compared the inhibition of the new variants to that of reference oligomer. The results show that both the position and nature of the anions strongly affected the inhibitory activity of the DNA mimics and that, in many cases, these effects differed between Top1 and HIV-1 IN (Figure 14b).[121] In a recent study, DNA-mimic foldamers have been shown to affect chromatin composition in vitro and in vivo, and disturb cell cycle progression (Figure 14c).[57] Altogether, these results further validate the novel concept of foldamer-based DNA surface mimicry

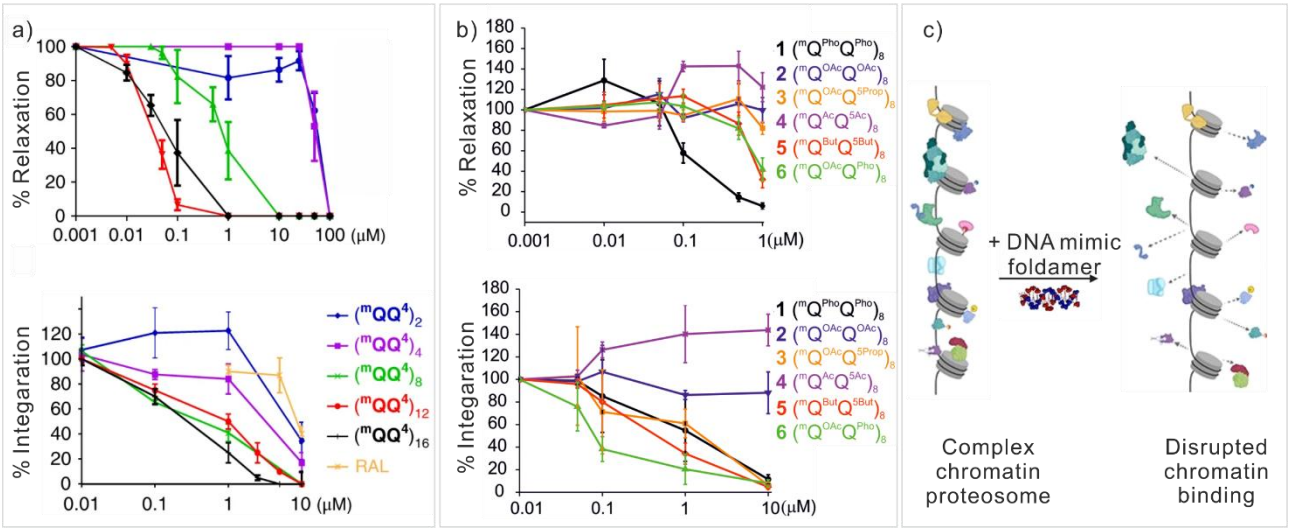


Figure 14. a) Inhibition of Top1 and HIV-1IN by DNA mimic with different length.[119] b) Inhibition of Top1 and HIV-1IN by DNA mimic foldamers with variant side chains.[121] Quantitation of Top1 inhibition (top lane) expressed as a percentage of relaxation, compared to the control, and in vitro HIV-1 IN integration (bottom lane). c) DNA mimic foldamer efficiently interferes with the chromatin-association of the origin recognition complex (ORC).[57]

4. Objectives

The prospect of achieving competitive inhibition is new when it comes to targeting the DNA-binding proteins. The common inhibition approaches usually involve DNA ligands[122-124] or 'interfacial inhibitors'[125-126] that bind the protein-DNA interface. Here we propose to generalize DNA mimics as competitive inhibitors of protein-nucleic acid interactions (PNIs) (Figure15). Despite being a homologous sequence, $(^{m}QQ^{4})_n$ oligomers showed a certain degree of selectivity. Several non-sequence selective enzymes were found not to be inhibited (e.g. deoxyribonuclease 1, S1 nuclease and benzonase) and some were weakly inhibited (topoisomerase 2, Flap endonuclease 1).[119] Slight changes in side chain composition and position led to selective topoisomerase 1 vs. HIV integrase inhibition.[121]

It is so far not yet clear why $(^{m}QQ^{4})_n$ helices bind some DNA-binding proteins much better than DNA itself. Thus, it is essential to understand which structural differences matter the most. Here we focus on the structural requirements for these foldamers to achieve stronger and more selective inhibition to DNA binding proteins.

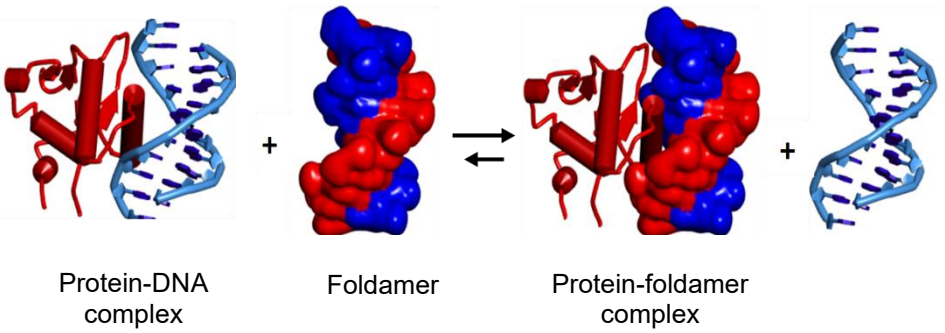


Figure 15. Principle of competitive inhibition of PNIs using a foldamer that mimic DNA surface features.

Sequence-specific DNA-binding proteins (DBPs) commonly use structural motifs like the helix-turn-helix (HTH),[127] zinc finger,[128] or leucine zippers[129] to interact with the DNA helix. Similarly, foldamers can be designed with specific groove features to adopt an α-helix, a protein motif widely involved in PNIs. For instance, functional groups can be introduced into the major grooves of DNA mimics, serving as hydrogen bond donors or acceptors (Figure 16a). Making the major groove deeper (by removing some rings) and wider can avoid steric hindrance that would prevent the protein binding. Stereo genic centers in these compounds will allow for quantitative helix handedness control (Figure 16b). Where needed, the flexible unit can serve as hinges to favor a local kink. Side chains can be varied, as it is recognized that not all the anions are needed for tight binding, and too many anions

reduce selectivity. Linkers will be developed to produce palindromic DNA mimic (Figure 16c), which increases the chances of crystallization at early stages when sequence selectivity is not strong. Alternatively, linkers can be used to produce DNA-foldamer hybrids, which can themselves be made palindromic (Figure 16d).

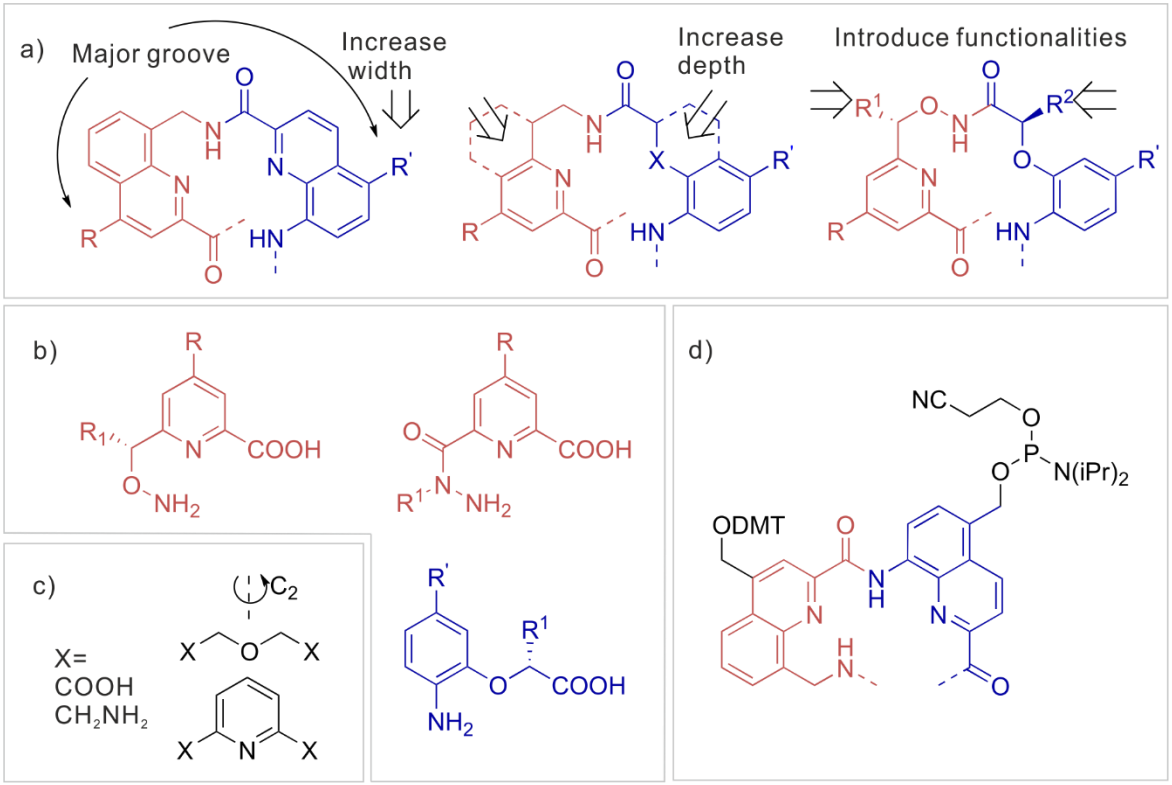


Figure 16. a) Modifications of the **Q^{m}Q** dimer are considered to change major groove width, depth and functionalities. b) **^{m}Q** analogues (in red) and **Q** analogues (in blue). c) diamine and diacid linkers to generate palindromic DNA mimic foldamers. d) **^{m}QQ** to be integrated in a hairpin oligonucleotide with a benzylic amine to attach the helical foldamer.

Last but not least, the DNA mimic foldamers, like many other polyanions, exhibits poor cell-membrane permeability. Preliminary studies have shown that DNA mimic foldamers are toxic to DNA topoisomerase I only in the presence of a transfection agent. Therefor it remains necessary to promote cellular delivery of DNA mimic foldamers.

In summary, future investigation alone this line will focus on different projects: 1) designing DNA mimic foldamers with more diverse features, such as chirality and symmetry for higher chance of crystal growth and affinity to DNA binding proteins. 2) gaining structural insight into the interactions between DNA-binding proteins and foldamers; 3) improving the cellular delivery efficiency of DNA

mimic foldamers; 4) optimizing DNA mimic foldamers for sequence-selective binding to protein targets.

5. Enhancing the features of DNA mimic foldamers for structural investigations

DNA mimic foldamers have been shown to bind to DNA-binding proteins with affinities that are sometimes orders of magnitude better than DNA itself. However, there are limited structural information about how these foldamers interact with DNA-binding proteins at a molecular level. To overcome the challenge of co-crystallizing these complexes and to facilitate the structural investigation, we pursued three strategies to enhance the features of our DNA mimic foldamers (Figure 17).

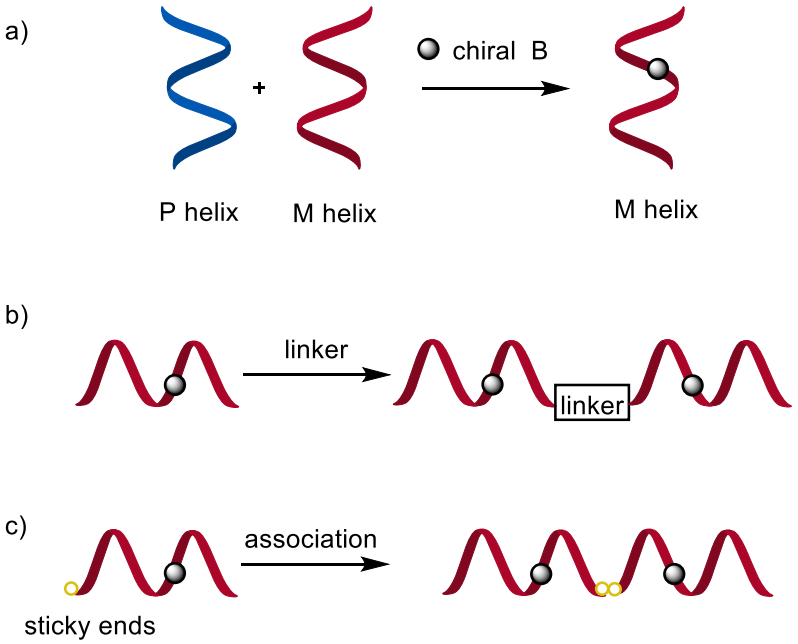


Figure 17. a) Handedness control by a chiral B monomer. b) Ligation chemistry to make C_2-symmetrical DNA mimic foldamers. c) Incorporation of sticky ends to promote helix stacking in crystal lattice.

First, we sought to achieve quantitative control over the foldamer's helicity. The original design based on repetitive building blocks with no stereogenic centre. As a result, foldamers exist in in a racemic mixture of right-handed (P) and left-handed (M) helical conformation in solution. This lack of helical uniformity can prevent selective recognition by DNA-binding proteins, thereby reducing the chance of crystal growth. To overcome this, we introduced a chiral B monomer at different position along the foldamer backbone. The following NMR and CD spectra confirmed that a B monomer placed in a specific position can bias the handedness quantitatively.

Second, we also developed ligation chemistry to make C_2-symmetrical DNA mimic foldamers. This design mimics palindromic DNA sequences and is particularly advantageous for proteins that recognize such symmetrical structures. Importantly, the recognition of proteins to C_2-symmetrical DNA mimic foldamer doesn't depend on the N→C orientation of the foldamer with respect to the protein surface. This symmetry might favour the crystal growth by reducing the structural heterogeneity of the complex.

Finally, we attempt to incorporate sticky ends into our foldamers, which might promote helix stacking and bring cohesion to the crystal lattice. This is a common strategy in DNA crystallography that we adapted for our foldamer system.

Our work on this topic has been published on *Chem.E.J*. In summary, we introduced new features in DNA mimic foldamer to promote the structural investigation of protein-foldamer complex. Benefiting from these enhancements, we obtained the first crystal structure of a DNA mimic foldamers bearing phosphonate side chain. Our findings demonstrated that 1) foldamer helix handedness can be quantitatively biased by a chiral B monomer; 2) the foldamer structure can be made C_2-symmetrical as in palindromic B-DNA sequences, by using a diamine or diacid linker; 3) associations between foldamer helices can be promoted by the C-terminal residues that act as sticky ends in B-DNA structures.

Contributions: The project was planned by IH. Synthetic monomer precursors have been provided by D. Gill. Monomer and foldamer synthesis have been performed by JW, VC and ML. The measurement of CD spectra and UV spectra was performed by VC. The energy minimized models were calculated by JW. Crystallographic studies and structure refinement were performed by DD. The research was supervised by IH. The manuscript was written by JW, VC and IH. JW, VC, ML, DD, IH and LA proofread and improved the manuscript. This work was supported the China Scholarship Council (CSC, predoctoral fellowship to JW.).

5.1. Publication

Enhancing the features of DNA mimic foldamers for structural investigations

Authors: V. Corvaglia, **J. Wu**, D. Deepak, M. Loos and I. Huc*.

Published: *Chem. Eur. J*. **2024**, 30, e202303650.

DOI: doi.org/10.1002/chem. 202303650.

Chemistry–A European Journal
Research Article
doi.org/10.1002/chem.202303650
Chemistry Europe
European Chemical Societies Publishing
www.chemeurj.org

Hot Paper

Enhancing the Features of DNA Mimic Foldamers for Structural Investigations

Valentina Corvaglia,[a, b] Jiaojiao Wu,[a] Deepak Deepak,[a] Manuel Loos,[a] and Ivan Huc*[a]

DNA mimic foldamers based on aromatic oligoamide helices bearing anionic phosphonate side chains have been shown to bind to DNA-binding proteins sometimes orders of magnitude better than DNA itself. Here, we introduce new features in the DNA mimic foldamers to facilitate structural investigations of their interactions with proteins. Thirteen new foldamer sequences have been synthesized and characterized using NMR, circular dichroism, molecular modeling, and X-ray crystallography. The results show that foldamer helix handedness can be quantitatively biased by means of a single stereogenic center, that the foldamer structure can be made C_2-symmetrical as in palindromic B-DNA sequences, and that associations between foldamer helices can be promoted utilizing dedicated C-terminal residues that act as sticky ends in B-DNA structures.

Introduction

DNA mimic foldamers are aromatic oligoamides bearing anionic side chains that adopt helically folded conformations in water, reproducing the shape and charge surface distribution of the B-DNA double helix.[1] They have been shown to bind to some DNA-binding proteins better than DNA itself and thus stand as candidates for the competitive inhibition of protein-DNA interactions. Conceptually, DNA mimic foldamers relate to naturally occurring DNA mimic proteins,[2] *i.e.*, proteins that mimic the shape and surface features of DNA and that inspire the design of nonnatural proteins to interfere with DNA-protein interactions.[3] DNA mimic foldamers may also be compared to the so-called decoy oligonucleotides that have been proposed to target DNA-binding proteins such as transcription factors.[4] More remotely, analogies can be drawn with sulfated polysaccharides such as heparin. These sulfated polysaccharides bind to DNA-binding proteins, and are used for their purification by affinity chromatography.[5] An original feature of DNA mimic foldamers is therefore their abiotic nature. Their chemical constitution is distinct from those of peptides, nucleotides, and saccharides. In addition, the way their structure can be modulated is also distinct.

The first proposed DNA mimic foldamers consisted of oligoamide sequences with alternating 8-amino-2-quinolinecarboxylic acid monomer Q^{Pho} and 8-aminomethyl-2-quinolinecarboxylic acid monomer M^{Pho}, both bearing negatively charged phosphonate residues (Figure 1).[1a] Carboxylate residues have been introduced subsequently.[1b] These sequences adopt stable single helical conformations stabilized by hydrogen bonds between amide NH protons and adjacent endocyclic quinoline nitrogen atoms and by the hydrophobic effect associated with

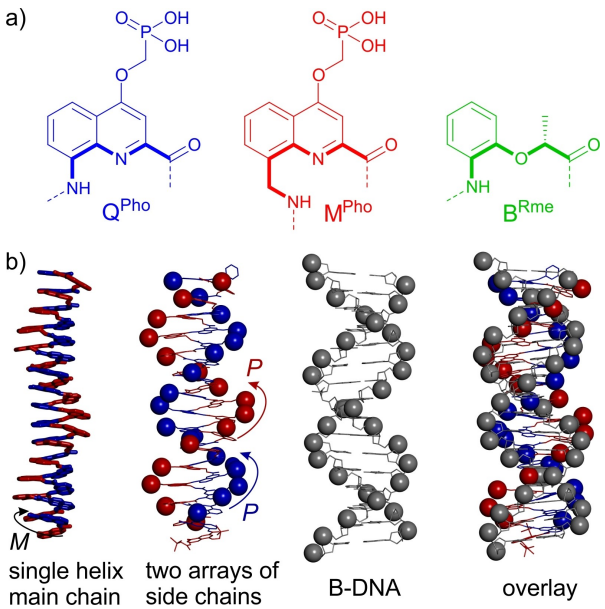


Figure 1. a) Amino acid monomers used to produce oligoamide sequences **1**–**7**. Bonds in bold indicate the inner rim of the helices involving these monomers. b) Crystal structure of the single helix of oligoamide $(M^{Pho}Q^{Pho})_{16}$ with protected ethyl phosphonate (left: aromatic oligoamide main chain only in tube representation; center left: same view with side chains),[1a] and of an ideal computationally-generated 16 base-pair B-DNA duplex (center right) and overlay of the two (right). Colors in b) are the same as in a). The structures are shown at the same scale in tube representation. Phosphorus atoms are shown in space-filling representation. Arrows show the helical sense of the main chain helix (*M*, in black) and of the arrays of side chains (*P* in red and blue).

[a] *Dr. V. Corvaglia, J. Wu, D. Deepak, M. Loos, Prof. Dr. I. Huc*
Department of Pharmacy
Ludwig-Maximilians-Universität München
Butenandtstr. 5–13, 81377 Munich (Germany)
E-mail: ivan.huc@cup.lmu.de

[b] *Dr. V. Corvaglia*
Current address: Institute for Stem-Cell Biology, Regenerative Medicine and Innovative Therapies, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Italy) & Center for Nanomedicine and Tissue Engineering (CNTE), ASST Grande Ospedale Metropolitano Niguarda, 20162 Milan (Italy)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202303650

aromatic stacking. The single helical nature of the main chain can be seen in Figure 1b (left). At the surface of the single helix, the phosphonate side chains of Q^{Pho} and M^{Pho} form a double helical array that matches the positions of phosphates in B-DNA (red and blue spheres in Figure 1b). The single helical nature of DNA mimic foldamers means they cannot be involved in Watson-Crick base pairing interactions, unlike many other DNA analogs such as peptide nucleic acids and locked nucleic acids.[6] It follows that the DNA mimic foldamers do not undergo dissociation, a feature that can be put to an advantage. For example, $(M^{Pho}Q^{Pho})4$ is a stable mimic of four base-pair DNA duplexes, that is, a duplex so short that it would not be stable at submillimolar concentrations.

The structural features of the DNA mimic foldamers enable them to tightly bind to some DNA-binding proteins, particularly nonsequence selective proteins that recognize DNA through its overall shape and charges. In some cases, $(M^{Pho}Q^{Pho})_n$ oligomers outcompeted DNA and bound to the proteins even in the presence of a large excess of DNA. This led, for example, to the strong inhibition of therapeutically relevant enzymes such as human topoisomerase 1 (Top1) and human immunodeficiency virus 1 integrase (HIV-1 IN).[1] In a recent study, DNA mimic foldamers have been shown to affect chromatin composition and perturb cell cycle progression.[7] These discoveries call for further developments, in particular with regard to the elucidation of the interaction modes between the foldamers and DNA-binding proteins. However, $(M^{Pho}Q^{Pho})_n$ foldamers are not ideally suited for this purpose. For instance, they do not possess any stereogenic center and thus exist as racemic mixtures of right-handed (*P*) and left-handed (*M*) helices, potentially producing mixtures of diastereomeric complexes with proteins. In addition, they possess an N terminus and a C terminus which may lead to alternate protein binding modes depending on the N→C orientation of the foldamer with respect to the protein surface. With DNA, this potential problem can be alleviated by using C_2-symmetrical palindromic sequences. Furthermore, growing crystals of protein-DNA complexes often benefit from promoting associations between DNA strands using overhanging bases that mediate inter-duplex base-pairing.[8] The DNA mimic foldamers possess may not have the same ability to self-assemble. Here, we present the implementation of various modifications of the DNA mimic foldamers in order to bias their handedness, make them C_2-symmetrical, *i.e.*, palindromic-like, and promote their association through helix cross-sections. Some of these enhanced features have already helped to solve a crystal structure of a complex between a DNA mimic foldamer and a protein.[9]

Results and Discussion

Biasing Handedness in DNA Mimic Foldamers

Natural B-DNA contains D-deoxyribose and is a right-handed (*P*) double helix. To mimic the structure of B-DNA, the single helical foldamers should thus also display a *P* double helical array of side chains at the surface of the main chain foldamer single helix. These arrays are called exo-helices. It turns out that *P* exo-helicity is achieved when the aromatic oligoamide main chain has *M* helicity, that is with an opposite sense (Figure 1b). This may not be intuitive and can be explained as follows: the repeat units of the foldamer single helix, $M^{Pho}Q^{Pho}$ dimers, span ~0.9 helix turn. Within a sequence, two consecutive $M^{Pho}Q^{Pho}$ dimers thus have an angular shift of ~0.1 turn while stacking on top of each other, which results in a vertical rise of 3.4 Å. These are the same vertical rise and angular shift as between consecutive base pairs in B-DNA duplexes, hence the resemblance between DNA mimic foldamers and B-DNA. However, because $M^{Pho}Q^{Pho}$ dimers span less than a turn, the 0.1 turn angular shift of the next $M^{Pho}Q^{Pho}$ dimer takes place "backwards" along the helix backbone. As a result, the handedness of the side chain exo-helices is opposite to that of the main chain. Exo-helices and main chain helix would have the same handedness if the repeat unit would span more than a turn. One might, for example, achieve B-DNA mimicry with a repeat unit spanning 1.1 turn, which would also result in a 0.1 angular shift, this time "forward", between repeat units.

Quantitative handedness control has been achieved in the context of helical oligoamides of 8-amino-2-quinolinecarboxylic acid Q, *i.e.*, analogs of Q^{Pho} bearing various types of side chains, through appended chiral groups at the N or C termini.[10] Recently, we have shown that a 2-(2-aminophenoxy)propionic acid unit B^{Rme} (Figure 1a) within a Q_n sequence is also efficient at biasing *M* handedness in water.[11] However, it was unknown whether this approach would also be effective in the context of $(M^{Pho}Q^{Pho})_n$ sequences, in particular with regards to the additional flexibility imparted by the methylene groups of M^{Pho} which increase the number of main chain rotatable bonds. We thus designed DNA mimic foldamers **2**–**7** as analogs of reference sequence **1** to assess the effect of B^{Rme} in various sequence contexts (Figure 2a).

In **2** and **3**, B^{Rme} replaces one Q^{Pho} or one M^{Pho}, respectively, within a $(M^{Pho}Q^{Pho})_n$ motif. In constrast, in **4** and **5**, B^{Rme} is inserted after Q^{Pho} or M^{Pho}, respectively. Because an insertion results in a frameshift in the alternation of Q^{Pho} and M^{Pho}, sequences **6** and **7** were also designed where the insertion of B^{Rme} is compensated by the insertion of an additional M^{Pho} later or earlier in the sequence, respectively. For the preparation of these oligomers by solid phase synthesis (SPS), new monomers Fmoc-Q^{Pho}-OH and Fmoc-M^{Pho}-OH were synthesized with their phosphonate side chain protected as diethyl esters (Schemes S1–S2), a protection previously validated for solution phase synthesis.[1a] The di-*tert*-butyl esters of these Fmoc-acid precursors have been introduced earlier,[1b] but the high acid lability of the *tert*-butyl groups make the synthesis and handling of the monomers more delicate. SPS was performed on low loading Wang resin (100–200 mesh) using previously reported protocols (Scheme S3).[12] Fmoc-acid building blocks were activated *in situ* by generating the respective acid chlorides prior to coupling. The oligomers were cleaved from the resin with trifluoroacetic acid (TFA) and purified using semipreparative reverse phase high performance liquid chromatography (RP-HPLC) under acidic conditions (water/acetonitrile containing 0.1% TFA). Removal of the diethylphosphonate protections was then

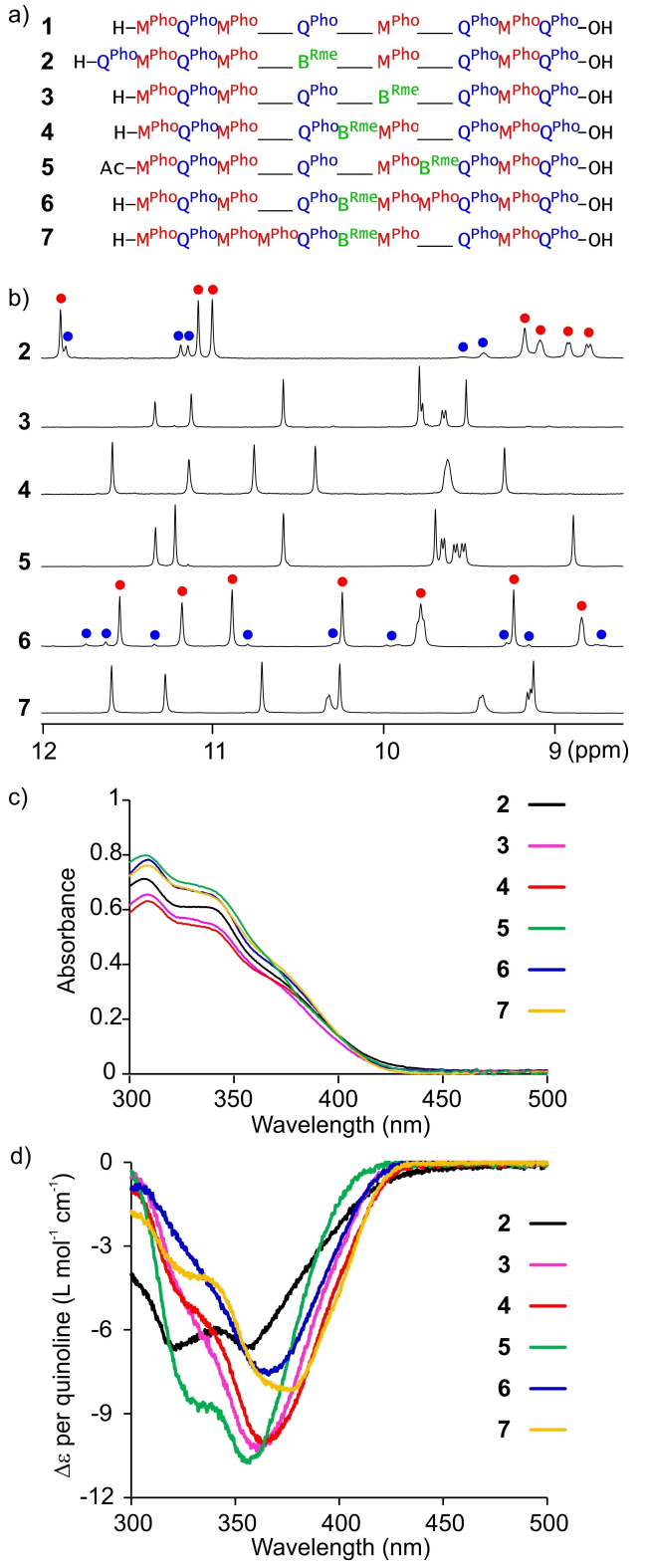


Figure 2. a) Foldamer sequences synthesized to investigate the handedness induction properties of B^{Rme}. Long dashes indicate that there is no monomer and serve to align the identical segments of the various sequences. b) Part of the ^1H NMR (500 MHz) spectra at 333 K of **2**–**7** in 50 mM NH_4HCO_3 (pH 8.5) in H_2O/D_2O (9:1, v/v) showing amide N*H* resonances. For **2** and **6**, major and minor sets of signal are marked with red and blue circles, respectively. c) UV absorption spectra of **2**–**7** (60 μM, 333 K in 50 mM NH_4HCO_3 pH 8.5) in the region of interest for the CD spectra. d) CD spectra of **2**–**7** (60 μM, 333 K in 50 mM NH_4HCO_3 pH 8.5). Molar extinction (Δε) is normalized for the number of quinoline (Q^{Pho} or M^{Pho}) units.

performed using trimethylsilyl bromide (TMSBr), and the oligomers were further purified by semipreparative RP-HPLC using a basic triethylammonium acetate buffer (pH 8.5). In a final step, they were submitted to ion exchange to deliver the side chains as water-soluble ammonium phosphonate salts providing the foldamers in 10–15% isolated yields (after two HPLC purifications and ion exchange) and a purity > 97%.

Helix handedness bias by B^{Rme} was assessed by nuclear magnetic resonance (NMR) and circular dichroism (CD) spectroscopies (Figure 2b, d). Both CD and NMR confirmed that all sequences are helically folded. NMR spectra show sharp amide and aromatic protons signals distributed over a wide range of chemical shift values (from 12 to 9 ppm for amide protons and from 8.5 to 6.5 ppm for aromatic protons, Figures 2b, S1) in agreement with previous observations.[1,12a] All CD spectra showed an intense negative band near 360 nm, typical of helically folded quinoline oligoamides with predominant *M* helicity (Figure 2d).[11,13] As a consequence of the different sequence composition and the different monomer contacts in the helices, the CD profiles show some differences. The λ_{max} values vary slightly from compound to compound and so does the relative contribution of a second negative band at 320–330 nm. Compound **2** has a distinct aromatic amine chromophore and compound **5** is acetylated at its N terminus whereas other sequences have a benzylic ammonium terminus. These features may perhaps relate to the distinct CD signatures of **2** and **5** which both have bands near 330 and 360 nm of comparable intensities. However, finding a trend and assigning the intensities to specific interactions within the helices is difficult and was not attempted. For example, sequences **6** and **7** differ minimally through the position of one M^{Pho} unit. Yet the contribution of the weaker CD band near 330 nm is barely visible in the spectrum of **6** while it is very clear for **7**. For comparison, the UV-vis absorption spectra of the various sequences are similar (Figure 2c).

Sequences **2**–**7** contain eight to ten residues and should all span over three full helix turns. For such long sequences in water, equilibrium between *M* and *P* diastereomeric conformers is expected to be slow on the NMR timescale. This is illustrated in the spectrum of achiral **1** (Figure S1) in which the signals of $C\underline{H}_2$ protons form AB systems reflecting their anisochronicity. Under a fast exchange regime, these signals would average. For chiral oligomers **2**–**7**, *P* and *M* conformational diastereomers should thus appear as distinct sets of NMR signals. The single set of resonances observed for **3**, **4**, **5**, and **7** therefore demonstrates that handedness bias towards the *M* helix is quantitative – as far as NMR can detect – in these cases (Figure 2b). In contrast, the second set of signals amounting to 30% in the spectrum of **2** and that, less intense (4%), in the spectrum of **6**, indicate the presence of some *P*-helical conformer. The incomplete handedness bias in **2** and **6** is also reflected in less intense CD bands for these compounds (Figure 2d). However, CD intensity is unreliable to quantitatively assess handedness bias because intensity also varies as a function of sequences even when handedness bias is quantitative. An illustration is the weaker CD band of **7** (Figure 2d). Its normalized CD intensity is 25% smaller than that of **5** whereas

these two compounds absorb similarly (Figure 2c). One should point that the Δε normalized "per quinoline ring" reported in Figure 2d does not consider possible differences in the contributions of Q^{Pho} and M^{Pho} monomers to the CD.

In summary, B^{Rme} efficiently biases helix handedness in DNA mimic foldamers but only in the context of certain sequence patterns. Of note, the partial handedness bias observed for **2** and **6** concerns the only two sequences where the B^{Rme} monomer is part of three consecutive more flexible, methylene-containing, monomers ($M^{Pho}B^{Rme}M^{Pho}$ or $B^{Rme}M^{Pho}M^{Pho}$).

Structural Consequences of a B^{Rme} Monomer in DNA Mimic Foldamer Sequences

B^{Rme} is a δ-amino acid and may thus be considered to be a structural analog of Q^{Pho} rather than ε-amino acid M^{Pho}. The Q^{Pho}/B^{Rme} mutation, as in **2**, might appear to be the most desirable way to introduce B^{Rme} in a sequence while preserving DNA mimicry. Yet, we have seen that handedness bias is insufficient in this case. To evaluate how the replacement of Q^{Pho} or M^{Pho} by B^{Rme} or the insertion of B^{Rme} alter the shape of DNA mimic foldamers, energy minimized molecular models were produced (Figures 3, S2–S4).

For a better visualization of the double helical arrangement of negative charges and of the major and minor grooves, model oligomers were elongated up to 48 units. Segments corresponding to **2**–**7** were introduced in the middle of these sequences and final objects were compared to a reference $(M^{Pho}Q^{Pho})_{24}$ DNA mimic foldamer with no stereogenic center. Three representative examples are shown in Figure 3, including the replacement of M^{Pho} by B^{Rme}, the insertion of B^{Rme}, and the double insertion of B^{Rme} and M^{Pho}. As predicted, the energy-minimized model of the sequence corresponding to an extended-**2** shows the smooth insertion of the B^{Rme} unit replacing one Q^{Pho} (Figure S2). Interestingly, molecular modeling also predicts a smooth insertion when B^{Rme} replaces M^{Pho} (extended-**3** in Figure 3a, b). At the scale of these large objects, replacing Q^{Pho} or M^{Pho} does not cause significant structural differences. In both extended-**4** (Figure 3c, d) and extended-**5** (Figure S3), the inserted B^{Rme} causes a swap of the major and minor grooves of the DNA mimic. We have described in earlier publications that, in the case of $(M^{Pho}Q^{Pho})_n$ sequences, and unlike B-DNA, the two grooves have similar widths.[1,14] Because of these matching widths, the swap then has limited structural consequences for what concerns the overall shape. It remains that M^{Pho} and Q^{Pho} line opposite sides of each groove and that this role is reverted if there is a groove swap. Furthermore, other DNA mimics have been designed where the groove widths differ more[1] and for which a swap of grooves may constitute an important structural change. Nevertheless, one cannot exclude potential benefits of such a swap to recognize some DNA-binding proteins. In extended-**6** (Figure 3e, f) and extended-**7** (Figure S4), the grooves do not swap so groove swapping can be avoided using such sequences. Initial attempts to crystallize sequences **2**–**7** to consolidate the results of

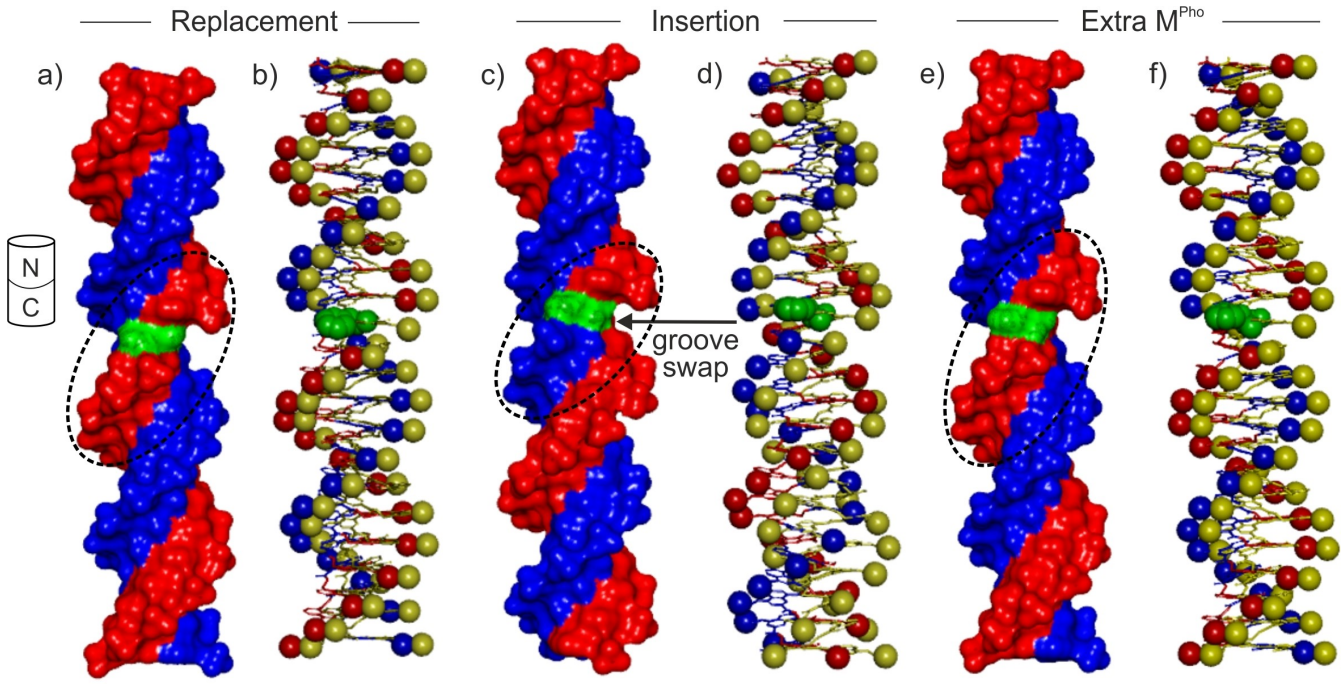


Figure 3. Energy minimized molecular models (Maestro, MMFFs force field, water as implicit solvent)[15] of: extended-**3** (a) and its overlay with $(M^{Pho}Q^{Pho})_{24}$ (b); extended-**4** (c) and its overlay with $(M^{Pho}Q^{Pho})_{24}$ (d); and extended-**6** (e) and its overlay with $(M^{Pho}Q^{Pho})_{24}$ (f). In a), c), and e), the molecule isosurface is shown with monomers color-coded as in Figure 1. In b), d) and f), molecules are shown in tube representation except the phosphorous atoms which are shown in space-filling representation. $(M^{Pho}Q^{Pho})_{24}$ is in yellow and the other molecule is color-coded as in Figure 1. The area near the B^{Rme} monomer is encircled by a dashed line in a), c), and e). The arrow in c) indicates a minor/major groove swapping on each side of the B^{Rme} monomer: monomers on either side of the B^{Rme} monomer have different colors. In contrast, this does not occur in a) and e).

molecular modeling by experimental observations did not yield diffracting crystals. However, a crystal structure was eventually obtained upon adding further features to the sequence, as described in the last section below.

Palindromic-Like C_2-Symmetrical DNA Mimic Foldamers

The next DNA mimic foldamer feature we envisaged is C_2 symmetry, by analogy with palindromic DNA sequences,[16] *i.e.*, DNA sequences that are self-complementary and can thus form C_2-symmetrical duplexes. Note that palindromes also exist in peptides but with a definition that does not entail symmetry.[17] Interest in C_2-symmetrical DNA mimic foldamers stemmed from the prospect of targeting proteins that recognize palindromic DNA, such as some transcription factors[18] and restriction enzymes,[19] which are often themselves C_2-symmetrical homodimers. Furthermore, the geometry of complexes between palindromic DNA and a given protein does not depend on the 5′-3′ DNA orientation. This degeneracy is desirable for crystal growth where the presence of distinct complexes can be an impediment. Thus, palindromic DNA sequences have been used to grow crystals of DNA-protein complexes, even with proteins that do not typically target such sites.[20] We therefore anticipated that this feature would be useful for DNA mimic foldamers as well.

C_2 symmetry was introduced in the DNA mimic foldamers by means of central diamine or diacid linkers that locally revert the C→N strand polarity. Four different linkers were considered, leading to the design of sequences **8**–**11** (Figure 4). Many others may be conceived depending on the desired outcome. The aliphatic linker of **8** and the aromatic linker of **9** are expected to contribute similarly to helix curvature. The rigid 2,6-pyridinedicarboxylic acid linker of **9** is a classical helicogenic monomer in aromatic oligoamide foldamers.[21] The linker of **8** is more flexible, even though its central oxygen atom may form hydrogen bonds with neighbor amide protons, stabilizing a helical structure as does the endocyclic nitrogen atom of **9**. The diamine linkers of **10** and **11** have been explored previously in the context of organic solvent-soluble Q_n oligomers.[22] Molecular models of **8**–**11** highlight their predicted symmetrical structures (Figure 4). The linkers provide various alternatives to put the grooves of the two halves of the molecules more or less in register. They may allow to deviate from ideal DNA mimicry, depending on what is desired. In this respect, the more flexible aliphatic linkers may provide a range of acceptable local helix curvature. They may also play the role of a hinge and permit a local kink in the foldamer structure to target proteins known to kink DNA.[23]

To prepare palindromic-like DNA mimic foldamers **8**–**11**, we developed a solution-phase fragment condensation approach using chiral sequence **4** as a building block (Figure 5a). The diethylphosphonate-protected precursor of **4** prepared by SPS could be cleaved from the resin before or after acetylation of the N terminus to provide fragments A and B, respectively, which were purified by RP-HPLC. Fragment A was reacted with diacid linkers L1 activated as *N*-hydroxysuccinimide (NHS) esters

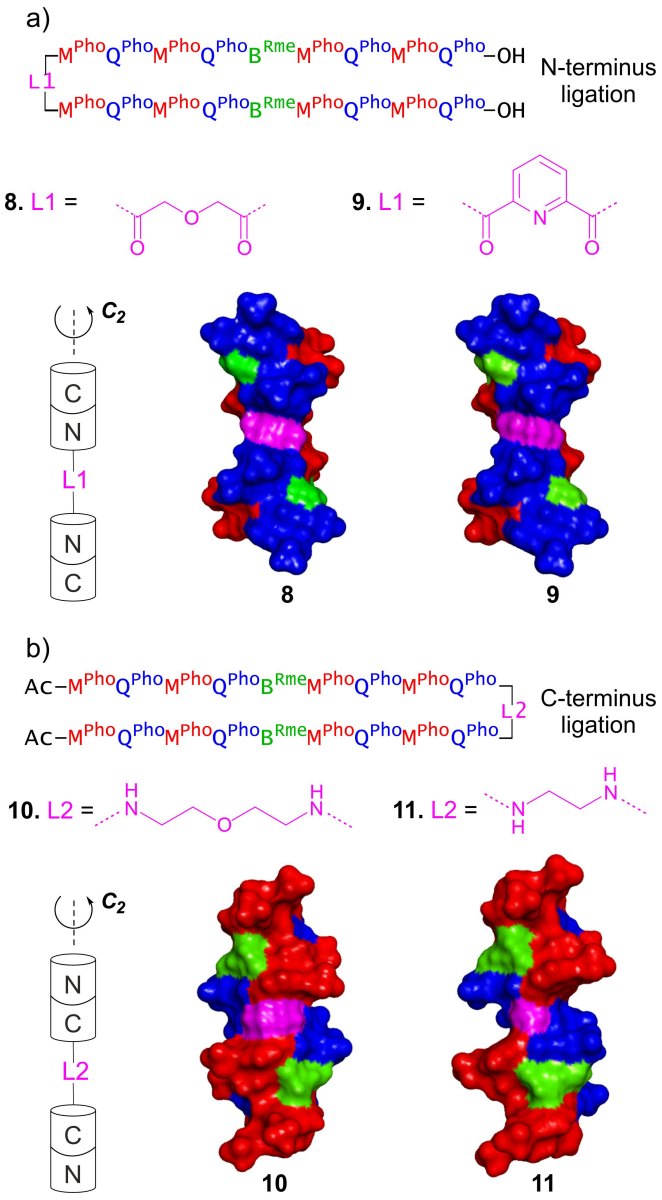


Figure 4. Palindromic-like DNA mimic foldamers. a) Formulas and energy-minimized molecular models of **8** and **9** with central diacid linkers. b) Formulas and energy-minimized molecular models of **10** and **11** with central diamine linkers. Models were obtained with Maestro (MMFFs force field, water as implicit solvent).[15] They are shown at the same scale down their C_2 symmetry axis (through the linker in purple) as solvent accessible isosurfaces color-coded according to the type of monomer as in Figure 1.

to yield the precursors of **8** and **9** which were further purified by RP-HPLC (Scheme S4). Conversely, acetylated fragment B was coupled to the diamine linkers L2 using benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) activation to yield the protected precursors of **10** and **11** which were also purified by RP-HPLC (Scheme S5). Cleavage of the diethylphosphonate esters with TMSBr yielded the final products which were again purified by RP-HPLC and subjected to ion exchange to generate ammonium salts.

The crude HPLC traces shown in Figure 5b highlight the high conversion yields of the double couplings on the linkers

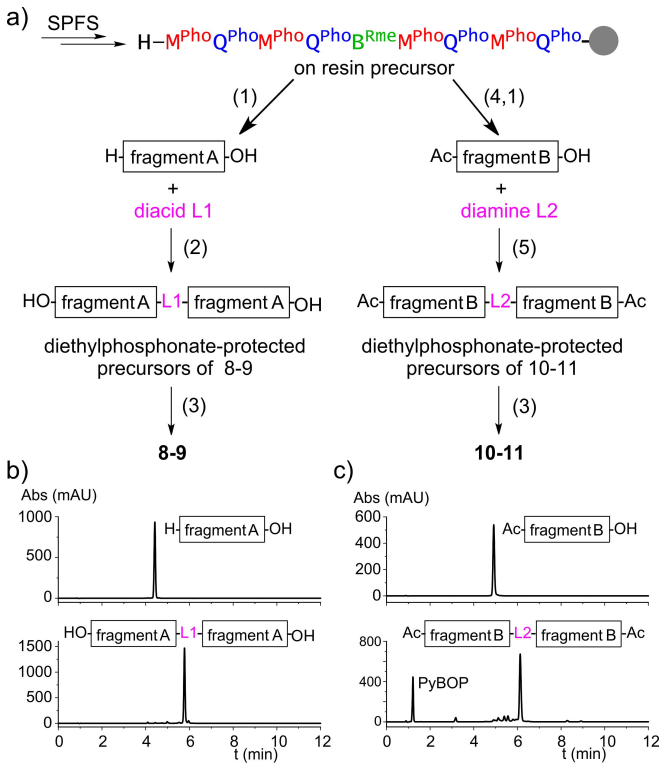


Figure 5. a) Scheme for the solution phase synthesis of **8**–**11** from a common precursor produced by SPS. (1) TFA-mediated resin cleavage; (2) double condensation of the N terminus of fragment A on diacid linkers activated as bis-NHS esters; (3) deprotection of diethyl-phosphonate esters with TMSBr; (4) acetylation of N terminus; (5) double condensation of diamine linkers on the C terminus of fragment B activated with PyBOP. b) Representative examples of the RP-HPLC traces of crude products after step (2) (left) or (5) (right). The chromatograms of the pure fragment precursors are shown above for comparison. Conditions: linear gradient from 30% to 80% solvent B in 15 min, and C8 column; A: H_2O + 0.1% TFA and B: acetonitrile + 0.1% TFA.

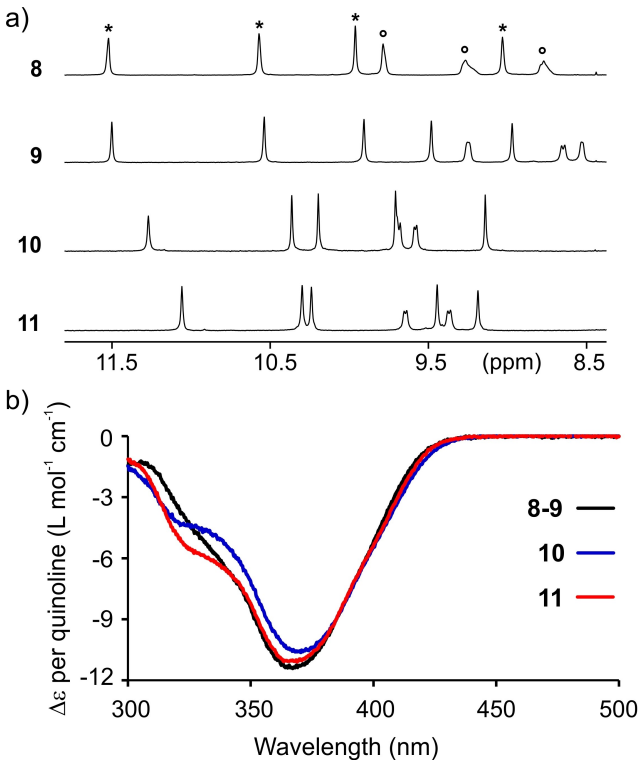


Figure 6. a) Part of the ^{1}H NMR (500 MHz) spectra at 333 K of **8**–**11** in 50 mM NH_4HCO_3 (pH 8.5) in H_2O/D_2O (9:1, vol/vol) showing amide resonances. Signals assigned to aromatic N*H*s and aliphatic N*H*s are indicated by stars and circles, respectively. The signals of other aliphatic N*H*s are found at higher fields and overlap with aromatic protons. b) CD spectra of **8**–**11** recorded at 333 K in 50 mM aqueous NH_4HCO_3 (pH 8.5). The molar extinction coefficient ($\Delta\varepsilon$) is normalized for the number of quinoline units (Q^{Pho} and M^{Pho}). The linkers and B^{Rme} units do not absorb in this region.

(see also the Supporting Information). With both diamine and diacid linkers, intermediates where the linker had reacted only once were consumed or kept to negligible amounts. In turn, the high yields greatly facilitated HPLC purification and the final sequences were obtained in 7–11% isolated yield from initial resin loading with >97% purity. Even though these reactions are not the first fragment couplings between foldamer segments prepared on solid phase,[24] the conversion yields reflect particularly well-behaved couplings without requiring excess reagent. This chemistry is currently being successfully extended to foldamer segments twice as long as **8**–**11**. Progress will be reported in due course.

The folding and symmetry of **8**–**11** are reflected in their ^{1}H NMR and CD spectra (Figure 6). All NMR spectra show five degenerate aromatic N*H* resonances indicating an overall symmetrical structure (Figure 6a). Not all resonances N*H* can be distinguished in the 1D ^{1}H NMR spectra because some overlap with aromatic C*H* signals. However, 2D ^{15}N–^{1}H heteronuclear single quantum coherence (HSQC) spectra show the expected number of resonances for symmetrical structures, that is, half the total number of NH functions (Figure S5). The CD spectra confirm the *M* helicity (Figure 6b). The single sets of ^{1}H NMR signals show that handedness bias is quantitative. The combination of *M* helicity and overall symmetry can be fulfilled only with an average C_2 symmetry axis; a center of inversion or a symmetry plane can be excluded. The similar ^{1}H NMR spectra of **8** and **9** as well as their perfectly overlapping CD spectra suggest that the two diacid linkers produce a similar arrangement of the two halves of the molecule. In contrast, the slight differences in the CD spectra of **10** and **11** can be interpreted as resulting from different aryl-aryl contacts due to the different lengths of the diamine linkers.

Adding Sticky Ends to DNA Mimic Foldamers

In B-DNA, sticky ends, as opposed to blunt ends, refer to overhanging unpaired nucleotides at the 3′ or 5′ of a strand within a duplex. They occur naturally from the staggered cut of B-DNA by restriction enzymes. Even when such overhangs contain as few as two bases, they may promote base pairing at high (millimolar) concentrations. This property can be advantageous when crystallizing DNA alone or in complex with proteins. The overhanging bases promote inter-duplex base-pairing and may favor the formation of continuous stacks of B-DNA duplexes that bring cohesion to the crystal lattices.[8] We

devised that inter-helix associations may also be implemented in DNA mimic foldamers utilizing different types of sticky ends. For example, some aromatic oligoamide sequences have been shown to promote dimerization via anti-parallel double helix formation.[25] Other sequences mediate the head-to-head stacking of the C-terminal cross-section of aromatic single helices.[26] Sequence **12** (Figure 7a) was thus conceived to contain an $(M^{Pho}Q^{Pho})_6$ DNA mimic segment at its N terminus, handedness control by a B^{Rme} unit, and a $Q^{Ala}Q^{Ala}Q^{Asp}$ C-terminal segment known to promote dimerization by stacking of single helices and that amount here to a sticky end. The lack of a charged side chain in the penultimate residue at the C terminus has been shown to be important for dimerization to occur.[26] The ^{1}H NMR spectrum of **12** was recorded at different concentrations (Figure S6). As with previous oligomers,[26] two sets of signals were observed whose proportions vary with concentration, in agreement with a monomer-dimer equilibrium. Sequences such as **1** show no such effect.

Evidence of head-to-head dimerization via stacking the C-terminal cross-section of two helices was obtained from the solid state structure of **13**. This sequence is an analog of **12** where a Q^{Sem} replaces a Q^{Ala}. The selenium-containing monomer was intended to facilitate X-ray structure elucidation using anomalous scattering, but this proved unnecessary for **13**. Crystals of **13** were obtained using the hanging drop method and diffracted up to 3 Å (see Supporting Information). The structure was solved by molecular replacement using an energy-minimized molecular model. The asymmetric unit contained two molecules (Figure S7). It confirmed the handedness bias imparted by the B^{Rme} monomer as well as dimerization mediated by the sticky ends at the C terminus (Figure 7c box). As with earlier structures, helix-helix interactions bring the two C-terminal carboxylates in close proximity, from which we infer that one of the two must be in its carboxylic acid form and the interaction be mediated by a hydrogen bond.

The crystal structure of **13** is the first of a DNA mimic foldamer bearing phosphonate side chains. Previously, structures of the diethylphosphonate precursors have been obtained after growing crystals from organic solvents.[1a] A structure of a carboxylate analog had also been obtained using a crystal grown from an aqueous solution.[1b] However, all prior attempts to solve the structure of the parent series with phosphonate side chains had been unsuccessful until now. At best, crystals that did not diffract were obtained. Whether the better result with **13** reflects a contribution from the sticky end, for example, by stiffening the lattice, cannot be ascertained from only one example. Obviously, the sticky end was not an impediment. The crystal lattice features remarkably large pores with up to 77% solvent content (Figure 7c, S8) that explain the low diffraction intensity and low resolution of the collected data. One may speculate that the strong negative charge density of DNA mimic foldamers with phosphonate side chains – each side chain may be up to two negative charges – may be responsible for the large pores of the structure of **13** and for the general difficulty in obtaining diffracting crystals. In nucleic acid solid-state structures, negative charge density is often partially screened by divalent metal cations. Such cations were also

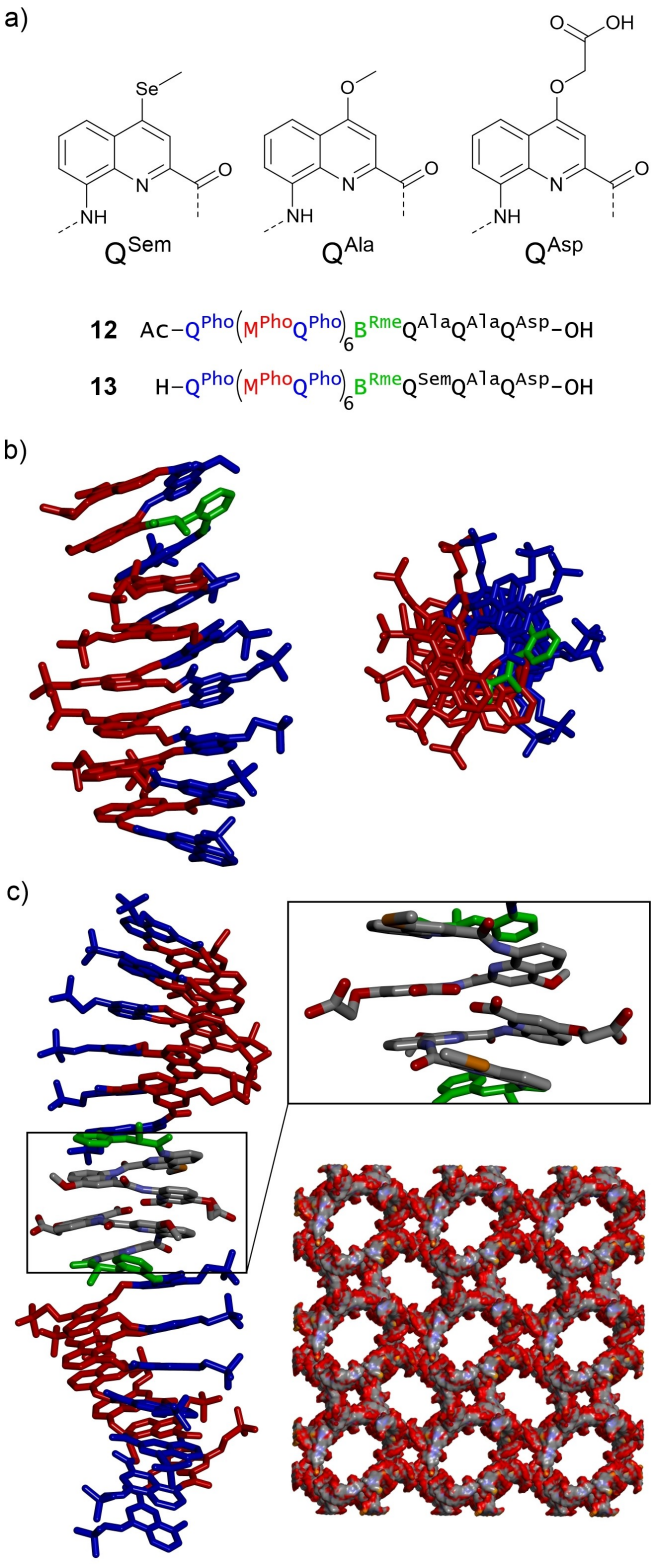


Figure 7. a) Formula of Q^{Sem}, Q^{Ala} and Q^{Asp} monomers and of sequences **12** and **13**. b) Solid state structure of **13** (PDB# 8QHM) showing the side view and top view of one of the two molecules found in the asymmetric unit. c) Crystal packing. Two helices stacked head-to-head via their C-terminal cross-sections are shown on the right. The box enlarges the contact area. On the left, a view of the crystal lattice down the *c* crystallographic axis highlights the very large pore of the structure.

involved in the structure of the carboxylate functionalized DNA mimic foldamers.[1b] However, this effect has until now not been helpful to crystallize phosphonate-functionalized aromatic helices.

Conclusions

In conclusion, we have successfully implemented new features in DNA mimic foldamers that are typical of nucleic acid double helices. Specifically, handedness control, palindromic-like C_2 symmetry, and chain extensions that promote helix-helix associations akin to sticky ends in nucleic acids can all be implemented in the DNA mimics. These features are useful to elucidate the structures of DNA-protein complexes and we anticipate that they will be similarly helpful to investigate the structures of complexes between DNA mimic foldamers and DNA-binding proteins. Steps in this direction have already been made.[9]

Supporting Information

The data that support the findings of this study are available in the supplementary material of this article. The authors have cited additional references within the Supporting Information.[27–33]

Acknowledgements

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) via projects HU1766/2-1 and 325871075 (CRC1309-C7), from the European Research Council (ERC) under the European Union's Horizon Europe Framework Programme (Grant Agreement No. ERC-2021-ADG-320892), from the German Academic Exchange Service (DAAD, predoctoral fellowship to D. D.) and from the China Scholarship Council (CSC, predoctoral fellowship to J. W.). We thank D. Bindl for providing the B^{Rme} monomer and L. Allmendinger for assistance with NMR measurements. Synchrotron data were collected at beamline ID23-1 at ESRF Grenoble, France. We thank A. McCarthy for his assistance in using the beamline and J. Sigl for his help with the generation of the geometric restraint library. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The crystallographic data that support the findings of this study are openly available in Protein Data Bank at https://doi.org/10.2210/pdb8QHM/pdb. Other data may obtained from the authors upon reasonable request.

Keywords: Aromatic oligoamide · DNA mimicry · Foldamers · Helix handeness control · Structure elucidation

[1] a) K. Ziach, C. Chollet, V. Parissi, P. Prabhakaran, M. Marchivie, V. Corvaglia, P. P. Bose, K. Laxmi-Reddy, F. Godde, J. M. Schmitter, S. Chaignepain, P. Pourquier, I. Huc, *Nat. Chem.* **2018**, *10*, 511–518; b) V. Corvaglia, D. Carbajo, P. Prabhakaran, K. Ziach, P. K. Mandal, V. D. Santos, C. Legeay, R. Vogel, V. Parissi, P. Pourquier, I. Huc, *Nucleic Acids Res.* **2019**, *47*, 5511–5521.

[2] a) S. S. Hegde, M. W. Vetting, S. L. Roderick, L. A. Mitchenall, A. Maxwell, H. E. Takiff, J. S. Blanchard, *Science* **2005**, *308*, 1480–1483; b) P. A. Tsonis, B. Dwivedi, *Biochim. Biophys. Acta Mol. Cell Res.* **2008**, *1783*, 177–187; c) H.-C. Wang, C.-C. Chou, K.-C. Hsu, C.-H. Lee, A. H. J. Wang, *IUBMB Life* **2019**, *71*, 539–548; d) H.-C. Wang, C.-H. Ho, K.-C. Hsu, J.-M. Yang, A. H. J. Wang, *Biochemistry* **2014**, *53*, 2865–2874; e) H.-C. Wang, H.-C. Wang, T.-P. Ko, Y.-M. Lee, J.-H. Leu, C.-H. Ho, W.-P. Huang, C.-F. Lo, A. H. J. Wang, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20758–20763; f) M. D. Walkinshaw, P. Taylor, S. S. Sturrock, C. Atanasiu, T. Berge, R. M. Henderson, J. M. Edwardson, D. T. F. Dryden, *Mol. Cell* **2002**, *9*, 187–194; g) F. Ye, I. Kotta-Loizou, M. Jovanovic, X. Liu, D. T. F. Dryden, M. Buck, X. Zhang, *eLife* **2020**, *9*, e52125.

[3] a) D. Yüksel, P. R. Bianco, K. Kumar, *Mol. BioSyst.* **2016**, *12*, 169–177; b) F. Haque, C. Freniere, Q. Ye, N. Mani, E. M. Wilson-Kubalek, P.-I. Ku, R. A. Milligan, R. Subramanian, *Nat. Cell Biol.* **2022**, *24*, 1088–1098; c) D. T. F. Dryden, *Trends Biotechnol.* **2006**, *24*, 378–382; d) H.-C. Wang, C.-H. Ho, C.-C. Chou, T.-P. Ko, M.-F. Huang, K.-C. Hsu, A. H. J. Wang, *Nucleic Acids Res.* **2016**, *44*, 4440–4449.

[4] a) R. Crinelli, M. Bianchi, L. Gentilini, M. Magnani, *Nucleic Acids Res.* **2002**, *30*, 2435–2443; b) B. Johari, M. Moradi, in *Methods in Molecular Biology, Vol. 2521* (Ed.: W. Walther), Humana, New York, NY, **2022**, pp. 207–230; c) B. Johari, J. Zargan, *Cell Biol. Int.* **2017**, *41*, 1335–1344; d) G. Casas, F. Perche, P. Midoux, C. Pichon, J. M. Malinge, *Mol. Ther. Nucleic Acids* **2022**, *29*, 162–175.

[5] a) M. C. Z. Meneghetti, A. J. Hughes, T. R. Rudd, H. B. Nader, A. K. Powell, E. A. Yates, M. A. Lima, *J. R. Soc. Interface* **2015**, *12*, 20150589; b) A. K. Powell, E. A. Yates, D. G. Fernig, J. E. Turnbull, *Glycobiology* **2004**, *14*, 17R-30R.

[6] a) A. Porcheddu, G. Giacomelli, *Curr. Med. Chem.* **2005**, *12*, 2561–2599; b) A. Gupta, A. Mishra, N. Puri, *J. Biotechnol.* **2017**, *259*, 148–159; c) D. A. Braasch, D. R. Corey, *Chem. Biol.* **2001**, *8*, 1–7; d) B. Vester, J. Wengel, *Biochemistry* **2004**, *43*, 13233–13241.

[7] V. Kleene, V. Corvaglia, E. Chacin, I. Forne, D. B. Konrad, P. Khosravani, C. Douat, C. F. Kurat, I. Huc, A. Imhof, *Nucleic Acids Res.* **2023**, *51*, 9629–9642.

[8] a) P. A. Rice, S.-w. Yang, K. Mizuuchi, H. A. Nash, *Cell* **1996**, *87*, 1295–1306; b) S. C. Schultz, G. C. Shields, T. A. Steitz, *Science* **1991**, *253*, 1001–1007; c) S. C. Schultz, G. C. Shields, T. A. Steitz, *J. Mol. Biol.* **1990**, *213*, 159–166; d) P. H. Winegar, O. G. Hayes, J. R. McMillan, C. A. Figg, P. J. Focia, C. A. Mirkin, *Chem* **2020**, *6*, 1007–1017; e) A. Shamim, N. Parveen, V. K. Subramani, K. K. Kim, *Crystals* **2020**, *10*, 1093.

[9] D. Deepak, V. Corvaglia, J. Wu, L. Allmendinger, I. Huc, *Chemarxiv.*, DOI: 10.26434/chemrxiv-2023-9wnb8. This content is a preprint and has not been peer-reviewed.

[10] a) A. M. Kendhale, L. Poniman, Z. Dong, K. Laxmi-Reddy, B. Kauffmann, Y. Ferrand, I. Huc, *J. Org. Chem.* **2011**, *76*, 195–200; b) S. J. Dawson, Á. Mészáros, L. Pethő, C. Colombo, M. Csékei, A. Kotschy, I. Huc, *Eur. J. Org. Chem.* **2014**, *2014*, 4265–4275; c) L. Zheng, D. Zheng, Y. Wang, C. Yu, K. Zhang, H. Jiang, *Org. Biomol. Chem.* **2019**, *17*, 9573–9577; d) L. Yang, C. Ma, B. Kauffmann, D. Li, Q. Gan, *Org. Biomol. Chem.* **2020**, *18*, 6643–6650; e) Z. Liu, X. Hu, A. M. Abramyan, Á. Mészáros, M. Csékei, A. Kotschy, I. Huc, V. Pophristic, *Chem. Eur. J.* **2017**, *23*, 3605–3615.

[11] D. Bindl, E. Heinemann, P. K. Mandal, I. Huc, *Chem. Commun.* **2021**, *57*, 5662–5665.

[12] a) B. Baptiste, C. Douat-Casassus, K. Laxmi-Reddy, F. Godde, I. Huc, *J. Org. Chem.* **2010**, *75*, 7175–7185; b) V. Corvaglia, F. Sanchez, F. S. Menke, C. Douat, I. Huc, *Chem. Eur. J.* **2023**, *29*, e202300898.

[13] C. Dolain, H. Jiang, J.-M. Léger, P. Guionneau, I. Huc, *J. Am. Chem. Soc.* **2005**, *127*, 12943–12951.

[14] The major and minor grooves of the DNA mimic foldamers can be distinguished from the fact that $Q^{Pho}M^{Pho}$ contains a methylene bridge whereas $M^{Pho}Q^{Pho}$ does not. The distance, at the surface of the mimic, from the side chain of Q^{Pho} to the side chain of the M^{Pho} forward in the sequence is a little longer than the distance from the side chain of M^{Pho} to the side chain of Q^{Pho} forward in the sequence. However, groove widths only differ by the contribution of this methylene group.
[15] *Maestro*, *Schrödinger*, LLC, New York, NY, **2021**.
[16] M. K. Ganapathiraju, S. Subramanian, S. Chaparala, K. B. Karunakaran, *Hum. Genome Var.* **2020**, *7*, 40.
[17] M. Giel-Pietraszuk, M. Hoffmann, S. Dolecka, J. Rychlewski, J. Barciszewski, *J. Protein Chem.* **2003**, *22*, 109–113.
[18] a) Y. Fujii, T. Shimizu, T. Toda, M. Yanagida, T. Hakoshima, *Nat. Struct. Biol.* **2000**, *7*, 889–893; b) W. Keller, P. König, T. J. Richmond, *J. Mol. Biol.* **1995**, *254*, 657–667.
[19] a) N. C. Horton, J. J. Perona, *Biochemistry* **2004**, *43*, 6841–6857; b) M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, A. K. Aggarwal, *Nature* **1994**, *368*, 660–664.
[20] a) C. Park, J. L. Campbell, W. A. Goddard, 3rd, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4892–4896; b) C. Y. Chen, T. P. Ko, T. W. Lin, C. C. Chou, C. J. Chen, A. H. Wang, *Nucleic Acids Res.* **2005**, *33*, 430–438; c) Z. Zhang, Y. Gong, L. Guo, T. Jiang, L. Huang, *Mol. Microbiol.* **2010**, *76*, 749–759.
[21] I. Huc, *Eur. J. Org. Chem.* **2004**, *2004*, 17–29.
[22] N. Delsuc, L. Poniman, J.-M. Léger, I. Huc, *Tetrahedron* **2012**, *68*, 4464–4469.
[23] a) H. Robinson, Y. G. Gao, B. S. McCrary, S. P. Edmondson, J. W. Shriver, A. H. Wang, *Nature* **1998**, *392*, 202–205; b) K. K. Swinger, K. M. Lemberg, Y. Zhang, P. A. Rice, *EMBO J.* **2003**, *22*, 3749–3760; c) R. Sánchez-Giraldo, F. J. Acosta-Reyes, C. S. Malarkey, N. Saperas, M. E. Churchill, J. L. Campos, *Acta Crystallogr. Sect. D* **2015**, *71*, 1423–1432.
[24] a) J. Fremaux, L. Fischer, T. Arbogast, B. Kauffmann, G. Guichard, *Angew. Chem. Int. Ed.* **2011**, *50*, 11382–11385; *Angew. Chem.* **2011**, *123*, 11584–11587; b) R. David, R. Günther, L. Baumann, T. Lühmann, D. Seebach, H.-J. Hofmann, A. G. Beck-Sickinger, *J. Am. Chem. Soc.* **2008**, *130*, 15311–15317.
[25] a) J. Shang, Q. Gan, S. J. Dawson, F. Rosu, H. Jiang, Y. Ferrand, I. Huc, *Org. Lett.* **2014**, *16*, 4992–4995; b) B. Teng, P. K. Mandal, L. Allmendinger, C. Douat, Y. Ferrand, I. Huc, *Chem. Sci.* **2023**, *14*, 11251–11260; c) V. Koehler, G. Bruschera, E. Merlet, P. K. Mandal, E. Morvan, F. Rosu, C. Douat, L. Fischer, I. Huc, Y. Ferrand, *Angew. Chem. Int. Ed.* **2023**, *62*, e202311639; *Angew. Chem.* **2023**, *135*, e202311639.
[26] D. Bindl, P. K. Mandal, L. Allmendinger, I. Huc, *Angew. Chem. Int. Ed.* **2022**, *61*, e202116509; *Angew. Chem.* **2022**, *134*, e202116509.
[27] J. Buratto, C. Colombo, M. Stupfel, S. J. Dawson, C. Dolain, B. Langlois d'Estaintot, L. Fischer, T. Granier, M. Laguerre, B. Gallois, I. Huc, *Angew. Chem. Int. Ed.* **2014**, *53*, 883–887; *Angew. Chem.* **2014**, *126*, 902–906.
[28] X. Hu, S. J. Dawson, Y. Nagaoka, A. Tanatani, I. Huc, *J. Org. Chem.* **2016**, *81*, 1137–1150.
[29] D. Nurizzo, T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J.-C. Biasci, S. McSweeney, E. Mitchell, *J. Synchrotron Radiat.* **2006**, *13*, 227–238.
[30] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, G. Bricogne, *Acta Crystallogr. Sect. D* **2011**, *67*, 293–302.
[31] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
[32] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr. Sect. D* **2019**, *75*, 861–877.
[33] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501.

Manuscript received: November 3, 2023
Accepted manuscript online: January 9, 2024
Version of record online: January 30, 2024

5.2. Supplementary Information

For:

Enhancing the features of DNA mimic foldamers for structural investigations

V. Corvaglia, J. Wu, D. Deepak, M. Loos and I. Huc*.

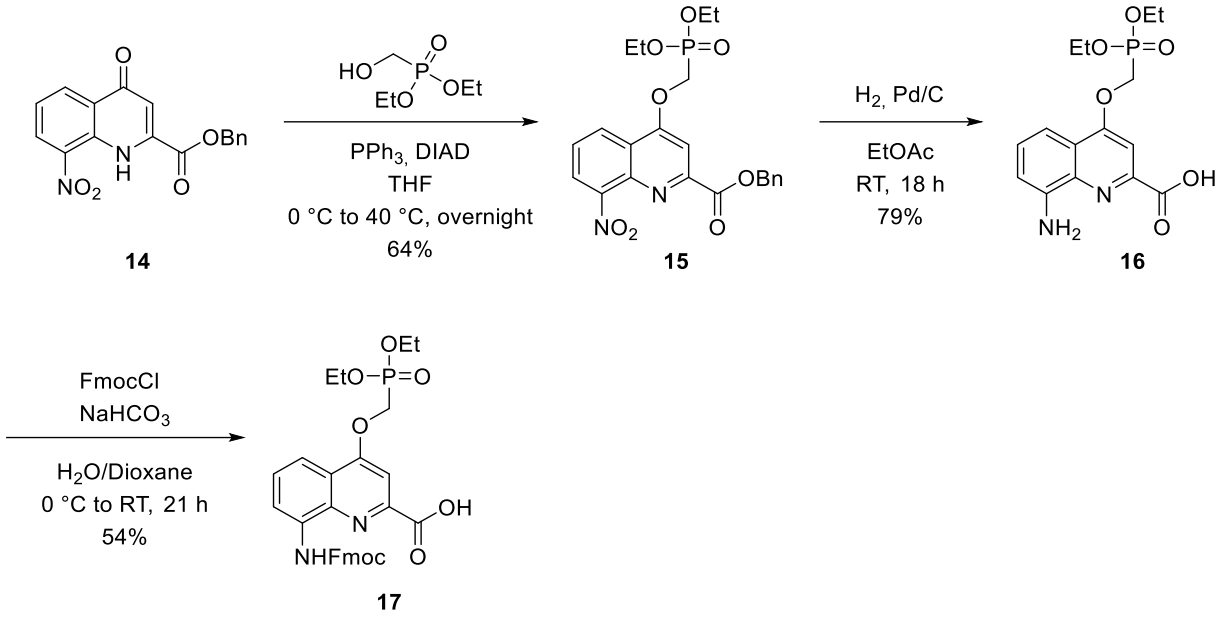
Constents

1. List of Abbreviations

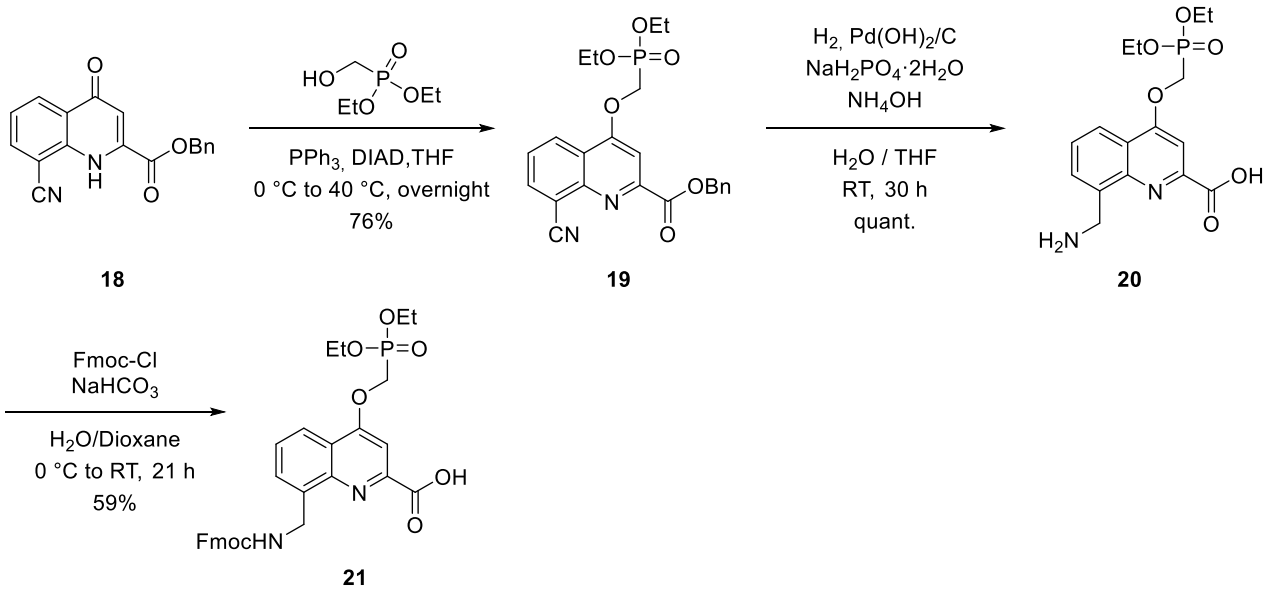
AcOH	acetic acid
Ac_2O	acetic anhydride
CD	circular dichroism
DCM	dichloromethane
DIAD	diisopropyl azodicarboxylate
DIPEA	*N,N*-diisopropylethylamine
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
ESI	electrospray ionization
EtOAc	ethyl acetate
Et_2O	diethyl ether
Fmoc-Cl	fluorenylmethyloxycarbonyl chloride
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
LC-MS	liquid chromatography-mass spectrometry
MeOH	methanol
NMR	nuclear magnetic resonance
PPh_3	triphenylphosphine
PyBOP	benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
RP	reversed phase
RT	room temperature
SPS	solid phase synthesis
TCAN	trichloroacetonitrile
TEAA	triethylammonium acetate
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TMSBr	trimethylbromosilane
UV	ultraviolet

2. Supplementary Schemes and Figures

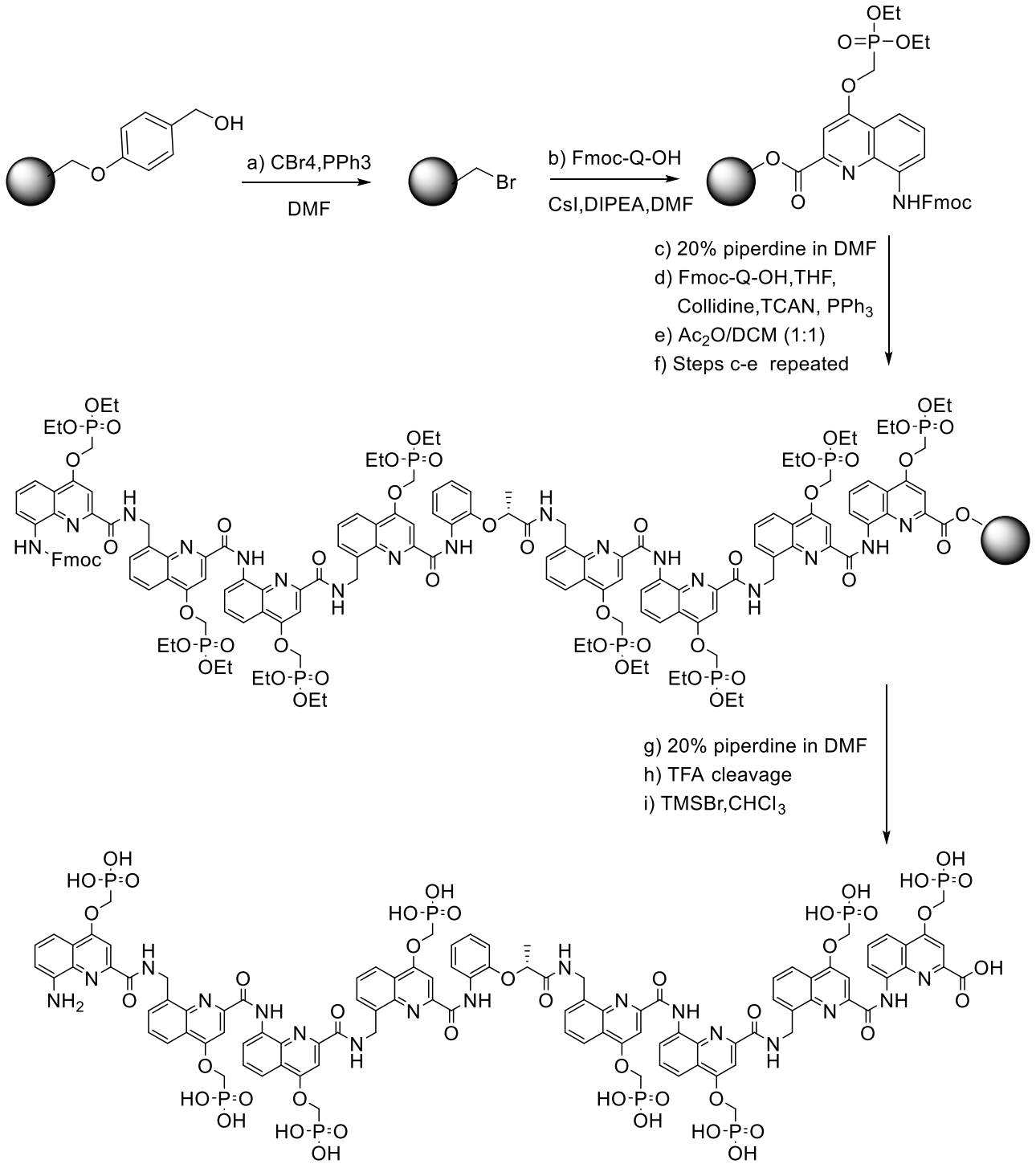
2.1 Supplementary Schemes



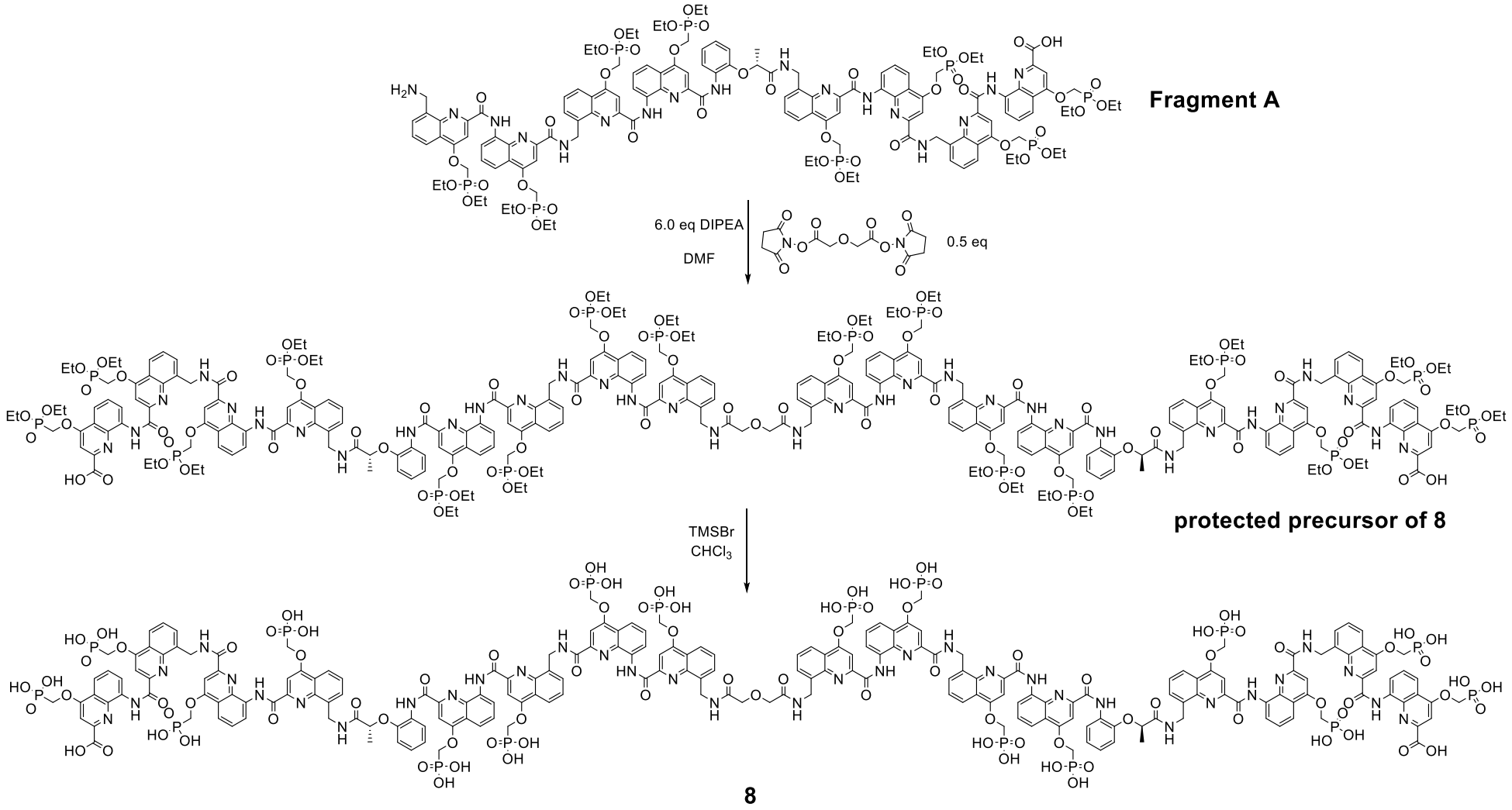
Scheme S1. Synthetic route to the Fmoc-Q^{Pho}-COOH monomer.



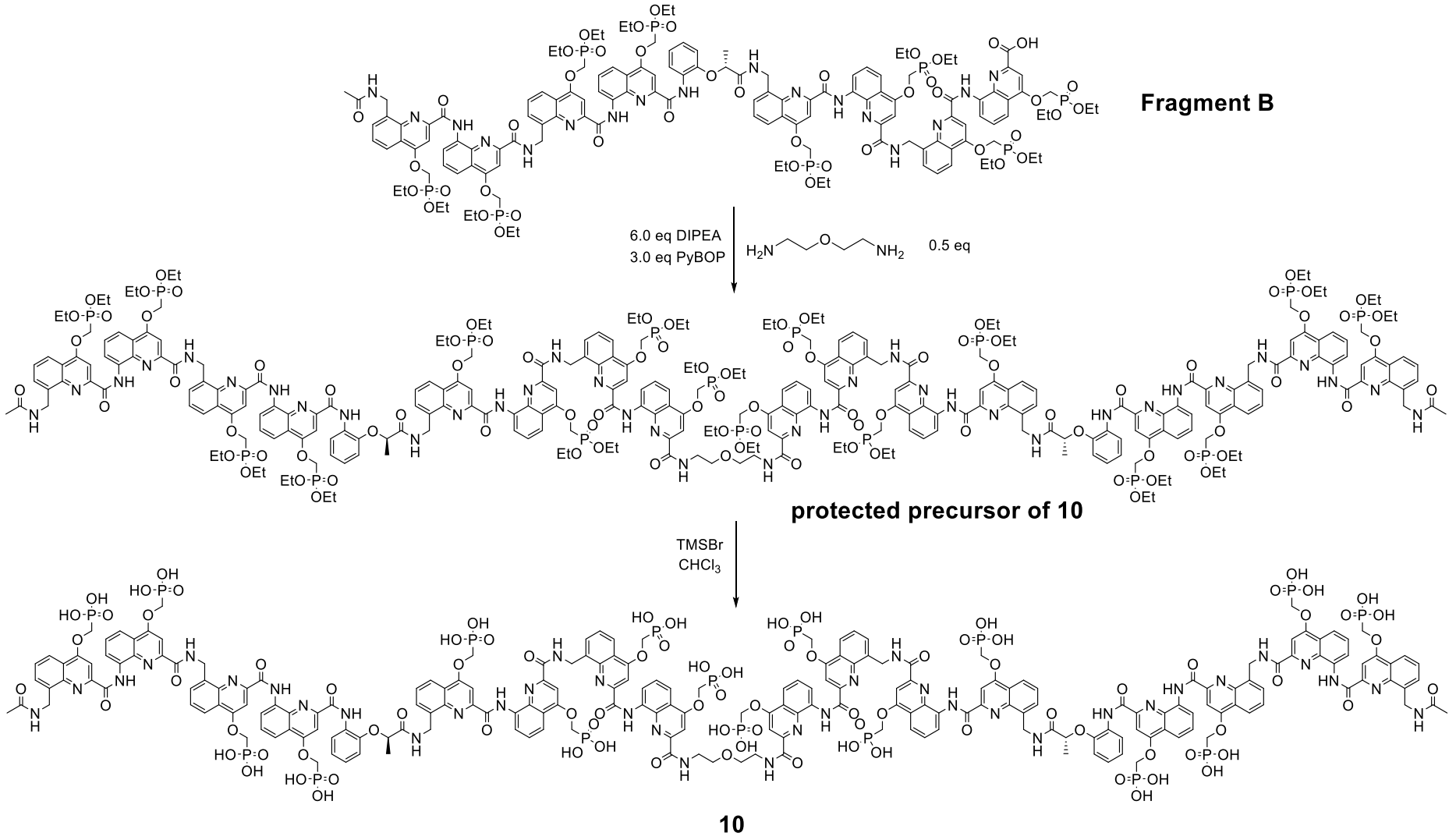
Scheme S2. Synthetic route to the Fmoc-M^{Pho}-COOH monomer.



Scheme S3. Representative example of SPS for oligomer **2**.



Scheme S4. Synthetic scheme for the N-terminus ligation to obtain oligomer **8**. A similar scheme leads to **9**.



Scheme S5. Synthetic scheme for the C-terminus ligation to obtain oligomer **10**. A similar scheme leads to **11**.

2.2 Supplementary Figures

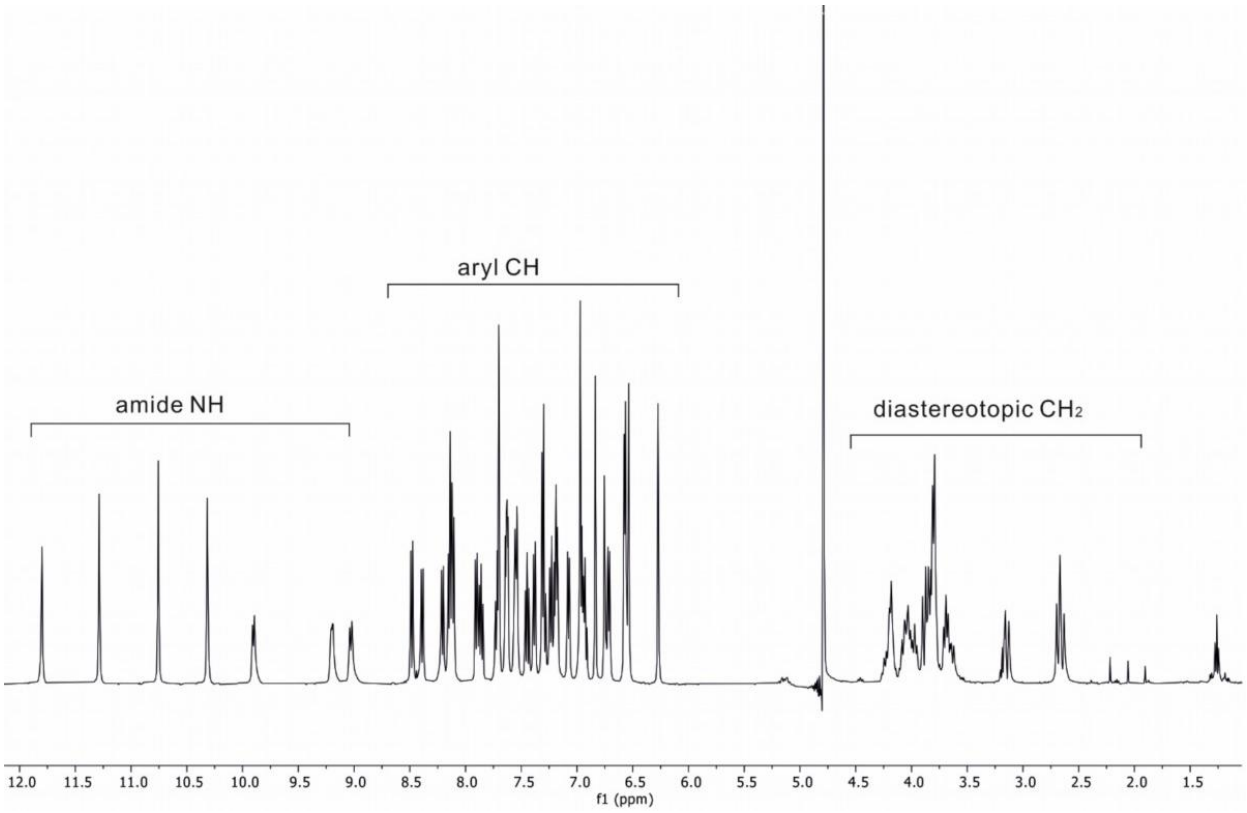


Figure S1. ^{1}H NMR spectrum of oligomer **1** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, H_2O suppression).

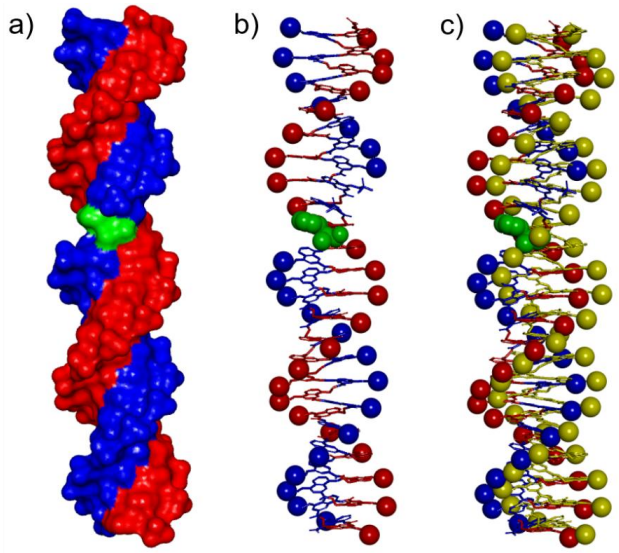


Figure S2. Energy minimized molecular models of extended-**2** as solvent accessible isosurface (a) and tube representation (b) except the phosphorous atoms which are shown in space-filling representation. The chiral B^{Rme} monomer is shown in bright green. The structure in (c) is an overlay

of (b) and the reference DNA mimic structure $(M^{pho}Q^{Pho})_{24}$ displayed in yellow. Hydrogen atoms have been omitted for clarity.

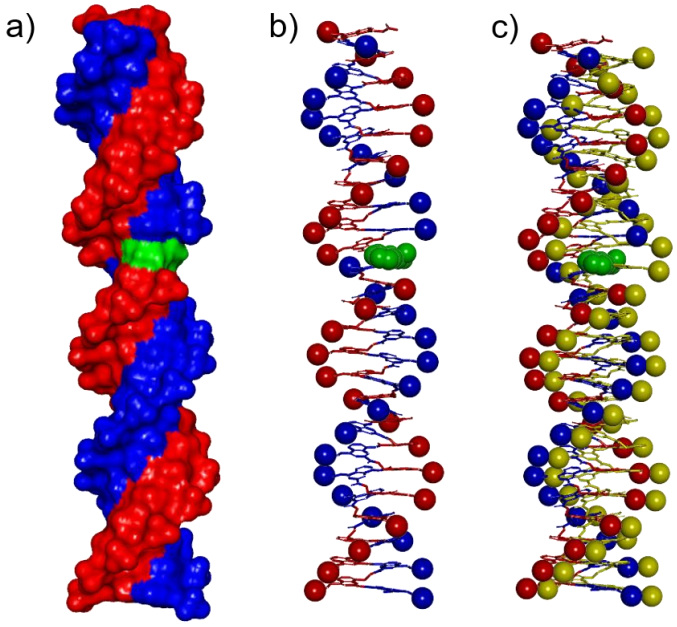


Figure S3. Energy minimized molecular models of extended-**5** as solvent accessible isosurface (a) and tube representation (b) except the phosphorous atoms which are shown in space-filling representation. The chiral B^{Rme} monomer is shown in bright green. The structure in (c) is an overlay of (b) and the reference DNA mimic structure $(M^{pho}Q^{Pho})_{24}$ displayed in yellow. Hydrogen atoms have been omitted for clarity.

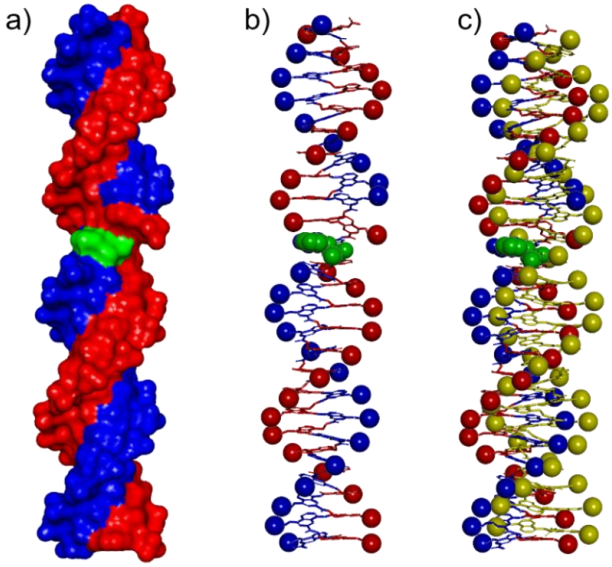


Figure S4. Energy minimized molecular models of extended-**7** as solvent accessible isosurface (a) and tube representation (b) except the phosphorous atoms which are shown in space-filling

representation. The double insertion of B^{Rme} and of M^{pho} creates minor distortions in the overall structure. The chiral B^{Rme} monomer is shown in bright green. The structure in (c) is an overlay of (b) and the reference DNA mimic structure $(M^{pho}Q^{Pho})_{24}$ displayed in yellow. Hydrogen atoms have been omitted for clarity.

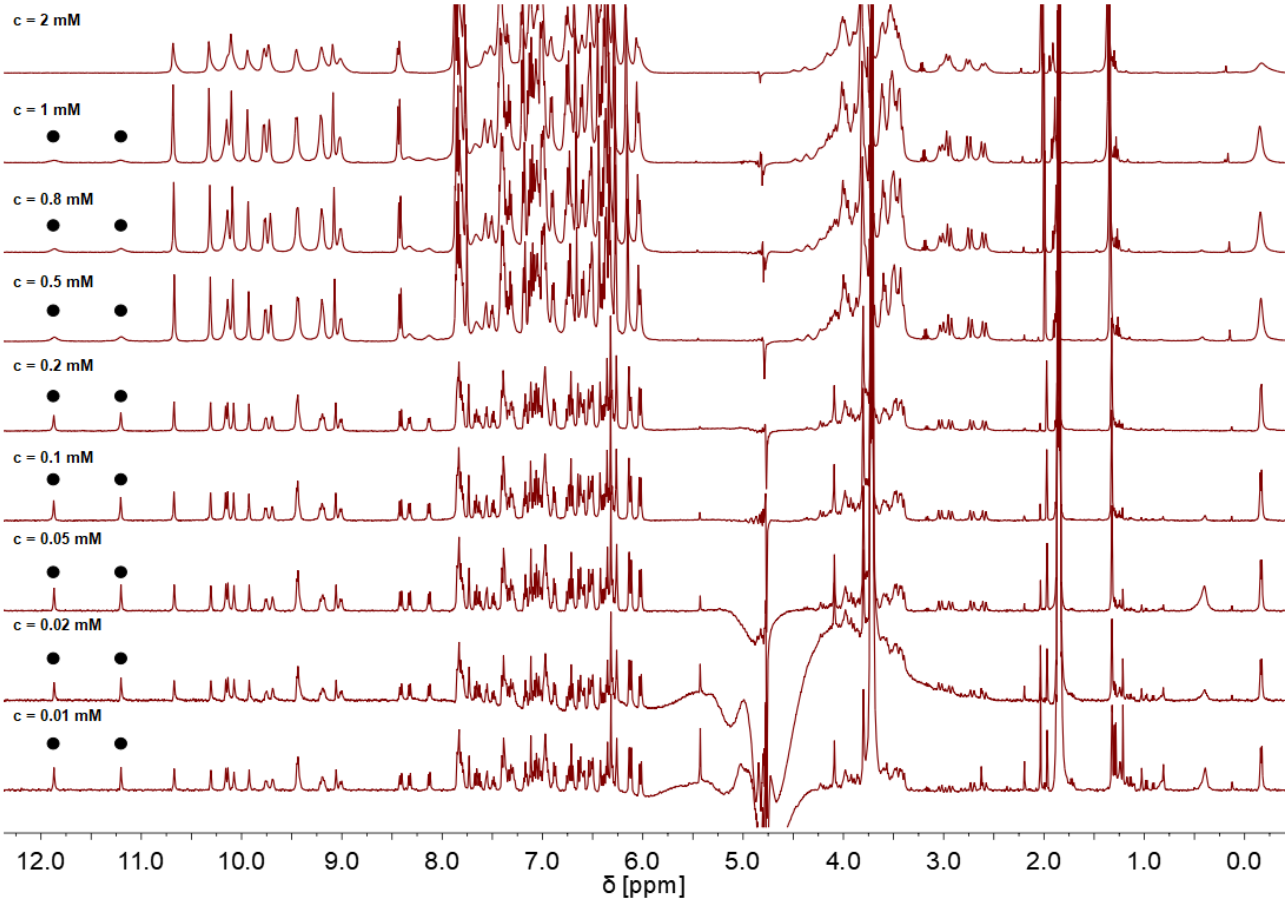


Figure S5. ^{1}H NMR dilution study of oligomer **12** (500 MHz, H_2O/D_2O 9:1 v/v, 27 mM sodium phosphate buffer, pH 7.2) from 2 mM to 0.01 mM. Black circles indicate the monomeric species forming upon dilution.

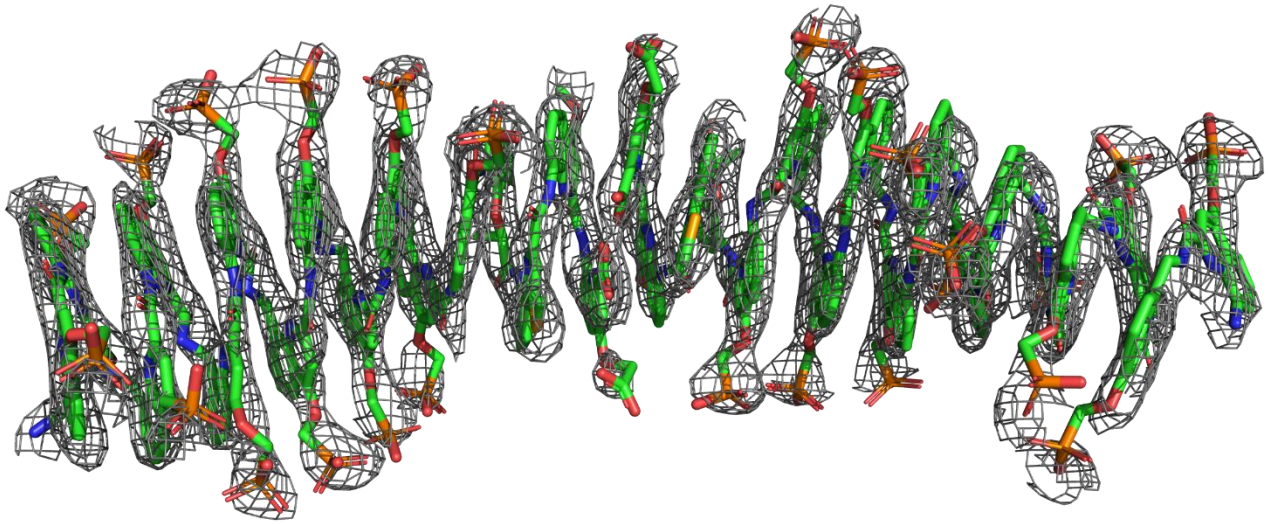


Figure S6. Sigma weighted $2F_o$-F_c electron density map (grey mesh) contoured at 1 σ superimposed on helices of **13**.

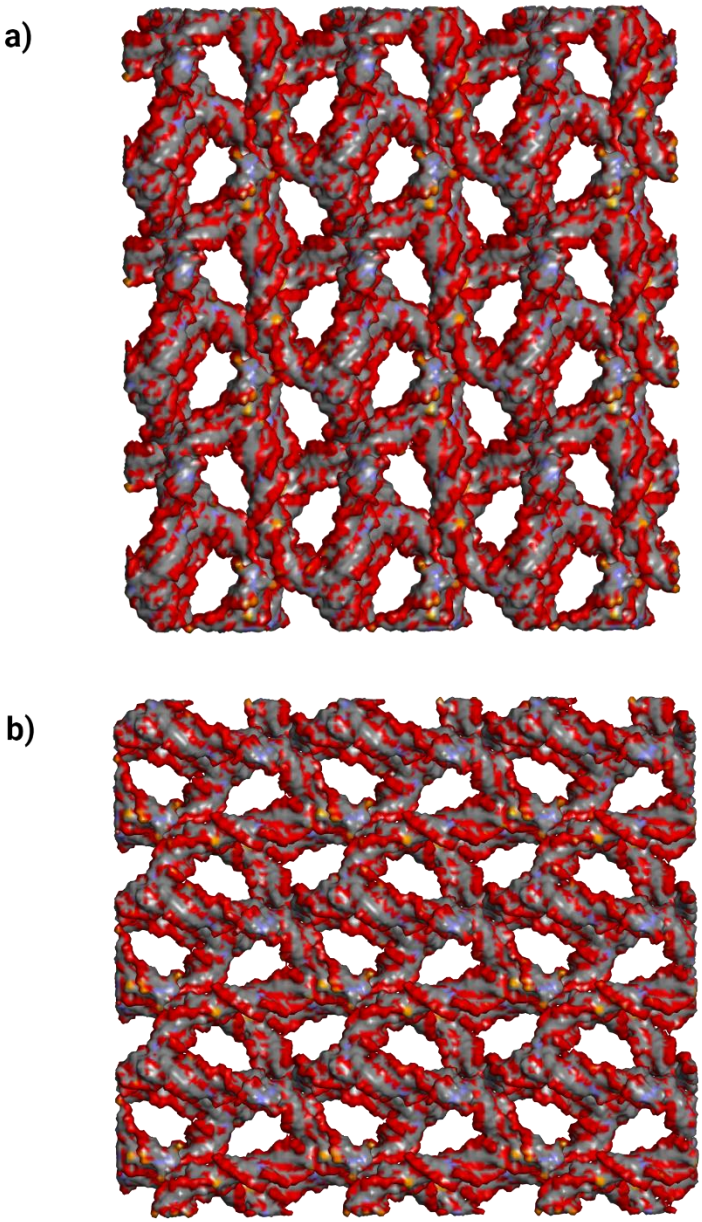


Figure S7. Packing of **13** (PDB # 8QHM) in the crystal lattice viewed down the a) a-axis and b) b-axis. Helices are shown in surface representation.

3. Materials and Methods

3.1 General

Chemicals and reagents were used as commercially supplied without any further purification unless otherwise stated. Low loading Wang resin (0.41 mmol g^{-1}) was purchased from Novabiochem. Analytical grade organic solvents were used for SPS. Anhydrous THF and DCM for SPS were dispensed from an *MBRAUN Solvent Purification System-800* solvent purification system. Reactions requiring anhydrous conditions were performed under nitrogen. Protected Fmoc-acid building blocks are shown in Figure S8.

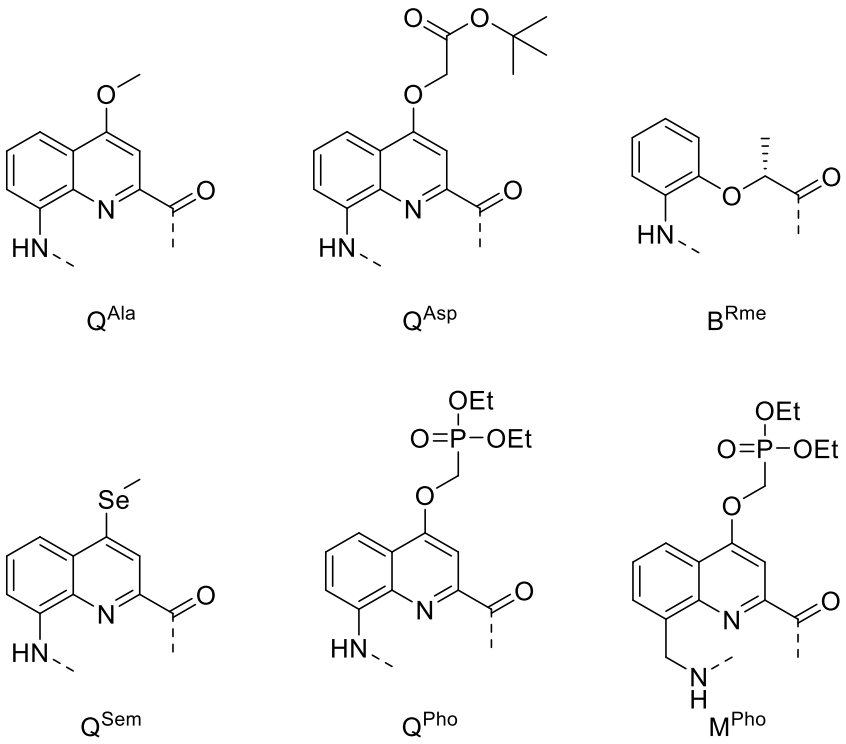


Figure S8. Side chain-protected Fmoc-acid building blocks used in this study. Fmoc-Q^{Ala}-OH,[27] Fmoc-Q^{Asp}-OH[12a] and Fmoc-B^{Rme}-OH[11] have been described previously. Fmoc-Q^{Sem}-OH will be described elsewhere. For a detailed procedure to Fmoc-Q^{Pho}-OH and Fmoc-M^{Pho}-OH, see section 3.2.

Analytical and semi-preparative RP-HPLC was performed on a Thermo Fisher Scientific Ultimate 3000 HPLC System using Macherey-Nagel Nucleodur C18 Gravity columns (4 × 100 mm, 5 µm and 10 × 250 mm, 5 µm) and Macherey-Nagel Nucleodur C8 Gravity columns (4 × 50 mm, 5 µm and 10 × 100 mm, 5 µm). When using acidic conditions, 0.1% TFA was added to aqueous mobile phase (referred to as mobile phase A) and to acetonitrile (referred to as mobile phase B). When using basic conditions, the mobile phase was composed of 12.5 mM TEAA in water at pH 8.5 (solvent A) and 12.5 mM TEAA in water: acetonitrile 1:2 vol/vol at pH 8.5 (solvent B). For RP-HPLC analyses, a

flow rate of 1.0 mL/min was applied; semi-preparative RP-HPLC purification were performed at a flow rate of 5.0 mL/min. UV absorbance was monitored at 300 nm if not stated otherwise.

NMR spectra were recorded on the *Avance III HD 500 MHz Bruker Biospin* spectrometer. DMSO-d_6 (δH: 2.50, δC: 39.4) and D_2O (δH: 4.79) were used as solvents. Water suppression was performed with excitation sculpting. Measurements were performed at 298 K unless stated otherwise. NMR spectra of the oligomers were recorded in H_2O/D_2O (9:1 v/v), 50 mM NH_4HCO_3. The raw data were evaluated using Mnova version 14.0.0 from Mestrelab Research. Signal multiplicities are abbreviated as s, singlet; d, doublet; t, triplet; q, quartet, and m, multiplet.

LC-MS spectra were recorded on a *Thermo Scientific Dionex UltiMate 3000* equipped with a Nucleodur C18 gravity column (2 x 50 mm, 1.8 µm) with a flow of 0.3 mL min^{-1}. 0.1% of formic acid in water (solvent A) and 0.1% of formic acid in acetonitrile (solvent B) were used as mobile phase for the ionization of the quinoline monomers. For water soluble oligomers, 12.5 mM aqueous NH_4OAc buffer adjusted to pH 8.5 and LC-MS grade acetonitrile were used as mobile phase for the ionization of the polyanionic foldamers. Elution was monitored by UV detection at 214, 254 and 300 nm with a diode array detector. The LC system was coupled to a *micrOTOF II* mass spectrometer by *Bruker Daltonics* and molecules were ionized by ESI.

CD spectra were recorded on a *J-815 Circular Dichroism spectrometer* by *Jasco* using quartz cells (2 mm optical path length). Scans were measured at 20 °C, over a wavelength range of 300–500 nm, with a response time of 0.5 s and a scanning speed of 50 nm/min. Molar extinction values were normalized per quinoline units. Foldamers (60 µM) were dissolved in 50 mM NH_4HCO_3 buffer pH 8.5.

3.2 Monomer synthesis procedures

Compound (15). Benzyl 8-nitro-4-quinolinone-2-carboxylate intermediate **14** was prepared according to previously described methods.[1a] Freshly dried compound **14** (12.0 g, 37.0 mmol, 1 eq.), PPh_3 (12.6 g, 48.1 mmol, 1.3 eq.) and diethyl hydroxymethyl phosphonate (6.84 g, 6 mL, 40.7 mmol, 1.1 eq.) were suspended in anhydrous THF under N_2 and cooled to 0 °C. DIAD (9.44 mL, 48.1 mmol, 1.3 eq.) was added dropwise over 20 min at 0 °C. The resulting solution was stirred at 0 °C for 1.5 h, at RT for another 1 h, and at 50 °C overnight. THF was removed *in vacuo* and co-evaporated with DCM (2 x 100 mL) and Et_2O (2 x 100 mL). The resulting solid was dried overnight under vacuum to remove residual THF. The crude product was purified by flash column chromatography (100% EtOAc, dry-load using silica gel). After evaporation of the solvent, the purified product was dissolved in DCM (5 mL), Et_2O (100 mL) was layered on top

and the product was crystallized at –14 °C overnight, filtered and washed with cold Et_2O (200 mL). Crystallization was repeated twice, until the HPLC analysis of an aliquot showed no residual triphenylphosphine oxide, to yield the title compound (11.2 g, 64%) as a white crystalline solid. **^{1}H NMR (500 MHz, $CDCl_3$):** δ [ppm] = 8.48 (dd, *J* = 7.5, 1.4 Hz, 1H), 8.13 (dd, *J* = 7.5, 1.4 Hz, 1H), 7.74 – 7.66 (m, 2H), 7.55 – 7.50 (m, 2H), 7.43 – 7.38 (m, 2H), 7.38 – 7.32 (m, 1H), 5.49 (s, 2H), 4.57 (d, *J* = 10.2 Hz, 2H), 4.28 (dq, *J* = 8.4, 7.0 Hz, 4H), 1.38 (t, *J* = 7.0 Hz, 6H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ [ppm] = 164.8, 162.4 (d, *J* = 13.2 Hz), 151.5, 148.7, 140.3, 135.5, 128.8, 128.5, 128.3, 126.7, 126.2, 125.4, 123.0, 102.4, 68.1, 63.4(d, *J* = 6.4 Hz), 63.0 (d, *J* = 171.5 Hz), 16.7(d, *J* = 5.6 Hz). **HRMS** (ESI^+) *m/z* calcd. for $C_{22}H_{23}N_2O_8P$: 475.1265 $[M+H]^+$; found: 475.1319.

Compound (17). Compound **15** (11.2 g, 23.4 mmol, 1 eq.) was dissolved in EtOAc (350 mL), split over two flasks under N_2 and the solvent was degassed for 15 min by bubbling N_2 through the solution while sonicating. Pd/C (1.30 g) was added and the solution was further degassed for 10 min. Under vigorous stirring, a H_2 balloon was placed on top of the round bottom flask and the reaction mixture was stirred for 18 h until full conversion. The catalyst was filtered with a paper filter and the residual solid washed with hot THF. The filtrate was removed *in vacuo,* the resulting solid suspended in water/acetonitrile (30 mL, 1:1 v/v), sonicated, and freeze-dried to give **16** (7.22 g, 79%) as a green solid that was used without further purification. Compound **16** (7.22 g, 18.5 mmol, 1.0 eq) was suspended in dioxane (350 mL) and $NaHCO_3$ (36.0 g, 426 mmol, 21 eq.) dissolved in water (370 mL) was added. The suspension was cooled to 0 °C and Fmoc-Cl (5.31 g, 20.5 mmol, 1.1 eq.) in dioxane (200 mL) was added dropwise over 1 h at 0 °C. The reaction mixture was stirred at 0 °C for 1 h and then stirred at RT for 21 h. The reaction mixture was concentrated by rotary evaporation to remove most of the dioxane and the resulting suspension was diluted with water to a volume of 800 mL. The reaction mixture was acidified to pH 3 with a saturated $KHSO_4$ solution, extracted with DCM (1 L, then 2 x 500 mL), and dried over Na_2SO_4. After filtration, the solvent was removed *in vacuo*, and the resulting solid was dried under vacuum to remove residual dioxane. The crude product was precipitated from acetonitrile (50 mL), sonicated shortly to allow full precipitation, filtered, and washed with cold acetonitrile (−14 °C). The resulting solid was purified by flash column chromatography (5% MeOH in DCM with 0.1% AcOH, dry-load with silica from DCM) and the combined fractions were concentrated and washed with water (3 x 500 mL). After removal of the solvent *in vacuo*, the resulting solid was precipitated from acetonitrile, washed with cold acetonitrile (−14 °C), and lyophilized to yield the title compound (6.25 g, 58%) as a pale yellow solid. **^{1}H NMR (500 MHz, DMSO-*d*$_6$):** δ [ppm] = 13.56 (s, 1H), 10.46 (s, 1H), 8.37 (s, 1H), 7.94 (d, *J* = 7.5, 2H),

7.84 (s, 1H), 7.82 – 7.76 (m, 2H), 7.76 (d, *J* = 1.3 Hz, 1H), 7.69 – 7.65 (m, 1H), 7.45 (td, *J* = 7.7, 1.0 Hz, 2H), 7.37 (td, *J* = 7.5, 1.2 Hz, 2H), 4.93 (d, *J* = 9.7 Hz, 2H), 4.63 (d, *J* = 6.8 Hz, 2H), 4.46 (t, *J* = 6.8 Hz, 1H), 4.19 (d, *J* = 8.4, 7.0 Hz, 4H), 1.28 (t, *J* = 7.0 Hz, 6H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ [ppm] = 165.3, 162.4 (d, *J* = 13.2 Hz), 153.5, 146.7, 143.7, 140.8, 137.5, 135.8, 128.9, 127.8, 127.2, 125.2, 121.6, 120.3, 116.6, 114.2, 101.3, 66.4, 62.5 (d, *J* = 5.9 Hz), 62.1(d, *J* = 164.3 Hz), 46.6, 16.3 (d, *J* = 5.4 Hz). **HRMS** (ESI^+) *m/z* calcd. for $C_{30}H_{29}N_2O_8P$: 577.1734 $[M+H]^+$; found: 577.1949.

Compound (20). Compound **19**[1a] (4.35 g, 9.61 mmol, 1.0 eq.) was dissolved in THF (360 mL) under N_2 and $NaH_2PO_4 \cdot 2H_2O$ (15.0 g, 96.1 mmol,10 eq.) dissolved in water (130 mL) was added. The solvent was degassed for 15 min by bubbling N_2 through the solution while sonicating, $Pd(OH)_2/C$ (0.41 g) and NH_4OH (20%, 5.49 mL, 28.8 mmol, 3.0 eq.) were added and the solution was further degassed for 10 min. Under vigorous stirring, a H_2 balloon was placed on top of the round bottom flask and the reaction mixture was stirred for 30 h at RT under H_2 until the reaction was complete. The catalyst was filtered with a paper filter and the residual solid washed with THF. The filtrate was removed *in vacuo* and the crude product was used in the next step without further purification (HPLC purity in acidic condition was 93%).

Compound (21): Compound **20** (7.07 g, 17.8 mmol, 1.0 eq.) was suspended in dioxane (350 mL) and 10% $NaHCO_3$ in water (340 mL) was added. The suspension was cooled to 0 °C and Fmoc-Cl (5.54 g, 21.4 mmol, 1.2 eq.) in dioxane (150 mL) was added dropwise over 1 h at 0 °C. The reaction mixture was stirred at 0 °C for 2 h and then stirred at RT for 13 h. The solvents were removed *in vacuo* and the resulting solid was suspended in water (500 mL). The suspension was acidified to pH 3 with a saturated $KHSO_4$ solution, extracted with DCM (3 x 500 mL) and dried over Na_2SO_4. After filtration, the solvent was removed *in vacuo*, and the resulting solid was dried under vacuum to remove residual dioxane. The crude product was precipitated from acetonitrile (50 mL), sonicated to allow full precipitation, filtered and washed with cold acetonitrile (−14 °C). Precipitation from the filtrates and freeze-drying of the combined solids, yielded the title compound (6.18 g, 59%) as a white solid. **^{1}H NMR (500z MHz, DMSO-d_6):** δ [ppm] = 13.12 (s, 1H), 8.06 (d, *J* = 8.3 Hz, 1H), 7.93 – 7.86 (m, 3H), 7.76 (s, 1H), 7.73 – 7.66 (m, 3H), 7.62 (d, *J* = 6.9 Hz, 1H), 7.42 (t, *J* = 7.5 Hz, 2H), 7.31 (t, *J* = 7.4 Hz, 2H), 4.89 (d, *J* = 9.9 Hz, 2H), 4.84 (d, *J* = 6.1 Hz, 2H), 4.40 (d, *J* = 6.7 Hz, 2H), 4.26 (d, *J* = 6.7 Hz, 1H), 4.22 – 4.15 (m, 4H), 1.27 (t, *J* = 7.0 Hz, 6H). **^{13}C NMR (126 MHz, DMSO-d_6):** δ [ppm] = 165.3, 162.1 (*J* = 13.1Hz), 153.5, 146.7, 143.7, 140.8, 137.5, 135.8, 128.9, 127.8, 127.2, 125.2, 121.6, 120.3, 116.6, 114.2, 101.3, 65.3, 62.4 (d, *J* = 6.0 Hz), 61.9 (d, *J* = 164.1 Hz), 46.8, 40.6, 16.3 (d, *J* = 5.4 Hz).**HRMS** (ESI^+) *m/z* calcd. for $C_{31}H_{31}N_2O_8P$: 591.1891 $[M+H]^+$; found: 591.2121.

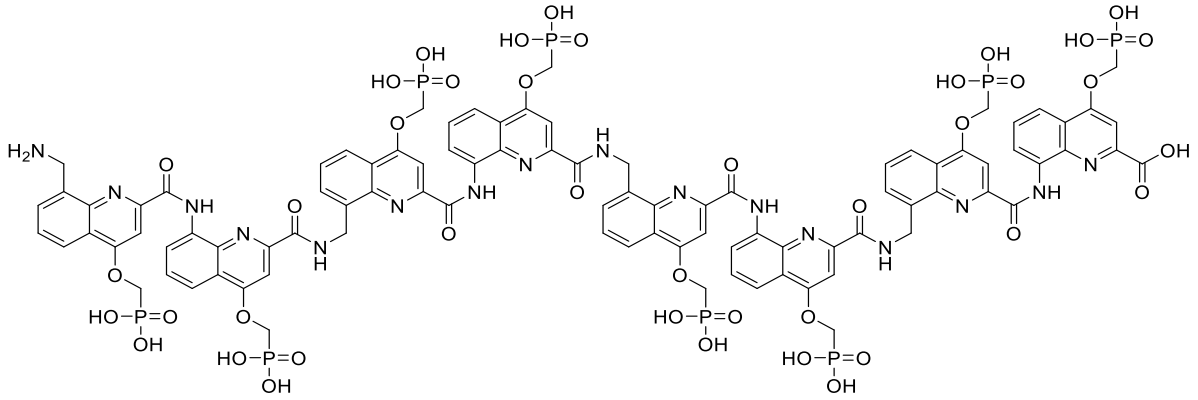
3.3 Oligomer synthesis procedures

Oligomers **2**–**7** were synthesized according to previously reported SPS protocols,[11,12a,28] see Scheme S3. Oligomers **8**–**9** and oligomers **10**–**11** were synthesized following Schemes S4 and Scheme S5, respectively. Oligomers **12**–**13** were synthesized by recently reported automated SPS procedures.[12b] Fmoc acid building blocks were activated *in situ* by generating the respective acid chlorides prior to coupling.

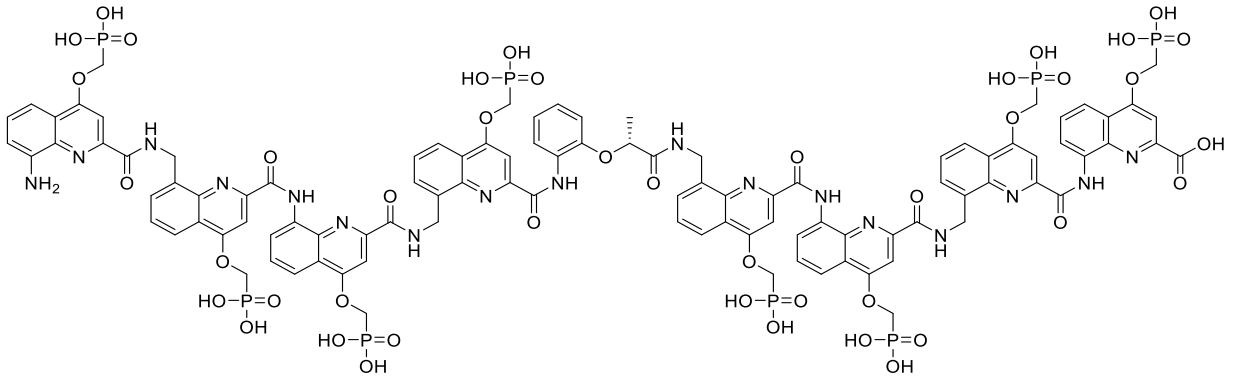
Acetylation: The resin (1.0 equiv.) was washed with DCM (3 x 3 mL) and incubated in Ac_2O/DCM (1:1 v:/v) for 10 min. Then, the resin was washed with DCM (2 x 3 mL) and DMF (3 x 3 mL).

Resin cleavage and Preparative HPLC purification: The resin-bound oligomer was placed in a syringe equipped with a filter, washed with DMF (3 x 3 mL), DCM (3 x 3 mL), and dried by passing N_2 flow through it. It was then suspended in a solution of TFA. The resin was next shaken for at least 2 h at RT and then filtered off and washed one time with TFA. The combined solvent was removed in vacuo. After precipitation in cold Et_2O, the crude oligomer with protecting groups was purified by semi prep RP-HPLC under acidic condition to give the oligomer as a yellow solid.

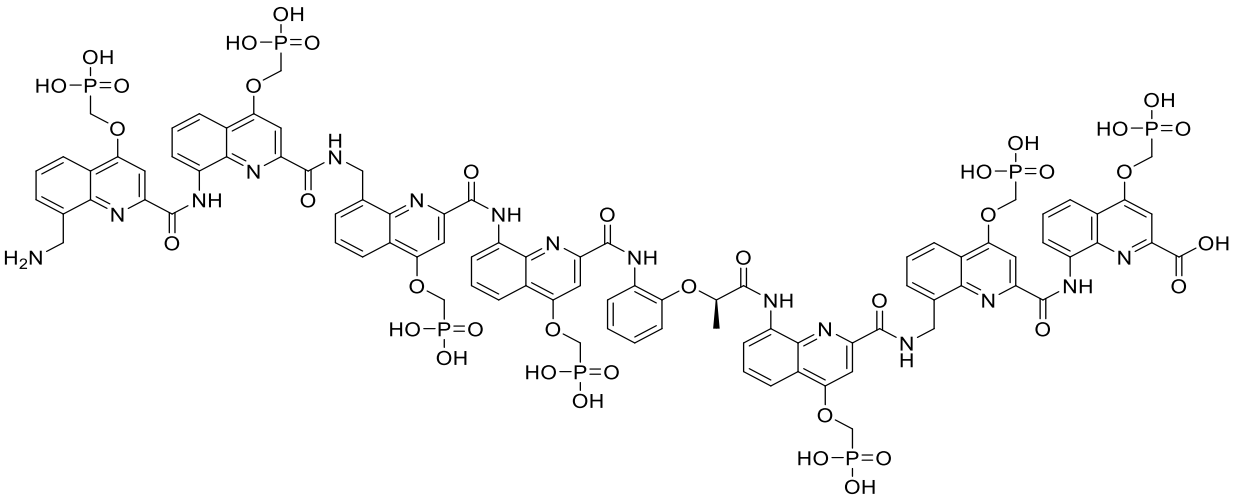
Synthesis of water-soluble oligomers: The previously purified oligomer was treated by TMSBr to remove the ethyl groups. Subsequently, the crude was purified by semi prep RP-HPLC under basic conditions (as described in section 3.1) to give the oligomer as a yellow solid. Following this, an ion exchange process was performed to obtain the side chains as water-soluble ammonium phosphonate salts. The removal of ethyl phosphonate protecting groups and ion exchange were performed as previously described.[1a]



Oligomer (1): The ^{1}H NMR spectrum of **1** in Figure S1 matched with that described.[1a]

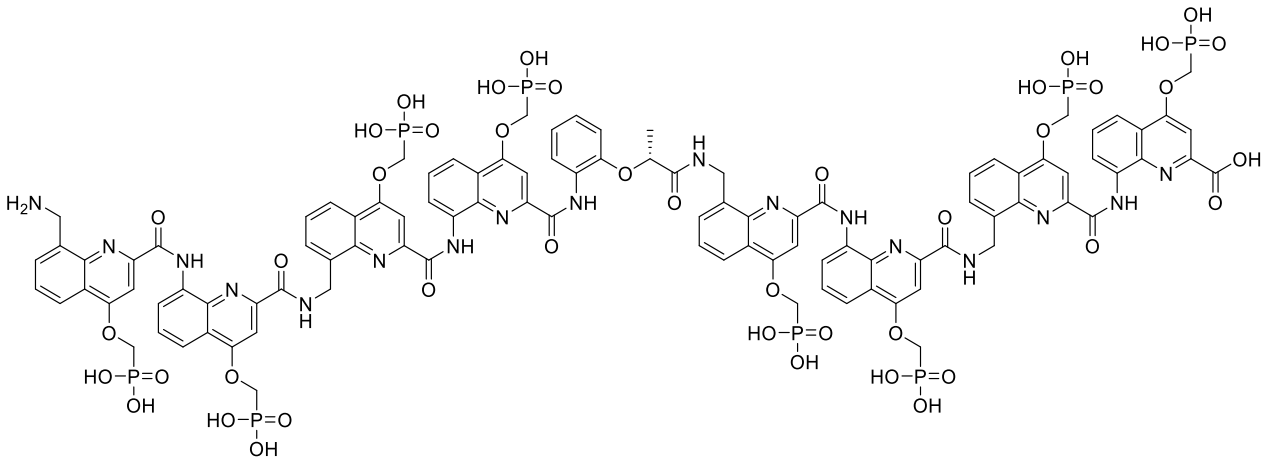


Oligomer (2): Oligomer **2** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following procedures reported previously.[11,12a,28] After ion exchange, the title compound was obtained as a light yellow solid (10 mg, 15%; HPLC-purity: >99%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.93 (s, 1H), 11.14 (s, 1H), 11.06 (s, 1H), 9.24 (s, 1H), 9.15 (s, 1H), 8.98 (d, *J* = 8.5 Hz, 1H), 8.86 (d, *J* = 11.9 Hz, 1H), 8.56 (dd, *J* = 15.9, 9.6 Hz, 1H); **aromatic CHs** δ [ppm] =8.39 (d, *J* = 9.4 Hz), 8.22 (d, *J* = 9.8 Hz), 8.18 – 8.12 (m), 8.02-7.92 (m), 7.79 (dd, *J* = 11.9, 8.7 Hz), 7.69 (d, *J* = 12.3 Hz), 7.58 (dd, *J* = 13.5, 8.8 Hz), 7.43-7.33 (m), 7.31 – 7.21 (m), 7.23 – 7.03 (m), 6.72 (t, *J* = 8.8 Hz), 6.67 (s), 6.60 (s), 6.53 (d, *J* = 6.1 Hz), 6.47 (s), 6.32 (s); **aliphatic CHs** δ [ppm] = 5.50 (s), 4.17 – 3.83 (m), 3.76 (t, *J* = 10.2 Hz), 3.71 (s), 3.29 (s), 3.10 – 2.88 (m), 2.00 (s), 1.32 – 1.12 (m), 0.21 (s). **HRMS** (ESI) *m/z* calcd. for: $C_{101}H_{91}N_{17}O_{43}P_8$: 1238.1623 $[M-2H]^{2-}$; found: 1238.2223.

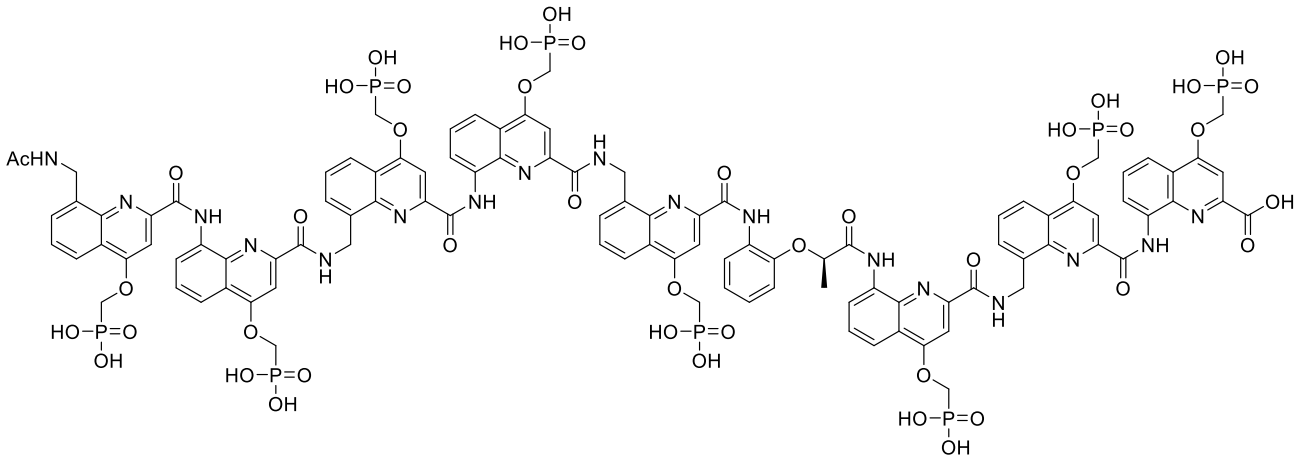


Oligomer (3): Oligomer **3** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following procedures reported previously.[11,12a,28] After ion exchange, the title compound was obtained as a light yellow solid (10 mg, 15%; HPLC-purity: >97%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.35 (s, 1H), 11.14 (s, 1H), 10.60 (s, 1H), 9.80 (d, *J* = 9.4 Hz, 2H), 9.67 (d, *J* = 9.4 Hz, 1H), 9.54 (s, 1H) , 8.46 (d, *J* = 9.2 Hz, 1H); **aromatic CHs** δ [ppm] = 8.26 – 8.18 (m), 8.07 (d, *J* = 8.6 Hz), 7.89 – 7.75 (m), 7.78 (d, *J* = 8.6 Hz), 7.67 (d, *J* = 9.2 Hz), 7.65 – 7.56 (m), 7.54 – 7.47 (m),

7.50 – 7.42 (m), 7.45 – 7.37 (m), 7.24 (t, *J* = 8.4 Hz), 7.17 (q, *J* = 9.0 Hz), 7.07 (s, 1H), 7.01 (t, *J* = 8.4 Hz), 6.98 – 6.92 (m), 6.84 (d, *J* = 8.0 Hz), 6.48 (d, *J* = 9.3 Hz), 6.36 (s), 6.08 (s); **aliphatic CHs** δ [ppm] = 4.25 – 4.06 (m), 3.96 (s), 3.94 – 3.79 (m), 2.63 - 2.59 (m), 2.22 (s), 1.35 – 1.22 (m), 0.17 (s). **HRMS** (ESI$^-$) *m/z* calcd. for: $C_{89}H_{80}N_{15}O_{38}P_7$: 1090.6403 $[M\text{-}2H]^{2-}$; found: 1090.6515.

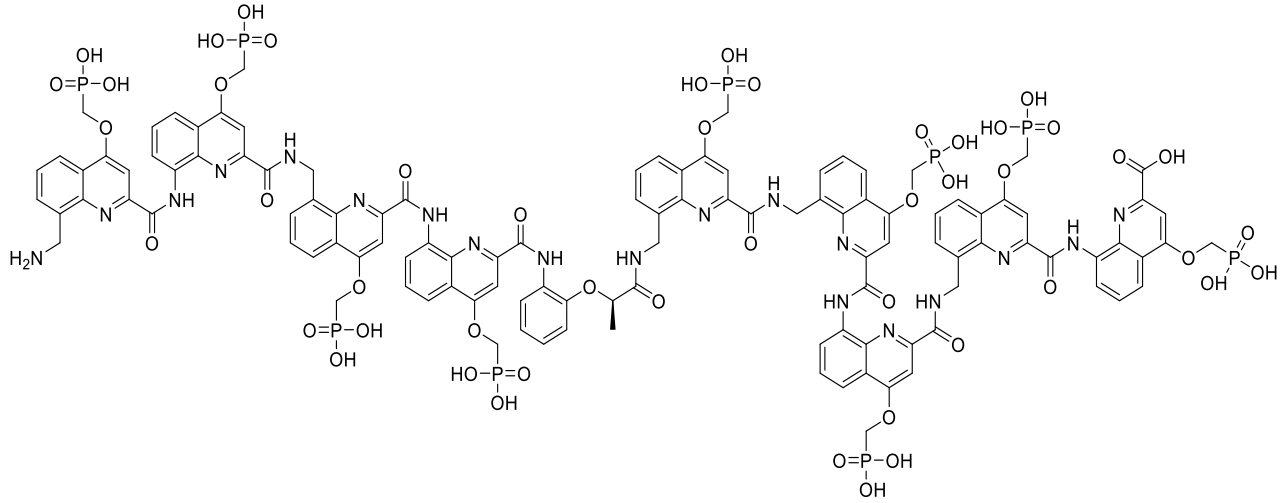


Oligomer (4): Oligomer **4** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following procedures reported previously.[11,12a,28] After ion exchange, the title compound was obtained as a light yellow solid (15 mg, 15%; HPLC-purity: >98%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1): amide NHs** δ [ppm] = 11.57 (s, 1H), 11.13 (s, 1H), 10.74 (s, 1H), 10.39 (s, 1H), 9.61 (s, 2H), 9.28 (s, 1H), 8.49 (d, *J* = 9.1 Hz, 1H); **aromatic CHs** δ [ppm] = 8.13 (d, *J* = 8.1 Hz), 8.05 – 8.00 (m), 7.97 (d, *J* = 8.9 Hz), 7.84 – 7.69 (m), 7.61 (d, *J* = 8.8 Hz), 7.58 (d, *J* = 6.1 Hz), 7.51 – 7.24 (m), 7.19 (s), 7.04 – 6.51 (m), 6.41 (s); **aliphatic CHs** δ [ppm] = 5.49 – 5.46 (m), 4.86 – 4.78 (m), 4.78 (s), 4.27 (s), 4.09 (d, *J* = 10.2 Hz), 3.98 (s), 3.86 (d, *J* = 14.3 Hz), 3.78 (d, *J* = 14.7 Hz), 3.75 – 3.67 (m), 3.63 – 3.55 (m), 3.32 – 3.24 (m), 3.12 – 2.59 (m), 2.21 (s), 2.05 (s), 1.33 – 1.21 (m), -0.40 (d, *J* = 6.9 Hz). **HRMS** (ESI$^-$) *m/z* calcd. for: $C_{101}H_{91}N_{17}O_{43}P_8$: 1238.1623 $[M\text{-}2H]^{2-}$; found: 1238.2073.

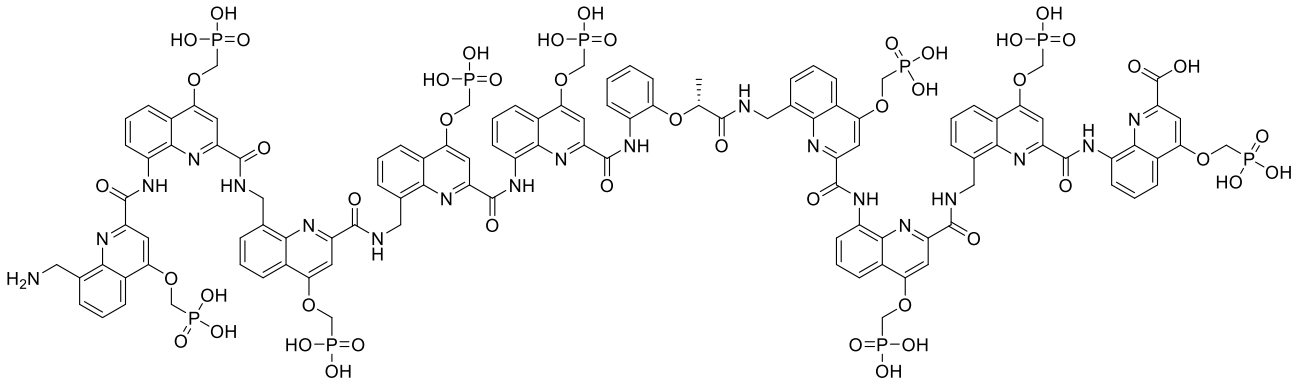


Oligomer (5): Oligomer **5** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following

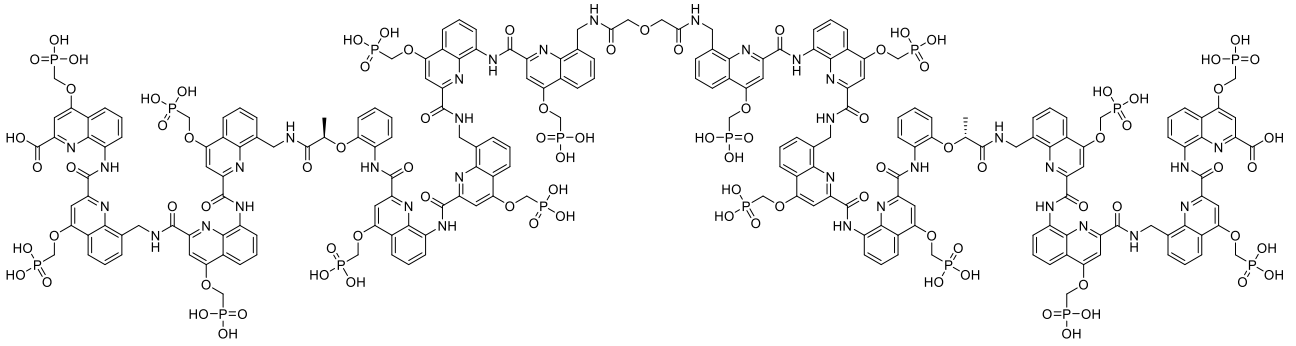
procedures reported previously.[11,12a,28] After ion exchange, the title compound was obtained as a light yellow solid (7 mg, 12%; HPLC-purity: >97%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.48 (s, 1H), 11.37 (s, 1H), 10.73 (s, 1H), 9.85 (s, 1H), 9.80 (d, *J* = 8.5 Hz, 1H), 9.73 (d, *J* = 9.3 Hz, 1H), 9.68 (d, *J* = 9.5 Hz, 1H), 9.04 (s, 1H), 8.40 (d, *J* = 9.2 Hz,1H); **aromatic CHs** δ [ppm] = 8.31 (d, *J* = 8.4 Hz), 8.11 (m), 8.02 (d, *J* = 8.6 Hz), 7.96 – 7.90 (m), 7.75 – 7.66 (m), 7.63 (d, *J* = 7.6 Hz), 7.51 – 7.44 (m), 7.37 – 7.23 (m), 7.17 – 7.06 (m), 6.96 (s), 6.83 – 6.79 (m), 6.77 (s), 6.69 – 6.64 (m), 6.57 – 6.52 (m), 6.31 – 6.27(m), 6.14 – 6.10 (m), 6.02 – 5.98 (m); **aliphatic CHs** δ [ppm] = 4.22 – 3.79 (m), 3.64 – 3.47 (m), 3.09 – 3.01 (m), 2.85 – 2.60 (m), 1.29 (s), 1.26 –1.22 (m), 0.30 (d, *J* = 6.6 Hz). **HRMS** (ESI$^-$) *m/z* calcd. for: $C_{103}H_{93}N_{17}O_{44}P_8$: 1259.1675 $[M-2H]^{2-}$; found: 1259.1859.



Oligomer (6): Oligomer **6** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following procedures reported previously.[11,12a,28] After ion exchange, the title compound was obtained as a light yellow solid (5 mg, 10%; HPLC-purity: >98%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.54 (s, 1H), 11.18 (s, 1H), 10.88 (s, 1H), 10.24 (s, 1H), 9.78 (t, *J* = 9.9 Hz, 2H), 9.24 (s, 1H), 8.84 (t, *J* = 4.8 Hz, 1H), 8.47 (d, *J* = 9.3 Hz, 1H); **aromatic CHs** δ [ppm] = 8.41 – 8.38 (m), 8.22 (d, *J* = 8.3 Hz), 8.17 (d, *J* = 9.0 Hz), 7.95 (d, *J* = 8.8 Hz), 7.75 (d, *J* = 8.5 Hz), 7.70 (d, *J* = 9.4 Hz), 7.59 – 7.26 (m), 7.24 (s), 7.13 – 6.75 (m), 6.65 (d, *J* = 5.5 Hz), 6.55 – 6.51 (m), 6.37 (d, *J* = 8.1 Hz); **aliphatic CHs** δ [ppm] = 5.78 (s), 5.58 (d, *J* = 9.4 Hz), 4.43 – 4.21 (m), 4.16 – 3.88 (m), 3.76 – 3.66 (m), 3.60 (s), 3.57 – 2.44 (m), 1.36 (s, 1H), 1.31 – 1.27 (m), -0.50 (d, *J* = 7.4 Hz). **HRMS** (ESI$^-$) *m/z* calcd. for: $C_{113}H_{102}N_{19}O_{48}P_9$: 1385.1825 $[M-2H]^{2-}$; found: 1385.2034.

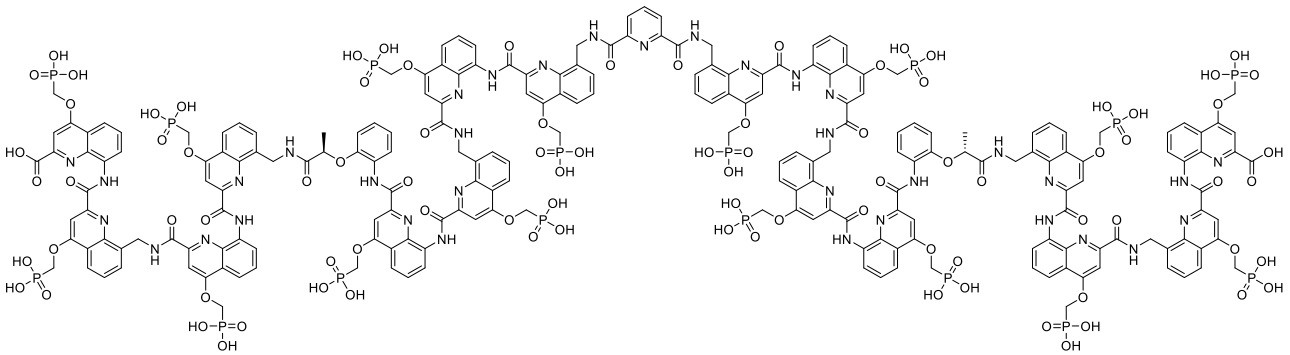


Oligomer (7): Oligomer **7** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following procedures reported previously.[11,12a,28] After ion exchange, the title compound was obtained as a light yellow solid (5 mg, 10%; HPLC-purity: >95%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.58 (s, 1H), 11.27 (s, 1H), 10.70 (s, 1H), 10.32 (d, *J* = 7.6 Hz, 1H), 10.25 (s, 1H), 9.42 (d, *J* = 9.7 Hz, 1H), 9.17 – 9.10 (m, 2H), 8.43 –8.39 (m, 2H); **aromatic CHs** δ [ppm] = 8.37 – 8.33 (m), 8.13 – 8.00 (m), 7.90 – 7.82 (m), 7.80 – 7.48 (m), 7.43 – 7.29 (m), 7.25 – 7.14 (m), 7.10 – 7.04 (m), 7.00 – 6.92 (m), 6.88 – 6.78 (m), 6.75 – 6.65 (m), 6.52 (s), 6.49 – 6.45 (m), 6.38 (s), 6.34 – 6.30 (m), 5.99 (s); **aliphatic CHs** δ [ppm] = 5.46 –5.42 (m), 4.35 – 4.26 (m), 4.22 – 4.05 (m), 4.01 – 3.92 (m), 3.83 – 3.70 (m), 3.70 – 3.60 (m), 3.42 – 3.17 (m), 3.09 – 2.99 (m), 2.77 – 2.69 (m), 1.98 (s), 1.35 – 1.29 (m), 1.28 – 1.24 (m), -0.58 (d, *J* = 7.6 Hz). **HRMS** (ESI^-) *m/z* calcd. for: $C_{113}H_{102}N_{19}O_{48}P_9$: 1385.1825 $[M-2H]^{2}$; found: 1385.2080.

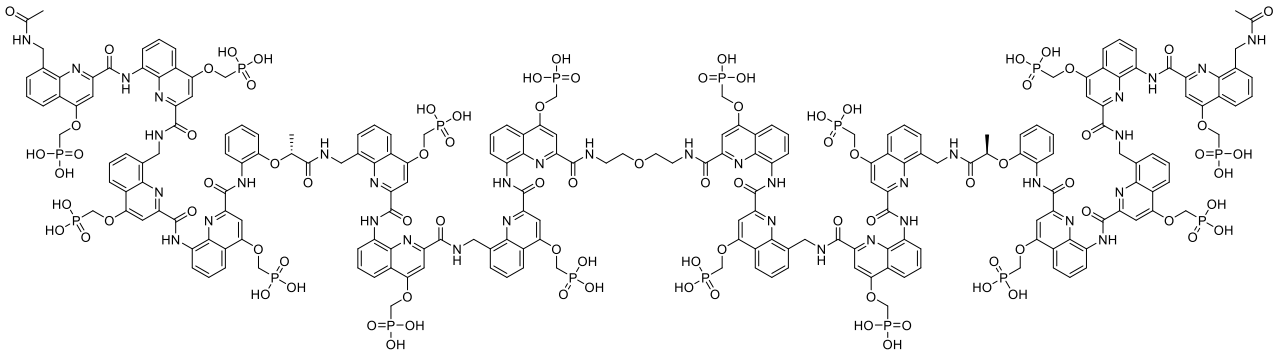


Oligomer (8): Fragment A (24 mg, 8.2 µmol, 2.0 equiv.) was dissolved in 1.35 mL DMF under N_2, DIPEA (17.1 µL, 6.0 eq) was added and wait for 15 min. 270 µL of linker solution (a stock solution of the linker was prepared by dissolving 2 mg of bis-*N*-succinimidyl diglycolic acid in 400 µL of DMF) were added dropwise over 40 min. The reaction mixture was stirred at RT for 2 h and monitored by RP-HPLC. The crude was lyophilized and purified on semi-preparative RP-HPLC with a gradient from 70% to 100% solvent B over 15 min at 50 °C (A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA). After RP-HPLC purification, the protected precursor of **8** was obtained as a light yellow solid (12 mg, 50%; HPLC-purity: >97%). After removal of ethyl phosphonate groups

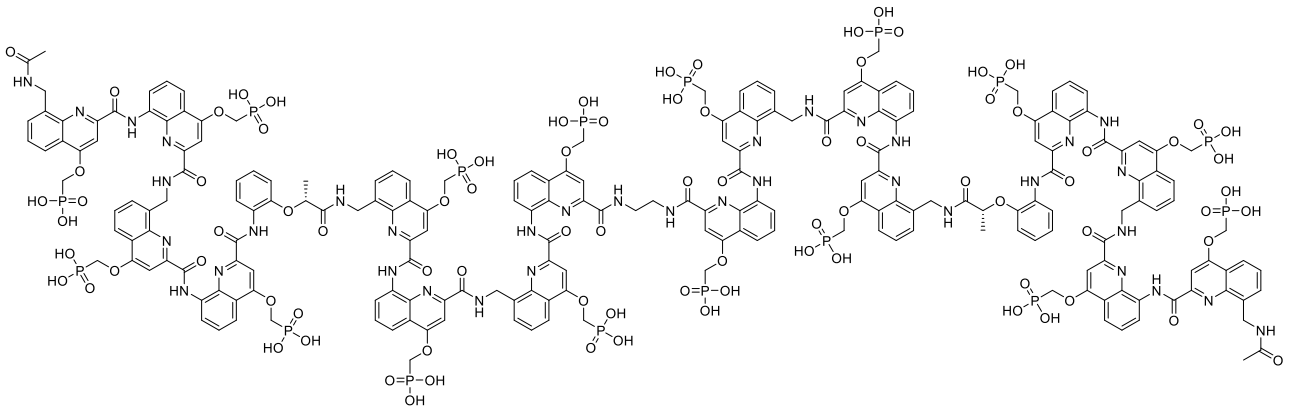
and ion exchange, oligomer **8** was obtained as a light yellow solid (9 mg, 20%; HPLC-purity: >98%). ^{1}H **NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.52 (s, 1H), 10.57 (s, 1H), 9.96 (s, 1H), 9.79 (s, 1H), 9.26 (s, 1H), 9.03 (s, 1H), 8.78 (d, *J* = 13.9 Hz, 1H), 8.09 (m, 1H); **aromatic CHs** δ [ppm] = 8.10 (m), 8.04 (m), 7.87 – 7.77 (m), 7.71 (m), 7.60 (m), 7.58 – 7.47 (m), 7.45 – 7.15 (m), 7.09 – 6.94 (m), 6.77 – 6.65 (m), 6.59 (m), 6.54 (m), 6.43 (m), 6.31 (s, 1H), 6.26 (s), 5.91 (m); **aliphatic CHs** δ [ppm] = 5.43 (m), 4.26 – 4.18 (m), 4.12 (m), 4.11 – 3.80 (m), 3.79 – 3.64 (m), 3.60 – 3.46 (m), 3.30 (s), 2.78 (s), 2.59 (m), 1.34 – 1.21 (m), 0.33 (m), -0.20 (d, *J* = 14.9 Hz), -0.49 (d, *J* = 7.7 Hz). **HRMS** (ESI^-) *m/z* calcd. for: $C_{206}H_{184}N_{34}O_{89}P_{16}$: 1683.8856 $[M-3H]^{3-}$; found: 1683.8949.



Oligomer (9): By using 2,6-Bis(2,5-dioxo-1-pyrrolidinyl) 2,6-pyridinedicarboxylate as a linker, oligomer **9** was synthesized as previously described for oligomer **8**. After removal of ethyl phosphonate groups and ion exchange, oligomer **9** was obtained as a light yellow solid (9 mg, 19%; HPLC-purity: >98%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.49 (s, 1H), 10.53 (s, 1H), 9.90 (s, 1H), 9.47 (s, 1H), 9.27 – 9.21 (m, 1H), 8.96 (s, 1H), 8.64 (d, *J* = 10.4 Hz, 1H), 8.52 (d, *J* = 6.9 Hz, 1H), 8.11 (m,1H); **aromatic CHs** δ [ppm] = 8.01 (m), 7.80 (m), 7.69 (m), 7.57 (m), 7.53 (m), 7.47 (m), 7.41 – 7.28 (m), 7.26 – 7.12 (m), 6.97 (m), 6.85 – 6.76 (m, 2H), 6.73 (s), 6.68 (s), 6.57 (m), 6.51 (m), 6.43 (m), 6.37 (s), 6.30 – 6.22 (m); **aliphatic CHs** δ [ppm] = 5.92 (s), 5.63 (m), 4.10 (m), 4.04 (s), 4.01 (m), 3.93 – 3.84 (m), 3.80 (m), 3.69 (m), 3.47 (m), 3.36 – 3.25 (m), 3.15 – 3.07 (m), 2.79 (m), 2.58 (s, 1H), 1.33 – 1.21 (m), -0.47 (d, *J* = 7.5 Hz). **HRMS** (ESI^-) *m/z* calcd. for: $C_{209}H_{183}N_{35}O_{88}P_{16}$: 1694.8857 $[M-3H]^{3-}$; found: 1694.9024.

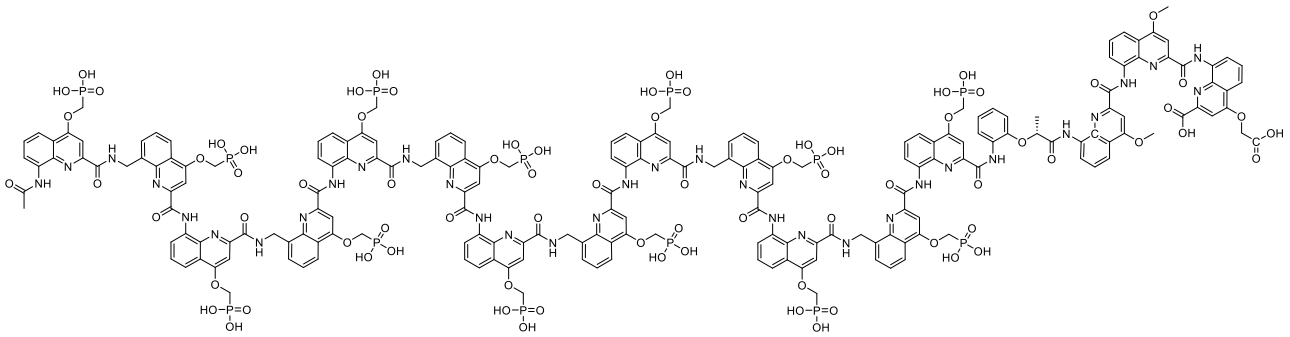


Oligomer (10): Fragment B (12 mg, 4 µmol, 2.0 equiv.) was dissolved in DMF (1.35 mL) under N_2, DIPEA (4.18 µL, 6.0 eq) and PyBOP (6.24 mg, 3.0 eq) were added and wait for 15 min. 100 µL of linker solution [a stock solution of the linker was prepared by dissolving 21.5 µL of 2,2'-oxybis(ethylamine) in 1mL of DMF] were added dropwise over 40 min. The reaction mixture was stirred at RT for 2 h and monitored by RP-HPLC. The crude was lyophilized and purified on semi-preparative RP-HPLC with a gradient from 30% to 100% solvent B over 20 min at 50 °C (A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA). After RP-HPLC purification, the protected precursor of **10** was obtained as a light yellow solid (6 mg, 50%; HPLC-purity: >99%). After removal of ethyl phosphonate groups and ion exchange, oligomer **10** was obtained as a light yellow solid (4.8 mg, 95%; HPLC-purity: >99%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] =11.26 (s, 1H), 10.36 (s, 1H), 10.19 (s, 1H), 9.68 (d, *J* = 15.2 Hz, 2H), 9.57 (d, *J* = 7.9 Hz, 1H), 9.13 (s, 1H), 8.24 (d, *J* = 9.1 Hz, 1H); **aromatic CHs** δ [ppm] = 7.79 – 7.46 (m), 7.39 – 7.29 (m), 7.26 (m), 7.18 – 7.05 (m), 6.95 (m), 6.72 (m), 6.64 (s), 6.58 (m), 6.52 (m), 6.43 (m), 6.26 (m); **aliphatic CHs** δ [ppm] = 4.08 (m), 4.00 (m), 3.93 (m), 3.63 (m), 3.44 (m), 3.31 (m), 3.18 (m), 2.91 (m), 2.81 (m), 2.73 – 2.65 (m), 2.37 (m), 1.34 – 1.23 (m), 1.19 (s), 1.16 (m), -0.64 (d, *J* = 7.6 Hz). **HRMS** (ESI^-) *m/z* calcd. for: $C_{210}H_{194}N_{36}O_{87}P_{16}$: 1701.9171 $[M-3H]^{3-}$; found: 1702.0112.

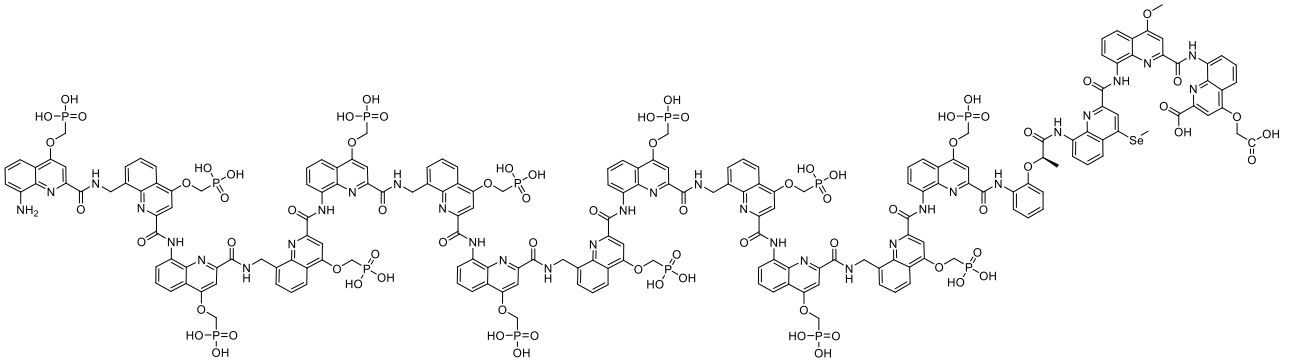


Oligomer (11): By using ethylenediamine as a linker, oligomer **11** was synthesized as above described for oligomer **10**. After removal of ethyl phosphonate groups and ion exchange, oligomer **11** was obtained as a light yellow solid (5.6 mg, 17%; HPLC-purity: >98%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.06 (s, 1H), 10.30 (s, 1H), 10.24 (s, 1H), 9.65 (d, *J* = 8.8 Hz, 1H), 9.44 (s, 1H), 9.37 (d, *J* = 9.6 Hz, 1H), 9.19 (s, 1H), 8.29 (d, *J* = 9.3 Hz, 1H); **aromatic CHs** δ [ppm] =7.99 (m), 7.94 (m), 7.81 (s), 7.82 – 7.75 (m), 7.69 (m), 7.63 (m), 7.58 (m), 7.51 – 7.41 (m), 7.39 (m), 7.33 (m), 7.24 – 7.14 (m), 7.12 – 7.00 (m), 6.85 (m), 6.81 (m), 6.75 (m), 6.67 (m), 6.59 (m), 6.49 – 6.40 (m), 6.38 (m), 6.14 (tm), 5.95 (s); **aliphatic CHs** δ [ppm] = 4.37 – 4.30 (m), 4.07 (m), 3.86 (m), 3.72 (m), 3.47 (m), 3.25 – 3.13 (m), 3.12 (s), 2.99 – 2.86 (m), 2.33 (m), 1.34 – 1.22 (m),

1.17 (m), -0.53 (d, *J* = 7.6 Hz). **HRMS** (ESI$^-$) *m/z* calcd. for: $C_{208}H_{190}N_{36}O_{86}P_{16}$: 1687.6663 $[M-3H]^{3-}$; found: 1687.6642.



Oligomer (12): Oligomer **12** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following recently reported procedures.[12b] After removal of the diethylphosphonate groups (scale: 6.39 µmol), the title compound was purified using a linear gradient of 0-15 B in A (A: 12.5 mmol NH_4OAc, B: acetonitrile), then lyophilized and obtained as a light yellow solid (17.7 mg, 62%; HPLC-purity: >99%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.90 (s, 1H), 11.22 (s, 1H), 10.66 (s, 1H), 10.21 – 10.16 (m, 2H), 9.94 (s, 1H), 9.80 (s, 1H), 9.65 (s, 1H), 9.63 – 9.57 (m, 2H), 9.49 (s, 1H), 9.39 (s, 1H), 9.28 (d, *J* = 8.2 Hz, 1H), 9.04 (d, *J* = 8.3 Hz, 1H), 9.01 (s, 1H), 8.94 (d, *J* = 9.6 Hz, 1H), 8.83 (d, *J* = 8.7 Hz, 1H); **aromatic CHs** δ [ppm] = 8.43 (d, *J* = 9.0 Hz, 1H), 8.35 (d, *J* = 8.0 Hz, 1H), 8.11 (d, *J* = 8.0 Hz, 1H), 8.06 – 8.00 (m, 4H), 7.96 (d, *J* = 9.1 Hz, 3H), 7.84 (d, *J* = 8.7 Hz, 1H), 7.77 (d, *J* = 8.1 Hz, 1H), 7.71 – 7.54 (m, 6H), 7.54 – 7.46 (m, 3H), 7.47 – 7.38 (m, 3H), 7.34 – 7.25 (m, 3H), 7.25 – 7.19 (m, 2H), 7.17 – 7.11 (m, 4H), 7.08 (d, *J* = 8.1 Hz, 1H), 7.05 – 6.93 (m, 10H), 6.93 – 6.84 (m, 4H), 6.81 – 6.74 (m, 2H), 6.71 (d, *J* = 8.3 Hz, 1H), 6.65 – 6.56 (m, 6H), 6.56 – 6.43 (m, 10H), 6.39 (t, *J* = 8.3 Hz, 1H), 6.34 (s, 1H), 6.32 – 6.26 (fm, 4H), 6.18 (s, 3H), 6.14 (s, 1H), 6.03 (d, *J* = 7.2 Hz, 3H); **aliphatic CHs** δ [ppm] = 4.05 – 3.85 (m, 6H), 3.81 – 3.44 (m, 10H), 3.43 – 3.15 (m, 2H), 2.83 – 2.70 (m, 1H), 2.70 – 2.60 (m, 1H), 2.59 – 2.51 (m, 1H), 2.41 – 2.30 (m, 1H), -0.08 (d, *J* = 6.8 Hz, 3H). **HRMS** (ESI$^-$) *m/z* calcd. for: $C_{194}H_{163}N_{33}O_{77}P_{13}^{3-}$ $[M-3H]^{3-}$: 1530.2175; found: 1530.2348.



Oligomer (13): Oligomer **13** was synthesized on Wang resin (0.41 mmol g^{-1}, 30 µmol scale) following recently reported procedures.[12b] After removal of the diethylphosphonate groups (scale: 3.74 µmol), the title compound was purified using a linear gradient of 0-15 B in A (A: 12.5 mmol NH_4OAc, B: acetonitrile), then lyophilized and obtained as a light yellow solid (7.13 mg, 41%; HPLC-purity: >98%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1): amide NHs** δ [ppm] = 11.93 (s, 1H), 11.19 (s, 25H), 10.92 (s, 1H), 10.20 (s, 1H), 10.17 (s, 1H), 9.97 (s, 1H), 9.88 (s, 1H), 9.80 (s, 1H), 9.68 (s, 1H), 9.55 (s, 1H), 9.34 (s, 1H), 9.29 (s, 1H), 9.21 (d, *J* = 7.7 Hz, 1H), 8.94 (d, *J* = 7.9 Hz, 1H), 8.80 (d, *J* = 8.3 Hz, 1H), 8.71 (d, *J* = 7.8 Hz, 1H); **aromatic CHs** δ [ppm] = 8.42 – 8.32 (m, 2H), 8.15 – 7.96 (m, 7H), 7.82 (d, *J* = 8.8 Hz, 1H), 7.69 (d, *J* = 8.8 Hz, 1H), 7.64 – 7.48 (m, 6H), 7.47 – 7.29 (m, 5H), 7.24 (d, *J* = 8.7 Hz, 1H), 7.13 (d, *J* = 8.1 Hz, 2H), 7.10 – 6.92 (m, 12H), 6.91 – 6.84 (m, 1H), 6.81 – 6.39 (m, 12H), 6.34 (s, 1H), 6.30 (s, 1H), 6.25 (s, 1H), 6.20 (d, *J* = 8.7 Hz, 3H), 6.05 (s, 1H), 6.00 (s, 1H); **aliphatic CHs** δ [ppm] = 4.09 (s, 2H), 4.03 – 3.32 (m, 20H), 2.69 – 2.45 (m, 2H), -0.03 (d, *J* = 6.8 Hz, 3H). **HRMS** (ESI^-) *m/z* calcd. for: $C_{192}H_{161}N_{33}O_{75}P_{13}Se^{3-}$ $[M-3H]^{3-}$: 1537.5212; found: 1537.6029.

4. NMR spectra and HPLC chromatograms

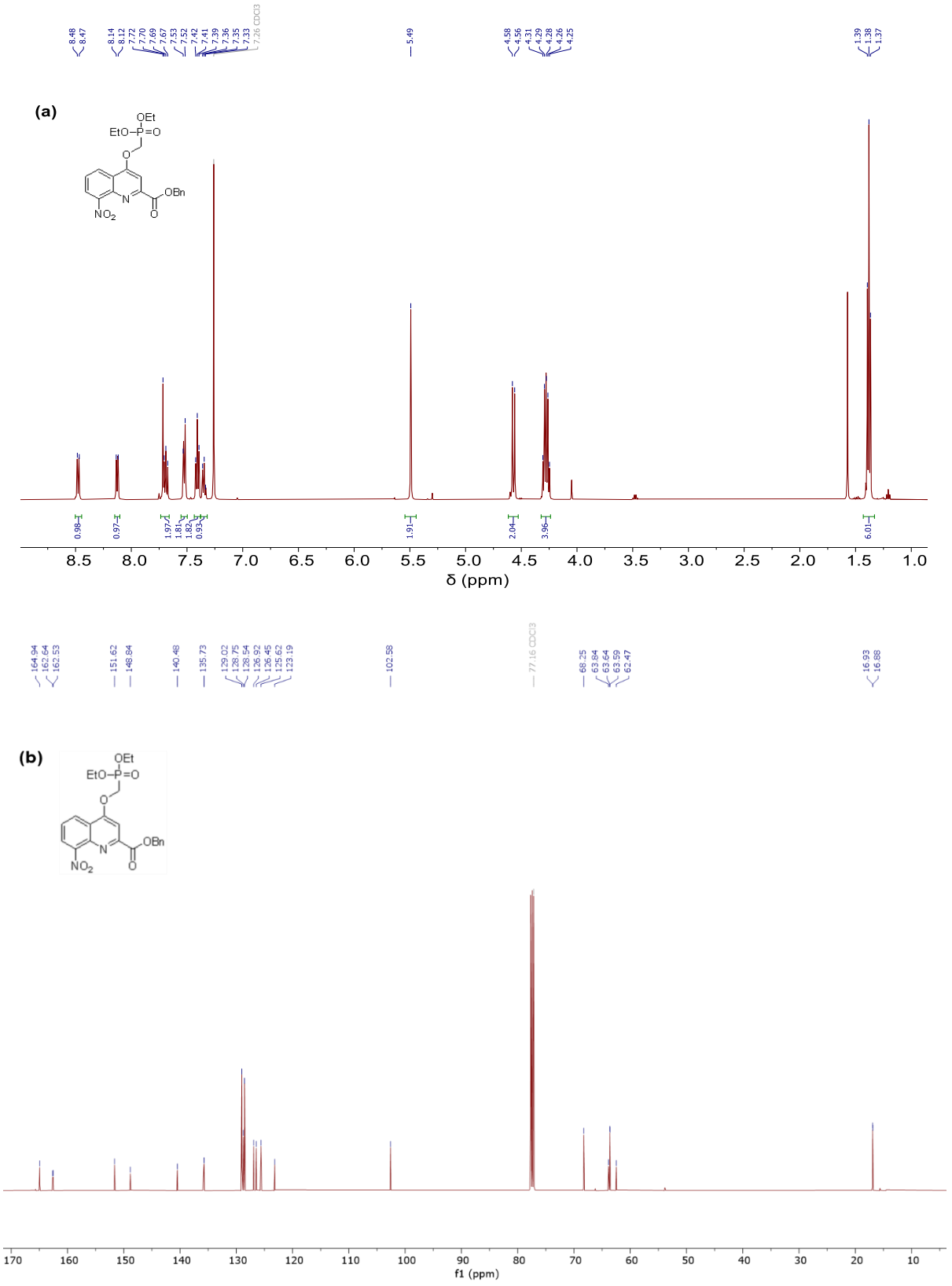


Figure S9. NMR spectra of compound **15**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

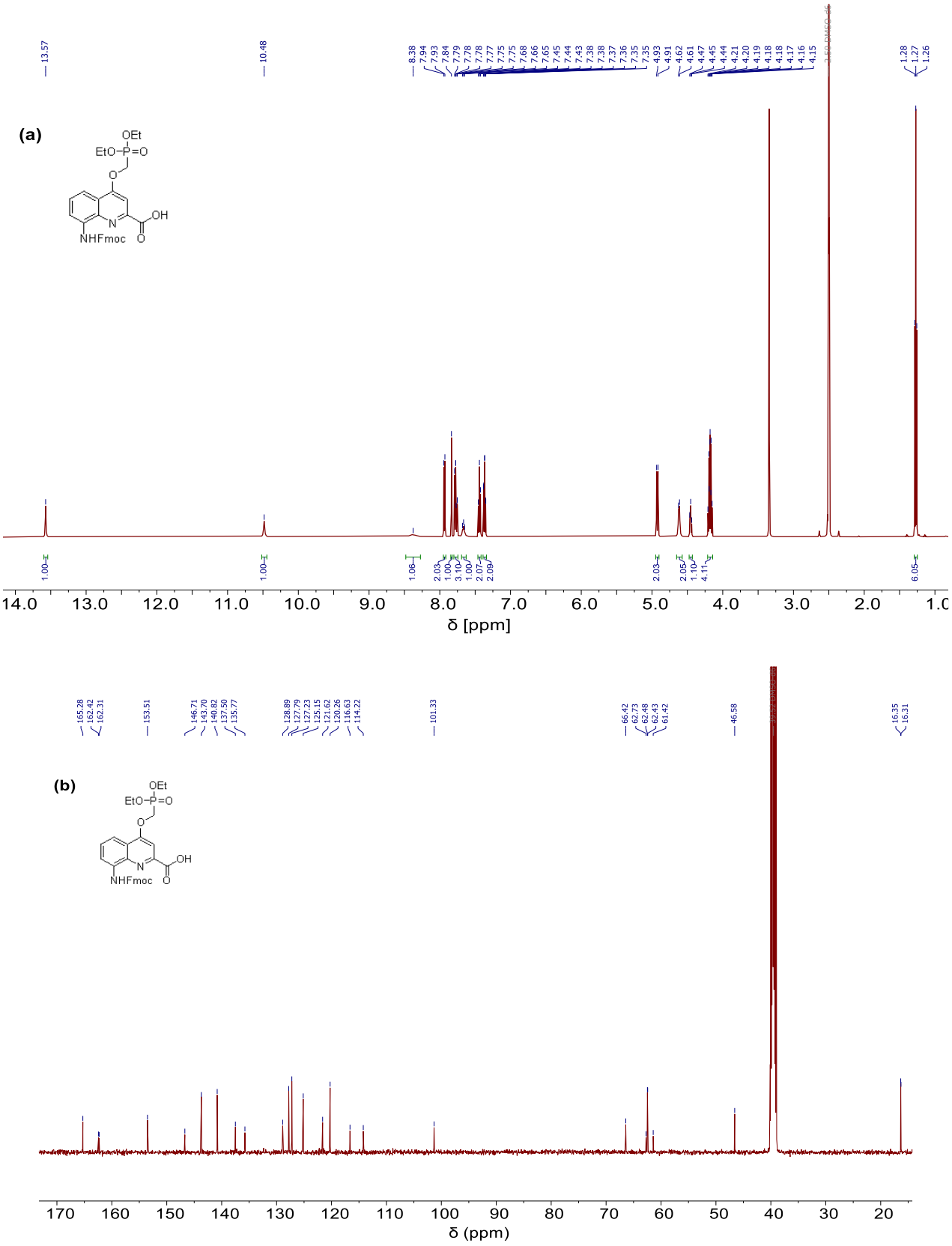


Figure S10. NMR spectra of compound **17**. (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

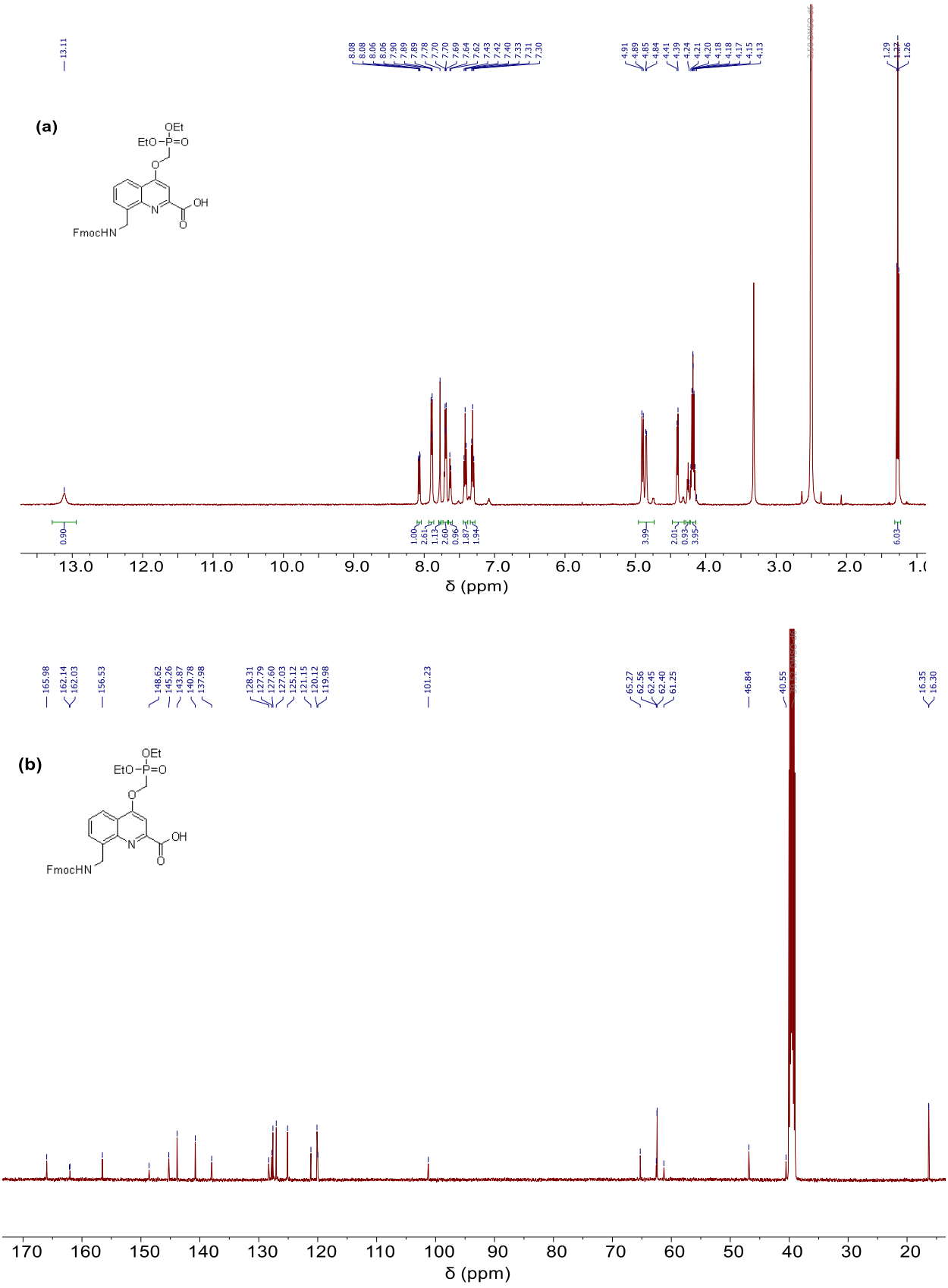


Figure S11. NMR spectra of compound **21**. (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

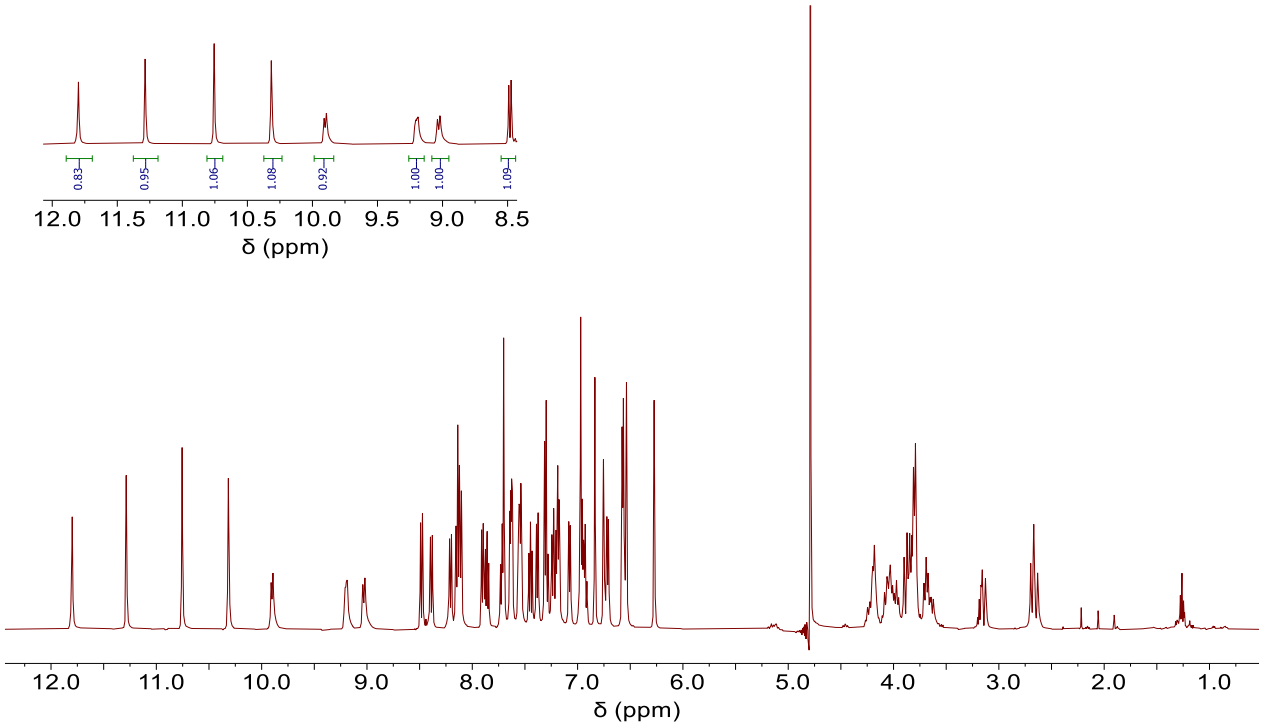


Figure S12. ^{1}H NMR spectrum of oligomer **1** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

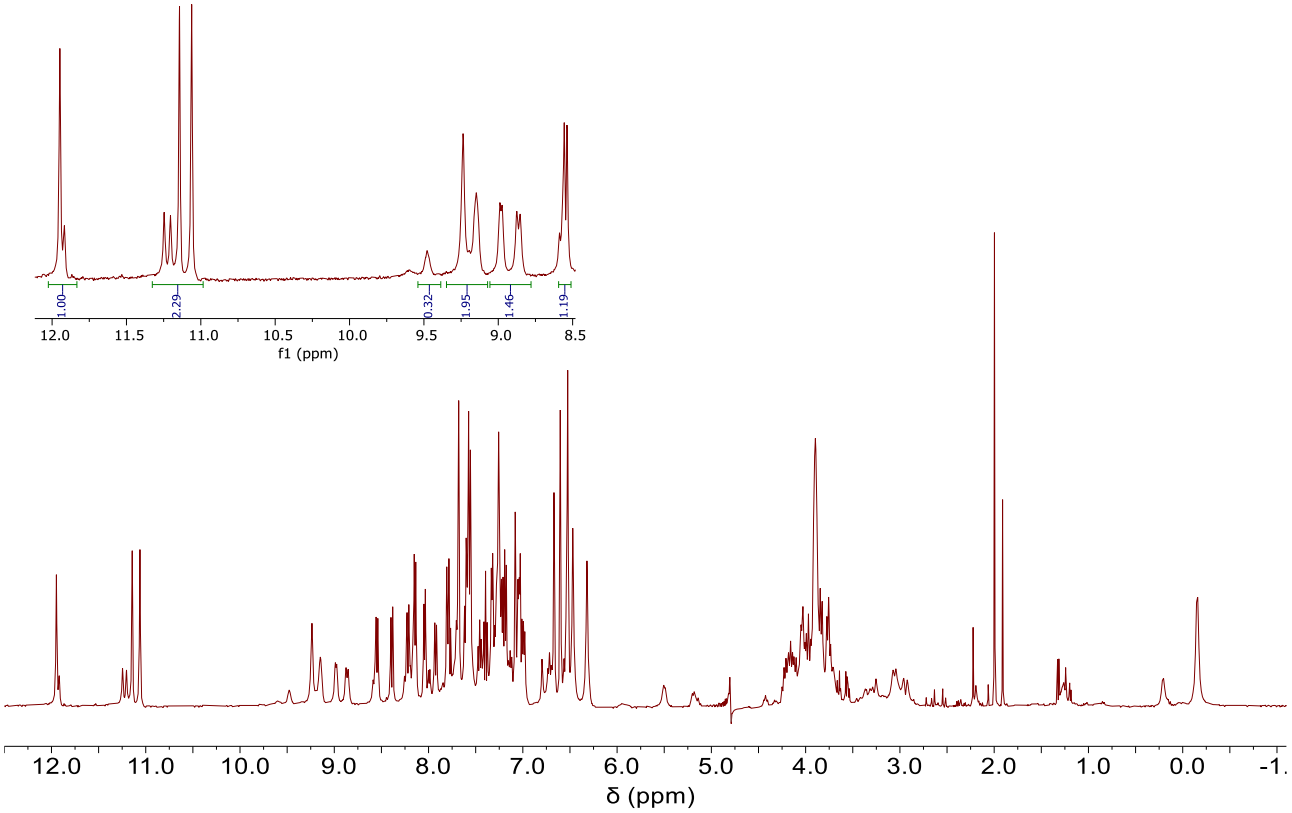


Figure S13. ^{1}H NMR spectrum of oligomer **2** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

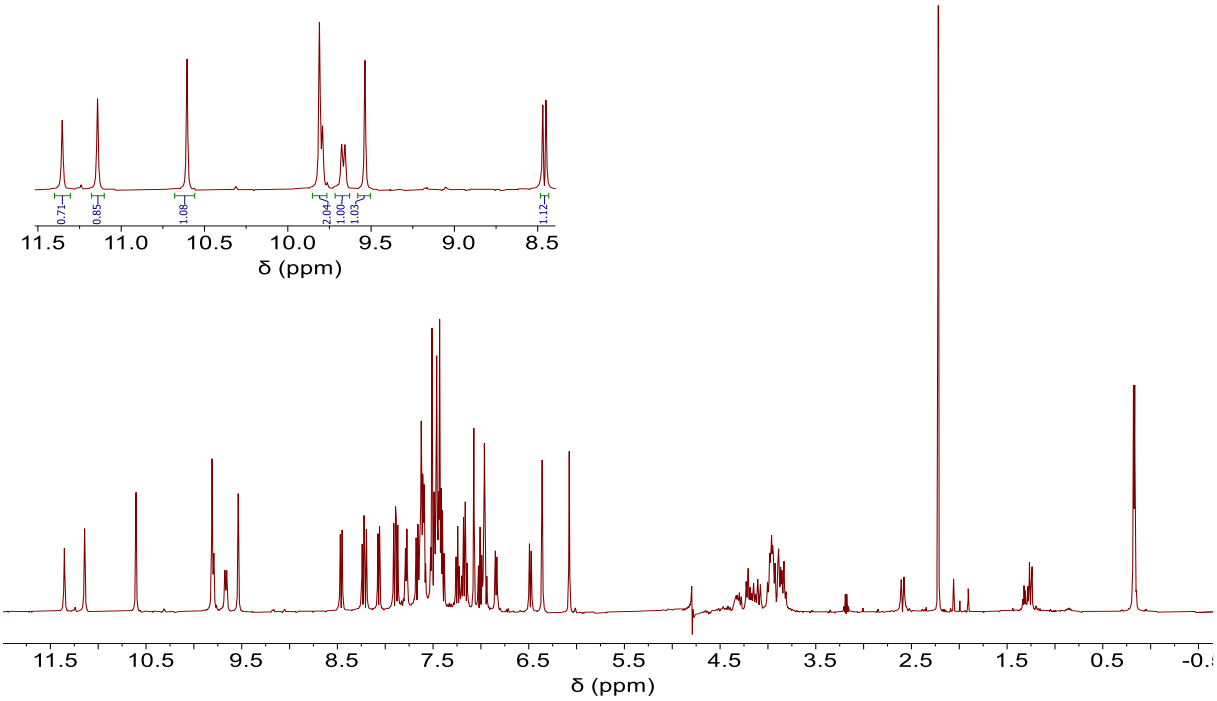


Figure S14. ^{1}H NMR spectrum of oligomer **3** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

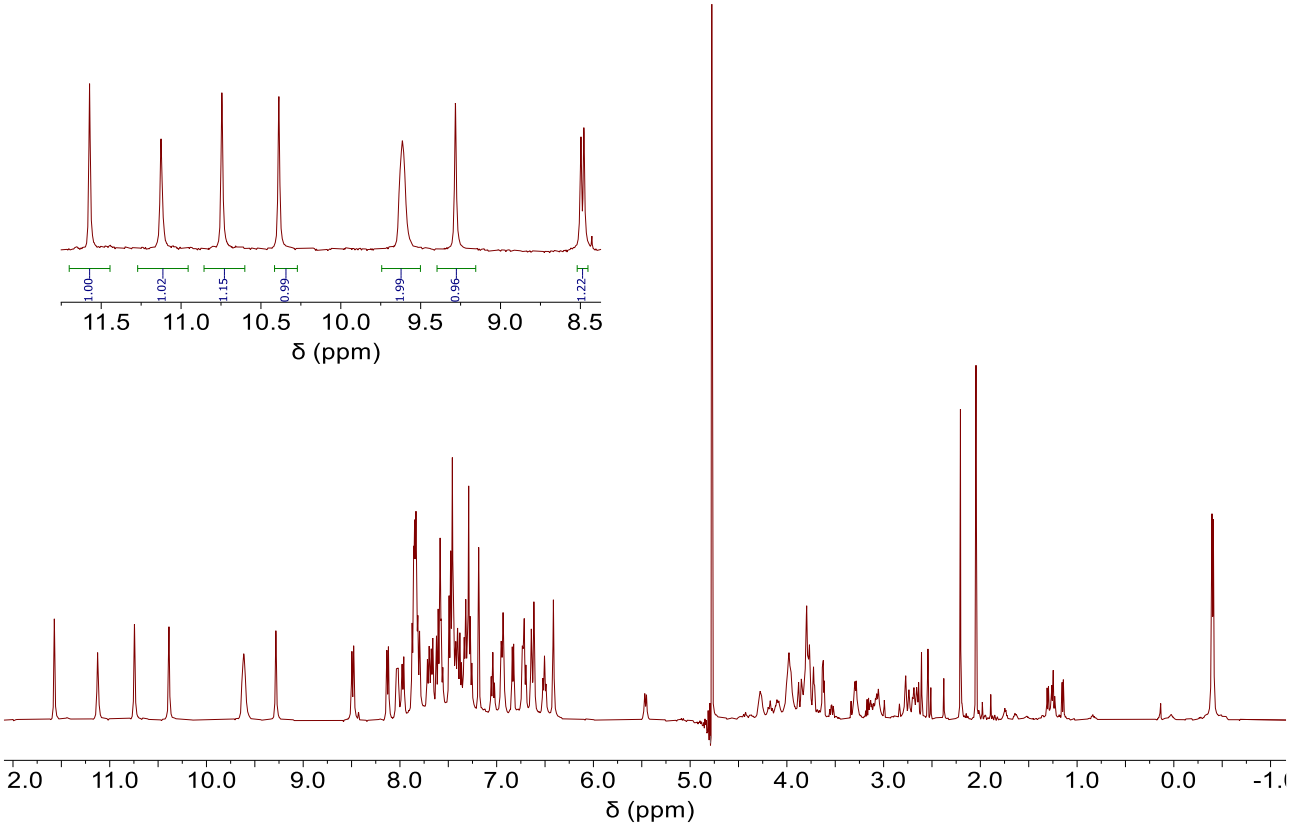


Figure S15. ^{1}H NMR spectrum of oligomer **4** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

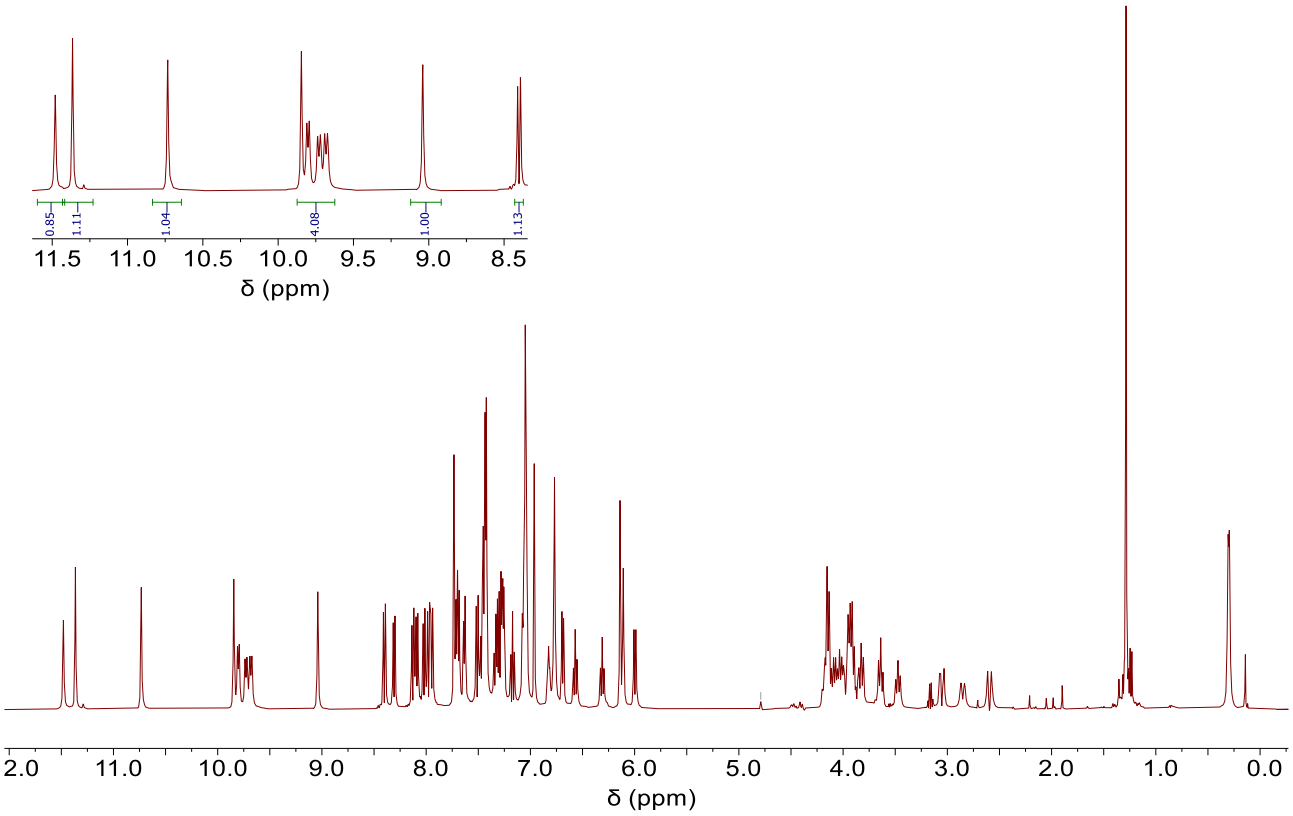


Figure S16. ^{1}H NMR spectrum of oligomer **5** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

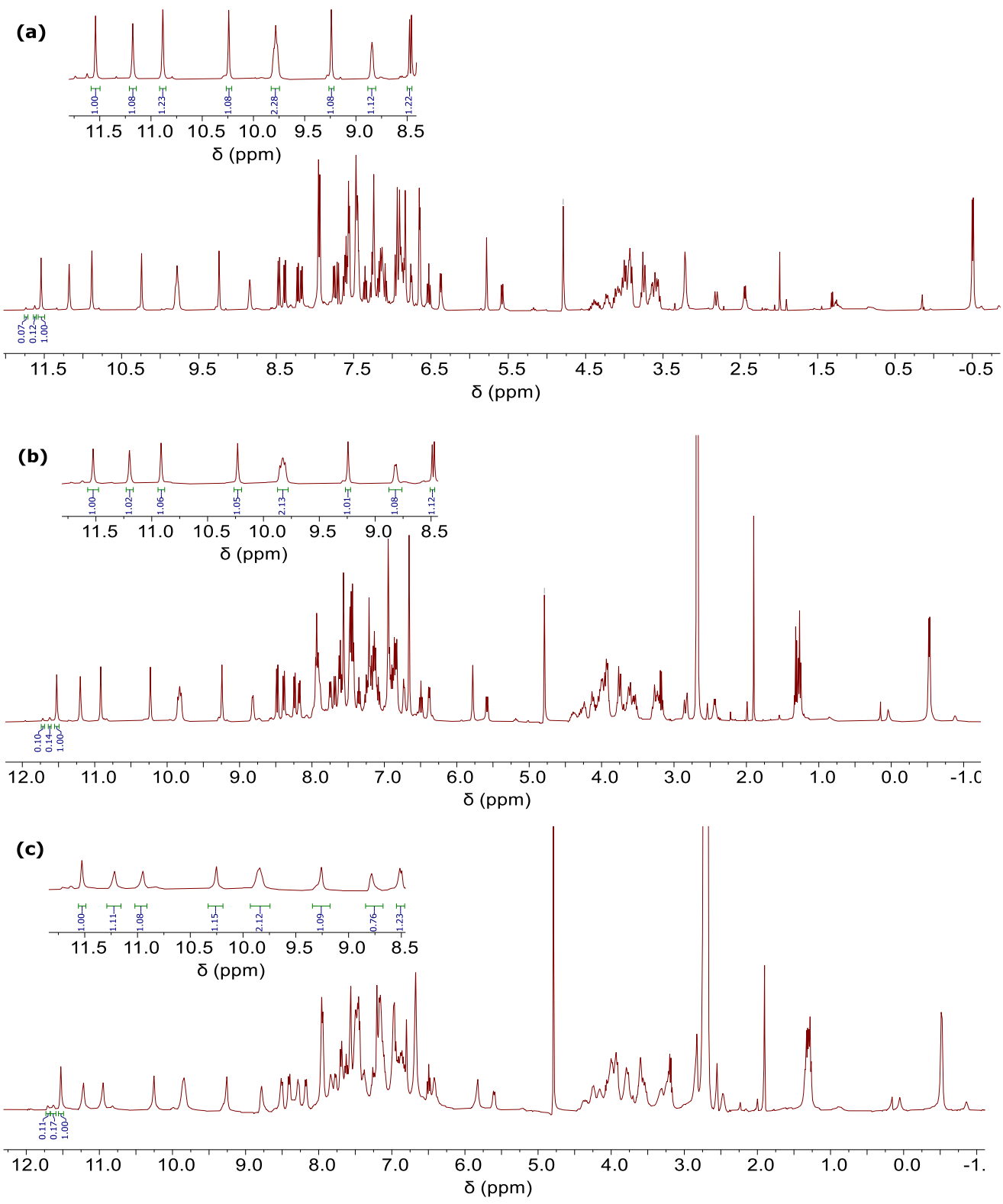


Figure S17. ^{1}H NMR spectra of oligomer **6**. (a) ^{1}H NMR (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5). (b) ^{1}H NMR (500 MHz, 81% NH_4HCO_3 + 9% D_2O + **10% DMSO-*d*$_6$**). (c) ^{1}H NMR (500 MHz, 63% NH_4HCO_3 + 7% D_2O + **30% DMSO-*d*$_6$**). The integrated protons are related to the NH amide region. The proportion between signal integration of the major set of signals and minor set of signals near 11.5 ppm changed as a function of the amount of DMSO-*d*$_6$ added to the

sample. The relative integration of the peaks from the minor signal set increases compared to the integration of the peaks of the major set of signals when the amount of DMSO changed from 0% to 30%. These data together with the clean HPLC profile support strong but not full handedness control of oligomer **6**.

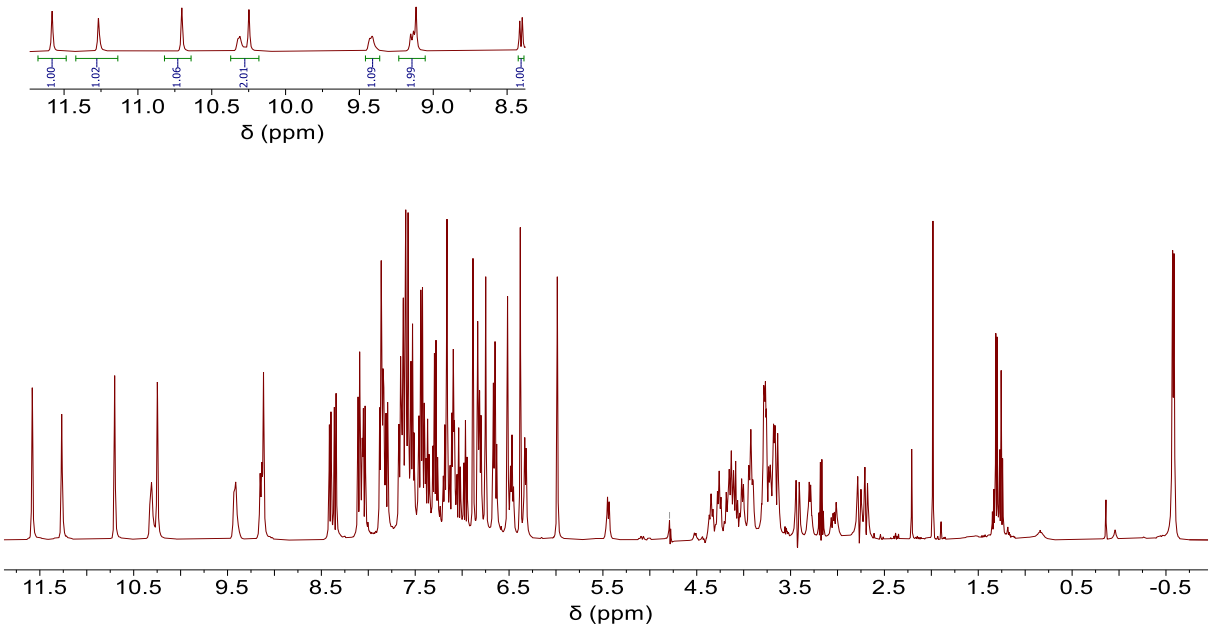


Figure S18. ^{1}H NMR spectrum of oligomer **7** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

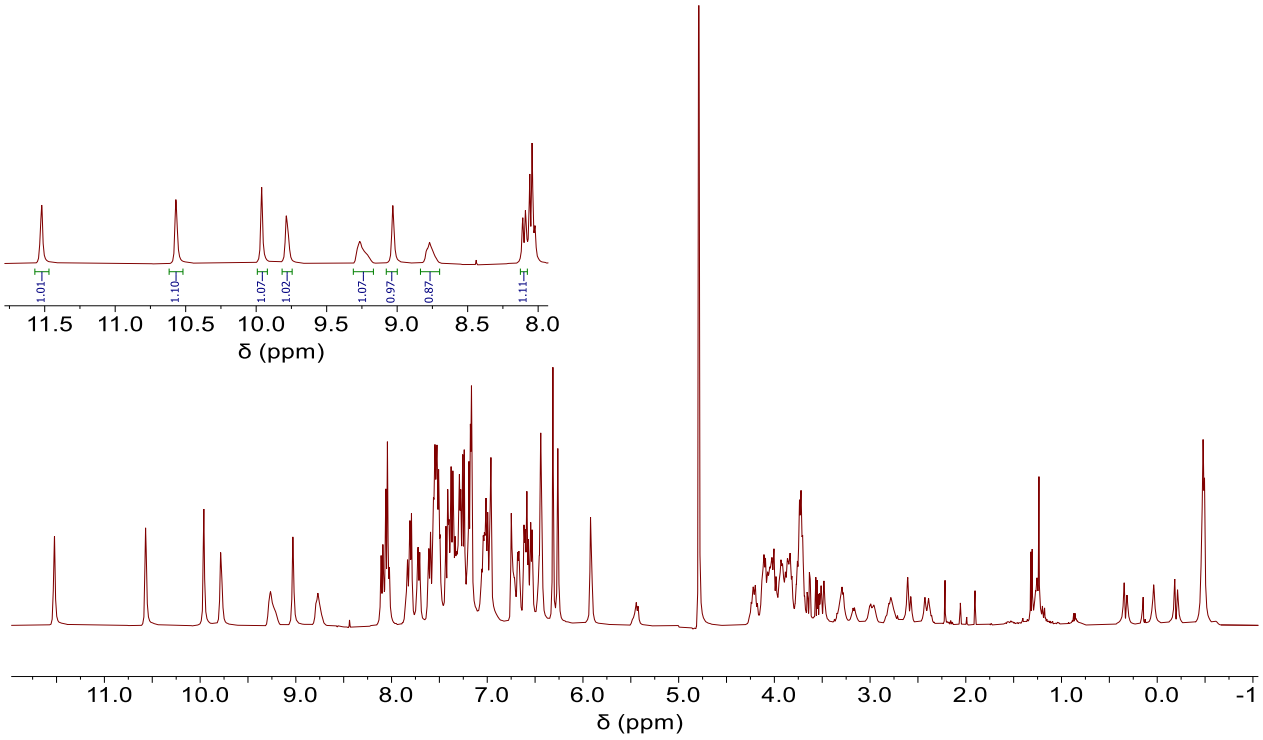


Figure S19. ^{1}H NMR spectrum of oligomer **8** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

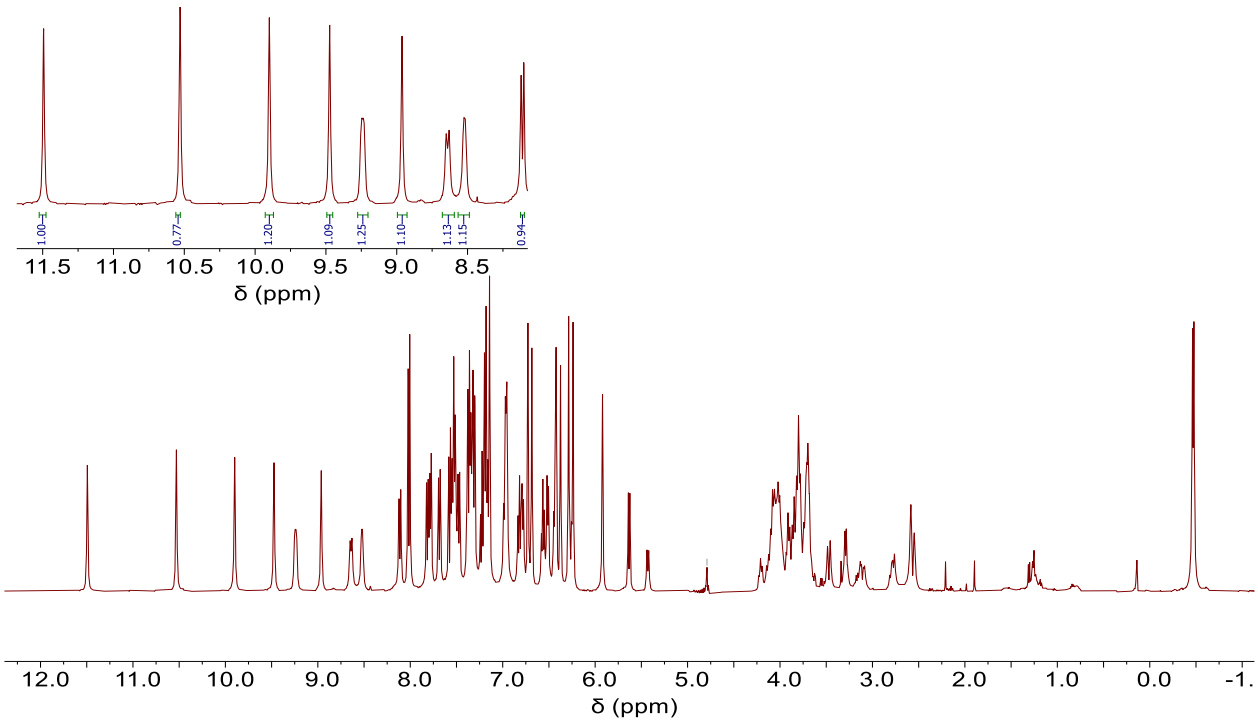


Figure S20. ^{1}H NMR spectrum of oligomer **9** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

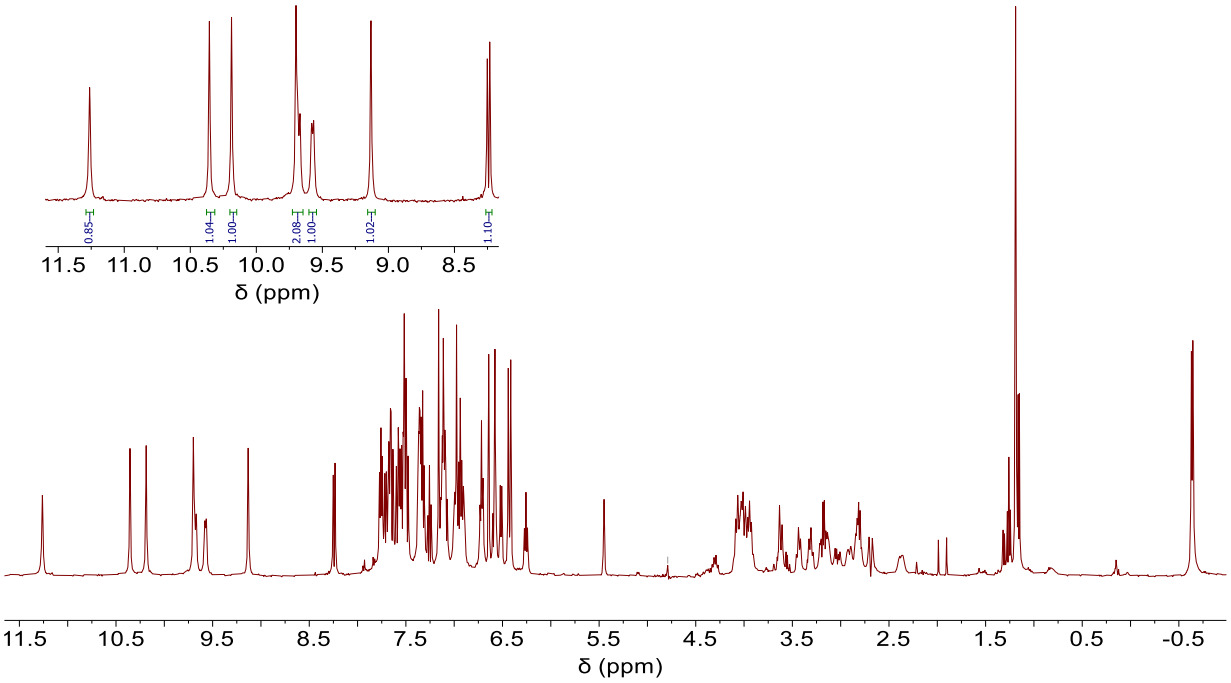


Figure S21. ^{1}H NMR spectrum of oligomer **10** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

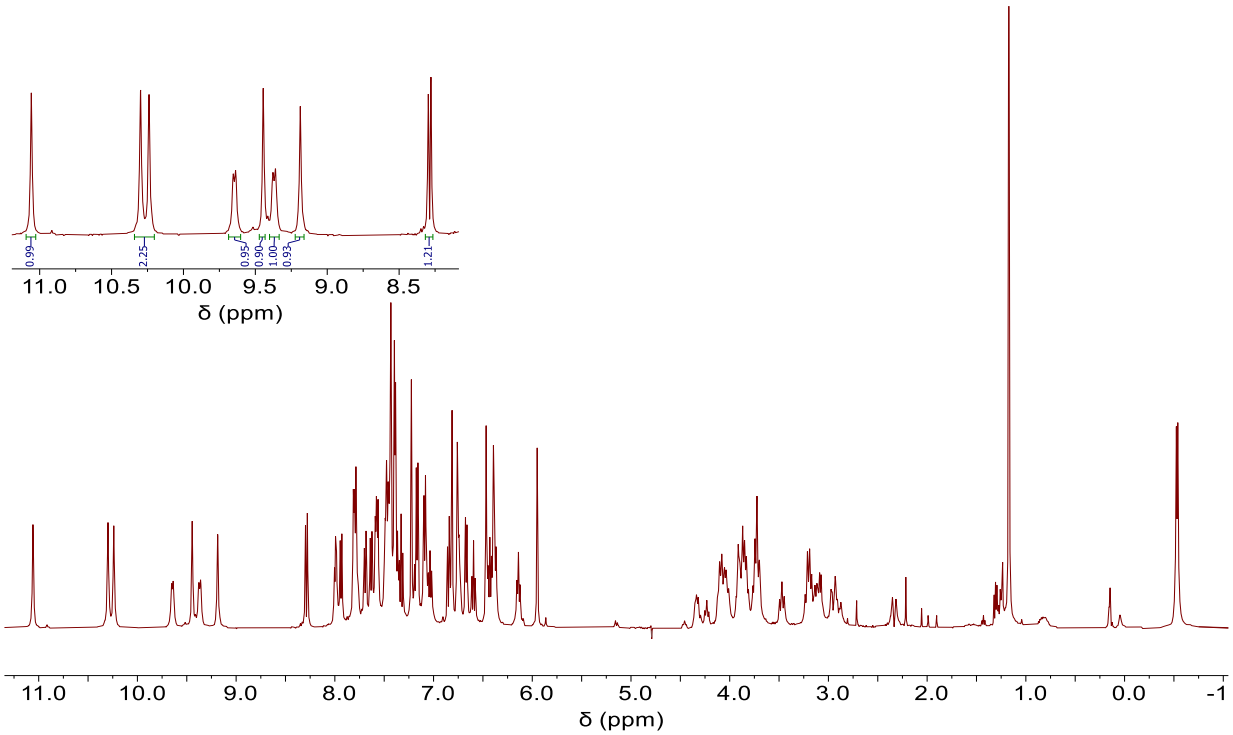


Figure S22. ^{1}H NMR spectrum of oligomer **11** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

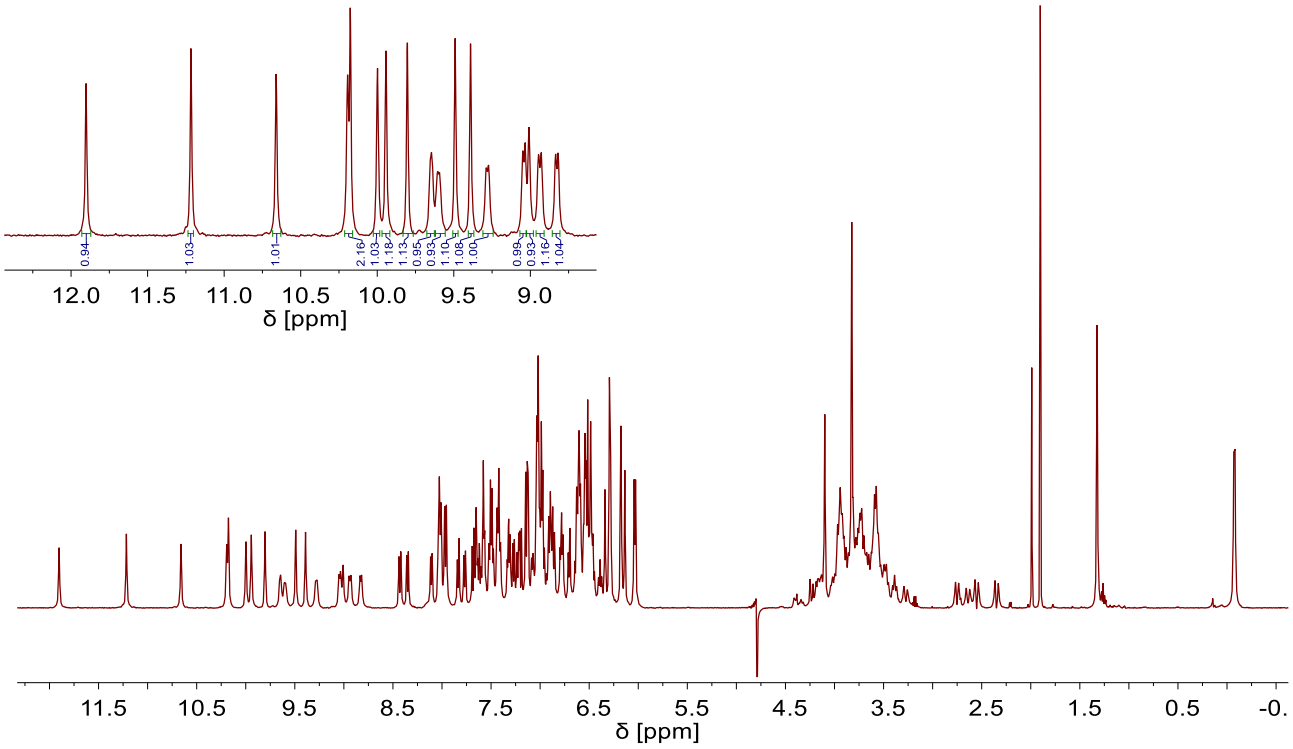


Figure S23. ^{1}H NMR spectrum of oligomer **12** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, water suppression). The integrated protons are related to the NH amide region.

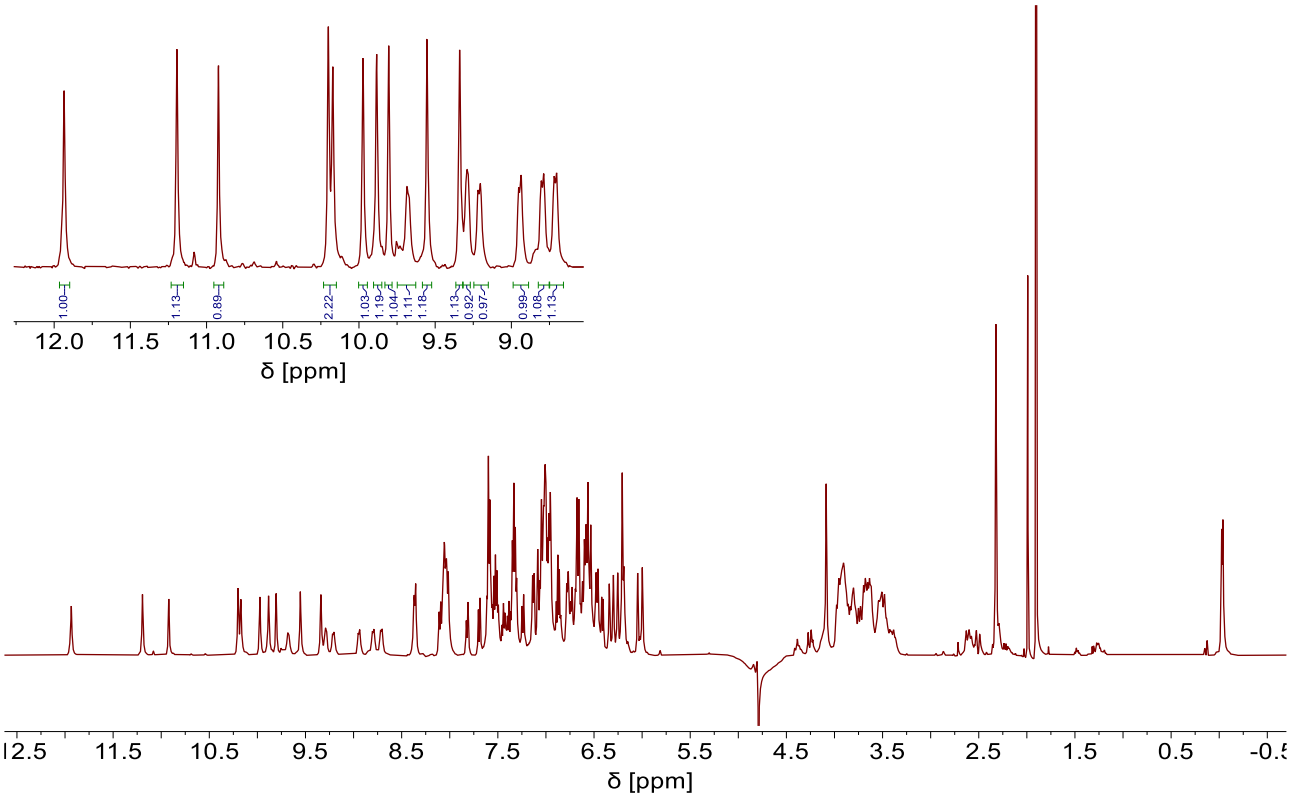


Figure S24. ^{1}H NMR spectrum of oligomer **13** (500 MHz, H_2O/D_2O 9:1 v/v, 50 mM NH_4HCO_3, pH 8.5, H_2O suppression). The integrated protons are related to the NH amide region.

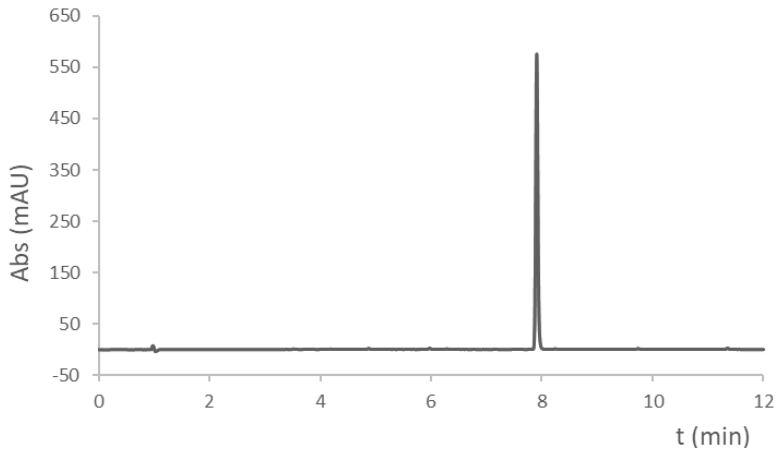


Figure S25. RP-HPLC chromatogram of **Fmoc-Q^{Pho}-COOH** using a linear gradient from 15% B to 100% B in 10 min; A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA.

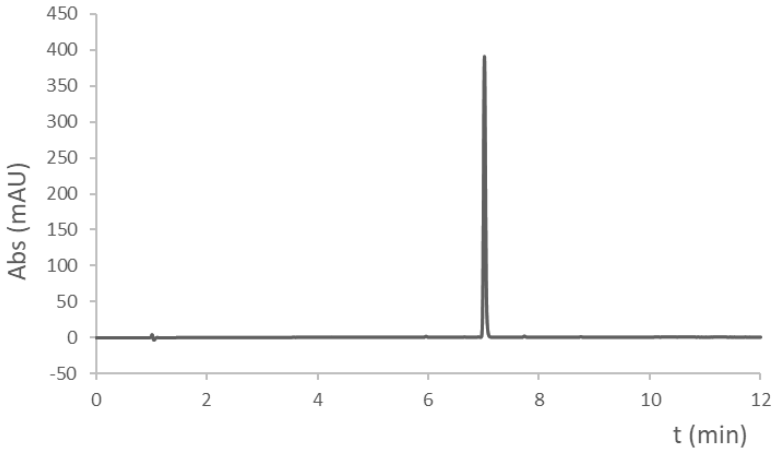


Figure S26. RP-HPLC chromatogram of **Fmoc-M^{Pho}-COOH** using a linear gradient from 15% B to 100% B in 10 min; A: water + 0.1% TFA and B: acetonitrile + 0.1% TFA.

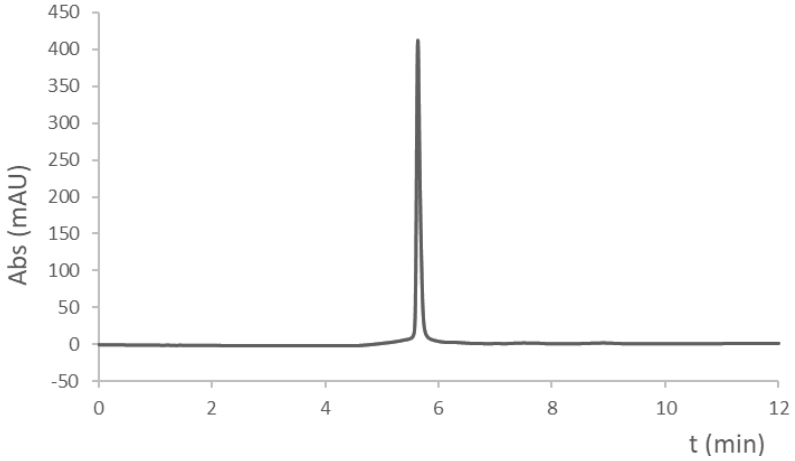


Figure S27. RP-HPLC chromatogram of oligomer **2** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

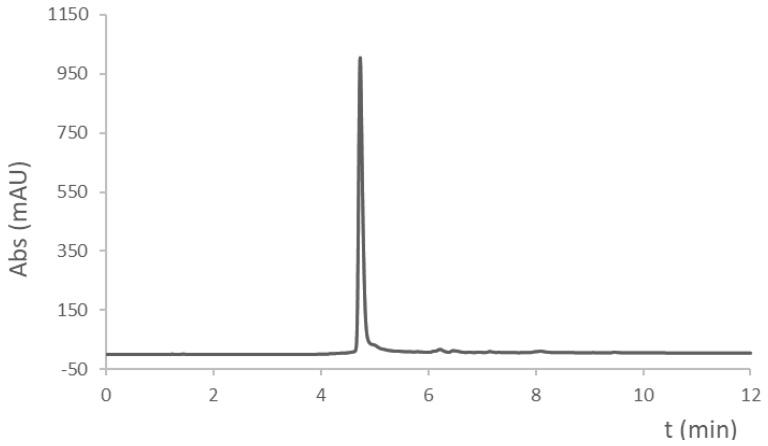


Figure S28. RP-HPLC chromatogram of oligomer **3** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

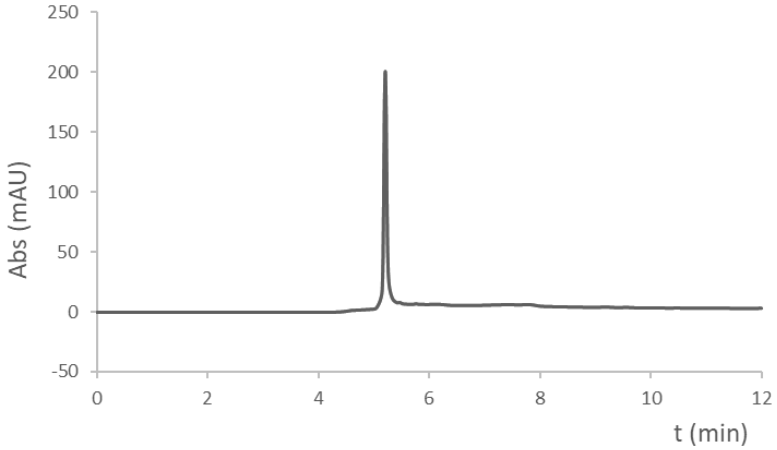


Figure S29. RP-HPLC chromatogram of oligomer **4** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

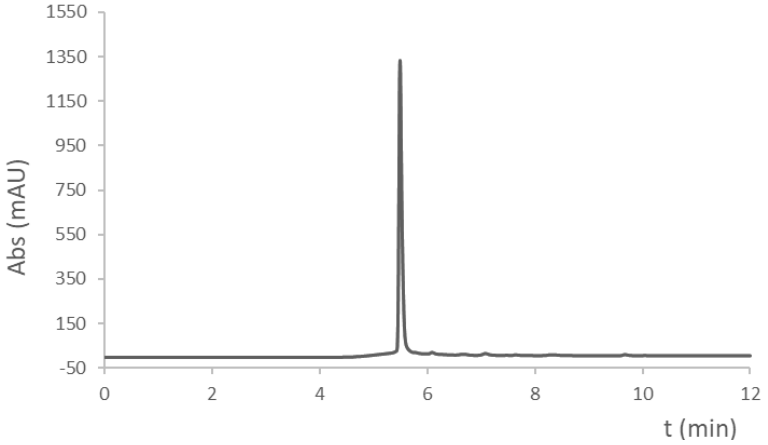


Figure S30. RP-HPLC chromatogram of oligomer **5** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile ((1:2), pH 8.5.

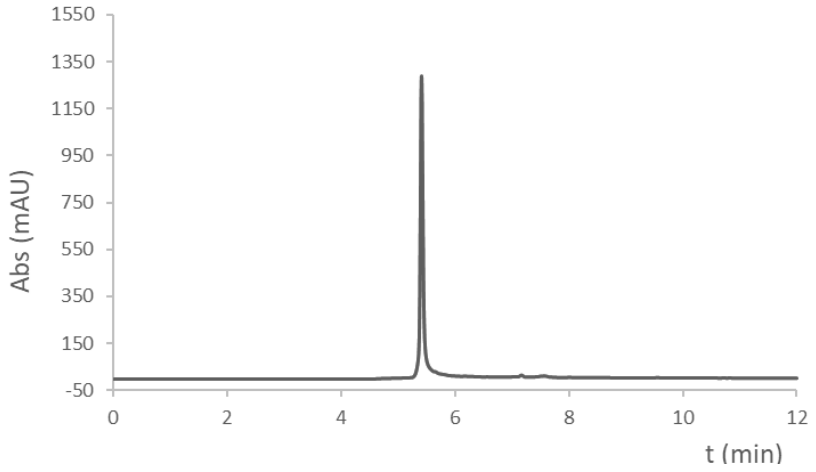


Figure S31. RP-HPLC chromatogram of oligomer **6** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

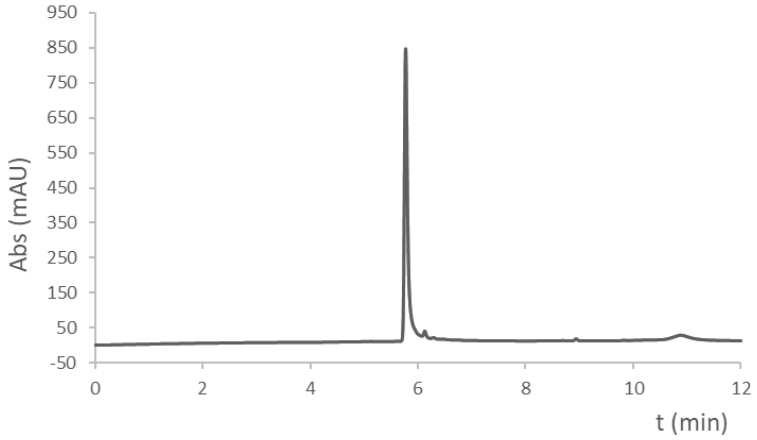


Figure S32. RP-HPLC chromatogram of oligomer **7** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

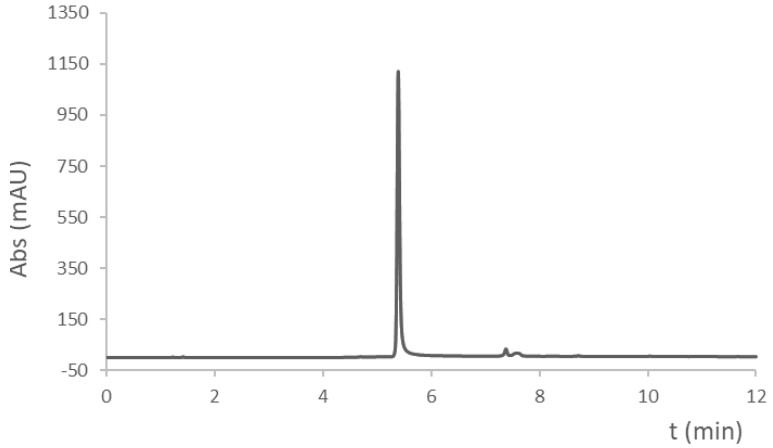


Figure S33. RP-HPLC chromatogram of oligomer **8** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

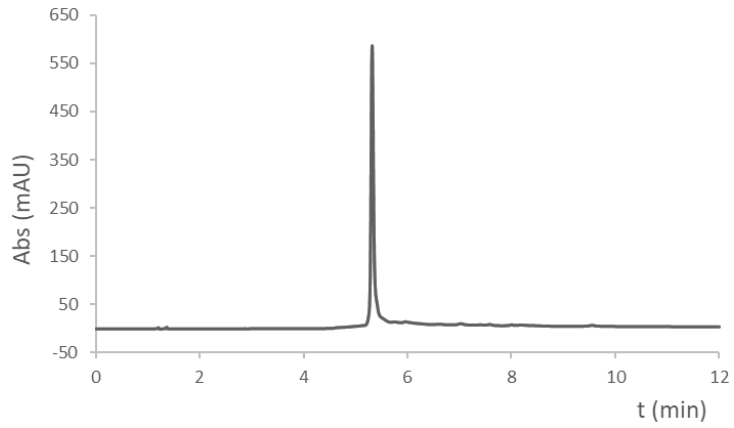


Figure S34. RP-HPLC chromatogram of oligomer **9** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

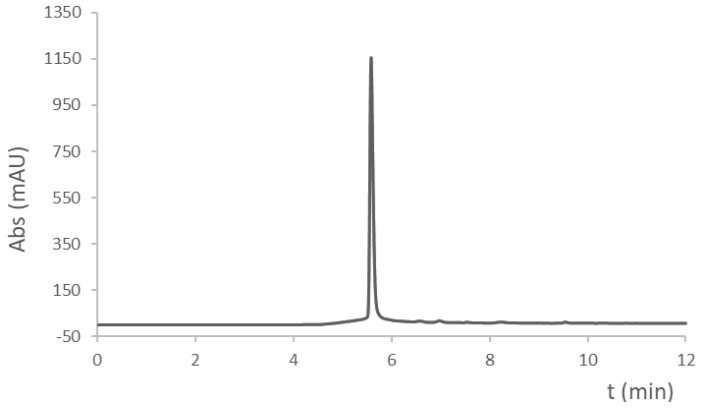


Figure S35. RP-HPLC chromatogram of oligomer **10** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

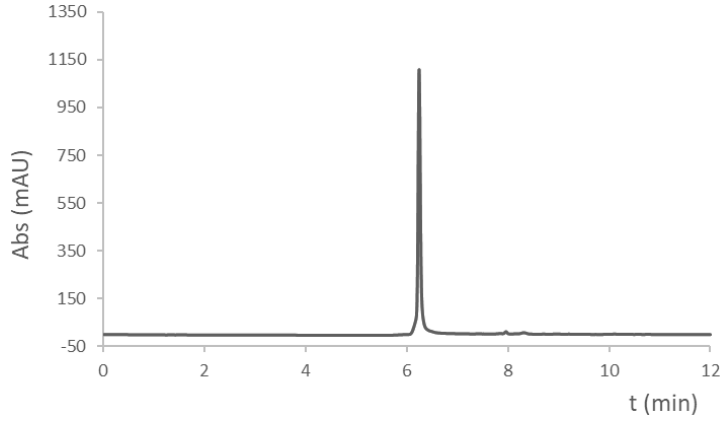


Figure S36. RP-HPLC chromatograms of oligomer **11** using a linear gradient from 5% B to 100% B in 10 min; A: 12.5 mM TEAA in water, pH 8.5; B: 12.5 mM TEAA in water:acetonitrile (1:2), pH 8.5.

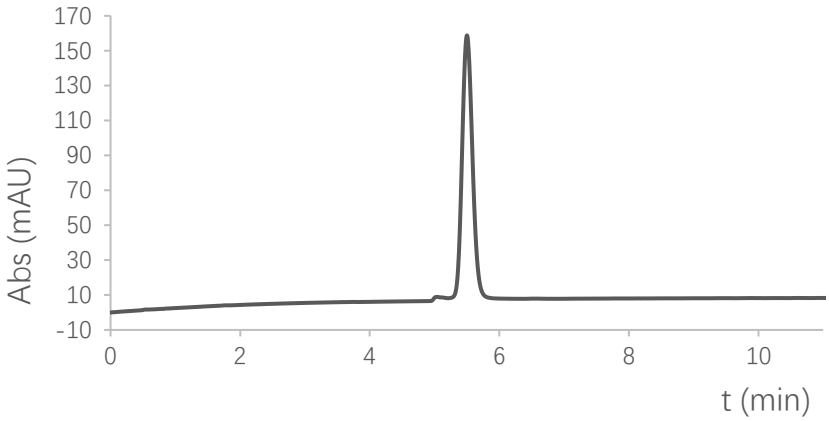


Figure S37. RP-HPLC chromatograms of oligomer **12** using a linear gradient from 0% B to 25% B in 10 min; A: 12.5 mmol NH_4OAc in water, pH 8.5; B: acetonitrile.

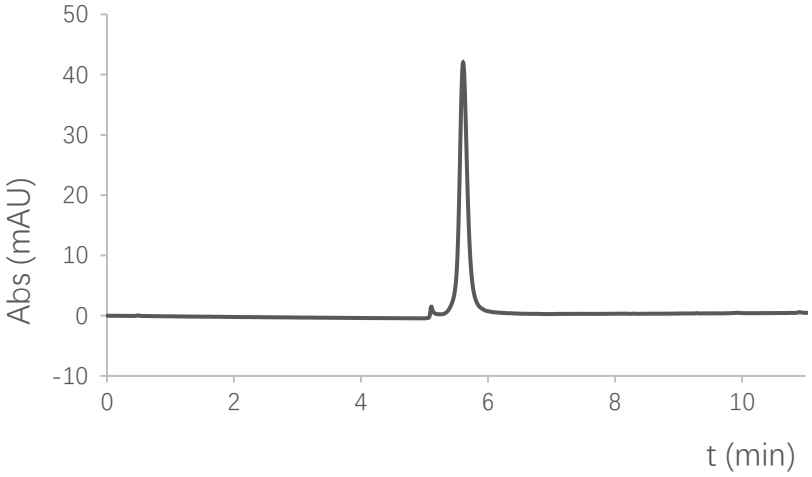


Figure S38. RP-HPLC chromatograms of oligomer **13** using a linear gradient from 0% B to 25% B in 10 min; A: 12.5 mmol NH_4OAc in water, pH 8.5; B: acetonitrile.

5. X-ray Crystallography

After preparative HPLC and ion exchange, the ammonium salt of oligomer **13** was dissolved in pure water to a concentration of 2.6 mM. Initially, **13** was mixed with Sac7d protein[20b] in 1:1 molar ratio to reach a 1.3 mM concentration. Various crystallization conditions were screened using the hanging drop vapor diffusion method. A precipitate formed in crystallization reagent composed of 20% PEG 6000, 0.1 M HEPES pH 7.0, 200 mM NaCl, but diamond-shaped crystals appeared in 30 days at 20 °C (Figure S39), which proved to be compound **13**. A single crystal was cryo-protected using 25% Glucose (*w/v*) prepared in crystallization reagent and flash-cooled in liquid nitrogen.

X-ray diffraction data were collected at 100K on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble) using a Dectris Eiger2 X 16M detector.[29] The dataset was processed using autoPROC pipeline (Global Phasing).[30] The crystal belonged to $P4_3\ 2_1\ 2$ (or $P4_1\ 2_1\ 2$) with 2 molecules in the asymmetric unit. The structure was solved by the Molecular Replacement (MR) method using PHASER[31] with a molecular model of compound **13**. The molecular model of **13** was built in Maestro (Version 11.5) based on previously crystallized foldamers containing a repeat of Q^{Pho} and M^{Pho} monomers (unpublished results) to which B^{Rme}, Q^{Sem}, Q^{Ala} and Q^{Asp} units were added at C terminus. Finally, the model was energy minimized using parameters described in Table S1. A single MR solution was found with a translation function Z-score (TFZ) of 16.0 and log-likelihood gain (LLG) of 223.12 in $P4_3\ 2_1\ 2$. Geometric restraints for each monomer were generated separately in eLBOW (Phenix suite).[32] Model building and refinement were performed in Phenix-Refine and Coot.[33] However, it is noteworthy that for chain F, the N-terminus monomer Q^{Pho} remained invisible in the electron density, even after multiple rounds of refinements. Consequently, it was not modeled. This observation raises the possibility that the Q^{Pho} monomer located at the N-terminus of chain F adopts an alternate conformation within the crystal lattice, leading to its absence in the observable electron density. In contrast, for chain A, the N-terminus Q^{Pho} monomer was readily discernible within the electron density maps during the refinement process and therefore, it was modeled accordingly. Data collection and structure refinement statistics are given in Table S2. The structure was deposited in PDB database with accession ID 8QHM.

Table S1. Parameters used to build the molecular model of compound **13**.

Forcefield	OPLS3
Solvent	Water
Charges from	Force Field
Cutoff	Extended
Method	PRCG
Converge on	Gradient
Convergence threshold	0.05
Minimization mode	Minimization of non-conformers
Maximum iterations	2500

Table S2. Crystallography data collection and structure refinement statistics for compound **13**.

Wavelength	0.95
Resolution range	19.43 - 3.0 (3.107 - 3.0)
Space group	P 43 21 2
Unit cell	74.146 74.146 83.693 90 90 90
Total reflections	21196 (1975)
Unique reflections	4874 (395)
Multiplicity	4.3 (4.1)
Completeness (%)	94.53 (81.28)
Mean I/sigma (I)	15.98 (5.02)
Wilson B-factor	14.79
R-merge	0.06324 (0.234)
R-meas	0.07193 (0.2674)
R-pim	0.03343 (0.1265)
CC1/2	0.998 (0.984)
CC*	0.999 (0.996)
Reflections used in refinement	4748 (395)
Reflections used for R-free	245 (15)
R-work	0.3479 (0.2287)
R-free	0.3821 (0.5667)
CC (work)	0.530 (0.572)
CC (free)	0.606 (0.519)
Number of non-hydrogen atoms	609
RMS (bonds)	0.029
RMS (angles)	5.63
Clash score	18.06
Average B-factor	19.74

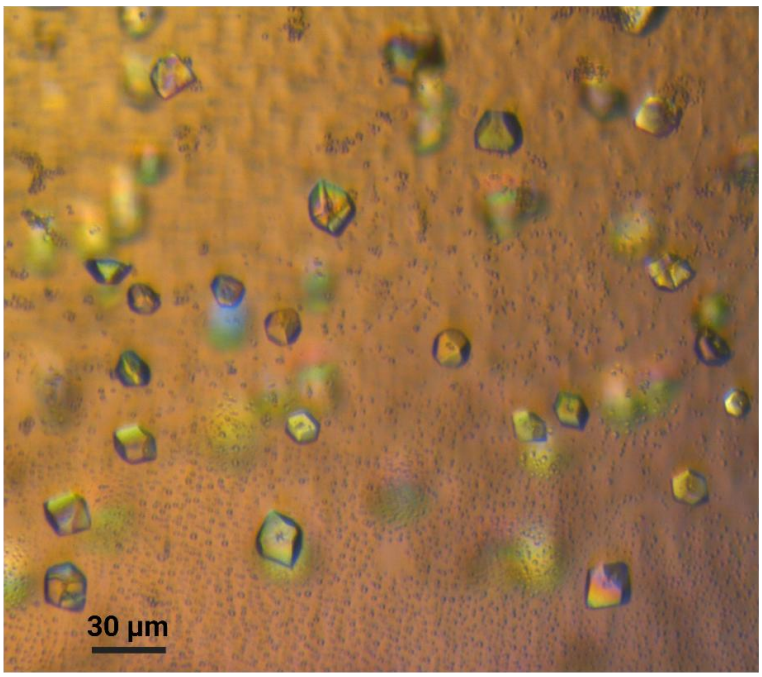


Figure S39. Crystals of compound **13** observed under crossed polarizing microscope.

6. References

[27] J. Buratto, C. Colombo, M. Stupfel, S. J. Dawson, C. Dolain, B. Langlois d'Estaintot, L. Fischer, T. Granier, M. Laguerre, B. Gallois, I. Huc, *Angew. Chem. Int. Ed.* **2014**, *53*, 883-887; *Angew. Chem.* **2014**, *126*, 902-906.

[28] X. Hu, S. J. Dawson, Y. Nagaoka, A. Tanatani, I. Huc, *J. Org. Chem.* **2016**, *81*, 1137-1150.

[29] D. Nurizzo, T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J.-C. Biasci, S. McSweeney, E. Mitchell, *J. Synchrotron Rad.* **2006**, *13*, 227-238.

[30] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, G. Bricogne, *Acta Cryst. D* **2011**, *67*, 293-302.

[31] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Cryst.* **2007**, *40*, 658-674.

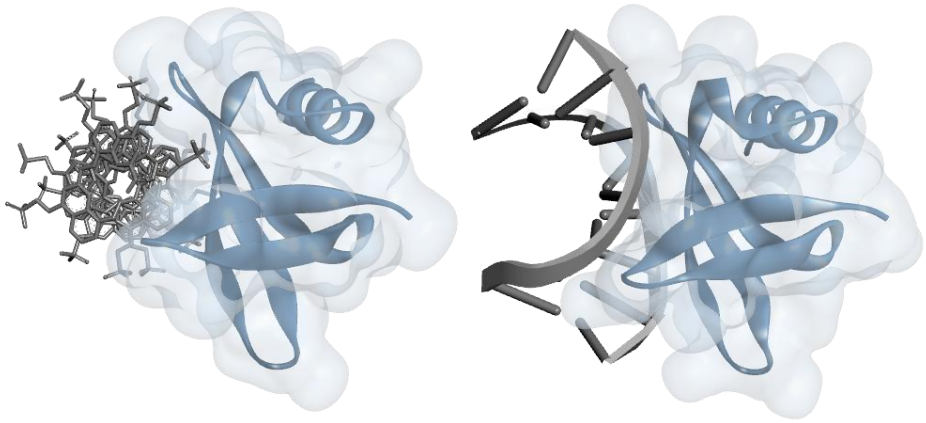
[32] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Cryst. D* **2019**, *75*, 861-877.

[33] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Cryst. D* **2010**, *66*, 486-501

6. DNA mimic foldamer recognition of a chromosomal protein

Based on our precious findings , we attempted to modify the foldamer structures to investigate their interaction with DNA-binding proteins. For this study, we chose Sac7d, a stable, soluble, and well-characterized chromosomal protein, as our model system. More importantly, it can be easily expressed in the laboratory. Our initial tests confirmed the strong binding of DNA mimic foldamers to Sac7d. A series of experiments including SPR, CD, ITC , AFM, NMR titrations and UV melting studies, demonstrated that foldamers not only act as a perfect "copy" of DNA, but also outcompete DNA for binding to Sac7d (Figure 18).

Inspired by these results, we set out to get structural information about how Sac7d interacts with DNA mimic foldamers. We included two new features, which are C_2-symmetry and handedness bias, into the foldamer design, which we expected to promote the crystal growth. To our surprise, Sac7d crystalize with a C_2-symmetrical 16-mer. The resulting crystal structure revealed that foldamers targeted the DNA binding site of Sac7d, but does so by a different orientation and without being kinked.



Sac7d-Foldamer Complexes ***VS.*** Sac7d-DNA Complexes

Figure 18. DNA mimic foldamers bind to some DNA-binding proteins better than DNA itself.

Our findings, published on *Angew. Chem. Int. Ed,* represent a significant advancement in the field. This work presents the first crystal structure of a protein in complex with a DNA mimic foldamer, which confirmed that the new design features are indeed beneficial for crystal growth. Furthermore, we have also visualize for the first time the foldamers outcompete DNA for binding to Sac7d, and

reduce DNA's thermal stability promoted by Sac7d. All these findings collectively demonstrated that 1) DNA mimic foldamers bind to Sac7d better than its DNA counterparts. 2) the interaction is diastereoselective and takes place at the DNA binding site, but through a different binding mode.

Contributions: DD and JW performed protein expression and purification. DD performed crystal growth and crystallographic analysis. VC and JW synthesized the foldamers. VC performed CD experiments. JW used ITC to measure the thermodynamic profile of foldamer binding to the protein. JW developed the UV melting studies. LA and VC led NMR experiments. JW assigned the 3D TOCSY and NOESY. MK and PT performed and analysed AFM observations. IH supervised the project. All authors commented and approved the manuscript.

6.1. Publication

DNA mimic foldamer recognition of a chromosomal protein

Authors: D. Deepak, **J. Wu**[+], V. Corvaglia[+], L. Allmendinger, M. Scheckenbach, P. Tinnefeld and I. Huc*

[+] These authors contributed equally

Published: *Angew. Chem. Int. Ed.* **2025**, 64, e202422958

DOI: doi.org/10.1002/anie.202422958.

How to cite: *Angew. Chem. Int. Ed.* **2025**, *64*, e202422958
doi.org/10.1002/anie.202422958

Foldamers

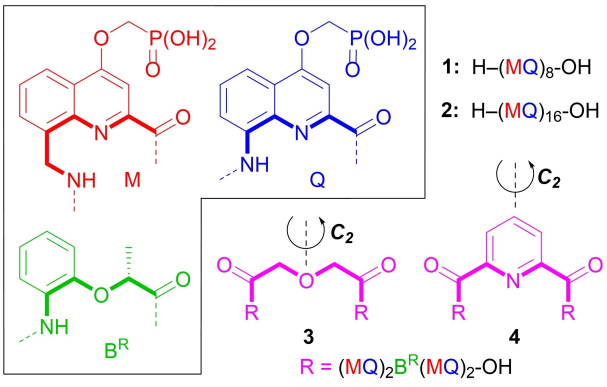
DNA Mimic Foldamer Recognition of a Chromosomal Protein

*Deepak Deepak, Jiaojiao Wu[+], Valentina Corvaglia[+], Lars Allmendinger, Michael Scheckenbach, Philip Tinnefeld, and Ivan Huc**

Abstract: Helical aromatic oligoamide foldamers bearing anionic side chains that mimic the overall shape and charge surface distribution of DNA were synthesized. Their interactions with chromosomal protein Sac7d, a non-sequence-selective DNA-binder that kinks DNA, were investigated by Surface Plasmon Resonance (SPR), Isothermal Titration Calorimetry (ITC), Circular Dichroism spectroscopy (CD), melting curve analysis, Atomic Force Microscopy (AFM), and Nuclear Magnetic Resonance (NMR), as well as by single crystal X-ray crystallography. The foldamers were shown to bind to Sac7d better than a DNA duplex of comparable length. The interaction is diastereoselective and takes place at the DNA binding site. Crystallography revealed that the DNA mimic foldamers have a binding mode of their own and that they can bind to Sac7d without being kinked.

The concept of biomolecular mimicry is familiar but not always fully understood. A mimic reproduces some features of the biomolecule from which it is originally inspired. However, a mimic is also generally intended to outcompete the original through its differences. A perfect "copy" that would match the original would not do better. Thus, designing mimics that are structurally remote from the original is challenging but it can be rewarding because differences may represent opportunities to outcompete the original. α-Helix mimetics[1] that target some protein surfaces and DNA analogs[2] that hybridize with DNA better than DNA itself are prominent examples of successful biomimicry. In contrast, molecules that mimic the surface of a DNA double helix and competitively inhibit DNA-protein interactions are underdeveloped despite their potential for pharmacological applications.

DNA mimicry exists in nature in the form of DNA mimic proteins typically rich in aspartic acid and glutamic acid residues.[3] Inspired by nature, the use of artificial proteins as DNA mimics has been initiated but not extensively pursued.[4] The so-called decoy oligonucleotides (ODNs) may also be used to target DNA-binding proteins such as transcription factors. These ODNs are modified to enhance their bioavailability, but they may not bind to their target better than their natural counterparts.[5] In this context, we have introduced abiotic aromatic oligoamides bearing anionic phosphonate side chains that fold in water into single helices whose shape and charge distribution mimic the shape and charge distribution of double-stranded B-DNA.[6] For example, sequences **1** and **2** (Scheme 1, Figure S1) are equivalent to eight and sixteen base-pair (bp) B-DNA duplexes, respectively. However, being already monomeric, they may not melt into two single strands as DNA does. Such foldamers do not possess sequence features other than the alternation of M and Q monomers



Scheme 1. Amino acid quinoline monomers M, Q, and B^R and oligoamide sequences **1**–**4** used in this study. Bold bonds indicate the inner rim of the helically folded conformations. Sequences **1** and **2** have both an N- and a C-terminus. Sequences **3** and **4** are C_2-symmetrical and have two C-termini.

[*] Dr. D. Deepak, J. Wu,[+] Dr. V. Corvaglia,[+] Dr. L. Allmendinger, Prof. Dr. I. Huc
Department of Pharmacy, Ludwig-Maximilians-Universität München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: ivan.huc@cup.lmu.de
M. Scheckenbach, Prof. Dr. P. Tinnefeld
Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstr. 5–13, 81377 München (Germany)
Dr. V. Corvaglia[+]
Current address: Institute for Stem-Cell Biology, Regenerative Medicine and Innovative Therapies, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Italy) & Center for Nanomedicine and Tissue Engineering (CNTE), ASST Grande Ospedale Metropolitano Niguarda, Milan (Italy)
[+] These authors contributed equally to this work

(Scheme 1). They have been shown to affect chromatin composition and perturb cell cycle progression[6c] and to bind to some non-sequence selective DNA-binding proteins with high affinity, leading to the competitive inhibition of therapeutically relevant enzymes such as topoisomerase 1 (Top1) and HIV integrase 1 (HIV-IN) even in the presence of a large excess of DNA substrate.[6b] These effects show some selectivity: binding is not effective on all non-sequence-selective DNA-binding proteins.[6b] Furthermore, structural modifications of the foldamers have been shown to enhance selectivity for Top1 or HIV-IN.[6a] However, the lack of detailed structural information on how these molecules recognize proteins hampers the development of better and more selective binders, including towards proteins that do not bind DNA sequence selectively. Here, we present the first structural investigation of interactions between a DNA mimic foldamer and, as a model system, Sac7d, a bacterial chromosomal protein known to bind and kink DNA non-sequence selectively in hyperthermophilic archaeon *Sulfolobus acidocaldarius*.[7] In this study, we demonstrate that foldamers outcompete DNA in targeting the DNA binding site of Sac7d by adopting a distinct binding orientation and without being kinked.

Sac7d was selected as a model protein for structural investigations with DNA mimic foldamers because of its high structural stability, solubility, and rich literature on how it interacts with DNA.[7b,e,g,j] As an initial test, we assessed the binding of **1** and **2** to Sac7d by SPR with the protein immobilized on the chip and each foldamer in the mobile phase (Figure 1a, Figures S2–S3). Steady-state data indicated a K_d in the one-digit micromolar range. This is weaker binding than that of similar foldamers to e.g. HIV-IN,[6b] but it is to be compared to the two-digit micromolar binding of Sac7d to calf thymus DNA.[7a] However, the sensorgrams did not fit well with the kinetics of a single binding event. A possible reason for this is that **1** and **2** do not contain any stereogenic center and thus exist as a racemic mixture of left-handed (*M*) and right-handed (*P*) helices, accounting for at least two types of interactions with the protein. ITC confirmed binding of Sac7d to **1** in the low to sub-micromolar range (Figure S4) and showed that the interaction is enthalpy driven ($\Delta H = -6.9$ kcal.mol^{-1}, $-T\Delta S =$ 1.57 kcal.mol^{-1}), in contrast with the Sac7d-DNA interaction which is entropy driven and endothermic.[7k,8]

Since **1** and **2** are achiral, their CD spectra show no signal. However, when incubating **1** and **2** with Sac7d, a CD band emerged at 365 nm as a function of time (Figure 1b, Figures S5–S6), indicating aromatic helix handedness bias by interconversion of one enantiomeric conformer into the other. The negative sign of the band implies a preference for *M* helicity,[9] that is, for the enantiomer that mimics the right-handed B-DNA double helix (Figure S1). Several conclusions can be drawn from this observation: (*i*) DNA mimic foldamer *P* and *M* helices interconvert, although slowly. This was unexpected, considering the lengths of **1** and **2** and the kinetic inertness of related foldamers.[10] The faster dynamics may result from negative charge repulsions; (*ii*) CD confirms that the foldamers bind to Sac7d; (*iii*) the interaction with Sac7d is diastereoselective and favors the

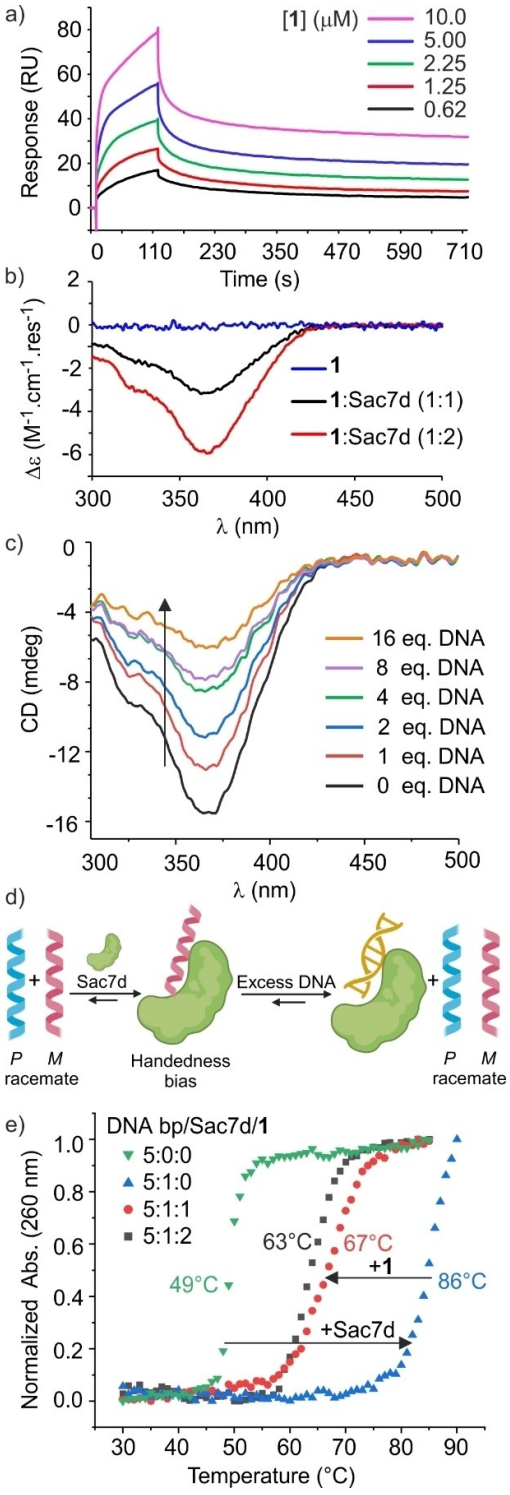


Figure 1. Evidence of binding of **1** to Sac7d. a) SPR sensorgrams of the interaction between His_6-tagged Sac7d and **1** (pH 7.4, 25 °C). Sensorgrams were plotted after subtraction of the signal of the reference flow cell. b) CD spectra of **1** (40 μM in 50 mM NH_4HCO_3, pH 8.5, 20 °C) in the presence of 0, 1, or 2 equiv. of Sac7d after 24 h equilibration. c) CD spectra of a 1:1 mixture of **1** and Sac7d in the presence of increasing amounts of a 10-bp DNA. d) Cartoon representation of helix handedness bias in **1** upon binding to Sac7d and of the competitive association of DNA. e) DNA melting profiles of poly[dAdT]*poly[dAdT] (500 μM bp concentration) in presence of different molar ratio of Sac7d and **1**, monitored at 260 nm.

enantiomer that mimics B-DNA, not its mirror image; and (*iv*) the intensity of the band (the value of Δε/residue in Figure 1b) suggests that bias is extensive (the estimated diastereomeric excess is 75 %).[11]

Fitting the CD data to a 1:1 binding isotherm (Figure S6) yielded a K_d of 34 μM, apparently higher than estimated by SPR.[12] Nevertheless, CD spectroscopy also confirmed that DNA binds less effectively than the foldamers. Thus, a competition experiment was set in which a 10-bp DNA duplex was added to the already equilibrated *M*-**1**-Sac7d complex. The intensity of the negative CD band at 365 nm gradually decreased as the DNA-Sac7d complex forms and the released foldamer helix racemizes (Figure 1c,d). From the quantity of DNA necessary to make CD intensity drop by half (~4 equiv. with respect to the foldamer), one can estimate that **1** binds at least tenfold better than the 10-bp DNA (see Supporting Information). Following this, we tested whether a DNA mimic foldamer could inhibit Sac7d functions. One effect of Sac7d binding to B-DNA is to considerably enhance DNA thermal stability, increasing the melting temperature of poly[dAdT]*poly-[dAdT] by more than 30 °C.[7f] Upon adding foldamer **1**, we found that this effect is largely diminished (Figure 1e). These experiments also showed that the foldamer undergoes no apparent melting (Figure S7). Another effect of Sac7d binding is to kink DNA which is thought to help package DNA in *Sulfolobus acidocaldarius*.[7] Inspired by work on Abf2p, another DNA-binding and compacting protein,[13] we used high-resolution AFM to demonstrate that Sac7d indeed promotes DNA compaction and that this is reverted in presence of foldamer **1** (Figures 2, S8).

We next investigated DNA mimic foldamer recognition of Sac7d using NMR spectroscopy. Foldamers **3** and **4** were designed and synthesized for this purpose (Scheme 1).[14] These sequences are based on a central diacid linker that makes them C_2-symmetrical, i.e., palindromic-like so that they would produce the same type of complex with Sac7d irrespective of their orientation. They also possess two chiral B^R residues (Scheme 1) that quantitatively bias their handedness towards the *M* helix.[11]

Sequence **3** possesses an aliphatic more flexible central linker than the aromatic linker of **4**. This difference was introduced in relation to the ability of Sac7d to kink DNA, a flexible linker possibly playing the role of a hinge. Several 2D and 3D NMR experiments on ^{15}N-labelled Sac7d using ^{1}H–^{15}N HSQC (Hetronuclear Single Quantum Coherence) spectroscopy, ^{15}N-HSQC-NOESY (Nuclear Overhauser Effect Spectroscopy) and ^{15}N-HSQC-TOCSY (Total Correlation Spectroscopy) allowed for the unambiguous assignment of the protein backbone (Figure S9). Titrations of Sac7d by **3** and **4** were then monitored by ^{1}H–^{15}N HSQC (Figures 3a–c, Figures S10–S13). No significant differences were observed between the two foldamers **3** and **4**. Both caused similar chemical shift variations, indicating molecular associations in fast exchange on the NMR timescale. Some chemical shift variations followed a monotonous trend (straight arrows in Figure 3a) which could in principle be fitted to a 1:1 binding isotherm. However, others were not monotonous (kinked arrows in Figure 3b), indicating that at least two complexes with different molecularities form, e.g. 1:1 and 1:2. Mapping the chemical shift perturbations (CSPs) on the surface of the Sac7d structure revealed that perturbations caused by the foldamers occur at and around the beta-sheet DNA binding site and involved some key residues for DNA recognition (W24, V26, M29, and R42). In contrast, the opposite face of the protein was essentially unaltered. This strongly supports that the DNA-mimic foldamers also interact with the DNA-binding region of Sac7d.

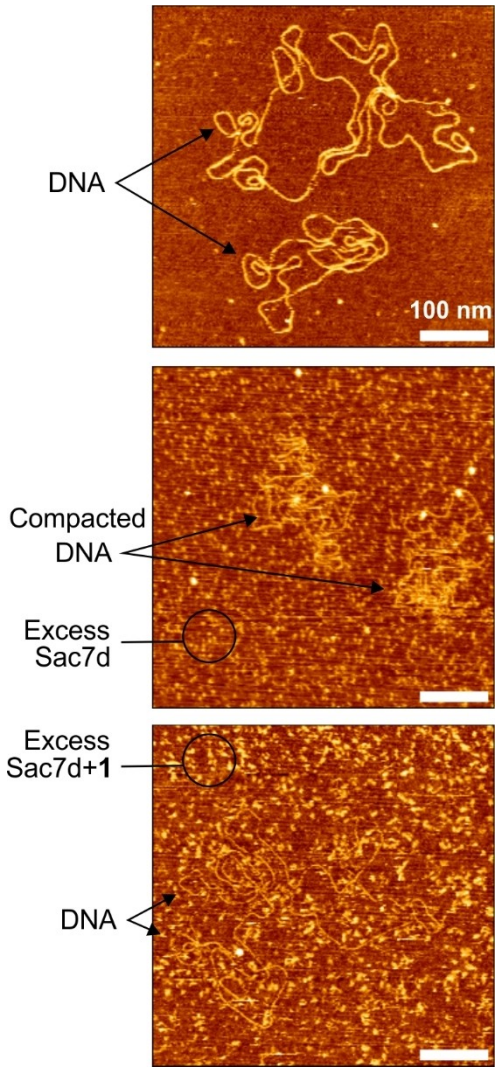


Figure 2. Representative high resolution AFM images showing DNA alone (top), compacted DNA in presence of Sac7d (middle) and uncompacted DNA in presence of Sac7d and **1** (bottom). Dots corresponding to excess Sac7d and Sac7d + **1** somewhat blur the middle and bottom images. See Figure S8 for details.

Finally, we endeavored to crystallize Sac7d-foldamer complexes. As for complexes between DNA and non-sequence selective DNA-binding proteins, this was challenged by the possible degeneracy of the binding modes via frame shifts of one bp (here one MQ dimer) that do not favor crystal growth. In **3** and **4**, this degeneracy was mitigated by the C_2-symmetry and the presence of other

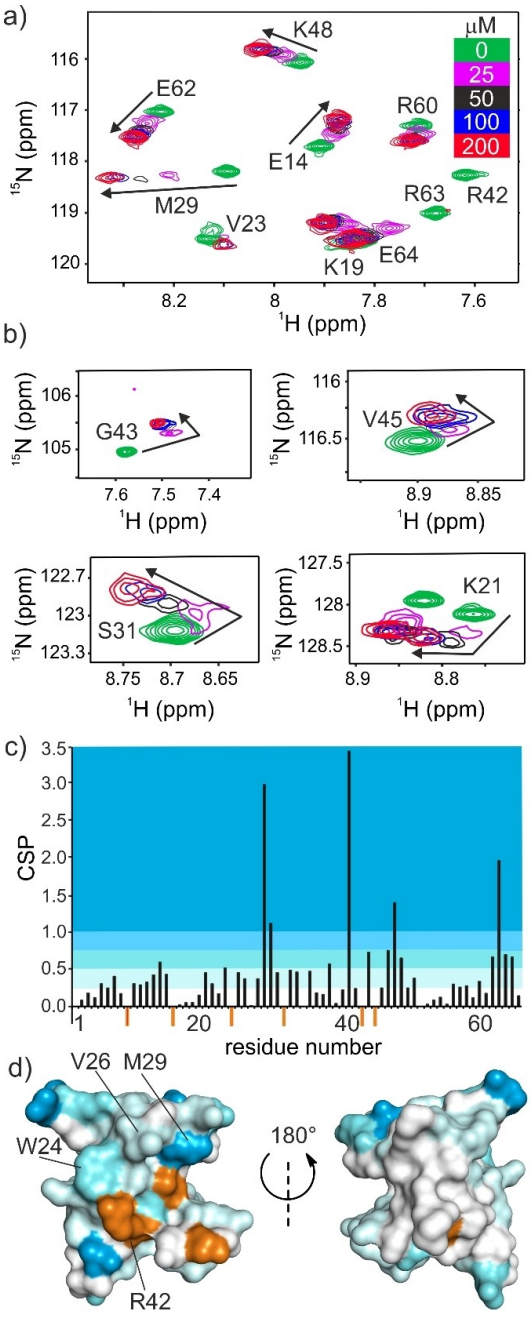


Figure 3. NMR spectroscopic evidence of binding of **3** to Sac7d at 100 μM concentration in 50 mM Tris-d_{11} buffer pH 7.5, 50 mM KCl, 10% D_2O. a, b) Part of ^1H–^{15}N HSQC titration of [^{15}N]-Sac7d with **3**. The colored scale indicates an increasing concentration of **3**. c) CSPs of [^{15}N]-Sac7d backbone amide ^1H-^{15}N HSQC in the presence of **3** (200 μM). CSPs were calculated as the root-mean-square deviation $((\Delta\delta_H)/0.14)^2+(\Delta\delta_N)^2)^{0.5}$. d) Protein surface of Sac7d crystal structure[7g] colored according to CSP values as in panel c. The signals of residues shown in orange broaden to the extent that they disappear during the titration, which is interpreted as a strong chemical shift perturbation.

units than M and Q. Crystals of a Sac7d in complex with **3** were obtained (Figures 4a, S14) and diffracted at 2.6 Å. The structure was solved by molecular replacement using the Sac7d structure[7g] and a molecular model of the foldamer fragment as the search model (see Supporting Information). The refined structure revealed the formation of a C_2-symmetrical 2:1 complex with two proteins binding to a molecule of **3** (Figure 4b). This stoichiometry may explain the trend of CSPs during NMR titrations (Figure 4b). Thus, a crystallographic C_2 axis crosses the middle of the linker of **3**. The asymmetric unit, therefore, contains half of this 2:1 complex, along with half of a second molecule of **3** not interacting with Sac7d. Crystal packing involved columnar pseudo-continuous stacks of foldamer helices resembling the stacks of DNA duplexes often observed in crystals of protein-DNA complexes (Figure S15). The structure confirmed that the foldamer extensively interacts with the DNA binding region of Sac7d, i.e. through contact area of 1500 Å. However, the orientation of the helix of **3** is almost perpendicular to that of DNA (Figure 4c,d). The protein-foldamer interface features multiple hydrogen bonds including charge-reinforced hydrogen bonds involving phosphonates and C-terminal carboxylates of **3**, as well as some hydrophobic contacts (Figure 4e–h). Many residues involved are also key residues for Sac7d-DNA interactions (e.g., Y8, W24, and R42, see Figure S16). Nevertheless, the Sac7d-**3** complex has a unique geometry reflecting the structural features of the foldamer. For example, the B^R forms a hydrophobic contact with the protein surface allowed by the lack of a side chain of that residue. If B^R was replaced by Q or M, it would lead to a steric clash.

The sharp 61° kinking of DNA by Sac7d is mediated by V26 and M29, which protrude from the protein and intercalate between bp's (Figure 4j).[7b] In the complex with **3**, the side chain of M29 sticks out of the intermolecular interface while the side chain of V26 fills a cavity in the foldamer structure created by the small size of the central linker (Figure 4i). Due to the C_2-symmetry axis crossing the linker, it is in fact, two V26 side chains from two proteins that fill this cavity. As a result, the foldamer helix binds to Sac7d without being kinked, a situation that is thought to be transient with DNA.[7b,j]

All attempts to crystallize **4** in complex with Sac7d failed: precipitates were obtained using conditions under which crystals of Sac7d-**3** grew. Clearly, the larger central linker of **4** would fill the space occupied by the two V26 side chains of the structure of **3**, hampering the binding geometry observed with **3**. Nevertheless, NMR showed that **4** also binds to Sac7d. The double mutant Sac7d (*V26A/M29A*), which lacks the side chains responsible for DNA kinking, was crystallized in complex with **3** in a structure otherwise identical to the Sac7d-**3** complex (Figure S17). Furthermore, Sac7d (*V26A/M29A*) also crystallized in complex with **4**. The structure is again similar to that of the Sac7d-**3** complex, with the central pyridine ring of **4** filling the space made available by the V26 A mutation (Figure S18).

In conclusion, DNA mimic foldamers represent a novel platform to target the large ensemble of proteins that interact with nucleic acids. They complement other aromatic foldamers[15] macrocycles[16] and tweezers[17] that have also been shown to bind to protein surfaces. The structure of the Sac7d-**3** complex provides a solid foundation for the structure-based design of the foldamer to enhance binding

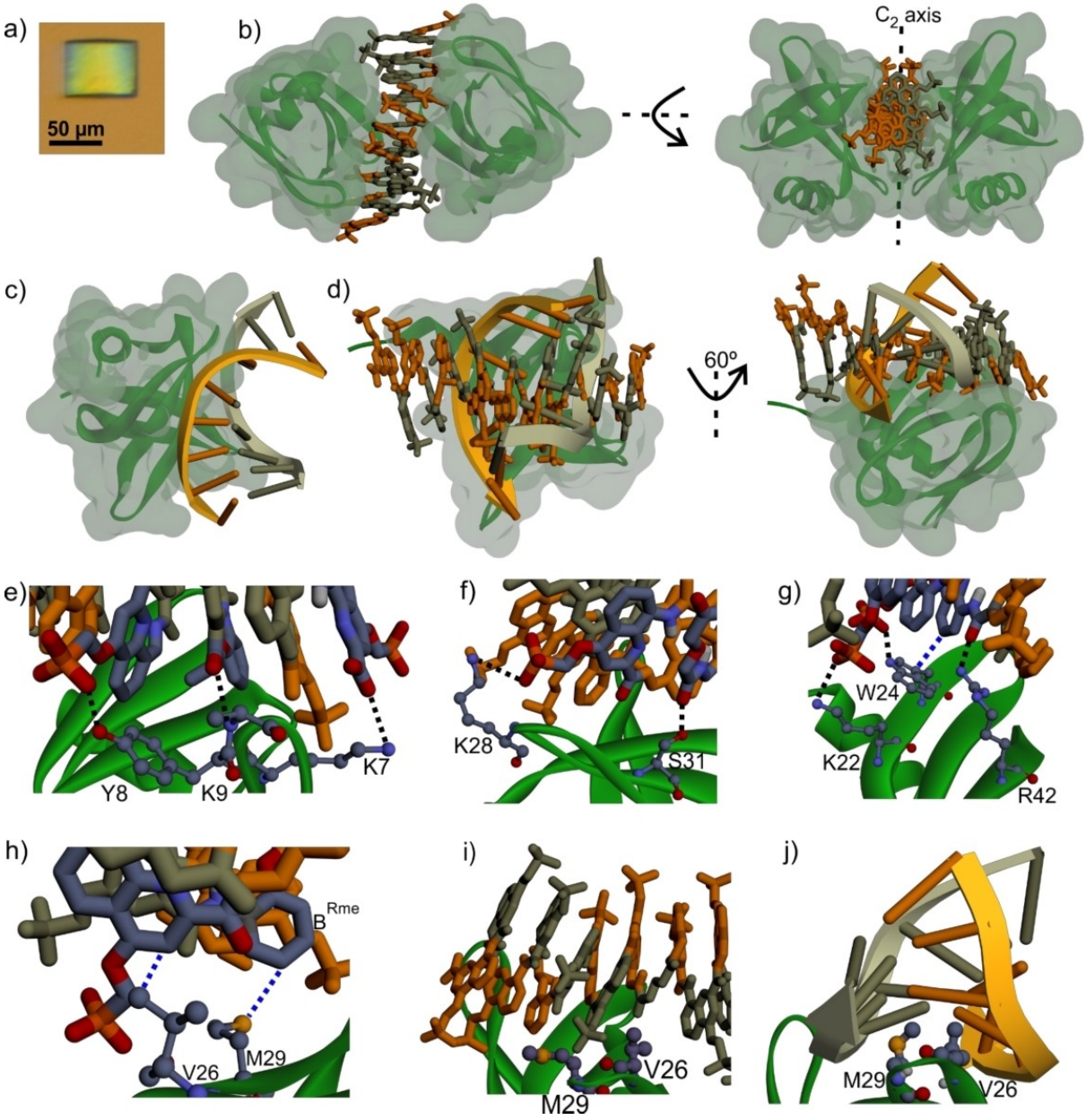


Figure 4. Crystal structure of Sac7d-**3** complex (PDB# 8CMN) and comparison with a Sac7d-DNA complex (PDB# 1AZQ).[7g] a) X-ray diffraction-quality Sac7d-**3** crystal. b) 2:1 Sac7d-**3** complex. c) Sac7d-DNA complex. d) Sac7d-**3** complex superimposed (superimposition α-carbons) with the Sac7d-DNA complex. e–i) Key interactions within the Sac7d-**3** complex. Blue and black dashed lines highlight hydrophobic contacts and H-bonds, respectively. j) DNA intercalation of V26 and M29 in the Sac7d-DNA complex. Sac7d is shown in green ribbon representation. A transparent green iso-surface has been added in b–d). Relevant residues are shown in ball and stick representation in e–j). In c), d), and j) DNA is shown in ribbon representation. Foldamer **3** is shown in tube representation with alternating beige and orange residues. In e–j), relevant residues have classical atom colors (C: dark gray, O: red, N: blue, orange: P).

affinity and specificity, exploiting the inherent modularity of foldamer sequences. For example, we consider the distinct protein-foldamer orientation as an opportunity to further design the foldamer so that it would not interact with other DNA-binding proteins. The specific features of the foldamer-Sac7d complex would have been hard to predict, rendering any further design attempts without structural information a rather elusive endeavor. We have now demonstrated that protein-DNA mimic foldamer complexes can be crystallized and deliver detailed structural information even in the absence of sequence specific recognition features. We are currently extending this work to other proteins with the long-term objective of creating DNA mimic foldamers highly specific to their protein target.

Supporting Information

The authors have cited additional references within the Supporting Information.[18–31]

Author Contributions

DD and JW performed protein expression and purification. DD performed crystal growth and crystallographic analysis. VC and JW synthesized the foldamers. VC performed CD experiments. JW performed ITC and melting studies. LA and VC led NMR experiments and analysis. JW was involved in TOCSY and NOESY assignment. MK and PT performed and analyzed AFM observations. IH supervised the project. All authors commented and approved the manuscript.

Acknowledgements

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) via projects HU1766/2-1 and 325871075 (CRC1309-C7), from the European Research Council (ERC) under the European Union's Horizon Europe Framework Programme (Grant Agreement No. ERC-2021-ADG-320892), from the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder through the ONE MUNICH Project Munich Multiscale Biofabrication, from the German Academic Exchange Service (DAAD, predoctoral fellowship to D. D.), and from the China Scholarship Council (CSC, predoctoral fellowship to J. W.). This work has benefited from the facilities and expertise of the Biophysical and Structural Chemistry platform (BPCS) at IECB, CNRS UMS3033, Inserm US001, and Bordeaux University. We thank L. Minder, T. Chakrabortty, and V. Morozov for assistance with SPR, protein expression and purification, and ITC, respectively, and P. K. Mandal for preliminary investigations. We thank L. McGregor (ID23-1, ESRF), M.S. Lopez (ID30B, ESRF), and I. Ibento (EMBL P13, Petra III, DESY) for assistance during data collection at synchrotron beamlines. We thank Prof. T. Carell for access to their UV/Vis spectrophotometer for T_m measurements. We also thank Prof. C. Y. Chen (Hubei University, Taiwan) for kindly providing us with the expression vector for Sac7d and Sac7d V26A/M29A proteins. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: Chromosomal Proteins · DNA Mimicry · Foldamers · Protein Recognition · Structure Elucidation

[1] a) I. Saraogi, A. D. Hamilton, *Biochem. Soc. Trans.* **2008**, *36*, 1414–1417; b) H. Wang, R. S. Dawber, P. Zhang, M. Walko, A. J. Wilson, X. Wang, *Chem. Sci.* **2021**, *12*, 5977–5993; c) V. Azzarito, K. Long, N. S. Murphy, A. J. Wilson, *Nat. Chem.* **2013**, *5*, 161–173; d) M. Pelay-Gimeno, A. Glas, O. Koch, T. N. Grossmann, *Angew. Chem. Int. Ed.* **2015**, *54*, 8896–8927; *Angew. Chem.* **2015**, *127*, 9022–9054; e) L.-G. Milroy, T. N. Grossmann, S. Hennig, L. Brunsveld, C. Ottmann, *Chem. Rev.* **2014**, *114*, 4695–4748.

[2] a) P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497–1500; b) W. Li, H. Shi, B. Dong, K. Nie, Z. Liu, N. He, *Curr. Med. Chem.* **2016**, *23*, 4681–4705; c) C. Sharma, S. K. Awasthi, *Chem. Biol. Drug Des.* **2017**, *89*, 16–37; d) A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen, J. Wengel, *Tetrahedron* **1998**, *54*, 3607–3630; e) S. Obika, D. Nanbu, Y. Hari, K. Morio, Y. In, T. Ishida, T. Imanishi, *Tetrahedron Lett.* **1997**, *38*, 8735–8738; f) P. H. Hagedorn, R. Persson, E. D. Funder, N. Albæk, S. L. Diemer, D. J. Hansen, M. R. Møller, N. Papargyri, H. Christiansen, B. R. Hansen, H. F. Hansen, M. A. Jensen, T. Koch, *Drug Discovery Today* **2018**, *23*, 101–114.

[3] a) S. S. Hegde, M. W. Vetting, S. L. Roderick, L. A. Mitchenall, A. Maxwell, H. E. Takiff, J. S. Blanchard, *Science* **2005**, *308*, 1480–1483; b) P. A. Tsonis, B. Dwivedi, *Biochim. Biophys. Acta Mol. Cell Res.* **2008**, *1783*, 177–187; c) H.-C. Wang, C.-C. Chou, K.-C. Hsu, C.-H. Lee, A. H. J. Wang, *IUBMB Life* **2019**, *71*, 539–548; d) H.-C. Wang, C.-H. Ho, K.-C. Hsu, J.-M. Yang, A. H.-J. Wang, *Biochemistry* **2014**, *53*, 2865–2874; e) H.-C. Wang, H.-C. Wang, T.-P. Ko, Y.-M. Lee, J.-H. Leu, C.-H. Ho, W.-P. Huang, C.-F. Lo, A. H. J. Wang, *Proc. Nat. Acad. Sci.* **2008**, *105*, 20758–20763; f) M. D. Walkinshaw, P. Taylor, S. S. Sturrock, C. Atanasiu, T. Berge, R. M. Henderson, J. M. Edwardson, D. T. Dryden, *Mol. Cell.* **2002**, *9*, 187–194; g) F. Ye, I. Kotta-Loizou, M. Jovanovic, X. Liu, D. T. Dryden, M. Buck, X. Zhang, *eLife* **2020**, *9*, e52125.

[4] a) D. Yüksel, P. R. Bianco, K. Kumar, *Mol. BioSyst.* **2016**, *12*, 169–177; b) F. Haque, C. Freniere, Q. Ye, N. Mani, E. M. Wilson-Kubalek, P.-I. Ku, R. A. Milligan, R. Subramanian, *Nat. Cell Biol.* **2022**, *24*, 1088–1098; c) D. T. Dryden, *Trends Biotechnol.* **2006**, *24*, 378–382; d) H. C. Wang, C. H. Ho, C. C. Chou, T. P. Ko, M. F. Huang, K. C. Hsu, A. H. Wang, *Nucleic Acids Res.* **2016**, *44*, 4440–4449.

[5] a) G. Casas, F. Perche, P. Midoux, C. Pichon, J. M. Malinge, *Mol. Ther. Nucleic Acids* **2022**, *29*, 162–175; b) B. Johari, J. Zargan, *Cell Biol. Int.* **2017**, *41*, 1335–1344; c) B. Johari, M. Moradi, *Methods Mol. Biol.* **2022**, *2521*, 207–230; d) R. Crinelli, M. Bianchi, L. Gentilini, M. Magnani, *Nucleic Acids Res.* **2002**, *30*, 2435–2443.

[6] a) V. Corvaglia, D. Carbajo, P. Prabhakaran, K. Ziach, P. K. Mandal, V. D. Santos, C. Legeay, R. Vogel, V. Parissi, P. Pourquier, I. Huc, *Nucleic Acids Res.* **2019**, *47*, 5511–5521; b) K. Ziach, C. Chollet, V. Parissi, P. Prabhakaran, M. Marchivie, V. Corvaglia, P. P. Bose, K. Laxmi-Reddy, F. Godde, J.-M. Schmitter, S. Chaignepain, P. Pourquier, I. Huc, *Nat. Chem.* **2018**, *10*, 511–518; c) V. Kleene, V. Corvaglia, E. Chacin, I. Forne, D. B. Konrad, P. Khosravani, C. Douat, C. F. Kurat, I. Huc, A. Imhof, *Nucleic Acids Res.* **2023**, *51*, 9629–9642.

[7] a) J. G. McAfee, S. P. Edmondson, I. Zegar, J. W. Shriver, *Biochemistry* **1996**, *35*, 4034–4045; b) C.-Y. Chen, T.-P. Ko, T.-W. Lin, C.-C. Chou, C.-J. Chen, A. H.-J. Wang, *Nucleic Acids Res.* **2005**, *33*, 430–438; c) Y.-G. Gao, S.-Y. Su, H. Robinson, S. Padmanabhan, L. Lim, B. S. McCrary, S. P. Edmondson, J. W. Shriver, A. H.-J. Wang, *Nat. Struct. Biol.* **1998**, *5*, 782–786;

d) V. Kalichuk, G. Béhar, A. Renodon-Cornière, G. Danovski, G. Obal, J. Barbet, B. Mouratou, F. Pecorari, *Sci. Rep.* **2016**, *6*, 37274; e) T.-P. Ko, H.-M. Chu, C.-Y. Chen, C.-C. Chou, A. H.-J. Wang, *Acta Crystallogr. Sect. D* **2004**, *60*, 1381–1387; f) J. G. McAfee, S. P. Edmondson, P. K. Datta, J. W. Shriver, R. Gupta, *Biochemistry* **1995**, *34*, 10063–10077; g) H. Robinson, Y.-G. Gao, B. S. McCrary, S. P. Edmondson, J. W. Shriver, A. H.-J. Wang, *Nature* **1998**, *392*, 202–205; h) R. K. Singh, A. Mukherjee, *J. Phys. Chem. B* **2022**, *126*, 1682–1690; i) J. Spiriti, A. van der Vaart, *ChemBioChem* **2013**, *14*, 1434–1437; j) M. Zacharias, *Angew. Chem. Int. Ed.* **2019**, *58*, 5967–5972; *Angew. Chem.* **2019**, *131*, 6028–6033; k) W. B. Peters, S. P. Edmondson, J. W. Shriver, *J. Mol. Biol.* **2004**, *343*, 339–360.

[8] The positive enthalpic term is thought to reflect the cost of kinking DNA in reference 7k. In contrast, Sac7d can bind to foldamers without kinking them.

[9] C. Dolain, H. Jiang, J.-M. Léger, P. Guionneau, I. Huc, *J. Am. Chem. Soc.* **2005**, *127*, 12943–12951.

[10] M. Vallade, P. Sai Reddy, L. Fischer, I. Huc, *Eur. J. Org. Chem.* **2018**, *2018*, 5489–5498.

[11] D. Bindl, E. Heinemann, P. K. Mandal, I. Huc, *Chem. Commun.* **2021**, *57*, 5662–5665.

[12] The equilibrium between the *P* and *M* helix of an achiral oligomer in solution make the apparent K_d towards the *M* helix only be twice larger than it really is. This value may thus be corrected by a factor of two.

[13] R. W. Friddle, J. E. Klare, S. S. Martin, M. Corzett, R. Balhorn, E. P. Baldwin, R. J. Baskin, A. Noy, *Biophys. J.* **2004**, *86*, 1632–1639.

[14] V. Corvaglia, J. Wu, D. Deepak, M. Loos, I. Huc, *Chem. Eur. J.* **2024**, *30*, e202303650.

[15] a) S. Kumar, M. Birol, D. E. Schlamadinger, S. P. Wojcik, E. Rhoades, A. D. Miranker, *Nat. Commun.* **2016**, *7*, 11412; b) P. S. Reddy, B. Langlois d'Estaintot, T. Granier, C. D. Mackereth, L. Fischer, I. Huc, *Chem. Eur. J.* **2019**, *25*, 11042–11047; c) J. P. Plante, T. Burnley, B. Malkova, M. E. Webb, S. L. Warriner, T. A. Edwards, A. J. Wilson, *Chem. Commun.* **2009**, 5091–5093; d) S. Kumar, A. D. Hamilton, *J. Am. Chem. Soc.* **2017**, *139*, 5744–5755.

[16] a) P. B. Crowley, *Acc. Chem. Res.* **2022**, *55*, 2019–2032; b) R. E. McGovern, H. Fernandes, A. R. Khan, N. P. Power, P. B. Crowley, *Nat. Chem.* **2012**, *4*, 527–533.

[17] a) X. Guillory, I. Hadrović, P. J. de Vink, A. Sowislok, L. Brunsveld, T. Schrader, C. Ottmann, *J. Am. Chem. Soc.* **2021**, *143*, 13495–13500; b) H. Shahpasand-Kroner, I. Siddique, R. Malik, G. R. Linares, M. I. Ivanova, J. Ichida, T. Weil, J. Münch, E. Sanchez-Garcia, F.-G. Klärner, T. Schrader, G. Bitan, *Pharmacol. Rev.* **2023**, *75*, 263–308.

[18] I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, D. Herschlag, *eLife* **2020**, *9*, e57264.

[19] J. L. Bedell, B. S. McCrary, S. P. Edmondson, J. W. Shriver, *Protein Sci.* **2000**, *9*, 1878–1888.

[20] L. Mureddu, G. W. Vuister, *FEBS J.* **2019**, *286*, 2035–2042.

[21] D. Nurizzo, T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J. C. Biasci, S. McSweeney, E. Mitchell, *J. Synchrotron Radiat.* **2006**, *13*, 227–238.

[22] W. Kabsch, *Acta Crystallogr. Sect. D* **2010**, *66*, 125–132.

[23] A. Vagin, A. Lebedev, *Acta Crystallogr.* **2015**, *A71*, s19.

[24] A. Correa, S. Pacheco, A. E. Mechaly, G. Obal, G. Béhar, B. Mouratou, P. Oppezzo, P. M. Alzari, F. Pecorari, *PLoS One* **2014**, *9*, e97438.

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

[26] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. Sect. D* **2010**, *66*, 486–501.

[27] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr. Sect. D* **2019**, *75*, 861–877.

[28] A. W. Schüttelkopf, D. M. F. van Aalten, *Acta Crystallogr. Sect. D* **2004**, *60*, 1355–1363.

[29] V. B. Chen, J. R. Wedell, R. K. Wenger, E. L. Ulrich, J. L. Markley, *J. Biomol. NMR* **2015**, *63*, 77–83.

[30] A. A. McCarthy, R. Barrett, A. Beteva, H. Caserotto, F. Dobias, F. Felisaz, T. Giraud, M. Guijarro, R. Janocha, A. Khadrouche, M. Lentini, G. A. Leonard, M. Lopez Marrero, S. Malbet-Monaco, S. McSweeney, D. Nurizzo, G. Papp, C. Rossi, J. Sinoir, C. Sorez, J. Surr, O. Svensson, U. Zander, F. Cipriani, P. Theveneau, C. Mueller-Dieckmann, *J. Synchrotron Radiat.* **2018**, *25*, 1249–1260.

[31] D. Nečas, P. Klapetek, *Open Phys.* **2012**, *10*, 181–188.

Manuscript received: November 25, 2024
Accepted manuscript online: December 23, 2024
Version of record online: January 10, 2025

6.2. Supplementary Information

For:

DNA mimic foldamer recognition of a chromosomal protein

D. Deepak, J. Wu[+], V. Corvaglia[+], L. Allmendinger, M. Scheckenbach, P. Tinnefeld and I. Huc*

Contents

1. List of Abbreviations

SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SPR	surface plasmon resonance
CD	circular dichroism
FPLC	fast protein liquid chromatography
EDTA	ethylenediamine tetraacetic acid
NMR	nuclear magnetic resonance
RT	room temperature
SEC	size exclusion chromatography
HSQC	heteronuclear single quantum coherence spectroscopy
TOCSY	total correlation spectroscopy
NOESY	nuclear overhauser effect spectroscopy
THF	tetrahydrofuran
UV	ultraviolet
D_2O	deuterium oxide
Tris	tris(hydroxymethyl)aminomethane

2. Supporting figures

2.1 Folding principle and DNA mimic foldamer design

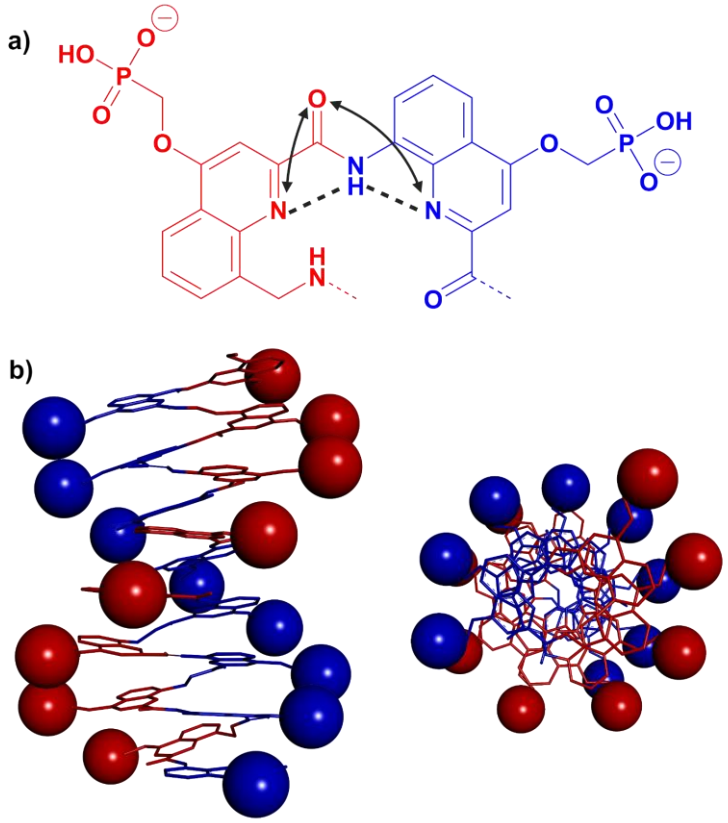


Figure S1. a) The folding principle of DNA mimic foldamers based on a MQ dimer. Dashed lines indicate hydrogen bonds, and arrows highlight electrostatic repulsions. These forces impart curvature to the main chain. Hydrophobic effects associated with aromatic stacking further stabilize the helical conformations. b) Molecular model of **1** with side view (left) and top view (right). M and Q units are color-coded in red and blue, respectively, in both a) and b). In **1** or **2,** an MQ dimer carries two phosphonate side chains and raises the helix by 3.4 Å along its axis and thus mimics a DNA base pair. Because of the geometrical parameters of M and Q, an MQ dimer spans a 0.9 helix turn,[6b] which means that the next MQ dimer in the sequence is shifted by a tenth of a turn backward along the helix backbone. Because the shift is backward, the handedness of the main chain single helix and the handedness of the double-helical array of phosphonate side chains are opposite. As shown in b), an M-foldamer helix displays a double P-exohelix of negatively charged side chains and thus mimics the P DNA double helix.

2.2 SPR and ITC binding studies

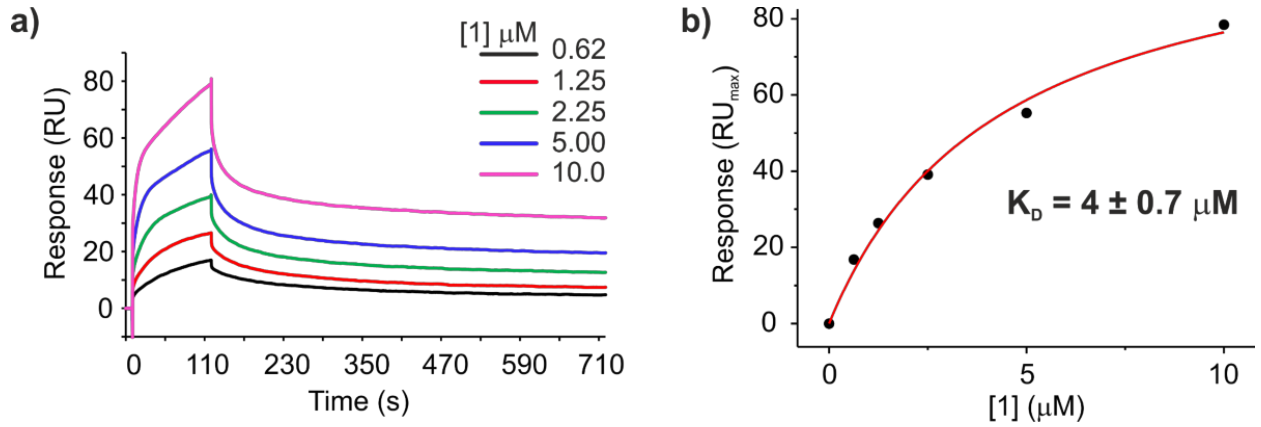


Figure S2. Binding of **1** to Sac7d quantified by SPR. a) SPR sensorgrams of the interaction between His_6-tagged Sac7d and **1** in HBS-EP buffer pH 7.4 at 25 °C. b) Curve fitting of the experimental data representing the maximum response unit values (steady state) plotted against foldamer concentration following the equation (1), see experimental section. K_d was found to be consistent across replicates. Note this value is only indicative of an order of magnitude of the association. The kinetic data do not fit well to a 1:1 binding model and indicate the system is more complex.

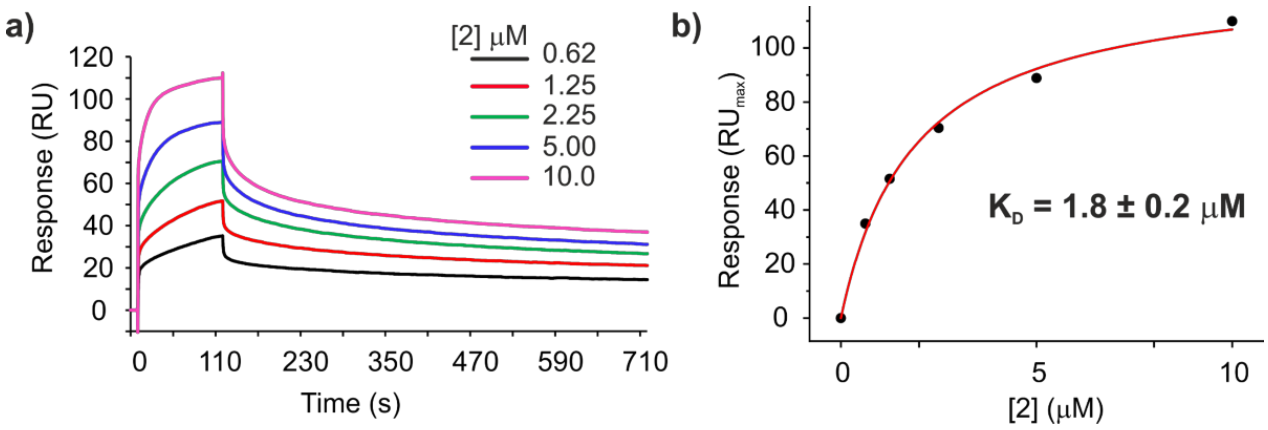


Figure S3. Binding of **2** to Sac7d quantified by SPR. a) SPR sensorgrams of the interaction between His_6-tagged Sac7d and **2** in HBS-EP buffer pH 7.4 at 25 °C. b) Curve fitting of the experimental data representing the maximum response unit values (steady state) plotted against foldamer concentration following the equation (1), see experimental section. K_d was found to be consistent across replicates. Note this value is only indicative of an order of magnitude of the association. The kinetic data do not fit well to a 1:1 binding model and indicate the system is more complex.

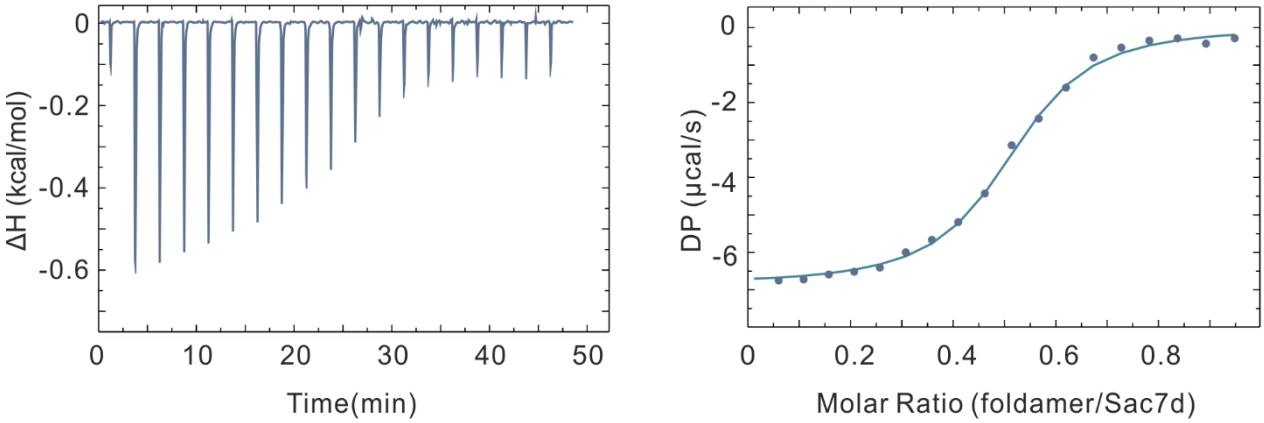


Figure S4. Isothermal titration calorimetry profile of foldamer binding to Sac7d at 25 ºC. Each peak in the upper panel resulted from the injection of **1** (0.25 mM) into a solution of Sac7d (0.1 mM). Both protein and foldamer were in 25 mM KCl and 10 mM KH_2PO_4 (pH 7). The panel at right shows the integrated peak intensities, giving a binding curve fitted by a single-site binding model, yielding a K_d value of 0.7 µM (two titrations showed reproducibility within less than 5%).

2.3 CD binding studies

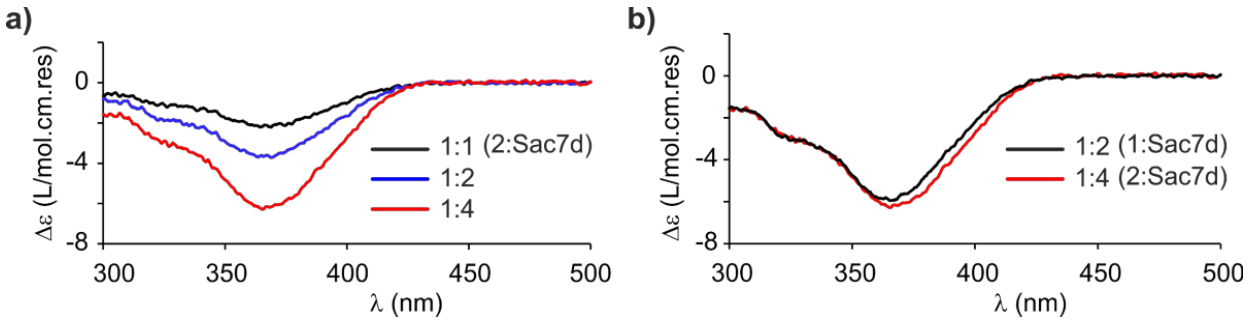


Figure S5. CD spectra of **2** in complex with Sac7d at 1:1, 1:2, and 1:4 foldamer: protein ratio (a) and of **1** and **2** overlaid (b) recorded in 50 mM NH_4HCO_3 buffer pH 8.5 at 20 °C after 24 h equilibration.

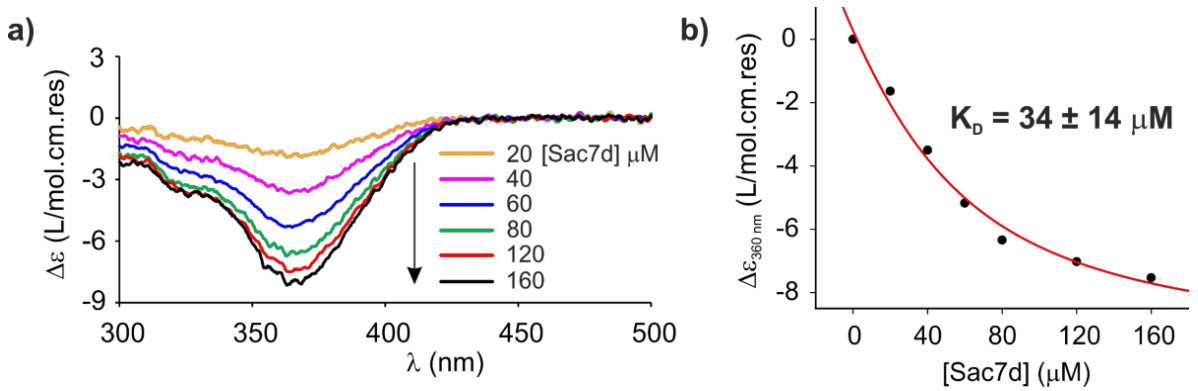


Figure S6. Binding of **1** to Sac7d quantified by CD. a) CD spectra of **1** (40 μM) in complex with Sac7d at different concentrations (from 20 to 160 μM) in 50 mM NH_4HCO_3 buffer pH 8.5 at 20 °C after 24 h equilibration. b) Curve fitting of the experimental data representing the maximum □□ values plotted against protein concentration to determine K_d following the quadratic binding equation.[18] K_d was found to be consistent across replicates.

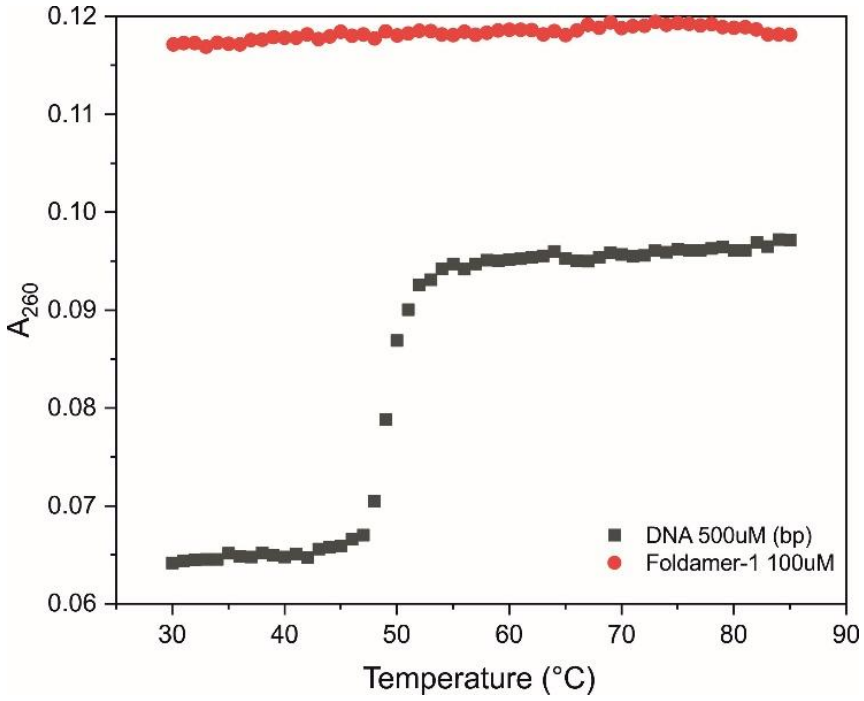


Figure S7. UV melting curves of DNA (in black) and foldamer 1 (in red) measured at 260 nm. Both DNA and foldamer 1 were in 10 mM KH_2PO_4 (pH 7).

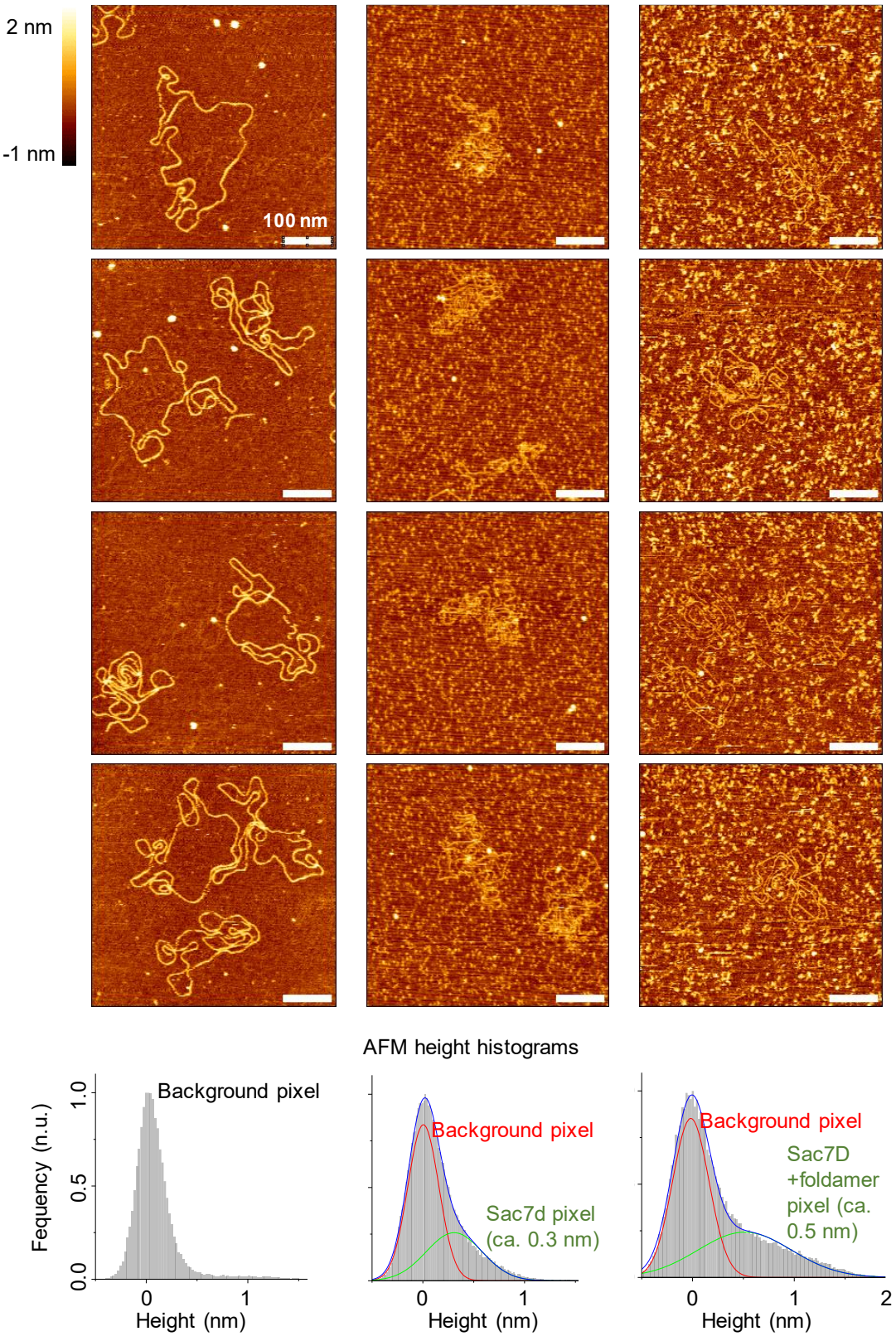


Figure S8. Representative high resolution AFM images of pBR322 DNA (22.9 nM, i.e. 100 µM b.p concentration, after 15 min incubation at 75° in PBS buffer). Left images: DNA alone. Middle images: DNA + Sac7D (100 µM). The DNA is compacted. Excess Sac7D blurs the background.

Right images: DNA + Sac7D (100 μM) + foldamer **1** (50 μM). The DNA is no longer compacted, the background is even more blurred. The height distributions (bottom graphs) show that the background particles are larger (higher) in presence of **1** (0.5 nm) than in its absence (0.3 nM), in agreement with the formation of Sac7D-foldamer complexes.

2.4 ^{1}H-^{15}N NMR titrations

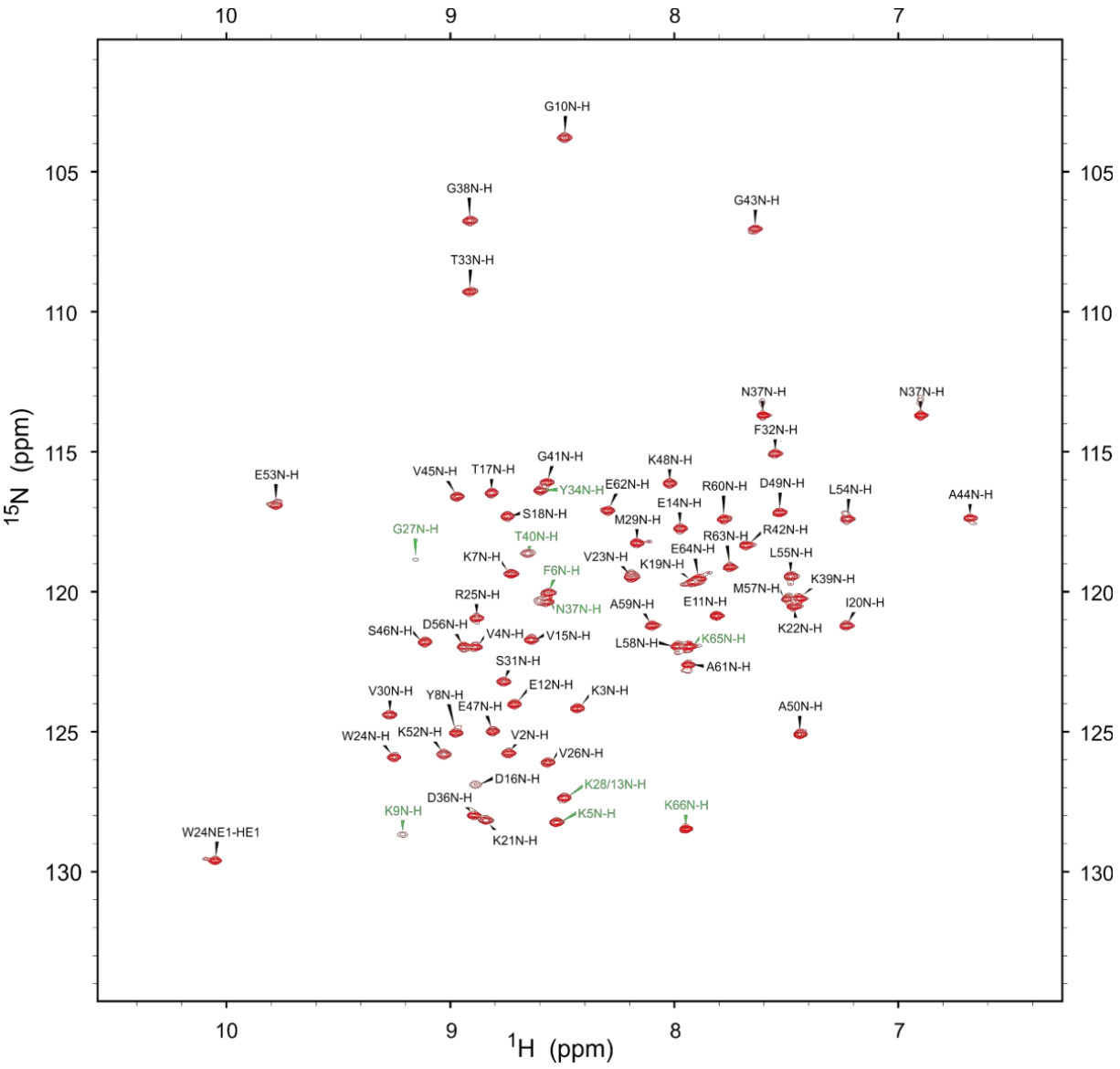


Figure S9. ^{1}H-^{15}N HSQC spectrum of [^{15}N]-Sac7d (500 µM) in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O. Assigned backbone amide signals were labeled with a single amino acid letter code followed by their position. In green are amide signals for which the assignment was not definite from our experiments. These signals were assigned by analogy with previously reported signals having similar chemical shift values. [19]

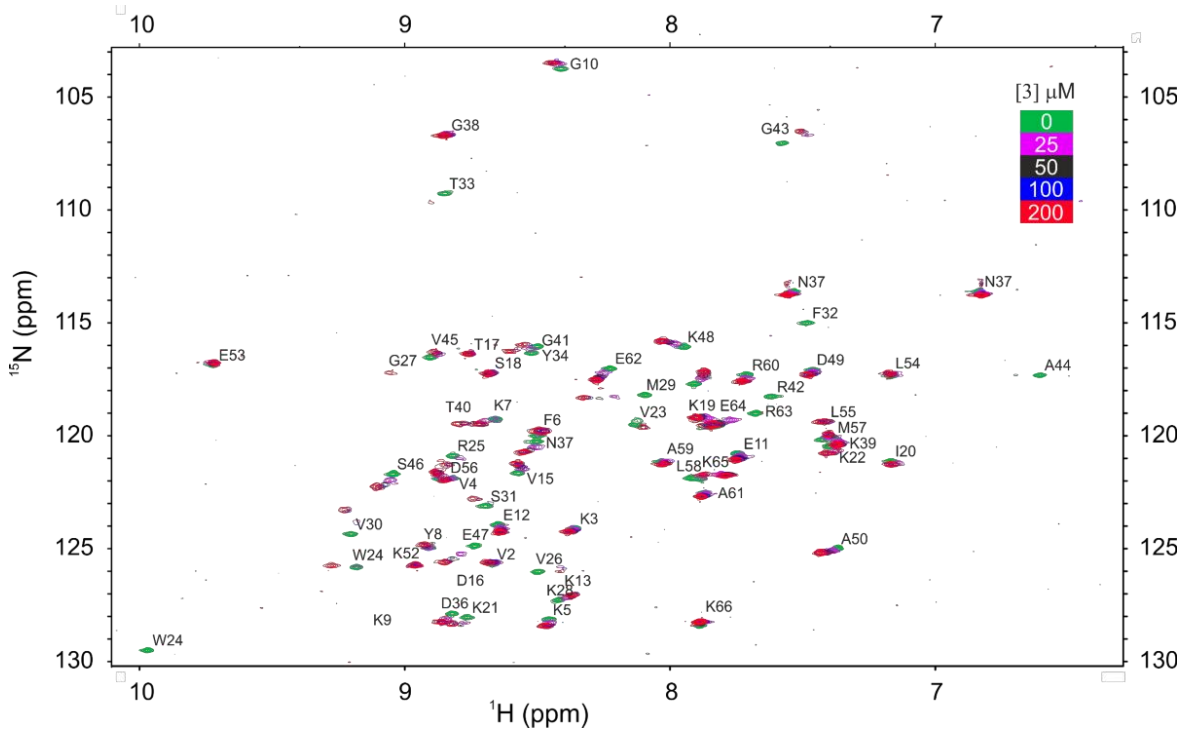


Figure S10. ^{1}H-^{15}N HSQC titration of [^{15}N]-Sac7d (100 µM) with **3** in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O. The colored scale indicates that an increasing concentration of **3** was added to Sac7d.

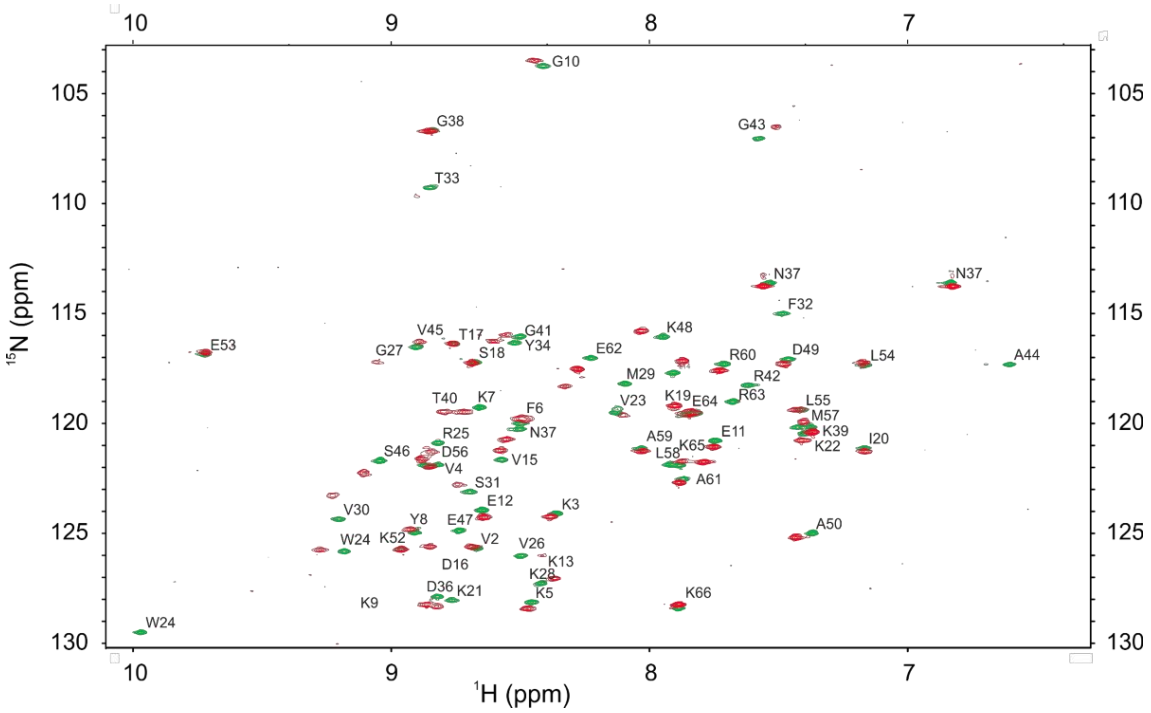


Figure S11. Superimposition of ^{1}H-^{15}N HSQC spectra of [^{15}N]-Sac7d (100 µM, green amide signals) and [^{15}N]-Sac7d with 2.0 eq. of **3** (200 µM, red amide signals) in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O.

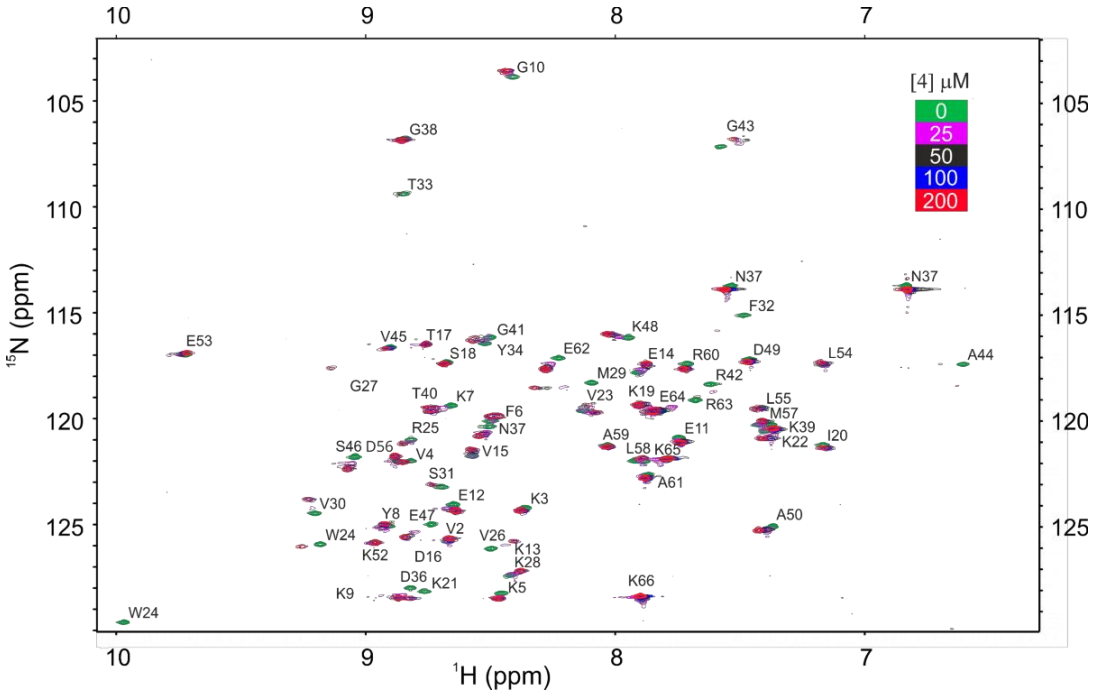


Figure S12. ^{1}H-^{15}N HSQC titration of [^{15}N]-Sac7d (100 µM) with **4** in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O. The colored scale indicates that an increasing concentration of **4** was added to Sac7d.

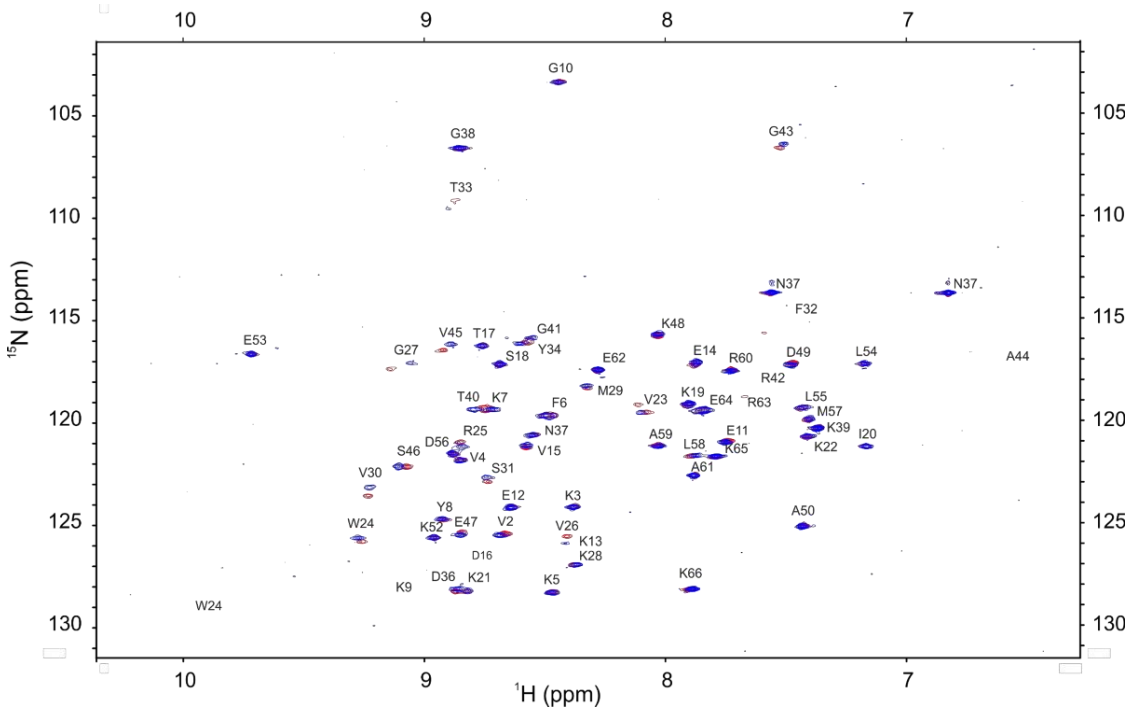


Figure S13. Superimposition of ^{1}H-^{15}N HSQC spectra of [^{15}N]-Sac7d (100 µM) with **3** (200 µM, blue amide signals) and **4** (200 µM, red amide signals) in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O.

2.5 X-ray crystallography

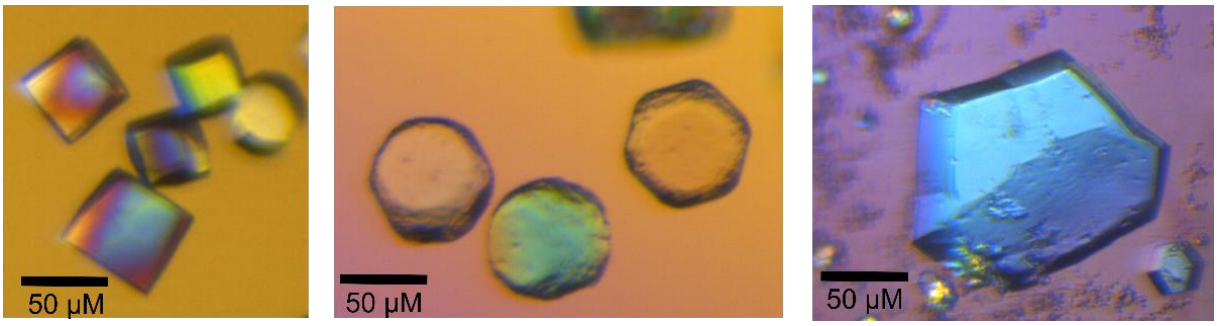


Figure S14. Crystals of Sac7d-**3**, Sac7d V26A/M29A-**3**, and Sac7d V26A/M29A-**4** from left to right observed under crossed polarizing microscope.

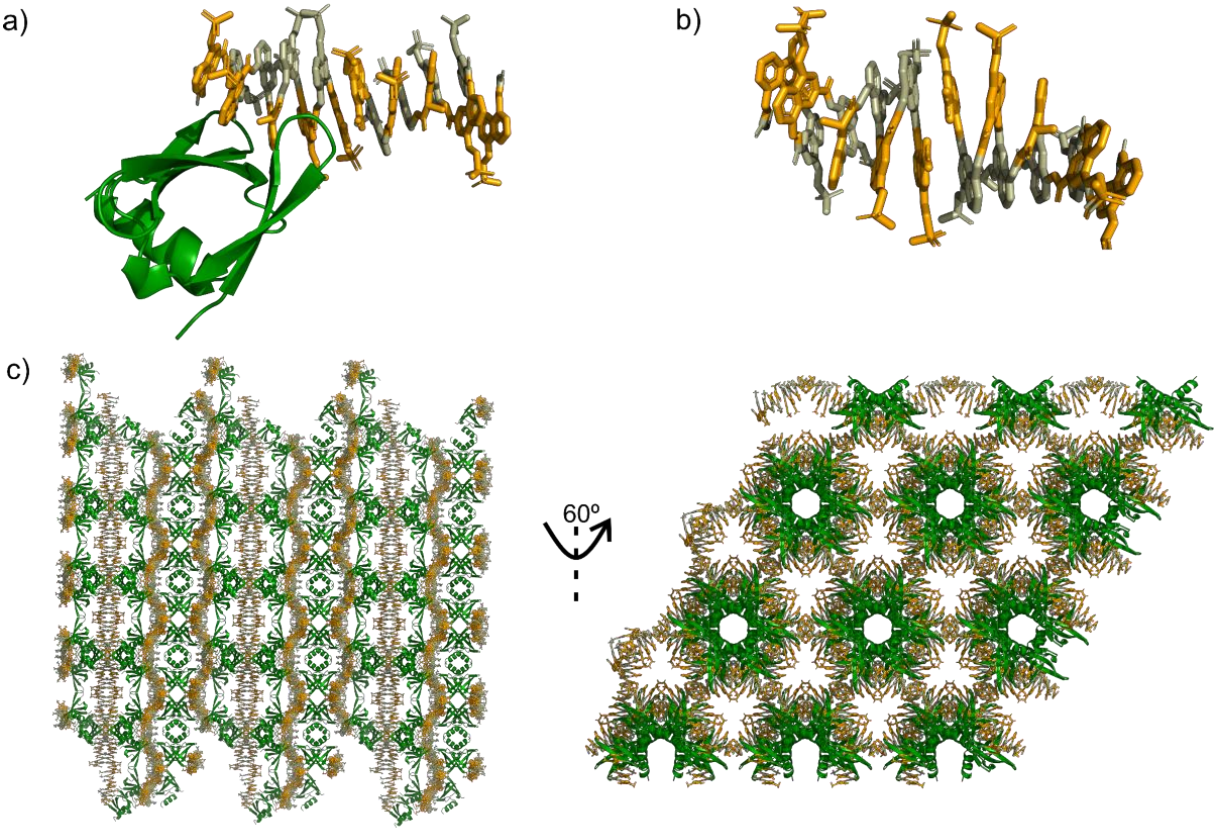


Figure S15. Sac7d-**3** crystal structure (PDB# 8CMN). a) Asymmetric unit showing Sac7d with two 9mer units from two 18mers of **3** in the ASU. b) C-terminal stacking of the two 9mer units in ASU of **3**. c) Two views of the Sac7d-**3** packing in the crystal lattice.

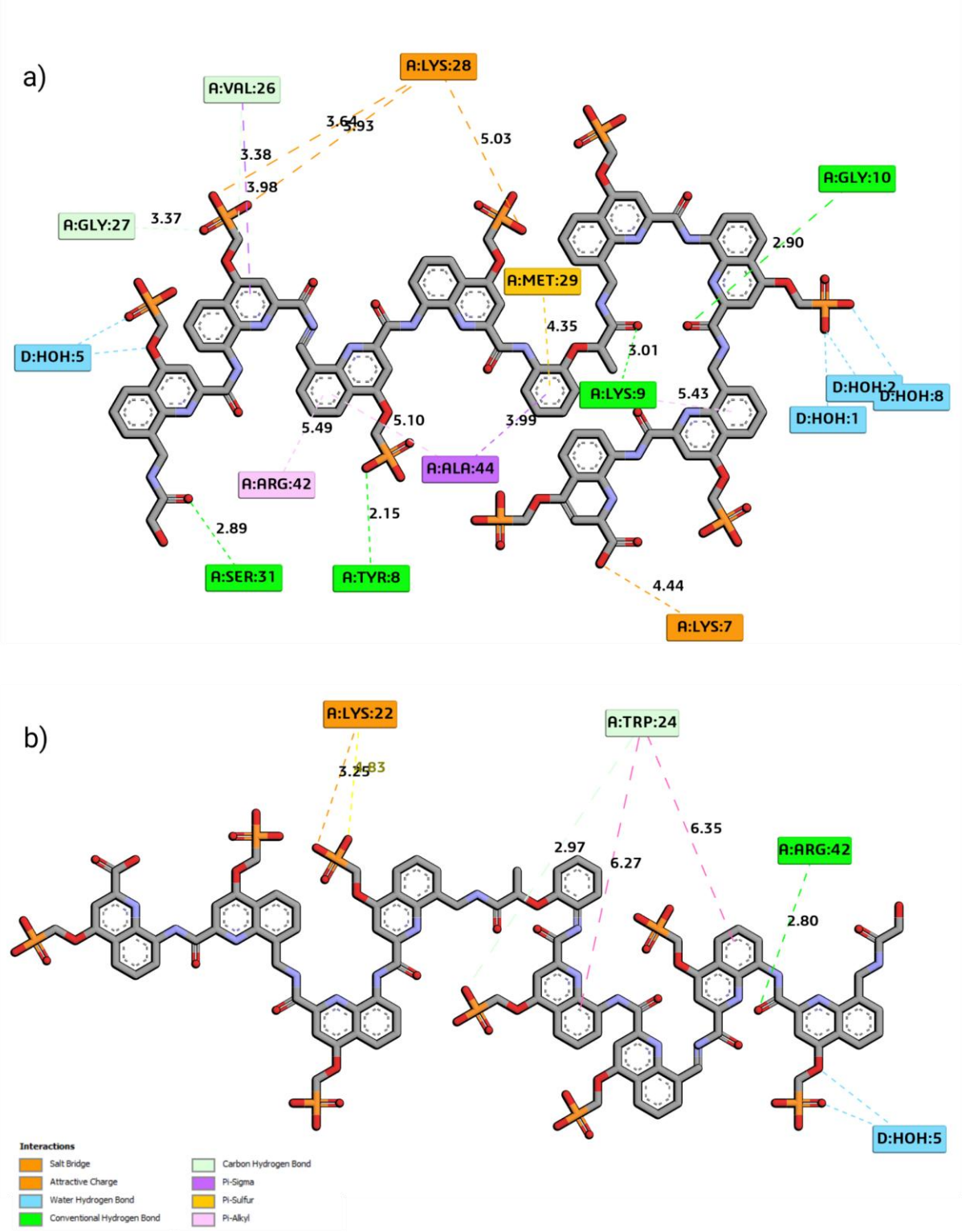


Figure S16. Ligplot-like representation of crystal contacts between Sac7d-**3**. a) Sac7d (chain A) interactions with **3** (chain B, 9mer). b) Sac7d (chain A) interaction with **3** (chain B' 9mer). The interaction map was generated using Discovery Studio (Dassault Systems). **3** is shown in 2D stick representation, colored according to elements.

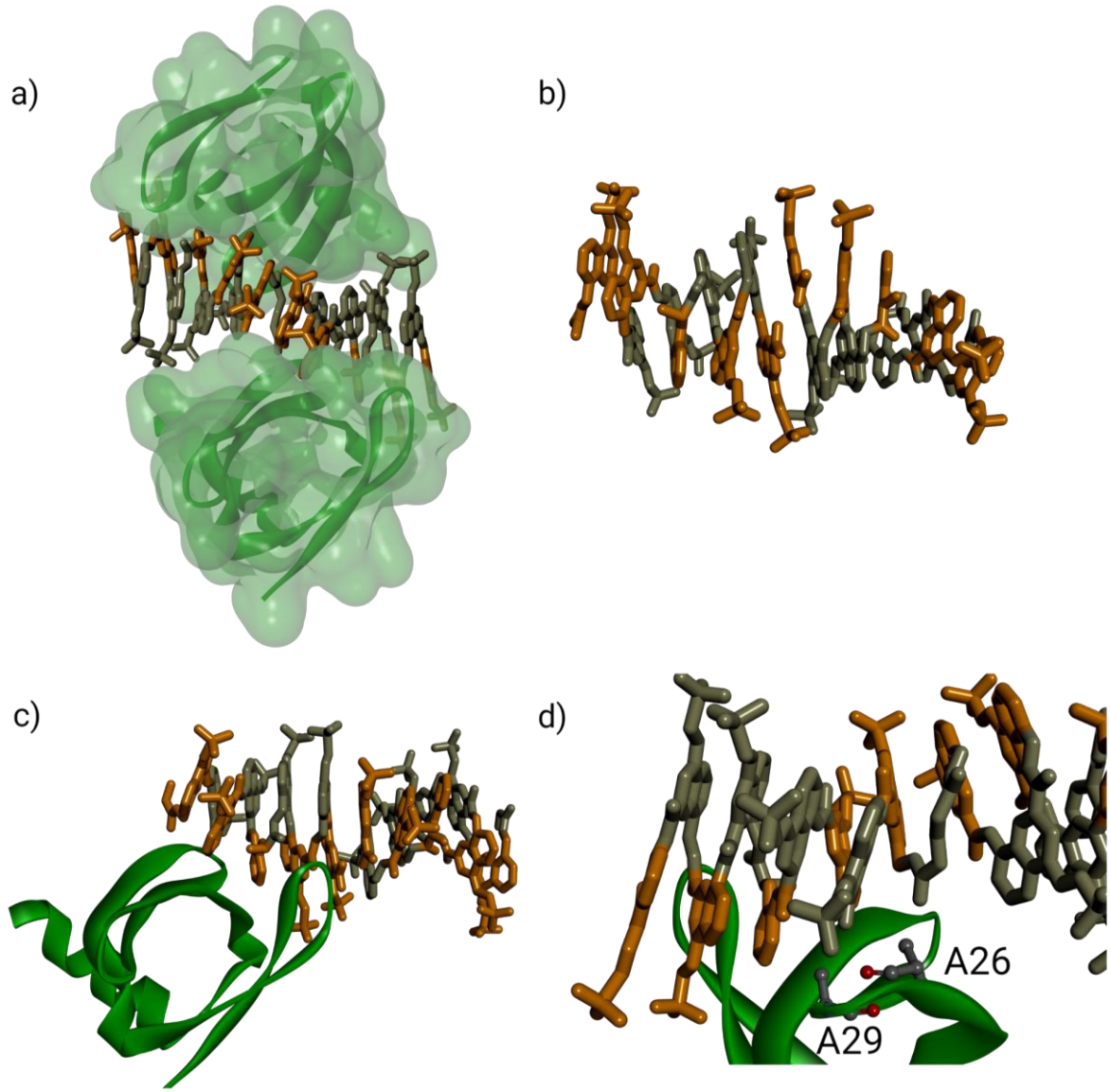


Figure S17. Sac7d V26A/M29A-**3** crystal structure (PDB# 8Q2M). a) Biological unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **3** in the ASU b) C-terminal stacking of the two 9mer units in ASU of **3**. c) Asymmetric unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **3** in the ASU. d) Contacts between **3** and A26 and A29 residues (mutated from V26 and M29) of Sac7d.

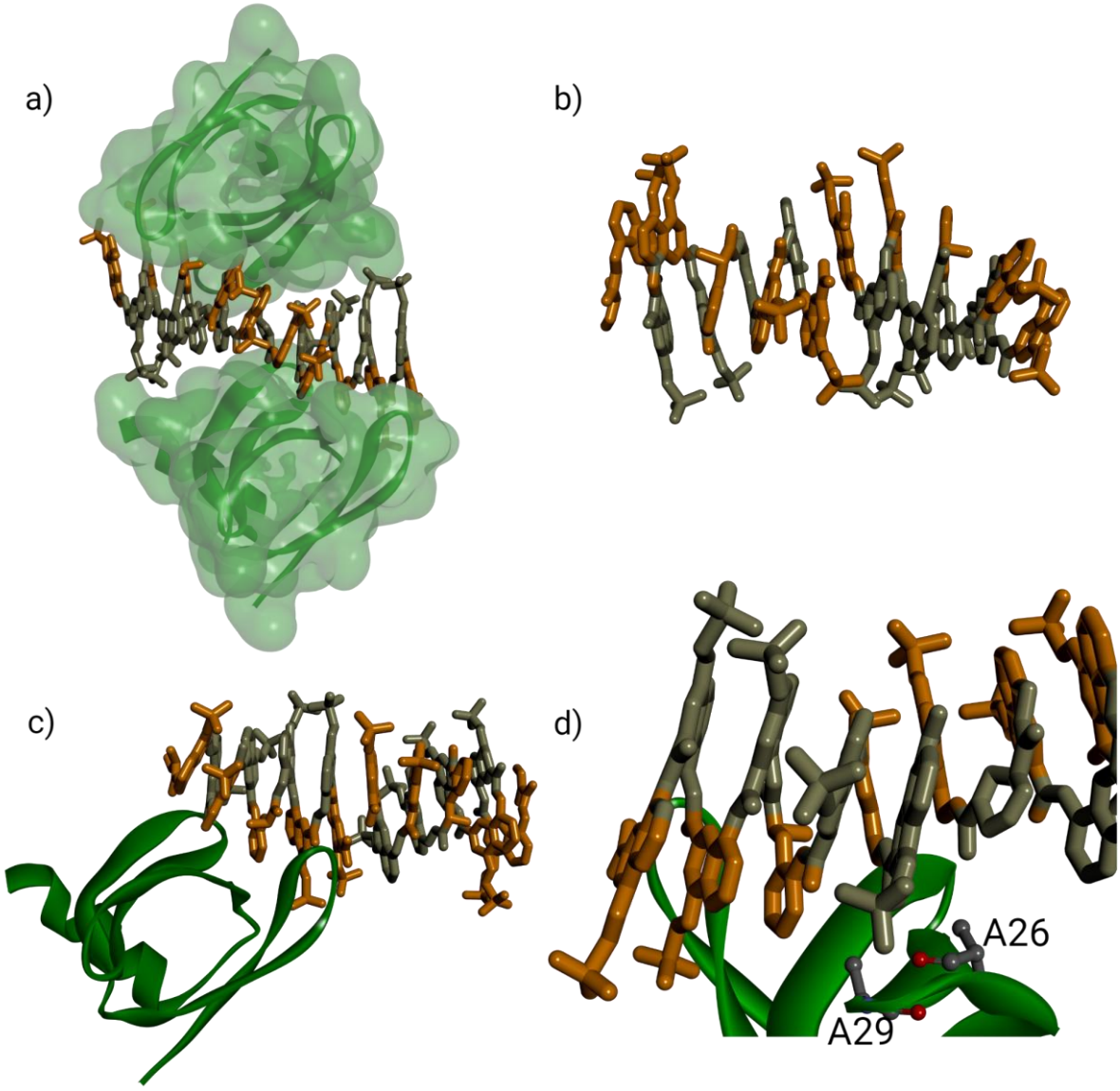


Figure S18. Sac7d V26A/M29A-**4** crystal structure (PDB# 8QPC). a) Biological unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **4** in the ASU b) C-terminal stacking of the two 9mer units in ASU of **4**. c) Asymmetric unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **4** in the ASU. d) Contacts between **4** and A26 and A29 residues (mutated from V26 and M29) of Sac7d.

3. Methods for SPR, CD, NMR, and X-ray crystallography

3.1 General

Chemicals and reagents were used as commercially supplied without any further purification unless otherwise stated. The One Shot™ BL21(DE3)pLysS Chemically Competent *E. coli* (C600003) procured from ThermoFisher Scientific. Bacterial culture media were sterilized by autoclaving. Bacterial cultures were grown using a MaxQ-6000 shaking incubator (ThermoFisher Scientific). HisPur™ Ni-NTA Superflow Agarose (25217) were purchased from ThermoFisher Scientific. Econo-Column® gravity flow columns (7372551) were obtained from BioRad. UV-Vis determination of protein concentration (280 nm), protein purity (260/280 nm), and OD600 were all measured on a NanoDrop™ OneC (ThermoFisher Scientific). Bacterial centrifugation was carried out on an Avanti JXN-26 Centrifuge (Beckman Coulter) using a JLA-8.1000 and Hitachi fixed angle rotor Himac P70AT. Cell sonication was carried out on a UP200St Ultrasonic Processor fitted with a S26d14 Sonotrode (Hielscher Ultrasonics) using 7 cycles of 2 min on (100% amplitude, 90% pulse) and 3 min rest. Sodium-dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out on a BioRad system, including a PowerPac™ HC High-Current Power Supply and a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell. Color Prestained Protein Standard (P7718S) was purchased from New England Biolabs. 2× 10-, 12-, or 15-well gels with a 15% resolving gel and 4% stacking gel. Gels were run at 75 V for 40 min then 120 V for 60 min and stained with Coomassie Brilliant Blue.

Size-exclusion chromatography (SEC) was performed in a cold cabinet (Unichromat 1500) maintained at 16°C on a Knauer fast protein liquid chromatography (FPLC) system coupled with a HiLoad® 16/600 Superdex® 75 pg column (Cytiva, 28989333). Protein elution was monitored by UV detection at 280 and 205 nm with a diode array detector and fractions were collected by a Foxy R1 Fraction Collector (Teledyne ISCO). Protein concentration and buffer exchange were performed using spin concentrators (Amicon, 3 kDa MWCO). Dialysis was performed using 3 kDa MWCO SlideA-Lyzer™ G2 dialysis cassettes (ThermoFisher Scientific).pH of the buffers was adjusted using HCl on a Mettler Toledo^{TM} SevenCompact pH meter. Ammonium-15N chloride with isotropic purity of ≥98 atom % ^{15}N (299251) was purchased from Sigma Aldrich. Tris-d_{11} (14978) was purchased from Deutrero.

3.2 Foldamer synthesis

Foldamers used in this study were prepared as previously described.[14]

3.3 SPR spectroscopy

Surface Plasmon Resonance (SPR). The measurements of the interaction between **1** and **2** with His_6-tagged Sac7d were performed on a Biacore™ T200. **1** and **2** used for SPR and CD studies were synthesized and characterized as previously described.[6b] His_6-tagged Sac7d was immobilized onto S Sensor Chip CM5 using HBS-EP, trademark Cytiva, USA (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% Tween-20, pH 7.4) as both running and analyte-binding buffer. The immobilization occurred through the activation of carboxymethyl groups on the dextran-coated chip by reaction with *N*-hydroxysuccinimide (7 min), followed by covalent bonding of the protein to the chip surface via amide linkages (7 min) and blockage of excess activated carboxylic groups with ethanolamine (7 min). Reference surfaces were prepared similarly, except that all the carboxylic groups on the chip were blocked, and no protein was added. The final concentration of bound protein was 2000 *RU*, expressed in response units (RU) and calculated by subtracting the reference RU from the protein RU. Binding experiments were performed at 25 °C and injected samples of **1** and **2** were prepared in the running buffer at different concentrations (0.62, 1.25, 2.25, 5, and 10 μM) and allowed to flow over the immobilized-protein surface for 2 min followed by a dissociation phase of 10 min. The chip surface was regenerated by removing foldamers with 50 mM NaOH for 30 s. The sensorgrams, which correspond to a variation of the SPR signal expressed in resonance units (RU) as a function of time (s), were plotted after subtraction from the signal on the reference flow cell. The kinetic data did not fit a 1:1 binding model (see main text for a discussion). However, the maximum response unit values (RU_{max}) plotted against the foldamer concentration ([F]) could be fitted to a 1:1 steady state model of the interaction ($RU_{eq} = RU_{max}$) using the following equation (1):

$$RU_{\mathrm{eq}} = RU_{\mathrm{max}} \cdot [F] \; / \, K_{\mathrm{d}} + [F] \qquad \text{(Eq. 1)}$$

K_d was found to be consistent across replicates. Note that the calculated values only give an order of magnitude of the interaction. They can be considered inherently poorly accurate since they derive from a 1:1 binding equation, while the kinetic data indicate that associations are more complex.

3.4 CD spectroscopy

Circular Dichroism (CD). CD data were recorded on a Jasco J-810 spectrometer with 1- or 2-mm quartz cuvettes. Scans were acquired at 20 °C, over the 300—500 nm range, with a 1 s response time and a 100 nm/min scanning speed. CD spectra were an average of three scans

and were baseline-corrected for signal contributions due to the buffer and protein. The exact concentration of Sac7d and **1** and **2** stock solutcons were determined using NanoDrop™ One^{C} (Thermo Fisher Scientific). Except for the DNA competition experiments, the raw CD data in millidegrees ($mdeg$) were converted to molar extinction ($\Delta\varepsilon$) per number of Q or M quinoline residues (*res*) by the following equation (2):

$$\text{(Eq. 1) } \Delta\varepsilon\ (L\,m^{-1}cm^{-1}res^{-1}) = mdeg\ /\ c\ \cdot\ l\ \cdot\ 32980 \cdot res$$

For binding studies, CD spectra of **1** and **2** in complex with 40 μM Sac7d at foldamer: protein ratio of 1:1, 1:2, and 1:4 (the latter only for **2**) were recorded in 50 mM NH_4HCO_3 buffer pH 8.5 after 24 h equilibration. To evaluate the binding affinity, seven samples were prepared in 50 mM NH_4HCO_3 buffer pH 8.5 containing 40 μM **1** and Sac7d at different concentrations (0, 20, 40, 60, 80, 120, 160 μM) and equilibrated for 24 h. For K_d calculation, maximum □□ values at 360 nm were plotted against protein concentration and fitted to the quadratic binding equation. K_d was found to be consistent across replicates and of the same order of magnitude as that determined through SPR measurements. DNA-binding competition experiments were performed in 50 mM NH_4HCO_3 buffer pH 8.5 using DNA decamer CCTATATAGG previously crystallized with Sac7d.[7e] To the equilibrated sample containing **1** and Sac7d (1:1, 40 μM), 10 bp DNA was added (1, 2, 4, 8, 16 eq.). Increasing amounts of DNA were added in the same sample containing the pre-formed complex of **1** and Sac7d waiting for 24 h equilibration between additions. The quantitative analysis of the competing experiment was done as explained below. We use the following abbreviations: S for Sac7d, D for the DNA duplex, F for the foldamer, M for the foldamer *M* helix and P for the foldamer *P* helix, SM for the Sac7d-*M* helix complex, and SD for the Sac7d-DNA complex.

We then define association constants:

K_1 = [SD]/[S][D] (association of the DNA to Sac7d)

K_2 = [SM]/[S][M] (association of the *M* helix of the foldamer to Sac7d, association to the *P* helix is neglected)

Unbound foldamer helices in solution have equal concentrations: [P] = [M]

The total concentrations of F, D, and S are:

$[F]_{tot}$ = [SM] + [M] + [P] = [SM] + 2[M]

$[D]_{tot}$ = [SD] + [D]

$[S]_{tot} = [SM] + [SD]$ (we assume saturation conditions and neglect the free Sac7d [S])

The CD signal at 360 nm is proportional to [SM]

Using the equations above, we can write:

$$\frac{K_2}{K_1} = \frac{2[\mathrm{SM}]}{[\mathrm{F}]\mathrm{tot} - [\mathrm{SM}]} \cdot \frac{[\mathrm{D}]\mathrm{tot} - [\mathrm{S}]\mathrm{tot} + [\mathrm{SM}]}{[\mathrm{S}]\mathrm{tot} - [\mathrm{SM}]}$$

Which may generally be solved as a quadratic equation. To simplify, we consider here the case where $[F]_{tot} = [S]_{tot}$ (foldamer and Sac7d concentrations are equal) and the situation where $[SM] = ½ [S]_{tot}$ (CD intensity is half its maximal intensity). The equation above can then be simplified to:

$$\frac{K_2}{K_1} = \frac{[\mathrm{S}]\mathrm{tot}}{\frac{1}{2}[\mathrm{S}]\mathrm{tot}} \cdot \frac{[\mathrm{D}]\mathrm{tot} - \frac{1}{2}[\mathrm{S}]\mathrm{tot}}{\frac{1}{2}[\mathrm{S}]\mathrm{tot}} = 4\frac{[\mathrm{D}]\mathrm{tot}}{[\mathrm{S}]\mathrm{tot}} - 2$$

Where $\frac{[\mathrm{D}]\mathrm{tot}}{[\mathrm{S}]\mathrm{tot}}$ is the number of equivalents of DNA with respect to the foldamer (or Sac7d) needed to reach half of the maximum CD intensity.

3.5 NMR spectroscopy

Compounds **3** and **4** used for NMR were synthesized and characterized as previously described.[14] NMR samples were prepared in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, and 10% D_2O. ^1H-^{15}N HSQC spectra were recorded at 298 K on a triple resonance Bruker Avance 800 MHz spectrometer equipped with a 5 mm cryoprobe for the detection of ^1H, ^{13}C, and ^{15}N.

Protein backbone assignment. The concentration of ^{15}N-labeled Sac7d ([^{15}N]-Sac7d) used was 500 µM and spectra were measured in a 5 mm NMR tube (Wilmad®). First, a ^1H-^{15}N HSQC spectrum was recorded using a standard pulse sequence including a watergate sequence with water flip-back pulses from the Bruker pulse sequence library (hsqcfpf3gpphwg). The sweep widths were 12800 (^1H) x 3333 (^{15}N) and 2K data points were collected for 512 increments of 4 scans per fid. Protein backbone assignment of [^{15}N]-Sac7d residues was performed based on already reported data[19] and with the help of additional (data not shown) 3D NMR spectra (two ^{15}N-TOCSY-HSQCs with different TOCSY mixing times, and one ^{15}N-NOESY-HSQC) on ^{15}N-labeled Sac7d and 2D NOESY on unlabelled Sac7d. ^{15}N-TOCSY-HSQC (mlevhsqcetf3gp3d) was collected as a series of 40 complex (t_2) data sets composed of 128 complex (t_3) values and 2K data points with 8 scans per fid. The mixing times were set to either 60 or 100 ms ^{15}N-

NOESY-HSQC (noesyhsqcfpf3gpsi3d) was collected as a series of 40 complex (t_2) data sets composed of 128 complex (t_3) values and 2K data points with 16 scans per fid. The noe mixing time was set to 120 ms 2D NOESY (noesyesgpph) on unlabeled Sac7d was measured with a sweep width of 12800 and 1K data points collected with 512 increments of 64 scans per fid. The noe mixing time was set to 120 ms All spectra were rudimentary processed with standard processing parameters from the Topspin (Bruker) processing library additionally applying zero filling to yield symmetrical either 3D or 2D matrices and then converted to UCSF format and further subjected to procession and evaluation with the open source software NMRFAM-SPARKY (version 3.190). This software allowed us to assign the Sac7d amide backbone through synchronization of 2D and 3D NMR spectra described above.

Chemical shift perturbation. Titrations were performed on independent samples of 100 µM ^{15}N-labeled Sac7d which were prepared to a final volume of 210 µL and titrated with 1—8 µL aliquots of 5 mM **3** or **4**. Each sample comprised 210 µL by supplementing the missing volume (after foldamer addition) with purified water to obtain a uniform concentration of 100 µM [^{15}N]-Sac7d in each NMR tube (see the following Table S1 for details). The prepared samples containing different concentrations of foldamer were kept at r.t. for 24 h to ensure equilibrium and were then transferred to 3 mm NMR tubes (Wilmad®). ^{1}H-^{15}N HSQC spectra for each foldamer concentration were then recorded with the same parameter set as for the backbone assignment mentioned above. All spectra were first processed using standard processing in Topspin and then converted to UCSF format. Chemical shift perturbations (CSPs, $\Delta\delta_H$ and $\Delta\delta_N$) were determined using CSP analysis module in CcpNmr Analysis Assign version 3.0.4 for Windows (http://www.ccpn.ac.uk/v3-software/downloads).[20]The obtained $\Delta\delta_H$ and $\Delta\delta_N$ values for the protein residues were then subjected to the equation $((\Delta\delta_H)/0.14)^2+(\Delta\delta_N)^2)^{0.5}$ in order to create a residue related CSP bar graph in Microsoft Excel®.

Table S1. Pipetting scheme of [^{15}N]-Sac7d-**3** and [^{15}N]-Sac7d-**4** for NMR spectroscopy.

	1 (No foldamer)	**2 25 µM foldamer**	**3 50 µM foldamer**	**4 100 µM foldamer**	**5 200 µM foldamer**
Sac7d (100 µM)	200 µL	200 µL	200 µL	200 µL	200 µL
Foldamer (3 or 4)	0 µL	1 µL	2 µL	4 µL	8 µL
Water	10 µL	9 µL	8 µL	6 µL	2 µL
Final volume	210 µL	210 µL	210 µL	210 µL	210 µL

3.6 Protein production

Expression and purification of Sac7d and Sac7d V26A/M29A proteins. The pET3b plasmid harboring Sac7d and Sac7d V26A/M29A gene was kindly received from Prof. Chin-Yu Chen.[7b] The pET3b-Sac7d plasmid was transformed into Escherichia coli BL21DE3 plys S cells. The cells were grown in Luria Broth to an OD at 600 nm of 0.8 and induced with 0.4 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) for 1 h induction at 37 °C. Afterward, cells were spun at 4000 g for 15 min. The supernatant was resuspended and sonicated in 20 mM Tris-HCl pH 7.5 supplemented with protease inhibitor cocktail (Thermo Scientific™ Halt™ Protease Inhibitor Cocktail) and 2 mM EDTA. This was then heated in a water bath at 68 °C for 30 min and centrifuged at 105000 g for 1 h. The supernatant was then loaded on SP cation exchange column (KNAUER Wissenschaftliche Geräte GmbH) and eluted with 20 mM Tris-HCl pH 7.5, 1000 mM NaCl. The purity was monitored by SDS-PAGE and LC-MS-ESI. In the end, pure Sac7d fractions were dialyzed against 20 mM Tris-HCl pH 7.5 and concentrated to 2.6 mM for crystallization. Sac7d V26A/M29A protein was expressed and purified using the same protocol described above for Sac7d.

[^{15}N]-Sac7d was expressed in minimal media supplemented with ^{15}N-labelled Ammonium Chloride (Sigma-Aldrich). The expression and purification procedure were similar to that described above for Sac7d.

3.7 Crystallization

Crystallization of Sac7d and Sac7d V26A/M29 with 3 and 4. Sac7d and Sac7d V26A/M29A protein were concentrated between 2.4 to 2.8 mM in 20 mM Tris-HCl pH 7.5 buffer. **3** and **4** were dissolved in pure water up to 5 mM concentration. Protein-foldamer complex was made as shown in Table S2 and incubated at 4 °C for 1 h prior to crystallization in the hanging drop vapor diffusion method.

Table S2 Crystallization parameters for Sac7d and Sac7d V26A/M29A protein with **3** and **4**.

Complex	Concentration	Crystallization reservoir solution	Crystallogenesis duration	Cryo-protectant solution
Sac7d-**3**	1.3 mM+1.3 mM	10%PEG 400, 0.1 M MES, pH 6.0	3 d at 20 °C	25% (w/v) glucose
Sac7d V26A/M29-**3**	1.3 mM+1.3 mM	10% PEG 400, 0.1 M MES, pH 6.0	2 d at 20 °C	25% (v/v) ethylene glycol
Sac7d V26A/M29A-**4**	1.3 mM+1.3 mM	10% PEG 400, 0.1 M MES, pH 6.0	2 d at 20 °C	30% (w/v) glucose

3.8 Data collection and structure refinement

Sac7d-3. X-ray diffraction data were collected from cryo-protected crystals at 100 K on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble) on a Dectris Eiger2 X 16M detector.[21] The collected data set was processed using XDS[22], and the structure was solved by MoRDa[23] using Affitin h4[24] (PDB ID 4CJ2) as a search model for protein and Phaser MR[25] using the energy minimized molecular model of **3**[14] (prepared on Maestro V11.5, Schrödinger) as the search model for the foldamer. The structure was refined in iterative rounds of manual model building and refinement in Coot[26] and Phenix Refine.[27] ProDRG[28] was used to generate restrain file for **3**. Foldamer was further adjusted in visible electron density at the final refinement stage. The structure was validated with MolProbity[29] and deposited in PDB under accession code 8CMN. Data collection and structure refinement statistics for Sac7d in complex with **3** are given in table S3.

The calculation of the binding surface area of **3** onto Sac7d surface was performed in Pymol using "get area" command.

Sac7d V26A/M29A-3. X-ray diffraction data were collected from cryo-protected crystals at 100 K on beamline P13 macromolecular X-ray crystallography at the Deutsches Elektronen Synchrotron (DESY) at European Molecular Biology Laboratory (EMBL, Hamburg) on an EIGER 16M detector.[21] The structure was solved by Phaser MR[25] using PDB 8CMN as a search model for protein and molecular model of **3** from PDB #8CMN. The structure was refined in iterative rounds of manual model building and refinement in Coot[26] and Phenix Refine.[27] ProDRG[28] was used to generate restrain file for **3**. Foldamer was further adjusted in visible electron density at the final refinement stage. The structure was validated with MolProbity[29] and deposited in PDB under accession code 8Q2M. Data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **3** are given in table S4.

Sac7d V26A/M29A-4. X-ray diffraction data were collected from cryo-protected crystals at 100 K on beamline ID30-B at the European Synchrotron Radiation Facility (ESRF, Grenoble) on a Dectris Eiger2 X 16M detector.[30] The structure was solved by Phaser MR[25] using PDB 8Q2M as a search model for protein. An energy minimized molecular model of **4**[14] (prepared on Maestro V11.5, Schrödinger) was used for molecular replacement of the foldamer. The structure was further refined in iterative rounds of manual model building and refinement in Coot[26] and Phenix Refine.[27] ProDRG[28] was used to generate restrain file for **4**. Foldamer was further adjusted in visible electron density at the final refinement stage. The structure was

validated with MolProbity[29] and deposited in PDB under accession code 8QPC. Data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **4** are given in table S5.

Table S3. Crystallography data collection and structure refinement statistics for Sac7d in complex with **3 (PDB #8CMN)**.

Parameters	**Value**
Wavelength	0.8856
Resolution range	42.34 – 2.65 (2.745-2.65)
Space group	P 64 2 2
Unit cell	71.41 71.41 116.17 90 90 120
Total reflections	37941 (3816)
Unique reflections	5508 (531)
Multiplicity	6.9 (7.2)
Completeness (%)	99.62 (99.62)
Mean I/sigma (I)	18.81 (2.13)
Wilson B-factor	74.99
R-merge	0.05791 (0.7586)
R-meas	0.06293 (0.818)
R-pim	0.02345 (0.2956)
CC1/2	0.998 (0.81)
CC*	1 (0.946)
Reflections used in refinement	5502 (531)
Reflections used for R-free	547 (52)
R-work	0.2613 (0.4296)
R-free	0.2979 (0.5179)
CC (work)	0.892 (0.718)
CC (free)	0.919 (0.313)
Number of non-hydrogen atoms	848
macromolecules	495
ligands	346
solvent	7
Protein residues	63
RMS (bonds)	0.008
RMS (angles)	2.69
Ramachandran favored (%)	98.36
Ramachandran allowed (%)	1.64

Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clash score	17.86
Average B-factor	66.50
macromolecules	67.78
ligands	64.68
solvent	65.40
Number of TLS groups	4

Table S4. Crystallography data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **3 (PDB #8Q2M)**.

Parameters	**Value**
Wavelength	0.9763
Resolution range	30.71 - 3.212 (3.327 - 3.212)
Space group	P 64 2 2
Unit cell	70.92 70.92 119.29 90 90 120
Total reflections	119569 (12369)
Unique reflections	3226 (310)
Multiplicity	37.1 (39.9)
Completeness (%)	99.54 (100.00)
Mean I/sigma (I)	23.56 (3.51)
Wilson B-factor	115.54
R-merge	0.08561 (1.222)
R-meas	0.08692 (1.238)
R-pim	0.01467 (0.1939)
CC1/2	1 (0.952)
CC*	1 (0.988)
Reflections used in refinement	3219 (310)
Reflections used for R-free	322 (31)
R-work	0.3238 (0.3609)
R-free	0.3267 (0.3914)
CC (work)	0.885 (0.823)
CC (free)	0.963 (0.921)
Number of non-hydrogen atoms	836
macromolecules	490
ligands	346
solvent	0
Protein residues	63

Nucleic acid bases	
RMS (bonds)	0.035
RMS (angles)	4.96
Ramachandran favored (%)	95.08
Ramachandran allowed (%)	4.92
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clash score	34.56
Average B-factor	132.72
macromolecules	145.47
ligands	114.67
solvent	
Number of TLS groups	3

Table S5. Crystallography data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **4 (PDB # 8QPC)**.

Parameters	**Value**
Wavelength	0.871
Resolution range	61.41 - 3.241 (3.356 - 3.241)
Space group	P 64 2 2
Unit cell	70.907 70.907 121.997 90 90 120
Total reflections	51589 (5784)
Unique reflections	2973 (244)
Multiplicity	17.4 (19.3)
Completeness (%)	89.51 (81.27)
Mean I/sigma (I)	9.81 (1.96)
Wilson B-factor	91.41
R-merge	0.1249 (1.559)
R-meas	0.1296 (1.601)
R-pim	0.03345 (0.3614)
CC1/2	0.999 (0.938)
CC*	1 (0.984)
Reflections used in refinement	2875 (243)
Reflections used for R-free	139 (9)
R-work	0.3356 (0.3776)
R-free	0.3470 (0.4302)
CC (work)	0.862 (0.927)
CC (free)	0.989 (0.773)

Number of non-hydrogen atoms	812
macromolecules	462
ligands	350
solvent	0
Protein residues	63
RMS (bonds)	0.028
RMS (angles)	3.81
Ramachandran favored (%)	95.08
Ramachandran allowed (%)	4.92
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clash score	19.31
Average B-factor	123.77
macromolecules	143.17
ligands	98.16
Number of TLS groups	3

3.9 High-resolution AFM imaging

General. AFM imaging in aqueous solution was performend on a NanoWizard® 3 ultra AFM (JPK Instruments AG). Measurements were performed in AC mode on a scan area of 500 x 500 nm with a micro cantilever (ν_{res} = 110 kHz, k_{spring} = 9 N/m, Olympus Corp.). For sample immobilization, a freshly cleaved mica surface (Quality V1, Plano GmbH) was incubated with a 10 mM solution of poly-L-ornithine (0.01% 30000 – 70000 g/mol, Sigma Aldrich) for 3 minutes. The mica was washed three times with ultra-pure water to get rid of unbound poly-L-ornithine and was blow-dried with air. Leveling, background correction and extraction of height histograms of obtained AFM images were realized with the software Gwyddion (version 2.60).[31]

Sample incubation and preparation. The relaxed plasmid DNA pBR322 (Inspiralis Limited, UK) was diluted in 1xPBS to *ca.* 23 nM, corresponding a concentration of DNA base pairs of *ca.* 100 µM, and shaken in an incubation shaker at 75°C for 15 min. For incubation with sac7D, 100 µM sac7D was added to the plasmid DNA and the mixture was shaken in an incubation shaker at 75°C for 15 min. For incubation with sac7D and foldamer **1**, 100 µM sac7D and 50 µM **1** were added to the plasmid DNA and the mixture was shaken in an incubation shaker at 75°C for 15 min. After incubation and shaking at 75°C, each sample was diluted 1:100 in 1xPBS resulting in a DNA concentration of *ca.* 0.23 nM and immediately incubated on freshly prepared

mica for 3 min. Excessive sample solution was taken off and the mica surface was washed three times with 1xPBS. AFM imaging was performed in 1xPBS.

Note: The temperature of incubation, 75°C, was chosen since it has been reported to be the temperature of maximum affinity of sac7D to DNA.[7k] Attempting to immobilize sample solution by diluting the 75°C incubation solutions with 75°C preheated 1xPBS did not result in efficient sample binding. Dilution with room temperature 1xPBS and immediate immobilization to prevent dissociation of sac7D from DNA led to efficient sample binding. It is not known how much of the sac7D dissociates from the DNA in the seconds between incubation in RT 1xPBS and immobilization on mica. DNA may possibly be even more compacted than observed.

3.10 Isothermal titration calorimetry

ITC titration experiments were performed at 25 °C on a *Malvern* MicroCal™ PEAQ-ITC instrument. The reference cell was filled with water, the stirring speed set to 750 rpm and the initial delay to 60s. Foldamer-to-Sac7d titrations were carried out using a 19-injection protocol, starting with a 0.4 µL injection followed by 18 injections of 2 µL each. The sample cell contained 100 µM Sac7d, while the syringe was filled with 250 µM foldamer solution. Both components were prepared in 25 mM KCl and 10 mM KH_2PO_4 (pH=7). The cell and syringe were equilibrated with the buffer before loading. To account for dilution effects, control experiments (foldamer injections into buffer) were performed, and the resulting data were subtracted from those of the foldamer titration into Sac7d. The ITC data were presented as differential power (DP) versus time and analyzed using Malvern MicroCal™ PEAQ-ITC analysis software.

3.11 UV melting curve

UV melting experiments were recorded on JASCO V-650 spectrophotometer with a PAC-743 6-position Peltier. DNA samples in 10 mM KH_2PO_4 (pH=7) were annealed and incubated with Sac7d and foldamer at room temperature. Measurements were taken in 0.1 mm quartz cuvettes (26 µL, Hellma) with silicone oil to prevent evaporation. Absorbance at 260 nm was recorded over a temperature range of 30 to 85 °C, with a heating rate of 0.1 °C/min. Each experiment was repeated at least three times, and the resulting melting curves were analyzed using JASCO data analysis software. The data were fitted to a two-state transition model with defined upper and lower baselines to determine the melting temperature (Tm).

4. References (in continuation of the main text)

[18] I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, D. Herschlag, *eLife* **2020**, *9*, e57264.

[19] J. L. Bedell, B. S. McCrary, S. P. Edmondson, J. W. Shriver, *Protein Sci.* **2000**, *9*, 1878-1888.

[20] L. Mureddu, G. W. Vuister, *FEBS J.* **2019**, *286*, 2035-2042.

[21] D. Nurizzo, T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J. C. Biasci, S. McSweeney, E. Mitchell, *J. Synchrotron Radiat.* **2006**, *13*, 227-238.

[22] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125-132.

[23] A. Vagin, A. Lebedev, *Acta Cryst.* **2015**, *A71*, s19.

[24] A. Correa, S. Pacheco, A. E. Mechaly, G. Obal, G. Béhar, B. Mouratou, P. Oppezzo, P. M. Alzari, F. Pecorari, *PloS One* **2014**, *9*, e97438.

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

[26] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr. D* **2010**, *66*, 486-501.

[27] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr. D Biol. Crystallogr.* **2019**, *75*, 861-877.

[28] A. W. Schüttelkopf, D. M. F. van Aalten, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1355-1363.

[29] V. B. Chen, J. R. Wedell, R. K. Wenger, E. L. Ulrich, J. L. Markley, *J. Biomol. NMR* **2015**, *63*, 77-83.

[30] A. A. McCarthy, R. Barrett, A. Beteva, H. Caserotto, F. Dobias, F. Felisaz, T. Giraud, M. Guijarro, R. Janocha, A. Khadrouche, M. Lentini, G. A. Leonard, M. Lopez Marrero, S. Malbet-Monaco, S. McSweeney, D. Nurizzo, G. Papp, C. Rossi, J. Sinoir, C. Sorez, J. Surr, O. Svensson, U. Zander, F. Cipriani, P. Theveneau, C. Mueller-Dieckmann, *J. Synchrotron Radiat.* **2018**, *25*, 1249-1260.

[31] D. Nečas, P. Klapetek, *Open Physics* **2012**, *10*, 181-188.

7. Tailoring the major groove of DNA mimic foldamers

Following our achievement of solving the first crystal structure of Sac7d-DNA mimic foldamer complex, we have moved toward to a new challenge: designing highly specific foldamers for their protein targets. Our previous work on the Sac7d-foldamer complex, while delivering detailed structural information, did not include sequence-specific recognition features. Inspired by natural protein-nucleic acid interactions (PNIs), where the protein's α-helix often fits into the major groove of DNA, we hypothesized that we could modify the groove features of our foldamers to accommodate an α-helix. (Figure 19).

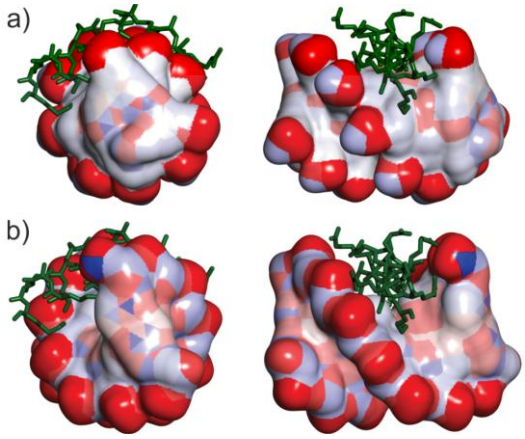


Figure 19. a) Two views of the α-helical Lys231-Lys246 segment of leucine zipper protein GCN4 docked in the major groove of a DNA mimic foldamer. b) Two views of the same peptide bound to the major groove of a B-DNA duplex extracted from the crystal structure of the a DNA-GCN4 complex (PDB: 2DGC). The peptide is shown as green tubes. DNA and foldamer are shown as solvent accessible surfaces color-coded with the parent atom charge.

To achieve this, we set out to develop a new chemical tool box to serve specific design objectives. We designed and synthesized four new monomers that are structurally different from the our first-generation building blocks. We then validated how these new monomers affect the folding behaviour and flexibility of the resulting oligoamide helices. Through CD and NMR spectroscopy, we confirmed that incorporating these monomers changes the foldamers' properties, particularly their folding stability.

For a long period of time, the limited number of foldamer building blocks restricted the sequence features we could create for our foldamers. My work in this chapter will be crucial for developing the next generation of highly selective DNA mimic foldamers, significantly increase the possibility of achieving specific α-helix recognition.

Contributions: The project was planned by IH. JW and VC synthesized the foldamers. JW developed all the new monomers. JW performed the UV and CD measurements. JW and PKM performed crystal growth. PKM performed the crystallographic analysis. IH supervised the project.

7.1. Publication (to be submitted)

Tailoring the major groove of DNA mimic foldamers

Authors: **J. Wu,** V. Corvaglia, P. K. Mandal, and I. Huc*

ARTICLE

Tailoring the major groove of DNA mimic foldamers

Jiaojiao Wu,[a] Valentina Corvaglia,[‡a] Pradeep K. Mandal,[a,b] and Ivan Huc[a*]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Single stranded helically folded aromatic oligoamides bearing anionic phosphonate side chains have been shown to bind to some DNA-binding proteins better than DNA itself. However, these DNA mimic foldamers have until now mainly consisted of a single repeat motif, like a poly(dA:dT) DNA duplex, and contained limited sequence information. Here, we introduce new monomers designed to display different chemical functionalities in the major groove of the DNA mimics. Four new Fmoc-protected amino acid monomers have been synthesized and incorporated into oligomers. Sixteen foldamer sequences were prepared on solid phase. Their conformation in solution and in the solid state and their conformational dynamics were investigated using NMR, circular dichroism, molecular modeling, and X-ray crystallography. The results show that three of four new monomers behaved as designed, and that their introduction enhances the conformational dynamics of the DNA mimic foldamers. In a fourth case, conformational behavior proved to be more complex than expected. These results showcase design strategies to manipulate large molecular biomimetics where not only side chains but also main chain components are varied. The new monomers pave the way to complex DNA mimic foldamer sequences targeting proteins that recognize sequence-selective DNA-binding proteins such as transcription factors or restriction enzymes.

Introduction

DNA-binding proteins control numerous essential biological processes, making protein-DNA interactions relevant targets for both fundamental research and therapeutic intervention.[1] In nature, protein-nucleic acid interactions are regulated by a wide array of mechanisms, including small molecule-induced allosteric effects in nuclear receptors,[2, 3] chromatin remodeling,[4] or post-translational modifications.[5] Nature has also developed so-called DNA mimic proteins, proteins that reproduce DNA's shape and surface features, and that bind to, and highjack DNA binding proteins.[6, 7] Artificial tools to interfere with DNA-protein interactions include small molecules able to block protein-DNA cross-links such as topoisomerase poisons,[8] molecules that recognize DNA and may prevent protein binding such as pyrrole-imidazole oligoamides,[9, 10] or DNA decoys, which are modified DNA strands that divert proteins from their natural DNA target.[11, 12]

Beyond these advances, DNA mimic foldamers offer a distinct and promising approach. Like DNA mimic proteins, they reproduce some surface features of DNA and bind to proteins that normally recognize DNA. Some peptide-based DNA mimics [13] and sulfated oligosaccharides such as heparin bind to DNA-binding proteins.[14, 15] Yet the main family of DNA mimic foldamers consist of single-stranded aromatic oligoamides bearing negatively charged side chains at positions that match the positions of phosphates in duplex B-DNA. In contrast with other DNA mimics that possess base-pairing abilities and target nucleic acids, *e.g.* peptide nucleic acids (PNA),[16] locked nucleic acids (LNA)[17] and other xenonucleic acids (XNAs),[18, 19] DNA mimic foldamers are specifically designed to directly compete with DNA for binding to proteins.[20-24]

The DNA mimic foldamer parent series consists of alternating 8-aminomethyl-2-quinolinecarboxylic acid monomer M and 8-amino-2-quinolinecarboxylic acid monomer Q^4, both bearing a negatively charged side chain (Fig. 1a).[20] Like many other aromatic foldamers,[25] these sequences adopt stable helical conformations stabilized by intramolecular hydrogen bonds, electrostatic repulsions and aromatic stacking. Since M and Q^4 lack stereogenic centers, $(MQ^4)_n$ helices exist as a racemic mixture of right-handed (*P*) and left-handed (*M*) conformers. Upon introducing a chiral B^R residue, handedness can be biased quantitatively to an *M* main chain helix that displays two *P* exo-helices of negatively charged side chains matching the negative charge distribution of B-DNA (Fig. 1b,c).[26, 27] These structural attributes, along with the construction of palindromic foldamer sequences,[27] enable DNA mimic foldamers to outcompete DNA in binding some DNA-binding proteins. For example, binding and inhibition has been demonstrated for topoisomerase 1 and HIV integrase,[20, 22] for bacterial chromosomal protein Sac7d,[21] and for the origin recognition complex.[28] $(MQ^4)_n$ helices have been shown to impact chromatin composition *in vitro* and *in vivo* and disrupt cell cycle progression.[28]

DNA mimic foldamer targets have until now been limited to proteins that recognize B-DNA through its overall shape and do not include proteins specific to a particular DNA sequence. Among shape selective-proteins, not all are efficiently bound by the DNA mimic foldamers. A certain degree of selectivity is thus observed.[20] Nevertheless, the large family of proteins that

[a] *Department Pharmazie, Ludwig-Maximilians-Universität München, Butenandtstr. 5-13, München 81377, Germany. E-mail: ivan.huc@cup.lmu.de.*
[b] *Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria*
‡ Present affiliation: Institute for Stem-Cell Biology, Regenerative Medicine and Innovative Therapies, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Italy) & Center for Nanomedicine and Tissue Engineering (CNTE), ASST Grande Ospedale Metropolitano Niguarda, Milan, Rotondo, Italy.
Electronic Supplementary Information (ESI) available: supplementary Figures, detailed experimental protocols, crystallographic studies, and characterisation of new compounds. CCDC 2514117, 2514118, 2286782, and 2478322. For ESI and crystallographic data in CIF or other electronic format see DOI:10.1039/x0xx00000x

selectively bind to specific DNA sequences, *e.g.* transcription factors and restriction enzymes, may not be targeted with simple $(MQ^4)_n$ oligomers. Indeed, sequence-selective DNA-binding proteins typically interact with the grooves of B-DNA where parts of the nucleobase are exposed.[1,29] In contrast, an $(MQ^4)_n$ helix has a constant repeat motif and may be compared to a poly(dA:dT) DNA duplex deprived of sequence information.

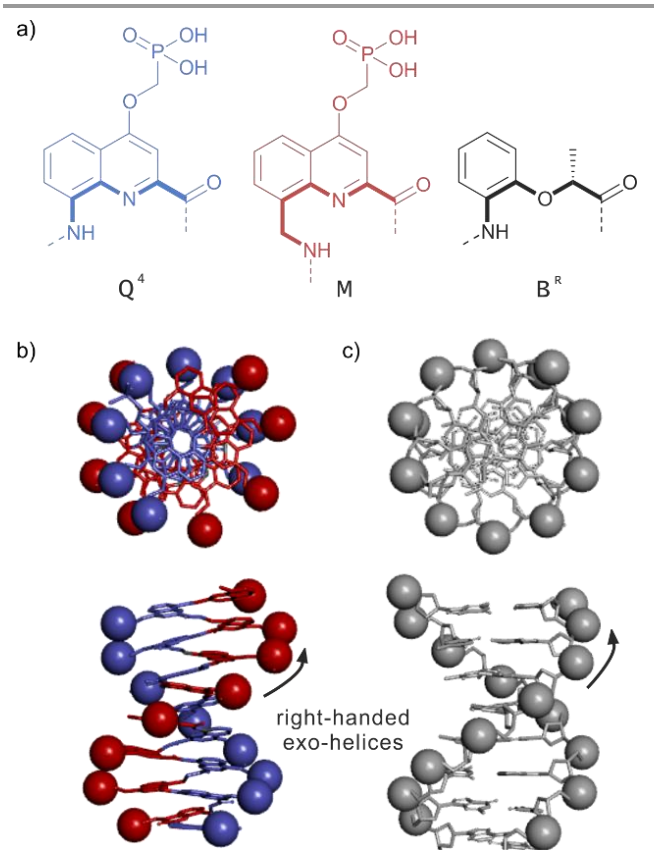


Fig. 1 (a) Structures of Q^4 , M and B^R amino-acid monomers. Top views and side views of molecular models of $(MQ^4)_8$ (b) and of an eight-base-pair B-DNA duplex (c). Models are shown at the same scale as stick representations, except phosphorus atoms which are shown as spheres. Monomers are color-coded in red and blue as in (a).

To address this challenge and advance the next generation of DNA mimic foldamers, we have developed chimeric molecules that combine foldamer and DNA segments and that showed promise in targeting protein-DNA interactions involving sequence-selective binding.[20, 30] Here, we introduce the first step of an alternate approach aiming at tailoring the major groove of the DNA mimics itself so that it contains sequence information through modifications of the foldamer main chain. We report the design and synthesis of novel monomers that are structural analogues of M and Q^4 units. Through solid-state and solution investigations, we validate that these monomers are compatible with the $(MQ^4)_n$ helical architecture while they modify groove shape, expose different functionalities in the major groove and allow for the modulation of the overall flexibility of the helix. The combination of the new units with M and Q^4 monomers results in aperiodicity in main chain composition, which is unusual in both biopolymers and synthetic foldamers. In the context of DNA mimic foldamers, it is hoped that it will enable the recognition of specific protein components such as the α-helices used by many transcription factors to read DNA base pair sequences in the major groove.

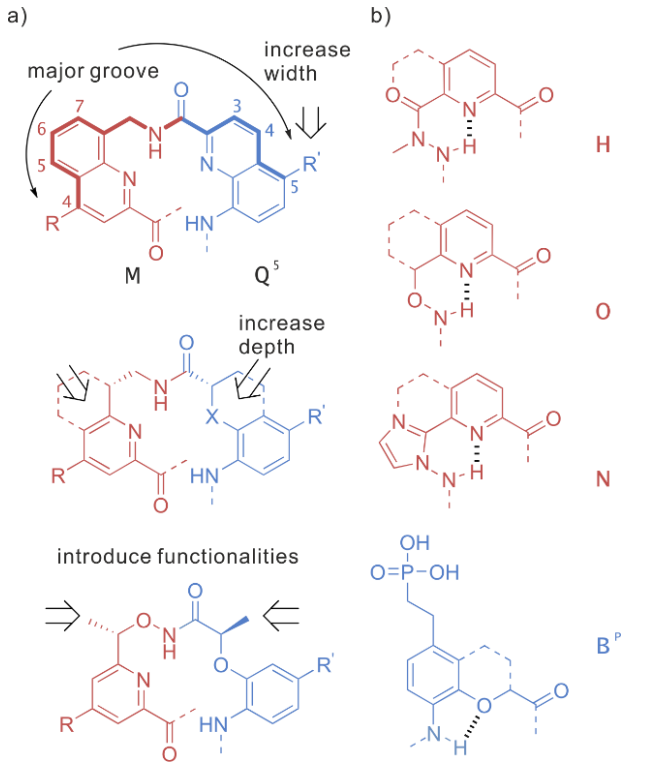


Fig. 2 (a) Q^5M dimer with the rim exposed in the major groove shown in bold lines. The side chain position in Q^5 differs from that of Q^4 (Fig. 1a). (b) M analogues H, O and N (in red), and Q analogue B^P (in blue). Hydrogen bonds are indicated as black hashed lines. Dashed lines indicate the parts of the aromatic rings of M and Q that have been removed.

Results and discussion

Monomer design and synthesis

The original $(MQ^4)_n$ helix has shallow grooves in which each MQ^4 dimer – the equivalent of a base pair in B-DNA – displays the same functionalities. This structural homogeneity makes it challenging for these foldamers to engage in sequence-selective binding with DNA-binding proteins. To overcome this limitation, we proposed to tailor the major groove of DNA mimic foldamers, as it is the primary site of interaction for a majority of DNA-binding proteins. In $(MQ)_n$ helices, the major groove can be defined as the area starting from the side chain of Q until the side chain of M (Fig. 2a). The functions exposed in the groove include part of the pyridine ring of Q (positions 3 and 4) as well as its carbonyl group and the main chain methylene of M as well as part of its benzene ring (positions 5, 6, and 7).

Our approach focused on three key objectives (Fig. 2a): 1) widening the major groove by repositioning the side chain of Q units from position 4, as Q^4 in Fig. 1a, to position 5, as Q^5 in Fig. 2a; 2) deepening the major groove by trimming the pyridine ring of Q or the benzene ring of M; and 3) introducing new functionalities to provide additional interaction sites. To this end, we designed and synthesized H, O and N, three pyridine

based ε-amino acid analogues of monomer M, and B^P, an aniline-based δ-amino acid analogue of monomer Q (Fig. 2b). The new features of these monomers include hydrogen-bond donors/acceptors and the methylene groups of O and B^P where other functionalities could be installed, including stereogenic centers.

Amino acid H was designed with a methyl-substituted hydrazide group: pyridyl-C(=O)N(Me)NH. The methyl group was intended to promote the formation of a six-membered hydrogen-bonded ring involving the adjacent NH proton and the pyridine endocyclic nitrogen atom (Fig. 2b) as with M. In the absence of methyl group, the *trans* conformation of the hydrazide C(=O)–NH bond would instead favor a five-membered hydrogen-bonded ring and the contribution to helix shape would differ from that of M. The H monomer lacks carbon atoms 6 and 7 of the benzene ring of M, which results in a deeper major groove. Instead of the carbon 5 of M, also in the major groove, H has a carbonyl oxygen atom, *i.e.* a hydrogen bond acceptor. Monomer H was synthesized in four steps starting from commercially available dimethyl pyridine-2,6-dicarboxylate (Fig. 3a) using described procedures for its mono-saponification into **1a** and subsequent activation into mono acid chloride **1b**.[31] Coupling with Fmoc-methylhydrazine afforded compound **1c**. Demethylation of **1c** with lithium iodide provided **1**, the Fmoc-protect form of H, in an overall 38% yield over four steps without requiring any chromatographic purification. This route can in principle be applied to variants of H bearing another substituent than methyl on the hydrazide or a side chain, *e.g.* a phosphonate, on the pyridine ring using chelidamic acid as a starting material, but this has not been tested yet.

Monomer O was designed by trimming M even more than for H (Fig. 2b). Two sp^3 centers, a methylene group and the oxygen atom of a hydroxylamine, are exposed in the major groove. The reduced number of sp^2 centers in the main chain may allow for easier bond rotation, that is, enhanced flexibility. The synthesis of O was carried in four steps from commercial methyl 6-(hydroxymethyl)picolinate (Fig. 3a). A hydroxylamine protected as phthalimide was introduced via a Mitsunobu reaction in 95% yield. Deprotection of the amine with hydrazine hydrate followed by methyl ester saponification yielded amino acid **2c** which was not purified and used directly in the final Fmoc installation step. Again, these relatively mild transformations are presumed to be compatible with a range of substituents on the pyridine ring.

Amino acid N can be viewed as an analogue of H in which the rotation about the C(O)–NMe bond is locked within an imidazole ring (Fig. 2b). N has the same number of rotatable bonds as M and no sp^3 center which we expect to contribute to rigidity. N was prepared from methyl 6-formylpicolinate in four steps,[32] starting by the construction of the imidazole ring using NH_4OAc and glyoxal, followed by an electrophilic N-amination, saponification of the methyl ester and Fmoc installation (Fig. 3c).

Finally, amino acid B^P carrying a phosphonic acid side chain was conceived as an analogue of Q[5], that is, with a wider major groove (Fig. 2b). In addition, the carbon atoms in position 3 and

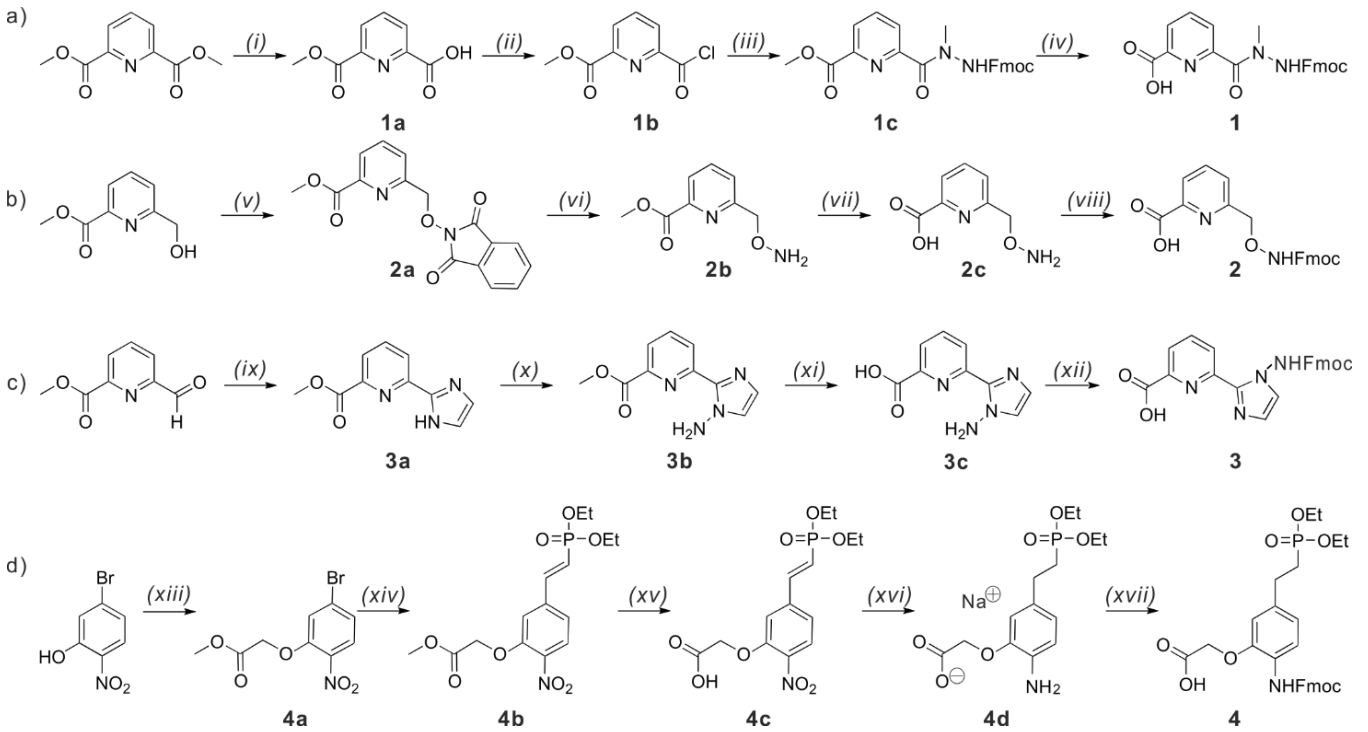


Fig. 3 (a) Synthesis of Fmoc-H-OH (**1**). (b) Synthesis of Fmoc-O-OH (**2**). (c) Synthesis of Fmoc-N-OH (**3**). (d) Synthesis of Fmoc-B^P-OH (**4**). Reagents and conditions: (i) KOH, MeOH, 0° C, 3h; (ii) Oxalyl chloride, DMF, r.t., 3h; (iii) Fmoc-methylhydrazine, triethylamine, dichloromethan, r.t., 2h; (iv) LiI, EtOAc, overnight; (v) *N*-Hydroxyphthalimide, diisopropyl-diazodicarboxylate, PPh_3, toluene, r.t., 3h; (vi) $N_2H_4.H_2O$, MeOH, r.t., 30 min; (vii) KOH, MeOH/H_2O, 0 °C, 1h; (viii) Fmoc-Osu, $NaHCO_3$, dioxane/H_2O, 0° C to r.t., overnight; (ix) NH_4OAc, glyoxal, MeOH, 50°C, 2h; (x) Lithium bis(trimethylsilyl)amide, O-diphenylphosphinyl hydroxylamine, THF/DMF, 0°C, 2h; (xi) LiOH, THF/H_2O, 0° C, 30 min; (xii) Fmoc-Osu, $NaHCO_3$, Dioxane/H_2O, 0° C to r.t., overnight; (xiii) Methyl bromoacetate, K_2CO_3, Acetone, 2h; (xiv) Diethyl vinylphosphonate, $Pd(PPh_3)_2Cl_2$, K_2CO_3, *o*-xylene, 125° C, 2h; (xv) LiOH, THF/H_2O, 0° C, 30 min; (xvi) H_2, Pd/C, Na_2CO_3, THF; (xvii) Fmoc-Cl, $NaHCO_3$, dioxane/H_2O, 0° C to r.t., overnight.

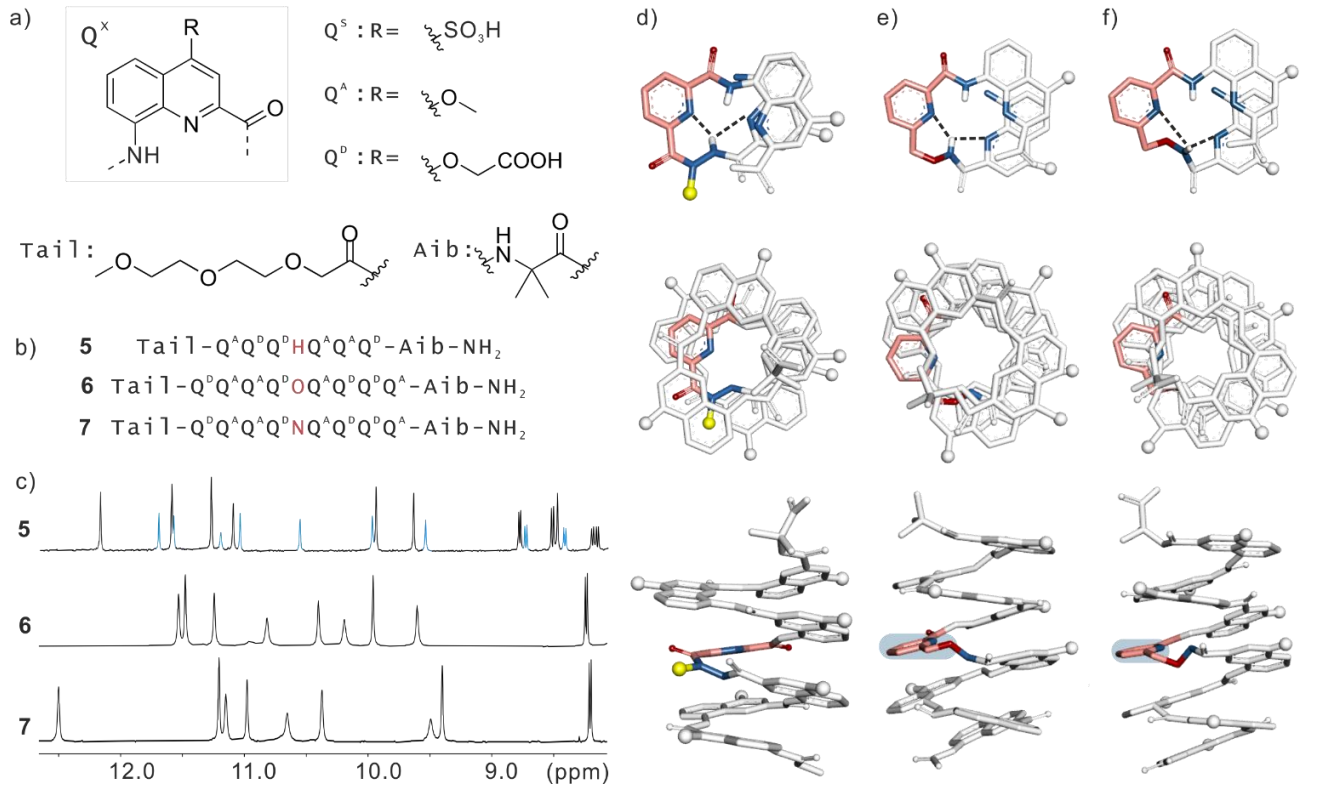


Fig. 4 (a) Structures of Q^X, and of N- and C-terminal groups Tail and Aib. (b) Aromatic foldamer sequences **5-7**. (c) Excerpts of the ^{1}H NMR spectra of sequences **5-7** (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C). A minor set of signals in the spectrum of **5** are highlighted in blue. (d) Crystal structure of sequence **5**. Top and side views show the conformation of H monomer. The *N*-methyl group carbon atom is shown as a yellow sphere. e, f) Two independent molecules of the crystal structure of **6**. Top and side views show different conformations of the O monomer. Atoms in the plane of the pyridine ring of O are highlighted by blue box. The first atom of each side chain is shown as a gray sphere to indicate its position. The rest of the side chains and most hydrogen atoms are omitted for clarity. Hydrogen bonds are indicated by black dashed lines.

4 of the quinoline have been removed to increase groove depth as well. B^P was synthesized in its diethylphosphonate and Fmoc protected form in five steps using protocols similar to those developed for other B monomers without side chains or with other side chains (Fig. 3d).[33] Noteworthy is the hydrogenation of the nitro group of **4c** that is performed on the sodium carboxylate, and not the carboxylic acid or the methyl ester to avoid lactam formation. Analogues of B^P could be conceived in which the main chain methylene group would bear a substituent, thus forming a stereogenic center, as in B^R (Fig. 1a).

Helix folding with H, O and N monomers

In order to investigate the conformations of these new building blocks both in solution and in the solid state, oligoamides **5**-**7** were first synthesized using previously reported solid phase synthesis (SPS) methods (Fig. 4a,b).[34] All consist of an H, O or N monomer flanked on both sides by three or four Q units which span over a helix turn. The sequences include a combination of acidic Q^D and neutral Q^A monomers to ensure solubility in aqueous media, but not to an extent that may hamper crystal growth. The positions of the neutral and acidic residues were strategically chosen on the different sides of the helix.[35] In addition, an Aib group was included at the C-terminus to prevent head-to-head stacking of the helices,[36] and a short diethylene glycol tail was added to the N-terminus for similar reasons.[35] The ^{1}H NMR spectra of these compounds in H_2O/D_2O showed sharp spectra with the signals distributed over a wide range of chemical shift values which is characteristic of helical folding (Fig. 4c).[37] Sequences **6** and **7** exhibited a single set of signals, indicative of a single, stable conformer.[38] In contrast, sequence **5** shows two distinct sets of signals, suggesting the presence of two species.

Single crystals suitable for X-ray diffraction analysis were obtained for **5** and **6** (not for **7**) and the structures in the solid state of these two oligomers could be elucidated. In both cases, the structures contained four molecules in the asymmetric unit, which represented a challenge in terms of refinement but also provided valuable information about the conformations of the new monomers. Both structures confirmed helical folding and new monomer conformations that match the original designs (Fig. 4d,e). The sequences are achiral and equal numbers of *P* and *M* helices were found in the lattices.

In the case of **6**, the four independent helices have similar overall shapes but differences are observed in the conformations of the O units. In one helix, the main chain sp^3 oxygen atom is found in the plane of the pyridine ring, meaning that there is almost no twist (7°) of the pyridine-methylene bond (Fig. 4e). In the other helices, the sp^3 oxygen atom is out of the plane of the pyridine ring – it is instead close to the plane of the preceding quinoline unit – due to torsions of up to 77° of the pyridine-methylene bonds (Fig. 4f). The N-O bonds are all twisted but take quite variable angle values, from 47° to 115°. Altogether, this point to an enhanced flexibility of monomer O, as predicted at the design step.

In the case of **5**, the four independent molecules all have very similar conformations, suggesting a robust pattern. As anticipated in the initial design, the H monomer's methyl group points away from the helix core, and its hydrazide proton forms a hydrogen bond with the endocyclic nitrogen atoms of the adjacent pyridine and quinoline rings (Fig. 4d). The pyridine-hydrazide bond is twisted (from 31 to 37°) which is not possible in M monomers. The N-N bond is *gauche* (MeN-NH torsion angles from 118 to 124°). This *gauche* conformation of the acylhydrazide (O=C-NR-NR-C=O) is one of the common forms.[39] It was also observed in the crystal structure of **1** and of a related QH dimer in which the N-N twist angle was close to 90° (Fig. S1). Completely flat *anti* conformations of acyl hydrazides have also been observed.[40] In any case, it is reasonable to assign the solid state structure of **5** to one of the two sets of signals observed in the NMR spectra.

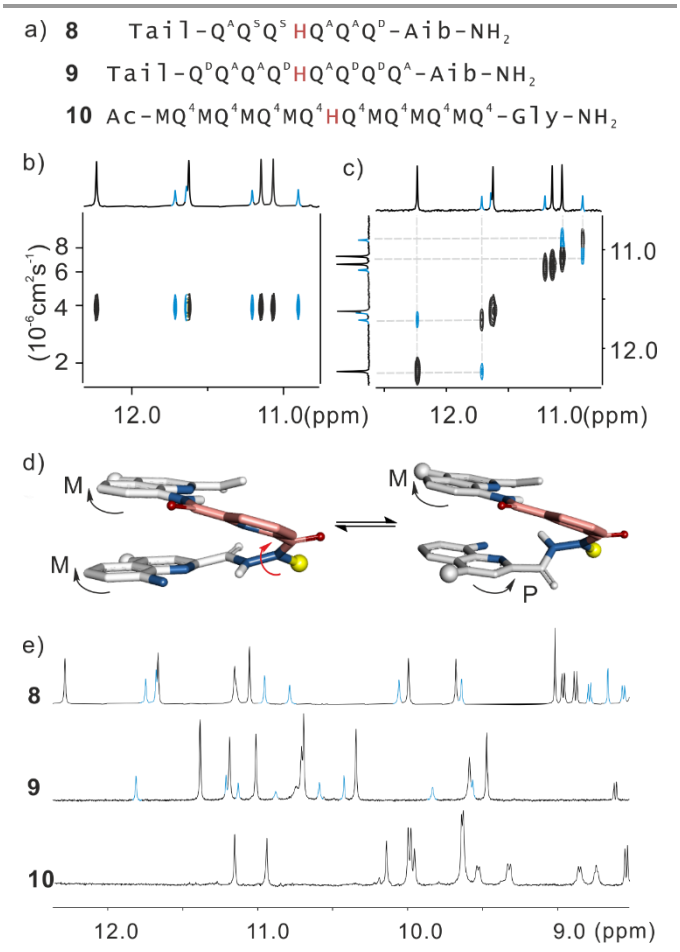


Fig. 5 (a) A summary of aromatic amide foldamer sequences **8**-**10**. Excerpt from ^{1}H DOSY NMR spectrum (b) and ^{1}H ROESY NMR spectrum (c) of sequence **8** in which two different species coexist (500 MHz, H_2O/CD_3CN 3:1 v/v, at 25°C). (d) Two proposed conformations of sequence **8**, with the H monomer shown in pink and the methyl group represented as a yellow sphere. (e) Excerpts of the ^{1}H NMR spectra of sequences **8**-**10** (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C). Signals that correspond to minor species are highlighted in blue.

Additional sequences **8**-**10** (Fig. 5a) were prepared to investigate what the other species present in the NMR spectra of sequence **5** may be. The moderate solubility of sequence **5** made it difficult to perform some 2D NMR experiments. These were instead performed using analogous sequence **8**, which contains two Q^S units instead of Q^D, resulting in better solubility. The ^{1}H NMR spectrum of **8** also showed two sets of signals (Fig. 5e). The relative intensities between the two sets of signals varied upon changing temperature and solvent (% of CD_3CN), demonstrating that the corresponding species interconvert (Figs. S2, S3). Diffusion-Ordered Spectroscopy (DOSY) revealed that two species possess comparable diffusion coefficients, hinting at two conformations rather than at some aggregation (Figs. 5b, S4). In agreement with this interpretation, the proportions are not concentration dependent (not shown). Rotating frame Overhauser effect spectroscopy (ROESY) also indicated exchange between the two species (Figs. 5c, S5). Nuclear Overhauser effect spectroscopy (NOESY) spectra were then measured. The ROESY spectra allowed us to tell for each NOE correlation whether transfer of nuclear spin polarization was due to exchange between the two species or due to a short distance between the corresponding protons. For the major species, the first (N-terminal) amide N*H* was identified through a correlation with diastereotopic CH_2. NOEs between N*H*s consecutive in the sequence allowed for their complete assignment (Fig. S6). Thus, the N*H* resonance at 8.5 ppm was assigned to the H monomer's amide, with the corresponding resonance of the minor species shifted downfield to 10.6 ppm, a difference of 2.1 ppm (Figs. S5, S6). The signal of the H monomer's methyl group was also shifted upfield in the major species ($\Delta\delta$ = 0.61 ppm). Another remarkable feature was the signal of an unassigned aromatic C*H* doublet found to be very different in the major species (7.0 ppm) and in the minor species (8.8 ppm) (Fig. S5). Altogether, these results point to distinct conformations at the H unit. A key information is a strong NOE correlation between the hydrazide N*H* and NCH_3 protons in the major species. This correlation is hardly compatible with the original design (Fig. 2b) and with the conformation of **5** in the solid state (Fig. 3d) where the H...H distance is larger than 3.2 Å. Instead, we propose an alternate conformation in which the N-N bond adopts the other *gauche* conformation (NH-NMe torsion of *ca.* 60° instead of *ca.* 120°) with a corresponding H…H distance of 2.5 Å (Fig. S7). In this conformation, intramolecular aromatic stacking was extensive when the two helix segments before and after the H unit had opposite handedness, the H unit thus promoting a local reversal of helix sense. This behavior contrasts with that of other monomers for which helix reversal also entails reduced intramolecular stacking.[38, 41]

The extended sequence **9** also displays two sets of ^{1}H NMR signals (Fig. 5e). The proportion of the minor species was reduced in comparison with **8**, hinting at length-dependent conformational preferences. Sequence **10** is a DNA mimic $(MQ^4)_n$ foldamer in which one M unit has been replaced by H. This sequence exists primarily as a single conformer resulting a single set of NMR signals (Fig. 5e) hinting at sequence-dependent conformational preferences. We made no attempt to assign the major species of **9** and **10** to one or the other conformer because it became clear at that stage that the H monomer complicates helix shape design and may not reliably be incorporated in foldamer helices without disturbing them.

DNA mimic helices with B^P monomers

Monomers like B with different side chains or stereogenic centers (*e.g.* B^R) have already been reported and their conformations in Q_n helices have been validated including in the solid state.[26, 33] Here, we assessed the behavior of B^P with a phosphonic side chain in the specific context of DNA mimic $(MQ^4)_n$ sequences. We prepared octaamides **12-14** as analogues of reference sequence $(MQ^4)_4$ **11** in which two, three, or four Q^4 units have been replaced by B^P (Fig. 6a). The oligomers were synthesized using low-loading Wang resin according to the SPS protocols mentioned above.[34] The final products were purified *via* RP-HPLC subsequently under acidic conditions for diethyl-phosphonate protected precursors and basic conditions for deprotected phosphonate products, achieving high purity (>99%) and good overall isolated yields (16-20%).

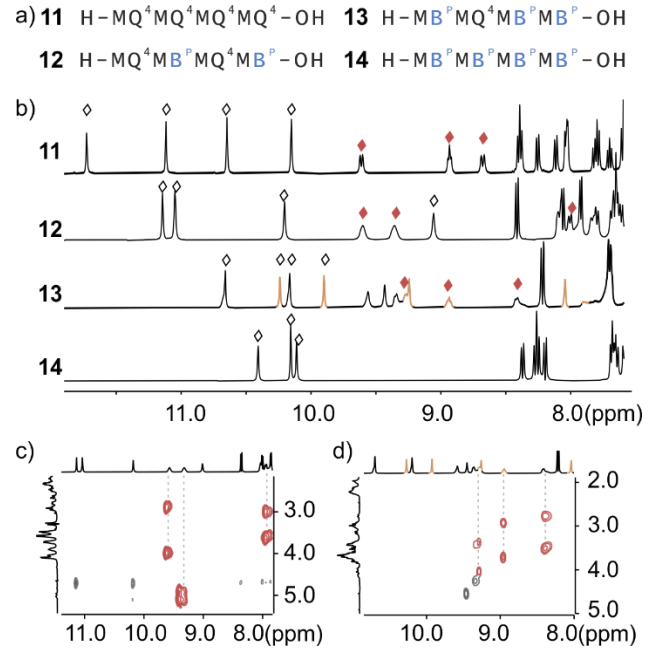


Fig. 6 (a) Aromatic amide foldamer sequences **11-14**. (b) Excerpts of the ^{1}H NMR spectra of sequences **11-14** (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C). Signals assigned to aromatic N*H*s and aliphatic N*H*s are indicated by black diamonds and red diamonds, respectively. (c) Excerpt from the ^{1}H-^{1}H TOCSY spectrum of sequence **12** and **13** showing the correlations between aliphatic N*H*s and diastereomeric pairs of CH_2 protons. Signals that correspond to two different species are highlighted in black and orange, respectively.

The ^{1}H NMR spectra of sequence **11**-**14** all display sharp amide and aromatic proton signals distributed over a wide range of chemical shift values (Fig. 6b). This pattern is consistent with previous studies on aromatic oligoamides and suggests helical conformations.[37] To further confirm helical folding in water, we probed the anisochronicity of the main chain NH-CH_2-aryl benzylic protons of M units. These protons are diastereotopic in helical conformations and may appear at distinct chemical shift values when *P* and *M* helices are under slow exchange on the NMR time scale. For both **12** and **13**, total correlation spectroscopy (TOCSY) spectra showed distinct signals for these protons for M monomers with Δδ values between 0.7 and 0.9 ppm (Fig. 6c,d). Additionally, NOE correlations between neighboring NHs further supported canonical helical folding of these two oligomers (Figs. S8, S9). In contrast, the aliphatic amide N*H* resonances of **14** cannot be seen in its ^{1}H NMR spectrum and multiple aliphatic resonances are also missing. Although this does not exclude helix folding, it suggests that faster dynamics are at play, leading to some sort of coalescence. Nevertheless, cooling down to 278 K did not result in significant changes in the spectra (Fig. S10).

In this series, **13** appeared to be an outlier as its ^{1}H NMR spectrum showed two sets of signals at higher concentration (Figs. 7, S9). The concentration dependence of the proportion of the two species as well as DOSY indicate some sort of aggregation in a discrete and well-defined species. The fact that this behavior was not observed with **12** and **14** which have the same two residues C-terminal residues as **13** exclude aggregation via the C-terminal cross-section (the reason why a C-terminal Aib has been included in **5**-**8**).[36] Aggregation of helical aromatic foldamers into multistranded helices, including in water, has been observed before,[42, 43] but not within segments that contain Q or M. It may well be that a new aggregation mode is at play here, which would bear particular interest for DNA mimics, but structural investigations have not yet been attempted.

Altogether, these results demonstrate that the B^P monomer does not hamper helix folding when combined with M monomers in the context of DNA mimic foldamers, that several B^P monomers may lead to an unidentified aggregation mode, and that they enhance conformational dynamics. Indeed, B^P contains more rotatable bonds than M and it possesses a smaller aromatic surface which is expected to reduce hydrophobic effects associated with aromatic stacking in the helix.

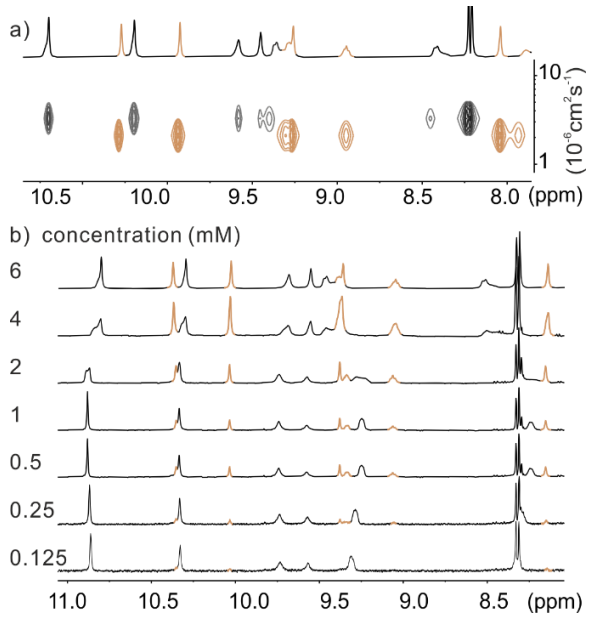


Fig. 7 (a) ^{1}H DOSY NMR spectrum of sequence **13**. (b) Excerpts of the ^{1}H NMR spectra of **13** at different concentrations. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v). Signals that correspond to two different species are highlighted in black and orange, respectively.

Quantitative assessment of conformational dynamics

The results above point to a high level of compatibility of the new monomers with the helical folding of DNA mimic $(MQ^4)_n$ sequences, at the exclusion of H which may promote alternate conformations. They also point to enhanced conformational dynamics when some of these monomers are introduced due to their reduced aromatic surface and larger number of rotatable bonds. We set to quantitatively assess DNA mimic helix stability in presence of H, O, N and B^P monomers using a recently developed assay.[44] In short, this assay consists in introducing a chiral B^R unit that quantitatively biases handedness to the *M* helix in aqueous media but only partially in presence of polar organic solvents. We allowed chiral foldamer solutions to first equilibrate in a 9:1 v/v mixture of DMSO and 15 mM aqueous NH_4OAc at 343 K. The samples were freeze-dried to obtain a solid containing both *M* and *P* conformers. These solids were then dissolved in water and the buildup of a circular dichroism (CD) band was monitored as a function of time as the effect of the B^R unit led to complete handedness bias.

For these experiments, a new series of sequences were designed and synthesized (**15-20** in Fig. 8a,b) all containing a chiral B^R unit at the same position. Sequences **15** and **17** served as references and the other sequences included two B^P units instead of Q^4 units, or two H, O, or N units instead of M units. Synthesis was performed on an automated synthesizer and the sequences were obtained in good yields. The NMR spectra of **15-19** in H_2O/D_2O 9:1 v/v showed a single set of signals (Fig. S11), indicating both helix folding and quantitative helix handedness bias. In contrast, sequence **20** exhibited at least one minor set of signals, consistent with the behavior of other H-containing sequences. The CD spectra of **15-20** in aqueous medium all showed a negative band near 360 nm characteristic of a prevalent *M* helix handedness. The CD band intensity varied despite the UV-absorbance at the same wavelength being the same (Fig. 8c,d). This tells that CD band intensity is sequence dependent and not a reliable tool to quantitatively assess handedness bias which NMR measurements already showed to be quantitative in **15-19**. We nevertheless note that **20** has the weakest CD band, consistent with possible helix handedness reversals at H units as suggested above. The CD spectra were also measured in DMSO/water to validate that handedness bias is weaker in this solvent (Fig. S12).

Table 1.Half-lives of helix handedness inversion of sequence **15-20** at different pH at 25°C.

Sequence	15-M	16-B^P	17- M	18-N	19-O	20-H
pH 5.5	- [a]	30 min	22 min	18 min	4.2 min	1.2 min
pH 8.5	54 min	4.2 min	10 min	9.0 min	2.4 min	- [b]

[a] Kinetics were too slow to monitor at 25°C. A value of 207 min has been reported at 45°C for this compound.[44] [b] Kinetics were too fast to monitor at 25°C.

The kinetic curves of CD band buildup in water at pH 5.5 are shown in Fig. 8e. These curves were fitted to single exponential decays (Figs. S13, S14) and the corresponding kinetic parameters are reported in Table 1. At pH 5.5, the phosphonic acid side chains are mostly monoanionic.[20] The helix handedness dynamics of reference sequence **15** were so slow that no kinetic parameters were extracted. In contrast, a half-life of helix handedness inversion of 30 min was measured for sequence **16**. The two B^P units of **16** thus enhance conformational dynamics but one may also point the considerable stability of **16** – 30 minutes is arguably a long time for a molecular conformational change.

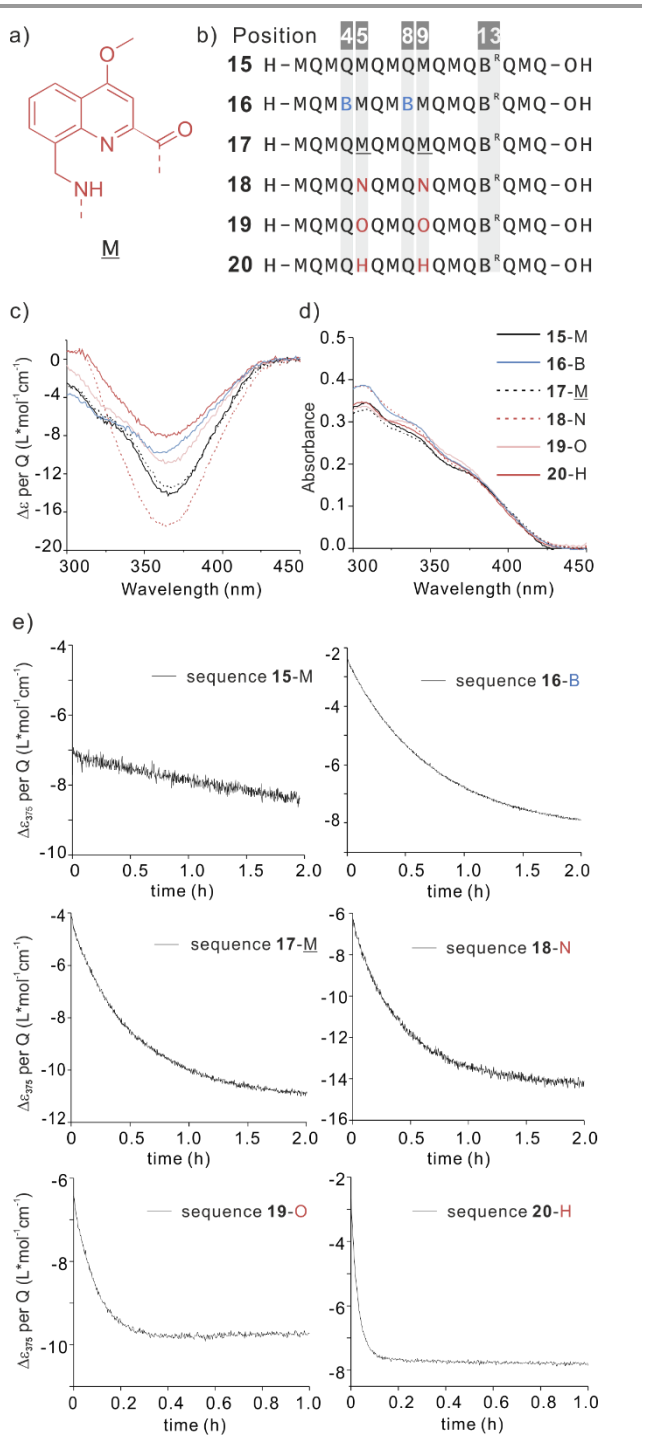


Fig. 8 (a) Structure of monomer M. (b) Foldamer sequences **15-20**. (c) CD spectra of **15-20** (40 µM, 50 mM NH_4HCO_3, pH=8.5, at 25°C). Molar extinction (Δε) is normalized for the number of quinoline rings. (d) UV absorption spectra of sequences **15-20** (40 µM, 50 mM NH_4HCO_3, pH=8.5 at 25°C). (e) Monitoring (375 nm) of *M* helix enrichment after

dissolving **15**-**20** (40 μM) in 50 mM aqueous NH_4OAc (pH 5.5) at 25°C. For clarity, the letter of a remarkable monomer that they contain is indicated after the compound number.

Sequence **17** is an analogue of **15** in which two negatively charged side chains have been replaced by neutral methoxy group in M units (Fig. 8a). Sequence **17** was designed to serve as a reference for **18**-**20** whose H, O and N units also do not contain a negatively charged side chain. The helix handedness dynamics of **17** were much faster than those of **15** and even faster than those of **16** ($T_{1/2}$ of 30 min). This result shows, somewhat unexpectedly, that the phosphonic acid side chains have a stabilizing effect. This effect has already been discussed elsewhere.[44] The kinetics of helix handedness inversion of **18** only marginally differed from those of **17** ($T_{1/2}$ of 22 min). M monomers may therefore be replaced by N monomers without significantly altering helix stability. This result is consistent with the design of N as an analogue of M that contains two aromatic rings and few rotatable bonds. It points to this design as a robust alternative even though it is the only monomer for which solid state structural evidence has not yet been obtained. In contrast, and in agreement with expectations, helix handedness inversion is faster in **19** and **20**, with $T_{1/2}$ values of 4.2 and 1.2 min, respectively. The fastest kinetics were thus observed when including monomer H, consistent with its ambivalent conformational behavior.

We also monitored helix handedness inversion dynamics by CD at pH 8.5. At this pH, the phosphonic acid side chains are largely dianionic.[20] This may lead to increased electrostatic repulsions and probably explains the observed half-lives of helix handedness inversion which all are shorter than those observed at pH 5.5. Nevertheless, the relative order of the contribution to helix stability remained unchanged: N > O > H. The effect of pH on helix stability appeared to be stronger for **15** and **16**. $T_{1/2}$ values are at least four times smaller at pH 8.5 than at pH 5.5 for these two sequences, and only two times smaller for **17**, **18**, and **19**. This larger effect may be due to the fact that **15** and **16** contain two more charged residues than the other sequences (Q^4 or B^P vs. M, N; O or H).

Conclusions

In summary, we designed and synthesized four new monomers as structural analogues for the quinoline-based M and Q units in order to tailor their contribution to the features of the major groove of $(MQ)_n$ DNA mimic foldamers. We validated the conformational behavior of these new monomers within oligoamide helices using both solid-state and solution studies. Our results show that three of the four monomers function as designed, H having a more complicated conformation behavior as anticipated. At the exception of N, the new monomers are more flexible than the original units and their incorporation leads to faster helix handedness inversion kinetics. This has moderate consequences when a more flexible monomers are introduced in a sequence. When all or almost all Q units are replaced by B^P the spectroscopic signature of the helix changes and a new self-assembled species was identified.

The new monomers expand the toolkit required to tailor DNA mimic foldamer groove features and pave the way to foldamer sequences designed to interact selectively with defined protein targets. Our results also showcase the delicate art of controlling folded structures through changes in main chain components. Indeed, structural and functional variations in biopolymers and foldamers generally derive from defined side-chain sequences on a constant main chain repeat motif. The rational arrangement of different main chain components to produce original structures, *e.g.* in peptide homologues and analogues,[45-48] in aromatic foldamers,[49] and in aliphatic-aromatic hybrid sequences,[33, 50, 51] represent an advanced level of design.

Author contributions++

JW and VC performed synthesis and experimental studies. JW performed CD kinetic experiments. JW and P.K.M performed crystal growth. P.K.M performed crystallographic analysis. IH supervised the research. JW and IH wrote the manuscript. All authors reviewed and edited the manuscript and approved its final version.

Conflicts of interest

There are no conflicts to declare.

Data availability

The supporting data has been provided as part of the Supplementary information. Crystallographic data for compound **1** and compound **1d**, oligomer **5** and oligomer **6** have been deposited at the CCDC with accession codes 2514117, 2514118, 2286782 and 2478322 respectively. The data is available upon request (www.ccdc.cam.ac.uk/).

Acknowledgements

We acknowledge financial support from the European Research Council (ERC) under the European Union's Horizon Europe Framework Programme (Grant Agreement No. ERC-2021-ADG-320892), and from the China Scholarship Council (CSC, predoctoral fellowship to J. W.). We thank L. Allmendinger for assistance with NMR measurements and P. Mayer for his assistance in solving the crystal structures of **1** and **1d**. Synchrotron data were collected at beamline ID30B at ESRF Grenoble, France and beamline P13 at EMBL Hamburg, Germany.

Notes and references

1 M. W. Gonzalez and M. G. Kann, *PLoS Comput. Biol.*, 2012, **8**, e1002819.
2 F. A. Meijer, I. A. Leijten-van de Gevel, R. M. J. M. de Vries and L. Brunsveld, *Mol. Cell. Endocrinol.*, 2019, **485**, 20-34.
3 V. Chandra, P. Huang, Y. Hamuro, S. Raghuram, Y. Wang, T. P. Burris and F. Rastinejad, *Nature*, 2008, **456**, 350-356.

4 Z. Zhai, A. Soman, J. Kuang, J. Guo, Y. Chen, L. Zhao, C. Prasanna, D. A. Putri, G. S. Glukhov, C.-P. Chng, C. Huang, L. Nordenskiöld, K. Xue and X. Shi, *Commun. Biol.*, 2025, **8**, 1758.
5 Q. Zhong, X. Xiao, Y. Qiu, Z. Xu, C. Chen, B. Chong, X. Zhao, S. Hai, S. Li, Z. An and L. Dai, *MedComm (2020)*, 2023, **4**, e261.
6 C. D. Putnam, M. J. Shroyer, A. J. Lundquist, C. D. Mol, A. S. Arvai, D. W. Mosbaugh and J. A. Tainer, *J. Mol. Biol.*, 1999, **287**, 331-346.
7 H. C. Wang, C. H. Ho, K. C. Hsu, J. M. Yang and A. H. Wang, *Biochemistry*, 2014, **53**, 2865-2874.
8 E. Michalczyk, K. Hommernick, I. Behroz, M. Kulike, Z. Pakosz-Stepien, L. Mazurek, M. Seidel, M. Kunert, K. Santos, H. von Moeller, B. Loll, J. B. Weston, A. Mainz, J. G. Heddle, R. D. Sussmuth and D. Ghilarov, *Nat. Catal.*, 2023, **6**, 52-67.
9 R. E. Bremer, E. E. Baird and P. B. Dervan, *Chem. Biol.*, 1998, **5**, 119-133.
10 D. M. Chenoweth, J. A. Poposki, M. A. Marques and P. B. Dervan, *Bioorg. Med. Chem.*, 2007, **15**, 759-770.
11 B. Johari and M. Moradi, *Methods Mol. Biol.*, 2022, **2521**, 207-230.
12 S. Yasuda, K. Morihiro, S. Koga and A. Okamoto, *Angew. Chem. Int. Ed.*, 2025, **64**, e202424421.
13 R. Brazil, *ACS Cent. Sci.*, 2023, **9**, 3-6.
14 S. Maity, M. Al-Ameer, R. K. Gundampati, S. Agrawal and T. K. S. Kumar, *Methods Mol. Biol.*, 2021, **2178**, 311-328.
15 A. Gomez Toledo, J. T. Sorrentino, D. R. Sandoval, J. Malmstrom, N. E. Lewis and J. D. Esko, *J. Histochem. Cytochem.*, 2021, **69**, 105-119.
16 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
17 A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, *Tetrahedron*, 1998, **54**, 3607-3630.
18 John C. Chaput, M. Egli and P. Herdewijn, *Nucleic Acids Res.*, 2025, **53**, gkaf635.
19 M. K. Skaanning, J. Bønnelykke, M. A. D. Nijenhuis, A. Samanta, J. M. Smidt and K. V. Gothelf, *J. Am. Chem. Soc.*, 2024, **146**, 20141-20146.
20 K. Ziach, C. Chollet, V. Parissi, P. Prabhakaran, M. Marchivie, V. Corvaglia, P. P. Bose, K. Laxmi-Reddy, F. Godde, J. M. Schmitter, S. Chaignepain, P. Pourquier and I. Huc, *Nat. Chem.*, 2018, **10**, 511-518.
21 D. Deepak, J. Wu, V. Corvaglia, L. Allmendinger, M. Scheckenbach, P. Tinnefeld and I. Huc, *Angew. Chem. Int. Ed.*, 2025, **64**, e202422958.
22 V. Corvaglia, D. Carbajo, P. Prabhakaran, K. Ziach, P. K. Mandal, V. D. Santos, C. Legeay, R. Vogel, V. Parissi, P. Pourquier and I. Huc, *Nucleic Acids Res.*, 2019, **47**, 5511-5521.
23 V. Corvaglia, I. Ait Mohamed Amar, V. Garambois, S. Letast, A. Garcin, C. Gongora, M. Del Rio, C. Denevault-Sabourin, N. Joubert, I. Huc and P. Pourquier, *Pharmaceuticals*, 2021, **14**, 624.
24 V. Kleene, V. Corvaglia, E. Chacin, I. Forne, D. B. Konrad, P. Khosravani, C. Douat, C. F. Kurat, I. Huc and A. Imhof, *Nucleic Acids Res.*, 2023, **51**, 9629-9642.
25 D. W. Zhang, X. Zhao, J. L. Hou and Z. T. Li, *Chem. Rev.*, 2012, **112**, 5271-5316.
26 D. Bindl, E. Heinemann, P. K. Mandal and I. Huc, *Chem. Commun.*, 2021, **57**, 5662-5665.
27 V. Corvaglia, J. Wu, D. Deepak, M. Loos and I. Huc, *Chem. Eur. J.*, 2024, **30**, e202303650.
28 V. Kleene, V. Corvaglia, E. Chacin, I. Forne, D. B. Konrad, P. Khosravani, C. Douat, C. F. Kurat, I. Huc and A. Imhof, *Nucleic Acids Res*, 2023, **51**, 9629-9642.
29 C. W. Garvie and C. Wolberger, *Mol. Cell*, 2001, **8**, 937-946.
30 M. Loos, F. Xu, P. K. Mandal, T. Chakrabortty, C. Douat, D. B. Konrad, M. Cabbar, J. Singer, V. Corvaglia, T. Carell and I. Huc, *Angew. Chem. Int. Ed.*, 2025, **64**, e202505273.
31 Z. Zhang, K. Tanaka and J. Q. Yu, *Nature*, 2017, **543**, 538-542.
32 S.-s. Jew, B.-s. Park, D.-y. Lim, M. G. Kim, I. K. Chung, J. H. Kim, C. I. Hong, J.-K. Kim, H.-J. Park, J.-H. Lee and H.-g. Park, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 609-612.
33 D. Bindl, P. K. Mandal and I. Huc, *Chem. Eur. J.*, 2022, **28**, e202200538.
34 V. Corvaglia, F. Sanchez, F. S. Menke, C. Douat and I. Huc, *Chem. Eur. J.*, 2023, **29**, e202300898.
35 X. Hu, S. J. Dawson, P. K. Mandal, X. de Hatten, B. Baptiste and I. Huc, *Chem. Sci.*, 2017, **8**, 3741-3749.
36 D. Bindl, P. K. Mandal, L. Allmendinger and I. Huc, *Angew. Chem. Int. Ed.*, 2022, **61**, e202116509.
37 C. Dolain, A. Grelard, M. Laguerre, H. Jiang, V. Maurizot and I. Huc, *Chem. Eur. J.*, 2005, **11**, 6135-6144.
38 C. Dolain, J.-M. Léger, N. Delsuc, H. Gornitzka and I. Huc, *Proc. Natl. Acad. Sci. U.S.A.* , 2005, **102**, 16146-16151.
39 P. Weng, X. Lin, Y.-L. Cai, X. Yan, J. Cao, Z. Li and Y.-B. Jiang, *Cell Rep. Phys. Sci.*, 2025, **6**, 102521.
40 N. Chandramouli, Y. Ferrand, G. Lautrette, B. Kauffmann, C. D. Mackereth, M. Laguerre, D. Dubreuil and I. Huc, *Nat. Chem.*, 2015, **7**, 334-341.
41 V. Maurizot, C. Dolain, Y. Leydet, J.-M. Léger, P. Guionneau and I. Huc, *J. Am. Chem. Soc.*, 2004, **126**, 10049-10052.
42 V. Koehler, G. Bruschera, E. Merlet, P. K. Mandal, E. Morvan, F. Rosu, C. Douat, L. Fischer, I. Huc and Y. Ferrand, *Angew. Chem. Int. Ed.*, 2023, **62**, e202311639.
43 B. Teng, P. K. Mandal, L. Allmendinger, C. Douat, Y. Ferrand and I. Huc, *Chem. Sci.*, 2023, **14**, 11251-11260.
44 M. Loos, L. Thurecht, J. Wu, V. Corvaglia, Z. Liu, V. Pophristic, M. Zacharias and I. Huc, *Chem. Sci.*, 2025, DOI: 10.1039/d5sc08567e.
45 I. M. Mandity, E. Weber, T. A. Martinek, G. Olajos, G. K. Toth, E. Vass and F. Fulop, *Angew. Chem. Int. Ed.* , 2009, **48**, 2171-2175.
46 J. Fremaux, L. Mauran, K. Pulka-Ziach, B. Kauffmann, B. Odaert and G. Guichard, *Angew. Chem. Int. Ed.* , 2015, **54**, 9816-9820.
47 L. Berlicki, L. Pilsl, E. Weber, I. M. Mandity, C. Cabrele, T. A. Martinek, F. Fulop and O. Reiser, *Angew. Chem. Int. Ed.*, 2012, **51**, 2208-2212.
48 S. H. Choi, I. A. Guzei, L. C. Spencer and S. H. Gellman, *J. Am. Chem. Soc.*, 2009, **131**, 2917-2924.
49 Y. Ferrand and I. Huc, *Acc. Chem. Res.*, 2018, **51**, 970-977.
50 R. V. Nair, K. N. Vijayadas, A. Roy and G. J. Sanjayan, *Eur. J. Org. Chem.*, 2014, **2014**, 7763-7780.
51 A. P. Moyer, T. A. Ramelot, M. Curti, M. A. Eastman, A. Kang, A. K. Bera, R. Tejero, P. J. Salveson, C. Curutchet, E. Romero, G. T. Montelione and D. Baker, *J. Am. Chem. Soc.*, 2024, **146**, 25501-25512.

7.2. Supplementary Information

For:

Monomer design for tailoring the groove features of DNA mimic foldamers

J. Wu, V. Corvaglia, P. K. Mandal, and I. Huc*

Contents

1. List of Abbreviations

Ac_2O	acetic anhydride
CD	circular dichroism
DBU	1,8-Diazabicyclo(5.4.0)undec-7-ene
DCM	dichloromethane
DIAD	diisopropyl azodicarboxylate
DIPEA	*N*,*N*-Diisopropylethylamine
DMF	*N,N*-dimethylformamide
DMSO	dimethyl sulfoxide
ESI	electrospray ionization
EtOAc	ethyl acetate
Et_2O	diethyl ether
Fmoc-Cl	fluorenylmethyloxycarbonyl chloride
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
LC-MS	liquid chromatography-mass spectrometry
MeOH	methanol
NMR	nuclear magnetic resonance
RP	reversed phase
PyBOP	benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
r.t.	room temperature
min	minutes
SPS	solid phase synthesis
TEAA	triethylammonium acetate
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TMSBr	trimethylbromosilane
UV	ultraviolet

2. Supplementary Figures

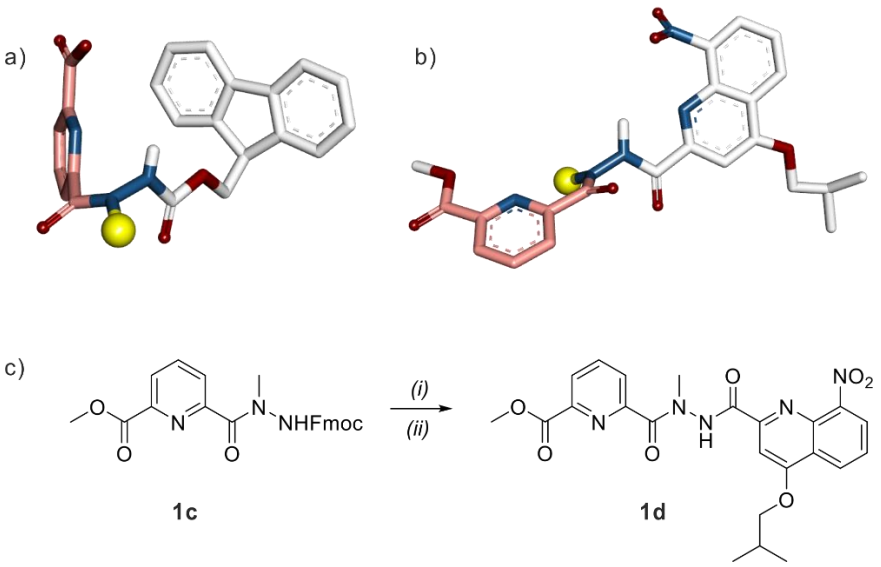


Figure S1. a) Crystal structure of compound **1** observed under crossed polarizing microscope. b) Crystal structure of compound **1d** observed under crossed polarizing microscope. H monomer is highlighted in pink. Methyl group represented as a yellow sphere. c) Synthetic route to the compound **1d**): i) DBU, $CHCl_3$, r.t., 2h; ii) 4-isobutoxy-8-nitroquinoline-2-carboxylic acid, PyBOP, DIPEA, $CHCl_3$, r.t., 2h.

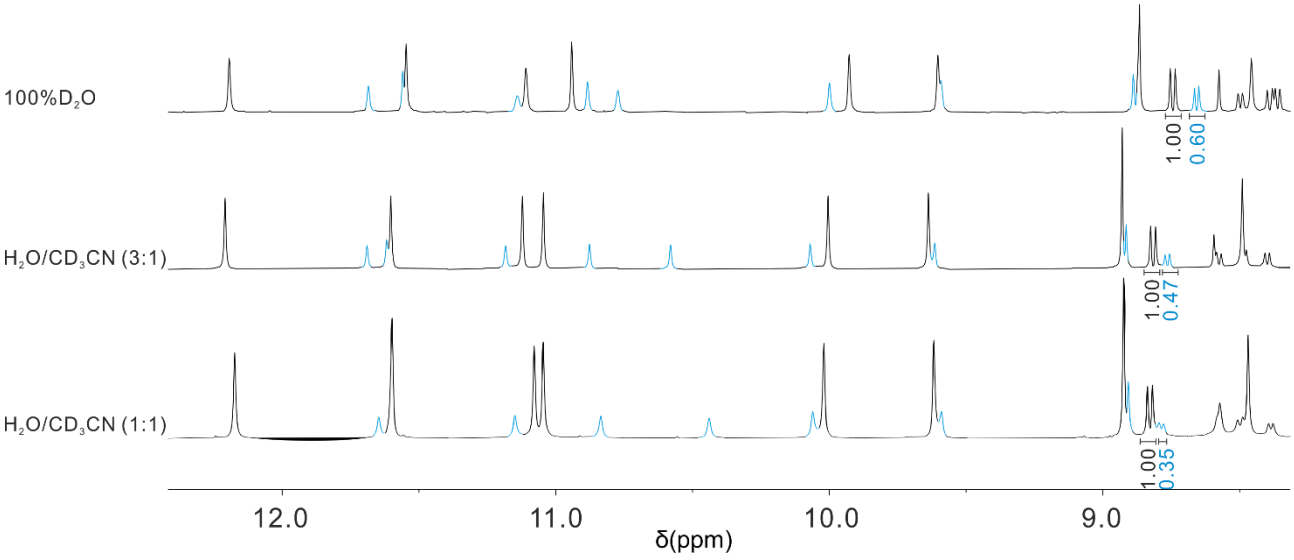


Figure S2. Excerpts of the ^{1}H NMR spectra of sequence **8** at different vol % of CD_3CN. (500 MHz, at 25°C, water suppression).

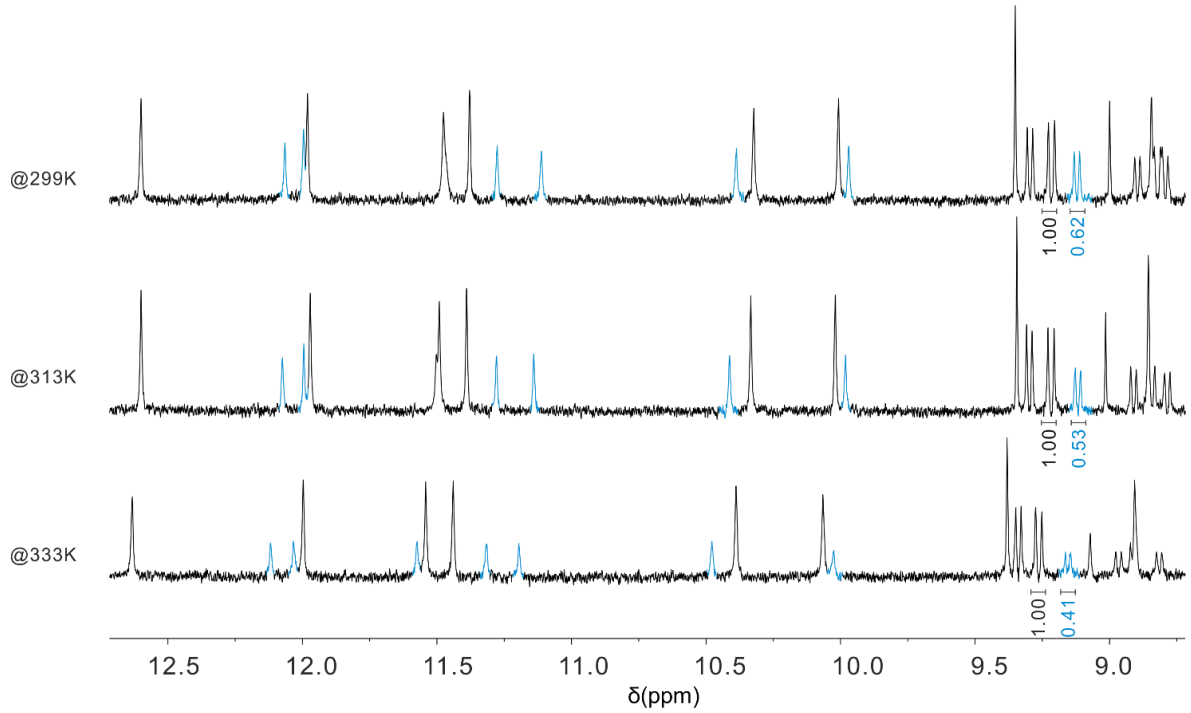


Figure S3. Excerpts of the ^{1}H NMR spectra of sequence **8** at different temperature. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

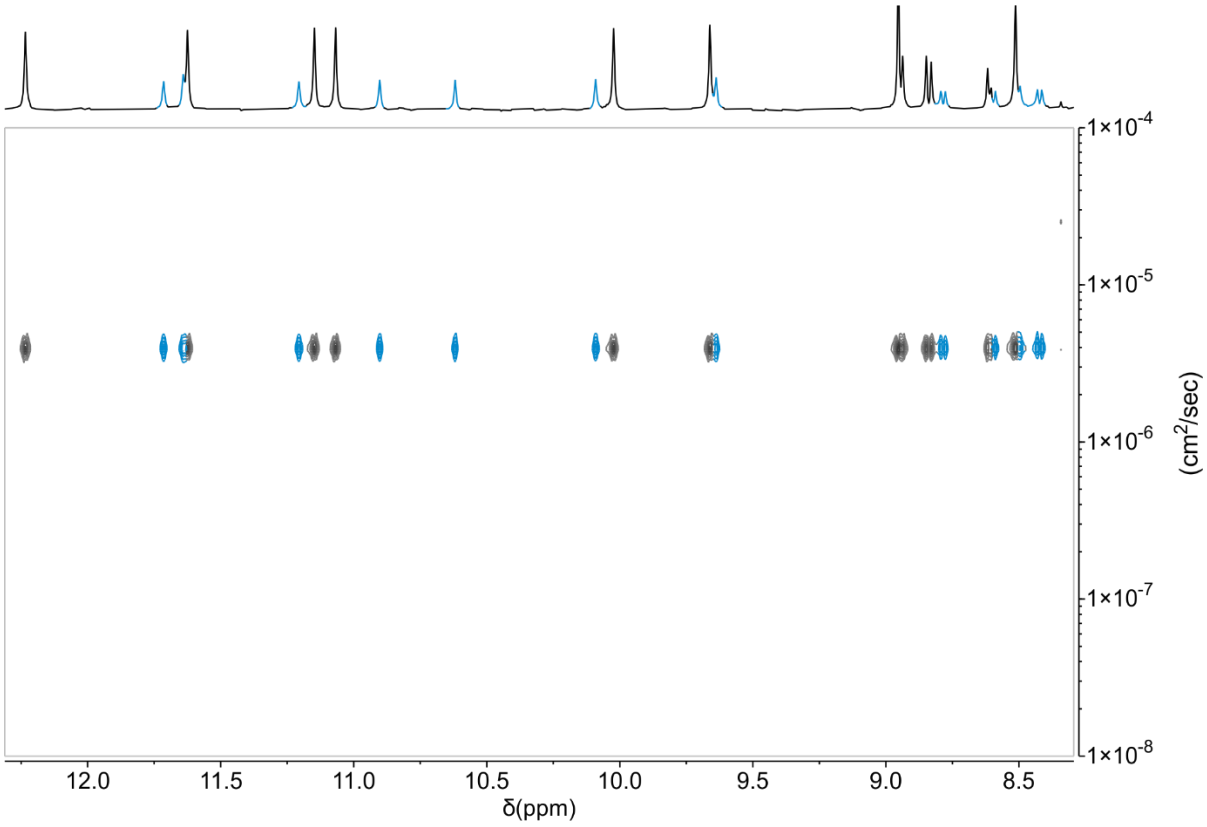
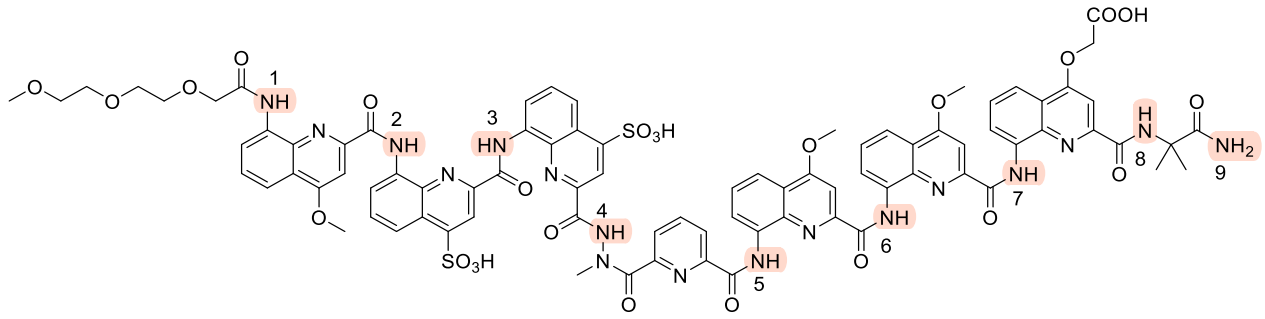


Figure S4. ^{1}H DOSY NMR spectrum of sequence **8** (500 MHz, H_2O/CD_3CN 3:1 v/v, at 25°C, water suppression).



Sequence 8

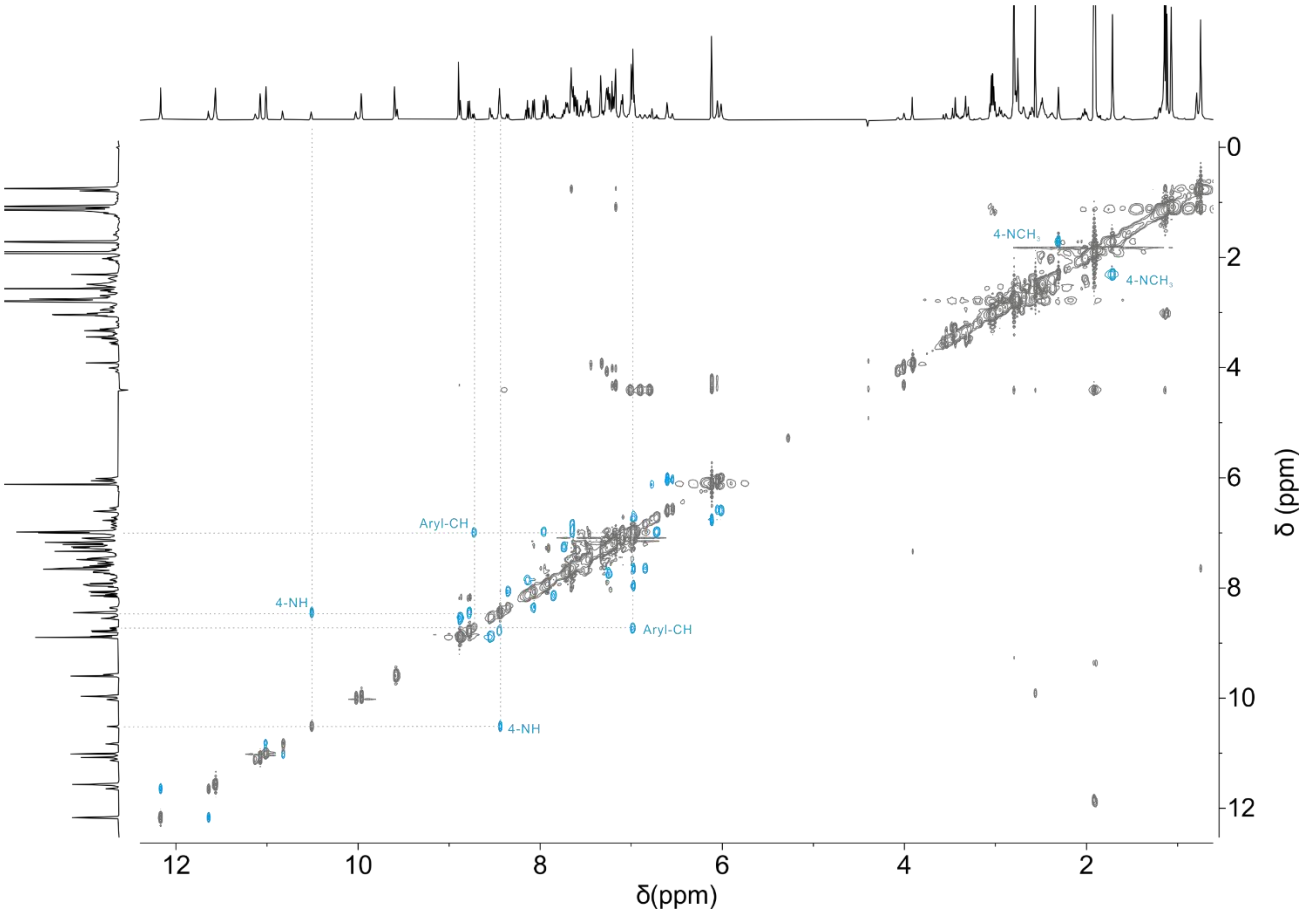


Figure S5. 2D ROESY NMR spectrum of sequence **8** (500 MHz, H_2O/CD_3CN 3:1 v/v, at 25°C, water suppression).

.

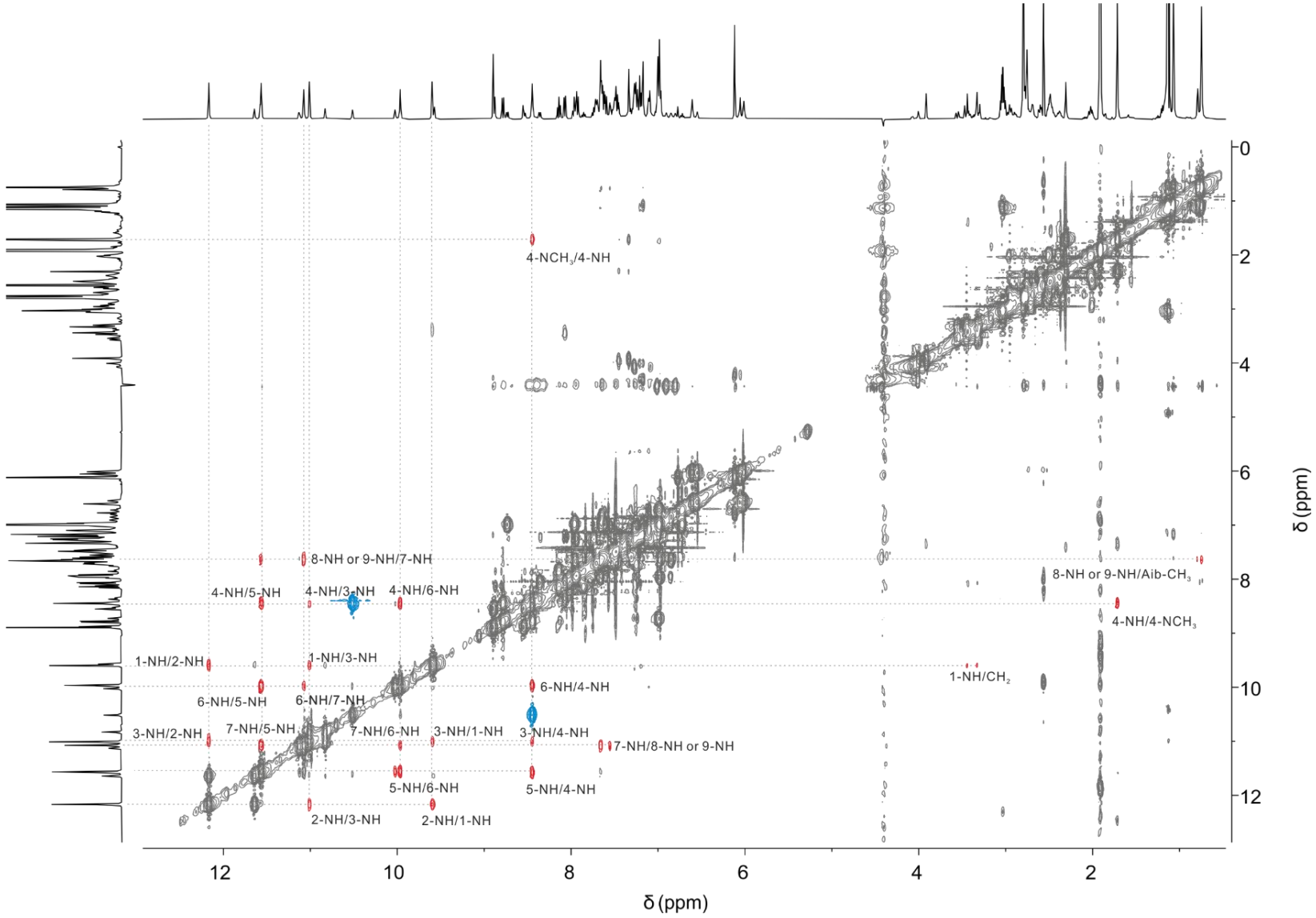


Figure S6. 2D NOESY NMR spectrum of sequence **8** (500 MHz, H_2O/CD_3CN 3:1 v/v, at 25°C, water suppression)

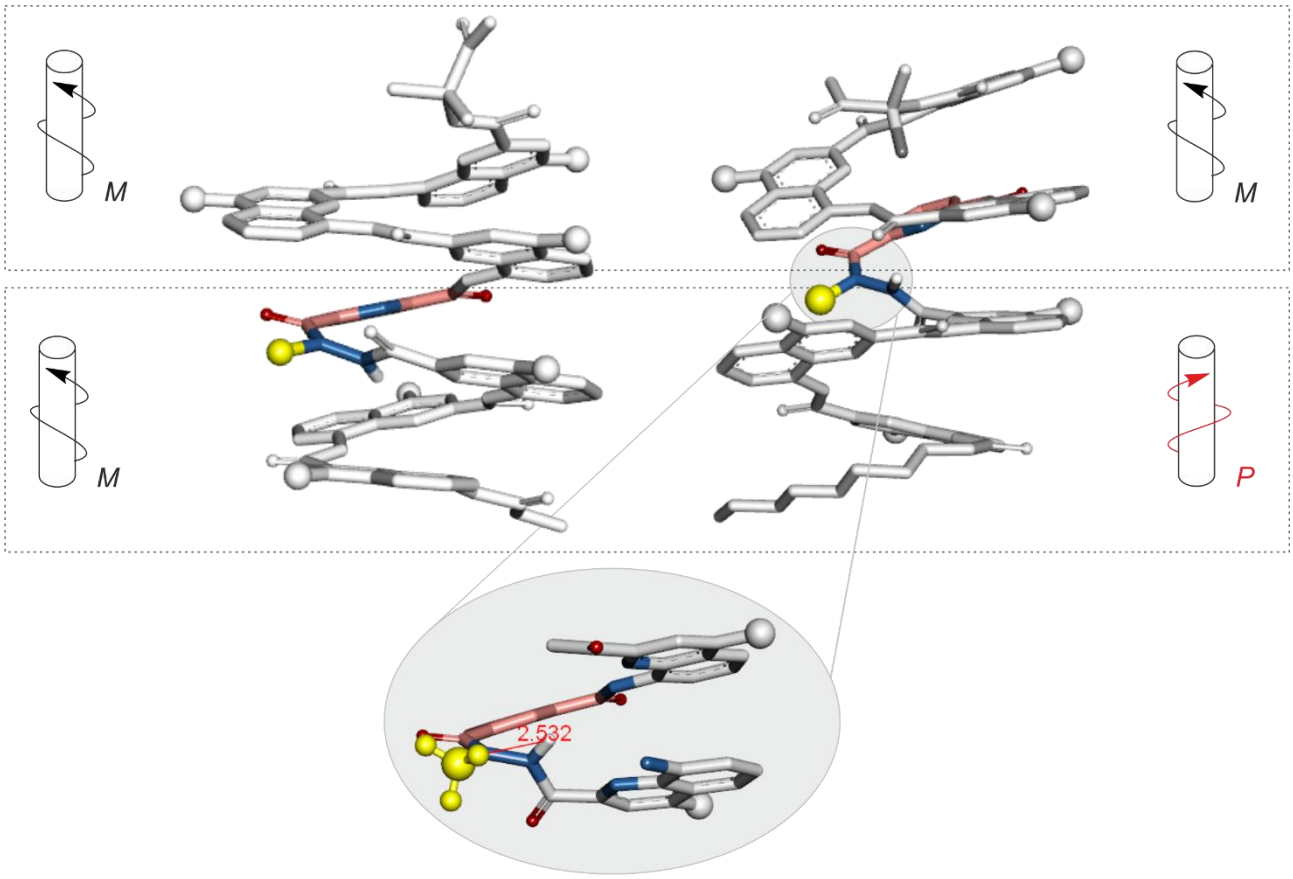


Figure S7. Left: Crystal structure of sequence **8** showing M-helicity. **Right:** Molecular model of a conformer of sequence **8** with opposite helicity on both sides of the H monomer (highlighted in pink). **Gray inset:** Close-up of the H monomer showing a flipped amide bond; the distance between the methyl group (yellow) and the NH proton is 2.5 Å. This correlation was observed by NOESY (see Fig. S6).

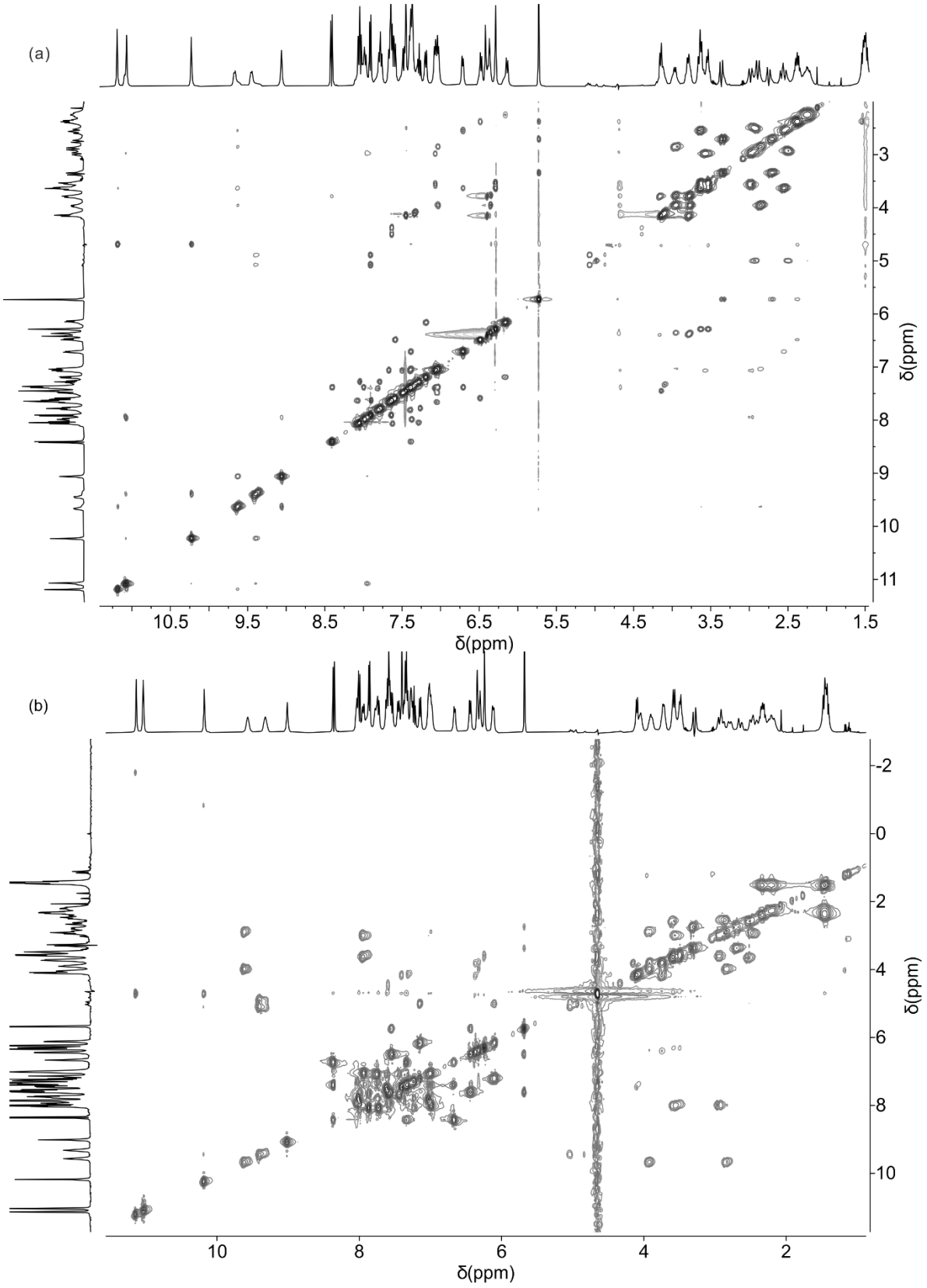


Figure S8. 2D NMR spectra of sequence **12** showing the helical folding. (a) ^{1}H-^{1}H NOESY; (b) ^{1}H-^{1}H TOCSY (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

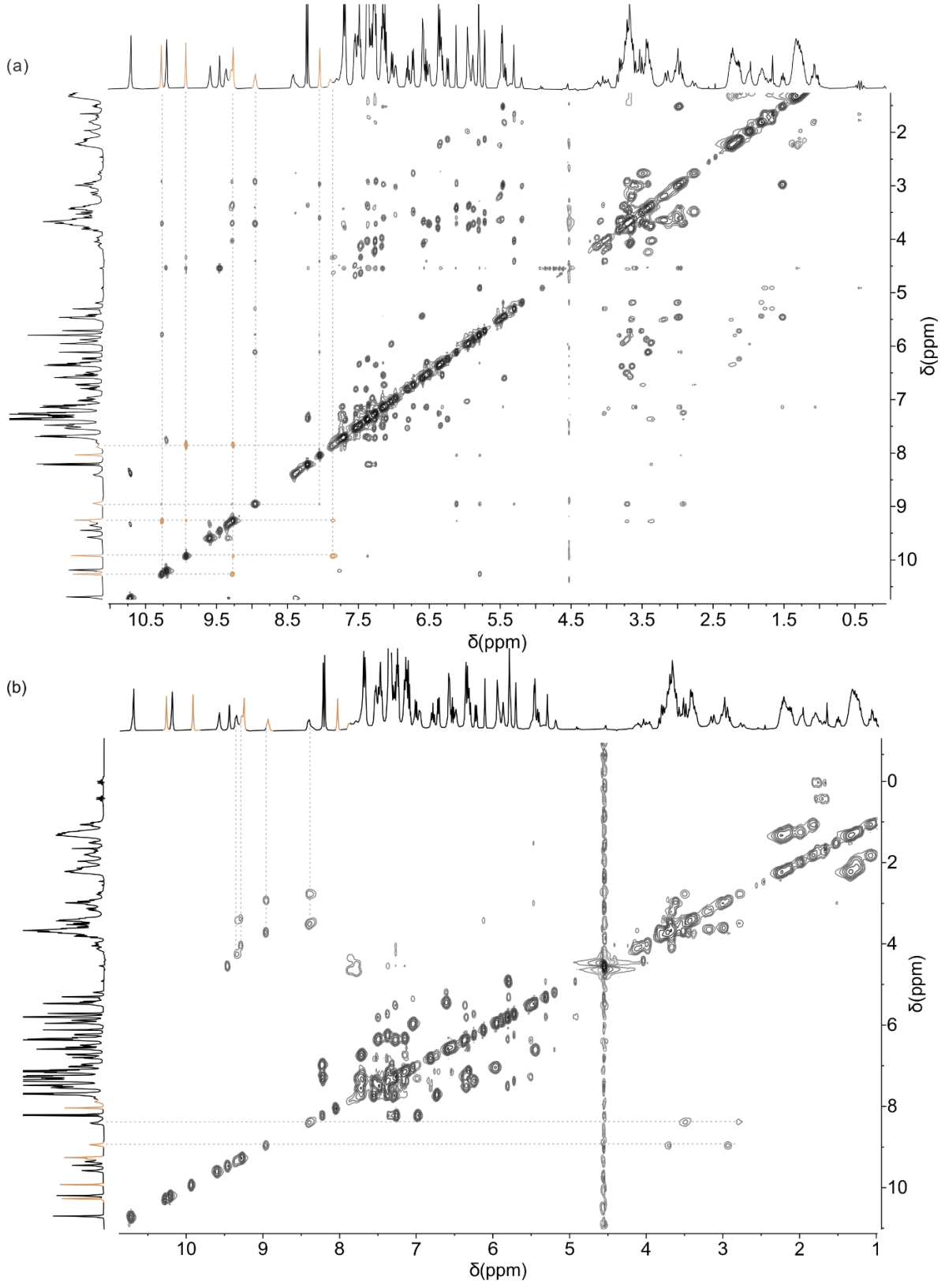


Figure S9. 2D NMR spectra of sequence **13** showing the helical folding and aggregation. (a) ^{1}H-^{1}H NOESY; (b) ^{1}H-^{1}H TOCSY (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression). Signals that correspond to two different species are highlighted in black and orange, respectively.

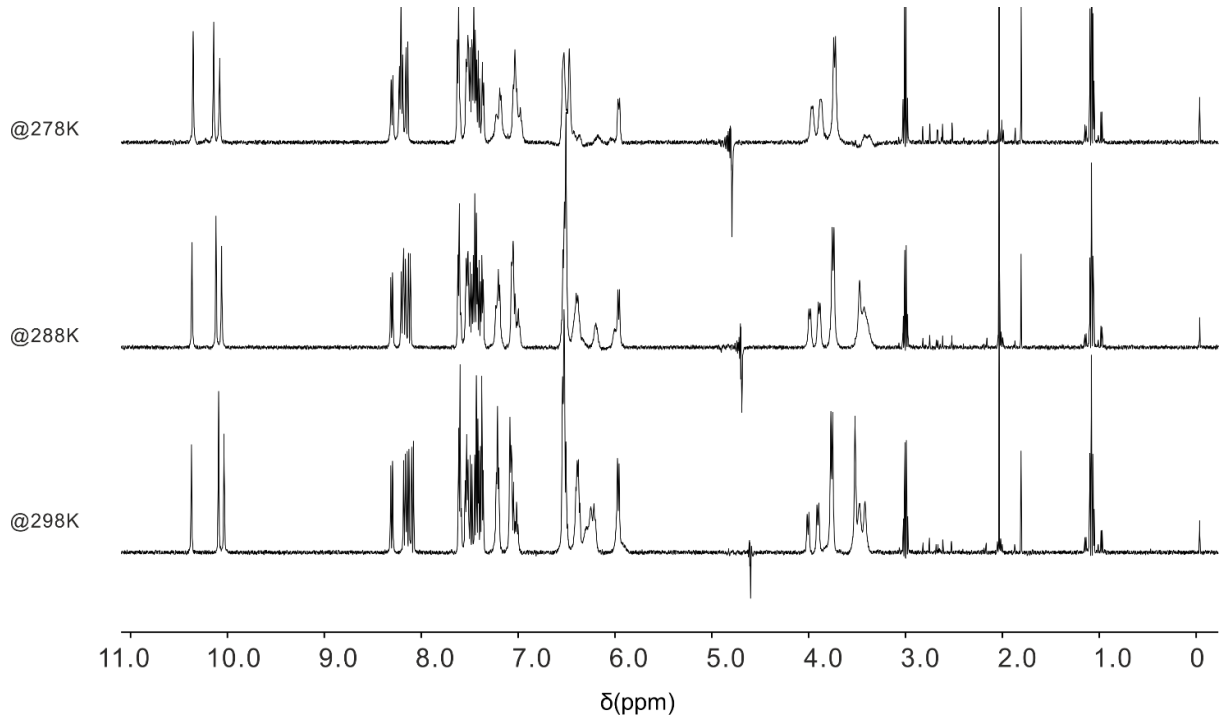


Figure S10. Excerpts of the ^{1}H NMR spectra of sequence **14** at different temperature. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

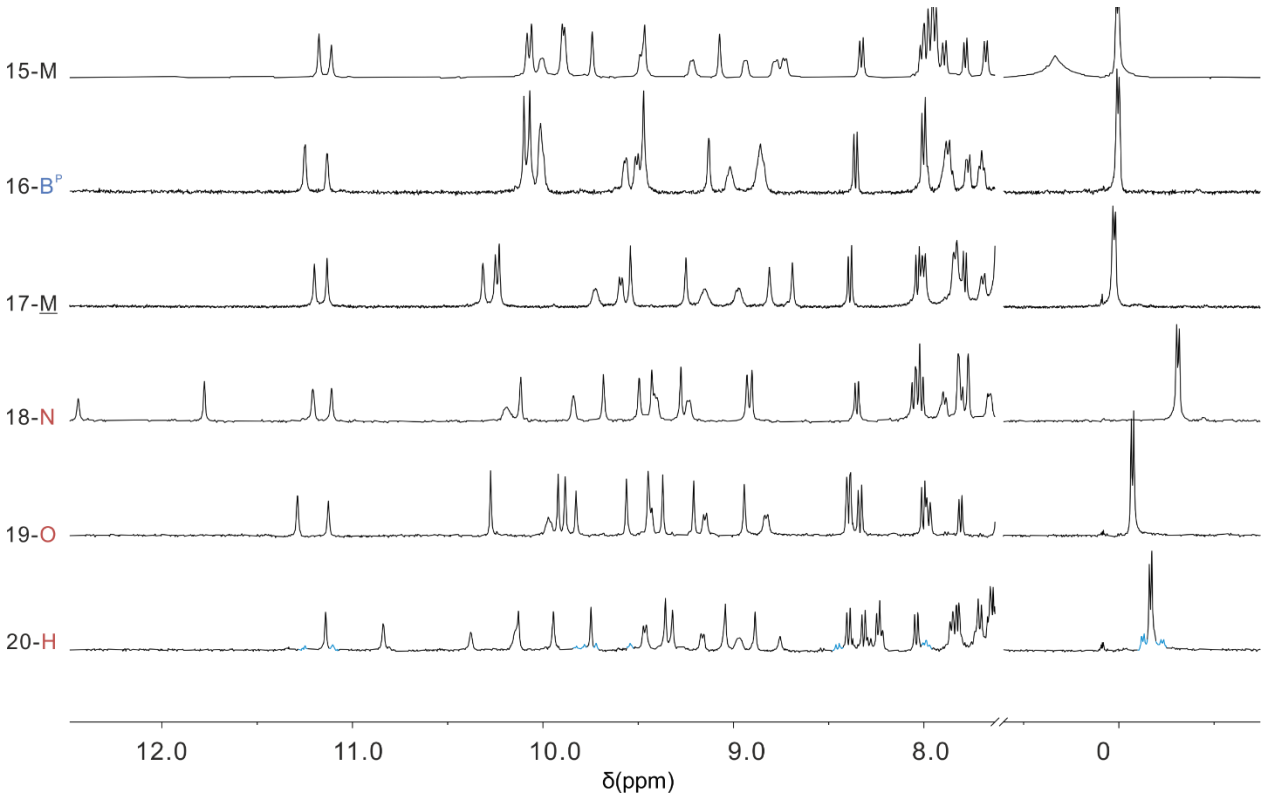


Figure S11. Excerpts of the ^{1}H NMR spectra of sequence **15-20**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression). Minor set of signals for sequence **20** is highlighted in blue.

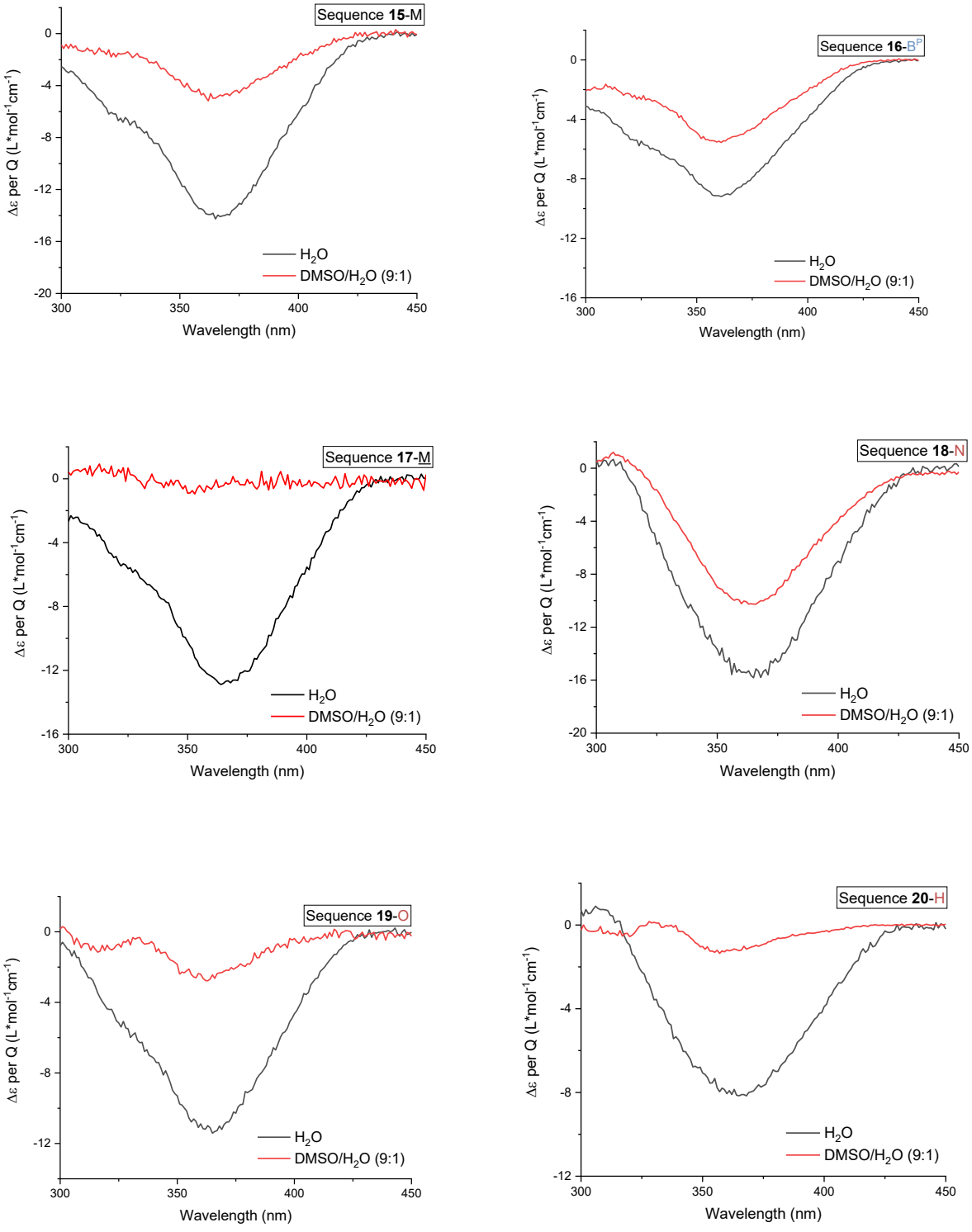


Figure S12. CD spectra of sequence **15-20** in H_2O and in DMSO/ H_2O (9:1 v/v) at 50 °C.

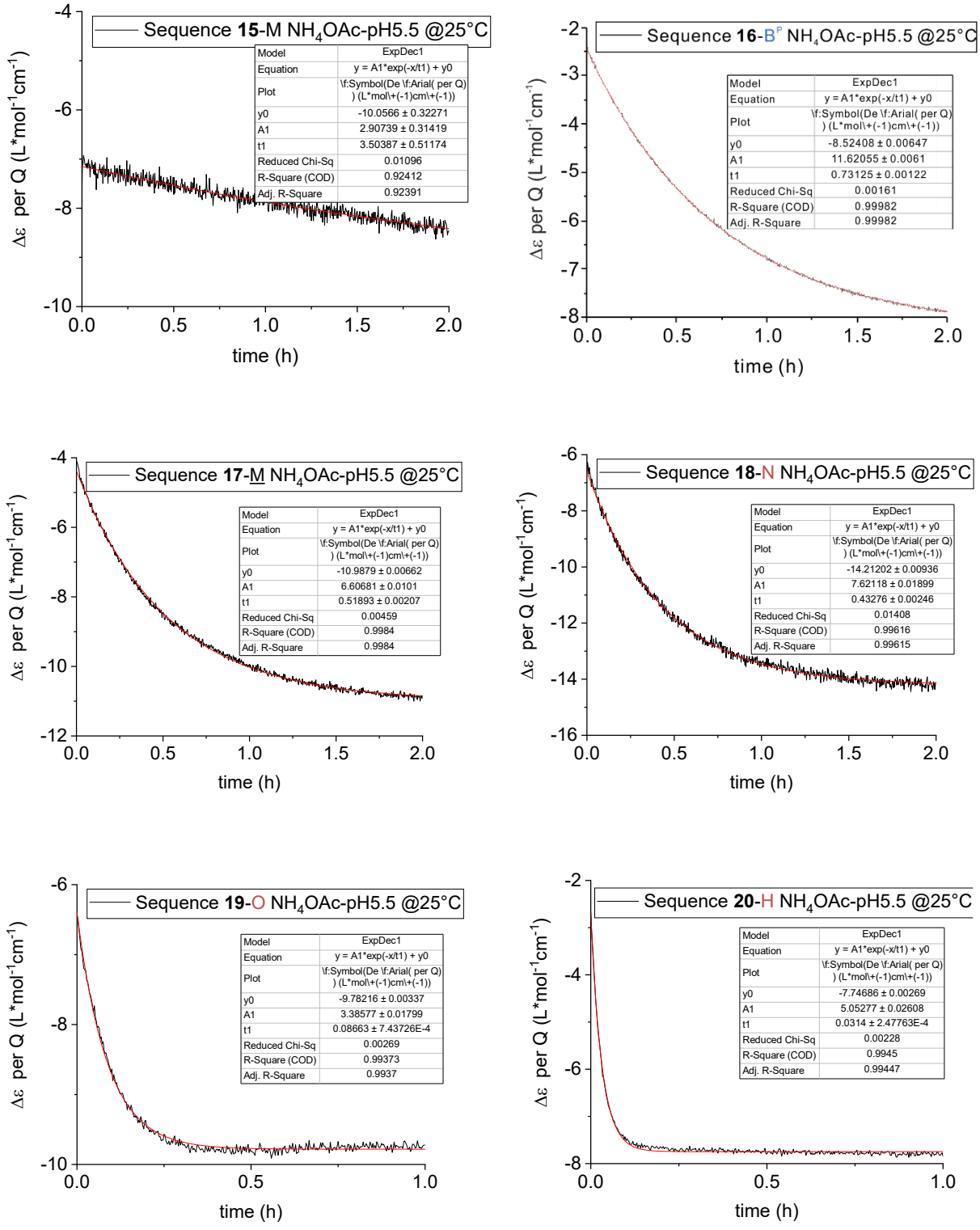


Figure S13. Helix-handedness enrichment of **15**-**20**, showing conversion of excess *P*-helix to *M*-helix (black), and a single-exponential decay fitted to the corresponding data (red) in 50 mM NH_4OAc at pH 5.5.

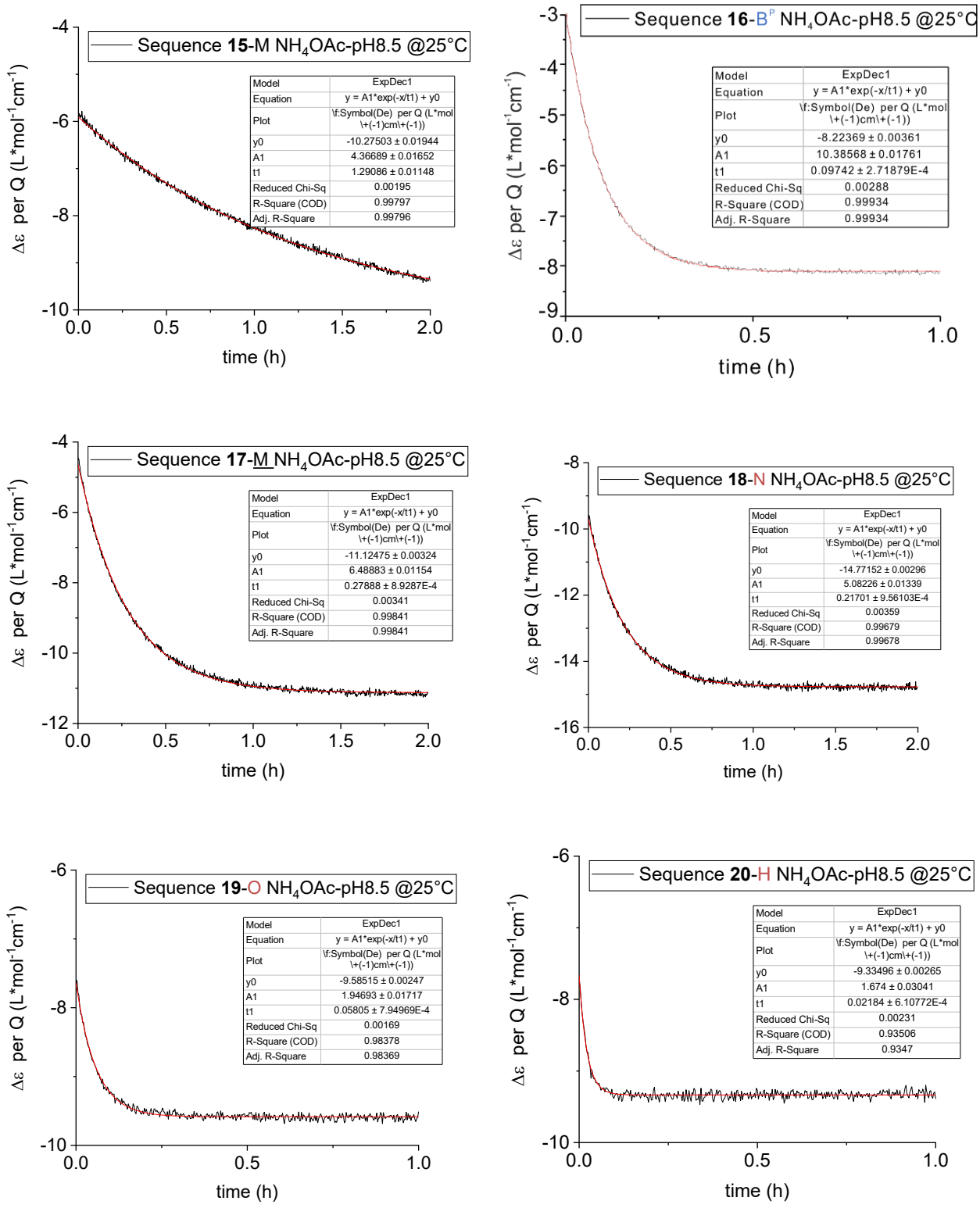


Figure S14. Helix-handedness enrichment of **15-20**, showing conversion of excess *P*-helix to *M*-helix (black), and a single-exponential decay fitted to the corresponding data (red) in 50 mM NH_4OAc at pH 8.5.

3. Materials and Methods

3.1 General

Chemicals and reagents were used as commercially supplied without any further purification unless otherwise stated. Low loading Wang resin (0.41 mmol g^{-1}) was purchased from Novabiochem. Analytical grade organic solvents were used for SPS. Anhydrous THF and DCM for SPS were dispensed from an *MBRAUN Solvent Purification System-800* solvent purification system. Reactions requiring anhydrous conditions were performed under nitrogen. Protected Fmoc-acid building blocks are shown in Figure S12.

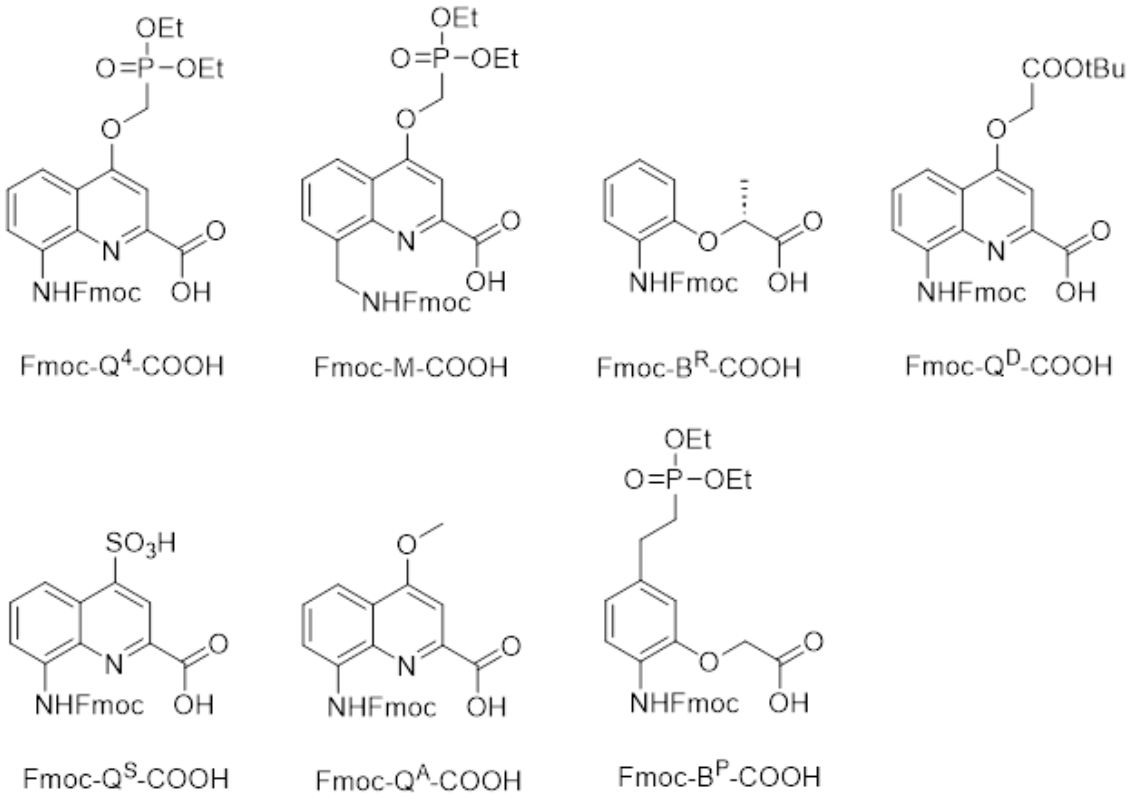


Figure S15. Side chain-protected Fmoc-acid building blocks used in this study. Fmoc-Q^4-OH, Fmoc-M-OH, Fmoc-B^R-OH,[1] Fmoc-Q^A-OH, [2] Fmoc-Q^S-OH,[3] Fmoc-Q^D-OH[4] have been described previously. For a detailed procedure to Fmoc-B-OH, see section 3.2.

Analytical RP-HPLC was performed on a Thermo Fisher Scientific Ultimate 3000 HPLC system equipped with a Macherey-Nagel Nucleodur C18 column (4.6 × 150 mm, 5 µm) at a flow rate of 1.0 mL/min. Semi-preparative RP-HPLC purification was carried out using a Kinetex C18 EVO column (10 × 100 mm, 5 µm) at a flow rate of 5.0 mL/min. For oligomer **5-9,** the mobile phase was composed of 12.5 mM NH_4OAc in water at pH 8.5 (solvent A) and acetonitrile (solvent B). For oligomer **12-20**, 0.1% TFA was added to aqueous mobile phase (referred to as mobile phase A) and to acetonitrile

(referred to as mobile phase B) under acidic conditions. When using basic conditions, the mobile phase was composed of 12.5 mM TEAA in water at pH 8.5 (solvent A) and 12.5 mM TEAA in water: acetonitrile 1:2 v/v at pH 8.5 (solvent B).

NMR spectra were recorded on different NMR spectrometers: (1) an Avance III HD NMR spectrometer 400 MHz (Bruker BioSpin) for ^{1}H NMR and ^{13}C NMR spectra of some small molecules. (2) an Avance III HD NMR spectrometer 500 MHz (Bruker BioSpin) with CryoProbe™ Prodigy for ^{1}H NMR, ^{13}C NMR, ^{1}H,^{15}N HSQC, and DOSY spectra of some small molecules and foldamers. All NMR measurements were performed at 25 °C unless specified. Chemical shifts are described in part per million (ppm, δ) relative to the ^{1}H residual signal of the deuterated solvent used – meaning DMSO-*d*$_6$ (δ_H: 2.50 ppm), $CDCl_3$ (δ_H: 7.26 ppm) and D_2O (δ_H: 4.79) For the H_2O/CD_3CN solvent mixtures, the chemical shifts were calibrated according to CD_3CN (δ 1.94 ppm). Water suppression was performed with excitation sculpting. ^{1}H NMR splitting patterns with observed first-order coupling are entitled as singlet (s), broad singlet (bs), doublet (d), triplet (t), doublet of doublets (dd) or multiplet (m). Coupling constants (J) are reported in Hz.

^{1}H,^{15}N-HSQC spectra were recorded with a phase-sensitive pulse sequence with flip-back pulse for water suppression applying a watergate sequence (hsqcfpf3gpphwg) from the Bruker pulse program library modified to make the sequence compatible with a 2-channel spectrometer. Data acquisition was performed with 512 (F2) x 64 (F1) data points in States-TPPI acquisition mode. The recycling delay was 1.0 s and 2048 transients per increment were applied at a sweep width of 8.5 kHz in F2 and 5 kHz in F1 resulting in an acquisition time of 0.062 s. Zero filling in F1 has been used to yield a final matrix of 512 x 512 real points.

The DOSY spectrum was recorded applying a pulse sequence with stimulated echo using bipolar gradient pulses for diffusion from the Bruker pulse program library (stebpesgp1s). The diffusion delay Δ (big delta) was set to 1 ms and the diffusion gradient pulse length δ (little delta) was set to 100 ms. The number of gradient steps were set to 32 with linear spacing starting from 2% reaching 95% of the full gradient strength in the final step. For each of the 32 gradient amplitudes, 64 transients of 65k complex data points were acquired. DOSY processing was performed with the DOSY processing tool from MestReNova employing the Peak Heights Fit algorithm including autocorrect peak position.

2D TOCSY spectra were recorded with a phase-sensitive pulse sequence using composite pulse scheme MLEV with water suppression employing an excitation sculpting element (mlevesgpph) from the Bruker pulse program library. Data acquisition was performed with 2048 (F2) x 256 (F1) data

points in States-TPPI mode. The recycling delay was 2.0 s and 8 transients per increment were applied at a sweep width of 8 kHz in both dimensions resulting in an acquisition time of 0.1283 s. The TOCSY mixing time was set to 80 ms. Special acquisition parameters regarding the water suppression element of the pulse sequence were adopted from the optimized parameter set of the respective one-dimensional experiment.

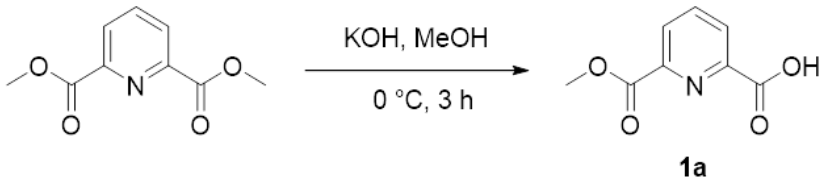
The 2D ROESY spectrum was recorded with a phase-sensitive pulse sequence with water suppression employing an excitation sculpting element from the Bruker pulse program library (roesyesgpph). Data acquisition was performed with 1K (F2) x 256 (F1) data points and a ROESY spinlock time of 0.2 s. The recycling delay was 1.0 s and 32 transients per increment were applied at a sweep width of 8 kHz in both dimensions resulting in an acquisition time of 1.64 s.

LCMS analysis was conducted on a Thermo Fisher Scientific Ultimate 3000 HPLC system with an NH_4OAc buffer system consisting of 12.5 mM NH_4OAc dissolved in ultra-pure water and adjusted to pH 8.5 with aqueous ammonia (referred to as mobile phase A) and LCMS-grade acetonitrile (referred to as mobile phase B) on a Kinetex C18 EVO column (2.1 x 50 mm, 1.8 µm) column and a flowrate of 0.33 mL/min. In all cases, elution was monitored by UV detection at 254 and 300 nm with a diode array detector. For LCMS analysis, the LC system was coupled to a micrOTOF II mass spectrometer by Bruker Daltonics and molecules were ionized by ESI.

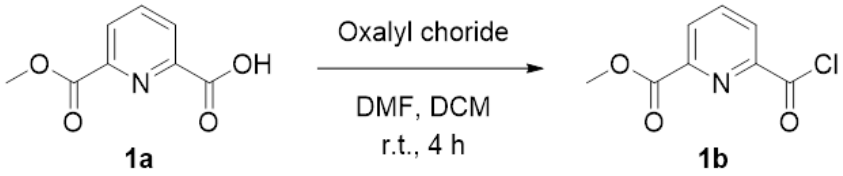
CD spectra were recorded on a Jasco J-1500 spectrometer with 1 mm quartz cuvette. The following parameters were used: wavelength range from 450 to 300 nm. Scan speed: 50 nm/min; accumulation: 2; response time: 1.0 s; bandwidth: 1; temperature: 20 °C; sensitivity: standard (100 mdeg); data pitch: 1 nm; nitrogen gas flow rate: 500 L/h. Molar extinction values were normalized per quinoline units. The data shown are the mean of two measurements and were smoothed using a Savitzky-Golay filter with a polynomal order of 3. Time-course measurements were recorded at 375 nm with a D.I.T of 2 seconds and a data pitch of 10.0 seconds with a Peltier element for temperature control.

UV-Vis spectra were measured on a Jasco V-750 spectrophotometer with a peltier element for temperature control. Spectra were recorded from 450 to 300 nm, a bandwidth of 2.00 nm, a continuous scanning mode with a scanning speed of 400 nm/min and a UV-Vis response of 0.06 s. All spectra were recorded in 1 mm quartz glass cuvettes at a concentration range of 30-40 µM for CD spectra and 50 µM for UV spectra. Baseline correction with the respective solvent or buffer used was implemented. DNA mimic foldamers are readily soluble in water. Concentrations were determined by UV-absorbance using an average ε value at 375 nm per monomer of 2506 $Lmol^{-1}cm^{-1}$ for $(MQ^4)_n$ sequences. [5]

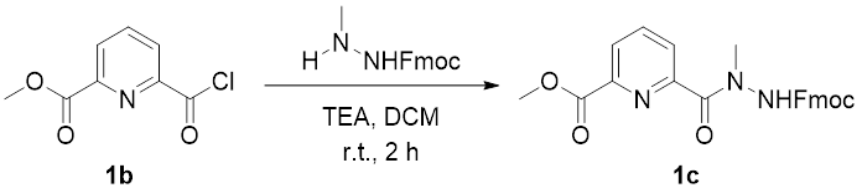
3.2 Monomer synthesis procedures



6-(Methoxycarbonyl)picolinic acid (1a): Dimethyl pyridine-2,6-dicarboxylate (3.8 g, 19.5 mmol, 1.0 eq) was dissolved in MeOH (150 mL) and the solution was cooled to 0 °C in an ice bath. KOH pellets (1.09 g, 19.5 mmol, 1.0 eq) were added portion wise, and the reaction mixture was stirred at 0 °C for 4 h. The solvent was then removed under reduced pressure. The solid residue was washed thoroughly with ethyl acetate, then dissolved in water and acidified to pH 2 using 4 M HCl. The aqueous layer was extracted with chloroform (3 × 50 mL), and the combined organic layers were dried over anhydrous $MgSO_4$. After filtration, the solvent was evaporated under reduced pressure to afford the product as a white solid (2.35 g, 67%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 13.11 (s, 1H), 7.67 – 7.41 (m, 3H), 3.21 (s, 3H). [6]

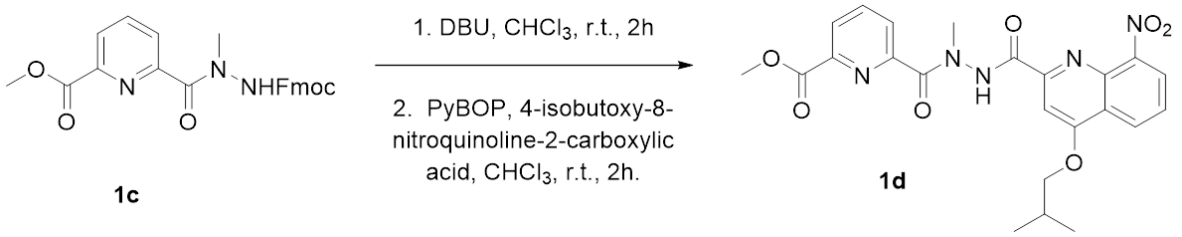


Methyl 6-(chlorocarbonyl)picolinate (1b): 6-(Methoxycarbonyl)picolinic acid **1a** (2.35 g, 13 mmol, 1.0 eq) was dissolved in anhydrous DCM (26 mL) and cooled to 0 °C under a N_2 atmosphere. To the stirred solution, two drops of DMF were added, followed by the dropwise addition of oxalyl chloride (1.34 mL, 15.5 mmol, 1.2 eq) over 10 min. The reaction mixture was stirred at r.t. for 3 h. The solvent and excess reagents were then removed under reduced pressure overnight. The crude acid chloride was used in the next step without further purification. **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.40 (dd, *J* = 7.8, 1.1 Hz, 1H), 8.29 (dd, *J* = 7.8, 1.1 Hz, 1H), 8.09 (t, *J* = 7.8 Hz, 1H), 4.06 (s, 3H). [7]

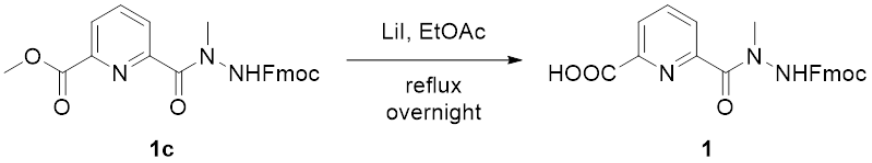


Methyl 6-(2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-1-methylhydrazine-1-carbonyl)picolinate (1c): Fmoc-methylhydrazine (3.83 g, 14.3 mmol, 1.1 eq) was dissolved in anhydrous DCM (40 mL), followed by the addition of triethylamine (2.00 mL, 14.3 mmol, 1.1 eq). A solution of methyl 6-

(chlorocarbonyl)picolinate **1b** in DCM (20 mL) was added dropwise to the reaction mixture at 0 °C. The reaction was then stirred at r.t. for 2 h. The reaction mixture was poured into water and extracted with DCM (3 × 50 mL). The combined organic layers were washed sequently with saturated $NaHCO_3$ solution and brine, then dried over Na_2SO_4. The solvent was removed under reduced pressure to afford the product as a white solid (3.2 g, 57%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.13 (d br, 1H), 7.89 (d br, 2H), 7.74 (d br, 2H), 7.40-7.24 (m, 6H), 4.30 (m, 3H), 3.98 (s br, 3H), 3.37 (s br, 3H) . **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 168.7, 164.9, 155.1, 152.8, 146.8, 143.4, 141.4, 138.4, 128.0, 127.1, 126.4, 124.8, 120.2, 67.7, 53.1, 47.0, 36.9. **HRMS** (ESI+) *m/z* calcd. for $C_{24}H_{21}N_3O_5$: 432.1913 $[M+H]^+$; found: 432.1910.

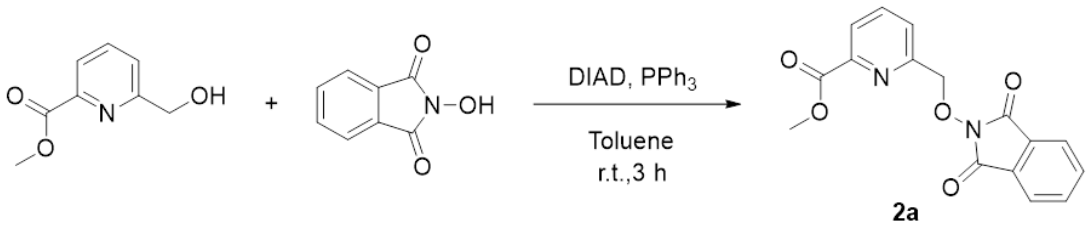


6-(2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-1-methylhydrazine-1-carbonyl)picolinic acid (1d): methyl 6-(2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-1-methylhydrazine-1-carbonyl)picolinate **1c** (130 mg, 0.3 mmol, 1.0 eq) and DBU (134 µL, 0.9 mmol, 3.0 eq) were dissolved in $CHCl_3$ (3 mL) at r.t.. After completion monitored by TLC, 4-isobutoxy-8-nitroquinoline-2-carboxylic acid (87 mg, 0.3 mmol, 1.0 eq), PyBOP (87 mg, 0.9 mmol, 3.0 eq) were added to the above mixture, and the reaction mixture was stirred at r.t. for 2h. The product was purified by column chromatography (100% EtOAc) to give the white solid (125 mg, 87%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 10.21 (s, 1H), 8.42 (dd, *J* = 8.4, 1.4 Hz, 1H), 8.05 (m, 2H), 7.90 (dd, *J* = 7.8, 1.4 Hz, 1H), 7.85 (t, *J* = 7.8 Hz, 1H), 7.62 (dd, *J* = 8.4, 7.5 Hz, 1H), 7.52 (s, 1H), 4.02 – 3.98 (m, 5H), 3.48 (s, 3H), 2.25 (dp, *J* = 13.2, 6.6 Hz, 1H), 1.09 (d, *J* = 6.6 Hz, 6H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 169.9, 165.4, 163.3, 162.1, 153.1, 151.2, 148.02, 146.7, 139.1, 138.2, 126.7, 126.5, 126.2, 126.0, 125.1, 123.6, 100.1, 76.0, 53.4, 36.5, 28.2, 19.3. **HRMS** (ESI+) *m/z* calcd. for $C_{23}H_{23}N_5O_7$: 482.1670 $[M+H]^+$; found: 482.1675.

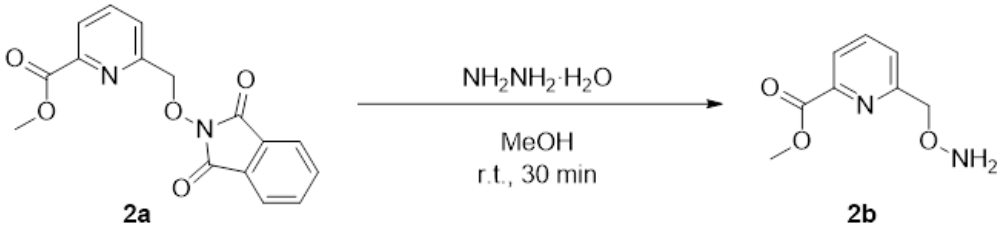


6-(2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-1-methylhydrazine-1-carbonyl)picolinic acid (1): methyl 6-(2-(((9H-fluoren-9-yl)methoxy)carbonyl)-1-methylhydrazine-1-carbonyl)picolinate **1c** (3.2 g, 7.4 mmol, 1.0 eq) and LiI (7.95 g, 59 mmol, 8.0 eq) were dissolved in degassed EtOAc (85 mL)

and heated to reflux under nitrogen. After completion monitored by TLC, the mixture was cooled to r.t., quenched with water, and acidified with 5% aqueous citric acid. The product was extracted with DCM (3 × 50 mL), and the combined organic layers were washed with brine, dried over Na_2SO_4, filtered, and concentrated under reduced pressure to give the white solid (2.38 g, 77%). **^{1}H NMR (500 MHz, CD_3OD):** δ (ppm) = 8.13 (d br, 1H), 7.89 (d br, 1H), 7.74 (d br, 2H), 7.68 (s br, 1H),7.40-7.24 (m, 6H), 4.30 (d br, 2H), 3.98 (dd br, 1H), 3.37 (s br, 3H). **^{13}C NMR (126 MHz, CD_3OD):** δ (ppm) = 172.3, 170.2, 157.3, 152.8, 152.3, 144.7, 142.5, 139.5, 128.8, 128.2, 126.9, 126.2, 125.9, 120.9, 68.4, 57.5, 48.1, 36.9. **HRMS** (ESI+) *m/z* calcd. for $C_{23}H_{19}N_3O_5$: 418.1397 $[M+H]^+$; found: 418.1391.

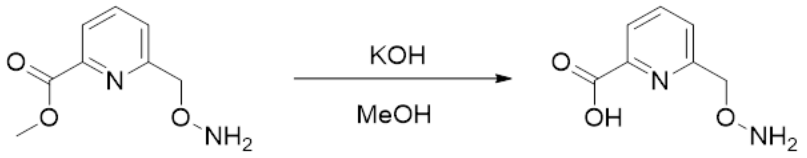


Methyl 6-(((1,3-dioxoisoindolin-2-yl)oxy)methyl)picolinate (2a): To a solution of methyl 6-(hydroxymethyl)picolinate (1.5 g, 9 mmol, 1.0 eq), PPh_3 (3.0 g, 11.7 mmol, 1.3 eq), and *N*-hydroxyphthalimide (1.61 g, 9.9 mmol, 1.1 eq) in anhydrous toluene (90 mL), DIAD (2.3 mL, 11.7 mmol, 1.3 eq) was added dropwise at r.t.. The reaction mixture was stirred for 3 h, then concentrated under reduced pressure. The resulting oil was covered with Et_2O and left overnight to give a white precipitate, which was collected by filtration (2.7 g, 95%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.12 (d, *J* = 7.7 Hz, 1H), 8.07 (d, *J* = 7.7 Hz, 1H), 7.96 (t, *J* = 7.7 Hz, 1H), 7.82 (m, 2H), 7.75 (m, 2H), 5.45 (s, 2H), 3.95 (s, 3H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 165.6, 163.5, 155.3, 147.5, 138.0, 134.7, 129.0, 126.8, 125.2, 123.8, 80.1, 53.1. **HRMS** (ESI+) *m/z* calcd. for $C_{16}H_{12}N_2O_5$: 313.0819 $[M+H]^+$; found: 313.0813.



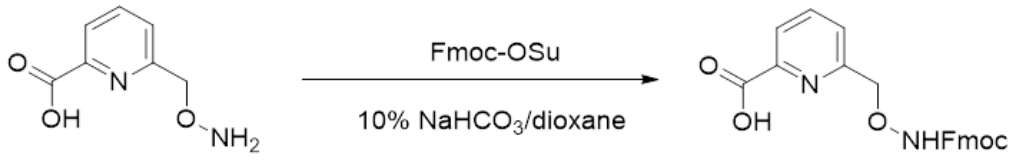
Methyl 6-((aminooxy)methyl)picolinate (2b): methyl 6-(((1,3-dioxoisoindolin-2-yl)oxy)methyl)picolinate **2a** (936 mg, 3 mmol, 1.0 eq) was suspended in MeOH (9 mL), and monohydrate (150 mL, 3 mmol, 1.0 eq) was added to the above solution in one portion. The reaction was monitored by TLC. After the reaction was complete, the precipitate was filtered off and the

remaining solution was purified by column chromatography with EtOAc, yielding a transparent oil. **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.06 (dd, *J* = 7.8, 1.1 Hz, 1H), 7.86 (t, *J* = 7.8 Hz, 1H), 7.61 (dd, *J* = 7.8, 1.1 Hz, 1H), 5.63 (s br, 2H), 4.92 (s, 2H), 4.00 (s, 3H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 165.9, 158.9, 147.8, 137.7, 125.1, 124.2, 78.2, 53.1. **HRMS** (EI+) *m/z* calcd. for $C_8H_{10}N_2O_3$: 182.0685 $[M]^{\cdot+}$; found: 182.0680. [8]



2b **2c**

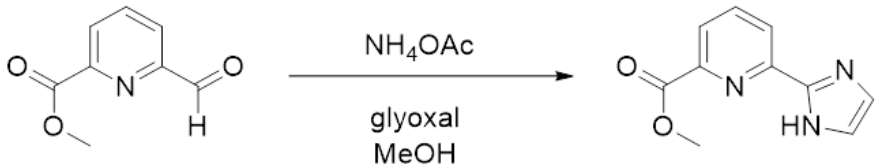
6-((aminooxy)methyl)picolinic acid (2c): methyl 6-((aminooxy)methyl)picolinate **2b** (546 mg, 3 mmol, 1.0 eq) was dissolved in 11 mL of MeOH, and a solution of KOH (168 mg, 3 mmol, 1.0 eq) in water (1 mL) was added. The reaction mixture was stirred at 0 °C and the reaction was monitored by TLC. Upon completion, the solvent was removed under reduced pressure. The residue was acidified with 1M citric acid to pH 3 and used directly in the next step without further purification. **^{1}H NMR (500 MHz, DMSO-d_6):** δ (ppm) = 7.98 (t, *J* = 7.7 Hz, 1H), 7.93 (dd, *J* = 7.7, 1.3 Hz, 1H), 7.65 (dd, *J* = 7.7, 1.3 Hz, 1H), 4.71 (s, 2H). **^{13}C NMR (126 MHz, DMSO- d_6):** δ (ppm) = 166.2, 159.0, 147.7, 137.8, 124.6, 123.2, 77.2. **HRMS** (EI+) *m/z* calcd. for $C_7H_8N_2O_3$: 168.0529 $[M]^{\cdot+}$; found: 168.0523



2c **2**

6-((((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)oxy)methyl)picolinic acid (2): 6-((aminooxy) methyl)picolinic acid **2c** (504 mg, 3 mmol, 1.0 eq) was suspended in dioxane (6 mL), and 10% $NaHCO_3$ solution (6 mL) was added. After cooling to 0 °C, Fmoc-*O*Su (1.11 g, 3.3 mmol, 1.1 eq) was added and the mixture stirred overnight. The reaction mixture was acidified to pH 3 with saturated $KHSO_4$, then extracted with DCM (3 × 100 mL) and dried over Na_2SO_4. The crude product was precipitated from acetonitrile (10 mL), sonicated, filtered, and washed with cold acetonitrile (−14 °C). Additional quantities of product were obtained from the filtrate after precipitation and freeze-drying, yielding the title compound as a white solid (1.07 g, 84%). **^{1}H NMR (500 MHz, DMSO-d_6):** δ (ppm) = 13.22 (s, 1H), 10.62 (s, 1H), 8.02 – 7.95 (m, 2H), 7.88 (d, *J* = 7.5 Hz, 2H), 7.66 (d, *J* = 7.5 Hz, 2H), 7.62 (s br, 1H), 7.40 (t, *J* = 7.4 Hz, 2H), 7.32 (t, *J* = 7.4, Hz, 2H), 4.85 (s,

2H), 4.45 (d, *J* = 6.6 Hz, 2H), 4.25 (t, *J* = 6.6 Hz, 1H). **^{13}C NMR (126 MHz, DMSO-*d*$_6$):** δ (ppm) = 166.0, 156.8, 156.3, 147.9, 143.6, 140.8, 138.1, 127.7, 127.1, 125.7, 125.1, 124.0, 120.2, 77.7, 65.8, 46.6. **HRMS** (ESI+) *m/z* calcd. for $C_{22}H_{18}N_2O_5$: 391.1289 $[M+H]^+$; found: 391.1286.

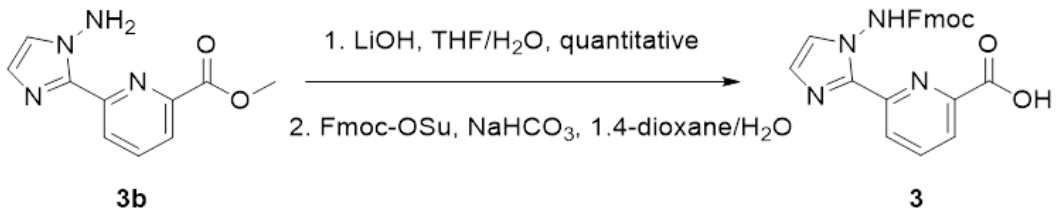


3a

Methyl 6-(1*H*-imidazol-2-yl)picolinate (3a): Methyl 6-formyl-2-pyridinecarboxylate (2.0 g, 12 mmol, 1.0 eq) and NH_4OAC (4.67 g, 60 mmol, 5.0 eq) were dissolved in MeOH (120 mL). The reaction mixture was stirred at 50°C for 30 min, then 40% glyoxal (1.41 mL, 12 mmol, 1.0 eq) was added, and stirring continued for 2 h. MeOH was removed under reduced pressure, and the crude product was purified by silica column chromatography (DCM/MeOH = 4:1), yielding a grey solid (1.5 g, 62%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 10.91 (s, 1H), 8.33 (dd, *J* = 7.9, 1.1 Hz, 1H), 8.02 (dd, *J* = 7.9, 1.1 Hz, 1H), 7.90 (t, *J* = 7.9 Hz, 1H), 7.25 (s, 1H) , 7.15 (s, 1H), 3.98 (s, 3H). **^{13}C NMR (126 MHz, $CDCl_3$)** δ (ppm) = 165.6, 149.0, 147.2, 145.7, 138.2, 130.8, 124.4, 123.2, 117.9, 52.9. **HRMS** (EI+) *m/z* calcd. for $C_{10}H_9N_3O_2$: 203.0689 $[M]^{\cdot+}$; found: 203.0686.

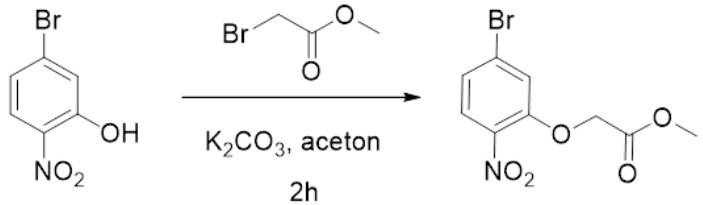
3a **3b**

Methyl 6-(1-amino-1*H*-imidazol-2-yl)picolinate (3b): Lithium hexamethyldisilazide (LiHMDS, 6 mL, 6 mmol, 1.2 eq, 1 M in THF) was added dropwise to methyl 6-(1H-imidazol-2-yl)picolinate **3a** (1.0 g, 5 mmol, 1.0 eq) dissolved in dry DMF (25 mL) at -10 °C, forming a yellow solution. After stirring for 10 min, *O*-(diphenylphosphinyl) (DPPH, 2.15 g, 10 mmol, 2.0 eq) was added at 0 °C, and the mixture was stirred at r.t. for 2 h. The reaction was quenched with water at 0 °C, and the solvent was removed under reduced pressure. The residue was washed with EtOAc (3 × 50 mL), and then combined organic phases were dried, concentrated, and purified by silica gel chromatography (DCM/MeOH = 80:1 to 40:1) to afford the product as a white solid (860 mg, 80%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 8.36 (dd, *J* = 8.0, 1.1 Hz, 1H), 8.04 (dd, *J* = 8.0, 1.1 Hz, 1H), 7.93 (t, *J* = 8.0 Hz, 1H), 7.17 (d, *J* = 1.1 Hz, 1H), 7.07 (d, *J* = 1.1 Hz, 1H), 6.67 (s, 2H), 4.00 (s, 3H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 165.2, 150.6, 146.0, 139.6, 138.3, 126.7, 124.7, 123.7, 123.5, 53.0. **HRMS** (EI+) *m/z* calcd. for $C_{10}H_{10}N_4O_2$: 218.0798 $[M]^{\cdot+}$; found: 218.0804



3b **3**

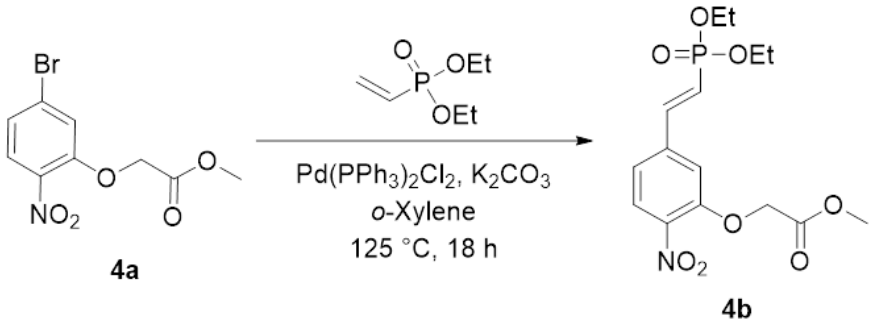
6-(1-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-1*H*-imidazol-2-yl)picolinic acid (3): methyl 6-(1-amino-1*H*-imidazol-2-yl)picolinate **3b** (1.05 g, 5 mmol, 1.0 equiv.) was dissolved in THF (80 mL). After the addition of LiOH (252 mg, 6 mmol, 1.2 equiv.) in H_2O (20 mL), the solution was stirred for 30 min at r.t.. Then, the mixture was acidified to pH 2 using 1 M citric acid. The solvents were removed under reduced pressure, and the resulting solid (1.02 g, 5.5 mmol, 1.0 equiv.) was dissolved in dioxane (97 mL). $NaHCO_3$ (9.7 g) was dissolved in H_2O (97 mL). The reaction solution was cooled to 0°C. Fmoc-Osu (1.85 g, 5.5 mmol, 1.0 equiv.) was dissolved in dioxane (47 mL) and added to the above solution. The organic solvent was removed under reduced pressure, and the remaining aqueous phase was acidified with 5% citric acid to pH 3-4, then extracted with DCM, which was dried over Na_2SO_4. After the removal of the solvent under reduced pressure, the crude product was precipitated from acetonitrile (10 mL), yielding the title compound as a white solid (1.61 g, 68%). **^{1}H NMR (500 MHz, DMSO-*d*$_6$):** δ (ppm) = 12.99 (s, 1H), 11.08 (s, 1H), 8.23 (d, *J* = 7.7 Hz, 1H), 8.07 (d, *J* = 7.7 Hz, 2H), 7.88 (s, 2H), 7.68 (s, 1H), 7.53 – 7.20 (m, 5H), 7.11 (s, 1H), 4.34 (s, 3H). **^{13}C NMR (126 MHz, DMSO-*d*$_6$):** δ (ppm) = 165.7, 148.5, 147.5, 143.5, 142.0, 140.7, 138.5, 127.7, 127.1, 126.5, 125.7, 125.2, 124.0, 120.1, 66.9, 46.4. **HRMS** (ESI+) *m/z* calcd. for $C_{24}H_{18}N_4O_4$: 427.1401 $[M+H]^+$; found:427.1393



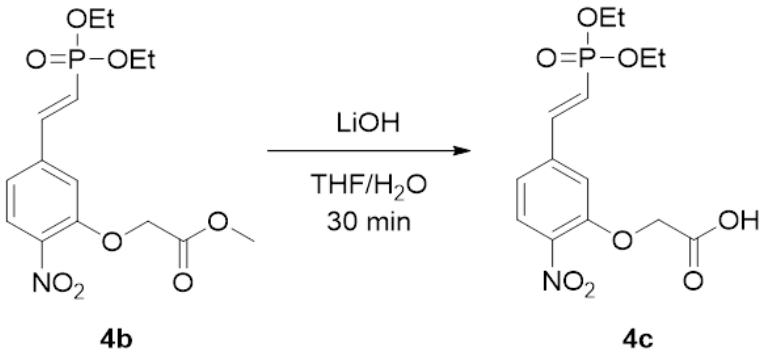
4a

Methyl 2-(5-bromo-2-nitrophenoxy)acetate (4a): 5-Bromo-2-nitrophenol (4.98 g, 23 mmol, 1.0 eq), methyl bromoacetate (2.39 mL, 25.3 mmol, 1.1 eq) and K_2CO_3 (3.50 g, 25.3 mmol, 1.1 eq) were suspended in dry acetone (90 mL). The reaction mixture was then refluxed for 2 h and filtered. The solvent was evaporated under reduced pressure and the resulting residue was extracted with water and DCM (2 x 50 mL). The compound was obtained as a light-yellow solid (12.8 g, 96%). **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm) = 7.78 (d, *J* = 8.7 Hz, 1H), 7.26 (dd, *J* = 8.7, 1.9 Hz, 1H), 7.14 (d, *J* = 1.9 Hz, 1H), 4.79 (s, 2H), 3.83 (s, 3H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 167.8, 151.9, 139.4,

128.4, 127.2, 125.3, 118.9, 66.7, 52.8. **HRMS** (ESI+) *m/z* calcd. for $C_9H_8BrNO_5$: 289.9659 $[M+H]^+$; found: 289.9654.

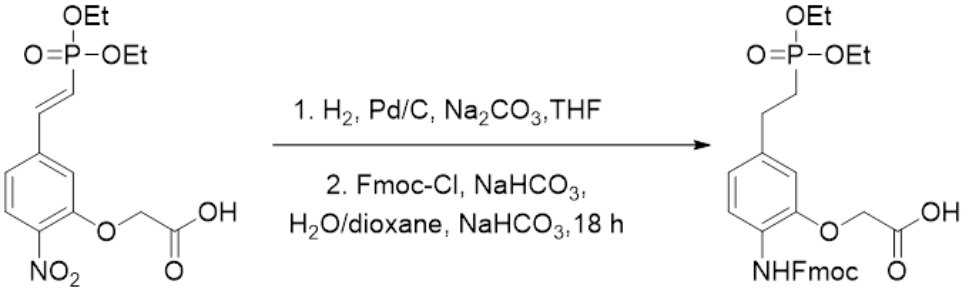


Methyl (*E*)-2-(5-(2-(diethoxyphosphoryl)vinyl)-2-nitrophenoxy)acetate (4b): In a dry flask, methyl 2-(5-bromo-2-nitrophenoxy)acetate **4a** (5.8 g, 20 mmol, 1.0 eq) and K_2CO_3 (2.76 g, 20 mmol, 1.0 eq) were suspended in dry *o*-xylene (72 mL) under N_2. Diethyl vinylphosphonate (3.68 mL, 1.2eq) and $Pd(PPh_3)_2Cl_2$ (421 mg,0.6 mmol, 0.03 eq) were added, and the reaction mixture was stirred at 125°C for 20 h. The solution was diluted with EtOAc (70 mL) and washed by 5% citric acid (3 x 50 mL). The organic phase was then washed by brine (80 mL) and dried over Na_2SO_4. The solvent was removed by vacuum and the crude was purified by RP-chromatography (from 15% acetonitrile to 100% acetonitrile) to give the target compound (6 g, 84%) as white solid. **^{1}H NMR (500 MHz, DMSO-*d*$_6$):** δ (ppm) = 7.91 (d, *J* = 8.4 Hz, 1H), 7.70 (s, 1H), 7.48 (d, *J* = 8.4 Hz, 1H), 7.40 (dd, *J* = 22.6, 17.3 Hz, 1H), 6.88 (t, *J* = 17.3 Hz, 1H), 5.09 (s, 2H), 4.11 – 3.97 (m, 4H), 3.71 (s, 3H), 1.27 (t, *J* = 7.1 Hz, 6H). **^{13}C NMR (126 MHz, DMSO-*d*$_6$):** δ (ppm)=168.3, 150.5, 145.1 (*J* = 6.8 Hz), 140.3 (*J* = 24.0 Hz), 140.0, 125.4, 120.8, 120.5 (*J* = 184Hz), 114.3, 65.48, 61.5 (*J* = 5.5 Hz), 52.0, 16.2 (*J* = 5.5 Hz). **HRMS** (ESI+) *m/z* calcd. for $C_{15}H_{20}NO_8P$: 374.1000 $[M+H]^+$; found: 374.1064.



(*E*)-2-(5-(2-(diethoxyphosphoryl)vinyl)-2-nitrophenoxy)acetic acid (4c): methyl (*E*)-2-(5-(2-(diethoxyphosphoryl)vinyl)-2-nitrophenoxy)acetate **4b** (8.2 g, 21.9 mmol, 1.0 equiv.) was dissolved in THF (320 mL). After the addition of 526 mg LiOH (21.9 mmol, 1.0 equiv.) in H_2O (80 mL), the solution was stirred for 30 min at r.t.. Then, the mixture was acidified to pH 2 using 1 M citric acid

in H_2O. The resulting solution was extracted with DCM (3 × 50 mL)) and dried over $MgSO_4$. The solvents were removed under reduced pressure, yielding 6.0 g (quant.) of the title compound as a yellow solid that was used without further purification. **^{1}H NMR (500 MHz, $CDCl_3$):** δ (ppm)= 7.89 (d, *J* = 8.4 Hz, 1H), 7.53 (dd, *J* = 22.6, 17.3 Hz, 1H), 7.22 (dd, *J* = 8.4, 1.6 Hz, 1H), 7.10 (d, *J* = 1.6 Hz, 1H), 6.35 (t, *J* = 17.3 Hz, 1H), 4.84 (s, 2H), 4.21 – 4.11 (m, 4H), 1.36 (t, *J* = 7.1 Hz, 6H). **^{13}C NMR (126 MHz, $CDCl_3$):** δ (ppm) = 151.99, 146.93, 140.60 (*J* = 24.0 Hz), 140.34, 126.49, 119.72, 118.77 (*J* = 195 Hz), 117.24, 115.11, 66.08, 62.95 (*J* = 5.5 Hz), 16.47 (*J* = 5.5 Hz). **HRMS** (ESI+) *m/z* calcd. for $C_{14}H_{18}NO_8P$: 360.0843 $[M+H]^+$; found: 360.0846.



2-(2-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-5-(2(diethoxyphosphoryl)ethyl)phenoxy) acetic acid (4): (*E*)-2-(5-(2-(diethoxyphosphoryl)vinyl)-2-nitrophenoxy)acetic acid **4c** (2.5 g, 6.96 mmol, 1.0 equiv.) and Na_2CO_3 (738 mg, 6.96 mmol, 1.0 equiv.) were suspended in MeOH (250 mL). After the solution was degassed by vacuum N_2 cycles (3x), 250 mg Pd/C (10%w) were added and the N_2 was replaced by H_2. The reaction mixture was stirred at r.t. for 6 h, filtered over celite and washed with MeOH. Solvents were evaporated under reduced pressure to obtain the crude amine as a white solid that was used for next step without further purification. The crude amine (4.6 g, 13.92 mmol, 1.0 equiv.) and $NaHCO_3$ (24.5 g, 292.3 mmol, 21.0 equiv.) were dissolved in H_2O (255 mL). Then, Fmoc-Cl (3.6 g, 13.92 mmol, 1.0 equiv.) in dioxane (100 mL) was added dropwise at 0 °C over a period of 1 h. The reaction mixture was stirred at 0 °C for 1 h and then at r.t. for 18 h. After the mixture was acidified to approximately pH 2 using 1 M citric acid in H_2O, it was extracted with DCM (3 x 50 mL), dried over $MgSO_4$ and solvents were removed under reduced pressure. The solid was precipitated from acetonitrile as a white solid (6.3 g, 80%). **^{1}H NMR (500 MHz, DMSO-*d*$_6$):** δ (ppm)= 13.07 (s, 1H), 8.61 (s, 1H), 7.91 (d, *J* = 7.4 Hz, 2H), 7.74 (d, *J* = 7.4 Hz, 2H), 7.53 (s, 1H), 7.43 (t, *J* = 7.5 Hz, 2H), 7.34 (t, *J* = 7.5 Hz, 2H), 6.94 (d, *J* = 1.8 Hz, 1H), 6.82 (d, *J* = 8.2 Hz, 1H), 4.74 (s, 2H), 4.40 (d, *J* = 7.0 Hz, 2H), 4.30 (t, *J* = 7.0 Hz, 1H), 3.99 (s, 4H), 2.74 – 2.68 (m, 2H), 2.08 – 1.99 (m, 2H), 1.23 (t, *J* = 7.0 Hz, 6H). **^{13}C NMR (126 MHz, DMSO-*d*$_6$):** δ (ppm) = 170.5, 153.6, 143.8, 140.7, 127.7, 127.1, 125.8, 125.3, 121.0, 120.1, 113.5, 66.1, 61.0 (*J* = 5.5 Hz), 46.6, 27.8 (*J* =

4.3 Hz), 26.7 (J = 137 Hz), 16.3 (J = 5.5 Hz). **HRMS** (ESI+) *m/z* calcd. for $C_{29}H_{32}NO_8P$: 554.1938 $[M+H]^+$; found: 554.1949.

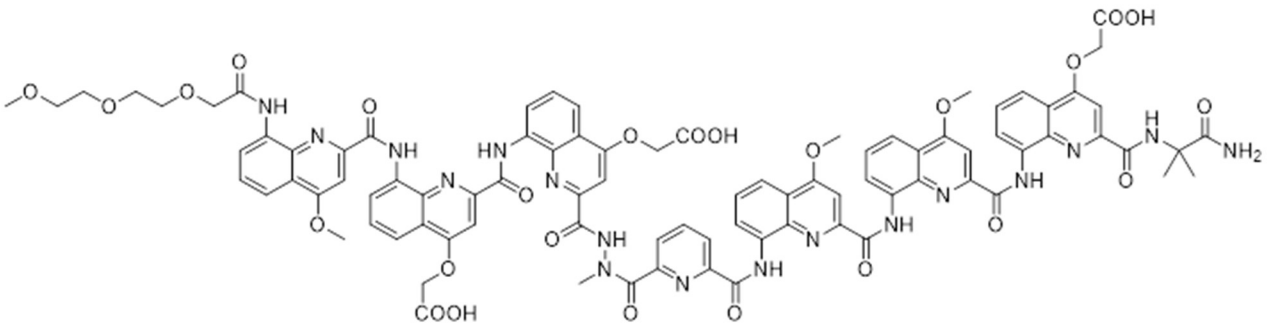
3.3 Oligomer synthesis procedures

Oligomers **5-20** were synthesized by recently reported automated solid phase foldamer synthesis (SPFS) procedures. [9] Fmoc acid building blocks were activated *in situ* by generating the respective acid chlorides prior to coupling.

Acetylation: The resin (1.0 eq.) was washed with DCM (3 x 3 mL) and incubated in Ac_2O/DCM (1:1 v:/v) for 10 min. Then, the resin was washed with DCM (2 x 3 mL) and DMF (3 x 3 mL).

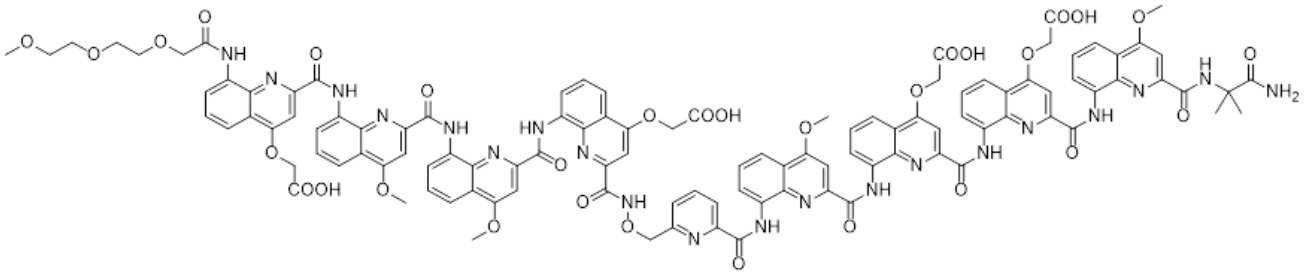
Resin cleavage and Preparative HPLC purification: The resin-bound oligomer was placed in a syringe equipped with a filter, washed with DMF (3 x 3 mL), DCM (3 x 3 mL), and dried by passing N_2 flow through it. It was then suspended in a solution of TFA. The resin was next shaken for at least 2 h at r.t. and then filtered off and washed one time with TFA. The combined solvent was removed in vacuo. After precipitation in cold Et_2O, the crude oligomer with protecting groups was purified by semi prep RP-HPLC under acidic condition to give the oligomer as a yellow solid.

Synthesis of water-soluble oligomers: The previously purified oligomer was treated by TMSBr to remove the ethyl groups. Subsequently, the crude was purified by semi prep RP-HPLC under basic fconditions (as described in section 3.1) to give the oligomer as a yellow solid. Subsequently, an ion exchange process was performed to obtain the side chains as water-soluble ammonium phosphonate salts. The removal of ethyl phosphonate protecting groups and ion exchange were performed as previously described.

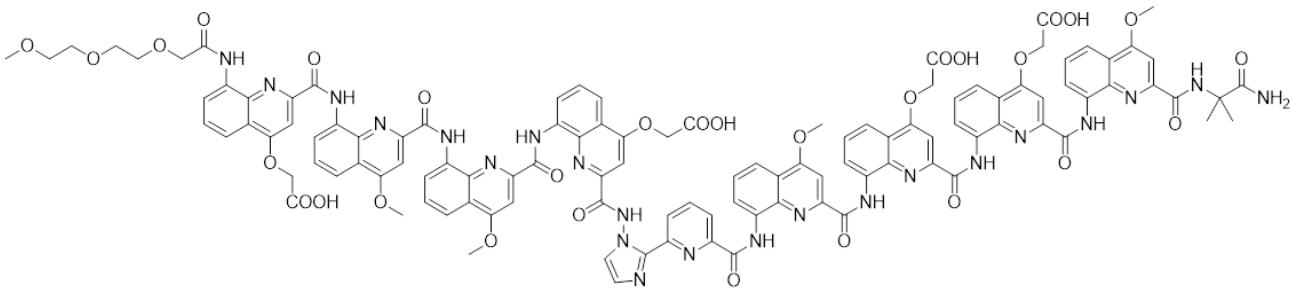


Tail-$Q^AQ^DQ^DHQ^AQ^AQ^D$-Aib-NH_2 (5): Oligomer **5** was synthesized by the automated SPFS. [9]The title compound was obtained as a white solid (15.2 mg, 44%).**^{1}H NMR (500 MHz, H_2O/D_2O 9:1):** δ (ppm) = 12.17 (s, 1H), 11.70 (s, 1H), 11.60 (s, 1H), 11.58 (s, 1H), 11.27 (s, 1H), 11.20 (s, 1H), 11.09 (s, 1H), 11.04 (s, 1H), 10.56 (s, 1H), 9.98 (s, 1H), 9.95 (s, 1H), 9.64 (s, 1H), 9.54 (s, 1H) 8.78 (d, J = 8.1 Hz, 1H), 8.73 (d, J = 8.1 Hz, 1H), 8.52 (d, J = 9.0 Hz, 1H), 8.48 (s, 1H), 8.42 (d, J = 8.0

Hz, 1H), 8.19 (d, *J* = 8.1 Hz, 1H), 8.16 (d, *J* = 8.9 Hz, 1H), 8.00 (dd, *J* = 20.1, 9.4 Hz, 5H), 7.89 (d, *J* = 8.2 Hz, 1H), 7.84 – 7.61 (m, 8H), 7.59 (d, *J* = 8.9 Hz, 1H), 7.54 – 7.46 (m, 2H), 7.42 (d, *J* = 8.8 Hz, 1H), 7.39 – 7.03 (m, 17H); 7.00 – 6.74 (m, 6H), 6.10 (s, 1H), 6.03 (s, 1H), 5.70 (s, 1H), 4.07 – 3.81 (m, 11H), 3.57 – 3.34 (m, 4H), 3.13 – 2.99 (m, 2H), 2.92 (t, *J* = 12.6 Hz, 2H), 2.84 – 2.62 (m, 12H), 2.55 – 2.44 (m, 2H), 2.34 (s, 2H), 1.97 (s, 2H), 1.90 (s, 2H), 1.71 (s, 3H), 1.21 – 1.15 (m, 4H), 1.13 (s, 3H), 1.00 (s, 2H), 0.91 (s, 3H). **HRMS** (ESI^-) *m/z* calcd. for $C_{88}H_{78}N_{17}O_{25}$: 1771.5282 $[M-H]^-$; found: 1771.5308.

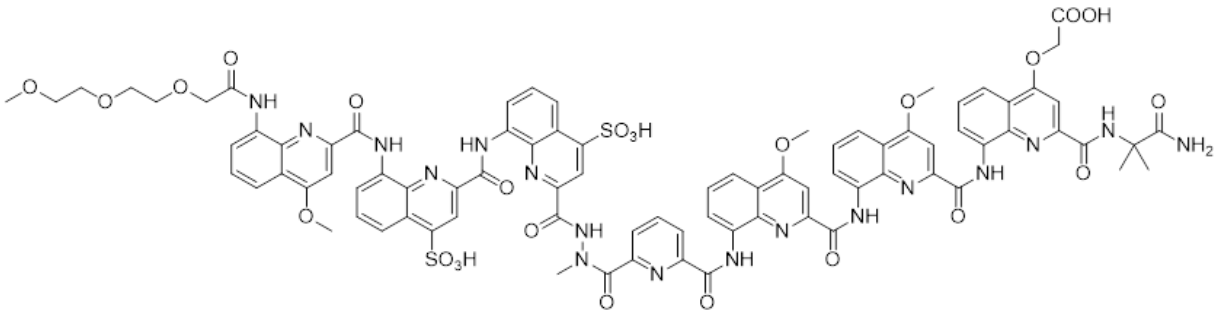


Tail-$Q^DQ^AQ^AQ^DOQ^AQ^DQ^DQ^A$-Aib-$NH_2$ (6): Oligomer **6** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (30 mg, 45%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1):** δ (ppm) = 11.53 (s 1H), 11.48 (s 1H) 11.24 (s, 1H), 10.95 (s, 1H), 10.82 (s, 1H), 10.40 (s, 1H), 10.19 (s, 1H), 9.96 (s, 1H), 9.60 (s, 1H), 8.24 (d, *J* = 8.0 Hz, 1H), 8.06 – 7.99 (m, 1H), 7.89 (m, 4H), 7.74 (m, 3H), 7.55 – 7.15 (m, 16H), 7.14 – 7.05 (m, 3H), 6.94 (s, 1H) 6.79 (s, 1H), 6.64 (d, *J* = 8.2 Hz, 1H), 6.42 (s, 1H), 6.38 (s, 1H), 6.23 (s, 1H), 6.06 (s, 1H), 4.11-3.99 (m, 6H), 3.89 (s, 3H), 3.76 - 3.73 (m, 2H), 3.57 -3.54 (m, 2H), 3.41 -3.35 (m, 2H), 2.97 -2.93 (m, 1H), 2.80 -2.55 (m, 6H), 2.41 - 2.21 (m, 4H), 2.66 (s, 3H), 2.56 (t, *J* = 4.8 Hz, 2H), 2.46 – 2.24 (m, 5H), 1.20 (s, 3H), 0.98 (s, 3H). **HRMS** (ESI^+) *m/z* calcd. for $C_{110}H_{92}N_{20}O_{31}$: 1095.3191 $[M+2H]^{2+}$; found: 1095.3201.

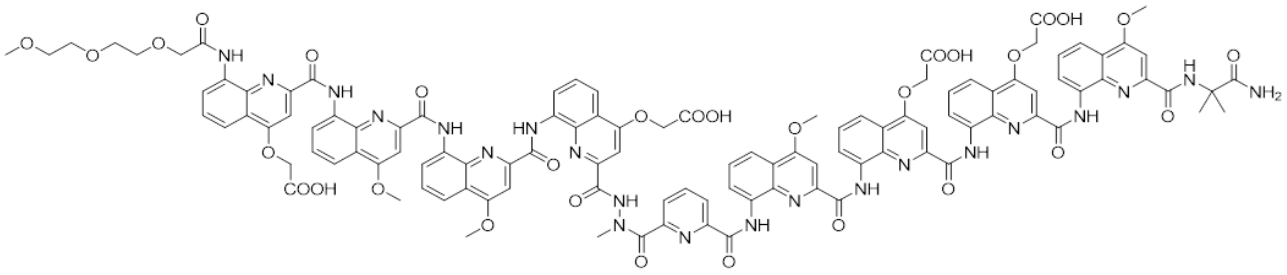


Tail-$Q^DQ^AQ^AQ^DNQ^AQ^DQ^DQ^A$-Aib-$NH_2$ (7): Oligomer **7** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (26.8 mg, 40%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1):** δ (ppm) = 12.50 (s, 1H), 11.21 (s, 1H), 11.15 (s, 1H), 10.98 (s, 1H), 10.65 (s, 1H), 10.37 (s, 1H), 9.49 (s, 1H), 9.40 (s, 1H) 8.20 (d, *J* = 7.9 Hz, 1H), 7.83 (d, *J* = 8.8 Hz, 1H), 7.79 - 7.71 (m, 4H), 7.63 - 7.60 (m, 2H), 7.47 - 7.44 (m, 2H), 7.32 - 7.23 (m, 4H), 7.19 - 7.13 (m, 4H), 7.06 – 6.73 (m,

7H) 6.55 - 6.20 (m, 6H), 5.75 (s, 1H), 4.83 – 4.65 (m, 1H), 3.99 (s, 1H), 3.91 (s, 2H), 3.75 (s, 2H), 3.32 (d, *J* = 14.6 Hz, 1H), 3.22 – 3.16 (m, 1H), 3.04 (q, *J* = 7.7 Hz, 1H), 2.73 (dd, *J* = 11.9, 9.1 Hz, 1H), 2.52 (s, 1H), 2.48 (s, 3H), 2.39 (t, *J* = 4.8 Hz, 2H), 2.23 – 2.12 (m, 2H), 2.11 – 2.05 (m, 1H), 1.67 – 1.60 (m, 1H), 1.19 – 1.07 (m, 2H), 0.98 (s, 3H), 0.78 (s, 3H), 0.70 (s, 1H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{112}H_{92}N_{22}O_{30}$: 2223.6277 [M-H]$^-$; found: 2223.6283.

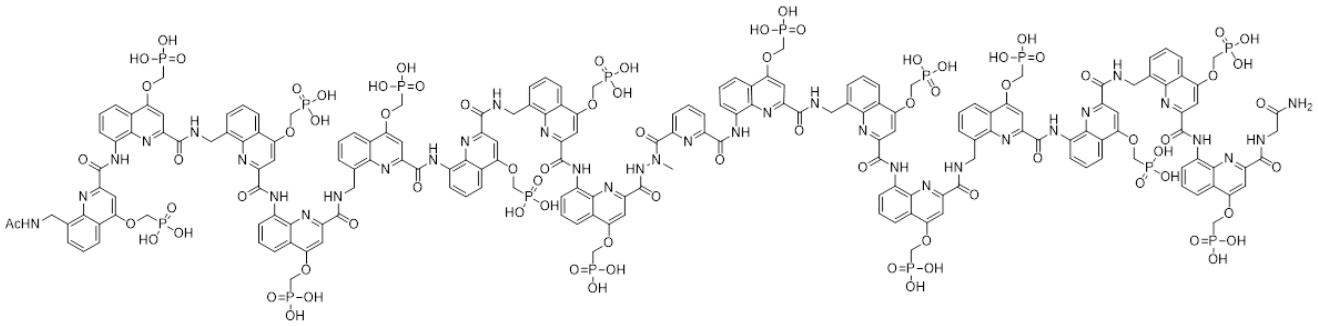


Tail-Q^{A}Q^{S}Q^{S}HQ^{A}Q^{D}Q^D-Aib-NH$_2$ (8): Oligomer **8** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (27.2 mg, 50 %).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):**δ (ppm) = 12.27 (s, 1H), 11.73 (s, 1H), 11.66 (s, 1H), 11.65 (s, 1H), 11.14 (s, 1H), 11.04 (s, 1H), 10.94 (s, 1H), 10.78 (s, 0H), 10.05 (s, 1H), 9.99 (s, 1H), 9.67 (s, 1H), 9.64 (s, 1H), 9.02 (s, 1H), 8.88 (d, *J* = 9.1 Hz, 1H), 8.79 (d, *J* = 8.1 Hz, 0H), 8.67 (s, 1H), 8.56 (d, *J* = 8.0 Hz, 1H), 8.53 – 8.44 (m, 2H), 8.28 (t, *J* = 8.6 Hz, 1H), 8.14 (d, *J* = 8.3 Hz, 1H), 8.11 – 7.84 (m, 4H), 7.83 – 7.52 (m, 6H), 7.50 – 7.31 (m, 5H), 7.24 (s, 1H), 7.22 (s, 1H), 7.19 –6.86 (m, 7H), 6.43 (s, 1H), 6.35 (s, 1H), 6.19 (s, 2H), 6.13 (s, 1H), 4.14 (s, 1H), 4.05 (s, 1H), 4.01 (s, 2H), 3.94 (s, 2H), 3.89 (s, 1H), 3.70 – 3.44 (m, 4H), 3.11 – 2.92 (m, 2H), 2.88 (s, 3H), 2.84 (s, 2H), 2.39 (s, 2H), 2.06 (s, 1H), 1.28 (s, 2H), 1.22 (s, 3H), 1.10 (s, 2H), 1.02 (s, 3H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{84}H_{73}N_{17}O_{25}S_2$: 1782.4332[M-H]$^-$; found: 1782.4340.

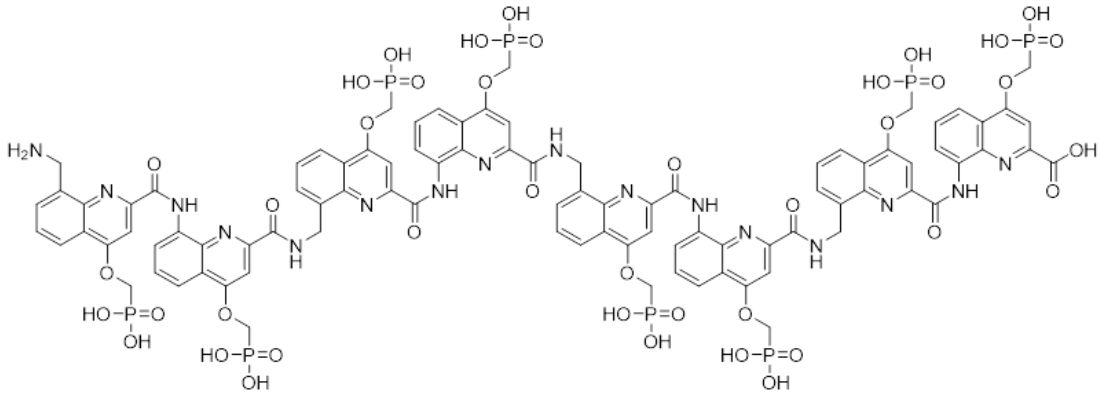


Tail-Q^{D}Q^{A}Q^{A}Q^{D}HQ^{A}Q^{D}Q^{D}Q^A-Aib-NH$_2$ (9): Oligomer **9** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (16.2 mg, 45%). **^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):**δ (ppm) = 11.38 (s, 1H), 11.19 (s, 1H), 11.01 (s, 1H), 10.70 (s, 1H), 10.69 (s, 1H), 10.34 (s, 1H), 9.59 (s, 1H), 9.47 (s, 1H), 8.41 (d, *J* = 7.8 Hz, 1H), 8.03 – 7.85 (m, 4H), 7.80 – 7.53 (m, 9H), 7.46 – 7.13 (m, 10H), 7.12 – 7.05 (m, 1H), 6.99 – 6.88 (m, 3H), 6.85 – 6.61 (m, 7H) 6.48 (s, 1H), 6.37-6.31

(m, 2H), 5.43 (s, 1H), 4.04 (s, 2H), 3.97 – 3.72 (m, 11H), 3.61 – 3.30 (m, 3H), 3.10 (q, *J* = 7.6 Hz, 1H), 2.94 – 2.81 (m, 1H), 2.78 – 2.67 (m, 2H), 2.62 (s, 1H), 2.57 (s, 3H), 2.49 (t, *J* = 4.7 Hz, 2H), 2.37 – 2.17 (m, 3H), 2.09 (s, 3H), 1.83 – 1.79 (m, 1H), 1.43 (s, 1H), 1.26 – 1.16 (m, 2H), 1.08 (s, 3H), 0.94 (s, 1H), 0.88 (s, 3H). **HRMS** (ESI^-) *m/z* calcd. for z: 2214.6274 $[M-H]^-$; found: 2214.6203

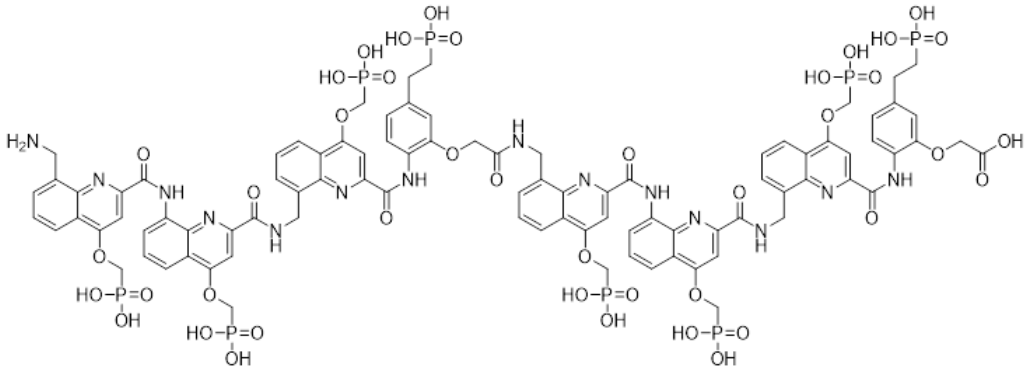


Ac-MQ^{4}MQ^{4}MQ^{4}MQ^{4}MQ^{4}HQ^{4}MQ^{4}MQ^4-Gly-NH$_2$ (10): Oligomer **10** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (14.2 mg, 45%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):**δ (ppm) = 11.15 (s, 1H), 10.94 (s, 1H), 10.14 (s, 1H), 10.04 – 9.93 (m, 3H), 9.64- 9.53 (m, 3H), 9.33 (d, *J* = 9.7 Hz, 1H), 8.86 (d, *J* = 8.7 Hz, 1H), 8.74 (d, *J* = 6.7 Hz, 1H), 8.55 (d, *J* = 8.6 Hz, 1H), 8.45 (t, *J* = 8.7 Hz, 3H), 8.32 – 8.10 (m, 8H), 8.06 – 7.90 (m, 7H), 7.83 - 7.67 (m, 10H), 7.55 – 7.26 (m, 12H), 7.23 – 7.02 (m, 13H), 7.01 – 6.78 (m, 13H), 6.76 – 6.63 (m, 8H), 6.59 – 6.44 (m, 12H), 6.35 (s, 1H), 6.30 (s, 2H), 3.25 – 3.08 (m, 3H), 3.03 – 2.89 (m, 2H), 2.77 – 2.31 (m, 5H), 2.04 (s, 4H), 1.37 – 1.22 (m, 9H). **HRMS** (ESI^-) *m/z* calcd. for $C_{184}H_{164}N_{35}O_{79}P_{15}$: 1146.8916 $[M-4H]^{4-}$; found: 1146.9021.

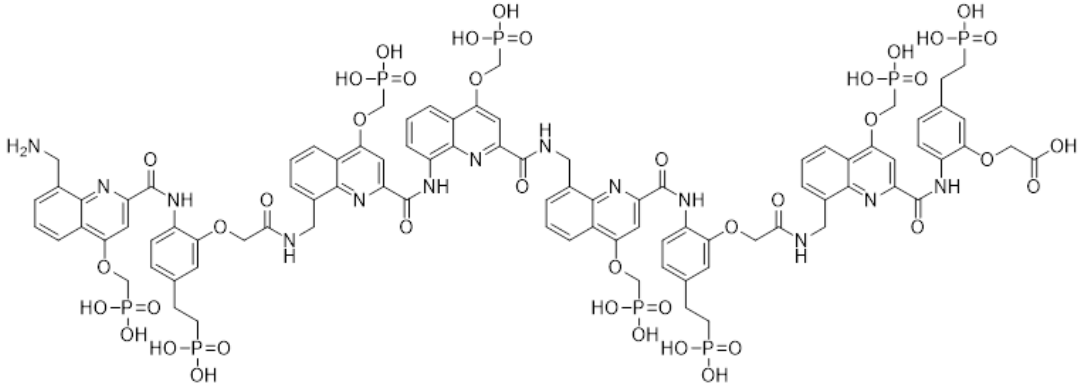


H-MQ^{4}MQ^{4}MQ^{4}MQ^4-OH (11): Oligomer **11** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (12.4 mg, 35%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):** δ (ppm) = 11.72 (s, 1H), 11.10 (s, 1H), 10.63 (s, 1H), 10.12 (s, 1H), 9.57 (d, *J* = 10.3 Hz, 1H), 8.88 (t, *J* = 6.9 Hz, 1H), 8.62 (d, *J* = 12.1 Hz, 1H), 8.33 (t, *J* = 8.5 Hz, 2H), 8.19 (d, *J* = 9.4 Hz, 1H), 8.05 (d, *J* = 9.4 Hz, 1H), 7.97 (dd, *J* = 8.0, 4.4 Hz, 2H), 7.74 (dd, *J* = 16.5, 8.4 Hz, 2H), 7.67 – 7.60 (m, 1H),

7.56 – 7.49 (m, 3H), 7.48 – 7.41 (m, 2H), 7.35 – 7.24 (m, 2H), 7.18 (d, *J* = 9.8 Hz, 3H), 7.17 – 7.11 (m, 1H), 7.06 (t, *J* = 8.6 Hz, 1H), 6.96 – 6.85 (m, 4H); 6.70 (s, 1H), 6.55 (d, *J* = 7.9 Hz, 1H), 6.48 (s, 1H), 6.42 (s, 1H), 6.35 (d, *J* = 5.3 Hz, 2H), 6.10 (s, 1H), 4.14 – 3.97 (m, 1H), 3.86 (d, *J* = 10.1 Hz, 1H), 3.84 – 3.48 (m, 8H), 3.31 (dd, *J* = 16.6, 11.5 Hz, 1H), 2.98 (d, *J* = 17.7 Hz, 1H), 2.45 – 2.37 (m, 2H), 1.84 (s, 3H), 1.76 (s, 2H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{92}H_{82}N_{16}O_{41}P_8$: 1156.1289 $[M-2H]^{2-}$; found: 1156.1331.

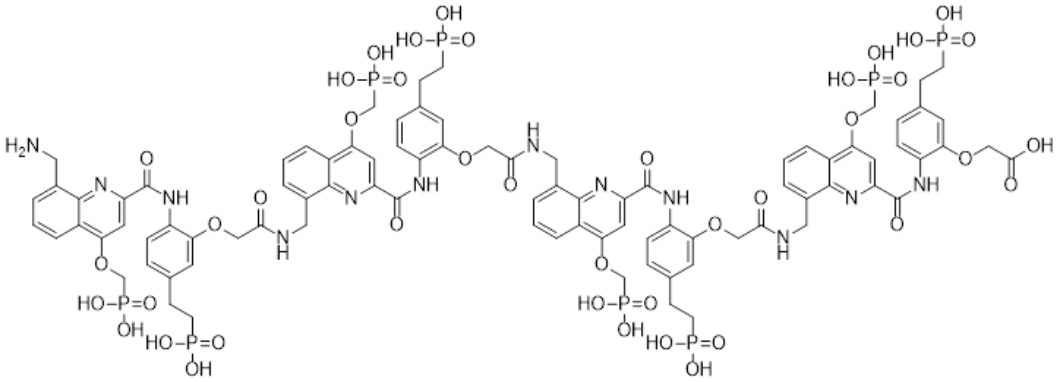


H-MQ^{4}MB^{P}MQ^{4}MB^P-OH (12): Oligomer **12** was synthesized by the automated SPFS.[9]The title compound was obtained as a white solid (11.3 mg, 34%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):**δ (ppm) =11.19 (s, 1H), 11.09 (s, 1H), 10.24 (s, 1H), 9.62 (s, 1H), 9.37 (s, 1H), 9.07 (s, 1H), 8.41 (d, *J* = 9.3 Hz, 1H), 8.10 – 7.03(m, 16H), 6.72 (d, *J* = 8.5 Hz, 1H), 6.50 (d, *J* = 9.7 Hz, 1H), 6.40 (s, 1H), 6.36 (s, 1H), 6.30 (s, 1H), 6.18 (d, *J* = 10.0 Hz, 1H), 5.73 (s, 1H), 4.16 -3.35 (m, 7H), 3.01 - 2.23 (m, 5H), 1.55 – 1.46 (m, 3H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{90}H_{88}N_{14}O_{41}P_8$: 1133.1493 $[M-2H]^{2-}$; found: 1133.1504.

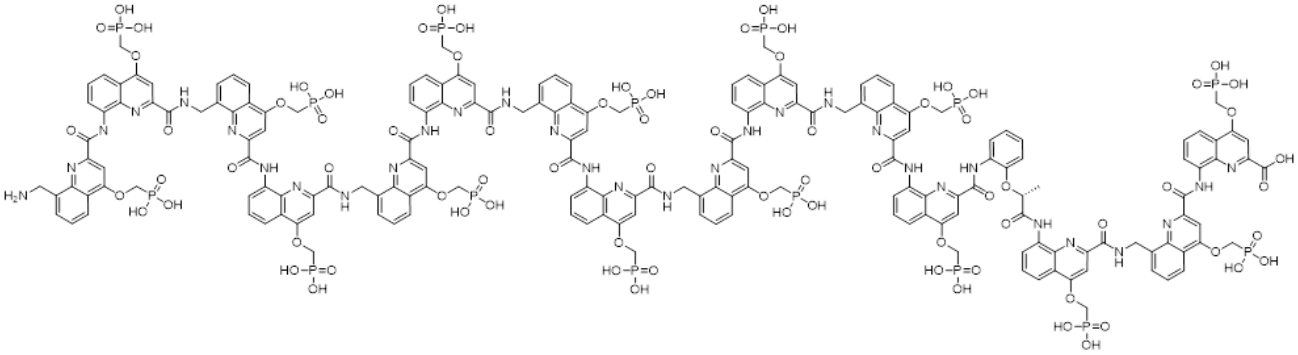


H-MB^{P}MQ^{4}MB^{P}MB^P-OH (13): Oligomer **13** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (14.6 mg, 38%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):** δ (ppm) = 10.71 (s, 1H), 10.27 (s, 1H), 10.20 (s, 1H), 9.93 (s, 1H), 9.58 (s, 1H), 9.45 (s, 1H), 9.37 - 9.28 (m, 3H), 8.95 (t, *J* = 7.4 Hz, 1H), 8.42 (d, *J* = 9.5 Hz, 1H), 8.22 (d, *J* = 9.5 Hz, 2H), 8.04 (s, 1H), 7.84 0 7.69 (m, 5H), 7.51 – 7.23 (m, 17H), 7.17 - 7.13 (m, 5H), 6.99 - 6.80 (m, 2H), 6.72 – 6.48 (m,

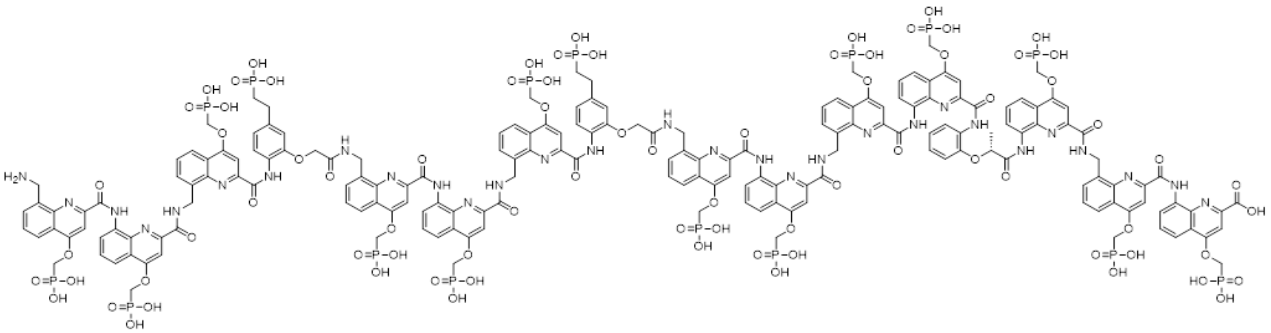
5H), 6.34 - 6.23 (m, 4H), 6.12 (s, 1H), 5.95 (d, *J* = 6.6 Hz, 2H), 5.88 (s, 1H), 5.79 (d, *J* = 7.2 Hz, 2H), 5.71 (s, 1H), 5.47 – 5.41(m, 2H), 3.86 – 3.75 (m, 1H), 3.73 (d, *J* = 8.3 Hz, 2H), 3.72 – 3.65 (m, 3H), 3.65 - 3.41 (m, 6H), 3.03 - 2.91 (m, 3H), 2.30 – 2.08 (m, 6H), 1.99 -1.81 (m, 3H), 1.41 – 1.17 (m, 7H), 1.12 – 0.98 (m, 1H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{89}H_{91}N_{13}O_{41}P_8$: 1121.6595 $[M\text{-}2H]^{2-}$; found: 1121.6540.



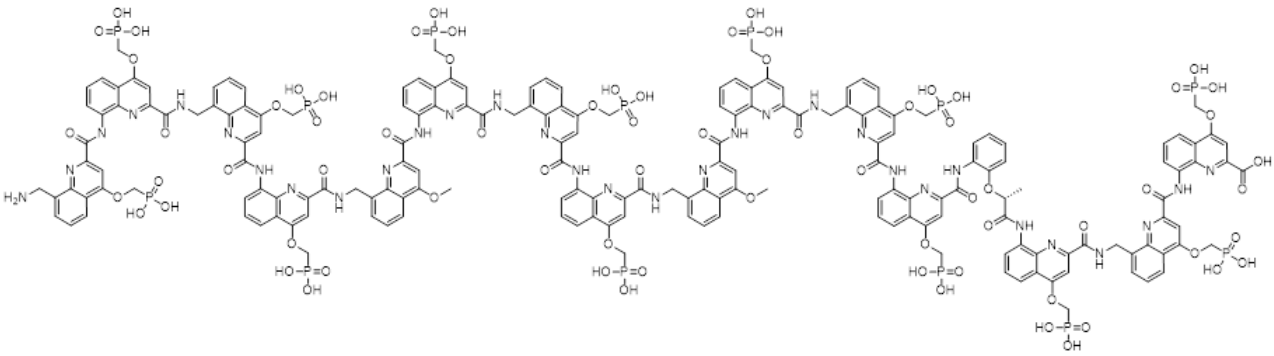
H-MB^{P}MB^{P}MB^{P}MB^P-OH (14): Oligomer **14** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (13.6 mg, 51%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):**δ (ppm) = 10.44 (s, 1H), 10.19 (s, 1H), 10.14 (s, 1H), 8.37 (d, *J* = 9.5 Hz, 1H), 8.26 (t, *J* = 10.0 Hz, 2H), 8.19 (d, *J* = 9.4 Hz, 1H), 7.70 – 7.61 (m, 3H), 7.60 – 7.41 (m, 7H), 7.31 - 7.26 (m, 3H), 7.19 (s, 1H), 7.14 (d, *J* = 9.2 Hz, 2H), 7.07 (t, *J* = 7.5 Hz, 1H), 6.62 - 6.57 (m, 6H), 6.50 - 6.43 (m, 3H), 6.39 (s, 1H), 6.32 – 6.24 (m, 2H), 6.04 - 4.09 (m, 3H), 3.97 (d, *J* = 9.8 Hz, 1H), 3.82 (d, *J* = 9.6 Hz, 3H), 3.55 - 3.51 (m, 5H), 3.05 (d, *J* = 8.6 Hz, 1H), 1.86 (s, 1H), 1.78 (s, 1H), 1.15 - 1.11 (m, 2H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{88}H_{94}N_{12}O_{41}P_8$: 1110.1697 $[M\text{-}2H]^{2-}$; found: 1110.1651.



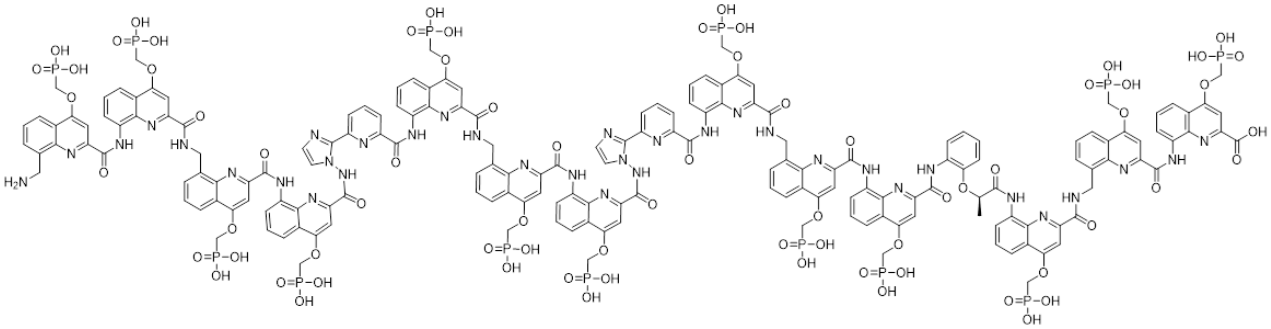
H-MQ^{4}MQ^{4}MQ^{4}MQ^{4}MQ^{4}MQ^{4}B^{R}Q^{4}MQ^4-OH(15): Oligomer **15** was previously reported.[10]



H-MQ^{4}MB^{P}MQ^{4}MB^{P}MQ^{4}MQ^{4}B^{R}Q^{4}MQ^4-OH(16): Oligomer **16** was synthesized by the automated SPFS.[137] The title compound was obtained as a white solid (17.3 mg, 46%).**^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):**δ (ppm)= 11.17 (s, 1H), 11.10 (s, 1H), 10.28 (s, 1H), 10.21 (d, *J* = 10.3 Hz, 2H), 9.69 (s, 1H), 9.56 (d, *J* = 7.7 Hz, 1H), 9.51 (s, 1H), 9.22 (s, 1H), 9.12 (s, 1H), 8.94 (s, 1H), 8.78 (s, 1H), 8.66 (s, 1H), 8.36 (d, *J* = 8.8 Hz, 1H), 8.01 -7.96 (m, 3H), 7.81 -7.75 (m, 4H), 7.71 – 7.54 (m, 6H), 7.50 – 7.27 (m, 8H), 7.26 – 6.84 (m, 17H), 6.84 – 6.61 (m, 10H), 6.53 (t, *J* = 9.0 Hz, 1H), 6.39 (s, 1H), 6.31 (s, 1H), 6.26 (d, *J* = 8.8 Hz, 1H), 6.03 – 5.91 (m, 3H), 5.84 (d, *J* = 9.3 Hz, 1H), 5.78 (s, 1H), 5.69 (d, *J* = 9.1 Hz, 1H), 4.20 – 3.46 (m, 14H), 3.23 (t, *J* = 10.7 Hz, 1H), 3.09 - 2.83 (m, 3H), 2.62 (s, 1H), 2.47 – 2.27 (m, 3H), 2.16 – 1.80 (m, 3H), 1.34 – 1.07 (m, 4H), -0.01 (d, *J* = 6.5 Hz, 3H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{179}H_{166}N_{29}O_{78}P_{15}$: 1476.8587 $[M-3H]^{3-}$; found: 1476.8543.

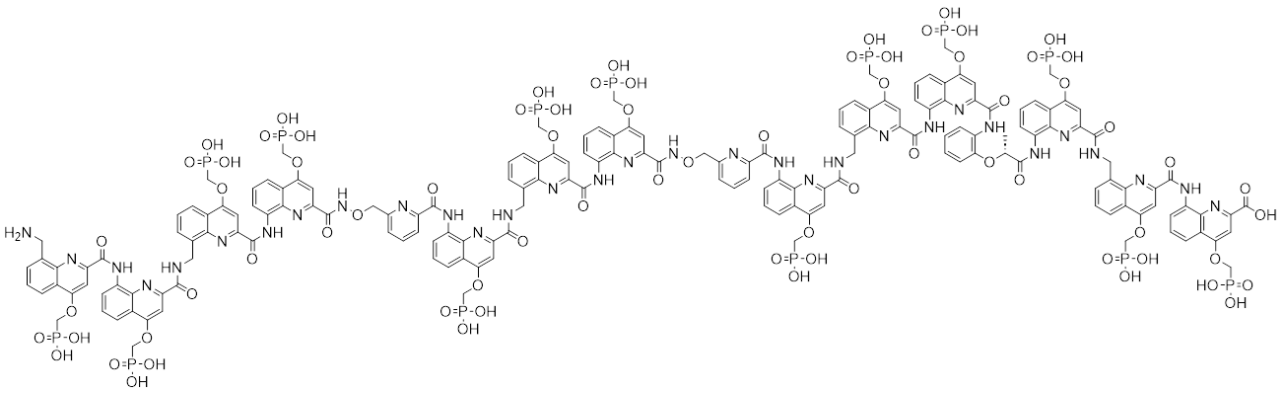


H-MQ^{4}MQ^{4}MQ^{4}MQ^{4}MQ^{4}MQ^{4}B^{R}Q^{4}MQ^4-OH(17): Oligomer **17** was previously reported.[10]

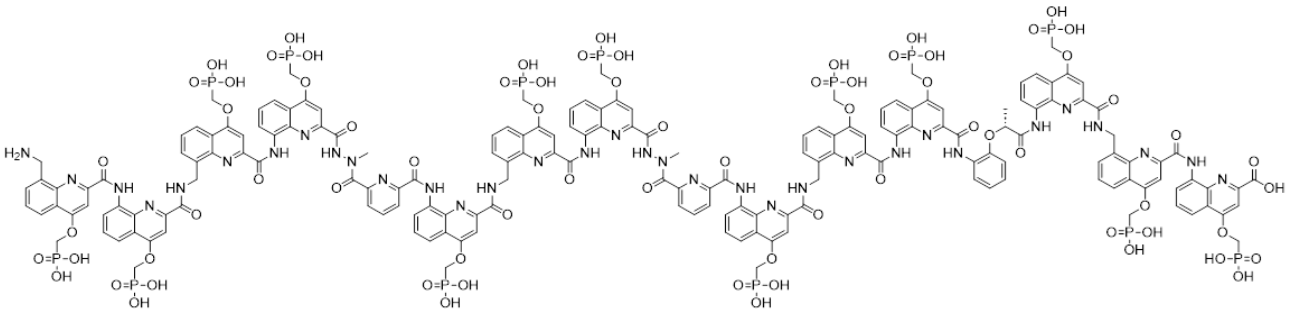


H-MQ^{4}MQ^{4}NQ^{4}MQ^{4}NQ^{4}MQ^{4}B^{R}Q^{4}MQ^4-OH (18): Oligomer **19** was synthesized by the automated

SPFS.[9] The title compound was obtained as a white solid (16.2 mg, 37%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1):** δ (ppm) =12.43 (s, 1H), δ 11.77 (s, 1H), 11.19 (s, 1H), 11.10 (s, 1H), 10.18 (s, 1H), 10.10 (s, 1H), 9.83 (s, 1H), 9.67 (s, 1H), 9.48 (s, 1H), 9.41 (d, *J* = 8.0 Hz, 2H), 9.26 (s, 1H), 9.22 (s, 1H), 8.90 (d, *J* = 12.6 Hz, 2H), 8.34 (d, *J* = 8.8 Hz, 1H), 8.07 – 7.97 (m, 3H), 7.88 (d, *J* = 8.7 Hz, 1H), 7.80 (d, *J* = 4.0 Hz, 2H), 7.75 (s, 1H), 7.64 (t, *J* = 6.4 Hz, 1H), 7.56 (d, *J* = 6.6 Hz, 1H), 7.51 – 7.45 (m, 2H), 7.44 – 7.34 (m, 2H), 7.36 – 7.17 (m, 11H), 7.14 (d, *J* = 8.4 Hz, 2H), 7.14 – 7.00 (m, 2H), 7.00 – 6.68 (m, 12H), 6.65 (d, *J* = 8.4 Hz, 2H), 6.63 – 6.38 (m, 11H), 6.34 - 6.25 (d, *J* = 3.6 Hz, 3H), 6.22 – 6.11 (m, 2H), 6.09 (s, 1H), 4.02 – 3.84 (m, 5H), 3.80 - 3.71 (d, *J* = 9.8 Hz, 5H), 3.60 – 3.52 (m, 2H), 3.41 – 3.06 (m, 3H), 3.01 – 1.15 (m, 3H), 0.32 (d, *J* = 6.6 Hz, 3H). **HRMS** (ESI^-) *m/z* calcd. for $C_{175}H_{150}N_{35}O_{70}P_{13}$: 1064.8888 $[M-4H]^{4-}$; found: 1064.8857.



H-MQ^{4}MQ^{4}OQ^{4}MQ^{4}OQ^{4}MQ^{4}B^{R}Q^{4}MQ^4-OH (19): Oligomer **19** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (15 mg, 42%). **^{1}H NMR (500 MHz, H_2O/D_2O 9:1):**δ (ppm)= 11.35 (s, 1H), 11.19 (s, 1H), 10.34 (s, 1H), 10.04 (d, *J* = 8.0 Hz, 1H), 9.99 (s, 1H), 9.95 (s, 1H), 9.89 (s, 1H), 9.63 (s, 1H), 9.50 (d, *J* = 9.3 Hz, 2H), 9.44 (s, 1H), 9.27 (s, 1H), 9.21 (d, *J* = 7.9 Hz, 1H), 9.01 (s, 1H), 8.92 – 8.87 (m, 1H), 8.46 (dd, *J* = 8.9, 2.8 Hz, 2H), 8.40 (d, *J* = 8.9 Hz, 1H), 8.08 -8.03 (m, 2H), 7.87 (d, *J* = 7.9 Hz, 1H), 7.71 – 7.63 (m, 2H), 7.63 – 7.58 (m, 2H), 7.61 – 7.57 (m, 1H), 7.54 (d, *J* = 6.9 Hz, 1H), 7.51 – 7.44 (m, 5H), 7.43 – 7.36 (m, 4H), 7.33 - 7.26 (m, 10H), 7.15 – 6.99 (m, 5H), 6.99 - 6.98 (s, 4H), 6.98 – 6.74 (m, 14H), 6.76 – 6.67 (m, 2H), 6.70 (s, 1H), 6.68 – 6.60 (m, 3H), 6.31- 6.14 (m, 4H), 5.98 (s, 1H), 5.73 (d, *J* = 7.2 Hz, 2H), 4.16 (d, *J* = 10.1 Hz, 2H), 4.14 – 4.05 (m, 3H), 4.05 – 3.83 (m, 5H), 3.80 (d, *J* = 9.3 Hz, 2H), 3.78 – 3.69 (m, 1H), 3.68 (s, 3H), 3.67 – 3.61 (m, 1H), 3.64 – 3.51 (m, 2H), 3.44 - 2.96 (m, 4H), 2.75 – 2.67 (m, 1H), 1.69 (t, *J* = 13.5 Hz, 2H), 1.34 – 1.23 (m, 2H), -0.01 (d, *J* = 6.7 Hz, 3H). **HRMS** (ESI^-) *m/z* calcd. for $C_{171}H_{150}N_{31}O_{72}P_{13}$: 1046.8832 $[M-4H]^{4-}$; found: 1046.8864.



H-MQ^{4}MQ^{4}HQ^{4}MQ^{4}HQ^{4}MQ^{4}B^{R}Q^{4}MQ^4-OH (20): Oligomer **20** was synthesized by the automated SPFS.[9] The title compound was obtained as a white solid (14.8 mg, 35%). **^{1}H NMR (500 MHz, H$_2$O/D$_2$O 9:1):** δ (ppm) = 11.31 (s, 1H), 11.01 (s, 1H), 10.55 (s, 1H), 10.30 (s, 1H), 10.12 (s, 1H), 9.92 (s, 1H), 9.64 (d, *J* = 7.8 Hz, 1H), 9.53 (s, 1H), 9.49 (s, 1H), 9.34 (d, *J* = 7.3 Hz, 1H), 9.22 (s, 1H), 9.14 (s, 1H), 9.06 (s, 1H), 8.93 (s, 1H), 8.60 – 8.37 (m, 4H), 8.21 (d, *J* = 8.6 Hz, 1H), 8.09 – 7.94 (m, 3H), 7.92 – 7.58 (m, 15H), 7.57 – 7.31 (m, 12H), 6.87 – 6.63 (m, 8H), 6.62 – 6.51 (m, 2H), 6.44 (d, *J* = 8.7 Hz, 1H), 6.41 (d, *J* = 3.4 Hz, 1H), 6.22 (s, 1H), 6.14 (s, 1H), 5.94 (s, 1H), 5.83 (s, 1H), 3.07 (s, 1H), 2.91 (s, 3H), 2.65 – 2.58 (m, 3H), 0.01 (d, *J* = 6.7 Hz, 3H). **HRMS** (ESI$^-$) *m/z* calcd. for $C_{173}H_{152}N_{33}O_{72}P_{13}$: 1060.3886 $[M\text{-}4H]^{4-}$; found: 1060.3857.

4. NMR spectra

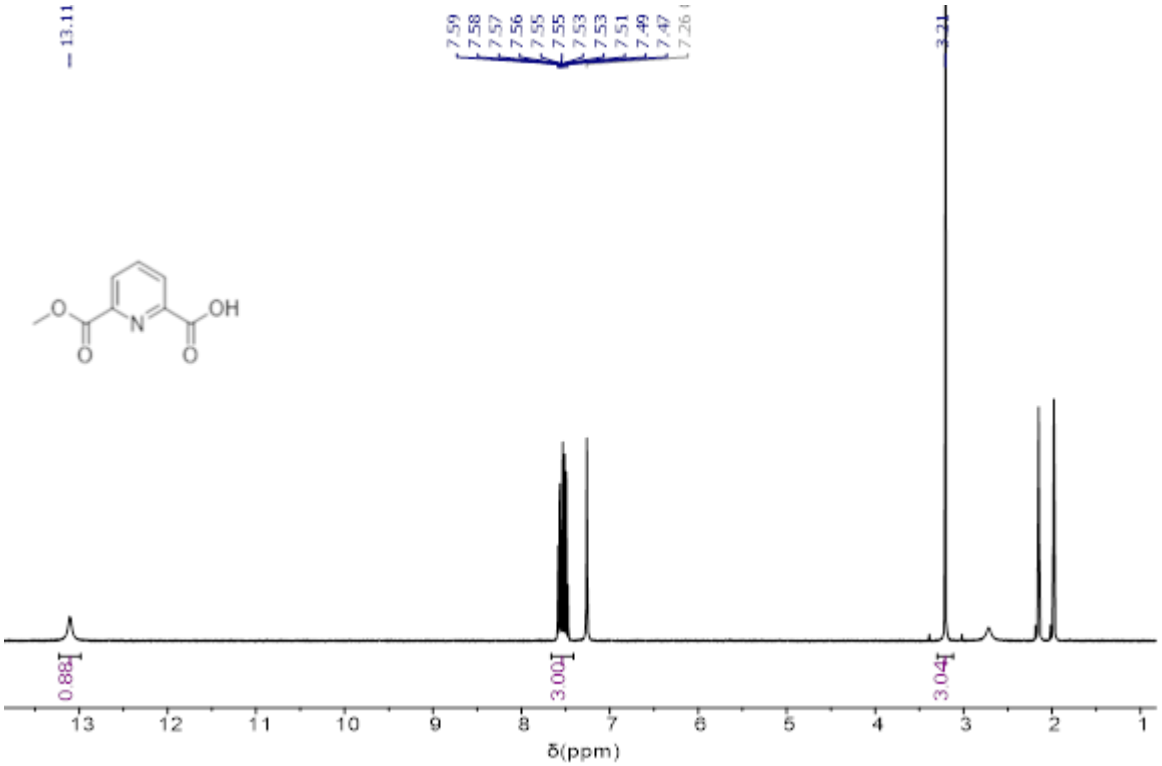


Figure S16. ^{1}H NMR spectrum of compound **1a**. (500 MHz, $CDCl_3$). According to ref. 6 at the start.

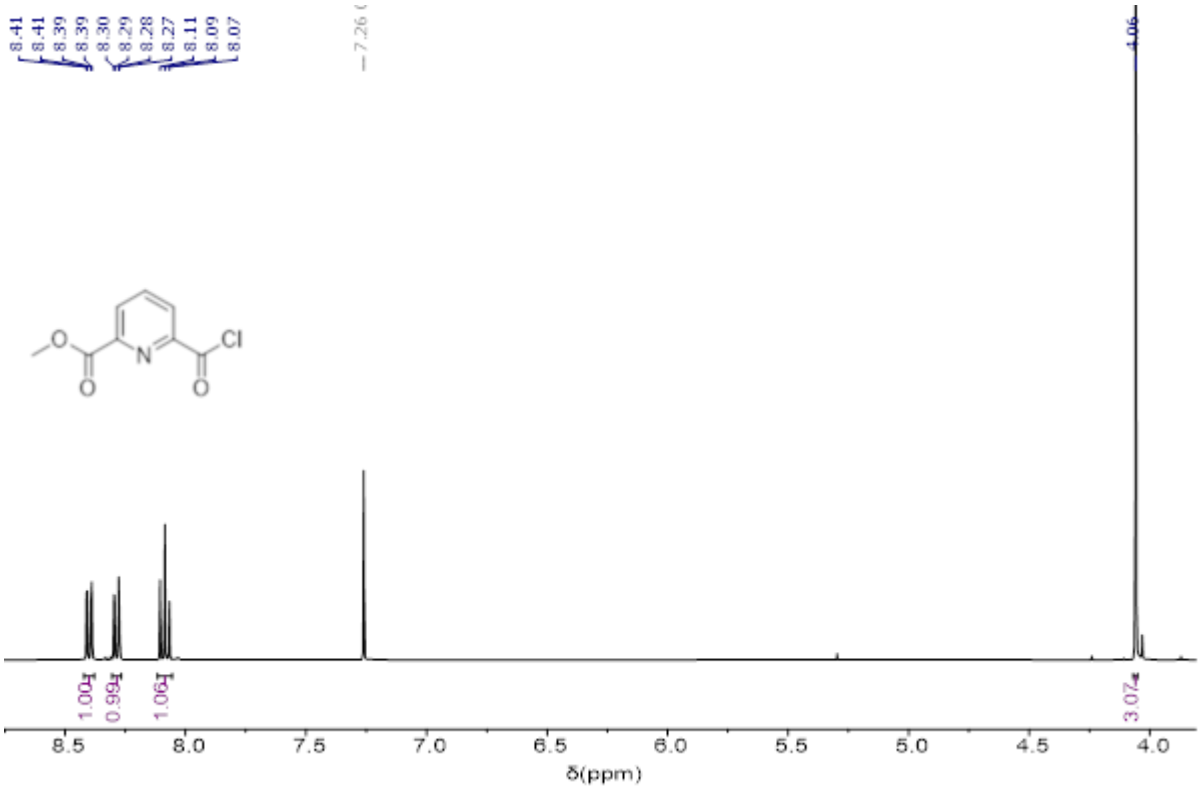


Figure S17. ^{1}H NMR spectrum of compound **1b.** (500 MHz, $CDCl_3$). According to ref. 7 at the start.

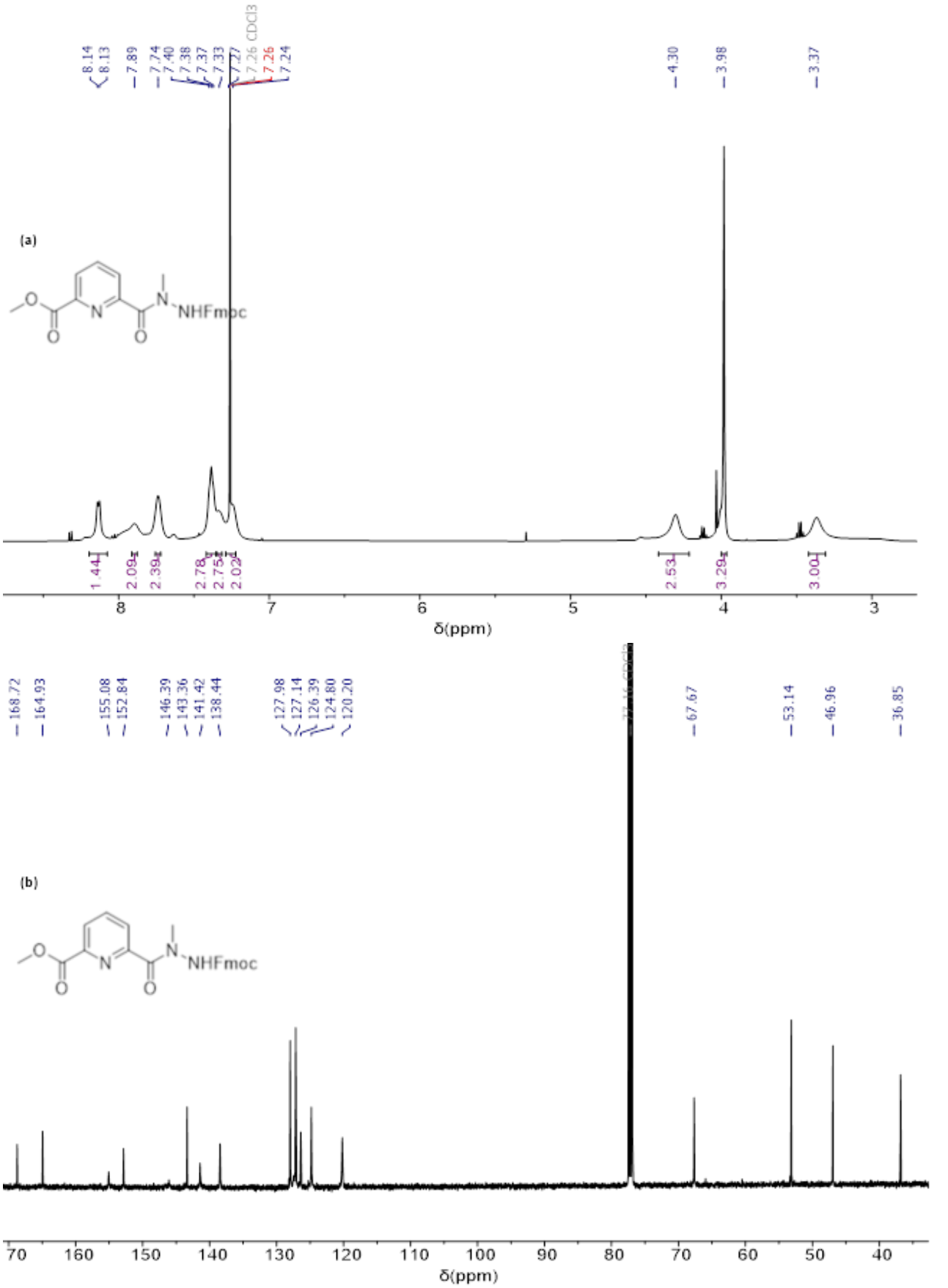


Figure S18. NMR spectra of compound **1c**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

Figure S19. NMR spectra of compound **1d**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

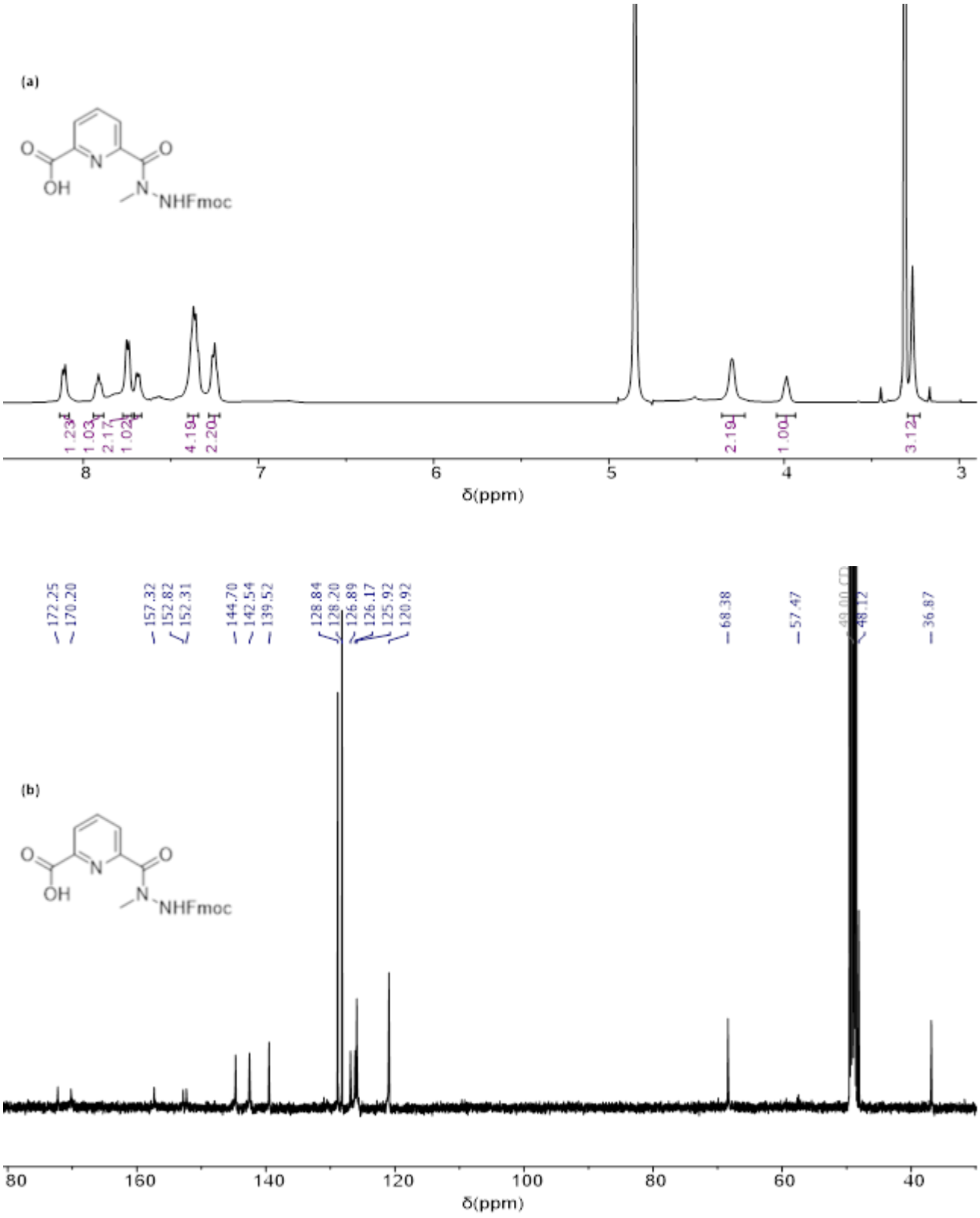


Figure S20. NMR spectra of compound **1**. (a) ^{1}H NMR (500 MHz, MeOD). (b) ^{13}C NMR (126 MHz, MeOD).

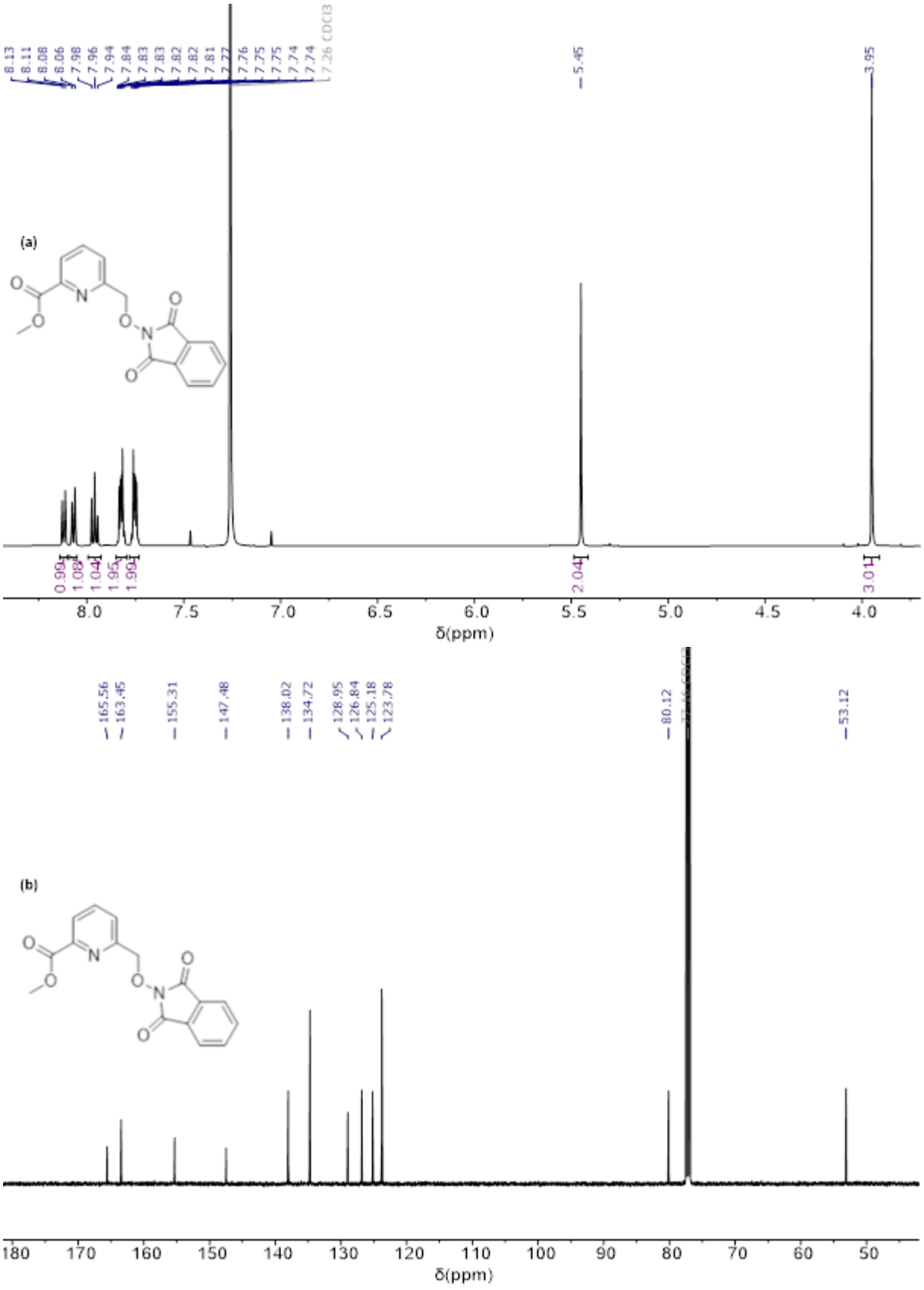


Figure S21. NMR spectra of compound **2a**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

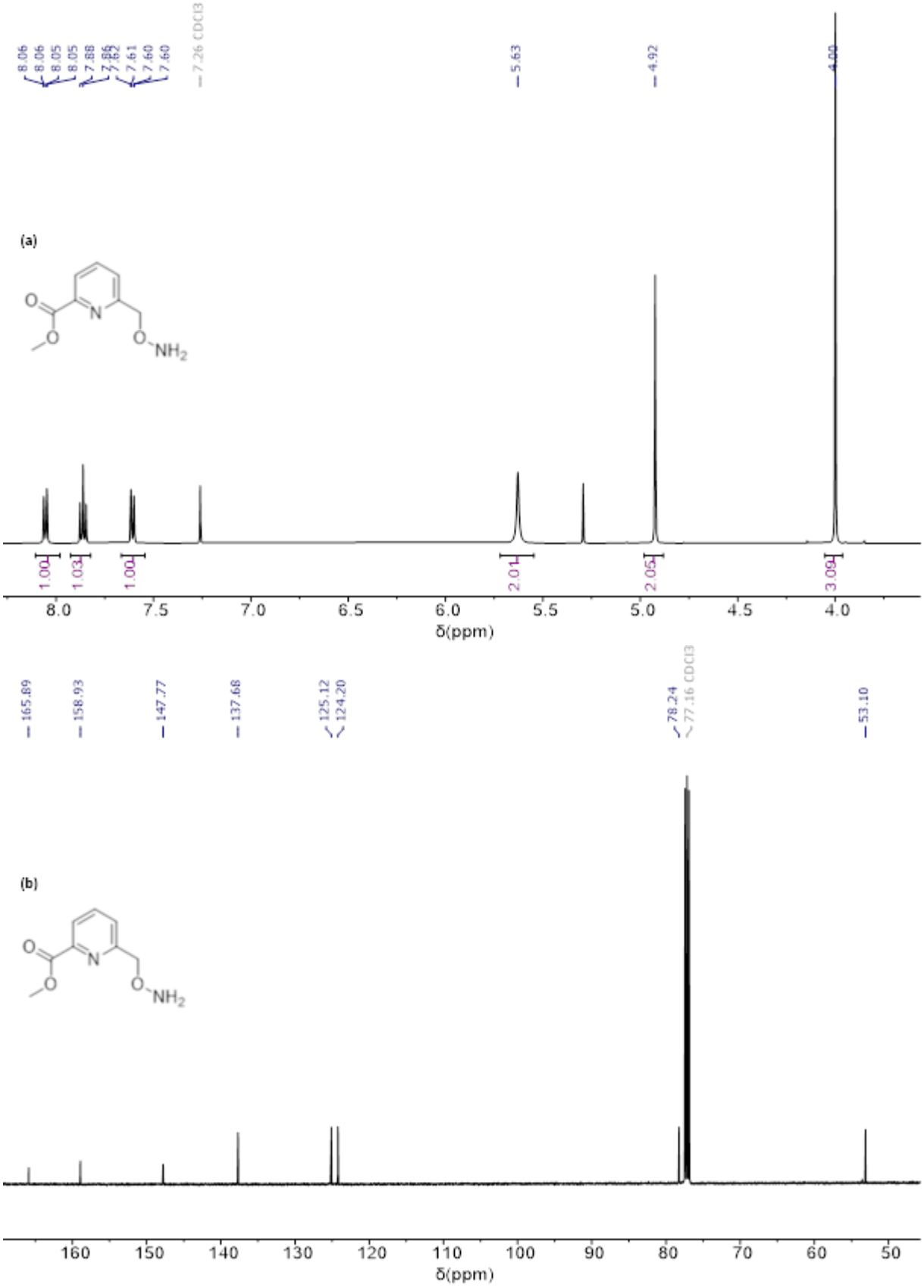


Figure S22. NMR spectra of compound **2b**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

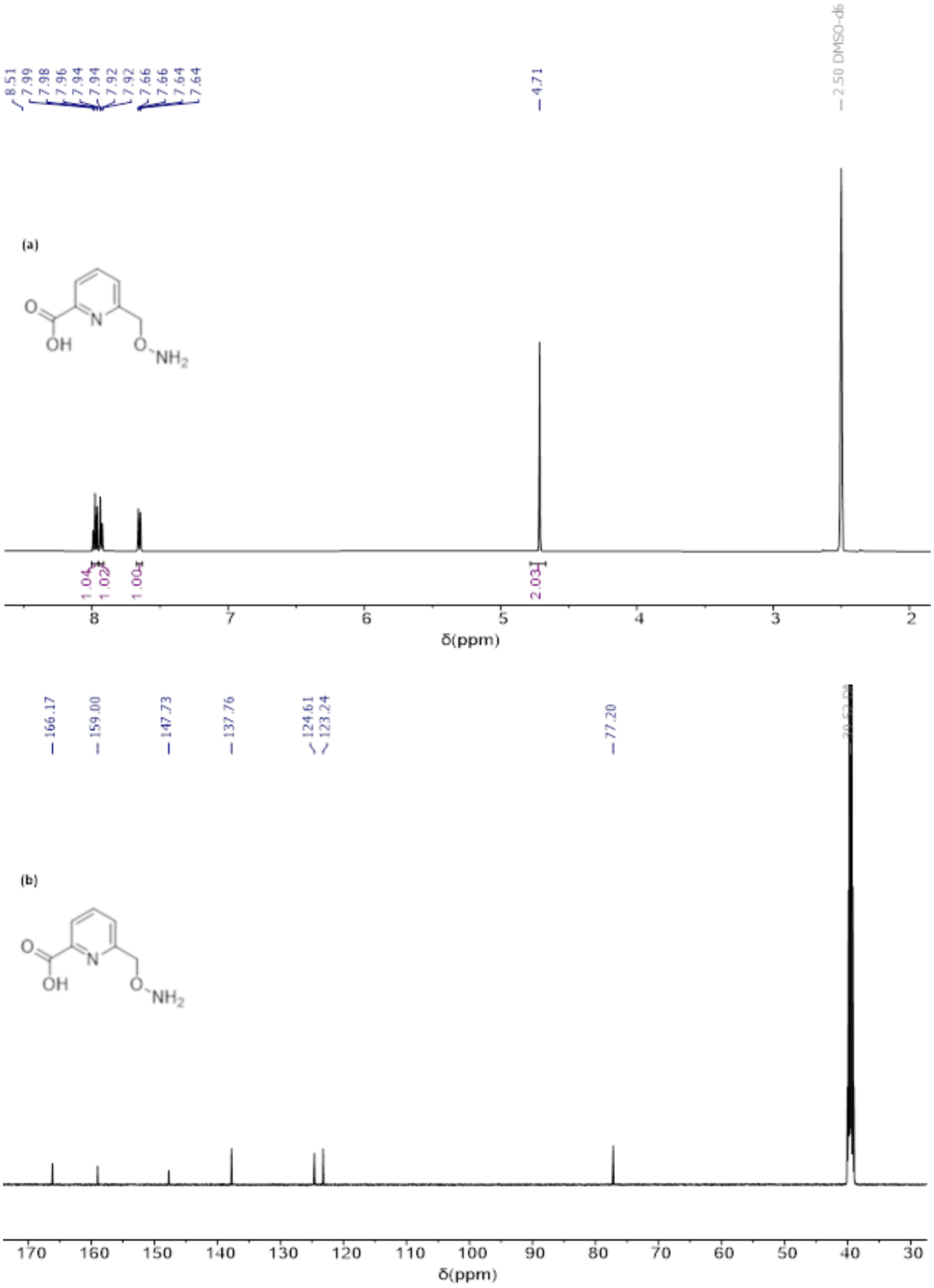


Figure S23. NMR spectra of compound **2c**. (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

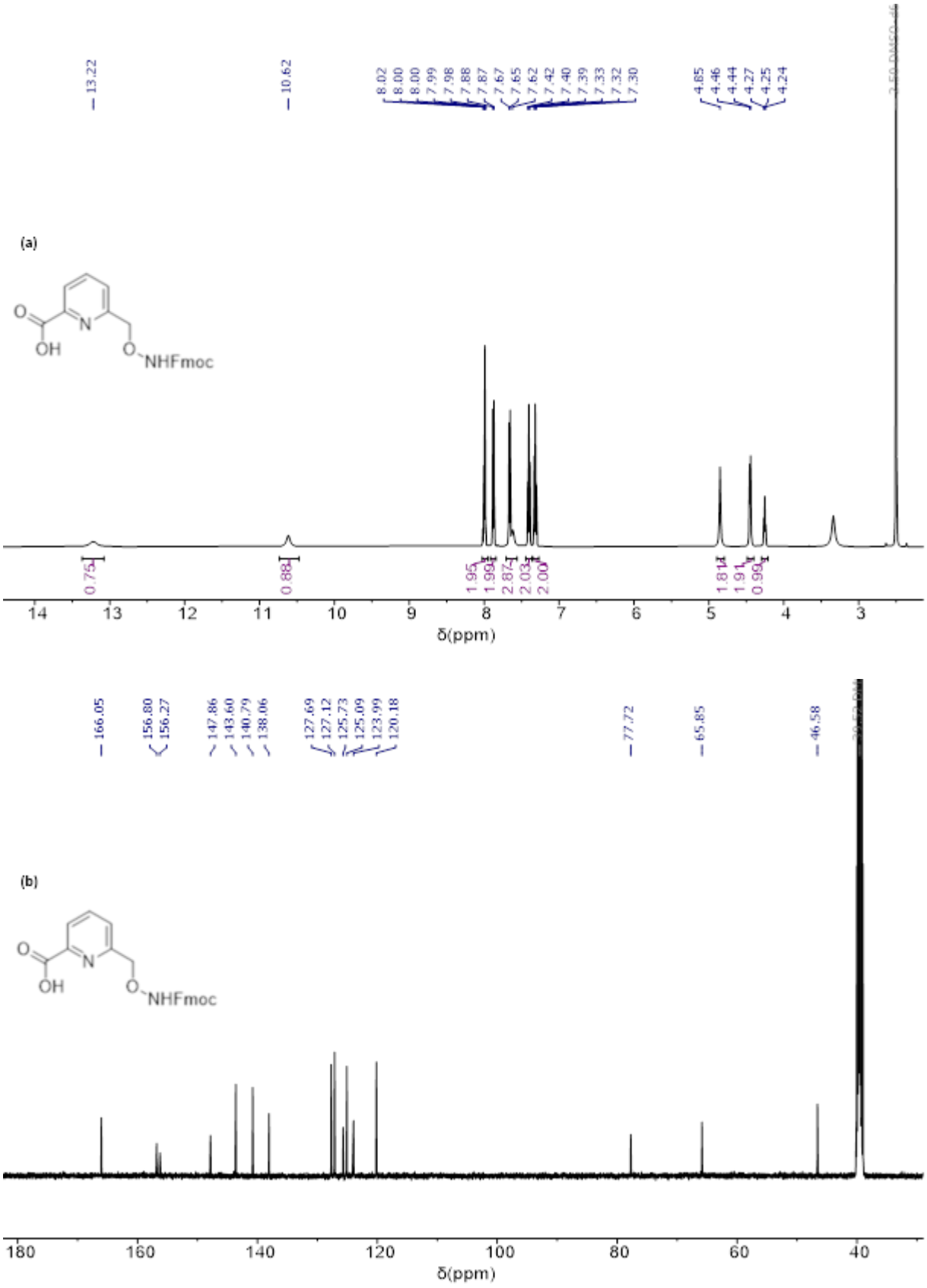


Figure S24. NMR spectra of compound **2**. (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

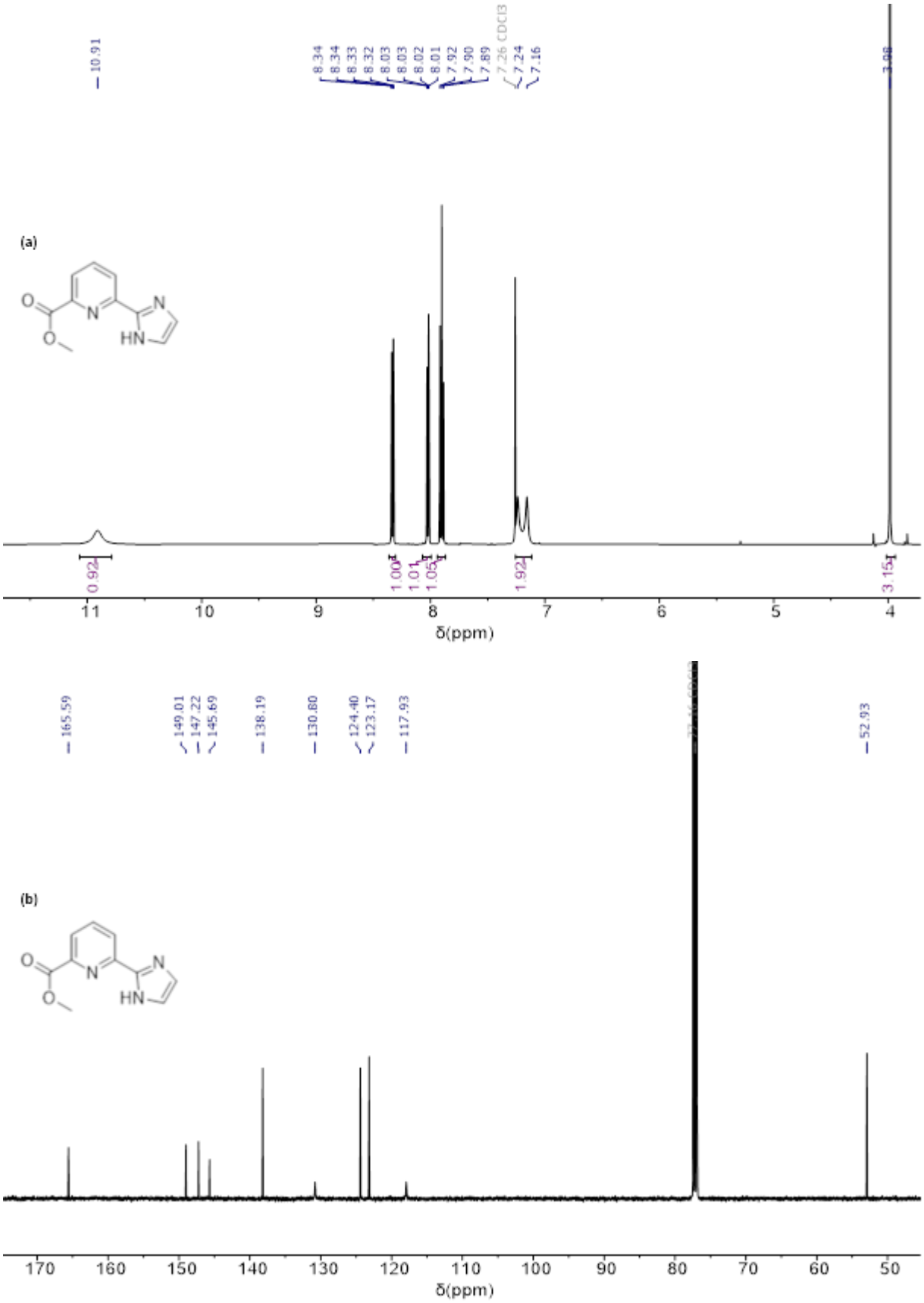


Figure S25. NMR spectra of compound **3a**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

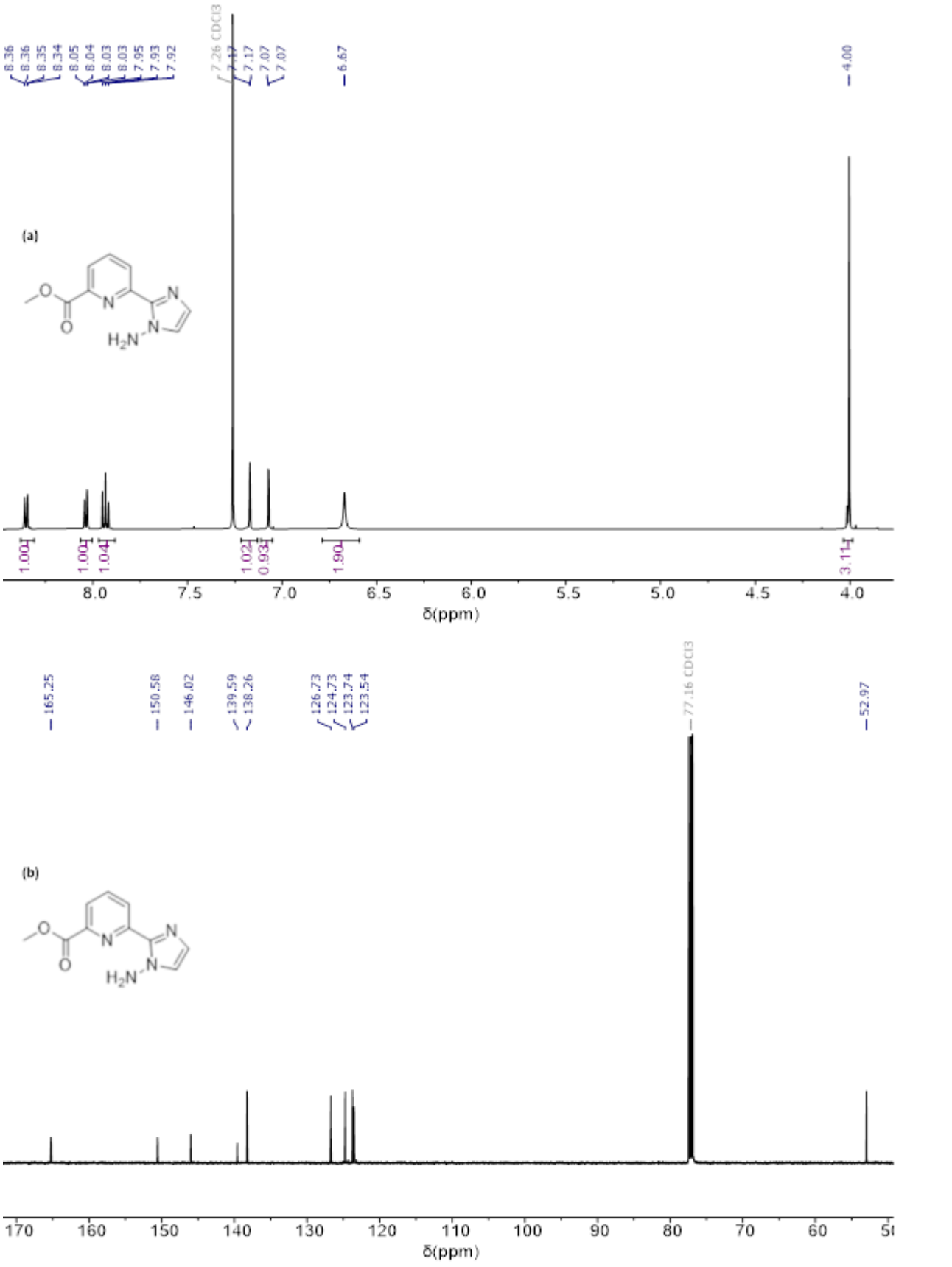


Figure S26. NMR spectra of compound **3b**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

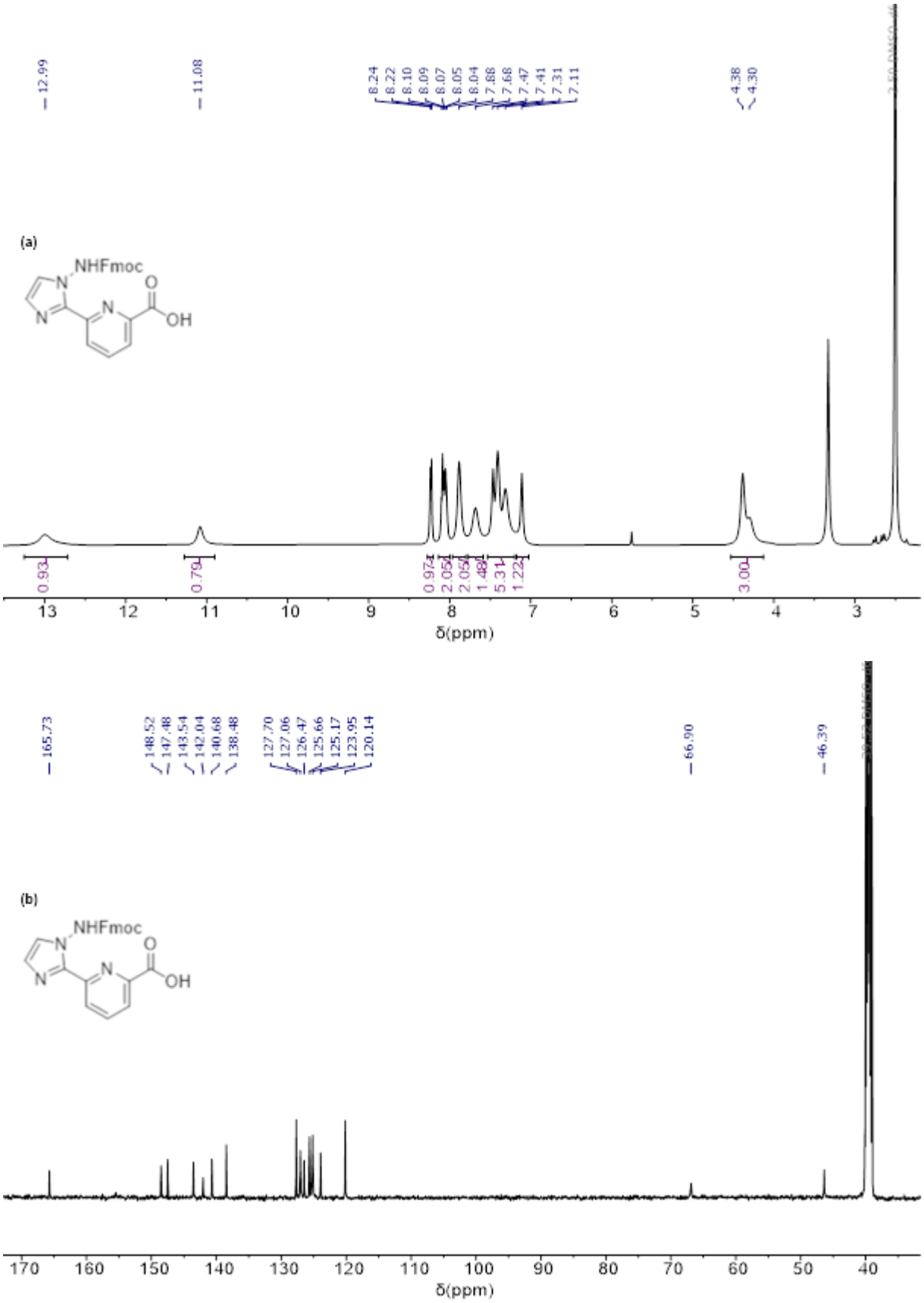


Figure S27. NMR spectra of compound **3.** (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

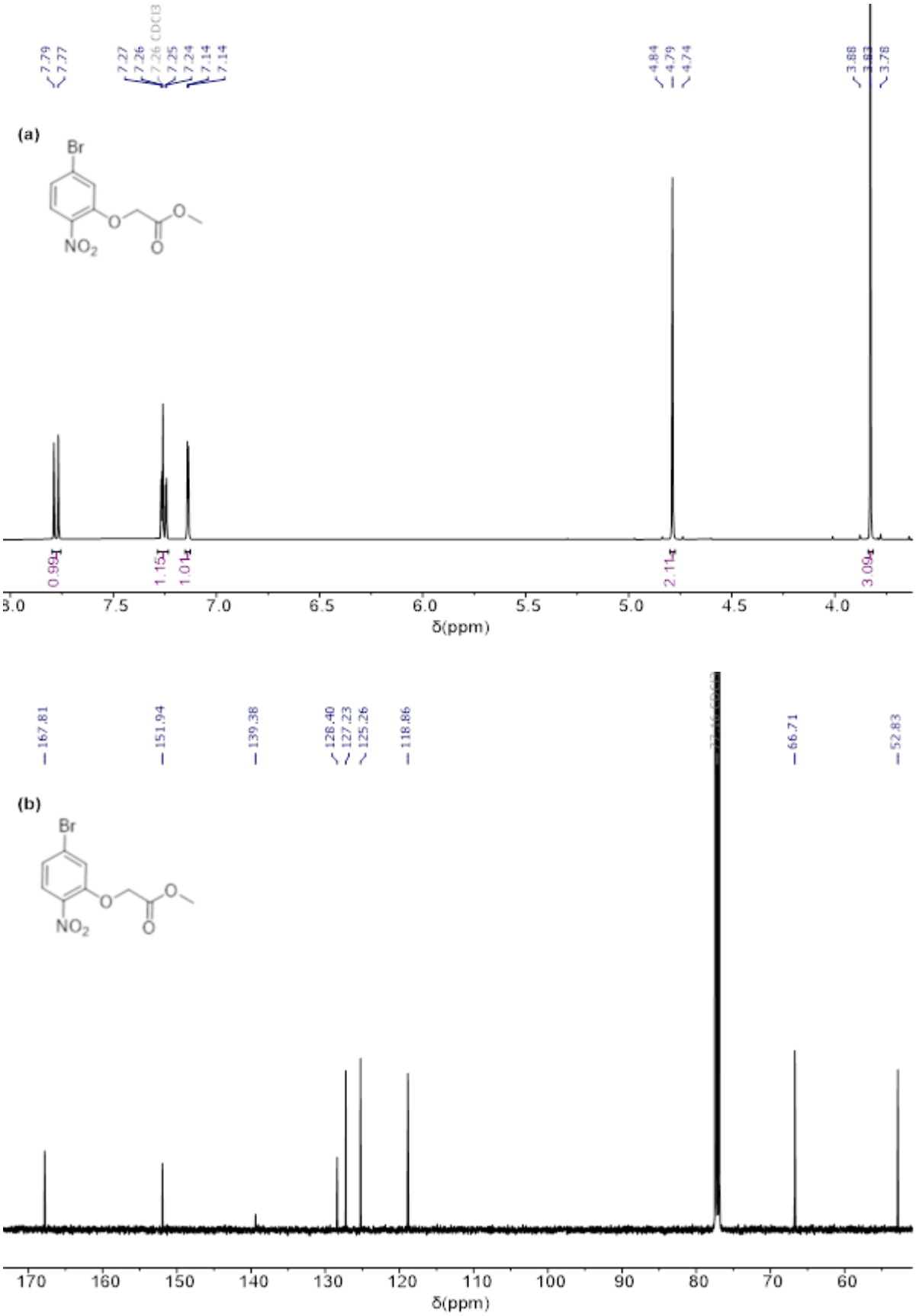


Figure S28. NMR spectra of compound **4a**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

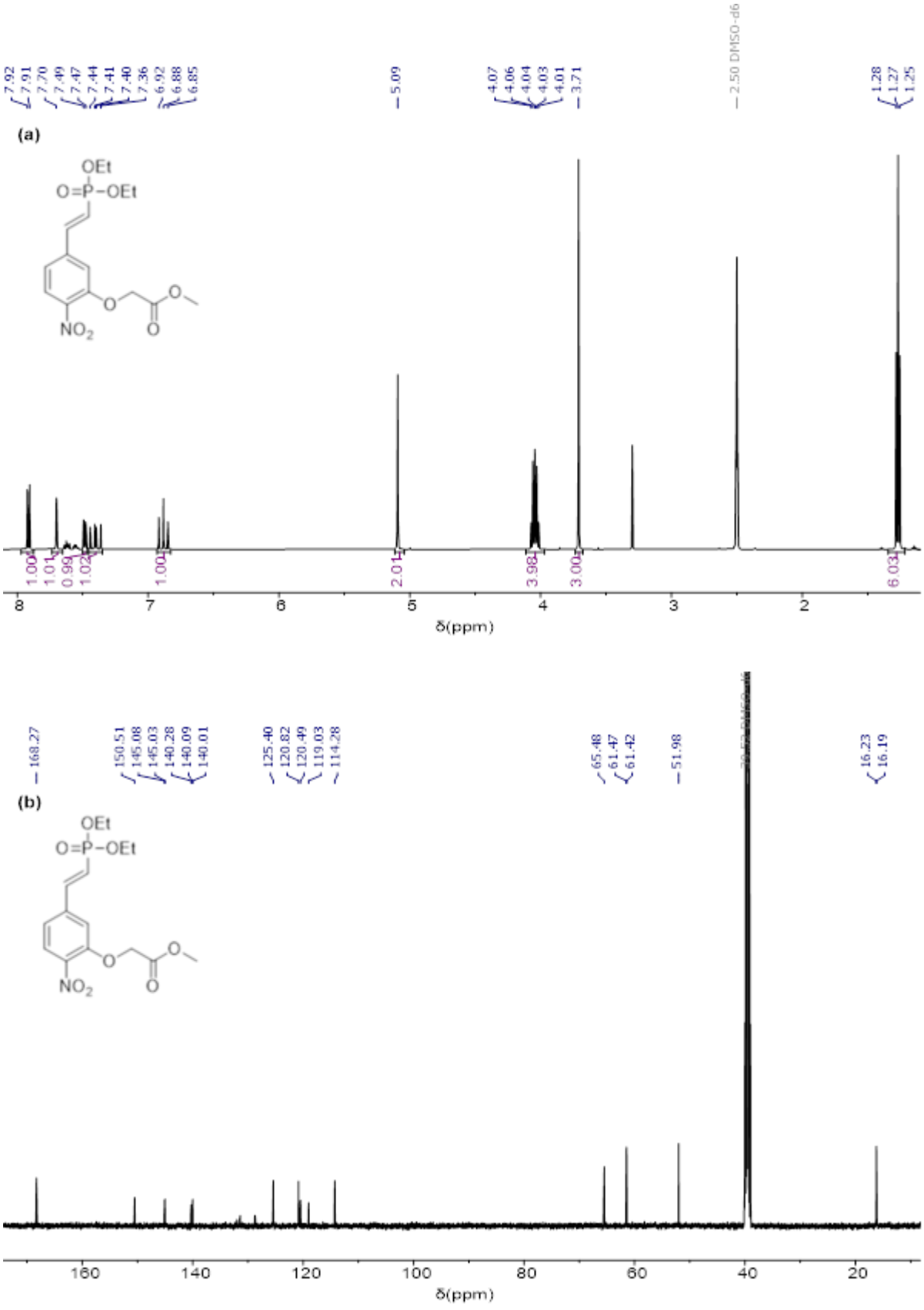


Figure S29. NMR spectra of compound **4b**. (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

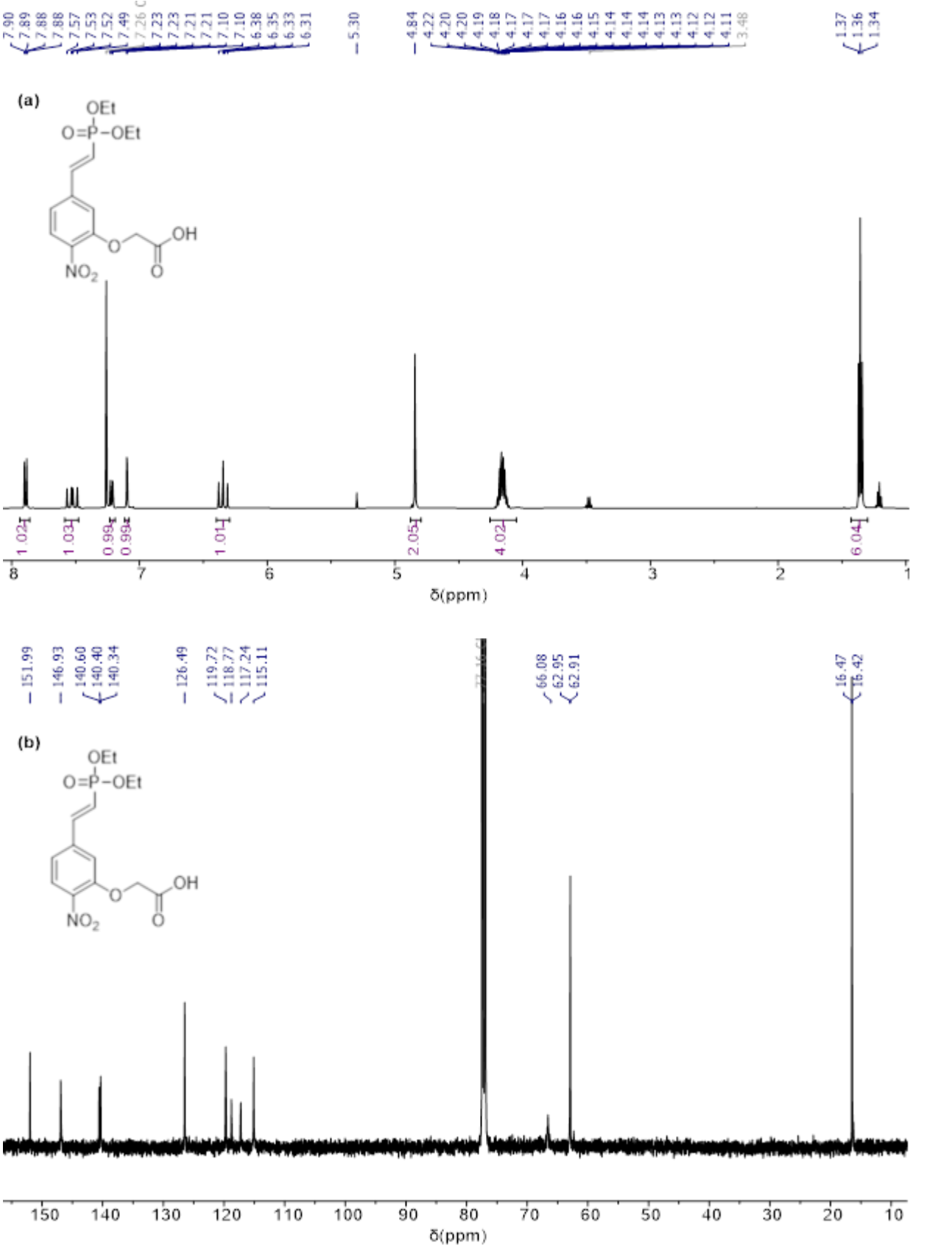


Figure S30. NMR spectrum of Compound **4c**. (a) ^{1}H NMR (500 MHz, $CDCl_3$). (b) ^{13}C NMR (126 MHz, $CDCl_3$).

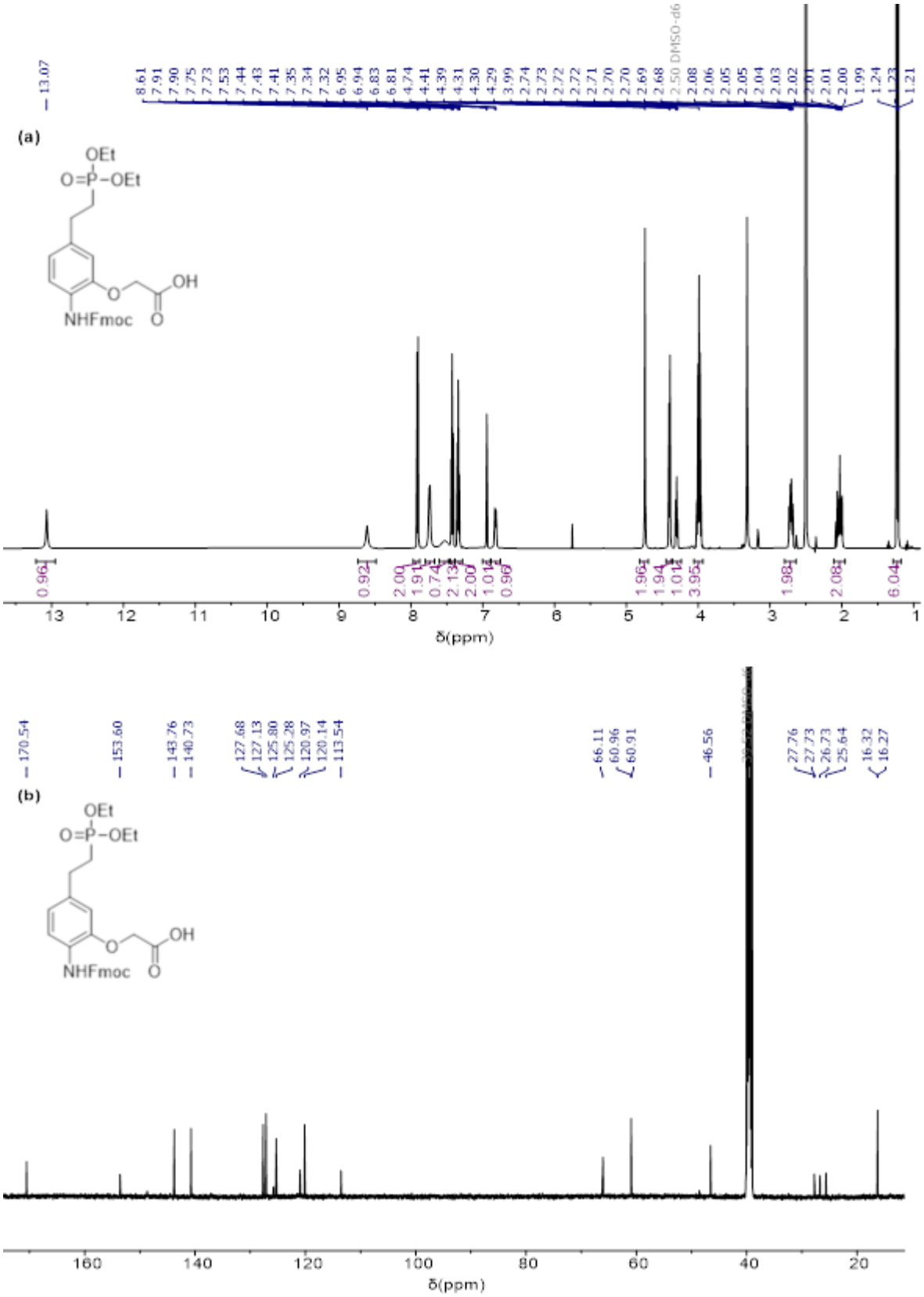


Figure S31. NMR spectra of Compound **4**. (a) ^{1}H NMR (500 MHz, DMSO-d_6). (b) ^{13}C NMR (126 MHz, DMSO-d_6).

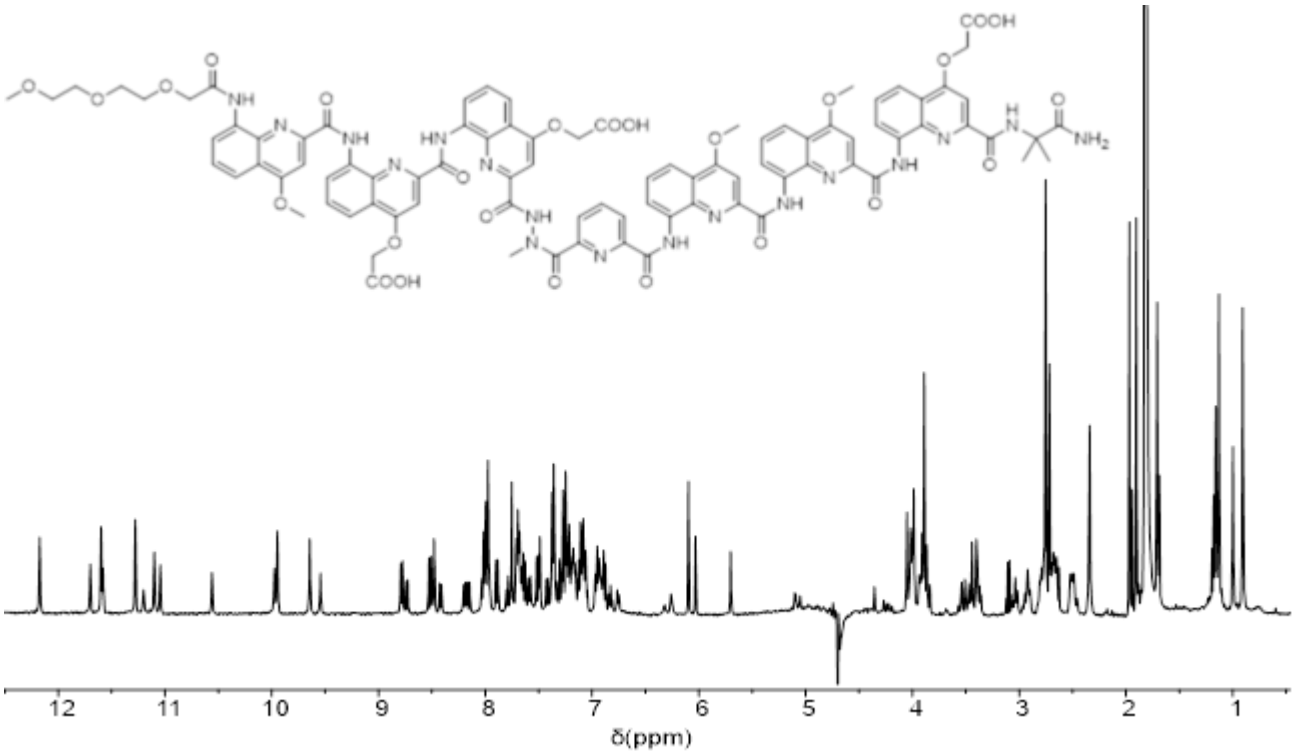


Figure S32. ^{1}H NMR spectrum of oligomer **5**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

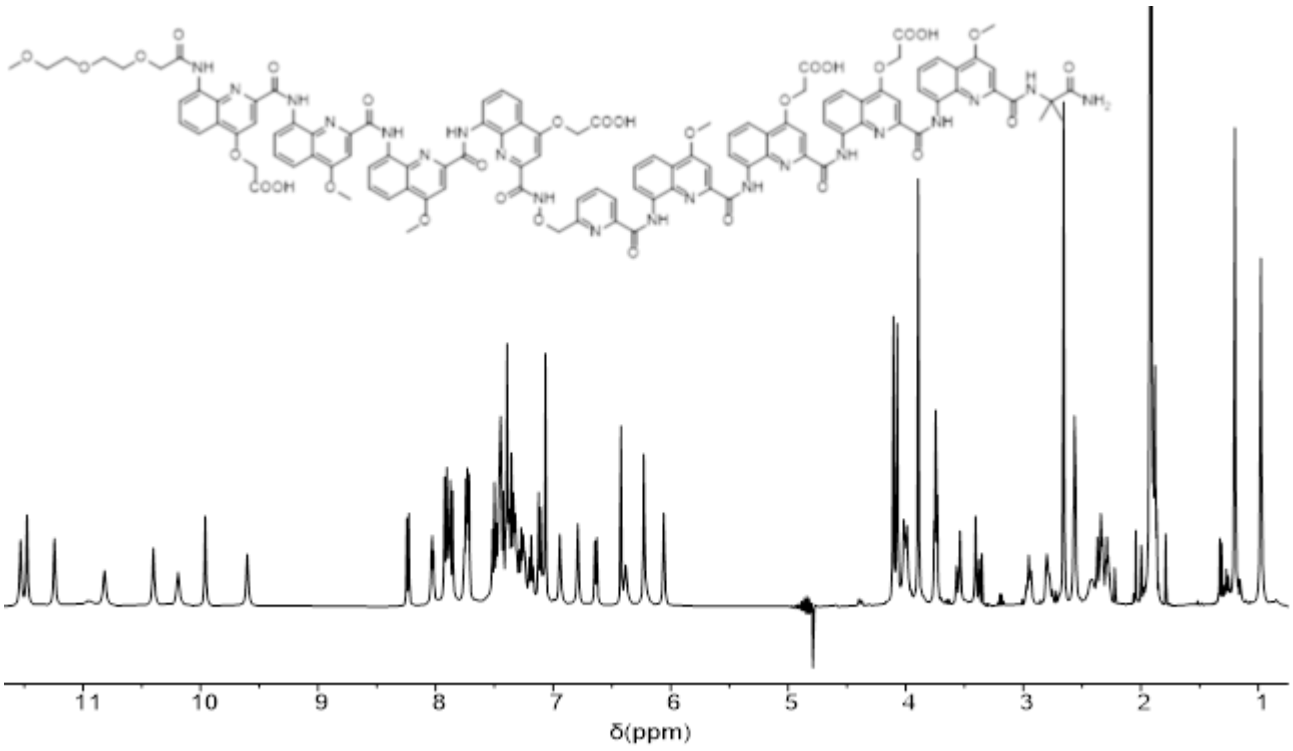


Figure S33. ^{1}H NMR spectra of oligomer **6**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

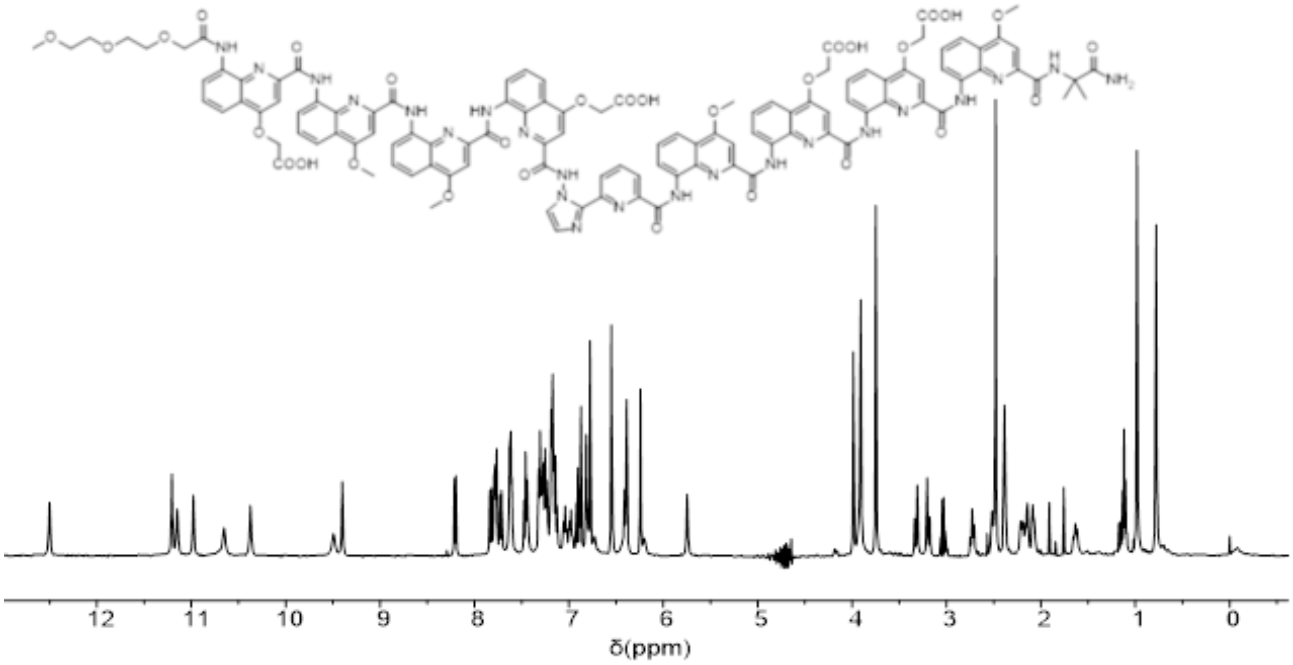


Figure S34. ^{1}H NMR spectra of oligomer **7**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

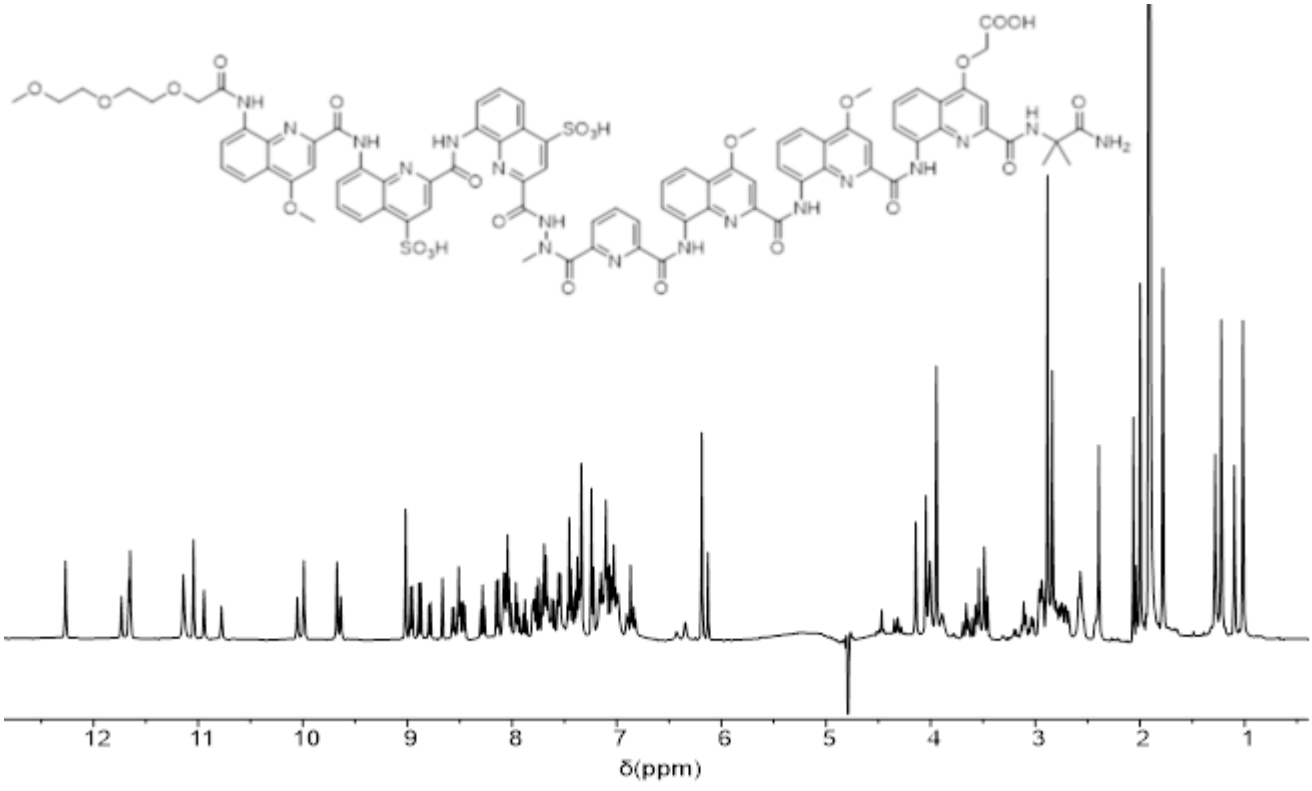


Figure S35. ^{1}H NMR spectrum of **oligomer 8**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

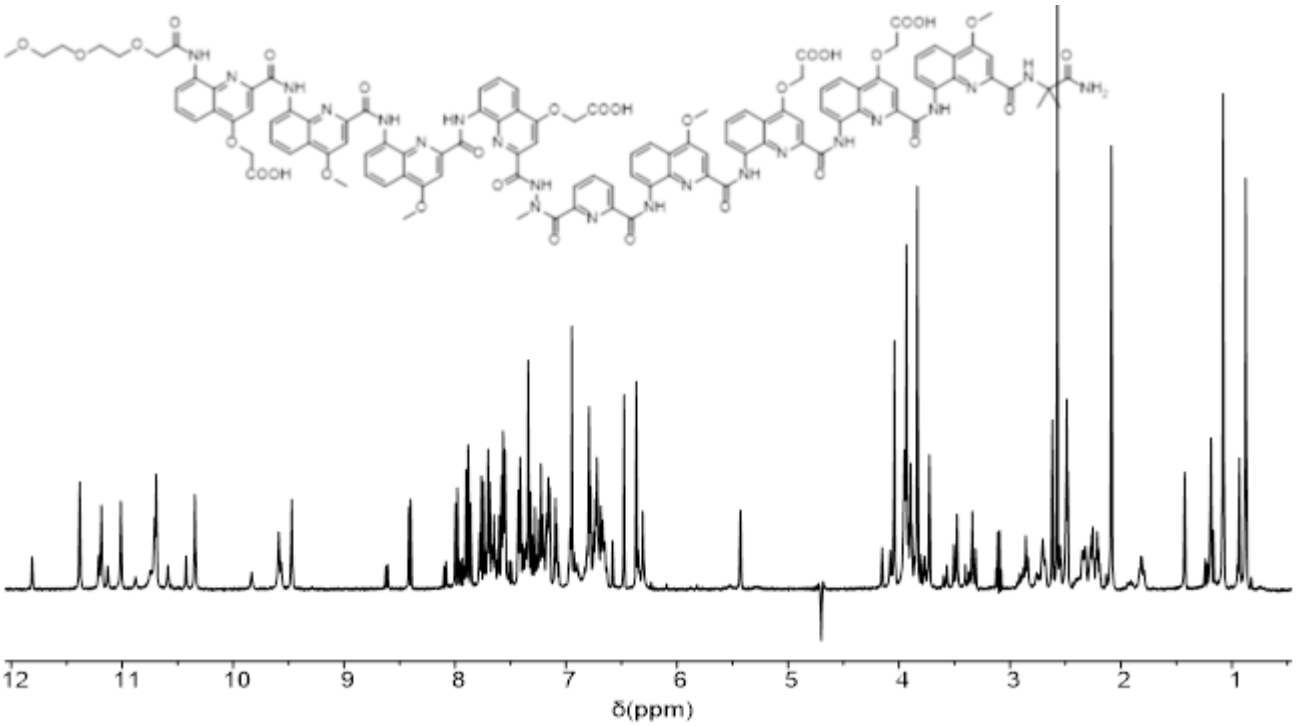


Figure S36. ^{1}H NMR spectra of oligomer **9**. (a) Part of the ^{1}H DOSY NMR. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

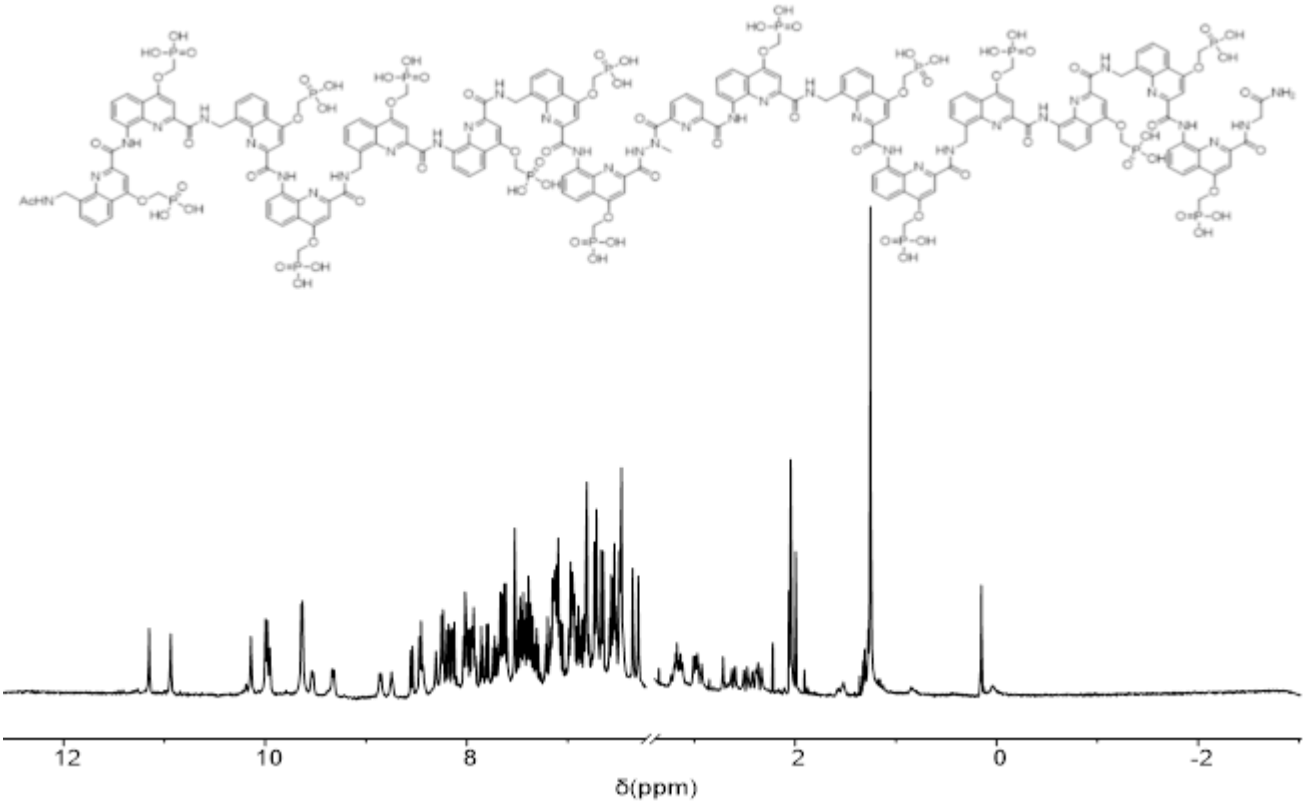


Figure S37. ^{1}H NMR spectra of oligomer **10**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

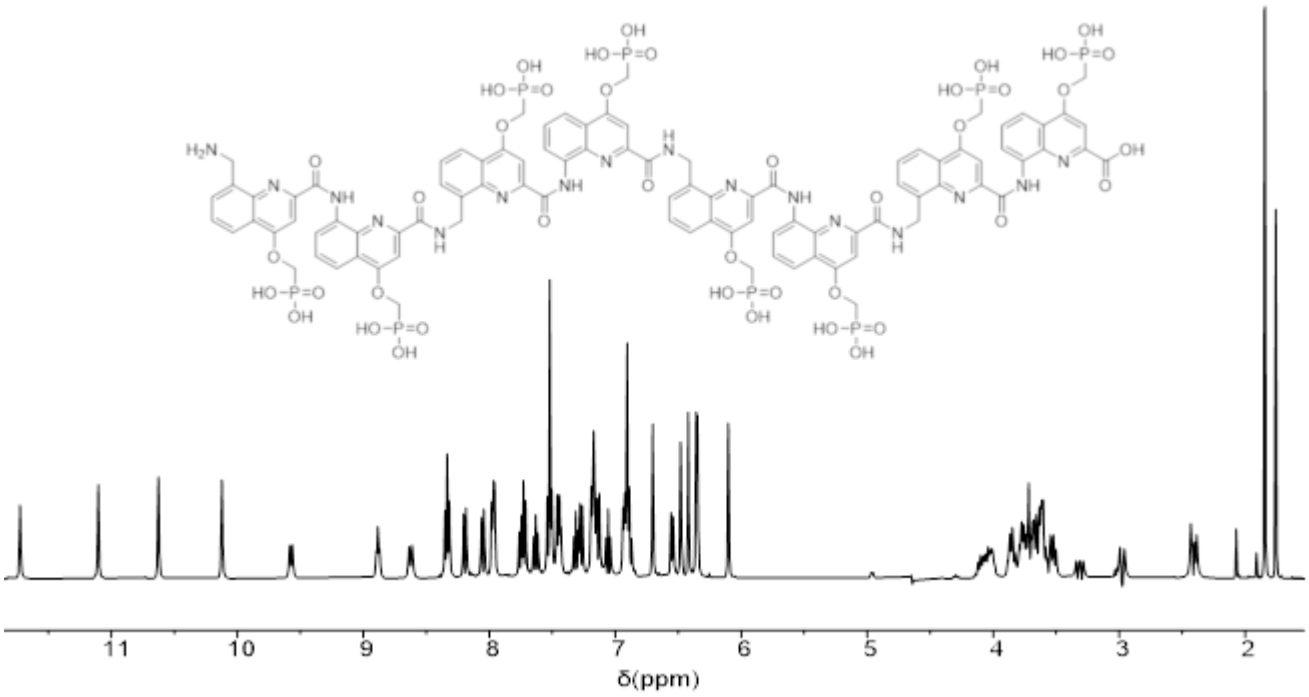


Figure S38. ^{1}H NMR spectra of **oligomer 11**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

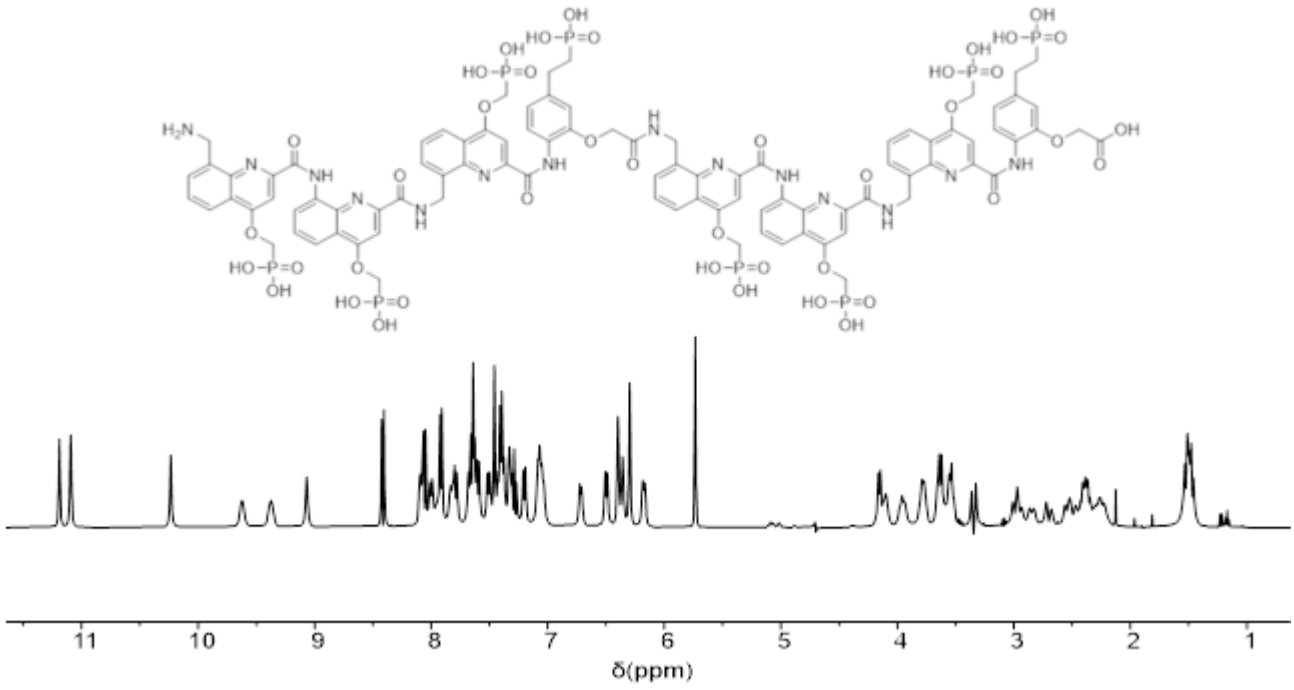


Figure S39. ^{1}H NMR spectra of **oligomer 12**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

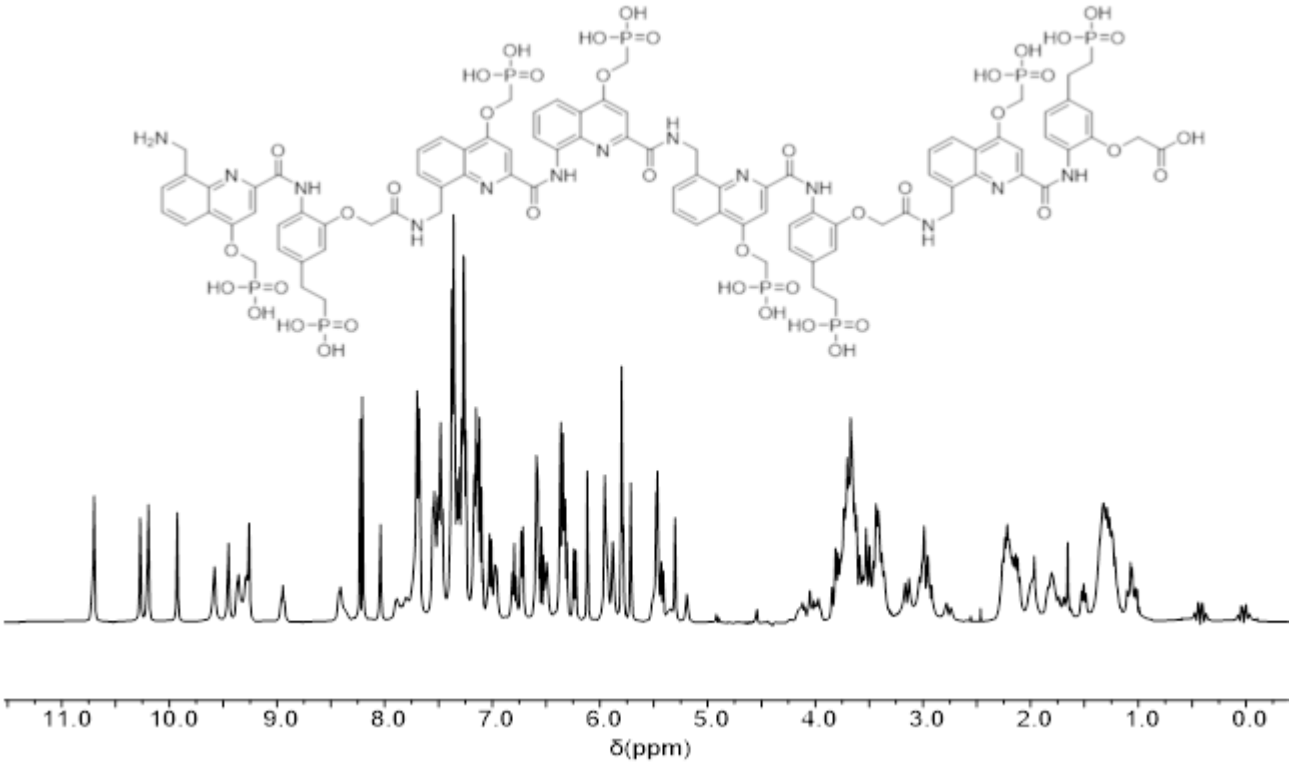


Figure S40. ^{1}H NMR spectra of **oligomer 13**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

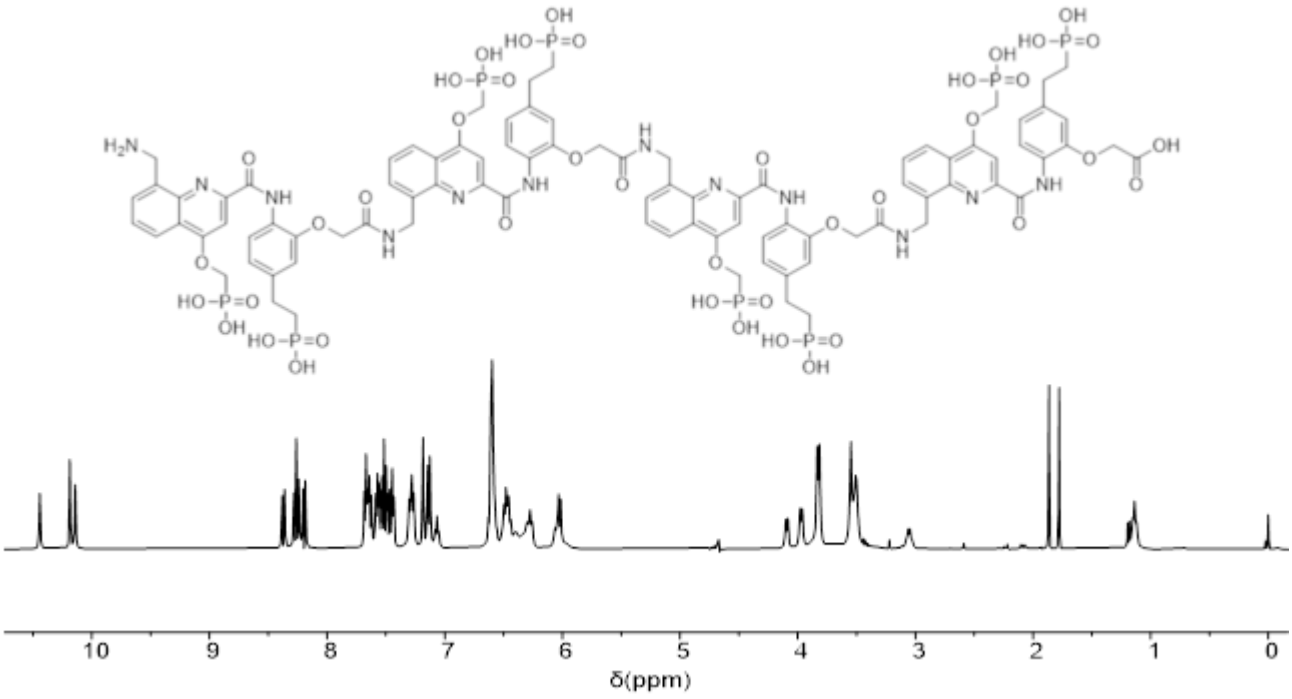


Figure S41. ^{1}H NMR spectrum of **oligomer 14**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

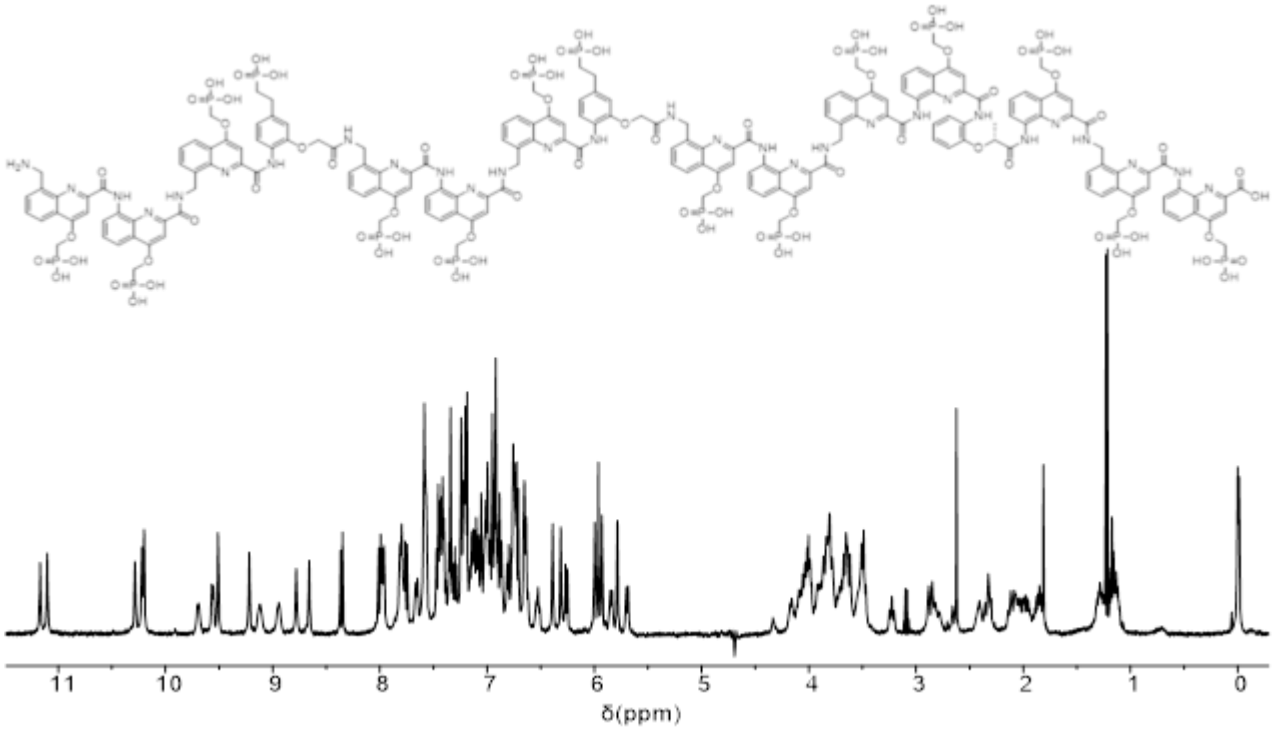


Figure S42. ^{1}H NMR spectrum of oligomer **16**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

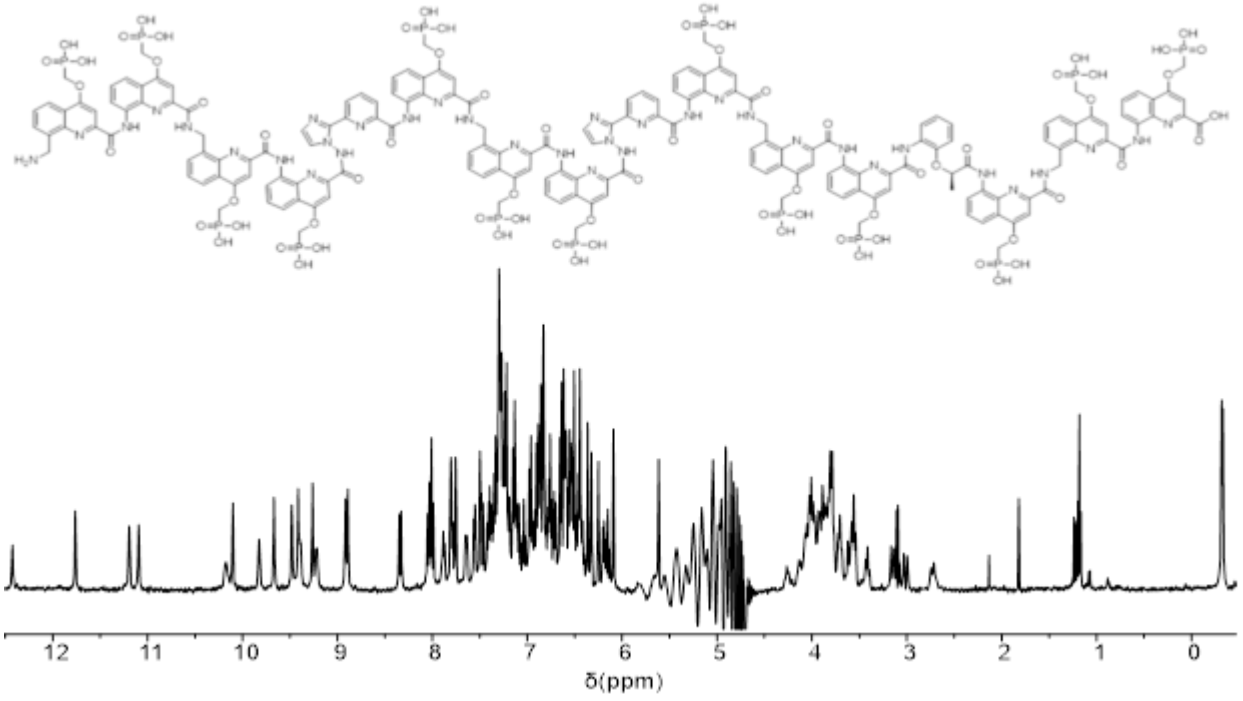


Figure S43. ^{1}H NMR spectrum of oligomer **18**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

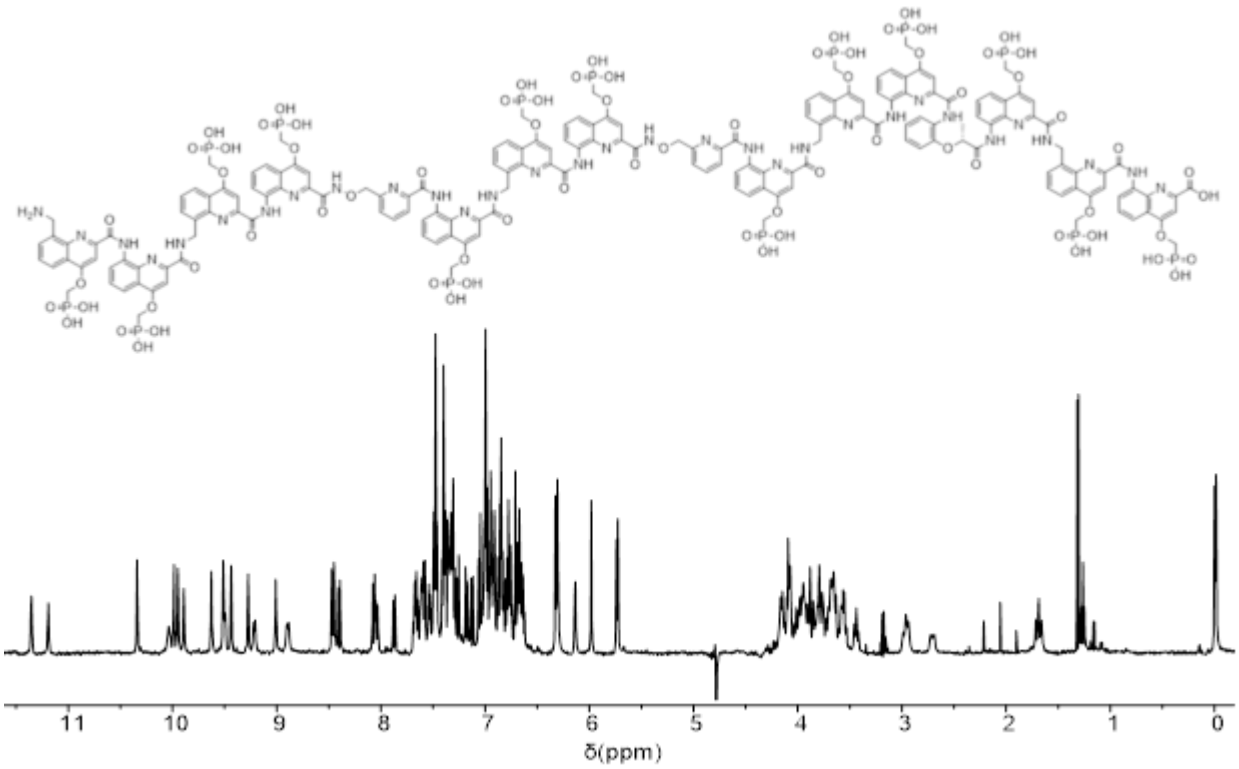


Figure S44. ^{1}H NMR spectrum of oligomer **19**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

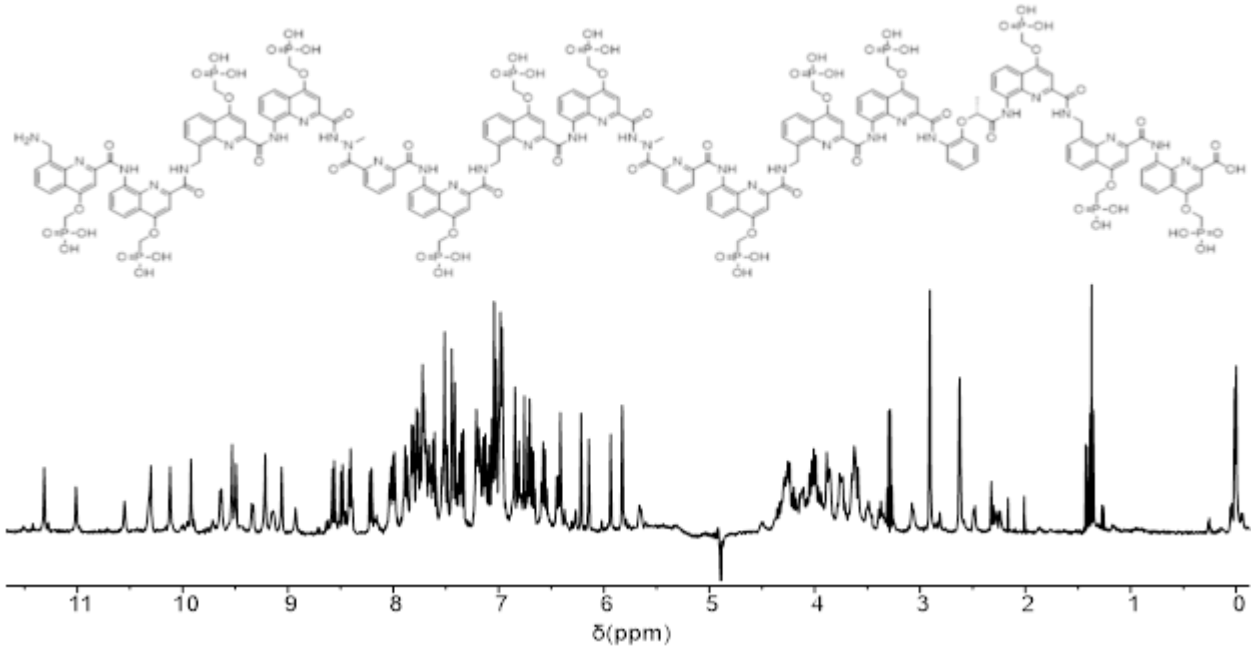


Figure S45. ^{1}H NMR spectrum of oligomer **20**. (500 MHz, 50 mM NH_4HCO_3, pH 8.5, H_2O/D_2O 9:1 v/v, at 25°C, water suppression).

5. X-ray Crystallography

Aqueous solution of oligomer **5** was prepared by dissolving the lyophilized powder in water and 12 mM ammonium acetate to a final concentration of 5mM. Aqueous solution of oligomer **6** was prepared in a similar way. Crystallization trials were carried with commercial sparse matrix screens using standard sitting drop vapor diffusion method at 293 K. X-ray quality crystals of oligomer **5** (Figure S38) were obtained after three months by the addition of 0.8 µl of oligomer **5** and 0.8 µl of 2.9 M 1,6-hexanediol, 0.050 M Tris hydrochloride pH 8.5 and 0.005 M magnesium sulfate in the reservoir. X-ray quality crystals of oligomer **6** (Figure S38) were obtained within 5 days by the addition of 0.8 µl of oligomer **6** and 0.8 µl of 30% w/v PEG 8000, 0.050 M Tris hydrochloride pH 8.5 and 0.2 M magnesium chloride in the reservoir. For low temperature diffraction measurement single crystals were fished using micro loop and plunged into liquid nitrogen. The mother liquor served as cryo-protectants for the crystals.

X-ray diffraction data for oligomer **5** was collected at the ID30B[11] beamline in European Synchrotron Radiation Facility (ESRF), Grenoble. Diffraction data was measured at T = 100 K, λ = 0.9184 Å. The crystal was exposed for 0.02 s and 0.2° oscillation per frame and a rotation pass of 360° was measured using an EIGER2 X 9M detector. Diffraction data was processed using the program XDS[12]. The crystal belonged to the space group Pbnb with four independent helices per asymmetric unit (Z = 32, Z' = 4).

X-ray diffraction data for oligomer **6** was collected at the beamline P13 operated by EMBL Hamburg, at the Petra III storage ring (DESY, Hamburg). [13] Diffraction data was measured at T = 100 K, λ = 0.8731 Å. The crystal was exposed for 0.008 s and 0.15° oscillation per frame and a rotation pass of 360° was measured using an EIGER 16M detector. Diffraction data was processed using the autoPROC pipeline. [12,14-16] The crystal belonged to the space group P1 with four independent helices per asymmetric unit (Z = Z' = 4).

Both structures were solved with SHELXD[17] structure solution program using dual space method and refined by full-matrix least-squares method on F2 with SHELXL-2014[18] within Olex2[19]. After each refinement step, visual inspection of the model and the electron-density maps were carried out using Olex2[19] and Coot[20]using 2Fo − Fc and difference Fourier (Fo − Fc) maps.

The initial structure solution of oligomer **5** revealed all main-chain atoms of four helices. Few Q^D side chains and N-terminus tail-atoms were severely disordered and omitted. The N and O atoms at C-terminus have been placed randomly. All non-H atoms were refined with anisotropic displacement parameters. After several attempts to model the disordered side chains and diethylene glycol tail, the

SQUEEZE[21] procedure was used to flatten the electron density map. Very disordered solvent molecules were removed. Calculated total potential solvent accessible void volume and electron count per cell are 32550.9 Å3 and 10144 respectively. Hydrogen atoms for oligomer **5** were placed at idealized positions.

The initial structure solution of oligomer **6** revealed majority of the main chain atoms and few side chains of three helices out of four. After several iterations of least square refinement, the main chain traces were established for three helices out of four. For one helix, only seven monomers out of nine could be traced. Throughout refinement, the terminal PEG tails and some side chain atoms of Q^D and Q^A were not modelled due to the poor quality of data. After several attempts to model the disordered side chains and diethylene glycol tail, the SQUEEZE[16] procedure was used to flatten the electron density map. Calculated total potential solvent accessible void volume and electron count per cell are 10171 Å3 and 3300 respectively. Hydrogen atoms for oligomer **6** were placed at idealized positions.

Statistics of data collection and refinement of oligomer **5** and oligomer **6** are described in Table S1. The final cif file of oligomer **5** and oligomer **6** were examined in IUCr's checkCIF algorithm. Due to the large volume fractions of disordered solvent molecules, weak diffraction intensity and poor resolution, a number of A- and B- level alerts remain in the checkCIF file. These alerts are inherent to the data set and refinement procedures. They are listed below and were divided into two groups. The first group demonstrates weak quality of the data and refinement statistics when compared to those expected for small molecule structures from highly diffracting crystals. The second group is concerned to decisions made during refinement and explained below. Atomic coordinates and structure factors of oligomer **5** and oligomer **6** were deposited in the Cambridge Crystallographic Data Centre (CCDC) with accession codes 2286782 and 2478322 respectively. The data is available free of charge upon request (www.ccdc.cam.ac.uk/).

CheckCIF validation of oligomer 5:

Group 1 alerts (these illustrate weak quality of data and refinement statistics if compared to small molecule structures from highly diffracting crystals):

THETM01_ALERT_3_A The value of sine(theta_max)/wavelength is less than 0.550
Calculated sin(theta_max)/wavelength = 0.4836
PLAT082_ALERT_2_A High R1 Value 0.24
PLAT084_ALERT_3_B High wR2 Value (i.e. > 0.25) 0.61
PLAT410_ALERT_2_A Short Intra H...H Contact
PLAT411_ALERT_2_A Short Intra H...H Contact

PLAT602_ALERT_2_A Solvent Accessible VOID(S) in Structure

PLAT241_ALERT_2_B High 'MainMol' Ueq as Compared to Neighbors

PLAT242_ALERT_2_B Low 'MainMol' Ueq as Compared to Neighbors

PLAT340_ALERT_3_B Low Bond Precision on C-C Bonds ..... 0.01624 Ang.

PLAT430_ALERT_2_A Short Inter D...A Contact

PLAT964_ALERT_2_B SHELXL WEIGHT Par. Values in CIF & RES Differ

Group 2 alert (is connected with decision made during refinement and explained below):

PLAT306_ALERT_2_B Isolated Oxygen Atom (H-atoms Missing ?) Check

Dummy O atom was introduced into refinement.

CheckCIF validation of oligomer 6:

Group 1 alerts (these illustrate weak quality of data and refinement statistics if compared to small molecule structures from highly diffracting crystals):

THETM01_ALERT_3_A The value of sine(theta_max)/wavelength is less than 0.550

Calculated sin(theta_max)/wavelength = 0.4348

PLAT082_ALERT_2_A High R1 Value 0.22

PLAT084_ALERT_3_B High wR2 Value (i.e. > 0.25) 0.53

PLAT412_ALERT_2_A Short Intra H...H Contact

PLAT414_ALERT_2_A Short Intra H...H Contact

PLAT602_ALERT_2_A Solvent Accessible VOID(S) in Structure

PLAT242_ALERT_2_B Low 'MainMol' Ueq as Compared to Neighbors

PLAT340_ALERT_3_B Low Bond Precision on C-C Bonds ..... 0.02305 Ang.

PLAT430_ALERT_2_A Short Inter D...A Contact

DIFMN02_ALERT_2_B The minimum difference density is < -0.1*ZMAX*1.00

_refine_diff_density_min given = -0.877 Test value = -0.800

PLAT097_ALERT_2_B Large Reported Max. (Positive) Residual Density 1.12 eA-3

PLAT098_ALERT_2_B Large Reported Min. (Negative) Residual Density -0.88 eA-3

PLAT315_ALERT_2_B Singly Bonded Carbon Detected

PLAT911_ALERT_3_B Missing FCF Refl Between Thmin & STh/L= 0.435

Table S1. Crystallographic data and refinement details for oligomer **5** and oligomer **6**.

Identification code	oligomer **5**	oligomer **6**.
Empirical formula	$C_{86.5}H_{69.8}Mg_{0.4}N_{17}O_{26.9}$	$C_{91}H_{58.3}N_{18.3}O_{18}$
Formula weight	1786.46	1691.31
Temperature	100 K	100.15 K
Wavelength	0.8731 Å	0.8731 Å
Crystal system	Orthorhombic	Triclinic
Space group	*P*bnb	*P*1
Unit cell dimensions	*a* = 34.10 (6) Å	*a* = 22.09 (2) Å
Volume	89734 (232) Å^3	16857 (26)
Z	32	4
Density (calculated)	1.058 g/cm^3	0.666 g/cm^3
Absorption coefficient	0.135 µ/mm^{-1}	0.079 µ/mm^{-1}
Colour and shape	Yellow, block	Yellow, needle
Crystal size	0.15 x 0.05 x 0.02 mm	0.20 x 0.03 x 0.02 mm
Index ranges	$-32 \leq h \leq 32$; $-42 \leq k \leq 43$	$-19 \leq h \leq 19$; $-21 \leq k \leq 21$
Reflections collected	40953	42226
R_{int}	0.0580	0.0323
Data/restraints/parameters	40953/2622/172	42226/930/474
Goodness-of-fit on F^2	2.352	2.602
Final R indexes [$I > 2\sigma(I)$]	$R_1 = 0.2980$; $wR_2 = 0.5669$	$R_1 = 0.2220$; $wR_2 = 0.4710$
Final R indexes [all data]	$R_1 = 0.2436$; $wR_2 = 0.6144$	$R_1 = 0.2448$; $wR_2 = 0.5272$
Largest diff. peak and hole	1.0/-0.6 e Å^{-3}	1.1/-0.8 e Å^{-3}
CCDC #	2286782	2478322

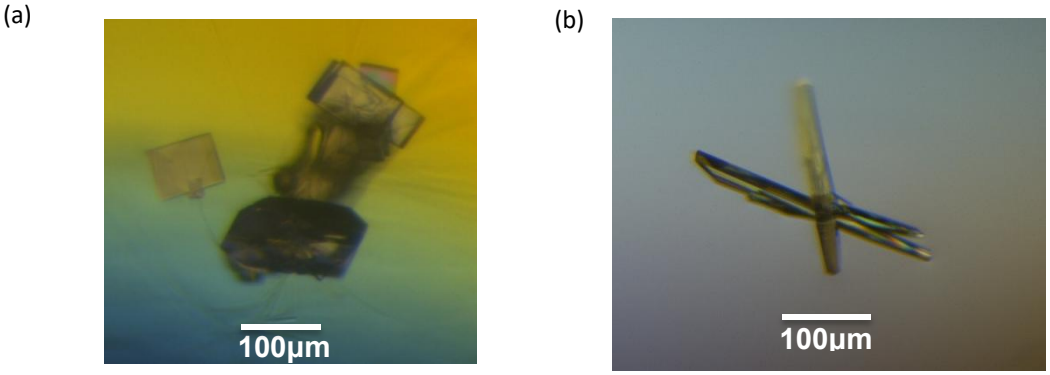


Figure S46. Crystals of oligomer **5** (a) and oligomer **6** (b) observed under crossed polarizing microscope.

The X-ray intensity data of compound **1** and compound **1d** were measured on a Bruker D8 Venture TXS system equipped with a multilayer mirror monochromator and a Mo Kα rotating anode X-ray tube (λ = 0.71073 Å). The frames were integrated with the Bruker SAINT software package. [22]Data

were corrected for absorption effects using the Multi-Scan method (SADABS).[23] The structure was solved and refined using the Bruker SHELXTL Software Package[18]. All C-bound hydrogen atoms have been calculated in ideal geometry riding on their parent atoms, the N- and O-bound hydrogen atoms have been refined freely. The structure has been refined as a 2-component perfect inversion twin. The figures have been drawn at the 25% ellipsoid probability level. [24] Atomic coordinates and structure factors of compound **1** and compound **1d** were deposited in the Cambridge Crystallographic Data Centre (CCDC) with accession codes: 2514117, 2514118. The data is available free of charge upon request (www.ccdc.cam.ac.uk/).

Table S2. Crystallographic data and refinement details for compound **1** and compound **1d**.

Identification code	Compound **1**	Compound **1d**
Empirical formula	$C_{23}H_{19}N_3O_5$	$C_{23}H_{23}N_5O_7$
Formula weight	417.41	481.46
crystal size/mm	0.140 × 0.100 × 0.080	0.090 × 0.060 × 0.020
T/K	173.(2)	173.(2)
radiation	MoKα	MoKα
diffractometer	'Bruker D8 Venture TXS'	'Bruker D8 Venture TXS'
crystal system	monoclinic	triclinic
space group	'P 1 c 1'	'P -1'
a/Å	8.905(7)	9.7590(7)
b/Å	10.403(9)	10.1656(6)
c/Å	10.793(9)	12.8980(9)
α/°	90	70.497(2)
β/°	95.66(3)	80.324(2)
γ/°	90	75.834(2)
V/Å^3	995.0(14)	1164.23(14)
Z	2	2
calc. density/g cm^{-3}	1.393	1.373
μ/mm^{-1}	0.100	0.104
absorption correction	Multi-Scan	Multi-Scan
transmission factor range	0.94–0.99	0.95–1.00
refls. measured	18936	12965
R_{int}	0.0485	0.0382
mean σ(*I*)/*I*	0.0441	0.0439
θ range	3.020–27.484	3.128–25.345
observed refls.	4118	3153
x, *y* (weighting scheme)	0.0340, 0.1700	0.0385, 0.5865
hydrogen refinement	mixed	mixed
Flack parameter	0.5	?

refls in refinement	4460	4242
parameters	289	348
restraints	2	0
$R(F_{obs})$	0.0348	0.0472
$R_w(F^2)$	0.0827	0.1107
S	1.042	1.022
shift/error$_{max}$	0.001	0.001
max electron density/e Å^{-3}	0.152	0.218
min electron density/e Å^{-3}	−0.175	−0.213
CCDC #	2514117	2514118

6. References

[1] D. Bindl, E. Heinemann, P. K. Mandal and I. Huc, *Chem. Commun.* **2021**, *57*, 5662-5665.

[2] J. Buratto, C. Colombo, M. Stupfel, S. J. Dawson, et al., *Angew. Chem. Int. Ed.* **2014**, *53*, 883-887.

[3] X. Hu, S. J. Dawson, P. K. Mandal, X. de Hatten, et al., *Chem. Sci.* **2017**, *8*, 3741-3749.

[4] B. Baptiste, C. Douat-Casassus, K. Laxmi-Reddy, F. Godde, et al., *J. Org. Chem.* **2010**, *75*, 7175-7185.

[5] V. Corvaglia, D. Carbajo, P. Prabhakaran, K. Ziach, et al., *Nucleic Acids Res.* **2019**, *47*, 5511-5521.

[6] L. Thunus and M. Dejardin-Duchene, *J. Pharm. Belg.* **1969**, *24*, 3-21.

[7] R. Young and C. K. Chang, *J. Am. Chem. Soc.* **1985**, *107*, 898-909.

[8] WO2008146774, **2008**.

[9] V. Corvaglia, F. Sanchez, F. S. Menke, C. Douat, et al., *Chem. Eur. J.* **2023**, *29*, e202300898.

[10] M. Loos, L. Thurecht, J. Wu, V. Corvaglia, et al., *Chem. Sci.* **2025**.

[11] A. A. McCarthy, R. Barrett, A. Beteva, H. Caserotto, et al., *J. Synchrotron Radiat.* **2018**, *25*, 1249-1260.

[12] W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125-132.

[13] M. Cianci, G. Bourenkov, G. Pompidor, I. Karpics, et al., *J. Synchrotron Radiat.* **2017**, *24*, 323-332.

[14] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, et al., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 293-302.

[15] P. R. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 72-82.

[16] P. R. Evans and G. N. Murshudov, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2013**, *69*, 1204-1214.

[17] G. M. Sheldrick, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2008**, *64*, 112-122.

[18] G. M. Sheldrick, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **2015**, *71*, 3-8.

[19] O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard, et al., *J. Appl. Crystallogr.* **2009**, *42*, 339-341.

[20] P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. Sect. D: Struct. Biol.* **2010**, *66*, 486-501.

[21] A. L. Spek, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2009**, *65*, 148-155.

[22] B. A. Inc in *Vol.* Bruker, Madison, Wisconsin, USA, **2012**.

[23] G. M. Sheldrick in *SADABS,* University of Göttingen, Germany, **1996**.

[24] L. Farrugia, *J. Appl. Cryst.* . **2012**, *45*, 849-854.

8. Summary and perspectives

8.1. Summary

Throughout this work, we have developed the foldamer library for protein recognition. We tested the foldamers with variable length against different protein targets (Figure 20a), and concluded that longer foldamers exhibit stronger binding and inhibition. To further optimize structural features of our DNA mimics, we attempted to control the handedness of foldamer helix (Figure 20b, 20f), or incorporate C_2-symetrical features to facilitate the structural investigation (Figure 20c, 20f). Furthermore, we attached fluorescent label to our DNA mimic foldamers for two purposes (Figure 20d, 20f): visualizing the foldamers' cellular localization and quantitatively measuring the binding affinity. In parallel, we have been working on improving the cellular delivery of our DNA mimics. To achieve this, we attached some special linkers (Figure 20e, 20f), which have been widely used in antibody conjugated drugs (ADCs), to the N-termini of the foldamers. These linkers, containing a double bond, allow our collaborators to ligate an antibody to the foldamer, facilitating targeted cellular uptake. The strategies mentioned above, together with our automated foldamer synthesis, establishes a diverse library of candidates. The continuous expansion of this library significantly accelerates our research on targeting DNA-binding proteins with DNA mimic foldamers.

a) 12-mer H-MQMQMQMQMQMQ-OH
16-mer H-MQMQMQMQMQMQMQMQ-OH
18-mer H-MQMQMQMQMQMQMQMQMQ-OH
24-mer H-MQMQMQMQMQMQMQMQMQMQMQMQ-OH

b) 16-mer Ac-MQMQ[B^R]QMQMQMQMQMQ-OH
16-mer Ac-MQMQ[B^S]QMQMQMQMQMQ-OH

c) 18-mer HO-QMQM[B^R]QMQM-[Linker]-MQMQ[B^R]MQMQ-OH
18-mer Ac-MQMQ[B^R]MQMQ-[Linker]-QMQM[B^R]QMQM-Ac

d) 12-mer [FAM]-[peg]-MQMQMQMQMQMQ-OH
16-mer [FAM]-[peg]-MQMQMQMQMQMQMQMQ-OH
16-mer [FAM]-[peg]-MQMQ[B^R]QMQMQMQMQMQ-OH
16-mer [FAM]-[peg]-MQMQ[B^S]QMQMQMQMQMQ-OH
18-mer [FAM]-[peg]-MQMQMQMQMQMQMQMQMQMQ-OH
24-mer [FAM]-[peg]-MQMQMQMQMQMQMQMQMQMQMQMQMQMQ-OH

e) 16-mer [Linker]-[peg]-MQMQMQMQMQMQMQMQMQ-OH

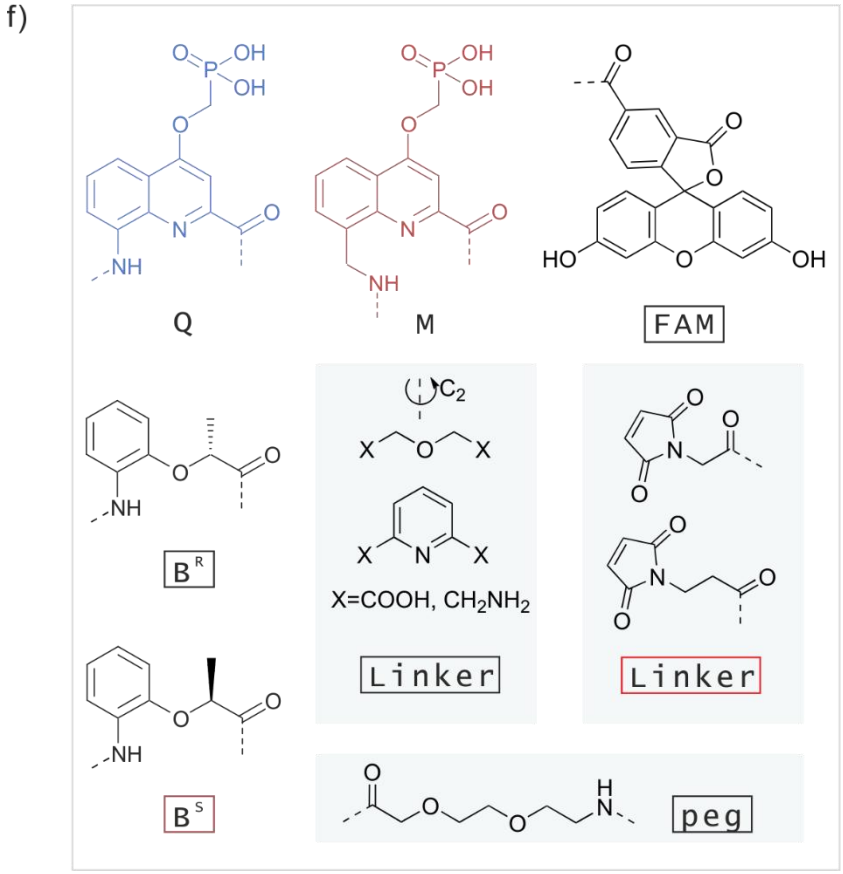


Figure 20. a)-d). The most commonly used DNA mimic foldamers from the foldamer library. f) Structure formulae of amino acid monomers Q, M, B^R, B^S and other motifs used in DNA mimic foldamers.

The work summarized in chapter 5-7, represents a milestone of our research on DNA mimic foldamers. For the first time, we managed to quantitatively bias the handedness of DNA mimic foldamers without disrupting the arrangement of charge distribution, which is crucial for the foldamers to act as DNA surface mimicry. Thanks to the chiral B monomer, the DNA mimic foldamers no longer exist in a racemic mixture of P- and M- helicity, but only in M-helicity with a double helical array that matches the positions of phosphates in duplex B-DNA. In addition to this improvement, we have also introduced another two new features in to the DNA mimic foldamers. To mimic palindromic DNA, we developed ligation chemistry and successfully synthesized C_2-symetrical DNA mimic foldamers with four linkers separately. The validation of this ligation chemistry pave the way for getting the crystal structure of Sac7d-foldmer complex. Furthermore, by introducing sticky ends in to the DNA mimic foldamers, we got the first crystal structure of DNA mimic foldamer bearing phosphonic side chains.

Based on these new features designed for DNA mimic foldamers, we set out to investigate the interaction between foldamers and a chromosomal protein, where we solved the first crystal structure of protein-foldamer complex. This is a major step towards structural studies of the interaction between DNA-binding proteins and DNA mimic foldamers. One important message from the crystal structure is that DNA mimics bind to Sac7d at the DNA binding site though a different binding mode. Beyond the structural insights, our binding assays show that the foldamers bind to Sac7d better than a DNA duplex of comparable length. Specifically, we adapted the melting curve studies to our foldamer research, that commonly used for measuring the melting temperature of DNA duplex. In this study, we found that our foldamers can reduce the thermal stability of DNA enhanced by Sac7d. Meanwhile, we managed to monitor the heat change of Sac7d binding to a 16-mer, and found that the thermodynamic of Sac7d binding to a 16-mer is different from binding to DNA. Finally, we visualized the process where the addition of 16-mer can reverse the compaction of a circular DNA promoted by the presences of Sac7d.

Alone this line, we attempted to optimize our DNA mimic foldamers into highly-selective candidates for protein recognition. To achieve this, we started to enrich the building blocks used for making DNA mimic foldamers. In this work, four new monomers has been designed and synthesized, providing the possibility to enhance the groove features of DNA mimic foldamers, which is required to target DNA-binding proteins in a sequence-selective way. We valid the conformation and behaviour of these new monomers within oligoamides in both solid-states and solution studies. The results showed that three out of four monomers behave in the way as designed. Importantly, their incorporation into the DNA mimic foldamers successfully altered the flexibility of the helical fold and introduced new groove features.

8.2. Future challenges

Despite demonstrating that DNA mimic foldamers can outcompete DNA for binding to DNA-binding proteins (eg: Top1, HIV integrase, Sac7d, TREX1 and HP1α), there are three main challenges to advancing toward therapeutic applications. First, Designing DNA mimics with high specificity remains extremely difficult due to the lack of a well-established rational design principle and the limited diversity of building blocks for foldamer design. So far, our success is primarily with proteins that bind DNA non-sequence selectively. Moving towards the sequence-selective binding, we need to enhance the groove feature of DNA mimic foldamers by developing diverse building blocks. Automated foldamer synthesis will accelerate the

expansion of our foldamer library, and the display selection will promote the high-throughput screening, providing essential feedback for effective rational optimization.

Another challenge is the cellular delivery of DNA mimic foldamers. For any nucleic acid mimicry to be effective, it must to be cell-permeable and able to reach their protein target inside the cell. While DNA mimics inherently offer the advantage of high stability due to their unique backbone ,which prevent enzymatic degradation, their poly-anionic nature and molecular size create a barrier for them to cross the cell membrane. To improve the cellular uptake, we can gain lessons from the delivery of nucleic acids or antibody conjugated drugs (ADCs). This involves designing the motifs for conjugation and cellular delivery.

Structural characterization is another major bottleneck that slow down our steps towards iterations of foldamer design. Understanding how the DNA mimic foldamer interacts with its target protein is crucial for rational design and optimization. But we’re struggling to obtain high-resolution structural data, such as through X-ray crystallography or cryo-electron microscopy. We sometimes encounter precipitation after mixing foldamers with various proteins, making the crystallization or sample preparation process difficult. Furthermore, success in solving the complex structure is often dependent on having a reliable computational model of the foldamer itself.

9. References

[1] S. Minchin and J. Lodge, *Essays Biochem.* **2019**, *63*, 433-456.

[2] D. L. Beveridge, T. E. Cheatham, 3rd and M. Mezei, *J. Biosci.* **2012**, *37*, 379-397.

[3] J. Choi and T. Majima, *Chem. Soc. Rev.* **2011**, *40*, 5893-5909.

[4] J. B. Krall, P. J. Nichols, M. A. Henen, Q. Vicens, et al., *Molecules.* **2023**, *28*, 843.

[5] E. Rivas and S. R. Eddy, *J. Mol. Biol.* **1999**, *285*, 2053-2068.

[6] Y. Tanaka, S. Fujii, H. Hiroaki, T. Sakata, et al., *Nucleic Acids Res.* **1999**, *27*, 949-955.

[7] Q. Zhou, M. Li, X. Wang, Q. Li, et al., *Int J Biol Sci.* **2012**, *8*, 118-123.

[8] W. Krause, *Biomedicines.* **2023**, *11*.

[9] J. C. Kaczmarek, P. S. Kowalski and D. G. Anderson, *Genome Med.* **2017**, *9*, 60.

[10] C. F. Bennett and E. E. Swayze, *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50*, 259-293.

[11] R. L. Juliano, *Nucleic Acids Res.* **2016**, *44*, 6518-6548.

[12] S. Balakrishnan, M. J. Scheuermann and N. J. Zondlo, *Chembiochem.* **2012**, *13*, 259-270.

[13] X. Han, D. W. Domaille, B. D. Fairbanks, L. He, et al., *Biomacromolecules.* **2018**, *19*, 4139-4146.

[14] S. Epple, A. Modi, Y. R. Baker, E. Węgrzyn, et al., *J. Am. Chem. Soc.* **2021**, *143*, 16293-16301.

[15] Y. Takahashi, K. Kakuta, Y. Namioka, A. Igarashi, et al., *J. Org. Chem.* **2023**, *88*, 10617-10631.

[16] H. Šipova, T. Špringer, D. Rejman, O. Šimak, et al., *Nucleic Acids Res.* **2014**, *42*, 5378-5389.

[17] V. S. Lelyveld, D. K. O'Flaherty, L. Zhou, E. C. Izgu, et al., *Nucleic Acids Res.* **2019**, *47*, 8941-8949.

[18] H. Challa and T. C. Bruice, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2423-2427.

[19] S. P. Pradeep, S. Malik, F. J. Slack and R. Bahal, *RNA.* **2023**, *29*, 434-445.

[20] R. Brazil, *ACS Cent. Sci.* **2023**, *9*, 3-6.

[21] S. T. Crooke, X. H. Liang, B. F. Baker and R. M. Crooke, *J. Biol. Chem.* **2021**, *296*, 100416.

[22] S. F. Dowdy, *Nat. Biotechnol.* **2017**, *35*, 222-229.

[23] D. R. Scoles, E. V. Minikel and S. M. Pulst in *Antisense oligonucleotides: A primer*, *Vol. 5* **2019**, p. e323.

[24] S. Ochoa and V. T. Milam, *Molecules.* **2020**, *25*.

[25] H. Peacock, A. Kannan, P. A. Beal and C. J. Burrows, *J. Org. Chem.* **2011**, *76*, 7295-7300.

[26] R. Owczarzy, Y. You, C. L. Groth and A. V. Tataurov, *Biochem.* **2011**, *50*, 9352-9367.

[27] M. Jasiński, J. Miszkiewicz, M. Feig and J. Trylska, *J Phys Chem B.* **2019**, *123*, 8168-8177.

[28] M. Petersen and J. Wengel, *Trends in Biotechnol.* **2003**, *21*, 74-81.

[29] M. C. Chakrabarti and F. P. Schwarz, *Nucleic Acids Res.* **1999**, *27*, 4801-4806.

[30] A. Carlesso, J. Hörberg, G. Deganutti, A. Reymer, et al., *NAR Genom. Bioinform.* **2024**, *6*, lqae058.

[31] V. Menchise, G. De Simone, T. Tedeschi, R. Corradini, et al., *Proc. Natl. Acad. Sci. U.S.A.* . **2003**, *100*, 12021-12026.

[32] P. L. Cornut, S. Boisset, J. P. Romanet, M. Maurin, et al., *Surv. Ophthalmol.* **2014**, *59*, 286-303.

[33] A. Loy, M. Horn and M. Wagner, *Nucleic Acids Res.* **2003**, *31*, 514-516.

[34] M. Hendling and I. Barisic, *Comput. Struct. Biotechnol. J.* **2019**, *17*, 1056-1065.

[35] M. Ashwini, S. B. Murugan, S. Balamurugan and R. Sathishkumar, *Mol. Biol.* **2016**, *50*, 3-9.

[36] B. G. de la Torre and F. Albericio, *Molecules.* **2024**, *29*.

[37] L. A. Buthelezi, S. Pillay, N. N. Ntuli, L. Gcanga, et al., *Cells.* **2023**, *12*.

[38] L. Moumne, A. C. Marie and N. Crouvezier, *Pharmaceutics.* **2022**, *14*.

[39] M. Moccia, B. Pascucci, M. Saviano, M. T. Cerasa, et al., *Int. J. Mol. Sci.* **2023**, *25*.

[40] X. Tang, H. Wang, Y. Yin and G. Zhong, *Gene Therapy.* **2025**, *32*, 163-171.

[41] I. M. Gabas and P. E. Nielsen, *Biomacromolecules.* **2020**, *21*, 472-483.

[42] R. Bahal, N. A. McNeer, D. H. Ly, W. M. Saltzman, et al., *Artificial DNA: PNA & XNA.* **2013**, *4*, 49-57.

[43] S. Qassem, D. Breier, G. S. Naidu, I. Hazan-Halevy, et al., *Mol. Ther. Nucleic Acids.* **2024**, *35*.

[44] M. C. Morris, J. Depollier, J. Mery, F. Heitz, et al., *Nat. Biotechnol.* **2001**, *19*, 1173-1176.

[45] S. Tomassi, C. Ieranò, M. E. Mercurio, E. Nigro, et al., *Bioorg. Med. Chem.* **2018**, *26*, 2539-2550.

[46] S. Volpi, U. Cancelli, M. Neri and R. Corradini, *Pharmaceuticals* **2020**, *14*.

[47] H. C. Wang, C. H. Ho, K. C. Hsu, J. M. Yang, et al., *Biochem.* **2014**, *53*, 2865-2874.

[48] H. C. Wang, C. C. Chou, K. C. Hsu, C. H. Lee, et al., *IUBMB Life.* **2019**, *71*, 539-548.

[49] D. Liu, R. Ishima, K. I. Tong, S. Bagby, et al., *Cell.* **1998**, *94*, 573-583.

[50] L. M. Parsons, D. C. Yeh and J. Orban, *Proteins.* **2004**, *54*, 375-383.

[51] C. D. Putnam, M. J. Shroyer, A. J. Lundquist, C. D. Mol, et al., *J. Mol. Biol.* **1999**, *287*, 331-346.

[52] M. D. Walkinshaw, P. Taylor, S. S. Sturrock, C. Atanasiu, et al., *Mol. Cell.* **2002**, *9*, 187-194.

[53] J. L. Asensio, L. Pérez-Lago, J. M. Lázaro, C. González, et al., *Nucleic Acids Res.* **2011**, *39*, 9779-9788.

[54] D. Yuksel, P. R. Bianco and K. Kumar, *Mol. Biosyst.* **2016**, *12*, 169-177.

[55] F. Crick, *Nature.* **1970**, *227*, 561-563.

[56] K. Kurtz, F. Martinez-Soler, J. Ausio and M. Chiva, *J Cell Biochem.* **2008**, *105*, 574-584.

[57] V. Kleene, V. Corvaglia, E. Chacin, I. Forne, et al., *Nucleic Acids Res.* **2023**, *51*, 9629-9642.

[58] E. Chacin, K.-U. Reusswig, J. Furtmeier, P. Bansal, et al., *Nature.* **2023**, *616*, 836-842.

[59] D. Kopytova, V. Popova, M. Kurshakova, Y. Shidlovskii, et al., *Nucleic Acids Res.* **2016**, *44*, 4920-4933.

[60] N. Chatterjee and G. C. Walker, *Environ. Mol. Mutagen.* **2017**, *58*, 235-263.

[61] B. Kaina, M. Christmann, S. Naumann and W. P. Roos, *DNA Repair.* **2007**, *6*, 1079-1099.

[62] F. Drabløs, E. Feyzi, P. A. Aas, C. B. Vaagbø, et al., *DNA Repair.* **2004**, *3*, 1389-1407.

[63] M. Schauperl, M. Podewitz, B. J. Waldner and K. R. Liedl, *J. Chem. Theory Comput.* **2016**, *12*, 4600-4610.

[64] C. M. Lukacs and A. K. Aggarwal, *Curr. Opin. Struct. Biol.* **2001**, *11*, 14-18.

[65] R. Rohs, X. Jin, S. M. West, R. Joshi, et al., *Annu. Rev. Biochem.* **2010**, *79*, 233-269.

[66] L. Yang, Y. Orenstein, A. Jolma, Y. Yin, et al., *Mol. Syst. Biol.* **2017**, *13*, 910.

[67] K. K. Swinger and P. A. Rice, *Curr. Opin. Struct. Biol.* **2004**, *14*, 28-35.

[68] N. M. Luscombe, R. A. Laskowski and J. M. Thornton, *Nucleic Acids Res.* **2001**, *29*, 2860-2874.

[69] Y. Mandel-Gutfreund, O. Schueler and H. Margalit, *J. Mol. Biol.* **1995**, *253*, 370-382.

[70] C. O. Pabo and R. T. Sauer, *Annu.l Revi.of Biochem.* **1992**, *61*, 1053-1095.

[71] C. K. Voong, J. A. Goodrich and J. F. Kugel, *Biomolecules.* **2021**, *11*.

[72] F. V. Murphy, R. M. Sweet and M. E. A. Churchill, *The EMBO Journal.* **1999**, *18*, 6610-6618-6618.

[73] H. Robinson, Y. G. Gao, B. S. McCrary, S. P. Edmondson, et al., *Nature.* **1998**, *392*, 202-205.

[74] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, et al., *Nature.* **1997**, *389*, 251-260.

[75] T. P. Ko, H. M. Chu, C. Y. Chen, C. C. Chou, et al., *Acta Crystallogr. D: Biol. Crystallogr.* **2004**, *60*, 1381-1387.

[76] D. Kamashev and J. Rouviere-Yaniv, *EMBO J.* **2000**, *19*, 6527-6535.

[77] J. O. Thomas, *Biochem Soc Trans.* **2001**, *29*, 395-401.

[78] P. Pellanda, M. Dalsass, M. Filipuzzi, A. Loffreda, et al., *EMBO J.* **2021**, *40*, e105464.

[79] S. C. Verma, A. Harned, K. Narayan and S. Adhya, *Mol. Microbiol.* **2023**, *119*, 439-455.

[80] S. N. Peterson, F. W. Dahlquist and N. O. Reich, *J. Mol. Biol.* **2007**, *369*, 1307-1317.

[81] Z. Wu, M. Nicoll and R. J. Ingham, *Exp. Hematol. Oncol.* **2021**, *10*, 4.

[82] M. L. Stolz and C. McCormick, *Viruses.* **2020**, *12*.

[83] B. E. Lonze and D. D. Ginty, *Neuron.* **2002**, *35*, 605-623.

[84] T. Vaijayanthi, G. N. Pandian and H. Sugiyama, *Advanced Therapeutics.* **2023**, *6*, 2300022.

[85] J. M. Gottesfeld, J. M. Turner and P. B. Dervan, *Gene Expr.* **2000**, *9*, 77-91.

[86] C. M. White, A. L. Satz, T. C. Bruice and T. A. Beerman, *Proc. Natl. Acad. Sci. U.S.A. .* **2001**, *98*, 10590-10595.

[87] T. J. Grant, J. A. Bishop, L. M. Christadore, G. Barot, et al., *Proc. Natl. Acad. Sci. U.S.A. .* **2012**, *109*, 4503-4508.

[88] Q. Li, R. M. Karim, M. Cheng, M. Das, et al., *Oncogene.* **2020**, *39*, 5187-5200.

[89] C. Li, Z. Zhou, C. Ren, Y. Deng, et al., *Front. Pharmacol.* **2022**, *13*, 1007723.

[90] M. M. Tehran, S. Rezaei, A. Jalili, S. H. Aghaee-Bakhtiari, et al., *Drug Discov. Today.* **2020**, *25*, 195-200.

[91] S. Li, X. Wang, J. Huang, X. Cao, et al., *Cell Death Dis.* **2025**, *16*, 197.

[92] N. C. Yeo, A. Chavez, A. Lance-Byrne, Y. Chan, et al., *Nat. Methods.* **2018**, *15*, 611-616.

[93] L. Wang, D. Ji, X. Liu, W. Lei, et al., *J. Med. Chem.* **2025**, *68*, 5055-5074.

[94] A. R. Chandrasekaran and D. A. Rusling, *Nucleic Acids Res.* **2018**, *46*, 1021-1037.

[95] R. R. Beerli, B. Dreier and C. F. Barbas, *Proc. Natl. Acad. Sci. U.S.A. .* **2000**, *97*, 1495-1500.

[96] Y. Li, Y. Jiang, H. Chen, W. Liao, et al., *Nat. Chem. Biol.* **2015**, *11*, 207-213.

[97] A. Olson, B. Basukala, S. Lee, M. Gagne, et al., *Viruses.* **2020**, *12*.

[98] G. Guichard and I. Huc, *Chem. Commun.* **2011**, *47*, 5933-5941.

[99] C. M. Goodman, S. Choi, S. Shandler and W. F. DeGrado, *Nat. Chem. Biol.* **2007**, *3*, 252-262.

[100] P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.

[101] A. M. Garcia, F. J. Romero-Salguero, D. M. Bassani, J.-M. Lehn, et al., *Chem. Eur. J.* **1999**, *5*, 1803-1808.

[102] J. L. Hou, X. B. Shao, G. J. Chen, Y. X. Zhou, et al., *J. Am. Chem. Soc.* **2004**, *126*, 12386-12394.

[103] M. Kudo and A. Tanatani, *New. J. Chem.* **2015**, *39*, 3190-3196.

[104] X. Yang, L. Yuan, K. Yamato, A. L. Brown, et al., *J. Am. Chem. Soc.* **2004**, *126*, 3148-3162.

[105] X. Hu, P. K. Mandal, B. Kauffmann and I. Huc, *ChemPlusChem.* **2020**, *85*, 1580-1586.

[106] J. M. Rogers, S. Kwon, S. J. Dawson, P. K. Mandal, et al., *Nat. Chem.* **2018**, *10*, 405-412.

[107] C. G. Pappas, P. K. Mandal, B. Liu, B. Kauffmann, et al., *Nat. Chem.* **2020**, *12*, 1180-1186.

[108] E. J. Petersson, C. J. Craig, D. S. Daniels, J. X. Qiu, et al., *J. Am. Chem. Soc.* **2007**, *129*, 5344-5345.

[109] C. G. Pappas, P. K. Mandal, B. Liu, B. Kauffmann, et al., *Nature Chemistry.* **2020**, *12*, 1180-1186.

[110] D. M. Bassani, J.-M. Lehn, G. Baum and D. Fenske, *Angew. Chem. Int. Ed.* **1997**, *36*, 1845-1847.

[111] L. Sebaoun, B. Kauffmann, T. Delclos, V. Maurizot, et al., *Organic Lett.* **2014**, *16*, 2326-2329.

[112] D. Bindl in *Folding properties, handedness control and aggregation behavior of helical aromatic δ-amino acid foldamers in water*, *Vol.* **2022**.

[113] M. Zwillinger, P. S. Reddy, B. Wicher, P. K. Mandal, et al., *Chemistry – A European Journal.* **2020**, *26*, 17366-17370.

[114] M. Horeau, G. Lautrette, B. Wicher, V. Blot, et al., *Angew. Chem. Int. Ed.* **2017**, *56*, 6823-6827.

[115] R. Hess, M. Brenet, H. Rajaonarivelo, M. Gauthier, et al., *Angew. Chem. Int. Ed.* **2024**, *63*, e202413977.

[116] J. Wang, B. Wicher, V. Maurizot and I. Huc, *Chem. Eur. J.* **2022**, *28*, e202201345.

[117] F. Totsingan and A. J. Bell, Jr., *Protein Sci.* **2013**, *22*, 1552-1562.

[118] J. Morris, S. Jayanthi, R. Langston, A. Daily, et al., *Protein Expr. Purif.* **2016**, *126*, 93-103.

[119] K. Ziach, C. Chollet, V. Parissi, P. Prabhakaran, et al., *Nat. Chem.* **2018**, *10*, 511-518.

[120] V. Corvaglia, F. Sanchez, F. S. Menke, C. Douat, et al., *Chem. Eur. J.* **2023**, *29*, e202300898.

[121] V. Corvaglia, D. Carbajo, P. Prabhakaran, K. Ziach, et al., *Nucleic Acids Res.* **2019**, *47*, 5511-5521.

[122] Y. Zhang, G. Sicot, X. Cui, M. Vogel, et al., *Biochem.* **2011**, *50*, 10431-10441.

[123] J. S. Kang, J. L. Meier and P. B. Dervan, *J. Am. Chem. Soc.* **2014**, *136*, 3687-3694.

[124] S. Asamitsu, Y. Kawamoto, F. Hashiya, K. Hashiya, et al., *Bioorg Med Chem.* **2014**, *22*, 4646-4657.

[125] Y. Pommier and C. Marchand, *Curr. Med. Chem. Anticancer Agents.* **2005**, *5*, 421-429.

[126] Y. Pommier, E. Kiselev and C. Marchand, *Bioorg. Med. Chem. Lett.* **2015**, *25*, 3961-3965.

[127] R. G. Brennan and B. W. Matthews, *J. Biol. Chem.* **1989**, *264*, 1903-1906.

[128] J. M. Berg, *J. Biol. Chem.* **1990**, *265*, 6513-6516.

[129] W. H. Landschulz, P. F. Johnson and S. L. McKnight, *Science.* **1988**, *240*, 1759-1764.

10. Acknowledgements

My deepest gratitude goes to Professor Ivan Huc for his supervision and for providing me the opportunities to begin this incredible PhD journey. His critical thinking, hard work, and deep passion for foldamer research have been a powerful example for me, inspiring me to pursue my own career with the same enthusiasm and dedication.

I would like to extend my thanks to Céline Douat for her encouragement and for her effort in providing us with functional HPLC, Mass and Chorus in the lab; Daniel Bindl for his kindness which means a lot at the beginning of my PhD; and Valentina Corvaglia for her supervision, the well-organized daily plans, rigorous attitude to research, and so many other lessons I have learnt from her have truly made me a different person than I was five years ago.

My gratitude also goes to Wei Tuo and Mathieu Dennis for their guidance in organic synthesis; Melis Caber for her patience and kindness with every request I had; Lars Allmendinger for his help in solving my NMR puzzles; and Tulika Chakrabortty for her supervision in the bio lab. I would also like to thank Vasileios Athanasopoulos for his encouragement to make me a stronger person, and Ignacio Muñoz Alonso for his willingness to take my playful punches, which helped me through some stressful moments. Finally, thank you to Lizeth Bodero for the kindest smile that always brightened my day.

To all the group members, thank you for creating a truly collaborative and supportive work environment.

I'm also thankful to my family and friends for their support. To my partner, Keli, thank you for your understanding, care, and companionship throughout this long journey. Lastly, I want to thank myself for not giving up and for making this wonderful journey a reality.