Sequence-selective DNA recognition with artificial ligands

Dissertation

Submitted to attain the academic degree of

Doctor of natural sciences

Dr. rer. nat.

presented by

Álvaro Serrano Fariñas

Faculty of Chemistry

University of Duisburg-Essen

Essen, 2021

DOI: 10.17185/duepublico/74742
URN: urn:nbn:de:hbz:464-20210831-112449-6

Sequenzselektive DNA-Erkennung mit künstlichen Liganden

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

Dr. rer. nat.

vorgelegt von

Álvaro Serrano Fariñas

Fakultät für Chemie

Universität Duisburg-Essen

Essen, 2021

The present work was carried out in the period from July 2017 to July 2021 in the working group of Prof. Dr. Thomas Schrader at the Institute of Organic Chemistry of the University of Duisburg-Essen.

First examiner: Prof. Dr. Thomas Schrader

Second examiner: Prof. Dr. Jochen Niemeyer

Chair of the examination board: Prof. Dr. Christian Mayer

The **oral defense** was on 19.08.21

Die vorliegende Arbeit wurde im Zeitraum von Juli 2017 bis Juli 2021 im Arbeitskreis von Prof. Dr. Thomas Schrader am Institut für Organische Chemie der Universität Duisburg-Essen durchgeführt.

Referent: Prof. Dr. Thomas Schrader

Korreferent: Prof. Dr. Jochen Niemeyer

Prüfungsvorsitzender: Prof. Dr. Christian Mayer

Tag der Disputation: 19.08.21

Affidavit

I hereby certify that I have written the present thesis entitled

"Sequence-selective DNA recognition with artificial ligands"

myself and that I have not used any auxiliary means or sources other than those indicated. Furthermore, I hereby declare that I have not submitted this dissertation in this or any other form to any other university.

Essen, July 2021 _____

Álvaro Serrano Fariñas

Eidesstattliche Erklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit mit dem Titel

„Sequenzselektive DNA-Erkennung mit künstlichen Liganden“

selbst verfasst und keine, außer den angegebenen, Hilfsmitteln und Quellen benutzt habe. Des Weiteren erkläre ich hiermit, dass ich diese Dissertation in dieser oder anderer Form noch bei keiner anderen Universität eingereicht habe.

Essen, Juli 2021 _____

Álvaro Serrano Fariñas

La he liao parda

Summary

Deoxyribonucleic acid (DNA) encodes the genetic data of most organisms. Abnormalities in the genome or the unregulated overexpression of certain genes can be the cause of multiple diseases, such as cancer. Modulating the expression of genes with external agents, called antigene agents, is a long-standing goal in biotechnology and human therapy. Over the last decades, some synthetic examples, such as PNAs and oligoamides, have shown great potential to bind to desired DNA sequences in a site-specific way. However, they also suffer from limitations and sequence constraints, reinforcing the need to develop new and more efficient approaches. Therefore, we present a novel modular system designed to bind to the major groove of DNA in a sequence-selective manner, to be used as an antigene agent.

Chapter 1 will provide the necessary theoretical background to understand the aim and the challenges faced during the development of the present work. Aspects such as the structure and biological relevance of DNA are briefly reviewed, followed by a summary of the state of the art techniques to interfere with DNA function, employing both modified biological systems and entirely artificial approaches.

After outlining the aim of this work Chapter 2, where the concept of our proposed approach is explained and our designed structures are presented, its detailed implementation is discussed in Chapter 3. The synthesis of unusual fused heteroaromatic scaffolds is reported, as well as challenges faced during their preparation and the preliminary NMR binding experiments.

In Chapter 4, the achieved results are summarized, and future challenges and potential improvements are commented.

At the end of this dissertation, experimental procedures are thoroughly detailed, together with an in-depth characterization and spectroscopic data of the synthesized compounds.

Zusammenfassung

Desoxyribonukleinsäure (DNS) kodiert die genetischen Daten der meisten Organismen. Anomalien im Genom oder die unregulierte Überexpression bestimmter Gene können die Ursache für verschiedene Krankheiten wie Krebs sein. Die Modulation der Genexpression mit externen Wirkstoffen, sogenannten Antigenen, ist ein langjähriges Ziel in der Biotechnologie und Humantherapie. In den letzten Jahrzehnten haben einige synthetische Beispiele, wie PNAs und Oligoamide, ein großes Potenzial zur ortsspezifischen Bindung an gewünschte DNA-Sequenzen gezeigt. Diese Beispiele bleiben jedoch eingeschränkt und weisen Sequenzbeschränkungen auf, was die Notwendigkeit verstärkt, neue und effizientere Ansätze zu entwickeln. Daher präsentieren wir ein neuartiges modulares System, das entwickelt wurde, um sequenzselektiv an die große Furche der DNA zu binden und als Antigen zu wirken.

Kapitel 1 bildet den notwendigen theoretischen Hintergrund, um das Ziel der vorliegenden Arbeit zu verstehen sowie die Herausforderungen, derer sich die vorliegende Arbeit angenommen hat. Aspekte wie die Struktur und biologische Relevanz der DNS werden besprochen, gefolgt von einer Zusammenfassung der etablierten Techniken zur Störung der DNS-Funktion, was sowohl durch modifizierte biologische Systeme als auch durch vollständig künstliche Ansätze erfolgen kann.

Nach der Erläuterung des Ziels dieser Arbeit und des Konzeptes des vorgeschlagenen Ansatzes mit den entworfenen Strukturen in Kapitel 2, beschreibt Kapitel 3 die detaillierte Implementierung dieses Konzeptes. Die Synthese außergewöhnlich verbrückter, heteroaromatischer Gerüste wird ebenso beschrieben wie die damit einhergehenden Hürden, die während ihrer Herstellung und den vorläufigen NMR-Bindungsexperimenten bewältigt wurden.

In Kapitel 4 werden die erzielten Ergebnisse zusammengefasst und zukünftige Herausforderungen und Verbesserungspotenziale aufgezeigt.

Am Ende dieser Dissertation werden die durchgeführten experimentellen Verfahren, gemeinsam mit der eingehenden Charakterisierung der synthetisierten Verbindungen und ihrer spektroskopischen Daten, ausführlich beschrieben.

Table of Contents

1. Introduction

1.1. DNA

Deoxyribonucleic acid (DNA) is the fundamental storage material of genetic information. The well-defined sequences of nucleobases in the DNA, called genes, dictate the proteins to be encoded, which in turn define multiple features such as physical traits or disease susceptibility. Essentially, genes encode proteins, and proteins dictate cell function.

The process begins when a copy of DNA is transcribed into messenger RNA (mRNA), which is later used to encode a specific protein sequence in a process called translation. This mechanism involves a dramatic amplification of the genetic information: a single copy of the source DNA sequence creates multiples copies of mRNA, each of which can be used to encode a protein multiple times. This amplification makes DNA an attractive medicinal target, as in principle only one molecule of a drug per cell would be necessary to block a target gene, while blocking the function of the multiple copies of the proteins encoded by that gene would require multiple drug molecules.[1]

Many diseases, like Alzheimer's disease or cancer, are expressed through protein action, but actually originate at the DNA level, in the sequence itself or in its processing. With the human genome fully sequenced,[2] attention is turning to establishing ways to control gene expression. Finding treatments for diseases caused by abnormalities in the genome is one of the most prominent challenges for modern medicine.

1.1.1. Structure of DNA

DNA consists of a right-handed double helix, which can be chemically characterized as a hetero-polymer constituted of different nucleotide monomers linked through their sugar-phosphate backbones (Figure 1a). Nucleotides (nt) are the fundamental building blocks of the genetic code. There are four different nucleotides in DNA, each of them containing a different nucleobase, namely adenine (A), thymine (T), guanine (G) and cytosine (C). A and G have analogous structures, consisting of a purine (a fused pyrimidine-imidazole) scaffold. They constitute the purine nucleobases. Similarly, T and C share the same pyrimidine core, and are thus termed pyrimidines. Uracil (U), which substitutes T in ribonucleic acid (RNA), is also included within the pyrimidines group. These bases are linked to a deoxyribose sugar (in DNA) or a ribose sugar (in RNA), forming a sugar-base combination called nucleoside. A phosphate group at the 5' carbon of the sugar completes

[1] Hannon, M. J. *Chem. Soc. Rev.* **2007**, *36*(2), 280–295.

[2] (a) International Human Genome Sequencing Consortium *Nature* **2004**, *431*, 931-945. (b) Pertea, M.; Salzberg, S. L. *Genome Biol.* **2010**, *11*(5), 206.

the nucleotide, which is linked to next nucleotide through a phosphodiester bond with its hydroxy group at the 3' carbon.

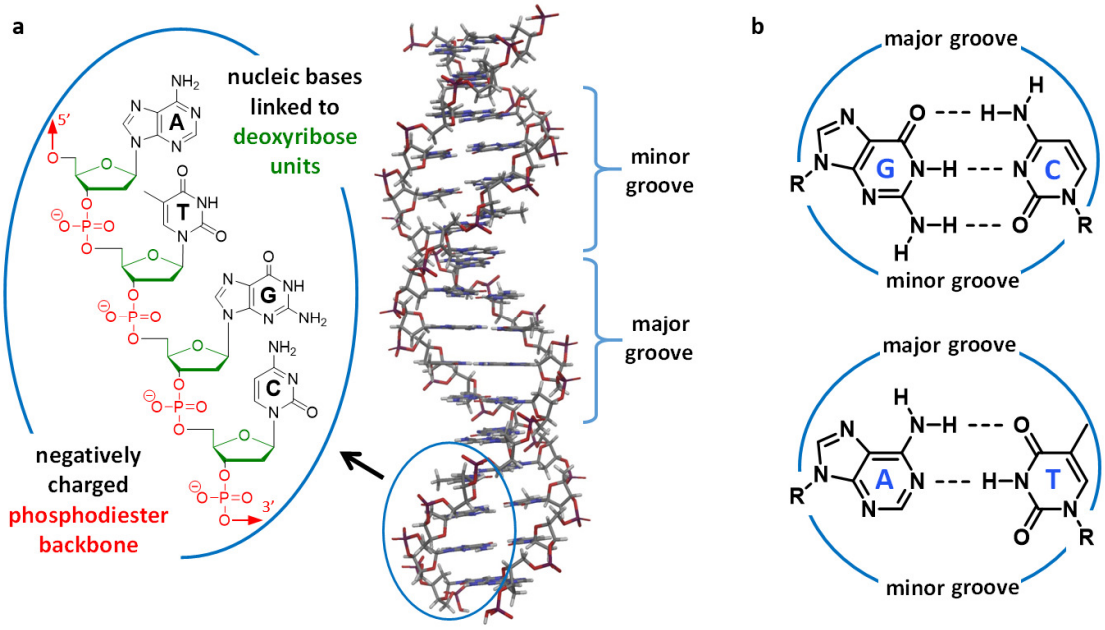


Figure 1. a) Structure of the DNA double helix. b) Watson-Crick base-pairing rules. R represents the sugar-phosphate backbone. Its asymmetrical linkage creates the major and minor grooves.

The sugar-phosphate backbone is located on the exterior of the double helix, while the nucleobases are projected towards the interior of the helix. The two strands of DNA are held together by weak interactions such as hydrophobic effects, van der Waals forces, ionic interactions and hydrogen bonds between the nucleobases (Figure 1b). These hydrogen bonds follow the Watson-Crick pairing rules: A pairs with T through two hydrogen bonds, and G pairs with C forming three hydrogen bonds. The strands are oriented in an antiparallel fashion and form a helical twist, which is required to minimize the distance between adjacent base pairs and maximize their hydrophobic stacking. As a result of the asymmetric linkage of the nucleobases to the sugar-phosphate backbone, the DNA presents two sides: the wider major groove, where the bases are completely exposed, and the narrower minor groove.

The double helix structure of DNA was solved in 1953 by James D. Watson and Francis H. C. Crick,[3] thanks to the work of Rosalind E. Franklin and Maurice H. F. Wilkins with X-ray diffraction of DNA, which provided essential evidence to support the double helix hypothesis.[4] However, until the structure of DNA was unveiled, many unsuccessful attempts were made towards its elucidation. Phoebus Levene is renowned for his "tetranucleotide

[3] Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

[4] (a)Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *171*, 740-741. (b) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. *Nature* **1953**, *171*, 738-740.

hypothesis",[5] which stated that DNA consisted of repeating sets of the four nucleotides, a tetranucleotide, and therefore having equal amounts of the four nucleobases (A, T, G, C). This belief prevailed for four decades, until proven wrong by Erwin Chargaff in 1949,[6] who experimentally showed that the quantity of each nucleobase in DNA followed A=T≠G=C, strengthening the idea of base pairing. In early 1953, right before the breakthrough of Watson and Crick, Linus Pauling and Robert Corey proposed a triple helix model, with the phosphate backbone on the inside, and the nucleotide bases pointing outwards (Figure 2).[7]

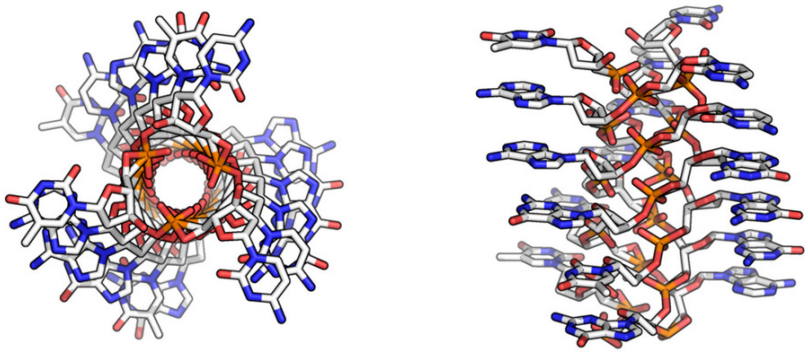


Figure 2. Top view (left) and front view (right) of the triple helix model proposed by Pauling and Corey.[8]

DNA can present itself in various forms, depending on the external conditions (Figure 3). The most common polymorph is B-DNA, which is often simply referred to as DNA. It exists under normal physiological conditions. One turn of the right handed double helix contains 10 base pairs (bp) perpendicular to the helical axis, with a length of 34 Å and a 20 Å diameter. A-DNA also consists of a right-handed double helix, but one turn can accommodate 11.6 bp (tilted 20°), with a length of 28.6 Å and a diameter of 26 Å. Therefore, A-DNA is wider but shorter than B-DNA. A-DNA appears under dehydration conditions (relative humidity below 75%). Z-DNA, on the other hand, is a left-handed double helix, winding in a zig-zag pattern. One turn has a length of 44 Å and an 18 Å diameter, making this the narrowest and longest of the three structures.[9] Although its role is unclear, it is thought to participate in some biological events, as some proteins bind specifically to it.[10]

---

[5] (a) Hargittai, I. *Struct. Chem.* **2009**, *20*, 753-756. (b) Frixione, E.; Ruiz-Zamarripa, L. *J. Biol. Chem.* **2019**, *294*(7), 2249-2255.

[6] (a) Chargaff, E.; Vischer, E. *J. Biol. Chem.* **1949**, *177*, 405-16. (b) Chargaff, E.; Zamenhof, S.; Green, C. *Nature* **1950**, *165*, 756-757.

[7] (a) Pauling, L.; Corey, R. B. *Proc. Nat. Acad. Sci. USA* **1953**, *39*(2), 84-97. (b) Pauling, L.; Corey, R. B. *Nature* **1953**, *171*(4347), 346-346.

[8] Taken from https://x3dna.org/highlights/pauling-triplex-model-of-nucleic-acids-is-available-in-3dna. Accesed June 26, 2021.

[9] Arnott, S. *Trends Biochem. Sci.* **2006**, *31*(6), 349-354.

[10] Rich, A.; Zhang, S. *Nat. Rev. Genet.* **2003**, *4*(7), 566-572.

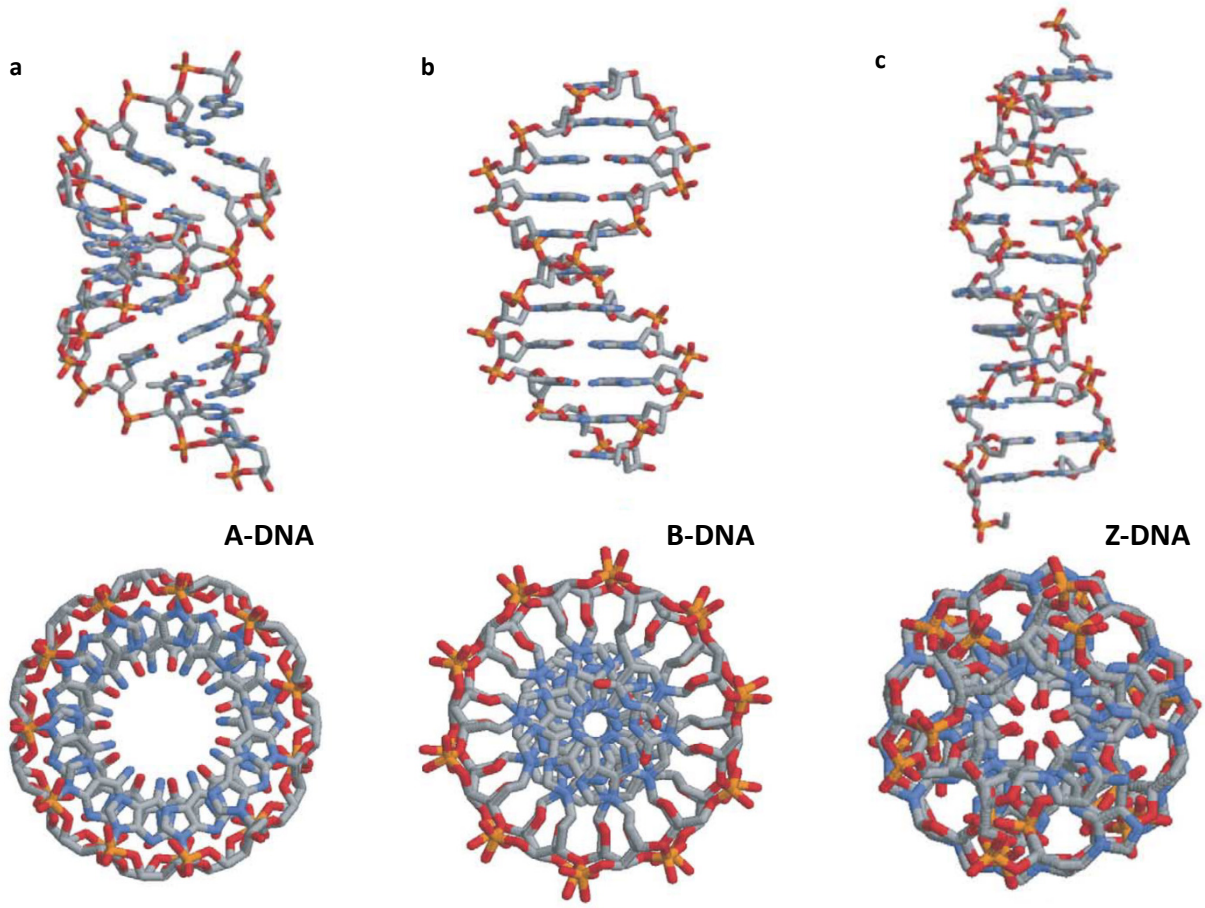


Figure 3. Structures of the different DNA polymorphs. a) A-DNA. b) B- DNA. c) Z-DNA. Each image shows the double helix (top) as well as the top view (bottom), revealing that the phosphates are always on the molecular periphery, while the position of the base pairs changes between DNA structures.[9]

1.1.2. Molecular recognition of DNA

Molecular recognition of DNA is essential to multiple biological processes, such as DNA repair, replication or transcription,[11] which rely on accurate and site-selective interactions between proteins and DNA. These interactions can be electrostatic, hydrophobic, van der Waals forces and hydrogen bonding. The recognition of the interface of DNA can occur at the minor groove, major groove, phosphate backbone, or a combination of all three. Similarly, DNA recognition by small molecules is driven by the same interactions. Therefore, the optimization of the same forces is required, involving the minimization of water-exposed hydrophobic surfaces, maximization of van der Waals interactions and optimization of hydrogen bond pairing. In this regard, the edges of the DNA base pairs in the minor and major groove provide an array of functionality that can be exploited to achieve a strong and selective binding to DNA. Particularly, the exposed hydrogen bond donors (**D**) and acceptors (**A**) represent an attractive target towards achieving a site-specific recognition

[11] Wolberger, C. *Ann. Rev. Biophys. Biomol. Struct.* **1999**, *28*(1), 29–56.

(Figure 4). The relatively flat surface of the major groove offers three hydrogen bonding possibilities in both GC and AT base pairs, with **AAD** and **ADA** patterns, respectively. The minor groove, on the other hand, shows an **ADA** arrangement on the GC base pair, and **AA** on AT (or **AAA**, if the second lone pair of the oxygen atom is considered). Differentiating all the possible base combinations (GC, CG, AT and TA) in the minor groove might be troublesome, as all pairs present the same acceptor groups in the same positions, with little information to distinguish one from another.

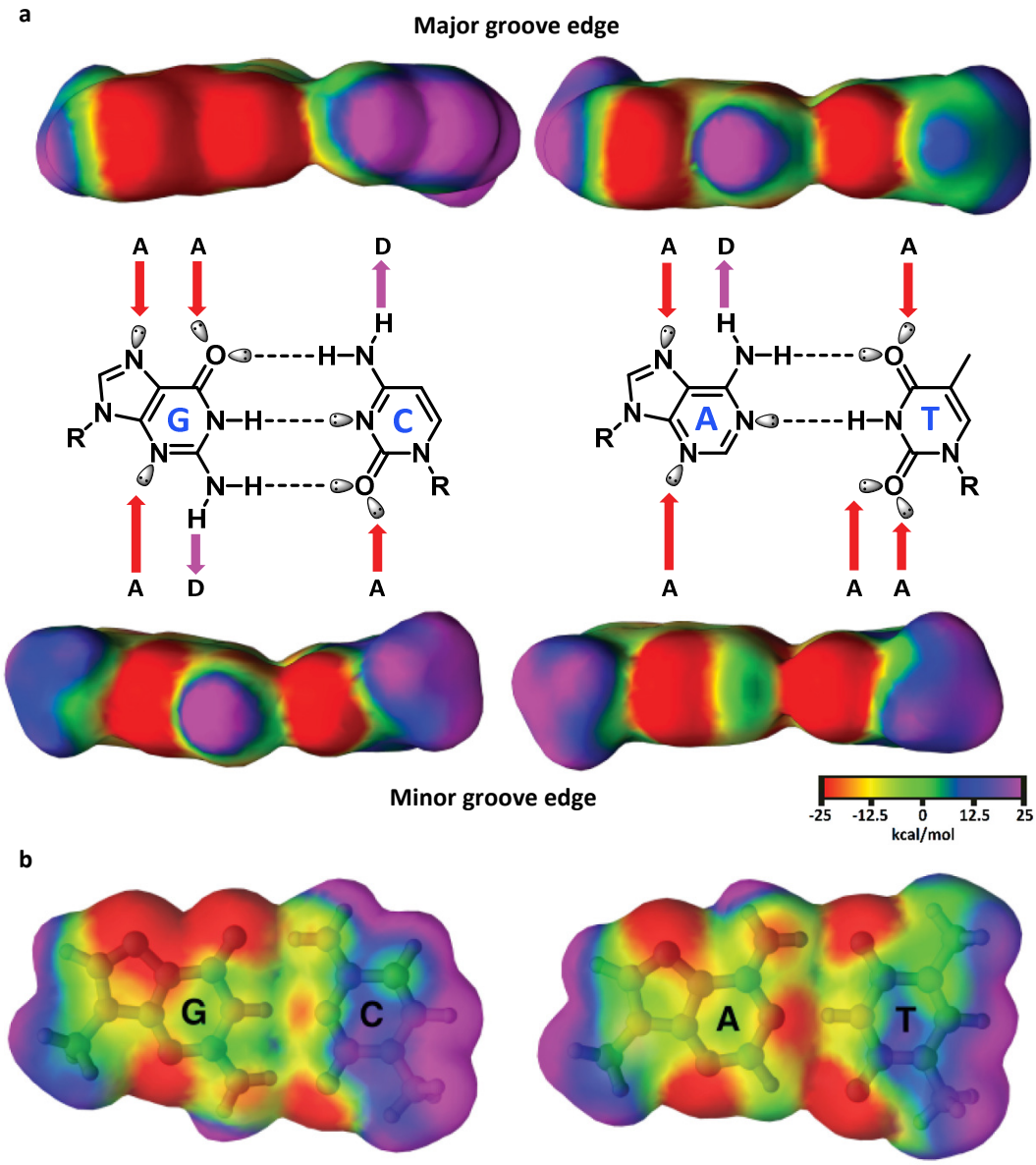


Figure 4. a) Electrostatic potential maps of the base pair edges at the major (top) and minor (bottom) grooves of DNA. Hidrogen bond donors are designated with the letter D and purple arrows, while A and red arrows are used for hydrogen bond acceptors. R represents the DNA backbone. b) Electrostatic potential maps of the top view of the Watson-Crick-paired base pairs.[12]

[12] Chenoweth, D. M., "Synthesys and structural studies of Py-Im polyamides", California, **2009**.

However, some other considerations need to be taken into account besides from the hydrogen bond pattern of the binders when designing groove-specific molecules, as the hydrogen bonding geometry differs significantly between grooves (Figure 5). The narrow minor groove has a very regular and linear arrangement of hydrogen bond donors and acceptors. On the other hand, the broad and open major groove displays binding sites of consecutive base pairs in an irregular fashion, with steric obstacles such as the methyl group of thymine.

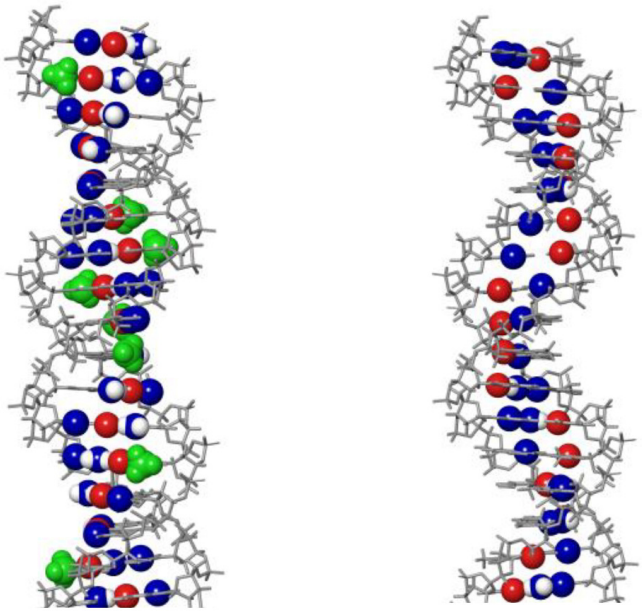


Figure 5. Arrangement of H-bond donors and acceptors in the major (left) and minor (right) grooves (nitrogen: blue, oxygen: red, hydrogen: white, thymine methyl: green).

In terms of molecule design, the major groove provides a richer yet more challenging hydrogen bond fingerprint, while the minor groove, despite being more regular and addressable, lacks differentiating information between base pairs to achieve sequence-selectivity.

1.2. Interfering with DNA function

1.2.1. Genome editing using engineered nucleases

One of the available sequence-selective mechanisms to interact with DNA is the ability to insert, delete or replace DNA fragments at precisely predetermined locations, termed genome editing, which is a type of genetic engineering that holds tremendous value for molecular biology, medicine and biotechnology. Genome editing is achieved by using artificially engineered nucleases or molecular scissors, leading to targeted genetic modifications.

Genome editing begins with DNA specific double-stranded break (DSB) at the target sequence, catalyzed by the engineered nucleases. All eukaryotic cells efficiently repair DSBs via the homology-directed repair (HDR) or non-homologous end joining (NHEJ) pathways.[13]

These highly conserved pathways can be harnessed to generate defined genetic outcomes across a wide range of cell types and species. The error-prone NHEJ repair rapidly and efficiently ligates the two broken ends, with the occasional gain or loss of genetic information; it can therefore be used to introduce small insertions and/or deletions (indels) at the site of the break, an outcome that can be exploited to disrupt a target gene (gene knockout). Alternatively, if an investigator-designed homologous single-strand oligodeoxynucleotides (ssODNs) or donor DNA is provided in combination with the engineered nuclease, information encoded on this template can be used to repair the DSB, thus resulting in gene correction (a few nucleotides changed at the endogenous site) or the addition of a new gene at the site of the break, respectively (Figure 6a). This mechanism is known as HDR.[14]

The repair of two concurrent DSBs induced by engineered nucleases can give rise to chromosomal rearrangements or to structural variations in a targeted manner (Figure 6b,c). Deletions, duplications and inversions of up to a few mega base pairs of chromosomal segments can be achieved in cells and whole organisms, which raises the possibility of correcting genetic defects caused by large chromosomal rearrangements in somatic cells or in stem cells. If DSBs are induced on two different chromosomes, chromosomal translocations can be generated. The possibility of unwanted chromosomal rearrangements that arise from off-target DNA cleavage should be considered when engineered nucleases are used for gene or cell therapy.[15]

[13] (a) Jackson, S.; Bartek, J. *Nature* **2009**, *461*, 1071-1078. (b) Lieber, M. R. *Annu. Rev. Biochem.* **2010**, *79*, 181-211. (c) Moynahan M. E.; Jasin, M. *Nat. Rev. Mol. Cell. Biol.* **2010**, *11*(3), 196-207.

[14] Urnov, D. F.; Rebar, E. J.; Holmes, M. C.; Zhang, H. S.; Gregory, P. D. *Nat. Rev. Genet.* **2010**, *11*, 636-646.

[15] Kim, H.; Kim, J. S. *Nat. Rev. Genet.* **2014**, *15*, 321-334.

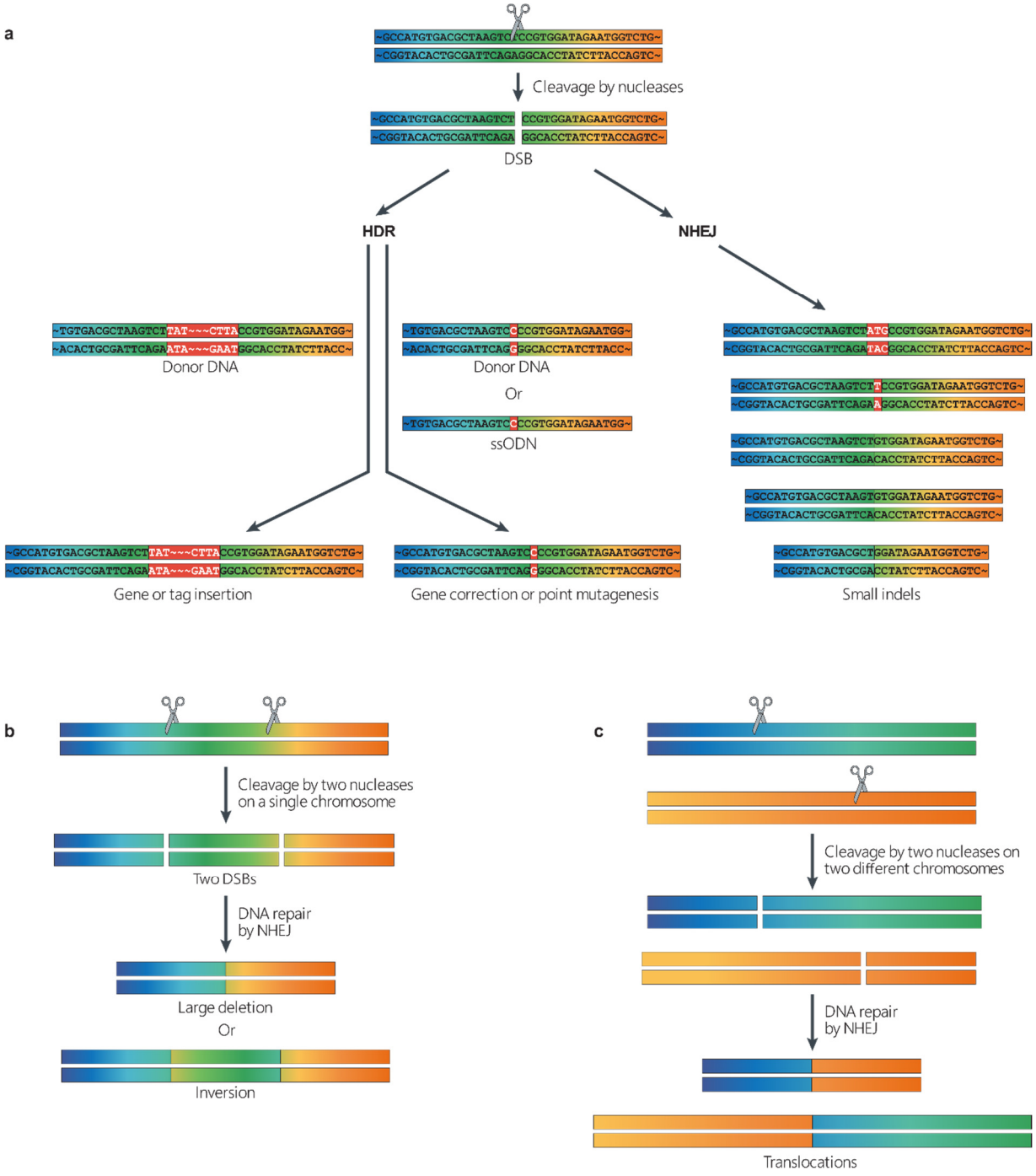


Figure 6. Outcome of genome editing using engineered nucleases. a) Sequence insertion, nucleotide correction or change (red box) through HDR in the presence of a donor DNA or ssODN. NHEJ does not require a template and consequently often leads to small indels. Typical indel sequences and the number of inserted (+3 and +1) or deleted (–2, –4 and –10) bases are shown. b) When two DSBs are generated on a single chromosome by engineered nucleases, the flanking region can be deleted or inverted. c) When two DSBs are generated on two different chromosomes, chromosomal translocations can be induced.[15]

To be useful for genome engineering, an endonuclease must exhibit a combination of qualities: specific recognition of target sequences (ideally, long enough for unique

occurrence in a eukaryotic genome) coupled with sufficient adaptability for retargeting to user-defined sequences.

Four main families of designer engineered nucleases can be distinguished: meganucleases, ZFNs, TALENs and CRISPR-Cas9. Meganucleases will not be discussed further as they have failed to achieve widespread use, being eclipsed by the other three more recent approaches.

1.2.1.1. Zinc Finger Nucleases

Zinc finger nucleases (ZFNs) were developed in the late 1990s and were first used for genome editing in 2003.[16] ZFNs have a modular structure that is composed of two domains: a DNA-binding zinc finger protein (ZFP) domain and a nuclease domain derived from the *Fok*I restriction enzyme.

The *Fok*I type IIS restriction endonuclease (naturally found in *Flavobacterium okeanokoites*) contains a DNA-binding domain and a cleavage domain. Since both domains are structurally separated, it was possible to replace its DNA-binding domain with ZFPs, resulting in a hybrid endonuclease with novel sequence specificities (a ZFN).[17] ZFPs determine the binding specificity of ZFNs, and consist of tandem arrays of zinc fingers. Each zinc finger recognizes a 3-bp DNA sequence, and 3-6 zinc fingers are used to generate a single ZFN subunit that binds to DNA sequences of 9-18 bp, directing the non-specific DNA-cleavage activity of *Fok*I to the right target.

The *Fok*I nuclease domain must dimerize to cleave DNA.[18] Thus, two ZFN monomers are required to form an active nuclease, each monomer binding to adjacent sites separated by spacers of 5-7 bp (Figure 7).This requirement for dimerization doubles the length of recognition sites (from 9-18 bp for monomers to 18-36 bp for dimers), which substantially increases the specificity of ZFNs. However, the *Fok*I domain can still form homodimers to cleave DNA when one monomer binds to DNA, which often leads to unwanted off-target effects. To prevent this, the *Fok*I dimeric interface has been artificially modified to generate obligate heterodimeric forms, which substantially reduces off-target effects and ZFN

[16] (a) Porteus, M. H.; Baltimore, D. *Science* **2003**, *300*(5620), 763. Bibikova, M.; Beumer, K.; Trautman, J. K.; Caroll, D. *Science* **2003**, *300*(5620), 764.

[17] Kim, Y. G.; Cha, J.; Chandrasegaran, S. *Proc. Natl. Acad. Sci. USA* **1996**, *93*(3), 1156-1160.

[18] Bitinaite, J.; Wah, D. A.; Aggarwal, A. K.; Schildkraut I. *Proc. Natl. Acad. Sci. USA* **1998**, *95*(18), 10570-10575.

cytotoxicity.[19] *Fok*I catalytic domain variants with enhanced cleavage activities have also been reported.[20]

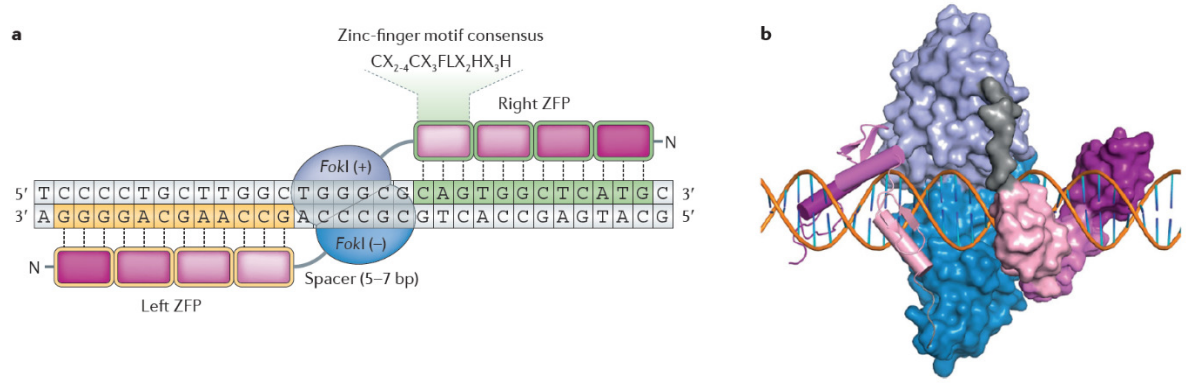


Figure 7. a) Schematic representation of a zinc finger nuclease (ZFN) pair, composed of a zinc finger protein (ZFP) at the N-terminus and the *Fok*I nuclease domain at the C-terminus. In the zinc finger motif consensus, X represents any amino acid.[15] b) Computer model structure of a ZFN pair bound to DNA. Each zinc finger is shown in shades of pink in ribbon (left) and space-filling (right) representations. The grey region represents the linker between the DNA-binding and catalytic domains. The *Fok*I catalytic domains are shown in shades of blue at the center using space-filling representations.[21]

There are several classes of zinc fingers according to the number of amino acids involved in zinc coordination. The Cys2-His2 finger motif is the most abundant DNA binding element in eukaryotes,[22] and is the one used in ZFNs. This motif was first identified in the transcription factor TFIII[23] and has also been found in transcription factors associated with RNA pol I. Among these, Sp1 from HeLa cells[24] and Zif268 from mice[25] are the most intensively studied.[26]

This zinc finger consists of approximately 30 amino acids, arranged as an antiparallel β-sheet, which contains a loop formed by the two cysteines, and an α-helix containing the His-His loop. The two structural units are held together by a Zn ion. The primary contacts with the DNA are made by the α-helix that binds in the major groove through specific hydrogen-bond interactions from amino acids at helical positions -1, 3 and 6 to three successive bases (a triplet) on one strand of the DNA. There is also an important secondary

[19] (a) Miller, J. C. *et al. Nat. Biotechnol.* **2007**, *25*, 778-785. (b) Szczepek, M.; Brondani, V.; Büchel, J.; Serrano, L.; Segal, D. J.; Cathomen, T. *Nat. Biotechnol.* **2007**, *25*, 786-793.

[20] Guo J.; Gaj, T.; Barbas, C. F. *J. Mol. Biol.* **2010**, *400*(1), 96-107.

[21] Carroll, D. *Genetics* **2011**, *188*(4), 773-782.

[22] Tupler, R.; Perini, G.; Green, M. R. *Nature* **2001**, *409*, 832-833.

[23] Hanas, J. S.; Hazuda, D. J.; Bogenhagen, D. F.; Wu, F. W.-H; Wu, C.-W. *J. Biol. Chem.* **1983**, *258*(23), 14120-14125.

[24] Dynan, W. S.; Tijan, R. *Cell* **1983**, *32*(3), 669-680.

[25] Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*(5007), 809-881.

[26] Leon, O.; Roth, M. *Biol. Res.* **2000**, *33*(1), 21-30.

interaction from helical position 2 to the complementary strand (Figure 8). Thus, the binding site for a finger is not simply a triplet of three successive bases, but a 4-bp site, overlapping with that of the preceding finger. There are still other interactions such as with the phosphate backbone of the DNA, but these do not play a direct role in specific recognition.[27]

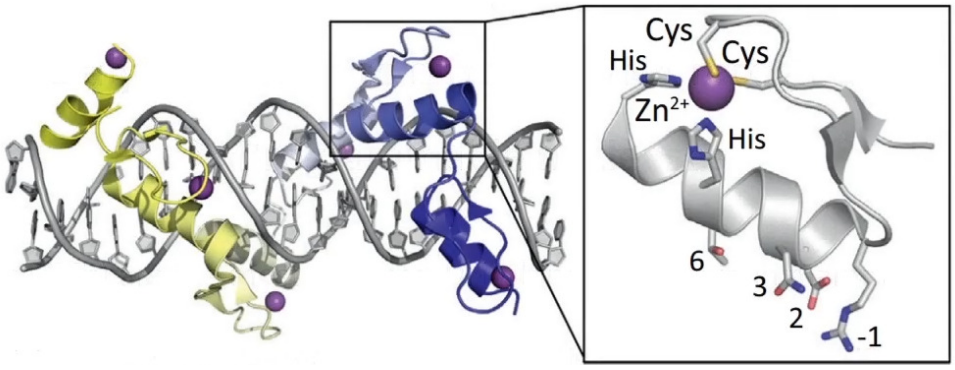


Figure 8. Structure of the Cys2-His2 zinc finger motif and its contact positions with DNA.

The specific amino acids at the positions -1, 3 and 6 determine which nucleotide triplet is targeted by the zinc finger.[28] Modifying these residues would allow building a zinc finger library to recognize all the 64 possible 3-bp triplet combinations. Then, in theory, connecting the right fingers one after the other would afford a ZFP capable of targeting the desired DNA sequence in a specific manner. This approach is known as modular assembly, and although conceptually simple, it neglects positional and context-dependent effects. The low success rate of modular assembly is most likely due to zinc finger domains not being truly independent in their DNA-binding activities. For example, the residue at position 2 makes contacts in the target sites of neighboring fingers, which can negatively affect the overall binding strength. Other context-sensitive selection strategies have been described, but the development of sequence-specific ZFNs is generally slow and costly, which has been the major reason why this technique has fallen out of favor.[29]

Despite these drawbacks, ZFNs have found some clinical applications, especially for the treatment of the Human Immunodeficiency Virus (HIV),[30] but also for some haemoglobinopathies.[31] Many phase I/II trials are now completed, but given the cost and

[27] Klug, A. *Q. Rev. Biophys.* **2010**, *43*(1), 1-21.

[28] Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *3*, 183-212.

[29] Cathomen, T.; Joung, J. K. *Mol. Ther.* **2008**, *16*(7), 1200-1207.

[30] (a) Perez, E. E. *et al. Nat. Biotechnol.* **2008**, *26*, 808-816. (b) Tebas P. *et al. N. Engl. J. Med.* **2014**, *370*, 901-910. (c) Gupta, R. K. *et al. Nature* **2019**, *568*, 244-248.

[31] (a) Moran, K. *et al. Blood* **2018**, *132*(1), 2190. (b) Sangamo Therapeutics, Inc. April 2, 2019. https://investor.sangamo.com/news-releases/news-release-details/sangamo-provides-clinical-development-update-including-early. Accessed June 16, 2021.

time required to reach that point, there is currently little activity to progress ZFN-based therapies toward phase III while long-term follow-up data continues to be gathered.

1.2.1.2. TALENs

In the early 2010s, transcription activator-like effector nucleases (TALENs) rapidly emerged as an alternative to ZFNs for genome editing and introducing targeted DSBs.

The general structural organization of TALENs is similar to that of ZFNs. Like ZFNs, TALENs contain the non-specific *Fok*I nuclease domain, and thus require dimerization to form an active nuclease. However, they use a different class of DNA-binding domain known as transcription activator-like effectors (TALEs), which are derived from the plant-pathogenic *Xanthomonas* spp. bacterium.[32] TALEs are composed of tandem arrays of 33–35 amino acid repeats, each of which recognizes a single base-pair in the major groove. The nucleotide specificity of each repeat domain is determined by two hypervariable residues at positions 12 and 13, which are called repeat variable diresidues (RVDs) (Figure 9a).[33] The only sequence requirement for TALE binding is that each target site be immediately preceded by a 5'-thymine for efficient DNA recognition, although modified proteins have been developed to accept other nucleotides at this position.[34]

Nearly all engineered TALE repeat arrays use four domains that contain the RVDs NN, NI, HD and NG for the recognition of guanine, adenine, cytosine and thymine, respectively. Another repeat with the hypervariable residues NK has been reported to be more specific for guanine than the NN containing repeat (which can also recognize adenine), but TALE repeat arrays using the NK repeats show less activity than those using NN containing repeats.[35]

Co-crystal structures of TALE DNA-binding domains bound to a DNA sequence have shown that individual repeats comprise two-helix v-shaped bundles that stack to form a superhelix around the DNA, and the RVDs at positions 12 and 13 are positioned in the DNA major groove. The residues at position 8 and 12 within the same repeat interact with each other, thereby possibly stabilizing the structure of the domain, whereas the residue at position 13 can make base-specific contacts with the DNA (Figure 9b).[36]

---

[32] Boch, J.; Bonas, U. *Annu. Rev. Phytopathol.* **2010**, *48*, 419-436.

[33] (a) Moscou, M. J.; Bogdanove, A. J. *Science* **2009**, *326*(5959), 1501. (b) Boch J.; Scholze, H.; Schornack, S. *et al. Science* **2009**, *326*(5959), 1509-1512.

[34] (a) Doyle, E. L.; Hummel, A. W.; Demorest, Z. L. *et al. PLoS One* **2013**, *8*(12), e82120. (b) Lamb, B. M.; Mercer, A. C.; Barbas, C. F. *Nucleic Acids Res.* **2013**, *41*(21), 9779-9785.

[35] Streubel, J.; Blücher, C.; Landgraf, A.; Boch, J. *Nat. Biotechnol.* **2012**, *30*(7), 593-595.

[36] (a) Mak, A. N.-S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L. *Science* **2012**, *335*(6069), 716-719. (b) Deng, D.; Yan, C.; Pan, X. *et al. Science* **2012**, *335*(6069), 720-723.

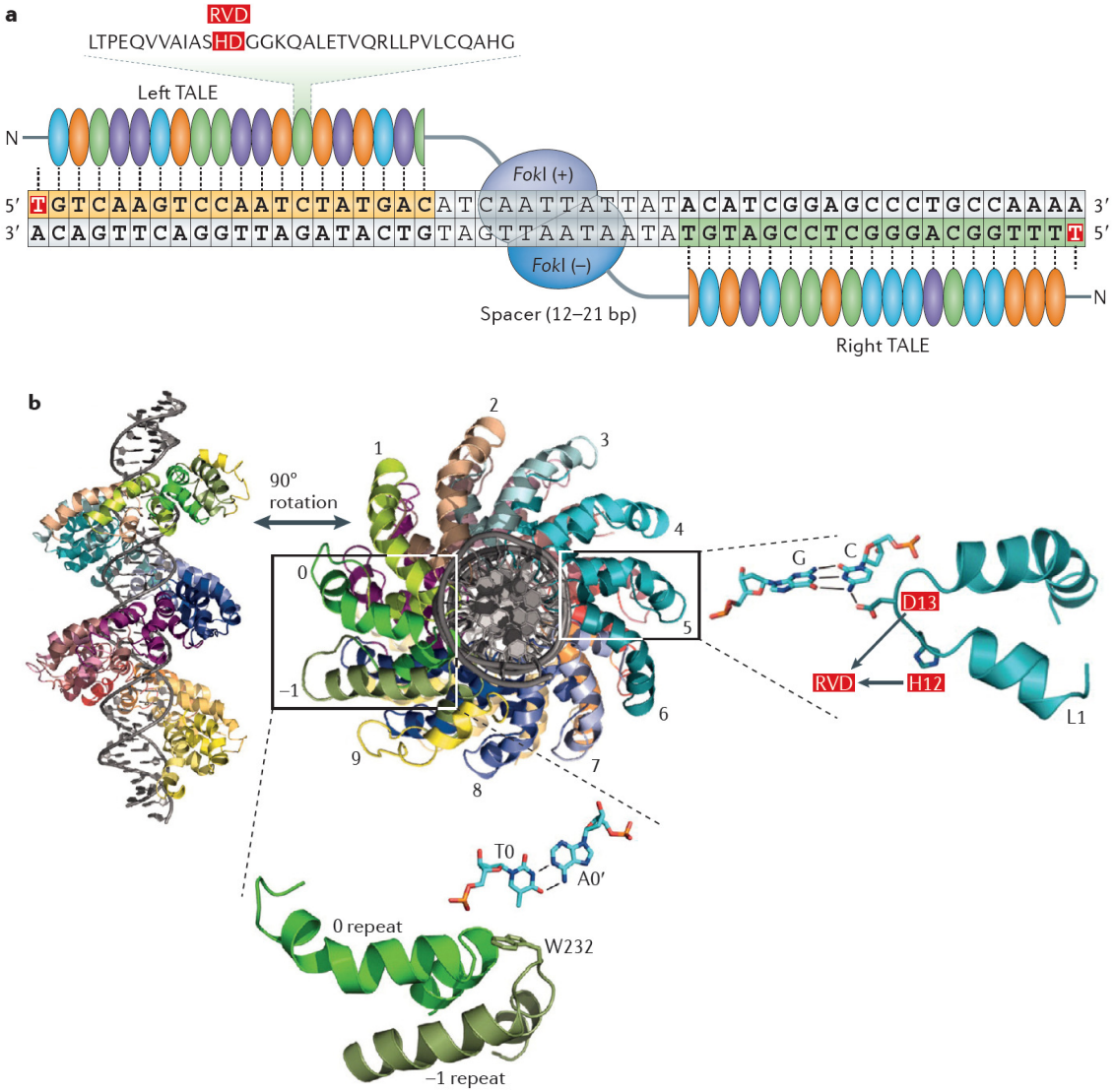


Figure 9. a) Schematic representation of a TALEN pair. Each TALEN is composed of TALEs at the N-terminus and the *Fok*I nuclease domain at the C-terminus. The repeat variable diresidue (RVD) is shown in red.[15] b) In the TALE–DNA co-crystal structure, the RVDs in TALE interact with DNA in the major groove. The amino-terminal repeats (designated as 0 and –1 in the box) contact 5′ thymine.[36a]

Like ZFNs, TALEN-based therapies are also in early clinical development, with some products entering phase I in late 2019. These focus exclusively on immuno-oncology, showing promise in autologous and antibody-mediated immunotherapies.[37]

[37] (a) Benjamin, R.; Graham, C.; Yallop, D.; Jozwik, A. *et al. Blood* **2018**, 132, 896. (b) Cellectis. January 15, 2020. https://www.cellectis.com/en/press/first-patient-dosed-with-cellectis-new-allogeneic-ucart123-product-candidate-for-relapsed-refractory-acute-myeloid-leukemia/. Accessed June 17, 2021.

1.2.1.3. CRISPR-Cas9

The CRISPR-Cas9 system has been in constant evolution since its discovery in late 2012,[38] and represents the state of the art regarding genome editing. In 2020, Jennifer A. Doudna and Emmanuelle Charpentier were jointly awarded the Nobel Prize in Chemistry for their contributions to the discovery and development of this technology.[39]

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. Although the name CRISPR was coined much later,[40] these repeat elements were initially noticed in *Escherichia coli* in 1987.[41] Interestingly, unlike typical tandem repeats in the genome, the CRISPR repeat clusters (23-55 bp) were separated by non-repeating DNA sequences (26-72 bp), which were called spacers. It took more than a decade for researchers to recognize the nature and origin of these spacer sequences. They noticed that the CRISPR sequences are present in more than 40% of sequenced bacteria and 90% of archaea, and that the CRISPR elements are adjacent to multiple well-conserved genes, called CRISPR-associated (Cas) genes. Most interestingly, the non-repeating spacer DNA sequences were recognized to belong to plasmids and phage genomes.[42] All these observations led to hypothesize that the CRISPR sequences were part of a prokaryotic adaptive immune system, which was confirmed experimentally soon after.[43]

There are three distinct phases in the CRISPR immunity process: adaptation, CRISPR RNA (crRNA) biogenesis, and interference (Figure 10). During adaptation, small fragments of foreign nucleic acids are selected, processed, and integrated into the CRISPR array as spacers, to provide a record of infection. The CRISPR array is then transcribed to produce a long precursor crRNA (pre-crRNA) that is processed within the repeat sequences to yield mature crRNAs. Upon subsequent infection, the interference machinery is guided by crRNAs to cleave complementary sequences, termed protospacers, in the foreign nucleic acids, leading to the degradation of the genetic material and, therefore, protection from infection.[44]

---

[38] A PubMed search for publications that had the words "CRISPR" or "Cas9" in the abstract or title showed 24,600 results, 6,000 of them from 2020. Each year the number of publications keeps increasing, illustrating the activity in this field.

[39] The Nobel Foundation. October 7, 2020. https://www.nobelprize.org/prizes/chemistry/2020/press-release/. Accesed June 20, 2021.

[40] Jansen, R.; van Embden, J. D. A.; Gaastra, W.; Schouls, L. M. *Mol. Microbiol.* **2002**, *43*(6), 1565-1575.

[41] Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. *J.Bacteriol.* **1987**, *169*, 5429–5433.

[42] (a) Pourcel, C.; Salvignol, G.; Vergnaud, Vergnaud, G. *Microbiology* **2005**, *151*, 653-663. (b) Mojica, F.J.; Díez-Villaseñor, C.; García-Martínez, J; Soria, E. *J. Mol. Evol.* **2005**, 60, 174-182. (c) Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, D. *Microbiology* **2005**, *151*, 2551-2561.

[43] Barrangou, R.; Fremaux, C.; Deveau, H. *et al. Science* **2007**, *315*(5819), 1709–1712.

[44] Hille, F.; Richter, H.; Wong, S. P.; Bratovič, M.; Ressel, S.; Charpentier, E. *Cell* **2018**, *172*, 1239-1259.

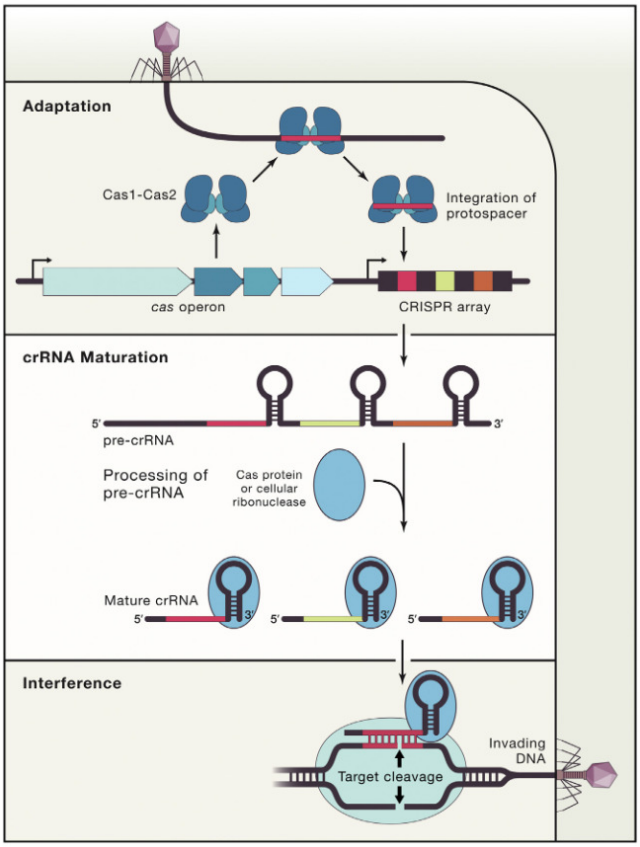


Figure 10. Stages of CRISPR immunity: adaptation, crRNA maturation and interference.[44]

According to the assortment of Cas genes and the nature of the interference complex, CRISPR-Cas systems have been divided in two classes, which are further subdivided into six types and several subtypes. Class 1 CRISPR-Cas systems (types I, III, and IV) employ multi-Cas protein complexes for interference, whereas in class 2 systems (types II, V, and VI), interference is accomplished by a single effector protein.[45] Class 2 type II has become the system of choice for the development of genome editing technologies, due to the simple architecture of the effector complexes using a Cas9 protein.

In addition to crRNA, Cas9 requires trans-activating crRNA (tracrRNA), a small RNA that bears complementarity to the repeat regions of crRNA (Figure 11a).[46] Once bound to mature dual RNA (tracrRNA:crRNA), Cas9 identifies target DNA through the recognition of a small region called the protospacer adjacent motif (PAM). The most commonly-used Cas9 from *Streptococcus pyogenes* recognizes the PAM sequence 5′-NGG-3′ (where "N" can be any nucleotide base), which must immediately precede the target site to be recognized. After a successful Cas9-pairing to the PAM region, the target sequence is checked for complementarity to the guide sequence (20 nt) of crRNA. Alternatively, CRISPR-Cas9 can

[45] Shmakov, S.; Abudayyeh, O. O.; Makarova, K. S. et al. Mol. *Cell* **2015**, *60*(3), 385-297.

[46] Deltcheva, E.; Chylinski, K.; Sharma, C. *et al.* M. *Nature* **2011**, *471*, 602-607.

also be guided by a single chimeric RNA formed by the fusion of tracrRNA and crRNA, called single-guide RNA (sgRNA), which the preferred form for genome engineering applications (Figure 11b).[47]

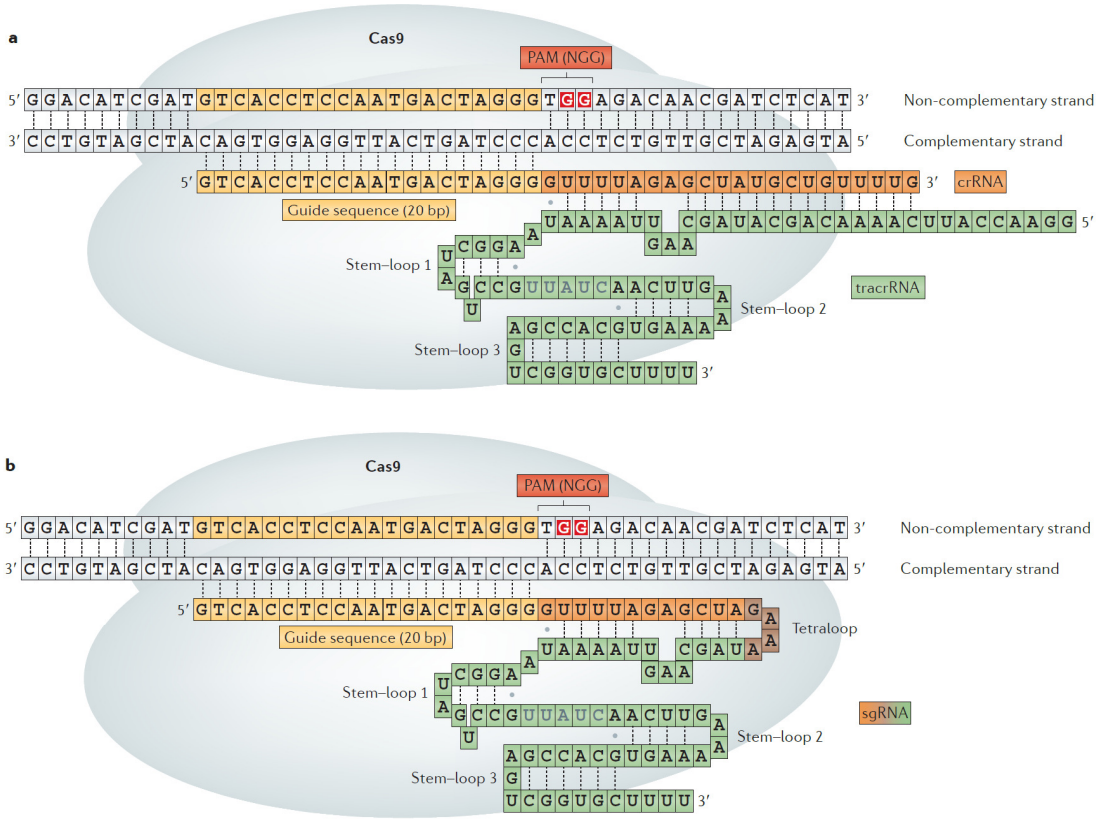


Figure 11. Schematic representation of the CRISPR-Cas9 system bound to the target DNA to be recognized. a) In nature, the crRNA and tracrRNA exist as two distinct units that form a complex. b) The crRNA and tracrRNA can be fused in a single chimeric unit to form a sgRNA, which is more convenient for genome editing applications.[15]

The structure of Cas9 in free form presents two distinct lobes, the recognition (REC) lobe and the nuclease (NUC) lobe, with the latter containing the HNH and RuvC nuclease domains (Figure 12). The Cas9 protein first searches for the PAM sequence, recognizing the GG dinucleotide by base-specific hydrogen-bonding interactions with two arginine residues in a PAM interacting site, which is disordered in the apo-form, but undergoes conformational rearrangements to reach a target-recognition mode after sgRNA binding. Once PAM is found, the flanking DNA is probed for complementarity to the seed (10 nt) of the sgRNA, which is the part of the 20-nt guide sequence that is most critical for complementarity. The interactions between PAM and Cas9/sgRNA lead to destabilization of the adjacent dsDNA, which in turn facilitates for sgRNA to invade the dsDNA. Once a stable RNA–DNA duplex, an R-loop, has been formed, Cas9 is activated for DNA cleavage.

[47] Junek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. *Science* **2012**, *337*(6096), 816-821.

Each of the two nuclease domains cleaves a strand of the target dsDNA at a specific site 3 bp upstream from the 5′-NGG-3′ PAM sequence and, in most cases, the ends that are formed are blunt.[48]

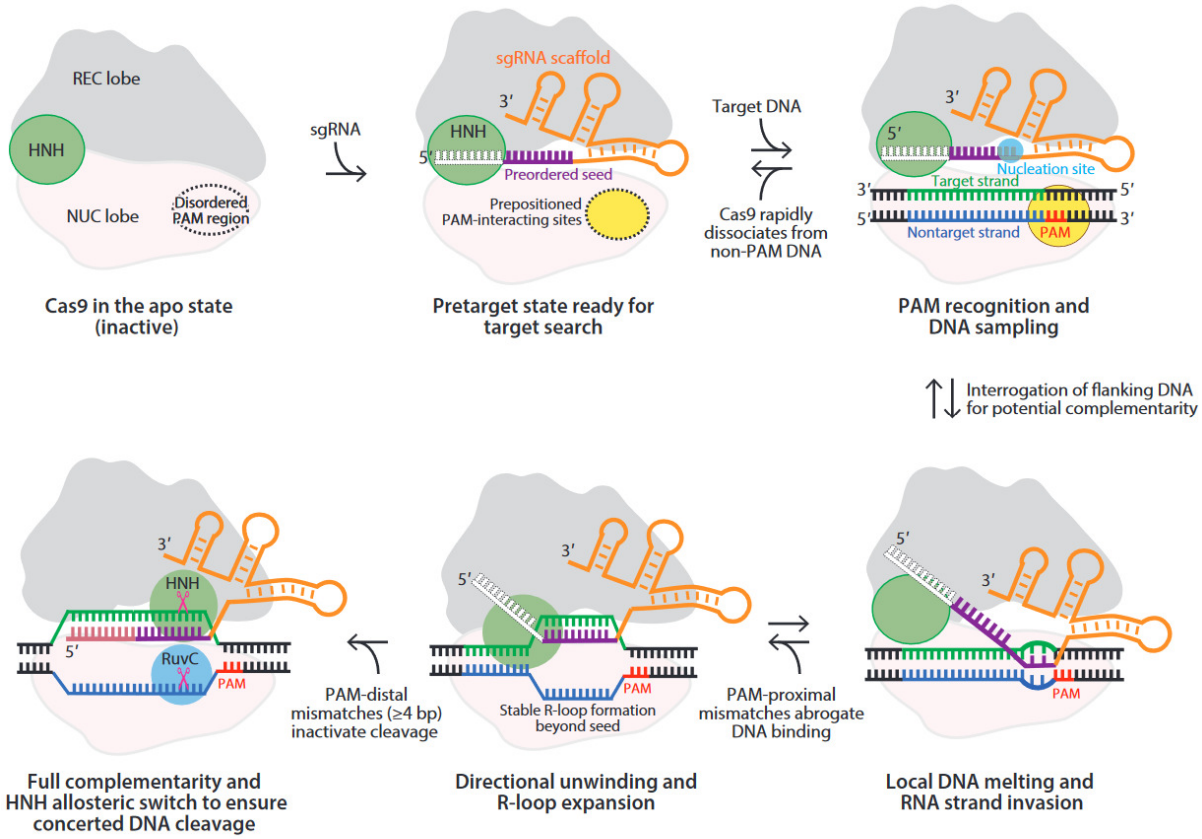


Figure 12. Schematic representation of the proposed mechanisms of CRISPR-Cas9-mediated target DNA recognition and cleavage.[47]

Mutant variants of the Cas9 protein have also found interesting applications. By inactivating one of the two nuclease domains of Cas9, a nickase can be formed, which cleaves only one strand of a DNA duplex. Nickases can be programmed to target opposite strands and thus make staggered cuts within the target DNA. In this way, a Cas9 nickase mutant, combined with a pair of sgRNA molecules, can introduce targeted double-strand breaks with very high sequence specificity.[49] Another variant of the Cas9 protein is the dead Cas9 (dCas9), which is completely devoid of its nuclease activity. The dCas9 protein is a useful tool for molecular biology experiments to regulate gene expression, since the CRISPR-dCas9 system can still bind to the target DNA sequence but is unable to cleave it. This can be used for live intracellular site-specific labeling, by fusing an enhanced green fluorescent protein (EGFP) to dCas9.[50]

---

[48] Jiang, F.; Doudna, J. A. *Annu. Rev. Biophys.* **2017**, *46*, 505-529.

[49] Ran, F. A.; Hsu, P. D.; Lin, C.-Y. *Cell* **2013**, *154*, 1–10.

[50] Chen, B.; Gilbert, L. A.; Cimini, B. A. *et al. Cell* **2013**, *155*, 1479–1491.

As in the case of TALENs, CRISPR therapies are still in phase I clinical trials. CRISPR-based cell therapies mainly focus on the use of autologous T-cells edited using CRISPR-Cas9 to knock out the immune checkpoint inhibitor programmed cell death-1 (PD1) prior to reinfusion back into the patient, for the treatment of different types of cancer.[51]

1.2.1.4. Safety and ethical concerns

The advent of genome editing technology promises to transform human health, livestock and agriculture, and to eradicate pest species. This transformative power demands urgent scrutiny and resolution of the ethical conflicts attached to the creation and release of engineered genomes.[52]

Concerns about these new scientific advancements were especially notorious after it was revealed in November 25, 2018 that He Jiankui, a Chinese scientist working at the Southern University of Science and Technology (SUSTech) in Shenzen, China, had created the first genetically modified human babies, the twins Lulu and Nana, born in October 2018.[53] He presented his research at the Second International Summit on Human Gene Editing, held in Hong Kong in November 27-29, 2018. He also announced that a third genetically modified baby was expected, whose birth was confirmed in 2019, although the exact date is unknown.

He's work, having crossed an internationally agreed upon ban on germline modification in humans, precipitated an international outcry and a call for a global moratorium on reproductive uses of human germline modification.[54] On December 30, 2018, He was sentenced to three years of prison, and was fined three million yuan (US$430,000). He was also banned from ever working with human reproductive technology again, and from applying for Chinese research funding.[55]

In his research, He took the sperm from the HIV-positive father, fertilized the eggs *in vitro* and modified the genes of the early embryos using CRISPR-Cas9 to attempt to mimic a natural deletion of the CCR5 gene, which encodes a protein that HIV employs to enter cells.[56] When disabled, an individual is immune from or greatly resistant to HIV infection.

---

[51] Lu, Y.; Xue, J.; Deng, T. *J. Clin. Oncol.* **2018**, 36, 3050.

[52] Eissenberg, J. C. *Biomol. Concepts* **2021**, *11*, 1-7.

[53] Greeli, H. T. *Journal of Law and the Biosciences* **2019**, *6*(1), 111–183.

[54] Lander, E. S.; Baylis, F.; Zhang, F.; Charpentier, E.; Berg, P.; Bourgain, C.; …; Winnacker, E.-L. *Nature* **2019**, *567*(7747), 165–168.

[55] Nature. January 3, 2020. https://www.nature.com/articles/d41586-020-00001-y. Accessed June 16, 2021.

[56] de Silva, E.; Stumpf M. P. H. *FEMS Microbiol. Lett.* **2004**, *241*(1), 1–12.

The germline editing He claimed to have carried out differs significantly from the somatic gene therapies that are currently changing the frontiers of medicine. To understand the implications of genome editing, it is important to note the differences between the two.

Somatic therapies target the genome of stem cells for the affected tissue of a patient. The edited gene is contained only in the target cell type, so no other types of cells are affected. Since these changes, including potential off-target effects, are limited to the treated individuals, the edited gene is not passed down to future generations. Somatic cell therapies have been researched and tested for over 20 years and are highly regulated

Germline modifications, on the other hand, are performed on gametes or early embryos, so that any change is copied into all the new cells (including sperm and eggs). Thus, the changes not only affect the treated individuals, but also their descendants. Although genome editing using embryos has successfully been performed in a variety of animal models,[57] human germline editing had never been reported before He's work.

Some people resist reproductive genetic modification, claiming that it leads to so-called "designer babies." They worry that people will use this technology not just to modify away genetic disorders and diseases, but to create children with the desired hair and eye color, enhanced intelligence or athletic abilities. Related to this are concerns about a new eugenics, employing the ability to eradicate individuals seen as bearing inferior traits. Others raise concerns about the fairness of access to these new technologies, and whether the gap between rich and poor can be further aggravated.[58]

Safety concerns are also a matter of discussion. Although novel technologies such as CRISPR-Cas9 are hailed as precise, they can make off-target, unintended changes to the genome. Safety issues include preventing genetic mosaicism (in which only some of the cells are modified), off-target cuts, and on-target cuts with unforeseeable bad results.[59]

In 2019, the World Health Organization (WHO) committee announced the creation of a global registry to track research in the field for national, local and international authorities to ensure human genome editing science progresses within agreed ethical boundaries.[60] It is indisputable that safety issues and the heritability of germline changes present new legal and ethical considerations that need to be addressed.

---

[57] Li, Q.; Qin, Z.; Wang, Q.; Xu, T.; Yang, Y.; He, Z. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 689–698.

[58] Rulli, T. *Bioethics* **2019**, *33*(9), 1072-1082.

[59] Bosley, K. S.; Botchan, M.; Bredenoord, A. L.; Carroll, D.; Charo, R. A.; Charpentier, E.; …; Zhou, Q. *Nat. Biotechnol.* **2015**, *33*, 478-496.

[60] Reuters. March 20, 2019. https://www.reuters.com/article/us-health-who-gene-editing/who-panel-calls-for-registry-of-all-human-gene-editing-research-idUSKCN1R02IC. Accessed June 16, 2021.

1.2.2. Gene modulation

Another pathway to control DNA function is the modulation of gene transcription into mRNA or the translation of this into proteins, employing antigene or antisense agents, respectively.[61] This method does not alter the nucleotide sequence of the target, but blocks its function in a reversible fashion.

1.2.2.1. Triplex forming oligonucleotides

Triplex forming oligonucleotides (TFOs), the classical antigene agents, are short single-stranded oligonucleotides (15-25 nt) that bind to the major groove of DNA, interacting with polypurine sequences through Hoogsteen base-pairing and forming triple helices (Figure 13b). These non-canonical DNA triplex structures were first reported in 1957.[62]

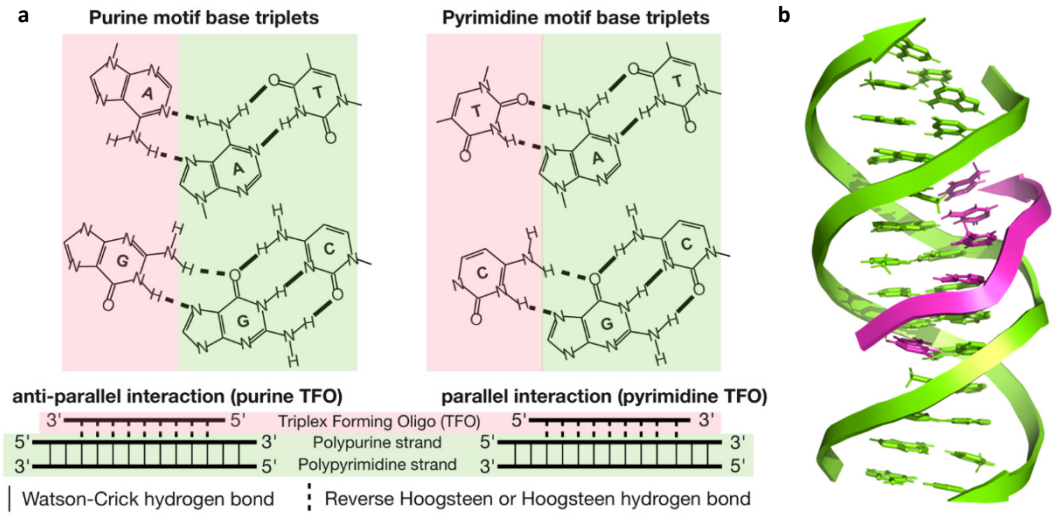


Figure 13. a) Schematic representation of the Hoogsteen base-pairing for the purine (left) and pyrimidine motifs (right). The DNA duplex and the TFO are highlighted in green and red, respectively. Watson-Crick hydrogen bonds are shown as continuous lines, while Hoogsteen hydrogen bonds as discontinuous lines.[63] b) A polypyrimidine TFO (pink) bound with a parallel orientation to the purine-rich strand of double-stranded (ds) DNA (green).[64]

Three different triplex-forming motifs have been described: (i) the T/U, C pyrimidine-rich TFO exhibits a parallel (forward Hoogsteen) alignment with respect to the polypurine strand orientation (Figure 13a, right);[65] (ii) G, A purine-rich sequences form anti-parallel (reverse Hoogsteen) bonds with respect to the polypurine strand (Figure 13a, left); and (iii)

[61] (a) Rakozcy, P. E. *Methods Mol. Med.* **2001**, *47*, 89-104. (b) Buchini, S; Leumann, C. J. *Curr. Opin. Chem. Biol.* **2003**, *7*(6), 717-726.

[62] Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024.

[63] Maldonado, R.; Filarsky, M.; Grummt, I.; Längst, G. *RNA* **2018**, *24*, 371-380.

[64] Taken from RCSB PDB. NDB: 1BWG.

[65] Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*(1), 63-80.

the G, T/U purine-pyrimidine TFO can adopt both parallel and anti-parallel binding configurations.[66]

As the three strands involved in Hoogstein base-pairing are negatively charged, due to the phosphate groups of the backbone, electrostatic repulsion forces represent the main obstacle to triplex formation. For this reason, larger sequences or the presence of divalent cations are essential to form stable complexes. Neutralization of charge repulsion is typically provided experimentally by levels of Mg^{2+} (5-10 mM)[67] that are much higher than the levels in cells.[68] Also, pyrimidine TFOs require the protonation of cytosine at N3 not only for the creation of the necessary hydrogen bond, but also to further stabilize the triplex.[69] However, protonation occurs at a relatively acidic pH (pK_a = 4.4),[70] which is far from the average physiological pH value of 7.40. Furthermore, triplex formation involves conformational changes and some distortion on the underlying DNA duplex.[71]

Although some modified TFOs have partially overcome these limitations, their inability to recognize mixed purine/pyrimidine sequences has limited their widespread use as antigene agents.

1.2.2.2. Peptide Nucleic Acids

Peptide Nucleic Acids (PNAs) were created in 1991 by Peter E. Nielsen *et al.* at the University of Copenhagen.[72] They sought to develop a DNA analog with a polyamide backbone that resembled the morphology of DNA with respect to the number of backbone bonds and distance between the backbone and nucleobases, to be able to recognize dsDNA. Similarly to TFOs, single-stranded PNAs bind to dsDNA forming a triplex structure through Hoogsteen base pairing in the major groove. Therefore, the recognition ability of PNAs, as in the case of TFOs, is also limited to polypurine sequences, which is well-ilustrated in Figure 13 (Section 1.2.2.1.).

PNAs are composed of *N*-(2-aminoethyl)-glycine (aeg) units linked by peptide bonds. The nucleobases are attached to this backbone via a methylene carbonyl linkage, rather than the traditional sugar moiety found in DNA or RNA (Figure 14).

---

[66] Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*(4999), 1360–1363.

[67] Blume, S. W.; Lebowitz, J.; Zacharias, W. *et al. Nucleic Acids Res.* **1999**, *27*(2), 695–702.

[68] Pesco, J.; Salmon, J.-M.; Vigo, J.; Viallet, P. *Anal. Biochem.* **2001**, *290*, 221–231.

[69] Asensio, J. L.; Lane, A. N.; Dhesi, J.; Bergqvist, S.; Brown, T. *J. Mol. Biol.* **1998**, *275*, 811-822.

[70] Verdolino, V.; Cammi, R.; Munk, B. H.; Schegel, H. B. *J. Phys. Chem. B* **2008**, *112*, 16860–16873.

[71] Esguerra, M.; Nilsson, L.; Villa, A. *Nucleic Acids Res.* **2014**, *42*(18), 11329–11338.

[72] Nielsen, P.; Egholm, M.; Berg, R.; Buchardt, O. *Science* **1991**, *254*(5037), 1497–1500.

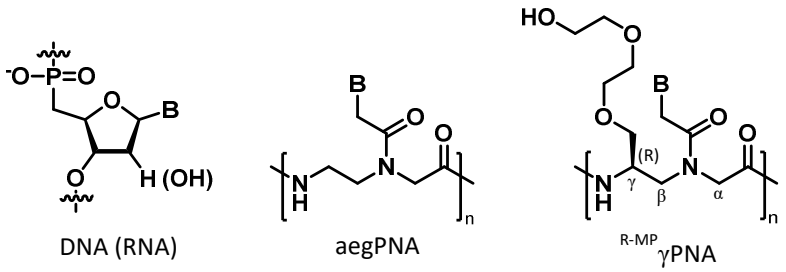


Figure 14. Chemical structures of DNA, unmodified aegPNA and Mini-PEG-modified PNA at the γ position. B represent the nucleobase.

This pseudopeptide backbone, due to its charge-neutral nature, allows for a stronger hybridization with complementary nucleic acids compared to TFOs,[73] which suffer from electrostatic repulsions between the phosphate groups of the backbone. The unnatural polyamide backbone of PNAs also renders them resistant to enzymatic degradation.[74]

Depending on the design of PNAs, different binding modes can occur. A dsDNA-PNA triplex is typically observed with cytosine rich PNAs. Polypurine PNAs, on the other hand, favor duplex invasion, where the PNA binds to one strand to the DNA through Watson-Crick base-pairing (Figure 15). PNA-DNA tends to be more stable than DNA-DNA, but it is less tolerant of base pair mismatches.[75] Triplex invasion, where a strand of DNA binds to two PNAs by both Hoogsteen and Watson-Crick hydrogen bonds, is observed for polypyrimidine PNAs.[76]

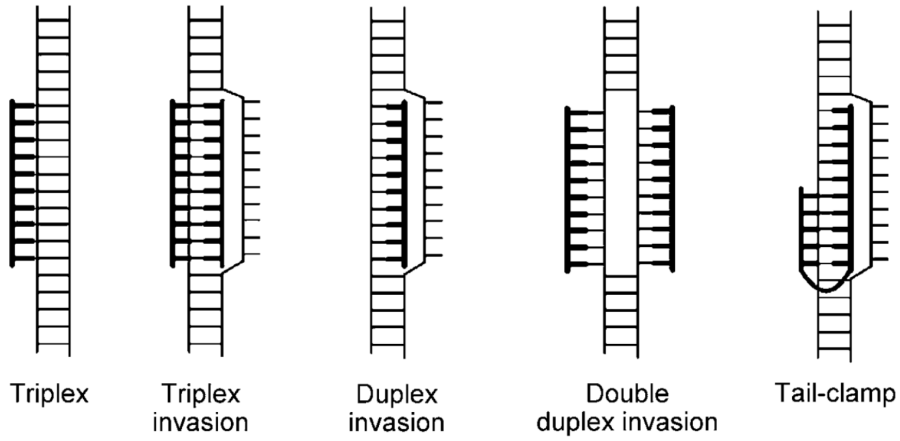


Figure 15. Schematic representation of the known dsDNA-PNA complexes. PNA is represented with thick black lines.[76]

[73] (a) Nielsen, P. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 167-183. (b) Ricciardi, A. S.; McNeer, N. A.; Anandalingam, K. K.; Saltzman, W. M.; Glazer, P. M. *Methods Mol. Biol.* **2014**, *1176*, 89-106.

[74] Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskil, M. D.; Egholm, M.; Buchard, O.; Sönnichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994**, *48*(6), 1310–1313.

[75] Nielsen, P. E.; Egholm, M. *Curr. Issues Mol. Biol.* **1999**, *1*(2), 89-104.

[76] Nielsen, P. E. *ChemBioChem* **2010**, *11*, 2073-2076.

PNAs have mainly been explored as antigene and antisense agents,[77] but they have also shown promise for gene editing.[78] Due to their stability, charge neutrality, and ease of design, PNAs have proven to be versatile molecules with a variety of applications. Furthermore, their ability to bind to target DNA or RNA by both Watson-Crick and Hoogsteen base pairing rules further enhances their therapeutic potential. However, their broad application in a clinical setting is still restricted by a number of limitations.

Unmodified aegPNAs suffer from low solubility, with a tendency to form aggregates.[79] PNAs have also shown poor cell permeability, indicating that passive diffusion of unmodified PNAs is not a suitable way of transport into biological cells.[80] This, combined with the fact that PNAs have poor bioavailability *in vivo*, being quickly (within 10-30 min in mice) excreted through the kidneys,[81] explains the slow progress in the therapeutic application of these molecules.

Various solutions have been developed to overcome this drawbacks. PNAs modified with lysine residues,[82] guanidinium functional groups,[83] and several other backbone modifications,[84] have shown improved solubility and cellular uptake, superior hybridization and enhanced discrimination for base pair mismatches in target sequences. Among them, it is worth mentioning the diethylene glycol (miniPEG) chiral substitution at the γ position (Figure 14), which not only drastically improved the solubility, but also pre-organized PNAs into a right-handed helix, increasing their affinity and specificity for DNA. This modification, however, was designed to bind solely through Watson-Crick base pairing rules, resulting in a duplex binding motif.[85] Cellular delivery has been improved with the use of polymeric nanoparticles as carriers, among other delivery systems.[86]

---

[77] (a) Janowski, B. A.; Kaihatsu, K.; Huffman, K. E. *et al. Nature Chem. Biol.* **2005**, *1*(4), 210-215. (b) Knudsen, H.; Nielsen, P. E. *Nucleic Acids Res.* **1996**, *24*, 494–500.

[78] Rogers, F. A.; Vasquez, K. M.; Egholm, M.; Glazer, P. M. *Proc. Nat. Acad. Sci. USA* **2002**, *99*(26), 16695-16700.

[79] Braasch, D. A.; Corey, D. R. *Methods* **2001**, *23*(2), 97-107.

[80] Wittung, P.; Kajanus, J.; Edwards, K.; Nielsen, P.; Nordén, B.; Malmström, B. G. *FEBS Letters* **1995**, 365, 27-29.

[81] (a) McMahon, B. M.; Mays, D.; Lipsky, J.; Stewart, J. A.; Fauq, A.; Richelson, E. *Antisense Nucleic Acid Drug Dev.* **2002**, *12*(2), 65-70. (b) Nielsen, P. E. *Q. Rev. Biophys.* **2005**, *38*(4), 345-350.

[82] Nielsen, P. E.; Haaima, G.; Lohse, A.; Buchardt, O. *Angew. Chem. Int. Ed.* **1996**, *35*(17), 1939-1942.

[83] Zhou, P.; Wang, M.; Du, L.; Fisher, G. W.; Waggoner, A.; Ly, D. H. *J. Am. Chem. Soc.* **2003**, 125, 6878-6879.

[84] Moccia, M.; Adamo, M. F. A.; Saviano, M. *Artif. DNA PNA XNA* **2014**, *5*(3), e1107176.

[85] (a) Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. *J. Org. Chem.* **2011**, *76*, 5614-5627. (b) Quijano, E.; Bahal, R.; Ricciardi, A.; Saltzman, W. M.; Glazer, P. M. *Yale J. Biol. Med.* **2017**, *90*(4), 583-598.

[86] (a) Park, J.-K.; Utsumi, T.; Seo, Y.-E.; Deng, Y.; Satoh, A.; Saltzman, W. M.; Iwakiri, Y. *Nanomedicine* **2016**, *12*(5), 1365-1374. (b) Deng, Y.; Yang, F.; Cocco, E.; Song, E. *et al. Proc. Nat. Acad. Sci. USA* **2016**, *113*(41), 11453-11458. (c) Gupta, A.; Bahal, R.; Gupta, M.; Glazer, P. M.; Saltzman, W. M. *J. Control. Release* **2016**, *240*, 302-311.

1.2.2.3. Minor groove binding polyamides

Peter B. Dervan is a Professor of Chemistry at the California Institute of Technology (Caltech), and he is known for his studies at the interface of chemistry and biology. He is an international authority and a pioneer in the field of sequence-specific DNA recognition with small organic molecules, the most important being pyrrole-imidazole (Py-Im) polyamides.

Py-Im polyamides were developed based on the natural antibiotics netropsin and distamycin A,[87] which bind in the middle of the minor groove of dsDNA at sites of 4-5 successive AT base pairs. They contain two and three Py moieties, respectively (Figure 16). Sequence-specificity is obtained through a number of hydrogen bonds between the amide hydrogens of the *N*-methylpyrrolecarboxamides and the N3 of adenine and O2 of thymine on the floor of the minor groove (Figure 17a). Netropsin preferentially binds with a 1:1 ligand-DNA stoichiometry, while distamicyn A can show 1:1 or 2:1 complexes.[88]

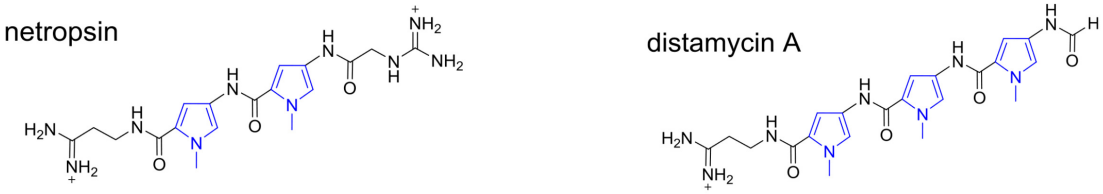


Figure 16. Structure of the natural antibiotics netropsin and distamycin A.

In 1992, Dervan expanded this concept by introducing modifications to be able to recognize sites other than AT-rich sequences. For this purpose, he introduced the imidazole (Im) ring, which generated an additional hydrogen bond between the exocyclic amino group on guanine and the N3 on imidazole, allowing for GC/CG recognition. They demonstrated that the netropsin analogue ImPyPyDp (Dp = dimethylaminopropylamide) recognized and bound to 5'-TGTCA-3' as a 2:1 ligand-DNA complex (Figure 17b). The polyamide formed an antiparallel homodimer in the minor groove of dsDNA.[89]

[87] (a) Finlay, A.C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. *J. Am. Chem. Soc.* **1951**, *73*, 341-343. (b) Arcamone, F.; Penco, S.; Orezzi, P.; Nicolella, V.; Pirelli, A. *Nature* **1964**, *203*, 1064-1065.

[88] (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Nat. Acad. Sci. USA* **1985**, *82*(5), 1376-1380. (b) Coll, M.; Frederick, C. A.; Wang, A. H.; Rich, A. *Proc. Nat. Acad. Sci. USA* **1987**, *82*(23), 8385-8389. (c) Pelton, J. G.; Wemmer, D. E. *Proc. Nat. Acad. Sci. USA* **1989**, *86*(15), 5723-5727.

[89] (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*(23), 8783–8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Nat. Acad. Sci. USA* **1992**, *89*(16), 7586–7590.

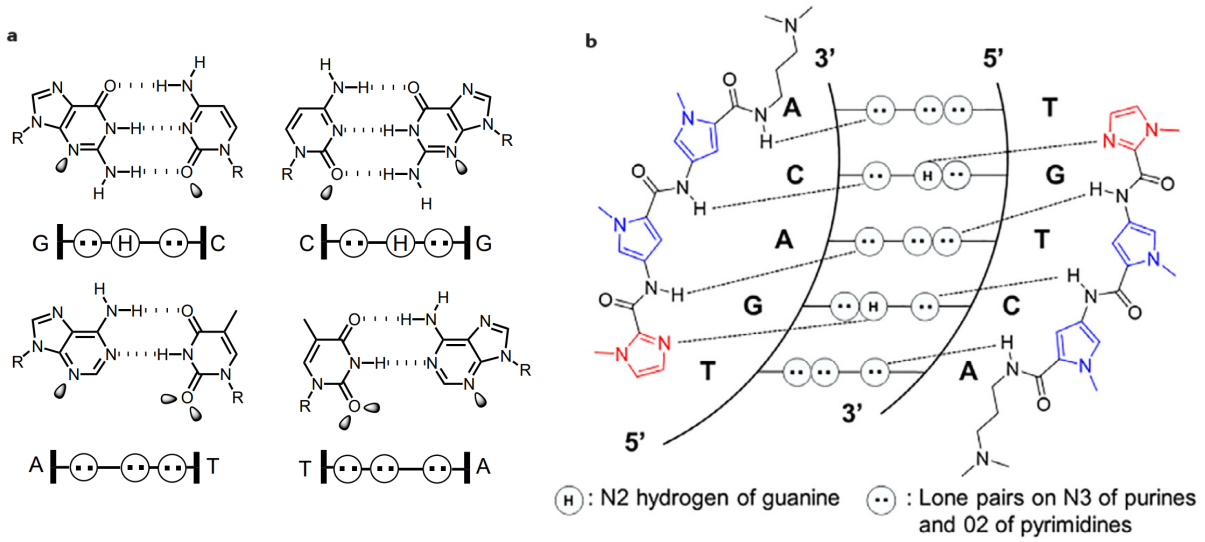


Figure 17. a) Minor groove hydrogen-bonding patterns of Watson-Crick base pairs. b) Molecular recognition of the minor groove of dsDNA by a netropsin derivative ImPyPyDp.

Importantly, the study established the fundamental pairing rules for DNA recognition with precisely oriented pairs of pyrrole and imidazole polyamides: the Im-Py pair recognizes GC base pairs, Py-Im recognizes CG, and Py-Py targets AT or TA. A hydroxypyrrole (Hp) ring, which was found to be selective for T, was used to create Py-Hp and Hp-Py pairs to recognize AT and TA, respectively. However, despite the increased selectivity, the binding affinities slightly decreased and the use of the Hp ring was limited due to its instability.[90]

In order to align Py-Im dimeric polyamides in a predictable way, Dervan introduced the hairpin motif, where the C-terminus of one polyamide and the N-terminus of the other moiety are covalently connected by a γ-aminobutyric acid (γ-turn) (Figure 18).[91] In this way, the aromatic ring pairs in a polyamide oligomer are unambiguously paired when folded in the minor groove, enhancing 100 fold their binding affinity. Eight-ring hairpin Py-Im polyamides were shown to bind 6-bp target sequences with equilibrium association constants of 10^8-10^{10} M^{-1}, which is comparable to those of DNA binding proteins.[92] Adding an (R)-amino substitution at the α position of the γ-turn residue enhanced tenfold the binding affinity, reaching 10^{11} M^{-1}.[93] Several other linking motifs were explored beyond the hairpin, such as the H-pin,[94] U-pin[95] and cycles.[96]

[90] Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*(3), 284–299.

[91] Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983–7988.

[92] Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.

[93] Wang, C. C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 653-657.

[94] Olenyuk, B.; Jitianu, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 4741–4751.

[95] Heckel, A.; Dervan, P. B. *Chem. Eur. J.* **2003**, *9*, 3353-3366.

[96] Melander, C.; Herman, D. M.; Dervan, P. D. *Chem. Eur. J.* **2000**, *6*, 4487-4497.

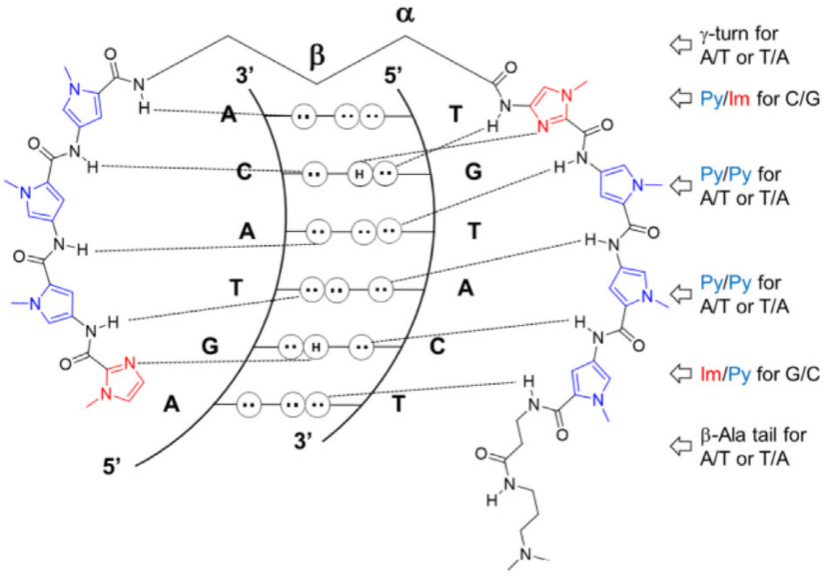


Figure 18. Molecular recognition of dsDNA by the hairpin ImPyPyPy-γ-ImPyPyPy-β-Dp polyamide.[97]

Hairpin polyamides were found to disrupt the binding of transcription factors to DNA, which was unexpected, as most of these bind in the major groove, while the polyamides occupy the minor groove.[98] However, a high resolution x-ray crystal structure of a cyclic polyamide bound to dsDNA showed a minor groove widening of up to 4 Å, with a simultaneous compression of the major groove and bending of the DNA helix towards the major groove by >15° (Figure 19).

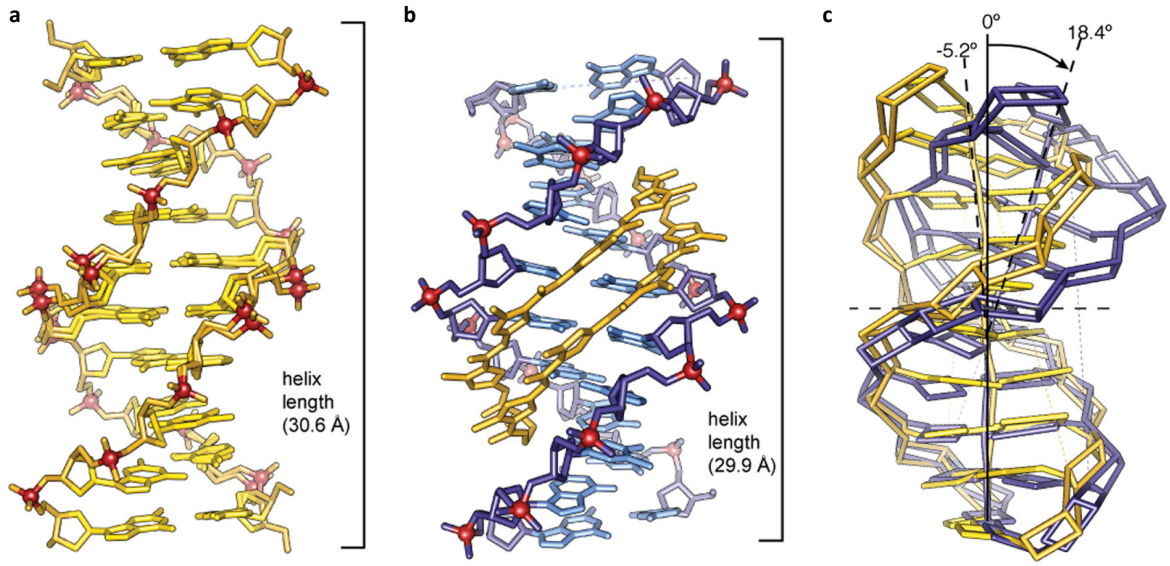


Figure 19. Comparison of native DNA to polyamide/DNA complex. a) Native DNA crystal structure at 0.98Å resolution. b) polyamide/DNA co-crystal structure at 1.18 Å resolution, showing a widened minor groove. c) Overlay of the calculated geometric helix model from each structure, showing a DNA bend of 18° in the polyamide/DNA complex (blue) compared to native DNA (yellow).[98]

[97] Kawamoto, Y.; Bando, T.; Sugiyama, H. *Bioorg. Med. Chem.* **2018**, *26*, 1393–1411.

[98] Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*(6629), 202–205.

This structural alteration enforces a major groove surface geometry incompatible with transcription factor-DNA binding.[99] Cell-permeable cyclic polyamides were shown to regulate the expression of androgen receptor target genes in cell culture studies.[100]

The use of Py-Im polyamides is however limited to the recognition of up to 7-bp sequences, since oligomers of more than four contiguous ring pairs no longer match the curvature of the DNA helix.[101] However, some modified polyamides have overcome this limitation. Two hairpin polyamides linked by a valeric acid linker, a tandem hairpin, can recognize 11-bp sequences with binding affinities of 10^{12} M^{-1}.[102] Also, some polyamides modified with flexible β-alanine residues have been reported to recognize up to 16 bp, forming partially overlapping homodimers.[103]

1.2.2.4. Major groove binding extended nucleobases

As previously discussed, DNA recognition with TFOs and PNAs using natural nucleobases is limited to polypurine sequences, since purines offer two hydrogen bonds possibilities in the Hoogsteen side. Relatively long polypurine target sequences are needed, which may only be interrupted by a few pyrimidines.[104] High stability triplex formation for the recognition of T or C pyrimidine sequences still poses a challenge, as they only present one hydrogen bond acceptor or donor, respectively, in the major groove.

A promising way to enhance binding affinities is to design extended unnatural nucleobases that form hydrogen bonds along the entire Hoogsteen side of the base-pairs. This approach offers the possibility of creating up to three simultaneous hydrogen bonds, potentially overcoming the above mentioned limitations. Dervan *et al.* pioneered the first design, where an artificial nucleobase (D_3) formed two hydrogen bonds with C and G (Figure 20).[105] Other groups presented similar designs shortly after.[106] Nevertheless, these early efforts shared common problems, such as reduced affinities compared to native triplexes or lack of pyrimidine discrimination.[107]

---

[99] Chenoweth, D. M.; Dervan, P. B. *Proc. Nat. Acad. Sci. USA* **2009**, *106*(32), 13175–13179.

[100] Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2009**, *131*, 7182-7188.

[101] Dervan, P. B. *Isr. J. Chem.* **2019**, *59*, 1–14.

[102] Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1999**, 5, 975-983.

[103] Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534–3535.

[104] Hari, Y.; Obika, S.; Imanishi, T. *Eur. J. Org. Chem.* **2012**, *2012*, 2875-2887.

[105] Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.

[106] (a) Huang, C. Y.; Miller, P. S. *J. Am. Chem. Soc.* **1993**, *115*, 10456-10457. (b) Huang, C.-Y.; Bi, G.; Miller, P. S. *Nucleic Acids Res.* **1996**, *24*, 2606-2613.

[107] Purwanto, M. G. M.; Weisz, K. *Curr. Org. Chem.***2003**, *7*, 427-446.

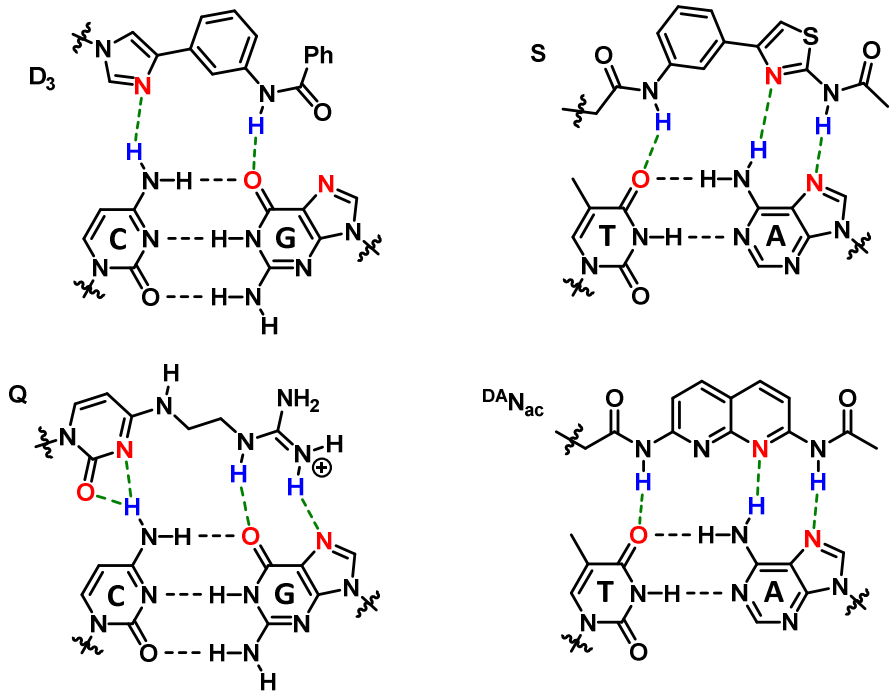


Figure 20. Examples of extended unnatural nucleobases. CG binders D_3 and Q, and TA binders S and $^{DA}N_{ac}$. Hydrogen bond donors are colored in blue, and acceptors in red. The bond with a curved line represents the attachment to the backbone.

Other studies evaluated the binding capabilities of the extended unnatural nucleobases with NMR experiments in organic solvents.[108] However, while these results offer valuable preliminary information to further modify and improve the binders, they might not translate into a stable and selective triplex. Other factors such as backbone constraints and nucleobase stacking need to be taken into account.[109]

In recent years, other promising candidates have emerged, such as Q for CG recognition and S and $^{DA}N_{ac}$ for TA recognition (Figure 20),[110] but the designs are still far from optimal. Some reports also replace the conventional sugar-phosphate backbone with artificial analogues, such as PNA.[111]

Triplex formation with oligonucleotides bearing unnatural nucleobases shows a great potential for the sequence-selective recognition of dsDNA in the major groove. Nonetheless, this approach is still in its early stages with much room for improvement, and the design of novel nucleobases remains an active area of research.

---

[108] (a) Zimmerman, S. C.; Schmitt, P. J. Am. Chem. Soc. **1995**, *117*, 10769-10770. (b) Mertz, E.; Mattei, S.; Zimmerman, S. C. *Org. Lett.* **2000**, *2*, 2931-2934. (c) Lengeler, D.; Weisz, K. Tetrahedron Lett. **2001**, *42*, 1479-1481.

[109] Mertz, E.; Mattei, S.; Zimmerman, S. C *Bioorg. Med. Chem.* **2004**, *12*(6), 1517-1526.

[110] (a) Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A. *et al. Biochemistry* **2010**, *49*, 7867-7878. (b) Guianvarc'h, D.; Benhida, R.; Fourrey, J.-L.; Maurisse, R.; Sun,J.-S. *Chem. Commun.* **2001**, 1814-1815. (c) Ohkubo, A.; Yamada, K.; Ito, Y.; Yoshimura, K. *et al. Nucleic Acids Res.* **2015**, *43*, 5675-5686.

[111] Kumpina, I.; Brodyagin, N.; MacKay, J. A.; Kennedy, S. D.; Katkevics, M.; Rozners, E. *J. Org. Chem.* **2019**, *84*, 13276-13298.

2. Aim of this work

Our research in the topic of DNA recognition focuses on the design and synthesis of antigene agents for the sequence-selective molecular recognition of dsDNA in the major groove. Previous attempts included major groove recognition along the DNA axis by dimeric calixarenes with an oligoamide spacer. However, selectivities remained small, and calculations demonstrated that a single oligoamide is too slim to occupy a specific segment inside the wide major groove.

Therefore, we turned our attention to the design of small molecules to recognize the major groove edge of each base pair individually, perpendicular to the DNA axis. Several biaryl binders were designed and subjected to molecular mechanics (MM) calculations by Max S. Peters. Finally, two promising AT and GC binder candidates were selected for further construction, each of them capable of making three linear hydrogen-bond contacts with the base pairs while maintaining a planar geometry (Figure 21). Preliminary NMR experiments with the 1st generation AT binder showed affinities of up to 117 M^{-1} at room temperature, which is comparable with other binders reported in literature.[112]

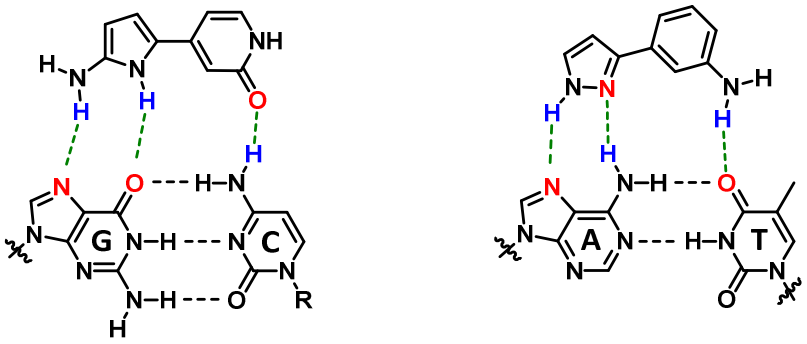


Figure 21. 1st generation 4-pyridonoaminopyrrole GC binder and 3-anilinopyrazole AT binder. Hydrogen bond donors are colored in blue, and acceptors in red. The bond with a curved line represents the attachment to the backbone.

Motivated by the promising preliminary affinities exhibited by the 1st generation binders, we decided to design a second, improved generation of binders. MM calculations conducted by Peters led to new molecules featuring a more planar, robust structure, due to the annulation of a 6-membered pyridone ring to a 5 membered heterocycle (Figure 22). These binders, termed 2nd generation, match perfectly the hydrogen-bonding pattern of the major groove edge of base pairs, each creating three contacts. Furthermore, we sought to create a modular system for the sequence-selective recognition of DNA. For that purpose, we envisioned linking these individual binders to a backbone, creating an oligomer that would

[112] Wang, W.; Purwanto, M. G. M.; Weisz, K. *Org. Biomol. Chem.* **2004**, *2*, 1194-1198.

bind to dsDNA, forming a triplex (Figure 23). This way of action is analogous to that of TFOs, but without polypurine sequence limitations (Section 1.2.2.1.).

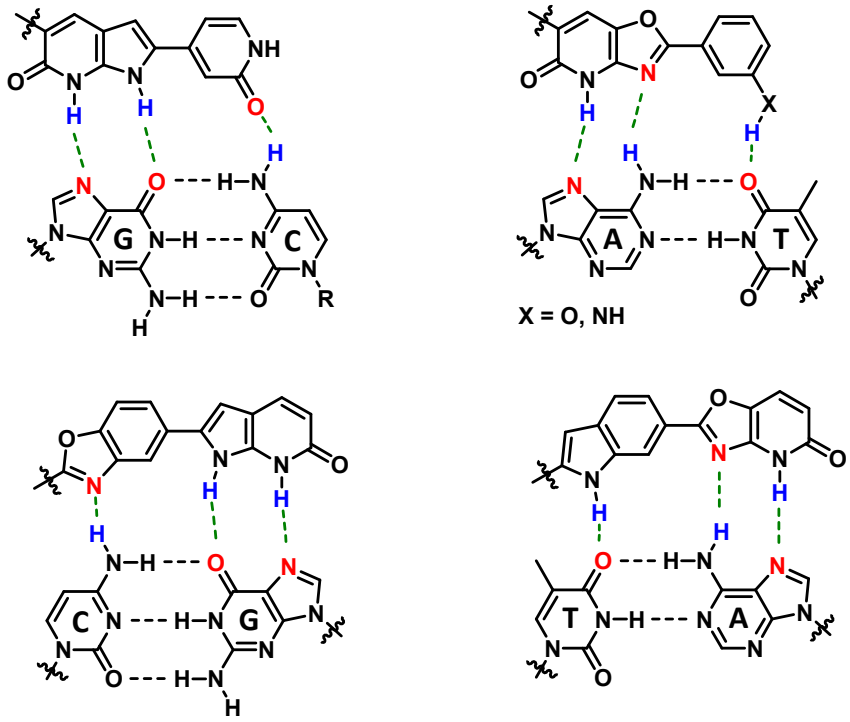


Figure 22. 2nd generation binders for all the possible base-pair combinations (GC, AT, CG and TA). Hydrogen bond donors are colored in blue, and acceptors in red. The bond with a curved line represents the attachment to the backbone.

Taking into account the accumulated knowledge in the field of DNA recognition described in Section 1.2.2., we decided to use the charge-neutral PNA backbone, as it would avoid electrostatic repulsions between phosphates groups, which leads to decreased affinities with the classical sugar-phosphate backbone TFOs. Analogous to the original PNA concept, only a methylene carbonyl linker separates the binder from the PNA backbone, eliminating any additional degree of rotational freedom (Figure 23b). Moreover, the extended aromatic surface of the binders leads to an increased pi-stacking.

For the challenging recognition of the inverted CG and TA binders, an extended linker containing a triple bond was designed, which not only generates the optimal distance to place the binders in the correct position, but also gracefully avoids any steric clash with the methyl group from thymine in the case of the TA binder.

A space-filling model reveals another key feature of this complexes: the major groove is completely filled, indicating a favorable compact packing of all base pair binders (Figure 23a). This resembles the naturally occurring DNA-protein complexes, which also tend to fill the major groove.

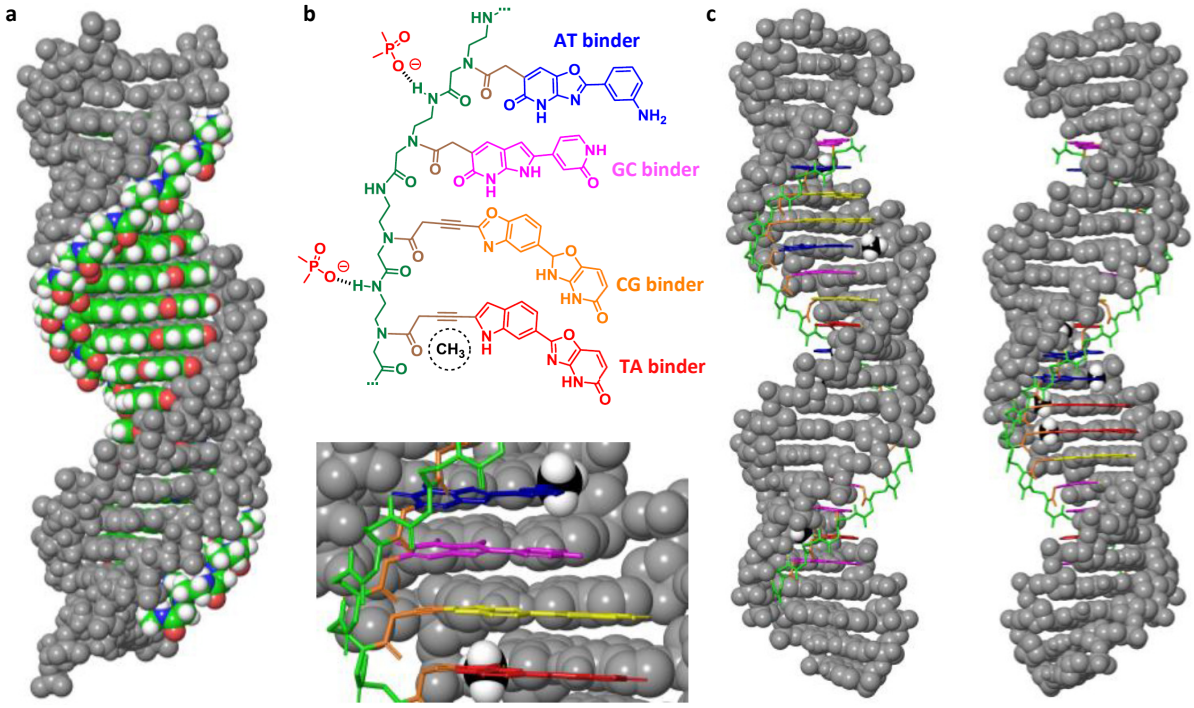


Figure 23. a) Space filling representation of a 2nd generation oligomer binding to a poly(TA) dsDNA. b) Schematic representation of a 2nd generation PNA oligomer, showing additional contacts to the phosphates of the DNA backbone (top). Notice how the methyl group from thymine is avoided by the TA binder (red, bottom). c) Front and back view of a 17mer 2nd generation PNA oligonucleotide bound to a randomized DNA sequence (MacroModel 10.2, OPLS-2005, water/GBA, 50,000 steps). AT binder: blue, GC binder: pink, CG binder: orange, TA binder: red, thymine methyl: black and white.

In order to improve the water-solubility of the oligomers and circumvent well-known issues with unmodified aegPNA scaffolds (Section 1.2.2.2.), the introduction of an aminomethylene group at the γ-carbon of each PNA monomer unit was envisaged (Figure 24).

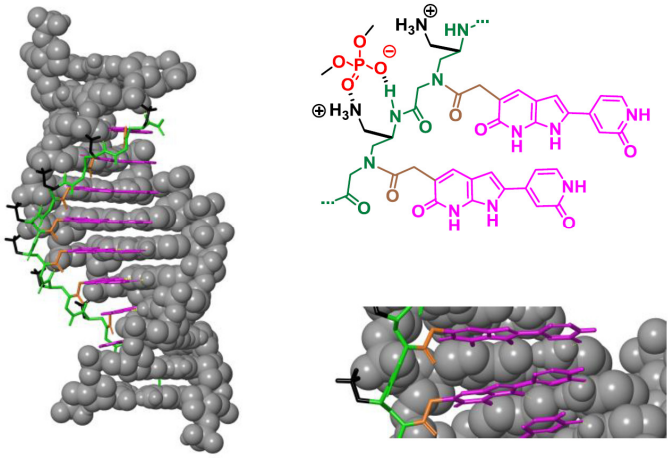


Figure 24. PNA oligomer with an additional protonated aminomethylene substituent on the PNA γ-carbon, binding a poly(GC) dsDNA. This modification leads to increased water solubility with further stabilization by an additional phosphate-ammonium ion pair. The aminomethylene substituent is represented in black, and the GC binder in pink.

This PNA derivative is also reported to improve cell membrane permeability and to increase helical stability.[113] Calculations between the $^{am}\gamma$-PNA hybrids and dsDNA suggest the formation of a phosphate-ammonium ion pair, further stabilizing the complex.

Initially, the isolated base pair binders will be synthesized, and their preliminary binding capabilities will be evaluated with NMR titrations. Then, the modified binders possessing a linker will be attached to the different PNA scaffolds, several oligomers will be constructed and their binding affinity and sequence-selectivity will be tested with techniques such as DAPI displacement assays, fluorescent titrations, measurement of melting points and circular dichroism (CD).

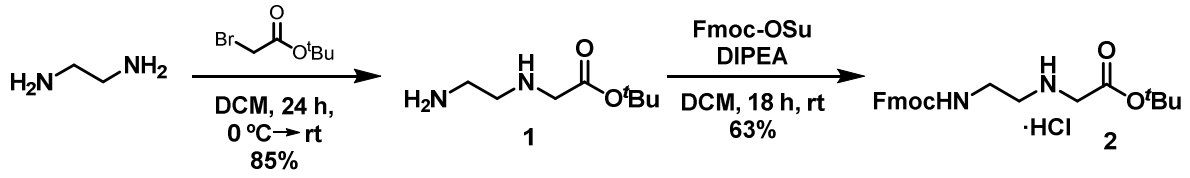
[113] Mitra, R.; Ganesh, K. N *J. Org. Chem.* **2012**, *77*, 5696-5704.

3. Results and discussion

3.1. Design and synthesis

3.1.1. PNA derivatives

PNA constitutes the charge-neutral backbone of our oligomers. As a preliminary approach, the traditional aegPNA scaffold was chosen, due to its synthetic simplicity and stability. Once the methodologies have been established, synthetically more demanding γ-substituted PNAs will be prepared, which are reported to have superior solubilizing properties.[113] The first PNA to be synthesized was the classical Fmoc-aeg-O^{t}Bu **2** reported by Thomson *et al.* (Scheme 1).[114] **R** Ethylenediamine was directly mono-alkylated with *tert*-butyl bromoacetate to afford **1**, which was subsequently Fmoc-protected and isolated as the hydrochloryde salt **2**.



Scheme 1. Synthetic pathway to obtain the Fmoc-aeg-O^{t}Bu **2**.

This PNA backbone bears two orthogonally addressable protecting groups: a base-labile Fmoc and an acid-labile *tert*-butyl. While this is necessary to ensure a successful peptide formation employing standard Fmoc peptide chemistry, it also limits the functionalization scope of the binders, since any acid-labile group would also be cleaved during the process. Therefore, a benzyl-protected PNA backbone was envisioned, which can be cleaved under hydrogenolytic conditions. This offers more flexibility towards the construction of the binders, allowing the presence of acid-labile functionalities.

This derivative was synthesized according to a procedure developed by Heemstra *et al.*,[115] affording Fmoc-aeg-OBn **5** with a significantly higher operational simplicity than previously reported methodologies, which proved to be difficult to reproduce.[116]

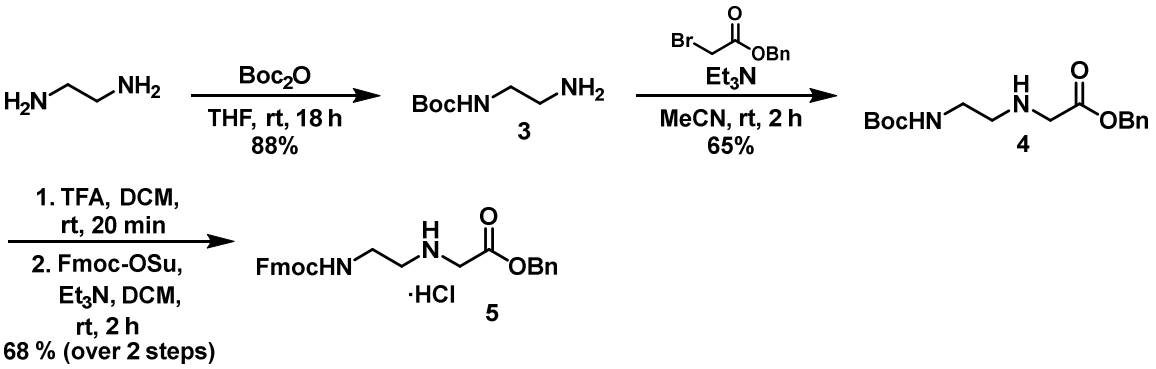
First, ethylenediamine was mono-protected with boc anhydride, obtaining **3,** followed by its alkylation with benzyl bromoacetate to afford **4** (Scheme 2). Unlike with the *tert*-butyl counterpart described above (Scheme 1), this additional boc-protection step was essential, since the authors reported that the direct alkylation of ethylenediamine only afforded benzyl

[114] Thomson, S. A.; Josey, J. A.; Cadilla, R. *et al. Tetrahedron* **1995**, *51*(22), 6179-6194.

[115] Feagin, T. A.; Shah, N. I.; Heemstra, J. M. *J. Nucleic Acids* **2012**, 354549.

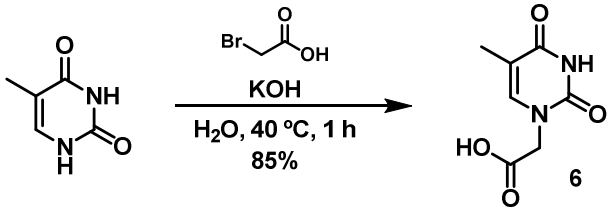
[116] Wojciechowski, F.; Hudson, R. H. E. *J. Org. Chem.* **2008**, *73*, 3807-3816.

alcohol, potentially product of unwanted intramolecular cyclizations. Finally, the boc group was cleaved and subsequently Fmoc-protected, and an additional treatment with a solution of HCl in diethyl ether furnished the hydrochloride salt **5**.



Scheme 2. Synthetic route towards the Fmoc-aeg-OBn **5**.

With the different PNA backbones in hand, the stage was set for the establishment of a method towards the construction of peptide-like structures using standard peptide chemistry. To this end, it was necessary to find a molecule to occupy the position where the unnatural binders synthesized in this project would be attached, acting as a replacement. Thymine was selected for this task, as its chemistry is well established and it is the only one of the four canonical DNA nucleobases that does not require the protection of any functional groups, making the attachment of the acetic acid linker a simple one-step process. Thus, thymine was alkylated with bromoacetic acid under basic conditions to afford the linker-attached thymine **6** (Scheme 3).[117]

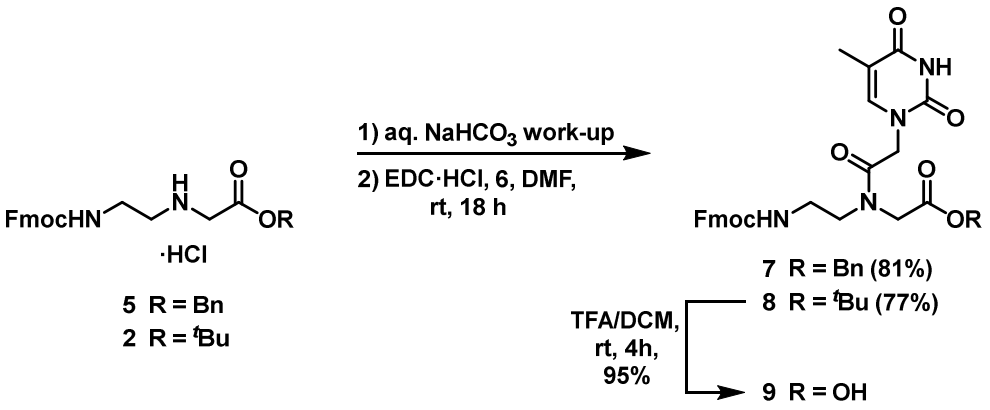


Scheme 3. Synthesis of the linker-attached thymine derivative **6**.

To obtain the different PNA monomers, **5** and **2** were first washed with a saturated aqueous $NaHCO_3$ solution to remove the hydrochloride salt and obtain the free amine, which was subsequently subjected to an EDC-mediated coupling with the thymine derivative **6**, yielding **7** and **8**, respectively (Scheme 4).[114] No significant differences in the yield and purity of both derivatives could be observed. Finally, the ester group had to be cleaved to furnish the free carboxylic acid. The *tert*-butyl-protected PNA monomer **8** was selected, as it can be easily hydrolyzed under mild acidic conditions and precipitated out of the reaction

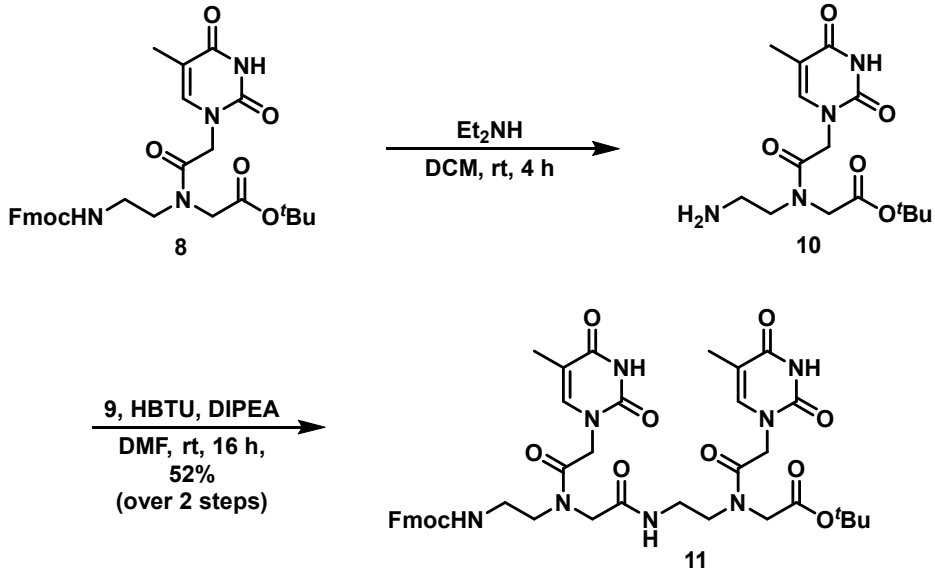
[117] Kosynkina, L.; Wang, W.; Liang, T. C. *Tetrahedron Lett.* **1994**, *35*(29), 5173-5176.

mixture, eliminating the need for purification. Thus, a treatment of **8** with diluted trifluoroacetic acid provided **9** with an excellent yield.



Scheme 4. Synthesis of the PNA monomer building block **9**.

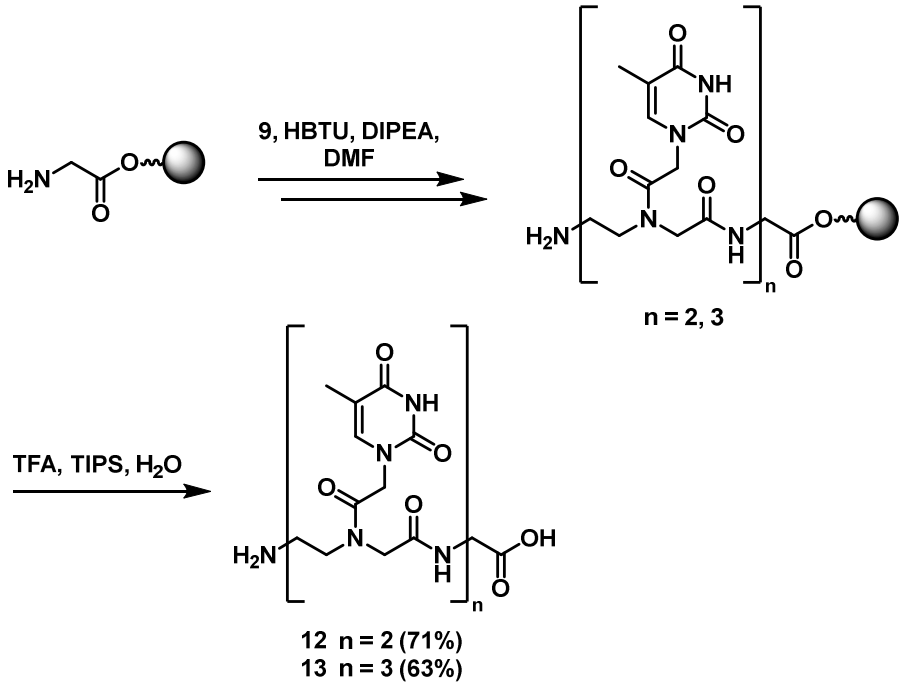
Once the thymine PNA monomer **9** was prepared, different ways for the construction of a peptide were explored. First, solution-phase peptide synthesis was briefly considered. To test this approach, first the dimer **11** was synthesized, cleaving first the Fmoc of a unit of **8** under basic conditions, and subsequently coupling it to a unit of **9**, without isolating **10** or removing the fulvene-diethylamine adduct resulting from the Fmoc deprotection (Scheme 5). Despite successfully obtaining the dimer **11**, the process was tedious and the purification challenging. Some procedures for the continuous synthesis of peptides in solution have been reported, were an aqueous buffer work-up between couplings removes side products and unreacted reagents.[118] However, no further attempts were made towards the establishment of solution phase-methodologies due to their operational complexity.



Scheme 5. Solution-phase synthesis of the dimer **11**.

[118] Carpino, L. A.; Ghassemi, S.; Ionescu, D. *Org. Proc. Res. Dev.* **2003**, *7*, 28-37.

Next, standard solid-phase peptide synthesis (SPPS) was tested. A Wang resin preloaded with a glycine residue was used as support for the construction of the dimer **12** and the trimer **13** (Scheme 6). Using standard Fmoc SPPS conditions, both compounds were obtained with a purities above 98% and with good yields, evidencing the reproducibility of this method.



Scheme 6. SPPS of the dimer **12** and trimer **13**. The gray spheres represent the resin support.

With a general method for the SPPS of PNA oligomers established, the next step was to synthesize oligomers bearing our unnatural binders.

3.1.2. AT binder

3.1.2.1. Isolated binder

The isolated 2nd generation AT binder represent the first synthetic challenge. The isolated binder will not only serve to determine its preliminary binding strength with NMR titrations, but also to establish synthetic methodologies towards the construction of the remaining binders, since all of them share similar scaffolds. In the case of the AT binder, three main sections with synthetic interest can be distinguished: a pyridone ring, an oxazole ring and an Ar-Ar bond between the oxazole and a phenyl ring (Figure 25). The pyridone and the oxazole are fused together, forming an oxazolopyridone scaffold which is not very frequent in the literature. To achieve the desired hydrogen bond pattern, either an amino or a hydroxy group can be used as a hydrogen bond donor, both of which performed equally well in MM calculations.

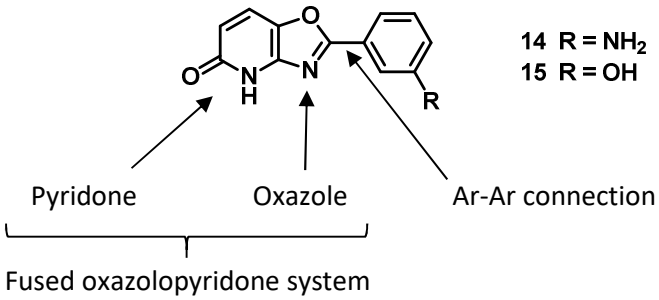
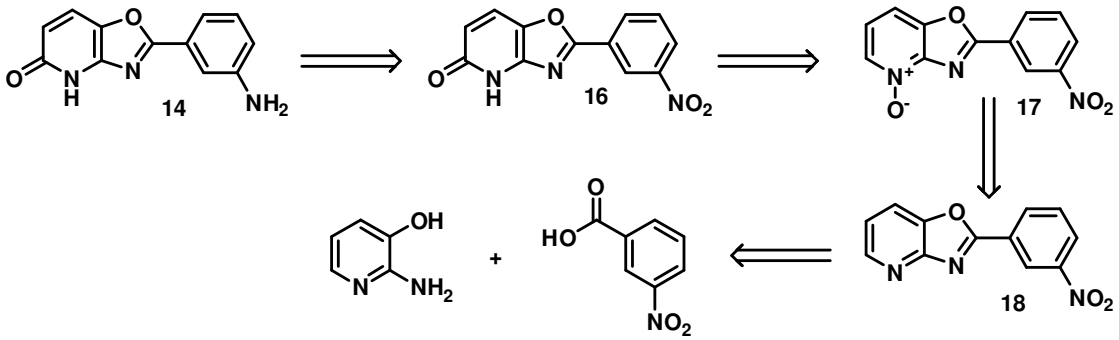


Figure 25. 2nd generation isolated AT binder.

First, the binder **14** bearing an amino group was selected for further construction. A nitro group was selected as a masked -NH_2 group, which could be reduced at a late stage to the free amine under hydrogenolytic conditions. For the construction of the pyridone moiety a rearrangement of the pyridine *N*-oxide **17** with acetic anhydride was envisioned, which is a well-known transformation widely reported in the literature over the last decades.[119] The heteroaromatic skeleton of **18** would be obtained from the condensation of a hydroxyamino pyridine and a carboxylic acid with polyphosphoric acid (PPA) (Scheme 7).

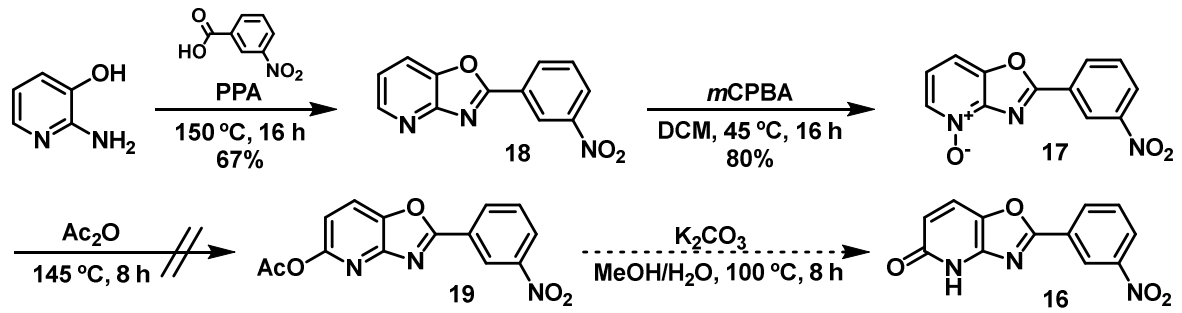


Scheme 7. Retrosynthetic analysis of the target AT binder **14**.

Firstly, 2-amino-3-hydroxypiridine and 3-nitrobenzoic acid were treated with PPA to form the oxazole derivative **18**.[120] Then, an *N*-oxidation with *m*CPBA afforded **17**, which had to be performed under reflux due to solubility issues. Finally, a rearrangement with acetic anhydride to obtain **19** was attempted, but the desired product could not be isolated (Scheme 8). Large amounts of 3-nitrobenzoic acid were detected in the crude reaction mixture by NMR spectroscopy, indicating a high degree of decomposition of the *N*-oxide **17** during this transformation. Reducing the reaction time or the temperature did not improve the outcome.

[119] (a) Markgraf, J. H.; Brown, H. B.; Mohr, S. C.; Peterson, R. G. *J. Am. Chem. Soc.* **1963**, *85*(7), 958-961. (b) Oae, S.; Kitao, T.; Kitaoka, Y. *J. Am. Chem. Soc.* **1962**, *84*(17), 3359-3362. (c) Tung, Y.-S.; Coumar, M. S.; Wu, Y.-S. *et al. J. Med. Chem.* **2011**, *54*, 3076-3080.

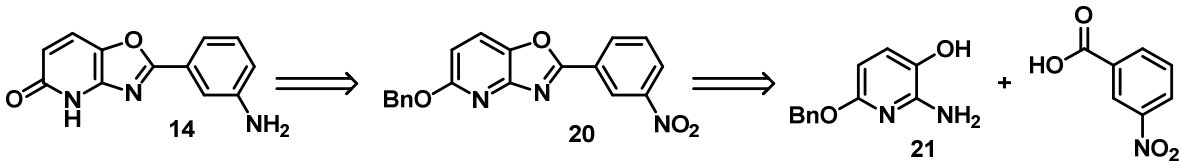
[120] Reen, G. K.; Kumar, A.; Sharma, P. *Med. Chem. Res.* **2017**, *26*(12), 3336–3344. Synthesized following a modified procedure employing PPA.



Scheme 8. Synthetic pathway towards derivative **16**.

It was hypothesized that the harsh acidic conditions might have had a negative effect on the stability of the oxazole ring. However, exposing the precursor **18** to the same reaction conditions showed no signs of degradation, evidencing the robustness of the oxazolopyridine core towards such conditions. Thus, it was clear that the rearrangement of the *N*-oxide **17** with acetic anhydride did not occur exclusively at position 6 of the pyridine as it was expected, leading to side-reactions that resulted in unwanted fragmentations.

In light of these events, another approach was envisaged. The pyridone functionality would be masked as an *O*-benzyl group, which could be cleaved under hydrogenolytic conditions to recover the free pyridones, reducing the nitro group at the same time. This *O*-benzyl substitution would be introduced at an early stage, synthesizing derivative **21**, which would be condensed with 3-nitrobenzoic acid to form the oxazole **20** (Scheme 9).



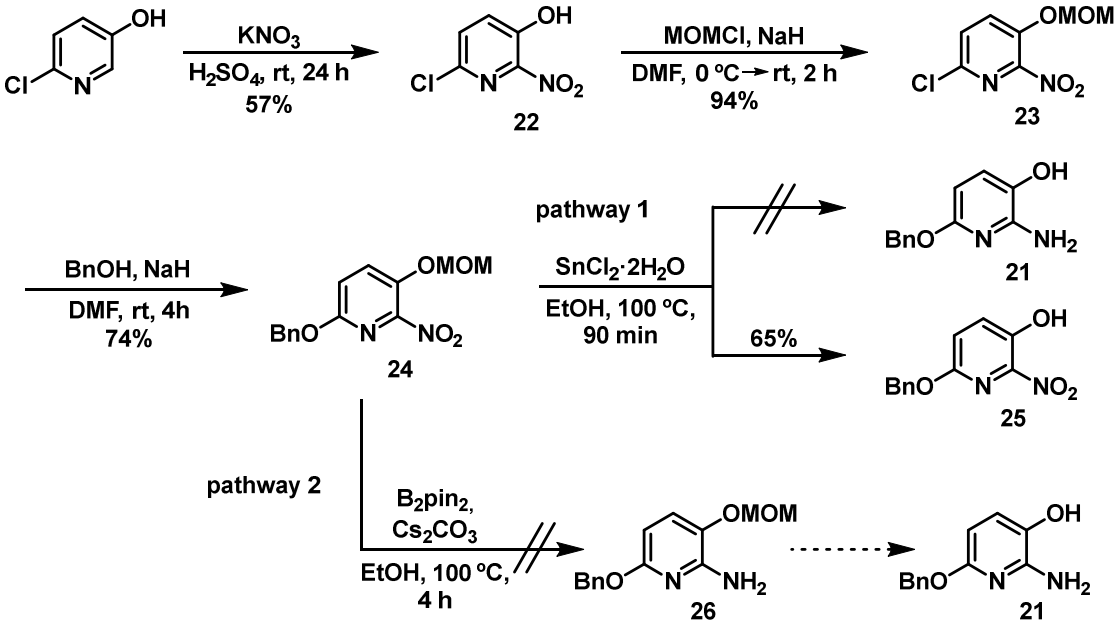
Scheme 9. Alternative retrosynthetic analysis of compound **14**.

To this end, first 6-chloropyridin-3-ol was selectively nitrated at position 2 of the pyridine ring with potassium nitrate, obtaining **22** (Scheme 10).[121] Then, the hydroxy group was protected with a MOM group, affording **23**, and a subsequent displacement of the chlorine atom with a benzyloxy group yielded **24**. To obtain compound **21**, two different approaches were envisioned: a one-pot procedure to reduce the nitro group and remove the MOM group simultaneously (pathway 1), and a stepwise process where each functional group would be individually addressed (pathway 2). However, all efforts to reduce the nitro group resulted in failure. The treatment of **24** with tin chloride, which reduces nitro functionalities through a sequence of single electron transfer (SET) steps,[122] resulted only in the cleavage of

[121] Gummadi, V. R.; Samajdar, S.; Mukherjee, S.; Bock, M. G.; Aurigene Discovery Technologies Limited. World patent WO 2017/009806 A1. 2017 Jan 19.

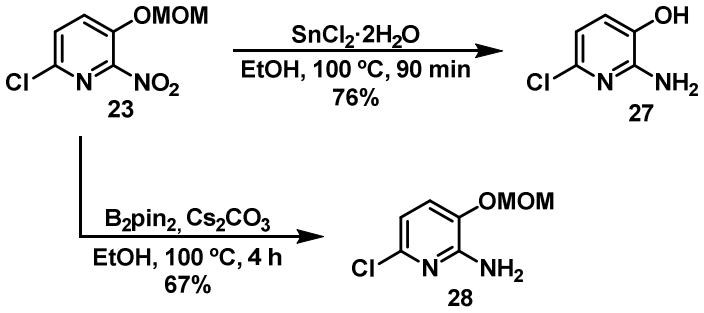
[122] Yamabe, S.; Yamazaki, S. *J. Phys. Org. Chem.* **2016**, *29*(7), 361-367.

the acid-labile MOM group, due to the acidic nature of the reaction. The nitro group was left intact, obtaining exclusively **25**.



Scheme 10. Synthetic strategy towards compound **21**.

Similarly, the metal-free procedure employing B_2pin_2 failed to afford **26**,[123] showing no reaction at all. It is important to note that hydrogenolysis was not considered, as the benzyl group would also be cleaved. Interestingly, when the precursor **23** was subjected to the same reaction conditions as above, the expected compounds **27** and **28** were obtained (Scheme 11),[124] evidencing that the introduction of the benzyloxy group hindered the reduction of the nitro group at the position 2 of the pyridine ring.



Scheme 11. Synthesis of compounds **27** and **28**.

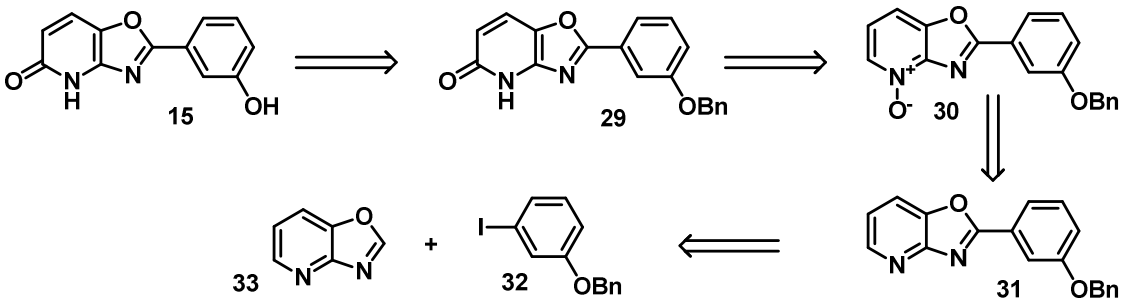
For all the previously stated reasons, further efforts towards the completion of this approach were abandoned. Furthermore, it was also unclear whether the benzyl group would survive the harsh acidic conditions of the PPA oxazole-formation step. Therefore, it was imperative

[123] Lu, H.; Geng, Z.; Li, J.; Zou, D.; Wu, Y.; Wu, Y. *Org. Lett.* **2016**, *18*, 2774–2776.

[124] Gummadi, V. R.; Samajdar, S.; Aurigene Discovery Technologies Limited. World patent WO 2015/104688 A1. 2015 Jul 16. Synthesized following a different synthetic pathway.

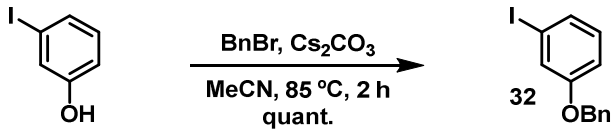
to find a milder procedure with a higher tolerance for a variety of functional groups. Additionally, it was noticed that the nitro group has a negative effect on the solubility of the molecules, which complicated their handling, reactivity and purification.

Thus, the target binder **15** was selected for further construction, which has a hydroxy group that could be masked with a benzyl group during the synthesis, potentially improving the solubility. The *N*-oxide rearrangement approach was considered again, but PyBroP-mediated, avoiding the harsh acidic conditions that led to degradation previously. **31** would be obtained from a cross-coupling reaction between the aryl iodide **32** and the oxazole derivative **33** (Scheme 12).



Scheme 12. Retrosynthetic strategy towards AT binder **15**.

First, the benzyl-protected derivative **32** was prepared reacting 3-iodophenol with benzyl bromide (Scheme 13).[125]



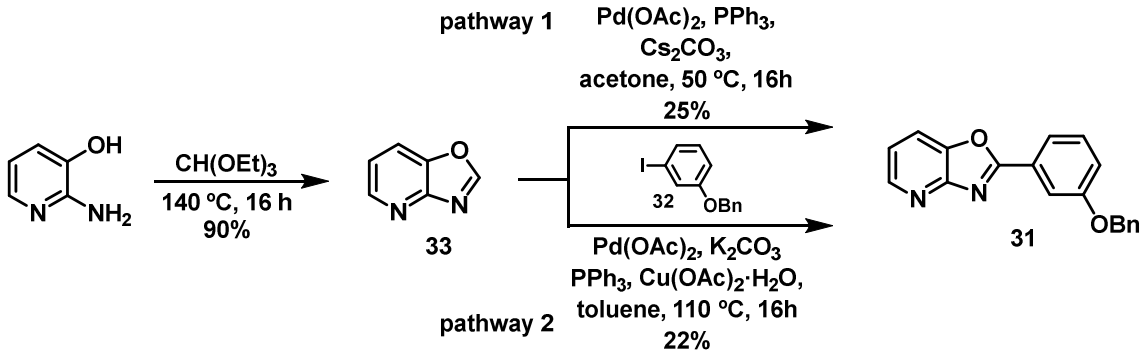
Scheme 13. Synthetic strategy towards compound **32**.

Then, the oxazolopyridine **33**, obtained reacting 2-aminopyridin-3-ol with triethyl orthoformate, was subjected to a Pd-catalyzed cross coupling reaction with **32** (Scheme 14, pathway 1).[126] The desired compound **31** was obtained, but the reaction was very low yielding, in spite of literature examples reporting higher yields for similar molecules.[127]

[125] Kim, Y. K.; Park, S. Y; Joo, H. W. *et al.*; LG CHEM, LTD. U.S. patent US 2017/0349594 A1. 2017 Dec 7.

[126] Li, Y.; Xie, Y.;Zhnag, R.; Jin, K.; Wang, X.; Duan, C. *J. Org. Chem.* **2011**, *76*, 5444–5449.

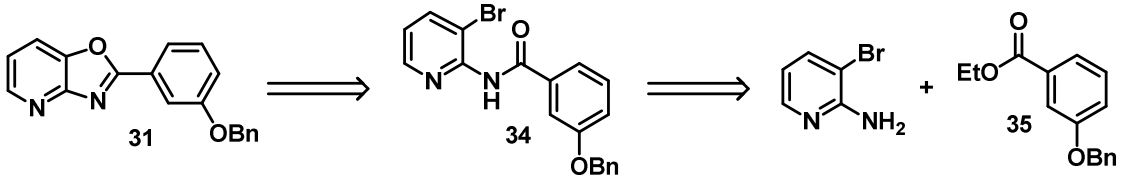
[127] Zhuravlev, F. A. *Tetrahedron Lett.* **2006**, *47*, 2929-2932.



Scheme 14. Synthetic pathway towards **31** through a Pd-catalyzed arylation.

Other reports mention a Pd(OAc)/Cu(II)/PPh_3 co-catalytic system that is able to successfully arylate benzoazoles with aryl bromides,[128] a non-trivial transformation that does not proceed with palladium catalysis alone.[129] Therefore, it was expected that by adding a Cu(II) co-catalyst to our more reactive reaction system, which uses aryl iodides, the arylation would proceed more efficiently. The outcome, however, was similar to that of the first approach using palladium catalysis alone. Performing the reaction under air or argon had no influence on the yield of the reaction (Scheme 14, pathway 2).

This disappointing findings prompted us to turn our attention to other approaches to construct the oxazole core, which is the key component of the heteroaromatic skeleton of the AT binder. An amidation and *O*-arylation sequence was envisioned to obtain the intermediate **31** (Scheme 15).[130] An $AlMe_3$-promoted amidation was selected,[131] as more traditional amide coupling conditions proved to be problematic for this transformation. Furthermore, the coupling reaction could be directly performed using esters, which reduced the number of synthetic steps.



Scheme 15. Synthetic pathway towards **31** through a Pd-catalyzed arylation.

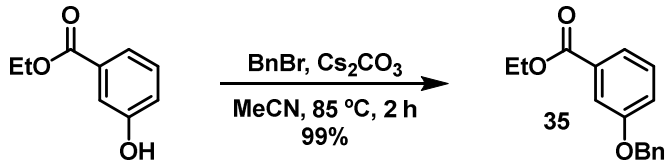
[128] Yan, X.-M.; Mao, X.-R.; Huang, Z.-Z. *Heterocycles* **2011**, *83*, 1371-1376.

[129] Alberico, D.; Scott, M. E.; Lautens, M. *Chem. Rev.* **2007**, *107*, 174-238.

[130] Xu, D.; Xu, X.; Liu, Z.; Sun, L.-P.; You, Q. *Synlett* **2009**, *7*, 1172-1174.

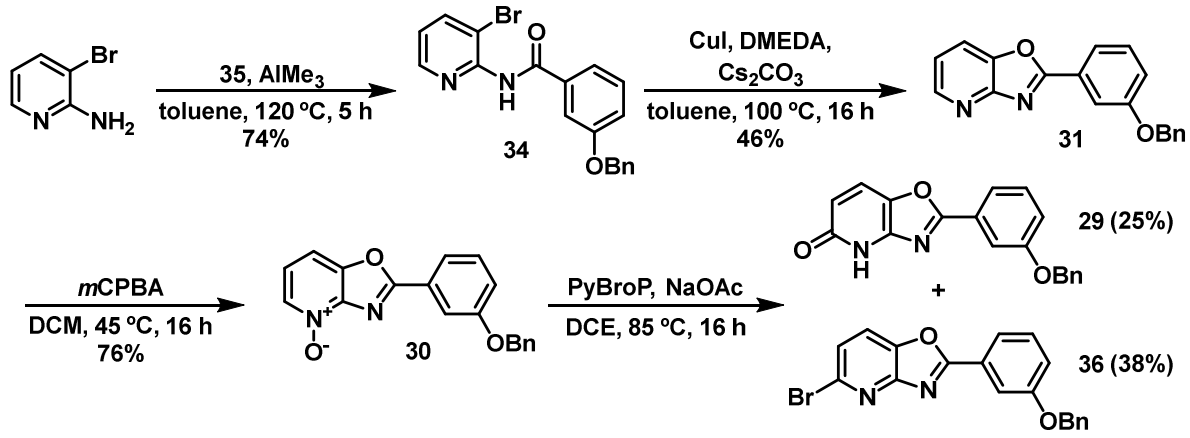
[131] (a) Basha, A.; Lipton, M.; Weinreb, S. M. *Tetrahedron Lett.* **1977**, *18*(48), 4171-4172. (b) Li, J.; Subramaniam, K.; Smith, D.; Qiao, J. X. *et al. Org. Lett.* 2011, *14*(1), 214-217.

First, the ethyl ester derivative **35** had to be prepared for the $AlMe_3$-promoted amidation reaction, which was conveniently synthesized in one step from commercially available ethyl 3-hydroxybenzoate (Scheme 16).[132]



Scheme 16. Synthetic pathway towards **31** through a Pd-catalyzed arylation.

Then, **35** was coupled with 2-amino-3-bromopyridine, yielding **34**. The Cu-catalyzed *O*-arylation to obtain the cyclized derivative **31** proceeded as expected, albeit with a moderate yield. *N*-oxide formation followed by a PyBroP-mediated rearrangement afforded the desired compound **29**.

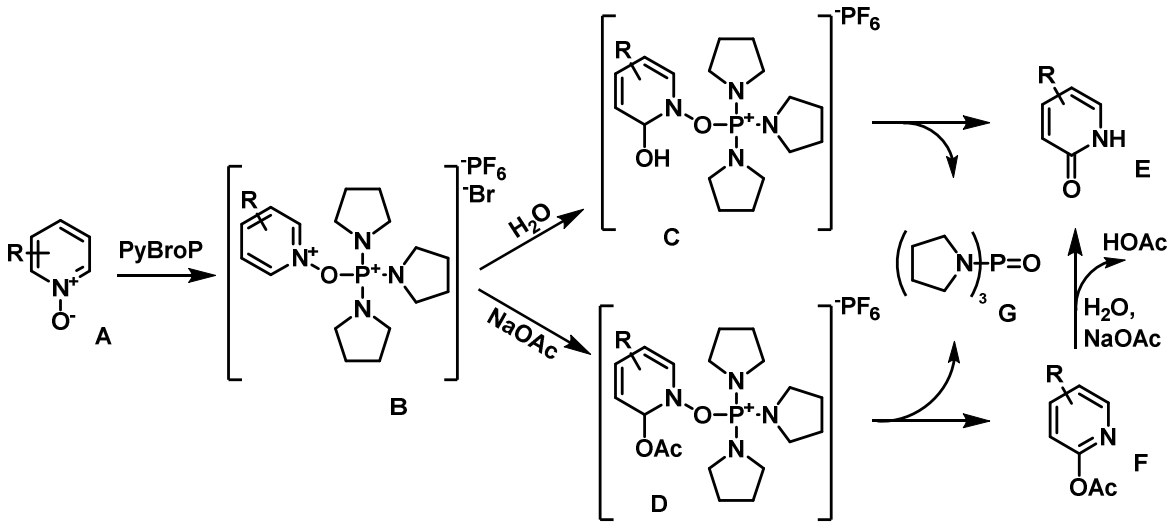


Scheme 17. Synthetic pathway towards **31** through a Pd-catalyzed arylation.

However, the desired pyridone **29** was obtained with just a 25% yield, being the brominated **36** derivative the major product. This outcome can be explained with the proposed mechanism for this reaction (Scheme 18).[133] First, the bromine atom of PyBroP is displaced by the pyridine *N*-oxide **A**, obtaining the activated complex **B**. A nucleophilic substitution with water or sodium acetate, both of which are present in the reaction mixture, furnishes **C** and **D**, respectively. A subsequent rearomatization affords the desired pyridone **E** or the acetate **F**, with the loss of phosphoryltripyrrolidine **G**. **F** is readily transformed into **E** due the basic reaction conditions.

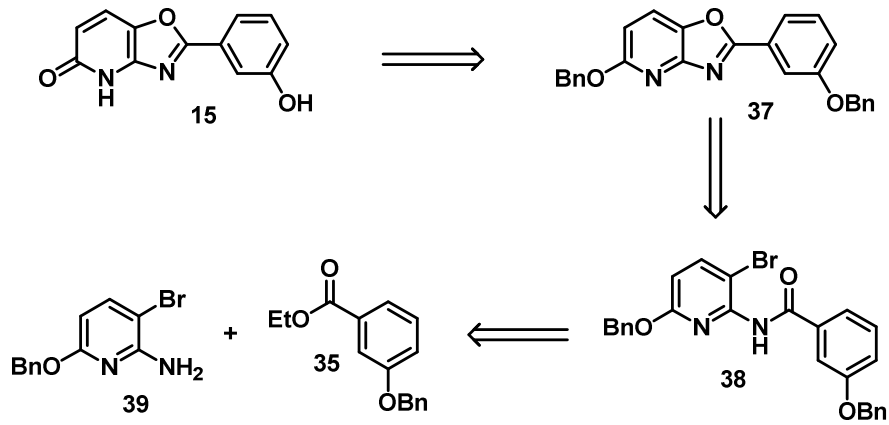
[132] Adjabeng, G.; Bifulco, N.; Davis-Ward, R. *et. al.*; Smithkline Becham Corporation. World patent WO 2009/076140 A1. 2009 Jun 18. Synthesized following a modified procedure.

[133] Wang, D.; Zhao, J.; Wang, Y.; Hu, J. *et al. Asian J. Org. Chem.* **2016**, *5*, 1442-1446.



Scheme 18. Proposed mechanism for the PyBroP-mediated rearrangement.

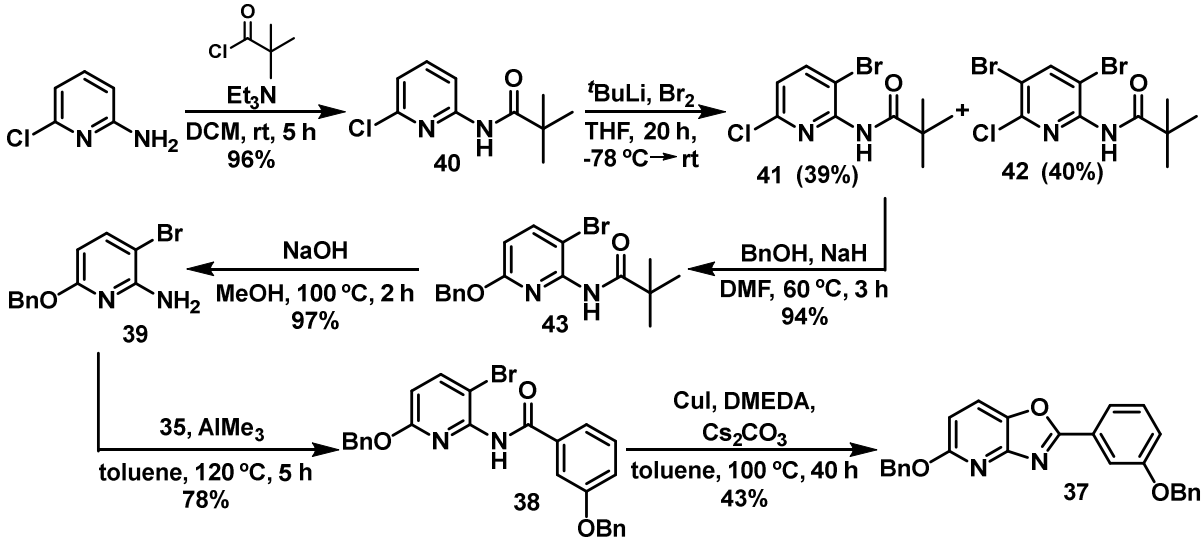
The formation of the brominated side-product **36** indicated that the nucleophilic substitution of the Br^- ion from PyBroP was more favorable than that of water or NaOAc, therefore obtaining a poor yield of the desired pyridone **29**. A late-stage rearrangement to obtain the pyridone moiety did not seem feasible. Given that the amidation and *O*-arylation sequence previously established did not rely on harsh conditions to form the oxazole core, as in the case of PPA, the early introduction of a benzylozy group as a masked pyridone seemed like an attractive option again. To that end, the key intermediate **39** would be prepared. After oxazole cyclization, the benzyl groups would be cleave under hydrogenolysis to render the AT binder **15** (Scheme 19).



Scheme 19. Alternative retrosynthetic analysis of the target AT binder **15**.

First, 2-amino-6-chloropyridine was reacted with pivaloyl chloride, to afford **40**. The pivaloyl group would act as a directing group in the subsequent directed *ortho* metalation reaction using *tert*-butyl lithium, where the lithiation occurred selectively at *ortho* respect to

the pivaloyl, and a subsequent reaction with bromine as the electrophile yielded **41**.[134] However, an equimolar amount of the side-product **42** was also obtained, diminishing the yield of the reaction. This behavior is indicative of the formation of radicals in the reaction mixture, which can be promoted by the interaction of bromine with THF.[135] Benzyloxy insertion followed by the cleavage of the pivaloyl group afforded **39**, which was subsequently coupled with the previously synthesized ethyl ester **35** to yield **38**. The intramolecular cyclization to form the oxazolo[4,5-*b*]pyridine skeleton in **37** did not proceed as expected. Extended reaction times and the addition of more equivalents of the reagents were necessary to achieve the full conversion of the starting material, and the formation of multiple side-products resulted in a poor yield.



Scheme 20. Modified synthetic pathway to obtain derivative **37**.

Despite successfully obtaining the desired derivative **37**, this synthetic pathway presents some major drawbacks that have a negative impact on the overall yield of the procedure, such as the formation of side-products and the low-efficiency Cu-catalyzed oxazole cyclization step. Therefore, a few modifications were introduced in the synthetic route.

First, we turned our attention to the *N*-heterocyclic carbene copper(I) complex [(*i*Pr)CuCl] (Figure 26), which has demonstrated excellent catalytic activity in the synthesis of 2-aryl benzoxazoles and benzothiazoles.[136] This group of catalysts is also air- and moisture-stable.

[134] Widdowson, K. L.; GlaxoSmithKline Intellectual Property Development Limited. World patent WO 2016/027249 A1. 2016 Feb 26. Synthesized following a modified procedure.

[135] Lo, J. M. H.; Marriott, R. A.; Giri, B. R.; Roscoe, J. M.; Klobukowski, M. *Can. J. Chem* **2010**, *88*, 1136-1145.

[136] Urzúa, J. I.; Contreras, R.; Salas, C. O.; Tapia, R. A. *RSC Advances* **2016**, *6*(85), 82401-82408.

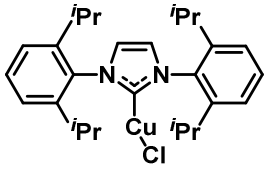
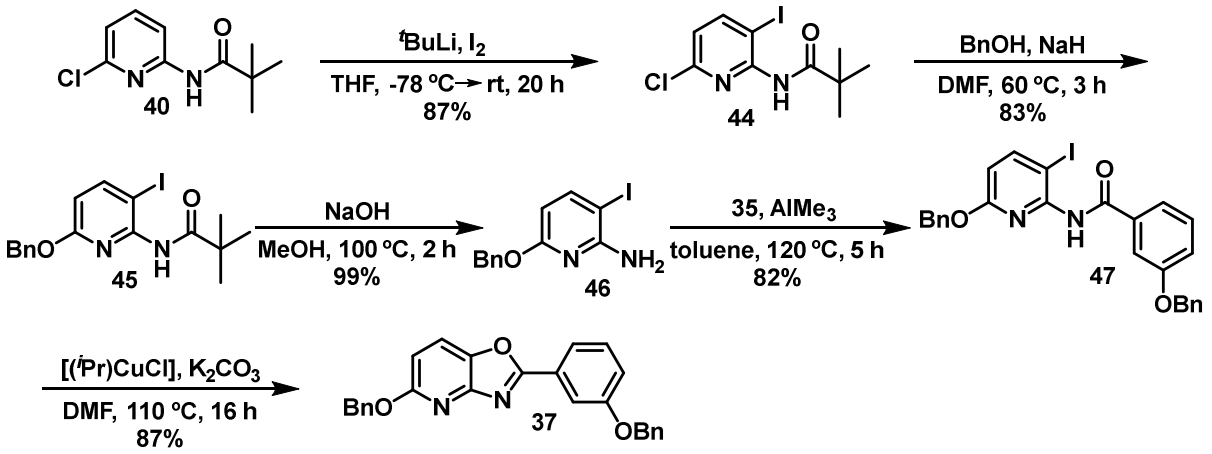


Figure 26. Structure of the *N*-heterocyclic carbene copper catalyst [(*i*Pr)CuCl].

Additionally, the inserted halogen atom would be iodine instead of bromine. This modification would not only avoid the formation of side-products during the directed *ortho* metalation step, as molecular iodine is perfectly compatible with THF, but also potentially facilitate copper insertion during oxazole formation.

Following a similar reaction sequence as previously, **40** was first iodinated selectively at the *ortho* position with respect to the pivaloyl group to obtain **44**, without the formation of any significant side-products.[137] Benzyloxy insertion followed by pivaloyl removal afforded **46**, which was subsequently coupled with **35** to yield **47**. The cyclization step to afford the oxazole **37** proceeded smoothly overnight, confirming the efficiency of *N*-heterocyclic carbene copper catalysts in this type of transformations.



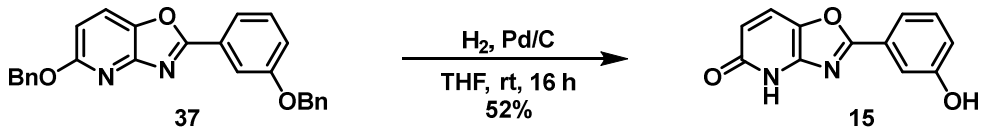
Scheme 21. Modified synthetic pathway to obtain derivative **37**.

With this optimized procedure, the overall yield over 6 steps to obtain **37** increased from 11% to 49%. Finally, the hydrogenolysis of both benzyl groups furnished the target AT binder **15**. THF was used as the solvent, since **37** was insoluble in the protic solvents normally used in this transformation, such as methanol. THF/MeOH solvent mixtures resulted in the precipitation of **37**. Although the hydrogenolysis rate in THF is supposed to be excellent,[138] 20 mol% of the Pd/C catalyst were necessary to afford **15** overnight, as

[137] Feng, Y; Chen, Y. T.; Sessions, H. *et al.*. World patent WO 2011/050245 A1. 2011 Apr 28.

[138] Dyson, P. J.; Jessop, P. G. *Catal. Sci. Technol.* **2016**, *6*(10), 3302-3316.

reducing the catalyst loading resulted in mixtures where the phenolic benzyl ether was not fully cleaved.



Scheme 22. Hydrogenolysis to obtain the isolated AT binder **15**.

Following this procedure, the isolated AT binder **15** was obtained with an overall yield of 25% over 7 steps. Unfortunately, only tetrahydrofuran was able to fully solubilize **15**, making the determination of its preliminary binding potential in suitable solvents for NMR titrations, such as chloroform or dichloromethane, a challenge (Section 3.2).

3.1.2.2. Binder with linker

As in the case of the isolated AT binder, two different derivatives were envisioned: one substituted with a nitro group, acting as a masked amino group, and one bearing an *O*-benzyl-masked hydroxy group. It is essential to keep these functionalities masked to avoid any side reactions during the oligomer formation with standard SPPS, which can be removed at a later stage. The binder is modified at position 5 of the pyridine ring with an acetic acid linker, which will be used to attach this unnatural nucleobase to the PNA backbone.

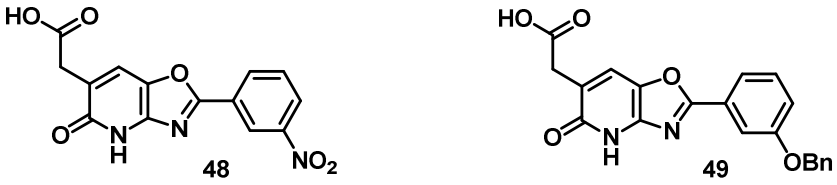
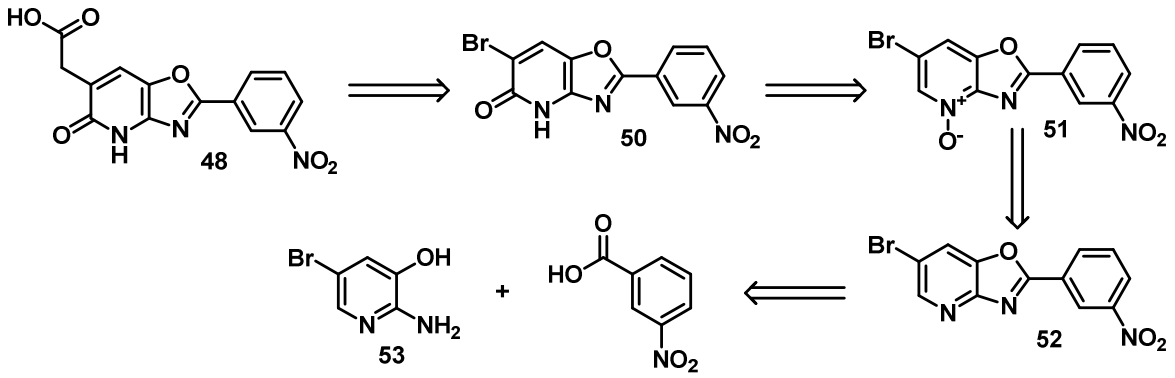


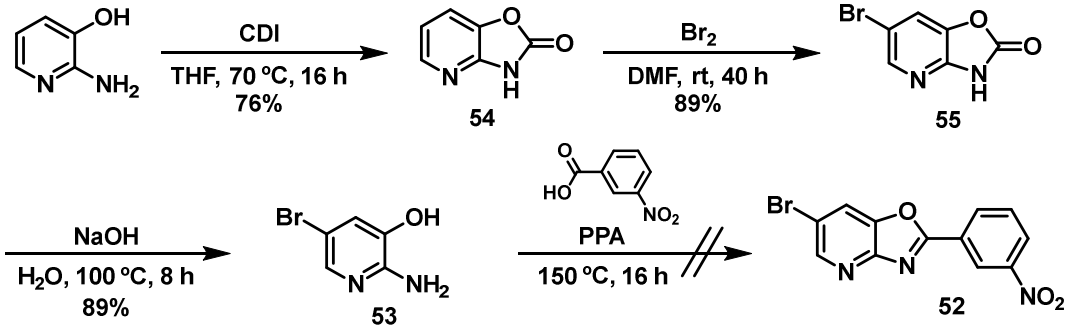
Figure 27. Structure of the linker-attached AT binders **48** and **49**.

The efforts towards the construction of the linker-attached AT binder were performed in parallel with the synthesis of the isolated binder. Therefore, the first approach also consisted in the late-stage pyridone formation via an acetic anhydride-mediated *N*-oxide rearrangement. In this case, however, the pyridine ring would be functionalized at position 5 with a bromine group from the beginning, which could later be used to install the linker at that position (Scheme 23).



Scheme 23. Retrosynthetic analysis of the target linker-attached AT binder **48**.

For this purpose, first 2-aminopyridin-3-ol was reacted with 1,1'-carbonyldiimidazole to protect both the amino and hydroxy group as an oxazolidinone, affording **54** (Scheme 24). Then, a selective bromination on the position 5 of the pyridine ring to obtain **55**, followed by the hydrolysis of the oxazolidinone under basic conditions, yielded the brominated pyridine derivative **53**.[139] However, the PPA-mediated condensation between **53** and 3-nitrobenzoic acid did not proceed as expected. Although TLC analysis showed one unique spot, LC-MS analysis of the crude revealed that some unidentified side-products were formed. Some amounts of the desired product **52** were detected, but could not be isolated from the crude mixture.



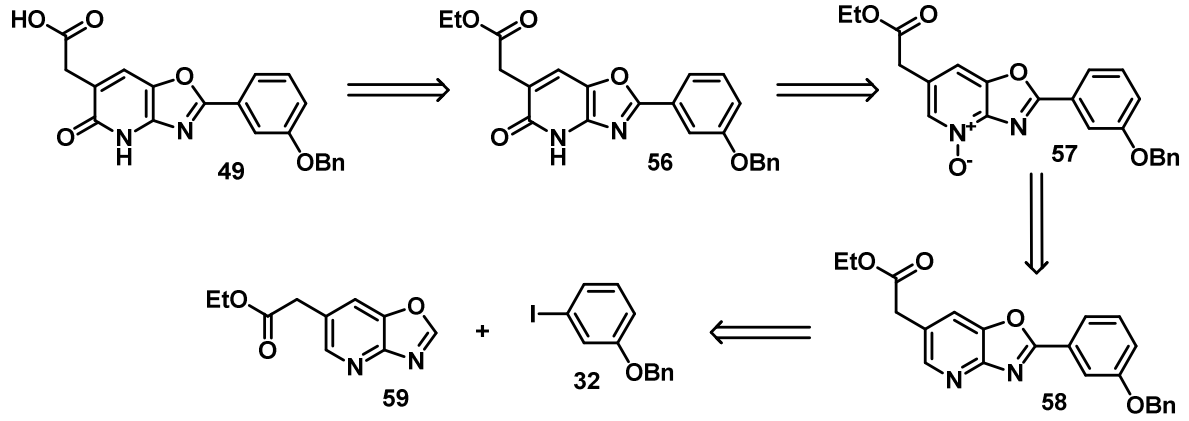
Scheme 24. Synthetic route towards derivative **52**.

These finding, together with the incompatibility of the acetic anhydride-mediated *N*-oxide rearrangement with the oxazolo[4,5-*b*]pyridine core noticed during the synthesis of the isolated AT binder, prompted us to look for alternative pathways.

Analogous to the synthetic strategy of the isolated AT binder, the new approach would also replace the amino group with a hidroxy group, masked as an OBn during the synthesis. It relied on a late-stage PyBroP-mediated *N*-oxide rearrangement to form the pyridone in **56**

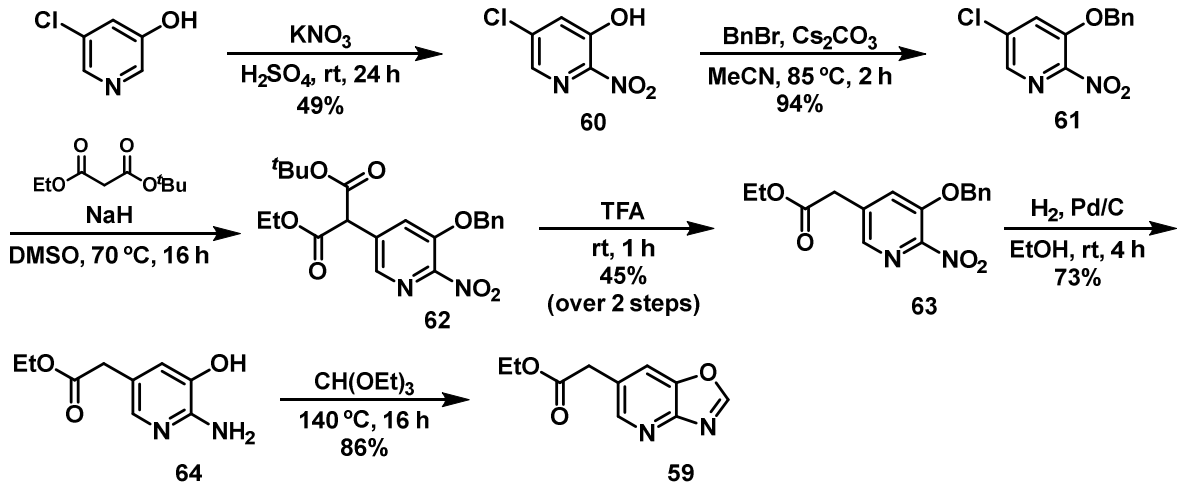
[139] Aronov, A.; Bandarage, U. K.; Cottrell, K. M. *et al.*; Vertex Pharmaceuticals Incorporated. U.S. patent US 8,524,906 B2. 2013 Sep 3.

(Scheme 25). The aromatic skeleton would be constructed via a Pd-catalyzed cross-coupling between **59** and the previously synthesized **32**. However, the linker functionalization would be performed at an early stage, potentially facilitating its introduction.



Scheme 25. Retrosynthetic pathway towards the linker-attached AT binder **49**.

To this end, first 3-chloropyridin-5-ol was nitrated at position 6 to obtain **60**. Although higher yields have been reported with the traditional nitration conditions employing nitric acid, potassium nitrate was used instead due to its safer manipulation and easier reaction setup.[140] The hydroxy group was protected with a benzyl group to yield **61**, followed by the insertion of *tert*-butyl ethyl malonate, which was subsequently decarboxylated under acidic conditions in a one-pot fashion without isolating **62**, obtaining compound **63**.[141] Finally, a hydrogenolysis of both nitro and benzyl groups followed by the oxazole ring formation with an orthoester afforded the target compound **59** (Scheme 26).



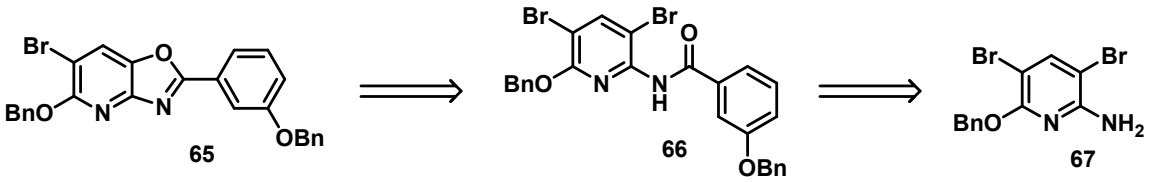
Scheme 26. Synthetic route towards the oxazolo[4,5-*b*]pyridine derivative **59**.

[140] Vakalopoulos, A.; Brockschneider, D.; Wunder, F.; Stasch, J.-P.; Marquardt, T.; Dietz, L.; Bayer Pharma AG. World patent WO 2016/087343 A1. 2016 Jun 09. Synthesized following a modified procedure employing KNO_3.

[141] Li, L.; Wang, Y.; Ren, P.; Liu, Y.; Der, R. K.; Kura Oncology Inc.. World patent WO 2018/106818 A1. 2018 Jun 14.

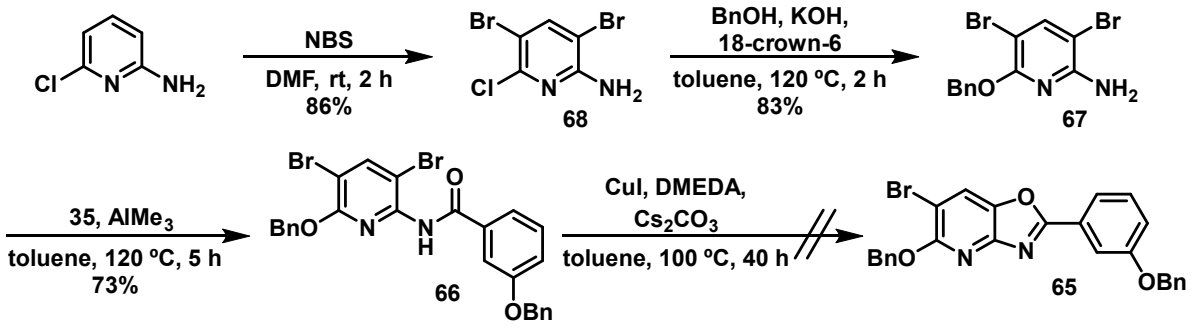
Despite successfully obtaining **59**, this approach was not completed after finding out the low efficiency of the Pd-catalyzed cross coupling and the PyBroP mediated rearrangement during the synthesis of the isolated AT binder. It was clear that the linker would have to be introduced at a later stage, therefore recovering the need for a halogen group that could eventually be functionalized.

For that purpose, an intramolecular copper-catalyzed cyclization of a dihalogenated derivative **63** was envisioned (Scheme 27). In this way, a halogen would still be available in **62** for the subsequent introduction of the linker after the formation of the oxazole ring. Additionally, to avoid the inconveniencies of *N*-oxide rearrangements to obtain the pyridone ring, a benzyloxy substitution would be introduced *ortho* to the pyridine nitrogen, which could be subjected to hydrogenolysis at a late stage to unmask the free pyridone.



Scheme 27. Retrosynthetic analysis for the production of the brominated derivative **65**.

Our first attempt following this approach started with the dibromination of 2-amino-6-chloropyridine, which afforded **68**.[142] Next, the chlorine atom was displaced with a benzyloxy group, obtaining **67**. To our delight, the nucleophilic substitution proceeded smoothly without the need of protecting the amino group, presumably due to the presence of the two electron-withdrawing halogens added to the inherent electron deficiency of pyridine. Then, an $AlMe_3$-promoted amidation reaction afforded **66**, which was then subjected to an intramolecular cyclization to form the oxazolo[4,5-*b*]pyridine skeleton in **65** (Scheme 28).



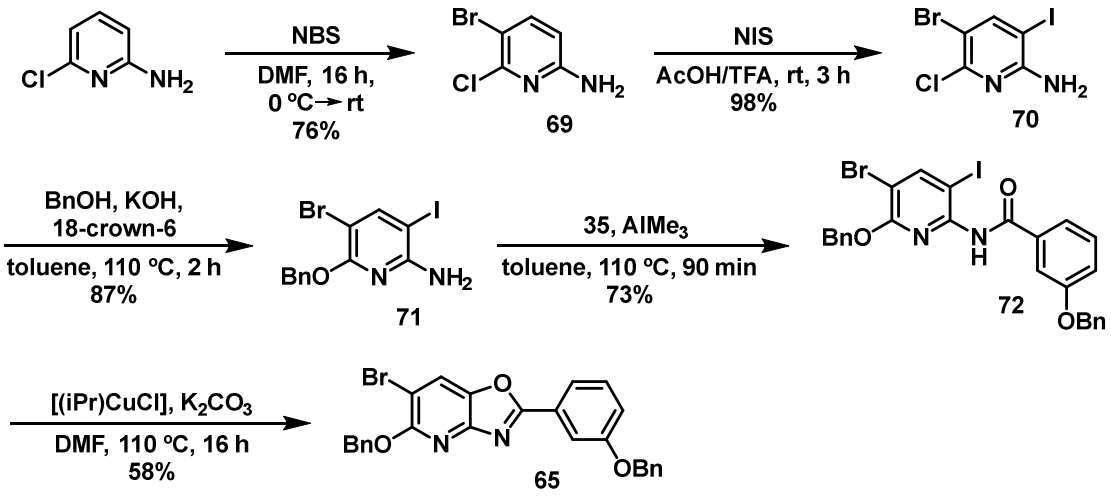
Scheme 28. First approach to obtain the brominated AT binder **65**.

[142] Bloomgren, P. A.; Currie, K. S.; Kropf, J. E. *et al.*; Gilead Connecticut, Inc.. U.S. patent 9,562,056 B2. 2017 Feb 7.

Unfortunately, the reaction did not proceed smoothly. More equivalents of each reagent had to be added to convert all the starting material **66**, even over an extended reaction time of 40 hours. Multiple side-products were formed and only trace amounts of the desired product **65** could be detected by NMR spectroscopy.

Following the modifications to the synthetic strategy towards the isolated AT binder, a similar approach was taken towards the synthesis of the brominated derivative **65**. Instead of a dibrominated pyridine derivative precursor, a heterohalogenated one bearing an iodine atom was used, presumably facilitating copper insertion during oxazole formation. The catalyst for this reaction was also replaced by the *N*-heterocyclic carbene copper(I) complex [(*i*Pr)CuCl], which showed an excellent performance during the synthesis of the isolated AT binder.

To synthesize the tri-hetero-halogenated pyridine derivative **70**, 2-amino-6-chloropyridine was first treated with *N*-bromosuccinimide, followed by *N*-iodosuccinimide (Scheme 29).[143] It is important to note that, during the reaction to obtain **69**, a precise control of the temperature and the addition rate of NBS as well as light deprivation are necessary to minimize dibromination, which would afford **68** instead. Then, a substitution of the chlorine with a benzyloxy group and a subsequent amidation reaction yielded **72**. In this particular case, both the reaction time and temperature had to be reduced and carefully controlled to reduce the formation of an unidentified side-product. Finally, employing [(*i*Pr)CuCl] as the catalyst, the cyclization reaction to obtain **65** proceeded with a 58% yield.

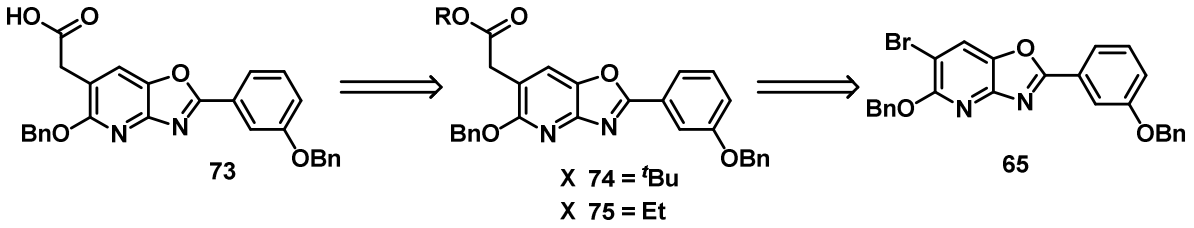


Scheme 29. Modified synthetic strategy towards the brominated AT binder **65**.

With compound **65** in hand, the stage was set for the linker introduction step, which would enable the attachment of the AT binder to a PNA backbone. To this end, a free carboxylic

[143] Ichikawa, M.; Matsunaga, H.; Oginoya, N. *et al.*; Japan patent JP 2011-178779 A. 2011 Sep. 15.

acid functionality is necessary, to couple it to the secondary amine from the backbone. Compound **73** with an acetic acid linker could be obtained from an acetate linker precursor (either ethyl or *tert*-butyl), which would be introduced exploiting the available halogen in **65** (Scheme 30).



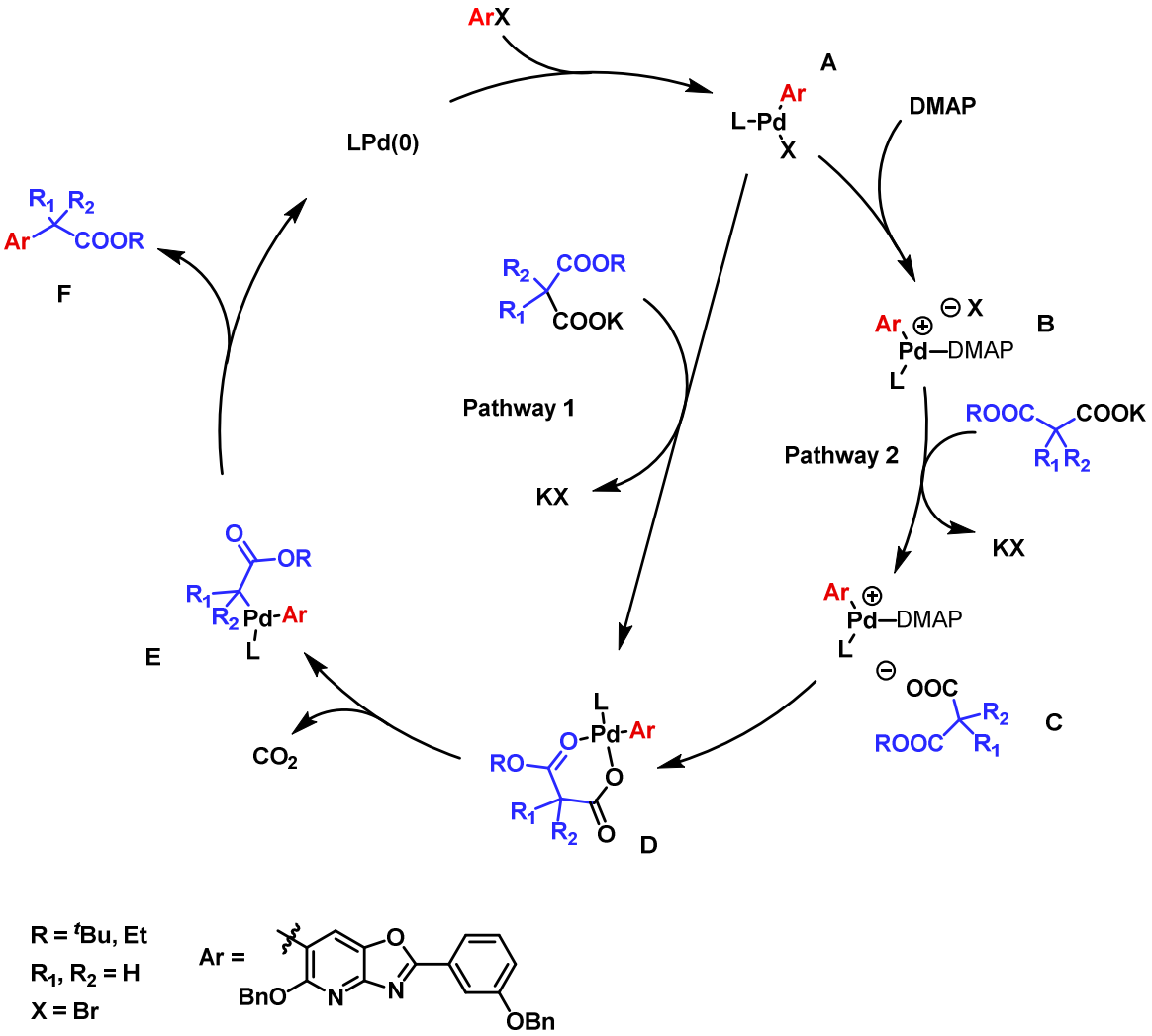
Scheme 30. Retrosynthetic analysis for the linker insertion.

After considering the available options to introduce the desired functionality, a palladium-catalyzed procedure introduced by Liu *et al.* and optimized by Xu *et al.* was chosen for this purpose.[144] It involves a cross-coupling reaction between an aryl halide and an alkyl potassium malonate salt, which is also decarboxylated during the catalytic cycle (Scheme 31).

First, **A** is formed through the oxidative addition of ArX to Pd(0)L. In the absence of DMAP, the halide anion from the palladium coordination sphere is exchanged by the carboxylate anion to form intermediate **D** (pathway 1). This transmetalation step may proceed smoothly when aryl chlorides are used,[136a] but can be sluggish with aryl bromide substrates.[145] DMAP may facilitate the transmetalation step by coordinating to palladium first, forming the cationic Pd complex **B** (pathway 2). The carboxylate anion reacts with **B** to form the intermediate **D**, and a subsequent decarboxylation provides the enolated complex **E**. A reductive amination affords the desired product **F** and regenerates the active Pd(0)L complex. In the case of α-alkylated malonate salts (R_1, $R_2 \neq H$), the transmetalation step may be challenging due to steric hindrance.

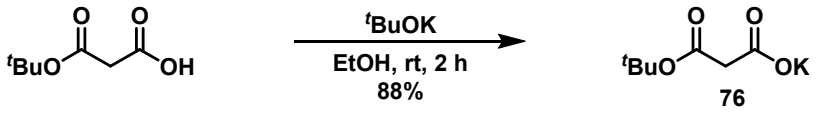
[144] (a) Shang, R.; Ji, D.-S.; Chu, L.; Fu, Y.; Liu, L. *Angew. Chem. Int. Ed.* **2011**, *50*, 4470-4474. *tert*-Butyl potassium malonate was synthesized following a modified procedure, starting from *tert*-butyl malonic acid. (b) Feng, Y.-S.; Wu, W.; Xu, Z.-Q.; Li, Y.; Li, M.; Xu, H.-J. *Tetrahedron* **2012**, *68*, 2113-2120.

[145] Fors, B. P.; Buchwald, S. L. *J. Am. Chem. Soc.* **2009**, *131*, 36, 12898-12899.



Scheme 31. Proposed mechanism for the decarboxylative cross-coupling reaction to insert the linker. The actual substituents that will participate in the reaction are also shown (bottom left).

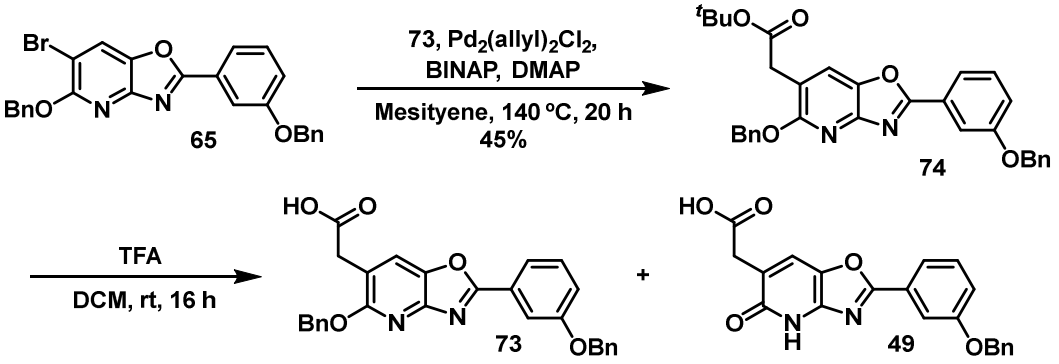
Two different derivatives were envisioned, having either a *tert*-butyl or an ethyl carboxyl-protection. The preparation of the the *tert*-butyl ester was carried out in the first place, due to its expected lability against mild acidic hydrolysis, potentially avoiding the need for chromatographic purification. *tert*-Butyl malonic acid was treated with potassium *tert*-butoxide at room temperature, obtaining **76** (Scheme 32).[144a]



Scheme 32. Synthesis of *tert*-butyl potassium malonate 76.

Then, **65** and **76** were cross-coupled employing the palladium-catalyzed procedure described previously (Scheme 33). Although the reaction provided the desired compound **74**, the yield was modest, and the formation of side-products could be observed. Some amounts of the debrominated **37** were also detected, meaning that hydrogen insertion

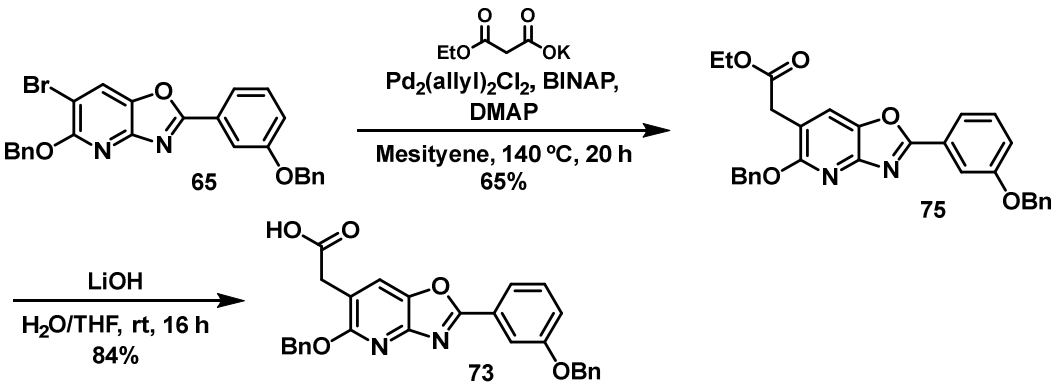
competes with the transmetalation step which, as mentioned before, can be challenging with aryl bromides.[145] Varying the equivalents of the reagents, temperature or reaction time had little effect on the outcome.



Scheme 33. Synthetic route towards **73** using *tert*-butyl potassium malonate **76**.

The hydrolysis of the *tert*-butyl group under diluted acidic conditions proved to be more challenging than initially expected. The benzyl group masking the pyridone was also readily cleaved under these conditions, affording mixtures of **73** and **49**. Increasing the reaction times did not lead to a full conversion to **49**, and reduced reaction times did not avoid its formation. The use of undiluted TFA to test if **49** could be exclusively obtained was not considered, as the phenolic benzyl ether was proven to be labile under those strongly acidic conditions.

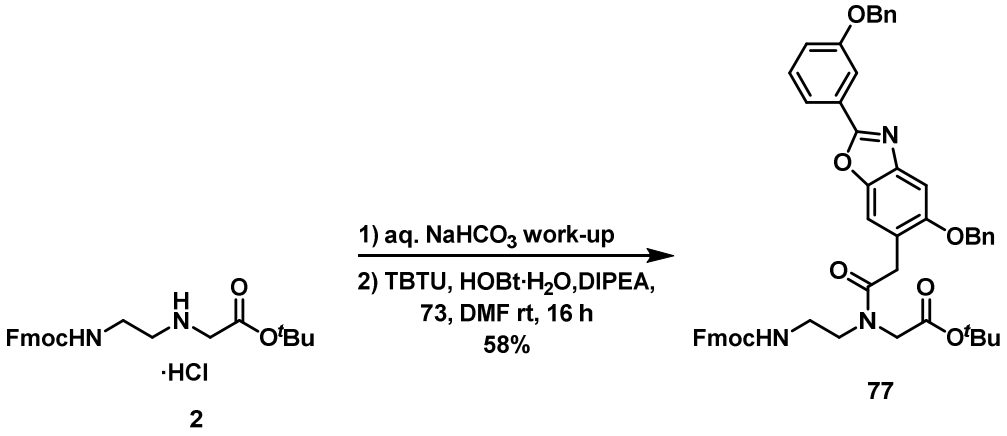
In view of these findings, the derivative bearing an ethyl ester was prepared, which would be hydrolized under basic conditions, therefore avoiding the cleavage of the benzyl groups. Thus, **65** was reacted with the commercially available ethyl potassium malonate to afford **75**. Under these conditions, less side-products were observed, resulting in an increased yield of 65%, compared to the 45% obtained using *tert*-butyl malonate. Hydrolysis of the ethyl ester under basic conditions provided the desired target binder **73** in a good yield, which could be isolated from the reaction mixture by filtration, taking advantage of its high insolubility.



Scheme 34. Synthetic route towards **73** using ethyl potassium malonate.

The target linker-attached AT binder **73** was obtained with an overall yield of 15% over 7 steps. Although this demanding synthetic strategy has been highly optimized, it still suffers from some moderate yielding steps, such as the oxazole formation from the dihalogenated derivative **72** and the cross-coupling reaction with ethyl malonate to obtain **52**. However, its operational simplicity and ease of purification make this an attractive approach towards the construction of linker-attached binders.

Additionally, the binder **73** was coupled with the previously synthesized PNA backbone **2**, yielding the AT binder PNA monomer **77** (Scheme 35). It was found that the previously employed coupling reagent (EDC·HCl) was not suitable for these artificial binders, failing to effect the desired coupling. Therefore, other conditions were explored, and a mixture of TBTU and HOBt in the presence of DIPEA was selected for this reaction, successfully affording **77** with a moderate yield.[146]



Scheme 35. Synthesis of the AT binder PNA monomer **77**.

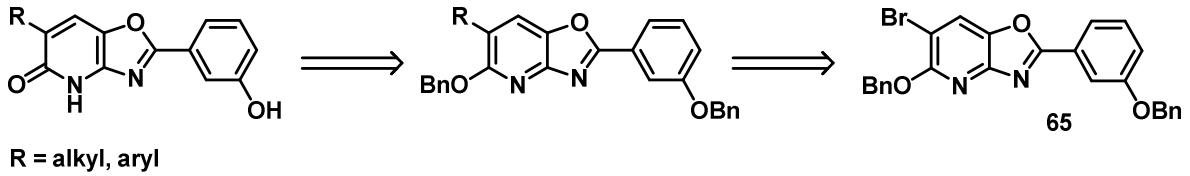
3.1.2.3. Solubilization attempts

While a slight insolubility of the isolated AT binder **15** in lower polarity solvents was to be expected, NMR analysis in deuterated chloroform or dichloromethane proved to be impossible, regardless of how low the concentration of the sample was. Only deuterated THF was able to solubilize the AT binder completely, which is not best suited for NMR (as discussed in Section 3.2). Therefore, efforts towards the construction of a modified AT binder with higher solubility were initiated.

It was imperative to find a molecule able to be functionalized with different solubilizing residues, ideally in one step and without the need to design a new synthetic strategy. The intermediate **65** synthesized during the preparation of the linker-attached AT binder was selected as the perfect candidate for this task, since it possesses a bromine suitable for Pd-

[146] MacKinnon, K. F.; Qualley, D. F.; & Woski, S. A. *Tetrahedron Lett.* **2007**, *48*(45), 8074-8077.

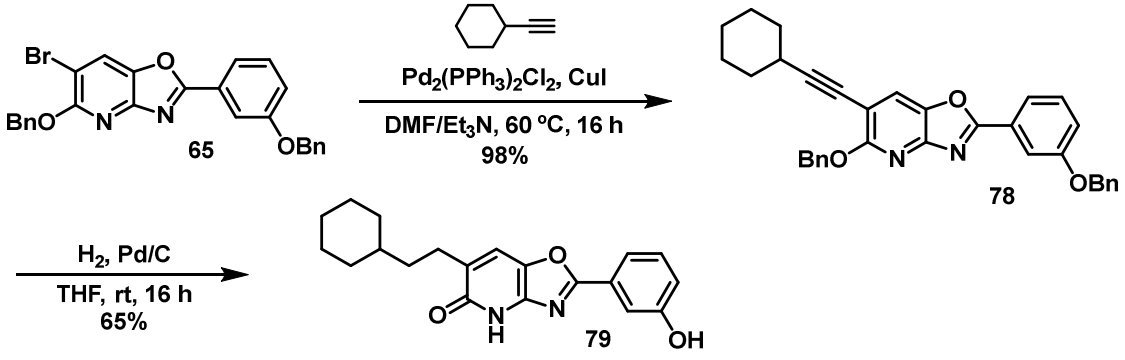
catalyzed cross-coupling reactions such as Sonogashira or Suzuki-Miyaura. This one-step functionalization, followed by the hydrogenolysis of both benzyl groups, would provide an accessible pathway towards the construction of potentially more soluble binders (Scheme 36).



Scheme 36. Retrosynthetic strategy towards AT binders with increased solubility.

Although alkyl-substituted aryl moieties have previously been used as solubilizing residues in our group, the improvement of the solubility of the obtained molecules in dichloromethane was marginal, as evidenced by the multiple obstacles that were faced during the construction of a soluble GC binder (Section 3.1.3.). Extension of the aromatic system and increased π-π stacking might potentially render the molecule less soluble, even in the presence of a bulky alkyl group. Therefore, we focused exclusively on alkyl substitutions.

To that end, ethynylcyclohexane was selected, which was subjected to a Sonogashira cross-coupling reaction with the brominated derivative **65** to afford **78** with an excellent yield (Scheme 37). It is important to note that previous attempts employing alkynes with a linear alkyl residue of varying lengths failed to render the desired molecule, resulting in an inseparable mixture of impurities, and therefore were not considered. A subsequent reduction of the triple bond and cleavage of the benzyl groups under a hydrogen atmosphere furnished the target binder **79**. Longer reaction times and higher loadings of the Pd/C catalyst were necessary to achieve full conversion to the desired compound.

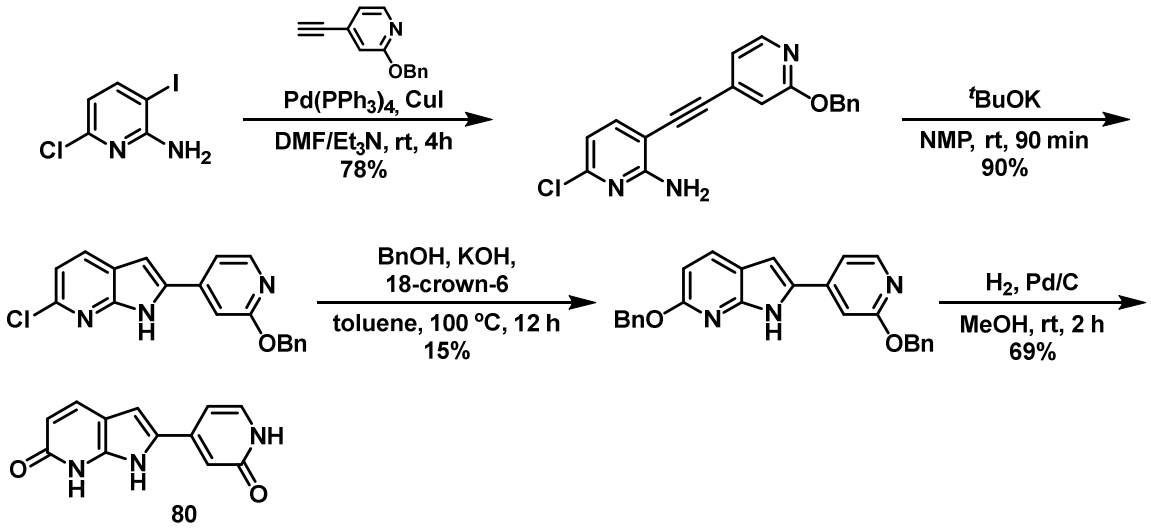


Scheme 37. Synthesis of the target AT binder with increased solubility **79**.

The noticeable increase in solubility allowed the meassurement of an H NMR spectrum of **79** in DCM. However, the maximum concentration was just 0.4 mM.

3.1.3. GC binder

The 2nd generation isolated GC binder was originally synthesized in our working group by Shubhankar Gadre.[147] The key steps of his approach are a Sonogashira cross-coupling reaction, followed by a base-catalyzed cyclization to form the pyrrole ring, which completes the construction of the polyheterocyclic skeleton of the GC binder (Scheme 38).



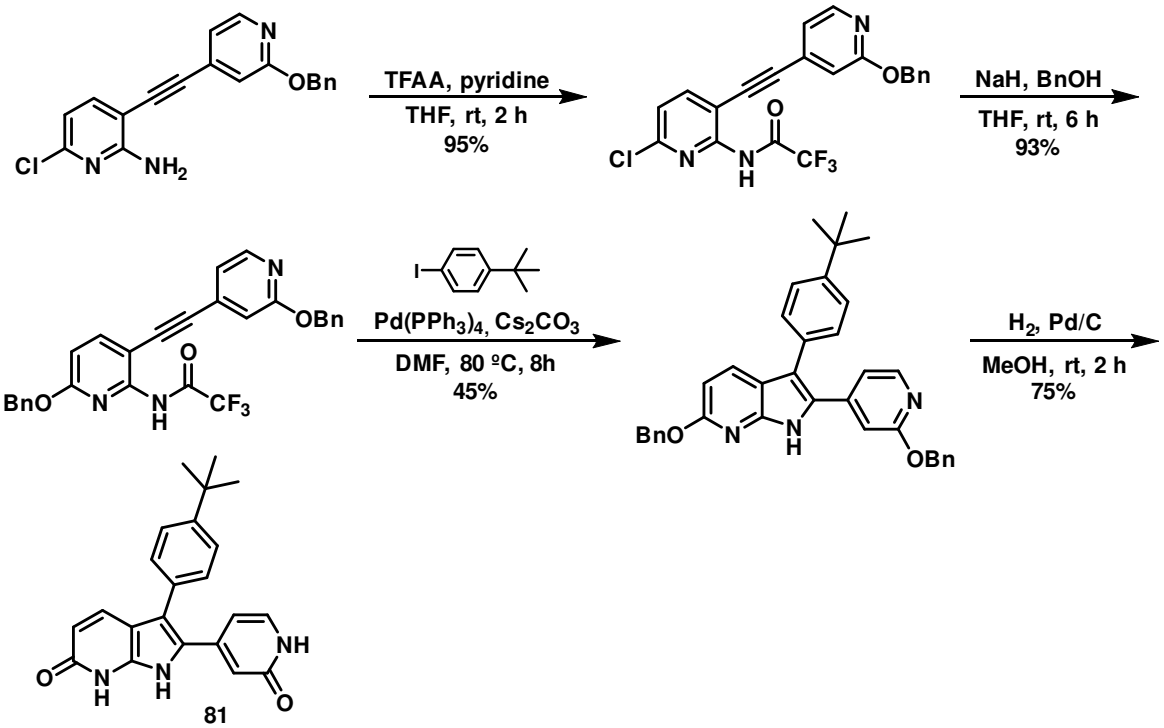
Scheme 38. Synthetic route developed by Gadre to obtain the isolated GC binder **80**.

This approach, however, relies on a low-yielding late-stage displacement of the chlorine atom by a benzyloxy group, making the route inefficient and hardly scalable. Despite this drawback, no further optimizations were explored, since it was found that the GC binder **80** was highly insoluble, even in polar solvents, most likely due to the presence of two pyridone rings. This proved to be a major problem, as the only purpose of the isolated GC binder is to estimate its binding potential with NMR titrations is a suitable organic solvent, such as DCM.

Gadre explored some additional functionalization of the GC binder to improve its solubility in organic solvents. The modified synthetic route featured a reaction developed by Cacchi *et al.*, which effects pyrrole cyclization, insertion of an aryl halide and trifluoroacetamide removal in a one-pot fashion (Scheme 39).[148]

[147] Gadre, S.; Peters, M. S.; Serrano, A.; Schrader, T. *Org. Lett.* **2018**, *20*, 6961–6964.

[148] (a) Arcadi, A.; Cacchi, S.; Marinelli F. *Tetrahedron Lett.* **1992**, *33*(27), 3915-3918. (b) Cacchi, S.; Fabrizi, G.; Parisi, L. M. *J. Comb. Chem.* **2005**, *7*(4), 510–512.



Scheme 39. Synthetic route developed by Gadre to obtain the modified GC binder **81**.

Unfortunately, although the solubility of the modified GC binder **81** in DCM was higher, it was not complete, making it still not suitable for NMR titrations. Gadre also experimented with different aryl halides containing alkyl chains of various lengths, but the debenzylated products after hydrogenolysis could not be successfully isolated. Further efforts towards a soluble GC binder were not considered.

Following the foundation laid by Gadre, as well as the experience acquired with the synthesis of the AT binder, a synthetic strategy towards the linker-attached GC binder was developed, which had not yet been successfully achieved. It features a pyrrolopyridone core linked to a second pyridone (Figure 28).

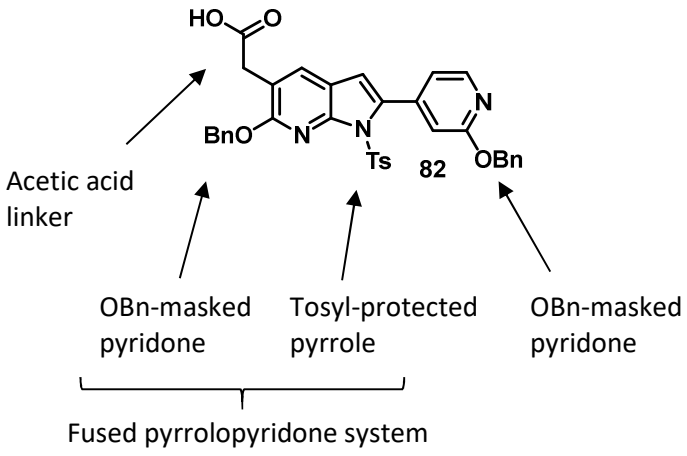
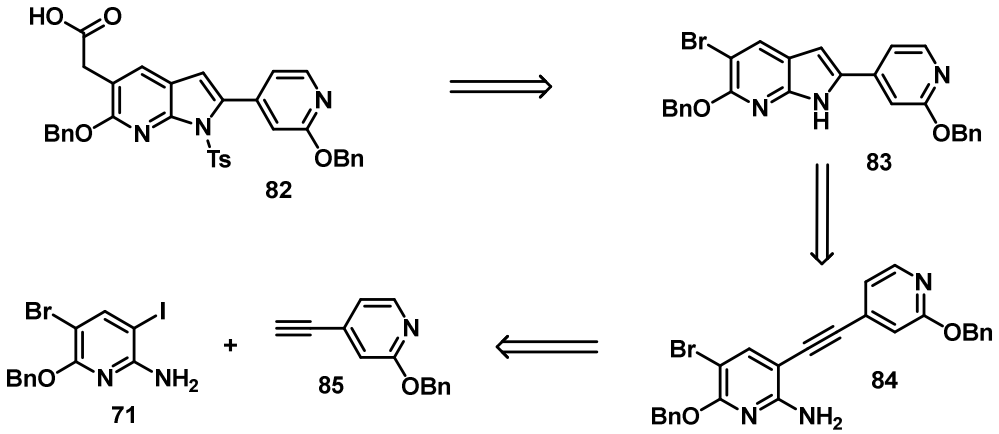


Figure 28. 2nd generation linker-attached GC binder **82**.

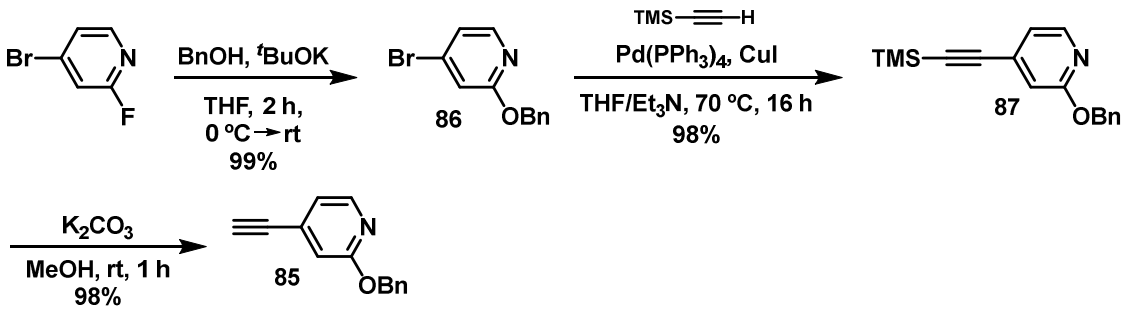
Both pyridones would be masked as benzyloxy groups during the synthesis, and the free pyrrole amino group would be protected with a tosyl group. The acetic acid linker would be the same as in case of the linker-attached AT binder.

For the construction of the target binder **82**, a brominated precursor **83** was envisioned, which could later be functionalized with the acetic acid linker. This molecule could be accessed following the same coupling-cyclization sequence established by Gadre, slightly optimising the reaction conditions. The previously synthesized di-heterohalogenated pyridine derivative **71** would be key for this approach, since it bears both iodine and bromine atoms, which could be orthogonally addressed in Pd-catalyzed cross-coupling reactions (Scheme 40).



Scheme 40. Retrosynthetic analysis towards the linker-attached GC binder **82**.

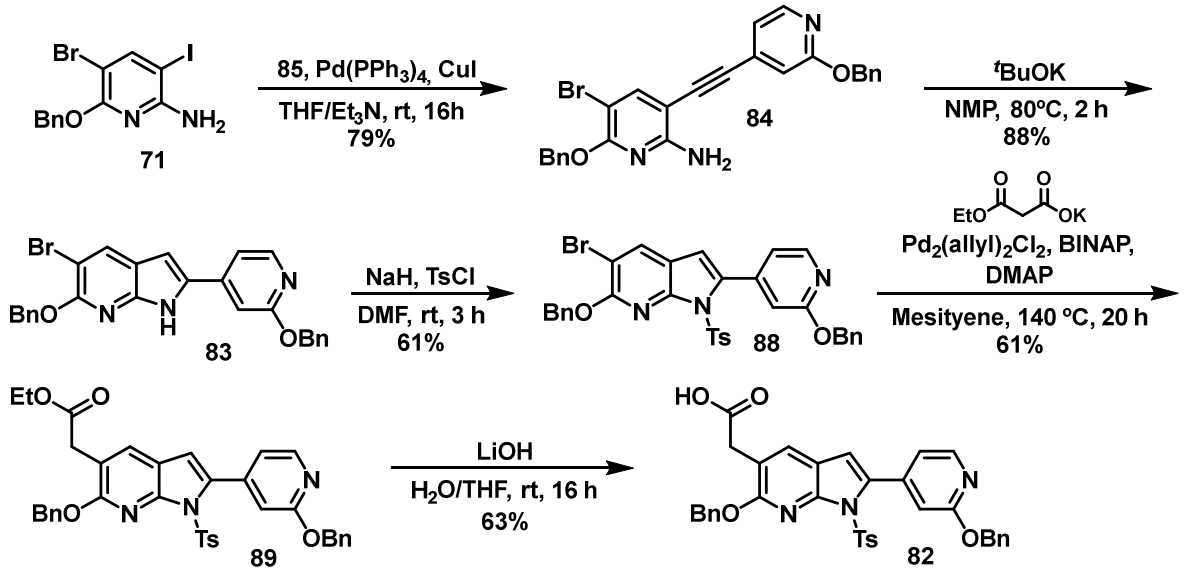
To this end, 4-bromo-2-fluoropyridine was subjected to a nucleophilic substitution of one of the halogens with a benzyloxy group, which occurred selectively on the fluorine atom due to its superior reactivity, affording **86** (Scheme 41).[149] Then, TMS-acetylene was introduced with palladium catalysis, replacing the remaining halogen, followed by the removal of the TMS group under mild catalytic basic conditions to obtain the ethynylpyridine derivative **85**.



Scheme 41. Synthetic strategy to obtain the ethynylpyridine derivative **85**.

[149] Holenz, J.; Karlström, S.; Kihlström, J. *et al.*; AstraZeneca AB. World patent WO 2011/002409 A1. 2016 Jan 6.

A Sonogashira cross-coupling reaction was used to join **85** to the pyridine derivative **71**, which selectively displaced the iodine atom taking advantage of the different reactivity of the halogens, yielding **84**. A base-catalyzed cyclization provided the brominated GC binder skeleton **83**. Next, the pyrrole -NH group was masked with a tosyl group, which proceeded with a surprisingly modest yield. This step was, however, essential for the successful introduction of the linker, as the free amine would have a negative impact on the linker introduction reaction. Thus, the tosyl-protected derivative **88** was subjected to a decarboxylative cross-coupling reaction, yielding **89**, followed by the hydrolysis of the ester to afford the linker-attached GC binder **82** (Scheme 42).



Scheme 42. Synthetic pathway to obtain the target linker-attached GC binder **82**.

Following this approach, the desired binder **82** was obtained with an overall yield of 15% over 8 steps. Although this route features multiple high-yielding transformations, there are some steps with room for optimization, such as the tosyl protection or the ester hydrolysis, the latter having a challenging purification due to the acetic acid linker.

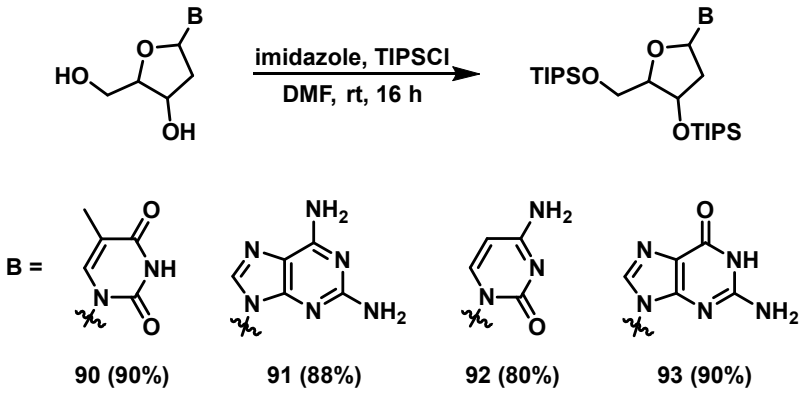
Coupling **82** with the PNA backbone would afford the GC binder PNA monomer, which could be used as a building block for the construction sequence-selective PNA oligomers, in combination with the rest of the binders. The benzyl and tosyl protecting groups should be kept during the SPPS of the oligomer to avoid any side-reactions, and can be removed under acidic conditions afterwards.

3.1.4. TA/CG binders

The synthesis and characterization of the TA and CG binders, both in their isolated and linker-attached forms, were carried out by Nahid S. Alavijeh, and will thus not be reported in this thesis.

3.1.5. Silylated nucleosides

To perform NMR titrations in suitable organic solvents, such as dichloromethane or chloroform, the base pairs also needed to be soluble in these solvents. For this purpose, nucleosides were employed, which were silylated at the two available hydroxy groups of the deoxyribose sugar, introducing bulky alkyl groups. Each nucleoside was treated with imidazole and TIPS chloride, which afforded the desired silylated nucleosides TIPS-thymidine **90** (T), TIPS-deoxyaminoadenosine **91** (AA), TIPS-deoxycytidine **92** (C), TIPS-deoxyguanosine **93** (G) (Scheme 43).[150]



Scheme 43. Synthesis of the sylilated nucleosides T (**90**), AA (**91**), C (**92**) and G (**93**).

3.2. NMR titrations

NMR spectroscopy is an excellent tool for the determination of the binding strength of isolated nucleobases in organic solvents, as hydrogen bonding can be directly analyzed through the variation in the chemical shifts of the involved protons. Although the obtained results may not translate into the formation of stable complexes at physiological conditions,[109] these measurements can provide valuable preliminary information about the binding capabilities of the binders.

For this purpose, the soluble silylated nucleosides synthesized previously (Section 3.1.5) would be used to determine the binding affinity of the isolated AT binder **15**. A modified aminoadenosine (AA) nucleoside would be use to provide an additional hydrogen bond with Thymidine (T), potentially strengthening their union. This base pair was experimentally shown by Peters to have a binding affinity of 253 M^{-1} at room temperature, determined with NMR spectroscopy in DCM (Figure 29).

[150] (a) Saudi, M.; Zmurko, J.; Kaptein, S.; Rozenski, J.; Neyts, J.; Van Aerschot, A. *Eur. J. Med. Chem.* **2014**, *76*, 98-109. (b) Burnham, B. S.; Wyrick, S. D.; Hall, I. H.; Sood, A.; Spielvogel, B. F. *J. Labelled Comp. Radiopharm.* **1991**, *29*(4), 469-473. (c) Marsh, A.; Alcock, N. W.; Errington, W.; Sagar, R. *Tetrahedron* **2003**, *59*(29), 5595-5601. All nucleosides were synthesided following the same modified procedure.

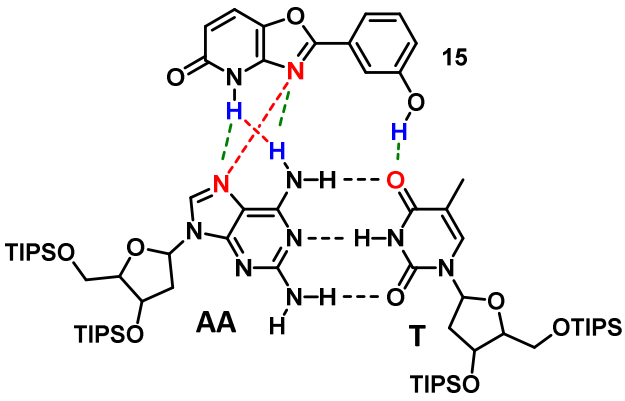


Figure 29. Hydrogen bond pattern of the AT binder **15**. Green dashed bonds represent the primary hydrogen bonds. Red dashed bonds represent repulsive secondary interactions.

It is essential to analyze the hydrogen bond pattern of **15**, which features a **DAD** arrangement. This alternating pattern of the hydrogen donors and acceptors provides the weakest contribution to the overall stability of the complex, according to the secondary interactions of neighboring atoms proposed by Jorgensen *et al.*.[151] In the particular case of **15**, the pyridone hydrogen and the neighboring oxazole N would suffer from these repulsive secondary interactions, since both of them find diagonally an atom with the same hydrogen bonding nature (donor or acceptor), which is not favorable. The same is true for the AA-T base pair, also presenting a **DAD** arrangement where all the possible secondary interactions are repulsive.

However, the poor solubility of **15** limited the choice of solvents for NMR titrations to just THF, which is not best suited for this measurements. In this solvent, it was not possible to obtain a binding curve for the titration of AA against T. It is imperative that the nucleosides form a strong pair to be able to treat them as a single species in the subsequent titration with the binder. Therefore, it was clear that an alternative had to be found.

To this end, the slightly more soluble modified AT binder **79** was prepared. The solubility of **79** in DCM was just of 0.4 mM which, despite being low, was enough to measure an H NMR. However, this spectrum revealed new peaks that were not present when the experiment was carried out in polar solvents such as DMSO. This observations were attributed to the potential self-association between pyridone rings, which is a well-known behavior in non-polar solvents.[152] Interestingly, when a 1:1:1 complex of **79** and the AA-T

[151] (a) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, 112, 5, 2008-2010. (b) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, 113, 8, 2810–2819.

[152] (a) Loppinet-Serani, A.; Charbonnier, F.; Rolando, C.; Huc, I. *J. Chem. Soc.*, Perkin Trans. 2, **1998**, 937-942. (b) Beak, P.; Covington, J. B.; Smith, S. G.; White, J. M.; Zeigler, J. M. *J. Org. Chem.* **1980**, *45*, 1354-1362.

base pair was analyzed by NMR, it was noticed that the small peaks had disappeared, possibly due to the dissolution of the self-association (Figure 31).

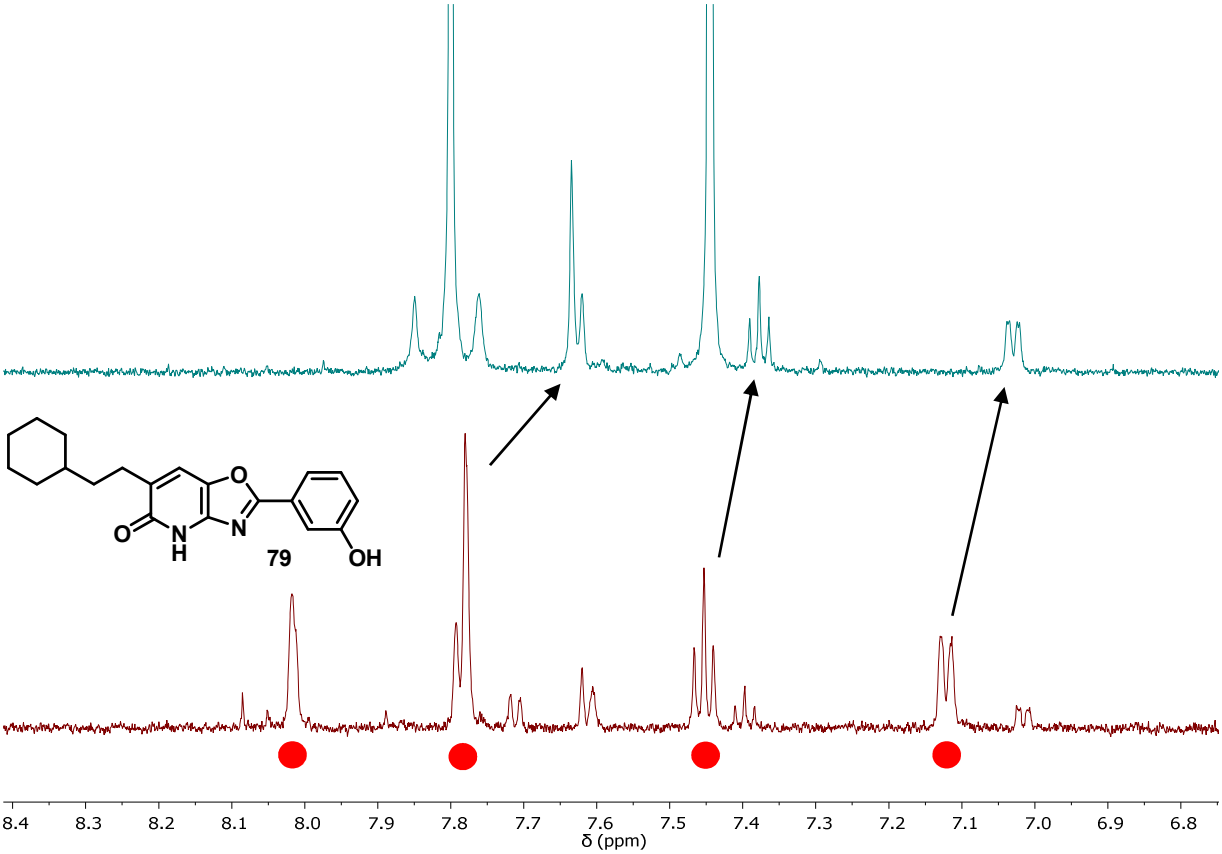


Figure 30. Bottom: aromatic region of an H NMR spectrum of **79** in DCM. The peaks marked with red circles belong to one molecule of **79**, while the minor peaks might be due to self-association. Top: aromatic region of an H NMR spectrum of a 1:1:1 complex of **79** and the AA-T base pair

It is of high importance to determine the nature of these interactions before performing NMR titrations, as the results could be affected and not be representative. The evaluation of this complex with a 2D NOE spectrum would provide valuable information, as intermolecular NOE cross peaks between protons would evidence their proximity, thus confirming the formation of the expected complex.[112] However, the measurement of 2D NOE spectra was not possible due to the low maximum concentration of **79** in DCM.

3.3. Photophysical properties

All the synthesized binders showed fluorescent properties. Specifically, the isolated AT binder **15** displayed a strong blue fluorescence when irradiated with a UV lamp at 365 nm (Figure 31).

Figure 31. Fluorescence of 100 µL solutions of the isolated AT binder **15** in THF (left) and MeOH (right), irradiated with a 365 nm UV lamp.

The fluorescence properties of **15** were characterized measuring its absorption and emission spectra (Figure 32).

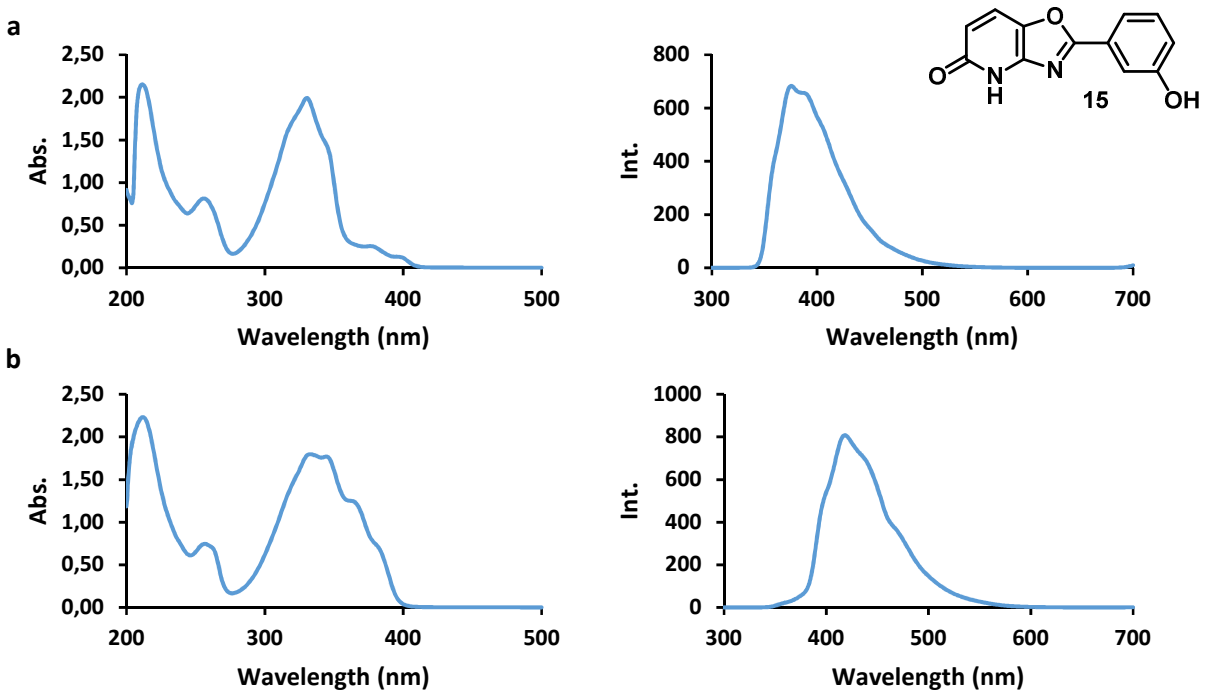


Figure 32. a) Absorption (left) and emission (right) spectra of a 100 µL solution of **15** in THF. b) Absorption (left) and emission (right) spectra of a 100 µL solution of **15** in MeOH. In the emission spectra, the excitation wavelength was fixed to the maximum absorption wavelength determined with the absorption spectra.

Additionally, the fluorescence quantum yield (Φ_f) was analyzed, which is the direct measure for the efficiency of the conversion of absorbed photons into emitted photons, with possible values ranging from 0 to 1.0. The results are summarized in Table 1, which suggest that this binder has the potential to be used as a fluorescent probe. Fluorescent nucleobases that

produce a distinguishable signal upon binding to the target DNA sequence have found a variety of biological and biomedical applications.[153]

Table 1. Photophysical properties of **15**.

Solvent	λ_{abs} (nm)	λ_{em} (nm)	Φ_f
THF	330	375	0.94
MeOH	344	418	0.46

[153] (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579-2619. (b) Krishna, M. S.; Toh, D.-F. K.; Meng, Z.; Ong, A. A. L.; Wang, Z.; Lu, Y.; Xia, K.; Prabakaran, M.; Chen, G. *Anal. Chem.* **2019**, *91*, 5331-5338. (c) Vilaivan, T. *Beilstein J. Org. Chem.* **2018**, *14*, 253-281.

4. Conclusions and Outlook

In this work, the synthesis and characterization of various binders was achieved, which brings us one step closer towards the goal of sequence-selective DNA recognition. The isolated AT binder **15** was successfully prepared with an overall yield of 25% over 7 steps. The synthetic methodologies employed laid the foundation for the construction of the remaining binders, which share some structural characteristics.

The preliminary determination of its binding potential with NMR titrations was not possible, due to its limited solubility. Only THF was able to fully dissolve **15**, but titrations in this solvent proved to be challenging. A modified AT binder **79** with a solubilizing residue was prepared and, although its solubility in DCM was significantly higher, it was still not enough for the determination of the binding affinity with NMR techniques.

The exceptional fluorescent properties of the isolated AT binder **15** were characterized measuring its absorption and emission spectra, as well as quantum yields.

Additionally, the linker-attached AT and GC binders (**73** and **82**) were synthesized, both with an overall yield of 15% over 7 and 8 steps, respectively. Both binders contain an acetic acid liker that can be attached to a PNA backbone through an amidation reaction. The linker-attached AT binder was successfully coupled to an aegPNA backbone, obtaining the AT binder PNA monomer that can be used for the construction of oligomers.

For the success of this work, it is essential to prepare the remaining PNA monomers. With the four fundamental building blocks (AT, TA, GC, CG PNA monomers) in hand, several oligomers (ideally 8-10mers) can be prepared, with a variety of sequences. As a first experiment, a homo-oligomer employing just one of the monomers should be prepared, the binding affinity of which would be tested against a fully complementary strand of DNA. Then, mixed sequences can be prepared, to recognize sequences with both perfect complementarity and some mismatches. A combination of the binding affinities and their variations when mismatches are introduced will determine the sequence-selectivity of the oligomers.

5. Experimental details

5.1. Materials and methods

5.1.1. Solvents and chemicals

Chemicals and solvents were purchased from the companies Acros Organics, Alfa Aesar, Bachem, Carbolution Chemicals, Fluka, Fluorochem, Sigma Aldrich, TCI chemicals and Thermo Fisher Scientific, and were used as provided without any additional purification, unless otherwise noted. All anhydrous solvents were commercially obtained, except for THF, which was distilled under standard benzophenone/sodium conditions.

5.1.2. Chromatography

The progress of the reactions was monitored with thin layer chromatography silica gel sheets type F254 (Macherey-Nagel), with a layer thickness of 0.2 mm. Compounds were detected with UV light at 254 nm and 365 nm. UV-inactive compounds were visualized using a potassium permanganate stain. Products were purified using flash column chromatography (Merck Silica gel 60, 230-400 mesh).

5.1.3. Nuclear magnetic resonance (NMR) spectroscopy

NMR spectra were recorded at room temperature using the following spectrometers: DMX300 (300 MHz), AV NEO 400 (400 MHz), DRX500 (500 MHz) and DRX600 (600 MHz), all of them from Bruker. Samples were prepared as solutions in deuterated solvents and referenced to the internal non-deuterated solvent peak. Chemical shifts were expressed in ppm (δ) downfield of tetramethylsilane. Coupling constants are given in Hertz (Hz). The following abbreviations were used: singlet (s), broad singlet (br s), doublet (d), doublet of doublets (dd), doublet of doublet of doublets (ddd), doublet of triplets (dt), doublet of quartets (dq), triplet (t), triplet of doublets (td), quartet (q), multiplet (m). The spectra were analyzed using the software MestReNova from Mestrelab Research (version 9.0.1).

5.1.4. High resolution mass spectrometry (HRMS)

HRMS was performed on a Bruker Maxis 4G using electrospray as ionization source (ESI) in positive or negative mode. The eluent was introduced into the ion course of the mass spectrometer. Nitrogen was used as nebulizing gas and dry gas. The nebulizer gas pressure was set to 29.0 psi and the dry gas was set to 5.0 L/min at 250 °C. The capillary voltage was set to - 4500 V. The end plate offset was set to -500 V. The acquisition was performed in scan mode in the m/z range 100 – 2900. The MS system was controlled by the software

Hystar version 3.2. The data was analysed using the software compass data analysis version 4.1.

5.1.5. High-performance liquid chromatography (HPLC)

HPLC measurements were performed on a Jasco with a UV-Vis detector (UV-975, DG-2080-53 solvent degasser, LG-980-02S 3-channel solvent mixer). A column from Phenomenex (model Luna® 5µM C18 (2) 100, Å 100/10 mm) was used, eluting with a gradient of water / acetonitrile with 0.1% TFA (A) for 60 minutes.

5.1.6. Melting point

Melting points were measured on a B-540 Büchi apparatus.

5.1.7. UV-Vis and fluorescence spectroscopy

Uv-Vis spectra were recorded in a V-550 spectrometer from Jasco, using 100 µL solutions in a quartz cuvette with a width of 1 cm.

Fluorescence spectra were recorded in a FP-6500 spectrometer from Jasco, using 100 µL solutions in a round quartz cuvette with a width of 1 cm. The excitation wavelength was fixed to the maximum absorption wavelength determined with the absorption spectrum.

5.1.8. Fluorescence quantum yield

Fluorescence quantum yields were measured with a RF-6000 spectrofluorophotometer from Shimadzu equipped with an integrating sphere. 100 µL solutions were used, measured in a Suprasil quartz Semi-Micro Cell Quvette from Hellma Analytics.

5.1.9. NMR titrations

NMR titrations were performed on a DRX500 spectrometer. The titration consists of two compounds, A and B. An NMR tube was filled with a 5mM solution of A in the deuterated solvent of choice. A stock solution containing 5mM of A and 50 mM of B in the deuterated solvent of choice was prepared, from which aliquots were taken and added to the NMR tube after each measurement. 18 H NMR spectra were measured, after adding the following additional amount of the stock solution B (in µL): 0, 10, 20, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 210, 240, 270, 300, 350. Once the data was collected, variations in the chemical shift of the different protons was analyzed.

5.1.10. General method for Solid-phase peptide synthesis (SPPS)

A wang resin preloaded with a Fmoc-protected amino acid was used to prepare the peptides by manual Fmoc solid-phase peptide synthesis. In order to ensure complete conversion or deprotection, the resin had to be swollen in DMF for 120 minutes before the actual solid phase synthesis started. 6.00 eq. PNA monomer, 5.72 eq. HBTU and 12.00 eq. DIPEA were used per coupling step. A 20% piperidine solution in DMF was used to cleave the Fmoc protecting group. The completeness of the individual coupling steps was checked with a Kaiser test, consisting of a solution of 75 µl phenol (80% phenol in ethanol), 100 µl KCN solution (2 mL of a 1 mM aqueous KCN solution in 98 mL pyridine) and 75 µl ninhydrin solution (5% ninhydrin in ethanol (w / v)). For this purpose, some resin balls were removed, washed thoroughly with DMF and mixed with the Kaiser test solution. The test solution was heated for 5 minutes at 90 °C in a thermomixer. A blue color indicates the completion of the coupling. If the reaction was incomplete (yellow color), the respective coupling step was repeated and checked again. Steps 2 and 3 (Table 6) were repeated until the desired oligomer was synthesized. Then, the resin was dried (step 4).

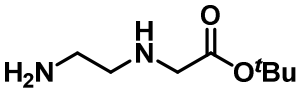
Table 2. Steps in the manual solid phase peptide synthesis.

Reaction step	Reagents	Time
1. Resin swelling	3 mL DMF	120 Minutes
2. Fmoc cleavage	1. 20% Piperidine in DMF 2. 20% Piperidine in DMF 3. Wash with 5x 500 µL DMF	1. 3 Minutes 2. 10 Minutes
3. Coupling	6.00 eq. PNA monomer in 5.72 eq. HBTU solution (0.5 M in DMF) 12.00 eq. DIPEA	1 h
4. Dehydration	1. 5x 500 µL DMF 2. 5x 500 µL DCM 3. 5x 500 µL Diethylether	

The cleavage cocktail containing trifluoroacetic acid (93%), triisopropylsilane (5%) and distilled water (2%) was added to the resin and stirred for 5 h at room temperature. The oligomer was precipitated from cold diethyl ether and washed three times with cold diethyl ether. Then, it was dissolved in 0.1% aqueous TFA and freeze-dried. The purity was checked by HPLC, NMR and mass spectrometry.

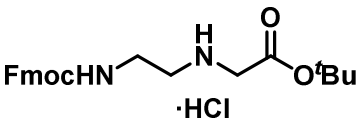
5.2. Synthetic procedures

***tert*-Butyl *N*-(2-aminoethyl)glycinate (1)**[114]



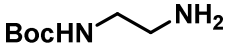
A solution of *tert*-butyl bromoacetate (20.00 g, 15.14 mL, 102.53 mmol, 1eq.) in dichloromethane (75 mL) was added dropwise over 5 hours to a vigorously stirred solution of ethylenediamine (49.30 g, 54.84 mL, 820.30 mmol, 8 eq.) in dichloromethane (365 mL) at 0 °C. The reaction mixture was allowed to warm slowly to room temperature and was further stirred for 24 h. Then, it washed with water (3 x 100 mL) and the combined aqueous phases were back-extracted with dichloromethane (100 mL). The combined organic extracts were dried over Na_2SO_4, filtered and evaporated under reduced pressure to afford **1** as a pale-yellow oil (15.14 g, 86.89 mmol, 85% yield). **R_f**: 0.09 (EtOAc/MeOH 9:1). **^{1}H NMR** (300 MHz, $CDCl_3$) δ (ppm): 3.26 (s, 2H), 2.76 (td, *J* = 6.4, 1.3 Hz, 2H), 2.65 (td, *J* = 6.5, 1.4 Hz, 2H), 1.90 (s, 3H), 1.42 (s, 9H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ (ppm): 172.06, 81.28, 52.24, 51.76, 41.89, 28.25. H**RMS** (ESI, m/z) calculated for $C_8H_{19}N_2O_2$ $[M + H]^+$: 175.1441, found: 175.1439; calculated for $C_8H_{18}N_2O_2Na$ $[M + Na]^+$: 197.1260, found: 197.1259.

***tert*-Butyl *N*-(2-(fluorenylmethoxycarbonyl)aminoethyl)glycinate hydrochloride (2)**[114]



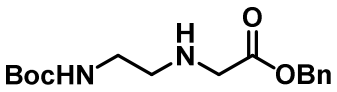
A solution of Fmoc succinimide (26.18 g, 77.61 mmol, 1eq.) in dichloromethane (145 mL) was added dropwise over 5 hours to a stirred solution of **1** (14.20 g, 81.49 mmol, 1.05 eq.) and DIPEA (10.03 g, 13.52 mL, 77.61 mmol, 1eq.) in dichloromethane (580 mL), and stirring was continued at room temperature for 18 hours. The reaction mixture was washed with 1 M HCl (5 x 100 mL) and brine (100 mL), dried over Na_2SO_4, filtered, and concentrated under reduced pressure to approximately 240 mL. After cooling overnight at -20 °C, the formed precipitate was collected by filtration, washed with cold dichloromethane until a colourless filtrate was obtained and dried under vacuum to afford **2** as a white solid (21.03 g, 48.57 mmol, 63% yield). **R_f**: 0.12 (DCM/MeOH 98:2). **m.p.**: 124.3 – 127.5 °C. **^{1}H NMR** (300 MHz, DMSO-d_6) δ (ppm): 9.46 (br s, 2H), 7.89 (d, *J* = 7.3 Hz, 2H), 7.70 (d, *J* = 7.3 Hz, 2H), 7.62 (t, *J* = 5.3 Hz, 1H), 7.41 (t, *J* = 7.1 Hz, 2H), 7.33 (t, *J* = 7.0 Hz, 2H), 4.32 (d, *J* = 6.3 Hz, 2H), 4.22 (t, *J* = 6.7 Hz, 1H), 3.86 (s, 2H), 3.35 (q, *J* = 5.9 Hz, 2H), 3.01 (t, *J* = 5.8 Hz, 2H), 1.45 (s, 9H). **^{13}C NMR** (75 MHz, DMSO-d_6) δ (ppm): 165.47, 156.19, 143.75, 140.68, 127.58, 127.02, 125.17, 120.06, 82.87, 65.65, 54.90, 47.17, 46.62, 46.36, 36.57, 27.57. **HRMS** (ESI, m/z) calculated for $C_{23}H_{29}N_2O_4$ $[M + H]^+$: 397.2122, found: 397.2119.

***tert*-Butyl *N*-(2-aminoethyl)carbamate (3)**[115]



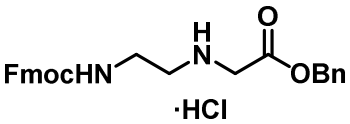
A solution of boc anhydride (10.00 g, 45.82 mmol, 1 eq.) in THF (80 mL) was added dropwise over 45 minutes to a vigorously stirred solution of ethylenediamine (55.07 g, 61.26 mL, 916.38 mmol, 20 eq.) in THF (120 mL) at room temperature. After 18 h of additional stirring, the reaction was quenched by addition of water (100 mL). The aqueous phase was saturated with solid K_2CO_3, the phases were separated and the organic phase was dried over Na_2SO_4, filtered, and evaporated under reduced pressure to give a pale-yellow oil. The oil was dissolved in toluene (2 x 150 mL) and concentrated under vacuum to remove the remaining ethylenediamine, affording **3** as a pale-yellow oil (6.44 g, 40.19 mmol, 88% yield). **R_f**: 0.17 (EtOAc/MeOH 9:1). **^{1}H NMR** (300 MHz, $CDCl_3$) δ (ppm): 5.00 (br s, 1H), 3.13 (q, *J* = 5.9 Hz, 2H), 2.75 (t, *J* = 5.9 Hz, 2H), 1.41 (s, 9H), 1.21 (s, 2H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ (ppm): 156.33, 79.25, 77.58, 77.16, 76.74, 43.55, 41.98, 28.49. **HRMS** (ESI, m/z) calculated for $C_7H_{17}N_2O_2$ $[M + H]^+$: 161.1285, found: 161.1284; calculated for $C_7H_{16}N_2O_2Na$ $[M + Na]^+$: 183.1104, found: 183.1102.

Benzyl *N*-(2-(*tert*-butoxycarbonyl)aminoethyl)glycinate (4)[115]



Benzyl bromoacetate (7.53 g, 5.21 mL, 32.89 mmol, 1eq.) was added dropwise over 5 minutes to a solution of **3** (5.27 g, 32.89 mmol, 1 eq.) and triethylamine (3.33 g, 4.59 mL, 32.89 mmol, 1 eq.) in acetonitrile (85 mL). After 2 hours, the reaction mixture was diluted with ethyl acetate (100 mL) and washed with 2 M aq. K_2CO_3 (100 mL) and brine (100 mL). The organic phase was dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc/Et_3N 1:1:0.01 → 1:4:0.01) to afford **4** as a pale-yellow solid (6.59 g, 21.36 mmol, 65% yield). **R_f**: 0.18 (Cy/EtOAc 1:1). **^{1}H NMR** (300 MHz, $CDCl_3$) δ (ppm): 7.38 – 7.29 (m, 5H), 5.15 (s, 2H), 5.06 (br s, 1H), 3.44 (s, 2H), 3.19 (q, *J* = 5.7 Hz, 2H), 2.73 (t, *J* = 5.8 Hz, 2H), 1.71 (s, 1H), 1.43 (s, 9H). **^{13}C NMR** (75 MHz, $CDCl_3$) δ (ppm): 172.45, 156.16, 135.64, 128.71, 128.52, 128.44, 79.25, 66.71, 50.57, 48.87, 40.28, 28.50. **HRMS** (ESI, m/z) calculated for $C_{16}H_{25}N_2O_4$ $[M + H]^+$: 309.1809, found: 309.1807; calculated for $C_{16}H_{24}N_2O_4Na$ $[M + Na]^+$: 331.1628, found: 331.1629.

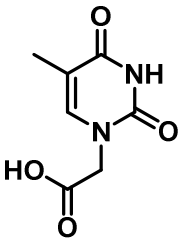
Benzyl *N*-(2-(fluorenylmethoxycarbonyl)aminoethyl)glycinate hydrochloride (5)[115]



To a solution of **4** (4.15 g, 13.45 mmol, 1 eq.) in dichloromethane (20 mL), trifluoroacetic acid (20 mL) was added at 0 °C. The reaction mixture was allowed to warm to

room temperature, and was further stirred for 20 minutes. Then, the reaction mixture was concentrated under vacuum, redissolved in toluene (2 x 100 mL) and concentrated under vacuum to remove the remaining trifluoroacetic acid. The resulting pale-yellow oil was dissolved in dichloromethane (80 mL) under argon and the solution was cooled to 0 °C. Fmoc succinimide (4.54 g, 13.45 mmol, 1 eq.) was added in one portion, and then triethylamine (4.08 g, 5.62 mL, 40.37 mmol, 3 eq.) was added dropwise over 5 minutes. The reaction mixture was allowed to warm to room temperature, and was further stirred for 2 hours. Then, it was washed with 1 M aq. K_2CO_3 (80 mL), and brine (80 mL). The organic phase was dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 1:3 → EtOAc). The obtained yellow oil was dissolved in diethyl ether (50 mL) and cooled to 0 °C, and a 2 M HCl solution in diethyl ether (18 mL) was added dropwise. After cooling at -20 °C for 2 hours, the formed precipitate was collected by filtration, rinsed with cold diethyl ether (25 mL) and dried under vacuum to afford **5** as a white solid (4.27 g, 9.14 mmol, 68% yield over two steps). **R_f**: 0.13 (DCM/MeOH 98:2). **m.p.**: 166.3 – 167.6 °C. **^{1}H NMR** (300 MHz, DMSO-*d*$_6$) δ (ppm): 9.46 (br s, 2H), 7.89 (d, *J* = 7.4 Hz, 2H), 7.69 (d, *J* = 7.3 Hz, 2H), 7.57 (t, *J* = 5.5 Hz, 1H), 7.37 (dq, *J* = 19.6, 7.7 Hz, 9H), 5.25 (s, 2H), 4.33 (d, *J* = 6.8 Hz, 2H), 4.22 (t, *J* = 6.6 Hz, 1H), 4.07 (s, 2H), 3.34 (s, 2H), 3.03 (t, *J* = 5.6 Hz, 2H). **^{13}C NMR** (75 MHz, DMSO-*d*$_6$) δ (ppm): 166.53, 156.26, 143.78, 140.72, 135.11, 128.45, 128.36, 128.21, 127.61, 127.05, 125.11, 120.11, 66.88, 65.61, 46.68, 46.63, 46.52, 36.57. **HRMS** (ESI, m/z) calculated for $C_{26}H_{27}N_2O_4$ $[M + H]^+$: 431.1965, found: 431.1964; calculated for $C_{26}H_{26}N_2O_4Na$ $[M + Na]^+$: 453.1785, found: 453.1783.

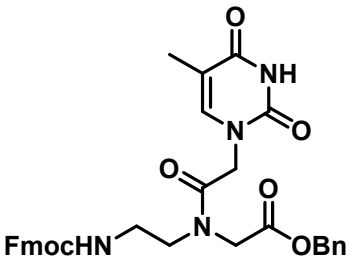
Thymin-1-yl acetic acid (6)[117]



Thymine (10.00 g, 79.29 mmol, 1 eq.) was dissolved in a solution of potassium hydroxide (16.90 g, 301.32 mmol, 3.8 eq.) in water (50 mL). While this solution was warmed to 40 °C, a solution of bromoacetic acid (16.53 g, 118.89 mmol, 1.5 eq.) in water (25 mL) was added over 30 minutes. The reaction was stirred for another 30 minutes at this temperature. It was allowed to cool to room temperature and the pH was adjusted to 5.5 with conc. HCl. The solution was then cooled in a refrigerator for 2 hours. Any precipitate (unreacted thymine) formed was removed by filtration. The solution was then adjusted to pH 2 with conc. HCl and stored in a freezer for 2 hours. The white precipitate was collected by filtration and dried under vacuum to afford **6** as a white solid (12.40 g, 67.33 mmol, 85% yield). **R_f**: 0.30 (EtOAc/MeOH 9:1). **m.p.**: 260.5 – 262.3 °C. **^{1}H NMR** (300 MHz, DMSO-*d*$_6$) δ (ppm): 13.09 (br s, 1H), 11.33 (s, 1H), 7.48 (s, 1H), 4.36 (s, 2H), 1.75 (s, 3H). **^{13}C NMR** (75 MHz, DMSO-*d*$_6$) δ (ppm): 169.66, 164.37, 151.00, 141.80, 108.37,

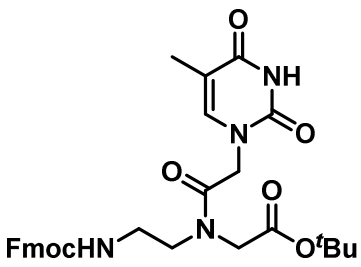
48.41, 11.89. **HRMS** (ESI, m/z) calculated for $C_7H_9N_2O_4$ [M + H]$^+$: 185.0557, found: 185.0552; calculated for $C_7H_8N_2O_4Na$ [M + Na]$^+$: 207.0376, found: 207.0370.

Benzyl *N*-(2 - (fluorenylmethoxycarbonyl) aminoethyl) - *N* - (thymin - 1 - yl) acetyl) glycinate (7)[114]



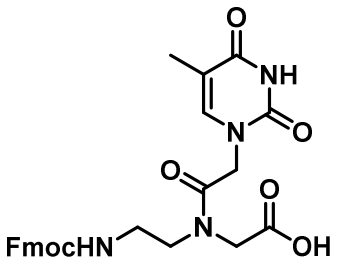
5 (1.06 g, 2.27 mmol, 1eq.) was suspended in ethyl acetate (100 mL), the mixture was washed with sat. aq. $NaHCO_3$ (4 x 50 mL) until the solids dissolved and the combined aqueous phases were back-extracted with ethyl acetate (50 mL). The combined organic extracts were washed with brine (50 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The obtained oil was dissolved in anhydrous DMF (13 mL), and **6** (878 mg, 4.77 mmol, 2.1 eq.) was added under an argon atmosphere. EDC·HCl (914 mg, 4.77 mmol, 2.1 eq.) was added in two equal portions over 30 min, and the mixture was stirred at room temperature for 18 hours. The solution was concentrated under vacuum, and water (25 mL) was added to the resulting thick oil. The mixture was shaken vigorously for several minutes until a precipitate was formed. The solids were collected by filtration, stirred with water (25 mL), filtered, washed with cold water, and dried under vacuum. Purification by column chromatography (Cy/EtOAc 1:4 → 1:6) afforded **7** as a white solid (1.09 g, 1.84 mmol, 81% yield). **R_f**: 0.16 (Cy/EtOAc 1:4). **m.p.**: 173.8 – 174.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 11.30 (s, 1H), 7.88 (d, J = 7.5 Hz, 2H), 7.71 – 7.60 (m, 2H), 7.45 – 7.25 (m, 11H), 5.16 (d, J = 31.3 Hz, 2H), 4.67 (s, 1H), 4.50 – 4.12 (m, 5H), 3.49 – 3.34 (m, 2H), 3.18 (dq, J = 59.4, 6.1 Hz, 2H), 1.73 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 169.33, 168.90, 167.68, 167.48, 164.39, 156.34, 156.13, 150.98, 142.04, 141.90, 140.74, 135.76, 135.55, 128.49, 128.45, 128.13, 128.08, 127.89, 127.62, 125.12, 120.12, 108.22, 108.11, 66.63, 65.97, 65.47, 65.36, 49.04, 47.98, 47.83, 47.60, 46.95, 46.73, 11.93. A precise NMR characterization was not possible due to the existence of multiple rotomers and peak overlapping. **HRMS** (ESI, m/z) calculated for $C_{33}H_{33}N_4O_7$ [M + H]$^+$: 597.2344, found: 597.2350; calculated for $C_{33}H_{32}N_4O_7Na$ [M + Na]$^+$: 619.2163, found: 619.2171.

***tert*-Butyl *N*-(2 - (fluorenylmethoxycarbonyl) aminoethyl) - *N*- (thymin - 1 - yl) acetyl) glycinate (8)**[114]



2 (895 mg, 2.07 mmol, 1eq.) was suspended in ethyl acetate (100 mL), the mixture was washed with sat. aq. $NaHCO_3$ (4 x 50 mL) until the solids dissolved and the combined aqueous phases were back-extracted with ethyl acetate (50 mL). The combined organic extracts were washed with brine (50 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The obtained oil was dissolved in anhydrous DMF (12 mL), and **6** (799 mg, 4.34 mmol, 2.1 eq.) was added under an argon atmosphere. EDC·HCl (832 mg, 4.34 mmol, 2.1 eq.) was added in two equal portions over 30 min, and the mixture was stirred at room temperature for 18 hours. The solution was concentrated under vacuum, and water (22 mL) was added to the resulting thick oil. The mixture was shaken vigorously for several minutes until a precipitate was formed. The solids were collected by filtration, stirred with water (22 mL), filtered, washed with cold water, and dried under vacuum. Purification by column chromatography (Cy/EtOAc 1:5 → 1:6) afforded **8** as a white solid (895 mg, 1.59 mmol, 77% yield). **R_f**: 0.19 (Cy/EtOAc 1:5). **m.p.**: 175.2 – 175.8 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 11.29 (d, J = 5.3 Hz, 1H), 7.89 (d, J = 7.5 Hz, 2H), 7.68 (dd, J = 7.9, 2.6 Hz, 2H), 7.44 – 7.24 (m, 6H), 4.64 (s, 1H), 4.45 (s, 1H), 4.32 (dd, J = 15.1, 6.9 Hz, 2H), 4.24 – 4.16 (m, 1H), 3.94 (s, 1H), 3.40 (t, J = 6.6 Hz, 1H), 3.17 (dq, J = 57.6, 6.2 Hz, 2H), 1.73 (d, J = 1.1 Hz, 3H), 1.45 (s, 3H), 1.39 (s, 6H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 168.42, 168.03, 167.62, 167.27, 164.40, 156.32, 156.12, 150.97, 143.88, 143.85, 142.11, 142.02, 140.75, 127.62, 127.06, 125.11, 120.13, 108.14, 108.06, 81.96, 80.93, 65.47, 65.36, 48.72, 47.75, 47.63, 46.91, 46.73, 27.68, 27.62, 11.91. A precise NMR characterization was not possible due to the existence of multiple rotomers and peak overlapping. **HRMS** (ESI, m/z) calculated for $C_{30}H_{35}N_4O_7$ $[M + H]^+$: 563.2500, found: 563.2506; calculated for $C_{30}H_{34}N_4O_7Na$ $[M + Na]^+$: 585.2320, found: 585.2326.

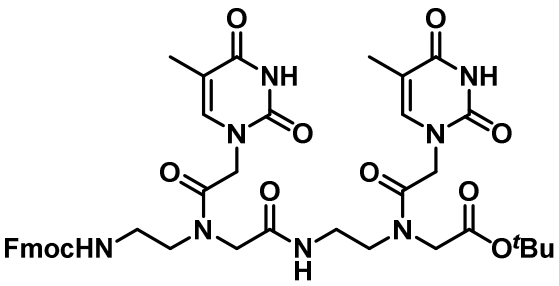
***N*-(2-(Fluorenylmethoxycarbonyl)aminoethyl)-*N*-(thymin-1-yl)acetyl)glycine (9)**[114]



8 (1.07 g, 1.90 mmol, 1eq.) was added to a mixture of dichloromethane (24 mL) and trifluoroacetic acid (8 mL). After stirring at room temperature for 4 hours, dichloromethane was evaporated and the solution was cooled to 0 °C. Ice-cold diethyl ether was added dropwise until a precipitate was formed. The solids were collected by filtration, washed with cold diethyl ether and dried under vacuum to afford **9** as a white solid (915 mg, 1.81 mmol, 95% yield). **R_f**:

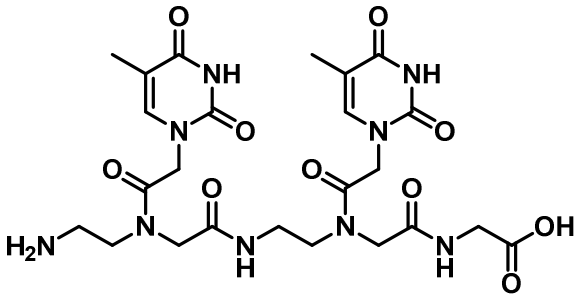
0.35 (EtOAc/MeOH 9:1). **m.p.**: 155.7 – 158.4 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 11.29 (s, 1H), 7.88 (d, J = 7.5 Hz, 2H), 7.68 (dd, J = 7.2, 2.8 Hz, 2H), 7.45 – 7.24 (m, 6H), 4.65 (s, 1H), 4.47 (s, 1H), 4.36 – 4.19 (m, 4H), 3.98 (s, 1H), 3.46 – 3.05 (m, 4H), 1.73 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 170.81, 170.45, 167.64, 167.25, 164.42, 156.35, 156.15, 151.01, 143.88, 142.13, 142.06, 140.76, 127.64, 127.09, 125.15, 120.15, 108.17, 108.08, 65.51, 65.41, 47.76, 47.69, 47.63, 46.81, 46.75, 11.90. A precise NMR characterization was not possible due to the existence of rotomers and peak overlapping. **HRMS** (ESI, m/z) calculated for $C_{26}H_{27}N_4O_7$ $[M + H]^+$: 507.1874, found: 507.1870; calculated for $C_{26}H_{26}N_4O_7Na$ $[M + Na]^+$: 529.1694, found: 529.1689.

Thymine PNA dimer (11)



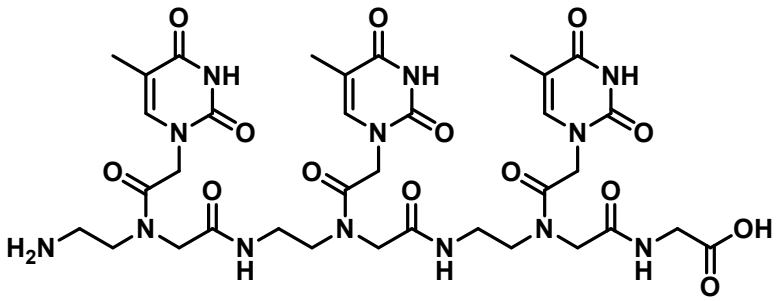
8 (56 mg, 0.10 mmol, 1 eq.) was dissolved in a Et_2NH/DCM mixture (1 mL/1 mL), and was stirred at room temperature for 4 hours. Then, the solvent was evaporated under reduced pressure, and the residue was redissolved in anhydrous DMF (1 mL), and **9** (101 mg, 0.20 mmol, 2 eq.), HBTU (75 mg, 0.20 mmol, 2 eq.) and DIPEA (51 mg, 69 µL, 0.40 mmol, 4 eq.) were added under an argon atmosphere. After stirring at room temperature for 16 hours, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (EtOAc/MeOH/H_2O 50:3:0.5) to afford **11** as a pale-yellow solid (43 mg, 0.05 mmol, 52% yield). **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 11.33 – 11.21 (m, 2H), 8.32 – 7.92 (m, 1H), 7.88 (d, J = 7.5 Hz, 2H), 7.67 (d, J = 7.3 Hz, 2H), 7.49 – 7.13 (m, 7H), 4.69 – 4.51 (m, 2H), 4.48 – 4.13 (m, 7H), 4.06 – 3.84 (m, 2H), 3.39 (d, J = 8.8 Hz, 2H), 3.26 – 3.09 (m, 2H), 1.77 – 1.65 (m, 6H), 1.47 – 1.34 (m, 9H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 168.18, 168.12, 164.43, 164.37, 164.26, 156.31, 156.08, 151.01, 143.85, 142.20, 142.07, 140.72, 127.61, 127.07, 125.15, 125.12, 120.12, 108.07, 108.01, 80.93, 80.78, 65.53, 65.47, 65.46, 46.71, 46.67, 27.69, 27.66, 27.63, 27.61, 11.89, 11.87, 11.84, 11.80, 11.77. **HRMS** (ESI, m/z) calculated for $C_{41}H_{49}N_8O_{11}$ $[M + H]^+$: 829.3525, found: 829.3524; calculated for $C_{41}H_{48}N_8O_{11}Na$ $[M + Na]^+$: 851.3335, found: 851.3342.

Thymine PNA dimer glycine (12)



A preloaded Fmoc-Gly-wang resin was used, with a loading number of 0.71 mml/g. The general method for SPPS was followed, using Fmoc-Gly-wang resin (104 mg, 0.074 mmol, 1 eq.), **9** (226 mg, 0.45 mmol, 6 eq.), HBTU (853 µL of a 0.5 M solution, 0.43 mmol, 5.72 eq.), DIPEA (116 mg, 156 µL, 0.90 mmol, 12 eq.). **12** was obtained as a white solid (32 mg, 0.053 mmol, 71% yield). **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 12.67 (br s, 1H), 11.36 – 11.24 (m, 2H), 8.62 – 8.15 (m, 2H), 7.89 (d, J = 23.9 Hz, 1H), 7.65 (br s, 1H), 7.42 – 7.21 (m, 2H), 4.72 – 4.64 (m, 1H), 4.54 – 4.40 (m, 2H), 4.21 – 3.76 (m, 7H), 3.64 – 3.19 (m, 15H), 3.01 (ddt, J = 91.4, 11.8, 5.5 Hz, 2H), 1.78 – 1.63 (m, 6H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 171.11, 169.60, 169.29, 169.11, 168.81, 168.77, 168.73, 168.71, 168.67, 168.56, 168.52, 168.12, 167.89, 167.58, 167.17, 167.15, 164.38, 164.35, 164.30, 157.91, 157.70, 151.25, 151.14, 151.12, 151.10, 151.08, 151.05, 151.02, 150.99, 142.27, 142.21, 142.17, 142.04, 142.03, 141.99, 141.94, 141.87, 118.08, 116.09, 108.37, 108.28, 108.22, 108.12, 107.97, 49.81, 49.67, 49.42, 49.36, 49.20, 49.17, 48.07, 47.98, 47.88, 47.77, 47.62, 46.89, 46.56, 46.07, 45.87, 45.41, 45.20, 45.05, 40.83, 40.62, 40.54, 37.00, 36.95, 36.91, 36.65, 36.43, 36.21, 35.99, 11.92, 11.89, 11.88, 11.86, 11.78, 11.69. A precise NMR characterization was not possible due to the existence of multiple rotomers and peak overlapping. **HPLC**: [A, 95→40], t_R= 1.750 min, (99%). **HRMS** (ESI, m/z) calculated for $C_{24}H_{34}N_9O_{10}$ $[M + H]^+$: 608.2423, found: 608.2423.

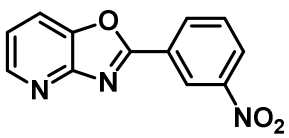
Thymine PNA trimer glycine (13)



A preloaded Fmoc-Gly-wang resin was used, with a loading number of 0.71 mml/g. The general method for SPPS was followed, using Fmoc-Gly-wang resin (105 mg, 0.074 mmol, 1 eq.), **9** (224 mg, 0.44 mmol, 6 eq.), HBTU (841 µL of a 0.5 M solution, 0.42 mmol, 5.72 eq.), DIPEA (114 mg, 154 µL, 0.88 mmol , 12 eq.). **13** was obtained as a white solid (41 mg, 0.046 mmol, 63% yield). **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 12.67 (br s, 1H), 11.36 – 11.18 (m, 3H), 8.68 – 8.00 (m, 3H), 7.96 – 7.86 (m, 1H), 7.66 (br s, 1H),

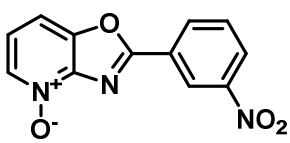
7.46 – 7.02 (m, 3H), 4.74 – 4.59 (m, 2H), 4.53 – 4.36 (m, 3H), 4.24 – 4.10 (m, 2H), 4.00 (t, *J* = 18.7 Hz, 3H), 3.91 – 3.75 (m, 4H), 3.69 – 3.17 (m, 16H), 3.12 – 2.84 (m, 4H), 1.79 – 1.48 (m, 9H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 171.11, 171.09, 170.18, 169.67, 169.56, 169.38, 168.98, 168.95, 168.88, 168.85, 168.79, 168.70, 168.67, 168.64, 168.61, 168.58, 168.54, 168.52, 168.46, 168.23, 168.19, 168.13, 168.05, 167.97, 167.91, 167.83, 167.79, 167.74, 167.69, 167.64, 167.62, 167.59, 167.31, 167.24, 167.17, 167.12, 164.46, 164.39, 164.35, 164.32, 158.02, 157.81, 157.60, 151.29, 151.21, 151.15, 151.06, 150.99, 150.94, 142.28, 142.25, 142.16, 142.13, 142.05, 141.90, 118.01, 116.03, 108.70, 108.39, 108.31, 108.29, 108.20, 108.16, 108.14, 107.87, 107.83, 107.78, 49.51, 49.37, 49.22, 49.05, 48.20, 48.10, 48.03, 48.00, 47.82, 47.76, 47.54, 47.26, 47.06, 46.68, 46.60, 46.24, 45.29, 45.20, 45.02, 40.84, 40.74, 40.55, 37.66, 37.55, 36.99, 36.91, 36.42, 36.29, 35.98, 35.86, 35.62, 11.93, 11.90, 11.87, 11.82, 11.76, 11.73, 11.69, 11.56, 11.52, 11.51, 11.38. A precise NMR characterization was not possible due to the existence of multiple rotomers and peak overlapping. **HPLC**: [A, 95→40], t_R= 1.700 min, (98%). **HRMS** (ESI, m/z) calculated for $C_{35}H_{48}N_{13}O_{14}$ $[M + H]^+$: 874.3438, found: 874.3438.

2-(3-Nitrophenyl)oxazolo[4,5-*b*]pyridine (18)[120]



2-Aminopyridin-3-ol (1.10 g, 10.00 mmol, 1 eq.) was added to a mixture of 3-nitrobenzoic acid (1.67 g, 10.00 mmol, 1 eq.) in polyphosphoric acid (50 g). After stirring at 150 °C for 16 hours, ice was added to the mixture, followed by NH_4OH (65 mL). The formed precipitate was fitered, dissolved in dichloromethane (100 mL) washed with brine and the aqueous phase was extracted several times with dichloromethane until the organic phase was colourless. The combined organic extracts were dried over Na_2SO_4, filtered and evaporated under reduced pressure. The obtained solid was recrystallized (EtOH) to afford **18** as a yellow solid (1.61 g, 6.67 mmol, 67 % yield). **R_f**: 0.66 (EtOAc). **m.p.**: 199.2 – 200.3 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.94 – 8.83 (m, 1H), 8.67 – 8.57 (m, 2H), 8.53 – 8.47 (m, 1H), 8.32 (dd, *J* = 8.2, 1.4 Hz, 1H), 7.99 – 7.90 (m, 1H), 7.53 (ddd, *J* = 8.0, 4.9, 1.4 Hz, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 162.80, 155.06, 148.28, 147.08, 143.04, 133.58, 131.36, 127.48, 126.95, 122.00, 121.48, 119.59. **HRMS** (ESI, m/z) calculated for $C_{12}H_8N_3O_3$ $[M + H]^+$: 242.0560, found: 242.0570.

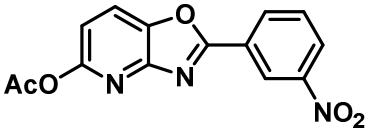
2-(3-Nitrophenyl)oxazolo[4,5-*b*]pyridine 4-oxide (17)



*m*CPBA (70 wt.%, 4.62 g, 18.74 mmol, 2 eq.) was added portionwise to a solution of **18** (2.26 g, 9.37 mmol, 1 eq.) in anhydrous dichloromethane (90 mL) at 0 °C under an argon atmosphere. After stirring at 45 °C for 16 hours, the reaction mixture was allowed to cool

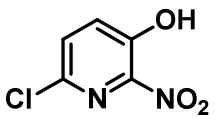
to room temperature. Dichloromethane (50 mL) and sat. aq. $NaHCO_3$ (100 mL) were added, the phases were separated and the aqueous phase was extracted several times with dichloromethane until the organic phase was colourless. The combined organic extracts were dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 98:2) to afford **17** as a pale-yellow solid (1.93 g, 7.50 mmol, 80% yield). **R_f**: 0.22 (DCM/MeOH 98:2). **m.p.**: 273.2 – 275.8 °C. **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 8.88 (t, J = 1.9 Hz, 1H), 8.63 (dt, J = 7.8, 1.3 Hz, 1H), 8.54 (ddd, J = 8.3, 2.3, 0.9 Hz, 1H), 8.40 (dd, J = 6.5, 0.6 Hz, 1H), 7.97 (t, J = 8.0 Hz, 1H), 7.92 – 7.90 (m, 1H), 7.50 (dd, J = 8.4, 6.6 Hz, 1H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 161.56, 148.34, 147.10, 146.31, 136.49, 133.68, 131.50, 127.25, 126.87, 122.42, 122.20, 109.50. **HRMS** (ESI, m/z) calculated for $C_{12}H_8N_3O_4$ $[M + H]^+$: 258.0509, found: 258.0518; calculated for $C_{12}H_7N_3O_3Na$ $[M + Na]^+$: 280.0329, found: 280.0334.

2-(3-Nitrophenyl)oxazolo[4,5-*b*]pyridin-5-yl acetate (19)



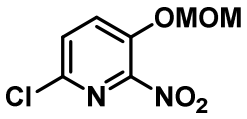
A solution of **17** (250 mg, 0.97 mmol) in acetic anhydride (16 mL) was stirred at 145 °C for 3 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was redissolved in dichloromethane (20 mL). The solution was washed with water (2 x 10 mL) and brine (10 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 99.7:0.3).The reaction resulted in a high degree of decomposition. Trace amounts of the desired product **19** were detected by LC-MS in the crude mixture, but none of the isolated fractions corresponded to it.

6-Chloro-2-nitropyridin-3-ol (22)[121]



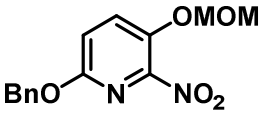
Potassium nitrate (7.02 g, 69.48 mmol, 1.8 eq.) was added in several portions to a solution of 6-chloropyridin-3-ol (5.00 g, 38.59 mmol, 1 eq.) in concentrated sulphuric acid (25 mL) at 0°C. The reaction was allowed to warm to room temperature and was further stirred for 24 hours. Then, ice was added to the reaction mixture and the formed precipitate was collected by filtration, washed with cold water and dried under high vacuum to afford **22** as a yellow solid (3.84 g, 22.00 mmol, 57% yield). **R_f**: 0.62 (Cy/EtOAc 1:1). **mp**: 101.5 – 103.4 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 11.90 (br s, 1H), 7.74 (d, J = 8.7 Hz, 1H), 7.70 (d, J = 8.7 Hz, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 146.79, 144.28, 136.38, 132.68, 130.87. **HRMS** (ESI, m/z) calculated for $C_5H_2ClN_2O_3$ $[M - H]^-$: 172.9748, found: 172.9758.

6-Chloro-3-(methoxymethoxy)-2-nitropyridine (23)



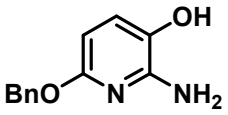
To a suspension of sodium hydride (60 wt.%, 756 mg, 18.91 mmol, 1.1 eq.) in anhydrous DMF (20 mL) at 0 °C under an argon atmosphere, a solution of **22** (3.00 g, 17.19 mmol, 1 eq.) in anhydrous DMF (15 mL) was added dropwise over 15 minutes. After stirring at 0 °C for 1 hour, MOM-Cl (1.52 g, 1.43 mL, 18.91 mmol, 1.1 eq.) was added. The reaction mixture was allowed to warm to room temperature, and was further stirred for 1 hour. The reaction was quenched by the addition of sat. aq. NH_4Cl (30 mL). Ethyl acetate was added (80 mL), the phases were separated and the aqueous phase was extracted with ethyl acetate (3 x 80 mL). The combined organic extracts were washed with brine (50 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was redissolved in toluene (2 x 100 mL) and concentrated under vacuum to remove the remaining DMF, and was purified by column chromatography (Cy/EtOAc 3:1) to afford **23** as a pale-yellow solid (3.53 g, 16.16 mmol, 94% yield). **R_f**: 0.27 (Cy/EtOAc 3:1). **mp**: 78.8 – 81.7 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.09 (d, J = 8.8 Hz, 1H), 7.91 (d, J = 8.8 Hz, 1H), 5.45 (s, 2H), 3.41 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 146.86, 144.17, 139.09, 131.14, 130.24, 95.43, 56.50. **HRMS** (ESI, m/z) calculated for $C_7H_7ClN_2O_4Na$ $[M + Na]^+$: 240.9987, found: 240.9987.

6-(Benzyloxy)-3-(methoxymethoxy)-2-nitropyridine (24)

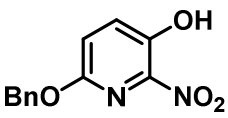


To a suspension of sodium hydride (60 wt.%, 198 mg, 4.94 mmol, 1.2 eq.) in anhydrous DMF (6 mL) at 0 °C under argon, benzyl alcohol (534 mg, 0.51 mL, 4.94 mmol, 2 eq.) was added dropwise over 15 minutes. The reaction mixture was allowed to warm to room temperature, and was further stirred for 1 hour. Then, a solution of **23** (900 mg, 4.12 mmol, 1 eq.) in anhydrous DMF (4 mL) was added. After stirring at the same temperature for 3 hours, the reaction mixture was quenched by the addition of sat. aq. NH_4Cl (20 mL). Dichloromethane was added (30 mL), the phases were separated and the aqueous phase was extracted with dichloromethane (3 x 30 mL). The combined organic extracts were washed with brine (20 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was redissolved in toluene (2 x 50 mL) and concentrated under vacuum to remove the remaining DMF, and was purified by column chromatography (Cy/EtOAC 10:1) to afford **24** as a colorless oil (884 mg, 3.05 mmol, 74% yield). **R_f**: 0.32 (Cy/EtOAc 10:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.51 – 7.45 (m, 3H), 7.42 – 7.30 (m, 3H), 7.03 (dd, J = 8.2, 0.9 Hz, 1H), 5.35 (s, 2H), 5.21 (s, 2H), 3.37 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 153.63, 140.01, 137.72, 136.49, 128.37, 128.22, 127.98, 126.68, 116.60, 94.90, 67.86, 55.90.

2-amino-6-(benzyloxy)pyridin-3-ol (21)

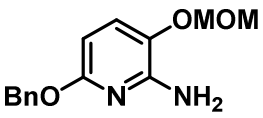


To a solution of **24** (560 mg, 1.93 mmol, 1 eq.) in ethanol (9 mL), $SnCl_2 \cdot 2H_2O$ (2.61 g, 11.58 mmol, 6 eq.) was added in one portion. After stirring at 100 °C for 90 minutes, the reaction mixture was allowed to cool to room temperature. Most of the ethanol was removed under vacuum, and water (15 mL) and ethyl acetate (15 mL) were added to the residue. Solid K_2CO_3 was carefully added to the resulting mixture until the pH was above 10. The organic phase was separated from the heterogeneous mixture and the aqueous phase was extracted with ethyl acetate (3 x 15 mL). The combined organic extracts were washed with brine (15 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 5:1). The desired product **21** was not obtained.



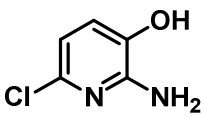
The reaction afforded exclusively 6-(benzyloxy)-2-nitropyridin-3-ol (**25**) as a pale-yellow oil (271 mg, 1.25 mmol, 65 % yield). $\mathbf{R_f}$: 0.33 (Cy/EtOAc 5:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.89 (s, 1H), 7.53 – 7.46 (m, 2H), 7.41 – 7.35 (m, 2H), 7.35 – 7.29 (m, 1H), 7.17 (dd, J = 8.0, 0.9 Hz, 1H), 6.90 (d, J = 8.0 Hz, 1H), 5.33 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 152.23, 140.83, 136.80, 134.40, 128.33, 128.14, 127.87, 125.01, 116.77, 67.53.

6-(benzyloxy)-3-(methoxymethoxy)pyridin-2-amine (26)[123]



Cs_2CO_3 (894 mg, 2.74, 2 eq.) was added to a solution of **24** (300 mg, 1.37 mmol, 1 eq.) and B_2pin_2 (1.08 g, 4.25 mmol, 3.1 eq.) in anhydrous ethanol under an argon atmosphere, and the mixture was stirred at 100 °C for 2 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was redissolved in ethyl acetate (20 mL). The solution was washed with water (2 x 10 mL) and brine (10 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 3:1). The reaction did not proceed, and only the starting material **24** was recovered.

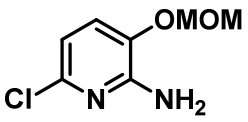
2-Amino-6-chloropyridin-3-ol (27)[124]



To a solution of **23** (300 mg, 1.37 mmol, 1 eq.) in ethanol (6 mL), $SnCl_2 \cdot 2H_2O$ (1.86 g, 8.23 mmol, 6 eq.) was added in one portion. After stirring at 100 °C for 90 minutes, the reaction mixture was allowed to cool to room temperature. Most of the ethanol was removed under vacuum, and water (10 mL) and ethyl acetate (10 mL) were added to the residue. Solid K_2CO_3 was carefully added to the resulting mixture until the pH was above 10. The organic phase was separated from the

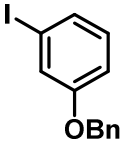
heterogeneous mixture and the aqueous phase was extracted with ethyl acetate (3 x 10 mL). The combined organic extracts were washed with brine (10 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM/EtOAc 9:1) to afford **27** as a pale-green solid (151 mg, 1.04 mmol, 76% yield). **R_f**: 0.20 (DCM/EtOAc 9:1). **m.p.**: 136.1 – 140.2 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.69 (s, 1H), 6.82 (d, J = 7.9 Hz, 1H), 6.37 (d, J = 7.8 Hz, 1H), 5.88 (br s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 150.66, 138.26, 136.64, 121.12, 110.24. **HRMS** (ESI, m/z) calculated for $C_5H_6ClN_2O$ $[M + H]^+$: 145.0163, found: 145.0165.

6-chloro-3-(methoxymethoxy)pyridin-2-amine (28)[123]



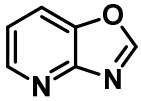
Cs_2CO_3 (894 mg, 2.74, 2 eq.) was added to a solution of **23** (300 mg, 1.37 mmol, 1 eq.) and B_2pin_2 (1.08 g, 4.25 mmol, 3.1 eq.) in anhydrous ethanol under an argon atmosphere, and the mixture was stirred at 100 °C for 2 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was redissolved in ethyl acetate (20 mL). The solution was washed with water (2 x 10 mL) and brine (10 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 3:1) to afford **28** as a brown oil, which solidified upon standing to form a pale-brown solid (173 mg, 0.92 mmol, 67% yield). **R_f**: 0.24 (Cy/EtOAc 3:1). **m.p.**: 58.7 – 60.5 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.12 (d, J = 8.1 Hz, 1H), 6.47 (d, J = 8.1 Hz, 1H), 6.21 (br s, 2H), 5.15 (s, 2H), 3.39 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 151.66, 139.56, 137.61, 122.61, 109.87, 94.59, 55.82. **HRMS** (ESI, m/z) calculated for $C_7H_{10}ClN_2O_2$ $[M + H]^+$: 189.0425, found: 189.0432.

1-(Benzyloxy)-3-iodobenzene (32)[125]



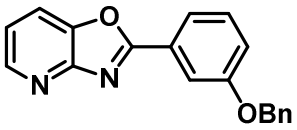
To a solution of 3-iodophenol (5.00 g, 22.73 mmol, 1 eq.) in acetonitrile (50 mL), benzyl bromide (3.89 g, 2.70 mL, 22.73 mmol, 1 eq.) and Cs_2CO_3 (14.81 g, 45.45 mmol, 2 eq.) were sequentially added, and the mixture was stirred at 85 °C for 2 hours. After cooling to room temperature, the reaction mixture was filtered through celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (Cy → Cy/EtOAc 15:1) to afford **32** as a white solid (7.04 g, 22.70 mmol, quantitative yield). **R_f**: 0.69 (Cy/EtOAc 15:1). **m.p.**: 51.2 – 52.8 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.51 – 7.25 (m, 7H), 7.12 – 7.00 (m, 2H), 5.10 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 159.04, 136.64, 131.31, 129.57, 128.44, 127.91, 127.72, 123.34, 114.68, 95.03, 69.37.

Oxazolo[4,5-*b*]pyridine (33)[126]



2-Aminopyridin-3-ol (3.00 g, 27.25 mmol, 1 eq.) was suspended in triethyl orthoformate (45 mL), and the reaction mixture was stirred at 140 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 1:2) to afford **33** as an off-white solid (2.93 g, 24.52 mmol, 90 % yield). **R_f**: 0.55 (EtOAc). **m.p.**: 70.5 – 75.2 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.03 (s, 1H), 8.57 (dd, J = 4.8, 1.4 Hz, 1H), 8.24 (dd, J = 8.2, 1.0 Hz, 1H), 7.48 (dd, J = 8.2, 4.8 Hz, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 157.19, 154.02, 146.57, 141.50, 121.01, 119.60. **HRMS** (ESI, m/z) calculated for $C_6H_5N_2O$ $[M + H]^+$: 121.0396, found: 121.0397; calculated for $C_6H_4N_2ONa$ $[M + Na]^+$: 143.0216, found: 143.0214.

2-(3-(Benzyloxy)phenyl)oxazolo[4,5-*b*]pyridine (31)



Method 1 (Pd-catalyzed cross coupling)[127]

Anhydrous acetone (15 mL) was added to an Ace pressure tube containing **33** (360 mg, 3.00 mmol, 1 eq.), **32** (1.40 g, 4.50 mmol, 1.5 eq.), $Pd(OAc)_2$ (40 mg, 0.18 mmol, 6 mol%), triphenylphosphine (189 mg, 0.72 mmol, 24 mol%.) and Cs_2CO_3 (1.95 g, 6.00 mmol, 2 eq.) under an argon atmosphere, and the mixture was stirred at 50 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 3:2) to afford **31** as an off-white solid (230 mg, 0.76 mmol, 25% yield).

Method 2 (Pd-catalyzed cross-coupling with Cu as co-catalyst)[128]

Anhydrous toluene (3 mL) was added to an Ace pressure tube containing **33** (120 mg, 1.00 mmol, 1 eq.), **32** (372 mg, 1.20 mmol, 1.2 eq.), $Pd(OAc)_2$ (2 mg, 0.01 mmol, 1 mol%), $Cu(OAc)_2 \cdot H_2O$ (40 mg, 0.20 mmol, 20 mol%), triphenylphosphine (131 mg, 0.50 mmol, 0.5 eq.) and K_2CO_3 (138 mg, 2.00 mmol, 2 eq.) under an argon atmosphere, and the mixture was stirred at 110 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 3:2) to afford **31** as a white solid (66 mg, 0.22 mmol, 22% yield).

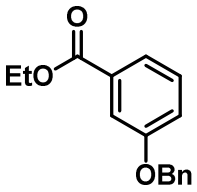
Method 3 (Cu-catalyzed cyclization)[130]

Anhydrous toluene (10 mL) was added to an Ace pressure tube containing **34** (2.00 g, 5.22 mmol, 1 eq.), CuI (99 mg, 0.52 mmol, 10 mol%), DMEDA (92 mg, 0.11 mL, 1.04 mmol, 0.2 eq.) and K_2CO_3 (1.44 g, 10.43 mmol, 2 eq.) under an argon atmosphere, and the

mixture was stirred at 100 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (DCM/EtOAc 15:1) to afford **31** as an off-white solid (726 mg, 2.40 mmol, 46% yield).

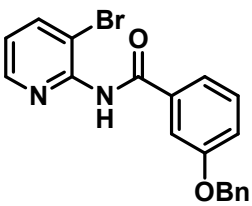
R_f: 0.29 (DCM/MeOH 98:2). **m.p.**: 148.3 – 151.8 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.56 (dd, J = 4.9, 1.4 Hz, 1H), 8.25 (dd, J = 8.2, 1.4 Hz, 1H), 7.87 – 7.83 (m, 2H), 7.60 – 7.55 (m, 1H), 7.53 – 7.46 (m, 3H), 7.45 – 7.40 (m, 2H), 7.38 – 7.32 (m, 2H), 5.26 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 164.59, 158.80, 155.43, 146.66, 142.80, 136.68, 130.76, 128.50, 127.96, 127.73, 127.18, 120.91, 120.26, 119.82, 119.16, 113.15, 69.51**HRMS** (ESI, m/z) calculated for $C_{19}H_{15}N_2O_2$ [M + H]$^+$: 303.1128, found: 303.1137; calculated for $C_{19}H_{14}N_2O_2Na$ [M + Na]$^+$: 325.0947, found: 325.0950.

Ethyl 3-(benzyloxy)benzoate (35)[132]



To a solution of ethyl 3-hydroxybenzoate (5.00 g, 30.09 mmol, 1 eq.) in acetonitrile (60 mL), benzyl bromide (5.15 g, 3.58 mL, 30.09 mmol, 1 eq.) and Cs_2CO_3 (21.41 g, 60.18 mmol, 2 eq.) were sequentially added, and the mixture was stirred at 85 °C for 2 hours. After cooling to room temperature, the reaction mixture was filtered through celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (10:1 Cy/EtOAc) to afford **35** as a colorless oil (7.63 g, 29.79 mmol, 99 % yield). **R_f**: 0.40 (Cy/EtOAc 10:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.59 – 7.26 (m, 9H), 5.16 (s, 2H), 4.30 (q, J = 7.1 Hz, 2H), 1.31 (t, J = 7.1 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 165.48, 158.37, 136.72, 131.28, 129.92, 128.44, 127.90, 127.69, 121.55, 119.92, 114.87, 69.41, 60.81, 14.11. **HRMS** (ESI, m/z) calculated for $C_{16}H_{17}O_3$ [M + H]$^+$: 257.1172, found: 257.1181; calculated for $C_{16}H_{16}O_3Na$ [M + Na]$^+$: 279.0992, found: 279.1000.

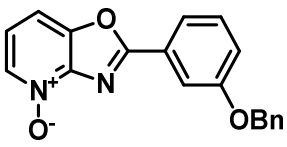
3-(Benzyloxy)-*N*-(3-bromopyridin-2-yl)benzamide (34)



To a solution of 2-amino-3-bromopyridine (2.00 g, 11.56 mmol, 1 eq.) in anhydrous toluene (60 mL), a solution of trimethyl aluminium (2.0 M in toluene, 5.78 mL, 11.56 mmol, 1 eq.) was added dropwise over 10 minutes at room temperature under an argon atmosphere, and it was further stirred for 1 hour. Then, **35** (2.96 g, 11.56 mmol, 1 eq.) in anhydrous toluene (10 mL) was added, and the reaction mixture was stirred at 120 °C for 5 hours. After cooling to room temperature, water was added until a precipitate formed, and the mixture was stirred for 10 minutes. The mixture was filtered over celite and washed several times with dichloromethane. The organic phase was separated, dried over Na_2SO_4,

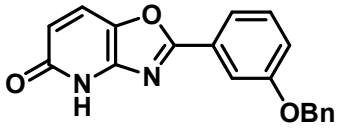
filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 1:1) to afford **34** as a pale-yellow oil, which solidified upon standing to form an off-white solid (3.27 g, 8.54 mmol, 74 % yield). **R_f**: 0.29 (Cy/EtOAc 1:1). **m.p.**: 97.5 – 98.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 10.71 (s, 1H), 8.51 (dd, J = 4.7, 1.6 Hz, 1H), 8.20 (dd, J = 8.0, 1.6 Hz, 1H), 7.66 – 7.58 (m, 2H), 7.50 – 7.38 (m, 5H), 7.37 – 7.30 (m, 2H), 7.29 – 7.24 (m, 1H), 5.19 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 165.11, 158.37, 149.69, 147.68, 142.07, 136.81, 134.88, 129.71, 128.48, 127.92, 127.75, 123.86, 120.36, 118.72, 118.68, 113.90, 69.45. **HRMS** (ESI, m/z) calculated for $C_{19}H_{16}BrN_2O_2$ $[M + H]^+$: 383.0390, found: 383.0392; calculated for $C_{19}H_{15}BrN_2O_2Na$ $[M + Na]^+$: 405.0209, found: 405.0208.

2-(3-(Benzyloxy)phenyl)oxazolo[4,5-*b*]pyridine 4-oxide (30)



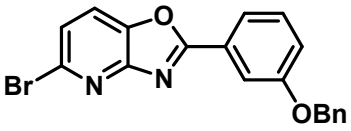
*m*CPBA (70 wt.%, 1.43 g, 5.83 mmol, 2 eq.) was added portionwise to a solution of **31** (882 mg, 2.91 mmol, 1 eq.) in anhydrous dichloromethane (25 mL) at 0 °C under an argon atmosphere. After stirring at 45 °C for 16 hours, the reaction mixture was allowed to cool to room temperature. Dichloromethane (25 mL) and sat. aq. $NaHCO_3$ (50 mL) were added, the phases were separated and the aqueous phase was extracted several times with dichloromethane until the organic phase was colourless. The combined organic extracts were dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (EtOAc/MeOH 9:1) to afford **30** as a pale-yellow solid (708 mg, 2.23 mmol, 76% yield). **R_f**: 0.28 (EtOAc/MeOH 9:1). **m.p.**: 240.1 – 241.7 (decomp.) °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.35 (dd, J = 6.6, 0.7 Hz, 1H), 7.86 – 7.79 (m, 3H), 7.59 (t, J = 8.0 Hz, 1H), 7.53 – 7.49 (m, 2H), 7.46 – 7.33 (m, 5H), 5.27 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.29, 158.82, 146.83, 146.48, 136.62, 136.23, 130.89, 128.50, 127.97, 127.69, 126.47, 121.83, 120.46, 120.28, 113.11, 109.21, 69.53. **HRMS** (ESI, m/z) calculated for $C_{19}H_{15}N_2O_3$ $[M + H]^+$: 319.1077, found: 319.1083.

2-(3-(Benzyloxy)phenyl)oxazolo[4,5-*b*]pyridin-5(4*H*)-one (29)[133]



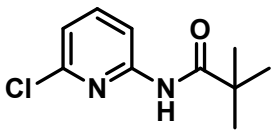
Anhydrous DCE (7 mL) was added to an Ace pressure tube containing **30** (580 mg, 1.82 mmol, 1 eq.), PyBroP (1.70 g, 3.64 mmol, 2 eq.), NaOAc (448 mg, 5.47 mmol, 3 eq.) and H_2O (492 mg, 0.49 mL, 27.33 mmol, 15 eq.) under an argon atmosphere, and the mixture was stirred at 85 °C for 16 hours. After cooling to room temperature, the mixture was filtered and the filter cake was washed with dichloromethane. The filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (DCM →

DCM/EtOAc 9:1 → EtOAc) to afford **29** as a pale-yellow solid (146mg, 0.46 mmol, 25% yield). R_f: 0.70 (EtOAc). m.p.: 325.8 – 332.3 °C (decomp.). ^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm): 11.31 (s, 1H), 8.06 (d, J = 8.9 Hz, 1H), 7.78 – 7.72 (m, 2H), 7.56 – 7.46 (m, 3H), 7.45 – 7.38 (m, 2H), 7.38 – 7.32 (m, 1H), 7.29 (ddd, J = 8.4, 2.5, 0.8 Hz, 1H), 6.63 (d, J = 8.9 Hz, 1H), 5.24 (s, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ (ppm): 163.60, 162.25, 158.78, 136.71, 130.69, 128.50, 127.96, 127.73, 127.54, 122.89, 119.76, 119.15, 112.70, 69.49. HRMS (ESI, m/z) calculated for $C_{19}H_{15}N_2O_3$ [M + H]$^+$: 319.1077, found: 319.1076; calculated for $C_{19}H_{14}N_2O_3Na$ [M + Na]$^+$: 341.0897, found: 341.3089.



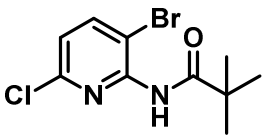
The reaction afforded the undesired 2-(3-(benzyloxy) phenyl)-5-bromooxazolo[4,5-*b*]pyridine (**36**) as the major product, as an off-white solid (267 mg, 0.70 mmol, 38% yield). **R_f**: 0.19 (DCM). **m.p.**: 211.5 – 213.1 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.25 (d, J = 8.4 Hz, 1H), 7.87 – 7.79 (m, 2H), 7.69 (d, J = 8.5 Hz, 1H), 7.59 (t, J = 8.0 Hz, 1H), 7.53 – 7.48 (m, 2H), 7.46 – 7.39 (m, 2H), 7.39 – 7.32 (m, 2H), 5.26 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 165.71, 158.81, 155.80, 142.68, 136.69, 136.63, 130.86, 128.51, 127.98, 127.73, 126.77, 124.55, 122.24, 120.46, 120.18, 113.31, 69.54. **HRMS** (ESI, m/z) calculated for $C_{19}H_{14}N_2O_2$ [M + H]$^+$: 381.0233, found: 381.0237; calculated for $C_{19}H_{13}N_2O_2Na$ [M + Na]$^+$: 403.0053, found: 403.0058.

***N*-(6-Chloropyridin-2-yl)pivalamide (40)**[137]

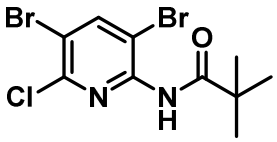


Pivaloyl chloride (4.13 g, 4.21 mL, 34.23 mmol, 1 eq.) was added dropwise over 10 minutes to a solution of 2-amino-6-chloropyridine (4.00 g, 31.11 mmol, 1 eq.) and triethylamine (3.94 g, 5.43 mL, 38.89 mmol, 1 eq.) in anhydrous dichloromethane (25 mL) at 0 °C under an argon atmosphere. The reaction mixture was allowed to warm to room temperature, and after 10 minutes a precipitate started forming. More dichloromethane (10 mL) was added, and the mixture was further stirred for 5 hours. Then, water was added, the phases were separated and the aqueous phase was extracted with dichloromethane (3 x 250 mL). The combined organic extracts were dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 5:1) to afford **40** as an off-white solid (6.35 g, 29.86 mmol, 96% yield). **R_f**: 0.47 (Cy/EtOAc 5:1). **m.p.**: 87.7 – 88.2 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 10.10 (s, 1H), 8.04 (d, J = 8.2 Hz, 1H), 7.81 (t, J = 8.0 Hz, 1H), 7.17 (d, J = 7.7 Hz, 1H), 1.22 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 177.43, 152.50, 147.82, 141.38, 118.94, 112.76, 39.49, 26.72. **HRMS** (ESI, m/z) calculated for $C_{10}H_{14}ClN_2O$ [M + H]$^+$: 213.0789, found 213.0791; calculated for $C_{10}H_{13}ClN_2ONa$ [M + Na]$^+$: 235.0609, found: 235.0610.

***N*-(3-Bromo-6-chloropyridin-2-yl)pivalamide (41)**[134]

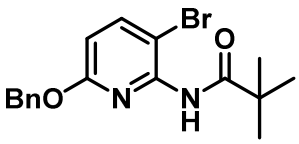


tert-Butyllithium (1.9 M in pentane, 16.31 mL, 31.03 mmol, 2.2 eq.) was added dropwise with a syringe pump over 1 hour to a solution of **40** (3.00 g, 14.11 mmol, 1 eq.) in anhydrous THF (45 mL) at -78 °C under an argon atmosphere. After stirring the reaction at -78 °C for 3 h, bromine (2.71 g, 0.87 mL, 16.93 mmol, 1.2 eq.) was added and the reaction mixture was allowed to warm to room temperature. After stirring for 16 hours, the reaction was quenched with 10% aq. $Na_2S_2O_3$ (500 mL) and the mixture was extracted with ethyl acetate (3 x 120 mL). The combined organic extracts were washed with sat. aq. $NaHCO_3$ (50 mL) and brine (50 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 5:1) to afford **41** as a white solid (1.60 g, 5.49 mmol, 39% yield). **R_f**: 0.25 (Cy/EtOAc 5:1). **m.p.**: 126.1 – 126.9 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.92 (s, 1H), 8.18 (d, J = 8.3 Hz, 1H), 7.39 (d, J = 8.3 Hz, 1H), 1.23 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 176.44, 150.11, 147.23, 144.87, 123.75, 117.62, 38.81, 26.99. **HRMS** (ESI, m/z) calculated for $C_{10}H_{13}BrClN_2O$ $[M + H]^+$: 290.9894, found: 290.9896; calculated for $C_{10}H_{12}BrClN_2ONa$ $[M + Na]^+$: 312.9714, found: 312.9718.



The reaction afforded the undesired side-product *N*-(3,5-dibromo-6-chloropyridin-2-yl)pivalamide **(42)** as a white solid (2.08 g, 5.61 mmol , 40% yield). **R_f**: 0.35 (Cy/EtOAc 5:1). **m.p.**: 121.5 – 122.1 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.98 (s, 1H), 8.67 (s, 1H), 1.22 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 176.52, 149.10, 146.92, 146.45, 117.79, 116.92, 38.86, 26.93. **HRMS** (ESI, m/z) calculated for $C_{10}H_{12}Br_2ClN_2O$ $[M + H]^+$: 370.8978, found: 370.8983; calculated for $C_{10}H_{11}Br_2ClN_2ONa$ $[M + Na]^+$: 392.8797, found: 392.8805.

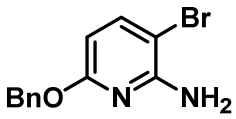
***N*-(6-(Benzyloxy)-3-bromopyridin-2-yl)pivalamide (43)**



To a suspension of sodium hydride (60 wt.%, 326 mg, 8.16 mmol, 2.8 eq.) in anhydrous DMF (5 mL) at 0 °C under an argon atmosphere, benzyl alcohol (851 mg, 0.81 mL, 7.87 mmol, 2.7 eq.) was added dropwise over 15 minutes. The reaction mixture was allowed to warm to room temperature, and was further stirred for 20 minutes. Then, a solution of **41** (850 mg, 2.91 mmol, 1 eq.) in anhydrous DMF (2 mL) was added. After stirring at 60 °C for 3 hours, the reaction mixture was cooled to room temperature and was quenched with sat. aq. NH_4Cl (20 mL). Ethyl acetate was added (20 mL), the phases were separated and the aqueous phase was extracted with ethyl acetate (5 x 20 mL). The combined organic extracts were dried over

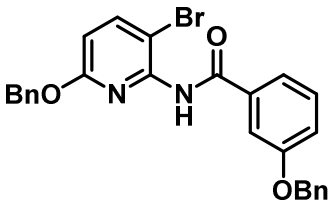
Na_2SO_4, filtered and evaporated under reduced pressure. The residue was redissolved in toluene (2 x 100 mL) and concentrated under vacuum to remove the remaining DMF, and was purified by column chromatography (Cy/EtOAc 7:1) to afford **43** as a white solid (999 mg, 2.75 mmol, 94% yield). **R_f**: 0.44 (Cy/EtOAc 5:1). **m.p.**: 84.0 – 84.7 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.60 (s, 1H), 7.97 (d, J = 8.6 Hz, 1H), 7.45 – 7.49 (m, 2H), 7.42 – 7.30 (m, 3H), 6.76 (d, J = 8.6 Hz, 1H), 5.30 (s, 2H), 1.25 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 176.23, 161.44, 147.37, 144.17, 136.63, 128.38, 128.28, 127.97, 110.38, 109.31, 67.63, 38.80, 27.17. **HRMS** (ESI, m/z) calculated for $C_{17}H_{20}BrN_2O_2$ $[M + H]^+$: 363.0703, found: 363.0714; calculated for $C_{17}H_{19}BrN_2O_2Na$ $[M + Na]^+$: 385.0522, found: 385.0530.

6-(Benzyloxy)-3-bromopyridin-2-amine (39)



43 (600 mg, 1.65 mmol, 1 eq.) was dissolved in MeOH (1.7 mL), and 5 M aq. NaOH (1.7 mL) was added to the solution. After stirring at 100 °C for 2 hours, the reaction mixture was cooled to room temperature. Water (10 mL) and ethyl acetate (20 mL) were added, the phases were separated and the aqueous phase was extracted with ethyl acetate (4 x 10 mL). The combined organic extracts were washed with brine (10 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure to afford **39** as a white solid (448 mg, 1.60 mmol, 97% yield). **R_f**: 0.53 (Cy/EtOAc 5:1). **m.p.**: 85.1 – 86.0 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.57 (d, J = 8.3 Hz, 1H), 7.45 – 7.40 (m, 2H), 7.40 – 7.28 (m, 3H), 6.16 (br s, 2H), 5.97 (d, J = 8.3 Hz, 1H), 5.24 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 161.55, 154.52, 142.69, 137.30, 128.33, 128.06, 127.73, 98.88, 92.47, 66.72. **HRMS** (ESI, m/z) calculated for $C_{12}H_{12}BrN_2O$ $[M + H]^+$: 279.0128, found: 279.0134; calculated for $C_{12}H_{11}BrN_2ONa$ $[M + Na]^+$: 300.9947, found: 300.9950.

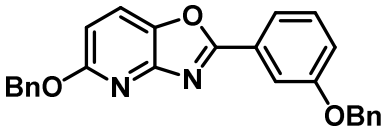
3-(Benzyloxy)-*N*-(6-(benzyloxy)-3-bromopyridin-2-yl)benzamide (38)



To a solution of **39** (279 mg, 1.00 mmol, 1 eq.) in anhydrous toluene (5 mL), a solution of trimethyl aluminium (2.0 M in toluene, 0.50 mL, 1.00 mmol, 1 eq.) was added dropwise over 10 minutes at room temperature under an argon atmosphere, and it was further stirred for 1 hour. Then, **35** (256 mg, 1.00 mmol, 1 eq.) in anhydrous toluene (1 mL) was added, and the reaction mixture was stirred at 120 °C for 5 hours. After cooling to room temperature, water was added until a precipitate formed, and the mixture was stirred for 10 minutes. The mixture was filtered over celite and washed several times with dichloromethane. The organic phase was separated, dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column

chromatography (Cy/EtOAc 7:1) to afford **38** as an off-white solid (382 mg, 0.78 mmol, 78 % yield). **R_f**: 0.31 (Cy/EtOAc 7:1). **m.p.**: 144.9 – 146.3 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ 10.57 (s, 1H), 8.05 (d, J = 8.6 Hz, 1H), 7.66 – 7.58 (m, 2H), 7.50 – 7.31 (m, 11H), 7.26 (dd, J = 8.3, 2.7, 1H), 6.83 (d, J = 8.6 Hz, 1H), 5.30 (s, 2H), 5.20 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 164.97, 161.50, 158.36, 147.02, 144.42, 136.82, 136.60, 134.95, 129.70, 128.48, 128.41, 128.18, 127.99, 127.92, 127.74, 120.38, 118.72, 113.92, 110.81, 109.00, 69.44, 67.63. **HRMS** (ESI, m/z) calculated for $C_{26}H_{22}BrN_2O_3$ [M + H]$^+$: 489.0808, found: 489.0801; calculated for $C_{12}H_9ClN_2O_3Na$ [M + Na]$^+$: 511.0628, found: 511.0619.

5-(Benzyloxy)-2-(3-(benzyloxy)phenyl)oxazolo[4,5-*b*]pyridine (37)



Method 1 (CuI as catalyst)[127]

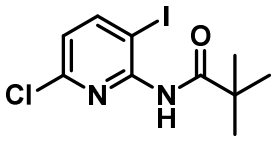
Anhydrous toluene (1.2 mL) was added to an Ace pressure tube containing **38** (224 mg, 0.46 mmol, 1 eq.), CuI (9 mg, 0.046 mmol, 10 mol%), DMEDA (8 mg, 10.0 µL, 0.091 mmol, 0.2 eq.) and K_2CO_3 (126 mg, 0.91 mmol, 2 eq.) under an argon atmosphere, and the mixture was stirred at 100 °C for 16 hours. TLC analysis revealed unreacted **X**, so more CuI (26 mg, 0.14 mmol, 30 mol%), DMEDA (24 mg, 29.6 µL, 0.27 mmol, 0.6 eq.) and K_2CO_3 (32 mg, 0.23 mmol, 0.5 eq.) were added, and the mixture was stirred at 100 °C for another 24 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/DCM 2:1) to afford **37** as a pale-yellow solid (80 mg, 0.20 mmol, 43% yield).

Method 2 ([(^{i}Pr)CuCl] as catalyst)

Anhydrous DMF (12 mL) was added to an Ace pressure tube containing **47** (1.00 g, 1.86 mmol, 1 eq.), [(^{i}Pr)CuCl] (91 mg, 0.19 mmol, 10 mol%) and K_2CO_3 (515 mg, 3.73 mmol, 2 eq.) under an argon atmosphere, and the mixture was stirred at 110 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (DCM) to afford **37** as a white solid (663 mg, 1.62 mmol, 87% yield).

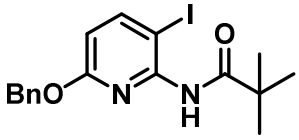
R_f: 0.50 (DCM). **m.p.**: 168.7 – 170.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.21 (d, J = 8.8 Hz, 1H), 7.83 – 7.75 (m, 2H), 7.59 – 7.47 (m, 5H), 7.44 – 7.29 (m, 7H), 6.95 (d, J = 8.8 Hz, 1H), 5.43 (s, 2H), 5.25 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 164.18, 161.48, 158.78, 152.54, 138.80, 137.05, 136.69, 130.72, 128.50, 128.39, 127.98, 127.95, 127.82, 127.71, 127.45, 122.62, 119.84, 119.41, 112.79, 108.57, 69.47, 67.55. **HRMS** (ESI, m/z) calculated for $C_{26}H_{21}N_2O_3$ [M + H]$^+$: 409.1547, found: 409.1551; calculated for $C_{26}H_{20}N_2O_3Na$ [M + Na]$^+$: 431.1366, found: 431.1367.

***N*-(6-Chloro-3-iodopyridin-2-yl)pivalamide (44)**[137]



tert-Butyllithium (1.7 M in pentane, 36.51 mL, 62.06 mmol, 2.2 eq.) was added dropwise with a syringe pump over 1 hour to a solution of **40** (6.00 g, 28.21 mmol, 1 eq.) in anhydrous THF (85 mL) at -78 °C under an argon atmosphere. After stirring the reaction at -78 °C for 3 h, a solution of iodine (8.59 g, 33.85 mmol, 1.2 eq.) in anhydrous THF (25 mL) was added, the reaction mixture was allowed to warm to room temperature and was further stirred for 16 hours. Then, the reaction was quenched with 10% aq. $Na_2S_2O_3$ (100 mL) and the mixture was extracted with ethyl acetate (3 x 250 mL). The combined organic extracts were washed with sat. aq. $NaHCO_3$ (100 mL) and brine (50 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 8:1 → 2:1) to afford **44** as an off-white solid (8.27 g, 24.43 mmol, 87% yield). $\mathbf{R_f}$: 0.26 (Cy/EtOAc 5:1). **m.p.**: 129.6 – 130.3 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 9.87 (s, 1H), 8.31 (d, J = 8.2 Hz, 1H), 7.21 (d, J = 8.2 Hz, 1H), 1.24 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 176.25, 153.17, 150.94, 148.30, 123.64, 93.67, 38.75, 27.03. **HRMS** (ESI, m/z) calculated for $C_{10}H_{13}ClIN_2O$ $[M + H]^+$: 338.9756, found: 338.9762; calculated for $C_{10}H_{12}ClIN_2ONa$ $[M + Na]^+$: 360.9575, found: 360.9593.

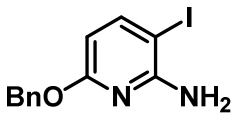
***N*-(6-(Benzyloxy)-3-iodopyridin-2-yl)pivalamide (45)**



To a suspension of sodium hydride (60 wt.%, 1.32 g, 33.08 mmol, 2.8 eq.) in anhydrous DMF (22 mL) at 0 °C under an argon atmosphere, benzyl alcohol (3.45 g, 3.30 mL, 31.90 mmol, 2.7 eq.) was added dropwise over 15 minutes. The reaction mixture was allowed to warm to room temperature, and was further stirred for 20 minutes. Then, a solution of **44** (4.00 g, 11.81 mmol, 1 eq.) in anhydrous DMF (11 mL) was added. After stirring at 60 °C for 3 hours, the reaction mixture was cooled to room temperature and was quenched with sat. aq. NH_4Cl (100 mL). Ethyl acetate was added (100 mL), the phases were separated and the aqueous phase was extracted with ethyl acetate (5 x 100 mL). The combined organic extracts were dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was redissolved in toluene (2 x 100 mL) and concentrated under vacuum to remove the remaining DMF, and was purified by column chromatography (Cy/EtOAc 10:1) to afford **45** as an off-white solid (4.02 g, 9.81 mmol, 83% yield). $\mathbf{R_f}$: 0.40 (Cy/EtOAc 5:1). **m.p.**: 92.0 – 92.4 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ 9.57 (s, 1H), 8.11 (d, J = 8.5 Hz, 1H), 7.49 – 7.44 (m, 2H), 7.41 – 7.30 (m, 3H), 6.62 (d, J = 8.5 Hz, 1H), 5.28 (s, 2H), 1.26 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 176.00, 162.39, 150.49, 149.98, 136.65, 128.35, 128.29, 127.94, 110.53, 83.78, 67.40, 38.73, 27.19. **HRMS** (ESI, m/z) calculated for

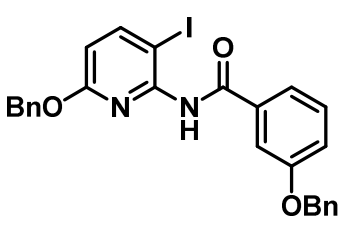
$C_{17}H_{20}IN_2O_2$ [M + H]$^+$: 411.0564, found: 411.0563; calculated for $C_{17}H_{19}IN_2O_2Na$ [M + Na]$^+$: 433.0383, found: 433.0382.

6-(Benzyloxy)-3-iodopyridin-2-amine (46)



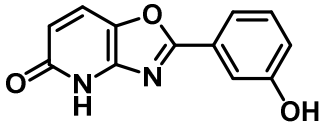
45 (3.81 g, 9.29 mmol, 1 eq.) was dissolved in MeOH (9.3 mL), and 5 M aq. NaOH (9.3 mL) was added to the solution. After stirring at 100 °C for 2 hours, the reaction mixture was cooled to room temperature. Water (50 mL) and ethyl acetate (100 mL) were added, the phases were separated and the aqueous phase was extracted with ethyl acetate (4 x 50 mL). The combined organic extracts were washed with brine (50 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure to afford **46** as a pale-yellow solid (3.00 g, 9.19 mmol, 99% yield). **R_f**: 0.51 (Cy/EtOAc 5:1). **m.p.**: 103.8 – 104.5 °C. **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ (ppm): 7.70 (d, *J* = 8.2 Hz, 1H), 7.45 – 7.41 (m, 2H), 7.39 – 7.28 (m, 3H), 6.02 (br s, 2H), 5.88 (d, *J* = 8.2 Hz, 1H), 5.24 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ (ppm): 162.53, 156.89, 148.92, 137.29, 128.30, 128.05, 127.70, 99.88, 66.51, 64.36. **HRMS** (ESI, m/z) calculated for $C_{12}H_{12}IN_2O$ [M + H]$^+$: 326.9989, found: 326.9990.

3-(Benzyloxy)-*N*-(6-(benzyloxy)-3-iodopyridin-2-yl)benzamide (47)



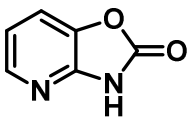
To a solution of **46** (2.00 g, 6.13 mmol, 1 eq.) in anhydrous toluene (30 mL), a solution of trimethyl aluminium (2.0 M in toluene, 3.06 mL, 6.13 mmol, 1 eq.) was added dropwise over 10 minutes at room temperature under an argon atmosphere, and it was further stirred for 1 hour. Then, **35** (1.57 g, 6.13 mmol, 1 eq.) in anhydrous toluene (5 mL) was added, and the reaction mixture was stirred at 120 °C for 5 hours. After cooling to room temperature, water was added until a precipitate formed, and the mixture was stirred for 10 minutes. The mixture was filtered over celite and washed several times with dichloromethane. The organic phase was separated, dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/DCM 1:3 → DCM) to afford **47** as an off-white solid (2.70 g, 5.03 mmol, 82 % yield). **R_f**: 0.09 (Cy/DCM 1:2). **m.p.**: 130.4 – 132.2 °C. **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ (ppm): 10.56 (s, 1H), 8.18 (d, *J* = 8.5 Hz, 1H), 7.68 – 7.65 (m, 1H), 7.61 (d, *J* = 7.8 Hz, 1H), 7.51 – 7.31 (m, 11H), 7.29 – 7.24 (m, 1H), 6.69 (d, *J* = 8.5 Hz, 1H), 5.30 (s, 2H), 5.20 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ (ppm): 164.95, 162.48, 158.36, 150.26, 150.15, 136.81, 136.63, 135.21, 129.68, 128.48, 128.40, 128.16, 127.96, 127.93, 127.75, 120.32, 118.56, 113.94, 111.03, 83.67, 69.46, 67.42. **HRMS** (ESI, m/z) calculated for $C_{26}H_{22}IN_2O_3$ [M + H]$^+$: 537.0670, found: 537.0678; calculated for $C_{26}H_{21}IN_2O_3Na$ [M + Na]$^+$: 559.0489, found: 559.0495.

2-(3-hydroxyphenyl)oxazolo[4,5-*b*]pyridin-5(4*H*)-one (15)



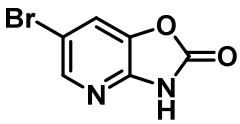
To a solution of **37** (150 mg, 0.36 mmol, 1 eq.) in TFH (5 mL), palladium on carbon (10 wt.%, 78 mg, 0.07 mmol, 20 mol%) was added and the reaction mixture was stirred under a hydrogen atmosphere (balloon) at room temperature for 24 hours. The reaction mixture was filtered through celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (Cy/THF 3:2) and washed with methanol to afford **15** as a pale-yellow solid (43 mg, 0.19 mmol, 52% yield). **R_f**: 0.17 (Cy/THF 2:1). **m.p.**: 295-5 – 305.7 °C (decomp.). **^{1}H NMR** (400 MHz, THF-d_8) δ (ppm): 10.10 (br s, 1H), 8.72 (s, 1H), 7.82 (d, *J* = 8.8 Hz, 1H), 7.66 (dt, *J* = 7.7, 1.1 Hz, 1H), 7.63 – 7.59 (m, 1H), 7.32 (t, *J* = 7.9 Hz, 1H), 6.95 (ddd, *J* = 8.1, 2.5, 0.9 Hz, 1H), 6.56 (d, *J* = 8.8 Hz, 1H). **^{13}C NMR** (101 MHz, THF-d_8) δ (ppm): 165.36, 163.37, 159.22, 138.75, 130.84, 129.26, 122.44, 119.78, 119.16, 114.80, 114.70, 108.56. Some unidentified triplet peaks were detected. **HRMS** (ESI, m/z) calculated for $C_{12}H_9N_2O_3$ $[M + H]^+$: 229.0608, found: 229.0608.

Oxazolo[4,5-*b*]pyridin-2(3*H*)-one (54)[139]



1,1'-carbonyldiimidazole (4.42 g, 27.25 mmol, 1.5 eq.) was added to a solution of 2-aminopyridin-3-ol (2.00 g, 18.16 mmol, 1 eq.) in THF (30 mL). After stirring at 70 °C for 16 hours, the reaction mixture was cooled to room temperature and concentrated under reduced pressure. The residue was dissolved in DCM (150 mL) and washed with 2 M aq. NaOH (3 x 50 mL). The combined aqueous layers were cooled to 0°C and acidified to a pH of 6 with 6 M aq. HCl. The formed precipitate was collected by filtration, washed with cold water (30 mL) and dried under high vacuum to afford **54** as a brown solid (1.88 g, 13.80 mmol, 76% yield). **R_f**: 0.70 (EtOAc). **m.p.**: 211.2 – 212.5 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 12.42 (br s, 1H), 8.02 (d, *J* = 5.3 Hz, 1H), 7.61 (d, *J* = 7.9 Hz, 1H), 7.09 (dd, *J* = 7.6, 5.6 Hz, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 153.48, 146.35, 142.48, 137.48, 117.86, 115.97. **HRMS** (ESI, m/z) calculated for $C_6H_5N_2O_2$ $[M + H]^+$: 137.0346, found: 137.0337.

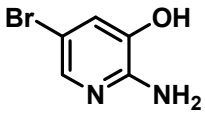
6-Bromooxazolo[4,5-*b*]pyridin-2(3*H*)-one (55)[139]



Bromine (1.29 g, 0.41 mL, 8.08 mmol, 1.1 eq.) was added dropwise over 20 minute to a solution of **54** (1.00 g, 7.35 mmol, 1 eq.) in anhydrous DMF (8 mL) under an argon atmosphere. After stirring at room temperature for 16 hours, more bromine was added (646 mg, 0.21 mL, 4.04 mmol, 0.55 eq.). The reaction mixture was further stirred for 24 hours, and then ice was added to it. The formed precipitate was collected by filtration, washed with cold water (30 mL) and

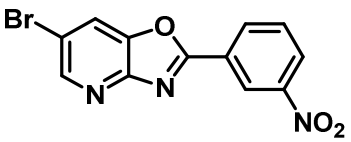
dried under high vacuum to afford **55** as a light-brown solid (1.41 g, 6.54 mmol, 89% yield). **R_f**: 0.60 (Cy/EtOAc 1:1). **m.p.**: 226.1 – 229.5 °C (decomp.). **^{1}H NMR** (300 MHz, DMSO-d_6) δ (ppm): 12.62 (br s, 1H), 8.17 (d, J = 2.0 Hz, 1H), 8.01 (d, J = 2.0 Hz, 1H). **^{13}C NMR** (75 MHz, DMSO-d_6) δ (ppm): 153.00, 145.45, 142.78, 137.87, 118.89, 111.85. **HRMS** (ESI, m/z) calculated for $C_6H_4BrN_2O_2$ $[M + H]^+$: 214.9451, found: 214.9448.

2-Amino-5-bromopyridin-3-ol (53)[139]



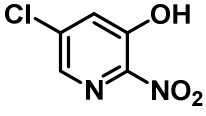
55 (2.00 g, 9.30 mmol, 1 eq.) was dissolved in 10% aq. NaOH (30 mL), and the resulting mixture was stirred at 100 °C for 8 hours. The reaction was cooled to 0°C, and 6 M aq. HCl was added until a precipitate formed (*ca.* pH 10). The solids were collected by filtration, washed with cold water (30 mL) and dried under high vacuum to afford **53** as a light-brown solid (1.56 g, 8.28 mmol, 89% yield). **R_f**: 0.22 (Cy/EtOAc 1:1). **m.p.**: 206.1 – 207.8 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 10.02 (s, 1H), 7.47 (d, J = 2.0 Hz, 1H), 6.93 (d, J = 2.0 Hz, 1H), 5.70 (br s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 149.80, 140.23, 137.10, 120.39, 104.38. **HRMS** (ESI, m/z) calculated for $C_5H_6BrN_2O$ $[M + H]^+$: 188.9658, found: 188.9652.

6-Bromo-2-(3-nitrophenyl)oxazolo[4,5-*b*]pyridine (52)



53 (500 mg, 2.64 mmol, 1 eq.) was added to a mixture of 3-nitrobenzoic acid (442 mg, 2.64 mmol, 1 eq.) in polyphosphoric acid (13 g). After stirring at 150 °C for 16 hours, ice was added to the mixture, followed by NH_4OH (20 mL). The formed precipitate was fitered, dissolved in dichloromethane (50 mL) washed with brine and the aqueous phase was extracted several times with dichloromethane until the organic phase was colourless. The combined organic extracts were dried over Na_2SO_4, filtered and evaporated under reduced pressure. LC-MS analysis of the obtained mixture revealed that some, together of a small amount of the desired product **52**, which could not be isolated.

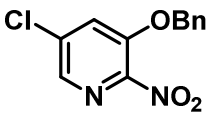
5-Chloro-2-nitropyridin-3-ol (60)[140]



Potassium nitrate (4.21 g, 41.68 mmol, 1.8 eq.) was added portionwise over 10 minutes to a solution of 3-chloropyridin-5-ol (3.00 g, 23.16 mmol, 1 eq.) in concentrated sulphuric acid (15 mL) at 0 °C. The reaction was allowed to warm to room temperature and was further stirred for 24 hours. Then, ice was added to the reaction mixture and the formed precipitate was collected by filtration, washed with cold water and dried under vacuum to afford **60** as an off-white solid (1.99 g,

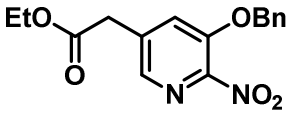
11.41 mmol, 49 % yield). **R_f**: 0.30 (Cy/EtOAc 6:1). **m.p.**: 82.1 – 83.1 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm): 10.30 (s, 1H), 8.13 (d, *J* = 2.1 Hz, 1H), 7.66 (d, *J* = 2.1 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm): 151.13, 140.72, 139.95, 139.56, 129.14. **HRMS** (ESI, m/z) calculated for $C_5H_2ClN_2O_3$ [M - H]$^-$: 172.9759, found: 172.9758.

3-(Benzyloxy)-5-chloro-2-nitropyridine (61)



To a solution of **60** (1.50 g, 8.59 mmol, 1 eq.) in acetonitrile (20 mL), benzyl bromide (1.47 g, 1.02 mL, 8.59 mmol, 1 eq.) and Cs_2CO_3 (5.60 g, 17.19 mmol, 2 eq.) were sequentially added, and the mixture was stirred at 85 °C for 2 hours. After cooling to room temperature, the reaction mixture filtered through celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 6:1) to afford **61** as an pale-yellow solid (2.14 g, 8.08 mmol, 94 % yield). **R_f**: 0.28 (Cy/EtOAc 6:1). **m.p.**: 108.2 – 109.8 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.31 (d, *J* = 1.9 Hz, 1H), 8.22 (d, *J* = 1.9 Hz, 1H), 7.47 – 7.34 (m, 5H), 5.40 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 146.80, 146.57, 137.74, 135.75, 134.94, 128.64, 128.47, 127.76, 125.69, 71.46. **HRMS** (ESI, m/z) calculated for $C_{12}H_9ClN_2O_3Na$ [M + Na]$^+$: 287.0194, found: 287.0198.

Ethyl 2-(5-(benzyloxy)-6-nitropyridin-3-yl)acetate (63)[141]

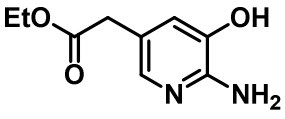


To a suspension of sodium hydride (60 wt.%, 605 mg, 15.11 mmol, 2 eq.) in anhydrous DMSO (15 mL) at room temperature under an argon atmosphere, *tert*-butyl ethyl malonate (2.84 g, 2.86 mL, 15.11 mmol, 2 eq.) was added dropwise at room temperature over 15 minutes. The mixture was stirred for 1 hour, and then a solution of **61** (2.00 g, 7.56 mmol, 1 eq.) in anhydrous DMSO (5 mL) was added. After stirring at 70 °C for 16 hours, the reaction mixture was cooled to room temperature and was quenched by the addition of sat. aq. NH_4Cl (20 mL). The mixture was extracted with diethyl ether (4 x 20 mL) and the organic extracts were washed several times with water to remove most of the DMSO. The combined aqueous phases were back-extracted with diethyl ether (2 x 20 mL), and the combined organic extracts were washed with brine (20 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure, which afforded a crude oil containing the intermediate **62**. **R_f**: 0.33 (Cy/EtOAc 4:1).

The crude mixture was dissolved in trifluoroacetic acid (22 mL), and was stirred at room temperature for 1 hour. Then, the reaction mixture was concentrated and the resulting oil was dissolved in ethyl acetate (50 mL) and washed with sat. aq. $NaHCO_3$ (2 x 20 mL). The combined aqueous phases were back-extracted with ethyl acetate (20 mL), and the combined organic extracts were washed with brine (20 mL), dried over Na_2SO_4, filtered and evaporated

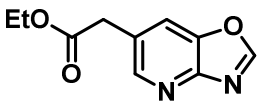
under reduced pressure. The residue was purified by column chromatography (Cy/EtOAc 3:1) to afford **63** as a yellow oil (1.08 g, 3.40 mmol, 45% yield over two steps). **R_f**: 0.24 (cHex/EtOAc 3:1). **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ (ppm): 8.06 – 8.03 (m, 2H), 7.46 – 7.33 (m, 5H), 5.33 (s, 2H), 4.12 (q, *J* = 7.1 Hz, 2H), 3.89 (s, 2H), 1.20 (t, *J* = 7.1 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ (ppm): 169.91, 147.27, 145.64, 140.15, 137.08, 135.27, 128.59, 128.38, 127.72, 126.49, 70.76, 60.81, 36.92, 14.02. **HRMS** (ESI, m/z) calculated for $C_{16}H_{17}N_2O_5$ $[M + H]^+$: 317.1132, found: 317.1135; calculated for $C_{16}H_{16}N_2O_5Na$ $[M + Na]^+$: 339.0951, found: 339.0978.

Ethyl 2-(6-amino-5-hydroxypyridin-3-yl)acetate (64)



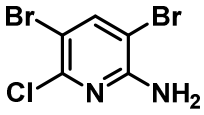
To a solution of **63** (884 mg, 2.79 mmol, 1 eq.) in EtOH (7 mL), palladium on carbon (10 wt.%, 297 mg, 0.28 mmol, 10 mol%) was added and the reaction mixture was stirred under a hydrogen atmosphere (balloon) at room temperature for 4 hours. The reaction mixture was filtered through celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (EtOAc/MeOH 98:2) to afford **64** as a yellow solid (399 g, 2.03 mmol, 73 % yield). **R_f** 0.21 (EtOAc/MeOH 98:2). **m.p.**: 145.1 – 146.8 °C (decomp.). **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ (ppm): 9.52 (br s, 1H), 7.29 (d, *J* = 1.9 Hz, 1H), 6.78 (d, *J* = 2.0 Hz, 1H), 5.37 (br s, 2H), 4.05 (q, *J* = 7.1 Hz, 2H), 3.40 (s, 2H), 1.17 (t, *J* = 7.1 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ (ppm): 171.57, 149.52, 138.99, 137.20, 119.63, 118.42, 60.17, 36.89, 14.12. **HRMS** (ESI, m/z) calculated for $C_9H_{13}N_2O_3$ $[M + H]^+$: 197.0921, found: 197.0919.

Ethyl 2-(oxazolo[4,5-*b*]pyridin-6-yl)acetate (59)



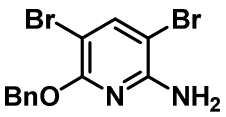
64 (374 mg, 1.91 mmol, 1 eq.) was suspended in triethyl orthoformate (3 mL), and the reaction mixture was stirred at 140 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 1:2) to afford **59** as a pale-yellow solid (338 mg, 1.64 mmol, 86% yield). **R_f**: 0.60 (EtOAc). **m.p.**: 99.1 – 101.5 °C. **^{1}H NMR** (400 MHz, DMSO-*d*$_6$) δ (ppm): 9.00 (s, 1H), 8.48 (d, *J* = 1.9 Hz, 1H), 8.22 – 8.14 (m, 1H), 4.10 (q, *J* = 7.1 Hz, 2H), 3.92 (s, 2H), 1.19 (t, *J* = 7.4 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-*d*$_6$) δ (ppm): 170.76, 157.28, 152.99, 147.73, 141.33, 128.13, 120.40, 60.61, 37.11, 14.03. **HRMS** (ESI, m/z) calculated for $C_{10}H_{11}N_2O_3$ $[M + H]^+$: 207.0764, found: 207.0761; calculated for $C_{10}H_{10}N_2O_3Na$ $[M + Na]^+$: 229.0584, found: 229.0579.

3,5-Dibromo-6-chloropyridin-2-amine (68)[142]



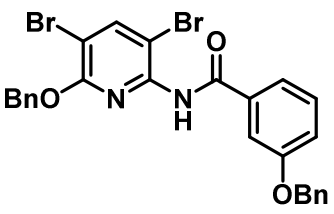
N-Bromosuccinimide (5.19 g, 29.17 mmol, 2.5 eq.) was added portionwise over 10 minutes to a solution of 6-chloropyridin-2-amine (1.50 g, 11.67 mmol, 1 eq.) in anhydrous DMF (15 mL) at 0 °C under an argon atmosphere. The mixture was allowed to warm to room temperature and was further stirred for 2 hours. Then, ice was added to the reaction mixture and the formed precipitate was collected by filtration, washed with cold water (100 mL) and cold dichloromethane (30 mL) and dried under high vacuum to afford **68** as a yellow solid (2.87 g, 10.03 mmol, 86 % yield). **R_f**: 0.37 (DCM). **m.p.**: 179.8 – 180.9 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.10 (s, 1H), 6.89 (br s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 155.37, 145.98, 144.43, 101.99, 101.81. **HRMS** (ESI, m/z) calculated for $C_5H_4Br_2ClN_2$ $[M + H]^+$: 286.8403, found: 286.8404.

6-(Benzyloxy)-3,5-dibromopyridin-2-amine (67)



Finely powdered KOH (1.14 g, 20.40 mmol, 3.3 eq.) was added to a solution of **68** (1.77 g, 6.18 mmol, 1 eq.), benzyl alcohol (836 mg, 0.80 mL, 7.73 mmol, 1.25 eq.) and 18-crown-6 (81.7 mg, 0.31 mmol, 5 mol%) in anhydrous toluene (15 mL) under an argon atmosphere. After stirring at 110 °C for 2 hours, the reaction mixture was cooled to room temperature, diluted with ethyl acetate (50 mL) and washed with water (2 x 25 mL). The combined aqueous phases were back-extracted with ethyl acetate (50 mL), and the combined organic extracts were washed with brine (25 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was purified by column chromatography (Cy/DCM 3:1) to afford **67** as a pale-yellow solid (1.83 g, 5.13 mmol, 83% yield). **R_f**: 0.17 (Cy/DCM 3:1). **m.p.**: 58.7 – 60.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.86 (s, 1H), 7.48 – 7.43 (m, 2H), 7.41 – 7.30 (m, 3H), 6.43 (br s, 2H), 5.32 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 156.90, 153.84, 144.02, 136.85, 128.37, 127.81, 92.71, 89.06, 67.47. **HRMS** (ESI, m/z) calculated for $C_{12}H_{11}Br_2N_2O$ $[M + H]^+$: 358.9213, found: 358.9218.

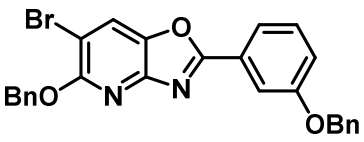
3-(Benzyloxy)-*N*-(6-(benzyloxy)-3,5-dibromopyridin-2-yl)benzamide (66)



To a solution of **67** (367 mg, 1.02 mmol, 1 eq.) in anhydrous toluene (5 mL), a solution of trimethyl aluminium (2.0 M in toluene, 0.51 mL, 1.02 mmol, 1 eq.) was added dropwise over 10 minutes at room temperature under an argon atmosphere, and it was further stirred for 1 hour. Then, **35** (263 mg, 1.02 mmol, 1 eq.) in anhydrous toluene (1 mL) was added, and the reaction mixture was stirred at 120 °C for 5 hours. After cooling to room temperature, water was added until a precipitate

formed, and the mixture was stirred for 10 minutes. The mixture was filtered over celite and washed several times with dichloromethane. The organic phase was separated, dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was purified by column chromatography (Cy/DCM 1:2) to afford **66** as an off-white solid (421 mg, 0.74 mmol, 72% yield). **R_f**: 0.19 (Cy/DCM 1:2). **m.p.**: 160.7 – 161.1 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 10.65 (s, 1H), 8.47 (s, 1H), 7.69 – 7.56 (m, 2H), 7.50 – 7.25 (m, 12H), 5.37 (s, 2H), 5.20 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 165.03, 158.37, 157.28, 146.29, 146.05, 136.79, 136.20, 134.74, 129.74, 128.47, 128.07, 127.93, 127.88, 127.73, 120.43, 118.84, 113.96, 108.92, 104.09, 69.45, 68.56. **HRMS** (ESI, m/z) calculated for $C_{26}H_{21}Br_2N_2O_3$ $[M + H]^+$: 568.9895, found: 568.9897; calculated for $C_{26}H_{20}Br_2N_2O_3Na$ $[M + Na]^+$: 590.9714, found: 590.9720.

5-(Benzyloxy)-2-(3-(benzyloxy)phenyl)-6-bromooxazolo[4,5-*b*]pyridine (65)



Method 1 (CuI as catalyst)

Anhydrous toluene (2.5 mL) was added to an Ace pressure tube containing **66** (360 mg, 0.63 mmol, 1 eq.), CuI (36 mg, 0.19 mmol, 30 mol%), DMEDA (34 mg, 41.1 µL, 0.38 mmol, 0.6 eq.) and K_2CO_3 (263 mg, 1.90 mmol, 3 eq.) under an argon atmosphere, and the mixture was stirred at 100 °C for 16 hours. TLC analysis revealed unreacted **66**, so more CuI (36 mg, 0.19 mmol, 30 mol%), DMEDA (34 mg, 41.1 µL, 0.38 mmol, 0.6 eq.) and K_2CO_3 (44 mg, 0.32 mmol, 0.5 eq.) were added, and the mixture was stirred at 100 °C for another 24 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/DCM 1:1). Only trace amounts of **65** were obtained, mixed with other impurities, as NMR analysis of the isolated fractions showed.

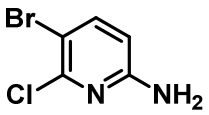
Method 2 ([(^{i}Pr)CuCl] as catalyst)

Anhydrous DMF (30 mL) was added to an Ace pressure tube containing **72** (2.90 g, 4.71 mmol, 1 eq.), [(^{i}Pr)CuCl] (230 mg, 0.47 mmol, 10 mol%) and K_2CO_3 (1.30 g, 9.42 mmol, 2 eq.) under an argon atmosphere, and the mixture was stirred at 110 °C for 16 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/DCM 1:1) to afford **65** as an off-white solid (1.34 mg, 2.76 mmol, 58% yield).

R_f: 0.21 (Cy/DCM 1:1). **m.p.**: 151.8 – 153.1 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.71 (s, 1H), 7.80 – 7.76 (m, 2H), 7.60 – 7.48 (m, 5H), 7.45 – 7.39 (m, 4H), 7.38 – 7.32 (m, 3H), 5.51 (s, 2H), 5.25 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 164.75, 158.79, 157.00, 151.67, 138.91, 136.65, 130.79, 128.51, 128.46, 127.97, 127.89,

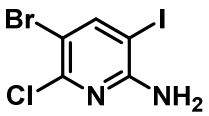
127.71, 127.58, 127.14, 125.52, 119.99, 119.68, 112.92, 102.66, 69.49, 68.57. **HRMS** (ESI, m/z) calculated for $C_{26}H_{20}BrN_2O_3$ [M + H]$^+$: 487.0652, found: 487.0652.

5-Bromo-6-chloropyridin-2-amine (69)[143]



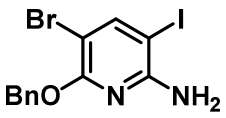
N-Bromosuccinimide (7.61 g, 42.78, 1.1 eq.) was added portionwise over 30 minutes to a solution of 2-amino-6-chloropyridine (5.00 g, 38.89 mmol, 1 eq.) in anhydrous DMF (85 mL) under an argon atmosphere protected from light. The reaction mixture was slowly warmed to room temperature over 2 hours, and it was further stirred at room temperature for 14 hours. Then, ice was added to the reaction mixture and the precipitate was filtered, washed with water and dried under high vacuum. The solid was purified by column chromatography (DCM) to afford **69** as an off-white solid (6.15 g, 29.64 mmol, 76% yield). **R_f**: 0.55 (DCM). **m.p.**: 155.5 – 156.1 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.63 (d, J = 8.6 Hz, 1H), 6.59 (br s, 2H), 6.35 (d, J = 8.6 Hz, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 158.78, 146.68, 142.74, 108.85, 101.76. **HRMS** (ESI, m/z) calculated for $C_5H_5BrClN_2$ [M + H]$^+$: 208.9297, found: 208.9295.

5-Bromo-6-chloro-3-iodopyridin-2-amine (70)[143]



To a solution of **69** (6.15 g, 29.64 mmol, 1 eq.) in a mixture of acetic acid (45 mL) and trifluoroacetic acid (45 mL), *N*-iodosuccinimide (8.00 g, 35.57 mmol, 1.2 eq.) was added in one portion. After stirring at room temperature for 3 hours, ice was added to the reaction mixture and the formed precipitate was collected by filtration, washed with water and dried under high vacuum to afford **70** as a pale-brown solid (9.72 g, 29.15 mmol, 98 % yield). **R_f**: 0.23 (DCM). **m.p.**: 181.5 – 181.9 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.16 (s, 1H), 6.72 (br s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 157.84, 150.55, 146.95, 102.42, 76.12. **HRMS** (ESI, m/z) calculated for $C_5H_4BrClIN_2$ [M + H]$^+$: 334.8263, found: 334.8266.

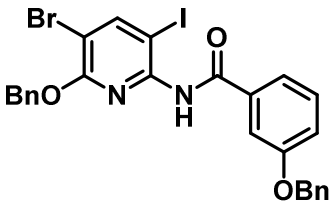
6-(Benzyloxy)-5-bromo-3-iodopyridin-2-amine (71)



Finely powdered KOH (3.71 g, 66.18 mmol, 3.3 eq.) was added to a solution of **70** (6.68 g, 20.05 mmol, 1 eq.), benzyl alcohol (2.39 g, 2.28 mL, 22.06 mmol, 1.1 eq.) and 18-crown-6 (265 mg, 1.00 mmol, 5 mol%) in anhydrous toluene (50 mL) under an argon atmosphere. After stirring at 110 °C for 2 hours, the reaction mixture was cooled to room temperature, diluted with ethyl acetate (100 mL) and washed with water (2 x 50 mL). The combined aqueous phases were back-extracted with ethyl acetate (100 mL), and the combined organic extracts were washed with brine (50 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The

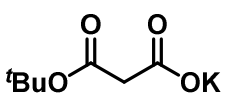
residue was purified by column chromatography (Cy/DCM 3:1) to afford **71** as a yellow oil, which solidified upon standing to form a pale-yellow solid (7.03 g, 17.37 mmol, 87% yield). **R_f**: 0.17 (Cy/DCM 3:1). **m.p.**: 53.5 – 54.2 °C. **^{1}H NMR** (400 MHz, $CDCl_3$) δ (ppm): 7.85 (s, 1H), 7.47 – 7.43 (m, 2H), 7.40 – 7.28 (m, 3H), 5.35 (s, 2H), 4.78 (br s, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ (ppm): 158.80, 155.02, 150.34, 136.95, 128.56, 127.91, 127.56, 93.61, 68.46, 64.38. **HRMS** (ESI, m/z) calculated for $C_{12}H_{11}BrIN_2O$ $[M + H]^+$: 404.9094, found: 404.9094.

3-(Benzyloxy)-*N*-(6-(benzyloxy)-5-bromo-3-iodopyridin-2-yl)benzamide (72)



To a solution of **71** (2.50 g, 6.17 mmol, 1 eq.) in anhydrous toluene (30 mL), a solution of trimethyl aluminium (2.0 M in toluene, 3.09 mL, 6.17 mmol, 1.1 eq.) was added dropwise over 10 minutes at room temperature under an argon atmosphere, and it was further stirred for 1 hour. Then, **35** (1.58 g, 6.17 mmol, 1.1 eq.) in anhydrous toluene (5 mL) was added, and the reaction mixture was stirred at 110 °C for 90 minutes. After cooling to room temperature, water was added until a precipitate formed, and the mixture was stirred for 10 minutes. The mixture was filtered over celite and washed several times with dichloromethane. The organic phase was separated, dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was purified by column chromatography (Cy/DCM 1:2) and the obtained solid was washed with warm cyclohexane (50 mL) to afford **72** as an off-white solid (2.74 g, 4.47 mmol, 72% yield). **R_f**: 0.19 (Cy/DCM 1:2). **m.p.**: 167.7 – 169.5 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 10.59 (s, 1H), 8.53 (s, 1H), 7.65 – 7.56 (m, 2H), 7.50 – 7.25 (m, 12H), 5.36 (s, 2H), 5.19 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 165.02, 158.38, 158.11, 151.46, 149.60, 136.80, 136.26, 135.00, 129.74, 128.49, 128.47, 128.06, 127.95, 127.88, 127.77, 120.35, 118.70, 113.95, 104.32, 84.03, 69.47, 68.35. **HRMS** (ESI, m/z) calculated for $C_{26}H_{21}BrIN_2O_3$ $[M + H]^+$: 614.9775, found: 614.9768; calculated for $C_{26}H_{20}BrIN_2O_3Na$ $[M + Na]^+$: 636.9594, found: 636.9590.

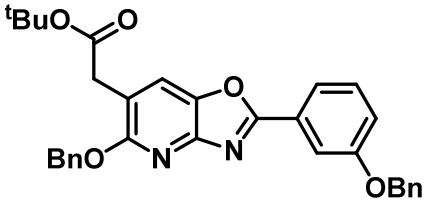
***tert*-Butyl potassium malonate (76)**[144a]



To a solution of mono-*tert*-butyl malonate (2.46 g, 15.35 mmol, 1 eq.) in EtOH (16 mL), a solution of ^{t}BuOK (1.72 g, 15.35 mmol, 1 eq.) in EtOH (21 mL) was added dropwise over 5 minutes under an argon atmosphere. After stirring at room temperature for 2 hours, the solvent was evaporated under reduced pressure. The residue was suspended in warm diethyl ether (10 mL), filtered and washed with more diethyl ether (5 mL) to afford **76** as an off-white solid (2.68 g, 13.53 mmol, 88% yield). **R_f**: 0.09 (EtOAc/MeOH 9:1). **m.p.**: 239.5 – 246.3 °C (decomp.). **^{1}H**

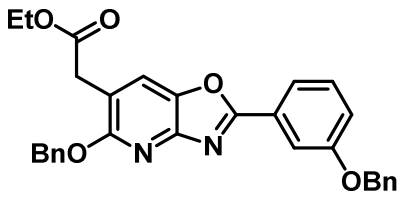
NMR (400 MHz, DMSO-d_6) δ (ppm): 2.76 (s, 2H), 1.36 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 169.45, 167.95, 78.06, 47.45, 27.93. **HRMS** (ESI, m/z) calculated for $C_7H_{11}K_2O_4$ [M + K]$^+$: 236.9926, found: 236.9924; calculated for $C_{14}H_{22}K_3O_8$ [2M + K]$^+$: 435.0220, found: 435.0227.

***tert*-Butyl 2-(5-(benzyloxy)-2-(3-(benzyloxy)phenyl)oxazolo[4,5-*b*]pyridin-6-yl)acetate (74)**[144b]



Mesitylene (3 mL) was added to an Ace pressure tube containing **65** (731 mg, 1.50 mmol, 1 eq.), **76** (446 mg, 2.25, 1.5 eq.), $Pd_2(allyl)_2Cl_2$ (11 mg, 0.03 mmol, 2 mol%), BINAP (56 mg, 0.09 mmol, 6 mol%) and DMAP (18 mg, 0.15 mmol, 0.1 eq.) under an argon atmosphere, and the mixture was stirred at 140 °C for 20 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (DCM) to afford **74** as a pale-yellow solid (352 mg, 0.67 mmol, 45% yield). **R_f**: 0.22 (DCM). **m.p.**: 120.3 – 122.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.17 (s, 1H), 7.81 – 7.77 (m, 2H), 7.57 – 7.46 (m, 5H), 7.44 – 7.29 (m, 7H), 5.44 (s, 2H), 5.25 (s, 2H), 3.71 (s, 2H), 1.33 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ 169.68, 163.75, 159.53, 158.77, 151.33, 138.59, 137.07, 136.70, 130.69, 128.50, 128.26, 127.96, 127.73, 127.64, 127.50, 127.31, 123.08, 119.79, 119.25, 116.31, 112.78, 80.31, 69.48, 67.60, 36.72, 27.62. **HRMS** (ESI, m/z) calculated for $C_{32}H_{31}N_2O_5$ [M + H]$^+$: 523.2227, found: 523.2228.

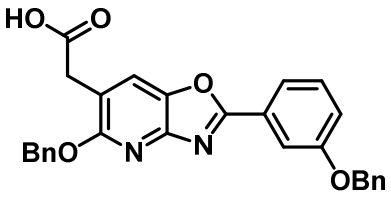
Ethyl 2-(5-(benzyloxy)-2-(3-(benzyloxy)phenyl)oxazolo[4,5-*b*]pyridin-6-yl)acetate (75)[144b]



Mesitylene (4 mL) was added to an Ace pressure tube containing **65** (975 mg, 2.00 mmol, 1 eq.), ethyl potassium malonate (511 mg, 3.00, 1.5 eq.), $Pd_2(allyl)_2Cl_2$ (15 mg, 0.04 mmol, 2 mol%), BINAP (75 mg, 0.12 mmol, 6 mol%) and DMAP (24 mg, 0.20 mmol, 0.1 eq.) under an argon atmosphere, and the mixture was stirred at 140 °C for 20 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (DCM) to afford **75** as an off-white solid (638 mg, 1.29 mmol, 65% yield). **R_f**: 0.19 (DCM). **m.p.**: 88.5 – 90.2 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.21 (s, 1H), 7.82 – 7.77 (m, 2H), 7.58 – 7.49 (m, 3H), 7.47 – 7.28 (m, 9H), 5.44 (s, 2H), 5.25 (s, 2H), 4.04 (q, J = 7.1 Hz, 2H), 3.80 (s, 2H), 1.11 (t, J = 7.1 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 170.50, 163.88, 159.46, 158.78, 151.45, 138.62,

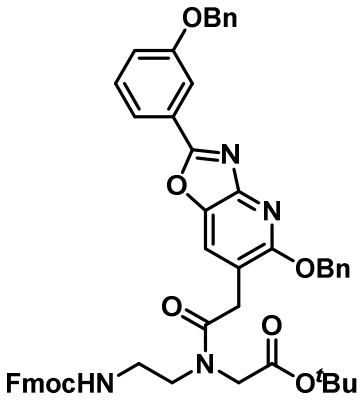
137.05, 136.71, 130.71, 128.51, 128.32, 127.96, 127.74, 127.67, 127.49, 127.30, 123.15, 119.83, 119.28, 115.92, 112.82, 69.49, 67.66, 60.38, 35.63, 14.01. **HRMS** (ESI, m/z) calculated for $C_{30}H_{27}N_2O_5$ $[M + H]^+$: 495.1914, found: 495.1921.

2-(5-(Benzyloxy)-2-(3-(benzyloxy)phenyl)oxazolo[4,5-*b*]pyridin-6-yl)acetic acid (73)



LiOH (90 mg, 3.78 mmol, 3 eq.) was added to a solution of **75** (623 mg, 1.25 mmol, 1 eq.) in THF/H_2O (16 mL/3.2 mL) at room temperature. After stirring the mixture at the same temperature for 16 hours, it was cooled in a freezer for 15 minutes and filtered. The solids were washed once with an ice-cold 5:1 mixture of THF/H_2O (10 mL) and dried under high vacuum to afford **73** as a white solid (490 mg, 1.05 mmol, 84% yield). **R_f**: 0.26 (EtOAc/MeOH 9:1). **m.p.**: 298.3 – 301.9 °C (decomp.). **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 8.01 (s, 1H), 7.79 – 7.75 (m, 2H), 7.55 – 7.49 (m, 5H), 7.42 (dd, J = 8.4, 6.8 Hz, 2H), 7.39 – 7.33 (m, 3H), 7.32 – 7.26 (m, 2H), 5.42 (s, 2H), 5.25 (s, 2H), 3.31 (s, 2H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 170.84, 162.54, 159.44, 158.78, 149.39, 138.93, 137.76, 136.76, 130.62, 128.50, 128.25, 127.94, 127.86, 127.74, 127.32, 127.13, 122.58, 122.00, 119.53, 118.89, 112.49, 69.46, 67.04. **HRMS** (ESI, m/z) calculated for $C_{28}H_{23}N_2O_5$ $[M + H]^+$: 467.1601, found: 467.1598; calculated for $C_{28}H_{22}N_2O_5Na$ $[M + Na]^+$: 489.1421, found: 489.1417.

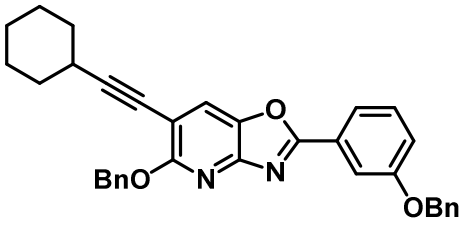
***tert*-butyl *N*-(2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)ethyl) - *N* - (2-(5-(benzyloxy)-2-(3-(benzyloxy)phenyl)oxazolo[4,5-*b*]pyridin-6-yl)acetyl)glycinate (77)[138]**



2 (228 mg, 0.53 mmol, 1eq.) was suspended in ethyl acetate (20 mL), the mixture was washed with sat. aq. $NaHCO_3$ (2 x 10 mL) until the solids dissolved and the combined aqueous phases were back-extracted with ethyl acetate (10 mL). The combined organic extracts were washed with brine (10 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The obtained oil was dissolved in anhydrous DMF (5 mL), and **73** (295 mg, 0.63, 1.2 eq.), TBTU (203 mg, 0.63 mmol, 1.2 eq.), HOBt·H_2O (98 mg, 0.63 mmol, 1.2 eq.) and DIPEA (82 mg, 0.11 mL, 0.63 mmol, 1.2 eq.) were added to the solution under an argon atmosphere. After stirring the reaction at room temperature for 16 hours, water (15 mL) was added, and the mixture was extracted with dichloromethane (3 x 20 mL). The combined organic extracts were washed with brine (10 mL), dried over Na_2SO_4, filtered, evaporated

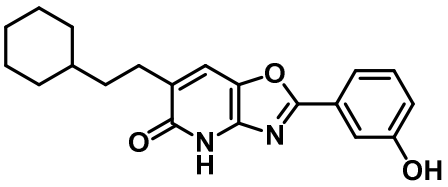
under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 99.5:0.5) to afford **77** as white solid (259 mg, 0.31 mmol, 58% yield). **R_f**: 0.47 (DCM/MeOH 98:2). **m.p.**: 77.9 – 81.3 °C. **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 8.02 (s, 1H), 7.85 (dd, J = 16.2, 7.5 Hz, 2H), 7.78 – 7.72 (m, 2H), 7.64 (t, J = 6.5 Hz, 2H), 7.55 – 7.23 (m, 18H), 5.42 (s, 2H), 5.23 (s, 2H), 4.34 – 4.13 (m, 4H), 3.98 – 3.68 (m, 3H), 3.48 – 3.36 (m, 2H), 3.16 (dq, J = 19.3, 6.4 Hz, 2H), 1.39 (s, 9H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 170.36, 170.02, 168.82, 168.47, 163.59, 159.25, 159.20, 158.76, 158.75, 156.27, 156.11, 151.05, 143.82, 140.75, 140.69, 138.67, 137.18, 137.14, 136.70, 130.65, 130.64, 128.49, 128.28, 128.22, 127.94, 127.72, 127.60, 127.56, 127.54, 127.52, 127.48, 127.41, 127.29, 127.01, 126.96, 125.08, 125.03, 122.66, 122.62, 120.12, 120.08, 119.74, 119.70, 119.18, 117.27, 116.94, 112.73, 81.62, 80.78, 69.47, 67.62, 67.53, 65.39, 65.36, 54.91, 50.97, 48.64, 47.98, 46.88, 46.70, 38.84, 38.21, 34.06, 33.77, 27.71, 27.58. A precise NMR characterization was not possible due to the existence of rotomers and peak overlapping. **HRMS** (ESI, m/z) calculated for $C_{41}H_{49}N_8O_{11}$ $[M + H]^+$: 829.3525, found: 829.3524; calculated for $C_{41}H_{48}N_8O_{11}Na$ $[M + Na]^+$: 851.3335, found: 851.3342.

5-(benzyloxy)-2-(3-(benzyloxy)phenyl)-6-(cyclohexylethynyl)oxazolo[4,5-*b*] pyridine (78)



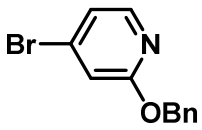
Triethylamine (2 mL) was added to an Ace pressure tube containing **65** (146 mg, 0.30 mmol, 1 eq.), $Pd(PPh_3)_2Cl_2$ (10 mg, 0.02 mmol, 5 mol%) and CuI (9 mg, 0.05 mmol, 15 mol%) under an argon atmosphere. Then, cyclohexylacetilene (45 mg, 54.9 µL, 0.42 mmol, 1.4 eq.) and anhydrous DMF (2 mL) were added. After stirring at 60 °C for 16 hours, the reaction mixture was cooled to room temperature and filtered through celite. The filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (Cy/DCM 1:2) to afford **78** as a pale-yellow solid (152 mg, 0.29 mmol, 98% yield). **R_f**: 0.39 (Cy/DCM 1:2). **m.p.**: 118.5 – 119.8 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.27 (s, 1H), 7.81 – 7.74 (m, 2H), 7.59 – 7.48 (m, 5H), 7.46 – 7.31 (m, 7H), 5.47 (s, 2H), 5.25 (s, 2H), 2.77 – 2.72 (m, 1H), 1.83 – 1.64 (m, 4H), 1.56 – 1.40 (m, 3H), 1.37 – 1.28 (m, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 164.88, 161.48, 158.78, 151.63, 138.28, 136.95, 136.66, 130.76, 128.51, 128.29, 127.96, 127.71, 127.31, 127.22, 124.05, 119.94, 119.57, 112.86, 104.76, 100.51, 75.75, 69.49, 67.87, 31.90, 28.88, 25.43, 23.83. **HRMS** (ESI, m/z) calculated for $C_{34}H_{31}N_2O_3$ $[M + H]^+$: 515.2329, found: 515.2332; calculated for $C_{34}H_{30}N_2O_3Na$ $[M + Na]^+$: 537.2149, found: 537.2149.

6-(2-cyclohexylethyl)-2-(3-hydroxyphenyl)oxazolo[4,5-*b*]pyridin-5(4*H*)-one (79)



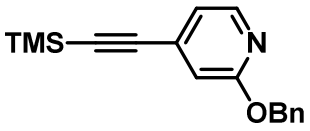
To a solution of **78** (120 mg, 0.23 mmol, 1 eq.) in THF (4 mL), palladium on carbon (10 wt.%, 50 mg, 0.05 mmol, 20 mol%) was added and the reaction mixture was stirred under a hydrogen atmosphere (balloon) at room temperature for 16 hours. TLC analysis revealed the reaction was incomplete, so more palladium on carbon (10 wt.%, 50 mg, 0.05 mmol, 20 mol%) was added, and the mixture was stirred at room temperature for another 24 hours. After that time, the reaction mixture was filtered through celite, the filtrate was concentrated under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 95:5) and washed with methanol to afford **79** as a pale-yellow solid (51 mg, 0.15 mmol, 65% yield). $\mathbf{R_f}$: 0.23 (DCM/MeOH 95:5). **m.p.**: 247.5 – 249.0 °C. **^{1}H NMR** (600 MHz, DMSO-d_6) δ (ppm): 11.40 (br s, 1H), 9.93 (s, 1H), 7.92 (s, 1H), 7.54 (d, *J* = 7.6 Hz, 1H), 7.50 (s, 1H), 7.39 (t, *J* = 7.9 Hz, 1H), 6.99 (dd, *J* = 8.1, 1.9 Hz, 1H), 2.56 (q, *J* = 6.5, 5.3 Hz, 2H), 1.76 (d, *J* = 12.3 Hz, 2H), 1.68 (dd, *J* = 9.3, 6.3 Hz, 2H), 1.61 (d, *J* = 11.7 Hz, 1H), 1.46 (q, *J* = 7.0 Hz, 2H), 1.26 – 1.11 (m, 4H), 0.92 (qd, *J* = 12.7, 2.7 Hz, 2H). **^{13}C NMR** (151 MHz, DMSO-d_6) δ (ppm): 157.89, 130.54, 127.47, 119.00, 117.61, 113.16, 36.88, 36.27, 32.77, 27.62, 26.20, 25.80. The structure was confirmed with additional 2D NMR spectra, such as HSQC and HMBC. The missing peaks in the 13C NMR spectrum were attributed to peak broadening. **HRMS** (ESI, m/z) calculated for $C_{20}H_{23}N_2O_3$ $[M + H]^+$: 339.1703, found: 339.1701; calculated for $C_{20}H_{22}N_2O_3Na$ $[M + Na]^+$: 361.1523, found: 361.1521.

2-(Benzyloxy)-4-bromopyridine (86)[149]



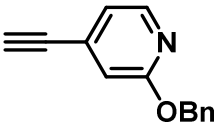
Potassium *tert*-butoxide (1.80 g, 16.06 mmol, 1.1 eq.) was added portionwise over 30 minutes to a solution of 4-bromo-2-fluoropyridine (2.57 g, 14.60 mmol, 1 eq.) and benzyl alcohol (1.74 g, 1.66 mL, 16.06 mmol, 1.1 eq.) in anhydrous THF (25 mL) at 0 °C under an argon atmosphere. The mixture was stirred 1 hour at 0 °C and further 30 minutes at room temperature. Then, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 25:1) to afford **86** as colorless oil (3.82 g, 14.45 mmol, 99% yield). $\mathbf{R_f}$: 0.35 (Cy/EtOAc 25:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.09 (d, *J* = 5.5 Hz, 1H), 7.46 – 7.29 (m, 5H), 7.25 (dt, *J* = 5.5, 1.5 Hz, 1H), 7.21 (d, *J* = 1.6 Hz, 1H), 5.36 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.74, 148.02, 136.80, 133.62, 128.38, 127.97, 127.87, 120.53, 113.70, 67.46. **HRMS** (ESI, m/z) calculated for $C_{12}H_{11}BrNO$ $[M + H]^+$: 264.0019, found: 264.0031.

2-(Benzyloxy)-4-((trimethylsilyl)ethynyl)pyridine (87)



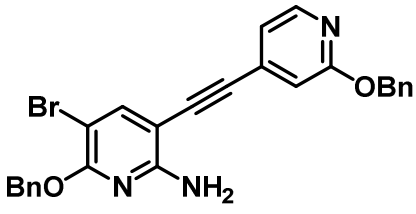
Trimethylsilylacetylene (1.60 g, 2.25 mL, 16.28 mmol, 1.5 eq.), $Pd(PPh_3)_4$ (627 mg, 0.54 mmol, 5 mol%) and CuI (103 mg, 0.54 mmol, 5 mol%) were added to a degassed solution of **86** (2.87 g, 10.87 mmol, 1 eq.) in a mixture of anhydrous THF/Et3N (63 mL/ 63 mL) under an argon atmosphere. After stirring at 70 °C for 16 hours, the reaction mixture was cooled to room temperature and filtered through celite. The filtrate was concentrated under reduced pressure and the residue was redissolved in ethyl acetate (100 mL). The solution was washed with water (2 x 50 mL) and brine (50 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 25:1) to afford **87** as a yellow oil (2.99 g, 10.64 mmol, 98% yield). **R_f**: 0.33 (Cy/EtOAc 25:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.17 (dd, J = 5.2, 0.8 Hz, 1H), 7.46 – 7.29 (m, 5H), 7.01 (dd, J = 5.3, 1.3 Hz, 1H), 6.91 (dd, J = 1.4, 0.8 Hz, 1H), 5.35 (s, 2H), 0.24 (s, 9H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.19, 147.43, 137.03, 132.73, 128.35, 127.81, 127.76, 119.15, 112.67, 102.05, 99.18, 67.09, -0.41. **HRMS** (ESI, m/z) calculated for $C_{17}H_{20}NOSi$ $[M + H]^+$: 282.1309, found: 282.1326.

2-(Benzyloxy)-4-ethynylpyridine (85)



To a solution of **87** (2.99 g, 10.64 mmol, 1 eq.) in MeOH (60 mL) K_2CO_3 (294 mg, 2.13 mmol, 0.2 eq.) was added, and the reaction mixture was stirred for 1 hour. Then, the solvent was removed under reduced pressure and the residue was redissolved in diethyl ether (100 mL). The solution was washed with water (2 x 50 mL) and brine (50 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure to afford **85** as a pale-yellow oil (2.18 g, 10.42 mmol, 98% yield). **R_f**: 0.32 (Cy/EtOAc 20:1). **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.19 (dd, J = 5.2, 0.7 Hz, 1H), 7.45 – 7.29 (m, 5H), 7.05 (dd, J = 5.2, 1.3 Hz, 1H), 6.96 (t, J = 1.1 Hz, 1H), 5.35 (s, 2H), 4.58 (s, 1H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.17, 147.41, 136.99, 132.48, 128.37, 127.93, 127.80, 119.35, 113.05, 85.21, 80.84, 67.14. **HRMS** (ESI, m/z) calculated for $C_{14}H_{12}NO$ $[M + H]^+$: 210.0913, found: 210.0935.

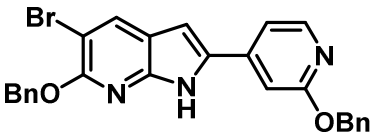
6-(Benzyloxy)-3-((2-(benzyloxy)pyridin-4-yl)ethynyl)-5-bromopyridin-2-amine (84)



A degassed solution of **85** (2.15 g, 10.28 mmol, 1.05 eq.) in anhydrous THF (20 mL) was added to a flask containing **71** (3.97 g, 9.80 mmol, 1 eq.), $Pd(PPh_3)_4$ (357 mg, 0.31 mmol, 3 mol%) and CuI (59 mg, 0.31 mmol, 3 mol%) under an argon atmosphere, and then

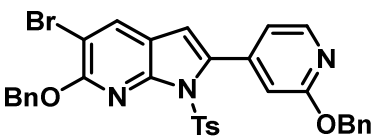
triethylamine (5 mL) was added. After stirring at room temperature for 16 hours, the reaction mixture was quenched with sat. aq. NH_4Cl. Dichloromethane (50 mL) was added, the phases were separated and the aqueous phase was extracted with dichloromethane (3 x 50 mL). The combined organic extracts were washed with water (50 mL) and brine (50 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM) to afford **84** as a pale-orange solid (3.75 g, 7.71 mmol, 79% yield). **R_f**: 0.19 (DCM). **m.p.**: 134.2 – 136.7 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.20 – 8.14 (m, 1H), 7.81 (s, 1H), 7.53 – 7.29 (m, 10H), 7.16 (d, J = 4.6 Hz, 2H), 6.82 (br s, 2H), 5.38 (s, 2H), 5.37 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.14, 158.38, 158.34, 146.99, 145.14, 137.20, 136.68, 133.58, 128.42, 128.36, 127.91, 127.87, 127.80, 127.74, 118.61, 112.01, 93.57, 91.69, 89.13, 89.10, 67.62, 66.99. **HRMS** (ESI, m/z) calculated for $C_{26}H_{21}BrN_3O_2$ $[M + H]^+$: 486.0812, found: 486.0815.

6-(Benzyloxy)-2-(2-(benzyloxy)pyridin-4-yl)-5-bromo-1*H*-pyrrolo[2,3-*b*]pyridine (83)



To a solution of **84** (3.75 g, 7.71 mmol, 1 eq.) in *N*-methyl-2-pyrrolidone (30 mL) potassium *tert*-butoxide (1.73 g, 15.42 mmol, 2 eq.) was added, and the reaction mixture was stirred at 80 °C for 2 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 99:1) to afford **83** as a pale-yellow solid (3.28 g, 6.75 mmol, 88% yield). **R_f**: 0.18 (DCM/MeOH 99:1). **m.p.**: 117.5 – 119.5 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 12.37 (s, 1H), 8.27 (s, 1H), 8.18 (d, J = 5.4 Hz, 1H), 7.54 – 7.44 (m, 5H), 7.44 – 7.30 (m, 7H), 7.10 (d, J = 2.0 Hz, 1H), 5.48 (s, 2H), 5.39 (s, 2H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.96, 155.34, 147.35, 145.85, 141.66, 137.35, 136.91, 134.67, 133.57, 128.42, 128.36, 127.94, 127.79, 127.74, 127.50, 116.61, 113.20, 105.18, 100.32, 98.93, 68.00, 66.97. **HRMS** (ESI, m/z) calculated for $C_{26}H_{21}BrN_3O_2$ $[M + H]^+$: 486.0812, found: 486.0817.

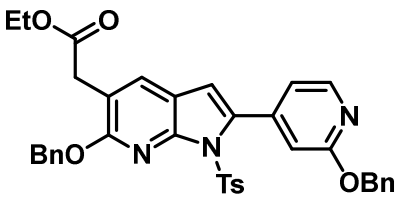
6-(Benzyloxy)-2-(2-(benzyloxy)pyridin-4-yl)-5-bromo-1-tosyl-1H-pyrrolo[2,3-*b*]pyridine (88)



To a suspension of sodium hydride (60 wt.%, 271 mg, 6.78 mmol, 1.1 eq.) in anhydrous DMF (12 mL) at 0 °C under argon, a solution of **83** (3.00 g, 6.17 mmol, 1 eq.) in anhydrous DMF (8 mL) was added dropwise over 15 minutes. The reaction mixture was allowed to warm to room temperature, and was further stirred for 1 hour. Then, a solution

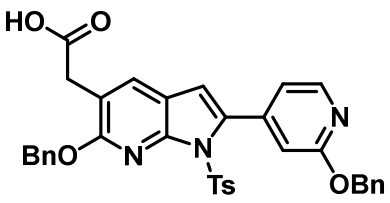
of tosyl chloride (1.76 g, 9.25 mmol, 1.5 eq.) in anhydrous DMF (5 mL) was added. After stirring at the same temperature for 3 hours, the reaction mixture was quenched by the addition of sat. aq. NH_4Cl (40 mL). Dichloromethane was added (80 mL), the phases were separated and the aqueous phase was extracted with dichloromethane (3 x 80 mL). The combined organic extracts were washed with brine (30 mL), dried over Na_2SO_4, filtered and evaporated under reduced pressure. The residue was redissolved in toluene (2 x 100 mL) and concentrated under vacuum to remove the remaining DMF, and was purified by column chromatography (DCM) to afford **88** as a yellow solid (2.43 g, 3.79 mmol, 61% yield). **R_f**: 0.24 (DCM). **m.p.**: 69.8 – 72.7 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.31 (d, J = 0.9 Hz, 1H), 8.25 (d, J = 5.3 Hz, 1H), 7.60 (d, J = 8.3 Hz, 2H), 7.52 (dd, J = 9.7, 7.6 Hz, 4H), 7.47 – 7.33 (m, 6H), 7.27 (d, J = 8.3 Hz, 2H), 7.20 (dd, J = 5.3, 1.4 Hz, 1H), 7.07 (s, 1H), 6.87 (s, 1H), 5.52 (s, 2H), 5.43 (s, 2H), 2.30 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 162.81, 155.70, 146.07, 145.94, 145.55, 142.96, 137.17, 137.10, 136.55, 135.70, 134.15, 129.87, 128.51, 128.37, 128.06, 127.97, 127.80, 127.58, 127.21, 118.33, 117.46, 110.97, 110.57, 102.82, 68.44, 67.17, 21.07. **HRMS** (ESI, m/z) calculated for $C_{33}H_{27}BrN_3O_4S$ $[M + H]^+$: 642.0885, found: 642.0889.

Ethyl 2-(6-(benzyloxy)-2-(2-(benzyloxy)pyridin-4-yl)-1-tosyl-1H-pyrrolo[2,3-*b*]pyridin-5-yl)acetate (89)[144b]



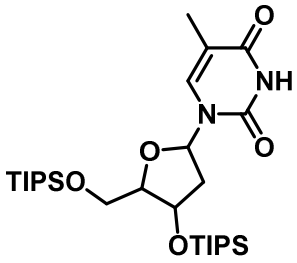
Mesitylene (0.75 mL) was added to an Ace pressure tube containing **88** (237 mg, 0.37 mmol, 1 eq.), ethyl potassium malonate (94 mg, 0.56, 1.5 eq.), $Pd_2(allyl)_2Cl_2$ (3 mg, 7 µmol, 2 mol%), BINAP (14mg, 22 µmol, 6 mol%) and DMAP (5 mg, 37 µmol, 0.1 eq.) under an argon atmosphere, and the mixture was stirred at 140 °C for 20 hours. After cooling to room temperature, the solvent was evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 99:1) to afford **89** as a yellow solid (147 mg, 0.23 mmol, 61% yield). **R_f**: 0.39 (DCM/MeOH 99:1). **m.p.**: 110.8 – 113.4 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 8.24 (dd, J = 5.3, 0.5 Hz, 1H), 7.87 (s, 1H), 7.64 (d, J = 8.4 Hz, 2H), 7.52 – 7.34 (m, 10H), 7.27 (d, J = 8.0 Hz, 2H), 7.20 (dd, J = 5.3, 1.4 Hz, 1H), 7.06 (dd, J = 1.3, 0.6 Hz, 1H), 6.92 (s, 1H), 5.44 (s, 2H), 5.43 (s, 2H), 3.99 (q, J = 7.1 Hz, 2H), 3.67 (s, 2H), 2.29 (s, 3H), 1.07 (t, J = 7.1 Hz, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 170.57, 162.80, 158.56, 146.49, 145.84, 145.31, 143.42, 137.23, 137.02, 135.86, 134.46, 133.67, 129.78, 128.39, 128.37, 128.08, 127.82, 127.74, 127.31, 118.37, 115.74, 115.17, 111.74, 110.42, 67.44, 67.13, 60.25, 35.25, 21.08, 13.95. **HRMS** (ESI, m/z) calculated for $C_{37}H_{34}N_3O_6S$ $[M + H]^+$: 648.2163, found: 648.2184.

2-(6-(benzyloxy)-2-(2-(benzyloxy)pyridin-4-yl)-1-tosyl-1H-pyrrolo[2,3-*b*] pyridin-5-yl)acetic acid (82)



LiOH (11 mg, 0.48 mmol, 3 eq.) was added to a solution of **89** (103 mg, 0.16 mmol, 1 eq.) in THF/H_2O (2 mL/0.4 mL) at room temperature. After stirring the mixture at the same temperature for 16 hours, the solvent was removed under reduced pressure and the residue was redissolved in dichloromethane (10 mL). The solution was washed with water (2 x 5 mL) and brine (5 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeOH 98:2) and washed with cyclohexane to afford **82** as an off-white solid (62 mg, 0.10 mmol, 63% yield). **R_f**: 0.21 (DCM/MeOH 98:2). **m.p.**: 100.2 – 103.1 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 12.38 (br s, 1H), 8.24 (d, *J* = 5.4 Hz, 1H), 7.85 (s, 1H), 7.63 (d, *J* = 8.4 Hz, 2H), 7.50 (t, *J* = 7.5 Hz, 4H), 7.44 – 7.38 (m, 4H), 7.38 – 7.32 (m, 2H), 7.26 (d, *J* = 8.1 Hz, 2H), 7.18 (dd, *J* = 5.3, 1.4 Hz, 1H), 7.05 (d, *J* = 6.4 Hz, 1H), 6.91 (s, 1H), 5.46 (s, 2H), 5.43 (s, 2H), 3.61 (s, 2H), 2.29 (s, 3H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 172.07, 162.79, 158.62, 146.35, 145.84, 145.29, 143.44, 137.23, 137.20, 135.73, 134.51, 133.62, 129.79, 128.39, 128.07, 127.81, 127.63, 127.31, 127.15, 118.37, 115.72, 111.72, 110.41, 67.32, 67.12, 35.23, 26.34, 21.09. **HRMS** (ESI, m/z) calculated for $C_{35}H_{30}N_3O_6S$ $[M + H]^+$: 620.1850, found: 620.1846.

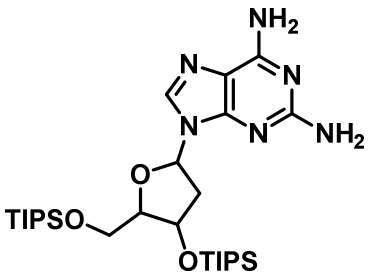
TIPS-Thymidine (90)[150a]



Thymidine (727 mg, 3.00 mmol, 1 eq.) and imidazole (1.37 g, 20.10 mmol, 6.7 eq.) were dissolved in anhydrous DMF (9 mL) under an argon atmosphere. The mixture was stirred at room temperature for 10 minutes, until the solution was clear, and then TIPS chloride (1.91 g, 2.11 mL 9.90 mmol, 3.3 eq.) was added. After stirring the reaction mixture at the same temperature for 16 hours, the solvent was removed under reduced pressure and the residue was redissolved in dichloromethane (100 mL). The solution was washed with water (2 x 50 mL) and brine (30 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 3:1) to afford **90** as a white solid (1.49 g, 2.69 mmol, 90% yield). **R_f**: 0.35 (Cy/EtOAc 3:1). **m.p.**: 128.1 – 128.8 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 11.34 (s, 1H), 7.41 – 7.36 (m, 1H), 6.17 (dd, *J* = 8.2, 5.9 Hz, 1H), 4.57 – 4.48 (m, 1H), 3.90 – 3.75 (m, 3H), 2.24 (ddd, *J* = 13.6, 8.1, 5.8 Hz, 1H), 2.13 (ddd, *J* = 13.1, 5.9, 2.4 Hz, 1H), 1.76 (d, *J* = 1.0 Hz, 3H), 1.15 – 1.00 (m, 42H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 163.61, 150.37, 135.47, 109.49, 87.11, 83.80,

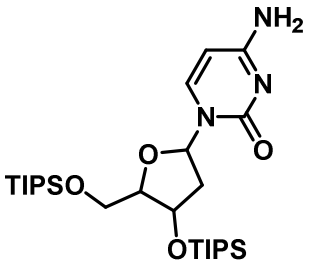
72.28, 63.19, 17.77, 17.74, 12.12, 11.50, 11.29. **HRMS** (ESI, m/z) calculated for $C_{28}H_{55}N_2O_5Si_2$ $[M + H]^+$: 555.3644, found: 555.3647; calculated for $C_{28}H_{54}N_2O_5Si_2Na$ $[M + Na]^+$: 577.3463, found: 577.3469.

TIPS-Deoxyaminoadenosine (91)



2'-Deoxy-2-aminoadenosine (799 mg, 3.00 mmol, 1 eq.) and imidazole (1.37 g, 20.10 mmol, 6.7 eq.) were dissolved in anhydrous DMF (9 mL) under an argon atmosphere. The mixture was stirred at room temperature for 20 minutes, until the solution was clear, and then TIPS chloride (1.91 g, 2.11 mL 9.90 mmol, 3.3 eq.) was added. After stirring the reaction mixture at the same temperature for 16 hours, the solvent was removed under reduced pressure and the residue was redissolved in dichloromethane (100 mL). The solution was washed with water (2 x 50 mL) and brine (30 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (Cy/EtOAc 1:3) to afford **91** as a white solid (1.53 g, 2.65 mmol, 88% yield). $\mathbf{R_f}$: 0.28 (Cy/EtOAc 1:3). **m.p.**: 70.1 – 77.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 7.87 (s, 1H), 6.71 (br s, 2H), 6.18 (dd, J = 8.1, 5.9 Hz, 1H), 5.76 (br s, 2H), 4.66 (dt, J = 5.0, 2.4 Hz, 1H), 3.93 – 3.88 (m, 1H), 3.84 (dd, J = 10.8, 5.8 Hz, 1H), 3.73 (dd, J = 10.8, 4.2 Hz, 1H), 2.80 (ddd, J = 13.3, 8.2, 5.4 Hz, 1H), 2.27 (ddd, J = 13.0, 5.8, 2.5 Hz, 1H), 1.12 – 0.98 (m, 42H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 160.29, 156.13, 151.61, 135.14, 113.32, 87.28, 82.22, 72.63, 63.37, 17.85, 17.77, 11.54, 11.32. **HRMS** (ESI, m/z) calculated for $C_{28}H_{55}N_6O_3Si_2$ $[M + H]^+$: 579.3869, found: 579.3874; calculated for $C_{28}H_{54}N_6O_3Si_2Na$ $[M + Na]^+$: 601.3688, found: , 601.3692.

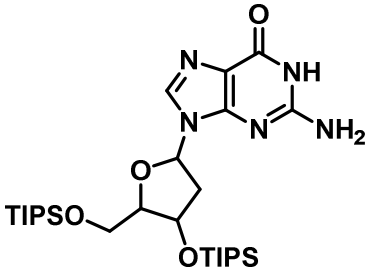
TIPS-Deoxycytidine (92)[150b]



2'-Deoxycytidine (682 mg, 3.00 mmol, 1 eq.) and imidazole (1.37 g, 20.10 mmol, 6.7 eq.) were dissolved in anhydrous DMF (9 mL) under an argon atmosphere. The mixture was stirred at room temperature for 10 minutes, until the solution was clear, and then TIPS chloride (1.91 g, 2.11 mL 9.90 mmol, 3.3 eq.) was added. After stirring the reaction mixture at the same temperature for 16 hours, the solvent was removed under reduced pressure and the residue was redissolved in dichloromethane (100 mL). The solution was washed with water (2 x 50 mL) and brine (30 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeCN 1:1) to afford **92** as a white solid (1.29 g, 2.39 mmol, 80% yield). $\mathbf{R_f}$: 0.23 (DCM/MeCN 1:1). **m.p.**: 101.5 – 107.8 °C. **^{1}H NMR**

(400 MHz, DMSO-d_6) δ (ppm): 7.70 (d, J = 7.4 Hz, 1H), 7.16 (br s, 2H), 6.19 (t, J = 6.5 Hz, 1H), 5.69 (d, J = 7.4 Hz, 1H), 4.51 (dt, J = 5.5, 2.9 Hz, 1H), 3.91 – 3.84 (m, 2H), 3.83 – 3.76 (m, 1H), 2.20 (ddd, J = 13.0, 5.9, 3.4 Hz, 1H), 2.09 – 2.00 (m, 1H), 1.15 – 1.01 (m, 42H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 165.49, 154.85, 140.25, 93.79, 86.98, 84.63, 71.93, 63.01, 40.92, 17.79, 17.77, 11.53, 11.29. **HRMS** (ESI, m/z) calculated for $C_{27}H_{54}N_3O_4Si_2$ [M + H]$^+$: 540.3647, found: 540.3652; calculated for $C_{27}H_{53}N_3O_4Si_2Na$ [M + Na]$^+$: 562.3467, found: 562.3469.

TIPS-Deoxyguanosine (93)[150c]



2'-Deoxyguanosine (802 mg, 3.00 mmol, 1 eq.) and imidazole (1.37 g, 20.10 mmol, 6.7 eq.) were dissolved in anhydrous DMF (9 mL) under an argon atmosphere. The mixture was stirred at room temperature for 20 minutes, until the solution was clear, and then TIPS chloride (1.91 g, 2.11 mL 9.90 mmol, 3.3 eq.) was added. After stirring the reaction mixture at the same temperature for 16 hours, the solvent was removed under reduced pressure and the residue was redissolved in dichloromethane (100 mL). The solution was washed with water (2 x 50 mL) and brine (30 mL), dried over Na_2SO_4, filtered, evaporated under reduced pressure and the residue was purified by column chromatography (DCM/MeCN 1:1) to afford **93** as a white solid (1.57 g, 2.71 mmol, 90% yield). **R_f**: 0.18 (DCM/MeCN 1:1). **m.p.**: 172.2 – 177.6 °C. **^{1}H NMR** (400 MHz, DMSO-d_6) δ (ppm): 10.63 (s, 1H), 7.86 (s, 1H), 6.47 (br s, 2H), 6.13 (dd, J = 8.2, 5.8 Hz, 1H), 4.63 (dt, J = 5.1, 2.5 Hz, 1H), 3.93 – 3.87 (m, 1H), 3.79 (dd, J = 10.9, 5.8 Hz, 1H), 3.72 (dd, J = 10.9, 4.4 Hz, 1H), 2.71 (ddd, J = 13.3, 8.2, 5.4 Hz, 1H), 2.29 (ddd, J = 13.1, 5.8, 2.6 Hz, 1H), 1.13 – 0.99 (m, 42H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ (ppm): 156.70, 153.74, 151.08, 134.79, 116.70, 87.46, 82.12, 72.55, 63.31, 17.84, 17.75, 11.52, 11.30. **HRMS** (ESI, m/z) calculated for $C_{28}H_{54}N_5O_4Si_2$ [M + H]$^+$: 580.3709, found: 580.3715; calculated for $C_{28}H_{53}N_5O_4Si_2Na$ [M + Na]$^+$: 602.3528, found: 602.3534.

6. References

1. Hannon, M. J. *Chem. Soc. Rev.* **2007**, *36*(2), 280–295.

2. (a) International Human Genome Sequencing Consortium *Nature* **2004**, *431*, 931-945. (b) Pertea, M.; Salzberg, S. L. *Genome Biol.* **2010**, *11*(5), 206.

3. Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737-738.

4. (a)Franklin, R. E.; Gosling, R. G. *Nature* **1953**, *171*, 740-741. (b) Wilkins, M. H. F.; Stokes, A. R.; Wilson, H. R. *Nature* **1953**, *171*, 738-740.

5. (a) Hargittai, I. *Struct. Chem.* **2009**, *20*, 753-756. (b) Frixione, E.; Ruiz-Zamarripa, L. *J. Biol. Chem.* **2019**, *294*(7), 2249-2255.

6. (a) Chargaff, E.; Vischer, E. *J. Biol. Chem.* **1949**, *177*, 405-16. (b) Chargaff, E.; Zamenhof, S.; Green, C. *Nature* **1950**, *165*, 756-757.

7. (a) Pauling, L.; Corey, R. B. *Proc. Nat. Acad. Sci. USA* **1953**, *39*(2), 84-97. (b) Pauling, L.; Corey, R. B. *Nature* **1953**, *171*(4347), 346-346.

8. Taken from https://x3dna.org/highlights/pauling-triplex-model-of-nucleic-acids-is-available-in-3dna. Accesed June 26, 2021.

9. Arnott, S. *Trends Biochem. Sci.* **2006**, *31*(6), 349-354.

10. Rich, A.; Zhang, S. *Nat. Rev. Genet.* **2003**, *4*(7), 566-572.

11. Wolberger, C. *Ann. Rev. Biophys. Biomol. Struct.* **1999**, *28*(1), 29–56.

12. Chenoweth, D. M., "Synthesys and structural studies of Py-Im polyamides", California, **2009**.

13. (a) Jackson, S.; Bartek, J. *Nature* **2009**, *461*, 1071-1078. (b) Lieber, M. R. *Annu. Rev. Biochem.* **2010**, *79*, 181-211. (c) Moynahan M. E.; Jasin, M. *Nat. Rev. Mol. Cell. Biol.* **2010**, *11*(3), 196-207.

14. Urnov, D. F.; Rebar, E. J.; Holmes, M. C.; Zhang, H. S.; Gregory, P. D. *Nat. Rev. Genet.* **2010**, *11*, 636-646.

15. Kim, H.; Kim, J. S. *Nat. Rev. Genet.* **2014**, *15*, 321-334.

16. (a) Porteus, M. H.; Baltimore, D. *Science* **2003**, *300*(5620), 763. Bibikova, M.; Beumer, K.; Trautman, J. K.; Caroll, D. *Science* **2003**, *300*(5620), 764.

17. Kim, Y. G.; Cha, J.; Chandrasegaran, S. *Proc. Natl. Acad. Sci. USA* **1996**, *93*(3), 1156-1160.

18. Bitinaite, J.; Wah, D. A.; Aggarwal, A. K.; Schildkraut I. *Proc. Natl. Acad. Sci. USA* **1998**, *95*(18), 10570-10575.

19. (a) Miller, J. C. *et al. Nat. Biotechnol.* **2007**, *25*, 778-785. (b) Szczepek, M.; Brondani, V.; Büchel, J.; Serrano, L.; Segal, D. J.; Cathomen, T. *Nat. Biotechnol.* **2007**, *25*, 786-793.

20. Guo J.; Gaj, T.; Barbas, C. F. *J. Mol. Biol.* **2010**, *400*(1), 96-107.

21. Carroll, D. *Genetics* **2011**, *188*(4), 773-782.

22. Tupler, R.; Perini, G.; Green, M. R. *Nature* **2001**, *409*, 832-833.
23. Hanas, J. S.; Hazuda, D. J.; Bogenhagen, D. F.; Wu, F. W.-H; Wu, C.-W. *J. Biol. Chem.* **1983**, *258*(23), 14120-14125.
24. Dynan, W. S.; Tijan, R. *Cell* **1983**, *32*(3), 669-680.
25. Pavletich, N. P.; Pabo, C. O. *Science* **1991**, *252*(5007), 809-881.
26. Leon, O.; Roth, M. *Biol. Res.* **2000**, *33*(1), 21-30.
27. Klug, A. *Q. Rev. Biophys.* **2010**, *43*(1), 1-21.
28. Wolfe, S. A.; Nekludova, L.; Pabo, C. O. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *3*, 183-212.
29. Cathomen, T.; Joung, J. K. *Mol. Ther.* **2008**, *16*(7), 1200-1207.
30. (a) Perez, E. E. *et al. Nat. Biotechnol.* **2008**, *26*, 808-816. (b) Tebas P. *et al. N. Engl. J. Med.* **2014**, *370*, 901-910. (c) Gupta, R. K. *et al. Nature* **2019**, *568*, 244-248.
31. *(a) Moran, K. et al. Blood* **2018**, *132*(1), 2190. (b) Sangamo Therapeutics, Inc. April 2, 2019. https://investor.sangamo.com/news-releases/news-release-details/sangamo-provides-clinical-development-update-including-early. Accessed June 16, 2021.
32. Boch, J.; Bonas, U. *Annu. Rev. Phytopathol.* **2010**, *48*, 419-436.
33. (a) Moscou, M. J.; Bogdanove, A. J. *Science* **2009**, *326*(5959), 1501. (b) Boch J.; Scholze, H.; Schornack, S. *et al. Science* **2009**, *326*(5959), 1509-1512.
34. (a) Doyle, E. L.; Hummel, A. W.; Demorest, Z. L. *et al. PLoS One* **2013**, *8*(12), e82120. (b) Lamb, B. M.; Mercer, A. C.; Barbas, C. F. *Nucleic Acids Res.* **2013**, *41*(21), 9779-9785.
35. Streubel, J.; Blücher, C.; Landgraf, A.; Boch, J. *Nat. Biotechnol.* **2012**, *30*(7), 593-595.
36. (a) Mak, A. N.-S.; Bradley, P.; Cernadas, R. A.; Bogdanove, A. J.; Stoddard, B. L. *Science* **2012**, *335*(6069), 716-719. (b) Deng, D.; Yan, C.; Pan, X. *et al. Science* **2012**, *335*(6069), 720-723.
37. (a) Benjamin, R.; Graham, C.; Yallop, D.; Jozwik, A. *et al. Blood* **2018**, 132, 896. (b) Cellectis. January 15, 2020. https://www.cellectis.com/en/press/first-patient-dosed-with-cellectis-new-allogeneic-ucart123-product-candidate-for-relapsed-refractory-acute-myeloid-leukemia/. Accessed June 17, 2021.
38. A PubMed search for publications that had the words "CRISPR" or "Cas9" in the abstract or title showed 24,600 results, 6,000 of them from 2020. Each year the number of publications keeps increasing, illustrating the activity in this field.
39. The Nobel Foundation. October 7, 2020. https://www.nobelprize.org/prizes/chemistry/2020/press-release/. Accesed June 20, 2021.
40. Jansen, R.; van Embden, J. D. A.; Gaastra, W.; Schouls, L. M. *Mol. Microbiol.* **2002**, *43*(6), 1565-1575.
41. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. *J.Bacteriol.* **1987**, *169*, 5429–5433.

42. (a) Pourcel, C.; Salvignol, G.; Vergnaud, Vergnaud, G. *Microbiology* **2005**, *151*, 653-663. (b) Mojica, F.J.; Díez-Villaseñor, C.; García-Martínez, J; Soria, E. *J. Mol. Evol.* **2005**, 60, 174-182. (c) Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, D. *Microbiology* **2005**, *151*, 2551-2561.
43. Barrangou, R.; Fremaux, C.; Deveau, H. *et al. Science* **2007**, *315*(5819), 1709–1712.
44. Hille, F.; Richter, H.; Wong, S. P.; Bratovič, M.; Ressel, S.; Charpentier, E. *Cell* **2018**, *172*, 1239-1259.
45. Shmakov, S.; Abudayyeh, O. O.; Makarova, K. S. et al. Mol. *Cell* **2015**, *60*(3), 385-297.
46. Deltcheva, E.; Chylinski, K.; Sharma, C. *et al.* M. *Nature* **2011**, *471*, 602-607.
47. Junek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. *Science* **2012**, *337*(6096), 816-821.
48. Jiang, F.; Doudna, J. A. *Annu. Rev. Biophys.* **2017**, *46*, 505-529.
49. Ran, F. A.; Hsu, P. D.; Lin, C.-Y. *Cell* **2013**, *154*, 1–10.
50. Chen, B.; Gilbert, L. A.; Cimini, B. A. *et al. Cell* **2013**, *155*, 1479–1491.
51. Lu, Y.; Xue, J.; Deng, T. *J. Clin. Oncol.* **2018**, 36, 3050.
52. Eissenberg, J. C. *Biomol. Concepts* **2021**, *11*, 1-7.
53. Greeli, H. T. *Journal of Law and the Biosciences* ***2019***, *6*(1), 111–183.
54. Lander, E. S.; Baylis, F.; Zhang, F.; Charpentier, E.; Berg, P.; Bourgain, C.; …; Winnacker, E.-L. *Nature* **2019**, *567*(7747), 165–168.
55. Nature. January 3, 2020. https://www.nature.com/articles/d41586-020-00001-y. Accessed June 16, 2021.
56. *de Silva, E.; Stumpf M. P. H. FEMS Microbiol. Lett.* ***2004***, *241*(1), 1–12.
57. Li, Q.; Qin, Z.; Wang, Q.; Xu, T.; Yang, Y.; He, Z. *Comput. Struct. Biotechnol. J.* **2019**, *17*, 689–698.
58. Rulli, T. *Bioethics* **2019**, *33*(9), 1072-1082.
59. Bosley, K. S.; Botchan, M.; Bredenoord, A. L.; Carroll, D.; Charo, R. A.; Charpentier, E.; …; Zhou, Q. *Nat. Biotechnol.* **2015**, *33*, 478-496.
60. Reuters. March 20, 2019. https://www.reuters.com/article/us-health-who-gene-editing/who-panel-calls-for-registry-of-all-human-gene-editing-research-idUSKCN1R02IC. Accessed June 16, 2021.
61. (a) Rakozcy, P. E. *Methods Mol. Med.* **2001**, *47*, 89-104. (b) Buchini, S; Leumann, C. J. *Curr. Opin. Chem. Biol.* **2003**, *7*(6), 717-726.
62. Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023–2024.
63. Maldonado, R.; Filarsky, M.; Grummt, I.; Längst, G. *RNA* **2018**, *24*, 371-380.
64. Taken from RCSB PDB. NDB: 1BWG.
65. Morgan, A. R.; Wells, R. D. *J. Mol. Biol.* **1968**, *37*(1), 63-80.

66. Beal, P. A.; Dervan, P. B. *Science* **1991**, *251*(4999), 1360–1363.
67. Blume, S. W.; Lebowitz, J.; Zacharias, W. *et al. Nucleic Acids Res.* **1999**, *27*(2), 695–702.
68. Pesco, J.; Salmon, J.-M.; Vigo, J.; Viallet, P. *Anal. Biochem.* **2001**, *290*, 221–231.
69. Asensio, J. L.; Lane, A. N.; Dhesi, J.; Bergqvist, S.; Brown, T. *J. Mol. Biol.* **1998**, *275*, 811-822.
70. Verdolino, V.; Cammi, R.; Munk, B. H.; Schegel, H. B. *J. Phys. Chem. B* **2008**, *112*, 16860–16873.
71. Esguerra, M.; Nilsson, L.; Villa, A. *Nucleic Acids Res.* **2014**, *42*(18), 11329–11338.
72. Nielsen, P.; Egholm, M.; Berg, R.; Buchardt, O. *Science* **1991**, *254*(5037), 1497–1500.
73. (a) Nielsen, P. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 167-183. (b) Ricciardi, A. S.; McNeer, N. A.; Anandalingam, K. K.; Saltzman, W. M.; Glazer, P. M. *Methods Mol. Biol.* **2014**, *1176*, 89-106.
74. Demidov, V. V.; Potaman, V. N.; Frank-Kamenetskil, M. D.; Egholm, M.; Buchard, O.; Sönnichsen, S. H.; Nielsen, P. E. *Biochem. Pharmacol.* **1994**, *48*(6), 1310–1313.
75. Nielsen, P. E.; Egholm, M. *Curr. Issues Mol. Biol.* **1999**, *1*(2), 89-104.
76. Nielsen, P. E. *ChemBioChem* **2010**, *11*, 2073-2076.
77. (a) Janowski, B. A.; Kaihatsu, K.; Huffman, K. E. *et al. Nature Chem. Biol*. **2005**, *1*(4), 210-215. (b) Knudsen, H.; Nielsen, P. E. *Nucleic Acids Res.* ***1996***, *24*, 494–500.
78. Rogers, F. A.; Vasquez, K. M.; Egholm, M.; Glazer, P. M. *Proc. Nat. Acad. Sci. USA* **2002**, *99*(26), 16695-16700.
79. Braasch, D. A.; Corey, D. R. *Methods* **2001**, *23*(2), 97-107.
80. Wittung, P.; Kajanus, J.; Edwards, K.; Nielsen, P.; Nordén, B.; Malmström, B. G. *FEBS Letters* **1995**, 365, 27-29.
81. (a) McMahon, B. M.; Mays, D.; Lipsky, J.; Stewart, J. A.; Fauq, A.; Richelson, E. *Antisense Nucleic Acid Drug Dev.* **2002**, *12*(2), 65-70. (b) Nielsen, P. E. *Q. Rev. Biophys.* **2005**, *38*(4), 345-350.
82. Nielsen, P. E.; Haaima, G.; Lohse, A.; Buchardt, O. *Angew. Chem. Int. Ed.* **1996**, *35*(17), 1939-1942.
83. Zhou, P.; Wang, M.; Du, L.; Fisher, G. W.; Waggoner, A.; Ly, D. H. *J. Am. Chem. Soc.* **2003**, 125, 6878-6879.
84. Moccia, M.; Adamo, M. F. A.; Saviano, M. *Artif. DNA PNA XNA* **2014**, *5*(3), e1107176.
85. (a) Sahu, B.; Sacui, I.; Rapireddy, S.; Zanotti, K. J.; Bahal, R.; Armitage, B. A.; Ly, D. H. *J. Org. Chem.* **2011**, *76*, 5614-5627. (b) Quijano, E.; Bahal, R.; Ricciardi, A.; Saltzman, W. M.; Glazer, P. M. *Yale J. Biol. Med.* **2017**, *90*(4), 583-598.

86. (a) Park, J.-K.; Utsumi, T.; Seo, Y.-E.; Deng, Y.; Satoh, A.; Saltzman, W. M.; Iwakiri, Y. *Nanomedicine* **2016**, *12*(5), 1365-1374. (b) Deng, Y.; Yang, F.; Cocco, E.; Song, E. *et al. Proc. Nat. Acad. Sci. USA* **2016**, *113*(41), 11453-11458. (c) Gupta, A.; Bahal, R.; Gupta, M.; Glazer, P. M.; Saltzman, W. M. *J. Control. Release* **2016**, *240*, 302-311.
87. (a) Finlay, A.C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. *J. Am. Chem. Soc.* **1951**, *73*, 341-343. (b) Arcamone, F.; Penco, S.; Orezzi, P.; Nicolella, V.; Pirelli, A. *Nature* **1964**, *203*, 1064-1065.
88. (a) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Nat. Acad. Sci. USA* **1985**, *82*(5), 1376-1380. (b) Coll, M.; Frederick, C. A.; Wang, A. H.; Rich, A. *Proc. Nat. Acad. Sci. USA* **1987**, *82*(23), 8385-8389. (c) Pelton, J. G.; Wemmer, D. E. *Proc. Nat. Acad. Sci. USA* **1989**, *86*(15), 5723-5727.
89. (a) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*(23), 8783–8794. (b) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Nat. Acad. Sci. USA* **1992**, *89*(16), 7586–7590.
90. Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*(3), 284–299.
91. Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983–7988.
92. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.
93. Wang, C. C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 653-657.
94. Olenyuk, B.; Jitianu, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 4741–4751.
95. Heckel, A.; Dervan, P. B. *Chem. Eur. J.* **2003**, *9*, 3353-3366.
96. Melander, C.; Herman, D. M.; Dervan, P. D. *Chem. Eur. J.* **2000**, *6*, 4487-4497.
97. Kawamoto, Y.; Bando, T.; Sugiyama, H. *Bioorg. Med. Chem.* **2018**, *26*, 1393–1411.
98. Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*(6629), 202–205.
99. Chenoweth, D. M.; Dervan, P. B. *Proc. Nat. Acad. Sci. USA* **2009**, *106*(32), 13175–13179.
100. Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. *J. Am. Chem. Soc.* **2009**, *131*, 7182-7188.
101. Dervan, P. B. *Isr. J. Chem.* **2019**, *59*, 1–14.
102. Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1999**, 5, 975-983.
103. Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534–3535.
104. Hari, Y.; Obika, S.; Imanishi, T. *Eur. J. Org. Chem.* **2012**, *2012*, 2875-2887.
105. Griffin, L. C.; Kiessling, L. L.; Beal, P. A.; Gillespie, P.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 7976-7982.
106. (a) Huang, C. Y.; Miller, P. S. *J. Am. Chem. Soc.* **1993**, *115*, 10456-10457. (b) Huang, C.-Y.; Bi, G.; Miller, P. S. *Nucleic Acids Res.* **1996**, *24*, 2606-2613.
107. Purwanto, M. G. M.; Weisz, K. *Curr. Org. Chem.***2003**, *7*, 427-446.

108. (a) Zimmerman, S. C.; Schmitt, P. J. Am. Chem. Soc. **1995**, *117*, 10769-10770. (b) Mertz, E.; Mattei, S.; Zimmerman, S. C. *Org. Lett.* **2000**, *2*, 2931-2934. (c) Lengeler, D.; Weisz, K. Tetrahedron Lett. **2001**, *42*, 1479-1481.
109. Mertz, E.; Mattei, S.; Zimmerman, S. C *Bioorg. Med. Chem.* **2004**, *12*(6), 1517-1526.
110. (a) Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A. *et al. Biochemistry* **2010**, *49*, 7867-7878. (b) Guianvarc'h, D.; Benhida, R.; Fourrey, J.-L.; Maurisse, R.; Sun,J.-S. *Chem. Commun.* **2001**, 1814-1815. (c) Ohkubo, A.; Yamada, K.; Ito, Y.; Yoshimura, K. *et al. Nucleic Acids Res.* **2015**, *43*, 5675-5686.
111. Kumpina, I.; Brodyagin, N.; MacKay, J. A.; Kennedy, S. D.; Katkevics, M.; Rozners, E. *J. Org. Chem.* **2019**, *84*, 13276-13298.
112. Wang, W.; Purwanto, M. G. M.; Weisz, K. *Org. Biomol.Chem.* **2004**, *2*, 1194-1198.
113. Mitra, R.; Ganesh, K. N *J. Org. Chem.* **2012**, *77*, 5696-5704.
114. Thomson, S. A.; Josey, J. A.; Cadilla, R. *et al. Tetrahedron* **1995**, *51*(22), 6179-6194.
115. Feagin, T. A.; Shah, N. I.; Heemstra, J. M. *J. Nucleic Acids* **2012**, 354549.
116. Wojciechowski, F.; Hudson, R. H. E. *J. Org. Chem.* **2008**, *73*, 3807-3816.
117. Kosynkina, L.; Wang, W.; Liang, T. C. *Tetrahedron Lett.* **1994**, *35*(29), 5173-5176.
118. Carpino, L. A.; Ghassemi, S.; Ionescu, D. *Org. Proc. Res. Dev.* **2003**, *7*, 28-37.
119. (a) Markgraf, J. H.; Brown, H. B.; Mohr, S. C.; Peterson, R. G. *J. Am. Chem. Soc.* **1963**, *85*(7), 958-961. (b) Oae, S.; Kitao, T.; Kitaoka, Y. *J. Am. Chem. Soc.* **1962**, *84*(17), 3359-3362. (c) Tung, Y.-S.; Coumar, M. S.; Wu, Y.-S. *et al. J. Med. Chem.* **2011**, *54*, 3076-3080.
120. Reen, G. K.; Kumar, A.; Sharma, P. *Med. Chem. Res.* **2017**, *26*(12), 3336–3344. Prepared with a modified procedure employing PPA.
121. Gummadi, V. R.; Samajdar, S.; Mukherjee, S.; Bock, M. G.; Aurigene Discovery Technologies Limited. World patent WO 2017/009806 A1. 2017 Jan 19.
122. Yamabe, S.; Yamazaki, S. *J. Phys. Org. Chem*. **2016**, *29*(7), 361-367.
123. Lu, H.; Geng, Z.; Li, J.; Zou, D.; Wu, Y.; Wu, Y. *Org. Lett.* **2016**, *18*, 2774–2776.
124. Gummadi, V. R.; Samajdar, S.; Aurigene Discovery Technologies Limited. World patent WO 2015/104688 A1. 2015 Jul 16. Synthesized following a different synthetic pathway.
125. Kim, Y. K.; Park, S. Y; Joo, H. W. *et al.*; LG CHEM, LTD. U.S. patent US 2017/0349594 A1. 2017 Dec 7.
126. Li, Y.; Xie, Y.;Zhnag, R.; Jin, K.; Wang, X.; Duan, C. *J. Org. Chem.* **2011**, *76*, 5444–5449.
127. Zhuravlev, F. A. *Tetrahedron Lett.* **2006**, *47*, 2929-2932.
128. Yan, X.-M.; Mao, X.-R.; Huang, Z.-Z. *Heterocycles* **2011**, *83*, 1371-1376.
129. Alberico, D.; Scott, M. E.; Lautens, M. *Chem. Rev*. **2007**, *107*, 174-238.

130. Xu, D.; Xu, X.; Liu, Z.; Sun, L.-P.; You, Q. *Synlett* **2009**, *7*, 1172-1174.
131. (a) Basha, A.; Lipton, M.; Weinreb, S. M. *Tetrahedron Lett.* **1977**, *18*(48), 4171-4172. (b) Li, J.; Subramaniam, K.; Smith, D.; Qiao, J. X. *et al. Org. Lett.* 2011, *14*(1), 214-217.
132. Adjabeng, G.; Bifulco, N.; Davis-Ward, R. *et. al.*; Smithkline Becham Corporation. World patent WO 2009/076140 A1. 2009 Jun 18. Synthesized following a modified procedure.
133. Wang, D.; Zhao, J.; Wang, Y.; Hu, J. *et al. Asian J. Org. Chem.* **2016**, *5*, 1442-1446.
134. Widdowson, K. L.; GlaxoSmithKline Intellectual Property Development Limited. World patent WO 2016/027249 A1. 2016 Feb 26. Synthesized following a modified procedure.
135. Lo, J. M. H.; Marriott, R. A.; Giri, B. R.; Roscoe, J. M.; Klobukowski, M. *Can. J. Chem* **2010**, *88*, 1136-1145.
136. Urzúa, J. I.; Contreras, R.; Salas, C. O.; Tapia, R. A. *RSC Advances* **2016**, *6*(85), 82401-82408.
137. Feng, Y; Chen, Y. T.; Sessions, H. *et al.*. World patent WO 2011/050245 A1. 2011 Apr 28.
138. Dyson, P. J.; Jessop, P. G. *Catal. Sci. Technol.* **2016**, *6*(10), 3302-3316.
139. Aronov, A.; Bandarage, U. K.; Cottrell, K. M. *et al.*; Vertex Pharmaceuticals Incorporated. U.S. patent US 8,524,906 B2. 2013 Sep 3.
140. Vakalopoulos, A.; Brockschneider, D.; Wunder, F.; Stasch, J.-P.; Marquardt, T.; Dietz, L.; Bayer Pharma AG. World patent WO 2016/087343 A1. 2016 Jun 09.
141. Li, L.; Wang, Y.; Ren, P.; Liu, Y.; Der, R. K.; Kura Oncology Inc.. World patent WO 2018/106818 A1. 2018 Jun 14.
142. Bloomgren, P. A.; Currie, K. S.; Kropf, J. E. *et al.*; Gilead Connecticut, Inc.. U.S. patent 9,562,056 B2. 2017 Feb 7.
143. Ichikawa, M.; Matsunaga, H.; Oginoya, N. *et al.*; Japan patent JP 2011-178779 A. 2011 Sep. 15.
144. (a) Shang, R.; Ji, D.-S.; Chu, L.; Fu, Y.; Liu, L. *Angew. Chem. Int. Ed.* **2011**, *50*, 4470-4474. *tert*-Butyl potassium malonate was synthesized following a modified procedure, starting from *tert*-butyl malonic acid (b) Feng, Y.-S.; Wu, W.; Xu, Z.-Q.; Li, Y.; Li, M.; Xu, H.-J. *Tetrahedron* **2012**, *68*, 2113-2120.
145. Fors, B. P.; Buchwald, S. L. *J. Am. Chem. Soc.* **2009**, *131*, 36, 12898-12899.
146. MacKinnon, K. F.; Qualley, D. F.; & Woski, S. A. *Tetrahedron Lett.* **2007**, *48*(45), 8074-8077.
147. Gadre, S.; Peters, M. S.; Serrano, A.; Schrader, T. *Org. Lett.* **2018**, *20*, 6961–6964.
148. (a) Arcadi, A.; Cacchi, S.; Marinelli F. *Tetrahedron Lett.* **1992**, *33*(27), 3915-3918. (b) Cacchi, S.; Fabrizi, G.; Parisi, L. M. *J. Comb. Chem.* **2005**, *7*(4), 510–512.

149. Holenz, J.; Karlström, S.; Kihlström, J. *et al.*; AstraZeneca AB. World patent WO 2011/002409 A1. 2016 Jan 6.
150. (a) Saudi, M.; Zmurko, J.; Kaptein, S.; Rozenski, J.; Neyts, J.; Van Aerschot, A. *Eur. J. Med. Chem.* **2014**, *76*, 98-109. (b) Burnham, B. S.; Wyrick, S. D.; Hall, I. H.; Sood, A.; Spielvogel, B. F. *J. Labelled Comp. Radiopharm.* **1991**, *29*(4), 469-473. (c) Marsh, A.; Alcock, N. W.; Errington, W.; Sagar, R. *Tetrahedron* **2003**, *59*(29), 5595-5601. All nucleosides were synthesided following the same modified procedure.
151. (a) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, 112, 5, 2008-2010. (b) Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991**, 113, 8, 2810–2819.
152. (a) Loppinet-Serani, A.; Charbonnier, F.; Rolando, C.; Huc, I. *J. Chem. Soc.*, Perkin Trans. 2, **1998**, 937-942. (b) Beak, P.; Covington, J. B.; Smith, S. G.; White, J. M.; Zeigler, J. M. *J. Org. Chem.* **1980**, *45*, 1354-1362.
153. (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579-2619. (b) Krishna, M. S.; Toh, D.-F. K.; Meng, Z.; Ong, A. A. L.; Wang, Z.; Lu, Y.; Xia, K.; Prabakaran, M.; Chen, G. *Anal. Chem.* **2019**, *91*, 5331-5338. (c) Vilaivan, T. *Beilstein J. Org. Chem.* **2018**, *14*, 253-281.

7. Appendix

7.1. List of abbreviations

A	Adenine
AA	TIPS-deoxyaminoadenosine
Å	Angstrom
Abs.	Absorbance
Ac	Acetyl
Ar	Aryl
aeg	*N*-(2-aminoethyl)-glycine
BINAP	2,2′-bis(diphenylphosphino)-1,1′-binaphthyl
bp	Base pair
Bn	Benzyl
Boc	*tert*-Butyloxycarbonyl
C	Cytosine, TIPS-deoxycytidine
Cas	CRISPR-associated
CD	Circular Dichroism
CDI	*N*,*N*′-Carbonyldiimidazole
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
Cy	Cyclohexane
Cys	Cysteine
D	Aspartic acid
dCas9	Dead Cas9
DCM	Dichloromethane
DIPEA	*N*,*N*-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMEDA	*N*,*N*-Dimetylethylenediamine

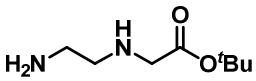
DMF	*N*,*N*-Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
Dp	Dimethylaminopropylamide
DSB	Double-stranded break
dsDNA	Double-stranded DNA
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EGFP	Enhanced Green Fluorescent Protein
Et	Ethyl
Fmoc	Fluorenylmethyloxycarbonyl
G	Guanine, TIPS-deoxyguanosine, glycine
h	hour(s)
HBTU	2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HDR	Homology-directed repair
HIV	Human Immunodeficiency Virus
His	Histidine
HOBt	1-Hydroxybenzotriazole
Hp	Hydroxypyrrole
HPLC	High-Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
Hz	Hertz
I	Isoleucine
Im	Imidazole
indel	Insertion/deletion
Int.	Intensity
^{i}Pr	Isopropyl
K	Lysine

LC-MS	Liquid Cromatograpy-Mass Spectrometry
M	Molar
*m*CPBA	*meta*-Cloroperbenzoic acid
Me	Methyl
min	minute(s)
MM	Molecular Mechanics
MOM	Methoxymethyl
m.p.	Melting point
mRNA	Messenger ribonucleic acid
N	Asparagine
NBS	*N*-Bromosuccinimide
NHEJ	Non-homologous end joining
NIS	*N*-Iodosuccinimide
nm	Nanometers
NMR	Nuclear Magnetic Resonance
nt	Nucleotide
NUC	Nuclease
PAM	Protospacer Adjacent Motif
Ph	Phenyl
pin	Pinacolato
PNA	Peptide Nucleic Acid
PPA	Polyphosphoric acid
pre-crRNA	Precursor CRISPR RNA
Py	Pyrazole
PyBroP	Bromotripyrrolidinophosphonium hexafluorophosphate
REC	Recognition
R_f	Retention factor
RNA	Ribonucleic acid

rt	Room temperature
RVD	Repeat Variable Diresidue
SET	Single electron transfer
SPPS	Solid-Phase Peptide Synthesis
sgRNA	single-guide RNA
ssODN	Single-strand oligodeoxynucleotide
Su	Succinimide
T	Thymine, TIPS-thymidine
TALE	Transcription Activator-Like Effector
TALEN	Transcription Activator-Like Effector Nucleases
TBTU	2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethylaminium tetrafluoroborate
^{t}Bu	*tert*-Butyl
TIPS	Triisopropylsilane
tracrRNA	trans-activating CRISPR RNA
TFA	Trifluoroacetic acid
TFAA	Trifluoroacetic anhydride
TFO	Triplex Forming Oligonucleotide
THF	Tetrahydrofuran
TLC	Thin-layer chromatography
TMS	Trimethylsilyl
Ts	Tosyl
U	Uracil
UV-Vis	Ultraviolet-visible
WHO	World Health Organization
ZFN	Zinc Finger Nuclease
ZFP	Zinc Finger Protein

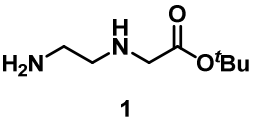
7.2. NMR spectra

^{1}H NMR (300 MHz, $CDCl_3$)



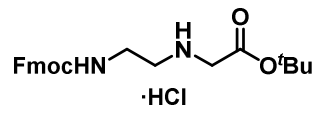
1

^{13}C NMR (75 MHz, $CDCl_3$)



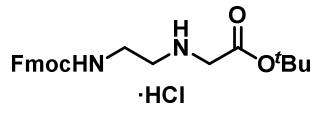
1

^{1}H NMR (300 MHz, DMSO-d_6)



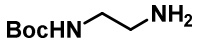
2

^{13}C NMR (75 MHz, DMSO-d_6)



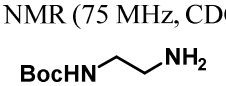
2

^{1}H NMR (300 MHz, $CDCl_3$)



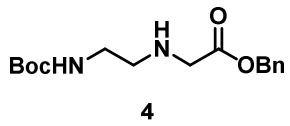
3

^{13}C NMR (75 MHz, $CDCl_3$)



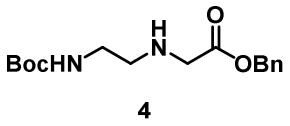
3

^{1}H NMR (300 MHz, $CDCl_3$)



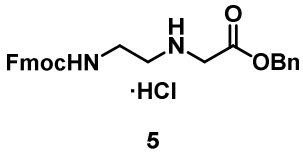
4

^{13}C NMR (75 MHz, $CDCl_3$)



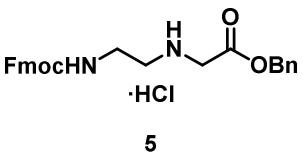
4

^{1}H NMR (300 MHz, DMSO-d_6)



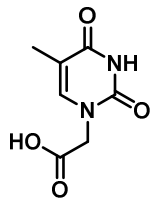
5

^{13}C NMR (75 MHz, DMSO-d_6)



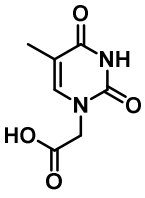
5

^{1}H NMR (300 MHz, DMSO-d_6)



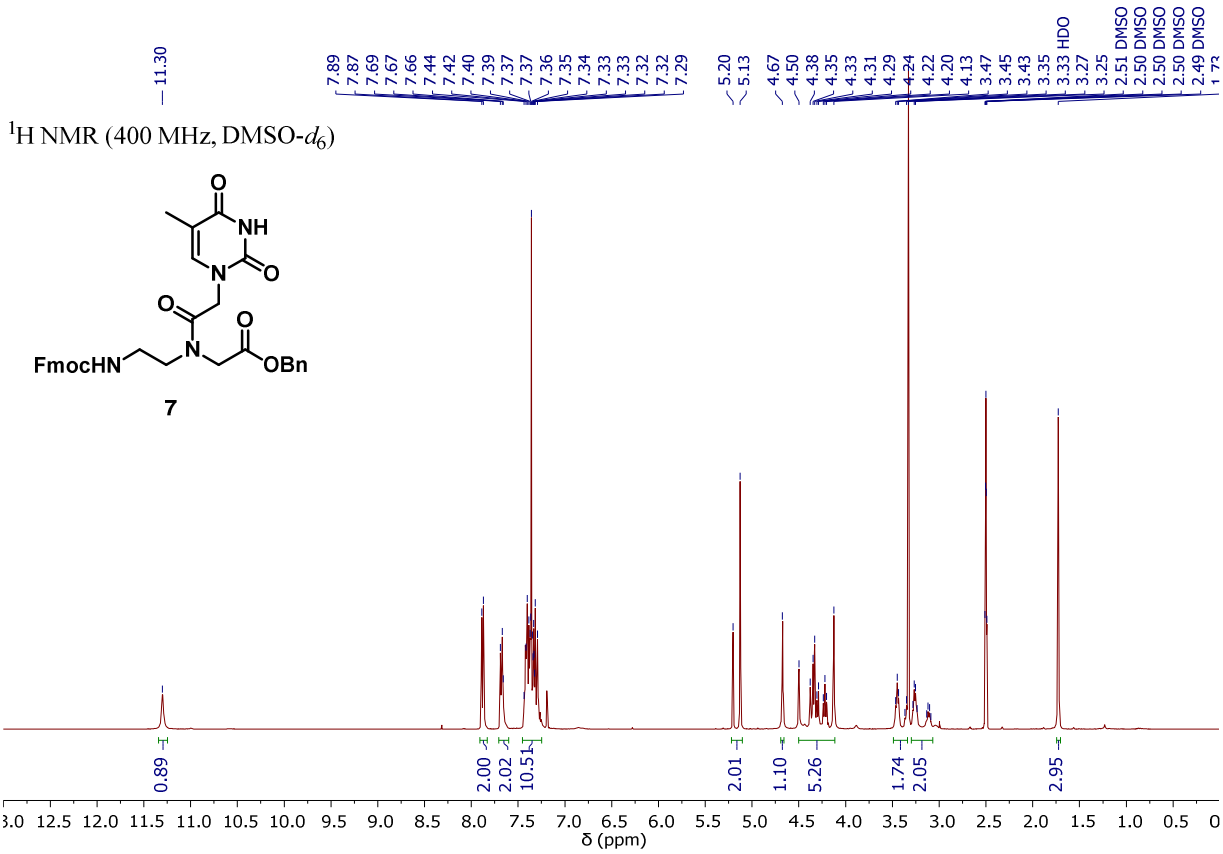
6

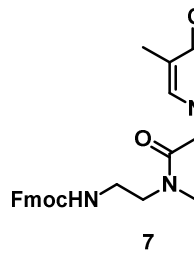
^{13}C NMR (75 MHz, DMSO-d_6)



6

^{1}H NMR (400 MHz, DMSO-d_6)



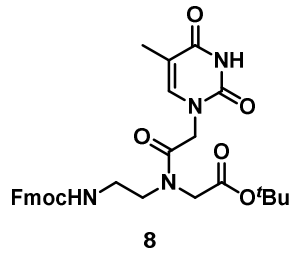


7

^{13}C NMR (101 MHz, DMSO-d_6)

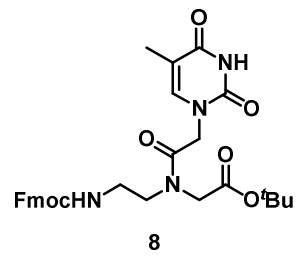
7

^{1}H NMR (400 MHz, DMSO-d_6)



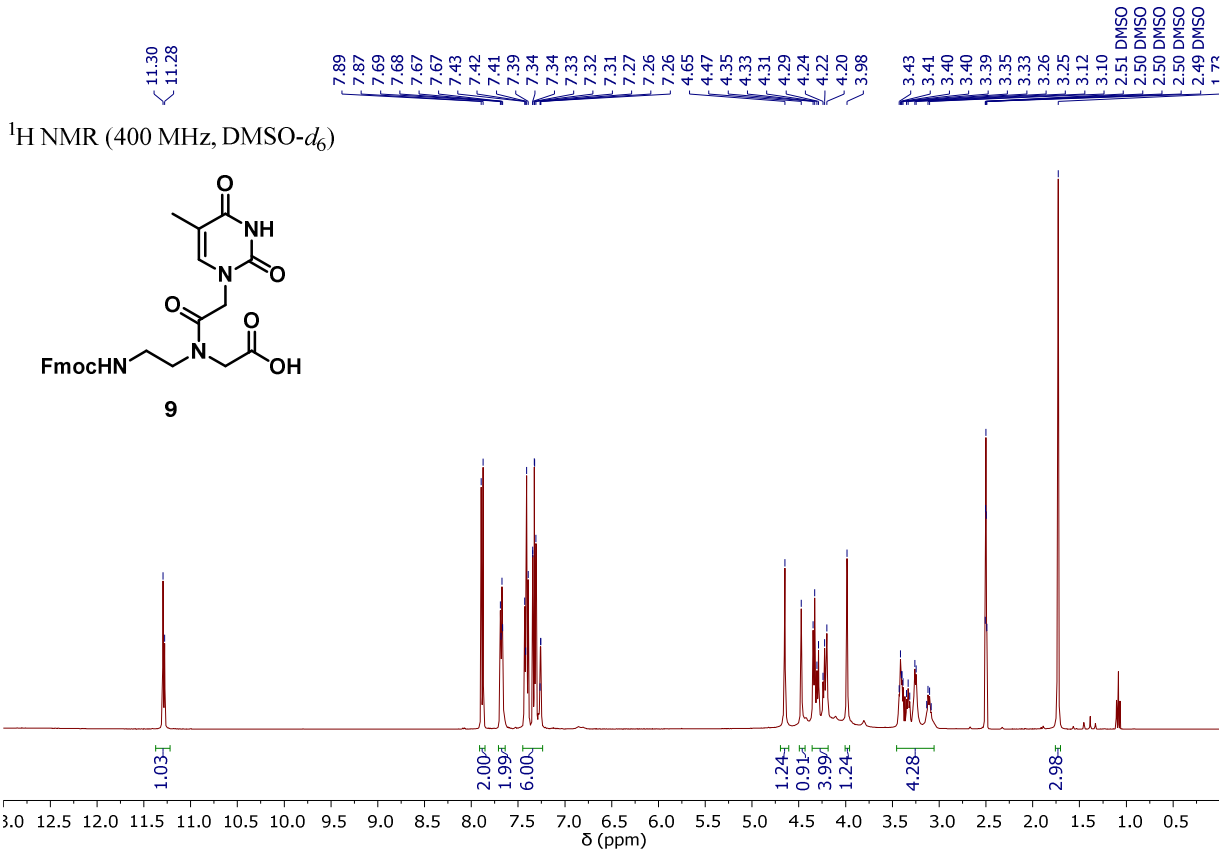
8

^{13}C NMR (101 MHz, DMSO-d_6)



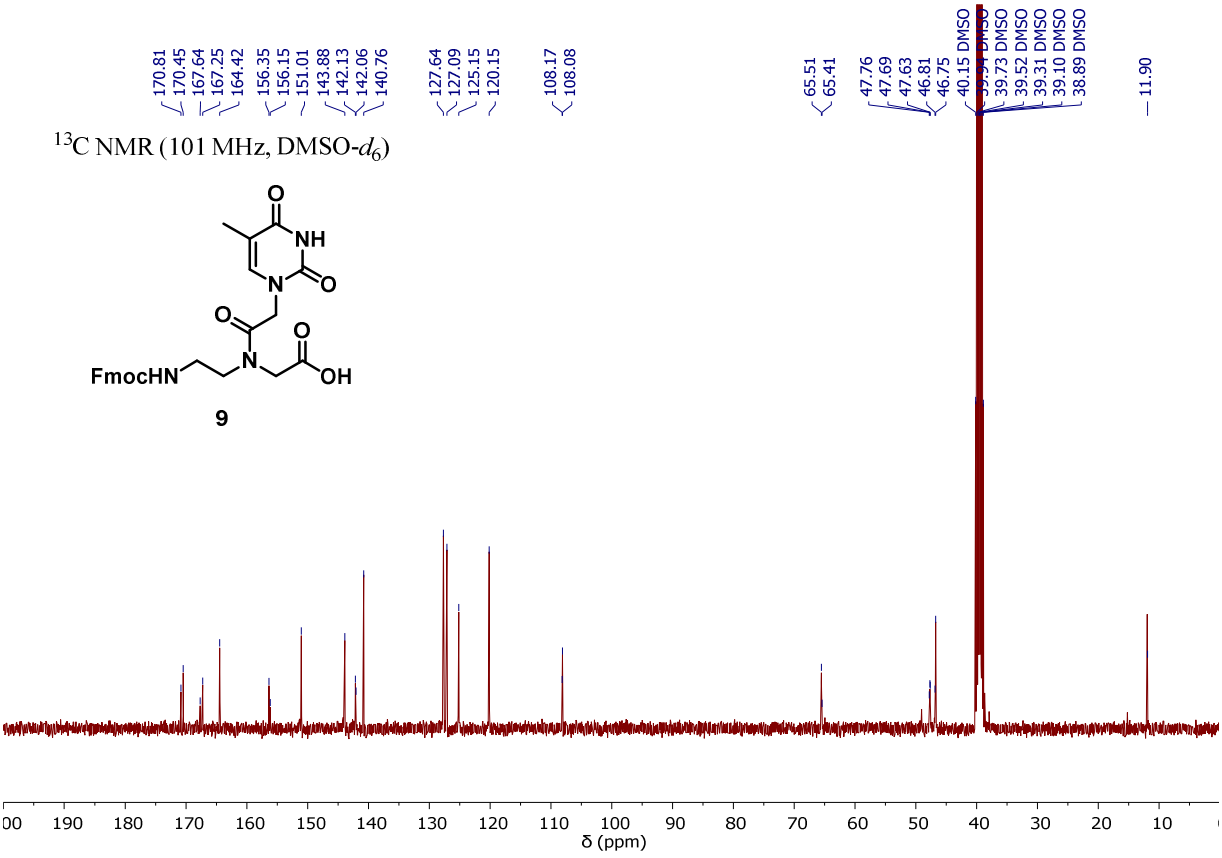
8

^{1}H NMR (400 MHz, DMSO-d_6)

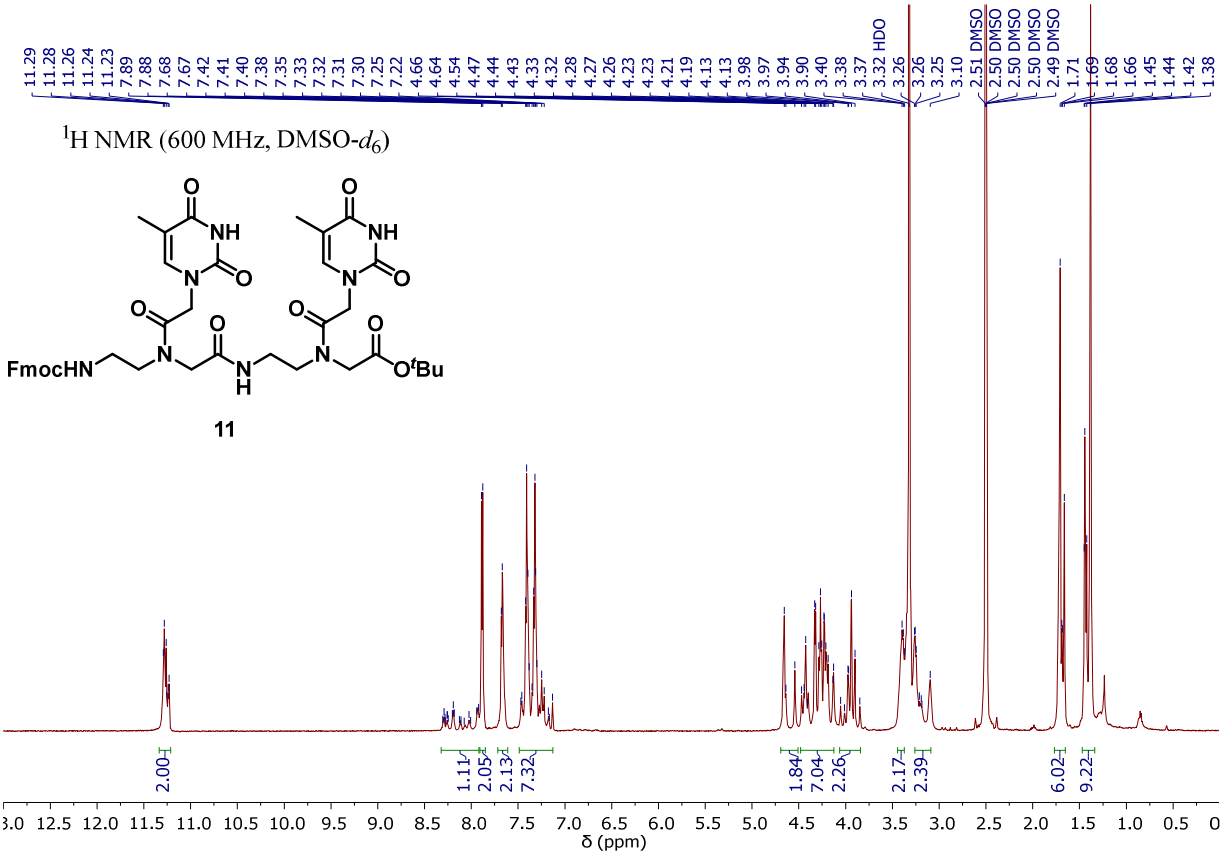


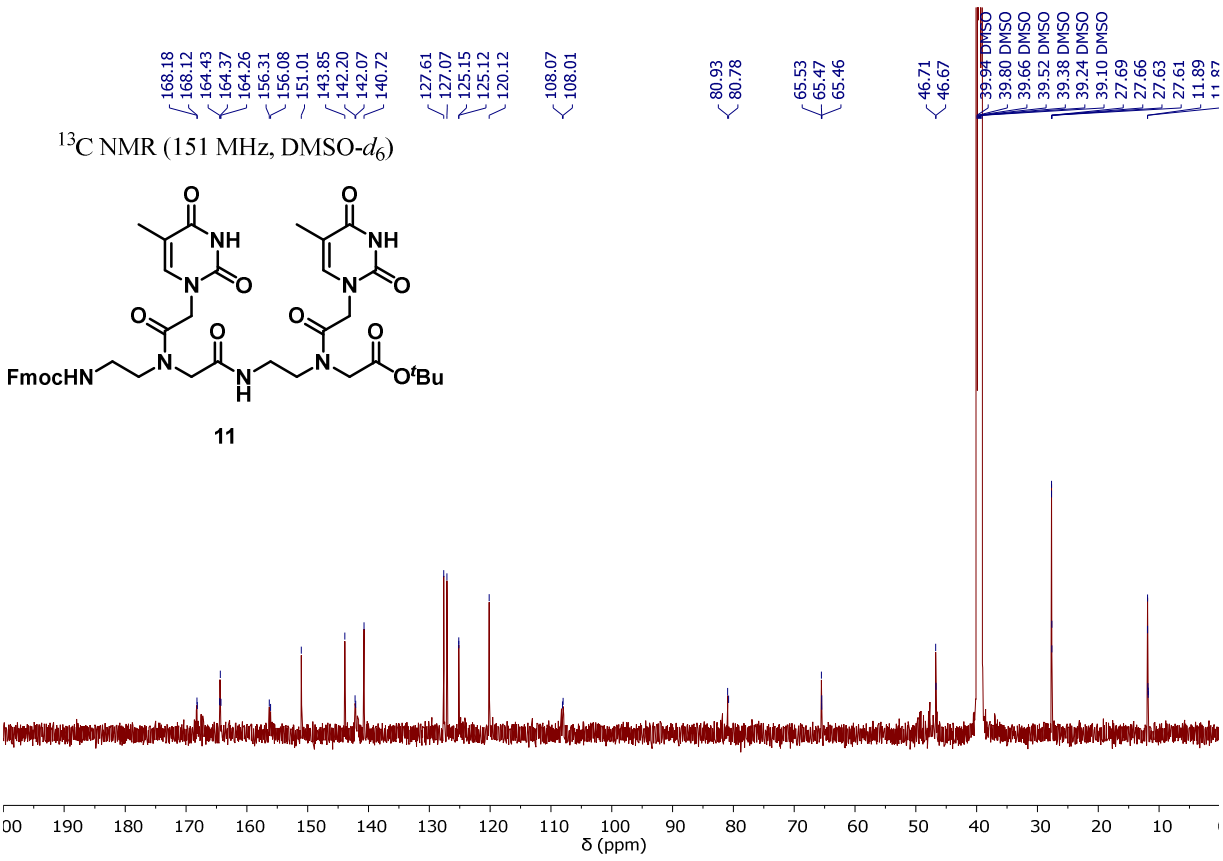
9

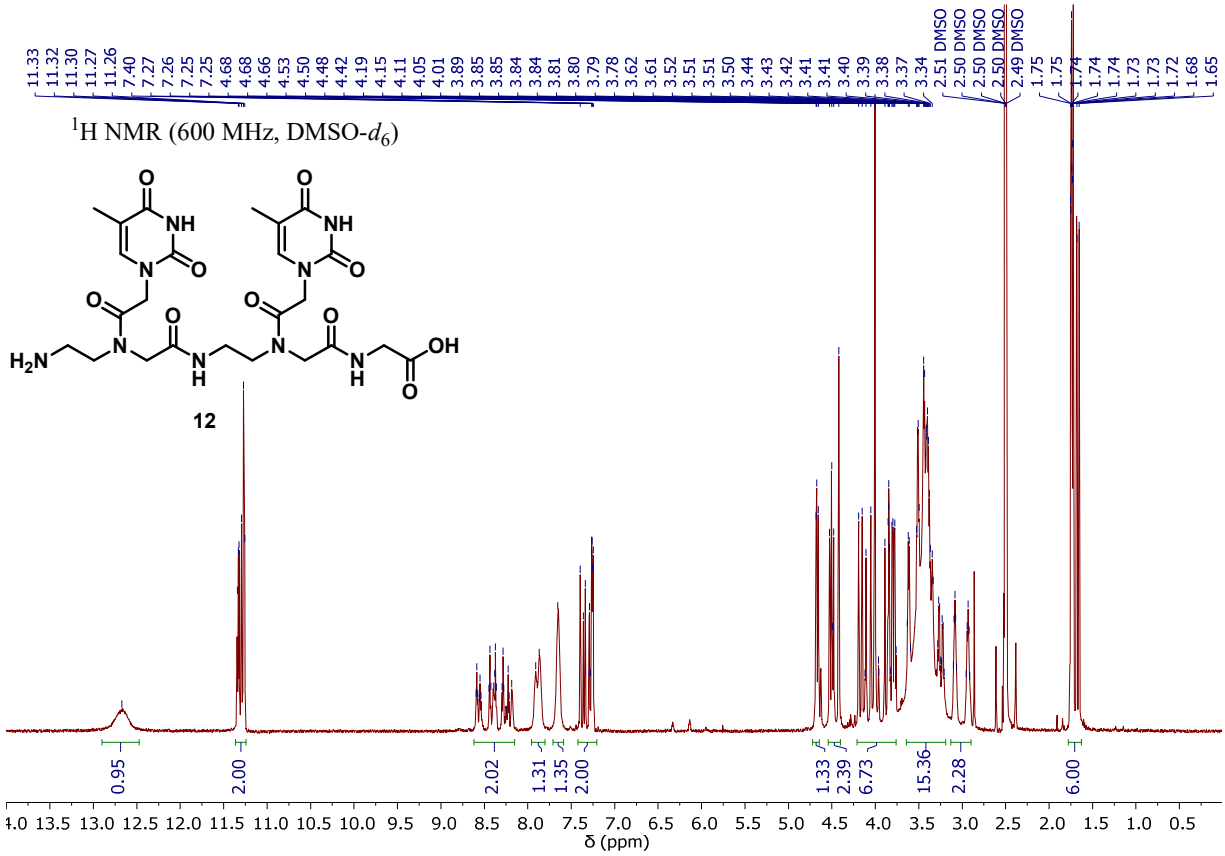
^{13}C NMR (101 MHz, DMSO-d_6)

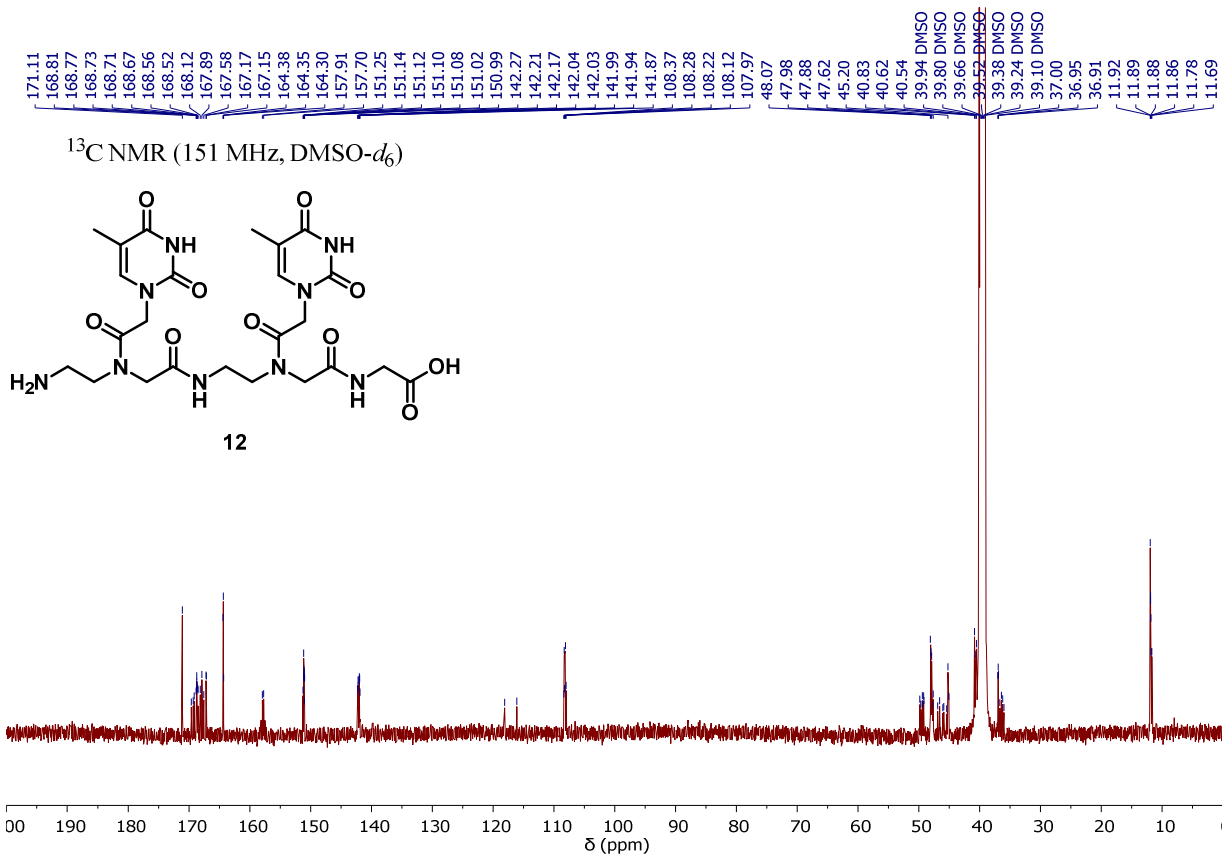


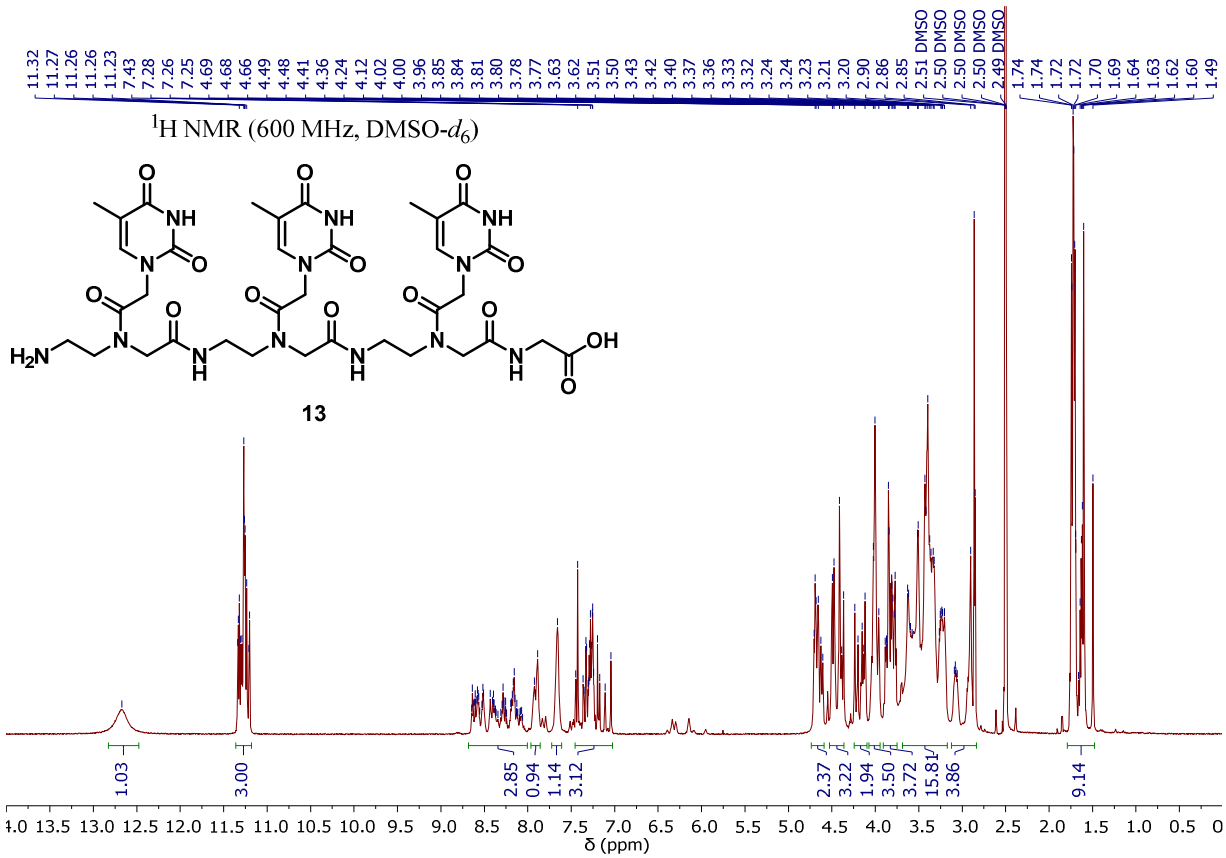
9

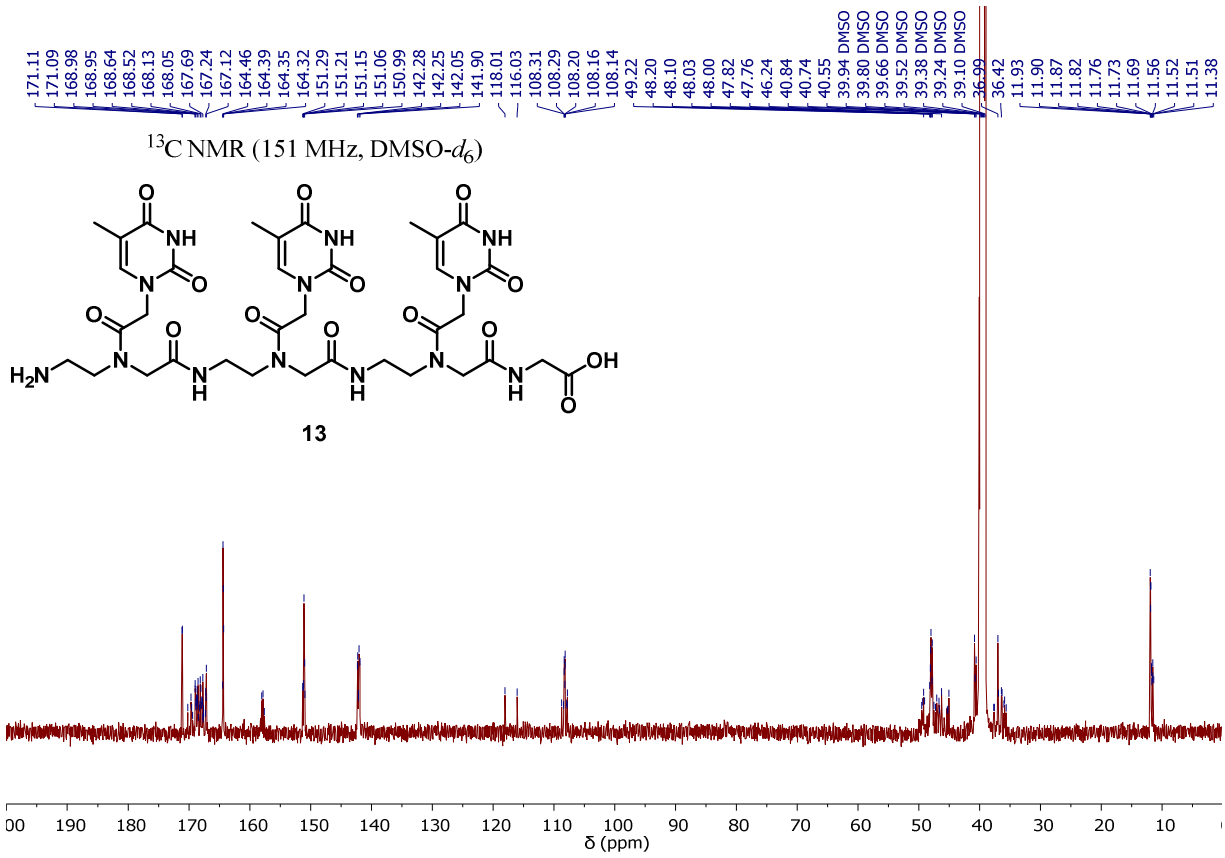

11.29 11.28 11.26 11.24 11.23 7.89 7.88 7.68 7.67 7.42 7.41 7.40 7.38 7.35 7.33 7.32 7.31 7.30 7.25 7.22 4.66 4.64 4.54 4.47 4.44 4.43 4.33 4.32 4.28 4.27 4.26 4.23 4.23 4.21 4.19 4.13 4.13 3.98 3.97 3.94 3.90 3.40 3.38 3.37 3.32 HDO 3.26 3.26 3.25 3.10 2.51 DMSO 2.50 DMSO 2.50 DMSO 2.50 DMSO 2.49 DMSO 1.71 1.69 1.68 1.66 1.45 1.44 1.42 1.38
^{1}H NMR (600 MHz, DMSO-d_6)
FmocHN
O^{t}Bu
11
2.00 1.11 2.05 2.13 7.32 1.84 7.04 2.26 2.17 2.39 6.02 9.22
δ (ppm)


168.18 168.12 164.43 164.37 164.26 156.31 156.08 151.01 143.85 142.20 142.07 140.72 127.61 127.07 125.15 125.12 120.12 108.07 108.01 80.93 80.78 65.53 65.47 65.46 46.71 46.67 39.94 DMSO 39.80 DMSO 39.66 DMSO 39.52 DMSO 39.38 DMSO 39.24 DMSO 39.10 DMSO 27.69 27.66 27.63 27.61 11.89
^{13}C NMR (151 MHz, DMSO-d_6)
FmocHN
O^{t}Bu
11
δ (ppm)

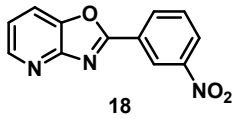

11.33 11.32 11.30 11.27 11.26 7.40 7.27 7.26 7.25 7.25 4.68 4.68 4.66 4.53 4.50 4.48 4.42 4.19 4.15 4.11 4.05 4.01 3.89 3.85 3.85 3.84 3.84 3.81 3.80 3.79 3.78 3.62 3.61 3.52 3.51 3.51 3.50 3.44 3.43 3.42 3.41 3.41 3.40 3.39 3.38 3.37 3.34 2.51 DMSO 2.50 DMSO 2.50 DMSO 2.50 DMSO 2.49 DMSO 1.75 1.75 1.74 1.74 1.74 1.73 1.73 1.72 1.68 1.65
1H NMR (600 MHz, DMSO-d6)
12
0.95 2.00 2.02 1.31 1.35 2.00 1.33 2.39 6.73 15.36 2.28 6.00
δ (ppm)


171.11 168.81 168.77 168.73 168.71 168.67 168.56 168.52 168.12 167.89 167.58 167.17 167.15 164.38 164.35 164.30 157.91 157.70 151.25 151.14 151.12 151.10 151.08 151.02 150.99 142.27 142.21 142.17 142.04 142.03 141.99 141.94 141.87 108.37 108.28 108.22 108.12 107.97 48.07 47.98 47.88 47.62 45.20 40.83 40.62 40.54 39.94 DMSO 39.80 DMSO 39.66 DMSO 39.52 DMSO 39.38 DMSO 39.24 DMSO 39.10 DMSO 37.00 36.95 36.91 11.92 11.89 11.88 11.86 11.78 11.69
13C NMR (151 MHz, DMSO-d6)
12
δ (ppm)

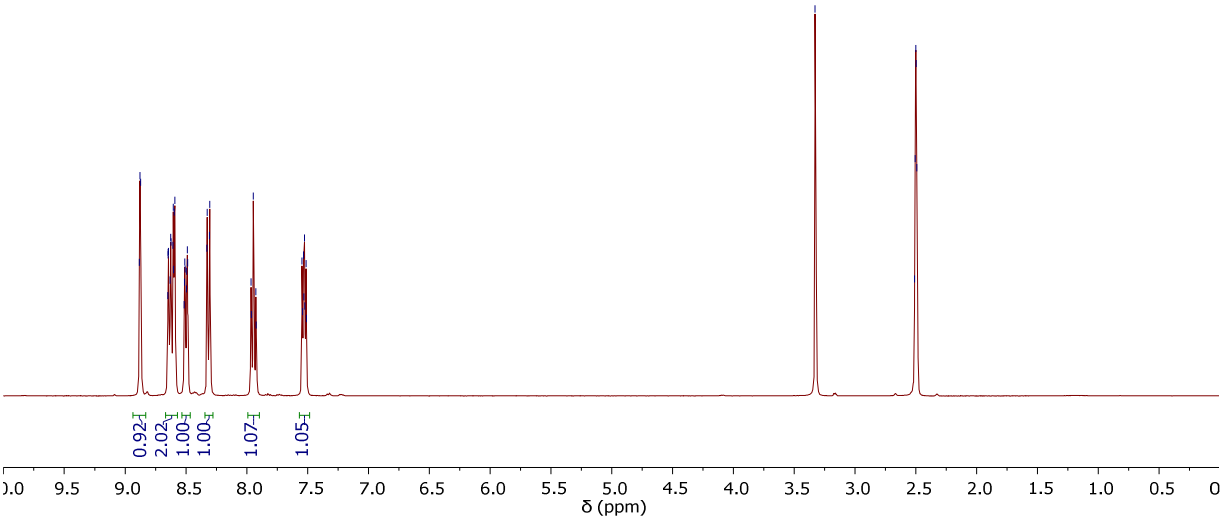

11.32 11.27 11.26 11.26 11.23 7.43 7.28 7.26 7.25 4.69 4.68 4.66 4.49 4.48 4.41 4.36 4.24 4.12 4.02 4.00 3.96 3.85 3.84 3.81 3.80 3.78 3.77 3.63 3.62 3.51 3.50 3.43 3.42 3.40 3.37 3.36 3.33 3.32 3.24 3.24 3.23 3.21 3.20 2.90 2.86 2.85 2.51 DMSO 2.50 DMSO 2.50 DMSO 2.50 DMSO 2.50 DMSO 2.49 DMSO 1.74 1.74 1.72 1.72 1.70 1.69 1.64 1.63 1.62 1.60 1.49
1H NMR (600 MHz, DMSO-d6)
13
1.03 3.00 2.85 0.94 1.14 3.12 2.37 3.22 1.94 3.50 3.72 15.81 3.86 9.14
δ (ppm)


171.11 171.09 168.98 168.95 168.64 168.52 168.13 168.05 167.69 167.24 167.12 164.46 164.39 164.35 164.32 151.29 151.21 151.15 151.06 150.99 142.28 142.25 142.05 141.90 118.01 116.03 108.31 108.29 108.20 108.16 108.14 49.22 48.20 48.10 48.03 48.00 47.82 47.76 46.24 40.84 40.74 40.55 39.94 DMSO 39.80 DMSO 39.66 DMSO 39.52 DMSO 39.38 DMSO 39.24 DMSO 39.10 DMSO 36.99 36.42 11.93 11.90 11.87 11.82 11.76 11.73 11.69 11.56 11.52 11.51 11.38
13C NMR (151 MHz, DMSO-d6)
13
δ (ppm)

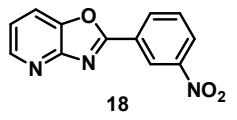
^{1}H NMR (400 MHz, DMSO-d_6)



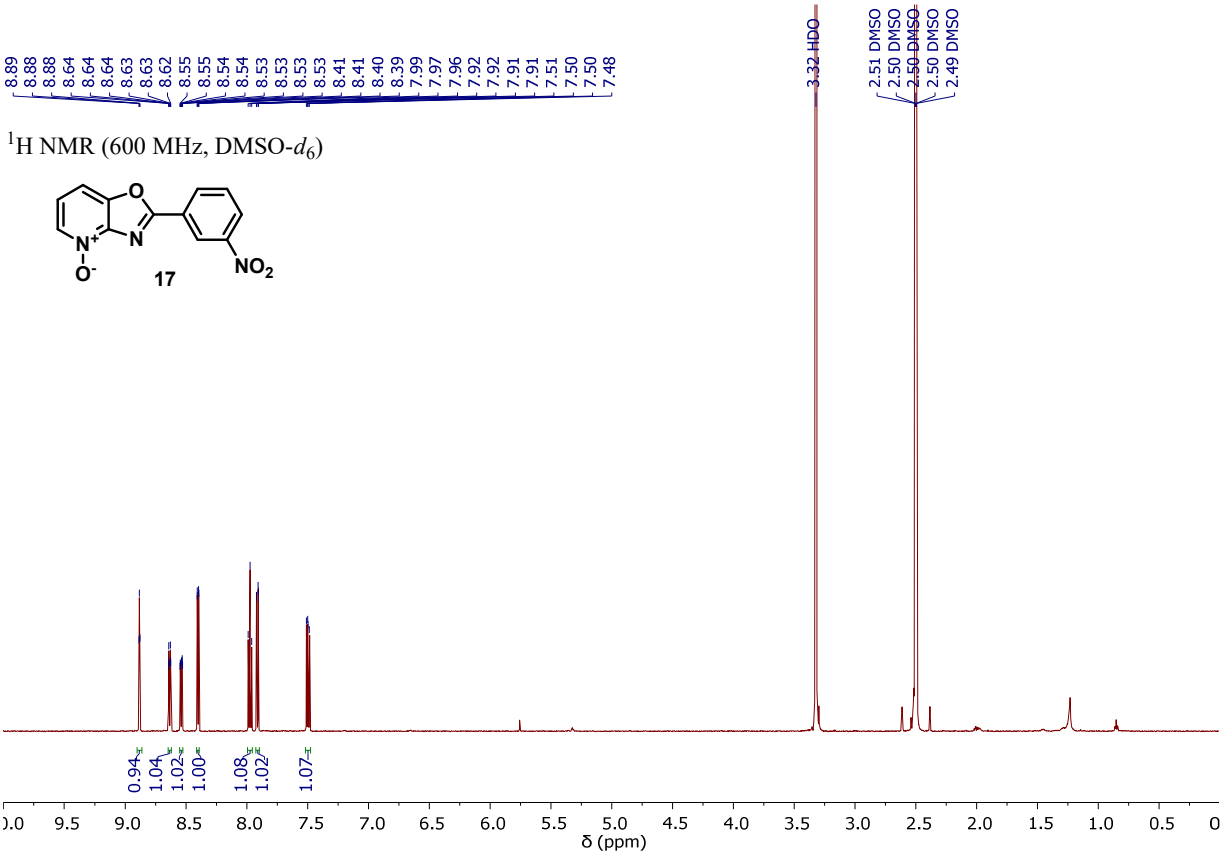
18



^{13}C NMR (101 MHz, DMSO-d_6)

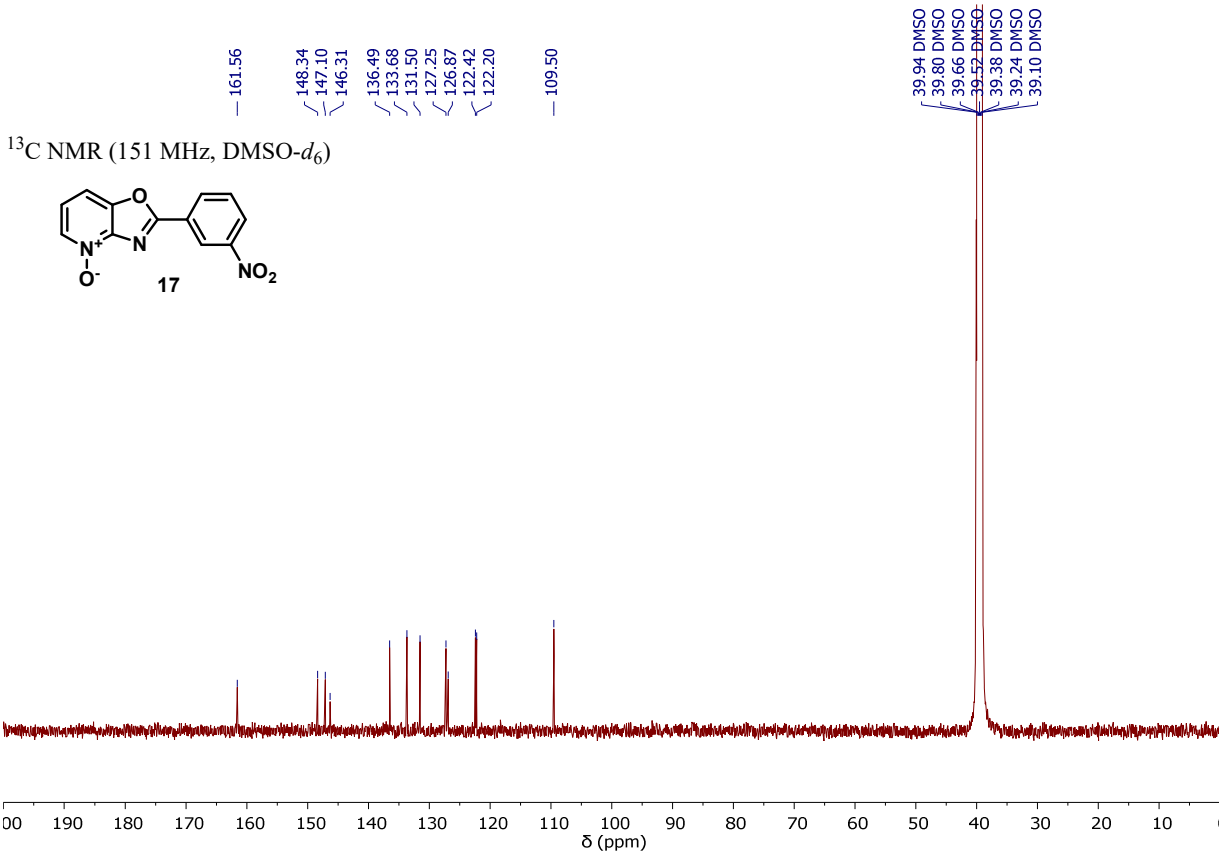


18



^{1}H NMR (600 MHz, DMSO-d_6)

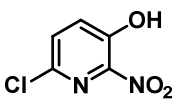
17



^{13}C NMR (151 MHz, DMSO-d_6)

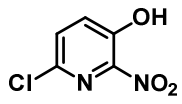
17

^{1}H NMR (400 MHz, DMSO-d_6)



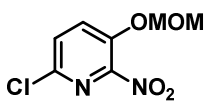
22

^{13}C NMR (101 MHz, DMSO-d_6)



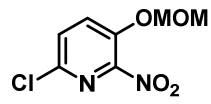
22

^{1}H NMR (400 MHz, DMSO-d_6)

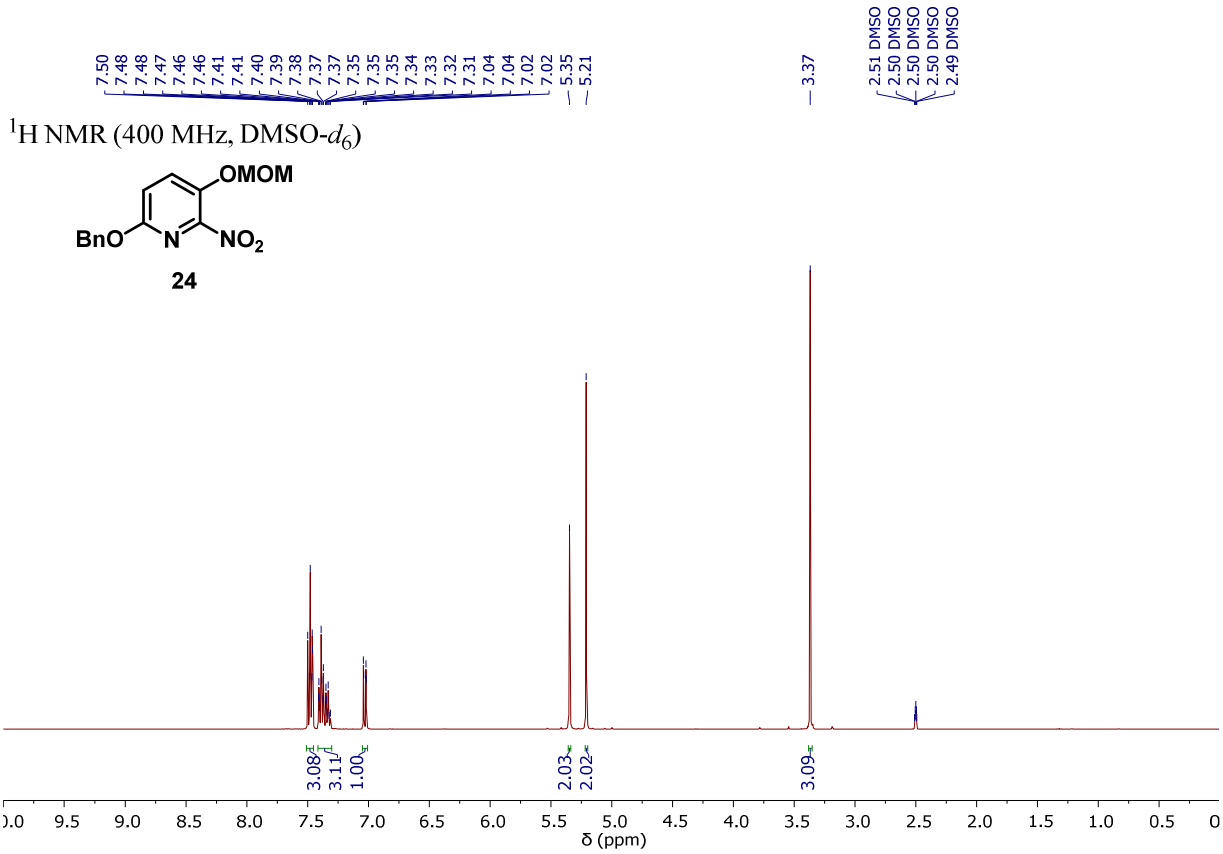


23

^{13}C NMR (101 MHz, DMSO-d_6)

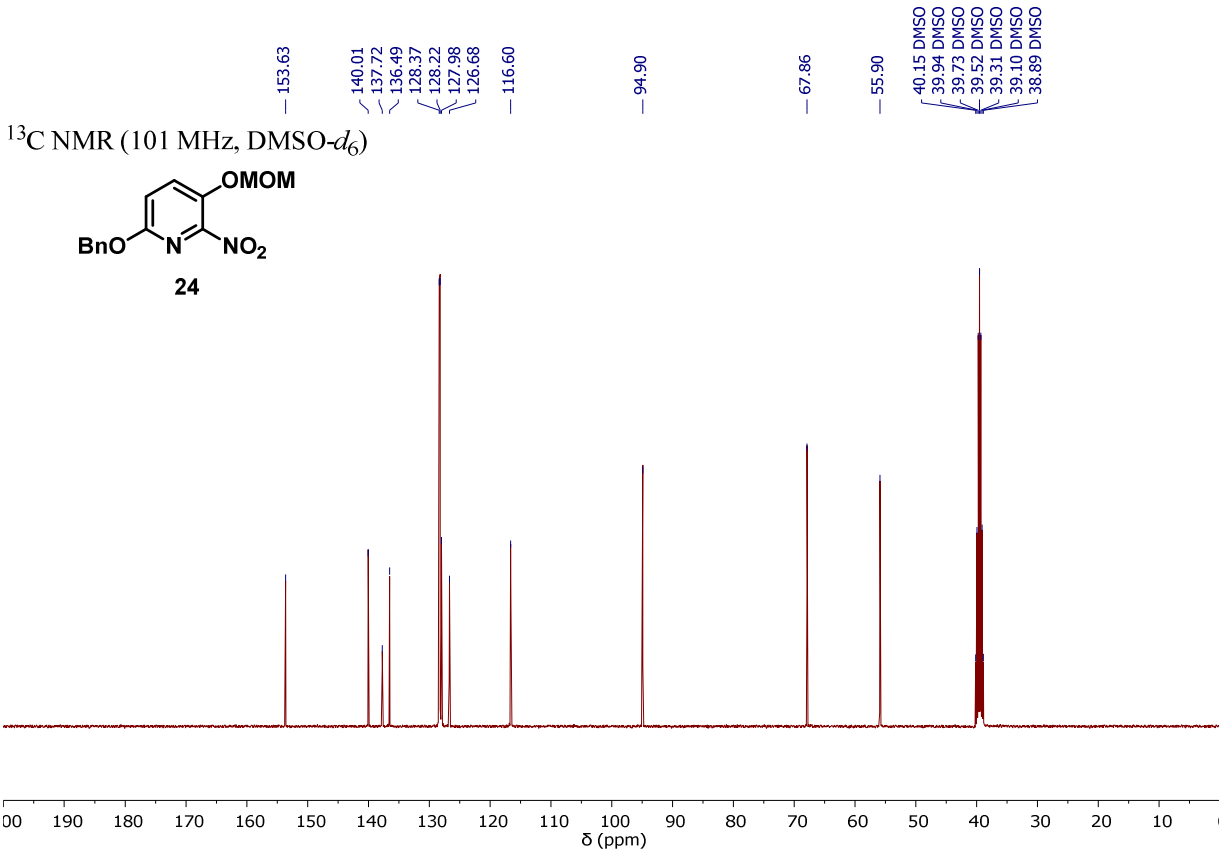


23



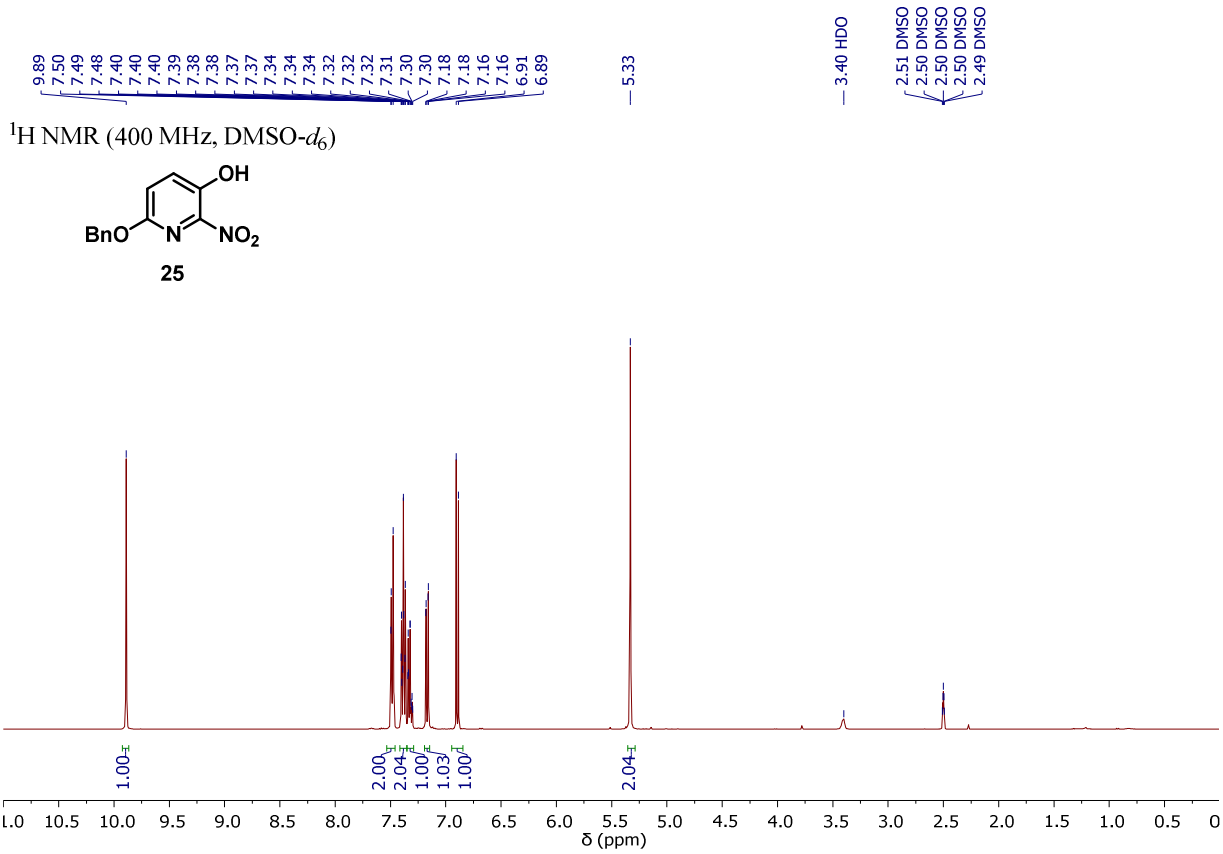
^{1}H NMR (400 MHz, DMSO-d_6)

24



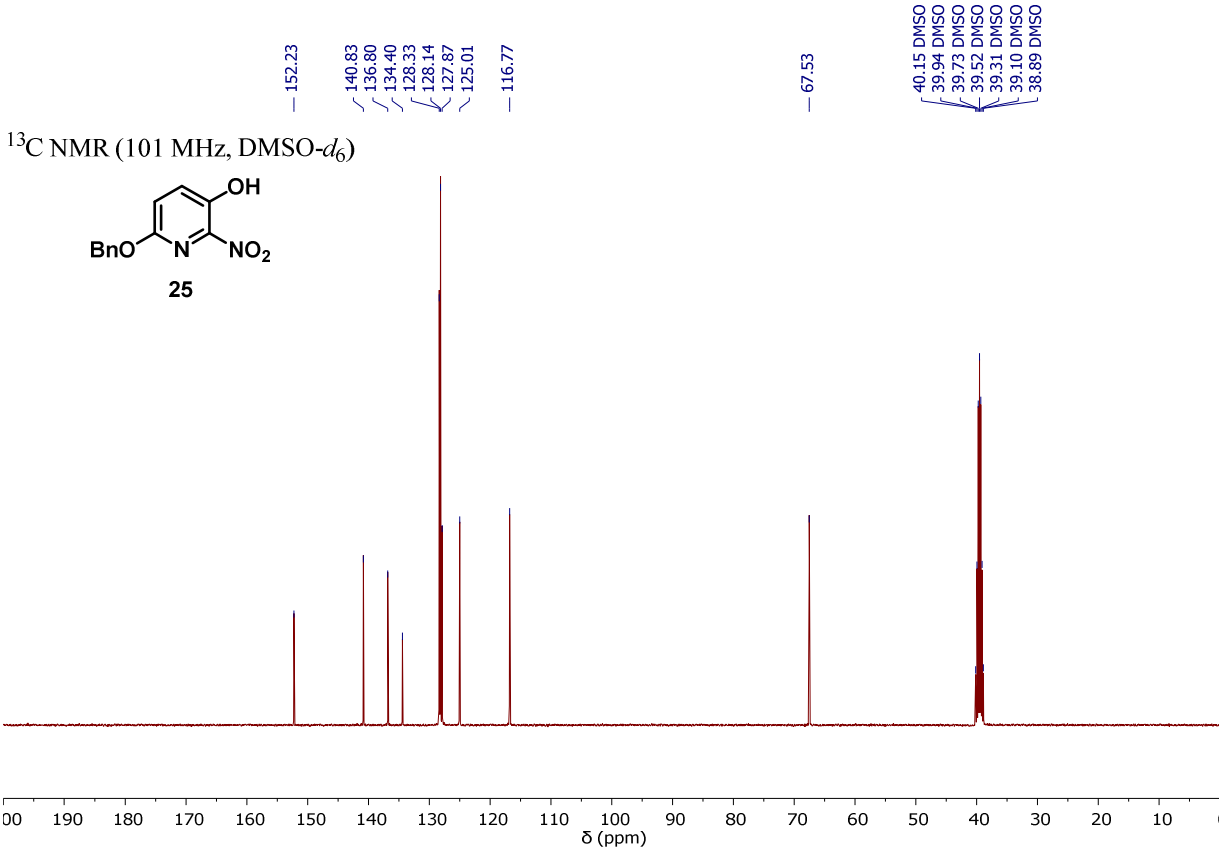
^{13}C NMR (101 MHz, DMSO-d_6)

24



^{1}H NMR (400 MHz, DMSO-d_6)

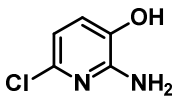
25



^{13}C NMR (101 MHz, DMSO-d_6)

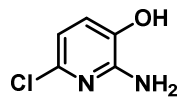
25

^{1}H NMR (400 MHz, DMSO-d_6)



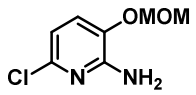
27

^{13}C NMR (101 MHz, DMSO-d_6)



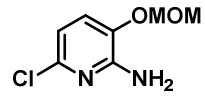
27

^{1}H NMR (400 MHz, DMSO-d_6)

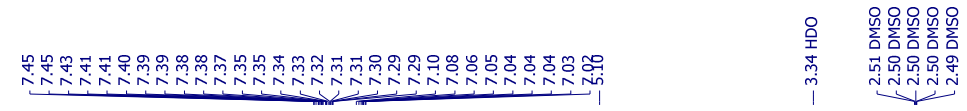


28

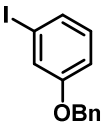
^{13}C NMR (101 MHz, DMSO-d_6)



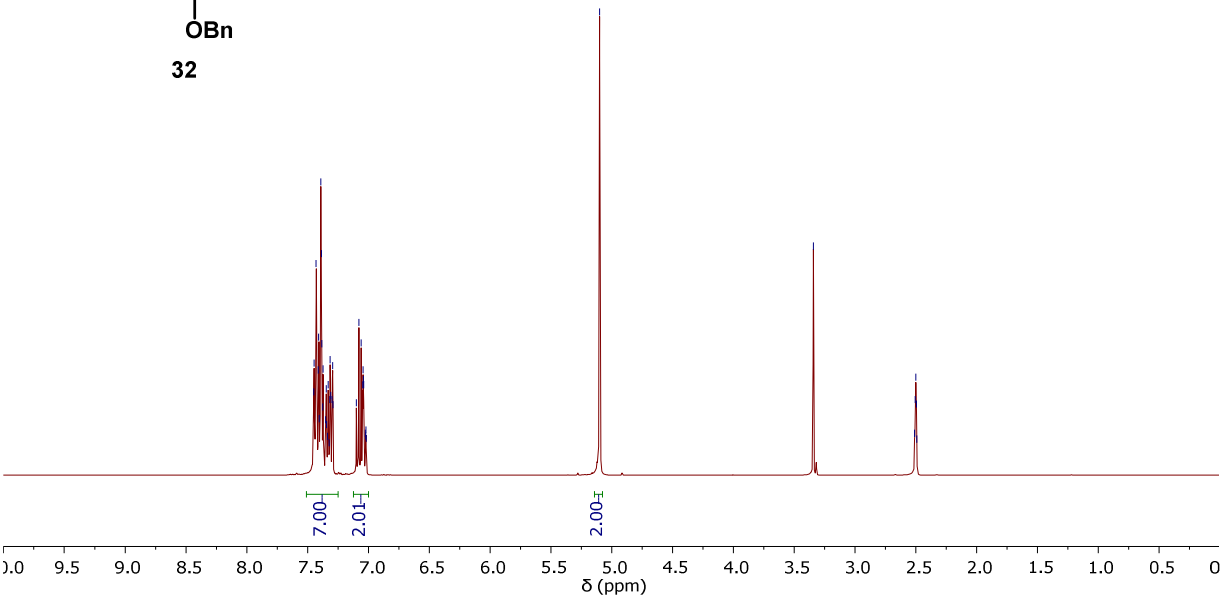
28

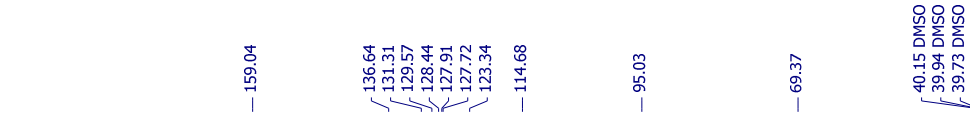


^{1}H NMR (400 MHz, DMSO-d_6)

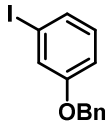


32

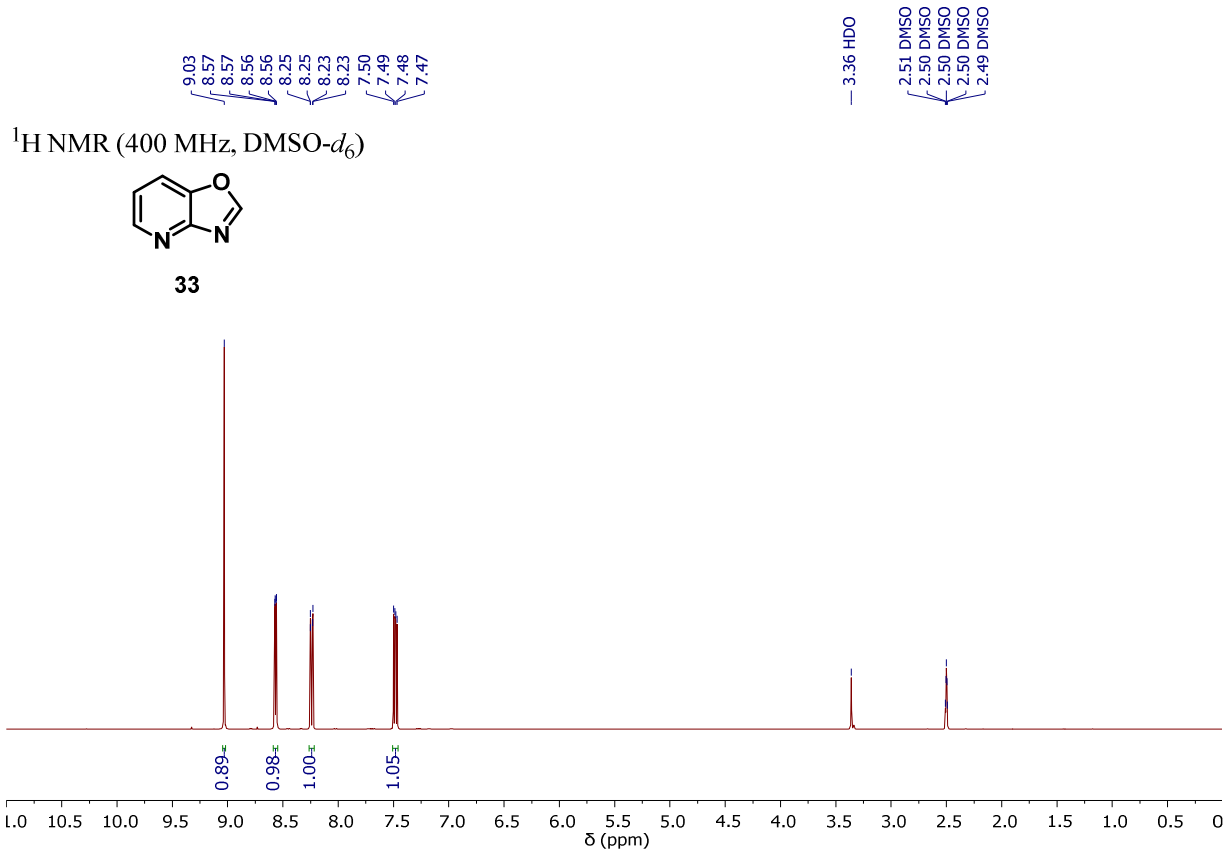




^{13}C NMR (101 MHz, DMSO-d_6)

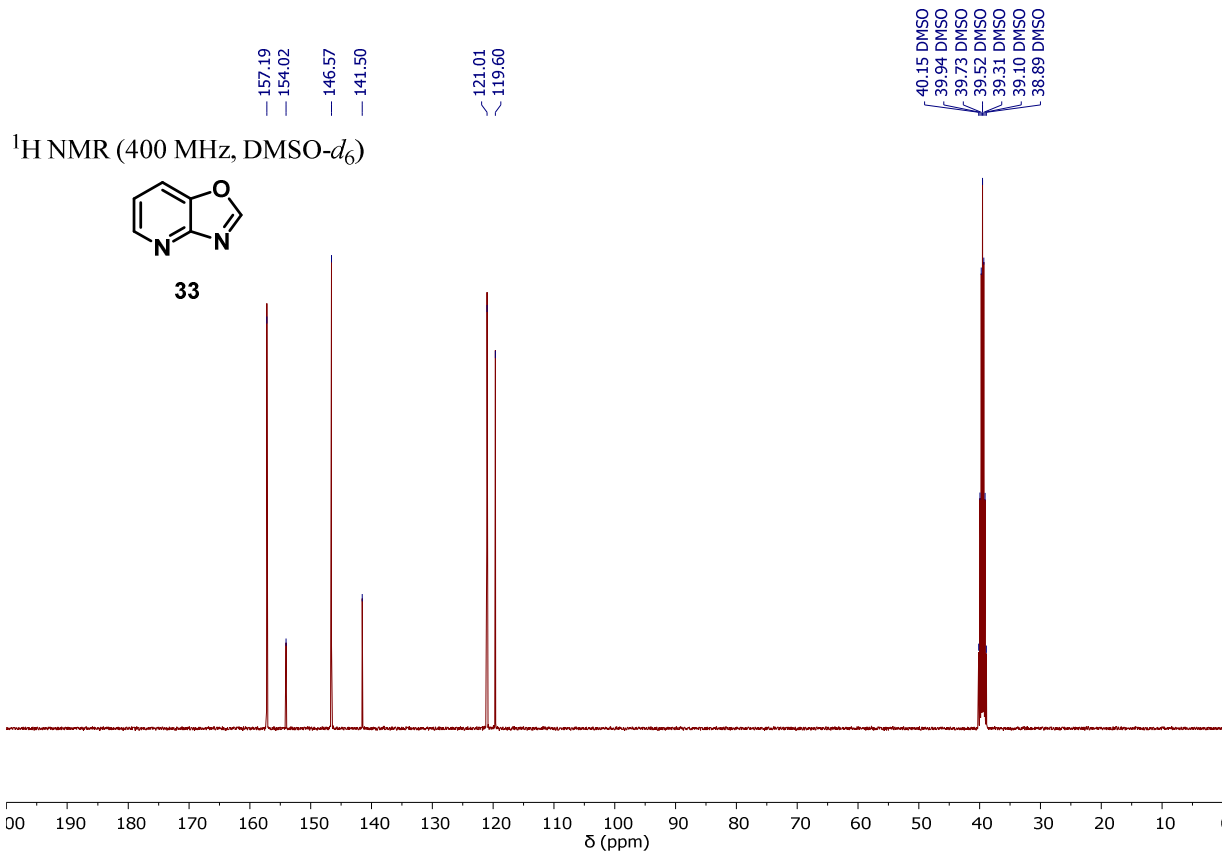


32



^{1}H NMR (400 MHz, DMSO-d_6)

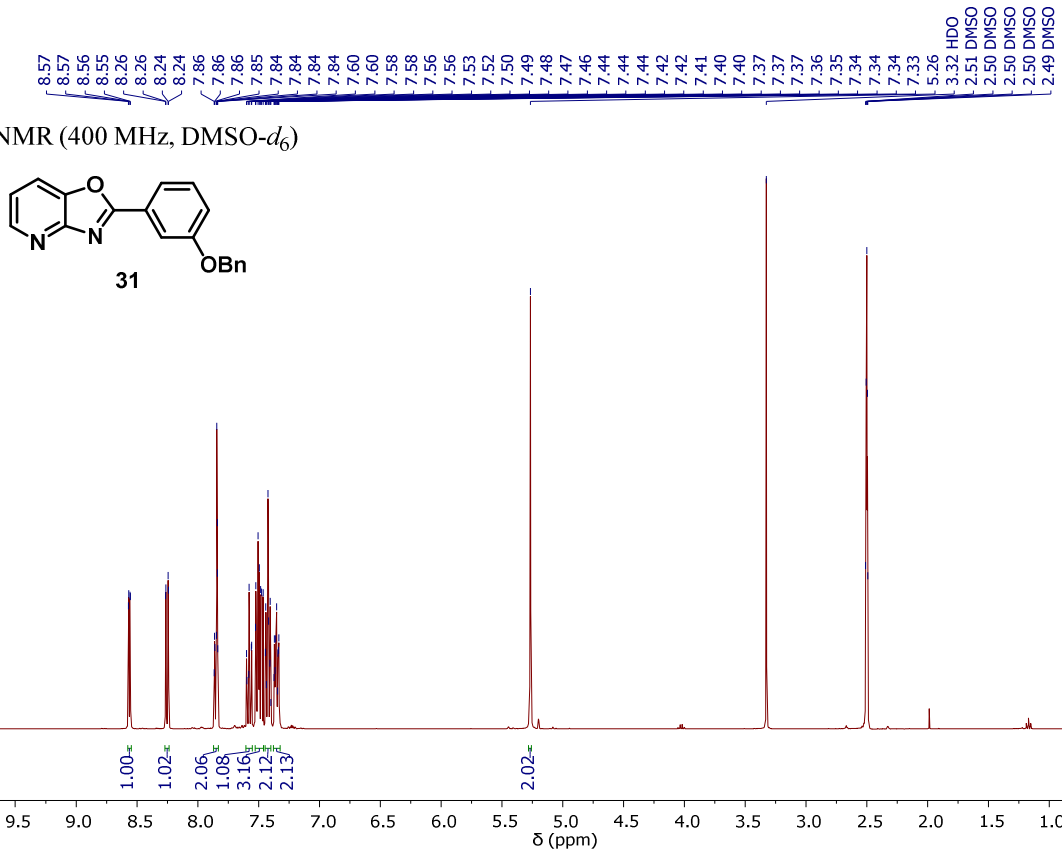
33



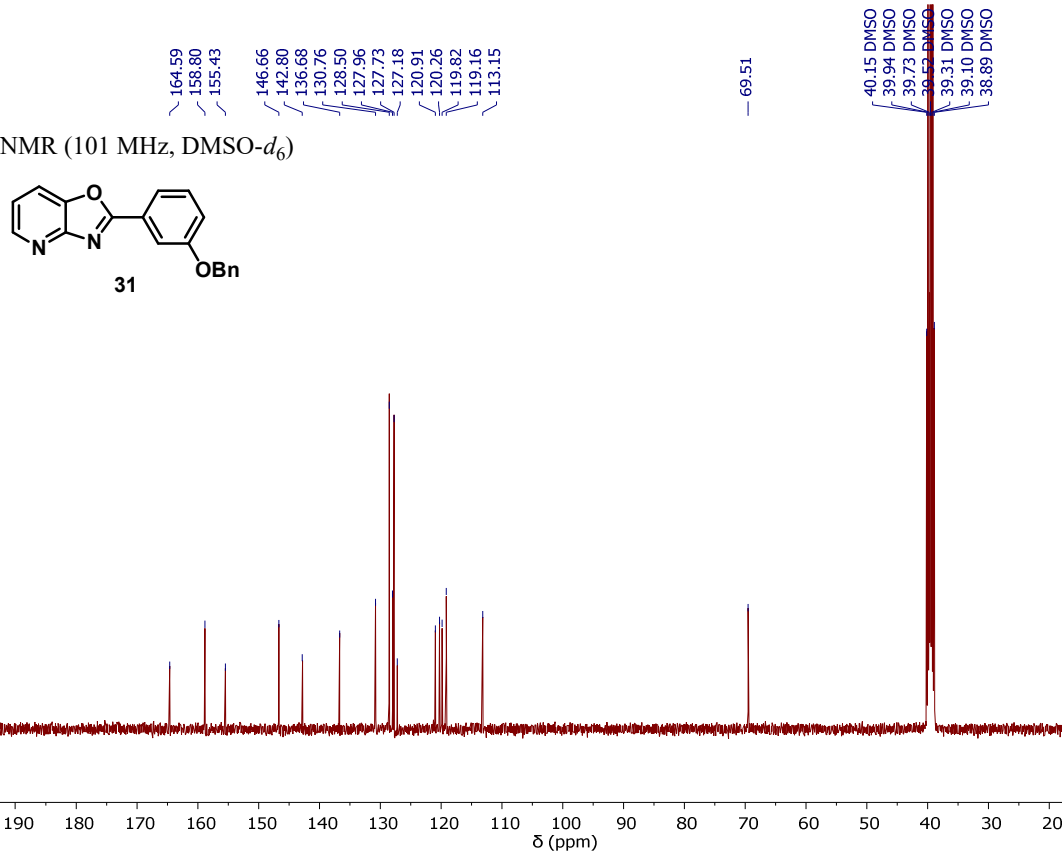
^{1}H NMR (400 MHz, DMSO-d_6)

33

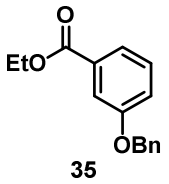
^{1}H NMR (400 MHz, DMSO-d_6)



^{13}C NMR (101 MHz, DMSO-d_6)

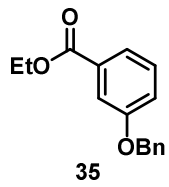


^{1}H NMR (400 MHz, DMSO-d_6)



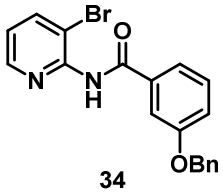
35

^{13}C NMR (101 MHz, DMSO-d_6)



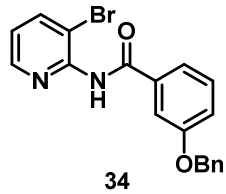
35

^{1}H NMR (400 MHz, DMSO-d_6)



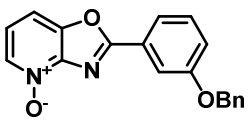
34

^{13}C NMR (101 MHz, DMSO-d_6)



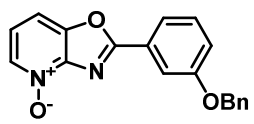
34

^{1}H NMR (400 MHz, DMSO-d_6)



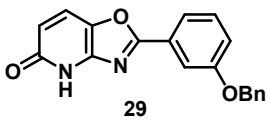
30

^{13}C NMR (101 MHz, DMSO-d_6)

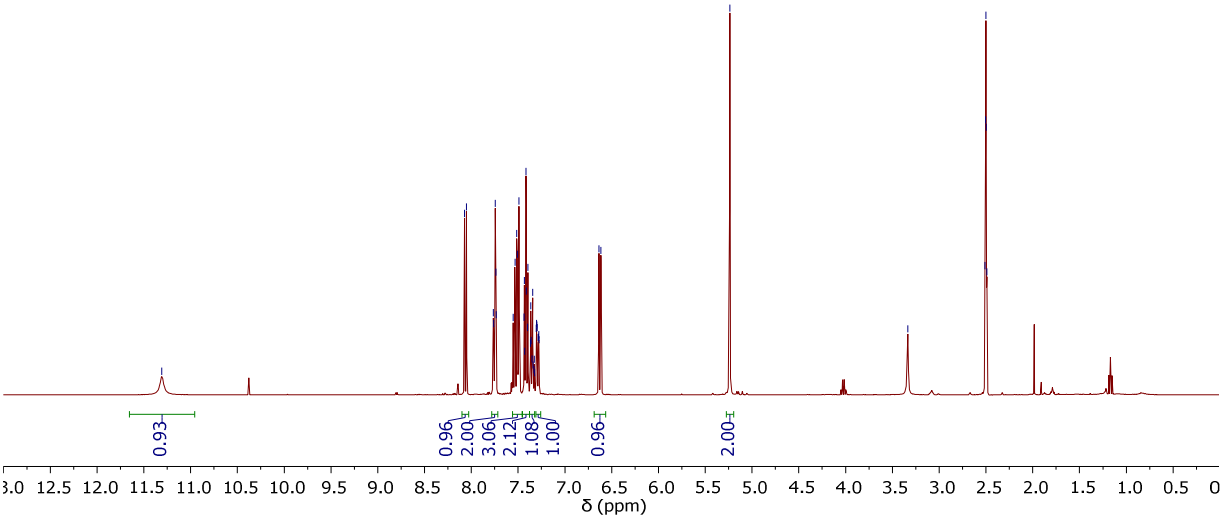


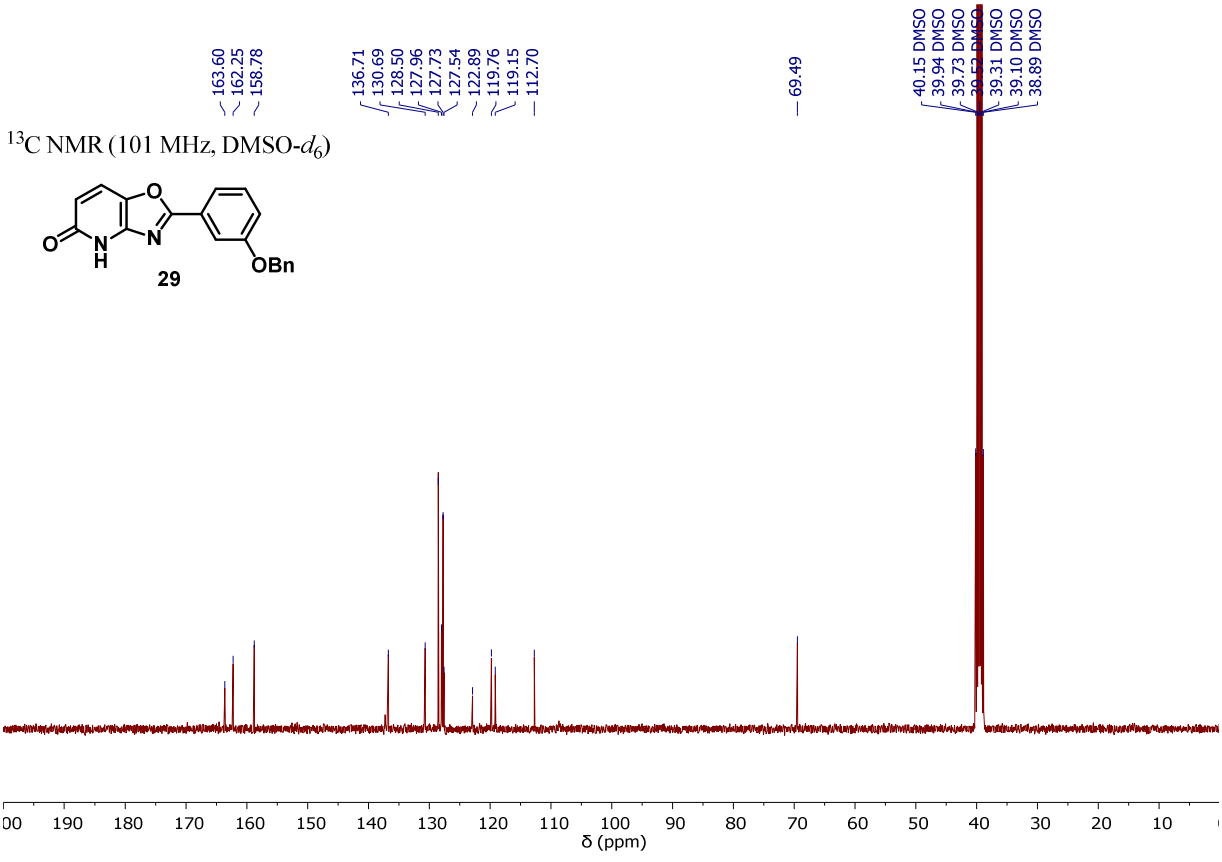
30

^{1}H NMR (400 MHz, DMSO-d_6)



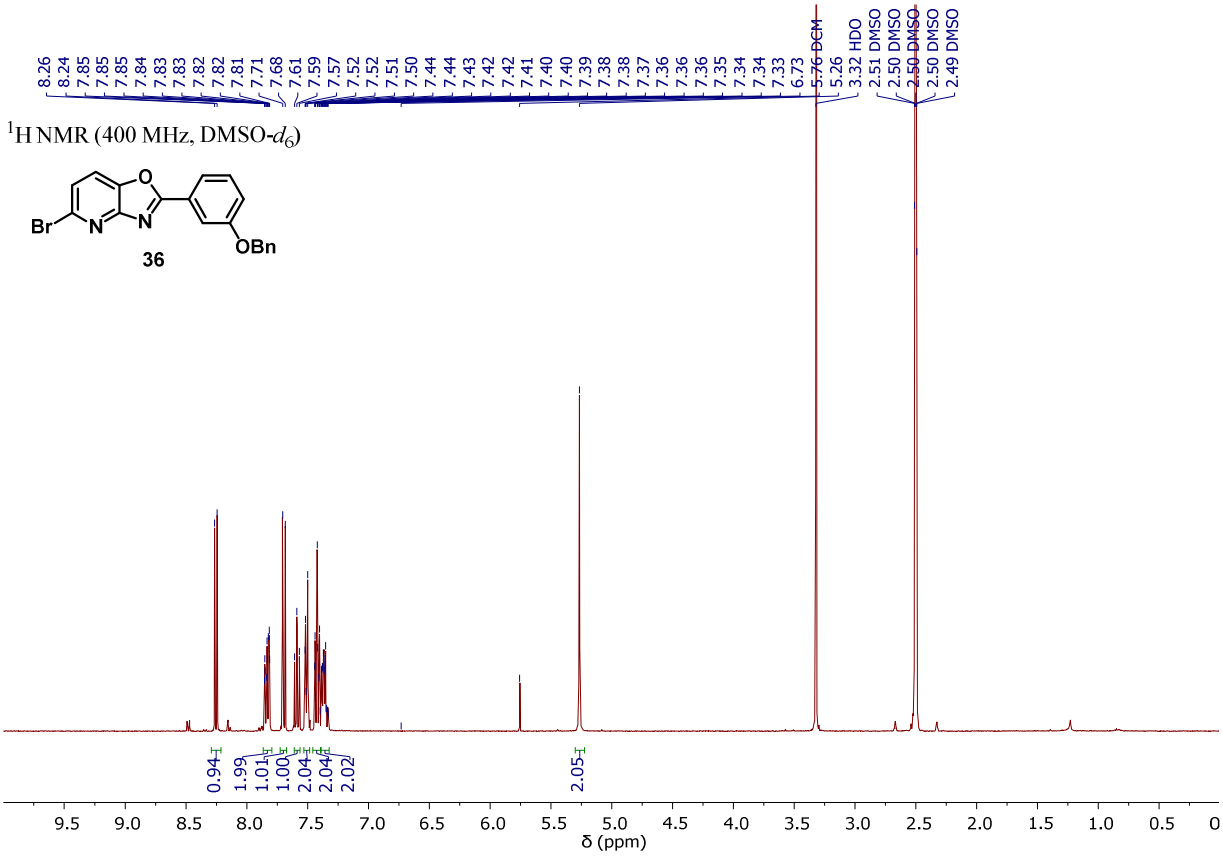
29





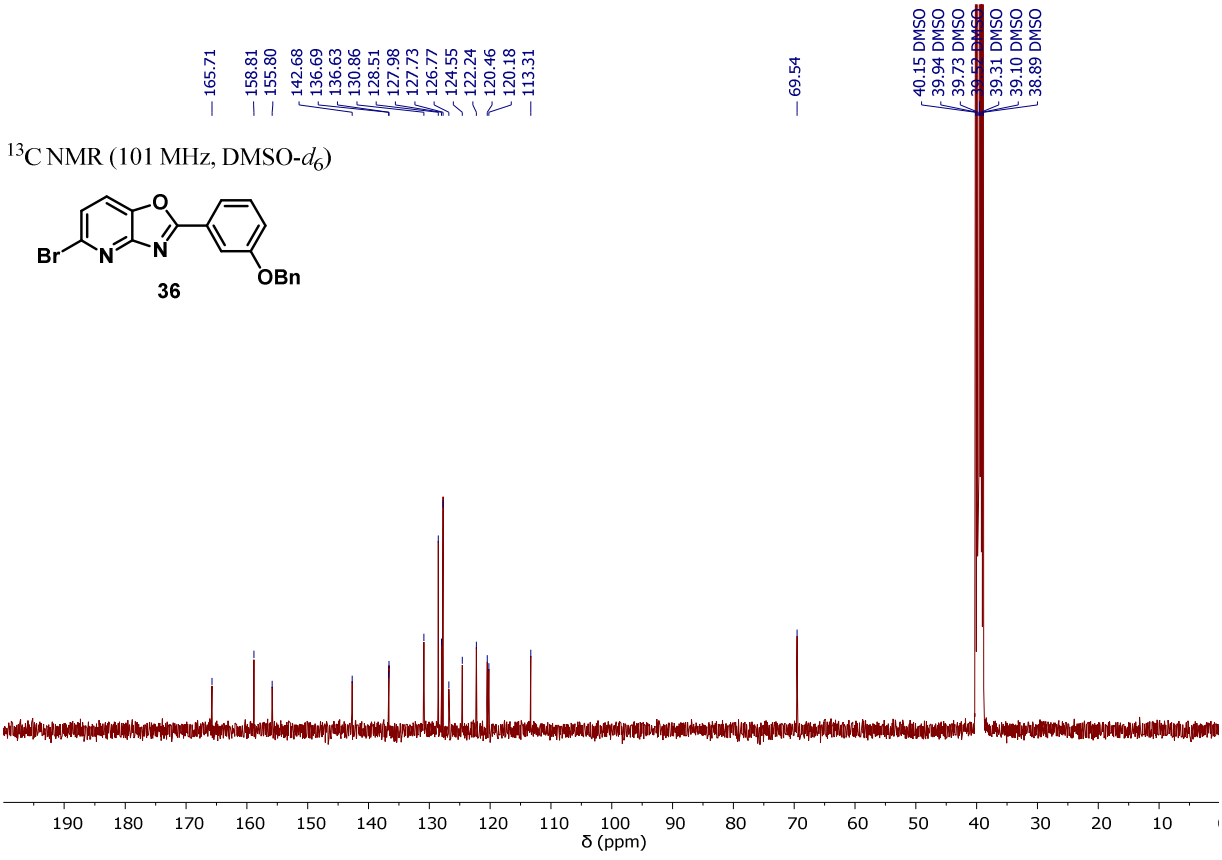
^{13}C NMR (101 MHz, DMSO-d_6)

29



^{1}H NMR (400 MHz, DMSO-d_6)

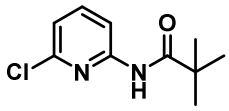
36



^{13}C NMR (101 MHz, DMSO-d_6)

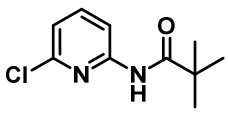
36

^{1}H NMR (400 MHz, DMSO-d_6)



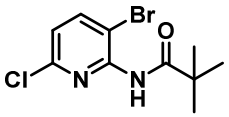
40

^{13}C NMR (101 MHz, DMSO-d_6)



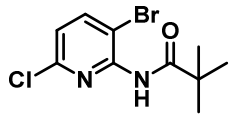
40

^{1}H NMR (400 MHz, DMSO-d_6)



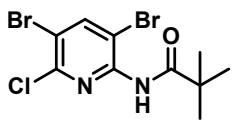
41

^{13}C NMR (101 MHz, DMSO-d_6)



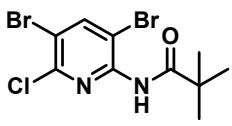
41

^{1}H NMR (400 MHz, DMSO-d_6)



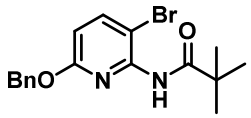
42

^{13}C NMR (101 MHz, DMSO-d_6)



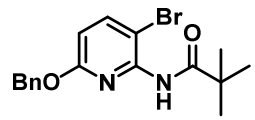
42

^{1}H NMR (400 MHz, DMSO-d_6)

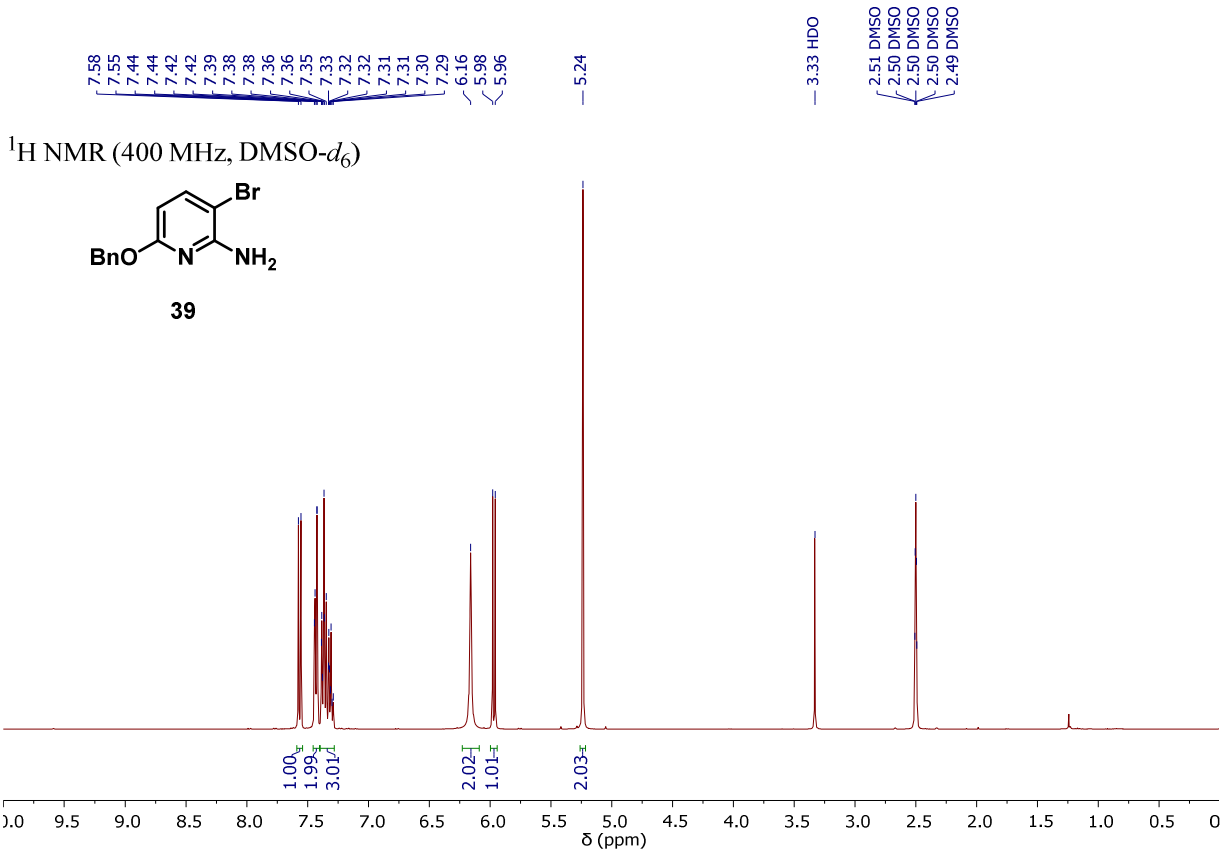


43

^{13}C NMR (101 MHz, DMSO-d_6)

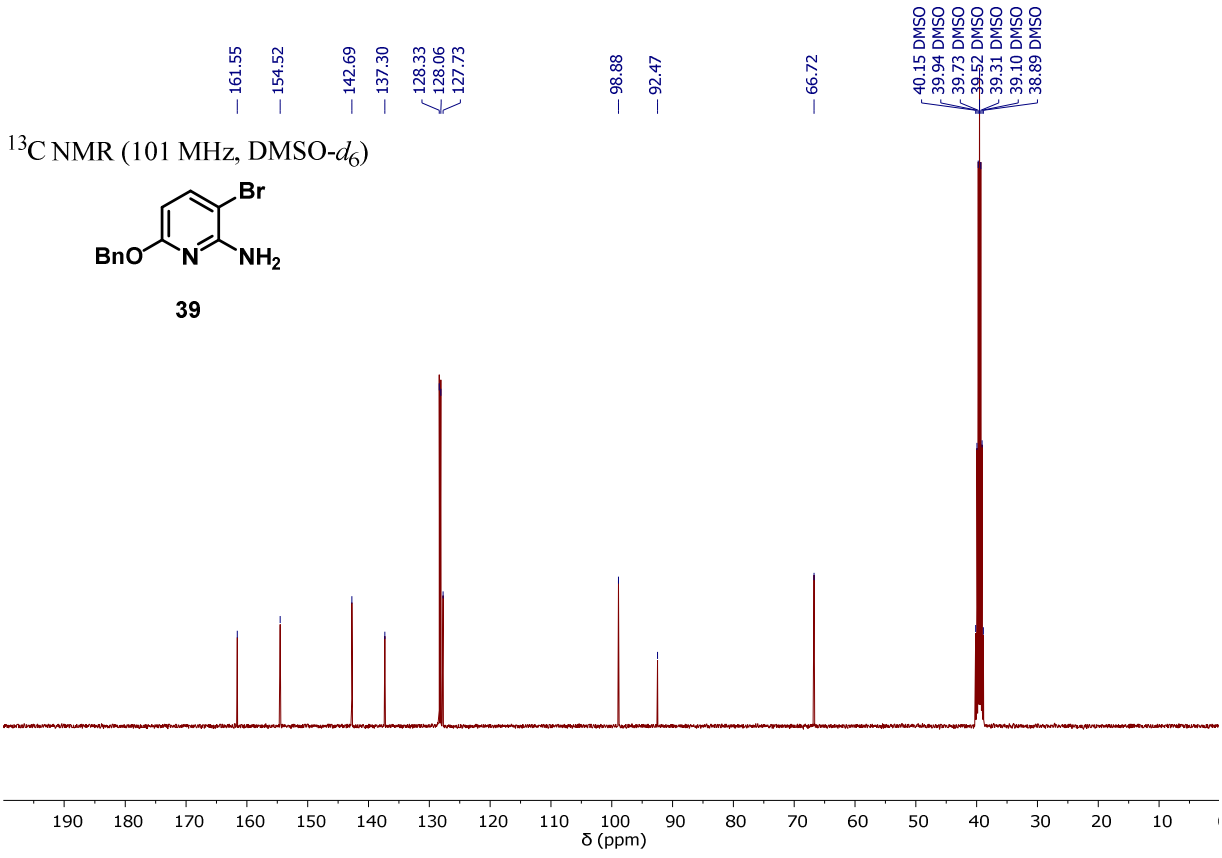


43



^{1}H NMR (400 MHz, DMSO-d_6)

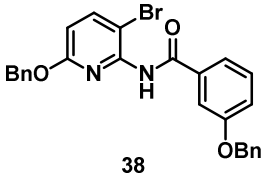
39



^{13}C NMR (101 MHz, DMSO-d_6)

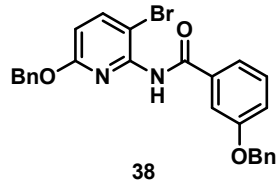
39

^{1}H NMR (400 MHz, DMSO-d_6)

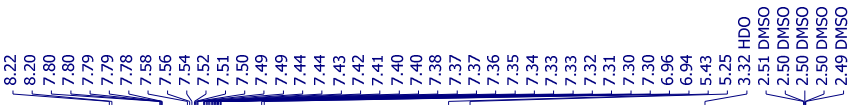


38

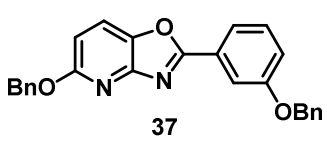
^{13}C NMR (101 MHz, DMSO-d_6)



38

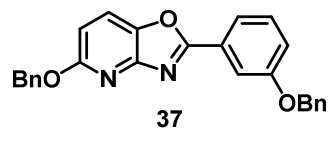


^{1}H NMR (400 MHz, DMSO-d_6)

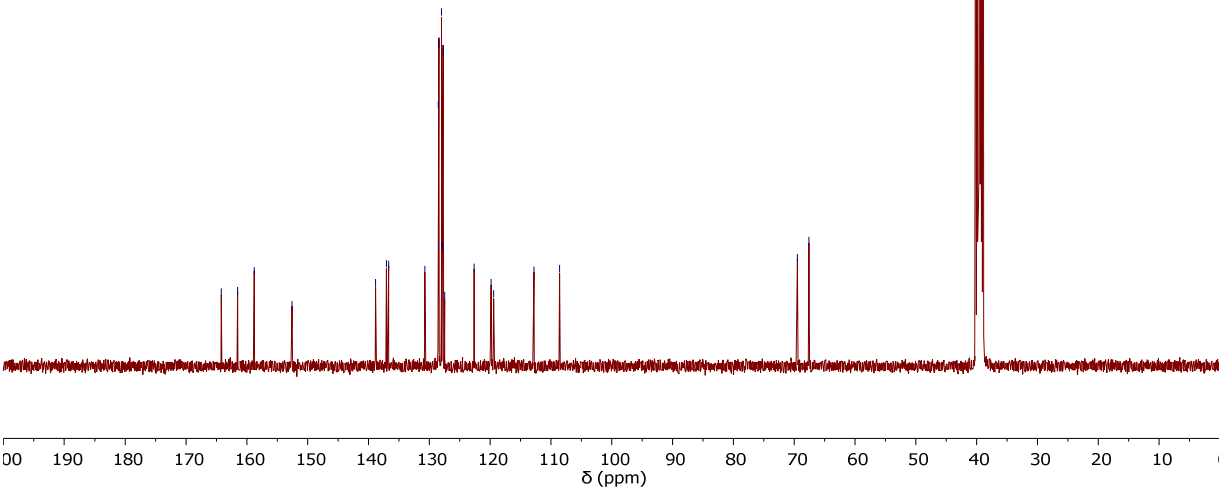


37

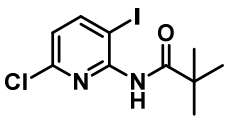
^{13}C NMR (101 MHz, DMSO-d_6)



37

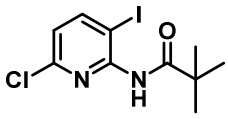


^{1}H NMR (400 MHz, DMSO-d_6)

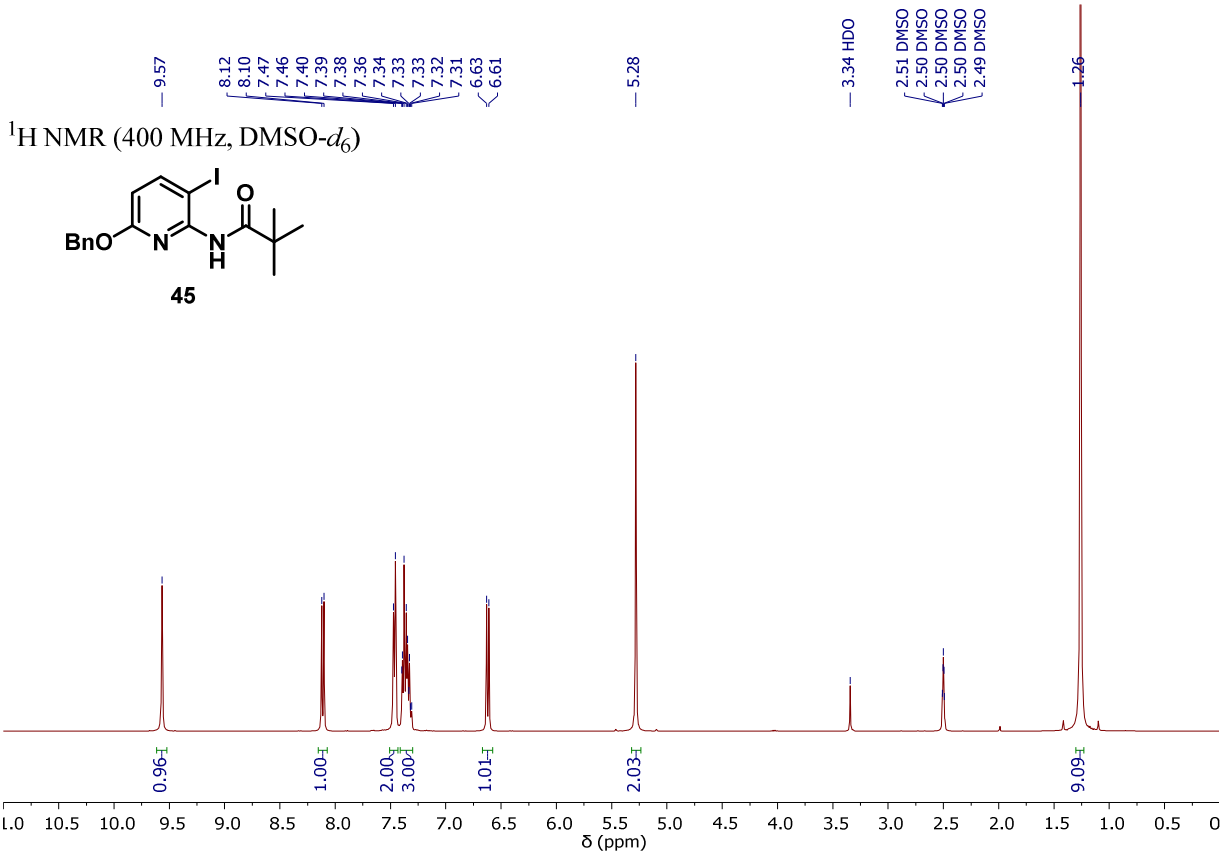


44

^{13}C NMR (101 MHz, DMSO-d_6)

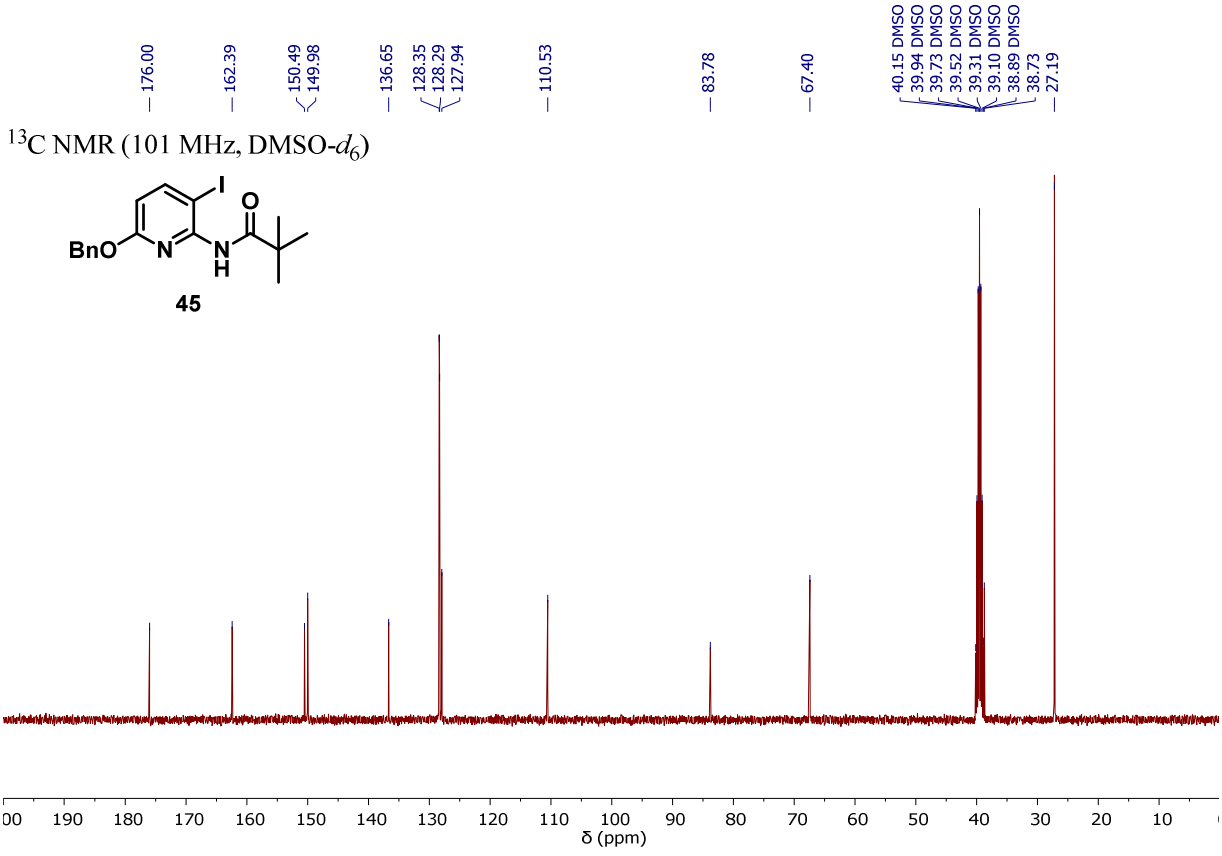


44



^{1}H NMR (400 MHz, DMSO-d_6)

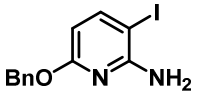
45



^{13}C NMR (101 MHz, DMSO-d_6)

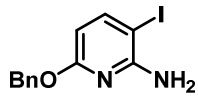
45

^{1}H NMR (400 MHz, DMSO-d_6)

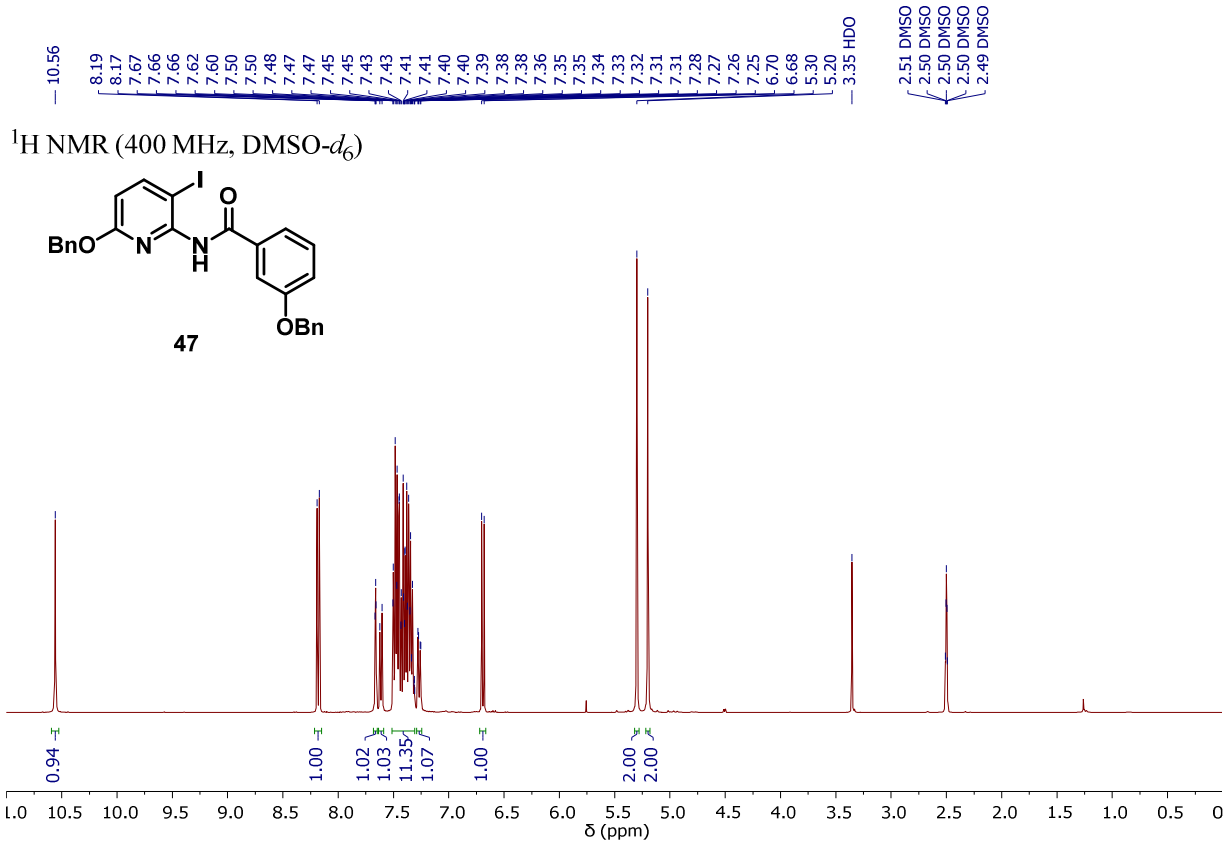


46

^{13}C NMR (101 MHz, DMSO-d_6)

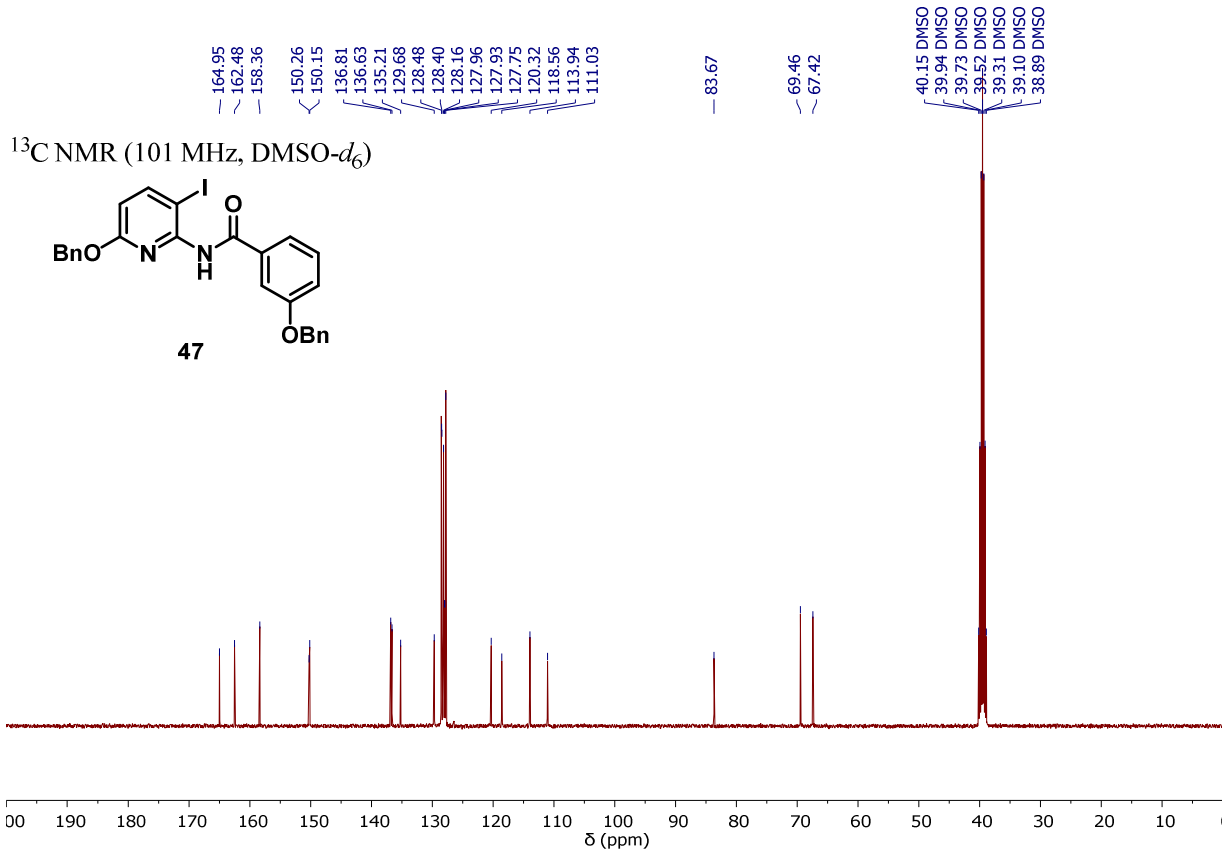


46



^{1}H NMR (400 MHz, DMSO-d_6)

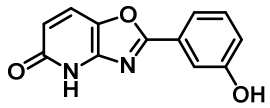
47



^{13}C NMR (101 MHz, DMSO-d_6)

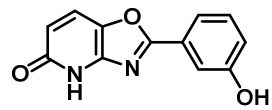
47

^{1}H NMR (400 MHz, THF-d_8)



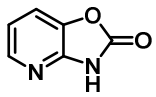
15

^{13}C NMR (151 MHz, THF-d_8)



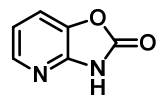
15

^{1}H NMR (400 MHz, DMSO-d_6)



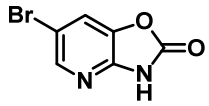
54

^{13}C NMR (101 MHz, DMSO-d_6)



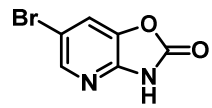
54

^{1}H NMR (400 MHz, DMSO-d_6)



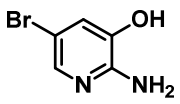
55

^{13}C NMR (101 MHz, DMSO-d_6)



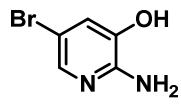
55

^{1}H NMR (400 MHz, DMSO-d_6)



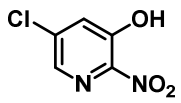
53

^{13}C NMR (101 MHz, DMSO-d_6)



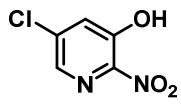
53

^{1}H NMR (400 MHz, $CDCl_3$)



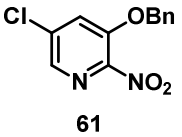
60

^{13}C NMR (101 MHz, $CDCl_3$)



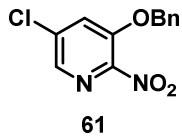
60

^{1}H NMR (400 MHz, DMSO-d_6)



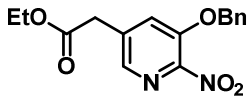
61

^{13}C NMR (101 MHz, DMSO-d_6)



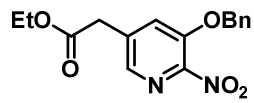
61

^{1}H NMR (400 MHz, DMSO-d_6)



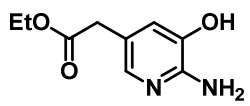
63

^{13}C NMR (101 MHz, DMSO-d_6)



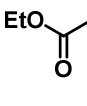
63

^{1}H NMR (400 MHz, DMSO-d_6)



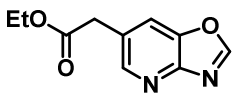
64

^{13}C NMR (101 MHz, DMSO-d_6)



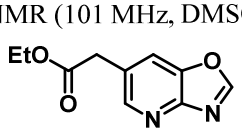
64

^{1}H NMR (400 MHz, DMSO-d_6)



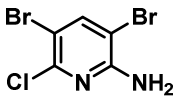
59

^{13}C NMR (101 MHz, DMSO-d_6)



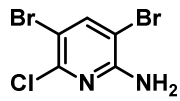
59

^{1}H NMR (400 MHz, DMSO-d_6)

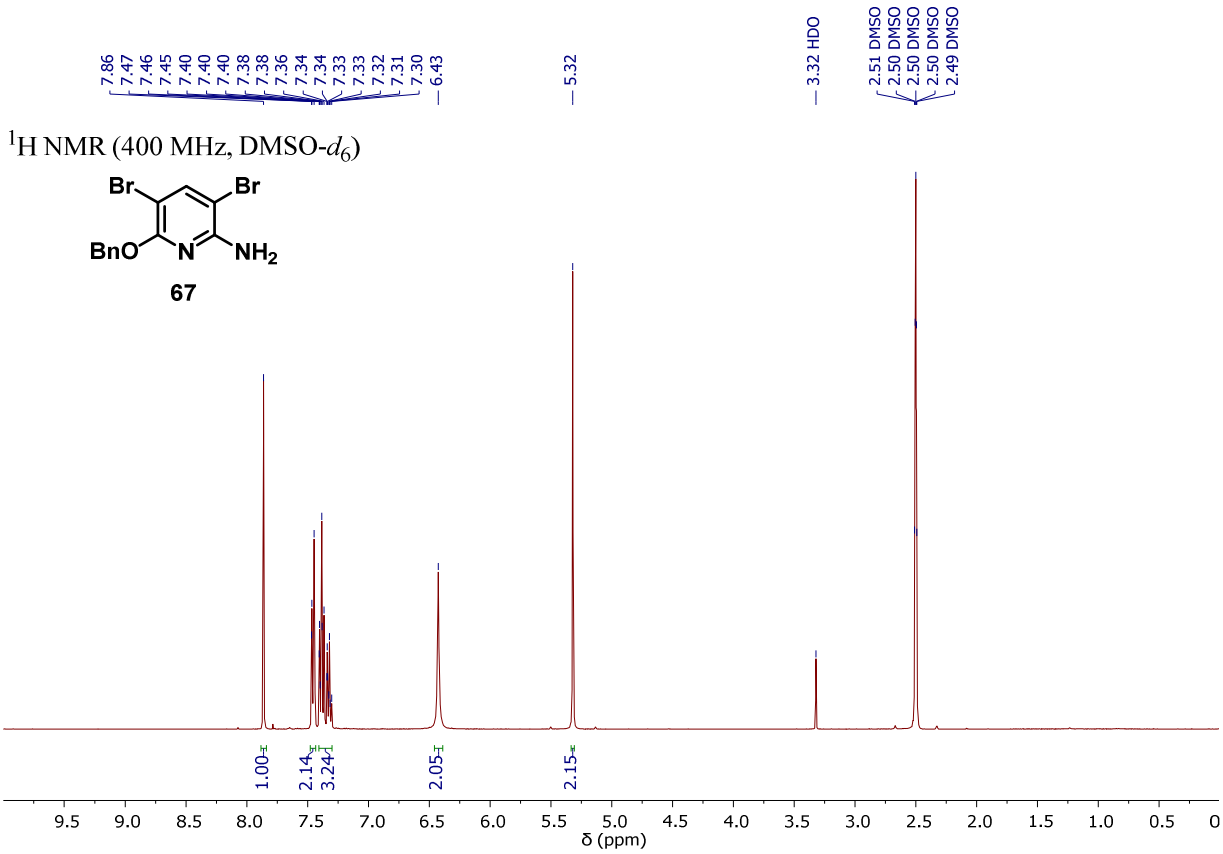


68

^{13}C NMR (101 MHz, DMSO-d_6)

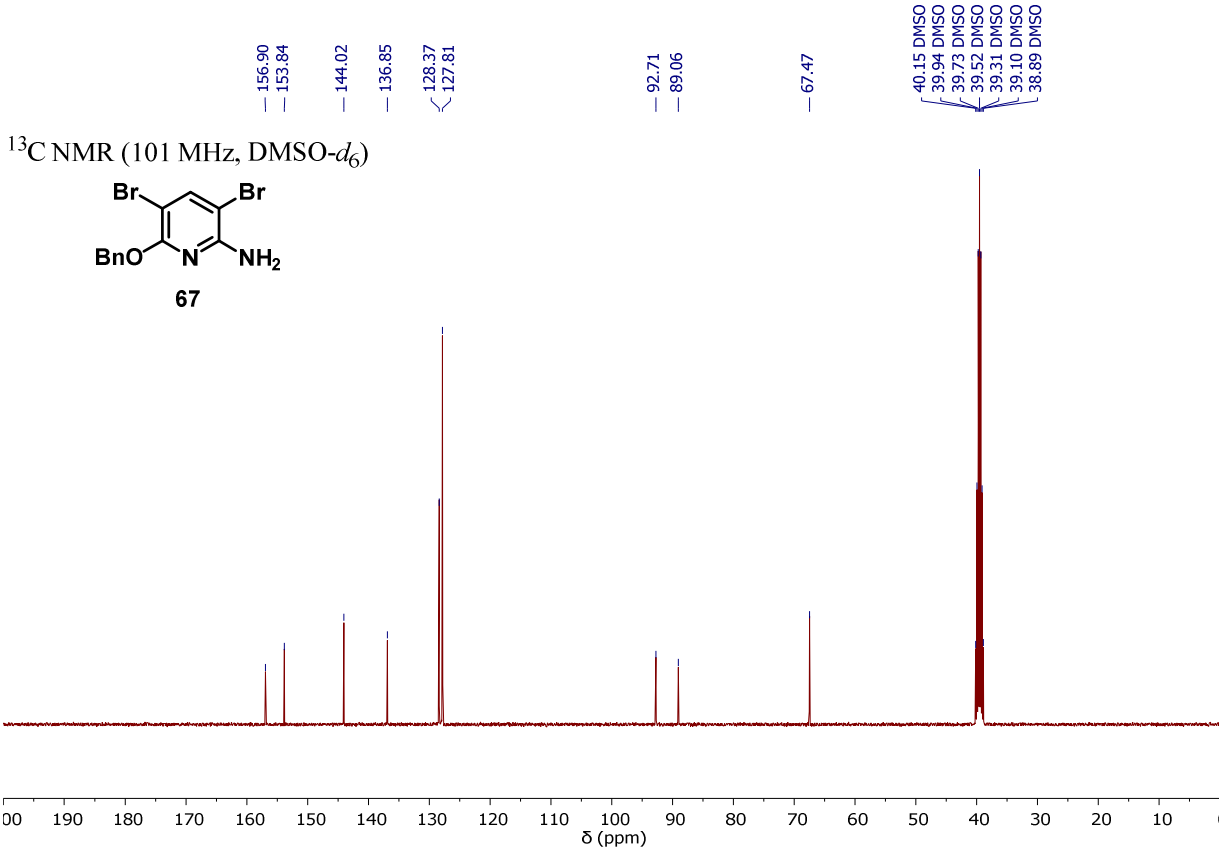


68



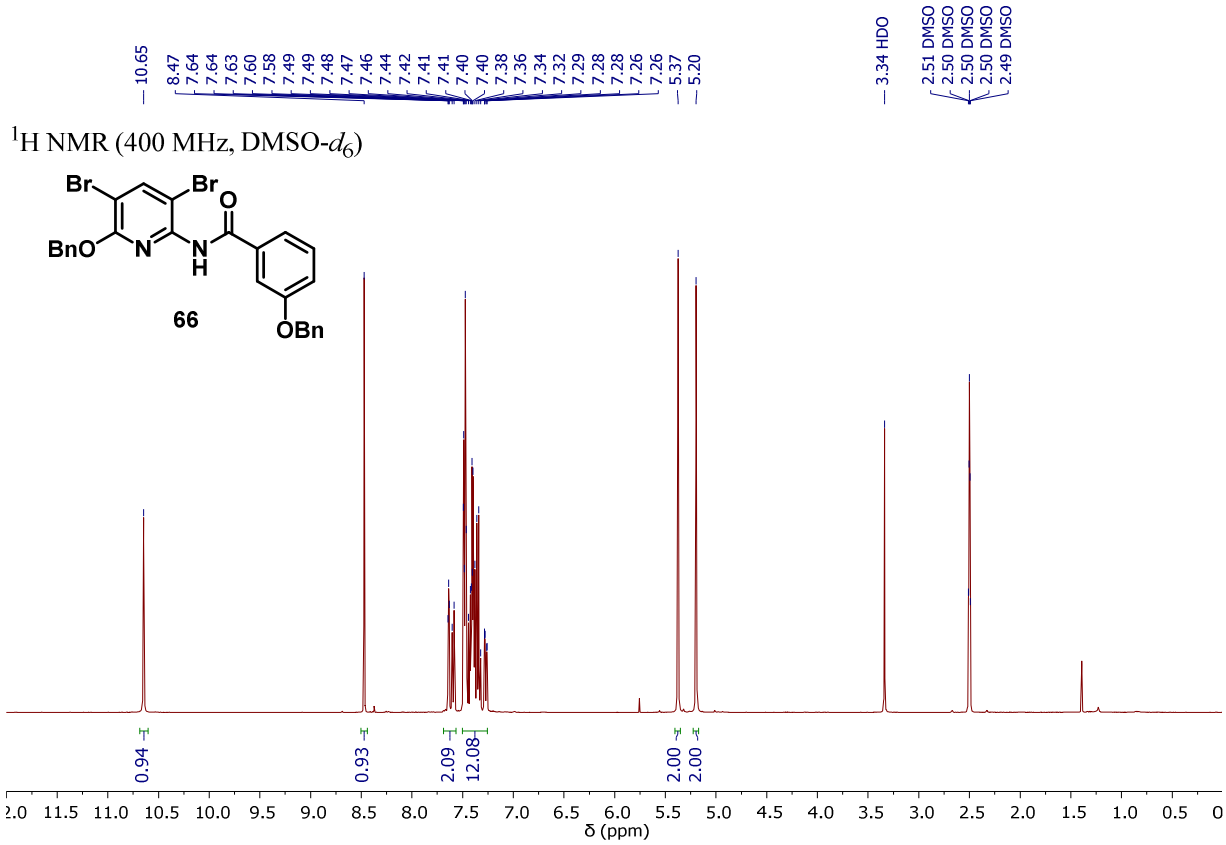
^{1}H NMR (400 MHz, DMSO-d_6)

67

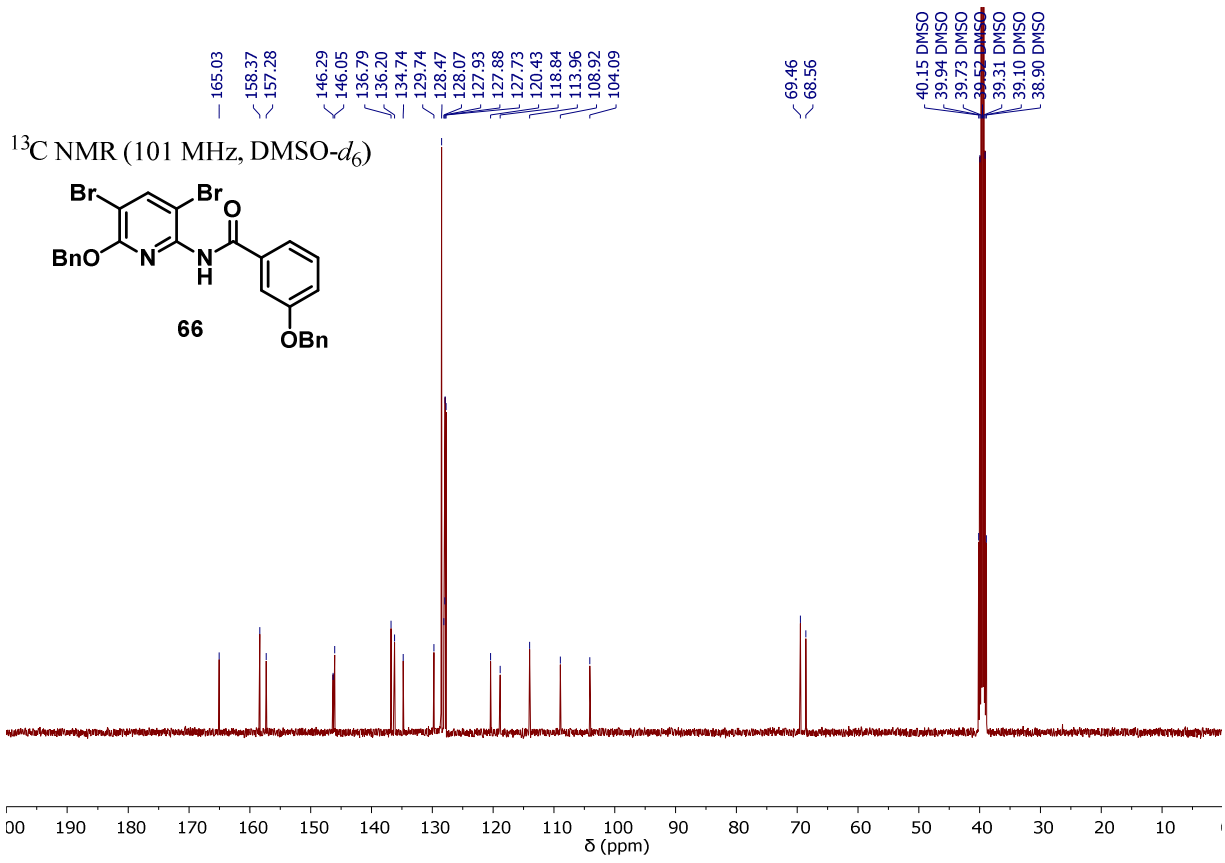


^{13}C NMR (101 MHz, DMSO-d_6)

67

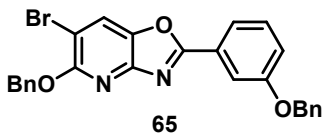


^{1}H NMR (400 MHz, DMSO-d_6)



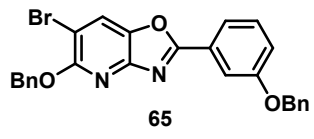
^{13}C NMR (101 MHz, DMSO-d_6)

^{1}H NMR (400 MHz, DMSO-d_6)



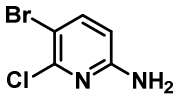
65

^{13}C NMR (101 MHz, DMSO-d_6)



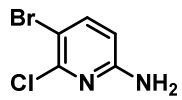
65

^{1}H NMR (400 MHz, DMSO-d_6)

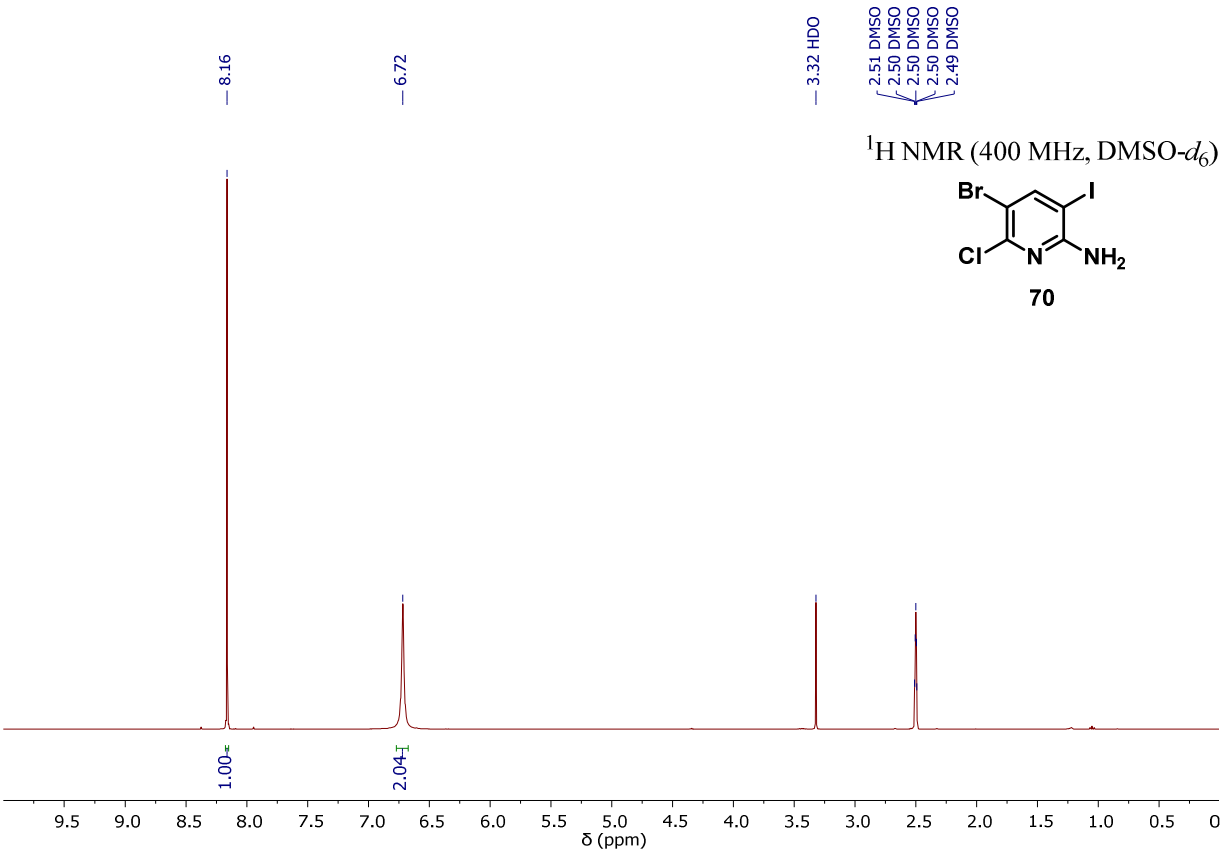


69

^{13}C NMR (101 MHz, DMSO-d_6)

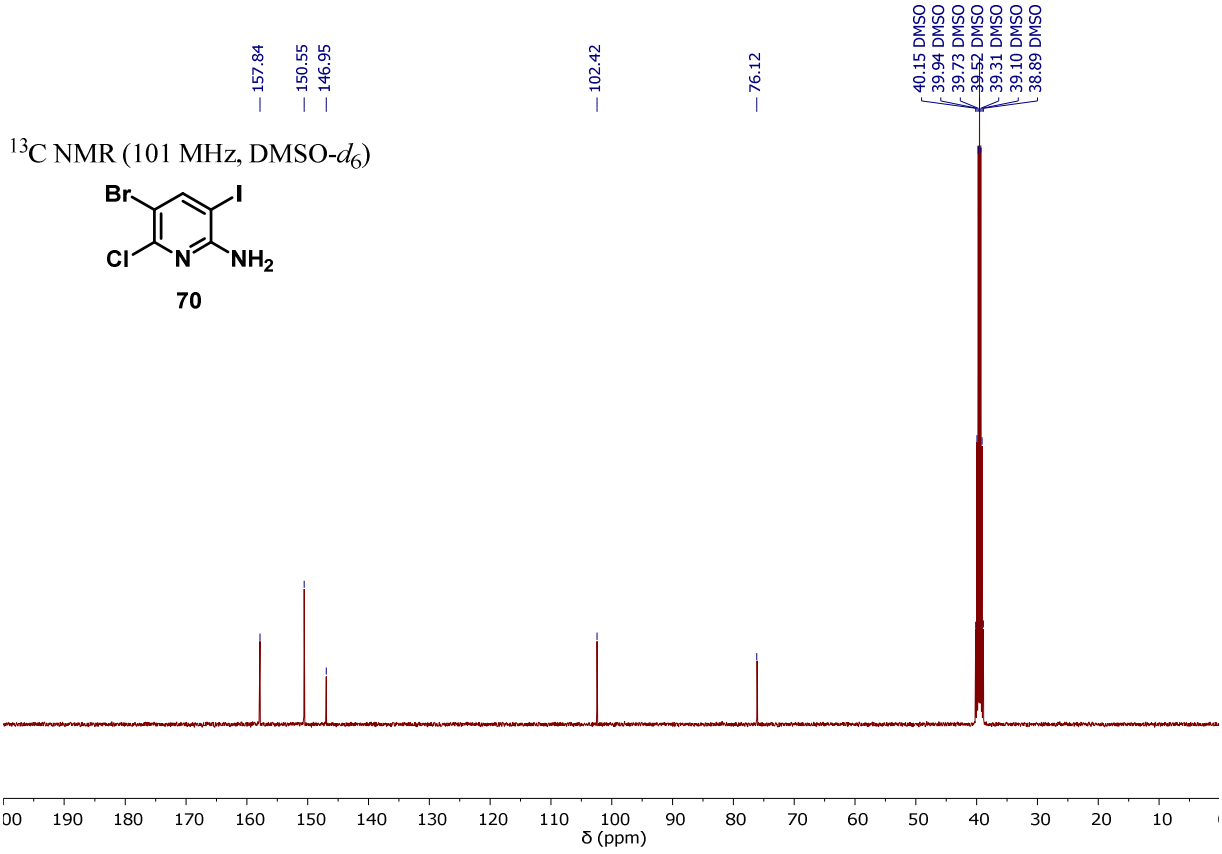


69



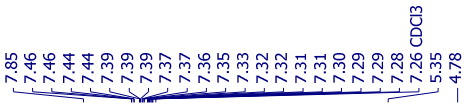
^{1}H NMR (400 MHz, DMSO-d_6)

70

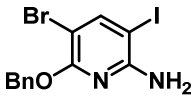


^{13}C NMR (101 MHz, DMSO-d_6)

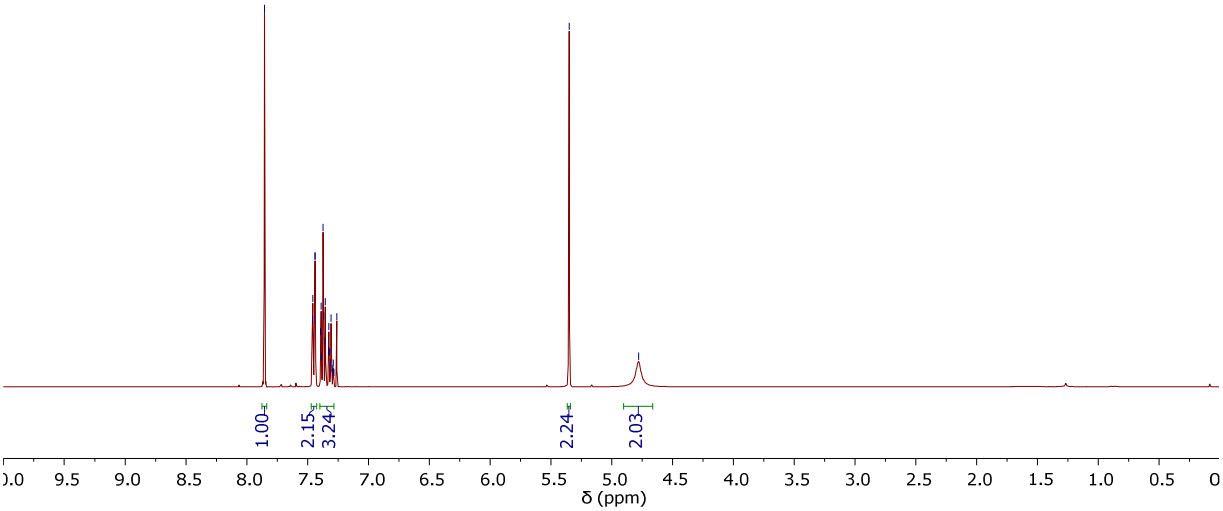
70



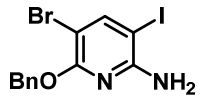
^{1}H NMR (400 MHz, $CDCl_3$)



71

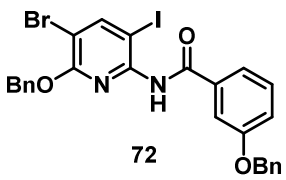


^{13}C NMR (101 MHz, $CDCl_3$)

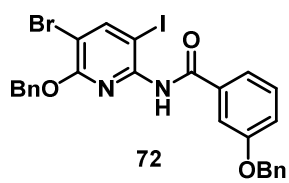


71

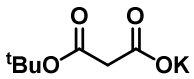
^{1}H NMR (400 MHz, DMSO-d_6)



^{13}C NMR (101 MHz, DMSO-d_6)

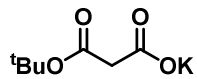


^{1}H NMR (400 MHz, DMSO-d_6)

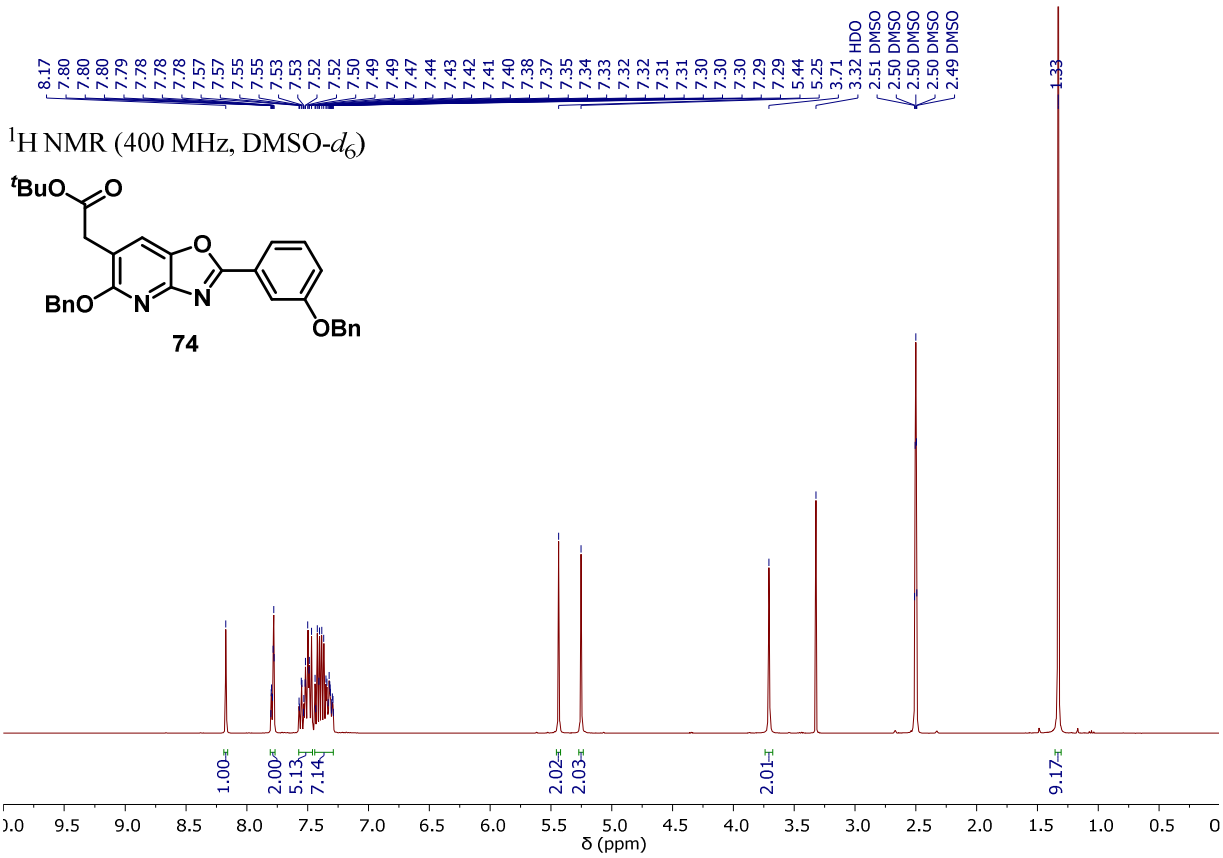


76

^{13}C NMR (101 MHz, DMSO-d_6)

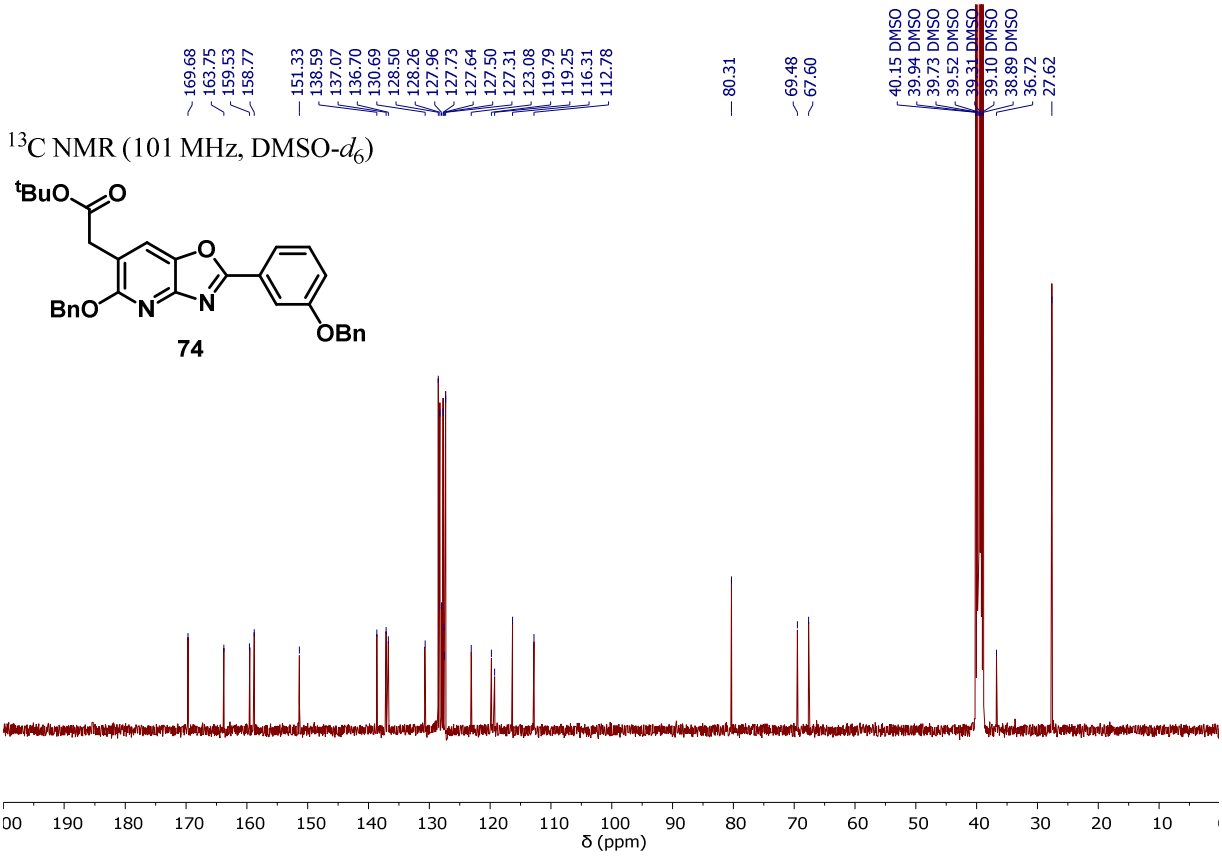


76



^{1}H NMR (400 MHz, DMSO-d_6)

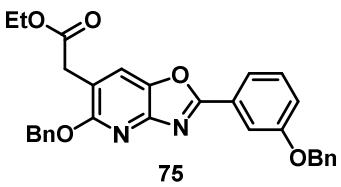
74



^{13}C NMR (101 MHz, DMSO-d_6)

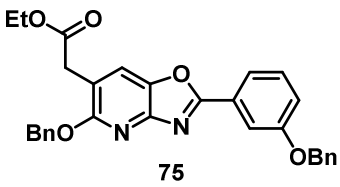
74

^{1}H NMR (400 MHz, DMSO-d_6)



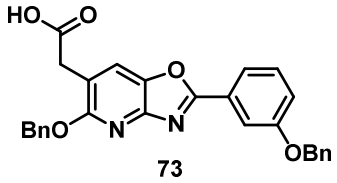
75

^{13}C NMR (101 MHz, DMSO-d_6)



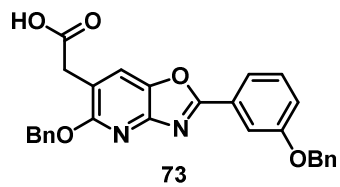
75

^{1}H NMR (600 MHz, DMSO-d_6)

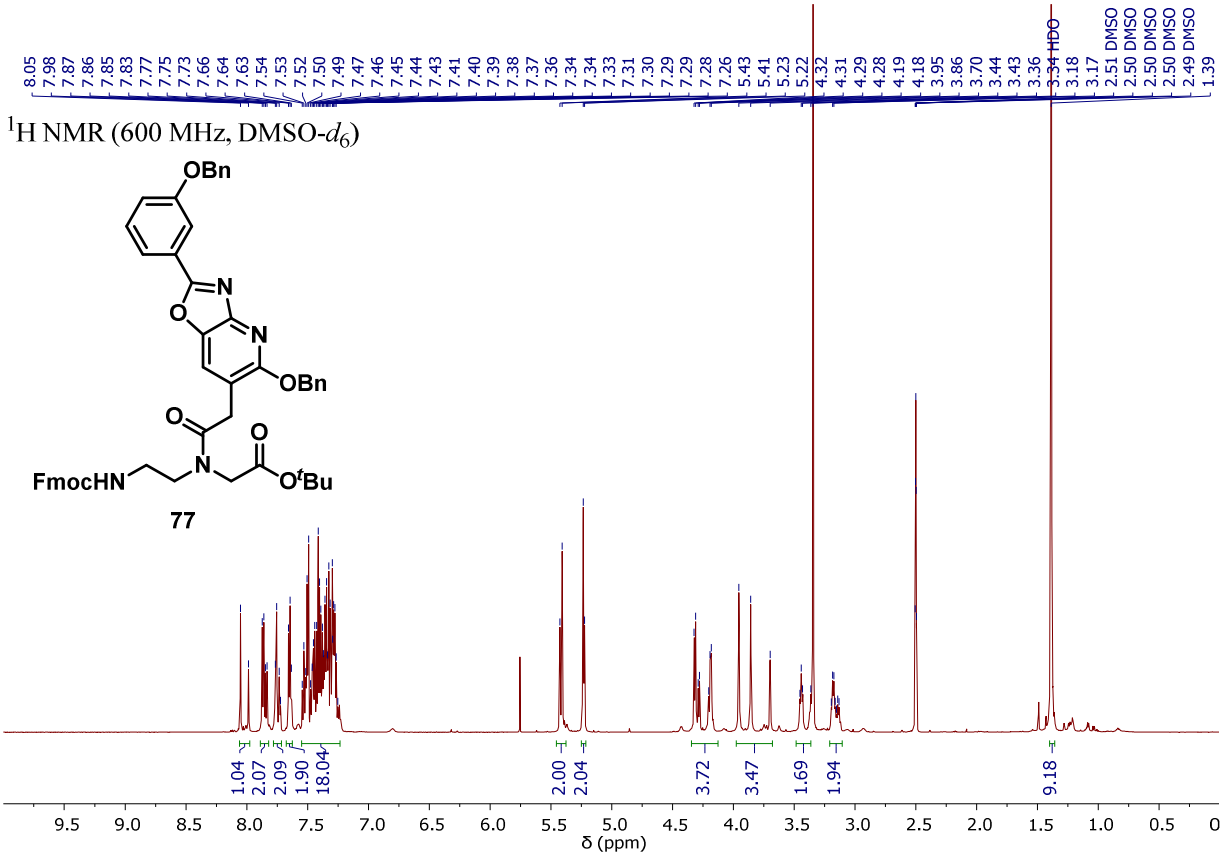


73

^{13}C NMR (151 MHz, DMSO-d_6)

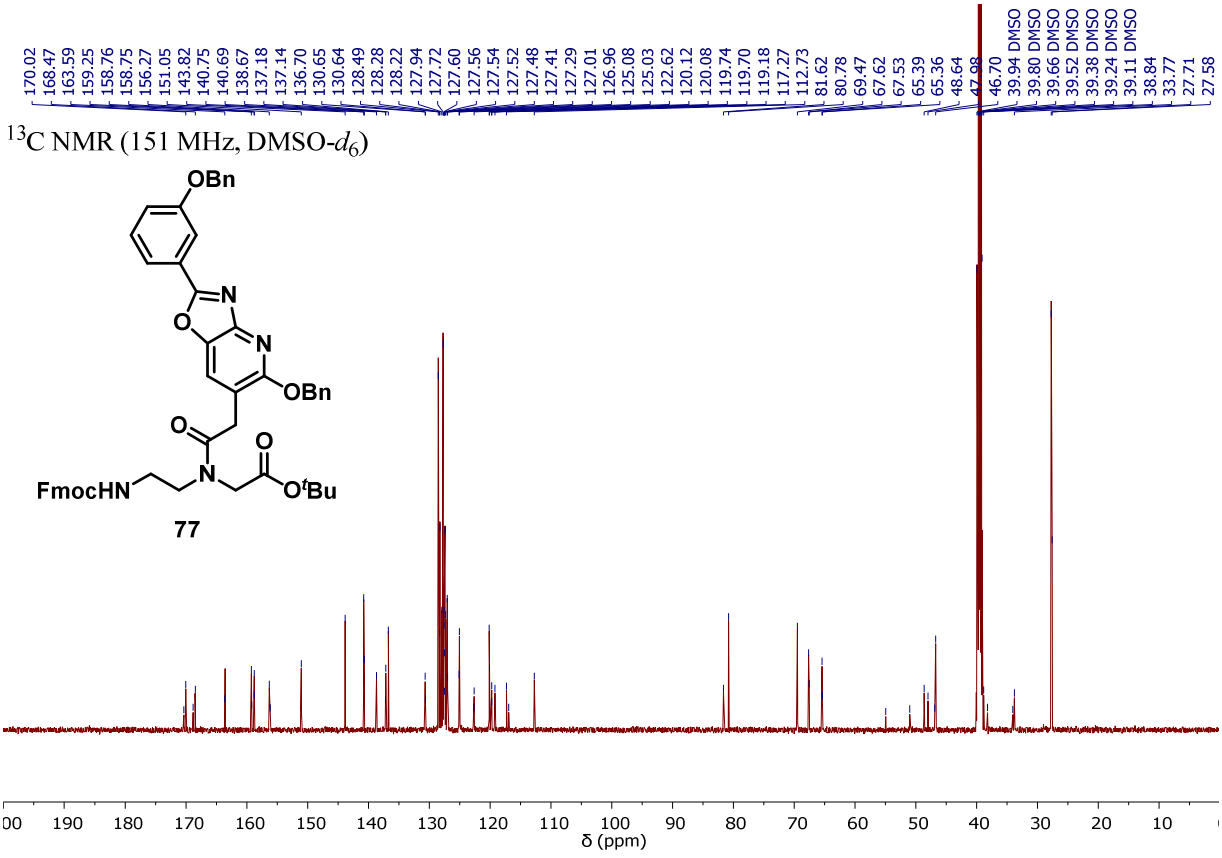


73



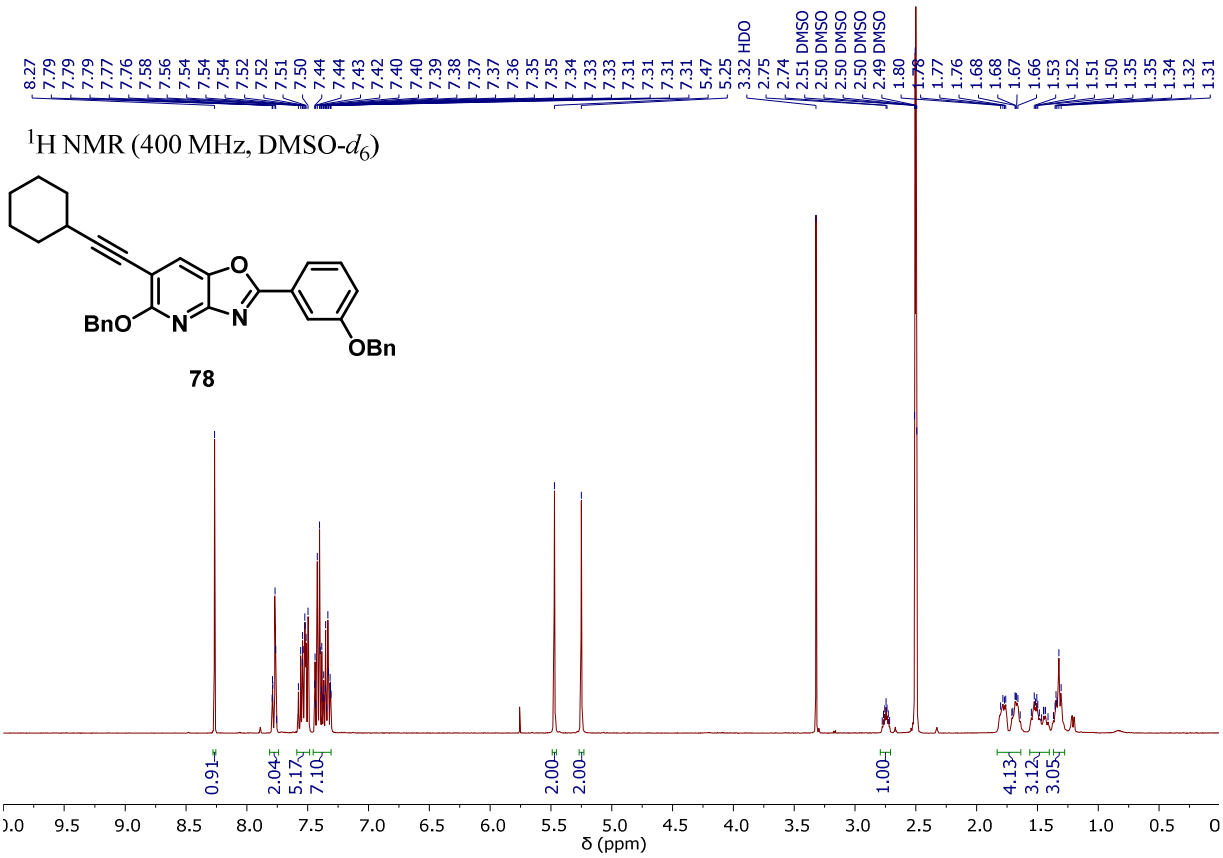
^{1}H NMR (600 MHz, DMSO-d_6)

77



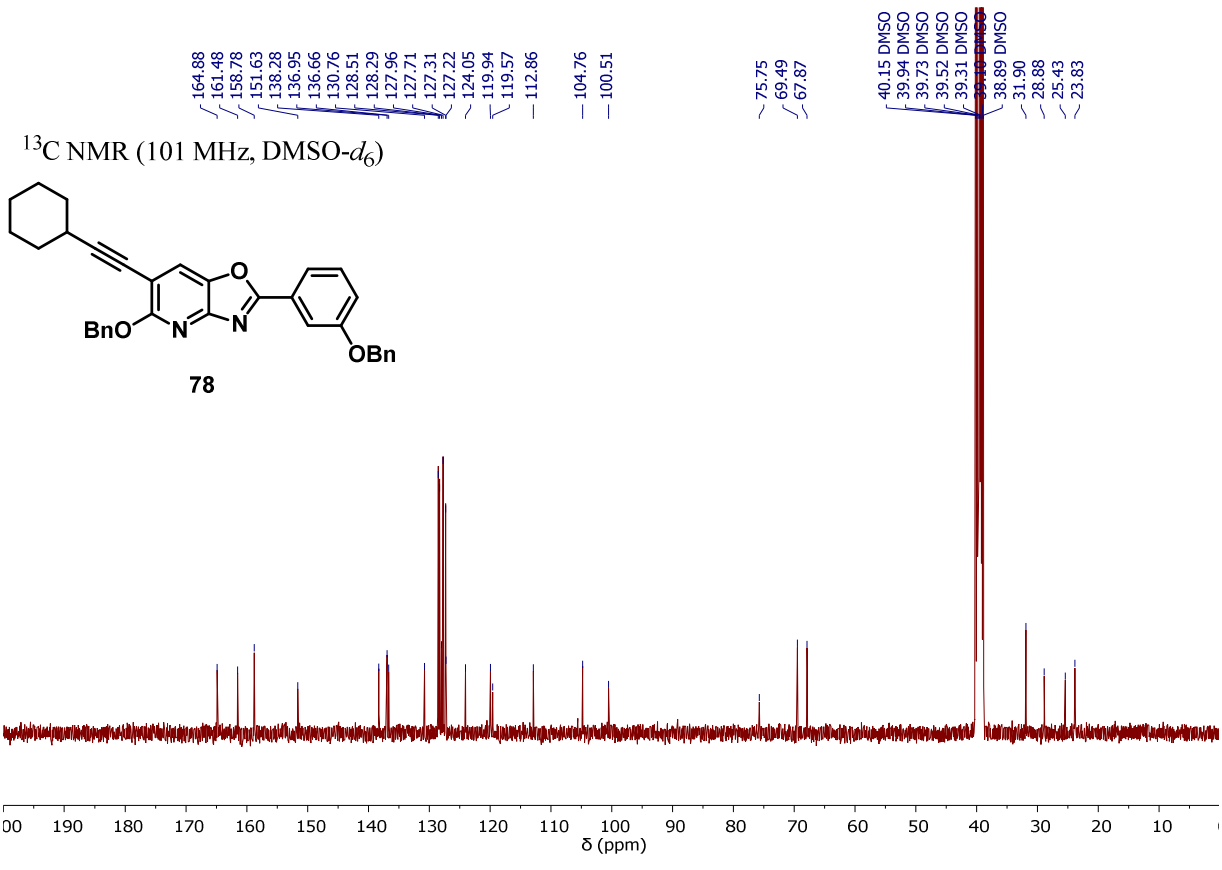
^{13}C NMR (151 MHz, DMSO-d_6)

77



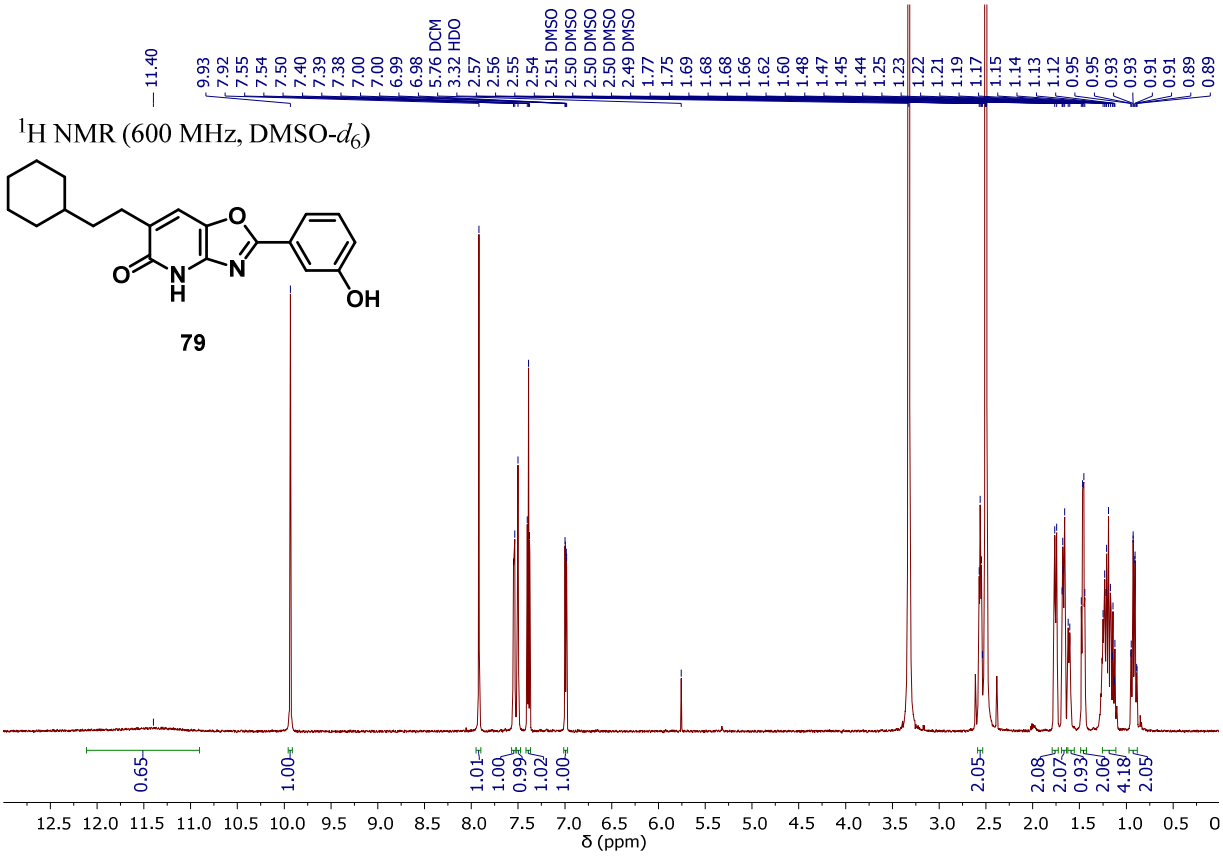
^{1}H NMR (400 MHz, DMSO-d_6)

78



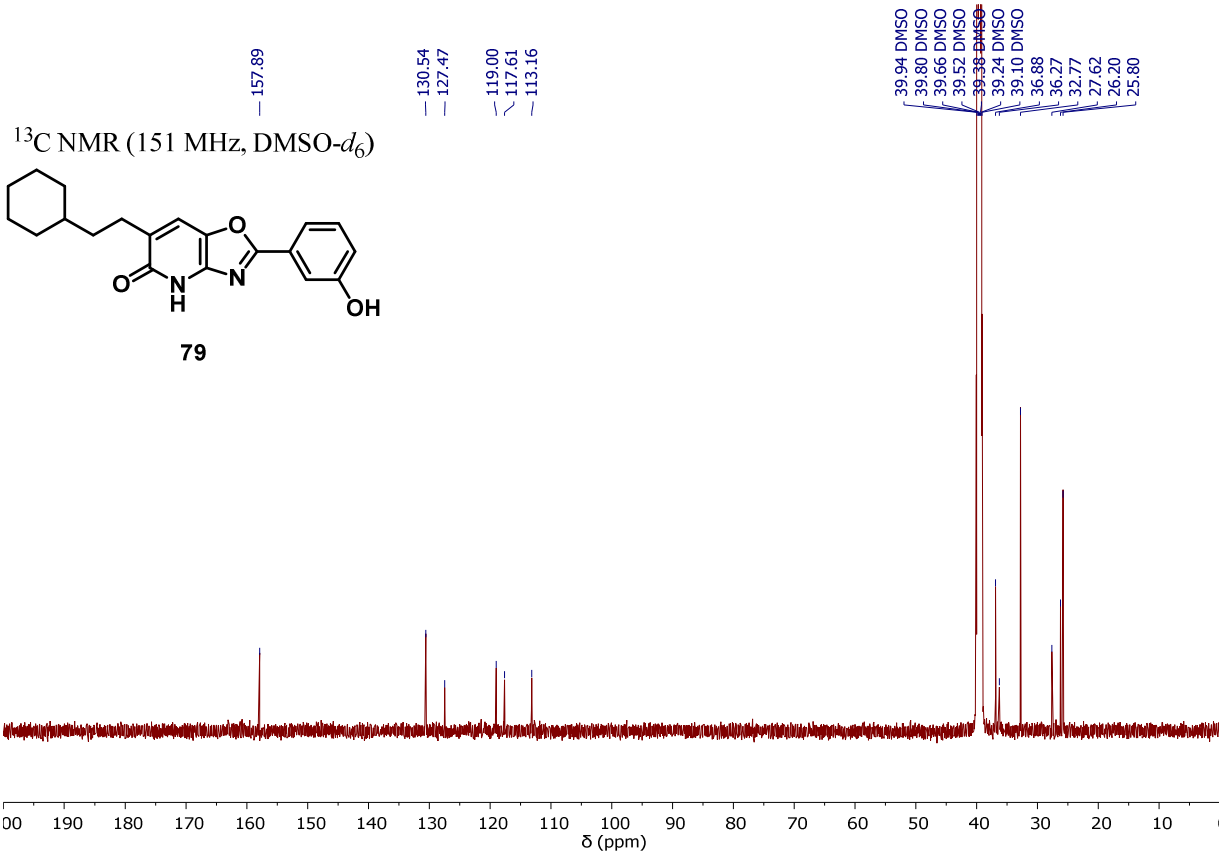
^{13}C NMR (101 MHz, DMSO-d_6)

78



^{1}H NMR (600 MHz, DMSO-d_6)

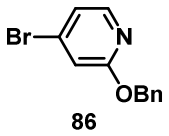
79



^{13}C NMR (151 MHz, DMSO-d_6)

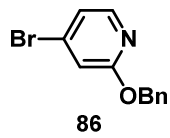
79

^{1}H NMR (400 MHz, DMSO-d_6)



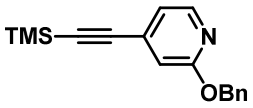
86

^{13}C NMR (101 MHz, DMSO-d_6)



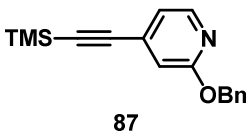
86

^{1}H NMR (400 MHz, DMSO-d_6)

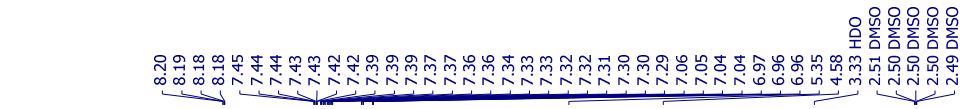


87

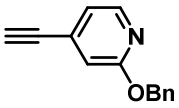
^{13}C NMR (101 MHz, DMSO-d_6)



87

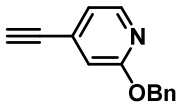


^{1}H NMR (400 MHz, DMSO-d_6)

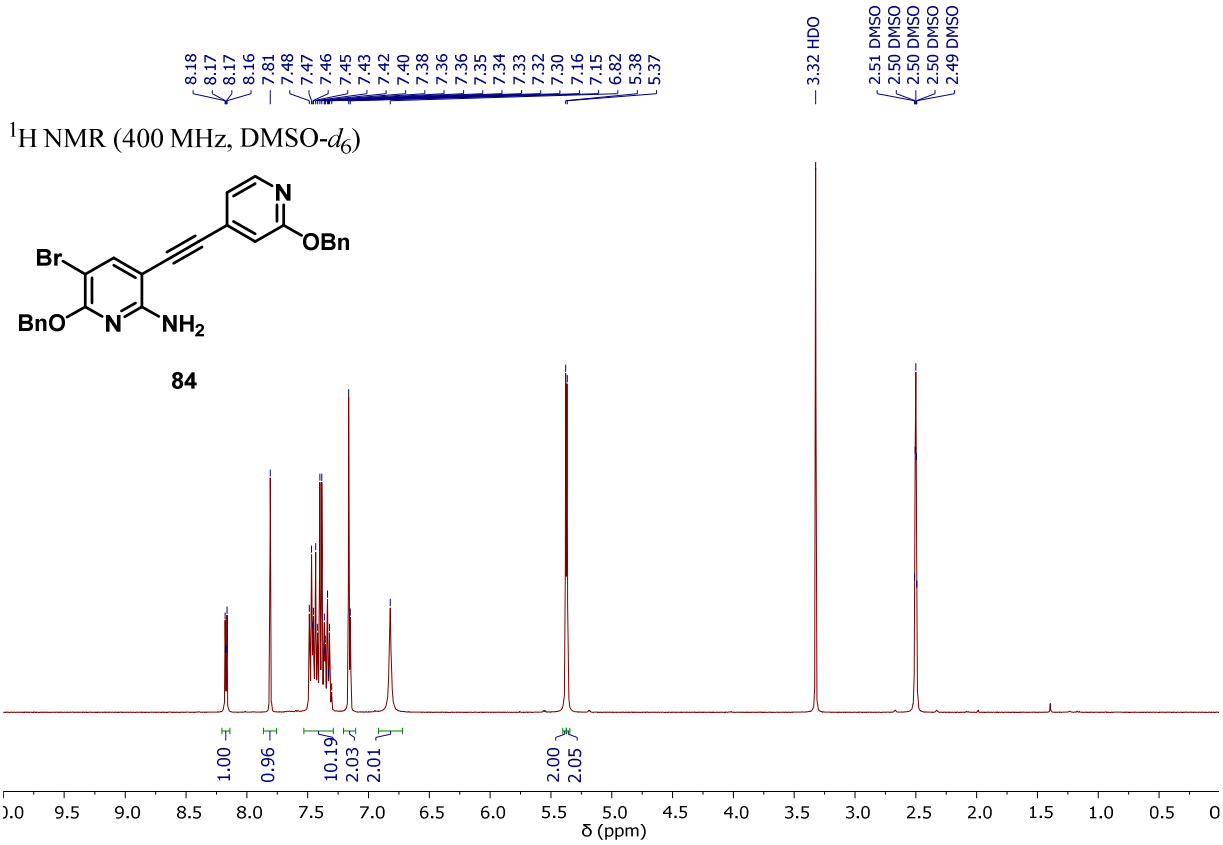


85

^{13}C NMR (101 MHz, DMSO-d_6)

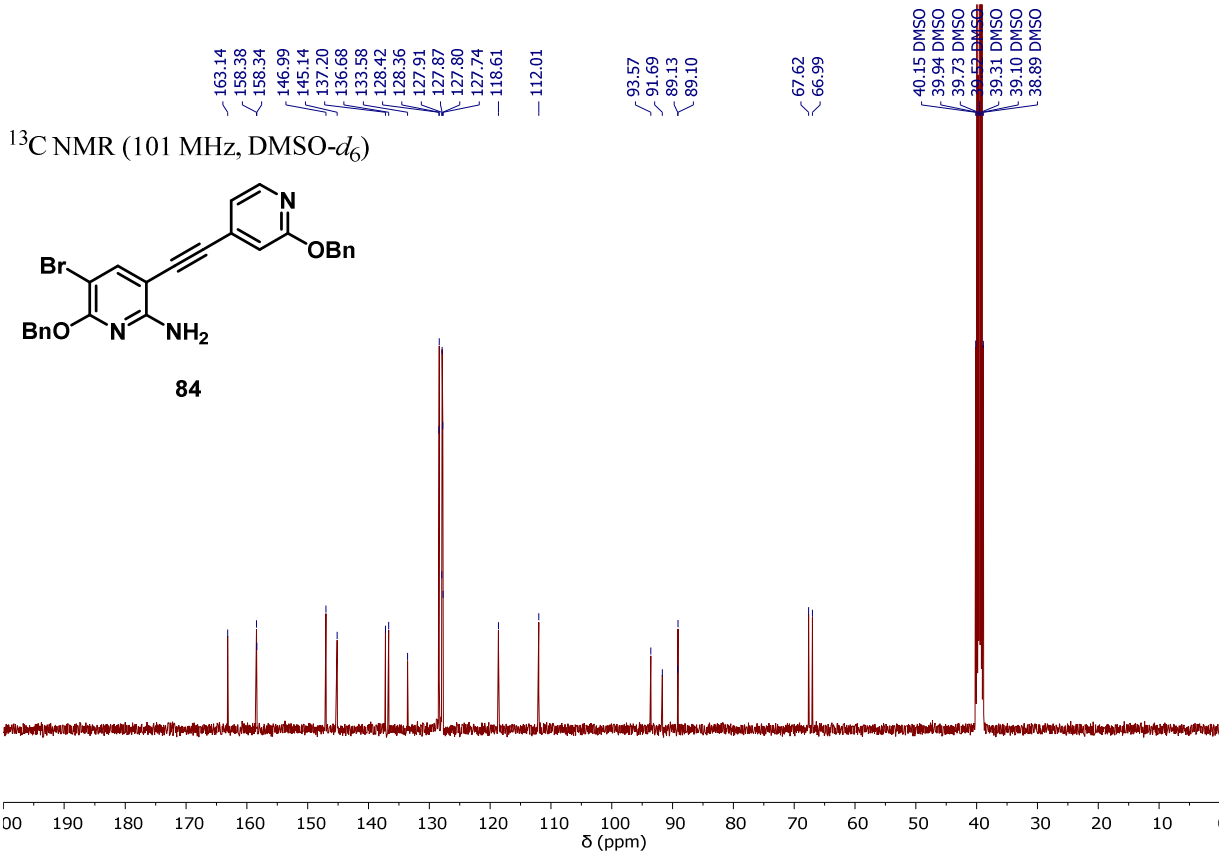


85



^{1}H NMR (400 MHz, DMSO-d_6)

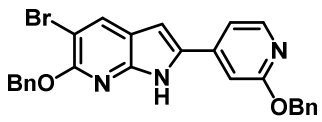
84



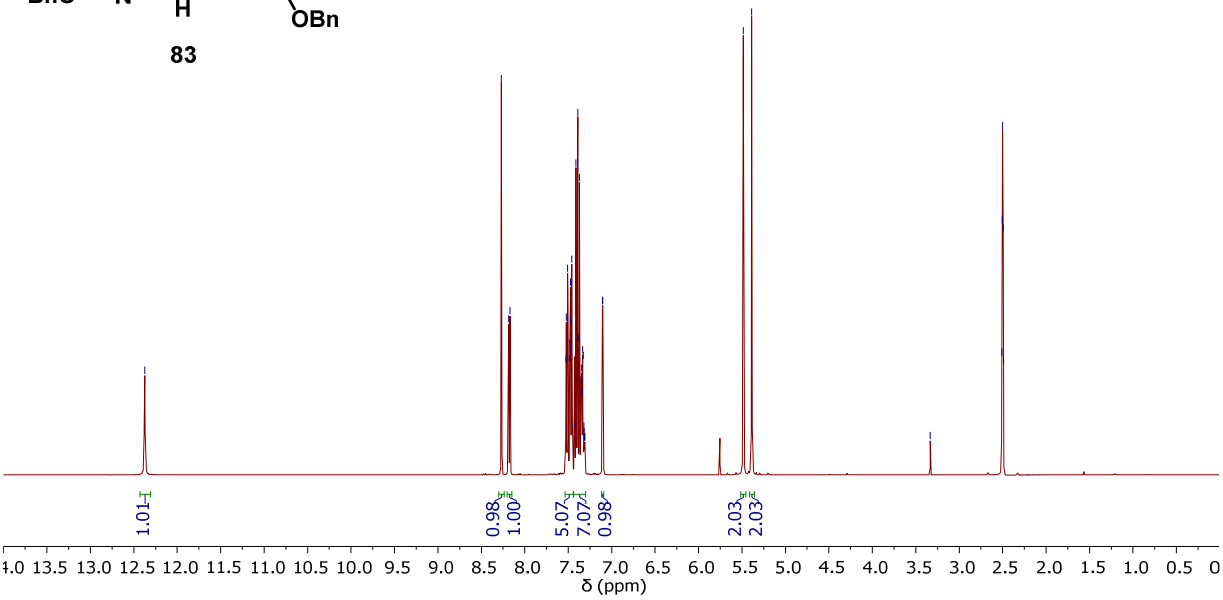
^{13}C NMR (101 MHz, DMSO-d_6)

84

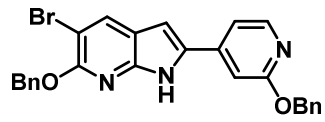
^{1}H NMR (400 MHz, DMSO-d_6)



83

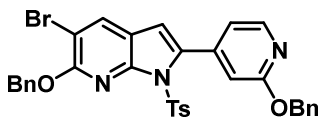


^{13}C NMR (101 MHz, DMSO-d_6)



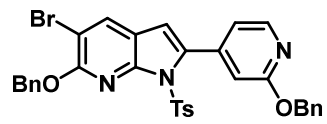
83

^{1}H NMR (400 MHz, DMSO-d_6)

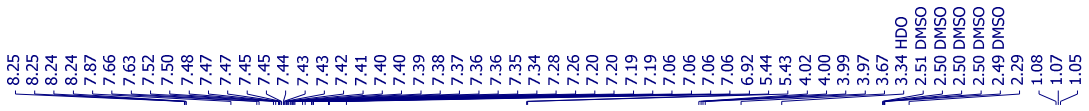


88

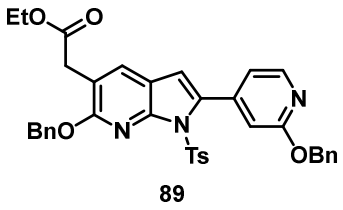
^{13}C NMR (101 MHz, DMSO-d_6)



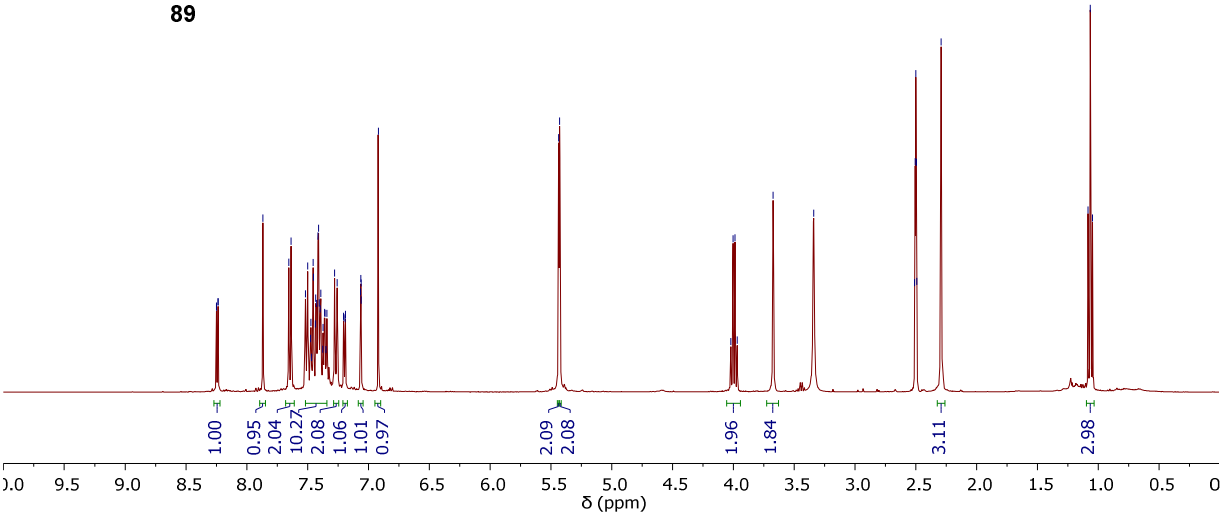
88



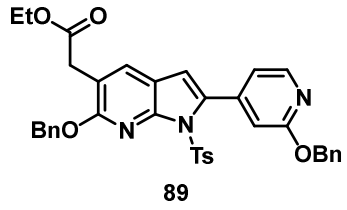
^{1}H NMR (400 MHz, DMSO-d_6)



89

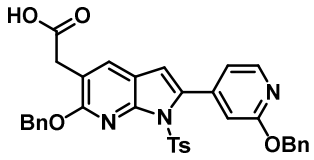


^{13}C NMR (101 MHz, DMSO-d_6)



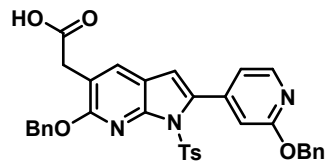
89

^{1}H NMR (400 MHz, DMSO-d_6)



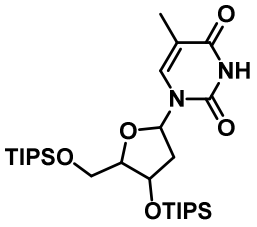
82

^{13}C NMR (101 MHz, DMSO-d_6)



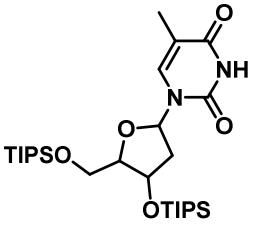
82

^{1}H NMR (400 MHz, DMSO-d_6)



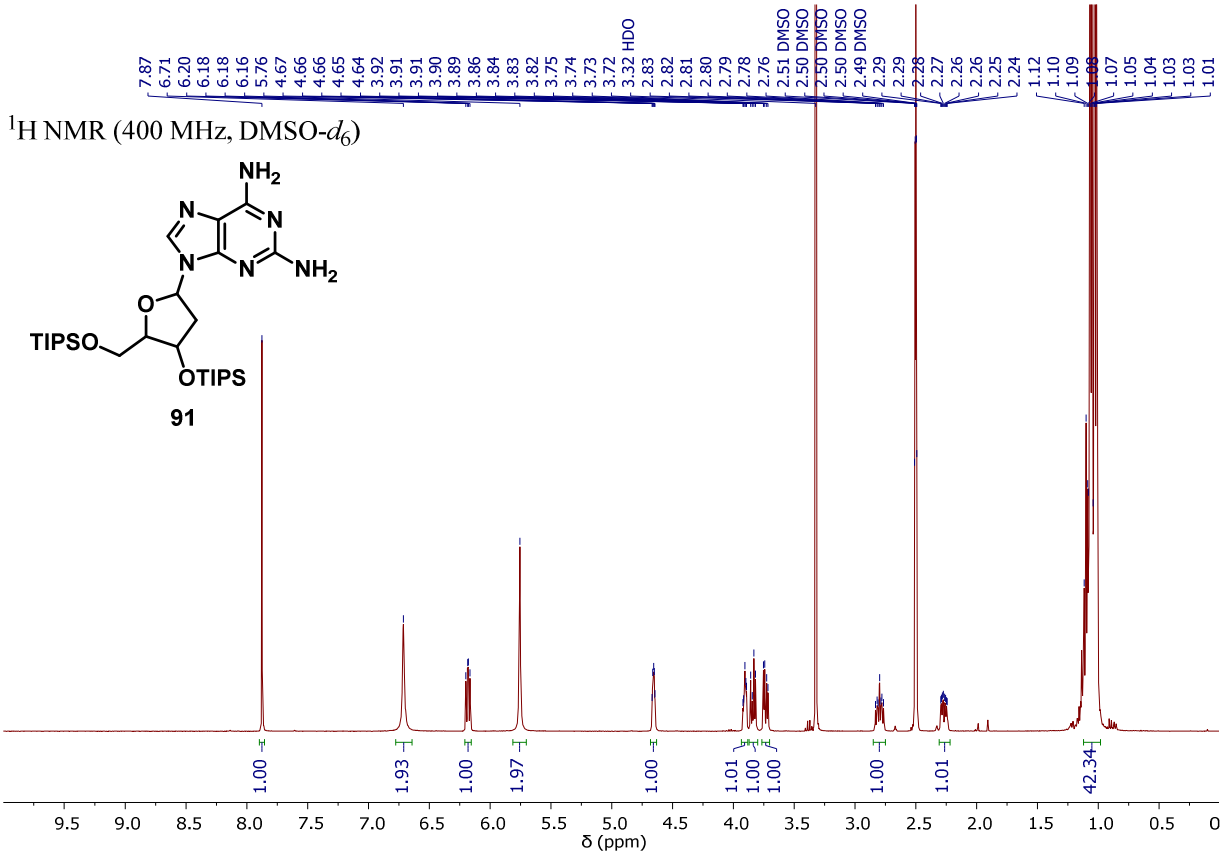
90

^{13}C NMR (101 MHz, DMSO-d_6)



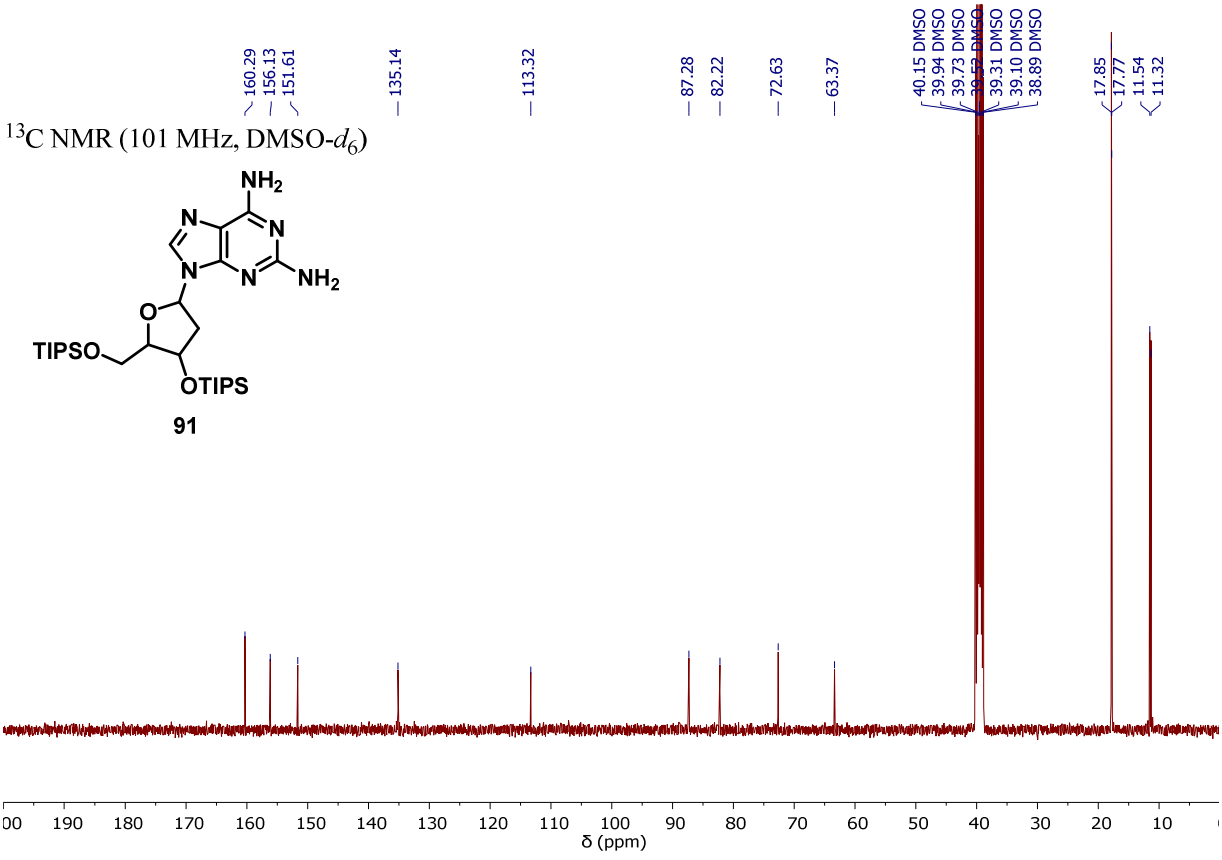
90

^{1}H NMR (400 MHz, DMSO-d_6)



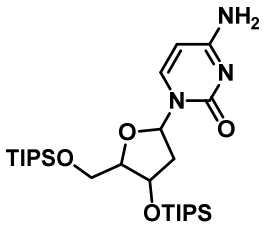
91

^{13}C NMR (101 MHz, DMSO-d_6)



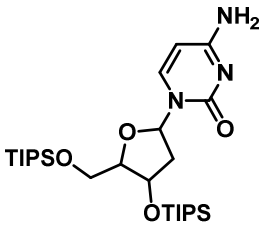
91

^{1}H NMR (400 MHz, DMSO-d_6)

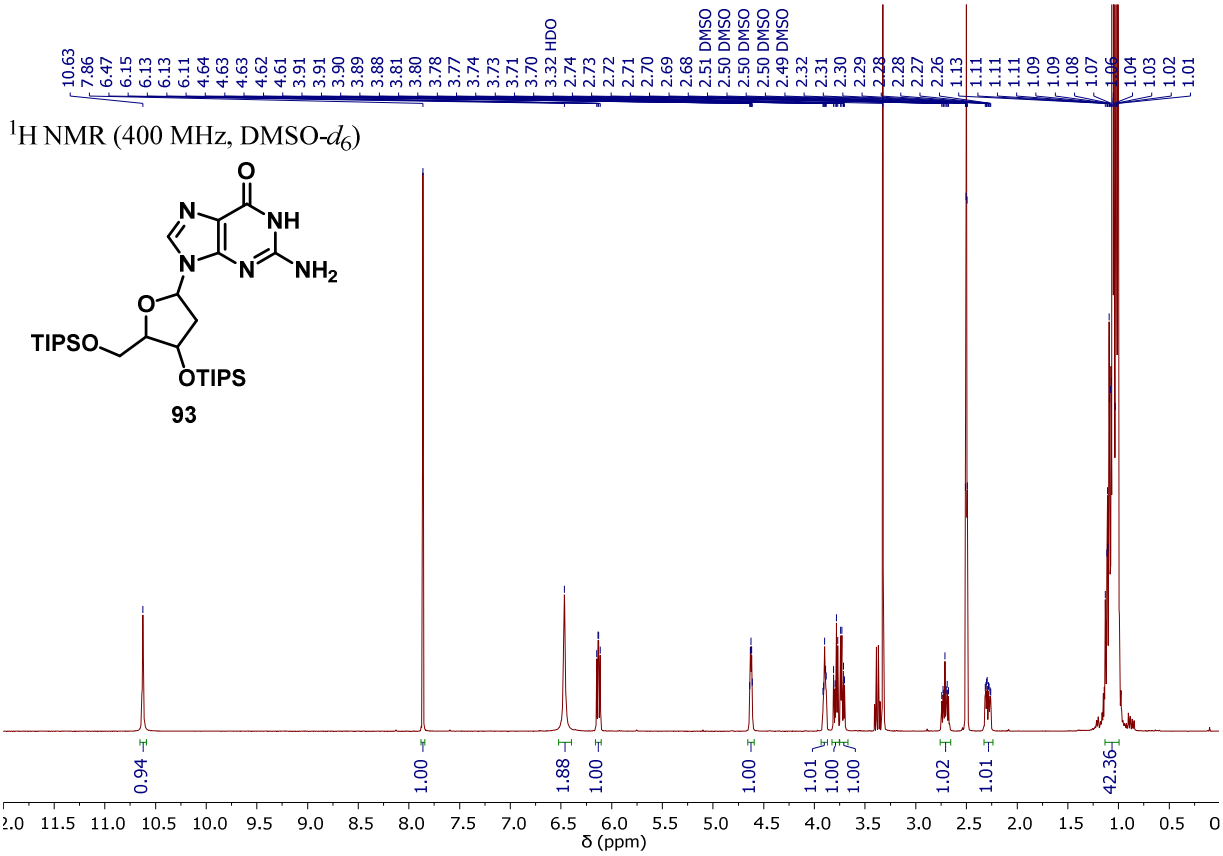


92

^{13}C NMR (101 MHz, DMSO-d_6)

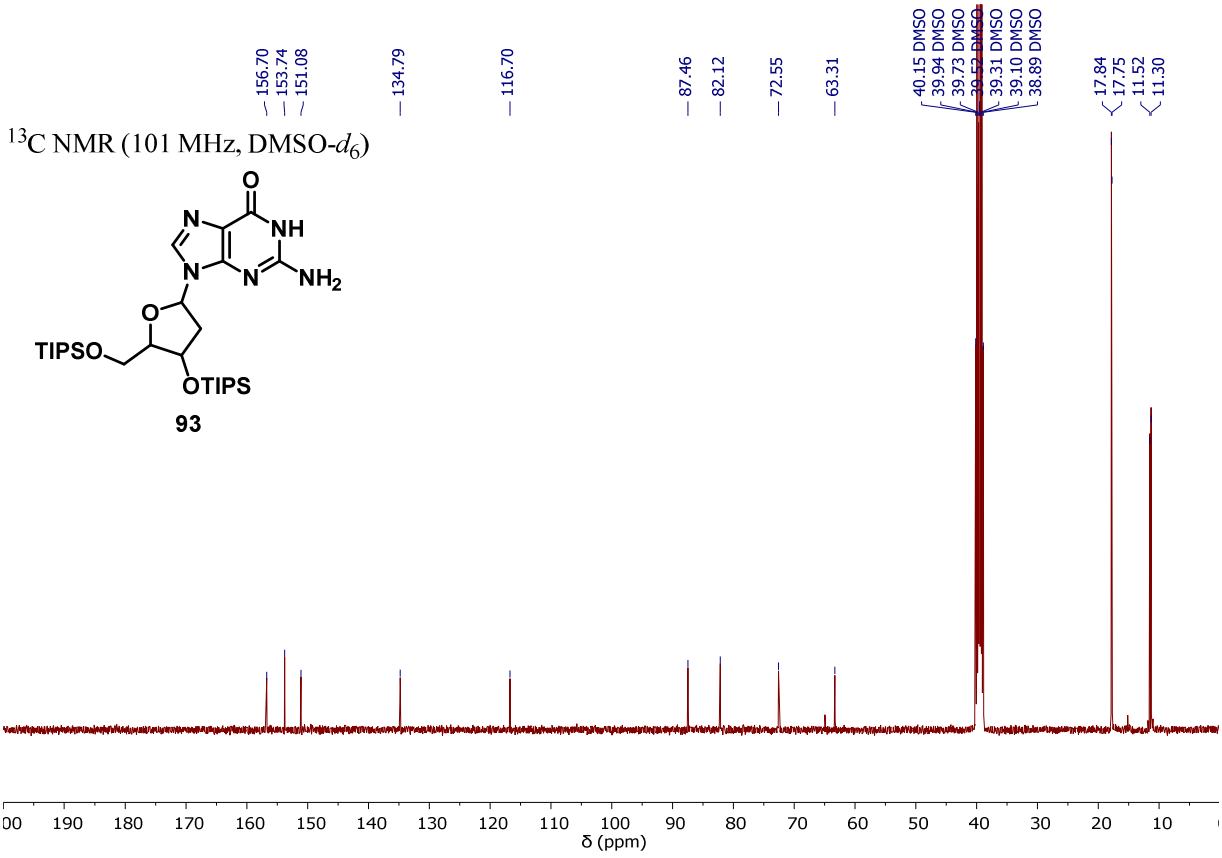


92



^{1}H NMR (400 MHz, DMSO-d_6)

93



^{13}C NMR (101 MHz, DMSO-d_6)

93

7.3. Curriculum Vitae

The Curriculum Vitae is not included in the online version for data protection reasons.

7.4. Acknowledgments/Agradecimientos

I am very grateful to have had the opportunity to do my Ph.D. thesis in Germany. These four years have meant a personal and professional growth that I cannot describe in words, but I will make an effort to...

First of all, I would like to thank my supervisor, Prof. Thomas Schrader, for giving me the opportunity to do my doctoral thesis in his group. He always allowed me to experiment and put my ideas into practice, even if many of them ended in failure. This allowed me to learn from my mistakes and to improve as a scientist. I am also especially grateful for his support on a personal level throughout the unfortunate events that have marked this thesis.

I would like to make a special mention to **Vanessa**, my work and life partner, with whom I have shared many good and bad moments during this PhD. Undoubtedly the difficult moments are more bearable with the right person by your side. I love you!

I have to mention the members of my lab, the S07 S05 D36, the "DNA lab" and certainly the best in the whole floor (and in the building, of course). To Shubankar, or "Shubi", for the good times we had during the time we shared during the beginning of my thesis. To Chris, my "Azubi", a great lab technician who is not only eager to learn, but also a great partner who enhances the atmosphere with his character. And of course to Nahid, my eternal lab partner, with whom I have laughed, fought, frustrated, and a myriad of experiences that are inevitable when you share almost half of your day with one person for three years. Working with her will certainly be missed.

I would also like to thank the rest of my colleagues. To Inesa, my first lab partner and great support during the bureaucratic chaos of moving to Germany. To Abbna and Sujenth, for all the good times we have shared together. To Philipp, for being my companion of frustrations during this thorny path called Ph.D., and for his infinite help with the German translations. To Antonio and Estelle, for bringing a bit of cultural diversity to our group. To Robert, My and Dennis, for being such good partners.

To Jessica Kunter, Heike Wöll and Don Marvin Voß for their repeated help during these 4 years.

To Dr. Torsten Schaller and Dr. Felix Niemeyer for their help with the (sometimes challenging) NMR characterization.

A very significant stage of my thesis, which many people underestimate, has been the supervision of Master's lab courses, which has not only allowed me to gain experience in teaching, but has also given me the opportunity to meet many great people. Among them I

want to highlight Louisa, a great person and scientist who surprisingly supports me despite my craziness. Also Mohammed, a great friend with whom I have shared countless laughs.

On a more personal level, I would like to thank my mother, as it has been her continued effort and dedication that has allowed me to get to where I am today. Also to my non-biological family, my in-laws and of course Morris, for their continued support during this adventure.

I cannot end the acknowledgements section without mentioning Alexandra Elbakyan, creator of Sci-Hub, for making science open access and not a business.

Estoy muy agradecido de haber tenido la oportunidad de realizar mi tesis doctoral en Alemania. Estos cuatro años han significado un crecimiento personal y profesional que no puedo describir con palabras, pero haré un esfuerzo...

En primer lugar, me gustaría agradecerle a mi supervisor, el Prof. Thomas Schrader, el haberme dado la oportunidad de realizar mi tesis doctoral en su grupo. Siempre me permitió experimentar y poner en práctica mis ideas, aunque muchas de ellas acabasen en fracaso. Ello me permitió aprender de mis errores y mejorar como científico. También le estoy especialmente agradecido por su apoyo a nivel personal a lo largo de los desafortunados acontecimientos que han marcado esta tesis.

Me gustaría hacer una mención especial a **Vanessa**, compañera de trabajo y de vida, con la que he compartido muchos buenos y malos momentos durante este doctorado. Sin duda los momentos difíciles son más llevaderos con la persona correcta a tu lado. ¡Te quiero!

Tengo que mencionar a los integrantes de mi laboratorio, el S07 S05 D36, el "laboratorio del ADN" y sin duda el mejor de toda la planta (y del edificio, por supuesto). A Shubankar, o "Shubi", por los buenos momentos que pasamos durante el tiempo que compartimos durante el comienzo de mi tesis. A Chris, mi "Azubi", un gran técnico de laboratorio que no solo tiene ganas de aprender, sino que también es un gran compañero que mejora el ambiente con su carácter. Y por supuesto a Nahid, mi eterna compañera de laboratorio, con la que me he reído, peleado, frustrado, y un sinfín de experiencias que son inevitables cuando compartes casi la mitad de tu día con una persona durante tres años. Sin duda se va a echar en falta trabajar con ella.

También quiero darle las gracias al resto de mis compañeros. A Inesa, mi primera compañera de laboratorio y gran apoyo durante el caos burocrático que es mudarse a alemania. A Abbna y Sujenth, por todos los buenos momentos que hemos compartido juntos. A Philipp, por

ser mi compañero de frustraciones durante este camino de espinas llamado doctorado, y por su infinita ayuda con las traducciones al alemán. A Antonio y Estelle, por traer un poco de diversidad cultural a nuestro grupo. A Robert, My y Dennis, por ser tan buenos compañeros.

A Jessica Kunter, Heike Wöll y Don Marvin Voß por su ayuda en repetidas ocasiones durante estos 4 años.

A Dr. Torsten Schaller y Dr. Felix Niemeyer por su ayuda con la caracterización por RMN.

Una etapa muy significativa de mi tesis, y la cual muchos menosprecian, ha sido la supervisión de prácticas de Máster, que no solo me ha permitido adquirir experiencia en la enseñanza, si no que me ha dado la oportunidad de conocer a mucha gente genial. Entre ellos quiero destacar a Louisa, gran persona y científica que sorprendentemente me soporta a pesar de mi locura. También a Mohammed, gran amigo con el que he compartido incontables risas.

A un nivel más personal, me gustaría darle las gracias a mi madre, ya que ha sido su continuo esfuerzo y dedicación lo que me ha permitido llegar a donde estoy hoy. También a mi familia no biológica, a mis suegros y por supuesto Morris, por su continuo apoyo durante esta aventura.

No puedo acabar la sección de agradecimientos sin mencionar a Alexandra Elbakyan, creadora de Sci-Hub, por hacer que la ciencia sea de libre acceso y no un negocio.