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

Structural Insights into DNA Mimic Foldamer Recognition of DNA Binding Proteins

Deepak

aus

Gatauli, India

2023

Erklärung

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

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 10.11.2023 _____

Deepak

Dissertation eingereicht am	07.12.2023
1. Gutachter:	Prof. Dr. Ivan Huc
2. Gutachter:	Prof. Dr. Daniel Merk
Mündliche Prüfung am	16.01.2024

Table of Contents

1. Abstract 5

2. Introduction 7

2.1 DNA binding proteins 10

2.2 Techniques to study DBPs-DNA complexes 18

2.3 X-ray crystallography 19

2.4 Importance of DBPs 25

2.5 Targeting protein-DNA interactions 25

2.6 Foldamers, going beyond biopolymers 35

2.7 Sac7d, a thermophilic marvel 41

3. Objectives and rationales 44

4. DNA mimic foldamer recognition of a chromosomal protein 48

4.1 Publication 50

5. Sac7d-foldamer adducts: a tethering approach 80

5.1 Design and methodology 83

5.2 Results and discussion 90

5.3 Conclusion and perspective 97

6. Targeting hcGAS by DNA mimic foldamers 100

6.1 Introduction 101

6.2 Design and methodology 104

6.3 Results and discussion 110

6.4 Conclusion and perspective 119

7. Targeting Dpo4 by DNA mimic foldamers 121

7.1 Objective of the study 124

7.2 Design and methodology 125

7.3 Results and discussion 128

7.4 Conclusion and perspective 138

8. Conclusion and perspective 140
8.1 Conclusion 141
8.2 Future perspective 142
8.3 Ongoing challenges 144
9. References 146
10. Appendix 159
10.1 Supplementary figures 160
10.2 Abbreviations 164
10.3 Index of figures 166
10.4 Index of tables 173
10.5 List of publications, PDB depositions, and conference contributions 174
10.6 Acknowledgment 175

1. Abstract

Abstract

The remarkable folding capabilities of biopolymers such as proteins and DNA underpin their exceptional performances in various biological processes, including molecular recognition, catalysis, and information storage. These natural biopolymers have the unique ability to precisely position functional groups in three-dimensional space, orchestrating their dynamic functions. Inspired by biopolymers, foldamers have been developed, molecules that fold into three-dimensional shapes and provide access to functions beyond the capabilities of biopolymers.

Taking inspiration from DNA mimic proteins, abiotic DNA mimics based on aromatic oligoamide foldamers were designed to mimic the shape and surface features of double-stranded DNA. These foldamers previously have been shown to interfere with protein-nucleic acid interactions (PNIs) and bind better than DNA itself. In this thesis, we present the design features of DNA mimic foldamers with features including C_2-symmetry (to mimic palindromic DNA sequences), chirality control (to mimic the B-DNA by introducing stereogenic center in the foldamer) suitable for biophysical and structural characterization with DNA binding proteins. Initial efforts made towards the recognition of Sac7d protein were performed by first characterizing the binding of the racemic DNA mimic foldamers using surface plasmon resonance and circular dichroism. Later, we characterized the binding of chiral C_2-symmetrical DNA mimic foldamers with Sac7d using X-ray crystallography and nuclear magnetic resonance. As evidenced by solid-state structure elucidation, DNA mimic foldamer finds a novel binding orientation on Sac7d despite maintaining the key interactions involved with Sac7d-DNA complex, which was also confirmed in solution by NMR spectroscopy.

Next, we investigated the binding of Dpo4 and hcGAS with DNA mimic foldamers. In this regard, we crystallized and solved the structure of the apo hcGAS protein. Initial crystals of the hcGAS-foldamer diffracted only to a lower resolution. For Dpo4, it was first crystallized, and structure was solved with its DNA sequence, and 4 Å data for Dpo4-foldamer was collected. However, efforts are ongoing to collect higher-resolution datasets and screen different lengths of foldamers with both proteins.

Put together, our results present DNA mimic foldamer as a potential molecular tool to interfere with and investigate protein-DNA interactions. The findings of this research may unlock new possibilities in understanding and manipulating protein-nucleic acid interactions, with broad implications for biology and pharmacology.

2. Introduction

Nucleic acids and proteins play central roles in numerous cellular and physiological processes.[1] The fundamental principle governing the flow of genetic information, known as the central dogma of life,[2] revolves around the intricate interactions between nucleic acids and proteins. Rather than existing in isolation, nucleic acids and proteins are intertwined components of this crucial biological framework. The central dogma begins with the process of deoxyribonucleic acid (DNA) transcription into ribonucleic acid (RNA), followed by translation to synthesize proteins. From an evolutionary perspective, nucleic acids, primarily RNA, can be viewed as the ancestral predecessors of proteins. According to the "RNA world hypothesis," in the early stages of life, genetic information was stored within RNA, and RNA also catalyzed chemical reactions crucial for cellular functions.[3] However, during the course of evolution, genetic information transitioned to being predominantly stored in DNA. DNA emerged as the preferred carrier of genetic information over RNA due to the presence of deoxyribose in its sugar-phosphate backbone. This deoxyribose component enhances the chemical stability of DNA when compared to RNA, allowing DNA chains to maintain their integrity over longer lengths without breakages.

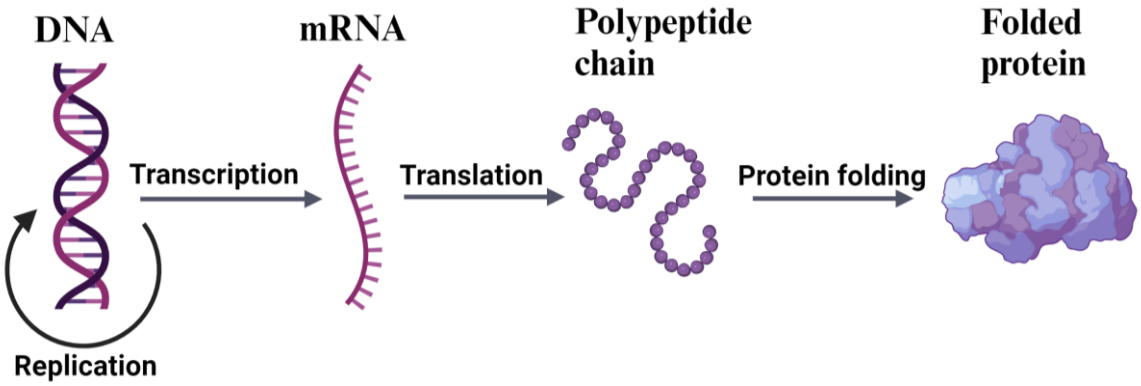


Fig. 1. Illustration of the central dogma of biology and key players involved.

On the other hand, proteins gradually took on the role of catalyzing biochemical reactions and became integral structural elements of life. Nonetheless, it is essential to emphasize that the interaction between DNA and proteins is of paramount significance for the proper functioning of cellular activities in all living organisms. (**Fig. 1**).

The key stabilizing forces in DNA and proteins are the same, e.g., hydrophobic contacts, Van der Waals interactions, and hydrogen bonds. However, nucleic acid's backbones differ from proteins as DNA is uniformly negatively charged, an attribute of phosphate charges (**Fig. 2**).

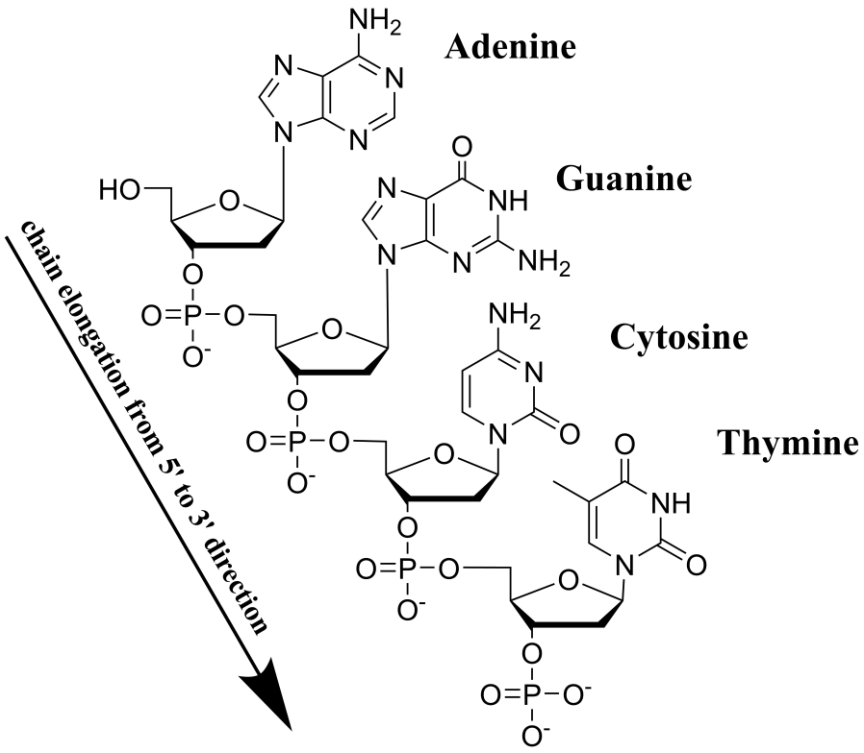


Fig. 2. Fragment of DNA primary structure shown with the sequence 5'-AGCT-3'.

Compared to proteins, secondary structures in nucleic acids are more stable than tertiary structures. Although it is possible to form tertiary structures, the double helix is still the predominant characteristic of nucleic acids, particularly DNA. In 1953, Watson and Crick first introduced the concept of base pairing between nitrogen bases of DNA strands, forming the double-stranded double helical fold of DNA.[4] A straight ladder conformation of DNA could also provide complementarity of bases, but the helical nature of double-stranded DNA is essential for efficient genetic storage and stability. It allows compact packaging, prevents unraveling, and enables precise replication and transcription of DNA.

DNA has the ability to adopt at least two right-handed double helical structures, which are commonly referred to as A-DNA and B-DNA. Watson and Crick's proposed structure of DNA was B form, which is most stable at high water content, but as the water content is reduced, it converts to A form. Moreover, this transition from A to B -DNA is also mainly exploited by various DNA-binding proteins. Under physiological conditions, the B form of DNA dominates compared to the A form. However, this could be altered by various proteins, for example, those involved in DNA bending and damage repair. Apart from these, DNA could also adopt a left-handed double helical form denoted as Z-DNA.[5] The base pairing in Z-DNA follows the classical Watson-Crick base pairing and non-canonical base pairing. Regardless of the form, the double helix of DNA is stable primarily due to the repulsive negative charge at the

phosphate backbone and the aromatic stacking of the nitrogen base pairs. DNA bending proteins exploit both features to induce a bend in the DNA (section **2.7.1**). However, it is worth mentioning here that long segments of DNA on their own can get a supercoiled structure when the ends of DNA are not accessible to move around. This feature of DNA can be seen in bacterial plasmids.

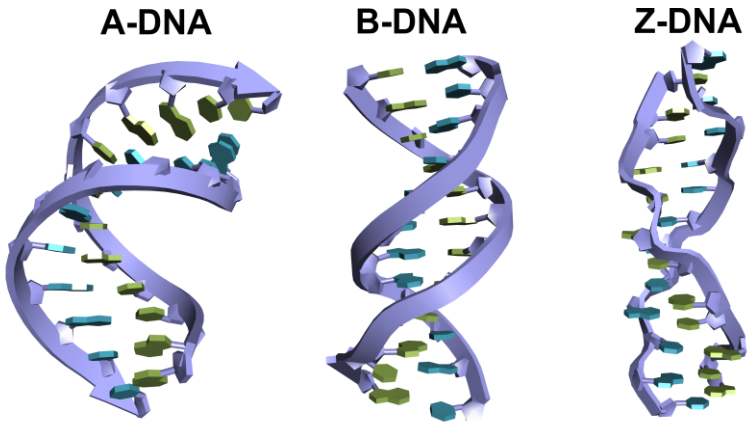


Fig. 3. Comparison of different forms of double-helical DNA. A-DNA in the left (PDB ID 1QPH), B-DNA in the middle (PDB ID 1BNA), Z-DNA in the right (PDB ID 4OCB)

The major difference between A-DNA and B-DNA is their grooves' overall shape and size. In A-DNA, the major groove is narrower and deeper as it is turned toward the interior of the helix and thus less accessible to the proteins. In contrast, in B-DNA, the major groove is wider and more accessible and provides more space for protein binding (**Fig. 3**).

2.1 DNA binding proteins

Several cellular processes require the interaction between DNA and proteins. This diverse class of proteins, which employ distinct binding motifs or domains, is collectively referred to as DNA binding proteins (DBPs).[6] Common examples of DBPs encompass DNA polymerases, helicases, transcription factors, DNA repair proteins, nucleases, and histones. The study of DBPs has a long and rich history. The first DBP discovered was the lac operon's repressor protein in *Escherichia coli*.[7] Since then, a multitude of DBPs have been discovered, significantly enhancing our comprehension of the fundamental principles governing DNA recognition by these proteins. DBPs can be classified as sequence-specific or non-sequence-specific based on their selectivity towards DNA.

2.1.1 Sequence-specific DBPs

Sequence-specific DBPs are essential in several cellular processes, such as DNA replication, recombination, repair, and transcription. Sequence-selective DBPs bind to a specific cognate DNA sequence, which means that the DNA stretch has to exhibit certain features and

functionalities that those DBPs can recognize. It is worth mentioning that although there are particular preferences between some amino acids and base pairs, a universal amino acid-base pair code does not exist.[8–10] Nevertheless, protein-DNA interactions are governed by similar forces involved in protein-protein interactions, e.g., van der Waals interactions and water-mediated hydrogen bonds.[11] The sum of these weak interactions makes a stable DNA-protein complex.

2.1.1.1 Direct Readout

Sequence-specific DNA recognition results from the protein's interaction with the major grooves of the DNA. The major groove is wider than the minor groove (**Fig. 4a**). Due to the width of the major groove, it exposes a longer stretch of DNA sequence and more functional groups such as H-bond donors, H-bonds acceptors, and other non-polar groups and imparts specificity (**Fig. 4b**). This binding mode can be classified as direct readout, as the DBPs interact directly with the exposed base pair functionalities.[1] Most transcription factors rely on direct readout to bring specificity. In 1976, it was proposed that hydrogen bonding between Protein and DNA base pairs could be used to differentiate base sequences and that specific amino acid residues recognize certain DNA bases through the major groove.[12] However, as stated earlier, no universal amino acid-base pair code exists.

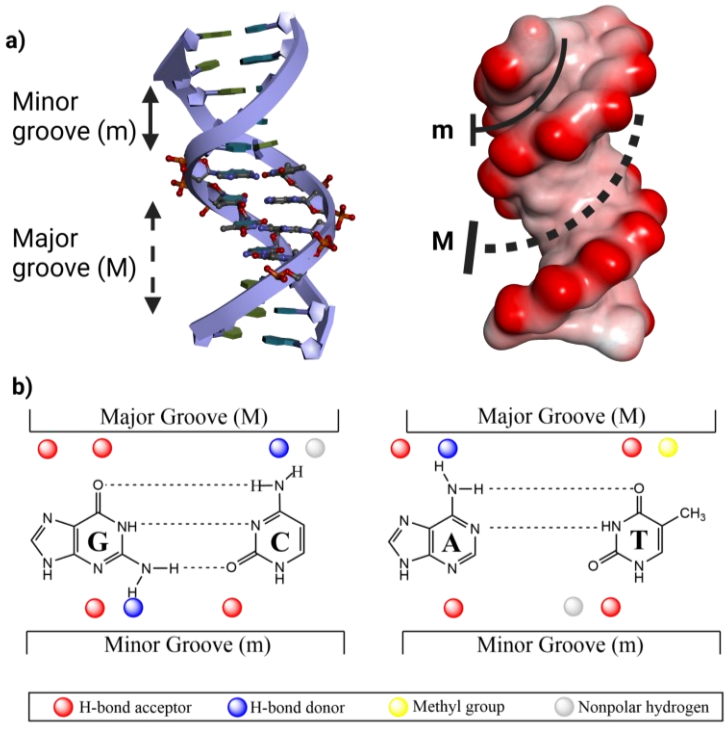


Fig. 4. Major groove and minor groove features of a double-stranded B-DNA. **a)** A 12-base pair double-stranded B-DNA with cartoon and surface representation to highlight the major groove (dashed line) and minor groove (solid line). **b)** Functional group presented by the major groove compared to minor groove on a GC and AT base pair.

However, compared to the major groove, the minor groove is more degenerate in terms of exposed functional groups and offers conformational changes that are then read by the DBPs and are often termed indirect readout. Thus, some proteins can bring sequence specificity without reading the base pair features by recognizing DNA structure, shape, and overall conformation. The indirect readout is often observed in DNA looping and chromatin organization.

2.1.1.2 Indirect Readout

As mentioned, certain proteins identify particular features of a DNA structure rather than just base pairs. Electrostatic interactions between protein and DNA significantly impact indirect readout, often termed intramolecular readout.[13] DBPs containing multiple positively charged amino acid residues, such as lysine and arginine, interact with the negatively charged DNA backbone. Water-mediated hydrogen bonding also plays a crucial role in the indirect readout of DNA features. Upon binding to DNA, some proteins undergo conformational changes that enhance their ability to recognize the DNA more effectively. In some cases, the shape of DNA is either bent or kinked to expose previously hidden binding features.[14]

2.1.2 Non-sequence specific DBPs

DBPs that bind to DNA with minimal or no sequence specificity are categorized as non-sequence-specific DBPs. Several biochemical processes, such as DNA packaging, maintenance, and regulation, depend on non-sequence-specific interactions between proteins and DNA. For example, it has been demonstrated that the chromosomal binding subfamily of the high mobility group (HMG) is capable of interacting with nucleosomes in a non-sequence-specific manner[15]. It is proposed to have a role in chromatin remodeling.[16] The crystal structure of HMG1 bound to cisplatin distorted DNA clearly shows a structure-specific interaction instead of a sequence-specific.[17]

Another example of a non-sequence-specific DBP is Sac7d, an archaeal chromosomal protein, which binds to the DNA's minor groove with no sequence specificity. The crystal structure shows that Sso7d, a Sac7d analog, interacts with the DNA phosphate backbone using lysine-mediated interactions.[18]

In yet another instance, the core octamer of histone has been shown to bind to 146 base pairs of DNA by recognizing structural features of the DNA. The DNA specificity that emerges arises from the inherent flexibility of the DNA sequence, which preferentially bends around the histone core.[19] It is important to note that despite a sequence-neutral binding in the above-

mentioned three examples, the affinity between DNA and proteins is achieved by a sum of all the involved interactions. In conclusion, non-sequence-specific DBPs primarily interact with the minor groove of DNA without forming sequence-specific hydrogen bonds.[20]

2.1.3 Classification of DBPs

DBPs can be classified based on their functional involvement in cellular processes, including DNA replication, repair, packaging, and transcription factors. However, this diverse class of DBPs consists of various DNA-binding domains or motifs that directly participate in DNA binding.[21] Based on the type of DNA-binding domain employed, DBPs can be further categorized into three families: those that bind to DNA via an alpha helix, a beta-sheet, or a combination of alpha helix and beta sheet (**Fig. 5**).

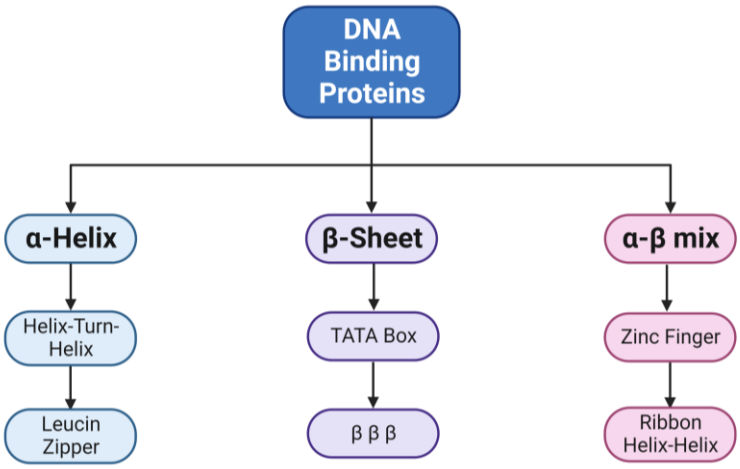


Fig. 5. Classification of DNA binding proteins based on the motif involved in DNA recognition.

2.1.3.1 Helix-Turn-Helix (HTH)

DBPs commonly feature the helix-turn-helix motif, which comprises two alpha helices joined by a short stretch of amino acids. The overall motif is composed of 20 amino acids. The first helix spans amino acids 1-7, while the second helix spans residues 12-20. The two helices are connected by a turn that bends at a 120-degree angle. The alpha helix of a DBP binds to the major groove of DNA. Helix-2 of the HTH motif is also called the DNA recognition helix that inserts into the major groove of DNA, creating specific contact with base pairs and the sugar-phosphate backbone. Helix 1 and, the turn, contribute to complex formation, but Helix-2 imparts specificity. HTH motifs are commonly found in prokaryotes and bacteriophages. Bacterial Trp repressor[14] and Bacteriophage lambda-repressor[22] were among the first DBPs characterized to have the HTH motif. HTH-containing transcription factors in bacteria and bacteriophages are generally homodimers (**Fig. 6a**).

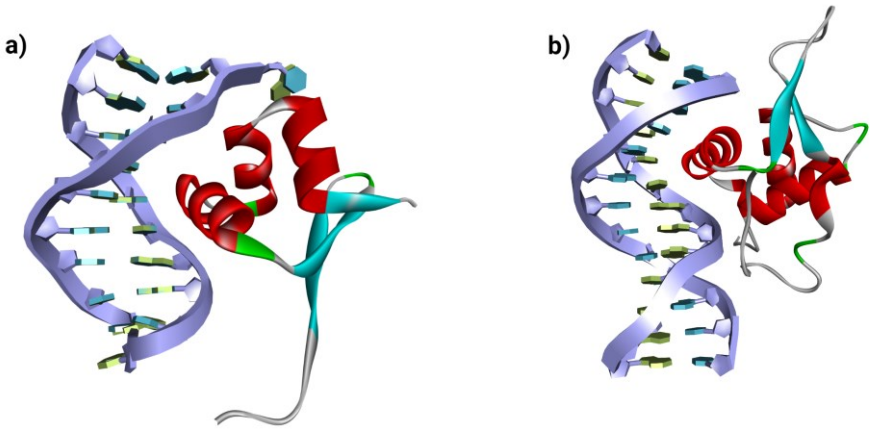


Fig. 6. Helix-turn-helix and winged-helix-turn-helix crystal structure. (a) The Trp-DNA complex shows a Helix-Turn-Helix motif binding the major groove of the DNA fragment (PDB ID 6CRO). (b) The HNF3-DNA complex shows the winged-helix-turn-helix. Notice the extra beta-sheet interaction with the minor groove and alpha helix interaction with the major groove of the DNA fragment. (PDB ID 1VTN)

An alternate form of the HTH motif is the winged helix-turn-helix (wHTH). In addition to the two helices, these motifs possess an antiparallel beta-sheet. Due to this extra beta-sheet, wHTH can interact with the minor groove of DNA (**Fig. 6b**).

2.1.3.2 Basic Leucine Zipper (bZIP)

In DBPs, bZIP is a frequently occurring DNA binding motif. It consists of two alpha helices connected by hydrophobic contacts between leucine residues at the C-terminus. These helices are arranged in a coiled-coil structure, and the leucine stretch creates a hydrophobic core that stabilizes the motif. bZIPs are thus dimers; their N-terminal domain, rich in basic residues, recognizes and binds to the DNA sequences (**Fig. 7**). A heptad repeat of leucine or isoleucine residues supports the zipper-like coiled-coil structure. Of the DNA binding domain, one coil makes contact with base pairs of DNA while the other coil interacts with the phosphodiester backbone.

Opposite from HTH, bZIP is only found in eukaryotes. Notably, the basic residues responsible for binding to DNA acquire a secondary structure only after binding to DNA. The DNA-dependent allosteric transition has been observed in numerous transcriptional regulators. The GCN4 protein is a typical example of a bZIP motif-containing protein.[23] The structure of GCN4 includes a leucine zipper at the C-terminus, forming a left-handed coiled-coil dimerization domain. This is followed by the basic region, which helps make specific contacts with the functional DNA groups in the major groove and the surrounding phosphodiester backbone.

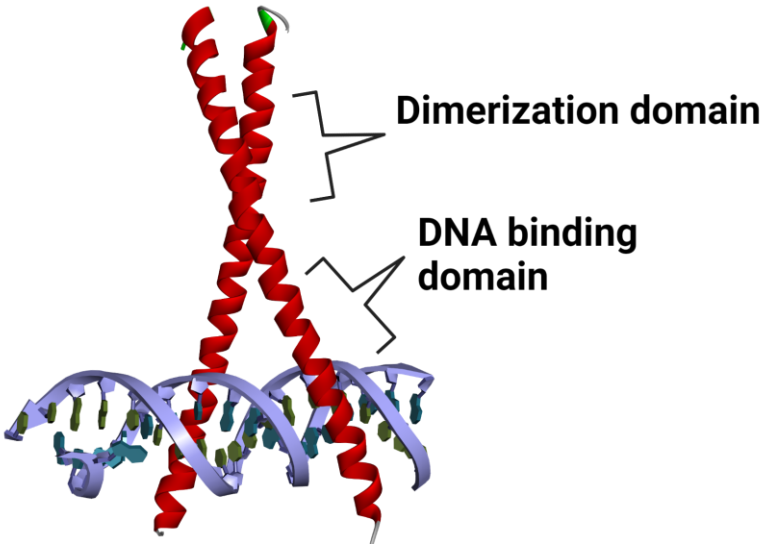


Fig. 7. A Basic leucine zipper motif. GCN4 protein bound to DNA segment (PDB ID 1YSA). Dimerization and DNA binding domains are marked individually.

Interestingly, GCN4 has been used to create designer proteins, and one such example is where Sac7d[24] was fused with GNC4 without losing its inherent function.[25]

2.1.3.3 Zinc Finger Motifs

Zinc finger motifs are another class of DNA binding motifs that occur most frequently in eukaryotes. It is worth noting that about 3% of the human genome codes for zinc finger domains. Zinc finger motifs are compact structural motifs made of 25 to 30 amino acids arranged in a loop-helix-loop structure that resembles a finger. The zinc ion is located in the palm of the finger, which stabilizes the loop-helix-loop structure, which is coordinated by two cysteine and two histidine residues (**Fig. 8**). The structure of each finger is comprised of an alpha helix and an antiparallel beta-sheet, coordinated by a zinc ion.

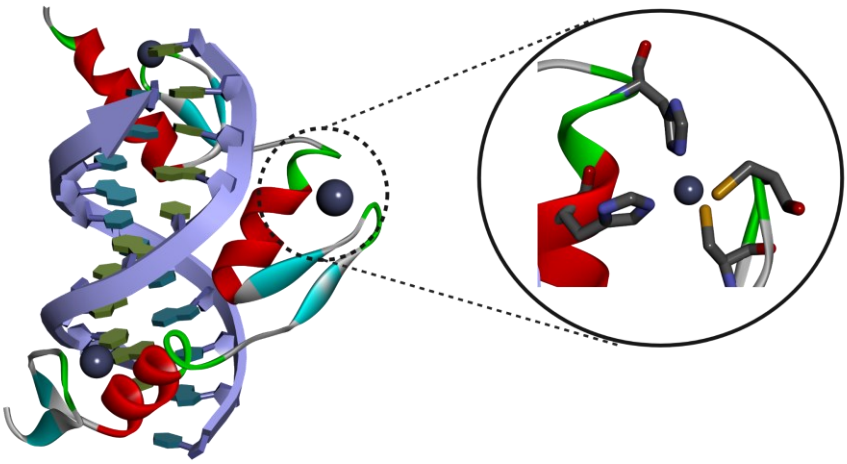


Fig. 8. Crystal structure of a ZIF2628 containing a zinc finger in complex with DNA (PDB ID 1ZAA) with a zoomed view at the zinc atom coordinated by two histidine and two cysteine residues. See also Fig. 67.

TFIIIA, a transcriptional factor from *Xenopus* oocytes, was the first zinc finger characterized.[26] TFIIIA contains between 3 to 15 copies of zinc finger motifs that impart sequence specificity against a vast stretch of DNA. The beta sheet makes backbone contacts, while the alpha helix of the zinc finger domain establishes base pair-specific contacts. Since multiple zinc fingers can recognize different DNA sequences, they are used to design and engineer proteins for therapeutic peptide development. Recently, ZFDesgin, an artificial intelligence model, was shown to design zinc finger motifs against any sequence of DNA, thus enabling transcription factor reprogramming.[27]

2.1.3.4 Beta-Ribbons

DNA recognition by beta ribbons is another commonly found motif found in prokaryotes and bacteriophage transcription factors. The TATA box binding protein (TBP) is one classical protein containing beta ribbons.[28] β-ribbons containing DBPs bind to DNA as a dimer. TBP binds to the minor groove of DNA by bending the TATA sequence by 80 degrees. Once the TATA sequence is distorted, kinked, and exposed, TBP binds to the TATA sequence. MET repressor from *E.coli* was the first DBP shown not to use an alpha helix but an antiparallel two-stranded beta ribbon motif.[29] The HU protein from prokaryotes also binds to DNA, utilizing the beta finger domain (**Fig. 9**).[30]

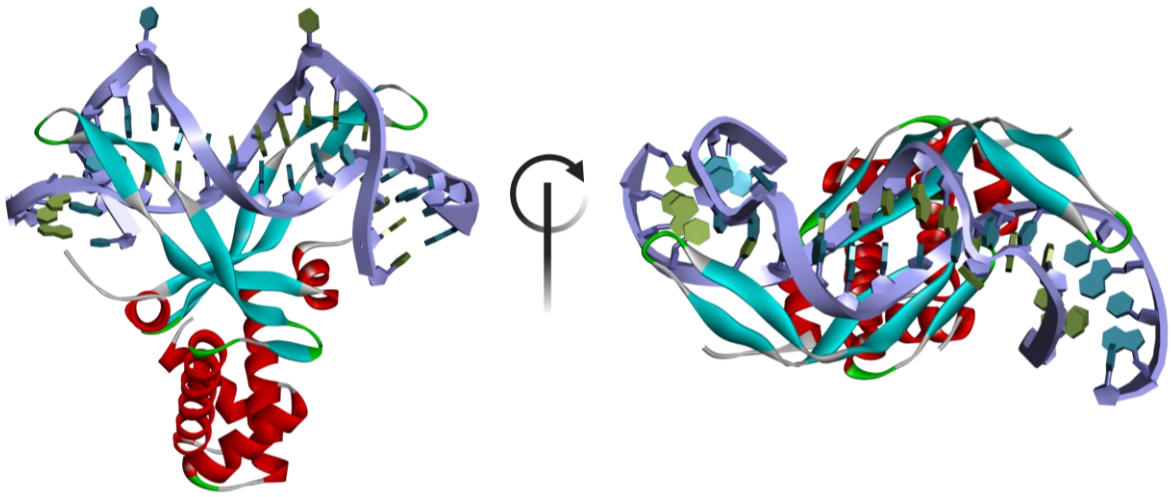


Fig. 9. Crystal structure of an Anabaena HU-DNA complex consisting of beta ribbon motifs (PDB ID 1P71).

2.1.3.5 High Mobility Group Domain (HMG Domain)

The family of high mobility group proteins contains the HMG domain, which is made of 80 to 90 amino acids characteristically arranged in three alpha helices that bind to minor grooves of the DNA. The HMG domain interacts with the bases and backbone of the DNA and can bend

the DNA upon binding to regulate the binding of other proteins (**Fig. 10**). The L-shaped fold of the three helices folds in a manner to accommodate a highly bent DNA and its minor groove.

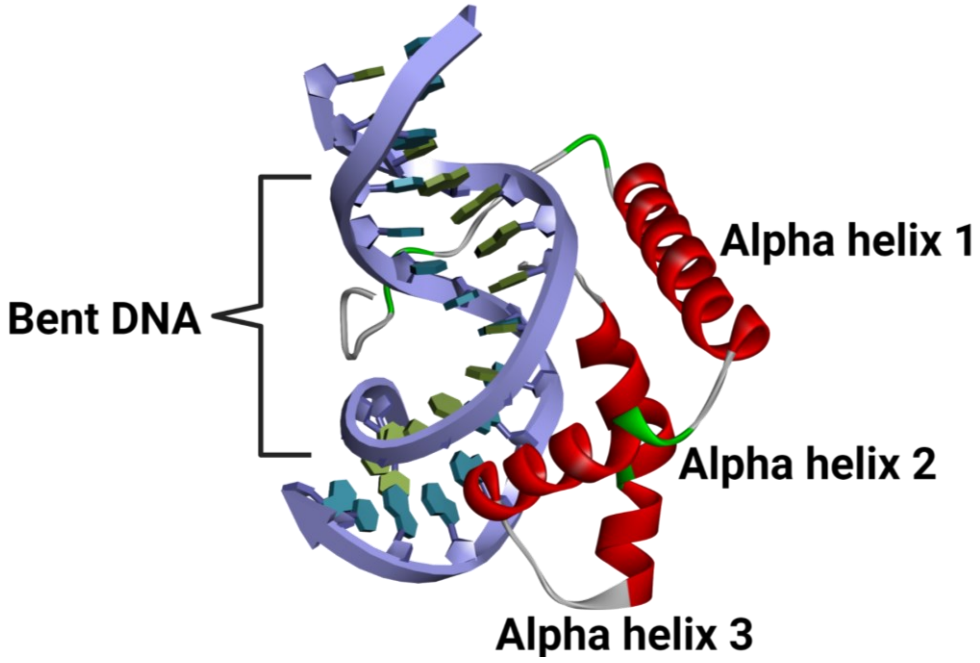


Fig. 10. NMR structure ensemble of LEF-1 protein bound with DNA (PDB ID 2LEF).

The HMG domain is found in the lymphoid enhancer factor (LEF-1) protein.[31] LEF-1 binds to T cell receptors and regulates the DNA expression, activating their transcription, which is essential for developing B and T cells.

2.1.3.6 Immunoglobulin Fold (Ig fold)

Immunoglobulin fold is a common structural motif in many proteins, including cell surface receptors and antibodies. The immunoglobulin fold comprises a beta-sandwich with two beta sheets and a connecting loop. This beta sheet usually contains 7-9 strands organized in a Greek key or jellyroll-like topology. Proteins consisting of the Ig fold use beta strands connecting loop as a critical factor for bringing sequence specificity while beta-sandwich presents the loop for DNA binding. The strands are kept stable by hydrogen bonding interactions. The P53 tumor suppressor protein is a classical immunoglobulin fold-containing protein (**Fig. 11**).[32]

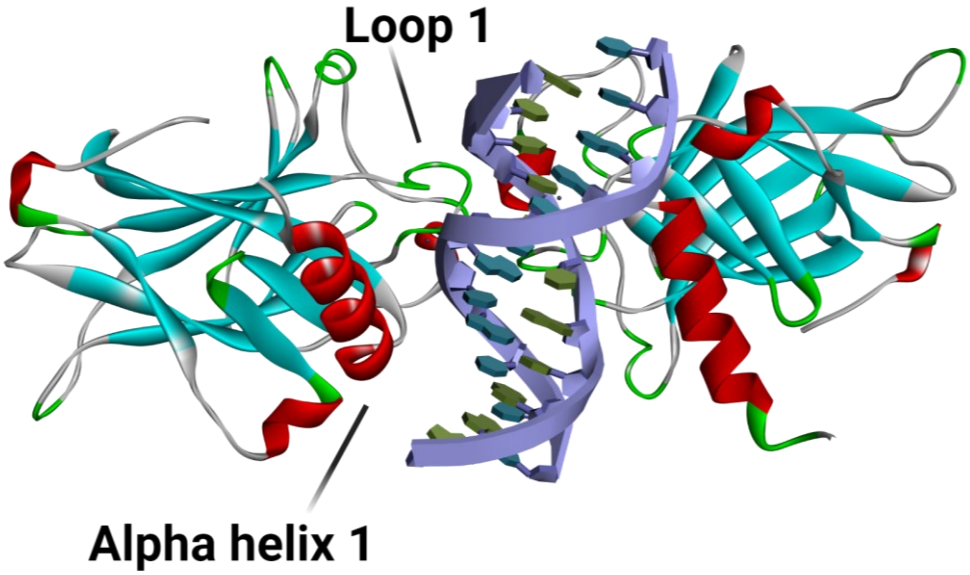


Fig. 11. Crystal structure of a P53 dimer containing immunoglobulin fold in complex with DNA (PDB ID 3EXJ).

2.2 Techniques to study DBPs-DNA complexes

Studying protein-DNA interactions is crucial to understanding the underlying mechanisms of various cellular processes, including replication, DNA damage repair, and transcription, among others. Techniques to study protein-DNA complexes have also been developed along with advancements in the discovery of new classes of DBPs. This section will discuss some key techniques commonly employed for investigating protein-DNA interactions.

Given the pivotal role of X-ray crystallography in the work presented in this thesis, a dedicated separate section (**section 2.3**) is allocated to describe the fundamental steps involved in using crystallography to study proteins and protein-ligand interactions.

2.2.1 Electrophoretic Mobility Shift Assay (EMSA)

EMSA, also known as gel shift assay, was among the first developed technologies in the 1960s to study protein-DNA interactions. EMSA involves the separation of protein-DNA complexes from free DNA using gel electrophoresis. Protein-DNA complexes would migrate slowly on gel compared to the free DNA. The current usage of EMSA is based on methods developed by M.M. Garner and A. Revzin in 1981.[33] In recent years, EMSA has been coupled with mass spectrometry to identify the binding of unknown proteins to DNA.[34,35]

2.2.2 DNA-footprinting

First introduced in the 1970s, it allowed the identification of DNA sequences bound to proteins. As the name suggests, this technique involves treating DNA with DNase I, which cleaves at the exposed regions of DNA contrary to uncleaved DNA, which remains bound to the protein.[36]

2.2.3 Chromatin Immunoprecipitation (ChIP)

Earlier versions of ChIP were developed in the 1980s to study protein-DNA in-vivo interactions.[37,38] ChIP involves cross-linking proteins to DNA in cells, followed by immunoprecipitation of proteins of interest with bound DNA sequences. ChIP has been used to study DNA sequences bound to transcription factors and histones, among other proteins. A range of comprehensive ChiP protocols have been devised to facilitate advancements in the realms of epigenomics and epigenetics.[39]

2.3 X-ray crystallography

With the advancement of other techniques, X-ray crystallography has also emerged as the technique of choice to provide structural information on the protein-DNA complexes. In 1958, John C. Kendrew solved the first protein structure, myoglobin,[40] utilizing X-ray Crystallography. This breakthrough marked the beginning of an era where macromolecular crystal structure elucidation gained significant traction, leading to over 208,000 structures being deposited in PDB[41] up to the present day. The process of structural elucidation using X-ray crystallography comprises multiple stages (**Fig. 12**), which are discussed in the following sub-sections.

2.3.1 Preparation of target Protein-ligand complexes

To begin with, it is essential to produce a substantial quantity of pure protein either by recombinant protein expression or from native organisms. Adjusting the target protein construct length or an alternate expression system could be utilized to increase the yields of the target protein.[42] The protein sample is then evaluated for stability, purity, and uniformity prior to crystallization.

2.3.2 Techniques of crystallization

Crystals of proteins and protein-ligand complexes can be grown by cautiously and gradually precipitating them from an aqueous solution. The first step involves dissolving the protein in an aqueous buffer containing a precipitant at a concentration lower than required to make the protein precipitate. Water is then evaporated in a controlled manner, leading to an increase in both protein and precipitant concentrations. This eventually results in precipitation or

crystallization. A slower precipitation process is more likely to yield bigger crystals, while a faster one leads to more precipitation. The crystallization process depends on various parameters, like protein concentration, pH, temperature, and ionic strength. Achieving the exact conditions for single crystal production usually takes a multitude of attempts and can take from several days to months.

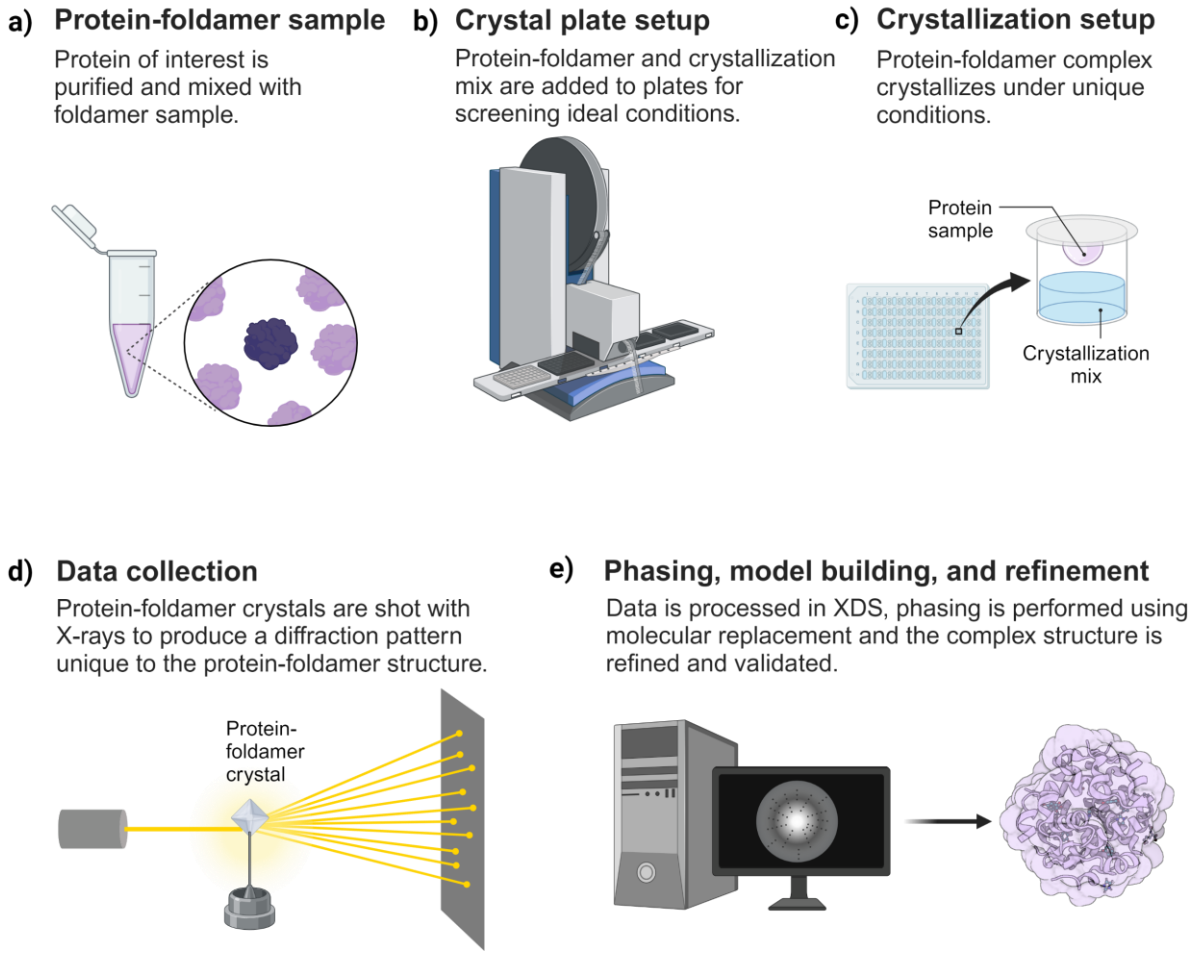


Fig. 12 General workflow of protein-ligand crystallography. Protein to Structure approach is shown in a stepwise workflow elucidating protein-ligand complex crystal structure.

2.3.3 Phase diagram and solubility curve

The crystallogenesis process can be grouped into three phases (**Fig. 13**): nucleation, crystal growth, and end of growth. To grow protein crystals, the solution must be supersaturated using a precipitant. At lower supersaturation, conditions are optimal for nucleation and crystal growth. A standard experiment involves reaching the nucleation zone, where critical nuclei are formed, then moving towards the metastable region, resulting in a few well-ordered crystals.[43]

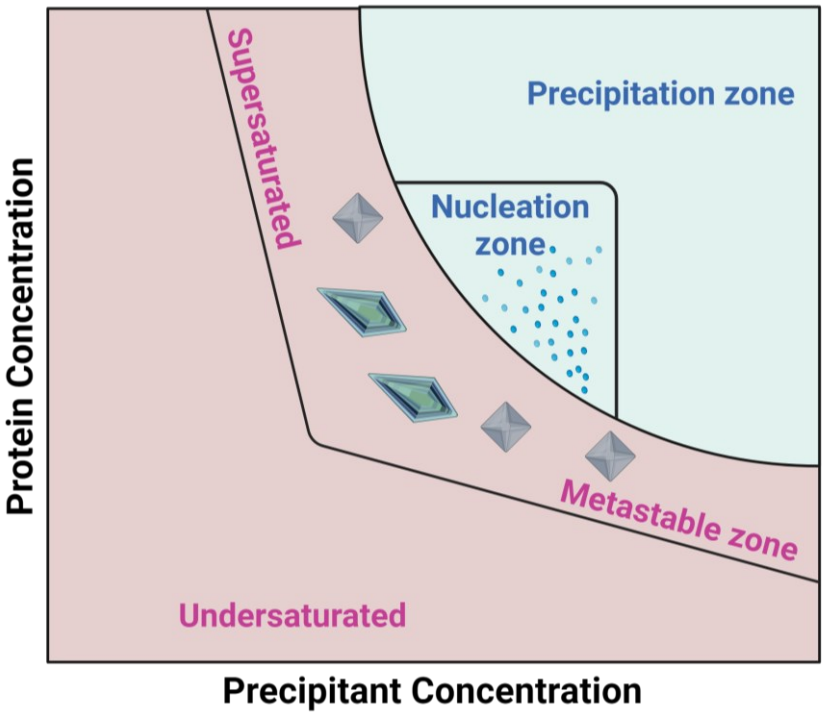


Fig. 13 A simplified phase diagram of a protein (ligand) crystallization experiment. Note the crystal growth in the metastable zone.

The phase diagram[44] in protein crystallography contains a solubility curve that distinguishes between undersaturated and supersaturated states based on the concentration of protein and precipitant at a fixed temperature (**Fig. 13**).[45] When the solution is in the supersaturated state, the solubility curve indicates the conditions that lead to phase separation, such as nucleation. Conversely, in the undersaturated state, the protein concentration remains below the saturation point without any phase separation. Essentially, the solubility curve illustrates the part of the phase diagram above saturation where the kinetic phenomenon governs the phase separation process.

2.3.4 Vapor diffusion

Proteins can be crystallized using the vapor diffusion method (hanging or sitting drop), which involves placing a droplet of protein and precipitant in a reservoir that contains the same precipitant but at a higher concentration. The system is then sealed, and as water vapor leaves the droplet and enters the reservoir, the reagent concentration becomes equal in both. This causes the protein to become supersaturated, which may eventually lead to crystallization. Other techniques, such as micro-batch, bulk and micro-dialysis, and free interface diffusion, are also utilized to produce high-quality crystals.

2.3.5 Data collection

Once crystals are formed, these crystals are generally vitrified using liquid nitrogen. They are then mounted on a goniometer head, maintained at a constant temperature of 100K using liquid nitrogen. Adjustments are made to screws to align the goniometer head with the X-ray beam. A single wavelength of an X-ray beam is then directed at the crystal, and a detector collects the diffraction data. The position of the detector and the exposure time are adjusted to ensure the best possible diffraction pattern is obtained. To obtain a complete dataset, crystals are oscillated through a small angle, with the amount of rotation required being inversely proportional to its symmetry.

2.3.6 Bragg's law

A crystal is composed of identical molecules arranged in a regular three-dimensional lattice formation. The unit cell is the fundamental, translational repeating unit that makes up the crystal. When X-rays with a particular wavelength are directed at a crystal, diffraction data images are obtained. The X-rays diffract off the crystal and form interferences. Bragg's law explains how X-rays are diffracted by a crystal lattice. It provides insights into the angles at which constructive interference occurs, leading to the formation of diffraction patterns.

$$\boldsymbol{n\lambda = 2d \sin(\theta)}$$

Here n is an integer of the diffraction peak, d is the distance between neighboring lattice planes in the crystal, θ is the angle of the incident, and λ is the wavelength of the diffracting X-ray. According to Bragg's law, when X-rays of a particular wavelength interact with a crystal lattice, they are scattered by the atoms in a way that causes constructive interference between the waves reflected from different lattice planes.[46] The angles at which these reflections occur are determined by the geometry of the crystal lattice and the wavelength of the radiation. If the path difference between the waves reflected from two adjacent lattice planes is an integer multiple of the wavelength, then the waves reinforce each other, resulting in a strong diffraction peak. If the path difference is not a multiple of the wavelength, destructive interference occurs, leading to a weak or absent diffraction signal. Thus, to obtain a complete data set, the crystal must be rotated in various ways to bring all sets of planes into reflection.

2.3.7 Data processing

The intensity of spots and the structure factor of a specific protein crystal are determined by processing its data set, which involves multiple steps. To begin with, the program should be used to locate the diffraction spots on the images. Once these spots are found, the autoindexing

process maps them onto reciprocal space. This automated process assigns the correct hkl to the reflection datasets, determines the crystal orientation, and estimates the unit cell parameters of a crystal, indicating crystallographic symmetry. The fast Fourier transform method is then used to identify the real space vectors, which are used to determine the best possible lattice arrangement among the 14 predefined Bravais lattices.[47] Determining the Bravais lattice involves auto-indexing, where a primitive triclinic cell is identified, and the best possible cells are determined using matrix transformation based on rules from the International Tables of Crystallography. The output is provided with a normalized quality index.

To eliminate any errors in diffraction data, it is essential to scale the intensity data to a standard scale. This includes adjustments for changes in X-ray intensity, anisotropic absorption of X-ray, detector response errors, and radiation damage. The scaling process involves comparing the spot intensities of symmetry-related reflections, merging them, and creating a set of unique reflections that are not related by symmetry operations. To merge the reflections, two-dimensional integration is used, as partial data collected on successive images must be combined into a single measurement for structure solution during phasing in the next step.

2.3.8 Phase problem

In X-ray crystallography, while it is possible to measure the amplitude (intensities) of the structure factor through experimental methods, the phase information is unfortunately missing and is known as the "phase problem."[48] The phases contain the crucial angles that dictate X-ray interference and generate the diffraction pattern. Without the phase information, it is impossible to recreate the electron density within the unit cell. Overcoming this problem has driven the development of innovative techniques, such as multiple isomorphous replacement, anomalous scattering, and advanced algorithms, enabling researchers to reconstruct molecular structures with high precision. In high-resolution crystallographic data, such as atomic resolution, the phase problem can be mathematically resolved using direct methods of solving crystal structure. However, when dealing with protein crystals that do not diffract to atomic resolution, as discussed earlier, indirect methods become essential for tackling the phase problem.

2.3.9 Molecular replacement

The phase information of an unknown protein can be determined by analyzing the phase information of a homologous protein structure or model.[49] The primary sequence of a protein contains information on its three-dimensional structure, making it probable that other proteins with homology or identity in amino acid sequence can also exhibit a similar three-dimensional

structure. The symmetry of the target protein can be used as a guide for placing the model, but this alone is not enough for phase estimation. To find the most suitable placement, every potential arrangement and orientation of the model within the new unit cell is examined, and the corresponding structure factor of the model is computed. The model's amplitudes are subsequently compared to the measured amplitudes derived from the diffraction intensities of the protein being studied. The initial phase is ascertained by selecting the orientation and position that most accurately corresponds to the computed phases. The Patterson map can give valuable insights into the orientation of a model within the unit cell of a target protein. If the model and target match orientation and symmetry in their unit cell dimensions, they will produce a similar Patterson map. MOLREP[50] and Phaser software[51] best serve molecular-replacement efforts due to their automation and user-friendliness. It is worth emphasizing that recent breakthroughs in the application of artificial intelligence and deep learning tools, such as AlphaFold, have made remarkable strides in advancing the field of protein structure prediction. AlphaFold's models have been instrumental in overcoming the challenges associated with solving protein structures.[52] Notably, these models have proven their efficacy by successfully facilitating molecular replacement when crystallographic structures for a homologous protein are unavailable.

2.3.10 Refinement

Experimental sources provide phase estimates, which are used to calculate an initial electron density map. Utilizing chemical knowledge of the protein and ligand can improve the map and enhance the phases. Due to the protein molecule's irregular shape, there are voids in the crystal that are filled with disordered solvents. The solvent can be smoothed to improve the electron density map, as the density in these areas lacks recognizable features. On the other hand, the side chain and other features of protein and ligands are fitted in the electron density. This cycle continues until the correlation between diffraction data and the generated model reaches its maximum, measured by the R-factor. One approach for cross-verification is to reserve 5-10% of diffraction data solely for computing R-free, which is not utilized in the refinement process.[53]

2.3.11 Validation and deposition

After ensuring that the R-factors, bond angles, lengths, and other geometric parameters align with the data, the protein structure is deposited into the Protein Data Bank. Over 208,000 structures thus far have been deposited in the Protein Data Bank. X-ray crystallography studies

of transcription factors have helped scientists understand how proteins recognize and bind to specific DNA sequences, thus designing drug candidates to target these DBPs.[54]

Nuclear Magnetic Resonance,[55] Surface Plasmon Resonance,[56] Fluorescence Resonance Energy Transfer,[57] and computational approaches are some other techniques that have also been extensively used to provide further insights into protein-DNA interactions.

2.4 Importance of DBPs

In the latter part of the 20th century, researchers in the field of molecular biology recognized the vital role that the association between protein and DNA plays in the functioning of living cells. Protein-DNA interactions have been a subject of significant scientific inquiry for several decades as experts endeavor to uncover the intricate and fundamental interactions that transpire when proteins and DNA bind together. Even before the DNA molecular structure was published, Stedman and Stedman 1950 identified histones as potential regulators of DNA activity.[58] Since then, the field of protein-DNA interaction has only gained attraction. We now know these interactions are crucial in cellular mechanisms such as transcription, transcription regulation, replication, DNA repair, packaging, and viral infection.[59] Transcription factors are essential DBPs that bind to DNA to regulate transcription.[60] Once bound to DNA, transcription factors promote or block RNA polymerase recruitment, thus regulating gene expression.

2.5 Targeting protein-DNA interactions

Thus far, we have undertaken proteins interacting with DNA, what classes of binding domains exist, and a set of techniques to study these interactions. Due to the great importance of protein-DNA interactions in many cellular processes, targeting these interactions has been a long-standing effort toward drug discovery, gene therapy, and biotechnology advancements. By manipulating these interactions, it is possible to regulate gene expression, treat genetic disorders, and create new therapeutic interventions. The following section will discuss the strategies used to target the protein-DNA interactions: fragments, small molecule inhibitors, antisense oligonucleotides, RNAi, and protein-protein interaction inhibitors (**Fig. 14**).

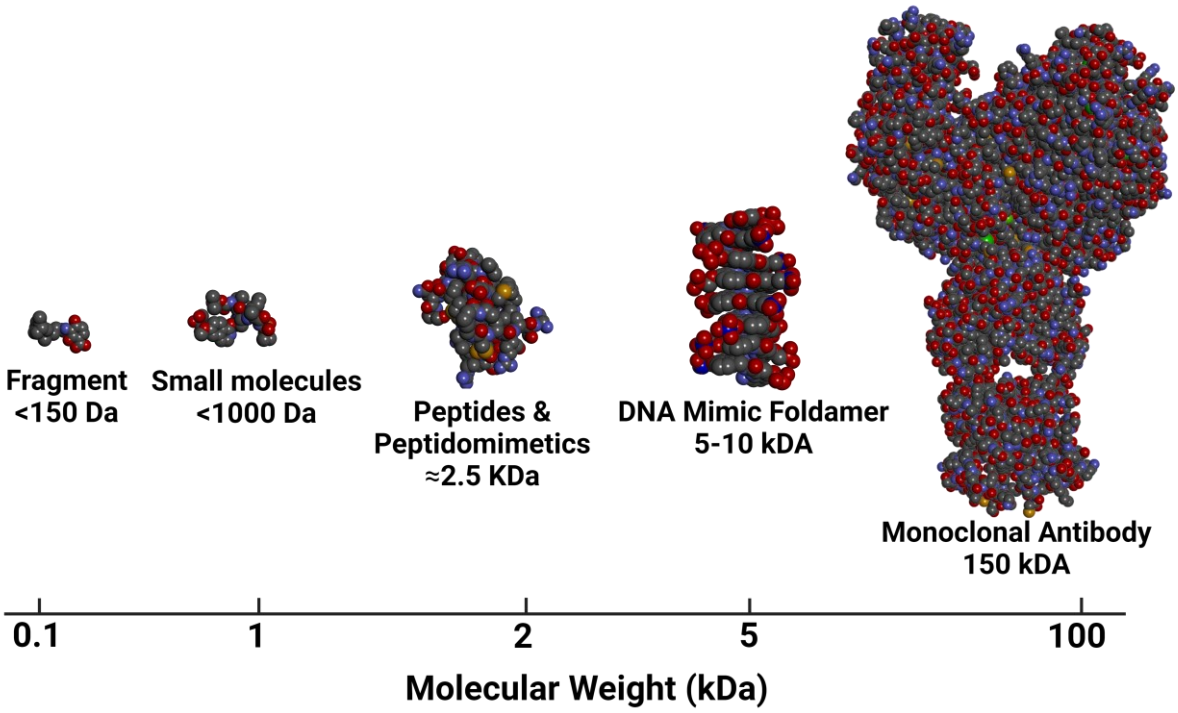


Fig. 14. Relative size of the molecules that target proteins.

2.5.1 Small molecule inhibitors

Transcription factors constitute one of the most essential classes of DBPs in eukaryotes.[61] Of the roughly 1600 transcription factors present in the human genome, about 19% have been associated with a phenotype related to a disease.[61] Transcription factors have been the target of therapeutic development in the context of anticancer drug development.[62] The deactivation of tumor suppressor transcription factors allows for the avoidance of apoptosis and promotes cancer cell survival.[63] FK506, sold under the brand name Prograf, is a 23-membered macrolide lactone that prevents T-cell proliferation by inhibiting a Ca^{2+} pathway required for interleukin-2 transcription by inhibiting NF-κB transcription factor.[64] FK506 is used to treat specific autoimmune diseases[65] and to prevent organ rejection.[66] Rapamycin is a macrolide compound commonly used to prevent rejection during organ transplants. Rapamycin blocks the mTOR transcription factor,[67] suppressing T cells by decreasing their responsiveness to interleukin-2.[68] Erlotinib[69], Actinomycin D,[70] and α-Amanitin[71] are some small molecule inhibitors, among others, that act either on transcription factors or DBPs.

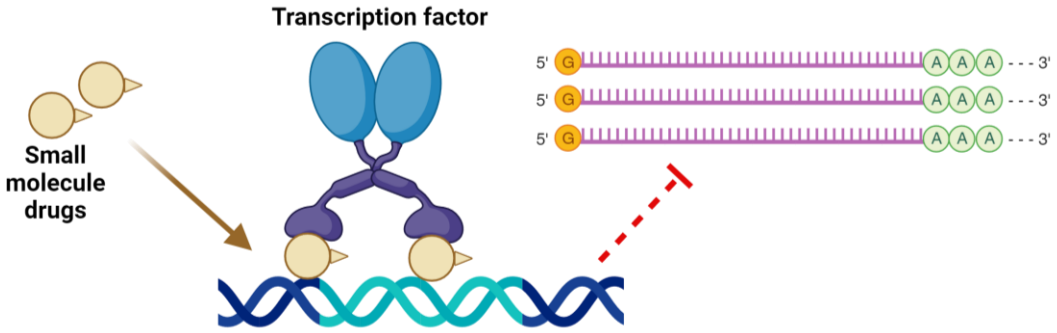


Fig. 15. Small molecule drugs inhibition mechanism rendition on transcription factor binding on DNA.

Despite being an attractive drug target, transcription factor has hurdles that impede the countless efforts at drugging them. A significant hurdle is that most transcription factors are intrinsically disordered and lack well-defined pockets for binding small molecules.[72,73] The reason behind their disorders is that transcription factors are involved in highly dynamic protein-DNA interaction. Thus, small-molecule inhibition of protein-DNA interactions has received less attention than protein-protein interactions. Developing compounds that can bind specific DNA sequences in the major groove may be highly beneficial, as this is where most transcription factors bind (**Fig. 15**).[74] Due to their dynamic nature, small molecule pockets are poorly defined and need other strategies to target these proteins.

2.5.2 Peptide-based therapies

With the rapid growth of bioinformatic tools and knowledge related to proteins, amino acid-based strategies to develop therapeutics have gained advances. Since insulin [75], more than 100 peptide structure-based therapies have been used worldwide.[76] Peptide-drug conjugates (**Fig. 16**), like antibody-drug conjugates, can be used as therapeutics, where peptide targets the ligand and conjugated drug brings cytotoxicity.[77] Compared to antibody conjugates, peptide conjugates can form nanoparticles, thus showing better retention and permeability at the tumor site.[78,79] Moreover, peptide-drug conjugates can also deliver significantly more drugs to tumors than antibodies and have emerged recently as novel anticancer drug candidates.[80–83]

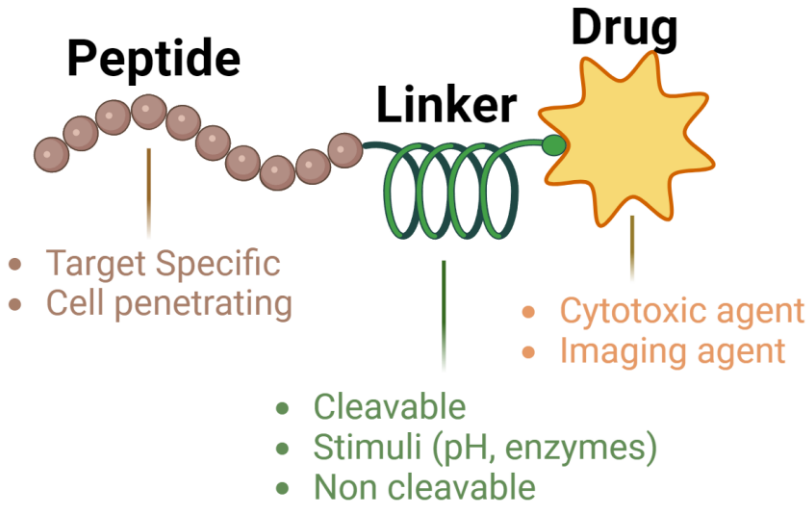


Fig. 16. Peptide-based modalities connected to a drug candidate.

Like protein-DNA interactions, protein-protein interactions are relevant in transcription factor regulation. P53, a transcription factor considered the "guardian of the genome," binds to E3 ligase MDM2 and thus gets downregulated in tumor cells.[84]. Like other transcription factors, P53 is highly disordered[85] and thus is difficult to be targeted by small molecules. An alternative approach is to develop an antagonist against MDM2 as the binding site of P53 as MDM2 is well defined. This understanding of the MDM2-P53 complex has enabled the development of several peptide antagonists of MDM2.[86,87] Peptide macrocycles offer additional options for utilizing peptide-based modalities to target protein-DNA interactions. Peptide macrocycles could be made from non-proteinogenic residues, significantly improving their metabolic stability. David Baker and colleagues, in 2022, reported the *de-novo* design of macrocycles containing non-canonical backbones that could permeate the membrane and are biologically active.[88] Peptide-foldamer macrocycles are other peptide-based modalities where a peptide and foldamer (a subject dealt with in detail in section 2.6) hybrid macrocycle is designed to where the foldamer helps in stabilization of peptide through stapling and contact with key residue with the peptide.[89] Despite the advances, peptide-based therapies face challenges in oral availability, immunogenicity, metabolic stability, and renal clearance.[90] In the future, there will hopefully be more development on potent peptide-based therapies.

2.5.3 Molecular glue as therapeutics

Another alternative to control transcription factor stability is molecular glues, which facilitate a non-native protein-protein interaction. Recently, it was found out that clinically approved thalidomide-based antitumor immunomodulatory imide drugs work by mediating the

interaction between Ikaros zinc-finger, a transcription factor, and the E3 ligase CRBN, which leads to degradation of the transcription factor by ubiquitin proteases.[91–94] Molecular glue for P53 transcription factor have been shown to covalently link P53 and the E3 ligase UBR7[95] which interestingly results in activation of P53 instead of degradation.

2.5.4 PROTACs

The discovery of molecular glues has coincided with the development of molecules that bind and degrade the target protein. Proteolysis targeting chimeras' technology offers a rational approach to designing such molecular degraders (**Fig. 17a**).[96,97] PROTACs work by mediating an interaction between E3 ligase and the target protein, which is then degraded by the ubiquitin-proteasome system (**Fig. 17b**). Compared to antagonists of a single domain of a transcription factor, PROTACs inhibit the activity of the transcription factor target altogether.[98–101] After the first example of PROTACs was described in 2001,[96] more and more PROTAC modalities have emerged in targeting MDM2, STAT3, BTK, and BRD9 proteins. By 2022, more than 130 proteins could be potentially targeted using PROTAC modalities, suggesting the PROTAC era has arrived.[102] PROTACs are being used to target transcription factors, with the first clinical trials focusing on androgen and estrogen nuclear receptors.[103] Although PROTAC molecules are currently used in clinical trials, several problems must be addressed. These include the design of PROTAC molecules and their ability to be effectively transported to target sites. This involves the use of new E3 ligase ligands and target protein ligands.[102] Efforts to expand the range of E3 ligases for use in the PROTAC toolset are ongoing.

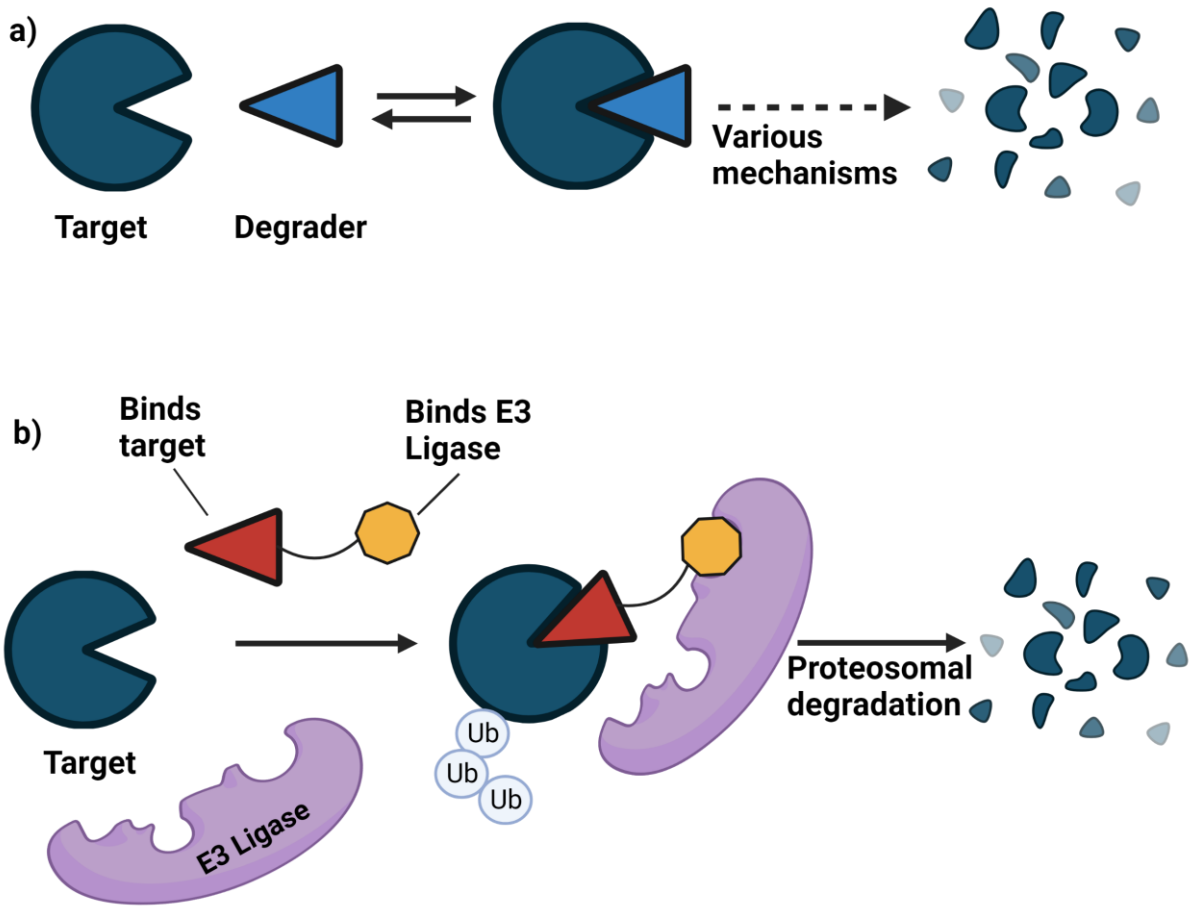


Fig. 17. Targeting DNA binding proteins using molecular degraders and PROTACS. a) Monomeric degrader binding to target DNA binding protein leads to its degradation with various mechanisms. b) PROTAC degraders work by linking a binder of the target protein (red) to an E3 ligase ligand (yellow) that leads to the degradation of the target DNA binding protein.

RNA interference (RNAi) is another strategy involved in transcriptional gene silencing, which prevents the gene from being transcribed into mRNA in the first place.[104] Chemical probes complementary to CRISPR-based approaches[105], minor groove-binding agents[106], and the use of Antisense oligonucleotides[107] are among other emerging technologies to target DBPs either by interfering with protein-protein or protein-DNA interactions.

Targeting transcription factors with drugs was previously deemed impossible. However, recent developments have proven otherwise. Many transcription factors can be effectively targeted with the proper tools and persistent effort. Promising strategies are currently in development, offering exciting opportunities for drug discovery in previously untapped areas of the transcription factor target class. However, despite numerous successes in developing transcription factor modulators in the last three decades, several obstacles remain unaddressed in targeting more DBPs. There is a dire need for alternative approaches to target protein-DNA interactions.

2.5.5 DNA mimic proteins

At present, only a limited number of proteins are known to utilize DNA mimicry to regulate the DBPs directly.[108,109] These proteins that mimic the DNA are referred to as "*DNA mimic proteins*" (DMPs). DMPs employ negatively charged amino acids to mimic the negatively charged phosphate backbone of DNA.[109] Most DMPs have a relatively high proportion of negatively charged amino acids, such as glutamic and aspartic acids, comprising around 15-20 % of the total amino acid content. Aspartic and glutamic acid abundance enables the protein to imitate DNA's negatively charged phosphate backbone (**Fig. 18**). Additionally, most DMPs have molecular weights below 25 kDa, conferring a specific DNA-like conformation. Another characteristic is the DNA-like distribution of negative charges on the surface of the DMP.[110] These features are crucial for the DMP to effectively compete for the nucleic acid-binding domain of its target protein. For example, AcrF2, a DMP, halts the CRISPR-Cas system in bacteria by interfering with protein-DNA interactions.[111–113]

Although the number of identified DNA mimics is still limited, their significance, in addition to biological functions, has been well-established in the realm of molecular biotechnology. Notably, the DMP Ocr has demonstrated its ability to enhance the transformation efficiency of unmodified DNA in bacteria by inhibiting the restriction and modification system.[114] Transcription factor P53, involved in tumor suppression, interacts with several DBPs, including replication protein A (RPA) that binds to single-stranded DNA.[115] Structural analysis reveals that the P53TAD, a region of P53, can mimic single-stranded DNA and compete with RPA for binding to DNA. This interaction may help P53 to control the activity of RPA.[110] Another transcription factor, TFIID, regulates the activity of other DBPS by mimicking the DNA. The N-terminal domain of TAFII230 (residues 11-77) binds to TATA-binding protein (TBP) by mimicking the minor groove surface of the TATA box.[89] This interaction likely contributes to the transcriptional control of genes downstream of the TATA box, as gene expression is suppressed when TBP binds to the TATA box. These noteworthy examples underline the critical role played by DNA mimicry in living cells and how they can be harnessed for therapeutic interventions. Despite the progress made in bioinformatics and proteomics, only around 20 DMPs have been identified, leaving many more DMPs and their mechanisms yet to be discovered.[116]

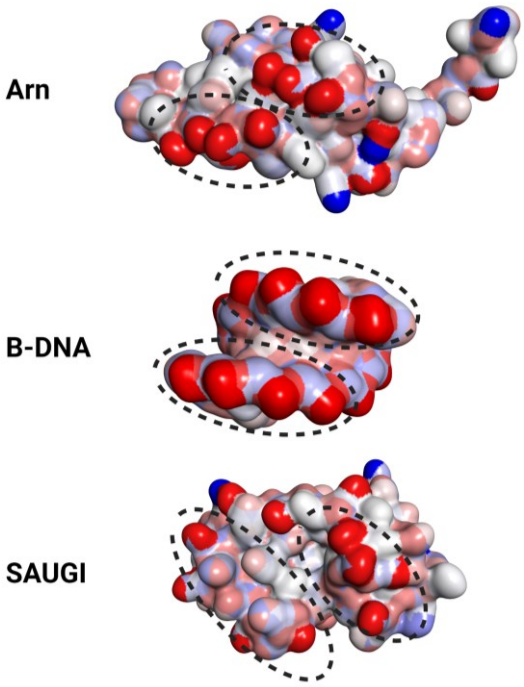


Fig. 18. Surface charge distribution of an 8BP B-DNA (PDB ID 6BEK) and two DNA binding proteins, Arn (PDB ID 3WX4) and SAUGI (PDB ID 3WDG). The negative charge distribution is highlighted with dashed circles.

Taking inspiration from naturally occurring DMPs, Yuksel and colleagues have described the de novo design of proteins mimicking DNA by utilizing natural amino acids and coiled-coil architectures that could be utilized in targeting systems such as DNA repair and replication.[117] In 2018, taking inspiration from DNA base pairing, David Baker and colleagues published an algorithm to design protein heterodimers driven by hydrogen bonding.[118] These studies indicate that there is still room for exploration in the realm of programmable biomolecule interactions. By designing additional artificial molecules that mimic DNA, protein-DNA interactions can be targeted innovatively, adding to the list of already available strategies.

2.5.6 Peptide nucleic acids

The convergence of peptide and nucleic acid chemistry has yielded a distinctive class of molecules called Peptide Nucleic Acids (PNAs). PNAs are synthetic polymers that mimic the structure of RNA or DNA, featuring a backbone composed of N-(2-aminoethyl) glycine units. PNAs were first introduced in 1991 by Nielsen and colleagues, where the sugar-phosphate backbone was replaced by N-(2-aminoethyl)glycine repeating units (**Fig. 19**).[119] Nucleobases in a PNA are attached to the polyamide chain via a carboxymethyl spacer. Thus, imparting structural hybridity enables them to pair the Watson-Crick base with complementary DNA and RNA strands. The lack of charged phosphate groups in the PNA backbone results in a more

robust binding between PNA/DNA strands compared to DNA/DNA duplex, as no electrostatic repulsion is present.

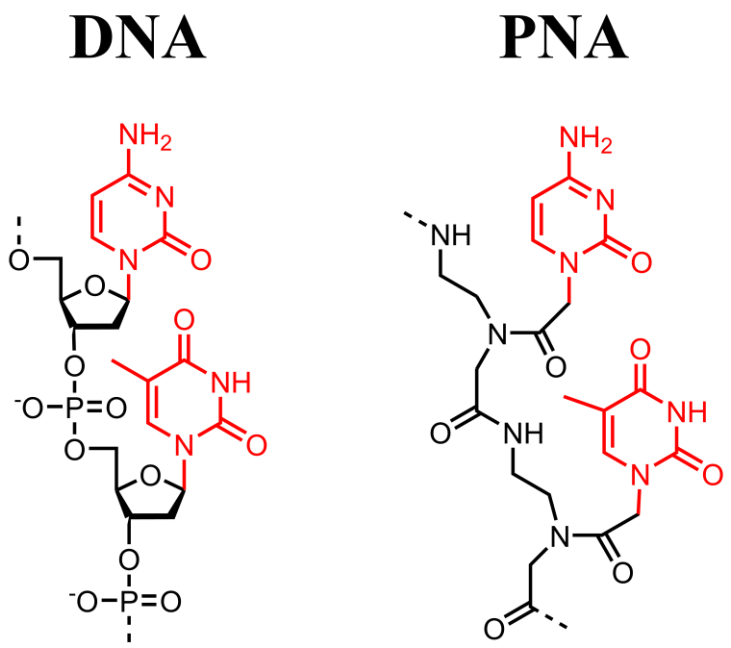


Fig. 19. Chemical difference between a DNA chain versus a PNA chain.

Because of the hybrid nature of PNAs, they are resistant to enzymatic degradation by proteases and nucleases. These atypical molecular features grant PNAs a pronounced edge over conventional DNA and RNA oligonucleotides. PNAs were originally designed to form duplexes to alter DNA or RNA functions. However, they have found long-standing applications in various research fields. For example, due to their longer lifetime, PNA has been used in vivo and in vitro applications such as molecular diagnostics.[120] A compelling avenue of inquiry within the realm of PNAs is their adeptness in DNA mimicry[121], characterized by their capacity to imitate the functions of DNA within biological systems. The affinity and recognition mechanisms between PNA and DNA stem from forming stable duplex structures facilitated by hydrogen bonding interactions with complementary nucleotide sequences. PNAs have been found to possess a greater efficacy in inhibiting reverse transcription when compared to phosphorothioate oligonucleotides.[122] This unique ability has sparked a growing interest in leveraging PNAs as agents for antisense and anti-gene drug therapies.[123] Moreover, PNAs have been engineered as biomolecular tools, molecular probes, and biosensors.[124] These advancements have opened up new avenues for exploring the potential applications of PNAs in molecular biology. However, using PNAs as therapeutics faces a long road ahead due to the lipid-based carriers currently in use. These carriers cause PNA accumulation within cells,

resulting in unsatisfactory cytosolic distribution. This limitation severely hinders the extensive use of PNAs in medicine and calls for other novel DNA mimic analogs.

2.5.7 Decoy nucleotides

The concept of decoy nucleotides is based on the idea of introducing synthetic nucleotide sequences that mimic the binding sites of specific regulatory elements, such as transcription factors or other DNA- or RNA-binding proteins. These synthetic sequences act as competitive inhibitors, selectively sequestering regulatory molecules and preventing them from interacting with their native target DNA or RNA. In 2019, Karni and colleagues showed the design and application of decoy RNA oligonucleotides targeting the splicing factors, thus inhibiting their biological activity.[125] The field of decoy nucleotides is moving towards development of therapeutic modalities targeting cancer. For example, a 2012 study by Fagard and colleagues demonstrated the use of hairpin decoy oligodeoxynucleotides that binds to STAT3 and induces apoptosis in colon cancer cell line.[126] These example furthers the idea of using nucleic acid mimics towards therapeutics targeting protein-nucleic acid interactions.

2.5.8 Applications of nucleic acid mimics

With the progress in technologies like polymerase chain reaction and nucleic acid chemistry, the development of nucleic acid mimics has also grown. Currently, the focus is on developing nucleic acid analogs that can overcome the limitations of natural nucleic acids and become powerful tools for biotechnology and diagnostics. Peptide nucleic acids and locked nucleic acids, among others, are nucleic acid mimics used in molecular diagnosis. PNAs are used in gene therapeutics as they have a higher affinity for matching mRNA. This can effectively hinder translation and remain stable within cells for longer due to their resistance against enzymatic degradation. Fluorescent PNAs were recently introduced and used in fluorescence in situ hybridization to gain brighter microscopy experiments.[127] Previous studies have demonstrated that PNA microarrays are more effective than DNA microarrays in detecting single and multiple base mismatches.[128] Singh and colleagues have shown that using PNA in nano biosensor arrays can increase sensitivity and effectiveness compared to DNA-based nano biosensors.[129] PNAs have been utilized for in-vivo imaging to identify the hybridization of KRAS2 mRNA in lung cancer cells through PNA fluorescence.[130] However, due to the hindrance in intercellular delivery of PNAs, their therapeutic use has been severely stalled. In summary, there is a massive potential for utilizing nucleic acid analogs in creating treatments, diagnostic tools, and assays for molecular biology research. Several research teams are striving to produce new synthetic DNA analogs that can overcome these limitations of PNAs and natural DMPs.

2.6 Foldamers, going beyond biopolymers

The complex structure of biomolecules, such as proteins and nucleic acids, has fascinated scientists for a long time. These natural biomolecules are involved in various cellular processes such as cell signaling, cell division, apoptosis, and catalysis. These functions, primarily by proteins, are achieved by acquiring a specific native fold that is thermodynamically and kinetically stable. Folding of proteins to their native state yield "active sites" that is used in the binding and recognition of target molecules. However, these natural biomolecules have certain limitations regarding design flexibility and specific functionalities. For instance, DNA can only be made using four repeating nitrogen base pairs (**Fig. 2**). To overcome these limitations, synthetic chemists have explored artificial backbones that can fold into well-defined structures based on the underlying principles governing folding in biomolecules. These synthetic molecules are known as *foldamers*. In other words, foldamers could be defined as synthetic oligomers that could adopt a well-defined conformation in solution. Foldamers offer a more comprehensive range of chemical functionalities than biomolecules. e.g., proteins are made of a limited set of 20 amino acids. The demonstration that folding is not limited only to biopolymers but molecules with artificial backbones could also fold and adapt desired conformations has paved the path for designing molecules that could fold and adopt specific conformations different from the biopolymer they were initially inspired from. Thus, new functionalities could be achieved by utilizing backbones different from biomolecules. The non-natural backbones in foldamer are an essential feature of the function they can perform because chemical functions are imparted by the nature of the backbone; for example, a protein cannot store the genetic information as DNA or a DNA cannot catalyze a reaction as does a protein. However, foldamers can be made with endless rational use of different monomers as building blocks. They can be classified based on the linkage between monomers such as amide, urea, or alkyne groups. Another way of classifying foldamers is to differentiate the backbone units themselves, which can be aromatic, a combination of aromatic and aliphatic parts, or purely aliphatic.

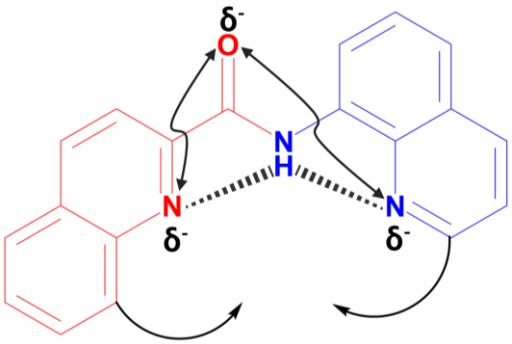


Fig. 20. Folding principle of aza-aromatic oligoamides. Hydrogen bonds are shown in dashed lines. Dipole repulsions are shown with two-headed arrows.

Taking inspiration from naturally occurring α-amino acids, β-peptides were designed and reported to be held together by H-bonding between the backbone amide functions (**Fig. 21a**).[131] However, unlike α-amino acid-containing peptides, β-peptides were found to be resistant to pepsin. As they remained close to their natural counterparts, these structures were termed *biotic foldamers*. However, scientists have ventured into using diverse, abiotic backbones to achieve more unique properties and functions. Examples of such backbones include aryl oligomers[132] and aromatic oligoamides (**Fig. 21b & Fig. 21c**).[133] Foldamers, thus, could be composed of two or more different monomers or a single building block, such as aromatic disulfides reported by Otto and coworkers (**Fig. 21d**).[134] The folding of a foldamer consisting of an aromatic oligoamide is governed by the electrostatic repulsion and hydrogen bonding, which gives rise to helical conformation (**Fig. 20**). Additional stability to the helix is brought by aromatic stacking intramolecular interactions.[135]

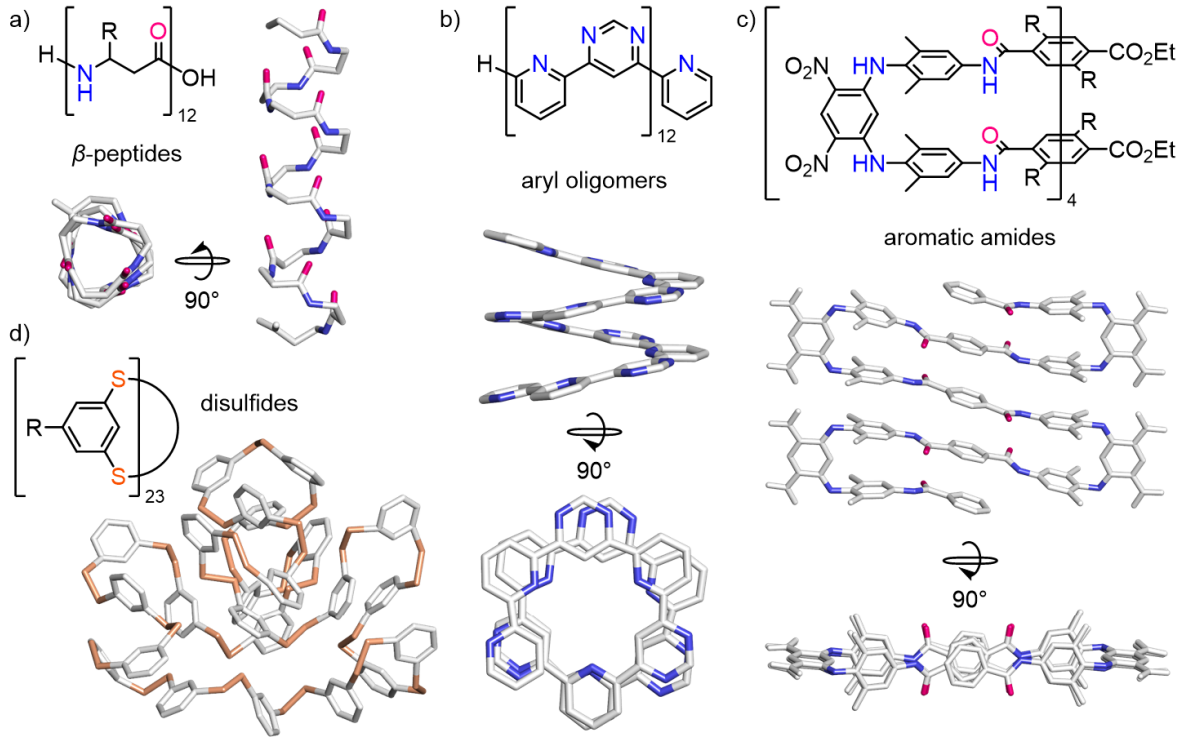


Fig. 21. An overview of the various foldamer families. a) Crystal structure of a β-peptide 12-mer. b) Crystal structure of a helical aryl oligomer. c) Crystal structure of a sheet-forming aromatic oligoamide. d) Crystal structure of an aryl disulfide macrocycle. The disulfide bonds are shown in orange, while the oxygen and nitrogen atoms are marked in purple and blue, respectively. The hydrogen atoms and side chains have been left out for clarity in all the structures. Figure adapted from figure number 2.1 from Bindl D.[136]

The fundamental forces governing the folding behavior of foldamer control the overall shape of the molecules. These forces are determined by the shape and size of monomers, which determines the intramolecular and intermolecular interactions between monomers.[137] As it is clear by now, creating a new foldamer involves identifying new polymeric backbones and designing them with chemical functionalities of interest. Due to this ability to achieve desired chemical functionalities, foldamer has gained interest from broad research areas ranging from synthetic chemistry to molecular biology. One such interest is to design foldamers that mimic their natural counterparts and use them as molecular tools and therapeutic agents. An early attempt to mimic biomolecules was mimicking proteins using peptoids.[138] Gellman and colleagues, in 1996, reported the ability of β-peptides to be more stable compared to α-peptides and fold in helical conformation. Since then, various backbones have been introduced in foldamer sequences, including aromatic oligoamides. The development of new heterogeneous foldamers (with more than one monomer) has led to the design and synthesis of molecules with the ability to attain tertiary and tetrameric conformations.[139] Helix bundles have dominated the field of tertiary conformations with α-peptidic helices, which has paved the path to the design of β-proteins.[140] As synthetic chemistry and molecular biology tools advance, the bridge between synthetic and bio-foldamers will be filled more quickly than ever before.

Over the last 25 years, foldamer research has rapidly expanded away from early efforts in peptidomimetics. Recently, efforts were made to design nucleic acid analogs toward a systematic exploration of folding in diverse structures that could be utilized in biological applications to foldamer-based materials.

2.6.1 DNA mimic foldamers

Although significant advancements in therapeutic development have been made, transcription factors have largely remained undruggable and are still primarily targeted by classical small molecules such as DNA ligands.[141–143] However, the dynamic nature of these DBPs poses a challenge when attempting to target them with small molecules. More difficulties are added because transcription factors, predominantly those that bind DNA with alpha helix, are not in a defined helical conformation in the absence of a DNA target, and thus, in the lack of an apo structure, they remain elusive. A potential solution involves replicating the interaction surfaces themselves. This requires the ability to mimic the primary structure of biomolecules and their secondary and possibly even tertiary structure.[137,144,145] Even before the term "foldamer" was coined, scientists had already successfully designed nucleic acid[119] and peptide[146] analogs that mimic the structures and biological properties of their natural counterparts. Molecules that could reproduce the topological features of either proteins or DNA could potentially be used to interfere with protein-DNA interactions. Despite the availability of DMPs in nature, no systematic study has been undertaken that could demonstrate the use of artificial architecture to mimic the topology of DNA.

In 2018, Ivan Huc and colleagues demonstrated using foldamer as DNA mimics for the first time.[147] *DNA mimic foldamers* are one group of molecules that aim to mimic DNA topology and thus interfere with the protein-DNA interactions.

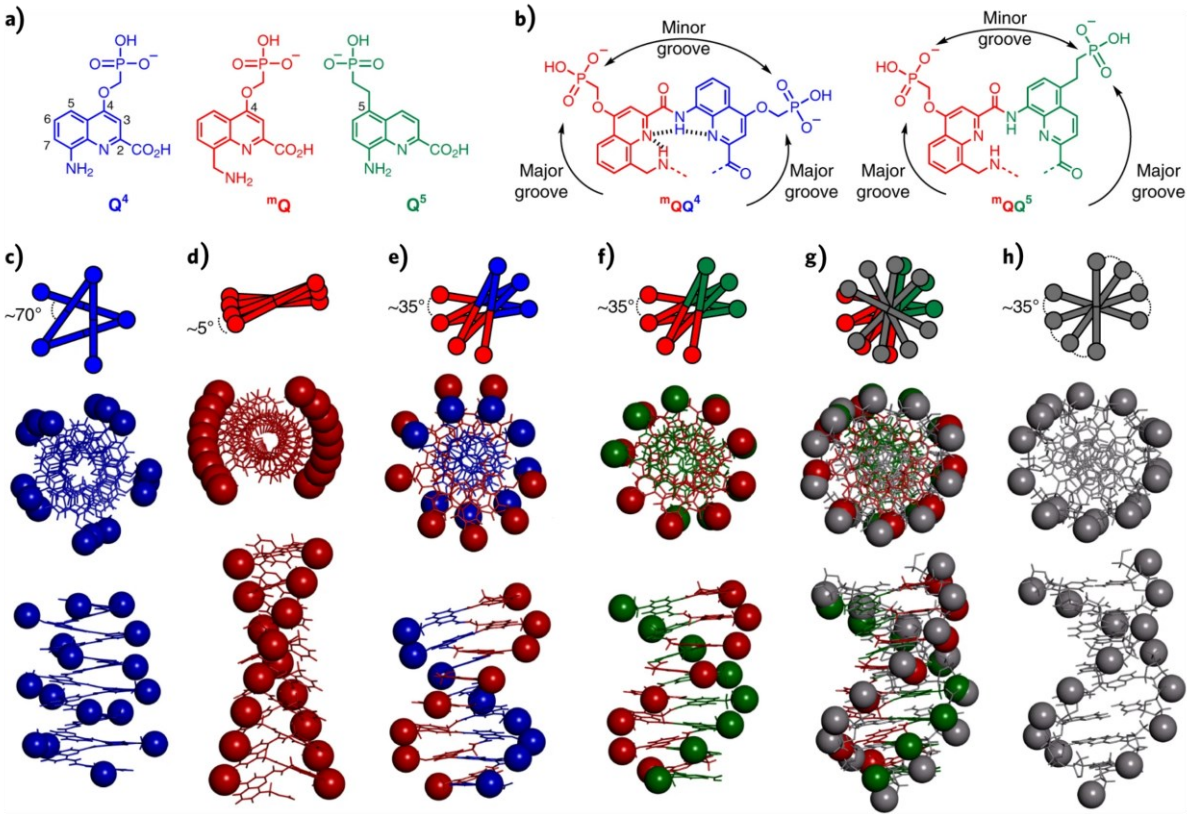


Fig. 22. DNA mimic foldamer design. a) Chemical formulae of monomers used in the synthesis of DNA mimic foldamers. b) Chemical formulae of $^{m}QQ^{4}$ and $^{m}QQ^{5}$ used in the synthesis of DNA mimic foldamers. c-f) Top view and side view of $(Q)_{16}$, $(^{m}Q)_{16}$, $(^{m}QQ^{4})_{8}$, and $(^{m}QQ^{5})_{8}$. g) Overlay of an $(^{m}QQ^{5})_{8}$ with an 8 base pair DNA showing the overlay of phosphorus atoms positions (shown in balls) h). Top view and side view of an 8 base pair double-stranded B-DNA. Figure adapted from Figure 1 of Ziach et al., 2018.[147]

DNA mimic foldamers reported in 2018 by Huc and colleagues remain the only foldamer class that could mimic the topology of B-DNA. These foldamers were synthesized based on aromatic oligoamides and folded into a single helix that replicates the charge distribution of a double-stranded B-DNA. The folding of these molecules is primarily influenced by the electrostatic repulsion and hydrogen bonding between the amide groups and neighboring residues. Additionally, the aromatic stacking within the helix contributes to increasing stability. In both solid state and solution, aromatic oligoamides derived from 8-amino-2-quinoline carboxylic acid **(Q)** adopt single helical conformations.[148,149] Interestingly, the $(Q)_n$ helix has an inner aromatic core with a diameter of approximately 9.4 Å between the C4 atoms of contiguous Q rings. This size matches a double-stranded B-DNA base pair with 9.0 Å between the purine and pyrimidine N1 atoms. Additionally, the $(Q)_n$ oligoamide helix has a pitch of 3.5 Å, similar to the base-pair distances in a double-stranded B-DNA. However, the helix curvature imparted by the Q monomer was reported to be 0.8 turns per two units, which differs from the position of phosphates in a double-stranded B-DNA. A new monomer, 8-aminomethyl-2-quinoline carboxylic acid **(^{m}Q)** combined with Q monomer (**Fig. 22a**), was reported to have a helix

curvature similar to a double-stranded B-DNA. In an **$(^{m}QQ^{4})_{n}$** helix, each $^{m}QQ^{4}$ unit has a curvature that spans 0.9 helix turns (**Fig. 22e**). Based on molecular modeling predictions, the angular shift between adjacent $^{m}QQ^{4}$ blocks was found to be one-tenth of a turn. This is equivalent to the angular shift between base pairs in double-stranded B-DNA, where a complete double helix turn consists of ten base pairs. As a result, the negatively charged residues on the surface of an $(^{m}QQ^{4})_{n}$ single helix will align in a double helical array corresponding to the phosphates' positions in double-stranded B-DNA (**Fig. 23**). Because of the geometrical parameters, an $^{m}QQ^{4}$ dimer spans 0.9 helix turn, which means that the next dimer in the sequence is shifted by a tenth of a turn backward. Because the shift is backward, the handedness of the main chain single helix and the handedness of the double-helical array of phosphonate side chains are opposite. Thus, an *M*-foldamer helix displays a double *P*-exohelix of negatively charged side chains and thus mimics the *P* DNA (B-DNA) double helix.

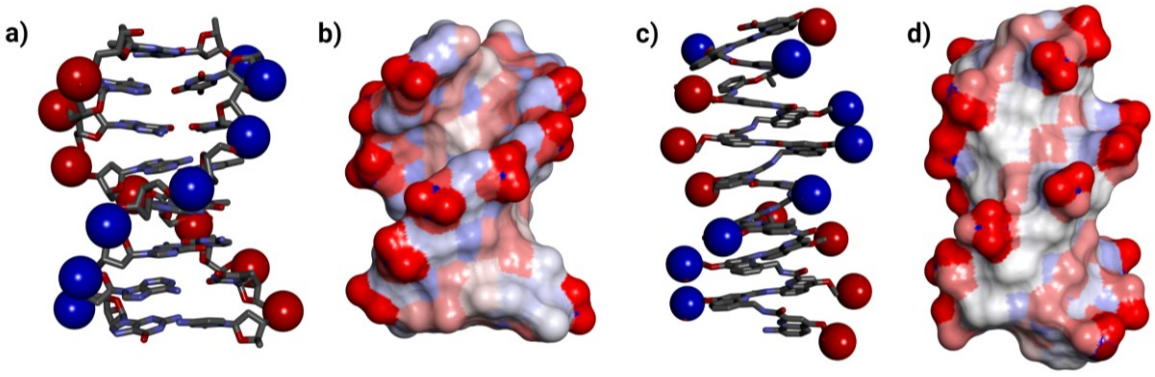


Fig. 23. DNA vs. DNA mimic foldamer. a) 8 base pair B-DNA shown in stick model representation with phosphate on each strand highlighted in red- and blue-colored balls. b) Surface representation of 8 base pair DNA. c) A $(^{m}QQ^{4})_{8}$ DNA mimic foldamer shown in stick representation with negatively charged phosphonates highlighted in red- and blue-colored balls. d) Surface representation of a $(^{m}QQ^{4})_{8}$ DNA mimic foldamer.

This design of the $(^{m}QQ^{4})_{n}$ oligomer also allows for the imitation of major and minor grooves similar to those found in double-stranded B-DNA. However, it was reported that the functionalization of the Q monomer with a negative charge at position 5 yielded a better mimic of double-stranded B-DNA, as evidenced by the molecular model (**Fig. 22**).[147]

Although double-stranded B-DNA and the $(^{m}QQ^{4})_{n}$-based foldamer share some similarities, there are also notable differences. Unlike DNA, where two covalent bonds connect base pairs, a single covalent bond connects the blocks of an $(^{m}QQ^{4})_{n}$ foldamer. This unique characteristic of DNA mimic foldamer allows for greater flexibility, potentially advantageous for binding and manipulation with DBPs such as transcription factors and histones.

2.7 Sac7d, a thermophilic marvel

In the realm of biological complexity, the underlying framework of life resides within the nucleus of each cell - a molecular tapestry of DNA and its associated proteins, collectively known as chromatin. Beyond its canonical role as a genetic repository, chromatin performs gene regulation that governs the diverse phenotypic manifestations across the spectrum of life forms. In eukaryotes, DNA is wrapped around a histone octamer, proteins that bind, condense, and regulate DNA.[150] Histone proteins are, however, absent in bacteria. Nucleoid-Associated Proteins (NAPs), a group of chromatin-binding proteins, take on the role of DNA condensation. Among the well-studied NAPs in *E. coli* are HNS, HU, Fis, IHF, and StpA (**Fig. 24**). These NAPs have the dual function of organizing bacterial chromosomes and regulating gene expression.[151] In contrast to bacteria, archaea contain many chromatin-binding proteins. Each species of archaea has multiple types of chromatin proteins that have different effects on the structure of DNA.

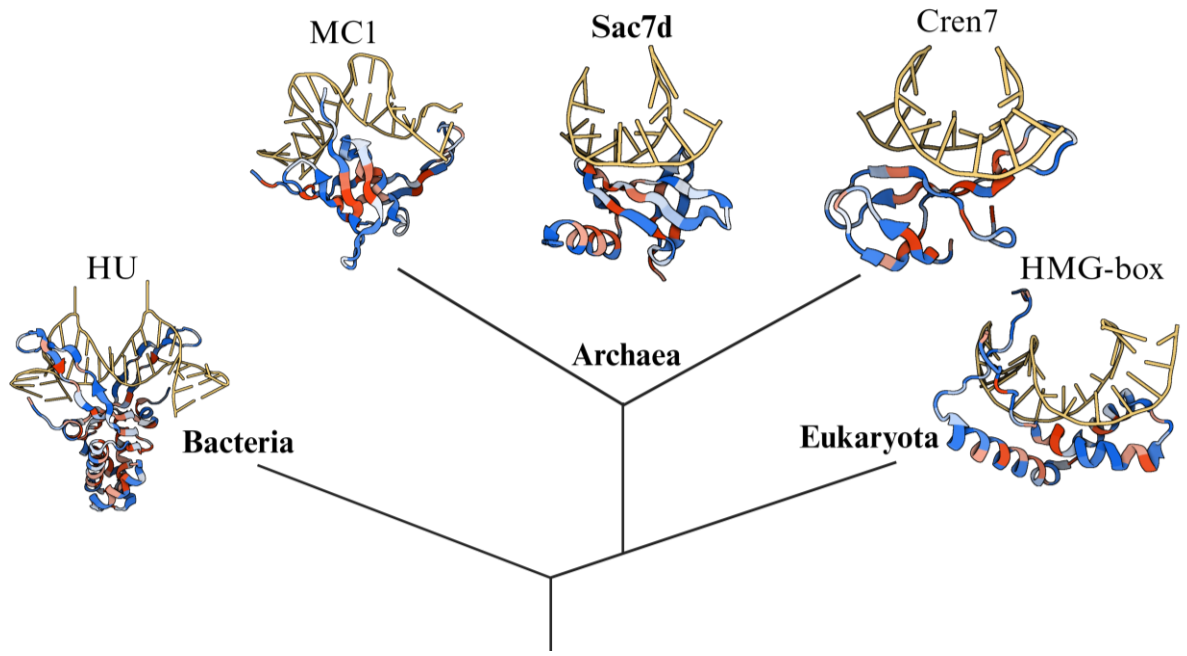


Fig. 24. A simple rendition of a phylogenetic tree showing Sac7d, among other DNA bending proteins ranging from HU in bacteria to HMG-box containing protein SRY in eukaryotes.

In the context of histones and their central importance in all forms of life, Sac7d protein emerges as an intriguing entity. Originally isolated from the extremophilic archaeon *Sulfolobus acidocaldarius,* Sac7d has garnered attention for its unique histone-like characteristics. Sac7d, although not a canonical histone, exhibits a structural architecture reminiscent of histone proteins. Its compact and positively charged surface, highlighted with basic amino acid residues, allows it to interact with DNA in a non-sequence-specific manner, forming stable

protein-DNA complexes (**Fig. 25**). The extreme conditions in which *S. acidocaldarius* thrive pose unique challenges for chromatin organization. The histone-like properties of Sac7d offer potential insights into how chromatin is adapted to function optimally under such extreme temperature conditions. Sac7d's role in DNA compaction and structural stability may protect the genetic material from harsh thermal and chemical stresses while allowing dynamic accessibility to regulate gene expression. Thus, sac7d is also essential for the genomic integrity of *S. acidocaldarius*. The ability of Sac7d to withstand harsh temperature conditions that denature most protein is attributed to its compact structure. Sac7d's structure is characterized by a single-domain α-helix bundle and five beta sheets spanning 66 amino acids (7 kDa). Due to its small size, chemical compatibility, and ability to be chemically synthesized, Sac7d lies somewhere in between a protein and peptide.

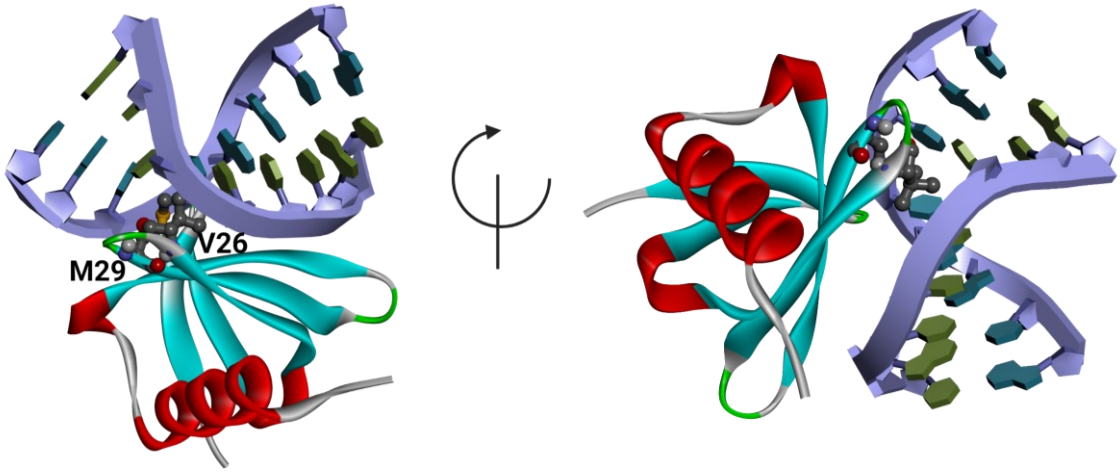


Fig. 25. Crystal structure of Sac7d in complex with an 8BP double-stranded B-DNA (PDB ID 1AZQ). Sac7d residues intercalating in DNA chains are shown in stick representation (V26 and M29).

Because Sac7d maintains its structural integrity even at temperatures above 100 °C and at pH 0, it has been used in various biotechnology and protein engineering applications.[152] Frédéric Pecorari and colleagues, in 2013, successfully re-engineered Sac7d, which originally binds to DNA segments to a protein binder.[152] Furthermore, they also demonstrated that site-directed mutagenesis of Sac7d conserving the structural integrity led to identifying ligands that were shown to bind to human IgG epitopes. This library of mutated wild-type Sac7d is commercially known as *Nanofitins*, antibody mimics, and developed as an antibody alternative biotechnology tool. Sac7d has also been shown to incorporate complementary determining regions from antibodies in its loop regions, thus being able to bind to the target antigen.[153] In 2005, a team led by Andrew H.-J. Wang created a multimeric DBP using Sac7d and leucine zipper as the basis. Their newly engineered protein was able to maintain the properties of the original Sac7d while also gaining new capabilities from the leucine zipper component, thus opening further

avenues for using Sac7d as a tool for therapeutic development and as a model protein for biological research.[25]

2.7.1 DNA bending versus DNA kinking

It is clear by now that protein-DNA complexes are highly dynamic, and the flexibility of DNA allows DBPs to take advantage of this feature and introduce deformities to DNA conformation to access recognition sites. In 1975, Crick and Klug[154] brought attention to the accessibility of base pairs at a kinked region by DBPs. Further studies have demonstrated that particular DNA sequences, such as AT-rich repeats, can induce curvature in the DNA molecule.[155] Some proteins induce a smooth bend compared to a sharp kink in the DNA structure. Sac7d, because it is involved in chromatin organization and transcription regulation, has been shown to induce a sharp kink of approximately 60 ° via the intercalation of Val26 and Met29 residues into the DNA helix (**Fig. 26**).[156]

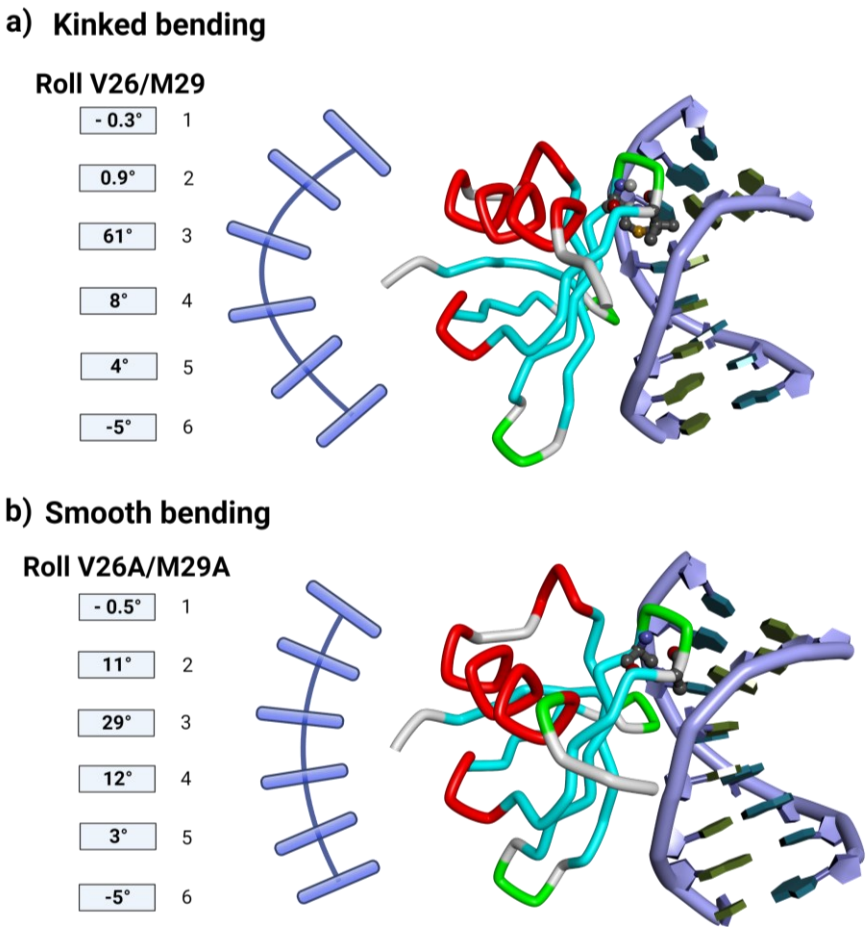


Fig. 26. Sharp kinking versus smooth bending of DNA by Sac7d. a) Wild-type Sac7d intercalates V26 and M29 residues in between 3rd and 4th base pair of an 8BP DNA strand inducing a sharp kink of 61 °. b) Sac7d V26A/M29A, due to shorter side chains at positions 26 and 29, induces a significantly smooth bending, resulting in a roll of 29 °.

3. Objectives and rationales

Competitive inhibitors, within the realm of biochemistry, are a class of molecules that exert their influence by occupying the active site of a protein. This occupation effectively outcompetes the substrate to access the protein's catalytic site. The mechanistic underpinning of competitive inhibition stems from the geometric compatibility of the inhibitor molecule with the active site's spatial and chemical features. By mimicking the structural attributes of the substrate, the competitive inhibitor effectively disguises itself to the protein, leading to its binding with a comparable or better affinity (**Fig. 27**). Competitive inhibitors are keenly explored for their potential applications in drug design and therapeutic interventions. By employing competitive inhibitors strategically, it becomes feasible to modulate enzyme activity, thereby controlling these pathways and potentially mitigating aberrant physiological processes.

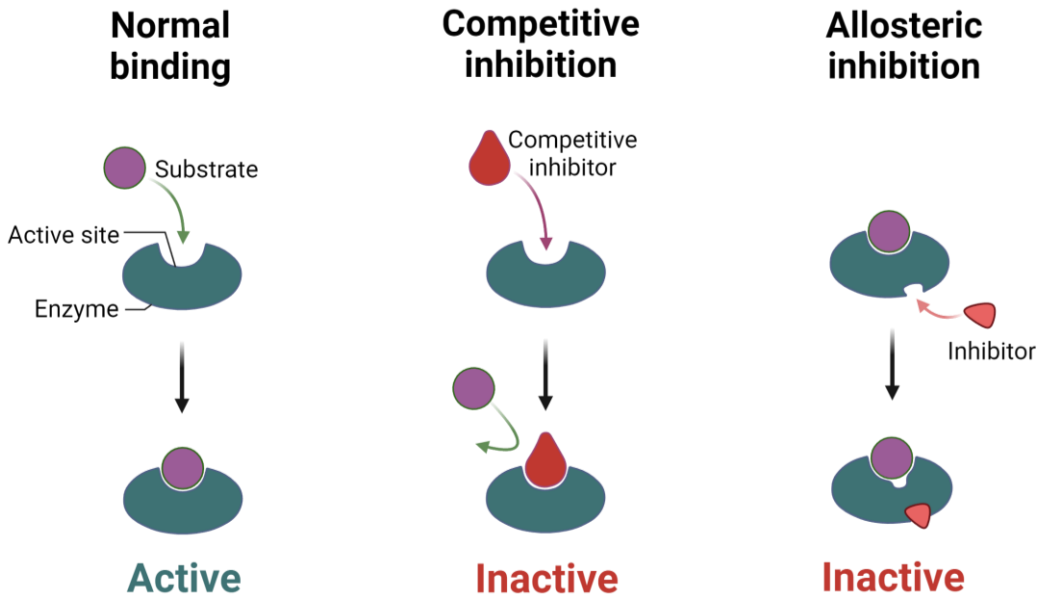


Fig. 27. General mechanism of enzyme activity inhibition by competitive inhibition (center) is shown along with normal binding (left) and non-competitive inhibition (right).

The 2018 study by Ivan Huc and colleagues successfully showed the use of $(^{m}QQ^{4})_{n}$-based DNA mimic foldamers to inhibit the activity of several DBPs.[147] It has been reported that $(^{m}QQ^{4})_{n}$-based DNA mimic foldamers strongly inhibit Topoisomerase 1 (Top1), an enzyme that cuts one strand of a double-stranded DNA to relieve supercoiling. Top1 inhibition by DNA mimic foldamers was performed in the presence of three orders of larger (in terms of base pairs) DNA substrate. Additionally, the human immunodeficiency virus 1 integrase (HIV-IN) enzyme, which has similarities in structure and mechanism to eukaryotic Top1, was also found to be inhibited by $(^{m}QQ^{4})_{n}$. Competitive enzyme inhibition by DNA mimic foldamers was reported at concentrations lower than 1 μM.[147] A DNA of similar length (8 base pair) to

($^{m}QQ^{4}$)$_8$ was reported to have little to no effect on enzymatic activity of both Top1 and HIV-IN. The mode of inhibition has been reported to be a direct interaction of DNA mimic foldamers with both Top1 and HIV-IN. Thus, the DNA mimic foldamers bind to Top1 and HIV-IN entirely differently than their poison drugs, CPT[157,158] and RAL[159]. This way, the mimicry of negative charges of DNA using an artificial architecture was proven critical for DBPs of therapeutic relevance. However, It has also been reported that ($^{m}QQ^{4}$)$_n$ DNA mimic foldamers had no inhibition of sequence-selective DBPs such as restriction enzymes, for example, XhoI and NdeI. Moreover, some sequence-neutral proteins were not inhibited, including deoxyribonuclease and S1 nuclease.[147]

Despite the excellent inhibition of HIV-IN and Top1 by DNA mimic foldamers, the 2018 study by Ivan Huc and colleagues left more questions than answers. In particular, how could the DNA mimic foldamer be designed to bind exclusively to a single protein target such as Top1 or HIV-IN? This is important because a drug candidate that binds to both viral (HIV-IN) and human (Top1) protein could cause complications. To develop DNA mimic foldamers as drug candidates, they must be designed against specific protein targets with the least off-target effects. Continuing from the 2018 study by Huc and colleagues, further pieces of evidence are needed to determine the specific structural requirements for DNA mimics to inhibit/bind DBPs in a sequence-selective manner. This entails conducting structural studies on foldamer-protein complexes and producing foldamer variants through solid-phase synthesis.

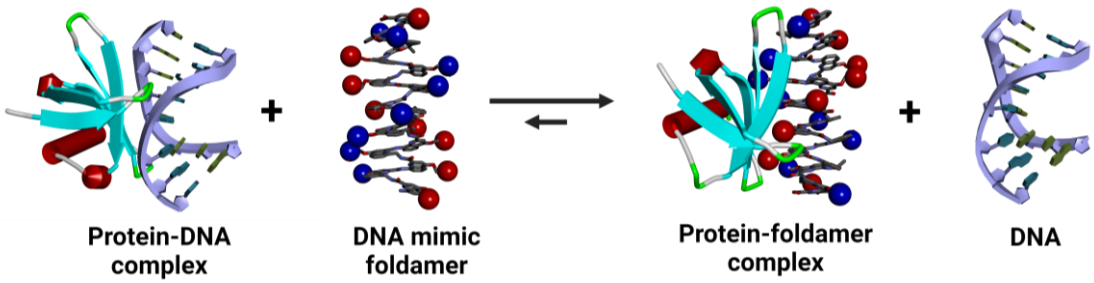


Fig. 28. Principle of competitive inhibition of protein-DNA interaction using foldamers that mimic the DNA surface features.

This thesis aims to reveal the factors that influence the binding of DNA mimic foldamers to DBPs. By studying the underlying principles that govern this process, we aim to gain a better understanding of how DNA mimic foldamers bind to a variety of DBPs, such as Sac7d, cGAS, and Dpo4, and establish DNA mimic foldamers as a competitive inhibitor of protein-DNA interactions (**Fig. 28**).

Sac7d is chosen to be a model DBP system to study the structural basis of interactions between DNA mimic foldamer and Sac7d. In addition to the remarkable features of Sac7d mentioned in section 2.7, it is also easy to express in the bacterial system and can be purified relatively easily.[24] Also, the crystallization[18] of Sac7d (and Sso7d, a homolog protein of Sac7d) with its cognate DNA is well established and characterized, along with the solution structure of Sac7d-DNA was previously solved using NMR spectroscopy.[160] These attributes render Sac7d a thoroughly characterized system well-suited for investigation using foldamers.

In order to understand the structural basis of interaction between DNA mimic foldamers and DBPs, we extensively rely on protein X-ray crystallography as the primary technique to reveal the details at the atomic level. It should be noted that there are inherent obstacles to using protein X-ray crystallography as a method of choice. The first hurdle lies in the fact that crystallization of DNA mimic foldamers with and without DBPs has never been explored with success. Even once the crystals are obtained, it is not guaranteed that there will also be diffraction data at a higher resolution. Once the data at a satisfactory resolution has been collected (between 2 Å to 3 Å), solving the structure by molecular replacement remains challenging due to the availability of limited molecular models for DNA mimic foldamers. Furthermore, even when the structure of DNA mimic foldamer (in complex with protein) has been solved, refinement of these structures is not readily usable in software suites that are predominantly meant to refine protein and DNA structures. For example, most available software packages that generate the restraint file (required in refinement) are meant for small molecule ligands, and generating restraint files for DNA mimic foldamer (molecular weight more than 5000 Da) could often lead to error-prone values. Despite these challenges, protein crystallography remains the best method (especially for proteins smaller than 80 kDa) to elucidate the high-resolution structure of protein-DNA mimic foldamer complexes. Although it is essential to mention here that protein X-ray crystallography only provides a single snapshot of the molecules in the solid state, crystallography studies were complemented with NMR spectroscopy and biophysical techniques to validate the binding of protein-DNA mimic foldamers.

In summary, this thesis aims to understand and reveal the structural basis of interaction between DNA mimic foldamers and DBPs using structure-based iterative design and pave the path for designing DNA mimic foldamers with capabilities to bind DBPs (such as transcription factors) that bind to DNA in a sequence-specific manner.

4. DNA mimic foldamer recognition of a chromosomal protein

Huc and colleagues successfully showed the use of aromatic oligoamide DNA mimic foldamer based on 8-amino-2-quinoline carboxylic acid (Q^4) and 8-aminomethyl-2-quinoline carboxylic acid (^mQ) units (**Fig. 22**) to inhibit the activity of several DBPs.[147] It has been reported that DNA mimic foldamers strongly inhibit Top1 and HIV-IN. However, it has also been reported that these foldamers had no inhibition of sequence-selective DBPs such as restriction enzymes. Moreover, some sequence-neutral proteins were not inhibited, including deoxyribonuclease and S1 nuclease.[147] Despite the excellent inhibition of HIV-IN and Top1, how could the DNA mimic foldamer be designed to bind exclusively to a single protein target? To develop DNA mimic foldamers as drug candidates, they must be designed against specific protein targets with the least off-target effects. Continuing from the 2018 study by Huc and colleagues, further evidence is needed to determine the specific structural requirements for DNA mimics to inhibit/bind DBPs in a sequence-selective manner. This entails conducting structural studies on foldamer-protein complexes and producing foldamer variants through solid-phase synthesis. Please note that in the manuscript Q^4 is called Q and ^mQ is called M.

The work presented in the following publication (submitted to Angewandte Chemie) shows for the first time how DNA mimic foldamers not only recognize Sac7d, a chromosomal protein, but outcompete DNA (of comparable length) for binding to Sac7d (**Fig. 29**). The three-crystal structure solved are presented in the publication highlighting the role of DNA mimic foldamers in binding to the DNA binding region of Sac7d but with a mode unique to foldamers only. This was further confirmed by various NMR spectroscopy experiments, which confirmed the interactions shown in crystal structures. This work, being the pioneer in structural studies of DNA mimic foldamers with DBPs, opens new avenues to design DNA mimic foldamers to target sequence-specific DNA binding proteins.

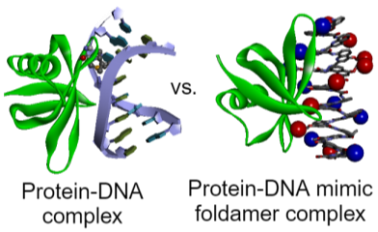


Fig. 29. Comparison of Sac7d-DNA complex versus the protein-DNA mimic foldamer complex (docked).

Contributions: The ideas for this work were developed by me in collaboration with I. Huc and V. Corvaglia. Protein expression (including ^{15}N labeled Sac7d) and purification were performed by me. All crystallization and structure solutions were also performed entirely by me. CD competition experiments were performed by me. Synthesis of foldamers was performed by V. Corvaglia and J. Wu. NMR spectroscopy experiments were performed in collaboration with L. Allmendinger. The manuscript was written by me in collaboration with V. Corvaglia and I. Huc.

4.1 Publication

DNA Mimic Foldamer Recognition of a Chromosomal Protein

Deepak Deepak,[a] Valentina Corvaglia+,[a] Jiaojiao Wu,[a] Lars Allmendinger,[a] and Ivan Huc*[a]

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

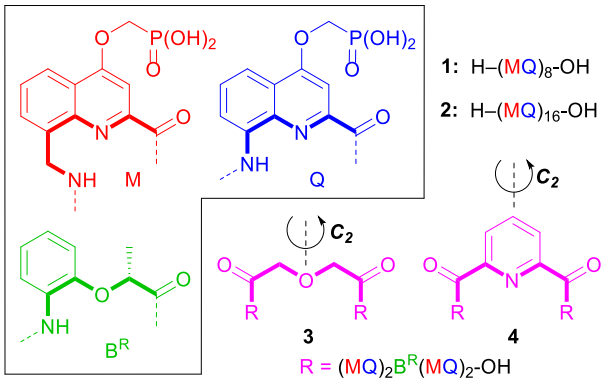
+ Current address: Institute for Stem-Cell Biology, Regenerative Medicine and Innovative Therapies, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo (Italy) & Center for Nanomedicine and Tissue Engineering (CNTE), ASST Grande Ospedale Metropolitano Niguarda, Milan (Italy).

Supporting information for this article is given via a link at the end of the document.

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

The concept of biomolecular mimicry is familiar but not always grasped in its full depth. A mimic reproduces some of the features of the biomolecule from which it is inspired. However, a mimic is also meant to outcompete the original, thanks to differences: a perfect “copy” would match the original and not outcompete it. Thus, the more structurally different the mimic, the more difficult it is to elicit properties similar to those of the original, and the more opportunities to outcompete the original. α-Helix mimetics[1] that target some protein surfaces and DNA analogs[2] that hybridize with DNA better than DNA itself are prominent examples of successful biomimicry. In contrast, synthetic molecules that would mimic the surface of an entire DNA double helix and highjack DNA-binding proteins are underdeveloped despite their potential to offer unique opportunities for pharmacological applications.

DNA mimicry exists in nature in the form of DNA mimic proteins typically rich in aspartic acid and glutamic acid residues.[3] Inspired by nature, the use of artificial proteins as DNA mimics has been initiated but not extensively pursued.[4] The so-called decoy oligonucleotides (ODNs) may also be used to target DNA-binding proteins such as transcription factors. These ODNs are modified to enhance their bioavailability, but they may not bind to their target better than their natural counterparts.[5] In this context, we have introduced abiotic aromatic oligoamides bearing anionic phosphonate side chains that fold in water into single helices whose shape and charge distribution mimic the shape and charge distribution of double-stranded B-DNA.[6] For example, sequences **1** and **2** (Scheme 1, Figure S1) are equivalent to eight and sixteen base-pair B-DNA duplexes, respectively. However, being monomeric, they do not melt into two single strands as DNA does. Such foldamers do not possess sequence features other than the alternation of M and Q monomers (Scheme 1). They have been shown to affect chromatin composition and perturb cell cycle progression[6c] and to bind to some non-sequence selective DNA-binding proteins with high affinity, leading to the competitive inhibition of therapeutically relevant enzymes such as topoisomerase 1 (Top1) and HIV integrase 1 (HIV-IN) even in the presence of a large excess of DNA substrate.[6b] These effects show some selectivity: binding is not effective on all non-sequence-selective DNA-binding proteins.[6b] Furthermore, structural modifications of the foldamers have been shown to enhance selectivity for Top1 or HIV-IN.[6a] However, the lack of detailed structural information on how these molecules recognize proteins hampers the development of better and more selective binders. Here, we present the first structural investigation of interactions between a DNA mimic foldamer and a non-sequence selective DNA-binding protein, Sac7d, a bacterial chromosomal protein known to kink and package DNA in hyperthermophilic archaeon *Sulfolobus acidocaldarius*.[7] In this study, we demonstrate that foldamers outcompete DNA in targeting the DNA binding site of Sac7d by adopting a distinct binding orientation and without being kinked.



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

Sac7d was selected as a model protein for structural investigations with DNA mimic foldamers because of its high structural stability, solubility, and rich literature on how it interacts with DNA.[7b, 7e, 7g, 7j] As an initial test, we assessed the binding of **1** and **2** to Sac7d by SPR with the protein immobilized on the chip and each foldamer in the mobile phase (Figure 1a, Figures S2-S3). Steady-state data indicated a K_d in the one-digit micromolar range. This is to be compared to the two-digit micromolar binding reported for calf thymus DNA.[7a] However, the sensorgrams did not fit well with the kinetics of a single binding event. A possible reason for this is that **1** and **2** do not contain any stereogenic center and thus exist as a racemic mixture of left-handed (*M*) and right-handed (*P*) helices, accounting for at least two types of interactions with the protein.

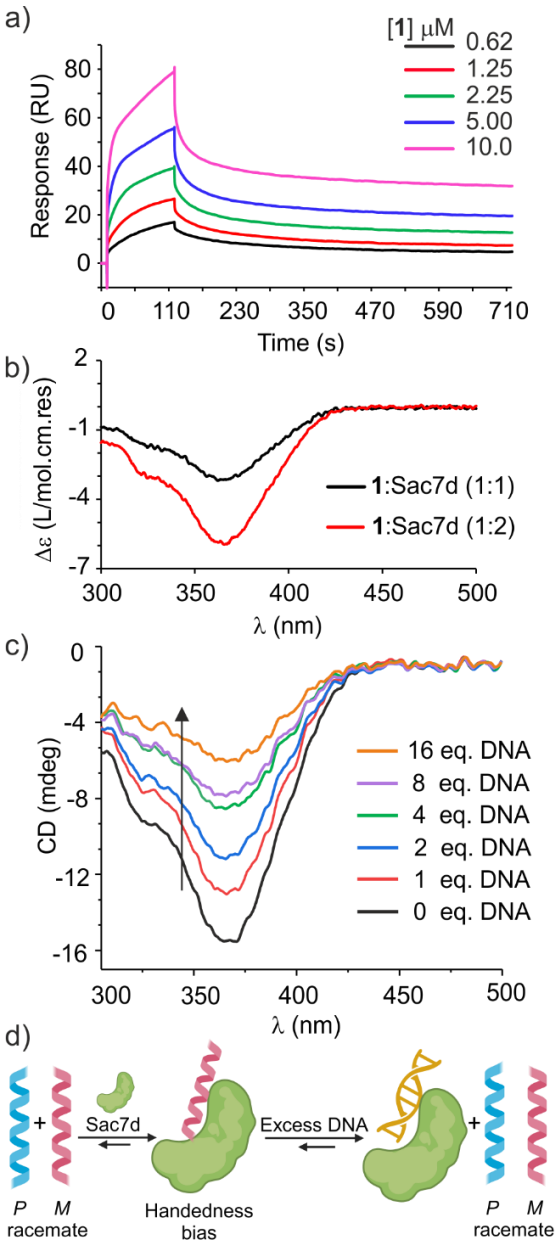


Figure 1. SPR and CD evidence of binding of **1** to Sac7d. a) SPR sensorgrams of the interaction between His_6-tagged Sac7d and **1** in HBS-EP buffer pH 7.4 at 25 °C (see SI for buffer composition). Sensorgrams were plotted after subtraction of the signal of the reference flow cell. b) CD spectra of **1** at 40 µM concentration in the presence of 1 or 2 equiv. of Sac7d in 50 mM NH_4HCO_3 buffer pH 8.5 at 20 °C after 24 h equilibration. c) CD spectra of a 1:1 mixture of **1** and Sac7d in the presence of increasing amounts of a 10-bp DNA. d) Cartoon representation of handedness bias of **1** upon binding to Sac7d and of the competitive association of DNA.

Because **1** and **2** are achiral, they have flat CD spectra. However, when incubating **1** and **2** with Sac7d, a CD band emerged at 365 nm as a function of time (Figure 1b, Figures S4-S5), indicating aromatic helix handedness bias by interconversion of one enantiomer into the other. The negative sign of the band implies a preference for *M* helicity,[8] that is, for the enantiomer that mimics the *P* B-DNA double helix (Figure S1). Several conclusions can be drawn from this observation: (*i*) DNA mimic foldamer *P* and *M* helices interconvert, although slowly. This was unexpected, considering the lengths of **1** and **2** and the kinetic inertness of related foldamers.[9] The faster dynamics may result from negative charge repulsions; (*ii*) CD confirms that the foldamers bind to Sac7d; (*iii*) the interaction is diastereoselective and favors the enantiomer that mimics B-DNA, not its mirror image; and (*iv*) the intensity of the band (the value of Δε/residue in Figure 1b) suggests that bias is extensive (the estimated diastereomeric excess is 75%).[10]

Fitting the CD data to a 1:1 binding isotherm (Figure S5) yielded a K_d of 34 µM, apparently higher than estimated by SPR. Nevertheless, CD spectroscopy also confirmed that DNA binds less effectively than the foldamers. Thus, a competition experiment was set in which a 10-base pair (bp) DNA duplex was added to the already equilibrated *M*-**1**-Sac7d complex. The negative CD band at 365 nm gradually vanished as the DNA-Sac7d complex forms and the released foldamer helix racemizes (Figure 1c, d). From the quantity of DNA necessary to make CD intensity drop by half (~4 equiv. with respect to the foldamer), one can estimate that **1** binds at least tenfold better than the 10 bp DNA (see Supporting Information).

We next investigated DNA mimic foldamer recognition of Sac7d using NMR spectroscopy. Foldamers **3** and **4** were designed and synthesized for this purpose (Scheme 1).[11] These sequences are based on a central diacid linker that makes them C_2-symmetrical, *i.e.,* palindromic-like so that they would produce the same type of complex with Sac7d irrespective of their orientation. They also possess two chiral B^R residues (Scheme 1) that quantitatively bias their handedness towards the *M* helix.[10] Sequence **3** possesses an aliphatic more flexible central linker than the aromatic linker of **4**. This difference was introduced in relation to the ability of Sac7d to kink DNA, a flexible linker possibly playing the role of a hinge. Several 2D and 3D NMR experiments on ^{15}N-labelled Sac7d using ^{1}H-^{15}N HSQC (Hetronuclear Single Quantum Coherence) spectroscopy, ^{15}N-HSQC-NOESY (Nuclear Overhauser Effect Spectroscopy) and ^{15}N-HSQC-TOCSY (Total Correlation Spectroscopy) allowed for the unambiguous assignment of the protein backbone (Figure S6). Titrations of Sac7d by **3** and **4** were then monitored by ^{1}H-^{15}N HSQC (Figures 2a-c, Figures S7-S10). No significant differences were observed between the two foldamers **3** and **4**. Both caused similar chemical shift variations, indicating molecular associations in fast exchange on the NMR timescale. Some

chemical shift variations followed a monotonous trend (straight arrows in Figure 2a) which could in principle be fitted to a 1:1 binding isotherm. However, others were not monotonous (kinked arrows in Figure

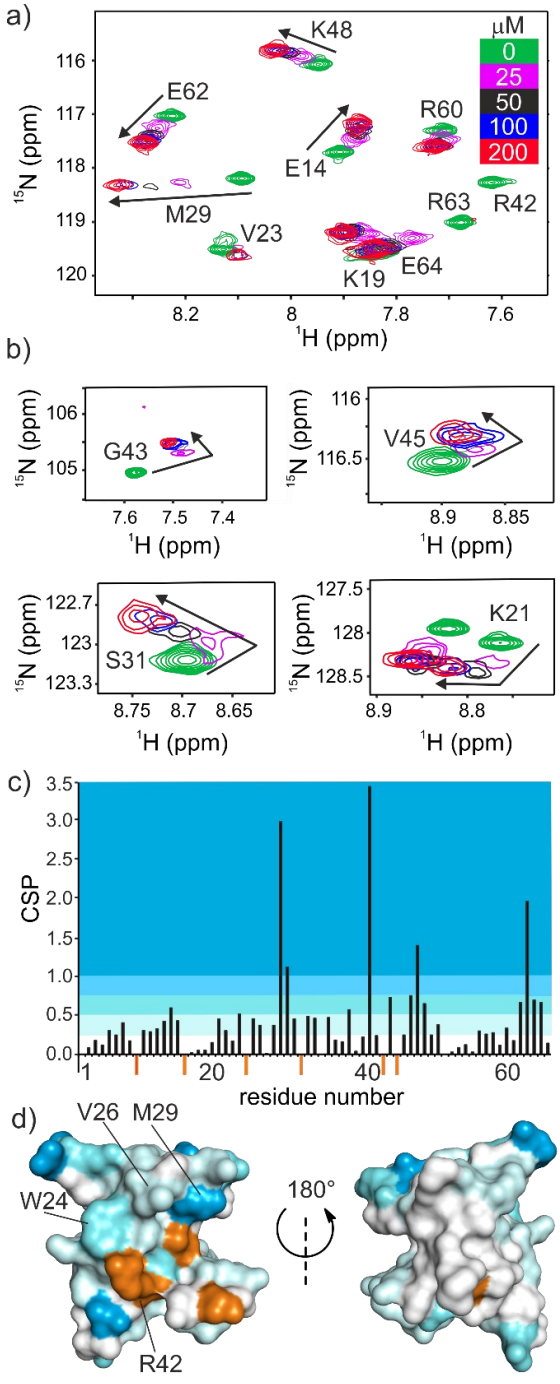


Figure 2. NMR spectroscopic evidence of binding of **3** to Sac7d at 100 μM concentration in 50 mM Tris-d_{11} buffer pH 7.5, 50 mM KCl, 10% D_2O. a, b) Part of ^{1}H-^{15}N HSQC titration of [^{15}N]-Sac7d with **3**. The colored scale indicates an increasing concentration of **3**. c) CSPs of [^{15}N]-Sac7d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

2b), indicating that at least two complexes with different molecularities form, *e.g.* 1:1 and 1:2. Mapping the chemical shift perturbations (CSPs) on the surface of the Sac7d structure revealed that perturbations caused by the foldamers occur at and around the beta-sheet DNA binding site and involved some key residues for DNA recognition (W24, V26, M29, and R42). In contrast, the opposite face of the protein was essentially unaltered. This strongly supports that the DNA-mimic foldamers also interact with the DNA-binding region of Sac7d.

Finally, we endeavored to crystallize Sac7d-foldamer complexes. As for complexes between DNA and non-sequence selective DNA-binding proteins, this was challenged by the possible degeneracy of the binding modes *via* frame shifts of one bp (here one MQ dimer) that do not favor crystal growth. In **3** and **4**, this degeneracy was mitigated by the C_2-symmetry and the presence of other units than M and Q. Crystals of a Sac7d in complex with **3** were obtained (Figure S11) and diffracted at 2.6 Å. The structure was solved by molecular replacement using the Sac7d structure[7g] and a molecular model of the foldamer fragment as the search model (see Supporting Information). The refined structure revealed the formation of a C_2-symmetrical 2:1 complex with two proteins binding to a molecule of **3** (Figure 3b). This stoichiometry may explain the trend of CSPs during NMR titrations (Figure 3b). Thus, a crystallographic C_2 axis crosses the middle of the linker of **3**. The asymmetric unit, therefore, contains half of this 2:1 complex, along with half of a second molecule of **3** not interacting with Sac7d. Crystal packing involved columnar pseudo-continuous stacks of foldamer helices resembling the stacks of DNA duplexes often observed in crystals of protein-DNA complexes (Figure S12). The structure confirmed that the foldamer extensively interacts with the DNA binding region of Sac7d, *i.e.* through contact area of 1500 Å. However, the orientation of the helix of **3** is almost perpendicular to that of DNA (Figure 3c, d). The protein-foldamer interface features multiple hydrogen bonds including charge-reinforced hydrogen bonds involving phosphonates and C-terminal carboxylates of **3**, as well as some hydrophobic contacts (Figure 3e-h). Many residues involved are also key residues for Sac7d-DNA interactions (*e.g.,* Y8, W24, and R42, see Figure S13). Nevertheless, the Sac7d-**3** complex has a unique geometry reflecting the structural features of the foldamer. For example, the B^R forms a hydrophobic contact with the protein surface allowed by the lack of a side chain of that residue. If B^R was replaced by Q or M, it would lead to a steric clash.

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

All attempts to crystallize **4** in complex with Sac7d failed: precipitates were obtained using conditions under which crystals of Sac7d-**3** grew. Clearly, the larger central linker of **4** would fill

the space occupied by the two V26 side chains of the structure of **3**, hampering the binding geometry observed with **3**. Nevertheless, NMR showed that **4** also binds to Sac7d. The double mutant Sac7d (*V26A/M29A*), which lacks the side chains responsible for DNA kinking, was crystallized in complex with **3** in

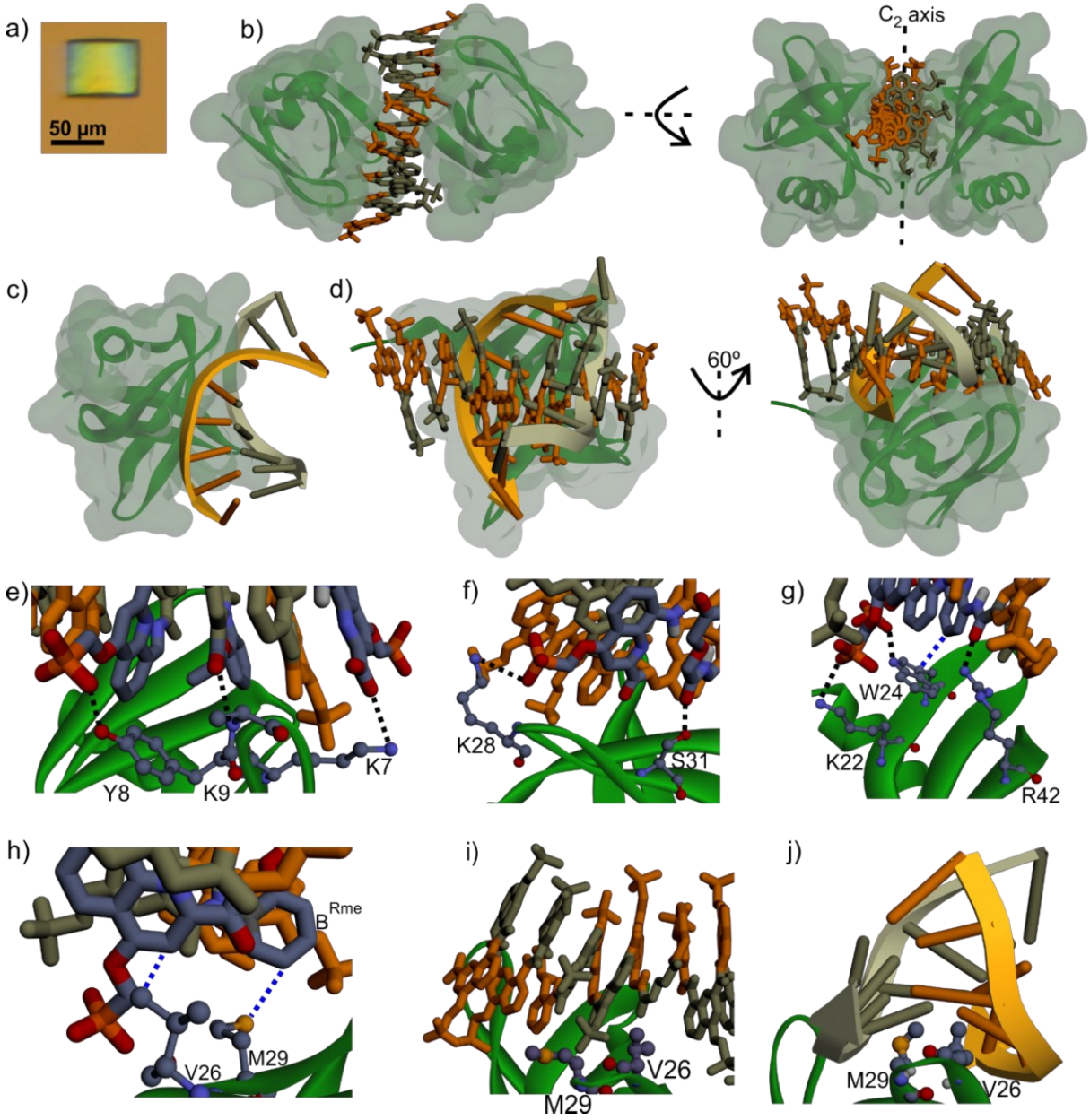


Figure 3. Crystal Structure of Sac7d-**3** complex (PDB# 8CMN) and comparison with a Sac7d-DNA complex (PDB# 1AZQ).[7g] a) X-ray diffraction-quality Sac7d-**3** crystal. b) 2:1 Sac7d-**3** complex. c) Sac7d-DNA complex. d) Sac7d-**3** complex superimposed (superimposition α-carbons) with the Sac7d-DNA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

a structure otherwise identical to the Sac7d-**3** complex (Figure S14). Furthermore, Sac7d (*V26A/M29A*) also crystallized in complex with **4**. The structure is again similar to that of the Sac7d-**3** complex, with the central pyridine ring of **4** filling the space made available by the V26A mutation (Figure S15).

In conclusion, DNA mimic foldamers represent a novel platform to target the large ensemble of proteins that interact with nucleic acids. They complement other aromatic foldamers[12] macrocycles[13] and tweezers[14] that have also been shown to bind to protein surfaces. The structure of the Sac7d-**3** complex provides a solid foundation for the structure-based design of the foldamer to enhance binding affinity and specificity, exploiting the inherent modularity of foldamer sequences. The specific features of this complex would have been hard to predict, rendering any

further design attempts without structural information a rather elusive endeavor. We have now demonstrated that protein-DNA mimic foldamer complexes can be crystallized and deliver detailed structural information even in the absence of sequence specific recognition features. We are currently extending this work to other proteins with the long-term objective of creating DNA mimic foldamers highly specific to their protein target.

Supporting Information

The authors have cited additional references within the Supporting Information.[15-27]

Acknowledgements

We acknowledge financial support from the Deutsche Forschungsgemeinschaft (DFG) via projects HU1766/2-1 and 325871075 (CRC1309-C7), from the European Research Council (ERC) under the European Union's Horizon Europe Framework Programme (Grant Agreement No. ERC-2021-ADG-320892), from the German Academic Exchange Service (*DAAD*, predoctoral fellowship to D. D.) and from the China Scholarship Council (CSC, predoctoral fellowship to J. W.). This work has benefited from the facilities and expertise of the Biophysical and Structural Chemistry platform (BPCS) at IECB, CNRS UMS3033, Inserm US001, and Bordeaux University. We thank L. Minder for assistance with SPR measurements and P. K. Mandal for preliminary investigations. We thank L. McGregor (ID23-1, ESRF), M.S. Lopez (ID30B, ESRF), and I. Ibento (EMBL P13, Petra III, DESY) for assistance during data collection at synchrotron beamlines. We also thank Prof. C.Y. Chen (Hubei University, Taiwan) for kindly providing us with the expression vector for Sac7d and Sac7d V26A/M29A proteins.

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

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

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

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

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

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

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

[7] a) J. G. McAfee, S. P. Edmondson, I. Zegar, J. W. Shriver, *Biochemistry* **1996**, *35*, 4034-4045; b) C.-Y. Chen, T.-P. Ko, T.-W. Lin, C.-C. Chou, C.-J. Chen, A. H.-J. Wang, *Nucleic Acids Res.* **2005**, *33*, 430-438; c) Y.-G. Gao, S.-Y. Su, H. Robinson, S. Padmanabhan, L. Lim, B. S. McCrary, S. P. Edmondson, J. W. Shriver, A. H.-J. Wang, *Nat. Struct. Biol.* **1998**, *5*, 782-786; d) V. Kalichuk, G. Béhar, A. Renodon-Cornière, G. Danovski, G. Obal, J. Barbet, B. Mouratou, F. Pecorari, *Sci. Rep.* **2016**, *6*, 37274; e) T.-P. Ko, H.-M. Chu, C.-Y. Chen, C.-C. Chou, A. H.-J. Wang, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1381-1387; f) J. G. McAfee, S. P. Edmondson, P. K. Datta, J. W. Shriver, R. Gupta, *Biochemistry* **1995**, *34*, 10063-10077; g) H. Robinson, Y.-G. Gao, B. S. McCrary, S. P. Edmondson, J. W. Shriver, A. H.-J. Wang, *Nature* **1998**, *392*, 202-205; h) R. K. Singh, A. Mukherjee, *J. Phys. Chem. B* **2022**, *126*, 1682-1690; i) J. Spiriti, A. van der Vaart, *Chembiochem* **2013**, *14*, 1434-1437; j) M. Zacharias, *Angew. Chem. Int. Ed.* **2019**, *58*, 5967-5972; *Angew. Chem.* **2019**, *131*, 6028-6033.

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

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

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

[11] V. Corvaglia, J. Wu, D. Deepak, M. Loos, I. Huc, *Chemarxiv.*, DOI: 10.26434/chemrxiv-2023-fz84h. This content is a preprint and has not been peer-reviewed.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[26] V. B. Chen, J. R. Wedell, R. K. Wenger, E. L. Ulrich, J. L. Markley, *J. Biomol. NMR* **2015**, *63*, 77-83.
[27] A. A. McCarthy, R. Barrett, A. Beteva, H. Caserotto, F. Dobias, F. Felisaz, T. Giraud, M. Guijarro, R. Janocha, A. Khadrouche, M. Lentini, G. A. Leonard, M. Lopez Marrero, S. Malbet-Monaco, S. McSweeney, D. Nurizzo, G. Papp, C. Rossi, J. Sinoir, C. Sorez, J. Surr, O. Svensson, U. Zander, F. Cipriani, P. Theveneau, C. Mueller-Dieckmann, *J. Synchrotron Radiat.* **2018**, *25*, 1249-1260.

Supporting Information

1 List of Abbreviations

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

2 Supporting figures

2.1 Folding principle and DNA mimic foldamer design

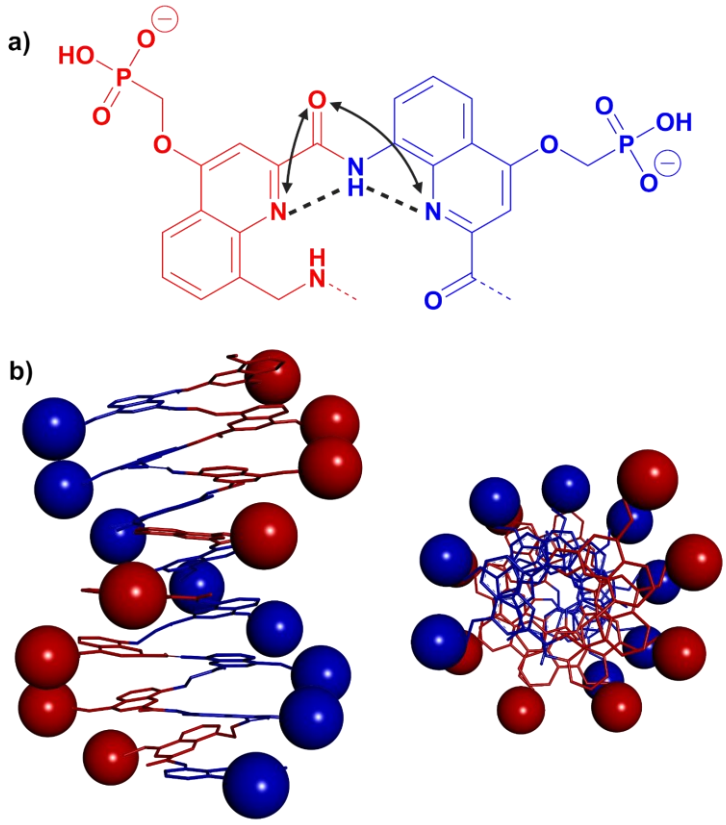


Figure S1. a) The folding principle of DNA mimic foldamers based on a MQ dimer. Dashed lines indicate hydrogen bonds, and arrows highlight electrostatic repulsions. These forces impart curvature to the main chain. Hydrophobic effects associated with aromatic stacking further stabilize the helical conformations.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

2.2 SPR binding studies

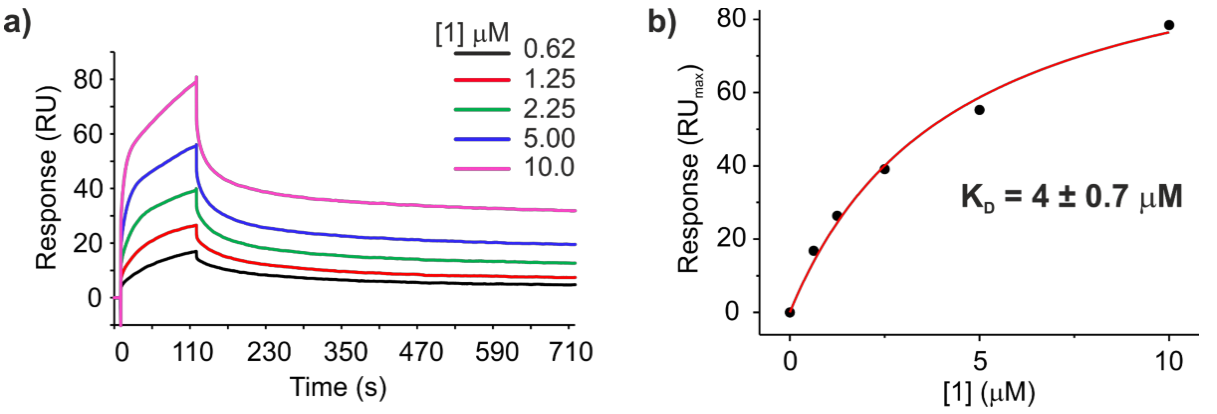


Figure S2. Binding of **1** to Sac7d quantified by SPR. a) SPR sensorgrams of the interaction between His_6-tagged Sac7d and **1** in HBS-EP buffer pH 7.4 at 25 °C. b) Curve fitting of the experimental data representing the maximum response unit values (steady state) plotted against foldamer concentration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

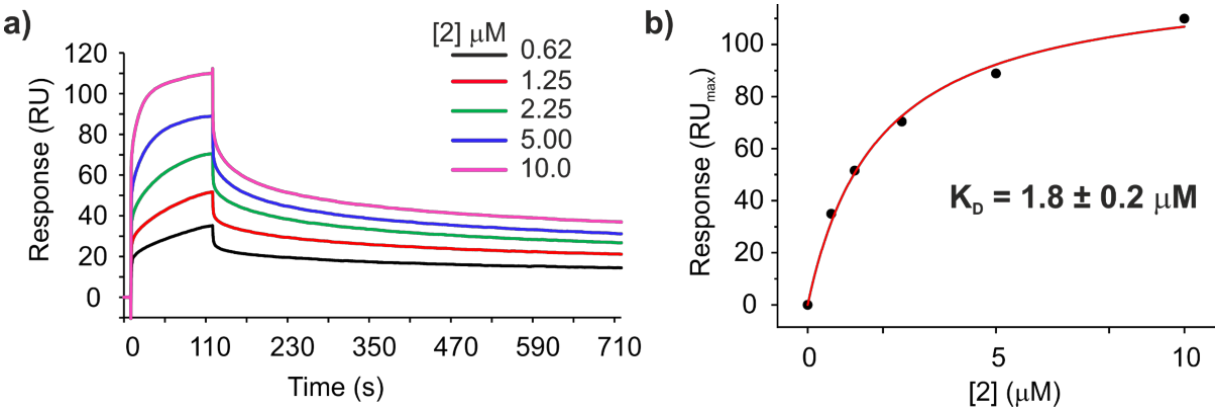


Figure S3. Binding of **2** to Sac7d quantified by SPR. a) SPR sensorgrams of the interaction between His_6-tagged Sac7d and **2** in HBS-EP buffer pH 7.4 at 25 °C. b) Curve fitting of the experimental data representing the maximum response unit values (steady state) plotted against foldamer concentration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

2.3 CD binding studies

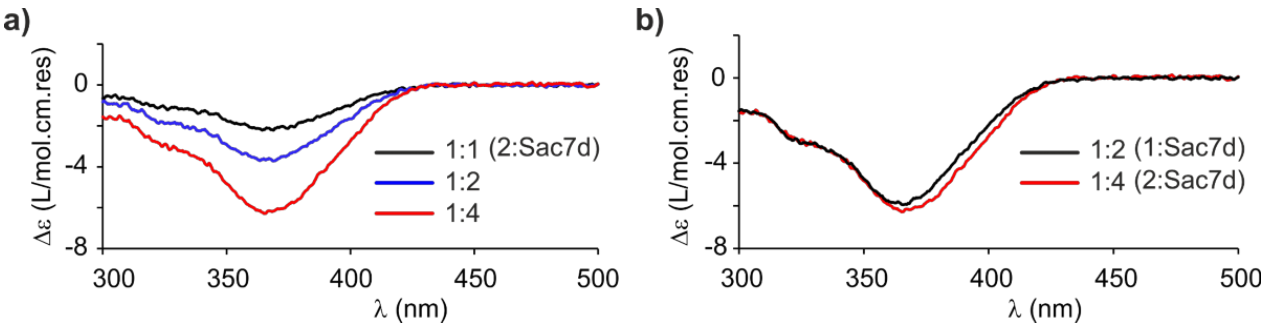


Figure S4. CD spectra of **2** in complex with Sac7d at 1:1, 1:2, and 1:4 foldamer: protein ratio (a) and of **1** and **2** overlaid (b) recorded in 50 mM NH_4HCO_3 buffer pH 8.5 at 20 °C after 24 h equilibration.

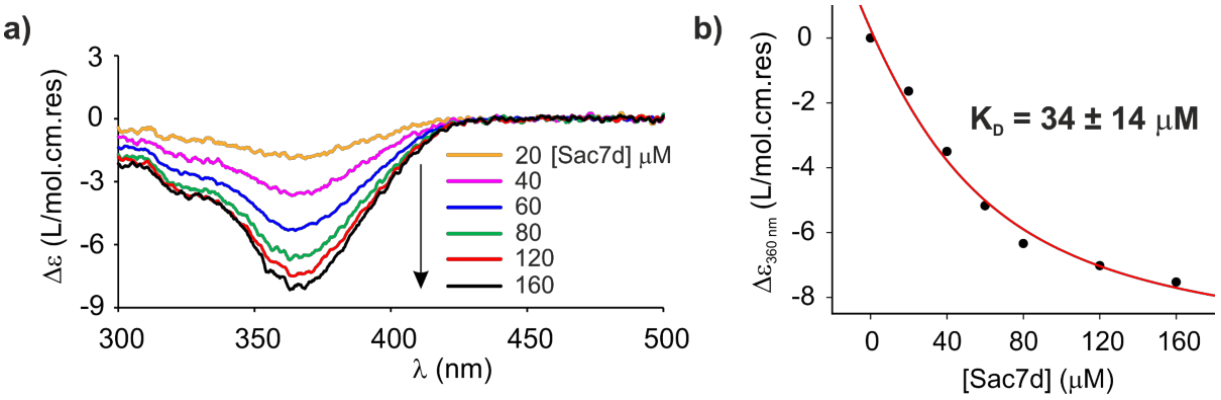


Figure S5. Binding of **1** to Sac7d quantified by CD. a) CD spectra of **1** (40 μM) in complex with Sac7d at different concentrations (from 20 to 160 μM) in 50 mM NH_4HCO_3 buffer pH 8.5 at 20 °C after 24 h equilibration. b) Curve fitting of the experimental data representing the maximum Δε values plotted against protein concentration to determine K_d following the quadratic binding equation.[15] K_d was found to be consistent across replicates.

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

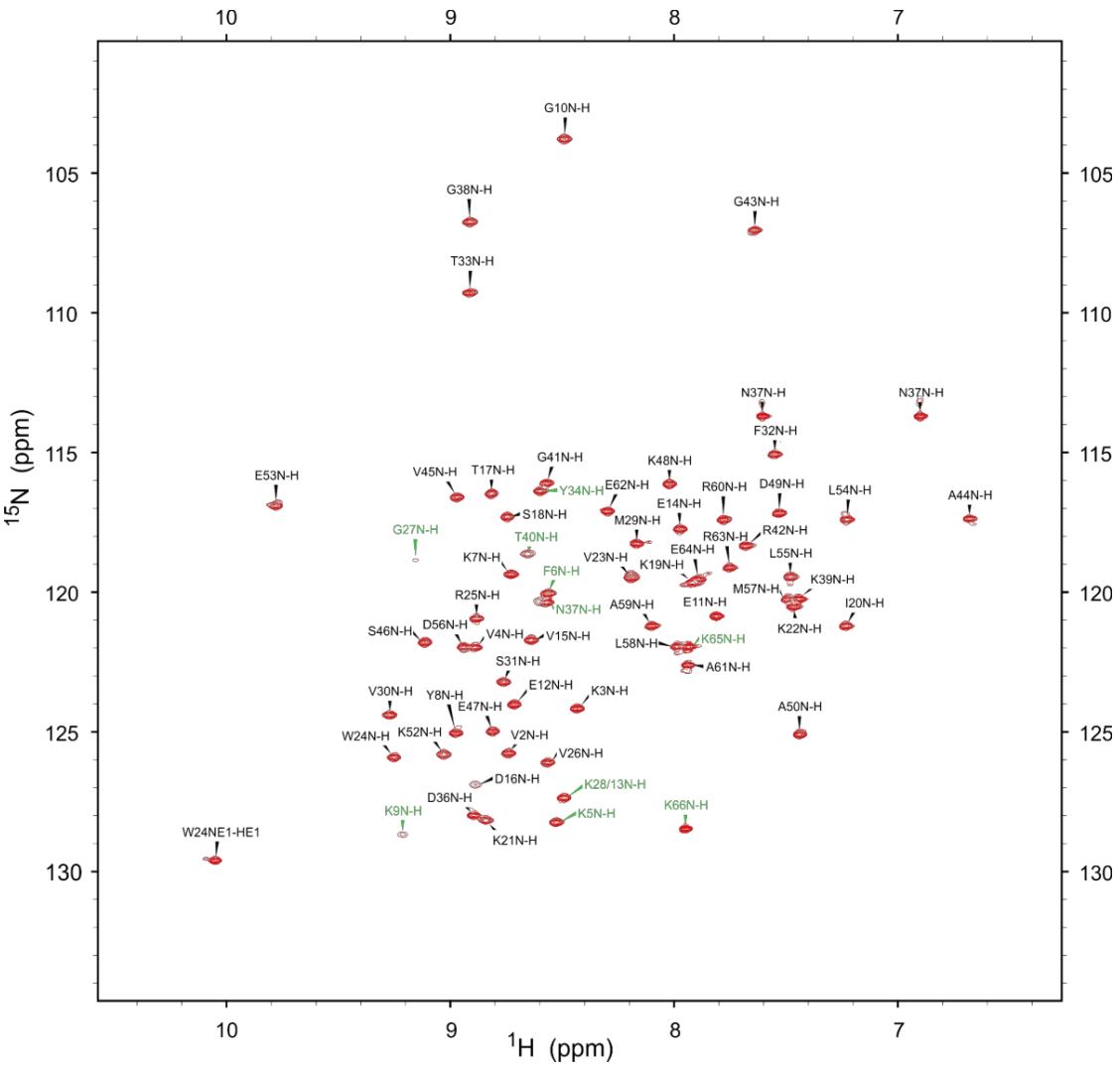


Figure S6. ^{1}H-^{15}N HSQC spectrum of [^{15}N]-Sac7d (500 µM) in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O. Assigned backbone amide signals were labeled with a single amino acid letter code followed by their position. In green are amide signals for which the assignment was not definite from our experiments. These signals were assigned by analogy with previously reported signals having similar chemical shift values and.[16]

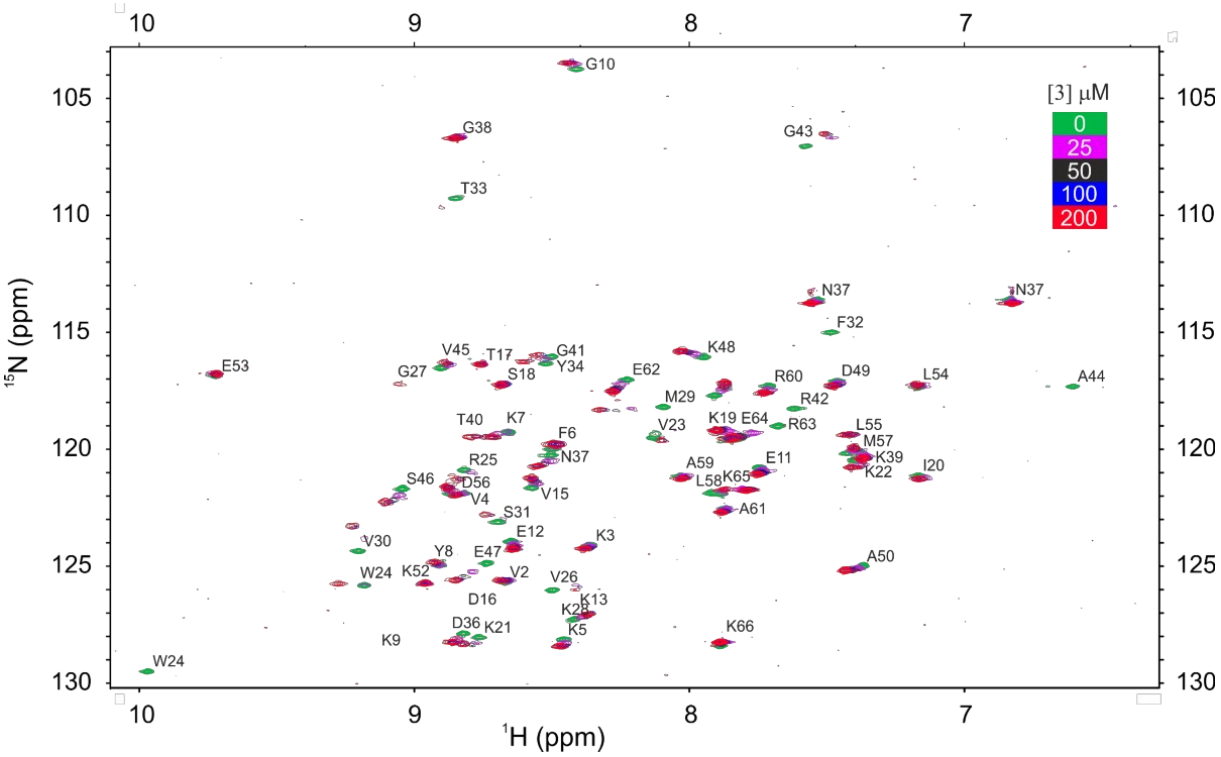


Figure S7. ^{1}H-^{15}N HSQC titration of [^{15}N]-Sac7d (100 µM) with **3** in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O. The colored scale indicates that an increasing concentration of **3** was added to Sac7d.

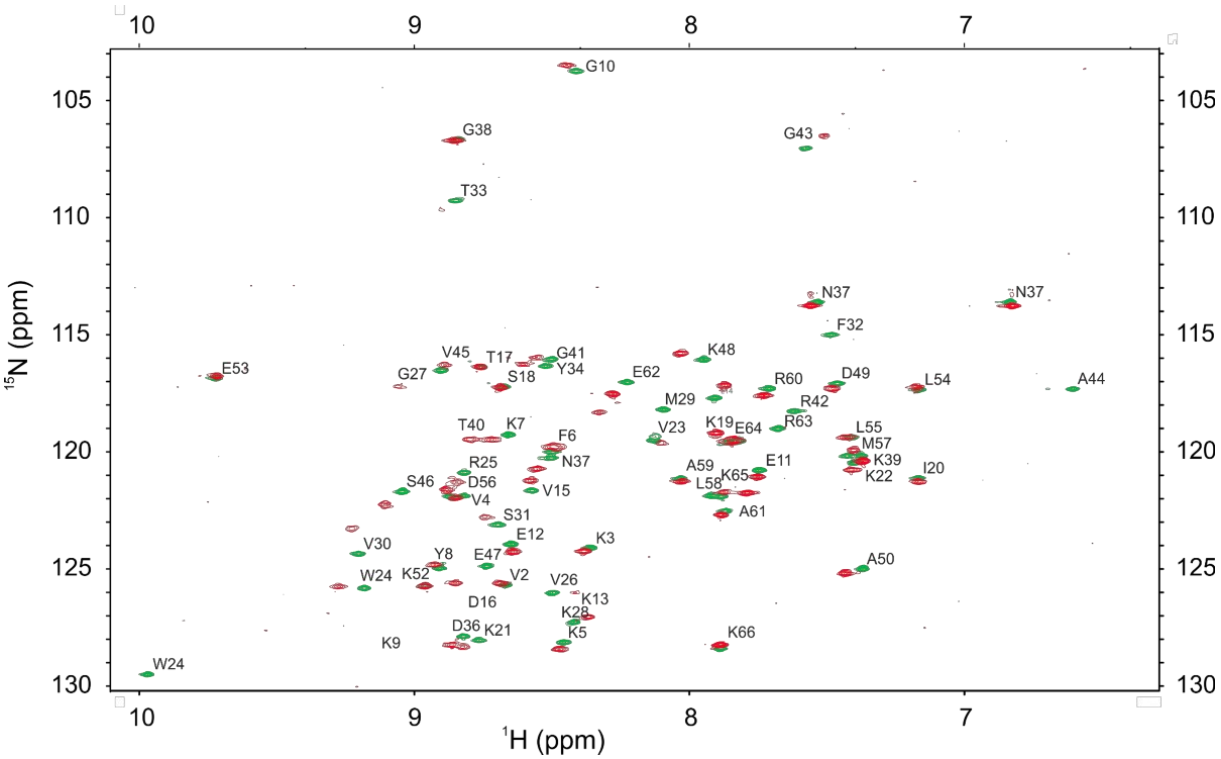


Figure S8. Superimposition of ^{1}H-^{15}N HSQC spectra of [^{15}N]-Sac7d (100 μM, green amide signals) and [^{15}N]-Sac7d with 2.0 eq. of **3** (200 μM, red amide signals) in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O.

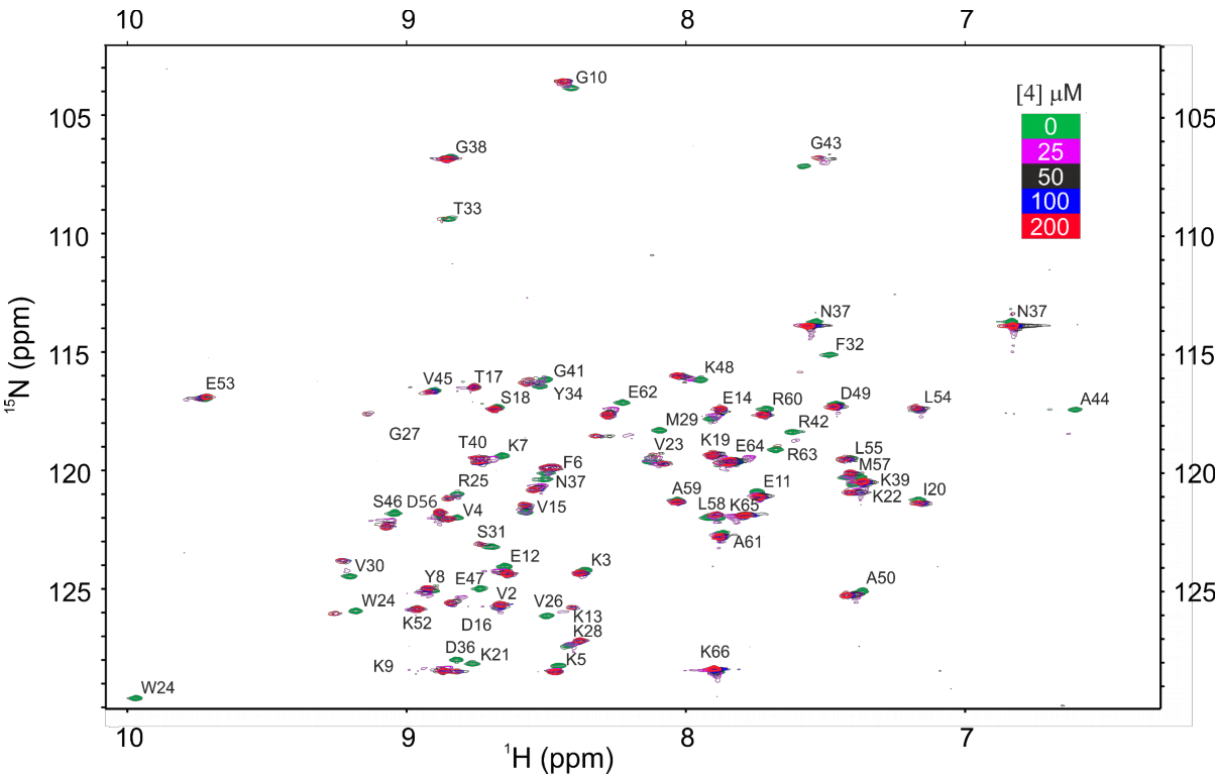


Figure S9. ^{1}H-^{15}N HSQC titration of [^{15}N]-Sac7d (100 µM) with **4** in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O. The colored scale indicates that an increasing concentration of **4** was added to Sac7d.

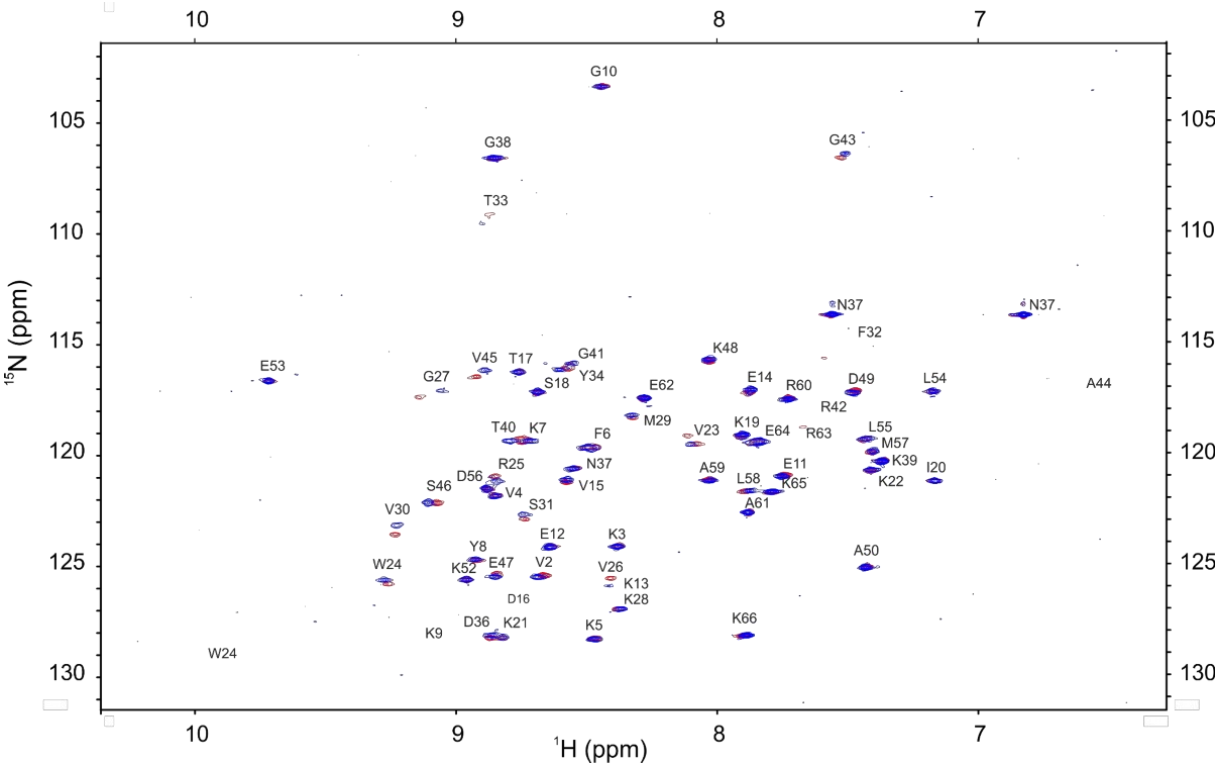


Figure S10. Superimposition of ^{1}H-^{15}N HSQC spectra of [^{15}N]-Sac7d (100 µM) with **3** (200 µM, blue amide signals) and **4** (200 µM, red amide signals) in Tris-d_{11}-HCl buffer pH 7.5, 50 mM KCl, 10% D_2O.

2.5 X-ray crystallography

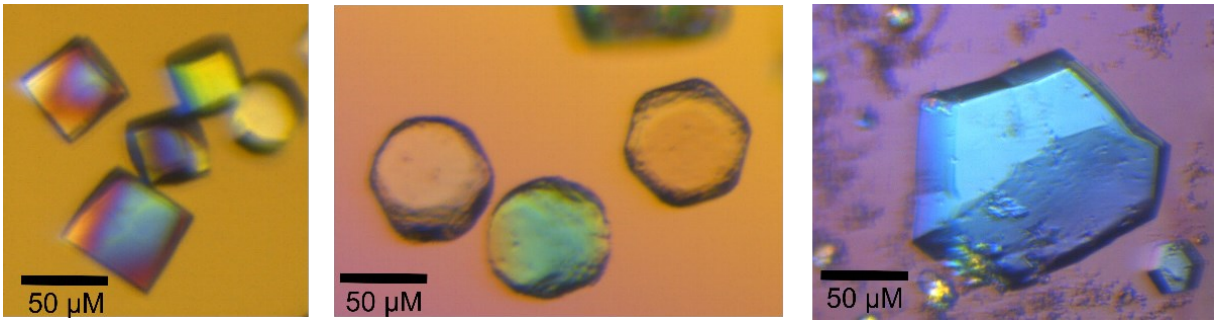


Figure S11. Crystals of Sac7d-**3**, Sac7d V26A/M29A-**3**, and Sac7d V26A/M29A-**4** from left to right observed under crossed polarizing microscope.

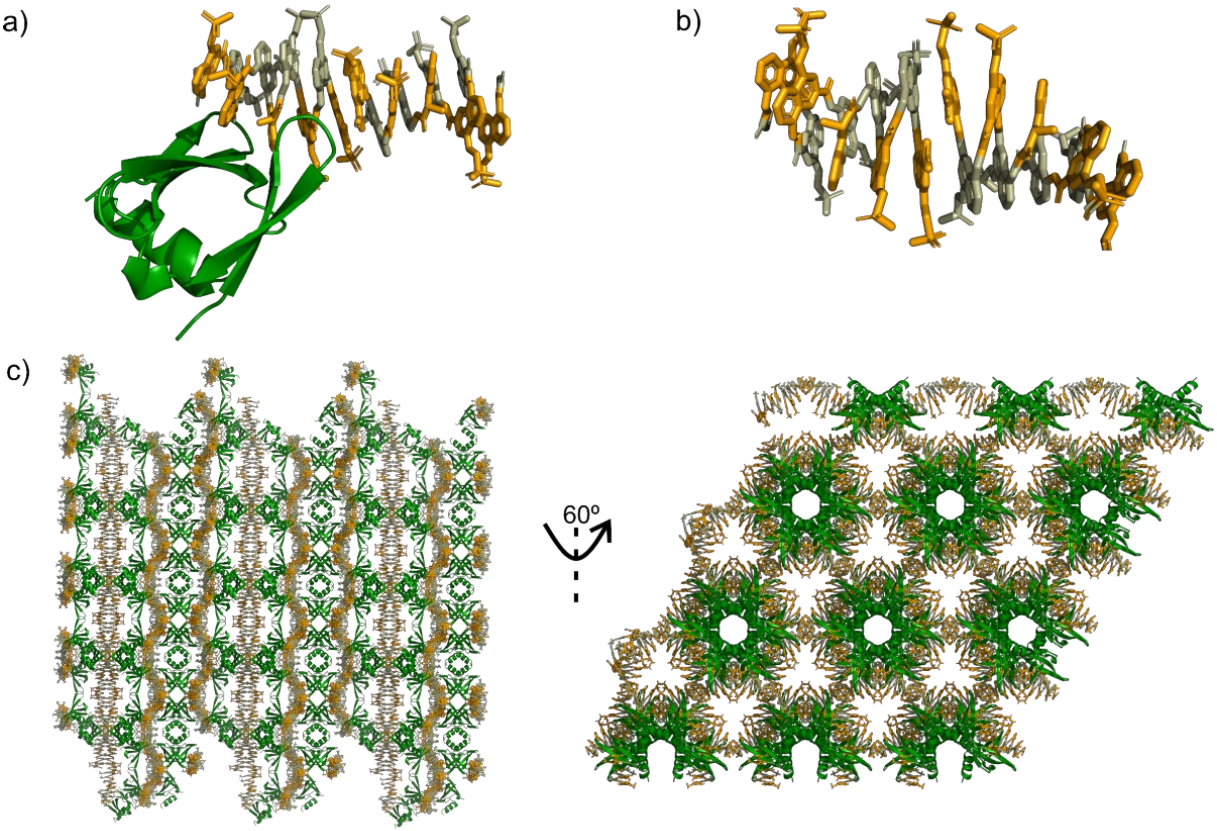


Figure S12. Sac7d-**3** crystal structure (PDB# 8CMN). a) Asymmetric unit showing Sac7d with two 9mer units from two 18mers of **3** in the ASU. b) C-terminal stacking of the two 9mer units in ASU of **3**. c) Two views of the Sac7d-**3** packing in the crystal lattice.

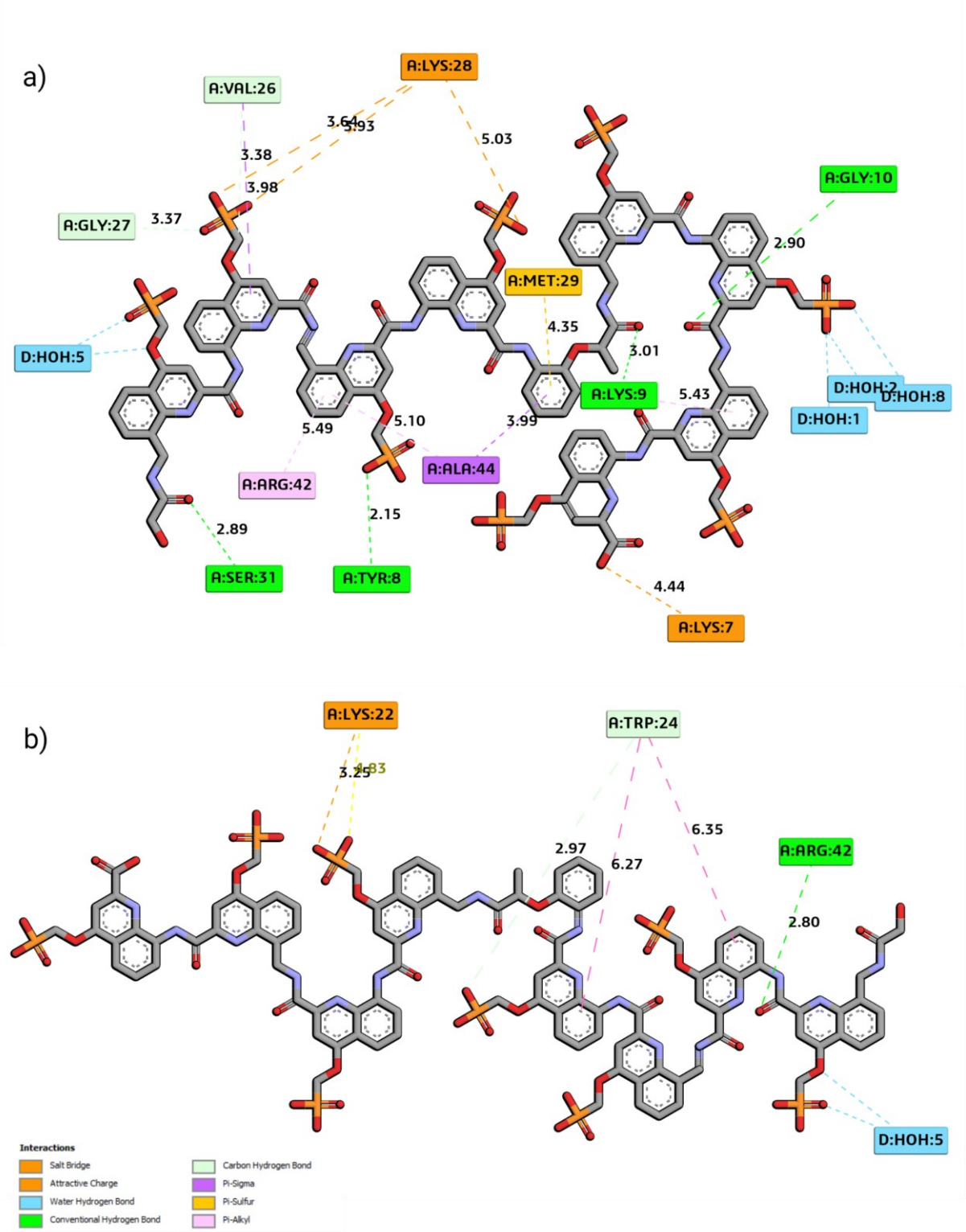


Figure S13. Ligplot-like representation of crystal contacts between Sac7d-**3**. a) Sac7d (chain A) interactions with **3** (chain B, 9mer). b) Sac7d (chain A) interaction with **3** (chain B' 9mer). The interaction map was generated using Discovery Studio (Dassault Systems). **3** is shown in 2D stick representation, colored according to elements.

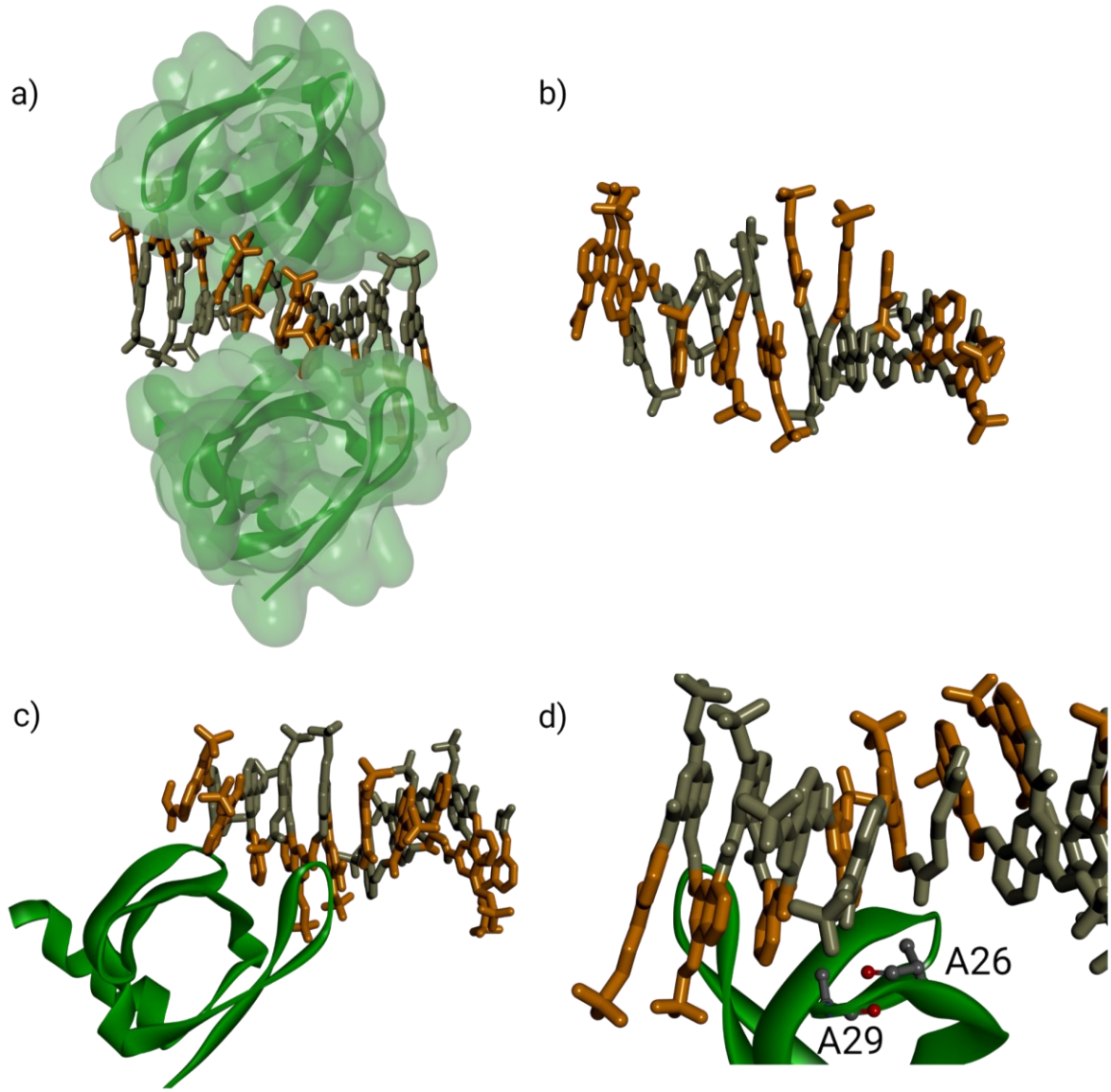


Figure S14. Sac7d V26A/M29A-**3** crystal structure (PDB# 8Q2M). a) Biological unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **3** in the ASU b) C-terminal stacking of the two 9mer units in ASU of **3**. c) Asymmetric unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **3** in the ASU. d) Contacts between **3** and A26 and A29 residues (mutated from V26 and M29) of Sac7d.

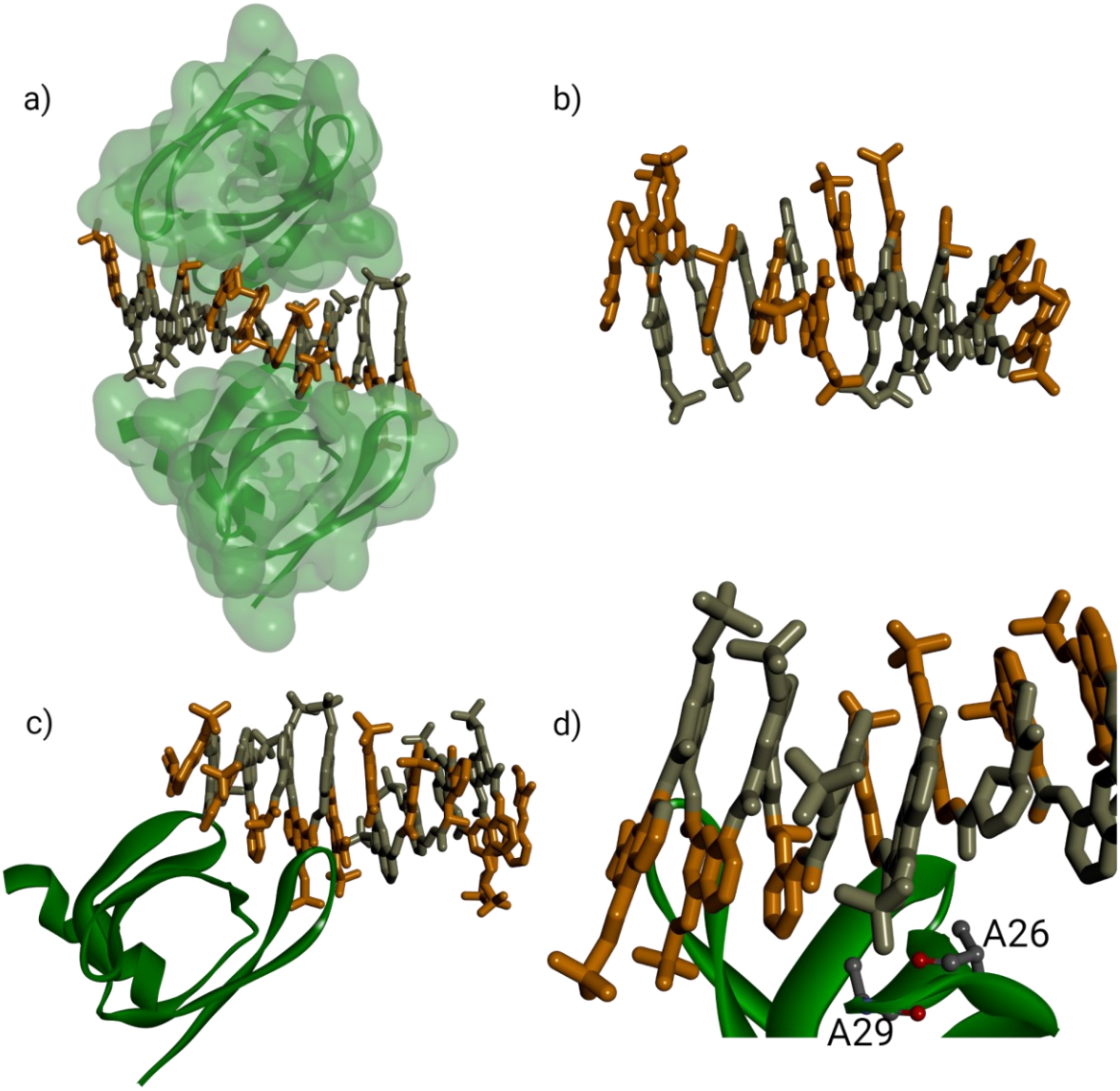


Figure S15. Sac7d V26A/M29A-**4** crystal structure (PDB# 8QPC). a) Biological unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **4** in the ASU b) C-terminal stacking of the two 9mer units in ASU of **4**. c) Asymmetric unit of Sac7d V26A/M29A with two 9mer units from two 18mers of **4** in the ASU. d) Contacts between **4** and A26 and A29 residues (mutated from V26 and M29) of Sac7d.

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

3.1 General

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

Size-exclusion chromatography (SEC) was performed in a cold cabinet (Unichromat 1500) maintained at 16°C on a Knauer fast protein liquid chromatography (FPLC) system coupled with a HiLoad® 16/600 Superdex® 75 pg column (Cytiva, 28989333). Protein elution was monitored by UV detection at 280 and 205 nm with a diode array detector and fractions were collected by a Foxy R1 Fraction Collector (Teledyne ISCO). Protein concentration and buffer exchange were performed using spin concentrators (Amicon, 3 kDa MWCO). Dialysis was performed using 3 kDa MWCO SlideA-Lyzer™ G2 dialysis cassettes (ThermoFisher Scientific).

pH of the buffers was adjusted using HCl on a Mettler Toledo™ SevenCompact pH meter. Ammonium-15N chloride with isotropic purity of ≥98 atom % ^{15}N (299251) was purchased from Sigma Aldrich. Tris-d_{11} (14978) was purchased from Deutrero.

3.2 Foldamer synthesis

Foldamers used in this study were prepared as previously described.[11]

3.3 SPR spectroscopy

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

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

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

3.4 CD spectroscopy

Circular Dichroism (CD). CD data were recorded on a Jasco J-810 spectrometer with 1- or 2-mm quartz cuvettes. Scans were acquired at 20 °C, over the 300—500 nm range, with a 1 s response time and a 100 nm/min scanning speed. CD spectra were an average of three scans and were baseline-corrected for signal contributions due to the buffer and protein. The exact concentration of Sac7d and **1** and **2** stock solutcons were determined using NanoDrop™ One^C (Thermo Fisher Scientific). Except for the DNA competition experiments, the raw CD data in millidegrees ($mdeg$) were converted to molar extinction ($\Delta\varepsilon$) per number of Q or M quinoline residues (*res*) by the following equation (2):

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

For binding studies, CD spectra of **1** and **2** in complex with 40 μM Sac7d at foldamer: protein ratio of 1:1, 1:2, and 1:4 (the latter only for **2**) were recorded in 50 mM NH_4HCO_3 buffer pH 8.5 after 24 h equilibration. To evaluate the binding affinity, seven samples were prepared in 50 mM NH_4HCO_3 buffer pH 8.5 containing 40 μM **1** and Sac7d at different concentrations (0, 20, 40, 60, 80, 120, 160 μM) and equilibrated for 24 h. For K_d calculation, maximum $\Delta\varepsilon$ values at 360 nm were plotted against protein concentration and fitted to the quadratic binding equation. K_d was found to be consistent across replicates and of the same order of magnitude as that determined through SPR measurements. DNA-binding competition experiments were performed in 50 mM NH_4HCO_3 buffer pH 8.5 using DNA decamer CCTATATAGG previously crystallized with Sac7d.[7e] To the equilibrated sample containing **1** and Sac7d (1:1, 40 μM), 10 bp DNA was added (1, 2, 4, 8, 16 eq.). Increasing amounts of DNA were added

in the same sample containing the pre-formed complex of **1** and Sac7d waiting for 24 h equilibration between additions. The quantitative analysis of the competing experiment was done as explained below. We use the following abbreviations: S for Sac7d, D for the DNA duplex, F for the foldamer, M for the foldamer *M* helix and P for the foldamer *P* helix, SM for the Sac7d-*M* helix complex, and SD for the Sac7d-DNA complex.

We then define association constants:

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

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

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

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

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

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

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

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

Using the equations above, we can write:

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

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

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

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

3.5 NMR spectroscopy

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

Protein backbone assignment. The concentration of ^{15}N-labeled Sac7d ([^{15}N]-Sac7d) used was 500 µM and spectra were measured in a 5 mm NMR tube (Wilmad®). First, a ^1H-^{15}N HSQC spectrum was recorded using a standard pulse sequence including a watergate sequence with water flip-back pulses from the Bruker pulse sequence library (hsqcfpf3gpphwg). The sweep widths were 12800 (^1H) x 3333 (^{15}N) and 2K data points were collected for 512 increments of 4 scans per fid. Protein backbone assignment of [^{15}N]-Sac7d residues was performed based on already reported data[16] and with the help of additional (data not shown) 3D NMR spectra (two ^{15}N-TOCSY-HSQCs with different TOCSY mixing times, and one ^{15}N-NOESY-HSQC) on ^{15}N-labeled Sac7d and 2D NOESY on unlabeled Sac7d. ^{15}N-TOCSY-HSQC (mlevhsqcetf3gp3d) was collected as a series of 40 complex (t_2) data sets composed of 128 complex (t_3) values and 2K data points with 8 scans per fid. The mixing times were set to either 60 or 100 ms ^{15}N-NOESY-HSQC (noesyhsqcfpf3gpsi3d) was collected as a series of 40 complex (t_2) data sets composed of 128 complex (t_3) values and 2K data points with 16 scans per fid. The noe mixing time was set to 120 ms 2D NOESY (noesyesgpph) on unlabeled Sac7d was measured with a sweep width of 12800 and 1K data points collected with 512 increments of 64 scans per fid. The noe mixing time was set to 120 ms All spectra were rudimentary processed with standard processing parameters from the Topspin (Bruker) processing library additionally applying zero filling to yield symmetrical either 3D or 2D matrices and then converted to UCSF format and further subjected to procession and evaluation with the open source software NMRFAM-SPARKY (version 3.190). This software allowed us to assign the Sac7d amide backbone through synchronization of 2D and 3D NMR spectra described above.

Chemical shift perturbation. Titrations were performed on independent samples of 100 µM ^{15}N-labeled Sac7d which were prepared to a final volume of 210 µL and titrated with 1—8 µL aliquots of 5 mM **3** or **4**. Each sample comprised 210 µL by supplementing

the missing volume (after foldamer addition) with purified water to obtain a uniform concentration of 100 µM [^{15}N]-Sac7d in each NMR tube (see the following Table S1 for details). The prepared samples containing different concentrations of foldamer were kept at r.t. for 24 h to ensure equilibrium and were then transferred to 3 mm NMR tubes (Wilmad®). ^{1}H-^{15}N HSQC spectra for each foldamer concentration were then recorded with the same parameter set as for the backbone assignment mentioned above. All spectra were first processed using standard processing in Topspin and then converted to UCSF format. Chemical shift perturbations (CSPs, $\Delta\delta_H$ and $\Delta\delta_N$) were determined using CSP analysis module in CcpNmr AnalysisAssign version 3.0.4 for Windows (http://www.ccpn.ac.uk/v3-software/downloads).[17]The obtained $\Delta\delta_H$ and $\Delta\delta_N$ values for the protein residues were then subjected to the equation $((\Delta\delta_H)/0.14)^2+(\Delta\delta_N)^2)^{0.5}$ in order to create a residue related CSP bar graph in Microsoft Excel®.

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

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

3.6 Protein production

Expression and purification of Sac7d and Sac7d V26A/M29A proteins. The pET3b plasmid harboring Sac7d and Sac7d V26A/M29A gene was kindly received from Prof. Chin-Yu Chen.[7b] The pET3b-Sac7d plasmid was transformed into Escherichia coli BL21DE3 plys S cells. The cells were grown in Luria Broth to an OD at 600 nm of 0.8 and induced with 0.4 mM Isopropyl β-d-1-thiogalactopyranoside (IPTG) for 1 h induction at 37 °C. Afterward, cells were spun at 4000 g for 15 min. The supernatant was resuspended and sonicated in 20 mM Tris-HCl pH 7.5 supplemented with protease inhibitor cocktail (Thermo Scientific™ Halt™ Protease Inhibitor Cocktail) and 2 mM EDTA. This was then heated in a water bath at 68 °C for 30 min and centrifuged

at 105000 g for 1 h. The supernatant was then loaded on SP cation exchange column (KNAUER Wissenschaftliche Geräte GmbH) and eluted with 20 mM Tris-HCl pH 7.5, 1000 mM NaCl. The purity was monitored by SDS-PAGE and LC-ESI-MS. In the end, pure Sac7d fractions were dialyzed against 20 mM Tris-HCl pH 7.5 and concentrated to 2.6 mM for crystallization. Sac7d V26A/M29A protein was expressed and purified using the same protocol described above for Sac7d.

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

3.7 Crystallization

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

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

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

3.8 Data collection and structure refinement

Sac7d-3. X-ray diffraction data were collected from cryo-protected crystals at 100 K on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble) on a Dectris Eiger2 X 16M detector.[18] The collected data set was processed using XDS[19], and the structure was solved by MoRDa[20] using Affitin h4[21] (PDB ID 4CJ2)

as a search model for protein and Phaser MR[22] using the energy minimized molecular model of **3**[11] (prepared on Maestro V11.5, Schrödinger) as the search model for the foldamer. The structure was refined in iterative rounds of manual model building and refinement in Coot[23] and Phenix Refine.[24] ProDRG[25] was used to generate restrain file for **3**. Foldamer was further adjusted in visible electron density at the final refinement stage. The structure was validated with MolProbity[26] and deposited in PDB under accession code 8CMN. Data collection and structure refinement statistics for Sac7d in complex with **3** are given in table S3.

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

Sac7d V26A/M29A-3. X-ray diffraction data were collected from cryo-protected crystals at 100 K on beamline P13 macromolecular X-ray crystallography at the Deutsches Elektronen Synchrotron (DESY) at European Molecular Biology Laboratory (EMBL, Hamburg) on an EIGER 16M detector.[18] The structure was solved by Phaser MR[22] using PDB 8CMN as a search model for protein and molecular model of **3** from PDB #8CMN. The structure was refined in iterative rounds of manual model building and refinement in Coot[23] and Phenix Refine.[24] ProDRG[25] was used to generate restrain file for **3**. Foldamer was further adjusted in visible electron density at the final refinement stage. The structure was validated with MolProbity[26] and deposited in PDB under accession code 8Q2M. Data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **3** are given in table S4.

Sac7d V26A/M29A-4. X-ray diffraction data were collected from cryo-protected crystals at 100 K on beamline ID30-B at the European Synchrotron Radiation Facility (ESRF, Grenoble) on a Dectris Eiger2 X 16M detector.[27] The structure was solved by Phaser MR[22] using PDB 8Q2M as a search model for protein. An energy minimized molecular model of **4**[11] (prepared on Maestro V11.5, Schrödinger) was used for molecular replacement of the foldamer. The structure was further refined in iterative rounds of manual model building and refinement in Coot[23] and Phenix Refine.[24] ProDRG[25] was used to generate restrain file for **4**. Foldamer was further adjusted in visible electron density at the final refinement stage. The structure was validated with MolProbity[26] and deposited in PDB under accession code 8QPC. Data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **4** are given in table S5.

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

Parameters	Value
Wavelength	0.8856
Resolution range	42.34 – 2.65 (2.745-2.65)
Space group	P 6_4 2 2
Unit cell	71.41 71.41 116.17 90 90 120
Total reflections	37941 (3816)
Unique reflections	5508 (531)
Multiplicity	6.9 (7.2)
Completeness (%)	99.62 (99.62)
Mean I/sigma (I)	18.81 (2.13)
Wilson B-factor	74.99
R-merge	0.05791 (0.7586)
R-meas	0.06293 (0.818)
R-pim	0.02345 (0.2956)
CC1/2	0.998 (0.81)
CC*	1 (0.946)
Reflections used in refinement	5502 (531)
Reflections used for R-free	547 (52)
R-work	0.2613 (0.4296)
R-free	0.2979 (0.5179)
CC (work)	0.892 (0.718)
CC (free)	0.919 (0.313)
Number of non-hydrogen atoms	848
macromolecules	495
ligands	346
solvent	7
Protein residues	63
RMS (bonds)	0.008
RMS (angles)	2.69
Ramachandran favored (%)	98.36
Ramachandran allowed (%)	1.64
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clash score	17.86
Average B-factor	66.50
macromolecules	67.78
ligands	64.68
solvent	65.40
Number of TLS groups	4

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

Parameters	Value
Wavelength	0.9763
Resolution range	30.71 - 3.212 (3.327 - 3.212)
Space group	P 6_4 2 2
Unit cell	70.92 70.92 119.29 90 90 120
Total reflections	119569 (12369)
Unique reflections	3226 (310)
Multiplicity	37.1 (39.9)
Completeness (%)	99.54 (100.00)
Mean I/sigma (I)	23.56 (3.51)
Wilson B-factor	115.54
R-merge	0.08561 (1.222)
R-meas	0.08692 (1.238)
R-pim	0.01467 (0.1939)
CC1/2	1 (0.952)
CC*	1 (0.988)
Reflections used in refinement	3219 (310)
Reflections used for R-free	322 (31)
R-work	0.3238 (0.3609)
R-free	0.3267 (0.3914)
CC (work)	0.885 (0.823)
CC (free)	0.963 (0.921)
Number of non-hydrogen atoms	836
macromolecules	490
ligands	346
solvent	0
Protein residues	63
Nucleic acid bases	
RMS (bonds)	0.035
RMS (angles)	4.96
Ramachandran favored (%)	95.08
Ramachandran allowed (%)	4.92
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clash score	34.56
Average B-factor	132.72
macromolecules	145.47
ligands	114.67
Number of TLS groups	3

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

Parameters	**Value**
Wavelength	0.871
Resolution range	61.41 - 3.241 (3.356 - 3.241)
Space group	P 6_4 2 2
Unit cell	70.907 70.907 121.997 90 90 120
Total reflections	51589 (5784)
Unique reflections	2973 (244)
Multiplicity	17.4 (19.3)
Completeness (%)	89.51 (81.27)
Mean I/sigma (I)	9.81 (1.96)
Wilson B-factor	91.41
R-merge	0.1249 (1.559)
R-meas	0.1296 (1.601)
R-pim	0.03345 (0.3614)
CC1/2	0.999 (0.938)
CC*	1 (0.984)
Reflections used in refinement	2875 (243)
Reflections used for R-free	139 (9)
R-work	0.3356 (0.3776)
R-free	0.3470 (0.4302)
CC (work)	0.862 (0.927)
CC (free)	0.989 (0.773)
Number of non-hydrogen atoms	812
macromolecules	462
ligands	350
solvent	0
Protein residues	63
RMS (bonds)	0.028
RMS (angles)	3.81
Ramachandran favored (%)	95.08
Ramachandran allowed (%)	4.92
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clash score	19.31
Average B-factor	123.77
macromolecules	143.17
ligands	98.16
Number of TLS groups	3

4 References (in continuation of the main text)

[15] I. Jarmoskaite, I. AlSadhan, P. P. Vaidyanathan, D. Herschlag, *eLife* **2020**, *9*, e57264.
[16] J. L. Bedell, B. S. McCrary, S. P. Edmondson, J. W. Shriver, *Protein Sci.* **2000**, *9*, 1878-1888.
[17] L. Mureddu, G. W. Vuister, *FEBS J.* **2019**, *286*, 2035-2042.
[18] D. Nurizzo, T. Mairs, M. Guijarro, V. Rey, J. Meyer, P. Fajardo, J. Chavanne, J. C. Biasci, S. McSweeney, E. Mitchell, *J. Synchrotron Radiat.* **2006**, *13*, 227-238.
[19] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125-132.
[20] A. Vagin, A. Lebedev, *Acta Cryst.* **2015**, *A71*, s19.
[21] A. Correa, S. Pacheco, A. E. Mechaly, G. Obal, G. Béhar, B. Mouratou, P. Oppezzo, P. M. Alzari, F. Pecorari, *PloS One* **2014**, *9*, e97438.
[22] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
[23] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr. D Biol. Crystallogr. D* **2010**, *66*, 486-501.
[24] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr. D Biol. Crystallogr.* **2019**, *75*, 861-877.
[25] A. W. Schüttelkopf, D. M. F. van Aalten, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1355-1363.
[26] V. B. Chen, J. R. Wedell, R. K. Wenger, E. L. Ulrich, J. L. Markley, *J. Biomol. NMR* **2015**, *63*, 77-83.
[27] A. A. McCarthy, R. Barrett, A. Beteva, H. Caserotto, F. Dobias, F. Felisaz, T. Giraud, M. Guijarro, R. Janocha, A. Khadrouche, M. Lentini, G. A. Leonard, M. Lopez Marrero, S. Malbet-Monaco, S. McSweeney, D. Nurizzo, G. Papp, C. Rossi, J. Sinoir, C. Sorez, J. Surr, O. Svensson, U. Zander, F. Cipriani, P. Theveneau, C. Mueller-Dieckmann, *J. Synchrotron Radiat.* **2018**, *25*, 1249-1260.

5. Sac7d-foldamer adducts: a tethering approach

Biomolecules, despite their remarkable complexity, are made from a relatively limited set of elementary building blocks, such as amino acids for proteins or nucleotides for nucleic acids. The spatial arrangement of these building blocks, which determines the shape and function of biomolecules, is encoded by the sequential arrangement of these monomers. This coding pattern is a rich source of inspiration for chemists engaged in foldamer research. Foldamers also utilize monomeric building blocks to program the three-dimensional structure and induce specific functions in oligomers. Interactions between folded molecules offer several advantages, particularly in their ability to cover extensive surfaces. These interactions can be homotypic (e.g., between proteins)[175–178] or heterotypic (e.g., between proteins and DNA)[179–182]. In this chapter, our objective is to design heterotypic interactions between protein surfaces and aromatic foldamers. Synthetic foldamers (artificial folded oligomers) have undergone rapid development driven mainly by curiosity and the hope that they could achieve functions that match or go beyond those of biopolymers.[137] Numerous and diverse molecular backbones that adopt well-defined folded conformations have been described. Among these, aromatic amide foldamers[183,184] stand out with their exceptionally predictable, tunable, and stable conformations in aqueous solution, relatively easy synthesis of secondary and tertiary-like objects as large as small proteins[185,186], and high amenability to crystal growth and structural elucidation. These features all point to aromatic amide foldamers as potential scaffolds that could eventually inhibit protein-DNA interactions. Earlier efforts towards protein surface ligand design have concerned peptidomimetics that mimic protein epitopes such as α-helices or larger motifs.[187,188] Recently, Ivan Huc and colleagues have developed foldamer-based mimics of the charged surface of B-DNA, successfully inhibiting protein-DNA interactions.[147] However, it is worth mentioning that these foldamers were found to bind and inhibit only a specific subset of proteins.[147]

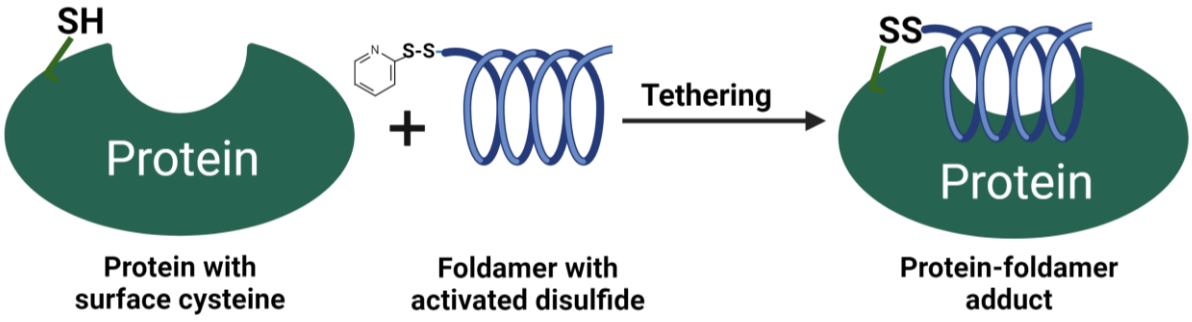


Fig. 30. Sketch of a protein-foldamer adduct with tethering approach.

In this chapter, we address the unsolved problem of ab initio design of ligands for targeting protein-DNA interactions. As a key concept, we aim to obtain structural information about

interactions at a DNA mimic foldamer-protein interface even in the absence of strong binding, providing some attachment links between the two molecules. This involves the covalent tethering of foldamers to target proteins via disulfide bridges between a cysteine residue introduced at the protein surface by site-directed mutagenesis and a pyridyl disulfide moiety on the foldamer (**Fig. 30**).[189] Analysis of adduct on mass spectrometry and structural studies using crystallography will reveal the atomic level details of the foldamer-protein interactions before any tight or very selective binding has been achieved. Enhancing interactions through structure-based iterative improvements would finally allow us to remove the tether and create and characterize the desired non-covalent foldamer-protein complex in the tethered protein-foldamer complex. The global objective in this chapter is to explore molecular recognition rules between large (4-5 kDa) aromatic oligoamide DNA mimic foldamers and protein surfaces and validate a novel iterative method based on combining covalent attachment of a foldamer to a model protein target followed by structural characterization.

The protein target chosen for tethering with DNA mimic foldamer includes Sac7d. Sac7d is found in the thermophilic archaea *S. acidocaldarius,* which thrives in hot temperatures.[24] Sac7d is involved in the genomic stability of the archaea as it folds the DNA and works as a histone-like protein. It has been shown that Sac7d's fold is very stable and could bear the effect of multiple site-directed mutagenesis while maintaining its characteristic 5 beta-sheets and one alpha-helix fold (**Fig. 31**).[152] With this knowledge about Sac7d, we endeavored to perform site-directed mutagenesis to introduce cysteine residue onto the Sac7d surface such that it could make a disulfide bond with a DNA mimic foldamer.

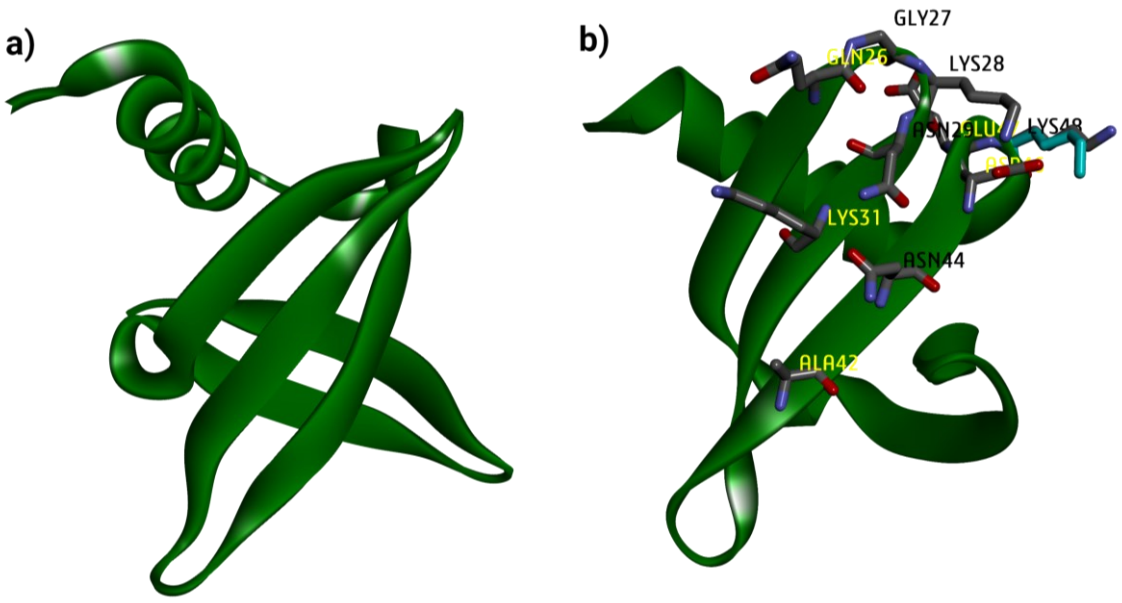


Fig. 31. Comparison of Sac7d with Nanofitin a) Sac7d is shown in ribbon representation (PDB ID 1AZQ, DNA not shown). b) Nanofitin (Sac7d, derived scaffold) is shown with mutated residues highlighted in stick representation compared to the wild-type Sac7d (PDB ID 2XIW).

5.1 Design and methodology

As stated in section 5, in the absence of a strong binding, tethering Sac7d with DNA mimic foldamer could lead to the design of a specific Sac7d foldamer complex from which, ultimately, the tether could be removed. In order to achieve these objectives, a DNA mimic foldamer with activated disulfide will be designed and synthesized along with a library of Sac7d surface cysteine mutants.

5.1.1 Designing Sac7d cysteine mutant library

S. acidocaldarius is the native source of Sac7d, a hyper-thermostable protein that protects the DNA of the archaea and is often termed a histone-like protein.[24] Because of this property, Sac7d kinks the DNA with a sharp 61° bend. However, the Sac7d V26A/M29A double mutant has been shown to induce a significantly less bend in the DNA at 29 °.[165] The fold of Sac7d has been shown to be resistant to multiple mutations, some even on the DNA binding site.[169] More than 10 mutations on Sac7d have been shown to maintain the same fold as wild-type Sac7d. Due to this property, Sac7d was chosen as a target protein for tethering with the DNA mimic foldamer.

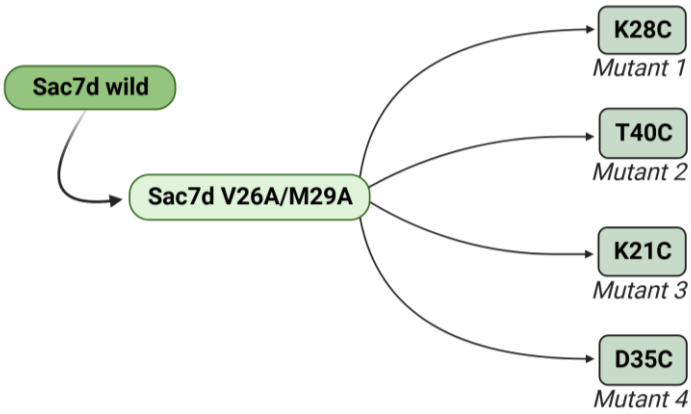


Fig. 32. Scheme following the design of Sac7d cysteine mutant library.

In order to generate cysteine mutants, we decided to use Sac7d V6A/M29A double mutant as it induces less kink to the binding partner (DNA). The $(^{m}QQ^{4})_{n}$-based DNA mimic foldamers (**Fig. 22**) were believed to be rigid, and it was hypothesized that Sac7d cysteine mutant tethered to a DNA mimic foldamer might not pack well in a crystal due to the rigidity of the foldamer not allowing Sac7d to intercalate V26 and M29 residues. Thus, the Sac7d cysteine library was based on Sac7d V26A/M29A (*Mutant 0*). In total, four different single cysteine mutants were designed. Following clones were designed (**Fig. 32**) and used for tethering with foldamer, Sac7d V26A/M29A/K28C (Mutant 1), Sac7d V26A/M29A/T40C (Mutant 2), Sac7d V26A/M29A/K21C (Mutant 3), Sac7d V26A/M29A/D35C (Mutant 4*)*. K21, K28, and D35 have previously been mutated and shown not to affect the fold of Sac7d.[169]

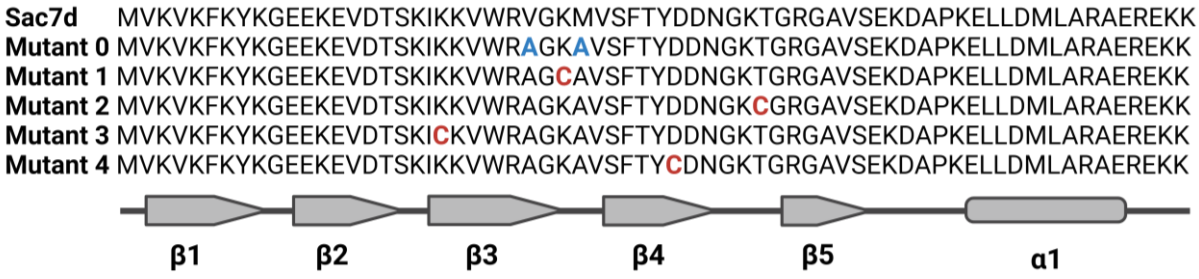


Fig. 33. Amino acid sequence of wild-type Sac7d and Sac7d V26A/M29A aligned with cysteine mutant library.

Mutant 1 was near the residues that kink the DNA, while mutant 2, mutant 3, and mutant 4 were on the other side of the beta-sheet (**Fig. 34**). All the mutants were synthesized and cloned in plasmid pET3b by GenScript Biotech, USA. The amino acid sequence of wild-type Sac7d and all the cysteine mutant libraries is shown in **Fig. 33**.

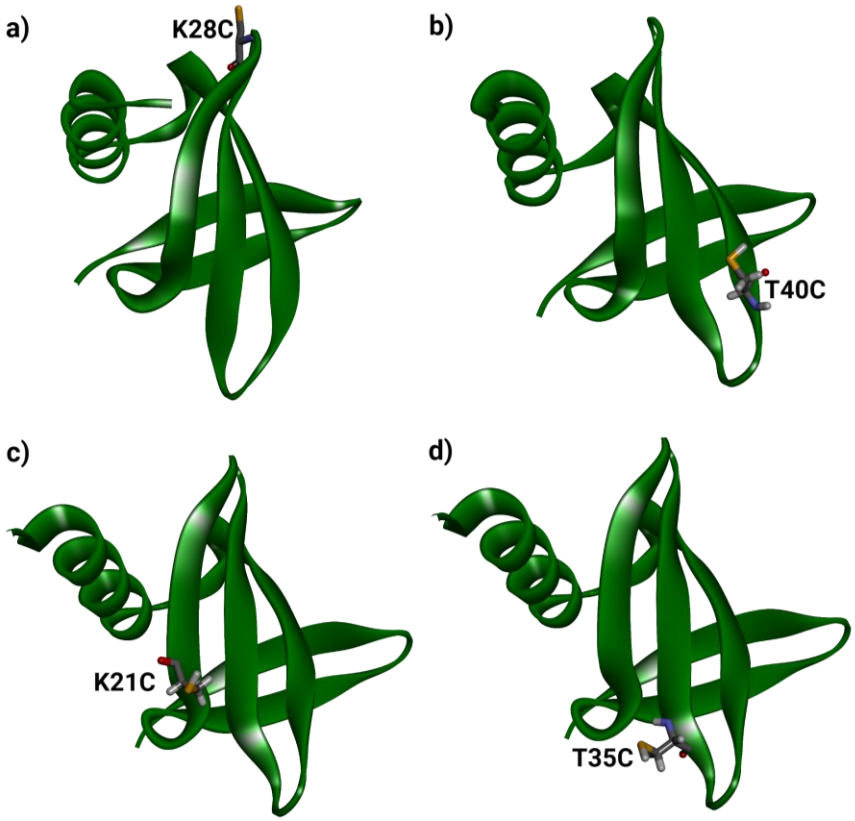


Fig. 34. Ribbon representation of Sac7d cysteine mutant (shown in stick representation). a) Mutant 1, b) Mutant 2, c) Mutant 3, d) Mutant 4. Note that these cysteine residues were mutated based on wild-type Sac7d (PDB ID 1AZQ) in Coot. Bound DNA to Sac7d is not shown for clarity purposes.

5.1.2 Design and synthesis of disulfide-activated DNA mimic foldamer

Foldamers used in this chapter were synthesized by Manuel Loos.

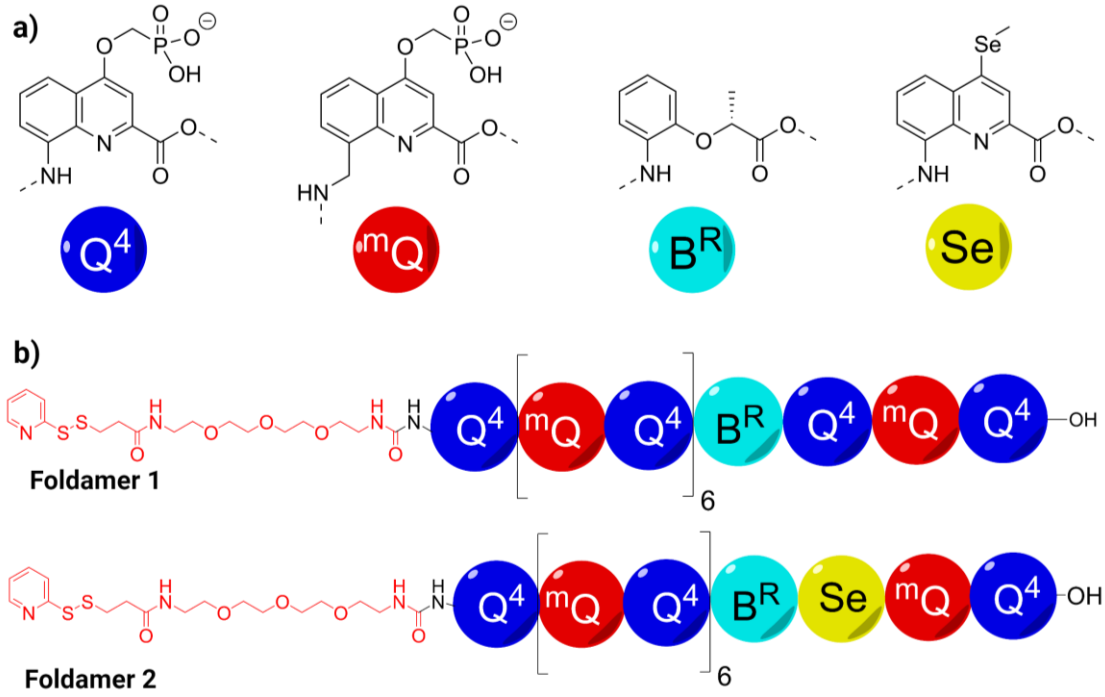


Fig. 35. a) Different building blocks used to synthesize the foldamers. b) N-terminal activated disulfide containing chiral DNA mimic foldamer.

The chiral DNA mimic foldamers **1** and **2** bearing an active disulfide moiety (**Fig. 35**) were synthesized by the microwave-assisted solid-phase synthesis (SPS). Compared to fragment condensation in solution, SPS allows for rapid and easy synthesis along with easier ways to change existing designs by altering only specific monomers. A brief description of the different steps involved in foldamer synthesis is discussed below and shown in **Fig. 36**.

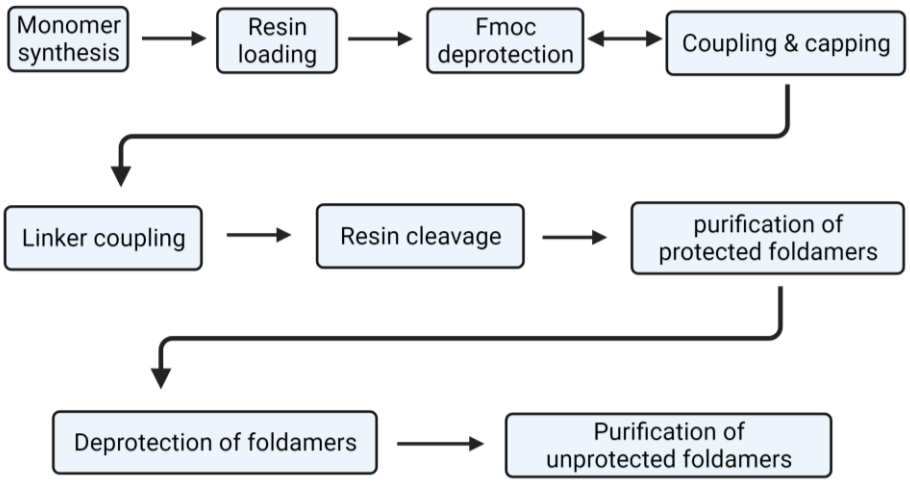


Fig. 36. Workflow and different steps involved in solid phase synthesis of foldamer.

Note that monomer synthesis is not shown. Foldamer synthesis was performed by SPS on low-loading Wang resin. Before loading the first monomer unit, the resin was brominated with PPh_3 and CBr_4 via an Appel reaction. Loading was measured by spectroscopic analysis of an aliquot of freshly Fmoc deprotected resin. Foldamer crude purity was measured by high performance liquid chromatography (HPLC) analysis. The activated disulfide linker was coupled to the free amine of the resin-bound foldamer after transforming it into an isocyanate by triphosgene. The resin-bound isocyanate was then reacted with a primary amine to form a urea bond. Fmoc deprotection was performed in N-methyl-2-pyrrolidone (NMP) to avoid residual DMF. After resin cleavage in TFA, the crude foldamers were purified by semi-preparative reverse phase HPLC to retrieve ethyl-protected phosphonate foldamers. The side chain phosphonate ethyl esters were removed by 20-fold excess of TMSBr per phosphonate ethyl ester. Later, the deprotected foldamers were purified on reverse phase HPLC (**Fig. 37**).

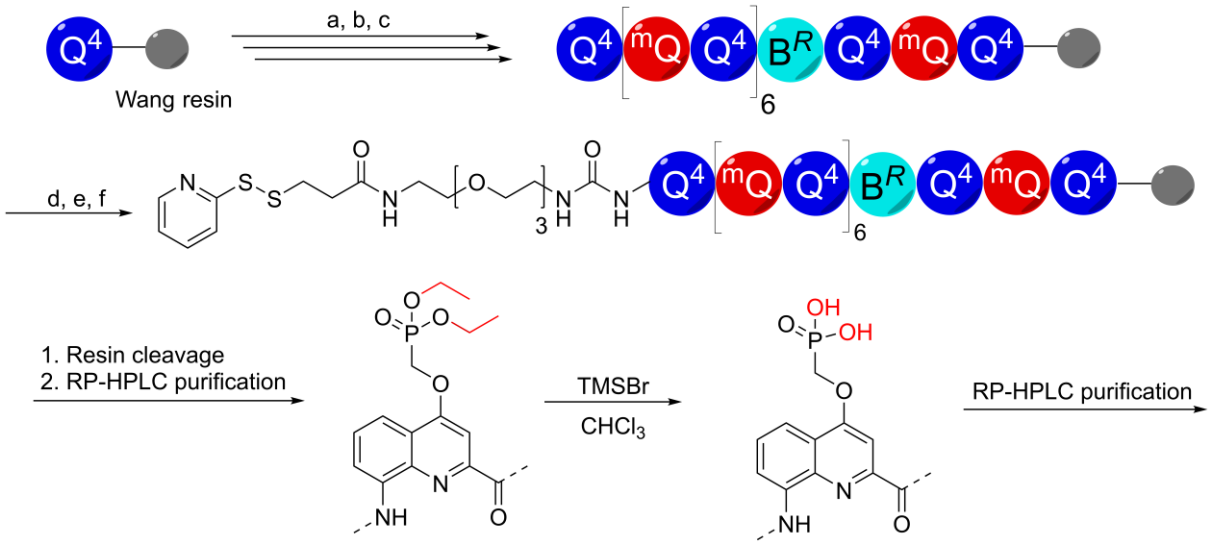


Fig. 37. Solid phase synthesis for a Q⁴-based sequence. Various steps and reagents involved are a) 20% piperidine in DMF; b) Monomer, Cl_3CCN, PPh_3, collidine, THF/NMP; c) 50% Ac_2O in DCM; d) piperidine in NMP; e) triphosgene, DIPEA, THF; f) Linker-NH2, DIPEA, THF.

5.1.3 Molecular modeling and MD simulation of tethered adduct

Molecular Modeling and MD simulation were performed using the Schrödinger 2018-1 suite. The structures were sketched in Maestro 11.5 and subsequently minimized with MacroModel. Energy minimizations were carried out using the OPLS3 as the force field and Steepest Descent (SD) followed by Polack-Ribiere Conjugate Gradient (PRCG) algorithms. The minimized models of the mutant 1-foldamer **2** adduct were subjected to molecular dynamic (MD) simulation at 300 K with a simulation time of 100 pico seconds in an OPLS3 force field in water as a solvent with SD as an energy minimization algorithm.

5.1.4 Protein over expression and purification

pET3b-Sac7d bacterial expression vector was kindly received from Prof. Chin-Yu Chen, department of life sciences, National Central University, Taoyuan, Taiwan.[165] Sac7d cysteine mutant genes were synthesized and cloned in bacterial expression vector pET3b by GenScript Biotech, USA (**Fig. 38**). The pET3b-Sac7d mutant 1 was transformed into *E. coli* BL21(DE3)pLysS cells. The cells were grown in Luria Broth (LB) to an OD at 600 nm of 0.8 and induced with 0.4 mM Isopropyl β- d-1-thiogalactopyranoside (IPTG) for 1 h induction at 37 °C.

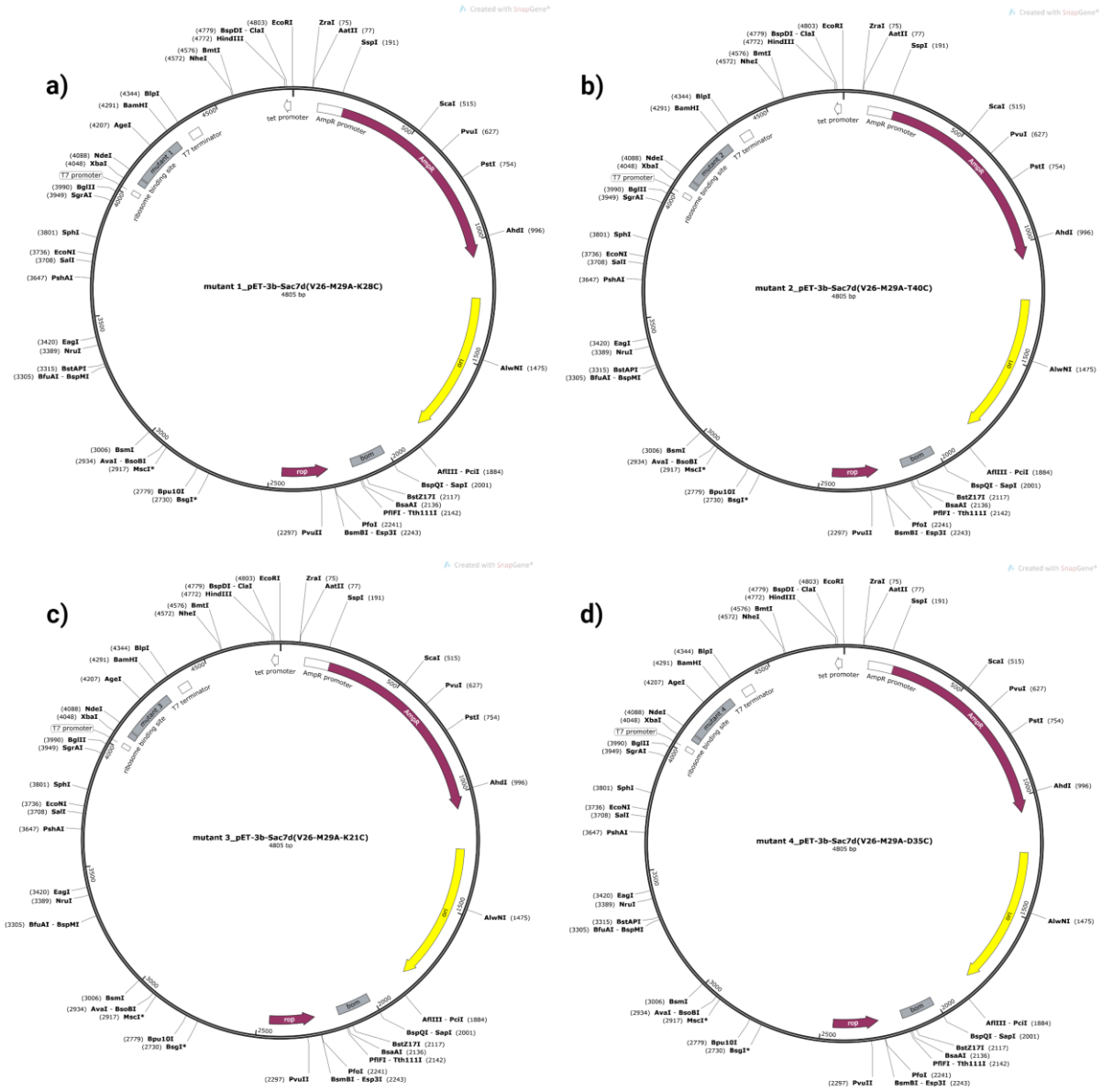


Fig. 38. Sac7d cysteine mutant plasmid maps. a) Mutant 1 (pET3b-Sac7d V26A/M29A/K28C) b) Mutant 2 (pET3b-Sac7d V26A/M29A/T40C) c) Mutant 3 (pET3b-Sac7d V26A/M29A/K2C) d) Mutant 4 (pET3b-Sac7d V26A/M29A/D35C). Please see the **section 10.1** for enlarged plasmid maps.

Afterward, cells were spun at 4000g for 15 min. The supernatant was resuspended and sonicated in 20 mM Tris pH 7.5 supplemented with protease inhibitor cocktail (Thermo Scientific™ Halt™ Protease Inhibitor Cocktail) and 2 mM EDTA. This was then heated in a water bath at 68 °C for 30 minutes and centrifuged at 105000 g for 1 hour. The supernatant was loaded on an SP cation exchange column (KNAUER Wissenschaftliche Geräte GmbH) and eluted with 20 mM Tris pH 7.5, 1 M NaCl. The purity was monitored by SDS-PAGE and LC-ESI-MS (**Fig. 39**). In the end, pure mutant 1 fractions were dialyzed against 20 mM Tris pH 7.5 and concentrated for making adducts with foldamer. mutant 2, mutant 3, and mutant 4 proteins were expressed and purified using the same protocol described above for mutant 1.

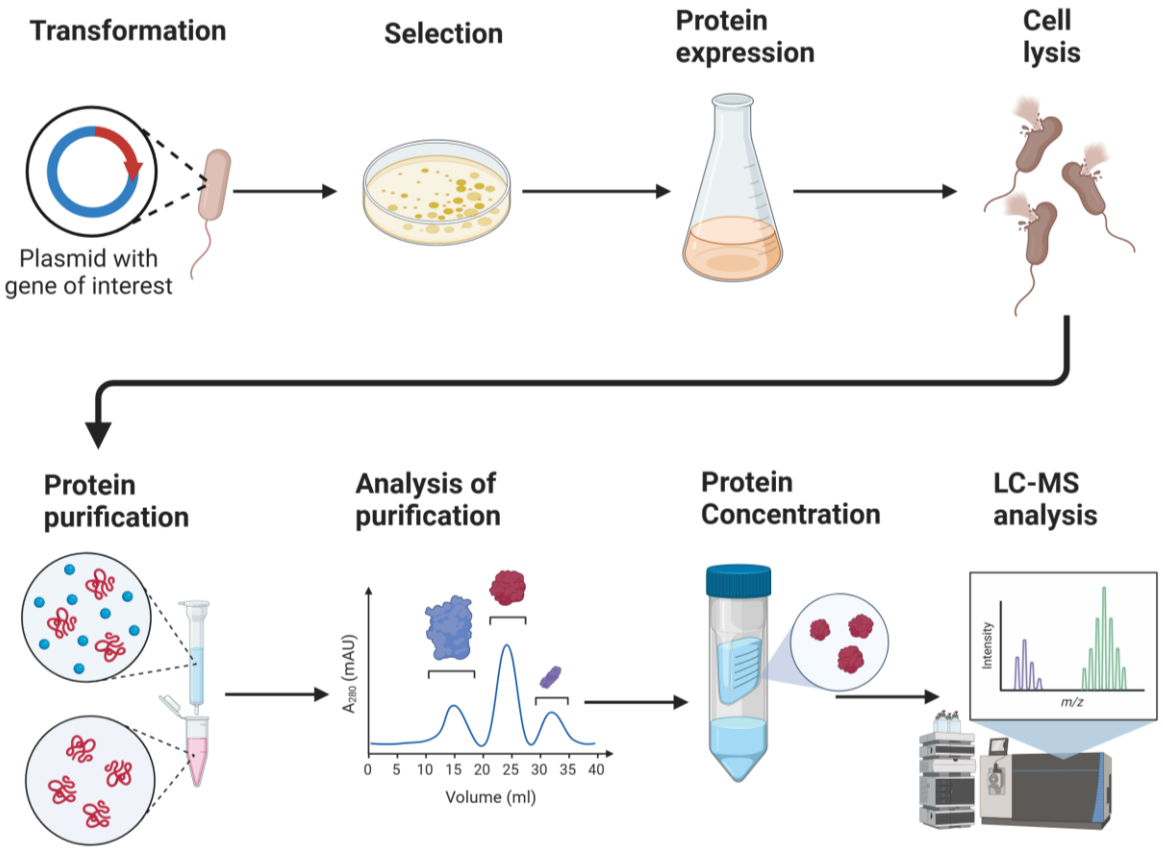


Fig. 39. Scheme of various steps involved in the expression and purification of mutant 1, mutant 2, mutant 3, and mutant 4.

5.1.5 Tether formation and purification

Once purified, Sac7d cysteine mutant 1 was reduced with 20 mM dithiothreitol (DTT) overnight at 4 °C. After 12 hours of reduction of the disulfides, reduced mutant 1 was applied to NAP-5 DNA purification columns (Cytiva Life science, USA). Reduced mutant 1 protein free from DTT was eluted with 20 mM Tris pH 7.5 and 50 mM NaCl and was used immediately to form a tether with foldamer. mutant 1: foldamer **2** (1:1) tether was produced in a concentration range of 100 µM to 250 µM up to a volume of 4 ml. This adduct was incubated at 4 °C for 2 hours and analyzed on LC-ESI-MS. Once the adduct formation was confirmed, it was purified on a HiLoad® 16/600 Superdex® 75 pg column (Cytiva Life science, USA) in 20 mM Tris pH 7.5, 150 mM NaCl. Once collected, fractions corresponding to mutant 1-foldamer **2** adduct were again analyzed on LC-ESI-MS to assess the purity before crystallization.

5.1.6 Tethered adduct crystallization and data collection

Pure fractions corresponding to mutant 1-foldamer **2** adduct were concentrated up to 1.3 mM, and initial crystallization broad screening experiments were performed at the Max Planck Institute of Biochemistry, Planegg, Germany. Initial broad screening conditions were optimized to yield crystals consistently in 18 % PEG 2000, 100 mM 3-(N-morpholino) propane sulfonic

acid (MOPS) pH 7.2 at 20 °C in a span of 2-3 days. Single crystals were cryo-protected for 10 seconds with 25 % glucose (w/v) in the crystallization solution before vitrification in liquid nitrogen for data collection. X-ray diffraction data were collected from cryo-protected crystals at 100K on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble) on a Dectris Eiger2 X 16M detector.[166]

5.2 Results and discussion

Designing a library of Sac7d cysteine mutants was envisioned to increase the chance of getting protein crystallography data with the tethered adduct. In this regard, foldamer **1** and foldamer **2** were synthesized. Foldamer **2** differed from foldamer **1** with the insertion of a selenium (Se) atom in the Q^4 monomer for the purpose of protein crystallography. As a heavy atom, selenium could help determine crystal structure via multiple-wavelength anomalous diffraction (MAD) phasing.[190] Selenomethionine derivatives have been previously shown to revolutionize the structure solution of proteins via MAD phasing.[191–194] However, in this case, instead of derivatizing all four Sac7d cysteine mutants with selenomethionine, we decided to incorporate a selenium atom in the foldamer **2**. This insertion of a selenium atom into the foldamer resulted in a net zero change in the conformation of the foldamer.

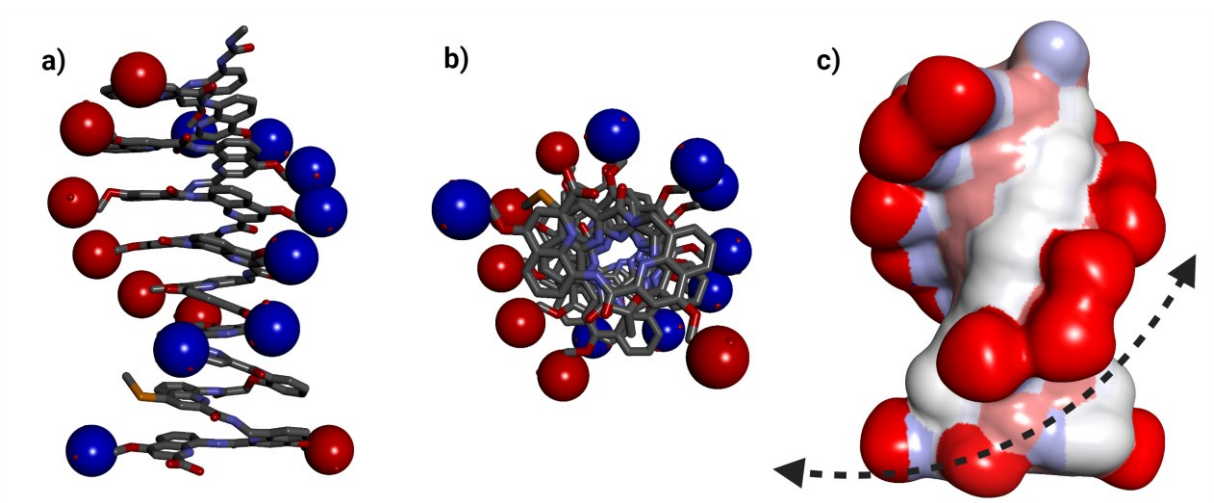


Fig. 40. a) Front view and (b) bottom view of the energy minimized model of the chiral Q^4Pho-based foldamer **2** with monoanionic phosphonate side chains and the chiral B^R unit in position 14. Alternate phosphorus atoms are colored dark blue and red. (c) The solvent-accessible surface of the energy-minimized model. Please note the double headed arrow showing the right-handed helix of the negative charge mimicking the right handed B-DNA.

Foldamer **2** (with selenium) was modeled using a racemic $(^mQQ^4)_8$ crystallized with TREX protein. 12 units of the TREX-$(^mQQ^4)_8$ foldamer were used, and other units were sketched on top of the helix, completing the foldamer **2**, which was then subjected to energy minimization in MacroModel. Energy minimizations were carried out using the OPLS3 as the force field and Steepest Descent (SD) followed by Polack-Ribiere Conjugate Gradient (PRCG) algorithms. As

shown in **Fig. 40**, the energy-minimized conformation followed a single left-handed helical conformation with the exo-helix (double helical) mimicking the topology of a right-handed idealized double-stranded B-DNA.

Sac7d V26A/M29A/K28C (mutant 1) model with DNA replaced by the foldamer **2** was energy minimized with restraints on the foldamer **2**. The MD simulation performed for 100 picoseconds yielded overall stable adduct trajectories (**Fig. 41**) during the span of the simulation. The main chain of the foldamer **2** maintained close contact with mutant 1, further approving the formation of a stable tethered adduct.

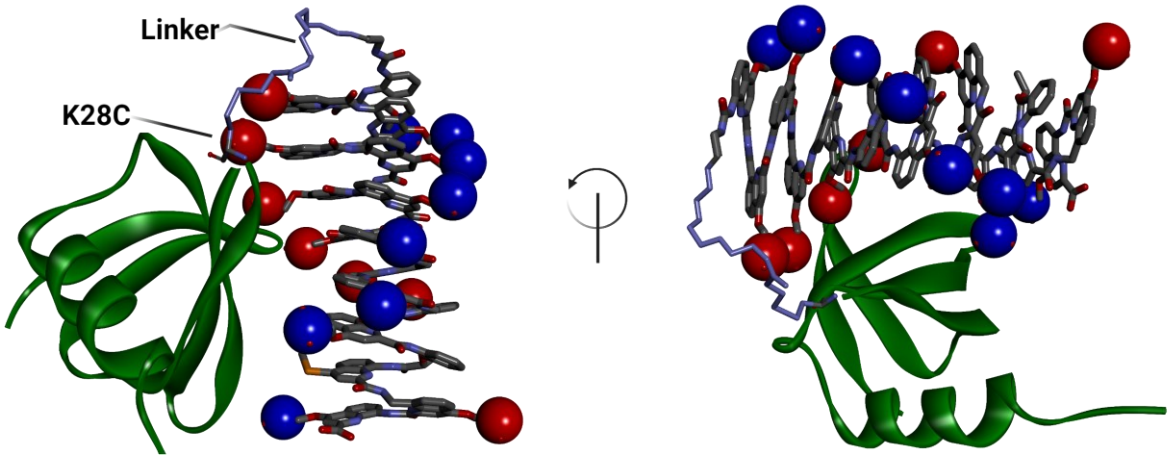


Fig. 41. Energy minimization followed by MD simulation showing one of the trajectories of tethered foldamer **2** (shown in stick representation with phosphorus atoms in blue and red balls) with mutant 1 (shown in green ribbons).

After the simulations, Sac7d cysteine mutants cloned in plasmid pET3b were expressed and purified, and all the fractions corresponding to pure proteins were analyzed on a 15 % sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) gel. It is worth mentioning that wild-type Sac7d and its cysteine mutants were found to be toxic to the expression cell line, and thus, an overnight induction with IPTG did not yield enough protein for the purpose of tethering. Thus, expression conditions were optimized with an induction period of only 1 hour at 200 RPM shaking at 37 °C. These conditions gave a good yield of each cysteine mutant that was enough to try tethering with foldamers 1 and 2. Please note that mutant 2 chromatogram is shown in time (minutes) on the x-axis with a flow rate of 0.5 ml per minute, and thus the 2nd peak in the chromatogram of mutant 2 corresponds to an elution volume of 75 ml to 90 ml (**Fig. 42**).

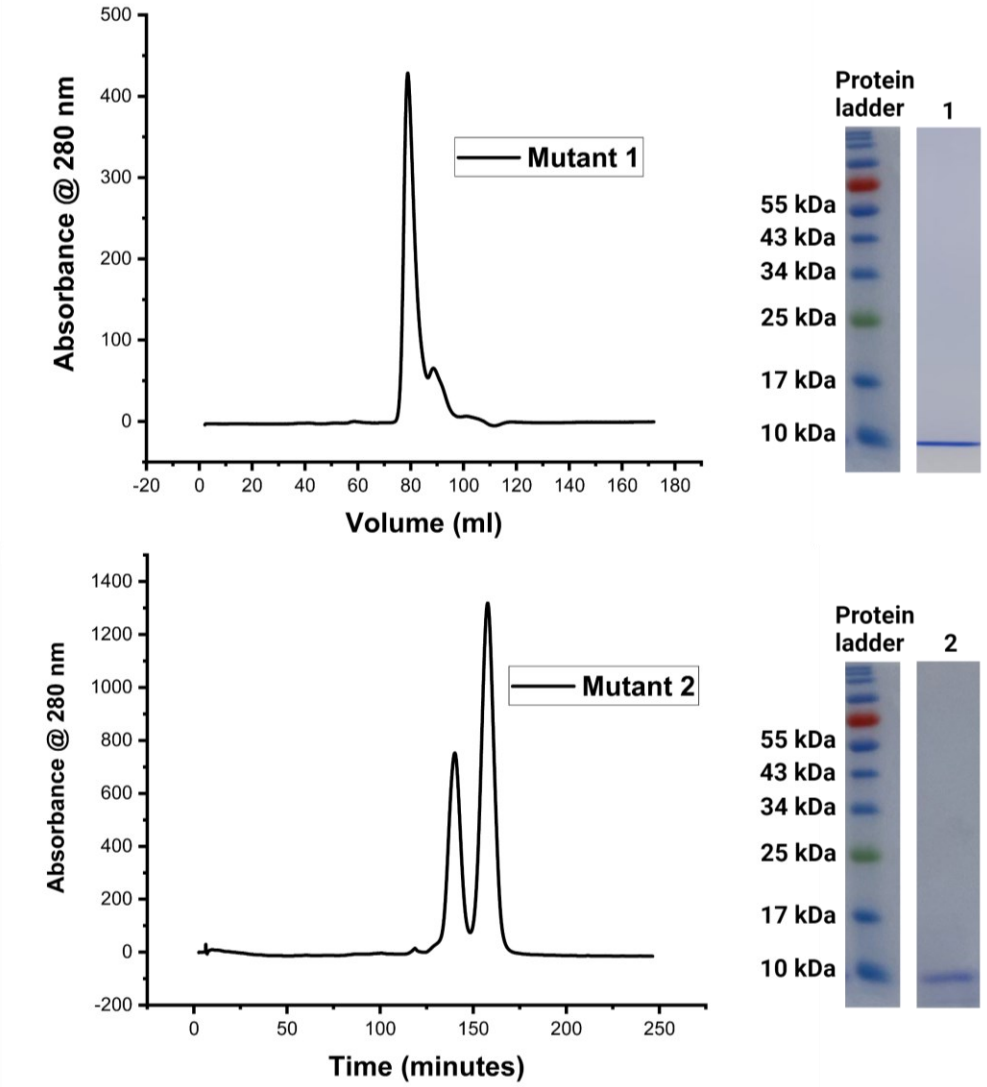


Fig. 42. Size exclusion chromatograms and 15% SDS-PAGE profiles of mutant 1 (top) and mutant 2 (bottom) purified on Superdex 75 pg column.

Mutant 3 and mutant 4 were also expressed under similar conditions as mutant 1 and mutant 2. During the cell lysis, the lysis buffer was supplemented with a protease inhibitor cocktail (Thermo Scientific™ Halt™ Protease Inhibitor Cocktail) and 2 mM EDTA, as it was previously seen that due to contamination, some unspecific protease had cleaved mutant 3 during the course of purification. Cell lysis using sonication was performed on ice for 15 minutes (5×3 minutes), then cells were heated in a water bath at 68 °C. This step ensured that most of the proteins of the expression host were denatured at such high temperatures, except the Sac7d cysteine mutant and some heat shock proteins from the expression system.

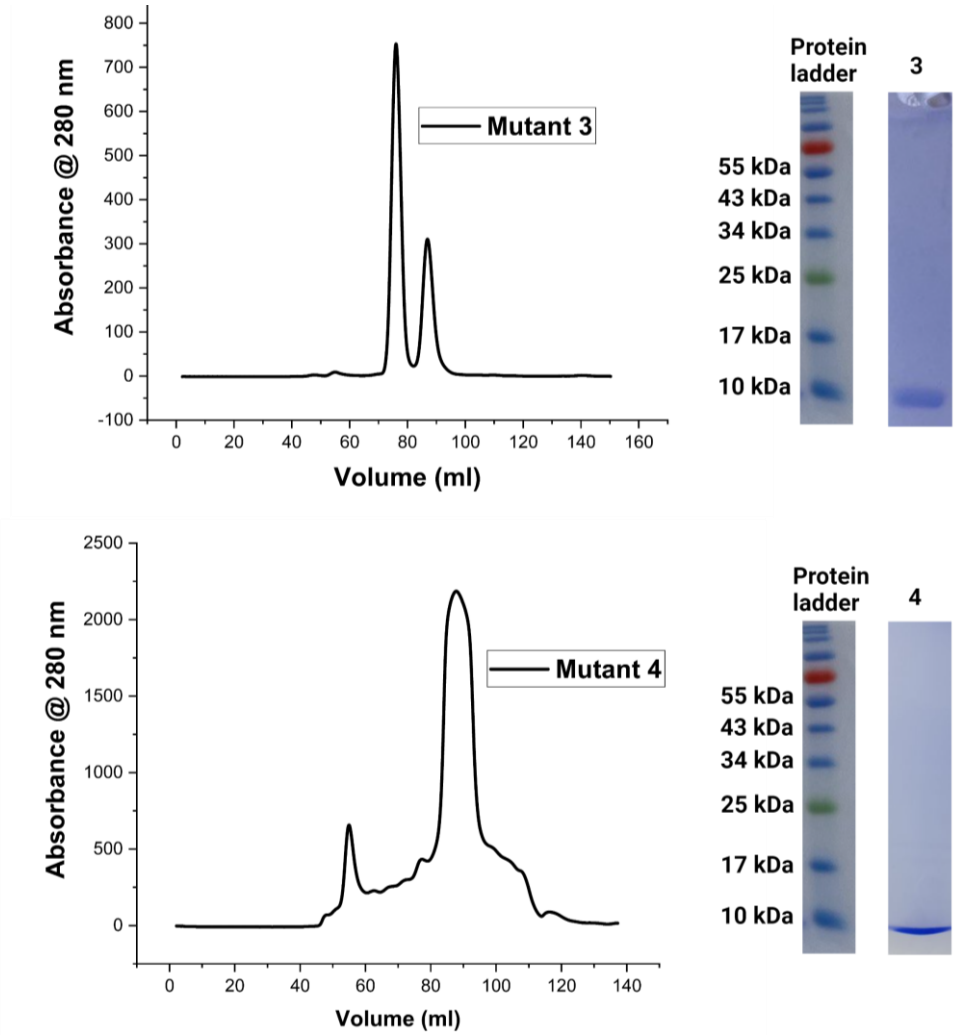


Fig. 43. Size exclusion chromatograms and 15% SDS-PAGE profiles of mutant 3 (top) and mutant 4 (bottom) purified on Superdex 75 pg column.

Later, cells were centrifuged at 105000 g for 1 hour to remove most of the denatured contaminant proteins from the pellet. Supernatant after centrifugation was filtered through a 0.2 µm PVDF filter paper, and the clear protein solutions were loaded on an SP cation exchange column, yielding more than 90 % purity. However, to achieve maximum purity (more than 95 %), each mutant protein was loaded on Superdex 75 pg column (**Fig. 43**), and fractions corresponding to pure protein were concentrated, analyzed on LC-ESI-MS (**Fig. 44** and **Fig. 45**), and stored at 4 °C until further use to tether with foldamer. Tethering for adduct formation was initially performed between mutant 3 and foldamer **2** (1:1) in a concentration of 150 µM in 4 ml volume. After 2 hours of incubation at room temperature, the crude adduct was purified on size exclusion chromatography using Superdex 75 pg column (**Fig. 46**).

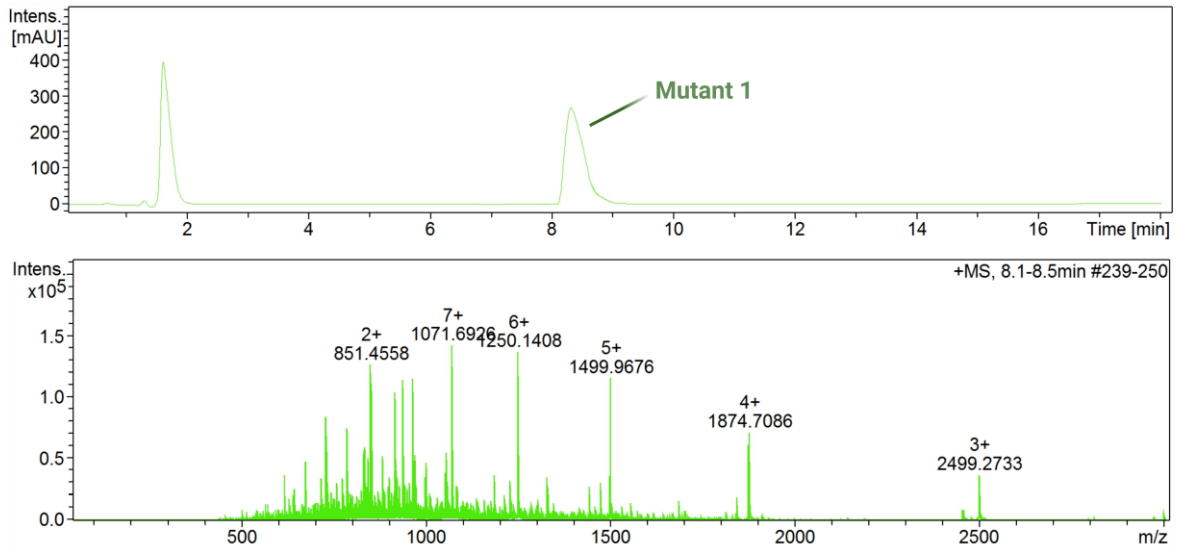


Fig. 44. LC chromatogram at UV 280 nm and ESI-MS spectrum of Sac7d V26A/M29/K28C (Mutant 1).

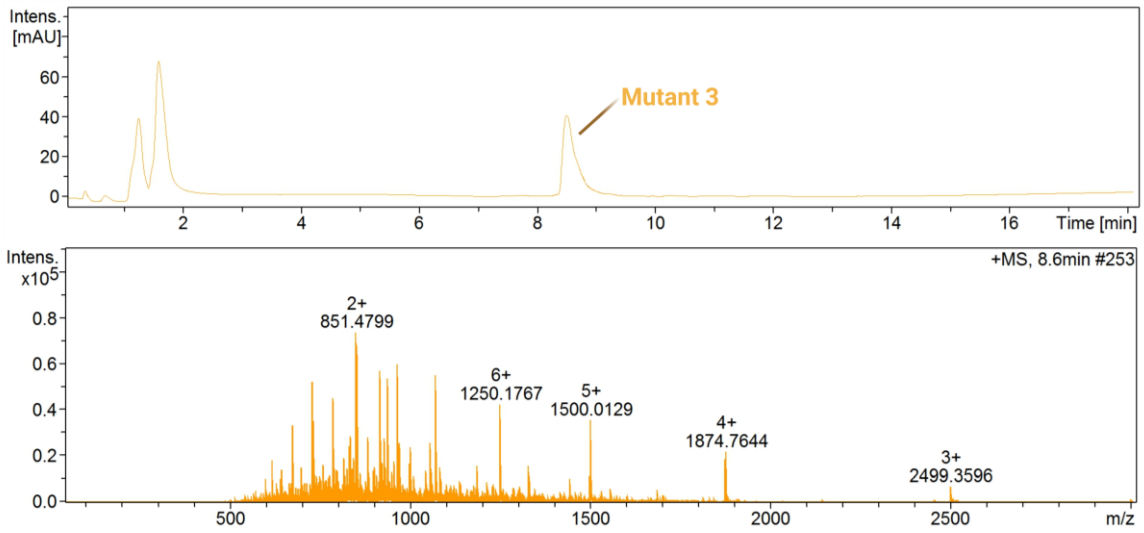


Fig. 45. LC chromatogram at UV 280 nm and ESI-MS spectrum of Sac7d V26A/M29/K21C (Mutant 3).

Fractions corresponding to the tethered adduct were analyzed by LC-ESI-MS (**Fig. 47**) with a C4 column, in a 5 % to 95 % gradient of water to acetonitrile (0.1 % formic acid), or with a C18 Kinetex column, in a 0 % to 70 % gradient of NH_4OAc (12.5 mM, pH 8.5) to acetonitrile. Once the mutant 3-foldamer **2** adducts were confirmed for purity, they were subjected to broad screening for crystallization conditions. However, after having screened more than 2000 conditions, no hits were observed. Moreover, making the adduct of mutant 1-foldamer **2** was initiated at this stage. This was the precise reason behind having a library of cysteine mutant Sac7d instead of working on a single mutant so that another mutant could be easily tested in case of roadblocks at any step.

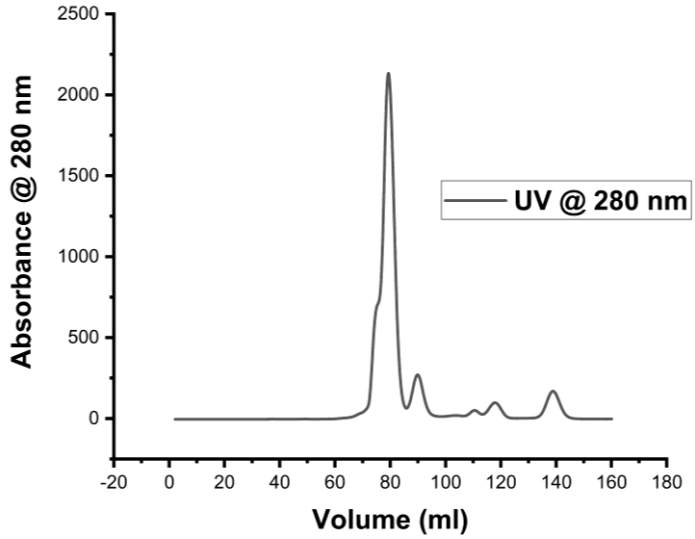


Fig. 46. Size exclusion chromatogram of mutant 3-foldamer **2** tethered adduct. Fractions corresponding to adduct eluted between 70 ml and 85 ml.

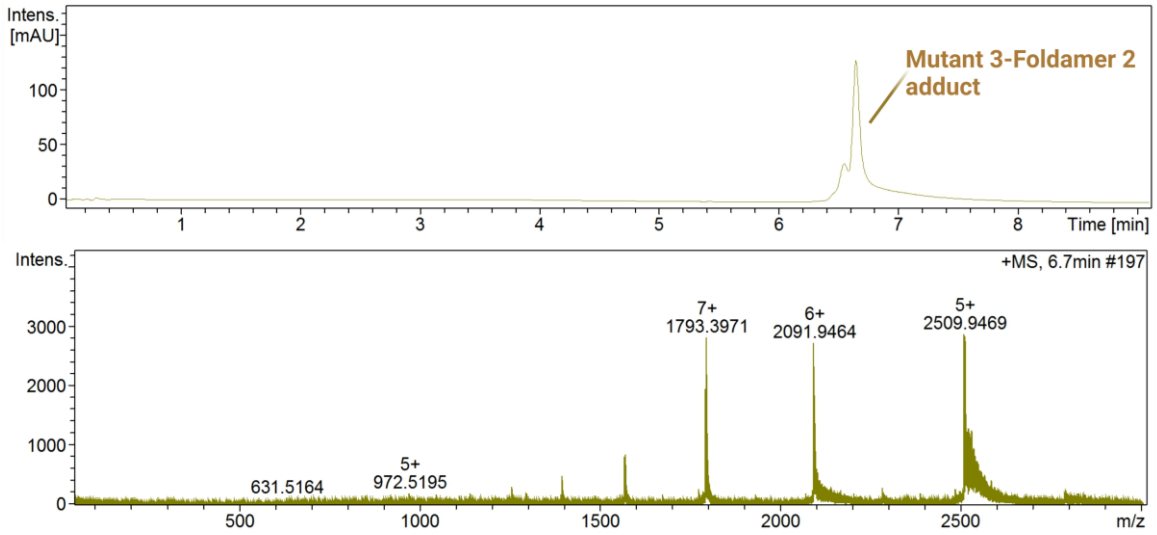


Fig. 47. LC chromatogram at UV 300 nm and ESI-MS spectrum of mutant 3-foldamer **2** adducts.

mutant 1-foldamer **2** adduct was formed using similar conditions as mutant3-foldamer **2**. Adduct purification was also performed on the Superdex 75 pg column (**Fig. 48**), and pure fractions corresponding to the adduct were analyzed on LC-ESI-MS (**Fig. 49**) for purity before setting up crystallization drops. It is worth noting that the purification of adduct was initially tried under HPLC conditions on a C18 column; however, yields were low. The amount of the recovered pure adduct was of insignificant to be tested for crystallization. Compared to HPLC, FPLC purification yielded larger amounts of more than 95 % pure adduct.

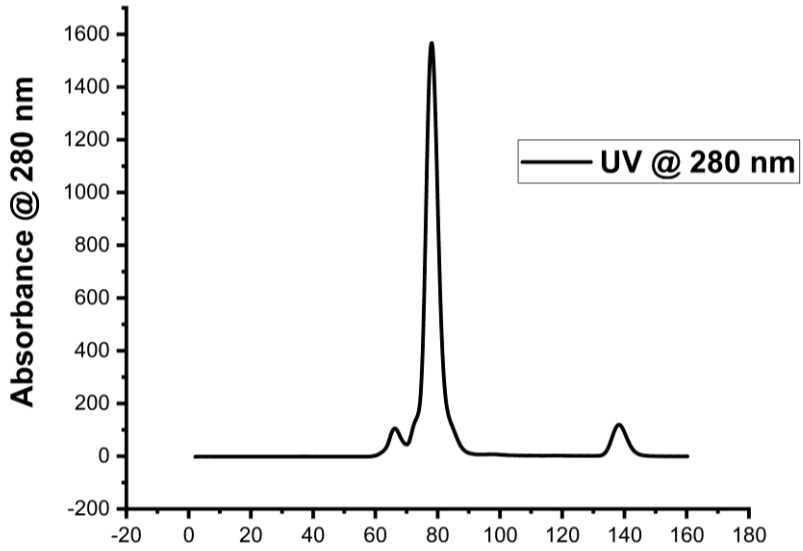


Fig. 48. Size exclusion chromatogram of mutant 1-foldamer **2** tethered adduct. Fractions corresponding to adduct eluted between 70 ml to 85 ml.

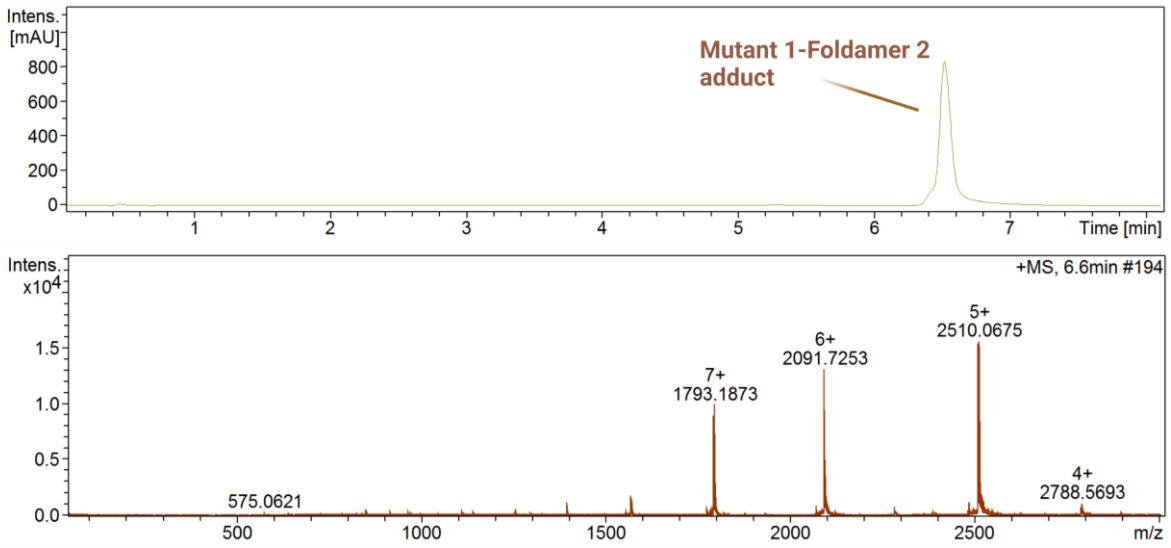


Fig. 49. LC chromatogram at UV 300 nm and ESI-MS spectrum of mutant 1-foldamer **2** adduct.

Pure fractions of mutant 1-foldamer **2** adduct were crystallized in 18 % PEG 2000, 100 mM MOPS pH 7.2 at 20 °C (**Fig. 50**). Crystals diffracted up to 2.99 Å and were auto-processed (unit cell parameters shown in **Table 7**) using "grenandes_parallelproc." However, the R merge in the higher resolution shell was excessively high and molecular replacement performed using various molecular models such of Sac7d, foldamer **2**, and Sac7d-foldamer **2** adduct all lead to an incorrect molecular replacement models.(**Table 6**).[167] New crystals with improved statistics in terms of resolution, I/sigma, and CC1/2 value could potentially help overcome the problem of a wrong molecular replacement model as discussed in **section 5.3**.

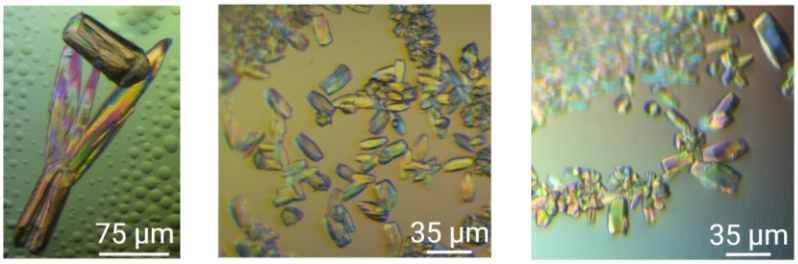


Fig. 50. Crystal of mutant 1-foldamer **2** adducts observed under cross-polarizing microscope.

Table 6. Initial statistics of auto-processed crystal data of mutant 1-foldamer **2** adduct.

P 3 Space group	Completeness (%)	Resolution (Å)	Rmerge
Overall	94.3	47 – 3.0	17.3
Inner	97.4	47 – 9.0	13.5
Outer	75.3	3.17 – 2.99	3666

Table 7. Unit cell parameters of mutant 1-foldamer **2** adduct from the initial auto-processed crystal data.

a	b	c
39.96 Å	39.96 Å	331.83 Å
α	**β**	**γ**
90 °	90 °	120 °

5.3 Conclusion and perspective

In conclusion, our study has successfully substantiated methodologies for covalently tethering helically folded aromatic oligo amides, which mimics the topology of B-DNA to cysteine residues introduced through site-directed mutagenesis on the surface of the Sac7d protein. Through analytical methods, we have corroborated the efficacy of ligation by forming disulfide bridges. Tethering DNA mimic foldamers onto Sac7d cysteine mutant proteins was broken down into four distinct tasks (**Fig. 51**). In Task 1, we designed, expressed, and purified four distinct single cysteine mutant versions of Sac7d V26A/M29A. In Task 2, the synthesis and purification of the DNA mimic foldamers, overseen by Manuel Loos, were successfully accomplished. These designed foldamers, featuring an activated disulfide moiety, were strategically synthesized to enable the formation of covalent adducts between the foldamers and Sac7d, facilitating the exploration of their structural interactions. Task 3 involved the meticulous tethering of foldamers to the protein surface, followed by a rigorous process of screening and purification of the resulting foldamer-protein adducts. In Task 4, the

crystallogenesis and subsequent crystallography of the foldamer-protein adducts were pursued (**Fig. 52**)

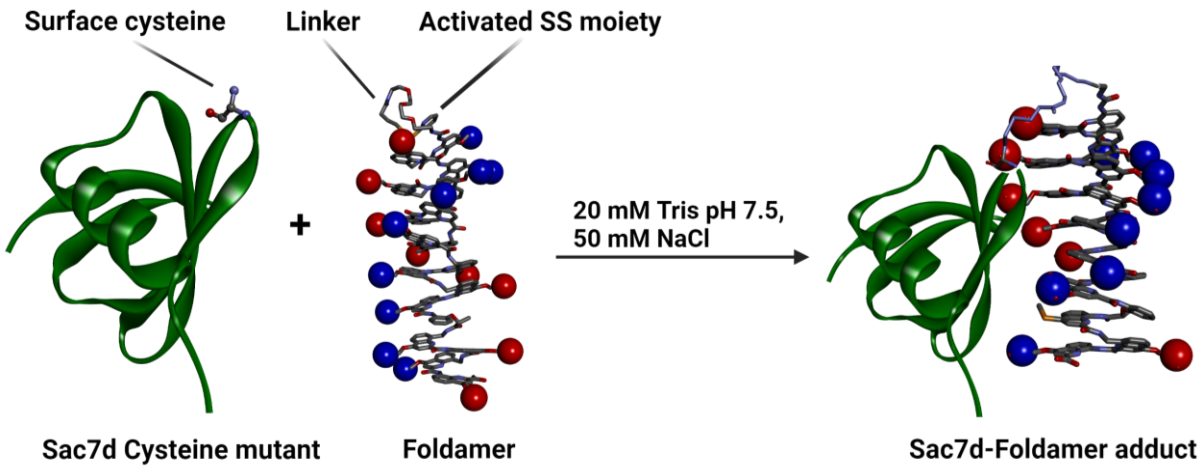


Fig. 51. An example of a ligation experiment between mutant 1 and foldamer.

The efforts invested yielded successful outcomes, with the design, expression, and purification of the single cysteine mutant versions of Sac7d being achieved as planned.

Furthermore, the synthesis and activation of DNA mimic foldamers and their successful covalent attachment to Sac7d were confirmed through characterization on mass spectrometry (ESI-MS). A crystallization campaign was initiated after successfully establishing and purifying the Sac7d-foldamer **2** covalent adducts. This resulted in crystals that exhibited X-ray diffraction up to 3.5 Å resolution and displayed consistent unit cell characteristics. However, these crystals presented challenges related to low-resolution and incomplete X-ray datasets, which impeded their suitability for molecular replacement-based structure determination. Despite these limitations, the consistent diffraction patterns observed in these crystals hold promise, and efforts are underway to collect improved resolution data to elucidate the Sac7d-foldamer **2** adduct's structure.

A pivotal aspect of this project lies in the iterative feedback loop to be established between high-resolution structural insights and foldamer design optimization (**Fig. 52**). Iterative rounds of structure-based analysis will aid in identifying critical structural features essential for the successful binding of foldamers to Sac7d. Ultimately, this process is anticipated to yield foldamers capable of binding to Sac7d even without needing a covalent tether in an orientation different from what was observed in chapter 4. The position of the cysteine mutants and the length of the tether linker used in this study do not allow the binding mode (orientation of foldamer on Sac7d surface) observed in the previous chapter. Mutant 0 (Sac7d V26A/M29A),

the basis of all the cysteine mutants, could compensate for the absence of the central aliphatic linker used in the previous chapter and allowing for a new binding mode on Sac7d.

This innovative tethering approach could be extrapolated to other DNA-binding proteins, enabling the tailored synthesis of DNA mimics targeting specific proteins, potentially outcompeting natural DNA substrates. The versatility of our DNA mimics in competing against protein-DNA interactions positions them as invaluable tools, both within the realms of biology and pharmacology. Moreover, their potential for therapeutic intervention should be noticed, presenting an exciting avenue for future exploration.

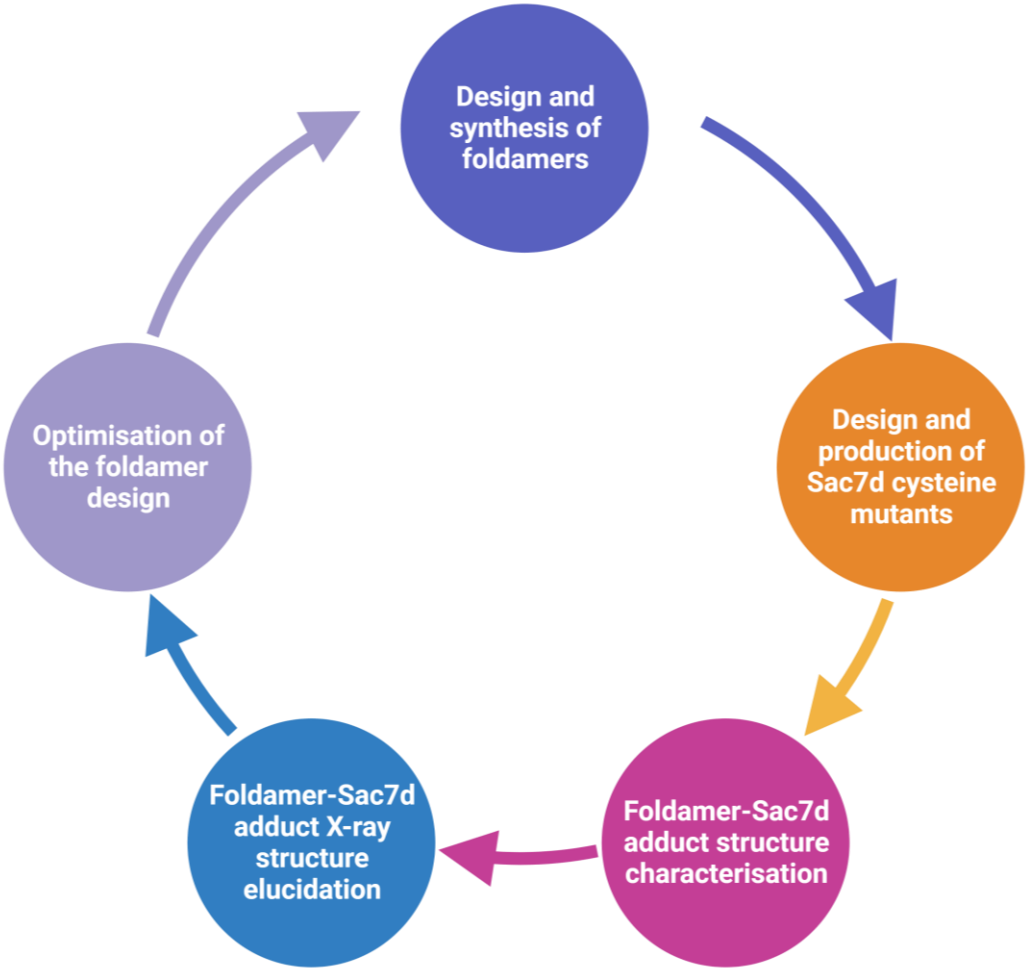


Fig. 52. Structure-based iterative design of foldamers as ligands for interfering with protein-DNA interactions.

In summary, the success of this study in validating the covalent tethering methods, along with the detailed insights to be gained into the structural aspects of foldamer-protein interactions, opens up new horizons for advancing our understanding of biomolecular interactions and designing foldamers with a wide range of applications.

6. Targeting hcGAS by DNA mimic foldamers

6.1 Introduction

Defense mechanisms developed against the recognition and targeting of foreign molecules and pathogens are crucial for the survival of the host organisms such as humans. This defense system is termed the immune system. For example, nucleases present in bacteria serve as a defense mechanism against the DNA of invading organisms such as viruses. The recognition of invading pathogens by molecular patterns is carried out by the innate immune system of various organisms, including humans and other vertebrates, which is represented by NK cells, dendritic cells, and macrophages.[195] On the other hand, the adaptive immune system works as a second line of defense against the invading pathogens found only in vertebrates. T and B lymphocytes are the cell types representing the adaptive immune system. A combination of innate and adaptive immune responses makes a robust defense mechanism to survive against invading pathogens.

6.1.1 DNA sensing by cGAS

The main targets of the innate immune system are molecular patterns associated with pathogen or self-damage, for example, cyclic dinucleotides, flagellin, or nucleic acids arising from the pathogen invasion or self-damage. The innate immune system recognizes these molecular patterns arising from pathogen or self-damage using pattern recognition receptors. The cGAS-STING pathway involves a family of pattern recognition receptors in mammalian cells that are involved in sensing cytosolic DNA.[196] In this pathway, the binding of cGAS to double-stranded DNA (dsDNA) activates its catalytic activity, producing 2′3′ cyclic GMP-AMP (cGAMP). This molecule acts as a potent activator of STING and functions as a second messenger.[197–199] The cGAS–STING pathway is unique in that its activation is triggered by DNA rather than possessing any pathogen-specific attributes.[200] As a result, cGAS can identify and respond to a wide range of DNA types, whether foreign or of self-origin. The human cGAS (hcGAS) protein has a C-terminal catalytic domain with the nucleotidyltransferase domain. The catalytic domain possesses DNA-binding areas that are positively charged, with a primary site and two additional sites. These areas bind to the DNA's sugar-phosphate backbone. Upon DNA binding to the primary site, it alters the protein's structure, modifying the enzyme's catalytic pocket and enabling optimal interaction with ATP and GTP substrates. This conformation change to hcGAS is crucial to binding to the adjacent B-site to create the active core 2:2 cGAS DNA complex. It has also been shown that in the presence of longer stretches of dsDNA, cGAS dimers form a ladder-like structure to form phase-separated organelles (**Fig. 54b**).[201] Because of this spatial arrangement, cGAS signaling is only activated in the presence of longer dsDNA,

such as by an invading pathogen. Once the cGAMP is produced, it is detected by STING, an endoplasmic reticulum membrane protein[202], which then, via a signaling cascade, leads to the production of type 1 interferons, which plays a critical role in halting the viral propagation. It is essential to mention that DNA in eukaryotic cells is confined within the boundaries of the nucleus and mitochondria and degraded rapidly once in the cytosol by various nucleases. So, following an infection by a pathogen, the increase in cytosolic DNA is sensed and acted upon by the cGAS-STING pathway (**Fig. 53**).[203]

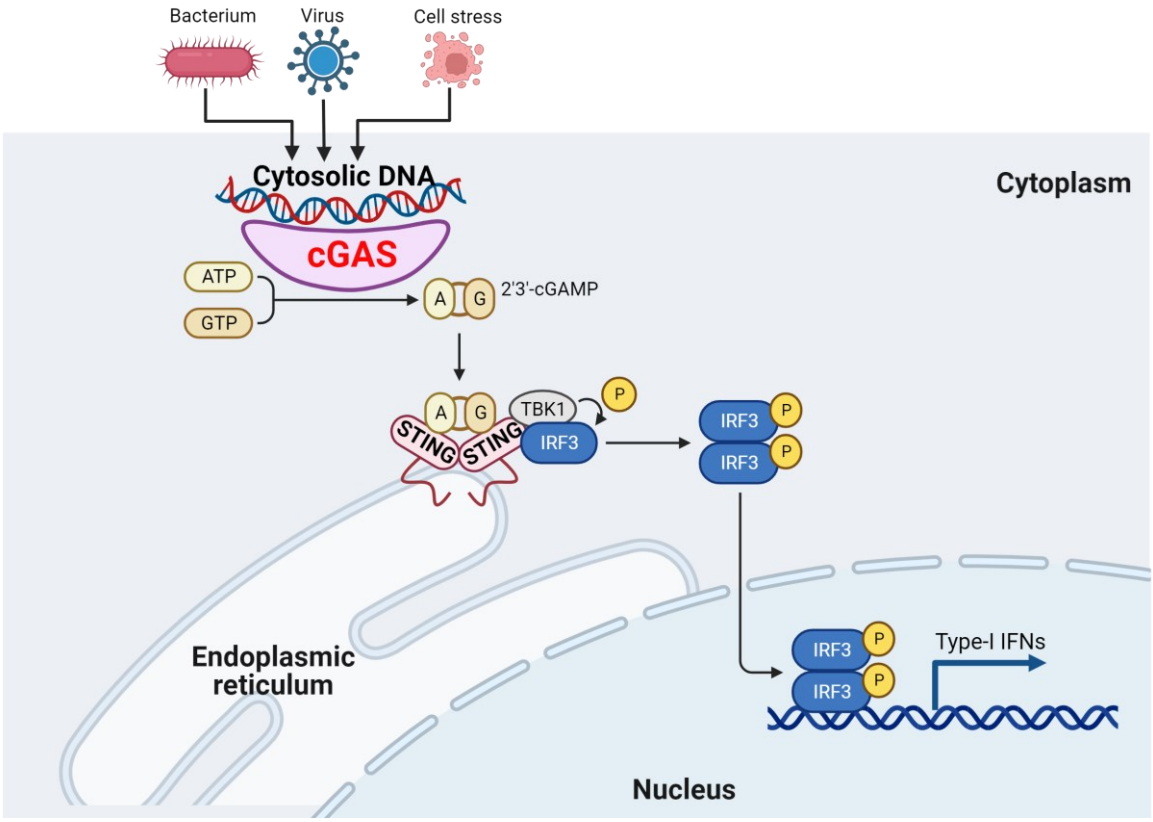


Fig. 53. Overview of cGAS-STING signaling pathway showing the dsDNA-induced activation of cGAS, which leads to the production of cGAMP that binds to STING to express the type 1 interferon and other chemokines.

In recent years, it has been observed that not only foreign DNA from viruses and bacteria can activate the cGAS-STING pathway but also mitochondrial and nuclear DNA that enters the cytosol. This activation can cause chronic inflammation and pathology due to higher levels of cytosolic DNA resulting from factors like mitotic stress in cancer and autoimmune disorders.[204,205]

6.1.2 cGAS structure

cGAS is a ~520 amino acid-containing protein with a basic N-terminal domain (160 amino acid long) and a ~360 amino acid long C-terminal domain. The catalytic domain of cGAS comprises two structural lobes that contain the active site in between. The first lobe includes a core β-sheet

of NTases (Nucleotidyl Transferases, responsible for synthesis of cyclic GMP-AMP). This lobe also contains acidic residues that are conserved and are involved in the Mg^{2+}-dependent transfer of nucleoside phosphate onto the hydroxyl acceptor. The second lobe completes the active site and provides additional interactions for binding the nucleoside-triphosphate donor. It has been by various independent studies shown that the apo form of hcGAS does not form an active site for the catalysis of the reaction; it is only upon the binding of DNA that a conformational change is induced, which leads to the active site formation that can efficiently bind to the ATP and GTP.[206,207] As previously described, cGAS binds to DNA as a dimer, forming a ladder-like structure by covering 16-18 bp of DNA with no specificity towards DNA sequence (**Fig. 54b** and **Fig. 55**).

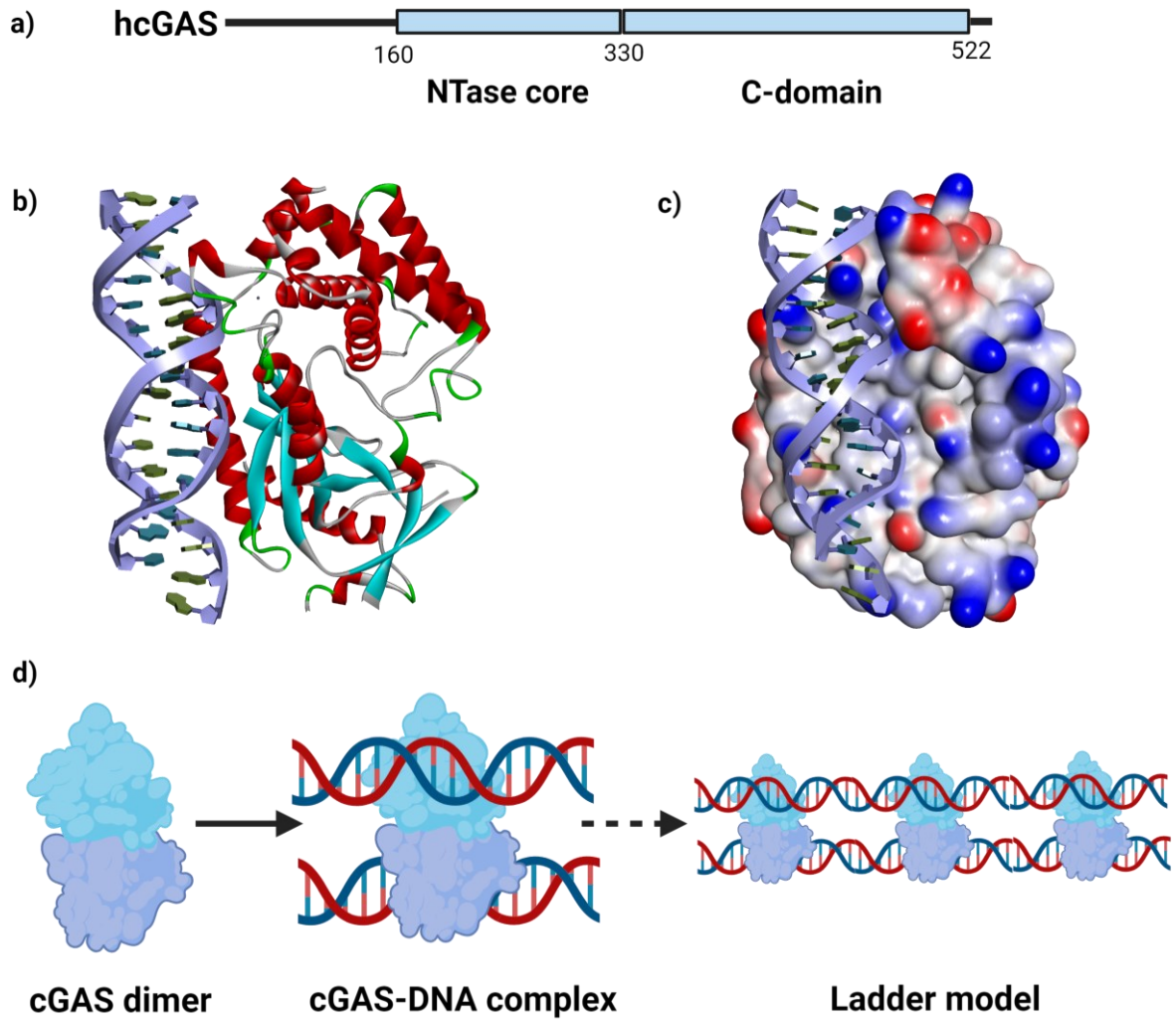


Fig. 54. Crystal structure of hcGAS-DNA. a) Schematic representation of hcGAS domains organization. b) Crystal structure of hcGAS-DNA. c) Crystal structure shown in surface representation highlighting the fit of DNA to the binding site on hcGAS (PDB ID 6EDC).

The C-domain of hcGAS is a helix bundle with a zinc finger motif, which mediates DNA binding and dimer formation. The leucine residue at the zinc finger motif controls the

production of cGAMP upon binding with DNA. DNA binding to hcGAS is driven by a positively charged cleft present on the hcGAS surface.

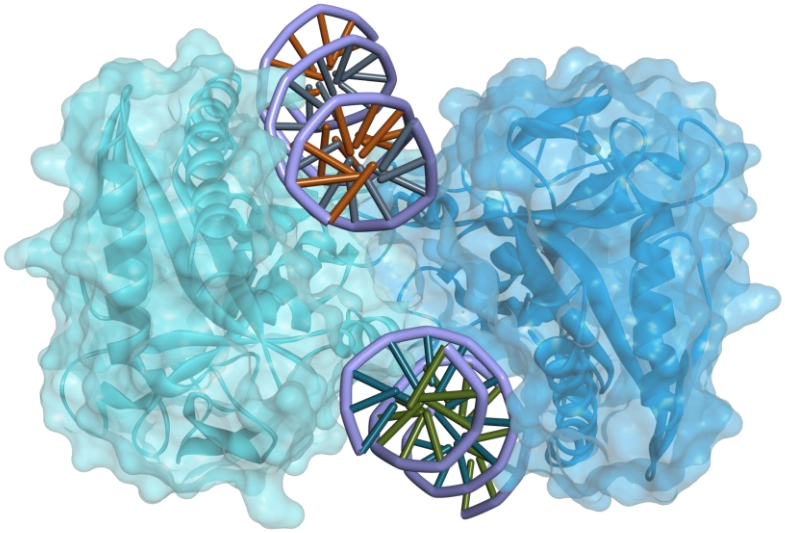


Fig. 55. Crystal structure of mcGAS dimer in complex with DNA (PDB ID 4O6A).

6.2 Design and methodology

hcGAS was chosen as the target protein by DNA mimic foldamers because hcGAS binds to DNA in a sequence-independent manner without any preference for specific base pairs. Thus, it was hypothesized to be a target protein for DNA mimic foldamers. Additionally, hcGAS has undergone extensive structural studies with DNA or other ligands, providing a foundation for further exploring its interaction with DNA mimic foldamers as potential inhibitors.

6.2.1 Foldamer design and synthesis

Foldamers used in this chapter were synthesized by Dr. Valentina Corvaglia and Jiaojiao Wu. DNA mimic foldamers possess unique structural attributes that enable them to interact with some DNA-binding proteins. These protein targets are non-sequence selective and recognize DNA based on its shape and charges. In some instances, oligomers of 8-amino-2-quinoline carboxylic acid (Q^4) and 8-aminomethyl-2-quinoline carboxylic acid (^mQ) have been observed to bind to proteins, even in the presence of excess DNA, thereby outcompeting it. This has resulted in the potent inhibition of significant enzymes such as human topoisomerase 1 (Top1) and human immunodeficiency virus 1 integrase (HIV-1 IN).[147] Recent research has demonstrated that DNA mimic foldamers can influence chromatin composition and alter cell cycle progression.[208] However, $(^mQQ^4)_n$ foldamers are not ideal for structural studies with DBPs such as hcGAS due to their N- and C-termini and lack of a stereogenic center. These factors can lead to mixtures of diastereomeric complexes with proteins and alternate protein

binding modes, depending on the N->C orientation of the foldamer. To address these limitations, modifications were made to bias their handedness, promote their association through helix cross-sections, and make them C_2-symmetrical (**Fig. 56** and **Fig. 57**). These enhancements have already proven effective in solving the crystal structure of a complex between a DNA-mimic foldamer and Sac7d (**Chapter 4**).[209] Please note that in the manuscript (**Chapter 4**) Q^4 is called Q and ^mQ is called M.

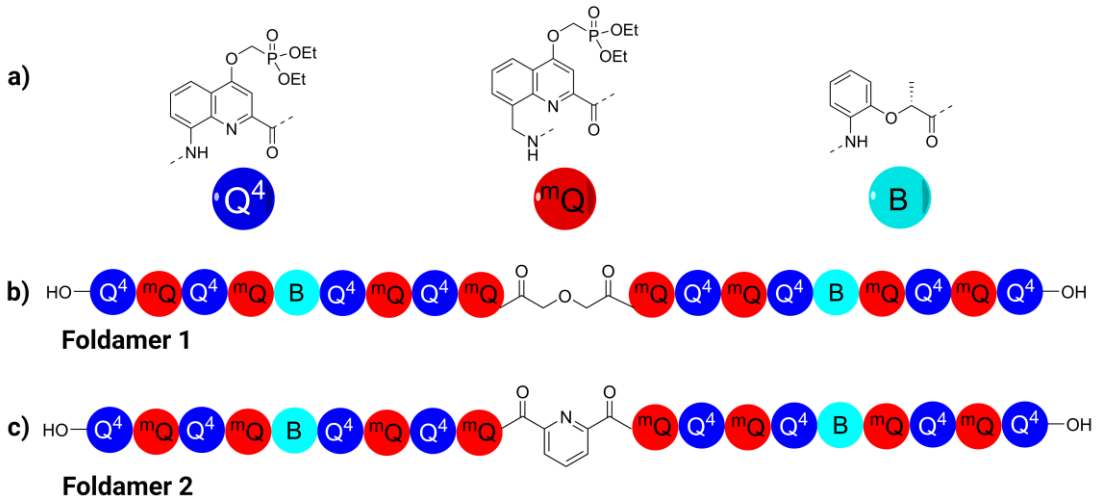


Fig. 56. C_2-Symmetrical DNA mimic foldamers. a) chemical formula of monomers used in DNA mimic foldamers synthesis. b, c) C_2-symmetrical DNA mimic foldamers with a chiral B monomer and central C_2-symmetrical moiety.

6.2.2 Synthetic strategy for C_2-symmetrical foldamers

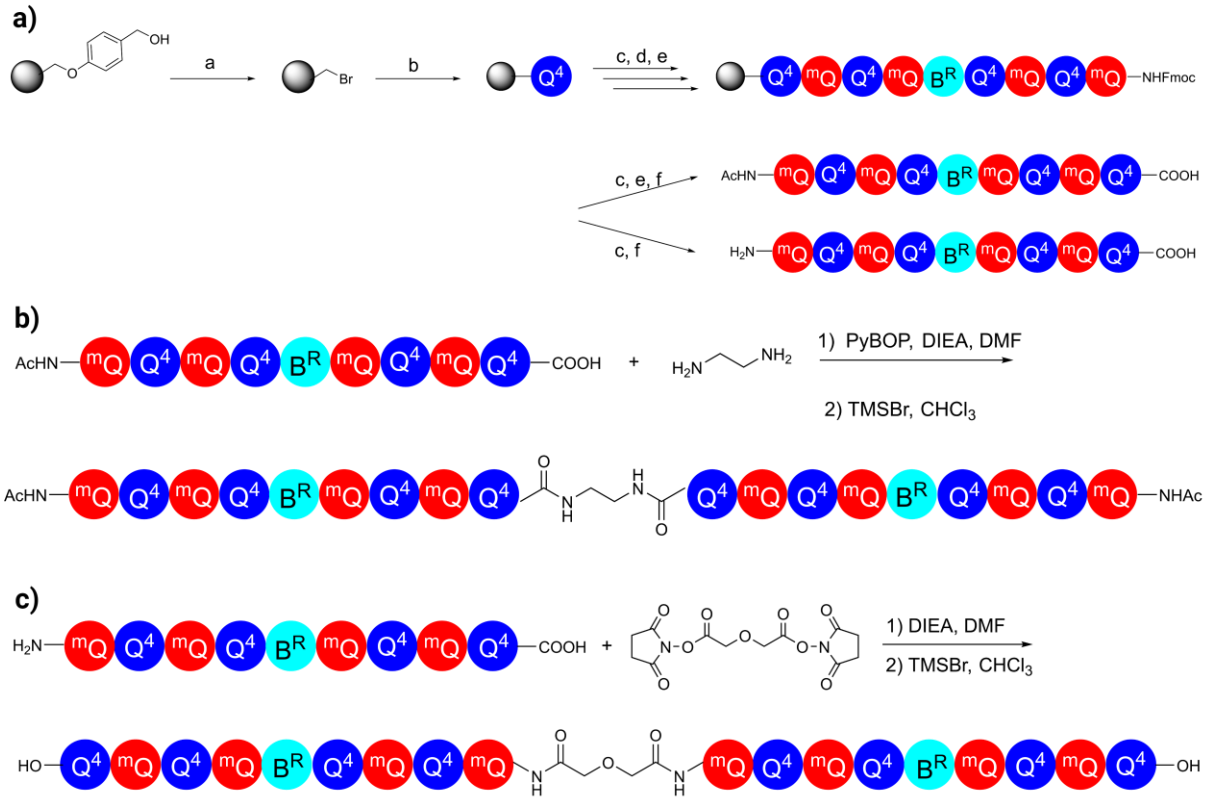


Fig. 57. a) Solid phase synthesis of a 9-mer chiral DNA mimic: a) CBr_4, PPh_3, DMF; b) Fmoc-Q-OH, CsI, DIEA, DMF; c) 20% piperidine in DMF; d) Monomer, TCAN, PPh_3, collidine, THF/$CHCl_3$; e) 50% Ac_2O in DCM; f) TFA. **b**) Chiral C_2-symmetrical DNA mimic foldamer synthesized using a diamine linker. **c**) Chiral C_2-symmetrical DNA mimic foldamer synthesized using an activated diacid linker.[210]

6.2.3 hcGAS construct design and over expression

The expression system for hcGAS was purchased from addgene.org with catalog number 127162 (courtesy of Prof. Thomas Tuschl's laboratory, The Rockefeller University, New York, USA). hcGAS (K427E/K428E) (gene name MB21D1) was cloned in plasmid pRSFDuet (Novagen) between BamHI and XhoI restriction sites with insert size of 5028 base pairs with an N-terminal His_6-SUMO tag with a ubiquitin-like protease (ULP1) cleavage sequence separating the hcGAS (K427E/K428E) from the His_6-SUMO tag (**Fig. 58**). Recombinant plasmid was received from addgene.org in DH5alpha cells which were first inoculated on a luria agar petri plate supplemented with Kanamycin antibiotic (50 µg/ml). The plasmid was isolated using the PureLink™ quick plasmid miniprep Kit. The pRSFDuet-sumo-hcGAS (K427E/K428E) was transformed into *E.coli* BL21-CodonPlus (DE3)-RIL competent cells. Bacteria were grown in luria broth to an OD at 600 nm of 0.8 and were then induced with 0.3 mM IPTG for overnight induction at 18 °C.[211]

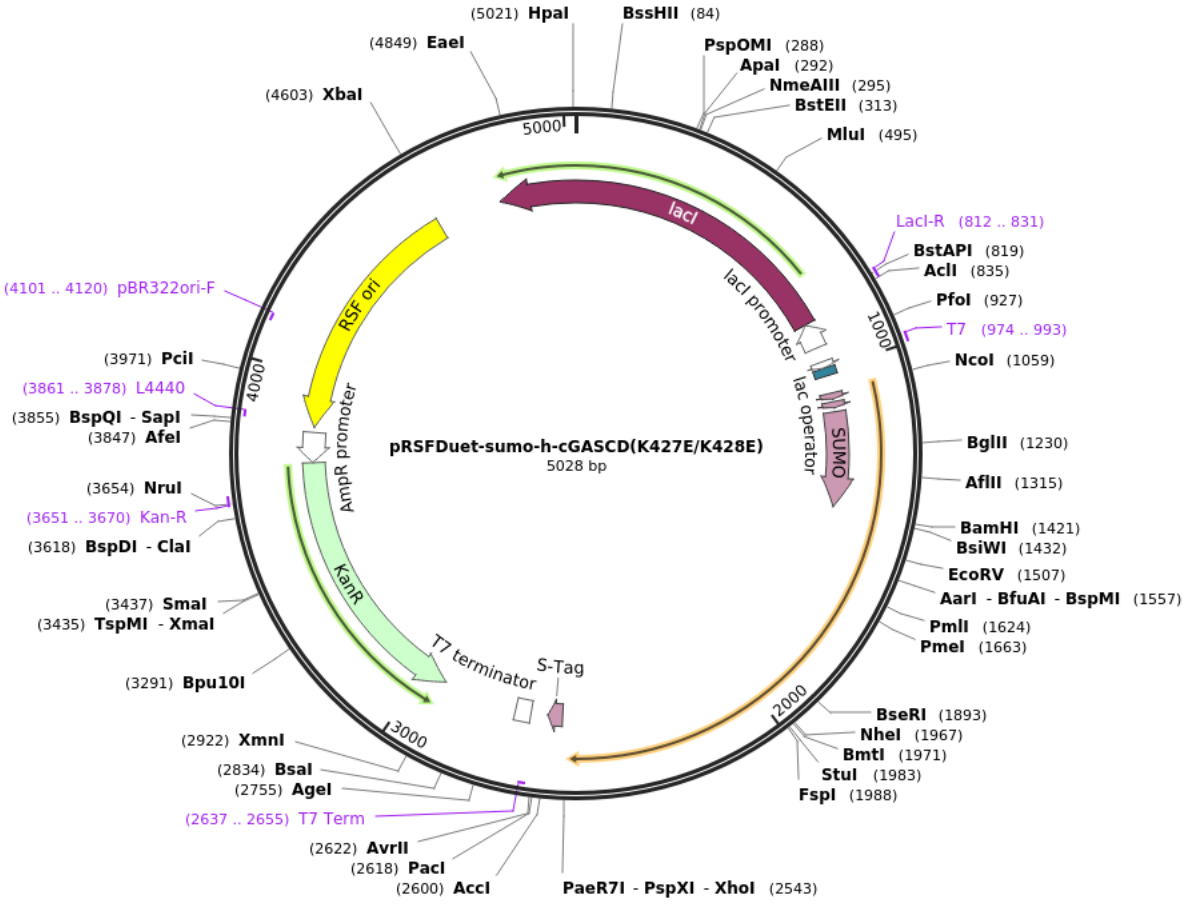


Fig. 58. Plasmid map of pRSFDuet-sumo-hcGAS (K427E/K428E).

6.2.4 hcGAS purification

6.2.4.1 Affinity chromatography

Proteins tagged with histidine tag have a strong affinity to nitrilotriacetic acid (NTA) resin that has been immobilized with Ni^{2+} ions. The histidine residues in the His_6 tag bind specifically to the vacant coordinate ion site of the immobilized coordinate ion sphere of Ni^{2+}. However, some untagged proteins from bacterial cell lysate may also bind to Ni-NTA if their histidine residues remain in close proximity to the surface, although the binding affinity of such untagged proteins is lower than that of tagged proteins, which have six consecutive histidine residues. These non-specific proteins can be removed without affecting the target protein, and the target protein can be eluted by increasing the concentration of imidazole or lowering the pH of the buffer.

Overnight induced culture of bacteria was harvested by centrifugation at 4 °C. Cells were resuspended in buffer A (500 mM NaCl, 20 mM imidazole, and 50 mM Tris-HCl, pH 7.5) supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF) protease inhibitor and 5 mM β-mercaptoethanol and lysed using sonication. Lysed cells were subjected to centrifugation at

20000 rpm (31304 *g*) in Beckman Coulter JA-25.50 rotor at 4 °C for 1 h and subjected to affinity chromatography using Ni-NTA resin.

The supernatant from centrifugation was added to HisPur™ Ni-NTA resin (Thermo Fisher Scientific, USA) and washed thoroughly with buffer A. The target protein was then eluted with buffer A supplemented with 500 mM imidazole. The ULP1 enzyme was used to remove the His_6-SUMO tag during dialysis at 4 °C overnight in a buffer containing 300 mM NaCl, 1 mM DTT, 20 mM Tris-HCl, pH 7.5, and 5% glycerol. The tags were removed by chromatography using HisPur™ Ni-NTA resin and concentrated for injection on size exclusion chromatography. In order to crystallize the hcGAS-DNA complex, the His_6-SUMO tag was not cleaved.

6.2.4.2 Size exclusion chromatography

Size exclusion chromatography was used to further purify the hcGAS protein after the tag cleavage and 2nd round of affinity chromatography. The sample corresponding to cleaved hcGAS was loaded onto the Knauer AZURA FPLC system coupled with a 16/600 Superdex 200 column (GE Healthcare, USA) in buffer C (300 mM NaCl, and 1 mM DTT, 20 mM Tris-HCl, pH 7.5). Before loading the protein, the column was pre-equilibrated with buffer C. Fractions corresponding to pure hcGAS were analyzed on SDS-PAGE and concentrated for crystallization (apo and in complex with foldamers).

6.2.5 Crystallization

Crystallization of His_6-SUMO hcGAS (K427E/K428) was attempted with a DNA sequence previously crystallized with hcGAS (**Fig. 59**).[212] For the purpose of crystallization with DNA, the affinity tag was not cleaved from hcGAS. His_6-SUMO hcGAS (K427E/K428) was mixed with dsDNA in 0.22 mM: 0.27 mM (protein: DNA). Crystallization screening was performed in sitting drop vapor diffusion method at 20 °C in 100 nl + 100 nl (protein-DNA complex + crystallization solution) at the Max Planck Institute of Biochemistry, Planegg, Germany.

Sense strand	**5'**	A A A T T G C C G A A G A C G A A **3'**
Antisense strand	**5'**	T T T C G T C T T C G G C A A T T **3'**

Fig. 59. DNA sequence used in crystallization with His_6-SUMO hcGAS (K427E/K428).

Initial crystal hits were obtained in 3 days under a crystallization solution containing 150 mM ammonium sulfate, 0.1 M 2-ethanesulfonic acid (MES), pH 6.0.

Crystallization with foldamer **1** was performed with hcGAS (K427E/K428), where the His_6-SUMO tag was cleaved off prior to crystallization setup. hcGAS (K427E/K428) was mixed with foldamer **1** in 0.3 mM:0.35 mM, respectively, and initial crystallization screening was

performed at the Max Planck Institute of Biochemistry, Planegg, Germany. After initial crystallization screening, reproducible rod-shaped crystals were obtained at 20 °C in hanging drop vapor diffusion method in a 15-well crystallization plate under 35 % 2-Methyl-2,4-pentanediol (MPD), 0.M Tris pH 7.0, 0.2M NaCl and 35 % 2-Methyl-2,4-pentanediol (MPD), 0.M N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) pH 7.0, 0.2M NaCl in a 1 µl + 1 µl mixture (protein-foldamer complex + crystallization solution).

6.2.6 Data collection

Crystals from the His_6-SUMO hcGAS (K427E/K428)-DNA complex were looped out from the crystallization drop using a 20 µm nylon loop and flash cooled in liquid nitrogen at 100 K. These crystals were cryoprotected using 25 % PEG 400 prepared in crystallization solution. X-ray diffraction data was collected at beamline P13 at the Deutsches Elektronen Synchrotron (DESY) at the European Molecular Biology Laboratory (EMBL, Hamburg).[213]

Crystals belonging to the hcGAS (K427E/K428)-foldamer **1** complex were looped out from the crystallization drop using a 100 µm nylon loop and flash cooled in liquid nitrogen at 100 K. These crystals were diffracted at beamline P13 the Deutsches Elektronen Synchrotron (DESY) at European Molecular Biology Laboratory (EMBL, Hamburg)[213] and beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble).[166]

6.2.7 Structure determination

X-ray diffraction data from crystals of the His_6-SUMO hcGAS (K427E/K428)-DNA complex was processed by the autoPROC pipeline.[214] At first, the asymmetric unit analysis was determined by calculating the Matthews coefficient and solvent content using Xtriage. PDB ID 6EDC was used in Phaser for molecular replacement to solve the phases.[51] The determination of Phaser's success in solving the structure was initially assessed through the evaluation of LLG (Log-Likelihood Gain) and TFZ (Translation Function Z-score) scores, and subsequent adjustments were made to the initial model (in the form of a .pdb file) based on these scores. Subsequently, a single PDB file containing the best solutions was employed for subsequent refinement steps, as determined by LLG, TFZ, and RFZ (Rotation Function Z-score). Following molecular replacement, the refinement process was conducted using Phenix Refine[171] with refining the coordinates (specifically rigid body refinement and XYZ coordinate refinement) along with atomic displacement parameters (isotropic B-factors).

6.3 Results and discussion

6.3.1 Over expression and purification of hcGAS

The pRSFDuet-sumo-hcGAS (K427E/K428E) plasmid harbored in DH5 alpha cells was inoculated for overnight growth at 37 °C and the isolated plasmid was transformed in *E. coli* BL21-CodonPlus (DE3)-RIL competent cells (**Fig. 60**). A single colony from transformed *E. coli* BL21-CodonPlus (DE3)-RIL cells was then inoculated for over expression of the protein.

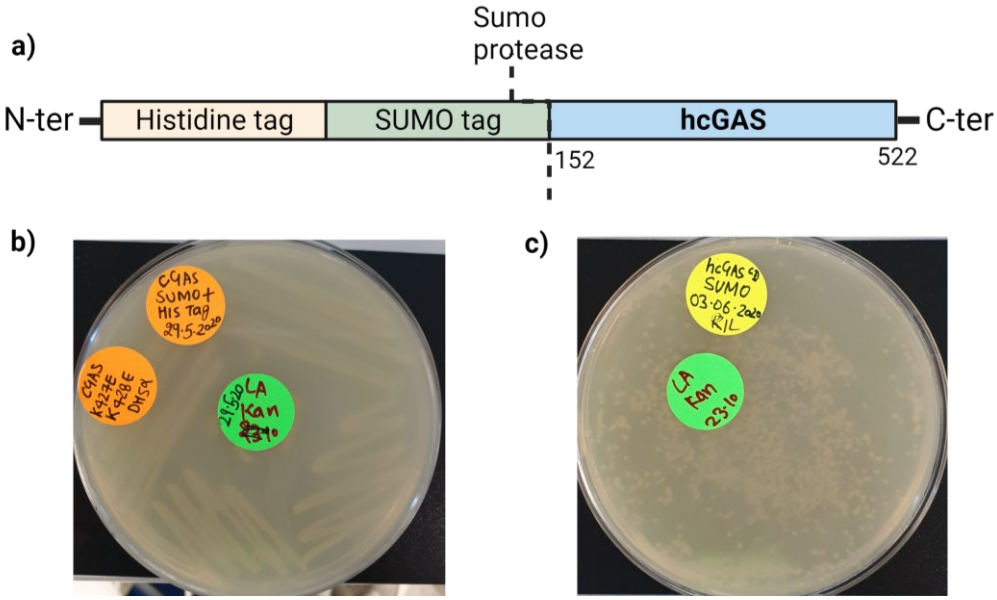


Fig. 60. a) hcGAS construct design with N terminus histidine and SUMO tag. b) pRSFDuet-sumo-hcGAS (K427E/K428E) transformed in DH5 alpha cells for plasmid isolation. c) pRSFDuet-sumo-hcGAS (K427E/K428E) transformed in *E. coli* BL21-CodonPlus (DE3)-RIL cells for over expression.

After the overnight expression, cells were lysed using sonication for 6×1 minutes on ice with a 2-minute break after each round of sonication. It is essential to mention that longer sonication durations lead to precipitation of proteins; thus, a total of 6 minutes of sonication protocol was optimal. Crude whole lysate (10 µl) sample after sonication was analyzed on SDS-PAGE. The rest of the post-sonication sample was centrifuged, and the supernatant was analyzed on SDS-PAGE. As can be seen in **Fig. 61**, His_6-SUMO hcGAS (K427E/K428E) expressed well, and the majority was found in the soluble supernatant fraction (**Fig. 61**, lane 3). Once the over expression in the soluble fraction was confirmed by SDS-PAGE, the supernatant, after sonication, was subjected to affinity purification on Ni-NTA resin using an increasing concentration of imidazole containing buffer A as eluent. Elution profiles are shown in **Fig. 62**.

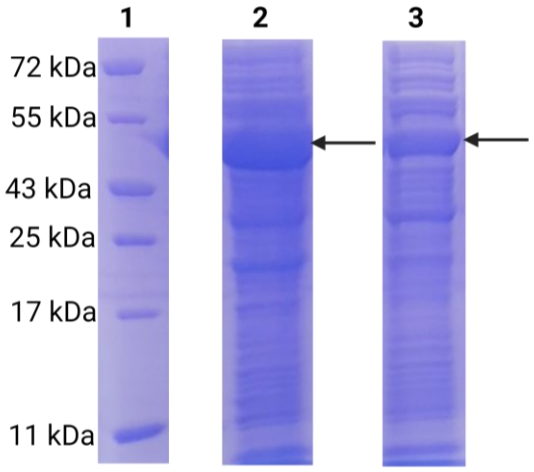


Fig. 61. Over expression profile of His_6-SUMO hcGAS (K427E/K428E). Recombinant protein over expressed with IPTG induction and analyzed on a 12 % SDS-PAGE. Lane 1 is the protein ladder marker; Lane 2 is the crude sample after sonication; Lane 3 is the supernatant after sonication.

To remove impurities of initial washes, the Ni-NTA resin were given with buffer A supplemented with 50 mM imidazole. Afterward, His_6-SUMO hcGAS (K427E/K428E) was eluted with 500 mM imidazole. In order to avoid precipitation of the protein due to the high concentration of imidazole, a dialysis exchange to buffer A was performed.

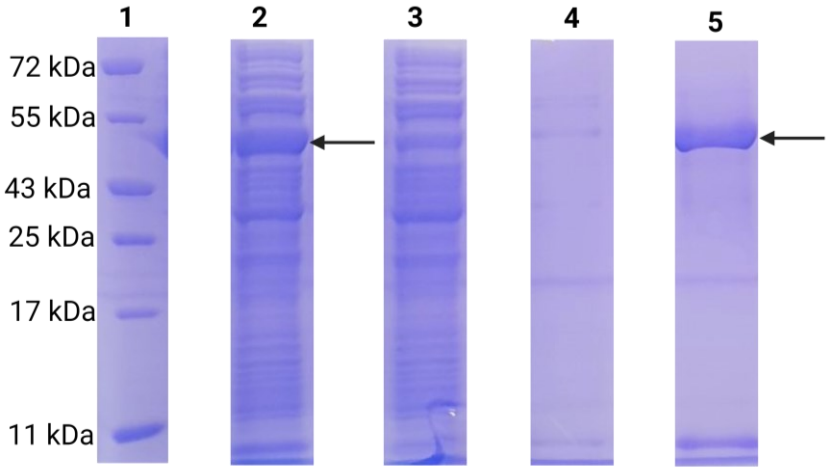


Fig. 62. Affinity purification profile of His_6-SUMO hcGAS (K427E/K428E) analyzed on a 12 % SDS-PAGE gel. Lane 1 is the protein ladder marker; Lane 2 is supernatant after sonication; Lane 3 is the flow-through fraction of Ni-NTA resin after binding to the protein of interest; Lane 4 is the wash fraction with buffer A supplemented with 50 mM imidazole; Lane 5 is the elution faction with buffer A supplemented with 500 mM imidazole. His-SUMO hcGAS (K427E/K428E) protein is highlighted with an arrow.

Once purified on Ni-NTA resin, the elution fractions were concentrated to 2 ml volume and purified on Superdex 200 size exclusion column eluting pure protein between 50 to 60 ml volume and analyzed on 12 % SDS-PAGE for purity (**Fig. 63**).

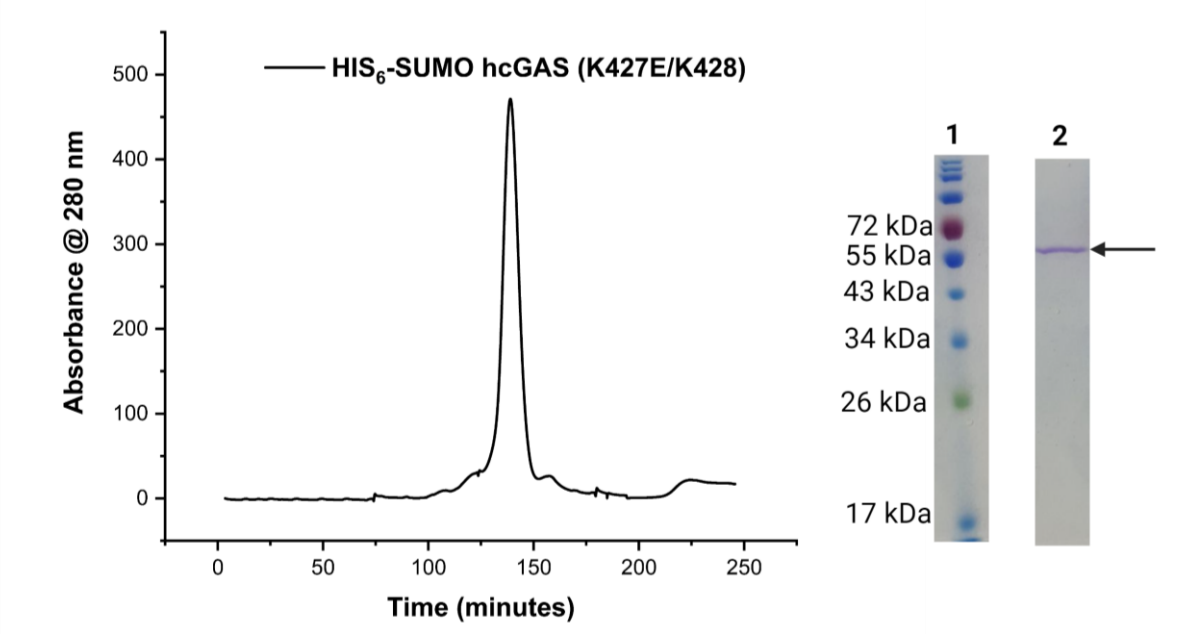


Fig. 63. Size exclusion chromatography (16/600 Superdex 200 column prep grade column) and 12 % SDS-PAGE profile of His_6-SUMO hcGAS (K427E/K428E). Please note that the x-axis on the chromatogram is in minutes. The purification was performed at a 0.4 ml/min flow rate. Lane 1 in SDS-PAGE is a pre-stained protein ladder; Lane 2 is the pure fraction of His_6-SUMO hcGAS (K427E/K428E).

The His6-SUMO tag was not cleaved from hcGAS (K427E/K428E) to preserve the yield. Subsequently, the His_6-SUMO hcGAS (K427E/K428E) protein, once purified through size exclusion chromatography, was utilized for crystallization with a dsDNA sequence, resulting in the successful generation of crystals. The details of the X-ray diffraction data collected from these crystals will be discussed later in this chapter. However, our attempts to crystallize His_6-SUMO hcGAS (K427E/K428E) in conjunction with foldamers did not yield any successful crystal formations. Based on this observation, we decided to proceed by removing the His_6-SUMO tag. To achieve this, we expressed and purified His_6-SUMO hcGAS (K427E/K428E) in a 4 L LB media (four times bigger scale compared to the previous expression) using the same protocol as previously described in this section.

Following the first round of purification using Ni-NTA chromatography, the elution fraction was subjected to dialysis in combination with ULP1 protease to cleave off the His_6-SUMO tag. As the ULP1 protease possessed an N-terminal histidine tag, we performed a second Ni-NTA purification step. In this step, the cleaved hcGAS (K427E/K428E) was found in the flow-through fraction, while the cleaved His_6-SUMO tag and the ULP1 protease remained bound to the Ni-NTA resin. To assess the purity of the resulting samples, we conducted an SDS-PAGE (**Fig. 64**) analysis before proceeding with the crystallization experiments involving the foldamers.

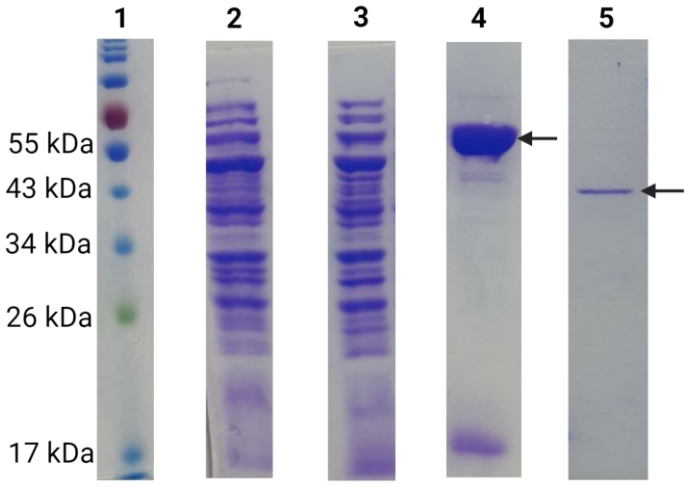


Fig. 64. Over expression and purification profile of His_6-SUMO hcGAS (K427E/K428E) analyzed on a 12 % SDS-PAGE. Lane 1 is the pre-stained protein ladder marker; Lane 2 is the crude sample after sonication; Lane 3 is the supernatant after sonication; Lane 4 is the elution profile after the first round of Ni-NTA purification; Lane 5 is flow through of cleaved hcGAS (K427E/K428E) after the second round of Ni-NTA purification.

6.3.2 hcGAS crystallization, data collection, and structure solution

6.3.2.1 His_6-SUMO hcGAS (K427E/K428E)

Efforts toward crystallizing His6-SUMO hcGAS (K427E/K428E) with DNA led to crystal formation within 3 days at 20 °C (**Fig. 65**). X-ray diffraction data (3600 images) collected for these crystals were auto-processed by autoPROC[214] pipeline in P 6_2 (reindexed to P 6_4) crystallographic space group with a resolution of 3.5 Å.

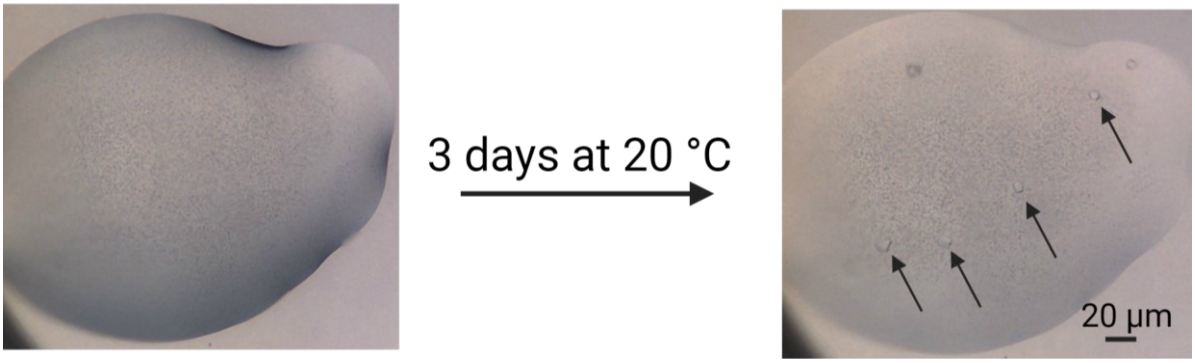


Fig. 65. Crystals of His_6-SUMO hcGAS (K427E/K428E) visualized under the microscope.

Data analysis performed using Xtriage showed no significant defect in the data, except the data appeared to have one or more ice rings. Solvent content and Matthews coefficient analysis were performed and indicated the presence of one copy of the His_6-SUMO hcGAS (K427E/K428E) in the asymmetric unit at 41.7 % solvent content. The structure was solved by molecular replacement using PDB ID 6MJX[211], which gave a top LLG score of 1000.025 and top TFG score of 36.0 with a single solution in the P 6_4 space group, which was then used further for refinement of the structure. After 10 rounds of initial refinement, the r-factor dropped to 30 %

with acceptable values of bonds and angle deviation with a clash score of 3.97 and no Ramachandran outliers or rotamer outliers as predicted by MolProbity.[173] The refinement strategy was further adjusted with TLS parameters and secondary structure restraints to yield an R-work of 29.6 % and an R-free of 33.3 %.

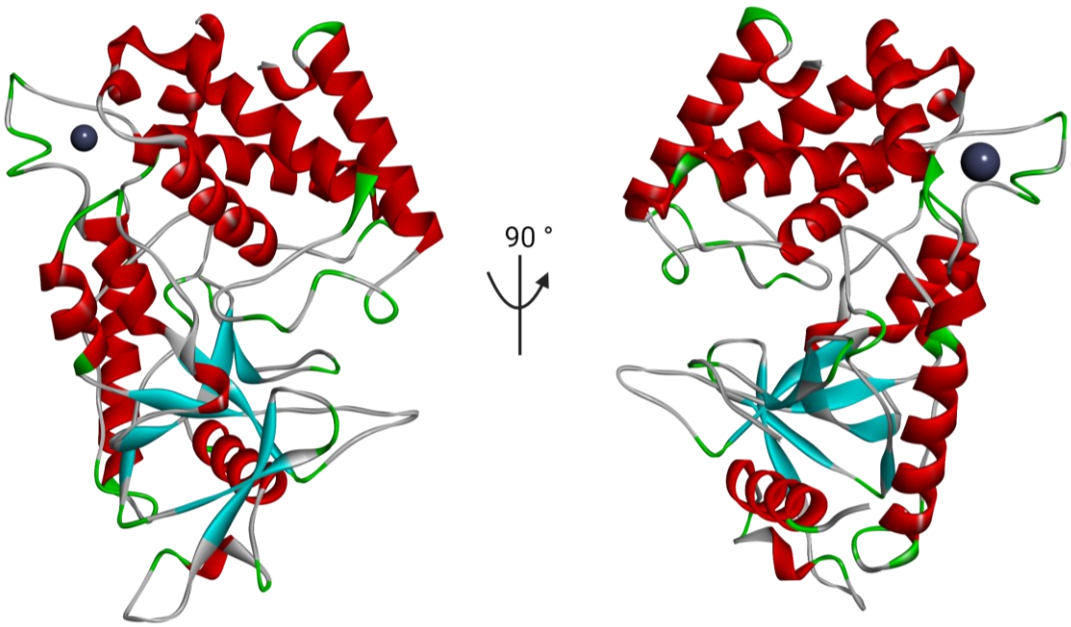


Fig. 66. Crystal structure of hcGAS (K427E/K428E) solved at 3.5 Å. Note that the His_6-SUMO tag is not resolved due to the lack of electron density around the tag.

During our attempt to crystallize His_6-SUMO hcGAS (K427E/K428E) in a complex with a 17-base pair DNA molecule, we encountered a crystallization outcome where only the protein component exhibited crystalline structure. Notably, electron density corresponding to the DNA was absent, strongly indicating the crystallization of the apo form of the protein (**Fig. 66**). This interpretation was substantiated through rigorous structural refinement, resulting in an R-value below 29% at a resolution of 3.5 Å (**Table 8**). This alignment between the experimental data and the provided model provided further support for the apo protein crystal structure.

The overall protein fold remained consistent with the previously crystallized structure of hcGAS (bound with cGAMP, PDB ID 6MJX). In particular, the zinc finger motif maintained its structural integrity, with coordination involving the residues H390, C396, C397, and C404, along with Zn^{2+}, as observed in previous structures (**Fig. 67**).

Table 8. Crystallography data collection and structure refinement statistics for His_6-SUMO hcGAS (K427E/K428E).

Resolution range	41.71 - 3.51 (3.635 - 3.51)
Space group	P 6_4
Unit cell	116.407 116.407 59.803 90 90 120
Total reflections	114766 (12458)
Unique reflections	5798 (595)
Multiplicity	19.8 (20.9)
Completeness (%)	97.84 (100.00)
Mean I/sigma(I)	16.18 (2.55)
Wilson B-factor	118.48
R-merge	0.1451 (1.42)
CC1/2	0.999 (0.756)
CC*	1 (0.928)
Reflections used in refinement	5788 (595)
Reflections used for R-free	270 (24)
R-work	0.2908 (0.3373)
R-free	0.3416 (0.4769)
CC (work)	0.917 (0.631)
CC (free)	0.742 (0.452)
Number of non-hydrogen atoms	2864
macromolecules	2863
ligands	1
Protein residues	353
RMS (bonds)	0.002
RMS (angles)	0.5
Ramachandran favored (%)	92.55
Ramachandran allowed (%)	7.16
Ramachandran outliers (%)	0.29
Rotamer outliers (%)	0
Clash score	8.04
Average B-factor	137.18
macromolecules	137.19
ligands	101.29
Number of TLS groups	3

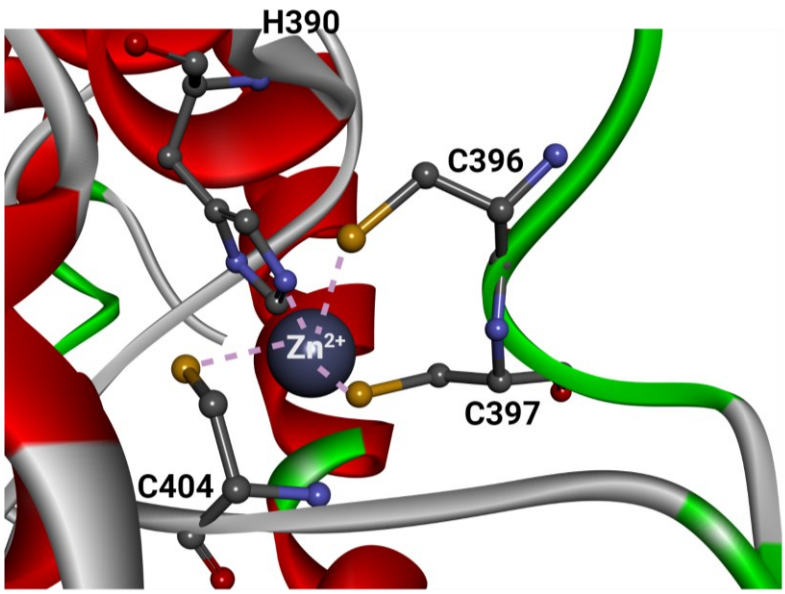


Fig. 67. Zoomed view of the zinc finger motif in the crystal structure of the hcGAS (K427E/K428E).

The primary objective behind crystallizing hcGAS (with and without DNA) was to validate that the protein had maintained a well-folded, structurally stable state following the purification procedures. This step was crucial in ensuring that the hcGAS was of crystallization quality before attempting crystallization experiments with the foldamers.

6.3.2.2 His_6-SUMO hcGAS (K427E/K428E)-foldamer 1

Attempts toward crystallization of His_6-SUMO hcGAS (K427E/K428E)-foldamer complexes were initially made by screening broad chemical space using the sitting drop vapor diffusion method. After initial hits in broad screening plates, conditions were optimized to give reproducible crystals for His_6-SUMO hcGAS (K427E/K428E)-foldamer **1** in 3 days at 20 °C using the hanging drop vapor diffusion method (**Fig. 68**).

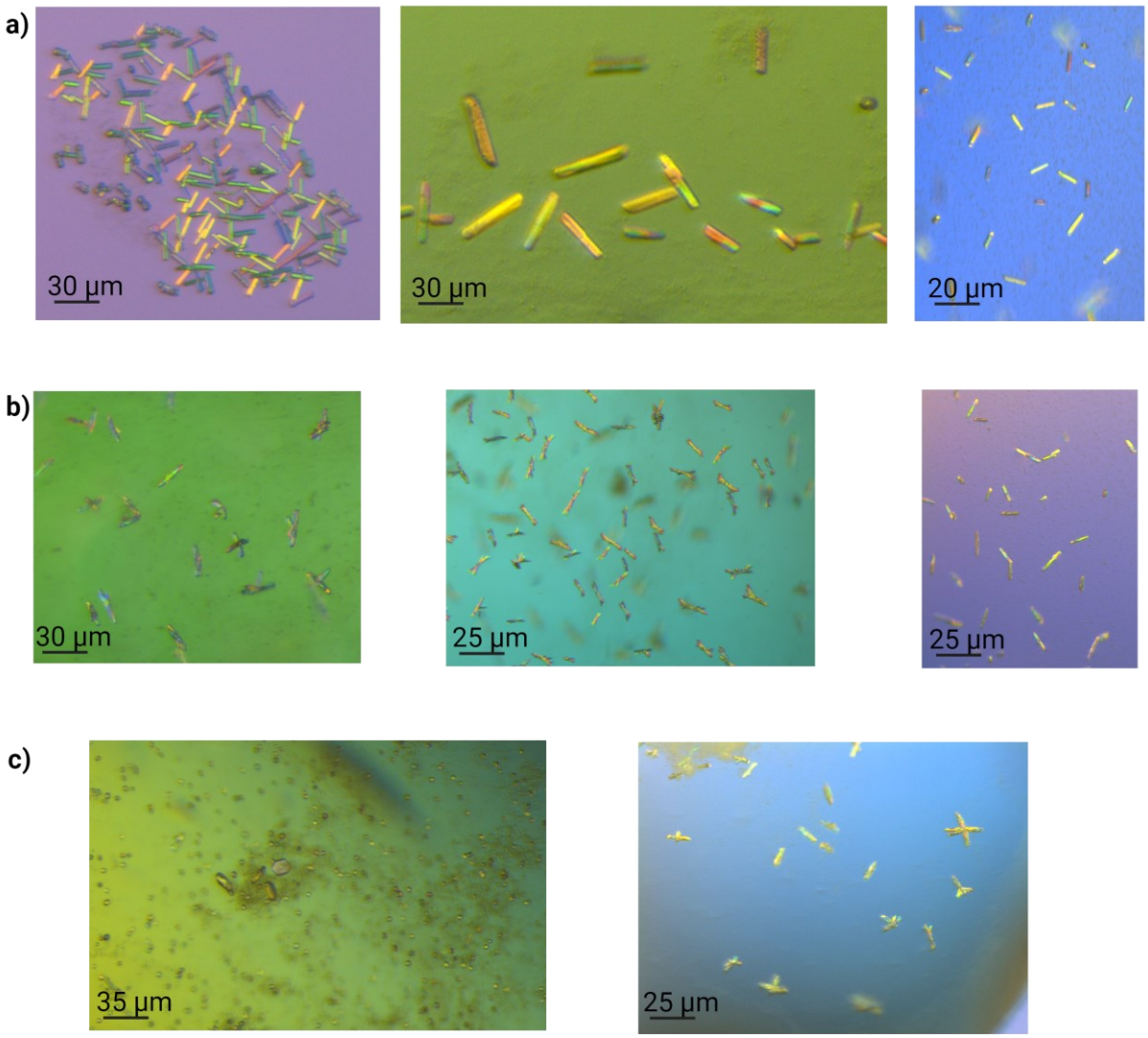


Fig. 68. Crystals of hcGAS (K427E/K428E)-foldamer **1** observed under crossed polarizing microscope grown in different crystallization solutions. a) 35% (v/v) MPD, 100 mM HEPES/ Sodium hydroxide pH 7.5, 200 mM Sodium chloride. b) 35% (v/v) MPD, 100 mM Tris base/ Hydrochloric acid pH 7.0, 200 mM Sodium chloride. c) 50% (v/v) Polyethylene glycol 200, 100 mM TRIS pH8.0.

Several crystals grew using various crystallization reagents and different concentrations of the protein-foldamer complex. These crystals were subsequently subjected to X-ray diffraction experiments at synchrotron beamlines. However, despite numerous attempts, these crystals only diffracted to a lower resolution (~15 Å). We also attempted to improve the diffraction quality by incorporating cofactors like zinc and testing various cryoprotection conditions during the flash cooling of crystals in liquid nitrogen. However, these strategies did not yield better results.

To achieve improved resolution datasets, we experimented with different temperatures for crystal growth: 4 °C, 16 °C, and 20 °C, using conditions that had previously yielded crystals successfully but with poor resolution diffraction (**Fig. 69**).

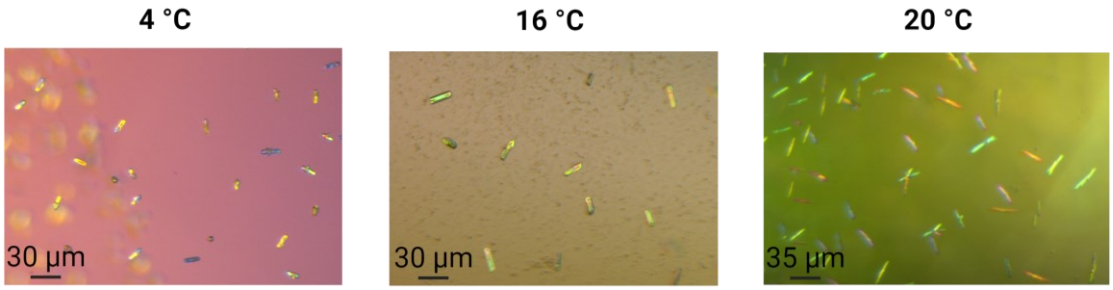


Fig. 69. Crystals of hcGAS (K427E/K428E)-foldamer **1** observed under crossed polarizing microscope grown in different temperatures in 35% (v/v) MPD, 100 mM HEPES/ Sodium hydroxide pH 7.5, 200 mM Sodium chloride as crystallization solution.

Notably, crystals grew at all three temperatures when utilizing a crystallization solution comprising 35% (v/v) MPD, 100 mM HEPES/Sodium hydroxide at pH 7.5, and 200 mM Sodium chloride. These crystals exhibited variations in morphology and size (**Fig. 69**). Subsequently, these crystals were subjected to X-ray diffraction experiments for data collection at the ID 23-1 beamline at the ESRF synchrotron facility in Grenoble (**Fig. 70**).

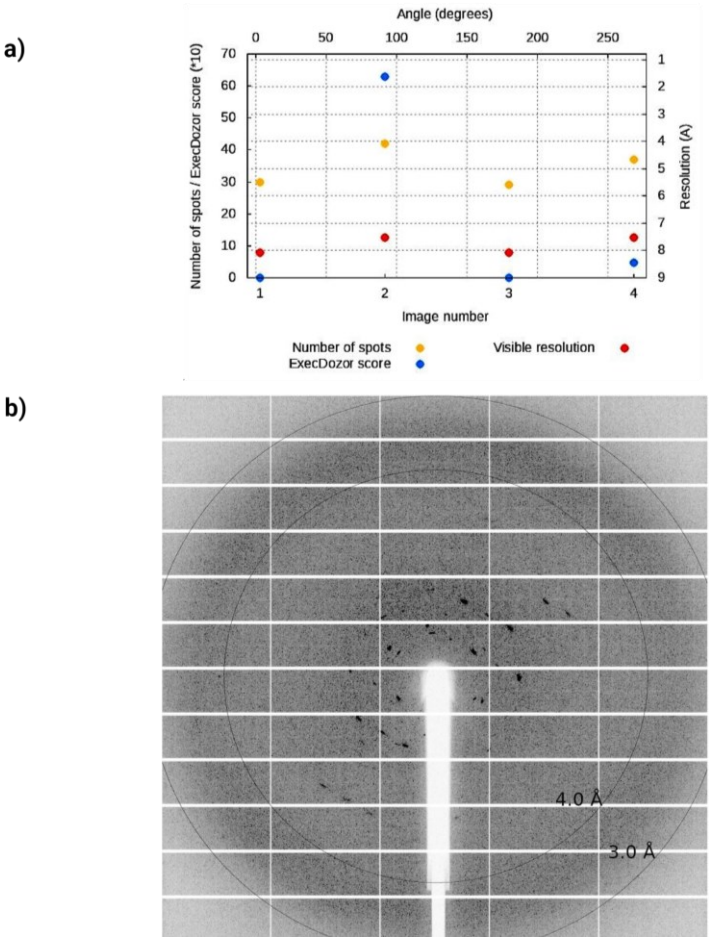
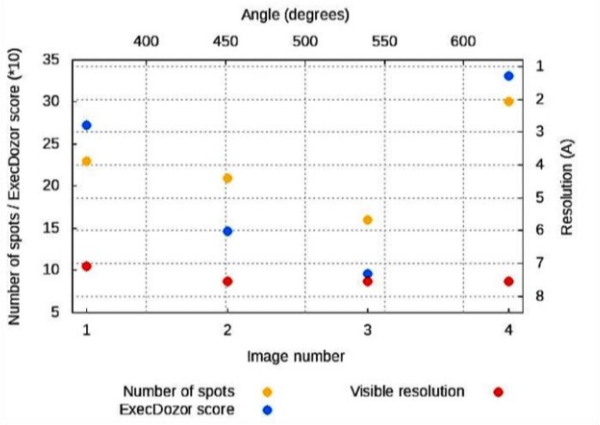


Fig. 70. a) Graph highlighting the resolution versus number of spots during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** grown at 4 °C in 35% (v/v) MPD, 100 mM HEPES/Sodium hydroxide at pH 7.5, 200-mM Sodium chloride. b) Diffraction map image (1 of 4) during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** complex.

a)



b)

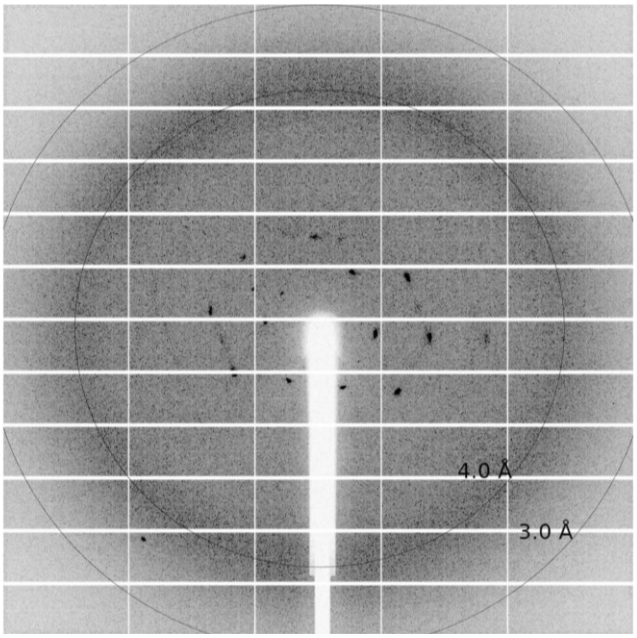


Fig. 71. a) Graph highlighting the resolution versus number of spots during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** grown at 20 °C in 50% (v/v) Polyethylene glycol 200, 100 mM TRIS pH8.0. b) Diffraction map image (1 of 4) during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** complex.

Through variations in crystal growth temperatures and crystallization conditions, we managed to attain a resolution ranging from 6 to 7 Å during the characterization phase of the diffraction process (**Fig. 70** and **Fig. 71**). Despite this improvement in resolution, the crystals still fell short of sufficient resolution for successful indexing. Consequently, we were unable to extract information concerning the unit cell parameters and space group from these crystals.

6.4 Conclusion and perspective

In conclusion, we discussed the over expression, purification, and crystallization of hcGAS protein. Our efforts led to the successful over expression of His_6-SUMO hcGAS (K427E/K428E) and the optimized protocol with affinity purification using Ni-NTA resin resulted in the isolation of pure protein, which was further purified through size exclusion chromatography. An important strategic decision was made to retain the His_6-SUMO tag,

primarily due to concerns regarding yield loss during tag cleavage. This decision proved instrumental in facilitating the crystallization of the apo-protein. We thus successfully crystallized His_6-SUMO hcGAS (K427E/K428E), achieving crystals suitable for X-ray diffraction data collection. However, attempts to crystallize this protein with foldamers did not yield successful crystal formations, prompting us to consider tag removal. Thus, we cleaved off the His_6-SUMO tag using ULP1 protease and subjected it to crystallization with the DNA mimic foldamers. Despite our best efforts, crystallization with foldamers did not yield high-resolution data. We also explored different crystallization conditions, temperatures, and concentrations, which improved the resolution but still fell short of the indexing criteria.

Presently, we are actively engaged in exploring new chemical spaces that are conducive to crystal growth, with the aim of potentially obtaining higher-resolution datasets. Simultaneously, we are undertaking efforts to crystallize hcGAS in complex with different DNA mimic foldamers, including C_2-symmetrical and sticky end foldamers (**Fig. 72**).

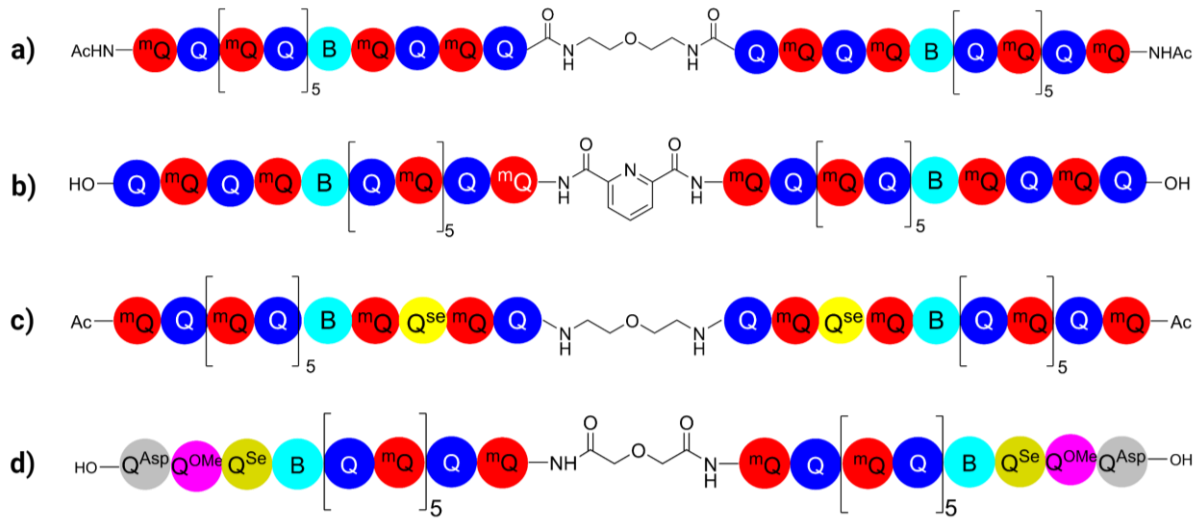


Fig. 72. New generation of foldamer sequences to be screened with hcGAS for crystallization. a,b) C_2-symmetrical DNA mimic foldamer (34 mer) with flexible and rigid C_2-symmetrical linker. c) C_2-symmetrical DNA mimic foldamer (34 mer) with Q-selenium monomer. d) C_2-symmetrical DNA mimic foldamer (32 mer) with sticky ends.

This is similar to the approach in protein-DNA crystallization, where systematically, a range of DNA lengths are screened to determine the optimal number of base pairs that produce crystals with a high-resolution dataset. Likewise, we are extending similar efforts to crystallize hcGAS with DNA mimic foldamers of varying lengths (**Fig. 72**). Foldamer sequences shown in **Fig. 72** a), b), c) are a mimic of 17 base pair DNA, and d) is a mimic of 16 base pair DNA. These foldamers present the ideal length to be recognized by hcGAS.

7. Targeting Dpo4 by DNA mimic foldamers

Cell division is a fundamental biological process essential for the growth, development, and maintenance of multicellular organisms. During cell division, DNA polymerases copy genetic material with remarkable accuracy, replicating millions of nucleotides. DNA polymerases are pivotal enzymes responsible for the replication and repair of DNA, playing a crucial role in maintaining genetic integrity and stability within living organisms. These enzymes facilitate the synthesis of new DNA strands by catalyzing the addition of complementary nucleotides to the template DNA strand during DNA replication.[215,216] DNA polymerases ensure the faithful transmission of genetic information from one generation to the next, as errors in DNA synthesis can lead to mutations and potentially harmful genetic changes. *E. coli*'s main replication enzyme, DNA polymerase III, achieves astonishing fidelity by making an error in only about one out of every 10^5 base pairs. This high level of accuracy is reinforced by innate exonucleolytic proofreading and post-replicative mismatch correction mechanisms, resulting in a minuscule overall error rate of approximately ~10^{-10}. These DNA polymerases have a shape similar to the right hand, consisting of thumb, palm, and finger domains (**Fig. 73**), which all wrap around the DNA fragment.

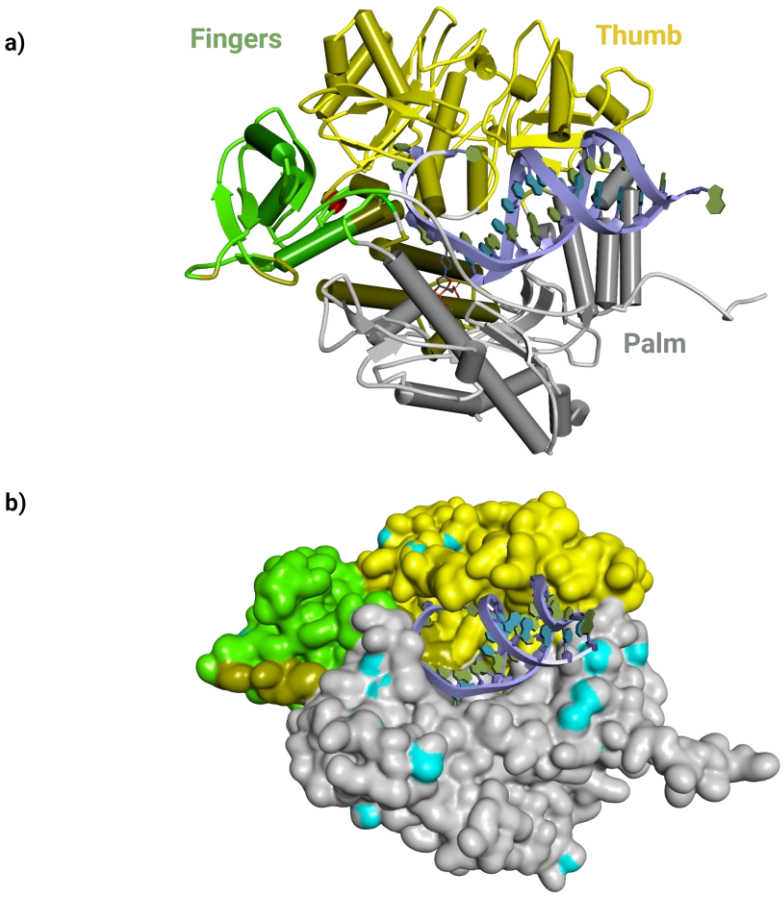


Fig. 73. General "hand" like structural repersentation of a DNA polymerase ternary complex (PDB ID 3NCI) a) and b) shown as schematic and surface representation respectively.

However, despite this impressive accuracy, high-fidelity DNA polymerases can be sensitive to even minor deviations or distortions in the template DNA and incoming nucleoside triphosphate molecules.[217] Two decades ago, it was discovered that specialized DNA polymerases can be utilized to temporarily navigate and bypass these DNA lesions on damaged DNA strands.[218,219] These specialized polymerases serve as substitutes for the high-fidelity polymerase in such instances. Many of the DNA polymerases involved in copying damaged DNA are phylogenetically related. Originally grouped as the "UmuC/DinB/Rev1/Rad30 superfamily," they are now collectively known as the Y-family of DNA polymerases or translesion polymerases.[220] Unlike other polymerases, Y-family polymerases lack the built-in capability to proofread DNA. In addition to their role in copying damaged DNA templates or handling bulky DNA modifications, these polymerases possess a distinct trait: when replicating undamaged DNA, they introduce errors at a rate 10 to 100 times higher than that observed in the other families of polymerases.[217,221] This elevated error rate is a defining characteristic of the Y-family polymerases, setting them apart in the realm of DNA replication. The first crystal structure of a Y-family polymerase was solved of a truncated fragment (but active) of Dbh (DinB homolog) from the *Sulfolobus solfataricus* P1.[222] Another DinB ortholog was identified from the *S. solfataricus* P2 genome and was termed DNA polymerase IV (Dpo4) (**Fig. 74**).[223] It was reported that Dpo4 has limited and non-specific interactions with the replicating base pair, resulting in relaxed base selection and, thus, a lesion bypass and error-prone DNA replication.[224]

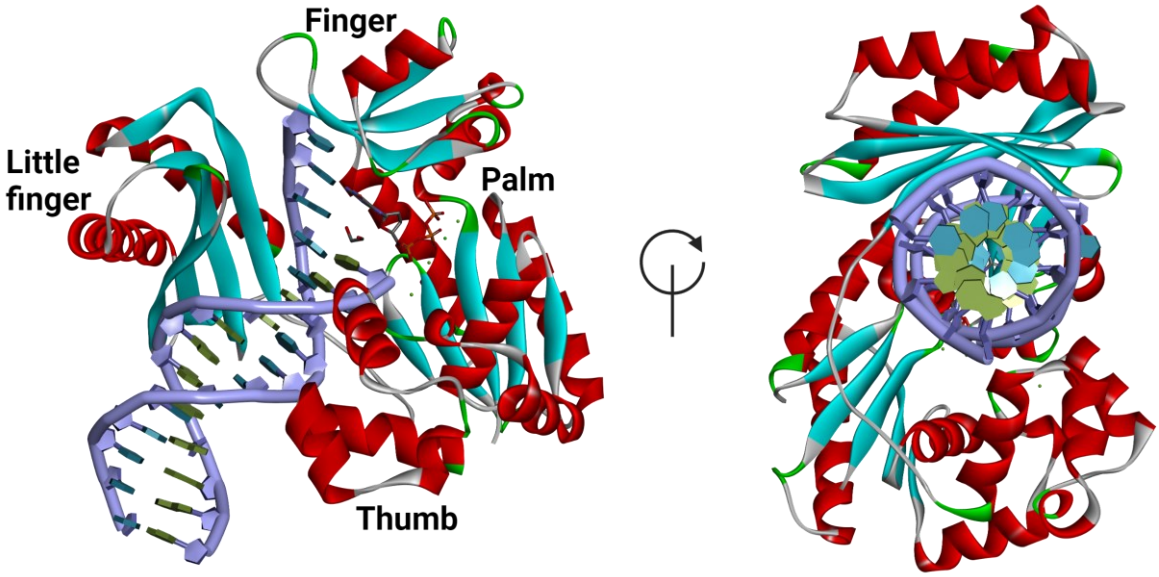


Fig. 74. Crystal structure of the Dpo4 ternary complex.

7.1 Objective of the study

Dpo4, as a polymerase, binds to a d-DNA in order to replicate the DNA strand. However, a recent study by Ahn and colleagues showed that Dpo4 could also bind to l-DNA, as revealed by the crystal structure where a Dpo4 dimer (induced by the little finger domain) results in forming a binding site for l-DNA.[225]

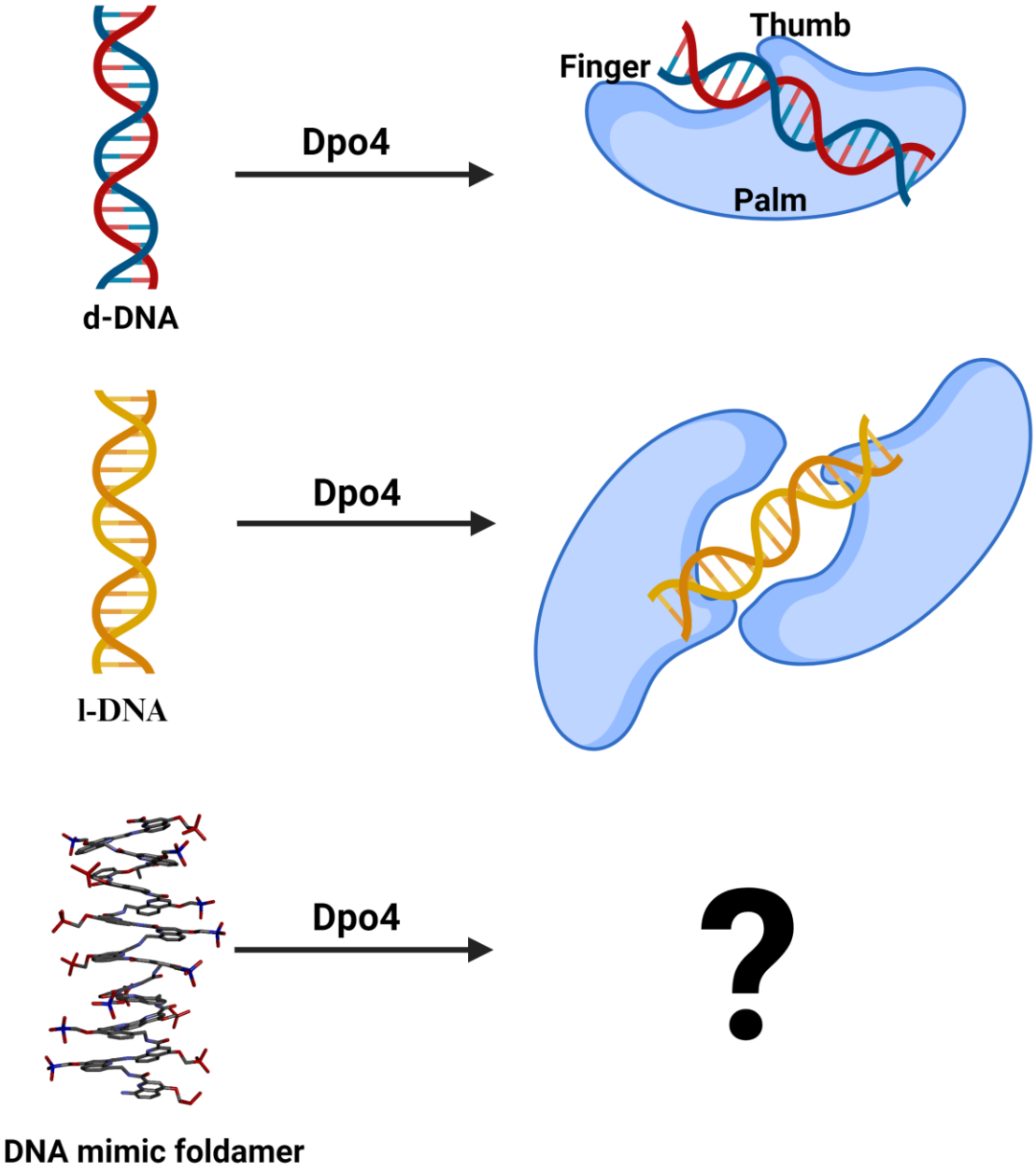


Fig. 75. Schematic representation of the "hand" model of the Dpo4 binding to d-DNA and l-DNA

Dpo4 has two notable properties: the ability to bind to lesion-DNA fragments and the ability to bind to both d-DNA and l-DNA. The promiscuity of Dpo4 makes it an intriguing protein target for studying its interaction with DNA mimic foldamers (**Fig. 75**). DNA mimic foldamers can

be designed to fold in a left-handed main chain helix with right-handed exo-helices of phosphonate side chains (mimicking a d-DNA), and a right-handed main chain helix then (mimicking an l-DNA). We aimed to express and purify the Dpo4 protein and then establish its crystallization with d-DNA and DNA mimic foldamers. This was done to determine whether the foldamer could bind to Dpo4 in a unique binding mode or on a site different from the DNA binding domain. Our study aimed to understand how the Y-family's Dpo4 polymerase could bind to a DNA mimic foldamer. Thus, foldamers may work as molecular tools to interfere with polymerase-DNA interactions.

7.2 Design and methodology

Foldamers used in this chapter are already described in **section 6.2.1** and **Fig. 77**. Foldamers used in this chapter were synthesized by Dr. Valentina Corvaglia and Jiaojiao Wu.

7.2.1 Dpo4 expression and purification

The pET21b plasmid carrying codon-optimized S. solfataricus Dpo4 was kindly received from Prof. Dae-Ro Ahn (University of Science and Technology, Korea).[225] The recombinant plasmid pET21b-Dpo4-His_6 was transformed into the OverExpress™ C41(DE3) competent cells. A single colony was inoculated in 50 ml of LB media supplement with 100 µg/ml ampicillin for overnight expression, which was then inoculated in 1 L of culture. Once the OD at 600 nm reached 0.6, induction was performed for protein expression with 1 mM IPTG at 18 °C for 20 hours. Overnight-induced culture of bacteria was harvested after overnight induction by centrifugation at 4 °C. Following the previously reported procedure,[226] cells were resuspended in buffer A (10 mM KHPO4 (pH 7.0), 10 mM MgAc2, 10 % glycerol, 50 mM NaCl, and 0.1 % 2-mercaptoethanol) and lysed using sonication. The lysate was heated to 78 °C for 12 minutes in a water bath to precipitate thermolabile host proteins, which were subsequently removed through ultracentrifugation at 105000g for 1 hour. The supernatant from centrifugation was added to HisPur™ Ni-NTA resin (Thermo Fisher Scientific, USA) and washed thoroughly with buffer A. The target protein was then eluted with buffer A supplemented with 300 mM imidazole. The Dpo4-containing fractions were concentrated with a 10K cut-off centrifugal filter in the buffer of 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT and purified on a HiLoad® 16/600 Superdex® 75 pg column (Cytiva Life science, USA) and analyzed on SDS-PAGE gels and LC-ESI-MS. The concentration of purified Dpo4 was determined at 280 nm using its calculated extinction coefficient of 24,058 M^{-1} cm^{-1} via spectrophotometry. Dpo4-His_6 protein sequence is shown below.

10	20	30	40	50	60
MIVLFVDFDY	FYAQVEEVLN	PSLKGKPVVV	CVFSGRFEDS	GAVATANYEA	RKFGVKAGIP
70	80	90	100	110	120
IVEAKKILPN	AVYLPMRKEV	YQQVSSRIMN	LLREYSEKIE	IASIDEAYLD	ISDKVRDYRE
130	140	150	160	170	180
AYNLGLEIKN	KILEKEKITV	TVGISKNKVF	AKIAADMAKP	NGIKVIDDEE	VKRLIRELDI
190	200	210	220	230	240
ADVPGIGNIT	AEKLKKLGIN	KLVDTLSIEF	DKLKGMIGEA	KAKYLISLAR	DEYNEPIRTR
250	260	270	280	290	300
VRKSIGRIVT	MKRNSRNLEE	IKPYLFRAIE	ESYYKLDKRI	PKAIHVVAVT	EDLDIVSRGR
310	320	330	340	350	360

TFPHGISKET AYSESVKLLQ KILEEDERKI RRIGVRFSKF IEAIGLDKFF DTLEHHHHHH

7.2.2 Dpo4 crystallization

Crystallization of Dpo4 was attempted with the previously co-crystallized DNA sequence (**Fig. 76**).[225] For crystallization with DNA, the affinity tag was not cleaved from Dpo4. Dpo4 was mixed with d-DNA in 0.08 mM: 0.95 mM (protein: DNA). Crystallization screening was performed in sitting drop vapor diffusion method at 20 °C in 100 nl + 100 nl (protein-DNA complex + crystallization solution) at the Max Planck Institute of Biochemistry, Planegg, Germany.

Template strand 5' TCA CGG AAT CCT TCC CCC 3'
Primer strand 5' GGG GGA AGG ATT CC 3'

Fig. 76. DNA sequence used in crystallization with Dpo4.

Initial crystal hits were obtained in 3 days under a crystallization solution containing (0.1 M BIS-TRIS (pH 7.0), 0.1 M calcium acetate, 2 % glycerol, and 9 % (w/v) PEG3350.

Crystallization of Dpo4 with different foldamer candidates was also performed with initial screenings at the Max Planck Institute of Biochemistry. Hits were later optimized in hanging drop crystallization plates. A range of concentrations for the complex of Dpo4: foldamer was screened from 0.08 mM to 0.3 mM. Dpo4-foldamer **1** and Dpo4-foldamer **2** crystallization hits with their X-ray diffraction are discussed in section **7.3.2.2**.

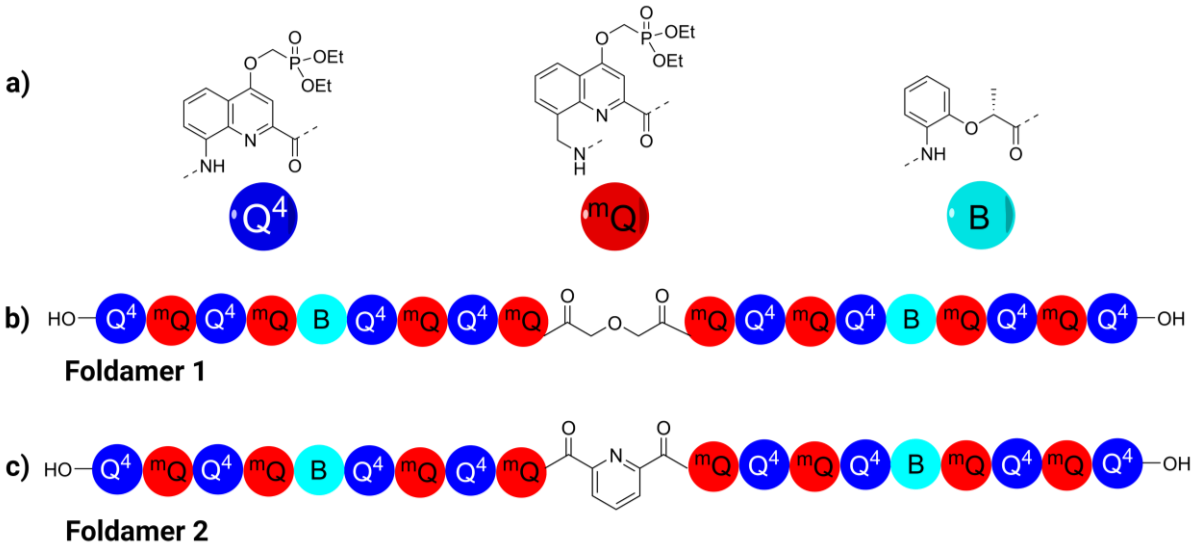


Fig. 77. C_2-Symmetrical DNA mimic foldamers used in crystallization with Dpo4. a) chemical formula of monomers used in DNA mimic foldamers synthesis. b, c) C_2-symmetrical DNA mimic foldamers with a chiral B monomer and central C_2-symmetrical moiety.

7.2.3 Data collection

Crystals from the Dpo4-DNA complex were looped out from the crystallization drop using a 20 µm nylon loop and flash cooled in liquid nitrogen at 100 K. Prior to flash cooling, crystals were soaked in 0.1 M BIS-TRIS (pH 7.0), 20 % PEG3350, 100 mM CaOAC2, and 25 % Glycerol. X-ray diffraction data was collected at the PX beamline at Swiss Light Source (SLS), Switzerland.

Crystals belonging to the Dpo4-foldamer **1** complex were looped out from the crystallization drop using a 20-50 µm nylon loop and flash cooled in liquid nitrogen at 100 K. Prior to flash cooling, crystals were soaked in a variety of cryo-protectants such as 30 % PEG 400, 30 % Glucose, 25 % Ethylene Glycol (all prepared separately in the respective crystallization solution). Crystals were diffracted at beamline P13 at the Deutsches Elektronen Synchrotron (DESY) at the European Molecular Biology Laboratory (EMBL, Hamburg)[213] and beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble).[166]

7.2.4 Structure determination

X-ray diffraction data from crystals of the Dpo4-DNA complex was processed by the XDS.[167] At first, the asymmetric unit analysis was determined by calculating the Matthews coefficient and solvent content using Xtriage. PDB ID 6L84 was used in Phaser for molecular replacement to solve the phases.[51] The determination of Phaser's success in solving the structure was initially assessed through the evaluation of LLG (Log-Likelihood Gain) and TFZ (Translation Function Z-score) scores, and subsequent adjustments were made to the initial model (in the

form of a .pdb file) based on these scores. Subsequently, a single PDB file containing the best solutions was employed for subsequent refinement steps, as determined by LLG, TFZ, and RFZ (Rotation Function Z-score). Following molecular replacement, the refinement process was conducted using Phenix Refine[171] with refining the coordinates (specifically rigid body refinement and XYZ coordinate refinement) along with atomic displacement parameters (isotropic B-factors).

7.3 Results and discussion

7.3.1 Over expression and purification of Dpo4

The pET21b-Dpo4-His_6 plasmid harbored in DH5 alpha cells was inoculated for overnight growth at 37 °C and the isolated plasmid was transformed in OverExpress™ C41(DE3) competent cells and *E. coli* BL21 (DE3)pLysS cells (**Fig. 78**). A single colony from transformed C41 (DE3) cells (Sigma-Aldrich) was then inoculated for over expression of the protein. It is important to note that the expression of Dpo4 in *E. coli* cells is toxic to the bacteria and thus leads to insufficient protein expression. However, the OverExpress™ competent cells are a type of *E. coli* that are highly effective at expressing toxic proteins. These cells have undergone genetic mutations specifically selected for their ability to tolerate toxic proteins. The C41(DE3) strain is a derivative of BL21(DE3) and has at least one mutation that prevents cell death and is associated with the expression of many recombinant toxic proteins. These cells suppress the basal expression of T7 RNA polymerase before induction, which helps to stabilize recombinants that encode particularly toxic proteins such as Dpo4.

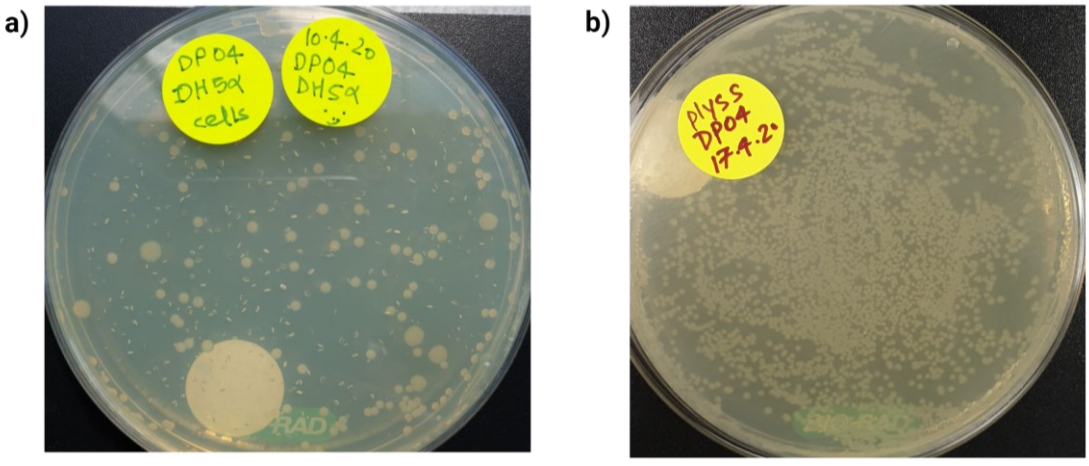


Fig. 78. pET21b-Dpo4-His_6 transformed in DH5 alpha cells for plasmid isolation. c) pET21b-Dpo4-His_6 transformed in *E. coli* BL21 (DE3)pLysS cells for over expression of Dpo4 protein.

After the overnight expression, cells were lysed using sonication for 10×1 minutes on ice with a 2-minute break after each round of sonication. The sonicated crude sample was subjected to

heat treatment at 78 °C for 12 minutes in a water bath to precipitate thermolabile host proteins. Dpo4 comes from *S. solfataricus*, a thermophilic archaea first isolated in 1980 from the hot springs of solfatara, Italy. Due to this, Dpo4 can withstand extremely high temperatures up to 80 °C, maintaining the native fold. However, at such high temperatures, all the *E. coli*-borne proteins will precipitate, leaving Dpo4, which was then subjected to ultracentrifugation to separate the precipitated proteins from active Dpo4 fraction and analyzed on a 12 % SDS-PAGE gel (**Fig. 79**).

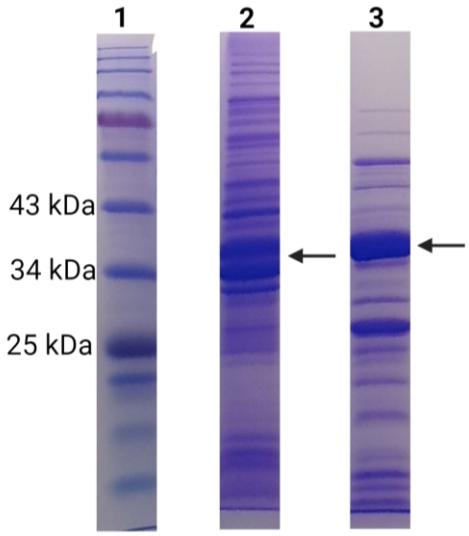


Fig. 79. Over expression profile of Dpo4. Recombinant protein over expressed with IPTG induction and analyzed on a 12 % SDS-PAGE. Lane 1 is the protein ladder marker; Lane 2 is the crude sample after sonication and heating; Lane 3 is the supernatant after ultracentrifugation.

Once the over expression in the soluble fraction was confirmed by SDS-PAGE, the supernatant, after sonication and ultracentrifugation, was subjected to affinity purification on Ni-NTA resin using the increasing concentration of imidazole containing buffer B as eluent. For the removal of impurities of initial washes, the Ni-NTA resin were given with buffer A supplemented with 50 mM imidazole. Afterward, Dpo4 was eluted with 300 mM imidazole. Elution profiles are shown in figure **Fig. 80**. In order to avoid precipitation of the protein due to the high concentration of imidazole, a dialysis exchange to buffer A was performed. This ensured the removal of imidazole and was later concentrated for further purification steps. Once purified on Ni-NTA resin, the elution fractions were concentrated to 2 ml volume and purified on Superdex 75pg size exclusion column, eluting pure protein between 60 to 70 ml volume (**Fig. 81**).

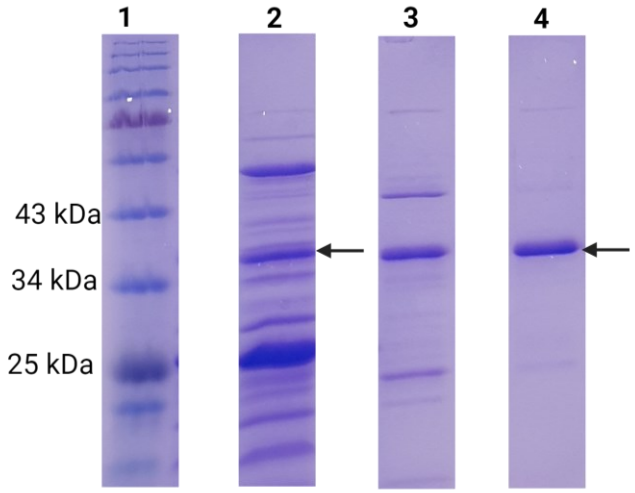


Fig. 80. Affinity purification profile of Dpo4 analyzed on a 12 % SDS-PAGE gel. Lane 1 is the protein ladder marker; Lane 2 is the flow-through fraction of Ni-NTA resin after binding to the protein of interest; Lane 3 is the wash fraction with buffer A supplemented with 50 mM imidazole; Lane 4 is the elution faction with buffer A supplemented with 500 mM imidazole. Note that the Dpo4 protein is highlighted with an arrow.

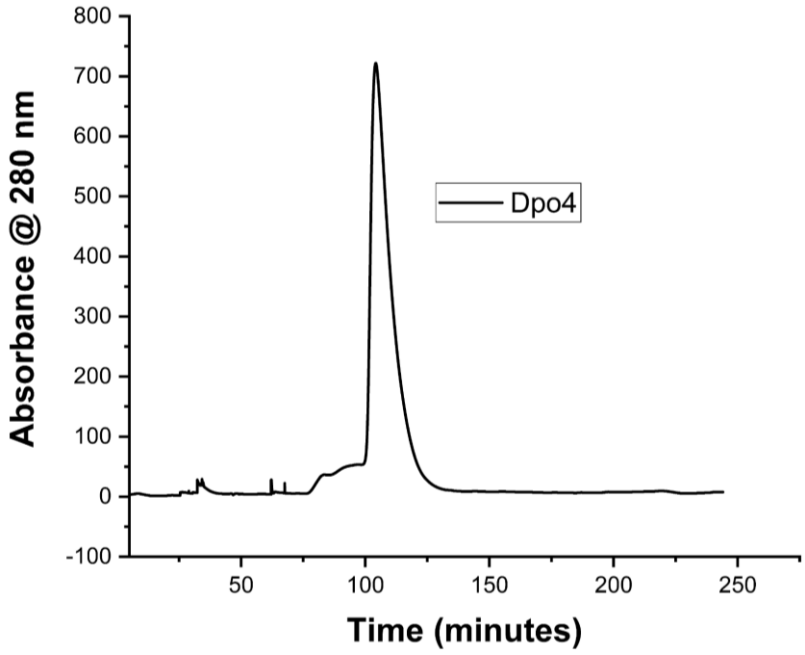


Fig. 81. Size exclusion chromatogram (16/600 Superdex 75 column prep grade column of Dpo4. Please note that the x-axis on the chromatogram is in minutes. The purification was performed at a 0.6 ml/min flow rate.

Fractions from size exclusion chromatography corresponding to Dpo4 were pooled together and analyzed on LC-ESI-MS under the C4 column, 5 % - 95 % in water to acetonitrile (0.1 % formic acid), for purity and identification of the Dpo4. As can be seen in **Fig. 82**, a single peak was visible (at 11 ml volume) corresponding to the Dpo4 protein, which was confirmed by ESI-MS. In our LC-ESI-MS analysis, the mass of the protein was determined to be m/z 2751.5152 Da for the 15+ ion [M+H]+. The reported mass represents the average of two replicates with a standard deviation of ± 0.02 Da.

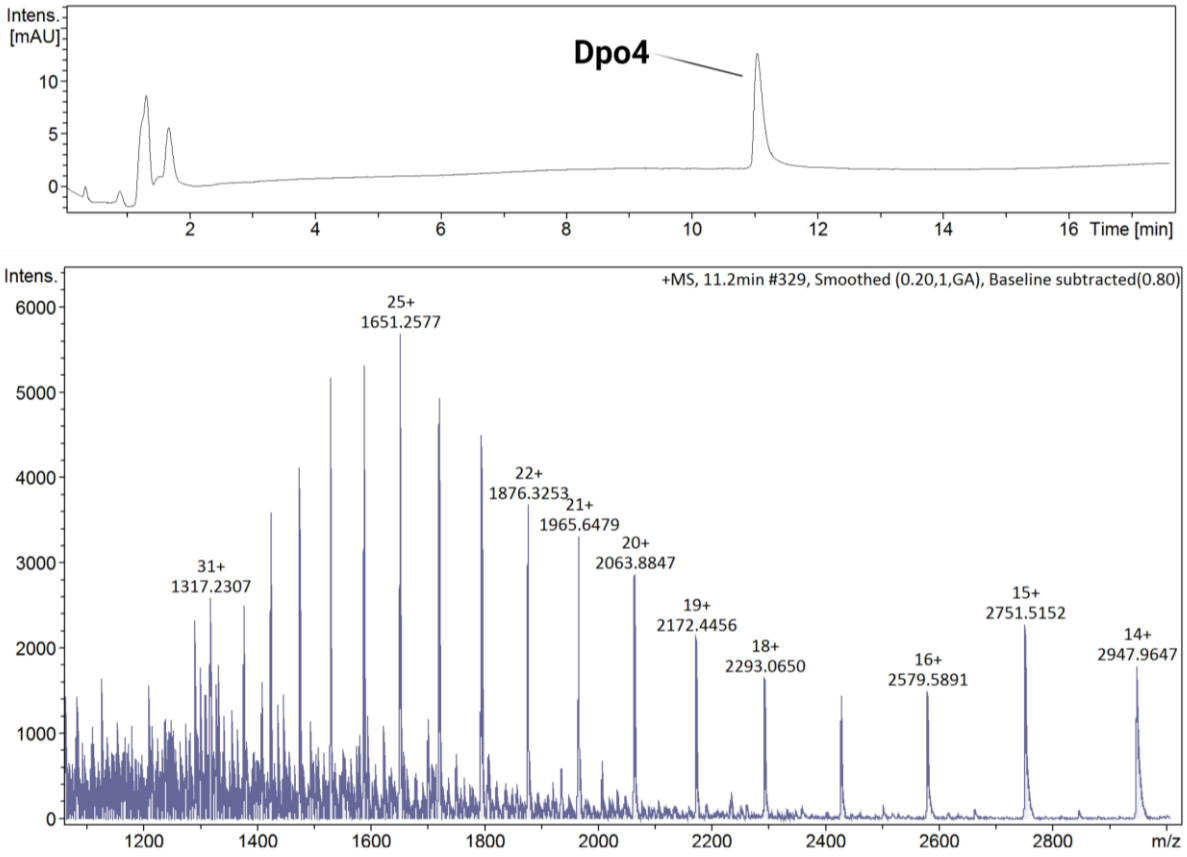


Fig. 82. LC chromatogram at UV 280 nm and ESI-MS spectrum of Dpo4.

7.3.2 Dpo4 crystallization, data collection, and structure solution

Once confirmed with LC-ESI-MS, fractions corresponding to Dpo4 were concentrated up to 11 mg/ml for crystallization with DNA and foldamers.

7.3.2.1 Dpo4-d-DNA crystallization

Efforts toward crystallizing Dpo4 with DNA led to crystal formation within 1 day at 20 °C (**Fig. 83**). The initial crystals that formed were smaller in size, making it impractical to retrieve them using a loop for flash freezing in liquid nitrogen. Consequently, we explored various complex concentrations, which led to single crystals with dimensions ranging from 40 to 50 µm. X-ray diffraction data collected for a single Dpo4-DNA crystal was processed using XDS[167] and indexed in P 2_1 2_1 2 crystallographic space group with a resolution of 2.6 Å.

Data analysis using Xtriage showed no significant defect, except the data appeared to have one or more ice rings. Solvent content and Matthews coefficient analysis were performed and indicated the presence of one copy of the Dpo4-DNA in the asymmetric unit at 52.1 % solvent content. The structure was solved by molecular replacement using PDB ID 6L84[225], which gave a top LLG score of 2272.135 and a top TFG score of 46.2 with a single solution in the P 2_1 2_1 2 space group, which was then used further for refinement of the structure. After 8 rounds

of initial refinement, the r-factor dropped to 30.5 % with acceptable values of bonds and angle deviation with a clash score of 13.58, as predicted by MolProbity.[173] The refinement strategy was further adjusted with TLS parameters and secondary structure restraints to yield an R-work of 24.5 % and an R-free of 29.8 % (**Table 9**).

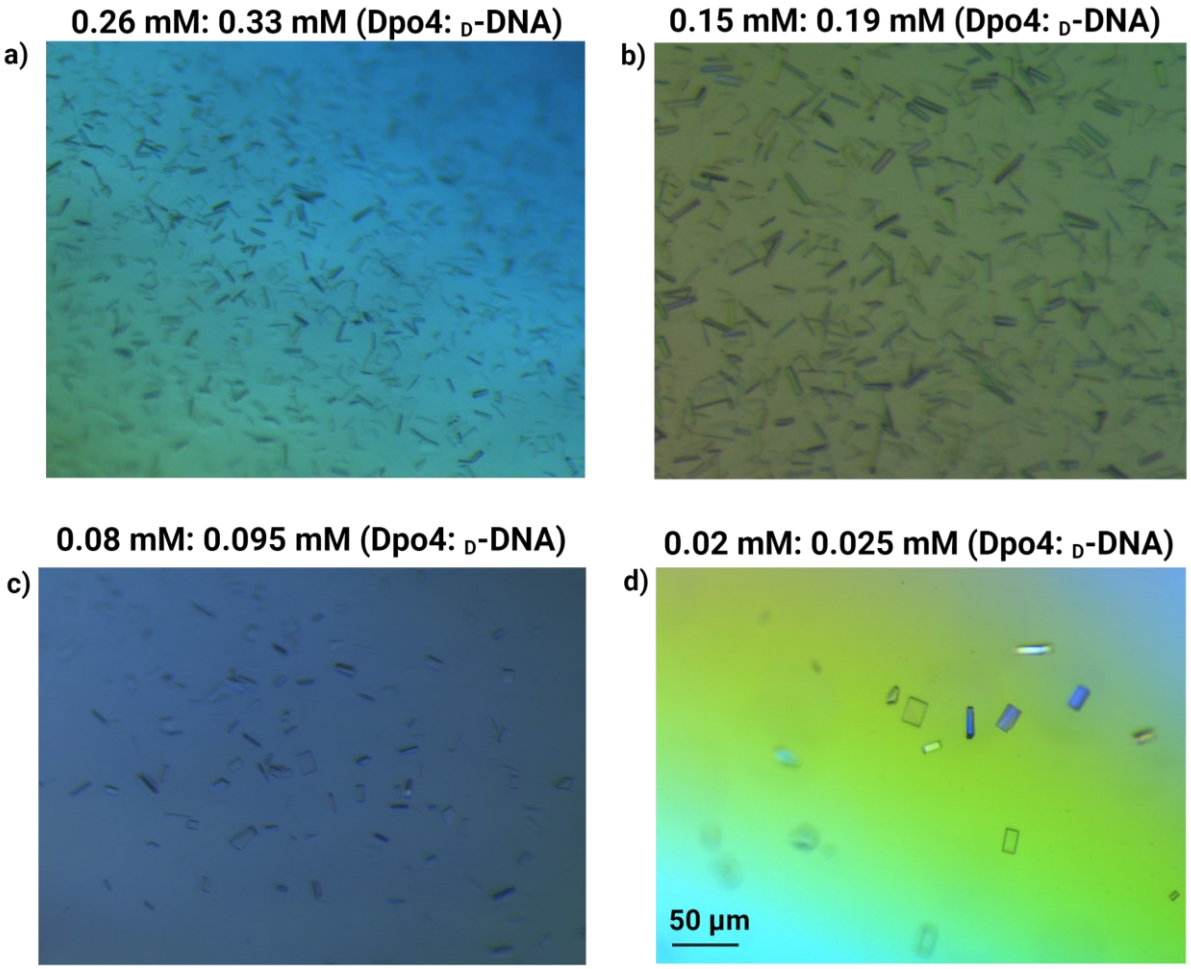


Fig. 83. Crystals of Dpo4-d-DNA complex visualized under the microscope. Note the change from a) to d) in the number of crystals as the concentration of the complex was decreased to yield single crystals.

The structure of Dpo4 has the d-DNA duplex held within its four domains, resembling a hand gripping a rope (**Fig. 84**). This is consistent with other crystal structures of Y-family DNA polymerases with ds-DNA complexes. The minor groove of the d-DNA duplex interacts with the thumb domain. The palm domain is where active site residues bound to divalent ions ($Ca^{2+,}$ yellow spheres in **Fig. 85**) are located, and they play a crucial role in polymerization activity. The finger domain helps interact with incoming nucleotides positioned at the 3′-end of the primer, ready for elongation. For the polymerization of d-DNA, the little finger domain enhances the processivity of Dpo4. It provides a binding interface between the polymerase and the major groove of the d-DNA duplex. The finger domain makes a narrow gap for protruding the 5′-end of the template strand.

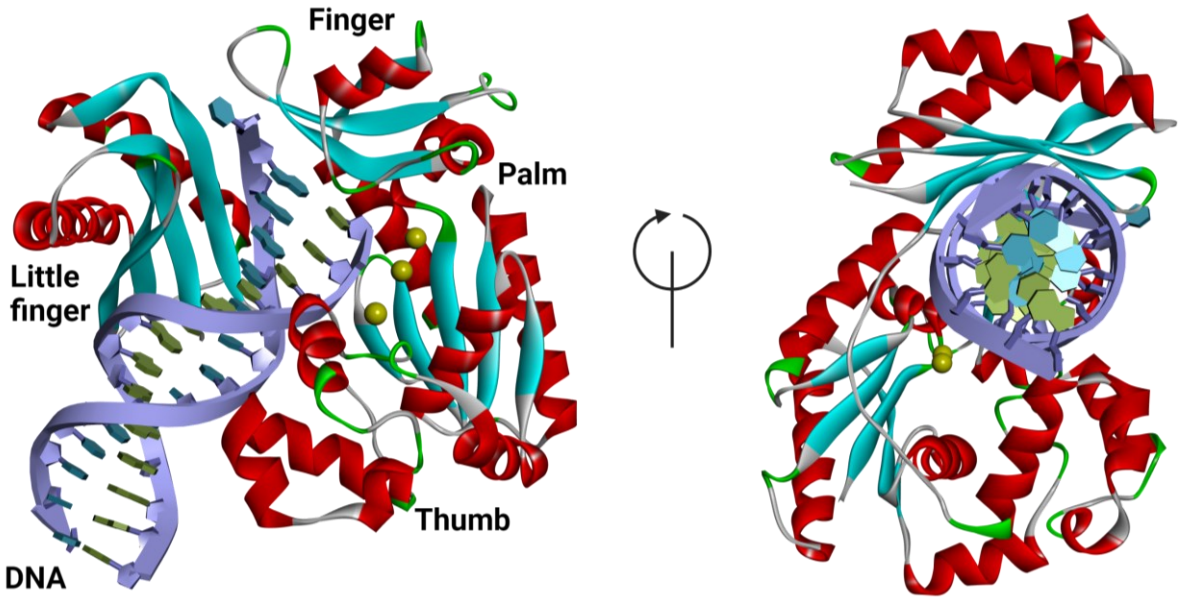


Fig. 84. Crystal structure of the Dpo4-d-DNA solved at 2.6 Å. Note the hand-like structure of Dpo4 grabbing the d-DNA.

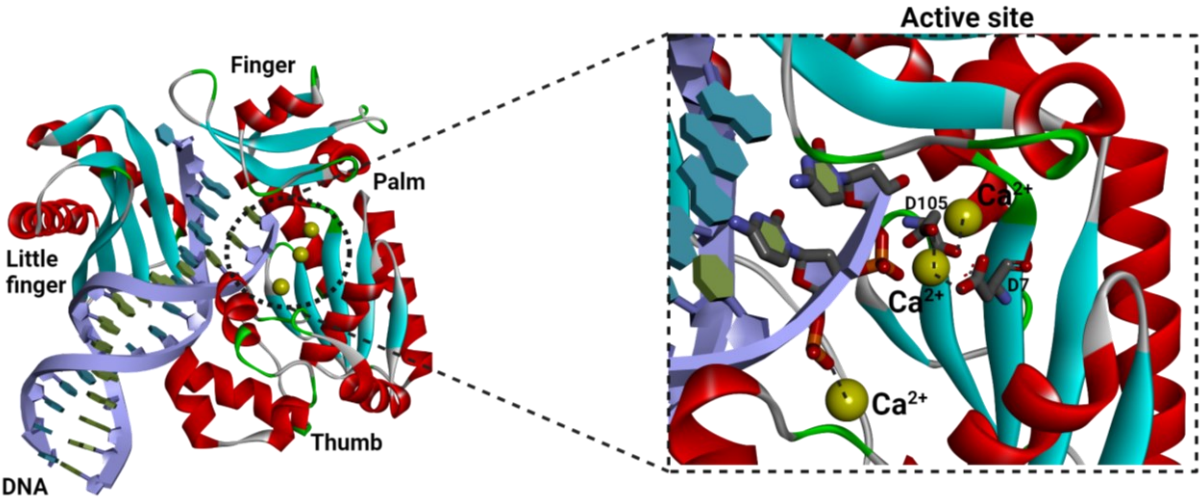


Fig. 85. Crystal structure of the Dpo4- d-DNA complex with a zoom on the active site where Dpo4 residues and DNA interact with the Ca^{2+} ions (yellow spheres).

Table 9. Crystallography data collection and structure refinement statistics for Dpo4-d-DNA.

Resolution range	46.47 - 2.642 (2.736 - 2.642)
Space group	P 2_1 2_1 2
Unit cell	97.863 101.092 52.328 90 90 90
Total reflections	176419 (5381)
Unique reflections	15109 (823)
Multiplicity	11.7 (6.2)
Completeness (%)	92.46 (53.58)
Mean I/sigma (I)	7.18 (0.49)
Wilson B-factor	92.01
R-merge	0.2269 (2.779)
CC1/2	0.998 (-0.124)
CC*	0.999 (-0.533)
Reflections used in refinement	14627 (823)
Reflections used for R-free	733 (42)
R-work	0.2439 (0.6336)
R-free	0.2990 (0.6487)
CC (work)	0.961 (0.065)
CC (free)	0.989 (0.017)
Number of non-hydrogen atoms	3336
macromolecules	3333
ligands	3
solvent	0
Protein residues	341
RMS (bonds)	0.005
RMS (angles)	0.81
Ramachandran favored (%)	92.63
Ramachandran allowed (%)	7.37
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	2.67
Clashscore	14.96
Average B-factor	119.90
macromolecules	119.06
ligands	105.95
Number of TLS group	9

7.3.2.2 Dpo4-foldamer crystallization

Attempts toward crystallization of Dpo4-foldamer **2** were initially made by screening broad chemical space using the sitting drop vapor diffusion method. After initial hits in broad screening plates, conditions were optimized to give reproducible crystals 3 to 7 days at 20 °C using the hanging drop vapor diffusion method.

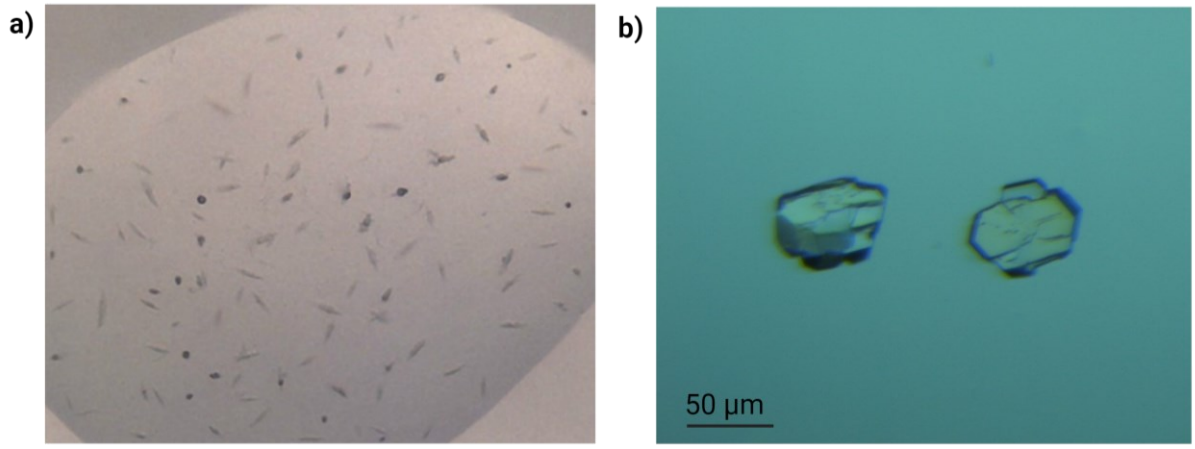


Fig. 86. **Crystals of Dpo4-foldamer 2 observed under crossed polarizing microscope grown in different crystallization solutions.** a) Crystals grown in 35 % MPD, 0.1 mM HEPES pH 7.5. b) Crystals grown in 10 % mPEG 5000, 0.1 M Tris pH 8.5.

Crystallization condition in **Fig. 86** b) led to crystals of the Dpo4-foldamer **2** and diffracted at beamline P13 the Deutsches Elektronen Synchrotron (DESY) at the European Molecular Biology Laboratory (EMBL, Hamburg)[213] to poor resolution.

Then, we moved toward the crystallization of Dpo4-foldamer **1**, which was also initially screened for broad chemical space using the sitting drop vapor diffusion method. After initial hits in broad screening plates, conditions were optimized to give reproducible crystals for Dpo4-foldamer **1** in 3 to 7 days at 20 °C using the hanging drop vapor diffusion method. Crystallization conditions in **Fig. 87** a), b), and c) were grown under different concentrations of the Dpo4-foldamer **1**, which significantly increased the size of the crystals, which, when tested for X-ray diffraction, did not yield even a poor resolution diffraction.

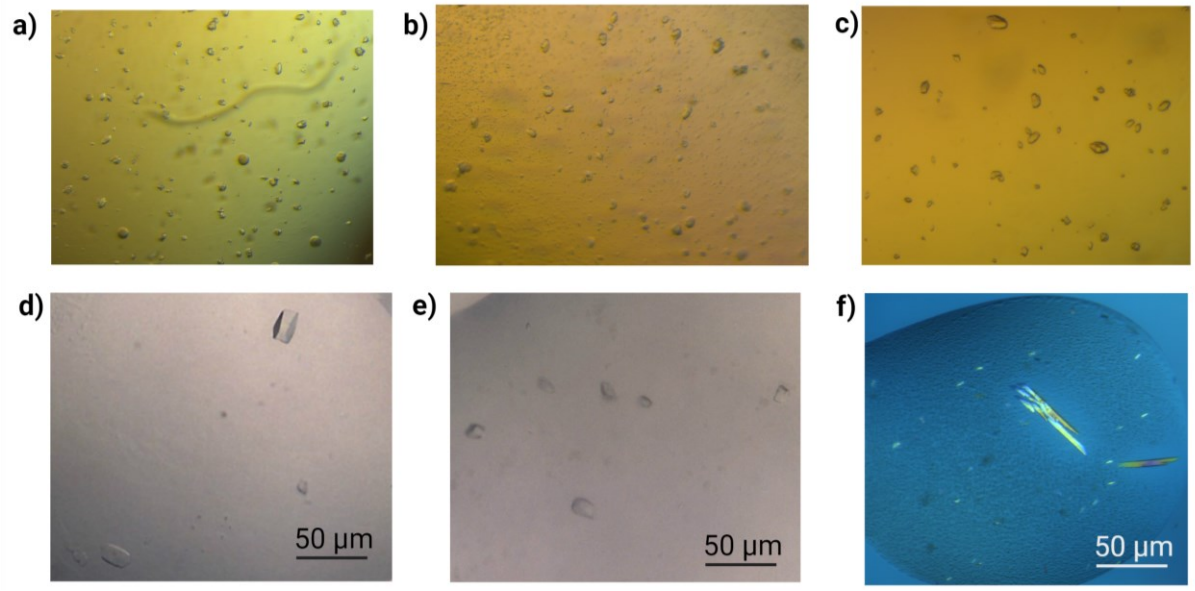


Fig. 87. Crystals of Dpo4-foldamer 1 observed under crossed polarizing microscope grown in different crystallization solutions. a), b), c) Crystals grown in 18 % PEG 4000, 0.1 M Tris pH 8.5 with Dpo4: foldamer **1** in 0.2 mM: 0.26 mM, 0.15 mM: 0.19 mM, and 0.10 mM: 0.13 mM respectively in a), b), and c). d) 1M Ammonium sulfate, 0.1 M Tris pH 8,0. e) 15 % PEG 6000, 0.1M Magnesium acetate, 0.1M Sodium cacodylate pH 6.5. f) 18 % PEG 6000, 0.1 M Tris pH 8.5.

Later on, crystals from **Fig. 87** d), e), and f) were grown under different conditions and diffracted at beamline P13 the Deutsches Elektronen Synchrotron (DESY) at the European Molecular Biology Laboratory (EMBL, Hamburg)[213] to poor resolution (**Fig. 88**).

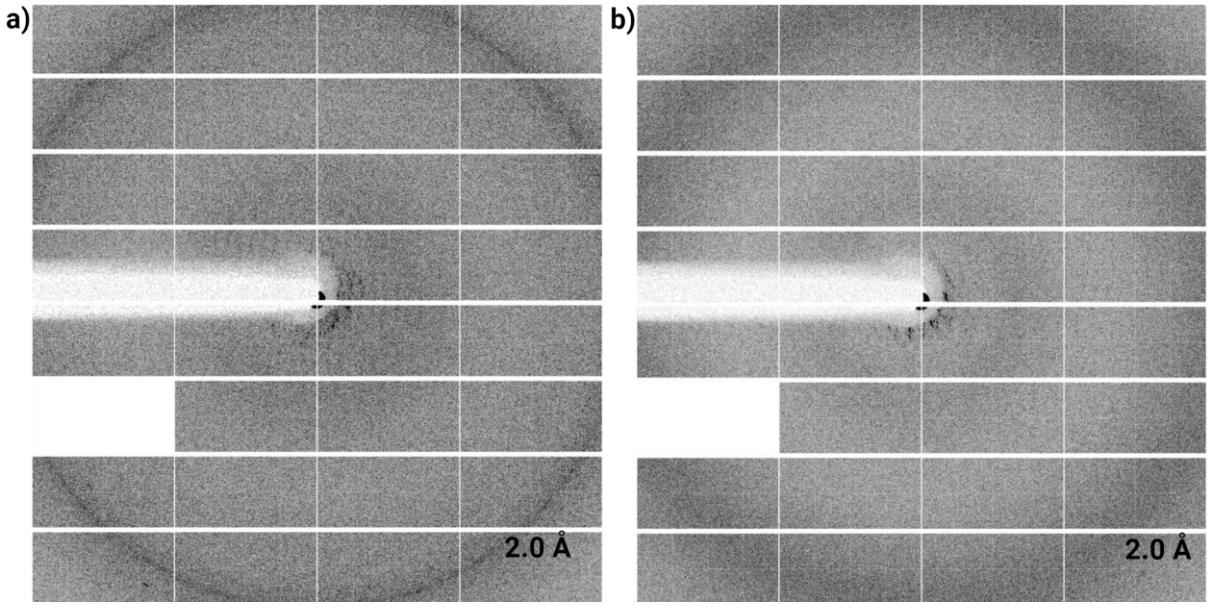


Fig. 88. a) Diffraction map image (1 of 4) during the characterization of crystals of Dpo4-foldamer **1** complex from Fig. 87 d). b) Diffraction map image (1 of 4) during the characterization of crystals of Dpo4-foldamer **1** complex from Fig. 87 e).

Later, new crystallization conditions were explored that lead to crystals of different morphological features, including long rod-shaped crystals that diffracted up to 5-6 Å.

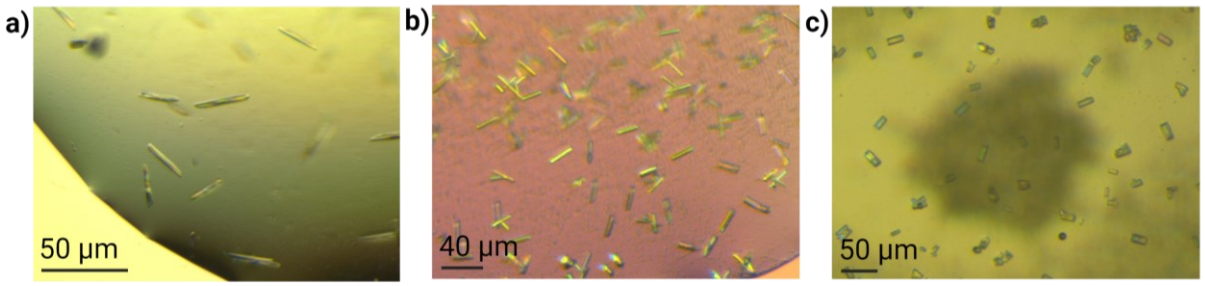


Fig. 89. Crystals of Dpo4-foldamer 2 observed under crossed polarizing microscope grown in different crystallization solutions. a) Crystals grown in 20 % PEG 3350, 0.2 M di-Sodium malonate. b) 20 % PEG 3350, 0.2 M Sodium fluoride. c) Crystals grown in 20 % PEG 3350, 100 mM BIS-TRIS propane pH 6.5, 200 mM Sodium acetate.

Note that the crystals shown in **Fig. 89** c) are similar in morphology and crystallization conditions to the Dpo4-d-DNA crystals shown in **Fig. 83**. These crystals, despite poor diffraction, are closer to yielding a better-resolution dataset. After multiple attempts at optimizing the cryo-conditions prior to flash cooling the crystals in liquid nitrogen, crystals cryo-protected in 25 % Glucose (w/v) led to diffraction upto 3.2 Å at the P13 the Deutsches Elektronen Synchrotron (DESY) at European Molecular Biology Laboratory (EMBL, Hamburg).[213] X-ray diffraction data collected for a single Dpo4-foldamer **1** crystal was processed using autoPROC[214] pipeline in P 6 2 2 crystallographic space group with a resolution of 3.2 Å at a completeness of 70 %.

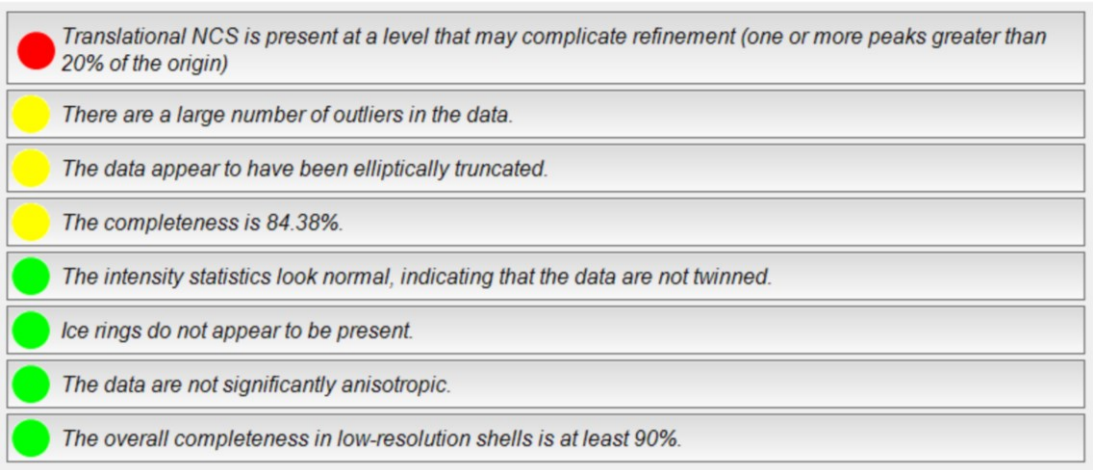


Fig. 90. Xtriage analysis of Dpo4-foldamer **1** crystals from Fig. 89 c).

Data analysis conducted using Xtriage revealed a significant data quality issue, as illustrated in **Fig. 90**. The presence of translational non-crystallography symmetry (TNCS) and a high number of outliers in the data posed substantial challenges for molecular replacement (MR). To gain insights into the crystal composition, solvent content and Matthews coefficient analyses were performed. These analyses suggested the presence of a single copy of the protein and foldamer **1**.

However, when attempting to locate a single copy of the protein for MR, no solutions were obtained. Subsequently, a search for a single copy of foldamer **1** failed to produce any MR solutions (**Table 10**).

Table 10. Crystallography data collection statistics for Dpo4-foldamer **1** crystals.

Resolution range	90.4 – 3.3 (4.15 – 3.29)
Space group	P 6 2 2
Unit cell parameters	40.35 Å 40.35 Å 90.42 Å 90 ° 90 ° 90 °
Multiplicity	9.0 (4.3)
Completeness (%)	82.9 (68.5)
Mean I/sigma (I)	3.8 (0.7)
CC (1/2)	100 (90)

7.4 Conclusion and perspective

In this study, we aimed to investigate the binding properties of Dpo4, a Y-family DNA polymerase from *S. solfataricus*, with both d-DNA and DNA mimic foldamers. We successfully demonstrated the expression and purification of Dpo4. We first crystallized Dpo4 in complex with d-DNA, revealing a crystal structure that resembles a hand gripping a rope. This structure confirms the binding of Dpo4 to d-DNA and provides insights into the key interactions between the polymerase and the DNA substrate. Furthermore, our findings shed light on the role of different domains of Dpo4 in facilitating DNA replication. Subsequently, we attempted to crystallize Dpo4 with DNA mimic foldamers to understand how Dpo4 interacts with these synthetic molecules. Although we encountered challenges in obtaining high-resolution diffraction data, our efforts have opened the door for future studies exploring the interaction between Dpo4 and DNA mimic foldamers.

This study marks an important step in unraveling the versatile binding properties of Dpo4 and its potential applications in DNA replication and repair. Beyond structural insights, functional assays can be conducted to assess the impact of DNA mimic foldamers on the enzymatic activity of Dpo4. Understanding how these synthetic molecules influence polymerase activity can have implications for DNA repair mechanisms. The study opens opportunities for designing novel DNA mimic foldamers with tailored properties to interact with Dpo4 or other DNA-

binding proteins. These molecules can serve as valuable tools for modulating protein-DNA interactions in various biological processes.

Further exploration to achieve higher resolution crystal structure is currently underway towards advancing our understanding of DNA polymerase function and its potential modulation with DNA mimic foldamers.

8. Conclusion and perspective

8.1 Conclusion

Protein-DNA interactions, as established earlier in this thesis, are integral to the proper functioning of cells.[1] Dysregulated protein-DNA interactions have been associated with various diseases.[227] Consequently, investigating protein-DNA interactions, particularly those involving transcription factors as potential therapeutic targets, holds great promise. However, traditional small-molecule drugs face significant challenges targeting these proteins due to their inherent dynamic conformation. In contrast, foldamers, designed to mimic the topology of B-DNA, have demonstrated their ability to disrupt and outcompete DNA binding to proteins like HIV-IN and Top1.[147] The primary objective of the work presented in this thesis was to investigate the structural basis of interactions between DNA mimic foldamers and DNA binding proteins. Understanding the structural basis of interaction between protein-foldamer complexes provides atomic-level details of the foldamers binding to proteins, thus establishing DNA mimic foldamers as specific binders.

The study commenced with Sac7d, chosen as the model DNA-binding protein due to its well-characterized structure with DNA and stability under various conditions.[162] We expressed Sac7d in a bacterial expression system, purified for biophysical and structural analyses, and subsequently first crystallized with an 8-base pair DNA, elucidating the atomic interactions within the Sac7d-DNA complex. This critical step validated the crystallizability of Sac7d and its native fold before progressing to crystallization with DNA mimic foldamers. Racemic DNA mimic foldamers, mimicking 8 and 16-base pair DNA, were designed for biophysical studies to confirm the binding of foldamer to Sac7d. Circular Dichroism (CD) studies confirmed foldamer binding to Sac7d as evidenced by the handedness conversion induced by Sac7d. Competitive experiments with a 10-base pair DNA and the Sac7d-foldamer complex demonstrated at least a ten-fold higher binding affinity of DNA mimic foldamers compared to DNA. This observation was corroborated by Surface Plasmon Resonance (SPR) data, where foldamers exhibited single-digit binding constants in contrast to double-digit constants for the Sac7d-DNA complex. Subsequently, for structural characterization, foldamers with handedness control (using chiral B monomer as stereogenic center) and C_2-symmetry (using aliphatic and aromatic C_2-symmetrical linker) were designed. NMR spectroscopy studies confirmed the binding of foldamers to the DNA binding beta-sheet region of Sac7d. This finding was further corroborated through X-ray crystallography, which revealed a distinctive binding mode on the beta-sheet region of Sac7d in which the foldamers were bound to the DNA binding beta-sheet region but at an approximately 90-degree tilt. It was also found that foldamers can bind to Sac7d

without being kinked, unlike DNA, where Sac7d induces a kink to the DNA strand by intercalation of V26 and M29 amino acid residues. The structure of the Sac7d double mutant (V26A/M29A, which displayed reduced DNA kinking) was solved with the foldamer. Interestingly, the structure of the wild-type Sac7d and the double mutant in complex with foldamers indicated that foldamers could bind to Sac7d without being kinked. This confirmed the potential of DNA mimic foldamers as tools capable of outcompeting DNA and perturbing protein-DNA interactions.

In the second phase of this study, we advanced by tethering DNA mimic foldamers to the surface of Sac7d, incorporating a surface cysteine mutant and an activated disulfide moiety on the foldamer. An adduct between a DNA mimic foldamer and a DNA binding protein was successfully created, providing a novel avenue for further structural elucidation of proteins exhibiting weak or initial binding with foldamers.

In the final research phase, the complexity of the study was increased with an examination of DNA polymerase Dpo4 and the DNA sensing protein hcGAS. Both proteins were expressed and purified for crystallization, with Dpo4 in complex with DNA and apo hcGAS. Although crystallization with chiral DNA mimic foldamers was attempted, poor resolution diffraction data was obtained, preventing the determination of the structure. Nonetheless, ongoing efforts in this direction hold promise for future research.

In summary, this work has significantly contributed to the understanding of the structural aspects of DNA mimic foldamers' interactions with the DNA-binding protein Sac7d. Efforts, discussed in section 8.2, are now being made to target sequences-specific DNA-binding proteins.

8.2 Future perspective

Despite the remarkable outcompetition of DNA by foldamers shown with Sac7d and other protein targets,[147] DNA mimic foldamer foldamers still lack the sequence features necessary for selectively targeting DNA-binding proteins.

To present DNA mimic foldamers as therapeutic modalities or agents for specific protein-DNA interactions, addressing off-target effects is imperative. Structure-based iterative design of foldamers is a promising strategy for conferring specific functionalities to the foldamers. However, challenges arise when foldamers exhibit weak or no binding to the target protein, necessitating the use of a tethering approach. This involves covalently linking the foldamer to the target protein's surface, as described in Chapter 5. Once significant binding is achieved

through structure-based iterative design, the covalent tether can be removed, allowing regulation by the protein-foldamer interaction. Efforts in this direction are extended to targeting HU, a bacterial histone-like protein, by Dr. Tulika Chakrabortty and Manuel Loos (**Fig. 91**).

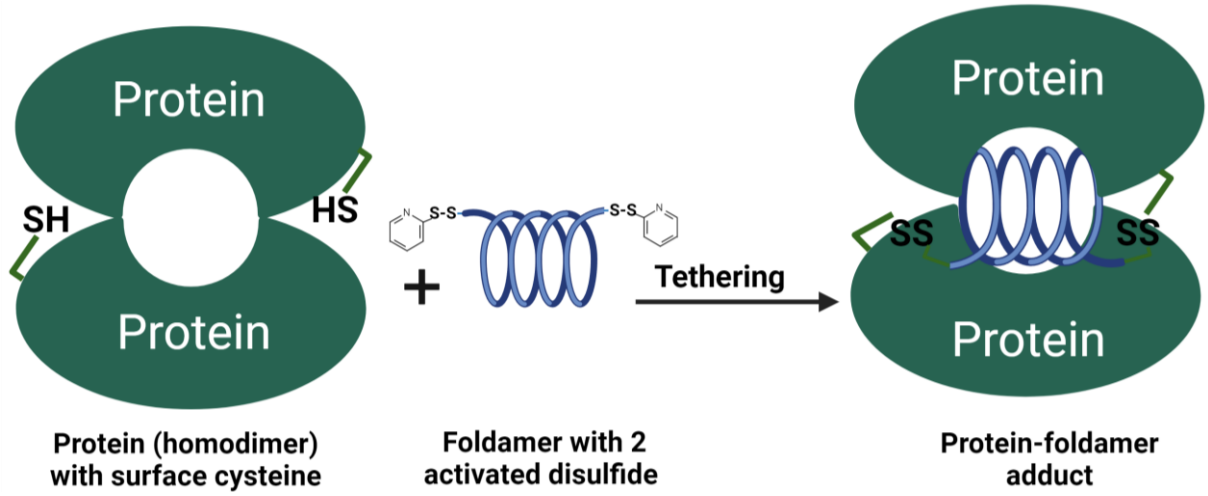


Fig. 91. Sketch of a protein-foldamer adduct with tethering approach.

Another approach to target sequence-specific DNA-binding proteins, such as transcription factors, is to hybridize foldamers with the DNA sequences recognized by these proteins. This strategy eliminates the need for a covalent linker, as the binding is DNA-driven, offering the potential for structural elucidation through X-ray crystallography.

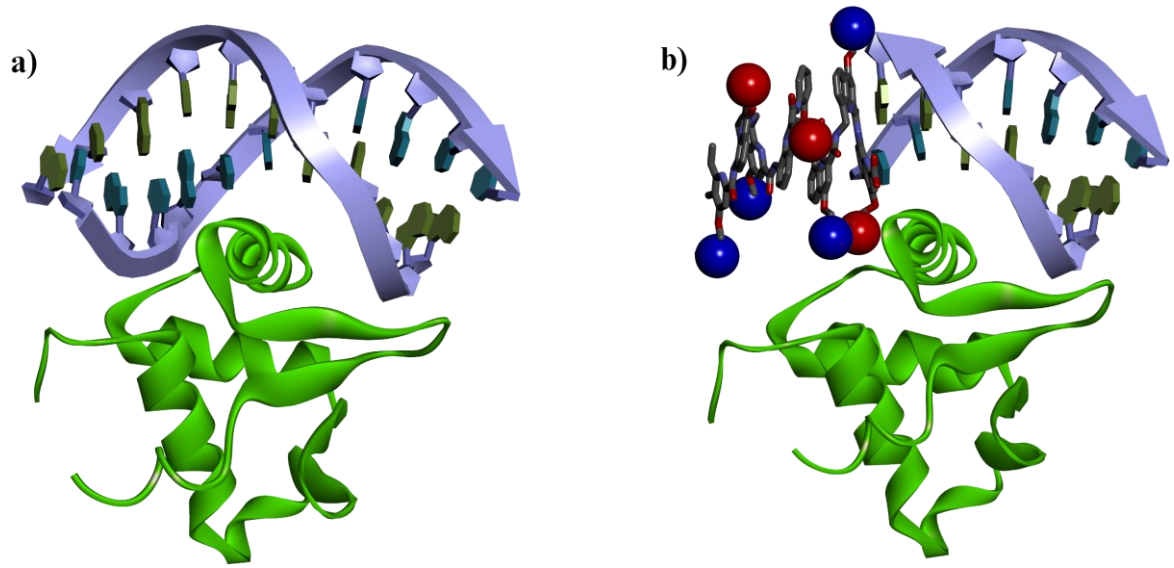


Fig. 92. a) Crystal structure of SAP1-DNA complex (PDB ID 1BC7). b) Simple cartoon representation of the foldamer (shown in sticks and ball representation) with DNA (shown in ladder and ring representation) on SAP1 protein (shown in ribbon representation).

Once the initial structure is obtained, it can guide adjustments to foldamer functionalities based on modeling, followed by the removal of DNA bases and extension of the foldamer monomer toward the DNA-binding site of the protein. However, challenges in achieving foldamer-DNA hybrids at a milligram scale, optimal for crystallization studies, must be addressed. These endeavors are currently underway in our research group by Dr. Tulika Chakrabortty and Manuel Loos, focusing on the development of foldamer-DNA hybrids for SAP1, a transcription factor (**Fig. 92**).

8.3 Ongoing challenges

The three crystal structures of Sac7d-foldamer complexes reported in Chapter 4 remain the only crystal structures of DNA mimic foldamers with protein targets. Consequently, specific challenges arise in the structural solution of protein-foldamer complexes using software such as Phenix and CCP4, primarily designed for proteins or protein-nucleic acids complexes. The generation of restrained libraries for foldamers, a prerequisite for refining protein-foldamer complex structures, is considered impossible in the current versions of this software, especially when foldamer lengths exceed 14mers, equivalent to 7 base pair B-DNA. Collaborative efforts with GlobalPhasing Limited are currently addressing the issue of restraint file generation.

Another challenge relates to the inherent helical nature of DNA mimic foldamers, which leads to a frameshift during molecular replacement, often resulting in the incorrect placement of the foldamer in the structure solution. This challenge becomes more pronounced at resolutions lower than 2.8 Å.

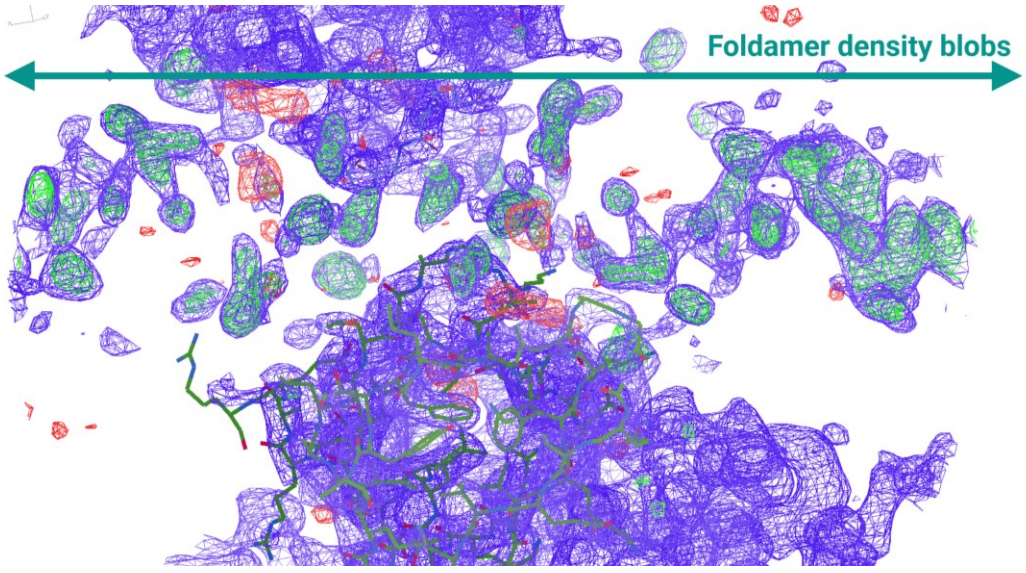


Fig. 93. Electron density map (blue and green) after the molecular replacement of Sac7d alone. Notice the electron density in green, showing the foldamer presence in the crystal structure.

However, at higher resolutions, this issue can be overcome by the presence of clear electron density immediately after molecular replacement of the protein alone (**Fig. 93**). It is important to note that, to date, only three crystal structures of chiral, C_2-symmetrical foldamers in complex with a protein exist. As more structures are determined in the future, the existing structures in this thesis can serve as a starting point for foldamer search models.

A third challenge pertains to the molecular modeling tools used to model DNA mimic foldamers. The structures reported here and in published work have revealed that molecular models often result in a slightly stretched version of DNA mimic foldamers, contrary to the slightly squeezed version observed in crystal structures (**Fig. 94**). Efforts to address this challenge are ongoing in the Huc research group.

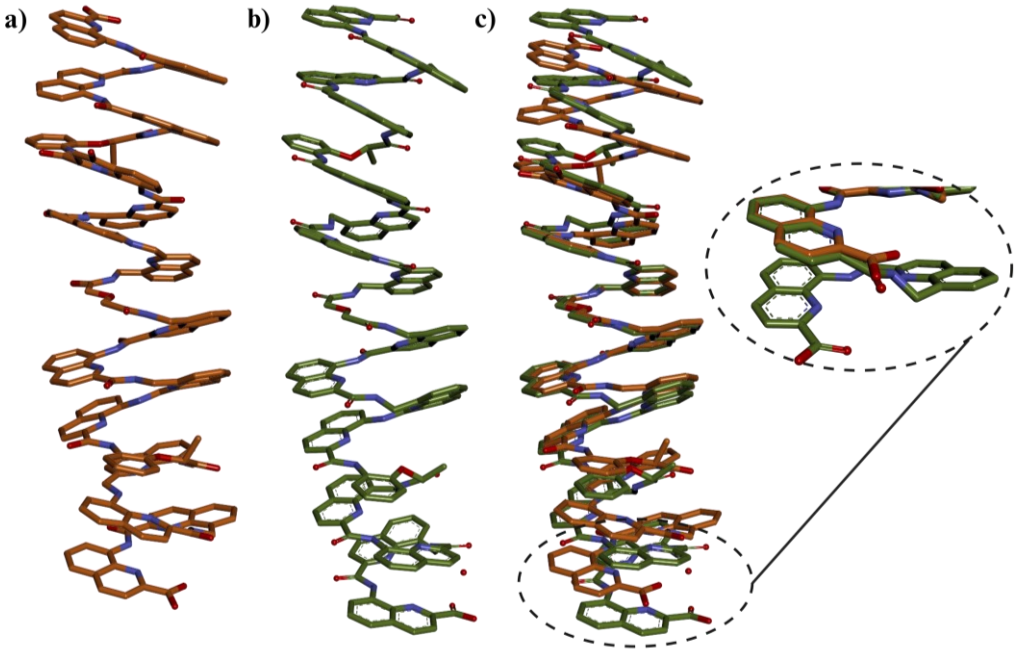


Fig. 94. Comparison of the crystal structure of DNA mimic foldamer versus the molecular model of DNA mimic foldamer. a) Crystal structure of 18mer DNA mimic foldamer (in complex with Sac7d). b) Molecular model of 18mer DNA mimic foldamer. c) Overlay of the a) and b), note the difference in length of the crystallized versus the molecular model in the zoom view on the right.

As the field of DNA mimic foldamer research advances and the current challenges are addressed, we anticipate DNA mimic foldamers targeting sequence-specific DBPs could ultimately serve as tools to interfere with protein-DNA interactions and towards therapeutics.

9. References

[1] M. M. Gromiha, R. Nagarajan, in *Adv. Protein Chem. Struct. Biol.* (Ed.: R. Donev), Academic Press, **2013**, pp. 65–99.

[2] F. Crick, *Nature* **1970**, *227*, 561–563.

[3] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, in *Mol. Biol. Cell 4th Ed.*, Garland Science, **2002**.

[4] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.

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

[6] W. H. Hudson, E. A. Ortlund, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 749–760.

[7] W. Gilbert, B. Müller-Hill, *Proc. Natl. Acad. Sci. U. S. A.* **1966**, *56*, 1891–1898.

[8] C. O. Pabo, L. Nekludova, *J. Mol. Biol.* **2000**, *301*, 597–624.

[9] B. W. Matthews, *Nature* **1988**, *335*, 294–295.

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

[11] P. H. von Hippel, *Science* **1994**, *263*, 769–770.

[12] N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 804–808.

[13] M. Michael Gromiha, J. G. Siebers, S. Selvaraj, H. Kono, A. Sarai, *J. Mol. Biol.* **2004**, *337*, 285–294.

[14] Z. Otwinowski, R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, P. B. Sigler, *Nature* **1988**, *335*, 321–329.

[15] K. Nightingale, S. Dimitrov, R. Reeves, A. P. Wolffe, *EMBO J.* **1996**, *15*, 548–561.

[16] M. Bustin, R. Reeves, in *Prog. Nucleic Acid Res. Mol. Biol.* (Eds.: W.E. Cohn, K. Moldave), Academic Press, **1996**, pp. 35–100b.

[17] U.-M. Ohndorf, M. A. Rould, Q. He, C. O. Pabo, S. J. Lippard, *Nature* **1999**, *399*, 708–712.

[18] Y.-G. Gao, S.-Y. Su, H. Robinson, S. Padmanabhan, L. Lim, B. S. McCrary, S. P. Edmondson, J. W. Shriver, A. H.-J. Wang, *Nat. Struct. Biol.* **1998**, *5*, 782–786.

[19] K. Luger, T. J. Richmond, *Curr. Opin. Struct. Biol.* **1998**, *8*, 33–40.

[20] F. V. Murphy, M. E. Churchill, *Structure* **2000**, *8*, R83–R89.

[21] Z. Zheng, Y. Wang, *Biomol. Concepts* **2011**, *2*, 293–303.

[22] L. J. Beamer, C. O. Pabo, *J. Mol. Biol.* **1992**, *227*, 177–196.

[23] T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **1992**, *71*, 1223–1237.

[24] J. G. McAfee, S. P. Edmondson, P. K. Datta, J. W. Shriver, R. Gupta, *Biochemistry* **1995**, *34*, 10063–10077.

[25] S.-W. Wu, T.-P. Ko, C.-C. Chou, A. H.-J. Wang, *Proteins* **2005**, *60*, 617–628.

[26] M. P. Foster, D. S. Wuttke, I. Radhakrishnan, D. A. Case, J. M. Gottesfeld, P. E. Wright, *Nat. Struct. Biol.* **1997**, *4*, 605–608.

[27] D. M. Ichikawa, O. Abdin, N. Alerasool, M. Kogenaru, A. L. Mueller, H. Wen, D. O. Giganti, G. W. Goldberg, S. Adams, J. M. Spencer, R. Razavi, S. Nim, H. Zheng, C. Gionco, F. T. Clark, A. Strokach, T. R. Hughes, T. Lionnet, M. Taipale, P. M. Kim, M. B. Noyes, *Nat. Biotechnol.* **2023**, 1–13.

[28] Y. Kim, J. H. Geiger, S. Hahn, P. B. Sigler, *Nature* **1993**, *365*, 512–520.

[29] W. S. Somers, S. E. V. Phillips, *Nature* **1992**, *359*, 387–393.

[30] S. E. V. Phillips, *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 671–701.

[31] J. J. Love, X. Li, D. A. Case, K. Giese, R. Grosschedl, P. E. Wright, *Nature* **1995**, *376*, 791–795.

[32] M. McCoy, E. S. Stavridi, J. L. F. Waterman, A. M. Wieczorek, S. J. Opella, T. D. Halazonetis, *EMBO J.* **1997**, *16*, 6230–6236.

[33] M. M. Garner, A. Revzin, *Nucleic Acids Res.* **1981**, *9*, 3047–3060.

[34] A. J. Woo, J. S. Dods, E. Susanto, D. Ulgiati, L. J. Abraham, *Mol. Cell. Proteomics* **2002**, *1*, 472–478.

[35] J. A. Stead, J. N. Keen, K. J. McDowall, *Mol. Cell. Proteomics MCP* **2006**, *5*, 1697–1702.

[36] D. J. Galas, A. Schmitz, *Nucleic Acids Res.* **1978**, *5*, 3157–3170.

[37] D. S. Gilmour, J. T. Lis, *Proc. Natl. Acad. Sci.* **1984**, *81*, 4275–4279.

[38] D. S. Gilmour, J. T. Lis, *Mol. Cell. Biol.* **1985**, *5*, 2009–2018.

[39] J. A. Dahl, P. Collas, *Stem Cells* **2007**, *25*, 1037–1046.

[40] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, D. C. Phillips, *Nature* **1958**, *181*, 662–666.

[41] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, *112*, 535–542.

[42] K. Kurpiewska, T. Borowski, *Acta Biochim. Pol.* **2021**, *68*, 535–546.

[43] N. Asherie, *Methods San Diego Calif* **2004**, *34*, 266–272.

[44] I. Russo Krauss, A. Merlino, A. Vergara, F. Sica, *Int. J. Mol. Sci.* **2013**, *14*, 11643–11691.

[45] M. Benvenuti, S. Mangani, *Nat. Protoc.* **2007**, *2*, 1633–1651.

[46] W. H. Bragg, W. L. Bragg, *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1997**, *88*, 428–438.

[47] H. R. Powell, *Biosci. Rep.* **2017**, *37*, BSR20170227.

[48] G. Taylor, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 1881–1890.

[49] M. G. Rossmann, *Acta Crystallogr. D Biol. Crystallogr.* **2001**, *57*, 1360–1366.

[50] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.

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

[52] T. G. Flower, J. H. Hurley, *Protein Sci. Publ. Protein Soc.* **2021**, *30*, 728–734.

[53] A. Ilari, C. Savino, *Methods Mol. Biol. Clifton NJ* **2008**, *452*, 63–87.

[54] R. G. Brennan, B. W. Matthews, *Trends Biochem. Sci.* **1989**, *14*, 286–290.

[55] S. Campagne, V. Gervais, A. Milon, *J. R. Soc. Interface* **2011**, *8*, 1065–1078.

[56] J. Majka, C. Speck, *Adv. Biochem. Eng. Biotechnol.* **2007**, *104*, 13–36.

[57] S. Blouin, T. D. Craggs, D. A. Lafontaine, J. C. Penedo, *Methods Mol. Biol. Clifton NJ* **2009**, *543*, 475–502.

[58] E. Stedman, E. Stedman, *Nature* **1950**, *166*, 780–781.

[59] Y. Ofran, V. Mysore, B. Rost, *Bioinforma. Oxf. Engl.* **2007**, *23*, i347-353.

[60] G. A. Maston, S. K. Evans, M. R. Green, *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 29–59.

[61] S. A. Lambert, A. Jolma, L. F. Campitelli, P. K. Das, Y. Yin, M. Albu, X. Chen, J. Taipale, T. R. Hughes, M. T. Weirauch, *Cell* **2018**, *172*, 650–665.

[62] J. E. Darnell, *Nat. Rev. Cancer* **2002**, *2*, 740–749.

[63] E. R. Kastenhuber, S. W. Lowe, *Cell* **2017**, *170*, 1062–1078.

[64] S. J. O'Keefe, J. Tamura, R. L. Kincaid, M. J. Tocci, E. A. O'Neill, *Nature* **1992**, *357*, 692–694.

[65] A. W. Thomson, P. B. Carroll, J. McCauley, J. Woo, K. Abu-Elmagd, T. E. Starzl, D. H. Van Thiel, *Springer Semin. Immunopathol.* **1993**, *14*, 323–344.

[66] S. Todo, N. Murase, Y. Ueda, L. Podesta, P. ChapChap, D. Kahn, K. Okuda, O. Imventarza, A. Casavilla, J. Demetris, L. Makowka, T. E. Starzl, *Transplant. Proc.* **1988**, *20*, 215–219.

[67] B. Seto, *Clin. Transl. Med.* **2012**, *1*, 29.

[68] S. Mukherjee, U. Mukherjee, *J. Transplant.* **2009**, *2009*, e701464.

[69] E. Raymond, S. Faivre, J. P. Armand, *Drugs* **2000**, *60*, 15–23.

[70] U. Hollstein, *Chem. Rev.* **1974**, *74*, 625–652.

[71] D. A. Bushnell, P. Cramer, R. D. Kornberg, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 1218–1222.

[72] J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, A. K. Dunker, *Biochemistry* **2006**, *45*, 6873–6888.

[73] J. H. Bushweller, *Nat. Rev. Cancer* **2019**, *19*, 611–624.

[74] J. L. Yap, J. Chauhan, K.-Y. Jung, L. Chen, E. V. Prochownik, S. Fletcher, *MedChemComm* **2012**, *3*, 541–551.

[75] D. A. Scott, C. H. Best, *Ind. Eng. Chem.* **1925**, *17*, 238–240.

[76] J. L. Lau, M. K. Dunn, *Bioorg. Med. Chem.* **2018**, *26*, 2700–2707.

[77] M. Alas, A. Saghaeidehkordi, K. Kaur, *J. Med. Chem.* **2021**, *64*, 216–232.

[78] J. Park, Y. Choi, H. Chang, W. Um, J. H. Ryu, I. C. Kwon, *Theranostics* **2019**, *9*, 8073–8090.

[79] Z. Jiang, J. Guan, J. Qian, C. Zhan, *Biomater. Sci.* **2019**, *7*, 461–471.

[80] C. H. Chau, P. S. Steeg, W. D. Figg, *Lancet Lond. Engl.* **2019**, *394*, 793–804.

[81] A. Thomas, B. A. Teicher, R. Hassan, *Lancet Oncol.* **2016**, *17*, e254–e262.

[82] R. He, B. Finan, J. P. Mayer, R. D. DiMarchi, *Mol. Basel Switz.* **2019**, *24*, 1855.

[83] E. I. Vrettos, G. Mező, A. G. Tzakos, *Beilstein J. Org. Chem.* **2018**, *14*, 930–954.

[84] S. N. Jones, A. E. Roe, L. A. Donehower, A. Bradley, *Nature* **1995**, *378*, 206–208.

[85] M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, D. I. Svergun, M. Blackledge, A. R. Fersht, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5762–5767.

[86] Y. S. Chang, B. Graves, V. Guerlavais, C. Tovar, K. Packman, K.-H. To, K. A. Olson, K. Kesavan, P. Gangurde, A. Mukherjee, T. Baker, K. Darlak, C. Elkin, Z. Filipovic, F. Z. Qureshi, H. Cai, P. Berry, E. Feyfant, X. E. Shi, J. Horstick, D. A. Annis, A. M. Manning, N. Fotouhi, H. Nash, L. T. Vassilev, T. K. Sawyer, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E3445-3454.

[87] N. Estrada-Ortiz, C. G. Neochoritis, A. Dömling, *ChemMedChem* **2016**, *11*, 757–772.

[88] G. Bhardwaj, J. O'Connor, S. Rettie, Y.-H. Huang, T. A. Ramelot, V. K. Mulligan, G. G. Alpkilic, J. Palmer, A. K. Bera, M. J. Bick, M. Di Piazza, X. Li, P. Hosseinzadeh, T. W. Craven, R. Tejero, A. Lauko, R. Choi, C. Glynn, L. Dong, R. Griffin, W. C. van Voorhis, J. Rodriguez, L. Stewart, G. T. Montelione, D. Craik, D. Baker, *Cell* **2022**, *185*, 3520-3532.e26.

[89] S. Dengler, R. T. Howard, V. Morozov, C. Tsiamantas, Z. Liu, C. Dobrzanski, V. Pophristic, S. Brameyer, C. Douat, H. Suga, I. Huc, **2022**, DOI 10.26434/chemrxiv-2022-5rlmv.

[90] A. A. Vinogradov, Y. Yin, H. Suga, *J. Am. Chem. Soc.* **2019**, *141*, 4167–4181.

[91] P. P. Chamberlain, A. Lopez-Girona, K. Miller, G. Carmel, B. Pagarigan, B. Chie-Leon, E. Rychak, L. G. Corral, Y. J. Ren, M. Wang, M. Riley, S. L. Delker, T. Ito, H. Ando, T.

Mori, Y. Hirano, H. Handa, T. Hakoshima, T. O. Daniel, B. E. Cathers, *Nat. Struct. Mol. Biol.* **2014**, *21*, 803–809.

[92] G. Lu, R. E. Middleton, H. Sun, M. Naniong, C. J. Ott, C. S. Mitsiades, K.-K. Wong, J. E. Bradner, W. G. Kaelin, *Science* **2014**, *343*, 305–309.

[93] J. Krönke, N. D. Udeshi, A. Narla, P. Grauman, S. N. Hurst, M. McConkey, T. Svinkina, D. Heckl, E. Comer, X. Li, C. Ciarlo, E. Hartman, N. Munshi, M. Schenone, S. L. Schreiber, S. A. Carr, B. L. Ebert, *Science* **2014**, *343*, 301–305.

[94] T. Ito, H. Ando, T. Suzuki, T. Ogura, K. Hotta, Y. Imamura, Y. Yamaguchi, H. Handa, *Science* **2010**, *327*, 1345–1350.

[95] Y. Isobe, M. Okumura, L. M. McGregor, S. M. Brittain, M. D. Jones, X. Liang, R. White, W. Forrester, J. M. McKenna, J. A. Tallarico, M. Schirle, T. J. Maimone, D. K. Nomura, *Nat. Chem. Biol.* **2020**, *16*, 1189–1198.

[96] K. M. Sakamoto, K. B. Kim, A. Kumagai, F. Mercurio, C. M. Crews, R. J. Deshaies, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 8554–8559.

[97] S.-L. Paiva, C. M. Crews, *Curr. Opin. Chem. Biol.* **2019**, *50*, 111–119.

[98] L. N. Gechijian, D. L. Buckley, M. A. Lawlor, J. M. Reyes, J. Paulk, C. J. Ott, G. E. Winter, M. A. Erb, T. G. Scott, M. Xu, H.-S. Seo, S. Dhe-Paganon, N. P. Kwiatkowski, J. A. Perry, J. Qi, N. S. Gray, J. E. Bradner, *Nat. Chem. Biol.* **2018**, *14*, 405–412.

[99] Z. I. Bassi, M. C. Fillmore, A. H. Miah, T. D. Chapman, C. Maller, E. J. Roberts, L. C. Davis, D. E. Lewis, N. W. Galwey, K. E. Waddington, V. Parravicini, A. L. Macmillan-Jones, C. Gongora, P. G. Humphreys, I. Churcher, R. K. Prinjha, D. F. Tough, *ACS Chem. Biol.* **2018**, *13*, 2862–2867.

[100] P. M. Cromm, K. T. G. Samarasinghe, J. Hines, C. M. Crews, *J. Am. Chem. Soc.* **2018**, *140*, 17019–17026.

[101] S. L. Degorce, O. Tavana, E. Banks, C. Crafter, L. Gingipalli, D. Kouvchinov, Y. Mao, F. Pachl, A. Solanki, V. Valge-Archer, B. Yang, S. D. Edmondson, *J. Med. Chem.* **2020**, *63*, 10460–10473.

[102] M. He, C. Cao, Z. Ni, Y. Liu, P. Song, S. Hao, Y. He, X. Sun, Y. Rao, *Signal Transduct. Target. Ther.* **2022**, *7*, 1–64.

[103] A. Mullard, *Nat. Rev. Drug Discov.* **2019**, *18*, 237–239.

[104] M. S. Weinberg, K. V. Morris, *Nucleic Acids Res.* **2016**, *44*, 6505–6517.

[105] C. H. Arrowsmith, J. E. Audia, C. Austin, J. Baell, J. Bennett, J. Blagg, C. Bountra, P. E. Brennan, P. J. Brown, M. E. Bunnage, C. Buser-Doepner, R. M. Campbell, A. J. Carter, P. Cohen, R. A. Copeland, B. Cravatt, J. L. Dahlin, D. Dhanak, A. M. Edwards, M.

Frederiksen, S. V. Frye, N. Gray, C. E. Grimshaw, D. Hepworth, T. Howe, K. V. M. Huber, J. Jin, S. Knapp, J. D. Kotz, R. G. Kruger, D. Lowe, M. M. Mader, B. Marsden, A. Mueller-Fahrnow, S. Müller, R. C. O'Hagan, J. P. Overington, D. R. Owen, S. H. Rosenberg, R. Ross, B. Roth, M. Schapira, S. L. Schreiber, B. Shoichet, M. Sundström, G. Superti-Furga, J. Taunton, L. Toledo-Sherman, C. Walpole, M. A. Walters, T. M. Willson, P. Workman, R. N. Young, W. J. Zuercher, *Nat. Chem. Biol.* **2015**, *11*, 536–541.

[106] D. E. Wemmer, P. B. Dervan, *Curr. Opin. Struct. Biol.* **1997**, *7*, 355–361.

[107] T. Zhou, J. Schmidt, Y. Kim, A. R. MacLeod, *Blood* **2017**, *130*, 3078.

[108] D. T. F. Dryden, *Trends Biotechnol.* **2006**, *24*, 378–382.

[109] C. D. Putnam, J. A. Tainer, *DNA Repair* **2005**, *4*, 1410–1420.

[110] H.-C. Wang, C.-H. Ho, K.-C. Hsu, J.-M. Yang, A. H.-J. Wang, *Biochemistry* **2014**, *53*, 2865–2874.

[111] J. Shin, F. Jiang, J.-J. Liu, N. L. Bray, B. J. Rauch, S. H. Baik, E. Nogales, J. Bondy-Denomy, J. E. Corn, J. A. Doudna, *Sci. Adv.* **2017**, *3*, e1701620.

[112] S. Chowdhury, J. Carter, M. F. Rollins, S. M. Golden, R. N. Jackson, C. Hoffmann, L. Nosaka, J. Bondy-Denomy, K. L. Maxwell, A. R. Davidson, E. R. Fischer, G. C. Lander, B. Wiedenheft, *Cell* **2017**, *169*, 47-57.e11.

[113] D. Dong, M. Guo, S. Wang, Y. Zhu, S. Wang, Z. Xiong, J. Yang, Z. Xu, Z. Huang, *Nature* **2017**, *546*, 436–439.

[114] G. Zhang, W. Wang, A. Deng, Z. Sun, Y. Zhang, Y. Liang, Y. Che, T. Wen, *PLoS Genet.* **2012**, *8*, e1002987.

[115] E. Bochkareva, L. Kaustov, A. Ayed, G.-S. Yi, Y. Lu, A. Pineda-Lucena, J. C. C. Liao, A. L. Okorokov, J. Milner, C. H. Arrowsmith, A. Bochkarev, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15412–15417.

[116] H.-C. Wang, C.-C. Chou, K.-C. Hsu, C.-H. Lee, A. H.-J. Wang, *IUBMB Life* **2019**, *71*, 539–548.

[117] D. Yüksel, P. R. Bianco, K. Kumar, *Mol. Biosyst.* **2015**, *12*, 169–177.

[118] Z. Chen, S. E. Boyken, M. Jia, F. Busch, D. Flores-Solis, M. J. Bick, P. Lu, Z. L. VanAernum, A. Sahasrabuddhe, R. A. Langan, S. Bermeo, T. J. Brunette, V. K. Mulligan, L. P. Carter, F. DiMaio, N. G. Sgourakis, V. H. Wysocki, D. Baker, *Nature* **2019**, *565*, 106–111.

[119] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497–1500.

[120] C. G. Janson, M. J. During, *Peptide Nucleic Acids, Morpholinos and Related Antisense Biomolecules*, Springer US, Boston, MA, **2006**.

[121] P. E. Nielsen, G. Haaima, *Chem. Soc. Rev.* **1997**, *26*, 73–78.

[122] M. Robaczewska, R. Narayan, B. Seigneres, O. Schorr, A. Thermet, A. J. Podhajska, C. Trepo, F. Zoulim, P. E. Nielsen, L. Cova, *J. Hepatol.* **2005**, *42*, 180–187.

[123] S. Montazersaheb, M. S. Hejazi, H. Nozad Charoudeh, *Adv. Pharm. Bull.* **2018**, *8*, 551–563.

[124] R. Patel, S. Sarma, A. Shukla, P. Parmar, D. Goswami, M. Saraf, *Mol. Biol. Rep.* **2020**, *47*, 8113–8131.

[125] P. Denichenko, M. Mogilevsky, A. Cléry, T. Welte, J. Biran, O. Shimshon, G. D. Barnabas, M. Danan-Gotthold, S. Kumar, E. Yavin, E. Y. Levanon, F. H. Allain, T. Geiger, G. Levkowitz, R. Karni, *Nat. Commun.* **2019**, *10*, 1590.

[126] I. Souissi, P. Ladam, J. A. Cognet, S. Le Coquil, N. Varin-Blank, F. Baran-Marszak, V. Metelev, R. Fagard, *Mol. Cancer* **2012**, *11*, 12.

[127] D. Hnedzko, D. W. McGee, E. Rozners, *Bioorg. Med. Chem.* **2016**, *24*, 4199–4205.

[128] F. Yang, B. Dong, K. Nie, H. Shi, Y. Wu, H. Wang, Z. Liu, *ACS Comb. Sci.* **2015**, *17*, 608–614.

[129] R. P. Singh, B.-K. Oh, J.-W. Choi, *Bioelectrochemistry Amst. Neth.* **2010**, *79*, 153–161.

[130] M. V. Sonar, M. E. Wampole, Y.-Y. Jin, C.-P. Chen, M. L. Thakur, E. Wickstrom, *Bioconjug. Chem.* **2014**, *25*, 1697–1708.

[131] D. Seebach, M. Overhand, F. N. M. Kühnle, B. Martinoni, L. Oberer, U. Hommel, H. Widmer, *Helv. Chim. Acta* **1996**, *79*, 913–941.

[132] I. Saraogi, A. D. Hamilton, *Chem. Soc. Rev.* **2009**, *38*, 1726–1743.

[133] Y. Hamuro, S. J. Geib, A. D. Hamilton, *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 446–448.

[134] C. G. Pappas, P. K. Mandal, B. Liu, B. Kauffmann, X. Miao, D. Komáromy, W. Hoffmann, C. Manz, R. Chang, K. Liu, K. Pagel, I. Huc, S. Otto, *Nat. Chem.* **2020**, *12*, 1180–1186.

[135] R. Gopalakrishnan, A. I. Frolov, L. Knerr, W. J. I. Drury, E. Valeur, *J. Med. Chem.* **2016**, *59*, 9599–9621.

[136] D. Bindl, Folding Properties, Handedness Control and Aggregation Behavior of Helical Aromatic Amino Acid Foldamers in Water, Ludwig-Maximilians-Universität München, **2022**.

[137] G. Guichard, I. Huc, *Chem. Commun.* **2011**, *47*, 5933–5941.

[138] R. J. Simon, R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 9367–9371.

[139] F. S. Menke, B. Wicher, L. Allmendinger, V. Maurizot, I. Huc, *Chem. Sci.* **2023**, *14*, 3742–3751.

[140] E. J. Petersson, C. J. Craig, D. S. Daniels, J. X. Qiu, A. Schepartz, *J. Am. Chem. Soc.* **2007**, *129*, 5344–5345.

[141] Y. Zhang, G. Sicot, X. Cui, M. Vogel, C. A. Wuertzer, K. Lezon-Geyda, J. Wheeler, D. A. Harki, K. A. Muzikar, D. A. Stolper, P. B. Dervan, A. S. Perkins, *Biochemistry* **2011**, *50*, 10431–10441.

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

[143] S. Asamitsu, Y. Kawamoto, F. Hashiya, K. Hashiya, M. Yamamoto, S. Kizaki, T. Bando, H. Sugiyama, *Bioorg. Med. Chem.* **2014**, *22*, 4646–4657.

[144] T. A. Martinek, F. Fülöp, *Chem. Soc. Rev.* **2012**, *41*, 687–702.

[145] I. M. Mándity, F. Fülöp, *Expert Opin. Drug Discov.* **2015**, *10*, 1163–1177.

[146] S. M. Miller, R. J. Simon, S. Ng, R. N. Zuckermann, J. M. Kerr, W. H. Moos, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 2657–2662.

[147] K. Ziach, C. Chollet, V. Parissi, P. Prabhakaran, M. Marchivie, V. Corvaglia, P. P. Bose, K. Laxmi-Reddy, F. Godde, J.-M. Schmitter, S. Chaignepain, P. Pourquier, I. Huc, *Nat. Chem.* **2018**, *10*, 511–518.

[148] H. Jiang, J.-M. Léger, I. Huc, *J. Am. Chem. Soc.* **2003**, *125*, 3448–3449.

[149] C. Dolain, A. Grélard, M. Laguerre, H. Jiang, V. Maurizot, I. Huc, *Chem. Weinh. Bergstr. Ger.* **2005**, *11*, 6135–6144.

[150] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, *389*, 251–260.

[151] S. C. Dillon, C. J. Dorman, *Nat. Rev. Microbiol.* **2010**, *8*, 185–195.

[152] G. Behar, M. Bellinzoni, M. Maillasson, L. Paillard-Laurance, P. M. Alzari, X. He, B. Mouratou, F. Pecorari, *Protein Eng. Des. Sel.* **2013**, *26*, 267–275.

[153] S. Pacheco, G. Béhar, M. Maillasson, B. Mouratou, F. Pecorari, *Protein Eng. Des. Sel.* **2014**, *27*, 431–438.

[154] F. H. C. Crick, A. Klug, *Nature* **1975**, *255*, 530–533.

[155] H.-S. Koo, H.-M. Wu, D. M. Crothers, *Nature* **1986**, *320*, 501–506.

[156] C.-Y. Chen, *Nucleic Acids Res.* **2005**, *33*, 430–438.

[157] Y. H. Hsiang, M. G. Lihou, L. F. Liu, *Cancer Res.* **1989**, *49*, 5077–5082.

[158] J. Y. Charcosset, S. Soues, F. Laval, *Bull. Cancer (Paris)* **1993**, *80*, 923–954.

[159] S. Hare, S. S. Gupta, E. Valkov, A. Engelman, P. Cherepanov, *Nature* **2010**, *464*, 232–236.

[160] S. P. Edmondson, L. Qiu, J. W. Shriver, *Biochemistry* **1995**, *34*, 13289–13304.

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

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

[163] T.-P. Ko, H.-M. Chu, C.-Y. Chen, C.-C. Chou, A. H.-J. Wang, *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 1381–1387.

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

[165] C.-Y. Chen, T.-P. Ko, T.-W. Lin, C.-C. Chou, C.-J. Chen, A. H.-J. Wang, *Nucleic Acids Res.* **2005**, *33*, 430–438.

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

[167] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.

[168] A. Vagin, A. Lebedev, IUCr, "MoRDa, an automatic molecular replacement pipeline," **2015**.

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

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

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

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

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

[174] A. A. McCarthy, R. Barrett, A. Beteva, H. Caserotto, F. Dobias, F. Felisaz, T. Giraud, M. Guijarro, R. Janocha, A. Khadrouche, M. Lentini, G. A. Leonard, M. Lopez Marrero, S. Malbet-Monaco, S. McSweeney, D. Nurizzo, G. Papp, C. Rossi, J. Sinoir, C. Sorez, J.

Surr, O. Svensson, U. Zander, F. Cipriani, P. Theveneau, C. Mueller-Dieckmann, *J. Synchrotron Radiat.* **2018**, *25*, 1249–1260.

[175] T. Berggård, S. Linse, P. James, *Proteomics* **2007**, *7*, 2833–2842.

[176] W. E. Stites, *Chem. Rev.* **1997**, *97*, 1233–1250.

[177] A. Miyanaga, R. Ouchi, F. Ishikawa, E. Goto, G. Tanabe, F. Kudo, T. Eguchi, *J. Am. Chem. Soc.* **2018**, *140*, 7970–7978.

[178] C. Faoro, L. Wilkinson-White, A. H. Kwan, S. F. Ataide, *PloS One* **2018**, *13*, e0200387.

[179] C. O. Pabo, R. T. Sauer, *Annu. Rev. Biochem.* **1984**, *53*, 293–321.

[180] P. J. Mitchell, R. Tjian, *Science* **1989**, *245*, 371–378.

[181] N. Cao, K. Tan, T. Annamalai, A. Joachimiak, Y.-C. Tse-Dinh, *Nucleic Acids Res.* **2018**, *46*, 7296–7308.

[182] T.-F. Lian, Y.-P. Xu, L.-F. Li, X.-D. Su, *Cell Rep.* **2017**, *19*, 1334–1342.

[183] D.-W. Zhang, X. Zhao, J.-L. Hou, Z.-T. Li, *Chem. Rev.* **2012**, *112*, 5271–5316.

[184] I. Huc, *Eur. J. Org. Chem.* **2004**, *2004*, 7–7.

[185] N. Delsuc, S. Massip, J.-M. Léger, B. Kauffmann, I. Huc, *J. Am. Chem. Soc.* **2011**, *133*, 3165–3172.

[186] S. De, B. Chi, T. Granier, T. Qi, V. Maurizot, I. Huc, *Nat. Chem.* **2018**, *10*, 51–57.

[187] T. A. Edwards, A. J. Wilson, *Amino Acids* **2011**, *41*, 743–754.

[188] M. K. P. Jayatunga, S. Thompson, A. D. Hamilton, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 717–724.

[189] M. Vallade, M. Jewginski, L. Fischer, J. Buratto, K. Bathany, J.-M. Schmitter, M. Stupfel, F. Godde, C. D. Mackereth, I. Huc, *Bioconjug. Chem.* **2019**, *30*, 54–62.

[190] J. Sheng, Z. Huang, *Int. J. Mol. Sci.* **2008**, *9*, 258–271.

[191] W. A. Hendrickson, *Science* **1991**, *254*, 51–58.

[192] W. A. Hendrickson, J. R. Horton, D. M. LeMaster, *EMBO J.* **1990**, *9*, 1665–1672.

[193] W. A. Hendrickson, *Trends Biochem. Sci.* **2000**, *25*, 637–643.

[194] S. R. Holbrook, *Curr. Opin. Struct. Biol.* **2005**, *15*, 302–308.

[195] R. Medzhitov, C. Janeway, *Immunol. Rev.* **2000**, *173*, 89–97.

[196] A. Decout, J. D. Katz, S. Venkatraman, A. Ablasser, *Nat. Rev. Immunol.* **2021**, *21*, 548–569.

[197] H. Ishikawa, Z. Ma, G. N. Barber, *Nature* **2009**, *461*, 788–792.

[198] A. Ablasser, M. Goldeck, T. Cavlar, T. Deimling, G. Witte, I. Röhl, K.-P. Hopfner, J. Ludwig, V. Hornung, *Nature* **2013**, *498*, 380–384.

[199] E. J. Diner, D. L. Burdette, S. C. Wilson, K. M. Monroe, C. A. Kellenberger, M. Hyodo, Y. Hayakawa, M. C. Hammond, R. E. Vance, *Cell Rep.* **2013**, *3*, 1355–1361.

[200] A. Ablasser, S. Hur, *Nat. Immunol.* **2020**, *21*, 17–29.

[201] L. Andreeva, B. Hiller, D. Kostrewa, C. Lässig, C. C. de Oliveira Mann, D. Jan Drexler, A. Maiser, M. Gaidt, H. Leonhardt, V. Hornung, K.-P. Hopfner, *Nature* **2017**, *549*, 394–398.

[202] H. Ishikawa, G. N. Barber, *Nature* **2008**, *455*, 674–678.

[203] K.-P. Hopfner, V. Hornung, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 501–521.

[204] D. Gao, T. Li, X.-D. Li, X. Chen, Q.-Z. Li, M. Wight-Carter, Z. J. Chen, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E5699-5705.

[205] E. E. Gray, P. M. Treuting, J. J. Woodward, D. B. Stetson, *J. Immunol. Baltim. Md 1950* **2015**, *195*, 1939–1943.

[206] F. Civril, T. Deimling, C. C. de Oliveira Mann, A. Ablasser, M. Moldt, G. Witte, V. Hornung, K.-P. Hopfner, *Nature* **2013**, *498*, 332–337.

[207] P. Gao, M. Ascano, Y. Wu, W. Barchet, B. L. Gaffney, T. Zillinger, A. A. Serganov, Y. Liu, R. A. Jones, G. Hartmann, T. Tuschl, D. J. Patel, *Cell* **2013**, *153*, 1094–1107.

[208] V. Kleene, V. Corvaglia, E. Chacin, I. Forne, D. B. Konrad, P. Khosravani, C. Douat, C. F. Kurat, I. Huc, A. Imhof, *Nucleic Acids Res.* **2023**, gkad681.

[209] D. Deepak, V. Corvaglia, J. Wu, L. Allmendinger, I. Huc, **2023**, DOI 10.26434/chemrxiv-2023-9wnb8.

[210] V. Corvaglia, J. Wu, D. Deepak, M. Loos, I. Huc, **2023**, DOI 10.26434/chemrxiv-2023-fz84h.

[211] L. Lama, C. Adura, W. Xie, D. Tomita, T. Kamei, V. Kuryavyi, T. Gogakos, J. I. Steinberg, M. Miller, L. Ramos-Espiritu, Y. Asano, S. Hashizume, J. Aida, T. Imaeda, R. Okamoto, A. J. Jennings, M. Michino, T. Kuroita, A. Stamford, P. Gao, P. Meinke, J. F. Glickman, D. J. Patel, T. Tuschl, *Nat. Commun.* **2019**, *10*, 2261.

[212] W. Zhou, A. T. Whiteley, C. C. de Oliveira Mann, B. R. Morehouse, R. P. Nowak, E. S. Fischer, N. S. Gray, J. J. Mekalanos, P. J. Kranzusch, *Cell* **2018**, *174*, 300-311.e11.

[213] M. Cianci, G. Bourenkov, G. Pompidor, I. Karpics, J. Kallio, I. Bento, M. Roessle, F. Cipriani, S. Fiedler, T. R. Schneider, *J. Synchrotron Radiat.* **2017**, *24*, 323–332.

[214] C. Vonrhein, C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack, G. Bricogne, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 293–302.

[215] A. Falaschi, A. Kornberg, *J. Biol. Chem.* **1966**, *241*, 1478–1482.

[216] I. R. Lehman, M. J. Bessman, E. S. Simms, A. Kornberg, *J. Biol. Chem.* **1958**, *233*, 163–170.

[217] T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529.

[218] M. F. Goodman, B. Tippin, *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 101–109.

[219] E. C. Friedberg, V. L. Gerlach, *Cell* **1999**, *98*, 413–416.

[220] H. Ohmori, E. C. Friedberg, R. P. P. Fuchs, M. F. Goodman, F. Hanaoka, D. Hinkle, T. A. Kunkel, C. W. Lawrence, Z. Livneh, T. Nohmi, L. Prakash, S. Prakash, T. Todo, G. C. Walker, Z. Wang, R. Woodgate, *Mol. Cell* **2001**, *8*, 7–8.

[221] R. E. Johnson, M. T. Washington, S. Prakash, L. Prakash, *J. Biol. Chem.* **2000**, *275*, 7447–7450.

[222] B.-L. Zhou, J. D. Pata, T. A. Steitz, *Mol. Cell* **2001**, *8*, 427–437.

[223] F. Boudsocq, S. Iwai, F. Hanaoka, R. Woodgate, *Nucleic Acids Res.* **2001**, *29*, 4607–4616.

[224] H. Ling, F. Boudsocq, R. Woodgate, W. Yang, *Cell* **2001**, *107*, 91–102.

[225] J. An, J. Choi, D. Hwang, J. Park, C. W. Pemble, T. H. M. Duong, K.-R. Kim, H. Ahn, H. S. Chung, D.-R. Ahn, *Chem. Commun.* **2020**, *56*, 2186–2189.

[226] K. A. Fiala, Z. Suo, *Biochemistry* **2004**, *43*, 2106–2115.

[227] J. S. Jiménez, *J. Alzheimers Dis. JAD* **2010**, *22*, 375–391.

10. Appendix

10.1 Supplementary figures

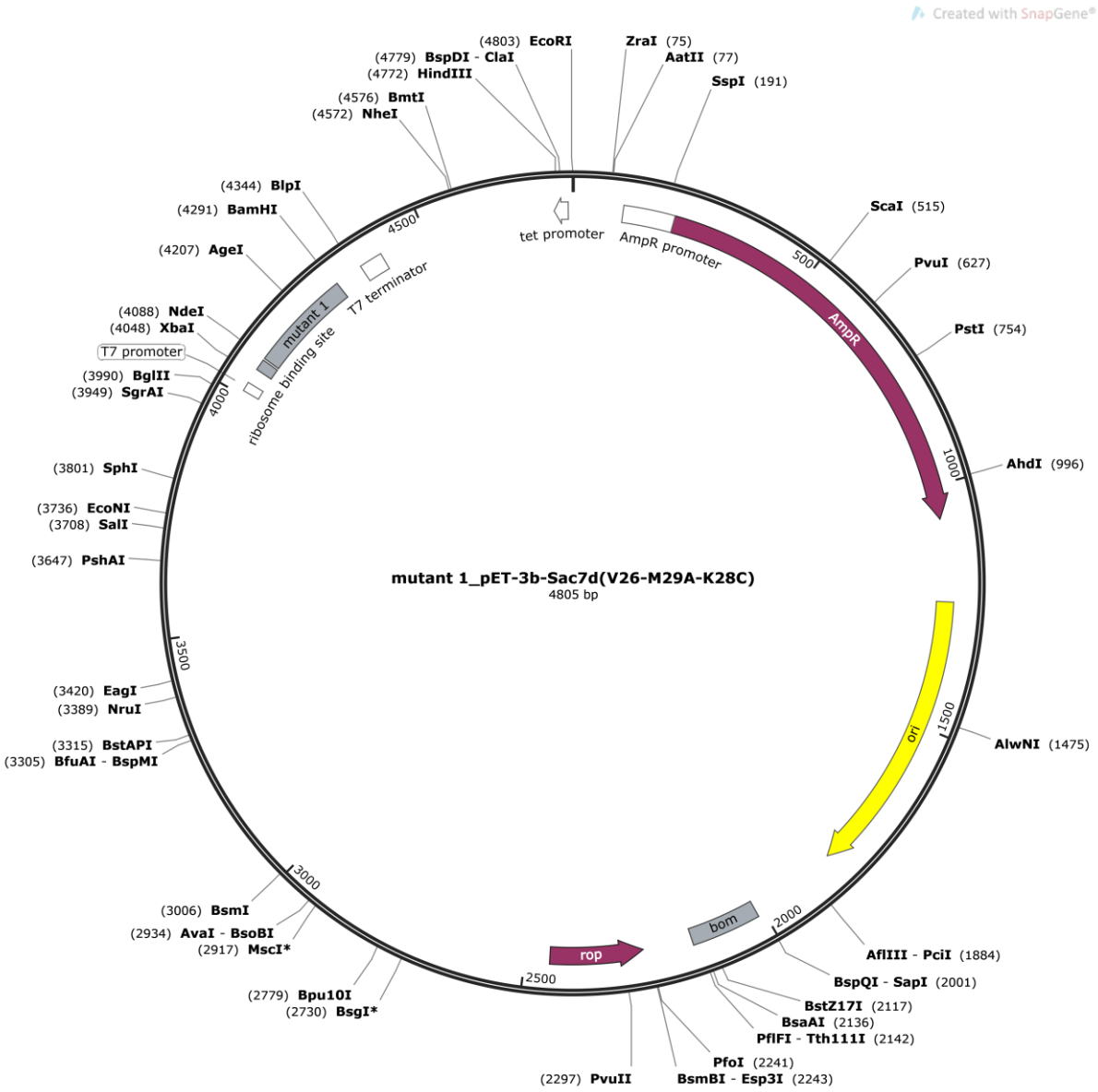


Fig. 95. Sac7d cysteine mutant plasmid map of mutant 1 (pET3b-Sac7d V26A/M29A/K28C).

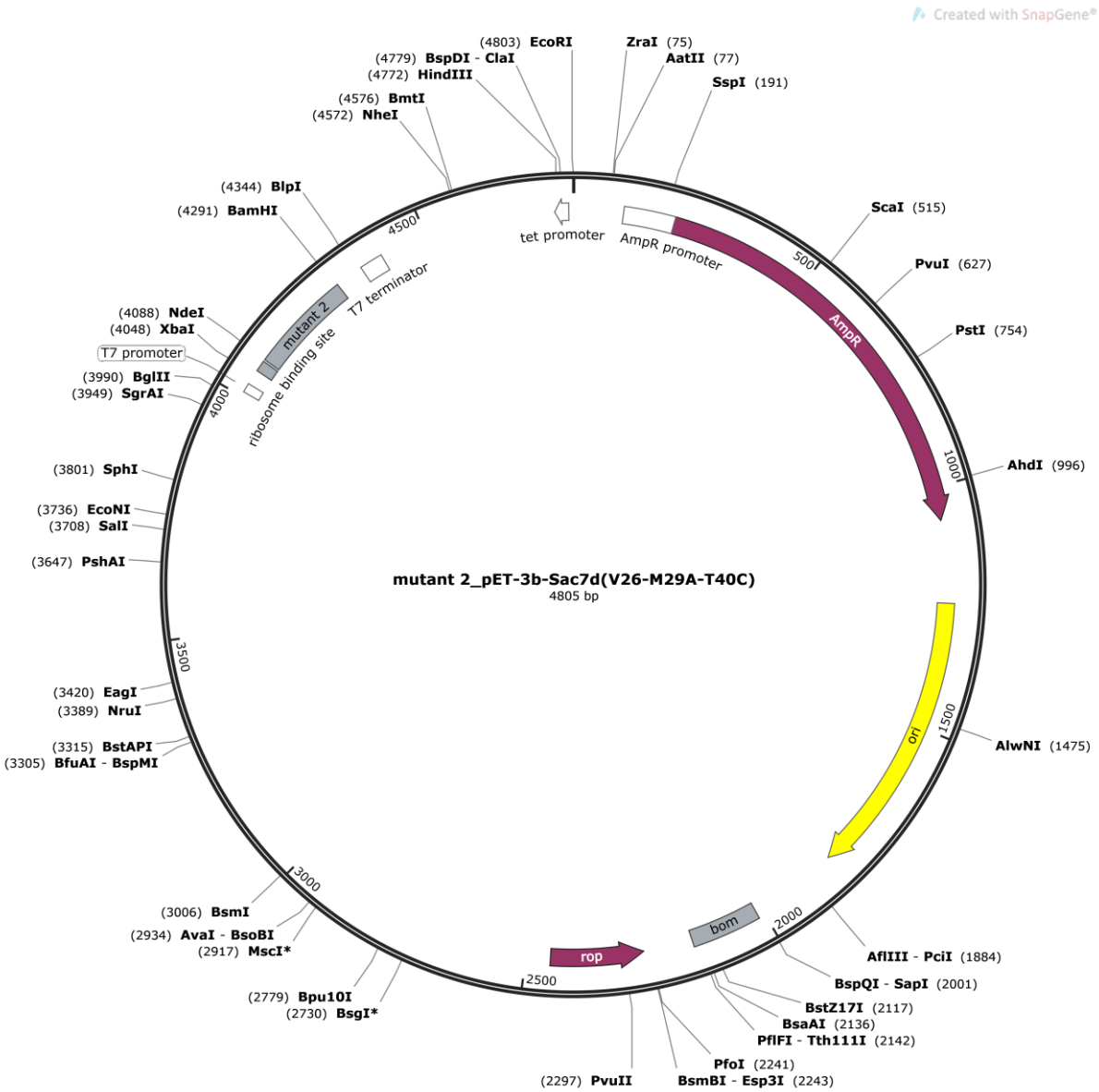


Fig. 96. Sac7d cysteine mutant plasmid map of mutant 2 (pET3b-Sac7d V26A/M29A/T40C).

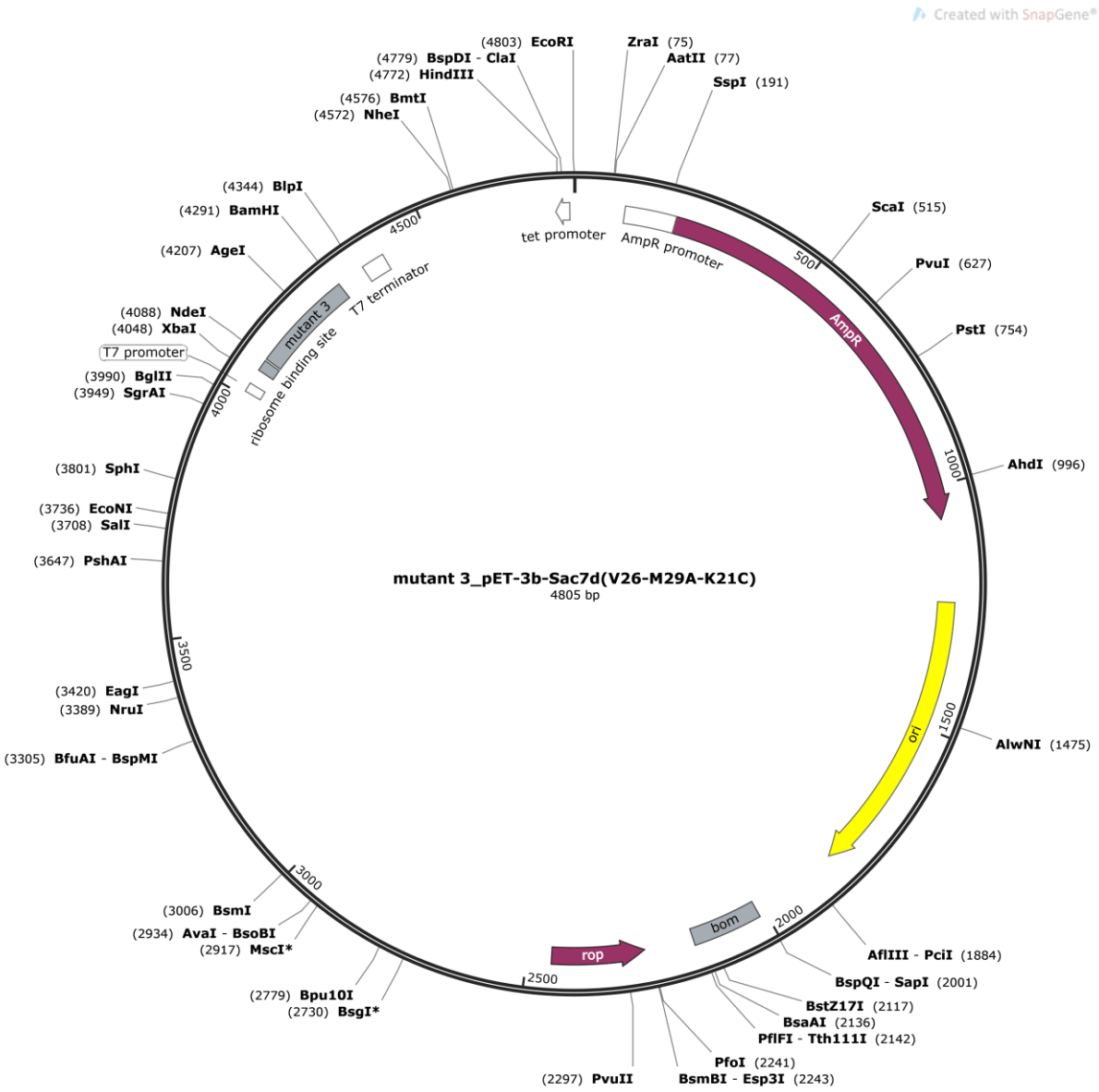


Fig. 97. Sac7d cysteine mutant plasmid map of mutant 3 (pET3b-Sac7d V26A/M29A/K21C).

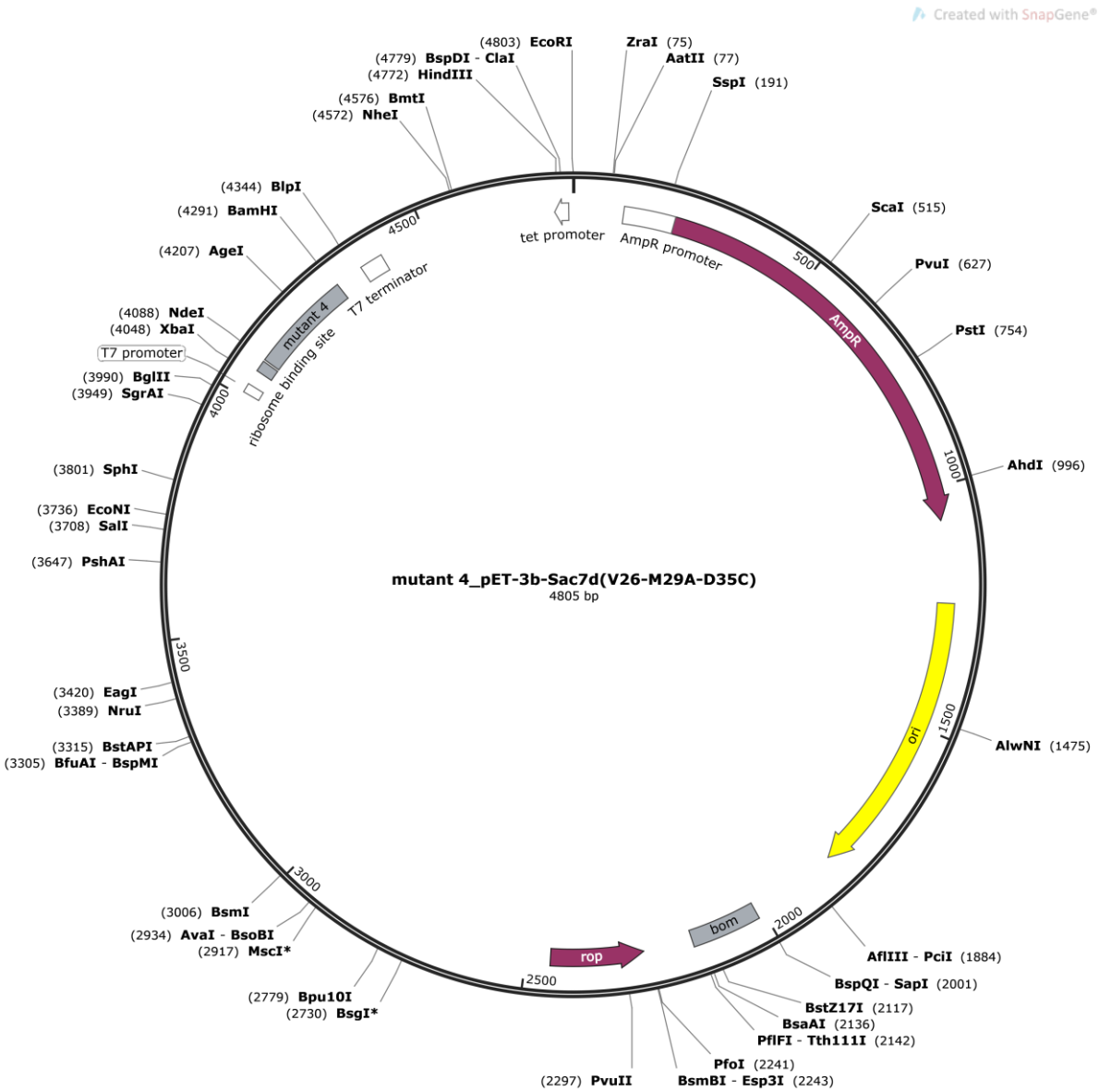


Fig. 98. Sac7d cysteine mutant plasmid map of mutant 4 (pET3b-Sac7d V26A/M29A/D35C).

10.2 Abbreviations

bZIP	basic leucine zipper
CD	circular dichroism
ChiP	chromatin immunoprecipitation
DBP	DNA binding protein
DNA	deoxyribonucleic acid
DESY	deutsches elektronen synchrotron
D_2O	deuterium oxide
EDTA	ethylenediamine tetraacetic acid
EMSA	electrophoretic mobility shift assay
EMBL	european molecular biology laboratory
ESI	electrospray ionization
ESRF	european synchrotron radiation facility
FPLC	fast protein liquid chromatography
HMG	high mobility group
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence spectroscopy
HTH	helix turn helix
Ig fold	immunoglobulin fold
IPTG	isopropyl β- d-1-thiogalactopyranoside
TBP	tata box binding protein
LCMS	liquid chromatography mass spectrometry
LEF-1	lymphoid enhancer factor
LLG	log-likelihood gain
MAD	multiple-wavelength anomalous diffraction
MD	molecular dynamic
MES	2-ethanesulfonic acid
NMR	nuclear magnetic resonance
NOESY	nuclear overhauser effect spectroscopy

Ni-NTA	nickel nitrilotriacetic acid
PDB	protein data bank
PRCG	polack-ribiere conjugate gradient
RFZ	rotation function Z-score
RNA	ribonucleic acid
RNAi	RNA interference
RPA	replication protein A
RT	room temperature
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SD	steepest descent
SEC	size exclusion chromatography
SPR	surface plasmon resonance
THF	tetrahydrofuran
TOCSY	total correlation spectroscopy
Tris	tris(hydroxymethyl)aminomethane
TFZ	translation function Z-score
UV	ultraviolet

10.3 Index of figures

Fig. 1. Illustration of the central dogma of biology and key players involved. 8

Fig. 2. Fragment of DNA primary structure shown with the sequence 5'-AGCT-3'. 9

Fig. 3. Comparison of different forms of double-helical DNA. A-DNA in the left (PDB ID 1QPH), B-DNA in the middle (PDB ID 1BNA), Z-DNA in the right (PDB ID 4OCB) 10

Fig. 4. Major groove and minor groove features of a double-stranded B-DNA. **a)** A 12-base pair double-stranded B-DNA with cartoon and surface representation to highlight the major groove (dashed line) and minor groove (solid line). **b)** Functional group presented by the major groove compared to minor groove on a GC and AT base pair. 11

Fig. 5. Classification of DNA binding proteins based on the motif involved in DNA recognition. 13

Fig. 6. Helix-turn-helix and winged-helix-turn-helix crystal structure. (a) The Trp-DNA complex shows a Helix-Turn-Helix motif binding the major groove of the DNA fragment (PDB ID 6CRO). (b) The HNF3-DNA complex shows the winged-helix-turn-helix. Notice the extra beta-sheet interaction with the minor groove and alpha helix interaction with the major groove of the DNA fragment. (PDB ID 1VTN) 14

Fig. 7. A Basic leucine zipper motif. GCN4 protein bound to DNA segment (PDB ID 1YSA). Dimerization and DNA binding domains are marked individually. 15

Fig. 8. Crystal structure of a ZIF2628 containing a zinc finger in complex with DNA (PDB ID 1ZAA) with a zoomed view at the zinc atom coordinated by two histidine and two cysteine residues. See also Fig. 67. . 15

Fig. 9. Crystal structure of an Anabaena HU-DNA complex consisting of beta ribbon motifs (PDB ID 1P71). . 16

Fig. 10. NMR structure ensemble of LEF-1 protein bound with DNA (PDB ID 2LEF). 17

Fig. 11. Crystal structure of a P53 dimer containing immunoglobulin fold in complex with DNA (PDB ID 3EXJ). 18

Fig. 12 General workflow of protein-ligand crystallography. Protein to Structure approach is shown in a stepwise workflow elucidating protein-ligand complex crystal structure 20

Fig. 13 A simplified phase diagram of a protein (ligand) crystallization experiment. Note the crystal growth in the metastable zone. 21

Fig. 14. Relative size of the molecules that target proteins. 26

Fig. 15. Small molecule drugs inhibition mechanism rendition on transcription factor binding on DNA. 27

Fig. 16. Peptide-based modalities connected to a drug candidate. 28

Fig. 17. Targeting DNA binding proteins using molecular degraders and PROTACS. a) Monomeric degrader binding to target DNA binding protein leads to its degradation with various mechanisms. b) PROTAC degraders work by linking a binder of the target protein (red) to an E3 ligase ligand (yellow) that leads to the degradation of the target DNA binding protein. 30

Fig. 18. Surface charge distribution of an 8BP B-DNA (PDB ID 6BEK) and two DNA binding proteins, Arn (PDB ID 3WX4) and SAUGI (PDB ID 3WDG). The negative charge distribution is highlighted with dashed circles. 32

Fig. 19. Chemical difference between a DNA chain versus a PNA chain. 33

Fig. 20. Folding principle of aza-aromatic oligoamides. Hydrogen bonds are shown in dashed lines. Dipole repulsions are shown with two-headed arrows. 36

Fig. 21. An overview of the various foldamer families. a) Crystal structure of a β-peptide 12-mer. b) Crystal structure of a helical aryl oligomer. c) Crystal structure of a sheet-forming aromatic oligoamide. d) Crystal structure of an aryl disulfide macrocycle. The disulfide bonds are shown in orange, while the oxygen and nitrogen atoms are marked in purple and blue, respectively. The hydrogen atoms and side chains have been left out for clarity in all the structures. Figure adapted from figure number 2.1 from Bindl D.[136]..... 37

Fig. 22. DNA mimic foldamer design. a) Chemical formulae of monomers used in the synthesis of DNA mimic foldamers. b) Chemical formulae of $^{m}QQ^{4}$ and $^{m}QQ^{5}$ used in the synthesis of DNA mimic foldamers. c-f) Top view and side view of $(Q)_{16,}$ $(^{m}Q)_{16,}$ $(^{m}QQ^{4})_{8,}$ and $(^{m}QQ^{5})_{8}$. g) Overlay of an $(^{m}QQ^{5})_{8}$ with an 8 base pair DNA showing the overlay of phosphorus atoms positions (shown in balls) h). Top view and side view of an 8 base pair double-stranded B-DNA. Figure adapted from Figure 1 of Ziach et al., 2018.[147]..... 39

Fig. 23. DNA vs. DNA mimic foldamer. a) 8 base pair B-DNA shown in stick model representation with phosphate on each strand highlighted in red- and blue-colored balls. b) Surface representation of 8 base pair DNA. c) A $(^{m}QQ^{4})_{8}$ DNA mimic foldamer shown in stick representation with negatively charged phosphonates highlighted in red- and blue-colored balls. d) Surface representation of a $(^{m}QQ^{4})_{8}$ DNA mimic foldamer. 40

Fig. 24. A simple rendition of a phylogenetic tree showing Sac7d, among other DNA bending proteins ranging from HU in bacteria to HMG-box containing protein SRY in eukaryotes. 41

Fig. 25. Crystal structure of Sac7d in complex with an 8BP double-stranded B-DNA (PDB ID 1AZQ). Sac7d residues intercalating in DNA chains are shown in stick representation (V26 and M29). 42

Fig. 26. Sharp kinking versus smooth bending of DNA by Sac7d. a) Wild-type Sac7d intercalates V26 and M29 residues in between 3rd and 4th base pair of an 8BP DNA strand inducing a sharp kink of 61 °. b) Sac7d V26A/M29A, due to shorter side chains at positions 26 and 29, induces a significantly smooth bending, resulting in a roll of 29 °. 43

Fig. 27. General mechanism of enzyme activity inhibition by competitive inhibition (center) is shown along with normal binding (left) and non-competitive inhibition (right). 45

Fig. 28. Principle of competitive inhibition of protein-DNA interaction using foldamers that mimic the DNA surface features. 46

Fig. 29. Comparison of Sac7d-DNA complex versus the protein-DNA mimic foldamer complex (docked). 49

Fig. 30. Sketch of a protein-foldamer adduct with tethering approach. 81

Fig. 31. Comparison of Sac7d with Nanofitin a) Sac7d is shown in ribbon representation (PDB ID 1AZQ, DNA not shown). b) Nanofitin (Sac7d, derived scaffold) is shown with mutated residues highlighted in stick representation compared to the wild-type Sac7d (PDB ID 2XIW). 83

Fig. 32. Scheme following the design of Sac7d cysteine mutant library. 84

Fig. 33. Amino acid sequence of wild-type Sac7d and Sac7d V26A/M29A aligned with cysteine mutant library. 84

Fig. 34. Ribbon representation of Sac7d cysteine mutant (shown in stick representation). a) Mutant 1, b) Mutant 2, c) Mutant 3, d) Mutant 4. Note that these cysteine residues were mutated based on wild-type Sac7d (PDB ID 1AZQ) in Coot. Bound DNA to Sac7d is not shown for clarity purposes. 85

Fig. 35. a) Different building blocks used to synthesize the foldamers. b) N-terminal activated disulfide containing chiral DNA mimic foldamer. 86

Fig. 36. Workflow and different steps involved in solid phase synthesis of foldamer. 86

Fig. 37. Solid phase synthesis for a Q^4-based sequence. Various steps and reagents involved are a) 20% piperidine in DMF; b) Monomer, Cl_3CCN, PPh_3, collidine, THF/NMP; c) 50% Ac_2O in DCM; d) piperidine in NMP; e) triphosgene, DIPEA, THF; f) Linker-NH2, DIPEA, THF. 87

Fig. 38. Sac7d cysteine mutant plasmid maps. a) Mutant 1 (pET3b-Sac7d V26A/M29A/K28C) b) Mutant 2 (pET3b-Sac7d V26A/M29A/T40C) c) Mutant 3 (pET3b-Sac7d V26A/M29A/K2C) d) Mutant 4 (pET3b-Sac7d V26A/M29A/D35C). Please see the **section 10.1** for enlarged plasmid maps. 88

Fig. 39. Scheme of various steps involved in the expression and purification of mutant 1, mutant 2, mutant 3, and mutant 4. 89

Fig. 40. a) Front view and (b) bottom view of the energy minimized model of the chiral Q^4Pho-based foldamer **2** with monoanionic phosphonate side chains and the chiral B^R unit in position 14. Alternate phosphorus atoms are colored dark blue and red. (c) The solvent-accessible surface of the energy-minimized model. Please note the double headed arrow showing the right-handed helix of the negative charge mimicking the right handed B-DNA. 90

Fig. 41. Energy minimization followed by MD simulation showing one of the trajectories of tethered foldamer **2** (shown in stick representation with phosphorus atoms in blue and red balls) with mutant 1 (shown in green ribbons). 91

Fig. 42. Size exclusion chromatograms and 15% SDS-PAGE profiles of mutant 1 (top) and mutant 2 (bottom) purified on Superdex 75 pg column. 92

Fig. 43. Size exclusion chromatograms and 15% SDS-PAGE profiles of mutant 3 (top) and mutant 4 (bottom) purified on Superdex 75 pg column. 93

Fig. 44. LC chromatogram at UV 280 nm and ESI-MS spectrum of Sac7d V26A/M29/K28C (Mutant 1). 94

Fig. 45. LC chromatogram at UV 280 nm and ESI-MS spectrum of Sac7d V26A/M29/K21C (Mutant 3). 94

Fig. 46. Size exclusion chromatogram of mutant 3-foldamer **2** tethered adduct. Fractions corresponding to adduct eluted between 70 ml and 85 ml. 95

Fig. 47. LC chromatogram at UV 300 nm and ESI-MS spectrum of mutant 3-foldamer **2** adducts. 95

Fig. 48. Size exclusion chromatogram of mutant 1-foldamer **2** tethered adduct. Fractions corresponding to adduct eluted between 70 ml to 85 ml. 96

Fig. 49. LC chromatogram at UV 300 nm and ESI-MS spectrum of mutant 1-foldamer **2** adduct. 96

Fig. 50. Crystal of mutant 1-foldamer **2** adducts observed under cross-polarizing microscope. 97

Fig. 51. An example of a ligation experiment between mutant 1 and foldamer **2**..... 98

Fig. 52. Structure-based iterative design of foldamers as ligands for interfering with protein-DNA interactions.99

Fig. 53. Overview of cGAS-STING signaling pathway showing the dsDNA-induced activation of cGAS, which leads to the production of cGAMP that binds to STING to express the type 1 interferon and other chemokines. 102

Fig. 54. Crystal structure of hcGAS-DNA. a) Schematic representation of hcGAS domains organization. b) Crystal structure of hcGAS-DNA. c) Crystal structure shown in surface representation highlighting the fit of DNA to the binding site on hcGAS (PDB ID 6EDC). 103

Fig. 55. Crystal structure of mcGAS dimer in complex with DNA (PDB ID 4O6A). 104

Fig. 56. C_2-Symmetrical DNA mimic foldamers. a) chemical formula of monomers used in DNA mimic foldamers synthesis. b, c) C_2-symmetrical DNA mimic foldamers with a chiral B monomer and central C_2-symmetrical moiety. 105

Fig. 57. a) Solid phase synthesis of a 9-mer chiral DNA mimic: a) CBr_4, PPh_3, DMF; b) Fmoc-Q-OH, CsI, DIEA, DMF; c) 20% piperidine in DMF; d) Monomer, TCAN, PPh_3, collidine, THF/$CHCl_3$; e) 50% Ac_2O in DCM; f) TFA. **b**) Chiral C_2-symmetrical DNA mimic foldamer synthesized using a diamine linker. **c**) Chiral C_2-symmetrical DNA mimic foldamer synthesized using an activated diacid linker.[210] 106

Fig. 58. Plasmid map of pRSFDuet-sumo-hcGAS (K427E/K428E). 107

Fig. 59. DNA sequence used in crystallization with His_6-SUMO hcGAS (K427E/K428). 108

Fig. 60. a) hcGAS construct design with N terminus histidine and SUMO tag. b) pRSFDuet-sumo-hcGAS (K427E/K428E) transformed in DH5 alpha cells for plasmid isolation. c) pRSFDuet-sumo-hcGAS (K427E/K428E) transformed in *E. coli* BL21-CodonPlus (DE3)-RIL cells for over expression. 110

Fig. 61. Over expression profile of His_6-SUMO hcGAS (K427E/K428E). Recombinant protein over expressed with IPTG induction and analyzed on a 12 % SDS-PAGE. Lane 1 is the protein ladder marker; Lane 2 is the crude sample after sonication; Lane 3 is the supernatant after sonication. 111

Fig. 62. Affinity purification profile of His_6-SUMO hcGAS (K427E/K428E) analyzed on a 12 % SDS-PAGE gel. Lane 1 is the protein ladder marker; Lane 2 is supernatant after sonication; Lane 3 is the flow-through fraction of Ni-NTA resin after binding to the protein of interest; Lane 4 is the wash fraction with buffer A supplemented with 50 mM imidazole; Lane 5 is the elution faction with buffer A supplemented with 500 mM imidazole. His-SUMO hcGAS (K427E/K428E) protein is highlighted with an arrow. 111

Fig. 63. Size exclusion chromatography (16/600 Superdex 200 column prep grade column) and 12 % SDS-PAGE profile of His_6-SUMO hcGAS (K427E/K428E). Please note that the x-axis on the chromatogram is in minutes. The purification was performed at a 0.4 ml/min flow rate. Lane 1 in SDS-PAGE is a pre-stained protein ladder; Lane 2 is the pure fraction of His_6-SUMO hcGAS (K427E/K428E). 112

Fig. 64. Over expression and purification profile of His_6-SUMO hcGAS (K427E/K428E) analyzed on a 12 % SDS-PAGE. Lane 1 is the pre-stained protein ladder marker; Lane 2 is the crude sample after sonication; Lane 3 is the supernatant after sonication; Lane 4 is the elution profile after the first round of Ni-NTA purification; Lane 5 is flow through of cleaved hcGAS (K427E/K428E) after the second round of Ni-NTA purification. 113

Fig. 65. Crystals of His_6-SUMO hcGAS (K427E/K428E) visualized under the microscope. 113

Fig. 66. Crystal structure of hcGAS (K427E/K428E) solved at 3.5 Å. Note that the His_6-SUMO tag is not resolved due to the lack of electron density around the tag. 114

Fig. 67. Zoomed view of the zinc finger motif in the crystal structure of the hcGAS (K427E/K428E). 116

Fig. 68. Crystals of hcGAS (K427E/K428E)-foldamer **1** observed under crossed polarizing microscope grown in different crystallization solutions. a) 35% (v/v) MPD, 100 mM HEPES/ Sodium hydroxide pH 7.5, 200 mM Sodium chloride. b) 35% (v/v) MPD, 100 mM Tris base/ Hydrochloric acid pH 7.0, 200 mM Sodium chloride. c) 50% (v/v) Polyethylene glycol 200, 100 mM TRIS pH8.0. 117

Fig. 69. Crystals of hcGAS (K427E/K428E)-foldamer **1** observed under crossed polarizing microscope grown in different temperatures in 35% (v/v) MPD, 100 mM HEPES/ Sodium hydroxide pH 7.5, 200 mM Sodium chloride as crystallization solution. 118

Fig. 70. a) Graph highlighting the resolution versus number of spots during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** grown at 4 °C in 35% (v/v) MPD, 100 mM HEPES/Sodium hydroxide at pH 7.5, 200-mM Sodium chloride. b) Diffraction map image (1 of 4) during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** complex..... 118

Fig. 71. a) Graph highlighting the resolution versus number of spots during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** grown at 20 °C in 50% (v/v) Polyethylene glycol 200, 100 mM TRIS pH8.0. b) Diffraction map image (1 of 4) during the characterization of crystals of hcGAS (K427E/K428E)-foldamer **1** complex. 119

Fig. 72. New generation of foldamer sequences to be screened with hcGAS for crystallization. a,b) C_2-symmetrical DNA mimic foldamer (34 mer) with flexible and rigid C_2-symmetrical linker. c) C_2-symmetrical DNA mimic foldamer (34 mer) with Q-selenium monomer. d) C_2-symmetrical DNA mimic foldamer (32 mer) with sticky ends. 120

Fig. 73. General "hand" like structural repersentation of a DNA polymerase ternary complex (PDB ID 3NCI) a) and b) shown as schematic and surface representation respectively. 122

Fig. 74. Crystal structure of the Dpo4 ternary complex. 123

Fig. 75. Schematic representation of the "hand" model of the Dpo4 binding to d-DNA and l-DNA 124

Fig. 76. DNA sequence used in crystallization with Dpo4. 126

Fig. 77. *C*$_2$-Symmetrical DNA mimic foldamers used in crystallization with Dpo4. a) chemical formula of monomers used in DNA mimic foldamers synthesis. b, c) C_2-symmetrical DNA mimic foldamers with a chiral B monomer and central C_2-symmetrical moiety..... 127

Fig. 78. pET21b-Dpo4-His$_6$ transformed in DH5 alpha cells for plasmid isolation. c) pET21b-Dpo4-His$_6$ transformed in *E. coli* BL21 (DE3)pLysS cells for over expression of Dpo4 protein. 128

Fig. 79. Over expression profile of Dpo4. Recombinant protein over expressed with IPTG induction and analyzed on a 12 % SDS-PAGE. Lane 1 is the protein ladder marker; Lane 2 is the crude sample after sonication and heating; Lane 3 is the supernatant after ultracentrifugation..... 129

Fig. 80. Affinity purification profile of Dpo4 analyzed on a 12 % SDS-PAGE gel. Lane 1 is the protein ladder marker; Lane 2 is the flow-through fraction of Ni-NTA resin after binding to the protein of interest; Lane 3 is the wash fraction with buffer A supplemented with 50 mM imidazole; Lane 4 is the elution faction with buffer A supplemented with 500 mM imidazole. Note that the Dpo4 protein is highlighted with an arrow..... 130

Fig. 81. Size exclusion chromatogram (16/600 Superdex 75 column prep grade column of Dpo4. Please note that the x-axis on the chromatogram is in minutes. The purification was performed at a 0.6 ml/min flow rate. 130

Fig. 82. LC chromatogram at UV 280 nm and ESI-MS spectrum of Dpo4..... 131

Fig. 83. Crystals of Dpo4-d-DNA complex visualized under the microscope. Note the change from a) to d) in the number of crystals as the concentration of the complex was decreased to yield single crystals. 132

Fig. 84. Crystal structure of the Dpo4-d-DNA solved at 2.6 Å. Note the hand-like structure of Dpo4 grabbing the d-DNA..... 133

Fig. 85. Crystal structure of the Dpo4- d-DNA complex with a zoom on the active site where Dpo4 residues and DNA interact with the Ca^{2+} ions (yellow spheres). 133

Fig. 86. **Crystals of Dpo4-foldamer 2 observed under crossed polarizing microscope grown in different crystallization solutions.** a) Crystals grown in 35 % MPD, 0.1 mM HEPES pH 7.5. b) Crystals grown in 10 % mPEG 5000, 0.1 M Tris pH 8.5. 135

Fig. 87. Crystals of Dpo4-foldamer 1 observed under crossed polarizing microscope grown in different crystallization solutions. a), b), c) Crystals grown in 18 % PEG 4000, 0.1 M Tris pH 8.5 with Dpo4: foldamer **1** in 0.2 mM: 0.26 mM, 0.15 mM: 0.19 mM, and 0.10 mM: 0.13 mM respectively in a), b), and c). d) 1M Ammonium sulfate, 0.1 M Tris pH 8,0. e) 15 % PEG 6000, 0.1M Magnesium acetate, 0.1M Sodium cacodylate pH 6.5. f) 18 % PEG 6000, 0.1 M Tris pH 8.5..... 136

Fig. 88. a) Diffraction map image (1 of 4) during the characterization of crystals of Dpo4-foldamer **1** complex from Fig. 87 d). b) Diffraction map image (1 of 4) during the characterization of crystals of Dpo4-foldamer **1** complex from Fig. 87 e). 136

Fig. 89. Crystals of Dpo4-foldamer 2 observed under crossed polarizing microscope grown in different crystallization solutions. a) Crystals grown in 20 % PEG 3350, 0.2 M di-Sodium malonate. b) 20 % PEG 3350, 0.2 M Sodium fluoride. c) Crystals grown in 20 % PEG 3350, 100 mM BIS-TRIS propane pH 6.5, 200 mM Sodium acetate. 137

Fig. 90. Xtriage analysis of Dpo4-foldamer **1** crystals from Fig. 89 c). 137

Fig. 91. Sketch of a protein-foldamer adduct with tethering approach. 143

Fig. 92. a) Crystal structure of SAP1-DNA complex (PDB ID 1BC7). b) Simple cartoon representation of the foldamer (shown in sticks and ball representation) with DNA (shown in ladder and ring representation) on SAP1 protein (shown in ribbon representation). 143

Fig. 93. Electron density map (blue and green) after the molecular replacement of Sac7d alone. Notice the electron density in green, showing the foldamer presence in the crystal structure. 144

Fig. 94. Comparison of the crystal structure of DNA mimic foldamer versus the molecular model of DNA mimic foldamer. a) Crystal structure of 18mer DNA mimic foldamer (in complex with Sac7d). b) Molecular model of 18mer DNA mimic foldamer. c) Overlay of the a) and b), note the difference in length of the crystallized versus the molecular model in the zoom view on the right. 145

Fig. 95. Sac7d cysteine mutant plasmid map of mutant 1 (pET3b-Sac7d V26A/M29A/K28C). 160

Fig. 96. Sac7d cysteine mutant plasmid map of mutant 2 (pET3b-Sac7d V26A/M29A/T40C). 161

Fig. 97. Sac7d cysteine mutant plasmid map of mutant 3 (pET3b-Sac7d V26A/M29A/K21C). 162

Fig. 98. Sac7d cysteine mutant plasmid map of mutant 4 (pET3b-Sac7d V26A/M29A/D35C). 163

10.4 Index of tables

Table S1. Pipetting scheme of [^{15}N]-Sac7d-**3** and [^{15}N]-Sac7d-**4** for NMR spectroscopy..... 73

Table S2. Crystallization parameters for Sac7d and Sac7d V26A/M29A protein with **3** and **4**..... 74

Table S3. Crystallography data collection and structure refinement statistics for Sac7d in complex with **3 (PDB #8CMN)**..... 76

Table S4. Crystallography data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **3 (PDB #8Q2M)**..... 77

Table S5. Crystallography data collection and structure refinement statistics for Sac7d V26A/M29A in complex with **4 (PDB # 8QPC)**..... 78

Table 6. Initial statistics of auto-processed crystal data of mutant 1-foldamer **2** adduct..... 97

Table 7. Unit cell parameters of mutant 1-foldamer **2** adduct from the initial auto-processed crystal data. 97

Table 8. Crystallography data collection and structure refinement statistics for His_6-SUMO hcGAS (K427E/K428E)..... 115

Table 9. Crystallography data collection and structure refinement statistics for Dpo4-d-DNA..... 134

Table 10. Crystallography data collection statistics for Dpo4-foldamer **1** crystals..... 138

10.5 List of publications, PDB depositions, and conference contributions

10.5.1 Publications

- **D. Deepak**, V. Corvaglia, J. Wu, L. Allmendinger, I. Huc, "DNA Mimic Foldamer Recognition of a Chromosomal Protein" (communicated to Angewandte Chemie)
- V. Corvaglia, J. Wu, **D. Deepak**, M. Loos, I. Huc, "Enhancing the Features of DNA Mimic Foldamers for Structural Investigations" (communicated to Chemistry – A European Journal)

10.5.2 PDB structure depositions:

- 8CMN, 8Q2M, 8QPC, and 8QHM.

10.5.3 Conferences

- **Oral presentation (short lecture)**, "Structure elucidation of a protein-DNA mimic foldamer complex" at Symposium on foldamers, Ludwig Maximilian University of Munich, Munich, Germany, 4 - 6 September 2023.
- **Poster presentation** (and flash talk), "Engineering DNA mimicry: Structural insights into DNA mimic foldamer-Sac7d interactions" at Interact science and translation conference, Ludwig Maximilian University of Munich, Munich, Germany, 30 - 31 March 2023.
- **Conducted workshop**, "Alpha-helix interaction with DNA major groove." at the 4th international workshop on Aromatic Foldamers: Structures and Functions at Ludwig Maximilian University of Munich, Munich, Germany, 8 – 10 June 2022.
- **Poster presentation**, "Structural investigation of protein-foldamer adducts: a tethering approach," at the 7th Symposium on Foldamers at European Institute of Chemistry and Biology, Bordeaux, France, 13 - 15 September 2021.
- **Conducted workshop**, "DNA binding proteins," at the 3rd international workshop on Aromatic Foldamers: Structures and Functions at European Institute of Chemistry and Biology, Bordeaux, France, 13 - 15 September 2021.
- **Workshop presentation**, "DNA mimicry: state of the art and prospects," at the international workshop on "Foldamers: Structures and Functions" at Ludwig Maximilian University of Munich, Munich, Germany, 28 - 30 October 2019.

10.6 Acknowledgment

First and foremost, I would like to express my deepest gratitude to my PhD supervisor and mentor, Prof. Dr. Ivan Huc. His unwavering support, visionary guidance, and inspiring ideas were instrumental in making the challenging and exciting projects I worked on possible. Throughout my entire PhD journey, his trust in my abilities and the joy of scientific discovery we shared were invaluable. I also appreciate the autonomy he provided, which allowed me to grow as a scientist and navigate challenging situations.

I am equally thankful to all the members of the Huc group for fostering a collaborative working environment. Their support and teamwork were always there when I needed them. Special thanks go to Dr. Valentina Corvaglia, Jiaojiao Wu, and Manuel Loos. Not only did they provide valuable foldamers, but we also collaborated on multiple experiments.

I extend my gratitude to Dr. Pradeep Mandal and Dr. Vasily Morozov for their crucial roles in establishing the biochemistry lab within the Huc group. Their expert guidance was instrumental in my research. I would also like to thank Johannes Sigl and Dr. Tulika Chakrabortty for their scientific contributions and interactions within the lab.

Next, I am deeply indebted to the German Academic Exchange Service (DAAD) for providing me with the PhD scholarship that supported my academic journey. This thesis is printed with the support of the DAAD.

I am also grateful for the unwavering support I received during my BSc at Delhi University from all the professors, particularly the late Prof. Dharmedra Mallick, who instilled in me a scientific mindset, and Prof. Pawan Dhar of JNU, who has been a continuous source of support and mentorship. It was a privilege to be mentored by Prof. Dharmendra Mallick and Prof. Pawan Dhar.

My heartfelt thanks also go to Karishma, who stood by my side throughout my PhD, providing valuable feedback and proofreading assistance for my thesis.

In the end, मेरी माँ, बाबू, बेबे, जीजा का ट्राली भर क धन्यवाद जो के मेरे गेल हमेशा खड़े रहे सै। मेरे माँ बाबू की मेहनत के बिना या पीएचडी कदे भी ना कर पांदा। और आखर मै सबतै जरुरी मेरा प्यारा लाडला भांजा मिठु जिसकी वजह त करड़ा टेम भी जमा आसानी त कट गया।