Universiteit Leiden
The Netherlands

The design of transcription factor-based inhibitors to target Myc: drop the Myc!

Ellenbroek, B.D.

Citation
Ellenbroek, B. D. (2026, June 10). *The design of transcription factor-based inhibitors to target Myc: drop the Myc!*. Retrieved from https://hdl.handle.net/1887/4305022

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4305022

Note: To cite this publication please use the final published version (if applicable).

The Design of Transcription Factor-based Inhibitors to target Myc

Drop the Myc!

Brecht David Ellenbroek

Cover description: The design process for all proteins described in this thesis to inhibit Myc. The figures are either generated with AlphaFold-Multimer or based on PDB 5i50 and modified with Pymol and Adobe Illustrator.

Cover design by Brecht D. Ellenbroek
Thesis layout by Brecht D. Ellenbroek
This thesis was printed by Ridderprint, the Netherlands
© Brecht David Ellenbroek, 2026

ISBN: 978-94-6537-447-5

All rights reserved. No part of this thesis may be reproduced in any form or by any means without permission of the author.

The Design of Transcription Factor-based Inhibitors to target Myc

Drop the Myc!

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op woensdag 10 juni 2026
klokke 14:30 uur

door

Brecht David Ellenbroek
geboren te Amsterdam, Nederland
in 1995

Promotores:
Prof. Dr. M. van Eck
Prof. Dr. N. I. Martin

Co-promotor:
Dr. S. J. Pomplun

Promotiecommissie:
Prof. Dr. L.H. Heitman
Prof. Dr. F. Thomas. (Universität Heidelberg)
Dr. M. Jbara (Tel Aviv University)
Prof. Dr. A. Kros
Prof. Dr. E.C.M. de Lange

The research in this thesis was performed at the devision of Medicinal Chemistry of the Leiden Academic Centre for Drug Research (LACDR), Leiden University, the Netherlands. The reseach in this PhD thesis was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (StG SyntTra-101039354) and additional financial support from Mr. H.J.M. Roels through a donation to Oncode Institute and the Dutch Cancer Society (KWF} financial support of Oncode Institute.

Table of contents

Chapter 1 **9**

General introduction

Chapter 2 **33**

Synthetic Peptides: Promising Modalities for the Targeting of Disease-Related Nucleic Acids

Chapter 3 **71**

Development of DuoMYC: A synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC

Chapter 4 **121**

ReCHEMbinant Stapling Enhances Intracellular Delivery and Bioactivity of Engineered Protein Inhibitors

Chapter 5 **159**

AlphaFold-Guided Discovery of an Overlapping MYC/Miz-1 Interface Enables Peptidomimetic Disruption of MYC/MAX

Chapter 6 **203**

Conclusion and future perspectives

Appendix A

List of abbreviations 223

Nederlandse samenvatting 225

English summary 229

List of publications 233

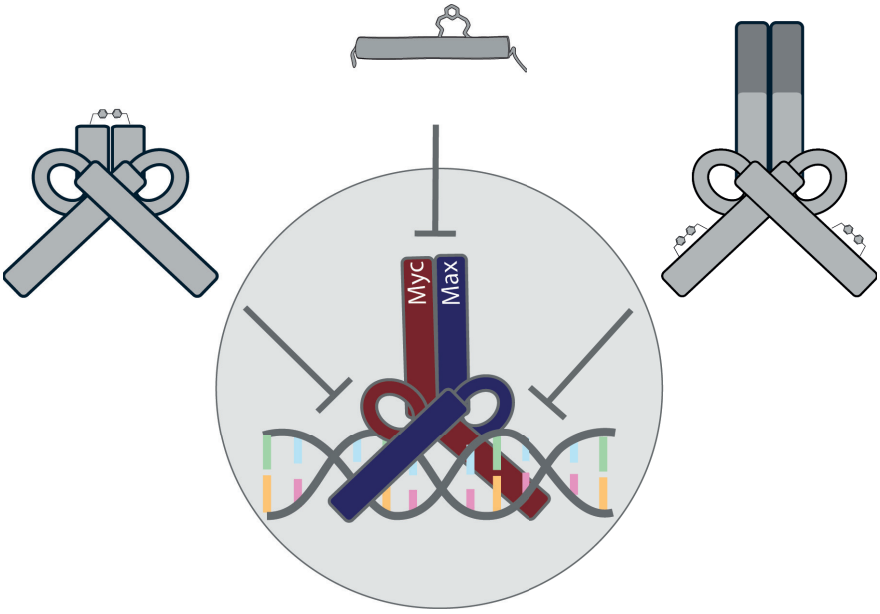
Curriculum Vitae 235

Acknowledgement 237

Chapter 1

1

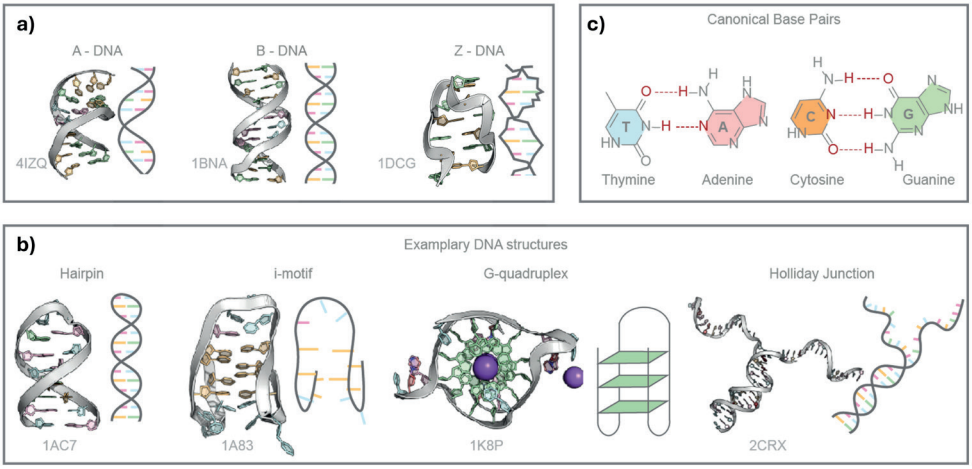
General introduction



Francis Crick described the central dogma of molecular biology as "*the exchange of genetic data, embedded in deoxynucleic acid (DNA), goes in only one direction to ribonucleic acid (RNA) to proteins.*"[1] The first step in this dogma is to read DNA and transcribe this to RNA. This step is performed by the transcription machinery that is recruited towards DNA. This protein complex reads DNA and simultaneously transcribes it to RNA. Subsequently, RNA is translated into the encoded protein.

DNA can adopt multiple conformations that allow DNA binding proteins to recognize DNA sequences. It generally consists as a right-handed double-stranded helical structure, called B-DNA, and was first described by Watson and Crick. [2] Other secondary DNA structures are Z-DNA, G-quadruplexes, i-motifs, hairpins and Holliday Junctions (**Figure 1.1a-b**). [3–7] These secondary structures are the result of canonical base pairing and the chemical composition of DNA. DNA contains a phosphor group, a sugar backbone and four complementary bases: adenine (A) is complementary toward thymine (T) and cytosine (C) is complementary towards guanine (G) (**Figure 1.1c**). The sequence composition combined with secondary elements are important for the flow of genetic data: they allow for specific gene recognition by transcription factors (TFs).

TFs are key proteins involved in the central dogma of molecular biology. They are a group of proteins that can bind a specific DNA sequence and regulate DNA transcription. [8–10] Therefore, these modalities appear to challenge the central dogma



▲**Figure 1.1: Exemplary DNA structures.** a) Possible DNA double helical confirmations. From left to right: A-DNA (PDB = 4IZQ), B-DNA (PDB = 1BNA), Z-DNA (PDB = 1DCG). b) Examples of identified DNA structures. From left to right: Hairpin DNA (PDB = 1AC7), i-loop (PDB = 1A83), G-quadruplex (PDB = 1K8P), Holliday Junction (PDB = 2CRX). For clarification, a schematic representation of each structure has been incorporated next to each structure. c) Canonical base pairing first described by Watson and Crick. Adenosine and Thymine form a couple, Cytosine and Guanine form second couple.

of molecular biology, as they regulate the flow of genetic information from protein to DNA. There are approximately 1600 human TFs that are sequence specific. [10] They often bind small binding sites (binding motifs) that are 6-12 bp in size. Considering that a nine bp sequence occurs ~15.000 times in the human genome and that a sequence becomes unique when it is sixteen bps long, TFs often need to work closely together in order to provide specificity. [11–15] Additionally, DNA methylation,[16] DNA shape[17] and post translational modifications[18–20] can also have an influence on specific DNA binding as well. TFs need to be precise as unspecific binding can result in the development of disease. It is estimated that dysregulation of ~19% of all TFs is involved in a disease, emphasizing their importance as therapeutic targets. [10]

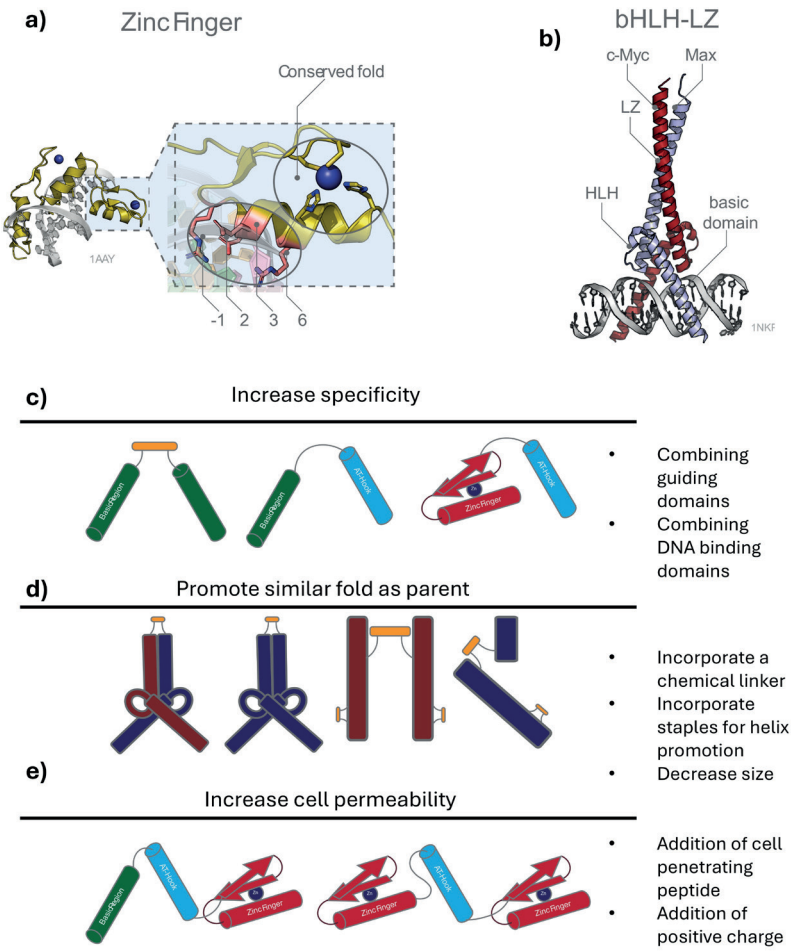
It is suggested that around 22% of the human genome are either intrinsically disordered proteins (IDPs) or have intrinsically disordered regions (IDRs). [21] Without a distinct and stabilized structure, proteins can adopt multiple conformations in solution. Moreover, with no binding pockets, it is difficult to develop small molecules that can recognize and bind this class of proteins.

TFs can contain disordered regions or can be completely disordered. Upon DNA binding, these regions undergo a conformational change and adopt an ordered structure. It is estimated that more than 80% of all TFs contain IDRs. [22] Due to their disordered nature and their location inside the cell, TFs were considered 'undruggable' for a long time. [22] As Johnson and McKnight described in their review in 1989, TFs first need to adopt a tertiary structure before they are able to bind DNA, which is similar for protein-protein interactions (PPI). [23] These structured DNA binding motifs enable another approach to target TFs: designed peptidomimetics and miniproteins based on endogenous TFs. Peptides and miniproteins can bind large surface areas and can recognize DNA or disordered proteins. Often, they are designed to target the DNA sequence of a TF. These artificial TFs therefore block the endogenous TF and thus function as a transcriptional repressor. **Chapter 2** describes an extensive overview of peptides and miniproteins designed to bind to disease-related sequences.

The aim of this thesis is to target the oncogenic TF c-Myc (from now on Myc) with inhibitors based on endogenous TFs. In **Chapter 3** and **4** miniproteins are developed based on a TF, whereas in **Chapter 5** the protein-protein interaction (PPI) between Myc and a binding partner is exploited for inhibitor design. In the following sections the main challenges for artificial TFs and peptide-based inhibitors are specified. Subsequently, Myc is described in detail to provide a better understanding on why we need to target Myc using peptide- and protein-based moieties.

1.1 Difficulties in artificial TF design

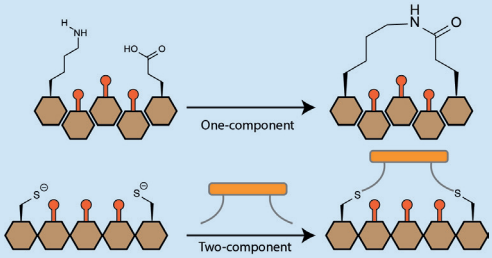
A strategy often observed in literature is to use defined DNA binding motifs of



▲**Figure 1.2: Solutions to overcome challenges in artificial TF design.** a) The crystal structure of Zinc Finger (ZF) 268. A trimeric ZF recognizes DNA inside the major groove. Each ZF contains conserved Histidines and Cysteines that coordinate around a zinc ion. This promotes the formation of one α-helix and a β-hairpin. A single ZF recognizes a triplet of DNA base pairs inside the major groove by residues at position -1, 2, 3 and 6. PDB = 1AAY. b) The crystal structure of c-Myc-Max. This heterodimeric complex adopts a basic Helix-loop-helix (bHLH) fold and utilizes the Leucine Zipper (LZ) domain to dimerize. Once dimerized this complex recognizes DNA. PDB = 1NKP. c) DBD attached via a linker and schematic representation of chimeric TFs developed by the group of Mascareñas. The GCN4 DBD is attached to the AT-Hook,[48] the GAGA ZF is attached to the AT-Hook. [48] d) Promoting similar fold as parent compound. Fully bHLH-LZ either as hetero- or as homodimers covalently attached. [49,50] bHLH-LZ (loop removed) or bZIP TFs covalently attached. [43,51] e) Combining two domains can increase specificity. Chimeric TF that contains the DBD of GCN4 that is attached to a ZF of GAGA by the AT-Hook,[52] Two GAGA ZFs are attached by the AT-Hook. [53] The AT-Hook functions as a cell penetrating peptide.

Info Box I: Stapling techniques

Peptide stapling entails covalently attachment of two residue sidechains to enforce a desired secondary structure within the peptide. This technique is often combined with SPPS. There are two strategies of stapling, one- and two-component stapling. [54] For one-component stapling, two sidechains within the peptide perform an intramolecular reaction to create a staple, whereas for a two-component staple an intermolecular reaction occurs between two sidechains and an external reagent. The latter often utilizes the reactivity of two cysteines. [46] By covalently linking two sidechains a peptide sequence is locked in a favored secondary conformation, depending on the position of the staple. For example, the Grossmann Group demonstrated that a metastasis staple promoted the formation of a β-sheet within their peptidomimetic. [55] For α-helix promotion, the side chain positions relatively towards each other are important. One α-helical turn is approximately four residues long. By incorporating a staple that spans four residues (*i, i+4*) a staple locks a single helical turn. [54] Additionally, a staple that spans two (*i, i+7*) or three (*i, i+11*) helical turns is also applied. Besides their ability to lock a peptide into a secondary formation, staples can also promote cell permeability and protease stability.



Info box Figure

Staples can be incorporated via one- or two-component stapling. The one-component strategy entails staple incorporation through an intramolecular reaction, whereas for a two-component staple an intermolecular reaction takes place that often involves cysteines.

endogenous TFs as starting point to inhibit TFs. This includes the use of Zinc Fingers (ZF, **Figure 1.2a**)[24] and motifs that fold upon and get stabilized by DNA binding, such as the Helix-Loop-Helix motif (HLH, **Figure 1.2b**). [25] A difficulty when designing artificial TFs using this strategy is loss of specificity due to the removal of domains from the parent protein. The designed TF needs to fold and act similarly compared to the parent protein. For example, basic Zipper (bZIP) TFs need to act as a dimer and need their basic region for DNA binding. [26] It was therefore envisioned that two basic regions must be linked together to target DNA as the dimerization domain was removed (**Figure 1.2c**). [27,28] To asses this hypothesis, two basic regions of GCN4, a classic bZIP TF, were covalently linked through a disulfide bond. [27] Another study used a light sensitive linker and showed that the *cis* isomer had the highest affinity. [28] Both studies demonstrated that the isolated bZIP domain has low selectivity to DNA and need a DNA guiding domain to coordinate the basic region correctly towards DNA. This was validated through the addition of a metal complex that conjugated two synthetic GCN4 basic regions together. [29,30] The metal complex could coordinate the two monomers differently relative to each other. Other strategies to retrieve the

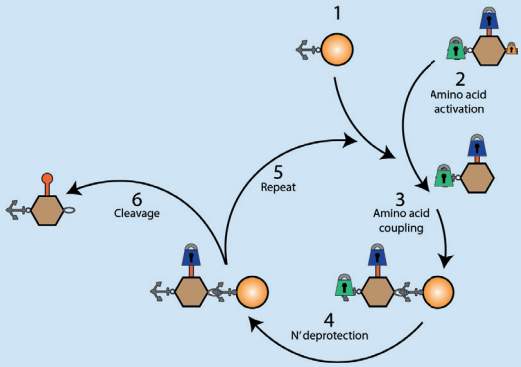
selectivity of the basic region of GCN4 entail the covalent attachment of DNA,[31] pyrrole-imidazole polyamides,[32,33] oligoguanidinium tail[34] or distamycin A[35] to the GCN4 bZIP domain. The latter resulted in a DNA binding hybrid with nanomolar affinity.

The designed peptidomimetic also needs to adopt a similar secondary structure as the parent compound. The correct structure promotes a peptide or miniprotein to recognize and bind specifically to DNA or protein partners. bZIP proteins need to form an α-helix, which normally occurs through dimerization. As a single bZIP domain can adopt multiple conformations in solution, α-helicity can be stabilized by the incorporation of a small peptide scaffold[43–45] or by connecting two residues covalently together (stapling). [41,46] Indeed, a stapled GCN4 DNA binding domain (DBD) was able to bind to DNA and enter RAW264.7 cells. [47] Instead of a two-component method where a staple is covalently attached (**See Info box I**), the incorporation of two histidines provided a non-covalent solution. [48] The addition of palladium resulted in the formation of an α-helix and DNA binding. Moreover, using this technique a reversible staple was created in such a way that DNA binding and cellular uptake could be tuned depending on the palladium concentration. Aside from structure stabilization, the incorporation of a metal also provided the possibility for self-assembly of a chimeric protein system containing a bZIP DBD and minor groove binder. [49,50] Solid phase peptide synthesis (SPPS) allows for the incorporation of unnatural amino acids that can be used for stapling (**See Info box II**). The Moellering group designed an all-carbon stapled artificial TF based on the X-box-binding protein 1 (**Figure 1.2d**). [41] Their miniprotein was able to inhibit hypoxia-dependent gene expression in vivo.

Next to structure, cell permeability is also a major challenge when designing peptide-based inhibitors. Generally, the bigger a compound, the lower the cell permeability. There are multiple techniques to increase permeability such as stapling and decreasing overall size. Additionally, it was observed that increased positive charge can enhance cell permeability. [51] Often cell-penetrating peptides are incorporated into the peptidomimetic design. For example, the AT-hook is a small positively charged AT-rich-minor groove binding motif originally occurring in HMG-I(Y). [52,53] Recent approaches involve the attachment of the AT-hook to the basic domain of GCN4,[36] two GAGA ZFs[38] or between a GCN4 and GAGA domain (**Figure 1.2e**). [37] Alone, the AT-hook has low affinity towards its target and, therefore, two low affinity DNA binders were combined with the AT-hook to increase selectivity and affinity (**Figure 1.2c**). [36] Additionally, it was observed that the AT-hook facilitated cellular penetration (**Figure 1.2e**)[36,37] and promoted selectivity. [37] Thus, the covalent attachment of a linker between DNA binding motifs could improve cellular uptake.

Info Box II: Solid phase peptide synthesis

Solid phase peptide synthesis (SPPS) is a technique that enables the straightforward assembly of peptides with virtually any residue sequence. SPPS, first developed by Merrifield in 1963,[61] applies an insoluble resin as starting point (Point 1) from which one by one a single amino acid is coupled until the final peptide or protein sequence is reached. A key aspect within SPPS is the use of orthogonal protection groups (PG) at the functional sidechain of each individual amino acid, compared to the PG at the α-amino group at the N-terminus. A standard coupling cycle is as follows: from the incoming amino acid the carboxyl is converted to an activated ester by an activating agent (e.g. Hexafluorophosphate Azabenzotriazole Tetramethyl Uronium, HATU, Point 2). In the presence of a weak base, the free amine group on the resin-bound amino acid reacts with the activated ester, thereby forming a peptide bond (Point 3). All unreacted products are then drained away and the resin is washed. The resin-bound PG at the N-terminus is removed (Point 4) so that the coupling cycle can start again (point 5) and the peptide can grow on the solid support. Finally, the peptide is cleaved from the resin (point 6). Depending on sequence composition and resin capacity, generally the size of a peptide that can be synthesized using this technique is 30-70 residues.

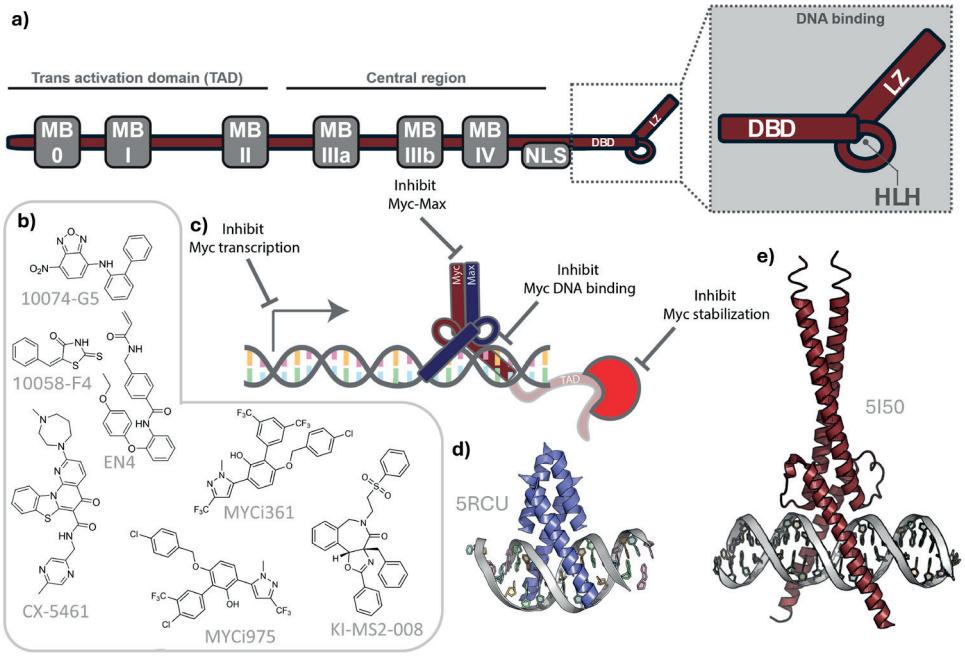


Info box Figure

The general procedure of solid phase peptide synthesis. An insoluble resin (1) is introduced that allows for the pre-activated (2) coupling of an amino acid (3). Next, the orthogonal N' terminal protection group is removed (4) that allows for the next cycle (5) or when finished cleavage of the peptide from the solid support (6).

1.2 Examples of successful ZF-based artificial TF designs

ZFs are a classic example for an artificial TF design and demonstrate the possibilities when a design is successful. They can be engineered to bind a designed sequence and moreover are intrinsic cell permeable. ZFs are small ~30 amino acids long peptides that can bind the major groove of DNA. Generally, a single ZF contains one α-helix and two β-sheets that form a β-hairpin (**Figure 1.2a**). [24] There are multiple classes of ZFs but the most common class is named after its four conserved residues that coordinate around a single zinc ion: Cys2His2. [56] This DNA binding motif contains two conserved histidines inside its α-helix and two cysteines inside its β-hairpin. Within ZFs, the helix is responsible for DNA recognition inside the major groove. Four specific residue positions are responsible for recognizing four bps in the major



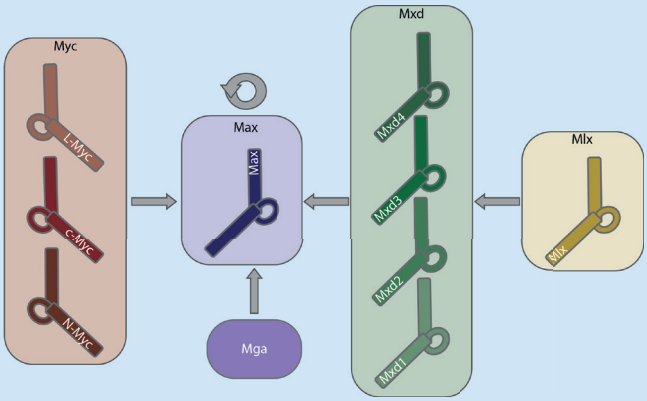
▲Figure 1.3: Myc is an oncogenic TF that can be targeted by protein modalities or small molecules. a) Myc is a largely disordered TF that has six conserved Myc Homology boxes (MB) within its Trans activation domain (TAD) and central region. The DBD domain adopts a classic bHLH-LZ fold. b) Examples of small molecules that can inhibit Myc activity. c) There are multiple ways to target Myc. Myc can be inhibited by preventing Myc transcription, stabilization, Myc-Max formation or DNA binding. d) Crystal structure of a Max-Max mimic that recognizes the E-box sequence (PDB = 5RCU). e) Crystal structure of Omomyc. Omomyc recognizes E-box DNA as a dimer (PDB = 5I50).

groove: position -1, 2, 3 and 6 (**Figure 1.2d**). To recognize specific DNA sequences, TFs often harbor multiple ZFs. Based on these findings, there has been research towards the development of artificial TFs based on ZFs that can be engineered to recognize specific sequences[24] and, moreover, can modify DNA.

A pioneer in the field of artificial ZFs was Carlos Barbas III. He envisioned that ZFs could be engineered to target every sequence. As ZFs have specific positions that are important for DNA recognition, he attempted to identify, through phage display, for each possible triplet the optimal α-helix composition. [57–60] This resulted in the identification of a specific α-helix composition for most DNA triplets and the development of Zinc Finger tools; a webserver that allows you to design ZFs based on the DNA sequence input. [62] With the possibility to design ZFs for specific DNA sequences, more complex artificial TFs could be designed. ZFs were designed to function as activators or as repressors, depending on the fused protein domain. [63]

Info box III: The Myc/Max/Mxd network

The Myc family consists of L-Myc, Myc and N-Myc. Myc belongs to the bHLH-LZ TF family and participates in the Myc/Max/Mxd network. Myc is disordered when not in complex with protein partners. It needs to interact with Max to bind DNA. Max can interact with all Mxd1-4 proteins, homodimerize and Mga. The Mxd1-4 proteins also interact with Mlx.



Info box Figure

A schematic representation of the Myc/Max/Mxd network.

Moreover, ZFs proved to be useful tools for the design of artificial TFs that can modify DNA sequences by means of methylation[64–66] or act as nucleases[67,68] or recombinases.[69]

All challenges described above are applicable on peptidomimetics and miniproteins. We needed to overcome these challenges to target the oncogenic TF Myc. Myc is a prime example of an 'undruggable TF' because it is intrinsically disordered without its protein partners and is located inside the nucleus. Moreover, it is highly involved in cancer and therefore an important pharmaceutical target.

1.3 The bHLH-LZ TF Myc

Myc is involved in important cellular processes such as differentiation, growth and death.[70–72] There are three family members in eukaryotes: Myc, N-Myc and L-Myc. Myc is involved in approximately 15% of all gene transcription.[73] Although structurally and functionally similar, these proteins differ in their expression pattern. Myc is widely expressed and the first discovered family member,[74,75] whereas N-Myc and L-Myc are mainly expressed in neural tissues and the lungs, respectively. The general structure of the Myc family is conserved between all three members. All members contain a C-terminal DNA binding and dimerization domain and an N-terminal disordered Trans-activation domain (TAD, **Figure 1.3a**). Within its disordered domain, the Myc family members have Myc homology boxes (MB) that allow other proteins to interact with Myc, regulate DNA binding and facilitate DNA

recognition (**Figure 1.3a**). [76]

Myc belongs to the HLH TF family. This structure contains two α-helices connected by a flexible loop (**Figure 1.2b**). [23,77] Often HLH proteins also contain a basic region (b) that allow them to bind to DNA. Additionally, HLH TFs need to form a dimeric complex to function and bind to DNA. This can be a hetero- or homodimeric complex. Myc has a leucine zipper (LZ) dimerization region. However, it cannot homodimerize and needs to interact with Myc-associated factor X (Max) to bind to DNA. Both Myc and Max are part of the Myc/Max/Mxd network (**See box III**). Other members of this protein network are Max- binding protein (MNT), Max-like protein X (Mlx) and Max gene-associated (Mga). Within this network, Myc can only interact with Max, whereas Max can homodimerize or interact with Mga and Mxd1-4. The Myc-Max complex recognizes Enhancer box (E-box) sequences (CACGTG) and when bound regulate downstream pathways.

Myc contains six MBs which mediate interactions with proteins partners (**Figure 1.3a**). [71] For example, Myc recruits Transactivation/Transformation-Domain Associated Protein (TRRAP) with MBII for histone acetylation and it interacts with Histone Deacetylase 3 (HDAC3) through its MBIIIa domain to promote histone deacetylation. [71,78] These PPI examples highlight why Myc is strictly regulated as it controls through chromatin remodeling a multitude of cellular processes via two contradictory protein partners.

Myc is regulated on two levels: on the mRNA and on the protein level. Both Myc RNA and the protein have short half-life's of approximately 10-20 min and 20 min, respectively. [79] Myc transcription is controlled by secondary DNA structures, including G-quadruplexes and Z-DNA. [70,79] A variety of TFs regulate Myc transcription by binding to Myc enhancer sites. [72,79,80] For example, Bromodomain protein 4 (BRD4) interacts with enhancer regions and stimulates increased Myc expression. [72,80] It was identified that JQ1, a BRD4 inhibitor, led to lower Myc expression levels. [80]

Myc is also regulated through phosphorylation[81] and protein partners. [79] Depending on phosphorylation patterns and cascades, Myc can either be stabilized or it can be marked for ubiquitination and therefore degradation. [81] Other proteins can stabilize Myc by interacting with its MBs, such as Aurora kinase A. [79] Another layer on Myc regulation is through Max. Max is essential for Myc to act as a transcriptional regulator. [70,71,82] By homodimerization or interacting with other members of the Myc/Max/Mxd network, Max monomers are sequestered away from Myc.

A major function of Myc is to sustain cellular proliferation by preventing cell cycle arrest and thereby cellular differentiation. Additionally, Myc is involved in diverse biological processes, such as metabolic reprogramming and angiogenesis. [72] Myc can function as a transcription amplifier that can bind to promotors close

to Transcriptional Start Sites. Contrarily, when interacting with Myc Interacting Zinc Finger 1 (Miz-1), Myc acts a transcriptional repressor, thereby inhibiting cell differentiation and apoptosis. [83,84] It is anticipated that 25-40% of genes repressed by Myc involve Miz-1. [83,85] When bound to Miz-1, Myc can repress genes, such as *p21*^{Cip1} and *p15*^{Ink4b}, that are involved in cell cycle progression.

Dysregulation of Myc can lead to the development of cancer. [79] Strikingly, Myc engages in more than 50% of all human cancers and moreover, cancer cells become addicted and heavily dependent on Myc activity. Thus, highlighting the importance of Myc inhibition as a pharmaceutical strategy. Under physiological levels, Myc binds to high-affinity promotors guided by PPIs that stabilize Myc-Max-DNA binding by interacting with MBs in the N-terminus of Myc (e.g. WDR5 binds MBIII). [86,87] In Myc driven cancers, Myc is highly overexpressed allowing it to occupy low-affinity promotors as well, which is called promoter invasion. [86] This can eventually lead to uncontrolled cell division. Exemplary cancer types with Myc involvement are lung, breast and brain cancer. [88]

Taken together, it is extremely difficult to inhibit Myc activity. Its complex regulation system, cellular location, involvement in PPIs and its disordered nature make it a challenging pharmaceutical target. This also explains why there is no direct Myc inhibitor on the clinical market while writing this thesis. However, there are attempts to target Myc using small molecules and peptidomimetics.

1.3.1 How to target Myc

Myc was originally deemed 'undruggable'. Improved insight into Myc function facilitated the development of novel approaches to suppress Myc activity. Although Myc is intrinsically disordered, there are small molecules and peptide/protein-based inhibitors developed that can target Myc. Two strategies are applied to inhibit Myc, either direct or indirect inhibition. Additionally, there are also inhibitors dubbed double negative inhibitors as they can inhibit Myc by direct and indirect inhibition. The following section describes examples how Myc is targeted using small molecules and peptide mimetics. The following reviews provide a more extensive overview on Myc inhibitors. [70,71]

1.3.2 Myc inhibition with small molecules

Myc gene transcription is regulated through secondary DNA elements that can be targeted. Within its promotor, Myc contains a G-quadruplex that functions as a gene silencer. [92,93] The G-quadruplex is untwined to activate Myc transcription. A Myc inhibition strategy is to stabilize the G-quadruplex in the Myc promotor and thereby prevent Myc transcription. [70] Small molecules have been identified that can stabilize the Myc G-quadruplex (**Figure 1.3b-c**). [92–96] One of these compounds, CX - 5461

(**Figure 1.3b**), is currently in clinical trials for BRCA-deficient breast cancer. [97] Additionally, peptides have been identified to stabilize the Myc G-quadruplex. [98,99] Other strategies entail the inhibition of upstream proteins of Myc, such as BRD4,[80] the interaction with Myc stabilizing proteins, such as Aurora A,[100] or the development of Proteolysis Targeting Chimeras (PROTACs). [101–104]

Small molecules have also been identified that target the Myc-Max interaction (**Figure 1.3c**). [71] Nomura *et al.* recently developed EN4 (**Figure 1.3b**). [105] EN4 is a covalent ligand that can target cysteine171 in Myc, whereas it is suggested that it is unable to interact with N- or L-Myc as this cysteine is not conserved among Myc family members. It was suggested that EN4 destabilizes Myc, which renders it unable to interact with Max. Han *et al.* developed small molecules that could inhibit Myc in cells by Myc-Max inhibition, hampering Myc expression and promoting Myc degradation through phosphorylation. [106] Moreover, their lead compound, MYCi361 and its improved version MYCi975 (**Figure 1.3b**), could suppress tumor growth in mice. Struntz *et al.* identified KI-MS2-008 that could stabilize the Max-Max homodimer, while simultaneously reducing both expression levels of Myc and proteins regulated by Myc (**Figure 1.3b**). [107] Other examples of small molecules that can prevent Myc-Max formation are 10058-F4 and 10074-G5 (**Figure 1.3b**). [108,109] These compound could bind Myc while not properly folded and prevented it from refolding to a bHLH-LZ fold with Max.

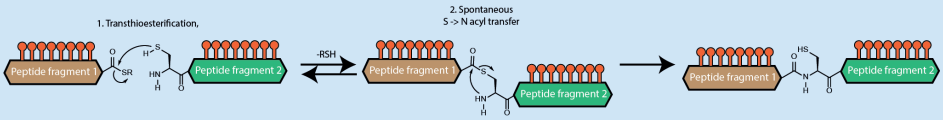
1.3.3 Myc inhibition with peptido- or protein mimetics

Transcription repressors have been designed to inhibit Myc activity based on endogenous Myc inhibitors (**Figure 1.3c-e**). Recently, Speltz *et al.* developed a transcription repressor based on Max that displayed Myc inhibition (**Figure 1.2d**, **1.3d**). [42] The flexible loop was removed from the Max sequence and the Zipper domain was decreased in size. The DBD and the remaining part of the Zipper region were covalently linked with a maleimide linker, thereby replacing the flexible loop. Additionally, the miniprotein ME47[110,111] and recombinant Mad1,[112] were reported to inhibit Myc activity as well. ME47 is a chimeric protein that contains the DBD of Max fused to the dimerization domain of E47, whereas Mad1 contains the first 147 residues of Mxd1. Other strategies entail peptides that are based on the bHLH-LZ domain of Max,[113] the LZ domain of Max,[114] covalently linked Myc-Max or Max DBDs[115,116] or covalently linked bHLH-LZ region of Myc-Max and Max-Max dimers (**Figure 1.2b**)[39,40]

One of the first attempts to use bHLH-LZ proteins as a template was based on the Myc-Max heterodimer by the Kent group in 1995. [117] Using Native Chemical Ligation (NCL, see **Info Box IV**), Myc and Max monomers were covalently attached, thereby forming the heterodimer. Firstly, through thioester formation the C-terminal HLH-LZ region was attached to the N-terminal basic region to generate the monomers.

Info Box IV: Native chemical ligation

As proteins often contain more than 70 amino acids that cannot be synthesized with SPPS, a different approach has been developed: native chemical ligation (NCL). [91] Here, the C-terminal thioester of fragment 1 reacts with the N-terminal cysteine of fragment 2 (point 1), which is followed by a spontaneous intramolecular S-to-N acyl shift (point 2). This entails that the α-amino group of the cysteine attacks the thioester, forming an amide bond and releasing the thiol group. After its development, this technique has been applied to the synthesis of proteins including for example Max and allows for the incorporation of post translational modification that can be studied in an in vitro context. [92,93]



Info box Figure

Native chemical ligation (NCL) is often applied for protein synthesis. First, a transthioesterfication occurs (1) followed by a spontaneous S->N acyl transfer shift (2) leading to the release of the thiol group, thereby retrieving the original cysteine.

Secondly, an oxime reaction provided the covalent attachment of both monomers that resulted in the heterodimer. Additionally, a one-pot thioester and oxime formation yielded in the formation of a Max-Max homodimer. Krylov *et al.*,[118] proposed a different strategy. This included the removal of the TAD and exchange of basic regions that would prevent correct DNA binding. Overexpression could then inhibit Myc-Max activity. A similar strategy was applied by Soucek *et al.* to develop Omomyc. [119]

Omomyc is based on Myc with four point mutations in its LZ domain that enables homodimerization (**Figure 1.3e**). [119] Omomyc can interact with Myc, Max and Miz-1, but is unable to interact with Mxd proteins. [120] The crystal structure of Omomyc displayed a DNA recognition triad consisting of a His12, Glu16 and Arg20. [86] Strategies to analyze Omomyc entail Omomyc expression in mammalian cells,[119] recombinant expression and purification,[86] flow chemistry[39,40] and NCL. [121] Strikingly, Omomyc has the strongest affinity for the E-box sequence as a homodimer and is unable to bind to E-box DNA when bound to Myc.[39] It is able to inhibit Myc activity when expressed inside cells[86,122] and it is claimed that it has cell penetrating properties. [123] There are, however, also reports showing that Omomyc is unable to enter cells and that a cell penetrating peptide is beneficial to assist cell permeability. [124] Jung *et al.* proposed that Omomyc mainly competes with Myc-Max for low-affinity promotors as the removal of the N-terminus renders it unable to interact with stabilizing PPIs. [86] By competing with Myc at low-affinity promotors, Omomyc counteracts aberrant expression behavior of Myc. Currently, Omomyc-103 is the first direct Myc inhibitor in clinical trials. [125]

Info Box V: AlphaFold

New and powerful AI methods have been developed to predict how proteins fold. This led to the development of AlphaFold2 by Deepmind. [127] AlphaFold is a deep-learning method that predicts how proteins fold based on their primary sequence. AlphaFold is built around a deep neural network that is trained on evolutionary, physical and geometric data of protein structures. [127] It contains two main stages. The first stage includes Evoformer, which is a novel architecture that operates on two datatypes: multiple sequence alignments (MSA) and pair representations. After multiple rounds of Evoformer processing, the obtained data is interpreted by the second stage that translates the Evoformer data into a 3D structure. The output of AlphaFold2 is an end-to-end structure prediction with additional prediction scores for each individual residue in the protein. AlphaFold has been applied to the whole human genome,[129] including IDPs through fragmentation[130] and PPIs. [131]

Myc is intrinsically disordered which hinders its characterization and crystallization. Additionally, IDPs and IDRs do not adopt defined binding pockets, which impedes the use of small molecules. Therefore, another approach is pursued to target disordered proteins: the use of deep-learning tools such as RFdiffusion[126] and AlphaFold-Multimer. [127] For example, recently the Baker lab used RFdiffusion to design IDR binding proteins. [128] In the absence of a crystal structure we used AlphaFold-Multimer[127] to predict how Myc and Miz-1 interact. This deep-learning tool predicts how proteins fold based on their primary sequence without the need for a crystal structure (**See Info Box V**). Based on the prediction we could design Miz-1 based Myc inhibitors.

1.4 Aims and outline of this thesis

Despite being involved in more than 50% of human cancers, there are no direct Myc inhibitors on the clinical market. Through substantial amount of process over the past decades we got a better understanding of Myc activity and Myc involvement in cancer. Because Myc is both an IDP and TF it becomes more difficult to inhibit. An increase in knowledge enabled scientists to develop inhibitors towards Myc via multiple strategies. Additionally, the development of AlphaFold2 revolutionized the field in protein structures. This thesis explores the field TF-based inhibitors via SPPS and protein expression. Additionally, it utilizes AlphaFold2 to design novel peptidomimetics based on a more than decade old unresolved PPI between Myc and Miz-1. The aim of this thesis is to design TF-based peptide modalities that can target Myc and validate these compounds through biochemical assays.

Myc is not the only disease related TF. Therefore, in **Chapter 2** a comprehensive overview is provided that entails synthetic peptide modalities able to tackle disease related DNA or RNA. **Chapter 3** describes the development of DuoMYC. Here, we

performed a structure-based study based on the Omomyc sequence that led to the design of a smaller bHLH miniprotein that functioned as an artificial transcriptional repressor and could inhibit Myc activity in vitro. In **Chapter 4** we explored the possibilities of expressed protein stapling to improve stability and cell permeability of Omomyc. A protocol was developed to express proteins that could be stapled, while remaining active. In **Chapter 5** a different approach was pursued to inhibit Myc. Whereas previous chapters explore the ability to inhibit Myc through artificial transcriptional repressors, in this chapter we use AlphaFold2 to predict how Myc and Miz-1 interact and used this prediction to develop peptidomimetics that can inhibit Myc-Max complex formation. **Chapter 6** combines all the findings of this thesis and describes the most important conclusions and future perspectives.

1.5. References

[1] F. Crick, "Central Dogma of Molecular Biology" *Nature* **1970**, *227*, 1970.

[2] J. D. WATSON, F. H. C. CRICK, "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid" *Nature* **1953**, *171*, 737–738.

[3] R. C. Duardo, F. Guerra, S. Pepe, G. Capranico, "Non-B DNA structures as a booster of genome instability" *Biochimie* **2023**, *214*, 176–192.

[4] S. Neidle, "Beyond the double helix: DNA structural diversity and the PDB" *Journal of Biological Chemistry* **2021**, *296*, 100553.

[5] A. Bansal, S. Kaushik, S. Kukreti, "Non-canonical DNA structures: Diversity and disease association" *Front Genet* **2022**, *13*, 959258.

[6] J. Spiegel, S. M. Cuesta, S. Adhikari, R. Hänsel-Hertsch, D. Tannahill, S. Balasubramanian, "G-quadruplexes are transcription factor binding hubs in human chromatin" *Genome Biol* **2021**, *22*, 1–15.

[7] V. J. Sahayasheela, M. Ooga, T. Kumagai, H. Sugiyama, "Z-DNA at the crossroads: untangling its role in genome dynamics" *Trends Biochem Sci* **2025**, *50*, 267–279.

[8] D. L. Fulton, S. Sundararajan, G. Badis, T. R. Hughes, W. W. Wasserman, J. C. Roach, R. Sladek, "TFCat: The curated catalog of mouse and human transcription factors" *Genome Biol* **2009**, *10*, 1–14.

[9] J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, N. M. Luscombe, "A census of human transcription factors: Function, expression and evolution" *Nat Rev Genet* **2009**, *10*, 252–263.

[10] 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, "The Human Transcription Factors" *Cell* **2018**, *172*, 650–665.

[11] M. Moore, A. Klug, Y. Choo, "Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units" *Proc Natl Acad Sci U S A* **2001**, *98*, 1437–1441.

[12] Q. Liu, D. J. Segal, J. B. Ghiara, C. F. Barbas, "Design of polydactyl zinc-finger proteins for unique addressing within complex genomes" *Proc Natl Acad Sci U S A* **1997**, *94*, 5525–5530.

[13] Z. Wunderlich, L. A. Mirny, "Different gene regulation strategies revealed by analysis of binding motifs" *Trends in Genetics* **2009**, *25*, 434–440.

[14] A. Jolma, Y. Yin, K. R. Nitta, K. Dave, A. Popov, M. Taipale, M. Enge, T. Kivioja, E. Morgunova, J. Taipale, "DNA-dependent formation of transcription factor pairs alters their binding specificity" *Nature 2015 527:7578* **2015**, *527*, 384–388.

[15] H. Göös, M. Kinnunen, K. Salokas, Z. Tan, X. Liu, L. Yadav, Q. Zhang, G. H. Wei, M. Varjosalo, "Human transcription factor protein interaction networks" *Nature Communications 2022 13:1* **2022**, *13*, 1–16.

[16] Y. Yin, E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schübeler, C. Vinson, J. Taipale, "Impact of cytosine methylation on DNA binding specificities of human transcription factors" *Science (1979)* **2017**, *356*, DOI 10.1126/SCIENCE.AAJ2239/ASSET/426CEDB2-D8EF-46C9-928A-E3A7CC62CCD2/ASSETS/GRAPHIC/356_AAJ2239_F7.JPEG.

[17] S. Inukai, K. H. Kock, M. L. Bulyk, "Transcription factor–DNA binding: beyond binding site motifs" *Curr Opin Genet Dev* **2017**, *43*, 110–119.

[18] R. V. Nithun, Y. M. Yao, X. Lin, S. Habiballah, A. Afek, M. Jbara, "Deciphering the Role of the Ser-Phosphorylation Pattern on the DNA-Binding Activity of Max Transcription Factor Using Chemical Protein Synthesis" *Angewandte Chemie - International Edition* **2023**, *62*, DOI 10.1002/anie.202310913.

[19] T. M. Filtz, W. K. Vogel, M. Leid, "Regulation of transcription factor activity by interconnected post-translational modifications" *Trends Pharmacol Sci* **2014**, *35*, 76–85.

[20] L. Seymour, N. Nuru, K. R. Johnson, J. M. V. Gutierrez, V. T. Njoku, C. C. Darie, A. N. Neagu, "Roles of Post-Translational Modifications of Transcription Factors Involved in Breast Cancer Hypoxia" *Molecules 2025, Vol. 30, Page 645* **2025**, *30*, 645.

[21] K. Kosoglu, Z. Aydin, N. Tuncbag, A. Gursoy, O. Keskin, "Structural coverage of the human interactome." *Brief Bioinform* **2023**, *25*, 1–11.

[22] M. J. Henley, A. N. Koehler, "Advances in targeting 'undruggable' transcription factors with small molecules" *Nat Rev Drug Discov* **2021**, *20*, 669–688.

[23] P. F. Johnson, S. L. McKnight, "EUKARYOTIC TRANSCRIPTIONAL REGULATORY PROTEINS" *Annu Rev Biochem* **1989**, *58*, 799–839.

[24] C. A. Gersbach, T. Gaj, C. F. Barbas, "Synthetic Zinc Finger Proteins: The Advent of Targeted Gene Regulation and Genome Modification Technologies" *Acc Chem Res* **2014**, *47*, 2309–2318.

[25] M. Kuravsky, C. Kelly, C. Redfield, S. L. Shammas, "The transition state for coupled folding and binding of a disordered DNA binding domain resembles the unbound state" *Nucleic Acids Res* **2024**, *52*, 11822–11837.

[26] J. J. Kohler, S. J. Metallo, T. L. Schneider, A. Schepartz, "DNA specificity enhanced by sequential binding of protein monomers" *Proceedings of the National Academy of Sciences* **1999**, *96*, 11735–11739.

[27] R. V. Talanian, C. J. Mcknight, P. S. Kim, "Sequence-Specific DNA Binding by a Short Peptide Dimer" *Science (1979)* **1990**, *249*, 769–771.

[28] A. M. Caamaño, M. E. Vázquez, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, "A Light-Modulated Sequence-Specific DNA-Binding Peptide" *Angewandte Chemie* **2000**, *39*, 3104–3107.

[29] B. Cuenoud, A. Schepartz, "Design of a metallo-bZIP protein that discriminates between CRE and AP1 target sites: selection against AP1." *Proceedings of the National Academy of Sciences* **1993**, *90*, 1154–1159.

[30] B. Cuenoud, A. Schepartz, "Altered Specificity of DNA-Binding Proteins with Transition Metal Dimerization Domains" *Science (1979)* **1993**, *259*, 510–513.

[31] D. Stanojevic, G. L. Verdine, "Deconstruction of GCN4/GCRE into a monomeric peptide-DNA complex" *Nat Struct Mol Biol* **1995**, *2*, 450–457.

[32] J. B. Blanco, M. E. Vázquez, L. Castedo, J. L. Mascareñas, "Sequence-Specific DNA Recognition by Monomeric bZIP Basic Regions Equipped with a Tripyrrole Unit on the N-Terminal Side. Towards the Development of Synthetic Mimics of Skn-1" *ChemBioChem* **2005**, *6*, 2173–2176.

[33] J. B. Blanco, V. I. Dodero, M. E. Vázquez, M. Mosquera, L. Castedo, J. L. Mascareñas, "Sequence-Specific DNA Binding by Noncovalent Peptide–Tripyrrole Conjugates" *Angewandte Chemie International Edition* **2006**, *45*, 8210–8214.

[34] J. Mosquera, M. I. Sánchez, J. Valero, J. de Mendoza, M. E. Vázquez, J. L. Mascareñas, "Sequence-selective DNA binding with cell-permeable oligoguanidinium–peptide conjugates" *Chemical Communications* **2015**, *51*, 4811–4814.

[35] M. E. Vázquez, A. M. Caamaño, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, "Design and Synthesis of a Peptide That Binds Specific DNA Sequences through Simultaneous Interaction in the Major and in the Minor Groove" *Angewandte Chemie International Edition* **2001**, *40*, 4723–4725.

[36] J. Rodríguez, J. Mosquera, J. R. Couceiro, M. E. Vázquez, J. L. Mascareñas, "The AT-Hook motif as a versatile minor groove anchor for promoting DNA binding of transcription factor fragments" *Chem Sci* **2015**, *6*, 4767–4771.

[37] J. Rodríguez, J. Mosquera, R. García-Fandiño, M. E. Vázquez, J. L. Mascareñas, "A designed DNA binding motif that recognizes extended sites and spans two adjacent major grooves" *Chem Sci* **2016**, *7*, 3298–3303.

[38] J. Rodríguez, S. Learte-Aymamí, J. Mosquera, G. Celaya, D. Rodríguez-Larrea, M. E. Vázquez, J. L. Mascareñas, "DNA-binding miniproteins based on zinc fingers. Assessment of the interaction using nanopores" *Chem Sci* **2018**, *9*, 4118–4123.

[39] M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, "Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents." *J Am Chem Soc* **2021**, *143*, 11788–11798.

[40] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow

Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription." *ACS Cent Sci* **2021**, *7*, 1408–1418.

[41] Z. Qiao, L. C. Nguyen, D. Yang, C. Dann, D. M. Thomas, M. Henn, A. Valdespino, C. S. Swenson, S. A. Oakes, M. R. Rosner, R. E. Moellering, "Direct inhibition of tumor hypoxia response with synthetic transcriptional repressors." *Nat Chem Biol* **2025**, *21*, 247–255.

[42] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, "Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains" *Nature Biotechnology 2022* **2022**, *41*, 1–11.

[43] L. Yang, A. Schepartz, "Relationship between folding and function in a sequence-specific miniature DNA-binding protein" *Biochemistry* **2005**, *44*, 7469–7478.

[44] N. J. Zondlo, A. Schepartz, "Highly Specific DNA Recognition by a Designed Miniature Protein" *J Am Chem Soc* **1999**, *121*, 6938–6939.

[45] J. W. Chin, A. Schepartz, "Concerted evolution of structure and function in a miniature protein" *J Am Chem Soc* **2001**, *123*, 2929–2930.

[46] H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. DeGrado, D. C. Greenbaum, "Development of α-helical calpain probes by mimicking a natural protein-protein interaction." *J Am Chem Soc* **2012**, *134*, 17704–17713.

[47] A. Iyer, D. Van Lysebetten, Y. Ruiz García, B. Louage, B. G. De Geest, A. Madder, "Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake" *Org Biomol Chem* **2015**, *13*, 3856–3862.

[48] S. Learte-Aymamí, N. Curado, J. Rodríguez, M. E. Vázquez, J. L. Mascareñas, "Metal-Dependent DNA Recognition and Cell Internalization of Designed, Basic Peptides" *J Am Chem Soc* **2017**, *139*, 16188–16193.

[49] S. Learte-Aymamí, J. Rodríguez, M. E. Vázquez, J. L. Mascareñas, "Assembly of a Ternary Metallopeptide Complex at Specific DNA Sites Mediated by an AT-Hook Adaptor" *Chemistry – A European Journal* **2020**, *26*, 8875–8878.

[50] M. I. Sánchez, J. Mosquera, M. E. Vázquez, J. L. Mascareñas, "Reversible Supramolecular Assembly at Specific DNA Sites: Nickel-Promoted Bivalent DNA Binding with Designed Peptide and Bipyridyl-Bis(benzamidine) Components" *Angewandte Chemie International Edition* **2014**, *53*, 9917–9921.

[51] N. Schmidt, A. Mishra, G. H. Lai, G. C. L. Wong, **2010**, DOI: 10.1016/j.febslet.2009.11.046.

[52] L. Aravind, "AT-hook motifs identified in a wide variety of DNA-binding proteins" *Nucleic Acids Res* **1998**, *26*, 4413–4421.

[53] T. Lund, K. H. Dahl, E. Mørk, J. Holtlund, S. G. Laland, "The human chromosomal protein HMG I contains two identical palindrome amino acid sequences" *Biochem Biophys Res Commun* **1987**, *146*, 725–730.

[54] Y. H. Lau, P. De Andrade, Y. Wu, D. R. Spring, "Peptide stapling techniques based on different macrocyclisation chemistries" *Chem Soc Rev* **2015**, *44*, 91–102.

[55] M. Wendt, R. Bellavita, A. Gerber, N. L. Efrém, T. van Ramshorst, N. M. Pearce, P. R. J. Davey, I. Everard, M. Vazquez-Chantada, E. Chiarparin, P. Grieco, S. Hennig, T. N. Grossmann, "Bicyclic β-Sheet Mimetics that Target the Transcriptional Coactivator β-Catenin and Inhibit Wnt Signaling" *Angewandte Chemie - International Edition* **2021**, *60*, 13937–13944.

[56] A. A. Fedotova, A. N. Bonchuk, V. A. Mogila, P. G. Georgiev, **2017**, Russian Federation Agency for Science and Innovation preprint, DOI: 10.32607/20758251-2017-9-2-47-58.

[57] D. J. Segal, B. Dreier, R. R. Beerli, C. F. Barbas, "Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences" *Proc Natl Acad Sci U S A* **1999**, *96*, 2758–2763.

[58] B. Dreier, R. P. Fuller, D. J. Segal, C. V. Lund, P. Blancafort, A. Huber, B. Koksch, C. F. Barbas, "Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors" *Journal of Biological Chemistry* **2005**, *280*, 35588–35597.

[59] B. Dreier, R. R. Beerli, D. J. Segal, J. D. Flippin, C. F. Barbas, "Development of Zinc Finger Domains for Recognition of the 5′-ANN-3′ Family of DNA Sequences and Their Use in the Construction of Artificial Transcription Factors" *Journal of Biological Chemistry* **2001**, *276*, 29466–29478.

[60] B. Dreier, D. J. Segal, C. F. Barbas, "Insights into the molecular recognition of the 5′-GNN-3′ family of DNA sequences by zinc finger domains" *J Mol Biol* **2000**, *303*, 489–502.

[61] R. B. Merrifield, "Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide" *J Am Chem Soc* **1963**, *85*, 2149–2154.

[62] J. G. Mandell, C. F. Barbas, "Zinc Finger Tools: Custom DNA-binding domains for transcription factors and nucleases" *Nucleic Acids Res* **2006**, *34*, W516.

[63] R. R. Beerli, D. J. Segal, B. Dreier, C. F. Barbas, "Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks" *Proc Natl Acad Sci U S A* **1998**, *95*, 14628–14633.

[64] W. Nomura, C. F. Barbas, "In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase" *J Am Chem Soc* **2007**, *129*, 8676–8677.

[65] X. Li, Z. Wang, J. Huang, H. Luo, S. Zhu, H. Yi, L. Zheng, B. Hu, L. Yu, L. Li, J. Xie, N. Zhu, "Specific zinc finger-induced methylation of PD-L1 promoter inhibits its expression" *FEBS Open Bio* **2019**, *9*, 1063–1070.

[66] A. G. Rivenbark, S. Stolzenburg, A. S. Beltran, X. Yuan, M. G. Rots, B. D. Strahl, P. Blancafort, "Epigenetic reprogramming of cancer cells via targeted DNA methylation" *Epigenetics* **2012**, *7*, 350–360.

[67] T. Gaj, J. Guo, Y. Kato, S. J. Sirk, C. F. Barbas, "Targeted gene knockout by direct delivery of zinc-finger nuclease proteins" *Nat Methods* **2012**, *9*, 805–807.

[68] S. Katayama, M. Watanabe, Y. Kato, W. Nomura, T. Yamamoto, "Engineering of Zinc Finger Nucleases Through Structural Modeling Improves Genome Editing Efficiency in Cells" *Advanced Science* **2024**, *11*, DOI 10.1002/advs.202310255.

[69] L. Mukhametzyanova, L. T. Schmitt, J. Torres-Rivera, T. Rojo-Romanos, F. Lansing, M. Paszkowski-Rogacz, H. Hollak, M. Brux, M. Augsburg, P. M. Schneider, F. Buchholz, "Activation of recombinases at specific DNA loci by zinc-finger domain insertions" *Nature Biotechnology 2024 42:12* **2024**, *42*, 1844–1854.

[70] Z. Deng, D. R. Richardson, "The Myc Family and the Metastasis Suppressor NDRG1: Targeting Key Molecular Interactions with Innovative Therapeutics" *Pharmacol Rev* **2023**, *75*, 1007–1035.

[71] S. K. Madden, A. D. de Araujo, M. Gerhardt, D. P. Fairlie, J. M. Mason, "Taking the Myc out of cancer: toward therapeutic strategies to directly inhibit c-Myc" *Molecular Cancer 2021 20:1* **2021**, *20*, 1–18.

[72] H. M. Hylton-McComas, A. Cordes, K. V. Floros, A. Faber, B. J. Drapkin, W. O. Miles, "Myc family proteins: Molecular drivers of tumorigenesis and resistance in neuroendocrine tumors" *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **2025**, *1880*, 189332.

[73] C. V. Dang, K. A. O'Donnell, K. I. Zeller, T. Nguyen, R. C. Osthus, F. Li, "The c-Myc target gene network" *Semin Cancer Biol* **2006**, *16*, 253–264.

[74] M. D. Cole, "The myc oncogene: its role in transformation and differentiation." *Annu Rev Genet* 1986, *20*, 361–84.

[75] B. Vennstrom,, D. Sheiness, J. Zabielski, A. J. M. Bishop2, "Isolation and characterization of c-myc, a cellular homolog of the oncogene (v-myc) of avian myelocytomatosis virus strain 29" *J Virol* **1982**, *42*, 773–779.

[76] S. Schütz, C. Bergsdorf, S. Hänni-Holzinger, A. Lingel, M. Renatus, A. D. Gossert, W. Jahnke, "Intrinsically Disordered Regions in the Transcription Factor MYC:MAX Modulate DNA Binding via Intramolecular Interactions" *Biochemistry* **2024**, *63*, 511.

[77] C. Murre, "Helix–loop–helix proteins and the advent of cellular diversity: 30 years of discovery" *Genes Dev* **2019**, *33*, 6–25.

[78] Y. Zhou, X. Gao, M. Yuan, B. Yang, Q. He, J. Cao, "Targeting Myc Interacting Proteins as a Winding Path in Cancer Therapy" *Front Pharmacol* **2021**, *12*, 748852.

[79] R. K. Jha, F. Kouzine, D. Levens, "MYC function and regulation in physiological perspective" *Front Cell Dev Biol* **2023**, *11*, 1268275.

[80] J. E. Delmore, G. C. Issa, M. E. Lemieux, P. B. Rahl, J. Shi, H. M. Jacobs, E. Kastritis, T. Gilpatrick, R. M. Paranal, J. Qi, M. Chesi, A. C. Schinzel, M. R. McKeown, T. P. Heffernan, C. R. Vakoc, P. L. Bergsagel, I. M. Ghobrial, P. G. Richardson, R. A. Young, W. C. Hahn, K. C. Anderson, A. L. Kung, J. E. Bradner, C. S. Mitsiades, "BET Bromodomain Inhibition as a Therapeutic Strategy to Target c-Myc" *Cell* **2011**, *146*, 904–917.

[81] X. X. Sun, Y. Li, R. C. Sears, M. S. Dai, "Targeting the MYC Ubiquitination-Proteasome Degradation Pathway for Cancer Therapy" *Front Oncol* **2021**, *11*, 679445.

[82] N. Meyer, L. Z. Penn, "Reflecting on 25 years with MYC" *Nature Reviews Cancer 2008 8:12* **2008**, *8*, 976–990.

[83] L. Zhang, J. Li, H. Xu, X. Shao, L. Fu, Y. Hou, C. Hao, W. Li, K. Joshi, W. Wei, Y. Xu, F. Zhang, S. Dai, P. Breslin, J. Zhang, J. Zhang, "Myc-Miz1 signaling promotes self-renewal of leukemia stem cells by repressing Cebpα and Cebpδ" *Blood* **2020**, *135*, 1133–1145.

[84] T. Möröy, I. Saba, C. Kosan, "The role of the transcription factor Miz-1 in lymphocyte development and lymphomagenesis-Binding Myc makes the difference" *Semin Immunol* **2011**, *23*, 379–387.

[85] N. Varlakhanova, R. Cotterman, K. Bradnam, I. Korf, P. S. Knoepfler, "Myc and Miz-1 have coordinate genomic functions including targeting Hox genes in human embryonic stem cells" *Epigenetics Chromatin* **2011**, *4*, 1–14.

[86] L. A. Jung, A. Gebhardt, W. Koelmel, C. P. Ade, S. Walz, J. Kuper, B. Von Eyss, S. Letschert, C. Redel, L. D'Artista, A. Biankin, L. Zender, M. Sauer, E. Wolf, G. Evan, C. Kisker, M. Eilers, "OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors" *Oncogene 2017 36:14* **2016**, *36*, 1911–1924.

[87] F. Lorenzin, U. Benary, A. Baluapuri, S. Walz, L. A. Jung, B. von Eyss, C. Kisker, J. Wolf, M. Eilers, E. Wolf, "Different promoter affinities account for specificity in MYC-dependent gene regulation" *Elife* **2016**, *5*, DOI 10.7554/ELIFE.15161.

[88] A. Sorolla, E. Wang, E. Golden, C. Duffy, S. T. Henriques, A. D. Redfern, P. Blancafort, "Precision medicine by designer interference peptides: applications in oncology and molecular therapeutics" *Oncogene* **2020**, *39*, 1167–1184.

[89] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, "Synthesis of Proteins by Native Chemical Ligation" *Science (1979)* **1994**, *266*, 776–779.

[90] O. Harel, M. Jbara, "Chemical Synthesis of Bioactive Proteins" *Angewandte Chemie* **2023**, *135*, DOI 10.1002/ange.202217716.

[91] R. V. Nithun, Y. M. Yao, O. Harel, S. Habiballah, A. Afek, M. Jbara, "Site-Specific Acetylation of the Transcription Factor Protein Max Modulates Its DNA Binding Activity" *ACS Cent Sci* **2024**, *2024*, 1295–1303.

[92] D. R. Calabrese, X. Chen, E. C. Leon, S. M. Gaikwad, Z. Phyo, W. M. Hewitt, S. Alden, T. A. Hilimire, F. He, A. M. Michalowski, J. K. Simmons, L. B. Saunders, S. Zhang, D. Connors, K. J. Walters, B. A. Mock, J. S. Schneekloth, "Chemical and structural studies provide a mechanistic basis for recognition of the MYC G-quadruplex" *Nature Communications 2018 9:1* **2018**, *9*, 1–15.

[93] X. Geng, Y. Zhang, S. Li, L. Liu, R. Yao, L. Liu, J. Gao, "Design, synthesis, and biological evaluation of novel benzimidazolyl isoxazole derivatives as potential c-Myc G4 stabilizers to suppress c-Myc transcription and myeloma growth" *J Mol Struct* **2023**, *1275*, DOI 10.1016/j.molstruc.2022.134673.

[94] K. M. Felsenstein, L. B. Saunders, J. K. Simmons, E. Leon, D. R. Calabrese, S. Zhang, A. Michalowski, P. Gareiss, B. A. Mock, J. S. Schneekloth, "Small Molecule Microarrays Enable the Identification of a Selective, Quadruplex-Binding Inhibitor of MYC Expression" *ACS Chem Biol* **2016**, *11*, 138–148.

[95] P. Sharma, K. Paul, "Selective Recognition of Oncogene Promoter C-Myc G-Quadruplex: Design, Synthesis, and In Vitro Evaluation of Naphthalimide and Imidazo[1,2-a]pyrazines for Their Anticancer Activity" *ACS Appl Bio Mater* **2025**, *8*, 1377–1396.

[96] H. Xu, M. Di Antonio, S. McKinney, V. Mathew, B. Ho, N. J. O'Neil, N. Dos Santos, J. Silvester, V. Wei, J. Garcia, F. Ka-

beer, D. Lai, P. Soriano, J. Banáth, D. S. Chiu, D. Yap, D. D. Le, F. B. Ye, A. Zhang, K. Thu, J. Soong, S. C. Lin, A. H. C. Tsai, T. Osako, T. Algara, D. N. Saunders, J. Wong, J. Xian, M. B. Bally, J. D. Brenton, G. W. Brown, S. P. Shah, D. Cescon, T. W. Mak, C. Caldas, P. C. Stirling, P. Hieter, S. Balasubramanian, S. Aparicio, "CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours" *Nat Commun* **2017**, *8*, 1–18.

[97] G. C. C. Koh, S. Boushaki, S. J. Zhao, A. M. Pregnall, F. Sadiyah, C. Badja, Y. Memari, I. Georgakopoulos-Soares, S. Nik-Zainal, "The chemotherapeutic drug CX-5461 is a potent mutagen in cultured human cells" *Nature Genetics 2023 56:1* **2023**, *56*, 23–26.

[98] A. Minard, D. Morgan, F. Raguseo, A. Di Porzio, D. Liano, A. G. Jamieson, M. Di Antonio, "A short peptide that preferentially binds c-MYC G-quadruplex DNA" *Chemical Communications* **2020**, *56*, 8940–8943.

[99] D. Dutta, M. Debnath, D. Müller, R. Paul, T. Das, I. Bessi, H. Schwalbe, J. Dash, "Cell penetrating thiazole peptides inhibit c-MYC expression via site-specific targeting of c-MYC G-quadruplex" *Nucleic Acids Res* **2018**, *46*, 5355–5365.

[100] Y. Li, X. Li, J. Pu, Q. Yang, H. Guan, M. Ji, B. Shi, M. Chen, P. Hou, "C-Myc Is a Major Determinant for Antitumor Activity of Aurora A Kinase Inhibitor MLN8237 in Thyroid Cancer" *Thyroid* **2018**, *28*, 1642–1654.

[101] S. R. Boyd, S. Chamakuri, A. J. Trostle, H. Chen, Z. Liu, A. Jian, J. Wang, A. Malovannaya, D. W. Young, "MYC-Targeting PROTACs Lead to Bimodal Degradation and N-Terminal Truncation." *ACS Chem Biol* **2025**, *20*, 896–906.

[102] S. Jia, H. Qi, Z. Shang, X. Hua, A. Liang, Y. Tian, D. Wu, Y. Jiang, X. Liu, C. Mao, "A novel designed fully peptide-based PROTAC, FPP29, demonstrates its potent cytotoxic effects to liver cancer HCCLM3 cells by targeting FOXM1" *Bioorg Chem* **2025**, *163*, DOI 10.1016/j.bioorg.2025.108626.

[103] M. Cornu, T. Lemaitre, C. Kieffer, A. S. Voisin-Chiret, **2025**, Elsevier Ltd preprint, DOI: 10.1016/j.drudis.2025.104376.

[104] X. Li, Z. Zhang, F. Gao, Y. Ma, D. Wei, Z. Lu, S. Chen, M. Wang, Y. Wang, K. Xu, R. Wang, F. Xu, J. Y. Chen, C. Zhu, Z. Li, H. Yu, X. Guan, "c-Myc-Targeting PROTAC Based on a TNA-DNA Bivalent Binder for Combination Therapy of Triple-Negative Breast Cancer" *J Am Chem Soc* **2023**, *145*, 9334–9342.

[105] L. Boike, A. G. Cioffi, F. C. Majewski, J. Co, N. J. Henning, M. D. Jones, G. Liu, J. M. McKenna, J. A. Tallarico, M. Schirle, D. K. Nomura, "Discovery of a Functional Covalent Ligand Targeting an Intrinsically Disordered Cysteine within MYC." *Cell Chem Biol* **2021**, *28*, 4-13.e17.

[106] H. Han, A. D. Jain, M. I. Truica, J. Izquierdo-Ferrer, J. F. Anker, B. Lysy, V. Sagar, Y. Luan, Z. R. Chalmers, K. Unno, H. Mok, R. Vatapalli, Y. A. Yoo, Y. Rodriguez, I. Kandela, J. B. Parker, D. Chakravarti, R. K. Mishra, G. E. Schiltz, S. A. Abdulkadir, "Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy." *Cancer Cell* **2019**, *36*, 483-497.e15.

[107] N. B. Struntz, A. Chen, A. Deutzmann, R. M. Wilson, E. Stefan, H. L. Evans, M. A. Ramirez, T. Liang, F. Caballero, M. H. E. Wildschut, D. V Neel, D. B. Freeman, M. S. Pop, M. McConkey, S. Muller, B. H. Curtin, H. Tseng, K. R. Frombach, V. L. Butty, S. S. Levine, C. Feau, S. Elmiligy, J. A. Hong, T. A. Lewis, A. Vetere, P. A. Clemons, S. E. Malstrom, B. L. Ebert, C. Y. Lin, D. W. Felsher, A. N. Koehler, "Stabilization of the Max Homodimer with a Small Molecule Attenuates Myc-Driven Transcription." *Cell Chem Biol* **2019**, *26*, 711-723.e14.

[108] M. J. Huang, Y. chih Cheng, C. R. Liu, S. Lin, H. E. Liu, "A small-molecule c-Myc inhibitor, 10058-F4, induces cell-cycle arrest, apoptosis, and myeloid differentiation of human acute myeloid leukemia" *Exp Hematol* **2006**, *34*, 1480–1489.

[109] X. Yin, C. Giap, J. S. Lazo, E. V. Prochownik, "Low molecular weight inhibitors of Myc-Max interaction and function" *Oncogene* **2003**, *22*, 6151–6159.

[110] J. Xu, G. Chen, A. T. De Jong, S. H. Shahravan, J. A. Shin, "Max-E47, a designed minimalist protein that targets the E-box DNA site in vivo and in vitro" *J Am Chem Soc* **2009**, *131*, 7839–7848.

[111] L. C. Lustig, D. Dingar, W. B. Tu, C. Lourenco, M. Kalkat, I. Inamoto, R. Ponzielli, W. C. W. Chan, J. A. Shin, L. Z. Penn, "Inhibiting MYC binding to the E-box DNA motif by ME47 decreases tumour

xenograft growth" *Oncogene 2017 36:49* **2017**, *36*, 6830–6837.

[112] M. J. Demma, M. J. Hohn, A. Sun, C. Mapelli, B. Hall, A. Walji, J. O'Neil, "Inhibition of Myc transcriptional activity by a mini-protein based upon Mxd1" *FEBS Lett* **2020**, *594*, 1467–1476.

[113] O. Harel, ‖ Ferran Nadal-Bufi, R. V Nithun, Y. M. Yao, A. Afek, M. Vendrell, M. Jbara, "Chemical Engineering of Transcription Factors Uncovered Cell-Permeable μMax Modulators" *J Am Chem Soc* **2025**, *15*, DOI 10.1021/JACS.5C13964.

[114] L. J. Draeger, G. P. Mullen, "Interaction of the bHLH-zip domain of c-Myc with H1-type peptides. Characterization of helicity in the H1 peptides by NMR." *J Biol Chem* **1994**, *269*, 1785–93.

[115] D. Verzele, A. Madder, "Synthetic Progress in cMyc-Max Oncoprotein Miniaturization: Semi-Online Monitoring Gives Solid-Phase Access to Hydrophobic b(-HLH-)ZIP Peptidosteroid Tweezers" *European J Org Chem* **2013**, *2013*, 673–687.

[116] Y. Ruiz García, Y. V. Pabon-Martinez, C. I. E. Smith, A. Madder, "Specific dsDNA recognition by a mimic of the DNA binding domain of the c-Myc/Max transcription factor" *Chemical Communications* **2017**, *53*, 6653–6656.

[117] L. E. Canne, A. R. Ferre- D'Amare, S. K. Burley, S. B. H. Kent, "Total Chemical Synthesis of a Unique Transcription Factor-Related Protein: cMyc-Max" *J Am Chem Soc* **1995**, *117*, 2998–3007.

[118] D. Krylov, K. Kasai, D. R. Echlin, E. J. Taparowsky, H. Arnheiter, C. Vinson, "A general method to design dominant negatives to B-HLHZip proteins that abolish DNA binding" *Proc Natl Acad Sci U S A* **1997**, *94*, 12274–12279.

[119] L. Soucek, M. Helmer-Citterich, A. Sacco, R. Jucker, G. Cesareni, S. Nasi, "Design and properties of a Myc derivative that efficiently homodimerizes." *Oncogene* **1998**, *17*, 2463–72.

[120] M. Savino, D. Annibali, N. Carucci, E. Favuzzi, M. D. Cole, G. I. Evan, L. Soucek, S. Nasi, "The Action Mechanism of the Myc Inhibitor Termed Omomyc May Give Clues on How to Target Myc for Cancer Therapy" *PLoS One* **2011**, *6*, e22284.

[121] Z. Z. Brown, C. Mapelli, I. Farasat, A. V Shoultz, S. A. Johnson, F. Orvieto, A. Santoprete, E. Bianchi, A. B. McCracken, K. Chen, X. Zhu, M. J. Demma, B. M. Lacey, K. A. Canada, R. M. Garbaccio, J. O'Neil, A. Walji, "Multiple Synthetic Routes to the Mini-Protein Omomyc and Coiled-Coil Domain Truncations." *J Org Chem* **2020**, *85*, 1466–1475.

[122] L. Soucek, J. Whitfield, C. P. Martins, A. J. Finch, D. J. Murphy, N. M. Sodir, A. N. Karnezis, L. B. Swigart, S. Nasi, G. I. Evan, "Modelling Myc inhibition as a cancer therapy" *Nature* **2008**, *455*, 679–683.

[123] M. E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. S. Del Pozo, C. Fiore, L. Foradada, V. C. Cano, M. Sánchez-Hervás, M. Guenther, E. R. Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. M. Alonso, J. R. Whitfield, P. Lavigne, L. Soucek, "Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy" *Sci Transl Med* **2019**, *11*, DOI 10.1126/scitranslmed.aar5012.

[124] E. Wang, A. Sorolla, P. T. Cunningham, H. M. Bogdawa, S. Beck, E. Golden, R. E. Dewhurst, L. Florez, M. N. Cruickshank, K. Hoffmann, R. M. Hopkins, J. Kim, A. J. Woo, P. M. Watt, P. Blancafort, "Tumor penetrating peptides inhibiting MYC as a potent targeted therapeutic strategy for triple-negative breast cancers" *Oncogene* **2018**, *38*, 140–150.

[125] E. Garralda, M. E. Beaulieu, V. Moreno, S. Casacuberta-Serra, S. Martínez-Martín, L. Foradada, G. Alonso, D. Massó-Vallés, S. López-Estévez, T. Jauset, E. Corral de la Fuente, B. Doger, T. Hernández, R. Perez-Lopez, O. Arqués, V. Castillo Cano, J. Morales, J. R. Whitfield, M. Niewel, L. Soucek, E. Calvo, "MYC targeting by OMO-103 in solid tumors: a phase 1 trial" *Nat Med* **2024**, *30*, 762–771.

[126] J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, "De novo design of protein structure and function with RFdiffusion" *Nature 2023 620:7976*

2023, *620*, 1089–1100.

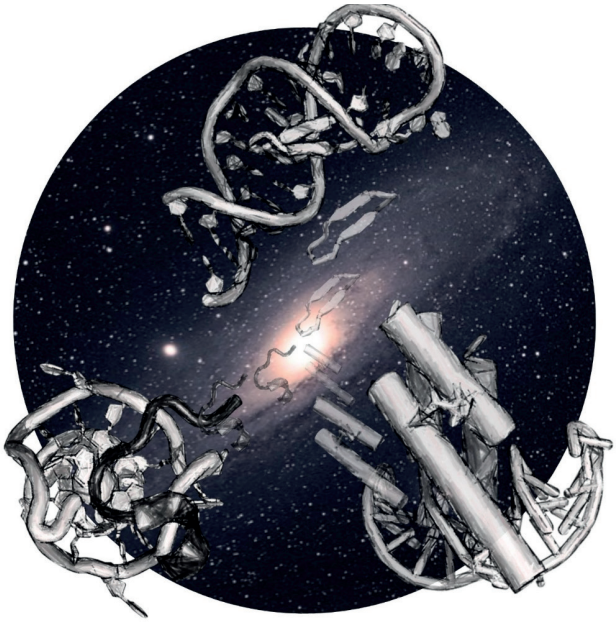
[127] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, "Highly accurate protein structure prediction with AlphaFold." *Nature* **2021**, *596*, 583–589.

[128] K. Wu, H. Jiang, D. R. Hicks, C. Liu, E. Muratspahić, T. A. Ramelot, Y. Liu, K. McNally, S. Kenny, A. Mihut, A. Gaur, B. Coventry, W. Chen, A. K. Bera, A. Kang, S. Gerben, M. Y.-L. Lamb, A. Murray, X. Li, M. A. Kennedy, W. Yang, Z. Song, G. Schober, S. M. Brierley, J. O'Neill, M. H. Gelb, G. T. Montelione, E. Derivery, D. Baker, **2024**, DOI: 10.1101/2024.07.15.603480.

[129] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, "Highly accurate protein structure prediction for the human proteome" *Nature 2021 596:7873* **2021**, *596*, 590–596.

[130] H. Bret, J. Gao, D. J. Zea, J. Andreani, R. Guerois, "From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2." *Nat Commun* **2024**, *15*, 597.

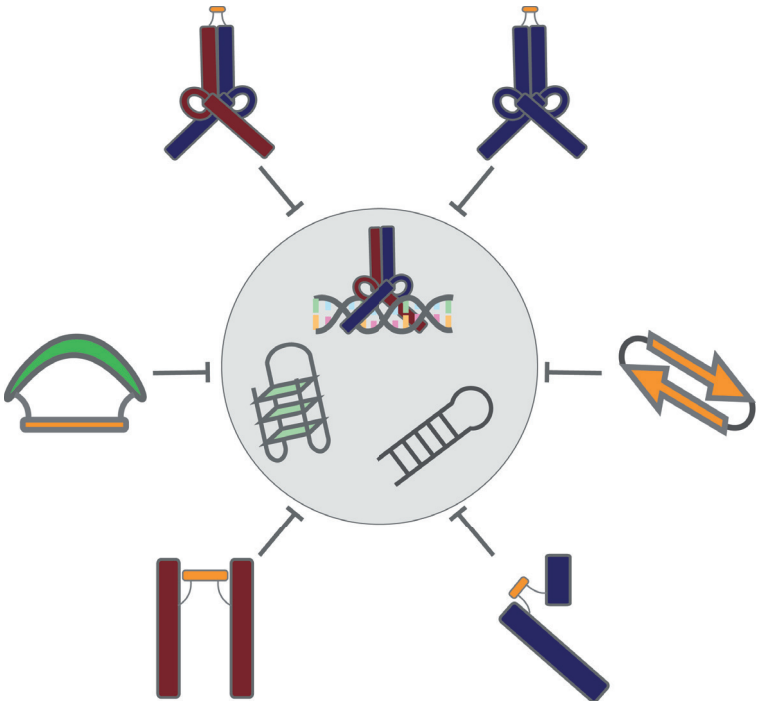
[131] P. Bryant, G. Pozzati, A. Elofsson, "Improved prediction of protein-protein interactions using AlphaFold2" *Nature Communications 2022 13:1* **2022**, *13*, 1–11.



Adapted from: B.D. Ellenbroek, J. P. Kahler, S. R. Evers, S. J. Pomplun, "Synthetic peptides: promising modalities for the targeting of disease related nucleic acids" *Angewandte Chemie International Edition* **2024**, DOI 10.1002/anie.202401704.

Chapter 2

Synthetic Peptides: Promising Modalities for the Targeting of Disease-Related Nucleic Acids



Authors:

Brecht D. Ellenbroek[‡], Dr. Jan Pascal Kahler[‡], Sophie R. Evers, Dr. Sebastian J. Pomplun

[‡]These authors contributed equally.

Abstract

DNA and RNA play pivotal roles in life processes by storing and transferring genetic information, modulating gene expression, and contributing to essential cellular machinery such as ribosomes. Dysregulation and mutations in nucleic acid-related processes are implicated in numerous diseases. Despite the critical impact on health of nucleic acid mutations or dysregulation, therapeutic compounds addressing these biomolecules remain limited. Peptides have emerged as a promising class of molecules for biomedical research, offering potential solutions for challenging drug targets. This review focuses on the use of synthetic peptides to target disease-related nucleic acids. We discuss examples of peptides targeting double-stranded DNA, including the clinical candidate Omomyc, and compounds designed for regulatory G-quadruplexes. Further, we provide insights into both library-based screenings and the rational design of peptides to target regulatory human RNA scaffolds and viral RNAs, emphasizing the potential of peptides in addressing nucleic acid-related diseases.

2.1 Introduction

Nucleic acids (NAs) play fundamental roles in all processes of life. Ribonucleic acids (RNA) have even been postulated as the potential linchpin in the origin of life itself. [1] Deoxyribonucleic acids (DNA) serve as repositories for the genetic blueprints inherent to every living organism and cell. As first described by Watson and Crick, the DNA houses the A, C, G, and T sequence in a double-helix arrangement,[2] delineating the unique code of each gene. Polymerases transcribe genes into RNA, which in the case of messenger RNA (mRNA) subsequently can undergo ribosomal translation, dictating the amino acid sequence of a final protein product. This unidirectional genetic flow, from DNA to RNA and from RNA to protein, has historically been termed the 'central dogma of molecular biology'.[3] Discoveries from the last decades, encompassing regulatory DNA elements (e.g., G-quadruplexes or G4s,[4] depicted in **Figure 2.3**), a variety of non-coding RNAs[5] (ncRNAs, such as micro RNAs (miRNA),[6] as depicted in **Figure 2.5**), and the intricate role of transcription factor (TF) proteins,[7] collectively illuminate the intricate and multifaceted nature of nucleic acids' functions and the genetic information cascade, far surpassing the unidirectionality of the traditional 'central dogma'.

Despite their fundamental role in health and disease, only few therapeutic compounds targeting specific disease related NAs have been developed and approved for use in humans. Classical small molecule drugs usually target proteins, and specifically proteins that have well-defined binding pockets, such as receptors or enzymes. DNA in its double helix does not present 'druggable' binding pockets for small molecules. DNA-targeted small molecules, indeed, usually have sequence agnostic unspecific affects,[8] ultimately leading to DNA damage and cell death.[9] An exceptional class of small molecules, the pyrrole-imidazole polyamides, mainly discovered and investigated by Dervan *et al.*, can target specific DNA sequences, via minor groove binding and base pair recognition.[10] However, the advancement towards therapeutic use has remained limited.[11] RNA can fold to several more or less defined three-dimensional structures.[12] Compared to proteins these motifs are less defined and structurally more flexible, but, nevertheless, targeting these motifs with small molecules has recently shown intriguing potential.[13,14]

Chemical modalities beyond small molecules can target specific NA sequences. Antisense oligonucleotides (ASOs) can be designed to hybridize with complementary RNA sequences, thereby modulating gene expression through different mechanisms.[15] However, challenges associated with poor pharmacokinetics, limited cellular uptake, and inefficient delivery to target tissues have constrained the translation of ASOs into successful therapies. Nevertheless, to date, nine ASO-based drugs have secured regulatory approval for clinical use.[16] Other approaches include the use of engineered DNA-binding proteins, such as zinc finger (ZF) proteins[17] and transcription activator-like effector nucleases (TALENs).[18] While these proteins

can be engineered to bind specific DNA sequences, obstacles like suboptimal cellular entry and issues related to specificity have impeded their widespread implementation in clinical settings. The revolutionary CRISPR-Cas system presents another avenue for nucleic acid manipulation.[19] This technology holds immense potential for precise genome editing, but hurdles related to efficient delivery of CRISPR components into target cells remain a substantial challenge. Developing effective delivery strategies that ensure the transport of Cas proteins and guide RNAs into the desired cells while minimizing off-target effects is a critical area of ongoing research.[20]

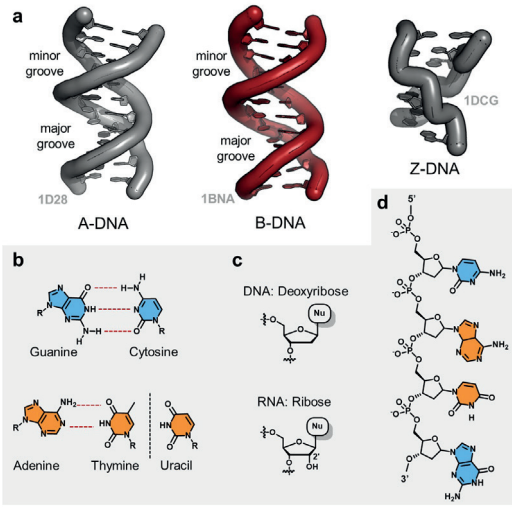
Peptides have emerged as a compelling class of molecules with the potential to tackle challenging drug targets in biomedical research.[21] Peptides offer a unique balance of specificity, affinity, and structural diversity. Their larger size and capacity for intricate three-dimensional structures enable precise interactions with target sites that may be beyond the reach of conventional small molecules.[22] At the same time, their size is significantly reduced compared to other biomolecular drugs, making their production and handling more straightforward. While there are several challenges related to peptides' pharmacokinetic properties, these modalities hold potential to tackle intracellular targets. For the discovery and the development of bioactive peptides there are a variety of robust strategies: when structural insights are available, the rational design of short active peptides can be pursued, often stemming from larger protein structures.[23] Alternatively, powerful affinity selection approaches, such as phage display,[24] or affinity selection mass spectrometry,[25] can be leveraged to screen libraries of billions of sequences for high-affinity hits, all in one experiment. A plethora of non-canonical amino acid building blocks can be used to improve the affinity and overall properties of peptide hits. Generally speaking, peptides present a versatile platform for addressing a wide range of therapeutic challenges, from targeting protein-protein interactions to modulating complex nucleic acid structures.

In this review we highlight the utilization of synthetic peptides to target disease related nucleic acids. In a first section we will focus on proteomimetic peptides able to specifically target E-Box DNA and inhibit the oncogenic transcription factor Myc. Notably, one peptidic Myc inhibitor, a variant of the miniprotein Omomyc,[26] is currently in human clinical trials, underlining the promise of this targeting strategy. Several synthetic analogs have been developed and show intriguing potential.[27–33] Further, we will describe peptides targeting DNA G-quadruplex structures, modulating gene transcription.[34–37] In a second section we will describe the development and discovery of peptides targeting diverse RNA structures involved in disease. Specifically, we will focus on compounds targeting viral RNAs and non-coding microRNA (miRNA) involved in cancer. For these targets we will present proteomimetic strategies, library selection platforms and unnatural modifications such as nucleobase amino acids and peptide stapling, leading to intriguing bioactive candidates.

2.2 Targeting of DNA with synthetic peptides

DNA stores genetic information and can be called the code of life. Replication of cellular DNA is necessary for cell division and proliferation. Amongst other functions, DNA encodes for proteins which are synthesized by first transcription of DNA into RNA followed by translation of RNA into the respective protein. DNA transcription is regulated by many different proteins such as transcription factors which can have an effect on the level of transcription of a certain protein and thus indirectly on the amount of protein eventually synthesized.

DNA sequences can form secondary structures and the biologically most prevalent one is the canonical double-helix, so-called B-DNA, proposed by Watson and Crick in 1953.[2] Here, two complementary single strands of DNA running in opposite directions (ssDNA) form double-stranded DNA (dsDNA) through canonical pairing of the individual bases which constitute the DNA building blocks: adenine pairs with thymine and guanine pairs with cytosine (**Figure 2.1b**).[38] In this right-handed double helix, the base pairs lie in one plane approximately vertically to the helix axis with their aromatic rings stacking on top of each other, stabilizing the helix (**Figure 2.1a, middle**). As the backbone carbons of both strands are not positioned



▲**Figure 2.1. The structure and components of nucleic acids. (a)** DNA secondary structures with B-DNA being the most common one (PDB: A-DNA 1D28; B-DNA 1BNA; Z-DNA 1DCG, Structures generated using PyMol). **(b)** Canonical base pairing. Adenine pairs with thymine through two H-bonds and guanine with cytosine through three H-bonds. In RNA thymine is replaced by uracil. R denotes the connection with the sugar-phosphate backbone **(c)** Structures of deoxyribose and ribose that form part of the backbone of DNA and RNA, respectively. **(d)** The structure of DNA with the ribose-phosphate backbone.

directly opposite each other on the axis of the helix two grooves of different size, major and minor groove, are running around the B-DNA.[39] The wider major groove is larger and provides more sequence information than the minor groove. It is thus more easily accessible to and is where most DNA binding proteins bind.[38]

Further, dsDNA can adopt an A- and Z-form as well as other motifs such as G-quadruplexes (G4s, see 2.1). The A-form of DNA, too, is a right-handed double helix which adopts a wider helix with a less accessible major groove (**Figure 2.1a, left**). Z-DNA on the other hand, is a left-handed helix with almost equally-sized grooves (**Figure 2.1a, right**).[38]

As DNA transcription and its regulation directly impact the eventual protein levels, focusing on targeting DNA represents a strategic approach to influencing disease-related proteins at an earlier level than directly targeting the proteins themselves.

2.2.1 Targeting dsDNA with synthetic peptides

Many strategies have been developed to target dsDNA and various dsDNA targeting small molecules such as cisplatin, doxorubicin and etoposide are widely used in cancer therapy. However, these small molecules do not recognize specific sequences and act as general cytotoxic agents. Conversely, proteins can distinguish between different DNA sequences. Nevertheless, entire proteins are not cell permeable and therefore of limited use when it comes to developing therapeutics for intracellular targets. Peptides are positioned between small molecules and proteins and might offer intriguing possibilities to develop and advance DNA-targeted therapeutics. Early on, Dervan *et al.* developed DNA binding molecules based on the structures of natural products Netropsin and Distamycin A, resulting in DNA-binding peptide-like compounds.[40]

A generally used method to obtain peptides that can target DNA is to modify existing natural TFs.[41] Most of these studies were aimed at a broader understanding of TFs and their binding to DNA. For instance, Mascareñas *et al.* introduced two His mutations in the basic region of yeast TF GCN4 which acted as a metal-dependent staple inducing formation of an α helix and allowing for specific binding to dsDNA and internalization.[42,43] In a different study, stapling of the same peptide by various cysteine crosslinkers afforded peptides binding to dsDNA with enhanced cell penetration and proteolytic stability.[44] All known DNA targeting peptides were developed based on natural TFs or by screening large peptide libraries. Designing sequence specific peptides *de novo* has not been achieved to date.

2.2.1.1 Targeting dsDNA with synthetic peptides derived from Myc and Max

The c-Myc TF plays a role in a multitude of cellular processes such as cell proliferation

and survival, differentiation, apoptosis, metabolism and DNA repair.[45] It belongs to the basic helix-loop-helix leucine zipper (bHLHLZ) family of TFs. The full length protein consists of an N-terminal transactivation domain followed by a central domain with a nuclear localization signal and the C-terminal bHLHLZ domain. While monomeric Myc is unable to bind to DNA, it heterodimerizes with its partner Myc-associated protein X (Max) through coiling of their bHLHLZ domains. This enables the Myc/Max heterodimer to bind to a DNA sequence called enhancer box (E-box) with the canonical sequence CACGTG (**Figure 2.2a, left**).[45] Activation of gene transcription is achieved by recruitment of different cofactors such as histone acetyltransferase complexes (TTRAP, GCN5, TIP60, TIP48) eventually leading to histone acetylation and allowing RNA polymerase II binding.[32,46–48]

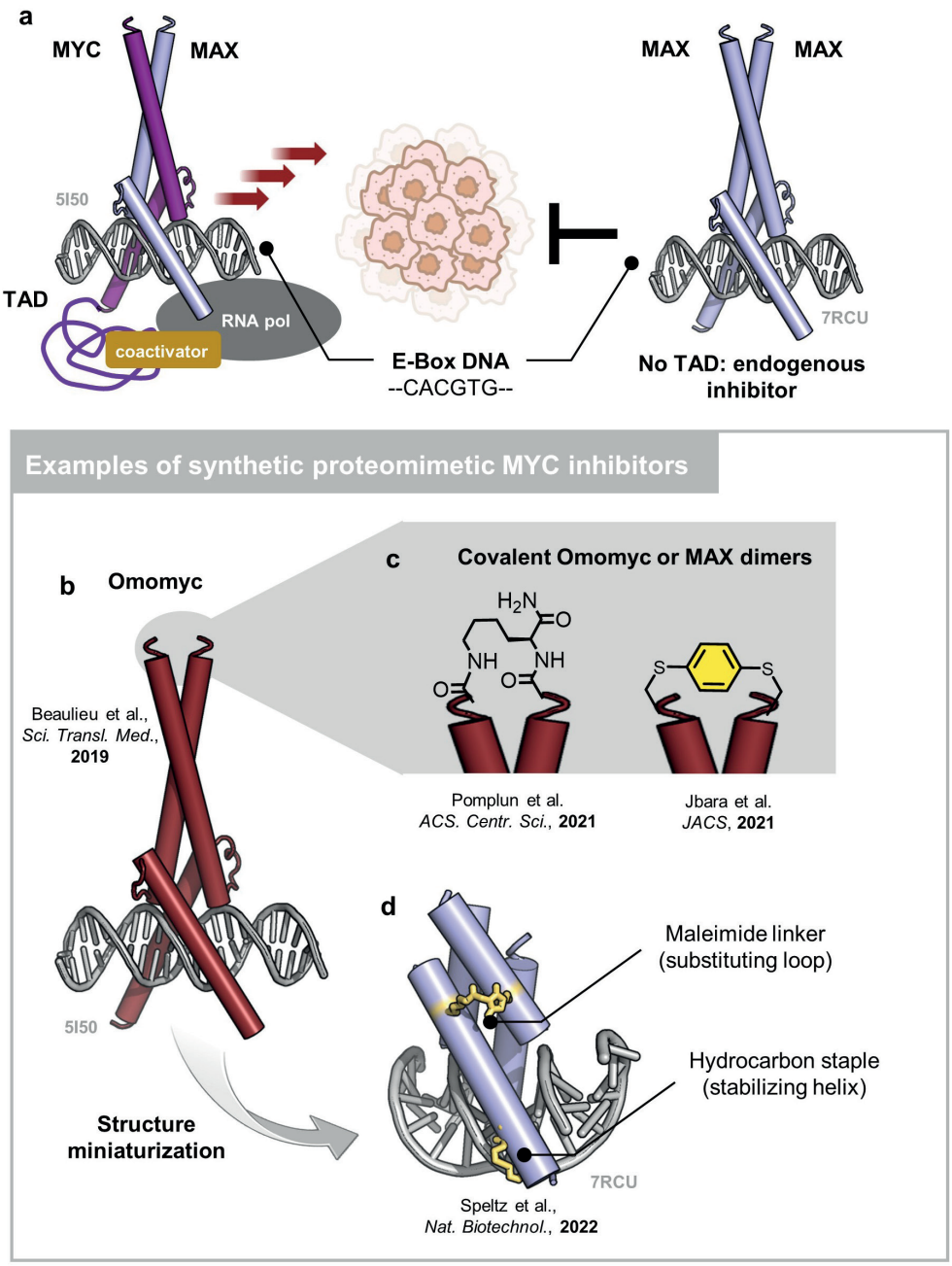
Acting within the Myc/Max/Mxd network, Myc can only heterodimerize with Max whereas Max itself can homodimerize as well as heterodimerize with other bHLHLZ proteins in the network, such as Mxd1-4 and others. These interactions sequester Max, effectively competing with Myc for Max binding. The Max/Max homodimer can also bind to E-box sites and block Myc/Max from binding (**Figure 2.2a, right**).[32]

The Myc TF modulates about 15% of the entire genome and it is highly abundant in proliferating cells. Under normal circumstances the level of Myc is tightly controlled by several mechanisms on the genetic, mRNA and protein level. Aberrant Myc control and Myc overexpression is observed in over 50% of human cancers. When abnormal Myc levels occur in combination with defective regulation of stress response checkpoints, non-canonical E-box sites become occupied by Myc resulting in activation of otherwise silent genes. This eventually leads to Myc involvement in all hallmarks of cancer such as proliferation angiogenesis and invasion.[32,46,49,50] Targeting Myc and Myc-controlled expression thus constitutes a promising pathway in cancer therapy and given its universal role, might make it an especially interesting target in cancers with poor prognosis and limited treatment options.

However, Myc is an intrinsically disordered protein (IDP) and by itself lacks features such as defined pockets which would be paramount for the development of classical small molecule drugs. Therefore, only a handful of small molecule drugs have been developed. These include the Max homodimer stabilizer KI-MS2-008,[51] EN4 which reduces Myc stability by covalently targeting C171[52] as well as Myci361 and Myci975 which disrupt Myc/Max dimerization and destabilize the Myc protein.[53] An interesting new approach is a recently developed bivalent Proteolysis targeting chimera (PROTAC) aimed at Myc which combines a TNA aptamer with an E-box DNA motif to achieve selectivity.[54]

2.2.1.2 Omomyc

Developed by Soucek *et al.* in the late 90s, Omomyc (**Figure 2.2b**) is a Myc-derived



protein which, in contrast to its parent protein, is able to homodimerize. However, as it does not contain a transactivation domain, it is unable to activate gene transcription. Omomyc's ability for homodimerization is achieved by four distinct mutations in its leucine zipper region which prevents homodimerization of canonical MYC: Two

◀**Figure 2.2 The Myc/Max system in gene transcription and synthetic transcription factors that inhibit Myc**. **(a)** Myc heterodimerizes with Max to bind to the E-box DNA sequence. The transactivation domain (TAD) acts as binding site for further cofactors eventually allowing RNA polymerase to bind, leading to gene transcription and subsequently cell proliferation. Conversely, Max can homodimerize and bind the E-box sequence (PDB 7RCU). As it is lacking a TAD, no gene transcription and cell proliferation is initiated. The Max homodimer essentially acts as an endogenous inhibitor. **(b)** Omomyc is based upon the structure of Myc but possesses four mutations, allowing it to homodimerize and bind to the E-box sequence (PDB 5I50). Importantly, cell permeability of Omomyc has been shown by microscopy, cell-based assays and in an animal model.[26] **(c)** Omomyc made by fast-flow SPPS has been synthetically dimerized and the dimers used to show DNA binding. Also the synthetically dimerized Omomyc displays cell permeability as shown by fluorescence microscopy and cell-based assays.[27,28] **(d)** The Max basic helix together with a minimal zipper helix has been linked with a maleimide linker and stapled using hydrocarbon stapling in an attempt to minimize the structure of synthetic transcription factors and improve cell penetration (PDB 7RCU). Cell penetration has been shown by fluorescence microscopy and in cell-based assays.[70] Representations of crystal structures were generated using PyMol.

arginines lead to unfavorable equally charged contacts in a Myc homodimer and were replaced by the AA found in Max at their respective positions (R77Q and R78N). The same was performed mutating E70I. Finally, E63 was replaced, not with asparagine, as found in Max but with a threonine as this lead to a better complementarity of the two monomers.[55] Omomyc has been found to form homodimers as well as heterodimers with Myc and Max. Whereas Omomyc homodimers and Omomyc/Max-heterodimers can bind DNA at the E-box motif thereby preventing Myc gene regulation, Omomyc/Myc heterodimers do not bind DNA. Omomyc thus competes with Myc for DNA binding while at the same time sequestering Myc away from DNA.[26],[56]

When conditionally expressed in cells, Omomyc has been found to bind to the E-box sequence thereby blocking Myc binding and thus transactivation of genes. It does, however, have little effect on or even potentiates Myc gene repression.[57] It has been suggested that this could be by Omomyc interaction with Myc interacting zinc finger 1 (Miz-1).[58] Yet, blocking of repression has been found as well.[56]

Omomyc displays a Myc-dependent apoptotic effect on cells.[57] In mouse skin, Omomyc was able to restore normal tissue homeostasis and counteract the Myc-induced oncogenic activity.[59] When expressed in transgenic mice inducible by doxycycline, Omomyc lead to regression and even eradication of lung carcinoma, whereas the effects on normal tissue are reversible and mostly found in highly proliferating cells.[60,61] It also leads to tumor regression in transgenic mice with an SV40-driven pancreatic islet tumor, suggesting that endogenous Myc is not only necessary to maintain the tumor environment in Myc-driven tumors but also in tumors driven by other oncogenic processes.[62] Similar effects have been found with

glioma.[63] The reason for the remarkable effects on tumor growth while simultaneously exhibiting little side effects has been suggested to be rooted in Omomyc blunting Myc promoter invasion. Omomyc is believed to mainly affect low affinity Myc-binding sites that are only occupied at high Myc levels whereas Myc-binding to high affinity binding sites also occupied at physiological Myc-levels are hardly affected.[56]

Due to the need to conditionally express Omomyc in these studies, its use was mainly limited to being a proof of principle and research tool instead of being used as a potential cancer treatment. More than 20 years after its emergence, however, Soucek *et al.* found that Omomyc possesses cell penetrating properties, most likely conferred by the basic region and its generally high arginine content.[26] Omomyc is considered to mainly enter the cell by clathrin-mediated endocytosis and macropinocytosis. Treatment of a mouse NSCL xenograft with Omomyc slowed tumor progression and in combination with paclitaxel almost led to tumor abrogation.[26] Although another study in which Omomyc is fused to a cell-penetrating peptide (CPP) indicates less favorable cell penetrating properties for Omomyc when not combined with the CPP,[64] cell penetration of Omomyc has been corroborated by Demma *et. al.* in an extensive study in which they also further examined the Omomyc binding to E-box sites.[65]

Several solid phase peptide synthesis (SPPS) approaches have been used to produce Omomyc[27–29,31] (as well as Myc and Max).[30,66–68] Brown *et al.* synthesized the whole protein using either microwave or infrared irradiation during SPPS and also explored native chemical ligation (NCL) to be able to make Omomyc from two fragments. Furthermore, they developed a high-throughput microbial expression workflow to introduce mutations and improve Omomyc binding affinity. They were able to significantly improve the K_D towards a fluorescently-labelled E-box DNA and found that the coiled-coil dimerization domain could be almost entirely removed and replaced with a disulfide bond.[69]

Omomyc, Max and the DNA binding domain of Myc have been synthesized by fast-flow SPPS. In a first approach all six combinations of homo- and heterodimers were obtained using an S-arylation approach with palladium oxidative addition complexes. All dimers except Myc/Myc and Omomyc/Myc do bind DNA in electrophoretic mobility shift assays (EMSAs), confirming earlier findings (**Figure 2.2c, right**).[27] In a second approach, Pomplun *et al.* obtained homodimers of Max and Omomyc as well as Myc/Max and Myc/Omomyc heterodimers by fast-flow synthesis using a special linker to the resin allowing their parallel (homodimers) or sequential (heterodimers) synthesis (**Figure 2.2c, left**).[28] The dimers display binding in EMSAs and inhibit cell proliferation. Penetration into the nucleus is shown using a fluorescently-tagged analog and can be improved by adding a nuclear penetration sequence to the protein. Remarkably, the size of these synthetic cell penetrating modalities ranges from 160 to 230 amino acid residues. In both studies RNA sequencing and gene set enrichment

analysis (GSEA) were performed and show significant modulation of Myc-target genes.

In an attempt to find new stapling and S-alkylation strategies, a Myc-derived 38mer was used in Zn^{2+}-promoted stapling reactions. This method was then expanded to S-palmitylate full length Omomyc which might increase cell permeability and serum stability as the palmityl-group promotes albumin binding.[71]

2.2.1.3 Max-derived mini proteins

A miniprotein derived from a fusion of Max and E47, the miniprotein ME47, has been developed in an attempt to inhibit Myc-binding to DNA using a "smaller protein with a simplified structure".[72] The chimeric ME47 comprises the Max basic region in conjunction with the HLH of TF E47. The strongly homodimerizing HLH leads to stable ME47 dimers which bind DNA with affinity in the low nanomolar range but, due to the differences between its HLH and the HLHzip of Myc and Max, no interaction with other proteins of the Myc/Max system is possible.[33] The crystal structure has been solved to a resolution of 1.7 Å and shows that ME47 usually exists as a dimer in solution, although about 17% of it forms tetramers. Interestingly, no monomeric ME47 is found, further highlighting that no interaction with Myc or Max is to be expected. As is to be expected with intrinsically disordered regions, when crystallized in the presence of dsDNA, alpha-helicity increases.[72] ME47 reduces cell proliferation when ectopically expressed in TNBCs MDA-MB-231 and delays tumor growth in mouse xenografts. Using FLAG-tagged ME47 in a ChIP-qPCR experiment, it is shown to bind and compete with Myc for binding with E-box as Myc-binding is reduced in a competitive ChIP-qPCR experiment.[73]

Using phage-assisted continuous evolution (PACE), Inamoto *et al.* further evolved ME47. First they mutated R12A and C29A and eventually added the LZ domain of their previously developed FosW protein to stabilize dimerization and correct folding of the protein which shows a tendency to misfold. This yields MEF which they term a 'franken-protein' as it is composed of domains from different classes of TFs, preventing heterodimerization. MEF displays slightly increased E-box binding and strong cooperative binding.[74]

In recent work Moellering and coworkers used the basic helix along with the minimal zipper helix from Max to target E-Box DNA with high affinity. The individual helices harbor olefin containing amino acids allowing for side-chain stapling by ring-closure-metathesis and appropriate amino acids for ligation with a maleimide linker. This resulted in a synthetic TF composed of stapled basic and minimal zipper helices connected by a maleimide linker termed 'cross-dimer' (**Figure 2.2d**). Their synthetic TFs exhibit an effect on Myc-regulated genes, are cell permeable and, owing to stapling, are more resistant to proteolytic cleavage. Most interestingly, they could

further expand their design to make synthetic TFs targeting different DNA sequences by sequence alignment.[70]

Similar to Omomyc, O'Neil and coworkers at MSD developed Mad, an artificial transcriptional repressor based on Mxd1, a Max binder and Myc antagonist. Mad is composed of the first 146 AAs of Mxd1 including an S146A mutation to prevent phosphorylation, ubiquitination and subsequent degradation along with a domain for interaction with mSin3, a transcriptional repressor. Mad exhibits better properties than Omomyc in cell proliferation assays and penetrates the cell where it can be found in the nucleus. It further interacts with Max but not Myc as seen in coimmunoprecipitation experiments and displays stronger downregulation of Myc-upregulated genes than Omomyc. Lastly, a ChIP assay confirms that Mad binds to E-box motifs in Myc-regulated genes and actually competes with Myc for binding.[75]

2.2.2 Targeting G-quadruplexes with synthetic peptides

2.2.2.1 Structure and biological relevance of G-quadruplexes

Guanine-rich sequences of ssDNA (e.g. unwound dsDNA) as well as RNA are prone to form guanine-quadruplex (G4) secondary structures (**Figure 2.3**). In these G4s four guanines assemble into a planar quartet formed by Hoogsteen-type base-pairing and the G4 is then established by more than two tetrads stacking on top of each other making use of π-π-interactions. The tetrads are further stabilized by binding of positively-charged ions such as Na^+ and K^+, with the latter being the most stabilizing one.[76,77] Depending on the direction of the four strands forming the edges of the G4, different topologies of them can form.[78] Moreover, G4s can be intra- and intermolecular. The cytosine-rich ssDNA complementary to G4s has been found to forge secondary structures termed imotif[79] but whether these form under the same conditions as G4s or are mutually exclusive has not been conclusively proven thus far.

The consensus sequence for G4s to assemble is $G_{3\text{-}5}N_{1\text{-}7}G_{3\text{-}5}N_{1\text{-}7}G_{3\text{-}5}N_{1\text{-}7}G_{3\text{-}5}$ with the N-termed nucleotides functioning as linkers between the columns of guanines. Simply scanning the human genome for this sequence suggested that there may be about 300,000 potential G4s.[80] However, two guanines from within a column can be separated by other nucleotides forming a bulge between the guanidine stretch thereby leading to non-consensus G4 sequences.[81] Thus, using Illumina sequencing in a method called g4-seq a total of 716,310 potential G4-forming sequences were identified.[82]

G4s are more abundant at telomeres, promoter regions, 5' and 3' untranslated regions (UTRs) along with ribosome binding sites (RBSs) suggesting a biological function and thereby making them potential therapeutic targets.[37,83] Generally, G4s might play a role in cellular functions such as transcription, replication, genomic stability

and others.[78] For instance, in telomeres, G4s might prevent normal function of telomerases by segregating the 3' end and hampering its extension.[84] Stabilizing telomeric G4s is thus a valuable strategy to target telomerase activity in cancer. Nevertheless, it has also been found that telomerase can unwind telomeric G4s and might play a role in their recruitment making it crucial to establish when stabilization of G4s is beneficent.[78,85]

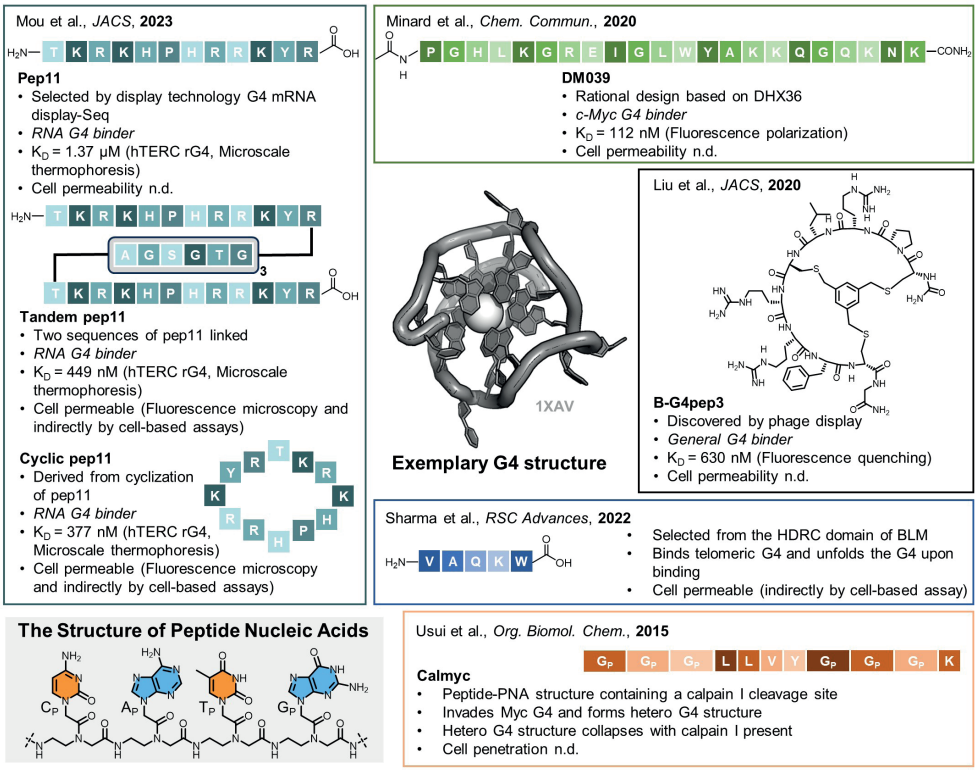
G4 enrichment in oncogenic promoter regions such as the Myc oncogene[86] suggests a role in regulating transcription. The Myc oncogene promoter entails a G4 and treatment with G4 stabilizing porphyrin TMPyP4 reduces MYC expression,[87] hinting on G4s functioning as negative regulators of gene expression. Similarly, stabilizing a G4 in the KRAS promoter with TMPyP4 results in reduced expression[88] and small molecules binding to the c-kit G4 reduced c-kit expression in MCF7 and GIST882 cells[89,90] exemplifying the value of G4 binding molecules as pharmacological compounds. However, many small molecules make use of π-stacking interactions with the surface of G4s which is a good strategy for gaining selectivity over dsDNA but does hardly allow to differentiate between different G4s. Peptides offer a more complex array of possible interactions and are thus a valuable tool to target and differentiate between different G4s.[36]

2.2.2.2 Targeting G4s with synthetic peptides

Many proteins bind and either stabilize, unwind or degrade G4s.[37] For instance, nucleolin, a protein binding to histones and playing a role in chromatin decondensation has been found to stabilize and fold G4s. It is involved in formation of the G4 in the Myc promoter region thereby reducing promoter activity and acting as tumor suppressor.[91] Furthermore, topoisomerases promote G4 formation and helicases unwind G4s catalyzing ssDNA formation.[37] The multitude of G4 binding proteins thus provide a logical and promising starting point for the design and development of G4 targeting peptides and numerous peptides have been developed. Therefore, this chapter will only highlight some recently developed peptides (for a comprehensive review see Sharma, Saxena and Kaur[37]).

Generally, peptides binding G4s can be classified according to different categories, one of which being their binding mode. As such they can intercalate into the planes of the G4 and they can bind to the loops and bulges between Guanidines. Additionally, not directly binding to the G4s, they can also bind at the dsDNA-ssDNA junction and bind the complementary cytosine-rich ssDNA.[37]

Making use of the crystal structure of bovine helicase DHX36 bound to the G4 found in the Myc oncogene promoter region[92], Minard *et al.*[36] designed a peptide targeting this G4. The 22 AA peptide termed DM039 spanning residues 57-78 as the minimal Myc-binding domain of DHX36 bound the Myc G4 selectively not only



▲**Figure 2.3. Peptides targeting G-quadruplexes.** General G4 structure (PDB 1XAV, middle) and examples of G4-targeting peptides. Lower left panel shows the structure of Peptide Nucleic Acids (PNAs) and lower right panel shows a peptide-PNA structure. Subscript P denotes PNA. Representation of crystal structure was generated using PyMol. n.d. = not determined. Top left: Mou *et al.*[97]; Top right: Minard *et al.*[36]; Top middle right: Liu *et al.*[94]; Lower middle right: Sharma et. al.[98]; Bottom right: Usui *et al.*[34]

over dsDNA and non G4 ssDNA but also over other G4s such as the c-kit1 and ckit2 G4s. However, when they introduced either an i, i+4 or an i, i+7 staple the peptide bound with reduced affinity to the Myc G4 and the selectivity over other G4s was lost. This suggests that a certain amount of flexibility is required for binding and thus an induced-fit binding mechanism.[36]

The group of Shankar Balasubramian[35,93] used phage display from a library of 4x4 and 3x3 bicyclic peptides to screen for G4 binders. They cyclized the linear peptides the phages displayed using TBMB and screened on biotinylated G4s. They found a bicycle with nanomolar affinity towards the c-kit2 G4 and by computational modeling demonstrate the affinity is derived from molecular preorganization.[94]

Exploring an interesting concept Usui *et al.*[34] employed peptide nucleic acids (PNAs, **Figure 2.3 lower left panel**) for the formation of switchable heteromolecular G4s in the Myc promoter region (see Brodyagin *et al.*[95] for a review of PNA and their application). They designed calmyc, a guanine-rich PNA construct comprising a calpain I protease cleavage site. Calmyc is able to invade the Myc G4 and form two 3-tetrad PNA-DNA hybrid G4s with ssDNA from the Myc promoter. These hybrid G4s displayed higher stability than the DNA G4 but collapsed upon calpain I addition allowing for control over the DNA secondary structure and stability.[34] They later employed calmyc to break G-wires.[96]

Saxena and coworkers demonstrated that not only stabilization but also unwinding of G4s can have beneficial effects, highlighting the complicated roles G4s might play. They employed a peptide library based upon the bloom syndrome protein helicase domain (BLM) to target telomeric G4. BLM unfolds telomeric G4s enabling normal telomer replication which is defective in Bloom's syndrome leading to telomere shortening and genomic instability.[99] The 5-mer BLM-peptide, VAQKW, derived from the library binds to the G4 loop with the C-terminal tryptophan intercalating into the G4-planes. This intercalation destabilizes the stacking interactions and leads to unwinding of the G4. Interestingly, the peptide displays an antiproliferative effect on MDA-MB-231 breast cancer cells.[98]

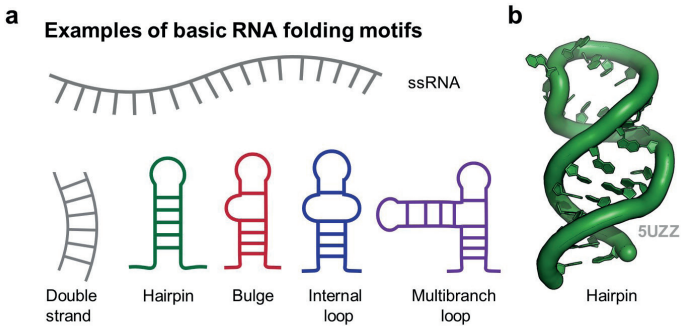
In a different project, they developed a 10-mer peptide, QW10 (QQWQQQQWQQ), consisting of glutamine for hydrogen-bond interaction with guanine bases and tryptophane for intercalation into the G4 planes. QW10 stabilizes telomeric[100] as well as the c-Myc[101] G4s in the presence of potassium. Conversely, another human telomeric G4 is destabilized by this QW10.[102] Moreover, they then shorten QW10 to a 5-mer peptide (QQWQQ) which destabilizes and unfolds human telomeric G4s by tryptophan intercalation.[103]

Recently, an RNA display platform termed G4-mRNA display-seq has been developed for screening of peptide binders to RNA G4s. They found a peptide binder, pep11, that bound specifically to RNA G4s. Furthermore, binding could be increased by cyclization of the peptide, showing that sequences found by this technique can be valuable as starting points for rationally designing peptide binders to G4s.[97] Although this method was developed using RNA G4s, it should be amenable to DNA G4s.

2.3 Targeting of RNA with synthetic peptides

In the process of transcription, RNA polymerases, assisted by various enzymes, transcribe genetic DNA into RNA. A key molecular distinction between RNA and DNA oligonucleotides lies in the presence of a 2' hydroxyl group in RNA and the use of the uracil nucleobase (RNA) instead of thymidine (DNA). RNA serves diverse functions within cells. mRNA undergoes translation to synthesize amino

acid sequences, ultimately forming proteins. A substantial portion of transcribed RNA is classified as ncRNA. ncRNA can fold into various (semi)stable secondary and tertiary structures (**Figure 2.4**), and performs a multitude of functions.[104,105] Prominent examples of ncRNA include transfer RNA (tRNA) and ribosomal RNA, both crucial for the translation process. Another noteworthy category of ncRNA comprises microRNAs (miRNAs), serving as crucial regulatory elements in gene expression (elaborated upon in the subsequent chapter). Numerous long and short ncRNAs harbor yet unknown functions.[106,107]



▲**Figure 2.4. RNA secondary structures** (**a**) graphic representation of common secondary RNA structures. (**b**) Crystal structure of an RNA hairpin. (PDB 5UZZ, structure created using PyMol).

Typical RNA motifs include double stranded RNA (dsRNA), hairpin RNA, stemloops, bulges and internal loops.[107,108] Longer ncRNAs can also fold into tertiary structures. [109] RNA structures are less rigid and defined than protein structures, posing challenges in the discovery of selective high-affinity ligands. Nevertheless, there is a growing interest to identifying modalities that specifically target RNA. [13,14] Disney *et al.* have been pioneers in developing strategies for discovering small molecules that target specific RNA motifs.[110] Given the expansive surface area of RNA scaffolds and their relatively low structural diversity, larger modalities like peptides and peptidomimetics emerge as intriguing options for achieving high-affinity and specific RNA targeting. This chapter delves into the exploration of peptidic modalities designed to target disease-related RNAs, such as pre-miRNAs and viral RNAs. For an extensive overview on cyclic peptides that target RNA, we refer to a recent review by Pal and 't Hart.[111]

2.3.1 miRNA

MiRNAs play pivotal roles in gene regulation and many miRNAs have been implicated in disease.[112–117] These small, approximately 22-nucleotide-long, single-stranded oligonucleotides primarily modulate the translation of specific mRNAs, often by inhibiting protein synthesis.[112,115,118] It is estimated that miRNAs influence over 60% of all genes. Elevated levels of specific miRNAs have been linked

to diseases such as cancer, metabolic disorders, vascular diseases, neurological conditions, and Alzheimer's disease, making them attractive targets for therapeutic interventions. [115,116,119,120] Antisense strategies have been proposed to target miRNAs, exemplified by the clinical candidate Miravirsen, blocking the hepatitis C virus miRNA-122 interaction.[121–123]

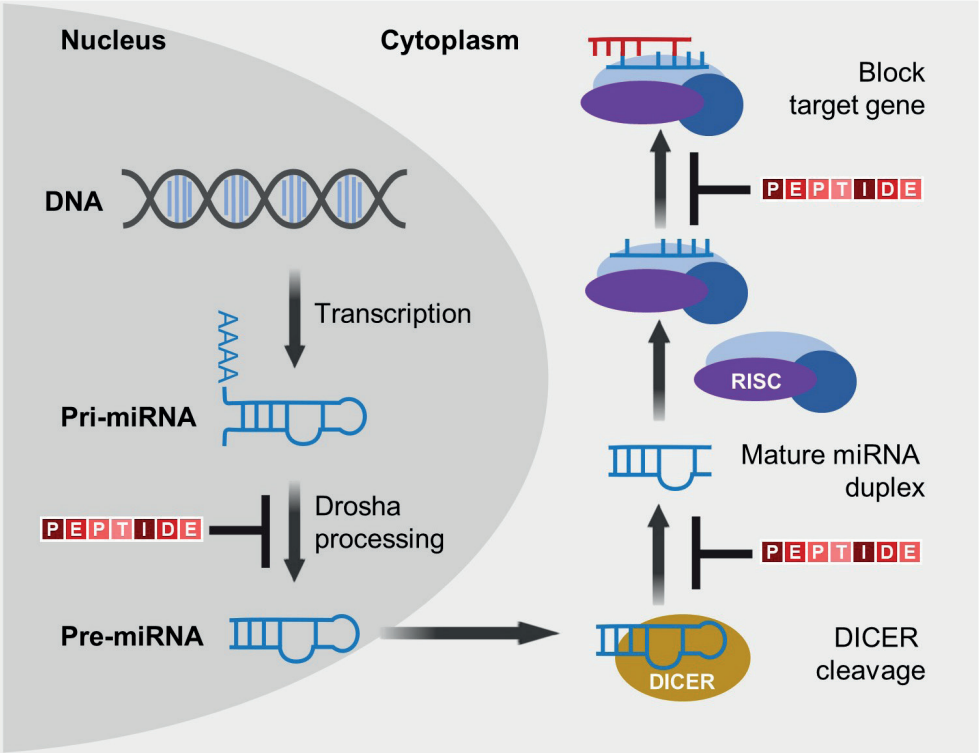
Two key precursors of mature miRNAs, namely primary miRNA (pri-miRNA) and precursor miRNA (pre-miRNA) hairpins, represent preferential targets for inhibiting miRNA activity.[112] Pri-miRNAs possess a large stem-loop structure that undergoes capping, polyadenylation, and cleavage by the RNAse Drosha in the final step, generating hairpin-shaped pre-miRNA in the nucleus. Subsequently, pre-miRNA is translocated from the nucleus to the cytoplasm where it undergoes cleavage by the RNAse Dicer resulting in the formation of mature miRNA duplexes. The guiding strand of these miRNA duplexes can interact with the RNA-induced silencing complex (RISC), forming the miRNA-RISC complex, which can bind to mRNA and either block or degrade targeted genes (**Figure 2.5**). The relatively defined structures of these miRNA precursors offer feasible targets for ligand development. Inhibiting the interaction between Drosha or Dicer and the RNA scaffolds blocks production of mature miRNA and with that miRNA activity. In this chapter we describe rational design and library-based strategies for the discovery of peptides targeting disease related pri- and pre-miRNAs (**Figure 2.6a**).

2.3.1.1 miR-21

miRNA-21 (miR-21) is widely implicated in cancer development, and therefore also referred to as an oncomir.[124] It blocks genes that are involved in apoptosis. Inhibition of miR-21 is therefore an interesting pharmaceutical approach. In the following sections we describe both rational design and combinatorial library approaches that led to peptidic miR-21 inhibitors.

Chirayil *et al.* performed a screening with a combinatorial peptoid library containing 7680 members to identify binders for the apical loop of pri-miR-21, the binding site of Drosha.[125,126] They identified a peptoid with an affinity of 1.9 μM against pri-miR-21 (**Figure 2.6b, left**).[126] However, in the presence of magnesium concentrations needed for microprocessor activity, which is necessary for RNA cleavage, no binding was observed. Building on these findings, Diaz and collaborators identified a novel peptoid through a library screen, utilizing the peptoid identified by Chirayil *et al.* as the library scaffold. [125] The screen resulted in a compact peptoid with an affinity of 12 μM (**Figure 2.6b, left**).

Bose *et al.* utilized phage display to discover high affinity peptides binding to pre-miR-21.[113] The best hit, a 12-mer peptide with the sequence ALWPPNLHAWVP demonstrated a binding affinity of 12.7 nM (**Figure 2.6b, upper middle**). The bulges



▲**Figure 2.5. The biogenesis of mature miRNA.** Pri-miRNA is formed by RNA polymerase II transcription in the nucleus and processed by Drosha resulting in pre-miRNA. After being transported to the cytoplasm, pre-miRNA is cleaved by DICER, generating mature miRNA which is subsequently incorporated into the RISC complex to degrade or block targeted genes. Maturation steps that can be inhibited with peptides are highlighted by a red peptide representation.

in the pre-miR-21 secondary structure are necessary for peptide recognition. Removal of the bulge near the terminal loop resulted in weak binding of the peptide, thereby demonstrating that the secondary structure is important for RNA recognition. The peptide was able to inhibit Dicer processing of pre-miR-21 to mature miR-21, as demonstrated by Dicer cleavage assays. Additionally, the TUNEL assays showed an increase of 30% in apoptosis caused by their hit compound.

Given the advantageous properties of macrocycles, such as stability and cell permeability,[127] the Varani group has focused on the design of macrocyclic peptides that can target RNA, including pre-miR-21.[128–130] They developed β-hairpin mimetics with a d-Pro/l-Pro template that are on average 12 amino acids in size.[128–130] This specific shape allows the peptides to mimic the orientation of the sidechains in an α-helix.[131] This strategy was used to identify peptides that can target pre-miR-21[129], pre-miR-20b[130], transactivation response (TAR) RNA and Rev Response Element (RRE) RNA.[131–138]

Dicer binds pre-miRNA at the apical loop. A library, which was originally designed for the BIV TAT RNA binding domain (RBD)[134] was screened against pre-miR-21 by Shortridge *et al.*[129] They identified a peptidomimetic that could bind the minor groove of pre-miR-21 with an affinity of 200 nM (**Figure 2.6b, lower middle**). L50 could also inhibit Dicer cleavage and prevent miR-21 maturation.

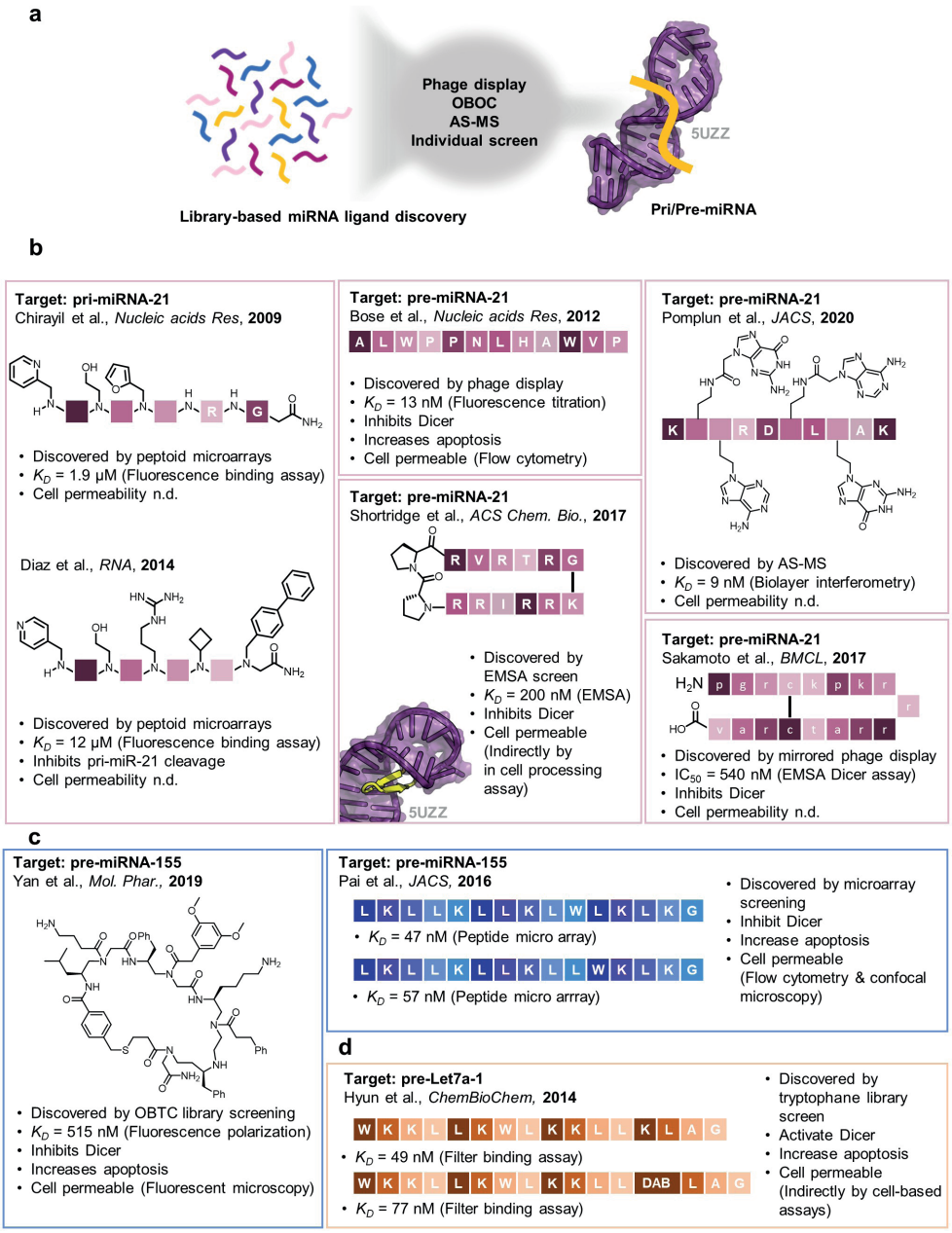
To identify a mirror image peptide that is able to inhibit miR-21, Sakamoto *et al.* screened a phage display library against l-pre-miR-21.[139] The obtained hit was resynthesized using d-amino acids and used in a Dicer assay (**Figure 2.6b, lower right**) that showed an IC_{50} of 540 nM and Dicer inhibition.

In 2020 Pomplun *et al.* designed a combinatorial library platform to target miRNA with biohybrid peptidomimetics displaying amino acid side chains and nuceobase functionalities.[140] Combining 12 nucleobase building blocks with 18 natural amino acids provided a combinatorial 100 million-membered library of 10-mer nucleobase peptides. The authors demonstrated how the nucleobase side chains increase overall affinity for nucleic targets, compared to using natural peptides. A hit compound with an affinity of 9 nM towards pre-miR-21 was identified from the selection. The biohybrid hit compound contains 4 nucleobases and 6 natural side chains. (**Figure 2.6b, upper right**) Overall this work shows that the RNA binders can be challenging as only one hit was selected from a library with 100 million compounds. At the same time the study points out the importance of ultralarge libraries and smart library design.

Structure-based approaches have been performed to design double-stranded RNA binders.[144,145] Kuepper *et al.* performed a structure-based approach to target mature miR-21 and thereby prevent miRNA-RISC interaction.[146] As template they used the *tomato aspermy virus* 2b (TAV2B) protein that can bind double-stranded RNA non-specifically and acts as a gene silencer.[146] Trimming this protein resulted in a peptide sequence that consists of 33 amino acids displaying an affinity of 1.19 μM towards palyndromic RNA (pal-RNA). Kuepper *et al.* showed that their peptidomimetic could bind miR-21 as well. Remarkably, the implementation of two hydrocarbon staples significantly improved affinity to 18 nM against double stranded miR-21, 75 nM against pre-miR-21 and 70 nM against a pal-RNA duplex. Additionally, the peptide was able to inhibit Dicer cleavage of pre-miR21. In another study the Grossmann group coupled the unstapled TAV2B peptidomimetics to complementary DNA to generate peptide-oligonucleotide conjugates that were able to form DNA/RNA/peptide complexes with single stranded RNA and make them more sequence specific.[147]

2.3.1.2 miR-155

Another miRNA with important implications in cancer is miR-155.[148] miR-155 is involved in apoptosis mechanisms and it has been shown that the inhibition of miR-



155 prevents cancer development in mice.[148] Several strategies for the targeting of pre-miR-155 using library screening techniques have been described. It has been shown that amphiphilic peptides have increased affinity towards hairpin RNA. [143,149,150] As such, Pai *et al.* developed a 185 membered amphiphilic peptide library based on the sequence LKKLLKLLKKLLKLKG, a peptide originally designed to

◀**Figure 2.6. Peptide sequences that target disease-related miRNAs. (a)** Different methods can be used for the screening of peptide-libraries ultimately resulting in pre/pri-miRNA binders (PDB 5UZZ). **(b-d)** Hit peptides identified in screenings against pri/pre-miR with their corresponding affinities and downstream effects. b) left: Chirayil *et al.* and Diaz *et al.*[125,126]; upper middle: Bose *et al.*[113]; lower middle: Shortridge *et al.*[129], PDB 5UZZ; upper right: Pomplun *et al.*[140]; lower right: Sakamoto *et al.*[139] ; c) left: Yan *et al.*[141] ; upper right: Pai *et al.*[142]; d) Hyun *et al.*[143]. Representations of crystal structures were made using PyMol. Dab = l-2,4- diaminobutyric acid. n.d. = not determined.

bind calmodulin[142], and screened it in a micro array format against fluorescent pre-miR-155.[114] This first screening round identified twelve peptides that displayed affinity towards pre-miR-155. A Dicer inhibition assay identified two peptides showing IC_{50}s of 0.77 μM and 1.6 μM and K_ds of 47 nM and 57 nM, respectively (**Figure 2.6c, right**). These two hits promoted apoptosis through the caspase pathway.

Yan *et al.* designed a cyclic γ-substituted-N-acylated-N-aminoethylamino acids (γ-AA) library using the one-bead-two-compounds (OBTC) strategy.[141,151] They screened a 320,000 membered library and identified a cyclic peptidomimetic able to bind pre-miR-155 and inhibit Dicer activity *in vitro* and *in vivo* with an IC_{50} of 22 μM (**Figure 2.6c, left**).

2.3.1.3 MiRNA Let7a-1 and miRNA-23b

Upregulation of miRNAs does not necessarily directly contribute to cancer development in all cases. Let7a-1 is an example of a miRNA that promotes caspase dependent apoptosis, thereby preventing uncontrolled cell growth.[152,153] Decreased let7a-1 maturation could lead to the development of cancer. Similar to the library designed by Pai et al,[114] an amphiphatic library was designed, in which in the sequence LKKLLKLLKKLLKLAG each leucine was replaced by a tryptophane.[143] Multiple scanning rounds resulted in the identification of two peptides binding pre-let7a-1 with affinities of 49 and 77 nM (**Figure 2.6d**). More importantly, one of these peptides increased Dicer induced pre-let7a-1 cleavage resulting in reduced activation of target genes in cells. The researchers suggest that binding of the peptidomimetic causes a conformational change in pre-Let7a-1, thereby making the Dicer cleavage site more accessible. This research demonstrates that peptidomimetics can also be utilized to promote Dicer activity.

Another miRNA involved in cancer development is miR-23b. An interesting design was developed by Krishnamurthy *et al.*[154] They developed helix-threading peptides (HTPs) using ring-closing metathesis. HTPs are heterocyclic peptides that can intercalate into RNA structures. RNA footprinting showed that two HTPs were able to selectively bind and protect hairpin RNA from nuclease activity. One peptide was able to bind pre-miR-23b with an affinity of 2.33 μM. However, as this peptide could bind multiple RNA sequences, more optimization is necessary for a selective

miR- 23b binder.

Various techniques exist for identifying peptidomimetics with selectivity towards miRNA. Library screening has proven to be a robust method uncovering multiple peptidomimetics capable of disrupting miRNA maturation at different stages of its biogenesis. While many strategies focus on the apical loop to inhibit Dicer or Drosha activity, peptidomimetics can also be employed to upregulate Dicer activity, influencing downstream processes. The growing accessibility of diverse unnatural amino acids further broadens the chemical space available for miRNA targeting.

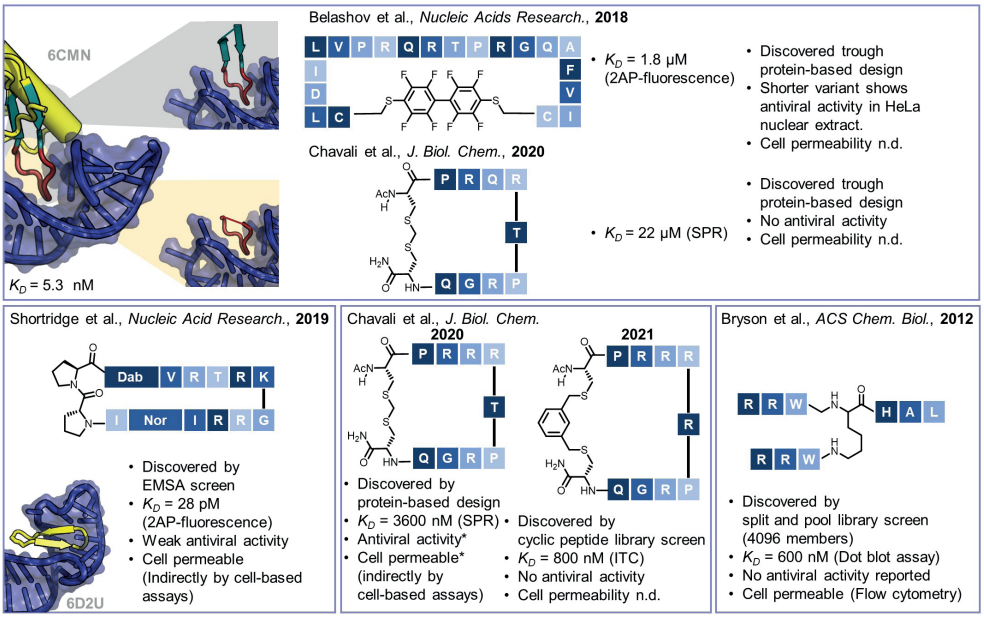
2.3.2 Viral RNA

Viral RNAs, especially those of the Human Immunodeficiency Virus type-1 (HIV-1), have been investigated as compelling targets for therapeutic intervention. Following the entry of HIV-1 into the host cell, its genomic RNA undergoes transcription into complementary DNA (cDNA) which is then transported to the nucleus and integrated into the host DNA.[155] RNA polymerase II (pol II) is then recruited to the HIV-1 promoter in the viral 5′ Long Terminal Repeat (LTR), allowing the transcription of the viral genome back into genomic RNA.[155] Two essential regulatory proteins, TAT and Rev, play a pivotal role in the growth and replication of HIV-1.[156] The effectiveness of both proteins relies on their binding to specific RNA elements; Tat binds to TAR RNA, and Rev binds to RRE RNA. Consequently, peptides that disrupt the TAT:TAR or Rev:RRE interaction have been explored as a potential strategy to address HIV-1 infection.[157]

2.3.2.1 Peptidomimetics targeting TAR

TAT, a cell penetrating transcriptional transactivator protein, plays a central role in the early development of HIV-1.[158] The target of TAT is TAR, which is an RNA hairpin located at the 5’ end of viral transcripts.[159] TAT interacts with the major groove of TAR through an arginine-rich motive (ARM). Upon binding TAR, TAT recruits the superelongation complex (SEC), which contains the positive transcription elongation factor b (P-TEFb). P-TEFb consist of Cdk9 and Cyclin T1 that can phosphorylate RNA pol II.[156,159] As such, when TAT binds the conserved TAR hairpin, Pol II is activated and the HIV-1 transcripts are formed. Whereas TAT can only bind to TAR, P-TEFb is involved in gene activation for multiple genes of the host. Therefore, TAT/TAR is an interesting pharmaceutical target.

As TAT/TAR binding is critical for the formation of HIV-1 mRNA in the early stages, much effort has been performed to block this interaction using peptidomimetics. Early publications focused on the analogous bovine immunodeficiency virus (BIV)-derived TAR stem-loop interaction with TAT because of limited structural information.[160,161] Later, linear peptides[149,150,162,163] or peptoids[164] were developed to block TAT/TAR



▲Figure 2.7. Peptides targeting TAR RNA with their corresponding affinities and downstream effects. Upper panel Belashov *et al.*[166] and Chavali *et al.*[167], PDB 6CMN; lower panel left Shortridge *et al.*[132], PDB 6D2U; middle Chavali *et al.*[167,168]; right Bryson *et al.*[169] Representations of crystal structures were generated using PyMol. Dab = l-2,4- diaminobutyric acid. Nor = norarginine. n.d. = not determined. *as reported by Chavali *et al.*, *J. Biol. Chem.* **2021**

or Rev/RRE. However, because of their high flexibility SARs remained limited. As previously mentioned, Varani and co-workers designed a d-Pro/l-Pro framework to target RNA using cyclic peptidomimetics. They were also successful in targeting RRE and TAR RNA using a similar cyclic framework, initially based on the BIV TAT peptide,[131–138] or a γ-AA framework.[163] To obtain TAT mimicking peptides, as starting point the BIV TAT was used and multiple optimization rounds were performed to design cyclic peptides with high affinity towards HIV TAR, with their most recent peptide targeting HIV-1 TAR within the picomolar range (**Figure 2.7, lower left**).[132] Nevertheless, cellular data showed only limited inhibition of binding between TAR RNA and P-TEFb. The researchers suggest that targeting the TAR loop alone is not sufficient to block the formation of the SEC.

High affinity binders towards RNA can be obtained using an RNA binding protein as a blueprint. Crawford *et al.* used laboratory evolution to alter the RNA Recognition Motif (RRM) of U1A to find a selective binder of the TAR hairpin.[165] They identified a selective TAR-Binding-Protein (TBP) with low nanomolar affinity that could interfere with TAT/TAR binding (**Figure 2.7, upper panel**). The Wedekind lab crystallized the most promising candidate, TBP6.7, in complex with the TAR hairpin which revealed an arginine trio located in a β2-β3 loop (**Figure 2.7, upper panel**).[166] Remarkably,

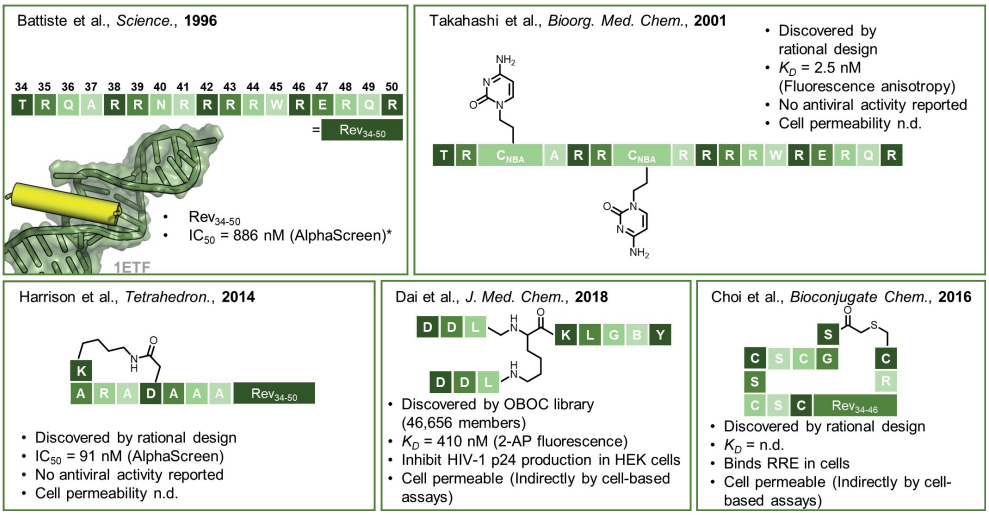
constrained isolated β2-β3 loop peptides could also bind TAR RNA with a K_D of 1.8 μM and inhibit TAT binding (**Figure 2.7, upper panel**).[166] Further research using variants of TBP6.7 revealed the importance of the spacing and position of the arginines located in the 'arginine fork'.[167] Moreover, cyclization of the β2-β3 loop of TBP6.9 resulted in an affinity of 3.6 μM against TAR RNA (**Figure 2.7, middle**). A follow-up cyclization study showed that napthalene and methylene linkers improved affinity.[168] Whereas the napthalene linker showed increased affinity, the methylene-linked peptides showed higher antiviral activity in cells and lower toxicity. One noteworthy observation was that introducing an Arg-to-Lys mutation in the cyclic peptide led to a tenfold reduction in TAR binding. This finding suggests that TBP-derived cyclic peptides employ an arginine-fork motif to recognize the TAR major groove. Moreover, these cyclic peptides demonstrated the ability to block the TAR binding by the TAT ARM and release the TAT ARM peptide from its bound state. This suggests a potential overlap between the site of cyclic peptide binding to TAR and TAT ARM-RNA binding site.

SPPS gives straightforward access to a vast chemical space and a variety of scaffolds beyond linear peptides. For example, branched peptides are able to have multivalent interactions with RNA that can provide high selectivity and affinity. Bryson *et al.* developed an on-bead library screen containing 4096 branched peptides, screened them against labeled TAR RNA and recovered hits using *de novo* sequencing by MALDI/MS resulting in the identification of a selective binder towards TAR RNA (**Figure 2.7, lower right**).[169] Most hits contained N-terminal arginines. The linear counterpart of the strongest binder showed a decrease in affinity, thereby illustrating the influence of branching. Later research demonstrated that these branched peptides are cell permeable.[170]

2.3.2.2 Peptidomimetics targeting RRE

Rev is involved in a later stage of HIV-1 development. Rev facilitates the transport of unspliced and single-spliced mRNAs from the nucleus to the cytoplasm where they encode structural viral proteins.[171,172] Similar to TAT, Rev also contains an ARM that provides cell permeability and RRE binding. A difference compared to TAT is that Rev can bind RNA multivalently.[172] The ARM of Rev (Rev_{34-50}) is in total 17 amino acids long and folds into an alpha helix in solution (**Figure 2.8, upper left**). [173,174] Rev_{34-50} can bind the stem-loop IIB of RRE in its major groove using a specific asparagine, threonine and four arginines.

Multiple studies have effectively induced alpha helicity in RRE binding proteins. The group of Frankel designed a combinatorial library against RRE[175,176] and showed that it binds partly as an alpha helix when bound to RNA.[177,178] One of their hit-compounds had a 15-fold increase in specificity towards RRE compared to Rev_{34-50}.[175] Remarkably, this peptide contained a proline in the middle of the helix but was



▲Figure 2.8. Peptides binding RRE RNA with their corresponding affinities and downstream effects. Upper left panel Battiste *et al.*[174], PDB 1ETF; Upper right panel Takahashi *et al.*[195]; Lower left panel Harrison *et al.*[181]; lower middle panel Dai *et al.*[189]; lower right panel Choi *et al.*[186] Representations of crystal structures were generated using PyMol. B= lysine boronic acid. n.d. = not determined. *as reported by Harrison *et al.*, *Tetrahedron.*, **2014**[181]

still able to partly form an α-helix, whereas the parent peptide adopted a random coil. Another combinatorial library using an ARM as framework identified the specific RRE binding sequence R_6QR_7.[179] Optimization of this sequence using a macrolactam staple resulted in an IC_{50} of 150 nM.[180] Harrison *et al.* also introduced a lactam bridge in their peptidomimetic design.[181] They designed Rev analogs by appending the small lactam constrained peptide, c-(1,5-*cyclo*)-[KARAD]-NH_2 known for inducing alpha helicity,[182] to the Rev_{34-50} peptide. To avoid steric clashes between the cyclic peptide and RNA, various linkers between the C- and N-terminus of Rev_{34-50} were modeled, resulting in a linker consisting of three alanines (**Figure 2.8, lower left**). Additionally, the lactam bridge was also introduced at non-essential positions inside the Rev_{34-50} sequence yielding eight peptides in total. The most promising peptide featured the N-terminal c-(1,5-*cyclo*)-[KARAD]-NH_2 sequence with a three residue linker. These works demonstrate that the introduction of a staple, such as a lactam bridge, can increase the affinity of peptidomimetics. However, the specific positioning of these staples is critical, as they can also interfere with binding.

In nature, alpha helicity can involve the use of metal ions. Zinc fingers, for example, fold around a zinc ion through the interaction of two conserved histidines and two conserved cysteines.[183] This DNA recognition motif adopts a structural configuration consisting of an α-helix and two β-sheets, with the α-helix specifically interacting with the major groove of the DNA. McColl *et al.* utilized this zinc finger design by introducing Rev_{34-50} in the alpha helix of the well-studied zinc finger 268, whereas

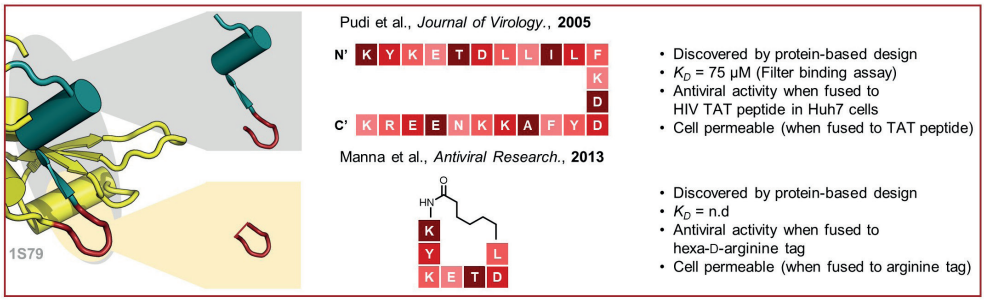
the β-sheets remained intact.[184] This peptidomimetic was able to recognize RRE *in vitro* and *in vivo*. Cowan and co-workers reported metal chelate Rev conjugates with tenfold higher binding affinities for the RRE compared to Rev.[185] These chelate-Rev conjugates formed coordination complexes with metals, and the arginine-rich Rev_{34-50} facilitated localization to the RRE stem loop-IIB site. Choi *et al.* engineered self-assembling cyclic peptide gold nanoparticles, incorporating the Rev ARM for RNA recognition and a cysteine-rich domain for gold binding (**Figure 2.8, lower right**).[186] These elements were strategically separated by a flexible linker. The cyclic and linear peptides showed similar affinities for RRE RNA in EMSA assays. However, the cyclic variant displayed enhanced selectivity and the capacity for multivalent binding to RRE RNA. Additionally, the cyclic peptide could recognize RRE RNA in both the cytoplasm and the nucleus, whereas its linear counterpart only recognized nuclear RRE RNA. These studies show the influence that metal ions can have on affinity and that they can be used as an alternative way to promote alpha helicity.

Branched peptide libraries were devised targeting RRE, incorporating boronic acid to facilitate reversible interaction with the 2'-hydroxyl of RNA. This led to the development of selective branched peptides with binding affinities in the micromolar range.[187–192] To further optimize binding, a branched peptide library was designed that introduced both boronic acid and acridine.[187] The latter was introduced because of its intercalating properties. Additionally, a library was designed that contained merely unnatural amino acids.[189] To determine RRE binding, the formation of p24, a HIV-1 capsid protein, was monitored resulting in seven hits in the micromolar range. The strongest binder had an affinity of 410 nM (**Figure 2.8, lower middle**).[189] Additionally, it was shown that this peptide binds not only to the stem-loop II regions of RRE, but also in the bulge region of stem loop I, competing for Rev protein primary and secondary binding site.[188] In conclusion, branched peptides exhibit promising results with regard to RRE binding. Their ability to perform multivalent and long-range interactions with RNA make them interesting scaffolds for the development of RNA binding therapeutics.

The Mihara group implemented NBA into Rev_{34-50}.[193–197], resulting in peptidomimetics that exhibited increased α-helicity in comparison to their PNA counterparts.[197] From a small library, hit compounds were designed that have affinity towards RRE in the nanomolar range and showed higher affinity towards RRE compared to the native Rev_{34-50}. Introduction of PNA moieties in Rev_{34-50} resulted in lower affinity towards RRE compared to Rev_{34-50} but still in the nanomolar range (**Figure 2.8, upper right**). [193,194] Overall, incorporating both PNA and NBA proved to be a compelling strategy for targeting RNA, potentially leading to the development of highly effective RNA binders.

3.2.3 Hepatitic C virus

Another virus that is considered a pharmaceutical target is Hepatitic C (HCV). HCV interacts with ribosomes through their internal ribosome entry sites (IRES).[198] In order to promote viral RNA translation, it is essential for HCV to also interact with the human La protein using their IRESs.[199] The human La protein is characterized as an initiator factor and can bind HCV RNA with its three RRM sites.[198] Pudi *et al.* isolated a 24-mer from the second RRM that was still able to interact with RNA (**Figure 2.9**). Moreover, this peptide could compete with the natural La protein and prevent HCV translation when fused to TAT. The peptide was trimmed down by Mondal *et al.* to its core sequence only consisting of seven amino acids eventually. [200] As the heptamer adopts a β-turn, Manna *et al.* prepared a cyclized variant (**Figure 2.9**). [201] To facilitate cell permeability, both the linear and cyclized heptamer were fused to hexa-arginine and administered to Huh7 cells. The cyclized peptide displayed an increase in affinity and bigger decrease in HCV RNA. Similar to the Wedekind lab this protein guided approach yielded a peptide that could interact and interfere with viral RNA maturation and proves that proteins can be utilized as a blueprint towards RNA binding peptidomimetics.



▲**Figure 2.9. Peptide binders of HCV IRES RNA,** based on the human La protein with their corresponding affinities and downstream effects developed by Pudi et al.[199] and Manna et al.[201], PDB 1S79. Representation of crystal structure was made using PyMol. n.d. = not determined.

2.4 Summary and outlook

In this review article we have summarized recent developments of peptides and peptidomimetics as modalities that can target different types of nucleic acids involved in disease. With their molecular weight and size in between small molecules and large biologics, peptides represent an intriguing option for recognizing DNA and RNA sequences or structures with specificity, while maintaining the potential to reach intracellular targets.

The two main discovery strategies for nucleic acid binding peptides are proteomimetic design and screening of large combinatorial libraries. In the proteomimetic design

strategy the DNA or RNA binding domains of natural proteins are excised to produce minimalistic peptide scaffolds. When isolated from their complete protein context these peptides often loose the secondary structure necessary for nucleic acid binding. Chemical modifications, such as stapling or cyclization, are often used to stabilize such structural features in the short peptides. These modifications often also help to improve certain properties of the peptides, such as cell penetration and proteolytic stability.

Several types of combinatorial libraries and screening or affinity selection platforms have been used to discover DNA and RNA binding peptides and peptidomimetics. These platforms include combinatorial one-bead-one-compound (or one-bead-two-compound) libraries, magnetic bead based affinity selections, phage display and microarrays. Interestingly, powerful platforms such as mRNA display or DNA encoded libraries, are almost never used in this context. The exposed nucleic acid tag, emblematic for these types of libraries, would make screenings for nucleic acid binding compounds rather complicated. However, with careful experimental design, DNA or mRNA encoded libraries can be used in specific cases.[202]

Selectivity is a major challenge for nucleic acid targeting modalities. Interestingly, in several studies, described in this review and elsewhere, peptides originally developed for one specific nucleic acid target, are used to address more or less unrelated sequences. Examples include: the miR-21 binder discovered by Pomplun *et al.*, binds to miR-155 with similar affinity;[140] the Grossmann group utilized peptides derived from the viral suppressor of RNA silencing, TAV2b, to target human miR-21;[146] the Arora group used DNA binding domains of TAV2b and the transcriptional repressor Max to design RNA binding modalities.[144] Based on these and other reports and on our own experience with nucleic acid binding modalities we would like to call for caution when it comes to making or interpreting claims about selectivity. Also natural TFs and other NA binding proteins usually have a certain degree of promiscuous generic affinity for any NA. However, in the complex cellular environment they ultimately recognize preferentially their specific target sequences. Assays the aim to showcase selectivity should be carefully designed and include biological evaluation in a cellular context.

Cell penetration is a general limitation for peptidic drug modalities. Many of the compounds reported in this review have not been tested for cell penetration or cellular activity. It is likely that these compounds might require extensive further development, in order to achieve bioactivity. On the other side a common trait of many nucleic acid binding peptides is a high arginine content which leads to polycationic sequences with intrinsic cell penetration. Remarkably, even very large polypeptides, such as Omomyc (~90 amino acids or ~180 amino acid residues for the dimer), display intrinsic cell penetrating activity triggered by their polycationic DNA binding domains.[26–28] A similar observation had been made by Carlos Barbas III,

who found multivalent zinc finger proteins to be, to a certain extent, intrinsically cell penetrating.[203,204] However, in order to achieve potent activity inside cells improved cell penetration of peptide-based modalities is necessary and will require further research and development.

Recent developments in engineering nucleic acid binding peptides have shown the potential of these modalities. A particularly prominent example is the development of Omomyc derivatives to treat MYC-driven cancers. An analog called Omomyc-103 is currently in clinical trials.[205] To date there are no approved drugs based on peptidic modalities targeting nucleic acids. Omomyc showcases the feasibility of the approach and the potential of peptidic modalities to address targets generally considered undruggable. With continuous developments in the field of peptide engineering and improved understanding on how to improve their pharmacokinetic properties,[127] we believe that peptide-based modalities can play an important role in identifying nucleic acid targeted therapies.

2.5 References

[1] G. F. Joyce, *Curr. Biol.* **1996**, *6*, 965–967.

[2] F. Crick, J. Watson, *Nature* **1953**, *171*, 737.

[3] F. Crick, *Nature* **1970**, *227*, 561.

[4] D. Varshney, J. Spiegel, K. Zyner, D. Tannahill, S. Balasubramanian, *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474.

[5] F. J. Slack, A. M. Chinnaiyan, *Cell* **2019**, *179*, 1033–1055.

[6] L. F. R. Gebert, I. J. MacRae, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 21–37.

[7] F. Spitz, E. E. M. Furlong, *Nat. Rev. Genet.* **2012**, *13*, 613–626.

[8] S. Venugopal, V. Sharma, A. Mehra, I. Singh, G. Singh, *Chem. Biol. Drug Des.* **2022**, *100*, 580–598.

[9] H. K. Liu, P. J. Sadler, *Acc. Chem. Res.* **2011**, *44*, 349–359.

[10] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.

[11] J. Lin, H. Nagase, *Biomolecules* **2020**, *10*, DOI 10.3390/biom10040544.

[12] Q. Vicens, J. S. Kieft, *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, 1–9.

[13] J. L. Childs-Disney, X. Yang, Q. M. R. Gibaut, Y. Tong, R. T. Batey, M. D. Disney, *Nat. Rev. Drug Discov.* **2022**, *21*, 736–762.

[14] A. Di Giorgio, M. Duca, *Medchemcomm* **2019**, *10*, 1242–1255.

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

[16] Y. Kim, *Biomol. Ther.* **2023**, *31*, 241–252.

[17] C. A. Gersbach, T. Gaj, C. F. Barbas, *Acc. Chem. Res.* **2014**, *47*, 2309–2318.

[18] S. Becker, J. Boch, *Gene Genome Ed.* **2021**, *2*, 100007.

[19] M. Chavez, X. Chen, P. B. Finn, L. S. Qi, *Nat. Rev. Nephrol.* **2023**, *19*, 9–22.

[20] A. P. Chandrasekaran, M. Song, K. S. Kim, S. Ramakrishna, *Different Methods of Delivering CRISPR/Cas9 Into Cells*, Elsevier Inc., **2018**.

[21] L. Wang, N. Wang, W. Zhang, X. Cheng, Z. Yan, G. Shao, X. Wang, R. Wang, C. Fu, *Signal Transduct. Target. Ther.* **2022**, *7*, DOI 10.1038/s41392-022-00904-4.

[22] W. Cabri, P. Cantelmi, D. Corbisiero, T. Fantoni, L. Ferrazzano, G. Martelli, A. Mattellone, A. Tolomelli, *Front. Mol. Biosci.* **2021**, *8*, 1–21.

[23] M. Pelay-Gimeno, A. Glas, O. Koch, T. N. Grossmann, *Angew. Chemie Int. Ed.* **2015**, *54*, 8896–8927.

[24] G. P. Smith, V. Petrenko, *Chem. Rev.* **1997**, 391–410.

[25] J. M. Mata, E. van der Nol, S. J. Pomplun, *J. Am. Chem. Soc.* **2023**, *145*, 19129–19139.

[26] M. E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. S. Del Pozo, C. Fiore, L. Foradada, V. C. Cano, M. Sánchez-Hervás, M. Guenther, E. R. Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. M. Alonso, J. R. Whitfield, P. Lavigne, L. Soucek, *Sci. Transl. Med.* **2019**, *11*, 1–14.

[27] M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, *J. Am. Chem. Soc.* **2021**, *143*, 11788–11798.

[28] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, *ACS Cent. Sci.* **2021**, *7*, 1408–1418.

[29] Z. Z. Brown, C. Mapelli, I. Farasat, A. V Shoultz, S. A. Johnson, F. Orvieto, A. Santoprete, E. Bianchi, A. B. McCracken, K. Chen, X. Zhu, M. J. Demma, B. M. Lacey, K. A. Canada, R. M. Garbaccio, J. O'Neil, A. Walji, *J. Org. Chem.* **2020**, *85*, 1466–1475.

[30] L. E. Canne, A. R. Ferré-D'Amaré, S. K. Burley, S. B. H. Kent, *J. Am. Chem. Soc.* **1995**, *117*, 2998–3007.

[31] R. Calo-Lapido, C. Penas, A. Jiménez-Balsa, M. E. Vázquez, J. L. Mascareñas, *Org. Biomol. Chem.* **2019**, *17*, 6748–6752.

[32] S. K. Madden, A. D. de Araujo, M. Gerhardt, D. P. Fairlie, J. M. Mason, *Mol. Cancer* **2021**, *20*, 1–18.

[33] J. Xu, G. Chen, A. T. De Jong, S. H. Shahravan, J. A. Shin, *J. Am. Chem. Soc.*

2009, *131*, 7839–7848.

[34] K. Usui, A. Okada, K. Kobayashi, N. Sugimoto, *Org. Biomol. Chem.* **2014**, *13*, 2022–2025.

[35] J. A. Schouten, S. Ladame, S. J. Mason, M. A. Cooper, S. Balasubramanian, *JACS* **2003**, 5594–5595.

[36] A. Minard, D. Morgan, F. Raguseo, A. di Porzio, D. Liano, A. G. Jamieson, M. di Antonio, *Chem. Commun.* **2020**, 8940–8943.

[37] T. Sharma, S. Saxena, S. Kaur, *Pept. Sci.* **2023**, 1–26.

[38] A. Travers, G. Muskhelishvili, *FEBS J.* **2015**, *282*, 2279–2295.

[39] S. Minchin, J. Lodge, *Essays Biochem* **2019**, *0*, 433–456.

[40] W. S. Wade, M. Mrksich, P. B. Dervan, *JACS* **1992**, *46*, 8783–8794.

[41] B. Cuenoud, A. Schepartz, *Science (80-.).* **1993**, *259*.

[42] M. I. Sμnchez, J. Mosquera, M. E. Vμzquez, J. L. Mascareças, *Angew. Chemie - Int. Ed.* **2014**, *53*, 9917–9921.

[43] S. Learte-Aymamí, N. Curado, J. Rodríguez, M. E. Vázquez, J. L. Mascareñas, *JACS* **2017**, DOI 10.1021/jacs.7b07422.

[44] A. Iyer, D. Van Lysebetten, Y. R. García, B. Louage, B. G. De Geest, A. Madder, *Biomol. Chem.* **2015**, *13*, 3856–3862.

[45] N. M. Karadkhelkar, M. Lin, L. M. Eubanks, K. D. Janda, *J. Am. Chem. Soc.* **2022**, DOI 10.1021/jacs.2c12732.

[46] C. V Dang, *Cell* **2012**, *149*, 22–35.

[47] S. R. Frank, T. Parisi, S. Taubert, P. Fernandez, M. Fuchs, H. Chan, D. M. Livingston, B. Amati, *EMBO Rep.* **2003**, *4*, 5–10.

[48] H. Chen, H. Liu, G. Qing, *Signal Transduct. Target. Ther.* **2018**, *3*, 1–7.

[49] J. R. Whitfield, M. Beaulieu, L. Soucek, *Front. cell Dev. Biol.* **2017**, *5*, 1–13.

[50] D. Massó-Vallés, L. Soucek, *Cells* **2020**, *9*, 883.

[51] N. B. Struntz, A. Chen, A. Deutzmann, R. M. Wilson, E. Stefan, H. L. Evans, M. A. Ramirez, T. Liang, F. Caballero, M. H. E. Wildschut, D. V Neel, D. B. Freeman, M. S. Pop, M. McConkey, S. Muller, B. H. Curtin, H. Tseng, K. R. Frombach, V. L. Butty, S. S. Levine, C. Feau, S. Elmiligy, J. A. Hong, T. A. Lewis, A. Vetere, P. A. Clemons, S. E. Malstrom, B. L. Ebert, C. Y. Lin, D. W. Felsher, A. N. Koehler, *Cell Chem. Biol.* **2019**, *26*, 711-723.e14.

[52] L. Boike, A. G. Cioffi, F. C. Majewski, J. Co, N. J. Henning, M. D. Jones, G. Liu, J. M. McKenna, J. A. Tallarico, M. Schirle, D. K. Nomura, *Cell Chem. Biol.* **2020**, 1–10.

[53] H. Han, A. D. Jain, M. I. Truica, J. Izquierdo-ferrer, J. F. Anker, B. Lysy, V. Sagar, *Cancer Cell* **2019**, *36*, 483-497.e15.

[54] X. Li, Z. Zhang, F. Gao, Y. Ma, D. Wei, Z. Lu, S. Chen, M. Wang, Y. Wang, K. Xu, R. Wang, F. Xu, J. Chen, C. Zhu, Z. Li, *JACS* **2023**, DOI 10.1021/jacs.3c02619.

[55] L. Soucek, M. Helmer-Citterich, A. Sacco, R. Jucker, G. Cesareni, S. Nasi, *Oncogene* **1998**, *17*, 2463–2472.

[56] L. A. Jung, A. Gebhardt, W. Koelmel, C. P. Ade, S. Walz, J. Kuper, B. Von Eyss, S. Letschert, C. Redel, L. Artista, A. Biankin, *Oncogene* **2017**, *36*, 1911–1924.

[57] L. Soucek, R. Jucker, L. Panacchia, R. Ricordy, F. Tatò, S. Nasi, *Cancer Res.* **2002**, *62*, 3507–3510.

[58] M. Savino, D. Annibali, N. Carucci, E. Favuzzi, M. D. Cole, I. Gerard, L. Soucek, S. Nasi, *PLoS One* **2011**, *6*, DOI 10.1371/journal.pone.0022284.

[59] L. Soucek, S. Nasi, G. I. Evan, *Cell Death Differ.* **2004**, *11*, 1038–1045.

[60] L. Soucek, J. Whitfield, C. P. Martins, A. J. Finch, D. J. Murphy, N. M. Sodir, A. N. Karnezis, L. B. Swigart, S. Nasi, G. I. Evan, *Nature* **2008**, *455*, 679–683.

[61] L. Soucek, J. R. Whitfield, N. M. Sodir, D. Masso, E. Serrano, A. N. Karnezis, L. B. Swigart, G. I. Evan, *Genes Dev.* **2013**, *27*, 504–513.

[62] N. M. Sodir, L. B. Swigart, A. N. Karnezis, D. Hanahan, G. I. Evan, L. Soucek, *Genes Dev.* **2011**, 907–916.

[63] D. Annibali, J. R. Whitfield, E. Favuzzi, T. Jauset, E. Serrano, I. Cuartas, S. Redondo-Campos, G. Folch, A. Gonzàlez-Juncà, N. M. Sodir, D. Massó-Vallés, M. E.

Beaulieu, L. B. Swigart, M. M. Mc Gee, M. P. Somma, S. Nasi, J. Seoane, G. I. Evan, L. Soucek, *Nat. Commun.* **2014**, *5*, 1–11.

[64] E. Wang, A. Sorolla, P. T. Cunningham, H. M. Bogdawa, S. Beck, E. Golden, R. E. Dewhurst, L. Florez, M. N. Cruickshank, K. Hoffmann, R. M. Hopkins, J. Kim, A. J. Woo, P. M. Watt, P. Blancafort, *Oncogene* **2019**, *38*, 140–150.

[65] M. J. Demma, C. Mapelli, A. Sun, S. Bodea, B. Ruprecht, S. Javaid, D. Wiswell, E. Muise, S. Chen, J. Zelina, F. Orvieto, A. Santoprete, S. Altezza, F. Tucci, E. Escandon, B. Hall, K. Ray, A. Walji, J. O'Neil, *Mol. Cell. Biol.* **2019**, *39*, 1–27.

[66] X. Lin, R. V. Nithun, R. Samanta, O. Harel, M. Jbara, *Org. Lett.* **2023**, *25*, 4715–4719.

[67] R. V. Nithun, Y. M. Yao, X. Lin, S. Habiballah, A. Afek, M. Jbara, *Angew. Chemie - Int. Ed.* **2023**, *62*, DOI 10.1002/anie.202310913.

[68] X. Lin, O. Harel, M. Jbara, *Angew. Chemie - Int. Ed.* **2023**, *202317511*, DOI 10.1002/anie.202317511.

[69] Z. Z. Brown, C. Mapelli, I. Farasat, A. V Shoultz, S. A. Johnson, F. Orvieto, A. Santoprete, E. Bianchi, A. B. Mccracken, K. Chen, X. Zhu, M. J. Demma, B. M. Lacey, K. A. Canada, R. M. Garbaccio, J. O. Neil, A. Walji, *J. Org. Chem.* **2020**, DOI 10.1021/acs.joc.9b02467.

[70] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, *Nat. Biotechnol.* **2023**, *41*, DOI 10.1038/s41587-022-01504-x.

[71] A. D. De Araujo, H. T. Nguyen, D. P. Fairlie, *ChemBioChem* **2021**, 1784–1789.

[72] F. Ahmadpour, R. Ghirlando, A. T. de Jong, M. Gloyd, J. A. Shin, A. Guarné, *PLoS One* **2012**, *7*, DOI 10.1371/journal.pone.0032136.

[73] L. C. Lustig, D. Dingar, W. B. Tu, C. Lourenco, M. Kalkat, I. Inamoto, R. Ponzielli, W. C. W. Chan, J. A. Shin, L. Z. Penn, *Oncogene* **2017**, *36*, 6830–6837.

[74] I. Inamoto, I. Sheoran, S. C. Popa, M. Hussain, J. A. Shin, *ACS Chem. Biol.* **2021**, *16*, 35–44.

[75] M. J. Demma, M. J. Hohn, A. Sun, C. Mapelli, B. Hall, A. Walji, J. O'Neil, *FEBS Lett.* **2020**, *594*, 1467–1476.

[76] Y. Ma, K. Iida, K. Nagasawa, *Biochem. Biophys. Res. Commun.* **2020**, *531*, 3–17.

[77] D. Sen, W. Gilbert, *Nature* **1990**, *344*.

[78] J. Spiegel, S. Adhikari, S. Balasubramanian, *Trends Chem.* **2020**, *2*, 123–136.

[79] J. Amato, N. Iaccarino, A. Randazzo, E. Novellino, B. Pagano, *ChemMedChem* **2014**, 2026–2030.

[80] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2005**, *33*, 2908–2916.

[81] V. T. Mukundan, A. T. Phan, *JACS* **2013**.

[82] V. S. Chambers, G. Marsico, J. M. Boutell, M. Di Antonio, G. P. Smith, S. Balasubramanian, *Nat. Biotechnol.* **2015**, *33*, 877–881.

[83] J. L. Huppert, S. Balasubramanian, *Nucleic Acids Res.* **2007**, *35*, 406–413.

[84] A. M. Zahler, J. R. Williamson, T. R. Cech, D. M. Prescott, *Nature* **1991**, *350*, 718–720.

[85] A. L. Moye, K. C. Porter, S. B. Cohen, T. Phan, K. G. Zyner, N. Sasaki, G. O. Lovrecz, J. L. Beck, T. M. Bryan, *Nat. Commun.* **2015**, DOI 10.1038/ncomms8643.

[86] T. Simonsson, P. Pecinka, M. Kubista, *Nucleic Acids Res.* **1998**, *26*, 1167–1172.

[87] A. Siddiqui-jain, C. L. Grand, D. J. Bearss, L. H. Hurley, *PNAS* **2002**, *99*, 11593–11598.

[88] S. Cogoi, L. E. Xodo, *Nucleic Acids Res.* **2006**, *34*, 2536–2549.

[89] M. Bejugam, S. Sewitz, P. S. Shirude, R. Shahid, S. Balasubramanian, T. Uni, V. Chemical, V. Uni, L. Road, *JACS* **2007**, 12926–12927.

[90] D. A. Sanders, M. Bejugam, M. Gunaratnam, M. Sebastian, S. Sewitz, J. A. Fletcher, S. Neidle, S. Balasubramanian, *ACS Med. Chem. Lett.* **2010**, 306–310.

[91] V. González, L. H. Hurley, *Biochemistry* **2010**, *49*, 9706–9714.

[92] M. C. Chen, R. Tippana, N. A. Demeshkina, P. Murat, S.

Balasubramanian, S. Myong, A. R. F. Amaré, *Nature* **2018**, *558*, 465–469.

[93] J. J. Green, S. Ladame, L. Ying, D. Klenerman, S. Balasubramanian, *JACS* **2006**, 9809–9812.

[94] K. C. Liu, K. R. Der, C. Mayer, S. Adhikari, D. J. Wales, *JACS* **2020**, *142*, 8367–8373.

[95] N. Brodyagin, M. Katkevics, V. Kotikam, C. A. Ryan, E. Rozners, *Beilstein J. Org. Chem.* **2021**, *17*, 1641–1688.

[96] K. Usui, A. Okada, S. Sakashita, M. Shimooka, *Molecules* **2017**, *22*, 1–12.

[97] X. Mou, C. K. Kwok, *JACS* **2023**, *145*, DOI 10.1021/jacs.3c04534.

[98] T. Sharma, N. Kundu, S. Kaur, A. Chakraborty, A. K. Mahto, R. P. Dewangan, J. Shankaraswamy, S. Saxena, *RSC Adv.* **2022**, *12*, 21760–21769.

[99] C. Barefield, J. Karlseder, *Nucleic Acids Res.* **2012**, *40*, 7358–7367.

[100] S. Tyagi, S. Saxena, N. Kundu, T. Sharma, A. Chakraborty, S. Kaur, D. Miyoshi, J. Shankaraswamy, *RSC Adv.* **2019**, *9*, 40255–40262.

[101] N. Kundu, T. Sharma, S. Kaur, M. Singh, V. Kumar, U. Sharma, A. Jain, J. Shankaraswamy, D. Miyoshi, S. Saxena, *RSC Adv.* **2022**, *12*, 7594–7604.

[102] N. Kundu, T. Sharma, S. Kaur, A. K. Mahto, R. P. Dewangan, J. Shankaraswamy, S. Saxena, *J. Biomol. Struct. Dyn.* **2023**, *41*, 7119–7127.

[103] T. Sharma, N. Kundu, S. Kaur, V. Tandon, S. Saxena, *J. Biomol. Struct. Dyn.* **2023**, *41*, 9977–9986.

[104] C. K. Kwok, Y. Tang, S. M. Assmann, P. C. Bevilacqua, *Trends Biochem. Sci.* **2015**, *40*, 221–232.

[105] Y. Wan, M. Kertesz, R. C. Spitale, E. Segal, H. Y. Chang, *Nat. Rev. Genet.* **2011**, *12*, 641–655.

[106] R. C. Spitale, R. A. Flynn, E. A. Torre, E. T. Kool, H. Y. Chang, *Wiley Interdiscip. Rev. RNA* **2014**, *5*, 867–881.

[107] P. C. Bevilacqua, L. E. Ritchey, Z. Su, S. M. Assmann, *Annu. Rev. Genet.* **2016**, *50*, 235–266.

[108] B. Liu, J. L. Childs-Disney, B. M. Znosko, D. Wang, M. Fallahi, S. M. Gallo, M. D. Disney, *BMC Bioinformatics* **2016**, *17*, 112.

[109] S. E. Butcher, A. M. Pyle, *Acc. Chem. Res.* **2011**, *44*, 1302–1311.

[110] M. D. Disney, B. G. Dwyer, J. L. Childs-Disney, *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a034769.

[111] S. Pal, P. 't Hart, *Front. Mol. Biosci.* **2022**, *9*, 883060.

[112] S. Lin, R. I. Gregory, *Nat. Rev. Cancer 2015 156* **2015**, *15*, 321–333.

[113] D. Bose, S. Nahar, M. K. Rai, A. Ray, K. Chakraborty, S. Maiti, *Nucleic Acids Res.* **2015**, *43*, 4342–4352.

[114] J. Pai, S. Hyun, J. Y. Hyun, S. H. Park, W. J. Kim, S. H. Bae, N. K. Kim, J. Yu, I. Shin, *J. Am. Chem. Soc.* **2016**, *138*, 857–867.

[115] M. Esteller, *Nat. Rev. Genet.* **2011**, *12*, 861–874.

[116] A. Zampetaki, M. Mayr, *Circ. Res.* **2012**, *110*, 508–522.

[117] V. Rottiers, A. M. Näär, *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 239–250.

[118] F. Wahid, A. Shehzad, T. Khan, Y. Y. Kim, *Biochim. Biophys. Acta - Mol. Cell Res.* **2010**, *1803*, 1231–1243.

[119] L.-A. MacFarlane, P. R. Murphy, *Curr. Genomics* **2010**, *11*, 537.

[120] N. M. McLoughlin, C. Mueller, T. N. Grossmann, *Cell Chem. Biol.* **2018**, *25*, 19–29.

[121] H. L. A. Janssen, H. W. Reesink, E. J. Lawitz, S. Zeuzem, M. Rodriguez-Torres, K. Patel, A. J. van der Meer, A. K. Patick, A. Chen, Y. Zhou, R. Persson, B. D. King, S. Kauppinen, A. A. Levin, M. R. Hodges, *N. Engl. J. Med.* **2013**, *368*, 1685–1694.

[122] M. Lindow, S. Kauppinen, *J. Cell Biol.* **2012**, *199*, 407–412.

[123] C. Chakraborty, A. R. Sharma, G. Sharma, S. S. Lee, *J. Adv. Res.* **2021**, *28*, 127–138.

[124] Y. H. Feng, C. J. Tsao, *Biomed. Reports* **2016**, *5*, 395–402.

[125] J. P. Diaz, R. Chirayil, S. Chirayil, M. Tom, K. J. Head, K. J. Luebke, *Rna* **2014**, *20*, 528–539.

[126] S. Chirayil, R. Chirayil, K. J. Luebke, *Nucleic Acids Res.* **2009**, *37*, 5486–5497.

[127] X. Ji, A. L. Nielsen, C. Heinis, *Angew. Chemie Int. Ed.* **2023**, *202308251*, DOI 10.1002/anie.202308251.

[128] J. A. Robinson, *Acc. Chem. Res.* **2008**, *41*, 1278–1288.

[129] M. D. Shortridge, M. J. Walker, T. Pavelitz, Y. Chen, W. Yang, G. Varani, *ACS Chem. Biol.* **2017**, *12*, 1611–1620.

[130] Y. T. Sun, M. D. Shortridge, G. Varani, *ChemBioChem* **2019**, *20*, 931–939.

[131] K. Moehle, Z. Athanassiou, K. Patora, A. Davidson, G. Varani, J. A. Robinson, *Angew. Chemie Int. Ed.* **2007**, *46*, 9101–9104.

[132] M. D. Shortridge, P. T. Wille, A. N. Jones, A. Davidson, J. Bogdanovic, E. Arts, J. Karn, J. A. Robinson, G. Varani, *Nucleic Acids Res.* **2019**, *47*, 1523–1531.

[133] T. C. Leeper, Z. Athanassiou, R. L. A. Dias, J. A. Robinson, G. Varani, *Biochemistry* **2005**, *44*, 12362–12372.

[134] Z. Athanassiou, R. L. A. Dias, K. Moehle, N. Dobson, G. Varani, J. A. Robinson, *J. Am. Chem. Soc.* **2004**, *126*, 6906–6913.

[135] A. Davidson, T. C. Leeper, Z. Athanassiou, K. Patora-Komisarska, J. Karn, J. A. Robinson, G. Varani, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 11931–11936.

[136] M. S. Lalonde, M. A. Lobritz, A. Ratcliff, M. Chamanian, Z. Athanassiou, M. Tyagi, J. Wong, J. A. Robinson, J. Karn, G. Varani, E. J. Arts, *PLOS Pathog.* **2011**, *7*, e1002038.

[137] Z. Athanassiou, K. Patora, R. L. A. Dias, K. Moehle, J. A. Robinson, G. Varani, *Biochemistry* **2007**, *46*, 741–751.

[138] A. Davidson, K. Patora-Komisarska, J. A. Robinson, G. Varani, *Nucleic Acids Res.* **2011**, *39*, 248–256.

[139] K. Sakamoto, K. Otake, T. Umemoto, *Bioorg. Med. Chem. Lett.* **2017**, *27*, 826–828.

[140] S. Pomplun, Z. P. Gates, G. Zhang, A. J. Quartararo, B. L. Pentelute, **2020**, DOI 10.1021/jacs.0c08964.

[141] H. Yan, M. Zhou, U. Bhattarai, Y. Song, M. Zheng, J. Cai, F. Sen Liang, *Mol. Pharm.* **2019**, *16*, 914–920.

[142] J. A. Cox, M. Comte, J. E. Fitton, W. F. DeGrado, *J. Biol. Chem.* **1985**, *260*, 2527–2534.

[143] S. Hyun, A. Han, M. H. Jo, S. Hohng, J. Yu, *ChemBioChem* **2014**, *15*, 1651–1659.

[144] J. G. Kwok, Z. Yuan, P. S. Arora, *Angew. Chemie Int. Ed.* **2023**, e202308650.

[145] N. M. McLoughlin, M. A. Albers, E. Collado Camps, J. Paulus, Y. A. Ran, S. Neubacher, S. Hennig, R. Brock, T. N. Grossmann, *Angew. Chemie Int. Ed.* **2023**, DOI 10.1002/ANIE.202308028.

[146] A. Kuepper, N. M. McLoughlin, S. Neubacher, A. Yeste-Vázquez, E. C. Camps, C. Nithin, S. Mukherjee, L. Bethge, J. M. Bujnicki, R. Brock, S. Heinrichs, T. N. Grossmann, *Nucleic Acids Res.* **2021**, *49*, 12622.

[147] N. M. McLoughlin, A. Kuepper, S. Neubacher, T. N. Grossmann, *Chem. – A Eur. J.* **2021**, *27*, 10477–10483.

[148] C. J. Cheng, R. Bahal, I. A. Babar, Z. Pincus, F. Barrera, C. Liu, A. Svoronos, D. T. Braddock, P. M. Glazer, D. M. Engelman, W. M. Saltzman, F. J. Slack, *Nature* **2015**, *518*, 107–110.

[149] S. Hyun, J. K. Hyun, J. L. Nam, H. L. Kyung, Y. Lee, R. A. Dae, K. Kim, S. Jeong, J. Yu, *J. Am. Chem. Soc.* **2007**, *129*, 4514–4515.

[150] Y. Lee, S. Hyun, H. J. Kim, J. Yu, *Angew. Chemie Int. Ed.* **2008**, *47*, 134–137.

[151] Y. Shi, S. Challa, P. Sang, F. She, C. Li, G. M. Gray, A. Nimmagadda, P. Teng, T. Odom, Y. Wang, A. van der Vaart, Q. Li, J. Cai, *J Med Chem* **2017**, *60*, 9290–9298.

[152] W. P. Tsang, T. T. Kwok, *Apoptosis* **2008**, *13*, 1215–1222.

[153] S. Roush, F. J. Slack, *Trends Cell Biol.* **2008**, *18*, 505–516.

[154] M. Krishnamurthy, K. Simon, A. M. Orendt, P. A. Beal, *Angew. Chemie - Int. Ed.* **2007**, *46*, 7044–7047.

[155] A. P. Rice, *Curr. Pharm. Des.* **2017**, *23*, DOI 10.2174/1381612823666170704130635.

[156] L. Chaloin, F. Smagulova, E. Hariton-Gazal, L. Briant, A. Loyter, C. Devaux, *J.*

Biomed. Sci. **2007**, *14*, 565–584.

[157] M. J. Gait, J. Karn, *Trends Biochem. Sci.* **1993**, *18*, 255–259.

[158] M. Green, M. Ishino, P. M. Loewenstein, *Cell* **1989**, *58*, 215–223.

[159] M. Ott, M. Geyer, Q. Zhou, *Cell Host Microbe* **2011**, *10*, 426–435.

[160] J. B. H. Tok, R. C. Des Jean, J. Fenker, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 43–46.

[161] L. A. Pallansch, C. S. Lackman-Smith, M. A. Gonda, *J. Virol.* **1992**, *66*, 2647–2652.

[162] M. A. Gelman, S. Richter, H. Cao, N. Umezawa, S. H. Gellman, T. M. Rana, *Org. Lett.* **2003**, *5*, 3563–3565.

[163] Y. Niu, A. J. Jones, H. Wu, G. Varani, J. Cai, *Org. Biomol. Chem.* **2011**, *9*, 6604.

[164] F. Hamy, E. R. Felder, G. Heizmann, J. Lazdins, F. Aboul-Ela, G. Varani, J. Karn, T. Klimkait, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 3548–3553.

[165] D. W. Crawford, B. D. Blakeley, P.-H. Chen, C. Sherpa, S. F. J. Le Grice, I. A. Laird-Offringa, B. R. McNaughton, *ACS Chem. Biol.* **2016**, *11*, 2206–2215.

[166] I. A. Belashov, D. W. Crawford, C. E. Cavender, P. Dai, P. C. Beardslee, D. H. Mathews, B. L. Pentelute, B. R. McNaughton, J. E. Wedekind, *Nucleic Acids Res.* **2018**, *46*, 6401–6415.

[167] S. S. Chavali, S. M. Mali, J. L. Jenkins, R. Fasan, J. E. Wedekind, *J. Biol. Chem.* **2020**, *295*, 16470–16486.

[168] S. S. Chavali, S. M. Mali, R. Bonn, A. S. Anitha, R. P. Bennett, H. C. Smith, R. Fasan, J. E. Wedekind, *J. Biol. Chem.* **2021**, *297*, 101390.

[169] D. I. Bryson, W. Zhang, W. K. Ray, W. L. Santos, *Mol. Biosyst.* **2009**, *5*, 1070.

[170] D. I. Bryson, W. Zhang, P. M. McLendon, T. M. Reineke, W. L. Santos, *ACS Chem. Biol.* **2012**, *7*, 210–217.

[171] J. Fernandes, B. Jayaraman, A. Frankel, *RNA Biol.* **2012**, *9*, 6–11.

[172] V. W. Pollard, M. H. Malim, *Annu. Rev. Microbiol.* **1998**, *52*, 491–532.

[173] R. Tan, L. Chen, J. A. Buettner, D. Hudson, A. D. Frankel, *Cell* **1993**, *73*, 1031–1040.

[174] J. L. Battiste, H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, A. D. Frankel, J. R. Williamson, *Science (80-.).* **1996**, *273*, 1547–1551.

[175] K. Harada, S. S. Martin, R. Tan, A. D. Frankel, *Proc. Natl. Acad. Sci.* **1997**, *94*, 11887–11892.

[176] K. Harada, S. S. Martin, A. D. Frankel, *Nature* **1996**, *380*, 175–179.

[177] Q. Zhang, K. Harada, H. S. Cho, A. D. Frankel, D. E. Wemmer, *Chem. Biol.* **2001**, *8*, 511–520.

[178] Y. Gosser, T. Hermann, A. Majumdar, W. Hu, R. Frederick, F. Jiang, W. Xu, D. J. Patel, *Nat. Struct. Biol. 2001 82* **2001**, *8*, 146–150.

[179] R. Tan, A. D. Frankel, *Proc. Natl. Acad. Sci.* **1998**, *95*, 4247–4252.

[180] N. L. Mills, M. D. Daugherty, A. D. Frankel, R. K. Guy, *J. Am. Chem. Soc.* **2006**, *128*, 3496–3497.

[181] R. S. Harrison, N. E. Shepherd, H. N. Hoang, R. L. Beyer, G. Ruiz-Gómez, M. J. Kelso, W. Mei Kok, T. A. Hill, G. Abbenante, D. P. Fairlie, *Tetrahedron* **2014**, *70*, 7645–7650.

[182] N. E. Shepherd, H. N. Hoang, G. Abbenante, D. P. Fairlie, *J. Am. Chem. Soc.* **2005**, *127*, 2974–2983.

[183] Q. Liu, D. J. Segal, J. B. Ghiara, C. F. Barbas, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 5525–5530.

[184] D. J. McColl, C. D. Honchell, A. D. Frankel, *Proc. Natl. Acad. Sci.* **1999**, *96*, 9521–9526.

[185] J. C. Joyner, J. A. Cowan, *J. Am. Chem. Soc.* **2011**, *133*, 9912–9922.

[186] J. S. Choi, S. Han, H. Kim, Y. Lim, *Bioconjug. Chem.* **2016**, *27*, 799–808.

[187] J. E. Wynn, W. Zhang, D. M. Tebit, L. R. Gray, M.-L. Hammarskjold, D. Rekosh, W. L. Santos, *Medchemcomm* **2016**, *7*, 1436–1440.

[188] Y. Dai, A. N. Peralta, J. E. Wynn, C. Sherpa, H. Li, A. Verma, S. F. J. Le Grice, W. L. Santos, *Bioorg. Med. Chem.* **2019**, *27*, 1759–1765.

[189] Y. Dai, J. E. Wynn, A. N. Peralta, C.

Sherpa, B. Jayaraman, H. Li, A. Verma, A. D. Frankel, S. F. Le Grice, W. L. Santos, *J Med Chem* **2018**, *61*, 9611–9620.

[190] J. E. Wynn, W. Zhang, D. M. Tebit, L. R. Gray, M.-L. Hammarskjold, D. Rekosh, W. L. Santos, *Bioorg. Med. Chem.* **2016**, *24*, 3947–3952.

[191] W. Zhang, D. I. Bryson, J. B. Crumpton, J. Wynn, W. L. Santos, *Org. Biomol. Chem.* **2013**, *11*, 6263–6271.

[192] W. Zhang, D. I. Bryson, J. B. Crumpton, J. Wynn, W. L. Santos, *Chem. Commun.* **2013**, *49*, 2436–2438.

[193] T. Takahashi, I. Kumagai, K. Hamasaki, A. Ueno, H. Mihara, *Nucleic Acids Symp. Ser.* **1999**, *42*, 271–272.

[194] I. Kumagai, T. Takahashi, K. Hamasaki, A. Ueno, H. Mihara, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 377–379.

[195] T. Takahashi, K. Hamasaki, A. Ueno, H. Mihara, *Bioorg. Med. Chem.* **2001**, *9*, 991–1000.

[196] I. Kumagai, T. Takahashi, K. Hamasaki, A. Ueno, H. Mihara, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1169–1172.

[197] T. Takahashi, K. Hamasaki, I. Kumagai, A. Ueno, H. Mihara, *Chem. Commun.* **2000**, *1*, 349–350.

[198] S. Ali, A. Siddiqui, *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2249–2254.

[199] R. Pudi, S. S. Ramamurthy, S. Das, *J. Virol.* **2005**, *79*, 9842–9853.

[200] T. Mondal, U. Ray, A. K. Manna, R. Gupta, S. Roy, S. Das, *J. Virol.* **2008**, *82*, 11927–11938.

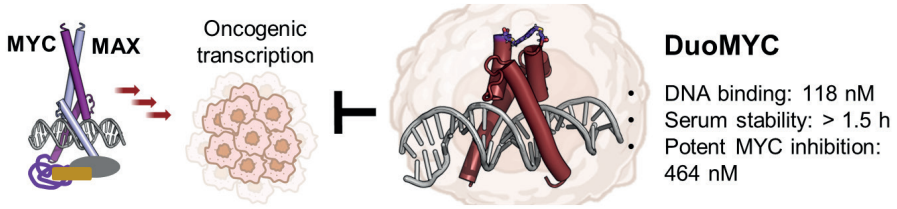
[201] A. K. Manna, A. Kumar, U. Ray, S. Das, G. Basu, S. Roy, *Antiviral Res.* **2013**, *97*, 223–226.

[202] R. I. Benhamou, B. M. Suresh, Y. Tong, W. G. Cochrane, V. Cavett, S. Vezina-Dawod, D. Abegg, J. L. Childs-Disney, A. Adibekian, B. M. Paegel, M. D. Disney, *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, 1–8.

[203] J. Liu, T. Gaj, M. C. Wallen, C. F. Barbas, *Mol. Ther. - Nucleic Acids* **2015**, *4*, e232.

[204] T. Gaj, J. Liu, K. E. Anderson, S. J. Sirk, C. F. Barbas, *ACS Chem. Biol.* **2014**, *9*, 1662–1667.

[205] E. Garralda, M. E. Beaulieu, V. Moreno, S. Casacuberta-Serra, S. Martínez-Martín, L. Foradada, G. Alonso, D. Massó-Vallés, S. López-Estévez, T. Jauset, E. Corral de la Fuente, B. Doger, T. Hernández, R. Perez-Lopez, O. Arqués, V. Castillo Cano, J. Morales, J. R. Whitfield, M. Niewel, L. Soucek, E. Calvo, *Nat. Med.* **2024**, DOI 10.1038/s41591-024-02805-1.

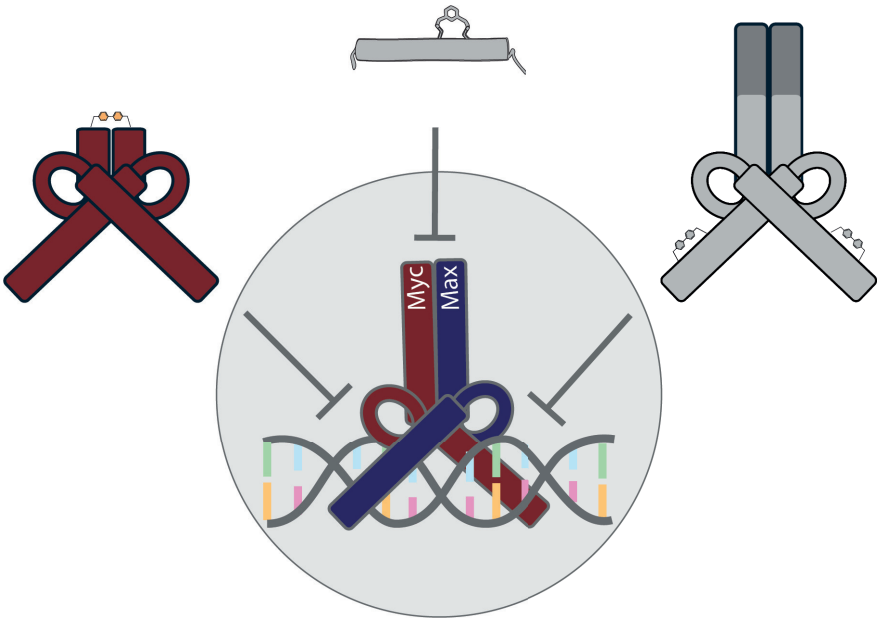


Adapted from: B. D. Ellenbroek, J. P. Kahler, D. Arella, C. Lin, W. Jespers, E. A. Züger, M. Drukker, S. J. Pomplun, "Development of DuoMYC: a synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC" *Angewandte Chemie International Edition* **2025**, *64*, DOI 10.1002/anie.202416082.

The data not included in this thesis (uncut gels and list of differently expressed genes) can be found online: https://onlinelibrary.wiley.com/doi/10.1002/anie.202416082

Chapter 3

Development of DuoMYC: A synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC



Authors:

Brecht D. Ellenbroek[‡], Dr. Jan Pascal Kahler[‡], Damiano Arella, Cherina Lin, Dr. Willem Jespers, Eliane Ann-Katrin Züger, Dr. Micha Drukker, Dr. Sebastian J. Pomplun

[‡]These authors contributed equally.

Abstract

The master regulator transcription factor MYC is implicated in numerous human cancers, and its targeting is a long-standing challenge in drug development. MYC is a typical 'undruggable' target, with no binding pockets on its DNA binding domain and extensive intrinsically disordered regions. Rather than trying to target MYC directly with classical modalities, here we engineer synthetic miniproteins that can bind to MYC's target DNA, the enhancer box (E-Box), and potently inhibit MYC-driven transcription. We crafted the miniproteins via structure-based design and a combination of solid phase peptide synthesis and site-specific crosslinking. Our lead variant, DuoMYC, binds to EBox DNA with high affinity (K_D ~ 0.1 μM) and is able to enter cells and inhibit MYC-driven transcription with submicromolar potency (IC_{50} = 464 nM) as shown by reporter gene assay and confirmed by RNA sequencing. Notably, DuoMYC surpasses the efficacy of several other recently developed MYC inhibitors. Our results highlight the potential of engineered synthetic protein therapeutics for addressing challenging intracellular targets.

3.1 Introduction

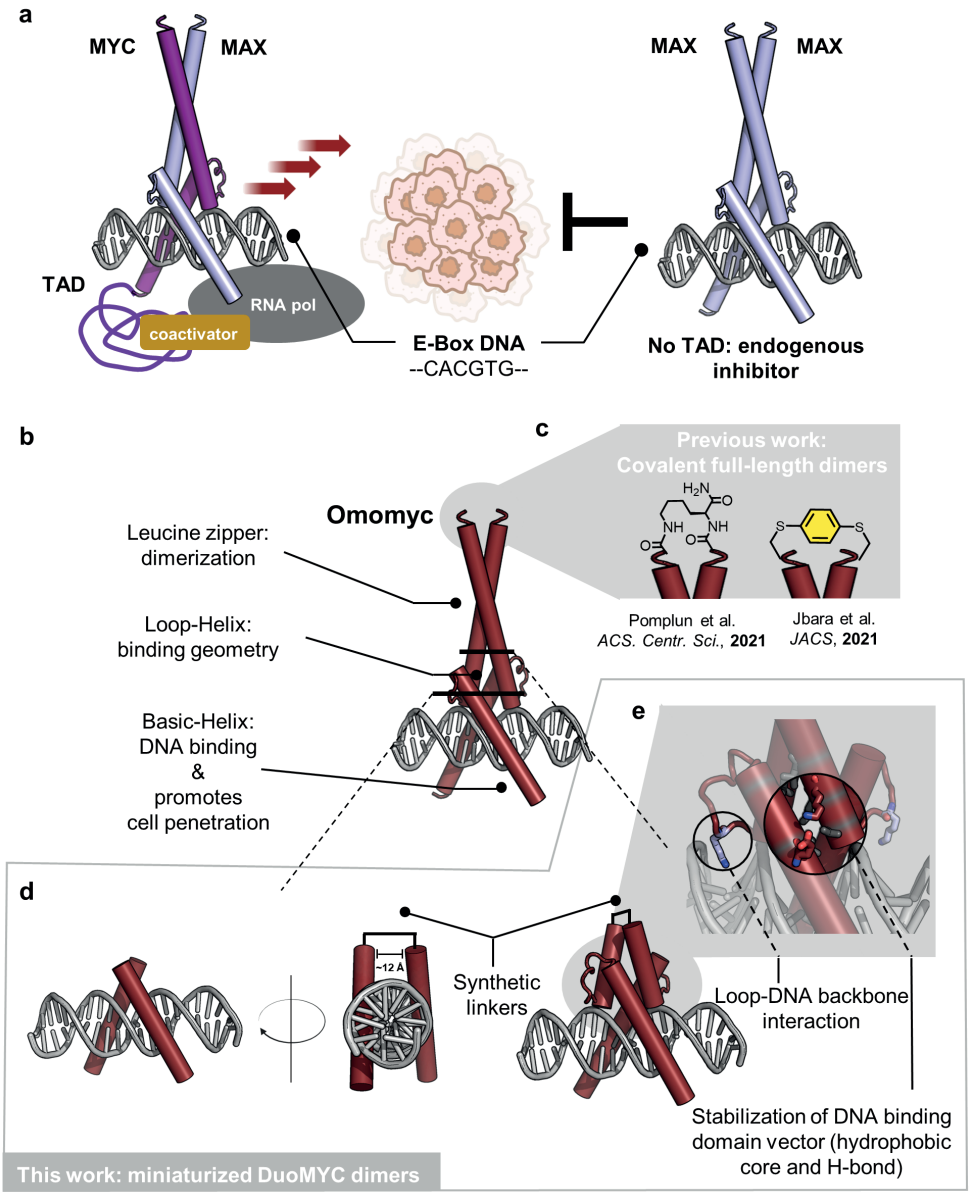
Transcription factors (TFs) orchestrate the delicate equilibrium between gene activation and repression and are crucial for cellular function.[1] TFs mainly act via protein-protein and protein-nucleic acid interaction and display extensive intrinsically disordered regions used to recruit components of the transcriptional machinery.[2] These traits make TFs challenging drug targets[3] and highlight the need for innovative approaches in drug discovery and design.

MYC is a prime example of an undruggable TF. It is involved in the pathomechanisms of over 50% of all human cancers.[4–7] To exert its activity, MYC heterodimerizes with its partner MAX and as a complex they bind to enhancer box (E-Box) DNA.[8] MYC's intrinsically disordered transactivation domain (TAD) then recruits the transcriptional machinery and activates gene programs related to cell growth, proliferation and survival.[9] MAX can also homodimerize and occupy the same E-Box sequence as a non-productive inhibitory complex.[10–13] Under healthy conditions MAX/MAX and MYC/MAX are well balanced, while MYC overexpression is involved in many types of cancer. The basic helix-loop-helix leucine zipper motif (bHLH-LZ) of MYC and its intrinsically disordered TAD do not offer any obvious binding sites for inhibitors and no MYC targeted therapeutics are available for patients, to date.

Only recently, a few small molecules with innovative modes of action have shown potential in initial investigations. These include the MAX/MAX stabilizer KI-MS2-008, developed in the Koehler lab,[14] and two compounds inducing MYC instability and degradation: EN4, discovered by Nomura and coworkers,[15] and MYCi361, reported by the group of Abdulkadir.[16] In addition to these approaches, a MYC-targeting bicyclic peptide was recently discovered from a combinatorial library selection.[17]

An intriguing strategy to target MYC is based on the use of proteins that mimic the inhibitory activity of MAX and occupy E-Box DNA.[18–24] However, while protein based modalities generally hold immense potential in targeting 'undruggable' interactions, their usage is usually limited to extracellular targets, because they are unable to cross cell membranes.[25–27] Indeed, also the protein-based MYC inhibitors, of which Omomyc is the most studied representant, have mainly been explored as research tools and utilized via ectopic expression in cells or xenografts.[18,20] However, recently, the Soucek group discovered that purified Omomyc has some intrinsic cell penetrating activity, propelling it from tool compound to a viable therapeutic modality.[28] An Omomyc variant (Omo-103) is currently being investigated in human clinical trials for the treatment of various forms of tumors.[29]

The intrinsic cell penetration of Omomyc makes it a promising therapeutic modality, but at the same time, cell penetration remains its biggest limitation. Wang *et al.,* e.g.,



found that Omomyc exhibits negligible cellular uptake unless conjugated to a cell-penetrating modality.[30] These observations correlate well with the fact that, while Omomyc has a low nanomolar binding affinity for E-Box DNA (~25 nM, see **Figure S3.2**), its reported cellular activities are usually 100-1000-fold weaker, when delivered exogenously.[24,28,31,32]

In this study we aim to close this gap and show the development of a synthetic miniprotein that not only has a high affinity for E-Box DNA but, as shown

◀**Figure 3.1 Design of miniaturized protein dimers for MYC inhibition**. a) Schematic representation of MYC's mode of action: MYC upon heterodimerization with MAX binds to E-Box DNA. The MYC-TAD recruits the transcriptional machinery and activates gene transcription. MYC-driven gene programs lead to proliferation, cell growth and survival, all common traits of cancer. MAX can also heterodimerize, occupy the same E-Box DNA and act as an endogenous MYC inhibitor. b) Omomyc is a mutated variant of MYC that can also inhibit MYC by occupying E-Box DNA. c) Recent examples of Omomyc variants with covalently linked dimerization domains. d) Design of a miniaturized E-Box binding dimer, consisting of only the two basic helices of MYC. e) Design of a miniaturized E-Box binding dimer, including the loop helix domain for correct and stable orientation of the basic helices into the major grooves of the E-Box sequence. All structures were generated with Pymol based on PDB 5I50.

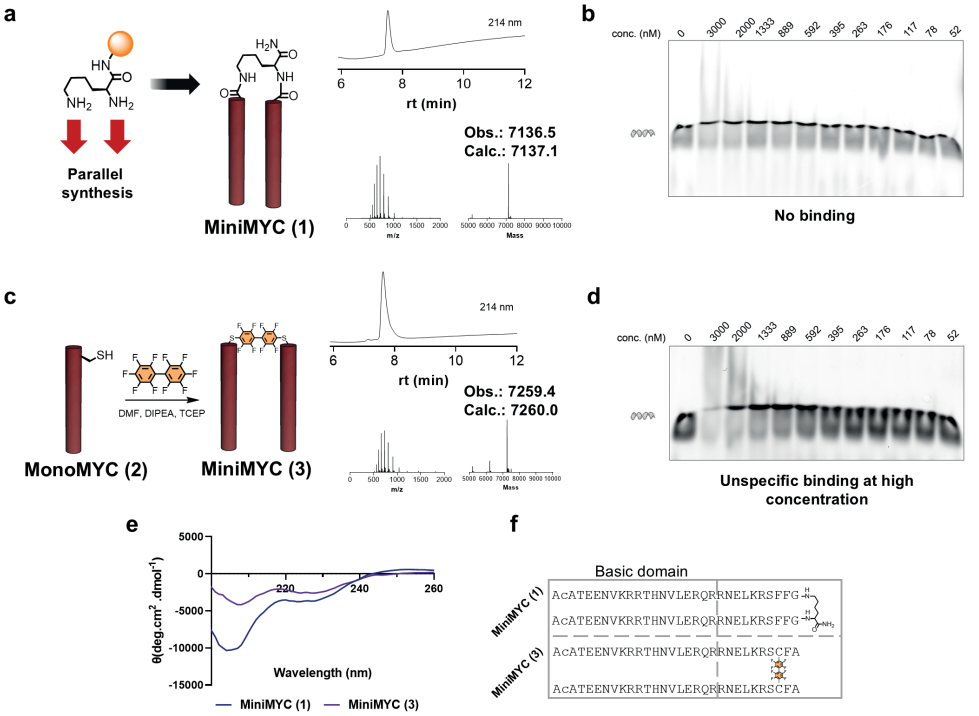
by functional cellular assays, can effectively enter cells and inhibit MYC driven transcription with submicromolar potency.

3.2 Results and discussion

To engineer such a bioactive miniprotein, we analyzed the structural features of E-Box binding proteins. The MYC/MAX, MAX/MAX, and Omomyc dimers belong to the bHLH-LZ protein family, where the LZ facilitates dimerization, and the basic helix mediates DNA binding. Dimeric configurations are requisite for DNA interaction, whereas monomers lack such affinity.[31,33] Situated between the basic helix and the LZ, the loop helix domain stabilizes tertiary and quaternary structures, aligning the basic helices appropriately with the major grooves of their target E-Box DNA (**Figure 3.1**). Notably, the basic helix domain drives not only DNA binding but also Omomyc's inherent cell penetration, as evidenced by the loss of this activity upon arginine-to-alanine substitutions within this region.[28]

We reasoned that by miniaturizing Omomyc's structure, we could create a compact dimeric scaffold with improved cell penetration and biological activity. Given the dual role of the basic helix in cell penetration and DNA binding, we envisioned that the development of synthetically dimerized variants of that portion of the protein could lead to optimized MYC inhibitors. The Moellering group recently showed that miniaturized proteomimetics derived from the MAX bHLH domain can tightly bind to E-Box (3.5 nM) and show improved cell penetration compared to full length MAX. [34] However, these compounds required > 1000-fold higher concentrations (20 μM) in order to enable significant effects in cellular assays (e.g. reporter gene). Missing a stable dimerization domain, these compounds might not be able to properly dimerize under physiological conditions inside cells. Similarly, compounds dimerized solely via disulfide linkage are likely to be reduced and to disassemble inside cells.[32,35]

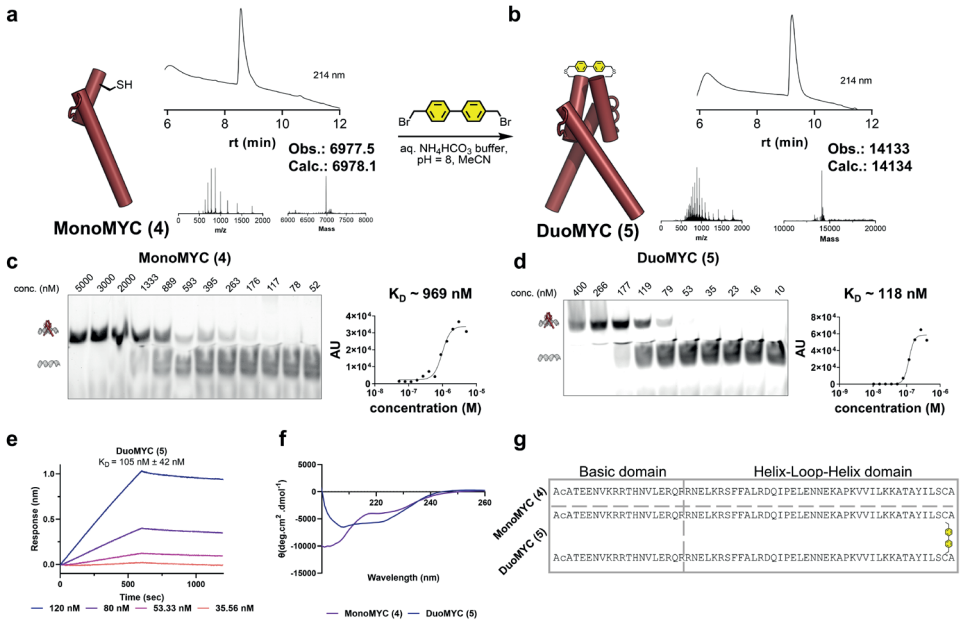
Thus, we here hypothesize that in order to obtain compounds with potent cellular



▲**Figure 3.2 The basic helix domain of Omomyc is not sufficient for E-Box DNA binding.** a) Parallel synthesis of MiniMYC **1** via Fmoc-based SPPS with LCMS analysis of the final product. b) EMSA gel of MiniMYC **1** shows no binding to E-Box DNA. c) Dimerization of MiniMYC **3** with LCMS analysis. d) EMSA gel of MiniMYC **3** shows no binding to E-Box DNA. At high concentration unspecific DNA binding was observed. e) CD spectra of MiniMYC **1** and **3**. f) Sequences of MiniMYC **1** and **3**. General EMSA protocol: labelled dsDNA construct (IRD700-ACCCCACCACGTGGTGCCT, only 5'-strand labelled, final concentration 4 nM) was preincubated with protein in EMSA buffer (20 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL of BSA, 1 mM DTT and 0.05% NP-40) for 30 minutes at rt followed by 15 minutes on ice and subsequently run on a 10% acrylamide TBE gel under native conditions. General CD protocol: Protein 10/20 μM in 20 mM K_2HPO_4 solution (pH = 7.4) was measured at 37 °C between 190 nm and 260 nm.

activity it is necessary 1) to identify the minimal structure required for specific DNA binding and 2) to covalently link these fragments for a compact and structurally stable scaffold.

We analyzed the DNA bound crystal structure of Omomyc[19] and designed a first miniaturized variant encompassing only the two symmetrical basic helix domains (27 residues). We prepared the dimeric miniMYC **1** via parallel automated solid-phase peptide synthesis (SPPS) using lysine (substituted both on the α- and on the ε-amine) as the covalent linkage between the two chains (**Figure 3.2a**). We first coupled the trifunctional building block Fmoc-Lys(Fmoc)-OH to the solid



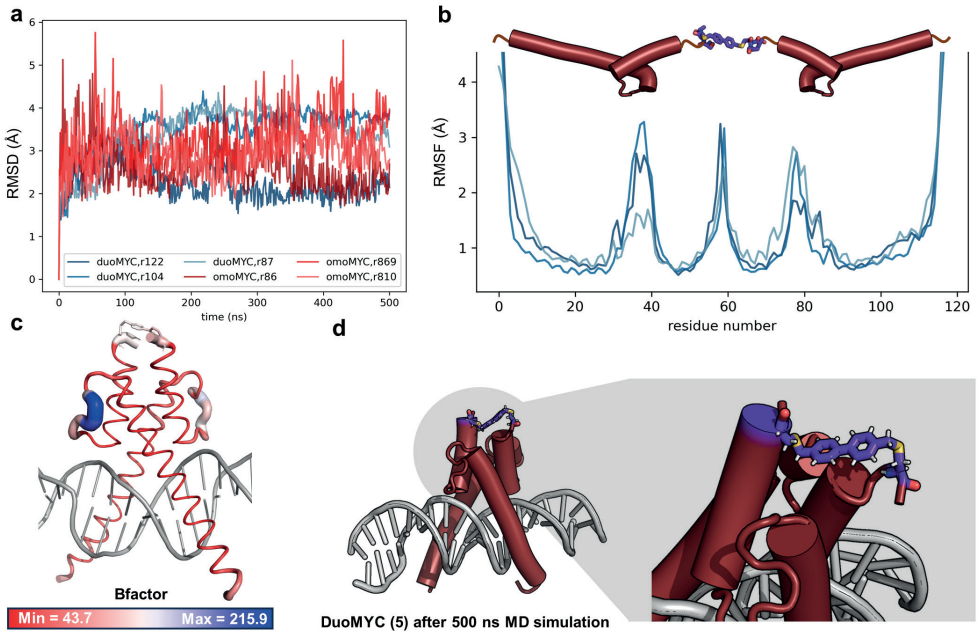
▲**Figure 3.3 DuoMYC 5 binds to E-Box DNA with nanomolar affinity**. a) LCMS analysis of MonoMYC **4** after SPPS and purification. b) Dimerization of **4** with bis(bromomethyl)biphenyl leads to pure DuoMYC **5** after purification. c) EMSA shows MonoMYC **4** binding to E-Box with a K_D ~ 969 nM. d) EMSA shows DuoMYC **5** binding to E-Box with a K_D ~ 118 nM. e) Binding evaluation by biolayer interferometry shows a K_D of 105 nM for DuoMYC **5**. f) DuoMYC **5** displays the CD spectrum of a partial α-helix while MonoMYC **4** exhibits the one of a random coil with partial helical structure when measured in absence of DNA. g) Sequences of MonoMYC **4** and DuoMYC 5. General SPPS coupling conditions: peptidyl resin incubated with Fmoc-AA-OH (10 eq.), HATU (9 eq.) and DIPEA (29 eq.) in DMF for 8 minutes at 70 °C. Fmoc was removed with 20% piperidine + 2% formic acid in DMF (4 minutes at 70 °C). General EMSA protocol: labelled dsDNA construct (IRD700-ACCCCACCACGTGGTGCCT, final concentration 4 nM) was preincubated with protein in EMSA buffer (20 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL of BSA, 1 mM DTT and 0.05% NP-40) for 30 minutes at rt followed by 15 minutes on ice and subsequently run on a 10% acrylamide TBE gel under native conditions. General BLI protocol: Protein was dissolved in kinetic buffer (1x PBS, 0.1% BSA, 0.02% Tween-20) with a final volume of 200 µL and measured against a 5' biotinylated hairpin with the sequence: TT CCT CAC GTG GCA TTT GGG TGC CAC GTG AGG. General CD protocol: Protein 10/20 µM in 20 mM K_2HPO_4 solution (pH = 7.4) was measured at 37 °C between 190 nm and 260 nm.

support and, upon removal of both Fmoc protecting groups, we coupled two glycines and performed a parallel synthesis of the 2 x 27mer sequences (**Figure 3.2a**). The Gly-Lys(Gly) linker should effectively bridge the 12 Å distance between the two helices. The dimeric 7 kDa MiniMYC **1** was obtained in good purity, as

shown by LC-MS. However, in electro mobility shift assay (EMSA), we detected no binding between E-Box DNA and MiniMYC **1** (**Figure 3.2b**). We sought to increase the rigidity of the linker for a variant with a stabilized structure, hoping to improve the binding to E-Box DNA. We synthesized monomeric monomer **2** with a Cys close to the C-terminus and linked two monomers using a decafluoro biphenyl reagent. The dimerization reaction worked efficiently and we isolated pure MiniMYC **3** (**Figure 3.2c**). Unfortunately, also this variant did not show any specific association to E-Box DNA (**Figure 3.2d**). For both MiniMYC **1** and **3** circular dichroism (CD) spectra show a partial α-helical structure (**Figure 3.2e**).

We hypothesized that the helix loop domain plays a crucial role in orienting the basic helices into the E-Box major grooves in the correct angle (**Figure 3.1e**). We thus envisioned a medium-sized analog encompassing the helix-loop-helix-domain (HLHD) in addition to the DNA binding helices. Residue 59, just outside the HLHD, appeared as an optimal choice for dimerization with a covalent linker. A direct parallel synthesis of this medium-sized analog failed (data not shown). Therefore, we synthesized a monomeric variant of the 59mer with a single cysteine close to the C-terminus **4** (**Figure 3.3a**). Encouragingly, this medium-sized variant, even without a covalent dimerization linker, exhibited E-Box DNA binding with an affinity of ~969 nM measured by EMSA (**Figure 3.3c**). The loop-helix domain might thus drive protein dimerization, at least to some extent, in good agreement with previous reports.[36] Recognizing, however, that the full-length Omomyc dimerization is mainly driven by the leucine zipper domain which we removed, we again decided to introduce a covalent linker for enhanced structural reinforcement. We cross-linked the two monomers with a bis(bromomethyl)biphenyl linker and successfully generated the synthetic 14 kDa miniprotein dimer, DuoMYC **5** (**Figure 3.3b**). Notably, DuoMYC **5** displayed improved E-Box DNA binding, with a K_D of ~118 nM measured by EMSA (**Figure 3.3d**) and 105 nM measured by biolayer interferometry (**Figure 3.3e, Figures S3.3**), representing an 8-fold enhancement over the monomeric variant. The CD spectrum of DuoMYC **5** shows clear α-helical patterns with defined minima at 208 and 222 nm. We also prepared a second variant, DuoMYC **7**, using a decafluorobiphenyl reagent for dimerization, and obtained a K_D of ~174 nM (**see Figure S3.1** for synthesis details and EMSA, **Figure S3.4** for BLI data). While we do not have a structural proof of DuoMYC's four helix core structure, its strong binding affinity for the E-Box domain and the recently solved structure of a similar miniprotein based on MAX[36] point towards this functional fold.

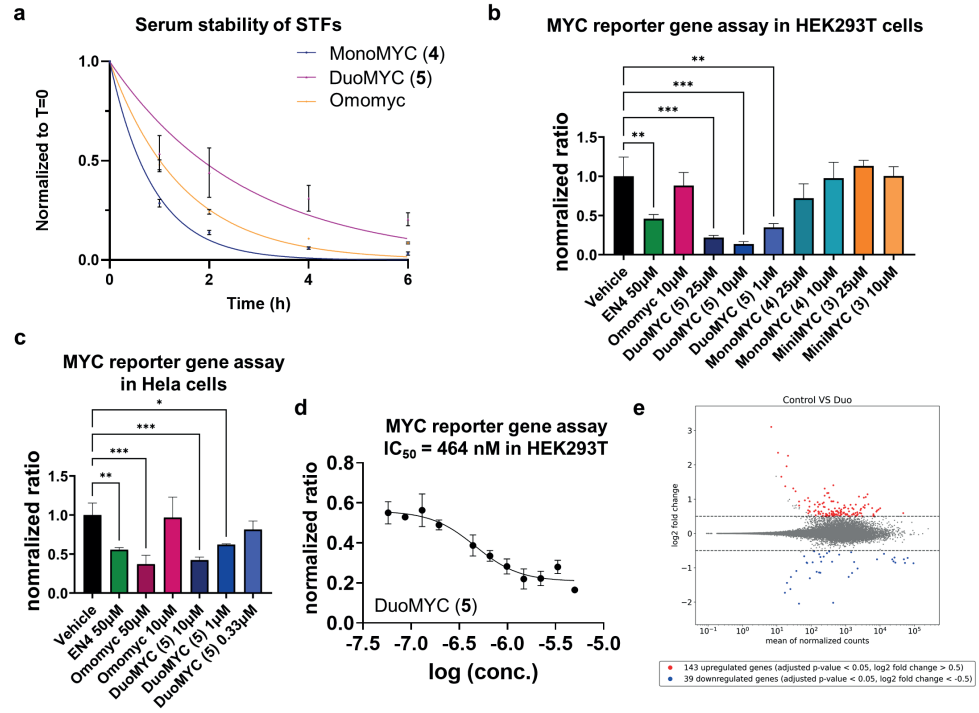
We next compared the stability of the full length Omomyc/Omomyc-DNA complex with the DuoMYC **5**-DNA complex using molecular dynamic (MD) simulations. The MD simulation was performed using Schrödinger's Desmond using the OPLS4 force field under the NPγT ensemble. Each system was run in triplicates, generating simulations of 500 ns. We found that both Omomyc and DuoMYC **5** remain in a stable DNA bound conformation in each of the three replicate systems, with overall residue fluctuations

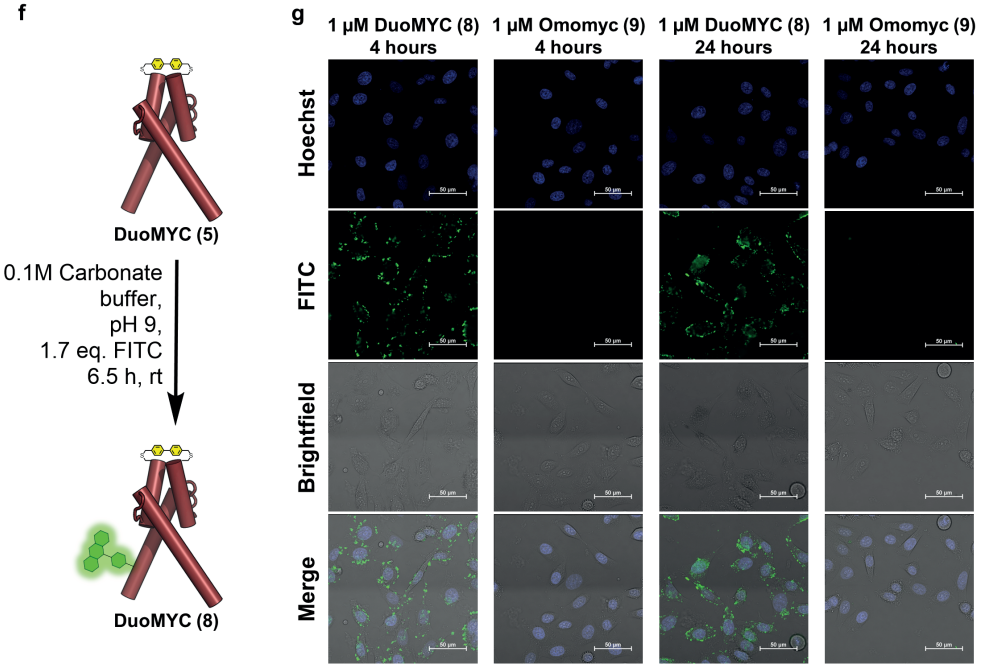


▲**Figure 3.4 MD simulation shows stability of DuoMYC 5 DNA complex.** a) Root-mean-square-deviation (RMSD) plot of DuoMYC **5** and Omomyc, in blue and red, respectively. Each triplicate simulation was run for 500 ns. b) RMSF plot for DuoMYC 5. c) RMSF plot translated to b-factor and shown as putty representation. d) DuoMYC **5** structure after 500 ns MD simulation with linker region highlighted.

between 2.2 and 3.5 Å compared with their starting coordinates (**Figure 3.4a**).The root-mean-square-fluctuation (RMSF) plot evidences the structural rigidity of the two helical domains in DuoMYC **5**, while the loop connecting the basic helix to the C-terminal helix displays significantly stronger fluctuations (**Figure 3.4b**). A very similar trend could be observed in Omomyc (**Figure S3.5**). The flexibility of the C-terminal extremities of the DuoMYC **5** dimer enables the correct arrangement of the covalent biphenyl linker, connecting the two helices (**Figure 3.4c-d**). Overall, the MD simulation highlights how the compact covalent linker (MW < 200) can efficiently replace the extensive leucine zipper domain (MW > 7000), still resulting in a stable DNA binding scaffold.

Since proteolytic instability is one of the major limitations of peptide-based therapeutics we next assessed the half-life of MonoMYC **4** and DuoMYC **5** in human serum, in comparison to Omomyc. We incubated the compounds in human serum and measured their stability by LC-MS at different timepoints normalizing to an internal standard. MonoMYC **4** was rapidly degraded with an overall half-life of 0.6 hours. DuoMYC **5** in turn, presented an improved stability with 50% compound still intact after 1.8 hours (**Figure 3.5a**), approximately two times longer than Omomyc


a
Serum stability of STFs
MonoMYC (4)
DuoMYC (5)
Omomyc
Normalized to T=0
Time (h)
b
MYC reporter gene assay in HEK293T cells
nomralized ratio
Vehicle
EN4 50μM
Omomyc 10μM
DuoMYC (5) 25μM
DuoMYC (5) 10μM
DuoMYC (5) 1μM
MonoMYC (4) 25μM
MonoMYC (4) 10μM
MiniMYC (3) 25μM
MiniMYC (3) 10μM
c
MYC reporter gene assay in Hela cells
nomralized ratio
Vehicle
EN4 50μM
Omomyc 50μM
Omomyc 10μM
DuoMYC (5) 10μM
DuoMYC (5) 1μM
DuoMYC (5) 0.33μM
d
MYC reporter gene assay
IC50 = 464 nM in HEK293T
normalized ratio
DuoMYC (5)
log (conc.)
e
Control VS Duo
log2 fold change
mean of normalized counts
143 upregulated genes (adjusted p-value < 0.05, log2 fold change > 0.5)
39 downregulated genes (adjusted p-value < 0.05, log2 fold change < -0.5)


f
DuoMYC (5)
0.1M Carbonate buffer, pH 9, 1.7 eq. FITC 6.5 h, rt
DuoMYC (8)
g
1 μM DuoMYC (8) 4 hours
1 μM Omomyc (9) 4 hours
1 μM DuoMYC (8) 24 hours
1 μM Omomyc (9) 24 hours
Hoechst
FITC
Brightfield
Merge

◀**Figure 3.5 DuoMYC 5 inhibits MYC driven transcription with submicromolar potency and displays good serum stability and cell penetration.** a) MonoMYC **4** displays a half-life of 0.6 hours, DuoMYC **5** a half-life of 1.8 hours and Omomyc a half-life of 1 hour. General serum stability protocol: proteins were dissolved in 10% human serum in DPBS to a final concentration of 60 µM, mixed and after the indicated time aliquots were taken and inactivated with 20% TFA in H_2O. The samples were analyzed via high-resolution LC-MS and curves plotted in Graphpad Prism 9. The standard deviations plotted are from n = 3 independent samples. b,c) MYC reporter gene assay with single concentrations of several compounds. HEK293T (b) or Hela (c) cells were transfected with a dual reporter gene, containing a MYC dependent firefly luciferase gene and a MYC independent, constitutively expressing renilla luciferase construct. The cells were treated with compounds at the indicated concentrations for 24 h. Subsequently, the cells were lysed, luciferase luminescence measured and the data were normalized to the control. Statistical significance details: one-way ANOVA was performed and revealed that there was a statistically significant difference in MYC-responsive luminescence between at least two groups (b) $F(9, 14) = 12.65$, $p < 0.0001$; (c) $F(6, 14) = 11.33$, $p < 0.0001$. Dunnett's multiple comparisons test was performed to test for a significant difference of the various groups from vehicle. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. d) Concentration curve of DuoMYC **5** reporter gene activity in transfected HEK293T cells. Curve shown is one of three independent experiments (see **Figure S3.6** for all experiments). e) RNA sequencing of HCT116 cells reveals that 182 genes were differently expressed compared to the non-treated cells. (f) DuoMYC **5** was reacted with FITC in carbonate buffer to yield unspecifically fluorescein-labelled DuoMYC **8**. (g) Confocal microscopy images show cell penetration of DuoMYC **8**. Hela cells were treated with 1 µM DuoMYC **8** or 1 µM FITC-labelled Omomyc for 24 h or 4 h and images of live cells acquired after washing. Hoechst dye was used for nuclear staining. Images shown were stitched from four acquired images.

(**Figure 3.5a, Figure S3.7**). The higher stability could be related to the fact that the crosslinked, dimeric scaffold is less accessible to proteases.

To assess whether the tight binding activity and the miniaturized, stable structure of DuoMYC **5** would translate to improved biological *in cell* activity we performed a MYC reporter gene assay. The reporter is based on a firefly luciferase gene activatable by MYC/MAX-dimers. A second constitutively-expressing Renilla luciferase vector serves as an internal positive control for normalization. We transfected HEK293T as well as HeLa cells with the reporter gene DNA and treated them with our compounds for 24 h (**Figure 3.5b-c**). Besides DuoMYC **5**, we included the recently discovered covalent small molecule MYC inhibitor EN4 as a positive control for MYC inhibition as it had shown activity in the same assay format.[15] We additionally included recombinant Omomyc (see SI 4.2-4.6 for expression protocol and **Figure S3.2** for EMSA validation) and in HEK293T cells also MniMYC **3** and MonoMYC **4**. In HEK293T cells only EN4 and DuoMYC **5** treatments resulted in statistically significant reduction of MYC-dependent transcription (**Figure 3.5b**). While 50 µM EN4 reduced MYC transcription by ~ 50%, DuoMYC **5** showed stronger inhibition even at 50-fold lower concentration (1 µM). Omomyc did not show any effect on MYC transcription in HEK293T cells. In HeLa cells Omomyc caused significant

reduction of MYC transcription only at its highest concentration (50 μM, **Figure 3.5c**). We next measured a full concentration curve of DuoMYC **5** and determined an IC_{50} of 464 nM in HEK293T (**Figure 3.5d, Figure S3.6**). The MYC-independent control reporter showed a stable signal over the whole concentration range of DuoMYC **5**, indicating that the compound is likely specifically inhibiting MYC driven transcription, rather than non-specifically shutting down cellular gene transcription.

To further assess DuoMYC **5**'s activity under more native conditions, we performed RNA-sequencing on HCT116 cells treated with either DuoMYC **5** (5 μM) or Omomyc (10 μM) for 24 hours, followed by whole RNA isolation. RNA-sequencing of DuoMYC-treated samples, compared to untreated controls, revealed 182 differentially expressed genes (**Figure 3.5e**). Gene set enrichment analysis (GSEA) using the Molecular Signatures Database (MSigDB)[37] identified several significantly modulated pathways, many of which are related to MYC activity and cancer biology. Notably, pathways such as apoptosis, hypoxia, and the G2-M checkpoint were among those significantly impacted, underscoring DuoMYC's role in modulating critical cellular processes that are typically influenced by MYC. The full list of modulated gene sets is provided in Supplementary Information (**Figures 3.S9-17, Table S3.1-4**). Additionally, a significant overlap in differentially expressed genes was observed between DuoMYC and Omomyc treatments (**Figure S3.17**), indicating their common mechanisms of action. We next wanted to confirm and visualize DuoMYC **5**'s uptake into cells. We prepared FITC labeled variants of DuoMYC **5** and Omomyc (**Figure 3.5f**). We then treated Hela cells with either 1 μM or 5 μM FITC-DuoMYC **8** or FITC-Omomyc **9** for 4 has well as 24 h, washed cells and acquired confocal microscopy images of live cells. Already at 1 μM DuoMYC **8** displays clear cellular uptake after 4 h as well as after 24 h treatment. Omomyc **9** at this concentration hardly shows any cell penetration at both treatment times (**Figure 3.5g**, second set of images in **Figure S3.8a**). The same effect with seemingly higher cellular uptake can be seen when cells are treated with 5 μM protein (**Figure S3.8b, Figure S3.8c-d** for untreated cells).

3.3 Conclusion

Taken together, here we have shown the development of a synthetic miniprotein that can bind E-Box DNA with high affinity and inhibit MYC-driven oncogenic transcription. Although MYC is a long sought after drug target, classical small molecule drugs cannot efficiently target MYC because of its lack of binding pockets. At the same time, protein drugs have limited activity due to their poor cell permeability. Our engineered miniprotein DuoMYC **5** combines a strong binding affinity and potent activity in cell assays. Furthermore DuoMYC **5** exhibits promising serum stability. We rationally engineered DuoMYC **5** based on the crystal structure of a larger E-Box targeting modality, Omomyc.[19] Surprisingly, our initial efforts to miniaturize the structure toward a dimeric miniprotein containing only the DNA

binding helices of Omomyc was unsuccessful while similar approaches had shown success in miniaturized GCN4 mimetics.[38,39] The mandatory components for our lead compound include the DNA binding helices, the loop-helix domains and a stable covalent linkage. This scaffold ultimately resulted in an efficient MYC inhibitor. Given the interest in MYC as a cancer drug target,[40] DuoMYC **5** can be considered a highly promising modality. Furthermore, our results highlight the possibility of designing innovative chemical modalities, such as synthetic miniproteins, able to address challenging intracellular targets.

3.4 References

[1] T. I. Lee, R. A. Young, "Transcriptional regulation and its misregulation in disease" *Cell* **2013**, *152*, 1237–1251.

[2] A. Boija, I. A. Klein, B. R. Sabari, A. Dall'Agnese, E. L. Coffey, A. V Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, R. A. Young, "Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains" *Cell* **2018**, *175*, 1842-1855.e16.

[3] J. H. Bushweller, "Targeting transcription factors in cancer — from undruggable to reality" *Nat Rev Cancer* **2019**, *19*, 611–624.

[4] N. Meyer, L. Z. Penn, "Reflecting on 25 years with MYC" *Nat Rev Cancer* **2008**, *8*, 976–990.

[5] S. K. Madden, A. D. de Araujo, M. Gerhardt, D. P. Fairlie, J. M. Mason, "Taking the Myc out of cancer: toward therapeutic strategies to directly inhibit c-Myc" *Mol Cancer* **2021**, *20*, 1–18.

[6] S. K. Das, B. A. Lewis, D. Levens, "MYC: a complex problem" *Trends Cell Biol* **2023**, *33*, 235–246.

[7] B. Ellenbroek, J. P. Kahler, S. R. Evers, S. J. Pomplun, "Synthetic Peptides: Promising Modalities for the Targeting of Disease-Related Nucleic Acids" *Angewandte Chemie (International Edition in English)* **2024**, DOI 10.1002/anie.202401704.

[8] E. M. Blackwood, R. N. Eisenman, "Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc" *Science (1979)* **1991**, *251*, 1211–1217.

[9] M. Kalkat, D. Resetca, C. Lourenco, P. K. Chan, Y. Wei, Y. J. Shiah, N. Vitkin, Y. Tong, M. Sunnerhagen, S. J. Done, P. C. Boutros, B. Raught, L. Z. Penn, "MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis" *Mol Cell* **2018**, *72*, 836-848.e7.

[10] E. M. Blackwood, R. N. Eisenman, "Max : A Helix-Loop-Helix Zipper Protein That Complex with Myc" *Science (1979)* **1991**, *251*, 1211–1217.

[11] R. V. Nithun, Y. M. Yao, X. Lin, S. Habiballah, A. Afek, M. Jbara, "Deciphering the Role of the Ser-Phosphorylation Pattern on the DNA-Binding Activity of Max Transcription Factor Using Chemical Protein Synthesis" *Angewandte Chemie - International Edition* **2023**, *62*, DOI 10.1002/anie.202310913.

[12] A. R. Ferré-D'Amaré, G. C. Prendergast, E. B. Ziff, S. K. Burley, "Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain" *Nature* **1993**, *363*, 38–45.

[13] S. K. Nair, S. K. Burley, "X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors" *Cell* **2003**, *112*, 193–205.

[14] N. B. Struntz, A. Chen, A. Deutzmann, R. M. Wilson, E. Stefan, H. L. Evans, M. A. Ramirez, T. Liang, F. Caballero, M. H. E. Wildschut, D. V Neel, D. B. Freeman, M. S. Pop, M. McConkey, S. Muller, B. H. Curtin, H. Tseng, K. R. Frombach, V. L. Butty, S. S. Levine, C. Feau, S. Elmiligy, J. A. Hong, T. A. Lewis, A. Vetere, P. A. Clemons, S. E. Malstrom, B. L. Ebert, C. Y. Lin, D. W. Felsher, A. N. Koehler, "Stabilization of the Max Homodimer with a Small Molecule Attenuates Myc-Driven Transcription" *Cell Chem Biol* **2019**, *26*, 711-723.e14.

[15] L. Boike, A. G. Cioffi, F. C. Majewski, J. Co, N. J. Henning, M. D. Jones, G. Liu, J. M. McKenna, J. A. Tallarico, M. Schirle, D. K. Nomura, "Discovery of a Functional Covalent Ligand Targeting an Intrinsically Disordered Cysteine within MYC" *Cell Chem Biol* **2020**, 1–10.

[16] H. Han, A. D. Jain, M. I. Truica, J. Izquierdo-Ferrer, J. F. Anker, B. Lysy, V. Sagar, Y. Luan, Z. R. Chalmers, K. Unno, H. Mok, R. Vatapalli, Y. A. Yoo, Y. Rodriguez, I. Kandela, J. B. Parker, D. Chakravarti, R. K. Mishra, G. E. Schiltz, S. A. Abdulkadir, "Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy" *Cancer Cell* **2019**, *36*, 483-497.e15.

[17] Z. Li, Y. Huang, T. I. Hung, J. Sun, D. Aispuro, B. Chen, N. Guevara, F. Ji, X. Cong, L. Zhu, S. Wang, Z. Guo, C. Chang, M. Xue, "MYC-Targeting Inhibitors Generated from a Stereodiversified Bicyclic Peptide Library" **2023**, DOI 10.1021/jacs.3c09615.

[18] L. Soucek, M. Helmer-Citterich, A. Sacco,

R. Jucker, G. Cesareni, S. Nasi, "Design and properties of a Myc derivative that efficiently homodimerizes" *Oncogene* **1998**, *17*, 2463–2472.

[19] L. A. Jung, A. Gebhardt, W. Koelmel, C. P. Ade, S. Walz, J. Kuper, B. Von Eyss, S. Letschert, C. Redel, L. D'Artista, A. Biankin, L. Zender, M. Sauer, E. Wolf, G. Evan, C. Kisker, M. Eilers, "OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors" *Oncogene* **2017**, *36*, 1911–1924.

[20] L. C. Lustig, D. Dingar, W. B. Tu, C. Lourenco, M. Kalkat, I. Inamoto, R. Ponzielli, W. C. W. Chan, J. A. Shin, L. Z. Penn, "Inhibiting MYC binding to the E-box DNA motif by ME47 decreases tumour xenograft growth" *Oncogene* **2017**, *36*, 6830–6837.

[21] M. Montagne, N. Beaudoin, D. Fortin, C. L. Lavoie, R. Klinck, P. Lavigne, "The max b-HLH-LZ can transduce into cells and inhibit c-Myc transcriptional activities" *PLoS One* **2012**, *7*, 2–10.

[22] X. Lin, O. Harel, M. Jbara, "Chemical Engineering of Artificial Transcription Factors by Orthogonal Palladium(II)-Mediated S-Arylation Reactions" *Angewandte Chemie - International Edition* **2023**, *202317511*, DOI 10.1002/anie.202317511.

[23] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription" *ACS Cent Sci* **2021**, *7*, 1408–1418.

[24] M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, "Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents" *J Am Chem Soc* **2021**, *143*, 11788–11798.

[25] O. Harel, M. Jbara, "Chemical Synthesis of Bioactive Proteins" *Angewandte Chemie - International Edition* **2023**, *62*, DOI 10.1002/anie.202217716.

[26] Leader B, Baca Q J, Golan D E, "Protein therapeutics: a summary and pharmacological classification" *Nat Rev Drug Discov* **2008**, *7*, 21–39.

[27] B. Khatri, I. Pramanick, S. K. Malladi, R. S. Rajmani, S. Kumar, P. Ghosh, N. Sengupta, R. Rahisuddin, N. Kumar, S. Kumaran, R. P. Ringe, R. Varadarajan, S. Dutta, J. Chatterjee, "A dimeric proteomimetic prevents SARS-CoV-2 infection by dimerizing the spike protein" *Nat Chem Biol* **2022**, *18*, 1046–1055.

[28] M. E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. S. Del Pozo, C. Fiore, L. Foradada, V. C. Cano, M. Sánchez-Hervás, M. Guenther, E. R. Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. M. Alonso, J. R. Whitfield, P. Lavigne, L. Soucek, "Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy" *Sci Transl Med* **2019**, *11*, 1–14.

[29] E. Garralda, M. E. Beaulieu, V. Moreno, S. Casacuberta-Serra, S. Martínez-Martín, L. Foradada, G. Alonso, D. Massó-Vallés, S. López-Estévez, T. Jauset, E. Corral de la Fuente, B. Doger, T. Hernández, R. Perez-Lopez, O. Arqués, V. Castillo Cano, J. Morales, J. R. Whitfield, M. Niewel, L. Soucek, E. Calvo, "MYC targeting by OMO-103 in solid tumors: a phase 1 trial" *Nat Med* **2024**, DOI 10.1038/s41591-024-02805-1.

[30] E. Wang, A. Sorolla, P. T. Cunningham, H. M. Bogdawa, S. Beck, E. Golden, R. E. Dewhurst, L. Florez, M. N. Cruickshank, K. Hoffmann, R. M. Hopkins, J. Kim, A. J. Woo, P. M. Watt, P. Blancafort, "Tumor penetrating peptides inhibiting MYC as a potent targeted therapeutic strategy for triple-negative breast cancers" *Oncogene* **2019**, *38*, 140–150.

[31] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription" *ACS Cent Sci* **2021**, *7*, 1408–1418.

[32] Z. Z. Brown, C. Mapelli, I. Farasat, A. V Shoultz, S. A. Johnson, F. Orvieto, A. Santoprete, E. Bianchi, A. B. McCracken, K. Chen, X. Zhu, M. J. Demma, B. M. Lacey, K. A. Canada, R. M. Garbaccio, J. O'Neil, A. Walji, "Multiple Synthetic Routes to the Mini-Protein Omomyc and Coiled-Coil Domain Truncations" *Journal of Organic Chemistry* **2020**, *85*, 1466–1475.

[33] J. G. Kwok, Z. Yuan, P. S. Arora, "An Encodable Scaffold for Sequence-Specific Recognition of Duplex RNA" *Angewandte Chemie International Edition* **2023**, e202308650.

[34] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, "Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains" *Nat Biotechnol* **2023**, *41*, DOI 10.1038/s41587-022-01504-x.

[35] N. M. McLoughlin, M. A. Albers, E. Collado Camps, J. Paulus, Y. A. Ran, S. Neubacher, S. Hennig, R. Brock, T. N. Grossmann, "Environment-Responsive Peptide Dimers Bind and Stabilize Double-Stranded RNA" *Angewandte Chemie International Edition* **2023**, DOI 10.1002/ANIE.202308028.

[36] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, "Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains" *Nat Biotechnol* **2023**, *41*, DOI 10.1038/s41587-022-01504-x.

[37] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov, P. Tamayo, "The Molecular Signatures Database Hallmark Gene Set Collection" *Cell Syst* **2015**, *1*, 417–425.

[38] B. Cuenoud, A. Schepartz, "Altered Specificity of DNA-Binding Proteins with Transition Metal Dimerization Domains" *Science (1979)* **1993**, *259*.

[39] J. Mosquera, M. I. Sánchez, M. E. Vázquez, J. L. Mascareñas, "Ruthenium bipyridyl complexes as photocleavable dimerizers: Deactivation of dna-binding peptides using visible light" *Chemical Communications* **2014**, *50*, 10975–10978.

[40] D. Massó-Vallés, L. Soucek, "Blocking Myc to Treat Cancer: Reflecting on Two Decades of Omomyc" *Cells* **2020**, *9*, 883.

3.5 Experimental

3.5.1 Synthesis procedures

3.5.1.1 Automated solid phase peptide synthesis (SPPS) – general protocol

All peptides were prepared on a Syro I XP synthesizer. The synthesis was performed on Chemmatrix rink amide resin (typically 100 mg, loading capacity of 0.41 mmol/g, 41 µmol scale). The synthesizer was charged with Fmoc-protected amino acid (Fmoc-AA-OH) solutions (0.5 M in DMF), HATU (0.45 M in DMF), Fmoc removal cocktail (20%/2%/78% = piperidine/formic acid/DMF) and pure DIPEA. As the first step in the peptide synthesis, the peptidyl resin was incubated at 70 °C in DMF while shaking for 10 min, after which the following coupling cycle was repeated until completion of the synthesis:

Coupling cycle

Coupling: Fmoc-AA-OH (10 eq., 800 µL for the 41 µmol scale), HATU (9 eq., 800 µL for the 41 µmol scale) and DIPEA (28.7 eq., 100 uL for the 41 umol scale) were sequentially added to the peptidyl resin and incubated at 70 °C for 8 minutes, with vigorous interval vortexing, followed by vacuum based draining of the resin.

Washing: DMF (1.2 mL) was added to the peptidyl resin, incubated for 1 min at room temperature (RT) with vigorous vortexing followed by vacuum based draining of the resin. The step was repeated 3 times.

Fmoc removal: Fmoc removal cocktail (1.5 mL) was added to the peptidyl resin incubated at 70 °C for 4 minutes, with vigorous interval vortexing.

Washing: DMF (1.2 mL) was added to the peptidyl resin, incubated for 1 min at RT with vigorous vortexing followed by vacuum based draining of the resin. The step was repeated 3 times.

After completion, the peptidyl resin was washed with DMF (3x), DCM (3x), dried under vacuum and either stored at -20 ˚C or directly acetylated.

3.5.1.2 Acetylation

After completion of the SPPS, the peptides were acetylated. First, the dry peptidyl resin was swollen in DMF (10 min) and subsequently vacuum-based drained. To the DMF soaked peptidyl resin the acetylation mixture (10%/10%/80% = Ac_2O/DIPEA/DMF) was administered (3 mL for 41 µL scale). The resin was thorougly shaken for 15 min at RT. The resin was subsequently washed with DMF (3x), DCM (3x), dried

under vacuum and either stored at -20 ˚C or cleaved immediately.

3.5.1.3 Full cleavage

The peptide was cleaved for 2 h using ~10 mL of 'special reagent K' cleavage cocktail (82.5%/5%/5%/5%/2.5% = TFA/H_2O/Cresol/Thioanisole/1,2-ethanedithiol). TFA was evaporated under a gentile nitrogen stream. Subsequently, the peptide was triturated (2x) using ice cold diethylether, spun down for 5 min at 10000 RPM. The pellet was dissolved in H_2O (10 mL for 41 µmol scale) and lyophilized.

3.5.1.4 Preperative HPLC (Prep-HPLC) – General protocol

For prep-HPLC a BESTA-Technik system attached to a Reprosil Gold 120 C18 column 10 µm (250 x 25 mm) and an ECOM Flash 10 DAD 800 UV detector set on 214 nm was used. The mobile phases used were solvent A (94.9%/5%/0.1% = H_2O/ MeCN /TFA) and solvent B (5%/94.9%/0.1% = H_2O/ MeCN /TFA).

3

Method: 0% solvent B over 3 min, next 0% to 15% solvent B over 1 min, followed by a linear gradient 15% to 90% solvent B over 86 min and 100% solvent B over 3 min. The obtained fractions where analyzed using a Sciex X500b QTOF LC-MS (see LC-MS high resolution protocol, method B). Pure fractions were combined and stored as lyophilized powders at -20 °C.

3.5.1.5 LC-MS

LC-MS chromatograms and associated mass spectra were acquired using a Shimazu LCMS-2020 system (Method A) or, for high resolution mass spectrometry data, a Sciex X500b QTOF ESI-QToF mass spectrometer couple to a Shimadzu Nexera UHPLC LC40DX3 (Method B). Mobile phases used for LC-MS analysis are solvent A (0.1% formic acid in H_2O) and solvent B (0.1% formic acid in acetonitrile). Solvent C (94.9%/5%/0.1% = H_2O/ MeCN /TFA) was used to dissolve protein mixtures to a concentration of 1 mg/mL.

The following LCMS methods were used:

Method A

Column: Kinetex® 2.6µm XB-C18 100 Å LC Column (50 x 3 mm) UV detector: 214 nM.

LC Method: 0% solvent B over 1 min, followed by a linear gradient 0% to 70% solvent B over 10 min, followed by 70% solvent B over 1.5 min, followed by 70% to 0% solvent B over 4.5 min, flowrate 0.55 mL/min.

Method B (high resolution)

Column: Phenomenex Synergi™ 4 µm Fusion-RP 80 Å LC Column (50 x 2 mm).

LC Method: 0% solvent B over 1 min, followed by a linear gradient 0% to 60% solvent B over 3.5 min, followed by a linear gradient 60% to 95% solvent B over 0.1 min, followed by 95% solvent B over 0.4 min, followed by a linear gradient 95% to 0% solvent B over 0.5 min, followed by 0% solvent B over 1.5 min, flowrate 0.5 mL/min.

MS parameters: General parameters: Method duration: 5 min; Total scan time: 0.276 sec; Estimated cycles: 1086; Intact protein mode: False; Decrease detector voltage: False; Large protein (>70 kDa): False; Ion Source: Source name: TurbolonSpray; Curtain gas: 35 psi; Ion source gas 1: 60 psi; Ion source gas 2: 60 psi; Temperature: 500 ˚C; Experiment: Scan type: TOF MS; Polarity: Positive; Spray voltage: 5500 V; CAD gas: 7; Time bins to sum: 4; Channel 1-4: True; TOF start mass: 350 Da; TOF stop mass 1500 Da; Accumulation time: 0.25; Declustering potential: 80V; Declustering potential spread: 0 V; Collision energy: 10V; Collision energy spread: 0 V; Override Qjet RF value: False.

Method C (high resolution)

When performing the serum stability assay, the following changes were made on the protocol Method B:

LC Method: 0% solvent B over 0.5 min, followed by a linear gradient 0% to 90% solvent B over 5.5 min, followed by 90% solvent B over 2 min, followed by a linear gradient 90% to 0% solvent B over 0.5 min, followed by 0% solvent B over 2 min, flowrate 0.5 mL/min.

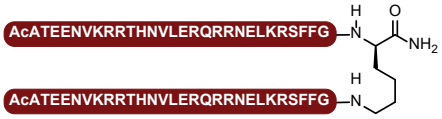
3.5.2 Protein synthesis with LCMS spectra

3.5.2.1 MiniMYC with lysine linker 1

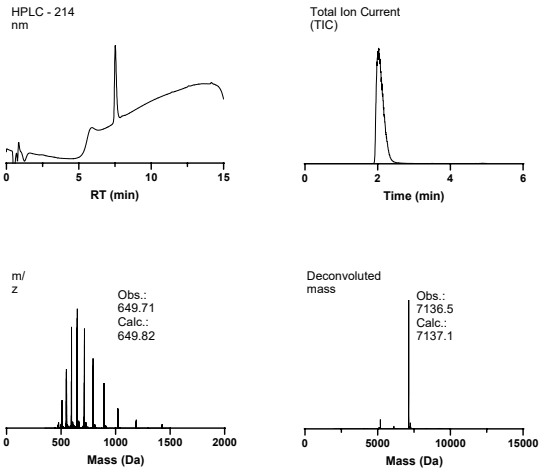
Fmoc-Lys(Fmoc)-OH (295 mg, 0.5 mmol, 1 eq.) was dissolved in a HATU solution (0.45M, 0.9 eq.) in DMF (1 mL). Subsequently, DIPEA (200 µL, 2.3 eq.) was added to the amino acid solution. The solution was incubated and shaken for 30 seconds and then added to the Chemmatrix Rink amide resin (50 mg, 41 mmol/g). After 10 min incubation at RT, with occasional shaking, the resin was drained and washed with DMF (3 x 5 mL). Next, the peptidyl resin was washed with 20% piperidine in DMF (5 mL) followed by 5 min incubation of 20% piperidine in DMF (5 mL). The resin was drained by vacuum and washed with DMF (5x 5 mL). The dimeric sequence was completed using the standard SPPS protocol. The final AA was acetylated and cleaved according to the general protocols.

MiniMYC (**1**) was purified using a Biotage® Selekt Flash Purification System coupled to a reverse phase Biotage® Sfär C18 D - Duo 100 Å 30 µm 25g column, further revered to as reversed phase flash chromatograpy (RP-FC). The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1% TFA). The UV detector was set to 214 nm. The following method was applied to the crude protein mixture:

RP-FC method: 10% solvent B over 1 column volumes (CV), followed by a linear gradient of 10% solvent B to 40% solvent B over 25 CV, followed by a linear gradient of 40% solvent B to 70% solvent B over 2 CV. The fractions were analyzed via LC-MS (Method A). Pure fractions were combined and lyophilized. Yield: 14.7 mg, 2.1 µmol, 16%.



Sequence : $(AcATEENVKRRTHNVLERQRRNELKRSFFG)_2K$



▲LCMS analytics of MiniMYC **1**. The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

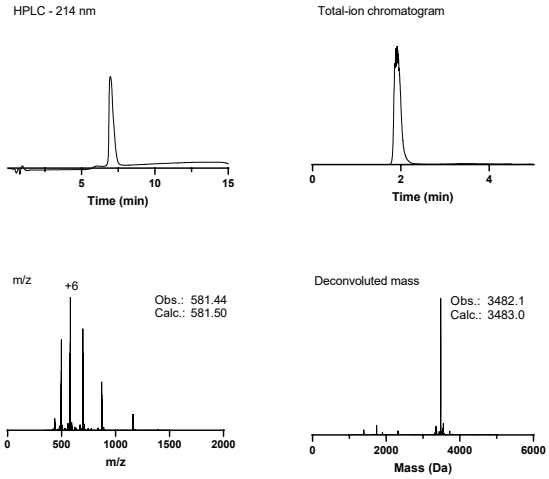
3.5.2.2 MiniMYC monomer 2

MiniMYC monomer **2** was synthesized, acetylated and cleaved according to the general SPPS, acetylation and cleavage protocol, respectively.

MiniMYC monomer **2** was purified using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm 25g column. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). The UV detector was set to 214 nm. The following method was applied to the crude protein mixture:

RP-FC method: 10% solvent B over 1 CV, followed by a linear gradient of 10% solvent B to 50% solvent B over 25 CV, followed by a linear gradient of 50% solvent B to 70% solvent B over 2 CV. The obtained fractions were analyzed according to the LC-MS – low resolution general protocol. Pure fractions were combined and stored as lyophilized powders. Yield: 23.4 mg, 6.7 2.1 µmol , 16%.

Sequence : AcATEENVKRRTHNVLERQRRNELKRSCFA



▲LCMS analytics of MiniMYC monomer **2**. The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

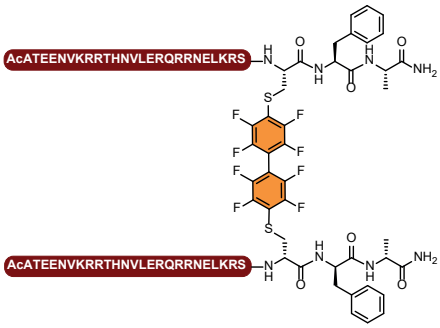
3.5.2.3 MiniMYC with octafluorobiphenyl linker 3

To prepare peptide **3**, stock solutions were prepared in DMF for decafluorbiphenyl (10 mM), DIPEA (80 mM) and TCEP (20 mM). Peptide **2** (11 mg, 3.16 µmol) was dissolved in 314 µL DMF. From each stock solution 314 µL were added to the peptide solution, resulting in the following final concentrations: peptide (3.16 µmol, 2.5 mM, 1 eq.), decafluorbiphenyl (3.16 µmol, 2.5 mM, 1 eq.), DIPEA (25.3 µmol, 20 mM, 8

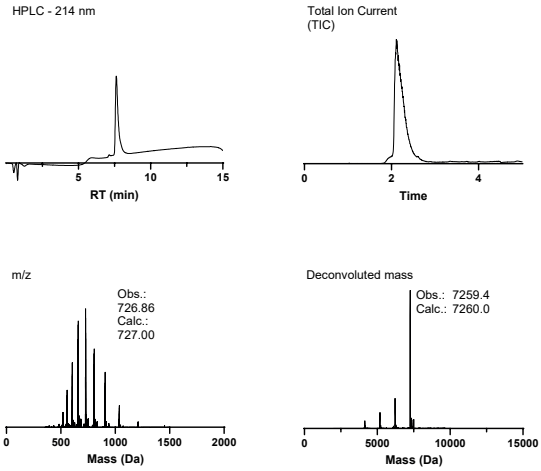
eq.) and TCEP (6.32 µmol, 5 mM, 2 eq.) in a final volume of 1.26 mL. The solution was shaken for 24 h after which it was quenched by the addition of 1% TFA in H_2O (10 mL) and lyophilized.

MiniMYC dimer (**3**) was purified using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm 10 g column. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). The UV detector was set to 214 nm. The following method was applied to the crude protein mixture:

RP-FC method: 10% solvent B over 1 CV, followed by a linear gradient of 10% solvent B to 50% solvent B over 25 CV, followed by a linear gradient of 50% solvent B to 70% solvent B over 2 CV. The obtained fractions were analyzed according to the LC-MS – low resolution general protocol. Pure fractions were combined and stored as lyophilized powders. Yield 3.5 mg, 482 nmol, 31%.



Sequence: $(AcATEENVKRRTHNVLERQRRNELKRSSCFA)_2$-Linker



▲LCMS analytics of MiniMYC **3**. The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3.5.2.4 MonoMYC 4

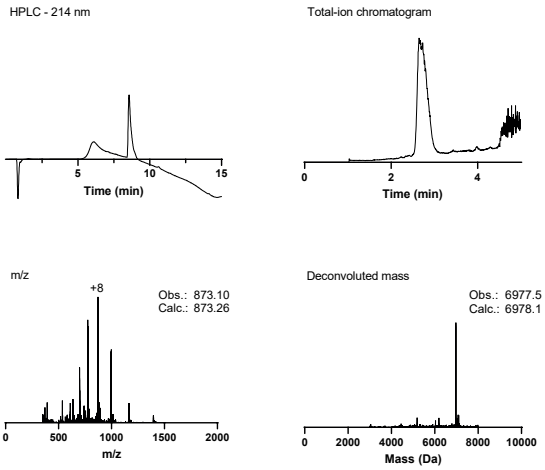
MonoMYC **4** was synthesized according to the general SPPS protocol on a 104 µmol scale. Instead of single couplings, double couplings were performed before removal of the Fmoc protection group at each cycle.

MonoMYC **4** was purified twice. The first purification step was performed using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 25g/50g. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). Solvent C (94.9%/5%/0.1% = H_2O/ MeCN /TFA) was used to dissolve crude protein mixtures. The UV detector was set to 214 nm. The following method was applied to the crude protein mixture:

RP-FC method: 10% solvent B over 2 CV, followed by 20% solvent B over 3 CV, followed by 25% solvent B over 2 CV, followed by 29% solvent B over 3 CV, followed by 32% solvent B over 4 CV, followed by 40% solvent B over 3 CV, followed by 98% solvent B over 5CV. The obtained semipure fractions where analyzed using a Sciex X500b QTOF LC/MS (see LC-MS high resolution protocol). Fractions containing product were combined and lyophilized.

The second purification was performed according to the general prep-HPLC protocol. Yield: 3.6 mg, 516 nmol, 0.5%.

Sequence: AcATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAP-KVVILKKATAYILSCA



▲LCMS analytics of MonoMYC **4**. The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3.5.2.5 DuoMYC with 4,4'-dimethyl-1,1'-biphenyl linker 5

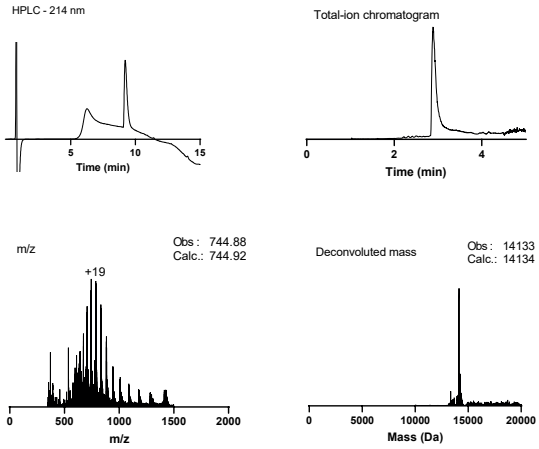
MonoMYC **4** (2.9 mg, 0.34 µmol, 1 eq.) was dissolved in H_2O (338 µL). A stock solution was prepared for 4,4′-Bis(bromomethyl)biphenyl in MeCN (0.4 mM). To the 1 mM MonoMYC **4** solution 2.2 mL NH_4HCO_3 buffer (100 mM, pH = 8) and 0.846 mL of the 4,4′-Bis(bromomethyl)biphenyl stock were added, resulting in a final volume of 3.4 mL and the following concentrations: MonoMYC **4** (0.34 µmol, 0.1 mM, 1 eq.) and 4,4′-Bis(bromomethyl)biphenyl (0.34 µmol, 0.1 mM, 1 eq.). The solution was stirred for 2h and subsequently lyophilized. The crude mixture was either purified according to the prep-HPLC protocol or with RP-FC. For RP-FC the following method was applied to the crude protein mixture:

For RP-FC a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 10g or 50g was used. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). Solvent C (89.9%/10%/0.1% = H_2O/ MeCN /TFA) was used to dissolve crude protein mixtures. The UV detector was set to 214 nm.

3

RP-FC method: 0% to 15% solvent B over 0.5 CV, followed by a linear gradient 15% to 56% solvent B over 12.7 CV and 95% solvent B over 2 CV. The obtained fractions where analyzed using a Sciex X500b QTOF LC/MS (see LC-MS high resolution protocol). Fractions containing product were combined and lyophilized. Yield: 1.11 mg, 78.5 nmol, 46%.

Sequence: (AcATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAP-KVVILKKATAYILSCA)$_2$-linker



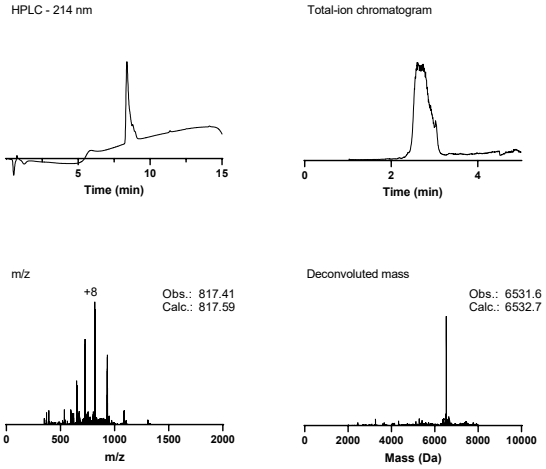
▲LCMS analytics of DuoMYC **5**. The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3.5.2.6 MonoMYC 6

MonoMYC **6** was synthesized according to the general SPPS, acetylation and cleavage protocols on a 41 µmol scale. MonoMYC **6** was purified using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 25g. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1% TFA). The UV detector was set to 214 nm. The following method was applied to the crude protein mixture:

RP-FC method: 20% solvent B over 2 CV, followed by a linear gradient of 20% solvent B to 30% for 3 CV, followed by a linear gradient of 30% solvent B to 40% over 22 CV, followed by 100% solvent B over 3 CV. The obtained pure fractions were analyzed using the low resulution LCMS protocol. Fractions containing product were combined and lyophilized. Yield: 17 mg, 2.6 2.1 µmol, 6%.

Sequence: AcVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILK-KATAYILSVA$_{\beta}$C



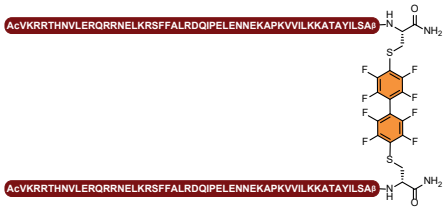
▲LCMS analytics of MonoMYC **6**. The 214nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3.5.2.7 DuoMYC with octafluorobiphenyl linker 7

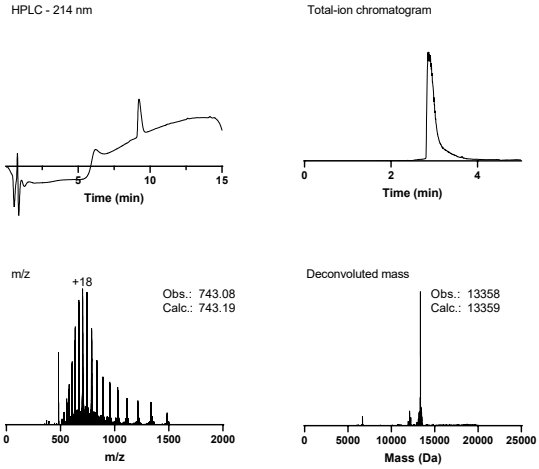
MonoMYC **6** (4.24 mg, 0.64 µmol, 1 eq.) was disolved in DMF (58 µL). The following stock solutions were prepared in DMF: DIPEA (0.5M), decalfluorobiphenyl (0.01M) and TCEP (0.1M). To the MonoMYC **6** solution was added from each stock solution: 13 µL of DIPEA (6.4 µmol, 10 eq., 60 mM), 19.5 µL TCEP (1.2 µmol, 0.3 eq., 17.7 mM) and 19.5 µL decafluorbiphenyl (0.19 µmol, 3 eq., 1.77 mM) resulting in a final peptide concentration of 6 mM. The reaction was shaken for 3 days at 37 °C. When precipitation occurred during the reaction, the reaction mixture was sonicated, vortexed and placed back in the shaker. Yield: 1.28 mg, 12%.

DuoMYC **7** was purified using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 10g. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). The UV detector was set to 214 nm. The following method was applied to the crude protein mixture:

HPFC method: 20% solvent B over 2 column volumes (CV), followed by a linear gradient of 20% solvent B to 30% for 3 CV, followed by a linear gradient of 30% solvent B to 40% over 22 CV, followed by 100% solvent B over 3 CV. The obtained semipure fractions were analyzed using the low resolution LCMS protocol. Fractions containing product were combined and lyophilized. Yield: 1.28 mg, 95.8 nmol, 12%.



Sequence: (AcVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILK KATAYILSA$_{\beta}$C)$_2$-linker



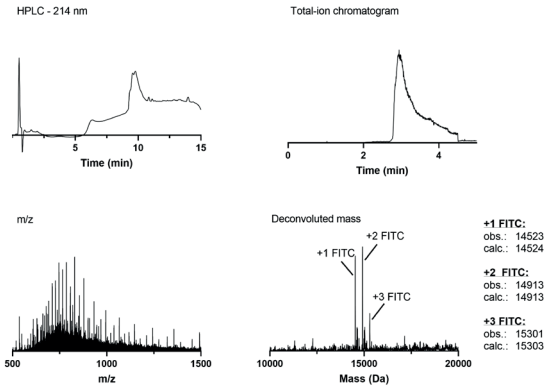
▲LCMS analytics of DuoMYC **7** The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3.5.2.8 Unspecifically FITC-labeled DuoMYC 8

DuoMYC **5** (1.55 mg, 0.089 µmol, 1 eq.) was dissolved in 0.1M carbonate buffer (1.55 mL, pH = 9.1), providing a 1 mg/mL solution. FITC in DMSO (51.5 µL of a 1 mg/mL solution, 0.12 µmol, 1.4 eq.) was added to the DuoMYC solution. The mixture was protected from light and reacted for 3.5h. An additional 0.3 eq. of FITC was added and reacted for an additional 3h. Labeled DuoMYC **8** was directly purified using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 10g. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). The UV detector was set off. The following method was applied to the crude protein mixture:

RP-FC method: 0% to 15% solvent B over 0.5 CV, followed by a linear gradient 15% to 56% solvent B over 12.7 CV and 95% solvent B over 2 CV. The obtained fractions where analyzed using a Sciex X500b QTOF LC/MS (see LC-MS high resolution protocol). Fractions containing product were combined and lyophilized yielding FITC-labelled DuoMYC (1.29 mg, 71.55 nmol, 80%) containing 1, 2 and 3 molecules

of fluorescein. For stock preparations an average molecular weight of 18027.37 g/mol was used.

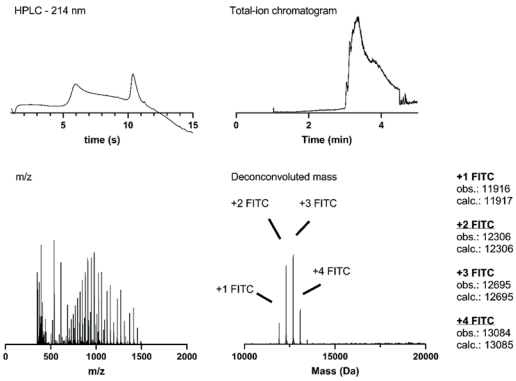


▲LCMS analytics of DuoMYC **8** The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3

3.5.2.9 Unspecifically FITC-labeled Omomyc 9

Omomyc expressed in E.Coli and purified (see 4.2, 4.4 mg, 287 nmol, 1 eq.) was dissolved in 440 µL 0.1M carbonate buffer pH 9.1 to a concentration of 1mg/mL. Subsequently, a 2.6mM solution of FITC in dmso (22 µL, 402 nmol, 1.4 eq.) was added and the solution was stirred at rt and protected from light for 4h. The crude product was then purified by reverse phase column using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 10g. The mobile phase solvents used were: solvent A (H_2O with 0.1% TFA) and solvent B (MeCN with 0.1%TFA). After column chromatography the product was further purified by spinning through an Advance spin column with a molecular weight cutoff 3 kD to remove residual FITC. This procedure was repeated three times followed by lyophilization yielding FITC-labelled Omomyc (0.87 mg, 58 nmol, 12%) containing 1, 2, 3, and 4 molecules of fluorescein. For stock preparations an average molecular weight of 12500.75 g/mol was used.



▲LCMS analytics of Omomyc **9** The 214 nm LC chromatogram was obtained using LC-MS method A. The TIC, MS and deconvoluted mass chromatograms were obtained using LC-MS method B.

3.5.3 Biochemistry

3.5.3.1 Overview of buffers used

Bacterial cell lysis buffer	20mM Tris-HCl, pH 8, 0.5 M NaCl, 3mM $MgCl_2$, 0.05% DNAse,complete EDTA free protease inhibitor (1 tablet freshly added in 50 mL buffer)
EK cleavage buffer	200 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 20 mM CaCl,
EK dilution buffer	20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 2 mM $CaCl_2$, 50% glycerol
Affinity wash buffer	20 mM Tris-HCl, pH 8, 0.5 M NaCl, 10mM imidazole
Affinity elution buffer	20 mM Tris-HCl, pH 8, 0.5 M NaCl, 500mM imidazole
EMSA binding buffer	20 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL of BSA, 1 mM DTT and 0.05% NP-40
EMSA running buffer	0.5x TBE, 50 mM Tris borate, pH 8.3, 1 mM EDTA

3.5.3.2 Expression of Omomyc in E.Coli

Plasmid pET-30 with Omomyc was gratefully received from the research group of Prof. Dr. Eilers at University of Wuerzburg (see annex for plasmid sequence) and sequenced before use.

Plasmids were transformed into competent ArcticXpress DE3 RIL cells (Agilent) by heat shock. 2 µL of 10% β-mercaptoethanol were mixed with 100 µL of competent cell suspension thawed on ice and incubated for 10 minutes on ice. Next, 25 ng of plasmid DNA were added and the cells incubated for another 30 minutes on ice. The cells were then heat-shocked in a water bath for 20 seconds at 42 °C and subsequently incubated on ice for 2 minutes followed by the addition of 0.9 mL SOC media and incubation at 37 °C, 250 rpm for 1 h. Cells were pelleted by centrifugation, 0.9 mL of the supernatant decanted and the pellet resuspended in the remaining 100 µL of media. Subsequently, cells were plated for selection on LB agar with kanamycin and gentamicin and incubated at 37 °C overnight.

Single colonies were picked from the plate and cultured overnight at 37 °C and 200 rpm in 100 mL of LB media containing kanamycin and gentamicin. Next, 3x2 L of LB media containing kanamycin and gentamicin in 5 L Erlenmeyer flasks were inoculated with 25 mL of preculture and grown at 37 °C, 200 rpm until an OD of 0.8 was reached. The temperature was then set to 14 °C and protein expression induced by addition of 100 µM of IPTG. Protein expression was conducted overnight for 18 h after which the cells were harvested by centrifugation (5,000g, 4 °C, 12 min.), the pellet resuspended in lysis buffer (20mM Tris-HCl, pH 8, 0.5 M NaCl, 3mM $MgCl_2$, 0.05% DNAse,complete EDTA free protease inhibitor) and the cells lysed by pressure lysis. Cell debris was removed by ultracentrifugation (35,000 rpm, 4°C, 45 min.).

3.6.3.3 Purification of his-tagged proteins

The supernatant after ultracentrifugation was purified using an ÄKTA start protein purification system equipped with a 5 mL HisTrap HP His tag protein purification column (Cytiva). After washing out unbound protein with wash buffer (20 mM Tris-HCl pH 8, 0.5 M NaCl, 10 mM imidazole) the protein was eluted using a gradient of 10 mM to 500 mM imidazole in the same buffer over 40-50 column volumes. Fractions with protein were analyzed for protein content and purity by SDS-PAGE.

3.5.3.4 General protocol for SDS-PAGE

Samples were mixed with the appropriate amount of 4x LB buffer containing β-mercaptoethanol and heated to 95 °C for 5 minutes to denature proteins. Samples were then loaded onto acrylamide gels of the desired percentage (made using acrylamide 37.5:1 acrylamide/Bis) and ran at room temperature at 150V for one hour. If desired, gels were stained for at least 20 minutes with coomassie blue staining solution followed by at least three rounds of destaining with destaining solution (50% MeOH, 40 % H_2O, 10% AcOH). Stained gels were scanned on a Bio-Rad ChemiDoc MP machine.

3.5.3.5 General protocol for buffer exchange

The combined fractions obtained from Ni-column purification were incubated with TCEP (500 µM-1mM) to break any possible formed disulfide bonds. The buffer was then exchanged by subjecting the protein to column chromatography on a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 25g or 50g column and a stepwise gradient of 0 % MeCN in water, 50 to 100% MeCN water). The combined fractions containing the protein were lyophilized, yielding the His-tagged or final protein as TFA salt.

3.5.3.6 His-tag cleavage by enterokinase

Omomyc was dissolved at 2 mg/mL in EK cleavage buffer (200 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 20 mM $CaCl_2$) and after addition of 10 u/mL enterokinase (previously diluted in dilution buffer, 20 mM Tris-HCl, pH 7.4, 200 mM NaCl, 2 mM $CaCl_2$, 50% glycerol), the protein was incubated overnight. Cleaved Omomyc was purified purified using an ÄKTA start protein purification system equipped with a 5 mL HisTrap HP His tag protein purification column with the protein being eluted in the flowthrough. Fractions were checked fro protein content and purity using SDS-PAGE and the buffer was exchanged using the general protol for buffer exchange.

3.5.3.7 Electromobility shift assay (EMSA)

For EMSAs, 14 µL (15 µL for no protein control) water were mixed with 1 µL of 20x FAM-labelled dsDNA construct (IRD700-ACC CCA C**CA CGT G**GT GCC T, one one strand labelled, (Integrated DNA technologies, Leuven, Belgium), 1 µL of 20x protein in water and 4 µL of 5x EMSA buffer (final buffer concentration: 20 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL of BSA, 1 mM DTT and 0.05% NP-40). The samples were incubated for 30 minutes at room temperature, placed on ice and incubated for another 15 minutes after which 15 µL of the samples were loaded onto a 10% acrylamide TBE gel which was pre run before for 1h at 75V at 4 °C in 0.5x TBE. Samples were run for 20 minutes at 120 V followed by 40 minutes at 100 V at 4 °C in 0.5x TBE and subsequently scanned on on a Bio-Rad ChemiDoc MP machine.

3.5.3.8 Cell culture

Cells were cultured at 37°C in 5% CO_2 atmosphere. Cell lines were cultured in ATCC recommended media and split twice a week before confluency was reached.

3.5.3.9 Reporter gene assay

3.5.3.9.1 Reporter gene assay in HEK293T cells

The reporter gene assay was performed using Cignal reporter assay (CCS-012L, Qiagen) according to the manufacturer's protocol with slightly prolonged incubation times. In brief, HEK293T cells were harvested and resuspended in OptiMEM media containing 5% FBS and 1% non essential amino acids. 40,000 cells were seeded per well in a 96-well plate and transfection cocktail of either cignal reporter or positive or negative control reporter along with attractene transfection reagent was added and the cells incubated overnight. Next, media was changed to assay media (OptiMEM, 0.5% FBS, 1% NEAA, Pen/Strep) and the cells were incubated for 8 hours after which the media was replaced by 75 µL assay media containing the different proteins/

control compounds at the required concentration and the cells were incubated for 16-24 hours. Luciferase assay was then performed using a luciferase assay kit (E2940, Promega). Cells were lysed by addition of 75 µL of DualGlo Luciferase assay reagent and incubated for 15 minutes after which Luciferase luminescence was measured in on a Bio-Rad ChemiDoc MP machine. Subsequently, 75 µL of DualGlo Stop & Glo reagent were added and the Renilla luciferase luminescence measured after 15 minutes of incubation time. Signal was normalized against cell number by calculating the ratio of firefly and renilla luminescence and eventually these signals were normalized against the untreated control. All experiments were done at least in technical duplicates.

3.5.3.9.2 Reporter gene assay in Hela cells

When performing the assay using Hela cells, the following changes were made to the protocol described in 4.9.1: 20,000 Hela cells were seeded per well and double the amount of DNA was used.

3.5.3.10 Serum stability assay

A 10% Human serum solution was prepared in DPBS (pH = 7.3). From the peptide stock (1 mM in H_2O) peptide was diluted to a final concentration of 60 µM in the 10% serum solution. The mixture was vortexed immediately after protein addition and 5 µL aliquotes were mixed with 5 µL of a 20% TFA solution in H_2O (T = 0) to quench the human serum, resulting in protein precipitating. Subsequently, the peptide serum solution was incubated at 37 ˚C using a BIO-RAD T100™ Thermal Cycler with the lid set on 95 °C. At indicated timepoints 5 µL aliquotes were mixed with 5 µL of 20% TFA in H_2O, respectively. The pellet was diluted 5.5x with addional DPBS to redissolve the precipitated protein pellet. This solution was analyzed according to the high resolution LCMS protocol (Method C). As internal standard (IS), the extracted-ion chromatogram (XIC) from 1233.5184 was used belonging to human serum albumin. After measurement, the XIC belonging to the protein, +8 peak for MonoMYC **4**, +19 peak for DuoMYC **5** and +13 peak for Omomyc, and IS were extracted from the total-ion chromatogram (TIC). Using Graphpad Prism 9, the area under the curve (AUC) from each timepoint was calculated, normalized against the AUC of IS followed by normalization against T = 0 and plotted. Next, a nonlinear regression – One phase decay analysis was performed to obtain $t_{1/2}$ with a plateau constant equal to 0 and Y_0 set to 1. The assay was performed in independent triplicates.

3.5.3.11 Circular Dichroism (CD)

All lyophilized proteins were dissolved in H_2O to a concentration of 1 mM. For CD measurements these proteins were diluted to 10 µM (MiniMYC **1**, MiniMYC **3**,

DuoMYC **5**) or 20 µM (MonoMYC **4**, MonoMYC **6**, DuoMYC **7**) in a 20 mM K_2HPO_4 solution (pH = 7.4) with a final volume of 200 µL. CD samples were measured at 37 °C using a Jasco J-815 CD spectrometer with a 1 mm path length quartz cuvette. The following parameters were used for a full wavelength scan: wavelength = 260-190 nm; Data pitch = 1 nm; scanning mode = continuous; scanning speed = 100 nm/min; response = 1, BW = 1, accumulation = 3. For analysis, all curves were smoothened using Graphpad Prism 9 by averaging 10 neighboring datapoints and a 4th order smoothing polynomial. All curves were normalized against their concentration (10 µM/20 µM), pathlength (0.1 mm) and peptide bonds (n) using the following formula:

$$\left[\emptyset\right] = \frac{100*\emptyset_{obs}}{c*n*l}$$

With θ in deg*cm²*dmol⁻¹, θ_{obs} in mdeg, concentration (c) in mM, peptide bonds (n) and path length of cuvette in cm. The graph was plotted using GraphPad Prism 10.1.0.

3.5.3.12 Biolayer interferometry (BLI)

DuoMYC **5** or **7** was diluted in H_2O to a concentration of 120 µM and diluted further in kinetic buffer (1x PBS, BSA 0.1%, Tween-20 0.02%) with a final volume of 200 µL. The following 5' biotinylated hairpin sequence was used: TT CCT **CAC GTG** GCA TTT GGG TGC **CAC GTG** AGG (Integrated DNA technologies, Leuven, Belgium). A 10 µM DNA stock in H_2O was incubated at 95 °C for 5 minutes, cooled down to RT and diluted to 1 µM in kinetic buffer, which could be loaded on the Sartorius Octet R4. Each experiment was performed in a 96-well plate. Before each measurement, the sensortips were incubated in 200 µL kinetic buffer for at least 10 minutes. All concentration curves were measured at 25 °C while shaking. During each experiment to following steps were performed: the sensortips were incubated for 1 minute in kinetic buffer, followed by a 1 minute DNA loading step, followed by 1 minute of incubation in kinetic buffer, followed by 10 minute sample incubation (association), followed by 10 minutes kinetic buffer incubation (dissociation). The same experiment was repeated without loading DNA on the sensortips to obtain the background signal that could be removed from the obtained BLI spectra. The experiment was performed in independent triplicates from which the average K_D was calculated.

3.5.3.13 Fluorescence microscopy

Hela cells were harvested and seeded at a density of 10,000 cells/well in a 96-well plate. In case of 24h compound treatment the cells were let to attach to the well for 6h and then treated with 75µL of the compounds at the desired concentration in MEM media for 24h. The media was then aspirated and the cells treated with 0.2µg/mL Hoechst nuclear stain for 30 minutes. In case of 4h compound treatment, the cells were let to attach to the wells overnight followed by treatment with 75 µL of the

compounds at the desired concentraton along with 0.2 µg/mL Hoechst nuclear stain in MEM for 4h. For both treatment times cells were then washed 2x with fresh MEM and the live cells immediately imaged on a Nikon Ti2 microscope equipped with a plan apo 40x/0.95 air objective using a 405nm laser for Hoechst and a 488nm laser for fluorescein excitation.

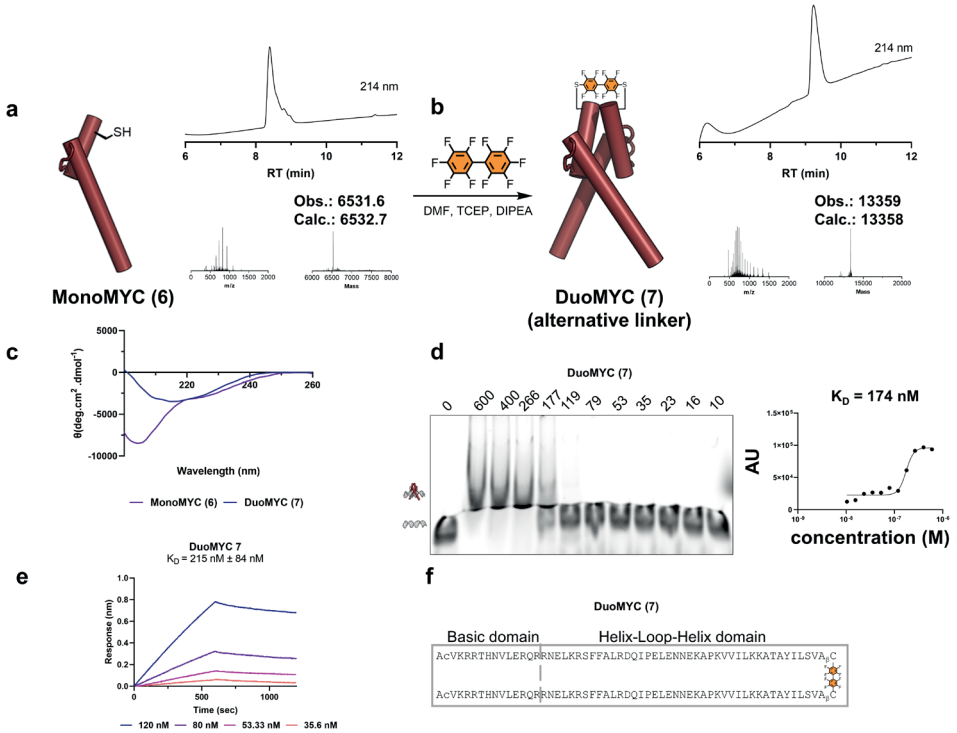
3.5.3.14 RNA sequencing

HCT116 cells were harvested and seeded at a density of 200,000 cells/well in 12-well plates. Cells were then grown overnight and subsequently treated for 24h with 5 µM DuoMyc, 10 µM Omomyc or vehicle in McCoy5a media supplemented with 10% FBS, 1% Pen/Strep and Glutamax. Total RNA was isolated using Qiagen RNeasy Plus Mini Kit (74134) and dissolved in RNAse free water. Samples were shipped to and sequencing was performed by Novogene GmbH (Planegg, Germany). Quality of the raw .fastq files was checked with FastQC v0.12.1. The BBTools software v3.62 (bbduk command) was used to trim adapter sequences and 3' quality decrease, remove PhiX contaminant reads and filter out low quality reads. Filtered reads were aligned to the human genome (Ensembl release 112, Homo_sapiens.GRCh38.112) with STAR v2.7.10b in -quantMode GeneCounts mode to quantify gene-level counts. Differential gene expression analysis between DuoMYC **5** or Omomyc treatment and untreated control was performed in python using the PyDESeq2 v0.4.10 package. Differentially expressed genes were selected based on shrunk Log2 fold change > 0.5 or < -0.5 and adjusted p-value < 0.05. Pre-ranked Gene Set Enrichment Analysis (GSEA) was performed with the gseapy v1.1.3 package with the Hallmark gene sets from the Molecular Signatures Database (MsigDB).

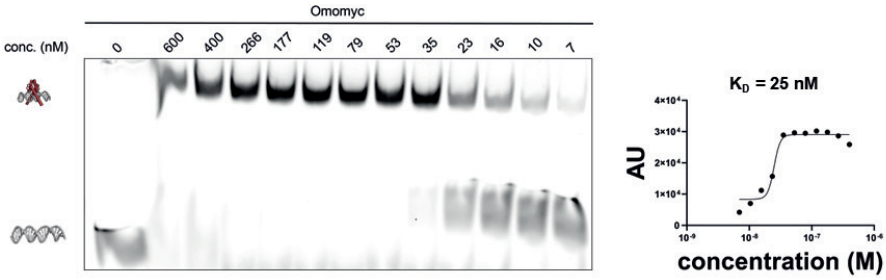
3.5.4 Molecular Dynamic simulations

The pdb structure of Omomyc bound to DNA was retrieved from the PDB (id=5I50) and subsequently prepared for MD using Maestro's 2022-3 protein preparation wizard. Thereafter, a system was build using the desmond system builder, MD simulations were performed using the OPLS4 forcefield under the NPgT ensemble at a temperature of 300K using 2fs timesteps. Simulations were performed in triplicate using a different random seed for the initial velocities, and the resulting 500ns trajectories were subsequently analyzed using the desmond simulation event analysis. All structural images were generated using PyMOL.

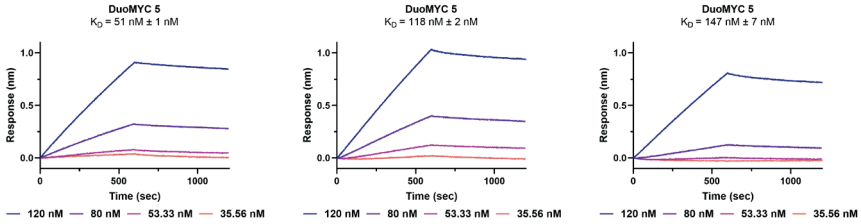
3.6 Supplementary Figures and Tables



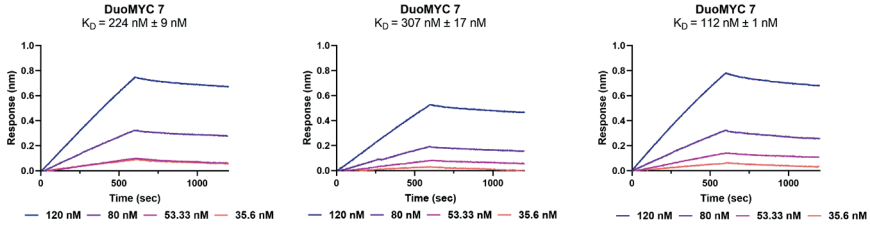
▲Figure S3.1. DuoMYC 7 with alternative (octafluorophenyl) linker binds E-Box with nanomolar affinity. (a) SPPS of monomeric variant **6** with LCMS analysis. b) Dimerization of **6** with decafluorobiphenyl, leading to DuoMYC **7**. c) DuoMYC **7** displays the CD spectrum of a partial α-helix while MonoMYC **6** exhibits the one of a random coil with partial helical structure when measured in absence of DNA. d) DuoMYC **7** binds to E-Box with a K_D = 174 nM. e) Binding evaluation by biolayer interferometry shows a KD of 215 nM for DuoMYC **7**. f) Sequence of DuoMYC **7**. General SPPS coupling conditions: peptidyl resin incubated with Fmoc-AA-OH (10 eq.), HATU (9 eq.) and DIPEA (29 eq.) in DMF for 8 minutes at 70 ^{0}C. Fmoc was removed with 20% piperidine + 2% formic acid in DMF (4 minutes at 70 ^{0}C). General CD protocol: Protein 10/20 µM in 20 mM K_2HPO_4 solution (pH = 7.4) was measured at 37 ^{0}C between 190 nm and 260 nm. General EMSA protocol: Fluorescently-labelled dsDNA construct (IRD700-ACCCCACCACGTGGTGCCT, final concentration 4 nM) was preincubated with protein in EMSA buffer (20 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM). General BLI protocol: Protein was dissolved in kinetic buffer (1x PBS, BSA 0.1%, Tween-20 0.02%) with a final volume of 200 µL and measured against a 5' biotinylated hairpin with the sequence: TT CCT **CAC GTG** GCA TTT GGG TGC **CAC GTG** AGG.



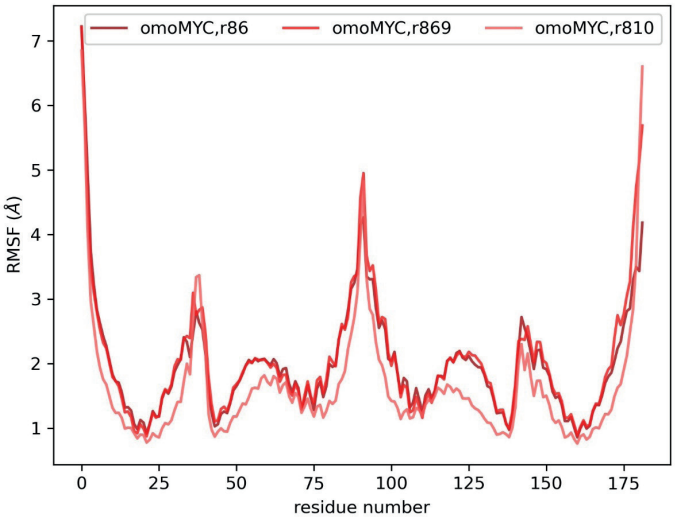
▲**Figure S3.2 Omomyc binds to E-box DNA with an affinity of 25 nM.** General EMSA protocol: Fluorescently-labelled dsDNA construct (IRD700-ACCCCACCACGTGGTGCCT, final concentration 4 nM) was preincubated with protein in EMSA buffer (20 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM)



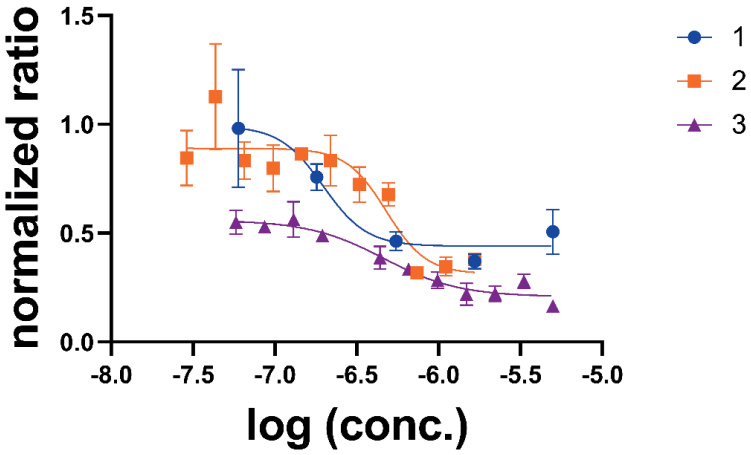
▲**Figure S3.3 Three independant experiments to determine the K_D of DuoMyc 5 in a BLI assay.** K_D (1) = 51 nM, K_D (2) = 118 nM, K_D (3) = 147 nM. General BLI protocol: Protein was dissolved in kinetic buffer (1x PBS, BSA 0.1%, Tween-20 0.02%) with a final volume of 200 µL and measured against a 5' biotinylated hairpin with the sequence: TT CCT **CAC GTG** GCA TTT GGG TGC **CAC GTG** AGG.



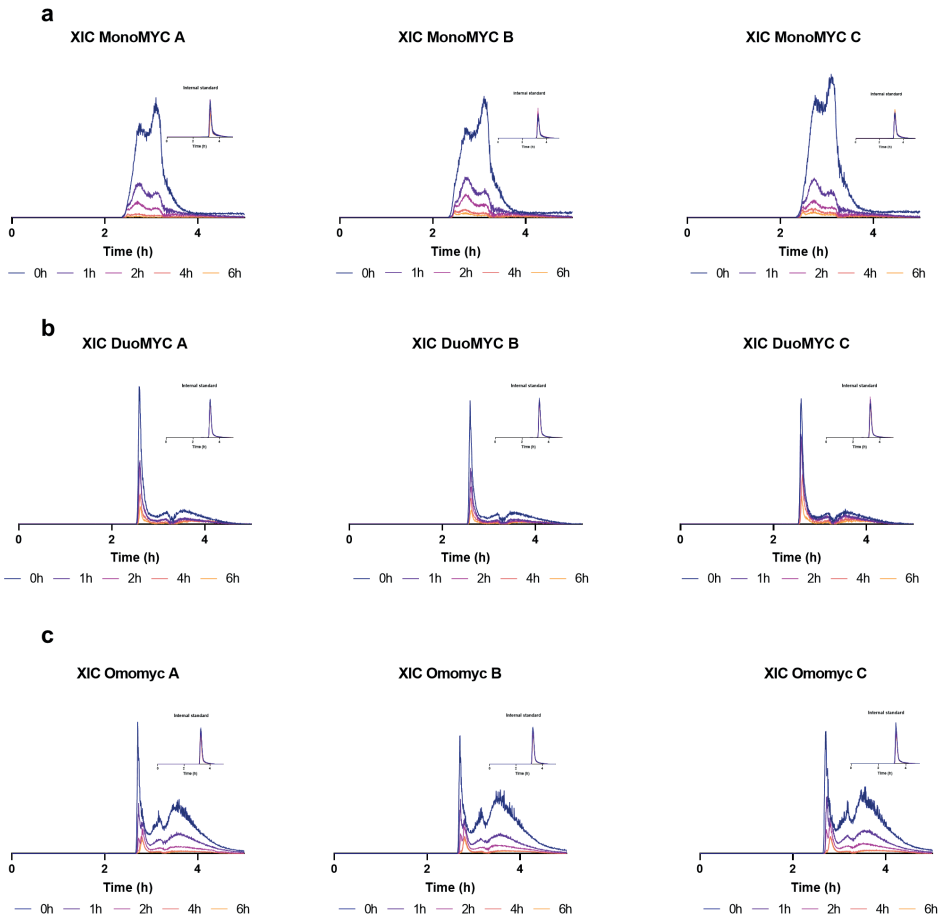
▲**Figure S3.4 Three independant experiments to determine the K_D of DuoMyc 7 in a BLI assay.** K_D (1) = 224 nM, K_D (2) = 307 nM, K_D (3) = 112 nM. General BLI protocol: Protein was dissolved in kinetic buffer (1x PBS, BSA 0.1%, Tween-20 0.02%) with a final volume of 200 µL and measured against a 5' biotinylated hairpin with the sequence: TT CCT **CAC GTG** GCA TTT GGG TGC **CAC GTG** AGG.



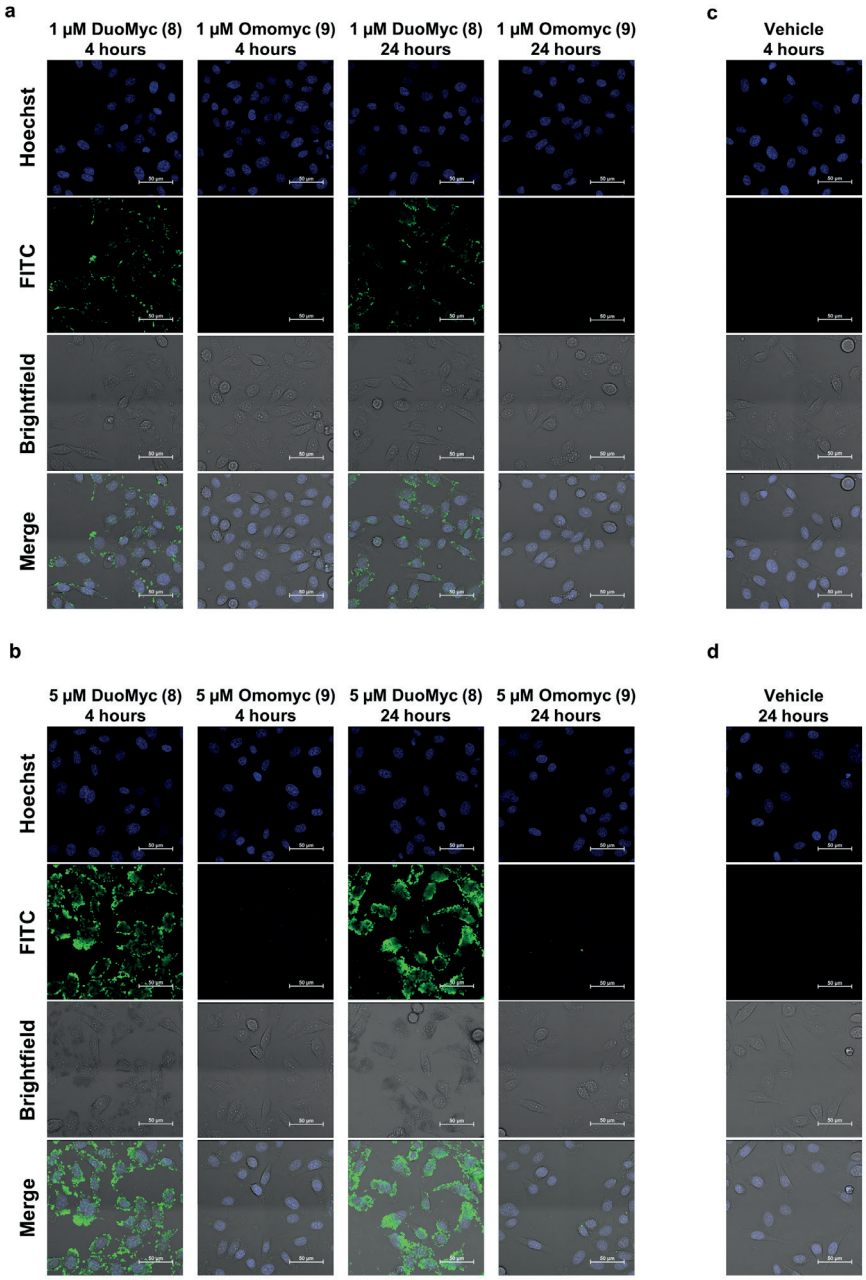
▲**Figure S3.5**. RSMF plot of Omomyc (500 ns MD simulation).



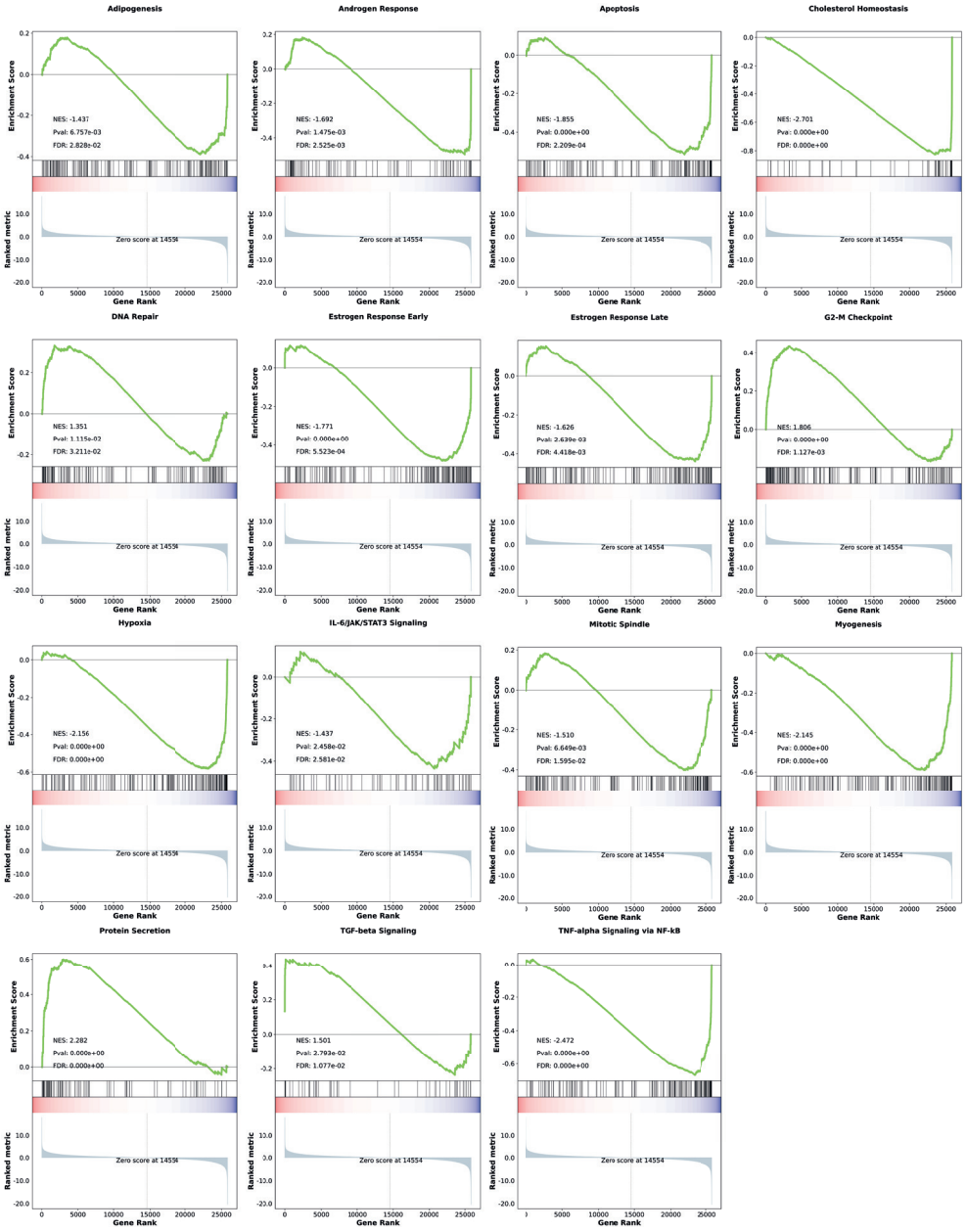
▲**Figure S3.6. Three independant experiments to determine the EC_{50} of DuoMyc 5 in a reporter gene assay.** Curves were fitted to logarithmized data and results indicate EC_{50}s in the nanomolar range. $EC_{50}(1)$ = 198 nM, $EC_{50}(2)$ = 480 nM, $EC_{50}(3)$ = 464 nM.



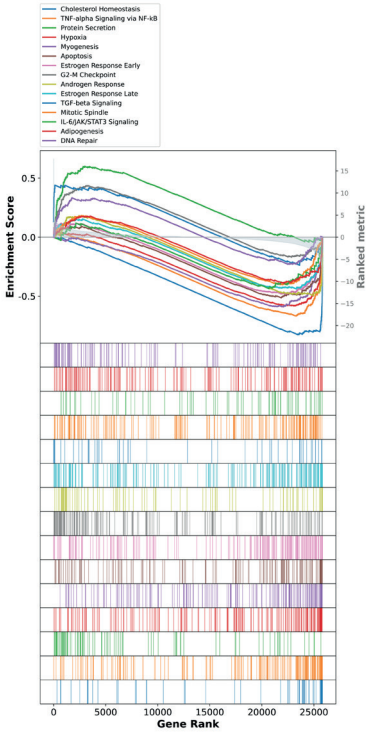
▲**Figure S3.7**. Extracted Ion currents with their corresponding IS from three independant experiments to determine the $t_{0.5}$ of a) MonoMYC **4** b) DuoMyc **5** and c) Omomyc in a serum stability assay.



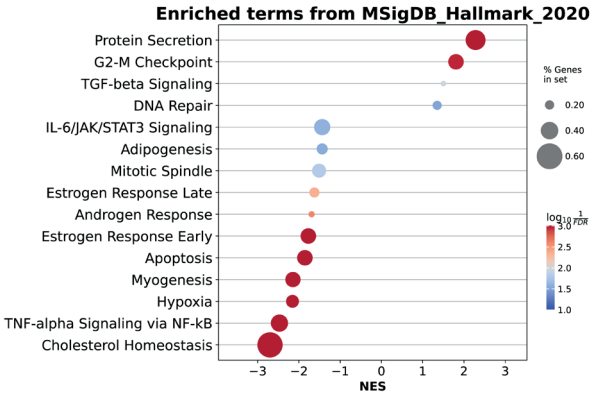
▲**Figure S3.8. Confocal fluorescence microscopy images of Hela cells treated with DuoMYC 8, fluorescein-conjugated Omomyc or vehicle at the indicated concentration for different times.** (a) A second set of images of Hela cells treated with proteins at 1 μM concentration. (b) Hela cells treated with proteins at 5 μM concentration. (c,d) Cells treated with vehicle for 4 h (c) and 24 h (d). Images shown were stitched from four acquired images.



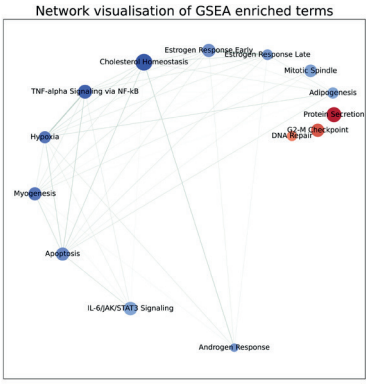
▲**Figure S3.9.** GSEA plots of differently expressed pathways by DuoMyc **5**.



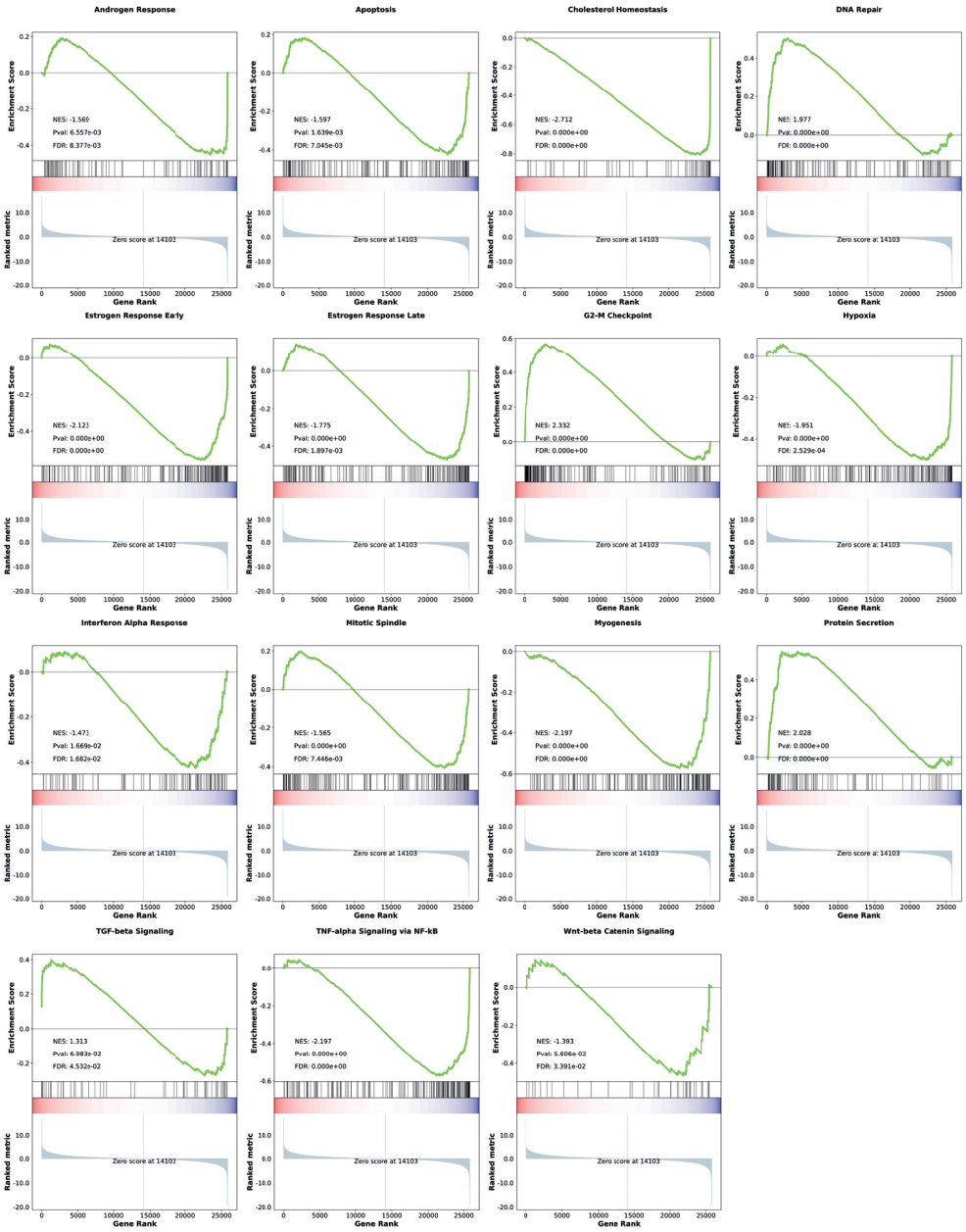
▲**Figure S3.10.** All GSEA plots of differently expressed pathways by DuoMYC **5** combined.



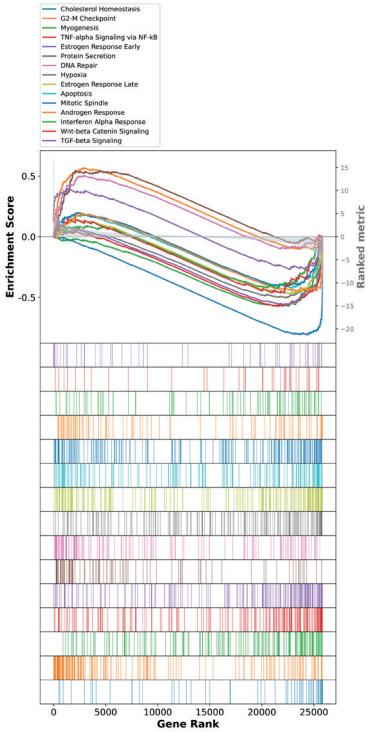
▲**Figure S3.11**. GSEA dotplot of differently expressed pathways by DuoMYC **5**



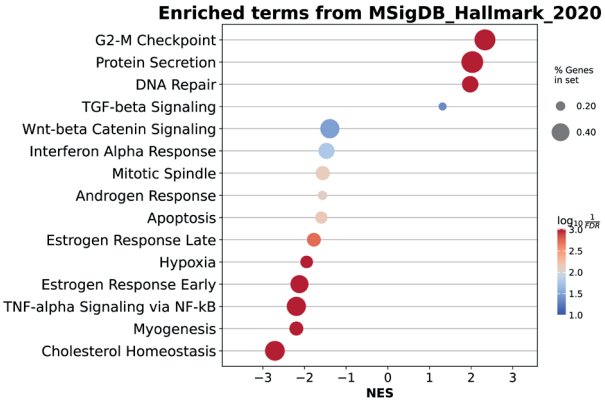
▲**Figure S3.12.** Network visualisation of GSEA enriched terms by DuoMYC **5**



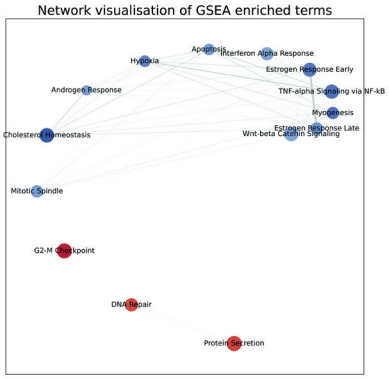
▲**Figure S3.13.** GSEA plots of differently expressed pathways by Omomyc.



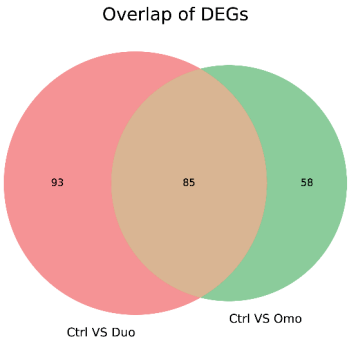
▲**Figure S3.14.** All GSEA plots of differently expressed pathways by Omomyc combined.



▲**Figure S3.15**. GSEA dotplot of differently expressed pathways by Omomyc.



▲**Figure S3.16.** Network visualisation of GSEA enriched terms by Omomyc.



▲**Figure S3.17.** Overlap of Differently expressed genes between DuoMYC **5** and Omomyc.

Table S3.1. List of differently expressed genes by DuoMYC **5**.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Symbol
ENSG00000228502	231.5547249	-0.677149607	0.175663757	4.531418911	5.86E-06	0.000525138	EEF1A1P11
ENSG00000135766	403.3851924	0.933252167	0.134210772	-7.355808651	1.90E-13	8.29E-11	EGLN1
ENSG00000117318	1236.353667	-1.266326031	0.072314822	17.68951726	5.05E-70	5.81E-66	ID3
ENSG00000160752	655.050198	0.731286092	0.109907653	-7.059502655	1.67E-12	6.34E-10	FDPS
ENSG00000117407	248.5514861	0.767308133	0.172536598	-5.051097079	4.39E-07	5.68E-05	ARTN
ENSG00000163041	2828.281156	0.719232345	0.069166433	-10.66255483	1.52E-26	1.95E-23	H3-3A
ENSG00000188290	894.3050678	0.741488778	0.106500938	-7.342785815	2.09E-13	9.02E-11	HES4
ENSG00000237330	134.6474068	0.674830464	0.240622974	-3.702726004	0.000213295	0.010110427	RNF223
ENSG00000175793	9147.604108	0.695786348	0.0510936	-13.81885137	1.96E-43	4.84E-40	SFN
ENSG00000158246	278.4923329	0.504439113	0.150683882	-4.076257055	4.58E-05	0.002968624	TENT5B
ENSG00000179862	30.22672621	1.311533431	0.523204432	-3.53060123	0.000414616	0.017238048	CITED4
ENSG00000132196	681.7037312	0.51331035	0.09891238	-5.673164207	1.40E-08	2.83E-06	HSD17B7
ENSG00000131584	477.1327101	0.548440978	0.1087839	-5.548004472	2.89E-08	5.17E-06	ACAP3
ENSG00000049249	21.26720974	2.262926605	0.536276125	-4.855182655	1.20E-06	0.000130109	TNFRSF9
ENSG00000225630	437.7967628	-2.020114691	0.143036122	14.33872583	1.25E-46	3.33E-43	MTND2P28
ENSG00000237973	407.362517	-0.824833415	0.145208205	6.169319504	6.86E-10	1.88E-07	MTCO1P12
ENSG00000229344	185.5931859	-1.145975921	0.210961522	5.946025057	2.75E-09	6.54E-07	MTCO2P12
ENSG00000248527	1781.157201	-0.833244334	0.098892019	8.757352547	2.00E-18	1.50E-15	MTATP6P1
ENSG00000198744	170.8838661	-1.137922662	0.207389113	6.005658593	1.91E-09	4.70E-07	MTCO3P12
ENSG00000187017	1985.286462	0.513915333	0.062136749	-8.57799432	9.65E-18	6.94E-15	ESPN
ENSG00000177606	1266.27815	0.641782312	0.071079043	-9.330214581	1.06E-20	1.01E-17	JUN
ENSG00000169174	3659.335143	0.822508825	0.054132618	-15.37799939	2.30E-53	7.93E-50	PCSK9
ENSG00000162496	654.5403177	-0.617407556	0.098303862	6.703869775	2.03E-11	6.67E-09	DHRS3
ENSG00000235847	19.85093328	-1.440792589	0.572571089	3.551374032	0.000383225	0.016146937	LDHAP7
ENSG00000204253	186.1866783	-0.686735834	0.184504029	4.427765805	9.52E-06	0.000802772	HNRNPCP2
ENSG00000144485	1706.874218	0.707564086	0.096135567	-7.76474083	8.18E-15	4.48E-12	HES6
ENSG00000172478	10.8900887	2.351048971	0.908130673	-3.554414594	0.000378822	0.016000457	MAB21L4
ENSG00000134324	565.5613349	0.65954789	0.112912365	-6.295144476	3.07E-10	8.98E-08	LPIN1
ENSG00000234072	89.19256337	0.779827477	0.272330702	-3.75784314	0.000171384	0.008512625	GTF3C2-AS2
ENSG00000115738	144.755946	-1.117909847	0.194253884	6.244228329	4.26E-10	1.21E-07	ID2
ENSG00000249992	179.0815529	0.846531789	0.187447167	-5.118207528	3.08E-07	4.16E-05	TMEM158
ENSG00000169908	44.50603735	0.930551644	0.416724387	-3.352889775	0.000799725	0.028102777	TM4SF1
ENSG00000134107	1234.20519	0.502626991	0.073272862	-7.2136474	5.45E-13	2.24E-10	BHLHE40
ENSG00000239445	62.85871615	-0.66600084	0.311397261	3.255584236	0.001131593	0.03711884	ST3GAL6-AS1
ENSG00000163702	81.54184715	0.674508987	0.290636646	-3.365736578	0.000763396	0.027242254	IL17RC
ENSG00000230409	37.21282564	-1.093731168	0.391711152	3.73658901	0.000186533	0.00910452	TCEA1P2
ENSG00000136603	288.4785348	-0.575603253	0.155848314	4.368005051	1.25E-05	0.001010944	SKIL
ENSG00000164105	301.4180674	0.520749996	0.131494368	-4.586224556	4.51E-06	0.000422076	SAP30
ENSG00000137440	490.6230144	0.7298345	0.10878201	-7.112820795	1.14E-12	4.41E-10	FGFBP1
ENSG00000118804	17.69755514	1.37367076	0.651678075	-3.286197131	0.001015499	0.033956245	STBD1
ENSG00000052802	3597.584659	0.747231147	0.082911966	-9.283395917	1.64E-20	1.53E-17	MSMO1
ENSG00000196656	19.3129026	1.955555708	0.603070367	-4.033260388	5.50E-05	0.003476603	
ENSG00000184985	35.26930665	1.227631245	0.436303083	-3.745180457	0.000180264	0.008911971	SORCS2
ENSG00000164211	1336.956198	0.503337533	0.085795041	-6.280628834	3.37E-10	9.70E-08	STARD4
ENSG00000112972	4630.614605	0.85085904	0.052982972	-16.20790035	4.43E-59	2.19E-55	HMGCS1
ENSG00000131495	337.1190849	-0.525398196	0.139677273	4.4231345	9.73E-06	0.000814784	NDUFA2
ENSG00000185641	158.9857277	0.775444147	0.216242583	-4.321793199	1.55E-05	0.001205569	
ENSG00000251493	390.4288688	0.649297237	0.139581745	-5.221153847	1.78E-07	2.59E-05	FOXD1
ENSG00000113161	7902.393155	0.635040295	0.054034178	-11.92092139	9.21E-33	1.59E-29	HMGCR
ENSG00000169220	280.8266316	0.560284555	0.151090515	-4.386784622	1.15E-05	0.0009407	RGS14
ENSG00000247627	76.68076647	-0.862018172	0.295816911	3.807248642	0.000140521	0.007280952	MTND4P12
ENSG00000164292	2241.754531	-1.021702083	0.062146574	16.60938295	5.96E-62	3.43E-58	RHOBTB3
ENSG00000120437	1995.729373	0.839948482	0.07603561	-11.27925883	1.66E-29	2.49E-26	ACAT2
ENSG00000181577	445.4144633	0.935362788	0.118695429	-8.240298332	1.72E-16	1.08E-13	LINC03040
ENSG00000112559	443.3927042	0.509480573	0.148500378	-4.137180635	3.52E-05	0.002383691	MDF1
ENSG00000289047	122.8453228	-0.520167436	0.236040061	3.246248068	0.001169369	0.037961039	
ENSG00000186480	3416.231027	1.042626683	0.061379901	-17.15176476	6.10E-66	5.26E-62	INSIG1
ENSG00000130675	197.7197087	0.640560433	0.178368957	-4.306639049	1.66E-05	0.001276748	MNX1
ENSG00000196456	135.4419583	0.529680252	0.219300715	-3.383722963	0.000715101	0.025893716	ZNF775
ENSG00000001630	219.5760083	0.511500389	0.198274333	-3.491716916	0.000479927	0.019415366	CYP51A1
ENSG00000128564	982.4353783	0.760945056	0.089133467	-8.861212179	7.91E-19	6.35E-16	VGF
ENSG00000176532	65.62810358	0.96383906	0.319999316	-3.883038244	0.000103159	0.005788329	PRR15
ENSG00000146834	422.2659949	0.538594872	0.118293774	-5.102191798	3.36E-07	4.47E-05	MEPCE
ENSG00000228668	6.91861558	3.103096778	1.310730062	-3.183164159	0.00145675	0.044843469	TRGV5P
ENSG00000147155	751.8149689	0.63192932	0.09032721	-7.378386645	1.60E-13	7.09E-11	FBP
ENSG00000232024	25.84236448	-1.269086299	0.552823293	3.407710023	0.000655105	0.024362564	LSM12P1
ENSG00000182759	90.11865355	0.935909029	0.274786828	-4.194243106	2.74E-05	0.001916379	MAFA
ENSG00000204839	1027.140792	0.540995076	0.082970868	-6.904555227	5.04E-12	1.79E-09	MROH6
ENSG00000179886	444.8976757	0.585313812	0.114257648	-5.626189672	1.84E-08	3.49E-06	TIGD5
ENSG00000079459	10204.90465	0.73319437	0.043463459	-16.98483193	1.06E-64	7.34E-61	FDFT1
ENSG00000198576	284.5964799	0.505938226	0.14990291	-4.094605763	4.23E-05	0.002800959	ARC
ENSG00000198435	715.8546023	0.533436704	0.095770476	-6.026820952	1.67E-09	4.18E-07	NRARP
ENSG00000167106	595.3319625	0.903537742	0.100195605	-9.33768241	9.85E-21	9.71E-18	EEIG1
ENSG00000160326	529.5893848	0.540866566	0.114675099	-5.252097717	1.50E-07	2.22E-05	SLC2A6
ENSG00000167136	532.4850942	0.647826541	0.115091265	-6.097003748	1.08E-09	2.85E-07	ENDOG
ENSG00000173457	5961.409218	0.738188416	0.064528473	-11.64172785	2.53E-31	3.97E-28	PPP1R14B
ENSG00000171219	605.5795648	0.534577677	0.107304783	-5.487497089	4.08E-08	7.07E-06	CDC42BPG
ENSG00000175602	2609.522732	0.597756331	0.085621112	-7.322728804	2.43E-13	1.04E-10	CCDC85B
ENSG00000188070	402.6203549	0.634770101	0.129442656	-5.442694874	5.25E-08	8.71E-06	ZFTA
ENSG00000184545	154.8136315	0.909819066	0.179472045	-5.611875785	2.00E-08	3.73E-06	DUSP8
ENSG00000167799	184.8800561	0.566006883	0.196052802	-3.734183227	0.000188325	0.009166053	NUDT8
ENSG00000244398	169.3144301	0.619893668	0.235263741	-3.567168307	0.00036086	0.015392518	RPL36AP37
ENSG00000168040	231.9496178	0.775876806	0.175838694	-5.017960955	5.22E-07	6.58E-05	FADD
ENSG00000172893	6668.327026	0.519921593	0.044632961	-11.86271887	1.85E-32	3.04E-29	DHCR7
ENSG00000149809	1200.360055	0.724072154	0.098170722	-7.719824881	1.16E-14	6.18E-12	TM7SF2
ENSG00000170955	170.4560238	0.843828676	0.208020789	-4.71878638	2.37E-06	0.000239392	CAVIN3
ENSG00000173599	622.366466	0.595534163	0.106380186	-6.058473172	1.37E-09	3.54E-07	PC
ENSG00000213260	17.26152316	-1.728144374	0.621814571	3.737778425	0.000185653	0.009074404	YWHAZP5
ENSG00000122861	3021.610212	0.506620584	0.056482904	-9.250697579	2.23E-20	2.03E-17	PLAU
ENSG00000107954	161.6353167	0.63224472	0.194707984	-4.028734102	5.61E-05	0.003537732	NEURL1
ENSG00000120549	164.7803085	0.573112587	0.195120352	-3.771393178	0.000162339	0.00821405	KIAA1217
ENSG00000077150	2283.218521	0.564197369	0.063164107	-9.221404461	2.93E-20	2.59E-17	NFKB2
ENSG00000110921	1578.739633	0.73875439	0.067910863	-11.12875848	9.09E-29	1.25E-25	MVK
ENSG00000167552	6362.891955	0.507618101	0.062671572	-8.383732489	5.13E-17	3.40E-14	TUBA1A
ENSG00000185950	902.5375034	0.609734254	0.09401347	-6.885875	5.74E-12	2.02E-09	IRS2
ENSG00000293057	91.71816609	-1.305444789	0.26467505	5.510384959	3.58E-08	6.27E-06	
ENSG00000257179	24.9360431	-1.621315591	0.480905292	4.181472849	2.90E-05	0.002006916	HMGN2P6
ENSG00000165804	545.4379426	0.556964426	0.112788186	-5.457480112	4.83E-08	8.13E-06	ZNF219
ENSG00000259020	44.37911172	-2.045641828	0.387820987	5.813559753	6.12E-09	1.38E-06	
ENSG00000100836	91.47722244	0.630260242	0.288475017	-3.269641279	0.001076839	0.035696039	PABPN1

Table S3.1. List of differently expressed genes by DuoMYC **5**.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Symbol
ENSG00000232573	232.8337439	0.901028106	0.18626645	-5.405177638	6.47E-08	1.04E-05	RPL3P4
ENSG00000137825	393.2801852	0.552878362	0.120590614	-5.140241663	2.74E-07	3.80E-05	ITPKA
ENSG00000256338	110.647322	-0.841577214	0.222730363	4.479995705	7.46E-06	0.000645258	RPL41P2
ENSG00000234797	196.4322686	-0.82276851	0.168917368	5.436753178	5.43E-08	8.96E-06	RPS3AP6
ENSG00000138623	395.3797771	0.573130199	0.125978138	-5.120354885	3.05E-07	4.14E-05	SEMA7A
ENSG00000140264	1461.09599	-0.548008471	0.082334272	7.042658889	1.89E-12	7.07E-10	SERF2
ENSG00000135736	216.7870469	0.509217324	0.162165555	-3.908143019	9.30E-05	0.005358144	CCDC102A
ENSG00000126603	260.3455188	0.502988139	0.154398444	-3.999882965	6.34E-05	0.003912171	GLIS2
ENSG00000158792	1105.761719	0.564869922	0.081321421	-7.3210941	2.46E-13	1.04E-10	SPATA2L
ENSG00000227827	127.5358408	0.841754912	0.222910466	-4.479976753	7.47E-06	0.000645258	PKD1P2
ENSG00000179588	345.2915684	0.524022343	0.135768688	-4.499660539	6.81E-06	0.000603775	ZFPM1
ENSG00000167508	4376.647092	1.140927141	0.056159148	-20.41087406	1.34E-92	4.62E-88	MVD
ENSG00000184207	1756.648795	0.551605398	0.075881496	-7.62511121	2.44E-14	1.24E-11	PGP
ENSG00000183971	2716.585436	0.679375601	0.068581331	-10.17938612	2.45E-24	2.82E-21	NPW
ENSG00000121104	486.593292	0.560913164	0.106006788	-5.7758906	7.65E-09	1.67E-06	FAM117A
ENSG00000189343	924.790806	0.818684267	0.19581988	-4.840679797	1.29E-06	0.000138608	RPS2P46
ENSG00000109107	985.9506263	0.566604888	0.078288483	-7.603957215	2.87E-14	1.42E-11	ALDOC
ENSG00000068137	713.410686	0.593336772	0.095612544	-6.630419863	3.35E-11	1.09E-08	PLEKHH3
ENSG00000185813	1693.42127	0.579434301	0.064082242	-9.339762035	9.66E-21	9.71E-18	PCYT2
ENSG00000161682	704.5112821	0.604801591	0.09304567	-6.904926245	5.02E-12	1.79E-09	FAM171A2
ENSG00000169710	47793.43142	0.591668505	0.038730423	-14.97576923	1.06E-50	3.32E-47	FASN
ENSG00000212724	269.1398349	0.569852639	0.149361582	-4.480927491	7.43E-06	0.000645258	KRTAP2-3
ENSG00000136463	276.5510604	0.529242368	0.140814991	-4.412773639	1.02E-05	0.000848603	TACO1
ENSG00000130584	122.9082355	0.642198043	0.250461808	-3.519996466	0.000431553	0.017792133	ZBTB46
ENSG00000125968	1551.855576	-1.15223023	0.074701448	15.61537494	5.72E-55	2.19E-51	ID1
ENSG00000172216	868.7133666	0.591211364	0.091770483	-6.842064554	7.81E-12	2.72E-09	CEBPB
ENSG00000227063	414.1756969	0.507738825	0.159816365	-3.935532181	8.30E-05	0.004921987	RPL41P1
ENSG00000088766	247.0165799	1.100125637	0.168779921	-6.948802058	3.68E-12	1.35E-09	CRLS1
ENSG00000131069	1400.502519	0.834006505	0.0695881	-12.21964967	2.44E-34	4.43E-31	ACSS2
ENSG00000160951	81.6317488	1.032413114	0.272725266	-4.504366469	6.66E-06	0.000592072	PTGER1
ENSG00000142279	276.9560526	0.557655715	0.152761387	-4.337922139	1.44E-05	0.001138416	WTIP
ENSG00000160469	163.4458425	0.639410596	0.197321059	-4.022756304	5.75E-05	0.003615546	BRSK1
ENSG00000197380	82.71752111	0.586888524	0.288443338	-3.167516254	0.001537471	0.046480584	DACT3
ENSG00000233493	248.0456034	1.307459463	0.170419189	-8.04440388	8.67E-16	5.16E-13	TMEM238
ENSG00000108106	8307.44442	0.667583626	0.055289693	-12.29093759	1.01E-34	1.94E-31	UBE2S
ENSG00000130522	3876.91953	1.059266613	0.054681683	-19.52185632	7.16E-85	1.24E-80	JUND
ENSG00000129968	2197.557767	0.616371871	0.07227457	-8.839945374	9.58E-19	7.51E-16	ABHD17A
ENSG00000213638	222.0524333	0.582457966	0.162152454	-4.296956354	1.73E-05	0.001324918	ADAT3
ENSG00000171223	1547.507543	0.54413062	0.069917823	-8.115129849	4.85E-16	2.94E-13	JUNB
ENSG00000141933	318.6311258	1.177675967	0.139245913	-8.794602353	1.44E-18	1.10E-15	TPGS1
ENSG00000174951	453.0068486	0.533843069	0.113565591	-5.230734911	1.69E-07	2.47E-05	FUT1
ENSG00000130203	2861.890541	0.58315692	0.067560961	-8.922756196	4.55E-19	3.74E-16	APOE
ENSG00000104881	334.2703606	0.503048617	0.134954026	-4.383339071	1.17E-05	0.000953449	PPP1R13L
ENSG00000181588	1949.959942	0.595031888	0.062504375	-9.798276509	1.15E-22	1.20E-19	MEX3D
ENSG00000142235	214.5983803	0.871407207	0.166125128	-5.76712887	8.06E-09	1.74E-06	LMTK3
ENSG00000142544	753.3707306	0.696970692	0.090690168	-8.037078751	9.20E-16	5.38E-13	CTU1
ENSG00000171119	219.2310769	0.511406013	0.185821018	-3.616783974	0.000298286	0.013272904	NRTN
ENSG00000115286	74.04604345	-0.847792303	0.276452551	3.920981336	8.82E-05	0.005149288	NDUFS7
ENSG00000070404	539.0006092	0.55144177	0.105537949	-5.713775467	1.10E-08	2.28E-06	FSTL3
ENSG00000063169	291.8908586	0.542993624	0.14685976	-4.365563944	1.27E-05	0.001018163	BICRA
ENSG00000176531	340.0521146	0.591104899	0.127046993	-5.206620322	1.92E-07	2.78E-05	PHLDB3
ENSG00000099625	436.676006	0.626729185	0.14156248	-5.016857464	5.25E-07	6.59E-05	CBARP
ENSG00000130751	90.39282556	0.649548988	0.273778218	-3.396069031	0.000683611	0.024989464	NPAS1
ENSG00000130748	717.6072024	0.65642527	0.114808134	-6.181060391	6.37E-10	1.77E-07	TMEM160
ENSG00000105327	599.443056	0.633009335	0.113432105	-6.052980165	1.42E-09	3.63E-07	BBC3
ENSG00000104856	768.9317272	0.555665945	0.104308584	-5.807617707	6.34E-09	1.41E-06	RELB
ENSG00000087074	1136.377849	0.547651084	0.085315419	-6.819431634	9.14E-12	3.15E-09	PPP1R15A
ENSG00000105376	324.7828012	0.700778117	0.133324951	-5.771414449	7.86E-09	1.71E-06	ICAM5
ENSG00000130164	3325.454457	0.769860787	0.053266305	-14.6419108	1.52E-48	4.36E-45	LDLR
ENSG00000118046	2148.531028	0.518546403	0.066547737	-8.120587584	4.64E-16	2.86E-13	STK11
ENSG00000128346	13.6013708	1.652585146	0.83207299	-3.21116497	0.00132198	0.041890627	C22orf23
ENSG00000234965	293.1512178	1.174382018	0.145638734	-8.415583351	3.91E-17	2.65E-14	SHISA8
ENSG00000244486	44.79138268	0.812639739	0.426851777	-3.146761154	0.001650898	0.04885864	SCARF2
ENSG00000250479	2190.047135	0.546703684	0.063099844	-8.964597805	3.11E-19	2.62E-16	CHCHD10
ENSG00000241360	338.471307	1.045647469	0.149064821	-7.417498626	1.19E-13	5.35E-11	PDXP
ENSG00000100078	387.3853599	0.937269127	0.132515931	-7.468655072	8.10E-14	3.73E-11	PLA2G3
ENSG00000188064	466.6501608	0.517882353	0.111172297	-5.194358499	2.05E-07	2.94E-05	WNT7B
ENSG00000198911	6478.286562	0.559946447	0.044293	-12.82421532	1.20E-37	2.59E-34	SREBF2
ENSG00000251322	860.8463556	0.599024279	0.088032157	-7.190236277	6.47E-13	2.63E-10	SHANK3
ENSG00000100290	442.9782556	1.403386766	0.113355425	-12.61421942	1.76E-36	3.58E-33	BIK
ENSG00000160285	2461.560651	1.07421771	0.067152491	-16.17184648	7.97E-59	3.44E-55	LSS
ENSG00000198888	13159.48781	-0.594063615	0.132465423	5.33225338	9.70E-08	1.51E-05	MT-ND1
ENSG00000198763	26614.62814	-0.702079284	0.136677877	5.383596506	7.30E-08	1.17E-05	MT-ND2
ENSG00000198804	94865.41161	-0.869397801	0.06378354	12.82849767	1.14E-37	2.59E-34	MT-CO1
ENSG00000198712	53617.40817	-0.859431899	0.110706169	7.891463143	2.99E-15	1.69E-12	MT-CO2
ENSG00000228253	171.9837818	-0.603822747	0.24460751	3.45289859	0.000554597	0.021673895	MT-ATP8
ENSG00000198899	28157.53597	-0.850358629	0.124885797	7.50824485	5.99E-14	2.79E-11	MT-ATP6
ENSG00000198938	49873.0884	-0.753901247	0.107608368	7.640688901	2.16E-14	1.11E-11	MT-CO3
ENSG00000198886	64138.84252	-0.713868051	0.125229429	7.063623356	1.62E-12	6.22E-10	MT-ND4
ENSG00000198786	22948.29882	-0.756262868	0.162043474	5.164754651	2.41E-07	3.38E-05	MT-ND5
ENSG00000198695	8257.40849	-0.755124321	0.110516185	7.222223208	5.11E-13	2.13E-10	MT-ND6
ENSG00000198727	24903.58321	-0.813565197	0.14369152	5.81970741	5.90E-09	1.35E-06	MT-CYB

Table S3.2. List of differently expressed genes by Omomyc.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Symbol
ENSG00000228502	231.5547249	-0.906950305	0.171412268	5.814086253	6.10E-09	4.10E-07	EEF1A1P11
ENSG00000135766	403.3851924	0.88130259	0.133652286	-7.010836275	2.37E-12	3.17E-10	EGLN1
ENSG00000117318	1236.353667	-1.18598749	0.071055122	16.87507041	6.86E-64	2.48E-60	ID3
ENSG00000162522	2238.690773	0.589747583	0.074595268	-8.230933882	1.86E-16	4.37E-14	NHSL3
ENSG00000117407	248.5514861	0.586543662	0.176160581	-4.070260406	4.70E-05	0.000704634	ARTN
ENSG00000228594	231.7405887	0.508312429	0.164995471	-3.839645522	0.000123212	0.00154477	FNDC10
ENSG00000224114	98.69811744	-0.632183514	0.363991443	3.025097505	0.00248553	0.016114048	
ENSG00000163041	2828.281156	0.738040275	0.068838184	-10.97449635	5.07E-28	2.61E-25	H3-3A
ENSG00000237330	134.6474068	0.611026073	0.24230234	-3.472878634	0.000514908	0.004735144	RNF223
ENSG00000185519	226.3217743	0.503925706	0.163600806	-3.83547857	0.00012532	0.001564046	FAM131C
ENSG00000263528	79.07064158	0.715278515	0.279443144	-3.524583256	0.00042415	0.00408065	IKBKE
ENSG00000160094	204.6155299	0.574808054	0.161695161	-4.245579877	2.18E-05	0.000380339	ZNF362
ENSG00000175793	9147.604108	0.556930207	0.051128941	-11.12758749	9.21E-29	4.99E-26	SFN
ENSG00000158246	278.4923329	0.558307608	0.147181232	-4.440390594	8.98E-06	0.000187791	TENT5B
ENSG00000132196	681.7037312	0.580620736	0.097312865	-6.393763561	1.62E-10	1.65E-08	HSD17B7
ENSG00000204060	128.5879628	0.770803946	0.213441226	-4.334099757	1.46E-05	0.000281615	FOXO6
ENSG00000225630	437.7967628	-1.617167301	0.138655084	11.91286872	1.01E-32	6.46E-30	MTND2P28
ENSG00000198744	170.8838661	-0.763131266	0.207633863	4.392058085	1.12E-05	0.00022821	MTCO3P12
ENSG00000158292	291.749485	0.700976005	0.132176948	-5.801168254	6.59E-09	4.36E-07	GPR153
ENSG00000169174	3659.335143	0.767210323	0.053957359	-14.41319606	4.27E-47	5.79E-44	PCSK9
ENSG00000116133	12856.50025	0.537091876	0.043835385	-12.45946642	1.24E-35	9.61E-33	DHCR24
ENSG00000204253	186.1866783	-0.942485225	0.179640285	5.778163656	7.55E-09	4.90E-07	HNRNPCP2
ENSG00000144485	1706.874218	0.50081256	0.097027796	-5.630429478	1.80E-08	1.05E-06	HES6
ENSG00000234072	89.19256337	0.605385153	0.285947735	-3.207735688	0.001337844	0.010078486	GTF3C2-AS2
ENSG00000115738	144.755946	-0.896210776	0.191146908	5.279089017	1.30E-07	5.81E-06	ID2
ENSG00000249992	179.0815529	0.699250062	0.189615198	-4.387373688	1.15E-05	0.000232009	TMEM158
ENSG00000137440	490.6230144	0.751670567	0.107535836	-7.375947996	1.63E-13	2.64E-11	FGFBP1
ENSG00000178458	793.0581631	0.515790618	0.107036529	-5.324061782	1.01E-07	4.76E-06	H3P16
ENSG00000052802	3597.584659	0.593674364	0.083215228	-7.452549207	9.16E-14	1.63E-11	MSMO1
ENSG00000123933	1325.123198	0.586206096	0.078140424	-7.843207992	4.39E-15	9.15E-13	MXD4
ENSG00000249855	104.4959411	-0.702634314	0.267385541	3.576001644	0.000348889	0.003496318	EEF1A1P19
ENSG00000112972	4630.614605	0.554045444	0.053200247	-10.63900081	1.96E-26	9.66E-24	HMGCS1
ENSG00000131495	337.1190849	-0.716364011	0.136260931	5.766517891	8.09E-09	5.22E-07	NDUFA2
ENSG00000164438	101.5774309	0.610999573	0.25941726	-3.359517071	0.000780788	0.006506369	TLX3
ENSG00000185641	158.9857277	0.785281547	0.213613087	-4.383872513	1.17E-05	0.000233889	
ENSG00000251493	390.4288688	0.672104692	0.137979958	-5.410512843	6.28E-08	3.26E-06	FOXD1
ENSG00000250182	342.7482363	-0.593239224	0.157469266	4.433496156	9.27E-06	0.000192415	EEF1A1P13
ENSG00000164292	2241.754531	-0.780643081	0.06119767	12.97486806	1.70E-38	1.67E-35	RHOBTB3
ENSG00000120437	1995.729373	0.597494084	0.076416615	-8.118119504	4.73E-16	1.07E-13	ACAT2
ENSG00000181577	445.4144633	1.109683485	0.116557978	-9.818823255	9.34E-23	3.49E-20	LINC03040
ENSG00000135604	85.39272502	0.537511112	0.27994605	-3.063430349	0.002188151	0.014605203	STX11
ENSG00000186480	3416.231027	0.841092239	0.061423884	-13.88868561	7.42E-44	8.93E-41	INSIG1
ENSG00000165215	241.218457	0.640856719	0.182052749	-4.228940578	2.35E-05	0.00040502	CLDN3
ENSG00000189143	1882.415566	0.522935654	0.062549369	-8.640023601	5.62E-18	1.56E-15	CLDN4
ENSG00000050327	348.4882086	0.51549062	0.129849225	-4.581049477	4.63E-06	0.000110151	ARHGEF5
ENSG00000130675	197.7197087	0.527560846	0.180725343	-3.72374625	0.000196288	0.002235105	MNX1
ENSG00000128564	982.4353783	0.608741979	0.08931285	-7.197428441	6.14E-13	9.11E-11	VGF
ENSG00000009950	110.6217194	0.771747015	0.236576291	-4.055253821	5.01E-05	0.000738116	MLXIPL
ENSG00000198910	505.385481	0.533824752	0.121513711	-4.951275552	7.37E-07	2.42E-05	L1CAM
ENSG00000182759	90.11865355	0.793191469	0.280542476	-3.730791619	0.000190879	0.002184072	MAFA
ENSG00000179886	444.8976757	0.620059385	0.112588549	-5.975663898	2.29E-09	1.77E-07	TIGD5
ENSG00000187954	204.5359387	0.540043744	0.165930308	-3.992974471	6.52E-05	0.000910888	ZFTRAF1
ENSG00000079459	10204.90465	0.67112732	0.043397678	-15.56730061	1.21E-54	2.19E-51	FDFT1
ENSG00000275342	1395.427567	0.527672073	0.081335239	-6.861714135	6.80E-12	8.64E-10	PRAG1
ENSG00000198576	284.5964799	0.501887162	0.148036119	-4.088318006	4.35E-05	0.000663279	ARC
ENSG00000198435	715.8546023	0.536552605	0.094799899	-6.100139012	1.06E-09	8.76E-08	NRARP
ENSG00000160326	529.5893848	0.695840265	0.111833283	-6.641235207	3.11E-11	3.58E-09	SLC2A6
ENSG00000196358	126.8372484	1.03673489	0.223892372	-5.221906594	1.77E-07	7.43E-06	NTNG2
ENSG00000173457	5961.409218	0.529501713	0.064786365	-8.42033694	3.75E-17	9.46E-15	PPP1R14B
ENSG00000181649	3156.996748	0.537570666	0.052842471	-10.35745393	3.87E-25	1.68E-22	PHLDA2
ENSG00000175567	112.8095575	0.535132572	0.255078218	-3.167317448	0.001538523	0.011246165	UCP2
ENSG00000129757	438.4045358	0.696116286	0.121029272	-6.196050315	5.79E-10	5.23E-08	CDKN1C
ENSG00000172893	6668.327026	0.586128051	0.044360731	-13.39530209	6.44E-41	6.98E-38	DHCR7
ENSG00000149809	1200.360055	0.895542308	0.096889157	-9.527410784	1.61E-21	5.29E-19	TM7SF2
ENSG00000173599	622.366466	0.623934549	0.10514343	-6.377334626	1.80E-10	1.82E-08	PC
ENSG00000181274	174.9348825	0.521627324	0.185558598	-3.64051627	0.000272092	0.002873119	FRAT2
ENSG00000107954	161.6353167	0.734805746	0.188386772	-4.571625579	4.84E-06	0.000113971	NEURL1
ENSG00000099194	35425.96812	0.576931618	0.045842619	-12.75490059	2.93E-37	2.44E-34	SCD
ENSG00000111319	942.3544039	0.626237262	0.091282085	-7.232159573	4.75E-13	7.15E-11	SCNN1A
ENSG00000186897	223.4724216	0.543357122	0.163867574	-4.036031853	5.44E-05	0.000784171	C1QL4
ENSG00000110921	1578.739633	0.687627726	0.067565519	-10.43792266	1.66E-25	7.51E-23	MVK
ENSG00000135114	133.3675076	0.504943616	0.219929538	-3.275750843	0.001053815	0.008289694	OASL
ENSG00000159335	7276.9351	0.56522062	0.04972025	-11.58685936	4.80E-31	2.89E-28	PTMS
ENSG00000167588	100.4517483	-0.796893059	0.222089366	4.320415823	1.56E-05	0.000294846	GPD1
ENSG00000167552	6362.891955	0.597228892	0.062332335	-9.797084653	1.16E-22	4.05E-20	TUBA1A
ENSG00000185950	902.5375034	0.552766045	0.093740836	-6.319230855	2.63E-10	2.61E-08	IRS2
ENSG00000293057	91.71816609	-1.755344008	0.263039813	7.103416663	1.22E-12	1.67E-10	
ENSG00000165804	545.4379426	0.539755071	0.111897507	-5.326649483	1.00E-07	4.71E-06	ZNF219
ENSG00000133935	1950.67499	0.534006255	0.071597197	-7.765062574	8.16E-15	1.61E-12	ERG28
ENSG00000232573	232.8337439	0.808551252	0.186703356	-4.93720715	7.92E-07	2.56E-05	RPL3P4
ENSG00000276168	300.8779606	0.621753463	0.157679185	-4.585201677	4.54E-06	0.000109184	RN7SL1
ENSG00000136378	180.395016	0.573010506	0.178706173	-3.96454255	7.35E-05	0.001008385	ADAMTS7
ENSG00000256338	110.647322	-0.995135337	0.217430837	5.181120774	2.21E-07	9.02E-06	RPL41P2
ENSG00000234797	196.4322686	-0.930265672	0.165968877	6.102918221	1.04E-09	8.75E-08	RPS3AP6
ENSG00000140264	1461.09599	-0.691512223	0.081589257	8.795476904	1.42E-18	4.29E-16	SERF2
ENSG00000204991	361.7753124	0.505743191	0.128676312	-4.540925637	5.60E-06	0.000128273	SPIRE2
ENSG00000118898	768.0483567	0.538475023	0.114666593	-5.213588072	1.85E-07	7.69E-06	PPL
ENSG00000179588	345.2915684	0.751557518	0.130455455	-6.225328818	4.81E-10	4.49E-08	ZFPM1
ENSG00000167508	4376.647092	0.969176323	0.056119594	-17.42899484	4.97E-68	2.69E-64	MVD
ENSG00000008710	248.0056433	0.509367243	0.179022395	-3.663277664	0.000249008	0.002680889	PKD1
ENSG00000103495	275.9037838	0.595242997	0.153752796	-4.518987053	6.21E-06	0.000139652	MAZ
ENSG00000183971	2716.585436	0.555463959	0.068619188	-8.404563463	4.29E-17	1.06E-14	NPW
ENSG00000196557	720.954687	0.529418791	0.089827481	-6.316151648	2.68E-10	2.64E-08	CACNA1H
ENSG00000121104	486.593292	0.662432996	0.103758157	-6.799366209	1.05E-11	1.26E-09	FAM117A
ENSG00000131094	125.9100296	0.750321331	0.209035563	-4.312063471	1.62E-05	0.000304185	C1QL1
ENSG00000189343	924.790806	0.778801368	0.19612932	-4.671694498	2.99E-06	7.76E-05	RPS2P46
ENSG00000109107	985.9506263	0.562928272	0.07753852	-7.620008514	2.54E-14	4.74E-12	ALDOC
ENSG00000185813	1693.42127	0.605898229	0.063479495	-9.811538676	1.00E-22	3.63E-20	PCYT2
ENSG00000161682	704.5112821	0.729277421	0.091313722	-8.324412069	8.47E-17	2.04E-14	FAM171A2
ENSG00000169710	47793.43142	0.652561398	0.038627948	-16.74332719	6.33E-63	1.72E-59	FASN
ENSG00000212901	610.0331719	-0.534879493	0.128658651	4.752246159	2.01E-06	5.60E-05	KRTAP3-1
ENSG00000006283	99.86910272	0.732852761	0.258007829	-3.729111402	0.000192156	0.00219581	CACNA1G
ENSG00000101493	553.4924869	0.516593201	0.10015384	-5.634737042	1.75E-08	1.04E-06	ZNF516
ENSG00000101132	847.2085792	-0.614021982	0.130217591	5.280997905	1.28E-07	5.80E-06	PFDN4
ENSG00000130702	341.354272	0.598497283	0.140407938	-4.858367232	1.18E-06	3.62E-05	LAMA5
ENSG00000130584	122.9082355	0.636808281	0.247423887	-3.511913412	0.000444893	0.004215488	ZBTB46
ENSG00000125968	1551.855576	-0.885729368	0.073442429	12.29435998	9.71E-35	6.58E-32	ID1
ENSG00000177732	1972.206033	0.634888761	0.067076952	-9.7371526	2.09E-22	7.09E-20	SOX12
ENSG00000172216	868.7133666	0.565383697	0.091200911	-6.601876723	4.06E-11	4.53E-09	CEBPB
ENSG00000227063	414.1756969	0.527910961	0.157547558	-4.079821446	4.51E-05	0.000682634	RPL41P1
ENSG00000088766	247.0165799	0.534647272	0.179790238	-3.775216621	0.000159869	0.001893903	CRLS1

Table S3.2 List of differently expressed genes by Omomyc.

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Symbol
ENSG00000131069	1400.502519	0.837906692	0.069046898	-12.36570198	4.01E-35	2.89E-32	ACSS2
ENSG00000099864	685.5219725	0.5091333	0.097113591	-5.710562203	1.13E-08	7.05E-07	PALM
ENSG00000160951	81.6317488	0.863361897	0.277952186	-3.940853437	8.12E-05	0.00109126	PTGER1
ENSG00000160469	163.4458425	0.590292557	0.197197501	-3.808537034	0.000139791	0.001714192	BRSK1
ENSG00000233493	248.0456034	1.236094604	0.16980363	-7.664697202	1.79E-14	3.47E-12	TMEM238
ENSG00000108106	8307.44442	0.57142213	0.055290247	-10.52573051	3.86E-26	1.82E-23	UBE2S
ENSG00000130522	3876.91953	1.0442947	0.054462921	-19.32214537	3.50E-83	3.79E-79	JUND
ENSG00000133250	221.0634428	0.541373298	0.187217933	-3.713783828	0.000204183	0.002296902	ZNF414
ENSG00000129968	2197.557767	0.595902889	0.071971195	-8.583869058	9.17E-18	2.37E-15	ABHD17A
ENSG00000183248	144.5506403	0.755187554	0.203935862	-4.408567286	1.04E-05	0.000213494	PRR36
ENSG00000141933	318.6311258	0.760094178	0.14215349	-5.845914668	5.04E-09	3.50E-07	TPGS1
ENSG00000177051	576.5200848	0.502006229	0.104797786	-5.297823386	1.17E-07	5.33E-06	FBXO46
ENSG00000104881	334.2703606	0.66762989	0.130271707	-5.638074052	1.72E-08	1.02E-06	PPP1R13L
ENSG00000079432	1352.873251	0.513385951	0.090304717	-6.124888492	9.07E-10	7.68E-08	CIC
ENSG00000129946	88.89869943	0.893627602	0.266490694	-4.138669722	3.49E-05	0.000552444	SHC2
ENSG00000181588	1949.959942	0.527260426	0.062260946	-8.777524292	1.67E-18	4.89E-16	MEX3D
ENSG00000142235	214.5983803	0.879649279	0.164346406	-5.860916505	4.60E-09	3.24E-07	LMTK3
ENSG00000063169	291.8908586	0.604623914	0.143553148	-4.81462445	1.47E-06	4.33E-05	BICRA
ENSG00000130513	6882.632022	0.548477768	0.049323752	-11.3591096	6.68E-30	3.81E-27	GDF15
ENSG00000099625	436.676006	0.671478985	0.13962633	-5.351921952	8.70E-08	4.23E-06	CBARP
ENSG00000105327	599.443056	0.654850899	0.112312045	-6.280274272	3.38E-10	3.21E-08	BBC3
ENSG00000069399	336.5760576	0.682579507	0.122636486	-6.04138937	1.53E-09	1.23E-07	BCL3
ENSG00000261221	468.7050989	0.510145163	0.110714538	-5.134821658	2.82E-07	1.10E-05	ZNF865
ENSG00000130164	3325.454457	0.668574957	0.053150449	-12.78996134	1.87E-37	1.68E-34	LDLR
ENSG00000100330	87.01161036	0.661472467	0.258921914	-3.505408619	0.000455907	0.004294644	MTMR3
ENSG00000099889	234.9609478	0.589859533	0.159969146	-4.357310992	1.32E-05	0.00025747	ARVCF
ENSG00000234965	293.1512178	1.112013915	0.14494194	-8.038645578	9.08E-16	2.01E-13	SHISA8
ENSG00000250479	2190.047135	0.616061041	0.06246946	-10.12823168	4.14E-24	1.66E-21	CHCHD10
ENSG00000241360	338.471307	0.938005482	0.149023281	-6.737335029	1.61E-11	1.92E-09	PDXP
ENSG00000100078	387.3853599	0.940378875	0.131443023	-7.542309182	4.62E-14	8.34E-12	PLA2G3
ENSG00000188064	466.6501608	0.661741866	0.108230058	-6.544181969	5.98E-11	6.55E-09	WNT7B
ENSG00000198911	6478.286562	0.630593385	0.044014391	-14.49591063	1.29E-47	1.99E-44	SREBF2
ENSG00000251322	860.8463556	0.642813357	0.087013277	-7.768487962	7.94E-15	1.59E-12	SHANK3
ENSG00000160285	2461.560651	1.032158319	0.066919235	-15.60368063	6.87E-55	1.49E-51	LSS

Table S3.3 Significantly differently expressed pathways by DuoMYC **5**

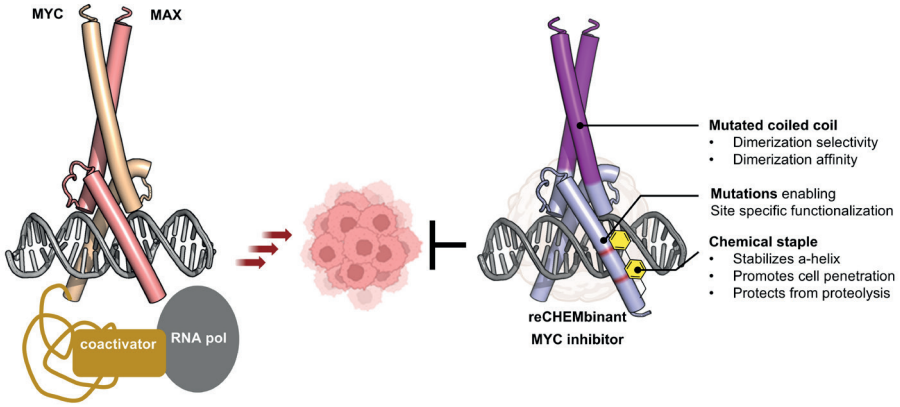
	Term	ES	NES	NOM p-val	FDR q-val	FWER p-val	Tag %	Gene %
0	Cholesterol Homeostasis	-0.82591	-2.70139	0	0	0	43/71	9.18%
1	TNF-alpha Signaling via NF-kB	-0.66797	-2.47178	0	0	0	79/191	9.98%
2	Protein Secretion	0.600019	2.2817	0	0	0	45/94	11.56%
3	Hypoxia	-0.58229	-2.1561	0	0	0	61/197	10.27%
4	Myogenesis	-0.59031	-2.1455	0	0	0	68/186	14.58%
5	Apoptosis	-0.51435	-1.8545	0	0.000221	0.001	57/153	15.32%
7	Estrogen Response Early	-0.48117	-1.77126	0	0.000552	0.003	71/191	15.06%
6	G2-M Checkpoint	0.436277	1.806425	0	0.001127	0.003	74/199	12.53%
8	Androgen Response	-0.49545	-1.69233	0.001475	0.002525	0.014	14/96	3.31%
9	Estrogen Response Late	-0.43921	-1.62558	0.002639	0.004418	0.03	44/182	9.68%
11	TGF-beta Signaling	0.441022	1.501438	0.027933	0.010767	0.043	jul-53	0.44%
10	Mitotic Spindle	-0.40223	-1.51045	0.006649	0.015954	0.121	67/199	14.97%
13	IL-6/JAK/STAT3 Signaling	-0.43601	-1.43656	0.024578	0.025806	0.223	30/77	19.89%
12	Adipogenesis	-0.38988	-1.43658	0.006757	0.028276	0.222	51/192	14.51%
14	DNA Repair	0.333783	1.351208	0.011152	0.032113	0.155	32/148	6.80%

Table S3.4 Significantly differently expressed pathways by Omomyc

	Term	ES	NES	NOM p-val	FDR q-val	FWER p-val	Tag %	Gene %
0	Cholesterol Homeostasis	-0.80753	-2.71233	0	0	0	33/71	5.10%
1	G2-M Checkpoint	0.567371	2.331756	0	0	0	98/199	11.29%
2	Myogenesis	-0.57259	-2.19715	0	0	0	61/186	12.54%
3	TNF-alpha Signaling via NF-kB	-0.57255	-2.19688	0	0	0	86/191	16.26%
4	Estrogen Response Early	-0.55963	-2.12312	0	0	0	81/191	14.43%
5	Protein Secretion	0.546611	2.028327	0	0	0	48/94	17.15%
6	DNA Repair	0.504444	1.976716	0	0	0	57/148	11.31%
7	Hypoxia	-0.50292	-1.95086	0	0.000253	0.001	58/197	13.44%
8	Estrogen Response Late	-0.47198	-1.77545	0	0.001897	0.007	59/182	11.83%
9	Apoptosis	-0.42526	-1.59697	0.001639	0.007045	0.036	44/153	11.38%
11	Mitotic Spindle	-0.40772	-1.56464	0	0.007446	0.05	66/199	13.18%
10	Androgen Response	-0.45156	-1.56892	0.006557	0.008377	0.05	21/96	9.06%
12	Interferon Alpha Response	-0.4253	-1.47311	0.016694	0.016817	0.125	35/93	16.80%
13	Wnt-beta Catenin Signaling	-0.46487	-1.39303	0.056058	0.03391	0.254	18/40	15.61%
14	TGF-beta Signaling	0.397714	1.313159	0.069825	0.045323	0.268	okt-53	5.11%

Table S3.5 Overview of all synthesized proteins with their respective calculated and observed masses and retention times

Miniprotein	Sequence	Yield (%)	Molecular weight calculated	Molecular weight observed	LCMS method	Retention time (min)	Observed ions
MiniMYC 1	$(AcATEENVKRRTHNVLERQRRNELKRSFFG)_2K$	16%	7137.1	3736.5	B	2.023	1190.3, 1020.4, 893.1, 793.9, 714.6, 649.7, 595.7, 549.9, 510.7, 476.7
MiniMYC monomer 2	AcATEENVKRRTHNVLERQRRNELKRSCFA	16%	3483.0	3482.1	B	1.895	1161.5, 871.7, 697.5, 581.4, 498.5, 436.2
MiniMYC 3	$(AcATEENVKRRTHNVLERQRRNELKRSSCFA)_2$-Linker	31%	7260	7259.4	B	2.125	1210.9, 1038.1, 908.3, 807.5, 726.9, 660.9, 605.9, 559.4, 519.5
MonoMYC 4	AcATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELEN-NEKAPKVVILKKATAYILSCA	0.5%	6978.1	6977.5	B	2.641	1396.6, 1163.8, 997.8, 873.1, 776.2, 698.7, 635.3,
DuoMYC 5	$(AcATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELEN-NEKAPKVVILKKATAYILSCA)_2$-linker	23%	14134	14133	B	2.897	1414.28, 1285.8, 1178.7, 1088.1, 1010.6, 943.2, 884.4, 832.4, 786.3, 744.9, 707.7, 674.0, 643.4
MonoMYC 6	$AcVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKV-VILKKATAYILSVA_{\beta}C$	6%	6532.7	6531.6	B	2.586	1089.5, 934.0, 817.4, 726.7, 649.0, 594.4
DuoMYC 7	$(AcVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAP-KVVILKKATAYILSA_{\beta}C)_2$-linker	12%	13359	13358	B	2.888	1114.0, 1028.5, 955.1, 891.1, 835.9, 786.8, 743.1, 704.1, 668.9, 637.1, 608.2, 581.7, 557.6
FITC labeled DuoMYC 8	$(AcATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELEN-NEKAPKVVILKKATAYILSCA)_2$-linker + 1/2/3 FITC	80%	1x FITC: 14524 2x FITC: 14913 3x FITC: 15303	1x FITC: 14523 2xFITC: 14913 3xFITC: 15301	B	2.942	1xFITC: 1453.39, 1321.35, 1211.26, 1118.14, 1038.4169, 969.12, 908.75, 855.23, 807.03, 765.41, 727.19, 692.52 2xFITC: 1492.19, 1356.72, 1243.74, 1147.99, 1066.21, 995.13, 933.06, 878.17, 829.44, 785.89, 746.60, 711.19 3xFITC: 1392.27, 1276.16, 1178.11, 1093.99, 1021.20, 957.37, 901.11, 851.00, 806.37, 765.47
FITC-labelled Omomyc 9	AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRD-QIPELENNEKAPKVVILKKATAYILSVQAETQKLISEIDLL-RKQNEQLKHKLEQLRNSCA +1/2/3/4x FITC	20%	1xFITC: 11917 2XFITC: 12306 3XFITC: 12695 4xFITC: 13085	1xFITC: 11916 2XFITC: 12306 3xFITC: 12695 4xFITC: 13084	B	3.3	1xFITC: 1324.92, 1190.49, 1084.30, 994.11, 917.56, 852.09. 794.42, 747.73 2xFITC: 1368.26, 1231.53, 1119.76, 1026.44, 947.64, 879.95, 821.36, 770.15. 724.84, 684.20 3xFITC: 1411.48, 1270.54, 1155.94, 1058.95, 977.57, 907.74, 847.42, 794.40, 747.72, 706.24, 669. 18 4xFITC: 1454.82, 1309.44, 1190.49, 1091.37, 1007.49, 935.60, 873.29, 818.71, 770.67, 727.91

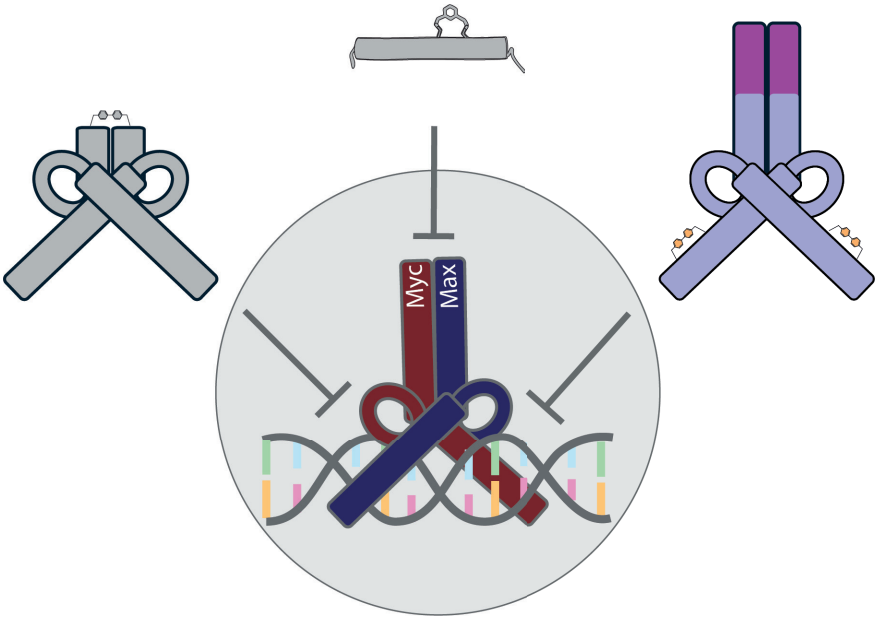


Adapted from: J. P. Kahler, B. D. Ellenbroek, V. E. van der Noord, B. van de Water, S. J. Pomplun, "ReCHEMbinant stapling enhances intracellular delivery and bioactivity of engineered protein inhibitors" *Chem* 2025, 102839.

The data not included in this thesis (uncut gels and list of differently expressed genes) can be found here: https://doi.org/10.1016/j.chempr.2025.102839

Chapter 4

ReCHEMbinant Stapling Enhances Intracellular Delivery and Bioactivity of Engineered Protein Inhibitors



Authors:

Jan Pascal Kahler, **Brecht D. Ellenbroek**, Vera E. van der Noord, Bob van de Water, Sebastian J. Pomplun

Abstract

Protein therapeutics have transformed drug discovery by enabling modulation of challenging targets inaccessible to small molecules. However, most proteins lack the ability to penetrate cells, where many critical drug targets reside. Here, we present reCHEMbinant protein engineering, a strategy designed to generate synthetically enhanced proteins with improved structural stability, serum resistance, and cellular uptake. Applying this approach to Omomyc, a protein-based MYC inhibitor, we developed several reCHEMbinant stapled variants (HeloMYCs) exhibiting low-nanomolar DNA-binding affinity. Notably, the i, i+7 biphenyl-stapled construct HeloMYC-1421 outperformed Omomyc across several functional assays, including potent inhibition of MYC-driven gene expression in luciferase reporter assays and selective antiproliferative effects in MYC-dependent cells. Live-cell imaging showed that these enhanced effects result from significantly improved cellular uptake. Transcriptional reprogramming was further confirmed by RNA-seq. Together, our findings establish reCHEMbinant engineering as a chemically defined strategy for stapling entire recombinant proteins to enhance their intracellular bioactivity.

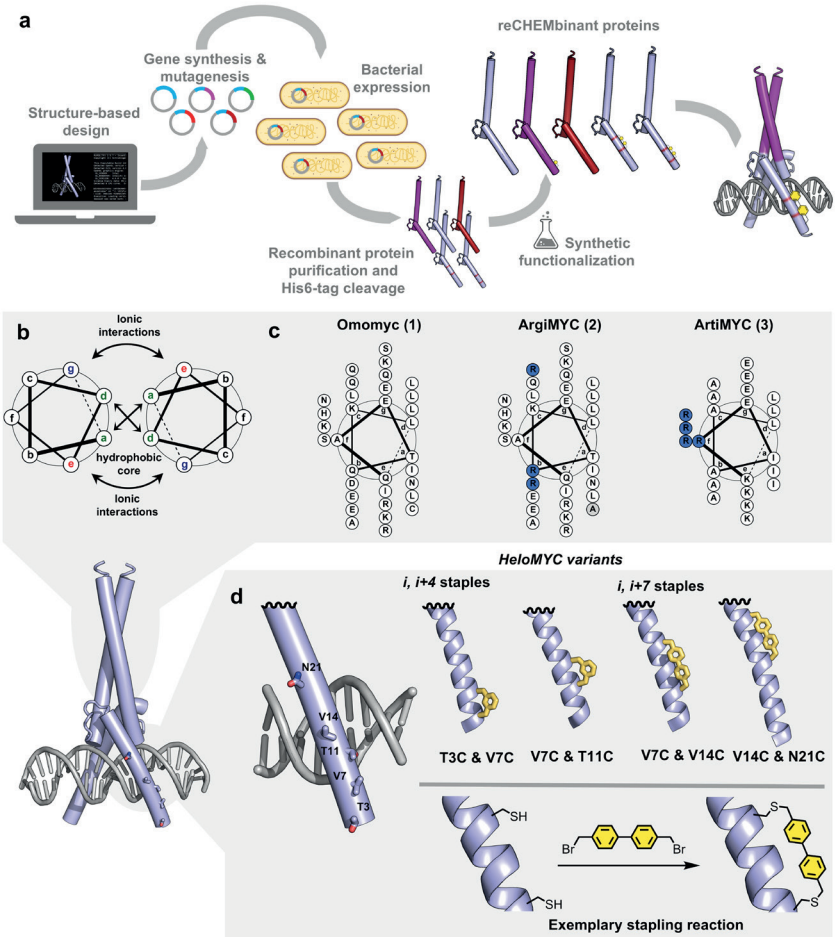
4.1 Introduction

Protein therapeutics have revolutionized drug discovery and expanded treatment options for challenging diseases.[1] With the advent of recombinant technology, protein drugs have enabled the targeting of drug targets previously elusive for small molecules.[2] Notable examples include antibodies like Herceptin[3] and Keytruda,[4] as well as other therapeutic proteins such as erythropoietin (EPO)[5] and interferon. [6] Due to their complex structures and extensive surfaces, protein drugs can recognize their targets with high specificity and potency. However, their large size and physicochemical properties prevent them from entering cells, making them effective primarily for extracellular targets. Many compelling drug targets, including thousands of protein-protein interactions (PPIs) and protein-nucleic acid interactions (PNAIs) do, however, reside inside cells, making them hardly accessible for protein based drugs.[7,8]

Strategies that enable the direct cytosolic delivery of therapeutic proteins would have transformative potential for targeting intracellular pathways. Traditional approaches have relied on fusion to cell-penetrating peptides (CPPs), such as cyclic polyarginines, to facilitate uptake.[9–11] These methods have enabled the delivery of cargos like nanobodies[9] and ubiquitin[10] into cells, primarily in model systems. However, CPPs often exhibit nonspecific toxicity and poor serum stability, limiting their translational potential. An alternative approach developed by the Raines group involves transient masking of surface-exposed carboxylates as esters to promote protein uptake, demonstrated using GFP.[12,13] While conceptually elegant, this strategy is hampered by the serum lability of esters, which precludes therapeutic application. Other efforts have focused on miniaturizing protein domains into proteomimetic peptides that mimic key binding motifs.[14–18] These constructs are often stabilized by chemical linkers or hydrocarbon staples and can achieve enhanced cell permeability thanks to their decreased size. However, replacing full-length proteins with short peptides can lead to reduced affinity or specificity, particularly when the target recognition depends on complex tertiary structures.

We sought to investigate a more general strategy for intracellular protein delivery by directly modifying full-length recombinant proteins. Our approach combines rationally introduced point mutations with compact, stable chemical modifications to enhance structural integrity and promote cell permeability, without relying on large fusion tags or inherently toxic motifs.

For studies, we selected Omomyc, a well-characterized, dominant-negative variant of the transcription factor MYC.[19–24] Omomyc forms homodimers that bind to the MYC recognition sequence (E-box) and competitively inhibit MYC's oncogenic transcriptional activity. It offers several advantages as a model system for evaluating intracellular protein delivery strategies. First, Omomyc is biophysically well



▲**Figure 4.1 ReCHEMbinant protein design.** (**a**) Design, generation and synthetic modification of reCHEMbinant proteins. After structure evaluation and mutagenesis his-tagged proteins are expressed in E. coli cells. The his-tag is then cleaved and, if applicable, the purified proteins are reacted with a staple. (**b**) Helical wheel showing heptad pattern and the interactions between the α-helices in a coiled-coil. (**c**) Helical wheel representation of the coiled-coil helices of Omomyc (**1**), ArgiMYC (**2**) and ArtiMYC (**3**). (**d**) DNA-binding basic part of Omomyc with residues highlighted that have been identified for mutation and stapling. Two residues are mutated into Cys and reacted with a staple.

understood, and quantitative assays such as DNA-binding electrophoretic mobility shift assays (EMSAs) enable precise evaluation of structural and functional integrity following chemical modification. Second, functional cellular readouts, including luciferase-based MYC reporter assays, provide rapid assessment of intracellular delivery and bioactivity. Third, Omomyc is recombinantly expressed in bacteria and

is of moderate size, facilitating high-resolution analytical characterization such as intact protein mass spectrometry. Notably, Omomyc exhibits some intrinsic cell-penetrating activity;[25] however, conflicting reports suggest that its native uptake is limited and insufficient for robust therapeutic effects.[11] These combined features, biochemical tractability, functional relevance, and borderline permeability, make Omomyc an ideal system for testing and optimizing chemically enhanced intracellular delivery strategies with potential therapeutic impact

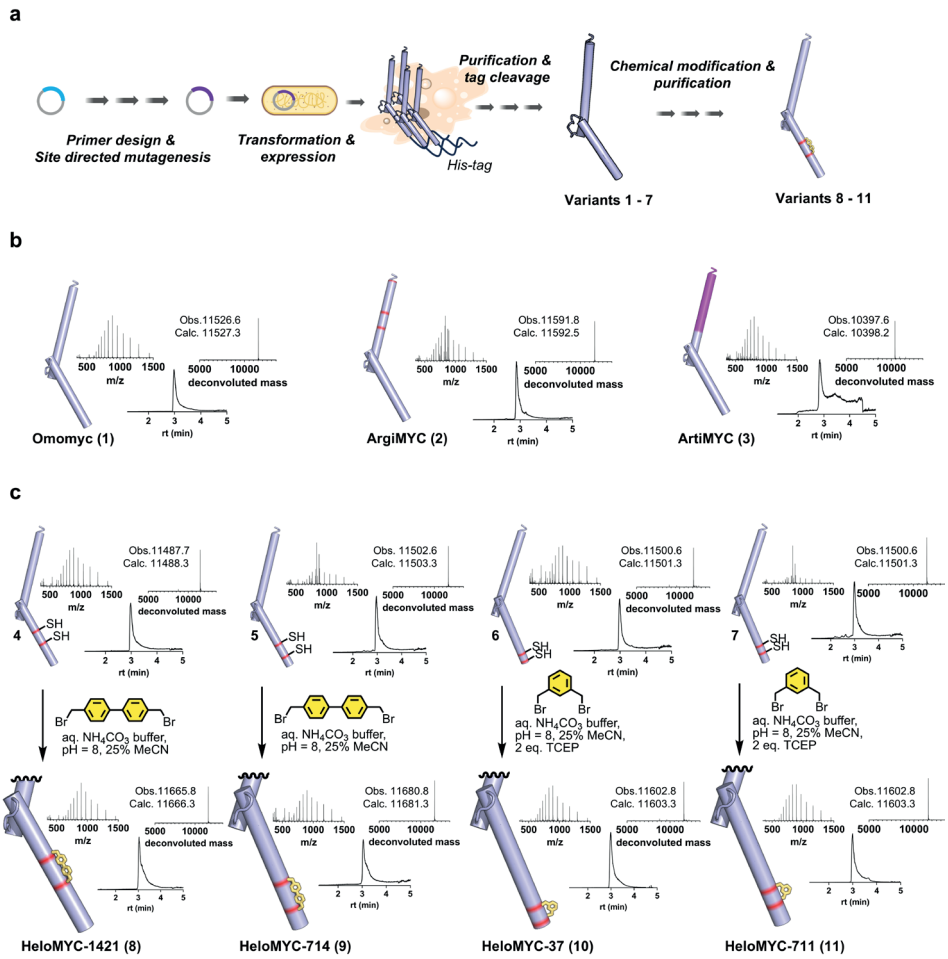
Here, we developed a reCHEMbinant protein engineering strategy that combines structure-guided design and recombinant expression of Omomyc analogs with targeted chemical modification. By introducing a series of sequence-based mutations and covalent stabilization elements, we generated stapled Omomyc variants with enhanced physicochemical and functional properties. This approach led to the identification of a reCHEMbinant class of compounds capable of efficient cellular uptake and potent inhibition of MYC-driven transcription. Our lead construct, HeloMYC-1421, exhibits low-nanomolar DNA-binding affinity and submicromolar potency in a MYC reporter assay, substantially outperforming unmodified Omomyc. Fluorescence imaging confirmed markedly improved cellular penetration, and further characterization demonstrated selective anti-proliferative effects in MYC-dependent cancer cells, with no observable toxicity in MYC-independent lines. These findings establish reCHEMbinant stapling as a promising strategy for enabling intracellular delivery and therapeutic modulation of recombinant protein biologics.

4.2 Results

4.2.1 Design and synthesis of reCHEMbinant transcription factors

To engineer protein variants with high affinity for E-Box DNA and improved bioactivity, particularly with a focus on cell penetration, we used Omomyc as a template. Our approach involved preserving its overall structural and functional characteristics while introducing targeted mutations to optimize bioactivity. We pursued two key strategies: (1) incorporating canonical sequence mutations to increase arginine content, and (2) introducing chemical modifications aimed at enhancing structural stability and promoting cell penetration.

The cell penetrating properties of polycationic arginine rich peptides have been extensively reported.[26] Studies about Omomyc's intrinsic cell penetrating properties have revealed that the arginines in the DNA binding helix are critical for this ability. [25] However, an excessive arginine content is associated with non-specific toxicity.[27–29] In an attempt to balance potential cell penetrating efficacy and safety, we decided to modify the coiled-coil region of Omomyc, matching its arginine content to that of its DNA-binding helix. We first selected solvent exposed residues, which based on structural analysis, are not involved in the protein dimerization and generated



▲**Figure 4.2 Generation of artificial transcription factors.** (**a**) General workflow for generation of HeloMYC synthetic transcription factors by recombinant expression and chemical modification. (**b**) Omomyc (**1**, left) and coiled-coil modified Omomyc variants ArgiMYC (**2**) and ArtiMYC (**3**) were obtained by recombinant expression in good purity. (**c**) Point-mutated Omomyc derivatives obtained by recombinant expression in excellent purity (upper part) were reacted with the desired staple to generate the different HeloMYC proteins **8-11**.

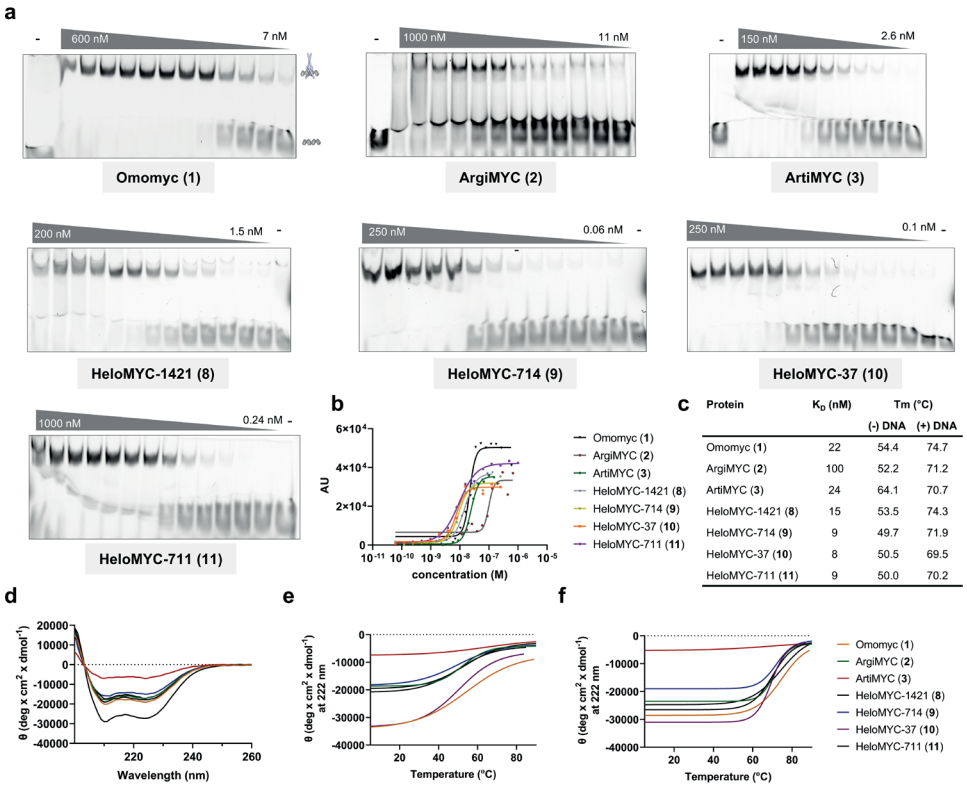
a variant, ArgiMYC, in which we mutated three residues (Q64, D71, and Q86) to arginine (**Figure 4.1c**, middle). In a second approach, we designed a fully artificial coiled coil (residues 61-83), also with increased arginine content compared to the wild-type protein (**Figure 4.1c**, right). The predicted AlphaFold structure of the resulting ArtiMYC dimer aligned well with wild-type Omomyc (**Figure S4.1**). ArgiMYC and ArtiMYC possess an overall slightly increased arginine content of 11.1% and 13.3%, respectively, compared to Omomyc with 9.7%.

To chemically enhance the protein variants, we employed a semisynthetic strategy we refer to as the *reCHEMbinant* approach. Chemical staples, which stabilize the α-helical conformation of peptides, have been shown to improve cell penetration, stability, and overall bioactivity.[30–32] These staples are covalent linkages between amino acid side chains spaced one or two helix turns apart (*i, i+4* or *i, i+7*, respectively). The enhanced cell penetration of stapled helices is attributed to increased lipophilicity from the staple and the locked α-helical structure, which engages all backbone amides in intramolecular hydrogen bonding.[31] In our *reCHEMbinant* approach, we adhered to the following steps: (1) identify suitable *i, i+4* and *i, i+7* positions based on structural analysis, (2) mutate the selected residues pairwise to cysteines, (3) recombinantly express and purify the proteins, and (4) staple the proteins using cysteine-reactive bifunctional reagents (**Figure 4.1a**).[33,34] We selected four pairwise mutations, two for *i, i+4* and two for *i, i+7* stapling, modifying the backside of the DNA-binding helix to avoid disrupting DNA recognition (**Figure 4.1d**).

We prepared all plasmids with the desired mutations and expressed the proteins in ArcticXpress cells. We purified via Ni-NTA affinity chromatography and subsequently cleaved the tags releasing the pure protein products. Omomyc (1), ArgiMYC (2) and ArtiMYC (3) were all obtained in satisfying yield (8-14 mg/L culture) and purity, as shown by LC-MS (**Figure 4.2a-b**). These three variants (1, 2 and 3) were not further synthetically modified. We also expressed and purified the bis-cysteine variants and further modified the proteins with either 4,4'-bis(bromomethyl)-1,1'-biphenyl or 1,3-bis(bromomethyl)benzene. The *i, i + 4* cysteine pairs required addition of reducing agent (TCEP) during the reaction, to prevent disulfide formation. All site specific protein functionalizations resulted in clean conversions and we obtained the four reCHEMbinant stapled proteins, HeloMYC-37, -711, -714 and -1421 (**Figure 4.2c**).

4.2.2 Biophysical characterization and binding evaluation

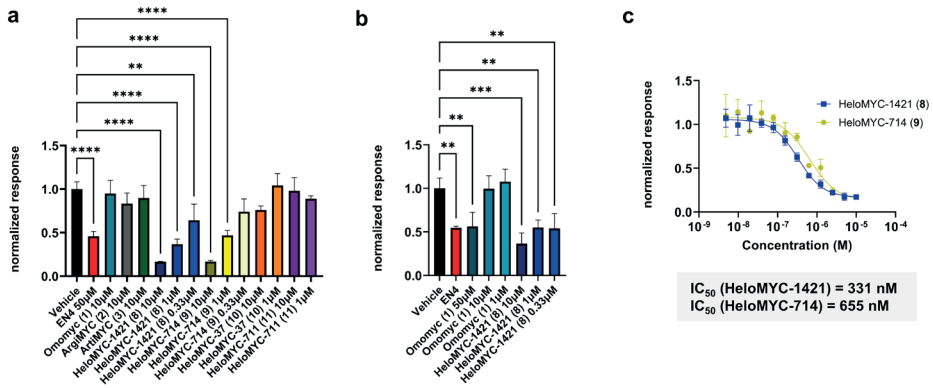
We measured the affinity to E-Box DNA of all reCHEMbinant transcription factors using EMSA. Based on their design, all our engineered protein variants are supposed to homodimerize, bind to E-Box DNA and result in a detectable shift when measured via native gel electrophoresis. To visualize the DNA and the protein-DNA complexes we incubated all variants with 5'-IRD700-labelled dsDNA containing the E-Box sequence CACGTG and resolved the mixture on a native gel. We then quantified the bound fraction of DNA and calculated K_D values. The control protein Omomyc binds DNA with a K_D of 22 nM. ArgiMYC (**2**) showed a K_D of 100 nM, but consistently resulted in smeared bands, potentially indicating partial unspecific binding. ArtiMYC (**3**) binds E-Box with 24 nM, an affinity comparable to Omomyc. The HeloMYC variants resulted in K_D values between 15 and 8 nM, which represents a 2-3 fold improvement in affinity compared to their unmodified parent compound Omomyc (**Figure 4.3a-c**).



Protein	K_D (nM)	Tm (°C) (-) DNA	Tm (°C) (+) DNA
Omomyc (**1**)	22	54.4	74.7
ArgiMYC (**2**)	100	52.2	71.2
ArtiMYC (**3**)	24	64.1	70.7
HeloMYC-1421 (**8**)	15	53.5	74.3
HeloMYC-714 (**9**)	9	49.7	71.9
HeloMYC-37 (**10**)	8	50.5	69.5
HeloMYC-711 (**11**)	9	50.0	70.2

▲**Figure 4.3 ReChembinant transcription factors bind to E-Box DNA with high potency.** (**a**) EMSAs of ReChembinant transcription factors with 5'-IRD700-labelled dsDNA (sequence 5′-IRD700-ACCCCAC**CACGTG**GTGCCT–3′) show DNA binding of Omomyc (**1**), ArgiMYC (**2**), ArtiMYC (**3**) and HeloMYC variants (**8-11**). DNA construct is incubated with miniprotein and the mixture resolved by native gel electrophoresis. DNA binding is seen as the DNA-protein complex running higher than free DNA. (**b**) Plot of curves used for K_D-value determination. AU values were obtained by quantifying the signal of the bound fraction of DNA. (**c**) Summary of all K_D-values as determined by EMSA and structural melting points as determined by CD in absence and presence of E-Box DNA. (**d**) CD spectra of all constructs. (**e-f**) CD melting curves of all constructs in absence (**e**) and presence (**f**) of E-Box DNA, measured at 222 nm.

We used circular dichroism (CD) to evaluate structural features and structural stability of all variants. CD spectra of all proteins exhibit defined minima at around 208 nm and 222 nm indicating α-helical folding (**Figure 4.3d**). Only ArtiMYC (**3**) showed a weaker CD signal, albeit still indicating alpha helicity. This effect might result from aggregation but we did not investigate further this phenomenon. We also performed CD melting experiments in absence and presence of DNA. In all cases the presence of DNA resulted in significant structural stabilization (approximately +20 °C), further confirming the functional DNA binding of these variants.

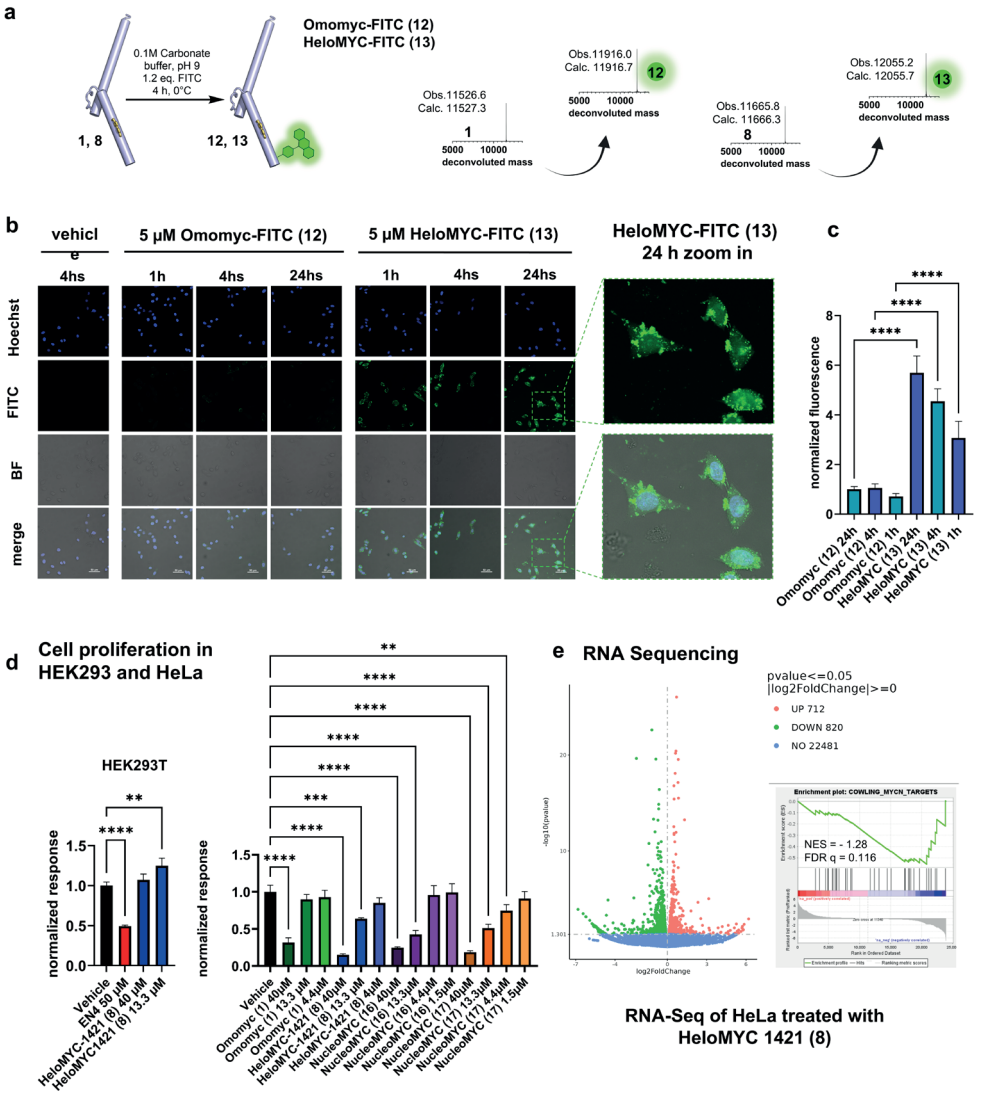


▲**Figure 4.4 Biphenyl-stapled HeloMYC miniproteins significantly reduce MYC-related gene transcription in a reporter assay.** (**a,b**) HEK293T (**a**) and HeLa (**b**) cells transiently transfected with a MYC-dependent luciferase gene were incubated with transcriptional repressor miniproteins or small molecule MYC inhibitor EN4 for one day and luciferase activity measured subsequently. HeloMYC-714 and HeloMYC-1421 show significant downregulation of MYC-related luciferase expression. (**c**) IC_{50} values for MYC-regulated luciferase expression inhibition by HeloMYC-1421 and HeloMYC-714 were measured in transiently transfected HEK293T cells. **Statistics.** A one way ANOVA was performed to compare the effect of miniprotein treatment on normalized signal showing that there was a statistical difference between treatments. (a) F (14, 30) = 22.76, P<0.0001; (b) F (7, 16) = 13.21, P<0.0001. * p<0.05, ** p<0.01, *** p<0.001.

Taken together, our set of biophysical assays indicate robust DNA binding activity and alpha helical folding for the four stapled HeloMYC variants (**8-11**). ArgiMYC and ArtiMYC showed less convincing results in the EMSA assay and CD analysis, respectively.

4.2.3 Evaluation of bioactivity

To evaluate the intracellular effects of our reCHEMbinant protein variants, we performed MYC-responsive reporter gene assays, which revealed that only the *i,i*+7-stapled constructs HeloMYC-714 and HeloMYC-1421 elicited significant cellular activity. These assays use a firefly luciferase gene activated by MYC/MAX dimers, with a constitutively expressed renilla luciferase under control of the CMV promoter for normalization. Following transfection of HEK293T with dual reporter DNA, cells were treated for 24 hours with Omomyc or our mutated reCHEMbinant transcription factors. We also included the small molecule covalent MYC inhibitor EN4 as a positive control.[35] Based on previous reports the main expected mechanism of action of Omomyc and analogs would be forming dimers and occupying the MYC/MAX DNA binding sites.[25] Among all protein variants only the *i, i+7*-stapled



proteins HeloMYC-714 and -1421 displayed a significant downregulation of MYC reporter activity (**Figure 4.4a**). Both HeloMYC derivatives exhibit roughly the same effect on MYC activity at 1 µM concentration as EN4 does at 50 µM. We then went on to measure their EC_{50} values in the reporter gene assay which we determined to be 655 nM for HeloMYC-714 and 331 nM for HeloMYC-1421 (**Figure 4.4c**, data from 3 independent experiments in **Figure S4.2**). To corroborate these findings, we tested a subset of compounds (HeloMYC-1421 (**8**), Omomyc (**1**), and EN4) in HeLa cells, where HeloMYC-1421 (**8**) again showed potent reporter downregulation at sub-micromolar concentrations, whereas Omomyc (**1**) was only effective at 50 µM (**Figure 4.4b**). In our control experiments we confirmed that the constructs had

◀**Figure 4.5 HeloMYC-1421 is cell permeable and inhibits cell proliferation of cancer cell lines.** **(a)** Omomyc (**1**) and HeloMYC-1421 (**8**) were reacted with FITC in carbonate buffer. **(b)** Fluorescence microscopy of live HeLa cells treated with HeloMYC-FITC (**13**) or Omomyc-FITC (**12**), imaged at 1 h, 4 h and 24 h. **(c)** Quantification of fluorescence shows superior cell permeability of HeloMYC-FITC (**13**) (**d**) Cell titer glow (CTG) assays to assess HeloMYC-1421 (**8**) activity proliferation of HEK293T and CTG assay in HeLa cells with HeloMYC-1421 (**8**) and NucleoMYC **16** and **17**. e) HeLa cells were treated with HeloMYC-1421 (**8**) at 10 **μ**M for 3 days. RNA was extracted and sequenced. Shown are up- (rosa) and down- (green) regulated genes with p value < 0.05 and on the right the GSEA plot for the COWLING_MYC_Targets is shown. **Statistics.** A one way ANOVA was performed to compare the effect of miniprotein treatment on normalized signal showing that there was a statistical difference between treatments. (b) F (5, 186) = 132.8, P<0.0001; (d) HEK293T F (3, 8) = 75.49, P<0.0001; HeLa F (14, 28) = 49.17, P<0.0001. * p<0.05, ** p<0.01, *** p<0.001.

no impact on the constitutively expressed CMV-driven renilla luciferase, but that HeloMYC-1421 and EN4 exhibited a negative effect on constitutively expressed CMV driven firefly luciferase. We could experimentally exclude that these compounds act as direct enzymatic inhibitors, by adding the constructs to control-transfected cells directly before cell lysis and luciferase read-out. These experiments therefore, in addition to direct MYC inhibition, hint at a likely more complex inhibition pathway mechanism (**Figure S4.3** for validation data). Despite this uncertainty, the MYC reporter gene assays indicated strong intracellular activity for the *i, i+7* biphenyl-stapled proteins HeloMYC-714 and HeloMYC-1421, while all other variants, including Omomyc, showed little to no effects.

To investigate whether the improved activity of the HeloMYC variants in the reporter gene assay is a result of improved cell penetration we turned to fluorescence microscopy. Given that the protein variants **2** and **3** with increased arginine content and the variants with the *i, i* + 4 staples had not shown any cell activity, we only proceeded with analyzing HeloMYC-1421 and comparing it to the unmodified parent protein Omomyc.

Fluorescence imaging revealed that HeloMYC-1421 enters cells more efficiently than Omomyc. We labeled HeloMYC-1421 and Omomyc with fluorescein isothiocyanate (FITC) and purified the resulting variants **12** and **13** (**Figure 4.5a**). We then performed live cells imaging in HeLa cells upon incubation with the fluorescent proteins (at 5 μM) for either 1 h, 4 h or 24 h. At all timepoints HeloMYC-FITC (**13**) is taken up into cells significantly more than Omomyc-FITC (**12**). We observed that the uptake increases over time and also enhanced colocalization with the nucleus (**Figure 4.5c**). This time course points toward an endocytosis bases cell entry mechanism followed by partial endosomal escape. **Figure 4.5b** shows that part of the compound remains localized in endosomal structures (green puncta), while a substantial fraction is also diffusely distributed throughout the cytosol and nucleus. The more efficient cell penetration compared to Omomyc can likely explain the enhanced efficacy of the

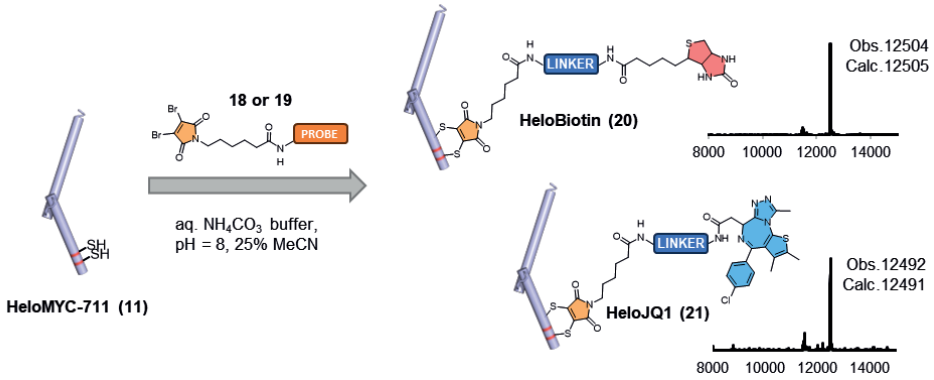
HeloMYC variants in the reporter gene assay.

As alpha helix stapling is often associated with improved resistance against proteases, we also performed a serum stability assay to compare unmodified Omomyc (1) to HeloMYC-1421 (**8**). We incubated both proteins in 10% human serum and assessed their half lives by LCMS (**Figure S4.4**). HeloMYC-1421 (**8**) displayed a significant improved stability compared to Omomyc (4.8 h compared to 1.3 h). In addition to the better cell uptake, also the stability can be part of the explanation of the improved bioactivity.

HeloMYC-1421 (**8**) inhibits the proliferation of cancer cells without inducing nonspecific cytotoxicity in MYC-independent cells. To assess potential off-target toxicity, we treated HEK293T cells—which do not rely on MYC for survival—with HeloMYC-1421 (**8**). The cells tolerated treatment well, showing no loss of viability after 3 days at concentrations up to 40 μM (**Figure 4.5d**). In contrast, MYC-dependent HeLa cells showed a significant reduction in proliferation at concentrations as low as 13.3 μM. To enhance nuclear delivery, we generated HeloMYC-1421 variants bearing nuclear localization sequences (NLS) at either the N- or C-terminus (NucleoMYC **16** and **17**, **Figure S4.5**). NucleoMYC **17** showed a modest improvement in potency, although overall the NLS addition had only a moderate effect. Collectively, these data show that HeloMYC-1421 and its NLS-tagged variants inhibit cancer cell proliferation at 13.3 μM (compounds **8** and **16**) and 4.4 μM (compound **17**) and, importantly, the reCHEMbinant HeloMYC did not exhibit cytotoxicity in MYC-independent HEK293T cells.

HeloMYC-1421 (8) modulates MYC activity in HeLa cells, leading to significant transcriptional changes. After 3 days of treatment, RNA-seq analysis revealed 712 upregulated and 820 downregulated genes. Gene Set Enrichment Analysis (GSEA) uncovered a trend toward coordinated downregulation of the Cowling MYC target gene set, reflecting partial suppression of MYC-driven transcription. Although the enrichment was modest, this subtle effect aligns with the expected mechanism of the inhibitor, which tempers MYC function rather than fully abolishing it, unlike the complete loss seen in knockout models used for the gene set comparisons. This nuanced modulation supports the therapeutic potential of HeloMYC-1421.

As a final experiment, we explored whether the stapling position in reCHEMbinant transcription factors could also serve as a site for the conjugation of additional functional groups, such as small molecules or affinity handles. To this end, we designed and synthesized two trifunctional probes containing a dibromomaleimide moiety for stapling via two cysteines, with an additional functionality introduced through substitution at the maleimide nitrogen. Probe 18 carried a biotin handle, while probe 19 contained the BET inhibitor JQ1, a compound known to modulate MYC expression. HeloMYC-711 (11) was reacted with these probes in aqueous



▲**Figure 6. Synthesis of reCHEMbinant protein conjugates.** The reaction between HeloMYC-711 (11) and trifuctional dibromomaleimide-biotin (18) or dibromomaleimide-JQ1 (19) resulted in efficient conversion to HeloBiotin (20) and HeloJQ1 (21).

buffer, resulting in efficient conversion to the corresponding conjugates, HeloBiotin and HeloJQ1 (**Figure 4.6**). Although we did not pursue a detailed characterization of these conjugates, the straightforward synthetic strategy demonstrates a feasible route to generate chemically diversified protein variants that may be of use in future biochemical studies or for probing functional effects.

4.3 Discussion and conclusions

In this study, we developed a chemically enhanced protein engineering strategy targeting the nuclear DNA binding site of the oncogenic transcription factor MYC. Efficient intracellular delivery of proteins remains a key challenge for protein therapeutics.[9,10,12,13,36] Starting from Omomyc, a known MYC inhibitor with some intrinsic cell-penetrating properties, we explored two distinct strategies to improve cell permeability and bioactivity: (1) increasing the arginine content, inspired by polycationic cell-penetrating peptides, and (2) chemically modifying the DNA-binding domains with peptide staples.

While the arginine-rich variants, ArgiMYC and ArtiMYC and the *i*, *i* + 4 staples showed no detectable improvements in cell activity, the *i*, *i* + 7 stapled proteins, HeloMYC-714 and HeloMYC-1421, exhibited significantly enhanced cellular activity. Via fluorescent microscopy we confirmed that the biphenyl staple leads to a substantially improved uptake in cells. Interestingly, despite minimal impact on helical stability, the *i*, *i* + 7 stapling led to pronounced gains in cellular uptake, potentially due to the amphipathic nature of these constructs. The *i*, *i* + 4 xylene staple might not add enough hydrophobicity to the helix backside to achieve the same effect. Also, just slightly increasing the arginine content, as attempted in ArgiMYC and ArtiMYC, did

not lead to the desired enhanced cellular uptake. Three more arginines distributed along such a long sequence might not be enough for such activity. While arginine rich peptide often are cell penetrating, the arginine residues might require closer clustering as opposed to the distributed placement in ArgiMYC and ArtiMYC.

Our finding highlight the unique potential of protein stapling for generating cell-penetrant protein drugs targeting intracellular complexes. While peptide staples have been used to stabilize α-helical structures, protect against proteolytic degradation and improve cell permeability, [31,32,37] applying stapling techniques to entire recombinant proteins, aiming at these properties, to date remains largely underexplored. Protein stapling has been reported, but mainly in proof of principle studies or mainly aiming at tertiary structure stabilization.[38–41] While there have been previous efforts toward the chemical synthesis of MYC analogs,[16,17,24,42–48] we developed a practical reCHEMbinant workflow, for directly modifying recombinant proteins. Our reCHEMbinant approach, which combines recombinant protein expression with targeted cysteine modification, enables rapid generation of stapled protein variants and may be applicable to other α-helical protein domains. Recent technology breakthroughs enable the de novo design of miniproteins binding to virtually any target.[49] For intracellular targets these approaches are of limited interest, but our reCHEMbinant strategy could unlock this potential.

Our cell proliferation experiments showed how HeloMYC's ability to inhibit the proliferation of cancer cells, without showing toxicity on MYC-independent HEK293 cells, indicating the absence of unspecific membrane toxicity, sometimes observed for cell penetrating peptides.[29] While our HeloMYC variants showed a clear superiority to the parent compound Omomyc, their current potency in inhibiting cell proliferation suggests that further optimization will be necessary to fully realize their clinical potential. Overall, our findings demonstrate that this approach can significantly enhance the therapeutic potential of protein drugs, expanding their applicability to previously undruggable intracellular targets.

4.4 Contribution to the chapter

In this chapter we performed a structure-based design study using the crystal structure of Omomyc. We regularly discussed results (daily to weekly basis) and I provided advice where possible. I helped regularly with protein expression and performed stability and folding assays for the compounds described within this chapter. The latter entails the development of the serum stability assay and characterization of all compounds using CD. The staple reaction protocol was based on a protocol developed by me for **Chapter 3.** I helped with revisions.

4.5 References

[1] B. Leader, Q. J. Baca, D. E. Golan, "Protein therapeutics: a summary and pharmacological classification" *Nature Reviews Drug Discovery 2007 7:1* **2008**, *7*, 21–39.

[2] K. Itakura, T. Hirose, R. Crea, A. D. Riggs, H. L. Heyneker, F. Bolivar, H. W. Boyer, "Expression in Escherichia coli of a Chemically Synthesized Gene for the Hormone Somatostatin" *Science (1979)* **1977**, *198*, 1056–1063.

[3] M. M. Goldenberg, "Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer" *Clin Ther* **1999**, *21*, 309–318.

[4] G. Kwok, T. C. C. Yau, J. W. Chiu, E. Tse, Y. L. Kwong, "Pembrolizumab (Keytruda)" *Hum Vaccin Immunother* **2016**, *12*, 2777–2789.

[5] H. F. Bunn, "Erythropoietin" *Cold Spring Harb Perspect Med* **2013**, *3*, a011619.

[6] S. Pestka, "The Interferons: 50 Years after Their Discovery, There Is Much More to Learn" *Journal of Biological Chemistry* **2007**, *282*, 20047–20051.

[7] M. Buyanova, D. Pei, **2022**, Elsevier Ltd preprint, DOI: 10.1016/j.tips.2021.11.008.

[8] F. Cozzolino, I. Iacobucci, V. Monaco, M. Monti, **2021**, American Chemical Society preprint, DOI: 10.1021/acs.jproteome.1c00074.

[9] H. D. Herce, D. Schumacher, A. F. L. Schneider, A. K. Ludwig, F. A. Mann, M. Fillies, M. A. Kasper, S. Reinke, E. Krause, H. Leonhardt, M. C. Cardoso, C. P. R. Hackenberger, "Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells" *Nat Chem* **2017**, *9*, 762–771.

[10] S. Mandal, G. Mann, G. Satish, A. Brik, "Enhanced Live-Cell Delivery of Synthetic Proteins Assisted by Cell-Penetrating Peptides Fused to DABCYL" *Angewandte Chemie - International Edition* **2021**, *60*, 7333–7343.

[11] E. Wang, A. Sorolla, P. T. Cunningham, H. M. Bogdawa, S. Beck, E. Golden, R. E. Dewhurst, L. Florez, M. N. Cruickshank, K. Hoffmann, R. M. Hopkins, J. Kim, A. J. Woo, P. M. Watt, P. Blancafort, "Tumor penetrating peptides inhibiting MYC as a potent targeted therapeutic strategy for triple-negative breast cancers" *Oncogene* **2019**, *38*, 140–150.

[12] K. A. Mix, J. E. Lomax, R. T. Raines, "Cytosolic Delivery of Proteins by Bioreversible Esterification" *J Am Chem Soc* **2017**, *139*, 14396–14398.

[13] J. V. Jun, Y. D. Petri, L. W. Erickson, R. T. Raines, "Modular Diazo Compound for the Bioreversible Late-Stage Modification of Proteins" *J Am Chem Soc* **2023**, *145*, 6615–6621.

[14] M. Wendt, R. Bellavita, A. Gerber, N. L. Efrém, T. van Ramshorst, N. M. Pearce, P. R. J. Davey, I. Everard, M. Vazquez-Chantada, E. Chiarparin, P. Grieco, S. Hennig, T. N. Grossmann, "Bicyclic β-Sheet Mimetics that Target the Transcriptional Coactivator β-Catenin and Inhibit Wnt Signaling" *Angewandte Chemie - International Edition* **2021**, *60*, 13937–13944.

[15] H. Adihou, R. Gopalakrishnan, T. Förster, S. M. Guéret, R. Gasper, S. Geschwindner, C. Carrillo García, H. Karatas, A. V Pobbati, M. Vazquez-Chantada, P. Davey, C. M. Wassvik, J. K. S. Pang, B. S. Soh, W. Hong, E. Chiarparin, D. Schade, A. T. Plowright, E. Valeur, M. Lemurell, T. N. Grossmann, H. Waldmann, "A protein tertiary structure mimetic modulator of the Hippo signalling pathway" *Nat Commun* **2020**, *11*, DOI 10.1038/s41467-020-19224-8.

[16] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, "Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains" *Nat Biotechnol* **2023**, *41*, DOI 10.1038/s41587-022-01504-x.

[17] B. D. Ellenbroek, J. P. Kahler, D. Arella, C. Lin, W. Jespers, E. A. Züger, M. Drukker, S. J. Pomplun, "Development of DuoMYC: a synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC" *Angewandte Chemie International Edition* **2024**, DOI 10.1002/anie.202416082.

[18] M. Pelay-Gimeno, A. Glas, O. Koch, T. N. Grossmann, "Structure-Based Design of Inhibitors of Protein-Protein Interactions: Mimicking Peptide Binding Epitopes" *Angewandte Chemie International Edition* **2015**, *54*, 8896–8927.

[19] L. Soucek, R. Jucker, L. Panacchia, R. Ricordy, F. Tatò, S. Nasi, "Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis" *Cancer Res* **2002**, *62*, 3507–3510.

[20] M. J. Demma, C. Mapelli, A. Sun, S. Bodea, B. Ruprecht, S. Javaid, D. Wiswell, E. Muise, S. Chen, J. Zelina, F. Orvieto, A. Santoprete, S. Altezza, F. Tucci, E. Escandon, B. Hall, K. Ray, A. Walji, J. O'Neil, "Omomyc Reveals New Mechanisms To Inhibit the MYC Oncogene" *Mol Cell Biol* **2019**, *39*, 1–27.

[21] D. Massó-Vallés, L. Soucek, "Blocking Myc to Treat Cancer: Reflecting on Two Decades of Omomyc" *Cells* **2020**, *9*, 883.

[22] L. A. Jung, A. Gebhardt, W. Koelmel, C. P. Ade, S. Walz, J. Kuper, B. Von Eyss, S. Letschert, C. Redel, L. D'Artista, A. Biankin, L. Zender, M. Sauer, E. Wolf, G. Evan, C. Kisker, M. Eilers, "OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors" *Oncogene* **2017**, *36*, 1911–1924.

[23] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription" *ACS Cent Sci* **2021**, *7*, 1408–1418.

[24] M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, "Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents" *J Am Chem Soc* **2021**, *143*, 11788–11798.

[25] M. E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. S. Del Pozo, C. Fiore, L. Foradada, V. C. Cano, M. Sánchez-Hervás, M. Guenther, E. R. Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. M. Alonso, J. R. Whitfield, P. Lavigne, L. Soucek, "Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy" *Sci Transl Med* **2019**, *11*, 1–14.

[26] N. Schmidt, A. Mishra, G. H. Lai, G. C. L. Wong, **2010**, DOI: 10.1016/j.febslet.2009.11.046.

[27] G. Tünnemann, G. Ter-Avetisyan, R. M. Martin, M. Stöckl, A. Herrmann, M. C. Cardoso, "Live-cell analysis of cell penetration ability and toxicity of oligo-arginines" *Journal of Peptide Science* **2008**, *14*, 469–476.

[28] E. M. López-Vidal, C. K. Schissel, S. Mohapatra, K. Bellovoda, C. L. Wu, J. A. Wood, A. B. Malmberg, A. Loas, R. Gómez-Bombarelli, B. L. Pentelute, "Deep Learning Enables Discovery of a Short Nuclear Targeting Peptide for Efficient Delivery of Antisense Oligomers" *JACS Au* **2021**, *1*, 2009–2020.

[29] C. Schissel, S. Mohapatra, J. Wolfe, C. Fadzen, K. Bellovoda, C.-L. Wu, J. Wood, A. Malmberg, A. Loas, R. Gómez-Bombarelli, B. Pentelute, "Interpretable Deep Learning for De Novo Design of Cell-Penetrating Abiotic Polymers" **2020**, DOI 10.1101/2020.04.10.036566.

[30] Q. Chu, R. E. Moellering, G. J. Hilinski, Y. W. Kim, T. N. Grossmann, J. T. H. Yeh, G. L. Verdine, "Towards understanding cell penetration by stapled peptides" *Medchemcomm* **2015**, *6*, 111–119.

[31] A. Chandramohan, H. Josien, T. Y. Yuen, R. Duggal, D. Spiegelberg, L. Yan, Y. C. A. Juang, L. Ge, P. G. Aronica, H. Y. K. Kaan, Y. H. Lim, A. Peier, B. Sherborne, J. Hochman, S. Lin, K. Biswas, M. Nestor, C. S. Verma, D. P. Lane, T. K. Sawyer, R. Garbaccio, B. Henry, S. Kannan, C. J. Brown, C. W. Johannes, A. W. Partridge, "Design-rules for stapled peptides with in vivo activity and their application to Mdm2/X antagonists" *Nat Commun* **2024**, *15*, DOI 10.1038/s41467-023-43346-4.

[32] Y. W. Kim, T. N. Grossmann, G. L. Verdine, "Synthesis of all-hydrocarbon stapled ±-helical peptides by ring-closing olefin metathesis" *Nat Protoc* **2011**, *6*, 761–771.

[33] P. Timmerman, W. C. Puijk, R. H. Meloen, "Functional reconstruction and synthetic mimicry of a conformational epitope using CLIPS™ technology" *Journal of Molecular Recognition* **2007**, *20*, 283–299.

[34] S. S. Kale, C. Villequey, X. D. Kong, A. Zorzi, K. Deyle, C. Heinis, "Cyclization of peptides with two chemical bridges affords large scaffold diversities" *Nat Chem* **2018**, *10*, 715–723.

[35] L. Boike, A. G. Cioffi, F. C. Majewski, J. Co, N. J. Henning, M. D. Jones, G. Liu, J.

M. McKenna, J. A. Tallarico, M. Schirle, D. K. Nomura, "Discovery of a Functional Covalent Ligand Targeting an Intrinsically Disordered Cysteine within MYC" *Cell Chem Biol* **2021**, *28*, 4-13.e17.

[36] J. Liu, T. Gaj, M. C. Wallen, C. F. Barbas, "Improved cell-penetrating zinc-finger nuclease proteins for precision genome engineering" *Mol Ther Nucleic Acids* **2015**, *4*, e232.

[37] D. P. Fairlie, A. Dantas de Araujo, "Review stapling peptides using cysteine crosslinking" *Biopolymers* **2016**, *106*, 843–852.

[38] X. H. Chen, Z. Xiang, Y. S. Hu, V. K. Lacey, H. Cang, L. Wang, "Genetically encoding an electrophilic amino acid for protein stapling and covalent binding to native receptors" *ACS Chem Biol* **2014**, *9*, 1956–1961.

[39] S. P. Brown, A. B. Smith, "Peptide/protein stapling and unstapling: Introduction of s-tetrazine, photochemical release, and regeneration of the peptide/protein" *J Am Chem Soc* **2015**, *137*, 4034–4037.

[40] S. Neubacher, J. M. Saya, A. Amore, T. N. Grossmann, "In Situ Cyclization of Proteins (INCYPRO): Cross-Link Derivatization Modulates Protein Stability" *Journal of Organic Chemistry* **2020**, *85*, 1476–1483.

[41] G. H. Hutchins, S. Kiehstaller, P. Poc, A. H. Lewis, J. Oh, R. Sadighi, N. M. Pearce, M. Ibrahim, I. Drienovská, A. M. Rijs, S. Neubacher, S. Hennig, T. N. Grossmann, "Covalent bicyclization of protein complexes yields durable quaternary structures" *Chem* **2024**, *10*, 615–627.

[42] L. E. Canne, A. R. Ferré-D'Amaré, S. K. Burley, S. B. H. Kent, "Total Chemical Synthesis of a Unique Transcription Factor-Related Protein: cMyc—Max" *J Am Chem Soc* **1995**, *117*, 2998–3007.

[43] R. Calo-Lapido, C. Penas, A. Jiménez-Balsa, M. E. Vázquez, J. L. Mascareñas, "A chemical approach for the synthesis of the DNA-binding domain of the oncoprotein MYC" *Org Biomol Chem* **2019**, *17*, 6748–6752.

[44] B. D. Ellenbroek, J. P. Kahler, S. R. Evers, S. J. Pomplun, "Synthetic Peptides: Promising Modalities for the Targeting of Disease-Related Nucleic Acids" *Angewandte Chemie International Edition* **2024**, *63*

[45] O. Harel, M. Jbara, "Chemical Synthesis of Bioactive Proteins" *Angewandte Chemie - International Edition* **2023**, *62*, DOI 10.1002/anie.202217716.

[46] X. Lin, O. Harel, M. Jbara, "Chemical Engineering of Artificial Transcription Factors by Orthogonal Palladium(II)-Mediated S-Arylation Reactions" *Angewandte Chemie - International Edition* **2023**, *202317511*, DOI 10.1002/anie.202317511.

[47] R. V. Nithun, Y. M. Yao, X. Lin, S. Habiballah, A. Afek, M. Jbara, "Deciphering the Role of the Ser-Phosphorylation Pattern on the DNA-Binding Activity of Max Transcription Factor Using Chemical Protein Synthesis" *Angewandte Chemie - International Edition* **2023**, *62*, DOI 10.1002/anie.202310913.

[48] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription" *ACS Cent Sci* **2021**, *7*, 1408–1418.

[49] M. Pacesa, L. Nickel, C. Schellhaas, J. Schmidt, E. Pyatova, L. Kissling, P. Barendse, J. Choudhury, S. Kapoor, A. Alcaraz-Serna, Y. Cho, K. H. Ghamary, L. Vinué, B. J. Yachnin, A. M. Wollacott, S. Buckley, A. H. Westphal, S. Lindhoud, S. Georgeon, C. A. Goverde, G. N. Hatzopoulos, P. Gönczy, Y. D. Muller, G. Schwank, D. C. Swarts, A. J. Vecchio, B. L. Schneider, S. Ovchinnikov, B. E. Correia, "One-shot design of functional protein binders with BindCraft" *Nature 2025* **2025**, 1–10.

4.6 Experimental

All reactions were carried out using commercially available reagents, unless noted otherwise. Reagents and solvents were purchased from Sigma-Aldrich (Merck), Fisher Scientific, or VWR chemicals.

Plasmid pET-30a with 6-his-tagged Omomyc was donated by the group of prof. dr. Eilers (University of Würzburg, Germany) and sequenced before being used. Primers for site-directed DNA mutagenesis were purchased from IDT (Leuven, Belgium). Plasmids with ArgiMYC, ArtiMYC and NucleoMYCs were obtained from genscript.

4.6.1 LC-MS

LC-MS chromatograms and associated mass spectra were acquired using a Shimazu LCMS-2020 system (Method A) or, for high resolution mass spectrometry data, a Sciex X500b QTOF ESI-QToF mass spectrometer coupled to a Shimadzu Nexera UHPLC LC40DX3 (Method B). Mobile phases used for LC-MS analysis are solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in acetonitrile).

The following LCMS methods were used:

Method A

Column: Kinetex® 2.6µm XB-C18 100 Å LC Column (50 x 3 mm) UV detector: 214 nm.

LC Method: 0% solvent B over 1 min, followed by a linear gradient 0% to 70% solvent B over 10 min, followed by 70% solvent B over 1.5 min, followed by 70% to 0% solvent B over 4.5, flowrate 0.55 mL/min

Method B (high resolution)

Column: Phenomenex Synergi™ 4 µm Fusion-RP 80 Å LC Column (50 x 2 mm)

LC Method: 0% solvent B over 1 min, followed by a linear gradient 0% to 60% solvent B over 3.5 min, followed by a linear gradient 60% to 95% solvent B over 0.1 min, followed by 95% solvent B over 0.4 min, followed by a linear gradient 95% to 0% solvent B over 0.5 min, followed by 0% solvent B over 1.5 min, flowrate 0.5 mL/min.

MS parameters: General parameters: Method duration: 5 min; Total scan time: 0.276 sec; Estimated cycles: 1086; Intact protein mode: False; Decrease detector voltage: False; Large protein (>70 kDa): False; Ion Source: Source name: TurbolonSpray; Curtain gas: 35 psi; Ion source gas 1: 60 psi; Ion source gas 2: 60 psi; Temperature: 500

˚C; Experiment: Scan type: TOF MS; Polarity: Positive; Spray voltage: 5500 V; CAD gas: 7; Time bins to sum: 4; Channel 1-4: True; TOF start mass: 350 Da; TOF stop mass 1500 Da; Accumulation time: 0.25; Declustering potential: 80V; Declustering potential spread: 0 V; Collision energy: 10V; Collision energy spread: 0 V; Override Qjet RF value: False

Method C (high resolution)

Column: Phenomenex Synergi™ 4 µm Fusion-RP 80 Å LC Column (50 x 2 mm) or Aeris™ 3.6 µm Widepore XB-C18.

LC Method: 0% solvent B over 1 min, followed by a linear gradient 0% to 90% solvent B over 6 min, followed by a 90% solvent B over 2 min, followed by a linear gradient of 90% to 0% solvent B over 0.5 min, followed by 0% solvent B over 1.5 min, flowrate 0.5 mL/min.

4.6.2 Synthesis

4.6.2.1 Stapling reactions

4.6.2.1.1 Stapling with i, i+7 staple (4,4′-Bis(bromomethyl)biphenyl)

The respective protein (1 eq., **4** or **5**) was dissolved in water (2 mM, e.g. 30 mg of **5** in 1.08mL) and then diluted to 100 µM into stapling buffer (15.12 mL for **5**, NH_4HCO_3, 100 mM, pH= 8). 4,4′-Bis(bromomethyl)biphenyl was dissolved at 4x the final concentration in MeCN (500 µM) and then added (1.25 eq., 5.4 mL for HeloMYC-1421) to the protein in stapling buffer. Final concentrations used were 100 µM (1 eq.) protein and 125 µM (1.25 eq.) staple. Final MeCN content was 25%.

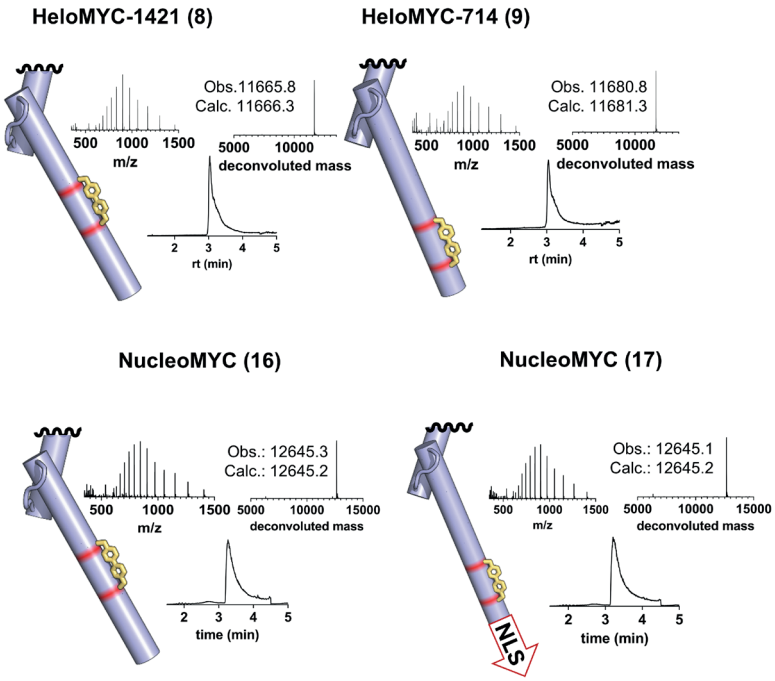
When the stapling reaction was completed as seen by LC-MS the product was purified using a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 10 g column on a gradient of water in MeCN (0-100%,with a flat gradient between 15% and 79%), both containing 0.1% TFA. The product containing fractions were lyophilized yielding the proteins as TFA salts.

Yield HeloMYC-1421 (**8**): 20.7 mg (69%)

Yield HeloMYC-714 (**9**): 1.94 mg (67%)

Yield NucleoMYC (**16**): 3.02 mg (60%)

Yield NucleoMYC (**17**): 3.71 mg (46%)



4

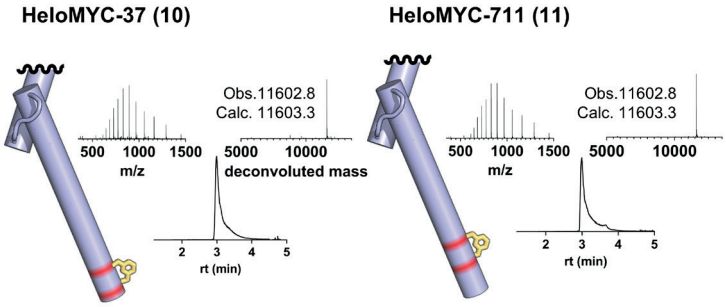
4.6.2.1.2 Stapling with i, i+4 staple (α,α-Dibromo-m-xylene)

The respective protein (1 eq., 6 or 7) was dissolved in water (2 mM, e.g. 1.96 mg of 7 in 70.53 µL) and then diluted into stapling buffer (973 µL for 7, NH_4HCO_3, 100 mM, pH= 8). TCEP was dissolved at 100x the final concentration in water (20 mM) and added to the stapling reaction (2 eq., 14.1 µL). α,α-Dibromo-m-xylene was dissolved at 4x the final concentration in MeCN (500 µM) and then added (1.25 eq., 353 µL for HeloMYC-711) to the protein in stapling buffer. Final concentrations used were 100 µM (1 eq.) protein, 125 µM (1.25 eq.) staple and 200 µM (2 eq.) TCEP. Final MeCN content was 25%.

When the stapling reaction was completed as seen by LC-MS the product was purified using a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 10 g column on a gradient of water in MeCN (0-100%,with a flat gradient between 15% and 79%), both containing 0.1% TFA. The product containing fractions were lyophilized yielding the proteins as TFA salts.

Yield HeloMYC-37 (**10**): 0.27 mg (38%)

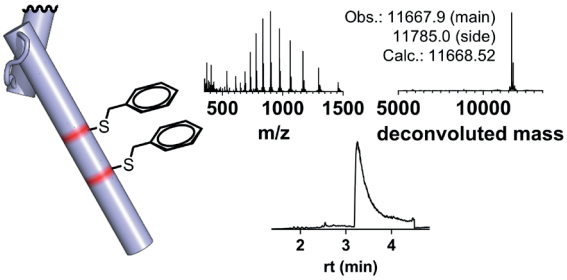
Yield HeloMYC-711 (**11**): 0.32 mg (26%)



4.6.2.1.3 Cys-capping of 4 with benzyl bromide

Protein **4** (2.8 mg, 202 nmol, 1eq.) was dissolved at 20x the final concentration in water (101 µL, 2 mM) and then diluted into stapling buffer (NH_4HCO_3, 100 mM, pH= 8). Benzyl bromide was dissolved at 4x the final concentration in MeCN (1 mM) and then added to the protein in stapling buffer (505 µL, 2.5 eq.). Final concentrations used were 100 µM (1 eq.) for the protein and 250 µM (2.5 eq.) for benzyl bromide. Final MeCN content was 25%.

When the stapling reaction was completed as seen by LC-MS the product was purified using a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 10g column on a gradient of water in ACN, both containing 0.1% TFA. The product containing fractions were lyophilized yielding BenzoMYC (**S1**) (1.06 mg, 75 nmol, 37%) as TFA salt. In addition to the main peak which corresponds to the desired product, a second deconvoluted mass corresponding to an additional benzyl substitution is detected in the product.

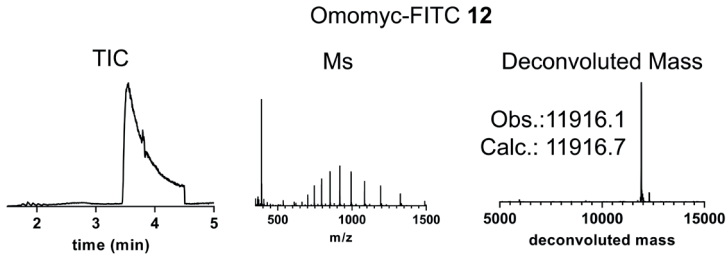


4.6.2.2 FITC labelling

4.6.2.2.1 Labelling Omomyc

Omomyc (4.19 mg, 300 nmol, 1 eq.) was dissolved in carbonate buffer (0.1 M, pH 9.1,

4.19 mL) to a concentration of 1 mg/mL and cooled to 0 °C. Subsequently, a solution of FITC in dmso (1.7 mM, 159 µg, 240 µL, 408 nmol, 1.36 eq.) was added slowly over the course of approximately 4 h. The crude product was then purified by reverse phase column chromatography using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 10 g applying a gradient of 0-100% MeCN in water (0-100% with a flat gradient between 20% and 48%) yielding FITC-labelled Omomyc-F (12) (0.93 mg, 65 nmol, 22%) as TFA salt after lyophilization.

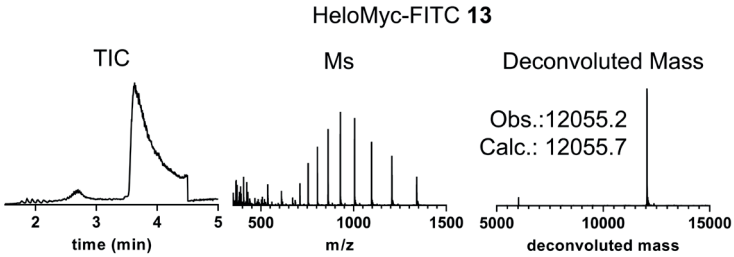


LC-MS spectrum with total ion count, extracted ion count and deconvoluted mass of FITC-labelled OmoMyc (**12**).

4

4.6.2.2.2 Labelling HeloMYC-1421

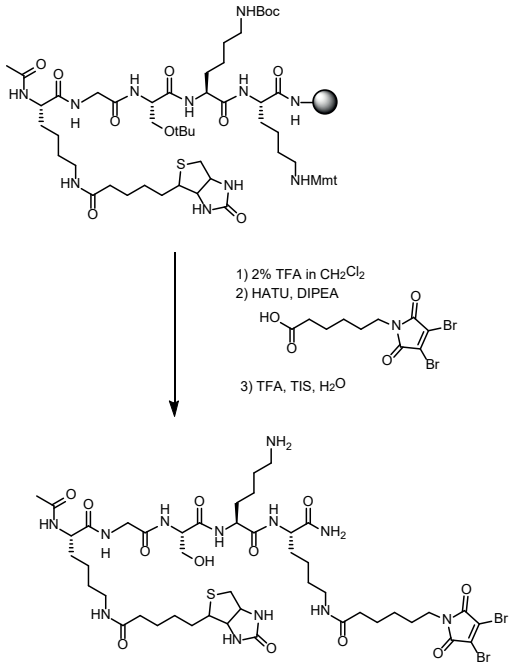
HeloMYC-1421 (1.35 mg, 96 nmol, 1 eq.) was dissolved in carbonate buffer (0.1 M pH 9.1, 1.35 mL) to a concentration of 1mg/mL and cooled to 0°C. Subsequently, a solution of FITC in dmso (1.56 mM, 67.7 µL, 106 nmol, 1.1 eq.) was added over the course of 4.5 h. The crude product was then purified by reverse phase column chromatography using a Biotage® Selekt Flash Purification System with a Biotage® Sfär C18 D - Duo 100 Å 30 µm column 10 g applying a gradient of 0-100% MeCN in water (0-100% with a flat gradient between 20% and 48%), yielding FITC-labelled HeloMYC-1421-F (**13**) (0.59 mg, 41 nmol, 43%) after lyophilization.

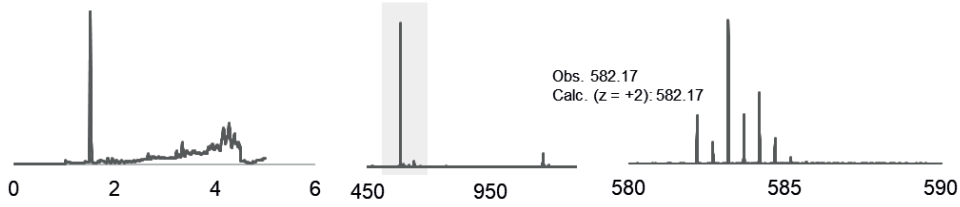


LC-MS spectrum with total ion count, extracted ion count and deconvoluted mass of FITC-labelled HeloMyc-1421 (**13**).

4.6.2.2.3 Synthesis of probe 18

Fmoc-Rink-amide-Protide resin (100 mg, 0.6 mmol/g loading) was incubated with piperidine (20% in DMF, 2 mL) for 10 minutes and then washed with DMF (5 x 3 mL). A HATU solution in DMF (0.4 M) was prepared and each amino acid building block was dissolved in this solution (appropriate amount to result in final amino acid concentration of 0.4 M, 500 µL). For each coupling cycle DIPEA (100 µL) was added to the amino acid HATU mix (500 µL) for preactivation. After 30 seconds this solution was added to the resin, stirred and then incubated for 15 minutes. The coupling mixture was drained under vacuum and the resin washed with DMF (3 x 3 mL), before Fmoc removal with piperidine (20% in DMF, 2 mL) for 10 minutes followed by washes with DMF (5 x 3 mL). As last linear building block we incorporate Fmoc-Lys(Biotin)-OH, followed by Fmoc removal. The peptide was then acetylated via treatment with acetic anhydride (10% in DMF with 10% DIPEA, 2 mL, 10 minutes). After washes with DMF (5 x 3 mL) and CH_2Cl_2 (3 x 3 mL), the resin was incubated with TFA (1% in CH_2Cl_2, 1 mL, 5 x 2 min) to remove the Mmt protecting group. Dibromomaleimide-hexanoic acid (74 mg) was dissolved in DMF containing 0.4 M HATU (500 µL), activated with DIPEA (100 µL) and then added to the peptidyl resin and incubated for 30 minutes. The resin was washed with DMF (5 x 3 mL) and CH_2Cl_2 (3 x 3 mL) before cleavage and global deprotection with TFA + 5% water, 60 minutes. The probe was precipitated with ice cold ether analyzed by LCMS and used without further purification.

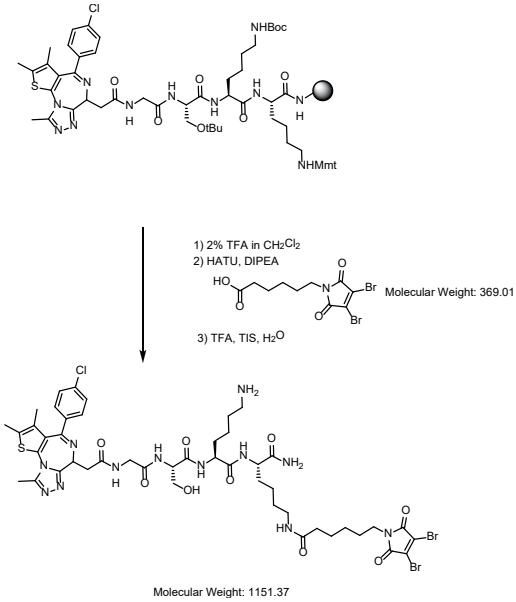


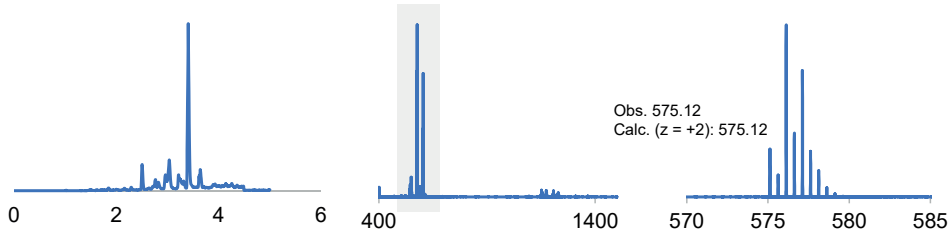


LCMS trace and spectrum of probe 18.

4.6.2.2.4 Synthesis of probe 19

The synthesis follows the same procedure as detailed above. As last linear building block JQ1-acid was incorporated, under the same coupling conditions as detailed for all amino acid building blocks.





LCMS trace and spectrum of probe 19.

4.6.2.2.5 Synthesis of protein conjugates 20 and 21

The protein (2 mg, 140 nmol) was dissolved in stapling buffer (500 µL, NH_4HCO_3, 100 mM, pH= 8), and then diluted with MeCN (250 µL). TCEP was dissolved in water (20 mM) and added to the stapling reaction (2 eq., 14.0 µL). Probe (18 or 19) was dissolved in MeCN (10 mM) and then added (2 eq., 28 µL) to the protein in stapling buffer. The reaction was incubated for 60 min. When the stapling reaction was completed as determined by LC-MS the product was purified using a size exclusion (cut off 7 kDa) zeba spin column.

4.6.3 Protein expression and characterization

4.6.3.1 Site-directed mutagenesis

For site-directed mutagenesis the QuikChange II site-directed mutagenesis kit (Agilent Technologies) was used and primers were designed using the QuikChange® Primer Design Program provided by Agilent Technologies. Mutagenesis was performed according to the manufacturer's protocol In brief, PCR reactions were prepared using pET-30a with Omomyc as template (0.5 µL, ~25 ng), forward and reverse primers (0.6 µL each, 10 µM), MQ water (18.8 µL) as well as the contents provided by the kit: NTP mix (1 µL), 10x reaction buffer (2.5 µL) and *PfuUltra* high fidelity DNA polymerase. In some cases, when the formation of primer dimers was seen, primer concentration was reduced and 1 µL of dmso was added. PCR was performed for 22 cycles (95 °C, 30s; 55 °C, 1 minute; 68 °C, 10 minutes) followed by *Dpn I* restriction for 2 hours at 37 °C (addition of 0.5 µL at 10 U/µL)

Mutated plasmids were then incorporated into XL-1 blue competent cells. Cells were thawed on ice and 8 µL of the *Dpn I*-treated DNA were added to 50 µL of cells. The mixture was incubated on ice for 30 minutes, followed by a 45 second heat pulse at 42 °C and incubation on ice for another 2 minutes. Subsequently, 0.5 mL SOC-medium were added and the cells shaken at 250 rpm, 37 °C for 1 hour. For selection, 100 µL of the mixture were plated on one half of an LB agar plate containing gentamicin and kanamycin. The rest of the cells was spun down, the supernatant removed except for 100 µL, the cells resuspended in these 100 µL and plated on the other half of the LB agar plate. The plate was incubated at 37 °C overnight.

About 5-10 colonies were picked from selection and grown overnight in 5 mL LB media supplemented with gentamicin and kanamycin and plasmid DNA was isolated using the QIAprep Spin Miniprep kit (Qiagen, Venlo, NL) according to the manufacturer's protocol. The isolated plasmid DNA was sequenced using a sanger sequencing service and analyzed using Benchling.

4.6.3.2 Primers used for site-directed mutagenesis

Mutation	Primer sequence 5'-3' (forward, mutation in bold)
C91A	caaacttgaacagctacggaactct**gct**gcgtaaggactc
T3C	gatatcggatccatggcg**tg**cgaggagaatgtcaagag
V7C	ccatggcgaccgaggagaat**tg**caagaggcgaacac
V7C T11C	ggcgaccgaggagaat**tg**caagaggcga**tgc**cacaacgtcttggagc
V14C	gaggcgaacacacaac**tg**cttggagcgccagagg
N21C	gagcgccagaggagg**tg**cgagctaaaacggag

4.6.3.3 Primers used for sequencing

Name	Sequence 5'-3'
T7 terbis – sequencing reverse	aacccctcaagacccg
pET upstream – sequencing forward	gatgcgtccggcgtagag

4.6.3.4 Expression of different Omomyc variants

4

Plasmids used were either obtained from site-directed mutagenesis of pET-30a with the omomyc gene sequence (for stapled variants termed HeloMYC) or genes as well as insertion of them by express cloning were ordered from GenScript and the plasmids used as delivered (for artificial coiled-coil and mutations in the coiled-coil termed ArtiMYC and ArgiMyc as well as NLS-HeloMYC fusions, see annex for plasmid sequences).

Plasmids were transformed into competent ArcticXpress DE3 RIL cells by heat shock according to the manufacturer's protocol. Briefly, 2 µL of 10% β mercaptoethanol were mixed with 100 µL of competent cell suspension thawed on ice and incubated for 10 minutes on ice. Next, 25 ng of plasmid DNA were added and the cells incubated for another 30 minutes on ice. The cells were then heat-shocked in a water bath for 20 seconds at 42 °C and subsequently incubated on ice for 2 minutes followed by the addition of 0.9 mL SOC media and incubation at 37 °C and 250 rpm for 1 h. Cells were then pelleted by centrifugation, 0.9 mL of the supernatant decanted and the pellet resuspended in the remaining 100 µL of media. Cells were plated for selection on LB agar with kanamycin and gentamicin and incubated at 37 °C overnight.

Single colonies were picked from the plate and cultured overnight at 37 °C and 180 rpm in 100 mL of LB media containing kanamycin and gentamycin. Next, 3x2 L of LB media containing kanamycin and gentamycin in 5 L Erlenmeyer flasks were inoculated with 25 mL of preculture and grown at 37 °C and 180 rpm until an OD of 0.8 was reached. The temperature was then set to 14 °C and protein expression

induced by addition of IPTG to a final concentration of 100 µM. Protein expression was conducted overnight for 18 h after which the cells were harvested by centrifugation (5,000g, 4 °C, 12 min.), the pellet resuspended in lysis buffer (20 mM Tris-HCl pH 8, 0.5 M NaCl, 10 mM imidazole, 3 mM $MgCl_2$, freshly added 1 cOmplete EDTA-free protease inhibitor cocktail tablet per 50 mL of buffer and 0.05-0.1% DNase) and the cells lysed by pressure lysis. Cell debris was removed by ultracentrifugation (35,000 rpm, 4℃, 45 min.)

4.6.3.5 Purification of His-tagged proteins

After ultracentrifugation, the supernatant was purified using an ÄKTA start protein purification system equipped with a 5 mL HisTrap HP His-tag protein purification column (Cytiva). After washing out unbound protein with wash buffer (20 mM Tris-HCl pH 8, 0.5 M NaCl, 10 mM imidazole) the protein was eluted using a gradient of 10 mM to 500 mM imidazole in the same buffer over 40-50 column volumes of buffer. Fractions with protein were analyzed for protein content and identity by LC-MS.

4.6.3.6 General protocol for buffer exchange

The combined fractions obtained from Ni-column purification were incubated with 5 mM TCEP for 1h to break any possible formed disulfide bonds. The buffer was then exchanged by subjecting the protein to column chromatography on a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 25g or 50g column and a stepwise gradient of 0 % MeCN in water followed by 50 to 100% MeCN in water). The combined fractions containing the protein were lyophilized, yielding the His-tagged or final protein as TFA salt.

4.6.3.7 Tag cleavage with enterokinase and subsequent purification

Lyophilized protein was dissolved at 2 mg/mL in EK cleavage buffer (200 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 20 mM $CaCl_2$) and after addition of 10 u/mL enterokinase the protein was incubated overnight. The protein was then purified using an ÄKTA start protein purification system equipped with a 5 mL HisTrap HP His tag protein purification column (Cytiva) applying a gradient of 10 mM to 500 mM imidazole in elution buffer (20 mM Tris-HCl pH 8, 0.5 M NaCl, 10-500 mM imidazole). The fractions were analyzed by LC-MS and the buffer of the combined protein containing fractions was exchanged as described above, yielding the final proteins as TFA salts after lyophilization.

4.6.3.8 Electromobility shift assay (EMSA)

For EMSAs, proteins were serial diluted with water to a final volume of 10 µL. Subsequently, 5 µL of 4x EMSA buffer (final buffer concentration: 20 mM HEPES

pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL of BSA, 1 mM DTT and 0.05% NP-40) followed by 5 µL of 4x FAM-labelled DNA construct (IRD700-ACC CCA CCA CGT GGT GCC T, final concentration 4 nM) were added.

The samples were incubated for 30 minutes at room temperature, placed on ice and incubated for another 15 minutes. Then 15 µL of the samples were loaded onto a 10% native acrylamide TBE gel which was pre run before for 1 h at 75 V at 4 °C in 0.5x TBE. Samples were run for 20 minutes at 120 V followed by 40 minutes at 100 V at 4 °C in 0.5x TBE and subsequently scanned on a Bio-Rad ChemiDoc MP machine. Bound protein signal was quantified using ImageJ and KD values were obtained using the inhibitor concentration vs. response variable slope model of nonlinear regression in GraphPad Prism 9.0.0.and the IC50 value reported as KD.

4.6.3.9 Circular Dichroism Spectroscopy

For CD measurements 1 mM protein stocks were diluted to 20 µM in Dulbecco's phosphate buffered saline (DPBS) to a final volume of 200 µL. For DNA containing samples, a 1 mM DNA stock in water was heated for 5 min to 95 °C, let cool down to RT and an equimolar amount of DNA was added to the protein samples dedicated to be measured with DNA. Additionally, a blank measurement with or without DNA was measured. Circular Dichroism (CD) samples were measured at 37 °C using a Jasco J-8151 CD spectrometer with a 1 mm path length quartz cuvette. The following parameters were used for a full wavelength scan: wavelength = 260-200 nm; Data pitch = 1 nm; scanning mode = continuous; scanning speed = 100 nm/min; response = 1, BW = 1, accumulation = 5. For CD melting curves the same samples were cooled down to 5 °C and slowly heated to 90 °C while measuring at 222 nm, with a heating speed of 2 °C/min.

For CD analysis, the blank measurement was subtracted from each spectrum and the mean residue molar ellipticity (θ, deg cm^2 dmol) was calculated using the following equation[2]:

$$\left[\emptyset\right] = \frac{100*\emptyset_{obs}}{c*n*l}$$

With θ_{obs} in mdeg, concentration (c) in mM, peptide bonds (n) and path length of cuvette in cm.

4.6.3.10 Serum stability assay

For serum stability assays 1 mM protein stocks were diluted to a final concentration of 60 µM in 10% human serum in DPBS. The mixture was vortexed immediately after protein addition and 5 µL aliquots were mixed with 5 µL of 20% TFA in water (t = 0) to quench the human serum, resulting in protein precipitating. Subsequently, the protein

serum solution was incubated at 37 °C using a BIO-RAD T100™ Thermal Cycler with the lid set to 95 °C. At indicated timepoints 5 µL aliquots were mixed with 5 µL of 20% TFA in water. The resulting pellet was then diluted 5.5x with additional DPBS to redissolve. This solution was analyzed according to the high resolution LCMS protocol (Method C). As internal standard (IS), the extracted-ion chromatogram (XIC) from m/z = 1233.52 was used belonging to human serum albumin. After measurement, the XIC obtained from the highest intensity peak belonging to each protein and IS were extracted from the total-ion chromatogram (TIC). Using Graphpad Prism 9, the area under the curve (AUC) from each timepoint was calculated, normalized against the AUC of the IS followed by normalization against t = 0 and plotted. Next, a nonlinear regression – One phase decay analysis was performed to obtain $t_{½}$ with a plateau constant equal to 0 and Y_0 set to 1. The assay was performed in triplicates. For Omomyc one set of outliers in the measurement was excluded.

4.6.3.11 Cell culture and cell assays

Cells were cultured at 37°C in 5% CO_2 atmosphere. Cell lines were cultured in ATCC recommended media and split twice a week before confluency was reached.

4.6.3.12 MYC reporter gene assay

The reporter gene assay was performed using Cignal reporter assay (CCS-012L, Qiagen) according to the manufacturer's protocol. In brief, HEK293T cells were harvested and resuspended in OptiMEM media containing 5% FBS and 1% non-essential amino acids (NEAA) as well as penicillin/streptavidin. 40,000 cells were seeded per well in a 96-well plate and 50 µL transfection cocktail of either cignal reporter or positive or negative control reporter along with attractene transfection reagent in OptiMEM without additives was added and the cells incubated overnight. Next, media was changed to assay media (OptiMEM, 0.5% FBS, 1% NEAA, penicillin/streptavidin) and the cells were incubated for 8 hours after which the media was replaced by 75 µL assay media containing the different proteins at the required concentration and the cells were incubated with the proteins for 24 hours.

Luciferase assay was then performed using a luciferase assay kit (E2940, Promega). Cells were lysed by addition of 75 µL of DualGlo luciferase assay reagent and incubated for 15 minutes after which Luciferase luminescence was measured on a Perkin Elmer EnVision 2104 Multilabel Reader. Subsequently, 75 µL of DualGlo Stop & Glo reagent were added and the renilla luciferase luminescence measured after 15 minutes of incubation time.

Signal was normalized against cell number by calculating the ratio of firefly and renilla luminescence and eventually these signals were normalized against the untreated control. All experiments were done in technical triplicates. Data was analyzed using

GraphPad Prism 9.0.0. applying the model of inhibitor concentration vs. response with variable slope model for nonlinear regression.

4.6.3.13 MYC reporter assay control for direct luciferase inhibition

For the control experiment testing direct luciferase inhibition by proteins, the cells were incubated for 24 hours in assay media and the media replaced by assay media containing the proteins directly before performing the luciferase assay as described above.

4.6.3.14 Cell proliferation assay

Cells were seeded in 100 µL of their respective media at 1500 cells/well (Figure 5.5d) or 2000 cells/well (Figure 5e) in a white opaque 96-well plate and let attach overnight. Media was then changed to 100 µL of media with the required protein at the appropriate concentration, the plate was covered with a membrane to prevent media evaporation and the cells were incubated with the proteins for 72 hours. Cell proliferation was then assessed using CellTiter-Glo (Promega) reading luminescence on a Perkin Elmer EnVision 2104 Multilabel Reader. All experiments were done in technical triplicates. Data was analyzed using GraphPad Prism 9.0.0.

4.6.3.15 Live cell microscopy with FITC-labelled HeloMYC and Omomyc

Hela cells were harvested and seeded at 12’000 cells/well in a 96-well plate and left to adhere overnight (4.5h in the case of 24h compound treatment). Cells were then treated with 5µM Omomyc-FITC (12) or HeloMYC-1421-FITC (13) for the desired time in full media. Subsequently, the media was aspirated cells were stained for 10 minutes with 1µg/mL Hoechst in DPBS, followed by three washes with full media. Live cells were imaged on a Nikon Ti2 microscope equipped with a Plan Apo VC 20x DIC N2 air objective using a 405nm laser for Hoechst and a 488nm laser for fluorescein excitation.

4.6.3.16 Quantification of signal in the green channel

For signal quantification the Hoechst channel was used and ROIs were determined by thresholding from 100-255. The ROIs were then dilated by 5 units to also cover the area surrounding the nucleus and a mask was created. The mask was applied to the FITC channel and the mean fluorescence measured per image. For each condition the mean of the mean fluorescence of the no treatment control was subtracted from the individual mean fluorescence and the values normalized to the desired condition.

4.6.3.17 RNA sequencing and GSEA

In a 12-well plate 50,000 HeLa cells/well were seeded. The next day the cells were treated with either 10 µM HeloMYC-1421 or vehicle in MEM media supplemented with 10% FBS, 1% Pen/Strep and Glutamax and incubated for 72 hours. Total RNA was isolated using Qiagen RNeasy Plus mini kit and dissolved in RNAse free water. Samples were shipped to and sequencing as well as differential gene expression analysis was performed by Novogene GmbH (Planegg, Germany). Gene set enrichment analysis (GSEA) was performed with the GSEA desktop application (Broad Institute) version 4.4.0 performing 10,000 permutations.

4.6.4 Protein sequences

Enterokinase cleavage sites shown with |

Omomyc MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSCA

Omomyc T3C V7C C91A HHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMA**C**EEN**C**KRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNS**A**A

Omomyc V7C T11C C91A MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMATEEN**C**KRR**C**HNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNS**A**A

Omomyc V7C V14C C91A MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMATEEN**C**KRRTHN**C**LERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNS**A**A

Omomyc V14C N21C C91A MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMATEENVKRRTHN**C**LERQRR**C**ELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNS**A**A

Omomyc Q64R D71R Q86R C91A - ArgiMyc MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQAET**R**KLISEI**R**LLRKQNEQLKHKLE**R**LRNS**A**A

ArtiMyc MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSV**KREIAALKREIAALKREIAALKRE**

NucleoMYC 16 MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIG-SMATEENVKRRTHN**C**LERQRR**C**ELKRSFFALRDQIPELENNEKAPKVVILKKATAYILSVQA-ETQKLISEIDLLRKQNEQLKHKLEQLRNS**AA**P**AAKRVKLD**

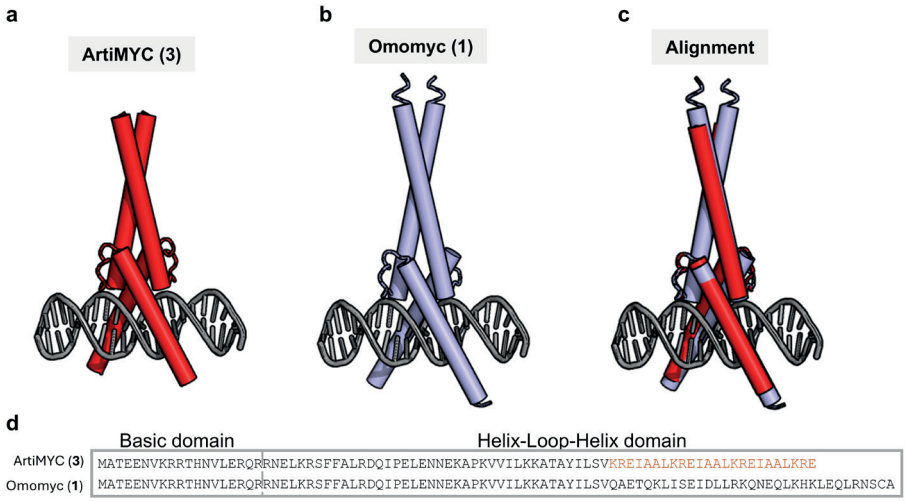
NucleoMYC 17 MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK|AMADIG-S**PAAKRVKLD**MATEENVKRRTHN**C**LERQRR**C**ELKRSFFALRDQIPELENNEKAPKVVILKKA-TAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNS**A**A

4.6.5 Overview of all produced miniproteins

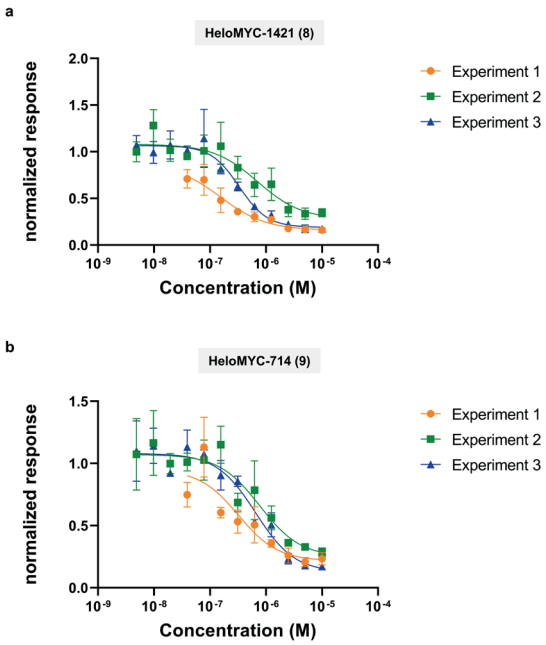
Protein	Sequence	Yield[1]	Molecular weight calculated	Molecular weight observed
Omomyc (**1**)	AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSCA	13.7 mg/L 4.5 mg/L	11527.3	11526.6
ArgMYC (**2**)	AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETRKLISEIRLLRKQNEQLKHKLERLRNSAA		11592.5	11591.8
ArtiMYC (**3**)	AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVKREIAALKREIAALKREIAALKRE	8 mg/L	10398.2	10397.6
Protein **4**	AMADIGSMATEENVKRRTHNCLERQRRCELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA	20.9 mg/L 10.4 mg/L	11488.3	11487.7
Protein **5**	AMADIGSMATEENCKRRTHNCLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA	6.2 mg/L	11503.3	11502.6
Protein **6**	AMADIGSMACEENCKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA	7.3 mg/L 2.3 mg/L	11501.3	11500.6
Protein **7**	AMADIGSMATEENCKRRCHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA	8 mg/L 4.4 mg/L	11501.3	11500.6
HeloMYC-1421 (**8**)	AMADIGSMATEENVKRRTHNCLERQRRCELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA (stapled with 1)	69%	11666.3	11665.8
HeloMYC-714 (**9**)	AMADIGSMATEENCKRRTHNCLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA(stapled with 1)	67%	11681.3	11680.8
HeloMYC-37 (**10**)	AMADIGSMACEENCKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA (stapled with 2)	28%	11603.3	11602.8
HeloMYC-711 (**11**)	AMADIGSMATEENCKRRCHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA (stapled with 2)	26%	11603.3	11602.8
Omomyc-FITC (**12**)	AMADIGSMATEENVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSCA*FITC	22%	11916.7	11916.1
HeloMYC-FITC (**13**)	AMADIGSMATEENVKRRTHNCLERQRRCELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA*FITC (stapled with 1)	43%	12055.7	12055.6
BenzoMYC (**S1**)	AMADIGSMATEENVKRRTHNCLERQRRCELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA (Cys-capped with Bz)	37%	11668.5	11667.9
NucleoMYC (**16**)	ADIGSMATEENVKRRTHNCLERQRRCELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAAPAAKRVKLD (stapled with 1)	60&	12645.2	12645.3
NucleoMYC (**17**)	ADIGSPAAKRVKLDMATEENVKRRTHNCLERQRRCELKRSFFALRDQIPELENNEKAPKV-VILKKATAYILSVQAETQKLISEIDLLRKQNEQLKHKLEQLRNSAA (stapled with 1)	46%	12645.2	12645.1

[1]First number represents yield of his-tagged protein, second number represents yield of tag-cleaved protein

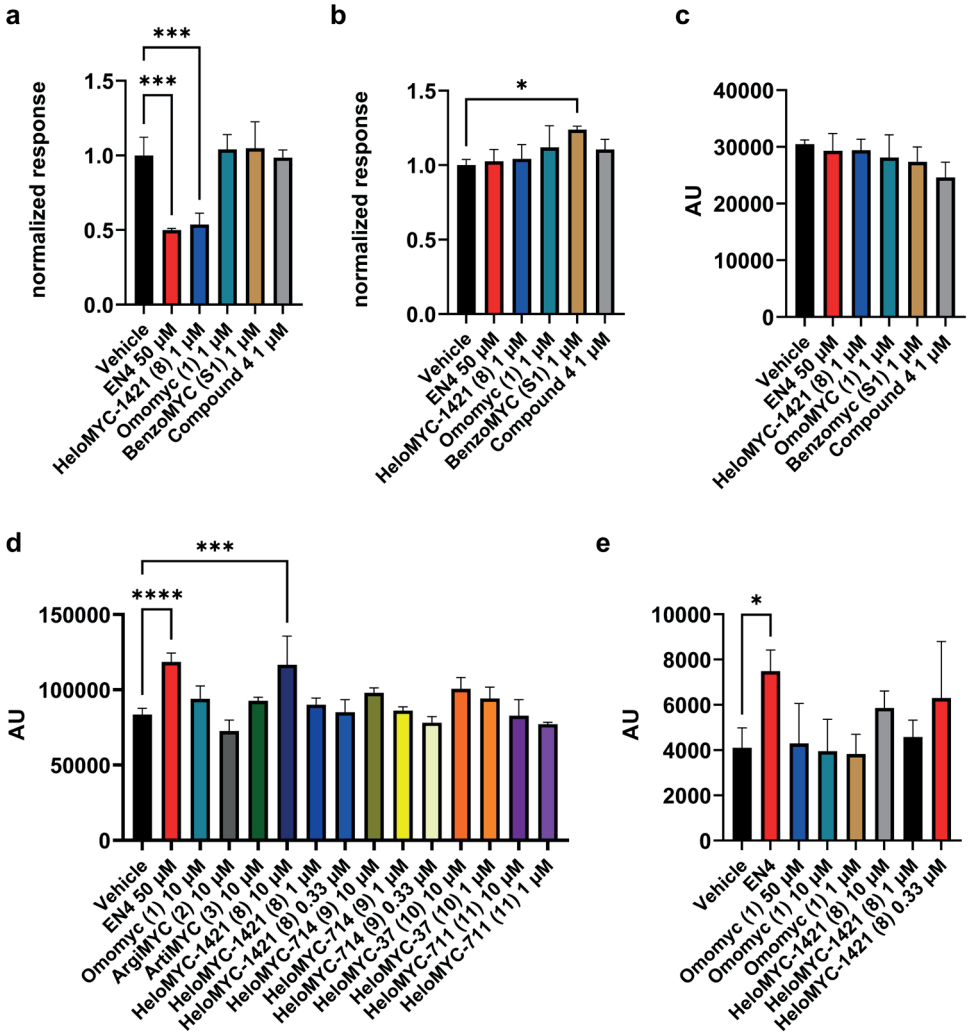
4.7 Supplementary Figures



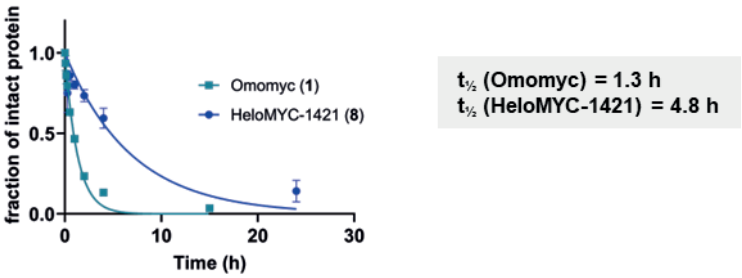
▲**Figure S4.1. Folding of Omomyc (1) and ArtiMYC (3) is very similar.** (**a-c**) Structures of ArtiMYC (**3**, a) and Omomyc (**1**, b) aligned (c). (**d**) The sequences of ArtiMYC (**3**) and Omomyc (**1**) with the artificial coiled-coil highlighted. Structure of ArtiMYC (**3**) was predicted using AlphaFold[1] and aligned to Omomyc (**1**) using Pymol. DNA shown in all panels is DNA from alignment with Omomyc (**1**).



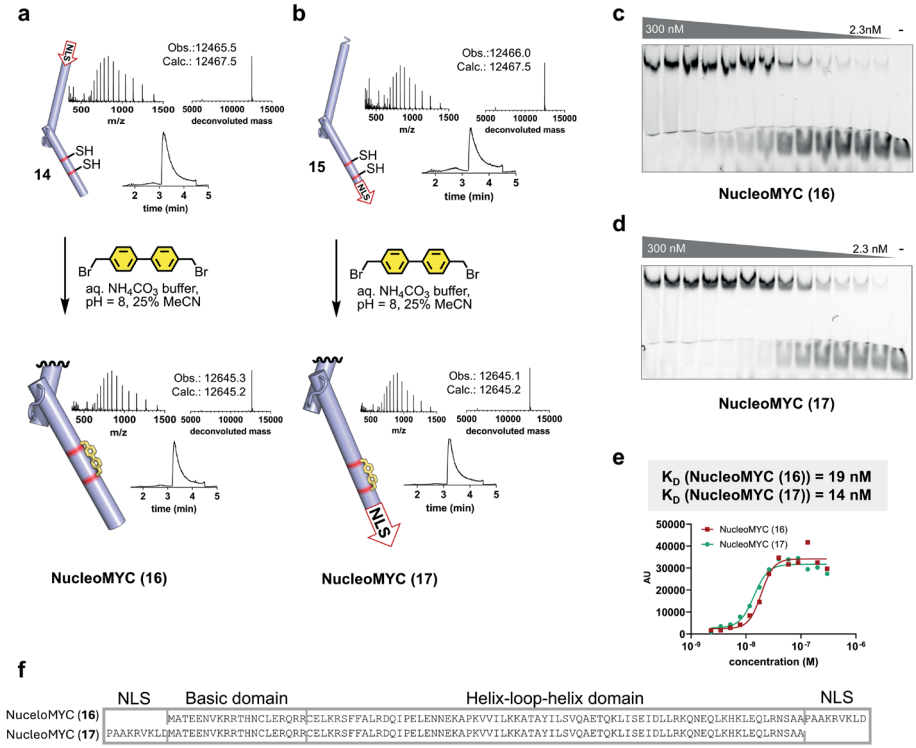
▲**Figure S4.2. Results of three independent MYC reporter gene assay experiments to determine the EC_{50} of HeloMYC-1421** (**8**, **a**) and HeloMYC-714 (**9**, **b**). Results shown have no outliers removed and might differ slightly from the one curve shown in the main paper figure 4.4b.



▲**Figure S4.3. Control experiments to validate reporter gene assay data. (a)** Positive control experiment shows effect of HeloMYC-1421 and EN4 on luciferase expression. HEK293T cells were transfected with a plasmid in which not only renilla luciferase expression but also firefly luciferase expression is under control of the CMV promoter instead of a MYC-responsive element. **(b)** Miniproteins are no direct luciferase inhibitors. HEK293T cells were transfected with the same plasmid as in panel a but treated with the indicated miniproteins shortly before cell lysis and addition of luciferase substrate to rule out any effect on gene transcription and expression. (**c**-**e**) Miniproteins have negligible effect on renilla luciferase expression under control of the CMV promoter. Shown is the renilla luciferase signal used for signal normalization of the experiments shown in figure S4.3a (panel c) and figure 4.4a and b (panels d and e, respectively) in the main text. Statistics**.** A one way ANOVA was performed to compare the effect of miniprotein treatment on firefly or renilla luciferase expression showing that there was a (a, b, d, e) or no (c) statistical difference between treatments (a) $F(5, 12) = 18.73$, $P<0.0001$; (b) $F(5, 12) = 3.115$, $P=0.0496$; (c) $F(5, 12) = 1.780$, $P=0.1915$; (d) $F(14, 30) = 8.662$, $P<0.0001$; (e) $F(7, 16) = 2.889$, $P=0.0373$.



▲**Figure S4.4** HeloMYC-1421 displays improved serum stability compared to Omomyc.

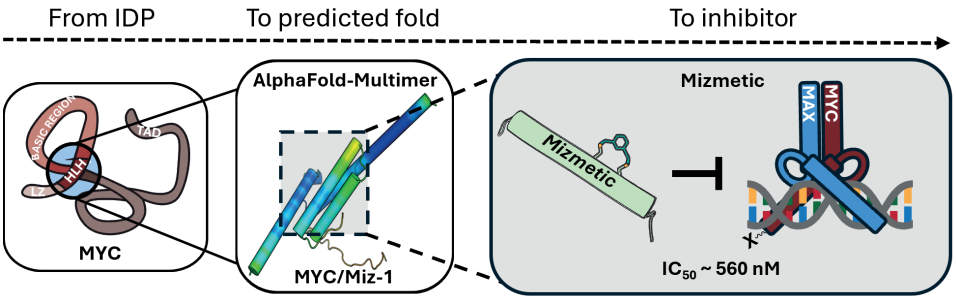


▲**Figure S4.5. NucleoMYC containing a nuclear localization sequence (NLS) fused to HeloMYC-1421 binds to E-Box DNA.** (**a, b**) An NLS sequence can be fused C- or N-terminally to the HeloMYC-1421 sequence, expressed and subsequently stapled, yielding NucleoMYC proteins **16** and **17**. (**c, d, e**) Both NucleoMYC proteins bind to E-BoxDNA in an EMSA with a K_D of 19 and 14 nM, respectively. (**f**) The sequence of NLS-containing NucleoMYC proteins.

4.8 References

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, (2021).

2. Crone, N. S. A., Kros, A. & Boyle, A. L. Modulation of Coiled-Coil Binding Strength and Fusogenicity through Peptide Stapling. *Bioconjug. Chem.* **31**, 834–843 (2020).

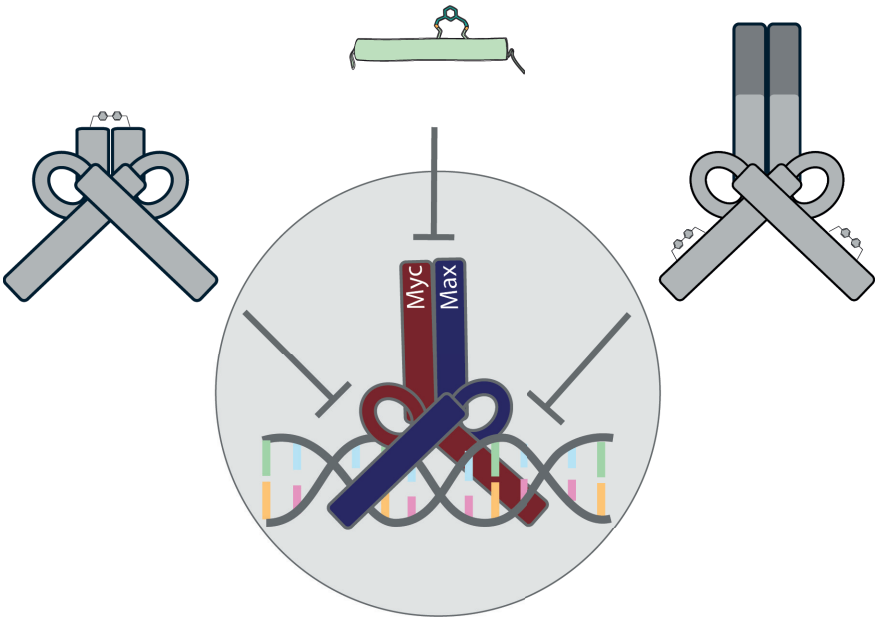


Adapted from: B. D. Ellenbroek, A. S. S. Tirtosentono, S. J. Pomplun, "AlphaFold-Guided Discovery of an Overlapping MYC/Miz-1 Interface Enables Peptidomimetic Disruption of MYC/MAX" *ChemMedChem* **2025**, e202500740.

The data not included in this thesis (e.g. uncut gels) can be found online: https://chemistry-europe.onlinelibrary.wiley.com/doi/10.1002/cmdc.202500740

Chapter 5

AlphaFold-Guided Discovery of an Overlapping MYC/Miz-1 Interface Enables Peptidomimetic Disruption of MYC/MAX



Authors:

Brecht D. Ellenbroek, Amardhiko S.S. Tirtosentono, Sebastian Pomplun

Abstract

Intrinsically disordered proteins (IDPs) lack stable structure and classical binding pockets, making them remarkably difficult to drug. However, upon interacting with partner proteins, IDPs can form transient yet specific and defined interfaces, which, if discovered, can become exploitable footholds for therapeutic intervention. Here, we use AlphaFold-Multimer to shine light on the underexplored oncogenic protein-protein interaction between MYC and Miz-1. Notably, AlphaFold-Multimer predicted that Miz-1 binds at a site overlapping with the MYC/MAX dimerization domain which is MYC's best-characterized oncogenic interface. Guided by this prediction, we designed and synthesized stapled peptidomimetics, targeting the MYC/Miz-1 and MYC/MAX interface. Our best variant, *Mizmetic* 2, binds MYC and alanine scanning identified key binding hotspots and promising sites for beneficial substitutions. The optimized variants disrupted MYC/MAX complex formation and prevented MYC from binding its canonical E-box DNA elements in vitro, with submicromolar potency. MAX/MAX DNA complex formation was also reduced, likely indicating some degree of unspecific binding. While still at a proof-of-concept stage, our findings showcase how AI-based structure prediction can illuminate disordered protein interactions and reveal new targetable sites on MYC.

5.1 Introduction

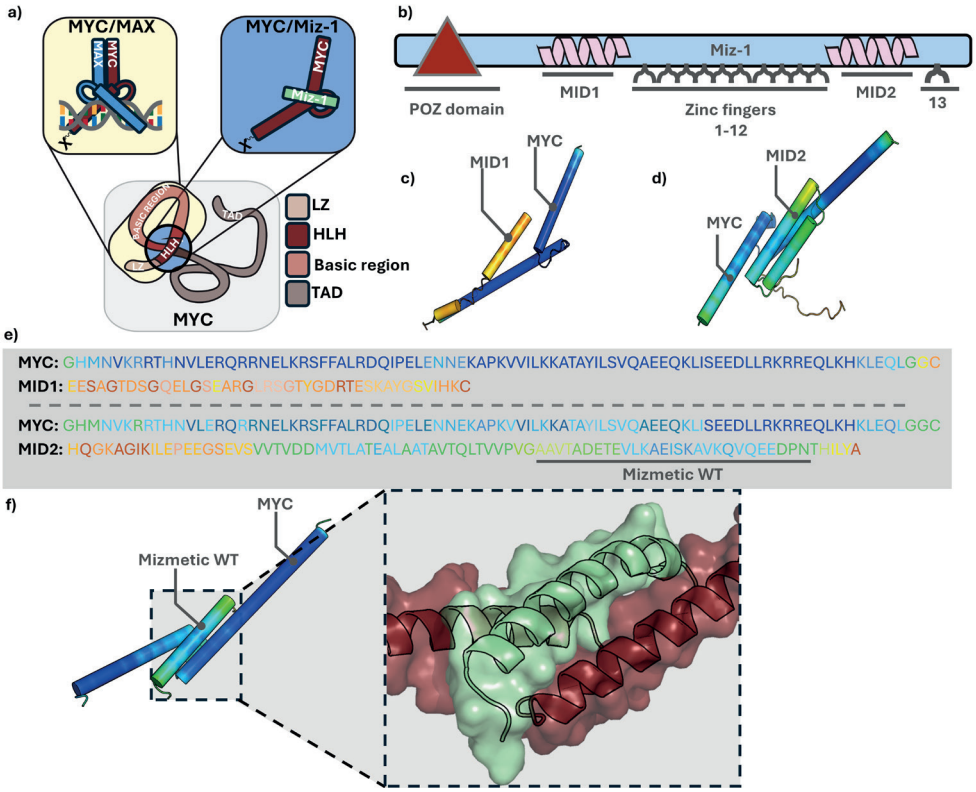
Protein-protein interactions (PPIs) usually entail large interaction interfaces. While many PPIs are involved in disease, these interactions are difficult to target. [1] The challenge becomes even more critical when intrinsically disordered regions (IDRs) or proteins (IDPs) are involved. IDRs and IDPs can adopt defined conformations when interacting with partner proteins, but their general disordered nature makes structure elucidation approaches very challenging. With approximately 22% of the human genome being an IDP or containing IDRs,[2] it is essential to identify and characterize how these proteins interact in order to target disease related IDPs. Recently more promising in silico strategies were reported, using AlphaFold[3] or RFDiffusion,[4] to predict how IDRs interact with protein partners with a high accuracy[5] or even design de novo interaction partners.[4] Yet, to enable therapeutic development further research is needed.

A well characterized disease related IDP is MYC. Dysregulation of MYC is involved in more than 50% of all human cancers. MYC belongs to the basic-helix-loop-helix leucine zipper (bHLH-LZ) protein family.[6,7] It needs to heterodimerize with protein partners, such as MYC associated protein X (MAX) (**Figure 5.1a**), to bind enhancer-box DNA (E-box, CACGTG) and regulate downstream genes through its disordered transactivation domain (TAD). When not in complex, MYC is disordered (**Figure 5.1a**).

MYC also drives tumor development in complex with MYC interacting Zinc Finger 1 (Miz-1, **Figure 5.1a**).[8,9] It was found that Miz-1 accumulates at MYC target sites.[8] Additionally, overexpression of the mutant MYC^{V394D}, unable to bind to Miz-1 but not impairing MYC/MAX dimerization, in transgenic mice resulted in delayed tumor development.[10,11] This indicated the involvement of the MYC/Miz-1 interaction in cancer. Through two-hybrid studies Peukert *et al.* identified that Miz-1 contains two MYC interacting domains (MID), MID1 and MID2, that interact with MYCs HLH domain (**Figure 5.1a-b)**, but not with MAX.[12] Structure predictions suggested that the Miz-1 MID2 domain adopts an amphipathic α-helix that can interact with the HLH domain of MYC.[12]

When MYC binds Miz-1, it becomes a repressor (e.g. of the negative cell cycle regulators $P21^{Cip1}$ and $P15^{ink4b}$) and inhibits cell differentiation. [13–15] The influence of MAX on this interplay is not fully understood yet. Whereas several literature reports found that a ternary complex between MYC/MAX and Miz-1 can be formed, other reports show that Miz-1 and MAX do compete for MYC.[8,12,16,17] In either scenario these data point out the MYC/Miz-1 interface as a promising site for therapeutic targeting.

Despite its critical involvement in cancer MYC is difficult to target with classical drugs. While there are small molecules developed that prevent MYC/MAX formation,[18,19] stabilize MAX/MAX formation,[20] promote MYC degradation[21] or protein-based



▲Figure 5.1: AlphaFold predicts that a section of 28 residues within MID2 can interact with MYCs HLH domain. a) A schematic representation of MYC and its interaction partners MAX and Miz-1. MAX and Miz-1 have an overlapping interaction site (Blue). MYC contains a transactivation domain (TAD, represented by X), basic region, Helix-Loop-Helix domain (HLH) and a Leucine Zipper (LZ) domain. MYC is disordered when not in complex. When interacting with protein partners it becomes ordered. b) A schematic representation of Miz-1. Miz-1 contains a Poxvirus and Zinc finger (POZ) domain, MYC interacting domain 1 (MID1), 13 zinc fingers and a MID2 domain. c) AlphaFold prediction between MID1 and MYC. ipTM = 0.07, pTM = 0.49. d) AlphaFold prediction between MID2 and MYC. ipTM = 0.33, pTM = 0.47 e) Sequences of MYC, MID1 and MID2 with corresponding coloring scheme for the AlphaFold predictions. Underlined: isolated 28mer, *Mizmetic* WT. f) AlphaFold prediction between *Mizmetic* WT and MYC. ipTM = 0.5 pTM = 0.59. Miz-1 (green) binds with its hydrophobic residues in the HLH domain of MYC (red). AlphaFold per-atom confident score color codes (pLDDT): Dark blue - Very high (pLDDT > 90), Cyano - Confident (90 > pLDDT > 70), Yellow - Low (70 > pLDDT > 50), Orange - Very low (pLDDT < 50). All AlphaFold predictions were performed with $MYC_{369\text{-}454}$ and the structures were generated with Pymol.

moieties that inhibit MYC function by blocking the E-box sequence,[7,22–27] no approved drugs are available. Moreover, no modalities targeting the MYC/Miz-1 interaction have been described and there is no crystal structure elucidated for this complex.

Here, we used AlphaFold-Multimer[4] to explore the structural features of the MYC/Miz-1 interaction and its potential druggability. AlphaFold predicted a region of 28amino acids in Miz-1 to bind to the HLH domain of MYC. We synthesized the peptide and demonstrated binding to MYC via biolayer interferometry (BLI). Alanine scanning and rational chemical modifications with helix stabilizing peptide staples guided us to an optimized peptidomimetic (*Mizmetic* 2) and variants able to not only bind to the Miz-1 site but also to disrupt MYC/MAX/DNA binding. Overall, our findings show that AlphaFold-Multimer is a powerful platform to analyze IDRs and their PPIs with potential for identifying novel druggable sites in challenging targets like MYC.

5.2 Results and discussion

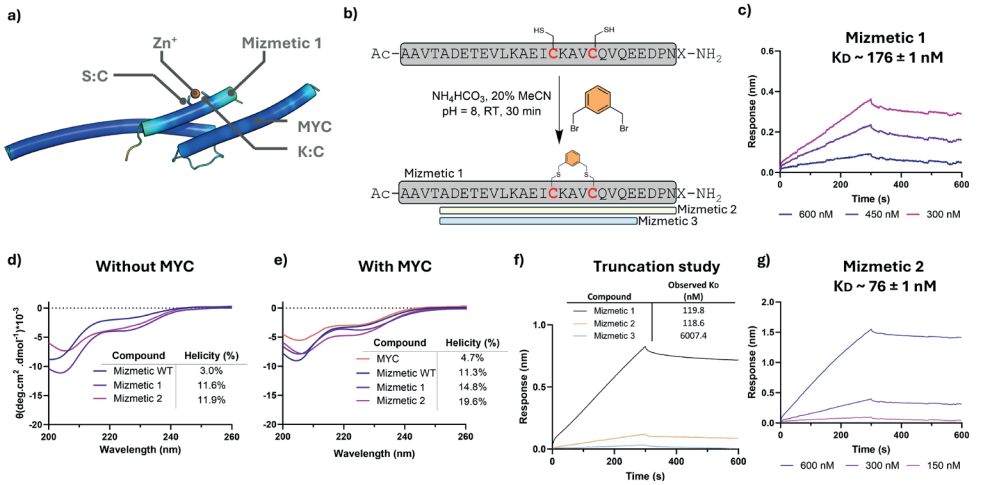
5.2.1 AlphaFold predicts MID2 to bind MYC in its HLH domain

The MYC–Miz-1 interaction is difficult to characterize experimentally due to the intrinsically disordered nature of MYC. Peukert *et al.* proposed that Miz-1 engages MYC through an amphipathic α-helix within MYC's HLH domain (**Figure 5.1a**).[15] Bédard *et al.* further showed that the MID2 region of Miz-1 (Miz-$1_{637\text{-}718}$, 84 amino acids) binds more strongly to MYC than MID1 (Miz-$1_{269\text{-}308}$, 42 amino acids).[20] However, the precise structural basis of these interactions remains unknown.

We used AlphaFold-Multimer to model MYC/MID1 (Miz-$1_{269\text{-}308}$, $MYC_{369\text{-}454}$, **Figure 5.1c&e**) and MYC/MID2 complexes (Miz-$1_{637\text{-}718}$, $MYC_{369\text{-}454}$,**Figure 5.1d&e**). AlphaFold did not predict any high-confidence binding mode for MID1 associated to MYC (**Figure 5.1c,** ipTM = 0.07, pTM = 0.49). In contrast, for MID2, AlphaFold predicted, with higher confidence (ipTM = 0.33**,** pTM = 0.47) a 28-residue α-helical segment binding in the hydrophobic groove formed by the MYC HLH domain (**Figure 5.1d**). This predicted binding mode aligns well with the hypothesis proposed by Peukert *et al.*[15] Based on these results, we focused subsequent analyses exclusively on MID2.

We performed a second AlphaFold prediction between MYC and the highest scoring sequence section of MID2 (Miz-$1_{685\text{-}712}$ = *Mizmetic* WT, $MYC_{369\text{-}454}$, ipTM = 0.5 pTM = 0.59, **Figure 5.1f**). This prediction showed a similar fold between *Mizmetic* WT and MYC with the hydrophobic residues of *Mizmetic* WT facing towards a hydrophobic region in the MYC HLH domain.

Experimental binding assays confirmed the MYC/MID2 interaction. To experimentally validate the binding between the two protein domains, we expressed and purified the bHLH-LZ region of MYC (**Figure S5.1-6**, **Table S5.1**, $MYC_{369\text{-}454}$). We synthesized *Mizmetic* WT via solid phase peptide synthesis (SPPS) and biotinylated the N-terminus (**Figure S5.1-6**, **Table S5.1**) for immobilization on the BLI biosensors. BLI indicated



▲Figure 5.2: The incorporation of an i, i+4 staple improves binding of *Mizmetic* to MYC. a) AlphaFold prediction between *Mizmetic* 1 and $MYC_{353–434}$. Serine16 and Lys20 are mutated to a cysteine allowing the incorporation of an artificial staple within AlphaFold. ipTM = 0.67, pTM = 0.68. b) Exemplary stapling reaction. X = β-alanine – Lysin(biotin) linker c) Binding evaluation of *Mizmetic* 1 by BLI. d) Structural characterization of *Mizmetic* WT, 1 and 2 with CD. *Mizmetic* WT shows mainly a random coil, *Mizmetic* 1 and 2 display a CD spectrum of a partial α-helix without MYC. e) Structural characterization of *Mizmetic* WT, 1 and 2 with CD in complex with MYC. A modest increase in α-helicity is observed around 222 nm. Helicities (%) were calculated with BestSEL analysis tool.[42,43] f) Binding evaluation of *Mizmetic* 1, 2 and 3 by BLI. All peptides were screened against 300 nM. *Mizmetic* 3 displayed a 50 times decrease in binding compared to *Mizmetic* 1. g) Binding evaluation of *Mizmetic* 2 with BLI. AlphaFold per-atom confident score color codes: Dark blue - Very high (plDDT > 90), Cyano - Confident (90 > pLDDT > 70), Yellow - Low (70 > pLDDT > 50), Orange - Very low (pLDDT < 50). All AlphaFold predictions were performed with $MYC_{369-454}$ and the structures were generated with Pymol.

that this peptide binds to MYC with an observed affinity of 4.6 µM ± 1.3 µM (**Figure S5.7**, **Table S5.2**), confirming an interaction between the two proteins. However, non-ideal binding behavior was observed by BLI, eventually caused by aggregation of the unstructured MYC domain to the tip when the analyte is bound.[31] We optimized the BLI assay conditions, in order to not observe any unspecific binding of MYC to the unfunctionalized biosensors. However, for the assays with Biotin-*Mizmetic* on the biosensors, further attempts to optimize the binding behavior, including increasing concentrations of blocking reagents or decreasing bound *Mizmetic* analytes, did not improve the shapes of the BLI curves, thereby limiting us in finding the true affinity value towards MYC. Keeping this limitation in mind, in the following text we will refer to observed KDs and in the following experiments we mainly made relative comparisons between the binding affinities rather than referring to absolute values.

5.2.2 Stapling of the Miz-1 helix improves binding to MYC

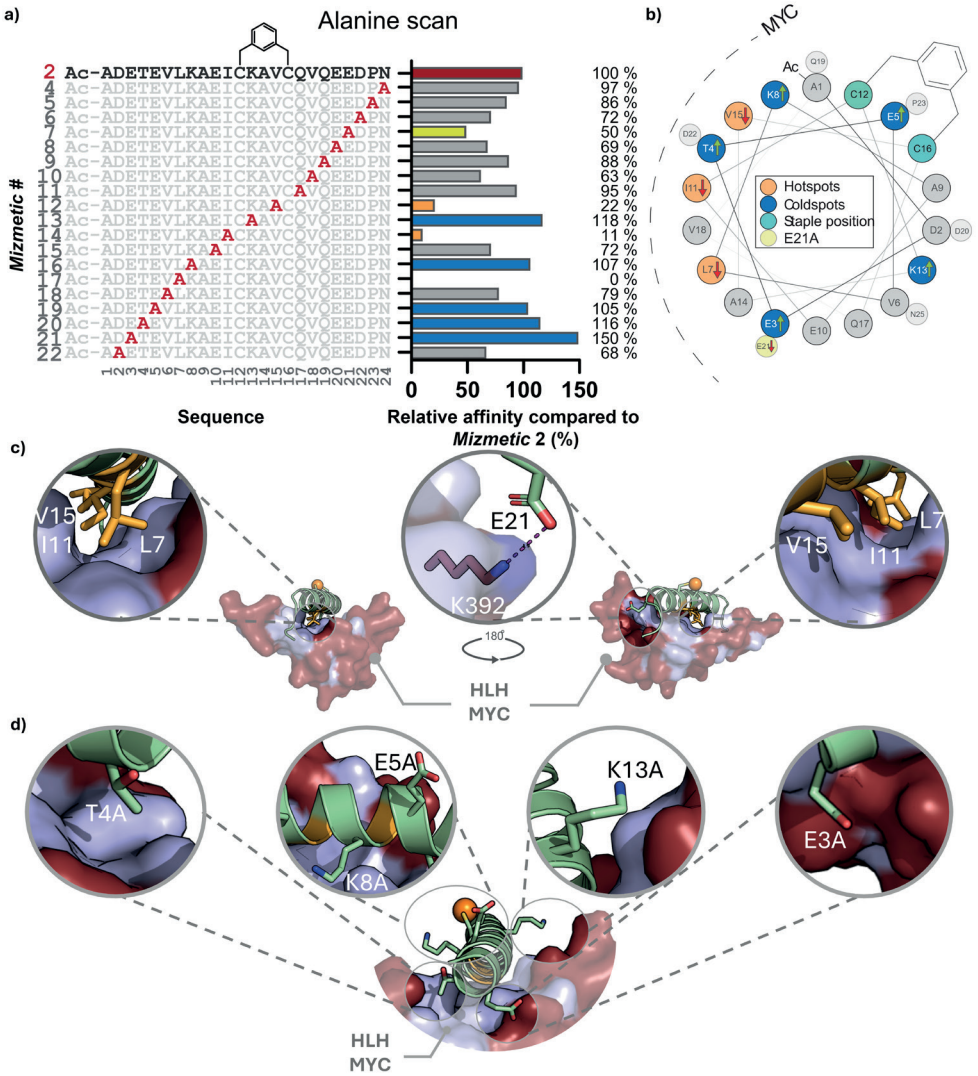
Both prior studies[15,20] and our AlphaFold predictions suggest that the MID2 region of Miz-1 adopts an α-helical conformation upon binding to MYC. We therefore hypothesized that conformational stabilization through α-helix stapling could enhance the structural rigidity of our *Mizmetic* construct and thereby improve its binding affinity for MYC.[28,32–40] Peptide stapling typically involves covalently linking two residues one or two helical turns apart (i, i + 4 or i, i + 7, respectively), locking the α-helical conformation. To explore optimal stapling positions, we employed AlphaFold modeling. Because AlphaFold does not support non-natural modifications, we mimicked helix stabilization by introducing two cysteines at positions i and i + 4 to coordinate a zinc ion. Cysteine – Zinc coordination is found as a helix stabilizing element in natural proteins, such as Zinc Fingers.[41] Ser16 and Lys20 were chosen as mutation sites, as these residues are located on the solvent-exposed "back" of the helix and are not predicted to participate in MYC binding (**Figure 5.2a**).

The AlphaFold prediction for the zinc-coordinated mutant showed an even higher confidence score compared to the unstapled construct (ipTM = 0.5, pTM = 0.59 vs. ipTM = 0.67, pTM = 0.68, **Figure 5.2a**). Based on these results, we synthesized *Mizmetic 1* with Ser16 and Lys20 replaced by cysteines, followed by thiol–aromatic crosslinking using *m*-bromoxylene to generate a covalent i, i + 4 staple (**Figure S5.1-10**, **Table S5.1**).[44,45] The reaction scheme and final product, *Mizmetic* 1, are shown in**Figure 5.2b**.

BLI indicated an increase in binding affinity for *Mizmetic* 1 compared to *Mizmetic* WT (**Figure 5.2c**). A full concentration curve displayed an observed K_D of ~ 176 ± 1 nM by BLI. When compared to *Mizmetic* WT (observed K_D ~ 4.6 μM ± 1.3 μM), this is a 26x improvement in binding, upon incorporation of the staple (**Table S5.2**).

Circular Dichroism (CD) spectroscopy was used to asses α-helicity (**Figure 5.2d-e**). Mizmetic WT displayed mainly a pattern that is characteristic for a random coil in the absence of MYC (**Figure 5.2d**, **Figure S5.8**), whereas Mizmetic 1 showed partial α-helicity with minima at 208 nm and 222 nm. BestSEL analysis[37,38] estimated helicity at 3.0% for Mizmetic WT and 11.6% for Mizmetic 1 (**Figure 5.2d**, **Figure S5.8**). MYC alone displayed 4.7% helicity (**Figure 5.2e**, **Figure S5.8**). Upon complex formation, both Mizmetic*s* showed increased helical content (**Figure 5.2e**, **Figure S5.8**).

To further optimize Mizmetic 1 we sought to determine the minimal sequence necessary for binding. Therefore, a stepwise truncation study was performed. This showcased that four amino acids at the N-terminus could be removed (Mizmetic 2, observed K_D ~ 118.6 nM, **Figure 5.2f**, **Table S5.3**) whereas an additional deletion of four amino acids at the C-terminus resulted in a drop in observed affinity according to BLI (Mizmetic 3, observed K_D ~ 6007.4 nM,



▲Figure 5.3: T The alanine scan highlighted three hotspot and five coldspot positions that are important for *Mizmetic*/MYC interaction. a) All stapled *Mizmetic* 2 alanine analogues were screened against 400 nM MYC. The alanine mutation is highlighted in red. The relative increase/decrease in affinity compared to the K_D of *Mizmetic* 2 are displayed. b) The α-helical wheel of *Mizmetic* 2 shows that the hotspots are facing the same side of the *Mizmetic* peptide and are flanked by E3, T4 and K8. c) On the left and right three hotspots are enlarged (orange) and are interacting in a hydrophobic region within the HLH domain. In the middle a potential ionic interaction between $Mizmetic^{E21}$ and MYC^{K392} is enlarged. d) Amplification of the coldspot residues with an affinity of >100% compared to *Mizmetic* 2. In red is the HLH domain of MYC depicted. In light blue the hydrophobic residues within the HLH domain. The HLH domain of MYC is portrayed in red with hydrophobic residues highlighted in light blue. All experiments with expressed $MYC_{369\text{-}454}$ and the structures were generated with Pymol.

Figure 5.2f, **Table S5.3**). A full concentration curve displayed an observed affinity of ~76 ± 1 nM for the stapled 24mer Mizmetic 2 according to BLI (**Figure 5.2g**).

Similar to Mizmetic 1, CD spectra indicated that Mizmetic 2 adopts a partially folded α-helix in the absence of MYC (**Figure 5.2d**, **Figure S5.8**). Notably, almost 20% of helicity was observed when incubated with MYC (**Figure 5.2e**, **Figure S5.8**). The formation of a partial α-helix combined with the increase in affinity indicate that the staple might promote the correct orientation of the interacting residues of Mizmetic 1 and 2 to bind MYC but is unable to promote a full α-helix as is observed when MYC and MAX are incubated together.[7,27]

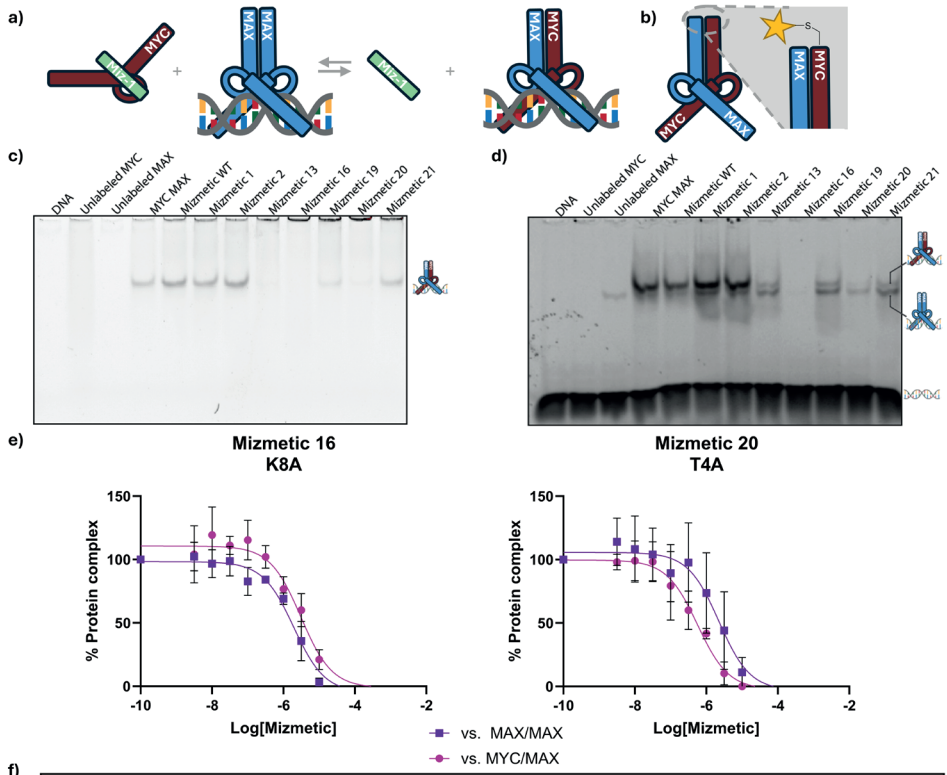
5.2.3 Alanine scanning showcased three hotspots and five coldspots

We next wanted to assess the influence of each individual amino acid of Mizmetic 2 by means of an alanine scan. We synthesized 19 alanine mutants (Mizmetic 4-22) and measured BLI at a single concentration to compare their relative affinity to Mizmetic 2 (**Figure 5.3a, Figure S5.9**, **Table S5.4**). For most positions the alanine substitution did not influence binding considerably. Alanine mutants for Leu7, Ile11 and Val15 significantly weakened binding, indicating the importance of these positions for binding to MYC. The helical wheel shows that these residues are facing the same side (**Figure 5.3b**). This observation supports the predicted AlphaFold simulation as these three hydrophobic residues were predicted to interact deeply within the hydrophobic surface of the HLH domain of MYC (**Figure 5.3c**, **orange**). Additionally, a 2-fold decrease in affinity was observed for E12A (**Figure 5.3a**). In the AlphaFold predictions this residue was facing towards MYC (**Figure 5.3c**, **middle**), potentially forming a salt bridge with K392.

Five alanine mutants displayed slightly increased binding: Glu3, Thr4, Glu5, Lys8 and Lys13. The α-helical wheel of Mizmetic 2 shows that the hydrophobic surface is flanked by Glu3, Thr4 and Lys8 (**Figure 5.3b**). Whereas Glu3 and Lys13 are predicted to face away from MYC, Thr4 and Lys8 are predicted to face towards a hydrophobic region (**Figure 5.3d**). This could explain the increase in affinity for T4A and K8A as these positions become more apolar when mutated to an alanine. E5A and K8A could potentially form an intramolecular salt-bridge that is unfavorable for Mizmetic/MYC binding (**Figure 5.3d**), resulting in alanine mutants with increased affinity. Overall, our alanine scan indicated hotspot residues, in good agreement with the AlphaFold prediction and in addition, unexpectedly it resulted in the identification of a few variants with increased affinity.

5.2.4 Mizmetic alanine analogues can inhibit MYC-MAX-DNA binding

While our Mizmetic compounds bind to the Miz-1 binding site on MYC, we set out to explore if these compounds facilitate the inhibition of the most prominent



f)

#	Sequence	IC_{50} MYC MAX (μM)	IC_{50} MAX MAX (μM)	Ratio
16	Ac-ADETEVL**A**AEICKAVCQVQEEDPN - X	3.08±1.94	1.94±0.71	0.6
20	Ac-ADE**A**EVLKAEICKAVCQVQEEDPN - X	0.56 ±0.22	2.25±1.93	4

▲Figure 5.4: Mizmetic alanine analogues can inhibit MYC/MAX and MAX/MAX complex formation. a) It is suggested that Miz-1 and Max compete to bind MYC. We hypothesized that Mizmetics can inhibit MYC/MAX complex formation. b) MYC and MAX can be labeled with Fluorescein-5-maleimide without interfering with complex formation. c) EMSA essay with labeled MYC (DNA construct: TTC CT**C ACG TG**G CAT TTG GGT GC**C ACG TG**A GG). MYC/MAX binding can be inhibited by Mizmetic 13, 16 and 20. d) EMSA assay with labeled DNA (DNA construct: IRD700-ACC CCA C**CA CGT G**GT GCC T). The MYC/MAX/DNA complex can be inhibited by Mizmetic 13, 16, 20 and 21. Mizmetic 13 and 20 show partial selectivity, whereas Mizmetic 16 shows complete inhibition. e&f) Concentration curves with Mizmetics 16 and 20. Competition assay was performed in independent triplicates. EMSA displays for Mizmetic 16 (K8A) no selectivity between MYC/MAX or MAX/MAX. Mizmetic 20 (T4A) displays moderate specificity for MYC over MAX. X = β-alanine + Lysine(biotin) linker.

cancer driving PPI: the MYC/MAX interaction. It has been proposed that a ternary complex forms in which MYC binds both Miz-1 and MAX.[8,12,16] Contrarily, it has also been proposed that Miz-1 and MAX compete for MYC interaction (**Figure 5.4a**).[17] Our AlphaFold simulations support the latter hypothesis, and therefore we hypothesized that our Mizmetic compounds could inhibit

the MYC/MAX dimer formation and with that their binding to E-Box DNA.

We used native electrophoretic mobility shift assay (EMSA) to study the potential inhibition of the protein-protein-DNA complex. We labeled MYC (MYC_F, **Figure S5.1-5.6**) with Fluorescein-5-maleimide at its C-terminal cysteine (**Figure 5.4b**). EMSA did confirm MYC/MAX/DNA formation (**Figure 5.4c** lane MYC/MAX and **Figure S5.10**). Subsequently, we incubated MYC_F and Max with either Mizmetic WT, 1, 2 as well as with the improved alanine analogues (Mizmetic 13, 16, 19-21, **Figure 5.4c**). While EMSA showed that Mizmetic WT, 1 and 2 did not inhibit MYC/MAX complex formation, variants 13, 16, and 20 resulted in complete disappearance of the MYC/MAX/DNA band, and also variants 19 and 21 show slight inhibition

In order to observe both MYC/MAX and MAX/MAX DNA complexes at the same time, we performed EMSAs also with unlabeled proteins and IRD700-labeled E-Box DNA. While of similar size, MAX/MAX/DNA and MYC/MAX/DNA caused bands at distinguished heights on the gel. With the hypothesis of Mizmetics binding to MYC[12,17] we expected to still see the MAX/MAX band (slightly below MYC/MAX) upon MYC/MAX inhibition. In accordance with the previous experiment we did not observe MYC/MAX/DNA inhibition for Mizmetic WT, 1 and 2. (**Figure 5.4d**), All alanine analogues, as seen before, displayed inhibition. For variants 13, 19, 20 and 21 the MYC/MAX/DNA band was substantially reduced (or even absent in 20), while a clear MAX/MAX/DNA band was visible. For Mizmetic 16 no protein/DNA complex was observed at all, potentially indicating some degree of unspecific binding.

To further investigate the specificity of the selected variants, we measured full DNA binding curves with inhibitor added to either MYC/MAX or MAX/MAX. For the visualization and quantification of the bands in the EMSA gels we used either labeled MYC (for MYC/MAX) or labeled MAX (MAX_F, **Figure S5.1-5.6**, for MAX/MAX) (**Figure 5.4e-f**, **Figures S5.11-5.13**). In accordance with the single point EMSA, Mizmetic 16 had no preference for MYC over MAX, displaying IC_{50}s of ~3 µM and ~2 µM, respectively. Mizmetic 20 inhibited MYC/MAX with an IC_{50} of 560 nM, 4-fold more potent than it's inhibitory activity measured for the MAX/MAX complex (**Figure 5.4e-f**).

After the observation that Mizmetic 16 and 20 also inhibited MAX/MAX/DNA we tested direct binding to MAX with BLI (**Figure S5.14**). However, no direct binding to MAX was observed at a similar concentration as MYC. MAX hypothetically exists as a dimer in solution and thus would block the binding site of both Mizmetics thereby preventing binding. Based on these data, we cannot fully explain what causes the unspecific binding in the EMSA competition assays.

5.3 Conclusion

Intrinsically disordered proteins (IDPs) remain one of the most elusive classes of

drug targets, yet their disorder often conceals well-defined structures that emerge upon binding in protein-protein complexes. Using AlphaFold-Multimer, we characterized the MYC/Miz-1 interaction, revealing a defined α-helical interface within Miz-1 MID2 that overlaps with the canonical MYC/MAX dimerization site. Guided by this model, we designed and optimized stapled Miz-1–derived peptidomimetics,improving binding affinity from micromolar to low-nanomolar potency. Alanine scanning validated AlphaFold's predicted hotspots and revealed substitutions that further enhanced activity. Notably, select variants disrupted MYC/MAX/DNA binding with submicromolar potency, highlighting a novel mode of intervention against a transcription factor long considered "undruggable." While our work remains at a proof-of-concept stage, it establishes a strategy for targeting transient but structurally precise IDP interfaces, offering a promising foothold for future therapeutic development against MYC-driven cancers.

5.4 References

[1] J. Oláh, T. Szénási, A. Lehotzky, V. Norris, J. Ovádi, "Challenges in Discovering Drugs That Target the Protein-Protein Interactions of Disordered Proteins." *Int J Mol Sci* **2022**, *23*, 1550.

[2] K. Kosoglu, Z. Aydin, N. Tuncbag, A. Gursoy, O. Keskin, "Structural coverage of the human interactome." *Brief Bioinform* **2023**, *25*, 1–11.

[3] C. Agoni, R. Fernández-Díaz, P. B. Timmons, A. Adelfio, H. Gómez, D. C. Shields, "Molecular Modelling in Bioactive Peptide Discovery and Characterisation." *Biomolecules* **2025**, *15*, 524.

[4] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, "Highly accurate protein structure prediction with AlphaFold." *Nature* **2021**, *596*, 583–589.

[5] K. Wu, H. Jiang, D. R. Hicks, C. Liu, E. Muratspahić, T. A. Ramelot, Y. Liu, K. McNally, S. Kenny, A. Mihut, A. Gaur, B. Coventry, W. Chen, A. K. Bera, A. Kang, S. Gerben, M. Y.-L. Lamb, A. Murray, X. Li, M. A. Kennedy, W. Yang, Z. Song, G. Schober, S. M. Brierley, J. O'Neill, M. H. Gelb, G. T. Montelione, E. Derivery, D. Baker, "Design of intrinsically disordered region binding proteins" *Science (1979)* **2025**, *389*, eadr8063.

[6] H. Bret, J. Gao, D. J. Zea, J. Andreani, R. Guerois, "From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2." *Nat Commun* **2024**, *15*, 597.

[7] A. Omidi, M. H. Møller, N. Malhis, J. M. Bui, J. Gsponer, "AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions." *Proc Natl Acad Sci U S A* **2024**, *121*, e2406407121.

[8] D. Piovesan, A. M. Monzon, S. C. E. Tosatto, "Intrinsic protein disorder and conditional folding in AlphaFoldDB." *Protein Sci* **2022**, *31*, e4466.

[9] S. K. Nair, S. K. Burley, "X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors." *Cell* **2003**, *112*, 193–205.

[10] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription." *ACS Cent Sci* **2021**, *7*, 1408–1418.

[11] S. Peter, J. Bultinck, K. Myant, L. A. Jaenicke, S. Walz, J. Müller, M. Gmachl, M. Treu, G. Boehmelt, C. P. Ade, W. Schmitz, A. Wiegering, C. Otto, N. Popov, O. Sansom, N. Kraut, M. Eilers, "Tumor cell-specific inhibition of MYC function using small molecule inhibitors of the HUWE1 ubiquitin ligase." *EMBO Mol Med* **2014**, *6*, 1525–1541.

[12] K. E. Wiese, S. Walz, B. von Eyss, E. Wolf, D. Athineos, O. Sansom, M. Eilers, "The role of MIZ-1 in MYC-dependent tumorigenesis." *Cold Spring Harb Perspect Med* **2013**, *3*, a014290.

[13] J. Van Riggelen, J. Müller, T. Otto, V. Beuger, A. Yetil, P. S. Choi, C. Kosan, T. Möröy, D. W. Felsher, M. Eilers, "The interaction between Myc and Miz1 is required to antagonize TGFβ-dependent autocrine signaling during lymphoma formation and maintenance" *Genes Dev* **2010**, *24*, 1281.

[14] S. Herold, M. Wanzel, V. Beuger, C. Frohme, D. Beul, T. Hillukkala, J. Syvaoja, H.-P. Saluz, F. Haenel, M. Eilers, "Negative regulation of the mammalian UV response by Myc through association with Miz-1." *Mol Cell* **2002**, *10*, 509–521.

[15] K. Peukert, P. Staller, A. Schneider, G. Carmichael, F. Hänel, M. Eilers, "An alternative pathway for gene regulation by Myc." *EMBO J* **1997**, *16*, 5672–5686.

[16] P. Staller, K. Peukert, A. Kiermaier, J. Seoane, J. Lukas, H. Karsunky, T. Möröy, J. Bartek, J. Massagué, F. Hänel, M. Eilers, "Repression of p15INK4b expression by

Myc through association with Miz-1." *Nat Cell Biol* **2001**, *3*, 392–399.

[17] J. Seoane, C. Pouponnot, P. Staller, M. Schader, M. Eilers, J. Massagué, "TGFβ influences Myc, Miz-1 and Smad to control the CDK inhibitor p15INK4b" *Nat Cell Biol* **2001**, *3*, 400–408.

[18] G. F. Claassen, S. R. Hann, "A role for transcriptional repression of p21CIP1 by c-Myc in overcoming transforming growth factor beta -induced cell-cycle arrest." *Proc Natl Acad Sci U S A* **2000**, *97*, 9498–503.

[19] S. Walz, F. Lorenzin, J. Morton, K. E. Wiese, B. von Eyss, S. Herold, L. Rycak, H. Dumay-Odelot, S. Karim, M. Bartkuhn, F. Roels, T. Wüstefeld, M. Fischer, M. Teichmann, L. Zender, C.-L. Wei, O. Sansom, E. Wolf, M. Eilers, "Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles." *Nature* **2014**, *511*, 483–7.

[20] M. Bédard, L. Maltais, M. Montagne, P. Lavigne, "Miz-1 and Max compete to engage c-Myc: implication for the mechanism of inhibition of c-Myc transcriptional activity by Miz-1." *Proteins* **2017**, *85*, 199–206.

[21] L. Boike, A. G. Cioffi, F. C. Majewski, J. Co, N. J. Henning, M. D. Jones, G. Liu, J. M. McKenna, J. A. Tallarico, M. Schirle, D. K. Nomura, "Discovery of a Functional Covalent Ligand Targeting an Intrinsically Disordered Cysteine within MYC" *Cell Chem Biol* **2021**, *28*, 4-13.e17.

[22] H. Han, A. D. Jain, M. I. Truica, J. Izquierdo-Ferrer, J. F. Anker, B. Lysy, V. Sagar, Y. Luan, Z. R. Chalmers, K. Unno, H. Mok, R. Vatapalli, Y. A. Yoo, Y. Rodriguez, I. Kandela, J. B. Parker, D. Chakravarti, R. K. Mishra, G. E. Schiltz, S. A. Abdulkadir, "Small-Molecule MYC Inhibitors Suppress Tumor Growth and Enhance Immunotherapy." *Cancer Cell* **2019**, *36*, 483-497.e15.

[23] N. B. Struntz, A. Chen, A. Deutzmann, R. M. Wilson, E. Stefan, H. L. Evans, M. A. Ramirez, T. Liang, F. Caballero, M. H. E. Wildschut, D. V Neel, D. B. Freeman, M. S. Pop, M. McConkey, S. Muller, B. H. Curtin, H. Tseng, K. R. Frombach, V. L. Butty, S. S. Levine, C. Feau, S. Elmiligy, J. A. Hong, T. A. Lewis, A. Vetere, P. A. Clemons, S. E. Malstrom, B. L. Ebert, C. Y. Lin, D. W. Felsher, A. N. Koehler, "Stabilization of the Max Homodimer with a Small Molecule Attenuates Myc-Driven Transcription." *Cell Chem Biol* **2019**, *26*, 711-723.e14.

[24] S. R. Boyd, S. Chamakuri, A. J. Trostle, H. Chen, Z. Liu, A. Jian, J. Wang, A. Malovannaya, D. W. Young, "MYC-Targeting PROTACs Lead to Bimodal Degradation and N-Terminal Truncation." *ACS Chem Biol* **2025**, *20*, 896–906.

[25] L. Soucek, M. Helmer-Citterich, A. Sacco, R. Jucker, G. Cesareni, S. Nasi, "Design and properties of a Myc derivative that efficiently homodimerizes" *Oncogene 1998 17:19* **1998**, *17*, 2463–2472.

[26] B. D. Ellenbroek, J. P. Kahler, D. Arella, C. Lin, W. Jespers, E. A.-K. Züger, M. Drukker, S. J. Pomplun, "Development of DuoMYC: a synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC." *Angew Chem Int Ed Engl* **2025**, *64*, e202416082.

[27] I. Inamoto, I. Sheoran, S. C. Popa, M. Hussain, J. A. Shin, "Combining Rational Design and Continuous Evolution on Minimalist Proteins That Target the E-box DNA Site." *ACS Chem Biol* **2021**, *16*, 35–44.

[28] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, "Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains." *Nat Biotechnol* **2023**, *41*, 541–551.

[29] Z. Z. Brown, C. Mapelli, I. Farasat, A. V Shoultz, S. A. Johnson, F. Orvieto, A. Santoprete, E. Bianchi, A. B. McCracken, K. Chen, X. Zhu, M. J. Demma, B. M. Lacey, K. A. Canada, R. M. Garbaccio, J. O'Neil, A. Walji, "Multiple Synthetic Routes to the Mini-Protein Omomyc and Coiled-Coil Domain Truncations." *J Org Chem* **2020**, *85*, 1466–1475.

[30] M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute,

"Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents." *J Am Chem Soc* **2021**, *143*, 11788–11798.

[31] N. Sherer, J.-H. Cho, "Overcoming Effects of Heterogeneous Binding on BLI Analysis." *ACS Omega* **2025**, *10*, 28422–28428.

[32] C. E. Schafmeister, J. Po, G. L. Verdine, "An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides" *J Am Chem Soc* **2000**, *122*, 5891–5892.

[33] A. Kuepper, N. M. McLoughlin, S. Neubacher, A. Yeste-Vázquez, E. Collado Camps, C. Nithin, S. Mukherjee, L. Bethge, J. M. Bujnicki, R. Brock, S. Heinrichs, T. N. Grossmann, "Constrained peptides mimic a viral suppressor of RNA silencing." *Nucleic Acids Res* **2021**, *49*, 12622–12633.

[34] T. Anananuchatkul, I. V. Chang, T. Miki, H. Tsutsumi, H. Mihara, "Construction of a Stapled α-Helix Peptide Library Displayed on Phage for the Screening of Galectin-3-Binding Peptide Ligands." *ACS Omega* **2020**, *5*, 5666–5674.

[35] P. Timmerman, J. Beld, W. C. Puijk, R. H. Meloen, "Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces." *Chembiochem* **2005**, *6*, 821–824.

[36] S. Mitra, J. E. Montgomery, M. J. Kolar, G. Li, K. J. Jeong, B. Peng, G. L. Verdine, G. B. Mills, R. E. Moellering, "Stapled peptide inhibitors of RAB25 target context-specific phenotypes in cancer." *Nat Commun* **2017**, *8*, 660.

[37] Z. Qiao, L. C. Nguyen, D. Yang, C. Dann, D. M. Thomas, M. Henn, A. Valdespino, C. S. Swenson, S. A. Oakes, M. R. Rosner, R. E. Moellering, "Direct inhibition of tumor hypoxia response with synthetic transcriptional repressors." *Nat Chem Biol* **2025**, *21*, 247–255.

[38] Q. Chu, R. E. Moellering, G. J. Hilinski, Y.-W. Kim, T. N. Grossmann, J. T.-H. Yeh, G. L. Verdine, "Towards understanding cell penetration by stapled peptides" *Medchemcomm* **2015**, *6*, 111–119.

[39] R. Mourtada, H. D. Herce, D. J. Yin, J. A. Moroco, T. E. Wales, J. R. Engen, L. D. Walensky, "Design of stapled antimicrobial peptides that are stable, nontoxic and kill antibiotic-resistant bacteria in mice." *Nat Biotechnol* **2019**, *37*, 1186–1197.

[40] A. M. Spokoyny, Y. Zou, J. J. Ling, H. Yu, Y.-S. Lin, B. L. Pentelute, "A perfluoroaryl-cysteine S(N)Ar chemistry approach to unprotected peptide stapling." *J Am Chem Soc* **2013**, *135*, 5946–5949.

[41] K. Kluska, J. Adamczyk, A. Krężel, "Metal binding properties of zinc fingers with a naturally altered metal binding site." *Metallomics* **2018**, *10*, 248–263.

[42] A. Micsonai, F. Wien, N. Murvai, M. P. Nyiri, B. Balatoni, Y.-H. Lee, T. Molnár, Y. Goto, F. Jamme, J. Kardos, "BeStSel: analysis site for protein CD spectra—2025 update" *Nucleic Acids Res* **2025**, *53*, W73–W83.

[43] A. Micsonai, É. Moussong, F. Wien, E. Boros, H. Vadászi, N. Murvai, Y.-H. Lee, T. Molnár, M. Réfrégiers, Y. Goto, Á. Tantos, J. Kardos, "BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy." *Nucleic Acids Res* **2022**, *50*, W90–W98.

[44] H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. DeGrado, D. C. Greenbaum, "Development of α-helical calpain probes by mimicking a natural protein-protein interaction." *J Am Chem Soc* **2012**, *134*, 17704–17713.

[45] P. Timmerman, W. C. Puijk, R. H. Meloen, "Functional reconstruction and synthetic mimicry of a conformational epitope using CLIPS technology." *J Mol Recognit* **2007**, *20*, 283–299.

5.5 Experimental

H_2O used in the experiments was Mili-Q water obtained from Milli-Q® Reference bought at Avantor and will from this point on be mentioned as H_2O.

5.5.1 Synthesis procedures

5.5.1.1 Automated Solid phase peptide synthesis (SPPS) – general protocol

Peptides were prepared using a Syro I XP synthesizer on a Chemmatrix rink amide resin at a 41 µmol, 20 µmol or 10 µmol scale. The same equivalents were utilized for all scales. Resin was always swollen in 1000 µL DMF for 10 min at 70 °C prior to the synthesis.

General 41 µmol SPPS protocol: Amino acids were dissolved to a concentration of 0.4 M in DMF and HATU to 0.36 M in DMF. For each coupling, amino acid solution (7.8 eq., 800 µL), HATU solution (7.0 eq., 800 µL for) and pure DIPEA (28.0 eq., 200 µL) were added to the resin (100 mg, loading capacity of 0.41 mmol/g). Coupling reactions were performed at 70 °C for 8 minutes, followed by washes with DMF (3x 1500 µL). Fmoc removal was performed with 20%:1%:79% = piperidine:Formic acid:DMF (1500 µL, 4 minutes, 70 °C) followed by washes with DMF (3x 1500 µL).

Fmoc-β-Alanine-OH coupling 41 µmol: For Fmoc-β-Alanine-OH coupling different conditions were performed. Fmoc-β-Alanine-OH was dissolved to a concentration of 0.25 M in DMF and HATU to 0.36 M in DMF. For each coupling, amino acid solution (4.9 eq., 800 µL), HATU solution (3.5 eq., 400 µL), pure DIPEA (14 eq., 100 µL) and DMF (500 µL) were added to the resin (100 mg, loading capacity of 0.41 mmol/g). Coupling reactions were performed at 70 °C for 15 minutes, followed by washes with DMF (3x 2000 µL). Fmoc removal was performed with 20%:1%:79% = piperidine:Formic acid:DMF (1800 µL, 7 minutes, 70 °C) followed by washes with DMF (3x 2000 µL).

General 20 µmol SPPS protocol (for 10 µmol 0.5x): Amino acids were dissolved to a concentration of 0.4 M in DMF and HATU to 0.36 M in DMF. For each coupling, amino acid solution (7.8 eq., 400 µL), HATU solution (7.0 eq., 400 µL), pure DIPEA (28.0 eq., 100 µL) and DMF (100 µL) were added to the resin (50 mg, loading capacity of 0.41 mmol/g). Coupling reactions were performed at 70 °C for 8 minutes, followed by washes with DMF (3x 1000 µL). Fmoc removal was performed with 20%:1%:79% = piperidine:Formic acid:DMF (1 mL, 4 minutes, 70 °C) followed by washes with DMF (3x 1000 µL).

Fmoc-β-Alanine-OH coupling 20 µmol scale (for 10 µmol 0.5x): For Fmoc-β-Alanine-OH coupling different conditions were performed. Fmoc-β-Alanine-OH

was dissolved to a concentration of 0.25 M in DMF and HATU to 0.36 M in DMF. For each coupling, amino acid solution (4.8 eq., 400 µL), HATU solution (3.5 eq., 200 µL), pure DIPEA (14.0 eq., 50 µL) and DMF (500 µL) were added to the resin (50 mg, loading capacity of 0.41 mmol/g). Coupling reactions were performed at 70 °C for 30 minutes, followed by washes with DMF (3x 2000 µL). Fmoc removal was performed with 20%:1%:79% = piperidine:Formic acid:DMF (1000 µL, 7 minutes, 70 °C) followed by washes with DMF (3x 2000 µL).

The final crude peptidyl resin was manually washed 3x with DMF followed by 3x DCM, dried by vacuum and either directly cleaved or stored at -20 °C.

5.5.1.2 Fmoc-Lys(biotin)-OH coupling protocol – general protocol

Fmoc-Lys(biotin) was coupled manually either as first building block for C-terminal incorporation or as last building block for N-terminal incorporation. Fmoc-Lys(biotin)-OH coupling was performed on a 10 µmol scale. Fmoc-Lys(Biotin)-OH (21.2 mg, 35.6 µmol, 3.5 eq.) and DIPEA (21.5 µL, 123.2 µmol, 12.3 eq.) were pre-incubated in 1000 µL DMF. To the mixture 1000 µL HATU solution in DMF (12.3 mg, 32.3 µmol, 2.9 eq.) was added and incubated for 1 min. Then, the full preactivated mixture was added to the peptidyl resin (10 µmol, 1eq.) and incubated at room temperature (RT) for 1 hour while shaking at 750 rpm. The resin was drained and washed with DMF (3x 1000 µL). Then, Fmoc removal cocktail 20%:1%:79% = piperidine:Formic acid:DMF, 1.5 mL) was added to the peptidyl resin and incubated at RT for 20 minutes while shaking at 750 rpm. The peptidyl resin was washed 3x with 1000 DMF.

5.5.1.3 Acetylation – general protocol

Dry peptidyl resin was swollen in DMF for 10 minutes prior to the addition of acetylation cocktail (10%:10%:80% = Ac_2O:DIPEA:DMF, 1.5 mL). The peptidyl resin was incubated in cocktail while shaking for 15 minutes at RT, drained under vacuum, washed with DMF (3x mL) and DCM (3x mL) and subsequently dried under vacuum. If not cleaved directly, peptidyl resin was stored at -20 °C.

5.5.1.4 Full cleavage – general protocol

Dry peptidyl resin was cleaved for 2 hours with 2 mL of the following cleavage cocktail: 92.5%:5%:2.5% = TFA:H_2O:TIPS. Then, cleavage cocktail was evaporated under a gentle nitrogen stream (to < 0.5 mL). To the peptide ice cold diethyl ether (~5 mL) was added and peptide was obtained by centrifugation for 5 minutes at 10.000 rpm. Crude peptide pellet was dissolved in 10%:0.1%:89.9% = MeCN:TFA:H_2O and lyophilized.

5.5.1.5 Purification – general protocol

All Mizmetic compounds were purified on a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 25 g column using UV detection at 214 nm. The following gradient was used: 15% solvent B over 2 column volumes (CV), 15% to 55% solvent B over 10 CV, 55% to 98% solvent B over 0 CV, 98% solvent B over 4 CV with H_2O with 0.1% TFA as Solvent A and MeCN with 0.1%TFA as Solvent B. Fractions were analyzed by LC-MS.

5.5.1.6 Cysteine *i, i+4* stapling with *1,3-bis(bromomethyl)benzene* – general protocol

Lyophilized Mizmetic 1-22 (1 eq., 10 mM) and 1,3-bis(bromomethyl)benzene (1 eq., 10 mM) were dissolved in 80% NH_4HCO_3 buffer (100 mM, pH = 8) containing 20% MeCN and let to react for 30 minutes at RT. The reaction was stopped by flash freezing in liquid nitrogen followed by lyophilization or, alternatively, the mixture was directly purified according to the general purification protocol.

5.5.1.7 LC-MS

LC-MS spectra were obtained with a Shimazu LCMS-2020 system (Method A) equipped with a Kinetex® 2.6µm XB-C18 100 Å LC Column (50 x 3 mm) and UV detector set to 214 nM. High resolution LC-MS spectra were obtained with a Sciex X500b QTOF ESI-QToF mass spectrometer coupled to a Shimadzu Nexera UHPLC LC40DX3 (Method B) equipped with a Phenomenex Synergi™ 4 µm Fusion-RP 80 Å LC Column (50 x 2 mm) or Aeris™ 3.6 µm Widepore XB-C18 set to 214 nM. For both Method A and Method B the solvent phases used were solvent A (0.1% formic acid in water) and solvent B (0.1% formic acid in MeCN).

Method A:

Column: Kinetex® 2.6µm XB-C18 100 Å LC Column (50 x 3 mm) UV detector: 214 nm.

LC Method: 0% solvent B for 1 min, 0% to 70% solvent B over 10 min 70% solvent B over 1.5 min, 70% to 0% solvent B over 4.5, flowrate 0.55 mL/min.

Method B (High resolution):

Column: Phenomenex Synergi™ 4 µm Fusion-RP 80 Å LC Column (50 x 2 mm) or Aeris™ 3.6 µm Widepore XB-C18.

LC Method: 0% solvent B for 1 min, 0% to 90% solvent B over 6 min, 90% solvent B

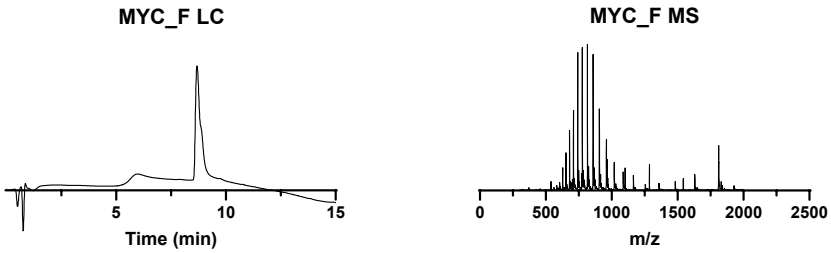
over 2 min, 90% to 0% solvent B over 0.5 min, 0% solvent B over 1.5 min, flowrate 0.5 mL/min.

MS parameters: General parameters: Method duration: 5 min; Total scan time: 0.276 sec; Estimated cycles: 1086; Intact protein mode: False; Decrease detector voltage: False; Large protein (>70 kDa): False; Ion Source: Source name: TurbolonSpray; Curtain gas: 35 psi; Ion source gas 1: 60 psi; Ion source gas 2: 60 psi; Temperature: 500 ˚C; Experiment: Scan type: TOF MS; Polarity: Positive; Spray voltage: 5500 V; CAD gas: 7; Time bins to sum: 4; Channel 1-4: True; TOF start mass: 350 Da; TOF stop mass 1500 Da; Accumulation time: 0.25; Declustering potential: 80V; Declustering potential spread: 0 V; Collision energy: 10V; Collision energy spread: 0 V; Override Qjet RF value: False

5.5.2 Protein synthesis with LCMS spectra

5.5.2.1 Fluorescent labeling of MYC (MYC_F)

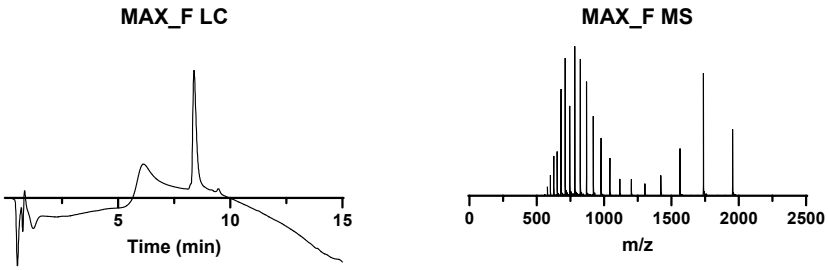
Dulbecco's Phosphate buffered saline (DPBS, pH = 6.75) was degassed using a gentle nitrogen stream for 15 minutes. MYC (2.28 mg, 143.7 nmol, 1 eq.) was dissolved in H_2O (228 µL, 10 mg/mL). A 10 mg/mL stock solution of N-(5-Fluoresceinyl) maleimide in DMSO was prepared. The complete MYC solution, was mixed with 6.14 µL of N-(5-Fluoresceinyl)maleimide stock (0.06 mg, 143.7 nmol, 1 eq.) and 905.9 µL degassed buffer resulting in a final concentration of 2 mg/mL of MYC and 0.05 mg/mL N-(5-Fluoresceinyl)maleimide. The reaction was quenched by flash freezing in liquid nitrogen followed by lyophilization after 15 minutes and purified on a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 25g column using UV detection at 214 nm with H_2O + 0.1% TFA (Buffer A) and MeCN+0.1% TFA (Buffer B). The following gradient was used: 0% buffer B for 2 CV, 0% to 80% buffer B over 11.4 CV, 95% Buffer B for 2 CV. The combined fractions containing the protein were lyophilized, yielding the protein as TFA salt. Yield: 1.47 mg, 72.5 nmol, 50%



UV LC-MS trace of MYC_F. LC-MS trace was obtained with Method A.

5.5.2.2 Fluorescent labeling of His-tagged MAX (MAX_F)

DPBS (pH = 6.75) was degassed using a gentle nitrogen stream for 15 minutes. MAX (7.97 mg, 524.2 nmol, 1 eq., MAX) was dissolved in H_2O (797 µL, 10 mg/mL). A 10 mg/mL stock solution of N-(5-Fluoresceinyl)maleimide in DMSO was prepared. MAX, was mixed with 22.4 µL of N-(5-Fluoresceinyl)maleimide stock (0.22 mg, 524.2 nmol, 1 eq.) and 3165.5 µL degassed buffer resulting in final concentration of 2 mg/mL of MAX and 0.06 mg/mL N-(5-Fluoresceinyl)maleimide. The reaction was quenched by flash freezing in liquid nitrogen followed by lyophilization after 30 minutes and purified on a Biotage® Selekt Flash Purification System equipped with a Biotage® Sfär C18 D - Duo 25 g column using UV detection at 214 nm with H_2O + 0.1% TFA (Buffer A) and MeCN+0.1% TFA (Buffer B). The following gradient was performed: 0% buffer B for 3 CV, 0% to 30% buffer B over 4.3 CV, 30% to 70% buffer B over 5.7 CV, 95% Buffer B for 2 CV. The combined fractions containing the protein were lyophilized, yielding the protein as TFA salt. Yield: 3.64 mg, 220 nmol, 67%



UV LC-MS trace of MYC_F. LC-MS trace was obtained with Method A.

5.5.2.3 Mizmetic peptides

All Mizmetic peptides were synthesized according to the protocols described in Supplemental info 2.1 -2.6. Yields can be found in section 5.1 and LC-MS traces can be found in section 7.1. All Mizmetic peptides were stored as a 1 mM solution in pure DMSO.

5.5.3. Biochemistry

5.5.3.1. Overview of buffers used

MYC Bacterial cell lysis buffer	20mM HEPES, pH 8, 0.5 M NaCl, 3mM $MgCl_2$, 0.05% DNAse, complete EDTA free protease inhibitor (1 tablet freshly added in 50 mL buffer)
MAX Bacterial cell lysis buffer	20 mM Tris-HCl, pH 8, 0.5 M NaCl, 3mM $MgCl_2$, 0.05% DNAse, complete EDTA free protease inhibitor (1 tablet freshly added in 50 mL buffer)
MYC extraction buffer	20 mM HEPES, 6M Gdn HCl, 0.5 M NaCl, 10 mM imidazole, pH = 8
MYC affinity wash buffer	20 mM HEPES, 6M urea, 0.5 M NaCl, 10 mM imidazole, pH = 8
MYC affinity elution buffer	20 mM HEPES, 6M urea, 0.5 M NaCl, 500 mM imidazole, pH = 8
MAX affinity wash buffer	20 mM Tris-HCl, 0.5 M NaCl, 10mM imidazole, pH = 8
MAX affinity elution buffer	20 mM Tris-HCl, 0.5 M NaCl, 500mM imidazole, pH = 8
EMSA loading buffer	20 mM HEPES, 100 mM NaCl, 5% glycerol, 1 mM EDTA, 20 mM $MgCl_2$, 1 mM DTT and 0.02% Tween-20, pH = 7
EMSA running buffer	25 mM Tris-HCl, 192 mM Glycine, pH = 8.3
BLI kinetic buffer	10 mM HEPES, 50 mM NaCl, 0.02% Tween-20, 0.1% BSA, pH = 7.4

5.5.3.2. Expression of MYC and MAX in E.Coli

Performed as described elsewhere.[1] pET-30 plasmid with His-tagged $MYC_{369-454}$ (Uniprot entry P01106) or His-tagged MAX were transformed into competent ArcticXpress DE3 RIL cells (Agilent) according to suppliers manual. All LB medium used was supplemented with kanamycin and gentamycin. Single colonies were cultured overnight at 37 °C and 200 rpm in 100 mL LB medium. Overnight culture was evenly distributed over 3x2L erlenmeyer flasks (total volume 6L) and allowed to reach an OD of 0.8 at 37 °C and 200 rpm. Then, temperature was cooled down to 14 °C before 100 µM IPTG was administered. After 18 hours, cells were harvested by centrifugation at 4 °C and 5000g for 12 minutes. Then, cell pellet was lysed with pressure lysis and lysis buffer (see overview of buffers 5.1). Lastly, cell debris, when MYC was expressed including inclusion bodies, was separated from supernatant by centrifugation at 4 °C and 50.000g for 45 minutes.

5.5.3.3. Purification of His-tagged MAX (MAX)

The supernatant after ultracentrifugation was purified using an ÄKTA start protein purification system attached to a 5 mL HisTrap HP His-tag protein purification column (Cytiva). Unbound protein was allowed to flow through at 100% wash buffer (see

Supplementary 5.1) before MAX was eluted with elution buffer (see Supplementary 5.1) over 40-50 CV. Pure fractions were analyzed by LC-MS method B.

5.5.3.4. Purification of His-tagged MYC (MYC) from inclusion bodies

The ultracentifuge pellet was first washed with 2M urea in Buffer A (500 mM NaCl, 20 mM HEPES, 10 mM imidazole, pH=8), then 1% Triton x-100 in buffer A and lastly with 1M NaCl in Buffer A following an extraction in 40 mL of extraction buffer (6M Gdn HCl, 500 mM NaCl, 20 mM HEPES, 10 mM imidazole, pH=8) and left overnight at 4 °C while rotating. The mixture was then centrifuged for 30 min at 10.000g and the supernatant was collected and stored at 4 °C for later use. The pellet was then again incubated in 15 mL of extraction buffer and stirred for 16h at RT followed by centrifugation for 30 min at 10.000g. The supernatant was collected and combined with the previously obtained supernatant. Next, the supernatant was loaded onto an Äkta pure purification system equipped with a 5 mL HisTrap HP His-tag protein purification column (Cytiva) that was primed with the extraction buffer with 4 mL/min. The system was subsequently washed with 20 mL of buffer A (20 mM HEPES, 6M urea, 500 mM NaCl, 10 mM imidazole, pH = 8) followed by a gradient of 10-250mM imidazole Buffer B (20 mM HEPES, 6M urea, 500 mM NaCl, 500 mM imidazole, pH = 8) over 30 CV. The system was then washed with 20 mL of 500mM imidazole Buffer B. All fractions where stored at -80 °C when not directly used and analyzed by SDS-PAGE and LC-MS.

5.5.3.5. Protocol for buffer exchange His-MAX (MAX)

Pure fractions were incubated with 0.5-1 mM TCEP for 1 hour before buffer exchange was performed with a Biotage® Selekt Flash Purification System attached to a Biotage® Sfär C18 D - Duo 25g or 50g column using UV detection at 214 nm. A stepwise gradient was performed with 100% H_2O (with 0.1% TFA) to 50% in MeCN (with 0.1% TFA), to 100% MeCN (with 0.1% TFA). Fractions were analyzed with LC-MS, combined and lyophilized yielding His-MAX as TFA salt.

5.5.3.6. Protocol for buffer exchange His-MYC (MYC)

Pure fractions were combined and buffer exchange was performed using reverse phase chromatography using a Biotage Selekt attached to a Biotage® Sfär Bio C18 D Duo 300 Å 20 µm 50g column using UV detection at 214 nm with H_2O +0.1% TFA (solvent A) and MeCN+0.1% TFA (Solvent B). The following gradient was performed: 0%B for 5CV, then 50%B for 4CV, then 100%B for 3 CV. Fractions were analyzed with LC-MS, combined and lyophilized. An unknown post translational modification was observed for MYC with an extra mass of +75. MYC was incubated with 100 eq. TCEP in H_2O for 2 hours at RT and buffer exchange was performed using the same protocol. The combined fractions containing the protein were lyophilized, yielding

the His-tagged protein as TFA salt.

5.5.3.7. General protocol for SDS-PAGE

Fractions from protein purification were 10-1000x diluted and mixed with 4x Laemmli buffer with β-mercaptoethanol. Samples were incubated 5 minutes for 95 °C, loaded on a 12.5% acrylamide gel (prepared with acrylamide 37.5:1 acrylamide/Bis) and ran for 45 minutes at 200 V at RT. Then, gel was stained with Coomassie blue for 1h or overnight and 3x destained for at least 15 minutes with refreshed destaining solution (50% MeOH, 40 % H_2O, 10% AcOH). Gels were analyzed with a Bio-Rad ChemiDoc MP machine.

5.5.3.8 Biolayer interferometry (BLI)

All Mizmetic compounds where directly diluted from their stock solution to 50 nM solution in kinetic buffer (10 mM HEPES, 50 mM NaCl, 0.02% Tween-20, 0.1% BSA, pH = 7.4) and screened against MYC in kinetic buffer. For a concentration curve MYC was stepwise diluted in kinetic buffer. To obtain an affinity for Mizmetic WT, the average of three separate measurements was calculated, screened against 5 µM MYC and the background removed. The following protocol was performed for a standard BLI assay:

The streptadavin coated tips where first soaked in kinetic buffer for at least 10 minutes, while the 96 well plate was shaking at 1000 rpm and 37 °C. The following steps were performed: baseline step (1 min), Mizmetic loading step (3 min), washing step (1 min), additional separate washing step (1 min), association step (5 min), dissociation step (5 min). All steps where performed at 37 °C while shaking at 1000 rpm. After measurement, the background measurement performed was without peptides loaded on the tips against the same MYC concentration and these background traces were subtracted. Fitting and kinetic K_Ds calculations were performed with the provided Sartorius Octed software. An Sartorius Octet R4 was used for the assay. The graph was plotted using GraphPad Prism 10.1.0.

5.5.3.9 Circular Dichroism (CD)

Lyophilized Mizmetic peptides were dissolved to 50 µM (Mizmetic 1C and 2) or 100 µM (Mizmetic WT) in 200 µL Tris acetate buffer (10 mM Tris acetate, 50 mM NaCl, pH = 7.4) and measured with a Jasco J-815 machine. Samples were measured in a High Precision Cell quartz cuvette (Hellma Analytics) with 1 mm Pathlength at 37 °C. Then, 10 µM of MYC was added to the same sample. Mizmetic and MYC were incubated in the machine for 10 minutes at 37 °C before measurement. The following CD parameters were utilized: start 260 nm, end 200 nm, Sensitivity standard, D.I.T. 1 sec., Data pitch 1 nm, scanning mode continuous, scanning speed 100 nm/min,

accumulations 5. The background was removed for all spectra. Spectra with MYC and Mizmetics were normalized twice, first for Mizmetic, second for MYC. The average values were used to portray the CD spectra. % helicity was calculated using the BestSEL software tool. [2,3] To calculate % helicity for Mizmetics together with MYC helicity was calculated normalized to both Mizmetic and MYC. The average value is reported. Spectra's were normalized according to the following formula:[4]

$$\left[\emptyset\right] = \frac{100*\emptyset_{obs}}{c*n*l}$$

With θ in deg*cm²*dmol⁻¹, θ_{obs} in mdeg, concentration (c) in mM, peptide bonds (n) and path length of cuvette in cm. The graph was plotted using GraphPad Prism 10.1.0.

5.5.3.10 Truncation comparison of Mizmetics

The general BLI protocol was performed, screened against 300 nM MYC. After subtraction of the background, the K_D was calculated and the K_D compared to Mizmetic 1**.** The graph was plotted using GraphPad Prism 10.1.0.

5.5.3.11 Alanine scan comparison of Mizmetics

For the alanine scan the general BLI protocol was performed, screened against 400 nM MYC. After subtraction of the background, the K_D was calculated and the relative K_D compared to Mizmetic 2 in the same assay was plotted in a barplot using GraphPad Prism 10.1.0. Following formula was used to calculate relative K_D.

$$\left[Relative\ \ K_D\right] = \frac{K_{D\ \ Mizmetic\ \ 2}}{K_{D\ \ variants}} * 100\,\%$$

5.5.3.12 General protocol for competition assay with labeled DNA

5x 7.5% native gels were prepared using 7.5 mL 30% Acrylamide/Bis solution (19:1 with 5% crosslinker), 7.5 mL 1.5M Tris pH = 8.8, 15 mL H_2O, 30 µL TEMED and 150 µL 10% APS. Mizmetic peptides were diluted to 40 µM in 1x loading buffer (20 mM HEPES, 100 mM NaCl, 5% glycerol, pH = 7, 20 mM $MgCl_2$, 1 mM DTT, 0.02% Tween-20, see Supplemental 5.1), MYC/MAX were separately diluted to 60 nM in 1x loading buffer. Labelled DNA construct (IRD700-ACC CCA C**CA CGT G**GT GCC T) was first incubated at 95 °C for 5 min, cooled down and diluted to 16 nM (4x) in 1x loading buffer. Next, 5 µL of 4x Mizmetic was incubated while shaking at RT with 5 µL of 4x MYC. MYC was also in a separate lane incubated with 200 µM 10058-F4, a known MYC/MAX inhibitor.[5] After 1h, 5 µL of 4x MAX and 5 µL labeled DNA construct were added to each well resulting in the following concentrations: Mizmetic 10 µM, MYC 15 nM, MAX 15 nM, DNA 4 nM, DMSO 1%. Lanes always contained

≤1% DMSO. This solution was incubated for 5 min at RT while shaking followed by placement on ice while loading 12.5 µL onto the gel which was pre-run before for 1h at 100V at 4 °C in 1x running buffer (25 mM Tris-HCl, 192 mM Glycine, pH = 8.3). Samples ran for 80 minutes at 75V at 4 °C. Gels were rinsed with demi water and DNA was visualized with Bio-Rad ChemiDoc MP machine.

5.5.3.13 General protocol for competition assay with labeled MYC (MYC_F) or MAX (MAX_F)

The standard protocol described in 4.13 was followed. For MYC labeled gels, MYC_F and MAX were separately diluted to 60 nM in 1x loading buffer. For MAX labeled gels, MAX_F was diluted to 120 nM in 1x loading buffer. The following unlabeled DNA hairpin construct was used: TTC CT**C ACG TG**G CAT TTG GGT GC**C ACG TG**A GG. MYC_F or MAX_F was visualize with a Bio-Rad ChemiDoc MP machine. Next, gels were incubated with SYBR Gold (1:50.000 in running buffer) for 5-40 minutes and again visualized with the Bio-Rad ChemiDoc MP machine. Lanes always contained ≤1% DMSO.

For a concentration curve, the Mizmetic peptide was first stepwise diluted in DMSO followed by a further dilution in loading buffer (20 mM HEPES, 100 mM NaCl, 5% glycerol, pH = 7, 20 mM $MgCl_2$, 1 mM DTT, 0.02% Tween-20) to a 4x concentration. The assay was performed with following final concentrations: Mizmetic 10-0.1 µM**,** MYC_F 15 nM, MAX 15 nM, DNA 40 nM or Mizmetic 10-0.1 µM**,** MAX_F 30 nM, DNA 40 nM. Gels were quantified with ImageJ 1.54f. For MYC gels the intensity of free MYC was removed from each lane. IC_{50} was calculated using GraphPad Prism 10.1.0 with the Nonlinear Regression One site – Fit logIC50 analysis. All concentration curves were performed in at least duplicates.

5.5.4 AlphaFold-Multimer inputs

MID1-MYC

MID1:	EESAGTDSGQELGSEARGLRSGTYGDRTESKAYGSVIHKC
MYC:	GHMNVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILK-KATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC

MID2-MYC

MID2:	HQGKAGIKILEPEEGSEVSVVTVDDMVTLATEALAATAVTQLTVVPV-GAAVTADETEVLKAEISKAVKQVQEEDPNTHILYA
MYC:	GHMNVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILK-KATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC

Mizmetic WT - MYC

Mizmetic WT:	AAVTADETEVLKAEISKAVKQVQEEDPN
MYC:	GHMNVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILK-KATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGGC

Mizmetic Zinc mutant - MYC

Mizmetic:	AAVTADETEVLKAEICKAVCQVQEEDPN
MYC:	GHMNVKRRTHNVLERQRRNELKRSFFALRDQIPELENNEKAPKVVILK-KATAYILSVQAEEQKLISEEDLLRKRREQLKHKLEQLGG
Ion	Zinc

5.5.5 Supporting Tables

Supporting table 5.1: All compounds with corresponding sequence, stapling, calculated mass, found mass and yield. Yield is reported over peptide synthesis and stapling.

Compound #		Sequence (N –> C)	Stapled	m/z calc.	m/z found	Yield (%)	Yield (mg)	Yield (µmol)
WT	NH_2	K(Biotin)A_βAAVTADETEVLKAEISKAVKQVQEEDPN	No	3437.87	3437.8	4	5.51	1.6
1	Ac	AAVTADETEVLKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3573.07	3572.1	1.5	1.1	0.9
2	Ac	----ADETEVLKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3230.67	3229.8	1.9	2.5	0.4
3	Ac	----ADETEVLKAEICKAVCQVQE----A_βK(Biotin)	Yes	2775.25	2774.9	1.3	0.72	0.3
4	Ac	----ADETEVLKAEICKAVCQVQEEDPAA$_\beta$K(Biotin)	Yes	3187.65	3187.4	6.7%	2.21	0.7
5	Ac	----ADETEVLKAEICKAVCQVQEEDANA$_\beta$K(Biotin)	Yes	3204.64	3204.4	5.9%	1.96	0.6
6	Ac	----ADETEVLKAEICKAVCQVQEEAPNA_βK(Biotin)	Yes	3186.67	3186.5	9.2%	3.02	0.9
7	Ac	----ADETEVLKAEICKAVCQVQEADPNA_βK(Biotin)	Yes	3172.64	3172.4	5.8%	1.9	0.6
8	Ac	----ADETEVLKAEICKAVCQVQAEDPNA_βK(Biotin)	Yes	3172.64	3172.4	7.2%	2.39	0.8
9	Ac	----ADETEVLKAEICKAVCQVAEEDPNA_βK(Biotin)	Yes	3173.62	3173.4	6.7%	2.22	0.7
10	Ac	----ADETEVLKAEICKAVCQAQEEDPNA_βK(Biotin)	Yes	3202.62	3202.4	12.2%	4.02	1.3
11	Ac	----ADETEVLKAEICKAVCAVQEEDPNA_βK(Biotin)	Yes	3173.62	3173.4	7.6%	2.52	0.8
12	Ac	----ADETEVLKAEICKAACQVQEEDPNA_βK(Biotin)	Yes	3202.62	3202.0	9.7%	3.2	1
13	Ac	----ADETEVLKAEICAAVCQVQEEDPNA_βK(Biotin)	Yes	3173.58	3173.2	0.6%	0.2	0.06
14	Ac	----ADETEVLKAEACKAVCQVQEEDPNA_βK(Biotin)	Yes	3188.59	3188.3	3.4%	1.12	0.4
15	Ac	----ADETEVLKAAICKAVCQVQEEDPNA_βK(Biotin)	Yes	3172.64	3172.2	1.7%	0.55	0.2
16	Ac	----ADETEVLAAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3173.58	3173.2	0.7%	0.98	0.3
17	Ac	----ADETEVAKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3188.59	3188.3	1.4%	0.46	0.1
18	Ac	----ADETEALKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3202.62	3202.4	4.9%	1.6	0.5
19	Ac	----ADETAVLKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3172.64	3172.4	0.4%	0.14	0.04
20	Ac	----ADEAEVLKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3200.65	3200.4	3.7%	1.22	0.4
21	Ac	----ADATEVLKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3172.64	3172.5	3.6%	0.12	0.04
22	Ac	----AAETEVLKAEICKAVCQVQEEDPNA_βK(Biotin)	Yes	3186.67	3186.5	1.5%	0.49	0.2

Supporting table 5.1: All compounds with corresponding sequence, stapling, calculated mass, found mass and yield. Yield is reported over peptide synthesis and stapling.

Compound #		Sequence (N –> C)	Stapled	m/z calc.	m/z found	Yield (%)	Yield (mg)	Yield (µmol)
His-MYC (MYC)	NH_2	MHHHHHHSSGLVPRGSGMKETAAAKFERQHMD-SPDLGTDDDDKAMADIGSNVKRRTHNVLERQRR-NELKRSFFALRDQIPELENNEKAPKVVILKKATAY-ILSVQAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA	N.A	15865.82	15865.5	N.A	N.A	N.A
MYC_F	NH_2	MHHHHHHSSGLVPRGSGMKETAAAKFER-QHMDSPDLGTDDDDKAMADIGSNVKRRTH-NVLERQRRNELKRSFFALRDQIPELEN-NEKAPKVVILKKATAYILSVQAEEQKLI-SEEDLLRKRREQLKHKLEQLRNSC(FITC)A	N.A	16293.4	16292.7	50	1.47	0.07
His-MAX (MAX)	NH_2	MHHHHHHSSGLVPRGSGMKETAAAKFERQHMD-SPDLGTDDDDKAMADIGSDKRAHHNALERKRRD-HIKDSFHSLRDSVPSLQGEKASRAQILDKATEYIQYM-RRKNHTHQQDIDDLKRQNALLEQQVRALGGC	N.A	15203.71	15203.5	N.A	N.A	N.A
MAX_F	NH_2	MHHHHHHSSGLVPRGSGMKETAAAKFERQHMD-SPDLGTDDDDKAMADIGSDKRAHHNALERKRRD-HIKDSFHSLRDSVPSLQGEKASRAQILDKATEYIQYMR-RKNHTHQQDIDDLKRQNALLEQQVRALGGC(FITC)	N.A	15631.27	15630.8	67	3.64	0.2

Supporting table 5.2: Observed affinities according to BLI for Mizmetic WT and Mizmetic 1 analogues.

Mizmetic #	KD (μM)	KD (nM)	Times difference compared to WT
WT	4.599	4599.2	1
1	0.176	176	26

Supporting table 5.3: Observed affinities according to BLI for Mizmetic 1 and truncated peptides. Percentages where calculated by dividing the affinity by the affinity of Mizmetic 1.

Mizmetic #	KD (nM)	Times difference compared to WT
1	119.8	1
2	118.6	1
3	6007.4	50

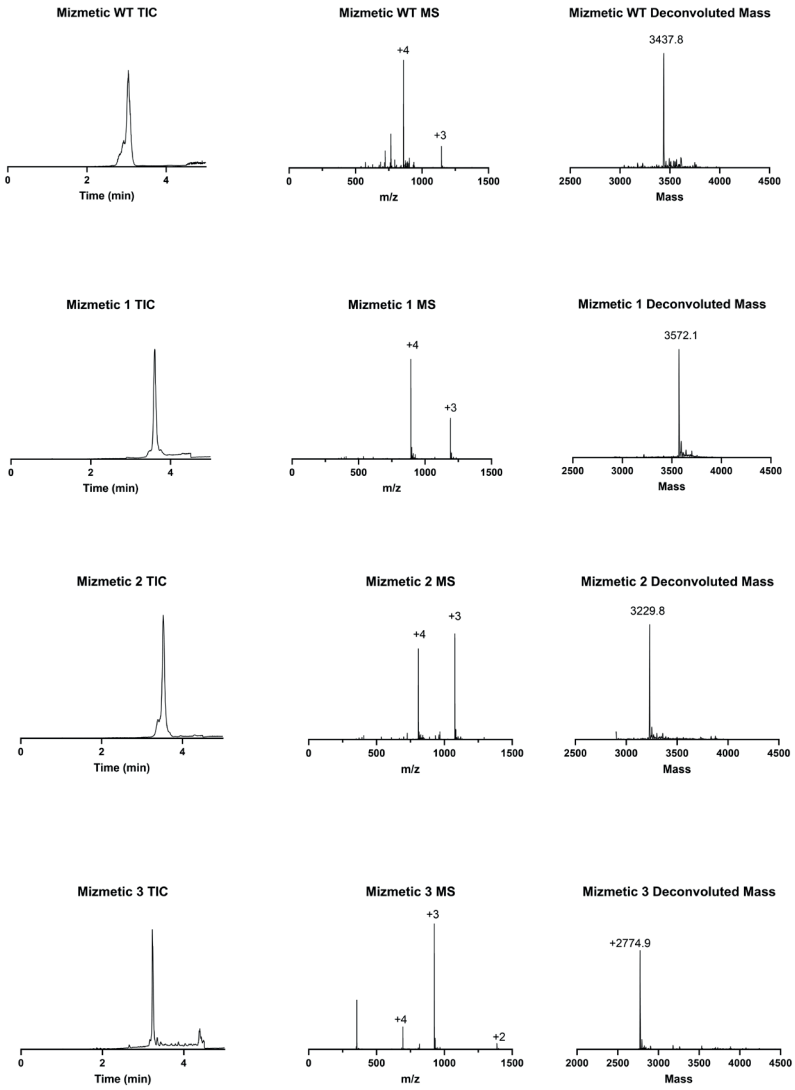
Supporting table 5.4 Observed affinities according to BLI for all Mizmetic 2 analogues compared to Mizmetic 2.

Mizmetic #	KD (nM)	Relative KD (%)
2	115.56	100%
4	118.72	97%
5	134.58	86%
6	159.91	72%
7	233.12	50%
8	167.29	69%
9	131.59	88%
10	184.54	63%
11	121.60	95%
12	529.56	22%
13	98.29	118%
14	1075.83	11%
15	161.49	72%
16	108.00	107%
17	38182.51	0%
18	146.00	79%
19	109.64	105%
20	99.49	116%

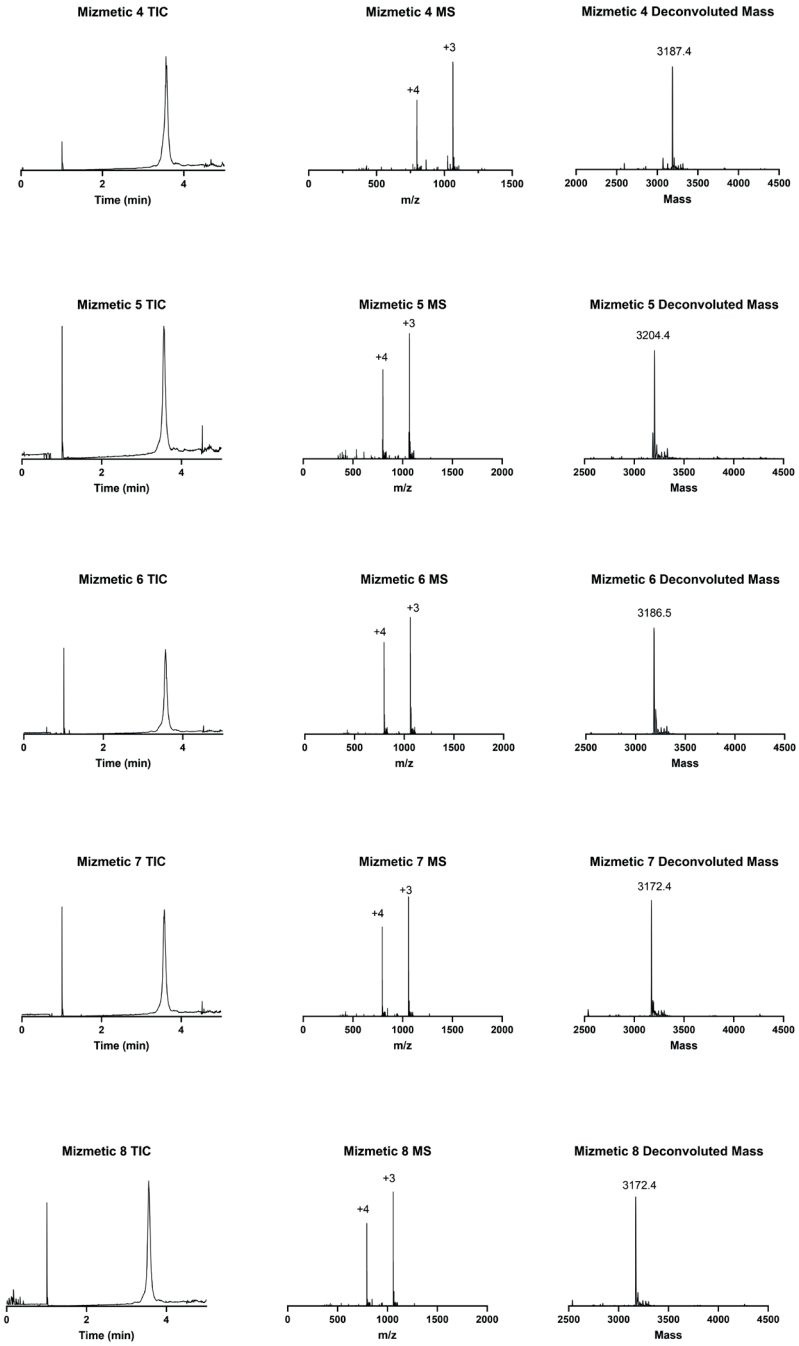
Supporting table 5.4 Observed affinities according to BLI for all Mizmetic 2 analogues compared to Mizmetic 2.

Mizmetic #	KD (nM)	Relative KD (%)
21	77.25	150%
22	171.03	68%

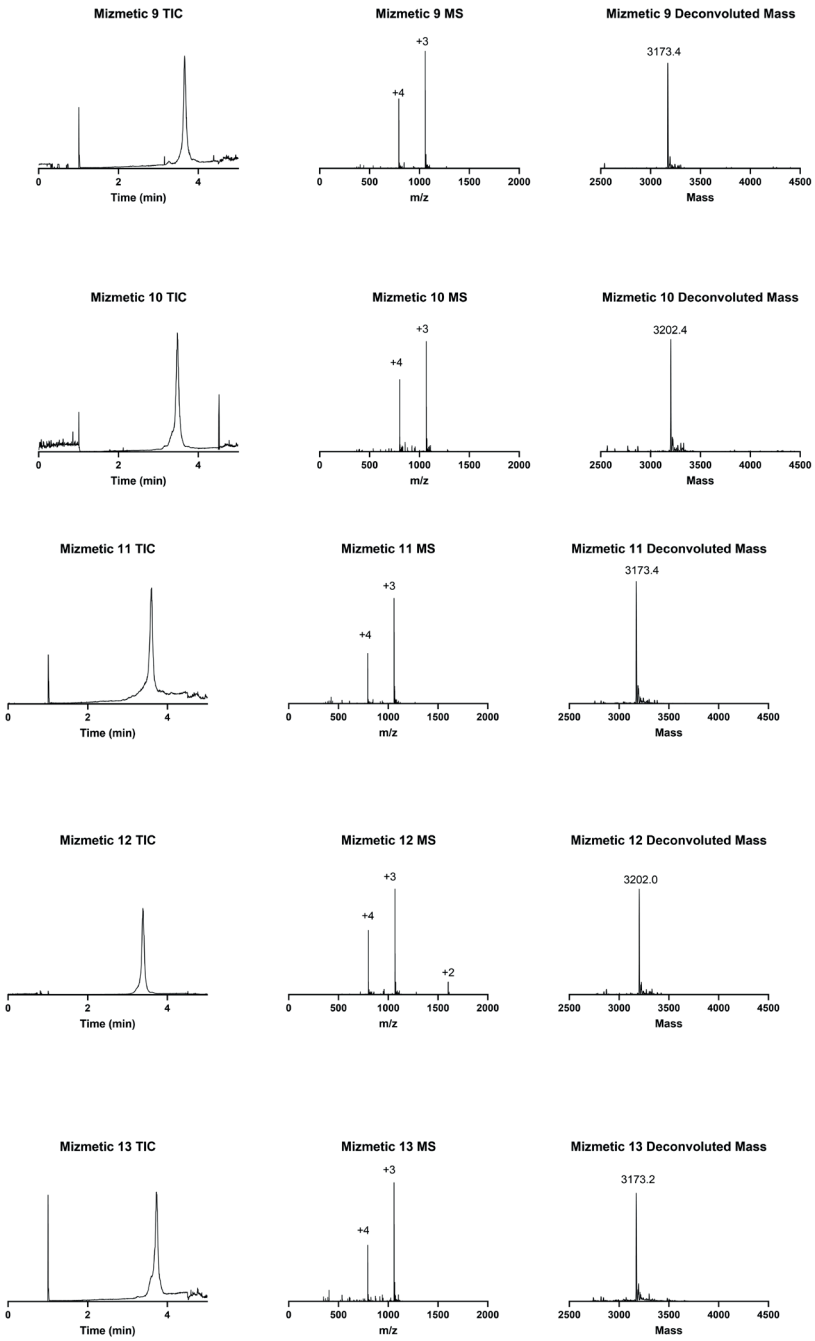
5.5.6 Supporting Figures



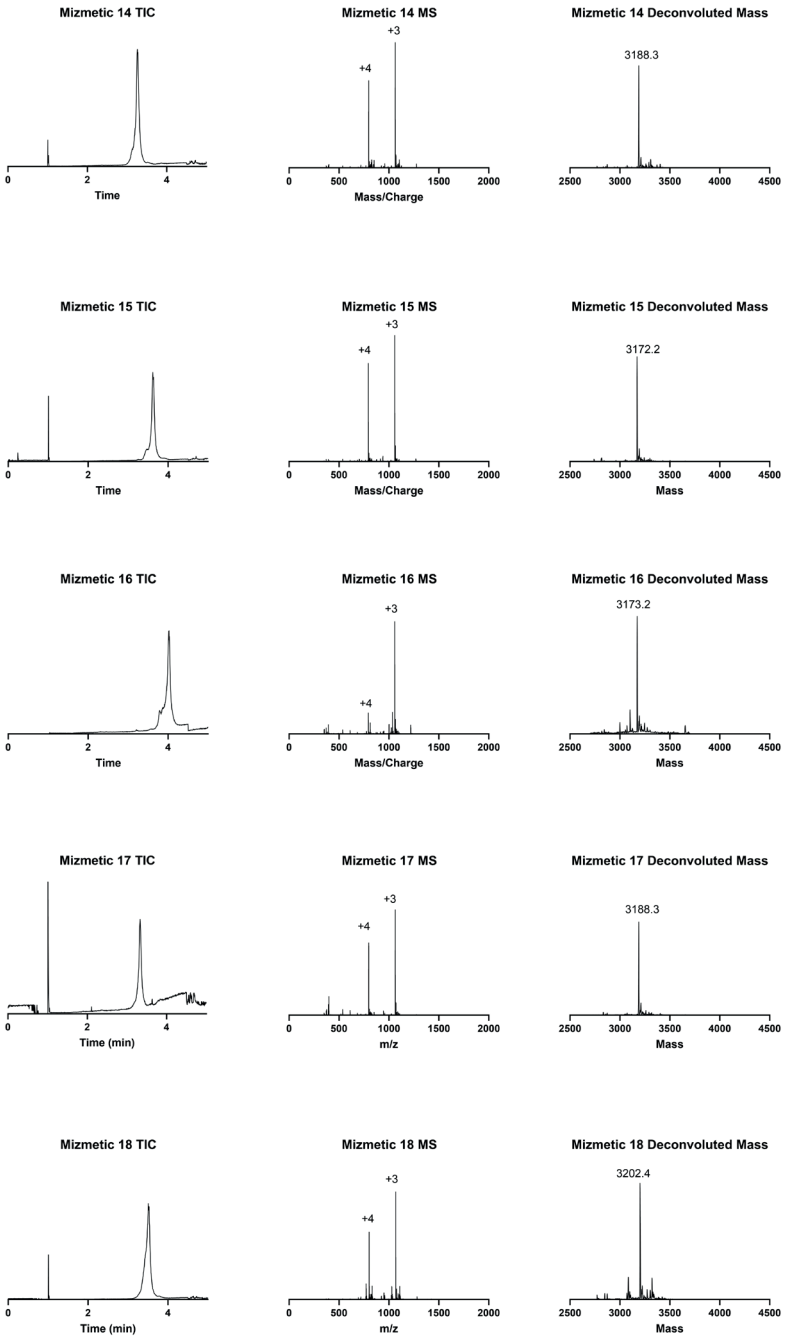
▲**Figure S5.1:** LC-MS traces of Mizmetic WT and 1-3. LC-MS traces were obtained with Method B.



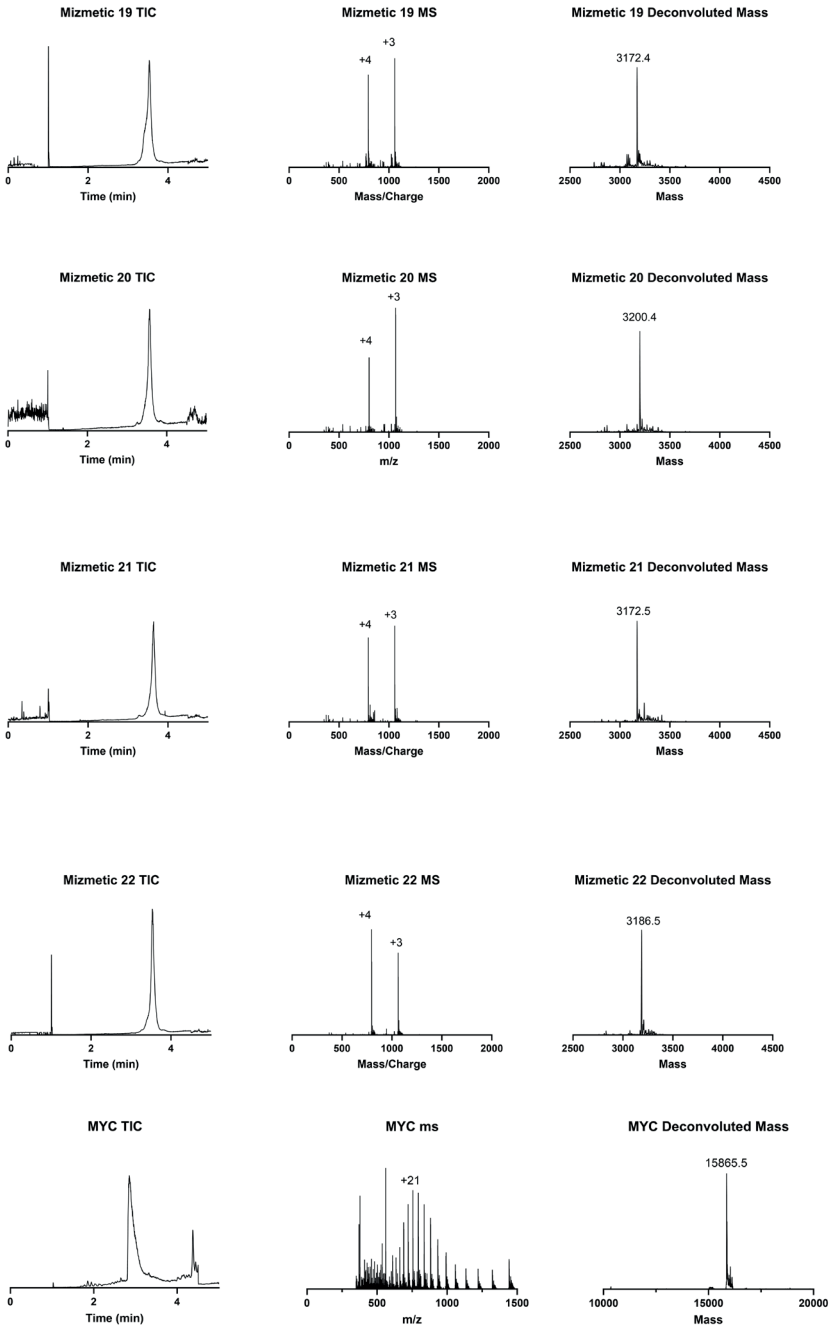
▲**Figure S5.2:** LC-MS traces of Mizmetic 4-8. LC-MS traces were obtained with Method B.



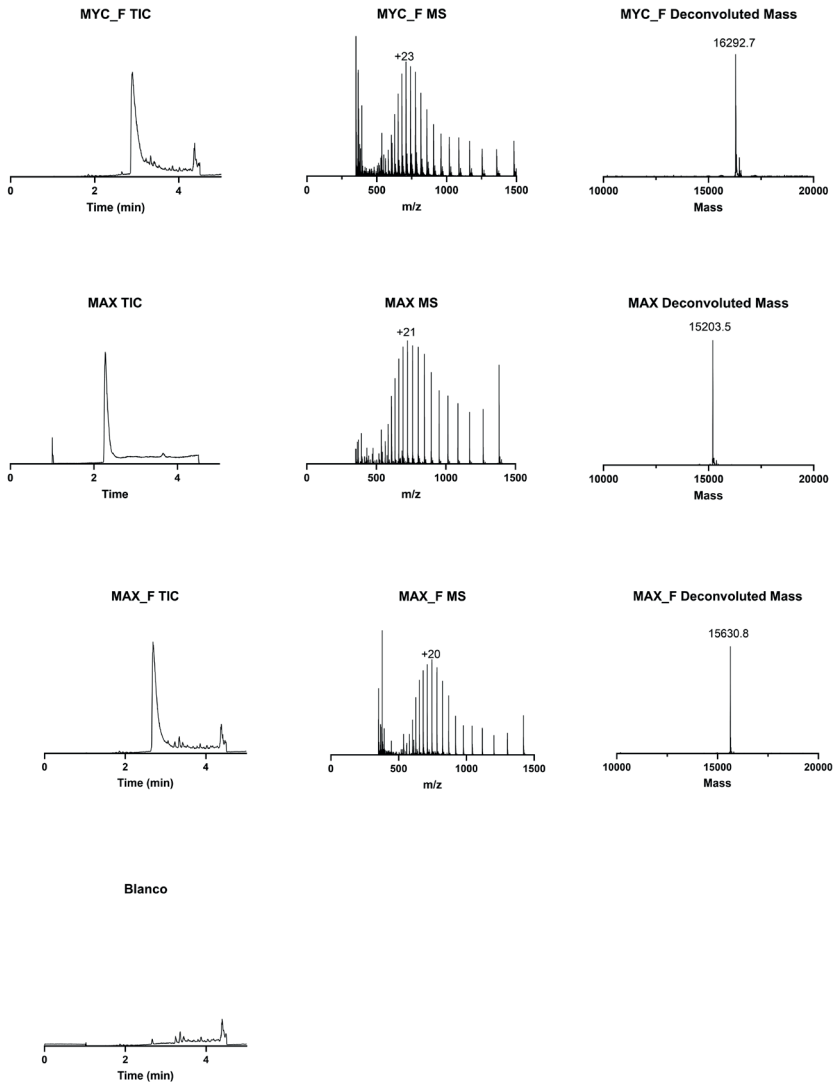
▲**Figure S5.3:** LC-MS traces of Mizmetic 9-13. LC-MS traces were obtained with Method B.



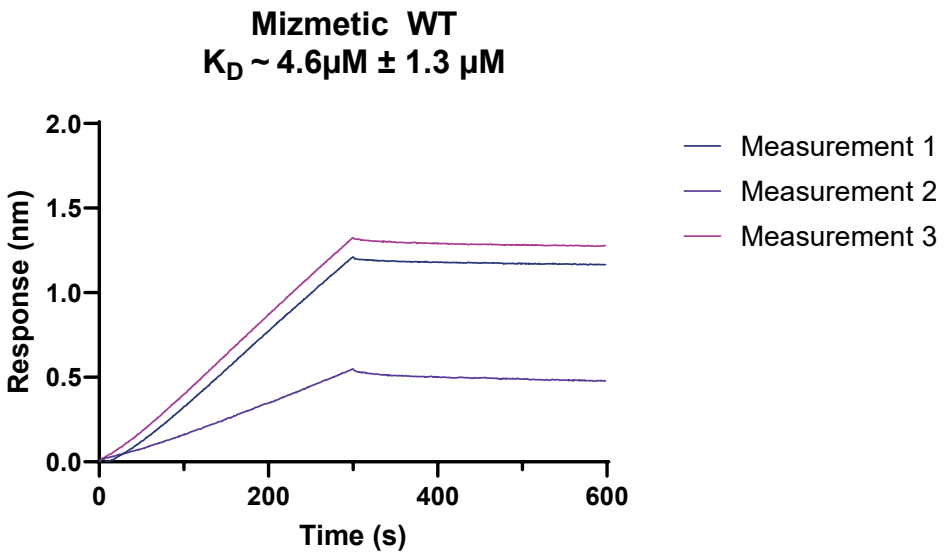
▲**Figure S5.4:** LC-MS traces of Mizmetic 14-18. LC-MS traces were obtained with Method B.



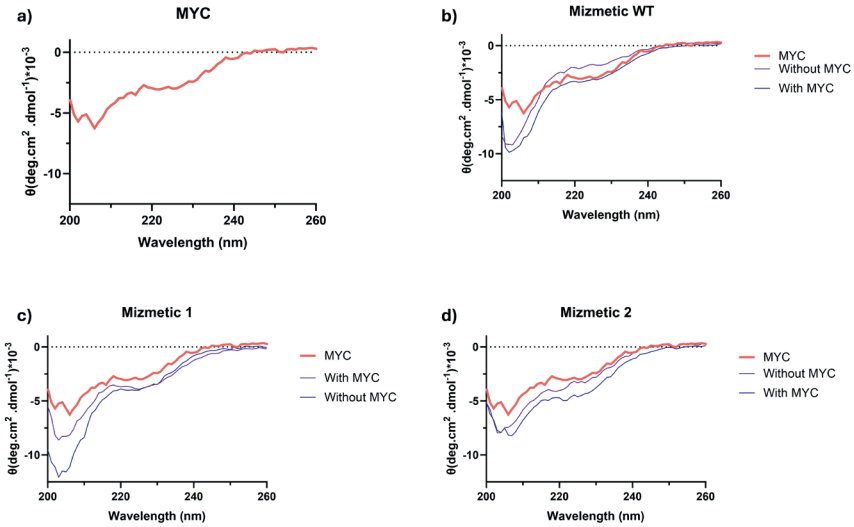
▲**Figure S5.5** LC-MS traces of Mizmetic 19-22, Myc and Myc-F. LC-MS traces were obtained with Method B.



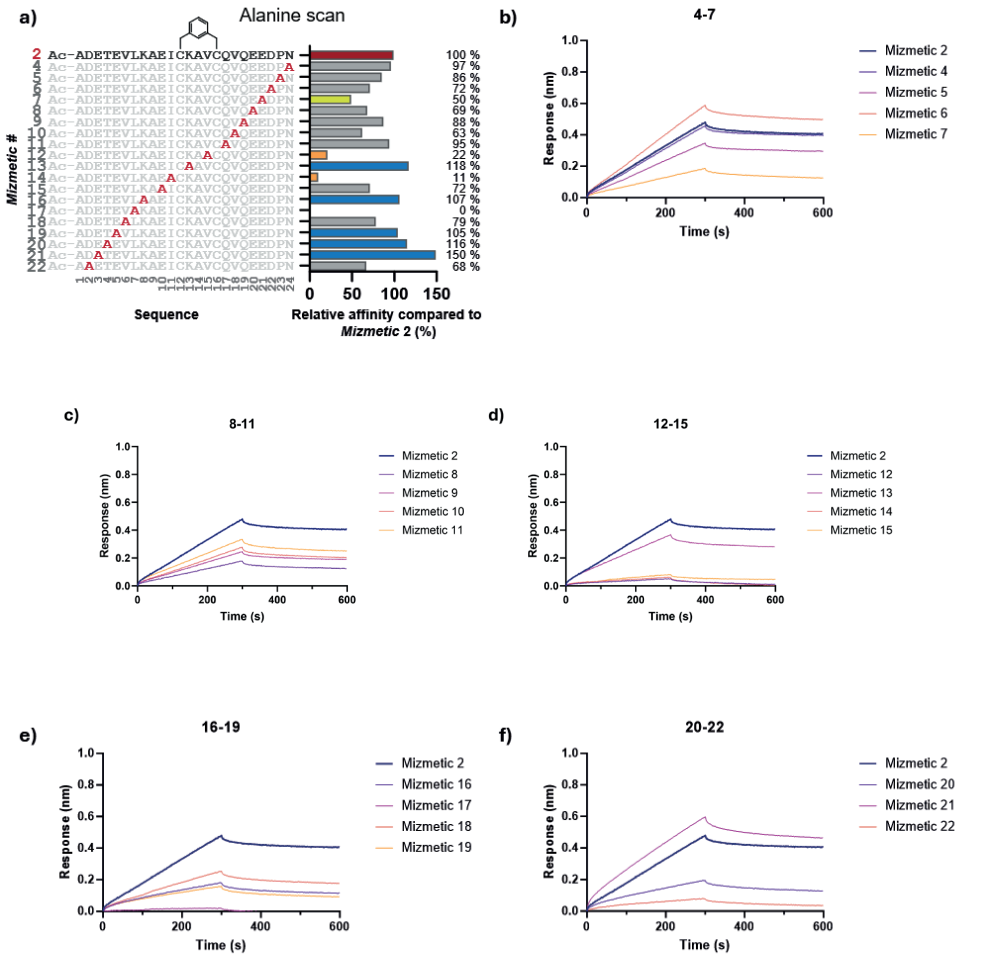
▲**Figure S5.6** LC-MS traces of MAX, Max-F and the blanco measured at the same day as MYC, MYC_F, MAX_F, Mizmetic 3. LC-MS traces were obtained with Method B.



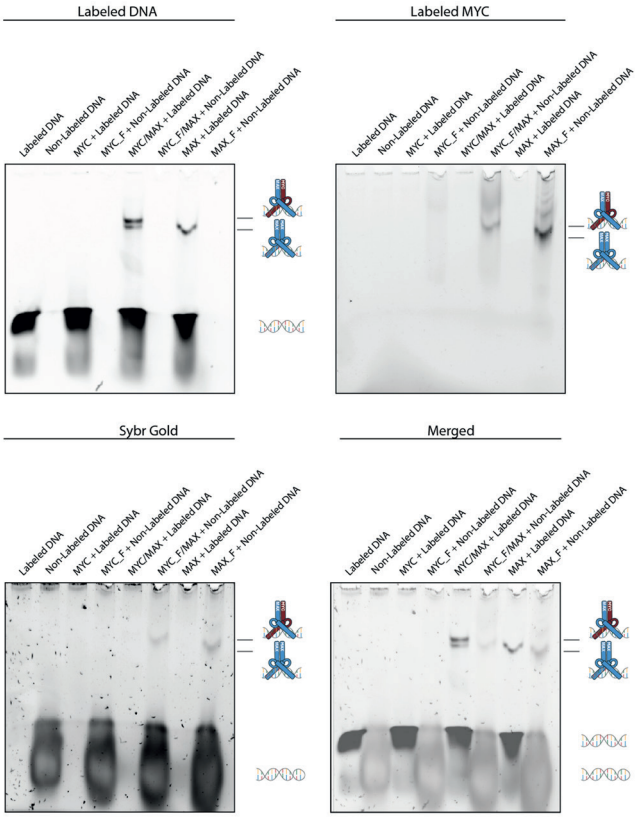
▲**Figure S5.7: BLI curves show that Mizmetics have affinity towards MYC.** Mizmetic WT displays an affinity of 4.6 μM ± 1.3 μM according to three independent BLI curves screened against 5 μM MYC.



▲**Figure S5.8: All Mizmetics display a partial α-helix with or without the presence of MYC.** a) CD spectrum of MYC without a Mizmetic. b) CD spectrum of Mizmetic WT with or without MYC. c) CD spectrum of Mizmetic 1 with or without MYC. d) CD spectrum of Mizmetic 2 with or without MYC. Lyophilized Mizmetics were diluted to 50 μM in 200 μL Tris acetate buffer (10 mM Tris acetate, 50 mM NaCl, pH = 7.4) with a Jasco J-815 at 37 °C. After the initial measurement, 10 μM MYC was added to the same cuvette and the cuvette was incubated for 10 minutes inside the machine at 37 °C before measurement.

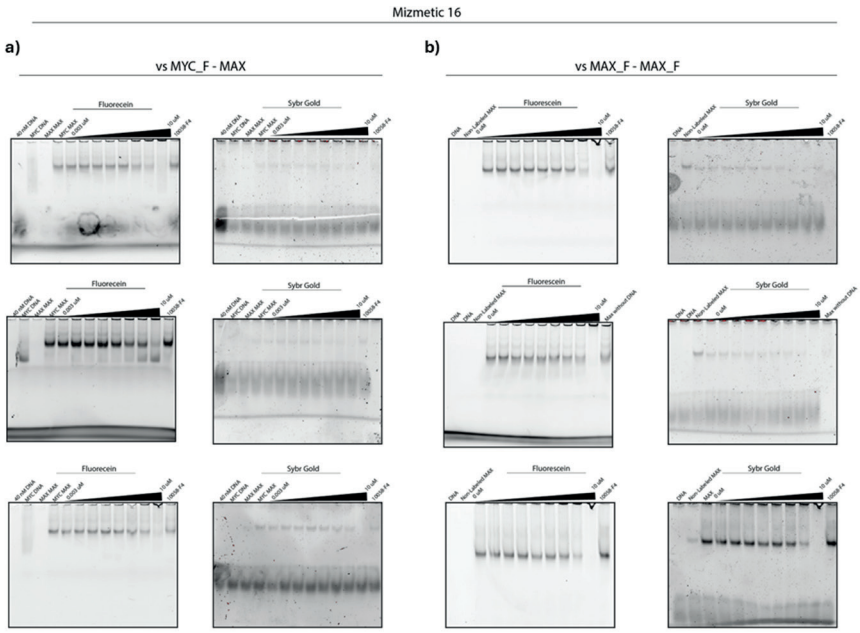


▲Figure S5.9: BLI curves display that Mizmetic 12, 14 and 17 have a decreased affinity towards MYC. Mizmetic 13, 16, 19-21 have increased affinity towards MYC. a) Comparison between affinities of alanine analogues of Mizmetic 2 compared to Mizmetic 2. b) BLI curves belonging to Mizmetic 2 and 4-7. c) BLI curves belonging to Mizmetic 2 and 8-11. d) BLI curves belonging to Mizmetic 2 and 12-15. e) BLI curves belonging to Mizmetic 2 and 16-19. f) BLI curves belonging to Mizmetic 2 and 20-22. All Mizmetics were screened against 400 nM of MYC.

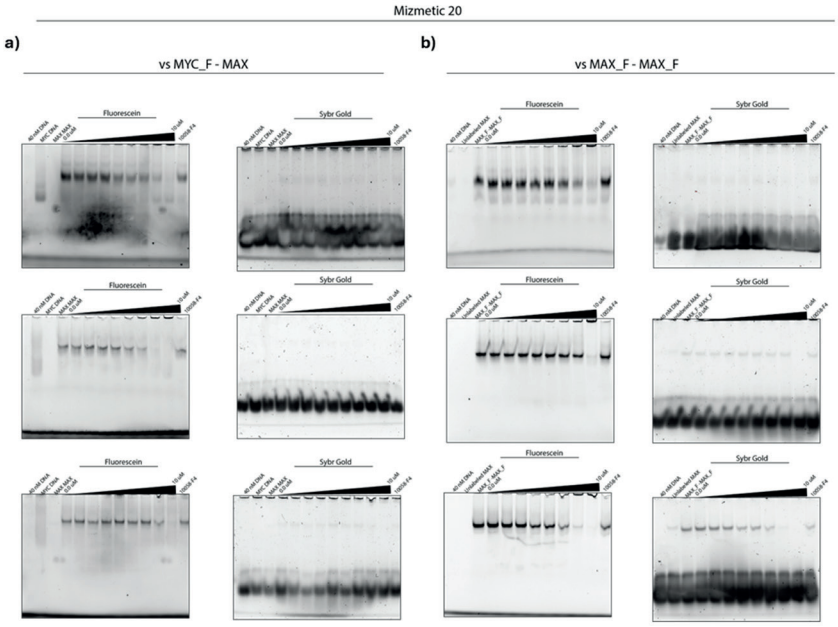


5

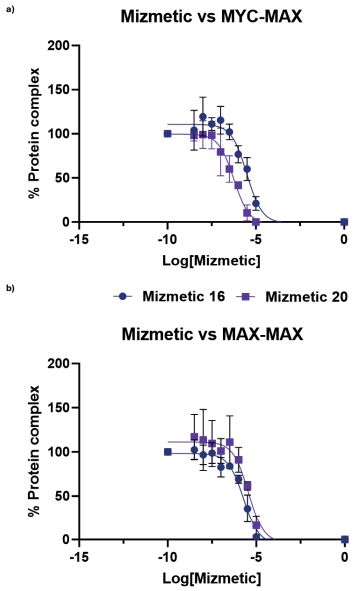
▲**Figure S5.10:** MYC/MAX complex is formed and visible with both labeled MYC_F or with labeled DNA.



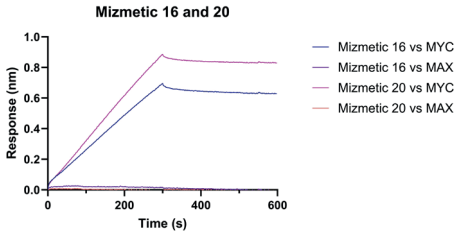
▲**Figure S5.11:** a) Concentration curve against MYC_F and MAX. An IC_{50} of 3.08 µM ± 1.94 µM was observed. b) Concentration curve against MYC_F and MAX. An IC_{50} of 1.94 µM ± 0.71 µM was observed. Mizmetics were stepwise diluted to 4x 10-0.003 µM in DMSO prior to incubation with 4x MYC_F (60 nM) in kinetic buffer. The mixture was incubated at RT for 1h followed by the addition of 4x MAX (60 nM) and 4x DNA (160 nM). The mixtures were incubated for 5 minutes and placed on ice while loading 12.5 µL on gel. Final concentrations Mizmetics (10-0.003 µM), MYC_F and MAX (both 15 nM), DNA (40 nM). Gels ran for 80 min at 75V at 4 °C and were imaged, followed by 15-40 min SYBR gold incubation to visualize DNA.



▲**Figure S5.12:** a) Concentration curve against MYC_F and MAX. An IC_{50} of 0.56 μM ± 0.22 μM was observed. b) Concentration curve against MYC_F and MAX. An IC_{50} of 2.25 μM ± 1.93 μM was observed. Mizmetics were stepwise diluted to 4x 10-0.003 μM in DMSO prior to incubation with 4x MYC_F (60 nM). The mixture was incubated at RT for 1h followed by the addition of 4x MAX (60 nM) and 4x DNA (160 nM). The mixtures were incubated for 5 minutes and placed on ice while loading 12.5 μL on gel. Final concentrations Mizmetics (10-0.003 μM), MYC_F and MAX (both 15 nM), DNA (40 nM). Gels ran for 80 min at 75V at 4 °C and were imaged, followed by 15-40 min SYBR gold incubation to visualize DNA.



▲Figure S5.13: Mizmetic alanine analogues can inhibit the formation of MYC-MAX or MAX-MAX. a) Quantification of concentration curves of Mizmetic 16 and 21 against MYC_F – MAX. b) Quantification of concentration curves of Mizmetic 16 and 21 against MAX_F – MAX. All experiments were performed at n = 3.



▲Figure S5.14: BLI curves show that Mizmetics have affinity towards MYC but not towards MAX. Mizmetics were screened against 1 μM MYC or MAX.

5.5.7 References

[1] B. D. Ellenbroek, J. P. Kahler, D. Arella, C. Lin, W. Jespers, E. A. Züger, M. Drukker, S. J. Pomplun, ""Development of DuoMYC: a synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC,"" *Angewandte Chemie International Edition* **2025**, *64*, , DOI 10.1002/anie.202416082.

[2] A. Micsonai, F. Wien, N. Murvai, M. P. Nyiri, B. Balatoni, Y.-H. Lee, T. Molnár, Y. Goto, F. Jamme, J. Kardos, ""BeStSel: analysis site for protein CD spectra—2025 update,"" *Nucleic Acids Res* **2013**, *1*, 13–14.

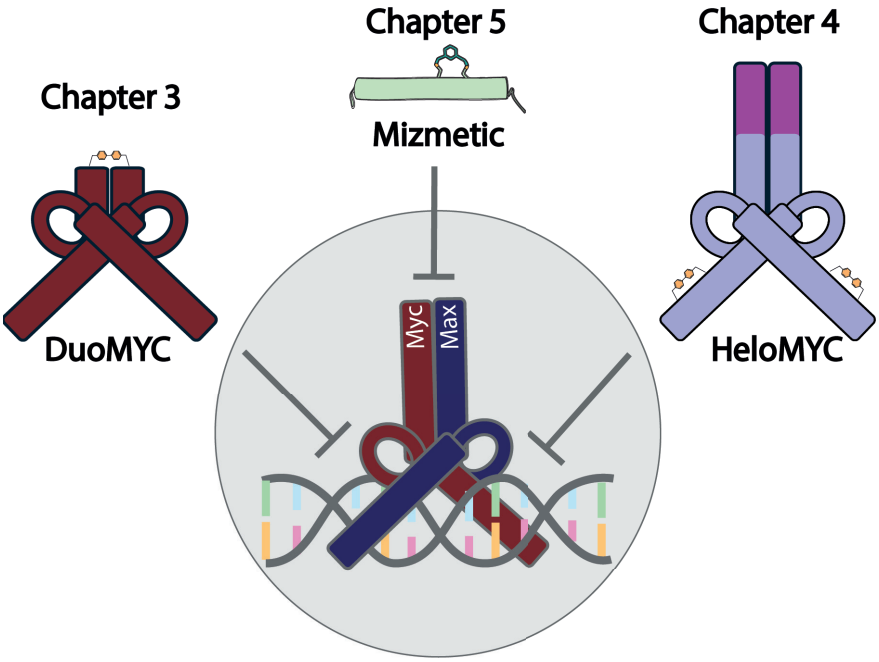
[3] A. Micsonai, É. Moussong, F. Wien, E. Boros, H. Vadászi, N. Murvai, Y. H. Lee, T. Molnár, M. Réfrégiers, Y. Goto, Á. Tantos, J. Kardos, ""BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy,"" *Nucleic Acids Res* **2022**, *50*, W90–W98.

[4] N. S. A. Crone, A. Kros, A. L. Boyle, ""Modulation of Coiled-Coil Binding Strength and Fusogenicity through Peptide Stapling,"" *Bioconjug Chem* **2020**, *31*, 834–843.

[5] M. J. Huang, Y. chih Cheng, C. R. Liu, S. Lin, H. E. Liu, ""A small-molecule c-Myc inhibitor, 10058-F4, induces cell-cycle arrest, apoptosis, and myeloid differentiation of human acute myeloid leukemia,"" *Exp Hematol* **2006**, *34*, 1480–1489.

Chapter 6

Conclusion and future perspectives



6

The main goal of this thesis is to *'Drop the Myc'*. We explored Myc inhibition using different approaches. Firstly, we developed a cell-permeable miniprotein that inhibits Myc activity by binding to Myc's designated DNA-binding site known as the Enhancer box (E-box, sequence: CACGTG). Secondly, aiming at a similar mechanism of action, we developed a method to stabilize and increase cell permeability of recombinant proteins, able to bind the E-box site. Lastly, AlphaFold was applied to predict an unresolved protein-protein interaction (PPI) that was utilized to design peptidomimetics that could prevent Myc-Max formation. All (mini)proteins and peptidomimetics were tested by a combination of biochemical characterization and cellular assays to analyze cell permeability, stability and affinity. This chapter summarizes the most important findings of each chapter and additionally highlights future possibilities for the design of Myc targeting peptide modalities.

6.1 Peptides are promising modalities for the inhibition of disease related nucleic acids

The central dogma of life entails that the genetic flow goes from DNA to RNA to proteins.[1] This process is strictly regulated by a network of PPIs, protein-DNA interactions and protein-RNA interactions. Disruption of this network can result in the development of diseases. Strikingly, approximately 19% of all human transcription factors (TFs) are involved in a disease, such as cancer.[2] Therefore, it is important to increase our understanding on how this class of proteins can be targeted. TF inhibition is particularly challenging because they are involved in protein-DNA interactions that often entail large interfaces and occur within the nucleus. The large interface limits the use of small molecules as they are not large enough. Consequently, alternative strategies are required to effectively target TFs.

Peptide-based modalities can be used to interact with large binding surfaces. **Chapter 2** describes the development and advantages of peptides that can target disease-related DNA and RNA. Peptides are advantageous compared to small molecules due to their larger size as this allows them to recognize large surface areas. Normally, an increase in molecular size can reduce cell permeability. However, nucleic acid (NA)-binding peptides often carry multiple positive charges that promote cell penetration, a characteristic also observed in cell-penetrating peptides. We observed two primary strategies for discovering specific NA-binding peptides: library screening and structure-based design. The latter strategy was applied in **Chapter 3** and **4**.

We highlight in **Chapter 2** that staple incorporation, or cyclisation, can be beneficial for cellular uptake, thereby locking NA-binding peptidomimetics in their active conformations. We observed similar effects in **Chapter 3** and **4**. In **Chapter 3** we could replace the full dimerization domain of Omomyc with a single covalent linker, thereby locking the protein in its dimeric state. Similarly, in **Chapter 4** we showed that staples in the basic region of Omomyc increased cell permeability and serum

stability. We observed a similar effect while designing the Mizmetic compounds in **Chapter 5.** Staple incorporation resulted in an increase in binding affinity towards Myc.

Selectivity is especially important for NA binders as off-target activity can have severe consequences for transcription and translation of genes. There are currently no FDA approved peptide-based drugs that target NAs but there are examples of NA binders that are in clinical trials, for example Omomyc-103.[3] Overall, peptide and protein mimetics can be promising modalities if they are carefully designed to selectively target a single NA sequence.

6.2 Natural TFs can serve as blueprints for creating artificial transcriptional repressors with improved drug likeness

In **Chapter 3,** we applied a structure-based design strategy on Omomyc to inhibit Myc. Omomyc is a basic Helix-Loop-Helix leucine zipper (bHLH-LZ) TF and needs to dimerize to act as a transcriptional repressor.[4] Cell permeability is a major challenge when designing miniproteins. Especially for Omomyc there is contradictory data regarding its intrinsic cell permeability. To address this, we designed several truncated Omomyc variants: MiniMYC 1, MiniMYC 3 and DuoMYC.

Our goal was to decrease the overall size of Omomyc, while increasing cell permeability and retaining DNA affinity. Both MiniMYCs only contained the basic region that were covalently linked to form a dimer. MiniMYC 1 and MiniMYC 3 were unable to selectively bind DNA. A more constrained decafluorbiphenyl linker displayed a moderate binding improvement compared to a flexible lysine linker but was not sufficient for a stable protein/DNA complex. Therefore, we designed DuoMYC, which also included the loop and a small segment of the LZ of Omomyc. DuoMYC displayed low nanomolar affinity towards its canonical E-box target sequence in two different binding assays. We incorporated a flexible and a constrained linker in DuoMYC. Strikingly, there was not a major difference in affinity between both DuoMYC constructs. This suggests that the additional loop is the primary contributor to accurate DNA binding, rather than linker rigidity.

The monomer of DuoMYC, MonoMYC, could also bind as a dimer to DNA before covalent dimerization. The incorporation of a covalent linker within DuoMYC increased affinity compared to its monomer. Whereas MonoMYC exists as a monomer in solution, DuoMYC exists as a dimer. The increase in affinity is presumably attributable to a reduced entropic penalty for DuoMYC, as MonoMYC must first form a dimer prior to DNA binding. DuoMYC could inhibit Myc (IC_{50} = 464 nM) and was cell permeable. Moreover, RNA sequencing displayed that DuoMYC influenced pathways involved in cancer.

This study emphasized that decreasing protein size can increase cell permeability. We showed that a major part of the protein can be removed but that there is a limit to how much can be removed. Too much removal can lead to loss of selectivity or function. Therefore, a careful design approach must be performed to find a cut-off in protein size with sufficient protein activity and increased cell penetration, as displayed by DuoMYC.

To improve cell penetration without decreasing protein size, we investigated two strategies to increase cell permeability of Omomyc in **Chapter 4**. The first approach entailed an increase in arginine content. Increase in arginine and thus positive charge, can lead to increased cell permeability.[5] The second approach introduced staples inside the basic region of Omomyc. Staples can amplify intrinsic cell permeability.[6,7] They enhance the overall lipophilicity and α-helicity of the peptide by promoting intramolecular hydrogen bonding within the backbone.[8] We applied the first approach on the LZ region and increased its arginine content. Additionally, we incorporated different staples at different positions inside the basic region. To obtain our designed Omomyc variants, we used recombinant expression and strategically placed cysteines inside the basic region.

Protein expression can be a delicate process to preserve complex protein structures. However, Omomyc can dimerize upon protein contact, even following denaturation. This is a characteristic treat shared among Myc- and Max-based TFs.[9–13] For example, denatured Myc was successfully used in our native binding assay in **Chapter 5**. Similarly, we could denature Omomyc during a buffer exchange, perform a two-component stapling reaction and, subsequently, test the compounds in a native binding assay. This approach resulted in the development of the ReCHEMbinent technique. Most designed Omomyc derivatives could bind to their canonical binding site, adopted an α-helical conformation and displayed increased stability when bound to DNA. Bioactivity assays showed that HeloMYC-1421 could inhibit Myc activity in both HEK293T and HeLa cells. Moreover, labeled HeloMYC-1421 displayed increased cell permeability compared to Omomyc. Overall, we demonstrated in **Chapter 4** that the characteristic ability of Omomyc to fold upon protein-partner contact can be exploited for a beneficial outcome.

6.3 AlphaFold predicts an interaction interface on Myc where Miz-1 and MAX overlap

In **Chapter 5** we developed a direct Myc inhibitor employing the AlphaFold-Multimer software.[14,15] Myc is difficult to target using a direct inhibitor because of its partially disordered nature in the absence of binding partners. We utilized AlphaFold-Multimer to predict how Myc and Myc interacting zinc finger 1 (Miz-1) interact, as no crystal structure is currently available. Miz-1 and Myc are together involved in cancer development but it is unknown how they physically interact.[16] AlphaFold-

Multimer predicted a sequence of 28 amino acids within Miz-1, located inside Myc interacting domain 2 (MID2), to interact with the HLH domain of Myc. As this was consistent with literature, we tested and validated that this small region displayed affinity towards Myc. Incorporation of a cysteine staple increased the affinity 26-fold. An alanine-scan revealed three hotspots and five beneficial mutations that were consistent with the AlphaFold prediction. AlphaFold predicted an interacting interface overlap between Max and Miz-1 on Myc. We therefore developed an activity assay to determine if our Mizmetic compounds could inhibit Myc-Max-DNA binding. Several alanine analogues could inhibit Myc-Max formation and Mizmetic 20 displayed increased specificity towards Myc over Max. Our Mizmetic compounds highlight that peptides can be good inhibitors for disordered proteins. Similar to what was discussed in **Chapter 2**, the large contact interface of peptides, compared to small molecules, is presumably the main driver of recognition of disordered Myc.

This project highlights the impact that artificial intelligence (AI) can have on research. We could design peptidomimetics based on an unresolved PPI without the presence of a crystal structure. It also demonstrates that AI and research can complement each other and that experimental validation remains important to confirm predictions. Especially for interactions that involve IDPs or IDRs. Despite the fact that AlphaFold is generally accurate, studies have reported that it becomes less reliable for regions exceeding 100 amino acids.[17]

To conclude, peptides are promising modalities to inhibit disease-related TFs. They can adopt multiple conformations and can be designed to target large NA interfaces that are able to assume complex secondary structures. The size of peptides can be challenging, mainly for cell permeability. Therefore, careful structure-based design strategies must be applied, which can be combined with novel AI-based tools, such as AlphaFold2.

6.4 Future perspectives

The main focus of this PhD thesis is the TF Myc. Myc is responsible for approximately 15% of all gene transcription in the cell.[18] Moreover, Myc engages in more than 50% of all human cancer types. Strikingly, there is currently no drug on the clinical market that can target this oncogenic TF directly. There are however, candidates that target Myc either directly (Omomyc-103)[3] or indirectly (CX-5461)[19] that are now in clinical trials. Small molecule CX-5461 prevents Myc expression and miniprotein Omomyc-103 functions as a transcriptional repressor. These two examples demonstrate that a protein previously considered 'undruggable' is becoming druggable, thanks to an improved understanding of Myc that allows for novel and innovative inhibition strategies. The compounds discussed in this thesis may attribute to novel strategies in Myc inhibition.

6.4.1 Development of a scaffold for NA recognition.

TFs are essential for healthy cells. Especially Myc participates in numerous vital cellular processes, such as cell differentiation.[20,21] In **Chapter 2** we provided an extensive overview on peptides that are designed to bind disease-related NAs. We highlighted the potential of peptides to target and recognize NA sequences with high precision. This includes secondary structural elements and the ability to discriminate between DNA or RNA sequences. The latter has been observed for Zinc Finger (ZF) motifs.[22] The general structure of a ZF is conserved but depending on the target, a ZF motif recognizes either DNA or RNA, for example ZF5 in TFIIIA.[23] This example highlights the complexity of NA binding. It brings forth the question 'How do we design artificial modalities that can recognize one specific DNA sequence?' To answer this question Barbas III identified rules for specific DNA recognition for ZFs[24–27] and Dervan for polyamides,[28] which have been applied for the design of artificial TFs.[29,30] More recently, the Moellering lab developed an innovative design based on bHLH-LZ[31] or bzipper[32] proteins that act as transcriptional repressors. Generally, TFs contain conserved structural folds that evolved to recognize DNA with remarkable specificity. Depending on the composition of the basic region, similar folded proteins can target different DNA sequences. Indeed, in most of the previous examples, a similar structure or scaffold was transferable to design artificial TFs that target different DNA sequences.

The DuoMYC scaffold can be utilized as a starting point for the design of new synthetic TFs. In theory, we could use DuoMYC's design on other bHLH-LZ TFs to target different DNA sequences. By utilizing known crystal structures of bHLH-LZs in complex with their canonical DNA sequence, DuoMYC's design and covalent linker can provide the basis for new artificial TFs. However, modifications can always result in loss of specificity, as was observed for the reduced Omomyc derivatives (MiniMYC, **Chapter 3**). Therefore, biochemical assays need to confirm specific DNA binding. Nevertheless, before DuoMYCs design can be utilized to design other TFs, its synthesis needs to be optimized.

6.4.2 Synthesis optimization

DuoMYC's synthesis resulted in a low yield, highlighting the need for further optimization if this scaffold is to be used for other artificial TFs or late-stage modifications. Studies demonstrated that artificial TFs targeting Myc can be synthesized using different approaches and that aggregation is a common challenge during synthesis.[11–13,20,31,33–39] Longer peptide sequences tend to aggregate or adopt secondary structures, reducing coupling efficiencies during synthesis. There are strategies to overcome aggregation during synthesis that include chaotropic salts, detergent, alternative solvents, temperature adjustments and different combinations of coupling reagents can improve solubility. [40,41] With the current optimized conditions,

we could try to implement chaotropic salts (e.g. urea or guanidine hydrochloride), detergents (e.g. Tween-20 or Triton X-100) or a mixture of solvents (e.g. addition of DMSO or trifluorethanol (TFE)) during each coupling step to prevent aggregation.

Another strategy to overcome aggregation is to use the SynTAG.[42] Bürgisser *et al.* demonstrated that, when combined with fast-flow peptide synthesis, the SynTag improved peptide solubility. This resulted in increased purity of typically difficult to synthesize peptides, including the TAD of Myc. The SynTAG combined with microwave-assisted peptide synthesis may offer a beneficial alternative approach due to shorter reaction times and generally improved yields compared to automated solid phase peptide synthesis (SPPS). Temperature can significantly influence side reactions that occur during microwave-assisted peptide synthesis, such as aspartic acid epimerization or δ-lactam formation in ariginines. Therefore, examining a temperature range could be beneficial when synthesizing DuoMYC using this approach.

Additionally, ReCHEMbinant expression could be investigated. Earlier attempts to recombinantly express DuoMYC were not viable. We were unable to remove the expression tag, presumably due to its disordered nature. This eventually limited dimerization and cell permeability. Previously, it has been demonstrated that urea combined with enterokinase could facilitate specific tag cleavage for a similar protein.[43] Therefore, urea addition may be an initial step in optimizing DuoMYC tag cleavage. Successful DuoMYC expression and tag cleavage would enable dimerization and thus the ReCHEMbinant platform.

Lastly, the Pentelute lab recently developed a strategy to synthesize proteins with SPPS followed by refolding techniques that promoted the most stable conformation.[44] Whereas incorrectly folded proteins tended to precipitate, correctly folded proteins did not. As such, their approach could be applied as an additional purification step for complicated syntheses, such as DuoMYC. However, this technique relies on protein folding and must be optimized for DuoMYC. To promote correctly folded DuoMYC, additives, such as TFE and DNA, may be incorporated to promote secondary structure formation. TFE is a helix stabilizer and is often used in circular dichroism experiments to promote helicity.[40] Similarly, we observed that DNA stabilized the overall protein structure of Omomyc in **Chapter 4**. In both examples, correctly folded DuoMYC should remain in solution, whereas misfolded proteins should precipitate out of solution.

6.4.3 Combining domains and incorporating staples may improve conformation and cell penetration

A challenge in artificial TFs design is their conformation and cell permeability. It is noteworthy that TFs contain conserved DNA binding motifs that can facilitate DNA

binding and cell permeability.[5] For example, isolated ZFs[45] and bHLH-LZs[11,13] are intrinsically cell permeable but sometimes a cell-penetrating peptide can be beneficial.[46,47] Nonetheless, removing a protein region from its native environment can lead to decreased or loss of function, as the active configuration is no longer sustained by the overall protein structure. Stabilizing the active conformation can increase specificity and cell penetration. There are approaches to guide peptides into the desired topology, such as the approach earlier described by the Pentelute lab.[44]

Combining multiple DNA binding domains can promote specificity and correct conformation.[48,49] As observed in **Chapter 3**, the incorporation of a covalent linker between two MonoMYC compounds increased affinity ~8 fold. To assess if the DuoMYC scaffold is transferable to other bHLH-LZ TFs, Myc-Max or Max-Max DuoMYC constructs would be a rational initial approach. Earlier work demonstrated that covalent Myc-Max and Max-Max dimers can bind E-box DNA.[11,13] Additionally, these constructs may have enhanced stability and helicity compared to DuoMYC, as these complexes naturally occur in cells.

TFs often bind DNA in an α-helical conformation that allows for the incorporation of staples. Staples increase helicity, protease resistance and cell permeability. Helicity is observed in Omomyc,[50] Myc-Max[51] and Max-Max[51] crystal structures in complex with their canonical binding site. Nonetheless, the basic region may adopt multiple conformations in the absence of DNA. In **Chapter 4** and literature it has been demonstrated that staples can be beneficial when incorporated into the basic region of Omomyc and Max.[12,31,52] As such, staples can also be introduced in the basic region of DuoMYC. However, as both MonoMYC dimerization and cysteine stapling rely on CLIPS™[53,54] chemistry, two synthesis strategies may be considered: performing stapling on solid support, or designing a route based on orthogonal protection groups. The latter allows for in-solution stapling and dimerization. For example, cysteine, protected with an acetamidomethyl, could be incorporated at the C-terminus for dimerization and two cysteines, protected with an acid labile trityl group, could be incorporated at the staple positions. Using this route the monomer is first stapled and subsequently dimerized after the removal of the acetamidomethyl group, all in solution. A similar strategy was applied by the Jbara group to staple Max. [52]

Helicity is also essential for Mizmetic peptides to bind Myc. The impact of different kind of staples, different size, different positions and multiple incorporated staples can be assessed in a follow-up study. Recently, Chandramohan *et al.* identified design rules for stapled peptides and suggested that an increase in hydrophobicity and rigidity is beneficial for in vivo activity.[8] All designed compounds could be assessed for helicity, Myc binding, cell penetration and Myc/Max inhibition.

6.4.4 Combining ReCHEMbinant with AI can be a powerful approach

A major difficulty when studying Myc or other TFs is their disordered nature. Whereas their disordered regions adopt a secondary conformation upon protein and DNA contact, they remain largely disordered without protein partners. This disordered character mitigates the identification of defined binding regions due to impaired structure elucidation. Before the development of AI tools, such as AlphaFold2[14] or RFDiffusion,[55] a significant part of the human proteome (~22% is IDP or IDR)[56] could not be visualized as crystal structures were not fully resolved. With AI we can predict with reasonable reliability protein structures based on their primary sequence. AI-based structure prediction software, such as RoseTTAFold[57] and AlphaFold,[14] are applied in the development of de novo protein design[55,58] and TF design.[59–61] The Baker lab are pioneers in de novo protein design. Recently, they developed a pipeline for the development of de novo designed IDR binding proteins.[55] AlphaFold on the other hand demonstrated to be a powerful tool as well, regarding IDRs.[17] AlphaFold2 allows for the incorporation of DNA which can be used for the development of artificial TF design. IDRs are extremely versatile and can adopt defined structures when they interact with protein partners.[62] Recently, a fragment scanning approach with AlphaFold2, similar to our work in **Chapter 5**, was successful in predicting IDR conformations.[17]

Using AlphaFold2 we may predict how IDRs interact with proteins. IDRs often interact with short motifs[63] that adopt secondary structures that can be stabilized through strategically placed staples. For regions larger than 60 residues the ReCHEMbinant technique may be applied. As discussed in **Chapter 4**, the ReCHEMbinant pipeline allows for the expression, purification and late-stage modification of proteins that fold upon contact. Depending on structure prediction, α-helices or β-sheets can be stabilized though the incorporation of a covalent constraint.[64,65] For example, we could apply the ReCHEMbinant approach on full length MID2. This should increase MID2 stability and affinity based on our findings in **Chapter 4** and **5**. Examples in the PDB highlight that IDPs can be crystallized when they are stabilized by protein partners (e.g. Myc-MaxDNA complex [51]). As the ReCHEMbinant pipeline produces stabilized structures this could hypothetically be utilized in combination with crystallization studies.

6.4.5 Other strategies to inhibit Myc

We demonstrated that the Myc-Max complex can be inhibited by a peptide identified with AlphaFold2 through an unresolved PPI first identified by Peukert *et al.* in 1997.[66] This fragment-based AI approach may also be suitable to target Myc in its complete disordered region. Myc contains a large, disordered region that with conserved Myc Homology boxes (MB).[67] The interaction between Myc and Max has been widely studied and most Myc inhibiting strategies either try to prevent DNA

binding or Myc-Max formation. For example, Struntz et al. developed a Max-Max stabilizer[68] Another approach would be to target the MBs. MBs regulate correct DNA recognition, PPIs and increase Myc stability.[69,70] There are small molecules identified that inhibit MB-protein interactions that would otherwise stabilize Myc, such as Aurora A.[71] By performing AlphaFold2 predictions with protein partners of Myc and isolated MBs, possible binding modes can be identified that could result into MB-based Myc inhibitors, such as our approach in **Chapter 5**.

Library designs are powerful for the identification of peptide-based NA binders.[72] A variety of library techniques have been used to identify new peptide modalities that target NA, such as phage display and combinatorial libraries.[73] A novel direct evolution technique called phage-assisted continuous evolution (PACE) was developed in 2011.[74] This technique utilizes the constant selection of the bacteriophage while promoting directed evolution regulated by the activity of interest.[74,75] This strategy was successfully implemented in combination with rational design on the structure of ME47.[76] ME47, similar to DuoMYC, is an artificial TF designed to target the E-box sequence.[77] PACE is a powerful technique that allows for quick protein evolution, which could be used to increase DuoMYC affinity to E-box DNA[76] or to repurpose it to target different disease-related NAs.[78] Similarly, it can be utilized to increase selectivity toward Myc when the Mizmetics are used as a template.[79]

Phage display can also be used to screen for macrocyclic peptides[80] or stapled peptides.[81] For example, recently the group of Barniol-Xicota developed a method to utilize phage display to screen for macrocyclic peptides through the incorporation and modification of strategically placed cysteines.[80] We could utilize stapled phage display to screen for stapled Mizmetic- or MB-based peptides that could function as a Myc inhibitor, provided that these peptides interact as an α-helix or β-sheet according to AlphaFold or other AI tools.

Lastly, covalent inhibition is a promising approach to target Myc. Covalent inhibitors, such as EN4,[82] form a covalent bond between reactive residues within the target. These residues often involve cysteines but other amino acids, such as lysine,[83] can be targeted.[84] The Mizmetic alanine analogues discussed in **Chapter 5** may be suitable as a starting point for a covalent inhibitor design. AlphaFold predicts Glu21 in Mizmetic WT in close proximity to Lys392 in Myc. Lysines are often utilized for their reactivity for a covalent inhibitor design.[85,86]The close proximity between Glu21 and Lys392 allows for the incorporation of a reactive warhead within Mizmetic 20. Resembling an earlier described staple protocol for DuoMYC, orthogonal protection groups can be utilized to first staple the Mizmetics on solid support and subsequently attach a warhead on position 21. An initial approach could incorporate Cys(Mmt). The Mmt group can be removed under mild acidic conditions, enabling on solid support stapling. At position 21 a Lys(alloc) could be incorporated. The alloc protection group can be removed after stapling and a reactive warhead can be attached to the amine

group of Lysine. The obtained compounds can be analyzed for covalent attachment using SDS-PAGE. Once a selective inhibitor is obtained, these can be evaluated using our developed native assay to observe Myc-Max inhibition.

6.5 Final note

In this thesis the inhibition of Myc was explored through two different strategies. Two different design strategies were investigated for the development of artificial TFs either with SPPS or recombinant expression. Additionally, in combination with AI, a structure activity relationship study was performed to design a new Myc/Max inhibitor based on an unresolved PPI. This work explored the field of protein design in combination with late-stage peptide modification that allowed for the replacement of a complete dimerization domain by a small covalent constraint. Additionally, this thesis tried to highlight the potential of peptides in the struggle against Myc. Altogether we hope that our work will be beneficial in the future development of artificial TFs, characterization of IDPs and IDRs and the complete field in protein architecture.

6.6 References

[1] F. Crick, "Central Dogma of Molecular Biology" *Nature* **1970**, *227*, 1970.

[2] 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, "The Human Transcription Factors" *Cell* **2018**, *172*, 650–665.

[3] E. Garralda, M. E. Beaulieu, V. Moreno, S. Casacuberta-Serra, S. Martínez-Martín, L. Foradada, G. Alonso, D. Massó-Vallés, S. López-Estévez, T. Jauset, E. Corral de la Fuente, B. Doger, T. Hernández, R. Perez-Lopez, O. Arqués, V. Castillo Cano, J. Morales, J. R. Whitfield, M. Niewel, L. Soucek, E. Calvo, "MYC targeting by OMO-103 in solid tumors: a phase 1 trial" *Nat Med* **2024**, *30*, 762–771.

[4] C. Murre, "Helix–loop–helix proteins and the advent of cellular diversity: 30 years of discovery" *Genes Dev* **2019**, *33*, 6–25.

[5] M. E. Beaulieu, T. Jauset, D. Massó-Vallés, S. Martínez-Martín, P. Rahl, L. Maltais, M. F. Zacarias-Fluck, S. Casacuberta-Serra, E. S. Del Pozo, C. Fiore, L. Foradada, V. C. Cano, M. Sánchez-Hervás, M. Guenther, E. R. Sanz, M. Oteo, C. Tremblay, G. Martín, D. Letourneau, M. Montagne, M. Á. M. Alonso, J. R. Whitfield, P. Lavigne, L. Soucek, "Intrinsic cell-penetrating activity propels omomyc from proof of concept to viable anti-myc therapy" *Sci Transl Med* **2019**, *11*, DOI 10.1126/scitranslmed.aar5012.

[6] H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. DeGrado, D. C. Greenbaum, "Development of α-helical calpain probes by mimicking a natural protein-protein interaction." *J Am Chem Soc* **2012**, *134*, 17704–17713.

[7] Y. H. Lau, P. De Andrade, Y. Wu, D. R. Spring, "Peptide stapling techniques based on different macrocyclisation chemistries" *Chem Soc Rev* **2015**, *44*, 91–102.

[8] A. Chandramohan, H. Josien, T. Y. Yuen, R. Duggal, D. Spiegelberg, L. Yan, Y.-C. A. Juang, L. Ge, P. G. Aronica, H. Y. K. Kaan, Y. H. Lim, A. Peier, B. Sherborne, J. Hochman, S. Lin, K. Biswas, M. Nestor, C. S. Verma, D. P. Lane, T. K. Sawyer, R. Garbaccio, B. Henry, S. Kannan, C. J. Brown, C. W. Johannes, A. W. Partridge, "Design-rules for stapled peptides with in vivo activity and their application to Mdm2/X antagonists" *Nat Commun* **2024**, *15*, 489.

[9] O. Harel, M. Jbara, "Chemical Synthesis of Bioactive Proteins" *Angewandte Chemie* **2023**, *135*, DOI 10.1002/ange.202217716.

[10] R. V. Nithun, Y. M. Yao, O. Harel, S. Habiballah, A. Afek, M. Jbara, "Site-Specific Acetylation of the Transcription Factor Protein Max Modulates Its DNA Binding Activity" *ACS Cent Sci* **2024**, *2024*, 1295–1303.

[11] S. Pomplun, M. Jbara, C. K. Schissel, S. Wilson Hawken, A. Boija, C. Li, I. Klein, B. L. Pentelute, "Parallel Automated Flow Synthesis of Covalent Protein Complexes That Can Inhibit MYC-Driven Transcription." *ACS Cent Sci* **2021**, *7*, 1408–1418.

[12] O. Harel, ‖ Ferran Nadal-Bufi, R. V Nithun, Y. M. Yao, A. Afek, M. Vendrell, M. Jbara, "Chemical Engineering of Transcription Factors Uncovered Cell-Permeable μMax Modulators" *J Am Chem Soc* **2025**, *15*, DOI 10.1021/JACS.5C13964.

[13] M. Jbara, S. Pomplun, C. K. Schissel, S. W. Hawken, A. Boija, I. Klein, J. Rodriguez, S. L. Buchwald, B. L. Pentelute, "Engineering Bioactive Dimeric Transcription Factor Analogs via Palladium Rebound Reagents." *J Am Chem Soc* **2021**, *143*, 11788–11798.

[14] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, "Highly accurate protein structure prediction with AlphaFold." *Nature* **2021**, *596*, 583–589.

[15] A. Omidi, M. H. Møller, N. Malhis, J. M. Bui, J. Gsponer, "AlphaFold-Multimer accurately captures interactions and dynamics of intrinsically disordered protein regions." *Proc Natl Acad Sci U S A* **2024**, *121*, e2406407121.

[16] S. Peter, J. Bultinck, K. Myant, L. A. Jaenicke, S. Walz, J. Müller, M. Gmachl, M. Treu, G. Boehmelt, C. P. Ade, W. Schmitz, A. Wiegering, C. Otto, N. Popov, O. Sansom, N. Kraut, M. Eilers, "Tumor cell-specific inhibition of MYC function using small molecule inhib-

itors of the HUWE1 ubiquitin ligase." *EMBO Mol Med* **2014**, *6*, 1525–1541.

[17] H. Bret, J. Gao, D. J. Zea, J. Andreani, R. Guerois, "From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2." *Nat Commun* **2024**, *15*, 597.

[18] C. V. Dang, K. A. O'Donnell, K. I. Zeller, T. Nguyen, R. C. Osthus, F. Li, "The c-Myc target gene network" *Semin Cancer Biol* **2006**, *16*, 253–264.

[19] G. C. C. Koh, S. Boushaki, S. J. Zhao, A. M. Pregnall, F. Sadiyah, C. Badja, Y. Memari, I. Georgakopoulos-Soares, S. Nik-Zainal, "The chemotherapeutic drug CX-5461 is a potent mutagen in cultured human cells" *Nature Genetics 2023 56:1* **2023**, *56*, 23–26.

[20] Z. Deng, D. R. Richardson, "The Myc Family and the Metastasis Suppressor NDRG1: Targeting Key Molecular Interactions with Innovative Therapeutics" *Pharmacol Rev* **2023**, *75*, 1007–1035.

[21] S. K. Madden, A. D. de Araujo, M. Gerhardt, D. P. Fairlie, J. M. Mason, "Taking the Myc out of cancer: toward therapeutic strategies to directly inhibit c-Myc" *Molecular Cancer 2021 20:1* **2021**, *20*, 1–18.

[22] M. Fu, P. J. Blackshear, "RNA-binding proteins in immune regulation: A focus on CCCH zinc finger proteins" *Nat Rev Immunol* **2017**, *17*, 130–143.

[23] T. M. T. Hall, "Multiple modes of RNA recognition by zinc finger proteins" *Curr Opin Struct Biol* **2005**, *15*, 367–373.

[24] D. J. Segal, B. Dreier, R. R. Beerli, C. F. Barbas, "Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5′-GNN-3′ DNA target sequences" *Proc Natl Acad Sci U S A* **1999**, *96*, 2758–2763.

[25] B. Dreier, R. P. Fuller, D. J. Segal, C. V. Lund, P. Blancafort, A. Huber, B. Koksch, C. F. Barbas, "Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors" *Journal of Biological Chemistry* **2005**, *280*, 35588–35597.

[26] B. Dreier, R. R. Beerli, D. J. Segal, J. D. Flippin, C. F. Barbas, "Development of Zinc Finger Domains for Recognition of the 5′-ANN-3′ Family of DNA Sequences and Their Use in the Construction of Artificial Transcription Factors" *Journal of Biological Chemistry* **2001**, *276*, 29466–29478.

[27] B. Dreier, D. J. Segal, C. F. Barbas, "Insights into the molecular recognition of the 5′-GNN-3′ family of DNA sequences by zinc finger domains" *J Mol Biol* **2000**, *303*, 489–502.

[28] P. B. Dervan, B. S. Edelson, "Recognition of the DNA minor groove by pyrrole-imidazole polyamides" *Curr Opin Struct Biol* **2003**, *13*, 284–299.

[29] E. A. Heiderscheit, A. Eguchi, M. C. Spurgat, A. Z. Ansari, "Reprogramming cell fate with artificial transcription factors" *FEBS Lett* **2018**, *592*, 888–900.

[30] A. Eguchi, M. J. Wleklinski, M. C. Spurgat, E. A. Heiderscheit, A. S. Kropornicka, C. K. Vu, D. Bhimsaria, S. A. Swanson, R. Stewart, P. Ramanathan, T. J. Kamp, I. Slukvin, J. A. Thomson, J. R. Dutton, A. Z. Ansari, "Reprogramming cell fate with a genome-scale library of artificial transcription factors" *Proc Natl Acad Sci U S A* **2016**, *113*, E8257–E8266.

[31] T. E. Speltz, Z. Qiao, C. S. Swenson, X. Shangguan, J. S. Coukos, C. W. Lee, D. M. Thomas, J. Santana, S. W. Fanning, G. L. Greene, R. E. Moellering, "Targeting MYC with modular synthetic transcriptional repressors derived from bHLH DNA-binding domains" *Nature Biotechnology 2022* **2022**, *41*, 1–11.

[32] Z. Qiao, L. C. Nguyen, D. Yang, C. Dann, D. M. Thomas, M. Henn, A. Valdespino, C. S. Swenson, S. A. Oakes, M. R. Rosner, R. E. Moellering, "Direct inhibition of tumor hypoxia response with synthetic transcriptional repressors." *Nat Chem Biol* **2025**, *21*, 247–255.

[33] L. C. Lustig, D. Dingar, W. B. Tu, C. Lourenco, M. Kalkat, I. Inamoto, R. Ponzielli, W. C. W. Chan, J. A. Shin, L. Z. Penn, "Inhibiting MYC binding to the E-box DNA motif by ME47 decreases tumour xenograft growth" *Oncogene 2017 36:49* **2017**, *36*, 6830–6837.

[34] M. J. Demma, M. J. Hohn, A. Sun, C. Mapelli, B. Hall, A. Walji, J. O'Neil, "Inhibition of Myc transcriptional activity by a mini-protein based upon Mxd1" *FEBS Lett* **2020**, *594*, 1467–1476.

[35] R. V. Nithun, Y. M. Yao, X. Lin, S. Habiballah, A. Afek, M. Jbara, "Deciphering the Role of the Ser-Phosphor-

ylation Pattern on the DNA-Binding Activity of Max Transcription Factor Using Chemical Protein Synthesis" *Angewandte Chemie International Edition* **2023**, *62*, e202310913.

[36] H. Masui, S. Fuse, "Recent Advances in the Solid- and Solution-Phase Synthesis of Peptides and Proteins Using Microflow Technology" *Org Process Res Dev* **2022**, DOI 10.1021/acs.oprd.2c00074.

[37] R. Calo-Lapido, C. Penas, A. Jiménez-Balsa, M. E. Vázquez, J. L. Mascareñas, "A chemical approach for the synthesis of the DNA-binding domain of the oncoprotein MYC" *Org Biomol Chem* **2019**, *17*, 6748–6752.

[38] Z. Z. Brown, C. Mapelli, I. Farasat, A. V Shoultz, S. A. Johnson, F. Orvieto, A. Santoprete, E. Bianchi, A. B. McCracken, K. Chen, X. Zhu, M. J. Demma, B. M. Lacey, K. A. Canada, R. M. Garbaccio, J. O'Neil, A. Walji, "Multiple Synthetic Routes to the Mini-Protein Omomyc and Coiled-Coil Domain Truncations." *J Org Chem* **2020**, *85*, 1466–1475.

[39] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, "Synthesis of Proteins by Native Chemical Ligation" *Science (1979)* **1994**, *266*, 776–779.

[40] M. Paradís-Bas, J. Tulla-Puche, F. Albericio, "The road to the synthesis of 'difficult peptides'" *Chem Soc Rev* **2016**, *45*, 631–654.

[41] A. El-Faham, F. Albericio, "Peptide Coupling Reagents, More than a Letter Soup" *Chem Rev* **2011**, *111*, 6557–6602.

[42] H. Bürgisser, E. T. Williams, A. Jeandin, R. Lescure, A. Premanand, S. Wang, N. Hartrampf, "A Versatile 'Synthesis Tag' (SynTag) for the Chemical Synthesis of Aggregating Peptides and Proteins" *J Am Chem Soc* **2024**, *146*, 34887–34899.

[43] S. H. Shahravan, X. Qu, I.-S. Chan, J. A. Shin, "Enhancing the specificity of the enterokinase cleavage reaction to promote efficient cleavage of a fusion tag" **2008**, DOI 10.1016/j.pep.2008.02.015.

[44] A. J. Callahan, A. Rondon, R. M. Reja, L. L. Salazar, S. Gandhesiri, J. Rodriguez, A. Loas, B. L. Pentelute, "Same Day Access to Folded Synthetic Proteins" *J Am Chem Soc* **2024**, *146*, 28696–28706.

[45] T. Gaj, J. Guo, Y. Kato, S. J. Sirk, C. F. Barbas, "Targeted gene knockout by direct delivery of zinc-finger nuclease proteins" *Nat Methods* **2012**, *9*, 805–807.

[46] A. F. L. Schneider, M. Kithil, M. C. Cardoso, M. Lehmann, C. P. R. Hackenberger, "Cellular uptake of large biomolecules enabled by cell-surface-reactive cell-penetrating peptide additives" *Nat Chem* **2021**, *13*, 530–539.

[47] E. Wang, A. Sorolla, P. T. Cunningham, H. M. Bogdawa, S. Beck, E. Golden, R. E. Dewhurst, L. Florez, M. N. Cruickshank, K. Hoffmann, R. M. Hopkins, J. Kim, A. J. Woo, P. M. Watt, P. Blancafort, "Tumor penetrating peptides inhibiting MYC as a potent targeted therapeutic strategy for triple-negative breast cancers" *Oncogene* **2018**, *38*, 140–150.

[48] J. Rodríguez, S. Learte-Aymamí, J. Mosquera, G. Celaya, D. Rodríguez-Larrea, M. E. Vázquez, J. L. Mascareñas, "DNA-binding miniproteins based on zinc fingers. Assessment of the interaction using nanopores" *Chem Sci* **2018**, *9*, 4118–4123.

[49] J. Rodríguez, J. Mosquera, R. García-Fandiño, M. E. Vázquez, J. L. Mascareñas, "A designed DNA binding motif that recognizes extended sites and spans two adjacent major grooves" *Chem Sci* **2016**, *7*, 3298–3303.

[50] L. A. Jung, A. Gebhardt, W. Koelmel, C. P. Ade, S. Walz, J. Kuper, B. Von Eyss, S. Letschert, C. Redel, L. D'Artista, A. Biankin, L. Zender, M. Sauer, E. Wolf, G. Evan, C. Kisker, M. Eilers, "OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors" *Oncogene 2017 36:14* **2016**, *36*, 1911–1924.

[51] S. K. Nair, S. K. Burley, "X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors." *Cell* **2003**, *112*, 193–205.

[52] X. Lin, O. Harel, M. Jbara, "Chemical Engineering of Artificial Transcription Factors by Orthogonal Palladium(II)-Mediated S-Arylation Reactions" *Angewandte Chemie International Edition* **2024**, *63*, e202317511.

[53] P. Timmerman, W. C. Puijk, R. H. Meloen, "Functional reconstruction and synthetic mimicry of a conformational epitope using CLIPS technology." *J Mol Recognit* **2007**, *20*, 283–299.

[54] P. Timmerman, J. Beld, W. C. Puijk, R. H. Meloen, "Rapid and quantitative cyclization of multiple peptide loops onto synthetic scaffolds for structural mimicry of protein surfaces." *Chembiochem* **2005**, *6*, 821–824.

[55] K. Wu, H. Jiang, D. R. Hicks, C. Liu, E. Muratspahić, T. A. Ramelot, Y. Liu, K. McNally, S. Kenny, A. Mihut, A. Gaur, B. Coventry, W. Chen, A. K. Bera, A. Kang, S. Gerben, M. Y.-L. Lamb, A. Murray, X. Li, M. A. Kennedy, W. Yang, Z. Song, G. Schober, S. M. Brierley, J. O'Neill, M. H. Gelb, G. T. Montelione, E. Derivery, D. Baker, **2024**, DOI: 10.1101/2024.07.15.603480.

[56] K. Kosoglu, Z. Aydin, N. Tuncbag, A. Gursoy, O. Keskin, "Structural coverage of the human interactome." *Brief Bioinform* **2023**, *25*, 1–11.

[57] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. Dustin Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. Van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. Christopher Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, "Accurate prediction of protein structures and interactions using a three-track neural network" *Science (1979)* **2021**, *373*, 871–876.

[58] J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, "De novo design of protein structure and function with RFdiffusion" *Nature 2023 620:7976* **2023**, *620*, 1089–1100.

[59] L. Ledesma-Dominguez, E. Carbajal-Degante, G. Moreno-Hagelsieb, E. Perez-Rueda, "DeepReg: a deep learning hybrid model for predicting transcription factors in eukaryotic and prokaryotic genomes" *Sci Rep* **2024**, *14*, 1–11.

[60] G. B. Kim, Y. Gao, B. O. Palsson, S. Y. Lee, "DeepTFactor: A deep learning-based tool for the prediction of transcription factors" *Proc Natl Acad Sci U S A* **2021**, *118*, e2021171118.

[61] J. Liebold, F. Neuhaus, J. Geiser, S. Kurtz, J. Baumbach, K. Newaz, "Transcription factor prediction using protein 3D secondary structures" *Bioinformatics* **2024**, *41*, DOI 10.1093/BIOINFORMATICS/BTAE762.

[62] T. Orand, M. R. Jensen, "Binding mechanisms of intrinsically disordered proteins: Insights from experimental studies and structural predictions" *Curr Opin Struct Biol* **2025**, *90*, 102958.

[63] K. Van Roey, B. Uyar, R. J. Weatheritt, H. Dinkel, M. Seiler, A. Budd, T. J. Gibson, N. E. Davey, "Short linear motifs: Ubiquitous and functionally diverse protein interaction modules directing cell regulation" *Chem Rev* **2014**, *114*, 6733–6778.

[64] C. S. Swenson, G. Mandava, D. M. Thomas, R. E. Moellering, "Tackling Undruggable Targets with Designer Peptidomimetics and Synthetic Biologics" *Chem Rev* **2024**, *124*, 13020–13093.

[65] A. Dangi, I. J. Angera, J. R. Del Valle, "Intrastrand Peptide Staples That Promote β-Sheet Folding, Self-Assembly, and Amyloid Seeding" *J Am Chem Soc* **2025**, *147*, 29930–29938.

[66] K. Peukert, P. Staller, A. Schneider, G. Carmichael, F. Hänel, M. Eilers, "An alternative pathway for gene regulation by Myc." *EMBO J* **1997**, *16*, 5672–5686.

[67] Y. Zhou, X. Gao, M. Yuan, B. Yang, Q. He, J. Cao, "Targeting Myc Interacting Proteins as a Winding Path in Cancer Therapy" *Front Pharmacol* **2021**, *12*, 748852.

[68] N. B. Struntz, A. Chen, A. Deutzmann, R. M. Wilson, E. Stefan, H. L. Evans, M. A. Ramirez, T. Liang, F. Caballero, M. H. E. Wildschut, D. V Neel, D. B. Freeman, M. S. Pop, M. McConkey, S. Muller, B. H. Curtin, H. Tseng, K. R. Frombach, V. L. Butty, S. S. Levine, C. Feau, S. Elmiligy, J. A. Hong, T. A. Lewis, A. Vetere, P. A. Clemons, S. E. Malstrom, B. L. Ebert, C. Y. Lin, D. W. Felsher, A. N. Koehler, "Stabilization of the Max Homodimer with a Small Molecule Attenuates Myc-Driven Transcription." *Cell Chem Biol* **2019**, *26*, 711-723.e14.

[69] X. X. Sun, Y. Li, R. C. Sears, M. S. Dai, "Targeting the MYC Ubiquitination-Proteasome Degradation Pathway for Cancer Therapy" *Front Oncol* **2021**, *11*, 679445.

[70] S. Schütz, C. Bergsdorf, S. Hänni-Holzinger, A. Lingel, M. Renatus, A. D. Gossert, W. Jahnke, "Intrinsically Disordered Regions in the Transcription Factor MYC:MAX Modulate DNA Binding via Intramolecular Interactions" *Biochemistry* **2024**, *63*, 511.

[71] Y. Li, X. Li, J. Pu, Q. Yang, H. Guan, M. Ji, B. Shi, M. Chen, P. Hou, "C-Myc Is a Major Determinant for Antitumor Activity of Aurora A Kinase Inhibitor MLN8237 in Thyroid Cancer" *Thyroid* **2018**, *28*, 1642–1654.

[72] J. M. Mata, E. van der Nol, S. J. Pomplun, "Advances in Ultrahigh Throughput Hit Discovery with Tandem Mass Spectrometry Encoded Libraries" *J Am Chem Soc* **2023**, *145*, 19129–19139.

[73] S. Pomplun, Z. P. Gates, G. Zhang, A. J. Quartararo, B. L. Pentelute, "Discovery of Nucleic Acid Binding Molecules from Combinatorial Biohybrid Nucleobase Peptide Libraries" *J Am Chem Soc* **2020**, *142*, 19642–19651.

[74] K. M. Esvelt, J. C. Carlson, D. R. Liu, "A system for the continuous directed evolution of biomolecules" *Nature* **2011**, *472*, 499–503.

[75] S. M. Miller, T. Wang, D. R. Liu, "Phage-assisted continuous and non-continuous evolution" *Nat Protoc* **2020**, *15*, 4101–4127.

[76] I. Inamoto, I. Sheoran, S. C. Popa, M. Hussain, J. A. Shin, "Combining Rational Design and Continuous Evolution on Minimalist Proteins That Target the E-box DNA Site." *ACS Chem Biol* **2021**, *16*, 35–44.

[77] J. Xu, G. Chen, A. T. De Jong, S. H. Shahravan, J. A. Shin, "Max-E47, a designed minimalist protein that targets the E-box DNA site in vivo and in vitro" *J Am Chem Soc* **2009**, *131*, 7839–7848.

[78] M. S. Packer, H. A. Rees, D. R. Liu, "Phage-assisted continuous evolution of proteases with altered substrate specificity" **n.d.**, DOI 10.1038/s41467-017-01055-9.

[79] A. H. Badran, V. M. Guzov, Q. Huai, M. M. Kemp, P. Vishwanath, W. Kain, A. M. Nance, A. Evdokimov, F. Moshiri, K. H. Turner, P. Wang, T. Malvar, D. R. Liu, "Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance" *Nature* **2016**, *533*, 58–63.

[80] H. Xiang, L. Bai, X. Zhang, T. Dan, P. Cheng, X. Yang, H. Ai, K. Li, X. Lei, "A facile strategy for the construction of a phage display cyclic peptide library for the selection of functional macrocycles" *Chem Sci* **2024**, *15*, 11847–11855.

[81] F. F. Faucher, K. Kristýna, B. Ková, S. Lovell, M. Bertolini, J. Herrero-Bourdieu, E. D. Cosco, M. Bogyo, M. Barniol-Xicota, "Macrocyclic Phage Display for Identification of Selective Protease Substrates" *J Am Chem Soc* **2025**, DOI 10.1021/JACS.5C04424.

[82] L. Boike, A. G. Cioffi, F. C. Majewski, J. Co, N. J. Henning, M. D. Jones, G. Liu, J. M. McKenna, J. A. Tallarico, M. Schirle, D. K. Nomura, "Discovery of a Functional Covalent Ligand Targeting an Intrinsically Disordered Cysteine within MYC." *Cell Chem Biol* **2021**, *28*, 4-13.e17.

[83] F. M. Paulussen, G. K. Schouten, C. Moertl, J. Verheul, I. Hoekstra, G. M. Koningstein, G. H. Hutchins, A. Alkir, R. A. Luirink, D. P. Geerke, P. Van Ulsen, T. Den Blaauwen, J. Luirink, T. N. Grossmann, "Covalent Proteomimetic Inhibitor of the Bacterial FtsQB Divisome Complex" *J Am Chem Soc* **2022**, *144*, 15303–15313.

[84] L. Petri, P. Ábrányi-Balogh, D. Vagrys, T. Imre, N. Varró, I. Mándity, A. Rácz, L. Wittner, K. Tóth, E. Z. Tóth, T. Juhász, B. Davis, G. M. Keserű, "A covalent strategy to target intrinsically disordered proteins: Discovery of novel tau aggregation inhibitors" *Eur J Med Chem* **2022**, *231*, 114163.

[85] M. Gehringer, S. A. Laufer, "Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology" *J Med Chem* **2018**, *62*, 5673–5724.

[86] J. Pettinger, K. Jones, M. D. Cheeseman, "Lysine-Targeting Covalent Inhibitors" *Angewandte Chemie International Edition* **2017**, *56*, 15200–15209.

Appendix A

List of abbreviations

A	Adenine
AI	Artificial intelligence
ARM	Arginine-righ motif
ASO	Antisense oligonucleotide
AUC	Area under the curve
b	basic region
bHL-HL-LZ	basic helix-loop-helix leucine zipper
BIV	Bovine immunodeficiency virus
BLI	Biolayer interferometry
BLM	Bloom syndrome protein helicase domain
BRD4	Bromodomain protein 4
bZIP	basic Zipper
C	Cytosine
CD	Circular Dichroism
cDNA	Complementary DNA
cpp	cell-penetrating peptide
CTG	Cell titer glow
CV	Colume volume
DBD	DNA binding domain
DCM	Dichloromethane
DIPEA	N,N-Diisopropylethylamine
DMF	N,N-Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxynucleic Acid
DPBS	Dulbecco's phosphate buffered saline
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
E-box	Enhancer box
EDTA	Ethylenediaminetetraacetic acid
EK	Enterokinase
EMSA	Electrophoretic mobility shift assay
EPO	Erythropoietin
FITC	Fluorescein isothiocyanate
Fmoc	Fluorenylmethyloxycarbonyl
G	Guanine
G4	G-quadruplex
Gly	Glysine
GSEA	gene set enrichment analysis
HATU	O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate
HCAC3	Histone Deacetylase 3
HCV	Hepatitic C virus
HIV-1	Human Immunodeficiency Virus type-1
HLH	Helix-loop-helix
HLHLD	Helix-loophelix-domain
HPLC	High-pressure Liquid chromatography
HTP	Helix-threading peptide
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IRES	Internal ribosome entry site
IS	Internal standard
LC-MS	Liquic Chromatography Mass spectroscopy
LTR	Long terminal repeat
Lys	Lysine
LZ	Leucine Zipper
Max	Myc-associated factor X
MB	Myc box
MD	Molecular Dynamics
MeCN	Acetonitrile
Mga	Max gene-associated
MID1/2	Myc interacting Domain 1/2

miR-21	miRNA-21
miRNA	Micro RNA
Miz-1	Myc interacting Zinc Finger 1
Mlx	Max-like protein X
MNT	Max-binding protein
mRNA	Messenger RNA
MSA	Multiple sequence alignment
MSigDB	Molecular Signatures Database
MW	Molecular weigth
NA	Nucleic acid
NCL	Native chemical ligation
ncRNA	Non-coding RNA
NEAA	Non-essential amino acids
NLS	Nuclear localization sequence
OBTC	One-bead-two-compounds
Omo-103	Omomyc-103
PACE	Phage-assisted continuous evolution
pal-RNA	Palyndromic RNA
PBS	Phosphate-Buffered Saline
PG	Protection group
PNA	Peptide nucleic acid
PNAI	Protein-nucleic acid interaction
Pol II	RNA polymerase II
POZ	Poxvirus and Zinc finger
PPI	Protein-protein interaction
pre-miRNA	precursor miRNA
pri-miRNA	Primary miRNA
PROTAC	Proteolysis Targeting Chimera
P-TEFb	Elongation factor b
QTOF	Quadrupole Time-of-Flight
RBD	RNA binding domain
RBS	Ribosome binding site
RISC	RNA-induced silencing complex
RMSD	Root-mean-square-deviation
RMSF	Root-mean-square-fluctuation
RNA	Ribonucleic Acid
RP-FC	Reversed phase flash chromatograpy
RRE	Rev Response Element
RRM	RNA recognition motif
RT	Room temperature
SEC	Superelongation complex
SPPS	Solid phase peptide synthesis
ssRNA	Single stranded DNA
T	Thymine
TAD	Transactivation domain
TALEN	Transcription activator-like effecotr nuclease
TAR	Transactivation response
TAV2b	*Tomato Aspermy virus 2b*
TBP	TAR-binding-protein
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TF	Transcription factor
TFA	Trifluoroacteic acid
TFE	Trifluorethanol
TIC	Total-ion chromatogram
tRNA	Transfer RNA
TRRAP	Transactivation/Transformation-Domain Associated Protein
UTR	Untranslated region
XIC	Extracted-ion chromatogram
ZF	Zinc Finger
γ-AA	γ-substituted-N-acylated-N-aminoethylamino acids

Nederlandse samenvatting

Het uitschakelen van de transcriptie factor (TF) Myc staat centraal in dit proefschrift. TFs zijn DNA bindende eiwitten en staan vaak bij de ontwikkeling van een ziekte centraal. Het is daarom belangrijk om onderzoek te doen naar aan ziektegerelateerde TFs. Myc is een schoolvoorbeeld van een TF die betrokken is bij de ontwikkeling van ziektes. Het is bijvoorbeeld bij meer dan 50% van alle kankertypes betrokken. Echter, het is erg lastig om Myc uit te schakelen. Myc is intrinsiek ongeordend en vormt alleen een stabiel eiwitstructuur in complex met zijn eiwitpartner Max. Zonder een stabiele structuur is het erg lastig om Myc specifiek te kunnen herkennen en uit te schakelen.

Dit proefschrift onderzoekt meerdere mogelijkheden om Myc uit te schakelen. Met behulp van kristalstructuren, kunstmatige intelligentie en eiwit expressie worden er nieuwe transcriptieremmers ontworpen, gesynthetiseerd en getest tegenover Myc. Enerzijds worden er technieken behandeld die de DNA-sequentie blokkeren waar Myc normaal gesproken aan bindt. Anderzijds wordt er besproken hoe de Myc/Max interactie geblokkeerd kan worden doormiddel van een derde eiwit.

In **Hoofdstuk 1** wordt achtergrondinformatie verstrekt dat de basis vormt voor het idee achter het onderzoek en dat dieper in gaat op het hoofddoelwit van dit proefschrift: Myc. Het hoofdstuk begint met een algemene introductie over TFs. Deze eiwitfamilie is belangrijk voor het activeren en inactiveren van genen. Echter, zijn ze ook vaak betrokken bij de ontwikkeling van ziektes, zoals kanker. Aangezien TFs veelal geen stabiele structuur hebben en diep in de celkern hun functie uitvoeren zijn het lastige eiwitten om uit te schakelen. In de introductie wordt een vaak gebruikte strategie besproken om TFs te inhiberen: het gebruik van peptiden en hoe deze gebruikt kunnen worden om TFs uit te schakelen, bijvoorbeeld door kunstmatige TFs te ontwerpen. Kunstmatige TFs zijn vaak gebaseerd op hun natuurlijke tegenhanger en kunnen daarom dezelfde DNA-sequentie binden en daarmee blokkeren. Naast verschillende ontwerpstrategieën om moeilijkheden te overbruggen wordt er ook dieper ingegaan op de biologische functie van de oncologische TF Myc en welke strategieën er al zijn ontwikkeld om Myc uit te schakelen. Alles bij elkaar genomen verstrekt dit hoofdstuk de informatie en wetenschappelijke context dat ten grondslag ligt aan dit proefschrift.

Hoofdstuk 2 gaat dieper in op het gebruik van peptiden om aan ziektegerelateerd DNA en RNA te beïnvloeden. Een voordeel dat peptiden hebben tegenover het

gebruik van 'kleine moleculen', is dat peptiden meerdere conformaties aan kunnen nemen waardoor ze efficiënter grote en lastig te herkennen DNA of RNA-oppervlaktes kunnen onderscheiden. Er wordt ingegaan op verschillende strategieën om deze peptiden te identificeren. Het screenen van grote bibliotheken vergroot de kans om werkzame peptiden te vinden. Daarnaast worden 'op structuur-gebaseerde studies' besproken die succesvol peptiden of mini-eiwitten hebben ontworpen. Ook wordt het gebruik van post-synthetische peptide modificatie besproken. Vaak worden peptiden gecycliseerd of in hun actieve conformatie gehouden doormiddel van de *staple* techniek. Deze techniek houdt in dat twee zijketens van aminozuren covalent aan elkaar worden gebonden waardoor de actieve conformatie in stand blijft. De bevindingen van dit hoofdstuk geven extra context voor het onderzoek dat wordt besproken in dit proefschrift.

In **Hoofdstuk 3** wordt een op structuur-gebaseerde studie toegepast om Myc uit te schakelen. In dit hoofdstuk beschrijven wij het ontwerpproces om van een kristalstructuur naar een mini-eiwit te komen dat in staat is om Myc uit te schakelen. Als startpunt werd de kristalstructuur van *Omomyc* gebruikt. *Omomyc* is volgens de literatuur in staat om Myc uit te schakelen, maar heeft een groot obstakel: het binnendringen van de cel. Onze hypothese was dat voornamelijk de grootte van *Omomyc* het binnendringen belemmerd. In dit onderzoek hebben wij meerdere varianten van verschillende groottes van *Omomyc* gemaakt. Doormiddel van biochemische experimenten hebben wij aangetoond dat enkel het domein dat DNA bindt niet voldoende is om een stabiel DNA/eiwit complex te vormen, maar dat er een extra domein nodig is voor verhoogde stabiliteit en affiniteit. Dit mini-eiwit, dat we *DuoMYC* hebben genoemd, bleek in staat te zijn om cellen binnen te dringen, Myc te inactiveren en kanker gerelateerde signaalcascades te beïnvloeden.

Hoofdstuk 4 gaat een stap verder omtrent het gebruik van de kristalstructuur van *Omomyc.* Naast het verkleinen van eiwitten, kan ook het inbrengen van *staples* er voor zorgen dat eiwitten efficiënter een cel binnendringen. Normaal gesproken kunnen gezuiverde eiwitten snel hun secondaire en tertiaire structuur verliezen. Dit is niet het geval voor *Omomyc* of zijn familieleden. Deze eiwitten hebben de karakteristieke eigenschap om de correcte secondaire structuur te vormen wanneer ze in contact komen met eiwitpartners. In **Hoofdstuk 4** proberen wij deze eigenschap in ons voordeel te gebruiken. Hier hebben wij varianten van *Omomyc* tot expressie gebracht in bacteriën om ze vervolgens vast te zetten in hun actieve conformatie door middel van de *staple* techniek. Deze techniek hebben wij de *ReCHEMbinant* techniek genoemd. *ReCHEMbinant* stelt de gebruiker in staat om makkelijk vouwende eiwitten met behulp van bacteriën te maken, te modificeren en te stabiliseren terwijl het eiwit zijn functie kan behouden. Als voorbeeld hebben wij *Omomyc* gemodificeerd tot *HeloMYC. HeloMYC,* was succesvol gesynthetiseerd en was in staat de cel binnen te dringen, Myc te inactiveren en kanker gerelateerde signaalcascades te beïnvloeden. Daarbij bleken de gemodificeerde eiwitten stabieler dan het origineel.

In **Hoofdstuk 5** benaderen we Myc inhibitie vanaf een andere kant. In dit hoofdstuk blokkeren we niet de DNA-sequentie van Myc, maar de interactie tussen Myc en Max. Met behulp van AlphaFold-Multimer analyseerden wij de interactie tussen Myc en een ander eiwitpartner: Miz-1. AlphaFold-Multimer voorspelt hoe eiwitten met elkaar interactie aan kunnen gaan aan de hand van de primaire aminozuur sequentie. De voorspelling van AlphaFold-Multimer kwam overeen met wat bekend is in de literatuur. Dit stelde ons in staat om een op Miz-1 gebaseerde Myc binder, *Mizmetic*, te ontwerpen. Biochemische experimenten valideerden de interactie tussen Myc en *Mizmetic* en de incorporatie van een *staple* resulteerde in een verhoogde affiniteit. Bovendien werden posities geïdentificeerd in *Mizmetic* die van essentieel belang zijn voor de interactie tussen Myc en *Mizmetic*. Ook werden mutaties geïdentificeerd, die verklaard konden worden door de AlphaFold voorspelling, die een versterkend effect hadden op de interactie tussen Myc en *Mizmetic*. Deze mutanten waren in staat om de interactie tussen Myc/Max/DNA te blokkeren. Ook konden sommigen beter onderscheid maken tussen Myc/Max en Max/Max.

In **Hoofdstuk 6** worden de belangrijkste bevindingen van dit proefschrift samengevat. Het minimaliseren of het stabiliseren van de *Omomyc* sequentie hebben geresulteerd in nieuwe Myc remmers. Ook het gebruik van kunstmatige intelligentie heeft geleid tot een nieuwe Myc remmer. De nieuwe ontwerpen en ontwerpstrategieën dragen bij aan het arsenaal om Myc uit te kunnen schakelen en dichter bij een remedie te komen. Ook wordt vervolgonderzoek besproken. Bijvoorbeeld hoe *DuoMYC* gebruikt kan worden om andere aan ziektegerelateerde TFs uit te schakelen. Ook wordt besproken hoe de *ReCHEMbinant* techniek gebruikt kan worden voor onderzoek naar intrinsiek ongeordende eiwitten en welke methodes er nog meer zijn om Myc uit te kunnen schakelen. Het proefschrift benadrukt daarmee zijn toegevoegde waarde in het onderzoeksveld van zowel Myc als intrinsiek ongeordende eiwitten.

Het gehele proefschrift bij elkaar genomen biedt een samenhangend geheel van rationeel ontworpen kunstmatige TFs tot aan het gebruik van kunstmatige intelligentie voor peptide ontwerp, altijd met Myc in het middelpunt. Zolang er geen remedie is ontwikkeld tegen Myc is het van belang om onderzoek te doen naar dit eiwit. Het onderzoek kan ook breder getrokken worden naar andere TFs en ongeordende eiwitten. De bevonden resultaten in dit proefschrift kunnen aan dit onderzoek bijdragen.

English summary

The aim of this thesis is to inhibit the transcription factor (TF) Myc. Myc is involved in more than 50% of all cancers. However, there is currently no treatment to inhibit Myc. As such, it is important to do research towards Myc or other disease-related TFs. As Myc is intrinsically disordered without its protein partner Max and located inside the nucleus, it is a difficult pharmaceutical target.

This thesis tries to inhibit Myc using different strategies. By using crystal structures, artificial intelligence and protein expression, novel Myc inhibitors were designed, synthesized and tested. We either prevent Myc from binding to its canonical target sequence or we prevent Myc to bind to its protein partner Max.

Chapter 1 goes into detail about background information that explains fundamental topics related to artificial TF and peptidomimetic design. It begins with a general introduction about TFs, that play an important role in gene regulation. However, they are also involved in multiple diseases, such as cancer. Despite their importance, TFs are difficult to inhibit due to their location deep inside the cellular nucleus and their partly disordered nature. The introduction outlines general strategies for inhibiting TFs and discusses how peptides can be used to overcome common challenges. Peptides can be engineered into artificial TFs that mimic their natural counterpart, target the same canonical DNA binding site and function as a transcriptional repressor. Additionally, Chapter 1 highlights the biological function of the oncogenic TF Myc and reviews strategies known to target this protein. Taken together, this chapter provides the scientific content to understand the topics discussed in this thesis.

Chapter 2 provides an extensive overview of peptides designed to target disease-related DNA or RNA. Compared to small molecules, peptides can adopt multiple conformations allowing them to recognize large DNA and RNA surfaces. We explain different strategies to discover peptides that can target DNA or RNA. Often large libraries are designed and screened against DNA or RNA of interest. Large libraries increase the possibility of finding peptides that can target a specific sequence. Additionally, structure-based design studies are discussed that used endogenous TFs as a template and successfully designed peptides or miniproteins. Lastly, late-stage peptide modifications can be beneficial. For example, cyclisation or stapling can increase affinity towards a specific sequence by locking the peptide in the active conformation. The findings in this chapter provide an extra context for the following chapters discussed within this thesis.

In **Chapter 3** we perform a structure-based design study, using Omomyc as a template, to design artificial TFs to inhibit Myc. Omomyc is known to inhibit Myc. However, not all studies show that Omomyc has intrinsic cell permeability. We hypothesize that Omomycs size is the main limitation and aim to decrease the overall structure to the core elements needed for successful DNA binding. We designed multiple variants that differed in size based on Omomyc and showed that the basic region alone is not sufficient for successful DNA binding. An additional domain is needed for a stable protein/DNA complex. This designed miniprotein, DuoMYC, is able to enter cells, inhibit Myc and influence cancer related pathways.

Chapter 4 uses the full crystal structure of Omomyc. Instead of decreasing overall size, we incorporate staples and positive charge to enhance intrinsic cell permeability. Normally, protein expression is a delicate process to retain the overall protein structure. However, Omomyc can fold upon protein partner contact which allowed us to incorporate staples after we expressed the proteins in bacteria. We called this technique the ReCHEMbinant technique. Using the ReCHEMbinant platform we can modify and stabilize easy-folding proteins after protein expression. We use Omomyc as a proof of principle protein and designed HeloMYC. HeloMYC could enter cells, inactivate Myc and influence cancer related pathways. Additionally, the modified proteins were more stable compared to their parent compound.

In **Chapter 5** we inhibit Myc via a different strategy. In this chapter we do not aim to block the canonical DNA binding site of Myc but the interaction between Myc and Max. By using AlphaFold-Multimer we analyzed the interaction between Myc and another protein partner of Myc: Miz-1. AlphaFold-Multimer predicts how proteins interact based on their primary sequence. Our Miz-1/Myc prediction was supported by literature and allowed us to design Miz-1-based Myc inhibitors, Mizmetics. Biochemical assays validated an interaction between Myc and Mizmetics and, additionally, the incorporation of a staple resulted in increased affinity. Moreover, we identified positions that were essential for the interaction between Myc and Mizmetic. Other mutations, that could be explained by AlphaFold, increased affinity between our designed peptides and Myc. These mutated peptides could inhibit Myc/Max/DNA interaction and became more specific towards Myc.

In **Chapter 6** the most important findings are summarized. Minimalizing and stabilizing Omomyc resulted in novel Myc inhibitors. Additionally, the use of artificial intelligence (AI) resulted in a newly identified Myc inhibitor. The novel designs and design strategies assist with inhibiting Myc and bring us closer to a cure towards Myc-related cancer. Additionally, future perspectives are discussed. For example, how DuoMYC can be used to target other disease-related TFs, how the ReCHEMbinant approach can be used to analyze disordered proteins or other strategies to inhibit Myc. This thesis highlights its value in the struggle against Myc.

The thesis provides a concise overview of rationally designed artificial TFs and AI-assisted peptide design, always with the aim to inhibit Myc. Without a treatment for Myc-related cancers, research remains important to get a better understanding of Myc. Moreover, the work described in this thesis can also be used to target other disease-related TFs and oncogenic proteins.

List of publications

Part of this thesis

B. D. Ellenbroek,[‡] J. P. Kahler,[‡] S. R. Evers, S. J. Pomplun, "Synthetic Peptides: Promising Modalities for the Targeting of Disease-Related Nucleic Acids" *Angewandte Chemie International Edition* **2024**, *63*

B. D. Ellenbroek,[‡] J. P. Kahler,[‡] D. Arella, C. Lin, W. Jespers, E. A. Züger, M. Drukker, S. J. Pomplun, "Development of DuoMYC: a synthetic cell penetrant miniprotein that efficiently inhibits the oncogenic transcription factor MYC" *Angewandte Chemie International Edition* **2025**, *64*

J. P. Kahler, **B. D. Ellenbroek**, V. E. van der Noord, B. van de Water, S. J. Pomplun, "ReCHEMbinant stapling enhances intracellular delivery and bioactivity of engineered protein inhibitors" *Chem* **2025**, 102839.

B. D. Ellenbroek, A. S. S. Tirtosentono, S. J. Pomplun, "AlphaFold-Guided Discovery of an Overlapping MYC/Miz-1 Interface Enables Peptidomimetic Disruption of MYC/MAX" *ChemMedChem* **2025,** *e202500740*

[‡]Contributed equally

Other publications (not part of thesis)

G. L. McLelland, M. Lopez-Osias, C. R. C. Verzijl, **B. D. Ellenbroek**, R. A. Oliveira, N. J. Boon, M. Dekker, L. G. van den Hengel, R. Ali, H. Janssen, J. Y. Song, P. Krimpenfort, T. van Zutphen, J. W. Jonker, T. R. Brummelkamp, "Identification of an alternative triglyceride biosynthesis pathway" *Nature* **2023**, *621*, 171–178.

J. Braun, M. I. Ariëns, B. T. Matsuo, S. de Vries, E. D. H. van Wordragen, **B. D. Ellenbroek**, C. M. L. Vande Velde, R. V. A. Orru, E. Ruijter, "Stereoselective Synthesis of Fused Vinylcyclopropanes by Intramolecular Tsuji-Trost Cascade Cyclization." *Org Lett* **2018**, *20*, 6611–6615.

Oral and poster communications

2025 **ChemBioParis** (oral)
Paris, France

2025 **Leiden Early Drug Discovery & Development (LED3) symposium** (oral)
Leiden, the Netherlands

2025 **LACDR Spring symposium** (poster)
Leiden, the Netherlands – Awarded with Poster Competition Prize

2025 **KNCV - SOC-MCCB symposium** (oral)
Lunteren, the Netherlands

2024 **European Peptide Conference** (poster)
Florence, Italy

2024 **Oncode Annual Meeting** (poster)
's-Hertogenbosch, the Netherlands

2024 **LACDR Spring symposium** (poster)
Leiden, the Netherlands

2024 **EFMC-ACSMEDI Medicinal Chemistry Frontiers symposium** (poster)
Utrecht, the Netherlands

2023 **LACDR Spring symposium** (poster)
Leiden, the Netherlands

2023 **Dutch Peptide Symposium** (poster)
Leiden, the Netherlands

2022 **LACDR Spring symposium** (poster)
Leiden, the Netherlands

Curriculum Vitae

Brecht Ellenbroek was born in Amsterdam, the Netherlands, on 20th of June 1995. He finished his high school education at the Sint Nicolaas Lyceum in Amsterdam (Noord-Holland, the Netherlands) in 2014. He started with the joint degree bachelor in Chemistry at the University of Amsterdam (UvA, Amsterdam, the Netherlands) and the Vrije Universiteit Amsterdam (VU, Amsterdam, the Netherlands). He did his bachelor internship in organic chemistry at the Synthetic and Bio-organic Chemistry group under supervision of Prof. Dr. Eelco Ruijter, Prof. Dr. Ir. Romanu Orru and John Braun resulting in a co-authorship on a publication (not part of this thesis).

After graduating in 2018, he began his master's program in Drug Discovery and Safety (now Drug Discovery Sciences) at the VU (Amsterdam, the Netherlands). Within his master's program he did three internships belonging to three disciplines: Biomolecular simulations, peptide/protein chemistry and molecular/cellular biology. He did his four-month biomolecular simulation internship under the supervision of Dr. Jocelyne Vreede (UvA) and his five-month internship in peptide chemistry under the supervision of Prof. Dr. Tom Grossmann and Dr. Niall Mcloughlin (VU). He followed his major internship of seven months in molecular and cellular biology under the supervision of Prof. Dr. Thijn Brummelkamp and Dr. Gian-Luca Mclelland at the Netherlands Cancer Institute. This resulted in a co-authorship on a publication in Nature (not part of this thesis). He obtained his master's degree *cum laude* in 2021.

He began his PhD in January 2022 at the Division of Medicinal Chemistry (Leiden Academic Centre for Drug Research (LACDR), Leiden University) under the supervision of Dr. Sebastian Pomplun, Prof. Dr. Nathaniel Martin and Prof. Dr. Miranda van Eck. His research focused on how to exploit endogenous transcription factors (TF) to design artificial DNA-binding proteins or TF-based inhibitors to target the oncogenic TF Myc. His aim is to pursue his scientific career in industry.

Acknowledgement

First and foremost I would like to show my gratitude towards **Sebastian**. I had the privilege of being your first PhD candidate and with your enthusiasm for science and great supervision we managed to make this PhD into a success. I learned a lot from you about how to start a project, how to plan projects, how to finish or when to abandon a project. It was not always easy, with proteins crashing out, behaving weird or refusing to be synthesized. But overall you helped me by keeping my spirit up and keeping me focused on the end goal. Since I started the lab expanded drastically and I am confident you will continue this progress. Thank you.

Nathaniel you always showed interest in my projects and in me personally, thank you. I also want to thank my other supervisors **Hubertus** and **Miranda**.

Jan (-Pascal) thanks, danke, bedankt, mercikes. Zonder jou was mijn PhD niet zo soepel verlopen. Met jouw hulp en je altijd vriendelijke en geduldige begeleiding stond je altijd klaar om met mij de discussie aan te gaan over onderzoek gerelateerde onderwerpen, maar ook over van alles en nog wat. Je hebt inderdaad gelijk. Het is VALdorpel. Ik wil ook graag **Khaled** en **Cas** bedanken. **Cas** het sparren over hardlopen, de fusie en fellowship avonden waren fantastisch. **Khaled** the past four years where amazing. Your relaxed attitude and you always being hyped about basically everything really inspired me. Additionally, I want to thank you for being my paranymph.

A special thanks to **Edith** and **Miguel**, together we were the first three members of this lab. Thanks for the fun times. Edith, good luck with your libraries, if somebody can do it, you can. Miguel I never met somebody with such a keen eye for detail when it comes to figure design like you. But more importantly, I am happy that somebody from Portugal likes bitterballen this much. To all other Pomplunlab members: **Jingming**, **Hugo**, **Thanasis**, **Sean**, **Monika**, **Mike**, **Gianluca**, **Patryk** thank you!

To all my students, without you guys I would not have been able to do all these projects. Thank you **Damiano**, **Cherina**, **Amardhiko**, **Sofie**, **Rozina** and **B'Elana**.

I also want to thank everyone who helped me with protein expression or peptide purification. Thank you **Vlad**, **Patrick**, **Anneloes** and **Monika.**

Of course I also want to thank the whole **Medchem** department. To prevent that I forget somebody I will not mention everybody by name. Everybody from **Gerards**,

Laura's, **Daans** and **Ads** group created such a supporting environment and made me feel welcome from my first day. A special thanks to **Linde** and **Inge**. **Linde** always stay this happy. **Inge** bedankt voor het luisteren naar Désenchantée ookal kon je er helaas niet van genieten. **Bert** ik ben blij dat ik nog iemand heb gevonden die zoveel flauwe woordgrappen maakt als ik.

Ik wil natuurlijk ook mijn familie en vrienden bedanken. Zonder jullie had ik dit niet gekund. **Pap**, **Mam** door jullie steun de afgelopen jaren ben ik hier gekomen. **Ward** en **Veerle** we zijn dan wel verschillend, maar dat brengt ons ook dichter bij elkaar. Bedankt voor jullie luisterend oor en advies tijdens de goede, maar ook de zwaardere momenten. **Veerle**, nu het mij is gelukt ben jij aan de beurt. Ik kan natuurlijk jou niet vergeten **Nienke**. Ik ben blij dat jij en **Ward** elkaar hebben gevonden. Ook wil ik jullie drie bedanken **Arjen**, **Ageeth** en **Maartje** voor jullie openheid en barmhartigheid vanaf dag één dat ik Anne heb leren kennen. Daarnaast wil ik natuurlijk al mijn vrienden bedanken. Wie had verwacht dat zoveel van ons een PhD zouden gaan doen. Bedankt **Stan**, **Lucien**, **Pim**, **Lars**, **Jelle**, **Dennis, David** en **Yorrick** en natuurlijk ben ik ook jullie niet vergeten **Rosa** en **Rhea**. Ook wil ik iedereen van Cohesie bedanken. Ik vind het fantastisch dat we na zoveel jaren nog zoveel met elkaar om gaan. Bedankt **Stef**, **Guido**, **Nordin**, **Jordan, Jente** en **Mees**. **Max** en **Thibo** ik wil jullie apart bedanken. Ondanks dat we compleet verschillende dingen zijn gaan doen, zijn we alleen maar dichter naar elkaar toe gegroeid. **Thibo**, mijn andere paranimf, bedankt. Als laatste wil ik Bijltjespad bedanken: **Sybren**, **Faan**, **Marnix** en **Nander**. Ik had mij geen betere huisgenoten kunnen wensen. **Mark** bedankt dat je mij aan de marathons hebt geholpen. Het is een cliché, maar een PhD is inderdaad vergelijkbaar.

Als laatste wil ik jou bedanken **Anne**. Bedankt voor je altijd luisterend oor, geduld, liefde en het aanhoren van mijn flauwe woordgrappen. Je bent mijn steun en toeverlaat en geeft mij een gevoel van rust.