THE UNIVERSITY OF CHICAGO

A SYNTHETIC MINIPROTEIN PLATFORM TARGETING BASIC HELIX-LOOP-HELIX TRANSCRIPTION FACTORS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

SEAN SHANGGUAN

CHICAGO, ILLINOIS

MARCH 2021

In Nature's infinite book of secrecy a little I can read.

— William Shakespeare

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to the following people, without whom the completion of this doctoral dissertation would be impossible.

First and foremost, I would like to express great appreciation to my doctoral advisor, Dr. Raymond Moellering. As a young scientist, Ray's confidence, ambition, enthusiasm, and extensive knowledge in the field have always inspired me to pursue this challenging topic throughout the years. Ray gave me great flexibility and trust to explore a wide variety of projects presented in this dissertation field while providing practical guidance whenever necessary. Under Ray's mentorship, I have acquired valuable hard and soft skills to ensure my success in scientific research as well as my future career. It is a great honor and privilege to work as one of the founding members of the Moellering lab.

I would also like to thank all members of our lab. I enjoyed being surrounded by a group of intelligent, motivated and supportive individuals with a shared endeavor. I am particularly grateful for my colleagues, Dr. Thomas Speltz and Mr. Zeyu Qiao, whose work directly influenced my projects and was partly presented in Chapter 4, and Dr. Somayeh Ahmadiantehrani, who assisted with figure and text editing for this dissertation.

I would also like to acknowledge my collaborators. Dr. Kevin White and Dr. Grace Yu from the Institute for Genomics and System Biology worked closely with me and provided material and technical assistance on the chromatin immunoprecipitation experiments. Dr. Geoffrey Greene and Dr. Sean Fanning from the Ben May Institute for Cancer Research initiated X-ray crystallography and analyzed the crystal structure, which is an integral part of this dissertation.

I would like to thank the members of my Ph.D. candidacy committee and dissertation committee, Dr. Joseph Piccirilli, Dr. Geoffrey Greene, Dr. Bryan Dickinson, and Dr. Sergey Kozmin, for their time and efforts in providing suggestions and commenting on my research leading up to this dissertation.

Finally, I am profoundly appreciative of the unwavering support of my friends and family. Throughout the enduring journey of graduate school, especially in the challenging year of 2020, their continuous care, encouragement, and understanding have always been a great source of comfort.

Abstract

Transcription factors (TFs) control gene expression by recognizing specific DNA sequences, thereby regulating most, if not all, aspects of cellular function. As such, many cancers are hallmarked by deregulation of TFs, which makes them ideal targets for cancer therapy. However, direct inhibition of the DNA binding of TFs remains largely untapped due to technical challenges in targeting protein-protein and protein-DNA interactions. In particular, c-Myc, a protooncogenic TF of the basic helix-loop-helix family, is deregulated in up to 70% of cancers but its inhibitor development has eluded researchers for decades.

Here we developed a modular platform to create fully synthetic miniproteins derived from bHLH TFs. As a proof-of-concept study, synthesis of heterodimeric, stabilized DNA-binding helices derived from Myc/Max yielded synthetic DNA-binding domains (sDBDs) that recapitulate the DNA binding activity of full-length TFs. Lead sDBDs showed increased structural stability, cellular uptake, and promising intracellular target accessibility. Extensive studies into the stabilization of secondary and tertiary structures provided insights into their structure activity relationship, and further optimizations of sDBDs led to the development of new orthogonal chemistries.

Next, we applied the synthetic platform to a synthetic transcription factor (STF) scaffold derived from the basic and leucine zipper domains of Max. STFs showed superior binding affinity and specificity to target DNA, and effectively competed with bHLH proteins *in vitro*. Vigilant stabilization of secondary, tertiary and quaternary structural elements in STFs simultaneously preserved biochemical activity and significantly improved structural, thermal and proteolytic

stability. Lead STFs are cell permeable, distribute throughout the cytosol and nucleus intact, and directly bind Myc target genes. We also solved the crystal structure of an STF-DNA complex, confirming that STFs assemble and recognize DNA in a manner similar to full length bHLH TFs.

Finally, we conducted thorough studies into the synthesis of dimeric STFs (dSTFs) using the multidimensional orthogonal chemistries in our platform. dSTFs possess enhanced activity and stability compared to STFs, and represent the largest stabilized miniprotein that have been made to date.

Collectively, these results validated the modularity and robustness of our synthetic platform, and multiple functional miniproteins from the platform, enabling the development of new solutions for every unique TF. Additionally, lead functional miniproteins that have emerged from our screens, possess great potentials for direct inhibition of Myc DNA binding, representing a critical first step towards overcoming the grand challenge of targeting Myc.

Chapter 1

Introduction

1.1 Transcription factors as targets for cancer therapy

Transcription factors (TFs) are proteins that bind to specific DNA promoter and enhancer sequences and subsequently control the transcription of genetic information from DNA to messenger RNA (mRNA), resulting in either activation or repression of target genes[1]. Although there are over 1,600 human transcription factors, their protein sequences, regulatory regions, and physiological roles are often deeply conserved[2, 3]. Given the central role served by TFs in all aspects of cellular function, aberrant TF activity is widely and unambiguously implicated in human disease. In 2002, Darnell identified a limited list of transcription factors are overactive in most human cancer cells[4]. This has been followed by a plethora of studies elucidating the important role that the deregulation of TFs plays in cancer[5]. Many cancers are hallmarked by direct genetic alteration of TFs by amplification, deletion, translocation or mutation[6, 7]. Additionally, cancers that do not harbor these direct alterations to TFs invariably rely on dysregulated upstream signaling pathways that ultimately impinge on TF function and gene expression programs[6]. As such, TFs are the most direct and hopeful targets for cancer therapy. The increased validation of TF targets has made. It clear that interrogation of their activity has substantial therapeutic potential, not only for cancer but for other human diseases as well.

Notwithstanding their great promise, TFs have proven to be particularly difficult to manipulate pharmacologically and been historically regarded as "undruggable". This view arises from the challenges associated with targeting either the protein-protein or protein-DNA interactions that mediate their function. The smoother surface and lack of more tractable binding pockets make them difficult targets for any small molecule drug development. In fact, only a small subset of TFs, nuclear hormone receptors, including estrogen receptor, androgen receptor, retinoic acid receptor and glucocorticoid receptor, have been routinely targeted by cell-permeable small molecules[8], due to the presence of binding sites for endogenous effector metabolites. On the other hand, large biomolecules, such as antibodies, are capable of blocking certain protein-protein interfaces, but their therapeutic values in targeting TFs are severely compromised, due to their large molecular size and hence their inability to pass through cell membranes. The significant gap in our therapeutic tools has consequently led to the historical lack of success by existing classes of molecules.

Meanwhile, multiple efforts have been made to target TFs via other avenues, which include modulating the expression of TFs, blocking essential protein-protein interactions in the TF complex, and promoting their proteasomal degradation. Nevertheless, considering that DNA binding is of vital importance to TFs' function and direct inhibition of such interactions would sequester TFs from their targets, there is a frustrating paucity of inhibitors that directly target the DNA binding of TFs. To date, the only class of synthetic molecules that has demonstrated a general capacity to directly target DNA, and by extension TF activity, is the engineered "polyamide" class introduced by Dervan et al. in the 1990s[9]. These molecules have been validated as direct competitors to TF-DNA binding interactions[10], as well as capable of reactivating gene expression near repetitive target sites in the genome[11]. However, polyamides function as minor groove DNA

binders [12, 13] in both open and closed chromatin, whereas most TFs employ archetypal DNA-interacting motifs of an α-helix oriented in the major groove. This discrepancy in mechanism is likely to lead to undesirable cellular activities, such as off-target effects[14]. Since their initial development, no significant progress has been made to demonstrate their clinical applicability. Therefore, this unifying common feature of major groove binding, which requires defined secondary, tertiary, and/or quaternary structures, must be considered in future ligand designs.

1.2 Myc is a proto-oncogenic transcription factor

1.2.1 Regulation of Myc/Max/Mad network and its role in cancer progression

Of the "undruggable" TFs that contain a widely conserved DNA binding domain, the c-Myc oncoprotein (or simply Myc hereafter) represents a prototypical and notorious driver of proliferation in numerous cancers[15, 16]. Myc is a part of the Myc/Max/Mad transcriptional network and belongs to a family of TFs that possess the basic helix-loop-helix (bHLH) motif[17, 18]. All three proteins form heterodimer or homodimer with Max via interactions between the leucine zipper domains. The dimerized bHLH domains recognize E-box DNA containing the sequence 5'-CANNTG-3'. Despite their structural similarities, Myc, Mad and Max trigger different effects in gene expression. In particular, upon binding to E-box, Myc activates transcription by associating its N-terminal transactivation domain with TRRAP, a member of the HAT complex SAGA[19]. Conversely, the recruitment of yeast transcriptional repressor Sin3 to the N-terminus of Mad, or lack of any N-terminal domains in Max, results in either active or passive transcriptional repression[20, 21].

As a TF, Myc acts as a global regulator of transcription and influences up to 15% of genes[22]. As such, Myc is often considered a master regulator of cellular functions. Over the past

two decades, development of high throughput screening techniques such as microarrays[23, 24], serial analysis of gene expression (SAGE)[25], and chromatin immunoprecipitation (ChIP)[26, 27] have generated a plethora of information regarding Myc responsive and target genes. In particular, Myc regulates specific classes of genes that are involved in cell cycle regulation, metabolism, ribosome biogenesis, cell adhesion, protein synthesis, mitochondrial function, and apoptosis[22, 28]. However, the target genes that distinguish between physiologic and tumorigenic functions of Myc remain largely elusive.

Table 1.1 Select Myc target genes and their functions

Gene	Function	References
ARF	Apoptosis	29
BIN1	DNA repair	30
CCNB1	Cell cycle	31
CCND1	Cell cycle	32, 33
CCND2	Cell cycle	34
CDK4	Cell cycle	35
COL1A1	Adhesion	36
E2F	Transcription	37
eIF4E	Translation	38
GLUT1	Metabolism	39
HK2	Metabolism	40
LDHA	Growth and metabolism	41
RPL proteins	Ribosomal biogenesis and protein synthesis	42, 43
SHMT	Metabolism	44

Since its discovery in the 1970s[45-47], numerous studies have been conducted to understand the Myc signaling pathway and its role in human cancer[15]. Under normal circumstances, the *MYC* oncogene is highly regulated as a downstream transduction pathway for receptor signaling. In particular, WNT receptor[48] and receptor tyrosine kinase[49] indirectly activate *MYC* expression via the β-catenin and MAPK pathway, respectively, whereas its expression is directly induced by T-cell receptors[50] or indirectly suppressed by TGF-β receptor[51]. The expression of *MYC* produces the Myc protein, a TF that regulates transcription of genes involved in cell growth and proliferation (Figure 1.1A). In case of an acute expression of *MYC*, checkpoints including Arf and p53 are

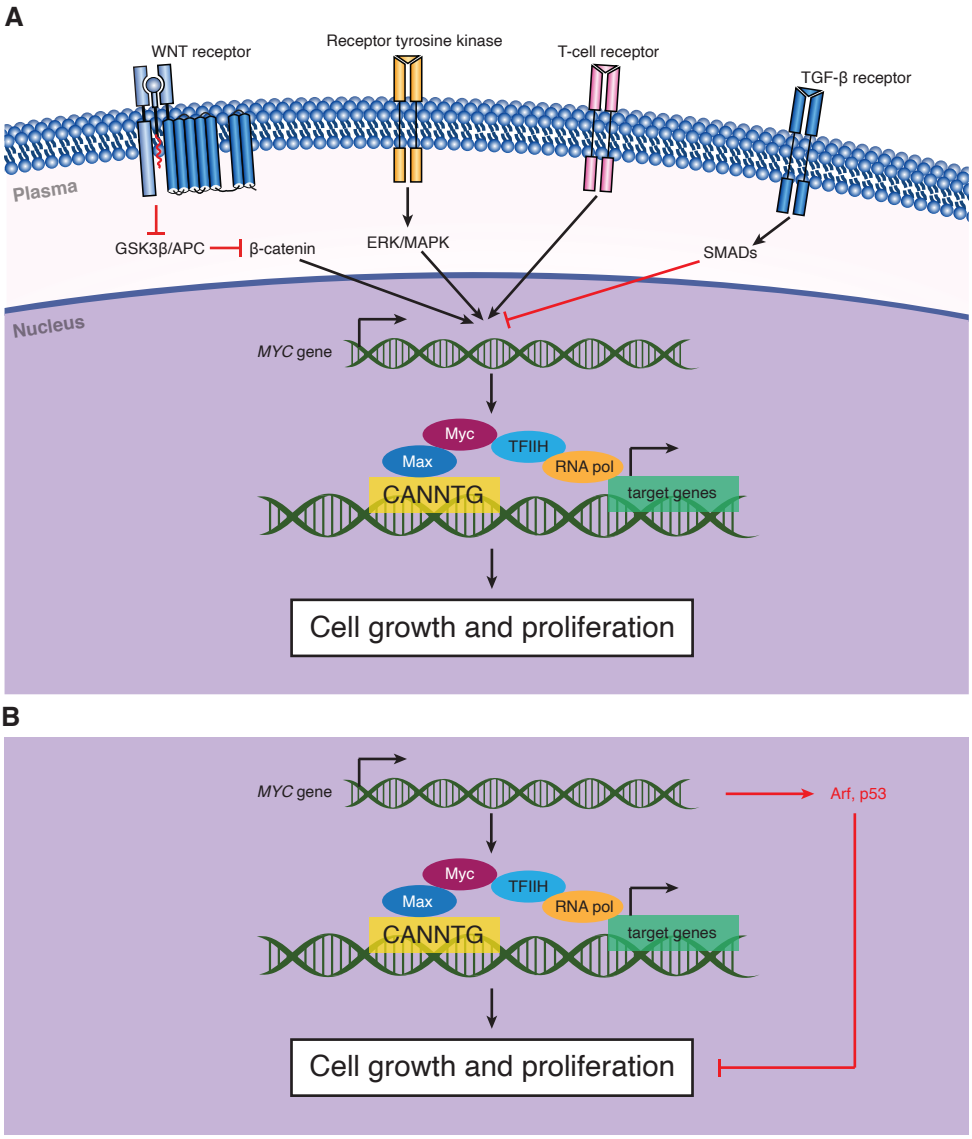


Figure 1.1 Myc regulates cell growth and proliferation

(**A**) The *MYC* oncogene is highly regulated by a series in normal cells. (**B**) Acute Myc expression triggers checkpoint regulation by tumor repressors Arf and p53. Loss of such checkpoints leads to tumorigenesis.

triggered, acting as tumor suppressors. In tumor cells, however, Myc frequently becomes independent of such regulations. For example, Myc-induced lymphomagenesis usually leads to a loss of functional checkpoint through mutations of p53 or Arf, unlocking Myc's full tumorigenic potential[29, 52] (Figure 1.1B). Myc deregulation can occur at any given stage of its expression, including gene amplification via retroviral insertion[53] or chromosomal translocation[54], mRNA stabilization, Myc protein stabilization due to mutations, and loss of upstream regulators. In fact,

Myc oncogene is deregulated in up to 70% of human cancers[15], and this deregulation is frequently associated with poor prognosis and unfavorable patient survival.

1.2.2 Therapeutic strategies targeting Myc and their limitations

Due to its deregulation and crucial role in most, if not all, human cancers, Myc is widely recognized as an ideal target for cancer therapy. However, just as many other TFs, targeting Myc poses significant pharmacological challenges and no Myc inhibitors have reached the clinic to date. Aside from the abovementioned technical issues, historical claims that Myc inhibition would cause catastrophic side effects such as inhibition of proliferation of normal tissues dominated the field for long until recent studies suggested otherwise. Particularly, in 2008, Soucek et al. used a transgenic mouse model expressing Omomyc, a dominant negative mutant, as way to model systemic Myc inhibition. In this proof-of-concept study, they demonstrated that Myc inhibition by Omomyc triggers rapid tumor regression while being reversable and well tolerated by normal tissues[55]. Since then, research interests in Myc inhibition have grown rapidly[56].

To date, numerous efforts have been made to target multiple stages in the Myc signaling pathway. These strategies generally fall into three large categories: blocking Myc expression, interfering with its transactivation, and inducing its degradation (Table 1.2).

Blocking the expression of Myc expression aims to reduce its production and can be achieved by targeting its transcription or translation, either in a direct or indirect manner. Examples for direct inhibition include G-quadruplex stabilizers[57], antisense oligonucleotides, and siRNA. G-quadruplex stabilizers are a class of small molecule ligands that stabilize G-quadruplexes in the *MYC* gene and thus repress its transcription. By contrast, antisense oligonucleotides[58, 59] and siRNA[60, 61] prevent Myc translation by interfering with the *MYC* mRNA.

On the other hand, indirect inhibition of Myc expression instead focuses on targeting other regulatory factors of Myc. For example, in 2011, Delmore et al. conducted a study using JQ1 for BET bromodomain inhibition in multiple myeloma and unexpectedly observed the downregulation of *MYC* transcription. This study identified *BRD4* as a regulator of *MYC* expression, and its inhibition led to robust inhibition of Myc, causing cell-cycle arrest, and cellular senescence, as well as resulting in significant anti-tumor activity in mouse models of multiple myeloma, and xenograft models of Burkitt's lymphoma and acute myeloid leukemia, diseases in which the *MYC* gene is amplified[62].

Interfering with Myc transactivation instead focuses on interrogating the protein-protein or protein-DNA interactions in the Myc/Max TF, which has led to a handful of developments of small molecule inhibitors. Lead candidates include 10058-F4, 10074-G5, and JY-3-094, which reportedly interfere with Myc/Max heterodimerization[63-65], and KSI-3716, which specifically inhibit Myc binding to DNA[66, 67]. Although a wide range of efficacy has been reported *in vitro*, these small molecule inhibitors' applicability *in vivo* has so far been limited by poor bioavailability, rapid metabolism, inadequate target site penetration, unclear inhibitory mechanism, and lack of specificity and selectivity in cells.

In addition to the development of small molecule inhibitors, miniproteins derived from the Myc family, as opposed to small molecule inhibitors, have demonstrated great efficacy *in vivo*. Particularly, Omomyc has been well characterized for its *in vivo* efficacy. Omomyc is a dominant negative mutant that comprises the bHLH domain of Myc with four key mutations that alter its dimerization specificity. Aside from dimerizing with Max, it also heterodimerizes with Myc as well as homodimerizes. As a result, Omomyc is capable of inhibiting Myc/Max DNA binding by sequestering Myc away from DNA and occupying the E-box with transcriptionally inactive

dimers[55, 68]. Multiple studies in mouse models of cancer demonstrated Omomyc's therapeutic impact in different types of cancer with minimal side effects[69-72]. Despite its proven efficacy, the large size of Omomyc (>60 amino acids) poses significant challenges on its cellular uptake, limiting their potential applicability other than gene therapy. Intriguingly, a recent study reveals that Omomyc penetrates lung cancer cells and reverts Myc-driven transcription, possibly via an adenosine triphosphate-dependent uptake mechanism[73], suggesting such DNA-binding miniproteins in conjunction with robust cellular uptake can deliver great pharmacological potentials.

Efforts to induce the degradation of Myc have been made by targeting ubiquitinases or phosphatases. Examples in this category include inhibiting the deubiquitinases that help stabilize Myc, such as USP28, USP38, and USP36[74], or triggering proteasomal degradation by inducing FBW7-mediated Myc breakdown[75]. Alternatively, inhibitors of SET and CIP2A have also been explored to activate the tumor suppressor protein phosphatase 2A[76, 77], which targets Ser62 of Myc and causes its destabilization[78].

Table 1.2 Existing efforts to target Myc

Strategy	Mechanism	Examples	References
inhibiting *MYC* expression	prevent *MYC* transcription	G-quadruplex stabilizers such as CX-3543	57, 79
	prevent *MYC* translation	antisense oligonucleotides; siRNA & microRNA	58-61
	indirectly prevent *MYC* transcription by inhibiting BET bromodomains	JQ1 and other derivatives	62
inhibiting Myc transactivation	interfere with protein-protein interaction between Myc and Max	small molecules such as 10058-F4, 10074-G5, and JY-3-094	63-65
	interfere with DNA binding of Myc	KSI-3716	66, 67
	disrupt Myc function with its dominant negative form	Omomyc	55, 68-73
inducing Myc degradation	target deubiquitinases that stabilize Myc	USP28, USP38, and USP36 inhibitors	74
	target Aurora family proteins that protect Myc from proteasomal degradation	MLN8237, MLN8054	80, 81
	activate protein phosphatase 2A	SET & CIP2A inhibitors	76, 77

Although remarkable progress has been achieved in the field of Myc, there is no reason to stay complacent. Many of the abovementioned approaches did not progress in early clinical studies, and other active areas are accompanied with intrinsic issues, ranging from low bioavailability and *in vivo* efficacy, lack of clear understanding mechanism, and poor cellular uptake. In particular, there is a paucity of inhibitors that target the protein-DNA interaction, a strategy that has shown great *in vivo* applicability in multiple transgenic models. As such, we believe more systemic and modular attempts need be made in this area in order to better identify opportunities to eventually overcome the grand challenge of targeting Myc.

1.3 Therapeutic potential of synthetic DNA-binding peptides and miniproteins

Containing our discussion at the end of 1.1, deliberately designed molecules directly targeting TFs' defined structures have potentials of successful direct inhibition of DNA binding of TFs. Synthetic biologics are an ideal solution to the creation of such synthetic scaffolds. As the name suggests, synthetic biologics are a class of scaffolds with synthetic accessibility. Emerged over the past two decades, synthetic biologics have demonstrated potentials of modulating a wide range of protein-protein interactions without compromising cell permeability, fulfilling the unmet therapeutic needs[82-85]. Their therapeutic potentials are the results of their distinct properties—The stabilized scaffolds, which often mimic large biomolecules, enable direct interrogation of protein-protein interactions not accessible by small molecules. On the other hand, their moderate, tunable size makes them capable of penetrating cell membranes via endocytosis or other mechanisms. An outstanding candidate among them is stapled peptides, developed by Grubbs[86] and Verdine[87] in the late 1990s and early 2000s. Featuring a covalent linkage ("staple") between two side chains, these peptide macrocycles have been shown to have improved structural stability, target binding

affinity and cell penetrability, making this stabilization strategy particularly relevant for clinical applications. The most classic example among all stapling techniques[88] is hydrocarbon stapling, which involves a ring-closing metathesis between two olefin-containing residues. As a proof-of-concept study, in 2004, Walensky et al. developed a series of BH3 peptides with stabilized α-helix structure and further demonstrated that lead peptides showed positive cell uptake, bound the intracellular BCL-2 family targets and induced the apoptosis of cancer cells[89]. Since then, the stapling technology has been applied to a wide variety of peptide families, enabling the study of many other PPIs using stapled peptides including cancer targets such as p53[90], β-catenin[91], NOTCH1[92], MCL-1 BH3[93], PUMA BH3[94], and RAB25[95]. Lead candidates of p53 stapled peptides inhibiting MDM2/MDMX have been actively investigated in multiple clinical trials by Aileron Therapeutics[82], proving their strong therapeutic potentials.

In spite of the early successes in targeting oncogenic protein-protein interactions by stapled peptides, very few studies have reported their direct interrogation of protein-DNA interfaces of TFs[96, 97]. One major caveat is that many TF functions involve cooperative DNA binding by multiple domains/proteins, which is intractable even by traditional peptidomimetic strategy. In the case of bHLH TFs, the added complexity from both protein-protein and protein-DNA interactions makes it particular challenging for proper inhibition by a single stapled peptide. In 2016, Edwards et al. designed a series of stapled peptides in an attempt to target OLIG2, a member of bHLH TFs[98]. Despite an exhaustive list of stapled α-helices that cover the entire HLH domain, inhibition of OLIG2-DNA interactions was not achieved. This study revealed key pitfalls to consider in developing such inhibitors and demonstrated the necessity for better designs.

Alternatively, in our search for potential inhibitors that target the protein-DNA interface, we have looked into the core architecture of TFs. Seminal studies have demonstrated that many

DNA binding domains are inherently modular, which permits retention of potent and specific DNA binding activity when domains are separated from the rest of their parent TF[99], or when grafted on to entirely new proteins[100, 101]. These early studies suggest that molecules that mimic the structural features of DNA binding domains could serve as minimal TF mimetics capable of interfering with or replacing endogenous TF activity. Based on our initial analysis, the outstanding gap between a single α-helix and full-length, functional proteins is the presence of multiple structural components such as leucine zipper, which do not directly interact with DNA but instead help form a proper tertiary structure that orients the DNA-binding helices for cooperative, sequence-specific DNA recognition. Therefore, we believe that more sophisticated designs encompassing both secondary and tertiary structures into consideration should theoretically proceed further along towards recapitulating the activity of full bHLH proteins and thus be capable of inhibiting the protein-DNA interaction. The resulting DNA-binding peptides or miniproteins, including their design, synthesis and deep understanding of their activities, have thus drawn our strong interest. Additionally, the development of such synthetic DNA-binding peptides and miniproteins may also be modular, wherein variation of DNA-binding residues within a core structural architecture could allow for specific targeting of alternative DNA sequences, thereby serving as a platform that sidesteps the development of new solutions for every unique TF.

1.4 Scope of this dissertation

In summary, synthetic biologics, particularly stabilized peptides and miniproteins, have huge potential for direct inhibition of TF-DNA due to their versatility and have drawn my particular interest during my doctoral studies. In this dissertation, I will primarily focus on three different areas:

1) Designing a series miniprotein scaffolds that attempt to mimic the DNA-binding helices of Myc/Max and other bHLH TFs.

2) Developing a modular platform for the de novo synthesis of miniprotein scaffolds with novel, orthogonal synthetic tools.

3) Establishing an assay pipeline to understand the *in vitro* and cellular activity of such synthetic miniproteins, extrapolating their structure activity relationship as guidelines for optimization and further designs.

Chapter 2 presents a proof-of-concept study of stabilization of both secondary and tertiary structures derived from the two basic helices of Myc/Max. These synthetic DNA-binding domains recapitulate full-length protein activity *in vitro* and possess positive cellular activities.

Chapter 3 uses guided approaches to optimize the design around the core structure of synthetic DNA-binding domains and further develops orthogonal chemistries in our platform to expand the scope of synthetic targets.

Chapter 4 presents an alternative scaffold that features proper alignment of basic and leucine zipper helices. These self-associating synthetic transcription factors show superior binding activity and recognize DNA in a similar fashion to bHLH proteins.

Chapter 5 describes a series of attempts to synthesize the dimeric synthetic transcription factors, the largest stabilized miniprotein scaffold we have ever made. Early results show significant improvement in activity and stability over their monomeric counterparts.

References

1. Lambert, S. A.; Jolma, A.; Campitelli, L. F.; Das, P. K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T. R.; Weirauch, M. T., The Human Transcription Factors. *Cell* **2018,** *172* (4), 650-665.

2. Bejerano, G.; Pheasant, M.; Makunin, I.; Stephen, S.; Kent, W. J.; Mattick, J. S.; Haussler, D., Ultraconserved elements in the human genome. *Science* **2004,** *304* (5675), 1321-5.

3. Carroll, S. B., Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **2008,** *134* (1), 25-36.

4. Darnell, J. E., Jr., Transcription factors as targets for cancer therapy. *Nat Rev Cancer* **2002,** *2* (10), 740-9.

5. Lee, T. I.; Young, R. A., Transcriptional regulation and its misregulation in disease. *Cell* **2013,** *152* (6), 1237-51.

6. Bradner, J. E.; Hnisz, D.; Young, R. A., Transcriptional Addiction in Cancer. *Cell* **2017,** *168* (4), 629-643.

7. Bushweller, J. H., Targeting transcription factors in cancer - from undruggable to reality. *Nat Rev Cancer* **2019,** *19* (11), 611-624.

8. Burris, T. P.; Solt, L. A.; Wang, Y.; Crumbley, C.; Banerjee, S.; Griffett, K.; Lundasen, T.; Hughes, T.; Kojetin, D. J., Nuclear receptors and their selective pharmacologic modulators. *Pharmacol Rev* **2013,** *65* (2), 710-78.

9. Trauger, J. W.; Baird, E. E.; Dervan, P. B., Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **1996,** *382* (6591), 559-61.

10. Kurmis, A. A.; Dervan, P. B., Sequence specific suppression of androgen receptor-DNA binding in vivo by a Py-Im polyamide. *Nucleic Acids Res* **2019,** *47* (8), 3828-3835.

11. Erwin, G. S.; Grieshop, M. P.; Ali, A.; Qi, J.; Lawlor, M.; Kumar, D.; Ahmad, I.; McNally, A.; Teider, N.; Worringer, K.; Sivasankaran, R.; Syed, D. N.; Eguchi, A.; Ashraf, M.; Jeffery, J.; Xu, M.; Park, P. M. C.; Mukhtar, H.; Srivastava, A. K.; Faruq, M.; Bradner, J. E.; Ansari, A. Z., Synthetic transcription elongation factors license transcription across repressive chromatin. *Science* **2017,** *358* (6370), 1617-1622.

12. Chenoweth, D. M.; Dervan, P. B., Allosteric modulation of DNA by small molecules. *Proc Natl Acad Sci U S A* **2009,** *106* (32), 13175-9.

13. Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C., Structural basis for G.C recognition in the DNA minor groove. *Nat Struct Biol* **1998,** *5* (2), 104-9.

14. Synold, T. W.; Xi, B.; Wu, J.; Yen, Y.; Li, B. C.; Yang, F.; Phillips, J. W.; Nickols, N. G.; Dervan, P. B., Single-dose pharmacokinetic and toxicity analysis of pyrrole-imidazole polyamides in mice. *Cancer Chemother Pharmacol* **2012,** *70* (4), 617-25.

15. Dang, C. V., MYC on the path to cancer. *Cell* **2012,** *149* (1), 22-35.

16. Schaub, F. X.; Dhankani, V.; Berger, A. C.; Trivedi, M.; Richardson, A. B.; Shaw, R.; Zhao, W.; Zhang, X.; Ventura, A.; Liu, Y.; Ayer, D. E.; Hurlin, P. J.; Cherniack, A. D.; Eisenman, R. N.; Bernard, B.; Grandori, C.; Cancer Genome Atlas, N., Pan-cancer Alterations of the MYC Oncogene and Its Proximal Network across the Cancer Genome Atlas. *Cell Syst* **2018,** *6* (3), 282-300 e2.

17. Amati, B.; Land, H., Myc-Max-Mad: a transcription factor network controlling cell cycle progression, differentiation and death. *Curr Opin Genet Dev* **1994,** *4* (1), 102-8.

18. Grandori, C.; Cowley, S. M.; James, L. P.; Eisenman, R. N., The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu Rev Cell Dev Biol* **2000,** *16*, 653-99.

19. McMahon, S. B.; Van Buskirk, H. A.; Dugan, K. A.; Copeland, T. D.; Cole, M. D., The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* **1998,** *94* (3), 363-74.

20. Ayer, D. E.; Lawrence, Q. A.; Eisenman, R. N., Mad-Max transcriptional repression is mediated by ternary complex formation with mammalian homologs of yeast repressor Sin3. *Cell* **1995,** *80* (5), 767-76.

21. James, L.; Eisenman, R. N., Myc and Mad bHLHZ domains possess identical DNA-binding specificities but only partially overlapping functions in vivo. *Proc Natl Acad Sci U S A* **2002,** *99* (16), 10429-34.

22. Dang, C. V.; O'Donnell, K. A.; Zeller, K. I.; Nguyen, T.; Osthus, R. C.; Li, F., The c-Myc target gene network. *Semin Cancer Biol* **2006,** *16* (4), 253-64.

23. Schuhmacher, M.; Kohlhuber, F.; Holzel, M.; Kaiser, C.; Burtscher, H.; Jarsch, M.; Bornkamm, G. W.; Laux, G.; Polack, A.; Weidle, U. H.; Eick, D., The transcriptional program of a human B cell line in response to Myc. *Nucleic Acids Res* **2001,** *29* (2), 397-406.

24. Coller, H. A.; Grandori, C.; Tamayo, P.; Colbert, T.; Lander, E. S.; Eisenman, R. N.; Golub, T. R., Expression analysis with oligonucleotide microarrays reveals that MYC regulates genes involved in growth, cell cycle, signaling, and adhesion. *Proc Natl Acad Sci U S A* **2000,** *97* (7), 3260-5.

25. Menssen, A.; Hermeking, H., Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. *Proc Natl Acad Sci U S A* **2002,** *99* (9), 6274-9.

26. Zeller, K. I.; Zhao, X.; Lee, C. W.; Chiu, K. P.; Yao, F.; Yustein, J. T.; Ooi, H. S.; Orlov, Y. L.; Shahab, A.; Yong, H. C.; Fu, Y.; Weng, Z.; Kuznetsov, V. A.; Sung, W. K.; Ruan, Y.; Dang, C. V.; Wei, C. L., Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc Natl Acad Sci U S A* **2006,** *103* (47), 17834-9.

27. Haggerty, T. J.; Zeller, K. I.; Osthus, R. C.; Wonsey, D. R.; Dang, C. V., A strategy for identifying transcription factor binding sites reveals two classes of genomic c-Myc target sites. *Proc Natl Acad Sci U S A* **2003,** *100* (9), 5313-8.

28. Dang, C. V., c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Mol Cell Biol* **1999,** *19* (1), 1-11.

29. Zindy, F.; Eischen, C. M.; Randle, D. H.; Kamijo, T.; Cleveland, J. L.; Sherr, C. J.; Roussel, M. F., Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. *Genes Dev* **1998,** *12* (15), 2424-33.

30. Elliott, K.; Sakamuro, D.; Basu, A.; Du, W.; Wunner, W.; Staller, P.; Gaubatz, S.; Zhang, H.; Prochownik, E.; Eilers, M.; Prendergast, G. C., Bin1 functionally interacts with Myc and inhibits cell proliferation via multiple mechanisms. *Oncogene* **1999,** *18* (24), 3564-73.

31. Yin, X. Y.; Grove, L.; Datta, N. S.; Katula, K.; Long, M. W.; Prochownik, E. V., Inverse regulation of cyclin B1 by c-Myc and p53 and induction of tetraploidy by cyclin B1 overexpression. *Cancer Res* **2001,** *61* (17), 6487-93.

32. Daksis, J. I.; Lu, R. Y.; Facchini, L. M.; Marhin, W. W.; Penn, L. J., Myc induces cyclin D1 expression in the absence of de novo protein synthesis and links mitogen-stimulated signal transduction to the cell cycle. *Oncogene* **1994,** *9* (12), 3635-45.

33. Philipp, A.; Schneider, A.; Vasrik, I.; Finke, K.; Xiong, Y.; Beach, D.; Alitalo, K.; Eilers, M., Repression of cyclin D1: a novel function of MYC. *Mol Cell Biol* **1994,** *14* (6), 4032-43.

34. Bouchard, C.; Dittrich, O.; Kiermaier, A.; Dohmann, K.; Menkel, A.; Eilers, M.; Luscher, B., Regulation of cyclin D2 gene expression by the Myc/Max/Mad network: Myc-dependent TRRAP recruitment and histone acetylation at the cyclin D2 promoter. *Genes Dev* **2001,** *15* (16), 2042-7.

35. Hermeking, H.; Rago, C.; Schuhmacher, M.; Li, Q.; Barrett, J. F.; Obaya, A. J.; O'Connell, B. C.; Mateyak, M. K.; Tam, W.; Kohlhuber, F.; Dang, C. V.; Sedivy, J. M.; Eick, D.; Vogelstein, B.; Kinzler, K. W., Identification of CDK4 as a target of c-MYC. *Proc Natl Acad Sci U S A* **2000,** *97* (5), 2229-34.

36. Tavtigian, S. V.; Zabludoff, S. D.; Wold, B. J., Cloning of mid-G1 serum response genes and identification of a subset regulated by conditional myc expression. *Mol Biol Cell* **1994,** *5* (3), 375-88.

37. Leone, G.; Sears, R.; Huang, E.; Rempel, R.; Nuckolls, F.; Park, C.-H.; Giangrande, P.; Wu, L.; Saavedra, H. I.; Field, S. J.; Thompson, M. A.; Yang, H.; Fujiwara, Y.; Greenberg, M. E.; Orkin, S.; Smith, C.; Nevins, J. R., Myc Requires Distinct E2F Activities to Induce S Phase and Apoptosis. *Molecular Cell* **2001,** *8* (1), 105-113.

38. Schmidt, E. V., The role of c-myc in regulation of translation initiation. *Oncogene* **2004,** *23* (18), 3217-21.

39. Osthus, R. C.; Shim, H.; Kim, S.; Li, Q.; Reddy, R.; Mukherjee, M.; Xu, Y.; Wonsey, D.; Lee, L. A.; Dang, C. V., Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. *J Biol Chem* **2000,** *275* (29), 21797-800.

40. Kim, J. W.; Gao, P.; Liu, Y. C.; Semenza, G. L.; Dang, C. V., Hypoxia-inducible factor 1 and dysregulated c-Myc cooperatively induce vascular endothelial growth factor and metabolic switches hexokinase 2 and pyruvate dehydrogenase kinase 1. *Mol Cell Biol* **2007,** *27* (21), 7381-93.

41. Shim, H.; Dolde, C.; Lewis, B. C.; Wu, C. S.; Dang, G.; Jungmann, R. A.; Dalla-Favera, R.; Dang, C. V., c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *Proc Natl Acad Sci U S A* **1997,** *94* (13), 6658-63.

42. van Riggelen, J.; Yetil, A.; Felsher, D. W., MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer* **2010,** *10* (4), 301-9.

43. Ruggero, D., The role of Myc-induced protein synthesis in cancer. *Cancer Res* **2009,** *69* (23), 8839-43.

44. Nikiforov, M. A.; Chandriani, S.; O'Connell, B.; Petrenko, O.; Kotenko, I.; Beavis, A.; Sedivy, J. M.; Cole, M. D., A functional screen for Myc-responsive genes reveals serine hydroxymethyltransferase, a major source of the one-carbon unit for cell metabolism. *Mol Cell Biol* **2002,** *22* (16), 5793-800.

45. Duesberg, P. H.; Vogt, P. K., Avian acute leukemia viruses MC29 and MH2 share specific RNA sequences: evidence for a second class of transforming genes. *Proc Natl Acad Sci U S A* **1979,** *76* (4), 1633-7.

46. Hu, S. S.; Lai, M. M.; Vogt, P. K., Genome of avian myelocytomatosis virus MC29: analysis by heteroduplex mapping. *Proc Natl Acad Sci U S A* **1979,** *76* (3), 1265-8.

47. Sheiness, D.; Bishop, J. M., DNA and RNA from uninfected vertebrate cells contain nucleotide sequences related to the putative transforming gene of avian myelocytomatosis virus. *J Virol* **1979,** *31* (2), 514-21.

48. He, T. C.; Sparks, A. B.; Rago, C.; Hermeking, H.; Zawel, L.; da Costa, L. T.; Morin, P. J.; Vogelstein, B.; Kinzler, K. W., Identification of c-MYC as a target of the APC pathway. *Science* **1998,** *281* (5382), 1509-12.

49. Zhang, W.; Liu, H. T., MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Res* **2002,** *12* (1), 9-18.

50. Weng, A. P.; Millholland, J. M.; Yashiro-Ohtani, Y.; Arcangeli, M. L.; Lau, A.; Wai, C.; Del Bianco, C.; Rodriguez, C. G.; Sai, H.; Tobias, J.; Li, Y.; Wolfe, M. S.; Shachaf, C.; Felsher, D.; Blacklow, S. C.; Pear, W. S.; Aster, J. C., c-Myc is an important direct

target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. *Genes Dev* **2006,** *20* (15), 2096-109.

51. Yagi, K.; Furuhashi, M.; Aoki, H.; Goto, D.; Kuwano, H.; Sugamura, K.; Miyazono, K.; Kato, M., c-myc is a downstream target of the Smad pathway. *J Biol Chem* **2002,** *277* (1), 854-61.

52. Eischen, C. M.; Weber, J. D.; Roussel, M. F.; Sherr, C. J.; Cleveland, J. L., Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes Dev* **1999,** *13* (20), 2658-69.

53. Akagi, K.; Suzuki, T.; Stephens, R. M.; Jenkins, N. A.; Copeland, N. G., RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res* **2004,** *32* (Database issue), D523-7.

54. Dalla-Favera, R.; Bregni, M.; Erikson, J.; Patterson, D.; Gallo, R. C.; Croce, C. M., Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc Natl Acad Sci U S A* **1982,** *79* (24), 7824-7.

55. Soucek, L.; Whitfield, J.; Martins, C. P.; Finch, A. J.; Murphy, D. J.; Sodir, N. M.; Karnezis, A. N.; Swigart, L. B.; Nasi, S.; Evan, G. I., Modelling Myc inhibition as a cancer therapy. *Nature* **2008,** *455* (7213), 679-83.

56. Whitfield, J. R.; Beaulieu, M. E.; Soucek, L., Strategies to Inhibit Myc and Their Clinical Applicability. *Front Cell Dev Biol* **2017,** *5*, 10.

57. Brooks, T. A.; Hurley, L. H., Targeting MYC Expression through G-Quadruplexes. *Genes Cancer* **2010,** *1* (6), 641-649.

58. Devi, G. R.; Beer, T. M.; Corless, C. L.; Arora, V.; Weller, D. L.; Iversen, P. L., In vivo bioavailability and pharmacokinetics of a c-MYC antisense phosphorodiamidate morpholino oligomer, AVI-4126, in solid tumors. *Clin Cancer Res* **2005,** *11* (10), 3930-8.

59. Webb, M. S.; Tortora, N.; Cremese, M.; Kozlowska, H.; Blaquiere, M.; Devine, D. V.; Kornbrust, D. J., Toxicity and toxicokinetics of a phosphorothioate oligonucleotide against the c-myc oncogene in cynomolgus monkeys. *Antisense Nucleic Acid Drug Dev* **2001,** *11* (3), 155-63.

60. Conde, J.; Tian, F.; Hernandez, Y.; Bao, C.; Cui, D.; Janssen, K. P.; Ibarra, M. R.; Baptista, P. V.; Stoeger, T.; de la Fuente, J. M., In vivo tumor targeting via nanoparticle-mediated therapeutic siRNA coupled to inflammatory response in lung cancer mouse models. *Biomaterials* **2013,** *34* (31), 7744-53.

61. Zhu, Q.; Feng, C.; Liao, W.; Zhang, Y.; Tang, S., Target delivery of MYCN siRNA by folate-nanoliposomes delivery system in a metastatic neuroblastoma model. *Cancer Cell Int* **2013,** *13* (1), 65.

62. Delmore, J. E.; Issa, G. C.; Lemieux, M. E.; Rahl, P. B.; Shi, J.; Jacobs, H. M.; Kastritis, E.; Gilpatrick, T.; Paranal, R. M.; Qi, J.; Chesi, M.; Schinzel, A. C.; McKeown, M. R.; Heffernan, T. P.; Vakoc, C. R.; Bergsagel, P. L.; Ghobrial, I. M.; Richardson, P. G.; Young, R. A.; Hahn, W. C.; Anderson, K. C.; Kung, A. L.; Bradner, J. E.; Mitsiades, C. S., BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **2011,** *146* (6), 904-17.

63. Yin, X.; Giap, C.; Lazo, J. S.; Prochownik, E. V., Low molecular weight inhibitors of Myc-Max interaction and function. *Oncogene* **2003,** *22* (40), 6151-9.

64. Yap, J. L.; Wang, H.; Hu, A.; Chauhan, J.; Jung, K. Y.; Gharavi, R. B.; Prochownik, E. V.; Fletcher, S., Pharmacophore identification of c-Myc inhibitor 10074-G5. *Bioorg Med Chem Lett* **2013,** *23* (1), 370-4.

65. Wang, H.; Chauhan, J.; Hu, A.; Pendleton, K.; Yap, J. L.; Sabato, P. E.; Jones, J. W.; Perri, M.; Yu, J.; Cione, E.; Kane, M. A.; Fletcher, S.; Prochownik, E. V., Disruption of Myc-Max heterodimerization with improved cell-penetrating analogs of the small molecule 10074-G5. *Oncotarget* **2013,** *4* (6), 936-47.

66. Jeong, K. C.; Kim, K. T.; Seo, H. H.; Shin, S. P.; Ahn, K. O.; Ji, M. J.; Park, W. S.; Kim, I. H.; Lee, S. J.; Seo, H. K., Intravesical instillation of c-MYC inhibitor KSI-3716 suppresses orthotopic bladder tumor growth. *J Urol* **2014,** *191* (2), 510-8.

67. Seo, H. K.; Ahn, K. O.; Jung, N. R.; Shin, J. S.; Park, W. S.; Lee, K. H.; Lee, S. J.; Jeong, K. C., Antitumor activity of the c-Myc inhibitor KSI-3716 in gemcitabine-resistant bladder cancer. *Oncotarget* **2014,** *5* (2), 326-37.

68. Savino, M.; Annibali, D.; Carucci, N.; Favuzzi, E.; Cole, M. D.; Evan, G. I.; Soucek, L.; Nasi, S., The action mechanism of the Myc inhibitor termed Omomyc may give clues on how to target Myc for cancer therapy. *PLoS One* **2011,** *6* (7), e22284.

69. Soucek, L.; Whitfield, J. R.; Sodir, N. M.; Masso-Valles, D.; Serrano, E.; Karnezis, A. N.; Swigart, L. B.; Evan, G. I., Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. *Genes Dev* **2013,** *27* (5), 504-13.

70. Sodir, N. M.; Swigart, L. B.; Karnezis, A. N.; Hanahan, D.; Evan, G. I.; Soucek, L., Endogenous Myc maintains the tumor microenvironment. *Genes Dev* **2011,** *25* (9), 907-16.

71. Annibali, D.; Whitfield, J. R.; Favuzzi, E.; Jauset, T.; Serrano, E.; Cuartas, I.; Redondo-Campos, S.; Folch, G.; Gonzalez-Junca, A.; Sodir, N. M.; Masso-Valles, D.; Beaulieu, M. E.; Swigart, L. B.; Mc Gee, M. M.; Somma, M. P.; Nasi, S.; Seoane, J.; Evan, G. I.; Soucek, L., Myc inhibition is effective against glioma and reveals a role for Myc in proficient mitosis. *Nat Commun* **2014,** *5*, 4632.

72. Galardi, S.; Savino, M.; Scagnoli, F.; Pellegatta, S.; Pisati, F.; Zambelli, F.; Illi, B.; Annibali, D.; Beji, S.; Orecchini, E.; Alberelli, M. A.; Apicella, C.; Fontanella, R. A.; Michienzi, A.; Finocchiaro, G.; Farace, M. G.; Pavesi, G.; Ciafre, S. A.; Nasi, S.,

Resetting cancer stem cell regulatory nodes upon MYC inhibition. *EMBO Rep* **2016,** *17* (12), 1872-1889.

73. Beaulieu, M. E.; Jauset, T.; Masso-Valles, D.; Martinez-Martin, S.; Rahl, P.; Maltais, L.; Zacarias-Fluck, M. F.; Casacuberta-Serra, S.; Serrano Del Pozo, E.; Fiore, C.; Foradada, L.; Cano, V. C.; Sanchez-Hervas, M.; Guenther, M.; Romero Sanz, E.; Oteo, M.; Tremblay, C.; Martin, G.; Letourneau, D.; Montagne, M.; Morcillo Alonso, M. A.; Whitfield, J. R.; Lavigne, P.; Soucek, L., Intrinsic cell-penetrating activity propels Omomyc from proof of concept to viable anti-MYC therapy. *Sci Transl Med* **2019,** *11* (484).

74. Sun, X. X.; Sears, R. C.; Dai, M. S., Deubiquitinating c-Myc: USP36 steps up in the nucleolus. *Cell Cycle* **2015,** *14* (24), 3786-93.

75. Huang, H. L.; Weng, H. Y.; Wang, L. Q.; Yu, C. H.; Huang, Q. J.; Zhao, P. P.; Wen, J. Z.; Zhou, H.; Qu, L. H., Triggering Fbw7-mediated proteasomal degradation of c-Myc by oridonin induces cell growth inhibition and apoptosis. *Mol Cancer Ther* **2012,** *11* (5), 1155-65.

76. Farrell, A. S.; Allen-Petersen, B.; Daniel, C. J.; Wang, X.; Wang, Z.; Rodriguez, S.; Impey, S.; Oddo, J.; Vitek, M. P.; Lopez, C.; Christensen, D. J.; Sheppard, B.; Sears, R. C., Targeting inhibitors of the tumor suppressor PP2A for the treatment of pancreatic cancer. *Mol Cancer Res* **2014,** *12* (6), 924-39.

77. Janghorban, M.; Farrell, A. S.; Allen-Petersen, B. L.; Pelz, C.; Daniel, C. J.; Oddo, J.; Langer, E. M.; Christensen, D. J.; Sears, R. C., Targeting c-MYC by antagonizing PP2A inhibitors in breast cancer. *Proc Natl Acad Sci U S A* **2014,** *111* (25), 9157-62.

78. Sears, R. C., The Life Cycle of C-Myc: From Synthesis to Degradation. *Cell Cycle* **2004,** *3* (9), 1131-1135.

79. Drygin, D.; Lin, A.; Bliesath, J.; Ho, C. B.; O'Brien, S. E.; Proffitt, C.; Omori, M.; Haddach, M.; Schwaebe, M. K.; Siddiqui-Jain, A.; Streiner, N.; Quin, J. E.; Sanij, E.; Bywater, M. J.; Hannan, R. D.; Ryckman, D.; Anderes, K.; Rice, W. G., Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res* **2011,** *71* (4), 1418-30.

80. Macarulla, T.; Cervantes, A.; Elez, E.; Rodriguez-Braun, E.; Baselga, J.; Rosello, S.; Sala, G.; Blasco, I.; Danaee, H.; Lee, Y.; Ecsedy, J.; Shinde, V.; Chakravarty, A.; Bowman, D.; Liu, H.; Eton, O.; Fingert, H.; Tabernero, J., Phase I study of the selective Aurora A kinase inhibitor MLN8054 in patients with advanced solid tumors: safety, pharmacokinetics, and pharmacodynamics. *Mol Cancer Ther* **2010,** *9* (10), 2844-52.

81. Brockmann, M.; Poon, E.; Berry, T.; Carstensen, A.; Deubzer, H. E.; Rycak, L.; Jamin, Y.; Thway, K.; Robinson, S. P.; Roels, F.; Witt, O.; Fischer, M.; Chesler, L.; Eilers, M., Small molecule inhibitors of aurora-a induce proteasomal degradation of N-myc in childhood neuroblastoma. *Cancer Cell* **2013,** *24* (1), 75-89.

82. Walensky, L. D.; Bird, G. H., Hydrocarbon-stapled peptides: principles, practice, and progress. *J Med Chem* **2014,** *57* (15), 6275-88.

83. Cromm, P. M.; Spiegel, J.; Grossmann, T. N., Hydrocarbon stapled peptides as modulators of biological function. *ACS Chem Biol* **2015,** *10* (6), 1362-75.

84. Joo, S. H., Cyclic peptides as therapeutic agents and biochemical tools. *Biomol Ther (Seoul)* **2012,** *20* (1), 19-26.

85. Vinogradov, A. A.; Yin, Y.; Suga, H., Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *J Am Chem Soc* **2019,** *141* (10), 4167-4181.

86. Blackwell, H. E.; Grubbs, R. H., Highly Efficient Synthesis of Covalently Cross-Linked Peptide Helices by Ring-Closing Metathesis. *Angew Chem Int Ed Engl* **1998,** *37* (23), 3281-3284.

87. Schafmeister, C. E.; Po, J.; Verdine, G. L., An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *Journal of the American Chemical Society* **2000,** *122* (24), 5891-5892.

88. Lau, Y. H.; de Andrade, P.; Wu, Y.; Spring, D. R., Peptide stapling techniques based on different macrocyclisation chemistries. *Chem Soc Rev* **2015,** *44* (1), 91-102.

89. Walensky, L. D.; Kung, A. L.; Escher, I.; Malia, T. J.; Barbuto, S.; Wright, R. D.; Wagner, G.; Verdine, G. L.; Korsmeyer, S. J., Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. *Science* **2004,** *305* (5689), 1466-70.

90. Bernal, F.; Wade, M.; Godes, M.; Davis, T. N.; Whitehead, D. G.; Kung, A. L.; Wahl, G. M.; Walensky, L. D., A stapled p53 helix overcomes HDMX-mediated suppression of p53. *Cancer Cell* **2010,** *18* (5), 411-22.

91. Grossmann, T. N.; Yeh, J. T.; Bowman, B. R.; Chu, Q.; Moellering, R. E.; Verdine, G. L., Inhibition of oncogenic Wnt signaling through direct targeting of beta-catenin. *Proc Natl Acad Sci U S A* **2012,** *109* (44), 17942-7.

92. Moellering, R. E.; Cornejo, M.; Davis, T. N.; Del Bianco, C.; Aster, J. C.; Blacklow, S. C.; Kung, A. L.; Gilliland, D. G.; Verdine, G. L.; Bradner, J. E., Direct inhibition of the NOTCH transcription factor complex. *Nature* **2009,** *462* (7270), 182-8.

93. Stewart, M. L.; Fire, E.; Keating, A. E.; Walensky, L. D., The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nat Chem Biol* **2010,** *6* (8), 595-601.

94. Edwards, A. L.; Gavathiotis, E.; LaBelle, J. L.; Braun, C. R.; Opoku-Nsiah, K. A.; Bird, G. H.; Walensky, L. D., Multimodal interaction with BCL-2 family proteins underlies the proapoptotic activity of PUMA BH3. *Chem Biol* **2013,** *20* (7), 888-902.

95. Mitra, S.; Montgomery, J. E.; Kolar, M. J.; Li, G.; Jeong, K. J.; Peng, B.; Verdine, G. L.; Mills, G. B.; Moellering, R. E., Stapled peptide inhibitors of RAB25 target context-specific phenotypes in cancer. *Nat Commun* **2017,** *8* (1), 660.

96. Iyer, A.; Van Lysebetten, D.; Ruiz Garcia, Y.; Louage, B.; De Geest, B. G.; Madder, A., Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake. *Org Biomol Chem* **2015,** *13* (13), 3856-62.

97. Payne, S. R.; Pau, D. I.; Whiting, A. L.; Kim, Y. J.; Pharoah, B. M.; Moi, C.; Boddy, C. N.; Bernal, F., Inhibition of Bacterial Gene Transcription with an RpoN-Based Stapled Peptide. *Cell Chem Biol* **2018,** *25* (9), 1059-1066 e4.

98. Edwards, A. L.; Meijer, D. H.; Guerra, R. M.; Molenaar, R. J.; Alberta, J. A.; Bernal, F.; Bird, G. H.; Stiles, C. D.; Walensky, L. D., Challenges in Targeting a Basic Helix-Loop-Helix Transcription Factor with Hydrocarbon-Stapled Peptides. *ACS Chem Biol* **2016,** *11* (11), 3146-3153.

99. Talanian, R. V.; McKnight, C. J.; Kim, P. S., Sequence-specific DNA binding by a short peptide dimer. *Science* **1990,** *249* (4970), 769-71.

100. Kouzarides, T.; Ziff, E., Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature* **1989,** *340* (6234), 568-71.

101. Blancafort, P.; Segal, D. J.; Barbas, C. F., 3rd, Designing transcription factor architectures for drug discovery. *Mol Pharmacol* **2004,** *66* (6), 1361-71.

Chapter 2

Synthetic DNA-Binding Domains

2.1 Design and synthesis of sDBDs

In Chapter 1, we have discussed early attempts to target bHLH TFs. In particular, few successful examples have been reported as effective means for direct inhibition of bHLH DNA binding due to the technical challenges in using small molecules to inhibit protein-DNA interactions. Miniproteins, by contrast, have shown great promise[1, 2] but their large size poses challenges both synthetically and pharmacologically. Therefore, when creating synthetic DNA-binding scaffolds, we seek the opportunity to interrogate the entire protein-DNA interfaces by effective means of mimicking the basic DNA-binding domain with the smallest possible size. On the other hand, mere stabilization of an individual α-helix would likely be insufficient due to lack of tertiary structure and binding cooperativity. An early study reveals that a short peptide dimer derived from GCN4, a bZIP TF, binds DNA in a sequence-specific manner similar to the full-length proteins[3]. Inspired by this design concept, we aim to stabilize the basic DNA-binding helices of Myc and Max, respectively, and link the two helices in a proper way that resembles the full DNA recognition domain. We expect the resulting scaffold, which we call synthetic DNA-binding domain (sDBD), to be fully functional *in vitro* and have the potential for targeting Myc/Max and other bHLH TFs *in situ* and *in vivo*.

2.1.1 Structural considerations

To come up with initial designs for sDBDs, we started by analyzing the protein-protein and protein-DNA interactions present in the crystal structure of the Myc/Max and Max/Max DNA complexes[4, 5]. The core structure of Myc/Max contains a leucine zipper helix connected to a basic DNA-binding helix through a flexible loop. The two proteins form a heterodimer through the leucine zipper helices, resulting in the formation of a stable tetrahelix core that orients two basic DNA-binding helices for interaction with the major groove of DNA (Figure 2.1A). This domain architecture is conserved across hundreds of bHLH TFs. The conserved and modular structure of this DNA binding domain suggested that it may be possible to create fully synthetic mimics of the basic DNA-binding region through synthetic preservation of 1) tertiary domain construction of dimeric, oriented DNA-binding helices (Figure 2.1B) and 2) local α-helix stabilization to retain or augment affinity, specificity and stability (Figure 2.1C). Therefore, by deconstructing the elements that result in productive DNA-binding, we hypothesized that synthesis of peptides containing the two minimal DNA-binding helices in a covalently dimerized and properly oriented form should be capable of DNA recognition. The proposed sDBD scaffold fully recapitulates the basic DNA-binding region of the Myc/Max heterodimer and thus should preserve the sequence-specific binding activity of the full-length proteins (Figure 2.1D).

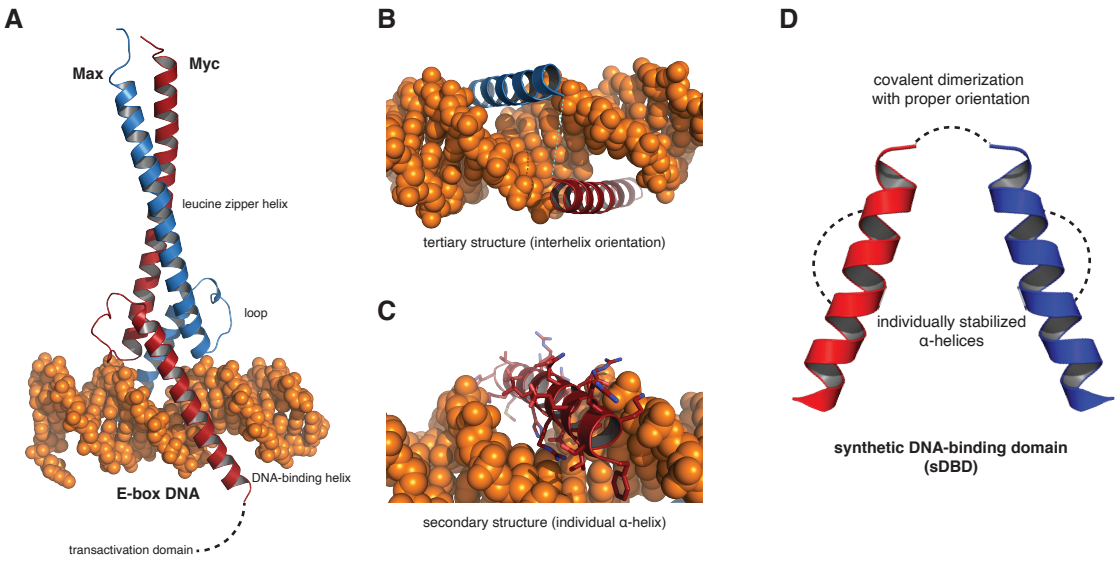


Figure 2.1 Structural analysis of Myc/Max-DNA complex leads to the initial design of sDBDs
(**A**) Crystal structure of Myc/Max proteins bound to target DNA (PDB ID: 1NKP). (**B-C**) Closer look of interhelix orientation and individual α-helices. (**D**) Proposed sDBD scaffold featuring stabilization of both secondary and tertiary structures.

Our initial hypothesis is that proper stabilization of the individual DNA-binding helices (secondary structure) as well as their orientation (tertiary structure) is critical to the DNA binding activity of sDBDs. To introduce stabilizing elements with minimal interference with the sDBD-DNA interactions, we first identified multiple residues on each α-helix with no direct contact with the DNA. These positions, spaced four residues or one helical turn apart, are pointing away from the DNA major groove (on the "back face") and can be used in different combinations to incorporate an *i*,*i* + 4 staple (Figure 2.2A, C). Meanwhile, helix orientation can be resembled by creating a linker between Phe374 and Phe43 of Myc and Max, respectively. The hydrophobic and π-π interactions between these two residues, along with several others in the loop and leucine zipper region, play a key role in properly orienting the helices[6], and their close proximity make them ideal anchor positions for linking the two helices (Figure 2.2B-C).

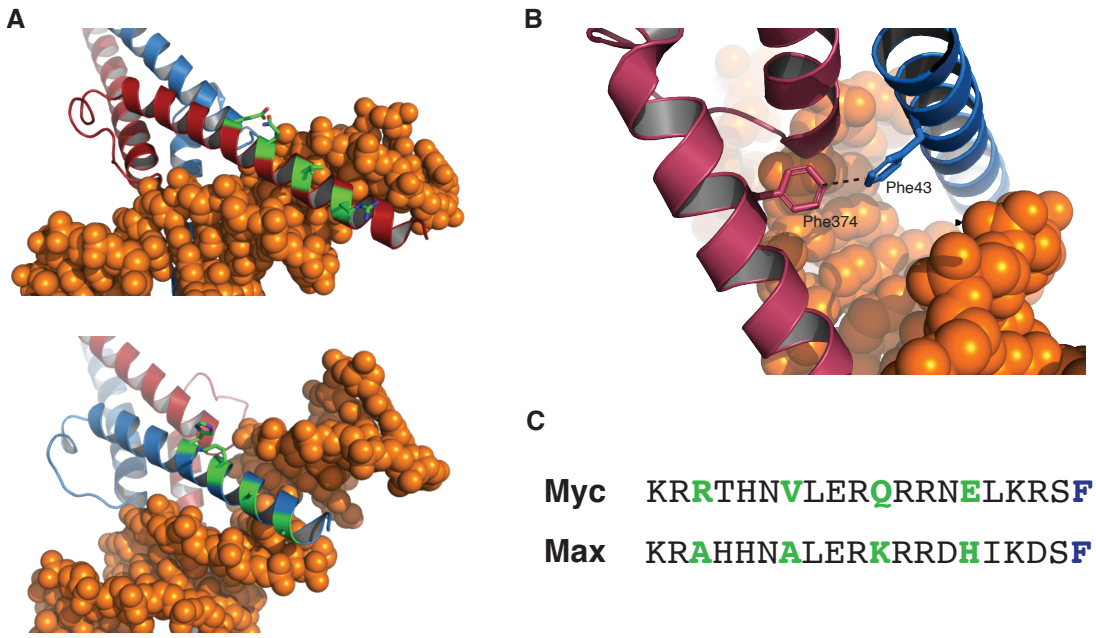


Figure 2.2 Further analysis reveals positions for incorporating stabilizing elements
(**A**) Back face of both Myc and Max proteins at the DNA binding site. Residues not directly involved in binding are depicted in green. (**B**) Two phenylalanine residues on both proteins are in close proximity and play a role in the formation of tertiary structure. (**C**) Possible positions for peptide stapling (green) and interhelix conjugation (navy).

2.1.2 Chemical considerations

The grand scheme of the sDBD design requires two types of stabilizing elements to be separately incorporated. The orthogonality of chemistries used in peptide and protein synthesis is thus very important and needs to be thoroughly considered. For individual α-helix stabilization, we elected a classic *i*,*i* + 4 hydrocarbon staple by ring-closing metathesis of two (*S*)-2-(4'-pentenyl)alanine residues, which has proven to be robust and fully compatible with solid phase peptide synthesis. The conjugation of the two helices, on the other hand, can be achieved by the ligation between two synthons on the C-termini (Figure 2.3A). The chemistry for such ligation needs to meet the following requirements: 1) Heterofunctionality: this ensures the heterodimerization of two different helices. 2) Biorthogonality and no cross reactivity with any natural amino acid side chains. 3) Compatibility with ring-closing metathesis. 4) High efficiency

with mild conditions. Among several candidates that are widely used in bioconjugation[7] (Figure 2.3B), we chose the thiol-maleimide Michael conjugation for this purpose, mainly because it can be conveniently incorporated into peptide synthesis schemes. A cysteine residue can provide the thiol group. Although the maleimide group may not be fully compatible with ring-closing metathesis due to the presence of the carbon-carbon double bond, it can be coupled onto a selectively deprotected amine on a lysine residue after peptide stapling.

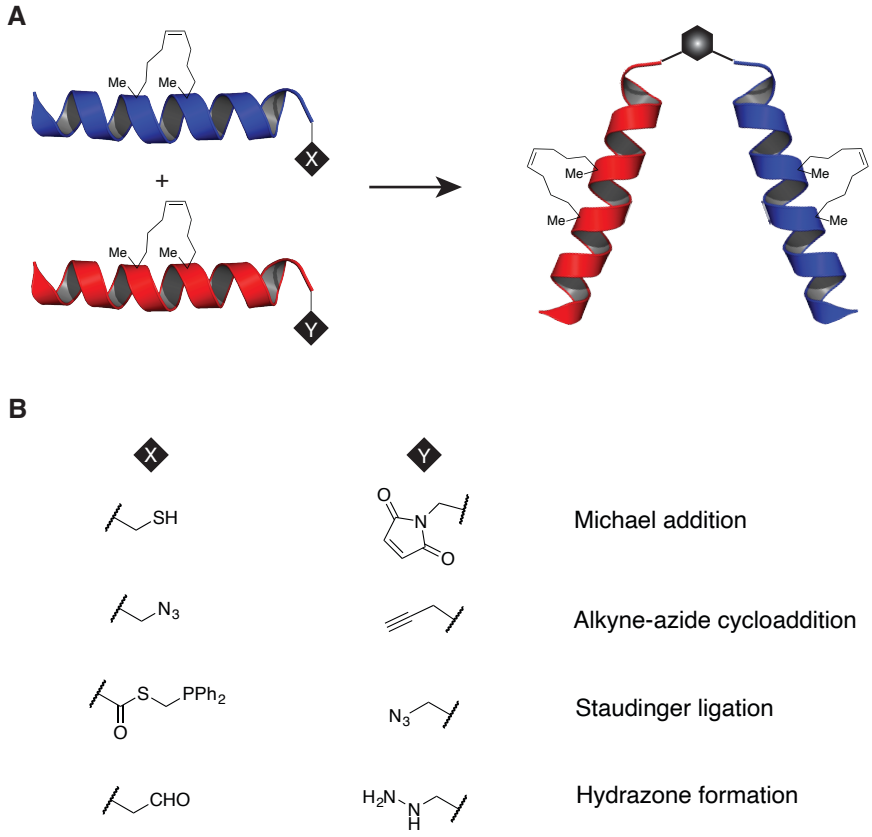


Figure 2.3 Orthogonal conjugation chemistries create the tertiary structure
(**A**) General synthetic strategy to conjugate two individually stapled peptides. (**B**) Potential reactive pairs.

2.2 Results

2.2.1 Synthesis of the initial sDBD library with increased structure and thermal stability

All things considered, we developed a two-step process for the preparation of sDBDs: 1) Individual stapled peptides were prepared by standard solid phase peptide synthesis followed by

ring-closing metathesis. Both helices contain a C-terminal reactive group that allows covalent heterodimerization. In particular, Myc-derived peptides feature a thiol group from cysteine, whereas Max-derived ones contain a C-terminal 4-monomethoxytrityllysine (Mmt-lysine) that is selectively deprotected on resin and then coupled with an *N*-hydroxylsuccinimide ester containing a maleimide (Figure 2.4A-B). The cyclohexane linker has moderate rigidity and an arm length of 8.3 Å, which is close to the modeled distance between the two residues (8.2 Å). 2) Upon cleavage and purification, the two peptides were conjugated in pH-neutral aqueous solution with high efficiency and quantitative yield (Figure 2.4C). The purity and molecular weight of the final sDBD product was confirmed by LC-MS (Figure 2.5).

With different stapling positions, we generated one version with the unmodified sequence and three stapled ones for both Myc- and Max-derived peptides. The different combinations of monomeric peptides yielded an initial library of one unmodified and nine stapled sDBDs, the sequences of which are listed in Table 2.1.

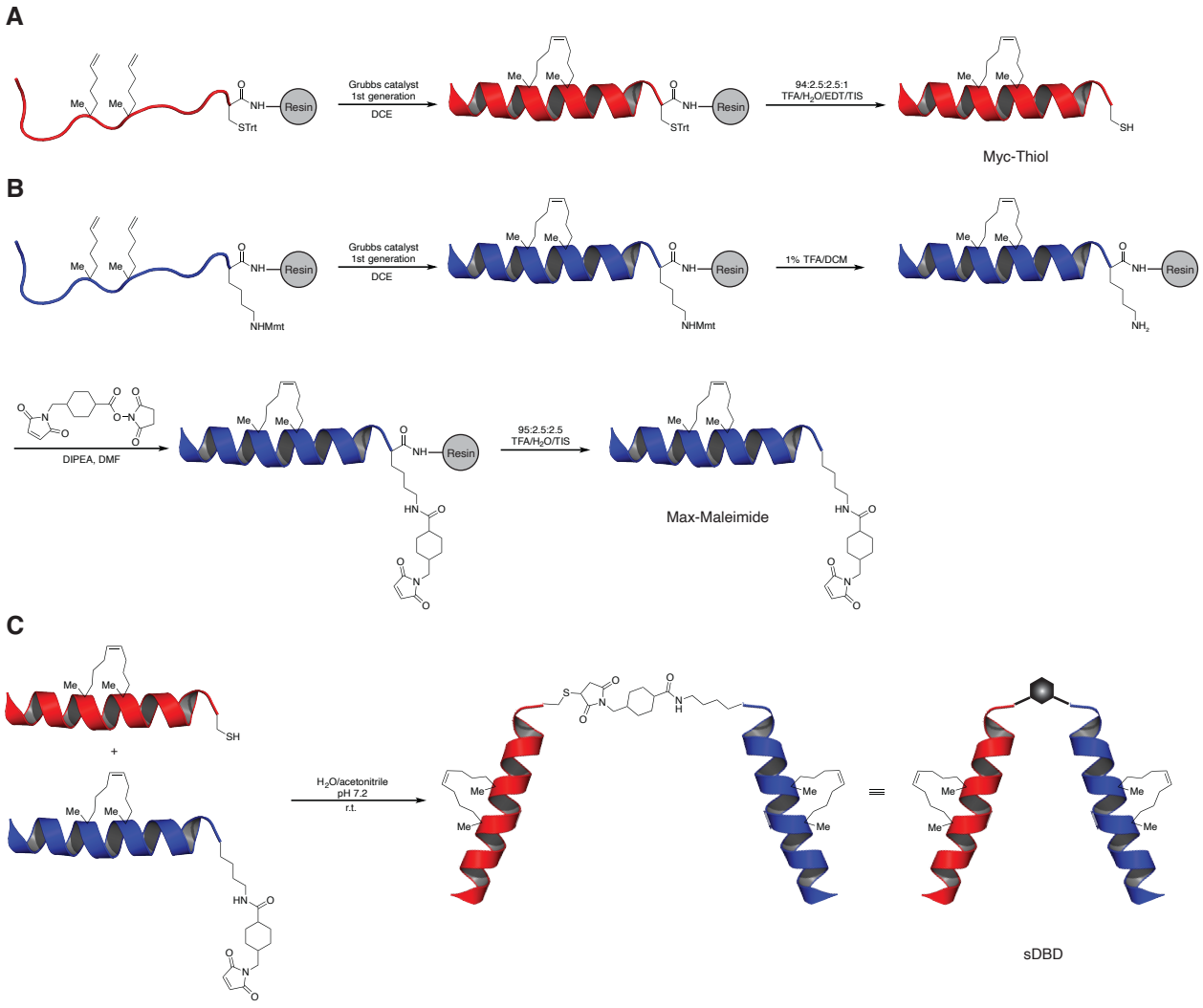
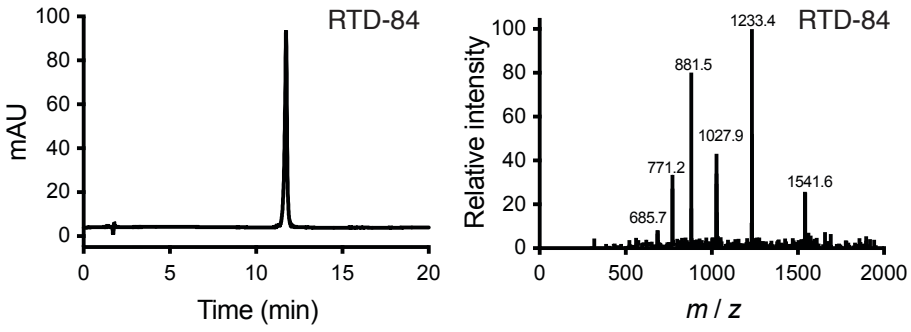


Figure 2.4 Synthetic scheme for sDBDs

(**A**-**B**) Solid phase synthesis of stapled peptides containing a C-terminal reactive group. (**C**) Conjugation of two stapled helices creates the sDBD scaffold.



Exact mass: 6163.4

Z	(M+Z)/Z
4	1541.9
5	1233.7
6	1028.2
7	881.5
8	771.4

Figure 2.5 LC-MS characterization of a representative sDBD

Table 2.1 Sequences of the initial sDBD library

Monomer	Sequence	Net charge	Helicity (%)
Myc series			
RTD-1	AcWβKRRTHNVLERQRRNELKRSβC	+6	18
RTD-6	AcWβKRS_5THNS_5LERQRRNELKRSβC	+5	48
RTD-4	AcWβKRRTHNS_5LERS_5RRNELKRSβC	+6	41
RTD-5	AcWβKRRTHNVLERS_5RRNS_5LKRSβC	+7	55
Max series			
RTD-3	AcWβKRAHHNALERKRRDHIKDSβK_{Mmt}	+4	15
RTD-8	AcWβKRS_5HHNS_5LERKRRDHIKDSβK_{Mmt}	+4	31
RTD-2	AcWβKRAHHNS_5LERS_5RRDHIKDSβK_{Mmt}	+3	37
RTD-7	AcWβKRAHHNALERS_5RRDS_5IKDSβK_{Mmt}	+3	44

C/K_{Mmt}: interhelix conjugation sites
S_5: (*S*)-2-(4'-pentenyl)alanine
β: β-alanine

The presence of the hydrocarbon stapled should improve the α-helicity and augment the structural stability of sDBDs. To confirm this effect, we employed circular dichroism (CD) spectroscopy to characterize the secondary structure of the individual α-helices. The data suggest that both unmodified peptides RTD-1 and RTD-3 is largely unstructured. The addition of the hydrocarbon staple significantly increases the helical content, resulting ellipticity minima at 208 and 222 nm (Figure 2.6A). In both cases, a staple at the C-terminus resulted in the highest helicity, whereas no specific correlations were observed for the other positions. Furthermore, we performed temperature-dependent CD and found that the all stapled sDBDs retained some level of α-helicity at up to 95 °C whereas RTD-31 remained unstructured throughout the wide temperature range (Figure 2.6B). Collectively, these results indicate significant improvement of secondary structure as well as thermal stability of stapled sDBDs over the unmodified version.

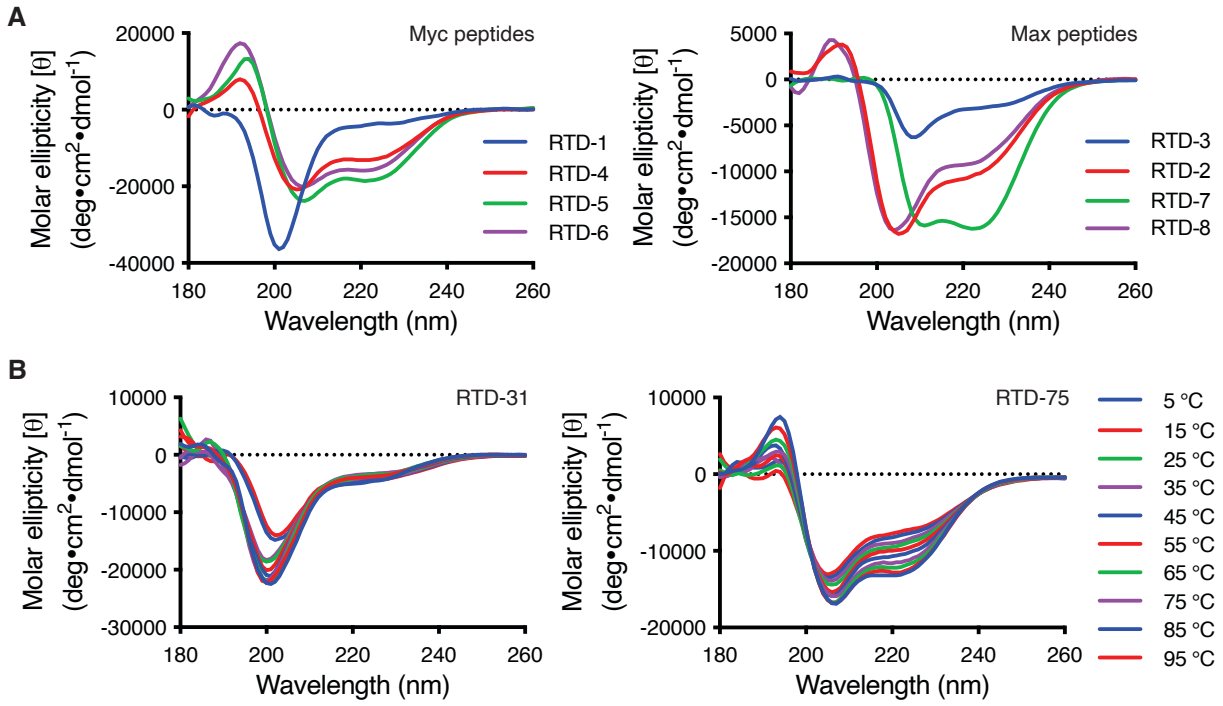


Figure 2.6 CD spectra of select sDBDs

(**A**) CD spectra of all monomeric peptides in both Myc and Max series. (**B**) Temperature-dependent CD scans indicating a change in secondary structure for unmodified (RTD-31) and stapled (RTD-75) sDBDs.

2.2.2 sDBDs bind E-box DNA with potency comparable to bHLH proteins

As the first step of our *in vitro* screening, we performed the electrophoretic mobility shift assays (EMSAs) to determine the DNA binding affinity of the sDBD library. An E-box-containing, infrared dye labeled 19-mer oligonucleotide was used as the fluorescent probe. We started with recombinant Myc/Max heterodimer and Max/Max homodimer, both of which showed stably bound DNA complex in a wide range of concentrations (Figure 2.7A). The apparent K_D's for both Myc/Max and Max/Max were 10–20 nM, consistent with existing studies[8-10].

We then proceeded with the entire sDBD library. While some compounds, notably those containing either RTD-2 or RTD-6, did not show meaningful binding with DNA, to our delight, several candidates, such as RTD-74, RTD-75, RTD-84, and RTD-85, gave gel shift patterns that resemble the ones for recombinant proteins, at comparable apparent K_D's in the low nanomolar

range (Figure 2.7B). For comparison, we also assayed RTD-1, RTD-4, and RTD-31. Despite reports that unmodified or "stapled" basic domain helices alone can specifically bind DNA[11-13], we found that neither RTD-1 nor RTD-4 showed any appreciable binding to E-box DNA. On the other hand, RTD-31, which features a conjugation of unstabilized DNA-binding helices, yielded smeared bands indicative of labile interactions (Figure 2.7C). These results suggest that both peptide stapling and conjugation, featured in our design rationale, are crucial to the DNA-binding activity.

We then move on to testing the inhibitory effects of sDBDs in a competitive format. In the presence of recombinant Myc/Max and oligonucleotide probe, where stable shifted bands were formed, increased concentrations of sDBDs resulted in diminishing protein-DNA binding (Figure 2.8A). the most potent among them are RTD-85, RTD-74, RTD-84, with IC_{50}'s of 300–600 nM (Figure 2.8B). This is consistent with the abovementioned results that they are also the strongest DNA binders in the library.

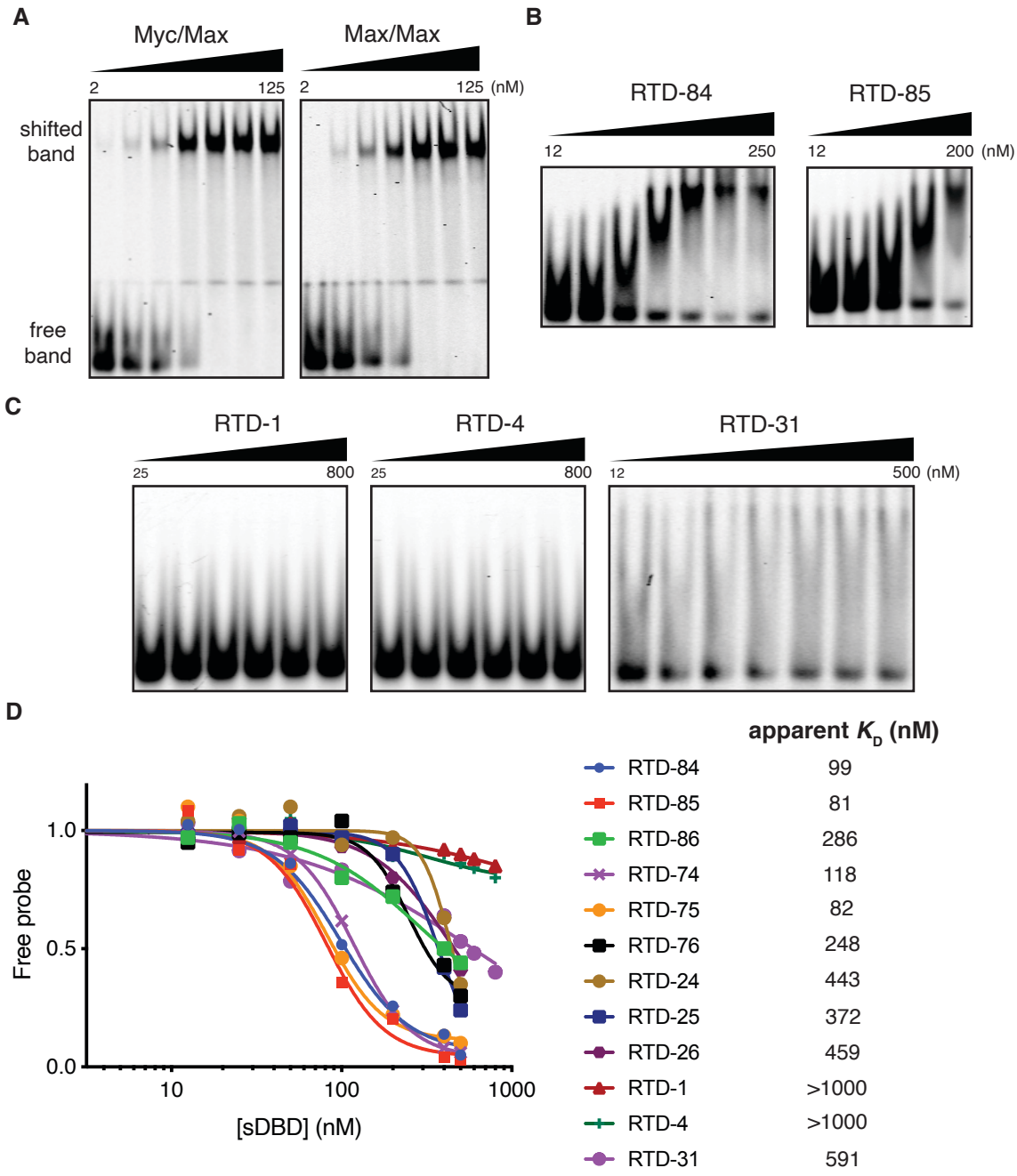


Figure 2.7 EMSA screening of bHLH TFs and sDBD library

(**A**-**C**) Representative EMSA gel images of recombinant bHLH proteins and sDBDs. (**D**) Quantification results for apparent K_D across different sDBDs.

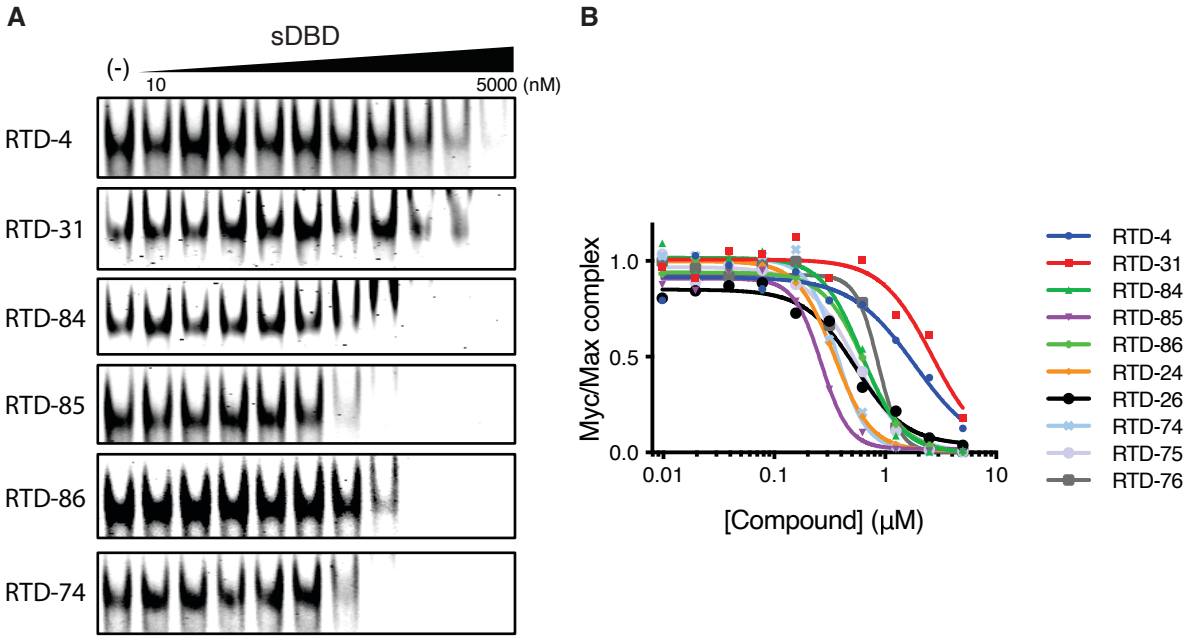


Figure 2.8 Competition of sDBD with Myc/Max protein
(**A**) The addition of sDBDs diminishes the shifted band that corresponds to Myc/Max-DNA complex. (**B**) Quantification of competition.

The kinetic profile of sDBD-DNA interaction was next evaluated using surface plasmon resonance (SPR). We immobilized a 5'-biotinylated E-box duplex on a NeutrAvidin-coated sensor chip surface which was then exposed to increasing concentrations of various peptides. The dose-dependent response data were best fit to a Langmuir kinetic model (Figure 2.9). The fitted results indicate uniformly rapid association rates (k_a) across the library, whereas dissociation rates (k_d) vary by orders of magnitude and hence contribute to the difference in activity of sDBDs (Table 2.2). The fact that dissociation rates play a larger role in differentiating the overall K_D is in accordance with DNA binding for most transcription factors revealed by existing studies[14, 15].

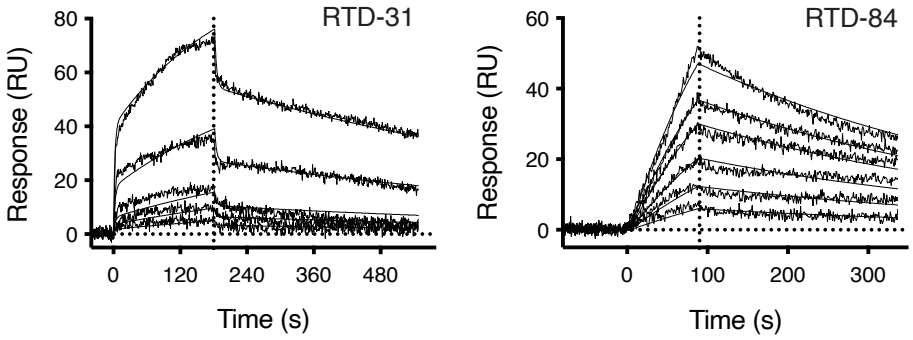


Figure 2.9 SPR sensorgrams of unmodified and stapled sDBDs

Table 2.2 sDBD binding kinetics from fitting of SPR curves

sDBD	k_a (10^4 $M^{-1}\cdot s^{-1}$)	k_d (10^{-3} s^{-1})	K_D (10^{-8} M)
RTD-31	7.10	268	377
RTD-84	3.61	2.25	6.24
RTD-85	3.54	8.58	24.2
RTD-86	6.91	87.1	126
RTD-74	3.19	2.63	8.26
RTD-75	3.18	0.651	2.04
RTD-76	5.64	76.5	136
RTD-24	5.39	56.3	104
RTD-26	7.59	92.3	122

Having demonstrated the DNA binding potency of sDBDs, we then focused on whether they process a comparable level of sequence specificity to that of recombinant proteins. We employed EMSA in a competitive format by adding non-fluorescently labeled oligonucleotides, either containing an E-box consensus or a mutated sequence, to see if the shifted fluorescent band can be diminished by the competitors. As shown in Figure 2.10, Myc/Max is highly specific. Addition of unlabeled competitor E-box consensus oligonucleotide (E-BoxC1, 5'-CACGTG-3') led to a diminishing shifted fluorescent band bound by the protein. A non-canonical E-box sequence (E-boxCM2, 5'-CA<u>TA</u>TG-3') with two internal mutations showed far less competition compared to the canonical E-box consensus sequence, whereas a random mutation (E-boxCM3, 5'-<u>ATATAT</u>-3') could not compete with the shifted band. The shifted band of RTD-84, on the other hand, was approximately equally responsive to all three competitors. While the results do

not necessarily preclude the possibility of sDBDs' sequence-specific binding in other experimental setup and biological context, such as in a cellular environment with multiple binding sites, they provide some evidence that the sequence specificity of sDBDs may not be as robust as full-length TFs.

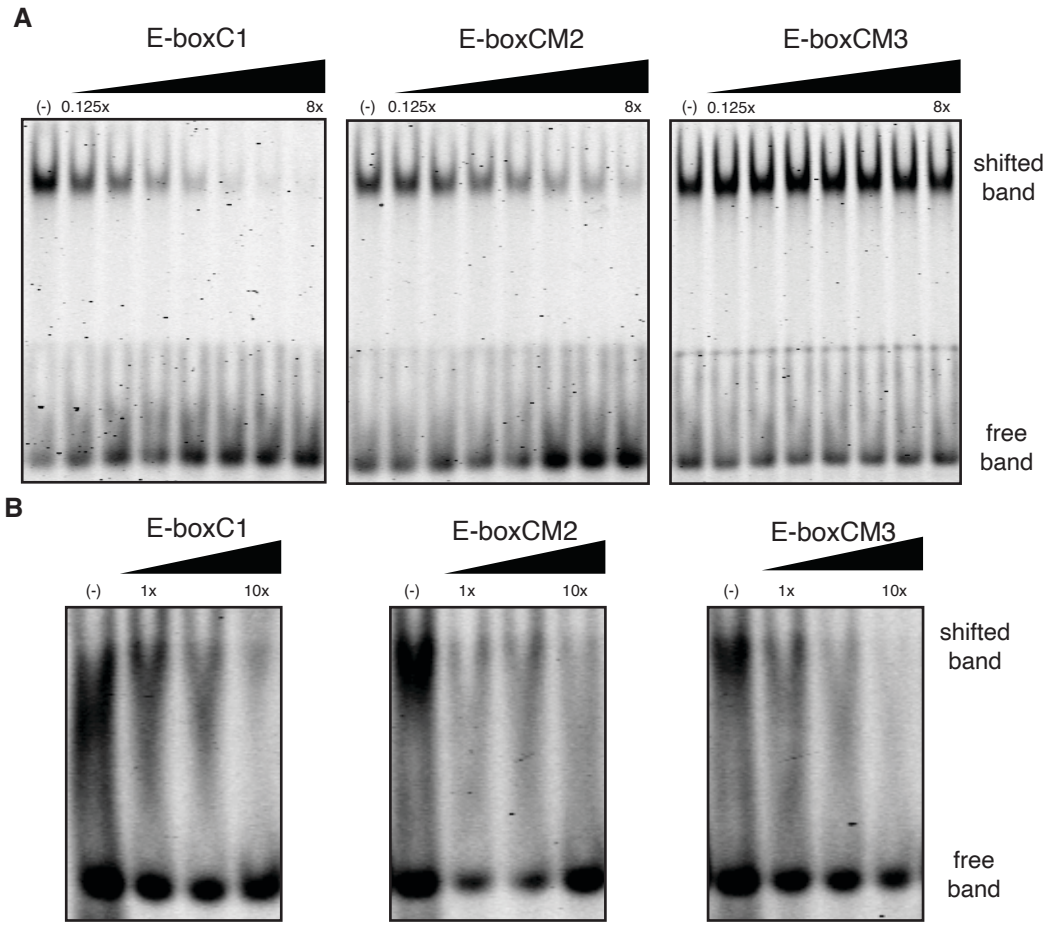


Figure 2.10 EMSA specificity assay results of Myc/Max protein and RTD-84
(**A**) Myc/Max and (**B**) RTD-84. Unlabeled competitor oligonucleotides of different sequences were added to sequester the fluorescent shifted band.

2.2.3 Altering the C-terminal dimerization motif affects DNA binding activity

The initial sDBD library delivered promising DNA binding results, validating our design principles. However, all compounds that we have made so far share the same C-terminal linker, the impact of which has not yet been fully elucidated. Currently, the C-terminal reactive group is

separated from the main sequence by a β-alanine, which breaks the α-helix and adds some structural flexibility on the C-terminus. To examine whether any subtle changes to the dimerization motif lead to differences in activity, we created two additional designs that are listed in Table 2.3. The first series, RTD-84g, replaces the β-alanine with a glycine, and consequently keeps the dimerization position within the α-helix. The added glycine residue, however, over-rotates both α-helices by 1/3.6 of a turn and thus leads to a misorientation of the DNA-binding surfaces. The second series, RTD-913, removes the β-alanine spacer and results in a somewhat rigid, properly oriented helix dimer (Figure 2.11A). EMSA results for these two compounds are consistent with our design rationale. RTD-84g showed a loss of binding affinity to E-box DNA due to the misorientation. Deletion of the glycine residues, on the other hand, restored the binding affinity in RTD-913. Furthermore, due to a more rigid orientation, RTD-913 appeared to be more potent than RTD-84 despite a narrower dynamic range, as characterized by a shifted band formation at 25–50 nM and lack thereof at higher concentrations.

Table 2.3 Sequences of sDBDs with altered C-terminal structure

sDBD	Sequence
RTD-84	
RTD-4	AcWβKRRTHNS_5LERS_5RRNELKRSβC
RTD-8	AcWβKRS_5HHNS_5LERKRRDHIKDSβK_{Mmt}
RTD-84g	
RTD-4g	AcWβKRRTHNS_5LERS_5RRNELKRSGC
RTD-8g	AcWβKRS_5HHNS_5LERKRRDHIKDSGK_{Mmt}
RTD-913	
RTD-13	AcWβKRRTHNS_5LERS_5RRNELKRSC
RTD-9	AcWβKRAHHNALERS_5RRDS_5IKDSK_{Mmt}

C/K_{Mmt}: interhelix conjugation sites
S_5: (*S*)-2-(4'-pentenyl)alanine
β: β-alanine

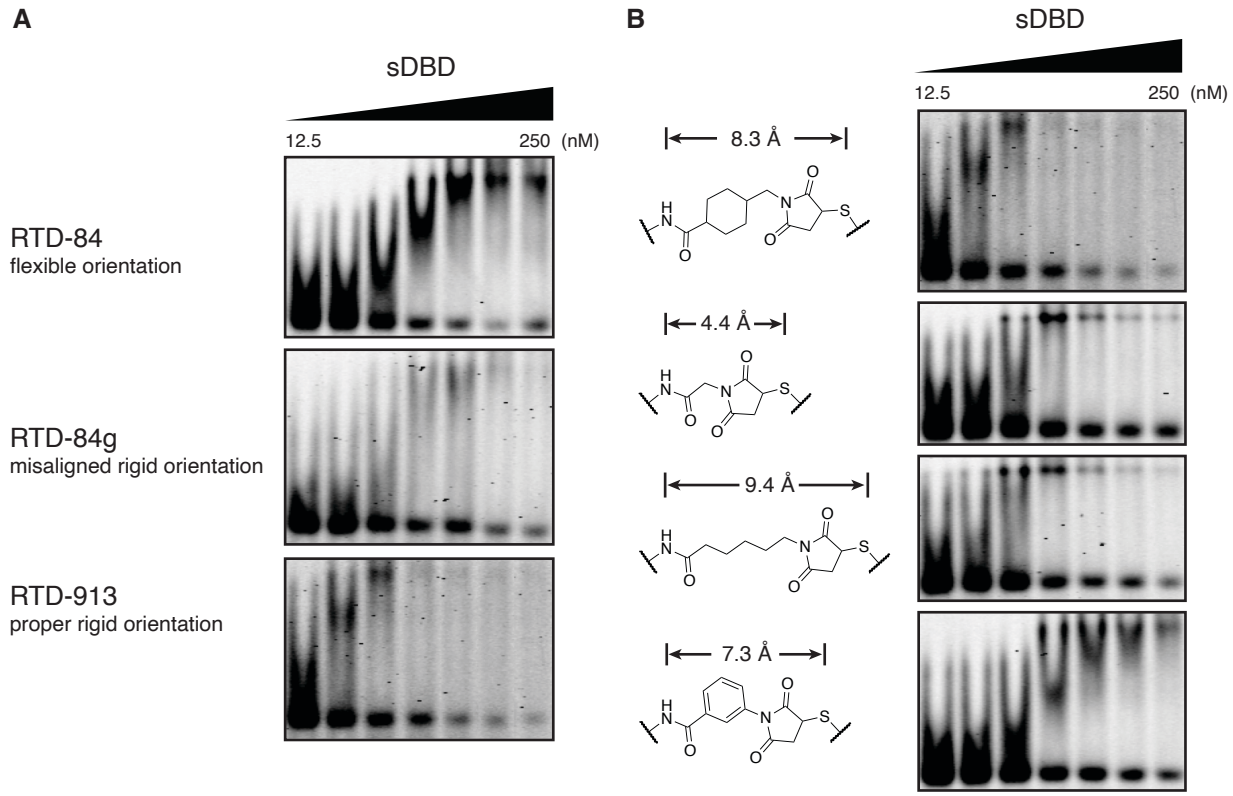


Figure 2.11 Altering the tertiary structure impacts sDBD activity
(**A**) Changes in helix orientation lead to expected loss of gain of activity. (**B**) Rigidity and arm length of the maleimide linker have an impact on the binding potency and dynamic range.

Because the increased rigidity leads to seemingly higher sensitivity to the linker structure, we also altered the arm length and rigidity of maleimide linker. Compared to the original cyclohexane linker, a more flexible alkane linker resulted in a decrease in activity, regardless of the arm lengths. An increase in linker rigidity with proper arm length, as in the case of an *m*-benzene ring, yielded a better binding profile with a wider dynamic range from low to mid nanomolar concentrations (Figure 2.11B). Collectively, these results suggest a strong correlation between the tertiary structure and the *in vitro* activity of sDBDs.

2.2.4 sDBDs are cell permeable and show nuclear localization

Stabilized peptides and cell-penetrating proteins interact with and enter cells via different mechanisms compared to many cell-permeable small molecules. Attributes such as secondary structure, charge, hydrophobicity, solubility and proteolytic stability have been shown to be important for productive cellular uptake and cytosolic access for different classes of stabilized peptides[16-19]. However, a potential issue arises that cell uptake for sDBDs may be impeded by their complex scaffold and increased size (over 6 kDa). By contrast, recent studies have reveal no direct correlation between the size of the peptides and their cell permeability[17]. To fully address these concerns, we used fluorescence microscopy to test the cell penetration of fluorescein isocyanate (FITC)-labeled sDBDs. To accurately assess the intracellular distribution of sDBDs, we also used 4',6-diamidino-2-phenylindole (DAPI) and Alexa Fluor 488 labeled wheat germ agglutinin (WGA) to stain the nuclei and cell membrane, respectively. As shown in Figure 2.12A, the images confirmed that the unmodified FITC-RTD-31 showed decent amount of FITC signal within the cells after 10 hours, despite its large size and lack of any secondary structure stabilization. Due to its predominant positive charge, it is plausible that it is able to translocate through the cell membrane via a mechanism similar to that of cell-penetrating peptides that are rich in arginine and lysine residues. FITC-RTD-84, on the other hand, showed far superior cellular uptake, with nearly six-fold increase in FITC signal compared to FITC-RTD-31 (Figure 2.12B). These results suggest that the incorporation of hydrocarbon staples on both helices significantly enhances cell penetration, and also indicate that sDBDs, despite their large size and structural complexity, can still engage in the endocytosis pathway that is responsible for stapled peptide internalization. Time-dependent experiments further revealed that higher cellular uptake for FITC-RTD-84 was associated with longer treatment (Figure 2.12C). Furthermore, in terms of the intracellular

localization of sDBDs, a moderate amount of FITC signal (green) overlapped with the nuclei (DAPI, blue), indicating that FITC-RTD-84 can further pass through the nuclear membrane and access the chromatin, which is critical to target binding and *in situ* effects sDBDs.

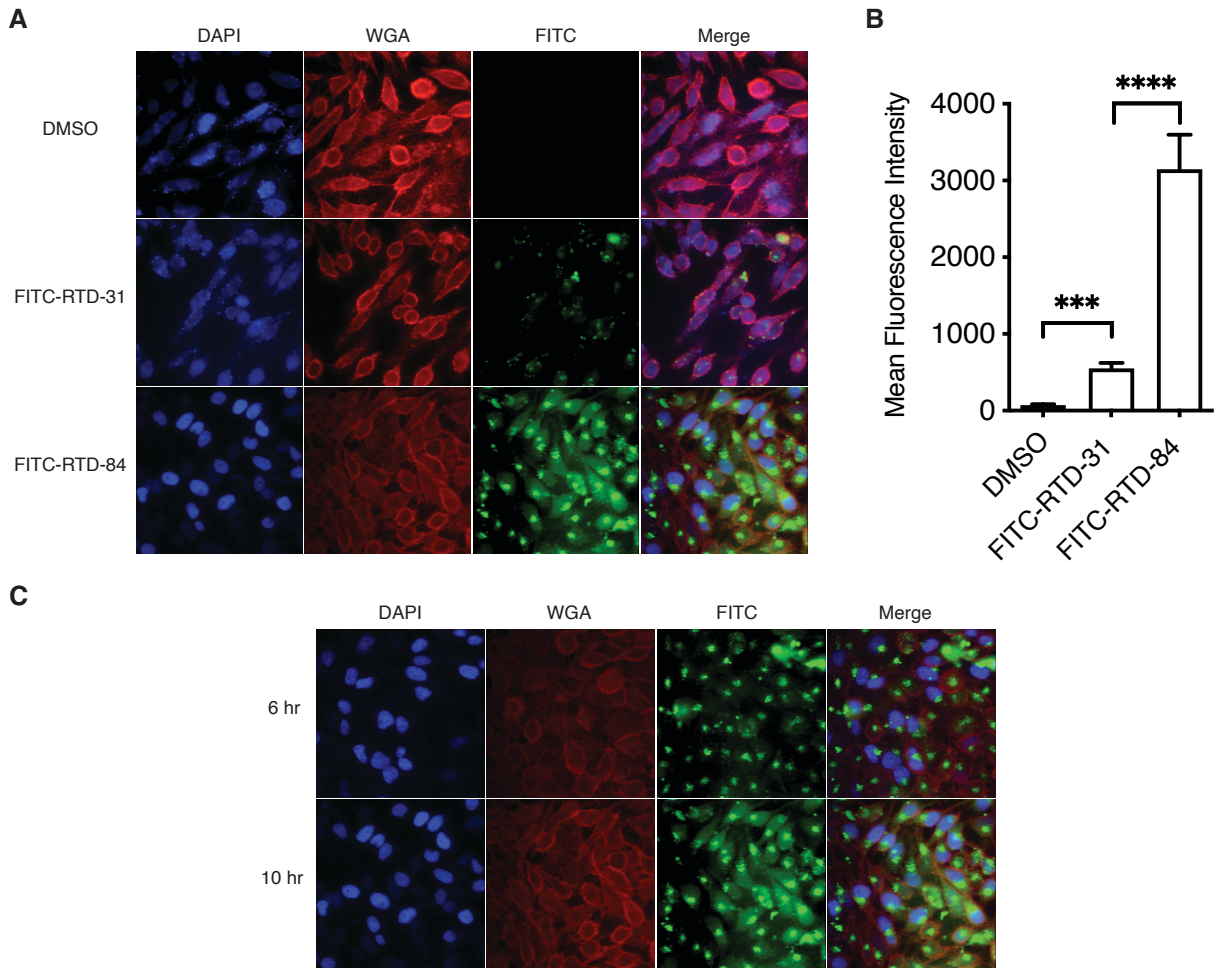


Figure 2.12 Fluorescent microscopy images of HeLa cells after treatment with FITC-labeled sDBDs
(**A**) Comparisons between unmodified FITC-RTD-31 and stapled FITC-RTD-84. (**B**) Quantification of fluorescent signal of the FITC channel across three conditions. (**C**) Time-dependent treatment with FITC-RTD-84.

2.2.5 sDBDs compete with endogenous Myc in cells

We then designed cell-based assays to evaluate lead sDBDs' effects on cellular functions. To begin with, we asked whether these compounds could bind E-box-containing genomic DNA and impact target binding of endogenous bHLH TFs. Myc-dependent chromatin immunoprecipation (ChIP) studies have validated the association of Myc and Max proteins with

specific E-box promoters and enhancers in numerous cell lines[20, 21]. Guided by these datasets, we performed ChIP-qPCR assays in HeLa cells to quantify the occupancy of Myc protein at five validated regions—*FBXW8*, *RPL37*, *CASP8*, *TOR1A*, and *MRPS15*—all of which showed significant enrichment over both the negative control regions (intergenic and *SNTG2*) and the isotype control (immunoprecipitated by IgG, Figure 2.13A)[21]. Upon treatment with 10 μM RTD-84 for 8 hours, we found that the occupancy of Myc across all five regions decreased by up to 50%, indicating that RTD-84 is capable of competing with the DNA binding of endogenous Myc at these loci (Figure 2.13B). To fully assess the kinetics and dose-response of the competition, we further examined a wide range of RTD-84 concentrations and different time points. In the time-dependent experiments, we observed decreased Myc occupancy in target regions as early as 2 hours, which further increased over time and remained steady after 8 hours of treatment (Figure 2.13C). In terms of dose-response, higher doses of RTD-84 showed more effective Myc competition, but as the concentration exceeded 5 μM, no insignificant improvement were observed (Figure 2.13D). We also evaluated three other lead compounds in the sDBD library and found that with the exception of RTD-74, both RTD-75 and RTD-85 resulted in competition profiles similar to RTD-84 (Figure 2.13E). Finally, to determine whether the competition was specific to Myc TF, we performed the competition ChIP experiment for estrogen receptor α and found no significant decrease in enrichment at its target locus, *pS2*/*TFF1*. Collectively, these data confirm that the lead sDBDs from our initial screening can access intracellular compartments and effectively compete with endogenous Myc for certain E-box-containing targets in the genome with some level of selectivity.

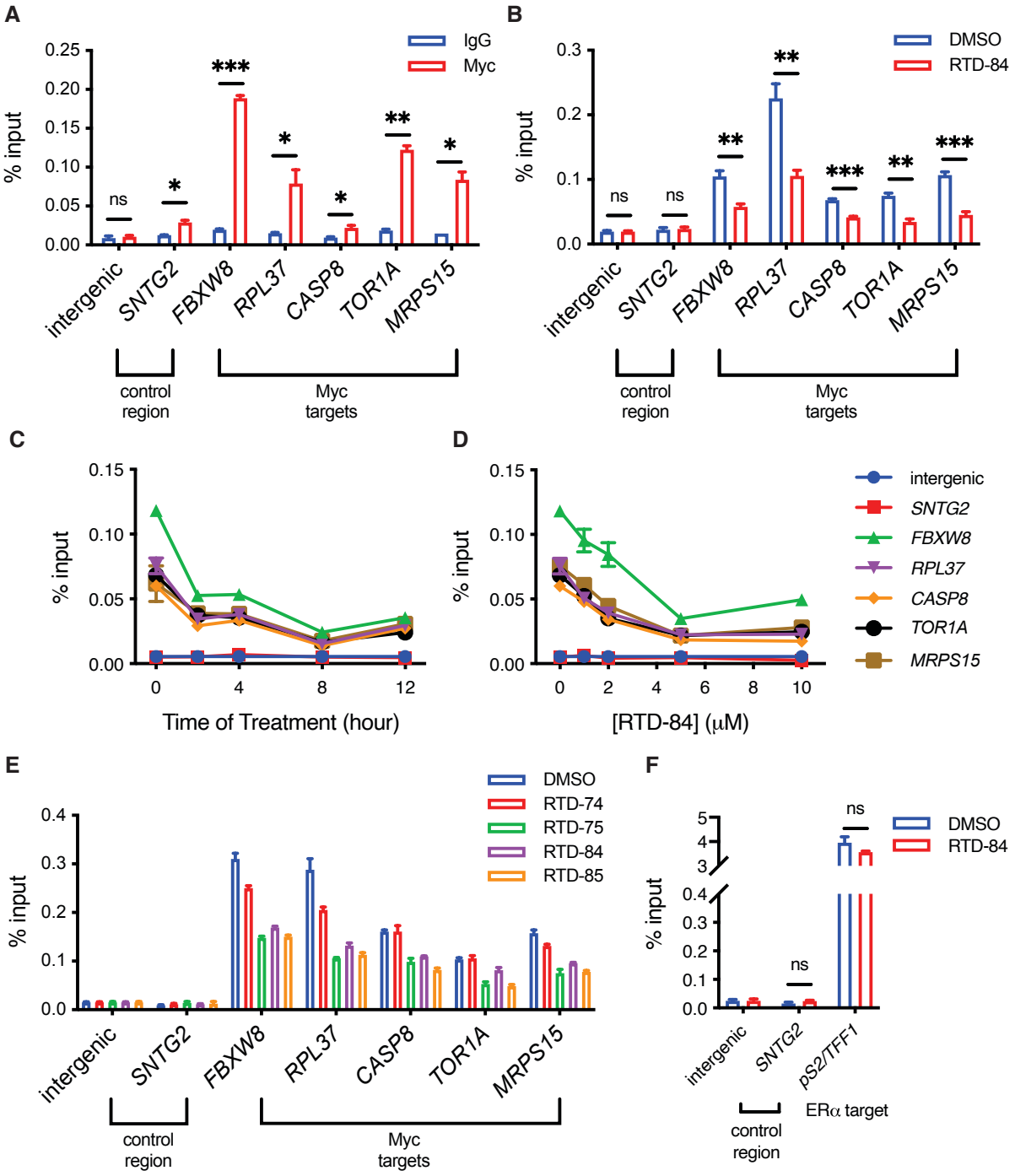


Figure 2.13 ChIP-qPCR in HeLa cells with various sDBD treatment conditions

(**A**) Validation of ChIP assay and Myc binding targets. (**B**) Treatment with RTD-84 decreases the enrichment level across all five Myc targets. (**C-E**) Time-, dose- and compound-dependent ChIP-qPCR experiments. (**F**) RTD-84 shows no significant competition with the target binding of ERα.

2.2.6 Some sDBDs are cytotoxic due to membrane disruption

Having confirmed that sDBDs can penetrate cell membrane and compete with endogenous Myc, we were interested in whether such effects could lead to any Myc-dependent downstream cellular responses. Given Myc's association with cell proliferation, we performed CellTiter-Glo viability assays to assess the compound's effects on viability. To begin with, we treated HeLa cells with varying concentrations of RTD-84 and RTD-85. RTD-84 showed no meaningful effects at up to 10 μM, whereas RTD-85 significantly reduced cell viability with an estimated IC_{50} of 4 μM. We further tested both compounds in a diverse set of cancer cell lines with varying levels of Myc regulation, and surprisingly found that such viability effects were nearly identical across multiple cell lines. We then questioned whether the observed cytotoxicity of RTD-85 was due to Myc-dependent transcriptional response or merely some universal effects. To confirm this hypothesis, we conducted LDH release assays to measure the cell membrane disruption of both compounds, a phenomenon that has been observed for some specific stabilized peptides and other cell penetrating peptide classes. While RTD-84 did not lead to the release of LDH above DMSO control at up to 10 μM, RTD-85 significantly increased LDH activity at low micromolar concentrations. Further screening of other sDBDs revealed similar results, namely any RTD-5-containing molecules resulted in higher LDH release. Interestingly, RTD-5 is the most helical and positively charged helix in the series, suggesting a possible correlation between α-helicity, excessive positive charge, and membrane disruption, which has been reported for other stapled peptide families in previous studies[18].

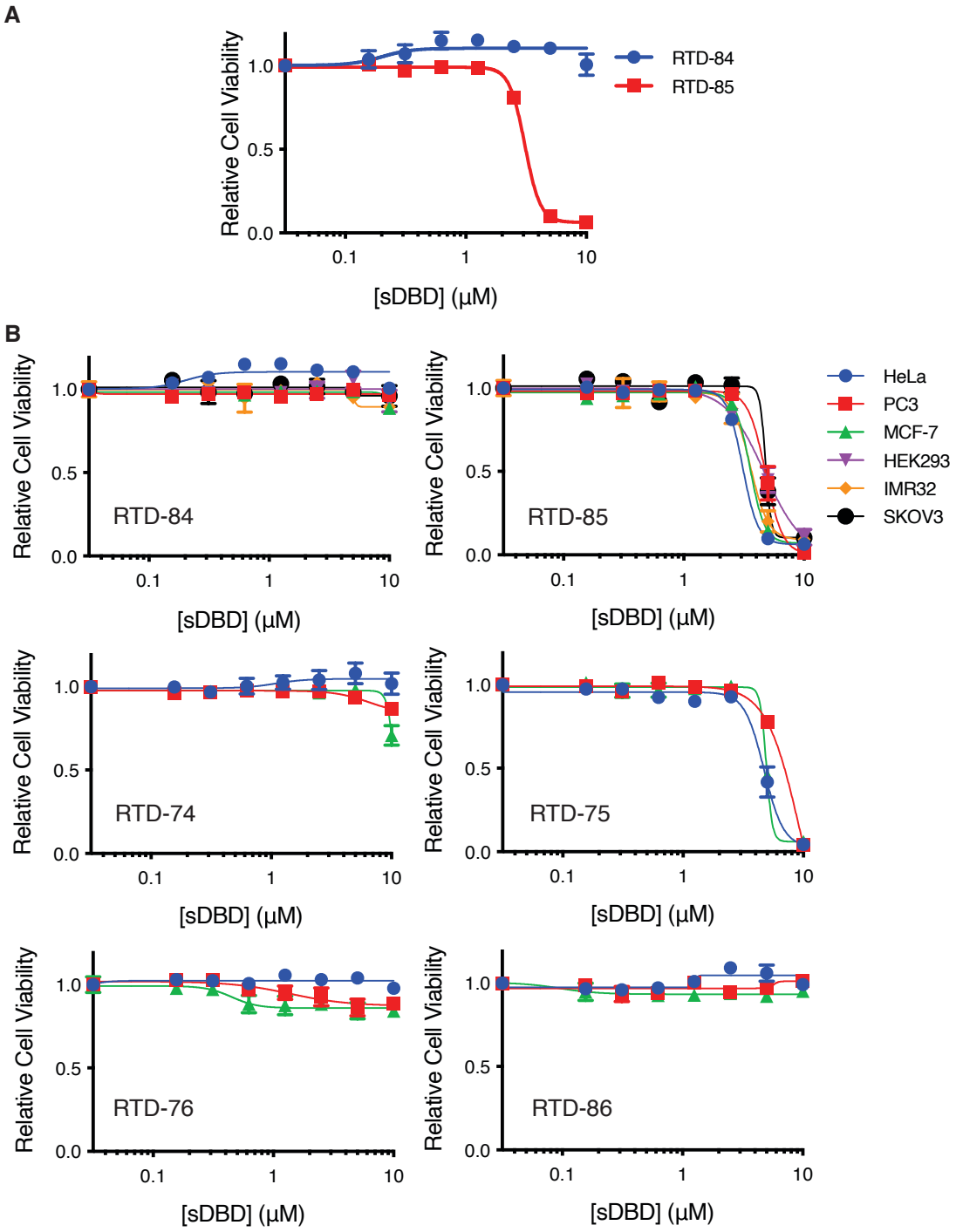


Figure 2.14 CellTiter-Glo viability assay results of sDBDs in a variety of cancer cell lines

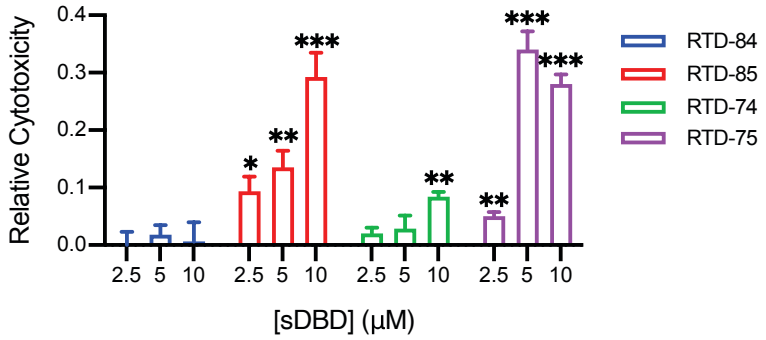


Figure 2.15 LDH release assay showed membrane disruption of select sDBDs

2.3 Discussion

2.3.1 A modular synthetic platform and an established assay pipeline reveal structure-activity relationship of sDBDs

In conclusion, we have developed a synthetic platform and generated an initial library of sDBDs. The chemistries are compatible and robust, and the platform is highly modular and applicable to a wide variety of synthetic targets. We have also established a pipeline of biochemical and cell biological assays for thorough screenings of sDBD activity, which have greatly expanded our knowledge of miniprotein design. Starting from the full-length TFs, we have fully recreated the DNA-binding activity by incorporating multiple stabilized structures, similar to a fragment-based ligand design process (Figure 2.16). A standalone DNA-binding α-helix, regardless of whether it is stapled or not, is inactive due to the lack of proper tertiary structure alignment. Similarly, a mere interhelix dimerization without stabilized secondary structures is also insufficient. Once we have both structural stabilizing elements in place, however, the resulting sDBD becomes functional and possesses a series of promising activities both *in vitro* and *in situ*.

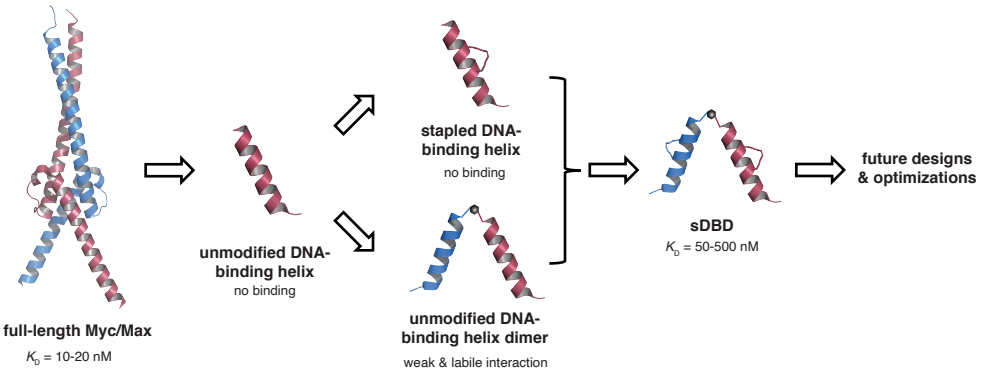


Figure 2.16 Recapitulating DNA binding activity of full-length proteins in a stepwise, fragment-based stabilization process

A deeper dive into the structure-activity relationship of sDBDs also provides important insights. In terms of secondary structure stabilization, for example, we have identified a strong correlation between α-helicity and *in vitro* binding activity, in accordance with the general notion that higher structure stabilization often leads to tighter target binding[22] (Figure 2.17). Perhaps the most noteworthy is that one particular data point—RTD-31—dominates the overall correlation, without which the trend would be less significant. This indicates that instead of a strict monotonic relationship, there may be a threshold helicity for target binding, above which contributions from other idiosyncratic factors, such as net charge, staple position, and hydrophobicity, become more important. In terms of tertiary structure stabilization, we have discovered that the rigidity and arm length of the linker as well as the interhelix orientation play a vital role in determining the sDBD's binding potency and dynamic range (Table 2.4). Proper understanding of these effects and efficient synthetic tools to manipulate the structures will serve as the guiding principles for any future designs and optimizations.

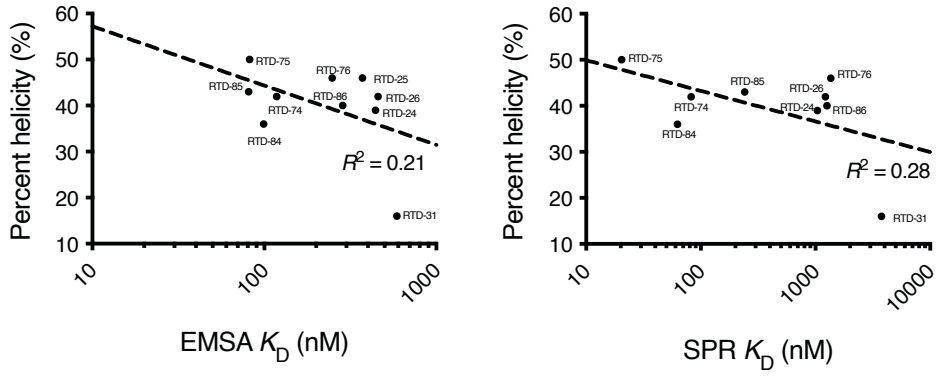


Figure 2.17 Correlation of sDBD activity with percent helicity

Table 2.4 Structure activity relationship of sDBDs with varying stabilizing elements

Representative sDBD	Description	K_D (μM)	Complex formation
RTD-31	Unmodified dimeric helices	>1.0	Labile
RTD-84	Flexible spacer, randomly-oriented dimeric helices	~0.2	Stable
RTD-84G	Rigid spacer, disoriented dimeric helices	~0.4	Labile → aggregation
RTD-913 with *m*-benzene linker	No spacer, well-oriented dimeric helices	~0.15	Stable

2.3.2 Limitations of the original sDBD library

Despite early success in this proof-of-concept design, there are several key limitations of the original sDBDs that prevent us from moving forward and further assessing these compounds in more sophisticated assays. First and foremost, even though we have demonstrated sDBDs' DNA binding affinity and effective competition with Myc, we have limited evidence of any sequence specificity of such interactions—In biochemical assays, lead sDBDs' binding patterns to several E-box mutant oligonucleotides were somewhat comparable to those for the canonical consensus sequence 5'-CACGTG-3', as indicated by both EMSA and SPR. Given their net positive charge and relatively flexible conformation, it is not entirely surprising that these sDBDs show some moderate affinity to other DNA sequences in a non-specific way. However, the lack of

differentiation of E-box consensus from other DNA sequences is contrary to our expectations that the specific component of the interactions would prevail and lead to higher affinities.

Second, while sDBDs are capable of competing with Myc and interfering with the formation of the Myc-DNA complex, we have yet to elucidate the exact mechanism of such competition. Ideally, we would expect a 1:1 ligand replacement at those binding sites, but as the additional of sDBDs led to the disappearance of the Myc/Max-DNA complex band on the EMSA gels, we did not observe any new bands that corresponded to the formation of stable sDBD-DNA complex. These results suggest that the competitive behavior of sDBDs is more complicated than a simple ligand replacement scheme and may involve the formation of nonspecific, higher-order complexes that could not be resolved by gel electrophoresis.

Third, the structural basis of sDBD-DNA interactions remains largely unknown. We have attempted to crystallize the DNA complexes of several lead sDBDs and solve their structure by X-ray crystallography. We were able to consistently obtain sizable crystals from multiple rounds of crystallization screening for RTD-84, but diffractions by these crystals failed to provide any meaningful structural information at <3 Å resolution. We suspect that the absence of other structural components, such as the loop and the leucine zipper region, resulted in a weakly packed crystal lattice with too much molecular movement within, which did not survive long enough upon exposure to the beam.

Finally, although preliminary data suggest that sDBDs are capable of penetrating cell membrane and compete with endogenous Myc, we have not revealed any direct connections between such in-situ properties and the downstream biological effects on the Myc-dependent transcriptome, proteome, and phenotypes. We developed an extensive biological assay pipeline to probe such effects, including luciferase reporter assay, reverse transcription PCR, and quantitative

proteomics, but concluded that lead sDBDs did not result in significant changes in Myc-dependent gene expression or proteome. Additionally, the lack of viability effects in a variety of cancer cell lines suggests that sDBDs are unlikely to cause phenotypic changes.

Taken all together, we believe that the above-mentioned limitations strongly suggest further improvements and optimizations are necessary. Instead of being complacent about current designs and hovering around more screenings, we decided to pivot to optimize the initial designs in search of more functional sDBDs.

2.4 Materials and methods

Solid phase peptide synthesis. All peptides were synthesized manually or by automated peptide synthesizer (Gyros Protein Technologies) using standard Fmoc-peptide chemistry on rink amide MBHA resin (EMD Millipore). Ring-closing metathesis of unnatural amino acids containing pentenyl side-chains was performed twice using 0.2 equivalents of Grubbs-I catalyst dissolved in dichloroethane under nitrogen for 2 hours. For Max peptides, C-terminal modification was achieved by selective deprotection of Mmt-lysine using 1% trifluoroacetic acid (TFA) in dichloromethane (DCM) for 5 × 5 min, followed by coupling with 2 equivalents of SMCC and 4 equivalents of *N*,*N*-diisopropylethylamine (DIPEA) in *N*,*N*-dimethylformamide (DMF). Prior to cleavage, peptides were subjected to one final round of deprotection and capping by standard methods with acetic anhydride, fluorescein isothiocyanate (FITC) or otherwise stated linkers. After completion, crude peptide was cleaved using 95:2.5:2.5 TFA/water/triisopropylsilane, dissolved in 1:1 acetonitrile/water, and purified by reverse-phase high-pressure liquid chromatography (HPLC) using a C18 column (Agilent). Compound identification and purity was assessed using coupled liquid chromatography mass spectrometry (LCMS) (Agilent).

Covalent conjugation of peptides. After purification by HPLC, thiol-containing Myc peptide and maleimide-containing Max peptide were mixed in equal amount in 4:1 water/acetonitrile with 0.1 M phosphate and pH 7.2. The reaction was monitored using LCMS and the final product was purified by HPLC.

Circular dichroism spectroscopy. A circular dichroism spectrometer (Jasco J-1500) was used to determine the α-helical content of all monomeric and dimeric peptides. Peptides were dissolved into 10 mM potassium phosphate buffer at pH 7.4 to a final concentration of 25–250 mM in a quartz cuvette with a path length of 0.1 cm. Ellipticity values were taken at 1 nm intervals between 180 and 260 nm. The percentage helicity was calculated from the ellipticity at 222 nm using helical models as previously reported.

Protein expression and purification. Human Myc (residues 355–434) and Max (residues 22–102) proteins were expressed with a N-terminal hexahistidine tag in *Escherichia coli* strain BL21(DE3) using a pET28c vector. Transformed bacteria was grown at 37°C and induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) at an A_{600} = 0.8. Cells were pelleted 14 h after induction and lysed in lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl, 8 M urea, 10 mM imidazole, pH 8.0) with cOmplete EDTA-free Protease Inhibitor (Roche) by sonication. The lysate was centrifuged to clear insoluble matter before loading onto Ni-NTA resin (Qiagen). After being washed once with lysis buffer and three times with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, 20 mM imidazole, pH 8.0), column-bound protein was eluted using elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, 250 mM imidazole, pH 8.0) and quantified using Bradford assay. The eluted Myc and Max proteins were then mixed in equal amount, dialyzed into desired buffers suitable for downstream applications, and further concentrated by centrifugation using a 10-kDa exclusion filter.

Electrophoretic mobility shift assay (EMSA). Recombinant protein or synthetic peptides were diluted to desired concentrations varying from 1 to 500 nM. A 5'-infrared dye-labeled double-stranded DNA oligonucleotide with the E-box consensus binding site is then added (5'-ACCCCACCACGTGGTGCCT-3'). The mixture was incubated at 4°C for 30 min. Final concentrations were: 20 mM HEPES (pH 8.0), 150 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1 mg/mL BSA, 5% glycerol, and 5 nM oligonucleotide probe. Complexes were resolved on 6% acrylamide gels (40 mM Tris, 45 mM borate, 1 mM EDTA, pH 8.3), and gels were scanned using an infrared imager (Li-Cor Odyssey). Gel images were analyzed using ImageJ software and binding results were plotted and fitted using GraphPad Prism 7.

Surface plasmon resonance. A ProteOn XPR36 Protein Interaction Array System (Bio-Rad) was used for all SPR-binding measurements. A 5'-biotin labeled double stranded DNA oligonucleotide with the E-box consensus binding site (5'-ACCCCACCACGTGGTGCCT-3') was diluted to 500 nM in binding buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 5 mM $MgCl_2$, 0.05% polysorbate-20, pH 7.6) and injected onto a ProteOn NLC sensor chip (Bio-Rad) at a flow rate of 30 µL/min for 2 min, reaching approximately 200 response units. Increasing concentrations of proteins or peptides (blanks followed by 25–400 nM) were subsequently injected at a flow rate of 30 µL/min for 2 min, followed by dissociation of 4 min. After each injection the sensor surface was regenerated with 1 M NaCl and 0.002% SDS to remove all bound analytes. A Langmuir binding model was applied to the processed dataset to determine kinetic parameters of the protein-DNA and peptide-DNA interactions.

Fluorescence microscopy. HeLa cells were plated on glass coverslips in RPMI-1640 media containing 10% FBS at 37°C under 5% CO_2 and allowed to settle overnight. Cells were washed with PBS twice prior to *in situ* treatment with DMSO control or 1 µM FITC-labeled

peptides in fresh growth media. After incubation for the indicated time, cells were washed four times with PBS and fixed at 25°C for 15 min in 3.7% methanol-free paraformaldehyde in PBS. Fixed cells were then stained with DAPI (Sigma Aldrich), Alexa Fluor 633-labeled wheat-germ agglutinin (Molecular Probes) according to manufacturers' recommendations for 1 h at 25°C. For image acquisition, processed coverslips were mounted on microscope slides and confocal images were acquired on a DSU spinning disk confocal microscope (Olympus) and analyzed using ImageJ software.

Cell culture. All cell lines were purchased from ATCC and cultured in RPMI-1640 with 10% FBS and 1% penicillin/streptomycin. All cell culture was performed under 37°C with 5% CO_2.

Chromatin immunoprecipitation (ChIP). HeLa cells were seeded in 200-mm dishes. After reaching 70% confluent, cells were treated with DMSO control or peptides in fresh growth media for the specified time. Upon completion, cells were crosslinked with 1% formaldehyde, fragmented by sonication, and incubated with c-Myc antibody (9E11, Abcam) overnight. The mixture was then immunoprecipitated with protein A beads (GenScript, pre-treated with 1% BSA for 1 hour) for 1 hour. Immunoprecipitated complexes were successively washed with Low Salt Wash Buffer I (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl, pH 8.0), High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl, pH 8.0), and LiCl Wash Buffer (250 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0). All washes were performed at RT for 8 min on a rotator. The complexes were eluted with 1% SDS at 30°C for 15 min, and then incubated at 65°C overnight to reverse crosslink protein-DNA complexes. After decrosslinking, DNA was purified using QIAQuick PCR Purification Kit (Qiagen) according to the manufacturer's instructions.

Cell viability assay. Approximately 2,000 cells were seeded in 50 μL media in a 96-well plate and let settle overnight. 25 μL media containing 3x compound or DMSO vehicle were added to experimental wells and the plate was incubated for the indicated time of experiment. Upon completion of treatment, 75 μL CellTiter-Glo reagent (Promega) was added to each well and the plate was gently shaken at room temperature for 10 minutes before luminescence signal was read on a plate reader.

LDH release assay. Approximately 5,000 HeLa cells were seeded in a 96-well plate. The following day an equal volume of media containing 2x compound or DMSO vehicle were added to experimental wells and the plate was incubated for the indicated time of experiment. A 2x volume of lysis buffer was added to additional wells 45 minutes prior to the final time-point and LDH activity in treatment medium was measured using the Pierce LDH Cytotoxicity Assay Kit according to manufacturer protocol.

References

1. Soucek, L.; Whitfield, J.; Martins, C. P.; Finch, A. J.; Murphy, D. J.; Sodir, N. M.; Karnezis, A. N.; Swigart, L. B.; Nasi, S.; Evan, G. I., Modelling Myc inhibition as a cancer therapy. *Nature* **2008,** *455* (7213), 679-83.

2. Savino, M.; Annibali, D.; Carucci, N.; Favuzzi, E.; Cole, M. D.; Evan, G. I.; Soucek, L.; Nasi, S., The action mechanism of the Myc inhibitor termed Omomyc may give clues on how to target Myc for cancer therapy. *PLoS One* **2011,** *6* (7), e22284.

3. Talanian, R. V.; McKnight, C. J.; Kim, P. S., Sequence-specific DNA binding by a short peptide dimer. *Science* **1990,** *249* (4970), 769-71.

4. Nair, S. K.; Burley, S. K., X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **2003,** *112* (2), 193-205.

5. Brownlie, P.; Ceska, T.; Lamers, M.; Romier, C.; Stier, G.; Teo, H.; Suck, D., The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* **1997,** *5* (4), 509-20.

6. Atchley, W. R.; Wollenberg, K. R.; Fitch, W. M.; Terhalle, W.; Dress, A. W., Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Mol Biol Evol* **2000,** *17* (1), 164-78.

7. Stephanopoulos, N.; Francis, M. B., Choosing an effective protein bioconjugation strategy. *Nat Chem Biol* **2011,** *7* (12), 876-84.

8. Blackwell, T. K.; Kretzner, L.; Blackwood, E. M.; Eisenman, R. N.; Weintraub, H., Sequence-specific DNA binding by the c-Myc protein. *Science* **1990,** *250* (4984), 1149-51.

9. Blackwood, E. M.; Eisenman, R. N., Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **1991,** *251* (4998), 1211-7.

10. Farina, A.; Faiola, F.; Martinez, E., Reconstitution of an E box-binding Myc:Max complex with recombinant full-length proteins expressed in Escherichia coli. *Protein Expr Purif* **2004,** *34* (2), 215-22.

11. Pazos, E.; Mosquera, J.; Vazquez, M. E.; Mascarenas, J. L., DNA recognition by synthetic constructs. *Chembiochem* **2011,** *12* (13), 1958-73.

12. Iyer, A.; Van Lysebetten, D.; Ruiz Garcia, Y.; Louage, B.; De Geest, B. G.; Madder, A., Stapling monomeric GCN4 peptides allows for DNA binding and enhanced cellular uptake. *Org Biomol Chem* **2015,** *13* (13), 3856-62.

13. Payne, S. R.; Pau, D. I.; Whiting, A. L.; Kim, Y. J.; Pharoah, B. M.; Moi, C.; Boddy, C. N.; Bernal, F., Inhibition of Bacterial Gene Transcription with an RpoN-Based Stapled Peptide. *Cell Chem Biol* **2018,** *25* (9), 1059-1066 e4.

14. Wang, H.; Ramakrishnan, A.; Fletcher, S.; Prochownik, E. V., A quantitative, surface plasmon resonance-based approach to evaluating DNA binding by the c-Myc oncoprotein and its disruption by small molecule inhibitors. *J Biol Methods* **2015,** *2* (2).

15. Wang, S.; Poon, G. M.; Wilson, W. D., Quantitative Investigation of Protein-Nucleic Acid Interactions by Biosensor Surface Plasmon Resonance. *Methods Mol Biol* **2015,** *1334*, 313-32.

16. Chang, Y. S.; Graves, B.; Guerlavais, V.; Tovar, C.; Packman, K.; To, K. H.; Olson, K. A.; Kesavan, K.; Gangurde, P.; Mukherjee, A.; Baker, T.; Darlak, K.; Elkin, C.; Filipovic, Z.; Qureshi, F. Z.; Cai, H.; Berry, P.; Feyfant, E.; Shi, X. E.; Horstick, J.; Annis, D. A.; Manning, A. M.; Fotouhi, N.; Nash, H.; Vassilev, L. T.; Sawyer, T. K., Stapled alpha-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc Natl Acad Sci U S A* **2013,** *110* (36), E3445-54.

17. Chu, Q.; Moellering, R. E.; Hilinski, G. J.; Kim, Y.-W.; Grossmann, T. N.; Yeh, J. T. H.; Verdine, G. L., Towards understanding cell penetration by stapled peptides. *MedChemComm* **2015,** *6* (1), 111-119.

18. Bird, G. H.; Mazzola, E.; Opoku-Nsiah, K.; Lammert, M. A.; Godes, M.; Neuberg, D. S.; Walensky, L. D., Biophysical determinants for cellular uptake of hydrocarbon-stapled peptide helices. *Nat Chem Biol* **2016,** *12* (10), 845-52.

19. LaRochelle, J. R.; Cobb, G. B.; Steinauer, A.; Rhoades, E.; Schepartz, A., Fluorescence correlation spectroscopy reveals highly efficient cytosolic delivery of certain penta-arg proteins and stapled peptides. *J Am Chem Soc* **2015,** *137* (7), 2536-2541.

20. Lin, C. Y.; Loven, J.; Rahl, P. B.; Paranal, R. M.; Burge, C. B.; Bradner, J. E.; Lee, T. I.; Young, R. A., Transcriptional amplification in tumor cells with elevated c-Myc. *Cell* **2012,** *151* (1), 56-67.

21. Walz, S.; Lorenzin, F.; Morton, J.; Wiese, K. E.; von Eyss, B.; Herold, S.; Rycak, L.; Dumay-Odelot, H.; Karim, S.; Bartkuhn, M.; Roels, F.; Wustefeld, T.; Fischer, M.; Teichmann, M.; Zender, L.; Wei, C. L.; Sansom, O.; Wolf, E.; Eilers, M., Activation and repression by oncogenic MYC shape tumour-specific gene expression profiles. *Nature* **2014,** *511* (7510), 483-7.

22. Pazgier, M.; Liu, M.; Zou, G.; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zella, D.; Tarasov, S. G.; Lu, W., Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proc Natl Acad Sci U S A* **2009,** *106* (12), 4665-70.

Chapter 3

Further Optimization of sDBDs and the Development of Orthogonal Peptide Chemistry

3.1 Rationale for further optimization of sDBDs

Before we came up with any specific optimizations and alternative designs of sDBDs, we would like to revisit the basic principles of protein-DNA interaction. A better understanding towards how certain proteins, especially bHLH TFs, interact and recognize particular DNA sequences would provide useful insights and guidelines into further structural optimizations. As a simplistic view, protein-DNA interactions can be segregated into nonspecific and specific components[1]. To facilitate the search of its specific DNA targets, a protein often associates with any nonspecific sequences and then diffuses along the DNA until specific contacts are formed[2, 3]. Therefore, most proteins are intrinsically involved in more or less nonspecific DNA interactions. Natural proteins have evolved to preserve the desired balance between nonspecific and specific interactions to ensure their functionality. In the case of sDBDs, however, the absence of additional structural components usually leads to greater molecular flexibility, and consequently, more nonspecific and/or less specific interactions with the target DNA compared to their natural counterparts. Therefore, one particular optimization strategy aims to disproportionately reduce non-specific interactions, which theoretically could lead to higher specificity to target DNAs.

From a thermodynamic point of view, protein-DNA binding events are associated with the decrease in total free energy in the system ($\Delta G < 0$). A favorable change in either the enthalpy ($\Delta H < 0$) or the entropy ($\Delta S > 0$) can contribute to the overall magnitude of the binding free energy. Therefore, we consider enthalpy and entropy the two driving forces for protein-DNA binding. The contributions of enthalpy and entropy to the binding free energy are usually inversely linked, a phenomenon known as the enthalpy-entropy compensation[4]. For example, a large negative enthalpy change due to the formation of multiple favorable interactions is usually accompanied by a negative entropy change due to the restriction of flexibility of the binding partners. Conversely, an enthalpic penalty is often necessary to disrupt interactions to achieve an increase in entropy. Previous studies have revealed that the DNA binding of bHLH proteins is primarily enthalpically driven throughout the physiological temperature range and characterized by an induced folding transition upon binding to the DNA[5]. These thermodynamic aspects provide us with two additional insights into how to optimize sDBD designs. First, we may increase the magnitude of the favorable enthalpic change by introducing additional contacts and interactions between the sDBD and the DNA. Second, we may use additional structural elements to induce the folding of the sDBD prior to binding to minimize the entropic penalty.

In this chapter, we will focus on the three optimization strategies identified above. Table 3.1 summarizes their rationale and synthetic considerations. While these strategies appear to be straightforward from a design perspective, their synthesis and the incorporation of new functionalities often require additional chemistries and thus pose significant challenges to our platform. Therefore, we also aim to expand our existing peptide chemistry repertoire as we create new designs and modifications of sDBDs.

Table 3.1 Strategies for sDBD optimizations

Strategy	Design approach	Synthetic considerations
Reduce nonspecific interactions while retaining specific interactions	Change peptide sequence to reduce nonspecific contacts with DNA	Existing synthetic tools can be used
Increase the magnitude of enthalpic change	Incorporate an additional DNA-binding motif to further lower the energy	Covalent coupling of a different motif onto sDBD, either on resin or in solution
Reduce the entropy penalty	Add a second linker to constrain the binding conformation	Additional orthogonal conjugation pair is likely required to add a second linker

3.2 Results

3.2.1 Altering peptide sequence reduces nonspecific interactions

The predominant nonspecific interactions between sDBDs and DNA are the electrostatic forces between the positively charged macromolecule and the negatively charged DNA, as well as specific hydrogen bonding between the side chains of polar amino acid residues and the phosphodiester backbone of the DNA. In Chapter 2, we have identified four residues on the back face of each helix are not predicted to directly contact DNA, which can be replaced with S_5 to create the hydrocarbon staple. In one effort to reduce nonspecific interactions by these residues, we synthesized a modified version of RTD-8 with mutations Lys11Ala and His15Ala. The resulting peptide, RTD-8M was then conjugated with RTD-4 to create RTD-84M, an sDBD with presumably lower nonspecific interactions than RTD-84. In a side-by-side comparison, to our surprise, RTD-84M showed no significant gain or loss of affinity compared to RTD-84, generating an almost identical EMSA binding profile, which suggests that mutations of two polar residues in such a large molecule are not significant enough to have a meaningful impact on its activity.

Table 3.2 Sequences with mutations on the back face

Monomer	Sequence
RTD-8	AcWβKRS_5HHNS_5LERKRRDHIKDSβK_{Mmt}
RTD-8M	AcWβKRS_5HHNS_5LERARRDAIKDSβK_{Mmt}

K_{Mmt}: interhelix conjugation site
S_5: (*S*)-2-(4'-pentenyl)alanine
A: alanine mutations
β: β-alanine

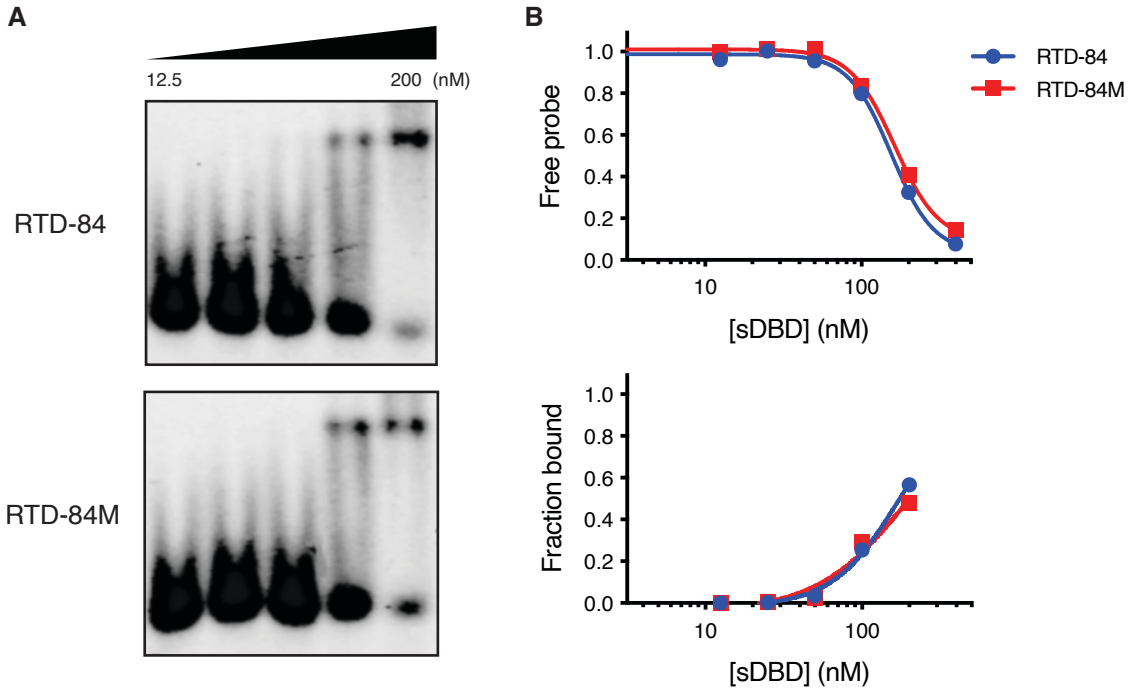
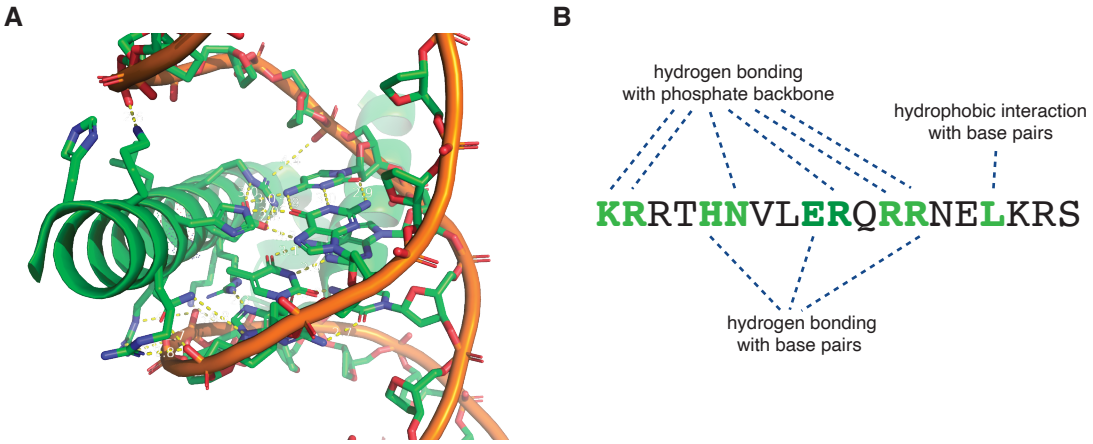


Figure 3.1 Affinity comparison of RTD-84 and RTD-84M
(**A**) EMSA gel images of both sDBDs binding to E-box DNA probe. (**B**) Quantification of affinity as measured in both free probe and fraction bound.

To test another extreme case, we wondered what would happen if we only kept the residues that are essential for DNA binding and recognition and replaced all others with alanine. I analyzed the crystal structure of Myc/Max-DNA complex and found that nine residues are in direct contact with target DNA through electrostatic interactions, hydrogen bonding and hydrophobic interactions. Intriguingly, these residues are also found to be highly conserved in multiple bHLH proteins across different species, indicating their critical role in target engagement. We constructed

both unstabilized and stapled polyalanine α-helices with these nine residues occupying the desired positions, which theoretically should only bind E-box consensus DNA due to the minimized nonspecific interactions from polar side chains. To confirm whether these artificially designed peptides remain in α-helical form, we performed CD spectroscopy and found that both showed about 60% α-helicity, even higher than their natural counterparts. Moreover, the introduction of the hydrocarbon staple did not increase α-helicity, which is not surprising given the high helical propensity of alanine. We then dimerized the individual polyalanine helices via a disulfide bond formation of the C-terminal cysteine, and tested the homodimers' DNA-binding activity in EMSA. The unstabilized homodimer had an apparent K_D of 5 μM, a nearly 30-fold decrease in affinity comparted to original sDBDs; the smeared shape of the shifted DNA band also indicates a very labile binding interaction. On the other hand, the stapled counterparts did not show any meaningful binding events at up to 10 μM despite a similar content. To further understand why the hydrocarbon staple caused a complete loss of affinity, we tested the stapled homodimer on SPR, which revealed a very fast dissociation rate. It is plausible that the increased hydrophobicity and lack of conformational flexibility from the hydrocarbon staple leads to poor solvation and unfavorable contacts between the sDBD and the DNA, destabilizing the complex.



C

	341	
human c-Myc	KCTSPRSSDTEENV	KRRTHNVLERQRRNELKRS	FFALRDQIPELENNEKA
mouse c-Myc	KCSSPRSSDTEEND	KRRTHNVLERQRRNELKRS	FFALRDQIPELENNEKA
rat c-Myc	KCSSPRSSDTEEND	KRRTHNVLERQRRNELKRS	FFALRDQIPELENNEKA
cat c-Myc	KCISPRSSDTEEND	KRRTHNVLERQRRNELKRS	FFALRDQIPELENNEKA
pig c-Myc	KCASPRSSDTEEND	KRRTHNVLERQRRNELKRS	FFARRDQIPELENNEKA
dog c-Myc	KCASPRSSDTEEND	KRRTHNVLERQRRNELKRS	FFALRDQIPELENNEKA
bovine c-Myc	KCASPRSSDTEEND	KRRTHNVLERQRRNELKRS	FFALRDQIPELENNEKA
	612		
fruit fly Myc	KDFVLGFDEADTIE	KRNQHNDMERQRRIGLKNL	FEALKKQIPTIRDKERA
	266		
human L-Myc	CHPKPVSSDTEDVT	KRKNHNFLERKRRNDLRSR	FLALRDQVPTLASCSKA
	10		
human Max	ESDEEQPRFQSAAD	KRAHHNALERKRRDHIKDS	FHSLRDSVPSLQGEKAS

Figure 3.2 Structural analysis reveals residues responsible for DNA binding and recognition
(**A**) Closer look of the protein-DNA interface with contact points labeled. (**B**) Summary of interactions between the Myc DNA-binding helix and E-box DNA. Contacting residues are labeled green. (**C**) Key contacting residues are highly conserved in bHLH proteins across species.

Table 3.3 Sequences of the polyalanine-based sDBDs

Monomer	Sequence
pA	AcWβAAAKRAAHNAAERARRAALAAAβC
st-pA	AcWβAAAAAAKRAAHNS_5AERS_5RRAALAAAβC

C: interhelix conjugation site
S_5: (*S*)-2-(4'-pentenyl)alanine
A: alanine mutations
β: β-alanine

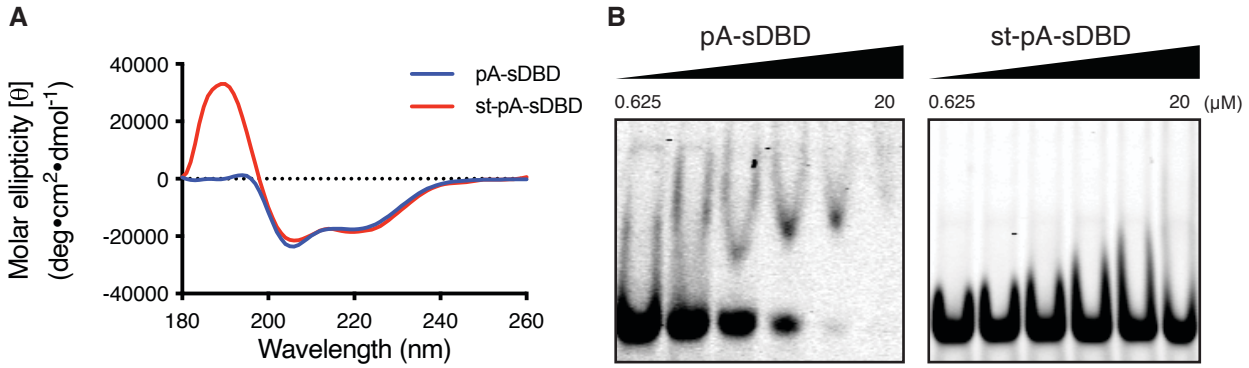


Figure 3.3 Characterization and activity of polyalanine-based sDBDs
(**A**) CD spectra and (**B**) EMSA gel images of both pA-sDBD and st-pA-sDBD.

Collectively, these preliminary results suggest that manipulating the primary sequence can result in substantial changes in the binding affinity of sDBDs. The magnitude of the changes depends on the number of residues being replaced. The results also indicate that replacing some residues will not only reduce nonspecific interactions, but also weaken the binding kinetics and stability of the complex—The crystal structure only captures a snapshot of the thermodynamic equilibrium; however, DNA binding is a dynamic process with molecular movements and conformational changes. Moreover, certain "back face" residues can also interact with water molecules and thus contribute significant solvation energy. Therefore, to find a balance between the competing effects and determine the optimal design, a more rigorous screening scheme to alter the peptide sequence, such as alanine scanning, will be necessary.

3.2.2 Additional binding motif expands the DNA interacting surface

Given that the bHLH-DNA interaction is already enthalpically driven, it would be challenging to create a more favorable enthalpic change by solely altering the peptide sequence. Instead, we aimed to create additional contacts between the two molecules to create more favorable enthalpic changes. Therefore, we proposed an scheme where a secondary DNA-binding motif is

incorporated into the sDBD core structure (Figure 3.4). The additional motif can either be a non-specific weak DNA binder, such as a DNA intercalator, or other structures targeting specific sequences.

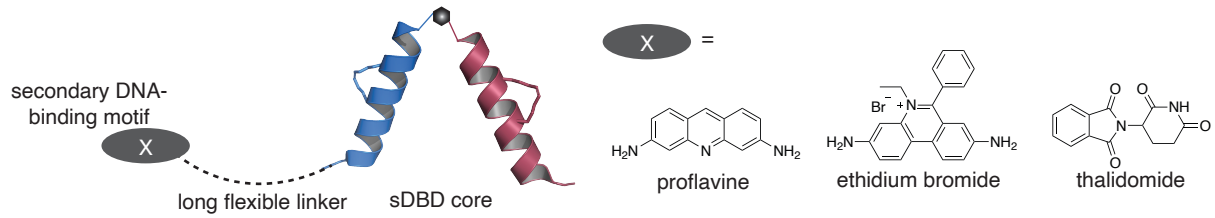


Figure 3.4 Proposed sDBD scaffold with an additional binding motif

The combination of sDBD and the additional motif should in theory create a stronger interaction with the DNA, similar to a fragment-based ligand design process. In theory, the added motif should serve as an anchor point that help dock the sDBD to the major groove, stabilizing the complex.

As a proof-of-concept study, we chose proflavine as the additional motif. Proflavine is an acriflavine derivative which is a disinfectant bacteriostatic against many gram-positive bacteria[6]. It binds to DNA by intercalating between nucleic acid base pairs with micromolar affinity[7]. We hypothesize the incorporation of this weak interaction would not dominate the DNA binding of the resulting molecule but still have meaningful impact on the activity.

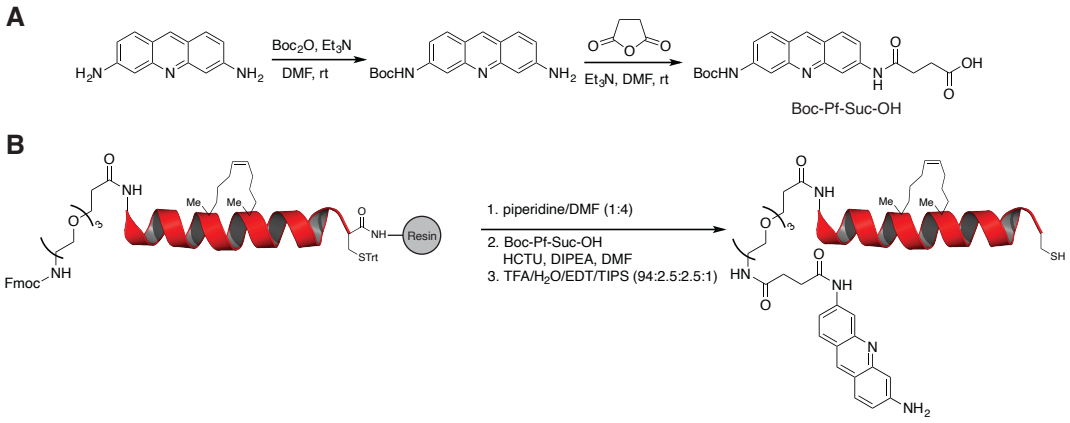


Figure 3.5 Synthetic scheme for proflavine-sDBDs
(**A**) Chemical synthesis of proflavine carboxylic acid. (**B**) Coupling of proflavine onto the N-terminus of the Myc-thiol peptide. The proflavine molecule and main helix is separated by a long PEG3 linker with enough spacing.

To couple the proflavine moiety on to the peptide, we chemically synthesized a Boc-protected carboxylic acid derivative of proflavine, Boc-Pf-Suc-OH, which is fully compatible with solid phase peptide synthesis and can be easily coupled on to the N-terminal of a stapled peptide, such as RTD-4. We utilized a PEG3 linker, consisting of a 13-atom long chain, which should give the proflavine molecule enough length and flexibility to "loop back" and bind to adjacent minor groove base pairs.

We performed EMSA to directly compare RTD-84 with proflavine-containing sDBD, Pf-RTD-84. Interestingly, the addition of proflavine resulted in a completely different binding profile (Figure 3.6), with labile interactions at lower concentrations and complete loss of stable complex as the concentration increases, usually an indication of narrower dynamic range and higher tendency for aggregation. While this preliminary result does not preclude better compounds may emerge in the future, it reveals pitfalls to consider in future trials. Adding a random DNA-binding motif doesn't represent an easy solution and more deliberate, rational designs are needed.

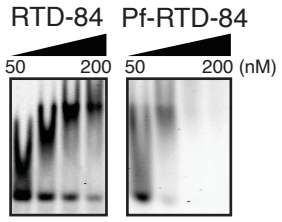


Figure 3.6 EMSA gel images of RTD-84 and Pf-RTD-84

3.2.3 Adding a second interhelix linker requires a rigorous design of new chemistry and constrains the binding conformation

As discussed in 3.1, the DNA binding of bHLH TFs is enthalpically favored but entropically disfavored. In the case of sDBDs, we would expect an even greater entropy penalty due to the absence of other structural domains that would otherwise help the basic DNA binding region fold into the proper confirmation. As depicted in Figure 3.7, the two basic helices covalently linked at the C-termini need to adopt a scissor-like *cis*-conformation for DNA binding, resulting in a loss of entropy. To mitigate this problem, we proposed to add another covalent linker at the adjacent helix turn, which can prevent the two helices from freely rotating and orient the sDBD into the desired *cis*-conformation. Consequently, such doubly linked sDBDs would not need to undergo the entropically disfavored conformational change upon DNA binding, and thus have higher binding affinity.

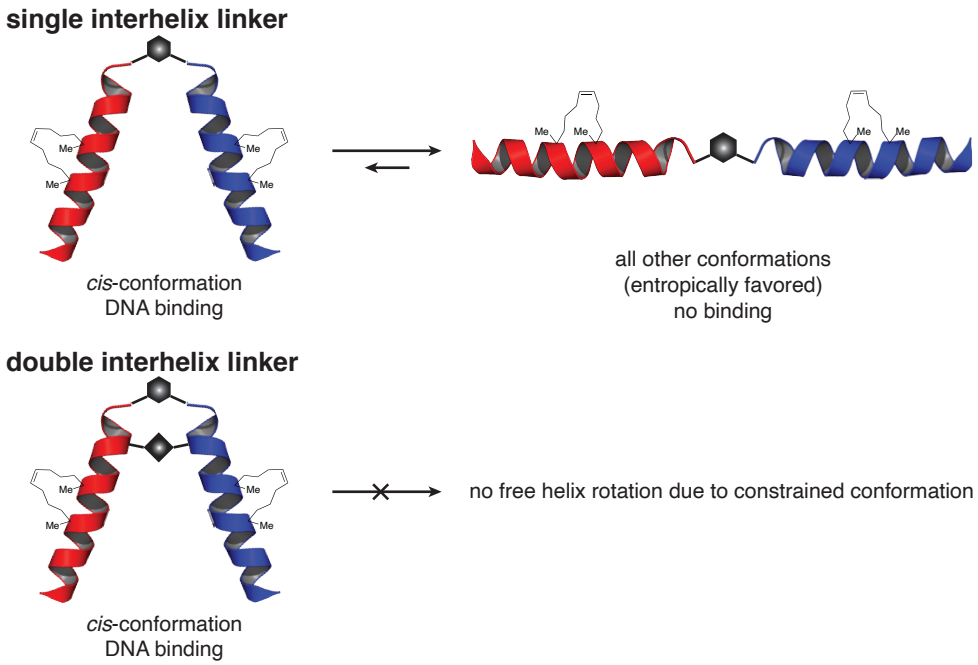


Figure 3.7 Proposed doubly linked sDBDs with constrained *cis*-conformation

The need of a second covalent linker poses additional synthetic challenges. The chemistry for such a transformation needs not only to be easy to incorporate into solid-phase peptide synthesis, but also compatible with ring-closing metathesis and thiol-maleimide conjugation. After careful considerations, we decided to utilize the copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) due to its versatility and wide applications in bioconjugations.

When incorporating CuAAC into our synthetic platform, we first attempted to identify any compatibility issues and explore alternative synthetic schemes. First of all, although unnatural amino acids containing a terminal alkyne group are commercially available, they are likely to interfere with the on-resin ring-closing metathesis due to the presence of a carbon-carbon triple bond. Therefore, we aimed to incorporate the alkyne group via post-metathesis modification on resin. We first synthetized a stapled peptide containing S-(*tert*-butylthio)cysteine, which was then selectively deprotected on resin under reductive conditions. The free cysteine then underwent an

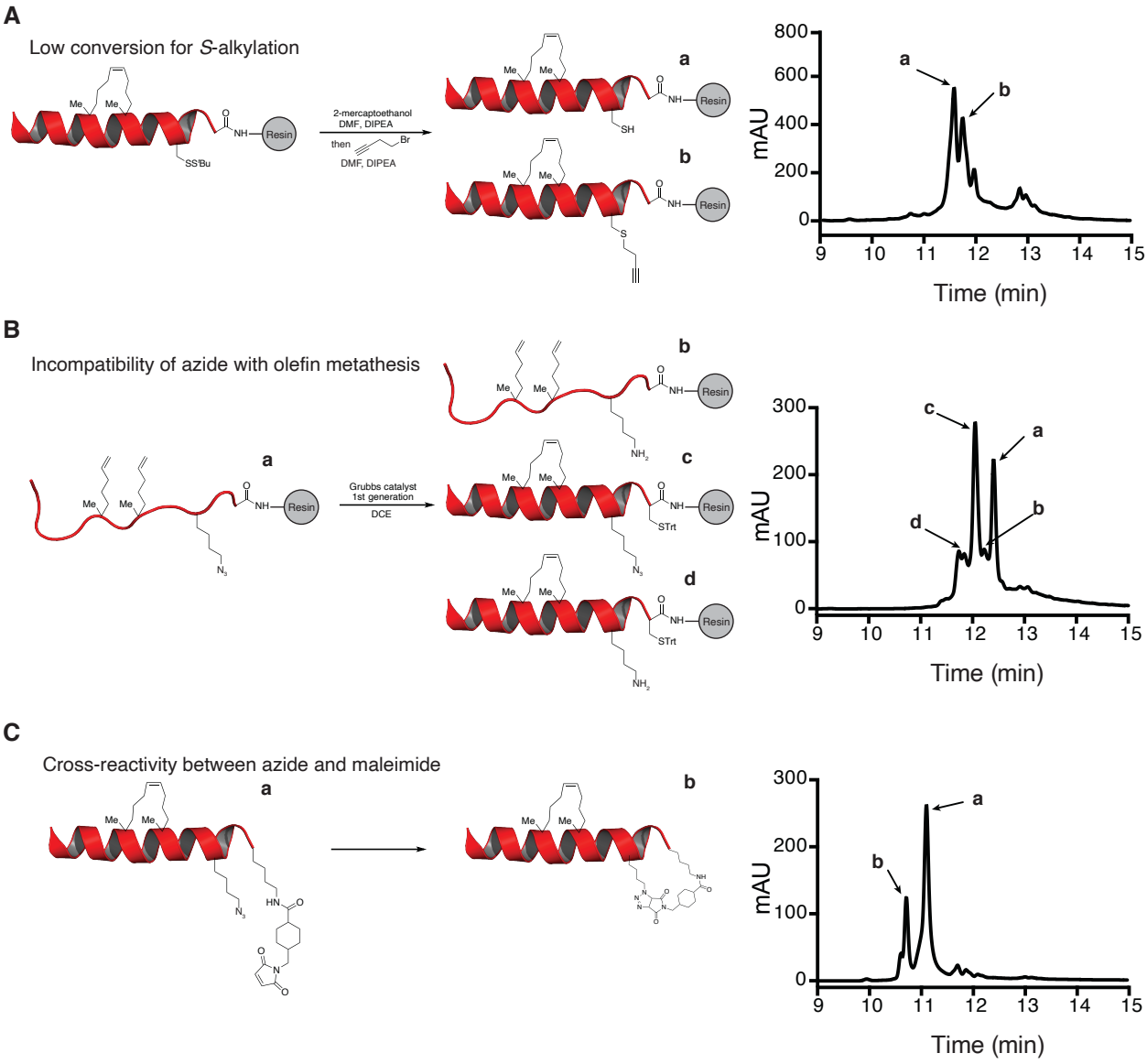


Figure 3.8 Synthetic challenges of incorporating CuAAC to the platform
(**A**) S-alkylation has low conversion with primary bromide. (**B**) Azide-containing side chain is incompatible with ring-closing metathesis, resulting in a reduction of azide to amine. (**C**) Intramolecular azide and maleimide have undesired cross-reactivity.

S_N2 alkylation to incorporate a terminal alkyne. In an effort to expand the scope of the alkyne group, we tested a series of alkyne-containing bromides with varying lengths and found that with the exception of propargyl bromide, all other primary bromides yielded unsatisfactory conversion due to their low reactivity (Figure 3.8A). With further optimizations, we found that the addition of

sodium iodide significantly catalyzed the alkylation with primary bromides and drove it to completion within 12 hours.

Due to a limited number of existing studies suggesting that azide group is compatible with olefin metathesis[8], we then questioned whether the commercially available N^{ε}-azidolysine can be directly used in stapled peptide synthesis. To confirm its compatibility, we performed ring-closing metathesis on a peptide containing a C-terminal N^{ε}-azidolysine and observed significant reduction of the azide group before the metathesis was complete (Figure 3.8B). These results suggest that the ruthenium complex leads to the reduction of the azide moiety—likely via a Staudinger reaction with the phosphine or a light-induced mechanism[9]—and that the azide functional group also needs to be incorporated after peptide stapling. Having already demonstrated that we can selectively deprotect Mmt-lysine in Chapter 2, we performed an on-resin diazotransfer reaction on the free amine to convert it to an azide group via as previously reported[10-12]. The reaction successfully went to completion after 3 hours.

After resolving the compatibility issue of CuAAC with ring-closing metathesis, we then prepared a series of monomeric peptides containing a combination of thiol/maleimide and alkyne/azide in an attempt to synthetize the doubly linked sDBDs. Intriguingly, in cases where azide and maleimide are groups present on the same peptide, we observed significant intramolecular reaction between the two groups, which was characterized by a loss of maleimide group (confirmed by addition of 2-mercaptoethanol) and a shift in LC retention time (Figure 3.8C). Therefore, when designing the precursors for doubly linked sDBDs, these two groups need to be placed on different peptides to prevent such cross-reactivity.

Taken all findings together, we created a robust scheme for the synthesis of doubly linked sDBDs, as shown in Figure 3.9. The Myc-derived peptide contains a thiol and an azide, whereas

the Max-derived one contains a maleimide and an alkyne. The two purified peptides first underwent a thiol-maleimide conjugation in pH-neutral buffer, and then CuAAC upon addition of the copper catalyst to complete the synthesis of doubly linked sDBDs.

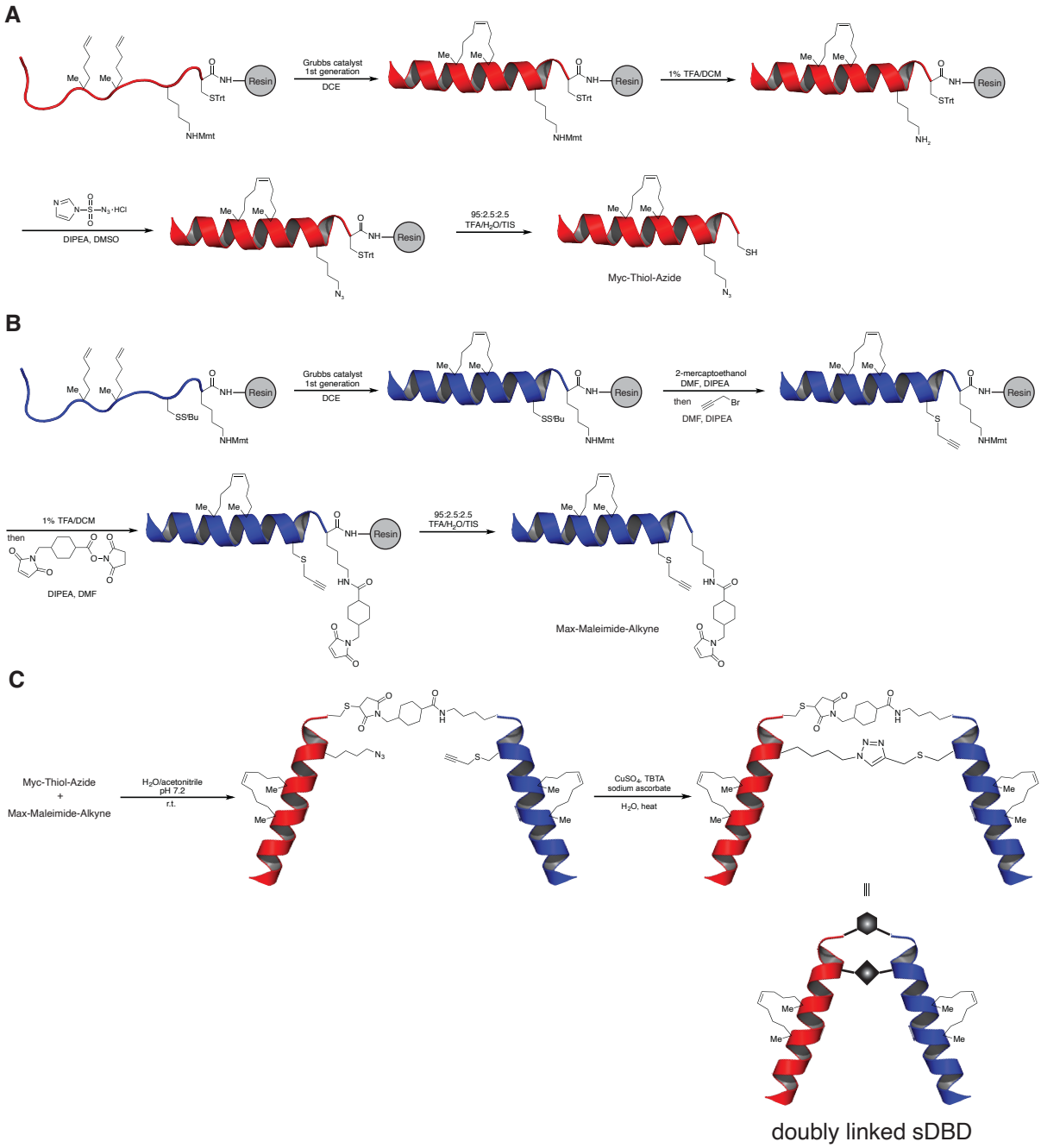


Figure 3.9 Synthetic scheme for doubly linked sDBDs
(**A**-**B**) Solid phase synthesis and modification of stapled peptides with orthogonal reactive groups. (**C**) Two-step conjugation creates two linkers between helices.

Under the developed scheme, we synthetized the doubly linked version of RTD-84, featuring an additional helical turn and a CuAAC triazole linker at the C-termini beyond the thiol-maleimide linker. Interestingly, the addition of a second linker did not improve RTD-84's activity; instead, the EMSA revealed that dlRTD-84 showed somewhat inferior binding affinity and did not form stable DNA complexes at up to 150 nM. It remains unclear why the doubly linked sDBDs did not improve the activity as expected. One possible explanation is that the stapled sDBD already has a more rigid secondary structure in place and thus is very sensitive to subtle changes to tertiary structures. Similarly, sDBDs may also require some level of flexibility to access their DNA target. A strict constraint on its conformation, although entropically favored in the binding equilibrium, may create undesirable effects on the kinetics, such as preventing proper association. We must also acknowledge that the *in vitro* results were based on one particular design and did not preclude any better compounds that may emerge from future designs. Therefore, an extensive screening of the length and rigidity of both covalent linkers might be necessary.

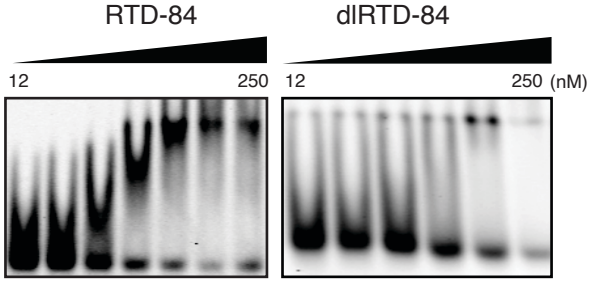


Figure 3.10 EMSA gel images of RTD-84 and dlRTD-84

3.3 Discussion

In summary, we have explored three different strategies to alter the activity of sDBDs by manipulating the primary sequence, extra binding motif and tertiary structure stabilizations. Early results suggest that these changes significantly altered the binding activity of sDBDs. However,

the compounds presented in this chapter have not been extensively explored in biochemical assays, and activity results were too preliminary for us to devise any meaningful structure activity relationships for these new sDBD designs. Rigorous studies of more compounds are thus necessary for elucidating their properties and providing guiding principles for further optimizations.

Although we have not yielded more functional sDBDs, we have further expanded our synthetic platform in our effort to prepare these newly designed scaffolds. In particular, we have established an additional dimension of orthogonality in our synthetic platform by incorporating CuAAC into our repertoire. Despite wide applications and presumable robustness of these biorthogonal chemistries, our study suggest that their use in stapled peptide and miniprotein synthesis, especially in combination, can raise a series of complications not known before and requires a trial and error learning process. Additionally, the development of new orthogonal chemistries is likely to have far-reaching impact on stapled peptide and miniprotein synthesis. To the best of our knowledge, no existing studies have extensively explored the modification and conjugation of hydrocarbon stapled peptides. The orthogonal synthetic tools we have developed in our platform are highly modular and robust, and provide access to a variety of different peptide and protein scaffolds beyond the scope of sDBDs and bHLH TFs.

3.4 Materials and methods

Solid phase peptide synthesis. All peptides were synthesized manually or by automated peptide synthesizer (Gyros Protein Technologies) using standard Fmoc-peptide chemistry on rink amide MBHA resin (EMD Millipore). Ring-closing metathesis of unnatural amino acids containing pentenyl side-chains was performed twice using 0.2 equivalents of Grubbs-I catalyst dissolved in dichloroethane under nitrogen for 2 hours. For Max peptides, C-terminal modification

was achieved by selective deprotection of Mmt-lysine using 1% trifluoroacetic acid (TFA) in dichloromethane (DCM) for 5 × 5 min, followed by coupling with 2 equivalents of SMCC and 4 equivalents of *N*,*N*-diisopropylethylamine (DIPEA) in *N*,*N*-dimethylformamide (DMF). Prior to cleavage, peptides were subjected to one final round of deprotection and capping by standard methods with acetic anhydride, fluorescein isothiocyanate (FITC) or otherwise stated linkers. After completion, crude peptide was cleaved using 95:2.5:2.5 TFA/water/triisopropylsilane, dissolved in 1:1 acetonitrile/water, and purified by reverse-phase high-pressure liquid chromatography (HPLC) using a C18 column (Agilent). Compound identification and purity was assessed using coupled liquid chromatography mass spectrometry (LCMS) (Agilent).

Covalent conjugation of peptides. After purification by HPLC, thiol-containing Myc peptide and maleimide-containing Max peptide were mixed in equal amount in 4:1 water/acetonitrile with 0.1 M phosphate and pH 7.2. The reaction was monitored using LCMS and the final product was purified by HPLC.

Circular dichroism spectroscopy. A circular dichroism spectrometer (Jasco J-1500) was used to determine the α-helical content of all monomeric and dimeric peptides. Peptides were dissolved into 10 mM potassium phosphate buffer at pH 7.4 to a final concentration of 25–250 mM in a quartz cuvette with a path length of 0.1 cm. Ellipticity values were taken at 1 nm intervals between 180 and 260 nm. The percentage helicity was calculated from the ellipticity at 222 nm using helical models as previously reported.

Electrophoretic mobility shift assay (EMSA). Recombinant protein or synthetic peptides were diluted to desired concentrations varying from 1 to 500 nM. A 5'-infrared dye-labeled double-stranded DNA oligonucleotide with the E-box consensus binding site is then added (5'-ACCCCAC<u>CACGTG</u>GTGCCT-3'). The mixture was incubated at 4°C for 30 min. Final

concentrations were: 20 mM HEPES (pH 8.0), 150 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1 mg/mL BSA, 5% glycerol, and 5 nM oligonucleotide probe. Complexes were resolved on 6% acrylamide gels (40 mM Tris, 45 mM borate, 1 mM EDTA, pH 8.3), and gels were scanned using an infrared imager (Li-Cor Odyssey). Gel images were analyzed using ImageJ software and binding results were plotted and fitted using GraphPad Prism 7.

Synthesis of Boc-Pf-Suc-OH. The synthesis of Boc-proflavine was performed as previously reported[1]. To a solution of di-*tert*-butyldicarbonate (4.4 mL, 19.12mmol) in dry acetone (150 mL), proflavine (2 g, 9.56 mmol) was added. The reaction mixture was heated at reflux for 8 hours and then concentrated up to dryness. The crude product thus obtained was purified by flash column chromatography using silica gel as adsorbent. With hexane-ethyl acetate (1:1) as eluent, followed by ethyl acetate and then ethyl acetate-acetone-triethylamine (5:5:1), (6-amino-3-acridinyl)carbamic acid *tert*-butyl ester (Boc-proflavine) was obtained (2.3 g, 59%). Boc-proflavine was then dissolved in DMF and added succinic anhydride (1.1 eq) and triethylamine (2 eq). The mixture was stirred at room temperature for 12 hours. Upon completion, the reaction was worked up with acidic aqueous solution and extracted with ethyl acetate. The combined organic phase was then concentrated and purified by flash column chromatography using ethyl acetate as eluent to obtain Boc-Pf-Suc-OH.

On resin *S*-alkylation. On a peptide containing *S*-(*tert*-butylthio)cysteine on resin, selective deprotection was achieved by agitating the resin under nitrogen flow with 2-mercaptoethanol/DMF/DIPEA (1:1:0.2) for 2 × 2 hours. Upon deprotection, alkyl bromide (10 eq.), DIPEA (20 eq.), sodium iodide (optional, 1 eq. when using primary alkyl bromide) and DMF were added to the resin and agitated under nitrogen flow for 12 hours.

Synthesis of imidazole-1-sulfonyl azide hydrochloride (ISA•HCl). The synthesis was performed as previously described[12]. Sulfuryl chloride (16.1 mL, 200 mol) was added drop-wise to an ice-cooled suspension of NaN_3 (13.0 g, 200 mmol) in MeCN (200 mL) and the mixture stirred overnight at room temperature. Imidazole (25.9 g, 380 mmol) was added portion-wise to the ice-cooled mixture and the resulting slurry stirred for 3 hours at room temperature. The mixture was diluted with EtOAc (400 mL), washed with H_2O (2 × 400 mL) then saturated aqueous $NaHCO_3$ (2 × 400 mL), dried over $MgSO_4$ and filtered. A solution of HCl in EtOH [obtained by the drop-wise addition of AcCl (21.3 mL, 300 mmol) to ice-cooled dry ethanol (75 mL)] was added drop-wise to the filtrate with stirring, the mixture chilled in an ice-bath, filtered and the filter cake washed with EtOAc (3 × 100 mL) to give ISA•HCl as colorless needles (26.4 g, 63%).

On resin diazotransfer. A previously reported protocol was modified for stapled peptides for this purpose[11]. On a peptide containing Mmt-lysine on resin, selective deprotection was achieved by treating the resin with 1% TFA in DCM for 5 × 5 min. The deprotected resin was then swollen with DMSO, before the addition of ISA·HCl (3 eq.), DIPEA (9 eq.), and $CuSO_4$ (0.01 eq.). The reaction mixture was gently agitated under nitrogen flow for 3 hours.

References

1. Travers, A., DNA-Protein Interactions. Springer Netherlands : Imprint : Springer,: Dordrecht, 1993; p. 1 online resource (192 pages). SpringerLink http://dx.doi.org/10.1007/978-94-011-1480-6 e-Book.

2. Dahirel, V.; Paillusson, F.; Jardat, M.; Barbi, M.; Victor, J. M., Nonspecific DNA-protein interaction: why proteins can diffuse along DNA. *Phys Rev Lett* **2009,** *102* (22), 228101.

3. Slutsky, M.; Mirny, L. A., Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys J* **2004,** *87* (6), 4021-35.

4. Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A., Structural and thermodynamic strategies for site-specific DNA binding proteins. *Structure* **2000,** *8* (10), 1015-23.

5. AG, K. u.; Sieber, M.; Meierhans, D.; Allemann, R. K., Thermodynamics of the DNA binding reaction of transcription factor MASH-1. *Biochemistry* **1998,** *37* (12), 4217-23.

6. Li, H. J.; Crothers, D. M., Relaxation studies of the proflavine-DNA complex: The kinetics of an intercalation reaction. *Journal of Molecular Biology* **1969,** *39* (3), 461-477.

7. Neidle, S.; Pearl, L. H.; Herzyk, P.; Berman, H. M., A molecular model for proflavine-DNA intercalation. *Nucleic Acids Res* **1988,** *16* (18), 8999-9016.

8. Kanemitsu, T.; Seeberger, P. H., Use of olefin cross-metathesis to release azide-containing sugars from solid support. *Org Lett* **2003,** *5* (24), 4541-4.

9. Chen, Y.; Kamlet, A. S.; Steinman, J. B.; Liu, D. R., A biomolecule-compatible visible-light-induced azide reduction from a DNA-encoded reaction-discovery system. *Nat Chem* **2011,** *3* (2), 146-53.

10. Marine, J. E.; Liang, X.; Song, S.; Rudick, J. G., Azide-rich peptides via an on-resin diazotransfer reaction. *Biopolymers* **2015,** *104* (4), 419-26.

11. Castro, V.; Blanco-Canosa, J. B.; Rodriguez, H.; Albericio, F., Imidazole-1-sulfonyl azide-based diazo-transfer reaction for the preparation of azido solid supports for solid-phase synthesis. *ACS Comb Sci* **2013,** *15* (7), 331-4.

12. Goddard-Borger, E. D.; Stick, R. V., An efficient, inexpensive, and shelf-stable diazotransfer reagent: imidazole-1-sulfonyl azide hydrochloride. *Org Lett* **2007,** *9* (19), 3797-800.

Chapter 4

Beyond the DNA-Binding Domain: Synthetic Transcription Factors[1]

4.1 Structural importance of the tetrahelix bundle and design of synthetic transcription factors

Up to this point, our core sDBD design shares a common scheme for tertiary structure stabilization—a covalent C-terminal linker, or two for doubly linked sDBDs, helps orient the two helices that bind to target DNA in a cooperative manner. Our studies suggest that the activity of sDBDs are highly sensitive to their tertiary structure, which can be hard to modulate by a single covalent linker. Although tertiary structure stabilization can theoretically be precisely controlled by proper use of two covalent linkers, our early results suggest that the reduced conformational flexibility may have adverse impact on sDBD's activity.

In an effort to explore alternative ways to program tertiary structure stabilization, we would like to revisit the way natural proteins behave. In Myc/Max and other bHLH proteins, the proper orientation of two DNA-binding helices is prompted by the formation of a tetrahelix core. A closer look into the protein-protein interfaces identified a number of hydrophobic residues buried in the

[1] The work presented in this chapter is currently in preparation for publication. Dr. Thomas Speltz and the author contributed equally to this work. Dr. Thomas Speltz, Dr. Raymond Moellering and the author contributed to the design and synthesis of all compounds, performed biochemical and cellular experiments, analyzed data, and wrote the manuscript for publication. Mr. Zeyu Qiao and Dr. Sean Fanning performed biochemical experiments, collected X-ray structure data, and analyzed data.

interior of the tetrahelix bundle, where they pack together and exhibit strong van der Waals interactions that stabilize the structure of homo- or heterodimers[1]. With these insights, we considered an alternative approach for assembling the tertiary structure, where shorter peptides encompassing the basic (B) and minimal leucine zipper (Z) helices are synthesized and then ligated to build the larger B-Z conjugate. Due to the preservation of the key hydrophobic residues in the resulting scaffold, which we call synthetic transcription factors (STFs), we hypothesize that it would be capable of forming a tetrahelix core via self-association in a similar fashion to bHLH TFs.

4.2 Results

4.2.1 Conjugation positions have a large impact on STF activity

Since Max can form DNA-binding homodimers[2], we synthesized a "natural" mimic, M1BZ, where the intervening loop between B and Z helices from the same monomer is replaced by an orthogonal interhelix linker between functionalized sidechains at specific locations. In parallel, we considered a non-natural approach, where B and Z helices from opposing monomers are connected at specific sites not involved in binding, resulting in a "cross-dimer" bHLH mimic that should self-associate in a sandwich fashion to recreate the tetrahelix core (Figure 4.1A). Functionally, we also predict that natural bHLH mimics could theoretically act as Max dominant-negatives by affecting both protein-protein and protein-DNA interactions, whereas the cross-dimer mimetics would be incapable of interfering with protein-protein interactions within the extended Myc/Max interaction network.

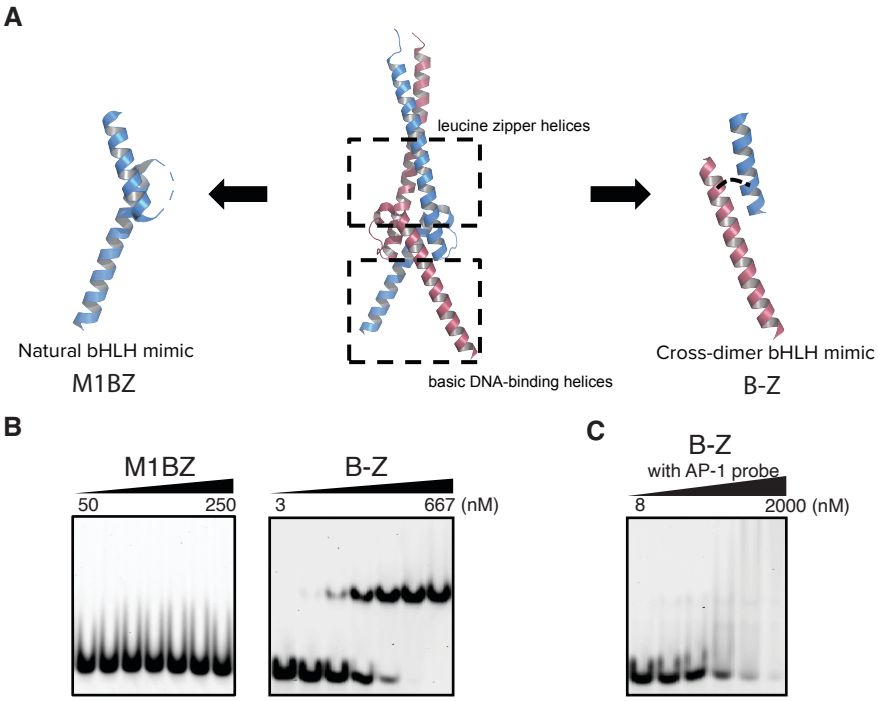


Figure 4.1 Design of STFs from Max

(A) Convergent ligation of leucine zipper (Z) and basic (B) helices from the same Max monomer (M1BZ, left) should dimerize to form the natural tetrahelix homodimeric DNA binding domain. Cross-dimer ligation of Z and B helices from opposing monomers (B-Z, right) would result in non-natural mimics that should assemble the tetrahelix bundle in a "sandwich" dimer. **(B)** Representative EMSA gel showing binding of an E-box containing oligonucleotide across a dose range of the natural and cross-dimer mimics M1BZ and B-Z, respectively. (**C**) EMSA gel showing binding of the cross-dimer mimic with AP-1 consensus probe, in a comparison with E-box consensus probe to show specificity.

Prototypes of each design were synthesized and tested by EMSA. Strikingly, we observed potent and stable DNA binding by the non-natural, cross-dimer B-Z mimic, whereas the natural bHLH M1BZ mimic did not show any appreciably binding in a similar concentration range (Figure 4.1B). Furthermore, as a basis to quantify specific DNA binding, we employed EMSA using a consensus E-box oligonucleotide and a control oligonucleotide that contained the unrelated AP-1 consensus binding site. To our delight, the basic-zipper helix hybrid containing no helix stabilizing elements (B-Z) potently bound E-box containing DNA with an apparent K_D of 16 nM, and no stable binding was observed with the control AP-1 oligonucleotide (5'-TGA C/G TCA-3', Figure 4.1C). These preliminary results indicate the STF scaffolds may possess superior DNA binding

affinity and sequence specificity compared to sDBDs, which represent important new avenues of further investigation.

4.2.2 STFs show superior affinity and specificity and effectively compete with Myc/Max DNA binding

Based on these results, and the potentially unique DNA-targeting capacity of the cross-dimer compounds, we devised synthetic schemes to create B-Z mimics with both secondary and tertiary domain stabilization, which was enabled by our established synthetic platform. Similar to the synthesis of sDBDs, each helix features an *i*,*i* + 4 hydrocarbon staple as well as an orthogonal conjugation synthon, which in this case included a thiol on the Z helix and a maleimide on the B helix. Convergent synthesis and modification of each stabilized helix followed by interhelix ligation in aqueous solution produce an STF of approximately 6 kDa (Figure 4.2).

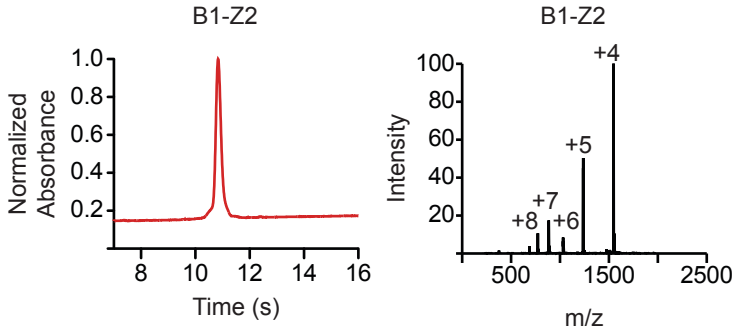


Figure 4.2 LC-MS characterization of a representative STF

We then undertook a focused medicinal chemistry campaign to understand the structural elements required for tight and specific DNA binding around the cross-dimer, B-Z domain architecture. We synthesized two parallel libraries of basic and zipper helices with promising stapling positions, peptide lengths and suitable modifications for stability (Table 4.1). Each individual modified zipper or basic helix peptide in the library was ligated with the corresponding

unstapled helix partner (B or Z alone) for controlled comparison of individual structural changes and corresponding changes in DNA binding.

Table 4.1 Sequences of STF library

Monomer	Sequence
Basic helix	
B	AcKRAHHNALERKRRDHIKDSFHK_{Mmt}LRDSVP
B1	AcKRS_5HHNS_5LERKRRDHIKDSFHK_{Mmt}LRDSVP
B2	AcKRAHHNALERKRRDHIKDSF$S_5$$K_{Mmt}LRS_5$SVP
B3	AcKRAHHNALERS_5RRDS_5IKDSFHK_{Mmt}LRDSVP
B4	AcKRAHHNS_5LERS_5RRDHIKDSFHK_{Mmt}LRDSVP
B5	AcKRS_5HHNS_5LERA_{ib}RRDHIKDSFHK_{Mmt}LRDSVP
B6	AcKRS_5HHNS_5LERkRRDHIKDSFHK_{Mmt}LRDSVP
B7	AcKRA_{ib}HHNS_5LERS_5RRDHIKDSFHK_{Mmt}LRDSVP
B8	AcKRA_{ib}HHNALERS_5RRDS_5IKDSFHK_{Mmt}LRDSVP
B9	AcHNALERKRRDHIKDSFHK_{Mmt}LRDSVP
BLI1	AcPRFQSAADKRAHHNALERKRRDHIKDSFHK_{Mmt}LRDSVP
B11	AcPRFQSAS_5DKRS_5HHNALERKRRDHIKDSFHK_{Mmt}LRDSVP
Zipper helix	
Z	AcSRAQILCKATEYIQYN_LRRKN
Z1	AcSRAS_5ILCS_5ATEYIQYN_LRRKN
Z2	AcSRAQILCKATEYIQ$S_5$$N_LRRS_5$N
Z3	AcSRAQILCQATEYIQ$S_5$$N_LRRS_5$N
Z4	AcSRA_{ib}QILCKATEYIQ$S_5$$N_LRRS_5$N
Z5	AcSRAQILCKATEYIQYN_LR
ZLI1	AcSRAQILCKATEYIQYN_LRRKNHTHQQDIDDLK
Z6	AcSRA_{ib}QILCKATEYIQ$S_5$$N_LRRS_5$LHTHE
Z8	AcSRAQILCKATEYIQYLRS_5KIHS_5LE

C/K_{Mmt}: interhelix conjugation sites
S_5: (*S*)-2-(4'-pentenyl)alanine
A_{ib}: 2-aminoisobutyric acid
N_L: norleucine
k: D-lysine

Within the basic helix library, we found that truncation of even a few N-terminal residues (B9) abrogated DNA-binding, whereas N-terminal extension alone (BLI1) did not significantly alter affinity. The location of *i*,*i* + 4 staples on the back face of the basic helix significantly impacted binding affinity. The most N- and C-terminal staple positions largely maintained tight binding affinity, and central staples reduced affinity by 10- to 20-fold (Figure 4.3A). Within the zipper helix library, truncation of the C-terminal residues (Z5), or introduction of the side chain staples around the inter-helix ligation site (Z1) significantly reduced affinity, whereas peptides

with C-terminal extension (ZLI1) or introduction of staples at other positions largely retained tight binding (Figure 4.3B).

Using these structure-activity relationship determinants as a guide, we subsequently synthesized a library of STFs that contained stabilizing modifications in both helices, and identified several lead compounds for further study (Figure 4.3C). These included B1-Z2, which was based on and had a similar binding affinity to the original unmodified B-Z scaffold. B6-Z4, with a key mutation on the B helix as discussed above, only showed a slight decrease in binding affinity compared to B1-Z2, which may have significant impact on its *in situ* and *in vivo* activity and stability. Meanwhile, the helix-templating A_{ib} residue did not seem to improve the binding affinity. Notably, two additional leads, B11-Z6 and B11-Z8, encompassed changes predicted to preserve or improve binding and augment metabolic stability, as discussed below. Notably, B11-Z6 had almost identical affinity to the progenitor B-Z (K_D = 18 nM), whereas B11-Z8 exhibited more potent binding (K_D = 3 nM) that is equivalent to reported affinities for full-length Myc/Max and Max/Max to E-box DNA[3-5]. The gain of activity in B11-Z8 is likely due to an elongation and a shift in staple positions at both ends (N-terminus of B and C-terminus of Z). The additional residues beyond the essential binding region have been reported to create positive impact on binding activity and complex stability[6]. Moreover, moving the two hydrocarbon staples further away from the interacting region possibly minimizes their disruptions to the binding surface, increasing the activity while still maintaining stabilized α-helical form.

Taken together, all these lead STFs showed potent binding across a wide dynamic range, with apparent K_D at low nanomolar concentrations, a significant improvement over sDBDs.

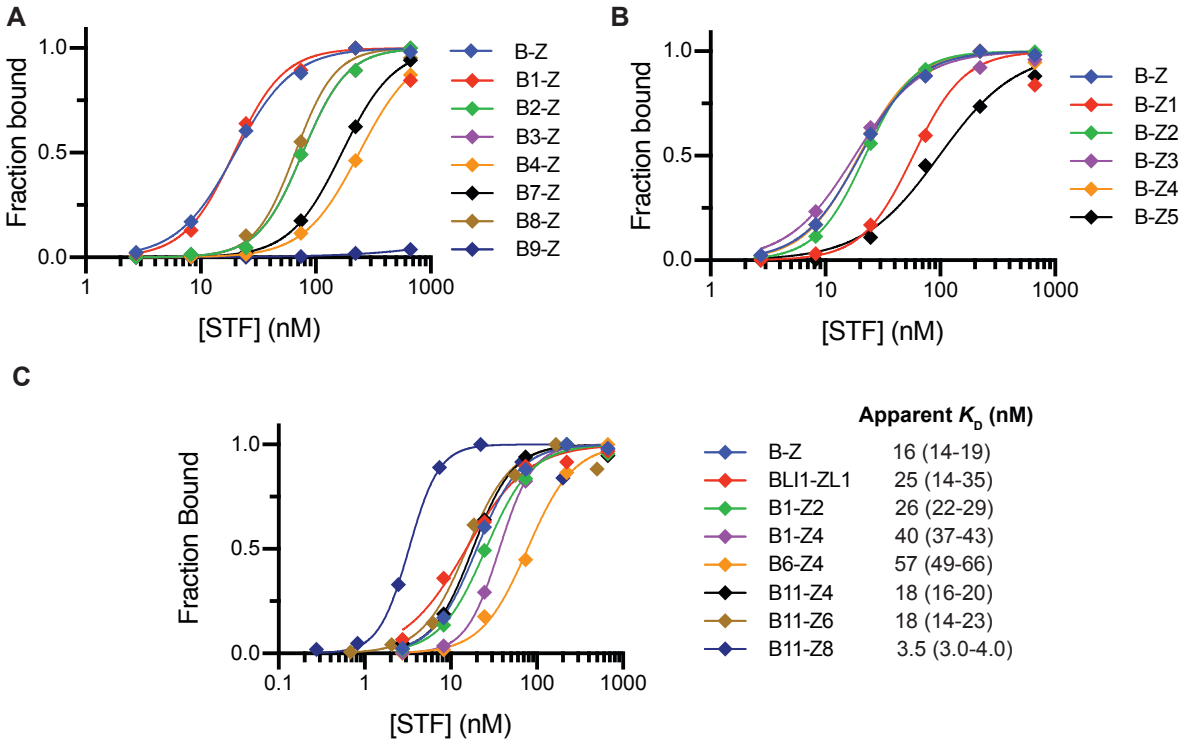


Figure 4.3 EMSA binding curves of the STF library

(**A**–**B**) Independent screening of each individual basic or zipper helix by quantifying the E-box DNA bound by their corresponding STF. (**C**) Binding curves of the lead STFs that emerged from the independent helix screening with their apparent K_D. The numbers in parentheses represent 95% C.I. from $n = 3$ independent replicates.

To test the sequence specificity of the STF library, we also employed EMSA in a competitive format similar to what we have used for sDBDs. Addition of unlabeled competitor E-box oligonucleotide (E-boxC1) confirmed specific binding of four B-Z, B1-Z2, B11-Z6, and B11-Z8 to the fluorophore-labeled E-box probe (Figure 4.4). Competitor DNA containing two mutations in the E-box site (E-boxCM2) or replacement of the E-box sequence entirely (E-boxCM2) showed greatly reduced competition for all four STFs, confirming that these STFs, unlike sDBDs, exhibit specific DNA binding that is comparable to full-length bHLH proteins. In summary, these structure activity relationship studies confirmed that stabilization of secondary and tertiary structure on both B and Z helices to template the formation of oriented DNA-binding helices is a more successful design compared to the original sDBDs.

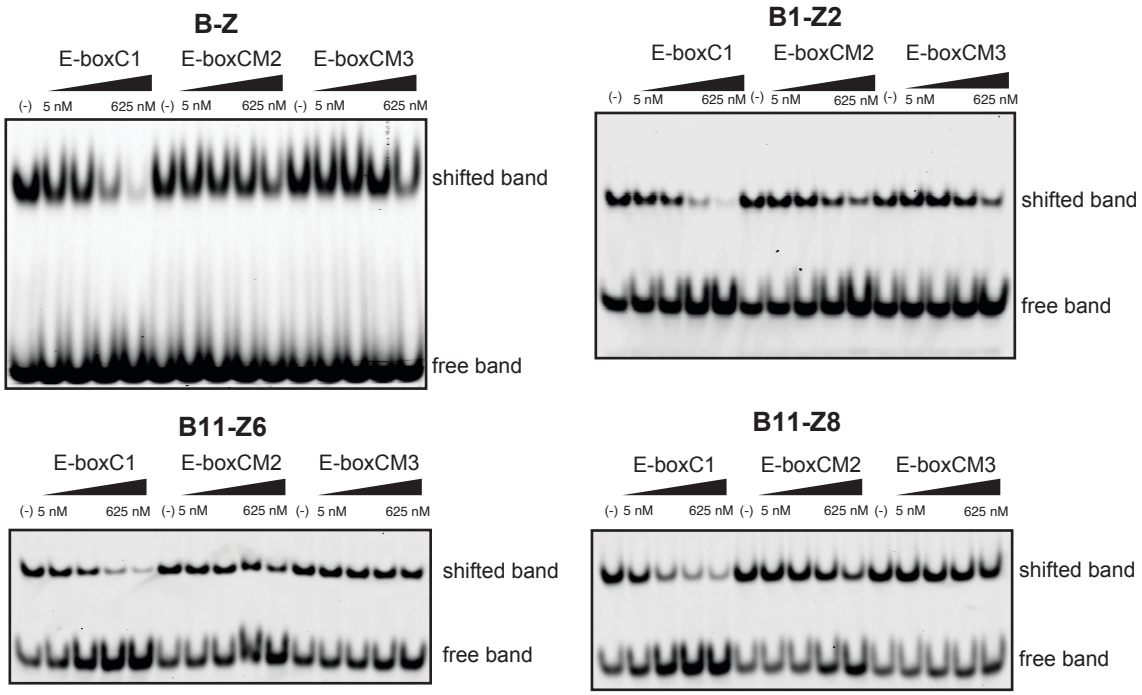


Figure 4.4 Competition EMSA gels of STFs binding to E-box oligonucleotide probe in the presence of unlabeled competitor DNAs

We next tested whether lead STFs can compete for E-box consensus with bHLH proteins. To determine their competitive properties, we added STFs in increasing concentration to the mixture of Myc/Max and E-box probe. Some STFs, such as B-Z and B1-Z2, despite their high affinity and specificity, did not show appreciable competition with Myc/Max at up to 1000 nM. By contrast, two lead STFs featuring extended helices, yielded promising competition pattern, as characterized by gradual replacement of Myc/Max-bound shifted band by STF-bound shifted band. The most effective compound, B11-Z8, with an IC_{50} of 165 nM, also happens to be the one with the lowest apparent K_D. Collectively, the results from direct DNA binding and Myc/Max competition experiments reconcile and indicate that the elongation designs we employed here significantly improved the potency of resulting STFs, making them promising inhibitors of bHLH-DNA binding *in vitro*.

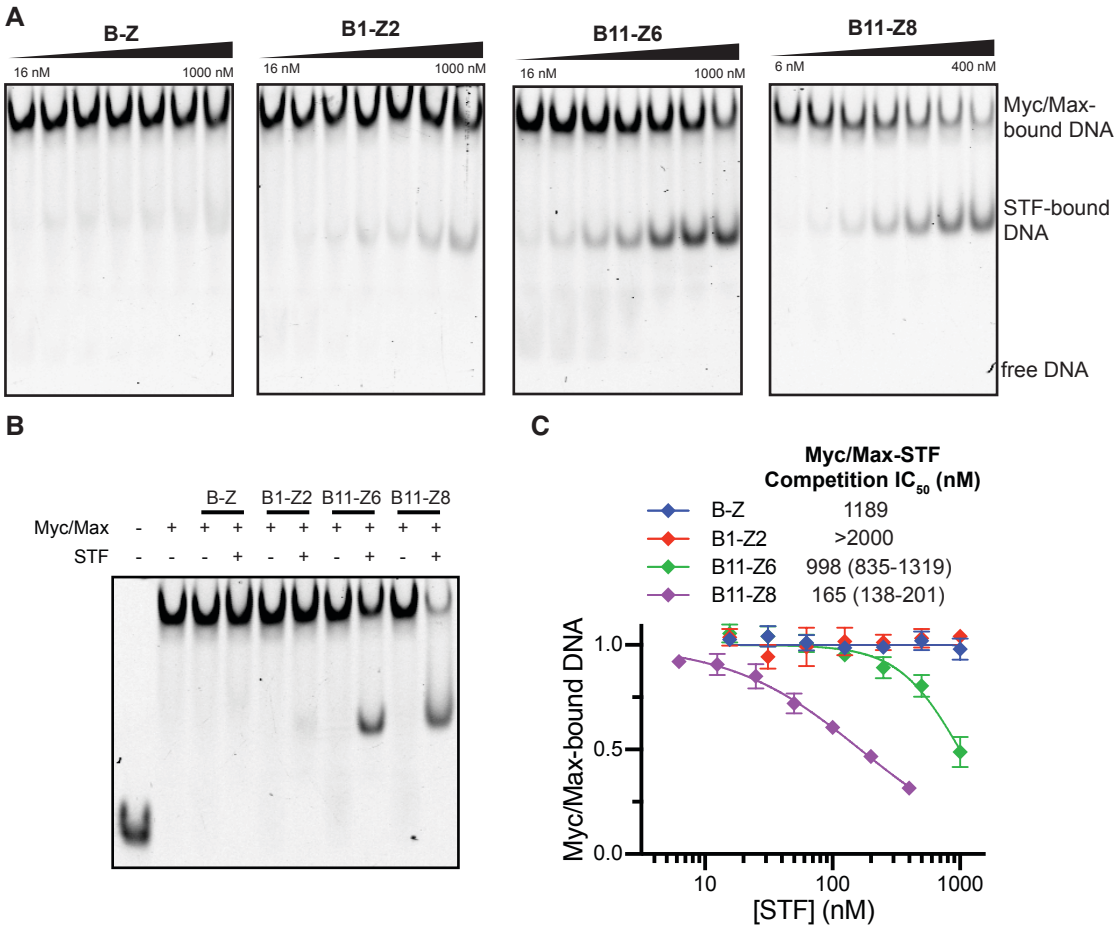


Figure 4.5 Competition of STFs with Myc/Max protein

(**A**) Dose-dependent competition of STF with Myc/Max-bound DNA. (**B**) Side-by-side comparison of four lead STFs in competition with Myc/Max. (**C**) Competition curves quantified from the dose-dependent competition. The IC_{50}'s were determined by a nonlinear fitting of the disappearance of the Max-bound shifted band. The numbers in parentheses represent 95% C.I. from $n = 3$ independent replicates.

4.2.3 Secondary and tertiary domain preorganization promotes thermal and proteolytic stability

We next performed CD spectroscopy to characterize the structural stability of the STFs. The spectrum of the non-stapled B-Z hybrid confirmed that it is largely unstructured (Figure 4.6A). By contrast, spectra of all other STFs with hydrocarbon staples showed strong ellipticity minima at 208 and 222 nm, consistent with a predominantly α-helical structure. Temperature-dependent

CD confirmed high thermal stability of the stabilized B1-Z2 hybrid, which showed little change across a wide temperature range and remained predominantly helical even at 85 °C (Figure 4.6B).

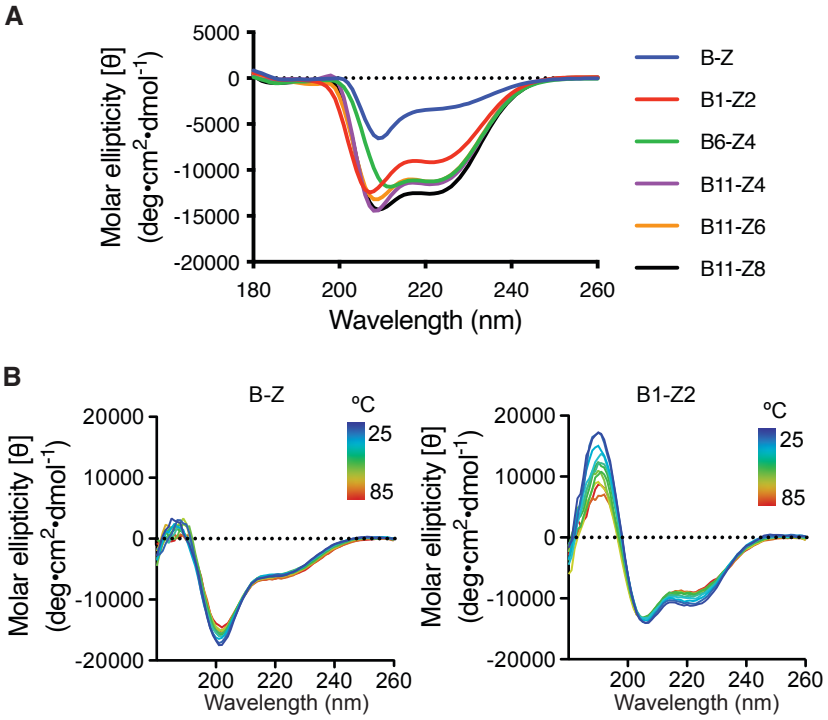


Figure 4.6 Secondary structure of STFs characterized by CD spectroscopy
(**A**) CD spectra of lead STFs. (**B**) Temperature-dependent CD scan indicating change in secondary structure and their thermal stability.

To ascertain how the structural stability impacts biological stability and activity, we performed a series of experiments aimed at quantifying the chemical and proteolytic sensitivity of unmodified and stabilized STFs. To start, we employed a kinetic trypsin stability assay, which we expected to be very active against arginine- and lysine-rich DNA binding domains. The wild-type B-Z hybrid was rapidly proteolyzed in this assay, with a measured half-life ($t_{1/2}$) of approximately 33 seconds (Figure 4.7). Mass spectrometry analysis of the resulting proteolysis products revealed that many trypsin sites were susceptible to cleavage in B-Z, consistent with largely disordered structure. Introduction of side-chain staples clearly protected internal and adjacent cleavage sites in stabilized molecules, leading to increases of 4–10 fold in half-life, however some distant sites

remained susceptible to cleavage. Targeted helix-stabilizing substitutions, such as 2-aminoisobutyric acid (A_{ib}) or synonymous mutations not recognized by trypsin, further reduce proteolytic sensitivity. Some changes, such as replacement of Ala with A_{ib} and Lys-to-Gln near the N-terminus of the zipper helix (e.g., B1-Z4), minimally affected binding, but protected adjacent sites from proteolysis. Combining these approaches, we found that introduction of a D-lysine into the same position—thereby generating a macromolecule that only differs from the precursor at one chiral center—significantly reduced proteolytic sensitivity of the helix core while largely maintaining strong binding affinity (B6-Z4, $t_{1/2}$ = 1,583 sec). The more potent lead molecules B11-Z6 and B11-Z8 embodied a combination of stabilizing modifications balanced with retained or improved binding affinity. Particularly, B11-Z6 demonstrated a nearly 10-fold improvement in proteolytic stability relative to the unstapled B-Z. Together these proteolytic data confirm that global and local stabilization of STF structure can significantly alter proteolytic stability, however these targeted modifications must be balanced with subtle changes in structure that can impact DNA-binding.

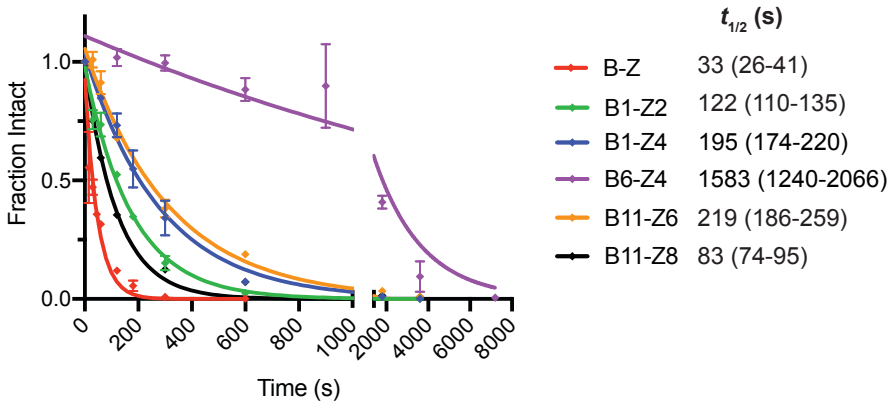


Figure 4.7 LC-MS quantification of STF stability to trypsin
The curves and $t_{1/2}$ were determined using a nonlinear one-phase decay fit. The numbers in parentheses represent 95% C.I. from n = 2 independent replicates.

While recombinant protease assays can pinpoint sensitive positions within the STF structure, it is not clear whether trypsin-like or other proteolytic activity in the cellular environment will be as robust as in this *in vitro* assay. Therefore, we studied the chemical and proteolytic stability of these molecules directly in the cellular environment. Previous studies employing thiol-maleimide conjugate chemistry, for example in antibody-drug conjugates, have reported appreciable hydrolysis and cleavage in physiologic environments[7, 8]. Incubation of B-Z and B1-Z2 in media supplemented with 10% fetal bovine serum (FBS) confirmed ring-opening hydrolysis after extended periods of time, however, we did not observe any cleavage of the B- and Z-helices, and hydrolysis did not affect DNA binding activity. To directly determine whether any proteolytic activity or hydrolysis encountered in conditioned media may impact the biochemical activity of these molecules, we directly assayed the STFs exposed to prolonged incubation in conditioned media for DNA binding activity (Figure 4.8A). Unmodified (B-Z) and stabilized (B1-Z2) STFs largely retained DNA binding activity after 24 hours in conditioned media. Extended incubation timepoints further highlighted marked differences between compounds, with B-Z losing all binding activity by 48 hours, which contrasted with the retention of active molecule after 72 hours for B1-Z2. Intriguingly, B11-Z6 and B11-Z8, both with strong potency and effective Max/Max competition, also showed increased stability in conditioned media with retention of activity after days of exposure (Figure 4.8B). Together, these data confirm that vigilant introduction of local and global stabilizing modifications can yield STFs with potent DNA binding activity and vastly improved protease resistance.

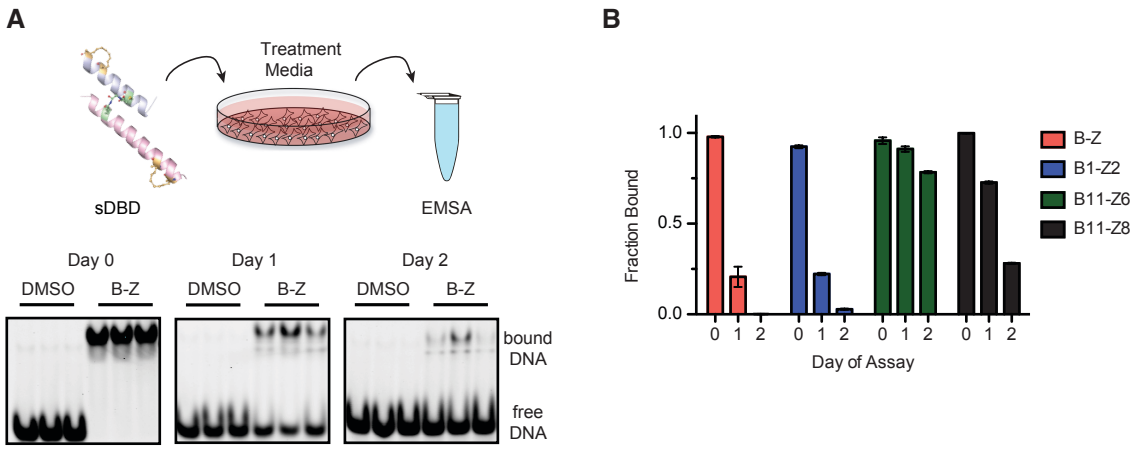


Figure 4.8 Conditioned media stability assay of STFs
(**A**) EMSA gels showing E-box DNA binding activity of B-Z after incubation in HeLa cell media for 0, 24, 48, or 72 hours. (**B**) Relative quantification of binding activity of B-Z, B1-Z2, B11-Z6, and B11-Z8 over incubation time.

4.2.4 STFs are cell permeable, distribute to the cytosol and nucleus intact, and specifically bind E-box-containing target genes

To directly visualize STF cell penetration and distribution, we synthesized FITC labeled versions of B-Z (B-Z-FITC) and B1-Z2 (B1-Z2-FITC), and imaged cells incubated with each molecule by confocal fluorescence microscopy. We observed time-dependent cellular uptake of both compounds that peaked after 8-12 hours. Consistent differences in sub-cellular distribution were observed, with significantly increased cellular and nuclear distribution being observed in cells treated with B1-Z2-FITC (Figure 4.9A). Quantification of fluorescence intensity and distribution on a per-cell basis highlighted predominantly punctate distribution of B-Z-FITC, compared to increased cellular and nuclear localization for B1-Z2-FITC (Figure 4.9B). We hypothesize that these distribution differences may stem from collective differences in proteolytic stability, endosomal escape and potential DNA binding activity.

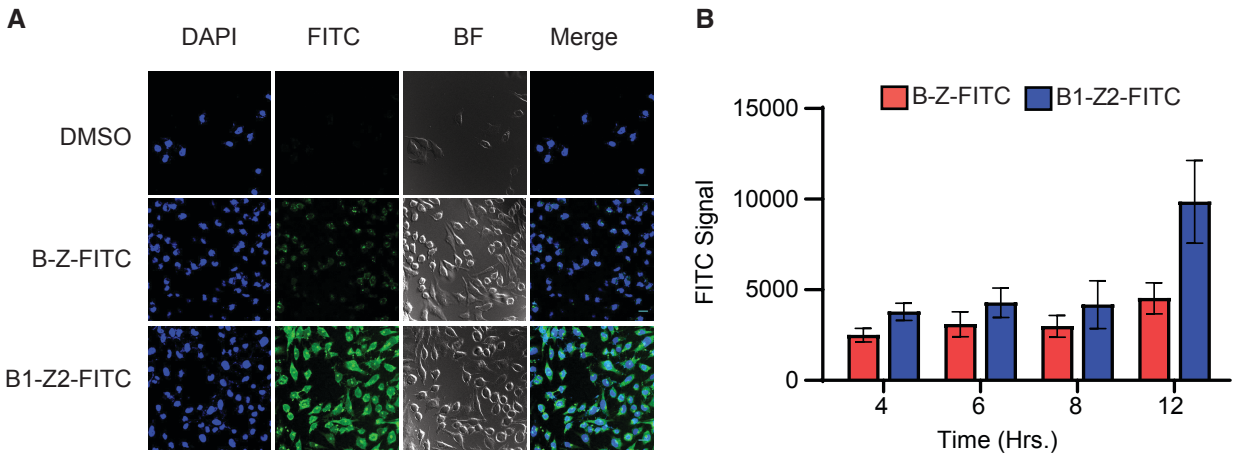


Figure 4.9 Fluorescent microscopy of STFs
(**A**) Representative images from confocal fluorescence microscopy of HeLa cells treated with B-Z-FITC, B1-Z2-FITC or DMSO for 12 hours. (**B**) Quantification of FITC signal for both compounds over time.

However, fluorescence imaging data alone cannot report on the discrete structure(s) present within cells. To directly determine whether intact B-Z-FITC and B1-Z2-FITC are present in cells—and at what relative concentrations—we extracted the intracellular contents of cells after treatment with each compound for visualization by gel electrophoresis. Full length B1-Z2-FITC was observed in cells after 24 hours of treatment, likely at concentrations in the 100–500 nM range based on comparison to a standard curve of purified B-Z-FITC (Figure 4.10). We observed two bands for B-Z-FITC at the same timepoint, suggesting that an appreciable fraction of the molecule is being proteolyzed in cells, and that full-length B-Z-FITC is present at much lower concentrations. This pattern was more pronounced at 48 hours, when full-length B1-Z2-FITC, but not B-Z-FITC, was still readily observed in cells.

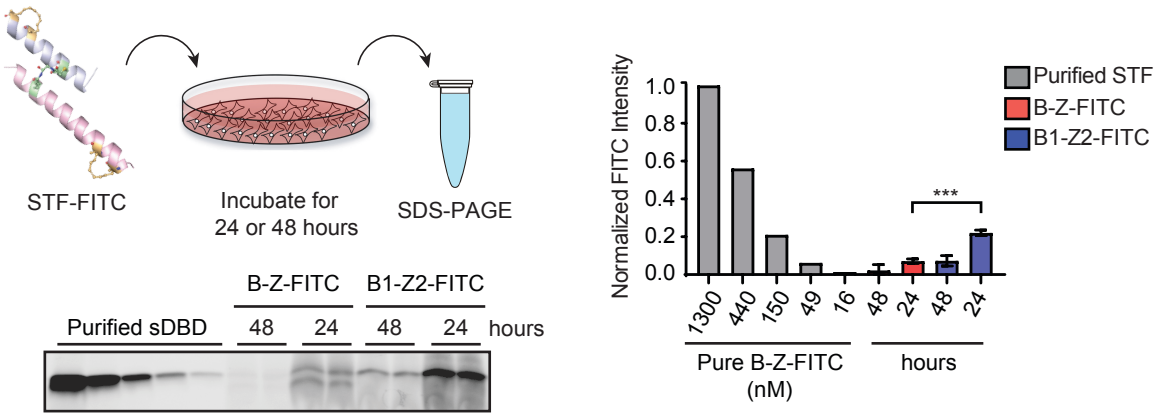


Figure 4.10 Cellular penetration and *in situ* stability of B-Z-FITC and B1-Z2-FITC measured by fluorescence gel analysis

Whole cell lysate from FITC-STF treated U2OS cells was resolved on an SDS-PAGE gel and imaged at 495 nm (FITC channel; technical duplicates are shown for each). A titration of purified B-Z-FITC is included for comparison. Graph on the right depicts relative fluorescence intensity of the intact FITC-STF band matched to concentrations in the pure dilution series.

Combined, this battery of stability and cellular uptake assays confirm that stabilized STFs like B1-Z2 can access cytosolic and nuclear compartments largely intact and should therefore be capable of binding target E-box DNA sequences in live cells. In Chapter 2, we have validated the enrichment at several Myc target loci by performing ChIP in HeLa cells. To directly assess whether B1-Z2 engages these target DNA sequences, we synthesized a biotinylated version of B1-Z2 (Bio-B1-Z2) and performed a direct Bio-ChIP using a modified streptavidin-biotin ChIP protocol[9]. We treated HeLa cells with 10 μM Bio-B1-Z2 for 12 hours and subsequently processed them for ChIP-qPCR using streptavidin enrichment. Relative to established control loci, including *SNTG2* and control intergenic region, Bio-B1-Z2 significantly enriched all Myc-bound target genes, with *RPL37* being the most robust for both endogenous Myc protein and Bio-B1-Z2 (Figure 4.11). Collectively, these data confirm that B1-Z2 can access intracellular compartments intact and specifically engage target E-box sites in the genome.

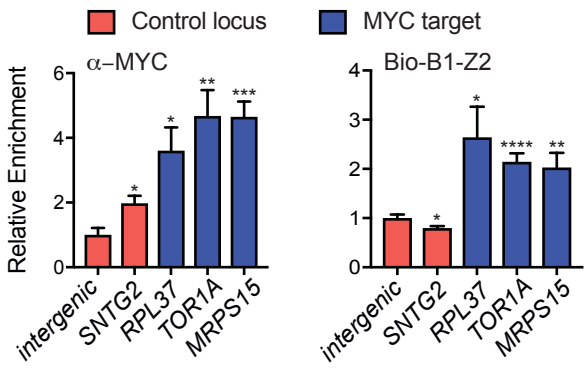


Figure 4.11 ChIP-qPCR quantification of endogenous Myc and Bio-B1-Z2 occupancy at control and E-box-containing target genes in HeLa cells

Enrichment values are calculated on a per-locus basis relative to mock enrichment from DMSO-treated cells.

4.2.5 Lead STFs suppress Myc-dependent cell proliferation

We performed phenotype screenings in order to evaluate STFs' effects on downstream cellular functions. In particular, we expect lead compounds, such as B11-Z6 and B11-Z8, to impact Myc-dependent transcriptome, proteome and phenotypic responses. To ascertain whether STFs have meaningful impact on Myc-dependent cellular functions, we elected the P493-6 cell line, which carries a Tet-off conditional expression of Myc that has been widely used for Myc-dependent cellular assays[10, 11]. We have successfully validated the system by activating supraphysiologic level of Myc expression, which led to cell proliferation (Figure 4.12A-B). Notably, both B11-Z6 and B11-Z8 treatment of "Myc-on" P493-6 cells phenocopies the effect of tetracycline-induced Myc blockade. By contrast, we do not observe impaired cell viability when these cells are in a "Myc-off" state, confirming a Myc-dependent effect of both STFs (Figure 4.12C).

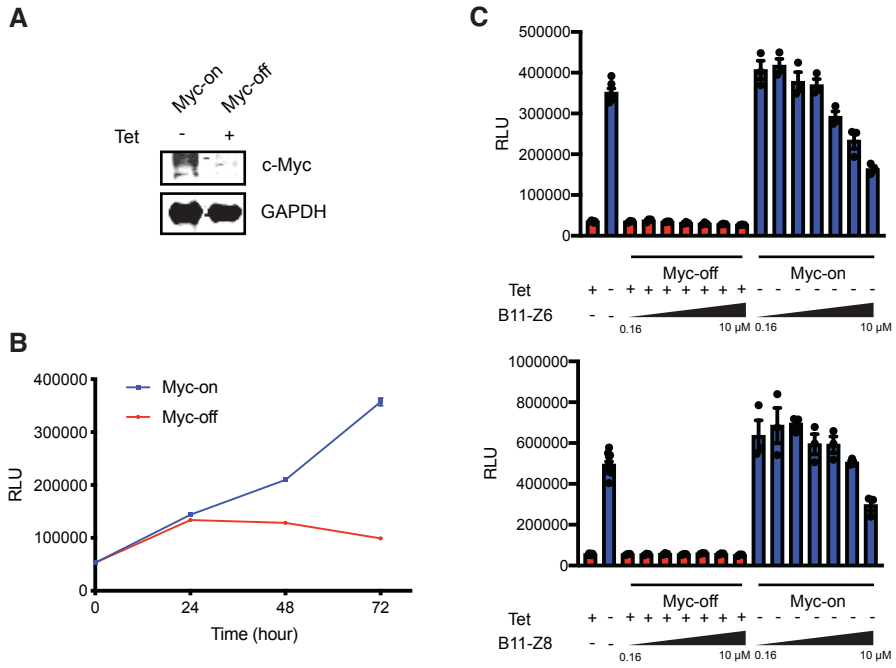


Figure 4.12 Viability effects of lead STFs on P493-6 cells under different levels of Myc expression
(**A**) Western blot analysis of c-Myc protein level and (**B**) cell proliferation curves of P493-6 cells under Myc-on and Myc-off states. (**C**) P493-6 cell viability after 72 hours of dose-dependent treatment with B11-Z6 and B11-Z8 under Myc-on and Myc-off states.

4.2.6 X-ray crystal structure confirms STFs mimic full-length transcription factors and specifically bind the DNA major groove

Thus far, the biophysical and biochemical structure activity relationships elucidated here strongly suggest that STFs assemble and recognize E-box DNA in a manner similar to full length bHLH TFs like Myc/Max or Max/Max. To directly confirm this hypothesis, we crystallized the B-Z STF with a 16-mer oligonucleotide containing a central 5'-CACGTG-3' sequence and single base pair overhangs on the 5'- and 3'- termini. A screen of crystallization conditions yielded reproducible rod-like crystals. Unlike the unsuccessful attempts for sDBDs, the crystal structure was solved by X-ray diffraction followed by molecular replacement using a published structure of the Max/Max ternary complex with E-box DNA (PDB ID: 1HLO)[2]. The asymmetric unit cell consisted of four B-Z dimers, each bound to a single duplex DNA, with crystal contacts observed

between inter-unit tetrahelix bundles and single base pair overhangs of adjacent oligonucleotide duplexes. As expected, the zipper and basic sequences in each B-Z STF monomer were completely α-helical and connected by a well-ordered glycyl-maleimide-cysteine adduct on the "back face" of each helix (Figure 4.13A). The interhelix crosslink was closely packed by surrounding residues on each helix, effectively locking the B and Z helices into a defined register relative to one another. In support of our non-natural cross-dimer design, each B-Z monomer forms a "sandwich-like" homodimer to create the tetrahelix bundle and orient the basic helices for sequence specific DNA binding (Figure 4.13B). The interface formed between the pseudo-symmetric homodimer buries approximately 1590 $Å^2$, and is mediated by extensive contacts between residues in both the B and Z helices. Anchoring this interface is an extensive hydrophobic core in the tetrahelix interior formed by *b*Ile39, *b*Lys40, *b*Phe43, *b*Leu46, *b*Arg47, *b*Val50, *b*Pro51, *z*Arg60, *z*Ile63, *z*Leu64, *z*Ala67, *z*Thr68, *z*Tyr70, *z*Ile71, *z*Nle74 and *z*Arg75, where *b* and *z* refer to the basic and zipper helix numbering from the parent Max protein (Figure 4.13C). Supporting this core is an additional layer of solvent exposed hydrophobic and polar contacts that contribute to the intermolecular tertiary and quaternary structure, including close packing between *z*Tyr70 and *z*Nle74 with *b*Val50 and *b*Pro51.

Like full length Myc and Max, the dimerized B-Z complex binds the E-box target DNA with each monomer interacting with half of the 5'-CACGTG-3' recognition sequence. Each basic helix makes numerous contacts to the phosphodiester backbone of DNA, as well as four sequence-specific contacts deep in the major groove. Backbone contacts are made by residues throughout the entire basic helix and encompass a 12 nucleotide span surrounding the E-box. These contacts include a *b*His27-PO_4 contact 3 nucleotides outside of the E-box, and *b*Arg25, *b*Asn29, *b*Arg33,

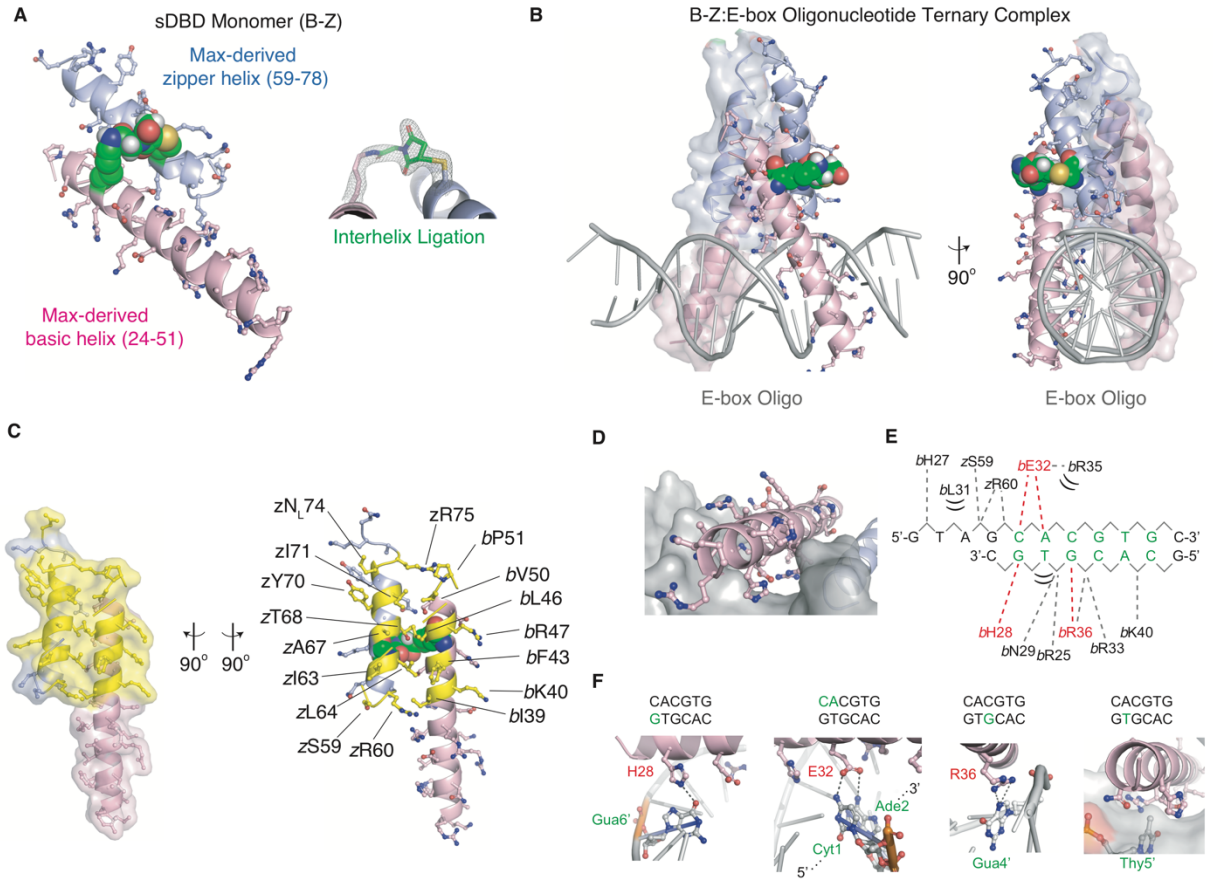


Figure 4.13 X-ray crystal structure of an STF:E-box DNA ternary complex

(**A**) Structure of B-Z monomer (left) with the basic and zipper helices colored in pink and blue, respectively. The glycylmaleimide interhelix linker is shown as green spheres (left) and as 2mFo-DFc difference map contoured to 1.5s (right). (**B**) Structure of the B-Z dimer bound to duplex DNA containing a central 5'-CACGTG-3' consensus sequence. One monomer is shown as helix cartoon, with the second monomer shown as a transparent surface. Overall ternary complex formation is highly similar to full length Max/Max and Myc/Max bound to DNA. (**C**) The B-Z dimeric interface involves extensive hydrophobic and polar interactions, highlighted in yellow and labeled, to form the intermolecular tetrahelix core. "Opened" view of the individual monomers reveals a sandwich like dimerization process to juxtapose the two DNA-binding helices. (**D**) B-Z DNA-binding helices interrogate the E-box major groove and make defined contacts to the backbone and E-box core. (**E**) Schematic of contacts between an individual B-Z monomer and the E-box containing oligonucleotide. Red and grey dashes denote polar sequence-specific and backbone contacts, respectively; double wedges denote Van der Waals interactions. (**F**) Individual sequence-specific B-Z:E-box contacts are shown in insets, with the relevant target nucleotides highlighted in the corresponding consensus sequence.

*b*Arg36, *b*Lys40, *z*Ser59 and *z*Arg60 all making contacts to phosphates within the core E-box sequence (Figure 4.13D-E). Each monomer makes hydrogen bond-mediated, sequence-specific contacts with both strands of the 5'-CAC-3' half-site (Figure 4.13F). The "antisense" contacts include *b*His28 and N7/C6 carbonyl of Guanosine-6 and *b*Arg36 with N7/C6 carbonyl of

Guanosine-4'. On the "sense" strand, *b*Glu32 makes close contact with N6 of Adenosine-2, N4 of Cytosine-3 and potentially N7 of the 5'-Guanosine outside of the E-box in this sequence (Figure 4.12F). Superimposing the B-Z and Max/Max structures reveals a striking congruence between the DNA binding residues, with an overall RMSD of 0.847 Å for the backbone of the entire DNA binding domain held in common (2.3 Å for entire bHLH structure). The interface between DNA binding domain of B-Z (1781 $Å^2$) is also comparable to that of Max homodimer (1726 $Å^2$).

Taken together, these data confirm that the hybrid STF architecture mimics the overall structure and sequence-specific DNA binding function of a full length transcription factor. Additionally, the crystal structure of STF-DNA complex also reveals, in retrospect, key drawbacks in our sDBD designs. Contrary to our original hypothesis, the Z helix, which is missing in the sDBD designs, plays an important role in DNA recognition and complex stabilization. In particular, the polar and hydrophobic residues on the Z helix, as shown in Figure 4.13C, help the formation of the "sandwich-like" homodimer and proper orientation of the B helix, which is crucial for sequence-specific binding. Additionally, several Z helix residues, such as *z*Ser59 and *z*Arg60, make direct contact with the DNA backbone, and thus stabilize the structure of the complex. The structural differences between the two designs explain the superior activities that we have observed in STFs.

4.3 Discussion

Here we developed a novel class of STFs derived from the bHLH protein Max. These non-natural, "cross-dimer" STFs recapitulate the cooperative association of individual DNA binding helices, as found in hundreds of known transcription factors. However, their hybrid architecture suggests they should not interfere with protein-protein interactions between endogenous bHLH

proteins. An extensive *in vitro* screening also confirmed that STFs have better DNA binding affinity and sequence specificity in a wider dynamic range than sDBDs, which highlights the importance of the leucine zipper helix for high affinity and specific DNA recognition. Notably, we observed comparable specificity, but not increased affinity, of the most active stapled hybrids (e.g., B1-Z2, B6-Z4, and B11-Z6) relative to the non-stapled progenitor B-Z. Combined with the fact that the stapled versions are considerably more helical and thermally stable than B-Z, these data reinforce published models of disordered-to-ordered search and binding by the basic helices as being integral to forming tight, specific complexes with DNA[12]. Preorganization of the individual helices may result in a thermodynamic trade-off between avoiding the energetic cost of folding to bind the major groove versus the restrictions on subtle conformational adjustments for optimal binding, which may be more accessible to non-stapled versions[13]. Most importantly, STFs based on the minimal B-Z core were incapable of competing with Myc/Max or Max/Max for E-box DNA binding. Optimized derivatives of this scaffold, including B11-Z6 and B11-Z8, were potent competitors with the full-length bHLH domain of both Myc/Max and Max/Max. This finding suggests that kinetic factors may be important determinants of effective competition, as B11-Z6 and B-Z have identical K_D values for E-box binding, but only the former competed with Max and formed a stable complex with DNA. Future efforts to understand these contributions and further optimize the dimer and DNA-binding interfaces, which were not a specific focus here, are warranted. To our knowledge, these are the most potent Myc and/or Max inhibitors reported to date, and the only molecules with unequivocal mechanism of action in targeting bHLH TF-DNA binding.

In addition to improved biochemical properties, we found that activity-guided stabilization of tertiary and secondary elements led to significant increases in structural and proteolytic stability

of several STFs. Targeted modification at and around sites of proteolysis within these molecules correlated with protection from proteolysis, however these changes had to be balanced with preservation of DNA binding. The net result of this enhanced stability for optimized STFs such as B11-Z6, at least as explored here, was enhanced protease resistance, cellular penetration and DNA binding activity and activity in cellular models. These data also suggest that DNA binding peptide domains comprised of canonical amino acids alone, which we confirmed here are largely unstructured and highly sensitive to proteases, are likely to suffer significant degradation inside and outside of cells. More generally, our data support the notion that modular synthesis of secondary and tertiary domain epitopes can be used to generate pharmacologically active mimetics, such as those targeting DNA in this study, and likely other proteins and biomolecules in the future.

What was most striking about the application of activity-guided stabilization of tertiary and secondary elements were the drastic increases in structural and proteolytic stability. CD spectroscopy, protease sensitivity, and cellular assays all confirmed the increased stability of STFs containing helix staples at optimized positions. These data also suggest that protein mimetics containing DNA binding peptide domains comprised of canonical amino acids alone, which we confirmed here are largely unstructured and sensitive to proteases, are likely to suffer significant degradation inside and outside of cells. More generally, our data support the notion that modular synthesis of secondary and tertiary domain epitopes can be used to generate pharmacologically active mimetics, such as those targeting DNA in this study, and likely other proteins and biomolecules in the future.

This iterative optimization approach enabled us to identify lead STF candidates with desirable biophysical and biochemical properties for potential downstream applications (Figure

4.14). Original designs such as B-Z and B1-Z2, showed promising E-box affinity but unsatisfactory results in terms of Myc/Max competition and proteolytic stability. The activity-guided optimization process significantly increased the stability (B11-Z6) or biochemical potency (B11-Z8) in later candidates without compromising other properties. The structure-activity relationship identified from this process can be further leveraged to optimize and search more well-rounded STFs in future iterations.

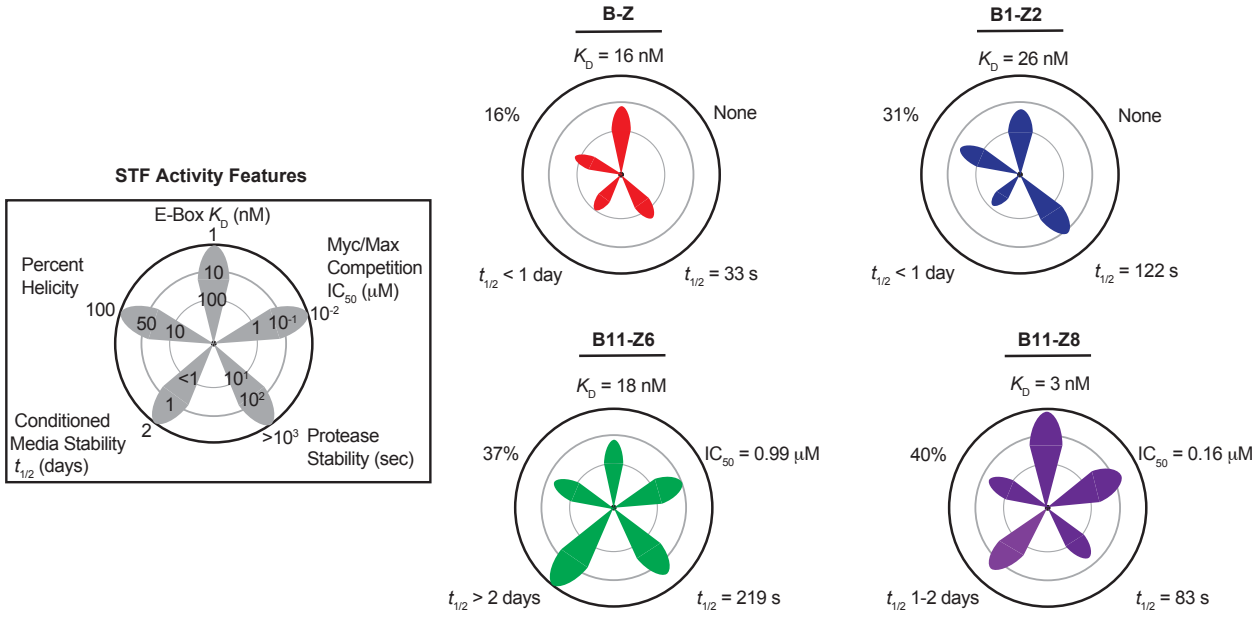


Figure 4.14 Petal charts for multiple activity features of representative STFs

Last but not least, preliminary cell-based assay results suggest that lead STFs, such as B11-Z6 and B11-Z8, lead to Myc-dependent phenotypic response in cancer cells, possibly by antagonizing Myc/Max in the cellular context. Future investigations are necessary to determine the role of STFs in the Myc signaling pathway in cancer cells and tumors. An extensive pipeline of cellular assays to probe the effect of STFs on the genome, transcriptome, proteome, and phenotypes is currently under way.

4.4 Materials and methods

Cell culture. HeLa and U2OS cells were purchased from ATCC. P493-6 cells were kindly provided by Dr. Sarki Abdulkadir from Northwestern University. HeLa and P493-6 cells were cultured in RPMI-1640 with 10% FBS and 1% penicillin/streptomycin. U2OS cells were cultured in DMEM with 10% FBS and 1% penicillin/streptomycin. All cell culture was performed under 37°C with 5% CO_2.

STF synthesis and purification. A Symphony X automated peptide synthesizer was used to prepare linear peptides on Rink amide MBHA resin. Fmoc-based solid phase chemistry, ring closing metathesis, and N-terminal modifications were carried out as previously described[14, 15]. Lysine residues bearing monomethoxy trityl (Mmt) side chain protecting groups were incorporated at cross-linking positions of basic helices. On-resin Mmt deprotection was carried out for 5 × 2 min consecutive cycles of 1% TFA/DCM solution mixed by N_2 bubbling. Deprotected lysine residues were functionalized with maleimide by 2 hours. treatment with a 0.1 M solution of 2-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)acetic acid (Mal-Gly-OH) (5 eq), HCTU (4.8 eq), and DIPEA (10 eq.) in DMF. Crude peptides cleaved from resin were purified on a Waters preparatory HPLC system using an Xbridge Prep C18 5μ OBN (19.5 × 150 mm) column; solvent A (0.1% TFA in H_2O); solvent B (MeOH); and a 10-min method with the following gradient (flowrate 30 mL/min): 35% B over 1 min; 35-85% B over 7 min; 95% B over 1 min; 35% B over 1 min. STF monomer ligation was performed in 50 mM sodium phosphate buffer pH 7.2 + 25% ACN as follows: a purified basic sequence bearing a maleimide (0.5 mL, 0.5 mM) and a purified zipper sequence with a free thiol (0.5 mL, 0.5 mM) were combined in a microcentrifuge tube and mixed by rotation for 2 hours at room temperature. The reaction mixture was diluted into 3 mL of 50% ACN/H_2O + 0.1% TFA and the ligated STF was purified using the same HPLC method as for individual

monomers. STF purity was confirmed by LC-MS using an Agilent system equipped with a Phenomonex C18, 5 μm (5.0 x 50 mm) column; solvent A (95:5:0.1 H_2O/ACN/TFA) and solvent B (95:5:0.1 ACN/H_2O/TFA); 0.5 ml min^{-1} flowrate, 0-2 min (0% B), 2-16 min (0-75% B), 16.5-18.5 min (100% B), 19 min (0% B). STF concentrations were quantified using 280 nm absorbance readings and compounds were stored as lyophilized powder or DMSO stocks.

Electrophoretic mobility shift assay (EMSA): For direct DNA binding experiments, STF were serial diluted (3-fold increments) at 2x concentration in 20 μL of 1X binding buffer (20 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL BSA, 1 mM DTT, 0.05% NP-40). 20 μL of 10 nM IRD700-labled E-box probe in 1x binding buffer was added and samples were incubated for 30 min at RT followed by 15 min at 4°C. 3.5 μL of each reaction was loaded on an 8% acrylamide, 0.5X TBE gel equilibrated to 4°C. Electrophoresis was carried out for 60 min at 110 V and 4°C with 0.5x TBE + 1 mM $MgCl_2$ running buffer. Gels were pre-run at 110 V for 60 min prior to sample loading. For competition experiments, 40 nM STF and 5 nM labeled E-box probe were incubated with 0, 5, 25, 125, or 625 nM unlabeled competitor oligo for 30 min at RT followed by 15 min at 4°C.

Gels were imaged using an Odyssey Li-COR. ImageJ was used to quantify band intensity and fraction bound DNA was calculated by dividing band intensity of bound DNA by the band intensity of the free DNA from a vehicle treated lane. A four-parameter dose-response curve fit to a plot of normalized fraction bound DNA vs. log B-Z concentration yields an IC50 which was reported as the apparent K_D. Mobility shift data were excluded from analysis when higher order binding species were observed.

Protein expression and purification. Human Myc (residues 355–434) and Max (residues 22–102) proteins were expressed with a N-terminal hexahistidine tag in *Escherichia coli* strain

BL21(DE3) using a pET28c vector. Transformed bacteria was grown at 37°C and induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) at an A_{600} = 0.8. Cells were pelleted 14 hours after induction and lysed in lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl, 8 M urea, 10 mM imidazole, pH 8.0) with cOmplete EDTA-free Protease Inhibitor (Roche) by sonication. The lysate was centrifuged to clear insoluble matter before loading onto Ni-NTA resin (Qiagen). After being washed once with lysis buffer and three times with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, 20 mM imidazole, pH 8.0), column-bound protein was eluted using elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, 250 mM imidazole, pH 8.0) and quantified using Bradford assay. The eluted Max protein was then dialyzed into desired buffers suitable for downstream applications, and further concentrated by centrifugation using a 10-kDa exclusion filter.

Circular dichroism spectroscopy. Lyophilized STF samples were resuspended in 20 mM phosphate buffer pH 7.4 and diluted to 20 μM. Circular dichroism spectra were obtained on a Jasco J-170 using a 0.1 cm quartz cuvette with the following settings: wavelength, 260-180 nm; data pitch, 1.0 nm; scan rate, 50 nm min^{-1}; accumulations, 3; temperature, 25-85°C with 6°C increments. Means-Movement smoothing at the lowest setting was applied to the recorded data.

***In vitro* trypsin stability assay.** Each STF, 30 μM, was dissolved in 330 μL of 20 mM phosphate buffer pH 7.4 in a microcentrifuge tube and heated to 37°C on a tabletop shaker (500 rpm). 30 μL of the reaction mixture was added to 60 μL of quenching solution (ACN + 3% formic acid with 500 nM Fmoc-lysine internal standard) for 0 s timepoint sample. Thermo-Pierce MS-Grade Trypsin was added to a final concentration of 0.5 μg/ml and additional 30 μL aliquots were quenched at indicated timepoints. Quenched samples were cooled to 4°C and centrifuged at 20,000 × *g* for 3 min. Sample injections of 35 μL were analyzed by LC-MS using an Agilent system

equipped with a Phenomonex C18 5µ (5.0 × 50 mm) column; solvent A (95:5:0.1 H_2O/ACN/TFA) and solvent B (95:5:0.1 ACN/H_2O/TFA); 0.5 ml min^{-1} flowrate, 0-2 min (5% B), 2-8.8 min (5-95% B), 9-11 min (95% B), 11.1 min (5% B). Intact STF was calculated by normalizing the background subtracted integrated area under the curve for the EIC of $(M + 4H^+)/4$, ± 0.5 mass units to the integrated area under the curve for the A_{280} peak of the internal standard. Fraction intact STF was calculated by dividing the intact STF by the normalized STF signal in the initial 0 s sample. GraphPad Prism was used to plot fraction intact STF vs time, and the proteolytic half-life was derived using a non-linear one-phase decay with the plateau constant set equal to 0.

Conditioned media binding assay. Approximately 6×10^4 HeLa cells were seeded into a 24-well plate and grown overnight. The media was aspirated, the cells were washed with 0.5 mL PBS and the treated with 10 µM STF in 0.5 mL fresh media. At 0, 24, 48 and 72 hours, treatment media (2 µL) was diluted into 38 µL of 1x EMSA binding buffer containing 5 nM E-box probe. DNA binding was measured using an electrophoretic mobility shift assay. Fraction bound to E-box probe was calculated by dividing the band intensity of bound DNA by the sum of the bound DNA + free DNA.

LDH membrane leakage assay. Approximately 5,000 HeLa cells were seeded in a 96-well plate. The following day an equal volume of media containing 2x compound or DMSO vehicle were added to experimental wells and the plate was incubated for the indicated time of experiment. A 2x volume of lysis buffer was added to additional wells 45 minutes prior to the final time-point and LDH activity in treatment medium was measured using the Pierce LDH Cytotoxicity Assay Kit according to manufacturer protocol (Thermo Scientific, #88953).

Fluorescence microscopy and quantitative analysis. HeLa cells were seeded in the 12-well chamber slide with 2500 cells per well (Ibidi, #81201). Cells at 80-90% confluency were

treated either with DMSO as the negative control or 10 μM FITC-labeled STF for indicated duration. Cells were washed with phosphate buffer saline (PBS) for three consecutive times, fixed by 4% formaldehyde in PBS at room temperature for 10 mins, and then washed twice with PBS. Cells were then stained by diluted DAPI (Thermo, #D1306) in PBS at room temperature for 3 mins, and washed twice with PBS. Rubber frame was then removed, and slide was dried in the dark at room temperature, covered with cover glass (Fisher, #12–545 M), mounted with 50 μL anti-fade mounting solution (Life technology, #P36961), and sealed with nail polish. Leica SP8 Laser Scanning Confocal was used to image a single focal plane to accurately detect the DAPI and FITC signal location using HyD detectors. Identical microscope acquisition parameters were set and used within experiments. Post-acquisition processing was performed using ImageJ software (NIH). Lossless TIFF files were employed to quantify fluorescence intensity. To simplify the image processing workflow, a Macro script to automatically process all images was created. The workflow was as follows: open all channels for each field of view; designate a color for each channel; adjust brightness/contrast for all channels (applying the same levels for all conditions within and between experiments to allow for direct comparison); merge the channels together; adjust the image unit from pixel to micrometer; add scale bars; export the processed TIFF files for quantification. For quantitative analysis, single-cell boundaries were identified manually using the DIC image. Then the "ROI Manager" tool in ImageJ was exploited to add all the cell outlines as a collection and overlay with the DAPI and FITC channel to measure per-cell fluorescence intensity. Typical quantitative comparisons were made using data from three or more independent fields of view per independent biological replicate condition.

SDS-PAGE gel analysis of STF uptake. Approximately 1×10^5 U2OS cells were seeded in each well of a 12-well plate. Cells were treated for 24 and 48 hours. with 10 μM B-Z-FITC or

10 μM B1-Z2-FITC. After the indicated treatment time, media was aspirated, cells were washed with PBS (2 x 1 mL) and treated with 0.25% trypsin (0.2 mL) for 5 min at 37°C. The trypsin was quenched with the addition of 1 mL of media and the detached cells were transferred to a microcentrifuge tube and centrifuged at 500 × *g* for 4 min. The media was aspirated, 20 μL of RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 0.25% deoxycholate, 1% NP-40, 1 mM EDTA) was added and cells were incubated in RIPA buffer for 10 min on ice. After lysis, 6.6 μL of 4x SDS loading buffer was added, samples were heated to 95°C for 10 minutes, cooled to RT and analyzed by SDS-PAGE using a tris-glycine buffer system with a 15% acrylamide gel.

Chromatin immunoprecipitation (ChIP). Myc ChIP*:* HeLa cells were seeded in 200-mm dishes. After reaching 70% confluence, cells were crosslinked with 1% formaldehyde, fragmented by sonication, and incubated with c-Myc antibody (9E11, Abcam) overnight. The mixture was then immunoprecipitated with protein A beads (GenScript, pre-treated with 1% BSA for 1 hour) for 1 hour. Immunoprecipitated complexes were successively washed with Low Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl, pH 8.0), High Salt Wash Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl, pH 8.0), and LiCl Wash Buffer (250 mM LiCl, 1% NP-40, 1% Sodium Deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0). All washes were performed at room temperature for 8 min on a rotator. The complexes were eluted with 1% SDS at 30°C for 15 min, and then incubated at 65°C overnight to reverse crosslink protein-DNA complexes. After decrosslinking, DNA was purified using QIAQuick PCR Purification Kit (Qiagen) according to the manufacturer's instructions.

Bio-STF ChIP: HeLa cells were seeded in 200-mm dishes. After reaching 70% confluent, cells were treated with DMSO control or biotinylated STF in fresh growth media for the specified

time. Upon completion, cells were crosslinked with 1% formaldehyde, fragmented by sonication, and incubated with streptavidin beads (Invitrogen, pre-treated with 1% BSA for 1 hour) for 4 hours. Immunoprecipitated complexes were successively washed with Wash Buffer I (2% SDS), Wash Buffer II (50 mM HEPES pH 7.5, 500 mM NaCl, 0.1% Deoxycholate, 1% Triton X-100, 1 mM EDTA), Wash Buffer III (10 mM Tris-HCl pH 8.1, 250 mM LiCl, 0.5% NP-40, 0.5% Deoxycholate, 1 mM EDTA,) and TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). All washes were performed at room temperature for 8 min on a rotator. SDS Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) was added and incubated at 65°C overnight to reverse crosslink protein-DNA complexes. After decrosslinking, DNA was purified using QIAQuick PCR Purification Kit (Qiagen) according to the manufacturer's instructions.

qPCR: ChIP DNA from both experiments was quantified using quantitative PCR on a LightCycler 480 (Roche). The sequences of the qPCR primers are listed in the supplemental table. Fold of enrichment for each target gene was normalized to the intergenic region negative control.

Cell viability assay. Approximately 2,000 cells were seeded in 50 µL media in a 96-well plate and let settle overnight (not necessary for suspension cell lines). 25 µL media containing 3x compound or DMSO vehicle were added to experimental wells and the plate was incubated for the indicated time of experiment. Upon completion of treatment, 75 µL CellTiter-Glo reagent (Promega) was added to each well and the plate was gently shaken at room temperature for 10 minutes before luminescence signal was read on a plate reader.

X-Ray crystallography and structural refinement. Purified peptide B-Z was dissolved in 50 mM HEPES pH 6.0, 200 mM NaCl and 10 mM $MgCl_2$ to yield a 200 µM solution. To this solution, 16-mer oligonucleotides (2.5 mM) containing E-box site in duplex buffer (100 mM potassium acetate, 30 mM HEPES, pH 7.5, Integrated DNA technologies, Lot #11-05-01-12) were

added to make the final concentration of oligonucleotides 100 μM. Co-crystals were generated using hanging drop vapor diffusion where 1 μL of complex solution was mixed with 1 μL well solution. Clear rectangular crystals emerged in 50 mM Tris pH 7.0, 30% 2-Methyl-2,4-pentanediol, 50 mM NaCl and 10 mM $MgCl_2$. Diffraction data was collected at the Advanced Photon Source, Argonne National Laboratories, Argonne, Illinois, at the SBC 19-BM beamline (0.97 Å). Data were indexed, scaled and merged using HKL-300016[16]. Molecular replacement was performed in Phenix using PDB: 1HLO as the search model with ligand removed[17]. The model was refined using iterative rounds of Phenix Refine and manual inspection with Coot[18]. Ligand constraints for interhelix were generated using Elbow. Final structure will be deposited in the Protein Data Bank upon publication (Accession: 6XRD). All x-ray crystal structure images were generated using PyMol.

References

1. Atchley, W. R.; Wollenberg, K. R.; Fitch, W. M.; Terhalle, W.; Dress, A. W., Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Mol Biol Evol* **2000,** *17* (1), 164-78.

2. Brownlie, P.; Ceska, T.; Lamers, M.; Romier, C.; Stier, G.; Teo, H.; Suck, D., The crystal structure of an intact human Max-DNA complex: new insights into mechanisms of transcriptional control. *Structure* **1997,** *5* (4), 509-20.

3. Blackwell, T. K.; Kretzner, L.; Blackwood, E. M.; Eisenman, R. N.; Weintraub, H., Sequence-specific DNA binding by the c-Myc protein. *Science* **1990,** *250* (4984), 1149-51.

4. Blackwood, E. M.; Eisenman, R. N., Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science* **1991,** *251* (4998), 1211-7.

5. Farina, A.; Faiola, F.; Martinez, E., Reconstitution of an E box-binding Myc:Max complex with recombinant full-length proteins expressed in Escherichia coli. *Protein Expr Purif* **2004,** *34* (2), 215-22.

6. Talanian, R. V.; McKnight, C. J.; Rutkowski, R.; Kim, P. S., Minimum length of a sequence-specific DNA binding peptide. *Biochemistry* **1992,** *31* (30), 6871-5.

7. Baldwin, A. D.; Kiick, K. L., Tunable degradation of maleimide-thiol adducts in reducing environments. *Bioconjug Chem* **2011,** *22* (10), 1946-53.

8. Lyon, R. P.; Setter, J. R.; Bovee, T. D.; Doronina, S. O.; Hunter, J. H.; Anderson, M. E.; Balasubramanian, C. L.; Duniho, S. M.; Leiske, C. I.; Li, F.; Senter, P. D., Self-hydrolyzing maleimides improve the stability and pharmacological properties of antibody-drug conjugates. *Nat Biotechnol* **2014,** *32* (10), 1059-62.

9. Bi, M.; Zhang, Z.; Jiang, Y. Z.; Xue, P.; Wang, H.; Lai, Z.; Fu, X.; De Angelis, C.; Gong, Y.; Gao, Z.; Ruan, J.; Jin, V. X.; Marangoni, E.; Montaudon, E.; Glass, C. K.; Li, W.; Huang, T. H.; Shao, Z. M.; Schiff, R.; Chen, L.; Liu, Z., Enhancer reprogramming driven by high-order assemblies of transcription factors promotes phenotypic plasticity and breast cancer endocrine resistance. *Nat Cell Biol* **2020,** *22* (6), 701-715.

10. Schuhmacher, M.; Kohlhuber, F.; Holzel, M.; Kaiser, C.; Burtscher, H.; Jarsch, M.; Bornkamm, G. W.; Laux, G.; Polack, A.; Weidle, U. H.; Eick, D., The transcriptional program of a human B cell line in response to Myc. *Nucleic Acids Res* **2001,** *29* (2), 397-406.

11. Pajic, A.; Spitkovsky, D.; Christoph, B.; Kempkes, B.; Schuhmacher, M.; Staege, M. S.; Brielmeier, M.; Ellwart, J.; Kohlhuber, F.; Bornkamm, G. W.; Polack, A.; Eick, D., Cell cycle activation by c-myc in a burkitt lymphoma model cell line. *Int J Cancer* **2000,** *87* (6), 787-93.

12. Metallo, S. J.; Schepartz, A., Certain bZIP peptides bind DNA sequentially as monomers and dimerize on the DNA. *Nat Struct Biol* **1997,** *4* (2), 115-7.

13. Kalodimos, C. G.; Biris, N.; Bonvin, A. M.; Levandoski, M. M.; Guennuegues, M.; Boelens, R.; Kaptein, R., Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science* **2004,** *305* (5682), 386-9.

14. Kim, Y. W.; Grossmann, T. N.; Verdine, G. L., Synthesis of all-hydrocarbon stapled alpha-helical peptides by ring-closing olefin metathesis. *Nature protocols* **2011,** *6* (6), 761-71.

15. Mitra, S.; Montgomery, J. E.; Kolar, M. J.; Li, G.; Jeong, K. J.; Peng, B.; Verdine, G. L.; Mills, G. B.; Moellering, R. E., Stapled peptide inhibitors of RAB25 target context-specific phenotypes in cancer. *Nat Commun* **2017,** *8* (1), 660.

16. Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* **2006,** *62* (Pt 8), 859-66.

17. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L. W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 2), 213-21.

18. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **2010,** *66* (Pt 4), 486-501.

Chapter 5

Pushing the Limit of Peptide Synthesis: Dimeric Synthetic Transcription Factors

5.1 Chemical complications of building a tetrahelix scaffold

In Chapter 4, we have developed a library of STFs that feature a "cross dimer" conjugation of two individually stabilized B and Z helices. Activity-guided optimizations and extensive studies revealed numerous structure-activity relationships that shed light on the design of better, more functional miniprotein scaffolds. One critical aspect of STFs' activity is their ability to self-associate via a number of hydrophobic residues on both B and Z helices, which enables the formation of a dimeric, tetrahelix core that binds the major groove of the E-box consensus, in a similar fashion to bHLH TF. Given its indispensable role in the DNA recognition process, we aimed to create a conjugate of two STF molecules at the C-termini (dimeric STF or dSTF) and hypothesize that the pre-dimerization guarantees a close proximity between two STFs, which in theory would facilitate the DNA binding and further improve the affinity (Figure 5.1). Moreover, dSTF's preorganized dimeric structure would also lead to better folding of the scaffold and lower protease accessibility, and consequently higher target specificity and stability.

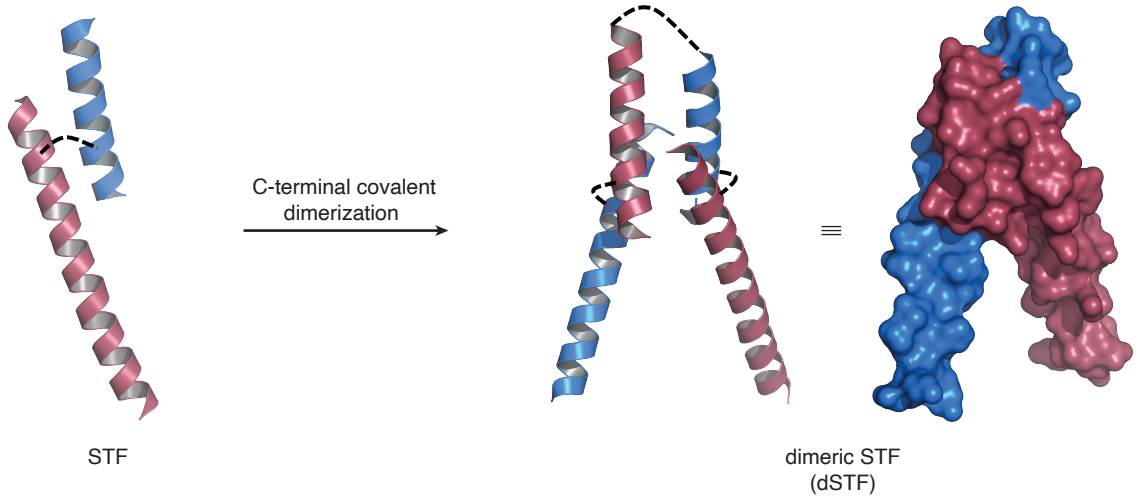


Figure 5.1 Proposed dSTF scaffold by covalently dimerizing two STF molecules at the C-termini

The synthesis of such dSTF scaffolds, however, is much more challenging than conventional peptide and protein synthesis. As we have learned from Chapter 3, novel miniprotein scaffolds, especially with added covalent conjugation and structural complexity, is usually accompanied by significant chemical complications. First, similar to our attempts to introduce a second covalent linker, the pre-dimerized tetrahelix scaffold requires at least three conjugation steps—one for each side of B-Z linkage, and one between the two zipper helices at the C-termini. Due to possible compatibility issues, the choice of chemistries and the order of ligations need to be carefully considered. Figure 5.2 lists two possible scenarios. Particularly, in the case where B-Z helices are identical on both sides (symmetric dSTFs), we can first create the C-terminal linker between two Z helices, followed by another conjugation with excess B helix. On the other hand, for a scaffold that consists of four completely distinct helices (asymmetric dSTFs), we need to conjugate the different B-Z pairs respectively before creating a linker at the C-termini.

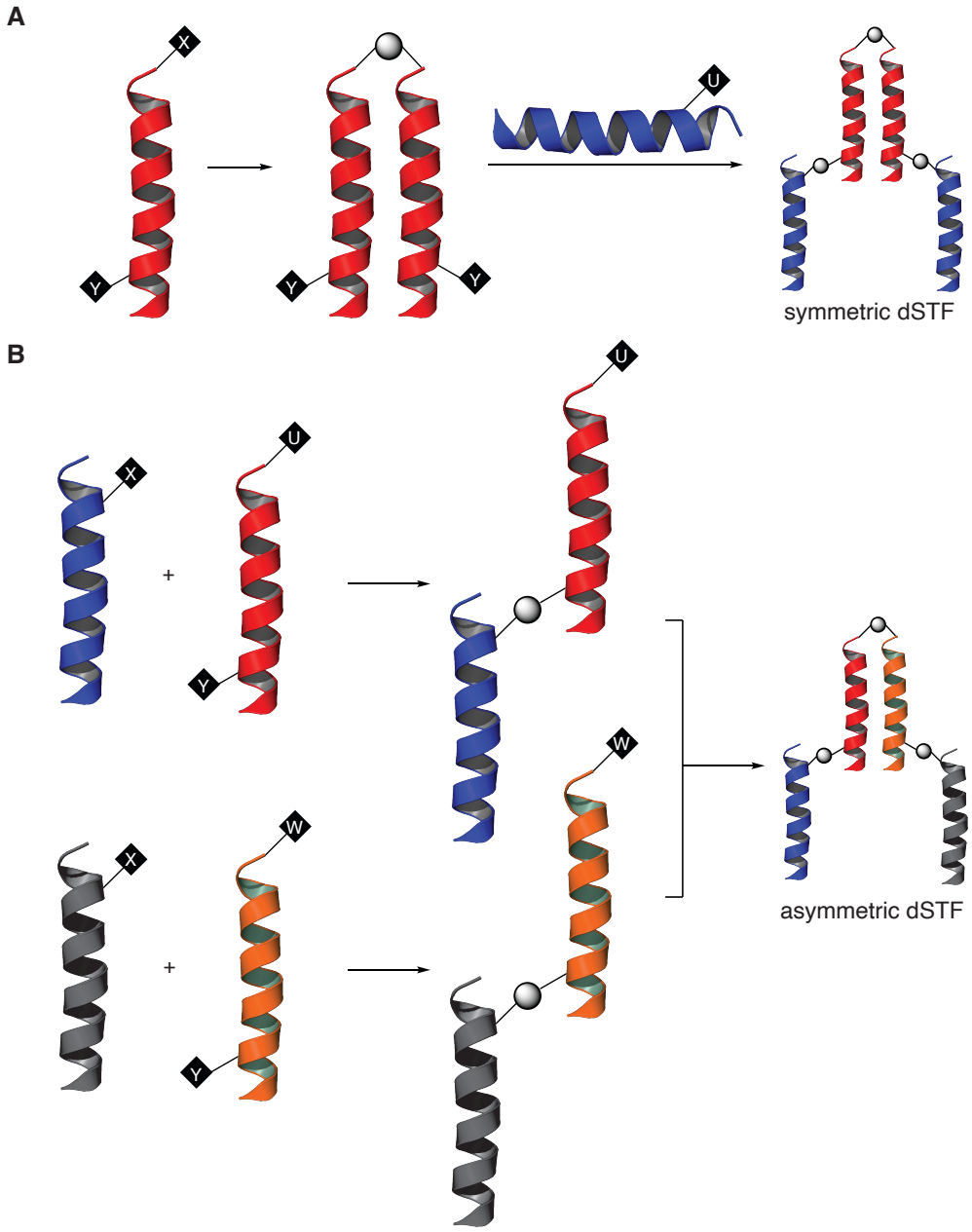


Figure 5.2 Different conjugation strategies for the synthesis of symmetric and asymmetric dSTFs

Second, even though we have so far established sufficient dimension of orthogonality in our synthetic platform, the size of the target poses significant challenges to our synthesis. dSTFs, even in their simplest form, are a 13-kDa, artificially designed miniprotein, a major test for the efficiency of peptide synthesis and the robustness of the orthogonal chemistries. Furthermore, the

overall yield may decrease with more linear steps. The solubility issue, as we have already observed in some STFs, may deteriorate due to increased folding. The separation and purification on C18-based reverse phase HPLC may also be problematic. Therefore, in efforts to create such a scaffold, we also need to strictly control reaction buffer conditions and sample handling and workup, which is often less of an issue for smaller peptides.

In the remainder of this chapter, we will primarily focus our discussion on the chemistries and synthetic strategies that we have employed in an attempt to prepare dSTFs, with some preliminary findings on the activity and stability of the resulting miniproteins.

5.2 Synthetic trials and findings

5.2.1 Symmetric dSTFs

To begin with, we aim to synthesize the symmetric dSTF scaffold that resembles the Max/Max homodimer. We opt for a two-step conjugation scheme for two main reasons. First, the overall yield will be higher with fewer conjugation steps. Additionally, the precursor peptides, each of which are approximately 4 kDa in size, are usually easier to handle compare to the alternative scheme with 6-kDa synthons.

Due to the symmetry of the zipper helix dimer, we chose a synthetic route toward a branched peptides featuring two identical Z helices. Due to the wide use of lysine in branched peptide synthesis[1-3], we first incorporated a C-terminal Mmt-lysine onto the resin. Subsequent on-resin deprotection of both N^{α}-Fmoc and N^{ε}-Mmt groups generated two free amino groups, which were then used as starting points for the solid phase synthesis of two identical zipper helices. Under regular SPPS coupling protocols, the synthesis was less efficient, resulting in <10% yield with too many side products that could not be resolved and purified by HPLC, mainly due to the steric

hindrance between the two branches in close proximity. However, the yield and efficiency significantly improved with a modified coupling protocol by increasing the temperature to 65 °C with IR heating. The resulted branched peptide with a free cysteine on each Z branch was then conjugated with two equivalent of maleimide-containing B helix to form the dSTF-B-Z scaffold (Figure 5.3).

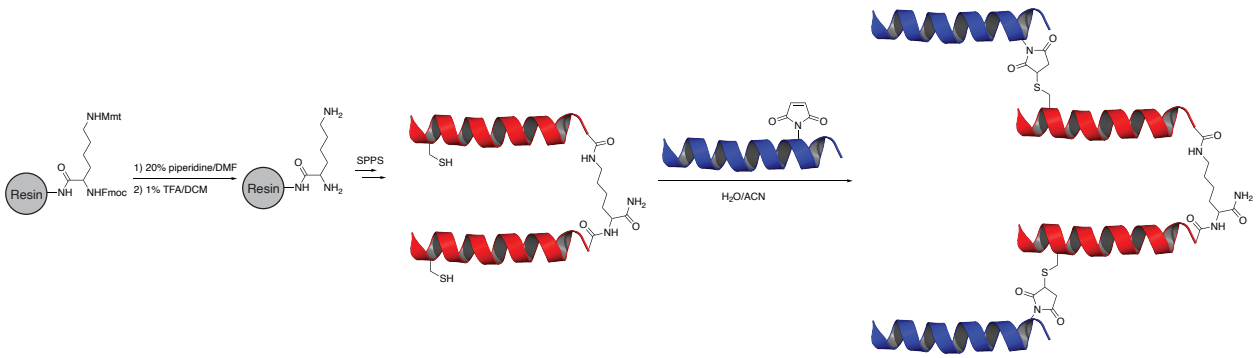


Figure 5.3 Synthetic scheme for dSTF-B-Z

In an attempt to extend the same synthetic scheme to dSTFs with stapled B and Z helices, such as dSTF-B1-Z2, we found unsatisfactory results even at higher coupling temperature. We believe that the added bulkiness from the bis-alkylated S_5 significantly reduces the efficiency and synthetic value of this branched synthesis route. Instead, we started with a single Z2 helix containing a C-terminal free cysteine and an internal S-(*tert*-butylthio)-cysteine and performed homodimerization at the C-terminal free cysteine in solution using a bis-maleimide linker. The disulfide bonds were then reduced to deprotect the internal cysteines, which then similarly underwent a conjugation with two equivalent of maleimide-containing B1 helix to form the stapled dSTF-B1-Z2 scaffold (Figure 5.4).

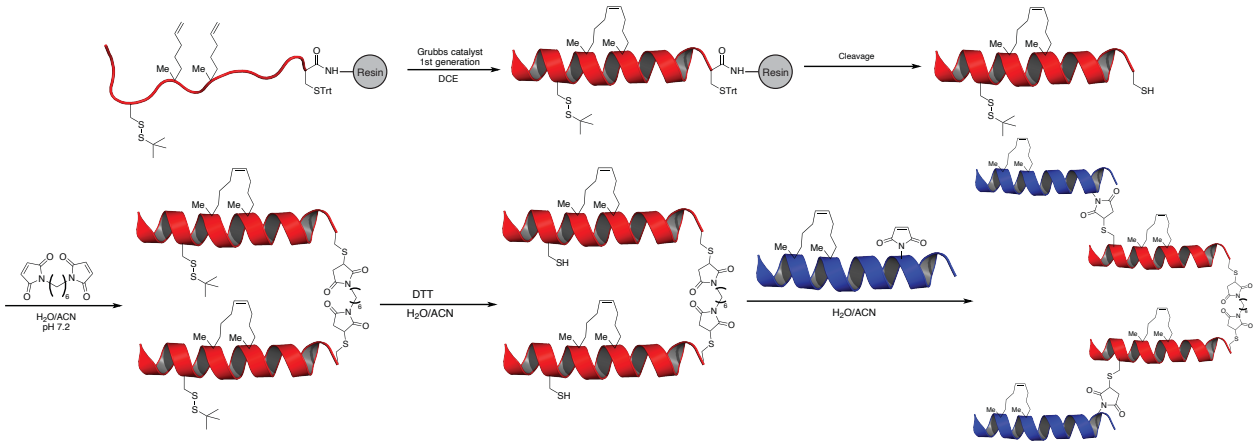
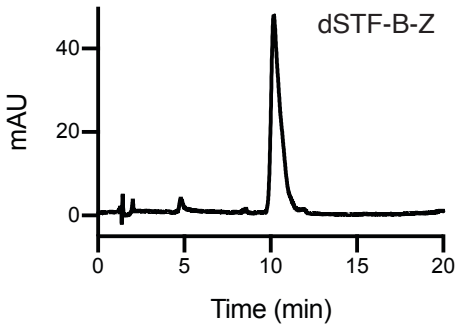


Figure 5.4 Synthetic scheme for dSTF-B1-Z2

In both cases, we were able to use reverse phase HPLC to separate the final dSTF product with reasonable purity, and ESI-MS confirmed their molecular mass of approximately 13 kDa (Figure 5.5), which were comparable to a small sized protein. Unlike recombinant proteins expressed from live organisms, our platform enables the *de novo* synthesis of chemically pure macromolecules with artificially designed structural components, a major breakthrough in miniprotein synthesis. Furthermore, with four stapled helices, deliberately oriented tertiary structure and a size of 13 kDa, dSTF represents the largest stabilized miniprotein scaffold that has ever been made to date, which further indicates that our platform is pushing the limit of peptide synthesis and expanding the scope of synthetic targets.



Exact mass: 12820.7

Z	(M+Z)/Z	Found
7	1832.5	1832.4
8	1603.6	1603.3
9	1425.5	1425.2
10	1283.1	1283.2
11	1166.5	1166.8
12	1069.4	1069.2
13	987.2	987.1
14	916.8	916.7
15	855.7	855.6

Figure 5.5 LC-MS characterization of dSTF-B-Z

5.2.2 Asymmetric dSTFs

Having solved the challenges for symmetric dSTFs, we aim to further explore the feasibility for asymmetric dSTF scaffolds derived from two different STF molecules. The heterodimeric structure makes them are an appealing synthetic target for various reasons.

First of all, unlike self-associating STFs, asymmetric dSTFs represent a preorganized heterodimeric structure with significant biological effects that differ from STFs. In Chapter 4, we have proposed a mechanism that some STFs preferentially target Max/Max homodimer, possibly because the homodimeric form better mimics the Max/Max structure. On the other hand, many other bHLH TFs, such as Myc/Max, are in a heterodimeric form. Synthetic assess to asymmetric dSTFs mimicking such structures thus will enable more specific and precise interrogation of these targets.

Moreover, the ability to "mix and match" different B and Z combinations on each side would allow us to create engineered scaffolds that bind different DNA sequences. For example, by selectively altering the DNA-recognizing sequences on one side, the resulting dSTF may bind non-palindromic DNA sequences[4]. It may also enable artificially designed combinations of TFs from different families[5], such as bHLH/zinc finger, bHLH/bZIP. Such combinations may possess intriguing properties that are not present in natural TFs.

With our existing orthogonal chemistries and lessons learned from earlier chapters, we proposed a synthetic scheme that utilizes two Z helices featuring a cysteine on the N-terminus and an alkyne or azide on the C-terminus. The bifunctional Z helices would be synthesized using the post-stapling modifications we have developed in Chapter 3. Once purified, each Z helix would be ligated with the corresponding B helix via thiol-maleimide conjugation. A final CuAAC conjugation at the C-termini would result in the desired asymmetric dSTF scaffold (Figure 5.6).

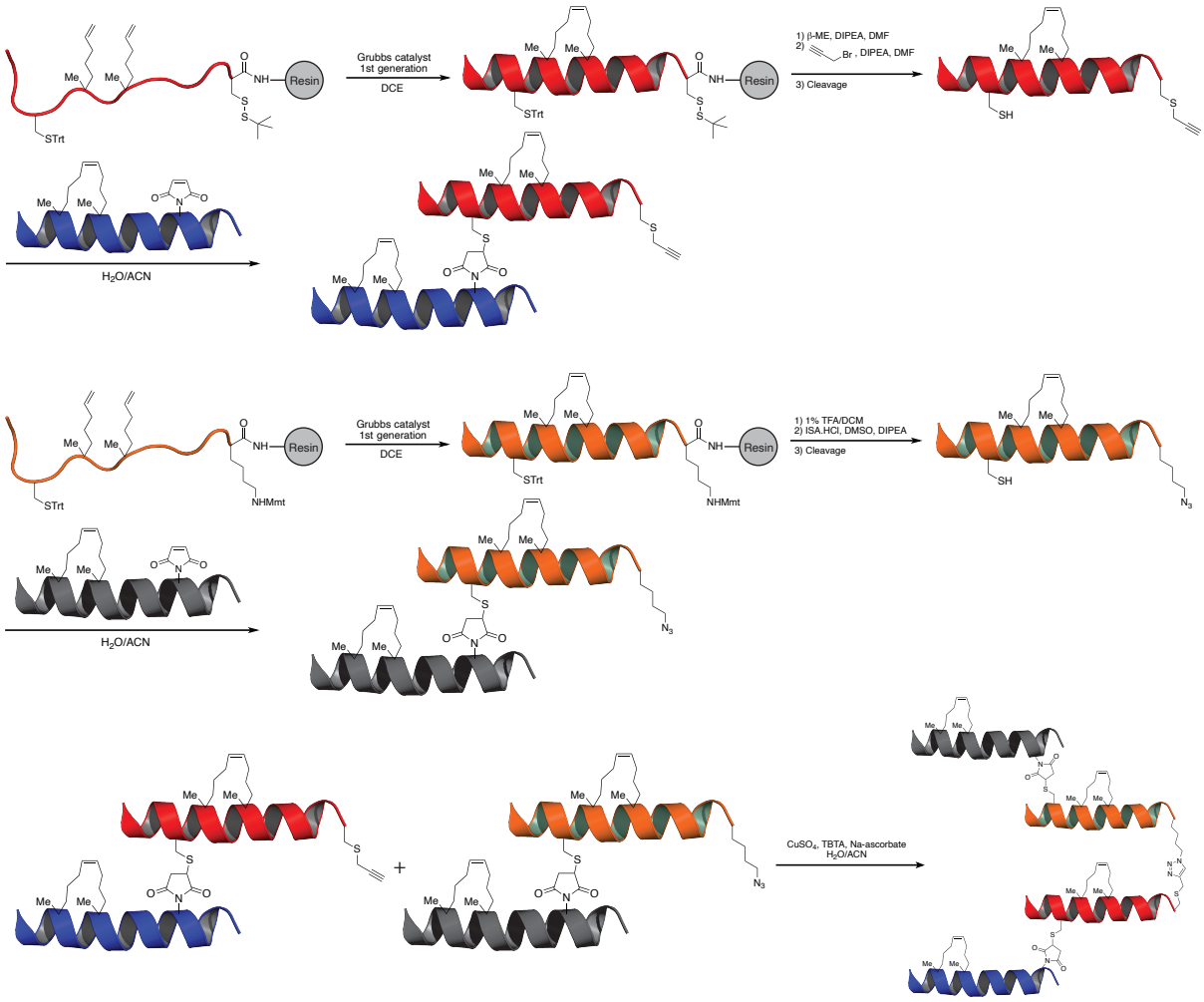


Figure 5.6 A proposed scheme for the synthesis of asymmetric dSTF scaffold

Under the proposed scheme, we were able to proceed to the penultimate step, generating alkyne- and azide-containing STF with high efficiency and purity. However, the final step was problematic. Similar to what we have observed for the synthesis of doubly linked sDBDs, the CuAAC conjugation did not work as expected. No conversions were observed after 6 hours with up to 2 equivalents of copper(I) catalyst, an amount which is sufficient for catalyzing the reaction between the shorter Z helices. We suspect that the lysine and arginine residues on the B helix interfere with the reaction by chelating the copper(I) catalyst. Further increasing the catalyst amount, on the other hand, resulted in precipitation of the peptides, possibly due to the formation

of insoluble complexes between excessive copper(I) ion and the polar residues, exposing the hydrophobic surface. Despite CuAAC's numerous applications in bioconjugation[6], such as in antibody-drug conjugates, our unsuccessful trials indicate that it is not as robust for our specific purpose of linking stabilized, amphipathic miniprotein scaffolds. Further condition optimization or different chemistries may be necessary for the synthesis of asymmetric dSTF scaffolds.

5.3 Pre-dimerization of STFs enhances activity and stability

With asymmetric dSTFs in hand, we proceeding to testing their DNA binding activity in EMSA. Remarkably, both the unstapled dSTF-B-Z and stapled dSTF-B1-Z2 showed potent DNA binding with an apparent K_D of 1–2 nM, a 10-fold increase compared to their monomeric counterparts (Figure 5.7). Given that the observed K_D is at the detection limit of the assay, the actual affinity of these dSTFs may be even higher. EMSA experiments with non-fluorescently labeled competitor oligonucleotides also revealed that the sequence specificity of both dSTFs, with the shifted fluorescent band being significantly competed by E-boxC1 but to a much less extent by either E-boxCM2 or E-boxCM3 (Figure 5.8).

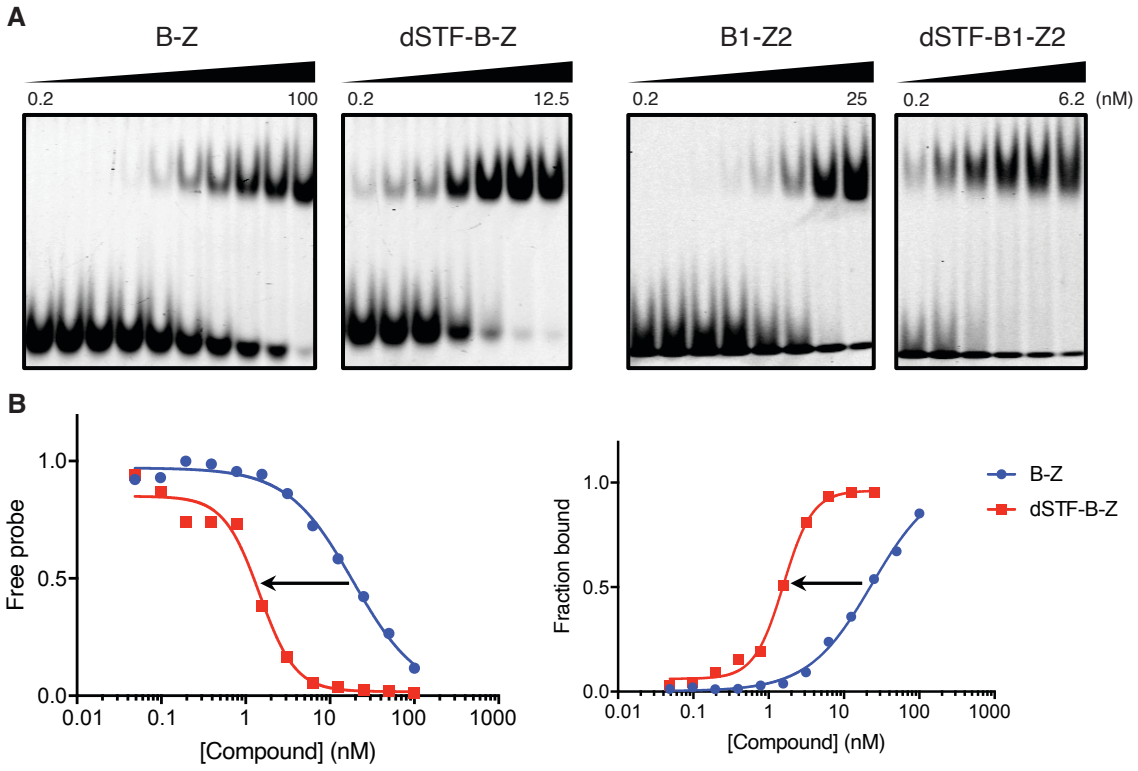


Figure 5.7 Direct comparison between STF and dSTF in EMSA
(**A**) EMSA gel images of both STFs and dSTFs binding to E-box DNA probe. (**B**) Quantification of EMSA results and comparison of potency between B-Z and dSTF-B-Z.

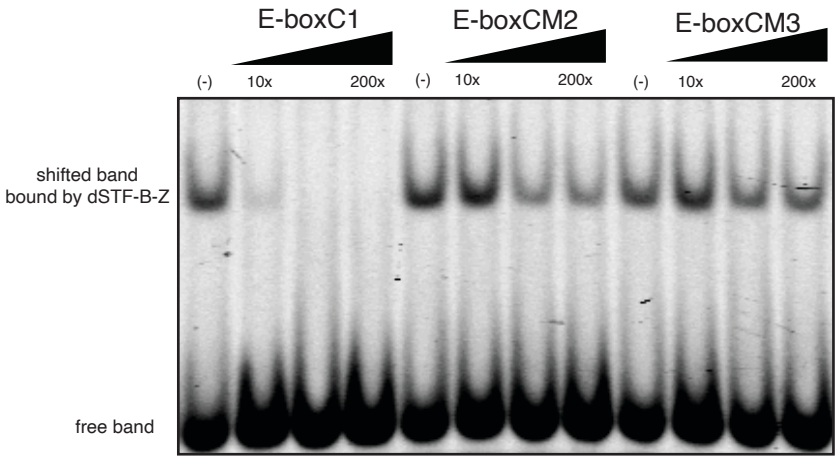


Figure 5.8 Competition EMSA gel image for dSTF-B-Z DNA binding specificity

The need for self-dimerization is critical to STF's activity and may compromise its capability of effectively competing with bHLH TFs. As we have discovered in the STF library,

some candidates with potent and specific DNA binding, such as B1-Z2, showed minimal competition with Max/Max homodimers. Given that the DNA binding of dSTFs no longer requires a self-associating, cooperative manner, we predict that it would lead to a significant stabilization effect of the dSTF-DNA complex, which would in turn make dSTFs compete with bHLH TFs more effectively. To confirm this hypothesis, we set up a competition EMSA assay. While dSTF-B-Z did not show appreciable competition with Max/Max, we observed a promising competition pattern for the stapled dSTF-B1-Z2, as characterized by the conversion of the Max/Max-DNA band into the dSTF-DNA band. These results suggest that the pre-dimerization of B1-Z2 enhances its ability to compete with Max/Max. Given that more candidates, such as B11-Z6 and B11-Z8, have emerged from additional screenings as more capable of competition with Max/Max, it would be intriguing to see if dSTFs derived from those monomers would possess any further enhanced activity. The results also confirm that, even though the hydrocarbon staples in B1-Z2 did not result in observable improvements in DNA binding affinity and specificity over the unmodified B-Z, the contributions from the secondary structure stabilization are likely to be what differentiates dSTF-B1-Z2 from dSTF-B-Z in their bHLH competition, highlighting the importance of stabilizing multiple structural components.

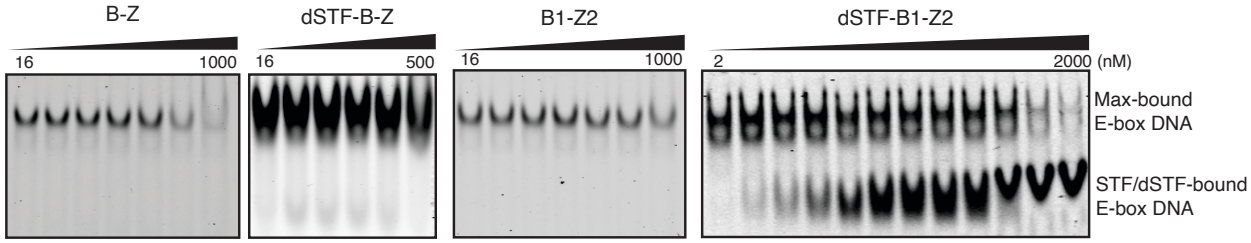


Figure 5.9 Competition of STFs and dSTFs with Max/Max protein

To determine how pre-dimerization of STFs impacts their proteolytic resistance, we performed a kinetic trypsin stability assay for a direct comparison between STFs and dSTFs. As

we have discussed in Chapter 4, B-Z was rapidly proteolyzed with a half-life of approximately 33 seconds. By contrast, dSTF-B-Z showed a prolonged half-life of approximately 90 seconds, a 3-fold increase. The increased stability is most likely a result of the increased protection of the proteolytic sensitive sites at the C-terminus, which represents roughly 30% of the cleavable sites. The presence of the C-terminal linker, in close proximity to the "RRK" motif, effectively prevents the protease from properly accessing these site, and thus improves the overall proteolytic stability.

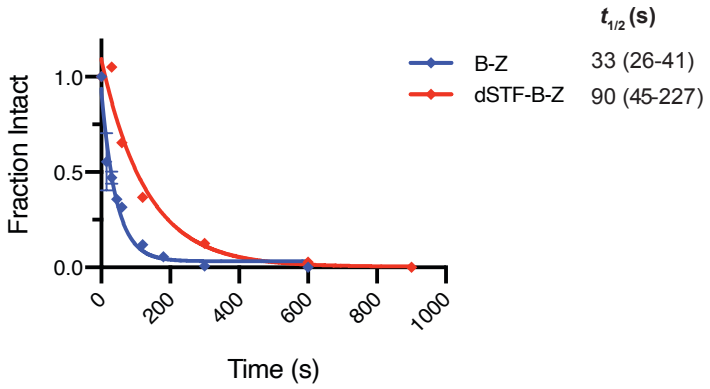


Figure 5.10 LC-MS quantification of STF and dSTF stability to trypsin
The curves and $t_{1/2}$ were determined using a nonlinear one-phase decay fit. The numbers in parentheses represent 95% C.I. from $n = 2$ independent replicates.

5.4 Discussion

In summary, we have developed synthetic schemes for the C-terminal dimerization of STFs. The resulting scaffolds, dSTFs, have shown significant improvement in their *in vitro* activity. Notably, pre-dimerization of both unmodified and stapled STFs increased their DNA binding affinity by at least 10 fold. Given that the EMSA conditions we employed could not precisely determine subnanomolar affinity, alternative conditions or assays with lower detection limit thus are needed to accurately quantify the improvement in affinity. What was most intriguing was that the C-terminal covalent dimerization resulted in a gain of function in terms of competition with bHLH TF DNA binding, validating our approach to further increase the potency of STFs.

Additionally, the preorganized folding from dimerization resulted in a better protection of residues prone to proteolysis, increasing its resistance to protease *in vitro*. While a library of dSTF has not yet been extensively explored, we believe that dSTFs derived from other lead STFs with better properties, such as B11-Z6 and B11-Z8, could possess stronger target affinities, higher propensities for competition with bHLH TFs, and better structural stability.

Despite enhanced properties *in vitro*, one particular concerns for dSTFs is their increased size, which may pose challenges in their cellular uptake and target accessibility. While there has been no evidence suggesting a hard size cutoff for cell penetration of stabilized peptides, existing studies have primarily focused on shorter stapled peptides much smaller than dSTFs[7], raising questions that would further test the limit of endocytosis-related cell penetration, an area worth future investigation.

From a synthetic perspective, with four individually stabilized α-helices and three interhelix linkers, dSTFs represent by far the largest and most complex stabilized miniprotein scaffolds that have been made to date. Continuous additions of orthogonal chemistries to our platform enabled modular and robust synthesis of such complex scaffolds, further pushing the limit of peptide synthesis. Although we have not yet established a viable route towards asymmetric dSTFs, a synthetic target with strong chemical and biological significance, alternative synthetic strategies are actively being explored in our lab.

5.5 Materials and methods

General Peptide Synthesis and Purification. A Symphony X automated peptide synthesizer was used to prepare linear peptides on Rink amide MBHA resin. Fmoc-based solid phase chemistry, ring closing metathesis, and N-terminal modifications were carried out as

previously described[8, 9]. Lysine residues bearing monomethoxy trityl (Mmt) side chain protecting groups were incorporated at cross-linking positions of basic helices. On-resin Mmt deprotection was carried out for 5 × 2 min consecutive cycles of 1% TFA/DCM solution mixed by N_2 bubbling. Deprotected lysine residues were functionalized with maleimide by 2-hour treatment with a 0.1 M solution of 2-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)acetic acid (Mal-Gly-OH) (5 eq), HCTU (4.8 eq), and DIPEA (10 eq) in DMF. Crude peptides cleaved from resin were purified on a Waters preparatory HPLC system using an Xbridge Prep C18 5µ OBN (19.5 × 150 mm) column; solvent A (0.1% TFA in H_2O); solvent B (MeOH); and a 10-min method with the following gradient (flowrate 30 mL/min): 35% B over 1 min; 35-85% B over 7 min; 95% B over 1 min; 35% B over 1 min. STF monomer ligation was performed in 50 mM sodium phosphate buffer pH 7.2 + 25% ACN as follows: a purified basic sequence bearing a maleimide (0.5 mL, 0.5 mM) and a purified zipper sequence with a free thiol (0.5 mL, 0.5 mM) were combined in a microcentrifuge tube and mixed by rotation for 2 hours. at room temperature. The reaction mixture was diluted into 3 mL of 50% ACN/H_2O + 0.1% TFA and the ligated dSTF was purified using the same HPLC method as for individual monomers. dSTF purity was confirmed by LC-MS using an Agilent system equipped with a Phenomonex C18, 5 µm (5.0 × 50 mm) column; solvent A (95:5:0.1 H_2O/ACN/TFA) and solvent B (95:5:0.1 ACN/H_2O/TFA); 0.5 ml min^{-1} flowrate, 0-2 min (0% B), 2-16 min (0-75% B), 16.5-18.5 min (100% B), 19 min (0% B). dSTF concentrations were quantified using 280 nm absorbance readings and compounds were stored as lyophilized powder or DMSO stocks.

Branched peptide synthesis. N^{α}-Fmoc-N^{ε}-Mmt-lysine was coupled onto rink amide MBHA resin. The resin was then subsequently treated with 20% piperidine/DMF (2 × 10 min) and 1% TFA/DCM (5 × 5 min) to deprotect both amino groups. The resulting peptide then was put on

a Symphony X automated peptide synthesizer with IR heating. The coupling of each subsequent amino acid was performed with 5 eq. Fmoc-amino acid, 4.8 eq. HCTU, 10 eq. DIPEA at 65 ºC for 10 minutes.

Electrophoretic mobility shift assays (EMSAs). For direct DNA binding experiments, dSTFs were serial diluted (3-fold increments) at 2x concentration in 20 µL of 1x binding buffer (20 mM HEPES pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 2 mM $MgCl_2$, 0.5 mg/mL BSA, 1 mM DTT, 0.05% NP-40). 20 µL of 10 nM IRD700-labled E-box probe in 1x binding buffer was added and samples were incubated for 30 min at RT followed by 15 min at 4°C. 3.5 µL of each reaction was loaded on an 8% acrylamide, 0.5x TBE gel equilibrated to 4°C. Electrophoresis was carried out for 60 min at 110 V and 4°C with 0.5x TBE + 1 mM $MgCl_2$ running buffer. Gels were pre-run at 110 V for 60 min prior to sample loading. For competition experiments, 2 nM dSTF and 5 nM labeled E-box probe were incubated with 0, 5, 25, 125, or 625 nM unlabeled competitor oligo for 30 min at RT followed by 15 min at 4°C.

Gels were imaged using an Odyssey Li-COR. ImageJ was used to quantify band intensity and fraction bound DNA was calculated by dividing band intensity of bound DNA by the band intensity of the free DNA from a vehicle treated lane. A four-parameter dose-response curve fit to a plot of normalized fraction bound DNA vs. log B-Z concentration yields an IC_{50} which was reported as the apparent K_D. Mobility shift data were excluded from analysis when higher order binding species were observed.

Max Protein Expression and Purification. Human Max (residues 22–102) protein was expressed with a N-terminal hexahistidine tag in *Escherichia coli* strain BL21(DE3) using a pET28c vector. Transformed bacteria was grown at 37°C and induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) at an A_{600} = 0.8. Cells were pelleted 14 h after induction and lysed in

lysis buffer (100 mM NaH_2PO_4, 10 mM Tris, 300 mM NaCl, 8 M urea, 10 mM imidazole, pH 8.0) with cOmplete EDTA-free Protease Inhibitor (Roche) by sonication. The lysate was centrifuged to clear insoluble matter before loading onto Ni-NTA resin (Qiagen). After being washed once with lysis buffer and three times with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, 20 mM imidazole, pH 8.0), column-bound protein was eluted using elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 8 M urea, 250 mM imidazole, pH 8.0) and quantified using Bradford assay. The eluted Max protein was then dialyzed into desired buffers suitable for downstream applications, and further concentrated by centrifugation using a 10-kDa exclusion filter.

***In vitro* trypsin stability assay.** STF or dSTF, 30 μM, was dissolved in 330 μL of 20 mM phosphate buffer pH 7.4 in a microcentrifuge tube and heated to 37°C on a tabletop shaker (500 rpm). 30 μL of the reaction mixture was added to 60 μL of quenching solution (ACN + 3% formic acid with 500 nM Fmoc-lysine internal standard) for 0 s timepoint sample. Thermo-Pierce MS-Grade Trypsin was added to a final concentration of 0.5 μg/ml and additional 30 μL aliquots were quenched at indicated timepoints. Quenched samples were cooled to 4°C and centrifuged at 20,000 xG for 3 min. Sample injections of 35 μL were analyzed by LC-MS using an Agilent system equipped with a Phenomonex C18 5μ (5.0 × 50 mm) column; solvent A (95:5:0.1 H_2O/ACN/TFA) and solvent B (95:5:0.1 ACN/H_2O/TFA); 0.5 ml min^{-1} flowrate, 0-2 min (5% B), 2-8.8 min (5-95% B), 9-11 min (95% B), 11.1 min (5% B). Intact STF/dSTF was calculated by normalizing the background subtracted integrated area under the curve for the EIC of $(M + 4H^+)/4$ (for STF) or $(M + 12H^+)/12$ (for dSTF), ± 0.5 mass units to the integrated area under the curve for the A_{280} peak of the internal standard. Fraction intact STF/dSTF was calculated by dividing the intact STF/dSTF by the normalized STF/dSTF signal in the initial 0 s sample. GraphPad Prism was used to plot

fraction intact STF/dSTF vs time, and the proteolytic half-life was derived using a non-linear one-phase decay with the plateau constant set equal to 0.

References

1. Bycroft, B. W.; Chan, W. C.; Chhabra, S. R.; Hone, N. D., A novel lysine-protecting procedure for continuous flow solid phase synthesis of branched peptides. *Journal of the Chemical Society, Chemical Communications* **1993,** (9), 778-779.

2. Vega, S. C.; Martinez, D. A.; Chala, M. D. S.; Vargas, H. A.; Rosas, J. E., Design, Synthesis and Evaluation of Branched RRWQWR-Based Peptides as Antibacterial Agents Against Clinically Relevant Gram-Positive and Gram-Negative Pathogens. *Front Microbiol* **2018,** *9*, 329.

3. Li, D.; Elbert, D. L., The kinetics of the removal of the N-methyltrityl (Mtt) group during the synthesis of branched peptides. *J Pept Res* **2002,** *60* (5), 300-3.

4. De Masi, F.; Grove, C. A.; Vedenko, A.; Alibes, A.; Gisselbrecht, S. S.; Serrano, L.; Bulyk, M. L.; Walhout, A. J., Using a structural and logics systems approach to infer bHLH-DNA binding specificity determinants. *Nucleic Acids Res* **2011,** *39* (11), 4553-63.

5. Blancafort, P.; Segal, D. J.; Barbas, C. F., 3rd, Designing transcription factor architectures for drug discovery. *Mol Pharmacol* **2004,** *66* (6), 1361-71.

6. Nwe, K.; Brechbiel, M. W., Growing applications of "click chemistry" for bioconjugation in contemporary biomedical research. *Cancer Biother Radiopharm* **2009,** *24* (3), 289-302.

7. Chu, Q.; Moellering, R. E.; Hilinski, G. J.; Kim, Y.-W.; Grossmann, T. N.; Yeh, J. T. H.; Verdine, G. L., Towards understanding cell penetration by stapled peptides. *MedChemComm* **2015,** *6* (1), 111-119.

8. Kim, Y. W.; Grossmann, T. N.; Verdine, G. L., Synthesis of all-hydrocarbon stapled alpha-helical peptides by ring-closing olefin metathesis. *Nature protocols* **2011,** *6* (6), 761-71.

9. Mitra, S.; Montgomery, J. E.; Kolar, M. J.; Li, G.; Jeong, K. J.; Peng, B.; Verdine, G. L.; Mills, G. B.; Moellering, R. E., Stapled peptide inhibitors of RAB25 target context-specific phenotypes in cancer. *Nat Commun* **2017,** *8* (1), 660.