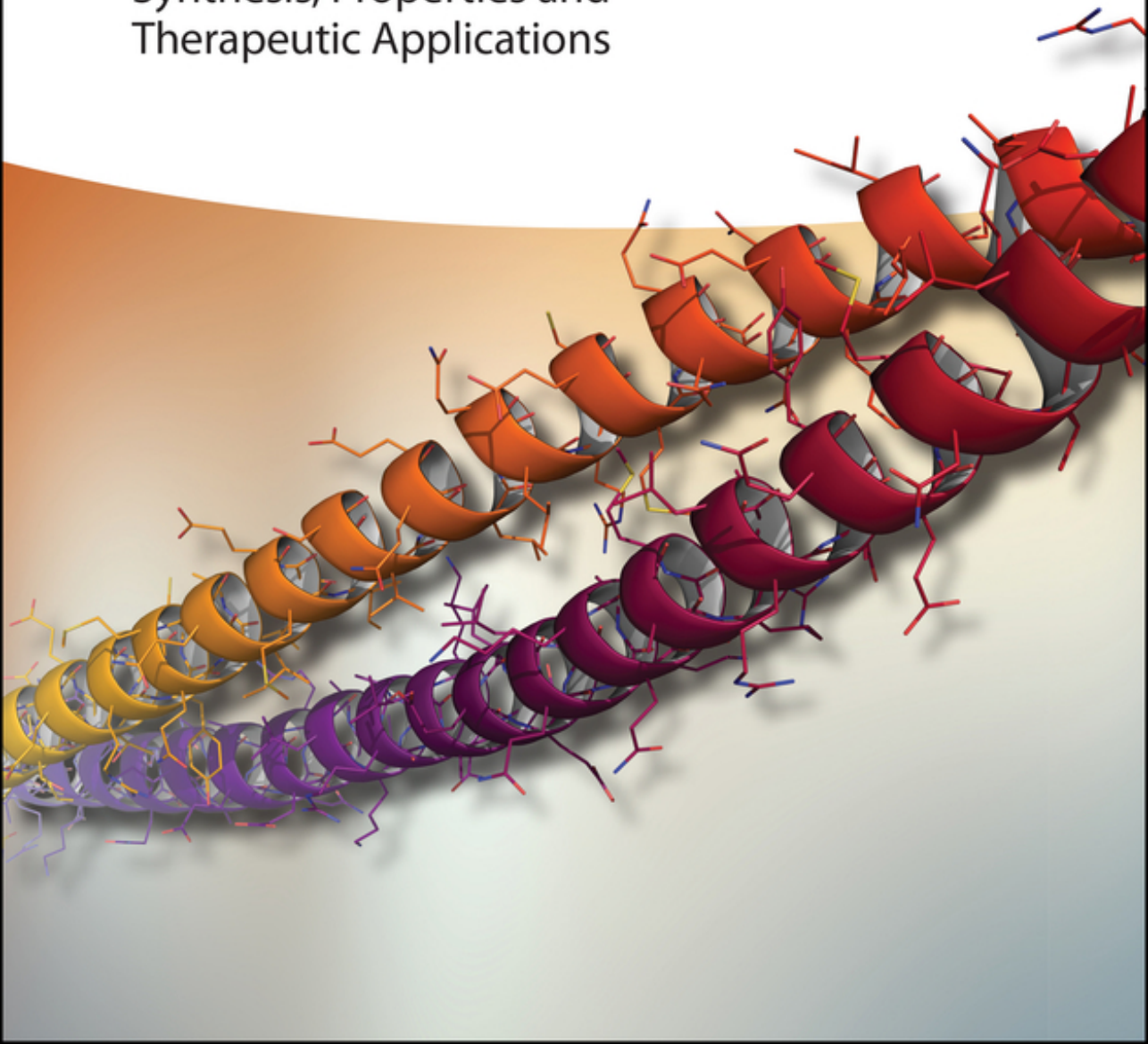
WILEY-VCH

Zigang Li, Hui Zhao, and Chuan Wan

Cyclized Helical Peptides

Synthesis, Properties and Therapeutic Applications



Cyclized Helical Peptides

Cyclized Helical Peptides

Synthesis, Properties, and Therapeutic Applications

Zigang Li
Hui Zhao
Chuan Wan

WILEY-VCH

Authors

Prof. Zigang Li
Shenzhen Graduate School of Peking University
Chemical Biology and Biotechnology
Lishui Road
Xili Town
518055 Shenzhen, Nanshan Disctrict
China

Dr. Hui Zhao
Shenzhen Graduate School of Peking University
Chemical Biology and Biotechnology
Lishui Road
Xili Town
Nanshan Disctrict
518055 Shenzhen
China

Dr. Chuan Wan
Shenzhen Graduate School of Peking University
Chemical Biology and Biotechnology
Lishui Road
Xili Town
Nanshan District
518055 Shenzhen
China

Cover Design: Wiley
Cover Image: © Molekuul/iStockphoto; © ivanastar/iStockphoto

Library of Congress Card No.:
applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.d-nb.de>.

Print ISBN: 978-3-527-34342-3
ePDF ISBN: 978-3-527-34347-8
ePub ISBN: 978-3-527-34345-4
oBook ISBN: 978-3-527-34343-0

Typesetting Straive, Chennai, India

Printing and Binding

Printed on acid-free paper

10 9 8 7 6 5 4 3 2 1

Contents

1

Introduction

1.1 Protein–Protein Interactions and Their Small-Molecule Modulators

1.1.1 Characteristics of Protein–Protein Interactions

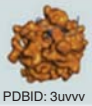
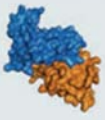
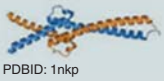
Proteins that work and degrade in highly congested and complex environments must be found by their partners in a large number of non-partners. It is estimated that human beings have 650 000 different pairs of interactions, which are responsible for a number of key biomolecular processes [1]. The surface of soluble proteins is covered by hydrophobic and hydrophilic residues, as well as by hydrophilic backbone. The highly specific physical contact between two or more protein molecules is mainly related to hydrophobic interactions, salt, and hydrogen bonds.

Protein–protein interactions have different affinity and longevity. Some complexes are weakly and instantaneously clustered; some may continue to form part of a larger protein complex, stabilized through multiple interactions; some reversible signal complexes have high pairing affinity, but only limited time; some complexes are stable, but have built-in timers; the presence of antibodies and antigens and protease and inhibitor complexes can take up to a day, some of which may be categorized as irreversible [2].

In addition, protein–protein interactions can be categorized according to the structural characteristics (Figure 1.1) [3]: the interaction between globular protein pairs, the interactions between globular proteins and individual peptide chains with continuous or discontinuous table position, and the interaction between two segments of peptide chains. Correspondingly, the polypeptide that participates in protein–protein interactions may adopt a combination of structures: the extension structure in the groove, β-sheet, α-helix, and even the poly-proline helix.

There is certain regularity in the presence of amino acid residues in proteins [3a]. In the general interface, leucine is the most common residue, followed by arginine. Furthermore, charged residues are more common than polar residues, and both, except for arginine and histidine, are generally abundant on the surface. Aromatic amino acids, except for tryptophan, have a very low abundance on the surface but have a high abundance at the interface. As is mentioned above, the frequency of

Cyclized Helical Peptides: Synthesis, Properties, and Therapeutic Applications, First Edition.
Zigang Li, Hui Zhao, and Chuan Wan.

PPI class	Description	Simplified illustration	Examples (target-displaced)	Examples structure
Globular protein-helical peptide, discontinuous epitope	Helix with a discontinuous epitope binding into a groove		*MDM2-p53 *BCL-X_L–BAD and BCL-X_L–BAK* *ZipA–FtsZ *S100B–p53 *β–Catenin–TCF3–TCF4 *MCL1–BH3 *SUR2–ESX	 Protein Data Bank (PDB) ID: 2xa0
Globular protein-peptide, continuous epitope	Continuous epitope on β-sheet or β-strand and loops binding into surface with pockets		*XIAP–SMAC* *HIV integrase–LEDGF *Integrins *RAD51–BRCA2 *PDZ domains *NRP1–VEGFA *Menin–MLL	 PDBID: 1g73
Globular protein-peptide, continuous epitope	Binding into pocket in a β-propeller		*KEAP1–NRF2* *WDRS–MLL	 PDBID: 2dyh
Globular protein-peptide, anchor residue	Peptide with an anchor residue owing to post-translational modification binding into a pocket		*Bromodomains* *PDEδ–KRAS *SH2 domains *PLK1 PBD–peptide *VHL–HIF1a	 PDBID: 3uvvv
Globular protein-globular protein, discontinuous epitope	Two proteins both presenting discontinuous epitopes		*IL-2-IL-2R* *TNF–TNF *E2–E1	 PDBID: 1z92
Peptide-peptide	A pair of helices with an elongated binding interaction		*MYC–MAX* *NEMO–IKK *Annexin II–P11 (also known as S100A10)	 PDBID: 1nkp

BAD, BCL-2-associated agonist of cell death: BAK, BCL-2 bomologousantagonist/killer: BCL B cell lymphoma: BH3, BCL-2 bolology domain 3: BRCAZ, breast Cancer type 2 susceptibility protein: HIF1a, hypoxia-inducible factor 1α: IL-Z, interleukin-2: IL-2R, interleukin-2 receptor: IKK, inhibitor of nuclear factor x8 kinase: KEAP1, kelch-like ECH- associated protein 1: LEDGF, lens epitheliurn-derived growth factor: MAX MYC-associated factor X: MCL1, myeloid cell leukaemia 1: MLL, mixed-lineage leukaemia: NEMO, nuclear factor κB essential modulator: NRF2, nuclear factor erythroid 2-related factor 2: NRP1, neuropilin 1: PLK1 PBD, polo-like kinase 1 polo box domain: S100A10, S100 calcium-binding protein A10: SH2, SRC homology 2: SMAC, second mitochondria- derived activator caspase:SUR2, mediator of RNA polymerase II transcription subunit 23: TCF3, HMG box trnscription factor 3: TNF, tumor necrosis factor: VEGFA vascular endothelial growth factor: VHL, Von Hippel-Lindau disease tumor suppressor: WDRS, WD repeat-containing protein 5: XIAP, X-linked inhibitor of apoptosis protein *Example structure illustrated in the column to the right.

Figure 1.1 Classification of protein–protein interactions and examples [3b]. Source: Scott et al. [3b]. © 2016, Springer Nature.

occurrence of hydrophobic residues is generally high at the interface and is low on the surface. Cysteine is particularly rare both on the surface and at the interface.

In addition, based on the results of alanine scanning mutagenesis, the residual base that has a great influence on the binding affinity is called "hot spot" [4]. Hot spots are almost always buried in the center of the core, not in contact with solvents. The hot spot processes the highest sequence conservation [5]. Tryptophan, arginine, and tyrosine are the most common, accounting for more than half of the total, as hot spots. These three versatile residues were able to form hydrophobic, aromatic, and polar interactions, all of which can be wrapped in complementary surfaces to meet unpaired hydrogen-bonded donors and receptors. In addition, the polar "π-cation" bond between arginine and tryptophan or tyrosine was found in more than 50% hot spots [6]. Apart from a "π-cation" bond with arginine, the traditional side chain interaction is more common for tyrosine. By contrast, the most common residual at the interface, leucine, is rarely found in hot spots, while isoleucine is rich.

In the complexes in the protein database, 62% has a helix on the interface [7]. However, the presence of a helix at the interface does not mean that the helix plays a

key role. Analysis shows that in about 60% of the interface, the hot residue is located on one side of the helix, one-third of the complexes with the hot spots on two faces of the helix, and about 10% of the complex with all three faces participating in the interaction with the target protein. In the protein database, the first four major types of function of protein–protein interactions, where helices are involved, are gene regulation, enzyme function, cell cycle, and signal transduction.

Analysis of the contribution of each helix residue to the interaction shows that leucine appears most in the interface area. This is not surprising, because in general, leucine is also the most common residue in proteins. After the normalization of natural abundance, aromatic amino acids, arginine, and leucine are of the highest frequencies at the helix interface as compared with polar residues [4, 8]. In addition, polar and charged residues are also important contributors to the interface.

1.1.2 Intervention of Protein–Protein Interactions Using Small Molecules

Abnormal protein–protein interactions are the basis of multiple diseases, and an increasing number of researchers are committed to developing molecules to modulate protein interactions for therapeutic purposes. Small molecule is a class of entity with potentially ideal therapeutic potentials. However, the contact surface of some of the protein interactions is large and shallow (about 1000–6000 $Å^2$), especially those featured by a linear peptide epitope 1–4 amino acids long, compared to the traditionally small and deep small-molecule binding pockets [9] (Figure 1.2). Therefore, the interface between proteins is sometimes regarded as a target of "undruggable." In establishing guidelines for the discovery of protein-protein interaction (PPI) inhibitors, clinical success cases should be considered in the context of the type of interface.

Work in recent years has begun to show that some protein–protein interactions are able to be suppressed by small molecules. Most of the developed inhibitors target PPIs, where hot spot residues are restricted to small binding pockets (250–900 $Å^2$) [11]. Some small-molecule inhibitors disrupt the interaction between a globular protein and a single peptide chain with a secondary or tertiary structure, through binding to the pocket on the globular protein. It is noteworthy that the secondary structural features processed by the peptide chain, such as α-helices and β-strands, have important implications for the design of inhibitors that mimic and replace these peptides. With a better understanding of the structural biology of the protein–protein interactions, it seems more promising and reasonable to discover drugs targeting protein–protein interactions with defined structures. In addition, the hot spots of the interaction interface can be targeted by inhibitors of protein–protein interactions. The interaction of the rigid globular protein with a polypeptide may be more suitable for small-molecule interruption because the polypeptide can contribute more to the binding energy and be replaced by the small molecule with good design. At present, there are many strategies to discover hits or leads that interfere with protein–protein interactions, the most notable of which is high-throughput screening, fragment screening, and optimization.

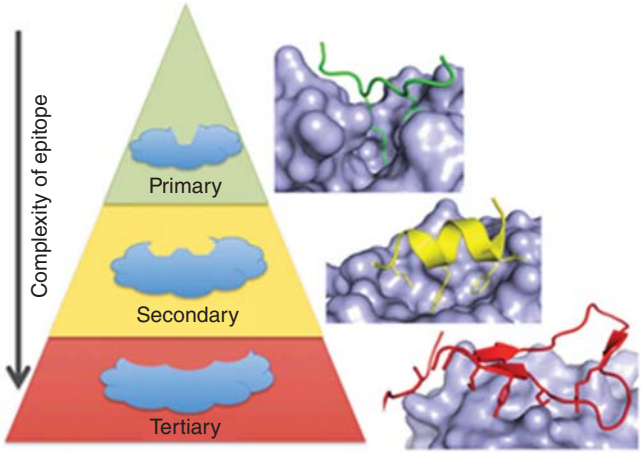


Figure 1.2 The complexity of the PPI interface affects druggability PPIs can be classified by whether one side of the interface consists of a primary (linear) protein sequence (green), a single region of secondary structure (such as an α-helix, yellow), or multiple sequences requiring tertiary structure (red). There are fewer examples of small-molecule inhibitors of PPIs as the interface becomes more complex (from primary to secondary to tertiary epitopes). Structures shown are BRDt/histone (green; Protein Data Bank [PDB]: 2WP1), MDM2/p53 (yellow; PDB: 1YCR), and IL-2/IL-2Ra (red; PDB: 1Z92) [10]. Source: Arkin et al. [10]. © 2014, Elsevier.

High-throughput screening is an effective way to find a hit in a traditional drug target. Most of the high-throughput screening strategies rely on assays such as fluorescence resonance energy transfer, amplified luminescent proximity homogeneous assay screen, surface plasmon resonance, or fluorescence polarization because they are highly efficient, sensitive, and reagent-available [12]. However, these methods can usually disrupt enzyme activity and lead to more false-positive signals. Another method is based on the label-free strategy, including the refractive index properties and mass spectrometry [12b]. Their applications may be more extensive, more quickly developed, and robust because they eliminate the steps associated with introducing and observing tags. Despite these established methods, it is still difficult to effectively generate protein–protein interaction inhibitors through high-throughput screening since the compounds used for screening are mainly targeting traditional drug targets. Traditional high-throughput screening faces some challenges in dealing with protein–protein interactions – low hit ratio, low activity, and hard to eliminate false positives [12b]. However, high-throughput screening has been successfully applied in the discovery of the analog of discontinuous epitope on an α-helix.

The fragment-based drug discovery is a strategy to discover molecules from smaller fragment of drugs or functional groups with low affinity, which can effectively explore the chemical space [13]. These fragments can simplify the calculation and analysis of ligand binding to improve affinity. The discovery of drug fragments in the past has become an effective way to target protein–protein interactions. Many protein interfaces have anchored residues to occupy the pockets of proteins, such as tyrosine, phenylalanine, tryptophan, or leucine [14]. The pockets of the short

peptide with well-defined structure can effectively become the target of the drug discovery based on fragments [15]. Fragment drug discovery screening usually consists of two steps. The first step involves using a surface plasmon resonance or differential scanning fluorescence for a preliminary rapid screening [16]. The second step includes more targeted validation of the hit molecule, the use of X-ray crystallography or protein-based nuclear magnetic resonance (NMR) to define the spatial aspects of the binding site, the thermodynamic parameters defined by isothermal calorimetry, and the surface plasmon resonance to define kinetics [15]. Fragment discovery methods combine fragment space with the enhanced hit ratio for lower complexity molecule, making itself a powerful lead generation tool. Compared with high-throughput screening, fragment-based drug discovery can capture more chemical structures with different hits, providing more hits for a larger number of protein targets, higher recognition rates and fewer false positives, and simpler and more reliable detection methods [17]. However, the need for a large number of proteins is also a problem that fragment-based drug discovery needs to address. In addition, fragments combine computational analysis aspects, requiring new hardware design or new concepts and great progress.

Virtual screening based on structure is an important tool to help the discovery and optimization of potential lead in a fast and cost-effective way based on structural drug discovery. The virtual screening based on structure is used to select the large-class drug compound library. Then, the screened out promising compounds were selected for experimental testing. In the method of de novo design, the three-dimensional (3D) structure of receptors is used to design novel molecules that have never been synthesized before using ligand growing programs and the intuition of medicinal chemists [18]. Compared with high-throughput screening, the discovery of computer-assisted drugs has the advantage of predicting new bioactive compounds and their receptor-binding structures, and in some cases having a greater hit rate.

So far, using the above methods, a number of small-molecule compounds targeting protein–protein interactions have entered clinical trials. Here are some successful examples of small-molecule inhibitors that interfere with protein–protein interactions.

1.1.2.1 Leukocyte Function-Associated Antigen-1

Leukocyte function-associated antigen-1 is a β2 integrin that participates in the activation and adhesion of T cells and is a target in the weakening of inflammatory immune response [19]. Lymphocyte function-associated antigen 1, the heterodimer consisting of an α-chain and a β-chain, binds to its ligand intercellular adhesion molecule-1, which is important for T cell–T cell interactions. The anti-lymphocyte function-associated antigen 1 antibody efalizumab, an immunosuppressant that inhibits lymphocyte activation and cell migration by binding to the CD11a subunit of lymphocyte function-associated antigen 1, had been approved for psoriasis and then withdrawn for immunosuppression-induced fatal viral infections [20]. Another leukocyte function-associated antigen-1 antagonist lifitegrast, which was discovered by Sunesis, and then developed clinically by SARcode/Shire, has been

approved as the only drug for the treatment of dry eye disease [21]. The mechanism of action of the molecule is under debate, which is the inhibition of leukocyte function-associated antigen-1 from binding to intercellular adhesion molecule 1, associated with either intercellular adhesion molecule site on the I domain or the related site on the I-like domain [22]

1.1.2.2 Inhibitor of Apoptosis Proteins

Apoptosis is a programmed cell death mediated by caspases activation. Inhibitor of apoptosis proteins has been expressed in tumor cells by inhibiting the activity of apoptosis-inducing protease, which regulates the fate of cells, including death and immunity of apoptotic cells. X-linked inhibitor of apoptosis proteins is the most effective inhibitor of apoptosis proteins, which interact with the initiator caspase-9 through the Baculoviral IAP repeat 3 structure domain and caspase-3/7 through the Baculoviral IAP repeat 1/2 domain [23]. Discovering new compounds that inhibit the interaction between X-linked inhibitor of apoptosis proteins and enzymes is thought to be a promising strategy for cancer treatment. Smac is a natural protein inhibitor of X-linked inhibitor of apoptosis proteins, which competes with caspase binding to Baculoviral IAP repeat domain through the alanine–valine–proline–isoleucine tetra-peptide at the nitrogen end [24]. Smac proteins have attracted the attention of academics and pharmaceutical companies to the design of small-molecule smac simulators [25]. There are currently two types of inhibitors, including univalent and bivalent inhibitors. Some of them have entered phase II clinical stage, such as LCL-161 by Novartis and Debio-1143 by Debiopharm [12a].

1.1.2.3 Bromodomains

The acetylation of lysine residues or the methylation of lysine and arginine residues can be "read," which undertakes central roles in epigenetic regulation. Bromodomains and histone interactions are important in controlling gene expression and DNA repair and in regulating inflammation and cancer. The bromodomains share a conservative structure consisting of four α-helical bundles, which are connected by different cyclic regions of variable charge and length. A hydrophobic pocket includes a conservative aspartic amide and five water molecules that can identify acetylation of lysine [26]. A quinazoline compound, apabetalone by Resverlogix, which is able to increase transcription of the ApoA-I gene by inhibiting bromodomain and extra terminal domain proteins, especially bromodomain-containing protein 4, is in phase III clinical trials, for the potential treatment of diabetes mellitus, renal impairment, and cardiovascular diseases [12a].

1.1.2.4 Human Immunodeficiency Virus Integrase

The homotetrameric protein human immunodeficiency virus integrase integrates the viral genome into human DNA, which is vital for human immunodeficiency virus replication. The protein–protein interactions between the human immunodeficiency virus 1 integrase and the growth factor/p75 of the host protein lens epithelium are key to this process, which makes integrase a target for human immunodeficiency virus. The structure of human immunodeficiency virus integrase

consists of three domains, nitrogen-terminal DNA binding domains, catalytic core domains, and carbon-terminal DNA binding domains. The catalytic core domain has several pockets selected as small molecular target for inhibition of enzyme activity [27]. Now several small molecules targeting the enzyme have been approved for the treatment of human immunodeficiency virus infections, such as raltegravir by Merck, dolutegravir by GSK, and elvitegravir by Tobacco, which inhibit the enzyme activity by binding to the active site. Furthermore, a series of 2-(quinolin-3-yl) acetic acid derivatives, including clinical compound BI-224436 by Gilead, have been developed to block the integration step, by inhibiting lens epithelium-derived growth factor/p75-integrase interaction, which displays a different resistance profile [28].

1.1.2.5 B-Cell Lymphoma-2 Family/B-Cell Lymphoma-2 Homology 3 Proteins Interaction

B-cell lymphoma family proteins, including members of the family promoting apoptosis and resistance to family members, are the central effectors of cell apoptosis. B-cell lymphoma-2 homology 3-containing promotes apoptotic proteins, such as B-cell lymphoma-2 homologous antagonist killer, through a single helix binding to hydrophobic pockets that inhibit the apoptosis of B-cell lymphoma-2 proteins. The use of small-molecule compounds to simulate the B-cell lymphoma-2 homology 3 domain has shown significant therapeutic potential [29]. Several B-cell lymphoma-2 homology 3 simulations were determined by the selection of NMR-based fragments and the optimization of structures. For example, ABT-737 has an affinity for B-cell lymphoma-2 and with nanometer mole range [30]. This compound occupies the same hydrophobic bag as a B-cell lymphoma-2 homologous antagonist killer-derived peptide, which has the same binding position as B-cell lymphoma-2 homologous antagonist killer's Leu78 and Ile85 to bind to the key residues. Another small-molecule therapeutic Venetoclax, which was granted Breakthrough Therapy Designation by USFDA, has been approved for the treatment of chronic lymphocytic leukemia.

1.1.2.6 Mouse Double Minute 2–p53 Interaction

The interaction between p53 and its negative regulatory protein mouse double minute 2 (MDM2) or MDMX is the target of anticancer treatment, and it is also a common model system to evaluate the new method of protein–protein interaction inhibition. This interaction is mediated by the short α-helix peptide sequence of p53, which binds to the globular domain of MDM2 or MDMX. Structurally, N-terminal domain of MDM2 binds to a short 15-residual α-helix peptide of p53, where three hydrophobic residues of p53 occupy a well-defined hydrophobic pocket on MDM2 [31]. These structural characteristics make the strategy of targeting MDM2–p53 protein–protein interaction feasible. Various methods had been used to determine the inhibitors of mdm2–p53 interactions, and a series of *cis*-imidazoline analogs, nutlins, were determined by screening the complex library [32]. These molecules are able to inhibit the interaction between p53 and mdm2 and adopt the same binding mode as the key residues of p53. Among these analogs, idasanutlin by Roche was in phase III clinical trial for the potential treatment of acute myelogenous leukemia.

1.2 Features of Peptide as Molecular Tools

1.2.1 Advantages of Peptides as Molecular Tools

Natural or artificial peptides or proteins play a central role in molecular processes, thanks to their strong molecular recognition capabilities. The strong recognition ability of peptides can be explained by a large number of different types of functional groups, which are easy to construct. Amino acids are rich in physicochemical properties. By polarity, they can be classified as basic, acidic, nonpolar, or polar amino acids, which provide hydrogen-bonded donors, receptors, or hydrophobic cores. According to rigidity, some amino acids are flexible, such as glycine, while others process fixed angles such as proline. This rich building block, combined with an easy combination of amide keys, makes polypeptides diverse and complex enough to form macromolecules of a particular nature and mediate important molecular processes accordingly. In addition, a large number of posttranslational modifications and unnatural amino acids greatly enhance their potential function.

In addition, peptide-mediated identification processes are ubiquitous in nature, including those between proteins/peptides and proteins/peptides, between proteins/peptides and nucleic acids, and between proteins and lipids, all of which involve all processes of biological systems. These structural information are known or readily available and can be designed according to complex structures of polypeptide modulators or further explore the development of small-molecule inhibitors.

Further, polypeptides and proteins are easy to screen and evolve. Because the amide bond is easy to construct, the peptide combinatorial library can be easily used in the screening of active sequences. Protein/polypeptide is located at the end of the central code, so their molecular evolution can easily be achieved through the appropriate size of DNA libraries, biological systems, and Darwinian choices. In contrast, the direct evolution of small molecules is difficult. An overview of the most common technologies is presented in Table 1.1. All techniques are somewhat related and share common steps. The common technical strategies for peptide screening are described below.

Multi-peptide arrays were synthesized by speckle technique. As a high-throughput research tool, peptide arrays are a new type of biochip that uses automated instrumentation, in situ synthesis, to design hundreds of or even thousands of polypeptides in very high density. This peptide chip can be incubated directly with a variety of different biological samples. After several washing steps, a secondary antibody, which is typically labeled by a fluorescent label and can be detected by a fluorescent scanner, is applied [44].

Protein/peptide evolution techniques are further modified to meet more application needs. The use of powerful techniques to generate and screen DNA-encoded protein libraries helps promote protein development as a drug ligand. However, their use as drug ligands is limited by their intrinsic characteristics. Two intrinsic limitations include rotational flexibility of the polypeptide backbone and a limited number

Table 1.1 Summary of key features of screening technologies commonly used for cyclic peptide discovery.

	Library size and its restriction	Screening host	Cyclic	Nonnatural amino acids/PTMs/backbone modification
One-bead-one-peptide [33]	106, library construction	In vitro	Various chemistries	Yes – all resin compatible chemistries
SICLOPPS[34]	109, Transformation efficiency	In cellulo	Head-to-tail cyclization via split-intein chemistry	Possible with amber codon suppression
Peptide on plasmid [35]	109, transformation efficiency	Bacteria	Possible	
Prokaryotic [36]	109, transformation efficiency	Bacteria	Yes – commonly by post translational cysteine alkylation	
Eukaryotic [37]	107, transformation efficiency	Yeast		
Phage [38]	109, transformation efficiency	Phage		
CIS [39]	1014, translation scale	In vitro	Yes – commonly	Yes – possible
Ribosome [39]				
mRNA [40]				
RaPID [41]/TRAP[42]			Yes – commonly *N*-acetyl chloride chemistry	Extensive reprogramming using the FIT system

Source: Obexer et al. [43]. © 2017, Elsevier.

(20) of natural amino acids. However, these restrictions can be overcome by using chemical modifications.

In the one-bead-one-peptide (1B1P) method, pioneered by Lam and Salmon in 1991 [45], split and pool synthesis techniques were developed to generate diverse libraries of beads (up to 107 compounds currently), each coated with multiple copies of a unique peptide. Resin-compatible chemistries had been exploited to make diverse backbones, peptoids, D-amino acids, and peptide cyclizations accessible. Pei Lab presented a method for synthesizing and screening a complex 1B1C Library of cyclic peptides for biological targets, such as proteins. In the Tentagel micro-beads, up to 10 million different cyclic peptides were synthesized rapidly by split-pool synthesis, and followed by multistage screening scheme, including fluorescent activated cell sorting, magnetic selection, the enzyme-linked reaction on beads, and the analysis of cyclic peptides in solution by fluorescence anisotropy. Finally, the most active hits are determined by the partial Edman degradation-mass spectrometry [33].

Split-intein circular ligation of peptides and proteins (SICLOPPS), an in vivo method for discovering head-to-tail cyclic peptides, is free from genetic code reprogramming. The method applies split-intein chemistry to cyclize randomized peptide sequences. The cyclic peptide library can potentially be of any size, and the peptide itself may contain unlimited random residues, including unnatural amino acids [46]. Plasmid propagation within cells bridge genotypes and phenotypes. Accordingly, transformation efficiency limits the achievable library size to 1079. Apart from being implemented in *Escherichia coli*, SICLOPPS has since been extended to eukaryotic cells [47].

Phage display was for the first time reported by George P. Smith in 1985 [38a]. Phage display technology is considered as a fast and effective method for screening small peptides. In this technique, a gene encodes an interest protein into the phage shell protein gene, and the phage displays the protein outside of it for binding force screening. Phage display is a useful tool for drug discovery, but there are some deficiencies. First, the library's capacity can only reach 109, which is limited by transfection efficiency. Second, we need to solve the diversity problem of the polypeptide library. Third, a small amount of peptide due to its hydrophobicity or because of the folding of the outer membrane protein cannot be displayed on the phage surface. During phage display, chemical epoxidation can be incorporated to directly evolve the cyclic peptide, including the direct evolution of polycyclic polypeptide and helical peptides. For example, in situ cyclization is easily realized by disulfide bridging or alkylation via cysteine or enzyme-mediated modifications [48].

Ribosome display techniques are used to perform protein evolution in vitro, producing proteins that can bind to an ideal ligand [40]. This technique optimizes the interaction of functional proteins through Plückthun laboratories. The ribosome shows the beginning of the polypeptide encoded from the DNA sequence's original library. Each sequence is transcribed and then translated into adult foreign peptides. The DNA library is fused to a lack of a stop codon interval sequence. The absence of a stop codon prevents the release factor from binding, triggering the dispersal of the transcription complex. Therefore, this interval sequence remains connected to the peptide tRNA, occupying the ribosome tunnel, which makes the protein of interest protruding from the ribosome and folds. The resulting mRNA, ribosomes, and protein complexes can be screened. The filtered mRNA was then transcribed to the cDNA and amplified by polymerase chain reaction (PCR). Since it is carried out completely in vitro, there are two main advantages over other alternative techniques. First, the diversity of the library is not limited by the transfection efficiency of bacterial cell but is only affected by the number of ribosomes and different mRNA molecules in the test tube. Second, random mutations can be easily introduced after every choice, making proteins evolve over several generations.

Traditionally, reprogramming was achieved via stop codon suppression or removal of canonical amino acids. More extensive genetic code reprogramming is particularly facile when carrying out in vitro display techniques, where enhanced library diversity can be achieved through reconstituted translation systems giving compositional freedom [49]. The flexible in vitro translation (FIT) method

remains the most versatile and least labor-intensive procedure for genetic code reprogramming in vitro. Furthermore, integrating the FIT system with mRNA display gave rise to the random nonstandard peptide integrated discovery (RaPID) system [50], the versatility of which is reflected in its broad use in the pharmaceutical industry. Bashiruddin et al. created tricycles, by combining bicycle bridging moiety with the classic *N*-acetyl chloride cyclisation of the RaPID system [51].

The ligand is an oligonucleotide or peptide molecule that binds to a specific target molecule. The aptamer is usually created by selecting them from a large random sequence pool, but the natural ligands also exist in riboswitches. Polypeptide ligands are artificial protein choices or are designed to bind to specific target molecules [52]. These proteins are represented by one or more polypeptide loops by a variable sequence of protein scaffolds. They are usually separated from the combinatorial library and are often modified by directional mutation or by mutation and selection of the variable region. In vivo, peptide ligands can bind to cell protein targets and exert biological effects, including interfering with the normal protein interactions between their target molecules and other proteins. The library of polypeptide ligands is used as a "mutation." In 2013, Shekhtman Laboratories developed a method to build a combinatorial library of improved peptide adaptation (clips) of high complexity, containing more than 3×10^{10} independent clones as molecular tools for the study of biological pathways [53]. The protein skeleton was modified to improve its solubility, and the aggregation of peptide was eliminated. Clips is used in yeast two-hybrid screening to determine the peptide adaptation to the late glycation end product of receptors in different domains. Cell function detection showed that, in addition to direct interference with the known binding sites, the combination of polypeptide and distal and ligand sites inhibited the signal transduction of rage ligand-induced signaling. The findings highlight the potential of using fragments to select biological targeting inhibitors.

1.2.2 Disadvantages of Peptides as Molecular Tools

Compared with the small molecule, the polypeptide has some disadvantages in the properties of proprietary medicines, which limited its bioavailability. First, amide bonds are fragile in vivo, making stability a fatal weakness of natural linear peptides. There are many ways to improve its stability, such as the insertion of a loop or unnatural module. In Chapter 2, the stability method of helix peptide is summarized and discussed. Furthermore, the penetration of the peptides is limited. The compartments in the organism are mainly composed of lipophilic substances, which provide the basis for the time and space control of biological processes, while peptides are generally hydrophilic in nature, which makes it difficult to cross compartments. Therefore, the penetration of peptides as well as absorption is often problematic. Regulating the physicochemical properties of peptides, such as substitution, modification, to increase their interaction with biofilms, can improve their penetrability. In Chapter 3, the factors affecting the permeability of polypeptides are discussed.

1.3 Helical Structures and Their Characterization

1.3.1 Different Types of Helices

1.3.1.1 α-Helix

A typical α-helix turn is composed of an average of 3.6 amino acid residues with dihedral angles (φ, ψ) in backbones close to −55° and −45°, respectively. As a rise of 1.5 Å/residue or 5.4 Å/turn, only $i+4$ and $i+7$ positions can make the side chains of a given residue at the i position and the other residue at the $i+n$ position are on the same face. α-Helices are stabilized by intramolecular $i \rightarrow i+4$ hydrogen bonds between a carbonyl group of the residue at position i and an amide proton at position $i+4$ in the main chain, with about 2.72 Å in the distance of nitrogen–oxygen and the side chains pointing away from the helix axis [54].

Protein–protein interactions are involved in lots of biological processes such as transcription, signal transduction, exocytosis, and so on [55]. α-Helix is the most abundant secondary structure motif in proteins, accounting for over 30% in nature. Meanwhile, α-helix is involved at interfaces of diverse protein–protein interactions, which was known for α-helix-mediated protein–protein interactions. For its significant proportion found in proteins' structures, it is not surprising to tell that α-helix is the most fundamental recognition motifs in diverse protein–protein interactions. According to the study of helical interfaces in protein–protein interactions based on the Protein Data Bank (PDB), about 13% multi-protein systems contained the helix interface, ranging from enzymatic activities to protein associations by classification of their functions, such as energy metabolism, protein synthesis, transcription, DNA binding, signaling, transport, immune system, and so on [56].

The structural characteristic of α-helix forces residues especially their side chains to extend out to the surrounding environment for selective and specific recognition, making it to be a template for designing small-molecule inhibitors or activators toward protein–protein interactions. The simplest system for α-helix-mediated protein–protein interactions between two proteins is that one partner binds to its partner protein by forming a short helical motif.

1.3.1.2 3_{10}-Helix

Besides classical α-helix and β-sheet conformations, the 3_{10}-helix is another important secondary structural motif occurring in natural proteins, which also plays significant roles in stabilizing proteins' conformations and maintaining their biological functions. Taylor first proposed the 3_{10}-helix structure in 1941. Since then this structure gained much attention and was studied fully [57]. The short name of 3_{10}-helix implies that the number of residues per turn is 3 and the number of atoms contained in each intramolecular hydrogen bond is 10, which indicates that 3_{10}-helix is more tightly packed than α-helix (also called 3.6_{13}-helix). The backbone torsion angles (φ, ψ) in 3_{10}-helices are approximately −60° and −30°, respectively, which are very close to that in α-helices ($\varphi = -55°$, $\psi = -45°$). However, 3_{10}-helices display significantly distinct hydrogen-bonding pattern of $i \rightarrow i+3$, while α-helices are stabilized by $i \rightarrow i+4$ intramolecular hydrogen bonds [58]. The 3_{10}-helix is

less stable than the α-helix because of its less favorable van der Waals energy and nonoptimal hydrogen bond geometry [59]. However, on account of the high structural similarity between the α-helix and 3_{10}-helix, it is proposed that the α-helix can be turned into the 3_{10}-helix when side chain interactions happen. Indeed, 3_{10}-helices are not rare and could be found in globular proteins like aconitase, dienelactone hydrolase, and phage T_4 lysozyme. Barlow and Thornton analyzed globular protein crystal structures in the database and suggested that at least 3.4% of the residues are involved in 3_{10}-helices. They also found that the location of 3_{10}-helices is often close to the N- or C-terminal of an α-helix [60]. Marshall et al. proved that Aib (α-aminoisobutyric acid or $C_{\alpha,\alpha}$-dimethyl-glycine) can promote the formation of 3_{10}-helices by calculations in 1971. Since then, many X-ray diffraction structures of peptides involving rich Aib indicate their structure preference of 3_{10}-helices [61]. It is worth noting that an α-helical peptide requests at least seven amino acid residues, while the formation of 3_{10}-helical peptides has no dependence on main chain length [62].

1.3.1.3 π-Helix

So far only three helix types α-, 3_{10}-, and π-helix were found in protein structures. Compared with α-helix (30%) and 3_{10}-helix (4%) in nature, π-helix seems particularly rare, which could be attributed to the instability of corresponding structures. To be specific, values of dihedral angles in π-helix were very close to the allowed minimum energy requirements indicated by Ramachandran plot and proven to be unfavorable [63]. Meanwhile, it was suggested that the required energy cost for stabilizing the intramolecular $i \rightarrow i+5$ hydrogen bond to form a helix was huge [64]. Therefore, many people believe that π-helices are unstable in nature. However, as researches on π-helix are moving forward, traditional concepts about π-helix are broke. Researchers found the formation of π-helices in molecular dynamics simulations of peptides [65]. More importantly, π-helices were observed in many protein structures. Most of naturally occurring π-helices contain at least seven amino acid residues and minimum two $i \rightarrow i+5$ hydrogen bonds and maximum seven H-bonds. Along with α-helix and 3_{10}-helix, π-helix can stably exist and may play important roles in maintaining lots of biological functions.

π-Helices are also called 4.4_{16}-helix where 4.4 is the number of residues in each turn and 16 is the number of atoms involved in a hydrogen bond [66]. π-Helices are stabilized by intramolecular $i \rightarrow i+5$ hydrogen bonds between a carbonyl group of the residue at position i and an amide proton at position $i+4$ in the main chain. α-Helices and 3_{10}-helices are stabilized by repeating $i \rightarrow i+4$ and $i \rightarrow i+3$ hydrogen bonds, respectively. Therefore, minimal number of residues in a single π-helix is one more than that in an α-helix and two more than that in a 3_{10}-helix. According to the structural analysis of π-helices in proteins in PDB, the mean values of dihedral angles (φ, ψ) observed in π-helices could be around $-76°$ and $-41°$, respectively. However, it could have slight distinctions according to different models on structure definition of π-helices. Besides, values of 1.2 Å in an average unit rise, 4.4 residues in each turn, and 83° in an average unit twist were observed in the helical geometry of π-helices. Like α-helices, π-helices have its featured amino acid preference in sequences. The

distributions of amino acid residues for π-helices showed that aromatic residues like Tyr, Trp, Phe, and His, as well as bulky aliphatic residues like Ile and Leu have higher propensities, while small amino acids like Ala, Gly, and Pro are less preferential. Also, there are amino acid residue preferences in their positions in sequences. For example, bulky residues such as Phe, Tyr, Trp, Ile, and Leu are more likely to be located at the beginning and at the end of π-helices. Besides hydrogen bond interactions, other factors facilitated the stabilization of π-helices. Compared with the α-helix, the π-helix had a lower unit rise (1.2 Å), whose side chains would be closer to each other in space. Therefore, other interactions between side chains such as the van der Waals, aromatic ring stacking, and electrostatic interactions became important contributors for the stabilization of π-helices. This is why aromatic and large aliphatic amino acids have higher propensities in π-helices [67].

It is worth noting that some researches revealed that there may be an evolutionary relationship between α-helices and π-helices. Based on the families of structurally similar proteins (FSSP) survey on all known π-helix-containing protein structures in databases [68], in 106 proteins with π-helices, 88 were found to exhibit the α-helix with one less amino acid residue, accounting for over 80%, which suggested that nature π-helices may originate from the insertion of one residue into the corresponding α-helices during evolution. Meanwhile, at least three residues could be found in designated α-helices in over 95% of the analyzed π-helices by FSSP method, which also suggested that there was a strong association between α-helices and π-helices [69]. The hypothesis on the originality of π-helices was further confirmed by the phenomenon of α-helix-to-π-helix conversion in some protein families. For example, in mercuric ion reductases, an α-helix-to-π-helix conversion, which was attributed to the insertion of a single residue compared to its ancient reductase member, occurred and put a catalytic Tyr residue into the binding site and triggered the Hg^{2+} detoxification by mercuric reductase [70].

Function of π-Helix Cellular function depends on highly specific interactions between biomolecules (proteins, RNA, DNA, and carbohydrates). A basic limitation of drug development is the inability of traditional "small-molecule" pharmaceuticals to specifically target large protein interfaces, many of which are desirable drug targets. α-Helices, ubiquitous elements of protein structures, play fundamental roles in many protein–protein interactions. Stable mimics of α-helices that can predictably disrupt these interactions would be invaluable as tools in chemical biology and as leads in drug discovery. There has been exciting progress in the molecular design of these protein domain mimetics and their remarkable potential to inhibit challenging interactions in the past decade. Key challenges in the field including identification of suitable targets and bioavailability of medium-sized molecules do not conform to empirical rules followed in traditional drug design. Stabilized α-helices avoid some of the strict limitations that have been placed on drug discovery. When designing potential drug candidates, medicinal chemists often adhere to the Lipinski rules, which stipulate that the molecular mass of a drug should not exceed 500 Da. Recent findings suggest that large synthetic α-helices can traffic into the cell and efficiently compete with cellular protein–protein