1

Introduction

1.1 Protein–Protein Interactions and Their Small-Molecule Modulators

1.1.1 Characteristics of Protein–Protein Interactions

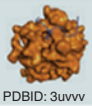
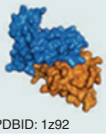
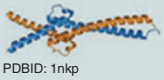
Proteins that work and degrade in highly congested and complex environments must be found by their partners in a large number of non-partners. It is estimated that human beings have 650 000 different pairs of interactions, which are responsible for a number of key biomolecular processes [1]. The surface 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.

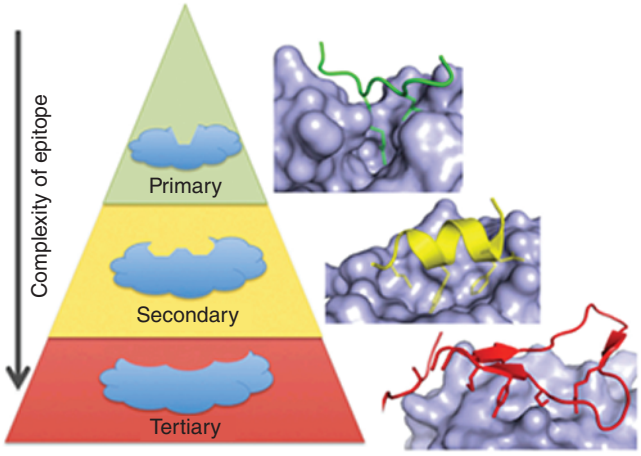


Figure 1.2 The complexity of the PPI interface affects druggability PPIs can be classified by whether one side of the interface consists of a primary (linear) protein sequence (green), a single region of secondary structure (such as an α-helix, yellow), or multiple sequences requiring tertiary structure (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

interactions, contrary to predictions based on the Lipinski rules. Although these molecules have undoubtedly proven their value as probes for decoding biological complexity, the next big question is whether these molecules can become therapeutics. This chapter discusses the applications of constrained helices, based on properties of protein–protein interactions.

Although π-helices had lower frequency observed in protein structure compared with other helix types, more and more results revealed that π-helices could exhibit distinct functional importance, especially in the chelatase family of proteins. For example, ferrochelatases from *Bacillus subtilis*, which catalyzes the insertion of Fe^{2+} into protoporphyrin to generate heme, contained a 10 amino acid residue-length π-helix.

This π-helix was located near the active site of the enzyme at the distance of 10 Å, whose residues were proven to have interactions with a hydrated magnesium complex in its bound state [71]. Similar as ferrochelatases from *B. subtilis* chelatase, the cobalt chelatase, which was responsible for the insertion of Co^{2+} into precorin-2 to generate cobalt-precorin, also possessed a π-helix of seven amino acid residue length [72]. Dioxygenase is another example, whose three histidine residues H494, H499, and H504 on the surface of the π-helix were involved in the formation of a metal binding site, which suggested that π-helices are the unique structural element for metal binding enzymes.

Besides the chelatase family, π-helices contribute to diverse protein functions. For example, an π-helix conformation was found in the hydroxylase component of methane monooxygenase hydroxylase (MMOH), which participates in the binding of a product analog 6-bromohexan-1-ol at the active site. The long π-helix contains two π-helices named piB and piD, respectively. In the process of ligand binding, a peristaltic shift of piD in an esophageal peristalsis-like manner was performed in the C-terminal direction [73]. Similar phenomenon of π-helical shift could be observed in the related toluene-4-mono-oxygenase (ToMO) hydroxylase structure during its binding of the regulatory component (ToMOD) [74], which may provide a new insight into mechanisms for the activation of bacterial multicomponent monooxygenases (BMMs) by their regulatory subunits [69].

1.3.2 Characterization of Helical Peptides

The study of the molecular structure can give fine details about the interface that enables the interaction between proteins. The molecular structures of many protein complexes have been unlocked by the technique of X-ray crystallography [75]. Later, NMR also started to be applied with the aim of unraveling the molecular structure of protein complexes and it is advantageous for characterizing weak protein–protein interactions [76].

1.3.2.1 Circular Dichroism

Circular dichroism (CD) is a simple and convenient method to analyze secondary structures of proteins and peptides. Only chiral molecules can exhibit featured absorption bands in their CD spectra while randomly oriented systems have no CD intensity. This is why proteins and peptides, which are composed of chiral

amino acid residues, show characteristic CD absorption according to their different secondary structures. Because of the sensitivity and convenience of CD measurement, CD has become a routine tool in many applications, especially in the study of determining the helical content of proteins and peptides [77]. Generally speaking, proteins and peptides in random or denatured structures usually have very weak signals in their CD spectra, while proteins and peptides with well-defined and stable conformations such as α-helix can give significant CD intensities and characteristics in CD spectra. The featured structural characteristics in α-helix, including $i/i+4$ hydrogen-bonding pattern, 3.6 amino acid residues per turn, 100° turn and 1.5 Å translation between two peptide units, make α-helix be distinguished from other structural motifs in proteins and peptides, which exhibits distinct band absorptions in CD spectra. Absorptions at different wavelengths indicate different kinds of electron transition such as $n \rightarrow \pi^*$, $\pi \rightarrow \pi^*$, $n \rightarrow \sigma^*$ occurring at 220, 207–190, and 175 nm in CD signals, respectively. The α-helices display two separate negative maximum signals at both 222 nm (the $n \rightarrow \pi^*$ transition) and 208 nm (part of the $\pi \rightarrow \pi^*$ transition) as well as another positive signal at 195 nm (part of the $\pi \rightarrow \pi^*$ transition). Compared with α-helices' featured absorptions, the other common structural unit β-sheets show a negative band at about 220 nm with another positive band at about 200 nm in the corresponding CD spectra. It is worth noting that the prediction of α-helices in proteins and peptides by a CD tool is very sensitive and has very high accuracy.

1.3.2.2 X-ray Crystallography

X-ray crystallography is a technique used for determining the atomic and molecular structure of a crystal, in which the crystalline atoms cause a beam of incident X-rays to diffract into many specific directions. The method has revealed the structure and function of many biological molecules, including vitamins, drugs, proteins, and nucleic acids. The crystal structures of proteins were first reported in the late 1950s, beginning with the structure of sperm whale myoglobin. X-ray crystallography is now used routinely by scientists to determine how a pharmaceutical drug interacts with its protein target and to unveil the detailed interface of protein–protein interactions [78]. Navitoclax is an inhibitor of both BCL-2 and BCL-X_L, but the concomitant on-target thrombocytopenia caused by BCL-X_L inhibition limits the efficacy achievable with this agent. Based on the crystal structure of this compound in complex with BCL-2 family proteins, Andrew J. Souers et al. re-engineered and optimized navitoclax to create a highly potent, orally bioavailable and BCL-2-selective inhibitor ABT-199, which showed potent inhibition activity on BCL-2-dependent tumors in vivo and spares human platelets [79].

The technique of single-crystal X-ray crystallography has three basic steps. The first step is to obtain an adequate crystal of the molecule under study. The crystal should be large enough (typically larger than 0.1 mm in all dimensions), regular in structure, and pure in composition, with no significant internal imperfections such as cracks or twinning, and this is often the most difficult step. In the second step, the crystal is illuminated with a finely focused monochromatic beam of X-rays, producing a diffraction pattern of regularly spaced spots known as reflections. As the

crystal is gradually rotated, previous reflections disappear, and new ones appear, the intensity of every spot is recorded at every orientation of the crystal. In the third step, the 2D images taken at different orientations are converted into a 3D model of the density of electrons within the crystal using the mathematical method of Fourier transforms; these data are combined computationally with complementary chemical information to produce and refine a model of the arrangement of atoms within the crystal. The crystal structures could give essential details about the structures of a molecule or macromolecular complex.

1.3.2.3 Nuclear Magnetic Resonance (NMR)

NMR is another fast and powerful method to determine conformations of proteins and peptides in solution. In 1D and 2D NMR spectroscopies, nuclear overhauser effects (NOEs), temperature dependence of the NMR spectrum, and coupling constants are three important indicators for the determination of peptides' secondary structures [80]. Besides, measurements of the chemical shift index and the rate of exchange with deuterons are also two regular methods for the conformation analysis of peptides.

There is a very close correlation between NOEs and the distance between protons. Generally speaking, NOE signals between two protons can arise when their distance is under 5 Å in space, indicating strong evidences of secondary structures [81]. The cross-peaks in a 2D scalar-correlated (COSY) spectrum indicate the coupling patterns for the resonances assigned to specific protons of amino acid residues. Specifically, sequential connectivities and NH–NH as well as NH–CH cross-peaks in 2D-ROESY spectra indicate key medium- and long-range ROEs, which strongly supports the identification of secondary structures. Some representative nonsequential medium cross-peaks such as $d_{\alpha N}(i, i+4)$, $d_{\alpha N}(i, i+3)$, and $d_{\alpha\beta}(i, i+3)$ in ROESY spectra support the adoption of helical conformations [82].

The temperature dependence of amide proton resonances is utilized for determining whether the backbone amino acid residues are involved in the intramolecular hydrogen-bonding formation. Generally speaking, amide proton resonances exhibit higher temperature coefficients with increasing temperature, which suggests no involvement of hydrogen bonds and the complete exposition to solvents. A value of temperature coefficient (< -4.0 ppb/K) indicates the specific amide proton's participation in the hydrogen-bonding in H_2O, which is so critical for the stabilization of an α-helix. Values of temperature coefficient for indicating the hydrogen-bonding are different in various hydrogen-bonding solvents. Compared with H_2O, temperature gradients for the chemical shift of the NH-signal is considered to be under -3 ppb/K in dimethyl sulfoxide (DMSO).

Vicinal coupling constant ^3J has a close relationship with dihedral angle originated from the Karplus equations and is very important to analyze conformations of proteins and peptides. Among these different types of coupling, the homonuclear $^3J_{HN\mathrm{C}\alpha H}$ is easy to be measured and thus particularly important. However, according to the Karplus curve, there is no one-to-one correspondence between measured coupling constants and bond angles in spite of the dependence of $^3J_{HN\mathrm{C}\alpha H}$ on the angle φ. Indeed, in most cases, one coupling constant can respond to four different bond

angles φ. Therefore, it is ambiguous and hard to achieve compelling conclusions from ${}^{3}J_{HNC\alpha H}$ measurements, thus viewed as a complement approach for analyzing conformations of peptides based on the bond angle φ [83]. It is generally acknowledged that the coupling constant ${}^{3}J_{HNC\alpha H}$ under 6 Hz for amino acid residues indicates the adoption of helical conformations.

Besides, the chemical shift is an equally important index for the determination of secondary structures of peptides and proteins. It is found that ^{1}H NMR chemical shifts have a very close connection with the character of secondary structures. Compared with the random coil value, the ^{1}H NMR chemical shift of the α-CH proton of all natural 20 amino acids shows an upfield shift when they are involved in the helical conformation. Conversely, a downfield shift of the corresponding ^{1}H NMR chemical shift is observed when they adopt the β-strand conformations. Therefore, these observations of ^{1}H NMR chemical shift characters are very helpful to determine the location, as well as compare contents of different secondary structural motifs in proteins and peptides. Compared with traditional NOE-based methods for secondary structure determination, this method simply requests the measurement of α-CH ^{1}H resonance assignments and was proven to be accurate and useful by a lot of examples [84].

Finally, the rate of exchange with deuterons like added D_2O is a convenient method for measuring molecular dynamics of peptides. Specifically, the exchange of amide protons with the deuteron solvent can tell whether these protons are involved in forming intramolecular hydrogen bonds by measuring the corresponding exchange rate. Generally speaking, the amide proton participating in intramolecular hydrogen bonds is expected to exhibit lower exchange rate in deuteron solvents while those protons absent from hydrogen bonds are exchanged more rapidly [85].

1.4 Stabilization of Peptides

This section describes the methods of stabilizing peptides, including the cyclic and main alteration.

1.4.1 Peptide Stabilization via Cyclization

1.4.1.1 Monocyclization

Macro cyclization uses additional covalent bonds to limit the distance between two points. The cyclic increase of peptide resistance to protease reduces the exposure of hydrogen bond donors or receptors to regulate the physical and chemical properties of peptides.

Disulfide bonds sensitive to redox reactions are common in nature. The specific pattern and length of the artificial disulfide bond can be used to stabilize the helix, hairpin, or loop. Their reversible properties can be used to study the folding of peptides or the uptake of cells [86]. At the same time, the biological system can directly evolve disulfide bonds. Similarly, the hydrophilic amide or ester bond is widely used

in the construction of peptide analog as a bridge connecting the natural clock. In particular, the lysine-aspartic acid stable helix shows the highest helicity in the water.

As a kind of high stability bridge connecting mode, the thiol ether bond is often used as the substitution of disulfide bonds to restrain peptide conformation. The thiol ether bond can be constructed by using a nucleophilic substitution reaction of halides or by using a free radical reaction of unsaturated hydrocarbons. In addition, compared with a component binding strategy, the two-component stabilization strategies can give the peptide with special characteristics, such as reversibility and modifying potentials [87]. Like disulfide, the two-ingredient strategy based on cysteine can be used in micro-proteins or combined with a screening system.

As a nonpolar connection, the all-hydrocarbon strategy increases the hydrophobicity of peptides. The stapled polypeptide, the all-hydrocarbon side chain stabilized helix, is constructed by ring-closing metathesis. Various types of stapling strategies have been developed and widely used. In addition, all hydrocarbon chains can be used in replace of hydrogen bond as an N-terminal template. In addition, azoles, constructed using bio-orthogonal reactions, can be used to stabilize the peptide, including the helix and hairpin structure [88]. In one-component way, the ring addition can derive a fluorescent bridge. While for two-component strategy, the functional stapled polypeptide can be established to further label [89].

1.4.1.2 Multicyclization

Increasing the number of polypeptide rings, such as construction of bicyclic or even polycyclic peptides, can further improve the therapeutic performance of peptides and even show the oral dosing potentials. There are many polycyclic peptides in nature, and in the process of studying these polycyclic peptides, many valuable techniques and molecules have been obtained. Here we will focus on the polycyclic polypeptide in nature. There are many bioactive polycyclic peptides in nature, including bacterial secretions, such as lantibiotics, toxic peptides, such as amatoxin and phallotoxin, and peptide toxins of animal secretions, such as α-toxins.

Lantibiotics is an important polypeptide antibiotic that is found and produced by a large number of Gram-positive bacteria (including *Streptococcus* and *Streptomyces*) and attacks other Gram-positive bacteria. The peptide contains the characteristics of multicyclic lanthionine or methyllanthionine and some unsaturated amino acid dehydroalanine and 2-aminoisobutyric acid as the basic building blocks. Lanthionine, abbreviation for lanthionine-containing peptide antibiotics, is one of the important building blocks of lantibiotics, consisting of two alanine residues, which are connected by sulfur to their β-carbon atoms [90]. Most of the lantibiotics are considered to be a class of bacteriocins synthesized by ribosome biosynthesis, with a pilot polypeptide sequence that is removed when the molecule is transported out of the synthetic cell. There are four enzymes involved in the construction of lanthionine rings. This process differs from most of the other natural antibiotics we have previously known [91]. Lipid II, a glycoprotein of this gram-positive bacterium, is the target for a large number of lantibiotics. Some lantibiotics causes destructive pores or inhibits the biosynthesis of peptides. Nisin is a lantibiotics widely used in today's commercial preservatives.

Cyclotides are a class of bioactive mini-proteins from plants that have the unique topological feature of a head-to-tail cyclic backbone combined with intramolecular disulfide bonds [92]. Because of this structure, they are ultra-stable against degradation at elevated temperatures or in the presence of proteolytic enzymes and have attracted interest as peptide-based templates for drug design. Their natural function in plants is acting as insecticidal agents, which provides potential applications in agriculture. Furthermore, they have a range of pharmaceutically relevant activities, including anti-HIV, antimicrobial, and uterotonic activities. Their exceptional stability and facile synthesis make them pharmaceutical templates that can be grafted into peptides with desired bioactivities.

Several poisonous peptides were found in several genera of poisonous mushrooms, such as phallotoxin and amatoxin [93]. Unlike many ingested poisons, phallotoxin and amatoxin are not destroyed by heat, so cooking poisonous mushrooms is no less lethal. Phallotoxins consists of at least seven amino acids, which are also called death cap mushrooms from the goose mushrooms. Amatoxins is present in the poisonous death cap and goose mushrooms, consisting of eight amino acid residues. This compound has a similar structure, with eight amino acid residues arranged in a conserved double-ring skeleton. Snake venom α-toxin (α-BTX) is a neurotoxin, which is a multicyclic peptide that is known to be competitive binding to the acetylcholine receptor (nachrs), which then causes paralysis of the victim, respiratory failure, and death. α-Toxin has 74 amino acids with 52 sulfide bridges. Compared with other snake venom α-toxin, it has a triple-fingered fold structure. Hydrogen bonds keep the second and third loops roughly parallel to allow parallel β-sheets. The three-fingered structure is conserved by four of two sulfide bridges, and the number of these bonds and secondary structures is responsible for the stability of the neurotoxin, which is not susceptible to degeneration and is proven to be resistant to boiling and low-pH environments [94]. On the basis of detailed studies of the structure and properties of these toxins, the researchers used these toxins as templates to embed other functional peptides into the template and to obtain a stable polypeptide with an interesting function.

In addition to the presence of polycyclic polypeptides in nature, researchers have developed a number of artificial techniques based on polycyclic peptides. Poly cyclization can enhance the effects of cyclization on the physical and chemical properties of peptides, such as the ability of researchers to obtain highly stable versatile peptides or peptides with better penetration. Existing natural peptides, mostly knottins and cyclotides, have been modified to derive peptides with new recognizing abilities (Figure 1.3). Cysteine knot structures, which is a structural motif comprising three disulfide bridges, with new molecular recognition properties could be engineered either by molecular grafting of peptide epitopes or by directed evolution strategies [96]. By replacing existing amino acids, cysteine knot-based protease inhibitors were converted into inhibitors of homologous proteases [97]. In addition, polycyclic polypeptide can be directly screened for evolution. In addition to the stable polycyclic polypeptide of pure disulfide, as mentioned above, on-phage peptide alkylation is able to realize the screening of a variety of two-ring peptide libraries, which can generate two-ring peptide ligands with good binding affinity

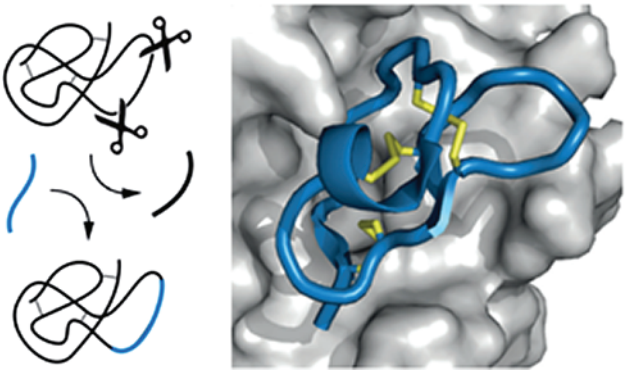


Figure 1.3 Epitope grafting in polycyclic peptides. (a) Schematic depiction of the approach of epitope grafting. A cysteine knot protein is chemically synthesized or recombinantly expressed with a peptide epitope inserted into one peptide loop. (b) Crystal structure of porcine pancreatic elastase (PPE) in complex with the engineered trypsin inhibitor EETI-II. The trypsin binding loop of the 28-amino acid inhibitor was substituted by a peptide sequence derived from the third domain of turkey ovomucoid inhibitor that was optimized to inhibit PPE (PDB entry: 1QNJ) [95]. Source: Baeriswyl and Heinis [95]. Reproduced with permission of Wiley.

[48a]. Finally, poly cyclic peptides could be obtained through rational design. Pei and coworkers used bicyclic polypeptide as a scaffold to combine cyclic targeting peptide and cyclic cell penetrating peptides [98]. The derived two-ring peptide has cellular permeability and retains the ability to identify intracellular targets. For stapled polypeptides, the addition of a hydrophobic ring can further improve the physical and chemical properties of peptides. Furthermore, the stability and permeability of the poly cyclic polypeptides are higher than those of the single-ring peptides.

1.4.2 Peptide Stabilization via Backbone Reconstruction

1.4.2.1 Methylation

Methylation can affect the physicochemical properties of peptides by dihedral immobilization and hydrogen-bonded donor blockage. The main form of N-methylation in nature is N-methylation in the polypeptide skeleton or side chain of lysine or arginine with free amino group. For natural N-methylated peptides, *N*-methyl is mainly in the trunk, especially the cyclic peptide. Many biological functional peptides are N-methylated peptides, which are often observed in non-ribosomal peptides. The N-methylation of lysine and arginine mainly occurs on functional proteins, which are not discussed here.

First, from a molecular point of view, the N-methylation of NH on the polypeptide skeleton blocks the potential hydrogen donor of NH, and if NH participates in the intramolecular hydrogen bond, it will have a great effect on the peptide's secondary structure. This effect is particularly important for long peptides, which introduce H-bonds to build peptide structures such as α-helix and β-slices. For

example, Kapurniotu and coworkers utilized N-methylation into β-sheet peptides to avoid its aggregation, which is assembled to form a fiber polymer [99]. However, N-methylation does not destroy the intramolecular hydrogen bond of peptides, so the additional *N*-methyl can help to stabilize the peptide from external disturbances and further optimize the biophysical properties of the polypeptide.

Second, the introduction of N-methylation can increase the resistance and stiffness of peptides. The addition of N-methylation on the polypeptide skeleton will strongly reduce the cis–trans equilibrium of the *N*-methyl amide bond [100]. In addition, as the 3D conformation of the whole peptide, the steric resistance of *N*-methyl substituted will further affect the adjacent residues. For example, natural n-methylated amino acid, proline, the only amino acid where secondary amine and the ring structure involved the backbone atom, plays an important role in protein folding. PRO provides CIS amide bonds, which are usually used for hairpin induction templates. The improvement of steric resistance and stiffness can further improve the resistance to protease degradation and can maintain biological function for a long time [101].

In addition, N-methylated peptide has a significant increase in cell permeability and oral bioavailability. This exciting improvement may come from increasing the hydrophobicity of the polypeptide skeleton and enhancing the interaction with the lipid layer. In fact, most natural n-methylated peptides have high oral bioavailability and are widely used as promising candidates for drug use. A typical n-methylated cyclic peptide, cyclosporin A, shows 19% oral bioavailability. This improvement is especially important for multiple N-methylation. Hoffman and coworkers have established a multiple *N*-methyl library, with different *N*-methyl substitutions on the cyclic hexa peptides. Caco-2 cells and parallel artificial membrane permeability assay (PAMPA) assay disclosed intestinal permeability of the methylated peptide, which can be further applied to the design of oral peptide therapy [102].

In addition, the addition of N-methylation can regulate the biological activity of peptides, such as binding affinity. For peptide inhibitor design, peptides usually bind to protein surfaces with certain conformation. The target's functional domain is typically specific, such as hydrophobicity. Therefore, the interaction of peptide target binding affinity can be improved by the regulation of the conformation by N-methylation and enhanced hydrophobicity.

According to the advantages of N-methylation, the researchers have done a great deal of design and exploration of N-methylation to improve the biophysical function of peptides.

Further, in addition to N-methylation, there is more general type of n-modified peptide skeleton that can modulate and optimize the biophysical properties of peptides called N-alkylation. In particular, the N-alkylation of polypeptide skeletons usually results in large changes in peptide structure. Therefore, in general, N-alkylation can be used as another kind of peptoid. With the use of different functional groups, N-alkylation seems to be more promising than single methylation. However, due to the difficulty of synthesis, the application of *N*-alkyl is mainly restricted. But N-alkylation peptides also show promising biological functions, such as antibacterial activity.

In the position of methylation, in addition to N-methylation, another methylation method, α-methylation, can also be considered a design strategy. Similarly, α-methylation can restrict dihedral angle by steric hindrance, which facilitates the formation of turns or helices, as is shown by stapled peptides, where the bridging module is α-methyl. In addition, Aib is often used in the induction of hairpin structures.

1.4.2.2 Foldamers

For a long time, researchers have wondered whether there are other skeletons, including modules that have not been chosen by biological evolution, and may have the ability to support identification, catalysis, or assembly activities while showing better applicability. The foldamer is based on the skeleton of de novo design, the conformational order of the epitope simulator. Since many of these activities seem to require precise spatial positioning of the side chains, foldamer studies tend to begin with the determination of a class of low polymer with a specific shape.

The foldamers can be divided into two types based on the frame selection of monomer unit. "Aliphatic" foldamers have a saturated carbon chain separating amide or urea groups, and the use of aromatic septal skeletons, such as the multiple pyrrole/imidazole DNA binding oligomer [103]. The foldamers of the aromatic skeleton is closer to the small molecule and thus has better penetrability. Initial monomer selection is usually influenced by their synthesis and the ease of structural characterization. In addition, systems containing identical monomer units are called "homogeneous" foldamers, while "heterogeneous" foldamers contain more than one type of subunit [104].

D-type amino acid is a kind of amino acid which is opposite to nature in α-carbon configuration. Examples of D-type amino acids in nature include opiates and antimicrobial peptides from frog skins, snail neuropeptides, shellfish hormones, and spider venom. These D-amino acids form when L-amino acids change after translation.

Molecular recognition in nature, such as the identification between enzymes and substrates, antibodies and antigens, is configuration specific. Although the presence of D-amino acids in nature, most proteins are made of L-amino acids, the rare but unrecognized D-amino acid insertion makes it difficult for polypeptides to be identified by shear enzymes or antibodies, which can be used to increase stability or to reduce the immunogenicity of peptides. On the other hand, it is also important to ensure the original activity of peptides while improving the stability or immunogenicity of polypeptide. Some special strategies have been developed, such as retro inverso peptide. The direct evolution of the D peptide sequence is also achieved through the image phage display of D protein, which enables us to obtain D-peptide with good stability and affinity [105].

Some peptides containing D-amino acids are biologically more powerful. The ω-agatoxins IVB and IVC peptides that contain D-serine showed higher inhibition of P-type calcium channel compared with those containing L-serine isomers [106].

As is known to all, β-hairpin is an important secondary conformation, which can be used to achieve the biological process control. Correspondingly, scientists have developed a number of ways to stabilize the structure of β-hairpins, mainly

including the nucleation of the turn and the hairpin [107]. As already mentioned in the preceding article, the induction nucleus requires a special dihedral module. Dihedral preferences of D-amino acids are different so that they lead to changes in the conformation of the peptide, which can be used to construct a peptide with a β-turn or β-hairpin [108]. The scientists found that turn position $i+1$ D-amino acid was used to promote the type II β-turn, thus supporting the formation of β-hairpins. Studies have shown that D-proline can be used as an auxiliary factor for β-turn [109]. Templates such as D-pro-l-pro and D-pro-gly dipeptide fragment are privileged peptides widely used to stabilize parallel β-hairpin, while D-proline Dadme (1,2-diamino-1,1-dimethylethane) provides parallel β-hairpin [110].

β-Amino acid is another kind of amino acid in nature, which are frequently found to be important components of bioactive natural products, such as Taxol, bleomycin, and microcystin [111]. Adding β-amino acids to L-amino acids in natural products produces unique structures with similar molecular polarity. In addition, similar to D-amino acids, difficulty to recognize the β-amino acid peptide by protease renders the peptide greater stability.

It is noteworthy that the use of β-amino acids in order to strengthen the β-sheet secondary structure has proved to be extremely challenging due to the severe mismatch between structural characteristics of α- and β-peptide. Because β-amino acids have no preference for dihedral angle, which leads to a lower conformation change and a higher flexibility. A single permutation mainly results in the partial stretching of the structure, and the α,α-di-peptide permutation only provides a modest result [112].

Recent studies have demonstrated that multiple a → b replacements at carefully selected positions are able to generate potent hormone analogs with in vivo activities. Cheloha et al. evaluated α/β-peptide analogs of parathyroid hormone [PTH (1–34)] that contained exclusively b3 residues in an aaab pattern (Figure 1.4) [114]. Johnson et al. examined α/β-peptide analogs of GLP-1, the activity of which could be rescued by using ring-constrained b residues [115]. An analog containing five cyclic b residues along with two Aib residues displayed high resistance to cleavage by dipeptidyl peptidase-4 or neprilysin, the two in vivo GLP-1 degrading proteases.

As an early example, Gellman and coworkers reported the substitution of β-amino acids in the phage-display-derived peptide, with hot residues in the last round of an α-helix and the following turn structures [116]. The final peptide with increased protease resistance proved the feasibility of the turn mimetic.

β-Amino acid insertion is a relatively less noticeable strategy that can be combined with more common side chain cross-linking to produce synergistic effects such as α/β-peptides based on the stapled BH3α-peptide, which contains a hydrocarbon cross-linking to improve α-helical stability [117]. A fixed α/β-peptide can mimic the structure and function of the mother's α-peptide in its ability to enter certain types of cells and block protein–protein interactions associated with apoptotic signals. However, the α/β-polypeptide is nearly 100 times more resistant to hydrolysis than parental stapled peptide. Similarly, β-amino acid insertion can improve the stability of peptide stability by hydrogen-bonding surrogate strategies [118].

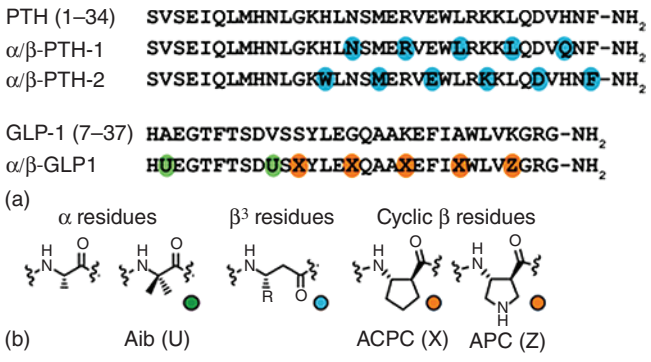


Figure 1.4 α/β-Peptide mimicry of G protein-coupled receptor agonists. (a) Primary sequences of PTH (1–34), GLP-1 (7–37), and biologically active α/β-peptide analogs. Colored circles indicate nonnatural residues, as indicated in (b). (b) Structures of a generic residue, the Aib residue (green), a generic β3 residue (blue), and cyclic β-residues ACPC and APC (orange) [113]. Source: Checco and Gellman [113]. © 2016, Elsevier.

The effects of β-residues replacement on the structure and stability of small proteins were studied by Kreitler et al. [119]. First, they evaluated the effect of α → β modification on the structure and stability of small-study villin-helmet subdomains. The original state of these 35 residual poly peptides consists of several α-helical segments wrapped around a small hydrophobic core. After that they examined the α → β replacement in four solvent exposure positions. In each case, the natural α-residue of the β3 homologs and the cyclic β-residues were evaluated. According to the variable temperature CD spectrum, all α → β3 substitutions result in severe instability of the tertiary structure, although in these locations, the replacement of β3 residues with cyclic β-residues improves stability. These findings contribute to a basic α-/β-peptide knowledge base that confirms that β3 amino acid residues can be used as effective homologous α-amino acid residues in structural simulations in natural tertiary structures, which support the rational design of the function of natural peptides α/β-analogs.

Foldamer based on heterogeneous backbone has some benefits relative to only relying on homogeneous skeleton. Heterogeneous methods allow many different combinations, which provide a potentially unique way to extend the side chain in space. By mimicking proteins, valuable foldamer activity is likely to depend on the placement of a particular 3D functional group; therefore, we can generate a molecule of clearer shape through foldamers, which is more likely to achieve any particular activity better. A variety of complementary scaffolds may be required to produce a wide range of foldamer functions, but the diversity of the skeleton is not sufficient to achieve this goal; people must also be able to decorate skeletons with different side chains. The heterogeneous backbone method can greatly promote the generation of foldamer sets with extensive side chain diversity.

The progress of folding structure synthesis technology, as well as the possible folding of the backbone in water, have opened up the way for the selective foldamer–biomolecular interaction and foldamers interfering biological function

[120]. Foldamers set many features in one, so that it can better function with biomolecules: medium size (MW = 500–5000 g/mol) and large contact areas are suitable for binding in the extended contact surface of protein, folding predictability, adjustable type, diversity, and expected resistance to protein hydrolysis. Because they are structurally well defined, foldamers can be used as scaffolding to accurately project a combination module in space. Some early work focused on the design of cationic amphiphilic foldamers, mimicking host–defense peptides and selectively disrupting bacterial membranes. A new and challenging foldamer application mimics the discovery of folding peptide fragments in proteins, particularly α-helices, to interfere with protein–protein interactions.

Arora et al. have developed oxopiperazine helix mimetics (OHMs) for modulating α-helical domain of HIF-1a-mediated protein–protein interactions [121] (Figure 1.5). These oligomers downregulated hypoxia-induced gene expression and reduced tumor volume in mice bearing MDA-MB-231 xenografts. Wilson and coworkers have reported significant recent advances in the development of aromatic oligoamide mimics of short helices containing *N*-alkylated *para*-aminobenzoate units, which was pioneered by Hamilton and coworkers [124]. The three alkyl groups were intended to reproduce the presentation of three spatially consecutive side chains along the side of an α-helix. Mimics that engage the p53-recognition cleft on HDM2 or the BH3-recognition cleft on Mcl-1 showed cell-membrane-penetrating activities.

Self-assembly has become a very effective method to produce large supramolecular containers with molecular recognition properties. Such containers may be

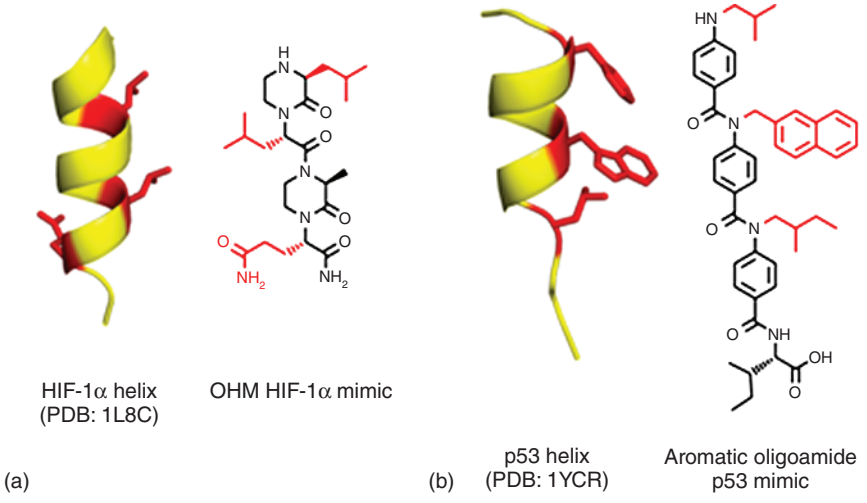


Figure 1.5 α-Helix-mimetic polyamides. (a) Cartoon representation of a portion of the HIF-1a α-helix (left, PDB: 1L8C) and drawing of an OHM mimic of this helix (right) [122]. Critical protein-contacting side chains are highlighted in red. (b) Cartoon representation of a portion of the p53 α-helix (left, PDB: 1YCR) and drawing of an aromatic oligoamide mimic (right) [123]. Critical protein-contacting side chains are highlighted in red. Source: Checco and Gellman [113]. Reproduced with permission of Wiley.

based on very simple building blocks and usually have high symmetry. Similarly, foldamers opens up new avenues for receptor design. The identification may occur on the foldamer surface or in the cavity of its folding structure. An important class of foldamer receptors includes a wide helix with a cavity. The groundbreaking work of Moore and coworkers shows the combination of hydrophobic objects in oligo-phenylene-ethynylene cavities [125]. Li, and coworkers reported that saccharides can be bound into aromatic oligomers [126]. These helical receptors are chiral in nature and can eventually discriminate between different enantiomers.

From the material point of view, foldamers can form materials with morphological characteristics on nano- or microscale by controlling the self-assembly of molecules. It is reported that some aliphatic and aromatic foldamers spontaneously form nano-fibers and nanoparticles [127]. For example, the hydrazide-based aromatic foldamers containing a long aliphatic side chain exhibits a two-mode assembly, forming vesicles in polar solvents and forming entangled fibers and gels in hydrocarbons. Several of the 14-helix β-decapeptides with different combinations of the lipophilic and hydrophilic side chains are assembled to form a lyotropic liquid crystal phases in the aqueous solution.

References

1 Stumpf, M.P., Thorne, T., de Silva, E. et al. (2008). *Proceedings of the National Academy of Sciences of the United States of America* 105: 6959–6964.

2 (a) Hunter, T. (2012). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367: 2513–2516. (b) Wolfenson, H., Lavelin, I., and Geiger, B. (2013). *Developmental Cell* 24: 447–458. (c) Rawlings, N.D., Tolle, D.P., and Barrett, A.J. (2004). *The Biochemical Journal* 378: 705–716.

3 (a) Ofran, Y. and Rost, B. (2003). *Journal of Molecular Biology* 325: 377–387. (b) Scott, D.E., Bayly, A.R., Abell, C., and Skidmore, J. (2016). *Nature Reviews. Drug Discovery* 15: 533–550.

4 Bogan, A.A. and Thorn, K.S. (1998). *Journal of Molecular Biology* 280: 1–9.

5 Keskin, O., Ma, B., and Nussinov, R. (2005). *Journal of Molecular Biology* 345: 1281–1294.

6 Crowley, P.B. and Golovin, A. (2005). *Proteins* 59: 231–239.

7 Jochim, A.L. and Arora, P.S. (2010). *ACS Chemical Biology* 5: 919–923.

8 (a) Guharoy, M. and Chakrabarti, P. (2007). *Bioinformatics*, vol. 23, 1909–1918. Oxford, England: Oxford University Press. (b) Jones, S. and Thornton, J.M. (1995). *Progress in Biophysics and Molecular Biology* 63: 31–65. (c) Kossiakoff, A.A. and Koide, S. (2008). *Current Opinion in Structural Biology* 18: 499–506.

9 (a) Smith, M.C. and Gestwicki, J.E. (2012). *Expert Reviews in Molecular Medicine* 14: e16. (b) Cheng, A.C., Coleman, R.G., Smyth, K.T. et al. (2007). *Nature Biotechnology* 25: 71–75.

10 Arkin, M.R., Tang, Y., and Wells, J.A. (2014). *Chemistry & Biology* 21: 1102–1114.

11 (a) Basse, M.J., Betzi, S., Bourgeas, R. et al. (2012). *Nucleic Acids Research* 41: D824–D827.

12 (a) Petta, I., Lievens, S., Libert, C. et al. (2016). *Molecular Therapy* 24: 707–718. (b) Gurard-Levin, Z.A., Scholle, M.D., Eisenberg, A.H., and Mrksich, M. (2011). *ACS Combinatorial Science* 13: 347–350.

13 Winter, A., Higueruelo, A.P., Marsh, M. et al. (2012). *Quarterly Reviews of Biophysics* 45: 383–426.

14 Meireles, L.M.C., Domling, A.S., and Camacho, C.J. (2010). *Nucleic Acids Research* 38: W407–W411.

15 Higueruelo, A.P., Jubb, H., and Blundell, T.L. (2013). *Current Opinion in Pharmacology* 13: 791–796.

16 Navratilova, I. and Hopkins, A.L. (2011). *Future Medicinal Chemistry* 3: 1809–1820.

17 Hajduk, P.J. and Greer, J. (2007). *Nature Reviews Drug Discovery* 6: 211–219.

18 Lionta, E., Spyrou, G., Vassilatis, D.K., and Cournia, Z. (2014). *Current Topics in Medicinal Chemistry* 14: 1923–1938.

19 Ford, M.L. and Larsen, C.P. (2009). *Immunological Reviews* 229: 294–306.

20 Schwab, N., Ulzheimer, J.C., Fox, R.J. et al. (2012). *Neurology* 78: 458–467.

21 (a) Gadek, T.R., Burdick, D.J., McDowell, R.S. et al. (2002). *Science (New York, NY)* 295: 1086–1089. (b) Zhong, M., Gadek, T.R., Bui, M. et al. (2012). *ACS Medicinal Chemistry Letters* 3: 203–206.

22 (a) Keating, S.M., Clark, K.R., Stefanich, L.D. et al. (2006). *Protein Science: A Publication of the Protein Society* 15: 290–303. (b) Shimaoka, M., Salas, A., Yang, W. et al. (2003). *Immunity* 19: 391–402.

23 Dubrez, L., Berthelet, J., and Glorian, V. (2013). *Oncotargets and Therapy* 6: 1285–1304.

24 Crisostomo, F.R.P., Feng, Y.M., Zhu, X.J. et al. (2009). *Bioorganic & Medicinal Chemistry Letters* 19: 6413–6418.

25 Sun, H.Y., Nikolovska-Coleska, Z., Yang, C.Y. et al. (2008). *Accounts of Chemical Research* 41: 1264–1277.

26 Filippakopoulos, P. and Knapp, S. (2014). *Nature Reviews Drug Discovery* 13: 339–358.

27 Cherepanov, P., Ambrosio, A.L.B., Rahman, S. et al. (2005). *Proceedings of the National Academy of Sciences of the United States of America* 102: 17308–17313.

28 Christ, F., Voet, A., Marchand, A. et al. (2010). *Nature Chemical Biology* 6: 442–448.

29 Arkin, M.R. and Wells, J.A. (2004). *Nature Reviews Drug Discovery* 3: 301–317.

30 Oltersdorf, T., Elmore, S.W., Shoemaker, A.R. et al. (2005). *Nature* 435: 677–681.

31 Chi, S.W., Lee, S.H., Kim, D.H. et al. (2005). *The Journal of Biological Chemistry* 280: 38795–38802.

32 Vassilev, L.T., Vu, B.T., Graves, B. et al. (2004). *Science (New York, NY)* 303: 844–848.

33 Qian, Z., Upadhyaya, P., and Pei, D. (2015). *Methods in Molecular Biology (Clifton, NJ)* 1248: 39–53.

34 Lennard, K.R. and Tavassoli, A. (2014). *Chemistry (Weinheim an der Bergstrasse, Germany)* 20: 10608–10614.

35 Cull, M.G., Miller, J.F., and Schatz, P.J. (1992). *Proceedings of the National Academy of Sciences of the United States of America* 89: 1865–1869.

36 Hansson, M., Samuelson, P., Gunneriusson, E., and Stahl, S. (2001). *Combinatorial Chemistry & High Throughput Screening* 4: 171–184.

37 Boder, E.T. and Wittrup, K.D. (1997). *Nature Biotechnology* 15: 553–557.

38 (a) Smith, G.P. (1985). *Science (New York, NY)* 228: 1315–1317. (b) Krumpe, L.R. and Mori, T. (2007). *Expert Opinion on Drug Discovery* 2: 525.

39 Mattheakis, L.C., Bhatt, R.R., and Dower, W.J. (1994). *Proceedings of the National Academy of Sciences of the United States of America* 91: 9022–9026.

40 Roberts, R.W. and Szostak, J.W. (1997). *Proceedings of the National Academy of Sciences of the United States of America* 94: 12297–12302.

41 Yamagishi, Y., Shoji, I., Miyagawa, S. et al. (2011). *Chemistry & Biology* 18: 1562–1570.

42 Ishizawa, T., Kawakami, T., Reid, P.C., and Murakami, H. (2013). *Journal of the American Chemical Society* 135: 5433–5440.

43 Obexer, R., Walport, L.J., and Suga, H. (2017). *Current Opinion in Chemical Biology* 38: 52–61.

44 Panse, S., Dong, L., Burian, A. et al. (2004). *Molecular Diversity* 8: 291–299.

45 Lam, K.S. and Salmon, S.E. (1991). *Nature* 354: 82.

46 Young, T.S., Young, D.D., Ahmad, I. et al. (2011). *Proceedings of the National Academy of Sciences of the United States of America* 108: 11052–11056.

47 (a) Kritzer, J.A., Hamamichi, S., McCaffery, J.M. et al. (2009). *Nature Chemical Biology* 5: 655–663. (b) Kinsella, T.M., Ohashi, C.T., Harder, A.G. et al. (2002). *Journal of Biological Chemistry* 277: 37512–37518.

48 (a) Heinis, C., Rutherford, T., Freund, S., and Winter, G. (2009). *Nature Chemical Biology* 5: 502–507. (b) McLafferty, M.A., Kent, R.B., Ladner, R.C., and Markland, W. (1993). *Gene* 128: 29–36. (c) Heinis, C. and Winter, G. (2015). *Current Opinion in Chemical Biology* 26: 89–98.

49 (a) Guillen Schlippe, Y.V., Hartman, M.C., Josephson, K., and Szostak, J.W. (2012). *Journal of the American Chemical Society* 134: 10469–10477. (b) Kawakami, T., Murakami, H., and Suga, H. (2008). *Journal of the American Chemical Society* 130: 16861–16863. (c) Kawakami, T., Ishizawa, T., and Murakami, H. (2013). *Journal of the American Chemical Society* 135: 12297–12304.

50 Hipolito, C.J. and Suga, H. (2012). *Current Opinion in Chemical Biology* 16: 196–203.

51 Bashiruddin, N.K., Nagano, M., and Suga, H. (2015). *Bioorganic Chemistry* 61: 45–50.

52 Colas, P., Cohen, B., Jessen, T. et al. (1996). *Nature* 380: 548–550.

53 Reverdatto, S., Rai, V., Xue, J. et al. (2013). *PLoS One* 8: e65180.

54 Koch, O., Cole, J., Block, P., and Klebe, G. (2009). *Journal of Chemical Information and Modeling* 49: 2388–2402.

55 Keskin, O., Gursoy, A., Ma, B., and Nussinov, R. (2008). *Chemical Reviews* 108: 1225–1244.

56 Jochim, A.L. and Arora, P.S. (2009). *Molecular BioSystems* 5: 924–926.

57 Taylor, H.S. (1941). *Proceedings of the American Philosophical Society* 85: 1–12.

58 Toniolo, C. (1980). *CRC Critical Reviews in Biochemistry* 9: 1–44.

59 Donohue, J. (1953). *Proceedings of the National Academy of Sciences of the United States of America* 39: 470–478.

60 Barlow, D.J. and Thornton, J.M. (1988). *Journal of Molecular Biology* 201: 601–619.

61 (a) Pavone, V., Di Blasio, B., Santini, A. et al. (1990). *Journal of Molecular Biology* 214: 633–635. (b) Benedetti, E. (1988). *Journal of Biomolecular Structure & Dynamics* 5: 803–817.

62 Toniolo, C. and Benedetti, E. (1991). *Trends in Biochemical Sciences* 16: 350–353.

63 Ramachandran, G.N. and Sasisekharan, V. (1968). *Advances in Protein Chemistry* 23: 283–438.

64 Rohl, C.A. and Doig, A.J. (1996). *Protein Science: A Publication of the Protein Society* 5: 1687–1696.

65 Gibbs, N., Sessions, R.B., Williams, P.B., and Dempsey, C.E. (1997). *Biophysical Journal* 72: 2490–2495.

66 Low, B.W. and Baybutt, R.B. (1952). *Journal of the American Chemical Society* 74: 691–699.

67 Fodje, M.N. and Al-Karadaghi, S. (2002). *Protein Engineering* 15: 353–358.

68 Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008). *Bioinformatics (Oxford, England)* 24: 2780–2781.

69 Cooley, R.B., Arp, D.J., and Karplus, P.A. (2010). *Journal of Molecular Biology* 404: 232–246.

70 Rennex, D., Cummings, R.T., Pickett, M. et al. (1993). *Biochemistry* 32: 7475–7478.

71 Al-Karadaghi, S., Hansson, M., Nikonov, S. et al. (1997). *Structure (London, England: 1993)* 5: 1501–1510.

72 Schubert, H.L., Raux, E., Wilson, K.S., and Warren, M.J. (1999). *Biochemistry* 38: 10660–10669.

73 Sazinsky, M.H. and Lippard, S.J. (2005). *Journal of the American Chemical Society* 127: 5814–5825.

74 Bailey, L.J., McCoy, J.G., Phillips, G.N. Jr., and Fox, B.G. (2008). *Proceedings of the National Academy of Sciences of the United States of America* 105: 19194–19198.

75 Lo Conte, L., Chothia, C., and Janin, J. (1999). *Journal of Molecular Biology* 285: 2177–2198.

76 Vinogradova, O. and Qin, J. (2012). *NMR of Proteins and Small Biomolecules* 326: 35–45.

77 Rodger, A. (2006). *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Ltd.

78 Scapin, G. (2006). *Current Pharmaceutical Design* 12: 2087–2097.

79 Souers, A.J., Leverson, J.D., Boghaert, E.R. et al. (2013). *Nature Medicine* 19: 202–208.

80 Dyson, H.J., Cross, K.J., Houghten, R.A. et al. (1985). *Nature* 318: 480–483.

81 Brandts, J.F., Halvorson, H.R., and Brennan, M. (1975). *Biochemistry* 14: 4953–4963.

82 Shepherd, N.E., Hoang, H.N., Abbenante, G., and Fairlie, D.P. (2005). *Journal of the American Chemical Society* 127: 2974–2983.

83 Bystrov, V.F., Arseniev, A.S., and Gavrilov, Y.D. (1978). *Journal of Magnetic Resonance (1969)* 30: 151–184.

84 Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). *Biochemistry* 31: 1647–1651.

85 Kessler, H. (1982). *Angewandte Chemie International Edition in English* 21: 512–523.

86 Miller, S.E., Kallenbach, N.R., and Arora, P.S. (2012). *Tetrahedron* 68: 4434–4437.

87 Assem, N., Ferreira, D.J., Wolan, D.W., and Dawson, P.E. (2015). *Angewandte Chemie International Edition* 54: 8665–8668.

88 Le Chevalier Isaad, A., Papini, A.M., Chorev, M., and Rovero, P. (2009). *Journal of Peptide Science: An Official Publication of the European Peptide Society* 15: 451–454.

89 Lau, Y.H., Wu, Y., de Andrade, P. et al. (2015). *Nature Protocols* 10: 585–594.

90 Chatterjee, C., Paul, M., Xie, L.L., and van der Donk, W.A. (2005). *Chemical Reviews* 105: 633–683.

91 (a) Goto, Y., Li, B., Claesen, J. et al. (2010). *PLoS Biology* 8: e1000339. (b) Zhang, Q., Yu, Y., Velasquez, J.E., and van der Donk, W.A. (2012). *Proceedings of the National Academy of Sciences of the United States of America* 109: 18361–18366. (c) Siegers, K., Heinzmann, S., and Entian, K.D. (1996). *The Journal of Biological Chemistry* 271: 12294–12301.

92 Poth, A.G., Chan, L.Y., and Craik, D.J. (2013). *Biopolymers* 100: 480–491.

93 (a) Wieland, T. (1979). *Hoppe-Seyler's Zeitschrift für Physiologische Chemie* 360: 1202. (b) Schafer, A.J. and Faulstich, H. (1977). *Analytical Biochemistry* 83: 720–723. (c) Malak, S.H.A. (1976). *Planta Medica* 29: 80–85. (d) Faulstic, H. and Wieland, T. (1971). *European Journal of Biochemistry* 22: 79.

94 (a) Tu, A.T. and Hong, B.S. (1971). *The Journal of Biological Chemistry* 246: 2772. (b) Chicheportiche, R., Vincent, J.P., Kopeyan, C. et al. (1975). *Biochemistry* 14: 2081–2091. (c) Chen, Y.H., Tai, J.C., Huang, W.J. et al. (1982). *Biochemistry* 21: 2592–2600.

95 Baeriswyl, V. and Heinis, C. (2013). *ChemMedChem* 8: 377–384.

96 (a) Kolmar, H. (2009). *Current Opinion in Pharmacology* 9: 608–614. (b) Craik, D.J., Cemazar, M., and Daly, N.L. (2006). *Current Opinion in Drug Discovery & Development* 9: 251–260. (c) Gould, A., Ji, Y., Aboye, T.L., and Camarero, J.A. (2011). *Current Pharmaceutical Design* 17: 4294–4307.

97 (a) Ay, J., Hilpert, K., Krauss, N. et al. (2003). *Acta Crystallographica Section D: Biological Crystallography* 59: 247–254. (b) Hilpert, K., Wessner, H.,

Schneider-Mergener, J. et al. (2003). *The Journal of Biological Chemistry* 278: 24986–24993.

98 Lian, W., Jiang, B., Qian, Z., and Pei, D. (2014). *Journal of the American Chemical Society* 136: 9830–9833.

99 Yan, L.-M., Tatarek-Nossol, M., Velkova, A. et al. (2006). *Proceedings of the National Academy of Sciences of the United States of America* 103: 2046–2051.

100 Kessler, H. (1970). *Angewandte Chemie International Edition in English* 82: 237–253.

101 Piriou, F., Lintner, K., Fermandjian, S. et al. (1980). *Proceedings of the National Academy of Sciences of the United States of America* 77: 82–86.

102 Doedens, L., Opperer, F., Cai, M. et al. (2010). *Journal of the American Chemical Society* 132: 8115–8128.

103 Dervan, P.B. (1986). *Science* 232: 464–471.

104 Horne, W.S. and Gellman, S.H. (2008). *Accounts of Chemical Research* 41: 1399–1408.

105 Schumacher, T.N., Mayr, L.M., Minor, D.L. Jr., et al. (1996). *Science (New York, NY)* 271: 1854–1857.

106 Bai, L., Sheeley, S., and Sweedler, J.V. (2009). *Bioanalytical Reviews* 1: 7–24.

107 (a) Khakshoor, O. and Nowick, J.S. (2008). *Current Opinion in Chemical Biology* 12: 722–729. (b) Nowick, J.S., Smith, E.M., and Pairish, M. (1996). *Chemical Society Reviews* 25: 401–415.

108 (a) Weckbecker, G., Lewis, I., Albert, R. et al. (2003). *Nature Reviews Drug Discovery* 2: 999–1017. (b) Aumailley, M., Gurrath, M., Muller, G. et al. (1991). *FEBS Letters* 291: 50–54. (c) Veber, D.F., Freidinger, R.M., Perlow, D.S. et al. (1981). *Nature* 292: 55–58.

109 (a) Karle, I.L., Awasthi, S.K., and Balaram, P. (1996). *Proceedings of the National Academy of Sciences of the United States of America* 93: 8189–8193. (b) Haque, T.S., Little, J.C., and Gellman, S.H. (1996). *Journal of the American Chemical Society* 118: 6975–6985.

110 (a) Athanassiou, Z., Dias, R.L.A., Moehle, K. et al. (2004). *Journal of the American Chemical Society* 126: 6906–6913. (b) Stanger, H.E. and Gellman, S.H. (1998). *Journal of the American Chemical Society* 120: 4236–4237. (c) Haque, T.S., Little, J.C., and Gellman, S.H. (1994). *Journal of the American Chemical Society* 116: 4105–4106.

111 (a) Wani, M.C., Taylor, H.L., Wall, M.E. et al. (1971). *Journal of the American Chemical Society* 93: 2325–2327. (b) Kudo, F., Miyanaga, A., and Eguchi, T. (2014). *Natural Product Reports* 31: 1056–1073.

112 (a) Lengyel, G.A., Frank, R.C., and Horne, W.S. (2011). *Journal of the American Chemical Society* 133: 4246–4249. (b) Reinert, Z.E., Lengyel, G.A., and Horne, W.S. (2013). *Journal of the American Chemical Society* 135: 12528–12531.

113 Checco, J.W. and Gellman, S.H. (2016). *Current Opinion in Structural Biology* 39: 96–105.

114 Cheloha, R.W., Maeda, A., Dean, T. et al. (2014). *Nature Biotechnology* 32: 653–655.

115 Johnson, L.M., Barrick, S., Hager, M.V. et al. (2014). *Journal of the American Chemical Society* 136: 12848–12851.

116 Haase, H.S., Peterson-Kaufman, K.J., Lan Levengood, S.K. et al. (2012). *Journal of the American Chemical Society* 134: 7652–7655.

117 Checco, J.W., Lee, E.F., Evangelista, M. et al. (2015). *Journal of the American Chemical Society* 137: 11365–11375.

118 Patgiri, A., Joy, S.T., and Arora, P.S. (2012). *Journal of the American Chemical Society* 134: 11495–11502.

119 Kreitler, D.F., Mortenson, D.E., Forest, K.T., and Gellman, S.H. (2016). *Journal of the American Chemical Society* 138: 6498–6505.

120 Lee, E.F., Sadowsky, J.D., Smith, B.J. et al. (2009). *Angewandte Chemie International Edition* 48: 4318–4322.

121 (a) Tošovská, P. and Arora, P.S. (2010). *Organic Letters* 12: 1588–1591.

122 Lao, B.B., Grishagin, I., Mesallati, H. et al. (2014). *Proceedings of the National Academy of Sciences of the United States of America* 111: 7531–7536.

123 Barnard, A., Long, K., Martin, H.L. et al. (2015). *Angewandte Chemie International Edition* 54: 2960–2965.

124 (a) Ernst, J.T., Becerril, J., Park, H.S. et al. (2003). *Angewandte Chemie International Edition* 115: 553–557.

125 Tanatani, A., Hughes, T.S., and Moore, J.S. (2002). *Angewandte Chemie International Edition* 41: 325–328.

126 Hou, J.L., Shao, X.B., Chen, G.J. et al. (2004). *Journal of the American Chemical Society* 126: 12386–12394.

127 (a) Cuccia, L.A., Ruiz, E., Lehn, J.M. et al. (2002). *Chemistry (Weinheim an der Bergstrasse, Germany)* 8: 3448–3457. (b) Martinek, T.A., Hetenyi, A., Fulop, L. et al. (2006). *Angewandte Chemie International Edition* 45: 2396–2400. (c) Pomerantz, W.C., Yuwono, V.M., Pizzey, C.L. et al. (2008). *Angewandte Chemie International Edition* 47: 1241–1244. (d) Cai, W., Wang, G.T., Xu, Y.X. et al. (2008). *Journal of the American Chemical Society* 130: 6936–6937.