Jianpeng Wang

Study of the Peptide–Peptide and Peptide–Protein Interactions and Their Applications in Cell Imaging and Nanoparticle Surface Modification

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria:

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

More information about this series at http://www.springer.com/series/8790

Jianpeng Wang

Study of the Peptide–Peptide and Peptide–Protein Interactions and Their Applications in Cell Imaging and Nanoparticle Surface Modification

{\underline{\mathcal{D}}} Springer

Author
Dr. Jianpeng Wang
Department of Chemistry
The Chinese University of Hong Kong
Hong Kong
China

Supervisor
Prof. Jiang Xia
Department of Chemistry
The Chinese University of Hong Kong
Hong Kong
China

Supervisor's Foreword

Protein–peptide interactions are nature's gift for chemists. This thesis creatively engineered naturally existing protein-peptide or peptide–peptide interactions to develop peptide tags for covalent protein labeling and protein assembly. This thesis shows for the first time that one can engineer covalent bond between peptide–peptide interactions based on the principle "affinity-guided covalent conjugation." The proof of this principle has opened a door to almost unlimited possibilities to develop proximity-induced reactions in many different interacting pairs. From fundamental science aspect, the investigation of this affinity-guided reactions will shed light on many natural important processes such as signal transduction. From more practical aspect, this thesis adds a new chapter in the field of bioconjugate chemistry. Further, these new chemical biology tools markedly advanced the fields of cell imaging and protein assemblies.

This thesis is an outstanding example of multidisciplinary research spanning organic chemistry, protein engineering, and cell biology. The major part of the thesis was reported in several publications and two of them have been highlighted in journal covers.

August 2016

Prof. Jiang Xia
Associate Professor of Chemistry
The Chinese University of Hong Kong

Parts of this thesis have been published in the following journal articles:

(*: corresponding author; †: co-first author)

1. **<u>Jianpeng Wang</u>**[†], Yunyu Nie[†], Yao Lu, Jiahui Liu, Jianhao Wang, Aisi Fu, Tiangang Liu, Jiang Xia[*], Assembly of multivalent protein ligands and quantum dots: a multifaceted investigation. *Langmuir*, **2014,** *30*, 2161–2169. (Equal contribution, cover paper)

2. **<u>Jianpeng Wang</u>**, Yongsheng Yu, Jiang Xia[*], Short peptide tag for covalent protein labeling based on coiled coils. *Bioconjugate. Chem.*, **2014,** *25*, 178–187. (Cover paper)

3. Yao Lu, **<u>Jianpeng Wang</u>**, Jianhao Wang, Lin Wang, Shannon Wing-Ngor Au, Jiang Xia[*], Genetically encodable design of ligand "bundling" on the surface of nanoparticles. *Langmuir*, **2012**, *28*, 13788–13792.

4. Yongsheng Yu, **<u>Jianpeng Wang</u>**, and Jiang Xia[*], Functional assembly of protein fragments induced by spatial confinement. *PLoS One*, **2015**, 10, e0122101.

Acknowledgements

Contents

Chapter 1
Introduction

1.1 Introduction to Chemical Biology

Chemical biology, which emerged over decades as a complex hybridization of bioorganic chemistry, biochemistry, cell biology, and pharmacology [1], is considered to be a modern interdisciplinary science. It involves the application of molecules from synthetic chemistry, as well as other chemical techniques and tools, to the understanding and exploring of biological problems. The past few decades have been a remarkable period for the chemical biology, with numerous intellectual ideas and methodological strategies coming to the center stage of the interface of chemistry and biology. It is among the fastest growing areas in natural sciences and in chemistry in particular.

Chemists have been developing methodologies to synthesize the biologically relevant molecules, ranging from as small molecules to proteins and nucleic acids. Meanwhile the development of analytical instruments and techniques also allows people to monitor the biological and biochemical processes in a much more precise manner [2]. Chemical biology often starts with a biological problem or phenomenon of interest, analyzes them with the assistance of modern techniques to convert them into chemical problems, and finally develops new tools or methodologies which apply back in the biological systems as shown in Fig. 1.1.

Much progress has been made during the last few years in chemical biology, for example the activity-based probes [3], RNAi in mammals [4], automated carbohydrate synthesis [5], noval protein fold design [6], in situ click chemistry [7], unprotected natural product synthesis [8], chromatin-level regulation of biosynthesis [9], PRIME fluorophore labeling [10]. And in this thesis, we will try to explore and understand several small issues addressed in chemical biology.

Within the past 4 years, I have been involved in several projects related to the design and application of peptides and proteins. Although whether these projects fall into the category of chemical biology is arguable, the main theme has been to utilize chemical knowledge and tools to solve biologically relevant problems.

Fig. 1.1 Interplay between
chemistry and biology in
chemical biology studies
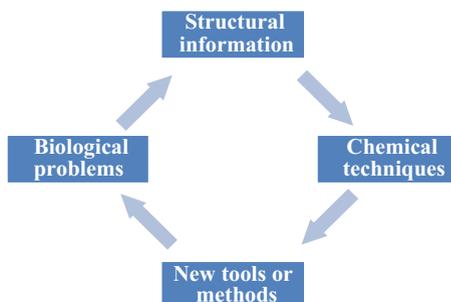
The projects involve organic synthesis of molecules, synthesis and characterization of inorganic nanoparticles, protein engineering, and cell imaging; the multidisciplinary nature of this thesis is thereby undoubted. One feature to distinguish my research from others' is that these projects do not fall into the realms of traditional chemistry or biochemistry; yet when it is deemed necessary, I learned and employed knowledge and techniques from other fields to solve problems. This, I believe, represents the essence of chemical biology.

1.2 Antimicrobial Peptides

Antimicrobial peptides are widely spread in nature, and playing fundamental role in the evolution of the multicellular organisms. There are more than 500 reported antimicrobial peptides [11]; their structures are so diverse that they could only be vaguely categorized by the secondary structure. Still, a converging feature of antimicrobial peptide is an "amphipathic" structure, with all the hydrophobic amino acids packed together and all the positively charged amino acids distributed on the surface of the molecule. For example, linear peptides, such as cecropin in silk moth [12], adopt an amphipathic α-helix when it encounters a plasma membrane. Other peptides such as defensins contain antiparallel β-sheets stabilized by multiple disulfide bonds [13].

The amphipathic nature is linked to the antimicrobial mechanism with respect to its membrane interaction, in many cases (but not all the cases). The plasma membrane of microbes contains negatively charged phospholipids while mammalian cell membrane is mainly composed of lipids with no net charges. This provides the antimicrobial peptides specific electrostatic interaction toward microbes but not mammalian cells. The hydrophobic portion of the amphipathic peptides then disrupt the membrane structure and thus kill the microbes. This is considered to be the widely accepted Shai—Matsuzaki—Huang (SMH) model for the antimicrobial peptide to kill microbes [14–16].

As antimicrobial peptides targets cell membrane, it is thereby more likely to circumvent the drug resistance pathways. Variants of antimicrobial peptides could also be engineered by amino acids replacement such as replacing natural L-amino acids by D-amino acids [17] or β amino acid [18].

1.3   Quantum Dots

The organic dyes and fluorescent proteins have wide application in fluorescent imaging, but they suffer from drawbacks such as broad absorption/emission profiles, low photobleaching thresholds. On the contrary, semiconductor nanocrystals (also known as quantum dots or QDs) hold several promising optical properties such as the high quantum yield, broad absorption with narrow symmetric photoluminescence spectra spanning from the UV to near-infrared, large effective Stokes shifts, high resistance to photobleaching and exceptional resistance to photo- and chemical degradation [19]. Moreover, QDs have tunable fluorescent emission spectra dependent on the size of the core; their broad excitation spectra also allow the simultaneously excitation different QDs at a single wavelength. QDs are thus widely used in energy conversion and storage [20], optoelectronic devices [21], fluorescent sensors [22, 23], and photocatalysis [24]. Specifically, CdSe-ZnS QDs are widely used in tissue imaging, biosensing and cell targeting. It contains a core shell structure with ZnS grown on the surface of CdSe, with the overall diameter of about 5 nm. The quantum yield could reach 30–50 % with a narrow band edge luminescence spanning most of the visible spectrum from 470 to 625 nm [25].

Protein coating of QDs has been a general method to modify and protect water soluble QDs to mediate its application in biological systems [26], One strategy to functionalize the surface of QDs by biomolecules is through metal-affinity-driven self-assembly [27]. Peelle et al. showed that histidine, cysteine, methionine and tryptophan have high tendency to bind to CdSe-ZnS QDs [28], and the metal–ligand coordination interaction is sufficiently stable in the case of oligohistidine [29]. Our lab further extended a single histag to a dendrimer of histags which showed a much stronger binding affinity toward QDs in both peptides and proteins [30, 31]. A systematic investigation of how the distribution and combination of histags contributes to the interaction between protein ligands and QDs is thereby intriguing.

1.4   Introduction to Chemical Tag and Probe System

Being the vital, if not the most important, building block of life, protein plays a fundamental role in almost every life process, varying from structural support, scaffolding, elements storage (such as N and Fe), to transportation, regulation, catalysis, defense and attack, etc. [32]. Thus to visualize, measure and track

proteins, especially in the in vivo conditions, has attracted much of biologists' attentions. In the past a few decades, various techniques, as well as corresponding instruments, have been developed to assist in the characterization and under-standing of the spatial distribution and temporal variations of proteins.

Generally speaking, the labeling of the protein of interest could be divided into two main strategies: the fusion of a reporter protein (auto fluorescent proteins (AFPs), luciferase, β-galactosidase, β-lactamase, etc.) and the chemical tag and probe system. Being one of the most important discoveries in protein science in the 1990s, AFPs had played an incredible role in monitoring and understanding protein localization, translocation, interactions and conformational change etc. [33]. When it was genetically fused to the host proteins, it provides biologists with great convenience and absolute specificity as well as some drawbacks listed below:

(1) Bulky size. AFPs contain ∼240 amino acids, comparable to the host protein, thus has the potential to interfere the structure and even the functionality of the host protein;

(2) Limited spectrums. Although GFP variants had been explored to give the corresponding Red (RFP), Blue (BFP), Cyan (CFP) and Yellow (YFP) [34], they still have limitation in offering spectrum libraries, for example no near-Infrared AFP has been developed. Besides, synthetic fluorescent dyes also beat AFP in the brightness and photostability.

(3) Limited fusion sites. So far as we know, GFP moiety could be fused only to the C- or N-terminus of the host protein, while some of the chemical tags could be fused to the middle of the host (TC tag, for example) [35].

(4) Limited environment. Having the conserved barrel structure, both the in vitro the in vivo locations of AFPs would be limited due to the pH value, envi-ronmental hydrophobicity, ion concentrations, etc.

So it had appeared to be of urgent need to explore a complementary strategy to label the protein of interest. And in the past two or three decades, numerous of efforts had been put into developing and evolution the chemical tag and probe system. As a rule of principle, the labeling should fulfill but is not limited to the following criteria:

(1) Specificity. There are great differences between the isolated proteins and that in the biological context. The labeling reaction should be highly specific so that it could be performed both in vitro and in vivo. More importantly, it should be capable of preventing the nonspecific interactions toward thousands of other proteins, DNA, RNA, carbohydrates, lipids, and other small mole-cules to provide the desired signal to noise ratio;

(2) Reactivity. It is known to all that the labeling rate determines the application area of the system. The faster the labeling reaction rate is the more biological processes it could trace. Besides, when it comes to the in vivo environment, everything is in fast dynamics so it requires the labeling to proceed before the target's degradation;

(3) Structural interference. The function of a protein is highly related to its well-defined structure. Neither the fusion of the potential targeting moiety nor the attachment of the binding ligand to the parental protein is supposed to alter the conformational property of the latter. Under this circumstance, the size of the tag or probe is better to be as small as possible, and the content to be biological relevant;

(4) Ease of manipulation. Due to the potential users of such a tag-probe approach have little background in synthetic chemistry, the commercial availability and the convenient manipulation are also in great demand.

1.4.1   Classification of Covalent Chemical Tag and Probe System

Usually, the artificial labeling of a protein could be divided into three major strategies: (1) based on a single amino acid; (2) based on a short peptide sequence; (3) based on a fusion protein (enzyme). Each of them would be discussed in the following content.

1.4.1.1   Labeling Methodologies Based on Single Amino Acid

Usually this kind of methodology takes advantage of a bioconjugation reaction which would introduce the synthetic probe to the side chain of one or more amino acid(s). In a protein with defined three-dimensional structure, the interior is non-solvent accessible which makes the inner amino acids poor candidates for the conjugation. Only the surface amino acids are potential reactive sites to the coming synthetic probe. Besides, because this kind of method would only recognize the potential reacting sites despite of the microenvironment, the probe would be anchored to a kind of functional groups with no or poor selectivity.

Modifications of Cys and Lys

While having a super wide spread in functional properties, the functional groups of a protein are usually limited to that in 20 natural amino acids and some post-translational modifications. Among them, the thiol group from Cys side chain and the amine group from Lys side chain or the N-terminus are the most common candidates due to their high reactivity. There are numerous reports for labeling Cys and Lys in the biological context [36] and even the reaction kits are highly commercialized. For the thiol group of Cys, reagents such as α-haloketones and maleimide derivatives are well developed, as well as the traditional disulfide exchange reaction. And for the ε-amine of Lys, the labeling reagents are even more explored to give a toolbox containing isothiocyanates, isocyanates, activated esters and epoxide, etc. [36]. Unlike Lys and other amino acids with reactive side chains,

Cys has a much lower abundance ratio in protein surface (2.3 %), making it a potential candidate for the selective labeling. But this is a case-dependent issue and only applicable for those proteins with no Cys at the surface. By single site mutagenesis, Cys could be easily introduced to the protein surface to function as the targeting site.

Modifications of rare amino acids

Although the bioconjugation reactions based on the side chains of Cys and Lys give fast and complete conversion, the selectivity and site-specificity are compromised due to the high reactivity of the electrophiles. Besides, when this methodology is applied to in vivo protein labeling, poor signal to noise ratio would be anticipated because of the high percentage of the Cys and Lys in proteins. Based on all these disadvantages, Francis et al. have been dedicated to the development of covalent bioconjugation reactions based on the rare amino acids.

Being both the largest and rarest natural amino acid, Trp is supposed to possess only 1.4 % occurrence in proteins, and the abundance ratio would be even lower in protein surface. A rhodium carbenoids based method was developed by Francis et al. in 2004 [37] which could achieve the selectively covalent attachment of a vinyl diazo compound to horse heart myoglobin in the presence of $Rh_2(OAc)_4$ as the catalyst. Although highly selective, the reaction requires long reaction time (7 h at room temperature), co-solvent (ethylene glycol) and large excessive reactant (100 equiv.). Besides, this kind of strategy was case dependent. The pH value had to be lower down to 1.5 for denature of the protein Carsberg to expose the buried Trp for the labeling reaction to proceed.

A three components Mannich reaction involves aldehyde and aniline was developed by the same group to covalently label Tyr (3.2 % in abundance). Due to the amphiphilic characteristics, Tyr is supposed to favor the protein surface location so that it could be easily targeted. Although longer reaction time was required (18 h at room temperature), this kind of reaction was highly selective and would proceed under mild conditions (pH5.5–6.5), compared to the traditional Tyr-labeling strategies which require large amounts of formaldehyde and heat denaturation [38]. To further accelerate the Tyr-labeling reaction, Francis et al. reported a protein alkylation method based on π-Allylpalladium complexes [39]. Upon adding of a rhodamine labeled allylic acetate, $Pd(OAc)_2$ and triphenylphosphine tris-(sulfonate), efficient Tyr-labeling of Chymotrypsinogen A could be observed within 45 min (room temperature) and mild pH condition (8.5–9.0).

Unnatural amino acid mutagenesis

Once a single amino acid becomes the targeting site, either on the protein surface or in the binding pocket, either in its original primary sequence or in a site-directed mutagenesis case, it would be incredibly difficult if this kind of reaction is going to be carried out in the in vivo environment, especially in the specific targeting manner. To overcome this drawback, unnatural amino acid incorporation technique based on amber condon was developed to introduce to the target protein a

functional group that does not originally exist in the biological system to bring absolute specificity. The azide, alkane, ketone, aldehyde, and many other groups could be introduced to the target protein by this very technique and could provide further orthogonal reactivity [40]. However, this technique was also suffered from the bulky and hydrophobic amino acid side chain so that its application was limited to some extent.

1.4.1.2 Labeling Strategies Based on Short Peptide Tag Sequence

Although the labeling strategies based on single amino acid provide least disruption of the protein of interest, the harsh reaction conditions, and the long reaction time significantly inhibit them to find applications in the in vivo protein targeting, or even the labeling of "difficult" proteins with poor stability becomes impossible. On the other hand, the labeling events taking advantages of the genetic fusion of an enzyme provide sufficient biocompatibility as well as greater structure distortion. Combine the two facts together, the labeling strategies that are based on a short peptide sequence as a potential reacting sites are becoming more and more popular. Generally speaking, the beauty of this kind of strategies lies not only in the comprising on both the biocompatibility and the less distortion of the POI, but also in the fast reaction rate, good specificities. Some of them could even achieve the covalent labeling that was considered to be highly important to prevent the signal losing problems during some of biological processes.

FLASH tag

It was the amazing work in the history of chemical tag and probe system pioneered by Tisen et al. in 1998 [33]. In their design, six natural amino acids (CCXXCC) were used as the targeting site, in which Cys1-Cys5 and Cys2-Cys6 were designed to locate in i and i + 4 positions in an α-helix respectively to form a parallelogram on one side of the helix. Upon addition of the nonfluorescent and cell-permeable small molecule, FlAsH-EDT2 (fluorescein arsenical helix binder, bis-EDT adduct), the tetra-Cys would replace EDT to form a fluorescent complex. In the following works, the undefined amino acids X were optimized to be Pro-Gly [41] to facilitate the formation of α-helix and enhance the binding affinity.

Being the first de novo designed chemical tag and probe system, FlAsH tag owns obvious advantages against the traditional fluorescent protein fusions: first, the target size is much smaller (only ~700D compared to 28KD for GFP) so that the potency and probability for it to detract the structure and function of the host protein is lower down to an acceptable level; second, theoretically it could be fused to any site of the target protein only if the function of the original protein remains; third, the four Cys are required to be in the reduced form so that the disulfide bonded surface Cystine would not contribute to the nonspecific binding. On the other hand, it also has several drawbacks such as the requirement for intensive washout because of the significant binding property for biarsenical compound

toward the isolated thiols. The background fluorescent signal would remain even after extensive washing. Besides, this method would fail to work in the oxidative environments such as the secretory pathway.

To overcome the drawbacks described before, numerous of work has been done in recent years. Although new fluorescent molecular structure has to be redesigned to generate a probe with an alternative spectrum, the recently developed ReAsH [42], CrAsH [43], and Cy3As [44] tags have greatly expanded the biological's fluorescent toolbox, which may even allow the pulse-chase imaging [45].

Furthermore, in the attempting to eliminate the arsenical caused high background labeling and potential cytotoxicities, Schepartz et al. derived the bis-arsenical based probe to bis-boronic acid based probe [46] in which the tetra-Cys sequence were replaced by tetra-Ser sequence(SSPGSS). Once bound to the RhoBo [(3-oxospirol[isobenzofuran-1(3H),9′-[9H]xanthene]-3′,6′-diyl)bis (ininomethylene-2,1-phenylene)]bis-(9CI)], the tetra-Ser tag is supposed to perform much better in photostability and brightness than ReAsH-EDT$_2$ tag. What is more, this tag and probe system was tested against a 377-member mammalian glycan microarray to show a great improvement in the selectivity.

D4 tag

Inspired by the interaction between Ni-NTA and $6\times$ His, Hamachi et al. developed the multinuclear Zn(II) complexes (Zn(II)-DpaTyrs) which derived from the Tyr scaffold and would coordinate to the genetically encoded oligo-aspartate (D4 tag) [47]. By taking advantage of the multivalence effects, the binding affinity between the dimerized Zn(II) complexes and sequential extended (D4)2 tag was dramatically increased by 1000 fold ($K_{app} = (1.8 \pm 0.3) \times 10^7 \text{ M}^{-1}$). The enzyme activity of the fused host protein RNase was not interfered by the small size of the tag. Whatis more interesting is that the interaction between Zn(II) complexes and D4 tag is orthogonal to that between Ni-NTA and $6\times$ His, making it a great candidate for the noncovalent cell surface protein labeling pair.

In order to further extend this labeling strategy to a nondissociable manner, they respectively install a Cys and an oligo Ala linker to the D4 tag, and an α-chloroacetyl functional group to the Zn(II) complexes. Proximity-induced nucleophilic reaction would take place to generate a covalent linkage between the tag and probe after the metal–ligand coordination [48]. Different from all the previously described tag-probe systems, this kind of strategy involves no enzyme, and the reaction would complete within 30 min. Due to the ultrasmall size of the D4 tag ($\sim$20 amino acids), it would generate little, if any, interference of the protein host. Based on this, fast labeling (within 30 min) of the G protein–coupled receptor [Bradykinin B$_2$ receptor (B$_2$R) and acetyl choline receptor (M1R)] could be achieved. By incorporation of a fluorescent dye, the agonist induced internalization of the GPCRs could be traced by fluorescent microscope [49].

1.4.1.3   Labeling Methodologies Based on Fusion Proteins (Enzymes)

For the tags developed based on the fusion of an enzyme, it would usually introduce a covalent linkage between the substrate or the enzyme itself and the coming ligand based on the catalytic reactions mediated by the enzyme. It is considered to have the following advantages, as well as some drawbacks:

(1) High biocompatibility. Since the enzyme itself is a functional protein, upon fusing to the target protein, no further compatible issues should be worried about;
(2) Fast. The enzyme catalyzed reactions usually proceed in a fast reaction rate manner, especially for some of the tags, the enzymes are even engineered to give a better performance in catalysis so that the labeling process usually complete to a satisfying degree within tens of minutes or even less;
(3) High specificity. Thanks to the enzyme, the reactions of this kind of labeling methodologies are of high specificity, only the conserved substrate and ligand could be recognized by the enzyme. In some cases, orthogonal reactions could be achieved even within similar substrate structure (for example, SNAP-tag and CLIP-tag);
(4) Interruption. Being the biggest drawbacks of this methodology, the structure of the protein of interest would sometimes be interrupted and the function might be lost due to the large size of the enzyme it is fused to.

hAGT tag

In the year of 2002, Kai Johnsson et al. reported a new covalent labeling strategy which took advantage of the irreversible transferation of an alky group from the substrate O^6-alkylguanine-DNA to one of the enzyme's Cys mediated by the human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT) [50].

Having low selectivity toward nucleobase O^6-benzylguanine (BG), derivatives of O^6-benzylguanine with substituted benzyl rings besides its original substrate, the hAGT enzyme protein (~ 210 amino acids in length) was supposed to capable of transferring the fluorescent probe from the engineered O^6-benzylguanine derivatives to the reactive Cys, as well as to the target protein, once it was fused to the protein of interest. Different fluorescent dyes (SNARF-1, fluorescein Oregon green, tetramethylrhodamine, etc.) [51] were conjugated to the BG group via a benzyl linker, and could be applied in site-specific incorporation in the AGT deficient living cells.

Besides, the hAGT tag has been fully optimized to be the highly commercialized SNAP-tag and it was further evolved to CLIP-tag which takes O^2-Benzylcytosine derivatives as the substrate [52] and could provide orthogonal labeling together with SNAP-tag. By the combination of SNAP-tag and CLIP-tag, multicolor and simultaneous labeling of different fusion proteins in vivo could be expected.

This kind of methodology greatly expanded the spectrum selectivity, and the labeling process could reach a quantitative complete degree within 5 min in *Escherichia coli*, yeast and mammalian cells in different cell compartments, including cytoplasm, nucleus and cell membrane, it also suffers from the bulky size of the hAGT protein ($\sim$22KD). Although in the optimized case, 30 amino acids deletion in the C terminus was proved not to interfere the functionality of the hAGT protein, this tag and probe system could not been applied to other demands. More importantly, all the endogenous AGT (widely spread in mammalian cells) would cause a significant background labeling due to the low specificity of different AGT proteins.

PCP tag and ACP tag

In 2004, Johnsson and Walsh developed the ACP and PCP tags independently [53, 54] which shorten the recognition protein tag sequence to $\sim$80 amino acids.

Inspired by the fatty acid biosynthesis pathway, in which phosphopantetheine transferase (PPTase) would transfer the 4′-phosphopantetheine from coenzyme A (CoA) to a reactive Ser side chain of Acyl Carrier Protein (ACP), Johnsson et al. fused the ACP sequence to a target membrane protein, in the presence of PPTase, the fluorescent probe attached to the CoA derivative would be transferred to the active Ser in ACP, thus resulted in the selective labeling of the target protein. Similarly, the posttranslational modification of PCP in *Bacillus subtilis* was catalyzed by Ppant transferase Sfp to transfer the fluorescently labeled CoA derivatives to the apo form of PCPs.

Although with high labeling specificity and labeling rate (within 10 min), these two tags still were limited by the size of external protein fusions. What is more, the requirement of the presence of a PPTase enzyme makes these two tags only applicable in the cell lysate or the cell membrane protein labeling.

Halo tag

It was reported that, haloalkane dehalogenases would remove the halide from aliphatic hydrocarbons. An ester intermediate would form and then followed by the base catalyzed hydrolysis to give the final alcohol product. A H272F mutation (*Xanthobacter*) would change the ester bond to a stable 1,2-dibromoethane intermediate [55]. Georgyi and coworkers applied an analogous mutation in *Rhodococcus* dehalogenase (DhaA) resulted in the trapping of the covalent intermediate formed by the enzyme and a chloroalkane linker [56]. By changing the length of the linker and key amino acid mutations, they developed a tag and probe pair (HaloTag) which possesses the binding rate comparable to that of common affinity based interactions (such as biotin-streptavdin). Furthermore, p65-HaloTag fusion protein would be specifically labeled by fluorescent appended ligand in living mammalian cells and the translocation of the protein could be monitored in real time.

1.4.2 Noncovalent Tag and Probe System

For this group of strategies, the interaction between the tag and probe is usually achieved by the hydrophobic interactions, hydrogen bonding, metal–ligand coordination interactions.

Antibody mediated tag

An antibody, also known as an immunoglobulin (lg), is a Y shaped protein produced by the B cells, it contains four disulfide stabilized polypeptide chains. It plays an important role in the immune system for binding to a unique part (called antigen) of the foreign objectives such as bacteria and viruses and activating the immune responses [57, 58]. The binding of antibody to antigen is one of the strongest, if not the strongest, in nature that the binding affinity could reach up to nanomolar or even picomolar. Generally, the antigen is short in sequence and could be optimized to be compatible with the host protein, the antibody is large in size, usually has a molecular weight up to around 150 K. The attachment of the antibody would probably disrupt the structure and the function of the target protein, despite of this, this kind of technology has found it great utility in the fields of protein purification, protein immunoprecipitation and protein targeting, etc. Being the first fully functionalized epitopic tag reported, the FLAG tag has been highly commercialized [59]. After that, numerous of other antibody mediated tag and probe systems, for example the HA tag and the Myc tag, have emerged to provide sufficient selectivity and convenience.

His-NTA tag

Oligohistidine sequences have been widely used in the assistance of the protein purification, with the combination of the metal-ion-chelating nitrilotriacetate (NTA) moiety [60]. The bivalent Zn^{2+} ion would provide 6 coordination sites while four of them are occupied by the three acetate groups and the nitrogen atom in the NTA moiety, leaving two for the coming ligands. Having an unsaturated nitrogen atom, the imidazole ring in the histidine side chain becomes a good coordination candidate, especially when 6–10 histidines are aligned together to dramatically enhance the binding affinity by the multivalence effects. Due to the reversible metal–ligand coordination interactions, the binding of the oligohistidine to the NTA moiety could also be competed by high concentration of imidazole solution. However, the beauty of this His-NTA interaction lies in the small size of both the tag and probe so that the oligohistidine could be fused to any position with the least disruption of the protein target, for example in the terminus or in the top of the loop. Horst Vogel et al. applied this technology to the superfast (within seconds) and reversible labeling of ligand-gated ion channel and G protein–coupled receptor. N-terminal fusion, C terminal fusion, as well as the fusion inside the protein sequence (in the loop structure) all indicated that the oligohistidine tag could be recognized by the fluorophore appended NTA probe [61].

Coiled coil tag

The widely existed biological functional motif coiled coil is a well characterized peptide/protein secondary structure, comprising of two or more α helix, aligned in either parallel or antiparallel orientation [62]. Matsuzaki et al. applied this strong peptide–peptide interaction into living cell membrane labeling [63] by fusing one peptide sequence (E3) of a de novo designed coiled coil pattern to the human-derived β_2-adrenergic receptor. The fluorescent dye appended partner sequence (K3 or K4) would recognize and bind to the fused target sequence in an ultrafast (within 10 min) and super sensitive (20 nM concentration of the probe) manner. Futaki et al. succeeded in the artificial control of the dimerization of E3 tag fused EGFR by adding a synthetic K4 dimeric scaffolding peptide [64].

References

1. Bucci M, Coodman C, Sheppard TL et al (2010) A decade of chemical biology. Nat Chem Biol 6:847–854
2. Vella F (2005) Chemical biology: a practical course: Waldmann, H., and Jenning P. Biochem Mol Biol EDU 33:313a–314
3. Liu YS, Patricelli MP, Cravatt BF et al (1999) Activity-based protein profiling: the serine hydrolases. Proc Natl Acad Sci USA 96:14694–14699
4. Elbashir SM, Harborth J, Lendeckel W et al (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411:494–498
5. Plante OJ, Palmacci ER, Seeberger PH et al (2001) Automated solid-phase synthesis of oligosaccharides. Science 291:1523–1527
6. Kuhlman B, Dantas G, Ireton GC et al (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302:1364–1368
7. Lewis WG, Green LG, Grynszpan F et al (2002) Click chemistry in situ: acetylcholinesterase as a reaction vessel for the selective assembly of a femtomolar inhibitor from an array of building blocks. Angew Chem Int Ed 41:1053–1057
8. Baran PS, Maimone TJ, Richter JM et al (2007) Total synthesis of marine natural products without using protecting groups. Nature 446:404–408
9. Bok JW, Chiang Y-M, Szewczyk E et al (2009) Chromatin-level regulation of biosynthetic gene clusters. Nat Chem Biol 5:462–464
10. Uttamapinant C, White KA, Baruah H et al (2010) A fluorophore ligase for site-specific protein labeling inside living cells. Proc Natl Acad Sci USA 107:10914–10919
11. Hancock RE, Patrzykat A (2002) Clinical development of cationic antimicrobial peptides: from natural to novel antibiotics. Curr Drug Targets Infect Disord 2:79–83
12. Steiner H, Hultmark D, Engostöm Å et al (1981) Sequence and specificity of two antibacterial proteins involved in insect immunity. Nature 292:246–268
13. Selsted ME, Harwig SS, Gans T et al (1985) Primary structure of three human neutrophil defensins. J Clin Invest 76:1436–1439
14. Matsuzaki K (1999) Why and how are peptide-lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. Biochim Biophys Acta 1462:1–10
15. Yang L, Weiss TM, Lehrer RI et al (2000) Crystallization of antimicrobial probes in membranes: magainin and protegrin. Biophys J 79:2002–2009
16. Shai Y (1999) Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α-helical antimicrobial and cell non-selective membrane-lytic peptides. Biochim Biophys Acta 1462:55–70

17. Maloy WL, Kari UP (1995) Structure-activity studies on magninins and other host defense peptides. Biopolymers 37:105–122
18. Porter EA, Wang X, Lee H-S et al (2000) Non-haemolytic β-amino-acid oligomers. Nature 404:565
19. Medintz IL, Uyeda HT, Goldman ER et al (2005) Quantum dot bioconjugates for imaging, labeling and sensing. Nat Mater 4:435–446
20. Nozik AJ (2002) Quantum dot solar cells. Physica E 14:115–120
21. Kamat PV (2007) Meeting the clean energy demand: nanostructure architectures for solar energy conversion. J Phys Chem C 111:2834–2860
22. Somers RC, Bawendi MG, Nocera DG et al (2007) CdSe nanoctrysal based chem-/bio-sensors. Chem Soc Rev 36:579–591
23. Walker GW, Sundar VC, Rudzinski CM et al (2003) Quantum-dot optical temperature probes. Appl Phys Lett 83:3555–3557
24. Kamat PV (1993) Photochemistry on nonreactive and reactive (semiconductor) surfaces. Chem Rev 93:267–300
25. Dabbousi BO, Bawendi MG (1997) (CdSe)ZnS core-shell quantum dots: synthesis and characterization of a size series of highly luminescent nanocrystallites. J Phys Chem B 101:9463–9475
26. Carter DC, Chang B, Ho JX et al (1994) Perliminary crystallographic studies of four crystal forms of serum albumin. Eur J Biochem 226:1049–1052
27. Slocik JM, Naik RR (2010) Probing peptide-nanomaterial interactions. Chem Soc Rev 39:3454–3463
28. Peelle BR, Krauland EM, Wittrup KD et al (2005) Design criteria for engineering inorganic material-specific peptides. Langmuir 21:6929–6933
29. Blanco-Canosa JB, Medintz IL, Farrell D et al (2010) Rapid covalent ligation of fluorescent peptides to water solubilized quantum dots. J Am Chem Soc 132:10027–10033
30. Wang J, Xia J (2011) Preferential binding of a novel polyhistidine peptide dendrimer ligand on quantum dots probed by capillary electrophoresis. Anal Chem 83:6323–6329
31. Lu Y, Wang J, Wang J et al (2012) Genetically encodable design of ligand "bundling" on the surface of nanoparticles. Langmuir 28:13788–13792
32. Chen Z (2005) General biology, 2nd edn. Higher Education Press, Beijing
33. Tsien RY (1998) The green fluorescent protein. Ann Rev Biochem 67:509–544
34. Shaner NC, Campbell RE, Steinbach PA et al (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. Red fluorescent protein. Nat Biotechnol 22:1567–1572
35. Griffin BA, Adams SR, Tsien RY et al (1998) Specific covalent labeling of recombinant protein molecules inside live cells. Science 281:269–272
36. Hermanson G (2008) Bioconjugate techniques, 2nd edn. Academic Press, London
37. Antos JM, Francis MB (2004) Selective tryptophan modification with rhodium carbenoids in aqueous solution. J Am Chem Soc 126:10256–10257
38. Joshi NS, Whitaker LR, Francis MB et al (2004) A three-component mannich-type reaction for selective tyrosine bioconjugation. J Am Chem Soc 126:15942–15943
39. Tilley SD, Francis MB (2006) Tyrosine-selective protein alkylation using π-allylpalladium complexes. J Am Chem Soc 128:1080–1081
40. Wang L, Schultz PG (2005) Expanding the genetic code. Angew Chem Int Ed 44:34–66
41. Griffin BA, Adams SR, Jones J et al (2000) Fluorescent labeling of recombinant proteins in living cells with FlAsH. Methods Enzymol 327:565–578
42. Adams SR, Campbell RE, Gross LA et al (2002) New biarsenical ligands and tetracycteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J Am Chem Soc 124:6063–6076
43. Cao H, Chen B, Squier TC et al (2006) CrAsH: a biarsenical multi-use affinity probe with low non-specific fluorescence. Chem Commun 42:2601–2603
44. Cao H, Xiong Y, Wang T et al (2007) A red Cy3-based biarsenical fluorescent probe targeted to a complementary binding peptide. J Am Chem Soc 129:8672–8673

45. Gaietta G, Deerinck TJ, Adams SR et al (2002) Multicolor and electron microscopic imaging of connexin trafficking. Science 296:503–507
46. Halo TL, Appelbaum J, Hobert EM et al (2009) Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. J Am Chem Soc 131:438–439
47. Ojida A, Honda K, Shinmi D et al (2006) Oligo-Asp tag/Zn(II) complex probes as a new pair for labeling and fluorescence imaging of proteins. J Am Chem Soc 128:10452–10459
48. Nonaka H, Tsukiji S, Ojida A et al (2007) Non-enzymatic covalent protein labeling using a reactive tag. J Am Chem Soc 129:15777–15779
49. Nonaka H, Fujishima SH, Uchinomiya SH et al (2010) Selective covalent labeling of tag-fused GPCR proteins on live cell surface with a synthetic probe for their functional analysis. J Am Chem Soc 132:9301–9309
50. Keppler A, Gendreizig S, Gronemeyer T et al (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat Biotechnol 21:86–89
51. O'Hare HM, Johnsson K, Gautier A et al (2007) Chemical probes shed light on protein function. Curr Opin Struct Biol 17:488–494
52. Gautier A, Juillerat A, Heinis C et al (2008) An engineered protein tag for multiprotein labeling in living cells. Chem Biol 15:128–136
53. George N, Pick H, Vogel H et al (2004) Specific labeling of cell surface proteins with chemically diverse compounds. J Am Chem Soc 126:8896–8897
54. Yin J, Liu F, Li X et al (2004) Labeling proteins with small molecules by site-specific posttranslational modification. J Am Chem Soc 126:7754–7755
55. Janssen DB (2004) Evolving haloalnane dehalognease. Curr Opin Chem Biol 8:150–159
56. Los GV, Encell LP, McDougall MG et al (2008) Halo tag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 3:373–382
57. Janeway C (2001) Immunobiology, 5th edn. Garland Publishing, New York
58. Litman GW, Rast JP, Shamblott MJ et al (1993) Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. Mol Biol Evol 10:60–72
59. Hopp TP, Prichett KS, Price VL et al (1988) A short polypeptide marker sequence useful for recombinant protein identification and purification. Nat Biotechnol 6:1204–1210
60. Lata S, Gavutis M, Tampé R et al (2006) Specific and stable fluorescence labeling of histidine-tagged proteins for dissecting multi-protein complex formation. J Am Chem Soc 128:2365–2372
61. Guignet EG, Hovius R, Vogel H et al (2004) Reversible site-selective labeling of membrane proteins in live cells. Nat Biotechnol 22:440–444
62. Apostolovic B, Danial M, Klok HA et al (2010) Coiled coils: attractive protein folding motifs for the fabrication of self-assembled, responsive and bioactive materials. Chem Soc Rev 39:3541–3575
63. Yano Y, Yano A, Oishi S et al (2008) Coiled-coil tag-probe system for quick labeling of membrane receptors in living cells. ACS Chem Biol 3:341–345
64. Nakase I, Okumura S, Tanaka G et al (2012) Signal transduction using an artificial receptor system that undergoes dimerization upon addition of a bivalent leucine-zipper ligand. Angew Chem Int Ed 51:7464–7467

Chapter 2
Biological Active Antifungal Peptides

2.1 Introduction

The emergence of drug resistant fungal pathogens urgently calls for new generation of antifungal drugs with improved pharmacokinetics and pharmacodynamics. Antimicrobial peptides (AMPs) are mainly cationic and amphiphilic peptides composed of less than 50 amino acids, produced by diverse organisms for killing various kinds of invade bacteria, fungi, and viruses while have low cytotoxicity toward the organism themselves [1]. Due to their new mechanisms of microbicidal action and scarce of resistance, they have been recognized as a gold mine of antimicrobial drugs.

Defensin is a widely spread antimicrobial peptide within plants and animals; it possesses a compact 3 dimensional structure stabilized by several disulfide bridges [2]. Plant defensin, the innate guarding line against invaders, harbors antifungal activity, while showing limited toxicity to mammalian cell lines. Specifically, it was reported that defensins isolated from the white cloud beans possess both antifungal and antibacterial activities which even retained after trypsin treatment [3]. Although defensins are potential drug candidates, some obstacles need to be addressed: (1) the protein is difficult to separate; (2) the biosynthesis of defensins has low efficiency; and (3) in vivo activity and resistance still need to be improved. So, the semisynthesis of the biologically active form of the peptide holds promise.

While biologists put their efforts in solving crystal structures and biological activities of defensin proteins, more and more attention has been paid to the partial sequence of this protein family [4]. Several structural parameters, such as amphi-pathicity, hydrophobicity, hydrophobic momentum, alpha helicity, and the net positive charges, play an important role in antibacterial peptides' bioactivities [5]. Accordingly, partial sequence of the protein might still possess biological property

of the parental protein. So we decided to perform a structural-activity analysis of the whole sequence of defensin.

One of the plant defensin has been crystallized and was shown in Fig. 2.1 [6]. It is a macrocyclic protein containing three β sheets and one α helix, stabilized by four disulfide bonds. As total synthesis of defensin is not cost effective, we decided to fragment the protein into basic units based on the secondary structures. In the crystal structure, the first short β sheet is linked to a rigid α helix, and then to two adjacent β sheets (the yellow line represents the disulfide bond). And to simplify the structure, the 1st β sheet was deleted in the first round optimization, and then there were two optimizing strategies: one was to omit the 2nd β sheet to form a two disulfide bonded, α helix-β sheet domain (Peptide I) and the other one was the omit the α helix to form a β loop domain (Peptide II).
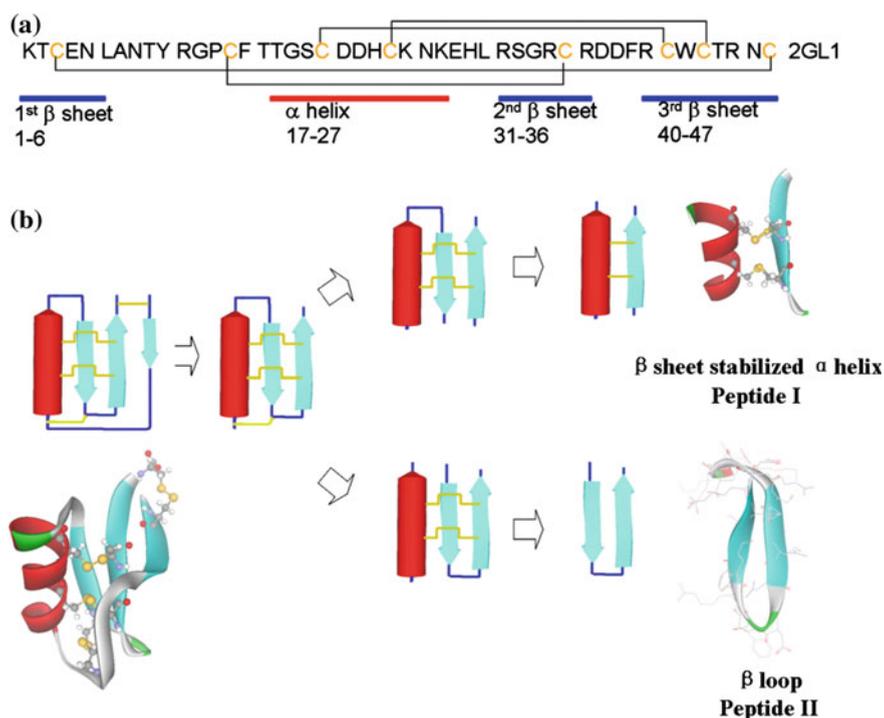
Fig. 2.1 Primary sequence and structure fragmentation of a plant defensin (PDB ID 2GL1). **a** Primary sequence of 2GL1. Cys are linked by disulfide bridges. *Red line* represents for the α helix and *blue line* represents for the β sheet. **b** 3-dimensional structure of 2GL1 and step by step structural fragmentation of the protein

2.2 Experimental Section

2.2.1 Peptide Synthesis and Purification

Peptides were manually synthesized based on standard FMOC solid phase peptide synthesis protocol using HBTU/HOBt activation strategy. Briefly, Rink Amide-ChemMatrix® resins (PCAS BioMatrix, Canada) with 0.1 mmol amine group were utilized in each synthesis. A solution containing Fmoc-protected amino acid, HBTU, HOBt, and DIPEA (with a ratio of 1:1:1:2 and 5 fold excess) in 2 mL was added to the resin and stirred for 30 min at RT. After confirming the completion of the coupling reaction by Kaiser Test, the resins were washed with DMF and deprotected in 5 mL 20 % piperidine in DMF (v/v) to remove the Fmoc group to allow the coupling of the following amino acid. Specifically in this chapter, Nα-Fmoc-S-acetaminomethyl-L-cysteine (Fmoc-Cys(Amc)-OH) and Nα-Fmoc-S-trityl-L-cysteine (Fmoc-Cys(Trt)-OH) were used to incorporate Cys with different side chain protecting groups. After the completion of the entire sequence, the terminal Fmoc was removed and the resins were washed by DMF, iso-propanol and n-hexane, respectively, and the resins were put in high vacuum overnight.

To every 100 mg resins, 2 mL final cleavage cocktail containing EDT, TIS, phenol, H_2O, and TFA (with a ratio of 1:2:2:2:33, v/v/w/v/v) was added. The cleavage reaction was allowed to proceed for 2 h at RT under stirring. After the resins were removed through filtration, ice-cold diethyl ester was added to the supernatant dropwise to precipitate the peptides. A final volume of 700 μL 50 % ACN in H_2O (v/v) was used to dissolve the peptide pellet. After being filtered through a 0.2 m filter, the peptide solution was injected to RP-HPLC (Shimadzu, DGU 20A5, Japan) equipped with a C18 column (Shimadzu, 250 L × 4.6, Japan). 0.1 % TFA in H_2O (v/v) and 0.1 % TFA in ACN (v/v) were used as the mobile phase A and B, respectively. For all the analytical HPLC trials, the total flow rate was set to be 1 mL/min and the B concentration raised from 0 to 95 % over 16 min following a linear gradient. For the purification of peptides in a larger scale by semi-prep HPLC columns (Grace, 218TP510, USA), the total flow rate was set to be 3 mL/min and the concentration of B raised linearly from 0 to 45 % over 24 min. The peptide peaks were collected, lyophilized, and validated by MALDI-TOF mass spectrometry analysis (Bruker, autoflex TOF/TOF, USA). For the conjugated peptide or peptide mixture, the mass spec was confirmed by LC-ESI. All the lyophilized peptides were stored at −20 °C.

2.2.2 Intermolecular Disulfide Bond Formation

6.3 mg alpha helix peptide (H-RC(Acm)WCTRNA-NH₂, pep 1), together with 8.5 mg bis(5-nitro-2-pyridyl) disulfide (DTNP, 4 equiv.) were dissolved in 1 mL acetic acid/H_2O mixture (3:1, v/v), stirring at RT, at different time points, aliquots of

the reaction mixture were quenched by 0.1 % TFA containing acetonitrile and the reaction was monitored by analytical RP-HPLC. After the completion of the reaction, the mixture were neutralized, filtered through a 0.2 μm membrane and purified by semi-prep HPLC and characterized by MALDI-TOF. 3.5 mg thiol activated α helix peptide (pep 2) was dissolved in 1 mL citric/phosphate buffer (pH 6), bubbled with argon for 5 min, to the solution was added 600 μL citric/phosphate buffer containing 4.4 mg β sheet peptide (H-TSNC(Acm)DDHCKNK-NH$_2$, equal equiv. pep 3) in the presence of argon to immediately give a bright yellow solution. The reaction was completed within 5 min as monitored by HPLC.

For the second disulfide bond formation, 1.1 mg heterodimeric peptide was dissolved in 400 uL acetic acid/H$_2$O mixture (4:1, v/v), to the solution was added 100 μL acetic acid/H$_2$O mixture containing 1.4 mg I$_2$, the reaction was allowed to proceed at RT with stirring for 3 h and quenched by the addition of 310 μL H$_2$O. And excess I$_2$ was reduced by the addition of 400 μL 10 mM ascorbic acid to avoid further oxidation. The reaction was also monitored by HPLC and characterized with LC-MS.

2.2.3 Intramolecular Disulfide Bond Formation

5.5 mg peptide IV was dissolved in 1 mL NaHCO$_3$/Na$_2$CO$_3$ buffer (0.1 mM, pH 9.4) bubbled with air and shaking at RT for 8 h, and the reaction was monitored by HPLC, after the completion of the reaction, the peptide was purified and lyophilized.

2.2.4 Biological Activity Assay

The assay of the synthetic antifungal peptides for antifungal activity toward *Candida Albicans*, which is human pathogen, was carried out in 10 cm petri dish containing 10 mL of potato dextrose agar. Fungal were incubated in 10 mL of nutrient broth in a thermal shaker for 12 h at 37 °C, and then 5 mL of this fungal suspension was transferred to 50 mL of nutrient broth and incubated for another 3–6 h in order to shift bacterial growth to the midlogarithmic phase. The fungal suspension was then centrifuged at 2000 g for 10 min, and the fungal pellet was resuspended in 1 mL normal saline and the solution was scraped on the petri dish. Sterile blank paper disks (0.625 cm in diameter) were placed at a distance of 0.5 cm away from the rim of edge of the petri dish and to the paper was dropped aliquots of 1 mg/mL antifungal peptides containing PBS buffer. And the petri dishes were placed in a 37 °C incubator to allow the growth of the fungal colonies. PBS only was conducted as the negative control and amphotericin was used as the positive control.

For the more accurate activity screening based on the liquid medium fungi culture method, peptides were dissolved in 10 mM PBS (pH 7.4) to give a final concentration of 1 mg/mL, the peptide solution was scraped on the agar surface and followed by the addition of the activated fungal medium as described above. The

petri dishes were then put into the 37 °C incubator and after 24 h the numbers of fungal colonies were counted.

2.3 Results and Discussion

2.3.1 Synthesis and Characterization of Peptide I

Based on the crystal structure (PDB ID 2GL1), the α helix and β sheet peptides need to be joined together through two disulfide bonds to form a hetrodimer. To introduce two disulfide bridges, the oxidations of thiol groups were arranged in a step by step manner to avoid nonspecific side products. As a result, the thiol groups in the corresponding Cys side chains were protected by different protecting groups so that they could be removed at the desired stages for selective oxidation (Fig. 2.2). And this strategy has found broad application in total synthesis of many difficult peptides and proteins [7].

In our strategy, trityl group was utilized to protect the first pair of Cys side chian, which could be cleaved by the TFA treatment. To increase the reactivity, as well as the specificity of the disulfide bond formation, DTNP was used to activate the thiol group in the β sheet peptide chain. The reaction was monitored by RP-HPLC (Fig. 2.3). A peak with the retention time of 17.4 min emerged and gradually increased as the reaction proceeded, accompanied by the decrease of the peak represented for pep 1. The reaction was completed in about 20 h and the intermediate peptide was purified by semi-prep HPLC. The preactivated pep 3 was mixed with the α-helix peptide in the presence of argon for a disulfide exchange reaction to give the intermediate pep 4. This reaction was completed within 1 min with good purity as monitored by HPLC. The acetaminomethyl (Acm) protecting group was removed by I_2 oxidation, with a simultaneous formation of the second disulfide bridge. However, the reaction gave complicated products as monitored by RP-HPLC. Over-oxidation might be the reason. Nevertheless, the desired product was found in the product as confirmed by LC-MS. The I_2 oxidation reaction needs further improvement. Therefore, pep 4 as well as the I_2 oxidation product was submitted to antifungal activity screening.

2.3.2 Synthesis and Characterization of Peptide II

It was reported that the β-loop structure containing β sheets β_2 and β_3 might harbor the antibacterial activity of defensins [8–10]. But, Cys seems to be not essential for the biological activity [11]. So we designed a small peptide library, based on the 2nd and 3rd β sheets sequence as the parental sequence (Peptide 1, sequence shown in Table 2.1, structure shown in Fig. 2.4). Given the fact that the Cys contributes little to the biological activity and with the purpose of preventing the formation of
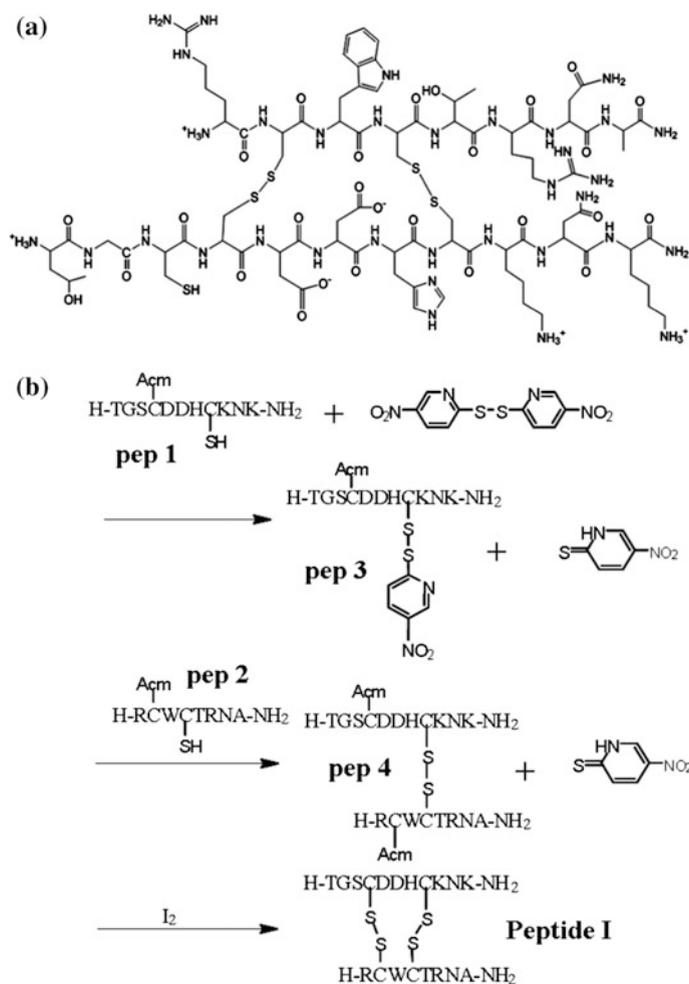
Fig. 2.2 Structural illustration of Pep I and the synthesis scheme. **a** Molecular structure of the two disulfide bonded β sheet stabilized α helix peptide. **b** Synthesis scheme of the β sheet stabilized α helix peptide

nonspecific disulfide bonded, all the three Cys were changed to Ala to yield peptide 2. As seen in the peptide sequence, two Asp were localized on the top corner of the β loop, and this structure was further stabilized by the hydrogen bonding interactions and the cation-π interactions. The two Asp were replaced by two continuous D-Pro and L-Pro, with all the Cys mutated into Ala to yield peptide 3. The sequence $_D$P$_L$P can force the peptide backbone to adopt a beta turn structure [8]. Furthermore, it might also reduce the negative charges of the peptide. In peptide 4, two cysteines can form an intramolecular disulfide bond to enhance the stability of the beta loop secondary structure.

Fig. 2.3 HPLC traces of the activation of Pep 1 and the conjugation of Pep 4.
a Activation reaction of pep 1 by DTNP as monitored by HPLC, the purified product was shown in green line.
b Conjugation reaction between pep 2 and pep 3 as monitored by HPLC, pep 4 was the purified product
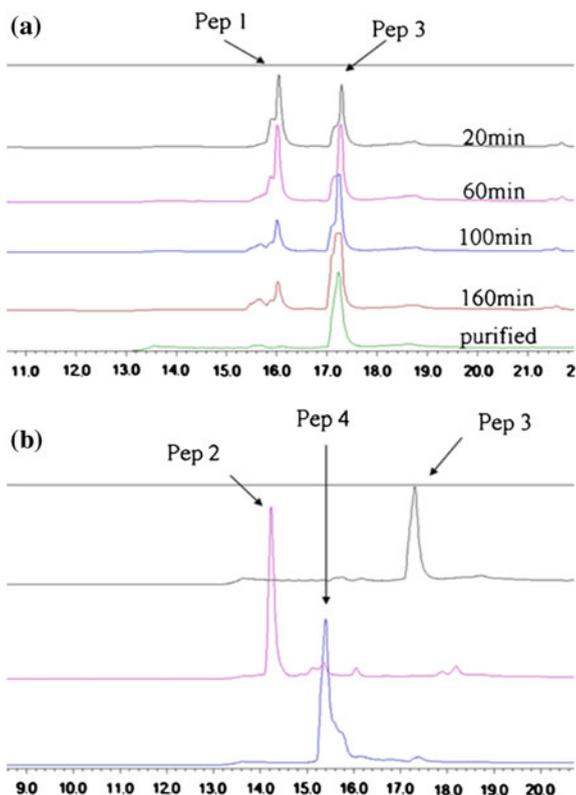
Table 2.1 Peptide sequences and mass spectral data

No.	Sequence[a]	M_{calc}	MALDI-TOF MS [M+H]$^+$	ESI MS [M+H]$^+$
1	H-KNKEHLLSGRCRDDFRCWCTR-NH$_2$	2622.248	2623.317	–
2	H-KNKEHLLSGRARDDFRAWATR-NH$_2$	2526.332	2527.459	2528
3	H-KNKEHLLSGRAR^{D}P^{L}PFRAWATR-NH$_2$	2490.391	2491.467	2491.6
4	H-KNKEHLLSGRCR^{D}P^{L}PFRCWATR-NH$_2$	2552.327	2553.335	–

[a]^{D}P represents for D-Pro, ^{L}P represents for L-Pro and for peptide 4, an intramolecular disulfide bond is formed

2.3.3 Structure–Activity Relationship Studies

All four peptides were prescreened for their antimicrobial activity against *Candida Albicans*. Peptides 3 and 4 inhibited the growth of the fungus, but not peptides 1 and 2. In each petri dish the fungus would grow on the paper, if the area does not contain inhibitory activity; or, if the area contains inhibitory molecules, the microbes will avoid it. As shown in Fig. 2.5, a clear blank rim could be observed in
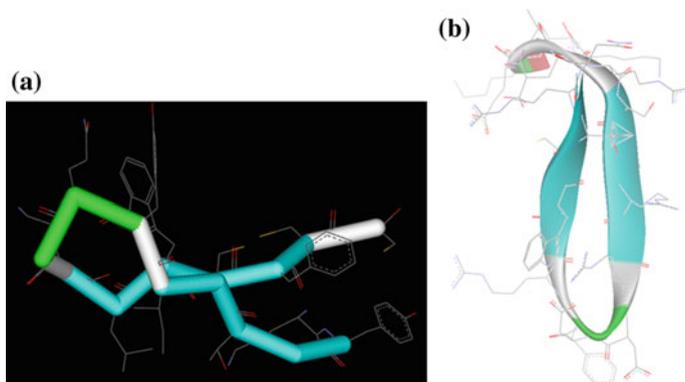
Fig. 2.4 3D structure of the β-loop peptide. Amino acids in green represent the turn structure. Basic amino acids are labeled in *blue*. **a** Side view. **b** Top view
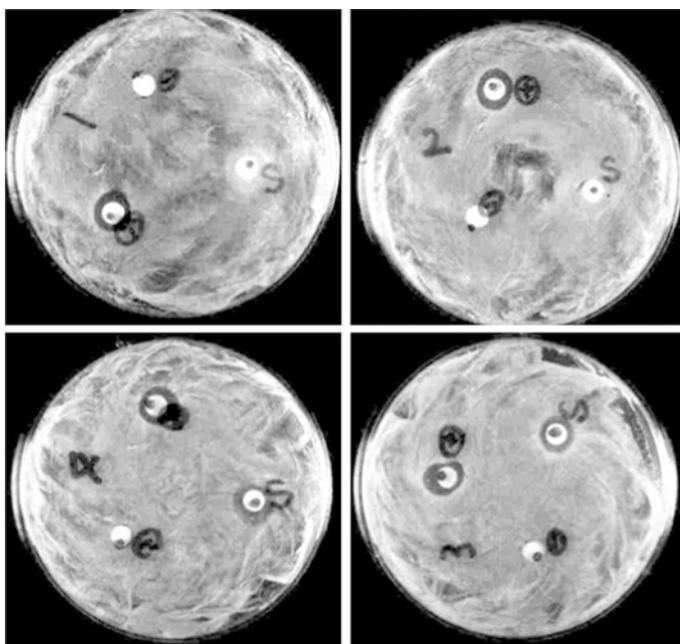
Fig. 2.5 Photographs of bioactivity screening of peptides *1–4* (Clockwise, start from top left corner). *S* stands for peptide sample. + stands for positive control. − stands for negative control

peptide 3 and peptide 4, but not in plates 1 and 2 which contain peptides 1 and 2, respectively.

To be more quantitative, peptide solutions were mixed with fungal culture and grown overnight at 37 °C in a thermal shaker. 200 μL of the mixture was then
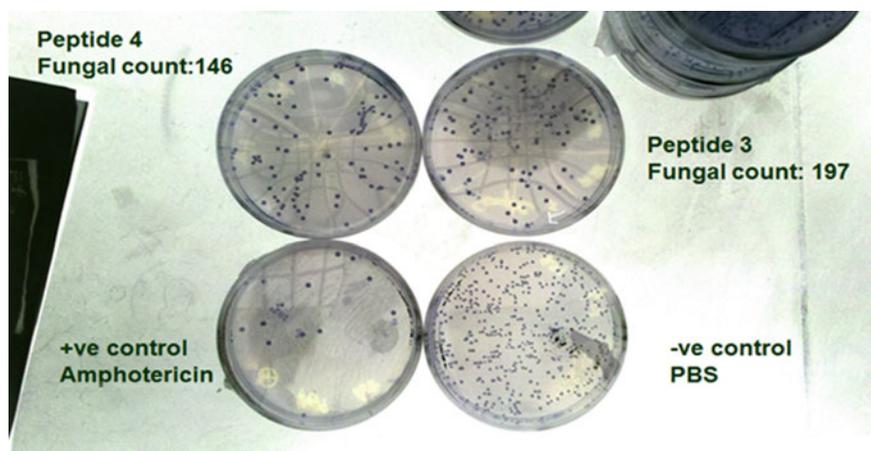
Fig. 2.6 Photographs of liquid fungal activity assay. Amphotericin and PBS buffer were used as positive and negative controls, respectively. Fungal cloning numbers were counted. (Clockwise, starting from top left corner, peptide 4, peptide 3, negative control and positive control)

spread on agar plates. After incubation at 37 °C for 24 h, colonies emerge, and were counted. Compared to the PBS negative control, plates containing peptide 3 and peptide 4 showed fewer colonies (Fig. 2.6), meaning that both peptides exhibited antifungal activities. Relatively, peptide 4 showed higher activity as evidenced by even fewer colonies (146 vs. 197).
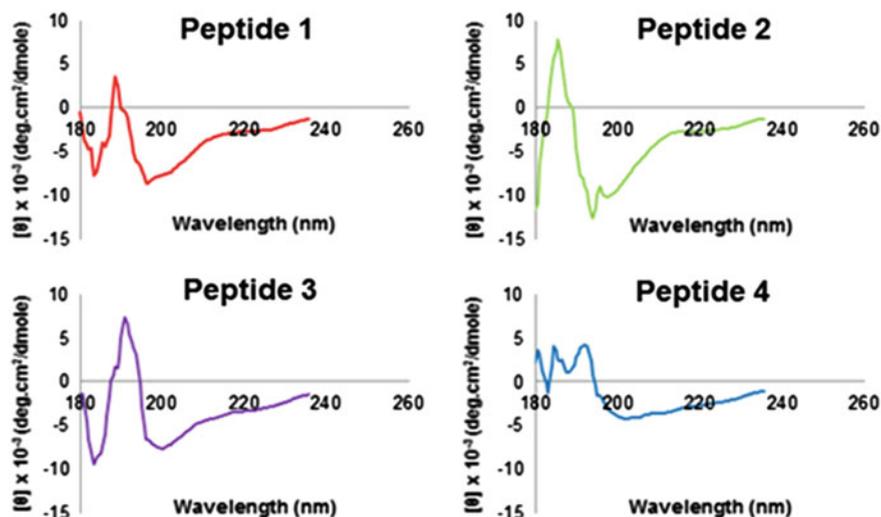
Fig. 2.7 CD spectra of the 4 peptides (Measured at the concentration of 50 μg/mL in 10 mM PBS, pH 7.4, the curve was the average value of ten scans)

We then utilized circular dichroism spectrum to characterize the secondary structure of the four peptides to find a correlation between secondary structure and antifungal activity (Fig. 2.7). Peptide 1 and peptide 2 mainly adopted the random coil structure while peptide 3 and peptide 4 possessed the beta loop structure due to a minimum band at 205 nm.

2.4   Conclusion

Two structural motifs, a β sheet-α helix unit stabilized by two disulfide bonds and the β loop of the white cloud bean defensin have been designed based on the crystal structure. Synthesis of both structures was attempted through solid phase peptide synthesis. Iodine oxidation of Acm protected cysteines failed to yield a clean product, so still it needs further optimization. Peptides mimicking the beta loop manifested desired antifungal activity against *Candida Albicans*. This work paved the road to further optimize the beta loop sequence to search for potential antifungal activity. We only examined the fungal strain *Candida Albicans*; in future experiments, we aim to include other fungi to examine whether peptides 3 and 4 have specificity toward different strains of fungi. We will further optimize the sequence of peptides 3 and 4 by including more positively charged and aromatic amino acids to increase its membranolytic activity. The mechanism of the antifungal activity is still under investigation; most likely the peptides disrupt the membrane of the cells. One should note that the biological activity of the defensin fragments might have significantly deviated from their parental peptide. In other words, new antimicrobial activity might be discovered in further structure–function relationship studies.

References

1. Zasloff M (2002) Antimicrobial peptides of multicellular organisms. Nature 415:389–395
2. Tomas G (2003) Defensins: antimicrobial peptides of innate immunity. Nat Rev Immunol 3:710–720
3. Wong JH, Zhang XQ, Wang HX et al (2006) A mitogenic defensin from white cloud beans (*Phaseolus vulgaris*). Peptides 27:2075–2081
4. Liu H, Boudreau M, Zheng J et al (2010) Chemical synthesis and biological activity of the neopetrosiamides and their analogues: revision of disulfide bond connectivity. J Am Chem Soc 132:1486–1487
5. Ahn H-S, Cho W, Kim J-M et al (2008) Design and synthesis of cyclic disulfide-bonded antibacterial peptides on the basis of the α helical domain of Tenecin 1, an insect defensin. 16:4127–4137
6. Lin K-F, Lee T-R, Tsai P-H et al (2007) Structure-based protein engineering for α-amylase inhibitory activity of plant defenisn. Proteins 68:530–540
7. Rajarathnam K, Lewis IC, Sykes BD et al (1999) ^{1}H NMR studies of interleukin 8 analogs: characterization of the domain essential for function. Biochemistry 38:7653–7658

8. Taylor K, Barran PE, Dorin JR et al (2008) Structure–activity relationships in β-defensin peptides. Bioploymers 90:1–7

9. Lee KH, Hong SY, Oh JE et al (1998) Identification and characterization of the antimicrobial peptide corresponding to C-terminal β-sheet domain of tenecin 1, an antibacterial protein of larvae of *Tenebrio molitor*. Biochem J 334:99–105

10. Cabral KM, Almeida MS, Valente AP et al (2003) Production of the active antifungal *Pisum sativum* defensin 1(*Ps*d1) in *Pichia pastoris*: overcoming the inefficiency of the STE13 protease. Protein Expr Purif 31:115–122

11. Schaaper WM, Posthuma GA, Plasman HH et al (2001) Synthetic peptides derived from the β$_2$-β$_3$ loop of *Raphanus sativus* antifungal protein 2 that mimic the active site. J Pept Res 57:409–418

Chapter 3
Protein Ligands Engineering

3.1 Introduction

Protein-nanoparticle assemblies as a type of hybrid biomaterials have found increasingly wide-ranging uses in catalysis, tissue imaging, biosensing, and cell targeting [1–4]. The inorganic nanoparticle cores grant the assemblies favorable physical properties such as optical, electrical and magnetic properties that organic or biological molecules normally do not possess, whereas protein ligands displayed on the periphery mediate the interaction between the particle and the biological environment [5–8]. As a ligand to functionalize the surface of nanoparticle, a protein is notably different from a small molecule, or a synthetic polymer. The first distinction lies in its size: proteins have similar dimensions as nanoparticles, with diameter often ranging between 3 and 6 nm, comparable to that of a nanoparticle. Therefore, while small molecules and polymers form a self-assembled monolayer on the surface of particles, monomeric proteins binds to quantum dots (QDs, as one example of nanoparticles) with a low stoichiometry around 16:1 (ligand:particle ratio) [9–11]. Secondly, featuring sophisticated three-dimensional structures, proteins are structurally asymmetric in shape, chirality, and chemical properties. These features are furthermore highly engineerable, thanks to the great advancement of recombinant technology and structural biology in recent decades. Therefore, one could base on the crystal structure of a protein to tailor-make protein ligands that have particular intermolecular interactions to affect specific controls on the properties of protein-nanoparticle assemblies, a degree of freedom that is difficult to achieve using synthetic small molecules or polymers [12–17].

Two properties are generally desirable in generating protein-functionalized nanoparticle assemblies. First, achieving a stable protein-nanoparticle assembly has practical significance. Weakly functionalized nanoparticles likely undergo a ligand exchange process on the surface in biological fluids or in the environment, which may lead to an alteration of the biological property of the particle assemblies as well as adverse effects. For example, plasma proteins can form a corona around particles,

the exact nature of which is still under investigation [18–23]. As protein binding on the surface of nanoparticles is often driven by electrostatic, hydrophobic, and coordination interactions between plasma proteins and the particles, a stable conjugate between ligand and nanoparticles could impede undesired ligand exchange. A self-assembly process driven by metal−hexahistidine peptide interaction has been developed as an efficient and site-directed method to functionalize the surface of CdSe–ZnS QDs with protein ligands [6, 9–11]. Still, protein-QD complexes formed through metal-hexahistidine interaction could be interfered through ligand exchange with histidine, cysteine, or methionine-rich proteins, or millimolar concentrations of thiol-containing molecules such as glutathione (GSH) [24–26]. The second capability is to control the stoichiometry of ligand-nanoparticle ratio in the complex [27–29]. Low-valency surface functionalization of nanoparticles has been achieved through several strategies. For example, reducing the size of the nanoparticle facilitates the fabrication of 1:1 protein-QD assembly [27]. Instead of utilizing ultrasmall nanoparticles, we can use large protein as ligands to effect the control of stoichiometry.

In this chapter, we adopted structure-guided protein design to engineer a collection of fluorescent proteins containing hexahistidine tags (histag, the QD-binding sequence) of different geometries and spatial distributions: monomers, a hinged dimer, a cross-shaped tetramer, together with the previously reported bundled tetramer. This collection allowed us to investigate for the first time how geometry of protein ligands affects protein-QD assembly. We further validated our discovery by a de novo design of a nanobelt protein which can transition from a random coil to a coiled-coil in the presence of a companion peptide.

3.2 Experimental Section

3.2.1 Plasmid Construction

The DNA fragment coding for mCherry protein sequence was amplified by PCR using primers mCherry-F (5′-ATTAGAATTCATGGTGA-GCAAGGGCGAGGA-3′) and mCherry-R (5′-TATACTCGAGTTACTTGTACAGCTCGTCCATG-3′) based on the vector pQL81-mCherry and sub-cloned into the *EcoR* I and *Xho* I sites of pET21a vector to yield pET21a-MC. An N terminal His tag, together with a linker sequence (KALEAQKQK) was introduced to the N terminus of TIP-1 DNA sequence by PCR using primers TIP-F (5′-ATATATACATATGCATCACCATCACCATCACAAA GCTCTTGAAGCTCAGAAACAGAAA-ATGTCCTACACCCCGGGCCA-3′) and TIP-R (5′TATAGAATTCAGACAG-CATGGACTGCTGTACA-3′) based on the ULD-TIP-1 plasmid (a gift from Prof. Zhimou Yang of Nankai University), the DNA fragment was then sub-cloned into *Nde* I and *EcoR* I sites of the pET21a-MC to yield pET21a-TIP-1-MC. Primers, mCherry-F and mCherry-R′ (5′-TATACTCGAGCT TGTACAGCTCGTCCATGCC-3′) were used to amplify mCherry protein, and the

DNA fragment was inserted into the *EcoR* I and *Xho* I sites of pET21a to yield pET21a-MCS. DNA fragment coding for ULD protein sequence was amplified from ULD-TIP-1 plasmid and sub-cloned into the *Nde* I and *EcoR* I sites of pET21a-MCS to yield pET21a-ULD-MCS. DNA coding for nanobelt protein was purchased from commercial supplier (Genscript, Nanjing) and sub-cloned into *BamH* I and *Hind* III restriction enzyme sites of pET21a vector to yield pET21a-NB using primers: NB-F (5′-GGATCCGGCCCGCATAAAATTGCGCAACTGAA-3′) and NB-R (5′-AAGC TTTTACAGCAAAGCAGAGATTTTGTGC-TCCAGG-3′), and followed by the insertion of the DNA fragment coding for mCherry at *Nde* I and *BamH* I sites to yield pET21a-MC-NB. All the plasmids constructed were confirmed by DNA sequencing (BGI, Shenzhen) (see Appendix 3.1).

3.2.2   Protein Expression and Purification

Plasmids were transformed into *Escherichia coli.* BL21 (DE3) competent cells. Single colony was grown overnight to give a starter culture. The starter culture was then used to inoculate a larger volume of LB medium (1 % v/v) and grown at 37 °C until OD_{600} reached 0.4–0.6 and IPTG was then added. After grown at 16 °C overnight, cells were harvested and resuspended in lysis buffer, sonicated, and centrifuged. Proteins were purified from cell supernatant using a HisTrap HP column on an ÄKTA prime FPLC (GE Healthcare, USA) with a linear increase of imidazole concentration from 10 to 500 mM. The protein eluent was buffer exchanged to storage buffer and stored at −20 °C until further use. Protein concentration was measured by Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA) based on an absorption coefficient constant ε_{587} of 72,000 cm^{-1} M^{-1} for a single mCherry moiety. The purity of the proteins was confirmed by SDS-PAGE (see Appendix 3.2).

3.2.3   Peptide Synthesis, Purification, and Characterization

Peptides were manually synthesized based on standard FMOC solid phase peptide synthesis protocol on Rink Amide-ChemMatrix® resins (PCAS BioMatrix, Canada). Briefly, coupling steps were done using a solution containing Fmoc-protected amino acid, HBTU, HOBt, and DIPEA in DMF (1:1:1:2:5, w/w/w/v/v). Deprotection of Fmoc was done in 20 % piperidine in DMF (v/v). After the completion of the peptide sequence, peptides were cleaved from the resin by a solution of EDT, TIS, phenol, H_2O, and TFA (1:2:2:2:33, v/v/w/v/v) and then precipitated by ice-cold diethyl ester. The peptide pellet was dissolved in 50 %

acetylnitrile in H_2O and purified by RP-HPLC (Shimadzu, DGU 20A5, Japan) equipped with a semi-prep C18 column (Shimadzu, 250L× 4.6, Japan) or an analytical column (Grace, 218TP510, USA). Peptide peaks were collected, lyophilized and validated by MALDI-TOF mass spectrometry analysis (Bruker, autoflex TOF/TOF, USA).

3.2.4   Preparation of TIP-1-MCherry Dimer

TIP-1 binding peptide CGGWRESAI was dissolved in 0.1 mM $Na_2CO_3/NaHCO_3$ buffer (pH 9.51) to a final concentration of 100 µM and agitated at R.T to form the dimer under monitor by HPLC. The dimer $(CGGWRESAI)_2$ was purified by semi-prep HPLC. Dimer peptide at 50 µM was mixed with TIP1-mCherry monomeric protein (10 µM) in pH 7.5 HEPES buffer at 4 °C overnight. The TIP1-mCherry dimer protein was purified and excess free peptide was removed by FPLC on a size-exclusion column (Superdex^{TM} 10/300GL, GE healthcare, USA).

3.2.5   Preparation of Water-Soluble QDs

GSH stabilized QDs were synthesized based on the previous protocols [24–26]. Briefly, oil-soluble CdSe–ZnS core-shell QDs purchased from (JIAYUAN Quantum Dots. Co. Ltd. Wuhan, China) was dissolved in trichloromethane. A basic glutathione solution (14.2 mg GSH and 5 mg KOH in methanol) was added with vigorous shaking for 2 h. Water-soluble QDs were then extracted using borate buffer (10 mM, pH 9.0) and washed. Fluorescent spectra were obtained on a Hitachi F-7000 fluorescence spectrophotometer in a quartz cuvette with an optical path length of 1 cm. The excitation wavelength was set to be 420 nm and the fluorescent emission spectrum was monitored from 500 to 700 nm.

3.2.6   Agarose Gel Electrophoresis

Proteins were preincubated with QDs at different ratios in a 20 µL reaction scale. 2 µL of DNA loading dye (TAKARA, Hong Kong) was added to the solution prior to loading to a precooled 3 % agarose gel. During the electrophoresis, the instrument set was maintained at low temperature by adding iced TAE cubic to prevent QDs from thermal quenching. The gel was illuminated under a UV–Vis imaginer acquire images.

3.3 Results and Discussion

3.3.1 A TIP1 Fusion Protein and Its Dimer as QD Ligands

Through a protein ligand GCN-mCherry that can form a bundled tetramer on the surface of QDs, we have elucidated that local clustering of four QD-binding sequences (histags) greatly enhances the stability of the QD-protein assembly [26]. Here we plan to expand the scope of the previous research and thoroughly explore how protein–protein interaction influences protein-QD assembly. We drew inspiration from the structure of a small protein, Tax-interacting protein-1 (TIP1) and utilized TIP1 as our first model protein. Belonging to the family of PDZ domains, TIP1 binds with a peptide WRESAI very strongly with a low dissociation constant K_d of 8.5 nM (Fig. 3.1a) [30, 31]. Such a protein–peptide binding interaction allows us to investigate the monomer–dimer effect.

We first designed a fusion protein TIP1-mCherry, with a fluorescent mCherry protein fused to the C terminus of TIP1 and a histag to the N terminus. TIP1-mCherry fusion protein will exist as a monomer in solution; a dimeric peptide ligand will drive the formation of a stable H-shaped dimer (Fig. 3.1a). The TIP1-mCherry dimer will have two histags at the N terminus to bind to QDs, as compared with only one in TIP1-mCherry monomer. The dimeric peptide (CGGWRESAI)$_2$ was synthesized by air oxidation of a monomeric peptide CGGWRESAI (Fig. 3.1b and Appendix 3.3) through a disulfide bond. Incubation of (CGGWRESAI)$_2$ with TIP1-mCherry resulted in a pure TIP1-mCherry dimer, owing to the high affinity between the WRESAI and TIP1 (Fig. 3.1c).

A low stoichiometry indicates that steric hindrance between protein ligands on the surface of QDs prevents more proteins from binding to the particle at the same time. Compared with histag-mCherry which has a molecular weight (MW) of about 30 KDa, TIP1-mCherry is significantly larger, having a MW about 45 KDa. Intuitively, proteins with large size and high rigidity will be less likely to share the space. Protein size is thereby a primary factor to the stoichiometry of protein-QD assembly.

Histag-containing proteins can assemble with glutathione stabilized ZnS–CdSe QDs (with an average diameter of 3.5 nm and a maximal emission wavelength of 565 nm) through metal-affinity driven assembly between histidines and QDs. This process can be readily monitored by the emergency of FRET signal at 610 nm (the emission of mCherry) and the decrease of the fluorescence (FL) signal at 565 nm (the emission of QDs) (Fig. 3.2A) [10, 26, 33]. Both TIP1-mCherry monomer and TIP1-mCherry dimer showed a low saturation stoichiometry of around 4:1 with QDs (protein counted as monomers), whereas a typical monomeric protein such as histag-mCherry bind to QDs with a saturation stoichiometry of 16:1 (Fig. 3.2B–D) [11, 26].
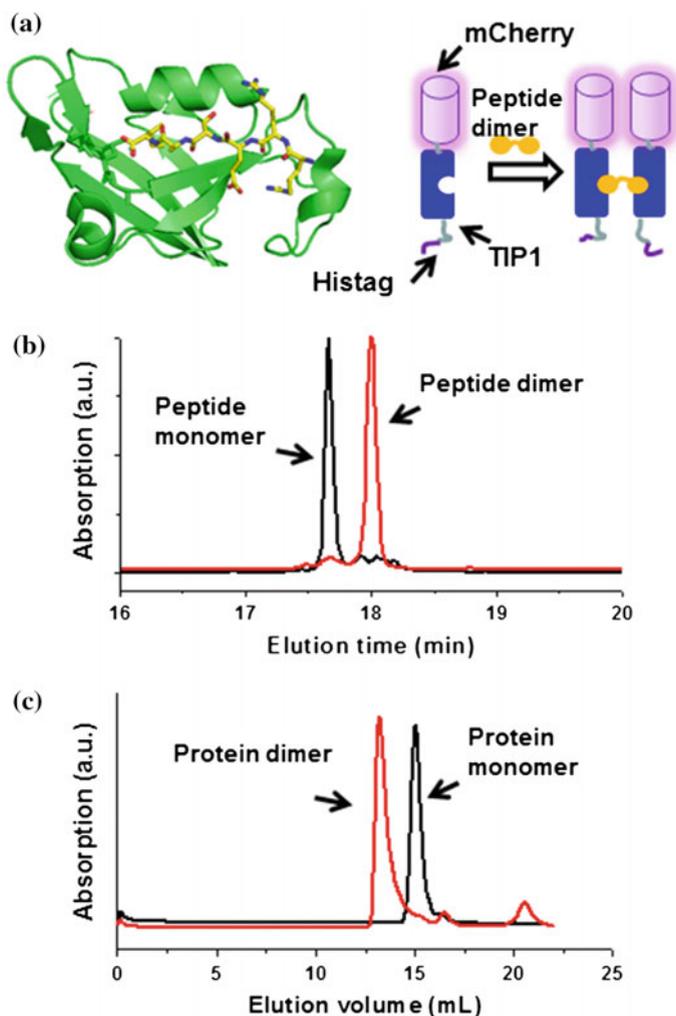
Fig. 3.1 Generation of TIP1-mCherry and its dimer. **a** Crystal structure of TIP1-RRESAI complex (PDB Id 3GJ9), and a schematic illustration of the formation of TIP1-mCherry dimer through peptide–protein interaction. **b** HPLC traces (with detection at UV 215 nm) of the peptides CGGWRESAI monomer (*black*) and (CGGWRESAI)2 dimer (*red*). **c** Size-exclusion chromatography traces of TIP1-mCherry monomer (*black*) and TIP1-mCherry dimer (*red*). The dimer solution contains a slight excess of peptide (CGGWRESAI)$_2$. Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

We then compared the stability of the protein-QD assembly in a ligand displacement assay. Widely used as a competitive ligand for histagged protein in immobilized metal-ion affinity chromatography (IMAC) for protein purification [34], imidazole competes with the coordination sites of ZnS on QDs and drives off
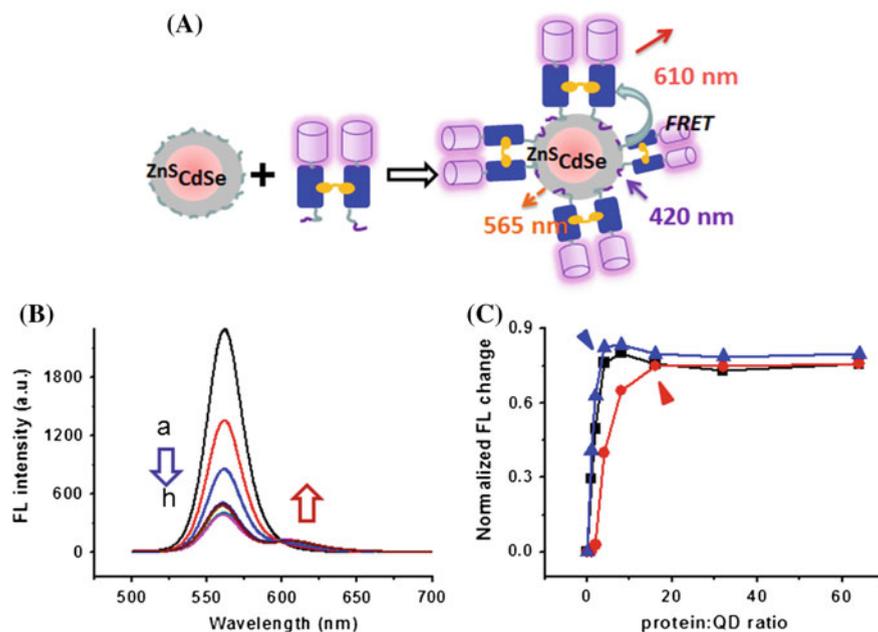
Fig. 3.2 Assembly of TIP1-mCherry and TIP1-mCherry dimer with QDs. **A** A schematic illustration showing that histagged proteins displace GSH ligands on the surface of QDs, and result in FRET signals. **B** FL spectra of QDs with an increasing ratio of TIP1-mCherry:QD. *a* QD alone, *b* 1:1, *c* 2:1, *d* 4:1, *e* 8:1, *f* 16:1, *g* 32:1, and *h* 64:1. **c** Concentration dependent decrease of the QD emission signal at 565 nm upon binding with increasing ratio of histag-mCherry (*red*), TIP1-mCherry (*blue*) and TIP1-mCherry dimer (*black*) (*arrows* indicate saturation). Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

the bound histagged proteins (Fig. 3.3A). Protein-QD assemblies were preformed at 16:1 protein:QD ratio, and then incubated with 200 mM or 400 mM imidazole solutions. Clearly, 400 mM imidazole resulted in 50 % decrease in the fluorescent intensity at 565 nm for TIP1-mCherry, whereas only 30 % decrease was seen for TIP1-mCherry dimer (Fig. 3.3B, C). This indicates that the dimer binds with a stronger affinity with QDs than TIP1-mCherry monomer, due to multivalency effect. Compared with previously reported bundled tetramer GCN-mCherry [26], the affinity enhancement seems to be lower, as imidazole could still compete off some bound dimer protein. This observation shows that two anchors with high flexibility in TIP1-mCherry dimer is less optimal to stably bind with QDs, whereas a bundled tetramer with high rigidity in GCN-mCherry provides a much higher binding affinity.
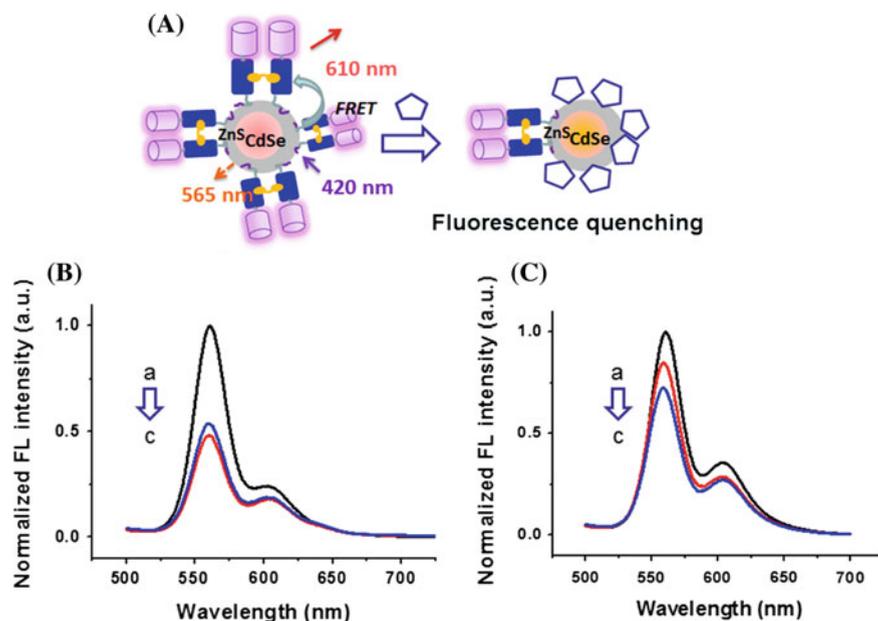
Fig. 3.3 Im displacement of protein-QD assemblies. **A** Imidazole competes with the surface-bound histag-containing proteins and leads to fluorescence trenching. **B** FL quenching of QD-TIP1-mCherry complexes in 200 mM (*red*) and 400 mM (*blue*) Im solutions. **C** FL quenching of QD-(TIP1-mCherry dimer) complexes in 200 mM (*red*) and 400 mM (*blue*) Im solutions. *Black traces* show QD-protein assemblies in the absence of Im. Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

3.3.2 Tetrameric Proteins with Different Spatial Distribution of QD-Binding Sites

Increasing the protein aggregation states from dimer to tetramer allows for a more detailed examination of how spatial distribution of histags affects QD-protein assembly. The ubiquitin-like domain (ULD) protein is known to adopt a "cross-shaped" tetrameric conformation in its native state, with the N terminus pointing toward the center and the C terminus pointing outside [30, 35]. This then represents a tetramer structure with maximal end-to-end distances. Fusing a mCherry protein containing a C terminal histag at the C terminus of ULD protein generates a cross-shaped tetramer with four histags at the periphery of complex, with a distance of about 5–6 nm (Fig. 3.4) [35]. The GCN-mCherry protein that we previously reported forms a bundled tetramer through a parallel coiled-coil sequence. The histags in GCN-mCherry tetramer are in close proximity, instead of being scattered. Therefore, these two protein tetramers feature the same amount of histags, but represent two extremely in term of histag–histag distance:
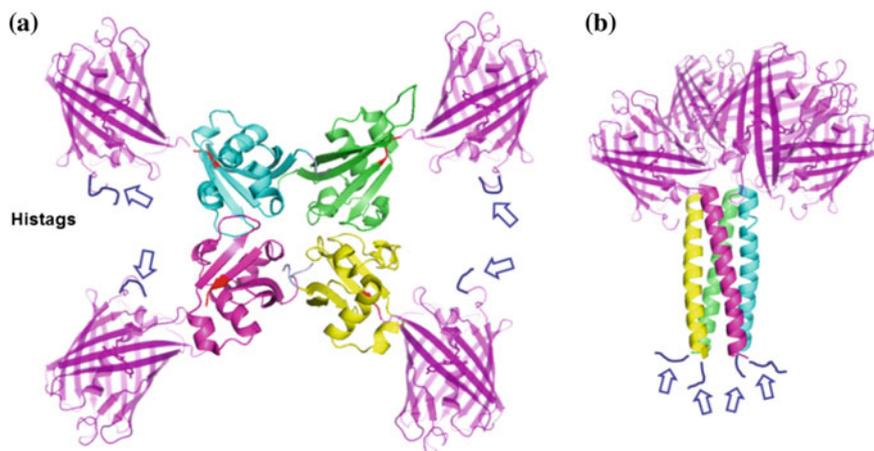
Fig. 3.4 Comparison of the structures of tetrameric protein ULD-mCherry (**A**) and GCN-mCherry (**B**). The structures of ULD and GCN are depicted based on PDB Id 3TU0 and 1GCL, respectively. Arrows indicate the positions of histags. Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

ULD-tetramer has the largest histag–histag distance, whereas GCN-mCherry has the smallest possible histag–histag distance.

Both tetrameric proteins bind with QDs (Fig. 3.5A, B). Although GCN-mCherry showed a normal stoichiometry of 16:1 (protein monomer:QD ratio), ULD-mCherry exhibited a slightly lower stoichiometry (estimably around 10) (Fig. 3.5C). The QD assembly of ULD-mCherry is also less stable than the assembly with GCN-mCherry as measured by imidazole displacement (Fig. 3.5D, E). Due to the rigid backbone structure and the large distance between each QD-binding sequence, the assemblies most likely assume a heterogeneous mixture of various multi-particle species. We also resolved the QD-protein assemblies in agarose gel by electrophoresis. Markedly, QD-GCN-mCherry assemblies showed discrete peaks at 8:1 and 16:1 ligand:QD ratios; however, QD-ULD-mCherry assemblies showed a smear band, indicating a mixture of heterogeneous compositions (Fig. 3.5F). This experiment manifested that besides aggregation state, spatial distribution of QD-binding sites also determines the protein-particle assembly mode.

3.3.3 Structural Transition of a Nanobelt Protein Ligand

Besides engineering protein ligands based on known structures, we took one step further to examine whether one could design a protein that can undergo structural transition from a flexible structure to a rigid one, and how such transition effects a
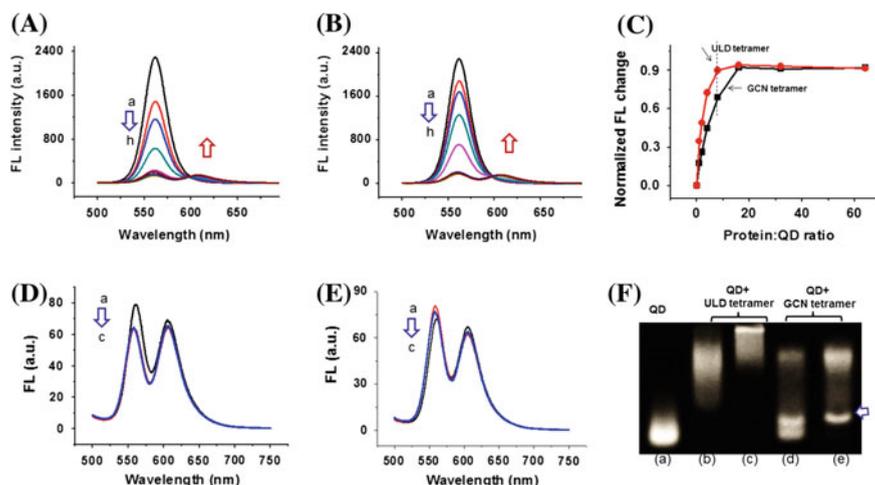
Fig. 3.5 Assembly of the tetrameric proteins, ULD-mCherry and GCN-mCherry with QDs. **A** FL spectra of QDs with an increasing ratio of ULD-mCherry:QD. **B** PL spectra of QDs with increasing ratio of GCN-mCherry dimer:QD. *a* QD, *b* 1:1, *c* 2:1, *d* 4:1, *e* 8:1, *f* 16:1, *g* 32:1, and *h* 64:1. **C** Concentration dependent decrease of the QD emission signal at 565 nm upon binding with increasing ratio of proteins (*arrows* indicate saturation). **D** FL quenching of QD-ULD-mCherry complexes in 200 mM (*red*) and 400 mM (*blue*) Im solutions. **E** FL quenching of QD-GCN-mCherry complexes in 200 mM (*red*) and 400 mM (*blue*) Im solutions. Black traces show QD-protein assemblies in the absence of Im. **F** Electropherograms of QD-ULD-mCherry and QD-GCN-mCherry complexes in agarose gel. *a* QD alone; *b* QD with ULD-mCherry, 8:1 ULD-mCherry:QD; *c* QD with ULD-mCherry, 16:1 protein:QD; *d* QD with GCN-mCherry, 8:1 protein:QD; *e* QD with GCN-mCherry, 16:1 protein:QD. *Arrow* indicates the position of discrete QD-GCN-mCherry complexes. Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

change in the assembly state. The 27-residue peptide **pep1**, HK VAQLKHEN QALEHE VASLEHK VSAL, belonging to the coiled-coil category, will adopt a random coil conformation in solution. But in the presence of its coiled-coil counterpart **pep2**, K EVQALEEK NAQLKEK VSALKKE VASLE, the two peptides will form a rigid coiled-coil structure, with hydrophilic side chains arranged at the periphery and hydrophobic in the interior. In random coil structure, the histidine residues in **pep1** is scattered along the sequence; in coiled-coil structure, all four histidines are aligned at one side of the coiled-coil bundle. The latter then allows for a more efficient assembly with the surface of QDs (Fig. 3.6).

We designed a multi-histidine protein dubbed "nanobelt" based on **pep1**, drawing inspiration from nanodisc technology, in which a long protein circumvents a bundle of lipid molecules [36, 37]. Containing 5 repeats of **pep1** hinged through proline residues, the 140-residue chain, (PHKIAQLKHENQALEHEIASLEHK ISAL)$_5$ spans a distance of 40–50 nm when fully stretched. When mixed with 5 eq. of the coiled-coil counterpart **pep2**, the protein will turn into 5 coiled-coil helices with an overall length of 15 nm (Fig. 3.6). As the perimeter of a QD is about
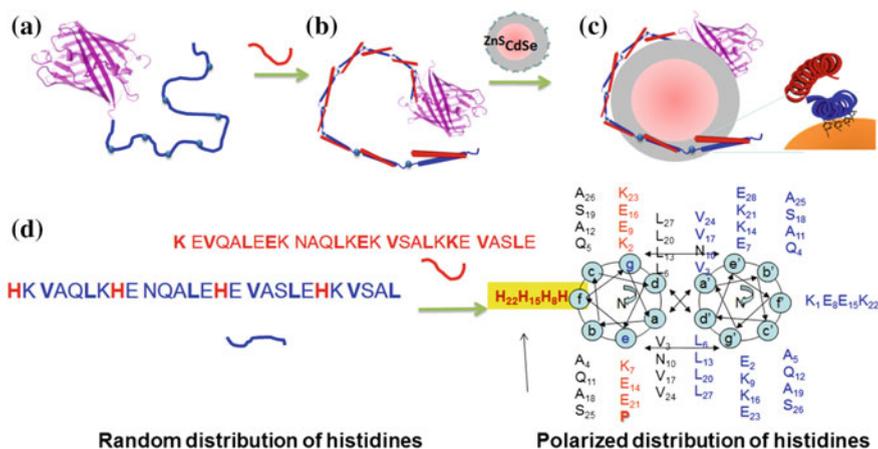
Fig. 3.6 Schematic illustration of the design of nanobelt protein. **a** Nanobelt protein in random coil conformation; **b** addition of **pep2** induces the formation of coiled-coils; **c** coiled-coils assemble with QDs. **d** Helical wheel depiction of the structure of coiled-coil shows the alignment of histidines on one side of the coiled-coil. Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

10 nm, the protein will then and expressed as a fusion protein with a mCherry at the C terminus. Although without a histag, the fusion protein binds to Ni-NTA resin and can be eluted by 70 mM imidazole solution. Despite a discontinuous distribution of histidine residues in random coil state, nanobelt-mCherry can assemble with QDs. Moreover, when nanobelt-mCherry was incubated with 5 eq. **pep2**, the QD-protein assembly showed a marked enhancement in FRET signal at 610 nm, indicating that the protein-peptide complex adopts a more compact structure on the surface of QDs (Fig. 3.7A). QD-protein assemblies were resolved at low protein: QD ratios, 1:1 or 2:1 by agarose gel. In the absence of **pep2**, QD-protein assembly showed a smear and broad band, indicating that a heterogenous mixture of QD-protein species with different stoichiometries was formed. The formation of a heteogenous mixture is ascribable to the long length of the nanobelt protein and the random coil structure, which can embrace multiple QD particles. In contrast, co-assembly of 5 eq. **pep2** showed a discrete band formed in 1:1 and 2:1 (protein: QD) ratio, which likely corresponds to a QD-protein assembly with 1:1 stoichiometry, or a QD species monofunctionalized by nanobelt-mCherry (Fig. 3.7B). As **pep2** does not contain any QD-binding amino acids, such a transition in protein-QD stoichiometry must be due to the structural transition of nanobelt-mCherry. The application of the low-stoichiometry nanobelt-QD complex in molecular imaging is under way.
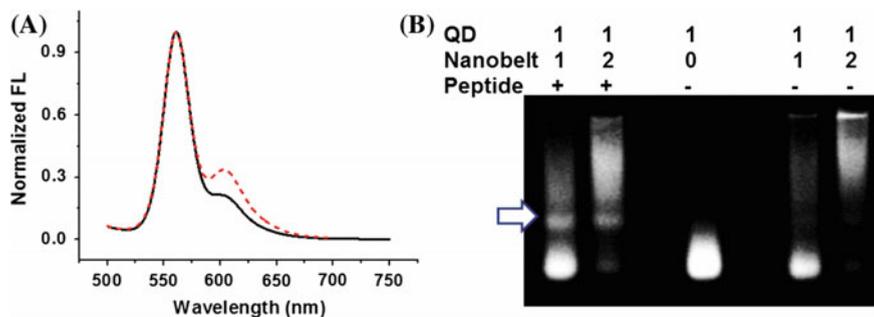
Fig. 3.7 Structural transition from random coil to coiled-coil effect the formation of a low-stoichiometry QD:nanobelt complex. **A** Peptide induced coiled-coils bind more effectively to QDs, *a* without peptide (*black*), *b* with peptide (*red*). **B** Electropherograms of QD-nanobelt complexes in agarose gel in the absence and presence of **pep2**. Arrow indicates the position of a discrete 1:1 complex between QD and nanobelt. Reprinted with permission from Ref. [32]. Copyright (2014) American Chemical Society

3.4   Conclusion

Self-assembly of proteins and QDs has become an important bioconjugation method to functionalize the surface of inorganic nanoparticles. Notwithstanding a great advance in recent years, there still lacks a systematic investigation on how the three-dimensional structures of proteins influences the properties of the assemblies. Here we unveiled unprecedented details of such a bio-nano assembly process from a unique protein engineering aspect. We designed a representative collection of protein ligands with diverse structural features. Some are well studied domains such as TIP1, GCN, and ULD, all with crystal structures available. The other one, nanobelt protein is a de novo design belonging to a new type of polyhistidine protein. Comparison of histag-mCherry with TIP1-mCherry clearly showed that a larger ligand favors lower stoichiometry in the assembly. Comparison of TIP1-mCherry, TIP1-mCherry dimer, and GCN-mCherry in imidazole displacement assay showed that four bundled histags binds with higher affinity to QDs than two histags, and single histag ligands bind QDs with lowest affinity.

The two protein tetramers represent two extremes of spatial distribution of the histags: one with the largest histag–histag distance and the other with smallest possible histag–histag distance. The two tetramers showed notably different behavior in binding to QDs. Clustering of the histags in a very confined space ensured a normal stoichiometry and the formation of discrete QD assemblies. Scattering the histags with a high rigidity results in lower stoichiometry and cross-linking of QD species to form a heterogenous mixture.

The nanobelt protein represents an example by which we applied our knowledge into practice. Further, the nanobelt protein is a new protein designed according to coiled-coil peptides, and has not been structurally characterized. The coiled-coil counterpart peptide triggered the structural transition from a random coil to rigid coiled-coils and subsequently resulted in a defined low-stoichiometry complex with QDs. Taken together, this diverse collection of engineered proteins allowed us to examine a multitude of structural features and their influence on protein-QD assembly. This knowledge will be greatly useful to guide the design of protein ligands for nanoparticles, and find application in molecular imaging.

Appendix 3.1 Plasmid Information of pET21a-TIP1-MC
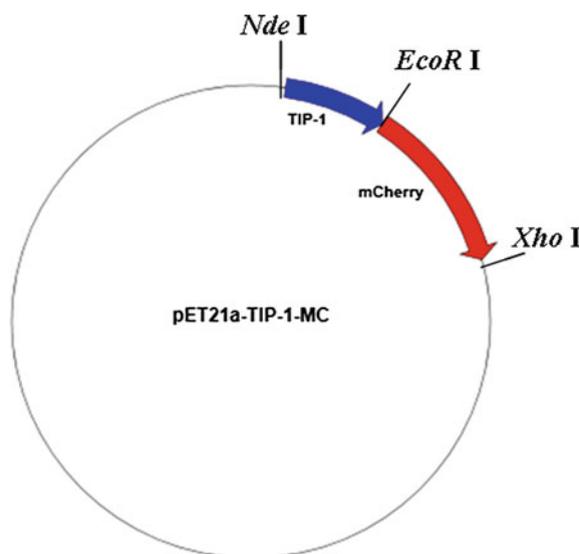
ATGCATCACCATCACCATCACaaagctcttgaagctcagaaacagaaaATGTCCTA
CACCCCGGGCCAGCCTGTCACCGCCGTAGTGCAAAGAGTTGAAATTCATAAGTTGC
GTCAAGGTGAGAACTTAATCTTGGGCTTCAGTATTGGAGGTGGGATCGACCAGGAC
CCGTCTCAGAATCCCTTCTCGGAAGATAAAACAGACAAGGGCATTTACGTCACACG
AGTATCAGAGGGAGGTCCTGCTGAAATTGCTGGGCTGCAGATTGGAGACAAGATCA
TGCAGGTGAATGGCTGGGACATGACCATGGTCACTCACGACCAGGCTCGGAAGCGG
CTCACCAAGCGCTCGGAGGAGGTGGTCCGCCTGCTGGTGACTCGGCAGTCTCTACA
AAAGGCTGTACAGCAGTCCATGCTGTCT*gaattc***ATGGTGAGCAAGGGCGAGGAGG
ATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCC
GTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGG
CACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGG
ACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCC**

**GACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGT
GATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGG
ACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGC
CCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCC
CGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCG
GCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTG
CCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTA
CACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGG
ACGAGCTGTACAAGTAActcgag**

Appendix 3.2 Plasmid Information of pET21a-ULD-MCS
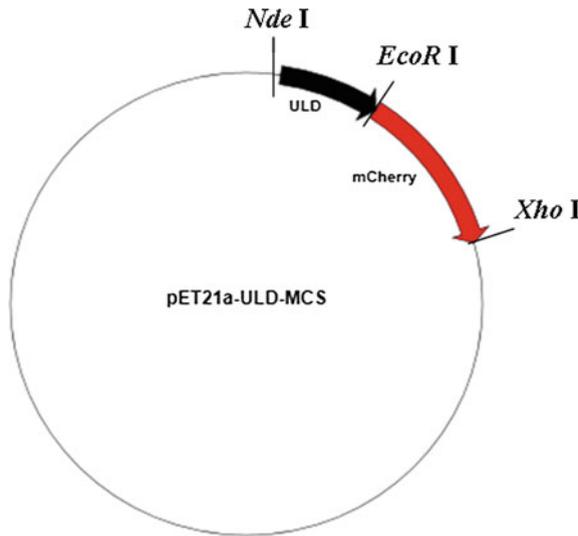
CATATGGGAACCATGTTACCAGTTTTTCTGCGTGGTGGAACATTATGAAAACGCCAT
TGAGTATGATTGCAAGGAGGAGCACGCGGAATTTGTATTGGTGAGAAAGGATATGC
TTTTCAACCAGCTGATAGAGATGGCGTTGCTGTCTCTAGGCTATTCACACAGCTCT
GCTGCCCAAGCCAAAGGGCTCATCCAGGTTGGGAAGTGGAATCCAGTTCCACTGTC
GTATGTGACAGATGCCCCTGATGCCACGGTGGCAGACATGCTTCAAGATGTGTATC
ATGTGGTCACCCTCAAAATTCAGTTACACAGTGAATTCATGGTGAGCAAGGGCGAG
GAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGG
CTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACG
AGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCC
TGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCC
CGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC
GCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTG
CAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGA

CGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGT
ACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGAC
GGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCA
GCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGG
ACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGC
ATGGACGAGCTGTACAAGCTCGAG

Appendix 3.3 Plasmid Information of pET21a-MC-NB

CATATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCG
CTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCG
AGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAG
GGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTC
CAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCC
CCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC
GTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG
CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGG
AGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAG
CAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTA
CAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGG
ACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAG
GGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGGGATCCGGCCCGCATAA
AATTGCGCAACTGAAACATGAAAACCAGGCTCTGGAACACGAAATTGCCTCCTTGG
AACACAAAATTTCTGCACTGCCACACAAGATCGCTCAGCTGAAGCACGAGAACCAA
GCCCTGGAACATGAGATCGCATCTCTGGAGCATAAGATCAGCGCGCTTCCGCACAA
AATCGCCCAGCTGAAACACGAAAACCAGGCACTCGAACATGAAATCGCCAGCCTGG

AACACAAGATTTCCGCCCTGCCACATAAAATTGCACAACTGAAGCATGAAAATCAA
GCTCTGGAGCACGAGATTGCATCCCTGGAACATAAAATCAGCGCACTCCCGCACAA
GATCGCGCAGCTTAAACACGAGAATCAGGCGCTGGAGCACGAAATCGCGAGCCTGG
AGCACAAAATCTCTGCTTTGCTGTAAAAGCTT

Appendix 3.4 Mass Spectrum of the Peptide
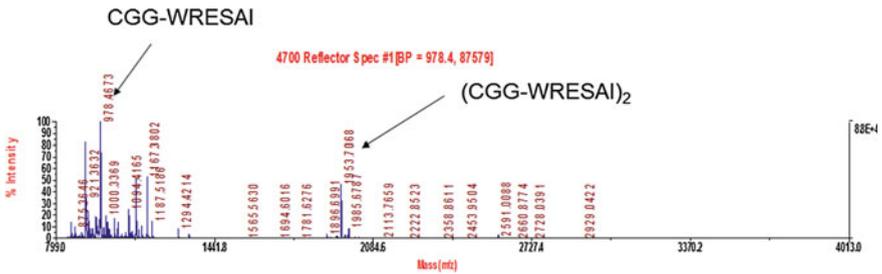
MALDI-TOF spectrum of the dimerization reaction showing the presence of the monomer CGGWRESAI and the dimer (CGGWRESAI)$_2$. CGGWRESAI, [M + H]$^+$, calculated 978.1, found 978.5. (CGGWRESAI)$_2$, [M + H]+, calculated 1953.2, found 1953.7.

Appendix 3.5 SDS-PAGE Results of the Proteins
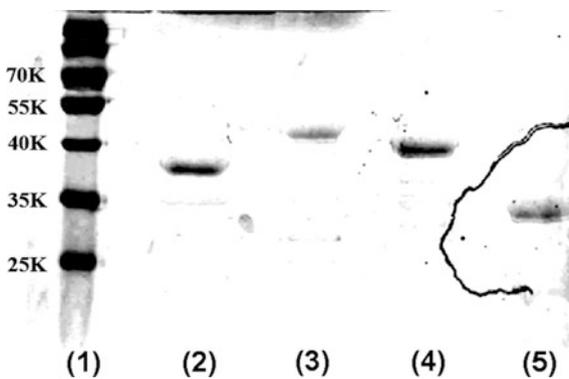
SDS-PAGE of the purified proteins: molecular weight marker (lane 1), GCN-mCherry (lane 2), TIP1-mCherry (lane 3), ULD-mCherry (lane 4), and histag-mCherry (lane 5).

Appendix 3.6 SDS-PAGE Results of Nanobelt-mCherry
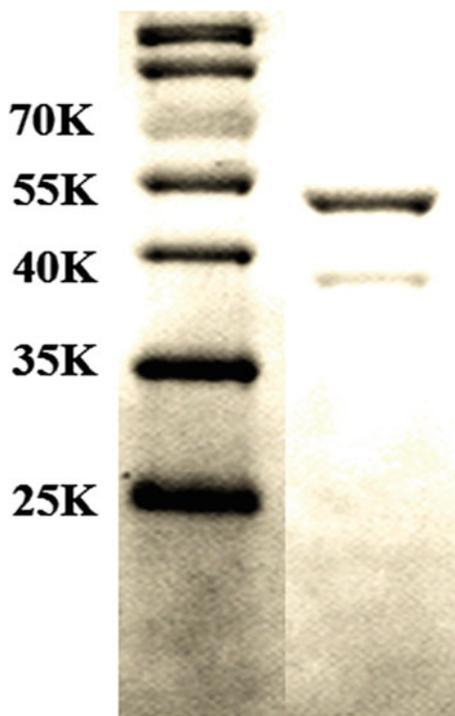
References

1. Dong A, Ye X, Chen J et al (2011) A generalized ligand-exchange strategy enabling sequential surface functionalization of colloidal nanocrystals. J Am Chem Soc 133:998–1006
2. Xiao Y, Patolsky F, Katz E et al (2003) "Plugging into enzymes": nanowiring of redox enzymes by a gold nanoparticle. Science 299:1877–1881
3. You C-C, Agasti SS, De M et al (2006) Modulation of the catalytic behavior of α-chymotrypsin at monolayer-protected nanoparticle surfaces. J Am Chem Soc 128:14612–14618
4. Yu JH, Kwon S-H, Petrášek Z et al (2013) High-resolution three-photon biomedical imaging using doped ZnS nanocrystals. Nat Mater 12:359–366
5. Gu H, Xu K, Xu C et al (2006) Biofunctional magnetic nanoparticles for protein separation and pathogen detection. Chem Commun 9:941–949

6. Sapsford KE, Algar WR, Berti L et al (2013) Functionalizing nanoparticles with biological molecules: developing chemistries that facilitate nanotechnology. Chem Rev 113:1904–2074

7. Reddy LH, Arias JL, Nicolas J et al (2012) Magnetic nanoparticles: design and characterization, toxicity and biocompatibility, pharmaceutical and biomedical applications. Chem Rev 112:5818–5878

8. Michalet X, Pinaud FF, Bentolila LA et al (2005) Quantum dots for live cells, in vivo imaging, and diagnostics. Science 307:538–544

9. Medintz IL, Uyeda HT, Goldman ER et al (2005) Quantum dot bioconjugates for imaging, labeling and sensing. Nat Mater 4:435–446

10. Clapp AR, Medintz IL, Mauro JM et al (2004) Fluorescence resonance energy transfer between quantum dot donors and dye-labeled protein acceptors. J Am Chem Soc 126:301–310

11. Sapsford KE, Pons T, Medintz IL et al (2007) Kinetics of metal-affinity driven self-assembly between protiens or peptides and CdSe-ZnS quantum dots. J Phys Chem 111:11528–11538

12. Hu M, Qian L, Brinas RP et al (2007) Assembly of nanoparticles-protein binding complexes: from monomers to ordered arrays. Angew Chem Int Ed 46:5111–5114

13. Allen M, Willits D, Mosolf J et al (2002) Protein cage constrained synthesis of ferrimagnetic iron oxide nanoparticles. Adv Mater 14:1562–1565

14. Ishii D, Kinbara K, Ishida Y et al (2003) Chaperonin-mediated stabilization and ATP-triggered release of semiconductor nanoparticles. Nature 423:628–632

15. Djalali R, Chem Y, Matsui H (2002) Au nanowire fabrication from sequenced histidine-rich peptide. J Am Chem Soc 124:13660–13661

16. Blum AS, Soto CM, Wilson CD et al (2004) Cowpea mosaic virus as a scaffold for 3-D patterning of gold nanoparticles. Nano Lett 4:867–870

17. Li F, Li K, Cui Z-Q et al (2010) Viral coat proteins as flexible nano-building-blocks for nanoparticle encapsulation. Small 6:2301–2308

18. Monopoli MP, Walczyk D, Campbell A et al (2011) Physical-chemical aspects of protein corona: relevance to in vitro and in vivo biological impacts of nanoparticles. J Am Chem Soc 133:2525–2534

19. Walczyk D, Bombelli FB, Monopoli MP et al (2010) What the cell "sees" in bionanoscience. J Am Chem Soc 132:5761–5768

20. Lundqvist M, Stigler J, Elia G et al (2008) Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. Proc Natl Acad Sci USA 105:14265–14270

21. Cedervall T, Lynch I, Lindman S et al (2007) Understanding the nanoparticle-protein corona using methods to quantify exchange rates and affinities of proteins for nanoparticles. Proc Natl Acad Sci USA 104:2050–2055

22. Klein J (2007) Probing the interactions of proteins and nanoparticles. Proc Natl Acad Sci USA 104:2029–2030

23. Amiri H, Bordonali L, Lascialfari A et al (2013) Protein corona affects the relaxivity and MRI contrast efficiency of magnetic nanoparticles. Nanoscale 5:8656–8665

24. Wang J, Xia J (2011) Preferential binding of a novel polyhistidine peptide dendrimer ligand on quantum dots probed by capillary electrophoresis. Anal Chem 83:6323–6329

25. Wang J, Jiang P, Han Z et al (2012) Fast self-assembly kinetics of quantum dots and a dendrimeric peptide ligand. Langmuir 28:7962–7966

26. Lu Y, Wang J, Wang J et al (2012) Genetically encodable design of ligand "bundling" on the surface of nanoparticles. Langmuir 28:13788–13792

27. Dif A, Boulmedais F, Piont M et al (2009) Small and stable peptidic PEGylated quantum dots to target polyhistidine-tagged proteins with controlled stoichiometry. J Am Chem Soc 131:14738–14746

28. Clarke S, Pinaud F, Beutel O et al (2010) Covalent monofunctionalization of peptide-coated quantum dots for single-molecule assays. Nano Lett 10:2147–2154

29. Clarke S, Tamang S, Reiss P et al (2011) A simple and general route for monofunction-alization of fluorescent and magnetic nanoparticles using peptides. Nanotechnology 22:175103–175113

30. Zhang X, Chu X, Wang L et al (2012) Rational design of a tetrameric protein to enhance interactions between self-assembled fibers gives molecular hydrogels. Angew Chem Int Ed 51:4388–4392

31. Yan X, Zhou H, Zhang J et al (2009) Molecular mechanism of inward rectifier potassium channel 2.3 regulation by tax-interacting protein-1. J Mol Biol 392:967–976

32. Wang J, Nie Y, Lu Y et al (2014) Assembly of multivalent protein ligands and quantum dots: a multifaceted investigation. Langmuir 30:2161–2169

33. Medintz IL, Mattoussi H (2009) Quantum dot-based resonance energy transfer and its growing application in biology. Phys Chem Chem Phys 11:11–45

34. Hochuli E, Bannwarth W, Döbeli H (1988) Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Nat Biotechnol 6:1321–1325

35. Wang Z, Yang X, Chu L et al (2012) The structural basis for the oligomerization of the N-terminal domain of SATB1. Nucleic Acids Res 40:4193–4202

36. Ritchie TK, Grinkova YV, Bayburt TH et al (2009) Reconstitution of membrane proteins in phospholipid bilayer nanodiscs. Methods Enzymol 464:211–231

37. Bayburt TH, Sligar SG (2009) Membrane protein assembly into nanodiscs. FEBS Lett 584:1721–1727

Chapter 4
Coiled-Coil Binding-Induced Covalent Cross-Linking

4.1 Introduction

Coiled-coil is a well characterized secondary structure in proteins. It was first described in 1953 by Crick as the main structural element of a large class of fibrous proteins (included keratin, myosin, and fibrinogen) based on the X-ray diffraction pattern of α-keratin [1, 2]. The first identification of the primary sequence of coiled-coil containing protein tropomyosin in 1972 [3] and the first high-resolution crystal structure characterization of a three stranded coiled-coil hemagglutinin and two stranded (CAP) in 1981 [4] gave scientists great inspiration and courage to further prove that the coiled-coil structure is the key dimerization element in a class of transcriptional factors—the Leucine Zipper proteins [5]. Since then, hundreds of coiled-coil structures were identified and even the high-resolution crystal structures were obtained.

Coiled-coil contains two or more right-handed α-helix, aligned in either parallel or antiparallel orientation to form a left-handed super helical structure. The structural twist slightly changes the periodicity of the α helix from 3.6 amino acids per turn to 3.5 amino acids, so each helix unit contains a repeat of 7 amino acids, called "heptad," with the representative form of (***abcdefg***)$_\mathbf{n}$ [6]. Usually the a and d positions are occupied by hydrophobic amino acids to form the hydrophobic interface, while other positions are occupied by polar or hydrophilic amino acids to form the solvent exposure part. With the increasing background knowledge of coiled-coil and with the assistance of X-ray diffraction and computer-based computation, people are capable of designing the coiled-coil sequence. Figure 4.1 shows the helical wheel structure of a typical two stranded, parallel coiled-coil structure, CCE and CCK. The *a* and *d* positions of CCE/CCK peptides are occupied by Val and Leu, respectively, to provide the Val–Val and Leu–Leu hydrophobic interactions. The *e* and *g* positions are occupied by the complementary charged Glu and Lys to form the interchain electrostatic interactions. The *b* and *c* positions are occupied by Ala and Ser to provide the flexibility and helical-forming potency.
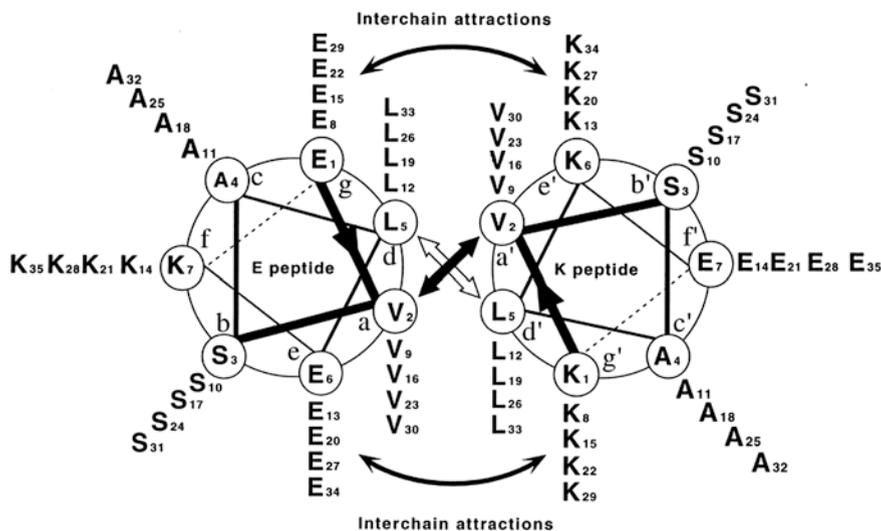
Fig. 4.1 Wheel structural illustration of a two stranded, parallel coiled-coil. Cross arrows represent for the hydrophobic interactions in the core; bent arrows represent for the interchain electrostatic interactions on the latter face; straight arrows represent for the helical direction of the peptide strand. Reprinted from Ref. [7], Copyright (1998), with permission from Elsevier

Lastly, the *f* positions are all occupied by intrachain complementary charged amino acid to balance the electrons in each peptide. It was reported that the affinity of coiled-coils with 5 heptads could reach 3.53×10^{-9} M [7].

In this chapter we will seek to achieve specific recognition from a coiled-coil binding interaction and the conversion of the noncovalent binding interaction into a site-specific covalent cross-linking reaction (Fig. 4.2).

4.2 Experimental Section

4.2.1 Peptide Synthesis

Peptides were manually synthesized based on standard FMOC solid phase peptide synthesis protocol. Briefly, Rink Amide-ChemMatrix® resins (PCAS BioMatrix, Canada) with 0.1 mmol amine group were utilized in each synthesis. A solution containing Fmoc-protected amino acid, HBTU, HOBt, and DIPEA (with a ratio of 1:1:1:2 and 5 fold excess) in 2 mL was added to the resin and stirred for 30 min at RT. After confirming the completion of the coupling reaction by Kaiser Test, the resins were washed with DMF and deprotected in 5 mL 20 % piperidine in DMF (v/v) to remove the Fmoc group. To incorporate fluorescent dyes, 5(6)-FAM (fl), 5(6)-TMR (tmr) or Cy5 free acids were activated by 3-fold EDC/HOBt, and then
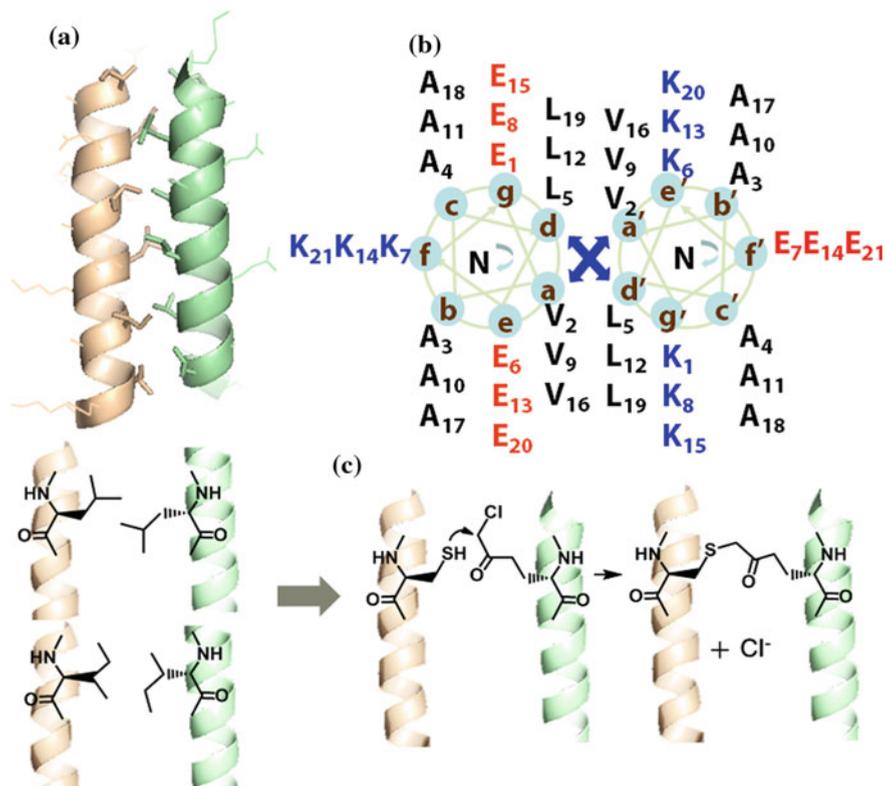
Fig. 4.2 Design of a cross-linking reaction induced by coiled-coil binding. **a** Coiled-coil structure of the CCE/CCK heterodimer (PDB ID 1U0I). **b** Helical wheel representation of the parallel CCE/CCK heterodimer. The coiled-coil is viewed in cross section, with both peptide chains propagating into the page from the NH_2 to the COOH terminus. Blue arrows denote the interhelical hydrophobic interactions at the *a–a'* and *d–d'* positions of the heptads. **c** The principle of a cross-linking reaction induced by coiled-coil binding. Replacing one pair of Leu–Leu or Ile–Ile residues at *a–a* or *d–d'* positions with Cys and an unnatural amino acid X generates coiled-coil derivatives that can still form heterodimers. The local proximity of Cys and X in the heterodimer induces the formation of an interstrand covalent bond. X denotes (2S)-2-amino-3-[(2-chloroacetyl) amino]propanoic acid. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

added to deprotected resins. The mixture was incubated for overnight. In order to construct the unnatural amino acid X, (2S)-2-amino-3-[(2-chloroacetyl)amino] propanoic acid, the resins were pretreated by 10 % acetic acid in DCM (v/v), and then Mtt group was removed in 1 % TFA and 5 % TIS in DCM (v/v). The side chain amine was subsequently coupled with ten fold excess of chloroacetic acid, EDC and HOBt in DMF for 1 h. For the synthesis of CCK-1-dimer, Fmoc-Lys (Fmoc)-OH was first conjugated to a Gly–Gly sequence on the resin. The Fmoc group was removed to expose two branches of amino groups on which a CCK-1

sequence was coupled. The purity of the peptide was confirmed later by reverse phase HPLC (Fig. 4.18) and mass spectrometry (Appendix 4.1).

4.2.2 Peptide Purification and Characterization

To every 100 mg resins, 2 mL final cleavage cocktail containing EDT, TIS, phenol, H_2O, and TFA (with a ratio of 1:2:2:2:33, v/v/w/v/v) was added. The cleavage reaction was allowed to proceed for 2 h at RT under stirring. After the resins were removed through filtration, ice-cold diethyl ester was added to the supernatant dropwise to precipitate the peptides. A final volume of 700 μL 50 % ACN in H_2O (v/v) was used to dissolve the peptide pellet. After being filtered through a 0.2 μm filter, the peptide solution was injected to RP-HPLC (Shimadzu, DGU 20A5, Japan) equipped with a C18 column (Shimadzu, 250L× 4.6, Japan). 0.1 % TFA in H_2O (v/v) and 0.1 % TFA in ACN (v/v) were used as the mobile phase A and B, respectively. For all the analytical HPLC trials, the total flow rate was set to be 1 mL/min and the B concentration raised from 0 to 95 % over 16 min following a linear gradient. For the purification of peptides in a larger scale by semi-prep HPLC columns (Grace, 218TP510, USA), the total flow rate was set to be 3 mL/min and the concentration of B raised linearly from 0 to 45 % over 24 min. The peptide peaks were collected, lyophilized and validated by MALDI-TOF mass spectrometry analysis (Bruker, autoflex TOF/TOF, USA) (Appendix 4.1).

4.2.3 Construction of pET28m-EGFP-CCE-1 Plasmid

Forward and reverse DNA fragments coding CCE-1 peptide (ECAALEKEVAAL EKEVAALEK) were synthesized by Life technology: forward, 5′-CAAATCT GAAGAGTC-TTATGAATGTGCTGCCTTAGAGAAGGAAGTTGCAGCGTTA GAGAAGGAAGTTGCTGCATTAGAGAAGTAGA-3′; reverse, 5′-AGCTTCTA CTTCTCTAATGCAGCAACTTCCTTC-TCTAACGCTGCAACTTCCTTCTCTA AGGCAGCACATTCATAAGACTCTTCAGATTTGAGCT-3′. The two fragments were annealed in annealing buffer (50 mM HEPES, 100 mM NaCl, pH 7.4), and subcloned into Sac I and Hind III sites of the pET28m-EGFP plasmid (a kind gift from Prof. Kowk Fai Lau of CUHK) to yield pET28m-EGFP-CCE-1. The plasmid was confirmed by DNA sequencing (BGI, Shenzhen) (Appendix 4.2).

4.2.4 Expression and Purification of EGFP-CCE-1 Proteins

The plasmids pET28m-EGFP and pET28m-EGFP-CCE-1 were transformed into *E. coli*. Rosetta 2 (DE 3) competent cells. Colonies were grown in LB medium at

37 °C for overnight and a starter culture was grown from a single colony for overnight. 600 mL LB medium was inoculated by the overnight starter culture (1:100 dilution) and allowed to grow at 37 °C until OD_{600} reached 0.4–0.6. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM for induction. After growing at 16 °C for 20 h, the cells were harvested by centrifugation at 6000 rpm for 15 min, and resuspended in 25 mL of lysis buffer (50 mM Tris, 300 mM NaCl, 4 mM β-mercaptoethonol, and 10 mM imidazole, pH 7.5). The cell suspension was sonicated on ice and centrifuged at 20,000 g for 2 h. The supernatant was collected, filtered, and incubated with Ni-NTA resins (GE healthcare, USA) on ice for 40 min to allow Histagged EGFP-CCE-1 protein to bind to the resins. After washes, the bound protein was eluted by elution buffers (50 mM Tris, 300 mM NaCl, 4 mM β-mercaptoethonol, pH 7.5) containing increasing concentrations of imidazole ranging from 50 mM to 500 mM. The protein eluent was exchanged from elution buffer to storage buffer (50 mM Tris, 300 mM NaCl, 15 % glycerol (v/v), pH 7.5) and stored at −20 °C. Protein concentration was measured by Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific Inc., USA) based on an absorption coefficient constant ε_{484} of 56,000 cm^{-1} M^{-1}.

4.2.5 Fitting the Cross-Linking Reaction to Second-Order Kinetics

The CCE/CCK conjugation reactions were fit into a second-order kinetic equation.

$$\text{peptide E} + \text{peptide K} \rightarrow \text{heterodimer E} - K$$

$$\frac{-d[E]}{dt} = \frac{d[E - K]}{dt} = k[E][K]$$

The concentrations of peptides E and K were normalized to 100, $[E]_0 = [K]_0 = 100$, so

$$[E] = 1/(kt + 0.01) \tag{4.1}$$

$$[E - K] = 100 - 1/(kt + 0.01) \tag{4.2}$$

The decrease of reactants E and K were then fit into Eq. (4.1) using Origin 8.1; the increase of product E – K was fit into Eq. (4.2). Half-life $t_{1/2}$ was calculated as $t_{1/2} = 0.01/k$.

4.2.6 Peptide Cross-Linking and Kinetics Measurement

Peptide was dissolved in HEPES buffer (50 mM HEPES, 150 mM NaCl, pH 7.4) and the concentration was measured by UV–Vis spectrometer (Varian, Cary 5G, USA) and calculated using the corresponding absorption coefficient constants (ε_{487nm} = 75,090 cm^{-1} M^{-1} for 5(6)-FAM, and ε_{555nm} = 91,000 cm^{-1} M^{-1} for 5(6)-TMR). 100 μM paired coiled-coil CCK and CCE peptides were dissolved in HEPES buffer containing 1 mM TCEP, incubated in the dark at room temperature, aliquots were taken at different time points and applied for RP-HPLC and monitored at 215 nm (for peptides), 448 nm (for 5(6)-FAM), and 565 nm for (5(6)-TMR), the kinetics of the reaction was analyzed according to the corresponding integration peak areas. For the covalently linked heterodimeric peptide, the mixture solution was applied to mass spectrometer.

4.2.7 In Vitro Protein Labeling and Kinetics Measurement

A final concentration of 5 μM EGFP-CCE-1 protein (or the EGFP control protein) was mixed with 20-fold excess of TMR-CCK-1 peptide (or TMR-CCK3 control peptide) in HEPES buffer (50 mM HEPES, 150 mM NaCl, pH 7.4) in the presence of 1 mM TCEP, and incubated in the dark at RT. Aliquots were taken at different time points and mixed with 5× protein loading buffer (0.313 M Tris-HCl (pH 6.8), 10 % SDS (w/v), 0.05 % bromophenol blue (w/v), 0.5 M DTT and 5 % glycerol (v/v)) and allowed for a further thermal denaturing at 100 °C for 10 min. Samples were then loaded to SDS-PAGE and the protein gel was applied to Typhoon TRIO+ Variable Mode Imager (GE Healthcare, USA) for in-gel fluorescence scanning. Fluorescent imagines were acquired by 488 nm excitation for EGFP and 532 nm excitation for TMR with an emission filter of 520 and 580 nm, respectively. And the fluorescent protein bands were quantified by pixel numbers using the software Image J.

For the on-beads labeling, 20 μL of protein (EGFP-CCE-1 or EGFP) bounded Ni-NTA beads were incubated with 100 μM fluorescent dye appended peptide ligands (TMR-CCK-1 or TMR-CCK3) in the presence of 1 mM TCEP in HEPES buffer in the dark at RT for 100 min. After centrifugation removal of the supernatant solution, the beads were washed extensively with HEPES buffer containing 1 mM TCEP (5 min for each wash). 2 μL of beads were loaded on a glass slide and applied to confocal fluorescent microscope and analyzed under TMR channel (G-2A, excitation 510–560 nm, dichroic mirror 575 nm, barrier filter 590 nm), after each washing trial. The fluorescent intensity was calculated using the confocal software based on the average of 20 beads.

4.3   Results and Discussion

4.3.1   Crosslinking Reaction on the Latter Face

The purpose of this research is to introduce additional chemical reactions between the two coiled-coil strands; such covalent reaction is induced by proximity of the amino acids in the two strands upon noncovalent interaction. Covalently linkage of the two coiled-coil peptides will then have application in covalent protein labeling. One could introduce chemical reactions to coiled-coil at two sets of positions: the electrostatic interaction sites and the hydrophobic core. In the first attempt, we tried to anchor two groups (such as an amine group and an aldehyde group) to $e–g$ positions of the coiled-coil structure. As mentioned before, the sequences for the two peptides that are capable of forming a typical coiled-coil structure are: CCK: H-KVAALKEKVAALKEKVAALKE-NH$_2$ and CCE: H-EVAALEKEVAALE* KEVAALEK-NH$_2$. With the purpose of the least interfering of the sequence and smallest distortion of the coiled-coil structure, the amino acid E* in the CCE sequence was mutated to a terminal Ser and then followed by NaIO$_4$ oxidation to give an α-aldehyde at the side chain. Upon mixing with the CCK peptide which has an amine group at the corresponding Lys side chain, the amino group and the aldehyde group were expected to form a Schiff base and could be further reduced to give an irreversible bond shown in Fig. 4.3 [9]. However, the formation of Schiff base might be too unstable in aqueous solution facing the competition of water, so covalently crosslinking of CCK and CCE derivatives were not observed.

4.3.2   Crosslinking Reaction in the Hydrophobic Core

Alternatively, covalent reaction could be designed at the hydrophobic core. Based on the crystal structure of GCN4 coiled-coil motif (pdb ID 1U01), the mean
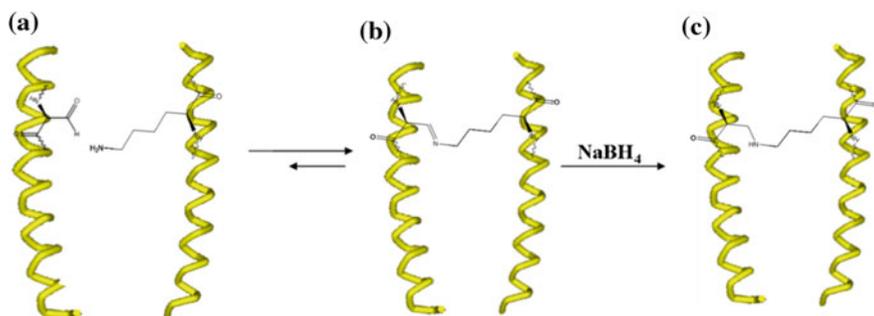
Fig. 4.3 Schematic illustration of crosslinking reaction taken place on the latter face of the coiled-coil. **a** Installation of an aldehyde group and an amine group to the two coiled-coil peptide; **b** Schiff base formation; **c** Irreversible covalent bond formation upon reduction

distances between the α carbons of the corresponding side chains of **a–a (Val–Val)** or **d–d (Leu–Leu)** interacting amino acids are measured and shown in Fig. 4.4, the average distances were estimated to be 5.73 and 3.88 Å, respectively, under this circumstance, there could only accommodate roughly five chemical bonds. Electrophiles with different reactivities toward thiol were screened to found the α-chloroacetyl and β-chloroprocetyl to be potential candidates while α-bromoacetyl was too reactive to lose specificity and the α-chloroprocetyl has too low reactivity. On installation of the halide group to the hydrophobic core, a Lys replace is required to provide a condensation site. Because after linkage, five chemical bonds are structural favored, Lys was changed to its analogue Dap (diaminopropionic acid), as shown in Figs. 4.4 and 4.5.
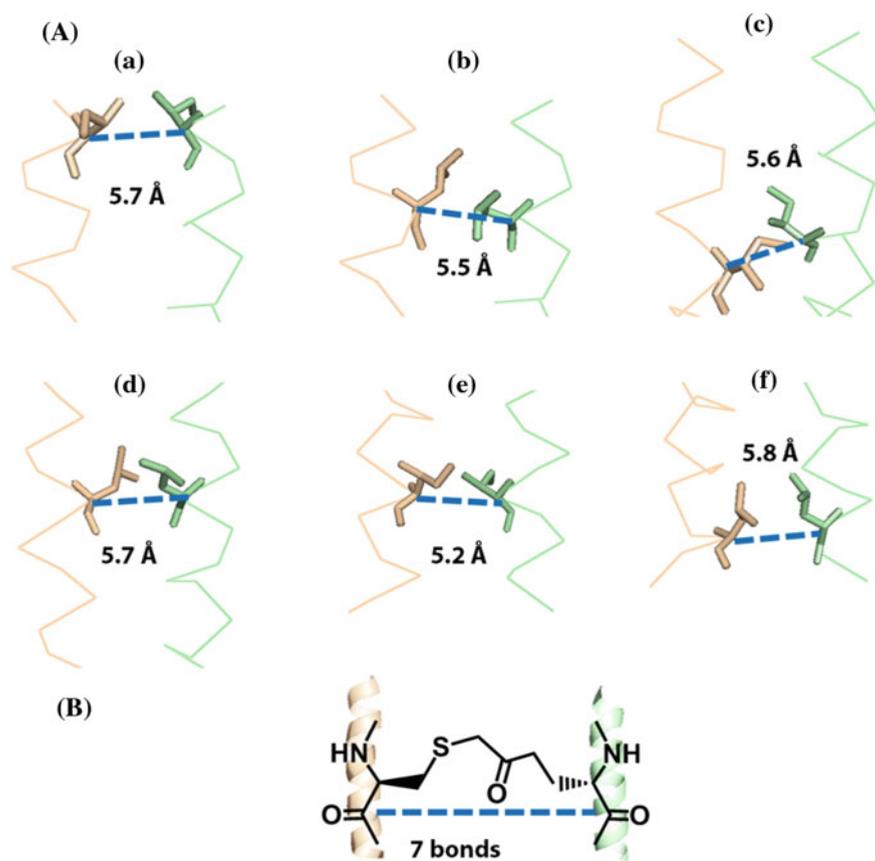

Fig. 4.4 Estimating the structural disturbance of the covalent linkage. **A** Distances between the α-carbons of the residues at a or d positions and those at corresponding a′ and d′ positions of a CCE/CCK heterodimer based on the structure (PDB ID 1U0I). (**A**-*a*). Ile–Ile at the first a–a′ position; (**A**-*b*). Leu–Leu at the first d–d′ position; (**A**-*c*). Ile–Ile at the second a–a′ position; (**A**-*d*). Leu–Leu at the second d–d′ position; (**A**-*e*). Ile–Ile at the third a–a′ position; (**A**-*d*). Leu–Leu at the third d–d′ position. **B** The structure of the thioether bond. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society
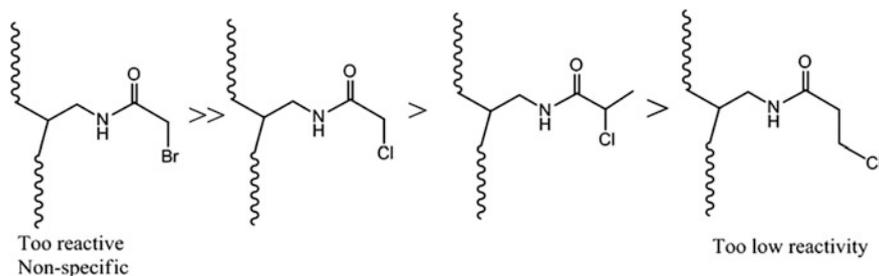
Fig. 4.5 Key amino acids side chain mutations of the hydrophobic core of coiled-coil. Different reactivity toward the thiol group was shown

It was reported that simultaneous replacement of the corresponding core amino acids in CCE and CCK sequences by Cys could give a disulfide bonded hetero-dimer, and the reaction rate of the disulfide formation and the stability of the heterodimer differs according to the mutation sites [10]. In our coiled-coil model with three heptads repeating units, all the **a–a** and **d–d** interaction pairs in the N-terminal and the middle heptads of sequence and were changed to the thiol-α-chloroacetyl pairs. And the coiled-coil sequences were even extended to the N terminus with mutations out-of-the heptads to yield the peptide library shown in Table 4.1, amino acid X represents the α-chloroacetyl containing amino acid derived from the unnatural amino acid Dap. All the CCK group peptides were labeled by 5(6)-carboxyfluorescein (FAM) and all the peptides in the CCE group were labeled by 5(6)-carboxytetramethylrhodamine (TMR) to facilitate their

Table 4.1 List of coiled-coil peptide derivatives

Name	Peptide sequence	Name	Peptide sequence	$t_{1/2}$ (min)
CCK-1	fl-GGGK **X**AALKEK VAALKEK VAALKE	CCE-1	tmr-GGGE **C**AALEKE VAALEKE VAALEK	13.8
		CCE-1′	tmr-KSEESYE**C**AALEKEVAALEKEVAALEK	16.9
CCK-2	fl-GGGK VAA**X**KEK VAALKEK VAALKE	CCE-2	tmr-GGGE VAA**C**EKE VAALEKE VAALEK	>1500
CCK-3	fl-GGK VAALKEK **X**AALKEK VAALKE	CCE-3	tmr-GGE VAALEKE **C**AALEKE VAALEK	920
CCK-4	fl-GGK VAALKEK VAA**X**KEK VAALKE	CCE-4	tmr-GGE VAALEKE VAA**C**EKE VAALEK	720
CCK-5	fl-GGGK VAALKEK **X**AALKEK VAALKEK VAALKE	CCE-5	tmr-GGGE VAALEKE **C**AALEKE VAALEKE VAALEK	>2000
CCK-6	fl-GGGK VAALKEK VAA**X**KEK VAALKEK VAALKE	CCE-6	tmr-GGGE VAALEKE VAA**C**EKE VAALEKE VAALEK	770
CCK-7	fl-GGK VAALKEK VAALKEK **X**AALKEK VAALKE	CCE-7	tmr-GGE VAALEKE VAALEKE **C**AALEKE VAALEK	>4000
CCK-8	fl-GGK VAALKEK VAALKEK VAA**X**KEK VAALKE	CCE-8	tmr-GGE VAALEKE VAALEKE VAA**C**EKE VAALEK	>4000

Note fl denotes 5(6)-carboxyfluorescein; *tmr* denotes 5(6)-tetramethylrhodamine
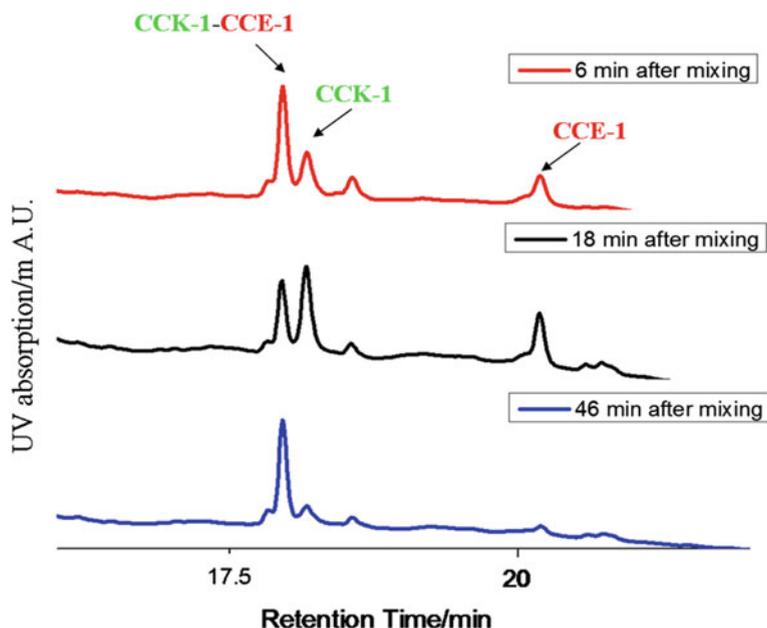
Fig. 4.6 The HPLC trace of the reaction of CCE and CCK pair

identification in HPLC. All the peptides were dissolved in HEPES buffer (50 mM HEPES, 150 mM NaCl, pH 7.5). To monitor the reaction between the peptide pairs, 100 μM peptides were mixed in HEPES buffer at room temperature, aliquots were taken at different time points for HPLC analysis.

The covalent crosslinking reaction of CCE and CCK pair was shown in Fig. 4.6 as one example. The two peptides gave different UV absorption signals due to the corresponding fluorophores. The peaks for CCK-1 and CCE-1 gradually decreased while a new peak emerged as the reaction proceeded. Because this new peak contained signals at both absorption channels, it represents the desired heterodimer. Under the denature condition of HPLC, this peak is thereby assigned as the covalent linked heterodimer. Also, homodimerization of CCE-1 could be observed as proven by incubation of only the CCE-1 peptide alone.

The reactivity of different pairs of coiled-coil peptides were then analyzed using this method. The reaction rates significantly differed as shown in Table 4.1, among which only the CCK-1 and CCE-1 group achieved fast reaction. In the three repeating units group, when the two reactive groups were placed in the middle heptad, almost no reaction took place; this might be caused by the greater distortion of the coiled-coil structure or the lower reactivity of the thiol group in the middle of the peptide. The kinetics of the reaction was also measured. 100 μM peptides were dissolved in HEPES buffer, aliquots at different time points were injected into HPLC and the integrated peak area was calculated. By defining the conversion rate of CCE-1 to the hetrodimer CCK-1-CCE-1 as the reaction rate, and the $t_{1/2}$ as the
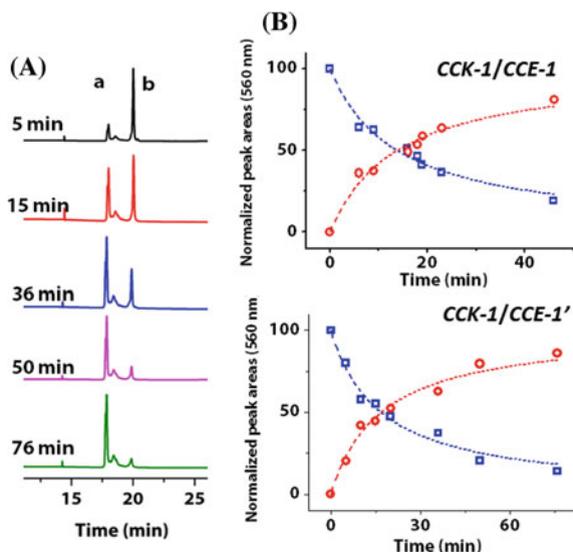
Fig. 4.7 Covalent cross-linking of coiled-coil peptides. **A** Reaction progress of CCE-1 and CCK-1 monitored by HPLC traces at 560 nm. Peak *a*, the heterodimer CCK-1–CCE-1; *b*, CCE-1. The small peak at 18 min is an impurity. **B** Reaction kinetics of the CCK-1/CCE-1 pair and the CCK-1/CCE-1′ pair. The *dashed lines* show curves fit to second-order reaction kinetics. [Peptide] = 50 μM. Red circles represent heterodimer, and *blue* squires represent CCE-1 or CCE-1′. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

equilibrium time for the conversion to achieve a complete degree of 50 %. As shown in Fig. 4.7, peak b represents the heterodimer while peak a represents the CCE-1 peptide. The calculated $t_{1/2}$ of CCK-1 and CCE-1 is 14 min which indicated a fast covalent bond formation. Fluorophores almost had little effect on the reaction rates.

In all the coiled-coil reactivity tests, the homodimerization of the cysteine-containing CCE peptides appeared to be an important side reaction. To reduce this side reaction, we added an extra peptide sequence (KSEESY) to the N terminus of CCE-1 peptide to yield CCE-1′. The negative charges in the extra sequence are expected to form a repulsion force to reduce CCE–CCE interaction. The addition of the extra sequence notably decreased the homodimerization side reaction. CCE-1 converted into the dimer form in noticeable quantity after 3 h while CCE-1′ stayed in the monomer form even after 10 h (Fig. 4.8). At the same time, the formation of the heterodimer between CCK-1 and CCE-1′ was not affected much ($t_{1/2}$ from ∼14 to 17 min).

Besides, we also tested the reaction rate of different CCK/CCE pairs with different mutation sites and they showed very different reaction kinetics, as shown in Figs. 4.9 and 4.10. Such a difference might be explained by several reasons. First, since hydrophobic residues at *a–a′* or *d–d′* positions are the primary drivers of coiled-coils, mutations at deeply buried internal *a–a′* or *d–d′* positions could
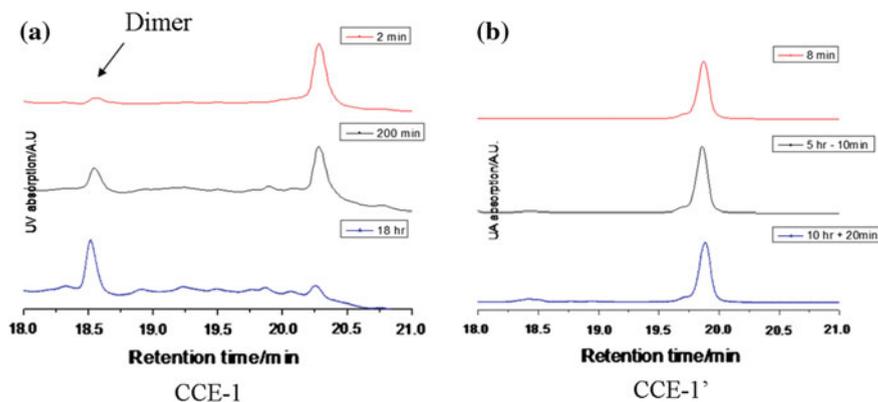
Fig. 4.8 Homodimerization of CCE peptides. CCE-1 peptide alone in the buffer would go homodimerization to give corresponding dimmers while CCE-1′ would stay in the monomeric state. **a** Homodimerization reaction of CCE-1 as monitored by HPLC; **b** homodimerization reaction of CCE-1′. [Peptide] = 50 μM
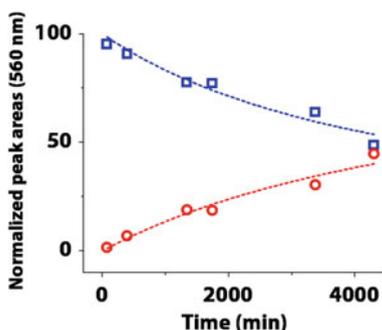
Fig. 4.9 Reaction kinetics of the CCK-7–CCE-7 pair as one example. [Peptide] = 50 μM. Red circles represent heterodimer, and blue squires represent CCE-7. The *dashed lines* show curves fit to second-order reaction kinetics. Half-life $t_{1/2}$ was calculated to be >4000 min. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

significantly disrupt the interaction between CCE and CCK derivatives. Second, the nucleophilic S_N2 reaction is driven by the departure of a Cl^- ion (Fig. 4.2c). Because the termini are more flexible, the chloride ion can be more easily hydrated and leave, driving the reaction forward. This notion is supported by an out-of-heptad crosslinking reaction (Fig. 4.10). The presence of the heterodimer was further confirmed by MALDI-TOF mass spectrum analysis (Fig. 4.11).

The α-chloroacetyl group likely reacts with other thiol containing molecules which are prevalent in biological fluids such as cysteine containing proteins. To examine the specificity between CCE-1 and CCK-1, the *fam*-CCK-1 peptide was incubated with 10-fold excessive of glutathione tri-peptide (GSH) in reduced form
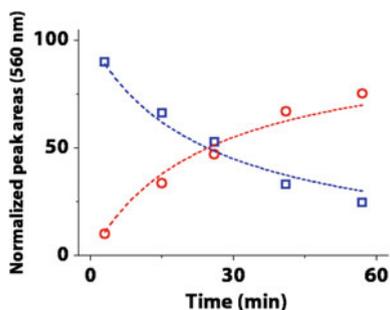
Fig. 4.10 Reaction kinetics of the "out-of-heptad" pair shows a $t_{1/2}$ of 25 min. [Peptide] = 50 μM. The sequences of the two peptides were *tmr*-KSEESY-C EKE VAALEKE NAALEKE VAALEK and *fl*-X KEK VAALKEK NAALKEK VAALKE (*blue*). The heterodimer was shown in *red*. The *dashed lines* show curves fit to second-order reaction kinetics. Half-life $t_{1/2}$ was calculated to be 25 min. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society
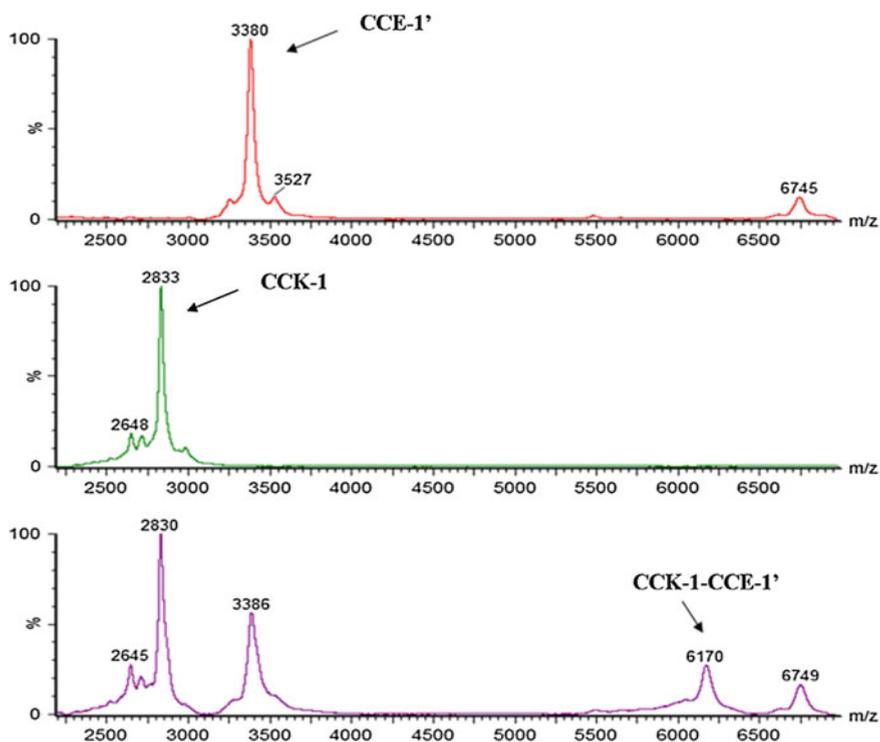
Fig. 4.11 MALDI-TOF mass spectra analysis of the CCE-1′ and CCK-1 reaction system. Top, *tmr*-CCE-1′ peptide only; middle, -*fam*-CCK-1 peptide only; bottom, the mixture of CCK-1 and CCE-1′
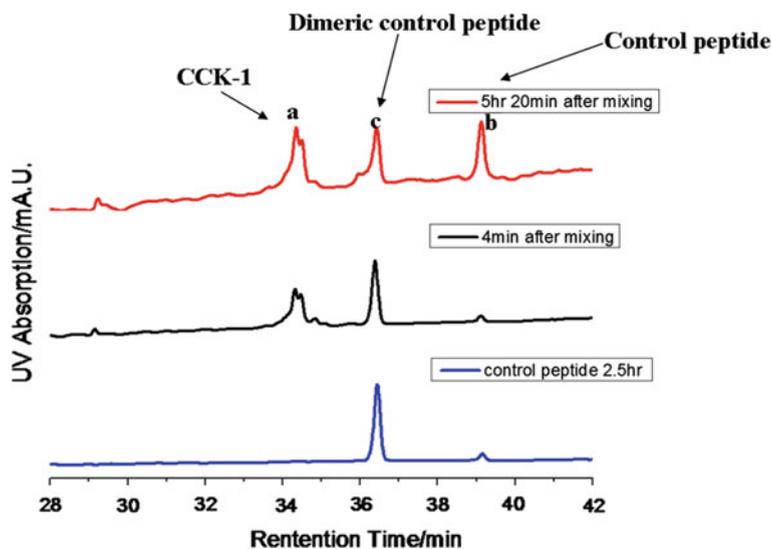
Fig. 4.12 *fam*-CCK-1 peptide (peak *a*) was incubated with a control peptide (peak *b*) containing a random Cys residue to give only the homodimer (peak *c*) of the control peptide (*upper line*) after 5.5 h′ incubation. The retention time of peak *c* was confirmed by incubating the control peptide along

at 37 °C. However, no reaction between CCK-1 and GSH was observed. Further, a thiol containing peptide TMR-LVPRGSGGC did not reaction with *fam*-CCK-1 either (Fig. 4.12), indicating that the reactivity of the α-chloroacetyl group toward the designed thiol group was specific and driven by proximity, and α-chloroacetyl group does not react nonspecifically with other thiol containing molecules.

4.3.3 Multi-component Labelings

The beauty of the coiled-coil interaction lies not only in its simplicity, but also in its extendability to heterospecific modules of multiple assembling components in synthetic biology [11]. Basing on an algorithm to search the specificity of coiled-coils, Bromley et al. designed coiled-coil tectons: three pairs of coiled-coil peptides that have maximal specificity toward their binding partners [12]. We envision that covalent cross-linking reactions based on tecton pairs should exhibit orthogonality (orthogonality here means that in a solution containing A and A′, and B and B′, only the covalent conjugates A–A′ and B-B′ could form, but not the heterodimers A–B′ or B–A′).

Having the tecton peptides p_1–p_6 as parents, we synthesized six derivatives p_{1X}, p_{2C}, p_{3X}, p_{4C}, p_{5C}, and p_{6X} (Table 4.2). In all three pairs, the Cys/**X** mutation was introduced at the N-terminal *a–a′* position. To differentiate the three sets of

Table 4.2 List of tecton peptide derivatives-Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

p_{1X}	*tmr*-GGE **X**AALKQE NQALEQK IAALKGY
p_{2C}	*tmr*-GGE **C**AALKQK NKYLKQE IQQLE
p_{3X}	*fl*-GGK **X**QALQQK IKQLKQK IAALKGY
p_{4C}	fl-GGQ CAALEQE IAALEQE IAALE
p_{5C}	Cy5-GGE CAALEQQ NKYLKQE IAALKGK
p_{6X}	Cy5-GGK XKALKQE NAYLQQE IQALK

cross-linking reactions, the six peptides were labeled with three different dyes, *fl*, *tmr*, and *cy5*, respectively, and monitored at the wavelengths of *fl* (448 nm), *tmr* (520 nm) and *Cy5* (650 nm) in a single HPLC run. By comparing the retention times of the cross-linked heterodimers with single-pair reaction systems as standards, we could assign each peak to monomers or heterodimers (Fig. 4.13). The cross-linking reactions were found to proceed faithfully within the specific pairs. Namely, a p_{1X}–p_{2C} heterodimer (peak **a**–**b**), a p_{3X}–p_{4C} heterodimer (peak **c**–**d**), and a p_{5C}–p_{6X} heterodimer (peak **e**–**f**), but not nonspecific heterodimers, could be observed (Fig. 4.14A).

To show more clearly the orthogonality of the reaction, we examined a four-peptide system containing the p_{1X}/p_{2C} pair and the p_{5C}/p_{6X} pair; fl-labeled peptides were excluded to obtain clearer spectra. Clearly, only the specific heterodimers p_{1X}–p_{2C} (peak **a**–**b**, with a retention time of 33.64 min) and p_{5C}–p_{6X} (peak e–f, with a retention time of 32.33 min) were detected (Fig. 4.14B). The UV–Vis spectra of the two heterodimer peaks were also indicative. Peak **a**–**b** showed the characteristic spectrum of *tmr* (with a slight "leaking" due to the difference in absorption coefficient), while peak **e**–**f** showed the characteristic spectrum of Cy5. Taken together, within our detection limit, the cross-linking reactions all happened within specific pairs and did not occur randomly. The orthogonality demonstrated here is consistent with the previous report that
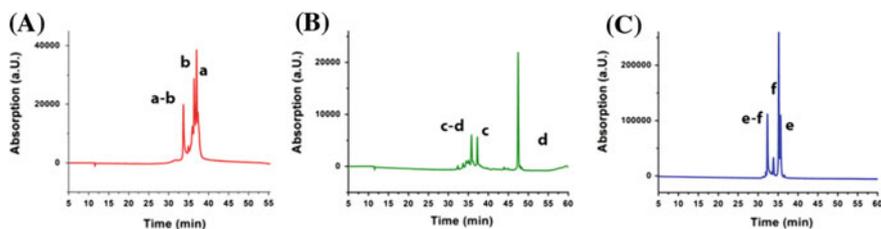
Fig. 4.13 Covalent cross-linking of coiled-coil peptide pairs. **A** Incubation of two peptides, p1X (peak *a*) and p2C (peak *b*) resulted in the heterodimer p1X–p2C (peak *a–b*). **B** Incubation of p3X (peak *c*) and p4C (peak *d*) resulted in the heterodimer p3X–p4C (peak *c–d*). **C** Incubation of p5C (peak *e*) and p6X (peak *f*) resulted in the heterodimer p5C–p6X (peak *e–f*). Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society
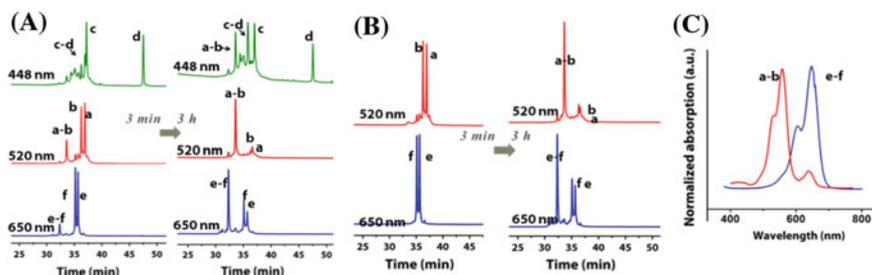
Fig. 4.14 Covalent cross-linking of coiled-coil peptides. **A** Progress of a reaction with equimolar amounts of six peptides, p1X (peak *a*), p2C (peak *b*), p3X (peak *c*), p4C (peak *d*), p5C (peak *e*), and p6X (peak *f*), after 3 h of incubation at RT and pH 7.4. The reaction was monitored by HPLC at 448 nm (*green trace*), 520 nm (*red trace*), and 650 nm (*blue trace*). Peaks *a–b*, *c–d*, and *e–f* were assigned as the heterodimers p1X–p2C, p3X–p4C, and p5C–p6X, respectively. The unassigned peaks at 448 nm were caused by "spectral leaking" of tmr and Cy5-labeled peptides. **B** Progress of a reaction with equimolar amounts of four peptides, p1X (peak *a*), p2C (peak *b*), p5C (peak *e*), and p6X (peak *f*). **C** UV–Vis spectra of peak *a–b* (retention time 33.64 min) and peak *e–f* (retention time 32.33 min) in **b** Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

cysteine-labeled tecton peptides p_1–p_6 could specifically cross-link with each partner [11]. Aside from marked orthogonality, the reaction rates of the three pairs in a mixture significantly differed. The p_{1X}/p_{2C} pair proceeded the most rapidly, while the p_{3X}/p_{4C} pair occurred the most slowly. The cause of this discrepancy is unknown to us at present.

4.3.4 Covalent Labeling of a Target Protein

We then try to label a protein using the CCE-1 and CCK-1 pair based on "proximity induced reactivity" effect. CCE-1 or CCE-1′ which contain only natural amino acids, were chosen as genetically encodable tags of a model protein, enhanced green fluorescent protein (EGFP). The labeled protein can be resolved on denaturing SDS-PAGE. We then utilized in-gel fluorescence scanning to examine the covalent labeling of EGFP protein by tmr labeled peptide.

EGFP-CCE-1 was cloned, expressed and purified. The protein and 20 fold excess *tmr*-CCK-1 were mixed in HEPES buffer (50 mM TCEP, 150 mM NaCl, pH 7.4) containing 1 mM TCEP at room temperature. The mixture was loaded to SDS-PAGE, separated by electrophoresis, and the protein gel was imaged under a Typhoon Trio fluorescent imager. Fluorescent imagines were acquired at 532 nm excitation wavelength for *tmr* with an emission filter of 580 nm (Fig. 4.15). A covalently conjugated complex between EGFP-CCE-1 and *tmr*-CCK-1 could be seen in the in-gel fluorescence scanning. EGFP without CCE-1 tag did not react with *tmr*-CCK-1, although it contains solvent accessible cysteines.
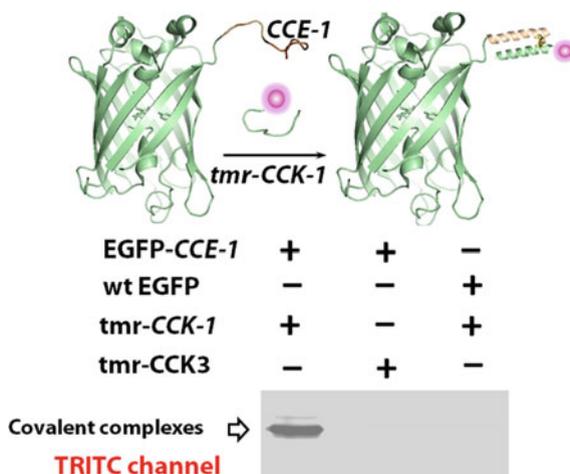
Fig. 4.15 Covalent labeling of *EGFP-CCE-1* in solution. The protein/peptide solution was thermally denatured and resolved by reducing SDS-PAGE. The gel was imaged with a Typhoon Imager at the TRITC channel. *CCK3*, the parental peptide without the α-chloroacetyl moiety, was included as a control. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

We also conducted the labeling reaction in a lysate of mouse brain containing a total protein up to 7.5 mg/mL. Only one band corresponding to the molecular weight of EGFP-CCE-1 was detected by the Typhoon imager (Fig. 4.16), further showing that covalent labeling of the CCE-1 tagged protein in vitro is site-specific.

Compared with the reaction of synthetic coiled-coil peptides, the covalent labeling of the EGFP-CCE-1 fusion occurred more slowly, possibly because the
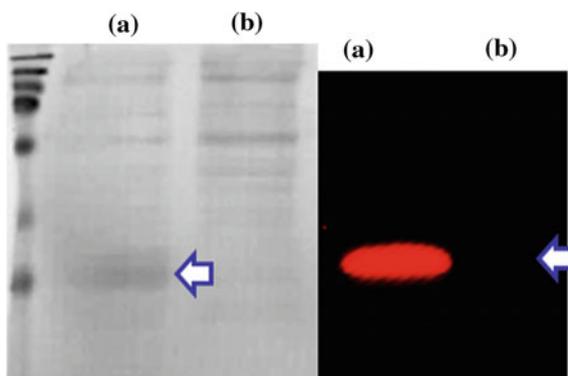
Fig. 4.16 Covalent labeling of EGFP-CCE-1 in mouse brain lysate (total protein concentration were 7.5 mg/ml). EGFP-CCE-1 protein was added to brain tissue homogenate in lane (**a**). Lane (**b**) without doped protein serves as a control. The gel was imaged with a Typhoon Imager at the TRITC channel (*right*) and then stained with Coomassie *blue dye* (*left*). Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

CCE-1 tag in the fusion was sterically hindered or had decreased rotational free-dom. The same effect was observed in the labeling reaction with the CA6D4 tag [13]. To improve the labeling efficiency, we followed the strategy of Hamachi et al. that used a bivalent interaction to more efficiently label cell surface receptors [14]. A bivalent CCK-1-dimer was synthesized using a lysine residue as the branching point (Figs. 4.17A and 4.18). CCK-1-dimer reacted with EGFP-CCE-1 roughly 5 times faster than the monovalent CCK-1 probe (38 min vs. 150 min by $t_{1/2}$) (Figs. 4.17B, C and 4.19). Interestingly, after EGFP-CCE-1 was covalently linked to CCK-1 or the CCK-1-dimer, its mobility in denaturing SDS-PAGE increased (Fig. 4.17B). As shown in the gel image stained by Coomassie blue, both the covalent conjugates (EGFP-CCE-1)-(tmr-CCK-1) and (EGFP-CCE-1)-(tmr-CCK-1-dimer) migrated faster than EGFP-CCE-1, despite their higher molecular weights. It could likely be explained by the postulation that covalently cross-linked CCE-1–CCK-1 heterodimer has a more compact structure than uncomplexed CCE-1 under denaturing conditions.

Still one question remained: what percentage of the complex formed between EGFP-CCE-1 and CCK-1 is linked through covalent linkage under native condi-tion? To better estimate this number, we conducted the labeling reaction on solid beads to exert a washing protocol and to mimic the labeling of cell surface receptors. Surface-immobilized EGFP-CCE-1 was labeled by TMR-CCK-1 or TMR-CCK3 on Ni-NTA resins. The resins were washed by PBS, and the
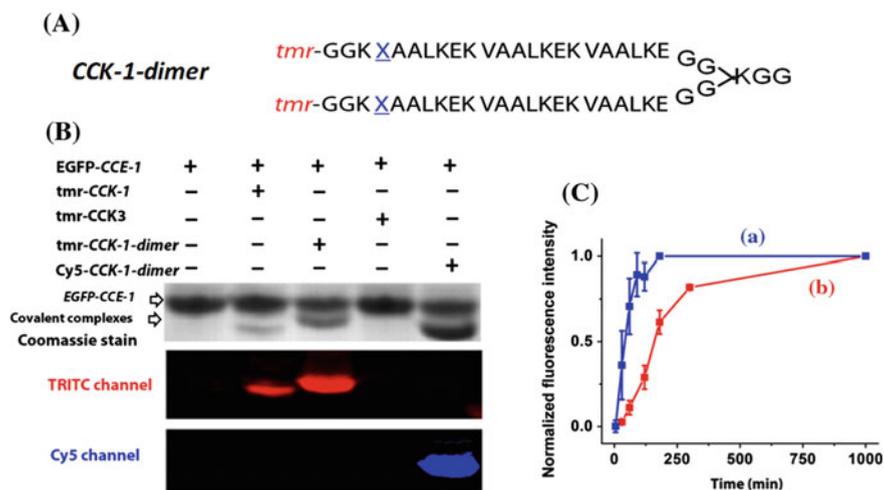
Fig. 4.17 The CCK-1-dimer probe labeled EGFP-CCE-1 more efficiently. **A** Structure of the CCK-1-dimer. **B** Covalent labeling of EGFP-CCE-1 by monomeric and dimeric peptides. The gel was imaged by a Typhoon Imager at the TRITC or Cy5 fluorescent channels, after which the gel was stained by *Coomassie blue*. [Protein] = 40 μM, [peptide] = 20 μM, 4 °C overnight. **C** The reaction kinetics of covalent labeling of EGFP-CCE-1 by the CCK-1-dimer (*a*) or CCK-1 (*b*). [protein] = [peptide] = 20 μM at RT. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society
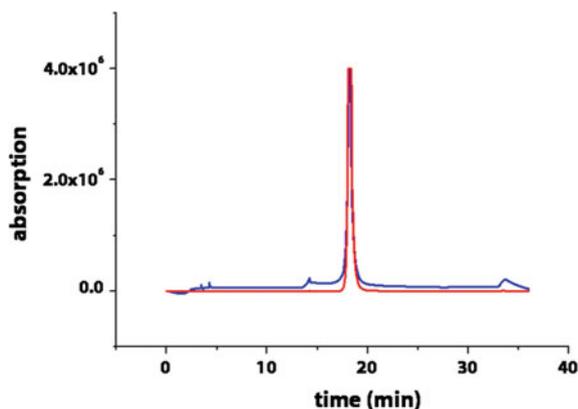
Fig. 4.18 RP-HPLC trace of the purified tmr-CCK-1-dimer peptide at 215 nm (*blue*) and 560 nm (*red*). Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society
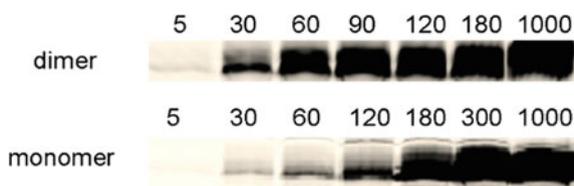
Fig. 4.19 Covalent cross-linking reactions between CCK-1-dimer probe (dimer) or CCK-1 probe (monomer) and EGFP-CCE-1 protein. Peptide and protein were mixed at RT ([protein] = [peptide] = 20 μM) to allow for covalent cross-linking reaction. At different time points (shown above in min), aliquots were taken and the reactions were quenched by heating at 95 °C for 10 min. The solutions were loaded in reducing SDS-PAGE. The gel was imaged by a Typhoon imager at TRITC channel. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

fluorescence of the resins was quantified immediately with the TRITC channel of a fluorescent microscope. We then washed the labeled particles again and took fluorescent images (Fig. 4.20a). This procedure was reiterated. After three repetitive washing steps, the fluorescent signal of noncovalent labeling by TMR-CCK3 decreased to the basal level. About 65 % of the fluorescent signal still remained after five washings of resins labeled by TMR-CCK-1 (Fig. 4.20b). This result suggests that before extensive washings were exerted under native condition, possibly two-thirds of the EGFP-CCE-1-CCK-1 complex were linked through covalent bond, whereas the remaining one-third of labeling was through noncovalent coiled-coil interaction. In addition, three rounds of rigorous washing were found to be sufficient to remove the noncovalently associated CCK-1 probe under this setting.
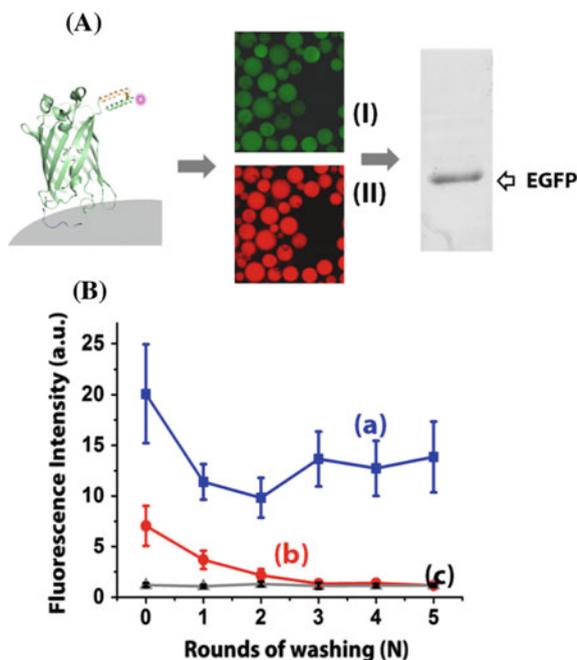
Fig. 4.20 Covalent labeling of immobilized EGFP-CCE-1. **A** EGFP-CCE-1 was immobilized on Ni-NTA resins. TMR-CCK-1 peptide was then added to label the protein-covered resins at RT. The resins were then washed and imaged by a fluorescent microscope at the FITC (I) and TRITC (II) channels. Pixels in the image from the TRITC channel were quantified as a measure of the fluorescent signal on the resins. The protein was lastly eluted by a solution of 500 mM imidazole and analyzed by denaturing SDS-PAGE. The gel was scanned by a Typhoon Imager at the TRITC channel. **B** Fluorescence intensity of the protein-loaded beads after each round of washing. *a* EGFP-CCE-1, labeled with TMR-CCK-1; *b* EGFP-CCE-1, labeled with TMR-CCK3; *c* wt EGFP, labeled with TMR-CCK-1. Reprinted with permission from Ref. [8]. Copyright (2014) American Chemical Society

4.4   Conclusion

In this chapter, we systematically envisioned the coiled-coil interaction, and successfully incorporated two reactive groups, the α-chloroacetyl and the thiol, into the side chains of two closely localized key amino acids of the two stranded coiled-coil partner (to yield one natural amino acid and one unnatural amino acid). Due to the "proximity induced reactivity" effect, the two groups, originally have low reactivity toward each other in aqueous buffer, would be accelerated significantly to give a covalent and nondissociable cross-linkage. Through a round of positional screening, the best reactivity, as well as the good selectivity could be achieved to give a pair of chemical tag and probe system (CCK-1 and CCE-1′).

Furthermore, CCE-1 was genetically fused to the C terminus of EGFP as a recognition site for CCK-1. The covalent labeling reaction was further accelerated

by the bivalent branched CCK-1-dimer. The labeling reaction was specific and could not be interfered by other thiol containing molecules. Notably CCE-1 tag contains only 25 amino acids and it could be fused to either N or C terminus of the target protein. This covalent labeling strategy achieves: (1) simple treatment (only incubation of the two peptides); (2) require no enzymes or ions; (3) fast ($t_{1/2}$ is about 15 min). It will find broad application in covalent and nonenzymatic protein labeling.

Appendix 4.1 List of the Synthetic Peptides and Their Molecular Weights

Name	Peptide sequence	Calculated	Found
CCK-1	fl-GGGK **X**AALKEK VAALKEK VAALKE	2828.3	2827.8
CCK-2	fl-GGGK VAA**X**KEK VAALKEK VAALKE	2814.3	2813.4
CCK-3	fl-GGK VAALKEK **X**AALKEK VAALKE	2771.3	2770.8
CCK-4	fl-GGK VAALKEK VAA**X**KEK VAALKE	2757.3	2756.3
CCK-5	*fl*-GGGK VAALKEK **X**AALKEK VAALKEK VAALKE	3567.8	3567.1
CCK-6	*fl*-GGGK VAALKEK VAA**X**KEK VAALKEK VAALKE	3553.7	3553.0
CCK-7	*fl*-GGK VAALKEK VAALKEK **X**AALKEK VAALKE	3510.7	3509.9
CCK-8	*fl*-GGK VAALKEK VAALKEK VAA**X**KEK VAALKE	3496.7	3495.9
CCE-1	*tmr*-GGGE **C**AALEKE VAALEKE VAALEK	2825.8	2824.7
CCE-1′	*tmr*-KSEESYE**C**AALEKEVAALEKEVAALEK	3379.1	3377.8
CCE-2	*tmr*-GGGE VAA**C**EKE VAALEKE VAALEK	2811.9	2810.7
CCE-3	*tmr*-GGE VAALEKE **C**AALEKE VAALEK	2768.8	2767.2
CCE-4	*tmr*-GGE VAALEKE VAA**C**EKE VAALEK	2755.8	2754.3
CCE-5	*tmr*-GGGE VAALEKE **C**AALEKE VAALEKE VAALEK	3566.3	3566.0
CCE-6	*tmr*-GGGE VAALEKE VAA**C**EKE VAALEKE VAALEK	3552.3	3551.9
CCE-7	*tmr*-GGE VAALEKE VAALEKE **C**AALEKE VAALEK	3509.3	3509.1
CCE-8	*tmr*-GGE VAALEKE VAALEKE VAA**C**EKE VAALEK	3495.3	3494.9
	tmr-GG **X**KEKVAALKEKNAALKEKVAALKE	3239.8	3237.9
	tmr-KSEESY**C**EKEVAALEKENAALEKEVAALEK	3750.3	3750.0

<div align="right">(continued)</div>

(continued)

Name	Peptide sequence	Calculated	Found
P1x	*tmr*-GG-EXAALKQENQALEQKIAALKGYK	3131.6	3130.5
P2c	*tmr*-GG-ECAALKQKNKYLKQEIQQLE	2930.0	2929.5
P3x	*fl*-GGKXQALQQKIKQLKQKIAALKGY	3059.4	3058.7
P4c	*fl*-GGQCAALEQEIAALEQEIAALE	2613.4	2635.2 (+Na$^+$)
P5c	*Cy5*-GG-ECAALEQQNKYLKQEIAALKGK	3226.7	3226.7
P6x	*Cy5*-GG-KXKALKQENAYLQQEIQALK	3157.2	3156.6
	tmr-GGG KXAALKEKVAALKEKVAALKE	2882.6	2881.8
CCK-1′-dimer	(*tmr*-GGKXAALKEKIAALKEKIAALKEGG)$_2$KGG	6160.4	6192.7 (+Na$^+$)

Appendix 4.2 Plasmid Information of pET28m-EGFP-CCE-1
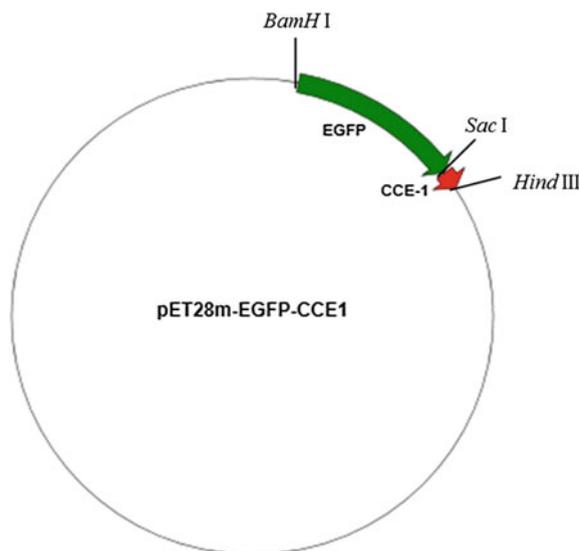
ATGCATCACCATCATCATCATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCG
C**GGATCC**ATGCATCACCATCATCATCATATGGCTAGCATGACTGGTGGACAGCAAA
TGGGTCGCGGATCCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCC
ATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA
GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCAT

GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAA
CTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGG
TGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCAC
TACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTA
CCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTAC
AAGGAGCTCAAATCTGAAGAGTCTTATGAATGTGCTGCCTTAGAGAAGGAAGTTGC
AGCGTTAGAGAAGGAAGTTGCTGCATTAGAGAAGTAGAAGCTT

References

1. Lupas A (1996) Coiled coil: new structures and new functions. TIBS 21:375–382
2. Crick FH (1953) The packing of α-helices: simple coiled-coils. Acta Cryst 6:689–697
3. Hodges RS, Sodek J, Smillie LB et al (1972) Tropomyosin: amino acid sequence and coiled-coil structure. Cold Spring Harbor Symp. Quant. Biol. 37:299–310
4. Wilson IA, Skehel JJ, Wiley DC et al (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature 289:366–373
5. Landschulz WH, Johnson PF, Mcknight SL et al (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764
6. Lupas A, Dyke MV, Stock J et al (1991) Predicting coiled coils from protein sequences. Science 252:1162–1164
7. Chao H, Bautista DL, Litowski J et al (1998) Use of a heterodimeric coiled-coil system for biosensor application and affinity purification. J Chromatorgr B 715:307–329
8. Wang J, Yu Y, Xia J et al (2014) Short peptide tag for covalent protein labeling based on coiled coils. Bioconju Chem 25:178–187
9. Siegel M, Xia J, Khosla C et al (2007) Structure-based design of α-amido aldehyde containing gluten peptide analogues as modulators of HLA-DQ2 and transglutaminase 2. Bioorg Med Chem 16:6253–6261
10. Zhou NE, Kay CM, Hodges RS et al (1993) Disulfide bond contribution to protein stability: positional effects of substitution in the hydrophobic core of the two-stranded α-helical coiled-coil. Biochemistry 32:3178–3187
11. Bromley EH, Channon K, Moutevelis E et al (2008) Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems. ACS Chem Biol 3:38–50
12. Bromley EH, Sessions RB, Thomson AR et al (2008) Design α-helical tectons for constructing multicomponent synthetic biological systems. J Am Chem Soc 131:928–930
13. Nonaka H, Tsukiji S, Ojida A et al (2007) Non-enzymatic covalent protein labeling using a reactive tag. J Am Chem Soc 129:15777–15779
14. Nonaka H, Fujishima S-H, Uchinomiya S-H et al (2010) Selective covalent labeling of tag-fused GPCR proteins on live cell surface with a synthetic probe for their functional analysis. J Am Chem Soc 132:9301–9309

Chapter 5
Cell Surface Receptor Labeling

5.1 Introduction

Genetic fusion of a reporter such as a fluorescent protein or an enzyme stamps a permanent mark on a protein of interest (POI) that imaging techniques can trace and visualize. Although widely utilized, the cumbersome size of the tags (for example, the green fluorescent protein is a 238-amino-acid polypeptide with a molecular weight of 26.9 kDa) can interfere with the intracellular transport or activity of the target protein [1, 2]. In contrast, covalent labeling based on chemical reactions at amino acids with specific chemical reactivities exhibits a variety of virtues, such as smaller tag sizes, a versatile choice of labels, and temporospatial control of the labeling event [3].

The most common covalent labeling reactions harness the chemical reactivities of natural amino acids [4]. Many chemical reactions that modify a single amino acid, such as cysteine [4, 5], lysine [4], tyrosine [6, 7], or tryptophan [8, 9], or several consecutive amino acids in a motif [10, 11], or one of the two termini of a POI [12–17], have been developed. Due to the multitude of chemical moieties with similar reactivities in a cellular environment, labeling reactions solely driven by chemical reactivity often face the challenge of selectivity. One tactic to overcome this hurdle is to introduce bioorthogonal chemical moieties by genetically incorporating unnatural amino acids [17–20]. Or chemical biologists seek help from enzymes, which have unparalleled specificity in recognizing and converting substrates. Covalent labels have been developed based on suicide inhibitors, cofactors, and substrates of various enzymes; examples include activity-based protein probes [21, 22], CoA-affinity-based kinase tags [23], the enzyme-suicide substrate-based SNAP/CLIP-tags [24, 25], the Halo-tag [26], the lactam-based-lactamase-tag [27], and the small molecule inhibitor-based TMP tag [28, 29]. But despite their outstanding specificity, the enzyme tags are still larger than 100 amino acids.

A third strategy uses short peptide sequences from natural substrates of ligases or transferases and converts them into covalent tags. Examples in this category include

a 15-amino-acid receptor peptide (AP, the substrate of the biotin ligase BirA) [30], a 13-amino-acid sequence LAP (the substrate of a mutant lipoic acid ligase) [31], a 7-residue Q-tag (the substrate of a transglutaminase) [32], the small sortagging motif LPXTG (the substrate of a sortase) [33], a LCTPSR formyl glycine tag (the substrate of a formyl glycine generating enzyme) [34], and the peptides A1 and S6 (the substrates of different PPTases) [35]. These labeling reactions, however, require the addition or co-expression of external enzymes. A fourth type of covalent labeling, affinity labeling, does not require enzymes. Although the general concept of "affinity labeling" encompasses many successful strategies [28, 29, 36], those that achieve covalent labeling by a short peptide are scarce. One such labeling reaction was reported by Hamachi et al. Designed to harness the phenomenon of "proximity-induced reactivity" [37–41], a spontaneous nucleophilic S_N2 reaction was found to rapidly cross-link a peptide tag, an N-terminal CA6D4 tag (11 amino acids), and a label, an N-α-chloroacetyl Zn(II)-DpaTyr molecule, when the two bound in vitro [42]. The same researchers recently improved the labeling efficacy by increasing the valency, producing an N-terminal CA6D4 $\times$ 2 tag (16 amino acids) and a tetranuclear Zn(II)-DpaTyr molecule carrying the α-chloroacetyl moiety, which together successfully labeled cell surface proteins [43]. The scarcity of more examples of this strategy reflects the difficulty in satisfying the tradeoff between labeling specificity and tag size, in particular in the context of covalent labeling. Here we sought to achieve specific recognition from a peptide–peptide binding interaction (more specifically the coiled-coil interaction) and the conversion of the noncovalent binding interaction into a site-specific covalent cross-linking reaction (Fig. 5.1). By screening specific positions in two paired coiled-coil peptides, we designed a covalent (irreversible) labeling reaction that is nonenzymatic, spontaneous, and site-specific, and it requires a small tag of only 21 proteinaceous amino acids.

Due to the impermeability of most of the probes, the labeling strategies are often limited to the cell surface protein targeting level. On the other hand, the membrane protein plays a fundamental role in numerous signal pathways as transportation proteins, adaptor proteins, receptor proteins, and enzymes [45]; to label and trace them also become very important for biologists to understand their spatial distribution and temporal variations.

Membrane proteins could be generally divided into integral proteins, peripheral proteins, and lipid-anchored proteins according to different interaction mechanisms with the phospholipid bilayer structure of membrane [46]. Integral protein (also called transmembrane protein), utilizes the hydrophobic amino acids to interact with the hydrophobic lipid tail of the membrane while the hydrophilic amino acids were aligned in one side or both sides of the membrane. Usually the hydrophobic amino acids adopt the alpha-helical structure, making this unique secondary structure to compose 25–50 % of the whole structure; peripheral protein is usually localized on the inside or outside of the membrane through the noncovalent interaction with the polar head of the lipid; while the lipid-linked protein is anchored to the outside of the membrane through a covalent bond with either the lipid bilayer or the membrane sugar.
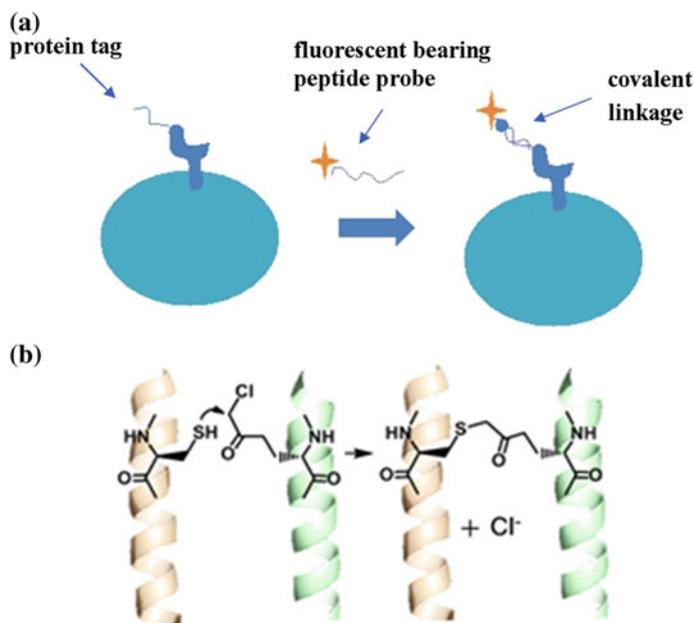
Fig. 5.1 Schematic illustration of the coiled-coil binding induced covalent cross-linking of cell surface proteins. **a** Receptor protein was recognized and then labeled by the fluorescent bearing synthetic peptide probe; **b** Nucleophilic attack reaction happened between the designed coiled-coil partners. Reprinted with permission from Ref. [44]. Copyright (2014) American Chemical Society

G protein–coupled receptors (GPCRs), also known as the seven-transmembrane domain receptors, are a big membrane protein family that mainly sense outside molecules and then activate the intracellular signaling pathways. It is involved in many diseases and is the largest potential recognition site in drug discovery (about 40 % of all modern medicinal drugs) [47]. It contains three major parts: the extracellular N termini that sense the stimuli, the seven helix transmembrane bundles and the intracellular C termini that bound to the G proteins. According to the different G proteins, GPCRs could be divided into Gs-coupled, Gq-coupled, and Gi-coupled subfamilies. The binding of extracellular ligand causes a conformational change of GPCR to release the α subunit of G protein from the β and γ subunits and activate the cAMP pathway or phosphatidylinositol signaling pathway [48].

The epidermal growth factor receptor (EGFR) is a member of the ErbB family of receptors, a subfamily of four closely related receptor tyrosine kinases: ErbB-1, ErbB-2, ErbB-3, and ErbB-4. It is also the cell surface receptor for epidermal growth factor family (EGF family) of extracellular protein ligands [49]. EGFR could also be divided into three major parts: the extracellular part containing domain I–IV, the transmembrane part, and the cytosolic part.

To apply our coiled-coil binding induced covalent labeling strategy described in Chap. 4 to the covalent labeling of membrane proteins, the CCE sequence should be first fused to the extracellular part of the membrane protein to make it accessible to the probe.

5.2   Experimental Section

5.2.1   Construction of B$_2$R Plasmids

B$_2$R gene DNA was purchased from commercial supplier and was subcloned into pCDNA3.1 vector at *BamH* I and *EcoR* I restriction sites. Oligo nucleotide primers were designed to code for the CCE-1′ sequence (KSEESY ECAALEKE VAALEKE VAALEK) together with the start codon and the digested enzyme sites of *Hind* III and *BamH* I, and they were annealed in the annealing buffer (100 mM NaCl, 50 mM HEPES, pH 7.5) to give the double strand DNA fragment. And the fragment was inserted into the digested pCDNA3.1 vector (*Hind* III and *BamH* I) to yield the pCDNA3.1-CCE-1′-B$_2$R plasmid. And the CCE-1′-B$_2$R fragment was amplified by PCR using the primers (forward: 5′-CGGTGGGAGGTCTATATAAGCAGAG-3′ and backward: 5′-CGATGTCGACTGTCTGCTCCCTGCCCAGTC-3′) and digested with *Sal* I and *Hind* III and subcloned into pEGFP-N1 vector to yield pEGFP-N1-CCE-1′-B$_2$R (Appendix 5.1).

Oligonucleotides coding for CCE-9 with a GSGD linker sequence and digested *Hind* III and *BamH* I sites (forward: 5′-AGCTTGGAATGTGCTGCCTTAGA GAAGGAAATTGCAGCGTTAGAGAAGGAAATTGCTGCATTAGAGAAGG GCTCTGGCTCG-3′ and backward: 5′-ACCTTACACGACGGAATCTCTTCC TTTAACGTCGCAATCTCTTCCTTTAACGACGTAATCTCTTCCCGAGACC GAGCCTAG-3′) were annealed and inserted into corresponding pBlueScript-α 7-His10-B$_2$R vector (a kind gift from Prof. Hamachi, Kyoto University) to yield pBS-α7-CCE-9-B$_2$R. DNA fragment coding for EGFP moiety with *BamH* I cleavage sites at both ends was digested from pCI-neo-α7-His10-EGFP-B$_2$R vector (a kind gift from Prof. Hamachi, Kyoto University), and inserted into the *BamH* I digested pBS-α7-CCE-9-B$_2$R vector which was pretreated with dephosphorylation reagent. Finally the DNA fragment coding for α7-CCE-9-EGFP-B$_2$R was digested with *Xho* I and *EcoR* I and subcloned into the corresponding cleavage sites of pCI-neo vector to yield pCI-neo-α7-CCE-9-EGFP-B$_2$R (Appendix 5.2).

5.2.2   Construction of EGFR and hIP Plasmids

The plasmid pDisplay-HA-E3-EGFR (a kind gift from Prof. Shiroh Futaki of Kyoto University) was used as the template to construct pDisplay-CCE-9-EGFR. The Ile to Cys mutation was introduced by a PCR reaction using primers EGFR-F

(5′-GAGCTAGCGAATGCG-CCGCGTTAGAG-3′) and EGFR-R (5′-GACACT CGAGTCATGCTCC-AATAAATTCAC-3′) and subcloned into *Nhe* I and *Xho* I sites of pDisplay-HA-E3-EGFR to yield pDisplay-CCE-9-EGFR (Appendix 5.3).

A CCE-9 peptide sequence (ECAALEKEIAALEKEIAALKE) was introduced to the N terminus of hIP in the pcDNA3.1-3HA-hIP plasmid (a kind gift from Prof. Helen Wise of CUHK) by PCR reaction using primers CCE-9-F (5′-CA AGCTTGCCACCATGGAATGTGCTGCCTTAGAGA-AGGAAATTGCAGCGT TAGAGAAGGAAATTGCTGCATTAGAGAAGATGTACCCATACGATGTTC C-3′) and hIP-R (5′-CGGGATCCTCAGCAGAGGGAGCAGG-CGACGCTG-3′) and subcloned into *Hind* III and *BamH* I sites of pcDNA3.1 (+) plasmid to yield pcDNA3.1-CCE-9-3HA-hIP (Appendix 5.4).

A plasmid expressing CCE-9-hIP-EGFP with the C terminal (cytosolic) end linked to an EGFP tag was also constructed by PCR reaction based on the pcDNA3.1-3HA-hIP using primers CCE-9-F (5′-CAAGCTTGCCACCATG GAATGTGCTGCCTTAGAGAAGGAAATTGCAGCGTTA-GAGAAGGAAAT TGCTGCATTAGAGAAGATGTACCCATACGATGTTCC-3′) and hIP-R-N1 (5′-CGGGATCCGCAGAGGGAGCAGGCGACGCTG-3′) and subcloned into *Hind* III and *BamH* I sites of pEGFP-N1 to yield pEGFP-N1-(CCE-9-hIP-EGFP) (Appendix 5.5).

5.2.3 Cell Culture, Transfection, and Labeling

Chinese Hamster Ovary (CHO) cells or Human Embryonic Kidney 293 (HEK293) cells were grown in DMEM/F12 medium (Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12, Life technology, USA) supplied with 10 % Fetal Bovine Serum (FBS, Life technology, USA) and antibiotics (penicillin/streptomycin, Life technology, USA) in a 10 cm culture dish (Corning, USA) and maintained at 37 °C in a humidified incubator supplied with 5 % CO_2.

For cell labeling, 1.5×10^5 cells were seeded in a 35 mm confocal dish (ibidi, Germany) 1 day prior to the transfection. At ∼40 % confluence, cells were transfected with a DNA (0.8 µg): PLUS reagent (0.8 µl): lipofectamine LTX (2 µl) mixture at a ratio of 1:1:2.5 with the final DNA concentration set to be 1 µg/mL in Opti-mem I (Life technology, USA). Cell medium was changed back to DMEM/F12 5 h posttransfection. After 48–50 h, the cells were pretreated with HEPES buffer containing 0.5 mM TCEP for 10 min at RT, and then incubated with fluorescent peptide probes for 20 min. The cells were then washed, fixed by 4 % paraformaldehyde in PBS (w/v), and then incubated with anti-HA-FITC antibody (in 1:300 dilution) (Sigma-Aldrich, USA) at RT for 1 h. The cells were washed for at least three times by PBS (10 min each) and imaged by confocal fluorescent microscope.

5.2.4 Extracting Covalently Labeled EGFR
for Gel Electrophoresis

1.5×10^6 CHO cells were seeded in a 10 cm dish (Corinng, USA) 1 day prior to the scale-up transfection. 60 h after transfection, the cells were pretreated with HEPES buffer containing 0.5 mM TCEP and then incubated with 100 μM CCK-9-dimer probe for 2 h at 37 °C. After being rinsed in PBS, the cells were scraped into 1.5 mL PBS and pelleted by centrifugation at 12000 rpm for 15 min at 4 °C. The cells were then lysed in RIPA buffer (Sigma, USA. 50 mM Tris-HCl, pH 8.0, with 150 mM sodium chloride, 1.0 % Igepal CA-630 (NP-40), 0.5 % sodium deoxycholate, and 0.1 % sodium dodecyl sulfate) containing protease inhibitor cocktail (Promega, USA) for 45 min at 4 °C. The supernatant was collected by centrifugation at 12000 rpm at 4 °C, mixed, incubated with $5 \times$ protein loading dye for 40 min at RT, denatured, and subjected to SDS-PAGE and subsequent in-gel fluorescence scanning.

5.2.5 cAMP Assay for Drug Responsiveness
of hIP Transfected Cells

Adenylyl cyclase activity was conducted based on previous reported protocols [50, 51]. 5×10^5 CHO cells were seeded in poly-D-lysine coated 12-well plates (Corning, USA) and cultured in DMEM with 10 % FBS for 1 day. Cells were transfected for 40 h, then replace cell medium with 1 mL DMEM with 1 % FBS (reducing serum to facilitate the uptake of [^{3}H]adenine). Add 10 μL [^{3}H]-adenine (1 μCi/10 μL) (Amersham Biosciences, Hong Kong) to each well and return plates to incubator for overnight. Aspirate the medium to remove free [^{3}H]-adenine and wash each well twice with 1 mL complete HEPES buffer. Then 0.5 mL assay solution (HEPES buffer containing 1 mM 3-Isobutyl-1-methylxanthine (IBMX), to inhibit cyclic nucleotide phosphodiesterase activity) was added to each well and incubated for 15 min in 37 °C water bath. 10 μL compound stock solution (forskolin, cicaprost or buffer) was added to each well, and incubated for 30 min in 37 °C water bath. The reaction was stopped by addition of ice-cold trichloroacetic acid and ATP, and [^{3}H]-cAMP was separated from [^{3}H]-ATP by column chromatography, and counted in a liquid scintillation counter using OptiPhase 'HiSafe' 3 scintillant (Pharmacia Biotech Far East Ltd, Hong Kong). The production of [^{3}H]-cAMP from cellular [^{3}H]-ATP was estimated as the ratio of radiolabeled cAMP to total AXP (i.e., adenosine, ADP, ATP and cAMP), and is expressed as [cAMP]/[total AXP] $\times$ 100 (i.e., % conversion). Cicaprost was a gift from Schering AG, Berlin, Germany. All assays were performed in triplicate.

5.3 Results and Discussion

5.3.1 Cell Labeling Based on B_2R

Bradykinin type II receptor (B_2R) is a known Gq-coupled GPCR [52]. After fusing the CCE-1′ sequence to the N terminus of B_2R, the C terminus of the protein was also installed an EGFP moiety to introduce an orthogonal fluorescent signal. To ensure the expressed EGFP protein was a fusion part of B_2R instead of a random fusion of other nonrelated proteins, HEK293 cells transiently expressing the CCE-1′-B_2R-EGFP protein were lysed and loaded to a SDS-PAGE and followed by the western blotting using antibodies of anti-EGFP and anti-B_2R, respectively. And the results showed that the recombinant protein of CCE-1′-B_2R-EGFP was expressed as a fusion with a molecular weight of ~ 70 kD, which was close to that of the recombinant protein (Fig. 5.2).

After confirming the expression of the recombinant protein, HEK293 cells expressing CCE-1′-B_2R-EGFP were grown on the glass cover slips in a six-well format, and pretreated with the reducing reagent (HBS buffer, containing 0.5 mM TCEP) to expose the active Cys, the cells were then incubated with gradient concentrations of TMR-CCK-1 probe and followed by the extensive PBS washing. The cover slip was fixed and viewed under a conventional fluorescent microscope. Although the red and green fluorescent signals showed a high degree of overlap, they seemed to be localized in the peri-nucleus region instead of cell surface (Fig. 5.3). So the labeling reaction failed.

Pictures taken from the conventional fluorescent microscope were accumulation results along Z-axis. We then used a confocal fluorescent microscope to give a more accurate resolution in live cells. HEK293 cells were seeded to a 35 mm confocal dish, and transfected with Fugene 6 reagent. 48–50 h after transfection, the cells were treated with TMR-CCK-1-dimer probe. Some cells were successfully transfected, because the membrane was lighten up with a green fluorescence under confocal microscope, but no red fluorescent signal (TMR) was observed (Fig. 5.4, middle). Some cells contained both EGFP and TMR fluorescent signals but they appeared to be dead cells (Fig. 5.4, bottom and top).

Fig. 5.2 Western blotting results of the CCE-1′-B_2R-EGFP protein. *Left*, protein bands stained with anti-B_2R antibody; *Right*, protein bands stained with anti-EGFP antibody. Both bands were of a molecular weight of about 70 K
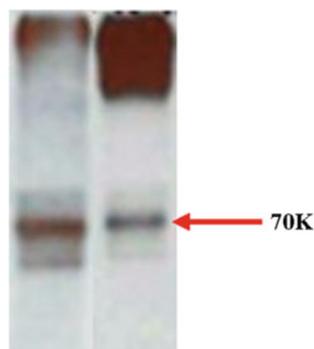
Fig. 5.3 Photographs showing the HEK293 cells expressing CCE-9-B$_2$R-EGFP treated with TMR-CCK-9 synthetic probe. *Left*, picture taken from EGFP channel; *right*, picture taken from TMR channel. [TMR-CCK-9] = 2 μM
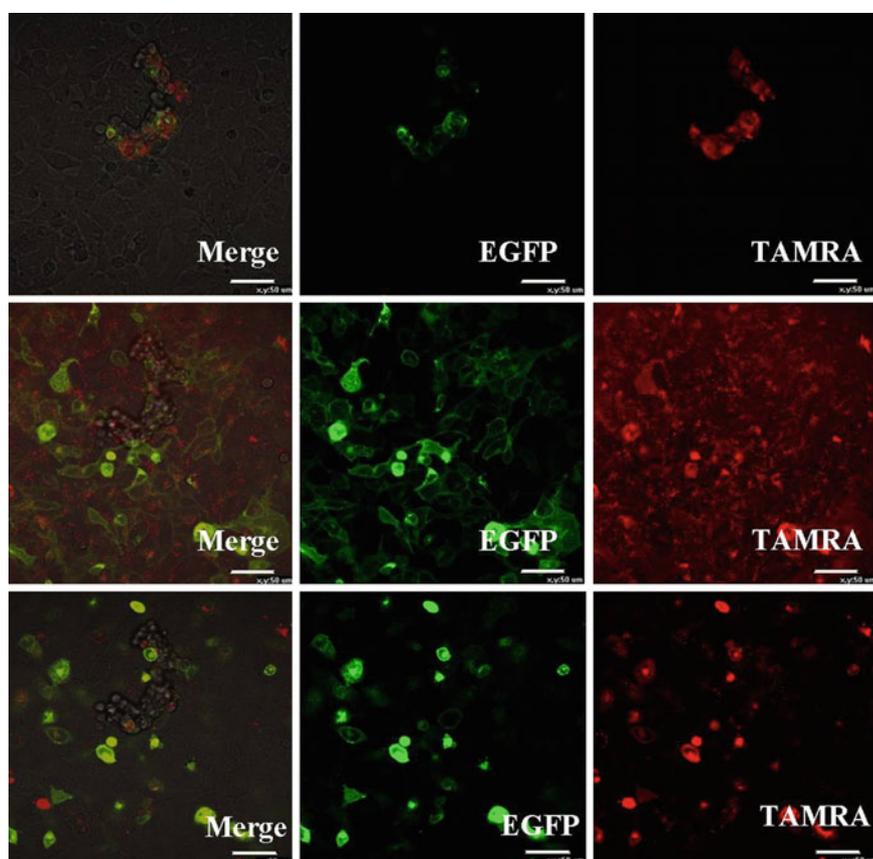
Fig. 5.4 Photographs showing the overview of the TMR-CCK-1-dimer treated HEK 293 cells. *Top*, floating cells with both EGFP and TMR signals; *middle*, adherent cells with only EGFP signal; *bottom*, rounded up cells with both EGFP and TMR signals. [TMR-CCK-1-dimer] = 2 μM, scale bar 50 μm

These experiments indicated a low transfection level and transfection of B$_2$R protein seemed to toxic to the cells. Figure 5.5a supports this notion as the expressing level of the recombinant protein increases, the morphology of the cell decreases, and finally the highly expression of the proteins could cause the disruption of the cell membrane and all the fluorescent dyes would go into the cells and light up the cells without selectivity. Also shown in Fig. 5.5b, cells with membrane compromised readily uptakes fluorescent peptide nonselectively.

The cytotoxicity of the plasmid/protein might come from the C terminal fusion of EGFP because of the large size of the moiety. The conformational change of B$_2$R might lead to low labeling efficacy. Besides, it is known that the typical GPCRs adopt the seven helical transmembrane structure with each of the helical contains about 20–30 hydrophobic amino acids. CCE-1′ sequence, being another amphiphilic helical peptide, likely interacts with the helical domain. We then moved the EGFP moiety to the N terminus. Also to improve the binding affinity of the tag and probe, the VAAL interacting model is changed to IAAL, which was reported to
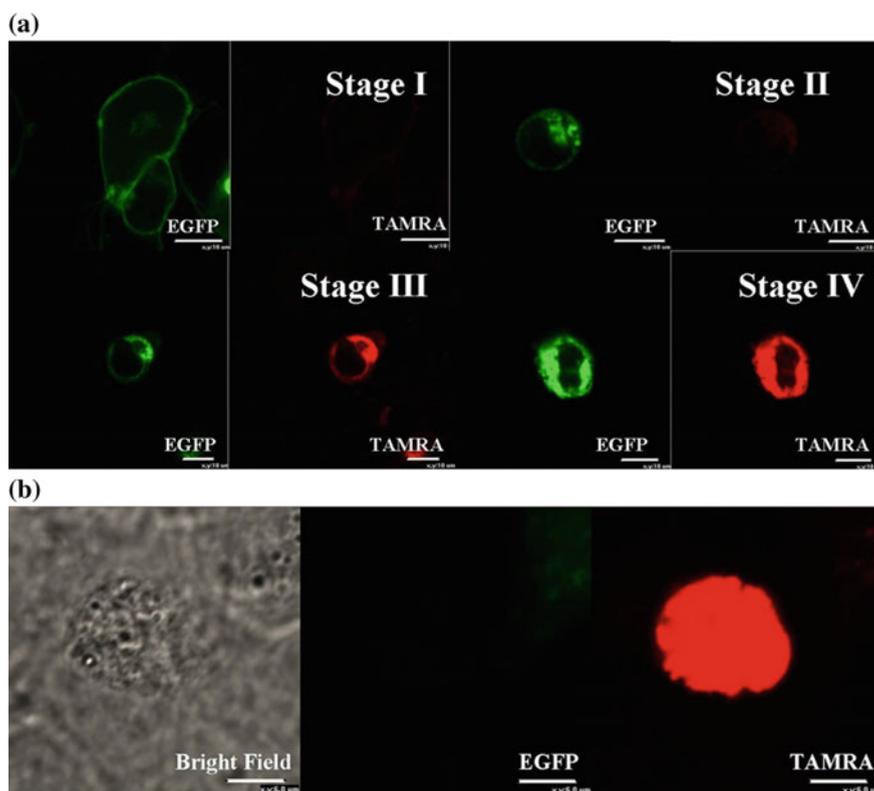
Fig. 5.5 Photographs showing the proposed step by step apoptosis process. **a** Cells were of different morphologies due to the different protein expressing levels; **b** A single cell was light up by the TMR fluorescent dye due to the nonselective permeation. Scale bar 10 μm

possess a higher affinity with CCK peptide. Moreover, an extra N-terminal fusion of a leading sequence (α7) was included to greatly enhance the membrane location of the B_2R. A new plasmid coding for α7-CCE-9-EGFP-B_2R was constructed and the procedure was described in experimental section.

After transfected with the newly constructed pCI-neo-α7-CCE-9-EGFP-B_2R plasmid, the CCE-9-EGFP-B_2R fusion protein expressed by HEK293 cells showed an obvious membrane distribution according to the indicated EGFP signal, and after the treatment of the cells with 1 μM dimeric peptide probe bearing the TMR fluorophores, some of the cells showed a well co-localized TMR signal with EGFP, indicating the recognition of the fusion protein by the synthetic probe and the achievement of the covalent reaction (Fig. 5.6).

5.3.2   Cell Labeling Based on hIP

However, we have observed a low labeling efficacy and the function of the receptor might also have been compromised. For example, EGFP tagging might cause problems in the translocation of GPCRs to the plasma membrane; the fusion protein might be trapped in cellular organisms such as Golgi apparatus [53]. It was also likely that the single strand of CCE (E IAALEKE IAALEKE IAALEK) might form a homodimer alone at high concentration. In view of these possible drawbacks, we decided to modify the system, by changing the target protein the Gq-coupled GPCR B_2R to a Gs-coupled GPCR the human prostacyclin receptor (hIP), and the EGFP tag to HA tag (only nine amino acids).

Since there were two tags to be labeled, HA tag by the antibody and the CCE-9 tag to be labeled by the peptide probe, the order of the labeling reactions needs to be carefully set. As antibody labeling requires cell fixing and the cell membrane is permeabilized during fixing, the following labeling step will not be specific at cell membrane (Fig. 5.7). Furthermore, it was observed that HEK293 cell line showed a strong nonspecific binding of the positively charged TMR dyes due to the highly
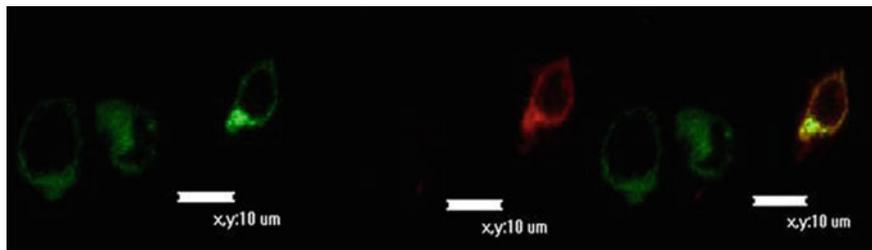
Fig. 5.6 Photographs showing the membrane localization of the CCE-9-EGFP-B_2R recombinant protein and the successfully labeling of the peptide probe to the CCE-9 tag sequence. *Left*, EGFP fluorescent channel; *middle*, TMR fluorescent channel; *right*, merged channels showing the co-localization of the both fluorescent signals. Scale bar 10 μm
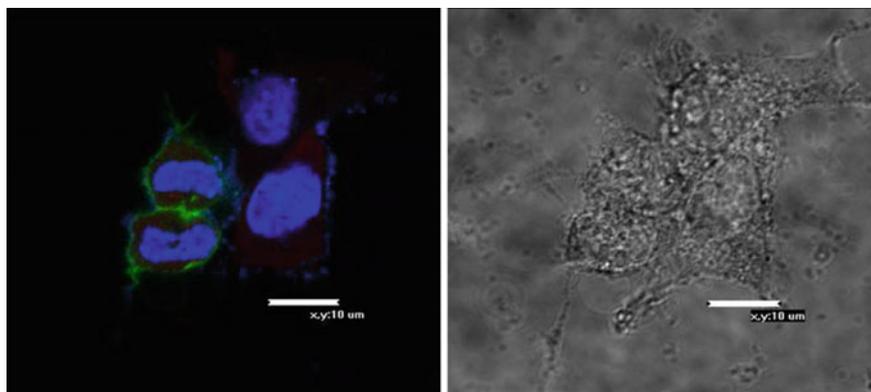
Fig. 5.7 Photographs showing the distribution of different fluorescent species in the antibody first labeling strategy. *Green* represents for the FITC-anti-HA antibody (membrane), *blue* represents for DAPI (nucleus), *red* represents for the TMR appended peptide probe (cytosol). And *left* was the merge of different fluorescent channels, *right* was the image taken from bright field. Scale bar 10 μm

negative charges presented on the cell membrane, so Chinese Hamster Ovary (CHO) cells are used instead.

The CHO cells were transfected with the new plasmid pcDNA3.1-CCE-9-3HA-hIP, as well as an empty pcDNA3.1 plasmid as a negative control and a pcDNA3.1-3HA-hIP as a positive control, and then labeled by the synthetic TMR-CCK-9-dimer probe, fixed and labeled by the FITC containing anti-HA antibody, imaged under a confocal microscope (Fig. 5.8). The 3HA-hIP group showed an obvious halo-shaped FITC signal, indicating that the 3HA-hIP protein could be expressed and displayed at the membrane, and it could be recognized by the anti-HA antibody. Both labels co-localize in the case of CCE-9-3HA-hIP protein (Fig. 5.8).

To exam whether the fusion of the CCE-9 sequence would interfere the functional property of the hIP receptor protein, the cAMP accumulation measurement was conducted. The mechanism of a Gs-coupled GPCR cAMP pathway was shown in Fig. 5.9a. Upon agonist binding and activation of the Gs-coupled GPCR, the conformational change of the GPCR causes the release of the β and γ subunits of the G protein. The released subunits then further activate another membrane protein enzyme the adenylyl cyclase which then converts ATP to cAMP. So the concentration of the cAMP indicates the functional property of the GPCR proteins.

5.3.3   Cell Labeling Based on EGFR

HEK293 cells were utilized due to its high protein expressing level for a cAMP assay. In all the cell groups forskolin stimulation of the adenylyl cyclase gave
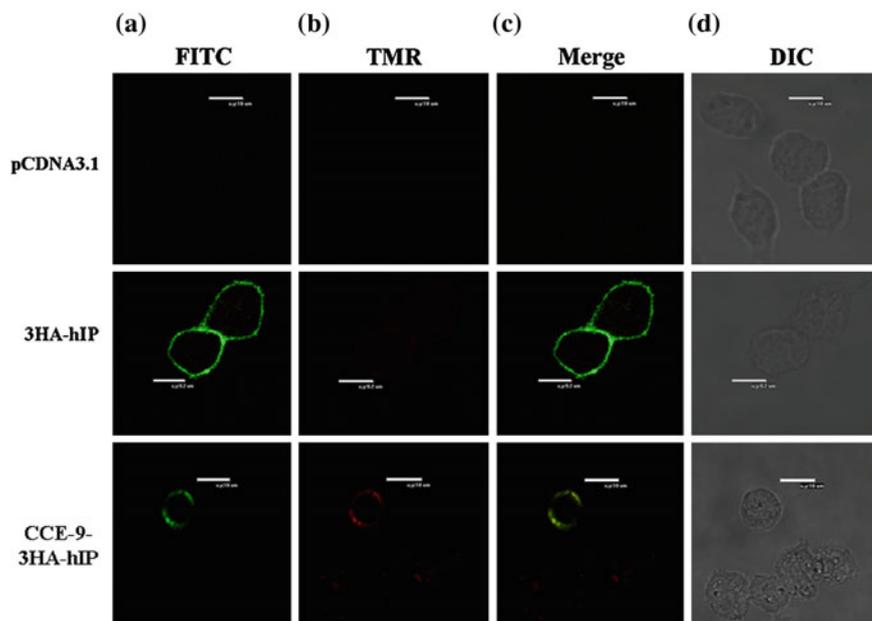
Fig. 5.8 Photographs showing the labeling of CCE-9-hIP on cell surface by the TMR-CCK-9-dimer and FITC-anti-HA antibody. **a** FITC channel; **b** TRITC channel; **c** overlay of (**a**) and (**b**); **d** bright field. *Top*, CHO cells transfected with empty pCDNA3.1 vector; *middle*, CHO cells transfected with pCDNA3.1-3HA-hIP vector; *bottom*, CHO cells transfected with pCDNA3.1-CCE-9-3HA-hIP vector. Scale bar 10 μm. Reprinted with permission from Ref. [44]. Copyright (2014) American Chemical Society

similar level of cAMP increase, indicating that the expression of CCE-9-3HA-hIP did not affect the function of the cells. Cicaprost stimulation of the hIP receptor generated comparable [³H]-cAMP conversion in 3HA-hIP and CCE-9-3HA-hIP groups, indicating that the genetic fusion of CCE-9 sequence to the hIP protein did not affect the functional property of hIP (Fig. 5.9b).

This result indicated that our tag has advantage due to the smaller size. The receptor remains active and functional after the recombinant fusion of the extra tag sequence. For comparison, we generated a pEGFP-CCE-9-3HA-hIP plasmid for the expression of the CCE-9-3HA-hIP-EGFP fusion. We observed that EGFP fusion proteins were trapped in the cytosol of the cells (Fig. 5.10). Notwithstanding the successful labeling, it is not proven that it is through covalent labeling. Futaki et al. reported that replacing the extracellular domains I, II, and III, and a part of domain IV of EGFR, with the CCE3′ sequence EIAALEKEIAALEKEIAALEK and an HA tag did not affect the function of the intracellular domain of the receptor. [54] Based on this we mutated the first Ile to Cys to yield CCE-9-EGFR, and the same experiments were conducted based on CHO cells transfected with pDisplay-CCE-9-EGFR and an empty pDisplay vector (Fig. 5.11). CHO cells expressing CCE-9-EGFR exhibited both green fluorescence (FITC-anti-HA antibody) and red fluorescence
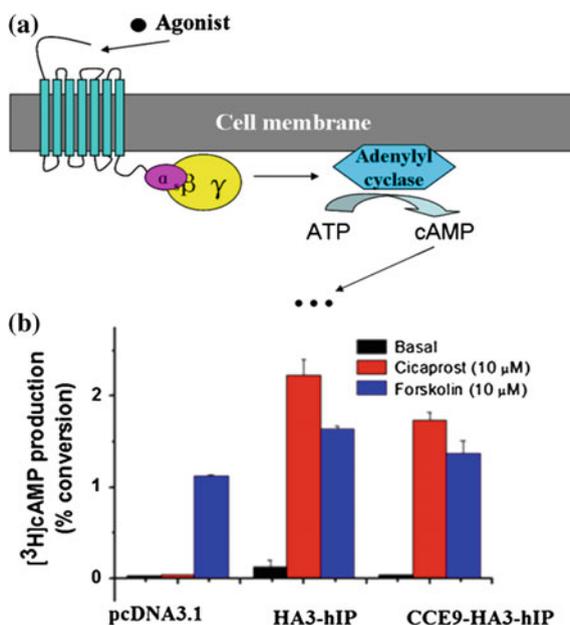
Fig. 5.9 Illustration of the biological activity of the engineered recombinant cell surface GPCRs. **a** Mechanism of the cAMP assay; **b** [^{3}H]-cAMP accumulation in HEK293 cells transfected with different DNA plasmids. Reprinted with permission from Ref. [44]. Copyright (2014) American Chemical Society

Fig. 5.10 C terminal tagging of hIP with an EGFP disrupt the membrane localization of the receptor. Also CHO cells transfected by pEGFP-N1-(CCE-9-hIP-EGFP) failed to be labeled by TMR-CCK-9-dimer probe. Scale bar = 10 μm. Reprinted with permission from Ref. [44]. Copyright (2014) American Chemical Society

(TMR from the CCK-9-dimer) around the plasma membrane, indicating that the CCE-9 tag on EGFR was successfully labeled (Fig. 5.11A). The fluorescent signal along the transection of a labeled cell also confirmed that the FITC signal and the TMR signal co-localized at the plasma membrane (Fig. 5.12). To confirm the covalent nature of the linkage between the label CCK-9-dimer and the tag CCE-9 on EGFR, surface-labeled CHO cells were harvested and lysed. At the molecular

weight of ∼145 kDa, a fluorescent band was clearly visible in the lane for
pDisplay-CCE-9-EGFR transfected cells but not in control lanes for pDisplay-
CCE3′-EGFR or pDisplay transfected cells (Fig. 5.11B). The CCK-9-dimer thereby
"locked" surface-expressed CCE-9-EGFR in dimeric form. The covalent linkage
showed marked advantage here as it allowed us to visualize the probe-tag complex
under denaturing condition.
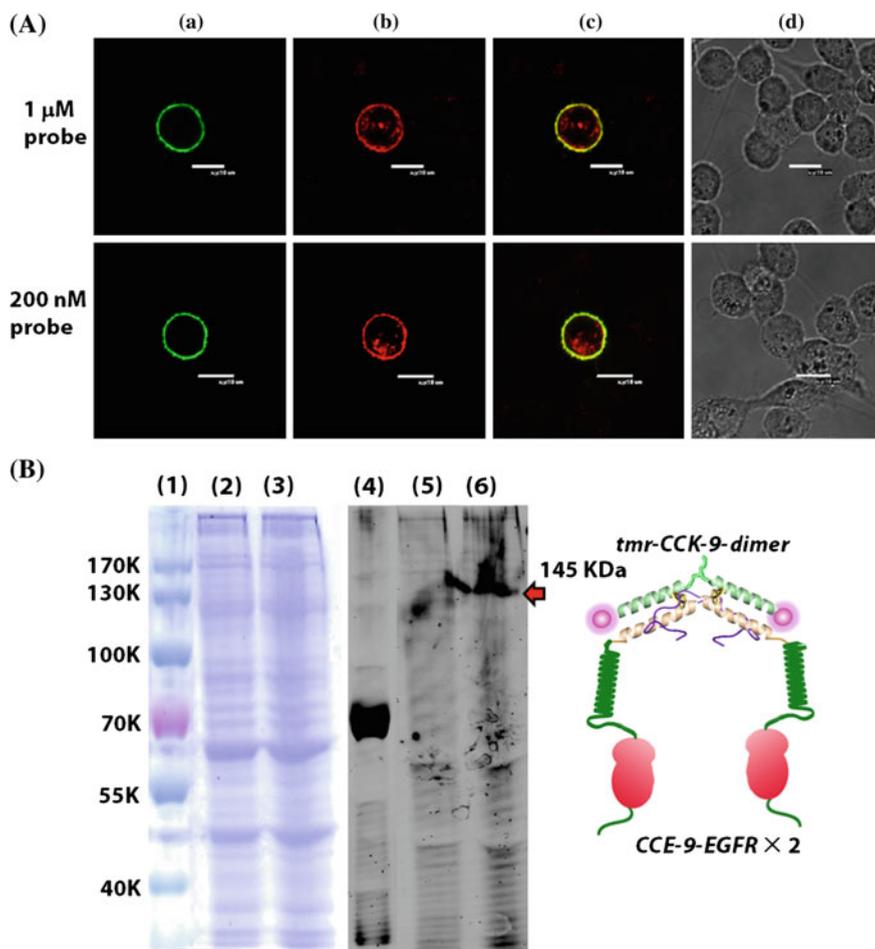
Fig. 5.11 Covalent labeling of CCE-9-EGFR on a cell surface. **A** Confocal fluorescent images of
CHO cells expressing CCE-9-EGFR and treated with 1 μM or 200 nM TMR-CCK-9-dimer and
FITC-anti-HA antibody. Scale bar = 10 μm. *a* FITC channel; *b* TRITC channel; *c* overlay of
(*a*) and (*b*); *d* bright field. **B** Covalently labeled CCE-9-EGFR in the membrane fraction by
TMR-CCK-9-dimer. Lanes (*1*)–(*3*) show the gel image stained by Coomassie blue; lanes (*4*)–(*6*)
show the corresponding fluorescent image acquired at TRITC channel of a Typhoon imager. Lane
(*1*) and (*4*) molecular weight markers; lane (*2*) and (*5*) cells transfected by pDisplay control; lane
(*3*) and (*6*) cells transfected by pDisplay-CCE-9-EGFR. Reprinted with permission from Ref. [44].
Copyright (2014) American Chemical Society
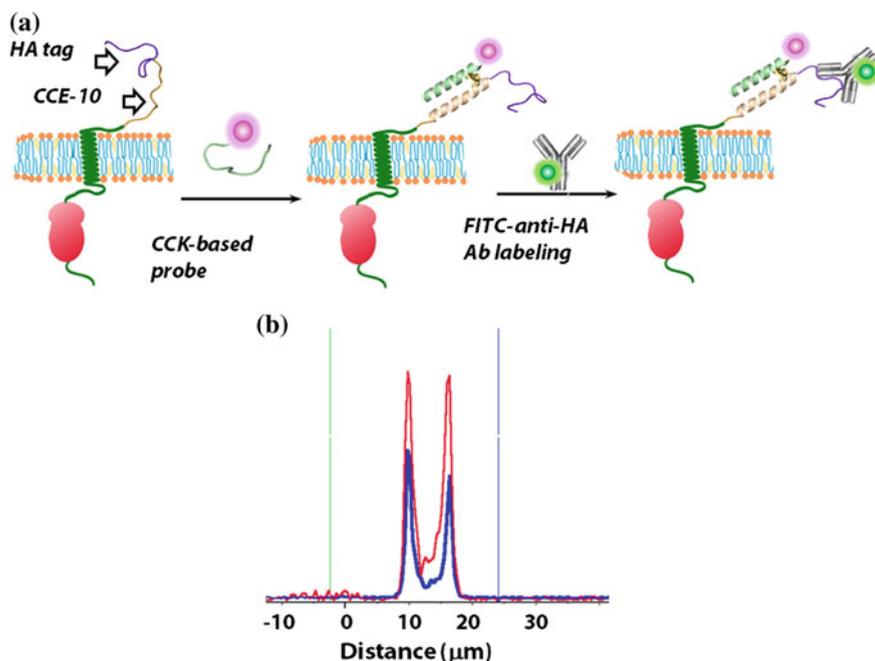
**Fig. 5.12** Covalent labeling of *CCE-9-EGFR* on a cell surface. **a** Scheme of the two-step labeling procedure. **b** Fluorescent signal of the cell intersection shows membrane labeling. *Red*, TRITC channel; *blue*, FITC channel. Reprinted with permission from Ref. [44]. Copyright (2014) American Chemical Society

5.4 Conclusion

In this chapter, we successfully applied the covalent cross-linking reaction to the labeling of surface proteins. Full length GPCR receptors, B_2R, and hIP and a truncated EGFR were examined. The specificity of the strategy harnesses the secondary structure of a protein (the coiled-coil), rather than primary (chemical reactivity) or three-dimensional (enzymes or other sophisticated protein structures) folds. We successfully addressed the tradeoff between the size of the tag and the selectivity of the reaction: the binding of a coiled-coil interaction provides the specificity and biocompatibility of the covalent reaction, and the Cys–chloroacetyl nucleophilic S_N2 reaction driven by enhanced local reactivity permanently and spontaneously "freezes" the interaction. One limitation of our strategy, which similar approaches will share, is the impermeability of the probe, which limits the labeling to cell surface receptors [24]. But the combined use of surface-specific probes and cell-permeable probes has the advantage of differentiating surface-presented versus intracellular proteins. Having demonstrated our cross-linking strategy, we are now engineering other short peptide tags in order to

simultaneously label multiple proteins of interest. Multiple orthogonal pairs of covalent probes might allow sequential labeling of cell surface receptors to track the fate of surface-presented receptors during internalization, intracellular targeting, and recycling.

Appendix 5.1 Plasmid Information of pEGFP-N1-CCE-1′-B₂R
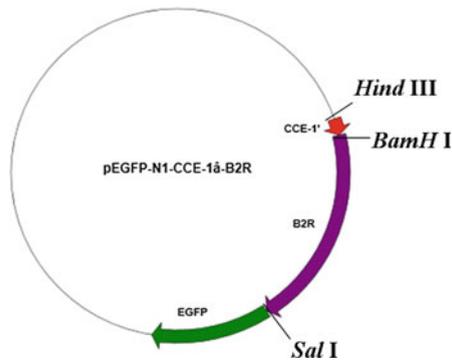
CTTATGAATCTGAGAGTCTTATGAATGTGCTGCCTTAGAGAAGGAAGTTGCAGCG
TTAGAGAAGGAAGTTGCTGCATTAGAGAAGGCGGATCCCTTCTCTCCCTGGAAGA
TATCAATGTTTCTGTCTGTTCGTGAGGACTCCGTGCCCACCACGGCCTCTTTCAG
CGCCGACATGCTCAATGTCACCTTGCAAGGGCCCACTCTTAACGGGACCTTTGCC
CAGAGCAAATGCCCCCAAGTGGAGTGGCTGGGCTGGCTCAACACCATCCAGCCCC
CCTTCCTCTGGGTGCTGTTCGTGCTGGCCACCCTAGAGAACATCTTTGTCCTCAG
CGTCTTCTGCCTGCACAAGAGCAGCTGCACGGTGGCAGAGATCTACCTGGGGAAC
CTGGCCGCAGCAGACCTGATCCTGGCCTGCGGGCTGCCCTTCTGGGCCATCACCA
TCTCCAACAACTTCGACTGGCTCTTTGGGGAGACGCTCTGCCGCGTGGTGAATGC
CATTATCTCCATGAACCTGTACAGCAGCATCTGTTTCCTGATGCTGGTGAGCATC
GACCGCTACCTGGCCCTGGTGAAAACCATGTCCATGGGCCGGATGCGCGGCGTG
CGCTGGGCCAAGCTCTACAGCTTGGTGATCTGGGGGTGTACGCTGCTCCTGAGC
TCACCCATGCTGGTGTTCCGGACCATGAAGGAGTACAGCGATGAGGGCCACAACG
TCACCGCTTGTGTCATCAGCTACCCATCCCTCATCTGGGAAGTGTTCACCAACAT
GCTCCTGAATGTCGTGGGCTTCCTGCTGCCCCTGAGTGTCATCACCTTCTGCACG
ATGCAGATCATGCAGTGCTGCGGAACAACGAGATGCAGAAGTTCAAGGGAGATCC
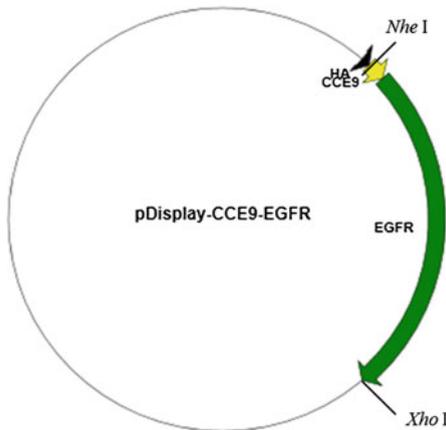AGACGGAGAGGAGGGCCACGGTGCTAGTCCTGGTTGTGCTGCTGCTATTCATCAT
CTGCTGGCTGCCC

Appendix 5.2 Plasmid Information
of pCI-neo-α7-CCE-9-EGFP-B₂R
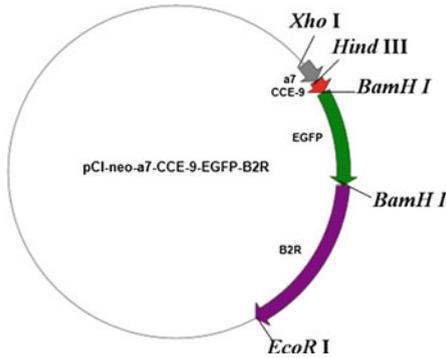
CTCGAGCCACCATGGGCCTCCGGGCGCTGATGCTGTGGCTGCTGGCGGCGGCGGG
GCTCGTGCGCGAGTCCCTGCAAGGAGAGTTCCAAAGGAAGCTGTACAAGGATCGA
GGaAGCTTGGAATGTGCTGCCTTAGAGAAGGAAATTGCAGCGTTAGAGAAGGAAA
TTGCTGCATTAGAGAAGGGCTCTGGCTCGGATCCAATGGTGAGCAAGGGCGAGG
AGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCT
GACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC
GTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCAC
CATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATAT
CATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAAC
ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCG
GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACC
GCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGCGGATCCAGCCCAGA
TGTTCTCTCCCTGGAAGATATCAATGTTTCTGTCTGTTCGTGAGGACTCCGTGCC
CACCACGGCCTCTTTCAGCGCCGACATGCTCAATGTCACCTTGCAAGGGCCCACT
CTTAACGGGACCTTTGCCCAGAGCAAATGCCCCCAAGTGGAGTGGCTGGGCTGGC
TCAACACCATCCAGCCCCCCTTCCTCTGGGTGCTGTTCGTGCTGGCCACCCTAGA
GAACATCTTTGTCCTCAGCGTCTTCTGCCTGCACAAGAGCAGCTGCACGGTGGCA
GAGATCTACCTGGGGAACCTGGCCGCAGCAGACCTGATCCTGGCCTGCGGGCTGC
CCTTCTGGGCCATCACCATCTCCAACAACTTCGACTGGCTCTTTGGGGAGACGCT
CTGCCGCGTGGTGAATGCCATTATCTCCATGAACCTGTACAGCAGCATCTGTTTC
CTGATGCTGGTGAGCATCGACCGCTACCTGGCCCTGGTGAAAACCATGTCCATGG
GCCGGATGCGCGGCGTGCGCTGGGCCAAGCTCTACAGCTTGGTGATCTGGGGGTG
TACGCTGCTCCTGAGCTCACCCATGCTGGTGTTCCGGACCATGAAGGAGTACAGC
GATGAGGGCCACAACGTCACCGCTTGTGTCATCAGCTACCCATCCCTCATCTGGG

```
AAGTGTTCACCAACATGCTCCTGAATGTCGTGGGCTTCCTGCTGCCCCTGAGTGT
CATCACCTTCTGCACGATGCAGATCATGCAGGTGCTGCGGAACAACGAGATGCAG
AAGTTCAAGGAGATCCAGACGGAGAGGAGGGCCACGGTGCTAGTCCTGGTTGTGC
TGCTGCTATTCATCATCTGCTGGCTGCCCTTCCAGATCAGCACCTTCCTGGATAC
GCTGCATCGCCTCGGCATCCTCTCCAGCTGCCAGGACGAGCGCATCATCGATGTA
ATCACACAGATCGCCTCCTTCATGGCCTACAGCAACAGCTGCCTCAACCCACTGG
TGTACGTGATCGTGGGCAAGCGCTTCCGAAAGAAGTCTTGGGAGGTGTACCAGGG
AGTGTGCCAGAAAGGGGGCTGCAGGTCAGAACCCATTCAGATGGAGAACTCCATG
GGCACACTGCGGACCTCCATCTCCGTGGAACGCCAGATTCACAAACTGCAGGACT
GGGCAGGGAGCAGACAGTGAGCAAACGCCAGCAGGGCTGCTGTGAATTC
```

Appendix 5.3 Plasmid Information
of pDisplay-CCE-9-EGFR


```
TATCCATATGATGTTCCAGATTATGCTGGGGCCGCTAGCGAATGCGCCGCGTTA
GAGAAAGAAATAGCGGCTCTAGAAAAGGAGATTGCAGCTCTTGAGAAGGACGCC
GGCCATGTGTGCCACCTGTGCCATCCAAACTGCACCTACGGATGCACTGGGCCAG
GTCTTGAAGGCTGTCCAACGAATGGGCCTAAGATCCCGTCCATCGCCACTGGGAT
GGTGGGGGCCCTCCTCTTGCTGCTGGTGGTGGCCCTGGGGATCGGCCTCTTCATG
CGAAGGCGCCACATCGTTCGGAAGCGCACGCTGCGGAGGCTGCTGCAGGAGAGGG
AGCTTGTGGAGCCTCTTACACCCAGTGGAGAAGCTCCCAACCAAGCTCTCTTGAG
GATCTTGAAGGAAACTGAATTCAAAAAGATCAAAGTGCTGGGCTCCGGTGCGTTC
GGCACGGTGTATAAGGGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCCG
TCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCT
CGATGAAGCCTACGTGATGGCCAGCGTGGACAACCCCCACGTGTGCCGCCTGCTG
GGCATCTGCCTCACCTCCACCGTGCAACTCATCACGCAGCTCATGCCCTTCGGCT
GCCTCCTGGACTATGTCCGGGAACACAAAGACAATATTGGCTCCCAGTACCTGCT
CAACTGGTGTGTGCAGATCGCAAAGGGCATGAACTACTTGGAGGACCGTCGCTTG
```

GTGCACCGCGACCTGGCAGCCAGGAACGTACTGGTGAAAACACCGCAGCATGTCA
AGATCACAGATTTTGGGCTGGCCAAACTGCTGGGTGCGGAAGAGAAAGAATACCA
TGCAGAAGGAGGCAAAGTGCCTATCAAGTGGATGGCATTGGAATCAATTTTACAC
AGAATCTATACCCACCAGAGTGATGTCTGGAGCTACGGGGTGACCGTTTGGGAGT
TGATGACCTTTGGATCCAAGCCATATGACGGAATCCCTGCCAGCGAGATCTCCTC
CATCCTGGAGAAAGGAGAACGCCTCCCTCAGCCACCCATATGTACCATCGATGTC
TACATGATCATGGTCAAGTGCTGGATGATAGACGCAGATAGTCGCCCAAAGTTCC
GTGAGTTGATCATCGAATTCTCCAAAATGGCCCGAGACCCCCAGCGCTACCTTGT
CATTCAGGGGGATGAAAGAATGCATTTGCCAAGTCCTACAGACTCCAACTTCTAC
CGTGCCCTGATGGATGAAGAAGACATGGACGACGTGGTGGATGCCGACGAGTACC
TCATCCCACAGCAGGGCTTCTTCAGCAGCCCCTCCACGTCACGGACTCCCCTCCT
GAGCTCTCTGAGTGCAACCAGCAACAATTCCACCGTGGCTTGCATTGATAGAAAT
GGGCTGCAAAGCTGTCCCATCAAGGAAGACAGCTTCTTGCAGCGATACAGCTCAG
ACCCCACAGGCGCCTTGACTGAGGACAGCATAGACGACACCTTCCTCCCAGTGCC
TGAATACATAAACCAGTCCGTTCCCAAAAGGCCCGCTGGCTCTGTGCAGAATCCT
GTCTATCACAATCAGCCTCTGAACCCCGCGCCCAGCAGAGACCCACACTACCAGG
ACCCCCACAGCACTGCAGTGGGCAACCCCGAGTATCTCAACACTGTCCAGCCCAC
CTGTGTCAACAGCACATTCGACAGCCCTGCCCACTGGGCCCAGAAAGGCAGCCAC
CAAATTAGCCTGGACAACCCTGACTACCAGCAGGACTTCTTTCCCAAGGAAGCCA
AGCCAAATGGCATCTTTAAGGGCTCCACAGCTGAAAATGCAGAATACCTAAGGGT
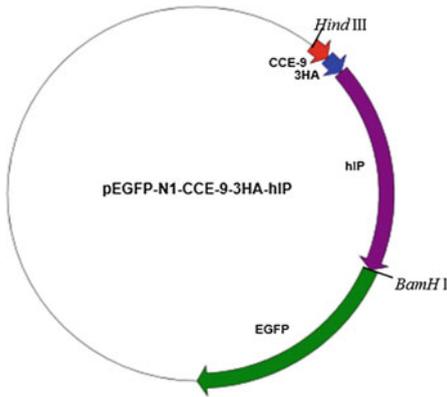CGCGCCACAAAGCAGTGAATTTATTGGAGCATGACCCGGGTAACTCGAG

Appendix 5.4 Plasmid Information of pcDNA3.1-CCE-9-3HA-hIP

CAAGCTTGCCACCATGGAATGTGCTGCCTTAGAGAAGGAAATTGCAGCGTTAGA
GAAGGAAATTGCTGCATTAGAGAAGATGTACCCATACGATGTTCCAGATTACGC
TTACCCATACGATGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGCT
GATGCGGATTCGTGCAGGAACCTCACCTACGTGCGGGGCTCGGTGGGGCCGGCCA

CCAGCACCCTGATGTTCGTGGCCGGTGTGGTGGGCAACGGGCTGGCCCTGGGCAT
CCTGAGCGCACGGCGACCGGCGCGCCCCTCGGCCTTCGCGGTGCTGGTGACCGGA
CTGGCGGCCACCGACCTGCTGGGCACCAGCTTCCTGAGCCCGGCCGTGTTCGTGG
CCTATGCGCGCAACAGCTCCCTGCTGGGCCTGGCCCGAGGCGGCCCCGCCCTGTG
CGATGCCTTCGCCTTCGCCATGACCTTCTTCGGCCTGGCGTCCATGCTCATCCTC
TTTGCCATGGCCGTGGAGCGCTGCCTGGCGCTGAGCCACCCCTACCTCTACGCGC
AGCTGGACGGGCCCCGCTGCGCCCGCCTGGCGCTGCCAGCCATCTACGCCTTCTG
CGTCCTCTTCTGCGCGCTGCCCCTGCTGGGCCTGGGCCAACACCAGCAGTACTGC
CCCGGCAGCTGGTGCTTCCTCCGCATGCGCTGGGCCCAGCCGGGCGGCGCCGCCT
TCTCGCTGGCCTACGCCGGCCTGGTGGCCCTGCTGGTGGCTGCCATCTTCCTCTG
CAACGGCTCGGTCACCCTCAGCCTCTGCCGCATGTACCGCCAGCAGAAGCGCCAC
CAGGGCTCTCTGGGTCCACGGCCGCGCACCGGAGAGGACGAGGTGGACCACCTGA
TCCTGCTGGCCCTCATGACAGTGGTCATGGCCGTGTGCTCCCTGCCTCTCACGAT
CCGCTGCTTCACCCAGGCTGTCGCCCCTGACAGCAGCAGTGAGATGGGGGACCTC
CTTGCCTTCCGCTTCTACGCCTTCAACCCCATCCTGGACCCCTGGGTCTTCATCC
TTTTCCGCAAGGCTGTCTTCCAGCGACTCAAGCTCTGGGTCTGCTGCCTGTGCCT
CGGGCCTGCCCACGGAGACTCGCAGACACCCCTTTCCCAGCTCGCCTCAGGGAGG
AGGGACCCAAGGGCCCCCTCTGCTCCTGTGGGAAAGGAGGGGAGCTGCGTGCCTT
TGTCGGCTTGGGGCGAGGGGCAGGTGGAGCCCTTGCCTCCCACACAGCAGTCCAG
CGGCAGCGCCGTGGGAACGTCGTCCAAAGCAGAAGCCAGCGTCGCCTGCTCCCTC
TGCTGAGGATCCCG

Appendix 5.5 Plasmid Information of pEGFP-N1-(CCE-9-hIP-EGFP)

CAAGCTTGCCACCATGGAATGTGCTGCCTTAGAGAAGGAAATTGCAGCGTTAGA
GAAGGAAATTGCTGCATTAGAGAAGATGTACCCATACGATGTTCCAGATTACGC
TTACCCATACGATGTTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGCT
GATGCGGATTCGTGCAGGAACCTCACCTACGTGCGGGGCTCGGTGGGGCCGGCCA

```
CCAGCACCCTGATGTTCGTGGCCGGTGTGGTGGGCAACGGGCTGGCCCTGGGCAT
CCTGAGCGCACGGCGACCGGCGCGCCCCTCGGCCTTCGCGGTGCTGGTGACCGGA
CTGGCGGCCACCGACCTGCTGGGCACCAGCTTCCTGAGCCCGGCCGTGTTCGTGG
CCTATGCGCGCAACAGCTCCCTGCTGGGCCTGGCCCGAGGCGGCCCCGCCCTGTG
CGATGCCTTCGCCTTCGCCATGACCTTCTTCGGCCTGGCGTCCATGCTCATCCTC
TTTGCCATGGCCGTGGAGCGCTGCCTGGCGCTGAGCCACCCCTACCTCTACGCGC
AGCTGGACGGGCCCCGCTGCGCCCGCCTGGCGCTGCCAGCCATCTACGCCTTCTG
CGTCCTCTTCTGCGCGCTGCCCCTGCTGGGCCTGGGCCAACACCAGCAGTACTGC
CCCGGCAGCTGGTGCTTCCTCCGCATGCGCTGGGCCCAGCCGGGCGGCGCCGCCT
TCTCGCTGGCCTACGCCGGCCTGGTGGCCCTGCTGGTGGCTGCCATCTTCCTCTG
CAACGGCTCGGTCACCCTCAGCCTCTGCCGCATGTACCGCCAGCAGAAGCGCCAC
CAGGGCTCTCTGGGTCCACGGCCGCGCACCGGAGAGGACGAGGTGGACCACCTGA
TCCTGCTGGCCCTCATGACAGTGGTCATGGCCGTGTGCTCCCTGCCTCTCACGAT
CCGCTGCTTCACCCAGGCTGTCGCCCCTGACAGCAGCAGTGAGATGGGGGACCTC
CTTGCCTTCCGCTTCTACGCCTTCAACCCCATCCTGGACCCCTGGGTCTTCATCC
TTTTCCGCAAGGCTGTCTTCCAGCGACTCAAGCTCTGGGTCTGCTGCCTGTGCCT
CGGGCCTGCCCACGGAGACTCGCAGACACCCCTTTCCCAGCTCGCCTCAGGGAGG
AGGGACCCAAGGGCCCCCTCTGCTCCTGTGGGAAAGGAGGGGAGCTGCGTGCCTT
TGTCGGCTTGGGGCGAGGGGCAGGTGGAGCCCTTGCCTCCCACACAGCAGTCCAG
CGGCAGCGCCGTGGGAACGTCGTCCAAAGCAGAAGCCAGCGTCGCCTGCTCCCTC
TGCCgGGATCCC
```

References

1. Thomas CL, Maule AJ (2000) Limitations on the use of fused green fluorescent protein to investigate structure-function relationships for the cauliflower mosaic virus movement protein. J Gen Virol 81:1851–1855
2. Zupan AL, Trobec S, Gaberc-Porekar V et al (2004) High expression of green fluorescent protein in Pichia pastoris leads to formation of fluorescent particles. J Biotechnol 109:115–122
3. Chen I, Ting AY (2005) Site-specific labeling of proteins with small molecules in live cells. Curr Opin Biotechnol 16:35–40
4. Hermanson G (2008) Bioconjugate techniques, 2nd edn. Academic Press, London
5. Pace NJ, Weerapana E (2013) Diverse functional roles of reactive cysteines. ACS Chem Biol 8:283–296
6. Joshi NS, Whitaker LR, Francis MB et al (2004) A three-component Mannich-type reaction for selective tyrosine bioconjugation. J Am Chem Soc 126:15942–15943
7. Tilley SD, Francis MB (2006) Tyrosine-selective protein alkylation using π-allylpalladium complexes. J Am Chem Soc 128:1080–1081
8. Antos JM, McFarland JM, Iavarone AT et al (2009) Chemoselective tryptophan labeling with rhodium carbenoids at mild pH. J Am Chem Soc 131:6301–6308
9. Antos JM, Francis MB (2004) Selective tryptophan modification with rhodium carbenoids in aqueous solution. J Am Chem Soc 126:10256–10257
10. Griffin BA, Adams SR, Tsien RY et al (1998) Specific covalent labeling of recombinant protein molecules inside live cells. Science 281:269–272

11. Adams SR, Campbell RE, Gross LA et al (2002) New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications. J Am Chem Soc 124:6063–6076
12. Yi L, Sun H, Wu Y-W et al (2010) A highly efficient strategy for modification of proteins at the C terminus. Angew Chem Int Ed 49:9417–9421
13. Yi L, Sun H, Itzen A et al (2011) One-pot dual-labeling of a protein by two chemoselective reactions. Angew Chem Int Ed 50:8287–8290
14. Stephanopoulos N, Francis MB (2011) Choosing an effective protein bioconjugation strategy. Nat Chem Biol 7:876–884
15. Gilmore JM, Scheck RA, Esser-Kahn AP et al (2006) N-terminal protein modification through a biomimetic transamination reaction. Angew Chem Int Ed 45:5307–5311
16. Scheck RA, Dedeo MT, Iavarone AT et al (2008) Optimization of a biomimetic transamination reaction. J Am Chem Soc 130:11762–11770
17. Cornish VW, Hahn KM, Schultz PG et al (1996) Site-specific protein modification using a ketone handle. J Am Chem Soc 118:8150–8151
18. Muir TW, Sondhi D, Cole PA et al (1998) Expressed protein ligation: a general method for protein engineering. Proc Natl Acad Sci USA 95:6705–6710
19. Kiich KL, Saxon E, Tirrell DA et al (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. Proc Natl Acad Sci USA 99:19–24
20. Hao Z, Hong S, Chen X et al (2011) Introducing bioorthogonal functionalities into proteins in living cells. Acc Chem Res 44:742–751
21. Cravatt BF, Sorensen EJ (2000) Chemical strategies for the global analysis of protein function. Curr Opin Chem Biol 4:663–668
22. Morell M, Nguyen DT, Willis AL et al (2013) Coupling protein engineering with probe design to inhibit and image matrix metalloproteinases with controlled specificity. J Am Chem Soc 135:9139–9148
23. Hwang Y, Thompson PR, Wang L et al (2007) A selective chemical probe for coenzyme a-requiring enzymes. Angew Chem Int Ed 46:7621–7624
24. O'Hare HM, Johnsson K, Gautier A et al (2007) Chemical probes shed light on protein function. Curr Opin Struct Biol 17:488–494
25. Gronemeyer T, Godin G, Johnsson K et al (2005) Adding value to fusion proteins through covalent labeling. Curr Opin Biotechnol 16:453–458
26. Los GV, Encell LP, McDougall MG et al (2008) Halotag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 3:373–382
27. Mizukami S, Watanabe S, Hori Y et al (2009) Covalent protein labeling based on noncatalytic β-lactamase and a designed FRET substrate. J Am Chem Soc 131:5016–5017
28. Chen Z, Jing C, Gallagher SS et al (2012) Second-generation covalent TMP-tag for live cell imaging. J Am Chem Soc 134:13692–13699
29. Gallagher SS, Sable JE, Sheetz MP et al (2009) An in vivo covalent TMP-tag based on proximity-induced reactivity. ACS Chem Biol 4:547–556
30. Chen I, Howarth M, Lin W et al (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. Nat Methods 2:99–104
31. Uttamapinant C, White KA, Baruah H et al (2010) A fluorophore ligase for site-specific protein labeling inside living cells. Proc Natl Acad Sci USA 107:10914–10919
32. Lin C-W, Ting AY (2006) Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. J Am Chem Soc 128:4542–4543
33. Antos JM, Chew G-L, Guimaraes CP et al (2009) Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. J Am Chem Soc 131:10800–10801
34. Carrico IS, Carlson BL, Bertozzy CR et al (2007) Introducing genetically encoded aldehydes into proteins. Nat Chem Biol 3:321–322

35. Zhou Z, Cironi P, Lin AJ et al (2007) Enetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases. ACS Chem Biol 2:337–346

36. Chen G, Heim A, Riether D et al (2003) Reactivity of functional groups on the protein surface: development of epoxide probes for protein labeling. J Am Chem Soc 125:8130–8133

37. Hobert EM, Schepartz A (2012) Rewriting kinase specificity with a synthetic adaptor protein. J Am Chem Soc 134:3976–3978

38. Corey DR, Schultz PG (1987) Generation of a hybrid sequence-specific single-stranded deoxyribonuclease. Science 238:1401–1403

39. Gartner ZJ, Liu DR (2001) The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. J Am Chem Soc 123:6961–6963

40. Lewis CA, Miller SJ (2006) Site-selective derivatization and remodeling of Erythromycin A by using simple peptide-based chiral catalysts. Angew Chem Int Ed 45:5616–5619

41. Xiao J, Broz P, Puri AW et al (2013) A coupled protein and probe engineering approach for selective inhibition and activity-based probe labeling of the caspases. J Am Chem Soc 135:9130–9138

42. Nonaka H, Tsukiji S, Ojida A et al (2007) Non-enzymatic covalent protein labeling using a reactive tag. J Am Chem Soc 129:15777–15779

43. Nonaka H, Fujishima S, Uchinomiya S et al (2010) Selective covalent labeling of tag-fused GPCR proteins on live cell surface with a synthetic probe for their functional analysis. J Am Chem Soc 132:9301–9309

44. Wang J, Yu Y, Xia J et al (2014) Short peptide tag for covalent protein labeling based on coiled coils. Bioconjugate Chem 25:178–187

45. Wang J (2003) Cell biology. Science Press, Beijing

46. Johnson JE, Cornell RB (1999) Amphitropic proteins: regulation by reversible membrane interactions (Review). Mol Membr Biol 16:217–235

47. Overington JP, AI-Lazikani B, Hopkins AL et al (2006) How many drug targets are there? Nat Rev Drug Discov 5:993–996

48. Ferguson SG (2001) Evolving concepts in G protein-coupled receptor endocytosis: the role in receptor desensitization and signaling. Pharmacol Rev 53:1–24

49. Herbst RS (2004) Review of epidermal growth factor receptor biology. Int J Radiat Oncol Biol Phys 59:21–26

50. Wise H (2006) Lack of interaction between prostaglandin E_2 receptor subtypes in regulating adenylyl cyclase activity in cultured rat dorsal root ganglion cells. Eur J Pharmacol 535:69–77

51. Ng KY, Wong YH, Wise H et al (2011) Glial cells isolated from dorsal root ganglia express prostaglandin E_2 (EP_4) and prostacyclin (IP) receptors. Eur J Pharmacol 661:42–48

52. Regoli D, Allogho SN, Rizzi A et al (1998) Bradykinin receptors and their antagonists. Eur J Pharmacol 348:1–10

53. Gurevich VV, Gurevich EV (2008) GPCR monomers and oligoers: it takes all kinds. Trends Neurosci 31:74–81

54. Nakase I, Okumura S, Tanaka G et al (2012) Signal transduction using an artificial receptor system that undergoes dimerization upon addition of a bivalent leucine-zipper ligand. Angew Chem Int Ed 51:7464–7467

Chapter 6
Summary and Conclusion

6.1 Conclusion

During my graduate studies, I have been involved in three different but related areas: (1) structure–activity relationship of antimicrobial peptide defensin, (2) structure-guided design of protein ligands for quantum dots, and (3) structure-guided design of new bioconjugation reaction for cell imaging. Notwithstanding their distinct diversity, all three projects converge on structure-guided design of peptide and proteins and their engineering through chemical and biochemical methods. As the projects were all driven by chemical basis and molecular understanding and the techniques involve sophisticated chemical synthesis, my work, is broadly categorized as chemical biology.

In the project described in Chap. 2, we attempted to explore and improve the biological active site of white cloud bean defensin. While biologists mainly focused on the entire protein structure or the partial primary sequence of defensins, we adopted a complementary fragmentation strategy based on the secondary structures. The β loop structure of the protein was identified to be a potential active site. By incorporating $_DP_LP$ turn, we further increased the biological activity of the candidate.

In the project described in Chap. 3, we exerted a systematic investigation on the interaction modes of protein and nanoparticle. By comparing histag-mCherry, TIP1-mCherry monomer and TIP1-mCherry dimer, we manifested that protein size significantly affected the stoichiometry of protein–nanoparticle assembly. Engineering of two tetrameric proteins, ULD-mCherry and GCN-mCherry, we have shown that the spatial distribution of the QD-binding sequences affect the aggregation states of protein-QD assemblies. Lastly, we designed a new protein, Nanobelt-mCherry and exhibited a peptide induced structural transition and studied how such structural transition affects protein-QD interaction.

Chapters 4 and 5 report how we developed a new covalent tagging system which found application in protein labeling and cell imaging. Incorporation of an

electrophilic group in coiled-coil peptides converted the noncovalent interaction was converted highly reactive and yet specific pairs. One of the peptides is composed of natural amino acids, so it could be fused to the protein of interest and serve as the recognition site for the other synthetic peptide probe. This technique has been applied to the covalent labeling of cell surface receptor proteins, GPCRs and EGFR.

Further refinement of the antifungal peptide derived from defensin is in progress. One direction is incorporating more positively charged and aromatic amino acids, natural and unnatural, to increase the amphiphilicity [1, 2]. On the other hand, the secondary structure of the peptide (specially the β loop and the cyclic structure) could also be refined. For example, by collaboration with Prof. Zigang Li, the disulfide bond could be replaced with a more stable linkage using their unique chemistry thiol-ene reaction. A more comprehensive peptide library will then be generated. Finally, all the peptide candidates need to go through more rigorous biological activity tests to fully reveal the structure–activity relationship.

Having proven the principle of binding-induced covalent crosslinking in coiled coils, we will take one step further to develop orthogonal peptide tags to simultaneously labeling two receptors. One candidate to explore is the *Dock and Lock* system (DNL) [3], a pair of peptides with high specificity and strong binding affinity. We will fuse two orthogonal tags to two cell surface receptors followed and label and trace them simultaneously.

References

1. Kohli RM, Walsh CT, Burkart MD et al (2002) Biomimetic synthesis and optimization of cyclic peptide antibiotics. Nature 418:658–661
2. Srinivas N, Jetter P, Ueberbacher BJ et al (2010) Peptidominetic antibiotics target outer-membrane biogenesis in *Pseudomonas aeruginose*. Science 327:1010–1013
3. Chang CH, Rossi EA, Goldenberg DM et al (2007) The dock and lock method: a novel platform technology for building multivalent, multifunctional structures of defined composition with retained bioactivity. Clin Cancer Res 13:5586–5591