

Molecular engineering strategies for visualizing low-affinity protein complexes

Qianqian Ming , David Gonzalez-Perez and Vincent C Luca 

Department of Drug Discovery, Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA

Corresponding author: Vincent C Luca. Email: Vince.Luca@moffitt.org

Impact statement

Low-affinity protein interactions, while biologically essential, have been difficult to visualize by traditional methods in structural biology. In this review, we describe a series of innovative molecular engineering strategies that have been used to stabilize weakly bound protein complexes for structure determination. By highlighting several examples from the literature along with potential advantages and disadvantages of the individual approaches, we hope to provide an introductory resource for structural biologists studying low-affinity systems.

Abstract

The growing availability of complex structures in the Protein Data Bank has provided key insight into the molecular architecture of protein–protein interfaces. The remarkable diversity observed in protein binding modes is paralleled by a tremendous variation in binding affinities, with interaction half-lives ranging from days to milliseconds. Within the protein interactome, low-affinity binding events have been particularly difficult to visualize by traditional structural methods, which has spurred the development of innovative strategies for reconstituting these short-lived yet biologically essential assemblies. An important takeaway from structural studies of low-affinity systems is that there is no universal solution for stabilizing protein complexes, and approaches such as single-chain fusions, biochemical linkages, and affinity-maturation have each been successful in certain contexts. In this article, we review how advances in molecular engineering have been used to capture weakly

associated complexes for structure determination, and we provide perspectives on how the continued application of these methods can shed new light on the “hidden world” of low-affinity interactions.

Keywords: Structural biology, X-ray crystallography, protein–protein interactions, cryoelectron microscopy, engineering, nuclear magnetic resonance

Experimental Biology and Medicine 2019; 244: 1559–1567. DOI: 10.1177/1535370219855401

Introduction

A vast network of physical and functional protein interactions maintains the survival of all biological organisms. Among these interactions, certain protein binding events may be classified as “low-affinity”, which generally refers to complexes with dissociation constants (K_D) in the micromolar range ($>1\mu\text{M}$) or to complexes with fast kinetic off-rates (half-lives $<0.1\text{ s}$).^{1,2} Compared to tightly bound complexes, low-affinity protein associations are more sensitive to variations in their environment and rapidly respond to changes in temperature, pH, and solvent composition. The biochemical sensitivity of weak interactions is important for their function and enables them to fine-tune processes such as receptor signal transduction and stress adaptation mechanisms.^{3–5} Low-affinity interactions have also been well-characterized for their roles in controlling receptor-ligand promiscuity,⁶ immune discrimination between

“self” and “non-self”,⁷ enzyme turnover,⁸ and protein phase transitions.⁹

Despite the fundamental importance of low-affinity protein interactions in molecular biology, visualizing these complexes by X-ray crystallography, nuclear magnetic resonance (NMR), and cryoelectron microscopy (cryoEM) has been a major challenge. The main bottleneck in the pipeline for determining low-affinity structures is the reconstitution of stable complexes. In contrast to high-affinity complexes, which may remain stable for hours or even days, low-affinity complexes have fast off-rates and dissociate rapidly in solution. Specific technical difficulties are also associated with each of the structure determination methods mentioned above. For example, in X-ray crystallography, crystallization solutions often contain harsh chemicals (e.g. high salt concentrations, acidic pH) that further diminish binding affinity, or one component of a weakly bound

complex may dissociate and crystallize in the absence of its binding partner. In cryoEM studies, samples are diluted to low concentrations to optimize particle separation on cryoEM grids, which may lead to the dissociation of low-affinity complexes prior to immobilization.

A variety of molecular engineering strategies have been employed to ensure that low-affinity complexes remain stable and homogeneous during structural analysis. Established approaches for enforcing weak interactions include the generation of single-chain fusions and the use of chemical crosslinkers, although engineering advances such as site-specific crosslinking and directed evolution have greatly expanded the structural biologist's toolkit for capturing low-affinity complexes. In this review, we highlight published examples in which protein engineering has enabled structural studies of low-affinity interactions, with a focus on recent breakthroughs that have led to the determination of high impact structures. Additionally, we discuss the advantages and disadvantages of existing approaches along with emerging methods for protein stabilization.

Single-chain fusions

Single-chain fusions were among the earliest approaches used to stabilize low-affinity protein complexes for structure determination. These constructs connect two binding partners via a genetically encoded linker such that the proteins are expressed as a single polypeptide. Mechanistically, single-chain fusions favor complex formation by forcing the linked proteins into close proximity, which enhances binding affinity by maintaining a high local concentration. An ideal linker is long enough to allow proteins to adopt their native binding mode, yet short enough that they substantially enforce interactions, and reported lengths have ranged from 0 (direct fusion) to >30 amino acids.¹⁰ The majority of linkers used in structural biology incorporate glycine residues for their flexibility and serine residues to enhance solubility. This prevalent "Gly-Ser" linker design was inspired by the (GGGS)₃ linkers that were originally used to generate single-chain variable fragments (scFvs).^{11,12} The scFv linker facilitates heterodimerization of variable heavy (V_H) and variable light (V_L) domains, which bind weakly outside the context of the full-length antibody, so that the two domains adopt the proper conformation for antigen binding.¹³ Crystal structures of scFvs have confirmed that scFv V_H and V_L domains are oriented equivalently to those of antibodies¹⁴ (Figure 1(a)), thus providing early validation for the single-chain fusion strategy in structural studies of low-affinity interactions.

Although the development of scFvs was an important breakthrough in antibody engineering, it was unclear whether the single-chain fusion approach would be broadly applicable in structural biology. However, in 1999 Wlodawer *et al.*¹⁵ successfully used a single-chain construct in a non-scFv system to co-crystallize the low-affinity complex between the phage minor coat gene 3 protein (g3p) and the bacterial co-receptor TolA (Figure 1(b)).¹⁶ Following this report, single-chain fusions have facilitated

structure determination of several other weakly bound complexes, including crystal structures of T-cell receptors (TCRs) bound to peptide-loaded major histocompatibility complexes (pMHCs; Figure 1(c)),¹⁷ a cryo-EM structure of the *Escherichia coli* signal particle (SRP) bound to its receptor FtsY (Figure 1(d)),¹⁸ and the NMR structure of the YES-associated protein (YAP) of the Hippo signaling pathway bound to PPXY peptides from large tumor suppressor (LATS) kinase (Figure 1(e)).¹⁹ A comprehensive overview of linkers in structural biology is discussed in a review from Reddy Chichili *et al.*¹⁰

Advantages and disadvantages. Compared to other methods for stabilizing low-affinity interactions, an advantage of the single-chain design is its simplicity, as both proteins are expressed from a single plasmid and will be present in equimolar quantities during structural analysis. In some cases, single-chain fusions have also been shown to enhance solubility, folding, and expression yield.^{20,21} A disadvantage of single-chain fusions is that, without *a priori* knowledge of the orientation of subunits in the complex, it may be difficult to optimize linker position (e.g. N- vs. C-terminus) and length. Linkers may also impair bioactivity if they sterically occlude functionally important regions of a protein, such as active sites or ligand-binding interfaces.

Site-specific crosslinking

Chemical crosslinkers have been widely used to stabilize intra- or intermolecular interactions for structural studies. Common crosslinking agents such as glutaraldehyde or bis (sulfosuccinimidyl)suberate contain amine-reactive groups at either end of a spacer arm. When mixed with a protein complex of interest, the crosslinkers react with N-termini or solvent exposed lysine residues to form covalent bonds between spatially adjacent regions. The non-specific reactivity of chemical crosslinkers often yields unpredictable results and necessitates a large degree of empirical optimization, which has led structural biologists to develop "site-specific" crosslinking solutions. In the sections below, we discuss innovative crosslinking methodologies that have led to the determination of various low-affinity complex structures.

Disulfide trapping

Disulfide trapping is a method for covalently stabilizing transient or low-affinity protein complexes through the genetic incorporation of cysteine residues. In this approach, two cysteines are introduced at opposing positions within protein binding interfaces to induce the formation of intermolecular disulfide bonds following co-incubation in an oxidizing environment²² (Figure 2(a)). Disulfide trapping has been particularly effective in studies of extracellular proteins, as intracellular proteins often contain free cysteines that can form non-specific crosslinks with mutated residues.²³ Thus far, disulfide trapping has facilitated structural studies of numerous macromolecular interactions, including enzyme-substrate complexes^{24,25} and receptor-ligand complexes.^{22,26,27} In addition to assisting with structure determination, disulfide trapping has also been used in the absence

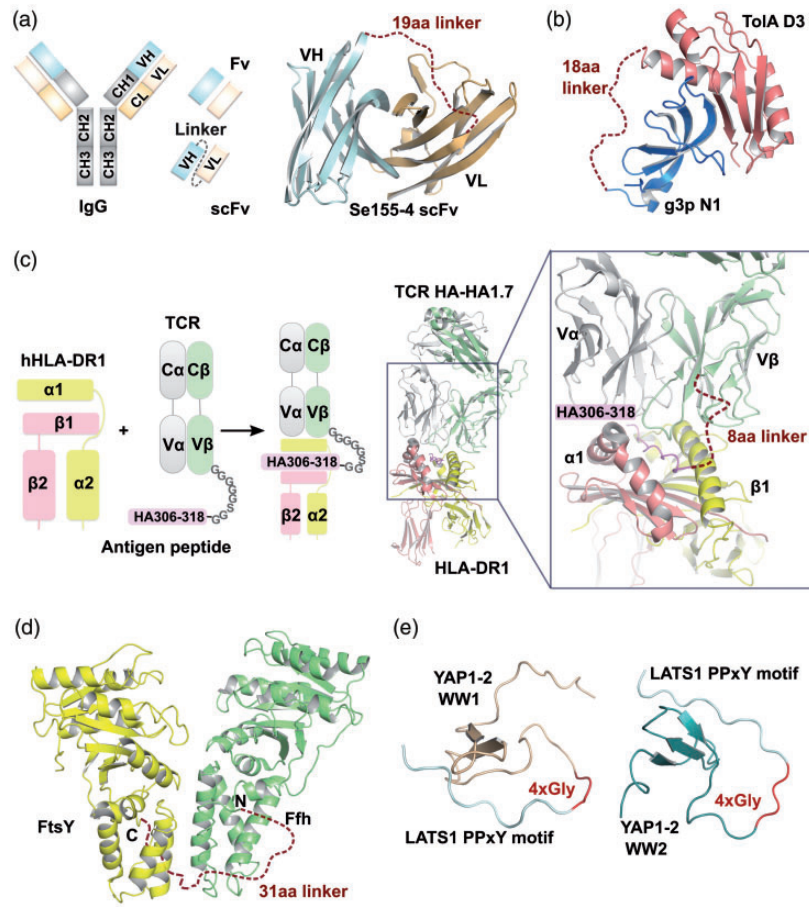


Figure 1. Structures of low-affinity protein complexes stabilized through single-chain fusion. (a) Schematic of an antibody and an scFv fragment (left) and crystal structure of the “Se155-4” scFv (right), PDB ID: 1MFA. CH1–3: heavy chain constant domains 1–3; CL: light chain constant domain; VH: heavy chain variable domain; VL: light chain variable domain. (b) Crystal structure of the g3p (blue)–TolA (red) complex linked by an 18 residue glycine-rich linker, PDB ID: 1TOL. (c) Schematic of a single-chain strategy used to stabilize an $\alpha\beta$ TCR/pMHCII complex. The antigen-derived peptide (HA306-318) was fused to the N-terminus of the TCR β -chain and binds in *trans* to the MHCII groove to enforce interactions (TCR C α and V α domains, gray; C β and V β domains, green; MHCII α 1 and α 2 domains, yellow; MHCII β 1 and β 2 domains, pink), PDB ID: 1FYT. (d) CryoEM structure of the *Escherichia coli* SRP protein Ffh (green) bound to its receptor FtsY (yellow), PDB ID: 2XKV. To stabilize the Ffh–FtsY complex, the FtsY C-terminus was fused to the Ffh N-terminus via a 31-residue Gly–Ser linker (red). (e) NMR structure of YAP2 WW domains (WW1, orange; WW2, teal) bound to the LATS1 PPxY motif (cyan), PDB IDs: 5YDX and 5YDY. The complex was stabilized using a 4xGly linker (depicted as a red line).

of structural information to map pairwise residue proximities based on the efficiency of crosslink formation.²⁶

Disulfide trapping has proven to be a powerful tool for understanding G-protein coupled receptor (GPCR) signaling^{22,26} and has enabled two landmark structural studies of GPCR–ligand interactions. In the first study, Rosenbaum *et al.* reported the crystal structure of an agonist bound to the β_2 adrenoceptor GPCR in the absence of a cognate G protein.²⁸ When interacting with agonists, GPCRs adopt either a low-affinity “G-protein free” conformation or a high-affinity “G-protein bound” conformation, and structural insight into both conformations is essential for understanding the complete mechanism of GPCR activation. To reconstitute the low-affinity complex, the authors introduced a cysteine residue in an extracellular loop of the β_2 adrenoceptor adjacent to the ligand-binding pocket (Figure 2(b)). The purified β_2 adrenoceptor was then coupled to an engineered β agonist containing a reactive ligand disulfide moiety (named FAUC50), which prevented FAUC50 from dissociating during co-crystallization.²⁸

In the second study, Qin *et al.* used disulfide trapping to determine the crystal structure of the C-X-C chemokine

receptor type 4 (CXCR4; a GPCR) bound to the viral chemokine viral macrophage inflammatory protein-II (vMIP-II).²⁷ Although vMIP-II binds tightly to cellular CXCR4, recombinant CXCR4 did not stably interact with vMIP-II when reconstituted in detergent. To solve this problem, cysteines were introduced at the predicted CXCR4:vMIP-II binding interface to generate a total of 37 possible cysteine pairs. Each mutant pair was then co-expressed and the pair that formed the highest percentage of covalent complex, CXCR4(D187C):vMIP-II(W5C), was purified and co-crystallized. The resulting structure (Figure 2(c)) revealed how the N-terminal chemokine “tail” extends into the GPCR “pocket” to fine-tune signaling and was a major breakthrough as it described one of the first visualizations of a GPCR bound to a protein ligand.^{27,29}

Advantages and disadvantages. One benefit of disulfide trapping is that the introduction of a single, solvent-exposed cysteine residue is unlikely to distort the overall protein fold. Disulfide trapping is also advantageous because disulfide bonds ($\sim 2\text{\AA}$) are much shorter than the

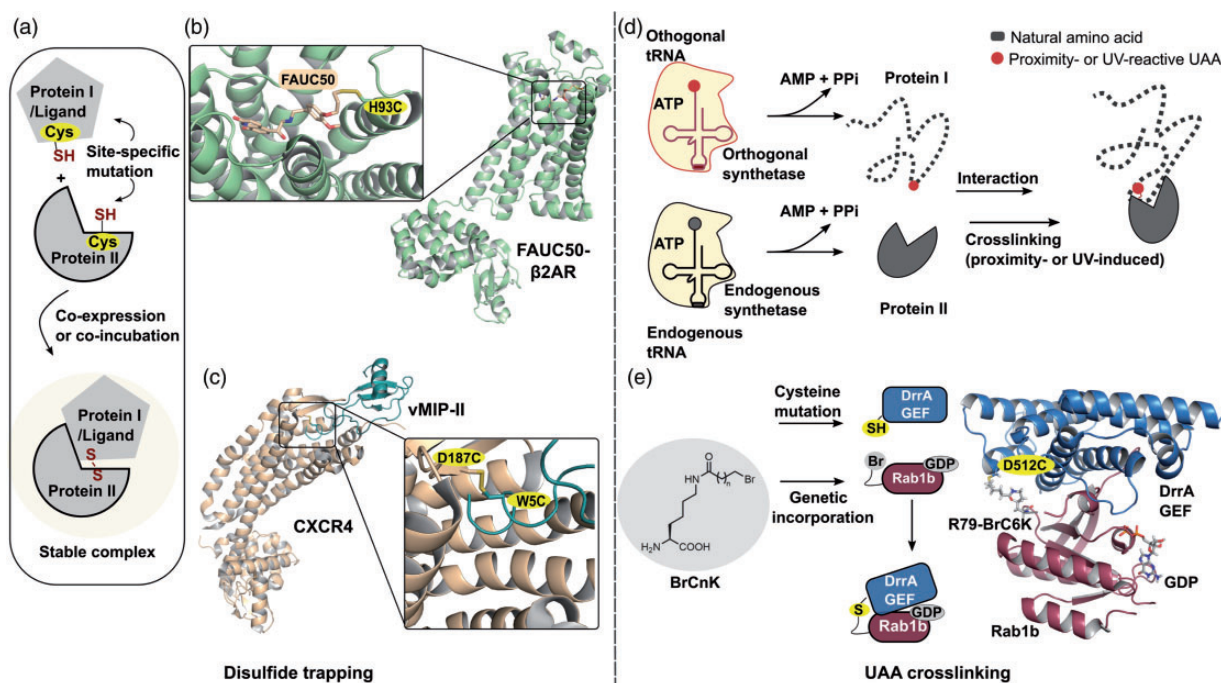


Figure 2. Disulfide trapping and site-specific crosslinking through unnatural amino acids. (a) Schematic describing the disulfide trapping method. Cysteine residues are introduced at opposing positions within the binding interface between two interacting proteins. Co-expression or co-incubation of the two binding partners then leads to the formation of disulfide bonds that covalently “trap” the complex. (b) Crystal structure of the β_2 adrenoceptor (green) bound to the cysteine-reactive β agonist molecule, FAUC50 (orange), PDB ID: 3PDS. (c) Crystal structure of the viral chemokine vMIP-II (dark green) bound to the chemokine receptor CXCR4 (wheat), PDB ID: 4RWS. The mutant cysteines, D187C (CXCR4) and W5C (vMIP-II), are highlighted in yellow. (d) Engineered orthogonal tRNA/aminoacyl-tRNA synthetase pairs allow for the genetic incorporation of UAAs. Proteins containing UAAs bearing reactive moieties may undergo site-specific, proximity- or photo-inducible crosslinking reactions. (e) Crystal structure of the transient, GDP-bound Rab1b-DrrA complex (Rab1b, dark red; DrrA, dark blue), PDB ID: 5O74. A native Arg79 residue of Rab1b was replaced by the proximity-reactive UAA R79-BrC6K.

spacers found in chemical crosslinkers, making them less likely to generate non-specific crosslinking artifacts. A disadvantage of disulfide trapping is that the approach is less effective on intracellular proteins containing free cysteines, as these unpaired residues may non-specifically bind to the mutant cysteines, and introduction of cysteines may cause protein misfolding due to improper disulfide bond formation. Furthermore, the method may be tedious if several mutants must be screened to identify an efficiently bonded pair.

Crosslinking via unnatural amino acids

The engineering of orthogonal tRNA/aminoacyl-tRNA synthase pairs³⁰ has allowed unnatural amino acids (UAAs) to be genetically incorporated within a protein sequence. At present, more than 70 different UAAs have been added to the genetic code of various strains of bacteria,³⁰ yeast,³¹ or mammalian cells.³² When introduced at protein interfaces, UAAs bearing proximity- or photo-reactive moieties such as aryl azide, diazirines, and benzophenones^{33,34} can be used as site-specific crosslinkers, and this technology has developed into a promising approach for determining low-affinity complex structures (Figure 2 (d)). In a proof-of-principle study designed to show that UAA-mediated crosslinking does not distort protein binding modes, Sato *et al.* determined the crystal structure of gankyrin crosslinked to its binding partner S6 proteasomal protein (S6C).³⁵ When compared to the native gankyrin-

S6C structure, there were no substantial structural differences in the “photobridged” complex despite the incorporation of 16 photoreactive *p*-benzoyl-L-phenylalanines (*p*Bpa) in the gankyrin protein.

Cigler *et al.* recently outlined a method by which UAA-mediated crosslinking may be used to determine structures of weakly bound complexes. In this study, proximity-reactive UAAs enabled the capture of a complex between the Ras-related GTP-binding (Rab1b) G-protein and the DrrA exchange factor from the bacterial pathogen *Legionella pneumophila*. When bound to GDP, Rab1b has a very low affinity for DrrA (K_D 10–30 μ M) and the complex cannot be easily purified without additional stabilization. To crosslink GDP-Rab1b and DrrA, the authors introduced UAAs bearing bromoalkyl moieties (BrCnK) within the DrrA-binding region of Rab1b, and a BrCnK-reactive cysteine residue (D512C) was introduced within the Rab1b-binding region of DrrA.³⁶ Co-expression of Rab1b-BrC6K and DrrA-D512C with an orthogonal RNA synthetase then yielded a stable complex that was subsequently co-crystallized (Figure 2(e)).

Advantages and disadvantages. A practical advantage of UAA-mediated crosslinking is that the reactions may be tightly controlled to minimize crosslinking artifacts. For example, photoreactive UAAs only crosslink proteins following UV irradiation, and UAAs such as BrCnK selectively react with cysteine residues. However, a significant disadvantage is the complexity and low-throughput

nature of UAA-based systems, as genes encoding for UAA-containing proteins must be customized to account for the expanded genetic code and expressed in engineered cell lines. The position and selection of UAAs must also be optimized to minimize steric clashes between the bulky UAA side chains and nearby interface residues.

Sortase ligation

Sortase A (hereafter referred to as sortase) is a transpeptidase from *Streptococcus aureus* that catalyzes the removal of

the C-terminal glycine from an LPXTG motif, followed by subsequent linkage to an N-terminal polyglycine repeat on second protein³⁷ (Figure 3(a)). The sortase reaction has been developed as a protein engineering tool to create C-to-N fusions between recombinant proteins, with LPXTG and polyglycine sequences functioning as C- and N-terminal tags, respectively. In biotechnology, sortase has been used for attaching non-native peptides to proteins, protein circularization, and cell surface labeling.³⁸ The ability to ligate proteins at a post-purification step is also a viable

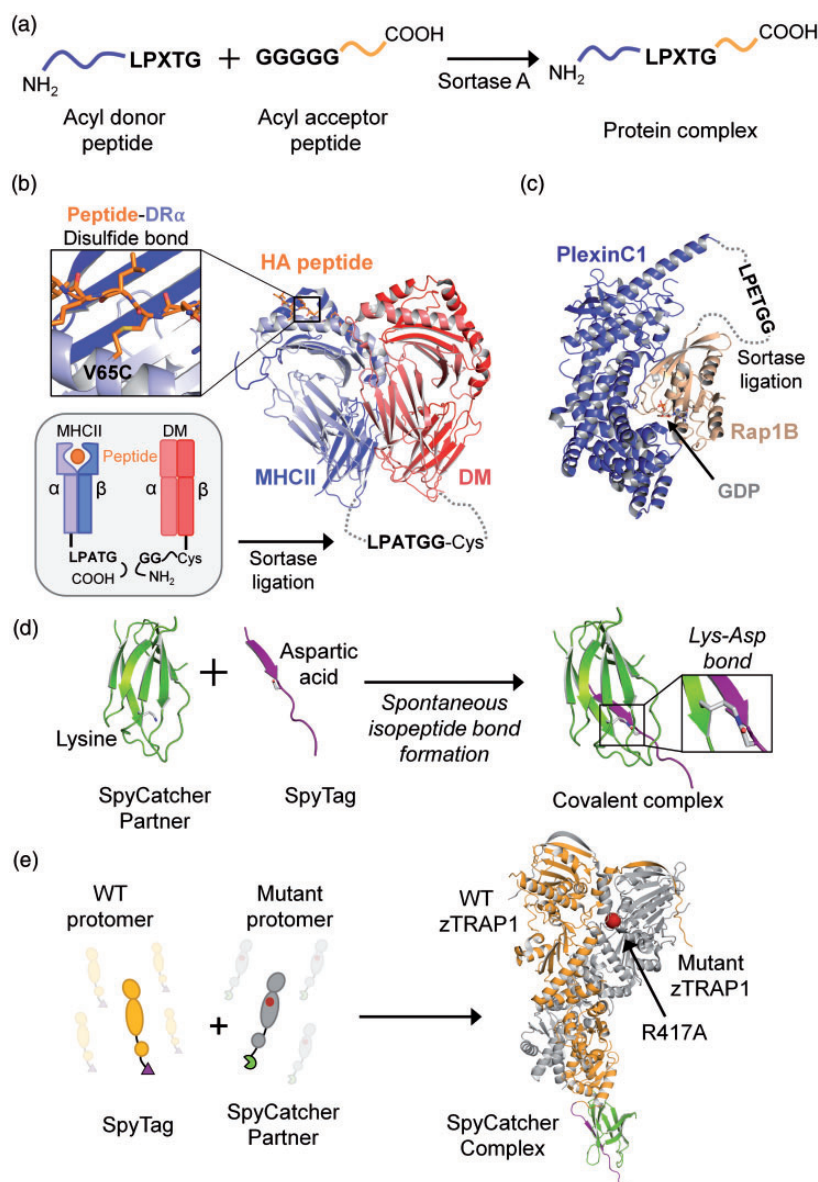


Figure 3. Sortase and SpyCatcher ligation for tethering protein complexes. (a) Schematic describing the sortase ligation reaction. The sortase transpeptidase enzyme covalently links peptides harboring a C-terminal LPXTG motif (where X is any amino acid) with N-terminal polyglycine sequences. (b) Crystal structure of the peptide: MHCII:DM “peptide selection” complex, PDB ID: 4GBX. A disulfide bond was introduced between the P6 position of a truncated HA peptide (GKQNC(L)KLAT) and residue 65 of the MHCII α -chain to prevent dissociation of the peptide (orange, stick representation) from the MHCII molecule (blue) and sortase ligation was used to enforce low-affinity interactions between the ECDs of MHCII and DM (red). To facilitate the sortase reaction, an LPATG sequence was fused to the C-terminus of the MHCII β -chain and a peptide with a polyglycine N-terminus was conjugated to the C-terminus of the DM β -chain. (c) Crystal structure of PlexinC1 (blue) bound to human Rap1B-GDP (brown), stabilized through sortase ligation. In this case, an LPTEG motif was attached to Rap-GDP C-terminus and a GG sequence was introduced at the N-terminus of PlexinC1, PDB ID: 4M8N. (d) Schematic describing the formation of an isopeptide bond between a Lys residue of SpyCatcher Partner (green) and an Asp residue of SpyTag (magenta), PDB ID: 4MLI. (e) The successful reconstitution of asymmetric WT:mutant zTRAP1 heterodimers was achieved using SpyCatcher. WT protomers (orange) with C-terminal SpyTags (magenta) were mixed with R417A mutant protomers (grey) with C-terminal SpyCatcher partners (green) to stabilize the WT:R417A zTRAP1 complex for co-crystallization PDB ID: 5TTH.

alternative to the single-chain constructs used in structural biology, especially in cases in which protein folding is impacted by the presence of the linker. Indeed, structural biologists have recently used sortase ligation to capture multiple intractable low-affinity protein complexes for co-crystallization.

In a breakthrough immunology study, Pos *et al.* employed both sortase ligation and disulfide trapping to capture a transient MHC class II (MHCII) “peptide selection complex” for co-crystallization.³⁹ MHCII molecules contribute to immune discrimination between “self” and “non-self” by presenting antigenic peptides to circulating T cells, and efficient peptide loading depends upon the concerted functions of two accessory proteins: class-II-associated invariant chain peptide (CLIP) and DM (HLA-DM in humans). CLIP fits into the MHCII peptide groove to prevent the binding of self-peptides during trafficking, and DM catalyzes the exchange of CLIP for foreign peptides in the endosome.^{40,41} Reconstitution of a stable peptide:MHCII:DM complex for structural studies is difficult for two reasons: (1) the binding of DM to MHCII destabilizes peptide binding and (2) the soluble extracellular domains (ECDs) of DM and MCHII bind weakly in the absence of their transmembrane domains. To solve the first problem, the authors used disulfide trapping to prevent a partial HA peptide from dissociating from the MHCII peptide-binding groove. To solve the second problem, the authors developed a sortase-based method to link the C-termini MHCII and DM, thereby mimicking their parallel orientation in the membrane. This atypical C-to-C linkage was generated by fusing a C-terminal MHCII sortase tag to the N-terminus of a polyglycine peptide that was conjugated to DM (Figure 3 (b)). Following covalent stabilization, the peptide:MHCII:DM complex was crystallized and the structure revealed how DM-induced rearrangements drive peptide release and stabilize the empty MHCII groove.

Sortase ligation also played a role in visualizing low-affinity interactions between the zebrafish neuronal signaling receptor PlexinC1 and the human intracellular signaling protein Rap1B (a Ras homolog). In this study, Wang *et al.* used the sortase system to fuse the C-terminus of Rap to the N-terminus of the plexin GTPase activating protein (GAP) domain.⁴² This fusion protein was validated in a GTP hydrolysis assay and exhibited improved catalytic activity relative to mixtures of the individual proteins. The sortase-ligated Rap-plexin complex was then co-crystallized to reveal how Rap binding improves GTP hydrolysis by promoting closure of the GAP domain active site (Figure 3(c)).

Advantages and disadvantages. The principal advantage of sortase ligation is that it provides a means of unidirectional coupling, as opposed to the bidirectional coupling of methods such as disulfide trapping and chemical crosslinking. Bidirectional crosslinking has a much higher probability of generating undesirable crosslinking artifacts, especially in systems where proteins may adopt homo- or heterodimeric conformations. A disadvantage of sortase ligation is the high temperature requirement for the

reaction (37°C), which may lead to the denaturation or precipitation of some protein samples. As with other site-specific crosslinking methods, sortase ligation also requires prior knowledge of the approximate location of the protein-protein binding interface.

SpyCatcher linkage

SpyCatcher is a unidirectional protein coupling system that covalently links protein tags called SpyCatcher partner (a 15-kDa protein fragment) and SpyTag (the peptide AHIVMVDAYKPTK; Figure 3(d)). When mixed together, these tags reconstitute a domain from the *Streptococcus pyogenes* protein FbaB that forms a spontaneous Lys-Asp isopeptide bond.^{43,44} Unlike sortase tags, which only generate N-to-C fusions, SpyTag is reactive at N-terminal, C-terminal, and internal sites within target proteins. In the first structural study to employ SpyCatcher, Elnatan *et al.* fused asymmetric heterodimers of the zebrafish Hsp90 (zTRAP1) chaperone to determine whether the zTRAP1 ATP-hydrolysis mechanism is cooperative, independent, or sequential.⁴⁵ Obtaining complexes that contain one “hydrolysis-dead” mutant (R417A) and one wild-type (WT) protomer was key to dissecting the roles of the individual subunits, yet R417A:WT heterodimers were difficult to separate from WT:WT or R417A homodimers during purification attempts. By attaching SpyCatcher partner and SpyTag to the C-termini of the mutant and WT protomers, respectively, the authors were able to successfully purify and co-crystallize the R417A:WT complex (Figure 3 (e)). The structure revealed that the ATP-sensing WT subunit adopted a “buckled” conformation, while the inactive R417A mutant adopted a “straight” conformation, which supported experimental observations indicating that ATP hydrolysis is sequential and deterministic. Although SpyCatcher has not yet been used to visualize low-affinity interactions, the above study demonstrates the potential of the system for future efforts in complex stabilization.

Advantages and disadvantages. A major strength of the SpyCatcher system is its versatility, as linkages may be introduced at either terminal or internal sites within proteins. Additionally, SpyCatcher reaction conditions are very mild (pH range of 5 to 8, temperature range of 4 to 37°C) when compared to those recommended for the comparable sortase method (37°C). A notable downside to the SpyCatcher system is that the linker contains an entire protein domain, which may make it less crystallizable than the small peptide or chemical linkers introduced through alternative methods.

Affinity-maturation by directed evolution

Directed evolution recently emerged as a powerful tool for determining structures of low-affinity complexes. In contrast to the methods highlighted above, which function through covalent tethering, directed evolution may be used to non-covalently enforce interactions through the introduction of affinity-enhancing mutations. This “affinity-maturation” process begins with the generation of a mutant library, typically containing thousands to

billions of variants, of the protein of interest. When no information about the binding interface is known, library mutations are randomly introduced using an error-prone polymerase chain reaction (PCR). Alternatively, when a binding surface has been identified, mutations may be introduced at specific hot-spot residues by using assembly PCR with degenerate primers. In either case, the library will then be expressed on a directed evolution platform such as bacteria,⁴⁶ yeast,⁴⁷ phage,⁴⁸ mRNA,⁴⁹ or ribosome displays⁵⁰ and subjected to several rounds of selection against the protein's binding partner (Figure 4(a)). Following the selections, high-affinity clones are sequenced, subcloned into expression vectors, and purified for structural analysis.

Affinity-maturation has played an instrumental role in visualizing receptor–ligand complexes from the Notch and interferon (IFN) pathways. In the Notch pathway, signaling is initiated when a transmembrane Delta-like or Jagged ligand engages a Notch receptor on an opposing cell. On the surface of cells, these low-affinity interactions are potentiated by catch bonds,⁵¹ clustering,⁵² and avidity-enhancement effects. However, Notch receptor and ligand ECDs bind with nearly undetectable affinities in solution, making the complexes difficult to capture for structural studies. To overcome this obstacle, Luca *et al.* used yeast surface display to evolve variants of Delta-like 4 (Dll4) and Jagged1 (Jag1) with enhanced affinity for the Notch1 receptor.^{51,53} The high-affinity Dll4 “SLP” (G28S, F107L, L206P)⁵³ and Jag1 “JV1” (S32L, R68G, D72N, T87R, Q182R)⁵¹ variants were selected from unbiased mutant libraries

generated through error-prone PCR. The DLL4^{SLP} and Jag1^{JV1} proteins were each co-crystallized with Notch1 to reveal ligand-specific binding modes with distinct “signatures” of Notch1 O-linked glycans contributing to the interfaces.^{51,53} Interestingly, multiple affinity-enhancing mutations were located in hinge-like regions outside of the receptor-binding interface, suggesting that they function by allosterically stabilizing the receptor-bound conformation⁵¹ (Figure 4(b)).

IFNs are immune cytokines that signal by multimerizing high- and low-affinity receptors on the cell surface. For both type II (IFN- γ) and type III (IFN- λ) IFNs, the rapid dissociation of low-affinity receptors had confounded attempts to determine structures of fully intact signaling complexes. Mendoza *et al.* solved this problem by using yeast display to evolve an IFN- λ variant, H11 (Q15R, E73D, H120R, T150A, V163E), with enhanced binding to the low-affinity receptor IL-10R β .⁵⁴ The stabilized IFN- λ :IL-10R β complex was then co-crystallized with IFN- λ R1, the high-affinity receptor, to reveal how IFN- λ binding coordinates interactions between receptor juxtamembrane domains to activate intracellular Jak kinases.⁵⁴ Recently, Mendoza *et al.* used a similar directed evolution-based approach to visualize interactions between IFN- γ and its receptors IFN- γ R1 and IFN- γ R2. In this case, the authors engineered a high-affinity IFN- γ R1 variant, F05 (T149I, M161K, Q167K, K174N, Q182R, H205N) to stabilize the complex for co-crystallization.⁵⁵ The hexameric IFN- γ :R1:IFN- γ :R2 structure revealed an IFN- γ binding mode nearly identical to that

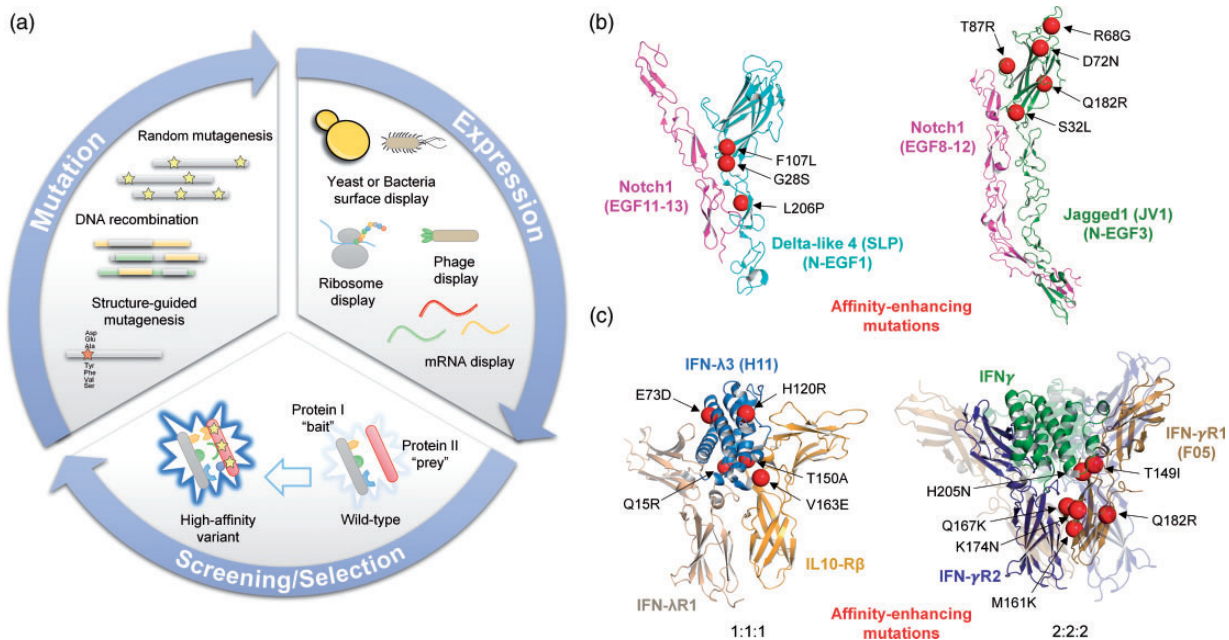


Figure 4. Affinity-maturation of protein–protein interactions by directed evolution. (a) Cartoon schematic of the affinity-maturation process. Briefly, a mutant library of a target protein (the “bait”) is generated and expressed on a directed evolution platform such as yeast display, bacteria display, phage display, mRNA display, or ribosome display. The mutant library is then selected against a known binding partner (the “prey”) to isolate high-affinity variants of the bait protein. (b) Crystal structures of notch receptor–ligand complexes stabilized through affinity-maturation. Left: structure of Notch1 EGF domains 11–13 (magenta) bound to the receptor-binding region of the DLL4 “SLP” variant (cyan), PDB ID: 4XL1. Right: structure of Notch1 EGF domains 8–12 (magenta) bound to the receptor-binding domains of the Jag1 “JV1” variant (green), PDB ID: 5UK5. Affinity-enhancing mutations are depicted as red spheres. (c) Crystal structures of IFN signaling complexes stabilized through affinity-maturation. Left: 1:1:1 complex of IFN- λ 3 (blue) bound to IL-10R β (yellow) and IFN- λ R1 (tan), PDB ID: 5T5W. Right: 2:2:2 complex of IFN- γ dimer (green) bound to two IFN- γ R1 receptors (brown) and two IFN- γ R2 receptors (blue), PDB ID: 6E3K. Affinity-enhancing mutations are depicted as red spheres.

of IFN- λ despite the different stoichiometries of the type II and type III IFN signaling complexes (2:2:2 vs. 1:1:1; ⁵⁵ Figure 4(c)).

Advantages and disadvantages. A key advantage of affinity-maturation method is that it does not require any prior knowledge of protein binding modes, as error-prone PCR allows for mutations to be interspersed throughout the entire coding sequence. Affinity-maturation also circumvents the use of covalent tethers that may interfere with protein function. As an added benefit, evolved proteins often have improved expression levels⁴⁸ and enhanced potency in functional assays,^{51–55} making them valuable tools for biomedical applications. A disadvantage to using directed evolution for protein stabilization is that selections may require expensive equipment and reagents (e.g. cell sorters, antibody-coated magnetic beads). Furthermore, selection strategies must be customized for individual proteins and protocols vary widely between the different directed evolution platforms.

Conclusions

The structural biology triumphs reviewed here showcase the outstanding potential of molecular engineering as a tool for visualizing low-affinity interactions. While no single strategy has emerged as a “one-size-fits-all” solution, an extensive repertoire of thoroughly vetted approaches is available for maximizing one’s chances of success. As cutting-edge methods such as cryoEM,⁵⁶ micro-electron diffraction,⁵⁷ and femtosecond X-ray nanocrystallography⁵⁸ become increasingly available for studying difficult targets, we anticipate that engineering complex stability will develop into an even more prevalent step in the sample preparation pipeline.

Author contributions: All authors participated in the composing and writing of the manuscript. QM, DGP, and VCL wrote the manuscript and generated the figures. VCL revised the manuscript.

ACKNOWLEDGMENTS

We thank Jennifer Binning for edits and helpful comments.



DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: We thank the Moffitt Cancer Center Department of Drug Discovery for funding. QM was supported by an R00 award from the NCI (NIH R00CA204738) and DG was supported by an award from the Rita Allen Foundation.

ORCID iDs

Qianqian Ming  <https://orcid.org/0000-0003-0296-7859>
Vincent C Luca  <https://orcid.org/0000-0001-9427-5520>

REFERENCES

- Nooren IMA, Thornton JM. Structural characterisation and functional significance of transient protein–protein interactions. *J Mol Biol* 2003;**325**:991–1018
- Nooren IMA, Thornton JM. Diversity of protein–protein interactions. *Embo J* 2003;**22**:3486–92
- Chien P, Gierasch LM. Challenges and dreams: physics of weak interactions essential to life. *Mol Biol Cell* 2014;**25**:3474–7
- Westermarck J, Ivaska J, Corthals GL. Identification of protein interactions involved in cellular signaling. *Mol Cell Proteomics MCP* 2013;**12**:1752–63
- Sukenik S, Ren P, Gruebele M. Weak protein–protein interactions in live cells are quantified by cell-volume modulation. *Proc Natl Acad Sci* 2017;**114**:6776–81
- Schreiber G, Walter MR. Cytokine receptor interactions as drug targets. *Curr Opin Chem Biol* 2010;**14**:511–9
- Garcia KC, Adams EJ. How the T cell receptor sees antigen – a structural view. *Cell* 2005;**122**:333–6
- Robinson PK. Enzymes: principles and biotechnological applications. *Essays Biochem* 2015;**59**:1–41
- Alberti S. Phase separation in biology. *Curr Biol* 2017;**27**:R1097–102
- Reddy Chichili VP, Kumar V, Sivaraman J. Linkers in the structural biology of protein–protein interactions. *Protein Sci Publ Protein Soc* 2013;**22**:153–67
- Bird RE, Hardman KD, Jacobson JW, Johnson S, Kaufman BM, Lee SM, Lee T, Pope SH, Riordan GS, Whitlow M. Single-chain antigen-binding proteins. *Science* 1988;**242**:423–6
- Huston JS, Levinson D, Mudgett-Hunter M, Tai MS, Novotný J, Margolies MN, Ridge RJ, Brucoleri RE, Haber E, Crea R. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc Natl Acad Sci U S A* 1988;**85**:5879–83
- Freund C, Ross A, Guth B, Plückthun A, Holak TA. Characterization of the linker peptide of the single-chain Fv fragment of an antibody by NMR spectroscopy. *FEBS Lett* 1993;**320**:97–100
- Zdanov A, Li Y, Bundle DR, Deng SJ, MacKenzie CR, Narang SA, Young NM, Cygler M. Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-Å resolution. *Proc Natl Acad Sci U S A* 1994;**91**:6423–7
- Lubkowski J, Hennecke F, Plückthun A, Wlodawer A. Filamentous phage infection: crystal structure of g3p in complex with its coreceptor, the C-terminal domain of TolA. *Structure* 1999;**7**:711–22
- Riechmann L, Holliger P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell* 1997;**90**:351–60
- Hennecke J, Carfi A, Wiley DC. Structure of a covalently stabilized complex of a human $\alpha\beta$ T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. *Embo J* 2000;**19**:5611–24
- Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C. Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat Struct Mol Biol* 2011;**18**:88–90
- Verma A, Jing-Song F, Finch-Edmondson ML, Velazquez-Campoy A, Balasegaran S, Sudol M, Sivaraman J. Biophysical studies and NMR structure of YAP2 WW domain – LATS1 PPxY motif complexes reveal the basis of their interaction. *Oncotarget* 2018;**9**:8068–80
- Kozono H, White J, Clements J, Marrack P, Kappler J. Production of soluble MHC class II proteins with covalently bound single peptides. *Nature* 1994;**369**:151–4
- Williams SJ, Ve T, Kobe B. A linker strategy for the production and crystallization of Toll/interleukin-1 receptor/resistance protein domain complexes. *Protein Eng Des Sel PEDS* 2015;**28**:137–45
- Buck E, Wells JA. Disulfide trapping to localize small-molecule agonists and antagonists for a G protein-coupled receptor. *Proc Natl Acad Sci U S A* 2005;**102**:2719–24

23. Hagemann IS, Miller DL, Klco JM, Nikiforovich GV, Baranski TJ. Structure of the complement factor 5a receptor-ligand complex studied by disulfide trapping and molecular modeling. *J Biol Chem* 2008;**283**:7763–75
24. Meenan NA, Sharma A, Fleishman SJ, MacDonald CJ, Morel B, Boetzel R, Moore GR, Baker D, Kleanthous C. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc Natl Acad Sci U S A* 2010;**107**:10080–5
25. Frouws TD, Barth PD, Richmond TJ. Site-specific disulfide crosslinked nucleosomes with enhanced stability. *J Mol Biol* 2018;**430**:45–57
26. Kufareva I, Gustavsson M, Holden LG, Qin L, Zheng Y, Handel TM. Disulfide trapping for modeling and structure determination of receptor: chemokine complexes. *Methods Enzymol* 2016;**570**:389–420
27. Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, Fenalti G, Wu H, Han GW, Cherezov V, Abagyan R. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* 2015;**347**:1117–22
28. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SG, Choi HJ, DeVree BT, Sunahara RK, Chae PS. Structure and function of an irreversible agonist- β 2 adrenoceptor complex. *Nature* 2011;**469**:236–40
29. Burg JS, Ingram JR, Venkatakrishnan AJ, Jude KM, Dukkupati A, Feinberg EN, Angelini A, Waghray D, Dror RO, Ploegh HL, Garcia KC. Structural basis for chemokine recognition and activation of a viral G protein-coupled receptor. *Science* 2015;**347**:1113–7
30. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. *Science* 2001;**292**:498–500
31. Deiters A, Cropp TA, Mukherji M, Chin JW, Anderson JC, Schultz PG. Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J Am Chem Soc* 2003;**125**:11782–3
32. Liu W, Brock A, Chen S, Chen S, Schultz PG. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* 2007;**4**:239–44
33. Brunner J. New photolabeling and crosslinking methods. *Annu Rev Biochem* 1993;**62**:483–514
34. Dorman G, Prestwich GD. Benzophenone photophores in biochemistry. *Biochemistry* 1994;**33**:5661–73
35. Sato S, Mimasu S, Sato A, Hino N, Sakamoto K, Umehara T, Yokoyama S. Crystallographic study of a site-specifically cross-linked protein complex with a genetically incorporated photoreactive amino acid. *Biochemistry* 2011;**50**:250–7
36. Cigler M, Müller TG, Horn-Ghetko D, von Wrisberg MK, Fottner M, Goody RS, Itzen A, Müller MP, Lang K. Proximity-triggered covalent stabilization of low-affinity protein complexes *in vitro* and *in vivo*. *Angew Chem Int Ed Engl* 2017;**56**:15737–41
37. Mazmanian SK, Liu G, Ton-That H, Schneewind O. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* 1999;**285**:760–3
38. Proft T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. *Biotechnol Lett* 2010;**32**:1–10
39. Pos W, Sethi DK, Call MJ, Schulze MS, Anders AK, Pyrdol J, Wucherpennig KW. Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection. *Cell* 2012;**151**:1557–68
40. Riberdy JM, Newcomb JR, Surman MJ, Barbosat JA, Cresswell P. HLA-DR molecules from an antigen-processing mutant cell line are associated with invariant chain peptides. *Nature* 1992;**360**:474–7
41. Sloan VS, Cameron P, Porter G, Gammon M, Amaya M, Mellins E, Zaller DM. Mediation by HLA-DM of dissociation of peptides from HLA-DR. *Nature* 1995;**375**:802–6
42. Wang Y, Pascoe HG, Brautigam CA, He H, Zhang X. Structural basis for activation and non-canonical catalysis of the Rap GTPase activating protein domain of plexin. *eLife* 2013;**2**:e01279
43. Kang HJ, Baker EN. Intramolecular isopeptide bonds: protein cross-links built for stress?. *Trends Biochem Sci* 2011;**36**:229–37
44. Zakeri B, Fierer JO, Celik E, Chittock EC, Schwarz-Linek U, Moy VT, Howarth M. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc Natl Acad Sci* 2012;**109**:E690–7
45. Elnatan D, Betegon M, Liu Y, Ramelot T, Kennedy MA, Agard DA. Symmetry broken and rebroken during the ATP hydrolysis cycle of the mitochondrial Hsp90 TRAP1. *eLife* 2017;**6**:e25235
46. Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss R 3rd. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol* 1997;**15**:29–34
47. Feldhaus MJ, Siegel RW, Opresko LK, Coleman JR, Feldhaus JM, Yeung YA, Cochran JR, Heinzelman P, Colby D, Swers J, Graff C. Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. *Nat Biotechnol* 2003;**21**:163–70
48. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 1985;**228**:1315–7
49. Seelig B. mRNA display for the selection and evolution of enzymes from *in vitro*-translated protein libraries. *Nat Protoc* 2011;**6**:540–52
50. Hanes J, Plückthun A. *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* 1997;**94**:4937–42
51. Luca VC, Kim BC, Ge C, Kakuda S, Wu D, Roein-Peikar M, Haltiwanger RS, Zhu C, Ha T, Garcia KC. Notch-jagged complex structure implicates a catch bond in tuning ligand sensitivity. *Science* 2017;**355**:1320–4
52. Bardot B, Mok LP, Thayer T, Ahimou F, Wesley C. The notch amino terminus regulates protein levels and delta-induced clustering of *Drosophila* notch receptors. *Exp Cell Res* 2005;**304**:202–23
53. Luca VC, Jude KM, Pierce NW, Nachury MV, Fischer S, Garcia KC. Structural biology. Structural basis for Notch1 engagement of delta-like 4. *Science* 2015;**347**:847–53
54. Mendoza JL, Schneider WM, Hoffmann HH, Vercauteren K, Jude KM, Xiong A, Moraga I, Horton TM, Glenn JS, de Jong YP, Rice CM. The IFN- λ -IFN- λ R1-IL-10R β complex reveals structural features underlying type III IFN functional plasticity. *Immunity* 2017;**46**:379–92
55. Mendoza JL, Escalante NK, Jude KM, Bellon JS, Su L, Horton TM, Tsutsumi N, Berardinelli SJ, Haltiwanger RS, Piehler J, Engleman EG. Structure of the IFN γ receptor complex guides design of biased agonists. *Nature* 2019;**567**:56–60
56. Bai X, McMullan G, Scheres S. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci* 2015;**40**:49–57
57. Rodriguez JA, Gonen T. High-resolution macromolecular structure determination by MicroED, a cryo-EM method. *Methods Enzymol* 2016;**579**:369–92
58. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, Hunter MS, Schulz J, DePonte DP, Weierstall U, Doak RB. Femtosecond X-ray protein nanocrystallography. *Nature* 2011;**470**:73–7