ACS chemical biology

Reviews

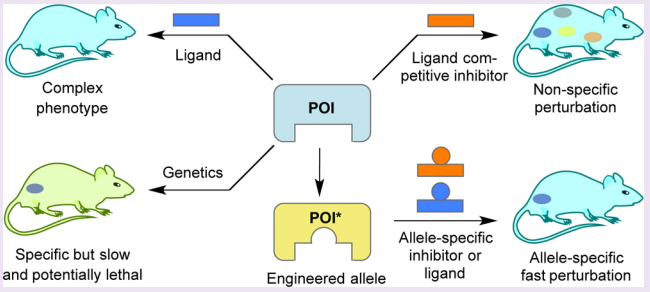
pubs.acs.org/acschemicalbiology

Allele-Specific Chemical Genetics: Concept, Strategies, and Applications

Kabirul Islam*

Department of Chemistry, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, United States

ABSTRACT: The relationship between DNA and protein sequences is well understood, yet because the members of a protein family/subfamily often carry out the same biochemical reaction, elucidating their individual role in cellular processes presents a challenge. Forward and reverse genetics have traditionally been employed to understand protein functions with considerable success. A fundamentally different approach that has gained widespread application is the use of small organic molecules, known as chemical genetics. However, the slow time-scale of genetics and inherent lack of specificity of small molecules used in chemical genetics have limited the applicability of these methods in deconvoluting the role of individual proteins involved in fast, dynamic biological events. Combining the advantages of both the techniques, the specificity achieved with genetics along with the reversibility and tunability of chemical genetics, has led to the development of a powerful approach to uncover protein functions in complex biological processes. This technique is known as allele-specific chemical genetics and is rapidly becoming an essential toolkit to shed light on proteins and their mechanism of action. The current review attempts to provide a comprehensive description of this approach by discussing the underlying principles, strategies, and successful case studies. Potential future implications of this technology in expanding the frontiers of modern biology are discussed.



Increasing diversity in structure and function of DNA, RNA, and proteins is central to the development and evolution of multicellular organisms. With only ~30 000 protein-coding genes, just 5-fold more than those present in *S. cerevisae*, the ability of humans to perform complex biochemical and cognitive processes comes from the bewildering complexity and enormous diversity of RNA and proteins coded by those genes.[1] RNA splicing and protein posttranslational modifications are two vital mechanisms that greatly increase the structural and catalytic complexity of biochemical networks.[2,3] This diversity-based mechanism, in which many products originate from relatively few genes, is central to the formation of closely related protein families. For example, more than 2000 ATP-dependent human proteins have emerged from a common ancestral protein that also served as precursor for other nucleotide triphosphate utilizing gene products such as GTPases.[4] This is true for other protein families as well. Approximately 1000 G-protein coupled receptors (GPCR) are expressed on cell surfaces to regulate proper cell signaling from myriad extracellular cues.[5] Although the emergence of such diversity from a common precursor is advantageous from an evolutionary viewpoint, the embedded similarity in ligand/cofactor binding pockets in receptors/enzymes and in protein−protein/DNA/RNA interacting domains presents a formidable challenge in efforts to deconvolute the role of a specific protein in biological processes.

Genetics has proven to be an essential tool to elucidate protein function, owing to its absolute specificity (single nucleotide change at a precise position in 30 billion base pairs in human DNA) and portability (can be applied to any organism). Forward and reverse genetics have been employed to elucidate gene/protein functions by creating random or specific mutations and correlating them to observed gain- or loss-of function phenotypes. At the molecular level, such mutations perturb active site or overall protein structure, resulting in misregulation of enzymatic functions or protein−protein interactions. In spite of the specificity and robustness of genetic manipulation, the lack of temporal control (hours for gene knockout and RNAi-base methods) makes genetic approaches problematic for the study of biological processes occurring on the millisecond to minute time scale (*e.g.*, posttranslational modification, signal transduction).[6] Because gene knockout and RNAi-based techniques result in the physical loss of the protein, it is challenging to decipher whether an observed phenotypic change results from a loss of enzymatic function or some other structural function of the protein (*e.g.*, involvement in a protein−protein binding interaction). The study of fast, dynamic biological processes demands the development of approaches that conditionally and selectively modulate a specific aspect (enzymatic, scaffolding, or allosteric) of a gene product in rapid and tunable manner.

Proteins and other biological macromolecules have evolved to bind small-molecule, such as coenzymes, neurotransmitters, prosthetic groups, hormones, and prostaglandins, for proper

Received: August 16, 2014
Accepted: December 1, 2014
Published: December 1, 2014

functions (Figure 1A). Such observation has led to a great deal of success in small-molecule mediated perturbation of protein function to probe underlying biology.[7−9] Advantages associated with small molecules, such as rapid action and reversibility, would seem to present a solution to the problems associated with genetic methods. However, small-molecule based approaches to explore biology, often referred to as chemical genetics,[10,11] are in reality far from ideal. A significant disadvantage is the lack of specificity originating from conserved active sites and folding topologies among members of a protein superfamily (Figure 1A). Therefore, unlike genetic methods that provide exquisite specificity in targeting a single gene product in an isoform-specific manner, chemical genetics falls short in identifying specific small-molecule inhibitors for every protein encoded in the human genome. An approach that could combine the specificity and portability of genetic methods with the rapid, tunable, conditional, and reversible nature of chemical genetics would constitute a powerful tool for elucidating protein functions in eukaryotic biology and disease. To address this challenge, allele-specific chemical genetics (ASCG) has been developed and successfully applied to a wide range of proteins (Figure 1B). This review provides a detailed account of the origin, development, and applications of this emerging technique in chemical biology.

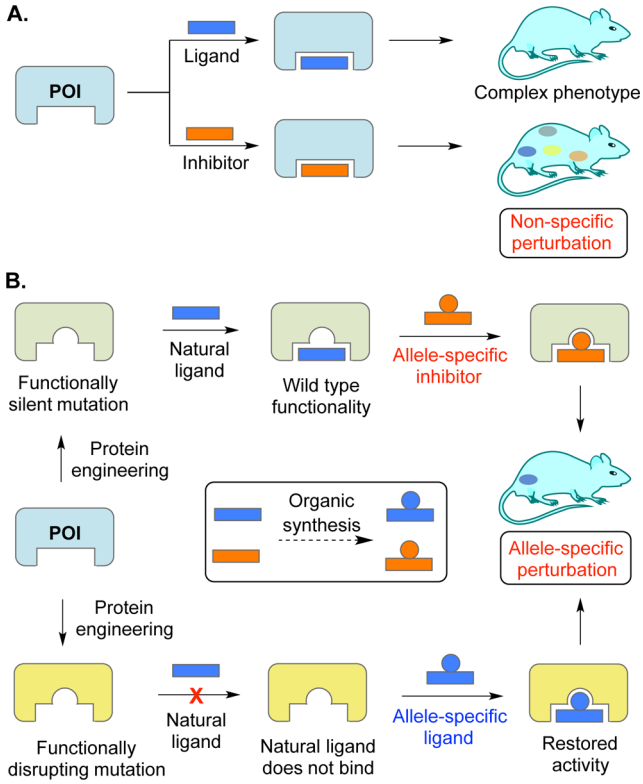


Figure 1. Chemical approaches to biology. (A) Functions of a protein are inherently depended on its binding to cofactors or ligands to regulate complex biological processes. A small-molecule based inhibitor (chemical genetics) of a protein of interest (POI) causes rapid inhibition of protein function resulting phenotypic changes. When ligand-competitive inhibitors nonspecifically engage a large number of proteins, it is difficult to elucidate the functions of a specific target. (B) In allele-specific chemical genetics, POI is engineered to be specifically sensitive to a designed inhibitor resulting in an isoform-specific phenotypic change. In some cases, the mutant allele can be selectively activated by cofactor or ligand analogues to produce orthogonal effects to the allele-specific activation. The inset shows that in most of the cases, the bumped cofactor/inhibitor can be synthesized directly from the corresponding unmodified counterparts.

■ ALLELE-SPECIFIC CHEMICAL GENETICS

Traditional approaches in pharmacology and chemical biology involve iterative optimization of small molecules to specifically interrogate protein targets, a practice facilitated by the ability of organic chemists to synthesize complex molecules. A conceptually novel approach, called allele-specific chemical genetics, has emerged from the work of multiple groups and involves simultaneous optimization of both a protein and its cognate ligand through interface engineering using recombinant DNA technology and organic synthesis.[12,13] A mutant allele is generated by site-directed mutagenesis that renders the mutant activated or inhibited by the complementarily modified ligands (Figure 1B). In some cases, the mutant allele may still be activated by its natural ligand; however, high mutual affinity between the mutant and cofactor/inhibitor analogues makes the engineered pair functionally orthogonal to the native protein−ligand pair. When the engineered allele fails to act on the natural ligand, it becomes in itself an orthogonal system that can be manipulated to explore its functions.

An early example of a successful allele-specific approach was reported in 1987 by Miller and colleagues. They demonstrated that designed mutation in the GTP-binding site of a GTPase switches the protein's cofactor selectivity to xanthine triphosphate (XTP), thus providing a unique tool to study the role of GTPases in the guise of XTPase.[14] This was followed in 1995 by the pioneering work of Schreiber and coworkers involving protein surface engineering for chemically inducible conditional association of desired proteins.[15] Since then, allele-specific chemical genetics has been routinely employed in numerous studies to specifically interrogate ligand−receptor, enzyme-cofactor/substrate, and protein−protein/DNA/RNA interactions. Several of these examples are discussed in the current review. Although this strategy has been consistently demonstrated its generality, successful implementation is contingent upon a few factors: (1) High-resolution structural information about a target protein−ligand complex must be available. This structural data is essential to enable the rational design of complementary ligand/receptor pairs and also sets ASCG apart from other protein engineering approaches that seek to improve protein function through random mutation (*e.g., in vitro* evolution). (2) The engineered protein must be structurally stable and functionally competent. (3) Biocompatibility of the mutant allele and bioavailability of the designed ligands are required for *in vivo* applications. Given that the typical mutant allele differs from the wild type by a single amino acid, lethality is uncommon. This overcomes a common problem associated with genetic knockout in addition to solving the specificity issue of chemical genetics.

■ DESIGN PRINCIPLES TO CONFER ALLELE SPECIFICITY

Multiple design strategies have been implemented to successfully create allele-specific systems for a large array of proteins and their ligands (Figures 2−5).[16] Common approaches include electrostatic, steric, and covalent complementation as well as unnatural amino acid mutagenesis. Innovative hybrid approaches have also been developed to generate allele-specific pairs. A general theme that underlies all

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

these methods is the attempt to confer specificity by introducing new interactions (noncovalent or covalent) between engineered protein−ligand pairs that are not present in the native system. Each of these techniques and their applications have been described below.

Charge Complementation. Molecular recognition between a ligand and its receptor is based on noncovalent forces, including hydrogen bonding and polar interactions, that are critical for conferring affinity and specificity. Miller's success in converting a GTPase to an XTPase was based on such electronic complementation (Figure 2A). D138 of GTPase

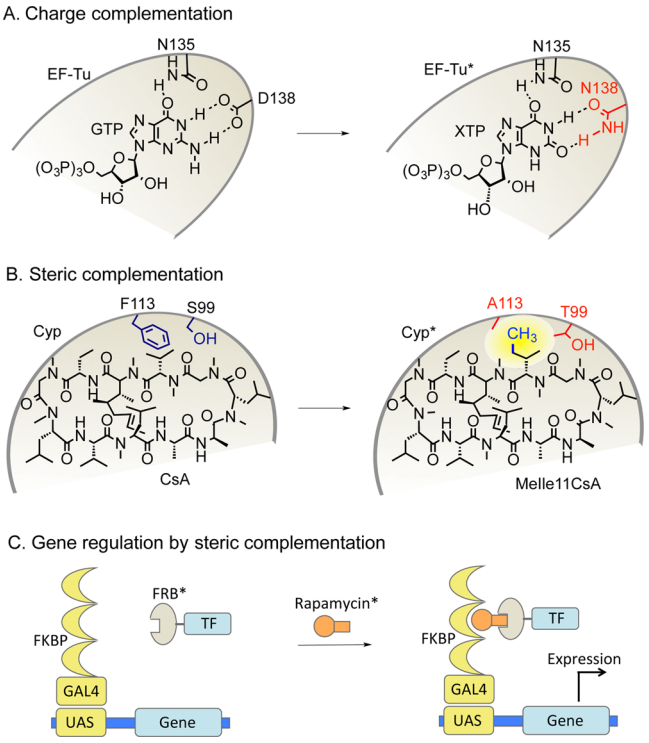


Figure 2. Approaches to generate ASCG pairs by charge and steric complementation. (A) Charge complementation by manipulations in hydrogen bonding interactions: Reversal of hydrogen bonding pattern through mutagenesis switches cofactor preferences of EF-Tu from GTP to XTP. (B) Steric complementation (bump−hole): S99T/F113A mutations cause expansion of the hydrophobic pocket in cyclophilin (Cyp) to accept a methyl "bump" in cyclosporine analogue Melle11CsA. Mutations create a "hole" in the active site represented by the yellow-colored oval box. (C) Inducible gene expression using "bumped" rapamycin (generated by replacing the C16-methoxyl group with (*R*)-isopropoxyl moiety) and "hole-modified" FRB (W2101F/T2098L/K2095P triple mutant). (Adapted with permission from ref 30. Copyright 1997, National Academy of Sciences.)

Elongation Factor Tu (EF-Tu) forms critical hydrogen bonds with the guanidinium moiety of GTP.[14] A D138N mutation in EF-Tu eliminates one hydrogen bond and at the same time introduces electronic repulsion between exocyclic amine of GTP and side-chain amide moiety of N138, resulting in almost 103-fold reduction in affinity for GTP (Figure 2A). However, the D138N mutant binds XTP with affinity comparable to that of EF-Tu toward GTP because of the newly established hydrogen bonding complementarity (Figure 2A).[14] An aspartic acid corresponding to D138 in EF-Tu is conserved in most of the GTP regulator proteins making Miller's approach highly generalizable across this protein family. This "nucleotide switch"-based allele specificity has been utilized to elucidate the biochemical role of EF-Tu in the elongation cycle of protein synthesis, where it led to the intriguing observation that two molecules of GTP are hydrolyzed for each peptide bond formed. The high cellular concentration of GTP and its unusually low dissociation constants (pM range) make GTPases challenging targets for inhibitor development. Miller's approach overcomes this challenge by targeting key intermolecular interactions, leading to development of the first allele-specific cofactor switch.

Introduction of new polar interactions has also been employed to change cofactor specificity. A Q102R mutation in lactate dehydrogenase (LDH) converts the enzyme to a malate dehydrogenase by introducing a salt-bridge between oxaloacetate and R102 with remarkable increase in catalytic efficiency.[17] In another study, Perham and colleagues demonstrated that introducing positively charged residues in the active site of pyruvate dehydrogenase could switch the cofactor specificity from NADH to NADPH.[18] Interestingly, elimination of such positively charged residues in glutathione reductase forces the enzyme to utilize NADH instead of its natural cofactor NADPH.[19]

These examples illustrate how manipulation of polar interactions can change cofactor specificity. However, because of the high intracellular concentration of these compounds, application of such methods is limited mostly to *in vitro* studies. Engineered mutant and cofactor analogues (non-native cofactors) have subsequently been developed for *in vivo* application. For example, abrogation of an intermolecular salt bridge between R269 of retinoic acid receptor β (RARβ) and a carboxylate group of its ligand retinoic acid (RA) has been used to generate an allele-specific system.[20] Charge neutralization by R269Q mutation caused a dramatic decrease in RA binding affinity but a substantial gain in affinity toward neutral retinol. Transactivation of chloramphenicol acetyltransferase (CAT) was achieved in CV-1 cells using the R269Q-retinol system to confirm the allele-specific activity of the mutant-cofactor pair. Because of the presence of many endogenous retinol-binding proteins with K_d in nM range, R269Q-retinol is still not orthogonal to the cellular environment.

Charge manipulation has been successfully employed to develop allele-specific pairs of GPCRs and their ligands. More than 1000 GPCRs are expressed differentially on the cell surface and can be activated by a wide range of ligands (e.g., biogenic amines, hormones, and lipids) to regulate diverse cell signaling processes. The diversity associated with GPCRs, their ligands, and response mechanisms have made it difficult to study their biological functions. Multiple strategies to stimulate a particular GPCR with a designed molecule in a cell- and tissue-specific manner have been explored.[21] These include RASSLs (receptors activated solely by synthetic ligands), TRECs (therapeutic receptor−effector complexes), neoreceptors, and DREADDs (designer receptors exclusively activated by designer drugs).[22] All these techniques essentially involve engineering GPCRs and their ligands either in combination or alone to create allele-specific systems. The first attempt in this direction was from Strader *et al.*, who mutated a conserved aspartic acid D113 in β2-adrenergic receptor (β2-AR). A D113S mutation in β2-AR makes the receptor inactive to its natural biogenic amine ligands (epinephrine, norepinephrine) due to the loss of a critical salt-bridge interaction.[23] Subsequent screening led to the identification of a synthetic compound (L-185,870) that could selectively activate the engineered receptor;

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

a key newly introduced contact was a hydrogen bond between the catechol group of L-185,870 and S113. Such targeted mutagenesis has been employed to generate RASSL systems for multiple receptors (κ-opioid, MC4-melanocortin, H1-histamine, 5-HT2A, serotonin, 5-HT4 serotonin, and α 2-adrenergic) by engineering in new polar interactions between the receptor allele and its specific ligand.[21] The neoreceptor−neoligand approach was based on a similar strategy of creating allele-specific pairs by making complementary changes in ionic interactions and hydrogen bonds in the receptor−ligand interfaces.[24] For example, H272E mutation in A3 receptor makes it inactive toward adenosine but active toward an amine-modified ligand that reestablishes the lost polar interaction. Aromatic appendages were added to neoligands to improve binding to the neoreceptor through hydrophobic interactions.[24] RASSL was further developed into DREADDs using chimeric proteins and *in vivo* evolution.[25,26] These allele-specific approaches have been extensively utilized to uncover the role of GPCRs in diverse biological processes such as muscle remodeling, sensory transduction, and neurobehavioral responses.

Steric Complementation (Bump−Hole Approach). Engineering the steric component of a ligand−receptor interface has been quite successful in improving allele-specificity by introducing shape complementarity between the modified ligand and its engineered receptor (Figures 1B, 2B).[12,13] In this approach, the ligand is modified with a large substituent that can only be accommodated by extra space in the receptor pocket created through site-directed mutagenesis. The bulky modification present in the unnatural ligand leads to a severe steric clash with the native receptor, thus providing an allele-specific activator for the mutant receptor. This design strategy, dubbed the "bump−hole" approach was first developed by Schreiber and colleagues to engineer the interface between ccyclosporine (ligand) and cyclophilin (receptor).[15] The resulting engineered ligand/receptor pair was used to introduce orthogonality in chemical inducer of dimerization (CID)-mediated manipulation of protein−protein interactions. Steric complementarity is typically achieved through manipulating hydrophobic interactions between a ligand and a receptor. A common strategy to generate a bump-hole system begins with replacement of bulky hydrophobic residues in the active site (e.g., phenylalanine, tyrosine, valine, isoleucine, methionine) with smaller side chains (e.g., alanine, glycine). The resulting void space (hole) in the protein−ligand complex is complemented by modification of the ligand with large hydrophobic substituent (bump) in order to restore the interaction.

Controlled manipulation of protein−protein interactions (PPIs) poses a significant challenge mostly because of the large number of interactions possible for a given protein in the complex cellular milieu,[27] yet, exerting control over PPIs is important in the functional elucidation of the role of specific contacts in biological processes. Schreiber and co-workers elegantly addressed this problem by using small organic molecules such as cyclosporin A (CsA), FK506, and rapamycin that can simultaneously bind two domains called immunophilins [cyclophilin, FK506 binding protein 12 (FKBP12), and the FKBP12-rapamycin-binding protein (FRBP)].[10] Proteins of interest were fused separately with these domains and spatiotemporally localized together by adding the small-molecule dimerizer. Such chemical genetic methods have been successfully employed to manipulate a wide range of signaling events such as gene regulation, protein transportation, and cell surface receptor dimerization. The inherent limitation of this approach, however, is the nonspecific binding of ligands used for CID with endogenous immunophilins. In the first "bump-hole" study involving CID, a "bumped" CsA was synthesized by putting a methyl group in isoleucine at position 11 and shown to bind the "hole-modified" double mutant (S99T/F113A) cyclophilin tightly (K_d = ∼2 nM) but not the wild-type receptor (K_d > 3000 nM) (Figure 2B).[15] The orthogonal receptor−ligand pair was subsequently employed to inhibit calcineurin-mediated dephosphorylation of nuclear factor of activated T cells (NFAT) in cell or tissue-specific manner.[28] Multiple related orthogonal pairs have been developed subsequently to control diverse signaling pathways, such as induction of rapid apoptosis by modulating the Fas receptor and conditional gene expression using orthogonal rapamycin-FRB pairs (Figure 2C).[10,29,30]

Another shape complementarity-based design strategy termed "knobs-into-holes", originally proposed by Crick to explain packing of amino-acid side chains between adjacent helices in coiled coils,[31] has been employed to remodel a protein−protein interface. Although not necessarily chemical genetics by the strictest definition of the term because a small organic molecule is not used, this approach has been effectively employed by Carter and colleagues to engineer an antibody by introducing complementary mutations of tyrosine to threonine (hole) and threonine to tyrosine (knob) in two adjacent heavy chains for forced heterodimerization to develop bifunctional therapeutics.[32]

Elegant application of the bump−hole approach to protein kinases by Shokat and co-workers demonstrated the ability of this strategy to probe enzymes in an isoform-specific manner.[33] This seminal work has brought about an avalanche of studies in a wide range of enzymes and cofactors and helped to elucidate their roles in diverse biological processes. The success of the bump−hole approach in elucidating enzyme functions can be attributed to the presence of a well-defined hydrophobic binding site for cofactors as well as high-resolution structural information. It is worth noting that such shape-complementary-based methods do not always generate orthogonal system. Although a designed ligand may selectively bind to the "hole-modified" receptor owing to the presence of a "bump", the engineered receptor may still retain substantial affinity for its natural ligands (Figure 1B). This can complicate the elucidation of a particular protein's role in signaling pathways. Nevertheless, allele-specific inhibitors that selectively bind mutant receptors have emerged as a powerful tool for studying the biological functions of proteins, a topic that has been discussed extensively later in this review.

Combined Charge and Steric Complementation. Simultaneous manipulation of both electronic and steric components in the interface between a ligand and its receptor has also been applied to develop allele-specific pairs (Figure 3A). An early example of combining charge manipulation with the bump−hole technique involved the modifications of retinoic acid (RA) and its receptor RARγ for conditional regulation of gene expression.[34] The carboxyl group of RA forms strong polar interactions with R278 and S289 of RARγ (Figure 3A). A double mutant comprising R278E and S289G displayed sufficient binding affinity to the neutral and bulky ethyl amide analog of retinoic acid by providing space (S289G) and creating a new polar interaction between the amide group and carboxyl group of E278 (Figure 3A). In an *in vivo* luciferase

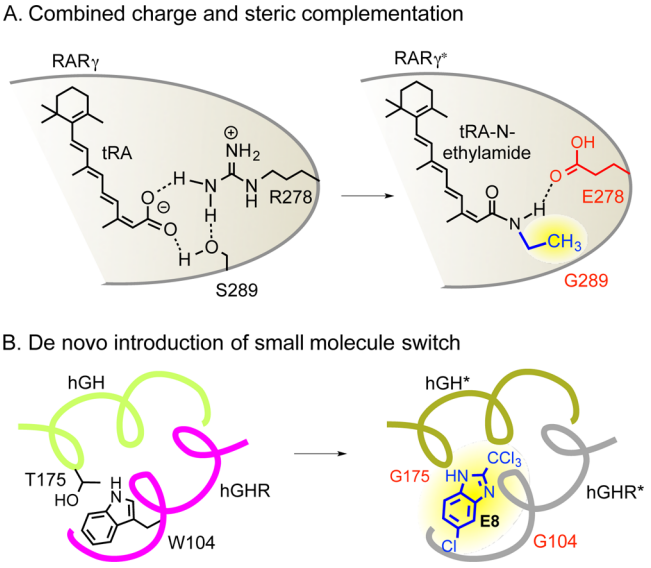


Figure 3. Approaches to generate ASCG pairs via combined complementarity. (A) Combined charge and steric complementation: E278/G289 mutations in RARγ display enhanced affinity to neutral analogue tRA-N-ethylamide of retinoic acid (tRA) by providing space (S289G) and new hydrogen bonding interactions (R278E). (B) *De novo* introduction of small molecule binding site: the void space generated through T175G and W104G mutations abolish the well-packed interface between human growth hormone (hGH) and its receptor (hGHR). This space can be filled with tryptophan side chain-like small molecule E8 to reestablish the protein−protein interaction.

reporter assay, it was demonstrated that the mutant allele could be activated selectively over the wild-type receptor.

Another successful example of combined charge and size manipulation to confer allele-specificity involves the modification of estradiol (E2) and human estrogen receptor (hER) interactions.[35,36] E353 of hER participates in polar interactions with hydroxyl group of E2 and R394 of hER. An E353A mutation lowered the binding affinity toward E2 by 400-fold. In one study led by Tedesco *et al.*, a hydrophobic steric bump in E2 activated the mutant by 34 times compared to the wild-type receptor.[35] In another study led by Koh and colleagues, an acrylate-functionalized E2 analog was shown to efficiently bind the E353A mutant with >95-fold preference over wild type hER.[36] Such improved selectivity is the direct result of both steric and electronic complementarity. The extra space necessary to accommodate an acrylate moiety is provided by a mutation (E353A), while nearby R394 participates in a salt bridge with the carboxylate in the ligand, compensating for the loss of a polar interaction with E2 due to E353A mutation. These allele-specific ligand−receptor pairs, generated via steric/electronic tuning individually or in combination, have played critical roles in precise temporal regulation of gene expression in various biological systems. The most widely used application probably is the development of conditional site-specific recombination by fusing mutant ligand-binding domain (LBD) of ER to the Cre recombinase. Although Cre-LoxP system has contributed greatly to the understanding of genetic control in mammalian development and physiology because of its ability to inactivate specific genes by homologous recombination, it lacks temporal control. Based on the earlier observation that G525R mutant of mouse ER LBD does not bind natural ligand E2 but recognizes synthetic ligands tamoxifen (T) and 4-hydroxytamoxifen (OHT), Chambon and co-workers developed a chimeric recombinase (CRE-ER^{T}) that can be specifically activated by T or OHT but insensitive to endogenous E2.[37] Triple mutation (G400V/M543A/L544A) in hydrophobic residues in hER resulted in more specific chimeric recombinase (CRE-ER^{T2}).[38] These systems have been widely employed for efficient spatiotemporally controlled gene knockout.

Corey and co-workers developed an orthogonal pair for retinoid X receptor (RXR) from a collection of compounds they termed "near drugs" that failed to generate effective pharmacological candidates for the wild-type receptors.[39] Based on the all-*trans* retinoic acid (atRA)- and 9-*cis* retinoic acid (9cRA)-bound structure of RAR, they generated a Q275C/I310M/F313I triple mutant of RXR that was efficiently stimulated by the synthetic ligand LG335 from the repertoire of their "near drugs". LG335 contains a bulky *n*-propyl group as opposed to the methyl in LG69, a ligand for wild-type RAR. The need for mutations at both polar (Q275C) and hydrophobic (I310M/F313I) residues suggest that orthogonality of the engineered RXR-LG335 pair is due to the combined engineering of polar and steric interactions with F313I mutation potentially facilitating the binding of bulky *n*-propyl group.

***De Novo* Introduction of Small Molecule Switch.** A novel approach for allele-specific modulation of protein−protein interactions was developed by the Schultz group, showing that a small molecule-binding pocket could be introduced into an otherwise well-packed protein−protein interface (Figure 3B).[40] Such modification has two consequences: the native PPI is abrogated due to the disrupted binding interface and the PPI can be restored by adding a designed small molecule with shape complementarity to the newly introduced pocket. In an early study, the structurally well-characterized interface between human growth hormone (hGH) and its receptor (hGHR)[41] was mutated (T175G in hGH and W104G in hGHR) to introduce a "hole" (Figure 3B).[40] Most of the lost affinity was regained upon filing the hole with E8, structurally similar to the side chain of tryptophan, identified from a rationally designed small molecule library.

The approach was later extended to control gene regulation by an engineered zinc finger (ZF) transcription factor, a class of protein that is present in more than 1000 genes in humans.[42] Based on the available structural information, Schultz and colleagues generated a structurally compromised double mutant F116A/H125G in the $\beta\beta\alpha$ fold of ZF protein C7, rendering it functionally inactive due to the formation of a hole around the active-site zinc ion. Through a luciferase-based reporter assay, a benzimidazole-based compound was identified that could fit into the pocket and reactivate the engineered transcription factor for gene regulation.[42]

Covalent Complementation. The process of ligand discovery and its subsequent optimization into a potent inhibitor for elucidation of protein function remains a major challenge. A recent strategy that can potentially accelerate the allele-specific ligand discovery process, introduces complementary reactivity in both protein and its otherwise nonspecific small molecule partner. In most examples of this approach, the unique nucelophilicity of the sulfhydryl group of cysteine has been employed. An early and successful example is a site-directed ligand discovery process (known as "tethering") pioneered by Wells and colleagues (Figure 4A).[43] In this approach, a cysteinylated mutant-allele (or a wild type protein carrying cysteine at suitable position) was used to capture and

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

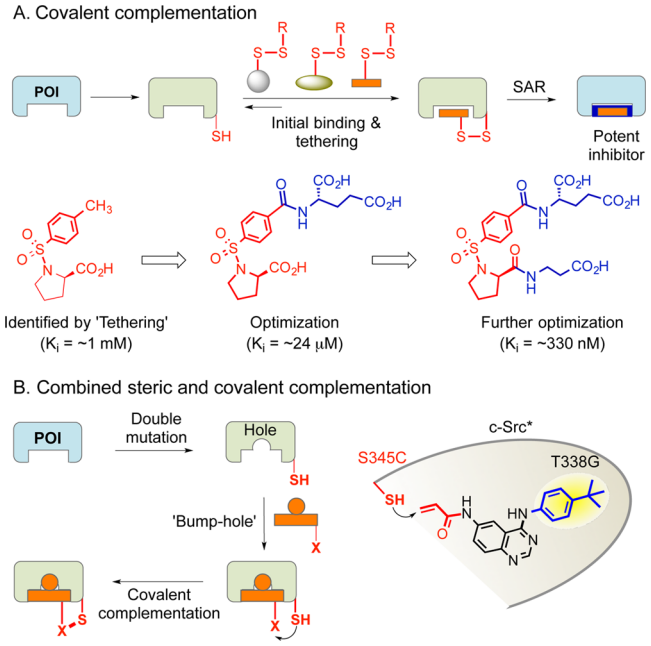


Figure 4. ASCG through complementary reactivity. (A) Covalent complementation by "tethering": a suitably introduced cysteine in protein of interest undergoes disulfide exchange with a weakly bound inhibitor to facilitate MS-based identification of the initial hit which can be subsequently optimized to a potent inhibitor via SAR study. A submicromolar inhibitor of thymidylate synthetase was developed from an initial hit of millimolar potency identified by "tethering". (B) Combined steric and covalent complementation: a systematically introduced double mutant (bearing a hole and cysteine in the active site) can be selectively inhibited by a craft inhibitor embellished with "bump" and a thiol-reactive functional group 'X'. Employing this approach, a quinazoline-based compound carrying acrylamide (cysteine reactive) and *tert*-butylphenyl (bump) groups was developed as a potent inhibitor of analog-sensitive T338G/S345C mutant of c-Src.

identify a weak inhibitor also bearing a thiol moiety through disulfide bond formation. The initial hit was then optimized via iterative structure−activity relationship (SAR) studies. As a representative example, an untethered ligand ($K_i \sim 1$ mM) of thymidylate synthase discovered through tethering was subsequently improved by 3000-fold.[43]

The Belshaw group adopted a similar 'proximity-induced alkylation' approach to develop a CsA-based inhibitor specific for cysteinylated cyclophillin.[44] In their study, an acrylamide-containing CsA analogue specifically inhibited the P105C mutant allele of cyclophillin but not the wild type or other mutants (A103C, G104C, and N106C), suggesting that proximity is essential for the development of covalent complementarity-based ASCG. This system was later employed for allele-specific labeling of fusion proteins *in vivo*.[45]

In an effort to develop a specific inhibitor for a protein kinase, Taunton and colleagues analyzed sequences of 491 kinases and found that a combination of a threonine at the "gatekeeper residue" and a cysteine in the glycine-rich loop near the active site of p90 ribosomal protein S6 kinase (RSK) act as selectivity filters.[46] By introducing a complementary reactive functionality such as α-halomethyl ketone to a nonspecific kinase inhibitor, the group synthesized an inhibitor selective for RSK1 and 2, which both display threonine and cysteine at the desired locations. Genetic incorporation of these selectivity filters into other kinases can sensitize the mutant alleles toward the reactive inhibitors. Subsequent studies demonstrated that the inhibition remained allele-specific toward the engineered protein kinases even in complex *in vivo* settings.[46]

Shokat *et al.* observed that, for a number of mutant kinases, shape complementarity leads to reduced kinase activity toward ATP due primarily to unoccupied space at the active site.[47] To overcome this issue, covalent complementarity was introduced. As an example, when conserved gatekeeper residue T338 in c-Src was replaced with cysteine, it was found that the mutant allele, although as active as wild type toward ATP, can be selectively inhibited by pyrazolopyrimidine-based compounds carrying a vinylsulfonamide electrophile.[47] When tested against a panel of 307 protein kinases, the inhibitor showed remarkably low off-target activity, making it one of the most selective chemical genetic inhibitors.

Recently, Bogyo *et al.* demonstrated the application of covalent complementary-based ASCG in studying matrix metalloproteinases (MMPs) and caspases by introducing noncatalytic cysteine at a suitable position to sensitize the mutant alleles to chloroacetamide- and acrylamide-containing electrophiles.[48,49] These inhibitors were further developed into activity-based probes for allele-specific imaging of only the engineered proteases in human cells in the presence of large numbers of wild type MMPs and caspases. The Cornish group developed a trimethoprim-based acrylamide-containing fluorophore that selectively underwent covalent addition to cysteinylated *E. coli* DHFR (L28C) suitable for cellular imaging and quantification.[50]

Combined Steric and Covalent Complementation. To further improve the efficacy and selectivity of allele-specific inhibitors in a kinome-wide manner, Shokat *et al.* introduced another approach where optimization of mutant-inhibitor pairs is guided through combined steric and covalent complementarity (Figure 4B).[51] In this strategy, a space-creating (hole) mutation was generated along with the introduction of cysteine (covalent) by replacing two selectivity filters on protein kinases so that the double mutant could be selectively sensitized by a designed inhibitor embellished with a properly positioned bulky substituent (bump) and an "electrophile" (covalent). A c-Src double mutant (T338G/S345C) was shown to be efficiently inhibited by quinazoline-based compounds bearing substituted anilines and acrylamide (Figure 4B). These compounds were much less effective toward either of the single mutants (T338G or S345C).[51] Such combined strategy was subsequently employed to examine the functions of EGFR in human cells as well as to study fission yeast Aurora kinase.[52]

Introduction of Engineered Domains. A conceptually novel approach to regulate the stability of a protein of interest in an allele-specific manner was developed by Wandless and colleagues.[53] In this ASCG approach, a genetically encoded tag provided the required specificity (genetics), and temporal control was achieved through a small molecule (chemical genetics) designed to confer stability to a fusion protein carrying a death domain (DD) (Figure 5A). The approach was based on a chance observation made by Crabtree and co-workers that "bumped" chemical inducers of dimerization (CIDs) in combination with "hole-modified" immunophilin domains could be employed for conditional degradation of fusion proteins.[54] However, this technique possessed limitations owing to the involvement of two dimerizing domains and synthetic difficulty in accessing CIDs. Wandless and co-workers addressed these issues by identifying a ligand-dependent single destabilizing domain in FRB* (space creating F36V mutation[55]

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

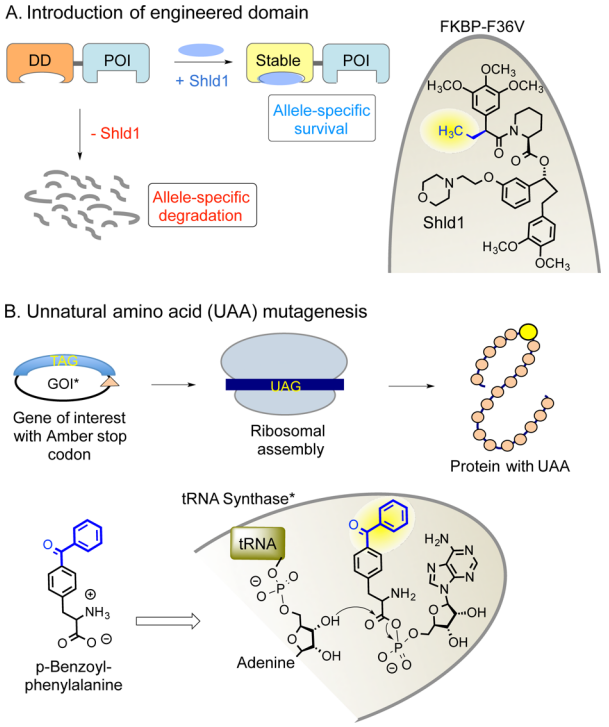


Figure 5. More approaches of ASCG. (A) Introduction of engineered domains: a protein of interest can be specifically manipulated by fusing it to an engineered death domain (DD) whose stability can be conditionally controlled using small-molecule ligand. Shld1 is a small molecule that protects POI from degradation by allele-specific binding to F36V variant of destabilized FKBP. (B) Allele-specificity through UAA mutagenesis: Amber stop codon (TAG) is introduced in gene of interest (GOI) to incorporate an UAA using an evolved orthogonal tRNA and its cognate synthetase that has been engineered to expand the active site cavity (bump-hole) to accommodate bulky UAAs such as *p*-benzoyl-phenylalanine.

in FRB).[53] Using *in vivo* selection and error prone PCR, they identified an L106P mutation that confers rapid instability to FRB* but can be rescued by a small molecule called Shield-1 (Shld1) (Figure 5A). Interestingly, Shld1 can bind to FRB* (but not wild-type FRB) in a shape-complementary manner (bump-hole), ensuring allele-specificity. Conditional stability of constitutively active mutants of various proteins fused with FRB*-L106P were achieved by controlled addition of Shld1 and well corroborated with the expected cellular phenotypes that were reversed upon retrieval of the ligand. A similar strategy was employed for the allele-specific aggregation of proteins and demonstrated its use in ligand-dependent conditional release of insulin from secretory glands.[56] Wandless group later developed another destabilizing domain by mutating *E. coli* dihydrofolate reductase (ecDHFR) that displayed trimethoprim (TMP)-dependent rapid, reversible and dose-dependent stability.[57] The system was employed to demonstrate conditional expression of green fluorescent protein (GFP) in rat brain.

Unnatural Amino Acid (UAA) Mutagenesis. Mutant alleles carrying UAAs with biologically orthogonal chemical properties are powerful systems for function elucidation (Figure 5B).[58] Complementary chemical or physical stimuli can be applied to reveal the roles of proteins bearing UAAs in complex *in vivo* environments. Using nonsense suppressor mutagenesis, the Schultz group successfully incorporated UAAs embellished with distinct reactivity into mutant alleles in a site-specific manner.[59] A highly evolved orthogonal pair of tRNA and its cognate synthetase accomplished such allele-specific incorporation of UAAs into proteins. Remarkably, steric complementarity was employed to expand the hydrophobic pocket of the synthetase to accommodate the "bump" in UAAs (Figure 5B).[60] Introduction of space in the active site also made the engineered synthetase orthogonal to the canonical amino acids. This powerful allele-specific chemical genetic approach has thus far been employed to incorporate more than 70 UAAs in a wide range of proteins having regulatory, structural, and/or enzymatic roles.[58] Incorporation of photoactivable UAAs into proteins, for example, allowed rapid activation of an allele in high spatiotemporal resolution using light to study its role in fast dynamic biological events or to identify and characterize transient protein−protein interactions.[61,62]

DEVELOPMENT AND APPLICATIONS OF BUMP−HOLE SYSTEMS FOR DIVERSE PROTEIN−SMALL MOLECULE PAIRS

Conserved cofactor-binding sites, identical catalytic mechanisms, and overall structural homology pose challenges in studying a single member of an enzyme family/superfamily. Allele-specific technologies, particularly the bump-and-hole approach, have been extremely successful in elucidating enzyme functions. The increasing popularity of this approach among chemical biologists is mostly due to the druggability of enzymes (presence of well-defined small-molecule binding site), wealth of structural information, and ready access to cofactor-based activators/inhibitors. Based on Shokat's seminal work on protein kinases, a number of enzyme-cofactor systems have been engineered to develop allele-specific pairs for functional elucidation. Many recent studies have combined bump−hole methodology with bioorthogonal ligations and "omics"-based profiling to identify substrates of a specific enzyme from complex cellular environments. These efforts are outlined in more detail below.

Protein Kinase. Early efforts to generate an analog-sensitive kinase allele involved replacing a conserved "gatekeeper" bulky hydrophobic residue in the active site with alanine or glycine (Figure 6A).[33] The underlying design principle was based on the structural analysis of the ATP- and/or inhibitor-bound protein kinases. For example, crystal structures of AMP-PNP (a nonhydrolyzable ATP analogue) and PP1 (a nonspecific kinase inhibitor) bound c-Src kinases clearly show that conserved T338 resides closely to the adenine moiety of these inhibitors and would undergo clash with the bulky compound such as 1-naphthyl analogue of PP1 (NA-PP1).[63] This analysis suggests that T338A/G mutations (hole) should release the steric repulsion by providing enough space to accommodate the larger substituents (bump) present in the inhibitor scaffold (Figure 6A). These mutations, however, do not change the ATP binding to significant degree and act much like the wild-type kinase because ATP binding does not rely on interactions with the hydrophobic pocket guarded by the gatekeeper residue. The steric bulk prevents these inhibitors from binding to the wild-type kinases, thus generating an allele-specific mutant-inhibitor pair. Initial work from Shokat group showed that I338A/G mutants of oncogenic tyrosine kinase v-Src (equivalent of T338A/G mutations in c-Src) could be potently inhibited by NA-PP1 or 1-naphthylmethylene analogue of PP1 (NM-PP1) (Figure 6B).[64] These analogues had no significant inhibitory activity against wild-type kinases. Sequence align-

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

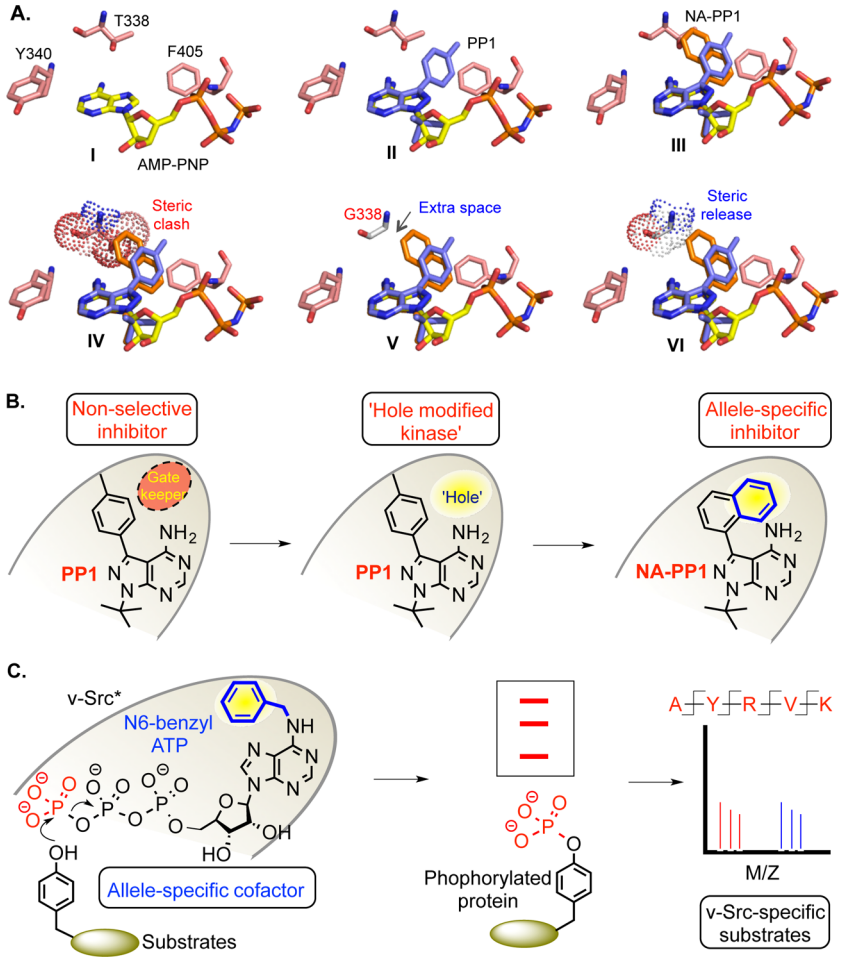


Figure 6. ASCG for protein kinases. (A) Structure-based guidance to develop allele-specific pair of c-Src kinase. Analyses of crystal structures of AMP-PNP and PP1-bound Src kinases (I and II, PDB codes: 2SRC and 1QCF) suggest that a bumped inhibitor such as NA-PP1 would undergo steric clash with the conserved tyrosine (T338) that acts as gatekeeper (III and IV, coordinates for NA-PP1 were obtained from PDB code 4K11).[63] Glycine mutation (hole) at this site relieves the steric clash and provides sufficient space to accommodate the bumped NA-PP1 (V and VI). Such structural information has been essential for the development of allele-specific cofactor and inhibitors for more than 80 kinases as well as other proteins. (B) Similar engineering approach for v-Src kinase: I338 in v-Src similar to T338 in c-Src acts a gatekeeper that restricts the binding of bulky NA-PP1. The bumped inhibitor, however, can selectively inactivate the I338G mutant. (C) The I338G mutant of v-Src binds N6-benzyl ATP analogues with wild-type efficiency and has been exploited to identify substrates specific to v-Src from cellular environment using proteomic approach.

ment of large numbers of eukaryotic protein kinases revealed that residues corresponding to position 338 in Src kinases were mostly bulky and hydrophobic, suggesting the generality of the approach and ensuring the potential orthogonality across the kinome.[33] Shokat's success in engineering protein kinases provided an important guideline for developing ASCG strategy: structural information is essential to make an intelligent guess on introducing steric complementarity at the protein−ligand interface. A powerful aspect of the strategy is that it did not rely on further structural validation. For majority of the protein classes (as will be delineated in this section), the initial best guess is often sufficient for finding the active orthogonal pairs. A small collection of synthesized cofactors/inhibitors analogues is usually screened against a number of "hole-modified" mutants in an activity assay.

To date analog-sensitive alleles have been successfully developed for more than 80 kinases from different families such as Src kinases, cyclin-dependent kinases, mitogen-activated kinases, p21-activated kinases, and calmodulin-dependent kinases to examine their roles in various organisms.[65] Notable applications include elucidation of the role of CaMKIIα in learning and memory in mice[66] and that of Cdc28 in cell-cycle regulation in budding yeast.[67] Of particular importance were efforts to interpret the role of Cdc28 in G2/M arrest using a 1NM-PPI analog-sensitive Cdc28 mutant.[67] These results contradicted an earlier observation that Cdc28's were involved in the G1/S phase transition, based on genetic approaches using a temperature-sensitive allele.[68] Because 1NM-PP1 does not show any off-target activity, the observed phenotypic difference is likely due to the fact that 1NM-PP1 blocks only catalytic activity of Cdc28, where the genetic method fundamentally disrupts the protein fold and, hence, all encoded functions. In another study, O'Shea and co-workers showed that phenotypes obtained using ASCG can be fundamentally different from those obtained by genetic approaches.[69] Using an analog-sensitive Pho85 kinase mutant and complementary small molecule inhibitor, they revealed an unexpected role of the kinase in glucose metabolism, which was not previously observed using a knockout strategy. However, by analyzing the complete set of Pho85-dependent gene expression profiles,

dx.doi.org/10.1021/cb500651d | ACS Chem. Biol. 2015, 10, 343−363

adaptation of gene expression over time was observed, consistent with the phenotypic differences between the two strategies and illustrating the usefulness of ASCG in studying processes regulated on a rapid time scale. In an attempt to understand the role of autophosphorylation of Ire1 to induce its endonuclease activity to trigger the unfolded-protein response (UPR) in endoplasmic reticulum, Papa *et al.* created "hole-modified" Ire1-L745G mutant that was expected to be sensitized by 1NM-PP1.[70] However, instead of inhibition, the mutant demonstrated 1NM-PP1-mediated activation and robust endonuclease activity most likely through required conformation changes.[70] Such observation, although unexpected, supports an unprecedented mechanism of the kinase domain (ligand binding instead of kinase activity) to propagate the UPR signal.

With ∼520 human protein kinases and an estimated 20 000 phosphorylation sites, the phosphoproteome constitutes a complex interacting network essential for cellular functions.[71,72] The identification of substrates for a specific kinase is an important first step toward deconvoluting its biological roles. In an elegant approach, Shokat and colleagues developed [γ-^{32}P]-labeled orthogonal ATP analogues containing sterically bulky groups in the adenosyl moiety of ATP (Figure 6C).[73] These ATP analogues were utilized by analog-sensitive kinase alleles as alternative cofactors to label their substrates with [^{32}P]-phosphate. Because only the mutant allele was active on these bulky ATP analogues and not the wild type, the orthogonal approach allowed identification of substrates for only the engineered kinase. Using such a strategy, novel candidate substrates of v-Src such as coflin, Dok-1, and calumenin were identified in NIH 3T3 cells. In another study, more than 200 substrates were identified for the cell-cycle regulator CDK1 in budding yeast, highlighting the importance of this kinase in myriad biochemical processes.[74] The Shokat group later developed another set of ATP analogues by replacing the [^{32}P]-phosphate with [γ-thiophosphate] and specifically enriched the thiophosphate-modified proteins using iodoacetyl-agarose. This technique has been successfully employed to identify >70 substrates of CDK1.[75]

Myosin and Kinesin. Following the success of ASCG strategies on protein kinases, two other ATP-utilizing enzymes, myosin and kinesin, were subjected to the same approach. These are members of the large family of molecular motors that use chemical energy of ATP-hydrolysis to carry out diverse mechanical processes such as cell motility, trafficking, organelle transport, spindle formation, chromosome segregation, and cytokinesis. The human genome encodes more than 40 myosin genes; however, functional analysis of these enzymes has remained a challenge owing to the large numbers of redundant isoforms with overlapping biochemical and catalytic properties.[76] Gillespie *et al.* employed the ASCG approach to demonstrate direct involvement of a specific myosin isoform in adaptation of the sensory hair cells of the inner ear.[77] It has long been thought that molecular motors such as myosins are involved in the adaptation of hair cells to sustained stimuli; however, extensive biochemical investigations could not determine the exact isoform due to the large number of myosin subtypes expressed in these cells. In their approach, Gillespie *et al.* developed N6-(2-methylbutyl)-ADP (NMB-ADP), which specifically inhibits a Y61G allele of myosin 1-β.[78] Their design was based on the rationale that myosins display strong product inhibition by ADP, as ADP release is the rate-determining step in the catalytic cycle (Figure 7A). In an *in vitro* motility assay, the mutant was active in the presence of ATP but was inhibited by NMB-ADP to stall motility, consistent with the observation that ADP-inhibited myosins strongly bind actin filaments. By microinjecting non cell-permeable NMB-ADP analogue into sensory epithelia derived from transgenic mice carrying Y61G mutation, the Gillespie group demonstrated that myosin 1-β-mediated adaptation of hair cells occurs over tens of milliseconds time scale.[77]

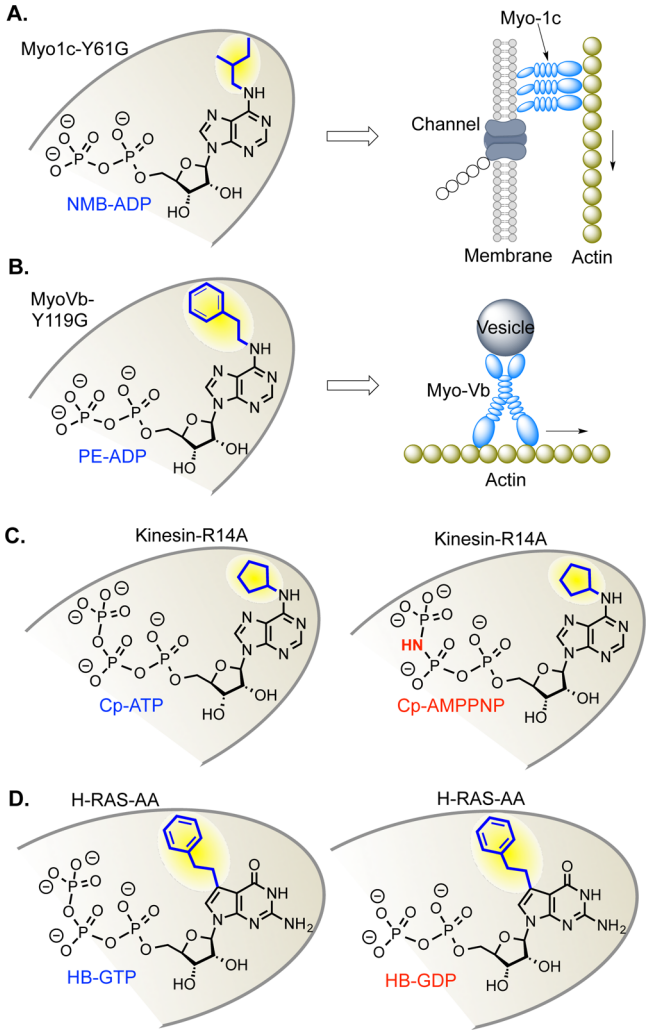


Figure 7. Applications of ASCG to ATPases and GTPases. (A) Using the orthogonal pair of Myo1c-Y61G (hole) and NMB-ADP (bump), it was demonstrated that myosin-1c drags tip link and transduction channel in response to sustained stimuli to achieve adaptation of hair cells and that the process occurs in tens of milliseconds (an example of application of ASCG to study fast, dynamic process). (B) A similar bump−hole pair where the "hole-modified" mutant and an ADP-based allele-specific inhibitor were employed to elucidate the role of myosin-Vb in intracellular transport. The arrows in A and B indicate the movement of myosins along actin filaments. (C) An R14A mutant of kinesin can be selectively modulated by cyclopentyl-ATP/ADP analogues. The cyclopentyl bump fits into the expanded cavity of the R14A mutant; (D) the orthogonal pairs based on small G-protein H-Ras: the picomolar binding affinity of GTPase-GT(D)P was engineered through double mutation of H-Ras and C7-homobenzyl analogues of GT(D)P. The pairs were employed to identify cancer-relevant effector proteins of H-Ras.

In a similar venture, Mercer and co-workers developed a Y119G mutant of myosin Vb capable of hydrolyzing ATP, which could be specifically inhibited by an ADP analog N6-(2-

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

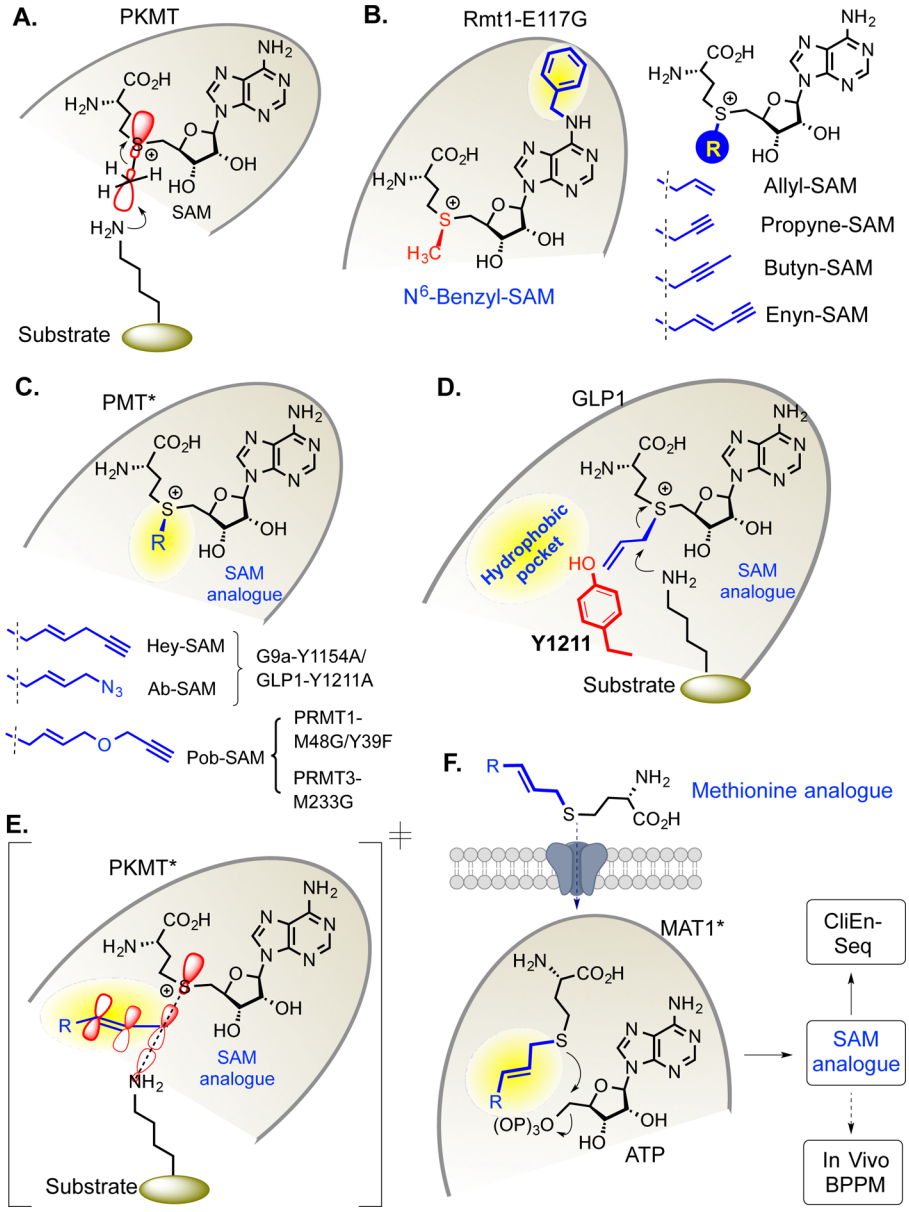


Figure 8. Development of ASCG systems for methyltransferases. (A) PKMT (protein lysine methyltransferase)-mediated methylation of the ε-amino group of lysine using SAM (S-adenosyl-L-methionine) as a cofactor follows an S_N2 mechanism involving a nucleophilic attack by amine nitrogen at the σ^* orbital of sp^3-hybridized sulfonium methyl group; (B) SAM analogues as alternative cofactors for wild type and mutant methyltransferases: N6-benzyl-SAM carrying modification on adenine is active with an Rmt1-E117G mutant. Multiple analogues such as allyl-, propyne-, enyn-SAM have shown their potential as alternative cofactors for a range of wild-type protein, DNA and RNA methyltransferases. (C) "Clickable" extended SAM analogues that are active toward "hole-carrying" PKMTs and PRMTs. These SAM analogues coupled with respective PMT mutants have been employed to modify protein substrates and identify them by proteomics. (D,E) Structural and kinetic characterization of "bump-hole" pairs: a conserved tyrosine (Y1154 of G9a and Y1211 of GLP1) acts as gatekeeper that restricts the ability of wild-type PKMTs to bind SAM analogues. Y1154A/Y1211A mutations remove the gatekeeper and allow SAM analogues to bind in the previously inaccessible hydrophobic cavity (D). Such rational was derived from the crystal structure of GLP1-Y1211A bound to the products (SAH and allylated peptide) of the enzymatic reaction as detailed in ref 105. The Cβ═Cγ double bond in SAM analogues participates in transition-state stabilization in the PKMT-mediated alkylation of lysine (E). In the proposed linear transition state, the sp^3-hybridized α-methylene carbon is rehybridized to sp^2 and the incipient p orbital is stabilized through delocalization with adjacent antibonding π orbitals, facilitating the S_N2 reaction. (F) *In vivo* synthesis of SAM analogues: "hole-modified" MAT1 can accept "bumped" methionine analogues to synthesize SAM analogues using endogenous ATP. *In situ* generated SAM analogue was recognized by a G9a-Y1154A mutant to modify chromatin for subsequent analysis by CliEn-Seq technology.

phenylethyl)-ADP (PE-ADP) (Figure 7B).[79] The engineered mutant−inhibitor pair was employed to decipher the role of myosin Vb in membrane trafficking between peripheral and pericentrosomal compartments. More recently, Ehlers and colleagues employed the mutant-inhibitor pair in conjunction with biochemical and genetic methods to demonstrate that myosin Vb captures and mobilizes recycling endosomes for trafficking of the AMPA receptor, an essential step in learning-related plasticity at excitatory synapses in the mammalian brain.[80]

Kinesins are microtubule-based ATP-driven motor proteins that play critical roles in processes such as intracellular

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

transport, neuronal development and chromosome segregation. Because of their mitosis-specific role, kinesins are preferred anticancer drug targets that could minimize the side effects of chemotherapeutics targeting tubulins/microtubules. However, detailed biological study of human kinesins has remained a challenge due to the large number of isoforms, conserved ATP-hydrolyzing domain, and cell- and tissue-specific expression. To address this, Mitchison and co-workers employed the ASCG approach by mutating a conserved arginine residue in the ATP pocket to either alanine or glycine to specifically accept N6-substituted ATP analogues (Figure 7C).[81] It is worth noting that removal of the guanidinium group completely abolished affinity toward ATP, but the mutant remained active toward a cyclopentyl-modified ATP analogue that recapitulates some of the binding affinity through steric complementation. The Mitchison group also developed a nonhydrolyzable ATP analogue bearing the N6-cyclopentyl moiety to specifically inhibit a mutant allele, consistent with the kinesin-mediated catalytic cycle where ATP hydrolysis is the rate-limiting step (Figure 7C).[81] Effectiveness of the engineered system was demonstrated in *in vitro* motility experiments.

Chaperonin GroEL. The chaperonin GroEL is an ATP-dependent megadalton molecular motor that plays a critical role in bacteria by assisting protein folding. Chapman and co-workers applied the bump-hole approach to develop small molecule inhibitors selective for a mutant allele of GroEL.[82] In an *in vitro* assay, an optimized pyrazolopyrimidine based compound, named EC3016, specifically inhibited the ATPase activity of an I493C variant but not wild-type GroEL. The mutant−inhibitor pair was tested for the ability to modulate the refolding of two GroEL substrates, bovine rhodnese and pig heart malate dehydrogenase.

GTPase. G-proteins constitute a large fraction of the human proteome and are involved in diverse cellular processes such as cellular differentiation and organelle trafficking. In a small G protein-mediated signaling process, the GTP-bound state typically corresponds to an "on" conformation that allows interaction with specific effector molecules to initiate downstream signaling cascades. A GDP-bound state switches "off" the process by inducing conformational changes in the receptor. Adding to the complexity of the signaling process being studied, the "on" and "off" states are governed by a set of interacting proteins such as GTPase-activating proteins (GAPs), GDP-dissociation inhibitors (GDIs), and guanine nucleotide exchange factors (GEFs). As mentioned earlier, Miller and co-workers developed an allele-specific approach to study G proteins *in vitro* by converting them to XTPases (Figure 3A).[14] However, this technique is not suitable for *in vivo* application, as the intracellular concentration of XTP must be kept very low to prevent its incorporation into nucleic acids. At such low concentrations of XTP and relatively high concentration of GTP, the mutant GTPases may still prefer to bind GTP (and GDP) thus making the allele-specific system ineffective in cells.

To orthogonally regulate GTPase-mediated signaling processes *in vivo*, Shah and colleagues developed a GTPase mutant specifically sensitive to GTP and GDP analogues as potential activator and inhibitor.[83] Careful analysis of the crystal structure of H-Ras and sequence alignment of multiple small G proteins led to the identification of L19 and N116 residues, situated close to the guanidine ring of GTP/GDP, as potential sites for mutagenesis. Subsequent protein engineering, analogue synthesis, and *in vitro* biochemical assays identified an H-Ras L19A/N116A double mutant (H-Ras AA) that accepted a C7-homobenzyl modified nucleotide derivative as alternative cofactor with efficiency comparable to the wild-type system (Figure 7D). A C7-homobenzyl GDP derivative was developed as a potent inhibitor for the mutant-allele. A series of experiments demonstrated that the allele-specific pairs efficiently switched the G-protein-mediated signaling process "on" and "off" much like the native system.[83] The utility of this engineered pair to elucidate the function of a specific G protein was further demonstrated by identifying Nol1 as putative downstream effector of H-Ras. Given the established role of Nol1 in human cancers, the work by Shah and colleagues suggested a potential mechanism by which Ras may regulate oncogenesis.

Methyltransferase. S-Adenosyl L-methionine (SAM)-dependent methylation of nucleic acids, proteins, and small molecules constitutes a major form of post synthetic modification that controls almost every aspect of cellular biology.[84] Most prominent are the epigenetic roles of nucleic acid and histone methylation in gene regulation, cellular differentiation, and organismal development. More than 100 human methyltransferases (MTs) are involved in methylation of lysine, arginine, cysteine, aspartic acid, glutamic acid, cytosine, adenine, and thymine in proteins, DNA, and RNA using essentially the same cofactor SAM (Figure 8A). These methylation events regulate localization, stability, and interactions of macromolecules.[85] Such large numbers of MTs coupled with their cell, tissue, and context-dependent dynamic activity have made elucidation of functions of a specific methyltransferase a challenging task. Candidate-based approaches have revealed multiple substrates of MTs, further adding to the complexity of their biological roles.[86] Unambiguous identification of the methylome (the collective target set) for a given MT has remained a challenge. The chemical inertness, lack of polarity, and small size of the methyl group have precluded development of high quality enrichment techniques for methylated biomolecules.

In recent years, a series of studies have surfaced in literature that address the above challenges by employing allele-specific systems of engineered MT and complementary SAM analogues (Figure 8B,C). In an early effort, Gray and co-workers demonstrated that the E117G mutant of yeast MT Rmt1 could accept an N6-benzyl modified SAM analogue.[87] The mutant showed catalytic efficiency 500-fold less than the wild-type system but was 65-fold more active toward SAM analogues than SAM (Figure 8B). This is another example where polar residue was replaced with glycine to create space to accommodate the hydrophobic benzyl moiety. The engineered pair was employed to label Rmt1 substrates *in vitro*.[87]

The Weinhold group demonstrated that native bacterial DNA methyltransferases (DNMTs) are capable of transferring bulky substituents from SAM analogues to the corresponding DNA substrates.[88−90] The implications of Weinhold's work are 2-fold. First, instead of modification at the N6 position, the sulfonium methyl group was replaced with the bulky moiety to uniquely modify the substrate (Figure 8B). Second, the work further improved catalytic efficiency for transfer of the bulky group by generating a hole-modified Q82A mutant of M. *Hha*I.[88] The group later expanded the engineering approach to other bacterial DNMTs.[91] The Weinhold group went on to develop Enyn-SAM, an analogue capable of transferring a terminal alkyne via a pent-2-en-4-ynyl moiety.[92] The ability of the terminal alkyne to act as a substrate in Cu-catalyzed azide

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

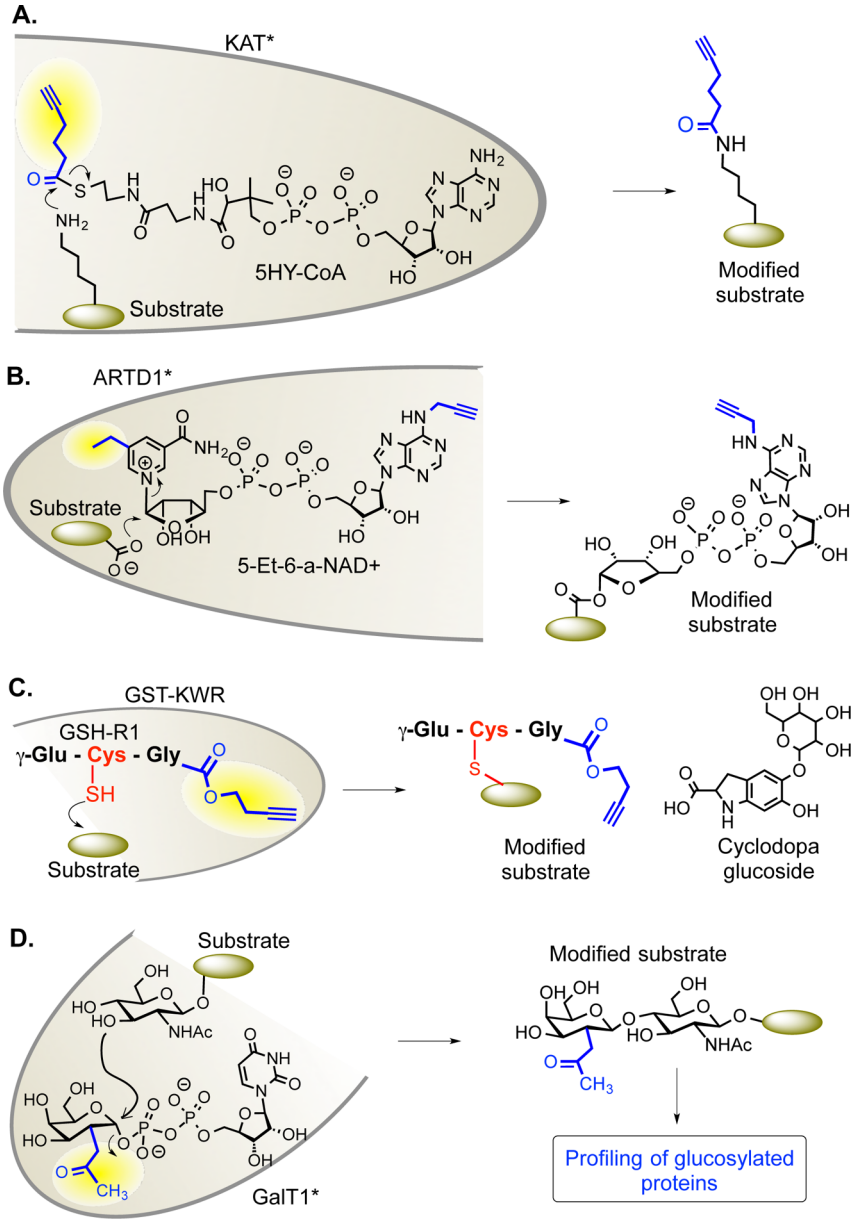


Figure 9. Applications of "bump−hole" approaches to profile substrates of multiple transferases: acetyltransferase (A), ADP ribosyltransferase (B), glutathione S-transferase (C), and β-galactosyltransferase (D). In all the cases, enzymes were engineered (hole-modified) to accept bulky cofactors (bump) to modify substrates with chemical functionalities amenable to bioorthogonal reactions such as copper-catalyzed azide−alkyne cycloaddition for subsequent enrichment and proteomic analyses.

alkyne cycloaddition (CuAAC)[93] made this result a significant advance in MT substrate profiling. Around the same time, Bertozzi and Gozani developed a long-sought but not so stable analogue propyne-SAM to label the substrate of the protein methyltransferase (PMT) SETDB1.[94] The Weinhold and Luo laboratories independently developed selenium-based SAM (ProSeAM) to circumvent the stability issue and demonstrated its complementarity to a range of wild-type PMTs.[95,96] However, because only a handful of native methyltransferases can transfer the extended moiety,[97−100] a general approach to profile substrates of a designated PMT/DNMT was still lacking.

To address this challenge, Luo and co-workers developed a strategy termed 'Bioorthogonal Profiling of Protein Methylation (BPPM)' for the discovery and characterization of PMT-specific methylation events in human cells.[101] The central component of the approach was to engineer the active site of a PMT to create a space to accommodate and catalytically transfer the bulky sulfonium group present in a SAM analogue to substrate molecules (Figure 8C). In a series of publications, the Luo group developed such an allele-specific catalytic system for multiple PMTs and SAM analogues.[102−106] More specifically, for lysine methyltransferases (PKMTs) G9a and GLP1, alanine mutation at a highly conserved tyrosine (Y1154 and Y1211, respectively) sensitized the mutant alleles to multiple bulky SAM analogues such as Hey-SAM and Ab-SAM with catalytic efficiencies comparable to wild-type pairs (Figure 8C).[105] Subsequent structural studies revealed that the conserved tyrosine acts as a gatekeeper in the active site, preventing the wild-type PKMTs from binding SAM analogues.[105] Interestingly, removal of the gatekeeper did not create a "hole" for the substrate but rather allowed bulky SAM

dx.doi.org/10.1021/cb500651d | ACS Chem. Biol. 2015, 10, 343−363

analogues to access a preexisting hydrophobic pocket nearby (Figure 8D). A similar mechanism was observed for an I338G mutant of v-Src protein kinase.[107] It is worth mentioning that the allele-specific activity was a result of a strong steroelectronic complementation that stabilized the transition state of an S_N2-type lysine methylation reaction (Figure 8E). Y1154/Y1211 in native G9a/GLP1 is known to form a critical hydrogen bond with the ε-amino group of the lysine undergoing methylation. Consistent with such observation, Y1154A/Y1211A mutation resulted in 250-fold loss in activity. However, the loss in hydrogen bonding was compensated by the "allylic stability" of the transition state through delocalization of β-sp2 electrons present in SAM analogues into the incipient p-orbital of the transferring methylene group. Complementing the electronic effect, Y1154A/Y1211A mutation provided space to house the extended group in the SAM analogues (Figure 8E).[105]

The engineered pairs (Y1154A/Y1211A-Hey SAM) were employed to identify more than one hundred novel substrates of G9a and GLP1, a subset of which was further validated *in vitro*.[105] To identify PMT substrates in a more relevant *in vivo* context, the Luo group made another seminal contribution. Hey-SAM was synthesized in living human cells by engineering the SAM synthetic machinery based on a shape complementarity approach (Figure 8F).[108] In biological systems, SAM is synthesized from ATP and methionine by SAM synthetase. Using an engineered SAM synthetase (MAT1-I117A mutant) and bulky Hey-methionine analogues, Luo and co-workers demonstrated the efficient synthesis of Hey-SAM in HEK293 cells. *In situ* generated Hey-SAM was further coupled with G9a-Y1154A/GLP1-Y1211A mutants to uniquely modify histone H3 in biologically relevant conditions (Figure 8F). Subsequently, "CliEn-Seq" technology was developed as an alternative to ChiP-Seq to enrich and sequence the underlying DNA revealing the genome-wide activity of highly homologous G9a and GLP1.[108] Recently, the Luo group showed that selenium-based SAM analogues have higher activity than their sulfur counterparts against a panel of PMT mutants.[109] The Zhou group have succeeded in tailoring viral SET-domain PMT (vSET) to process a novel 2′,3′-dibenzyl SAM analogue, thus increasing the repertoire of biologically active SAM analogues.[110]

Acetyltransferase. Protein acetylation is a posttranslational modification that controls diverse biological processes. Of particular importance to epigenetic control over gene regulation is histone acetylation. A class of enzymes called lysine acetyltransferases (KATs) catalyze protein acetylation by employing Ac-CoA as the sole acetyl donor. Biochemical and proteomic studies have shown that a large fraction of the human proteome undergoes acetylation, suggesting these modifications may impact a wide range of biological processes.[111,112] Acetylome profiling has also been performed through the transfer of terminal alkyne-containing Ac-CoA analogues catalyzed by promiscuous KATs.[113] To determine the substrates of a specific KAT, Zheng and colleagues employed an ASCG approach by engineering representative KATs to accept synthetic Ac-CoA surrogates with chemical handles for bioorthogonal ligation and subsequent detection and enrichment (Figure 9A).[114] Zheng *et al.* created double mutants of GCN5 (T612G/L531A) and MOF (I317A/H273A) and demonstrated their ability to transfer a six-carbon unit from the cofactor analogue 5HY-CoA to respective peptide substrates (Figure 9A). The potential of these allele-specific mutant-cofactor pairs in detecting the complete acylome of a given KAT was further validated by exogenously adding the engineered pairs to HEK293 cell lysates followed by "click" chemistry labeling and in-gel fluorescence analysis. Proteomic characterization of these substrates and development of cell-permeable Ac-CoA analogues for *in vivo* application would be the next step forward for functional annotation of various members of the KAT family.

ADP Ribosyltransferase. ADP diphosphate ribosylation is a posttranslational modification catalyzed by a class of 17 ADP ribosyltransferases (ARTDs). The enzyme employs nicotinamide adenine dinucleotide (NAD^+) as its cofactor, and the modification plays an essential role in DNA repair and energy metabolism. A conserved catalytic mechanism, overlapping target preferences, and functional redundancy make elucidation of direct targets of specific ARDTs a formidable task. Recently, Cohen and co-workers have developed allele-specific pairs of ARDT mutants and "clickable" NAD^+ analogues to address this challenge (Figure 9B).[115] In their approach, the Cohen group has identified a conserved lysine (K903 for ARTD1) that when mutated to alanine (KA-ARTD1) can be specifically inhibited by 5-ethyl nicotinamide. Based on this observation, an allele-specific cofactor for KA-ARTD1 was constructed by merging 5-ethyl nicotinamide with 6-a-NAD^+, a clickable NAD^+ derivative originally developed by the Lin group (Figure 9B).[116] The 5-Et-6-a-NAD^+ was also shown to be effective in KA-ARTD2. To demonstrate the utility of their approach, the engineered pairs were subsequently employed to isolate ARTD substrates from HEK293T cells via bioorthogonal "click" ligation and biotin–avidin enrichment. LC-MS/MS analysis of digested peptides led to the identification of several novel substrates of ARDT1 and 2, both enzymes having common and nonoverlapping targets. Immunoblot detection was used to further confirm that Ku80 is a common target of both ARDT1 and 2, while Catenin-δ and hnRNP Q/R are specific to ARDT2.

Glutathione S-transferase. Reversible nucleophilic addition of the sulfhydryl group of reduced glutathione (GSH) to various endogenous metabolites is essential in cellular detoxification and drug metabolism. This process is catalyzed by glutathione S-transferase (GST), which is present in various isoforms in both prokaryotes and eukaryotes. To understand the role of GSTs in cellular processes, particularly in drug metabolism, it is imperative to identify their specific set of substrates, which can prove challenging given the small molecule nature of these substrates. Contributing to the recent effort toward substrate profiling for a wide range of transferases, Deng and colleagues formulated a technique called 'bio-orthogonal identification of GST substrates' (BIGS) by developing allele-specific pair of GST mutant and modified reduced glutathione.[117] Upon analysis of the crystal structure of GSTM1 from *S. japonicum*, they generated a K44G/W40A/R41A triple mutant (GST-KWR) that was able to modify a homopropargylic ester of GSH (GSH-R1) as its substrate with efficiency comparable to the wild-type GSTM1-GSH pair (Figure 9C). Although the mutant was still active toward GSH, it showed preference toward modified GSH in a competitive assay, demonstrating its functional orthogonality. As an interesting application, the GST-KWR and GSH-R1 pair was used to identify multiple substrates, such as cyclodopaglucoside and callystatin A, from the Chinese herbal medicine Ganmaochongli using LC-MS/MS of the enriched sample (Figure 9C).[117] Recently, Ahn and colleagues have demonstrated that hole-bearing glutathione synthetase can synthesize a GSH analogue containing bumped azido-alanine, providing a

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343–363

versatile handle for profiling and characterizing cellular glutathionylation.[118]

β-Galactosyltransferase. β1,4-Galactosyltransferase I (Gal-T1) catalyzes the transfer of galactose (Gal) from UDP-Gal to *N*-acetylglucosamine (GlcNAc) attached to serine or threonine residues in glycoproteins. Gal-T1 is also capable of transferring GalNAc from UDP-GalNAc to GlcNAc, but with only 0.1% catalytic efficiency. To understand and improve this low activity, Qasba and Ramakrishnan analyzed the crystal structure of Gal-T1 and identified Y289 that resides near UDP-GalNAc and forms a hydrogen bond with the substrate acetyl group.[119] Among multiple mutants, Y289L exhibited enhanced GalNAc transferase activity that is nearly equal to the wild-type protein.[119] The Y289L mutant was effectively employed by Hsieh-Wilson and colleagues to profile the GlcNAcylated proteome, an important posttranslational modification that has implications in host−pathogen interactions, inflammation and oncogenesis (Figure 9D).[120] In their work, the Hsieh-Wilson group replaced the NAc functionality in UDP-GalNAc with an isosteric 2-oxopropyl moiety and demonstrated that the resulting ketone-containing Gal unit can be efficiently transferred to GlcNAc in peptide substrate by the Y298L mutant (Figure 9D).[120] Employing an aminoxy-biotin reporter capable of selective reaction with the ketone functionality, novel GlcNAcylated proteins associated with gene expression, neuronal signaling, and synaptic plasticity were identified from rat brain.[121]

In a related study, to identify the binding partners of GlcNAcylated proteins, Kohler group developed diaziridine-containing GlcNDAz-1-phosphate (bump) that can be exploited by F383G mutant (hole) of UDP-GlcNAc phosphorylase (AGX1) to synthesize UDP-GlcNDAz in living cells.[122] GlcNDAz-modified proteins would undergo cross-linking with the interacting molecules upon UV irradiation and could be identified by mass spectrometry. Using this approach, Kohlar group identified binding partners of GlcNAc-modified FG-repeat nucleoporins, suggesting a potential link between the modification recognition and nuclear transport.[122]

Phosphatase. Protein phosphorylation is arguably the predominant mechanism for maintaining cell signaling networks in eukaryotes. Two opposite acting enzymes achieve tight spatiotemporal control of protein phosphorylation: protein kinases that "write" the mark and phosphatases that "erase" it. Given that the human genome encodes more than 100 protein tyrosine phosphatases (PTPs),[123] it is a considerable challenge to understand the functions of PTPs. Adding to the complexity is the large number of complexes a catalytic subunit of given phosphatase can form by interacting with different scaffolding subunits. Bishop and co-workers successfully implemented ASCG techniques to develop a specific small-molecule inhibitor for the mutant allele of human phosphatase PTP1B.[124] Alanine mutants at highly conserved sites V49 and I219 displayed wild-type activity toward phosphorylated substrates but were specifically inhibited by the *N*-methylindole derivative of an otherwise nonselective PTP inhibitor. The "bumped" inhibitor showed at least 10-fold selectivity toward the mutant allele over wild-type PTP1B. The generality of this approach was demonstrated by extending it to another phosphatase, TCPTP, whose I220A mutant (comparable to I219A of PTP1B) was similarly inhibited by the same designed inhibitor with 30-fold selectivity.[124] The Bishop group also developed thiophene-based compounds that were more potent and selective toward PTP mutant alleles.[125] The finding that mutation at conserved isoleucine sensitizes PTPs to a variety of inhibitors suggests that this may act as "gatekeeper" and can be exploited for functional elucidation of PTPs in signaling pathways.

Protease. Hydrolysis of proteins is achieved by a large number of human proteases, some of which are highly promiscuous (e.g., trypsin) while others display absolute specificity and only cleave substrates with a certain sequence (e.g., thrombin). In an early effort, Wells and co-workers demonstrated that protease specificity can be redesigned in an allele-specific manner by introducing shape complementarity at an engineered interface between protease and substrate.[126] Replacement of the catalytic histidine with alanine (H64A) in the serine protease subtilisin resulted in loss of catalytic efficiency by a factor of a million against its substrate *N*-succinyl-Phe-Ala-Ala-Phe-p-nitroanilide (sFAAF-pNA, a chromogenic analogue of the natural substrate). However, the mutant could hydrolyze a properly positioned histidine-containing "bumped" substrate sFA**H**F-pNA 400 times faster than "nonbumped" substrates. In this 'substrate-assisted catalysis',[127] the "bump" in substrate complemented the "hole" in active site to fulfill a catalytic role.

Using a similar approach, Agard and colleagues engineered α-lytic, an extracellular bacterial serine protease, with remodeled substrates specificity.[128] The enzyme is specific for X = Ala in the p1 position of the peptide substrate s-AAP**X**-pNA. From the available crystal structure of α-lytic bound with a covalent peptidic inhibitor, it seemed that specificity for Ala was due to the side chain of M192 of α-lytic, which occluded what would otherwise be a deep binding pocket. Although a M192A mutant showed 2-fold reduction in catalysis toward s-AAP**A**-pNA peptide, the maximal activity increased by 840 000-fold toward sAAP**F**-pNA substrates compared to the wild-type enzyme. Removal of just three methylene groups (80 $Å^3$ volume) produced a specificity pocket large enough to accommodate a phenylalanine in the peptide substrate. Improved catalytic efficiency was mainly attributed to the enhanced ability of the mutant to form the Michaelis complex. A phenylalanine-containing "bumped" peptide inhibitor was developed that specifically inhibited the mutant allele of α-lytic protease.[128]

Receptors. Allele-specific chemical genetics studies of cell-surface and nuclear receptors have provided a great deal of information on their role in signaling processes, synaptic functions, gene expression, and neuronal activity.[22] As discussed above, these orthogonal strategies relied mainly on engineering the electrostatic interactions present in receptor−ligand interfaces. However, for a subset of receptors, steric complementary has been successfully employed to alter ligand specificity. An early example includes generation of a RASSL system to activate an F435A mutant of histamine H1 receptor (H1R) solely by synthetic trifluoromethylphenyl-histamine (CF_3PheHA) analogues.[129] Using an NF-κB-driven reporter gene assay, the synthetic ligand displayed agonist activity exceeding the natural ligand histamine by ∼70 000-fold, with subnanomolar efficacy. The increased affinity of the mutant toward ligand analogues was mostly due to the extra space in the binding pocket of F435A mutant that can accommodate the bulky trifluoromethylphenyl moiety of CF_3PheHA.[129] Detailed modeling analysis revealed that in addition to the steric complementarity, remodeled π−π interactions and hydrogen bonding may play roles in the heightened specificity.

Recently, a systematic chemical and genetic engineering approach has been employed by Sternson and colleagues to

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

develop an allele-specific pair for ligand-gated ion channels (LGICs) that displayed distinct ligand selectivity and ion conductance properties (Figure 10A).[130] As a first step, a chimeric LGIC was developed by fusing the ligand-binding domain (LBD) of α7 nicotinic acetylcholine receptor (nAChR) with the ion pore domain (IPD) of either a cation-selective serotonin 5-HT3 receptor (α7-5HT3) or anion-selective glycine receptor (α7-GlyR) (Figure 10A). This resulted in functional channels with the same pharmacological profile but different ion selectivity. To generate allele-specific chimeric ion channels, the ligand-binding site, situated in the protein–protein interface of pentameric complex of nAChR, was hole-modified to accommodate bumped ligand analogues (Figure 10A). Using the structure of snail acetylcholine (ACh) binding protein complexed to nicotine as a model and a small molecule ligand PNU-282987 as a starting point, a library of ligand–receptor pairs was designed, synthesized, and screened for orthogonal functionality. A series of *in vitro* experiments led to the identification of bulky PNU-282987 analogues that can selectively activate the hole-modified mutant channels but not the wild type (Figure 10A). The mutants were screened for lack of activation by acetylcholine and nicotine to ensure allele-specificity. The mutant LBDs were termed "pharmacologically selective actuator modules" (PSAMs) and the corresponding cognate synthetic ligands are called "pharmacologically selective effector molecules" (PSEMs). The engineered channel-ligand pairs were tested to ascertain their effectiveness in a range of *in vivo* settings.[130]

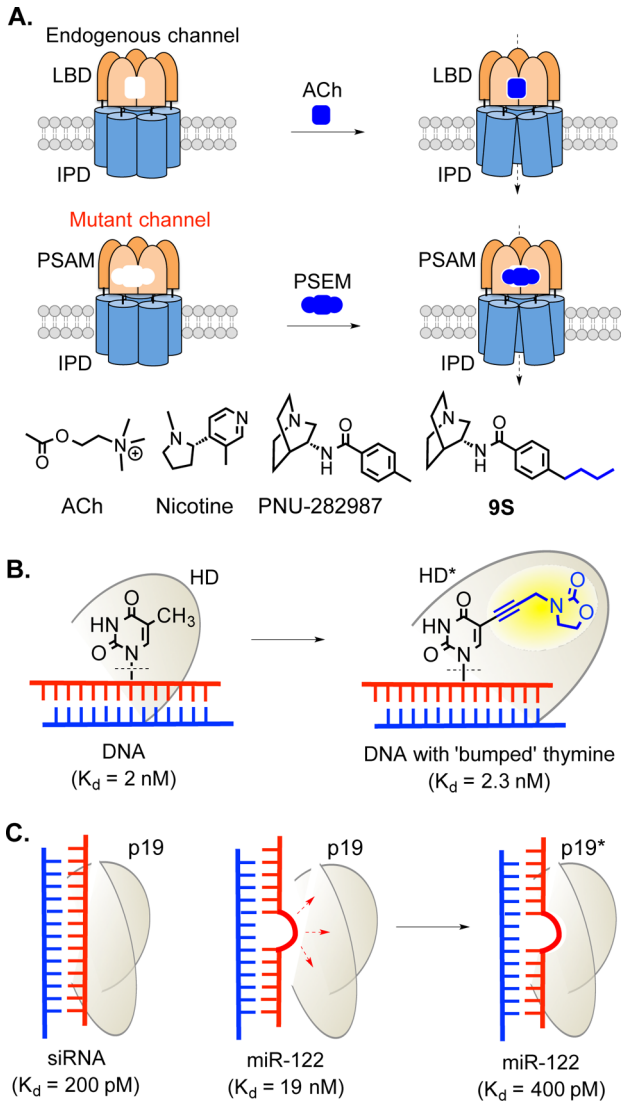


Figure 10. Applications of "bump-hole" approaches to regulate protein–protein, protein–DNA, and protein–RNA interactions. (A) Chimeric channels were generated by fusing the ligand-binding domain (LBD) of nAChR with the ion pore domain (IPD) of either a cation- or an anion-selective receptor. The ligand-binding site at the protein–protein interface was hole modified to accommodate bulky ligands such as **9S** to generate allele-specific ligand-gated ion channels (adapted from ref 130). (B) The DNA-binding site of a homoedomain zinc finger protein was expanded to accommodate bumped thymine containing DNA with wild-type binding affinity. (C) p19 protein was hole-modified to discriminate the binding between siRNA and miRNA. "Bulging" (equivalent to "bumped") structures of miRNA fit better into the "hole-carrying" p19 mutant allele (adapted from ref 133).

DNA- and RNA-Binding Proteins. The majority, if not all, cellular processes are manifestations of dynamic interactions of biomolecules. Although ASCG has mainly involved engineering protein-small molecule interfaces to examine biology of a given protein, similar approaches could be immensely useful to develop allele-specific DNA/RNA–protein interactions for functional elucidation. In one such study, Shokat and colleagues engineered a homeodomain (HD) protein to recognize unnatural nucleotide-containing DNA (Figure 10B).[131,132] HD is a DNA binding domain (DBD) protein that recognizes the nonpalindromic sequence TAATCC. Structural data suggests that the C5-methyl of thymine (T) contacts residue I47 of HD. Given that C5 substitution on thymine still maintains B-DNA structure, a propynyl oxazolidinone (bump) containing thymine was introduced into the DNA that completely abrogated the interaction with HD. An I47A mutant (hole) was able to release the steric clash to improve binding (Figure 10B). Further engineering led to a triple mutant I45V/I47G/K52M that demonstrated at least 40-fold preference to the bulky DNA analogues.

Small, noncoding RNAs are essential regulators of gene expression and seem to influence almost every aspect of eukaryotic biology. However, such regulation is largely a manifestation of a multitude of RNA–protein interactions. There has been effort to develop an allele-specific strategy akin to bump–hole to selectively perturb such interactions. In an effort to specifically regulate RNA–protein interactions, Pezacki and colleagues engineered a virus protein p19 that is known to bind both siRNAs and miRNAs, albeit in different affinity (Figure 10C).[133] A T111S mutant of p19 was generated and shown to improve the binding affinity toward miRNAs by 48-fold. The rationale for the improved binding was the bulging (bump) structure of miRNAs that fit better into the mutant p19 pocket (Figure 10C). The p19-T111S mutant was applied to sequester miR122 from human hepatoma cell lines.[133]

■ ALLELE-SPECIFIC CHEMICAL GENETICS IN PHARMACOLOGY

A large numbers of human diseases are associated with mutations in critical gene products that affect their normal functions either by misregulation of enzymatic activity (up- or down-regulation) or disruption of protein–protein interactions. In many cases, these mutations occur in the ligand/cofactor binding sites and change cofactor binding by manipulating steric and polar interactions between receptor and ligand

dx.doi.org/10.1021/cb500651d | ACS Chem. Biol. 2015, 10, 343–363

(Figure 11A). An engineered cofactor/ligand analogue that can reintroduce the lost interactions in a complementary manner

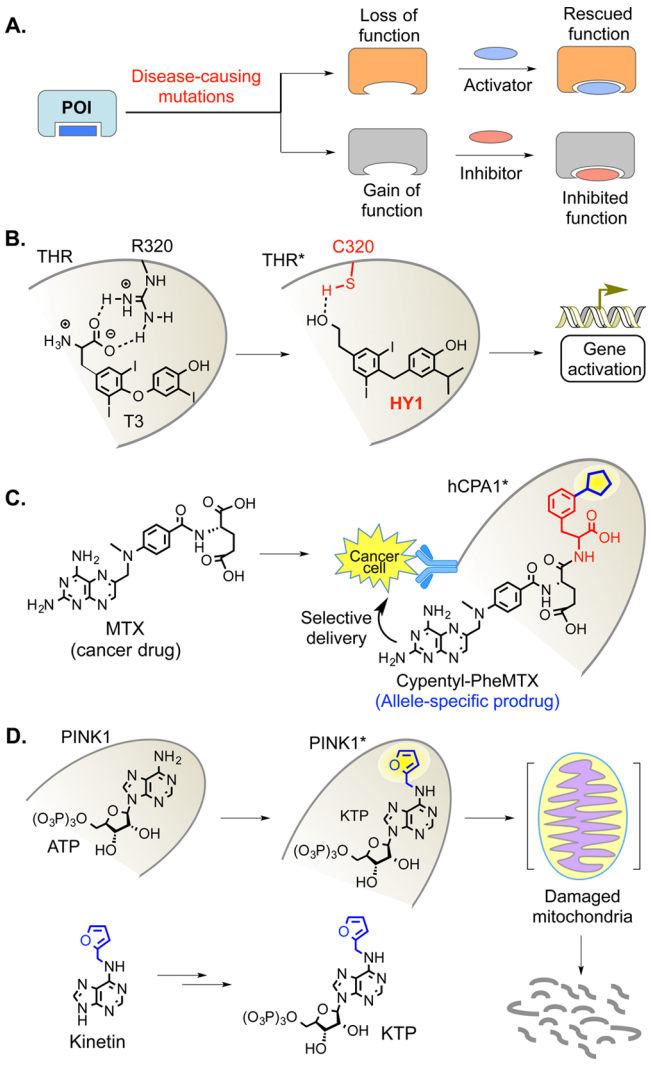


Figure 11. ASCG in pharmacology. (A) Mutations critical to human disease occur at the enzyme-cofactor/ligand−receptor interfaces disrupting normal functions. Correcting such loss-of-function or gain-of-function mutations could be an effective therapeutic approach. (B) The loss-of-function mutation in thyroid hormone receptor (THR) neutralizes charge interactions between the ligand T3 and residue R320 of THR. HY1, a synthetic neutral analogue of T3, restores the interaction and activates the otherwise inactive mutant in gene transcription. (C) "Hole-modified" human carboxypeptidase 1 can selectively cleave a "bumped" peptide substrate containing a MTX prodrug, thus improving selectivity and reducing immunogenicity associated with the foreign peptidase commonly used in similar prodrug therapies. (D) Kinetin triphosphate (KTP), a bumped ATP analogue that can be synthesized inside cells using kinetin, reactivates mutant PINK1 kinase to phosphorylate Mfn2 for degradation of the damaged mitochondria in Parkinson's disease.

would be an exciting approach for therapy (Figure 11A). An early effort in this direction came from Koh and colleagues in their design of thyromimetics specific to the disease-causing mutations of thyroid hormone receptor (THR).[134] R320C mutation in a THR subtype TRβ abolishes ligand-dependent transactivation function and is implicated in multiple human abnormalities. Because R320C mutation replaces the charged polar arginine residue with neutral cysteine and results in the loss of a favorable interaction with the carboxylic group of its ligand triiodothyronine (T3), Koh *et al.* synthesized a neutral polar analogue HY1 that reestablished interaction with the mutant allele with improved binding (Figure 11B).[134] In a transactivation assay, HY1 was capable of activating the mutant allele of TRβ. In a related work aimed at compensating for the loss of activity from H435Y mutation in TRβ (identified in pituitary carcinoma resistant to thyroid hormone), the Koh group developed ligand analogues that displayed hydrogen bonding complementarity with the mutant.[135] The phenolic group of the mutant Y435 and the nitrogen atom of QH13, a pyridinium analogue of T3, reestablished the lost hydrogen bond between H435 and the phenolic group of T3. QH13 showed 15-fold selectivity and modest potency toward the mutant receptor and may potentially rescue the tumor suppressor properties of thyroid receptor. Such complementarity-based recovery of mutant activity has been demonstrated to reverse the antiandrogen resistance activity of androgen receptor mutants.[136]

Allele-specific chemical genetics has been employed to reveal underlying mechanisms of drugs targeting receptor molecules. For example, detailed biochemical studies carried out by Gronemeyer and colleagues revealed how a glycine in position 722 in human progesterone receptor (hPR) is critical to binding of RU486, an abortifacient drug.[137] G722 is a part of the "11β-pocket" that provides enough space to accommodate bulky 4-N,N-dimethylphenyl group in 11β-substituted RU486. Glycine to cysteine mutation (G722C) lowers the binding of drug by 40 000-fold, consistent with the observation that chicken and hamster PRs that carry cysteine in comparable position do not bind the drug. Steric complementarity was clearly demonstrated using a series of 11β-substituted RU486 analogues.[137] Consistent with these observations, a mutation in position 722 on hPR was considered the most likely reason for the failure to respond to RU486 by approximately 1% of women.[138]

In a similar application, ASCG has been effective in confirming the polyspecific role of Gleevec in protein kinase inhibition.[139] Gleevec is an approved drug for chronic myelogenic leukemia (CML) and B-cell lymphoblastic leukemia (BCL).[140] Gleevec acts through the inhibition of the protein kinase BCR-ABL; however, off-target activities against KIT and potentially other protein kinases are hypothesized to be a contributing factor to its efficacy. A study led by Witte and colleagues demonstrated that sole inhibition of a "hole-modified" T315A mutant of BCR-ABL by "bumped" inhibitor NA-PP1 was as effective as Gleevec in mature myeloid cells and Pro/Pre B cells that do not express KIT.[139] However, NA-PP1 remained inactive in Kit-positive progenitors, confirming the polypharmacology displayed by Gleevec for effective cancer treatment.

Smith *et al.* applied an allele-specific approach in an antibody-directed enzyme prodrug therapy (ADEPT) to minimize the immunogenicity associated with using a foreign enzyme that unmasks the prodrug (Figure 11C).[141] In their work, Smith *et al.* synthesized methotrexate (MTX)-based prodrugs joined by a peptide linkage with bumped side chains such as cyclopentyl tyrosine. These "bumped" peptides can only be cleaved by "hole-bearing" T268G mutant of human carboxypeptidase A1 (hCPA1), and not by any wild type protease present in human cells, to release MTX (Figure 11C). By conjugating an hCPA1 mutant to a tumor specific antibody, HT-29 human colon

dx.doi.org/10.1021/cb500651d | *ACS Chem. Biol.* 2015, 10, 343−363

adenocarcinoma cells were effectively killed upon addition of the allele-specific prodrug.[141]

Pharmacological agents capable of improving the activity of an otherwise inactive kinase mutant found in human disease would be of paramount importance.[142] In a recent study, Shokat and co-workers developed an ATP analogue (neo-substrate) that can selectively activate mutant PINK1 kinase (neoenzyme) implicated in Parkinson's Disease (PD) (Figure 11D).[143] The activity of PINK1 is essential for Mfn2 phosphorylation that regulates the function of Parkin.[144] Parkin is an E3 ubiquitin ligase responsible for the degradation of damaged mitochondria. Inactivity of PINK1 kinase arises due to a G309D mutation in the inserted loop in the active site. Based on sequence analysis and prior experience from their own extensive work on protein kinases, Shokat and colleagues developed N6-furfuryl ATP (Kinetin ATP, KTP) as an efficient selective cofactor for mutant PINK1 (Figure 11D). "Bumped" ATP analogues thus far have found limited applications in studying protein kinases and motor proteins *in vivo* because of their inability to cross cell membranes. However, quite remarkably, the authors demonstrated that KTP can be enzymatically synthesized in human cells fed with N6-furfuryl to probe the *in vivo* activity of PINK1 in the context of PD (Figure 11D).[143] Shokat's work represents an unprecedented pharmacological approach toward developing neosubstrates for activating mutant kinases as opposed to the more traditional approach that relies on identifying inhibitors for kinase activity.

SUMMARY AND FUTURE DIRECTIONS

Studies over the past three decades have convincingly demonstrated that engineering small molecule−protein interfaces is a powerful method for elucidating protein function. Combining organic synthesis with protein engineering to create protein−small molecule pairs functionally orthogonal to wild-type systems is the central tenet of allele-specific chemical genetics. From its early foundations in converting a GTPase to an XTPase to the very development of KTP as a "cellular ATP" for mutant PINK1, a large and diverse set of engineered proteins (enzymes, receptors, etc.) and their small-molecule counterparts have been successfully implemented to address diverse biological problems. The ability to both manipulate molecular interactions at a protein−ligand interface with the precision of genetics as well as craft small molecules fine-tuned for specific interactions is what makes ASCG so successful.

Base-resolution sequencing of the human genome is only the first step in understanding how the instructions coded in DNA lead to a functioning human being. The challenge in deriving meaningful knowledge from DNA sequence is to assign biological functions of each encoded protein in an isoform-specific manner. If the success of the techniques described above is any indication, then ASCG is certainly well poised to make major contributions toward understanding protein functions in eukaryotic biology and human disease. State of the art techniques in structural biology used in characterizing enzymes, receptors, and multimeric protein complexes will further accelerate progress designing engineered protein−ligand interfaces for biological evaluation. Another exciting direction is the merging of ASCG, activity-based protein profiling (ABPP),[145] and quantitative proteomics for allele-specific detection, imaging, isolation, and identification of substrates/binding partners. Efforts in this direction have already yielded a wealth of information, particularly through the identification of novel substrates for protein kinases, methyltransferases, acetyltransferases, and other enzymes. These studies have opened new avenues for further biological exploration.

In recent years, synthetic biology has shown great promise in harnessing the synthetic power of chemistry as well as biology to develop a range of important materials such as pharmaceuticals and biofuels.[146] Being inherently a rational protein engineering approach, ASCG has not been fully explored in the context of synthetic biology. However, efforts in this direction have begun and a number of studies are underway to create designed molecules with required chemical, physical, and medicinal properties.[147]

Molecular understanding of epigenetics-mediated gene regulation in differentiation and development of multicellular organisms has recently been an active area of research.[148] A large ensemble of enzymatic and nonenzymatic proteins are involved in writing, reading, and erasing chemical marks in chromatin landscape and paly essential role to control gene transcription.[149] Application of ASCG to epigenetic enzymes such as PMTs and KATs has already demonstrated its potential. In a recent work by Ciulli and co-workers, successful development of "bumped" I-BET (inhibitor of bromodomain and extra terminal protein) for the engineered (hole-modified) epigenetic reader BRD4 (bromodomain containing protein 4) further indicates that epigenetics will remain a fertile ground for ASCG in deconvoluting how chromatin-templated chemistry regulates human biology and disease.[150]

Although the pharmacological relevance of ASCG in "real world" drug design is questionable,[29] as target proteins cannot be conveniently modified, recent success in activating otherwise inactive receptor and kinase mutants found in human diseases by designed molecules underscores the therapeutic potential of the technique. Development of specific small-molecule inhibitors of neomorphic activity of an oncogenic IDH mutant allele further illustrates such potential.[151−153] Human genome sequencing is making great strides in revealing the genetic basis of diseases by finding rare mutations, many of which change shape complementarity, charge distribution, and affinity at the protein−ligand interface. Such information creates an unparalleled opportunity to employ ASCG to develop highly selective modulators for these "naturally" occurring mutant alleles for effective disease treatment. A parallel approach, termed "allele-specific RNAi", has recently been developed to selectively target the mutant transcript in human diseases.[154,155] The availability of a diverse repertoire of small molecules and the high spatiotemporal resolution make ASCG a superior technology for probing disease-causing mutations.

Advances in structural biology, molecular biology and organic synthesis have made ASCG a sought-after technology that can delineate protein functions in the context of fast, dynamic cellular events, a task difficult to accomplish using any other existing strategies. Discoveries at the frontier of biology engender new questions to address and fuel the development of innovative technologies such as ASCG, which in turn contribute to uncovering the molecular blueprint of life.

AUTHOR INFORMATION

Corresponding Author

*Phone: 412-624-5061. E-mail: kai27@pitt.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The author thanks Prof. S. Horne and D. Chakraborty for the critical reading and editing of the manuscript. The author thanks all the reviewers for their critical comments and suggestions. The financial support from the University of Pittsburgh, the Central Research and Development Fund, and the Samuel and Emma Winters Foundation is gratefully acknowledged. The author sincerely apologizes to colleagues whose work could not be discussed because of the limited space.

■ REFERENCES

(1) International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature 431*, 931−945.
(2) Matlin, A. J., Clark, F., and Smith, C. W. (2005) Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol. 6*, 386−398.
(3) Walsh, C. (2006) *Posttranslational Modification of Proteins: Expanding Nature's Inventory*, Roberts and Co. Publishers, Englewood, CO.
(4) Leipe, D. D., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol. 317*, 41−72.
(5) Bjarnadottir, T. K., Gloriam, D. E., Hellstrand, S. H., Kristiansson, H., Fredriksson, R., and Schioth, H. B. (2006) Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics 88*, 263−273.
(6) Lampson, M. A., and Kapoor, T. M. (2006) Targeting protein stability with a small molecule. *Cell 126*, 827−9.
(7) Lampson, M. A., and Kapoor, T. M. (2006) Unraveling cell division mechanisms with small-molecule inhibitors. *Nat. Chem. Biol. 2*, 19−27.
(8) Cole, P. A. (2008) Chemical probes for histone-modifying enzymes. *Nat. Chem. Biol. 4*, 590−597.
(9) Lyssiotis, C. A., Lairson, L. L., Boitano, A. E., Wurdak, H., Zhu, S., and Schultz, P. G. (2011) Chemical control of stem cell fate and developmental potential. *Angew. Chem., Int. Ed. Engl. 50*, 200−242.
(10) Schreiber, S. L. (1998) Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem. 6*, 1127−1152.
(11) Walsh, D. P., and Chang, Y. T. (2006) Chemical genetics. *Chem. Rev. 106*, 2476−2530.
(12) Shah, K. (2005) Orthogonal chemical genetic approaches for unraveling signaling pathways. *IUBMB Life 57*, 397−405.
(13) Shokat, K., and Velleca, M. (2002) Novel chemical genetic approaches to the discovery of signal transduction inhibitors. *Drug Discovery Today 7*, 872−879.
(14) Hwang, Y. W., and Miller, D. L. (1987) A mutation that alters the nucleotide specificity of elongation factor Tu, a GTP regulatory protein. *J. Biol. Chem. 262*, 13081−13085.
(15) Belshaw, P. J., Schoepfer, J. G., Liu, K.-Q., Morrison, K. L., and Schreiber, S. L. (1995) Rational design of orthogonal receptor−ligand combinations. *Angew. Chem., Int. Ed. Engl. 34*, 2129−2132.
(16) Koh, J. T. (2002) Engineering selectivity and discrimination into ligand−receptor interfaces. *Chem. Biol. 9*, 17−23.
(17) Clarke, A. R., Atkinson, T., and Holbrook, J. J. (1989) From analysis to synthesis: New ligand binding sites on the lactate dehydrogenase framework. Part I. *Trends Biochem. Sci. 14*, 101−105.
(18) Bocanegra, J. A., Scrutton, N. S., and Perham, R. N. (1993) Creation of an NADP-dependent pyruvate dehydrogenase multienzyme complex by protein engineering. *Biochem. 32*, 2737−2740.
(19) Scrutton, N. S., Berry, A., and Perham, R. N. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature 343*, 38−43.
(20) Tairis, N., Gabriel, J. L., Soprano, K. J., and Soprano, D. R. (1995) Alteration in the retinoid specificity of retinoic acid receptor-β by site-directed mutagenesis of Arg269 and Lys220. *J. Biol. Chem. 270*, 18380−18387.
(21) Conklin, B. R., Hsiao, E. C., Claeysen, S., Dumuis, A., Srinivasan, S., Forsayeth, J. R., Guettier, J. M., Chang, W. C., Pei, Y., McCarthy, K. D., Nissenson, R. A., Wess, J., Bockaert, J., and Roth, B. L. (2008) Engineering GPCR signaling pathways with RASSLs. *Nat. Methods 5*, 673−678.
(22) Shapiro, M. G., Frazier, S. J., and Lester, H. A. (2012) Unparalleled control of neural activity using orthogonal pharmacogenetics. *ACS Chem. Neurosci. 3*, 619−629.
(23) Strader, C. D., Gaffney, T., Sugg, E. E., Candelore, M. R., Keys, R., Patchett, A. A., and Dixon, R. A. (1991) Allele-specific activation of genetically engineered receptors. *J. Biol. Chem. 266*, 5−8.
(24) Jacobson, K. A., Gao, Z. G., and Liang, B. T. (2007) Neoceptors: Reengineering GPCRs to recognize tailored ligands. *Trends Pharmacol. Sci. 28*, 111−116.
(25) Wess, J., Nakajima, K., and Jain, S. (2013) Novel designer receptors to probe GPCR signaling and physiology. *Trends Pharmacol. Sci. 34*, 385−392.
(26) Dong, S., Allen, J. A., Farrell, M., and Roth, B. L. (2010) A chemical-genetic approach for precise spatio-temporal control of cellular signaling. *Mol. Biosyst. 6*, 1376−1380.
(27) Azzarito, V., Long, K., Murphy, N. S., and Wilson, A. J. (2013) Inhibition of α-helix-mediated protein−protein interactions using designed molecules. *Nat. Chem. 5*, 161−173.
(28) Belshaw, P. J., and Schreiber, S. L. (1997) Cell-specific calcineurin inhibition by a modified cyclosporin. *J. Am. Chem. Soc. 119*, 1805−1806.
(29) Clackson, T. (1998) Redesigning small molecule−protein interfaces. *Curr. Opin. Struct. Biol. 8*, 451−458.
(30) Liberles, S. D., Diver, S. T., Austin, D. J., and Schreiber, S. L. (1997) Inducible gene expression and protein translocation using nontoxic ligands identified by a mammalian three-hybrid screen. *Proc. Natl. Acad. Sci. U.S.A. 94*, 7825−7830.
(31) Crick, F. H. (1952) Is α-keratin a coiled coil? *Nature 170*, 882−883.
(32) Ridgway, J. B., Presta, L. G., and Carter, P. (1996) 'Knobs-into-holes' engineering of antibody CH3 domains for heavy chain heterodimerization. *Protein Eng. 9*, 617−621.
(33) Bishop, A. C., Buzko, O., and Shokat, K. M. (2001) Magic bullets for protein kinases. *Trends Cell Biol. 11*, 167−72.
(34) Koh, J. T., Putnam, M., Tomic-Canic, M., and McDaniel, C. M. (1999) Selective regulation of gene expression using rationally-modified retinoic acid receptors. *J. Am. Chem. Soc. 121*, 1984−1985.
(35) Tedesco, R., Thomas, J. A., Katzenellenbogen, B. S., and Katzenellenbogen, J. A. (2001) The estrogen receptor: A structure-based approach to the design of new specific hormone−receptor combinations. *Chem. Biol. 8*, 277−287.
(36) Shi, Y., and Koh, J. T. (2001) Selective regulation of gene expression by an orthogonal estrogen receptor−ligand pair created by polar-group exchange. *Chem. Biol. 8*, 501−510.
(37) Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996) Ligand-activated site-specific recombination in mice. *Proc. Natl. Acad. Sci. U.S.A. 93*, 10887−10890.
(38) Feil, R., Wagner, J., Metzger, D., and Chambon, P. (1997) Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun. 237*, 752−757.
(39) Doyle, D. F., Braasch, D. A., Jackson, L. K., Weiss, H. E., Boehm, M. F., Mangelsdorf, D. J., and Corey, D. R. (2001) Engineering orthogonal ligand−receptor pairs from "near drugs". *J. Am. Chem. Soc. 123*, 11367−11371.
(40) Guo, Z., Zhou, D., and Schultz, P. G. (2000) Designing small-molecule switches for protein−protein interactions. *Science 288*, 2042−2045.
(41) Clackson, T., and Wells, J. A. (1995) A hot spot of binding energy in a hormone−receptor interface. *Science 267*, 383−386.

(42) Lin, Q., Barbas, C. F., 3rd, and Schultz, P. G. (2003) Small-molecule switches for zinc finger transcription factors. *J. Am. Chem. Soc. 125*, 612−613.
(43) Erlanson, D. A., Braisted, A. C., Raphael, D. R., Randal, M., Stroud, R. M., Gordon, E. M., and Wells, J. A. (2000) Site-directed ligand discovery. *Proc. Natl. Acad. Sci. U.S.A. 97*, 9367−9372.
(44) Levitsky, K., Ciolli, C. J., and Belshaw, P. J. (2003) Selective inhibition of engineered receptors via proximity-accelerated alkylation. *Org. Lett. 5*, 693−696.
(45) Krusemark, C. J., and Belshaw, P. J. (2007) Covalent labelling of fusion proteins in live cells via an engineered receptor−ligand pair. *Org. Biomol. Chem. 5*, 2201−2204.
(46) Cohen, M. S., Zhang, C., Shokat, K. M., and Taunton, J. (2005) Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science 308*, 1318−1321.
(47) Garske, A. L., Peters, U., Cortesi, A. T., Perez, J. L., Shokat, K. M., Chemical genetic strategy for targeting protein kinases based on covalent complementarity. *Proc. Natl. Acad. Sci. U.S.A. 108*, 15046−15052.
(48) Morell, M., Nguyen Duc, T., Willis, A. L., Syed, S., Lee, J., Deu, E., Deng, Y., Xiao, J., Turk, B. E., Jessen, J. R., Weiss, S. J., and Bogyo, M. (2013) Coupling protein engineering with probe design to inhibit and image matrix metalloproteinases with controlled specificity. *J. Am. Chem. Soc. 135*, 9139−9148.
(49) Xiao, J., Broz, P., Puri, A. W., Deu, E., Morell, M., Monack, D. M., and Bogyo, M. (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.
(50) Chen, Z., Jing, C., Gallagher, S. S., Sheetz, M. P., and Cornish, V. W. (2012) Second-generation covalent TMP-tag for live cell imaging. *J. Am. Chem. Soc. 134*, 13692−13699.
(51) Blair, J. A., Rauh, D., Kung, C., Yun, C. H., Fan, Q. W., Rode, H., Zhang, C., Eck, M. J., Weiss, W. A., and Shokat, K. M. (2007) Structure-guided development of affinity probes for tyrosine kinases using chemical genetics. *Nat. Chem. Biol. 3*, 229−238.
(52) Koch, A., Rode, H. B., Richters, A., Rauh, D., and Hauf, S. (2012) A chemical genetic approach for covalent inhibition of analogue-sensitive aurora kinase. *ACS Che. Biol. 7*, 723−731.
(53) Banaszynski, L. A., Chen, L. C., Maynard-Smith, L. A., Ooi, A. G., and Wandless, T. J. (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell 126*, 995−1004.
(54) Stankunas, K., Bayle, J. H., Gestwicki, J. E., Lin, Y. M., Wandless, T. J., and Crabtree, G. R. (2003) Conditional protein alleles using knockin mice and a chemical inducer of dimerization. *Mol. Cell 12*, 1615−1624.
(55) Clackson, T., Yang, W., Rozamus, L. W., Hatada, M., Amara, J. F., Rollins, C. T., Stevenson, L. F., Magari, S. R., Wood, S. A., Courage, N. L., Lu, X., Cerasoli, F., Jr., Gilman, M., and Holt, D. A. (1998) Redesigning an FKBP−ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci. U.S.A. 95*, 10437−10442.
(56) Rivera, V. M., Wang, X., Wardwell, S., Courage, N. L., Volchuk, A., Keenan, T., Holt, D. A., Gilman, M., Orci, L., Cerasoli, F., Jr., Rothman, J. E., and Clackson, T. (2000) Regulation of protein secretion through controlled aggregation in the endoplasmic reticulum. *Science 287*, 826−830.
(57) Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D., and Wandless, T. J. (2010) A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol. 17*, 981−988.
(58) Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem. 79*, 413−444.
(59) Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of *Escherichia coli*. *Science 292*, 498−500.
(60) Xie, J., and Schultz, P. G. (2006) A chemical toolkit for proteins—An expanded genetic code. *Nat. Rev. Mol. Cell Biol. 7*, 775−782.
(61) Baker, A. S., and Deiters, A. (2014) Optical control of protein function through unnatural amino acid mutagenesis and other optogenetic approaches. *ACS Chem. Biol. 9*, 1398−1407.
(62) Pham, N. D., Parker, R. B., and Kohler, J. J. (2013) Photocrosslinking approaches to interactome mapping. *Curr. Opinion Chem. Biol. 17*, 90−101.
(63) Zhang, C., Lopez, M. S., Dar, A. C., Ladow, E., Finkbeiner, S., Yun, C. H., Eck, M. J., and Shokat, K. M. (2013) Structure-guided inhibitor design expands the scope of analog-sensitive kinase technology. *ACS Chem. Biol. 8*, 1931−1938.
(64) Bishop, A. C., Kung, C.-Y., Shah, K., Witucki, L., Shokat, K. M., and Liu, Y. (1999) Generation of monospecific nanomolar tyrosine kinase inhibitors via a chemical genetic approach. *J. Am. Chem. Soc. 121*, 627−631.
(65) Lopez, M. S., Choy, J. W., Peters, U., Sos, M. L., Morgan, D. O., and Shokat, K. M. (2013) Staurosporine-derived inhibitors broaden the scope of analog-sensitive kinase technology. *J. Am. Chem. Soc. 135*, 18153−18159.
(66) Wang, H., Shimizu, E., Tang, Y. P., Cho, M., Kyin, M., Zuo, W., Robinson, D. A., Alaimo, P. J., Zhang, C., Morimoto, H., Zhuo, M., Feng, R., Shokat, K. M., and Tsien, J. Z. (2003) Inducible protein knockout reveals temporal requirement of CaMKII reactivation for memory consolidation in the brain. *Proc. Natl. Acad. Sci. U.S.A. 100*, 4287−4292.
(67) Bishop, A. C., Ubersax, J. A., Petsch, D. T., Matheos, D. P., Gray, N. S., Blethrow, J., Shimizu, E., Tsien, J. Z., Schultz, P. G., Rose, M. D., Wood, J. L., Morgan, D. O., and Shokat, K. M. (2000) A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature 407*, 395−401.
(68) Reed, S. I. (1980) The selection of *S. cerevisiae* mutants defective in the start event of cell division. *Genetics 95*, 561−577.
(69) Carroll, A. S., Bishop, A. C., DeRisi, J. L., Shokat, K. M., and O'Shea, E. K. (2001) Chemical inhibition of the Pho85 cyclin-dependent kinase reveals a role in the environmental stress response. *Proc. Natl. Acad. Sci. U.S.A. 98*, 12578−12583.
(70) Papa, F. R., Zhang, C., Shokat, K., and Walter, P. (2003) Bypassing a kinase activity with an ATP-competitive drug. *Science 302*, 1533−1537.
(71) Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science 298*, 1912−1934.
(72) Goel, R., Harsha, H. C., Pandey, A., and Prasad, T. S. (2012) Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. *Mol. Biosyst. 8*, 453−463.
(73) Shah, K., Liu, Y., Deirmengian, C., and Shokat, K. M. (1997) Engineering unnatural nucleotide specificity for *Rous sarcoma* virus tyrosine kinase to uniquely label its direct substrates. *Proc. Natl. Acad. Sci. U.S.A. 94*, 3565−3570.
(74) Ubersax, J. A., Woodbury, E. L., Quang, P. N., Paraz, M., Blethrow, J. D., Shah, K., Shokat, K. M., and Morgan, D. O. (2003) Targets of the cyclin-dependent kinase Cdk1. *Nature 425*, 859−864.
(75) Blethrow, J. D., Glavy, J. S., Morgan, D. O., and Shokat, K. M. (2008) Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc. Natl. Acad. Sci. U.S.A. 105*, 1442−1447.
(76) Berg, J. S., Powell, B. C., and Cheney, R. E. (2001) A millennial myosin census. *Mol. Biol. Cell 12*, 780−794.
(77) Holt, J. R., Gillespie, S. K., Provance, D. W., Shah, K., Shokat, K. M., Corey, D. P., Mercer, J. A., and Gillespie, P. G. (2002) A chemical-genetic strategy implicates myosin-1c in adaptation by hair cells. *Cell 108*, 371−381.
(78) Gillespie, P. G., Gillespie, S. K., Mercer, J. A., Shah, K., and Shokat, K. M. (1999) Engineering of the myosin-iβ nucleotide-binding pocket to create selective sensitivity to N(6)-modified ADP analogs. *J. Biol. Chem. 274*, 31373−31381.
(79) Provance, D. W., Jr., Gourley, C. R., Silan, C. M., Cameron, L. C., Shokat, K. M., Goldenring, J. R., Shah, K., Gillespie, P. G., and Mercer, J. A. (2004) Chemical-genetic inhibition of a sensitized mutant myosin Vb demonstrates a role in peripheral−pericentriolar membrane traffic. *Proc. Natl. Acad. Sci. U.S.A. 101*, 1868−1873.

(80) Wang, Z., Edwards, J. G., Riley, N., Provance, D. W., Jr., Karcher, R., Li, X. D., Davison, I. G., Ikebe, M., Mercer, J. A., Kauer, J. A., and Ehlers, M. D. (2008) Myosin Vb mobilizes recycling endosomes and AMPA receptors for postsynaptic plasticity. *Cell 135*, 535−548.
(81) Kapoor, T. M., and Mitchison, T. J. (1999) Allele-specific activators and inhibitors for kinesin. *Proc. Natl. Acad. Sci. U.S.A. 96*, 9106−9111.
(82) Chapman, E., Farr, G. W., Furtak, K., and Horwich, A. L. (2009) A small molecule inhibitor selective for a variant ATP-binding site of the chaperonin GroEL. *Bioorg. Med. Chem. Lett. 19*, 811−813.
(83) Vincent, F., Cook, S. P., Johnson, E. O., Emmert, D., and Shah, K. (2007) Engineering unnatural nucleotide specificity to probe G protein signaling. *Chem. Biol. 14*, 1007−1018.
(84) Luo, M. (2012) Current chemical biology approaches to interrogate protein methyltransferases. *ACS Chem. Biol. 7*, 443−463.
(85) Erce, M. A., Pang, C. N., Hart-Smith, G., and Wilkins, M. R. (2012) The methylproteome and the intracellular methylation network. *Proteomics 12*, 564−586.
(86) Zhang, X., Wen, H., and Shi, X. (2012) Lysine methylation: Beyond histones. *Acta Biochim. Biophys. Sin. 44*, 14−27.
(87) Lin, Q., Jiang, F., Schultz, P. G., and Gray, N. S. (2001) Design of allele-specific protein methyltransferase inhibitors. *J. Am. Chem. Soc. 123*, 11608−11613.
(88) Dalhoff, C., Lukinavicius, G., Klimasauskas, S., and Weinhold, E. (2006) Direct transfer of extended groups from synthetic cofactors by DNA methyltransferases. *Nat. Chem. Biol. 2*, 31−32.
(89) Klimasauskas, S., and Weinhold, E. (2007) A new tool for biotechnology: AdoMet-dependent methyltransferases. *Trends Biotechnol. 25*, 99−104.
(90) Lukinavicius, G., Lapiene, V., Stasevskij, Z., Dalhoff, C., Weinhold, E., and Klimasauskas, S. (2007) Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). *J. Am. Chem. Soc. 129*, 2758−2759.
(91) Lukinavicius, G., Lapinaite, A., Urbanaviciute, G., Gerasimaite, R., and Klimasauskas, S. (2012) Engineering the DNA cytosine-5 methyltransferase reaction for sequence-specific labeling of DNA. *Nucleic Acids Res. 40*, 11594−11602.
(92) Peters, W., Willnow, S., Duisken, M., Kleine, H., Macherey, T., Duncan, K. E., Litchfield, D. W., Luscher, B., and Weinhold, E. (2010) Enzymatic site-specific functionalization of protein methyltransferase substrates with alkynes for click labeling. *Angew. Chem., Int. Ed. Engl. 49*, 5170−5173.
(93) Hein, J. E., and Fokin, V. V. (2010) Copper-catalyzed azide−alkyne cycloaddition (CuAAC) and beyond: New reactivity of copper(I) acetylides. *Chem. Soc. Rev. 39*, 1302−1315.
(94) Binda, O., Boyce, M., Rush, J. S., Palaniappan, K. K., Bertozzi, C. R., and Gozani, O. (2011) A chemical method for labeling lysine methyltransferase substrates. *ChemBioChem 12*, 330−334.
(95) Willnow, S., Martin, M., Luscher, B., and Weinhold, E. (2012) A selenium-based click AdoMet analogue for versatile substrate labeling with wild-type protein methyltransferases. *ChemBioChem. 13*, 1167−1173.
(96) Bothwell, I. R., Islam, K., Chen, Y., Zheng, W., Blum, G., Deng, H., and Luo, M. (2012) Se-adenosyl-L-selenomethionine cofactor analogue as a reporter of protein methylation. *J. Am. Chem. Soc. 134*, 14905−14912.
(97) Stecher, H., Tengg, M., Ueberbacher, B. J., Remler, P., Schwab, H., Griengl, H., and Gruber-Khadjawi, M. (2009) Biocatalytic Friedel−Crafts alkylation using non-natural cofactors. *Angew. Chem., Int. Ed. Engl. 48*, 9546−9548.
(98) Lee, B. W., Sun, H. G., Zang, T., Kim, B. J., Alfaro, J. F., and Zhou, Z. S. (2010) Enzyme-catalyzed transfer of a ketone group from an S-adenosylmethionine analogue: A tool for the functional analysis of methyltransferases. *J. Am. Chem. Soc. 132*, 3642−3643.
(99) Motorin, Y., Burhenne, J., Teimer, R., Koynov, K., Willnow, S., Weinhold, E., and Helm, M. (2011) Expanding the chemical scope of RNA/methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic Acids Res. 39*, 1943−1952.
(100) Winter, J. M., Chiou, G., Bothwell, I. R., Xu, W., Garg, N. K., Luo, M., and Tang, Y. (2013) Expanding the structural diversity of polyketides by exploring the cofactor tolerance of an inline methyltransferase domain. *Org. Lett. 15*, 3774−3777.
(101) Wang, R., and Luo, M. (2013) A journey toward bioorthogonal profiling of protein methylation inside living cells. *Curr. Opin. Chem. Biol. 17*, 729−737.
(102) Islam, K., Zheng, W., Yu, H., Deng, H., and Luo, M. (2011) Expanding cofactor repertoire of protein lysine methyltransferase for substrate labeling. *ACS Chem. Biol. 6*, 679−684.
(103) Wang, R., Zheng, W., Yu, H., Deng, H., and Luo, M. (2011) Labeling substrates of protein arginine methyltransferase with engineered enzymes and matched S-adenosyl-L-methionine analogues. *J. Am. Chem. Chem. 133*, 7648−7651.
(104) Islam, K., Bothwell, I., Chen, Y., Sengelaub, C., Wang, R., Deng, H., and Luo, M. (2012) Bioorthogonal profiling of protein methylation using azido derivative of S-adenosyl-L-methionine. *J. Am. Chem. Chem. 134*, 5909−5915.
(105) Islam, K., Chen, Y., Wu, H., Bothwell, I. R., Blum, G. J., Zeng, H., Dong, A., Zheng, W., Min, J., Deng, H., and Luo, M. (2013) Defining efficient enzyme-cofactor pairs for bioorthogonal profiling of protein methylation. *Proc. Natl. Acad. Sci. U.S.A. 110*, 16778−16783.
(106) Guo, H., Wang, R., Zheng, W., Chen, Y., Blum, G., Deng, H., and Luo, M. (2014) Profiling substrates of protein arginine N-methyltransferase 3 with S-adenosyl-L-methionine analogues. *ACS Chem. Biol. 9*, 476−484.
(107) Liu, Y., Shah, K., Yang, F., Witucki, L., and Shokat, K. M. (1998) A molecular gate which controls unnatural ATP analogue recognition by the tyrosine kinase v-Src. *Bioorg. Med. Chem. 6*, 1219−1226.
(108) Wang, R., Islam, K., Liu, Y., Zheng, W., Tang, H., Lailler, N., Blum, G., Deng, H., and Luo, M. (2013) Profiling genome-wide chromatin methylation with engineered posttranslation apparatus within living cells. *J. Am. Chem. Chem. 135*, 1048−1056.
(109) Bothwell, I. R., and Luo, M. (2014) Large-scale, protection-free synthesis of Se-adenosyl-L-selenomethionine analogues and their application as cofactor surrogates of methyltransferases. *Org. Lett. 16*, 3056−3059.
(110) Li, J., Wei, H., and Zhou, M. M. (2011) Structure-guided design of a methyl donor cofactor that controls a viral histone H3 lysine 27 methyltransferase activity. *J. Med. Chem. 54*, 7734−7738.
(111) Smith, K. T., and Workman, J. L. (2009) Introducing the acetylome. *Nat. Biotechnol. 27*, 917−919.
(112) Kim, G. W., and Yang, X. J. (2011) Comprehensive lysine acetylomes emerging from bacteria to humans. *Trends Biochem. Sci. 36*, 211−220.
(113) Yang, Y. Y., Ascano, J. M., Hang, H. C. Bioorthogonal chemical reporters for monitoring protein acetylation. *J. Am. Chem. Chem. 132*, 3640−3641.
(114) Yang, C., Mi, J., Feng, Y., Ngo, L., Gao, T., Yan, L., and Zheng, Y. G. (2013) Labeling lysine acetyltransferase substrates with engineered enzymes and functionalized cofactor surrogates. *J. Am. Chem. Chem. 135*, 7791−7794.
(115) Carter-O'Connell, I., Jin, H., Morgan, R. K., David, L. L., and Cohen, M. S. (2014) Engineering the substrate specificity of ADP-ribosyltransferases for identifying direct protein targets. *J. Am. Chem. Chem. 136*, 5201−5204.
(116) Jiang, H., Kim, J. H., Frizzell, K. M., Kraus, W. L., and Lin, H. (2010) Clickable NAD analogues for labeling substrate proteins of poly(ADP-ribose) polymerases. *J. Am. Chem. Chem. 132*, 9363−9372.
(117) Feng, S., Zhang, L., Adilijiang, G., Liu, J., Luo, M., and Deng, H. (2014) Substrate profiling of glutathione S-transferase with engineered enzymes and matched glutathione analogues. *Angew. Chem., Int. Ed. Engl. 53*, 7149−7153.
(118) Samarasinghe, K. T., Munkanatta Godage, D. N., VanHecke, G. C., and Ahn, Y. H. (2014) Metabolic synthesis of clickable glutathione for chemoselective detection of glutathionylation. *J. Am. Chem. Chem. 136*, 11566−11569.

(119) Ramakrishnan, B., and Qasba, P. K. (2002) Structure-based design of β 1,4-galactosyltransferase I (β 4Gal-T1) with equally efficient N-acetylgalactosaminyltransferase activity: Point mutation broadens β 4Gal-T1 donor specificity. *J. Biol. Chem. 277*, 20833−20839.
(120) Khidekel, N., Arndt, S., Lamarre-Vincent, N., Lippert, A., Poulin-Kerstien, K. G., Ramakrishnan, B., Qasba, P. K., and Hsieh-Wilson, L. C. (2003) A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *J. Am. Chem. Chem. 125*, 16162−16163.
(121) Khidekel, N., Ficarro, S. B., Peters, E. C., and Hsieh-Wilson, L. C. (2004) Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain. *Proc. Natl. Acad. Sci. U.S.A. 101*, 13132−13137.
(122) Yu, S. H., Boyce, M., Wands, A. M., Bond, M. R., Bertozzi, C. R., and Kohler, J. J. (2012) Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. *Proc. Natl. Acad. Sci. U.S.A. 109*, 4834−4839.
(123) Andersen, J. N., Mortensen, O. H., Peters, G. H., Drake, P. G., Iversen, L. F., Olsen, O. H., Jansen, P. G., Andersen, H. S., Tonks, N. K., and Møller, N. P. (2001) Structural and evolutionary relationships among protein tyrosine phosphatase domains. *Mol. Cell. Biol. 21*, 7117−7136.
(124) Hoffman, H. E., Blair, E. R., Johndrow, J. E., and Bishop, A. C. (2005) Allele-specific inhibitors of protein tyrosine phosphatases. *J. Am. Chem. Soc. 127*, 2824−2825.
(125) Bishop, A. C., and Blair, E. R. (2006) A gatekeeper residue for inhibitor sensitization of protein tyrosine phosphatases. *Bioorg. Med. Chem. Lett. 16*, 4002−4006.
(126) Carter, P., and Wells, J. A. (1987) Engineering enzyme specificity by "substrate-assisted catalysis". *Science 237*, 394−399.
(127) Dall'Acqua, W., and Carter, P. (2000) Substrate-assisted catalysis: molecular basis and biological significance. *Protein Sci. 9*, 1−9.
(128) Bone, R., Silen, J. L., and Agard, D. A. (1989) Structural plasticity broadens the specificity of an engineered protease. *Nature 339*, 191−195.
(129) Bruysters, M., Jongejan, A., Akdemir, A., Bakker, R. A., and Leurs, R. (2005) A G(q/11)-coupled mutant histamine H(1) receptor F435A activated solely by synthetic ligands (RASSL). *J. Biol. Chem. 280*, 34741−34746.
(130) Magnus, C. J., Lee, P. H., Atasoy, D., Su, H. H., Looger, L. L., and Sternson, S. M. (2011) Chemical and genetic engineering of selective ion channel−ligand interactions. *Science 333*, 1292−1296.
(131) Simon, M. D., and Shokat, K. M. (2004) Adaptability at a protein-DNA interface: Re-engineering the engrailed homeodomain to recognize an unnatural nucleotide. *J. Am. Chem. Soc. 126*, 8078−8079.
(132) Simon, M. D., Feldman, M. E., Rauh, D., Maris, A. E., Wemmer, D. E., and Shokat, K. M. (2006) Structure and properties of a re-engineered homeodomain protein-DNA interface. *ACS Chem. Biol. 1*, 755−760.
(133) Cheng, J., Danielson, D. C., Nasheri, N., Singaravelu, R., and Pezacki, J. P. (2011) Enhanced specificity of the viral suppressor of RNA silencing protein p19 toward sequestering of human microRNA-122. *Biochem. 50*, 7745−7755.
(134) Ye, H. F., O'Reilly, K. E., and Koh, J. T. (2001) A subtype-selective thyromimetic designed to bind a mutant thyroid hormone receptor implicated in resistance to thyroid hormone. *J. Am. Chem. Soc. 123*, 1521−1522.
(135) Hassan, A. Q., and Koh, J. T. (2008) Selective chemical rescue of a thyroid-hormone-receptor mutant, TRβ(H435Y), identified in pituitary carcinoma and resistance to thyroid hormone. *Angew. Chem., Int. Ed. Engl. 47*, 7280−7283.
(136) McGinley, P. L., and Koh, J. T. (2007) Circumventing anti-androgen resistance by molecular design. *J. Am. Chem. Soc. 129*, 3822−3823.
(137) Benhamou, B., Garcia, T., Lerouge, T., Vergezac, A., Gofflo, D., Bigogne, C., Chambon, P., and Gronemeyer, H. (1992) A single amino acid that determines the sensitivity of progesterone receptors to RU486. *Science 255*, 206−209.
(138) Baulieu, E. E. (1989) Contragestion and other clinical applications of RU 486, an antiprogesterone at the receptor. *Science 245*, 1351−1357.
(139) Wong, S., McLaughlin, J., Cheng, D., Zhang, C., Shokat, K. M., and Witte, O. N. (2004) Sole BCR-ABL inhibition is insufficient to eliminate all myeloproliferative disorder cell populations. *Proc. Natl. Acad. Sci. U.S.A. 101*, 17456−17461.
(140) Wong, S., and Witte, O. N. (2004) The BCR-ABL story: Bench to bedside and back. *Annu. Rev. Immunol. 22*, 247−306.
(141) Smith, G. K., Banks, S., Blumenkopf, T. A., Cory, M., Humphreys, J., Laethem, R. M., Miller, J., Moxham, C. P., Mullin, R., Ray, P. H., Walton, L. M., and Wolfe, L. A., 3rd (1997) Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel *in vivo* stable prodrugs of methotrexate. *J. Biol. Chem. 272*, 15804−15816.
(142) Kleiner, R. E., and Kapoor, T. M. (2013) Bumping up kinase activity with an ATP-derived neo-substrate. *Cell 154*, 716−718.
(143) Hertz, N. T., Berthet, A., Sos, M. L., Thorn, K. S., Burlingame, A. L., Nakamura, K., and Shokat, K. M. (2013) A neo-substrate that amplifies catalytic activity of Parkinson's-disease-related kinase PINK1. *Cell 154*, 737−747.
(144) Narendra, D. P., Jin, S. M., Tanaka, A., Suen, D. F., Gautier, C. A., Shen, J., Cookson, M. R., and Youle, R. J. (2010) PINK1 is selectively stabilized on impaired mitochondria to activate Parkin. *PLoS Biol. 8*, e1000298.
(145) Cravatt, B. F., Wright, A. T., and Kozarich, J. W. (2008) Activity-based protein profiling: From enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem. 77*, 383−414.
(146) Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014) A brief history of synthetic biology. *Nat. Rev. Microbiol. 12*, 381−390.
(147) Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., and Lee, S. Y. (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol. 8*, 536−546.
(148) Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: A landscape takes shape. *Cell 128*, 635−638.
(149) Kouzarides, T. (2007) Chromatin modifications and their function. *Cell 128*, 693−705.
(150) Baud, M. G., Lin-Shiao, E., Cardote, T., Tallant, C., Pschibul, A., Chan, K. H., Zengerle, M., Garcia, J. R., Kwan, T. T., Ferguson, F. M., and Ciulli, A. (2014) A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. *Science 346*, 638−641.
(151) Rohle, D., Popovici-Muller, J., Palaskas, N., Turcan, S., Grommes, C., Campos, C., Tsoi, J., Clark, O., Oldrini, B., Komisopoulou, E., Kunii, K., Pedraza, A., Schalm, S., Silverman, L., Miller, A., Wang, F., Yang, H., Chen, Y., Kernytsky, A., Rosenblum, M. K., Liu, W., Biller, S. A., Su, S. M., Brennan, C. W., Chan, T. A., Graeber, T. G., Yen, K. E., and Mellinghoff, I. K. (2013) An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science 340*, 626−630.
(152) Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., Straley, K., Kernytsky, A., Liu, W., Gliser, C., Yang, H., Gross, S., Artin, E., Saada, V., Mylonas, E., Quivoron, C., Popovici-Muller, J., Saunders, J. O., Salituro, F. G., Yan, S., Murray, S., Wei, W., Gao, Y., Dang, L., Dorsch, M., Agresta, S., Schenkein, D. P., Biller, S. A., Su, S. M., de Botton, S., and Yen, K. E. (2013) Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science 340*, 622−626.
(153) Zheng, B., Yao, Y., Liu, Z., Deng, L., Anglin, J. L., Jiang, H., Prasad, B. V., and Song, Y. (2013) Crystallographic investigation and selective inhibition of mutant isocitrate dehydrogenase. *ACS Med. Chem. Lett. 4*, 542−546.
(154) Hohjoh, H. (2013) Disease-causing allele-specific silencing by RNA interference. *Pharmaceuticals 6*, 522−535.
(155) Jiang, J., Wakimoto, H., Seidman, J. G., and Seidman, C. E. (2013) Allele-specific silencing of mutant Myh6 transcripts in mice suppresses hypertrophic cardiomyopathy. *Science 342*, 111−114.