



NPR

Chemical proteomics approaches for identifying the cellular targets of natural products

| | |
|-------------------------------|--|
| Journal: | <i>Natural Product Reports</i> |
| Manuscript ID | NP-REV-01-2016-000001.R1 |
| Article Type: | Review Article |
| Date Submitted by the Author: | 07-Apr-2016 |
| Complete List of Authors: | Wright, Megan; Technische Universitat Munchen Fakultat fur Chemie Sieber, Stephan; Technische Universitat Munchen Fakultat fur Chemie |
| | |

SCHOLARONE™
Manuscripts

Chemical proteomics approaches for identifying the cellular targets of natural products

M. H. Wright,^a and S. A. S. Sieber^a

a. Department of Chemistry, Technische Universität München, Lichtenbergstraße 4, 85748, Garching, Germany

Deconvoluting the mode of action of natural products and drugs remains one of the biggest challenges in chemistry and biology today. Chemical proteomics is a growing area of chemical biology that seeks to design small molecule probes to understand protein function. In the context of natural products, chemical proteomics can be used to identify the protein binding partners or targets of small molecules in live cells. Here, we highlight recent examples of chemical probes based on natural products and their application for target identification. The review focuses on probes that can be covalently linked to their target proteins (either via intrinsic chemical reactivity or via the introduction of photocrosslinkers), and can be applied “in situ” – in living systems rather than cell lysates. We also focus here on strategies that employ a click reaction, the copper-catalysed azide-alkyne cycloaddition reaction (CuAAC), to allow minimal functionalisation of natural product scaffolds with an alkyne or azide tag. We also discuss ‘competitive mode’ approaches that screen for natural products that compete with a well-characterised chemical probe for binding to a particular set of protein targets. Fuelled by advances in mass spectrometry instrumentation and bioinformatics, many modern strategies are now embracing quantitative proteomics to help define the true interacting partners of probes, and we highlight the opportunities this rapidly evolving technology provides in chemical proteomics. Finally, some of the limitations and challenges of chemical proteomics approaches are discussed.

Introduction

Natural products (NPs) have always been the inspiration for the development of drugs and a source of tools to elucidate cellular mechanisms. Drug discovery typically takes one of two approaches: a target-based approach, where a compound is developed to target a particular enzyme, and phenotypic screening, where libraries of small molecules are screened against live cells with a phenotypic read-out. A target-based approach has the advantages that mode of action is, in principle, known right from the outset, enzyme assays are typically reliable and easy to perform, and compound design is often aided by structural biology. Phenotypic screening, on the other hand, already takes issues such as cell uptake and efflux into account, and non-specific cytotoxicity is more easily assessed. For the phenotypic approach in particular, it is very challenging to discover and validate the mode of action of a compound. An added complexity is that compounds may have different modes of action in different organisms or contexts, or may have pleiotropic effects.

Chemical proteomics is a rapidly evolving area of chemical biology that draws on synthetic chemistry to generate probes that inform on small molecule mode of action and protein function. In the context of understanding the mode of action of NPs and other small molecules, chemical proteomics primarily provides information on the proteins a compound interacts with. Chemical proteomic strategies for NP target identification were reviewed in 2010^{1,2} and several reviews since have touched on the different approaches for target identification, of natural products, drugs and other molecules of interest.³⁻⁵ Here we aim to survey and analyse studies from the last five years on chemical proteomics technologies for mode of action, focusing on the following aspects: natural products, ‘discovery’ chemical proteomics (i.e. where the mode of action and binding partner of the NP is unknown at the outset), in situ approaches using covalent or photoprobes, two-step labelling protocols that exploit bioorthogonal ligation chemistry, and finally the advantage that can be taken of recent advances in quantitative mass spectrometry-based proteomics methods. We define ‘in situ’ chemical proteomics as labelling with a ‘free’ probe in living cells, to distinguish it from classical affinity chromatography (where lysates are incubated with resin functionalised with the NP of interest and binding partners retained, whilst non-binders are washed away), or incubation of cell lysates with probes. Whilst affinity chromatography has been widely and, in some cases, very successfully applied for target identification, it suffers from several disadvantages: weak but true binders are often missed (false negatives), the in vitro setting is more artificial than applying a probe in live cells (and therefore potentially prone to false positives), and the ‘modification’ to the probe is large and may perturb protein binding. Incubation of a free probe with lysates avoids some of these issues but not the artificial setting; approaches that apply a probe in situ have the significant advantage of being more biologically relevant. However, it is generally accepted that in order to be used in situ, the probe should be functionalised to covalently link with binding partners. This can be accomplished by incorporating either an intrinsically chemically reactive group or a photoreactive group within the probe structure.

We will first cover some general points on probe design, validation and chemical proteomics workflows. The second part of the review covers recent case studies and we conclude with a discussion on the advantages, limitations and challenges of these approaches.

Two-step activity and affinity-based protein profiling

Chemical proteomics and bioorthogonal ligation chemistry. In situ chemical proteomics approaches take a small molecule of interest and create a probe by functionalising it with 1) a tag that will allow downstream analysis and, if necessary, 2) a handle to stabilise the interaction between the molecule and its protein binding partner (Figure 1). This second handle could be a photoreactive or intrinsically chemically reactive moiety. In some cases it is possible to directly attach a fluorophore or biotin affinity group in place of the tag; however, a two-step tag-then-capture approach is becoming increasingly popular. Two-step methods take advantage of bioorthogonal ligations, highly chemoselective reactions that take place under physiological conditions between two partners – probe-bound protein and fluorophore for example – in high yields and with high specificity.⁶ These reactions enable the tag

incorporated into the molecule of interest to be very small: for example, a terminal alkyne, azide or cyclopropene. This approach minimises the alteration to the structure of the small molecule and therefore reduces the chances that such functionalisation will perturb binding to the target protein, as well as preserving as much as possible physicochemical properties so that cellular uptake is relatively unaffected and the probe can be applied in the context of a live cell. The most widely used bioorthogonal ligation is the copper-catalysed azide-alkyne cycloaddition reaction (CuAAC) between an alkyne and azide.^{7, 8} These two functional groups are relatively easy to introduce synthetically, small, inert in biological contexts and the reaction is robust. The main disadvantage of CuAAC is that it is limited, largely, to use in lysates and fixed cells because the copper catalyst is toxic. However, for most identification strategies such toxicity is not a problem: the probe is applied in live cells ('in situ'), the cells lysed and CuAAC carried out in lysate. CuAAC also denatures and, sometimes, causes protein aggregation and precipitation, which may make it incompatible with some downstream applications, although some efforts have been made to optimise reagents to reduce this effect.⁹ Two-step protocols also allow the type of label introduced to be easily varied; for example, a biotin or FLAG group can be introduced for affinity enrichment of tagged proteins and subsequent identification by standard proteomic workflows, or a fluorophore can be attached for visualisation in SDS-PAGE gels or for imaging in cells (Figure 1). Alkynes have emerged as preferred tags due to their small size and the empirical observation that having the azide on the labelling partner during CuAAC sometimes results in lower background.¹⁰⁻¹² It is worth noting that removing excess probe can sometimes be important for efficiency of CuAAC and can be accomplished by washing cells or precipitating lysates. More importantly, unreacted CuAAC labelling reagent must usually be removed before affinity enrichment analysis, since excess reagent can bind to the beads and prevent efficient pull-down of labelled proteins. This introduces an extra clean-up step, usually protein precipitation, and any additional steps in a complex workflow can introduce variability. Therefore, in some cases, direct labelling may have advantages: for example, in labelling of secreted proteins, where cell permeability of probes is less of a concern. Due to its wide applicability compared to other bioorthogonal ligations, we will focus here on CuAAC approaches for NP target identification.

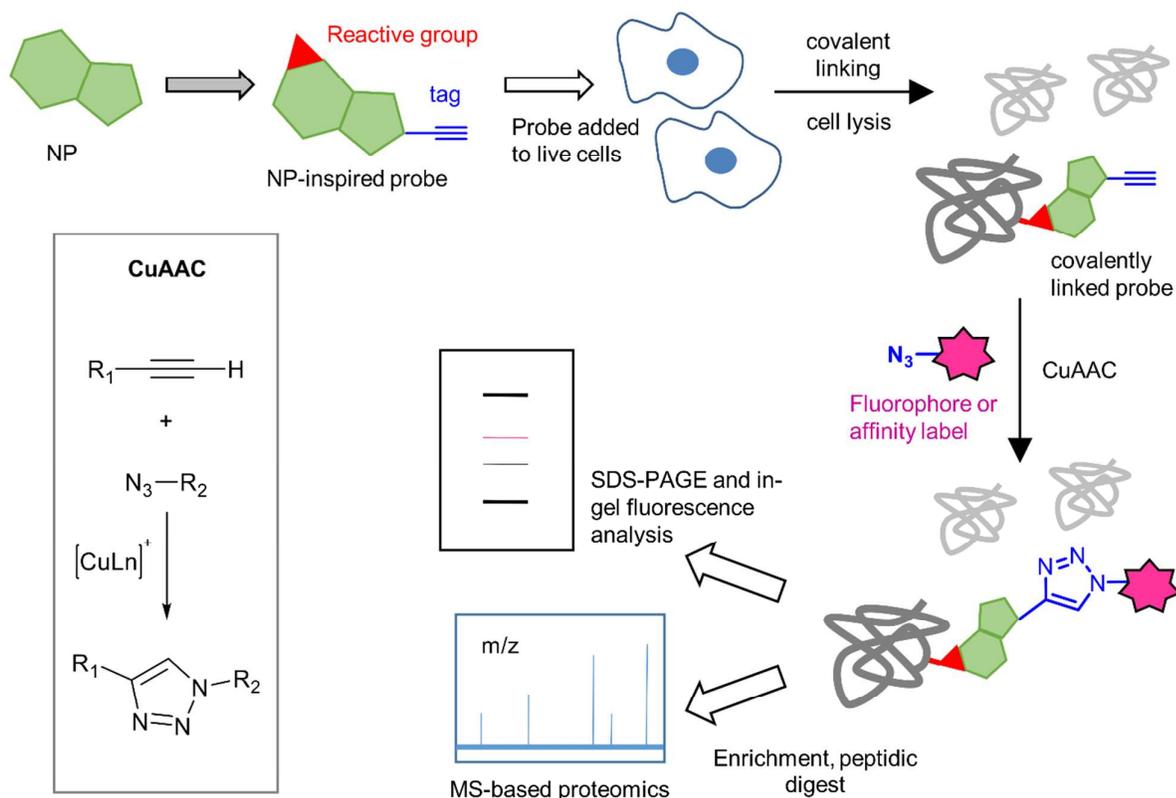


Figure 1: Compound-centric chemical proteomics approach for identifying the targets of natural products (NPs). The probe is designed based on the structure of the NP, is added to live cells and binds to its target protein. It reacts covalently (via an electrophilic trap or a photocrosslinking group) with the target protein. Cells are lysed and samples subject to copper-catalysed azide-alkyne cycloaddition (CuAAC) to attach a fluorophore or affinity label for downstream analysis.

Using intrinsic chemical reactivity. Use of CuAAC chemistry and in situ labelling usually requires a second handle in the molecule to stabilise the interaction with the target protein. Pioneered by the Cravatt and Bogoy group, activity-based protein profiling (ABPP) is now a well-established strategy that uses probes containing electrophilic groups that react covalently with enzyme active site nucleophiles.^{13, 14} The reaction is 'activity-based' because only those enzyme molecules that are in an active conformation with an exposed active site can be labelled with the probe. Some NPs are electrophilic, exerting their biological effect via this type of mechanism, and are therefore readily amenable to an ABPP approach: the NP is transformed into a probe itself (Figure 1). Alternatively, if the NP is thought to address a particular enzyme class, it can be screened in competitive mode against promiscuous ABPP probes that label that class (Figure 2): the probe labels a defined set of proteins and if the NP addresses one or more of these by binding to the same site as the probe, a reduction in labelling of those proteins is detectable. The probe may be directly functionalised with a detection tag or a two-step bioorthogonal ligation may be used. In this review we refer to the approach of

modifying a NP to generate a probe as 'compound centric' and the approach of screening against other probes as 'competition-based'. It is worth mentioning that not all reactive probes are strictly speaking 'activity-based'; as illustrated later, many probes appear to not only react with active site nucleophiles but also with other residues on enzymes, and other proteins. Whether this reactivity is biologically relevant is very case- and, probably, context-dependent.

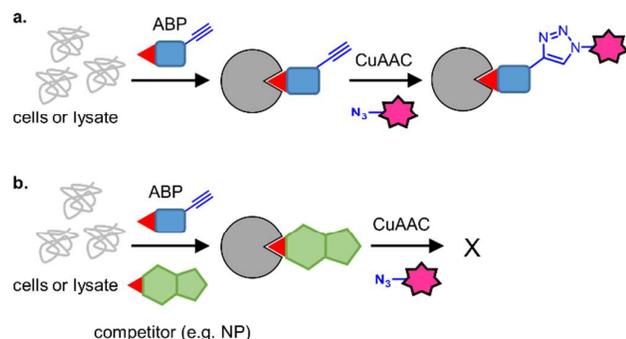


Figure 2: Principle of competitive ABPP. a. An activity-based probe (ABP) reacts with target proteins and can subsequently be detected by CuAAC. b. In the presence of a competitor, which may react with the protein or may be non-covalent, the ABP cannot bind and a reduced signal is detected.

Potential electrophilic traps include Michael acceptors, ring strain scaffolds (β -lactones, β -lactams, aziridines, epoxides), α -halo-carbonyls, isothiocyanates and many more (Figure 3). Some, such as β -lactones, are common motifs in natural products, whereas others, fluorophosphates for example, are not found in nature but have been discovered or designed by chemists. Tuning of electrophiles is an active area of research in chemical biology and will undoubtedly yield new reactive motifs in the future to expand the available toolbox.¹⁵

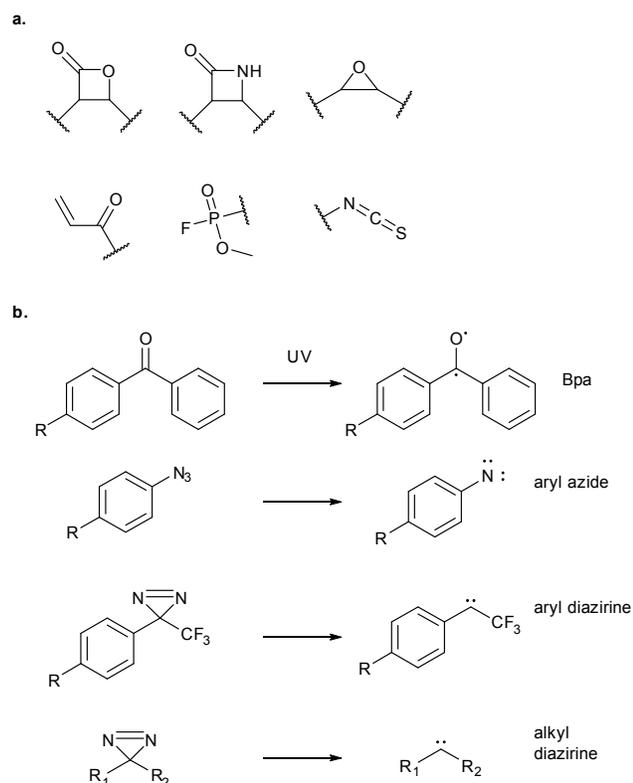


Figure 3: a. Common electrophilic traps that react with nucleophilic residues in proteins. b. Commonly used photoreactive groups and their reaction under UV light.

An ABPP approach is clearly most applicable to NPs that are intrinsically electrophilic. However, many NPs bind their targets non-covalently. An intriguing strategy, but one that has yet to be widely explored in the field, is that of 'weaponising' NPs with reactive traps. Porco, Cravatt and co-workers recently reported the synthesis of a β -lactone-functionalised derivative of the NP rocaglate, which was shown by competitive ABPP to target serine hydrolases.¹⁶ Such functionalisation of a NP has the potential to significantly alter its biological targets; however, even if this is the case, it may lead to new and unexpected useful tools for probing cellular function.

Photoaffinity labelling. A more intuitive and unbiased strategy for identifying the binding partners of unreactive NPs is to use photoaffinity labelling (PAL). PAL makes use of photoreactive moieties that are inert under standard synthetic chemistry and biological conditions but can be activated by UV light, generating highly reactive transient species that crosslink the probe to whatever molecule is in close proximity – including presumably the protein binding partner. Commonly used PAL groups are benzophenones, aryl azides and alkyl or aryl diazirines (Figure 3). They all have their advantages and disadvantages (reviewed in¹⁷), and a dearth of comparative studies means that no consensus has yet emerged on how to choose one over the other; indeed the

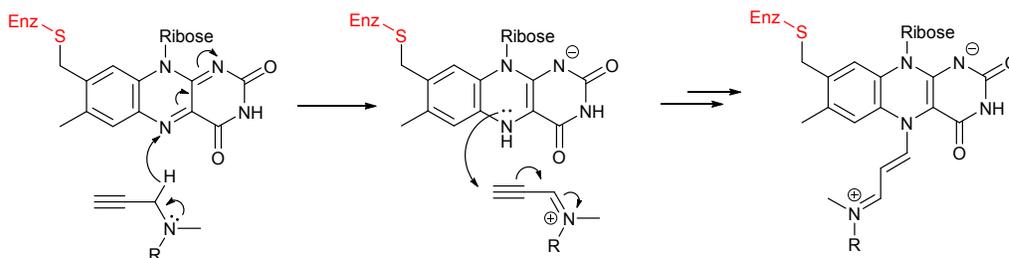
choice of photocrosslinker seems to be highly context dependent.¹⁸ There has been a recent resurgence of interest in alkyl diazirines, paralleled by improvements in synthetic routes to access building blocks,¹⁹ due to their small size, unbiased reactivity, and the relatively undamaging wavelengths required for their activation.

Design and validation of probes

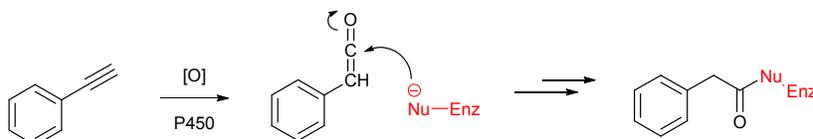
Probe design. Designing probes based on NPs in the absence of any information on the potential targets or origin of bioactivity is very challenging, and is compounded by the structural complexity of many NPs and the necessity of developing total syntheses to access them. Even small tags and handles such as alkynes or diazirines can perturb bioactivity and binding if installed in the ‘wrong’ position. However, there is almost always scope for modifying a NP without drastically changing its mode of action and in this context structure-activity relationship information can be very helpful. An alternative approach is to generate multiple probes with tags introduced at different positions. The synthesis of tag/handle-modified NPs is highly dependent on the NP in question and the reader is referred to a recent review on the topic.²⁰ Here, we will instead illustrate aspects of probe design throughout with reference to examples.

Intrinsic reactivity of alkynes. An aspect of probe design worth discussing is the potential reactivity of alkynes themselves. Terminal alkynes are found in several mechanism-based inhibitors of enzymes. One example is the inhibition of FAD-dependent enzymes by *N*-propargylamines: in the case of monoamine oxidases, the FAD cofactor, which is covalently bound to the enzyme active site via cysteine, oxidises the *N*-propargylamine of the bound inhibitor, and subsequently covalently attacks it (Figure 4); this approach was used to develop activity-based probes for this class of enzymes, which are important for neurological functions due to their roles in oxidative deamination of neurotransmitters.²¹ Similarly, terminal alkynes can act as mechanism-based inhibitors of P450 enzymes: oxidation of aryl acetylenes by the enzyme generates a highly reactive ketene, which is rapidly attacked by adjacent enzyme nucleophiles. Alternatively, alkyl acetylenes can be oxidised to generate Michael acceptors. Again, such a strategy has been exploited for the design of ABPs for p450s.^{22, 23} Finally, Ovaia et al. showed that an *N*-propargyl amide functionality attached to ubiquitin reacted selectively with the active site cysteine of deubiquitinating enzymes; furthermore, they showed that *N*-propargyl amides appended to other scaffolds also targeted the active site cysteines of other proteases highly selectively.²⁴ So although alkynes are considered highly bioorthogonal in most contexts, they can be rendered reactive by enzymes or can react directly – usually aided by binding in an enzyme pocket. The extent to which alkynes react or are destroyed in living cells is currently unknown, since such reaction results in removal of the detection tag.

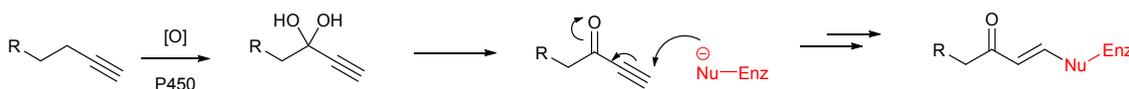
a.



b.



c.



d.

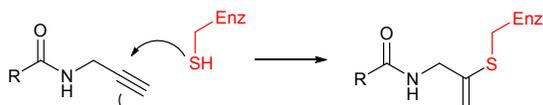


Figure 4: Reactive alkynes. a. Mechanism-based inhibition of FAD-dependent enzymes by propargyl amine moiety.²¹ b. Activation and inhibition of cytochrome p450 enzyme by aryl alkynes or c. alkyl alkynes.^{22,23} d. Inhibition of proteases reported by Ekkebus et al.²⁴

Probe validation. Probes are typically validated by direct comparison with the NP of interest in bioactivity assays. Some effect on bioactivity is not necessarily damning, since the probe may be slightly more or less active than the NP but still retain the same mode of action. Full proof of probe function usually proceeds together with identification of putative binding partners.

Chemical proteomic workflows

With probe in hand, the first steps are typically small scale labelling experiments to optimise time, concentration and other conditions with a simple and quick read-out by SDS-PAGE gel. A typical workflow is the following: incubation of probe ‘in situ’,

photo-irradiation if appropriate, lysis, CuAAC ligation to append a fluorophore or other detection label, and analysis by gel electrophoresis and fluorescence scanning or Western blot (to detect biotin). In-gel fluorescence is quicker and cleaner than blotting for biotin, since there are no additional signals from endogenously biotinylated proteins.¹¹ Probe incubation can alternatively be carried out 'in vitro' in cell lysates. Lysis conditions are important in this case since the proteins must remain in their functional and folded states in order to specifically bind probe molecules. In situ approaches avoid this potential concern and probes interact with a 'native' system. The caveat to an in situ approach is that probe labelling of proteins may be reduced by uptake issues or metabolism, and cytotoxic probes can also cause problems by killing cells and reducing the amount of material recovered. The latter problem can usually be avoided by the use of short labelling times. An alternative to this compound-centric strategy is to use a competitive approach, where a NP of interest is used to outcompete promiscuous probes that, for example, bind reactive cysteines (see later). Competitive profiling is usually applied in vitro but has advantages in that it completely circumvents the need to create a functional probe of each compound.

The conditions for probe incubation should clearly reflect the conditions of the biological assays being used to assess the phenotype of the NP. For microorganisms the life stage or growth stage and culture medium are important, since these factors can impact uptake of compounds as well as the proteome, and hence potential probe targets. Probe concentration is also often optimised. Electrophilic and photoreactive probes usually label promiscuously at high concentrations; for example, a recent paper by Cravatt and co-workers analysed the labelling and targets of covalent kinase inhibitors converted into activity-based probes by attachment of an alkyne tag: at high concentrations of probe, labelling increased dramatically and most bands could not be out-competed with parent inhibitors – indicating non-specific reaction of probe with proteins.²⁵ The authors linked this increased labelling with generic cytotoxicity. This study also illustrates how competition with parent inhibitor can be used effectively to assign proteins as true or off-targets of a probe and refine the list of hits.

Proteomic identification of hits. To identify potential hits of probes, experiments are performed with click chemistry ligation to an affinity label – usually biotin. Biotin binds strongly to the protein streptavidin and its variants, allowing enrichment or pull-down of labelled proteins onto streptavidin-coated beads. Several groups have made use of customised CuAAC reagents that incorporate both biotin and a fluorophore to enable simultaneous enrichment and facile detection.^{10, 26, 27} Enriched proteins can be eluted from the beads with denaturing conditions for gel-based proteomics, or directly digested on-bead with proteolytic enzymes. In a standard gel-based proteomics workflow, enriched samples are separated by gel, bands or slices cut and digested, usually by trypsin. This approach has been widely used historically and enables simple fractionation of the sample, but suffers from contamination with environmental proteins such as keratin and problems of bias – certain proteins simply do not separate well on a gel. It is also difficult to quantify between samples. For these reasons, and enabled by the improved sensitivity of mass spectrometers, gel-free approaches are becoming increasingly common. Direct on-bead digest is simple to perform, minimises environmental contamination and loss by reducing sample handling, and is less biased than gel-based methods. In both cases, the resulting mix of peptides can be analysed by standard shotgun LC-MS/MS workflows: peptides are separated by reverse phase chromatography, and injected on-line into a tandem mass spectrometer for measurement of ions, fragmentation and measurement of fragments. Identification of proteins is carried out by software packages that compare the acquired spectra with spectra obtained from in silico digest of proteins in a database (e.g. the total proteome of the organism of interest). A vehicle control (often DMSO) is processed in parallel to provide a benchmark list of proteins that bind non-specifically to the beads. Hits are therefore identified as those proteins enriched over background.

Quantitative chemoproteomics. Related to the shift to gel-free methods, recent chemical proteomic workflows often incorporate quantification: for example, to compare vehicle control and probe-incubated sample, competition with NP, different but related probes, biological conditions, replicates, and so on. Popular quantification approaches include SILAC (stable isotope labelling by amino acids in culture), iTRAQ (isobaric tag for relative and absolute quantitation), TMT (tandem mass tag), dimethyl labelling (DiMe), and label-free methods, such as spectral counting and MS¹ intensity-based methods.

Labelling approaches rely on the incorporation of isotopes into peptides that are then mixed and analysed together in the same LC-MS/MS run – peptides originating from different samples can then be distinguished in the mass spectrometer. Label-free approaches compare samples that are analysed in different runs and have greatly improved in recent years with the increasing sensitivity and resolution of LC-MS/MS as well as advances in computational algorithms.²⁸ DiMe, iTRAQ and TMT all involve chemical labelling of peptides²⁹ and so have the drawback that samples are only combined towards the end of the chemical proteomics workflow (Figure 5a), therefore not controlling for operator error or reproducibility of the protocol to that point. Label-free suffers from the same drawback. SILAC uses cultured cells where all proteins incorporate heavy, light or medium amino acids and is therefore often a popular choice where it is applicable, since it allows samples (e.g. DMSO vehicle control and probe) to be combined directly after lysis (Figure 5b). One disadvantage of labelling methods like SILAC is that only a limited number of samples can be combined; however, a 'spike-in' approach can be used to overcome this: in this approach a heavy standard sample is spiked into any number of samples which are run separately and then the standard used later in analysis for quantitative comparison.³⁰ Considerations in the choice of quantitative approach include: the complexity and robustness of the workflow (early mixing is an advantage), the number of samples to be analysed and their origin (e.g. vehicle control vs probe vs probe + competitor in a well-studied cell culture line is easily done via SILAC, but for comparing tens of clinical samples from diverse origins label-free may be more appropriate), cost (some labelling methods use expensive reagents), access to MS measurement time (label-free requires each sample to be measured separately whereas in labelled approaches samples are combined), and the MS instrument set up (some methods quantify at the parent ion MS¹ level, others at the MS² fragment ion level; reproducible chromatography and instrument performance are particularly important for label-free). For a full discussion of these issues the reader is referred to several recent reviews.^{29, 31}

In chemical proteomics, a quantitative approach is a significant improvement on simply determining the presence or absence of a protein in a particular sample and also allows for greater flexibility in the types of experiments that can be performed. Some probes are highly specific, whereas others will label many proteins, either due to promiscuity of the NP itself or to limitations in the methodology (non-specific photoreactivity or intrinsic reactivity). We and others have found that quantifying the proteins enriched in the presence and absence of a competitor such as the parent NP, for example, can provide a lot of information on the putative hits of a probe.

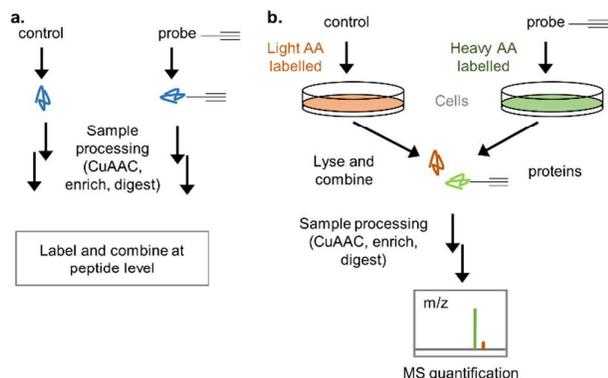


Figure 5: Quantitative approaches for chemical proteomics experiments. a. In many label-based approaches (ITRAQ, DiMe.), samples are processed separately, peptides are isotope labelled and then combined before mass spectrometry. b. SILAC (stable isotope labelling of amino acids in culture): Cells are cultured with medium containing heavy or light isotope labelled amino acids; cells are treated differently (e.g. probe or vehicle control/untreated), lysed and samples combined for all downstream analysis.

Cleavable linkers in chemical proteomics. In the majority of chemical proteomics experiments, binding of a probe to a protein is inferred indirectly from the presence of the protein (in fact its constituent peptides). This is because the probe-modified peptide is likely left attached to the beads after digest (in gel-free approaches) and/or the peptides are not separated well by HPLC or do not 'fly' well in the mass spectrometer due to the bulky modification. However, direct identification of the modification site adds confidence in the assignment of the protein as a hit and provides additional information about the putative binding site of the NP. Several reagents for CuAAC have been reported that incorporate a chemically, light, pH, enzyme or otherwise cleavable linker, enabling selective release of the probe-modified peptide from the beads, either with the total shotgun mix or in a separate step. Such approaches also have the potential to reduce contamination with non-specific background bead binders, since proteins linked via the CuAAC reagent can be selectively eluted. An early approach by the Cravatt lab termed TOP(Tandem Orthogonal Proteolysis)-ABPP^{32, 33} uses click reagent **1** (Figure 6) incorporating a TEV (tobacco etch virus) protease cleavable sequence that allows release of modified peptides from the beads orthogonal to trypsin (the enzyme typically used for the digest of proteins for MS). This reagent has been used successfully for identifying cysteines modified by simple electrophiles, such as an alkynylated iodoacetamide derivative.³⁴ Diazobenzene-based reagents cleaved by treatment with Na₂S₂O₄ have been used to identify peptides incorporating bioorthogonally tagged amino acids³⁵ and trypsin protease cleavable linkers were recently reported for the detection of modification sites after incorporation of alkyne-tagged fatty acids or AMP analogues into proteins.^{36, 37} Other reagents such as vicinal diol linkers cleavable with sodium periodate have been shown to reduce non-specific bead background³⁸ but have not yet been used for site identification. In general, few cleavable reagents have been applied to directly identify the site of modification of NPs, probably because the structural complexity, large size or lability of some NP probes is still a limiting factor for their direct detection. We have also found no report of direct detection of photolinked probe in a chemical proteomic context. This is perhaps unsurprising since photoprobes can in principle react at almost any position, with coincident losses or rearrangements, yielding a very complex problem for the search algorithms. Elution efficiency of the intact proteins for gel electrophoresis is also reported to be poor in some specific cases.³⁹

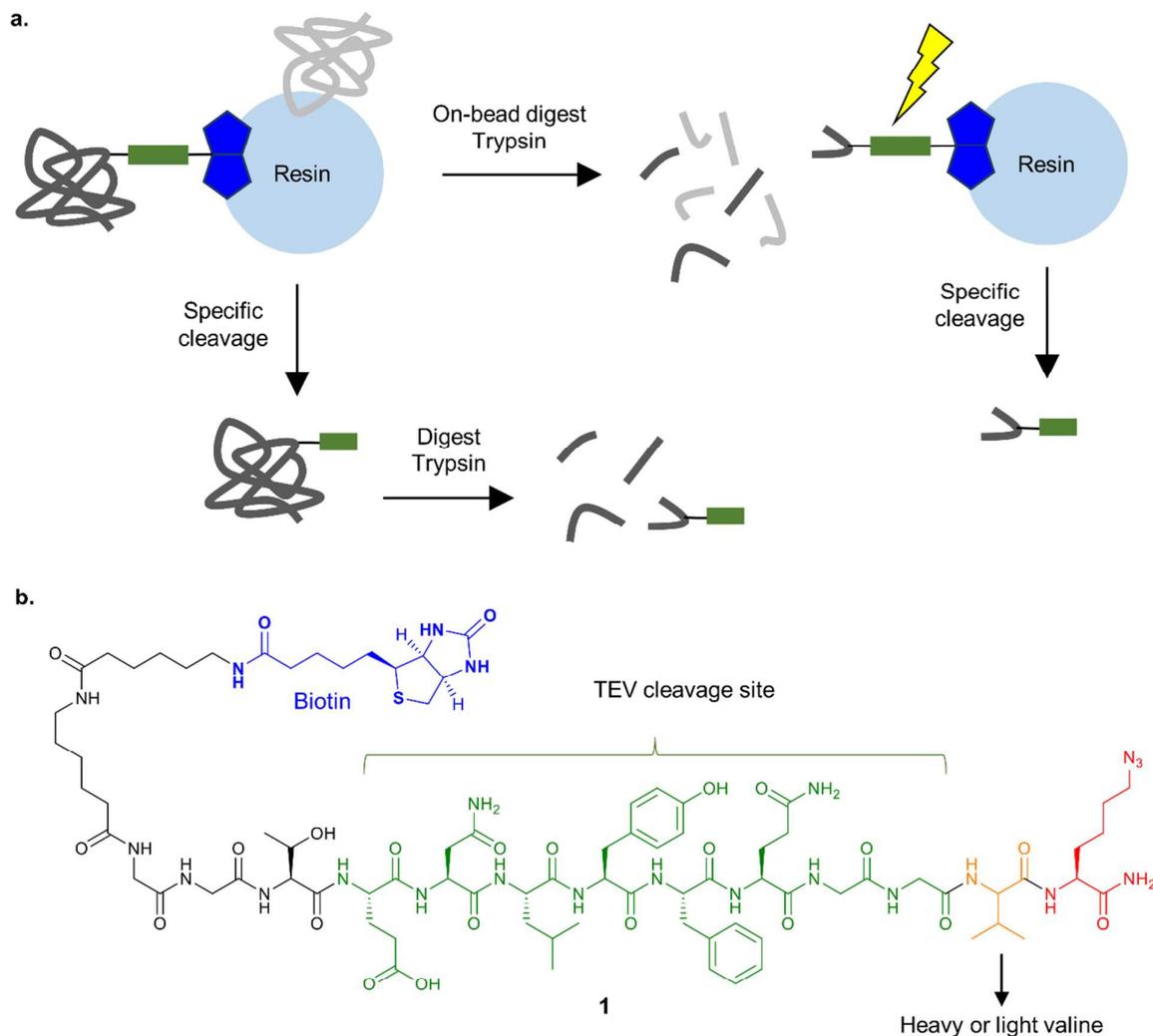


Figure 6: Cleavable linkers for chemical proteomics. a. Proteins enriched on resin e.g. by biotin-streptavidin, can be specifically released and then analysed by MS (bottom), or digested on bead and then the modified peptide selectively released (top). b. The isotopically-encoded protease (TEV)-cleavable linker **1** for CuAAC and enrichment reported by Cravatt and co-workers.⁴⁰

Alternative enrichment/detection strategies. Biotin-streptavidin is the most widely used enrichment system, due to the strength of the interaction and the ready commercial availability of reagents. However, this system does suffer from background binding of endogenously biotinylated proteins that can hinder detection of low abundance hits. An alternative immunoaffinity method developed by the La Clair group uses a coumarin label (immunoaffinity fluorescent label, IAF), which has a dual role as both a dye for cell imaging and an affinity label (Figure 7).^{41, 42} The specific antibody raised against the coumarin epitope was shown to enable much cleaner immunoprecipitation than biotin-streptavidin in a BSA (bovine serum albumin protein) test system.⁴¹ In the same study the authors systematically examined CuAAC two-step labelling for both biotin and IAF, with the azide and alkyne partners on either the protein or the labelling reagent. They observed better results with the protein-alkyne and IAF-azide combination, in line with other studies.¹⁰⁻¹² Interestingly, they noted that CuAAC likely requires a covalent linkage between probe and target protein for effective enrichment, consistent with the denaturing nature of the reaction: in applying their method to the NP glycyrrhetic acid, if the coumarin label was ready-attached to the probe, enrichment was possible, but two-step CuAAC did not result in any enrichment. The IAF method has been applied to identify the targets of the NPs seriniquinone⁴³ and spirohexenolide A,⁴⁴ amongst others.

Another approach that avoids biotin used a resin pre-coupled to an alkyne for simultaneous click and enrichment of an azide-tagged probe and its binding partners.⁴⁵ The NP scalarial **2** was functionalised with an azide, generating probe **3**, added to lysates or live cells, and then lysates incubated with the alkyne-resin and click reagents (Figure 7). The authors also directly compared the new method with a purely in vitro method where scalarial was directly attached to a resin and incubated with lysates. Western blot confirmed that the previously reported protein targets PRX1 and 14-3-3 ϵ were enriched by both methods, and gel-based proteomics was used to identify possible hits. Since scalarial presumably binds its protein targets non-covalently, this study suggests it may be possible to optimise on-bead click to work without introducing a photocrosslinker or reactive moiety. However, controls performed in the study showed that even in the absence of the reducing agent required for the click reaction, some PRX1 and 14-3-3 ϵ were detectable on the beads. These proteins are abundant and it remains to be seen whether such an approach will work for low abundance protein binders of a probe.

Sibbersen et al. recently reported a method combining click-on-resin with a chemically cleavable diazobenzene linker to facilitate release of enriched proteins.⁴⁶ The amine- and azide-functionalised diazo-unit was prepared by coupling to NHS-activated sepharose resin (Figure 7). The authors then applied their approach in several test systems and in comparison with either a standard biotin reagent or a diazo-biotin. They labelled cells with a promiscuous electrophilic probe, 2-oxohex-5-ynal, performed CuAAC and enriched proteins from lysates. All three methods performed comparably, giving a similar pattern of enriched proteins by Coomassie staining, with the difference that the methods employing biotin also resulted in contamination with significant amounts of monomeric streptavidin. It will be interesting to see whether this type of reagent proves useful for chemical proteomic identification of protein binders of a probe, and whether it proves possible to isolate non-covalent probe-protein adducts.

Several methods for on-resin capture and release have also been reported for peptides^{47, 48} and, most recently, a silane-based fluoride cleavable linker was used to capture by strain-promoted azide-alkyne cycloaddition (SPAAC) and then release a small molecule and the enzyme chymotrypsin.⁴⁹ Interestingly, the released enzyme still retained ~67% of its activity, consistent with the milder nature of SPAAC compared to CuAAC.⁴⁹

Biotin-streptavidin continues to be the most widely used method for enrichment due to the strength of the interaction and the ready availability of tools. Disadvantages include difficulties in distinguishing non-specific streptavidin background binders from true hits and challenges of identifying modified peptides. Numerous cleavable linkers have been reported to tackle the second problem, but further advances in sample preparation and bioinformatics workflows will likely be needed to expand their use. New technologies for enrichment are always a welcome addition to the chemical proteomic arsenal.

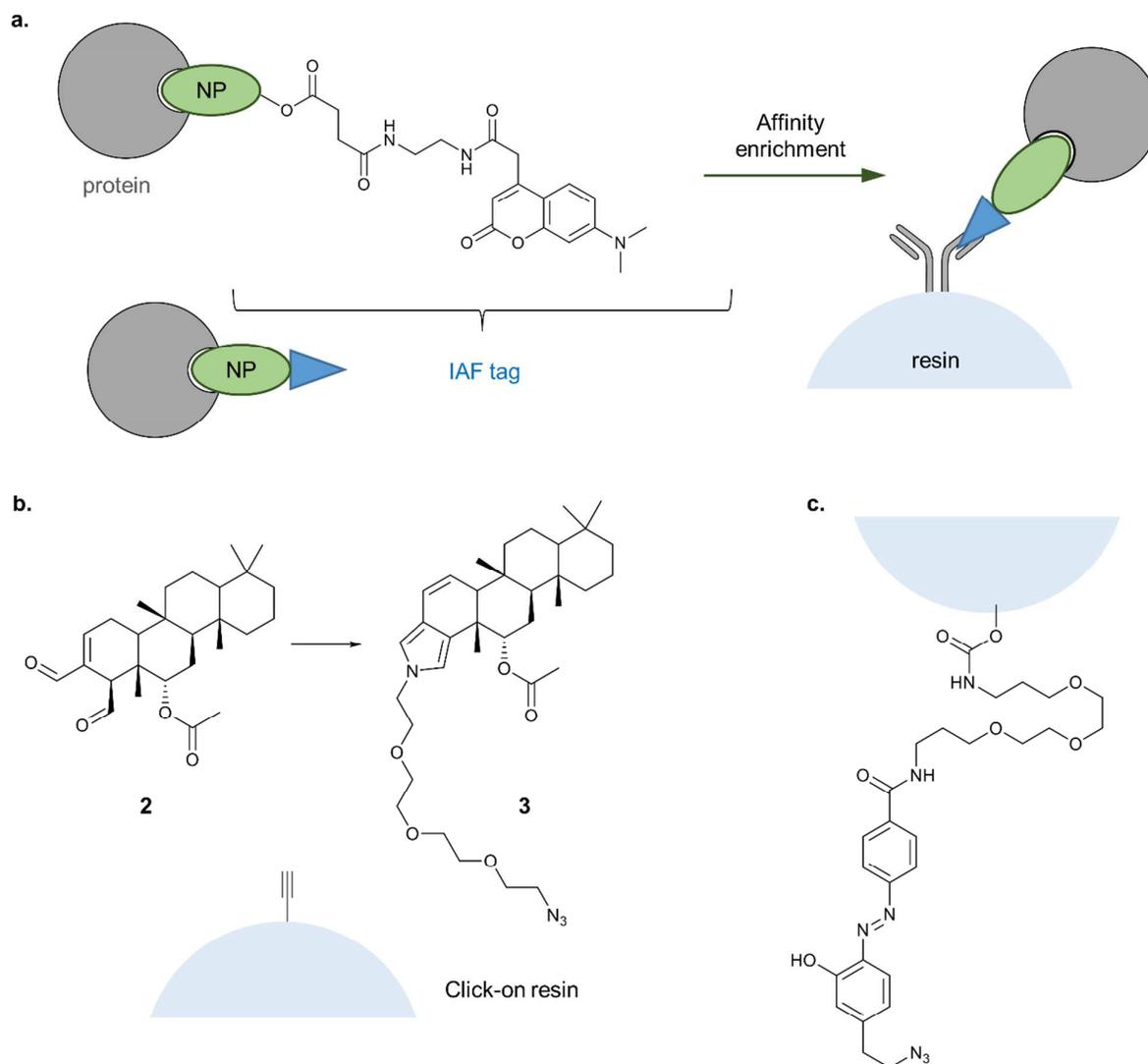


Figure 7: Alternative enrichment methods to biotin-streptavidin. a. IAF (immunoaffinity fluorescent tag) reported by La Clair et al. for antibody-based enrichment of probe-bound proteins. b. Click-on-resin approach reported for the NP (natural product) scalaradiol: scalaradiol **2** is functionalised to generate probe **3**, which can be attached in turn by CuAAC to an alkyne-functionalised resin.⁴⁵ c. On-resin cleavable linker reported by Sibbersen et al.⁴⁶

Target validation and mode of action

Once potential hits have been identified, the challenging tasks of validating that they are true hits and linking binding to mode of action begin. The first problem is, relatively, simple and is typically approached in several ways: if an antibody is available or can be

produced, labelling can be validated by Western blot to show that the protein of interest is enriched after CuAAC and affinity pull-down, or by immunoprecipitation and on-bead CuAAC to install a label for detection. Alternatively, the protein of interest can be obtained recombinantly and labelling shown on the purified protein or the protein in the context of the overexpression system or spiked into a lysate. Labelling of an isolated protein should be viewed with caution, however, as some promiscuous probes label almost any protein in such an artificial environment. To demonstrate that probe binding is dependent on a native, properly folded protein, a 'heat' control in which proteins are first heated with SDS to unfold them is sometimes used. Access to the purified protein also provides the opportunity to detect modification of the protein by the probe via intact protein MS or to digest and identify the binding site by LC-MS/MS in a simpler system than the total enriched proteome. Some probe-protein interactions can be studied by biophysical methods, such as ITC (isothermal titration calorimetry) and structural biology can also provide supportive evidence, in the form for example of a crystal structure of probe-bound protein. In the case of electrophilic probes, if the site of modification can be identified, the amino acid residue involved, e.g. Cys, Ser, can be mutated and the mutant protein analysed to demonstrate that it no longer binds the protein of interest.

Demonstrating a direct link between NP binding to a protein and mode of action of the NP can be very challenging. We will illustrate the types of experiments used to tackle this problem via specific examples. However, in general it is necessary to apply many different biological and chemical tools. For example, in the context of the chemical proteomics experiment, access to a structurally similar but biologically inactive NP and/or probe is one approach that can help distinguish biologically irrelevant but nonetheless genuine binders from those protein targets that mediate the bioactivity. Small libraries of probes, competition experiments, comparison of labelling in biological systems where the phenotype of response differs, chemical or genetic knockout or knockdown, are all used to demonstrate a link between binding and bioactivity.

Fishing for targets: probe-centric approaches

The following section discusses specific examples where CuAAC together with in situ chemical proteomics has been used to identify the binding partners of NPs.

Early in situ approaches to identify the targets of electrophilic natural products

Antibacterials. Prior to 2008, several ABPP-type approaches for identifying the targets of electrophilic NPs were reported, but studies typically used a fluorophore or biotin directly appended to the NP scaffold for detection. In 2008 we took inspiration from the fact that several classes of important antibiotics contain electrophilic β -lactone and β -lactam moieties to develop "minimal", alkyne-tagged ABPs from these structures.⁵⁰⁻⁵² Based on the β -lactam core, probes derived from penicillin, aztreonam, cephalosporin (**4**) (Table 1), as well as more simplified structures, were synthesised.⁵⁰ Labelling with these probes was profiled by in-gel fluorescence after 'in situ' incubation with live bacterial cells and 'in vitro' incubation in cell lysates. Gel-based proteomics was used to identify several penicillin-binding proteins as targets, as well as a β -lactamase enzyme; this enzyme class is important for antibiotic resistance in pathogenic bacteria. Recombinant overexpression and analysis of the catalytic activity of the β -lactamase confirmed that the probes indeed inhibited the enzyme. Subsequent studies used NP-inspired β -lactams to identify and compare enzymes in antibiotic-resistant and sensitive strains of *Staphylococcus aureus*, the cause of major MRSA-hospital-acquired infections.⁵¹ An identified MRSA-specific enzyme of unknown function was further characterised as a likely metallo- β -lactamase by enzyme assays, labelling with other tools specific for this enzyme class, and competition of labelling with structurally-unrelated protease inhibitors. Others have also applied β -lactam probes to label the nucleophilic active sites of other enzyme classes such as a specific lysosomal cysteine hydrolase,⁵³ or have developed cephalosporin C-based probes for imaging.⁵⁴ Simplified probes incorporating the 3-membered ring heterocycles oxirane, thirane and aziridine – other electrophilic motifs found in NPs – were also shown to label proteins in situ in pathogenic bacteria strains,⁵⁵ further expanding the toolbox of probes for functionally characterising enzymes in diverse systems.

One of the interesting targets of β -lactones identified in the original study by Böttcher and Sieber was the caseinolytic peptidase ClpP, whose activity is also linked to virulence.⁵² Some organisms, such as the human pathogen *Listeria monocytogenes*, have two isoforms of ClpP, although most β -lactones only label one. Zeiler et al. found that an alkynylated probe **5** based on the NP vibrallactone labelled both isoforms, providing a useful tool for biochemical and structural characterisation of the ClpP system in *Listeria*.^{56, 57} Our group has applied a similar ABPP approach to identify the targets of the antibacterial NPs showdomycin, rugulactone and hydroxyderricin. Showdomycin is a nucleoside antibiotic with an electrophilic maleimide moiety, a motif that commonly reacts with thiols. The probe **6** was synthesised by esterification of the free alcohol on the ribose (Table 1) and applied in situ in different pathogenic bacteria.⁵⁸ Proteins identified as targets by gel-based proteomics were oxidoreductases and transferases, including the enzyme MurA1, which catalyses an essential step in cell wall biosynthesis. Another putative target, the reductase AhpC, was recombinantly expressed, incubated with showdomycin, digested with chymotrypsin, and the peptides analysed by MS to identify the binding site of the NP.

The NP rugulactone also contains two, presumably cysteine-reactive, Michael acceptor systems. Nodwell et al. synthesised a rugulactone probe **7** possessing an alkyne tag introduced on the phenyl ring (Table 1) and a number of analogues with the Michael system reduced to assess the role of these electrophilic moieties.⁵⁹ This is an example of a strategy used to link the reactivity of a NP with protein binding and biological activity. Application of in situ labelling and gel-based proteomics led to the identification of a kinase, ThiD, involved in the biosynthesis of thiamine, as one of the prominent hits. By showing that antibacterial activity of the NP increased in thiamine-deficient medium, the authors provided evidence that the bioactivity is related, in part at least, to ThiD inhibition. Battenberg et al. also applied ABPP to identify the targets of the antibacterial chalcone 4-hydroxyderricin (probe **8**, Table 1), another Michael acceptor functionalised NP.⁶⁰ Labelling in the cell lysates of pathogenic bacterium *S. aureus* and gel-based proteomic identification facilitated by a diazobenzene-based cleavable linker to release bound proteins from streptavidin resin,

revealed seryl-tRNA synthetase (STS) as one of the putative hits. The enzyme, which is responsible for loading cognate tRNAs with serine, was inhibited in vitro by the NP, suggesting it is a functionally relevant target. However, identification of which of the five cysteines in STS is targeted by the probe proved difficult, suggesting that the compound may bind to multiple nucleophilic sites on the enzyme.

Cytotoxic and anti-cancer NPs. The ABPP approach has also been applied to identify the targets of reactive NPs in human cells. Inspired by NPs such as fumagillin and Lumanicin C, Cravatt and co-workers synthesised a library of alkynylated probes containing a reactive spiroepoxide moiety (e.g. **9**, Table 1) and screened them for cytotoxicity.⁶¹ The compound with the highest cytotoxicity in a human breast cancer cell line was further shown to label an additional band compared to biologically inactive probes in in situ profiling experiments. This band was identified as a glycolytic enzyme phosphoglycerate mutase 1 (PGAM1) by gel-based proteomics. Interestingly, PGAM1 was not labelled in vitro by the probe, highlighting the value of applying probes in a native in situ setting. The TOP-cleavable linker approach described above was used here on partially purified PGAM1 after labelling of the enzyme in an overexpression system, resulting in the identification of the region of modification by the spiroepoxide probe. A follow-up study clarified the site of modification as a lysine residue in the enzyme.⁶²

β -Lactone probes have also been applied in human cells. The β -lactone-containing anti-obesity drug Orlistat is a derivative of the NP lipstatin, an irreversible inhibitor of pancreatic lipase. To investigate the putative targets of orlistat in cancer cells, Yao and colleagues synthesised orlistat-like probes incorporating an alkyne handle at different positions (e.g. **10**, Table 1).⁶³ The anti-proliferative activity and cellular effects of probes were analysed in HepG2 liver cancer cells, confirming their bioactivity was comparable to that of orlistat, and then the probes were applied in labelling experiments. Labelling patterns in vitro were comparable to in situ, but background labelling was higher in vitro. Gel-based proteomics identified known target FAS (fatty acid synthase) as well as additional putative targets of the drug. Western blotting with antibodies after pull-down confirmed the enrichment of several hits, and some of these were overexpressed and shown to be labelled by the orlistat probe. A Cys-Ala mutant of identified putative target GAPDH, an abundant housekeeping protein, was not labelled by the probe, indicating probe labelling of this particular cysteine residue. GAPDH and additional proposed targets of orlistat have yet to be linked to mode of action of this drug, but at the very least orlistat-based probes have revealed that the compound covalently reacts with a number of cellular targets. In 2014, the Wenk and Yao groups further applied orlistat probes to investigate the bactericidal activity of the drug.⁶⁴ Gel-based proteomics after probe labelling of *Mycobacterium* cell lysates revealed lipid esterases as targets.

Chemical proteomics can reveal unexpected modes of action of NPs. Duocarmycins (e.g. **11**, Figure 8), a class of anti-tumour NP from *Streptomyces*, are known to act by alkylation of DNA via an electrophilic cyclopropane and derivatives use a seco-drug structure where the cyclopropane is formed in situ from an appended chloromethyl (Figure 8).⁶⁵ However, removal of the indole DNA-binding unit, gave a compound that was still highly cytotoxic, raising the possibility of alternative non-DNA related targets.⁶⁶ Wirth et al. synthesised a duocarmycin-inspired alkynylated probe **12** which retained cytotoxicity and labelled one prominent protein after in situ incubation with A549 cancer cells.⁶⁷ In yet another example of such an effect, the probe barely labelled this band in cell lysates. Gel-based proteomics identified the protein as aldehyde dehydrogenase ALDH1A1. A follow-up study applied additional probes lacking the DNA-binding motif as well as those based on a bifunctional scaffold; these probes also labelled ALDH1A1.⁶⁸ Further crystallographic, computational and biochemical studies of ALDH1A1 clarified the mechanism of inhibition and the binding site of duocarmycin-inspired compounds.⁶⁹

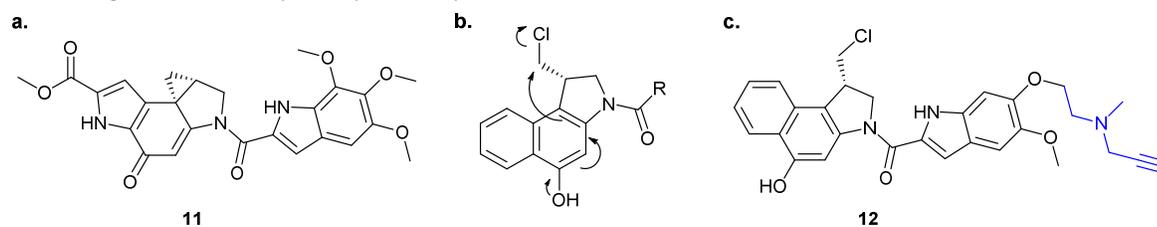


Figure 8: a. Cytotoxic NP duocarmycin **11**. b. Seco-drug version of duocarmycin and activation mechanism. c. Duocarmycin-inspired probe **12** reported by Wirth et al.⁶⁷

Food-derived natural products with potential anti-inflammatory properties are also interesting bioactive compounds. A recent study looked at the isothiocyanate iberin, which was identified from cabbage and onion extracts as an inhibitor of inflammation-mediating Toll-like receptors (TLRs).⁷⁰ As part of this wider study, alkynylated probes (**13**, Table 1) were synthesised and applied in situ. Target engagement was confirmed using Western blot for TLRs after enrichment of tagged proteins, although additional targets of the iberin probes were not identified by proteomics.

Going gel-free and quantitative

A survey of the examples given above makes it clear that although some reactive NPs are highly selective for a few targets, many probes label multiple proteins. Distinguishing between true hits related to bioactivity, hits that are irrelevant for bioactivity but nonetheless genuinely labelled, probe-specific off-targets, and background proteins from CuAAC or enrichment, is not trivial. As discussed previously, the increasing sensitivity and resolution of mass spectrometers, coupled with the improvement in algorithms for analysing complex datasets, has brought quantitative proteomics to the fore in chemical biology. Gel-free approaches enable unbiased identification of hits and quantification allows for far more sophisticated experimental set-ups, including measuring binding in situ by quantifying increasing concentrations of probe, comparing across different tissue or cell types, or quantifying response to inhibitors or competitors. The following recent examples of NP or NP-inspired probes highlight the use of gel-free and/or quantitative proteomics to study NP mode of action.

Studies using label-free methods. Pyrethroids are widely used insecticides derived from pyrethrins, botanical NPs, and act by blocking voltage-gated sodium channels. Ismail et al. developed a library of pyrethroid-inspired alkyne probes in order to profile the

cytochrome P450 (CYP) detoxifying enzymes that may be involved in pyrethroid metabolism and resistance.⁷¹ Aryl-alkynes are known mechanism-based inhibitors of CYPs (as described above) and probes incorporated an alkyne warhead as well as a second alkyne for click chemistry (**14**, Table 2). Incubation of mouse liver microsomes with probes and gel-based identification with semi-quantification by a gel-free method revealed the profile of CYPs labelled. This was followed up with gel-free analyses, which gave a similar list of CYP targets, although their rank differed compared to the gel-based data. The semi-quantitative method used, emPAI (Exponentially Modified Protein Abundance Index),⁷² is based on coverage of the protein sequence and was used here to indicate the relative abundance of different enzymes. Importantly, the gel-free approach enabled the identification of additional drug-metabolising enzymes labelled by pyrethroid probes.

Several recent studies have used chemical proteomics to profile the modification of proteins with the important antimalarial drug artemisinin (ART). ART is a plant-derived natural product with an unusual endoperoxide bridge (**15**, Figure 9); this feature is thought to be activated by iron species in the red blood cells where the malaria parasite resides, generating reactive radicals that can attack cellular structures or biomolecules. However, the identities of modified proteins and the mode of action of ART are not fully understood. Wang et al.⁷³ and Ismail et al.⁷⁴ independently reported several alkyne- or azide-tagged artemisinin probes (**16-18**, Figure 9); these were incubated with parasite cultures, proteins captured by CuAAC, enriched and identified, with semi-quantification via emPAI scoring in both cases. Wang et al. identified 124 enriched proteins and showed that several of the enzymes were inhibited by activated artemisinin in vitro; they further applied their alkynylated probe to study the mechanism of ART activation (an open question in the field) using iron chelators or inhibitors of haem production, concluding that haem is the dominant source of activation. Ismail et al. applied azido- and alkynyl-probes to identify around 60 enriched proteins; they also studied the effect of free iron chelators, observing some effect on labelling but also noting that this was not sufficient to explain ART modification of several proteins. Wang et al. also applied their probe in a human cancer cell line, since cancer cells have been proposed to be more susceptible to ART treatment than normal cells. A third recent study (Zhou et al.) also identified ART targets in cancer cells using an alkynylated probe **19**, similarly concluding that haem is the agent responsible for ART activation.⁷⁵ Together these three studies shed light on the likely protein targets of ART in live cells and provide tools to probe biological questions of the mode of action of this important drug.

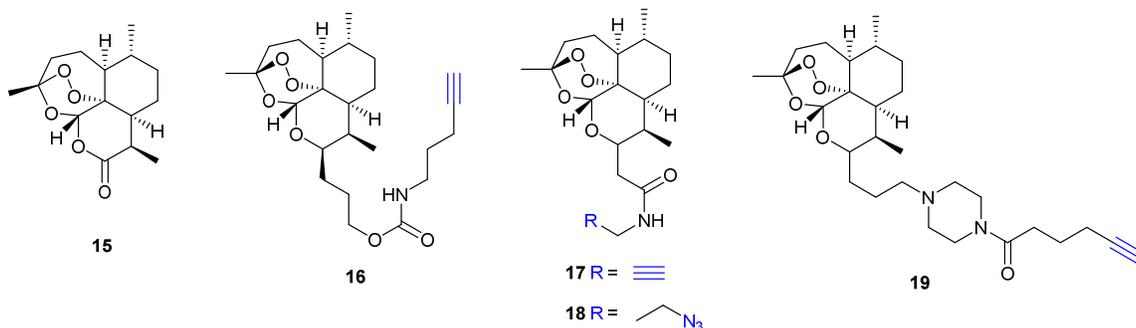


Figure 9: Antimalarial natural product artemisinin (ART, **15**) and reported probes **16-19**.⁷³⁻⁷⁵

Another label-free approach is spectral counting; this method estimates protein abundance by counting the number of MS² spectra assigned to a particular precursor ion and hence protein, although the method suffers from limited dynamic range and can only accurately quantify abundant proteins in a sample.⁷⁶ Zheng et al. recently used spectral counting to quantify the chemical proteomic targets of an alkyne-tagged version of the NP cerulenin (**20**, Table 2).⁷⁷ Cerulenin has been proposed to inhibit palmitoyltransferase (PAT) enzymes, which carry out the lipid modification of proteins known as palmitoylation, and fatty acid synthases by reaction of the α -keto-epoxide with active site cysteine residues. Zheng et al. modified the terminus of the long aliphatic chain of the NP with an alkyne, and carried out identification of targets of the cerulenin probe in melanoma cells. They compared this with cells pre-treated with cerulenin itself or another PAT inhibitor 2-bromopalmitate in competition experiments. Several proteins that responded to competition were indeed identified and further experiments validated that two of these were palmitoylated. The authors also profiled the labelling pattern in lysates from Hek293 cells transfected with different members of the DHHC family of PAT enzymes; however, mutation of the active site cysteines of DHHC3 and 4 did not abolish labelling, suggesting that the probe labels other sites and therefore is not 'activity-based'.

Spectral counting was also used by Pratt and co-workers in a study with alkynylated aspirin (**21**, Figure 10).⁷⁸ Well known as an anti-inflammatory drug, aspirin acts in part as an acetylating agent, covalently modifying key serine residues in cyclooxygenase enzymes, but the identities of other potential acetylation targets were unknown. HCT-15 colorectal cancer cells were labelled with aspirin probe and proteins identified by a standard chemical proteomics workflow. The authors performed three replicates and used spectral counting and other cut-off criteria to identify potential hits; acetylation was further confirmed for several histone proteins.

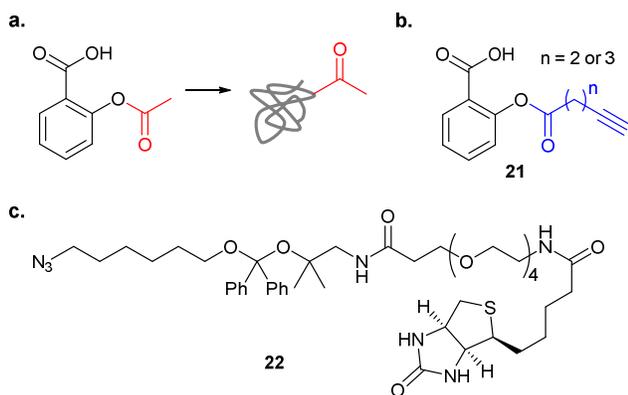


Figure 10: a. The NP aspirin alkylates proteins. b. Structure of aspirin-based probes **21** that have been reported.^{78,79} c. Cleavable linker **22** for the detection of aspirin probe-acylated peptide sites.⁷⁹

Label-based approaches: quantification at the peptide level. Apart from the bias of spectral counting towards larger proteins and the difficulties of quantifying low abundance hits, it is also challenging to define a cut-off when faced with a list of proteins and counts – to differentiate specific and non-specific binders for example. This is particularly true when dealing with probes that label many protein targets, as is the case for the examples of cerulenin and aspirin discussed above. Therefore, many recent studies have used alternative quantification methods. In an independent report, Wang et al. also applied alkyne-modified aspirin probes to identify protein targets of aspirin acetylation in cancer cells, but used quantification by label-based method iTRAQ.⁷⁹ In iTRAQ, isotope labelling is performed at the peptide level, in this case after peptide release by on-bead digest. Wang et al. also applied a cleavable linker strategy to orthogonally release modified peptides from the beads and simultaneously quantify sites of aspirin acetylation. In this case, the novel CuAAC reagent contained an acid-cleavable linker **22** (Figure 10), so that following on-bead digest and removal of most peptides, the modified peptides could be easily cleaved with 5% formic acid. For the ~1000 proteins quantified in the study, the modified site was identified in roughly half of cases, providing high confidence for the identification of these proteins as true hits of the probe. Although lysine was the main acetylated residue, the authors also found evidence for modification of other side chains, such as serine and threonine, although the high concentration of probe used may mean that some of these are not physiologically relevant – always a concern with chemical probes. Based on their high confidence hit list, Wang et al. predicted the pathways affected by aspirin acetylation and showed that aspirin induces autophagy and activates lysosomal function in cells.

Natural products with validated mode of action in one organism are sometimes transferable to less well characterised systems. In 2013, the Taunton et al. showed that the polyketide NP hypothemycin, which is known to inhibit several human kinases, is also active in *Trypanosoma brucei*, the protozoan parasite that causes the disease Human African Trypanosomiasis (HAT).⁸⁰ Hypothemycin contains an electrophilic Michael system which reacts with the nucleophilic thiol of select kinases. Taunton and colleagues created alkynylated probe version **23** via semi-synthetic modification of the NP - by simple replacement of a methyl ether with propargyl ether (Table 2). **23** had the same potency as the NP in parasite growth inhibition assays and labelled multiple bands after CuAAC. Only one clear band at 43 kDa showed dose-dependent saturation of signal intensity and could be out-competed by the parent NP, however, in an elegant example of using a gel-based read-out to assess specificity of probe labelling. Gel-based proteomics of the 43 kDa band revealed the kinase TbGSK3short as the dominant target but several other kinases were also identified. Noting that low abundance targets could be overlooked in proteomic analyses given that the probe also labelled abundant proteins non-specifically, the authors next carried out gel-free, iTRAQ-based quantitative chemical proteomics with samples incubated with their probe and increasing concentrations of the parent NP as competitor. Competition curves were then plotted with the quantified data for each protein, confirming TbGSK3short as a strong hit, but also showing that the kinase TbCLK1/2 was in fact more sensitive to modification by the NP. Follow-up experiments showed that the isoform TbCLK1, and not CLK2 or the more abundant kinase GSK3short, is likely the relevant target for parasite death. This study is therefore a nice example of how gel-free, quantitative proteomics combined with competition experiments and followed up with careful validation can deconvolute the biologically-relevant targets of a NP.

Further lessons on the promiscuity of some reactive NPs and the challenges of identifying the interaction responsible for bioactivity are provided by the NP andrographolide (Andro **24**, Figure 11), which has diverse anti-inflammatory and anti-cancer effects. Two independent studies have developed probes for Andro and applied them in diverse cancer cell lines.^{81,82} Andro possesses an α,β -unsaturated γ -lactone moiety and in the first study Wang et al. modified the pendant hydroxyl by esterification or replaced it with an amide for introduction of an alkyne (**25** and **26** Table 2).⁸¹ The amide probe gave far more intense labelling in cancer cell lines, likely because the ester is hydrolysed, resulting in loss of the alkyne tag; in fact elimination of the β -hydroxy group has been proposed to occur during reaction with proteins. After iTRAQ quantification of replicate experiments with the amide probe and filtering, the authors identified 75 possible hits, including one known target. Several hits were validated by enrichment and Western blot. Two proteins were then studied in more detail via overexpression and identification of cysteine residues as Andro binding sites. Finally, based on pathway analysis of identified proteins, the authors predicted that metastasis would be affected by Andro treatment; indeed, fairly low concentrations of Andro reduced migration and invasion in HCT116 cells. The second report on Andro probes took a different approach to probe design. Taking advantage of the susceptibility to elimination of any leaving groups appended to the α,β -unsaturated γ -lactone ring, the Yao group attached fluorogenic dyes at this position to provide a rapid read-out of probe reaction for imaging studies.⁸² The alkyne tag was instead attached to alcohol functionalities at the other end of the

molecule (Figure 11). Gel-based proteomics identified several hits; however, the authors also randomly selected recombinant proteins and showed that almost all were labelled by Andro in vitro. On this basis they concluded that Andro is a highly promiscuous compound, consistent with the study of Wang et al.

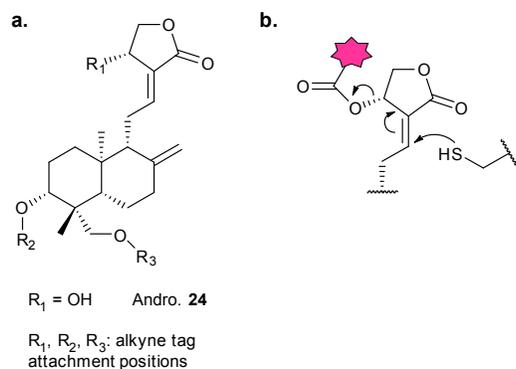


Figure 11: a. Andrographolide **24** and positions used to incorporate an alkyne in various reported probes (see also Table 2).^{81, 82} b. Fluorogenic probe reported by Yao et al.⁸² reacts to release a fluorophore upon attack by a nucleophilic cysteine residue.

iTRAQ uses reporter ion quantification: this means that the different reagents used to label peptides are isobaric – the same in mass – resulting in identical parent peptides which are differentiable only after fragmentation in the mass spectrometer (so-called ‘reporter ions’ are released). An alternative to iTRAQ for peptide-level mixing of samples is dimethyl labelling (DiMe), where quantification is carried out at the parent ion (MS^1) level and which has the advantage of using inexpensive reagents.⁸³ DiMe works by isotopic labelling of free amines in peptides via reductive amination with formaldehyde and sodium cyanoborohydride, both available with heavy isotopes incorporated such that peptides from different samples differ in mass by >4 Da (Figure 12a). A recent example from our lab used DiMe quantification to deconvolute the targets of fimbrolide NPs. Fimbrolides such as **27** (Figure 12b) are marine halogenated furanones that inhibit bacterial quorum sensing, intercellular communication via small molecules.⁸⁴ Zhao et al. synthesised alkyne-tagged variants of these halogenated furanones (e.g. **28**, Table 2) and applied them in live *Vibrio* cells – bacteria well-studied for their ability to engage in QS-induced bioluminescence.⁸⁵ After assessing the panel of probes for their ability to reduce bioluminescence in the bacteria, the fimbrolide and corresponding probe showing the best and most comparable activity were selected for further studies. DiMe was used to quantitatively compare proteins enriched from samples treated with 1) DMSO vehicle, 2) probe **28**, and 3) probe **28** following addition of excess NP **27** (competition). The competition experiment was used to select protein targets shared by the probe and NP, which included QS-related enzymes LuxS and LuxE. Binding of the probe to cysteine residues was further shown in these two cases via MS/MS analysis of digested recombinant protein. Several other targets were also identified, suggesting that fimbrolides act via multiple pathways in bacteria.

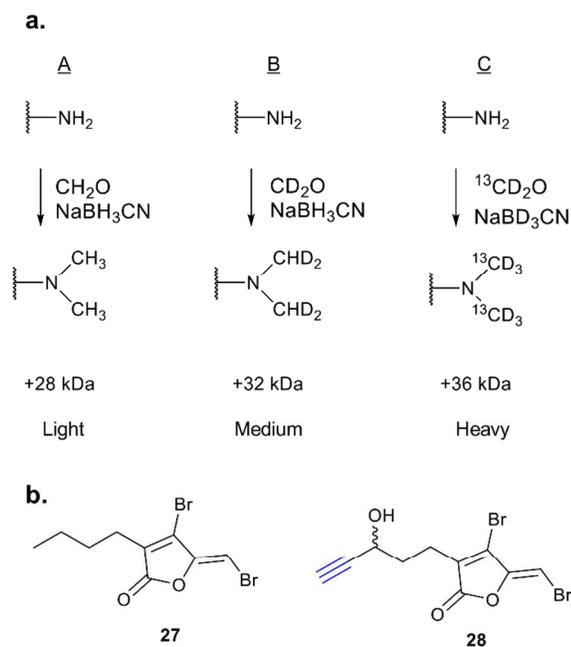


Figure 12: a. Dimethyl labelling (DiMe) on peptide amines for quantitative comparison of up to 3 samples. b. Fimbrolide **27** and corresponding probe **28**.⁸⁵

Label-based approaches: SILAC. Protein quantification by SILAC, where samples can be mixed at the lysate stage, is very attractive because it does not suffer from the problem that samples are combined at the peptide level – which is the case in methods such as iTRAQ and label free described above.

Acivicin (**29** Figure 13) is a NP with broad bioactivity that has been studied as a potential anti-cancer agent. Several enzymes had been identified as potentially inhibited by acivicin, but no unbiased and global analysis of potential protein targets had been carried out. The 4-chloroisoxazole moiety reacts with nucleophilic side chains in an addition-elimination reaction. An initial study by Orth et al. developed ACV-inspired 4-chloro- and 4-bromo-isoxazole probes and applied them in gel-based chemical proteomics analyses of potential targets in bacteria, identifying several dehydrogenases as likely targets.⁸⁶ Kreuzer et al. extended this approach in a liver cancer cell line and mouse tissue.⁸⁷ Since acivicin is a small compound with several potential sites for modification, Kreuzer et al. generated several different probes **30-33** to maximise the chances of identifying relevant targets. Gel-based chemoproteomics identified aldehyde dehydrogenases amongst the targets in mouse liver lysate and HepG2 human cells, and subsequent gel-based SILAC experiments, including quantification of competition with the NP, confirmed the hits in the latter. Interestingly, after prolonged incubation of cells with probe **30**, a new target, identified as carboxylesterase 1 (CES1), appeared. Recombinant CES1 itself was not labelled by **30**, however, although it was active, suggesting that a probe metabolite is the responsible agent. Follow-up in vitro studies and siRNA knockdown in human cells confirmed CES1 and one aldehyde dehydrogenase, ALDH4A1, as proteins important for viability.

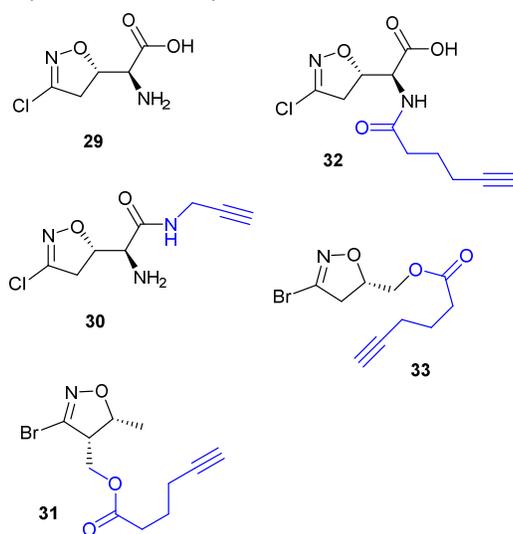


Figure 13: Acivicin **29** and probes **30-33**.⁸⁷

The marine diterpene eupalmerin acetate (EuPA) shows activity against cancer cell lines and in a mouse tumour xenograft model.⁸⁸ The mode of action of EuPA is presumed to arise from reaction of the exocyclic alkene with nucleophilic residues in proteins. Romo and colleagues explored the use of rhodium(II)-catalysed C–H amination chemistry to remodel and functionalise the structures of complex NPs.⁸⁹ EuPA was derivatised with an alkyne tag and the resulting probe **34** (Table 2) applied in HL-60 leukaemia cells; many bands were labelled but only a few of these could be outcompeted by the parent NP. To identify the responsive proteins, SILAC experiments were used to a) compare heavy and light EuPA-probe-treated cells, and b) compare cells treated with either probe alone or first treated with parent NP and then probe. Duplicate runs and a gel-free protocol were used, and hits selected on the basis of an enrichment of ~ 1.0 in experiment a), and a high ratio of probe/competition in experiment b). The three most promising hits were further validated by overexpression and labelling in Hek293 cells.

A few natural products already contain terminal alkynes that are potentially usable as tags. Nickel et al. devised a total synthesis of callyspongynic acid (**35**, Table 2) to test its potential as a probe.⁹⁰ Like other polyacetylenes, callyspongynic acid is lipid-like and was therefore hypothesised to act on lipid signalling pathways in the cell. In situ labelling in human cell lines revealed a number of bands and gel-free proteomics with SILAC in triplicates was used to quantify enrichment in NP-treated cells compared to a vehicle control. The authors note the high reproducibility of SILAC ratios between replicates and that the majority of the ~ 400 proteins quantified across all three replicates showed no enrichment, indicative of the very high sensitivity of modern MS instruments for detecting background binders on the enrichment beads. Gene ontology analysis of hits revealed that almost half were enzymes that could potentially bind callyspongynic acid or a metabolite for further metabolism or degradation. Non-enzyme putative targets were also identified and the authors speculate that these, mostly membrane, proteins may be enriched due to membrane perturbations of the NP.

As mentioned above, standard SILAC is limited by the available isotopically labelled amino acids to direct comparison of three samples mixed together, and cannot be used to analyse tissue samples, for example. However, use of a ‘spike-in’ standard can sometimes overcome these limitations. In spike-in SILAC a single reference sample is produced by SILAC labelling and then spiked into any number of different experimental samples.³⁰ Provided the spike-in reference has appropriate coverage of the proteome in the experimental samples, the reference can then act as an internal standard, allowing comparison across all samples, which can be produced independently and from different sources. Spike-in SILAC was used by the Tate group to identify the targets of the NP zerumbone, for quantifying concentration-dependent competition of alkyne probe with the parent NP.⁹¹ The cyclic sesquiterpene zerumbone, which has diverse anti-inflammatory activities, possess two electrophilic Michael acceptor systems with potential protein reactivity and so a probe **36** was synthesised with an alkyne attached at a distal site (Table 2). A large batch of SILAC heavy cells were labelled with the probe, and subsequent triplicate samples in standard cell culture medium were prepared with probe and increasing concentrations of parent NP. Statistical analysis of the enrichment ratios allowed a subtle analysis of the hits to

identify those where probe binding was most robustly out-competed. The profile of high confidence hits revealed proteins with diverse biological functions, underlining the likely pleiotropic mode of action of zerumbone.

Alternative quantification approaches. All the studies discussed above performed enrichment of proteins following CuAAC and then digested these for shotgun proteomics. In select cases, orthogonal release of the modified peptide was also applied to identify the site of reaction of the probe. However, an alternative approach is to digest proteins first and then carry out CuAAC and enrichment of peptides with a cleavable reagent. Although this results in a loss of information – since the majority of the peptides from a particular protein are not present – it is one strategy for focusing on the site of modification, and enrichment of peptides is also likely to be less biased and more efficient than enrichment of proteins. A recent study by Yang et al. applied such a peptidic site-centric approach to identify modification of proteins by the electrophilic lipid 4-hydroxy-2-nonenal (HNE).⁹² HNE is one of a number of endogenously produced lipid-derived electrophiles which modify proteins and thereby mediate responses to oxidative stress. In 2008, an azido-tagged HNE analogue was used to identify potential targets, including several heat shock proteins, by gel-based proteomics.⁹³ The follow-up study from Yang et al. then applied the peptide-centric approach with alkyne-tagged HNE **37** (Table 2) and novel isotopically-tagged, UV-cleavable CuAAC reagents (“Az-UV-biotin”).⁹² **37** was incubated with RKO colon cancer cells, cells lysed and proteins digested with trypsin; Az-UV-biotin reagents were then reacted with peptide mixtures, samples enriched on streptavidin beads and released by irradiation with UV light. The isotopic tagging in the Az-UV-biotin reagents facilitated identification of the peptides and quantification. As an example of how this could be applied, Yang et al. analysed pulse-chase type experiments in which cells were labelled with **37**, and then incubated in HNE-free medium for different periods of time. Using this approach they were able to measure the dynamics of HNE modification in live cells.

Photoaffinity labelling

This review has thus far focused on NPs that are intrinsically electrophilic, of which there are many examples. However, most NPs act by non-covalently binding their targets. Photoaffinity labelling (PAL) is a technology in which photoreactive groups are introduced into a probe so that a covalent link can be made between the probe and its target. PAL has been experiencing a resurgence in recent years, fuelled by the development of CuAAC chemistry allowing the introduction of minimal tags for target fishing. Benzophenone, aryl azide, aryl and alkyl diazirines are the most commonly used PAL groups (Figure 3).^{17, 18}

Benzophenone (Bpa) is a historically popular PAL group due to its stability towards chemical synthesis. Bpa-alkyne probes (e.g. **38**) of the glycopeptide antibiotic vancomycin was used to identify potential protein binders in pathogenic bacteria (Figure 14).⁹⁴ Eirich et al. synthesised several probes with the Bpa and alkyne moieties in different positions and applied them both in vitro and in situ. Two distinct protein hits were identified by gel-based proteomics in antibiotic-resistant *S. aureus* and *E. faecalis* as bifunctional autolysin (ATL) and a peptide ABC transporter respectively. Both hits were subsequently validated by recombinant expression and labelling, and the degradation of peptidoglycan by ATL shown to be inhibited by vancomycin. Eirich et al. reported another example of a Bpa photoclick probe **39**, based on pretubulysins, compounds derived from the NPs tubulysins (Figure 14).²⁷ Structure-activity relationships aided probe design, and the bulky Bpa was incorporated at a position known to be tolerant to modifications. Despite this careful design, there was a significant drop in activity. However, labelling was performed in HeLa cancer cells and one particularly prominent band was shown to respond to competition; this band was identified as tubulin, the known target of tubulysins. The authors further applied their probe for fluorescence imaging in cells.

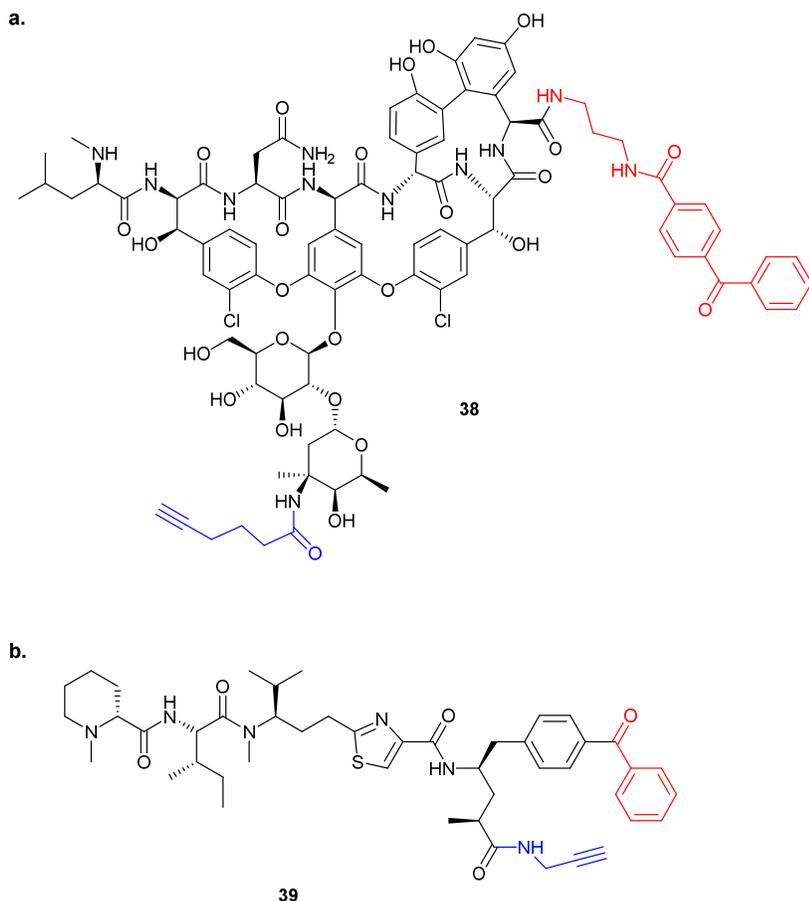


Figure 14: Benzophenone (Bpa) - containing photoprobes a. Probe **38** based on vancomycin.⁹⁴ b. Probe **39** based on pretubulysin.²⁷

The bulky size and possible propensity of Bpa to react preferentially with methionine⁹⁵ are disadvantages. Other PAL linkers such as aryl azides are best excited by short wavelengths of light (typically 265 to 275nm), potentially damaging proteins which absorb light in this range. The first example of incorporation of the small alkyl diazirine group into a NP was reported in 2007 by the Taunton group.⁹⁶ In this study, a probe version of a fungal cyclodepsipeptide (HUN-7293, **40**), which was known to inhibit cotranslational translocation, was designed. The probe, **41**, contained an alkyl diazirine-tagged unnatural amino acid derivative “photo-Leu” **42** and alkyne tag (Figure 15a). Crude ER microsomal fractions were incubated with the probe, irradiated, denatured with SDS and ligated to a fluorescent azide derivative by CuAAC. One clear major band was observed at a molecular weight consistent with that of the hypothesised target, Sec61 α , a subunit of the complex that forms the translocation channel at the ER membrane. This target was confirmed by labelling in microsomes depleted of Sec61 α . The authors concluded that the use of small photo- and click-tags which presumably minimally perturb the binding of the peptide to its target, together with the photoreactive properties of the diazirine, were key factors in the success of this identification strategy.

The diazirine-functionalised amino acid derivative photo-methionine **43** has also been reported.⁹⁷ A third unnatural diazirine-functionalised amino acid derivative, photo-proline **44**, was incorporated into a peptidomimetic with potent antibacterial activity against the gram-negative bacterium *Pseudomonas aeruginosa* (probe **45**, Figure 15).⁹⁸ The parent peptidomimetic was based on the structure of the naturally occurring membranolytic host-defense peptide protegrin I. A forward genetic screen suggested an outer membrane protein as a potential target, and to confirm direct binding in live cells the authors synthesised a biotin-functionalised analogue of the photoreactive peptidomimetic and showed that it did indeed label the putative target LptD.

Alkyl diazirines have also been incorporated into non-peptidic compounds. The NP staurosporine (STS) is a potent, promiscuous kinase inhibitor, which has found use as a general probe for kinases in different technological approaches: an STS analogue has been directly attached to resin,⁹⁹ and an aryl azide-based photocrosslinker and biotin-tagged probe synthesised for detecting binding partners in cell lysates.¹⁰⁰ In order to detect kinase targets of STS in live cells, Yao and colleagues prepared an STS analogue **46** with pendant alkyl diazirine and alkyne for click chemistry (Figure 15).¹⁰¹ Gel-based fluorescence and proteomic analysis of proteins labelled by the STS probe revealed overlapping but distinct profiles in lysates and live cells; five ‘in situ-specific’ targets were further validated by pull-down and Western blot. In addition to the 43 kinases identified in their approach, the authors also saw a large number of other proteins in the LC-MS/MS experiments. In this case, the authors chose to focus on kinases, although they note that kinase-related and -binding proteins could also be enriched by the probe. The Yao group expanded on this concept with the report of a series of versatile “minimal” alkyl-diazirine and alkyne-functionalised building blocks that can be incorporated into scaffolds of interest.¹⁰² They incorporated these into kinase inhibitors, including STS, and applied a cocktail of several probes to identify kinases in different cell and tissue types. However, in addition to the 94 kinases identified via the probes from rat kidney tissues, hundreds of non-kinases were apparently enriched. Whether these other putative targets represent true off-targets of the inhibitors or simply probe-specific background remained unexplored. This example illustrates a very real challenge of photoaffinity labelling for

discovery: how to identify true binders against a, potentially high, non-specific background. We discuss this point further in the conclusions, below.

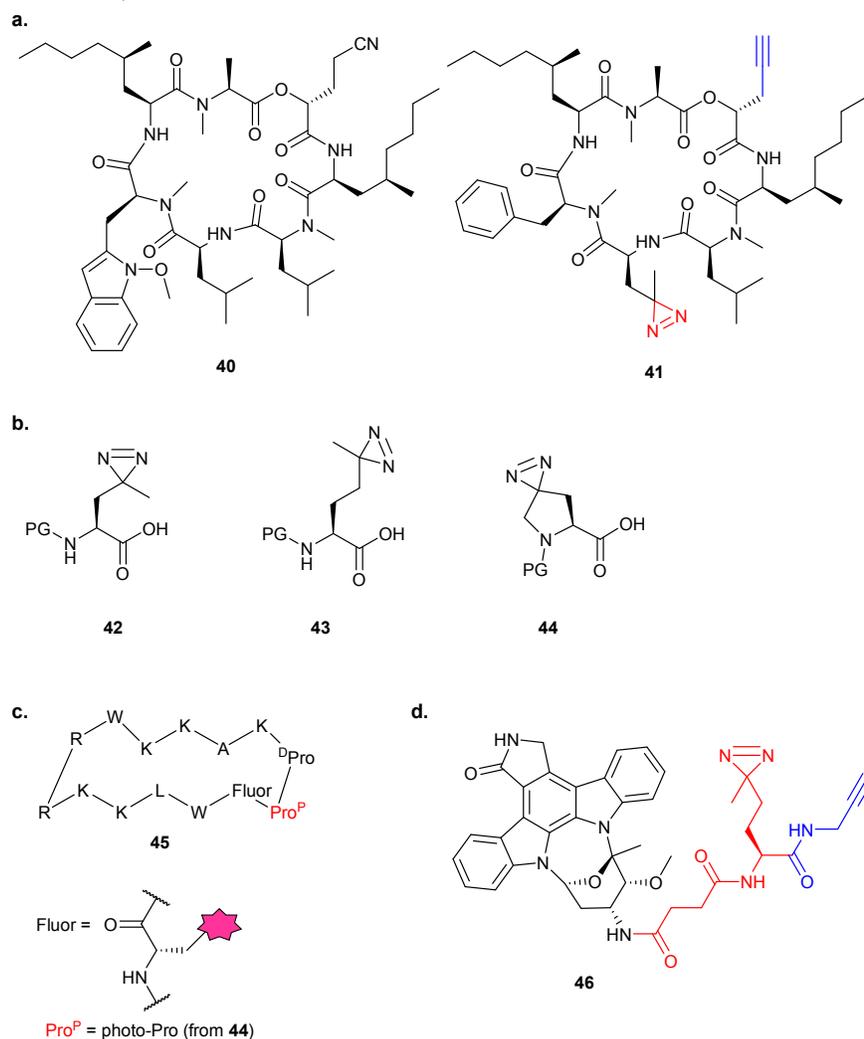


Figure 15: Diazirine-based photoprobes. **a.** Probe **41** based on a fungal depsipeptide NP **40**.⁹⁶ **b.** Amino acid analogues photo-Leu **42**, photo-Met **43** and photo-Pro **44** incorporating diazirines. PG = protecting group. **c.** Peptidomimetic probe **45**.⁹⁸ **d.** Staurosporine photoprobe **46**.¹⁰¹

Endogenous small molecules with unknown or unclear mode of action are also the subject of studies employing PAL. Acyl-homoserine lactones (e.g. **47**) produced by gram-negative bacteria are interesting compounds which, in addition to their roles in bacterial quorum sensing, also trigger responses in other organisms – such as yeast and human cells.¹⁰³ The Meijler group reported a series of alkyne-tagged probes incorporating a diazirine in different positions (**48-50** Figure 16a) and applied these tools to demonstrate labelling of the known bacterial target LasR.¹⁰⁴ Interestingly, the phenotypic response of bacteria and human cells to the probes differed – demonstrating that even the minimal alkyl diazirine modification can have strong influences in small molecules. A further example of probes based on endogenous small molecules is the application of alkyne- and diazirine-tagged cholesterol analogues to identify sterol-binding proteins in human cell lines by Cravatt and co-workers.¹⁰⁵ SILAC was used here to quantify enrichment over control, compare different probes and determine sensitivity of each protein via competition with native cholesterol.

Finally, our group recently applied PAL to identify the binding partners of the NP falcarinol.¹⁰⁶ Falcarinol and the related NP stipudiol are extracted from vegetables and have been reported to have anti-cancer properties, but falcarinol is also known to alkylate nucleophiles and isolated proteins.¹⁰⁷ Heydenreuter et al. synthesised the NP **51**, alkynylated probe **52** and an alkynylated version of the oxidised NP falcarinone **53** (Figure 16b).¹⁰⁶ In situ labelling in A549 cancer cells revealed a prominent band that was identified by gel-based proteomics as aldehyde dehydrogenase 2 (ALDH2). To further identify the possible reversible targets of the NP, the authors synthesised an alkyne probe **54** incorporating an alkyl diazirine. SILAC and gel-free MS was used to quantify the targets of the photoprobe. Several other proteins appeared in this analysis, in addition to the previously identified covalent target ALDH2; the latter was further shown to be inhibited by falcarinol in vitro.

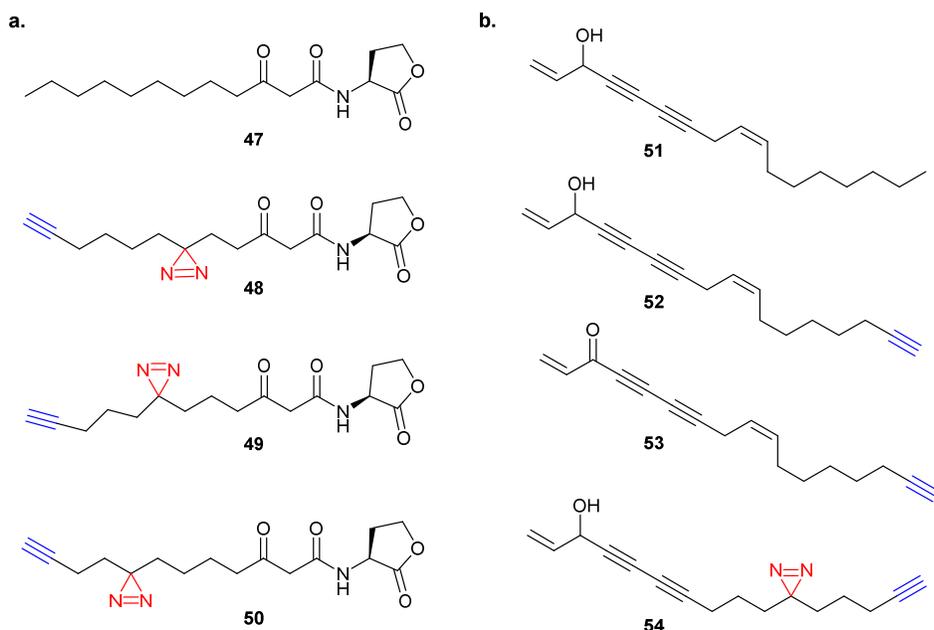


Figure 16: a. Acyl homoserine lactone **47** and reported probes **48-50**.¹⁰⁴ b. Falcarinol **51** and inspired probes **52-54**.¹⁰⁶

Competitive activity-based protein screening

Screening platforms

Historically, ABPP approaches have focused on compounds bearing electrophilic warheads that react with nucleophilic residues, often in enzyme active sites. This has led to the development of many probes that may preferentially label a specific enzyme class but are often quite promiscuous in labelling multiple proteins within that class. Many of the studies discussed above make use of competition experiments to provide evidence that a probe binds the same target as its parent compound. Combining these two concepts, small molecules of interest can be added as competitors to promiscuous probes to detect, for example, selective binders. Competitive ABPP (Figure 2) has not been very widely applied to elucidate the mode of action of natural products. However, given the potential complexity of probe synthesis and the potential perturbations to function from introducing even minimal tags into NPs, such approaches are worth considering. Many promiscuous ABPs preferentially react with cysteine residues. An example is iodoacetamide-alkyne **55** (Figure 17), which has been applied to profile cysteine reactive proteins in total proteomes.³⁴ Probes based on phenylsulfonate esters, epoxides, Michael acceptors and α -chloroacetamides all display proteome reactivity, with a preference for functional reactive nucleophilic residues.¹⁰⁸

Competitive mode for natural product target elucidation

Competitive ABPP can provide a different perspective and an orthogonal approach for identifying the binding partners of reactive endogenous molecules or NPs. The electrophilic lipid 4-hydroxy-2-nonenal (HNE **56**, Figure 17) has a Michael acceptor susceptible to reaction with cellular cysteines. Several probes of HNE have been reported that incorporate an azide or alkyne tag (see previously). However, in order to rapidly screen for 'hyper-reactive' cysteine residues that could act as cellular redox sensors by reacting with HNE, Cravatt and co-workers applied their isoTOP-ABPP (isotopic tandem orthogonal proteolysis activity-based protein profiling) method in combination with cysteine-reactive iodoacetamide-alkyne (IAA, Figure 17) for competitive ABPP.⁴⁰ isoTOP-ABPP uses heavy or light versions of an azide-TEV-biotin tag for click chemistry and enrichment of labelled proteins, allowing comparative quantification of samples.³⁴ Proteomes of a breast cancer cell line were treated with lipid electrophiles, including HNE, IAA added and samples analysed by gel-free MS.⁴⁰ Of the ~1000 cysteines quantified in the experiments, most were relatively unaffected, but a small subset responded to treatment with the lipids. To further investigate the susceptibilities of these cysteines, samples were treated with different concentrations of HNE and quantified relative to DMSO sample (here acting as the standard, by analogy with the spike-in SILAC approach). In this way, half-maximal inhibitory concentrations (IC_{50} values) could effectively be calculated. Follow-up investigations of the hit ZAK kinase revealed that HNE modification of a specific cysteine residue inhibited kinase activity. Furthermore, the previously reported alkyne-HNE **37** (Table 2)⁹³ also labelled this protein, providing further evidence that it is a target of lipid electrophiles. The authors discuss the value of applying these two complementary approaches, noting that the indirect nature of IAA profiling means that the competitor may in fact not bind the same cysteine as IAA but instead an adjacent residue that is sterically shielded by probe binding.

IAA is rarely used in living cells, being instead mostly applied in proteomes due to its high toxicity and the high concentrations required to saturate labelling. The Weerapana group recently reported caged bromomethyl ketone probe **57**, that is unreactive and can accumulate in live cells with low toxicity but is then unmasked by irradiation (Figure 17).¹⁰⁹ **57** was used to profile the effects of EGF stimulation, which enhances production of reactive oxygen species. Abegg et al. have reported another generic cysteine-reactive probe for use in live cells based on alkynyl benziodoxolone and carrying an azide tag for click chemistry (EBX, **58**, Figure 17).¹¹⁰ This probe proved to be more reactive than IAA and, although it was still toxic, the higher reactivity allowed lower concentrations to be used for in situ labelling. To demonstrate the utility of their probe, and compare it to IAA, Abegg et al. profiled the NP curcumin in competitive mode. Curcumin is found in the spice turmeric and has been reported to have anti-cancer and anti-inflammatory properties. The Michael acceptor of curcumin is potentially cysteine-reactive, and indeed did out-compete select

bands in both IAA and EBX treated samples. SILAC chemical proteomics was used to identify those labelled cysteines selectively outcompeted by curcumin, and this was further backed up by labelling with an alkyne-tagged curcumin derivative **59** (Figure 17). Of the 57 targets identified by competitive ABPP, 42 were also found as targets of **59**. The authors also noted that the complexity of the fragmentation patterns of alkyne-curcumin-labelled peptides meant that site identification with **59** was very challenging – whereas identifying the cysteine modification sites with the simpler IAA or EBX reagents was more straightforward. The hits identified with IAA and EBX also did not overlap completely, highlighting the value of having multiple reagents available to profile reactive cysteine residues. A further recent study independently applied the same probe **59** together with iTRAQ chemical proteomics to identify curcumin targets in colon cancer cells.¹¹¹ The authors identified nearly 200 enriched proteins, adding further evidence that curcumin is a promiscuous protein binder.

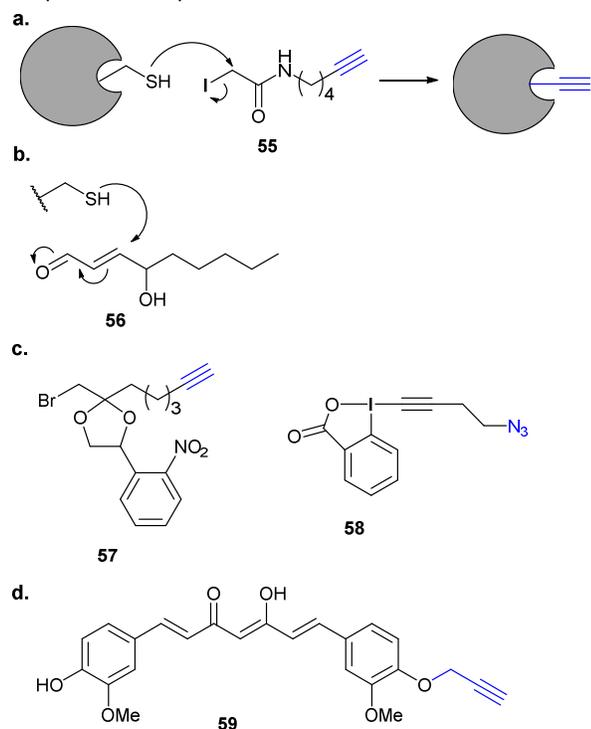


Figure 17: a. Reaction of iodoacetamide alkyne probes **55** with nucleophilic cysteine residues in proteins.³⁴ b. Reaction of 4-hydroxy-2-nonenal (HNE, **56**) with cysteine. c. Caged bromomethyl ketone electrophilic probe **57**¹⁰⁹ and alkynyl benzodioxolone reagent EBX, **58**, for labelling cysteines.¹¹⁰ d. Curcumin-alkyne probe **59** for compound-centric target identification.

Probes that are more specific than general cysteine-reactive compounds for particular subsets of proteins have also been applied in competitive ABPP. Many cellular proteins bind ATP, such as kinases and other ATPases; these include the chaperone HSP90, an important potential drug target in cancer. Nordin et al. applied an ATP acyl phosphate probe **60**,¹¹² which binds to the ATP pocket and acylates nearby lysine residues with a desthiobiotin tag, *in vitro* (Figure 18).¹¹³ Following enrichment, digest and LC-MS/MS, labelled peptides were identified directly based on MS² fragment spectra, aided by an inclusion list to improve detection. Competition with ATP was used to distinguish binding of the probe non-specifically to surface exposed lysines and binding in ATP sites. With this sensitive set-up in place, Nordin et al. profiled the ability of two NPs geldanamycin **61** and radicicol **62** (Figure 18), as well as several synthetic compounds, to inhibit HSP90 paralogs via binding to the ATP pocket. The compounds were applied to live cells and the authors also searched for off-targets of the compounds amongst other ATP-binding proteins.

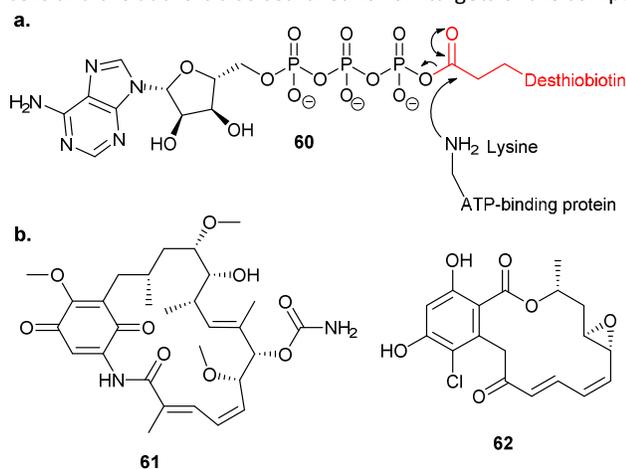


Figure 18: a. Reaction of lysine in active site of ATP-binding protein with ATP analogue probe **60**.¹¹³ b. NPs geldanamycin **61** and radicicol **62**.

In 2012, Stolze et al. combined competitive ABPP and fluorescent probes to reveal the mode of action of symplostatin 4, a secondary metabolite from cyanobacteria with potent anti-malarial activity.¹¹⁴ Symplostatin 4 features a potentially protein-reactive Michael acceptor and methyl-methoxy-pyrrolinone (mmp) unit. SAR studies first demonstrated the importance of the mmp unit for anti-parasitic activity, then the authors further characterised the phenotype arising from treatment of the malaria parasite *Plasmodium falciparum* with symplostatin 4 and derivatives. They observed a defect in the food vacuole, suggestive of a block in haemoglobin degradation. They synthesised probe **63** (Figure 19) and used the alkyne tag to attach a rhodamine fluorophore. Several proteins were covalently labelled with the rhodamine dye-coupled symplostatin 4 probe, leading the authors to hypothesise that proteases called falcipains, which play roles in parasite development within the red blood cell, were inhibited by the compound. Competitive labelling experiments using probe **64** based on the epoxide-containing cysteine protease inhibitor E-64 showed that symplostatin 4 inhibited specific members of this protease family. Quantification of the labelling revealed the sensitivities of the different proteases to the NP.

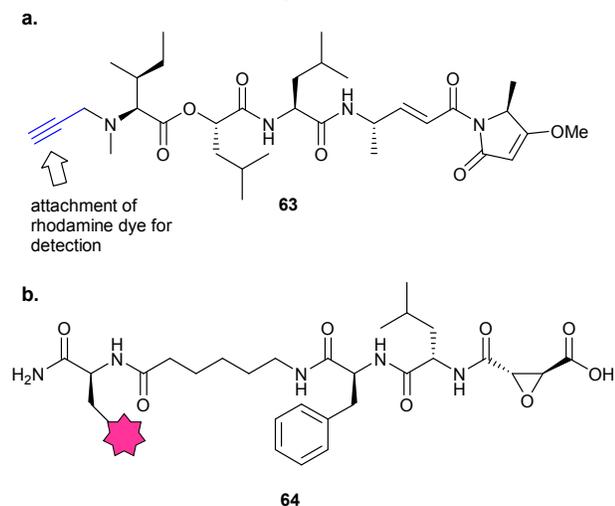


Figure 19: a. Symplostatin 4 precursor probe **63**. b. E64-inspired protease probe **64**.¹¹⁴

Competitive ABPP is intuitively most applicable to reactive NPs. A structural modification that converts an unreactive NP to a reactive one may change the bioactivity or mode of action, and so is a risky strategy for target identification. However, there has always been interest in generating modified versions of NPs and recently Porco, Cravatt and co-workers studied a derivative of the NP rocaglate **65** incorporating a reactive β -lactone moiety (**66**, Figure 20).¹⁶ Rocaglates have diverse bioactivity and β -lactones have proven value as electrophilic traps for nucleophilic enzyme residues.⁵² Hypothesising that the β -lactone would direct the compound to enzymes possessing nucleophilic residues, the authors conducted competitive ABPP against a promiscuous fluorophosphonate probe. SILAC quantification of enriched proteins with and without competition revealed four out of 40 serine hydrolases that were specifically and dose-dependently inhibited by the novel compound. Competition also worked well in situ, suggesting that the new rocaglate derivative could serve as a tool compound for future studies on specific serine hydrolases.

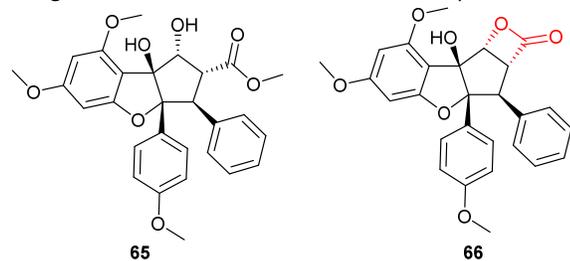


Figure 20: NP Rocaglate **65** and β -lactone-functionalised derivative **66**.¹⁶

Finally, Cravatt and co-workers developed a high-throughput competitive screening platform termed fluopol-ABPP to screen an enzyme in vitro for compounds that bind.¹¹⁵ The method works on the premise that if a relatively large ABP functionalised with a fluorophore is attached to a protein of interest, the extent of depolarisation of polarised light emitted by the excited fluorophore will be low. If instead of the probe, a competitor compound is bound, the probe will be free to tumble rapidly in solution and will emit depolarised light (Figure 21). Bachovchin et al. applied this approach to the putative serine hydrolase RBBP9. Approximately 18000 compounds were screened using RBBP9 and rhodamine-functionalised fluorophosphonate probe **67** that binds serine hydrolases, identifying several hits including the bioactive natural product emetine **68**. Gel-based fluorescence ABPP was further used to refine and study the hits.

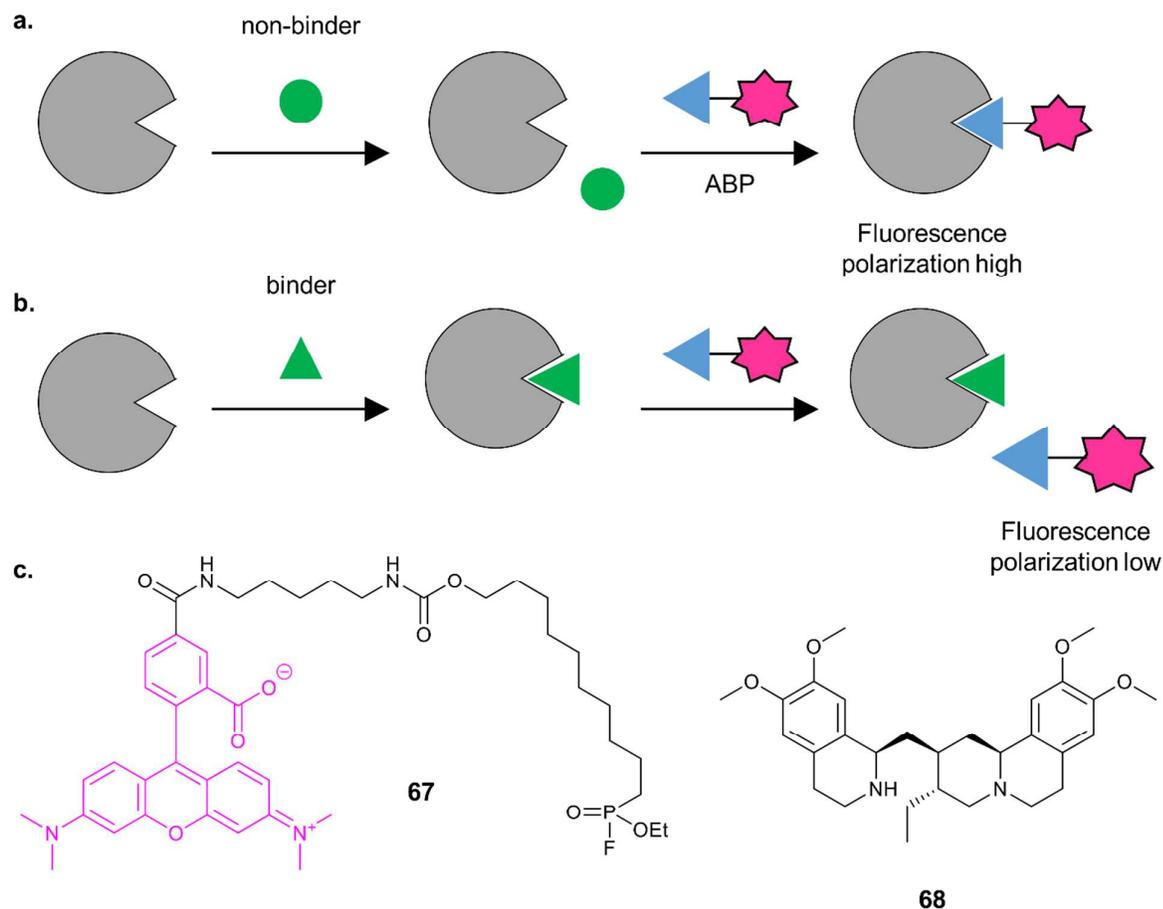


Figure 21: FluoPOL ABPP screening. a. If a compound does not bind the protein in the assay, the activity-based probe (ABP) can bind, resulting in highly fluorescently polarised light because the fluorophore tumbles slowly. b. In contrast, if a compound binds, the ABP does not and tumbles rapidly in solution, resulting in low fluorescence polarisation. Figure adapted from Bachovchin et al.¹¹⁵ c. Fluorophosphonate probe **67** and NP hit emetine **68**.¹¹⁵

Conclusions

The main strength of chemical proteomics is its ability to provide an unbiased, global and quantitative analysis of protein binding partners. By combining a cell-permeable probe with chemical proteomics, often facilitated by a two-step bioorthogonal ligation protocol, it is possible to work in live cells ('in situ') and with endogenous protein levels. Probes can also be applied in diverse systems. And using sensitive in-gel fluorescence detection with well-characterised probes, or using quantitative proteomics, it is possible to quantify dose-response, response to inhibitor or competitor and even measure IC_{50} of probe-protein binding directly in the cell.

These advantages make chemical proteomics a very powerful method for mode of action studies. However, as with all technologies, there are limitations and difficulties. We have mentioned many of these in the main part of the review, but in the following section we will explicitly discuss the main problems and some potential solutions.

Limitations, pitfalls and potential solutions

One common concern in chemical proteomics that we have highlighted throughout this review is how to deal with promiscuity and background binders. There are two issues here: firstly, at high concentrations, photoreactive and electrophilic probes will almost always label proteins non-specifically to some extent, and some proteins may be probe-specific hits (i.e. not targets of the parent compound). Secondly, some proteins will bind non-specifically to affinity resin, and there may be background from the CuAAC reaction, where used.

The second problem is often addressed by comparison with controls combined with quantitative proteomics. However, it can still be challenging to deal with both highly abundant protein hits (that may well be present in controls too) and low abundance proteins that are masked by background. Furthermore, choosing a cut-off, even with quantitative data, is difficult. Some resources are available to help assess whether a putative hit is a 'sticky' or abundant background binder; for example, in one study, the background binders from different resins and affinity enrichment protocols were analysed.¹¹⁶ The issue of background is in part due to the indirect nature of most chemical proteomics identification strategies: in general proteins are identified as hits by their enrichment in probe-treated sample over some control and not by direct identification of probe modification. As discussed here, use of a cleavable linker is one approach to this problem; examples discussed above include identification of the sites of modification of cysteine-reactive probes,⁴⁰ the use of isotopically labelled azide reagents to capture peptides from proteins labelled with a reactive lipid probe⁹² and the identification of sites of aspirin alkylation.⁷⁹ Such linkers work best with probes that are

incorporated metabolically at defined positions in a protein or probes that react predictably with certain residues (e.g. cysteine) and without excessive fragmentation. However, some of these limitations may be overcome as de novo-aided sequencing approaches improve and MS instruments become increasingly sensitive.

The issue of probe promiscuity is difficult to deal with because it is so context-dependent. Optimising probe concentration and other labelling conditions can sometimes help, and quantifying enrichment in the presence and absence of a competitor (typically the parent NP) is one approach widely used to test whether a protein is a probe-specific hit. This approach was used to help identify the specific targets of the NPs zerumbone,⁹¹ eupalmerin acetate,⁸⁹ hypothemicin⁸⁰ and fimbrolides,⁸⁵ for example. Multiple probes with modifications at different positions can be used to reduce the risk of probe-related off-targets and to ensure coverage of as many targets as possible (an approach used to identify possible targets of acivicin, for example).⁸⁷ Choice of photoreactive group may also have an influence on the profile of hits, although very few comparative studies exist, and it can also be challenging to introduce a photolinker into a small scaffold without affecting compound activity. One recent study analysed the hits of simple compounds incorporating a benzophenone, aryl azide or alkyl diazirine photocrosslinker, as well as alkyne tag, with the aim of providing a list of photolinker-specific targets.¹¹⁷ However, this study applied gel-based proteomics, which as discussed above suffers from several technical limitations. Further work in this area may be helpful in providing resources to aid researchers in assessing whether putative targets are genuine or related to the photocrosslinker moiety itself. Promiscuity of photoprobes is a common challenge: as discussed above in the context of kinase inhibitors, it is difficult to identify 'true' binders against a background of likely probe-specific targets. In our experience, chemical proteomics experiments, particularly those with technical challenges (e.g. photocrosslinkers) or where many proteins are addressed, benefit from the following: 1. The application of quantitative proteomics and rigorous statistical evaluation of data; 2. Control experiments, again performed quantitatively e.g. competition with parent compound; 3. Follow-up validation of putative targets.

Finally, newly described alternative approaches, such as CETSA (cellular thermal shift assay)¹¹⁸ or TPP (Thermal proteome profiling),¹¹⁹ for measuring compound-protein binding in cells complement chemical proteomics and are more readily applicable to compounds with non-covalent binding modes. CETSA or TPP relies on the fact that binding of a compound often stabilises the target protein in a manner that can be detected by a shift in the melting temperature of that protein. Thus, lysates or cells are treated with a compound, and sample aliquots heated to different temperatures; the aliquots are then lysed if necessary and centrifuged to remove aggregated (denatured) proteins; the soluble portion is analysed to detect the amount of undenatured protein remaining. These approaches are only applicable to protein-ligand interactions showing a thermal shift and, currently, to soluble proteins, but are nonetheless very powerful.

Linking target binding with bioactivity, and dealing with pleiotropic effects are broad challenges faced by any approach that seeks to identify the proteins that NPs bind. Structure-activity relationship studies (e.g. comparing the effects of compounds with different levels of bioactivity) and in-depth biological characterisation are typically used for this next level of mode-of-action deconvolution.

Current and future challenges

In 2012, Moellering and Cravatt proposed that chemical proteomics can help bridge the gulf between early drug discovery efforts that focus on identifying interesting drug targets, often genetically, and later efforts where the emphasis is on the chemistry or pharmacology of small molecules.¹²⁰ They note that chemical probes provide a way to detect engagement of a compound with its target protein in living systems, and suggest that this strength could be leveraged to bring chemistry-driven approaches into contact with early stages of drug discovery – validating drug targets chemically, identifying novel druggable proteins, and clarifying the mode of action of compounds identified in phenotypic screens. Although chemical proteomics has made huge contributions to our understanding of particular enzyme classes, large portions of the proteome, and many organisms, remain unexplored. It is also interesting that most clinically approved drugs target membrane receptors,¹²¹ but that relatively few chemical probes have been developed to address this large class of proteins. Affinity chromatography and chemical proteomics in cell lysates are difficult to apply to membrane proteins since cell lysis, particularly if detergents are used, can seriously disrupt these proteins. Other techniques such as CETSA/TPP also currently cannot address membrane proteins. However, in situ chemical probes should, in principle, be applicable to capturing membrane targets. Nevertheless, the vast majority of in situ chemical probes reported so far target either enzymes with fairly well-defined pockets (proteases, kinases) or are promiscuous electrophiles that bind soluble metabolic enzymes; this is likely due to the field's historical focus on activity-based probes. Several photoaffinity natural product-based probes that target membrane proteins have been reported, including mimics of vancomycin and a peptidomimetic, but the number of examples is still small. The lack of chemical probes targeting membrane receptors may be due to the facts that these proteins typically bind ligands non-covalently (necessitating the use of photoreactive probes or probes designed to react with a non-catalytic nucleophile in the binding pocket), often have multiple active and inactive conformations complicating ligand binding, and are typically of low abundance. As illustrated by the examples given above, with photoaffinity labelling it is particularly challenging to tease out true targets against background labelling, and this may have thus far hindered the development of probes for low abundance membrane receptors. However, in situ chemical proteomics is uniquely positioned to tackle this challenge. Improved sensitivity of mass spectrometry instrumentation may aid in the detection of low abundance hits or allow lower concentrations of probe to be used (which can reduce non-specific binding and hence off-targets) and focused efforts to mine chemical space should yield more specific high affinity probes. Design of covalent probes for G-protein coupled receptors is also an active area of research, for example.¹²² Natural products provide a rich resource of potential probes that could be applied in underexplored areas of chemical proteomics.

Chemical proteomics, by its nature, focuses on proteins, but small molecules also act on other biomolecules (DNA, RNA) or structures (membranes). However, many of the same principles of chemical proteomics could be developed for direct detection of binding of small molecules to oligonucleotides. For example, clickable Pt(II)-complexes were used to identify rRNA interactors of the drug picoplatin in *Saccharomyces cerevisiae*.¹²³ Studies applying such complexes were recently reviewed by White et al.¹²⁴

Finally, the indirect nature of many identification strategies is a limitation that we have discussed in depth here. However, cleavable linkers and advances in quantitative proteomic technology continue to improve. We anticipate that efforts to make these

technologies more broadly applicable and the development of innovative tools will continue to be at the cutting edge of chemical proteomics.

Acknowledgements

We gratefully acknowledge the European Commission FP7 programme grants 625528 and 259024 for financial support. We thank Dr. Pavel Kielkowski (Sieber lab) for proofreading.

Notes and references

1. T. Bottcher, M. Pitscheider and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 2680-2698.
2. J. J. La Clair, *Nat. Prod. Rep.*, 2010, **27**, 969-995.
3. S. Ziegler, V. Pries, C. Hedberg and H. Waldmann, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 2744-2792.
4. P. Yang and K. Liu, *Chembiochem*, 2015, **16**, 712-724.
5. M. Gersch, J. Kreuzer and S. A. Sieber, *Nat. Prod. Rep.*, 2012, **29**, 659-682.
6. D. M. Patterson, L. A. Nazarova and J. A. Prescher, *ACS Chem. Biol.*, 2014, **9**, 592-605.
7. C. W. Tornøe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057-3064.
8. V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem. Int. Ed. Engl.*, 2002, **41**, 2596-2599.
9. G. C. Rudolf and S. A. Sieber, *Chembiochem*, 2013, **14**, 2447-2455.
10. A. E. Speers and B. F. Cravatt, *Chem. Biol.*, 2004, **11**, 535-546.
11. G. Charron, M. M. Zhang, J. S. Yount, J. Wilson, A. S. Raghavan, E. Shamir and H. C. Hang, *J. Am. Chem. Soc.*, 2009, **131**, 4967-4975.
12. N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem. Biol.*, 2006, **1**, 644-648.
13. B. F. Cravatt, A. T. Wright and J. W. Kozarich, *Annu. Rev. Biochem.*, 2008, **77**, 383-414.
14. M. Fonovic and M. Bogoy, *Expert Rev. Proteomics*, 2008, **5**, 721-730.
15. D. A. Shannon and E. Weerapana, *Curr. Opin. Chem. Biol.*, 2015, **24**, 18-26.
16. N. J. Lajkiewicz, A. B. Coggnetta, 3rd, M. J. Niphakis, B. F. Cravatt and J. A. Porco, Jr., *J. Am. Chem. Soc.*, 2014, **136**, 2659-2664.
17. E. Smith and I. Collins, *Future Med. Chem.*, 2015, **7**, 159-183.
18. G. W. Preston and A. J. Wilson, *Chem. Soc. Rev.*, 2013, **42**, 3289-3301.
19. L. Dubinsky, B. P. Krom and M. M. Meijler, *Bioorg. Med. Chem.*, 2012, **20**, 554-570.
20. J. Lehmann, M. H. Wright and S. A. Sieber, *Chem. Eur. J.*, 2016, **22**, 4666-4678.
21. J. M. Krysiak, J. Kreuzer, P. Macheroux, A. Hermetter, S. A. Sieber and R. Breinbauer, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 7035-7040.
22. A. T. Wright and B. F. Cravatt, *Chem. Biol.*, 2007, **14**, 1043-1051.
23. A. T. Wright, J. D. Song and B. F. Cravatt, *J. Am. Chem. Soc.*, 2009, **131**, 10692-10700.
24. R. Ekkebus, S. I. van Kasteren, Y. Kulathu, A. Scholten, I. Berlin, P. P. Geurink, A. de Jong, S. Goerdalay, J. Neefjes, A. J. Heck, D. Komander and H. Ovaa, *J. Am. Chem. Soc.*, 2013, **135**, 2867-2870.
25. B. R. Lanning, L. R. Whitby, M. M. Dix, J. Douhan, A. M. Gilbert, E. C. Hett, T. O. Johnson, C. Joslyn, J. C. Kath, S. Niessen, L. R. Roberts, M. E. Schnute, C. Wang, J. J. Hulce, B. Wei, L. O. Whiteley, M. M. Hayward and B. F. Cravatt, *Nat. Chem. Biol.*, 2014, **10**, 760-767.
26. W. P. Heal, M. H. Wright, E. Thinon and E. W. Tate, *Nat. Protoc.*, 2012, **7**, 105-117.
27. J. Eirich, J. L. Burkhart, A. Ullrich, G. C. Rudolf, A. Vollmar, S. Zahler, U. Kazmaier and S. A. Sieber, *Mol. Biosyst.*, 2012, **8**, 2067-2075.
28. S. Nahnsen, C. Bielow, K. Reinert and O. Kohlbacher, *Mol. Cell. Proteomics*, 2013, **12**, 549-556.
29. M. Bantscheff, S. Lemeer, M. M. Savitski and B. Kuster, *Anal. Bioanal. Chem.*, 2012, **404**, 939-965.
30. T. Geiger, J. R. Wisniewski, J. Cox, S. Zanivan, M. Kruger, Y. Ishihama and M. Mann, *Nat. Protoc.*, 2011, **6**, 147-157.
31. F. Meissner and M. Mann, *Nat. Immunol.*, 2014, **15**, 112-117.
32. A. E. Speers and B. F. Cravatt, *J. Am. Chem. Soc.*, 2005, **127**, 10018-10019.
33. E. Weerapana, A. E. Speers and B. F. Cravatt, *Nat. Protoc.*, 2007, **2**, 1414-1425.
34. E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. Dillon, D. A. Bachovchin, K. Mowen, D. Baker and B. F. Cravatt, *Nature*, 2010, **468**, 790-795.
35. Y. Y. Yang, M. Gammel, A. S. Raghavan, G. Charron and H. C. Hang, *Chem. Biol.*, 2010, **17**, 1212-1222.
36. M. Broncel, R. A. Serwa, P. Ciepla, E. Krause, M. J. Dallman, A. I. Magee and E. W. Tate, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 5948-5951.
37. M. Broncel, R. A. Serwa, T. D. Bunney, M. Katan and E. W. Tate, *Mol. Cell. Proteomics*, 2015, DOI: 10.1074/mcp.O115.054429.
38. Y. Yang, H. Hahne, B. Kuster and S. H. Verhelst, *Mol. Cell. Proteomics*, 2013, **12**, 237-244.
39. A. L. Mackinnon and J. Taunton, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 55-73.
40. C. Wang, E. Weerapana, M. M. Blewett and B. F. Cravatt, *Nat. Methods*, 2014, **11**, 79-85.
41. W. L. Yu, G. Guizzunti, T. L. Foley, M. D. Burkart and J. J. La Clair, *J. Nat. Prod.*, 2010, **73**, 1659-1666.
42. M. D. Alexander, M. D. Burkart, M. S. Leonard, P. Portonovo, B. Liang, X. Ding, M. M. Joulie, B. M. Gullledge, J. B. Aggen, A. R. Chamberlin, J. Sandler, W. Fenical, J. Cui, S. J. Gharpure, A. Polosukhin, H. R. Zhang, P. A. Evans, A. D. Richardson, M. K. Harper, C. M. Ireland, B. G. Vong, T. P. Brady, E. A. Theodorakis and J. J. La Clair, *Chembiochem*, 2006, **7**, 409-416.
43. L. Trzoss, T. Fukuda, L. V. Costa-Lotufu, P. Jimenez, J. J. La Clair and W. Fenical, *Proc Natl Acad Sci U S A*, 2014, **111**, 14687-14692.
44. W. L. Yu, B. D. Jones, M. Kang, J. C. Hammons, J. J. La Clair and M. D. Burkart, *J. Nat. Prod.*, 2013, **76**, 817-823.
45. C. Cassiano, L. Margarucci, R. Esposito, R. Riccio, A. Tosco, A. Casapullo and M. C. Monti, *Chem Commun (Camb)*, 2014, **50**, 6043-6045.

46. C. Sibbersen, L. Lykke, N. Gregersen, K. A. Jorgensen and M. Johannsen, *Chem Commun (Camb)*, 2014, **50**, 12098-12100.
47. S. Wang, W. Xie, X. Zhang, X. Zou and Y. Zhang, *Chem Commun (Camb)*, 2012, **48**, 5907-5909.
48. M. A. Nessen, G. Kramer, J. Back, J. M. Baskin, L. E. Smeenk, L. J. de Koning, J. H. van Maarseveen, L. de Jong, C. R. Bertozzi, H. Hiemstra and C. G. de Koster, *J. Proteome Res.*, 2009, **8**, 3702-3711.
49. E. M. Schneider, M. Zeltner, V. Zlateski, R. N. Grass and W. J. Stark, *Chem Commun (Camb)*, 2015, DOI: 10.1039/c5cc07692g.
50. I. Staub and S. A. Sieber, *J. Am. Chem. Soc.*, 2008, **130**, 13400-13409.
51. I. Staub and S. A. Sieber, *J. Am. Chem. Soc.*, 2009, **131**, 6271-6276.
52. T. Bottcher and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2008, **47**, 4600-4603.
53. E. Romeo, S. Ponzano, A. Armirotti, M. Summa, F. Bertozzi, G. Garau, T. Bandiera and D. Piomelli, *ACS Chem. Biol.*, 2015, **10**, 2057-2064.
54. O. Kocaoglu, R. A. Calvo, L. T. Sham, L. M. Cozy, B. R. Lanning, S. Francis, M. E. Winkler, D. B. Kearns and E. E. Carlson, *ACS Chem. Biol.*, 2012, **7**, 1746-1753.
55. M. Pitscheider, N. Mausbacher and S. A. Sieber, *Chem. Sci.*, 2012, **3**, 2035-2041.
56. E. Zeiler, N. Braun, T. Bottcher, A. Kastenmuller, S. Weinkauff and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 11001-11004.
57. A. List, E. Zeiler, N. Gallastegui, M. Rusch, C. Hedberg, S. A. Sieber and M. Groll, *Angew. Chem. Int. Ed. Engl.*, 2013, **53**, 571-574.
58. T. Bottcher and S. A. Sieber, *J. Am. Chem. Soc.*, 2010, **132**, 6964-6972.
59. M. B. Nodwell, H. Menz, S. F. Kirsch and S. A. Sieber, *Chembiochem*, 2012, **13**, 1439-1446.
60. O. A. Battenberg, Y. Yang, S. H. Verhelst and S. A. Sieber, *Mol. Biosyst.*, 2013, **9**, 343-351.
61. M. J. Evans, A. Saghatelian, E. J. Sorensen and B. F. Cravatt, *Nat. Biotechnol.*, 2005, **23**, 1303-1307.
62. M. J. Evans, G. M. Morris, J. Wu, A. J. Olson, E. J. Sorensen and B. F. Cravatt, *Mol. Biosyst.*, 2007, **3**, 495-506.
63. P. Y. Yang, K. Liu, M. H. Ngai, M. J. Lear, M. R. Wenk and S. Q. Yao, *J. Am. Chem. Soc.*, 2010, **132**, 656-666.
64. M. S. Ravindran, S. P. Rao, X. Cheng, A. Shukla, A. Cazenave-Gassiot, S. Q. Yao and M. R. Wenk, *Mol. Cell. Proteomics*, 2014, **13**, 435-448.
65. D. L. Boger and D. S. Johnson, *Proc Natl Acad Sci U S A*, 1995, **92**, 3642-3649.
66. L. F. Tietze, J. M. von Hof, M. Muller, B. Krewer and I. Schuberth, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 7336-7339.
67. T. Wirth, K. Schmuck, L. F. Tietze and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 2874-2877.
68. T. Wirth, G. F. Pestel, V. Ganal, T. Kirmeier, I. Schuberth, T. Rein, L. F. Tietze and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 6921-6925.
69. M. F. Koch, S. Harteis, I. D. Blank, G. Pestel, L. F. Tietze, C. Ochsenfeld, S. Schneider and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2015, DOI: 10.1002/anie.201505749.
70. T. Shibata, F. Nakashima, K. Honda, Y. J. Lu, T. Kondo, Y. Ushida, K. Aizawa, H. Suganuma, S. Oe, H. Tanaka, T. Takahashi and K. Uchida, *J. Biol. Chem.*, 2014, **289**, 32757-32772.
71. H. M. Ismail, P. M. O'Neill, D. W. Hong, R. D. Finn, C. J. Henderson, A. T. Wright, B. F. Cravatt, J. Hemingway and M. J. Paine, *Proc Natl Acad Sci U S A*, 2014, **110**, 19766-19771.
72. Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber and M. Mann, *Mol. Cell. Proteomics*, 2005, **4**, 1265-1272.
73. J. Wang, C. J. Zhang, W. N. Chia, C. C. Loh, Z. Li, Y. M. Lee, Y. He, L. X. Yuan, T. K. Lim, M. Liu, C. X. Liew, Y. Q. Lee, J. Zhang, N. Lu, C. T. Lim, Z. C. Hua, B. Liu, H. M. Shen, K. S. Tan and Q. Lin, *Nat Commun*, 2015, **6**, 10111.
74. H. M. Ismail, V. Barton, M. Phanchana, S. Charoensutthivarakul, M. H. Wong, J. Hemingway, G. A. Biagini, P. M. O'Neill and S. A. Ward, *Proc Natl Acad Sci U S A*, 2016, **113**, 2080-2085.
75. Y. Zhou, W. Li and Y. Xiao, *ACS Chem. Biol.*, 2016, DOI: 10.1021/acscchembio.5b01043.
76. W. M. Old, K. Meyer-Arendt, L. Aveline-Wolf, K. G. Pierce, A. Mendoza, J. R. Sevinisky, K. A. Resing and N. G. Ahn, *Mol. Cell. Proteomics*, 2005, **4**, 1487-1502.
77. B. Zheng, S. Zhu and X. Wu, *ACS Chem. Biol.*, 2015, **10**, 115-121.
78. L. A. Bateman, B. W. Zaro, S. M. Miller and M. R. Pratt, *J. Am. Chem. Soc.*, 2013, **135**, 14568-14573.
79. J. Wang, C. J. Zhang, J. Zhang, Y. He, Y. M. Lee, S. Chen, T. K. Lim, S. Ng, H. M. Shen and Q. Lin, *Sci. Rep.*, 2015, **5**, 7896.
80. M. Nishino, J. W. Choy, N. N. Gushwa, J. A. Osés-Prieto, K. Koupparis, A. L. Burlingame, A. R. Renslo, J. H. McKerrow and J. Taunton, *Elife*, 2013, **2**, e00712.
81. J. Wang, X. F. Tan, V. S. Nguyen, P. Yang, J. Zhou, M. Gao, Z. Li, T. K. Lim, Y. He, C. S. Ong, Y. Lay, J. Zhang, G. Zhu, S. L. Lai, D. Ghosh, Y. K. Mok, H. M. Shen and Q. Lin, *Mol. Cell. Proteomics*, 2014, **13**, 876-886.
82. L. Li, H. Wijaya, S. Samanta, Y. Lam and S. Q. Yao, *Sci. Rep.*, 2015, **5**, 11522.
83. P. J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed and A. J. Heck, *Nat. Protoc.*, 2009, **4**, 484-494.
84. M. Hentzer, K. Riedel, T. B. Rasmussen, A. Heydorn, J. B. Andersen, M. R. Parsek, S. A. Rice, L. Eberl, S. Molin, N. Hoiby, S. Kjelleberg and M. Givskov, *Microbiology*, 2002, **148**, 87-102.
85. W. Zhao, N. Lorenz, K. Jung and S. A. Sieber, *Angew. Chem. Int. Ed. Engl.*, 2015, DOI: 10.1002/anie.201508052.
86. R. Orth, T. Bottcher and S. A. Sieber, *Chem Commun (Camb)*, 2010, **46**, 8475-8477.
87. J. Kreuzer, N. C. Bach, D. Forler and S. A. Sieber, *Chem. Sci.*, 2015, **6**, 237-245.
88. A. Iwamaru, E. Iwado, S. Kondo, R. A. Newman, B. Vera, A. D. Rodriguez and Y. Kondo, *Mol. Cancer Ther.*, 2007, **6**, 184-192.
89. J. Li, J. S. Cisar, C. Y. Zhou, B. Vera, H. Williams, A. D. Rodriguez, B. F. Cravatt and D. Romo, *Nat. Chem.*, 2013, **5**, 510-517.
90. S. Nickel, R. A. Serwa, F. Kaschani, S. Ninck, S. Zweerink, E. W. Tate and M. Kaiser, *Chem. Eur. J.*, 2015, **21**, 10721-10728.
91. K. A. Kalesh, J. A. Clulow and E. W. Tate, *Chem Commun (Camb)*, 2015, **51**, 5497-5500.
92. J. Yang, K. A. Tallman, N. A. Porter and D. C. Liebler, *Anal. Chem.*, 2015, **87**, 2535-2541.
93. A. Vila, K. A. Tallman, A. T. Jacobs, D. C. Liebler, N. A. Porter and L. J. Marnett, *Chem. Res. Toxicol.*, 2008, **21**, 432-444.
94. J. Eirich, R. Orth and S. A. Sieber, *J. Am. Chem. Soc.*, 2011, **133**, 12144-12153.
95. A. Wittelsberger, B. E. Thomas, D. F. Mierke and M. Rosenblatt, *FEBS Lett.*, 2006, **580**, 1872-1876.
96. A. L. MacKinnon, J. L. Garrison, R. S. Hegde and J. Taunton, *J. Am. Chem. Soc.*, 2007, **129**, 14560-14561.
97. M. Suchanek, A. Radzikowska and C. Thiele, *Nat. Methods*, 2005, **2**, 261-267.

98. N. Srinivas, P. Jetter, B. J. Ueberbacher, M. Werneburg, K. Zerbe, J. Steinmann, B. Van der Meijden, F. Bernardini, A. Lederer, R. L. Dias, P. E. Misson, H. Henze, J. Zumbur, F. O. Gombert, D. Obrecht, P. Hunziker, S. Schauer, U. Ziegler, A. Kach, L. Eberl, K. Riedel, S. J. DeMarco and J. A. Robinson, *Science*, 2010, **327**, 1010-1013.
99. M. Bantscheff, D. Eberhard, Y. Abraham, S. Bastuck, M. Boesche, S. Hobson, T. Mathieson, J. Perrin, M. Rida, C. Rau, V. Reader, G. Sweetman, A. Bauer, T. Bouwmeester, C. Hopf, U. Kruse, G. Neubauer, N. Ramsden, J. Rick, B. Kuster and G. Drewes, *Nat. Biotechnol.*, 2007, **25**, 1035-1044.
100. J. J. Fischer, O. Y. Graebner Baessler, C. Dalhoff, S. Michaelis, A. K. Schrey, J. Ungewiss, K. Andrich, D. Jeske, F. Kroll, M. Gliński, M. Sefkow, M. Dreger and H. Koester, *J. Proteome Res.*, 2010, **9**, 806-817.
101. H. Shi, X. Cheng, S. K. Sze and S. Q. Yao, *Chem Commun (Camb)*, 2011, **47**, 11306-11308.
102. Z. Li, P. Hao, L. Li, C. Y. Tan, X. Cheng, G. Y. Chen, S. K. Sze, H. M. Shen and S. Q. Yao, *Angew. Chem. Int. Ed. Engl.*, 2013, **52**, 8551-8556.
103. M. Teplitzki, U. Mathesius and K. P. Rumbaugh, *Chem. Rev.*, 2011, **111**, 100-116.
104. L. Dubinsky, A. Delago, N. Amara, P. Krief, J. Rayo, T. Zor, V. V. Kravchenko and M. M. Meijler, *Chem Commun (Camb)*, 2013, **49**, 5826-5828.
105. J. J. Hulce, A. B. Coggins, M. J. Niphakis, S. E. Tully and B. F. Cravatt, *Nat. Methods*, 2013, **10**, 259-264.
106. W. Heydenreuter, E. Kunold and S. A. Sieber, *Chem Commun (Camb)*, 2015, **51**, 15784-15787.
107. M. Leonti, L. Casu, S. Raduner, F. Cottiglia, C. Floris, K. H. Altmann and J. Gertsch, *Biochem. Pharmacol.*, 2010, **79**, 1815-1826.
108. E. Weerapana, G. M. Simon and B. F. Cravatt, *Nat. Chem. Biol.*, 2008, **4**, 405-407.
109. M. Abo and E. Weerapana, *J. Am. Chem. Soc.*, 2015, **137**, 7087-7090.
110. D. Abegg, R. Frei, L. Cerato, D. Prasad Hari, C. Wang, J. Waser and A. Adibekian, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 10852-10857.
111. J. Wang, J. Zhang, C. J. Zhang, Y. K. Wong, T. K. Lim, Z. C. Hua, B. Liu, S. R. Tannenbaum, H. M. Shen and Q. Lin, *Sci. Rep.*, 2016, **6**, 22146.
112. M. P. Patricelli, A. K. Szardenings, M. Liyanage, T. K. Nomanbhoy, M. Wu, H. Weissig, A. Aban, D. Chun, S. Tanner and J. W. Kozarich, *Biochemistry*, 2007, **46**, 350-358.
113. B. E. Nordin, Y. Liu, A. Aban, H. E. Brown, J. Wu, A. K. Hainley, J. S. Rosenblum, T. K. Nomanbhoy and J. W. Kozarich, *Biochemistry*, 2015, **54**, 3024-3036.
114. S. C. Stolze, E. Deu, F. Kaschani, N. Li, B. I. Florea, K. H. Richau, T. Colby, R. A. van der Hoorn, H. S. Overkleeft, M. Bogoy and M. Kaiser, *Chem. Biol.*, 2012, **19**, 1546-1555.
115. D. A. Bachovchin, S. J. Brown, H. Rosen and B. F. Cravatt, *Nat. Biotechnol.*, 2009, **27**, 387-394.
116. D. Mellacheruvu, Z. Wright, A. L. Couzens, J. P. Lambert, N. A. St-Denis, T. Li, Y. V. Miteva, S. Hauri, M. E. Sardu, T. Y. Low, V. A. Halim, R. D. Bagshaw, N. C. Hubner, A. Al-Hakim, A. Bouchard, D. Faubert, D. Fermin, W. H. Dunham, M. Goudreau, Z. Y. Lin, B. G. Badillo, T. Pawson, D. Durocher, B. Coulombe, R. Aebersold, G. Superti-Furga, J. Colinge, A. J. Heck, H. Choi, M. Gstaiger, S. Mohammed, I. M. Cristea, K. L. Bennett, M. P. Washburn, B. Raught, R. M. Ewing, A. C. Gingras and A. I. Nesvizhskii, *Nat. Methods*, 2013, **10**, 730-736.
117. J. Park, M. Koh, J. Y. Koo, S. Lee and S. B. Park, *ACS Chem. Biol.*, 2015, DOI: 10.1021/acscchembio.5b00671.
118. D. Martinez Molina, R. Jafari, M. Ignatushchenko, T. Seki, E. A. Larsson, C. Dan, L. Sreekumar, Y. Cao and P. Nordlund, *Science*, 2013, **341**, 84-87.
119. M. M. Savitski, F. B. Reinhard, H. Franken, T. Werner, M. F. Savitski, D. Eberhard, D. Martinez Molina, R. Jafari, R. B. Dovega, S. Klaeger, B. Kuster, P. Nordlund, M. Bantscheff and G. Drewes, *Science*, 2014, **346**, 1255784.
120. R. E. Moellering and B. F. Cravatt, *Chem. Biol.*, 2012, **19**, 11-22.
121. M. Rask-Andersen, M. S. Almen and H. B. Schioth, *Nat. Rev. Drug Discov.*, 2011, **10**, 579-590.
122. D. Weichert and P. Gmeiner, *ACS Chem. Biol.*, 2015, **10**, 1376-1386.
123. J. D. White, M. F. Osborn, A. D. Moghaddam, L. E. Guzman, M. M. Haley and V. J. DeRose, *J. Am. Chem. Soc.*, 2013, **135**, 11680-11683.
124. J. D. White, M. M. Haley and V. J. DeRose, *Acc. Chem. Res.*, 2015, DOI: 10.1021/acs.accounts.5b00322.

Table 1. NP-inspired probes applied for target protein identification studies using gel-based proteomics. Blue: modification to introduce bioorthogonal alkyne tag.

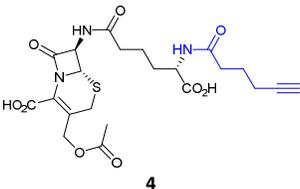
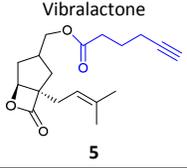
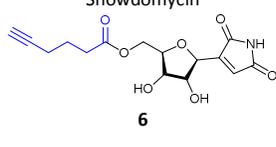
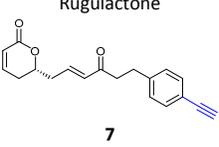
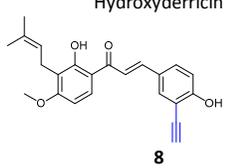
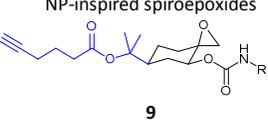
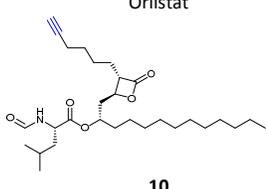
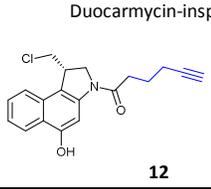
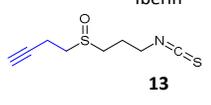
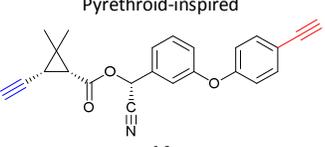
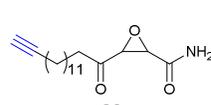
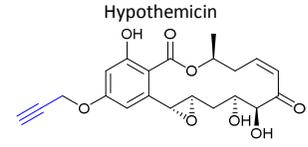
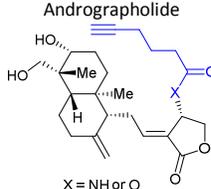
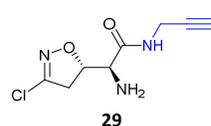
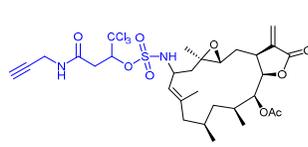
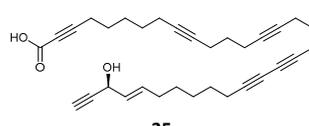
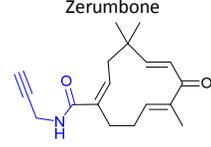
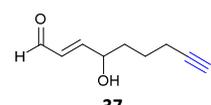
| Natural product and structure of inspired probe | Application, target(s) | Ref. |
|--|--|------|
| <p>Beta-lactam antibiotics.</p>  <p>4</p> | <p>Pathogenic bacteria.</p> <p>Penicillin binding proteins and β-lactamase identified as targets.</p> | 50 |
| <p>Vibrallactone</p>  <p>5</p> | <p><i>Listeria monocytogenes</i>.</p> <p>Multiple targets, including ClpP.</p> | 56 |
| <p>Showdomycin</p>  <p>6</p> | <p>Pathogenic bacteria.</p> <p>Targets: oxidoreductases and transferases, including the enzyme MurA1.</p> | 58 |
| <p>Rugulactone</p>  <p>7</p> | <p>Pathogenic bacteria.</p> <p>Targets: multiple, including kinase ThiD.</p> | 59 |
| <p>Hydroxyderricin</p>  <p>8</p> | <p><i>S. aureus</i></p> <p>Targets: multiple, including seryl-tRNA synthetase.</p> | 60 |
| <p>NP-inspired spiroepoxides</p>  <p>9</p> | <p>Human breast cancer cells.</p> <p>Target: phosphoglycerate mutase 1 (PGAM1).</p> | 62 |
| <p>Orlistat</p>  <p>10</p> | <p>HepG2 liver cancer cells.</p> <p>Targets: multiple, including FAS (fatty acid synthase) and GAPDH</p> | 63 |
| <p>Duocarmycin-inspired</p>  <p>12</p> | <p>A549 cancer cells.</p> <p>Target: Aldehyde dehydrogenase ALDH1A1</p> | 67 |
| <p>Iberin</p>  <p>13</p> | <p>HEK293 cells expressing TLRs.</p> <p>No proteomics – immunoaffinity methods identified binding to TLRs</p> | 70 |

Table 2. NP-inspired probes applied for target protein identification studies using gel-free and quantitative proteomics. Blue: modification to introduce bioorthogonal alkyne tag.

| Natural product-inspired probe | Application, methods, target | Ref. |
|--|---|------|
| <p>Pyrethroid-inspired</p>  <p>14</p> | <p>Mouse liver microsomes. EmPAI quantitative proteomics.</p> <p>Targets: cytochrome p450 enzymes</p> | 71 |
| <p>Cerulenin</p>  <p>20</p> | <p>Melanoma cells and Hek293 cells overexpressing PAT enzymes. Quantification by spectral counting.</p> <p>Multiple targets, including PATs, likely palmitoylated proteins</p> | 77 |
| <p>Hypochemicin</p>  <p>23</p> | <p><i>Trypanosoma brucei</i>. Gel-based and gel-free iTRAQ.</p> <p>Targets: kinases, including <i>TbGSK3short</i>, <i>TbCLK1</i> & 2.</p> | 80 |
| <p>Andrographolide</p>  <p>X = NH or O</p> <p>25-26</p> | <p>Human cancer cell lines. iTRAQ proteomics.</p> <p>Targets: multiple.</p> | 81 |
| <p>Fimbrilides</p>  <p>28</p> | <p><i>Vibrio</i> species. Dimethyl labelling proteomics.</p> <p>Targets: LuxS and LuxE, others</p> | 85 |
| <p>Acivicin</p>  <p>29</p> | <p>Liver cancer cell lines and mouse liver tissue. SILAC proteomics.</p> <p>Targets: multiple, including aldehyde dehydrogenases (esp. ALDH4A1) and CES1 (metabolite target).</p> | 87 |
| <p>Eupalmerin acetate</p>  <p>34</p> | <p>HL-60 leukaemia cells. SILAC proteomics.</p> <p>Targets: DERL1, CYB5B, TBXAS1</p> | 89 |

| | | |
|--|---|----|
| <p>Callyspongynic acid</p>  <p>35</p> | <p>HeLa and HEK293 cancer cells. SILAC proteomics.</p> <p>Targets: multiple membrane-associated proteins and proteins involved in lipid biosynthesis/metabolism</p> | 90 |
| <p>Zerumbone</p>  <p>36</p> | <p>HeLa cancer cells. Spike-in SILAC proteomics.</p> <p>Targets: multiple</p> | 91 |
| <p>Hydroxynonenal</p>  <p>37</p> | <p>RKO colon cancer cells. Peptide-centric approach; quantification based on isotopically-tagged, UV-cleavable CuAAC reagents.</p> <p>Targets: multiple – reactive cysteines.</p> | 92 |

