HHS Public Access

Author manuscript

Medchemcomm. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

Medchemcomm. 2016 April 1; 7(4): 576–585. doi:10.1039/c5md00518c.

Covalent Tethering of Fragments For Covalent Probe Discovery†

Stefan G. Kathman and **Alexander V. Statsyuk**

Center for Molecular Innovation and Drug Discovery, Chemistry of Life Processes Institute, Department of Chemistry, Northwestern University, Silverman Hall, 2145 Sheridan Road, Evanston, Illinois 60208, United States

Alexander V. Statsyuk: a-statsyuk@northwestern.edu

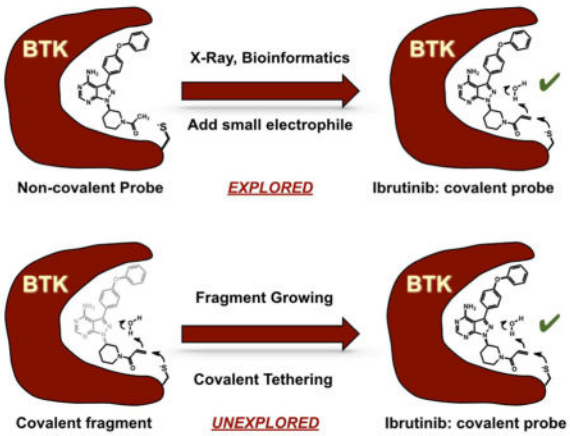
Abstract

Covalent probes and drugs have found widespread use as research tools and clinical agents. Covalent probes are useful because of their increased intracellular potency and because covalent labeling of cellular proteins can be tracked using click chemistry. Covalent drugs, on the other hand, can overcome drug resistance toward their reversible counterparts. The discovery of covalent probes and drugs usually follows two trajectories: covalent natural products and their analogues are used directly as covalent probes or drugs; or alternatively, a non-covalent probe is equipped with a reactive group and converted into a covalent probe. In both cases, there is a need to either have a natural product or a potent non-covalent scaffold. The alternative approach to discover covalent probes is to start with a drug-like fragment that already has an electrophile, and then grow the fragment into a potent lead compound. In this approach, the electrophilic fragment will react covalently with the target protein, and therefore the initial weak binding of the fragment can be amplified over time and detected using mass spectrometry. With this approach the surface of the protein can be interrogated with a library of covalent fragments to identify covalent drug binding sites. One challenge with this approach is the danger of non-specific covalent labeling of proteins with covalent fragments. The second challenge is the risk of selecting the most reactive fragment rather than the best binder if the covalent fragments are screened in mixtures. This review will highlight how covalent tethering was developed, its current state, and its future.

Graphical Abstract

†The authors declare no competing interests.

Correspondence to: Alexander V. Statsyuk, a-statsyuk@northwestern.edu.



1 Covalent Fragments: An Introduction

Covalent probes and drugs have gained wide popularity with the recent FDA approval of the covalent proteasome inhibitor carfilzomib and the covalent kinase inhibitors afatinib, and ibrutinib.[1] Carfilzomib is derived from an electrophilic natural product,[2] while afatinib and ibrutinib are synthetic drugs.[3, 4] The advent of covalent probes (both nucleophilic and electrophilic) for chemical biology has been assisted by the development of click chemistry methods.[5, 6] In fact, covalent probes are the most convenient probes to use, because their intracellular selectivity and potency, and the covalent labelling of off-target proteins can be easily estimated by conducting click chemistry experiments.[5, 6] Furthermore, click chemistry allows for tracking tissue and organ distribution of covalent probes *in vivo*.[6] Importantly, reversible interactions of covalent inhibitors with the protein target are also essential for their biochemical and cellular potency.[4] It is also important to keep in mind that covalent inhibitors can also bind reversibly to off-target proteins in the cases when the reactive groups are misaligned for the subsequent formation of the covalent bond.

In theory, the design of covalent probes follows two trajectories. In the first approach, the reversibly binding scaffold, which is usually potent, is equipped with the electrophile, which converts this scaffold into a covalent probe.[7] From a chemical perspective, this approach transforms a thermodynamic system into a kinetic system, in which one starts with a potent K^I and then builds in k_{inact}. In the second approach, a covalent fragment that contains ≤ 16 non-hydrogen atoms[8] and a reactive functional group is grown into a potent covalent probe by growing the fragment and improving its binding affinity (Figure 1B). In this case, one starts with a kinetic system and stays in the kinetic landscape during fragment optimization. Effectively one starts with a weak K_I and improves K_I while maintaining the same k_{inact} during optimization, thus improving the k_{inact}/K_I ratio, which is used to characterize covalent probe potency. Typical k_{inact}/K_I values for clinically useful covalent kinase inhibitors are in the range of 10^5–10^7 $M^{-1}s^{-1}$.[4] The second approach is particularly useful in cases when covalent drug binding sites on protein targets are unknown and need to be discovered. In this case, a protein of interest is treated with a mixture of covalent fragments, and if any of those fragments bind proximal to the nucleophilic residue on the protein (such as cysteine) they will be covalently trapped on the protein surface.

The first approach to design covalent probes is very well explored (see kinases as an example);[9] while the second approach ("covalent fragments") is relatively new and unexplored.[10–12] In this review, we will highlight the past, present, and future of covalent fragments, and outline emerging guidelines, challenges, and provocative questions when dealing with covalent fragments. We hope that this review will provide both academic and industrial communities with some guidelines on how to design and use covalent fragments to stimulate their widespread use and avoid failures in the future.

2 The Origin of Covalent Tethering

The concept of covalent fragments is based on the original tethering method, in which a library of disulfide containing molecules (MW ≈250 Da) is incubated with a protein of interest that contains an either native or engineered surface cysteine in the presence of 1 mM β-mercaptoethanol.[13] Small molecules that bind near the cysteine undergo a reversible disulfide bond exchange, and are covalently captured on the protein surface (Figure 2A). The reversibility of the system ensures that only specific binding events lead to the formation of the stable disulfide linked protein•fragment complex. If binding and the formation of the disulfide bond are non-specific, the large excess of β-mercaptoethanol will cleave non-specifically bound fragments from the protein surface, thus eliminating false positives. Covalent protein•fragment complexes can be identified using mass spectrometry. This strategy has been successfully used to discover many enzyme inhibitors, including a covalent inhibitor of the G12C K-Ras oncoprotein (Figure 2B).[14, 15] Importantly, in the last case only a small library of 500 fragments had to be screened to identify the initial hit. Taken together, disulfide tethering provided the first conceptually novel framework to tether drug like fragments to the protein surface. Disulfide tethering is especially useful to discover fragments that target native cysteines on their protein targets.

Such fragments can subsequently be elaborated into covalent probes, as in the case of K-Ras. To this end, we intended to improve upon disulfide tethering for our purposes, to address the following challenges that stimulated the development of the covalent tethering method:

- Disulfide fragment libraries are not commercially available
- Synthesis of disulfides requires working with thiols that have bad odor and are sensitive to oxidation.
- Lack of options for covalent capture, since a disulfide bond is the only option available.
- The disulfide tether needs to be replaced with a physiologically compatible electrophile. In such a case the resulting new fragment may not covalently react with its protein target due to the perturbed ligand binding mode and the different reactivity and geometry of the bound electrophile.

To address these challenges we asked if it would be possible to design a similar system in which fragments would already contain an electrophile (acrylamide, vinylsulfone, etc…) instead of disulfides. We would like to highlight the pioneering work by Jack Taunton and

co-workers that addresses this challenge by developing covalent reversible electrophiles such as α-cyanoacrylates, which can reversibly react with the cysteine.[16–20] However, we will focus on work related to irreversible covalent tethering in this review In such a case, there would be no need to replace the electrophile during optimization, while the fragment part could be elaborated into a more potent binder. We envisioned that the best binding fragment would covalently label a cysteine residue on the protein surface, while non-binders would not react with the cysteine (Figure 2C). Therefore we would not face the problem of electrophile switching, and would have many screening options since many cysteine reactive Michael acceptors are known.[21] Furthermore, synthesis of these libraries would be easy since only a one step amide bond formation reaction between the fragment and the electrophile is needed. The required Michael acceptor components can either be purchased or be prepared using a Horner–Wadsworth–Emmons reaction on a ~10 g scale.[12]

Finally we envisioned that such covalent fragments, if successfully developed, will become commercially available or otherwise will be synthesized and maintained by many academic groups. We also envisioned that such libraries of covalent fragments, if properly assembled, could be used in virtual docking studies.[16]

When our group began working on this approach in January 2011, it was not thought possible due to the major concern of non-specific covalent labeling of proteins with electrophilic fragments, and the fear that if screened in mixtures the most reactive fragment rather than the best binding fragment will be selected. No systematic studies and approaches to develop this screening method were known.

3 Initial Challenges and Design Rules

Three research groups: ours, D.J. Mann's group, and Pfizer, independently reported a covalent tethering approach.[10–12] The Pfizer group identified compounds with at least one electrophile from their internal collection, and filtered them by (1) quality control for purity > 95 %, (2) MW<350 but > 125, (3) clogP < 3.5, (4) total polar surface area < 140, (5) number of total rotatable bonds < 9, and (6) diversity analysis. The resulting compounds were screened against HIF-1α Cys^{255} and eight compounds were selected based on their ability to covalently modify their target protein. Out of these, seven compounds were hyper-reactive and non-specifically labelled other nucleophilic residues on the protein surface and therefore were eliminated from the screen (88% false positives). [10] David J. Mann and coworkers reported the screen of a small acrylamide library (10 compounds) against thymidylate synthase and identified a covalent inhibitor of this enzyme, thus providing the first proof of concept studies for covalent tethering.[11] In addition, one compound (**1**) in their library was hyper-reactive (10% rate of potential false positive), and had to be discarded before screening. We had a similar experience at the beginning, when we screened a small library of 10 acrylamides against the cysteine-containing HECT E3 Nedd4-1 and identified acrylamide **2** as a hyper reactive acrylamide fragment (Figure 3).

Perhaps compounds **1** and **2** are hyper-reactive because the lone pair of the acrylamide nitrogen is donated to the electron deficient aromatic rings and not the Michael acceptor, thereby increasing the reactivity of the acrylamide toward the nucleophiles. Such a high

frequency of hyper-reactive fragments (3/3 research groups identified hyper-reactive compounds in their libraries) prompted us to take a step back and begin outlining design rules for covalent fragment libraries. Ideally such libraries should have 0 hyper-reactive fragments, to ensure that there are 0% false positive results.

The following major criteria for the design of covalent fragment libraries emerged upon further consideration:

1. *The electrophile in the covalent fragments should be derived from known covalent inhibitors of enzymes that show broad SAR on the directing group.*[22] *Alternatively, the electrophile can be derived from natural products or FDA approved drugs which also show broad SAR.*[4, 23, 24] *This criterion ensures that covalent labeling of the protein will depend on the structure of the fragment, and the electrophile will be compatible with physiological conditions.*

2. *The intrinsic reactivity (the pseudo first order rate constant of the covalent reaction with the nucleophile of choice) of all covalent fragments in the library toward the nucleophile should ideally be the same. This criterion ensures that the protein target selects the best binding fragment (since the fragment part is variable) rather than the most reactive fragment.*[12]

3. Covalent fragments in the library should contain the same electrophile. It is not a good idea to screen a mixture of covalent fragments that contain acrylate, acrylamide, vinyl sulfone and other types of electrophiles, since they each have different reactivities. All covalent fragments should contain the same type of electrophile: either acrylates, acrylamides, vinyl sulfones, or another type of electrophile. This criterion ensures minimal variability in the system (i.e. only the fragment part changes) and ensures that there will be a minimal variability in the reactivity of the fragments. Furthermore, different electrophiles can contribute differently to the binding affinity of covalent fragments to their protein targets (different K_I values). If these conditions are met then the best binder rather than most reactive fragment will be selected.

4. Electrophile and fragment must be linked together by a minimal linker. It is better to keep the electrophile at the end of the molecule and not embedded in the molecule.

5. Ideally the required covalent fragments should be synthesized in one step using an amide bond formation reaction. This ensures the robustness, simplicity, and predictability of the covalent fragment synthesis.

It is generally accepted that approximately ~1000 fragments are needed to obtain high quality hits.[25] Thus it would be difficult to test criterion #2 at scale since it is not feasible for a small academic group to make 1000 fragments and then learn that they have wide variations in nucleophile reactivity and therefore cannot be used. To address this challenge we envisioned that ideally we should only synthesize 3 model compounds to test criterion #2. One compound would represent the most reactive fragment, the other would represent the least reactive fragment, and the reactivity of the third compound would be in between the first two. If the reactivity difference between the most reactive and the least reactive

compounds is negligible, we would predict that a larger library of covalent fragments would behave similarly, i.e. the reactivity difference between the most reactive and the least reactive fragment in a large ~1000 fragment library would be negligible, indicating that the intrinsic reactivity of all fragments in the library toward the nucleophile is similar (criterion #2).

We therefore developed a very useful experimental system that allowed us to predict the behavior of the large covalent fragment library using only a small set of experiments.[12] We took advantage of the well known inductive and mesomeric effects of $–NO_2$ and $–OCH_3$ groups when these are attached to aromatic rings. We envisioned that the NO_2-group containing compound would represent the most electron deficient and therefore most reactive fragment, while the compound that contains the electron donating CH_3O-group would be the least reactive in the series and therefore represent the least reactive fragment. With this in mind we prepared a series of compounds that contained NO_2-, CH_3O-, and H-functional groups with different arrangements of the electrophile. Among these series we found that acrylamides **3a–c** showed wide variations in their intrinsic reactivity toward the cysteine as judged by their pseudo first order reaction rate constant values. The NO_2-containing derivative **3c** was ~2000 fold more reactive than its CH_3O-counterpart **3b**. From these experiments we concluded that a large library of acrylamides prepared from aliphatic and aromatic amines would have large variations (~2000 fold) in their intrinsic reactivity toward thiols. Such a library would be enriched for hyper-reactive fragments leading to many false positives. In addition, it would be difficult to screen covalent fragments in mixtures, because the most reactive fragment rather than the best binder would be selected.

Interestingly, the analogous vinylsulfonamides **4a–c** showed only an 8-fold difference in their reactivity. Due to the poor conjugation of the nitrogen lone pair to the *d*-orbitals of sulfur, vinylsulfonamides are less sensitive to the electron donating effect of the nitrogen lone pair. Changing the linker between the electrophile and the fragment led to the series of compounds **5a–c** and **6a–c** that showed a narrow range of reactivities. In both cases the pseudo-first order reaction rate constants $k_{(\text{most reactive})}$ and $k_{(\text{least reactive})}$ were only ~1.5 fold different from each other, suggesting that a larger library of fragments would also display similar behavior. Both acrylates **5a–c** and vinylsulfones **6a–c** are promising candidates for covalent probe/drug design since both are orally bioavailable and non-toxic.[26, 27] Furthermore, vinylsulfone based covalent inhibitors have been recently reported to be 785 fold more toxic to *Trypanosoma brucei* when compared to human HL60 cells.[28] In addition, both are frequently used to design covalent inhibitors of cysteine proteases and show broad SAR on the directing group, suggesting that covalent labelling by these electrophiles will be sensitive to the structure of the attached fragment.[7, 29] With this in mind we assembled a library of 100 diverse fragments by simply conducting an amide bond formation reaction between acrylate **7** and commercially available carboxylic acids (~92,000 carboxylic acids are commercially available) (Figure 5A). We found that the synthesis of these acrylates is robust and proceeds with yields ranging from 10–90% with an average yield of 50%. In our experience 100 compounds can be prepared by a graduate student in a two month period. Automated synthesis of covalent fragments was also recently reported.[30] Since this is a one step synthesis, these fragments can be prepared on a 10–50 mg scale. Since the original

intention was to screen these fragments as mixtures of 10 at 100 μM each, large amounts of DMSO stock solutions can be prepared from these amounts (10–50 mg), and these DMSO stock solutions can be supplied to other research groups if needed. For example as of today our research group has sent DMSO stock solutions of covalent fragments to 9 research groups in the USA and one in Australia, and three research groups have successfully identified hit compounds. Furthermore, follow up studies are also feasible, since our research group was able to provide ~5 mg of selected compounds for follow up crystallography studies of the identified hits. As expected, the prepared library of covalent fragments showed a narrow range of reactivities toward the cysteine ($k_{(most\ reactive)}/k_{(least\ reactive)} = 2.4$), indicating the usefulness of our minimalist model system that predicts the behavior of the larger library of electrophiles (Figure 4).

Taken together we propose the following simple design rules for covalent fragments that would hopefully facilitate their further use in academic and industrial communities as well as the commercial availability of high quality libraries of covalent fragments (Figure 6). We suggest that in principle there could be two types of covalent fragments. **Type I** fragments are based on non-aromatic amines that can be converted into covalent fragments in one synthetic step.[11] The methylene linker between the R1 group and the amino group ensures the separation of the nitrogen from the fragment, to minimize the effect of the fragment on the lone pair of nitrogen. Therefore **Type I** covalent fragments must be designed with caution, and the reactivity of each fragment (rate constant) toward a thiol nucleophile (glutathione) must be measured and deposited into a database. Ideally each library of covalent fragments must have a database of pseudo first order reaction rate constants for each fragment. This will help to eliminate hyper-reactive acrylamide fragments from the collection. As a self check it is always useful to take p-NO_2, p-CH_3O, and benzylamine and couple them with the electrophiles of choice and then measure their reactivity (pseudo first order reaction rate constant) toward glutathione using NMR before making a large library of electrophiles. If the rate constant difference between p-NO_2 and p-CH_3O derivatives is small (<2 fold in our experience) then one can proceed with building a larger library of fragments. Most likely, the covalent fragments in this library will display a narrow range of reactivities, and therefore will be suitable for screening in mixtures or individually. If this difference is large (~2000 fold) then most likely the library of acrylamides will contain hyper-reactive fragments and will be difficult to screen under the original tethering conditions. There are many examples of acrylamides, vinylsulfonamides, and epoxides that show broad SAR on the directing group, and therefore covalent modification of the protein target with **Type I** covalent fragments will most likely depend on the chemical structure of the fragment.[4, 29, 31] Besides the discussed electrophiles, many other types of Michael acceptors can be used to assemble **Type I** covalent fragments.[21] **Type II** covalent fragments are based on carboxylic acids that can be coupled with the corresponding amines that carry Michael acceptors. Our practice shows that these are the most robust fragments that do not show large variations in reactivity across the library of fragments.[12] There are many protease inhibitors of this type that show good SAR on the directing R_1 group, and therefore covalent labeling of the protein target with these covalent fragments will most likely be driven by the fragment structure.[22] For both **Type I** and **Type II** covalent fragments, different electrophiles derived

from FDA approved drugs, known selective covalent probes, and natural products can serve as inspiration to select electrophiles for covalent fragments.[22, 23]

There are many thiol reactive electrophiles known (>20).[21] Thus in principle a library of 1000 fragments can be converted to >20,000 covalent fragments. We hope that in the future covalent fragment libraries will become commercially available, and it is advisable to group covalent fragments based on the electrophile these fragments contain. Ideally, one could provide libraries of covalent fragments that contain acryalmides, epoxides, vinylsulfones, alkynes, cyanoacrylates, etc.

4 Covalent Fragments As Enzyme Inhibitors

To evaluate the utility of the prepared covalent fragment library we chose the cysteine protease papain as a model enzyme. Papain has a highly reactive nucleophilic cysteine and several covalent inhibitors of papain are known. Upon treatment of 10 μM of papain with the mixtures of 10×100 μM covalent acrylate fragments (1 mM total concentration of electrophile) we identified three compounds that covalently and irreversibly modified the catalytic cysteine of papain after 1h incubation time, while the other 97 acrylates did not significantly modify papain (0% false positives).[12] This indicates that the covalent fragments that we designed are highly chemoselective, and the covalent modification of papain is driven by the structure of the fragment as we initially desired.

Since the only structurally variable part in all the acrylate fragments is the fragment itself, the observed selective covalent modification of papain with three fragments can be attributed to the fragment structure as was initially desired. Identified compounds **8–10** showed k_{inact}/K_I values comparable to known inhibitors of papain, and covalently labeled papain even in the presence of 10 mM of glutathione. Importantly, by screening fragments as mixtures and using intact protein mass spectrometry as a detection method we are able to screen ~100 compounds in one day, without the use of special robotic equipment. In yet another case a research group from Jackson State University successfully screened the same library of acrylate fragments using simple enzymatic assays as a screening method (10 μM final concentration of fragments) against the cysteine protease rhodesain.[32]

In this case seven hit compounds were identified, and the three most potent inhibitors were characterized. Due to the structural similarities between rhodesain and papain, the same compounds **9** and **8** were identified as rhodesain inhibitors. However, they were more potent inhibitors of rhodesain than papain as judged from k_{inact}/K_I values (Figure 7). Interestingly, the last example contradicts current practices in which “reactive compounds” always eliminated from any kind of HTS or FBDD screens. Here, in fact, only reactive compounds were screened. The take home lesson is that it is possible to screen reactive compounds. However, these compounds must be carefully designed using the criteria we outlined above for covalent fragments.

Taken together these early examples show that covalent fragments can be used to identify covalent enzyme inhibitors. Two screening methods of covalent fragments have emerged during these studies. First, covalent fragments can be screened using mass spectrometry

similar to the original disulfide tethering method. We find this detection method convenient, since it allows one person to easily screen ~100 compounds in one day. Second, covalent fragments can be screened individually using enzymatic assays such as in the case of the cysteine protease rhodesain. This is an important and unique feature of covalent tethering, since screening disulfide-containing fragments in enzymatic assays may not be possible. Therefore in laboratory settings where mass spectrometry is not available as a detection method, screening covalent fragments in enzymatic assays can become an alternative or even first line strategy. When screening fragments in an enzymatic assay, we looked for ≥85% inhibition at 10μM fragment concentration, which was enough to distinguish specific hits from non-specific ones. However, the ideal threshold may vary from enzyme to enzyme, so it is advisable to screen a non-specific electrophile which lacks a directing fragment as a control. Remarkably, the identified covalent inhibitors of papain and rhodesain are non-peptidic inhibitors, thus offering a path to non-peptidic covalent inhibitors of cysteine proteases, which have improved pharmacokinetic properties.[21, 33]

5 Covalent Fragments As Inhibitors of Protein-Protein Interactions

Over the past decade, protein-protein interactions (PPIs) have emerged as promising yet challenging targets.[34] Since functional protein complexes are needed to elicit physiological function, small molecule modulators of protein-protein interactions are a promising class of chemical probes and drugs. Several protein-protein interaction inhibitors are undergoing multiple stages of clinical and preclinical development.[34] Common features of protein-protein interaction inhibitors are a large molecular weight (MW >500 Da), increased hydrophobicity, and a large number of rings.[35] PPI inhibitors have to compete with the large protein-protein interaction interface, and this partly accounts for difficulties in developing PPI inhibitors. One solution to this challenge is to develop covalent inhibitors of protein-protein interactions. Because such inhibitors react with the protein target irreversibly, they would be more effective at disrupting the protein-protein interaction interface.

One large and completely unexplored class of enzymes that can be targeted by PPI inhibitors are enzymes that mediate the dynamic attachment of ubiquitin and ubiquitin-like proteins to their protein/lipid substrates.[36] The approximately ~800 known enzymes of this class regulate protein degradation, the activity of cell surface receptors, single transduction pathways, nuclear import/export, and gene transcription. In our attempts to develop inhibitors of these enzymes we focused on the E3 ligase Nedd4-1, which is a drug target to treat cancers, Parkinson's disease, obesity, and viral infections.[37–40] Nedd4-1 is a HECT E3 ligase (~28 known) that has a catalytic cysteine, and current studies suggest the simplified Nedd4-1 enzymatic mechanism shown in Figure 8.[41–45]

These studies have begun to reveal how Nedd4-1 receives Ub from the E2 enzyme, and how they transfer the first ubiquitin onto the lysine of the protein substrate, and how this enzyme elongates polyubiquitin chains. The catalytic domain of Nedd4-1 consists of a C-lobe and an N-lobe which are tethered to each other via a flexible hinge region. Upon binding of E2~Ub thioester the catalytic cysteine of the C-lobe faces E2~Ub thioester for the subsequent transthiolation reaction.[45, 46] Upon receiving the ubiquitin from the E2 enzyme, the C-lobe rotates such that the catalytic cysteine of Nedd4-1 is facing away from the E2 enzyme

binding site and adopts a ligation specific conformation.[44] Subsequent to the initial ubiquitin conjugation, polyubiquitin chains are elongated with the help of the second ubiquitin binding site at the N-lobe of the enzyme, which is thought to regulate the processivity of the enzyme.[41,43]

Analysis of the crystal structure of Nedd4-1 bound to ubiquitin at this processivity site revealed classic features of a protein-protein interaction interface (Figure 9A). Ile^{44} of ubiquitin forms hydrophobic contacts with Phe^{707} of Nedd4-1 in its catalytic HECT domain, and mutation of Phe^{707} of Nedd4-1 to alanine decreases the binding affinity of ubiquitin to Nedd4-1 (K_d drops from 11 μM to 340 μM).[43] The crystal structure of the Nedd4-1:Ub complex also reveals a nearby cysteine Cys^{627} that resides near the hotspot Ile^{44} and Phe^{707} residues. We therefore envisioned that the covalent modification of this cysteine residue with a covalent fragment would disrupt hydrophobic interactions between Phe^{707} of Nedd4-1 and Ile^{44} of the Ub.

In this case such an inhibitor would be one of the first examples of a covalent PPI inhibitor, which would also inhibit Nedd4-1 enzyme processivity. Accordingly, we used covalent tethering to screen 100 covalent fragments and identified two covalent modifiers of Nedd4-1 (0% false positives) and unexpectedly both of those covalently modified the non-catalytic Cys^{627} of Nedd4-1 at 1 mM inhibitor concentration even in the presence of the more reactive catalytic Cys^{867}.[47] This result highlights the remarkable specificity of the covalent tethering method relative to other cysteines. Subsequently, we were able to obtain the crystal structure of the Nedd4-1 HECT domain bound to covalent fragment **112** (Figure 9B). Subsequent optimization of the fragment has led to the more potent analogue **113**. Both compounds disrupted ubiquitin binding to the catalytic HECT domain of Nedd4-1 in a time and dose dependent manner as measured in fluorescence polarization assays (Figure 9C). k_{inact}/K_I values for the original fragment **112** and its improved analogue **113** were 0.089 $M^{-1}s^{-1}$ and 1.98 $M^{-1}s^{-1}$ respectively (22 fold improvement in potency). Current work to further improve the potency of **113** is ongoing. Thus in fact, covalent tethering can be used to identify covalent fragments that disrupt protein-protein interactions. The important lesson learned during these studies is that covalent fragment potency can be improved by structural optimization of the fragment. Subsequently, it was shown that covalent fragment **113** inhibits Nedd4-1 processivity, and in the presence of an antagonizing deubiquitinating enzyme USP8 indole **113** effectively inhibited the ability of Nedd4-1 to mono- and polyubiquitinate its protein substrate *in vitro*.[47]

6 Conclusions and Future Outlook

In summary, all these case studies have begun outlining major guidelines and design rules in using libraries of covalent fragments. When properly designed, covalent fragments can be used to discover initial leads for covalent enzyme inhibitors and covalent PPI inhibitors. In our own experience we had 0% false positive results and did not experience the non-specific covalent labeling of proteins when using a carefully designed library of covalent fragments. In our experience screening a small library of covalent fragments (100 fragments) against two different protein targets yielded three hit compounds against the cysteine protease papain, and two unique hit compounds against the HECT E3 ligase Nedd4-1. Thus, the

effective hit rate is 5%. Three other research groups that used our covalent fragment library (as of today 200 covalent fragments) have also identified multiple hit compounds different from our hits (unpublished). Such a high hit rate (≈5%) is typical for the fragment-based drug discovery approach. Successful examples showed that covalent fragment hits can be detected by either using mass spectrometry as a detection method (fragments screened as mixtures),[4, 11] or by simple enzymatic assays (fragments are screened individually).[32] In cases when the protein target is not an enzyme, screening using mass spectrometry is the method of choice.

When it is possible to design covalent fragment libraries that do not cause non-specific covalent labeling of proteins (**Type II** covalent fragments for example), it becomes possible to begin constructing libraries of covalent fragments for virtual docking studies.[48] In this case there is no need to make and store covalent fragments. However, since covalent fragments have reactive groups**,** they may be prone to hydration, and therefore have limited half-life under storage conditions. How long one can store covalent fragments remains to be determined. However, in our practice we found that we can store DMSO stocks of acrylates for at least a year without significant decomposition.

We envision that covalent fragments can target protein kinases (~200 kinases have cysteine near the ATP binding site).[8] Given that there are two FDA approved kinase inhibitor drugs, afatinib and ibrutinib, that contain acrylamide as a cysteine reactive functionality, acrylamide containing **Type I** covalent fragments could be the method of choice here. **Type II** covalent fragments can be used to target cysteine proteases (~150 known),[21] deubiquitinating enzymes (DUBs)/isopeptidases (~100 known),[49] and HECT/RBR E3s (~37 known).[50] Additionally, a recent example showed that **Type I** covalent inhibitors based on the E-64 epoxide electrophile can be used to design covalent inhibitors of SUMO deconjugating SENP enzymes, suggesting that covalent active site inhibitors of isopeptidases are feasible.[51] Taken together approximately ~500 enzymes can be targeted by covalent fragments, thus offering novel opportunities to target these enzymes. Perhaps other proteins can be targeted with covalent fragments as well, such as heat shock proteins; GTPases; epigenetic writers, readers and erasers; and other common PPI interfaces.

A unique advantage of covalent fragments is that the identified covalent fragment hits can be equipped with an alkyne tag and their intracellular potency and selectivity can be evaluated using click chemistry methods. Thus, early on, if several covalent fragments have been identified, it becomes possible to select the fragment which displays minimal off-target reactivity in cells and has higher intracellular potency. Different fragments may have different covalent protein labeling profiles as was shown earlier.[52]

We conclude this review by discussing a small set of provocative questions that pose some challenges for covalent fragment library design and fragment growth. *What is the equivalent of ligand efficiency for covalent fragments?*[53] Currently the potency of covalent fragments and their analogues during SAR studies should be evaluated by using k_{intact}/K_I values. *Once covalent fragment hits are identified, what is the strategy to grow the fragment?* In our practice we had some success by optimizing the reversibly binding fragment part, thereby building in K_I, and were able to achieve 22 fold improvement in the potency of Nedd4-1

inhibitors (Figure 8).[47] In our practice we had limited success with electrophile switching on the fragment or when we introduced any type of substituents into the electrophile. Thus it appears that it is not advisable to switch the electrophile on the covalent fragment after the fragment was identified. However, testing a panel of fragment analogues with different electrophiles could still be a good practice in other cases. *If one screens 100* ***Type II*** *vinyl sulfone fragments against papain and identifies three fragment hits, what happens if one screens the same set of* ***Type II*** *fragments that have a cyanoacrylate electrophile instead of vinylsulfone?* Would one identify different covalent fragment hits in this case? If so, it will significantly expand the utility of covalent fragment libraries since the same set of fragments can be coupled with multiple electrophiles, effectively producing a diverse library of compounds.

In summary, this review provides some perspective on the past, present, and future of covalent fragments and begins to introduce design rules for covalent fragment libraries (5 criteria) to facilitate the use and production of covalent fragment libraries. As this technology develops, further lessons can be learned to provide answers on provocative questions and further optimize the design rules and the use of covalent fragments.

Acknowledgments

Funding was provided from Northwestern University (A.V.S.), Pew Charitable Trusts (A.V.S.), National Institutes of Health Grants R01GM115632 (A.V.S.), T32GM105538 (S.G.K.), and the ACS Medicinal Chemistry Predoctoral Fellowship (S.G.K.). We thank Rama Mishra and the Center for Molecular Innovation and Drug Discovery for assisting with the initial design of the library of electrophilic fragments.

3. Bibliography

1. Singh J, Petter RC, Baillie TA, Whitty A. Nature Reviews Drug Discovery. 2011; 10:307–317. [PubMed: 21455239]
2. Kim KB, Crews CM. Natural Product Reports. 2013; 30:600. [PubMed: 23575525]
3. Pan ZY, Scheerens H, Li SJ, Schultz BE, Sprengeler PA, Burrill LC, Mendonca RV, Sweeney MD, Scott KCK, Grothaus PG, Jeffery DA, Spoerke JM, Honigberg LA, Young PR, Dalrymple SA, Palmer JT. Chemmedchem. 2007; 2:58. [PubMed: 17154430]
4. Schwartz PA, Kuzmic P, Solowiej J, Bergqvist S, Bolanos B, Almaden C, Nagata A, Ryan K, Feng JL, Dalvie D, Kath JC, Xu MR, Wani R, Murray BW. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:173. [PubMed: 24347635]
5. An H, Statsyuk AV. Journal of the American Chemical Society. 2013; 135:16948. [PubMed: 24138456]
6. Lanning BR, Whitby LR, Dix MM, Douhan J, Gilbert AM, Hett EC, Johnson TO, Joslyn C, Kath JC, Niessen S, Roberts LR, Schnute ME, Wang C, Hulce JJ, Wei B, Whiteley LO, Hayward MM, Cravatt BF. Nat Chem Biol. 2014; 10:760. [PubMed: 25038787]
7. Hanzlik RP, Thompson SA. Journal of Medicinal Chemistry. 1984; 27:711. [PubMed: 6547487]
8. Hall RJ, Mortenson PN, Murray CW. Prog Biophys Mol Biol. 2014; 116:82. [PubMed: 25268064]
9. Liu QS, Sabnis Y, Zhao Z, Zhang TH, Buhrlage SJ, Jones LH, Gray NS. Chemistry & Biology. 2013; 20:146. [PubMed: 23438744]
10. Cardoso R, Love R, Nilsson CL, Bergqvist S, Nowlin D, Yan J, Liu KK, Zhu J, Chen P, Deng YL, Dyson HJ, Greig MJ, Brooun A. Protein Sci. 2012; 21:1885. [PubMed: 23033253]
11. Nonoo RH, Armstrong A, Mann DJ. Chemmedchem. 2012; 7:2082. [PubMed: 23033251]
12. Kathman SG, Xu ZY, Statsyuk AV. Journal of Medicinal Chemistry. 2014; 57:4969. [PubMed: 24870364]

13. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, Wells JA. Proc Natl Acad Sci U S A. 2000; 97:9367. [PubMed: 10944209]
14. Erlanson DA, Wells JA, Braisted AC. Annual Review of Biophysics and Biomolecular Structure. 2004; 33:199.
15. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. Nature. 2013; 503:548. [PubMed: 24256730]
16. London N, Miller RM, Krishnan S, Uchida K, Irwin JJ, Eidam O, Gibold L, Cimermancic P, Bonnet R, Shoichet BK, Taunton J. Nature Chemical Biology. 2014; 10:1066. [PubMed: 25344815]
17. Bradshaw JM, McFarland JM, Paavilainen VO, Bisconte A, Tam D, Phan VT, Romanov S, Finkle D, Shu J, Patel V, Ton T, Li XY, Loughhead DG, Nunn PA, Karr DE, Gerritsen ME, Funk JO, Owens TD, Verner E, Brameld KA, Hill RJ, Goldstein DM, Taunton J. Nature Chemical Biology. 2015; 11:525. [PubMed: 26006010]
18. Serafimova IM, Pufall MA, Krishnan S, Duda K, Cohen MS, Maglathlin RL, McFarland JM, Miller RM, Frodin M, Taunton J. Nature Chemical Biology. 2012; 8:471. [PubMed: 22466421]
19. Miller RM, Paavilainen VO, Krishnan S, Serafimova IM, Taunton J. Journal of the American Chemical Society. 2013; 135:5298. [PubMed: 23540679]
20. Krishnan S, Miller RM, Tian B, Mullins RD, Jacobson MP, Taunton J. J Am Chem Soc. 2014; 136:12624. [PubMed: 25153195]
21. Flanagan ME, Abramite JA, Anderson DP, Aulabaugh A, Dahal UP, Gilbert AM, Li C, Montgomery J, Oppenheimer SR, Ryder T, Schuff BP, Uccello DP, Walker GS, Wu Y, Brown MF, Chen JSM, Hayward MM, Noe MC, Obach RS, Philippe L, Shanmugasundaram V, Shapiro MJ, Starr J, Stroh J, Che Y. Journal of Medicinal Chemistry. 2014; 57:10072. [PubMed: 25375838]
22. Ettari R, Micale N, Schirmeister T, Gelhaus C, Leippe M, Nizi E, Di Francesco ME, Grasso S, Zappala M. Journal of Medicinal Chemistry. 2009; 52:2157. [PubMed: 19296600]
23. Drahl C, Cravatt BF, Sorensen EJ. Angewandte Chemie-International Edition. 2005; 44:5788.
24. Gersch M, Kreuzer J, Sieber SA. Natural Product Reports. 2012; 29:659. [PubMed: 22504336]
25. Hall RJ, Mortenson PN, Murray CW. Progress in Biophysics & Molecular Biology. 2014; 116:82. [PubMed: 25268064]
26. Patick AK, Brothers MA, Maldonado F, Binford S, Maldonado O, Fuhrman S, Petersen A, Smith GJ 3rd, Zalman LS, Burns-Naas LA, Tran JQ. Antimicrob Agents Chemother. 2005; 49:2267. [PubMed: 15917520]
27. Li H, van der Linden WA, Verdoes M, Florea BI, McAllister FE, Govindaswamy K, Elias JE, Bhanot P, Overkleeft HS, Bogyo M. ACS Chemical Biology. 2014; 9:1869. [PubMed: 24918547]
28. Dunny E, Doherty W, Evans P, Malthouse JPG, Nolan D, Knox AJS. Journal of Medicinal Chemistry. 2013; 56:6638. [PubMed: 23952916]
29. Palmer JT, Rasnick D, Klaus JL, Bromme D. Journal of Medicinal Chemistry. 1995; 38:3193. [PubMed: 7650671]
30. Allen CE, Curran PR, Brearley AS, Boissel V, Sviridenko L, Press NJ, Stonehouse JP, Armstrong A. Organic Letters. 2015; 17:458. [PubMed: 25582165]
31. Schiefer IT, Tapadar S, Litosh V, Siklos M, Scism R, Wijewickrama GT, Chandrasena EP, Sinha V, Tavassoli E, Brunsteiner M, Fa' M, Arancio O, Petukhov P, Thatcher GRJ. Journal of Medicinal Chemistry. 2013; 56:6054. [PubMed: 23834438]
32. McShan D, Kathman S, Lowe B, Xu ZY, Zhan J, Statsyuk A, Ogungbe IV. Bioorganic & Medicinal Chemistry Letters. 2015; 25:4509. [PubMed: 26342866]
33. Renukuntla J, Vadlapudi AD, Patel A, Boddu SHS, Mitra AK. International Journal of Pharmaceutics. 2013; 447:75. [PubMed: 23428883]
34. Arkin MR, Tang YY, Wells JA. Chemistry & Biology. 2014; 21:1102. [PubMed: 25237857]
35. Higueruelo AP, Schreyer A, Bickerton GRJ, Pitt WR, Groom CR, Blundell TL. Chemical Biology & Drug Design. 2009; 74:457. [PubMed: 19811506]
36. Buckley DL, Crews CM. Angewandte Chemie-International Edition. 2014; 53:2312.

37. Tardiff DF, Jui NT, Khurana V, Tambe MA, Thompson ML, Chung CY, Kamadurai HB, Kim HT, Lancaster AK, Caldwell KA, Caldwell GA, Rochet JC, Buchwald SL, Lindquist S. Science. 2013; 342:979. [PubMed: 24158909]
38. Han ZY, Lu JH, Liu YL, Davis B, Lee MS, Olson MA, Ruthel G, Freedman BD, Schnell MJ, Wrobel JE, Reitz AB, Harty RN. Journal of Virology. 2014; 88:7294. [PubMed: 24741084]
39. Shi YJ, Wang JR, Chandarlapaty S, Cross J, Thompson C, Rosen N, Jiang XJ. Nature Structural & Molecular Biology. 2014; 21:522.
40. Li JJ, Ferry RJ, Diao SY, Xue BZ, Bahouth SW, Liao FF. Endocrinology. 2015; 156:1283. [PubMed: 25607895]
41. French ME, Kretzmann BR, Hicke L. J Biol Chem. 2009; 284:12071. [PubMed: 19252184]
42. Kim HC, Steffen AM, Oldham ML, Chen J, Huibregtse J. EMBO Reports. 2011; 12:334. [PubMed: 21399621]
43. Maspero E, Mari S, Valentini E, Musacchio A, Fish A, Pasqualato S, Polo S. EMBO Rep. 2011; 12:342. [PubMed: 21399620]
44. Kamadurai HB, Qiu Y, Deng A, Harrison JS, MacDonald C, Actis M, Rodrigues P, Miller DJ, Souphron J, Lewis SM, Kurinov I, Fujii N, Hammel M, Piper R, Kuhlman B, Schulman BA. Elife. 2013; 2:e00828. [PubMed: 23936628]
45. Maspero E, Valentini E, Mari S, Cecatiello V, Soffientini P, Pasqualato S, Polo S. Nature Structural & Molecular Biology. 2013; 20:696.
46. Kamadurai HB, Souphron J, Scott DC, Duda DM, Miller DJ, Stringer D, Piper RC, Schulman BA. Mol Cell. 2009; 36:1095. [PubMed: 20064473]
47. Kathman SG, Span I, Smith AT, Xu ZY, Zhan J, Rosenzweig AC, Statsyuk AV. Journal of the American Chemical Society. 2015; 137:12442. [PubMed: 26371805]
48. London N, Miller RM, Irwin JJ, Eidam O, Gibold L, Bonnet R, Shoichet BK, Taunton J. Biophysical Journal. 2014; 106:263a. [PubMed: 24411258]
49. Borodovsky A, Ovaa H, Kolli N, Gan-Erdene T, Wilkinson KD, Ploegh HL, Kessler BM. Chem Biol. 2002; 9:1149. [PubMed: 12401499]
50. Love KR, Pandya RK, Spooner E, Ploegh HL. ACS Chemical Biology. 2009; 4:275. [PubMed: 19256548]
51. Albrow VE, Ponder EL, Fasci D, Bekes M, Deu E, Salvesen GS, Bogyo M. Chemistry & Biology. 2011; 18:722. [PubMed: 21700208]
52. Banerjee R, Pace NJ, Brown DR, Weerapana E. Journal of the American Chemical Society. 2013; 135:2497. [PubMed: 23379904]
53. Hopkins AL, Keseru GM, Leeson PD, Rees DC, Reynolds CH. Nature Reviews Drug Discovery. 2014; 13:105.

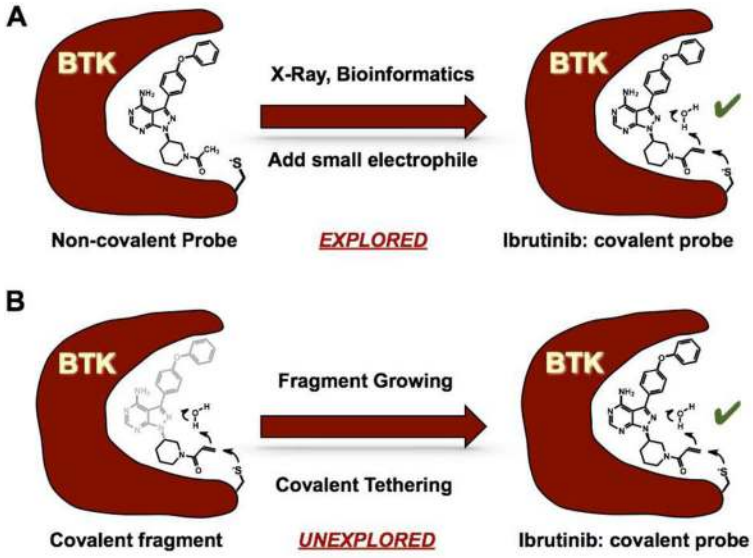


Figure 1.

The schematic/theoretical representation of two complementary strategies toward covalent probe design. Ibrutinib, a covalent inhibitor of Bruton's tyrosine kinase (BTK), is used as an example. A) A non-covalent scaffold is co-crystallized with the protein target to determine the relative position of the scaffold and the nucleophilic cysteine. The area of the non-covalent scaffold which is proximal to the target cysteine is equipped with the reactive group to convert the non-covalent probe into a covalent probe. B) A covalent fragment with mM-μM K_I forms a covalent bond with the target cysteine on the protein. Co-crystallizing the covalent fragment with the protein target facilitates subsequent SAR studies, to grow the fragment into a drug-lead by improving the K_I to nM range, thus leading to an increase in the k_{inact}/K_I ratio, which characterizes the potency of the covalent probe.

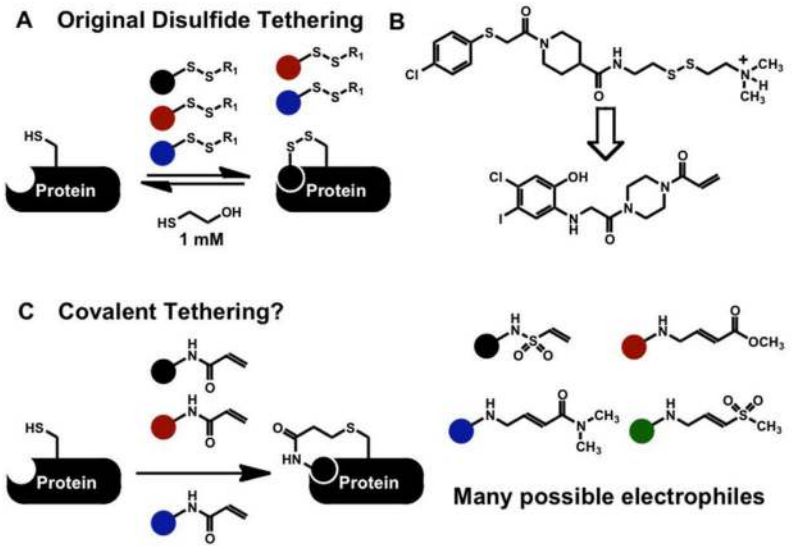


Figure 2.

A) The original disulfide tethering method. B) A covalent inhibitor of G12C K-Ras discovered using disulfide tethering. C) The proposed covalent tethering method. In contrast to disulfide tethering, many different types of fragments bearing different electrophiles are possible. In this case, fragments bearing the same electrophile could be screened in mixtures against the protein target.

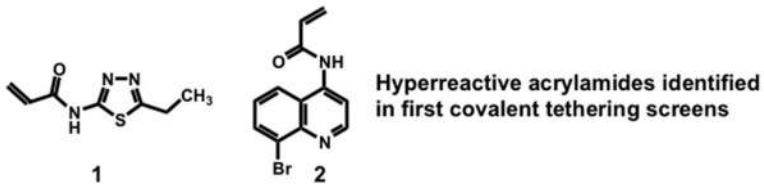


Figure 3.
Hyper-reactive acrylamide fragments identified by our (unpublished) and David J. Mann's laboratories.

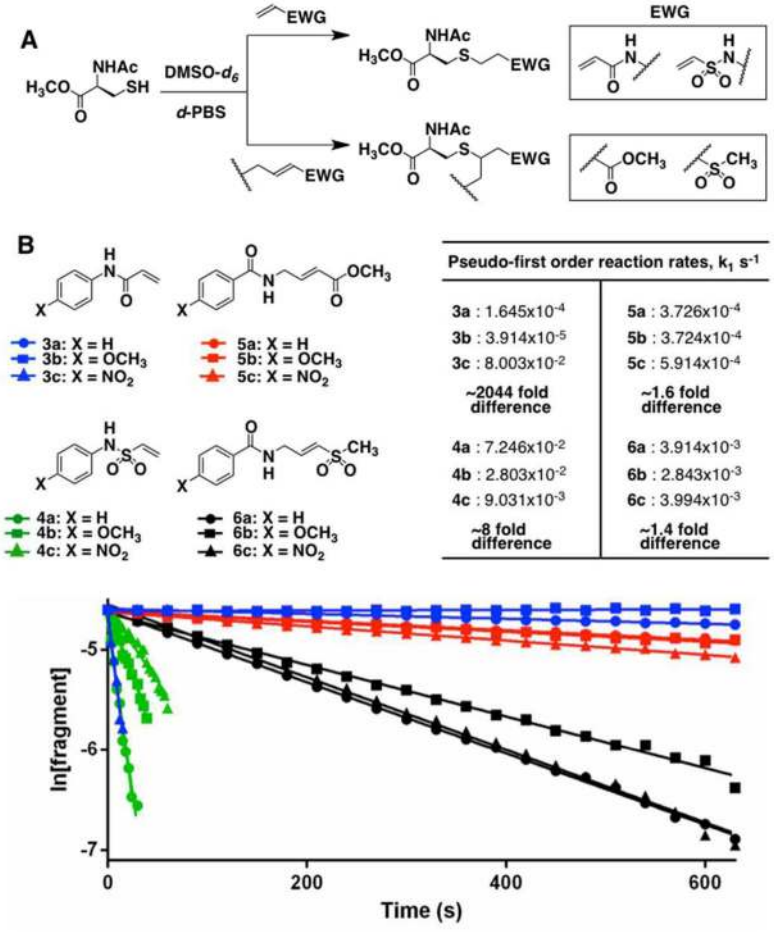


Figure 4.

A) General scheme of NMR rate studies. B) Chemical structures of the electrophiles **1**–**4** tested for suitability for covalent tethering and their pseudo-first order reaction rates with *N*-acetylcysteine methyl ester at pD 8.0 as measured by NMR spectroscopy. Reproduced from ref. 12.

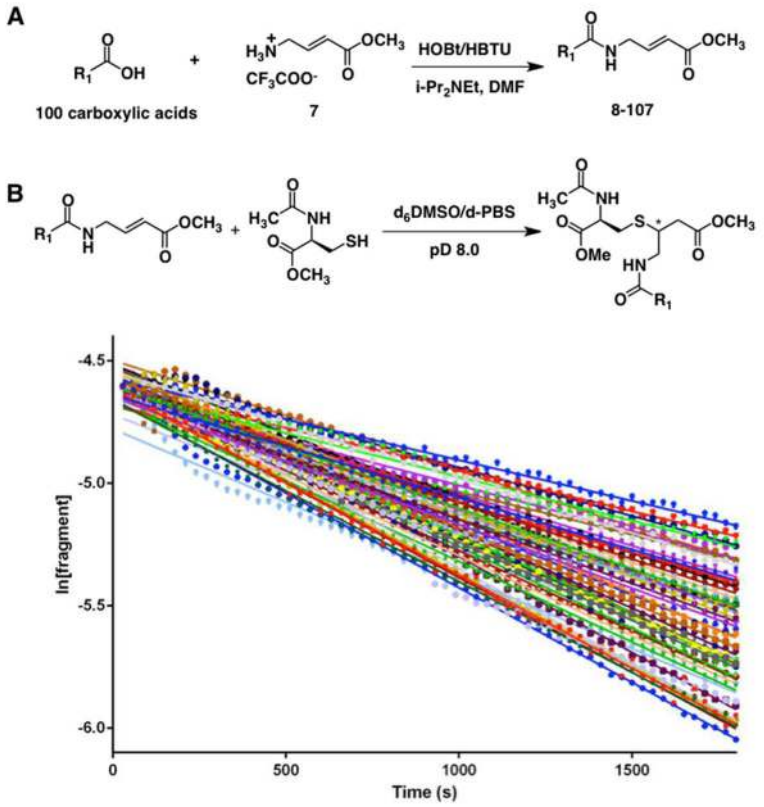


Figure 5.

(A) Design and synthesis of the fragment library. Electrophile **7** can be prepared on a 10g scale. (B). Pseudo-first order NMR rate plots of the reaction of 50 compounds with *N*-acetyl cysteine methyl ester. Different colors represent different fragments. Reproduced from ref. 12.

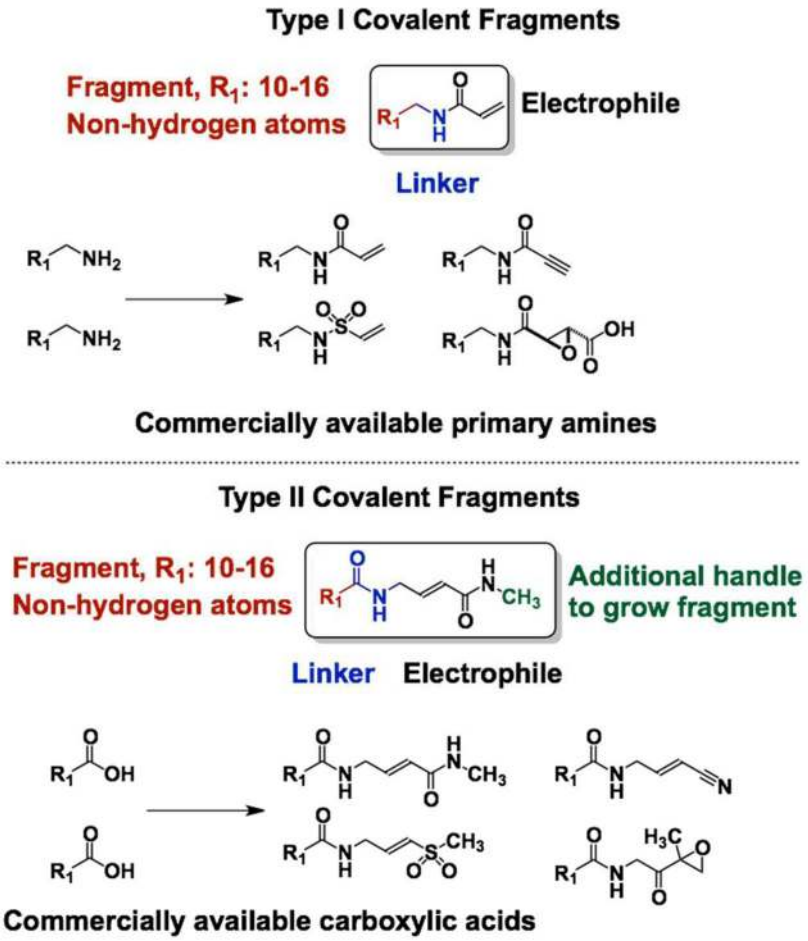


Figure 6.

Proposed design of covalent fragments. **Type I** covalent fragments are made from non-aromatic amines (~4000 commercially available) and contain the fragment part R_1 (10–16 non-hydrogen atoms), a linker, and the electrophile. R_1 is separated from the amino group with methylene as a linker to avoid the conjugation of the nitrogen lone pair to the fragment. This ensures that these types of acrylamides will have similar reactivity toward thiols. **Type II** covalent fragments are based on carboxylic acids (~92,000 commercially available) that are converted into covalent fragments. **Type II** fragments satisfy all 5 criteria and are therefore ideally suited for covalent tethering. In both cases FDA approved covalent drugs and reactive natural products can serve as inspiration and a guide to select electrophiles for covalent fragment library design.

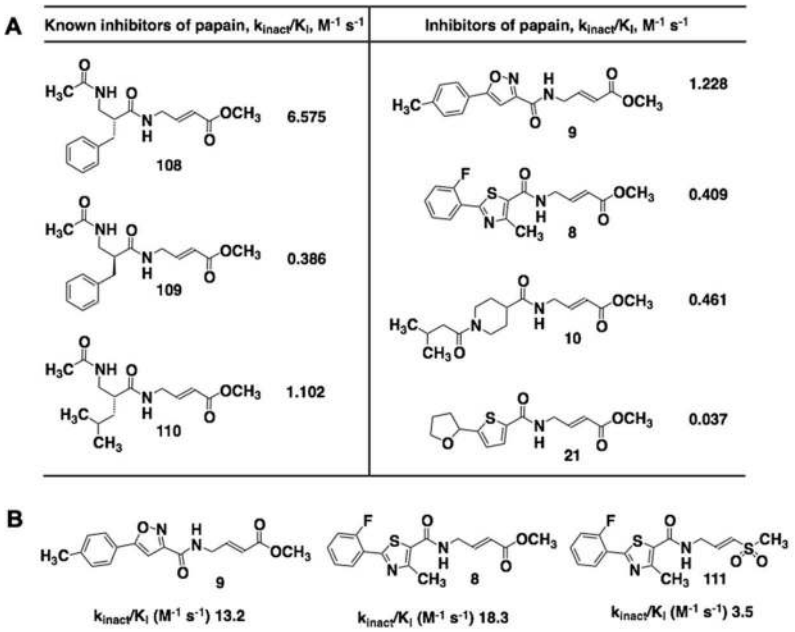


Figure 7.

A) Covalent inhibitors of papain identified using the covalent tethering method. Compound **21** is a negative control compound that did not covalently label papain. B) Covalent inhibitors of rhodesain identified using the covalent tethering method.

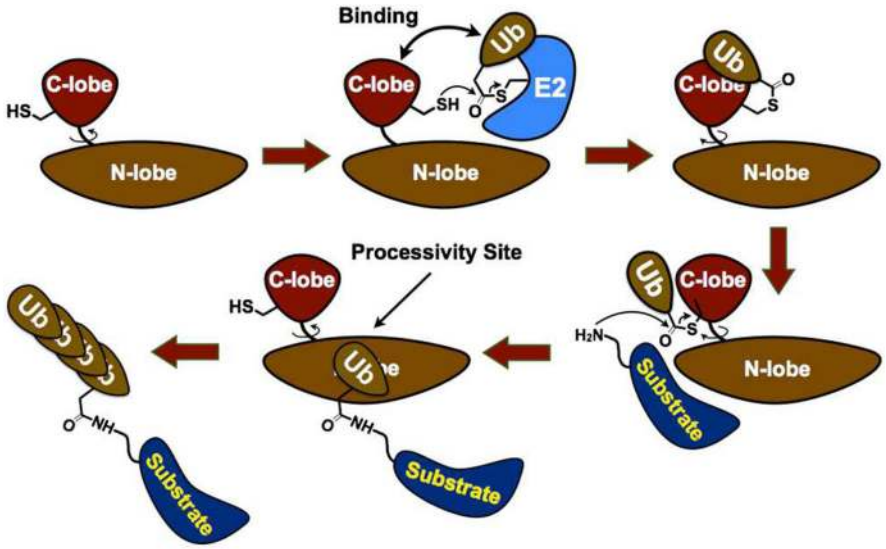


Figure 8.

Simplified model of HECT E3 mediated protein ubiquitination. E2~Ub thioester binds the catalytic HECT domain, followed by a transthiolation reaction, producing HECT E3~Ub thioester. Subsequently, HECT E3~Ub ligates Ub onto the lysine of the substrate, followed by polyubiquitin chain growth. The C- and N-lobes of the HECT domain rotate relative to each other during the catalytic cycle.

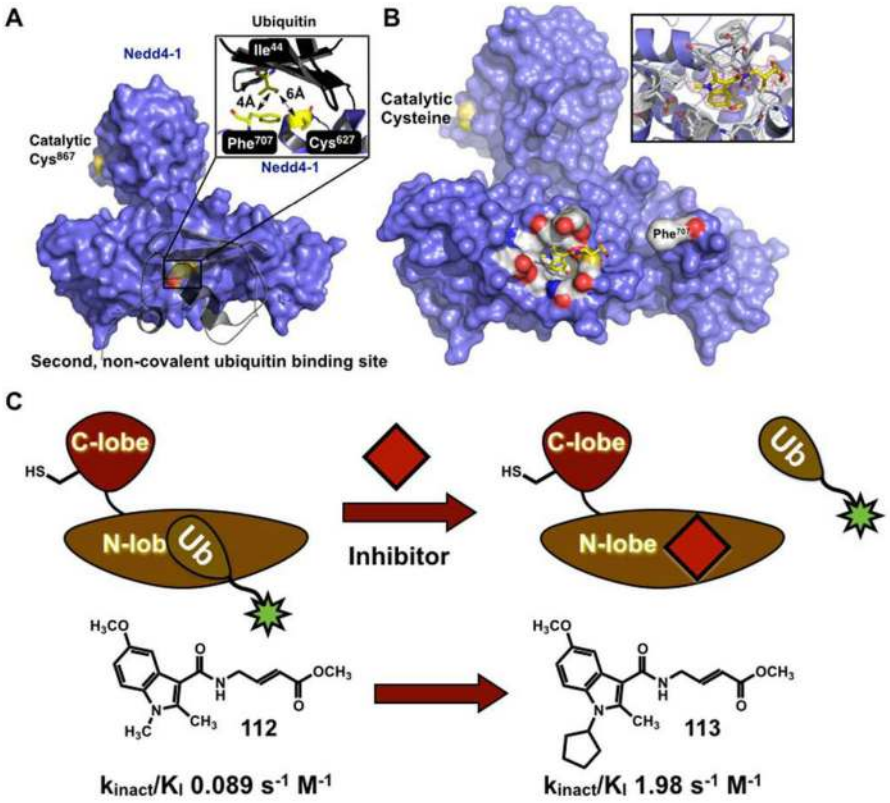


Figure 9.

Covalent fragments as PPI inhibitors. A) Co-crystal structure of Nedd4-1 catalytic HECT domain with ubiquitin. Phe^{707} of Nedd4-1 forms critical hydrophobic interactions with Ile^{44} of ubiquitin. Mutation of Phe^{707} of Nedd4-1 into alanine increases the K_d from ~11 μM to 340 μM. Therefore Phe^{707} and Ile^{44} represent a classic protein-protein interaction hot spot. In addition, Tyr^{605} of Nedd4-1 also forms critical binding interactions with Leu^{73} of Ub. B) Crystal structure of the indole **112** bound to the catalytic HECT domain of Nedd4-1 showing that compound **112** actually blocks Tyr^{605} of Nedd4-1 from interacting with Ub. C) The potency of compound **112** and its analogue **113** were investigated in fluorescence polarization assays to obtain k_{inact}/K_I values.