A Phenotypic Approach for the Identification of New Molecules for Targeted Protein Degradation Applications

Stacey, Peter; Lithgow, Hannah; Lewell, Xiao; Konopacka, Agnieszka; Besley, Stephen; Green, Georgina

Published in:
SLAS Discovery

DOI:
10.1177/24725552211017517

Publication date:
2021

Document Version
Peer reviewed version

A phenotypic approach for the identification of new molecules for targeted protein degradation applications

Peter Stacey [a]±, Hannah Lithgow [a][b]±, Xiao Lewell [a], Agnieszka Konopacka [a], Stephen Besley [a], Georgina Green [a], Ryan Whatling [a], Robert Law [a], Sascha Roth [d], Gopal Sapkota [d], Ian E. D. Smith [a], Glenn A. Burley [b], John Harling [a], Andrew B. Benowitz [a], Markus A. Queisser [a]*, Marcel Muelbaier [c]*

[a] Medicine Design, GlaxoSmithKline, Gunnels Wood Road, Stevenage, SG1 2NY, UK. [b] Department of Pure and Applied Chemistry, WestCHEM, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL UK. [c] Cellzome GmbH, a GSK company, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [d] MRC Protein Phosphorylation and Ubiquitylation Unit (PPU), University of Dundee, Dundee, DD1 5EH, UK.

±These authors contributed equally.

* Email: m.muelbaier@web.de or markus.a.queisser@gsk.com

KEYWORDS: Targeted Protein Degradation • E3 ubiquitin ligases • HaloTag • PROTACs • HTE • Phenotypic screening.

ABSTRACT

Targeted protein degradation (TPD) is an emerging new strategy for the modulation of intracellular protein levels with applications in chemical biology and drug discovery. One approach to enable this strategy is to redirect the ubiquitin-proteasome system (UPS) to mark and degrade target proteins of interest (POI) through the use of Proteolysis Targeting Chimeras

(PROTACs). Although great progress has been made in enabling PROTACs as a platform, there are still a limited number of E3 ligases that have been employed for PROTAC design. Herein we report a novel phenotypic screening approach for the identification of E3 ligase binders. The key concept underlying this approach is the high-throughput modification of screening compounds with a chloroalkane moiety to generate HaloPROTACs *in-situ,* which were then evaluated for their ability to degrade a GFP-HaloTag fusion protein in a cellular context. As proof of concept, we demonstrated that we could generate and detect functional HaloPROTACs *in situ,* using a validated von-Hippel Lindau (VHL) binder that successfully degraded the GFP-

Stacey, P. et al. (2021) 'A Phenotypic Approach for the Identification of New Molecules for Targeted Protein Degradation Applications', *SLAS Discovery*, vol. 26, no. 7, pp. 885-895.
https://doi.org/10.1177/24725552211017517.

HaloTag fusion protein in living cells. We then used this method to prepare and screen a library of approximately 2000 prospective E3 ligase recruiting molecules.

INTRODUCTION

Protein degradation has emerged as an important therapeutic strategy in which small molecules can recruit cellular protein degradation machinery to induce the degradation of selected proteins of interest (POIs). One important subset of degraders that are targeted to specific POIs are PROTACs, which are bifunctional molecules composed of a ligand to the POI linked to a second ligand that binds an E3 ligase[1-6]. Ternary complex formation between the target POI, the PROTAC, and the E3 ligase enables the POI to be covalently modified with poly-ubiquitin chains which are then recognised by the proteasome[7-10]. Many different POIs have been reported to be successfully degraded by PROTACs, and several PROTAC molecules have now entered clinical trials[11-13]. Currently reported PROTACs recruit only a small proportion of the known E3 ligases encoded by the human genome[14-19]. However, expanding the available portfolio of functional E3 ligase-recruiting ligands might provide additional versatility for the degradation of challenging targets, as well as potentially offering improvements in physicochemical properties and developability as compared to currently known E3 ligase chemical binders[20-22].

One approach to expanding the chemical matter available for the recruitment of E3 ligases is to conduct high throughput screening campaigns against specific individual E3 ligases using binding affinity methods such as encoded library technology (ELT)[23, 24]. This type of screening has the advantage of being high throughput, and well-established [25-27]. However, this approach also means that screened E3 ligases must generally be selected with limited knowledge about their substrate scope and tractability. Additionally, even before a functional PROTAC can be constructed from an E3 ligase screening hit, a medicinal chemistry effort is typically required to improve E3 ligase binding affinity and to prepare and test multiple putative PROTACs with various linkers and linker attachment configurations. In practice, this reductionist single-target screening approach for E3 ligases is costly, inefficient, and slow, and can realistically only be conducted on a small fraction of the total number of potential E3 ligases that might have utility in PROTAC design[28].

To address the inherent inefficiencies posed by serial screening of individual E3 ligases, we have developed and implemented an alternative strategy that combines a cell-based targeted protein degradation phenotypic screen with efficient 384-well high-throughput chemical synthesis, shown schematically in Figure **1**. We based our screening design strategy on the previously described HaloPROTAC method for the evaluation of VHL E3 ligase binders in a cellular context. HaloPROTAC molecules contain a linker with a terminal chloroalkane group that can covalently bind to the HaloTag protein[29], and we have previously demonstrated the degradation of a HaloTag-GFP (Green Fluorescent Protein) fusion protein upon cell treatment with a VHL-based HaloPROTAC[30]. We reasoned that the HaloPROTAC assay could be developed into a high-throughput phenotypic screening format, where new ligands would be identified based on their ability to act as functional degraders when tethered to the HaloTag-GFP protein in a cell. This format should be agnostic regarding the underlying mechanism of POI degradation, but would effectively multiplex a screening campaign of E3 ligases as well as other intracellular proteins and mechanisms relevant to targeted protein degradation. Any resulting hits will have already demonstrated functional target degradation in cells, and would in principle be suitable for incorporation into heterobifunctional degraders.

To realize the goal of conducting a high-throughput phenotypic HaloPROTAC screen, the preparation of a suitable HaloPROTAC screening library must be addressed, as we predicted that this step might be rate-limiting. We hypothesized that it might be feasible to prepare a HaloPROTAC library **3** *in situ,* via high yielding coupling reactions in 384-well low volume plates between a diverse array of amines **2** and an activated ester building block **1** containing a linker and a primary chloroalkane (Figure **1A**). This approach provides for significant diversity, since multiple HaloPROTAC screening libraries can be prepared from the same amine building blocks by employing different chloroalkane linkers.

Since only fully elaborated HaloPROTAC molecules should be capable of degrading the HaloTag-GFP fusion protein, we further hypothesized that it would be possible to screen the HaloPROTAC library *in situ,* without purification of the individual library compounds. We also speculated that in view of the multiplexed nature of this assay design, which in principle screens all of the E3 ligases and degradation mechanisms present in the cells simultaneously, it might be possible to identify a functional degrader hit from a much smaller compound library than would otherwise be required in an equivalent non-multiplexed single target screening campaign. Toward this goal, we describe herein the optimization of a high-throughput amide coupling reaction in low volume 384-well format, and the efficient synthesis of a

HaloPROTAC screening set starting from a curated library of approximately 3000 reactive amines. The resulting HaloPROTAC screening set was used to conduct a phenotypic screening campaign to detect functional HaloTag-GFP degraders in HeLa cells (Figure **1B**). A false-positive counter-screen was also implemented using a cell line with a mutated version of the HaloTag protein which does not covalently bind chloroalkanes (Figure **1C**). Our work demonstrates a novel application of HaloTag technology for a phenotypic screening approach to detect small molecule ligands for use in targeted protein degradation.

[INSERT FIGURE 1]

MATERIALS AND METHODS

General high throughput chemistry considerations

Experiments conducted under an inert atmosphere were performed inside a MBraun glovebox with a constant N_2-purge with oxygen typically <5 ppm. Stock solution reservoirs were polystyrene, 50 mL capacity Corning Costar cat. 4870. 384-well plates were polypropylene and either Greiner, v-bottom, cat. 781280 or Labcyte, flat-bottom, cat. P-05525. Plates were covered with a polystyrene lid where indicated (Corning, cat. 3098). Pipetting procedures were completed using an electronic multichannel pipette from stock solution reservoirs (Thermofisher, 12 channels, 1 to 30 µL, cat. 4671030BT with Thermofisher ClipTip cat. 94410103).

HaloPROTAC library design and synthesis

The HaloPROTAC screening library **5** was constructed via the high-throughput coupling of a diverse array of amines **2** with a single activated N-hydroxysuccinimde (NHS) ester[31] building block **4** bearing a primary alkyl chloride (Figure **2A**). Suitable amines were selected based on their predicted reactivity under the anticipated reaction conditions, the presence of a chromophore to facilitate UV detection, structural diversity, and the predicted cellular permeability of the corresponding fully elaborated HaloPROTACs. Based on these criteria, a total of 2934 amines (661 primary, 2267 secondary, and 6 compounds containing both primary and secondary amines) were selected for use in library construction from the GSK compound collection. Supplementary Figure **S1** provides an overview of the physicochemical property distribution of the selected amines, and details of the selection procedure are described in the Supplementary Material.

HaloPROTAC library synthesis was conducted directly on 384-well reactor plates containing the selected amines in random order. The plates were prepared to contain 10 μL of 10 mM DMSO stock solution per well, and were stored at –20 °C. The selected amine building blocks were delivered on nine plates (including 16 negative control DMSO wells and 16 wells containing the VHL ligand **6** corresponding to positive control VHL-HaloPROTAC **7** (Figure **3A**) which was also synthesized *in situ* (example plate layout shown in Figure **2B**). The plates were thawed at room temperature, centrifuged briefly at 200g and then placed in a glovebox prior to reaction dispensing. Reaction wells were then dispensed with stock solution 1 (**SS1**; prepared as described below), and the reactor plates were then sealed with lids (Greiner cat. 676001) and left to stand at room temperature for 16 h. The lids were then removed, DMSO (20 μL) was added to each well, and each plate was removed from the glovebox and resealed using a new metal foil lid. The plates were then placed into an automatic sample organizer and analyzed by LC/MS for reaction completion and the presence of the target mass in each well.

Stock solution 1 (**SS1**): A stock solution of the chloroalkane NHS ester 2,5-dioxopyrrolidin-1-yl 18-chloro-3,6,9,12-tetraoxaoctadecanoate **4** (Figure **2A**; 42.4 mg, 0.100 mmol) and *N*-methylmorpholine (109 μL, 0.991 μmol) in DMSO (10 mL) was prepared in a glovebox under a nitrogen atmosphere to give a 10 mM DMSO solution that was used immediately.

An analysis of the reaction outcomes for each of the nine plates is shown in Supplementary Figure **S4**.

[INSERT FIGURE 2]

Stable cell line preparation

A HeLa cell line expressing enhanced Green Fluorescent Protein (GFP) fused at the N-terminus with the previously described HaloTag protein was prepared[30]. Additionally, a second analogous HeLa cell line was prepared containing a D106A mutation in the HaloTag part of the fusion protein to enable counter-screening for false-positives. A pBabeD based retroviral vector system was used for cell line generation (modified version of Cell Biolabs pBabe plasmid, plasmids DU29230 and DU29215, respectively; MRC-PPU Reagents and Services https://mrcppureagents.dundee.ac.uk). In the event, HEK293-FT cells at ~60% confluence in a 10 cm dish were transfected with 6 μg of the respective pBabeD constructs, 3.8 μg of pCMV-GAG/POL (Clontech), and 2.2 μg pCM-VSVG (Clontech) using PEI (2 μg/ml). The medium was exchanged on the next day, and after a further 24 h, virus-containing medium was

harvested and cleaned using a 0.45 μm filter. HeLa cells (University of Dundee MRC-PPU stock, obtained from the American Type Culture Collection) were then transduced with the virus-containing medium supplemented with 10 μg/ml polybrene for 24 h. Cells were grown for an additional 24 h in normal growth medium and selected for 3 days in 2 μg/mL puromycin-containing medium. Finally, single cell clones were isolated using FACS sorting based on GFP fluorescence. The cell clones selected for screening and counter-screening both exhibited similar magnitude of fluorescence intensity, although the cells containing the HaloTag D106A mutation were marginally less bright. Stably transduced HeLa cells were subsequently cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat inactivated FBS, 2 mM GlutaMAX, 50 U/mL penicillin, 50 μg/mL streptomycin and 1 μg/mL puromycin, at 37°C with 5% CO_2.

Degradation screening assay

Samples of HaloPROTAC test compounds were transferred into black clear-base 384-well tissue culture assay plates (Greiner Bio-One). An Echo acoustic liquid handler (Labcyte) was used for this procedure, with transfer volumes of 200 nL and 20 nL for 10 μM and 1 μM screening, respectively. For concentration response curve format assays, typically a 250x test concentration dilution series of each compound was first prepared in 100% DMSO, with up to 11 concentration points differing in half-log unit increments. This series was next further diluted 25x in Fluorobrite DMEM (Thermofisher), and 5 μL was then transferred to the assay plate. Cultured cells were harvested at 80-90% confluency using TrypLE enzyme, resuspended in Fluorobrite DMEM supplemented with 10% heat inactivated FBS, 2 mM GlutaMAX, 50 U/mL penicillin, 50 μg/mL streptomycin, and seeded into assay plates containing test compounds at a density of 10,000 cells per well. The final assay volume was 50 μL per well. DMSO concentrations in the assays were 0.4% or less. Assay plates were then incubated at 37°C with 5% CO_2 for two days. Hoechst 33342 dye (0.5 μg/mL) was added 30 min before the end of the incubation, and the resulting GFP fluorescence in each well was then measured with a PHERAstar FSX microplate reader (BMG Labtech). Where stated (Figure **3** and Supplementary Figures **S5**, **S6**, and **S8**), GFP levels were also measured using fluorescence imaging or flow cytometry with an IN Cell Analyzer 6000 (GE Healthcare) or BD FACSCanto II (BD Biosciences). All assays included negative 0% effect (ZPE) control wells of DMSO vehicle-treated HaloTag-GFP cells, and positive 100% effect (HPE) control wells of untreated HeLa parental cells, and raw GFP fluorescence values from each well were converted to

percent remaining or percent degradation values for all methods by normalizing to these control wells. The HaloPROTAC screening library as described above also included additional positive control wells corresponding to the known VHL ligand **6** to evaluate the *in situ* chemical synthesis, compound transfer, and assay performance for each 384-well plate.

Hit selection

To select compounds for follow-up from the HaloPROTAC screening library, simple numerical filters were applied to the average (n=3) 10 μM test concentration target degradation percentage values, and hits were defined as compounds causing greater than 50% degradation of the HaloTag-GFP fusion protein, but less than 50% degradation of the HaloTag (D106A)-GFP fusion protein. To make the selection more inclusive, compounds causing greater than 40% degradation of HaloTag-GFP, but less than 30% degradation of HaloTag (D106A)-GFP were also included in the hit definition. A small percentage of outlier control wells were excluded before making the percentage degradation calculations.

RESULTS

Development of a HaloPROTAC - HaloTag-GFP degradation 384-well screen and false-positive counter-screen

The HaloPROTAC screen and false-positive counter-screen designs are shown schematically in Figures **1B** and **1C**. To enable the development and validation of the degradation assay and the false-positive counter-screen, known positive control VHL-HaloPROTACs **7** and **8** (Figure **3A**) were prepared (synthetic details described in the Supplementary Material). Target protein degradation was assessed by measuring the reduction in GFP fluorescence in response to cell treatment with these VHL-HaloPROTACs in both HaloTag-GFP and in HaloTag (D106A)-GFP false-positive counter-screen cell lines. The D106A HaloTag mutation abrogates the ability of the HaloTag protein to covalently bind chloroalkanes, which means that any GFP signal decrease observed in this counter-screen cell line cannot be due to on-target HaloPROTAC activity[32]. As expected, treatment with VHL HaloPROTACs **7** and **8** caused a concentration-dependent reduction in fluorescence signal intensity in the HaloTag-GFP cells, but it had no effect on the fluorescence signal in the HaloTag (D106A)-GFP counter-screen cells (Figure **3**). The fluorescence signal decrease in the HaloTag-GFP cells was more pronounced after two days of incubation as compared to one day (Figure **3C**), and no significant

difference in the normalised response curves was evident between HeLa HaloTag-GFP clones with different HaloTag-GFP expression levels (Supplementary Figures **S5** and **S6**).

To determine the suitability of different approaches to deploy for our screening strategy in 384-well plate format, we investigated a live cell whole-well fluorescence endpoint assay, an imaging-based endpoint assay, and a flow cytometry assay. We found that each of these methods provided approximately equivalent response curves, good reproducibility between technical replicates, and high Z' values (Supplementary Figure **S8**). These studies indicated that there were multiple robust, fit for purpose options to use for the detection of true positive degraders as well as false-positive results which might occur due to growth inhibition, cytotoxicity, or other non-specific effects. For the subsequent screening, brightly fluorescent HeLa HaloTag-GFP and HaloTag (D106A)-GFP mutant cell clones with normal HeLa morphology and growth rate were selected for use, along with a two-day assay incubation time. For data acquisition we selected the whole-well fluorescence endpoint method (PHERAstar microplate reader) because of its efficiency and good performance, and the positive control VHL-HaloPROTAC building block amine **6** (Figure **3A**) was employed as a per-plate standard to validate and control the consistency of the *in situ* chemistry, the acoustic compound transfer efficiency, and the overall assay performance for each 384-well plate.

[INSERT FIGURE 3]

High throughput chemistry optimization and generation of a HaloPROTAC compound screening library

As described above, we designed a simple, one-stage coupling protocol between a collection of diverse amine building blocks **2** and a single primary alkyl chloride N-hydroxysuccinimide (NHS) activated ester building block **4** (Figure **2**) to enable the preparation of a HaloPROTAC library in plate-based format. To optimize the reaction conditions for efficient coupling, a range of bases, reaction times, and reagent stoichiometry was evaluated employing the VHL-HaloPROTAC amine building block **6** (Figure **3A**) as a prototypical amine for array chemistry.

Extensive experimentation demonstrated that robust and reproducible coupling was obtained using DMSO as a solvent with a 1:1 molar ratio of amine **2** to NHS ester **4**, 10 equivalents of *N*-methyl morpholine (NMM) as base, and a reaction time of 16 h. Importantly, all reaction components were evaluated in our HaloTAG-GFP fluorescence assay system and were not observed to cause any detrimental effects on the cells or any decrease in fluorescence signal

(Supplementary Figure **S7**). We also observed that the HaloTag-GFP fluorescence assay results generated with *in situ* prepared HaloPROTAC control compound **7** gave a similar result to those generated using HaloPROTAC control compound **7** that was isolated and purified prior to testing (Supplementary Figure **S7**), suggesting that isolation and purification of the HaloPROTAC library compounds was not required prior to testing in cells. During the course of reaction optimization, it was discovered that airflow-mediated evaporation of NMM in a standard fume hood was associated with substantially incomplete reactions, especially in the corners and along the edges of the plates. To mitigate the poor reaction conversion due to evaporative loss of NMM, we elected to conduct all subsequent plate-based chemistry in a glovebox under a positive-pressure nitrogen atmosphere.

Following plate-based chemistry optimization, we next prepared the screening library of HaloPROTAC test molecules (Figure **2B**). We selected a total of 2934 amines for use as building blocks for HaloPROTAC library synthesis, and employing our optimized reaction conditions, we observed >50% conversion to product for 64% of amines, 20-50% conversion for 6% of amines, and <20% conversion for 30% of amines (Figure **2C**). Factors influencing reaction conversion appear to be both physicochemical (Supplementary Figure **S3**) and steric in nature (Figure **2D**), with steric effects dominating the reaction outcome. Successful conversion rates decreased in the order of secondary cyclic non-hindered > primary non-hindered > primary hindered > secondary linear non-hindered > secondary cyclic hindered > secondary linear hindered amines (Figure **2D** and Supplementary Figure **S2**). With over 1900 library compounds successfully prepared with over 50% conversion via high throughput chemistry, the plates were then assessed in the phenotypic screening assays.

Screening of HaloPROTAC compound library for HaloTag-GFP degrading compounds and false-positive counter-screening

The crude HaloPROTAC compounds at 10 μM concentration were tested for their ability to induce a reduction in GFP fluorescence in the HaloTag-GFP cells and in the HaloTag (D106A)-GFP counter-screen cells (n=3, tested on different days). Screening was also performed at 1 μM concentration (n=1) using the HaloTag-GFP cells in an attempt to detect more potent degraders. The hit activity histograms are shown in Figure **4**. The positive control VHL-HaloPROTAC **7**, which was prepared *in situ* in 16 wells on each assay plate, consistently caused a high level of fluorescence reduction (73% average across all control wells) in the HaloTag-GFP cells, but had minimal effect on fluorescence in the HaloTag (D106A)-GFP

counter-screen cells. We found that a significant proportion of the tested HaloPROTAC screening set compounds caused some reduction in HaloTag-GFP fluorescence, with 6% of test compound wells reducing GFP levels by more than 50%. However, the activity histogram for the HaloTag-GFP cells was similar to that for the HaloTag (D106A)-GFP counter-screen cells, indicating that many of the HaloPROTAC compounds inducing a reduction in GFP fluorescence were likely to be false-positives that were not acting through a targeted degradation mechanism.

To identify the potential functional degraders from the many apparent false-positive hits, we compared the average effects of 10 μM HaloPROTAC compounds in the screening and counter-screening cells (mean, n=3; Figure **4D**) and applied simple thresholds based on the percentage decrease in the GFP fluorescence signal, as described in the Methods section above. This resulted in nine HaloPROTAC compounds being identified as potential true degrader hits. These nine compounds were then individually reprepared from stock solutions in the same manner as described above and retested in a concentration-response format using both the screening and the counter-screening cell lines. We observed that two of these nine compounds (HaloPROTAC **A** and HaloPROTAC **B**) gave reproducibly (n=2) differentiated concentration response profiles between the HaloTag-GFP and the control HaloTag (D106A)-GFP cell lines that were consistent with the response expected for a specific HaloTag-GFP degrader (Figure **5**). To further evaluate these two putative hit compounds, we re-synthesized HaloPROTAC **A** and HaloPROTAC **B** again, this time starting from the *de novo* syntheses of the amine building blocks, and we also included a separate cell viability assay in the testing protocol. Following testing of these newly re-synthesized HaloPROTAC hits in both the screening and counter-screening cell lines, we observed that HaloPROTAC **A** appeared to be a cytotoxic false-positive hit, causing a significant reduction in the intracellular ATP levels in both test cell lines. Surprisingly, we observed that this compound was more cytotoxic against the HaloTag-GFP HeLa cells than against the HaloTag (D106A)-GFP HeLa cells (Supplementary Figure **S9**), which was a confounding result leading to the initial classification of this compound as a true degrader. Unfortunately, we were unable to reproduce the previously observed HaloTag-GFP degradation effect with the reprepared sample of HaloPROTAC **B**, and there was not enough of the original sample remaining to undertake further structural or biological analysis.

[INSERT FIGURE 4]

[INSERT FIGURE 5]

DISCUSSION

Targeted protein degradation is a field with significant potential for industrial and academic research applications, and the first examples of PROTACs entering clinical trials have recently been described[5, 8, 12, 33-35]. Currently reported PROTAC molecules employ a limited catalog of E3 ligases to induce ubiquitin transfer and subsequent target degradation, and increasing the available E3 ligase repertoire for PROTAC design represents a clear opportunity to expand the scope of PROTAC utility for medicines discovery[26]. However, this goal is challenging to realize due to difficulties typically associated with the expression, purification, and screening of biologically relevant E3 ligase complexes. Consequently, limited examples of successful high- or medium-throughput screening approaches for E3 ligases have been reported to date[36-39].

In contrast to a traditional affinity binding or biochemical E3 ligase screening campaign, we envisaged that a phenotypic approach might provide a straightforward solution for the identification of molecules capable of effecting target degradation in cells[40]. Such an approach would have the additional advantage of multiplexed target screening, since the entire cellular repertoire of E3 ligases and other protein degradation machinery would be sampled simultaneously, and we reasoned that a plate-based HaloPROTAC array could provide a good opportunity to implement this approach[41]. HaloPROTACs employ an E3 ligase binder with a linker bearing a primary chloroalkane, and they have been previously demonstrated to covalently bind to HaloTag-POI fusion proteins resulting in ubiquitin transfer and subsequent degradation of the HaloTag-POI fusion. To the best of our knowledge, the HaloPROTAC concept has not been previously used for screening purposes.

To enable our screening strategy, we generated a clonal HeLa cell line expressing the HaloTag-GFP fusion protein, and we tested the performance of these cells with the known VHL HaloPROTACs **7** and **8**. Since we anticipated that many off-target effects unrelated to proteasomal target degradation could result in decreases in the HaloTag-GFP fluorescence signal, we also generated a clonal cell line containing a mutant HaloTag (D106A)-GFP which is unable to covalently bind to HaloPROTAC molecules. This false-positive counter screening cell line was subsequently used to identify compounds causing loss of fluorescence signal due to mechanisms unrelated to protein degradation. We also incorporated an ATP-based cellular

viability assay into the follow up screening protocol to eliminate potentially cytotoxic compounds.

Chimeric molecules such as PROTACs are not widely represented in typical screening libraries, and the ability to access a library of HaloPROTAC degraders is envisaged to be useful for drug discovery and chemical biology purposes. To this end we developed a high-throughput chemistry protocol in 384-well plate format to combine a diverse selection of amine building blocks with a single linker bearing a primary chloroalkane HaloTag recognition element. During library construction, we observed conversion rates to product HaloPROTACs of >50% for approximately 64% of the starting amine building block library, which was deemed to be fit for purpose for the evaluation of an initial degradation screen. We also identified steric and physicochemical parameters such as basicity, solubility, size and flexibility of the amines to be important for controlling reaction rate and ultimate reaction success, which are envisioned to be useful for future library design and preparation.

Screening of the HaloPROTAC library in the HaloTag-GFP HeLa cells as well as counter-screening in the HaloTag (D106A)-GFP cells revealed that many compounds significantly reduced the fluorescence signal in both cell lines. Because the HaloPROTACs are unable to bind the HaloTag D106A mutant[42], we speculate that compounds reducing fluorescence in both cell lines are acting primarily through cytotoxicity or other nuisance mechanisms affecting the proliferation and viability of the cells during the 48 hour assay incubation, rather than through a true on-target degradation mechanism. This effect demonstrates the well-established need to accurately discern false-positives in phenotypic screening campaigns.

We identified nine HaloPROTAC compounds as potential true HaloTag-GFP degrader hits worthy of progression to confirmatory concentration-response testing. Out of these nine compounds, only HaloPROTACs **A** and **B** were found to cause reproducible concentration dependent reduction in HaloTag-GFP fluorescence with a clear window when compared to reduction in the control HaloTag (D106A)-GFP cell line fluorescence. However, upon *de novo* re-synthesis, neither of these compounds were found to be genuine HaloTag-GFP degraders that qualified for further investigation.

The identification of HaloPROTAC **A** as an initial hit appears to have been due to the differential cytotoxicity/cell viability potency of this compound between the HeLa HaloTag GFP screening cells, and the counter-screen HeLa HaloTag (D106A)-GFP cell clones (Supplementary Figure **S9**). Although surprising, this result is not inconsistent with previously

reported variability that has been observed between different HeLa clonal cell lines[43]. In contrast, our inability to reproduce HaloPROTAC **B**-mediated GFP degradation following *de novo* re-synthesis of HaloPROTAC **B** highlights an important caveat of following up hits directly from compound library stock solutions, and emphasizes the need to re-prepare hits on an individual *de novo* basis for assay confirmation. Unfortunately, depletion of the original compound stock solution prevents us from conducting a rigorous structural analysis of the library building blocks used for the preparation of HaloPROTAC **B**. Therefore, the actual structural identity of the HaloPROTAC **B** screening hit is unknown.

Since effective target degradation is dependent on many factors, we speculate that this screening approach can be further improved not only through the addition of more amine components, but also by employing additional linker building blocks with different lengths and compositions to increase the odds of effective ternary complex formation leading to degradation. Furthermore, screening HaloPROTAC libraries in different cell lines may also increase the odds of detecting functional degraders, since different cell lines can have differential expression of degradation machinery component proteins. Future iterations of this screening design might also benefit from other established counter-screening approaches such as co-expression of the HaloTag-GFP with a second fluorescent or bioluminescent control protein[6, 44-47].

Despite our inability to identify novel functional HaloPROTACs with on-target HaloTag-GFP degradation in this study, we believe that our phenotypic screening approach is fit for purpose for the identification of new chemical matter capable of recruiting E3 ligases or other elements of the cellular protein degradation machinery in an unbiased way, and will be a valuable new method for the discovery of target protein degraders.

ASSOCIATED CONTENT

The Supporting Information is available on the Sage Journals website.

AUTHOR INFORMATION

Corresponding Author

* Email markus.a.queisser@gsk.com or m.muelbaier@web.de

Author Contributions

±These authors contributed equally.

FUNDING SOURCES

This work was funded by GlaxoSmithKline PLC, GSK/University of Strathclyde Centre for Doctoral Training in Synthetic and Medicinal Chemistry, EPSRC via Prosperity Partnership EP/S035990/1 and UK MRC grant number MC_UU_00018/6.

ACKNOWLEDGMENT

The authors are grateful to Richard Kasprowicz for help with the analysis of fluorescence images and to Blandine McKay and David Battersby for helpful discussions.

ABBREVIATIONS

PROTAC: Proteolysis Targeting Chimera, VHL: Von Hippel-Lindau.

FIGURE LEGENDS

Figure **1**. Schematic overview. (A) HaloPROTACs **3** are synthesized *in situ* in 384 well plates in high-throughput format using an optimized amide coupling reaction between a reactive haloalkane containing linker **1** and a library of amines **2**. (B) HaloPROTACs **3** are screened in a 384 well format HeLa cell based phenotypic assay for the ability to degrade HaloTag-GFP. This may be achieved by recruitment of an E3 ligase such as with a PROTAC, or via recruitment of other cellular degradation mechanisms. (C) False positives are identified in a counter-screen using a D106A HaloTag mutant that is incapable of binding the HaloPROTAC.

Figure **2**. (A) HaloPROTAC **5** synthesis from the selected amines **2** through an optimised amide coupling with an NHS ester **4**. **(**B) Example LCMS analysis of plate 2 of 9, showing reactions with >50% conversion in yellow, <20% conversion in blue, and in-between in green. (C) Overall outcome for the reactions, out of the 2934 amines. (D) Computational analysis of the reactions, showing successful conversion in yellow and categorised by amine steric hinderance.

Figure **3**. (A) Structures of a reference VHL ligand **6** and two positive control VHL-HaloPROTACs **7** and **8**, and validation of the HeLa HaloTag-GFP screening and HaloTag (D106A)-GFP counter-screening cell lines. (B) Example fluorescence images taken in 384-well format showing differential GFP degradation (fluorescence reduction) from 1 μM control VHL-HaloPROTAC **7** treatment. (C) Concentration response profiles using different

incubation times with fluorescence imaging. Plots use the average values of two technical well replicates. Cells in this experiment were pre-clonal selection, 20,000 cells per well.

Figure **4**. Results of screening the HaloPROTAC set. (A-C) Hit activity distribution histograms. (D) Comparison of screening and counter-screening results using HaloTag-GFP and HaloTag (D106A)-GFP cells, respectively. Line is y = x. One outlier with a high raw fluorescence value is excluded from the figure. Numerical percent activity filters used to select hits are shown.

Figure **5**. Hit confirmation in concentration response format showing reproducible differential HaloTag-GFP vs counter screen HaloTag (D106A)-GFP fluorescence reduction profiles for the two out of nine hits that were confirmed.

REFERENCES

1. Burslem, G. M.; Crews, C. M. Proteolysis-Targeting Chimeras as Therapeutics and Tools for Biological Discovery. *Cell* **2020,** 181, 102-114.
2. Lai, A. C.; Crews, C. M. Induced protein degradation: an emerging drug discovery paradigm. *Nat Rev Drug Discov* **2017,** *16*, 101-114.
3. Churcher, I. Protac-Induced Protein Degradation in Drug Discovery: Breaking the Rules or Just Making New Ones? *J Med Chem* **2018,** *61*, 444-452.
4. Bondeson, D. P.; Mares, A.; Smith, I. E.; et al. Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat Chem Biol* **2015,** *11*, 611-7.
5. Nowak, R. P.; Jones, L. H. Target Validation Using PROTACs: Applying the Four Pillars Framework. *SLAS Discov* **2021**, 26, 474-483.
6. Kastl, J. M.; Davies, G.; Godsman, E.; et al. Small-Molecule Degraders beyond PROTACs-Challenges and Opportunities. *SLAS Discov* **2021**, 26, 524-533.
7. Zhang, Y.; Loh, C.; Chen, J.; et al. Targeted protein degradation mechanisms. *Drug Discov Today Technol* **2019,** *31*, 53-60.
8. Luh, L. M.; Scheib, U.; Juenemann, K.; et al. Prey for the Proteasome: Targeted Protein Degradation-A Medicinal Chemist's Perspective. *Angew Chem Int Ed Engl* **2020,** *59*, 15448-15466.
9. Gadd, M. S.; Testa, A.; Lucas, X.; et al. Structural basis of PROTAC cooperative recognition for selective protein degradation. *Nat Chem Biol* **2017,** *13*, 514-521.
10. Collins, G. A.; Goldberg, A. L. The Logic of the 26S Proteasome. *Cell* **2017,** *169*, 792-806.
11. Verma, R.; Mohl, D.; Deshaies, R. J. Harnessing the Power of Proteolysis for Targeted Protein Inactivation. *Mol Cell* **2020,** *77*, 446-460.
12. Mullard, A. Arvinas's PROTACs pass first safety and PK analysis. *Nat Rev Drug Discov* **2019,** *18*, 895.
13. Mullard, A. Targeted protein degraders crowd into the clinic. *Nat Rev Drug Discov* **2021,** 20, 247-250.
14. Bulatov, E.; Ciulli, A. Targeting Cullin-RING E3 ubiquitin ligases for drug discovery: structure, assembly and small-molecule modulation. *Biochem J* **2015,** *467*, 365-86.

15. Zengerle, M.; Chan, K. H.; Ciulli, A. Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem Biol* **2015,** *10*, 1770-7.
16. Lu, J.; Qian, Y.; Altieri, M.; et al. Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem Biol* **2015,** *22*, 755-63.
17. Tomoshige, S.; Hashimoto, Y.; Ishikawa, M. Efficient protein knockdown of HaloTag-fused proteins using hybrid molecules consisting of IAP antagonist and HaloTag ligand. *Bioorg Med Chem* **2016,** *24*, 3144-8.
18. Mares, A.; Miah, A. H.; Smith, I. E. D.; et al. Extended pharmacodynamic responses observed upon PROTAC-mediated degradation of RIPK2. *Commun Biol* **2020,** *3*, 140.
19. Hines, J.; Lartigue, S.; Dong, H.; et al. MDM2-Recruiting PROTAC Offers Superior, Synergistic Antiproliferative Activity via Simultaneous Degradation of BRD4 and Stabilization of p53. *Cancer Res* **2019,** *79*, 251-262.
20. Chamberlain, P. P.; Hamann, L. G. Development of targeted protein degradation therapeutics. *Nat Chem Biol* **2019,** *15*, 937-944.
21. Martin-Acosta, P.; Xiao, X. PROTACs to address the challenges facing small molecule inhibitors. *Eur J Med Chem* **2021,** *210*, 112993.
22. Schapira, M.; Calabrese, M. F.; Bullock, A. N.; et al. Targeted protein degradation: expanding the toolbox. *Nat Rev Drug Discov* **2019,** *18*, 949-963.
23. Madsen, D.; Azevedo, C.; Micco, I.; et al. An overview of DNA-encoded libraries: A versatile tool for drug discovery. *Prog Med Chem* **2020,** *59*, 181-249.
24. Goodnow, R. A., Jr.; Dumelin, C. E.; Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat Rev Drug Discov* **2017,** *16*, 131-147.
25. Zhu, Z.; Grady, L. C.; Ding, Y.; et al. Development of a Selection Method for Discovering Irreversible (Covalent) Binders from a DNA-Encoded Library. *SLAS Discov* **2019,** *24*, 169-174.
26. Ishida, T.; Ciulli, A. E3 Ligase Ligands for PROTACs: How They Were Found and How to Discover New Ones. *SLAS Discov* **2021**, 26, 484-502.
27. Goodnow, R., Jr. DNA-Encoded Library Technology (DELT) After a Quarter Century. *SLAS Discov* **2018,** *23*, 385-386.
28. Simard, J. R.; Lee, L.; Vieux, E.; et al. High-Throughput Quantitative Assay Technologies for Accelerating the Discovery and Optimization of Targeted Protein Degradation Therapeutics. *SLAS Discov* **2021,** 26, 503-517.
29. Los, G. V.; Encell, L. P.; McDougall, M. G.; et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol* **2008,** *3*, 373-82.
30. Buckley, D. L.; Raina, K.; Darricarrere, N.; et al. HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem Biol* **2015,** *10*, 1831-7.
31. Boss, C.; Hazemann, J.; Kimmerlin, T.; et al. The Screening Compound Collection: A Key Asset for Drug Discovery. *Chimia (Aarau)* **2017,** *71*, 667-677.
32. Neklesa, T. K.; Crews, C. M. Chemical biology: Greasy tags for protein removal. *Nature* **2012,** *487*, 308-9.
33. Roth, S.; Macartney, T. J.; Konopacka, A.; et al. Targeting Endogenous K-RAS for Degradation through the Affinity-Directed Protein Missile System. *Cell Chem Biol* **2020,** *27*, 1151-1163 e6.
34. Gao, H.; Sun, X.; Rao, Y. PROTAC Technology: Opportunities and Challenges. *ACS Med Chem Lett* **2020,** *11*, 237-240.
35. Ding, Y.; Fei, Y.; Lu, B. Emerging New Concepts of Degrader Technologies. *Trends Pharmacol Sci* **2020,** *41*, 464-474.
36. Landre, V.; Rotblat, B.; Melino, S.; et al. Screening for E3-ubiquitin ligase inhibitors: challenges and opportunities. *Oncotarget* **2014,** *5*, 7988-8013.

37. Foote, P. K.; Statsyuk, A. V. Monitoring PARKIN RBR Ubiquitin Ligase Activation States with UbFluor. *Curr Protoc Chem Biol* **2018,** *10*, e45.
38. Rossi, M.; Rotblat, B.; Ansell, K.; et al. High throughput screening for inhibitors of the HECT ubiquitin E3 ligase ITCH identifies antidepressant drugs as regulators of autophagy. *Cell Death Dis* **2014,** *5*, e1203.
39. Tian, M.; Zeng, T.; Liu, M.; et al. A cell-based high-throughput screening method based on a ubiquitin-reference technique for identifying modulators of E3 ligases. *J Biol Chem* **2019,** *294*, 2880-2891.
40. Mayor-Ruiz, C.; Bauer, S.; Brand, M.; et al. Rational discovery of molecular glue degraders via scalable chemical profiling. *Nat Chem Biol* **2020,** *16*, 1199-1207.
41. Heidary, D. K.; Fox, A.; Richards, C. I.; et al. A High-Throughput Screening Assay Using a Photoconvertable Protein for Identifying Inhibitors of Transcription, Translation, or Proteasomal Degradation. *SLAS Discov* **2017,** *22*, 399-407.
42. Neklesa, T. K.; Tae, H. S.; Schneekloth, A. R.; et al. Small-molecule hydrophobic tagging-induced degradation of HaloTag fusion proteins. *Nat Chem Biol* **2011,** *7*, 538-43.
43. Hu, W. E.; Zhang, X.; Guo, Q. F.; et al. HeLa-CCL2 cell heterogeneity studied by single-cell DNA and RNA sequencing. *PLoS One* **2019,** *14*, e0225466.
44. Wolff, N. C.; Pavía-Jiménez, A.; Tcheuyap, V. T.; et al. High-throughput simultaneous screen and counterscreen identifies homoharringtonine as synthetic lethal with von Hippel-Lindau loss in renal cell carcinoma. *Oncotarget* **2015,** *6*, 16951-62.
45. Zanella, F.; Lorens, J. B.; Link, W. High content screening: seeing is believing. *Trends Biotechnol* **2010,** *28*, 237-45.
46. Koren, I.; Timms, R. T.; Kula, T.; et al. The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell* **2018,** *173*, 1622-1635.e14.
47. Paguio, A.; Stecha, P.; Wood, K. V.; et al. Improved dual-luciferase reporter assays for nuclear receptors. *Curr Chem Genomics* **2010,** *4*, 43-9.

SUPPLEMENTARY

A phenotypic approach for the identification of new molecules for targeted protein degradation applications

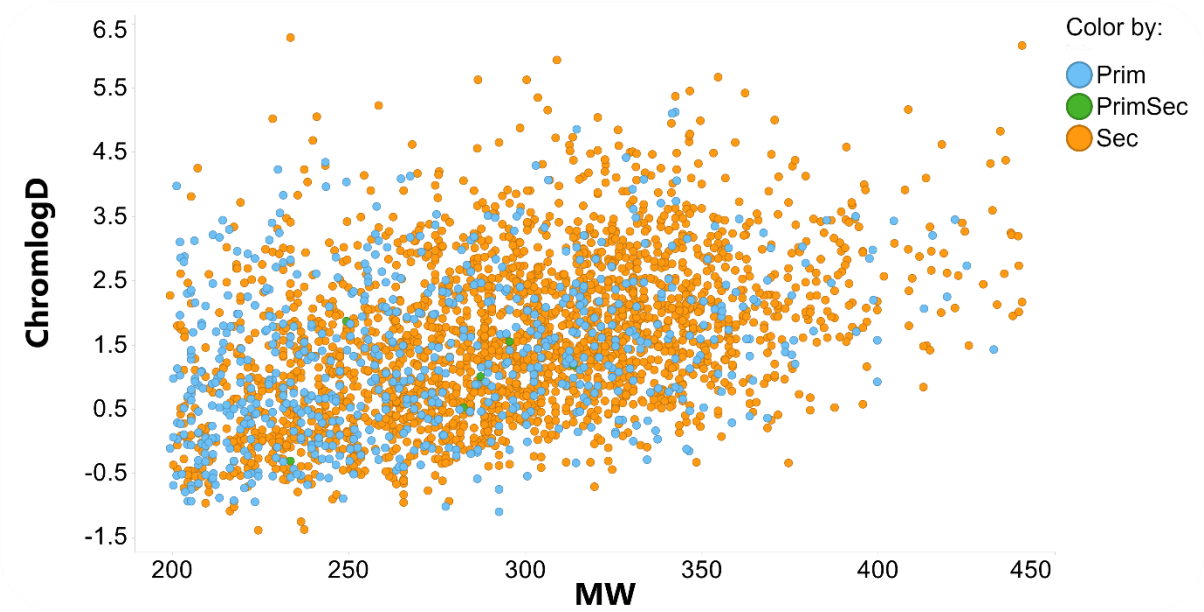
Peter Stacey [a]±, Hannah Lithgow [a][b]±, Xiao Lewell [a], Agnieszka Konopacka [a], Stephen Besley [a], Georgina Green [a], Ryan Whatling [a], Robert Law [a], Sascha Roth [d], Gopal Sapkota [d], Ian E. D. Smith [a] ,Glenn A. Burley [b], John Harling [a], Andrew Benowitz [a], Markus A. Queisser [a]*, Marcel Muelbaier. [c]*

[a] GSK Medicines Research Centre, Gunnels Wood Road, Stevenage, SG1 2NY (UK). [b] Department of Pure and Applied Chemistry, WestCHEM, University of Strathclyde, 295 Cathedral Street, Glasgow, G1 1XL (UK). [c] Cellzome GmbH, GlaxoSmithKline, Meyerhofstrasse 1, D-69117 Heidelberg, Germany. [d] MRC Protein Phosphorylation and Ubiquitinylation Unit (PPU), University of Dundee, Dundee, DD1 5EH, UK.

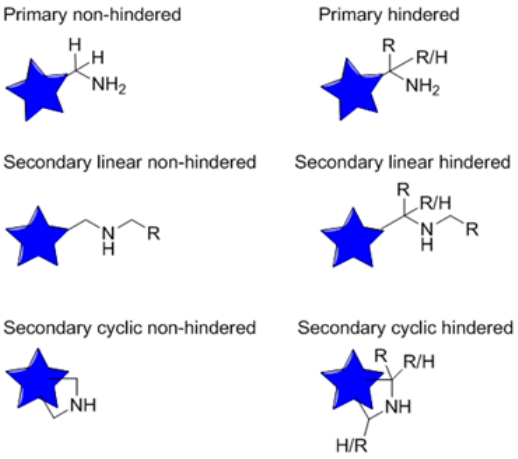
±These authors contributed equally.

* Corresponding author mail: m.muelbaier@web.de or markus.a.queisser@gsk.com

Supplementary Figure S1. Physical chemical space (ChromLogD versus MW) occupied by the selected amines. The proportion of primary to secondary amines was 661:2267.



Supplementary Figure S2. Six types of amines to classify their steric hindrance. Classifications are coded using Daylight SMARTS notation (*https://www.daylight.com/).*



Primary non-hindered: [NH2][CH2,CH3]

Primary hindered: [NH2][X4]([!H])([!H])[H,*]

Secondary linear non-hindered: [NH1;!R]([CH2,CH3])[CH2,CH3]

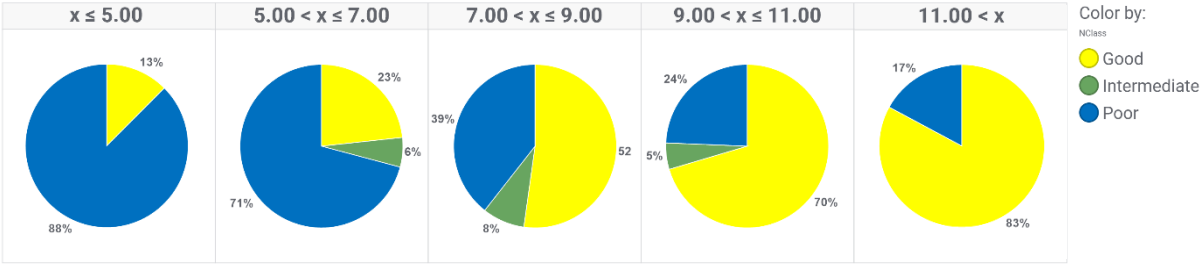
Secondary linear hindered: [NH1;!R]([X4;!S])[X4;!S]([!H])([!H])[H,*]

Secondary cyclic non-hindered: [NH1;R]([CH2])[CH2]

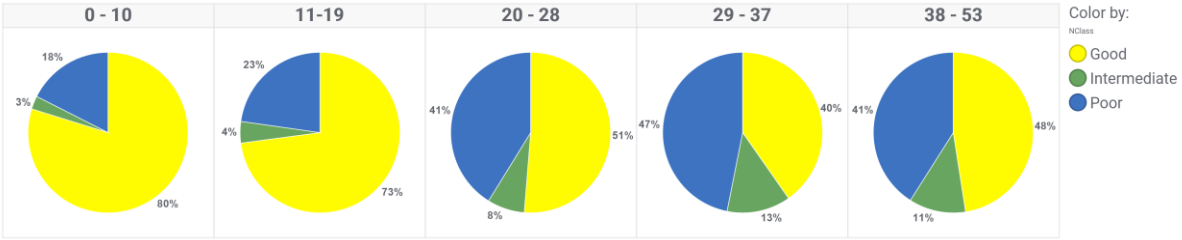
Secondary cyclic hindered: [NH1;R]([X4;!S])[X4;!S]([!H])([!H])[H,*]

Supplementary Figure S3. Apart from steric hindrance, several additional factors of the amines (basicity, flexibility, size, solubility) were found to influence reaction conversions. However, the highest occupied molecular orbital (HOMO) does not appear to affect reaction conversion.

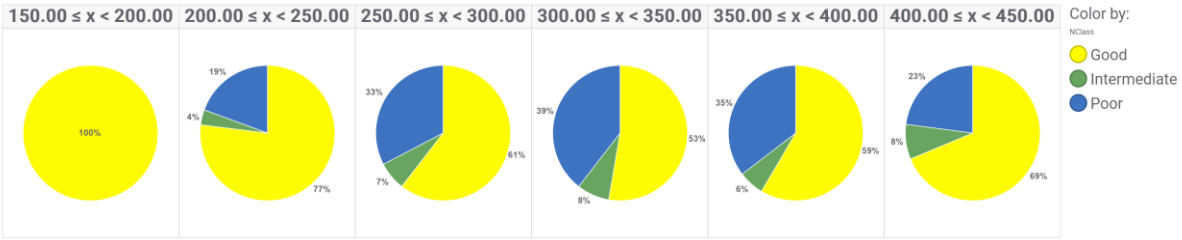
(a) The more basic the amine the more reaction conversion to the desired product mass. Proportion of the amines with good (>50%), intermediate (20-50%), and poor (<50%) conversion. Binned pKa (X=pKa) shows that amines with pKa<5 generally had poor conversion (first pie), whereas amines with pKa>11 generally had good conversion (last pie). (*https://docs.chemaxon.com/display/docs/pKa+calculation#pKacalculation-References)*;



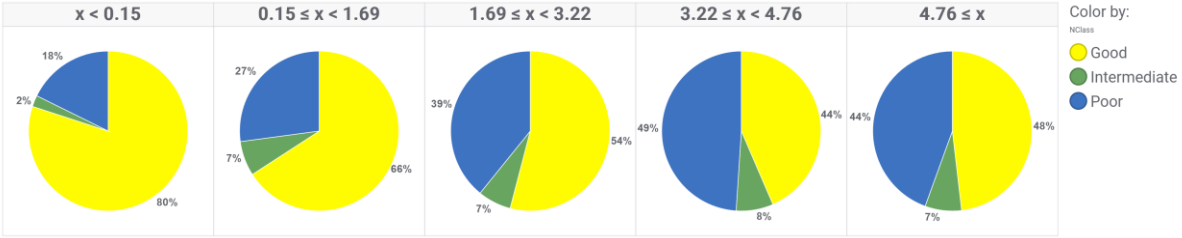
(b) Less flexible amines resulted in better conversion: Binned flexibility for the amines. Flexibility is defined as 100*rotatable bonds/total bonds.



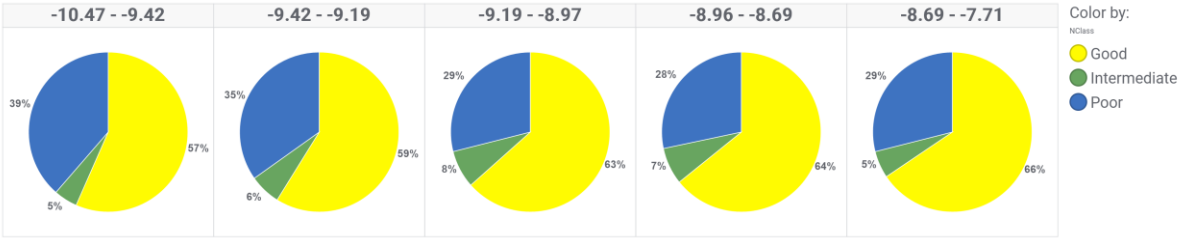
(c) Smaller amines gave better conversion: Binned molecular weight (x=MW) versus proportion of amines with conversion categories.



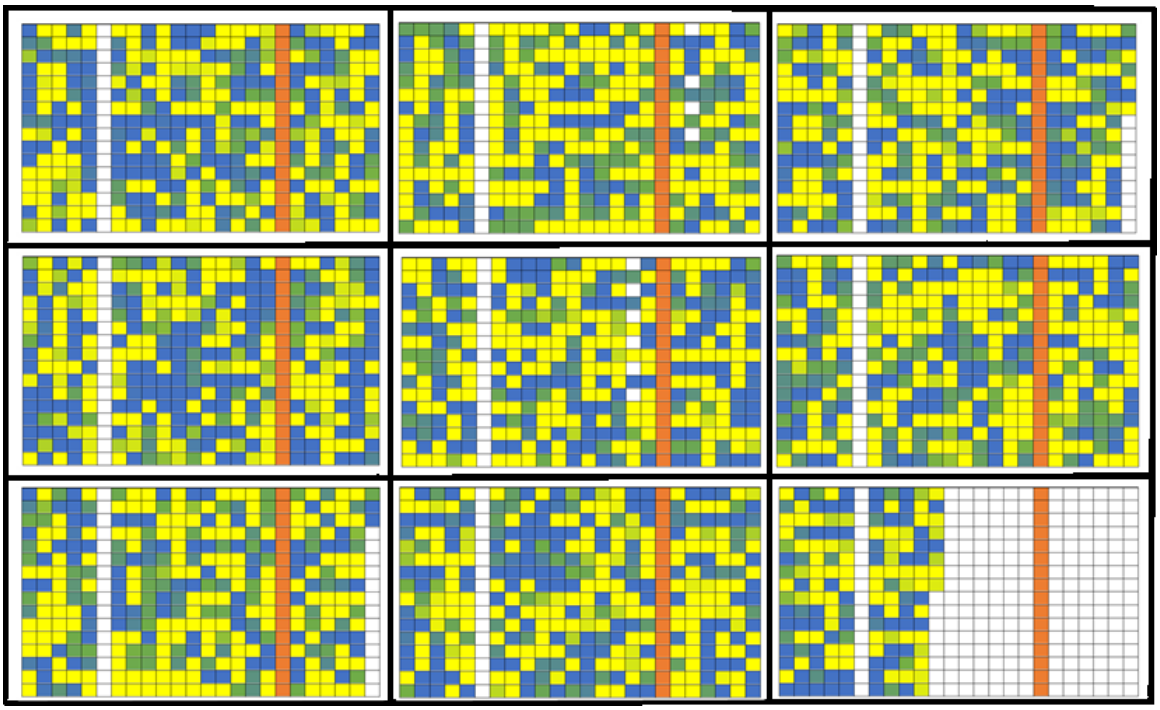
(d) Amines with lower ChromLogD gave better conversions. Binned ChromLogD (x=ChromlogD) of the amines versus conversion categories.



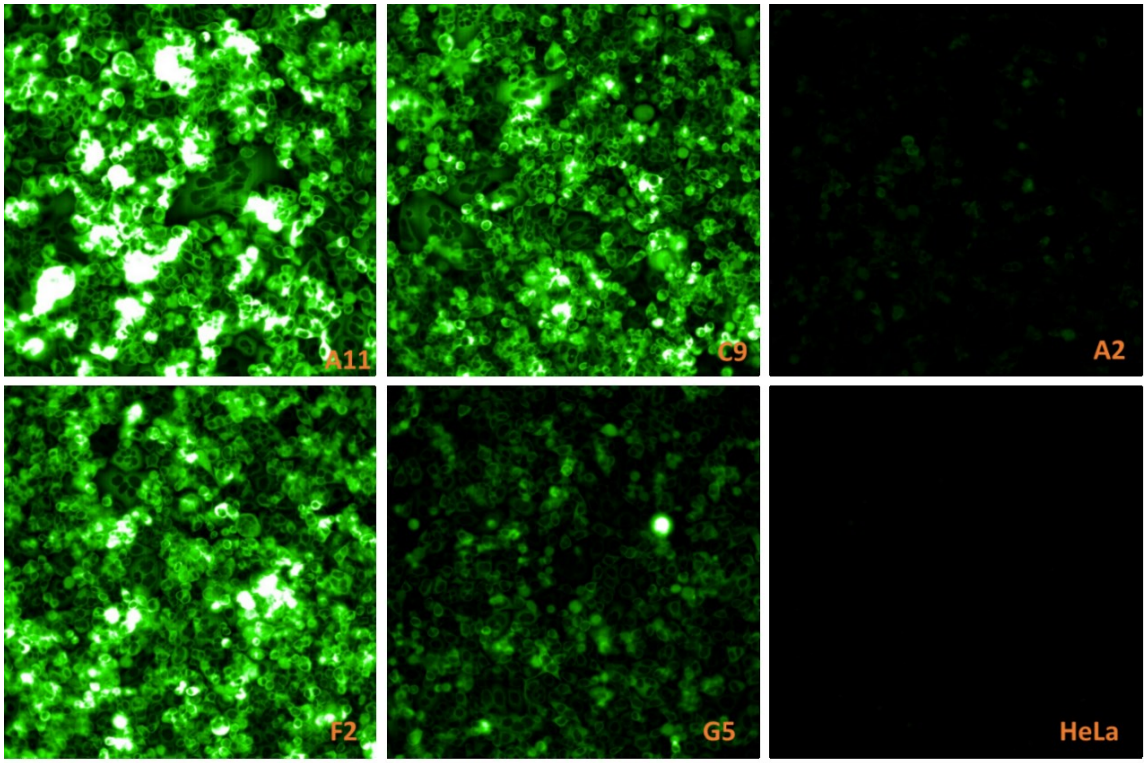
(e) HOMO of the amine does not appear to affect conversion. Binned HOMO energies in eV calculated using AM1 method provided in MOE2018. *(MOE2018: https://www.chemcomp.com/)*



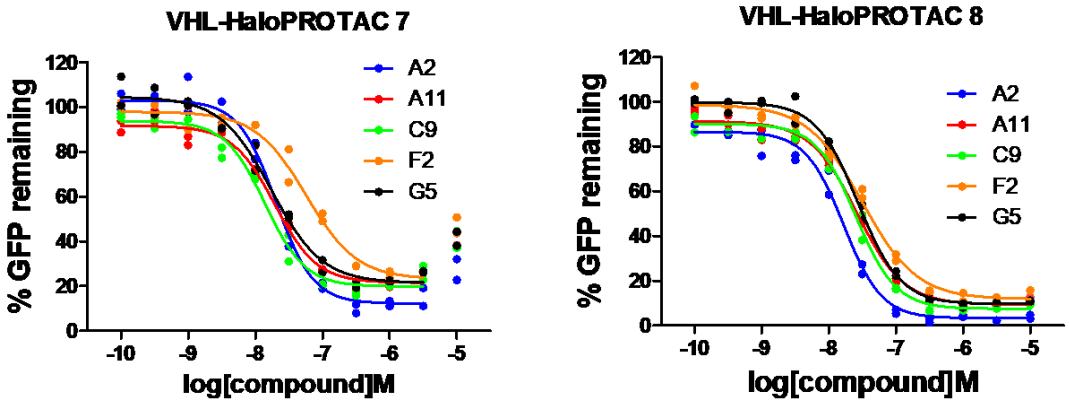
Supplementary Figure S4: High throughput chemistry analysis of the amines showing LCMS conversion to the desired mass, indicative of HaloPROTAC formation. Plates shown with conditional formatting using a three point colour scale from lowest value “0% conversion” in blue to highest value “100% conversion” in yellow with the mid-point percentile “50% conversion” in green. White indicates a DMSO control well, and orange indicates a positive control well containing the VHL HaloPROTAC amine building block (**6**).



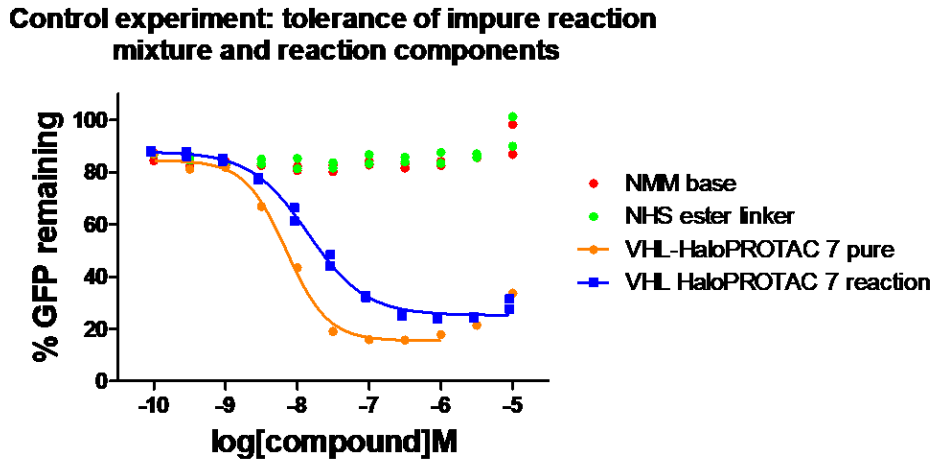
Supplementary Figure S5. Fluorescence images of cells in 384 well format showing the range of HaloTag-GFP expression in 5 HeLa clones. Fluorescence images were generated using an IN Cell 6000 imager.



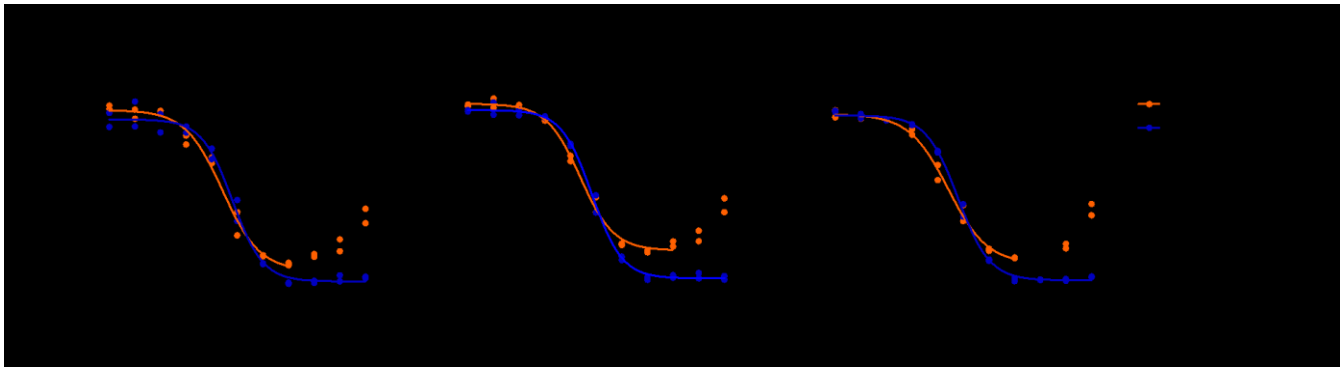
Supplementary Figure S6. Control VHL-HaloPROTAC **7** and **8** responses were similar using alternative HeLa cell clones with different HaloTag-GFP expression levels. The graphs show representative response profiles using an IN Cell 6000 imager in 384 well format. Two well technical replicates are shown for each concentration in each response curve.



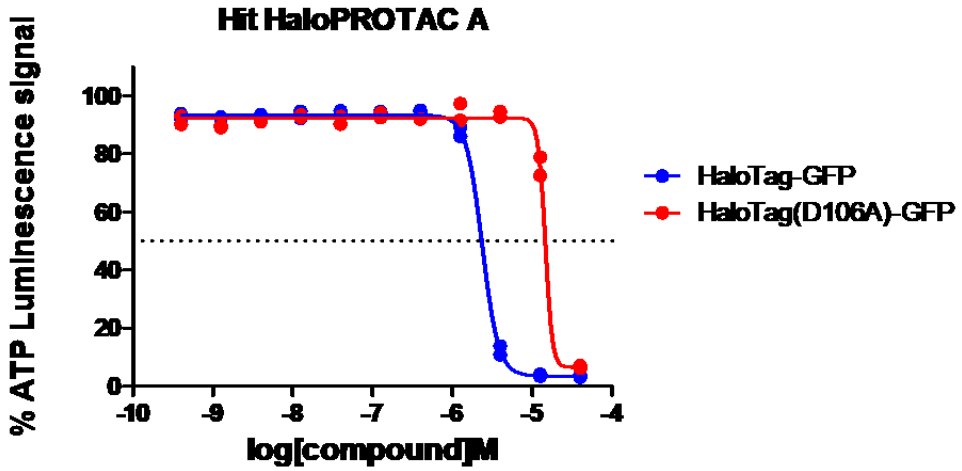
Supplementary Figure S7. Testing of non-purified VHL-HaloPROTAC **7** and tolerance of the assay to the NMM base and the NHS-linker over 72 h incubation time. PHERAstar endpoint reader. Replicate points where shown are same experiment alternative well technical replicates.



Supplementary Figure S8. Alternative 384 well GFP measurement methods gave equivalent HaloTag-GFP degradation response profiles to positive control VHL-HaloPROTACs **7**, and **8**.



Supplementary Figure S9. HaloPROTAC **A** was determined to be a cytotoxic false positive. Reductions in cellular ATP levels in HaloTag-GFP and counter-screen HaloTag (D106A)-GFP cells were measured using the CellTitre-Glo method (Promega) following a 2-day incubation with re-synthesised HaloPROTAC **A**.



Supplementary Chemistry Section

Independent Chemical Synthesis of VHL control compounds

General Considerations

All temperatures are in °C. All commercial solvents, reagents and building blocks are of reagent grade and were used as received without further purification unless otherwise specified.

Nuclear Magnetic Resonance (NMR)

NMR spectra were recorded using a Bruker DPX400, DPX500, AV400, or AVIII600 (with cryoprobe). Chemical shifts (δ) are reported in parts per million (ppm) relative to tetramethylsilane and coupling constants (*J*) in Hz. The following abbre*via*tions are used for multiplicities: s = singlet; br. s = broad singlet; d = doublet; t = triplet; q = quartet; spt = septet; m = multiplet; dd = doublet of doublets. If not specifically stated, the NMR experiments were run at 30 °C.

Liquid Chromatography Mass Spectroscopy (LCMS)

LCMS Method A

The analysis was conducted on an Acquity UPLC BEH C18 column (50 mm × 2.1 mm internal diameter 1.7 µm packing diameter) at 40 °C.

The solvents employed were:

A = 0.1 % v/v solution of formic acid in water.

B = 0.1 % v/v solution of formic acid in acetonitrile.

The gradient employed was as follows:

Time (min)	Flow rate (mL/min)	% A	% B
0	1	97	3
1.5	1	5	95
1.9	1	5	95
2.0	1	97	3

The UV detection was an averaged signal from wavelength of 210 nm to 350 nm and mass spectra were recorded on a Waters ZQ mass spectrometer using alternate-scan positive and negative mode electrospray ionization (ES +ve and ES -ve).

LCMS Method B

The analysis was conducted on an XBridge C18 column (50 mm × 4.6 mm internal diameter 3.5 μm packing diameter) at 30 °C. The solvents employed were:

A = 10 mM ammonium bicarbonate in water adjusted to pH 10 with ammonia solution

B = acetonitrile

The typical gradient employed was as follows:

Time (min)	Flow rate (mL/min)	% A	% B
0	1	97	3
0.05	1	97	3
1.50	1	5	95
1.90	1	5	95
2.00	1	97	3

The UV detection was an averaged signal from wavelength of 210 nm to 350 nm and mass spectra were recorded on a Waters ZQ mass spectrometer using alternate-scan positive and negative mode electrospray ionization (ES +ve and ES -ve).

High Resolution Mass Spectroscopy (HRMS)

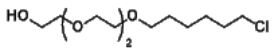
ESI (+) high resolution mass spectra were obtained on a Micromass Q-Tof 2 hybrid quadrupole time-of-flight mass spectrometer, equipped with a Z-spray interface, over a mass range of 100 – 1500 Da, with a scan time of 0.9 s and an interscan delay of 0.1 s. Reserpine was used as the external mass calibrant ($[M + H]^+ = 609.2812$ Da). The Q-Tof 2 mass spectrometer was operated in W reflection mode to give a resolution (FWHM) of 16000 – 20000. Ionization was achieved with a spray voltage of 3.2 kV, a cone voltage of 50 V, with cone and desolvation gas flows of 10-20 and 600 L/h, respectively. The source block and desolvation temperatures were maintained at

120 ºC and 250 ºC, respectively. The elemental composition was calculated using MassLynx v4.1 for the $[M + H]^+$ and the mass error quoted as ppm.

Mass Directed Auto-Preparative (MDAP)

"Mass directed automated preparative HPLC" (MDAP) was conducted on a system such as a Waters FractionLynx system comprising of a Waters 600 pump with extended pump heads, Waters 2700 autosampler, Waters 996 diode array and Gilson 202 fraction collector on an XBridge C18 column (100 mm × 30 mm i.d. 5 µm packing diameter) at ambient temperature, eluting with 10 mM ammonium bicarbonate in water adjusted to pH 10 with ammonia solution (solvent A) and acetonitrile (solvent B) using the appropriate elution gradient. The UV detection was a summed signal from wavelength of 210 nm to 350 nm. The mass spectra were recorded on a Waters ZQ spectrometer using electrospray positive and negative mode (ES +ve and ES -ve). The software used was MassLynx 3.5 with OpenLynx and FractionLynx option or using equivalent alternative systems. Similar systems using Sunfire C18 columns and gradient of solvents such as formic acid (or TFA) in water (solvent A) and acetonitrile (solvent B) were also employed.

2-(2-(2-((6-Chlorohexyl)oxy)ethoxy)ethoxy)ethan-1-ol, **(156)**.



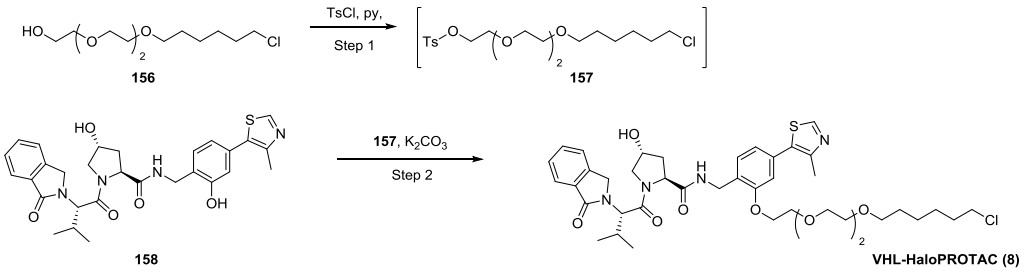
156

To a stirred and cooled 0 °C solution of 2,2'-(ethane-1,2-diylbis(oxy))bis(ethan-1-ol) (3.66 g, 24.3 mmol) in THF (8 mL) and DMF (8 mL) was added portion wise sodium hydride (0.811 g, 20.3 mmol). After 30 minutes at this temperature 1-chloro-6-iodohexane (2.00 g, 8.11 mmol) was added. The resulting mixture was warmed to room temperature over 16 h. The reaction mixture was quenched with water (10 mL) and diluted with 2M HCl aq. (10 mL). The reaction mixture was extracted with CH_2Cl_2 (3 × 50 mL) and the combined organics were washed with sat. aq. LiCl (100 mL) and passed through a hydrophobic frit and concentrated *in vacuo*. The reaction mixture was purified by column chromatography (0–100% cyclohexane in EtOAc) and fractions containing the pure product were concentrated *in vacuo* to afford the desired product **(156)** (1.50 g, 69%) as a yellow oil.

HRMS (ES) calcd for $C_{12}H_{25}ClO_4$ $(M + H)^+$ 269.1520 found 269.1514. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 3.76 – 3.58 (m, 12H), 3.54 (app t, *J* = 6.6 Hz, 2H), 3.51 – 3.44 (m, 2H), 2.51 (br. s, 1H), 1.85 – 1.74 (m, 2H), 1.66 – 1.56 (m, 2H), 1.52 – 1.33 (m, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 72.5, 71.3, 70.6, 70.6, 70.4, 70.1, 61.8, 45.0, 32.6, 29.4, 26.7, 25.4.

Consistent with literature data.[58]

(2*S*,4*R*)-*N*-(2-(2-(2-(2-((6-Chlorohexyl)oxy)ethoxy)ethoxy)ethoxy)-4-(4-methylthiazol-5-yl)benzyl)-4-hydroxy-1-((*S*)-3-methyl-2-(1-oxoisoindolin-2-yl)butanoyl)pyrrolidine-2-carboxamide, **VHL-HaloPROTAC (8)**.



Step 1:

To a stirred solution of 2-(2-(2-((6-chlorohexyl)oxy)ethoxy)ethoxy)ethan-1-ol **(156)** (100 mg, 0.372 mmol) in THF (0.7 mL) was added pyridine (60.2 µL, 0.744 mmol) and 4-methylbenzenesulfonyl chloride (70.9 mg, 0.372 mmol). The resulting mixture was stirred at room temperature for 16 h. The resulting mixture was diluted with CH_2Cl_2 (50 mL) and was washed with 5% Cu(II)SO_4 aq. solution (2 × 50 mL) and brine (50 mL). The organic layer was filtered through a hydrophobic frit and concentrated *in vacuo* to give the crude product **(157)** which was used without further purification.

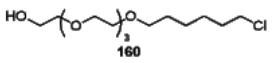
Step 2:

The crude product **(157)** was dissolved in DMF (3 mL) and was added (2*S*,4*R*)-4-hydroxy-*N*-(2-hydroxy-4-(4-methylthiazol-5-yl)benzyl)-1-((*S*)-3-methyl-2-(1-oxoisoindolin-2-yl)butanoyl)pyrrolidine-2-carboxamide **(158)** (202 mg, 0.369 mmol) and potassium carbonate (127 mg, 0.922 mmol). The resulting mixture was stirred at 60 ºC for 16 h. The reaction mixture was cooled to room temperature and purified directly via MDAP (TFA) and fractions containing the pure product were concentrated under a stream of nitrogen to afford the desired product **VHL-HaloPROTAC (8)** (85 mg, 35% over two steps) as a yellow gum.

LCMS (Method B) (ES +ve) *m/z* 799.2 $[M + H]^+$ Rt 1.18 min (93% pure). **HRMS** (ES) calcd for $C_{41}H_{55}ClN_4O_8S$ $(M + H)^+$ 799.3507 found 799.3499. **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 8.77 (s, 1H), 7.84 – 7.75 (m, 1H), 7.58 – 7.48 (m, 1H), 7.49 – 7.41 (m, 2H), 7.37 – 7.30 (m, 2H), 6.99 (dd, *J* = 1.7, 7.6 Hz, 1H), 6.91 (d, *J* = 2.0 Hz, 1H), 4.80 – 4.73 (m, 2H), 4.63 – 4.39 (m, 6H), 4.27 – 4.15 (m, 2H), 3.98 – 3.86 (m, 2H), 3.81 – 3.56 (m, 8H), 3.55 – 3.49 (m, 3H), 3.41 (app t, *J* = 6.6 Hz, 2H), 2.55 (s, 3H), 2.46 – 2.33 (m, 2H), 2.14 – 2.04 (m, 1H), 1.81 – 1.71 (m, 2H), 1.60 – 1.51 (m, 2H), 1.48 – 1.29 (m, 4H), 0.98 – 0.87 (m, 6H). One exchangeable not observed. **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 170.62, 170.09, 169.49, 156.88, 150.69, 147.89, 142.14, 132.14, 131.85, 131.75, 131.62, 129.82, 127.99, 127.13, 123.82, 122.85, 122.03, 112.73, 71.20, 70.79, 70.55, 70.54, 70.02, 69.61, 68.00, 58.69, 58.57, 56.06, 47.47, 45.01, 39.23, 36.37, 36.34, 32.52, 29.37, 28.81, 26.67, 25.38, 19.09, 15.76. Note that ^{13}C NMR data are reported to 2 decimal places to differentiate signals. **IR** ν_{max} (neat) 3328, 2941, 2862, 1667, 1643, 1469, 1446, 1408, 1124 cm^{-1}.

Consistent with literature data.[58]

18-Chloro-3,6,9,12-tetraoxaoctadecan-1-ol, **(160)**.



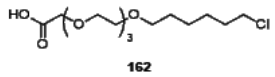
160

To a cooled 0 °C solution of 2,2'-((oxybis(ethane-2,1-diyl))bis(oxy))bis(ethan-1-ol) (4.73 g, 24.3 mmol) in THF anhydrous (40 mL) was added portion wise sodium hydride (0.811 g, 20.3 mmol). The resulting mixture was stirred at this temperature for 0.5 h before 1-chloro-6-iodohexane (2.00 g, 8.11 mmol) was added dropwise. The resulting mixture was warmed to room temperature over 16 h. The reaction was quenched with H_2O (50 mL), diluted with 2M HCl aq. (50 mL) and extracted with chloroform (3 × 100 mL) and the organic layers were passed through a hydrophobic frit before concentrating *in vacuo*. The crude product was then purified via silica FCC (washed with 50:50 EtOAc in cyclohexane and eluted with 100% EtOAc) and fractions containing the pure product were concentrated *in vacuo* to afford the desired product **(160)** (1.00 g, 39%) as a light-yellow oil.

^{1}H NMR (400 MHz, $CDCl_3$) δ = 3.74 – 3.68 (m, 2H), 3.68 – 3.55 (m, 14H), 3.52 (app t, *J* = 6.6 Hz, 2H), 3.45 (app t, *J* = 6.6 Hz, 2H), 2.74 (br. s, 1H), 1.84 – 1.71 (m, 2H), 1.62 – 1.53 (m, 2H), 1.50 – 1.31 (m, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ = 72.47, 71.12, 70.52, 70.50, 70.26, 70.01, 61.62, 44.91, 32.46, 29.33, 26.59, 25.32. Note that ^{13}C NMR data are reported to 2 decimal places to differentiate signals.

Consistent with literature data.[252]

18-Chloro-3,6,9,12-tetraoxaoctadecanoic acid, (162).

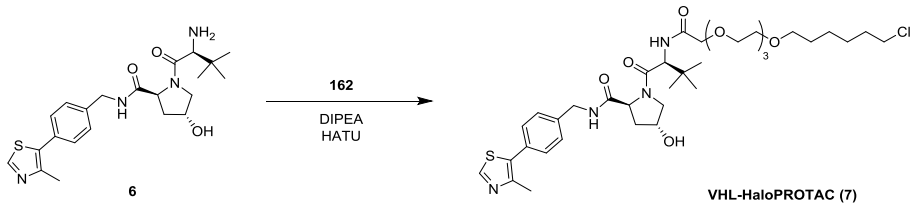


162

To a stirred and cooled 0 °C solution of 18-chloro-3,6,9,12-tetraoxaoctadecan-1-ol **(160)** (1.40 g, 4.48 mmol) in acetone (10 mL) and $NaHCO_3$ (sat. aq. 12.5 mL) was added TEMPO (69.9 mg, 0.448 mmol) and potassium bromide (107 mg, 0.895 mmol). To this resulting mixture was added trichloroisocyanuric acid (2.06 g, 8.95 mmol) over 15 minutes. The reaction mixture was warmed to room temperature over 16 h. The reaction mixture was then diluted with H_2O (50 mL) and 2M HCl aq. (50 mL) and extracted with chloroform (3 × 100 mL) and the combined organics were concentrated *in vacuo* to give the desired product **(162)** (1.32 g, 90%) as a colourless oil.

HRMS (ES) calcd for $C_{14}H_{27}ClO_6$ $(M + H)^+$ 327.1574 found 327.1581. **^{1}H NMR** (400 MHz, DMSO-d_6) δ = 12.46 (br. s, 1H), 4.02 (s, 2H), 3.64 – 3.46 (m, 14H), 3.38 (app t, J = 6.4 Hz, 2H), 1.77 – 1.66 (m, 2H), 1.50 (app quint, J = 7.0 Hz, 2H), 1.43 – 1.27 (m, 4H). **^{13}C NMR** (101 MHz, DMSO-d_6) δ = 171.54, 70.13, 69.78, 69.75, 69.70, 69.43, 67.58, 45.29, 31.97, 29.00, 26.06, 24.88. Note that ^{13}C NMR data are reported to 2 decimal places to differentiate signals. Two signals not observed, potentially due to overlapping frequencies of PEG chain carbons.

(2*S*,4*R*)-1-((*S*)-2-(*tert*-Butyl)-21-chloro-4-oxo-6,9,12,15-tetraoxa-3-azahenicosanoyl)-4-hydroxy-*N*-(4-(4-methylthiazol-5-yl)benzyl)pyrrolidine-2-carboxamide, VHL-HaloPROTAC (7).



To a stirred solution of (2*S*,4*R*)-1-((*S*)-2-amino-3,3-dimethylbutanoyl)-4-hydroxy-*N*-(4-(4-methylthiazol-5-yl)benzyl)pyrrolidine-2-carboxamide **(6)** (39.5 mg, 0.0920 mmol) in DMF (1 mL) was added 18-chloro-3,6,9,12-tetraoxaoctadecanoic acid **(162)** (30.0 mg, 0.0920 mmol), DIPEA (64.1 µL, 0.367 mmol) and HATU (34.9 mg, 0.0920 mmol) sequentially. The resulting mixture was stirred at room temperature for 1 h and purified directly via MDAP (FOR) and appropriate fractions were concentrated under a stream of nitrogen to afford the desired product **VHL-HaloPROTAC (7)** (12 mg, 18%) as a yellow gum.

LCMS (Method B) (ES +ve) *m/z* 739.2 $[M + H]^+$ Rt 1.15 min (>95% pure). **HRMS** (ES) calcd for $C_{36}H_{55}ClN_4O_8S$ $(M + H)^+$ 739.3507 found 739.3502. **^{1}H NMR** (400 MHz, DMSO-d_6) δ = 8.70 (br. s, 1H), 7.41 – 7.33 (m, 4H), 7.30 – 7.29 (m, 1H), 4.76 (app t, J = 8.1 Hz, 1H), 4.61 – 4.46 (m, 3H), 4.39 – 4.31 (m, 1H), 4.14 – 4.11 (m, 1H), 4.07 – 3.95 (m, 2H), 3.70 – 3.50 (m, 15H), 3.45 (app t, J = 6.6 Hz, 2H), 2.64 – 2.56 (m, 1H), 2.53 (s, 3H), 2.19 – 2.09 (m, 1H), 1.85 – 1.73 (m, 2H), 1.65 – 1.55 (m, 2H), 1.51 – 1.34 (m, 4H), 0.96 (s, 9H). 2 exchangeables not observed. **^{13}C NMR** (101 MHz, DMSO-d_6) δ = 171.53, 170.66, 170.58, 150.35, 148.42, 138.12, 131.63, 130.93, 129.53, 128.16, 77.20, 71.25, 71.20, 70.61, 70.59, 70.48, 70.38, 70.16, 70.07, 58.32, 57.26, 56.63, 45.01, 43.29, 35.68, 34.76, 32.54, 29.44,

26.69, 26.40, 25.41, 15.99. Note that ^{13}C NMR data are reported to 2 decimal places to differentiate signals. **IR** ν_{max} (neat) 3312, 2931, 2865, 1633, 1524, 1435, 1105 $cm^{-1.}$

Supplementary Computational Chemistry Section

Computational Selection of the Amine Set

In a first step, suitable chemical matter to undergo the anticipated amide coupling reaction was identified. Amines were chosen to ideally ensure that the compounds would be reactive under the anticipated reaction conditions, would be detected by UV analytical techniques, were diverse in structure and were predicted to be cell permeable. Starting from the GSK compound collection, the following selection criteria were applied to identify such amines. In a first step, the collection was filtered for aliphatic amines, because of previous unsuccessful or very low-yielding reactions of anilines or heteroaromatic amines with the active ester of choice, *N*-hydroxy succinimide (NHS) (*Boss, C.; Hazemann, J.; Kimmerlin, T.; von Korff, M.; Luthi, U.; Peter, O.; Sander, T.; Siegrist, R. Chimia (Aarau). 2017, 71, 667).* Second, to allow for easy reaction monitoring by LCMS combined with UV detection, a requisite of 1 to 3 aromatic rings was added as additional structural filter. This subset was then further refined in a third step using an in-house chemical diversity selection algorithm. Finally, the resulting compounds were filtered based on physicochemical properties (ChromLogD <7, PFI <7, PSA <140, HD <5, MW<450). Overall, a total of 2934 amines met the above criteria and were available in sufficient quantities as DMSO stock solutions from the GSK chemical stores.

Computational Analysis of Factors Influencing Reaction Conversion Rate

Out of the 2934 amines used for synthesis, 64% of amines had >50% product conversion, 6% had 20-50% conversion, and 30% had <20% conversion. To improve our library design for future screening iterations, we analyzed factors influencing conversion rates. Typically, the rate and yield of chemical reactions depend on several external and intrinsic factors. External factors, such compound concentration, equivalents of reactants, and reaction atmosphere are held constant under high throughput chemistry conditions for all individual reactions. Intrinsic factors depend on the properties of the reaction centers, i.e. the amine. A range of simple physicochemical descriptors were calculated (Supplementary Figure **S3**). Additionally, it was noted that steric hindrance around the amine functionality might also play a role in reaction conversion rates and therefore we have included steric hindrance using Daylight SMARTS notation to define various types of amines (Supplementary Figure **S2**).

Our analysis shows that although most physicochemical factors have an influence on reaction conversion (Supplementary Figure **S3**), steric hindrance is the primary driver. Successful conversion

rates decreased in the order of secondary cyclic non-hindered > primary non-hindered > primary hindered > secondary linear non-hindered > secondary cyclic hindered > secondary linear hindered (Figure **2D**). Interestingly, secondary linear non-hindered amines were less reactive than expected, in particular when compared with secondary cyclic non-hindered amines. This could potentially be explained by the more flexible conformation making the amine functionality less exposed under given reaction conditions.