



A Complete Pipeline for the Discovery and Development of Haptenizing Chimeras at Scale

Citation

Dhaliwal, Rohil. 2025. A Complete Pipeline for the Discovery and Development of Haptenizing Chimeras at Scale. Bachelors Thesis, Harvard University Engineering and Applied Sciences.

Link

<https://dash.harvard.edu/handle/1/42719542>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.

Please share how this access benefits you. [Submit a story](#)

A Complete Pipeline for the Discovery and Development of Haptenizing Chimeras at Scale

A thesis presented by

Rohil Dhaliwal

to

the Faculty of

Arts and Sciences and the Harvard John A. Paulson School of Engineering and Applied Sciences

in partial fulfillment of the requirements for

the Bachelor of Arts degree with honors in

Chemistry and Physics and Biomedical Engineering

Faculty Adviser: Professor Amit Choudhary

Harvard University

Cambridge, MA

March 28, 2025

In submitting this thesis to the Harvard John A. Paulson School of Engineering and Applied Sciences in partial fulfillment of the requirements for the degree with honors of Bachelor of Arts, I affirm my awareness of the standards of the Harvard College Honor Code.

Name: Rohil Dhaliwal

Signature: _____

The Harvard College Honor Code Members of the Harvard College community commit themselves to producing academic work of integrity – that is, work that adheres to the scholarly and intellectual standards of accurate attribution of sources, appropriate collection and use of data, and transparent acknowledgement of the contribution of others to their ideas, discoveries, interpretations, and conclusions. Cheating on exams or problem sets, plagiarizing or misrepresenting the ideas or language of someone else as one's own, falsifying data, or any other instance of academic dishonesty violates the standards of our community, as well as the standards of the wider world of learning and affairs.

ACKNOWLEDGMENTS

Science is supremely collaborative. Thank you to everyone in my lab who has guided me, taught me, and made this work possible. Thank you, Endri, for being my first mentor, for introducing me to chemistry, and for being exceedingly patient as my inexperience affected many experiments. Your kindness allowed me to learn and grow rather than fear failure. Thank you, Kien, for taking on the time-consuming (and probably painful) task of guiding me for the past year. Your openness to answering all my incessant questions, both in the middle of your work and in the middle of the night, has allowed me to remain reassured that I am going in the right direction. Thank you, Prashant, for taking the time to not just work with me but explain every aspect of each biophysical and bacterial experiment that we did, and for always advocating for me. Thank you, Rajaiah, for accommodating my sporadic and inconsistent schedule that is often incompatible with the life of cells, and for reassuring me that it was okay (even though it wasn't) that I dropped a 96-well-plate with all your results. Thank you, Shaimaa, Sameek, Manuel, and Karishma, for helping me out in lab and speaking with me about my classes, my interests, graduate school, and everything else scientific and non-scientific. I appreciate every single one of our conversations. Thank you, Sunny, for keeping me company as we make peptide solutions, and for allowing the lab to feel less daunting. And thank you, Professor Choudhary, for choosing to respond to my email two years ago and allowing me to join and work as part of this lab. Thank you to the Broad Institute for supporting this work, and to the American Chemical Society for funding part of this project.

Thank you to my roommates for making sure that I have a home to return to after long nights in lab. Thank you to my tutor, Sam, for being such an inspiring role model, for helping me understand the most confusing scientific papers, and for holding crepe study breaks. Thank you to my friends for making Harvard what it is. The best part of my time here has been the people. Thank you to my family for making sure all my worries in life are little ones.

It is both a joyful and painful time to be a scientist. The possibilities of work like this make me optimistic that we will continue to find breakthroughs and novel ways to treat those in need. At the same time, the current administration has attacked science to an unprecedented and catastrophic extent. During the course of my work, many of my closest and most hardworking colleagues and friends have lost funding and lost their jobs. Much of the work of this thesis has had to adapt

significantly to funding changes. We have been waiting on funding to order the UniTEs and reagents needed to continue this work. Experiments in the lab had been paused for months as the lab devoted its time to grant writing. Science is what makes humans great. I hope for a future in which it is embraced as such.

STATEMENT OF WORK

It is important to note that much of this work was done collaboratively. Peptide and small molecule synthesis was performed under the supervision of Kien Tran (Postdoctoral Associate) and Endri Karaj (Postdoctoral Associate). T2 cell assay and other cell-based assays were performed under the supervision of Rajaiah Pergu (Postdoctoral Associate) and Chaiheon Lee (Postdoctoral Associate). Computational work was done with Ananthan Rajagopalan (Graduate Student). MHC refolding assays were performed under the supervision of Prashant Singh (Postdoctoral Associate). FKBP-based peptide synthesis and MOE docking analysis was done with Sunny Shi (Undergraduate Student). Ongoing hydrophobic linker work (biphenyl and silyl groups) is being done by Kien Tran and Sunny Shi. Complete HaCs synthesis was done by Endri Karaj and Shaimaa Sindi (Postdoctoral Associate). Figures were made in conjunction with Endri Karaj, Shaimaa Sindi, Prashant Singh, and Rajaiah Pergu. Much of this work is part of a paper that has been submitted to the Journal of the American Chemical Society (JACS) for publication. I cannot directly cite the paper since it is not yet published, but many of the figures and brief parts of the text are taken from this manuscript (of which I am an author).

ABSTRACT

Covalent inhibitors are increasingly popular therapeutics. However, these drugs suffer from acquired resistance. A new therapeutic approach to treating cancer builds upon existing covalent inhibitors to repurpose them as Haptenizing Chimeras (HaCs), immunotherapeutic agents, to not just combat acquired resistance but function as a novel treatment modality to combat cancers of all types. The strength of HaCs lies in their versatility and scalability. We report a complete pipeline that leverages computational tools, cell-based assays, and solid-phase peptide synthesis to vastly scale the scope and efficiency of the discovery and development of HaCs. This pipeline greatly expanded the therapeutic reach of HaCs, having identified tens of thousands of new and promising targets. Utilization of the pipeline has also contributed to a near 3-times increase in efficacy of the molecules through linker exploration and discovery. It also allows us to respond to patient needs in real time. Finally, the pipeline takes a step towards turning the intracellular space into extracellular markers, potentially revolutionizing the ability to develop targeted therapeutics for different cancers.

Table of Contents

1	INTRODUCTION	8
1.1	BACKGROUND	8
1.2	COVALENT INHIBITORS AND THE IMMUNE SYSTEM.....	10
1.3	OUR APPROACH.....	12
1.4	INITIAL WORK.....	14
1.5	AIMS.....	16
2	RESULTS	18
2.1	T2 CELL ASSAY FOR RAPID LINKER TESTING	18
2.2	HAPTENATED PEPTIDE SYNTHETIC SCHEME	19
2.3	ADDITIONAL WORKFLOW FOR ACID-LABILE COMPOUNDS	23
2.4	INITIAL VALIDATION OF SYNTHESIS AND T2 ASSAY	26
2.5	HAPLOTYPE DEPENDENCY IN T2 CELLS	30
2.6	COMPUTATIONAL PREDICTIONS	31
2.7	FINAL VALIDATION.....	33
2.8	LYSINE COMPATIBILITY	36
2.9	HIGH THROUGHPUT IDENTIFICATION OF HACs.....	37
2.10	REAL-TIME PATIENT RESPONSE.....	39
2.11	BIOPHYSICAL ASSAY DEVELOPMENT.....	40
2.12	HYDROPHOBIC LINKERS AND POCKET TARGETING.....	44
3	MATERIALS AND METHODS.....	47
3.1	PEPTIDE SYNTHESIS.....	47
3.2	GENERAL CHEMICAL SYNTHESIS.....	48
3.3	T2 CELL ASSAY	48
3.4	COMPLETE HAC COMPOUND CELL ASSAY.....	48
3.5	COMPUTATIONAL PREDICTIONS	49
3.6	MHC REFOLDING.....	50
3.7	STATISTICAL ANALYSES	51
3.8	FIGURES.....	52
4	DISCUSSION.....	53
4.1	FUTURE WORK	53
4.2	CONCLUSION	55
5	REFERENCES.....	56
6	SUPPLEMENTARY FIGURES.....	67

1 INTRODUCTION

1.1 BACKGROUND

Despite being the most researched disease, cancer still claims 10 million lives annually (1). Cancers are often driven by somatic mutations in genes referred to as proto-oncogenes (2). Until recently, a large group of the oncogenic proteins in many cancers were considered undruggable (3,4). These undruggable targets represented a large pitfall in the field of oncology, as mutations in these oncoproteins are found in a large share of cancers (3,4). The paradigmatic example of this is the Kirsten rat sarcoma (KRAS) gene that encoded for the KRAS guanosine triphosphatase (GTPase) which is part of the rat sarcoma (RAS) signaling pathway. Somatic KRAS mutations are the most common activating mutation in human cancer (3–6). KRAS mutations are found in 11.6% of all carcinomas, with wide variation between tumor types and different geographic areas, and are especially present in Western countries (7,8). KRAS mutations are found in nearly 82% of patients with pancreatic ductal adenocarcinoma, 40% of patients with colorectal cancer, and over 21% of patients with non-small cell lung cancer (NSCLC) (7–10). Cancers with these mutations tend to be quite difficult to treat and are associated with poor response to therapy and poor outcomes (5).

KRAS was considered undruggable given its picomolar affinity for GTP/GDP and its lack of any known allosteric sites and binding pockets (which are the typical target for most drugs) (3–5). KRAS mutations inhibit GTP hydrolysis, leading to higher residence time of GTP, and thus an increased amount of the active, GTP-bound KRAS isoform (5). In 2014, Kevan Shokat and colleagues discovered a novel pocket in the mutant KRAS and demonstrated mutant-specific inhibition of KRAS (5). Specifically, their compound targeted the G12C mutation in KRAS (5). The thiol side chain of the mutated cysteine residue provided a nucleophilic target with which the molecule could covalently bind to KRAS. The mechanism of action of the compound allowed it to selectively inhibit mutant KRAS G12C while leaving wild-type KRAS unharmed (5). This work would lead to the creation of Amgen’s Sotorasib which has significantly prolonged survival in patients with NSCLC, especially in groups with poor prognosis (11). Since then, new KRAS G12C drugs such as Adagrasib have also received FDA approval and have been used successfully as a monotherapy (10,12).

These drugs are known as covalent inhibitors due to the nature of the bond formed with their targets. Covalent inhibitors bind irreversibly to the side chains of specific amino acids in a target protein, usually through the nucleophilic thiol or amine side chains of a cysteine or lysine residue respectively (5,13,14). The covalence of the bond formed leads to increased affinity compared to non-covalent inhibitors, which form transient complexes with their targets through hydrogen bonding, pi stacking, van der Waals force, and other non-covalent bonds, which may result in both lower selectivity and inhibition ability as they dissociate from their complexes (14–18). Covalent inhibitors subsequently benefit from increased residence time and thus sustained inhibition compared to non-covalent alternatives. The target is ‘permanently’ inhibited, up until degradation (15). This can lead to benefits such as lower dosages and improved compliance with the drug (13,14). Additionally, non-covalent inhibitors can typically be competed off by high concentrations of the endogenous ligand (19). This is especially problematic with kinases as intracellular concentrations of ATP are quite high (19). An important and novel covalent inhibitor is Nirmatrelvir, which inhibits SARS-CoV-2 through the targeting of a cysteine residue, and is part of Paxlovid which has been widely used for COVID-19 treatment (14).

The reach of covalent inhibitors is not limited to undruggable targets. Covalent inhibitors have been effectively utilized for targeting mutations in various other proteins, include Bruton Tyrosine Kinase (BTK), Epidermal Growth Factor Receptor (EGFR), Fibroblast Growth Factor Receptor (FGFR), Interleukin-2-Inducible T cell Kinase (ITK), and more (20–30). Covalent inhibitors represent increasingly popular therapeutics, with eight key covalent inhibitors being approved for cancer treatments between 2013 and 2021 alone (26). These include the BTK inhibitors zanubrutinib and ibrutinib (a World Health Organization essential medicine), and Pfizer’s Dacomitinib, an EGFR inhibitor (22,26).

Despite their potential therapeutic benefits, large groups of patients do not respond to covalent inhibitors (11,31,32). Additionally, even when initially effective, covalent inhibitors suffer from a dramatic loss of efficacy known as acquired resistance (12,33–35). In almost all cases, the cancer develops an additional mutation that renders the treatment ineffective, thus ‘acquiring resistance’ against the drug (12,24,36,37). These patients typically have quite poor outcomes (12).

1.2 COVALENT INHIBITORS AND THE IMMUNE SYSTEM

In the case of KRAS, however, it has been observed that acquired resistance does not typically emerge from mutations in the drug-binding pocket (on-target mutations) (12,37). Instead, most cases occur due to a mutation downstream in the RAS signaling pathway or in a parallel pathway (12,37). As a result, inhibiting the target protein (such as KRAS) is no longer sufficient to disrupt the signaling and block tumor growth (12). Notably, the G12C mutation is preserved in over 80% of cases (12,37). The conservation of the mutation and binding pocket indicates that the inhibitors can still bind to their targets. This crucial observation presents an opportunity to repurpose these now ineffective covalent inhibitors as powerful immunotherapeutic agents.

After a protein is degraded by the proteasome, the resulting peptides are transported into the endoplasmic reticulum the Transporter associated with Antigen Processing (TAP) and subsequently loaded onto the Major Histocompatibility Complex (MHC) Class I, which processes the peptides (38). These peptides are then presented extracellularly by the MHC I (Figure 1A) (38). A cytotoxic CD8+ T-cell can then recognize the peptide-MHC complex on the cell surface, and, if it identifies the peptide as foreign, will release cytotoxic granules such as perforin and granzymes that lead to the targeted death of that cell (38). It was observed for the case of mutant KRAS, as well as EGFR and other oncoproteins, that peptides containing the mutation (e.g. G12C) are often presented extracellularly by the MHC I and recognized by T-cells as a neoantigen (39,40). This phenomenon has been observed in human patients with mutant KRAS (39,41,42). This led to the development of antibodies that target the mutant-peptide-MHC complex (39,40). In theory, these antibodies would specifically recognize only mutant-peptide complexes. However, given that these are usually single amino acid mutations, it is quite difficult to ensure that the antibody binds the mutant while exhibiting no reactivity towards the wild-type peptide (39,43–46). This leads to poor target specificity (39,43–46). In addition, the mutant peptides are not presented by the MHC at a high enough rate to serve as a useful marker.

Covalent inhibitors are particularly useful here. Due to the irreversible bond, they remain linked to the target amino acid residue (and thus to the peptide) even after the protein has been degraded by the proteasome (39,40). This is in contrast to non-covalent inhibitors that would freely dissociate as soon as the protein is degraded. The benefit is twofold: firstly, the peptide to which the inhibitor is bound (which is indeed the peptide containing the target mutation) is presented at an increased rate as

compared to the mutant peptide without the bound drug (39,40). This was not a trivial finding, as it was unclear whether the bound drug would interfere with peptide processing and presentation by TAP and MHC I (39). Secondly, when these neoantigens are presented extracellularly, a novel chemical moiety (i.e. the inhibitor-peptide-MHC-complex) becomes present on the surface of the cancerous cell (Figure 1B) (39,40). This presents an opportunity for a targeted therapeutic modality.

Antibodies that recognize the inhibitor-peptide-MHC-complex can be developed. These antibodies can then be linked to another antibody (specifically a single-chain variable fragment) that binds to the CD3 region of T-cells (39,40). These molecules induce proximity between the cell of interest and the T-cell, leading to an ‘immune synapse’ and the degradation of the target cell (Figure 1B) (39). These dual antibody compounds are referred to as Bispecific T-cell Engagers (BiTEs) (Figure 1B). The BiTEs thus increase both the recruitment and residence time of T-cells, leading to greater cytotoxicity in the tumor region (39,40). These BiTEs have incredible specificity to the target cells. In order to bind to a cell, the covalent inhibitor must first bind to the intracellular target and then be presented extracellularly by the MHC I for recognition (39,40). This essentially allows us to create extracellular markers on cancer cells using intracellular targets (Figure 1B). These BiTEs show promising data in both cells and in mice (39,40).

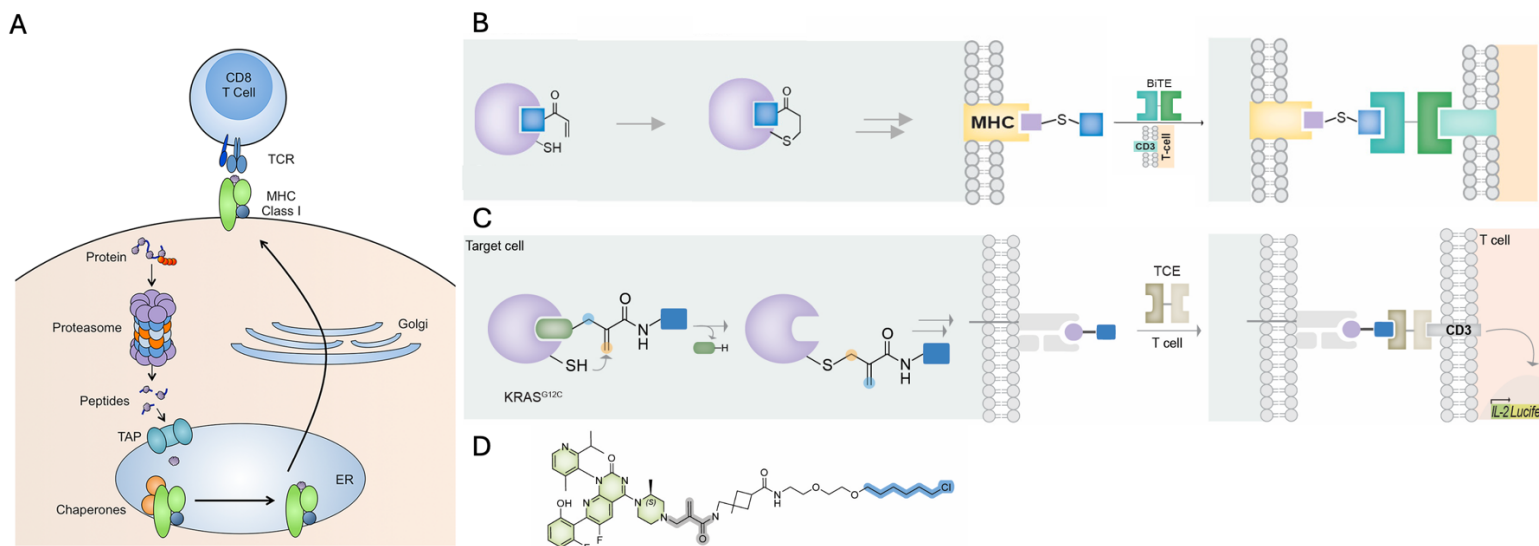


Figure 1: **A)** MHC-mediated peptide presentation pathway from McCarthy et al., 2015 (47). **B)** Original hapten-targeting approach. **C)** HaC platform approach. **D)** Example HaC molecule.

1.3 OUR APPROACH

My thesis, and the aim of the HaCs project at the Choudhary Lab, has built upon this idea, expanding, improving, and scaling the work significantly. The first task was to broadly improve the system as a whole. There are a few inefficiencies of the current platform. Most importantly, this approach is not easily scalable and cannot readily be expanded to different covalent inhibitors as it requires the development of new BiTEs (i.e. new antibodies) for each new inhibitor and target protein. The method is also haplotype-specific, as the BiTE engages with the MHC complex itself which can vary significantly from person to person (48,49). Everyone expresses a certain subset of MHC I alleles (48–50). These are typically written as HLA-(A/B/C)*X:Y where A/B/C refer to the gene locus, X refers to the allele group, and Y to the specific protein within that group (note that there are additional subcategories not mentioned) (49). Over 35,000 MHC alleles have been identified, indicating that hundreds of different BiTEs may be required to treat a large patient population (51,52). Additionally, the BiTE must show high specificity for the inhibitor-peptide-MHC-complex, not just the inhibitor itself. If it has any affinity for the inhibitor on its own, the BiTE will be competed off the complex by the free drug given the high concentration of free drug outside the cell, especially compared to the presence of the inhibitor-peptide-MHC-complex (39). Finally, the affinity of the inhibitor-bound-peptide for the MHC is entirely dependent on the existing structure of the inhibitor and thus cannot easily be enhanced (39).

Our vision was to instead create a universal T-cell engager (UniTE) that would function like a BiTE but be compatible with the neoantigens from all covalent inhibitors and oncoproteins. In order to standardize the BiTE, we needed to modify the various covalent inhibitors to ensure the presentation of the same chemical moiety (i.e. hapten). This led to the development of haptenizing chimeras (HaCs). HaCs repurpose covalent inhibitors as immunotherapeutic agents. Simply put, HaCs are existing covalent inhibitors with an appended hapten of interest (Figure 1C). Specifically, we utilize a methacrylamide functional group that allows for specific group transfer chemistry that was developed in the lab. The thiol on a mutant cysteine, for example, can then perform a Michael 1,4-addition-elimination reaction, allowing the hapten to become covalently linked to a specific residue, while the remaining part of the inhibitor is released (Figure 1C). Thus, when this modified peptide, referred to as a haptenated peptide, is presented to the MHC complex, our reactive group of choice becomes

positioned on the cell surface. A UniTE can then bind specifically to this hapten as well as to a T-cell, inducing T-cell proximity and cancer cell death (Figure 1C).

This platform is readily scalable to a variety of current cancer drugs (Figure 2A,B). By appending the linker to any established covalent inhibitor, almost all of which contain Michael acceptors, the same hapten can be displayed by the MHC on the cell surface (20–30). The same UniTE can then be used across a variety of drugs, making the approach applicable to any cancer that can be treated with a covalent inhibitor, and potentially promising a cost-effective system to improve the efficacy of current cancer treatments and reduce acquired resistance. Once developed, it could quickly be expanded to additional targets with minimal cost, especially compared to the antibody-based approach that others have attempted which requires the development of a new BiTE for each new target (39,40). Finally, the size of the UniTE can potentially be ~50% smaller than a BiTE due to the removal of the neoantigen recognizing antibody in favor of a smaller hapten-binding group. This reduction in size would likely improve tissue penetration which may be especially useful for solid tumors (53). Since the UniTE only engages with the linker and hapten and not the MHC itself (Figure 1C), the platform does not run into any of the current haplotype issues, contributing to its incredible versatility.

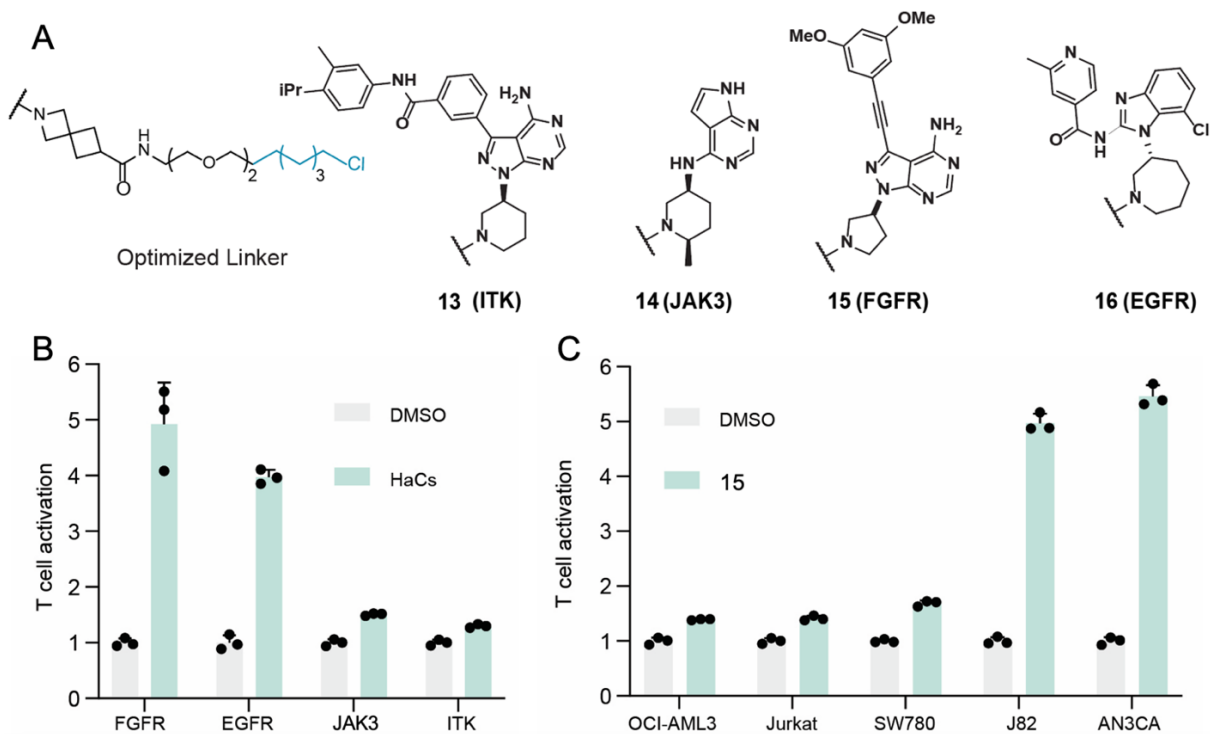


Figure 2: A) Design of HaCs for different protein targets. **B)** Efficacy of HaCs in cells. **C)** Efficacy of FGFR HaC in different cell lines.

1.4 INITIAL WORK

First-generation HaCs attempted to use bio-orthogonal chemistry. For example, a tetrazine group was one of the initial haptens chosen in conjunction with a UniTE containing trans-cyclooctene (TCO) (Figure 3). The goal was to perform a click reaction between the UniTE and hapten. However, the cell permeability of these groups was a major issue. Thus, we switched to the HaloTag hapten, a chloroalkane that binds with high specificity to the HaloTag protein (54,55). While other haptens were also tested, the HaloTag hapten was most commonly used as it exhibited a high degree of efficacy and is stable in most chemical conditions and therefore compatible with many different synthetic schemes (more detail below). Additionally, while some covalent inhibitors target lysine residues, we chose to focus on cysteine-targeting inhibitors to develop the platform. As discussed later, the platform can easily be expanded to include lysine-targeting inhibitors.

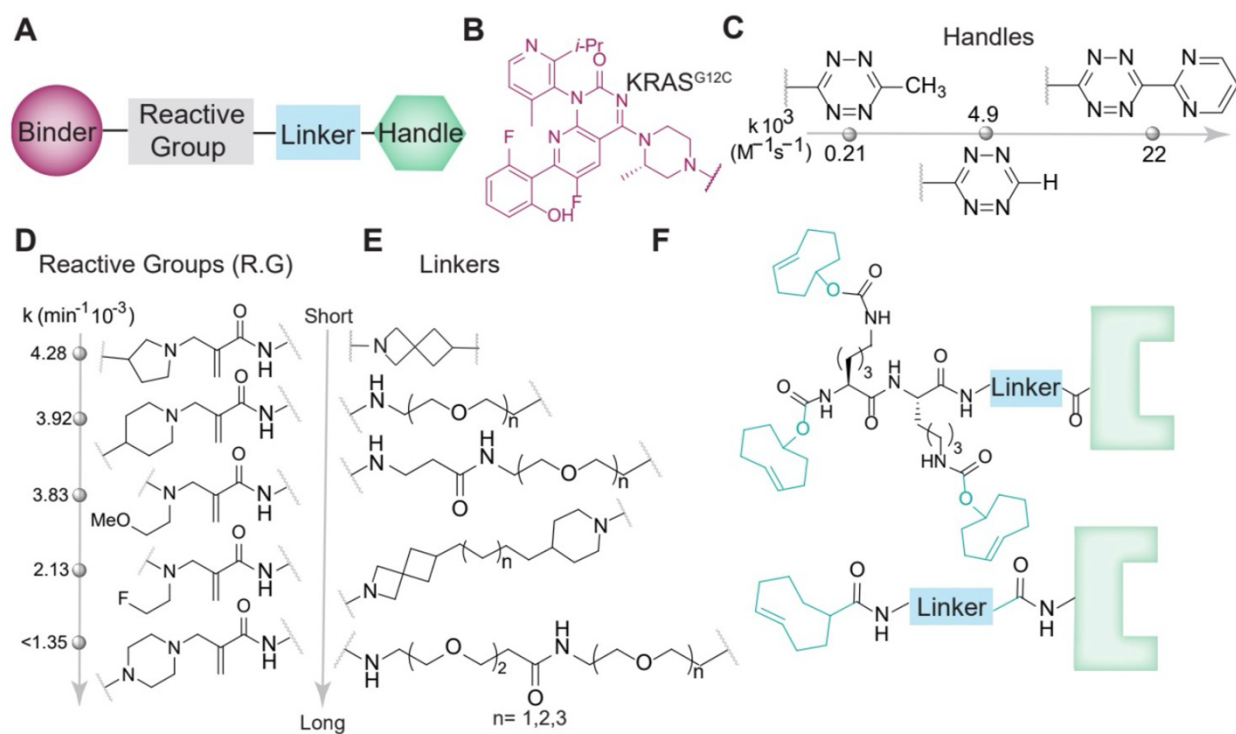


Figure 3: **A)** HaC design scheme. **B)** Example binder. **C)** Click-based tetrazine haptens. **D)** Methacrylamide reactive groups. **E)** Linker designs. **F)** Click-based TCO UniTE structure.

The lab began by creating HaCs targeting the KRAS G12C mutation. This was done by linking Sotorasib with the HaloTag chloroalkane through a methacrylamide (Figure 1D, 4E). Previous cell-

based assays in the lab confirmed the proposed mechanism of action of HaCs (Figure 4). The KRAS HaC was tested in MIA PaCa-2 cells that expressed KRAS^{G12C}, as well as other cell lines that either did not contain KRAS, or did not contain the mutation (either wild type KRAS or a different mutation, such as G12V and G12D) (56) (Figure 4A). The T-cell activation generated by the compounds in each cell line was measured using a luciferase assay and luminescence readout, which is explained in more detail below. As expected, T-cell response was only observed in the MIA PaCa-2 cell line (Figure 4A). In addition, a control compound was synthesized with the same linker and chloroalkane hapten but with the KRAS^{G12C} binder (Sotorasib) replaced with an Ibrutinib scaffold that exhibits no affinity for KRAS (Figure 4B) (29). This control compound did not result in T-cell activation as expected (Figure 4B,F). Finally, we compared T-cell activation from cells bearing the KRAS^{G12C} or KRAS^{G12D} mutations, with the latter lacking the cysteine necessary for group transfer. As expected, we observed little T-cell activation from KRAS^{G12D} cells compared to KRAS^{G12C} cells (Figure 4C). Finally, competition of the KRAS HaC with unmodified Sotorasib led to a decrease in T-cell activation, also as expected (Figure 4D).

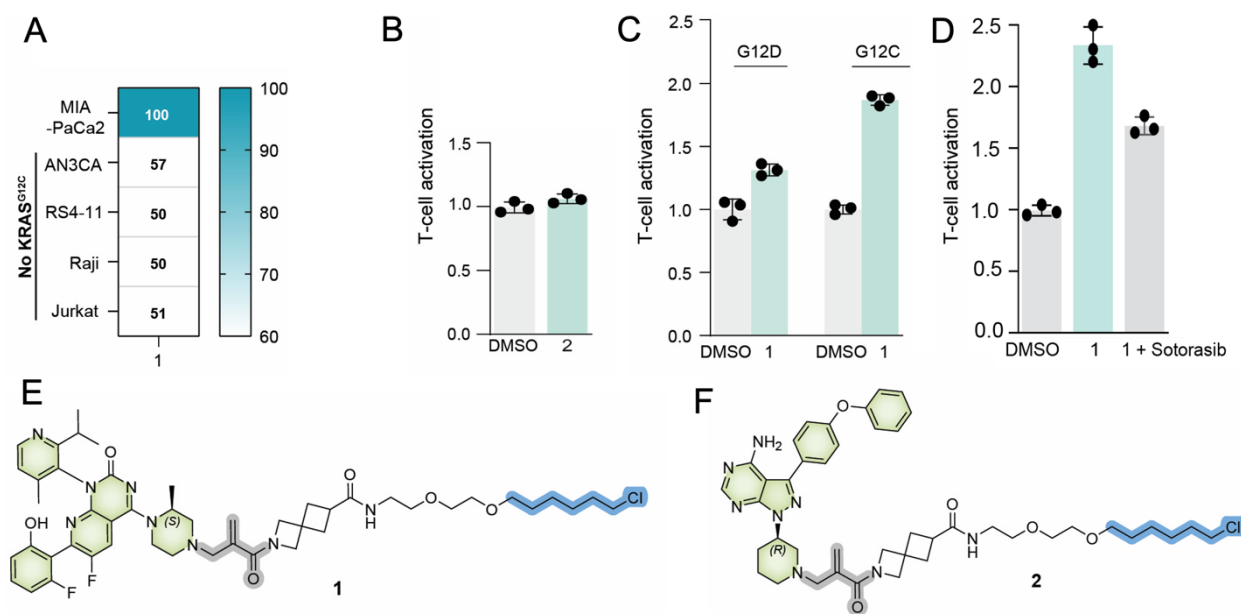


Figure 4: **A)** Evaluation of KRAS^{G12C} HaC in cell lines expressing wild-type KRAS or different KRAS mutants. **B)** T-cell response to Ibrutinib-substituted HaC. **C)** T-cell response in cells expressing KRAS^{G12C} vs. KRAS^{G12D}. **D)** Competitive inhibition of KRAS^{G12C} HaC with Sotorasib. **E)** Structure of KRAS^{G12C} HaC. **F)** Structure of Ibrutinib-substituted HaC compound.

In addition to KRAS, HaCs for a variety of other high-profile targets were synthesized: EGFR^{L858R/790M}, FGFR, JAK3, BTK (not shown), and ITK (Figure 2A) (20,21,30,57). Although the FGFR, JAK3, BTK, and ITK HaCs do not target mutant versions of the proteins, these targets are sufficiently overexpressed in cancer cells such that HaCs remain highly specific (21,29,58). Each of these compounds were tested in a cell line that contained the target protein (and mutation, if necessary). For example, the MIA PaCa-2 cell line was used for KRAS^{G12C}. AN3CA was used for FGFR, Raji for BTK, Jurkat for ITK, and H1975 (double mutant) for EGFR^{L858R/790M} (Figure 2B). These compounds worked to varying degrees, all exhibiting significant increases in T-cell recruitment (Figure 2B).

1.5 AIMS

1. We observed that small modifications to the linker lead to drastic differences in T-cell activation (Figure 5A,B, Figure 7E). There are three components to a HaC – the target binder (e.g. Sotorasib for KRAS), the linker, and the hapten (e.g. HaloTag chloroalkane) (Figure 3A, Figure 5A). The least flexible component is the binder, which is chosen based on previously established and highly specific covalent binders. Although certain targets have multiple inhibitors, this component is usually predetermined and unreceptive to alteration. The most flexible component is the linker, which can be extensively modified in nearly any manner. The hapten is also variable but is still chosen from a group of known compounds and has limitations including bio-orthogonality (crucial to avoid unwanted interactions *in vivo*), cell permeability (a significant issue with groups like tetrazine), and size. Additionally, the hapten will become functionally unmodifiable once a certain UniTE is chosen since the same hapten needs to be displayed by all HaCs for the UniTE to be universally effective. We aimed to greatly improve HaCs by modifying and testing the linkers used, as well as trying different haptens. To that end, we hoped to explore a large number of linkers. Previously, entire HaC molecules would have to be synthesized for each different linker and hapten combination we wished to test (Figure 5). This approach suffered from a lack of scalability. Synthesizing the complete HaC molecules was very time-consuming, costly, and low throughput. The need for a more efficient workflow became apparent, and we aimed to develop a method to greatly improve the speed of testing new linker-hapten combinations.

2. Additionally, we hypothesize that one of the most important components in determining the efficacy of a HaC was the ability of the haptenated peptide to stabilize the MHC complex. Previous studies have shown that peptide affinity greatly influences MHC presentation (59,60). If the haptenated peptide binds with high affinity to the MHC, it will have greater extracellular presentation and thus greater concentration of the hapten on the cell surface, leading to more T-cell activation. We aimed to directly assess and improve peptide-MHC-complex stability as the main driving force for improving the efficacy of HaCs.

3. Finally, we believed that HaCs could have far-reaching therapeutic value, with hundreds of potentially viable targets. Although we successfully designed HaCs for some well-characterized targets with known inhibitors (Figure 2A), identifying targets at scale was a challenge. In addition, we wished to gauge the likelihood that a HaC for a particular target would be successful. The problem of finding the optimal target for a HaC exists in a wide combinatorial space. We hypothesized that there are three key features of a successful HaC: good ligand binding, short distance from binding site to reactive amino acid residue, and strong peptide affinity for the MHC complex. We aimed to create a tool that allows for a thorough and high throughput search for the most promising HaC targets.

2 RESULTS

2.1 T2 CELL ASSAY FOR RAPID LINKER TESTING

To address aims 1 and 2, we developed an assay that allowed us to test different linker-hapten combinations as well as MHC binding ability without the need to synthesize complete HaC molecules: the T2 cell assay for peptides. T2 cells are cells that are deficient in TAP (39,61). This ensures that their MHC I cannot present their own endogenous peptides (Figure 5D) (39). Instead, we can incubate these cells with an exogenous peptide of interest. Specifically, we can treat the cells with a haptenated peptide containing any linker and hapten of interest (Figure 5D). The haptenated peptide is internalized into the cell, where it is loaded onto the MHC I and subsequently presented extracellularly (Figure 5D). We then treat these cells with a UniTE consisting of the hapten binder (e.g. HaloTag protein) on one side, and an anti-CD3 antibody on the other (Figure 5D). Finally, we add mutant Jurkat cells, a type of CD4⁺/CD8⁻ T-helper cell line. When activated, T-helper cells release cytokines (such as interleukins) that trigger an immune signaling cascade. These Jurkat cells were genetically edited to express the luciferase gene downstream of the interleukin-2 (IL-2) promoter (Figure 5D). When the Jurkat cells are activated, the IL-2 promoter is activated, and the amount of luciferase produced corresponds to the level of activation of the T-cell (Figure 5D). The quantity of luciferase can easily be measured using a standard luminescence assay readout (39,62,63). A similar approach has been used by Shokat et al. and was shown to accurately assess the immunogenicity of the haptenated peptide (39). Since the goal of this assay is to test peptides with the hapten pre-attached, we only need to synthesize the linker and hapten rather than the entire structure of the HaC.

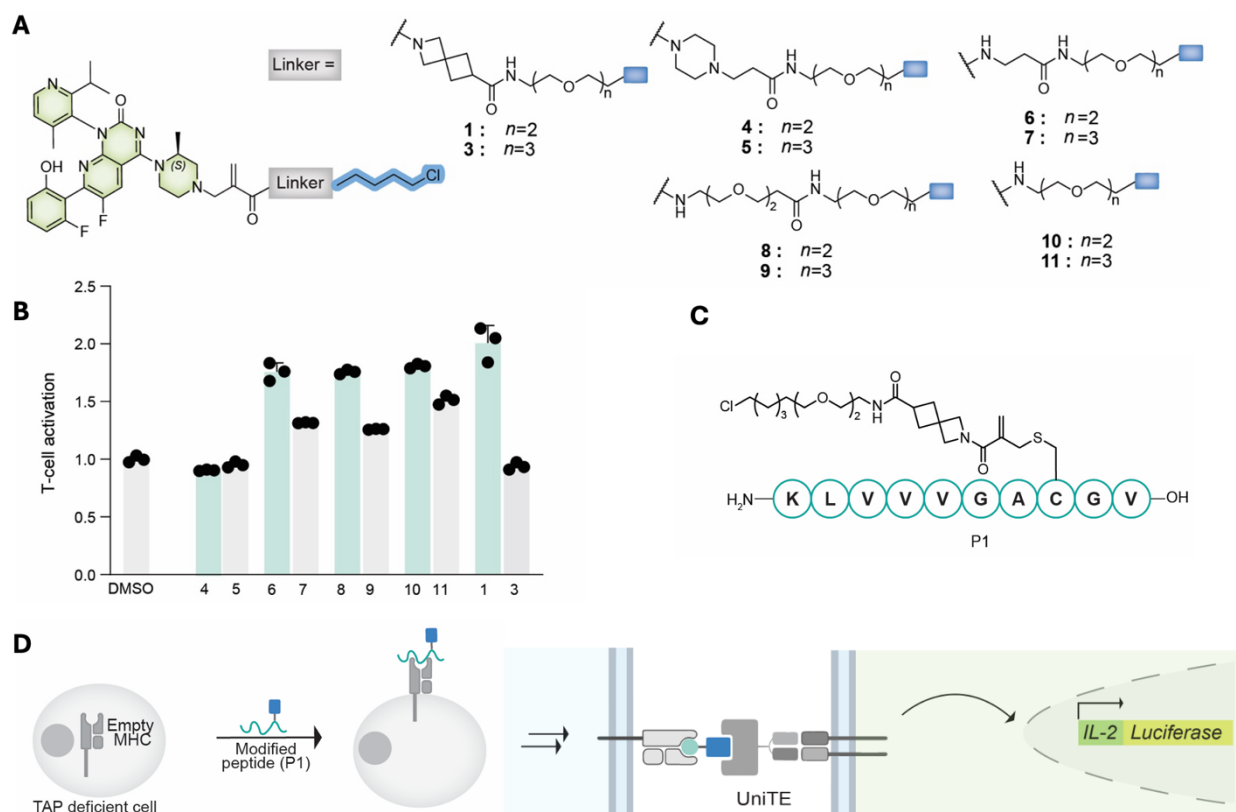


Figure 5: **A)** Various explored linkers. **B)** Efficacy of different linkers. **C)** Example haptenated peptide. **D)** T2 cell assay workflow.

2.2 HAPTENATED PEPTIDE SYNTHETIC SCHEME

In order to utilize this assay in a high throughput manner, we needed to synthesize large amounts of peptides with different linker-hapten combinations. To accomplish this, we first prepared peptides using solid phase peptide synthesis (SPPS) with fluorenylmethyloxycarbonyl (Fmoc) chemistry on 2-chlorotrityl chloride (2-CTC) resin. The peptide was subsequently cleaved from the resin, and any acid-labile protecting groups were lost. Our initial approach was to make a ‘pseudo-HaC,’ and then couple the peptide to the HaC in solution (Figure 6A). The hapten was first coupled to the linker, after which it was coupled to a bromo-methyl acrylic acid to form an amide. The resulting compound was alkylated with morpholine, which would serve as a ‘stand-in’ for the covalent inhibitor during this reaction. This pseudo-HaC intermediate resembled the structure of a HaC (Figure 3A, 6A). The peptide would then be added to this pseudo-HaC and stirred overnight to allow the addition-elimination reaction to take place. This approach resulted in very low yields. The free cysteine of the peptide would often get oxidized, and the reactivity towards the methacrylamide appeared to be low.

Additionally, it was very difficult to ensure regioselectivity of the alkylation, as any nucleophilic side chains (i.e. lysine) as well as the N-terminus needed to remain protected in both the basic and strongly acidic conditions of the synthesis.

To achieve much better yields, scale the synthesis, and create a generalizable protocol, we decided to utilize a solid phase approach (Figure 6B). Here, the peptide is synthesized using SPPS, but not cleaved from the resin. To do this we changed from the more labile 2-CTC resin to Wang resin that is stable in most conditions and cleaved only in high concentrations of acid (64). The N-terminus is subsequently *tert*-butoxycarbonyl (BOC) protected (Figure 6B). We then needed a method to selectively deprotect the thiol side chain of the cysteine while leaving the protecting groups on other side chains intact. The previous peptides were synthesized with trityl-protected cysteine (Fmoc-L-Cys(Trt)-OH). We tried selectively deprotecting the thiol with 20% trifluoroacetic acid (TFA), but this resulted in low yields because a) the trityl was not fully cleaved, and b) BOC and other protecting groups were lost in this process. Instead, we use the monomethoxytrityl (MMT) protecting group which contains a methoxide substituted on the *para* position of one of the phenyl groups of the trityl, further stabilizing the carbocation through resonance and allowing for cleavage in very mildly acidic conditions (65). This allows for selective removal of the MMT group which is orthogonal to other protecting groups, such as BOC, *tert*-butyl, trityl, and more (66). Next, the peptide is alkylated to 2-bromomethylacrylic acid (Figure 6B). The regioselectivity of this reaction was verified to ensure that alkylation was occurring on the thiol of the cysteine and not on an amine (e.g. the N-terminus or lysine residue) that had been mistakenly deprotected. To do so, the intermediates generated from the previous reaction were coupled overnight with excess (10 eq.) of both Hexafluorophosphate Azabenzotriazole Tetramethyl Uronium (HATU) and glycine (Fmoc-L-Gly-OH) (Figure 6C). Here, if any amine had been deprotected and either alkylated or not, it would be coupled to the carboxylic acid of the glycine and the product would be observed (Figure 6C). Using liquid chromatography-mass spectrometry (LC-MS), we observed that the alkylation had occurred on only the thiol (90% as measured by area under the curve (AUC) of the chromatogram), or on both the thiol and an amine (double-alkylated product), but never selectively on an amine (Figure 6C).

At this point, any linker of interest with an amine group (or otherwise) can be coupled to the intermediate to form an amide (using HATU, for example), allowing for a huge variety of linker-hapten combinations to be synthesized and tested (Figure 6B). That is, to test a variety of linkers, we

simply synthesize a large quantity of the intermediate described above. This is easily done since each step is highly efficient (over 90% efficiency) and does not require any purification steps given the solid phase approach – in fact, the entire synthesis is done inside a single syringe (see Materials and Methods). Then, various linkers of interest can be coupled to this intermediate, generating a large library of the same peptide with different linkers. The peptide compounds are finally cleaved off the resin and freed of all remaining protecting groups after two hours of deprotection in a 95:5 solution of TFA:TIPS (tri-isopropyl silane) (Figure 6B). The haptenated peptides are then purified to 95% purity using reverse-phase high-performance liquid chromatography (HPLC). The peptides are lyophilized and then lyophilized again with 0.1% TFA as it was observed that forming a TFA salt greatly increased peptide solubility in DMSO compared to the formic acid salt that was generated from the HPLC. Each of these peptide compounds can then be tested in the T2 cell assay described above, and their efficacies compared. The linkers that lead to the most T-cell activation can subsequently be used to synthesize complete HaC compounds. Note that linkers are often easy to synthesize (Figure 6B, Figure 7D) and large libraries of linkers are commercially available (67).

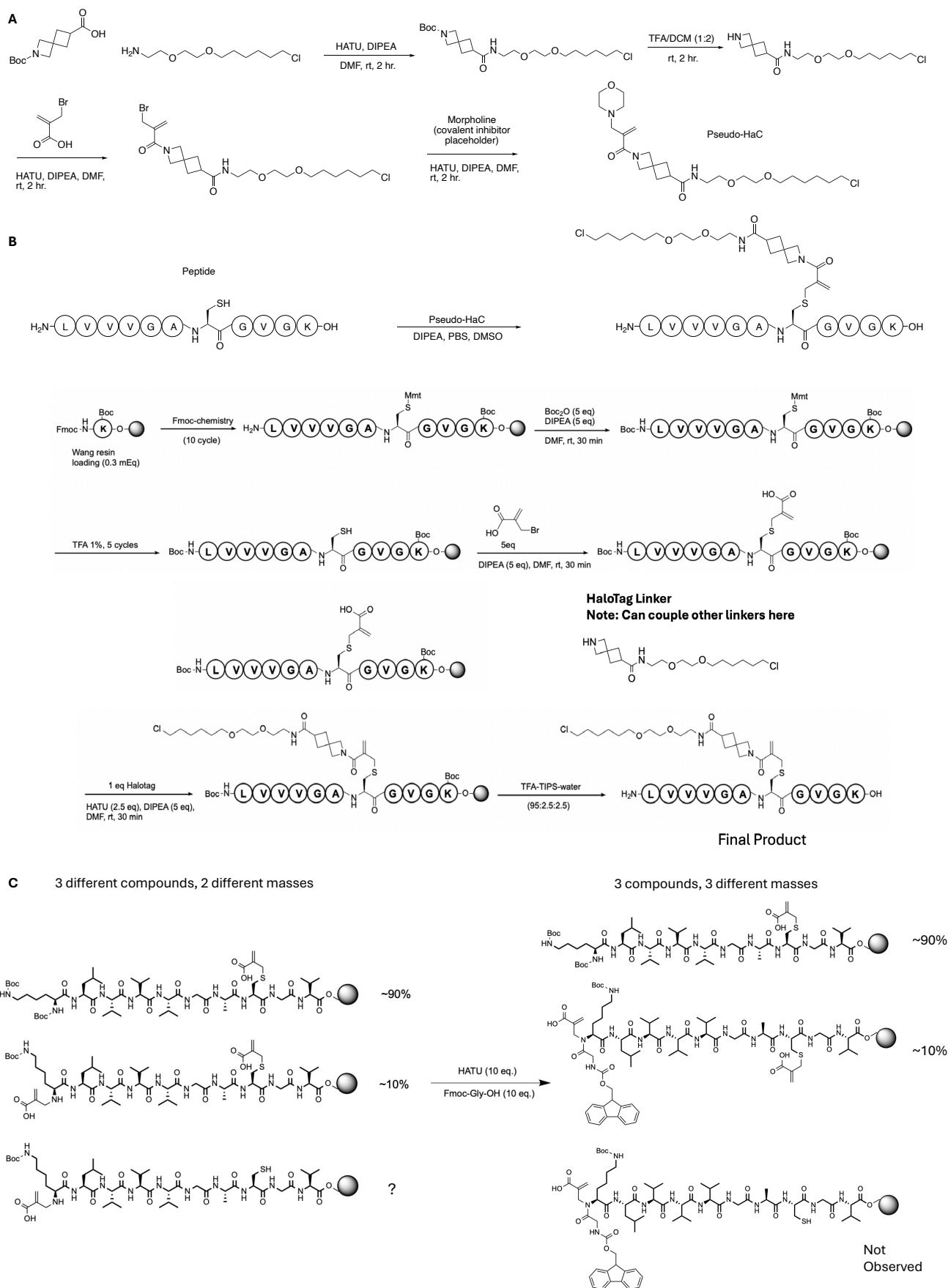


Figure 6: A) Initial synthetic approach. B) Solid-phase workflow. C) Regioselectivity verification.

2.3 ADDITIONAL WORKFLOW FOR ACID-LABILE COMPOUNDS

While this workflow is highly efficient and effective for a large number of linkers and haptens, some very acid-sensitive molecules require special treatment. An example of this is ortho-AP1867, a ligand that binds non-covalently to FKBP12^{F36V} (Figure 8B) (68,69). The ligand exhibits high affinity for FKBP12^{F36V} and low affinity for wild-type FKBP, allowing it to be a good candidate for a bio-orthogonal hapten (68,69). In addition, the UniTE composed of FKBP12^{F36V} (~40kDa) is ~33% smaller than HaloTag's UniTE (~60 kDa), which may allow for better tissue penetration (53). Additionally, HaloTag is of bacterial origin and may be immunogenic, while FKBP12 is of human origin and may be better tolerated *in vivo* (68,69). This FKBP binder had already been validated to function effectively in HaCs (Figure 7E). However, the ligand is quite acid-labile, which we assessed using LC-MS monitoring. The molecule degrades instantly in 95% TFA conditions, the deprotection conditions of the protocol above, and is nearly 80% degraded in 50% TFA conditions, breaking into two pieces (Figure 8A, Supplementary Figure 1). Even the meta-substituted ligand, which is slightly more stable in acidic conditions, was 75% degraded after 30 minutes in 95% TFA, and 50% degraded after 30 minutes in 50% TFA (Figure 8A, Supplementary Figure 1). In addition, the ligand is quite expensive (~\$4000 for 1 gram), which became a crucial constraint. This further emphasized the importance of achieving a high yield for the synthesis. Therefore we developed a separate protocol for this and similarly acid-labile molecules.

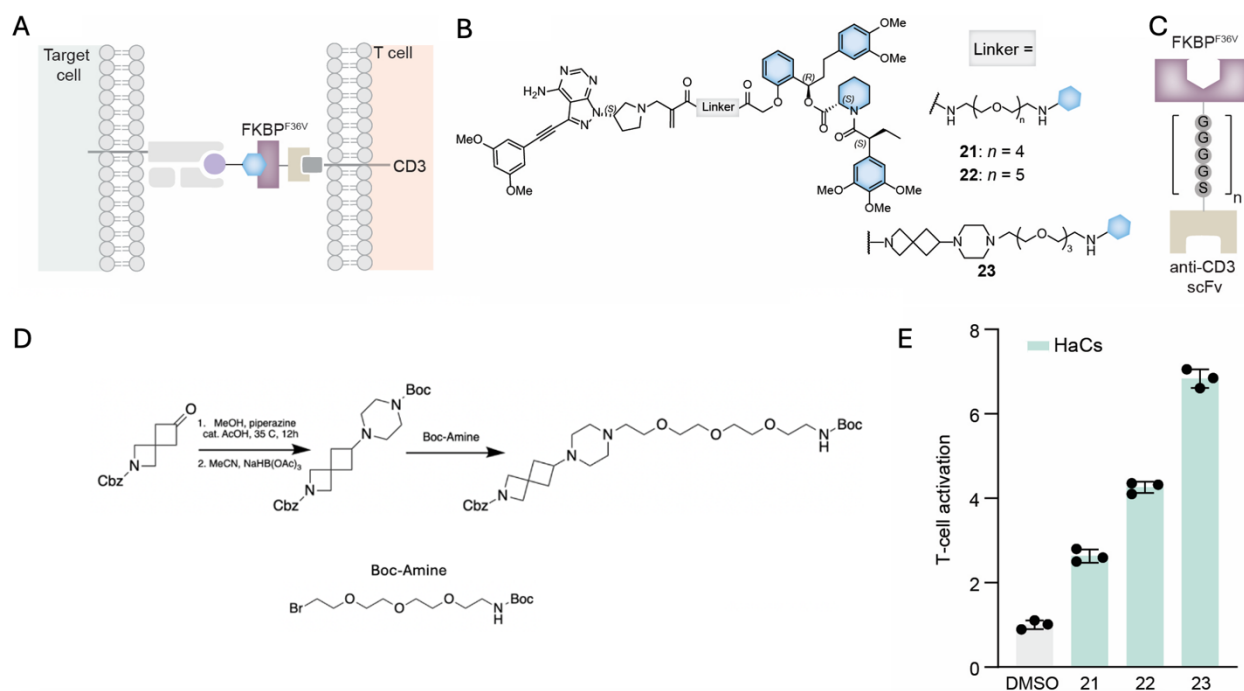


Figure 7: A) FKBP-based UniTE. **B)** FGFR targeting HaC with FKBP ligand. **C)** UniTE structure. **D)** Linker 23 synthesis. **E)** T-cell activation for different linkers.

We wanted to ensure that the new synthetic scheme mostly remained on solid phase as this leads to good yields and keeps the workflow as similar as possible to the previously established conditions. We also wanted the addition of the FKBP binder to be the last step due to the reagent's substantial cost. To do this, we developed a rough protocol consisting of first protecting the amines (N-terminus and lysines) of the peptide with an acid-stable protecting group, then completing the alkylation until the linker step as previously described (Figure 8C). This intermediate would be cleaved off the resin before being coupled to ortho-AP1867 (Figure 8C). Finally, the amine protecting group would be removed (Figure 8C). The protecting group had two key requirements: 1) it must tolerate strongly acidic conditions (95% TFA), and 2) its deprotection conditions must not affect the rest of the compound, including the FKBP12 ligand. After an extensive review of available amino acid protecting groups (70), three candidates were identified: Benzyloxycarbonyl (Cbz), 2-(trimethylsilyl)ethoxycarbonyl (TEOC), and 6-nitroveratryloxycarbonyl (NVOC) (Figure 8D). Cbz is the best established and most commonly used of these groups, but did not work in the synthesis because it was partially cleaved under the strong acidic conditions and also because the hydrogenation required to deprotect Cbz affected the methacrylamide (70). TEOC is deprotected using the fluoride ion (i.e. TBAF), which meets criteria 2, but was also not compatible with the synthesis due to its instability in acidic conditions

of 20% or greater TFA (70). NVOC is stable under highly acidic conditions, and is deprotected using photolysis with $\lambda > 320\text{nm}$ (71,72). Using LC-MS, we verified that the FKBP binder and the methacrylamide were stable under photolysis for up to 6 hours. Additionally, the NVOC group was stable in extremely acidic conditions (95% TFA) for upwards of 3 hours. Thus, using NVOC, we designed a complete synthetic scheme for peptides with acid-labile linkers and haptens: amines are NVOC protected, coupling is performed in solution, and NVOC is cleaved with photolysis as the final step (Figure 8C). This approach was used to synthesize FKBP-based peptide compounds.

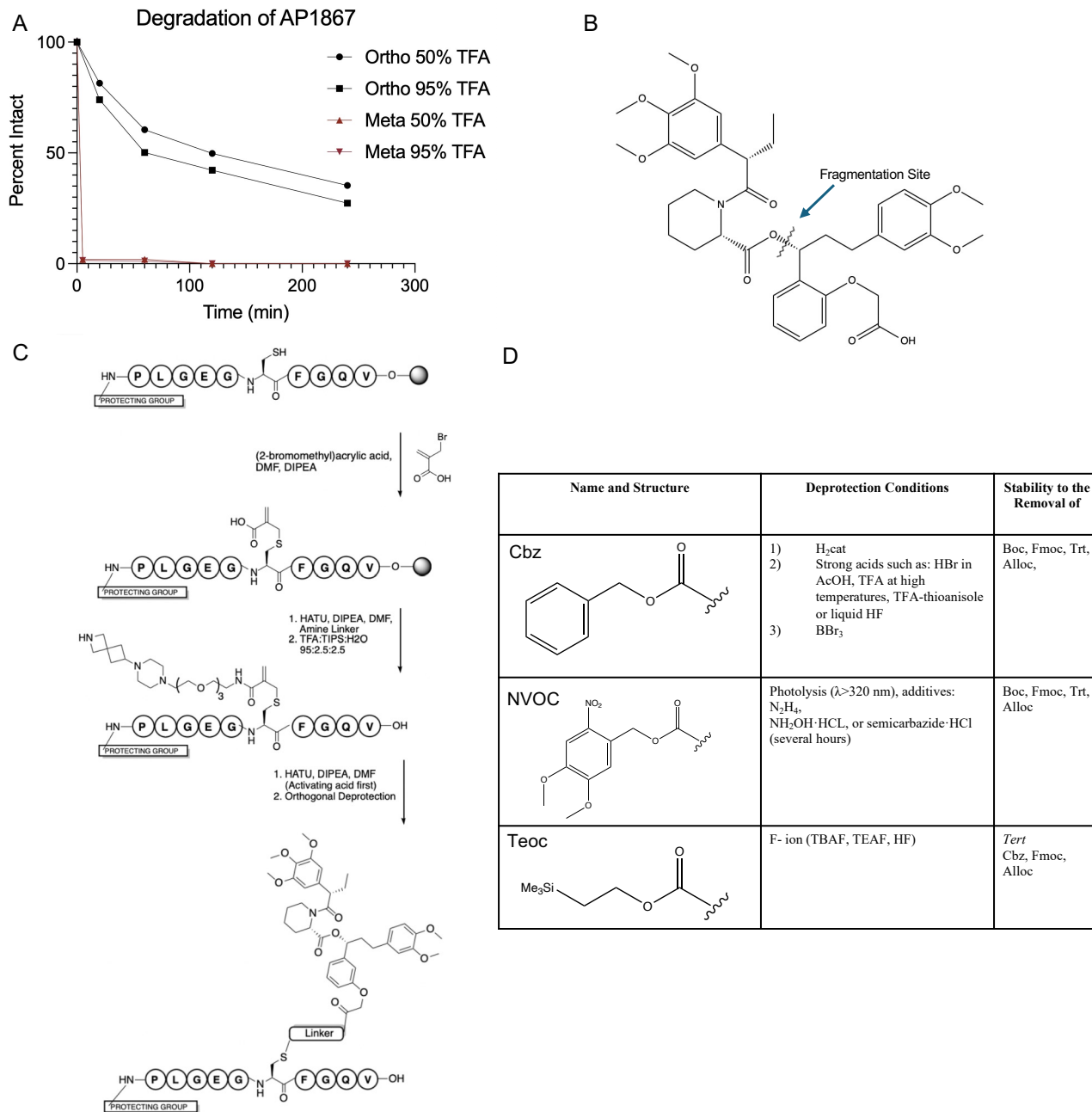


Figure 8: **A)** Degradation of *meta*- and *para*-substituted FKBP over time ligands in highly acidic conditions. **B)** Structure of *Ortho*-AP1867. **C)** Synthetic scheme for acid-labile compounds. **D)** Protecting group options.

2.4 INITIAL VALIDATION OF SYNTHESIS AND T2 ASSAY

To validate the platform for linker comparison, we aimed to compare the difference between a HaloTag-chloroalkane-based spiroamine linker and an FKBP-based polyethylene glycol (PEG) linker.

Previous HaCs designed with these linker-hapten combinations found that the two had roughly the same efficacy (Figure 7E, Figure 2B). The HaloTag linker-hapten was roughly 20% more effective than the FKBP PEG-5 linker-hapten (5-fold vs. 4-fold increase compared to DMSO control) in HaCs targeting FGFR (Figure 7E, Figure 2B). Ultra-low-affinity FGFR peptides (see Section 2.6) were synthesized for these linker-hapten combinations (Figure 9) and tested in a T2 cell assay in triplicate. There was generally minimal variance among data points within each triplicate not just for this experiment but for all T2 assays conducted, indicating high reproducibility and robustness (Figure 9, 10). Cell viability was monitored after incubation with the compound and the number of live cells were standardized across conditions. Higher concentrations of compound typically led to more cell death, with cell viability ranging from as low as 60% for the highest doses up to 80% for DMSO-only conditions. The low affinity of the peptides reduced baseline MHC presentation and T-cell activation and thus background noise, allowing changes introduced by the linker and hapten to be more clearly observable. T-cell activation was compared to both a DMSO control and peptide controls (non-alkylated versions of the peptides), and fold-change was assessed compared to the DMSO control (Figure 9). The FKBP peptides generated a 1.5-fold increase in T-cell activation while the HaloTag peptides generated a ~2.2-fold increase, resulting in nearly 50% better performance for the HaloTag linker-hapten combination (Figure 9). Ultra-low-affinity KRAS peptides were similarly tested (Figure 10A). The same peptide with the FKBP linker-hapten led to a 1.49-fold increase in T-cell activation compared to a 1.55-fold increase for the HaloTag peptide (Figure 10A). Finally, a high-affinity KRAS peptide was similarly tested. In this case, a dose-response assessment was conducted, and the FKBP peptide outperformed the HaloTag peptide at all concentrations (Figure 10B). At 100 μ M, the FKBP peptide generated a 3.23-fold increase while the HaloTag peptide only generated a 2.63-fold increase (Figure 10B). Since this is the exact haptenated peptide reported in the literature to be presented by MHC I (39), these results suggest that an FKBP-based KRAS HaC may be more effective than the HaloTag-based KRAS HaC (despite HaloTag generally exhibiting higher efficacy), potentially due to specific interactions between this particular peptide sequence, ortho-AP1867, and the MHC. This FKBP KRAS HaC is currently being synthesized for testing. Note that additional linker comparison experiments (especially with FKBP) were planned but had to be canceled due to the high cost of the reagents needed (see Acknowledgments).

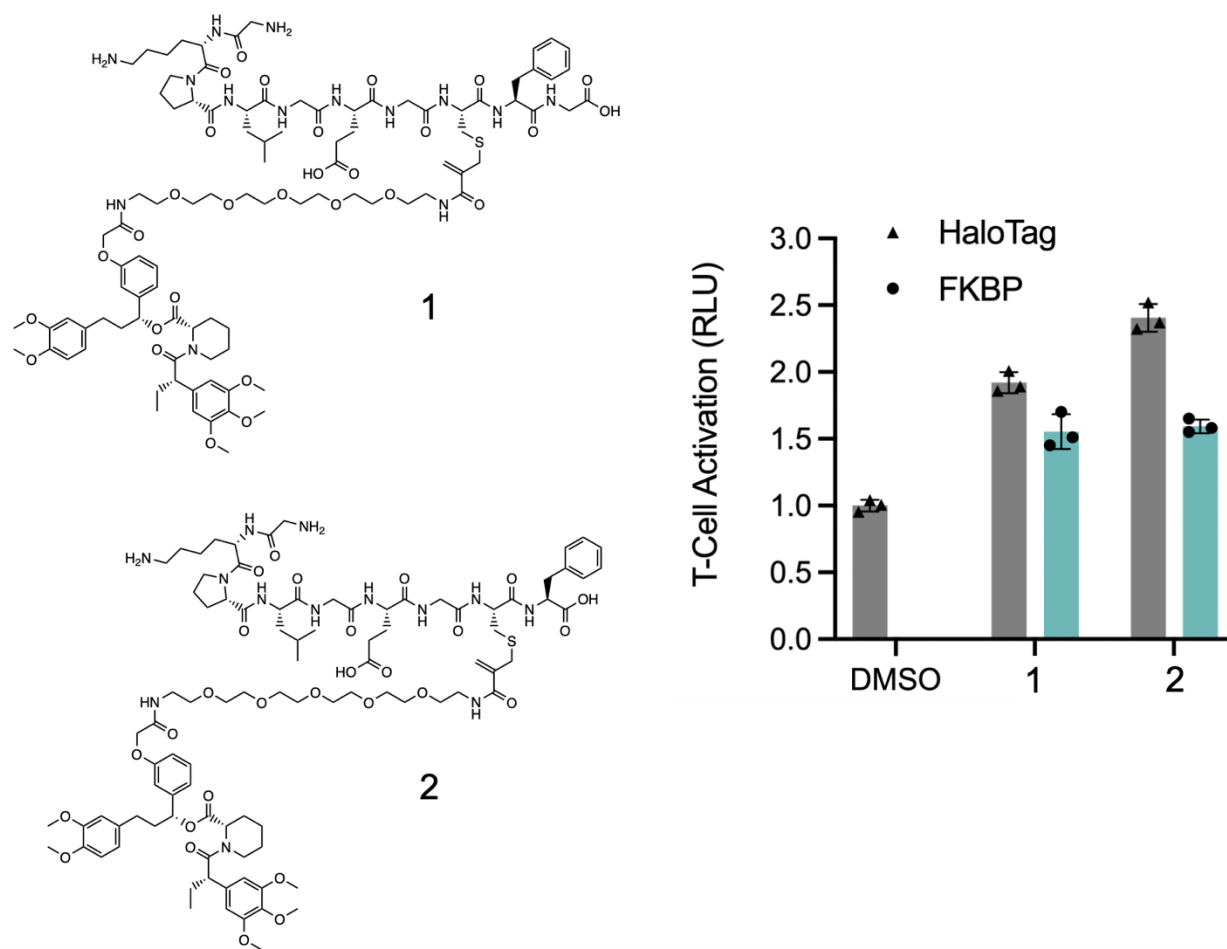


Figure 9: Low-affinity FGFR peptides and respective T-cell activation

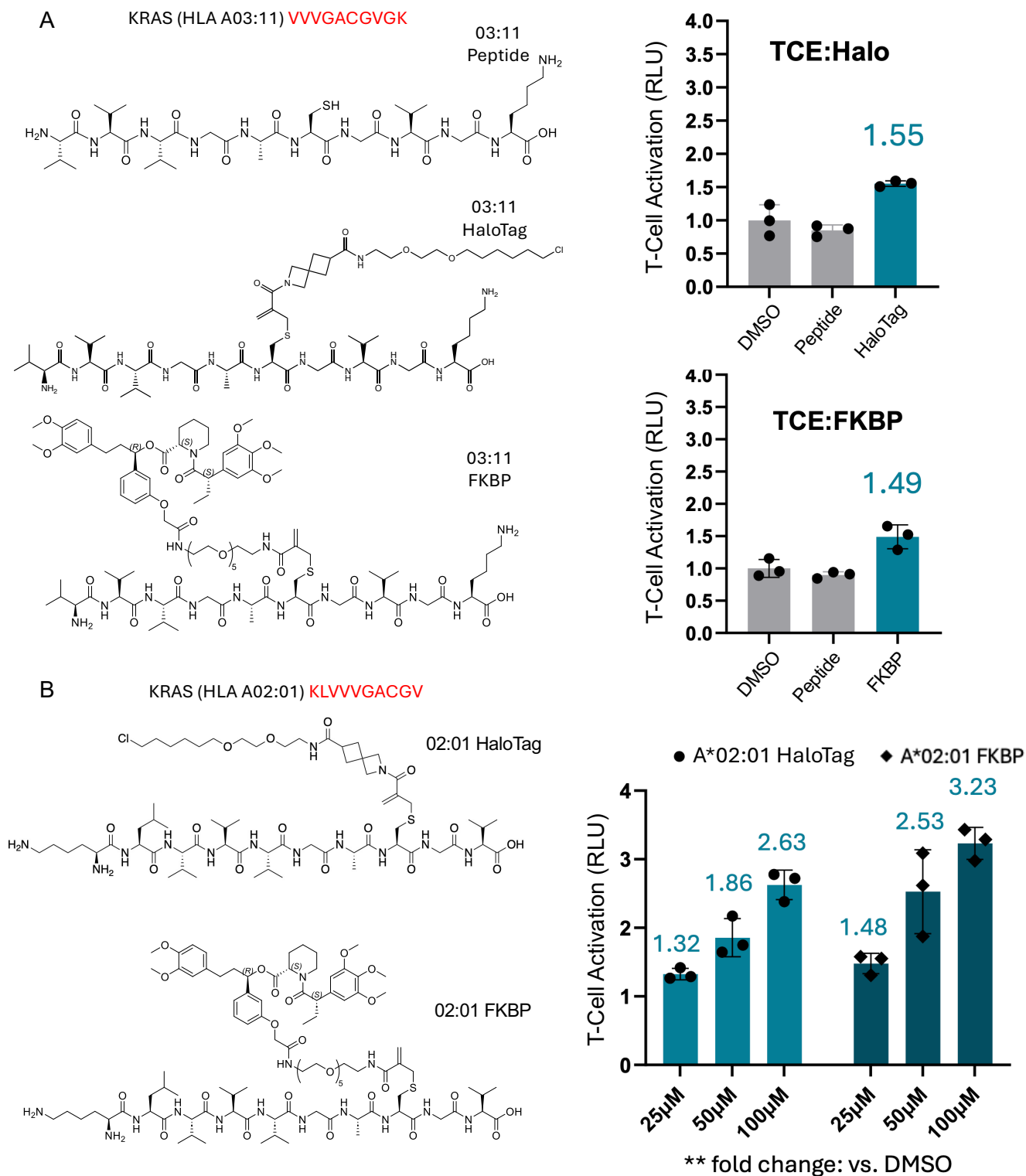


Figure 10: A) Low affinity peptides and respective T-cell activation at 100 μ M. **B)** High-affinity peptides and dose-based T-cell activation.

2.5 HAPLOTYPE DEPENDENCY IN T2 CELLS

One potential shortcoming of the T2 assay is its lack of haplotype diversity. Different alleles have drastically different affinities for the same peptide (Figure 11), which is critical to be aware of when interpreting our data (48,49). T2 cells express the HLA-A*02:01 allele and thus T-cell activation is proportional to the affinity of a peptide for the A*02:01 allele (61). Subsequently, peptides must be optimized for this allele. Thus, when comparing the T-cell activation between different oncoprotein targets, differences in peptide affinities must be accounted for. However, there are advantages to this as well. Any cell line (or animal model) will contain some ‘random’ subset of alleles, and since different HaCs must be tested in different cell lines, there is no way to efficiently or effectively standardize for haplotype. These cell lines and animal models will also have inherent differences in the rate of MHC presentation, leading to additional confounding variables (73). Along with variance in oncoprotein expression levels, variance in MHC expression helps explain why different cell lines exhibit different responses to the same HaC compound (Figure 2C). Differences in target protein expression levels, haplotype, and rate of MHC presentation make it difficult to compare the efficacy of HaCs for different targets since this involves comparison across cell lines, which may be flawed. The standardization of the T2 cell line creates a better metric for a more direct comparison between these HaCs. Additionally, since linkers are tested on identical peptide sequences, baseline peptide affinity does not account for differences in T-cell activation between different linkers, allowing us to effectively isolate the linker-mediated portion of the response. In addition, HLA-A*02:01 is one of the most commonly found MHC alleles in humans (52,74), and computational predictions on the allele are fairly accurate given a large dataset of binders for this allele (see Section 2.6) (75), making it a very reasonable choice as the standard allele to use.

To test the haplotype dependency, we compared the T-cell activation of two KRAS peptides reported in the literature to be strongly displayed, one on HLA-A*03:11 (not expressed in T2 cells) and another on HLA-A*02:01 (expressed in T2 cells) (39). As predicted, T-cell activation for the A*02:01 peptide was twice that of the A*03:11 peptide across multiple replicates (Figure 11). Compared to the DMSO control, the A*03:11 peptide generated 1.1-1.6 times the T-cell activation, while the A*02:01 peptide generated 2-3 times as much (Figure 11), indicating that even very-low affinity peptides can be presented, generating moderate T-cell activation. Control peptides (non-alkylated versions of each

peptide) exhibited no more T-cell activation than the DMSO control. This underscores the large difference in peptide affinity between alleles.

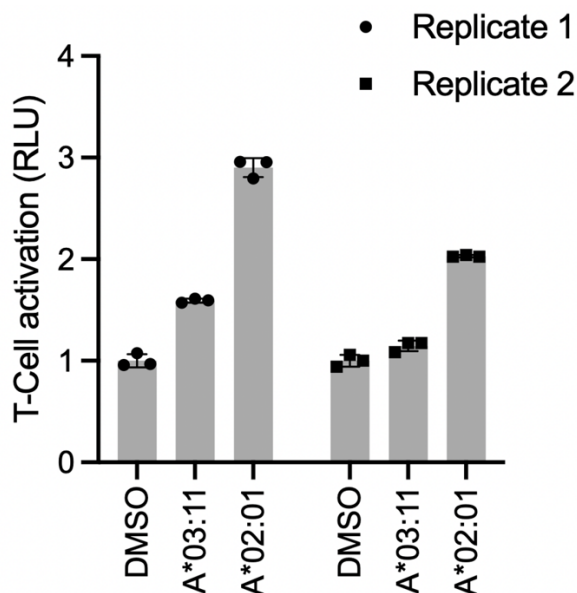


Figure 11: Haplotype dependence of T2 assay.

2.6 COMPUTATIONAL PREDICTIONS

To further scale the project, we moved to computational tools. We hypothesized that MHC display algorithms in the Immune Epitope Database (IEDB) do not predict the display of peptides containing the targeted cysteines that are haptenated by HaCs, especially for diverse MHC haplotypes. To test the hypothesis, we used IEDB tools, and chose the NetMHCpan prediction network (75,76). NetMHCpan is an artificial neural network for the prediction of peptide affinities for any group of MHC alleles given only the primary structure of the peptide (75,77,78). All MHC alleles are available, and predictions can be generated for various animal models in addition to humans (75). NetMHC has demonstrated high levels of accuracy (> 90%), and is the most accurate MHC prediction software to-date (75,79). It consistently performs the best on weekly benchmarks conducted by the IEDB (76), and is widely established in the literature (79). It is also one of the few prediction tools that produces quantitative IC50 binding value predictions. It has been trained on data from over 850,000 peptide assays, including both elution-based assays and competitive inhibition assays (75). We first implemented the tool to generate predictions for peptides containing the cysteines targeted by our

HaCs. We were specifically interested in the predicted affinities of these peptides since they have the potential to be haptened and presented by the MHC. As discussed, the affinity of this peptide to MHC plays a key role in the efficacy of the HaC. For each target, every possible 8, 9, 10, and 11mer peptide that could be generated from the protein sequence was considered. These lengths were chosen as MHC Class I ligands are almost always of this size (80,81). Outliers to this size rule exist, typically in the form of larger peptides of length up to 14, but are rare (81). This means that for a sequence AABBCCDXDEEFFGG, there would be eight 8mer peptides: AABBCCDX, ABBCCDXD, BBCCDXDE, ..., XDEEFFGG. In our case, we were interested in all target-cysteine containing peptides, i.e. peptides with residue **X** being the target cysteine **C**. The affinity of each of these peptides for a chosen group of alleles would then be predicted, meaning each peptide receives a predicted affinity for each of the alleles.

Subsequently, we generated predictions for all possible peptides from each of the protein targets (KRAS, FGFR, BTK, and ITK). Alleles chosen were those present in the respective cell lines for each target. For example, the KRAS HaCs were tested in the MIA-PaCa-2 cell line which expresses HLA-A*24:02, HLA-B*14:02, HLA-B*46:01, and HLA-C*03:02, so affinity of KRAS peptides were predicted for these alleles. This resulted in hundreds of thousands of predicted affinities (Figure 12A). In general, a lower IC50 value corresponds to a better binder, with IC50 less than 50nM indicating a very strong binder. Peptides with IC50 less than 500nM are predicted to be good binders presented by the MHC, a benchmarked value for 80-90% of immunogenic peptides (82–86). In addition, there is a strict upper bound of 5µM (5000nM), above which there exists no known T-cell epitope (82,83,85). Thus, any peptide above this threshold is considered a non-binder. We observed that most of the target-cysteine-containing peptides for all targets fell comfortably outside of this range, with almost all target-cysteine-containing peptides predicted to be non-binders (Figure 12A). This data suggests that covalent modification of the cysteines greatly enhances MHC affinity, potentially through hydrophobic interactions (see Section 2.12) or other mechanisms, and may also change the type of peptides being displayed.

Additionally, most prediction tools seem to systematically underestimate the prevalence of cysteine-containing peptides (87). This may be due to the ability of cysteine to become oxidized and a lack of training data with cysteine-containing peptides (85). For example, for a small set of 71 test peptides specifically for the HLA*A02:01 allele, Sachs et al. found NetMHC to be 90% accurate for the entire

data set, but only 71% accurate for the cysteine-containing peptides (87). However, this study looked at an exceedingly small data set (only 14 of the peptides contained cysteines), and other research has shown that the bias against cysteine-containing peptides in NetMHC is less impactful than previously thought (88). Regardless, even if NetMHC is underestimating the affinity of cysteine-containing peptides, the impact on our work is not substantial and may even be beneficial (see Section 2.9).

2.7 FINAL VALIDATION

To validate the T2 assay and computational predictions, we evaluated four peptides from three different targets (EGFR, KRAS, BTK), each with various affinities for HLA-A*02:01 ranging from 23nM to 24.3μM. We found that T-cell recruitment exactly correlated with each peptide's predicted affinity for A*02:01 (Figure 12B). The EGFR peptide (sequence: QLMPFGCLL) and one of the KRAS peptides (sequence: KLVVVGAGCGV) have been reported in the literature to be strong binders (39). This aligned with our predicted binding values, which estimated IC50 values of 23.15nM and 166.86nM respectively. As anticipated, these two peptides generated the most and second most T-cell activation respectively (Figure 12B). These two known binders both had an affinity less than the 500nM cutoff as expected, and the peptide with the most affinity indeed showed the most T-cell activation. Additionally, the next highest affinity peptide was the BTK peptide with predicted IC50 of 2339.73nM - above the 500nM cutoff but below the 5μM strict upper bound. This peptide generated the next most T-cell activity (Figure 12B). The last peptide was a KRAS peptide with strong affinity for HLA-A*03:11 but terrible predicted affinity of 24.3μM for A*02:01 (well above the 5μM strict upper bound). As expected, this peptide resulted in the least T-cell activation. Also as expected, the peptide was still able to generate a ~1.5-fold increase in T-cell activation, which seems to be the baseline value for HaC peptides across all assays (Figure 12B). This once again indicates that haptenization appears to increase peptide affinity for and presentation by the MHC. Note that T2 cell results generally agree with the results from complete HaC molecule testing, although this is not always the case (Figure 12B, 2B, 4C). As explained previously, differences are likely due to oncoprotein expression levels and haplotype discrepancies, as well as varying rates of MHC presentation by different cell types. For the vast majority of oncoproteins, there is no literature on which peptides containing target cysteines are presented by the MHC. Thus, NetMHC can be leveraged as a powerful

tool to analyze any oncoprotein and determine the most (and least) immunogenic target-containing peptides.

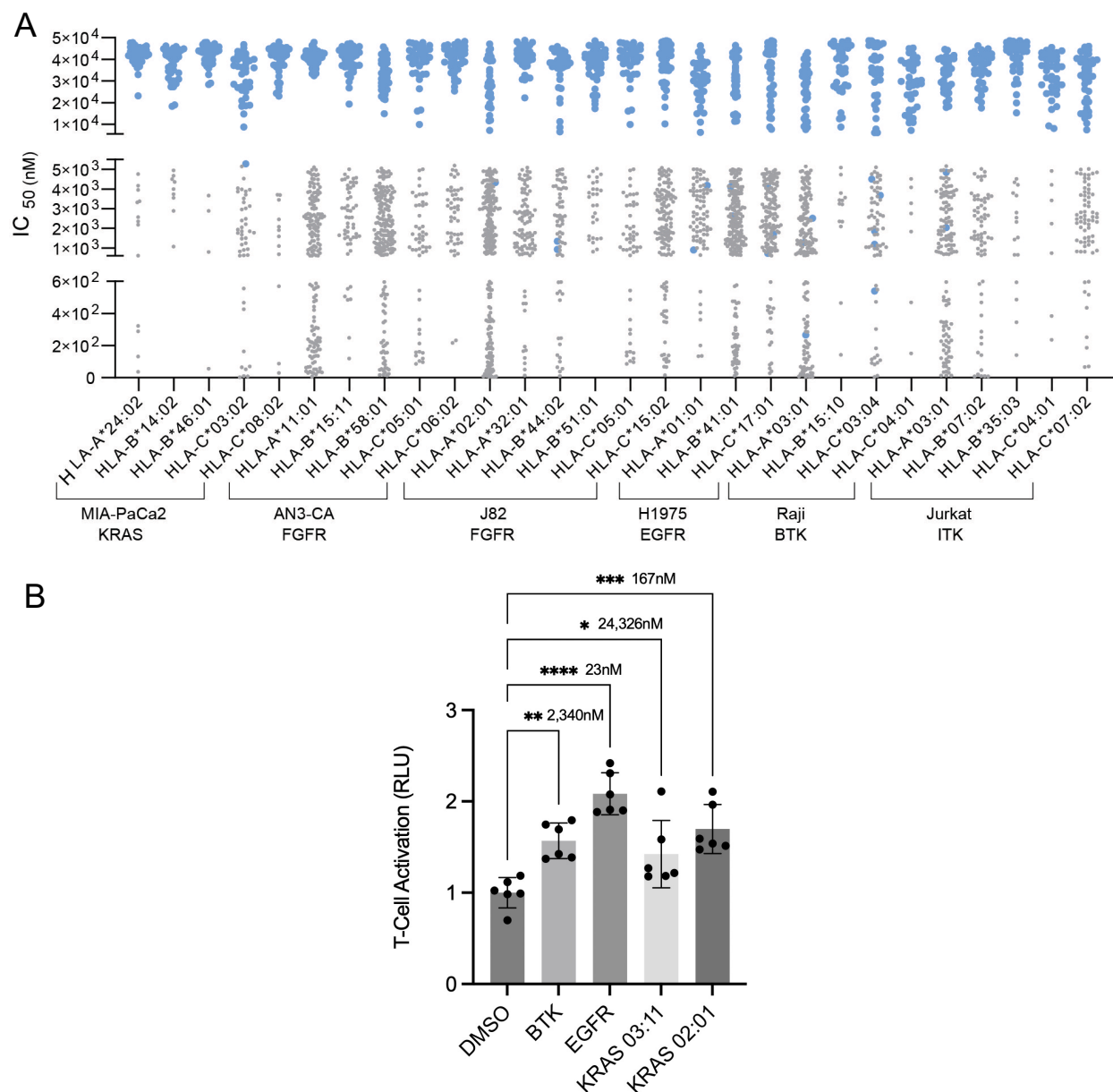


Figure 12: A) High-throughput screening of target-cysteine-containing peptides (blue) vs. all others (gray). **B)** T2 cell assay to validate computational predictions.

To demonstrate the limitations of the computational tool, we turned to the FGFR protein, for which no specific haptenated peptides (or target-cysteine-containing peptides in general) have been reported in the literature. Using NetMHC, we generated an exhaustive list of all possible target-cysteine-

containing peptides (42 possible peptides) (Supplementary Table 1). These peptides had affinities ranging from 4.3 μ M to over 47 μ M. This indicates all FGFR cysteine-containing-peptide are likely non-binders for HLA-A*02:01 (given the 500nM cutoff), with only one peptide having any chance of being a low affinity binder (given the 5 μ M strict cutoff). Six of these peptides were chosen across the range of predicted binding affinities: two with the highest (4.3 μ M, 7.1 μ M), two in the middle (23.1 μ M, 27.1 μ M), and two with the lowest affinities (42.6 μ M, 44.8 μ M) (Figure 13B). Given the poor affinity predictions, we expected that none of these peptides would bind effectively, and that the difference between the best and the worst of these peptides would be marginal. Since they are all expected to be non-binders, the relative ranking of these peptides is likely inconsequential. All the peptides were alkylated with the standard HaloTag spiroamine linker and hapten using the same synthetic approach described previously. As expected, T-cell activation did not exhibit any clear pattern corresponding to the predicted affinities of the peptides (Figure 13A). Peptide compounds generated 1.4-2.4 times the T-cell activation when cells were treated with 100 μ M of the compounds, and 1.2-2 times T-cell activation when treated with 50 μ M concentrations (Figure 13A). This validates the established literature that 500nM should be used as a cutoff for binders, while lower predicted affinities, especially 5 μ M or higher, are relatively arbitrary (82,83,85). This experiment helped determine the strengths and limitations of NetMHC.

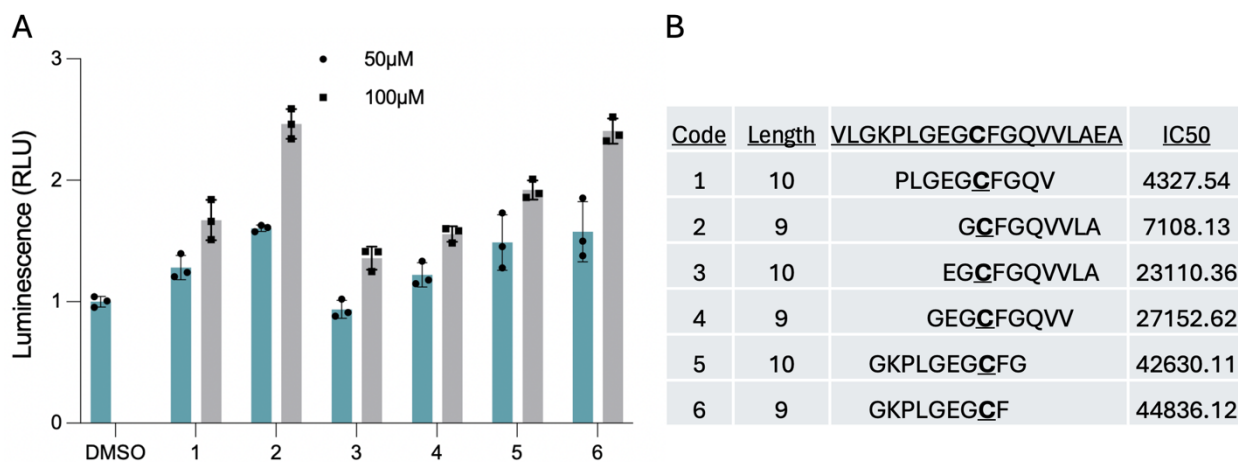


Figure 13: A) T2 results of extremely low-affinity peptides. **B)** FGFR peptides tested.

2.8 LYSINE COMPATIBILITY

While all examples discussed above focused on haptenating cysteine residues, this platform can and has been used with lysines as well (Figure 14). The amine side chain of a lysine residue provides a nucleophile (much like the thiol side chain of the cysteine) which can perform an addition-elimination reaction using an N-acyl-N-alkyl-sulfonamide activated with a nearby electron withdrawing group (Figure 14A), leading to a haptenated protein. All steps of the pipeline can easily be applied to lysine residues. For example, simply using Fmoc-Lys(MMT)-OH allows for the synthesis of lysine-haptenated-peptides while keeping all other steps of the solid-phase protocol the same. Similarly, computational predictions can readily be made for lysine-containing peptides and may be more accurate than those for cysteine-containing peptides (87). This expands the scope of possibilities for HaCs even further.

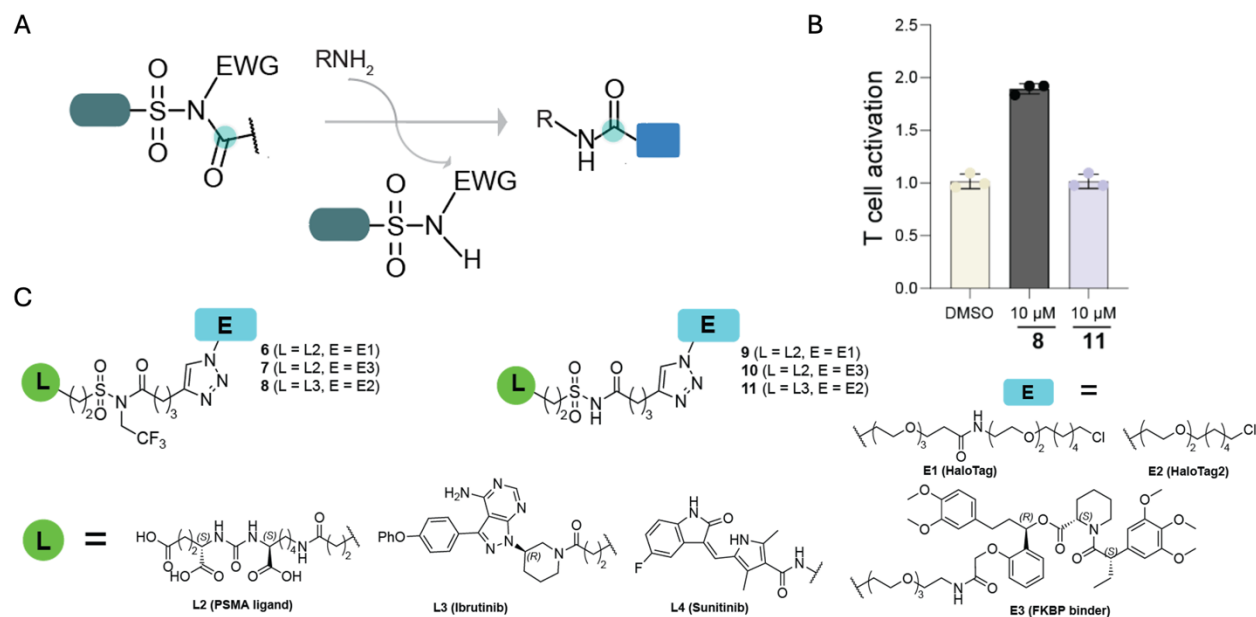


Figure 14: A) Lysine-based group transfer chemistry. **B)** Efficacy of lysine-targeting HaCs in cells. **C)** Structure of lysine-targeting HaCs.

In order for a molecule to serve as a HaC, it must bind to a target with high specificity, and it must be able to covalently attach to this target through a nucleophilic side chain (i.e. cysteine or lysine) that attacks an electrophile in the molecule (i.e. methacrylamide or sulfonamide group). Covalent inhibitors were easiest to repurpose given that they already bind covalently and many of them already contain an acrylamide functional group (or a similar electrophile) that directly interacts with the target cysteine

or lysine side chain. However, the HaC modality is not limited to covalent inhibitors — it can easily be applied to non-covalent modalities. If a non-covalent ligand of any kind (inhibitor or otherwise) induces proximity between the molecule and a nearby nucleophilic residue, it can be repurposed as a HaC. For example, if there are cysteine or lysine residues in the ligand binding pocket, these proximal cysteines and lysines can serve as the nucleophile to react with a HaC made from this non-covalent modality.

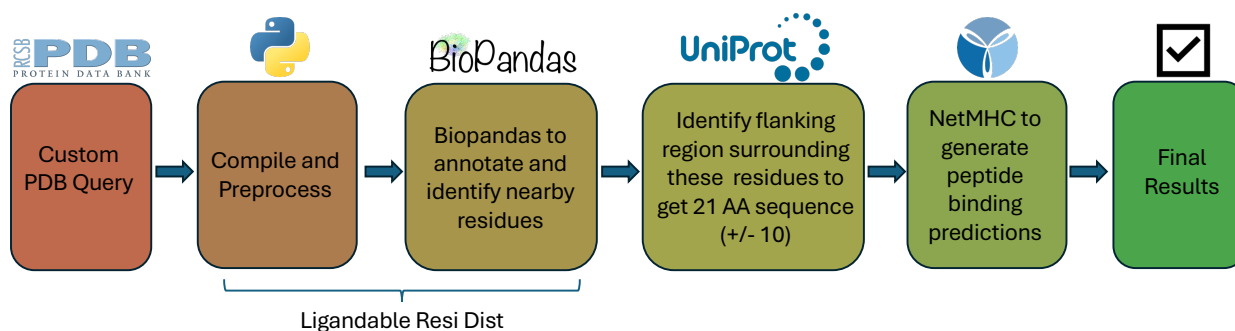
2.9 HIGH THROUGHPUT IDENTIFICATION OF HACs

In conjunction with this observation, we theorized that we could complete the pipeline using an innovative approach: a comprehensive and systematic search for all potentially viable targets for successful and effective HaCs (Figure 15A). To do this, we first performed a search of the entire Protein Data Bank (PDB) for all known, high-resolution protein structures containing ligands (Figure 15A). This generated a list of over 8,600 structures with their ligands. Note that any custom query can be used to restrict the types of inhibitors, targets, etc. as desired. We aimed to perform the most comprehensive search possible. Subsequently, we used BioPandas to iteratively search through all of these structures for “ligandable residues,” that is, cysteine and lysine residues present in the binding pocket 15Å or less from the ligand, and to annotate the distance from the ligand to the residue. This generated a list of tens of thousands of “ligandable residues.” For cysteine alone, over 4,000 residues in various enzymes were identified as potential targets, including over 1,500 residues in potentially oncogenic kinases.

Next, the primary sequence of each of these peptides was extracted to get the surrounding (flanking) region of each of these ligandable residues. The 10 amino acids before and after the target residue were considered, generating sequences of length 21 (e.g. XXXXXXXXXXXX**C**XXXXXXXXXXXX where **C** can also be **K**) (Figure 15A). For each of these 21 residue sequences, all 8, 9, 10, and 11-mers that contained the target residue were considered (38 peptides in total). This generated over 1.8 million potentially ‘haptizable’ peptides. The affinity of these peptides was then tested against the 27 most common alleles found in humans (52,74,76), leading to over 49.2 million predictions for cysteine residues alone. To maximize compatibility with the T2 cell assay, we focused on the A*02:01 allele for which there were 1.8 million predictions. The complete list resulting from this search is currently being

kept proprietary, but a snapshot is shown (Supplementary Figure 2). This list contains key information: the target from which the peptide originates, the ligand for the target, the affinity of the ligand for the target, the distance from the ligand to the target residue, the affinity of the peptide for the given MHC allele, and general protein and ligand information (molecular weight, structure, etc.) (Supplementary Figure 2). We predict that the best targets and ligands to be repurposed as HaCs will include those with the following three criteria: the highest peptide affinities, the smallest distances from the ligand to the target residues, and the highest affinity ligands for the target. Some targets have multiple high affinity peptides and multiple proximal ligandable residues, offering even more promise. Some other practical considerations also include the viability of repurposing the ligand structure and what impact it might have on affinity for the target. This massive and exhaustive list is currently being used in the lab to explore a huge new therapeutic area that can lend itself to developing HaCs for all kinds of targets. One of these targets currently being explored is the p53 transcription factor, which has extremely high therapeutic value (89). Rezatapopt is a ligand targeting the Y220C mutation in p53, but we have shown that it retains some affinity for the wild-type structure (90). The computational search revealed that p53 has a separate, conserved cysteine (C229) located 9.9Å from the Rezatapopt ligand (Figure 15B) and five promising hapteneizable peptides containing this residue: two very high affinity (21nM, 28nM), two strong binders (148nM, 151nM), and one additional potential binder (584nM). These peptides are then synthesized, alkylated with a library of linkers, and tested in T2 cells to allow for rapid generation of HaCs. Hundreds more of these targets are now actively being explored, prompted by this data.

A



B



Figure 15: A) Pipeline for identification of promising targets. **B)** p53 and proximal cysteine.

2.10 REAL-TIME PATIENT RESPONSE

In addition to a comprehensive global search, the pipeline allows us to respond to patient needs and develop HaCs for specific targets in real time. Recently, a doctor treating a patient with an aggressive form of colorectal cancer approached the lab, seeking help in treating the disease. The cancer is driven by B-Rapidly Accelerating Fibrosarcoma (BRAF), and current treatments are limited and face acquired resistance (91). Thus, we are seeking to help by developing a HaC for BRAF. To do so, we are utilizing the newly developed pipeline. First, we identified a high affinity ligand: an imidazopyridinyl benzamide inhibitor that exhibits an IC₅₀ of 0.2nM for BRAF (PDB ID 6B8U). We next identified a proximal cysteine (C532) 3.72Å away from this ligand (Figure 16A) and identified one peptide containing this cysteine with affinity 946.57nM for HLA-A*02:01 (Figure 16B). We synthesized this peptide with the

HaloTag spiroamine linker and subsequently tested it in T2 cells (Figure 16B,C). This haptenated peptide generated a 1.64 and 1.87-fold increase in T-cell activation at 50 μ M and 100 μ M respectively (Figure 16C). Having established this as a baseline, we are now using the solid-phase synthetic scheme to test a large library of linkers and improve upon this response. The most efficacious linkers will be used to synthesize complete HaC molecules (which is already being done in parallel). Within a week of starting our search, we had begun testing and optimizing linkers in cells. This was not previously possible, as the timescale required to identify, synthesize, and develop a HaC was simply too long. With the new tools we have created we can help patients like these, demonstrating the immense translatability of the work and the importance of having such an efficient workflow. We are hopeful that this pipeline will continue to help lots of patients in real time.

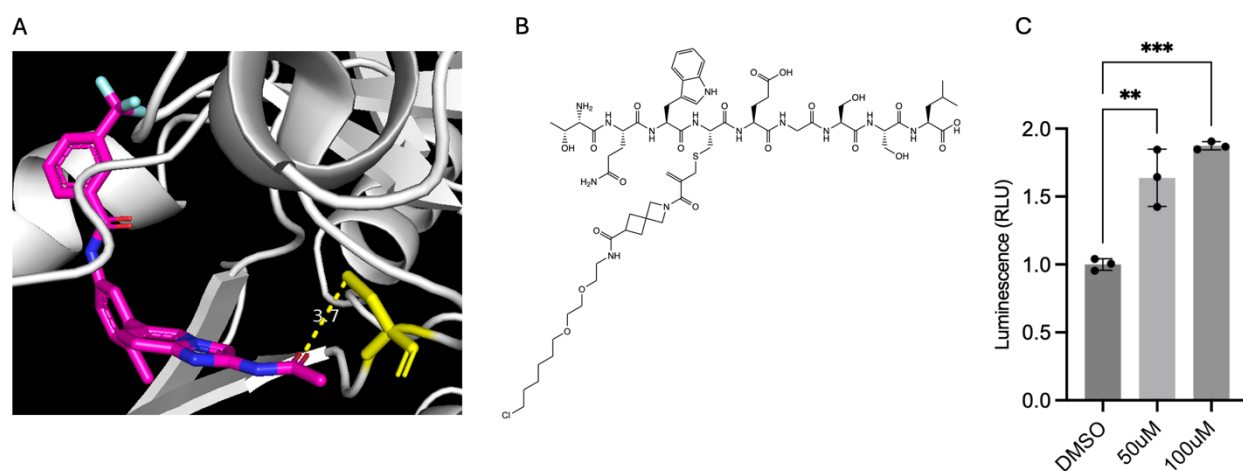
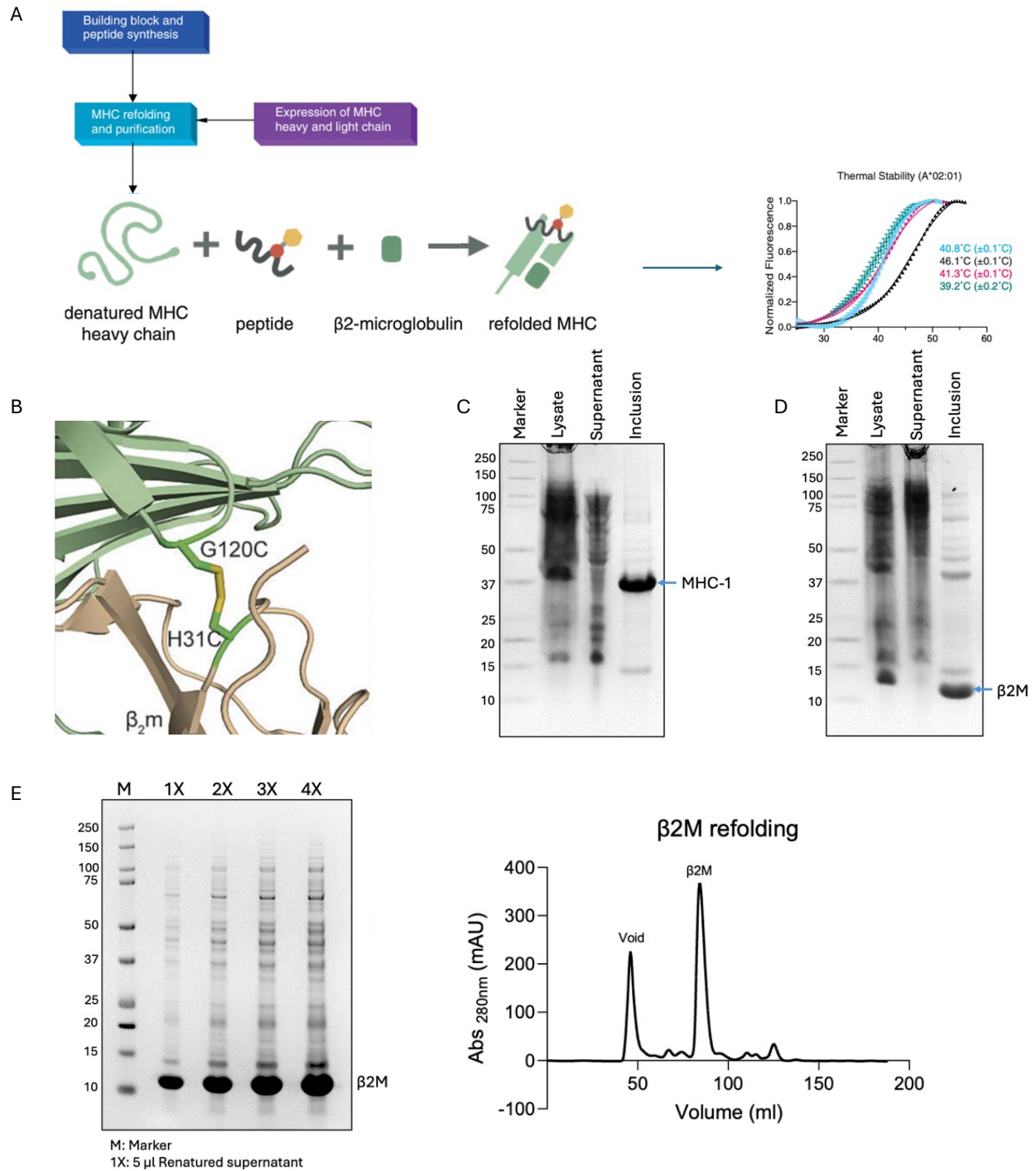


Figure 16: **A)** Proximal cysteine to BRAF ligand. **B)** Structure of BRAF peptide. **C)** T2 assay of BRAF peptide.

2.11 BIOPHYSICAL ASSAY DEVELOPMENT

To bolster the pipeline, further validate our results, and give additional insight into the peptide-MHC binding interaction, we also hope to develop a biophysical assay that is orthogonal to the T2 assay that directly and quantitatively measures haptenated-peptide affinity for the MHC I complex. The MHC I complex is a heterodimer consisting of its heavy chain non-covalently associated with its light chain, the beta-2 microglobulin (β 2M) (39,92,93). This heterodimer is very unstable and readily dissociates without a peptide to stabilize its conformation. Our first strategy was to try and use ‘empty’ MHC complexes (just the heavy and light chains without any peptide) and use surface plasmon resonance (SPR) to measure affinity. This was performed by immobilizing the biotinylated ‘empty’ MHC on a

streptavidin chip and flowing over our haptenated peptide of interest, allowing it to associate with the ‘empty’ MHC. This did not work due to the instability of the MHC complex, which quickly dissociated without a stabilizing peptide. We realized that we needed to generate the peptide-MHC complex first and then perform either a competition assay or a thermal stability assay, such as differential scanning fluorimetry (DSF). Competition assays have been established in the literature as a valid method to measure peptide-MHC affinity (94), and DSF thermal shift was used by Shokat et. al. to measure the relative affinity of haptenated-peptides for the MHC (Figure 17A) (39,95,96). In order to do this, we needed to generate a large amount of haptenated-peptide-MHC-complex. Most groups, including Shokat et al., did this by using denatured (unfolded) MHC heavy chain and β 2M, and refolding them in the presence of a very high concentration (large excess) of the peptide of interest (39,93). However, this method is extremely low throughput and requires very large quantities of haptenated peptide, both of which are antithetical to the pipeline as we aim to test a diverse set of haptenated peptides with different linkers in a high throughput manner. Note that we did initially try this method but found extremely low yields. Subsequently, we developed the following workflow: we identified and used a mutant version of the β 2M and MHC heavy chain that each contained cysteine mutations in non-binding regions (Figure 17B) (92,97). This does not affect the MHC’s ability to bind peptides but instead allows the two chains to form a disulfide bridge and thus a more stable complex (92,97). These were mutants of the HLA-A*02:01 allele, maximizing compatibility with the T2 assay. This complex is still not stable on its own but is stable when in complex with a known very low affinity ‘placeholder’ peptide (sequence: LFGYPVYV) (92). This peptide can then be competed off simply by incubating with a different peptide of interest (92). Thus we can generate large scale quantities of this placeholder-MHC-complex, and then easily generate haptenated-peptide-MHC-complexes by taking aliquots from this stock and simply incubating with any haptenated peptide of interest in a high throughput manner (Figure 17A). These complexes would subsequently be used in DSF assays to determine their relative stability and identify which linkers stabilized the structures the best.



To generate these complexes, plasmids for each of these disulfide mutants (Supplementary Figure 3) were expressed in *E. Coli* in inclusion bodies, and both the unfolded $\beta 2M$ and MHC heavy chain were isolated (Figure 17C,D). Previous literature indicated that the complex refolding reaction should take place with denatured MHC heavy chain but refolded $\beta 2M$ (93), so $\beta 2M$ was subsequently refolded in the presence of glutathione after 3 rounds of dialysis, after which it was purified with size exclusion chromatography (SEC) (Figure 17E). Three complex refolding reactions were run: one with denatured MHC heavy chain, renatured $\beta 2M$, and the placeholder peptide (Figure 18A), another without the placeholder peptide (Figure 18B), and one with just MHC heavy chain (Figure 18C). In addition, we ran a fourth refolding reaction with the placeholder peptide and denatured versions of both the MHC heavy chain and $\beta 2M$ (Figure 18D). After refolding, each of the conditions were subjected to SEC. Although complex formation was identified in the chromatogram (Figure 18E), only individual components could be recovered and intact complexes were not isolated for any of the conditions (Figure 18A,B,D). Thus, additional work is being conducted to optimize the protocol and generate intact complexes, which will bolster the pipeline further.

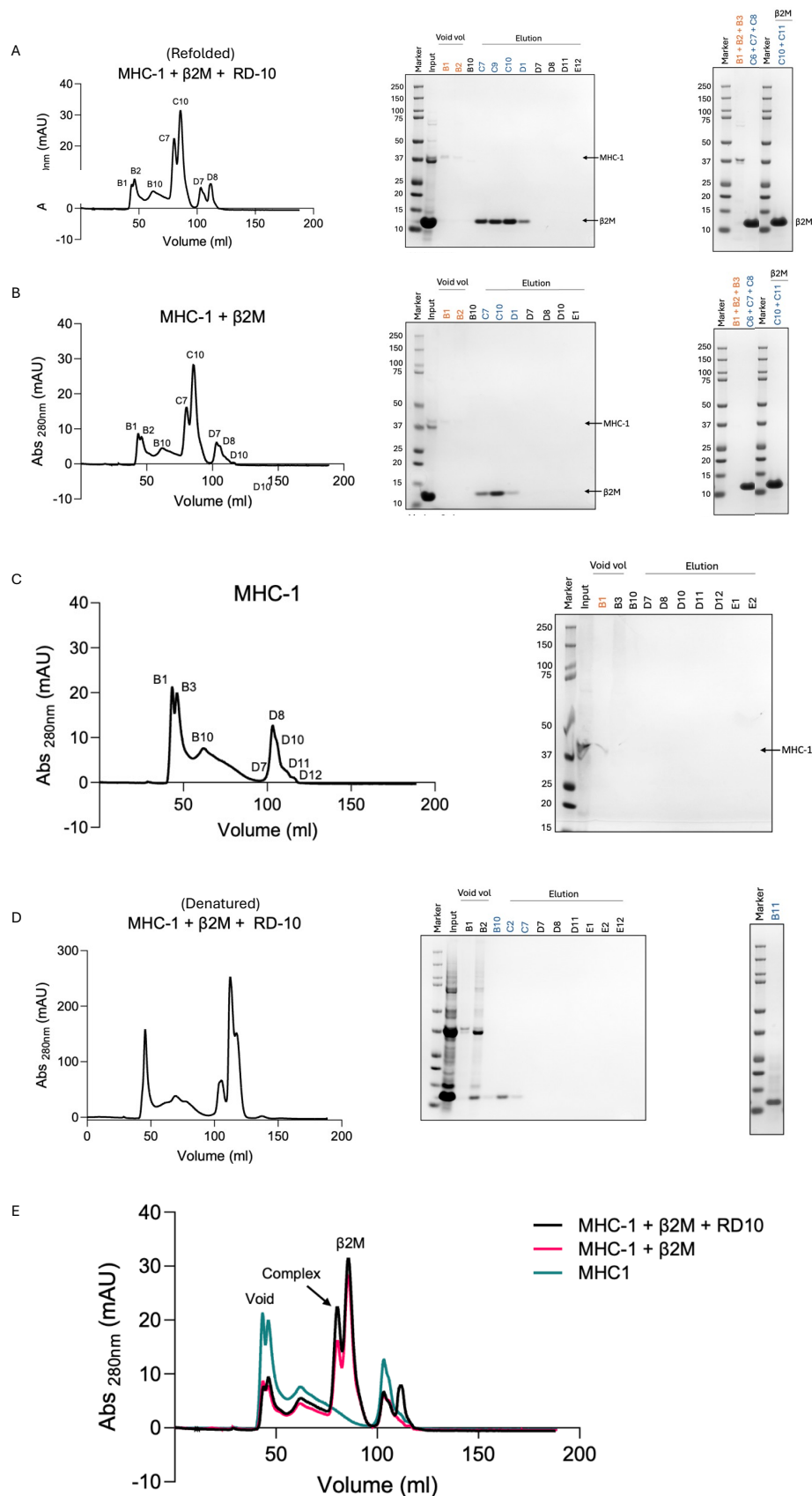


Figure 18: **A)** Refolding reactions with MHC, refolded B2M, placeholder peptide. **B)** Refolding reactions with MHC, B2M. **C)** Refolding reactions with MHC only. **D)** Refolding reactions with MHC, denatured B2M, placeholder peptide. **E)** Combined chromatograms from all purification runs.

2.12 HYDROPHOBIC LINKERS AND POCKET TARGETING

A recent structure of an MHC-I peptide covalently bound to a KRAS^{G12C} inhibitor reveals a shallow MHC binding pocket, near the peptide binding pocket, being partially occupied by the inhibitor (98). We believe we can target these shallow pockets of the MHC and have begun performing structure analysis and docking studies to explore linker designs that may bind to these pockets (Figure 19C). As mentioned previously, we also believe the mechanism enhancing the affinity of haptenated peptides for the MHC is driven by hydrophobic interactions. Therefore, we hypothesized that highly hydrophobic groups may get the haptenated peptides to “stick” better to the MHC. To test this, we are synthesizing linkers with silyl and biphenyl groups and testing them in T2 cells (Figure 19A). These hydrophobic designs have worked to great degrees of success, with some compounds leading to nearly 6.5-fold increases in T-cell activation, the most seen by any haptenated peptide thus far (Figure 19B). Note that previous designs before implementation of the pipeline showed 2.5-fold increases in activation at best (Figure 11, 12), demonstrating the ability of the pipeline to lead to great advances in the design and efficacy of HaCs. These hydrophobic groups show great promise, and we continue to build upon this work, iteratively improving HaCs to new heights at rapid speeds. Note that with each improvement in linker design, we take a medicinal chemistry approach and perform structure-activity-relationship studies to further understand these peptide-MHC interactions. These guide the designs of the subsequent linkers which we synthesize and test.

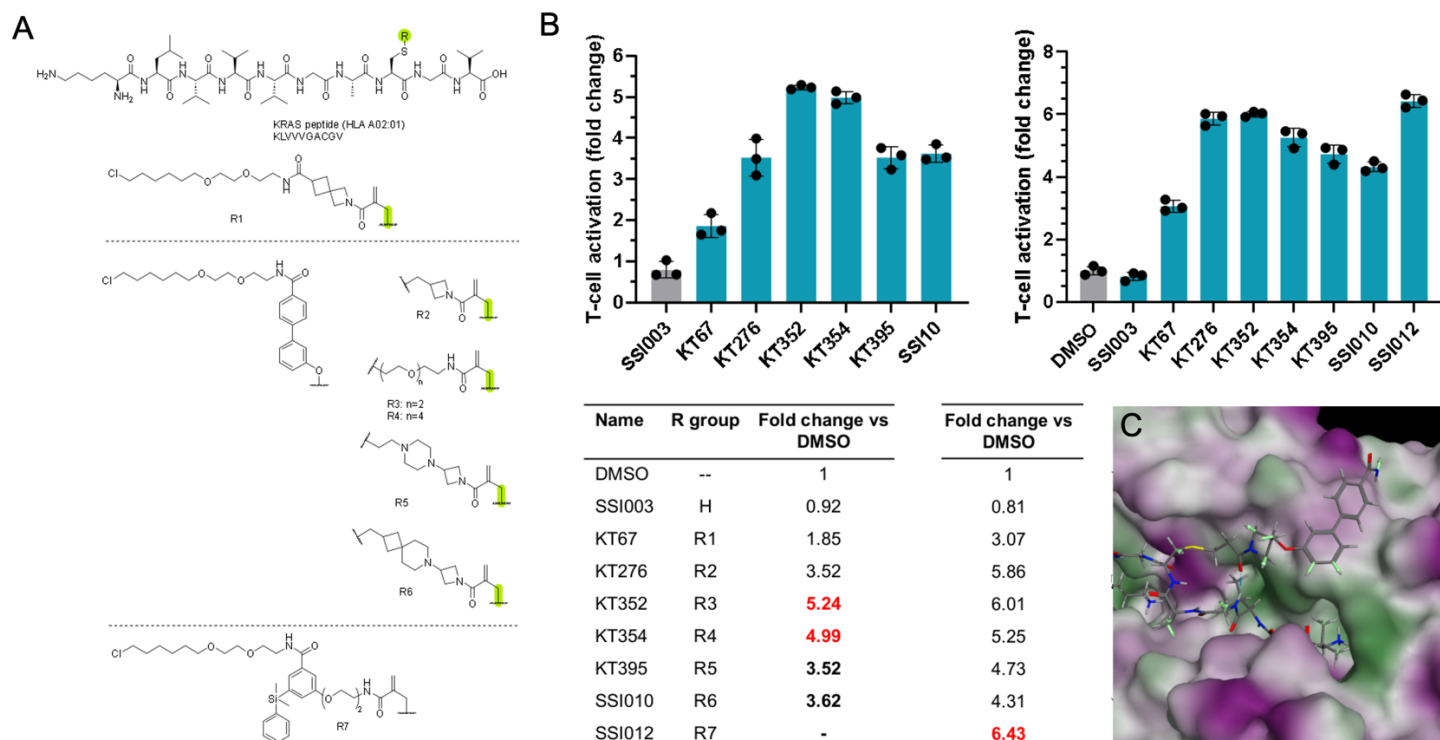


Figure 19: A) Structure of hydrophobic linkers. **B)** T2-cell assay results with hydrophobic linkers. **C)** MHC pocket bound to covalent inhibitor.

3 MATERIALS AND METHODS

3.1 PEPTIDE SYNTHESIS

The precursor peptide was synthesized at 100 μ mol scale using microwave peptide synthesizer Liberty Blue 2.0 (CEM, Matthews, NC). The Fmoc protecting group was removed with piperidine 20% in dimethyl formamide (DMF) (75°C, 15s; followed by 90°C, 50s) while peptide coupling was carried out with amino acid (5 equiv.), DIC (5 equiv.), oxyma (5 equiv.) (75°C, 15s; followed by 90°C, 110s) in DMF. Cysteine was introduced using Fmoc-L-Cys(Mmt)-OH. Resin was washed with DMF between coupling and deprotection cycle or after every other reaction. When the last amino acid was coupled, the Fmoc was deprotected by piperidine 20% and the N-terminus was capped with Boc₂O (equiv.), di-isopropyl ethylamine (DIPEA) (5 equiv.) in DMF for 30 min. The resin was washed thrice with dichloromethane (DCM) and the monomethoxytrityl (Mmt) group was selectively removed from the cysteine by treating the resin five times with the mixture DCM/TFA/TIPS (95:1:4) at room temperature, 2 min each treatment. Residual trifluoroacetic acid (TFA) was removed by washing thoroughly with DMF. The free cysteine residue was alkylated with 2-(bromomethyl)acrylic acid (5 equiv.), DIPEA (10 equiv.) in dry DMF for 30 min. The HaloTag chloroalkane (1 equiv.) was coupled using HATU (2.5 equiv.), DIPEA (5 equiv.) in dry DMF. The peptide was cleaved with TFA-water-TIPS (95:2.5:2.5) for 2h (4 mL for 200 mg resin). The liquid was collected by filtration and concentrated under nitrogen flow until < 1mL. The peptide was precipitated in 8 mL cold ether. The solid was isolated by centrifugation (2000 rpm, 10 min at 4°C), dried for 5 min under nitrogen flow before redissolved in DMSO for purification. The product was purified by preparative HPLC using gradient 10-35% acetonitrile (ACN) in water (+0.1% formic) for 30 min on XBridge Prep C18 5 μ m 19x250mm OBD column (Waters, Milford, MA) with AccQPrep HP150 system (Teledyne, Thousand Oaks, CA). The peptide was lyophilized and then lyophilized again in 1:1 Water:ACN +0.1% TFA. The peptide purity was assessed using Acquity UPLC-MS system (Waters, Milford, MA), equipped with BEH C18 1.7 μ m 2.1x50mm column. The gradient used ACN and water with 0.1% HCOOH (0 $\rightarrow$ 0.25 min, 5% ACN; 0.25 $\rightarrow$ 4.25 min, 5% $\rightarrow$ 95%; 4.25 $\rightarrow$ 4.75 min, 95%; 4.75 $\rightarrow$ 5.0 min, 95% $\rightarrow$ 5%). Peptides displayed purity over 95% before being used for testing.

3.2 GENERAL CHEMICAL SYNTHESIS

For chemical synthesis, all reagents were purchased and used as received from commercial sources without further purification. Reactions were performed in round-bottom flasks or pressure relief vials (purchased from Chemglass) stirred with Teflon®-coated magnetic stir bars. Moisture and air-sensitive reactions were performed under nitrogen atmosphere. Moisture and air-sensitive liquids or solutions were transferred via nitrogen-flushed syringes. Flash column chromatography was performed using silica gel (60 Å mesh, 20–40 µm) on a Teledyne Isco CombiFlash Rf system. HPLC purification was performed on Teledyne ISCO ACCQPrep™ HP150 equipped with an XBridge BEH Prep Column (19 mm x 250 mm). UPLC-MS was performed on a Waters ACQUITY UPLC I-Class 15 PLUS system with an ACQUITY SQ Detector 2.

3.3 T2 CELL ASSAY

This assay was a 3-day long protocol. T2 cells were cultured in IMDM + 20 % Hi-FBS supplemented with PenStrep (1x). Peptide solutions were prepared at 10-20 mM in DMSO. Day 1: Wash T2 cells were washed once with Aim V medium and resuspend at 1×10^6 cells /mL. Peptide solutions were prepared at 2x concentrations (200µM) in Aim V medium. DMSO controls were prepared by adding DMSO to Aim V medium such that the concentration of DMSO was the same as the peptide solutions. 300 µL T2 cells (for three technical replicates together) were mixed with 300 µL peptide solution, transferred to a 24-well plate, and incubated at 37°C for 18h. Day 2: unbound peptides washed out (at 200 x g, for 3 min) twice with 1 ml of RPMI-1640 medium and pelleted down the T2 cells at the end of the spin. T2 cells (pelleted down in the above step) were counted and co-cultured with 60K T-cells (at 1:1 ratio) in the presence of UniTE (at 2µM), “Halo-3(G4S)-aCD3-ScFV” for HaloTag-based peptides for 24 hours. Day 3: Bright-Glo luciferase assay reagent was added to each well and incubated for 3-5 min. Luminescence was recorded and normalized to that of control wells to report the relative luminescence units (RLUs) as fold change.

3.4 COMPLETE HAC COMPOUND CELL ASSAY

Cell lines used were as follows: KRAS: MIA PaCa-2, FGFR: AN3CA, BTK: Raji, ITK: Jurkat, EGFR:

H1975 (double mutant). For the T-cell activation assay using live target cells, the following steps were followed: Day 1: Target (adherent) cells were seeded at 0.1×10^6 cells/well in 100 μ L media in a 96-plate and incubated for two days. Day 3: Target suspension cells (0.1×10^6 cells/well) and the day 1 seeded adherent cells were incubated with 1-10 μ M (in 100 μ L) of the indicated compounds and the control (DMSO only) for 24 hours. Day 4: The cells were then washed gently four times with complete media (200 μ L/wash) and cocultured with 0.6×10^6 Jurkat cells (IL2-Luciferase reporter cell line) at the [target: effector] ratio of 1:6 in the presence of UniTEs, “Halo-3G4S-aCD3-ScFV” for HaloTag based HaCs and “FKBP-3G4S-aCD3-ScFV” for FKBP based HaCs, at 2 μ M for 24 hours. Day 5: Bright-Glo luciferase assay reagent was added to each well and incubated for 3-5 min. Luminescence was recorded and normalized to that of control wells to report the relative luminescence units (RLUs) as fold change.

3.5 COMPUTATIONAL PREDICTIONS

NetMHC was first accessed using TepiTool, a pipeline for computational prediction of T-cell epitope candidates provided by the Immune Epitope Database (IEDB). TepiTool was accessed online through IEDB at <http://tools.iedb.org/tepitool/>. TepiTool was obtained with an academic license. Python scripts and Linux 64-bit binaries were run on Ubuntu 22.10 through a UTM virtual machine. NetMHCpan 4.1 is available at <https://services.healthtech.dtu.dk/services/NetMHCpan-4.1/> and was implemented in Ubuntu 22.4 using the UTM virtual machine. Protein sequences were taken as the canonical sequence from UniProt with their respective mutations made, and MHC alleles for each cell line were taken from Scholtabers et al. (AN3-CA, J-82, H-929, H1975, Jurkat) (99), Boegel et al. (Raji) (100), or Kawaoka et al. (MIA PaCa-2) (101). Parameters used for the prediction include Human Class I alleles, with specific alleles being selected based on the cell line. Screening was performed for all 8-11 amino acid length peptides in the sequence. Peptides were scored using predicted IC_{50} by the NetMHCpan 4.1 prediction method. This method was used as it is the ‘IEDB Recommended’ method and consistently has the most accurate binding predictions on weekly benchmarks performed by IEDB (<http://tools.iedb.org/mhci/help/#Method>). 500 and 5000nM were used as cutoff IC_{50} values for binders.

Protein Sequence - UniProt (isoform chosen as the canonical sequence):

1. FGFR 1 (<https://www.uniprot.org/uniprotkb/A0A0S2Z3Q6/entry>)

2. FGFR 2 (<https://www.uniprot.org/uniprotkb/P21802/entry#sequences>)
3. FGFR 3 (<https://www.uniprot.org/uniprotkb/P22607/entry#sequences>)
4. FGFR 4 (<https://www.uniprot.org/uniprotkb/P22455/entry#sequences>)
5. KRAS (changed G12 → C)
(<https://www.uniprot.org/uniprotkb/P01116/entry#sequences>)
6. EGFR (changed L858→ R, T790→M)
(<https://www.uniprot.org/uniprotkb/P00533/entry#sequences>)
7. ITK (<https://www.uniprot.org/uniprotkb/Q08881/entry#sequences>)

Alleles - Cellosaurus:

1. https://www.cellosaurus.org/CVCL_0028 (AN3-CA)
2. https://www.cellosaurus.org/CVCL_0359 (J-82)
3. https://www.cellosaurus.org/CVCL_1600 (H-929)
4. https://www.cellosaurus.org/CVCL_0428 (MIA PaCa-2)
5. https://www.cellosaurus.org/CVCL_0511 (Raji)
6. https://www.cellosaurus.org/CVCL_1511 (H1975)
7. https://www.cellosaurus.org/CVCL_0367 (or https://www.cellosaurus.org/CVCL_0065)
(Jurkat)

Part of the code can be found at <https://github.com/as1000/LigandableResiDist>. Note that much of the code is being submitted for copyright. Thus, it is not publicly available at this time.

3.6 MHC REFOLDING

Protocol was adapted from Rodenko et al., 2006, and Sun et al. 2023 (92,93). Plasmids were obtained from AddGene (Watertown, MA) (Plasmid #215569, Plasmid #215657). High-efficiency competent *E. Coli* were obtained from New England Biolabs. Bacteria were grown in luria broth (LB) media. After expression, heavy-chain-expressing bacteria were grown in the presence of kanamycin (50µg mL⁻¹) while β2M-expressing bacteria were grown in the presence of ampicillin (100µg mL⁻¹). A single colony was inoculated in 10mL of LB. After 4 hours of growth at 37°C, OD600 was around 0.7 at which point bacteria were diluted to 1L of LB and grown to an OD600 of around 0.6 (~4 hours). The

resulting bacteria were split into two flasks and 200 μ L of IPTG was then added to each to begin induction. Cells were grown at 37°C for 4 hours. Pellets were collected with centrifugation, divided into 50mL falcon tubes, and centrifuged again. 8mL of lysis buffer was added for every 5g of cells, after which the cells were sonicated. 20mg of lysozyme was added and cells were incubated on ice for 30 minutes. 20mL of detergent buffer was added, and the cells stored at -80°C. Pellet was washed with a wash buffer (0.05% Tween 20 (cat. no. P1379, Sigma-Aldrich) in PBS), resuspended, and then centrifuged again. This was repeated thrice more.

For β 2M refolding, pellets were suspended in 10mL denaturing buffer (8 M urea, 100 mM Tris-HCl (pH 8)), after which protease inhibitor cocktail was added, as well as 40 more mL of denaturing buffer (0.2 M NaCl, 1% deoxycholic acid, 1% NP40, 20 mM Tris-HCl (pH 7.5), 2 mM EDTA.). This solution was then taken to dialysis with MW cut-off 3.5 kDa dialysis tubing and dialyzed overnight against 2L of 10 mM Tris-HCl (pH 7). Dialysis was repeated and solution was concentrated with centrifugation in Centrplus filters (MW cutoff 3.5 kDa) and then purified. All purification was done on Cytiva AKTA Pure, with HiLoad 16/600 75 μ g column at 4°C with injection: 2ml, flow rate: 0.8 ml/min, buffer: 20mM Tris (pH 7.5, 100 mM NaCl).

For refolding, pellets of the heavy chain were dissolved in 1 mL of denaturing buffer and resuspended. The suspension was centrifuged and transferred to a fresh tube. In 50mL of PBS, 20mM stock of peptide was added for a final concentration of 40 μ M, refolded β 2M (refolded as described above), and heavy chain (as described previously). Refolding occurred at 4°C for 72 hours. The refolded sample was concentrated to 4mL and purified using SEC as described above.

3.7 STATISTICAL ANALYSES

Statistical analyses were performed using GraphPad Prism 10. Significance analysis was done using an ordinary one-way ANOVA (no matching or pairing), assuming a normal distribution with equal standard deviations across groups. Variances were confirmed to be homogenous using the Brown-Forsythe test and Bartlett's test. Dunnett's multiple comparisons test was performed to compare each treatment group against DMSO. Adjusted p-values were used to demonstrate significance as follows: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.0001$ (***)

3.8 FIGURES

Figures were made in Adobe Illustrator, ChemDraw, and GraphPad Prism 10.

4 DISCUSSION

4.1 FUTURE WORK

In addition to stabilizing the haptenated-peptide-MHC-interaction, the lab is also looking at other mechanisms of improving efficacy. Some groups have found that stabilizing the peptide interaction with TAP leads to increased MHC presentation, as more of the peptide is presented to the MHC complex (102,103). Additionally, we are looking to make HaCs from proteolysis targeting chimeras (ProTaCs), another emerging and promising modality, by using an E3-ligase-recruiting-group as the hapten (Figure 20A) (104). This would cause increased ubiquitination and degradation of the targeted protein, leading to increased intracellular concentrations of haptenated peptides, and thus enhanced extracellular presentation (Figure 20A). Finally, as we aim to increase the translatability of this work, *in vivo* animal models are required to validate much of the data, and mouse models will be developed for these compounds.

Extracellular markers have long served as the cornerstone of cancer treatment (105,106). They represent the largest category of drug targets, with two-thirds of all approved drugs targeting cell surface proteins (105,106). They enable drugs to achieve specificity to cancer cells without harming healthy tissue, paving the way for the advent of antibody-drug-conjugates (ADCs) and other targeted therapeutics. However, extracellular markers are quite hard to come by (39,107). As of 2023, only 11 of the roughly 5000 extracellular proteins had led to FDA approved ADCs due to insufficient selectivity of most targets (105,108). In addition to the limited diversity of the extracellular space, only around a quarter of the proteasome is extracellular, meaning there are likely three times as many intracellular proteins as extracellular ones (6,109). This platform aims to vastly expand the number of targetable extracellular markers by enabling the presentation of intracellular proteins to the cell surface. Since these mutant intracellular targets are either uniquely present or highly overexpressed in cancer cells, the specificity of these antigens is also dramatically improved. Additionally, the enhanced immunogenicity of these peptides allows us to unlock even more of this intracellular space (Figure 12A). Acquired cysteines and lysines represent a large portion of oncogenic mutations, especially gain-of-function mutations (110,111), further enhancing the viability of HaCs. Although it is likely that these haptens will not be as abundant as native extracellular markers due to the limited rate of MHC

presentation, opening up this set of diverse and disease-specific intracellular targets would dramatically change the face of targeting cancer. The generalizability and reach of HaCs could be revolutionary. Any druggable target with a proximal lysine or cysteine in the binding pocket could become treatable with HaCs. Our hope is that we may be able to one day produce HaCs for a majority of druggable targets, and we will need high-throughput pipelines like this one to achieve our goal.

Finally, our vision of standardizing the hapten across all developed HaCs potentially removes the need for UniTEs entirely, replacing them with chimeric antigen receptor T-cells (CAR-T cells). These CAR-T cells would have the hapten binding region of the UniTE as their antigen receptor (Figure 20B). Large populations could be preemptively treated with these CAR-T cells, which would remain inert in the body (given high specificity of the receptor) as their corresponding antigen is not endogenously present. If the patient develops cancer, they can be treated with a HaC for that specific target, and the CAR-T cells would recognize the hapten, leading directly to targeted cell death without the need for a UniTE to mediate the immune synapse (Figure 20B). Since the hapten is standardized across HaCs, cancers due to any druggable oncoprotein could be treated in this way and the CAR-T treatment would be universally applicable. This would necessitate developing a large library of HaCs for a wide array of targets. While this vision is exceedingly far from the current reality, we are hopeful that we can slowly work towards it.

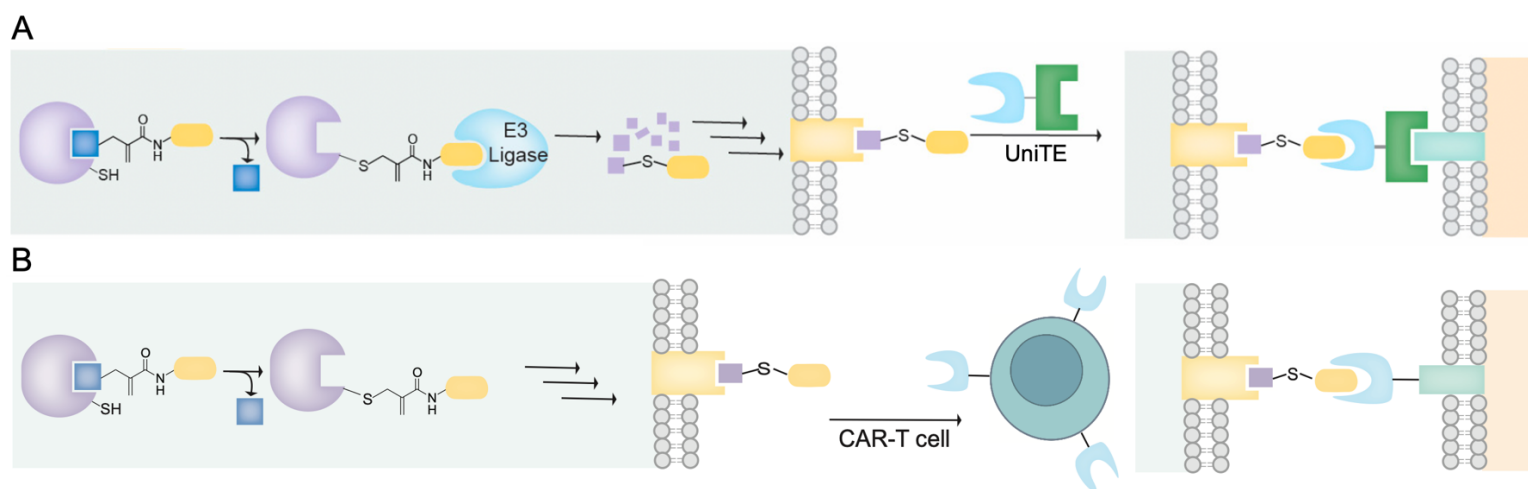


Figure 20: A) HaCs from ProTaCs. **B)** HaCs with universal CAR-T cells.

4.2 CONCLUSION

HaCs are a promising modality that can be used not just to combat acquired resistance but serve as independent immunotherapeutic agents. The beauty of HaCs lies in their scalability, as they take advantage of the vast library of readily available and highly specific inhibitors and provide a universal method of repurposing them. We report the development of a complete pipeline being used to scale HaCs to great heights in a cost-effective and efficient manner. The pipeline allows us to start from a comprehensive search for all possible targets, identify the most promising candidates, and subsequently test and optimize a wide array of linker designs in a high throughput manner. Utilizing the pipeline has allowed HaCs to achieve 6.5-fold increases in T-cell activation, nearly tripling their efficacy compared to past-generation compounds. It allows us to both identify promising new targets and respond to specific patient needs in real time. Finally, it opens up the vast potential of the intracellular space to serve as cell surface markers for targeted cancer therapies that have the ability to become universal.

5 REFERENCES

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
2. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91**, 355–358 (2004).
3. Cox, A. D. & Der, C. J. “Undruggable KRAS”: druggable after all. *Genes Amp Dev.* (2024). doi:10.1101/gad.352081.124
4. Huang, L., Guo, Z., Wang, F. & Fu, L. KRAS mutation: from undruggable to druggable in cancer. *Signal Transduct. Target. Ther.* **6**, (2021).
5. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–551 (2013).
6. Prior, I. A., Lewis, P. D. & Mattos, C. A Comprehensive Survey of Ras Mutations in Cancer. *Cancer Res.* **72**, 2457–2467 (2012).
7. Yang, Y., Zhang, H., Huang, S. & Chu, Q. KRAS Mutations in Solid Tumors: Characteristics, Current Therapeutic Strategy, and Potential Treatment Exploration. *J. Clin. Med.* **12**, 709 (2023).
8. Lim, T. K. H. *et al.* KRAS G12C in advanced NSCLC: Prevalence, co-mutations, and testing. *Lung Cancer* **184**, 107293 (2023).
9. Strickler, J. H. *et al.* Prevalence of KRAS G12C Mutation and Co-mutations and Associated Clinical Outcomes in Patients With Colorectal Cancer: A Systematic Literature Review. *The Oncologist* **28**, e981–e994 (2023).
10. Qunaj, L., May, M. S., Neugut, A. I. & Herzberg, B. O. Prognostic and therapeutic impact of the KRAS G12C mutation in colorectal cancer. *Front. Oncol.* **13**, (2023).

11. Dy, G. K. *et al.* Long-Term Outcomes and Molecular Correlates of Sotorasib Efficacy in Patients With Pretreated KRAS G12C-Mutated Non–Small-Cell Lung Cancer: 2-Year Analysis of CodeBreaK 100. *J. Clin. Oncol.* **41**, 3311–3317 (2023).
12. Awad, M. M. *et al.* Acquired Resistance to KRAS G12C Inhibition in Cancer. *N. Engl. J. Med.* **384**, 2382–2393 (2021).
13. Zheng, L. *et al.* Development of covalent inhibitors: Principle, design, and application in cancer. *MedComm – Oncol.* **2**, (2023).
14. Xie, X. *et al.* Recent advances in targeting the “undruggable” proteins: from drug discovery to clinical trials. *Signal Transduct. Target. Ther.* **8**, (2023).
15. Bauer, R. A. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov. Today* **20**, 1061–1073 (2015).
16. Zhang, T., Hatcher, J. M., Teng, M., Gray, N. S. & Kostic, M. Recent Advances in Selective and Irreversible Covalent Ligand Development and Validation. *Cell Chem. Biol.* **26**, 1486–1500 (2019).
17. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
18. Boike, L. *et al.* Discovery of a Functional Covalent Ligand Targeting an Intrinsically Disordered Cysteine within MYC. *Cell Chem. Biol.* **28**, 4-13.e17 (2021).
19. Raouf, Y. S. Covalent Inhibitors: To Infinity and Beyond. *J. Med. Chem.* **67**, 10513–10516 (2024).
20. Goyal, L. *et al.* TAS-120 Overcomes Resistance to ATP-Competitive FGFR Inhibitors in Patients with FGFR2 Fusion–Positive Intrahepatic Cholangiocarcinoma. *Cancer Discov.* **9**, 1064–1079 (2019).

21. Zapf, C. W. *et al.* Covalent Inhibitors of Interleukin-2 Inducible T Cell Kinase (Itk) with Nanomolar Potency in a Whole-Blood Assay. *J. Med. Chem.* **55**, 10047–10063 (2012).
22. Alu, A., Lei, H., Han, X., Wei, Y. & Wei, X. BTK inhibitors in the treatment of hematological malignancies and inflammatory diseases: mechanisms and clinical studies. *J. Hematol. Amp Oncol.* **15**, (2022).
23. Zhong, L. *et al.* Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. *Signal Transduct. Target. Ther.* **6**, (2021).
24. Engel, J., Lategahn, J. & Rauh, D. Hope and Disappointment: Covalent Inhibitors to Overcome Drug Resistance in Non-Small Cell Lung Cancer. *ACS Med. Chem. Lett.* **7**, 2–5 (2015).
25. Schwartz, P. A. *et al.* Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance. *Proc. Natl. Acad. Sci.* **111**, 173–178 (2013).
26. Sutanto, F., Konstantinidou, M. & Dömling, A. Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **11**, 876–884 (2020).
27. Tan, L. *et al.* Development of covalent inhibitors that can overcome resistance to first-generation FGFR kinase inhibitors. *Proc. Natl. Acad. Sci.* **111**, (2014).
28. Uribe, M. L., Marrocco, I. & Yarden, Y. EGFR in Cancer: Signaling Mechanisms, Drugs, and Acquired Resistance. *Cancers* **13**, 2748 (2021).
29. Honigberg, L. A. *et al.* The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc. Natl. Acad. Sci.* **107**, 13075–13080 (2010).
30. Lelais, G. *et al.* Discovery of (R,E)-N-(7-Chloro-1-(1-[4-(dimethylamino)but-2-enoyl]azepan-3-yl)-1H-benzo[d]imidazol-2-yl)-2-methylisonicotinamide (EGF816), a Novel, Potent, and WT

- Sparing Covalent Inhibitor of Oncogenic (L858R, ex19del) and Resistant (T790M) EGFR Mutants for the Treatment of EGFR Mutant Non-Small-Cell Lung Cancers. *J. Med. Chem.* **59**, 6671–6689 (2016).
31. Zhou, Q. *et al.* Potent covalent irreversible inhibitor of KRAS G12C IBI351 in patients with advanced solid tumors: First-in-human phase I study. *Eur. J. Cancer* **212**, 114337 (2024).
 32. Herrera, A. F. & Jacobsen, E. D. Ibrutinib for the Treatment of Mantle Cell Lymphoma. *Clin. Cancer Res.* **20**, 5365–5371 (2014).
 33. Zhao, Y. *et al.* Diverse alterations associated with resistance to KRAS(G12C) inhibition. *Nature* **599**, 679–683 (2021).
 34. Akhave, N. S., Biter, A. B. & Hong, D. S. Mechanisms of Resistance to KRASG12C-Targeted Therapy. *Cancer Discov.* **11**, 1345–1352 (2021).
 35. Tanaka, S. *et al.* Dynamic Cooperativity of Ligand–Residue Interactions Evaluated with the Fragment Molecular Orbital Method. *J. Phys. Chem. B* **125**, 6501–6512 (2021).
 36. Eberlein, C. A. *et al.* Acquired Resistance to the Mutant-Selective EGFR Inhibitor AZD9291 Is Associated with Increased Dependence on RAS Signaling in Preclinical Models. *Cancer Res.* **75**, 2489–2500 (2015).
 37. Ash, L. J. *et al.* KRAS: Biology, Inhibition, and Mechanisms of Inhibitor Resistance. *Curr. Oncol.* **31**, 2024–2046 (2024).
 38. Dalchau, N. *et al.* A Peptide Filtering Relation Quantifies MHC Class I Peptide Optimization. *PLoS Comput. Biol.* **7**, e1002144 (2011).
 39. Zhang, Z. *et al.* A covalent inhibitor of K-Ras(G12C) induces MHC class I presentation of haptenated peptide neoepitopes targetable by immunotherapy. *Cancer Cell* **40**, 1060-1069.e7 (2022).

40. Hattori, T. *et al.* Creating MHC-Restricted Neoantigens with Covalent Inhibitors That Can Be Targeted by Immune Therapy. *Cancer Discov.* **13**, 132–145 (2022).
41. Tran, E. *et al.* T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *N. Engl. J. Med.* **375**, 2255–2262 (2016).
42. Wang, Q. J. *et al.* Identification of T-cell Receptors Targeting KRAS-Mutated Human Tumors. *Cancer Immunol. Res.* **4**, 204–214 (2016).
43. Dao, T. *et al.* Targeting the Intracellular WT1 Oncogene Product with a Therapeutic Human Antibody. *Sci. Transl. Med.* **5**, (2013).
44. Hsiue, E. H.-C. *et al.* Targeting a neoantigen derived from a common TP53 mutation. *Science* **371**, (2021).
45. Li, D. *et al.* Development of a T-cell Receptor Mimic Antibody against Wild-Type p53 for Cancer Immunotherapy. *Cancer Res.* **77**, 2699–2711 (2017).
46. Low, L., Goh, A., Koh, J., Lim, S. & Wang, C.-I. Targeting mutant p53-expressing tumours with a T cell receptor-like antibody specific for a wild-type antigen. *Nat. Commun.* **10**, (2019).
47. McCarthy, M. K. & Weinberg, J. B. The immunoproteasome and viral infection: a complex regulator of inflammation. *Front. Microbiol.* **6**, (2015).
48. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. The major histocompatibility complex and its functions. (2000).
49. Naito, T. & Okada, Y. HLA imputation and its application to genetic and molecular fine-mapping of the MHC region in autoimmune diseases. *Semin. Immunopathol.* **44**, 15–28 (2021).
50. Traherne, J. A. Human MHC architecture and evolution: implications for disease association studies. *Int. J. Immunogenet.* **35**, 179–192 (2008).
51. Robinson, J. *et al.* IPD-IMGT/HLA Database. *Nucleic Acids Res.* (2019).
doi:10.1093/nar/gkz950

52. Sanchez-Mazas, A. *et al.* The most frequent HLA alleles around the world: A fundamental synopsis. *Best Pract. Amp Res. Clin. Haematol.* **37**, 101559 (2024).
53. Vandemoortele, G., Eyckerman, S. & Gevaert, K. Pick a Tag and Explore the Functions of Your Pet Protein. *Trends Biotechnol.* **37**, 1078–1090 (2019).
54. Urh, M. HaloTag, a Platform Technology for Protein Analysis. *Curr. Chem. Genomics* **6**, 72–78 (2013).
55. England, C. G., Luo, H. & Cai, W. HaloTag Technology: A Versatile Platform for Biomedical Applications. *Bioconjug. Chem.* **26**, 975–986 (2015).
56. Yunis, A. A., Arimura, G. K. & Russin, D. J. Human pancreatic carcinoma (mia paca-2) in continuous culture: Sensitivity to asparaginase. *Int. J. Cancer* **19**, 128–135 (1977).
57. Purcell, A. W., Ramarathnam, S. H. & Ternette, N. Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* **14**, 1687–1707 (2019).
58. Du, S., Zhang, Y. & Xu, J. Current progress in cancer treatment by targeting FGFR signaling. *Cancer Biol. Amp Med.* **20**, 490–499 (2023).
59. Chen, W., Khilko, S., Fecondo, J., Margulies, D. H. & McCluskey, J. Determinant selection of major histocompatibility complex class I-restricted antigenic peptides is explained by class I-peptide affinity and is strongly influenced by nondominant anchor residues. *J. Exp. Med.* **180**, 1471–1483 (1994).
60. Hobohm, U. & Meyerhans, A. A pattern search method for putative anchor residues in T cell epitopes. *Eur. J. Immunol.* **23**, 1271–1276 (1993).
61. Steinle, A. & Schendel, D. J. HLA class I alleles of LCL 721 and 174XCEM.T2 (T2). *Tissue Antigens* **44**, 268–270 (1994).

62. Weaver, J. R., Good, K., Walters, R. D., Kugel, J. F. & Goodrich, J. A. Characterization of the sequence and architectural constraints of the regulatory and core regions of the human interleukin-2 promoter. *Mol. Immunol.* **44**, 2813–2819 (2007).
63. Janssen, A. W. F. *et al.* Determination of in vitro immunotoxic potencies of a series of perfluoralkylsubstances (PFASs) in human Namalwa B lymphocyte and human Jurkat T lymphocyte cells. *Front. Toxicol.* **6**, (2024).
64. Wang, S.-Sun. p-Alkoxybenzyl Alcohol Resin and p-Alkoxybenzyloxycarbonylhydrazide Resin for Solid Phase Synthesis of Protected Peptide Fragments. *J. Am. Chem. Soc.* **95**, 1328–1333 (1973).
65. Dubowchik, G. M. & Radia, S. Monomethoxytrityl (MMT) as a versatile amino protecting group for complex prodrugs of anticancer compounds sensitive to strong acids, bases and nucleophiles. *Tetrahedron Lett.* **38**, 5257–5260 (1997).
66. Stathopoulos, P., Papas, S., Sakka, M., Tzakos, A. G. & Tsikaris, V. A rapid and efficient method for the synthesis of selectively S-Trt or S-Mmt protected Cys-containing peptides. *Amino Acids* **46**, 1367–1376 (2014).
67. Troup, R. I., Fallan, C. & Baud, M. G. J. Current strategies for the design of PROTAC linkers: a critical review. *Explor. Target. Anti-Tumor Ther.* **1**, (2020).
68. Clackson, T. *et al.* Redesigning an FKBP–ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci.* **95**, 10437–10442 (1998).
69. Nabet, B. *et al.* The dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.* **14**, 431–441 (2018).
70. Wuts, P. G. M. & Greene, T. W. *Greene's Protective Groups in Organic Synthesis*. (Wiley, 2006).
doi:10.1002/0470053488

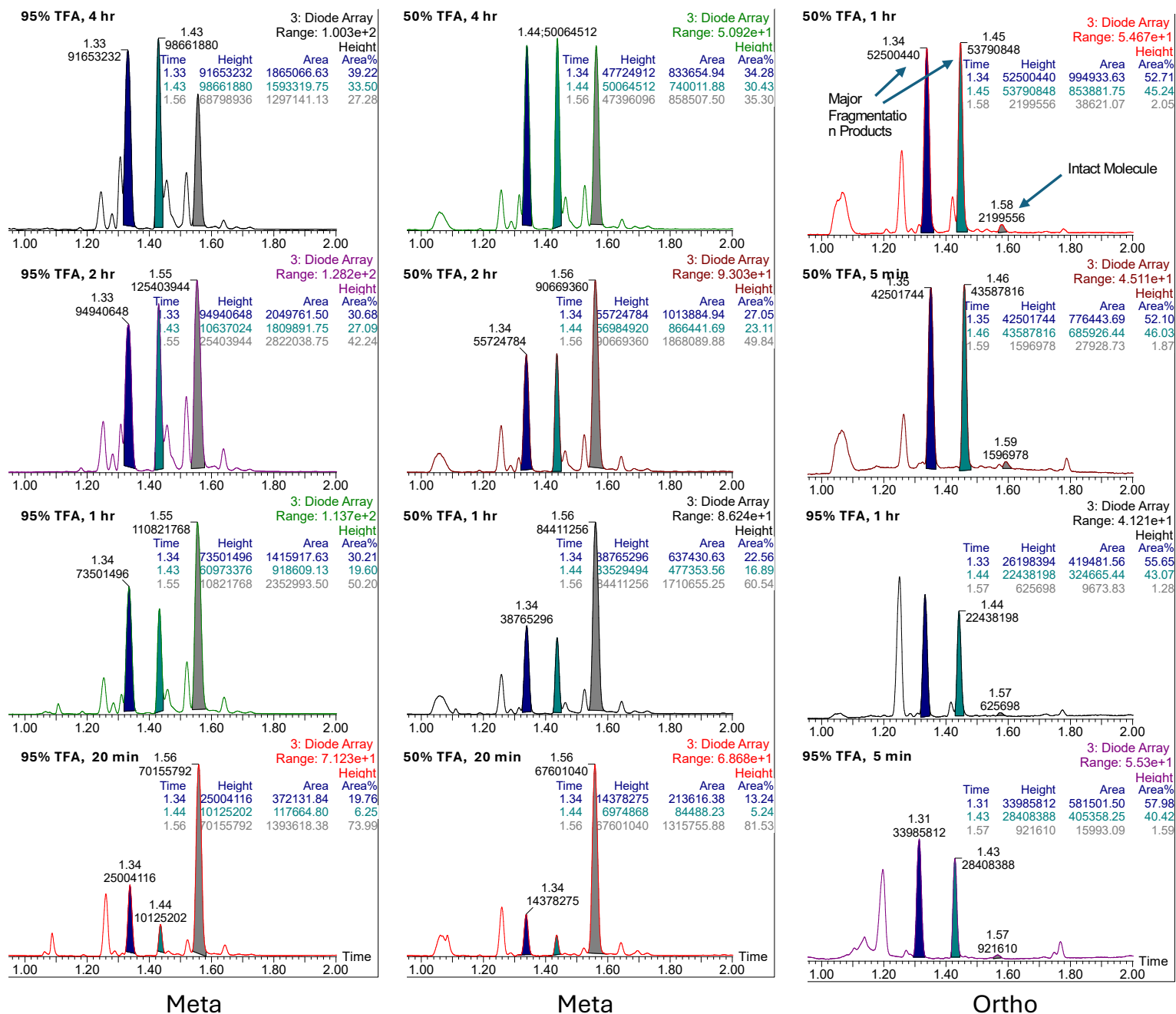
71. Bochet, C. G. Photochemical release of various functional groups. *Pure Appl. Chem.* **78**, 241–247 (2006).
72. Klán, P. *et al.* Photoremovable Protecting Groups in Chemistry and Biology: Reaction Mechanisms and Efficacy. *Chem. Rev.* **113**, 119–191 (2012).
73. Caron, E. *et al.* Analysis of Major Histocompatibility Complex (MHC) Immuno-peptidomes Using Mass Spectrometry*. *Mol. Amp Cell. Proteomics* **14**, 3105–3117 (2015).
74. Papassavas, E. C. *et al.* MHC class I and class II phenotype, gene, and haplotype frequencies in Greeks using molecular typing data. *Hum. Immunol.* **61**, 615–623 (2000).
75. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
76. Paul, S., Sidney, J., Sette, A. & Peters, B. TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr. Protoc. Immunol.* **114**, (2016).
77. Nielsen, M. *et al.* NetMHCpan, a Method for Quantitative Predictions of Peptide Binding to Any HLA-A and -B Locus Protein of Known Sequence. *PLoS ONE* **2**, e796 (2007).
78. Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1–13 (2008).
79. Zhao, W. & Sher, X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLOS Comput. Biol.* **14**, e1006457 (2018).
80. Lundegaard, C., Lund, O. & Nielsen, M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics* **24**, 1397–1398 (2008).

81. Trolle, T. *et al.* The Length Distribution of Class I–Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele–Specific Binding Preference. *J. Immunol.* **196**, 1480–1487 (2016).
82. Yi, X. *et al.* caAtlas: An immuno-peptidome atlas of human cancer. *iScience* **24**, 103107 (2021).
83. Fleri, W. *et al.* The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. *Front. Immunol.* **8**, (2017).
84. Paul, S. *et al.* HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *J. Immunol.* **191**, 5831–5839 (2013).
85. Sette *et al.* The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol. Baltim. Md 1950* **153**, (1994).
86. Assarsson, E. *et al.* A Quantitative Analysis of the Variables Affecting the Repertoire of T Cell Specificities Recognized after Vaccinia Virus Infection. *J. Immunol.* **178**, 7890–7901 (2007).
87. Sachs, A. *et al.* Impact of Cysteine Residues on MHC Binding Predictions and Recognition by Tumor-Reactive T Cells. *J. Immunol.* **205**, 539–549 (2020).
88. Jurtz, V. *et al.* NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J. Immunol.* **199**, 3360–3368 (2017).
89. Wang, H., Guo, M., Wei, H. & Chen, Y. Targeting p53 pathways: mechanisms, structures and advances in therapy. *Signal Transduct. Target. Ther.* **8**, (2023).
90. Vu, B. T. *et al.* Discovery of Rezatapopt (PC14586), a First-in-Class, Small-Molecule Reactivator of p53 Y220C Mutant in Development. *ACS Med. Chem. Lett.* **16**, 34–39 (2024).
91. Patel, H. *et al.* Current Advances in the Treatment of BRAF-Mutant Melanoma. *Cancers* **12**, 482 (2020).

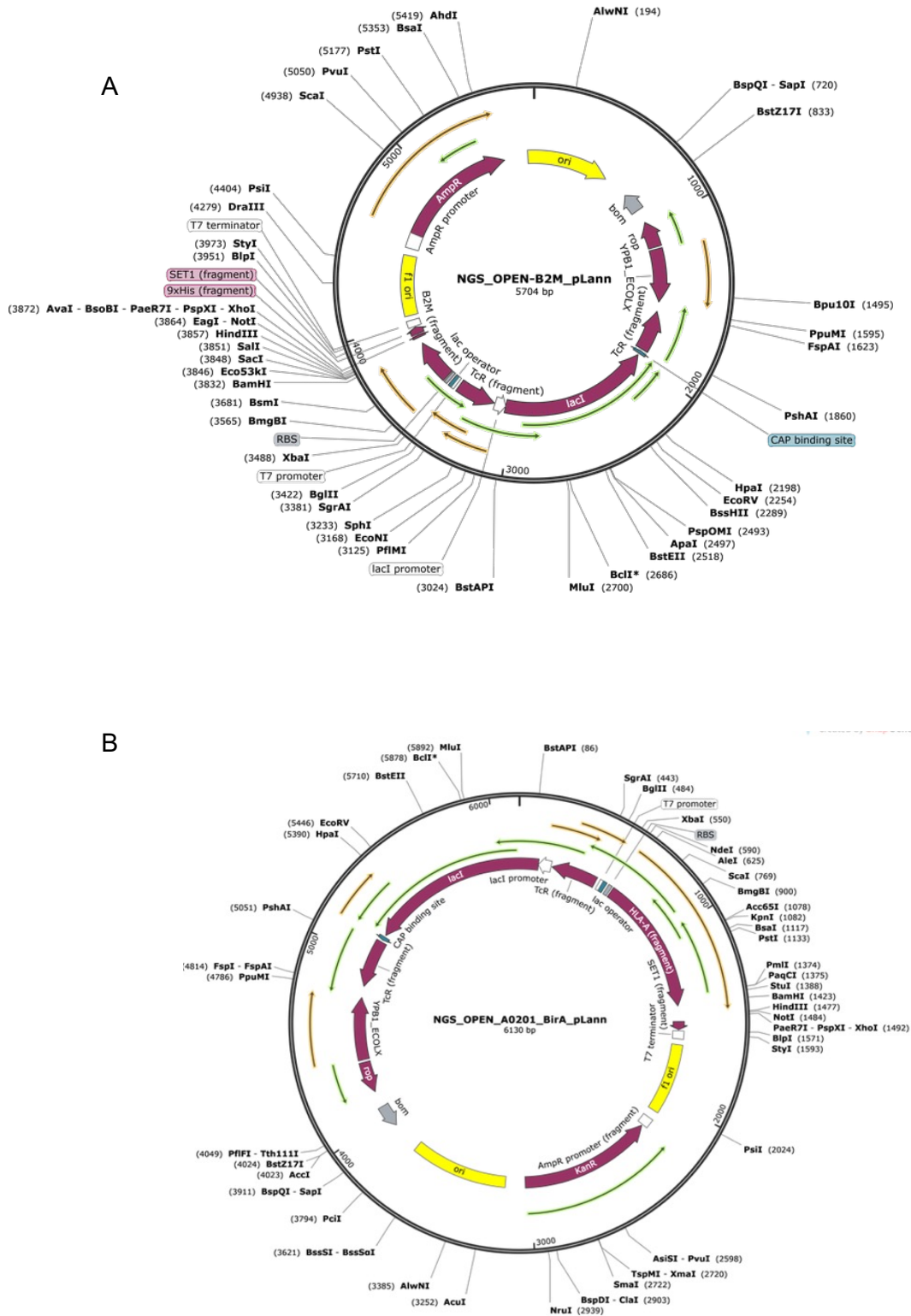
92. Sun, Y. *et al.* Universal open MHC-I molecules for rapid peptide loading and enhanced complex stability across HLA allotypes. *Proc. Natl. Acad. Sci.* **120**, (2023).
93. Rodenko, B. *et al.* Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nat. Protoc.* **1**, 1120–1132 (2006).
94. Sidney, J. *et al.* Measurement of MHC/Peptide Interactions by Gel Filtration or Monoclonal Antibody Capture. *Curr. Protoc. Immunol.* **100**, (2013).
95. Hellman, L. M. *et al.* Differential scanning fluorimetry based assessments of the thermal and kinetic stability of peptide–MHC complexes. *J. Immunol. Methods* **432**, 95–101 (2016).
96. Niebling, S. *et al.* FoldAffinity: binding affinities from nDSF experiments. *Sci. Rep.* **11**, (2021).
97. Anjanappa, R. *et al.* Disulfide-stabilized recombinant MHC class I molecules allow the generation of peptide-receptive MHC tetramers. *Mol. Immunol.* **150**, 25 (2022).
98. Maso, L. *et al.* Molecular basis for antibody recognition of multiple drug–peptide/MHC complexes. *Proc. Natl. Acad. Sci.* **121**, (2024).
99. Scholtalbers, J. *et al.* TCLP: an online cancer cell line catalogue integrating HLA type, predicted neo-epitopes, virus and gene expression. *Genome Med.* **7**, 118 (2015).
100. Boegel, S., Löwer, M., Bukur, T., Sahin, U. & Castle, J. C. A catalog of HLA type, HLA expression, and neo-epitope candidates in human cancer cell lines. *Oncoimmunology* **3**, e954893 (2014).
101. Kawaoka, T. *et al.* Adoptive immunotherapy for pancreatic cancer: cytotoxic T lymphocytes stimulated by the MUC1-expressing human pancreatic cancer cell line YPK-1. *Oncol. Rep.* **20**, 155–163 (2008).
102. Mantel, I., Sadiq, B. A. & Blander, J. M. Spotlight on TAP and its vital role in antigen presentation and cross-presentation. *Mol. Immunol.* **142**, 105–119 (2022).

103. Hewitt, E. W. The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology* **110**, 163–169 (2003).
104. Bekes, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discov.* **21**, 181–200 (2022).
105. Bausch-Fluck, D. *et al.* The in silico human surfaceome. *Proc. Natl. Acad. Sci.* **115**, (2018).
106. Laurent, A., Allard, A. & Fillet, M. Cell surface proteomics: Analytical challenges and clinical applications in cancer. *TrAC Trends Anal. Chem.* **184**, 118143 (2025).
107. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
108. Gogia, P., Ashraf, H., Bhasin, S. & Xu, Y. Antibody–Drug Conjugates: A Review of Approved Drugs and Their Clinical Level of Evidence. *Cancers* **15**, 3886 (2023).
109. Gonzalez, R. *et al.* Screening the mammalian extracellular proteome for regulators of embryonic human stem cell pluripotency. *Proc. Natl. Acad. Sci.* **107**, 3552–3557 (2010).
110. Visscher, M., Arkin, M. R. & Dansen, T. B. Covalent targeting of acquired cysteines in cancer. *Curr. Opin. Chem. Biol.* **30**, 61–67 (2016).
111. Chen, L. *et al.* Pan-Cancer Analysis Reveals the Functional Importance of Protein Lysine Modification in Cancer Development. *Front. Genet.* **9**, (2018).

6 SUPPLEMENTARY FIGURES



Supplementary Figure 1: LC-MS Monitoring of *meta*- and *para*-substituted FKBP Ligands in high concentrations of TFA



Supplementary Figure 3: A) Plasmid for disulfide mutant B2M (Addgene). **B)** Plasmid for disulfide mutant MHC heavy chain (Addgene).

Supplementary Table 1: FGFR Peptide Affinity Predictions.

allele	Seq_num	start	end	length	peptide	ic50
HLA-A*02:01	1	5	14	10	PLGEGCFGQV	4327.54
HLA-A*02:01	1	11	19	9	FGQVVLAEA	4804.18
HLA-A*02:01	1	9	17	9	GCFGQVVLA	7108.13
HLA-A*02:01	1	5	15	11	PLGEGCFGQVV	11639.25
HLA-A*02:01	1	1	10	10	VLGKPLGEGC	11900.43
HLA-A*02:01	1	10	19	10	CFGQVVLAEA	14090.08
HLA-A*02:01	1	6	14	9	LGEGCFGQV	17475.6
HLA-A*02:01	1	9	19	11	GCFGQVVLAEA	18055.48
HLA-A*02:01	1	6	15	10	LGEGCFGQVV	19411.72
HLA-A*02:01	1	1	9	9	VLGKPLGEG	20085.58
HLA-A*02:01	1	7	16	10	GEGCFGQVVL	22245.53
HLA-A*02:01	1	12	19	8	GQVVLAEA	22527.95
HLA-A*02:01	1	8	17	10	EGCFGQVVLA	23110.36
HLA-A*02:01	1	4	14	11	KPLGEGCFGQV	23476.29
HLA-A*02:01	1	7	17	11	GEGCFGQVVLA	24193.68
HLA-A*02:01	1	4	13	10	KPLGEGCFGQ	24370.23
HLA-A*02:01	1	7	15	9	GEGCFGQVV	27152.62
HLA-A*02:01	1	1	11	11	VLGKPLGEGCF	27383.94
HLA-A*02:01	1	9	16	8	GCFGQVVL	27740
HLA-A*02:01	1	9	18	10	GCFGQVVLAE	28898.35
HLA-A*02:01	1	6	16	11	LGEGCFGQVVL	29129.72
HLA-A*02:01	1	8	16	9	EGCFGQVVL	29526.05
HLA-A*02:01	1	10	17	8	CFGQVVL	30057.89
HLA-A*02:01	1	5	13	9	PLGEGCFGQ	32378.34
HLA-A*02:01	1	1	8	8	VLGKPLGE	32586.06
HLA-A*02:01	1	3	13	11	GKPLGEGCFGQ	34092.91
HLA-A*02:01	1	10	18	9	CFGQVVLAE	35246.68
HLA-A*02:01	1	11	18	8	FGQVVLAE	38647.41
HLA-A*02:01	1	4	12	9	KPLGEGCFG	39208
HLA-A*02:01	1	8	15	8	EGCFGQVV	39356.35
HLA-A*02:01	1	2	10	9	LGKPLGEGC	39683.01
HLA-A*02:01	1	2	11	10	LGKPLGEGCF	40044.46
HLA-A*02:01	1	7	14	8	GEGCFGQV	40090.41
HLA-A*02:01	1	8	18	11	EGCFGQVVLAE	40649.95
HLA-A*02:01	1	6	13	8	LGEGCFGQ	41227.53
HLA-A*02:01	1	5	12	8	PLGEGCFG	41884.89
HLA-A*02:01	1	2	12	11	LGKPLGEGCFG	41970.18
HLA-A*02:01	1	3	12	10	GKPLGEGCFG	42630.11

HLA-A*02:01	1	2	9	8	LGKPLGEG	44063.27
HLA-A*02:01	1	4	11	8	KPLGEGCF	44072.31
HLA-A*02:01	1	3	11	9	GKPLGEGCF	44836.12
HLA-A*02:01	1	3	10	8	GKPLGEGC	47005.95