High-Throughput Protein Perturbation Screens for AI-Designed Degraders

Lin Zhao,[1] **Aastha Pal,**[1] **Tong Chen,**[2]
Pranam Chatterjee[1,2,†]

[1]Department of Bioengineering, University of Pennsylvania
[2]Department of Computer and Information Science, University of Pennsylvania
[†]Corresponding author: pranam@seas.upenn.edu

Abstract

The development of specific protein binders is crucial for biologics and targeted protein degradation (TPD) therapies. However, current techniques rely on low-throughput, labor-intensive approaches and often use non-human display systems such as phage, yeast, or mRNA display, limiting their translational relevance. To address this, we developed a high-throughput, human cell-based binder screening platform that enables the functional evaluation of artificial intelligence (AI)-designed peptide binders in a mammalian context for protein perturbation. Our approach utilizes genetically engineered, doxycycline-inducible ubiquibodies (uAbs) that fuse libraries of computationally designed "guide" peptides to an E3 ubiquitin ligase domain, enabling modular, CRISPR-like TPD. Using target-expressing cells fused with an mCherry reporter to monitor protein degradation, we screened and validated AI-generated binders in a physiologically relevant setting. We successfully applied this platform to identify functional peptide degraders for β-catenin, GFAP, and EWS::FLI1, oncogenic and disease-associated proteins that have remained challenging to target with conventional approaches. Overall, this platform enables scalable discovery of AI-designed peptide binders for proteome perturbation and therapeutic development.

1 Introduction

A substantial fraction of disease-driving protein, including transcription factors and fusion oncoproteins, remain intractable to small-molecule inhibition due to disordered domains and inaccessible binding pockets [Chen et al., 2023]. Biologics such as antibodies, nanobodies, and peptides offer a compelling alternative for functional protein perturbation, as they can bind diverse epitopes without requiring classical active sites Wang et al. [2022], Mieczkowski et al. [2023]. Advances in generative AI have further accelerated the design of programmable binders, enabling precise perturbation of both structured and disordered targets [Watson et al., 2023, Chen et al., 2025a, Pacesa et al., 2025].

A recent approach to proteome editing leverages targeted protein degradation (TPD), where short, sequence-designed peptide "guides" are fused to the ubiquitin conjugation domain of the human CHIP E3 ligase, yielding genetically encodable chimeras called "ubiquibodies" (uAbs) [Brixi et al., 2023, Bhat et al., 2025, Chen et al., 2025a]. These CRISPR-like tools act as programmable perturbation agents, degrading specific protein targets inside cells. The rise of generative peptide design models, including *PepMLM*, *PepPrCLIP*, *SaLT&PepPr*, *moPPIt*, *SOAPIA*, *PepTune*, and *MOG-DFM* enables rapid generation of such uAb binders directly from target sequences [Bhat et al., 2025, Chen et al., 2025a, Brixi et al., 2023, Chen et al., 2025b, Vincoff et al., 2025, Tang et al., 2025, Chen et al.,

The NeurIPS 2025 Workshop on AI Virtual Cells and Instruments: A New Era in Drug Discovery and Development (AI4D3 2025), San Diego, California, USA, 2025.

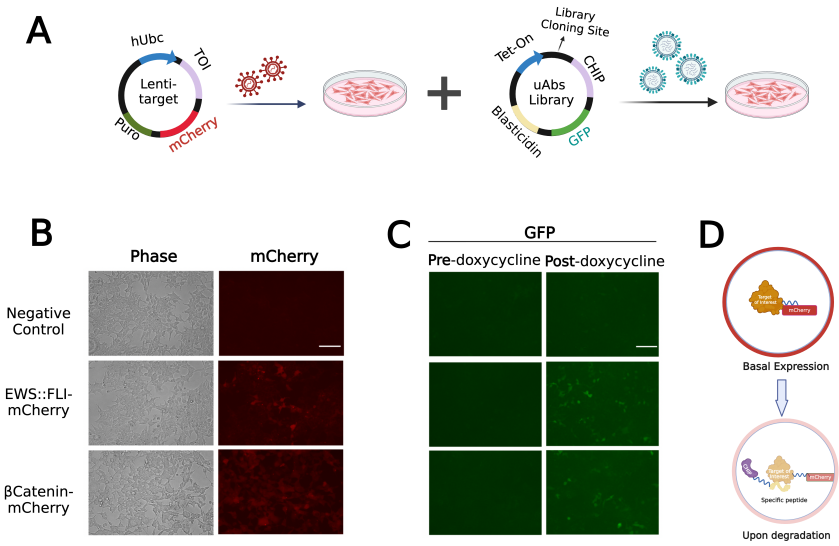

Figure 1: **Reporter cell line generation for screening of functional binder peptides.** A) Reporter cell line integration for both target-mCherry and uAb-GFP constructs. B) Target expression validation by mCherry fluorescence. C) uAb induction confirmed by GFP fluorescence following doxycycline addition. D) Schematic of uAb-mediated degradation in the target-mCherry reporter cell line upon peptide engagement.

2025c]. However, selecting and validating peptides that induce functional perturbation remains a key bottleneck.

In parallel, high-throughput CRISPR screening (CRISPR-HTS) has become a standard for scalable genomic perturbation using programmable nucleases such as Cas9 and dCas9 [Bock et al., 2022, Gilbert et al., 2014, Konermann et al., 2015]. These screens systematically assess perturbation outcomes by delivering sgRNA libraries to cells, applying selection pressure, and identifying enriched or depleted guides via deep sequencing [Shalem et al., 2015, Yang et al., 2023].

Motivated by this strategy, we establish a CRISPR-like high-throughput screening (HTS) platform for functional proteome perturbation using uAbs. We engineer HEK293T cells expressing mCherry-tagged target proteins and introduce a doxycycline-inducible uAb library, where each variant encodes a language model–designed peptide fused to the CHIPΔTPR ligase domain [Brixi et al., 2023, **?**, Bhat et al., 2025]. Functional peptide binders induce uAb-mediated degradation of the target, leading to loss of mCherry signal and enabling FACS-based selection of perturbed cells for sequencing. Using this screen, we successfully identified peptides that perturb disease-relevant proteins including the transcriptional coactivator β-catenin [Yu et al., 2021], the type III intermediate filament protein GFAP implicated in Alexander disease [Lin et al., 2024], and the fusion oncoprotein EWS::FLI1 [May et al., 1993]. Together, our platform enables scalable functional evaluation of AI-designed peptide perturbagens and establishes a foundation for CRISPR-like high-throughput proteome perturbation.

2 Methods

2.1 Binder Design

Novel binding peptides designed in this study were generated either by the target sequence-conditioned PepMLM algorithm [Chen et al., 2025a] (`https://huggingface.co/ChatterjeeLab/PepMLM-650M`), by the generative-discriminative PepPrCLIP model [Bhat et al., 2025] (`https://huggingface.co/ubiquitx/pepprclip`) by the motif-specific peptide generator moPPit-v2 [Chen et al., 2025b] algorithm (`https://huggingface.co/ChatterjeeLab/moPPIt-v2`).

2.2 Lentiviral production

For target-reporter packaging and pooled peptide packaging, HEK293T cells were seeded in a 6-well plate and transfected at approximately 50% confluency. For each well, 0.5 μg pMD2.G (Addgene #12259), 1.5 μg psPAX2 (Addgene #12260), and 0.5 μg of the target-mCherry reporter transfer vector were transfected with Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol. The medium was exchanged 8 hours post-transfection, and the viral supernatant was harvested at 48 and 72 hours post-transfection.

2.3 Target-mCherry reporter monoclonal cell line generation

For target-reporter cell line generation, 1×10^5 HEK293T cells were mixed with 20 μL of the concentrated virus in a 6-well plate. Media was changed 24 hours post-transduction. Antibiotic selection was started 36 hours post-transduction by adding 2 μg/mL puromycin (Sigma, P8833). Cells were harvested for sorting at 5 days post-antibiotic selection, and a single *mCherry*-positive cell was plated in a 96-well plate. Genomic PCR was performed after cell growth to validate the genotype of the monoclonal cell line.

2.4 Generation of plasmids

The uAb plasmid was generated from the standard pcDNA3 vector, harboring a cytomegalovirus (CMV) promoter and a C-terminal P2A-eGFP cassette, as well as a human phosphoglycerate kinase 1 (hPGK) promoter followed by a blasticidin selection cassette. An Esp3I restriction cloning site was introduced immediately upstream of the CHIPΔTPR CDS and a flexible GSGSG linker via the KLD Enzyme Mix (NEB) following PCR amplification with mutagenic primers (Genewiz).

2.5 Peptide library design and cloning

For each of the targets, the library included a set of polyG and polyA sequences as negative controls, along with 999 peptides generated using the PepMLM algorithm [Chen et al., 2025a] and 999 peptides generated using the moPPit algorithm [Chen et al., 2025b]. In total, 2000 peptides were synthesized by Twist as single-stranded DNA, PCR amplified, and cloned into the uAb architecture using Gibson assembly.

2.6 Peptide pooled lentiviral titration

The titer of the lentiviral peptide library pools for the single peptide libraries was determined by transducing 6×10^4 cells with serial dilutions of lentivirus and measuring GFP expression four days post-transduction using an Accuri C6 flow cytometer (BD). All lentiviral titrations were performed in the target-mCherry cell line.

2.7 Functional peptide binders screens

Each target-specific binder screen was performed in duplicate with independent transductions. For each replicate, 100 ng/μL doxycycline was added to induce uAb expression. Cells were harvested 5 days post-doxycycline treatment for sorting. A total of 1×10^7 target-mCherry cells were washed once with 1$\times$ PBS, dissociated using Accutase, filtered through a 30 μm CellTrics filter, and resuspended in FACS buffer (0.5% BSA [Sigma, A7906], 2 mM EDTA [Sigma, E7889] in PBS). The highest and lowest 10% of cells were sorted based on mCherry expression using a SH800 FACS Cell Sorter (Sony Biotechnology). After sorting, genomic DNA was extracted with the Monarch Genomic DNA Purification Kit (NEB, T3010).

2.8 Peptide library sequencing

The peptide libraries were amplified from each genomic DNA sample in 100 μL PCR reactions using Q5 Hot Start Polymerase (NEB, M0493) with 1 μg of genomic DNA per reaction. PCR amplification was performed according to the manufacturer's instructions, using 25 cycles at an annealing temperature of 60 °C with the following primers:

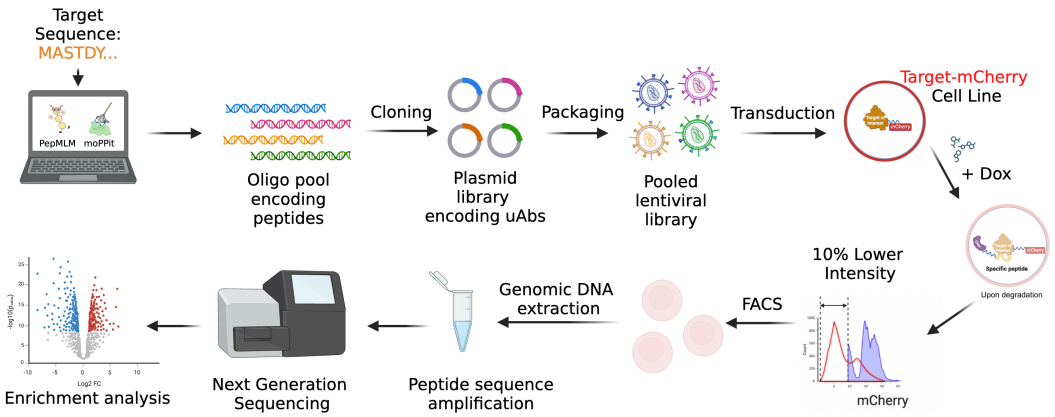

Figure 2: **High-throughput screening of AI-designed peptide-guided uAbs.** Schematic workflow from computational peptide design, cloning, lentiviral packaging, pooled transduction, FACS isolation of mCherry-low cells, and sequencing-based enrichment analysis.

```
Fwd: ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGGCTAGCGGATCCGCCACCATG
Rev: GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGTTCAGCCGGCCAGAACCGCTGCC
```

The amplified libraries were purified with the QIAquick Gel Extraction Kit and the amplicons were sent to Genewiz for next-generation sequencing.

2.9 Data analysis

We used MAGeCK and MAGeCK-VISPR to analyze peptide enrichment across experimental conditions [Li et al., 2015]. The pipeline normalizes sequencing counts, models read distributions across replicates, and identifies significantly enriched peptides by comparing post-sort and pre-sort populations. For each peptide, MAGeCK reports the $\log_2$ fold change ($\log_2$fc), reflecting relative enrichment, and an associated p-value indicating statistical significance. To visualize these results, we generated volcano plots showing both the magnitude of effect ($\log_2$fc) and statistical significance ($-\log_{10} p$-value) for each peptide. Peptides meeting the thresholds of $\log_2\text{fc} > 2$ and $p < 0.01$ were considered significantly enriched (red), while all others were designated as non-significant (gray).

2.10 Cell fractionation and immunoblotting

Cells were harvested, detached with 0.05% trypsin-EDTA, and washed twice with ice-cold PBS. Pellets were lysed in RIPA buffer (ThermoFisher, Cat #89900) supplemented with protease inhibitor cocktail (Millipore Sigma, Cat #P8340, 1:100), incubated at 4 °C for 30 min, and centrifuged at 15,000 rpm for 10 min. Protein concentration was determined using the Pierce BCA Assay Kit (ThermoFisher, Cat #23227).

Twenty micrograms of protein were mixed with 4× LDS sample buffer containing 5% β-mercaptoethanol (3:1), denatured at 95 °C for 10 min, separated on Bolt™ Bis-Tris Plus gels (ThermoFisher, Cat #NW04125BOX), and transferred using iBlot™ 2 stacks. Membranes were blocked in 5% milk-TBST and probed overnight at 4 °C with mouse anti-beta-Catenin (R&D Systems, Cat #MAB13292, 1:1000) and anti-GAPDH (Santa Cruz, Cat #sc-47724, 1:10,000), followed by goat anti-mouse HRP secondary (ThermoFisher, Cat #32230, 1:5000). Detection was performed by chemiluminescence (BioRad ChemiDoc™ Touch). Band intensities were quantified using FIJI gel analysis, normalized to GAPDH, and expressed relative to polyG uAb vector controls.

2.11 Real-time quantitative RT-PCR

Total RNA was isolated using a Monarch Total RNA Miniprep Kit (NEB, T2010). One microgram of total RNA was reverse transcribed using PrimeScript™ RT Master Mix (Takara Bio, RR036A).

Quantitative real-time PCR (qPCR) analysis using the fluorescent SYBR Green method (Bio-Rad, Richmond, CA) was performed according to the manufacturer's instructions.

The primers used were as follows:

- EWSAT1: 5′-TGTGCATTCCGCATCCAGGTGT-3′ (forward) and 5′-GCTTCAGCAGAGATGTTGCAGG-3′ (reverse)
- ACTINB: 5′-CACCATTGGCAATGAGCGGTTC-3′ (forward) and 5′-AGGTCTTTGCGGATGTCCACGT-3′ (reverse)

The PCR program consisted of enzyme activation at 95 °C for 10 min, followed by 40 amplification cycles (95 °C for 30 sec, 61 °C for 30 sec, 72 °C for 30 sec). Data were acquired using an iCycler iQ Real-Time PCR Detection System (Bio-Rad) and normalized against ACTINB.

2.12 TOPFlash Luciferase assay

DLD1 cells were seeded in 96-well plates at a density of 1×10^4 cells per well and incubated for 20–24 h prior to transfection. On the day of transfection, each well received a total of 100 ng plasmid DNA at a ratio of TOPFlash luciferase:uAb plasmids of 1:3. The plasmid DNA was mixed with Lipofectamine 2000 (Invitrogen) in serum-free Opti-MEM (Gibco) and incubated for 20 min at room temperature before being added dropwise to each well. After 72 h of incubation, the Dual-Luciferase® Reporter Assay (Promega) was performed according to the manufacturer's protocol.

2.13 Cell apoptosis assay

TC71 cells were seeded in opaque 96-well plates at a density of 1×10^4 cells per well and incubated for 20–24 h prior to transfection. On the day of transfection, each well received 100 ng of uAb plasmid DNA mixed with Lipofectamine 2000 (Invitrogen) in serum-free Opti-MEM (Gibco). The mixture was incubated for 20 min at room temperature and added dropwise to each well. After 72 h of incubation, 20 μL of the Cytotoxicity Reagent from the ApoTox-Glo™ Triplex Assay (Promega) was added directly to each well and incubated for 30 min at 37 °C, protected from light. Luminescence was measured using a microplate reader to quantify caspase-3/7 activity as an indicator of apoptosis. All measurements were performed in triplicate, and results were normalized to untreated control wells.

3 Results

3.1 Generation of a dual reporter cell line for functional peptide screening

To establish a mammalian screening platform, we engineered a dual-reporter system (Figure 1A). Coding sequences of target proteins of interest were cloned into lentiviral vectors fused to an mCherry reporter and linked to a puromycin resistance cassette for selection. Following antibiotic selection, we derived monoclonal target-mCherry reporter cell lines to ensure stable and uniform expression of each target protein (Figure 1B).

Target-fused mCherry HEK293T cell lines were maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 100 U/mL penicillin, 100 μg/mL streptomycin, and 10% fetal bovine serum (FBS). For lentiviral peptide library integration selection, 8 μg/mL blasticidin was added 72 hours post-transduction.

To enable pooled screening of peptide-guided ubiquibodies (uAbs), we introduced an oligonucleotide library into the uAb vector architecture. This construct is driven by a doxycycline-inducible Tet-on promoter and included a GFP reporter for real-time monitoring of uAb induction. A blasticidin resistance cassette was incorporated to enforce dual-selection of both target-mCherry and uAb-GFP constructs. As can be observed in Figure 1C, doxycycline supplementation induces uAb expression, visualized by GFP fluorescence. As such, functional peptide binders will trigger uAb-mediated ubiquitination and degradation of the mCherry-tagged target protein, detectable as a loss of red fluorescence (Figure 1D).

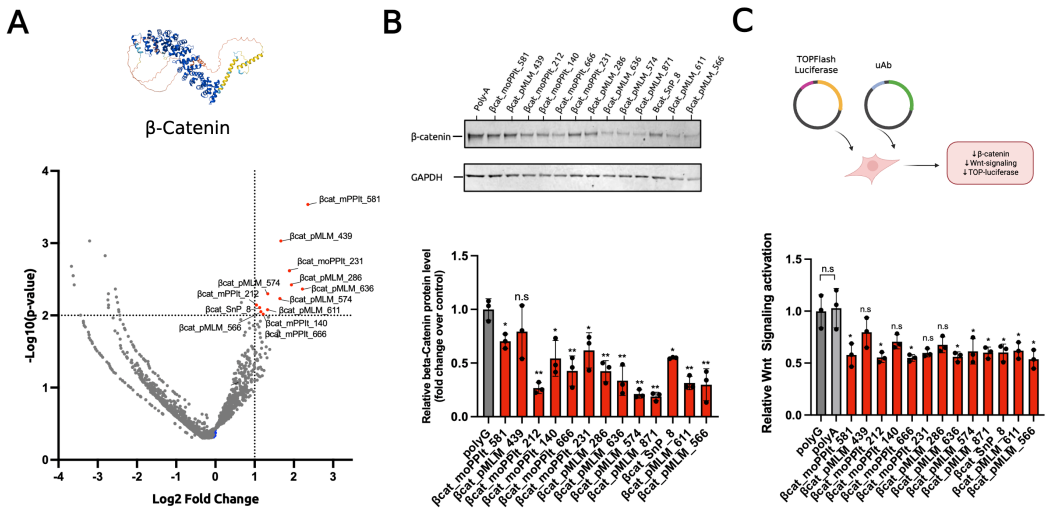

Figure 3: **High-throughput screening and validation of functional peptide binders for β-catenin.** (A) Volcano plot showing enriched peptides in the β-catenin–mCherry reporter line. Red dots indicate peptides significantly enriched ($p < 0.01$, $\log_2$fc > 2). (B) Immunoblot confirming degradation of endogenous β-catenin in DLD1 cells. (C) TOPFlash luciferase assay quantifying Wnt signaling in DLD1 cells. Degradation of β-catenin reduces Wnt-driven transcriptional activity. Data are shown as mean $\pm$ s.d. from three independent replicates ($n = 3$).

3.2 High-throughput screening identifies target-specific peptide degraders

We implemented our pooled high-throughput functional screen to identify peptide binders for three oncogenic targets: (i) β-catenin, a transcriptional coactivator whose dysregulation drives tumorigenesis [Yu et al., 2021]; (ii) GFAP, a type III intermediate filament protein whose abnormal accumulation causes Alexander disease [Lin et al., 2024]; and (iii) the fusion oncoprotein EWS::FLI1, a driver of Ewing sarcoma [Braun et al., 1995]. A total of 2000 peptides were designed using either PepMLM [Chen et al., 2025a], PepPrCLIP Bhat et al. [2025], or moPPIt-v2 Chen et al. [2025b], synthesized as an oligonucleotide pool, cloned into the uAb vector, packaged into lentivirus, and transduced into target–mCherry reporter cell lines. Cells expressing uAbs that bind and degrade their targets exhibit reduced mCherry fluorescence. As a standard protocol, we isolated the lowest 10% of the population based on mCherry intensity, followed by genomic DNA extraction and next-generation sequencing (NGS) of peptide-encoding sequences (Figure 2).

To validate the assay, we first performed a pilot screen targeting β-catenin. Peptide frequencies in post-sort and pre-sort populations were compared using the MAGeCK pipeline [Li et al., 2015], which normalizes read counts and calculates differential enrichment statistics. Peptides were ranked by $\log_2$fc (treatment versus control) and associated p-values. Positively enriched peptides (red) and non-enriched peptides (gray) are shown in the volcano plot for β-catenin (Figure 3A). Immunoblotting confirmed that several uAbs reduced endogenous β-catenin levels in DLD1 colorectal cancer cells (Figure 3B). We further corroborated these results by co-transfecting the top β-catenin–targeting uAb constructs into DLD1 cells and quantifying Wnt signaling using a TOPFlash luciferase reporter assay [Korinek et al., 1997]. Cells expressing these uAbs exhibited significantly reduced Wnt-dependent luciferase activity compared to polyG controls (Figure 3C).

Encouraged by the observed β-catenin degradation, we next examined GFAP, a type III intermediate filament protein whose abnormal accumulation causes Alexander disease [Lin et al., 2024]. Twenty peptide binders displayed significant enrichment in the volcano plot (Figure 4A). Immunoblotting further demonstrated that several GFAP-targeting uAbs decreased endogenous GFAP levels in U251 glioblastoma cells (Figure 4B). Ongoing studies aim to evaluate the downstream cellular effects of GFAP degradation, including impacts on cell proliferation and filament organization.

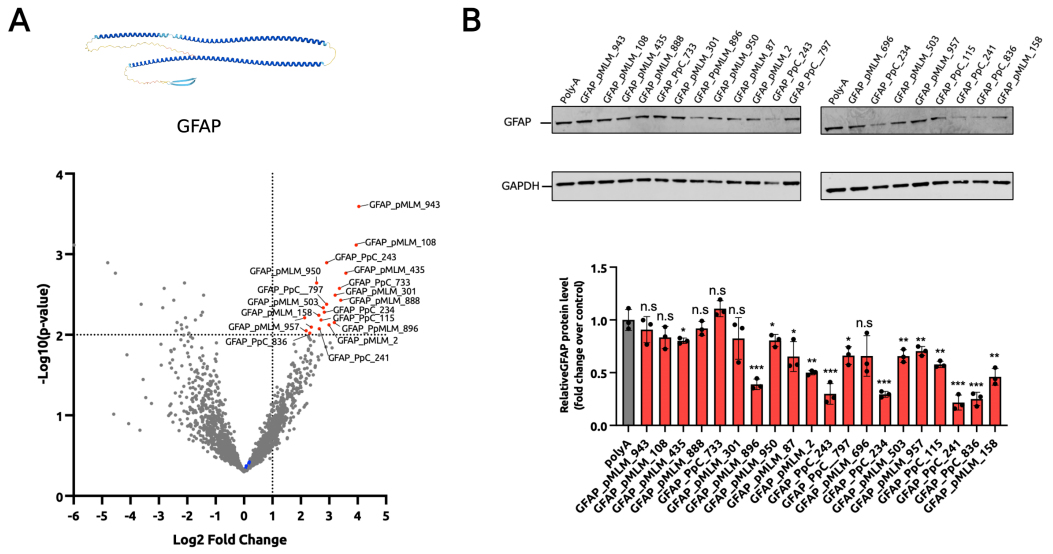

Figure 4: **High-throughput screening and validation of functional peptide binders for GFAP.** (A) Volcano plot of enriched peptides in the GFAP–mCherry reporter line. Red dots indicate peptides significantly enriched ($p < 0.01$, $\log_2$ fold change > 2). (B) Immunoblot confirming degradation of endogenous GFAP in U251 cells.

Finally, we sought to target the oncogenic fusion protein EWS::FLI1, a key driver of Ewing sarcoma [Braun et al., 1995]. Sixteen enriched peptide binders were selected for validation based on the volcano plot results (Figure 5A). The corresponding uAb constructs were co-transfected into TC71 cells, a human Ewing sarcoma line commonly used in EWS::FLI1 studies [Guan et al., 2005]. Quantitative RT–PCR analysis revealed that 13 of the 16 uAbs significantly downregulated *EWSAT1*, a canonical transcriptional target of EWS::FLI1 [Howarth et al., 2014], consistent with increased apoptosis observed in functional assays (Figure 5C). Together, these findings demonstrate that our pooled screening approach reliably identifies functional peptide degraders across both cytoskeletal and nuclear oncogenic targets.

4 Discussion

The CRISPR-Cas system has revolutionized genome editing due to its modular and programmable nature. The ability to design an sgRNA for a new genomic sequence enables precise Cas-mediated editing. Beyond gene disruption, CRISPR has been adapted for single-base editing, gene insertion, and transcriptional modulation [Pacesa et al., 2024, Wang and Doudna, 2023]. Inspired by this RNA-guided system, our peptide-guided uAb degraders are genetically encodable, allowing integration into a CRISPR-like high-throughput screening platform to identify functional peptide binders.

In this study, we established a mammalian high-throughput screening (HTS) platform for identifying peptide binders of targets of interest. The results showed that enriched candidates induced significant degradation of β-catenin and functional perturbation of EWS::FLI1. Beyond target protein degradation, these results motivated us to extend the platform toward target protein stabilization (TPS) via utilizing engineered dual antibodies [Hong et al., 2025], which can enable proteome-wide modulation and support combinatorial screening of protein activation and inhibition, analogous to CRISPRa and CRISPRi approaches for genetic perturbation Gilbert et al. [2014]. With recent advances in protein language model (pLM) architectures, such as SOAPIA for fusion-driven oncoprotein targeting Vincoff et al. [2025]and PTM-Mamba for post-translational modification (PTM)-aware protein design Peng et al. [2025], our platform can be further applied to previously undruggable oncofusion proteins. Additionally, by integrating diverse PTM domains, including kinases, phosphatases, acetylases, and deglycosylases, our HTS system can potentially facilitate targeted editing of protein PTMs. Overall, we anticipate that this approach will accelerate the development of therapeutic modalities based

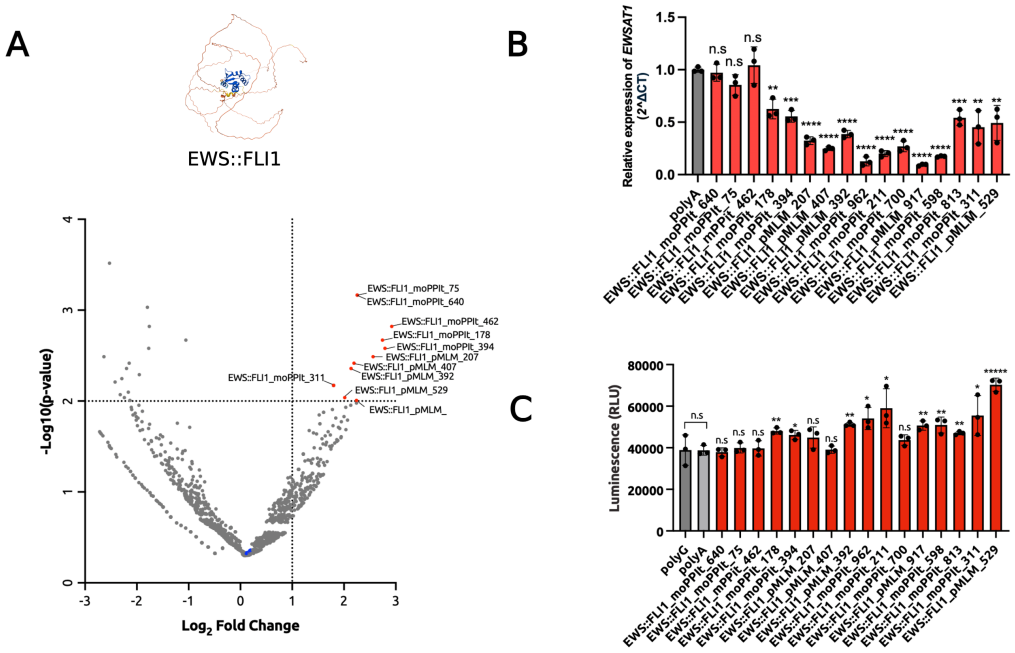

Figure 5: **High-throughput screening and validation of functional peptide binders for EWS::FLI1.** (A) Volcano plot of enriched peptides in the EWS::FLI1–mCherry reporter line. Red dots indicate peptides significantly enriched ($p < 0.01$, $\log_2$ fold change > 2). (B) RT–PCR showing downregulation of *EWSAT1* in TC71 cells following transfection with enriched EWS::FLI1 uAbs. (C) Cell apoptosis assay confirming the functional impact of EWS::FLI1 degradation.

on selective target binders and broaden the scope for designing novel interventions across multiple disease contexts.

References

Tianlai Chen, Lauren Hong, Vivian Yudistyra, Sophia Vincoff, and Pranam Chatterjee. Generative design of therapeutics that bind and modulate protein states. *Current Opinion in Biomedical Engineering*, 28:100496, 2023.

Lei Wang, Nanxi Wang, Wenping Zhang, Xurui Cheng, Zhibin Yan, Gang Shao, Xi Wang, Rui Wang, and Caiyun Fu. Therapeutic peptides: current applications and future directions. *Signal Transduction and Targeted Therapy*, 7(1), February 2022. ISSN 2059-3635. doi: 10.1038/s41392-022-00904-4. URL `http://dx.doi.org/10.1038/s41392-022-00904-4`.

Carl Mieczkowski, Xuejin Zhang, Dana Lee, Khanh Nguyen, Wei Lv, Yanling Wang, Yue Zhang, Jackie Way, and Jean-Michel Gries. Blueprint for antibody biologics developability. *mAbs*, 15(1), March 2023. ISSN 1942-0870. doi: 10.1080/19420862.2023.2185924. URL `http://dx.doi.org/10.1080/19420862.2023.2185924`.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Leo Tianlai Chen, Zachary Quinn, Madeleine Dumas, Christina Peng, Lauren Hong, Moises Lopez-Gonzalez, Alexander Mestre, Rio Watson, Sophia Vincoff, Lin Zhao, et al. Target sequence-conditioned design of peptide binders using masked language modeling. *Nature Biotechnology*, pages 1–9, 2025a.

Martin Pacesa, Lennart Nickel, Christian Schellhaas, Joseph Schmidt, Ekaterina Pyatova, Lucas Kissling, Patrick Barendse, Jagrity Choudhury, Srajan Kapoor, Ana Alcaraz-Serna, Yehlin Cho,

Kourosh H. Ghamary, Laura Vinué, Brahm J. Yachnin, Andrew M. Wollacott, Stephen Buckley, Adrie H. Westphal, Simon Lindhoud, Sandrine Georgeon, Casper A. Goverde, Georgios N. Hatzopoulos, Pierre Gönczy, Yannick D. Muller, Gerald Schwank, Daan C. Swarts, Alex J. Vecchio, Bernard L. Schneider, Sergey Ovchinnikov, and Bruno E. Correia. Bindcraft: one-shot design of functional protein binders. *Nature*, October 2025.

Garyk Brixi, Tianzheng Ye, Lauren Hong, Tian Wang, Connor Monticello, Natalia Lopez-Barbosa, Sophia Vincoff, Vivian Yudistyra, Lin Zhao, Elena Haarer, et al. Salt&peppr is an interface-predicting language model for designing peptide-guided protein degraders. *Communications Biology*, 6(1):1081, 2023.

Suhaas Bhat, Kalyan Palepu, Lauren Hong, Joey Mao, Tianzheng Ye, Rema Iyer, Lin Zhao, Tianlai Chen, Sophia Vincoff, Rio Watson, et al. De novo design of peptide binders to conformationally diverse targets with contrastive language modeling. *Science Advances*, 11(4):eadr8638, 2025.

Tong Chen, Yinuo Zhang, Zachary Quinn, and Pranam Chatterjee. moppit: De novo generation of motif-specific peptide binders via conditional uniform discrete diffusion. In *Learning Meaningful Representations of Life (LMRL) Workshop at ICLR 2025*, 2025b.

Sophia Vincoff, Oscar Davis, Alexander Tong, Joey Bose, and Pranam Chatterjee. Soapi: Siamese-guided generation of off-target-avoiding protein interactions. In *Learning Meaningful Representations of Life (LMRL) Workshop at ICLR 2025*, 2025.

Sophia Tang, Yinuo Zhang, and Pranam Chatterjee. Peptune: De novo generation of therapeutic peptides with multi-objective-guided discrete diffusion. In *Forty-second International Conference on Machine Learning*, 2025. URL `https://openreview.net/forum?id=FQoy1Y1Hd8`.

Tong Chen, Yinuo Zhang, Sophia Tang, and Pranam Chatterjee. Multi-objective-guided discrete flow matching for controllable biological sequence design. In *ICML 2025 Generative AI and Biology (GenBio) Workshop*, 2025c. URL `https://openreview.net/forum?id=8YIMLoHP9J`.

Christoph Bock, Paul Datlinger, Florence Chardon, Matthew A Coelho, Matthew B Dong, Keith A Lawson, Tian Lu, Laetitia Maroc, Thomas M Norman, Bicna Song, et al. High-content crispr screening. *Nature Reviews Methods Primers*, 2(1):8, 2022.

Luke A Gilbert, Max A Horlbeck, Britt Adamson, Jacqueline E Villalta, Yuwen Chen, Evan H Whitehead, Carla Guimaraes, Barbara Panning, Hidde L Ploegh, Michael C Bassik, et al. Genome-scale crispr-mediated control of gene repression and activation. *Cell*, 159(3):647–661, 2014.

Silvana Konermann, Mark D Brigham, Alexandro E Trevino, Julia Joung, Omar O Abudayyeh, Clea Barcena, Patrick D Hsu, Naomi Habib, Jonathan S Gootenberg, Hiroshi Nishimasu, et al. Genome-scale transcriptional activation by an engineered crispr-cas9 complex. *Nature*, 517(7536): 583–588, 2015.

Ophir Shalem, Neville E Sanjana, and Feng Zhang. High-throughput functional genomics using crispr–cas9. *Nature Reviews Genetics*, 16(5):299–311, 2015.

Xin Yang, Shoufu Duan, Zhiming Li, Zhe Wang, Ning Kon, Zhiguo Zhang, and Wei Gu. Protocol of crispr-cas9 knockout screens for identifying ferroptosis regulators. *STAR protocols*, 4(4):102762, 2023.

Fanyuan Yu, Changhao Yu, Feifei Li, Yanqin Zuo, Yitian Wang, Lin Yao, Chenzhou Wu, Chenglin Wang, and Ling Ye. Wnt/β-catenin signaling in cancers and targeted therapies. *Signal Transduction and Targeted Therapy*, 6(1):307, 2021.

Ni-Hsuan Lin, Wan-Syuan Jian, Natasha Snider, and Ming-Der Perng. Glial fibrillary acidic protein is pathologically modified in alexander disease. *Journal of Biological Chemistry*, 300(7), 2024.

William A May, Stephen L Lessnick, Benjamin S Braun, Micheal Klemsz, Brain C Lewis, Lynn B Lunsford, Robert Hromas, and Christopher T Denny. The ewing's sarcoma ews/fli-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than fli-1. *Molecular and cellular biology*, 13(12):7393–7398, 1993.

Wei Li, Johannes Köster, Han Xu, Chen-Hao Chen, Tengfei Xiao, Jun S Liu, Myles Brown, and X Shirley Liu. Quality control, modeling, and visualization of crispr screens with mageck-vispr. *Genome biology*, 16(1):281, 2015.

Benjamin S Braun, Richard Frieden, Stephen L Lessnick, William A May, and Christopher T Denny. Identification of target genes for the ewing's sarcoma ews/fli fusion protein by representational difference analysis. *Molecular and Cellular Biology*, 15(8):4623–4630, 1995.

Vladimir Korinek, Nick Barker, Patrice J Morin, Dick Van Wichen, Roel De Weger, Kenneth W Kinzler, Bert Vogelstein, and Hans Clevers. Constitutive transcriptional activation by a β-catenin-tcf complex in apc-/- colon carcinoma. *Science*, 275(5307):1784–1787, 1997.

Hui Guan, Zhichao Zhou, Hua Wang, Shu-Fang Jia, Wenbiao Liu, and Eugenie S. Kleinerman. A small interfering rna targeting vascular endothelial growth factor inhibits ewing's sarcoma growth in a xenograft mouse model. *Clinical Cancer Research*, 11(7):2662–2669, April 2005. ISSN 1557-3265. doi: 10.1158/1078-0432.ccr-04-1206. URL `http://dx.doi.org/10.1158/1078-0432.CCR-04-1206`.

Michelle Marques Howarth, David Simpson, Siu P Ngok, Bethsaida Nieves, Ron Chen, Zurab Siprashvili, Dedeepya Vaka, Marcus R Breese, Brian D Crompton, Gabriela Alexe, et al. Long noncoding rna ewsat1-mediated gene repression facilitates ewing sarcoma oncogenesis. *The Journal of clinical investigation*, 124(12):5275–5290, 2014.

Martin Pacesa, Oana Pelea, and Martin Jinek. Past, present, and future of crispr genome editing technologies. *Cell*, 187(5):1076–1100, 2024.

Joy Y Wang and Jennifer A Doudna. Crispr technology: A decade of genome editing is only the beginning. *Science*, 379(6629):eadd8643, 2023.

Lauren Hong, Tianzheng Ye, Tian Z Wang, Divya Srijay, Howard Liu, Lin Zhao, Rio Watson, Sophia Vincoff, Tianlai Chen, Kseniia Kholina, et al. Programmable protein stabilization with language model-derived peptide guides. *Nature Communications*, 16(1):3555, 2025.

Fred Zhangzhi Peng, Chentong Wang, Tong Chen, Benjamin Schussheim, Sophia Vincoff, and Pranam Chatterjee. Ptm-mamba: A ptm-aware protein language model with bidirectional gated mamba blocks. *Nature Methods*, 22(5):945–949, 2025. doi: 10.1038/s41592-025-02656-9.

Table 1: Summary of translated amino acid sequences with associated p-values and $\log_2$fc.

Name	**Amino Acid Sequence**	p**-value**	$\log_2$ **FC**
βcat_moPPIt_581	DAC EQVQLCQRQLAEQAK	0.00029121	2.3573
βcat_pMLM_439	SPEGSRDALADNLLVAVLL	0.00093245	1.6682
βcat_moPPIt_212	DKIQEQLNASRLANLAELQAK	0.0024091	1.8876
βcat_moPPIt_140	HKYREQVTAAPNERLARLQAQ	0.0024091	1.8876
βcat_moPPIt_666	DNARKERLSASRQEELAA LQG K	0.0024091	1.8876
βcat_moPPIt_231	AQYQEQLTAERNALAEEQA K	0.0024091	1.8876
βcat_pMLM_286	EPVVRQMSSSLLLDLLVLLQ	0.0037798	1.9364
βcat_pMLM_636	DPEVRQLASDDQLNAVVVLK	0.0043269	2.2151
βcat_pMLM_574	SPRVEEVSSEDLDLDAVVVAQ	0.0050152	1.3357
βcat_pMLM_871	DPVRVEMLAEDLLDLLVLL	0.0058447	1.6440
βcat_SnP_8	PTAPPYDSLLVFDYEGGSGS	0.0078037	1.1268
βcat_pMLM_611	REVQVRMSAEDDLDLAVLVK	0.0083450	1.3289
βcat_pMLM_566	DEVQREMSELQDNDVVVLL	0.0088921	1.1565
GFAP_pMLM_943	VKRGEEPLAIIIKKYLLLLQ	0.00025521	4.0392
GFAP_pMLM_108	KSRREEDLSSLITLDDEILA	0.00077057	3.9489
GFAP_pMLM_435	KKVRVEDLIILLKYYQMQAL	0.0012760	3.5840
GFAP_pMLM_888	KVRRTEPAASLYKYDLMLLA	0.0017171	3.4014
GFAP_PpC_733	GGNKVEASCRNLQKEGVLAL	0.0022721	3.3602
GFAP_pMLM_301	FKVGTLPLSRLITYLLELAQ	0.0026536	3.2069
GFAP_pMLM_896	KVRRVEPLSSLYYYYDMILL	0.0032235	3.1774
GFAP_pMLM_950	KVRRTRRLSSIITYLLMIQQ	0.0037240	3.1012
GFAP_pMLM_87	KKLGVRPLSSLIKYDQMILA	0.0041799	3.0304
GFAP_pMLM_2	KKLGVLDLSSLYTLLDELAQ	0.0046259	2.9916
GFAP_PpC_243	RVWNEIEEQMHQKRNKMVWD	0.0052404	2.9076
GFAP_PpC_797	MNTKAIAAELLKIIVAERRR	0.0057260	2.9060
GFAP_pMLM_696	KVRGEEPLSILIKYYQEILA	0.0061472	2.8833
GFAP_PpC_234	WLEKRDMCEEVEKKIKEEEE	0.0065585	2.8256
GFAP_pMLM_503	FSLVEEDLSRLYKKLDEQAA	0.0070392	2.7915
GFAP_pMLM_957	KVVVTLPLISIIKYLLEQLA	0.0075348	2.7307
GFAP_PpC_115	KKEIREWRKRRAFWAWWVMN	0.0080204	2.7122
GFAP_PpC_241	EQIKTGELRRIRGEIGEKWD	0.0084317	2.6451
GFAP_PpC_836	WDERKRQKGMWWLYWGIKID	0.0089223	2.6450
GFAP_pMLM_158	FSLVVERLSIILTYLQELLA	0.0095417	2.6327
EWS::FLI1_moPPIt_640	HPPPPKPSMSDYYYHDDWHY	0.00068272	2.2562
EWS::FLI1_moPPIt_75	CPPPPKEEDEYYDKHWDNHV	0.00068272	2.2562
EWS::FLI1_moPPIt_462	NPLPQQSPSDSYTQQQAQQQ	0.0015068	2.9209
EWS::FLI1_moPPIt_178	WPPKPKEKDEYYDKHRDNHW	0.0021384	2.7441
EWS::FLI1_moPPIt_394	NPPTPKEKSAWYDYSWDNWV	0.0026256	2.7924
EWS::FLI1_pMLM_207	PMPPLIAAADYYTAQAAAQQ	0.0032572	2.5615
EWS::FLI1_pMLM_407	NPQPQIRPSDYIRDNLAQQS	0.0038347	2.1937
EWS::FLI1_pMLM_392	PPQPQKRLSESITQQLALAQ	0.0043881	2.1367
EWS::FLI1_moPPIt_962	NPPPPKETSNYGDYHWDNWV	0.0067220	1.7985
EWS::FLI1_moPPIt_211	VPPPPKEKDCYYDYHNDNWV	0.0067220	1.7985
EWS::FLI1_moPPIt_700	VPPPPKEKCNYYDNGWDNWY	0.0067220	1.7985
EWS::FLI1_pMLM_917	NPPPQKRASSEISDLQQAQQ	0.0067220	1.7985
EWS::FLI1_moPPIt_598	NPPPPKTKDEYYDHHWRNCD	0.0067220	1.7985
EWS::FLI1_moPPIt_813	YPPPPKEKDEYYDWFGSNWY	0.0067220	1.7985
EWS::FLI1_moPPIt_311	VPPPPKCKDEYHEKHWDNWV	0.0067220	1.7985
EWS::FLI1_pMLM_529	DVLVLKRAADEGRAQLQQQQ	0.0091521	2.0155