Journal of proteome research

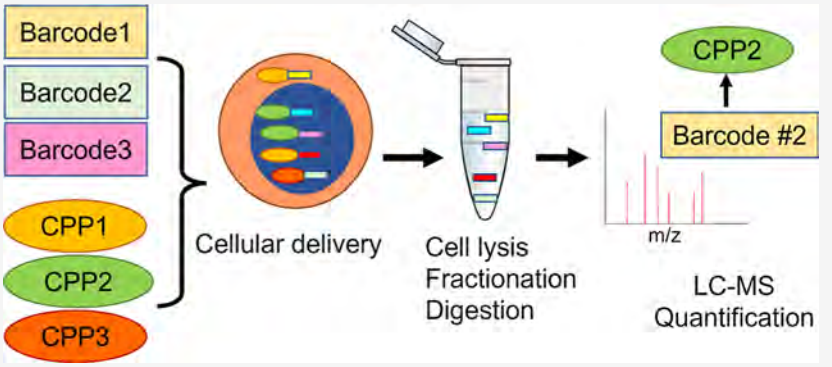
pubs.acs.org/jpr Article

Proteomic Barcoding Platform for Macromolecular Screening and Delivery

Ning Wang, Nicole A. Mcneer, Elliot Eton, Josh Fass, and Alex Kentsis*

Cite This: https://doi.org/10.1021/acs.jproteome.4c00068 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Engineered macromolecules offer compelling means for the therapy of conventionally undruggable interactions in human disease. However, their efficacy is limited by barriers to tissue and intracellular delivery. Inspired by recent advances in molecular barcoding and evolution, we developed BarcodeBabel, a generalized method for the design of libraries of peptide barcodes suitable for high-throughput mass spectrometry proteomics. Combined with PeptideBabel, a Monte Carlo sampling algorithm for the design of peptides with evolvable physicochemical properties and sequence complexity, we developed a barcoded library of cell penetrating peptides (CPPs) with distinct physicochemical features. Using quantitative targeted mass spectrometry, we identified CPPS with improved nuclear and cytoplasmic delivery exceeding hundreds of millions of molecules per human cell while maintaining minimal membrane disruption and negligible toxicity in vitro. These studies provide a proof of concept for peptide barcoding as a homogeneous high-throughput method for macromolecular screening and delivery. BarcodeBabel and PeptideBabel are available open-source from https://github.com/kentsisresearchgroup/.

KEYWORDS: *peptide barcoding, mass spectrometry proteomics, cell penetration, molecular screening, macromolecular drug delivery*

INTRODUCTION

Macromolecular drugs have emerged as a promising category of treatment for severe diseases, including neurodegenerative disorders and cancers, both as a complement and an alternative to small molecules. Macromolecules in general, and polypeptides specifically, offer many advantages to small-molecule therapeutics due to greater selectivity and specificity and are thus able to target otherwise undruggable factors such as protein−protein interactions.[1−5]

However, the relatively large size and polar features of many biological macromolecules render them impermeable to cellular lipid membranes. The recognition of naturally occurring moieties that allow for membrane penetration has led to the development of various cell penetrating peptides (CPPs) and protein transduction domains (PTDs) to assist with drug delivery, which are thought to work by membrane pore formation and other interfacial processes, endocytosis with or without involvement of transmembrane proteins, and translocation via inverted micelles followed by their intracellular dissolution.[6−10] However, the diversity and complexity of these mechanisms have been a major barrier to the development of explicit structure−activity relationships, which are necessary for the ultimate design and development of efficient, selective, and safe CPPs and PTDs for macromolecular drug delivery.[11,12] An efficient and robust method to quantify the subcellular localization of CPPs is therefore highly desired.

Recent advances in molecular evolution and high-throughput screening methods permit the identification of molecules with specific biological activities. In particular, Hoffmann et al. and Kauffmann et al. applied phage and RNA display and molecular evolution techniques, respectively, to develop improved

Received: February 2, 2024
Revised: April 5, 2024
Accepted: May 15, 2024

CPPs.[13,14] Similarly, recent advances in mass spectrometry (MS) proteomics permit quantitative and sensitive measurements of diverse molecules in complex biological samples. Egloff et al. used genetically encoded peptide barcodes for measuring binding of libraries of engineered proteins.[15] Peptide barcoding has also been implemented recently for diversity-oriented small molecule screening, demonstrating its increased stability and information coding capacity as compared to that of nucleic acid barcoding.[16] In principle, it is possible to track CPPs and other bioactive peptides directly without barcoding, but this is often challenging and, in some cases, like the cationic polyarginine TAT CPP, is not possible due to the lack of specific signature ions.

Inspired by these distinct approaches, we sought to develop peptide barcoding as a homogeneous method for macromolecular screening and delivery. Implemented in open-source algorithms BarcodeBabel and PeptideBabel, this approach permits the design of libraries of peptide barcodes and novel CPPs optimized for screening and analysis by quantitative MS. With this approach, we identified CPPs with improved nuclear and cytoplasmic delivery exceeding hundreds of millions of molecules per human cell while maintaining minimal membrane disruption and negligible toxicity in vitro. These methods are useful for a wide variety of molecular evolution and screening applications.

MATERIALS AND METHODS

Reagents

The Fmoc amino acids OxymaPure and Rink Amide ProTide Resin used for solid-phase peptide synthesis were purchased from CEM Corporation (Charlotte, NC, USA). *N*,*N*′-Diisopropylcarbodiimide (DIC), 2,2′-(ethylenedioxy) diethanethiol (DODT), triisopropylsilane (TIPS), sodium nitrite, sodium phosphate monobasic, 4-mercaptophenylacetic acid (MPAA), trifluoroacetic acid (TFA), piperidine, dimethylformamide (DMF), dichloromethane (DCM), diethyl ether, guanidine hydrochloride (Gn·HCl), 2-mercaptoethanol, and tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl) were purchased from Sigma-Aldrich (St. Louis, MO, USA). 2,2′-Azobis[2-(2-imidazolin-2-yl)propane] dihydrochloride (Va-044) was purchased from Wako Chemicals (Richmond, VA, USA). HPLC-grade reagents, including water, methanol, acetonitrile, and formic acid, were purchased from Thermo Fisher Scientific (Waltham, MA, USA). Digitonin was purchased from MilliporeSigma (Burlington, MA, USA). Complete protease inhibitors were obtained from Roche Diagnostics GmbH (Mannheim, Germany).

Synthetic Chemistry

Peptide barcodes and CPPs were synthesized using the Liberty Blue HT12 microwave peptide synthesizer, according to the manufacturer's instructions (CEM Corporation, Charlotte, NC).[17,18] 0.2 M amino acids were made in DMF. 20% piperidine in DMF was used as the Fmoc-deprotecting solvent. 1 M OxymaPure ethyl 2-cyano-2-(hydroxyimino)acetate was used with 0.5 M DIC in the carbodiimide approach to form peptide bonds, with inhibition of racemization and improved coupling efficiencies.[19] All the barcode sequences were designed with N-terminal alanine (A) and C-terminal arginine (R) and one tryptophan (W) in their sequences to aid in UV absorbance detection, as listed in Table S1. In order to proceed with native chemical ligation, all of the barcode peptides were initially synthesized with N-terminal cysteine (C), instead of alanine (A), using rink amide resin, resulting in C-terminal amides. CPP peptides were synthesized using 2-chlorotrityl chloride (CTC) resin to prepare C-terminal hydrazide (R-NHNH$_2$).[20] The synthesized peptides were cleaved off the resin using the TFA/H_2O/TIPS/DODT solution (92.5:2.5:2.5:2.5) and purified by adding a 10-fold volume of cold diethyl ether to the TFA cleavage cocktail. Short barcode peptides (Table S1) used for generating signal−response standard curves were synthesized as SpikeTides (JPT Technologies, Berlin, Germany).

Barcoded CPP libraries were generated using native chemical ligation.[21] All reactions were carried out in 0.2 M phosphate buffer containing 6 M guanidine hydrochloride, pH 3 (NCL buffer). For each peptide, 3.6 mg of CPPs was dissolved in 0.4 mL of NCL buffer and stirred at −15 °C for 15 min. Separately, 13.6 mg of MPAA and 1.25 mg of barcode peptides were dissolved in 0.4 mL of NCL buffer, pH 6.5. The precooled CPP peptide solution was oxidized with 40 μL of 0.5 M $NaNO_2$ by stirring at −15 °C for 15 min. MPAA and barcode mixtures were added dropwise into the tube containing CPPs, the tube was then warmed to room temperature, and its pH was adjusted to 6.8−7.0. After overnight incubation with stirring at room temperature, reaction mixtures were reduced by the addition of 0.4 mL of 0.1 M TCEP (NCL buffer, pH 6.0−7.0) and stirring for 20 min.[22] Cysteine desulfurization was carried out as previously described.[23] Specifically, 2 mg of NCL-ligated barcoded-CPPs was dissolved in 300 μL of NCL buffer, into which 300 μL of aqueous 0.5 M TCEP·HCl, 20 μL of 2-mercaptoethanol (10% v/v), and 20 μL of 0.1 M VA-044 were added. The pH of the reaction mixture was adjusted to neutral, and the reaction was stirred at 45 °C for 45 min. All peptides were purified by high-performance liquid chromatography using an XBridge Peptide BEH C18 OBD Prep Column (#186008193, Waters, Milford, MA) and an acetonitrile gradient of 5−40% with 0.1% TFA to achieve at least 90% purity as measured by LC−MS.

Cell Culture

Kasumi-1, HEK293T, and HUVEC cells were obtained from the American Type Culture Collection (ATCC, Manassas, Virginia, USA). Huh-7 cells were kindly provided by Dr. Hao Zhu from the UT Southwestern Medical Center. All cell lines were verified by short tandem repeat analysis (Integrated Genomics Operation, Memorial Sloan Kettering Cancer Center, New York, NY, USA). Cultures were confirmed to be free of mycoplasma contamination by using Lonza MycoAlert (Lonza Walkersville, Inc., Walkersville, MD, USA). Cells were cultured at a concentration of 1×10^6 cells/mL in 5% CO_2 in a humidified atmosphere at 37 °C in complete media supplemented with 10% fetal bovine serum, 100 U/mL penicillin, and 100 μg/mL streptomycin. RPMI-1640 medium (Corning Life Science, Corning, NY, USA) was used for Kasumi-1 cells; Dulbecco's modified Eagle medium (DMEM) (Corning Life Science, Corning, NY, USA) was used for HEK293T and Huh-7 cells; and EGM-2 endothelial cell growth medium (Lonza Inc., Morristown, NJ, USA) was used for HUVEC cells.

Membrane Permeabilization and Cytotoxicity Measurements

Membrane stability was evaluated using the lactate dehydrogenase (LDH) release assay (no. ab65393, AbCam, Cambridge, UK), and cell viability was measured using the CellTiter-Glo luminescent assay, according to the manufacturer's instructions (no. G7571, Promega Corporation, Madison, WI, USA). Cells were counted with the Countess II Automated Cell Counter

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

(Thermo Fisher Scientific, Waltham, MA, USA), and 10,000 cells were aliquoted into each well of 96-well cell culture plates. Suspension Kasumi-1 cells were aliquoted immediately before treatment, while adherent HEK293T, Huh-7, and HUVEC cells were plated 17 h before treatment. 10 μL of barcoded CPP stock solutions (with varying concentrations) was added into 90 μL cell suspensions, resulting in final barcoded CPP concentrations ranging from tens of nanomolar to hundreds of micromolar. Then, the cells were incubated in complete media in 5% CO_2 in a humidified atmosphere at 37 °C for 3 and 24 h, respectively, and assayed using the TECAN Infinite M1000 Pro microplate reader (Tecan Group Ltd., Männedorf, Switzerland).

Cell Internalization and Fractionation

Cells were aliquoted into 12-well cell culture plates. Suspension Kasumi-1 cells (300 K) were aliquoted immediately before treatment, while adherent HEK293T (175 K), Huh-7 (240 K), and HUVEC (280 K) cells were plated 17 h before treatment. 10 μL of 100 μM barcoded CPP stock solutions was added into 990 μL cell suspensions, resulting in 1 μM final barcoded CPP concentration. Then, the cells were incubated in 5% CO_2 in a humidified atmosphere at 37 °C for 3 and 24 h, respectively. Kasumi-1 cells were pelleted by centrifugation, while the other three adherent cell lines were harvested using a cell scraper and then spun down. The cell pellets were washed once in PBS. For cell fractionation, a cell pellet of 2−3 million cells was resuspended in 90−150 μL of hypotonic buffer (10 mM HEPES, pH 7.9, 10 mM NaCl, 1 mM $MgCl_2$, 0.5 mM DTT, cOmplete protease inhibitors) containing 0.1% digitonin and was incubated at 25 °C for 5 min. The suspension was then centrifuged for 15 min at 3300g at 4 °C, and the cytoplasmic supernatant was collected. The nuclei pellet was then resuspended in 500 μL of sucrose resuspension buffer containing 0.25 M sucrose and 10 mM $MgCl_2$ (PBS buffer, pH 7.4). The suspension was layered onto a 500 μL sucrose cushion buffer (0.88 M sucrose and 0.05 mM $MgCl_2$ in PBS, pH 7.4) and centrifuged at 1200 g at 4 °C for 11 min. The nuclei pellet thus obtained was subjected to nuclear lysis and protein extraction in 70−110 μL of lysis buffer (6 M guanidine hydrochloride in PBS buffer, pH 7.4), sonicated by a Covaris S220 ultrasonicator (Covaris, LLC., Woburn, MA, USA) at a peak power of 125 W, a duty factor of 10%, a cycle/burst of 200, and a duration of 360 s. The protein content in both nuclear and cytoplasmic fractions was quantified using a Pierce BCA Protein Assay Kit, according to the manufacturer's instructions (Thermo Fisher Scientific, Waltham, MA, USA). Purified samples were stored at −80 °C.

Proteomics Sample Preparation

Purified protein extracts were digested with sequencing grade modified trypsin protease (Promega Corporation, Madison, WI, USA), with a protein-to-trypsin mass ratio of 10:1. The digestion mixture was incubated at 37 °C overnight. Digestion was halted by the addition of formic acid to 3.36% v/v, and 650 fmol of the Pierce Retention Time Calibration Mixture (PRTC) was added into each sample as internal standards. Tryptic peptides were purified using solid-phase extraction with BioPureSPN MIDI columns (#HEM S18 V, Nest Group, Southborough, MA, USA) according to the manufacturer's protocol. Briefly, the spin column was washed with 200 μL of methanol, activated with 200 μL of acetonitrile, and then equilibrated twice with 200 μL of 0.1% formic acid in water. Samples were loaded onto equilibrated columns and then washed twice with 200 μL of 0.1% formic acid in water. Peptides were eluted with 60% acetonitrile containing 0.1% formic acid and lyophilized by vacuum centrifugation. For analysis, purified peptides were resuspended in 0.1% formic acid in water. Calibration curves of barcoded CPPs were established by adding variable amounts of synthetic peptides to cell lysates, followed by their digestion and purification as described above.

Nanoscale Liquid Chromatography and Nanoelectrospray Ionization Mass Spectrometry

All nanoscale LC experiments were performed by using the Eksigent nanoLC425 chromatography system (Sciex, Framingham, MA, USA). The 50 cm uPAC pillar array reverse phase column (#COL-NANO050G1B, Thermo Fisher Scientific, San Jose, CA, USA) was used for the LC−MS signal−response study of the pilot 96 peptide barcodes. Capillary reverse-phase columns were used for cellular delivery quantitation. These capillary columns were fabricated by pressure filling the stationary phase into silica capillaries fritted with K-silicate, as previously described.[24] Samples were resolved with a constant flow rate of 300 nL/min using a 5−40% linear gradient of acetonitrile in water (both with 0.1% v/v/formic acid) over 90 min. Ionization was accomplished using laser-fabricated emitters with a terminal opening diameter of 2−3 μm, made from 50 μm silica capillaries, and connected to the outlet of the reversed-phase column using a metal union that also served as the electrospray current electrode, as described.[25] Eluting peptides were transferred to an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) via the DPV-566 PicoView nanoelectrospray ion source (New Objective, Woburn, MA, USA).

For data dependent acquisition (DDA), the full-scan spectra were acquired in the positive ion mode at a fixed resolution of 120,000 and a mass range of 300−1200 m/z. Fragmentation spectra were acquired at a resolution of 7500, and the precursor ions were isolated using 1.6 m/z width and fragmented with a fixed higher-energy collisional dissociation (HCD) collision energy of 30%.

For parallel reaction monitoring (PRM) scans, full-scan spectra were acquired in the positive ion mode at a mass range of 300−1200 m/z and 120,000 resolution. Precursor ions were isolated in the quadrupole using 1.2 m/z windows and fragmented by HCD with a normalized collision energy of 30% (stepped collision energy ±5%), before analysis of fragment ions in the Orbitrap at 7500 resolution. The PRM method was scheduled using 10 min acquisition windows set for each barcode based on the DDA retention time results.[26]

For selected ion monitoring (SIM) scans, the precursor ions were isolated in the quadrupole using 0.7 m/z window and were detected in the Orbitrap at 240,000 resolution. The maximum injection time mode was set as dynamic with six minimum points across the peak. The SIM method was scheduled using 10 min acquisition windows set for each barcode based on the DDA retention time results.

The mass spectra obtained from all methods were analyzed using Skyline[27] (version 21.2.0.425). The target list was first established in Skyline, and then, the MS raw data files were imported to extract the target chromatographs. DDA MS1 spectra were filtered with a resolving power of 120,000 at m/z 400. SIM MS1 spectra were filtered with a resolving power of 240,000 at m/z 200. PRM MS2 spectra were filtered using 60,000 resolving power at m/z 200, and at least 3 fragments per precursor are required for valid quantitation. For the LC−MS signal−response study, the MS1 and MS2 ion intensities of each

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

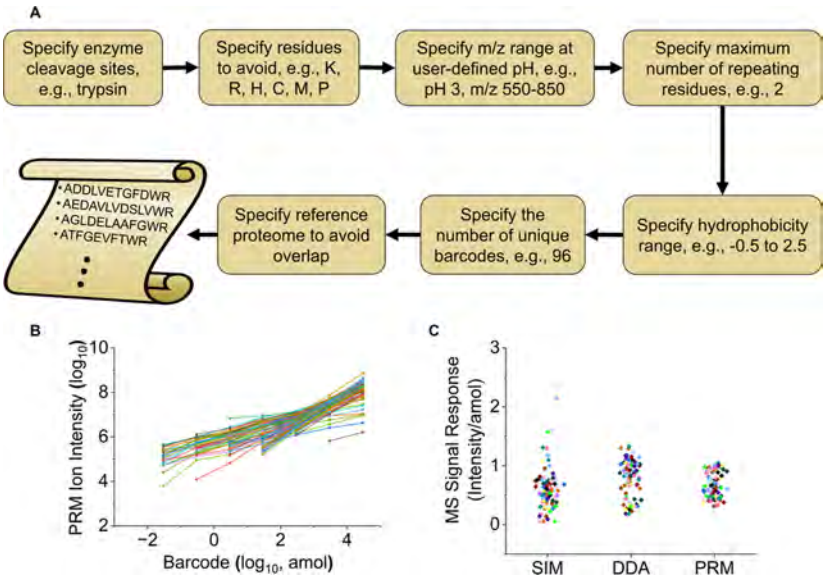


Figure 1. BarcodeBabel algorithm generates unique peptide barcodes that are readily detectable by high-resolution MS. (A) Schematic of BarcodeBabel to generate barcode libraries. (B) Signal−response functions for the library of 96 designed barcodes using PRM. The MS2 XIC signal is used as the *y* axis. (C) Signal response (MS signal intensity/amol of barcodes) of the signal−response functions of each barcode obtained from SIM, DDA, and PRM.

barcode precursor in all serially diluted samples were extracted and linearly fitted to obtain a signal−response function for each barcode. For the absolute quantitation of digested cell fractions, MS2 ion intensities of all barcodes were exported and normalized to the ion intensities of the PRTC ionization standards. A baseline sample (untreated cell lysate) was measured, and any ion intensities of the cell fractions that were lower than or comparable to the baseline signal were considered nondetectable. Only valid signals from the cell fractions were calculated by using the calibration curves to obtain the final delivery quantities.

Numerical and statistical analyses were performed using Origin Pro (OriginLab Corporation, Northampton, MA).

BarcodeBabel

BarcodeBabel is a Python algorithm to generate libraries of peptide barcodes with user-defined features for optimal detectability by nanoscale LC and tandem MS. User-selected rules include the m/z range, absence of homopolymers, specific hydrophobicity range, enzyme cleavage sites, residue frequencies, and library size. In addition, users can specify a reference proteome to remove any naturally occurring interfering sequences. The default enzyme for barcode cleavage is trypsin. Specific residues are programed to be avoided: lysine (K), arginine (R), and histidine (H) were omitted to prevent trypsin cleavage within the barcode sequences; methionine (M) and cysteine (C) were omitted to avoid oxidation and cross-linking events; proline (P) was omitted as it may skew fragmentation (though could be included to promote fragmentation at the proline residue); glutamine (Q) and asparagine (N) were avoided due to their propensity for deamidation; and isoleucine (I) was omitted because it cannot be distinguished from leucine (L) by conventional MS. Peptide properties were calculated at a default pH of 3 (for conventional positive ion mode electrospray). The default m/z range was set to 550−850 to fall within the optimal detection window of high-field Orbitrap mass analyzers. The default hydrophobicity range (−0.5 to 2.5) is specified to allow for consistent peptide elution in the middle of conventional reverse-phase chromatography gradients. Users can also list specific reference proteomes as well as common contaminants to avoid sequence overlap. BarcodeBabel is implemented open-source via https://github.com/kentsisresearchgroup/BarcodeBabel.

PeptideBabel

PeptideBabel is a Python algorithm for the generation of novel bioactive peptide sequences using Metropolis−Hastings sampling. This algorithm generates novel peptides by exploring the sequence space around a set of seed sequences using Markov chain Monte Carlo sampling (Metropolis−Hastings) with the permutation of seed sequences mapped to a density function based on physicochemical features or k-mer sequence complexity, as specified by users. This allows efficient generation of hundreds to millions of peptide sequences for subsequent empirical validation, tailored to application based on input peptides. Briefly, after the user uploads a list of references peptides (seed library), PeptideBabel identifies key properties of the seed library, including sequence length, hydrophobicity (windowed Kyte−Doolittle scale), isoelectric point (Bjellqvist), and secondary structure propensity (fraction helical, turn-like, and sheet-like based on residue composition using the Garnier−Osguthorpe−Robson method). The sampling algorithm permutes peptide sequences (substitutions, insertions, and deletions) at user-defined sequence length constraints and sampling density and implements a random-walk Metropolis−Hastings, favoring steps moving up the density function at user-defined probability to generate new sequences or down the density function to sample similar sequences. PeptideBabel is implemented open-source via https://github.com/kentsisresearchgroup/PeptideBabel.

RESULTS

MS proteomics enables the detection and quantitation of specific macromolecules based on tandem fragmentation and high-accuracy measurements of their mass and charge. This enables the resolution of unique polypeptide sequences,

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

differing by as little as one amino acid. To enable the generation of libraries of unique peptide barcodes for studies of engineered macromolecules in complex biological samples, we developed BarcodeBabel, a python script and the corresponding Jupyter notebook, which computes arbitrary numbers of unique amino acid peptide sequences, flanked by specific enzyme cleavage sites for their release in complex biological samples, e.g., trypsin (Figure 1A). The script can exclude user-defined internal amino acids, which may interfere with correct enzymatic cleavage and consistent mass spectrometric detection or fragmentation, i.e., lysine, arginine, histidine, cysteine, methionine, proline, and isoleucine. Users may also specify frequencies of amino acids to generate libraries with specific features, such as inclusion of tryptophan residue for optical measurements as well as specific ranges of m/z values and amino acid lengths. The script also implements a hydrophobicity estimator based on the Kyte−Doolittle scale for optimal reverse-phase separations that accompany high-resolution MS measurements. To calibrate this parameter, we analyzed 175,579 unique tryptic peptides of length from 5 to 15 residues[28] and found that peptides with hydrophobicity scores of −0.5 to 2.5 exhibit monotonically variable retention times in reverse-phase chromatography most often coupled with modern high-resolution MS instruments, which we implement as the default values for BarcodeBabel (Figure S1). Finally, BarcodeBabel is configured with a user-specified reference proteome to ensure that the designed barcode sequences do not match any naturally occurring or common contaminant proteins.

To test BarcodeBabel, we designed a library of 96 unique barcode sequences, synthesized them using solid-phase synthesis, and serially diluted the purified peptides in whole-cell extracts of human OCI-AML2 cells. The abundance of peptides in whole-cell proteomes was quantified by using SIM, data-dependent acquisition (DDA), and PRM, with limits of quantitation (LOQs) determined using Skyline. We found that PRM exhibited superior LOQs, as compared to SIM and DDA (mean LOQ of 75 versus 550 and 280 amol, respectively; Figure 1B). We also found that PRM exhibited a more uniform signal−response for the quantitation of barcode peptide abundance, as compared to SIM and DDA (mean 0.67 ± 0.22 versus 0.60 ± 0.34 and 0.80 ± 0.30 intensity/amol, respectively; Figure 1C). Thus, BarcodeBabel permits the construction of libraries of specific peptide barcodes, which can be quantitatively deconvoluted using high-resolution MS.

Pioneering studies of first-generation CPPs and PTDs for macromolecular delivery used naturally inspired peptides derived from TAT and penetratin.[29,30] Since then, a variety of cationic and amphipathic CPPs have been identified experimentally, as most recently catalogued in the CPPsite 2.0 database.[9,31] Comparative studies of specific CPPs have identified several key propensities, such as the optimal number of eight guanidine side chains for cationic polyarginine CPPs.[32] However, the development of explicit structure−activity relationships for efficient, selective, and safe CPPs and PTDs has been challenging, at least in part due to the diversity and complexity of CPPs and their membrane penetration and cellular internalization mechanisms.

We reasoned that the mechanisms of CPP membrane penetration and cellular internalization ultimately can be learned from large-scale structure−function studies. To generate libraries of candidate CPPs for high-throughput studies, we implemented a Monte Carlo sampling algorithm, PeptideBabel. PeptideBabel uses the Metropolis−Hastings algorithm to introduce random changes in the amino acid composition of seed peptide sequences, followed by their acceptance or rejection based on user-specified density functions of either the physicochemical properties of peptides or their k-mer sequence complexity (Figure 2). The current version of

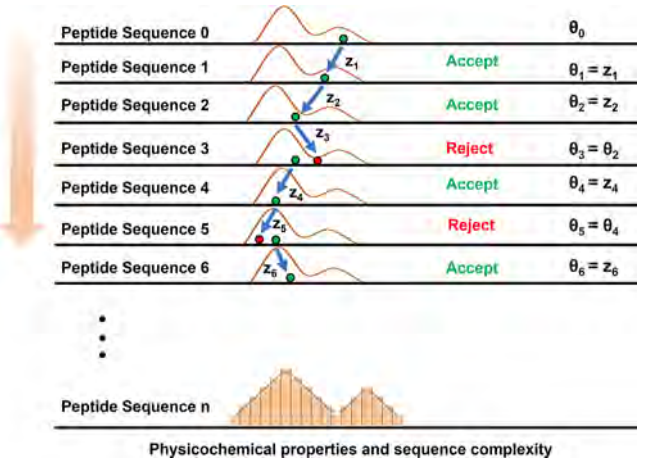


Figure 2. Schematic of the PeptideBabel algorithm, which generates bioactive peptide sequences using Markov chain Monte Carlo sampling with permutation of seed sequences (sequence 0) mapped to a density function. PeptideBabel implements a random-walk Metropolis−Hastings algorithm, accepting steps moving up and rejecting steps moving down the density function using user-defined probability functions. The model parameters (θ) and proposed parameters (z) are based on the k-mer space that indicates sequence complexity or the physiochemical space (e.g., sequence length, isoelectric point, secondary structure propensity, and hydrophobicity). The figure is based on Jin et al.[35]

PeptideBabel includes 1552 seed sequences of putative membrane penetration domains from diverse viral pathogens,[33] as well as curated members from CPPsite 2.0 (Figure S2 and Table S4). PeptideBabel samples the chemical space according to the physicochemical scoring function of the seed sequences: it is based on the linear combination of peptide length, estimated isoelectric point, hydrophobicity (Kyte−Doolittle), and secondary structure propensity (Garnier−Osguthorpe−Robson). Alternatively, PeptideBabel can sample peptide sequences based on their sequence complexity, as measured using k-mer scoring.[34] The Monte Carlo sampling and design can be executed to generate unique sequences with physicochemical or complexity properties that are either similar to the seed sequences or, alternatively, those that are of increasing diversity from the seed sequences. Convergence of sampling can be assessed using multiple independent calculations, starting from different initial conditions.[35] Thus, PeptideBabel is expected to permit the construction of diverse libraries of candidate bioactive peptides.

As a proof of concept of this strategy, we designed a library of seven barcoded CPPs, including representative cationic and amphipathic CPPs, as well as novel chimeric CPPs, and their anionic negative controls that do not transit across anionic mammalian cell membranes (Table 1). We chose unique barcodes with diverse physicochemical properties, as estimated by their m/z, net charge, and hydrophobicity values (Figure S3). Barcode peptides and CPPs were synthesized using solid-phase synthesis and assembled using native chemical ligation.[22] Because native chemical ligation involves cysteines that can

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. CPP Candidates Chosen for Chemical Synthesis and the Cellular Delivery Assay[a]

ID	CPP	sequence	CPP type	sequence source	barcode
1	TAT	YGRKKRRQRRR	cationic	prototypic CPP, NLS from HIV TAT protein	B64
2	P14	RKKRWFRRRRPKWKK	cationic	TAT/penetratin hybrid	B91
3	TAT-P-ebola	GAAIGLAWIPYFGPAAYPRKKRRQRRR	hydrophobic + cationic (moderate hydrophobicity)	good penetrator in pilot study, chimera design of TAT + Ebola coat	B121
4	TAT-G-EBV	IYNGWYAYGRKKRRQRRR	hydrophobic + cationic (moderate hydrophobicity)	TAT + Epstein−Barr Virus coat	B107
5	KLAL-TAT	KLALKLALKALKAALKLAGCYGRKKRRQRRR	amphipathic	model amphipathic peptide + TAT	B82
6	RLAL-TAT	RLALRLALRALRAALRLAGCYGRKKRRQRRR	amphipathic	model amphipathic peptide + TAT	B108
7	badTAT	YGEKKEEQRRR	negative ctrl	negative control	B55

[a]NLS: nuclear localization sequence.

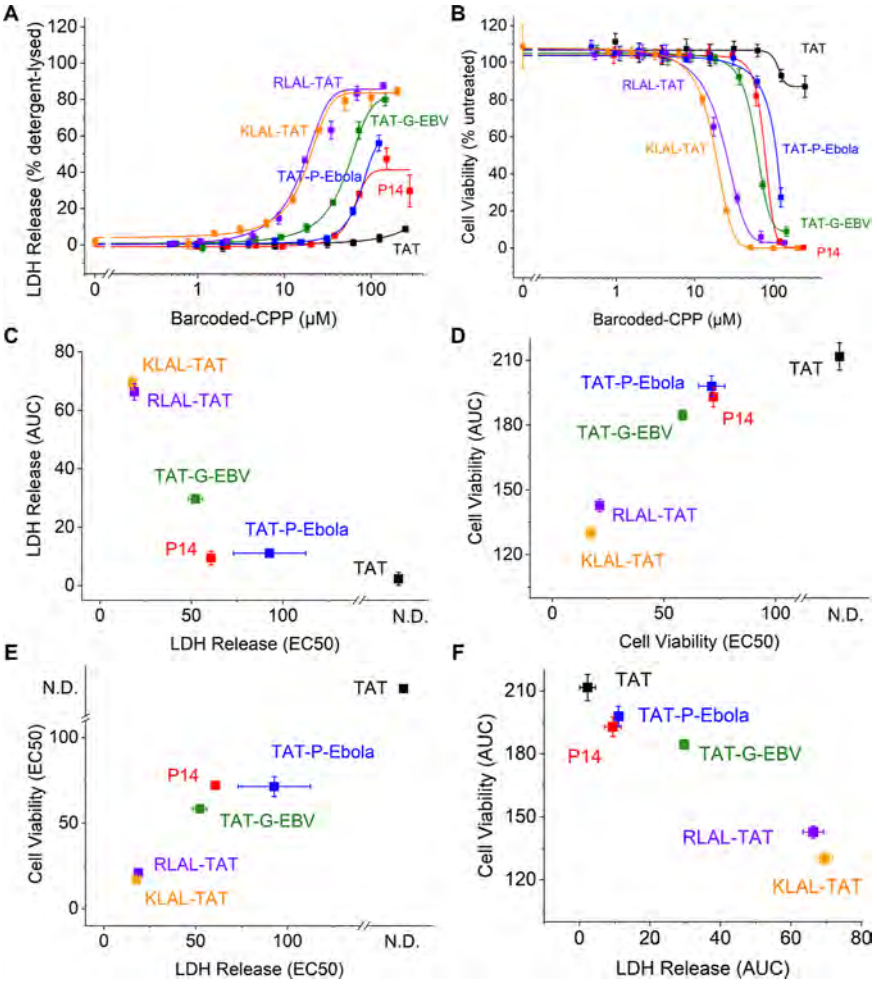


Figure 3. Membrane destabilization and cytotoxicity measurements of CPPs. (A) LDH release of Huh-7 cells after 3 h incubation of barcoded CPPs. (B) Cell viability of Huh-7 cells after 3 h incubation of barcoded CPPs. (C) EC_{50} values as a function of the AUC of the LDH release measurements. (D) EC_{50} values as a function of the AUC of the cell viability. (E) EC_{50} values of LDH release as a function of EC_{50} values of cell viability. (F) AUC of LDH release as a function of the AUC of cell viability. Symbols and whiskers represent mean and standard deviation values of three biologic replicates, respectively.

undergo chemical reactions in cells, we used desulfurization to convert them into alanines.[23] This allowed us to generate barcoded CPPs up to 42 amino acids in length with >90% purity, as confirmed using LC−MS (Figure S4).

CPPs can disrupt cellular membranes, causing both cytotoxicity and artifactual internalization due to cell death. Therefore, we measured the effects of individual barcoded CPPs on cell viability, as assessed with cellular ATP content and membrane stability with assays of LDH release using a panel of four biologically diverse cell lines: hematopoietic Kasumi-1 cells, mesenchymal HEK293T cells, endothelial HUVEC cells, and epithelial Huh-7 cells (Figures 3, S5, and S6). We found that the TAT-[barcode64] CPP showed no measurable cytotoxicity up to 15 μM either at 3 or 24 h of exposure on Kasumi-1 cells and up to 100 μM across the other three cell lines tested. In contrast, 3 h incubation of chimeric KLAL-TAT-[barcode82] CPP promoted LDH release and impaired cell viability with IC_{50} values of 1.4 and 2.3 μM, respectively, in HUVEC cells and 34

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

and 10 μM, respectively, in HEK293T cells, with the other two cell lines having intermediate values (Figures 3, S5, and S6). We found that Kasumi-1 cells were most sensitive to various CPPs, whereas Huh-7 and HEK293T cells were largely insensitive. Importantly, all barcoded CPPs had negligible cytotoxicity and membrane disruption at 1 μM, thereby establishing this dose for cell delivery studies in vitro (Figures 6, 7, and S9−S11). Though many prior studies have used 10−100 μM CPP cell treatments,[36] we chose to use this lower concentration in order to avoid any potential confounding effects from membrane disruption and cytotoxicity.

To assess membrane penetration and subcellular delivery of various CPPs, we first sought to establish a robust cellular fractionation method to measure nuclear versus cytoplasmic accumulation of barcoded CPPs. We found that plasma membrane extraction using 0.1% digitonin, followed by sucrose density sedimentation, produced specific separation of nuclear versus cytoplasmic compartments, as validated by Western immunoblotting against Lamin B1, Histone H3, and GAPDH, respectively (Figures 4 and S7; Tables S2 and S3). We then established a quantitative procedure for measuring the absolute abundance of specific barcoded CPPs using cell fractionation combined with targeted PRM MS, with variation in ionization efficiency controlled by normalization to synthetic PRTC peptides (Figure 5). Figure 5b shows the signal−response function for representative barcode B91.

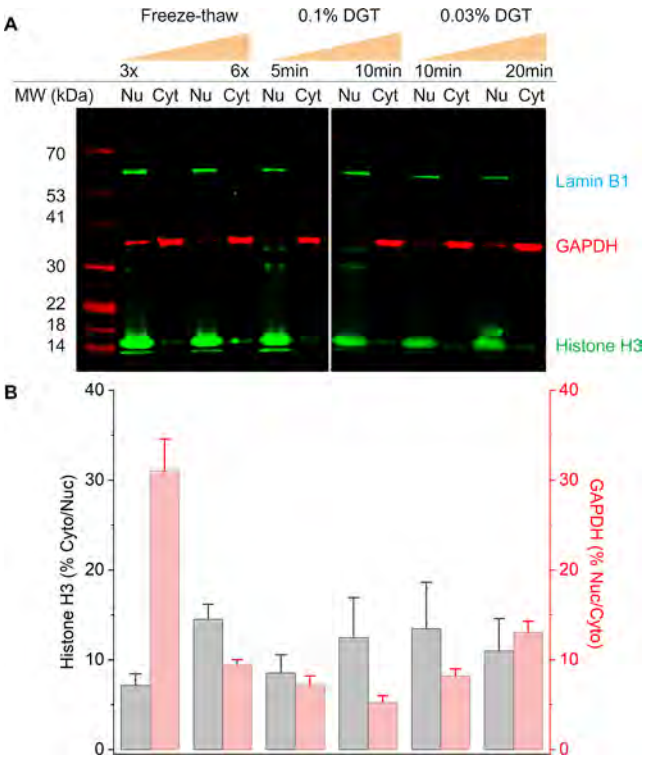


Figure 4. Efficient cell fractionation achieved using digitonin and sucrose density sedimentation. (A) Representative Western blot with cytoplasmic (Cyt) marker GAPDH and nuclear (Nu) markers Lamin B1 and Histone H3, as a function of various numbers of freeze−thaw cycles (left), time of treatment with 0.1% digitonin (DGT), and time of treatment with 0.03% DGT, as indicated. B. Fluorescence densitometry quantitation of 0.1% DGT-treated cells of Histone H3 abundance (gray) and GAPDH abundance (pink) in nuclear (Nuc) and cytoplasmic (Cyto) fractions. The *x*-axis corresponds with the Western blot columns in panel A. Histone H3 (% Cyto/Nuc) is the % ratio calculated by the band intensity of Histone H3 in the cytoplasmic fraction divided by band intensity of Histone H3 in the nuclear fraction. Lower ratios (%) correspond to more efficient cell fractionation. Bars and whiskers represent mean and standard deviation values of three biologic replicates, respectively.

Using this approach, we first sought to determine the potential contribution of barcode peptides to the cellular penetration activity of the chimeric TAT-P-ebola CPP, chosen because it contains both cationic and amphipathic CPP components (Figure 6). Library of TAT-P-ebola GAAIGLAWIPYFG-PAAYPRKKRRQRRR-[barcode] CPPs containing 8 diverse peptide barcodes showed no significant variation in their membrane destabilization or cytotoxicity, as measured by LDH release and ATP content, respectively (mean IC_{50} = 22 ± 6.0 and 12 ± 2.2 mM, respectively; Figure 6a,b). In contrast, TAT-P-ebola variants with AFSVDAETLWR [barcode82] and AGLDELAAFGWR [barcode91] barcodes tested at nontoxic 1 μM concentrations exhibited significantly enhanced cytoplasmic accumulation, as compared to that of B55, B64, B81, B107, B108, and B121 barcodes, which were nearly exclusively nuclear (mean cytoplasmic abundance = 22 and 9.9 millions of molecules/cell, two-tailed unpaired Student's *t*-test p = 0.018 and 0.022, respectively; Figure 6c). These results indicate that peptide barcoding can be used both to quantitatively measure and to modulate the biologic properties of CPPs.

We then investigated the nuclear and cytoplasmic delivery of barcoded CPPs in our panel of cell lines, tested at 1 μM concentrations with negligible cytotoxicity and membrane disruption (as shown in Figures 3, S5, and S6). As a negative control, we used the anionic version of TAT, termed badTAT, in which key arginine residues have been replaced with glutamates (YGEKKEEQRRR-[barcode55]), preventing its membrane translocation, which consistently led to a lack of measurable accumulation of badTAT in either cytoplasmic or nuclear fractions (Figure 7a). We detected nuclear accumulation of TAT-[barcode64] and P14-[barcode91] CPPs, which is consistent with prior studies, but at relatively low levels, due to the low 1 μM concentration of treatment, in order to be more relevant for future therapy development and to minimize the confounding effects of cytotoxicity; TAT and P14 CPPs are used at >10 μM concentrations in many prior studies.[14,37]

Notably, the novel chimeric CPPs TAT-P-ebola-[barcode121], TAT-G-EBV-[barcode107], and RLAL-TAT-[barcode108] exhibited significantly higher nuclear accumulation (mean nuclear abundance = 403, 251, and 196 millions of molecules/cell, two-tailed unpaired Student's *t*-test p < 0.0001 versus TAT-[barcode64]; Figure 7a). In addition, TAT-P-ebola-[barcode121] and RLAL-TAT-[barcode108] but not TAT-G-EBV-[barcode107] exhibited a time-dependent increase in nuclear accumulation after 24 h of cell exposure (Figure 7b). Interestingly, KLAL-TAT-[barcode82] also exhibited significant time-dependent cytoplasmic accumulation, as compared to other barcoded CPPs (mean cytoplasmic abundance = 27 and 42 millions of molecules/cell at 3 and 24 h, two-tailed unpaired Student's *t* = test p = 0.0001 and 0.0004 versus TAT-[barcode64], respectively; Figure 7a,b). This effect may be potentiated by the specific contribution of the AFSVDAETLWR B82 barcode to the apparent CPP activity (Figure 6c).

We also found that the activity of novel barcoded CPPs was cell type-specific. For example, KLAL-TAT-[barcode82] showed increased cytoplasmic accumulation in mesenchymal HEK293T, epithelial Huh-7, and hematopoietic Kasumi-1 cells but not in endothelial HUVEC cells (mean cytoplasmic

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

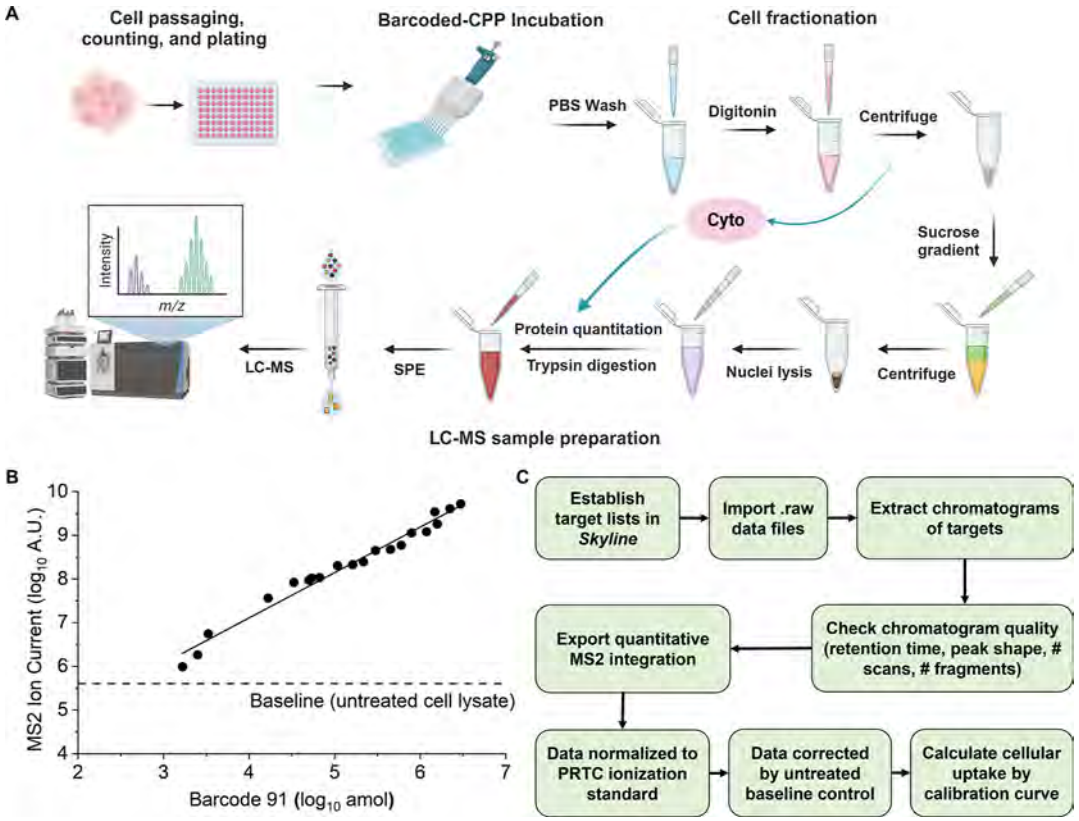


Figure 5. Barcoding and screening strategy results for absolute quantitation of CPP penetration. (A) Schematic of the cell treatment, fractionation, and LC−MS analysis. (B) Signal−response function of representative barcode 91, showing the MS2 ion current intensity as a function of peptide abundance. Solid line indicates the linear fit; dashed line indicates the baseline noise value from the untreated cell lysate. (C) Schematic for data analysis to calculate extracted ion chromatograms, normalized for variation in ionization efficiency using PRTC reference standards and background noise from untreated control samples.

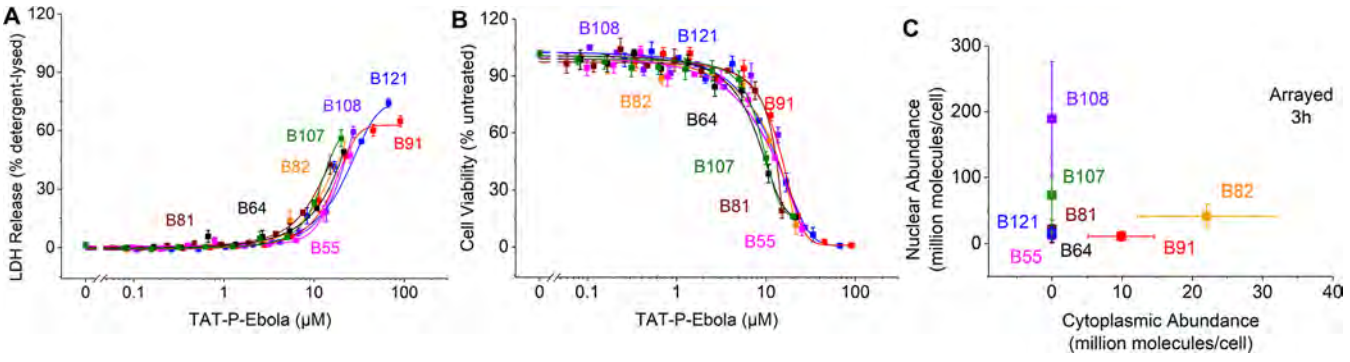


Figure 6. Barcoded TAT-P-ebola peptides exhibit efficient cellular penetration without apparent membrane disruption and cytotoxicity in vitro. (A) LDH release measurements of Kasumi-1 cells upon 3 h incubation with barcoded TAT-P-ebola peptides with various barcodes, as indicated. (B) Cell viability of Kasumi-1 cells upon 3 h incubation with barcoded TAT-P-ebola peptides with various barcodes, as indicated. (C) Abundance of barcoded TAT-P-ebola CPPs in nuclear and cytoplasmic fractions upon 1 mM treatment for 3 h. Symbols and whiskers indicate the mean and standard deviation values of three biologic replicates, respectively. The specific sequence of TAT-P-ebola and the barcodes are listed in Table 1 and Supporting Information Table S1.

abundance = 31, 27, and 15 versus 0 millions of molecules/cell, respectively; Figure 7c). In contrast, TAT-P-ebola-[barcode121] exhibited increased nuclear accumulation in Huh-7 and HUVEC cells, as compared to HEK293T and Kasumi-1 cells (mean nuclear abundance = 420 and 291 versus 112 and 30.5 millions of molecules/cell, respectively; Figure 7d). Thus, the combination of peptide barcoding and de novo CPP design can be used to discover CPPs with improved cellular penetration activities and reduced toxicities (Figure 8). In addition, comparative studies can reveal time-dependent and cell type-specific differences in activity, thereby identifying potential targets for mechanistic studies.

DISCUSSION

Nucleic acid barcoding has become a highly enabling technology for diverse high-throughput biological studies, such as binding studies using diversity-oriented synthesis and DNA barcoding and peptide and protein engineering using RNA display and RNA barcoding, among others.[38] Peptide barcoding is

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

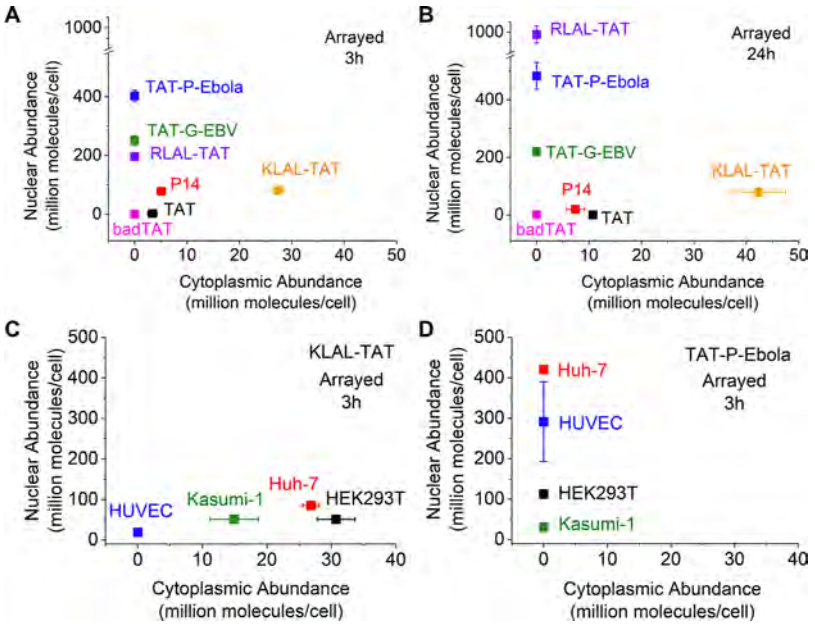


Figure 7. Chimeric CPPs exhibit improved cell type-specific subcellular penetration. (A) Abundance of barcoded CPPs in nuclear and cytoplasmic fractions of Huh-7 cells upon 3 h treatment. TAT (black) serves as a conventional cationic CPP. badTAT (pink) serves as nonpenetrating negative control. (B) Abundance of barcoded CPPs in nuclear and cytoplasmic fractions of Huh-7 cells upon 24 h treatment. (C,D) Abundance of barcoded KLAL-TAT CPP (C) and TAT-P-ebola CPP (D) in nuclear and cytoplasmic fractions of HUVEC, Kasumi-1, Huh-7, and HEK293T cells, as indicated, upon 3 h treatment. Symbols and whiskers represent the mean and standard deviation values for three biologic replicates, respectively. The specific sequence of the CPP candidates and the barcodes are listed in Table 1 and Supporting Information Table S1.

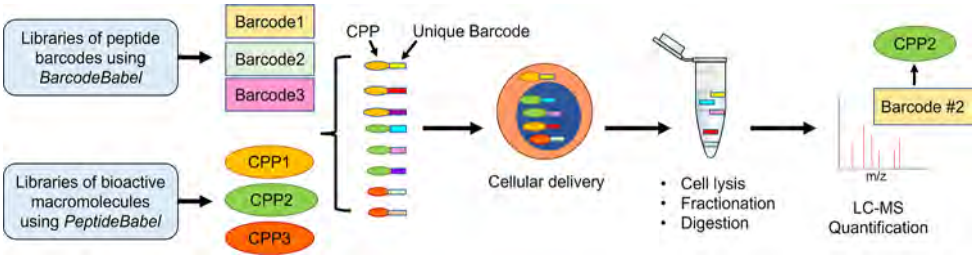


Figure 8. Schematic for peptide barcoding and screening for profiling of bioactive macromolecules using BarcodeBabel and PeptideBabel. The subcellular delivery of specific CPPs is deconvoluted using quantitative MS of the peptide barcodes.

particularly compelling for the engineering, screening, and other studies of biological molecules because of their homogeneous biochemical properties, superior stability, and information content, in contrast to mixed macromolecules with nucleic acid barcodes.[16]

Here, we developed an open-source algorithm, BarcodeBabel, designed to construct libraries of unique peptide barcodes with optimal properties for high-throughput MS proteomics studies. This enables the design of unique sequences not present in canonical biological proteomes, e.g., human tissues, with optimal ionization, separation, and fragmentation properties, as empirically validated using a library of 96 designed peptide barcodes.

BarcodeBabel is accompanied by the open-source algorithm PeptideBabel, which implements Monte Carlo sampling of user-defined seed sequences to generate novel libraries of peptides with diverse physicochemical and sequence complexity properties and to create bioactive macromolecules that can be uniquely identified and quantified in complex biological environments. We tested this approach empirically using a pilot library of barcoded CPPs. We found that peptide barcoding can be used to quantitatively measure cell penetration and subcellular distribution of CPPs, as validated by using known CPPs and their inactive negative controls. This work also implemented a targeted MS method suitable for quantitative studies of absolute molecular abundance of peptide barcodes in complex biological samples.

Using this proof-of-concept study, we identified novel CPPs with improved nuclear and cytoplasmic delivery exceeding hundreds of millions of molecules per human cell with distinct cell type specific activities while maintaining minimal membrane disruption and negligible toxicity in vitro. While we observed improved nuclear and cytoplasmic delivery of novel chimeric CPPs such as TAT-P-ebola and KLAL-TAT, future studies will be required to define specific structure−activity relationships for CPPs and their precise mechanisms of membrane penetration and subcellular distribution, potential cellular receptors, and mechanisms of cell type-specific internalization. Interestingly, we found that in some cases, the specific barcode modulated the apparent CPP activity, suggesting that peptide barcodes may themselves be incorporated into the design of bioactive molecules. This also indicates that high-throughput screens should utilize multiple independent barcodes in order to discern

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX

specific biological activities, as is practiced with nucleic acid barcoding and other high-throughput technologies.

We anticipate that BarcodeBabel and PeptideBabel should be useful for diverse screening, design, and analytical studies. For example, this approach may be used to design novel protein binders and quantify their binding affinities and kinetics using libraries of purified barcoded proteins in vitro.[15,39] Given the high sensitivity and resolving power of modern mass spectrometers, similar screens may also be performed with libraries of barcoded CPPs injected intravenously and quantified using proteomics of specific tissues and organs in vivo. BarcodeBabel and PeptideBabel enable the construction of barcoded libraries of peptidic macromolecules with varied biological activities and thus should be useful for a wide variety of molecular evolution and screening applications.

■ ASSOCIATED CONTENT

Data Availability Statement

All raw and processed MS data as well as Skyline chromatogram documents are available via ProteomeXchange with the identifier PXD048412.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.4c00068.

Design strategy and design parameters of BarcodeBabel, the pilot 96 barcode sequences designed by BarcodeBabel, the workflow to identify candidate CPPs by cataloguing viral peptide sequences, physiochemical properties of the 96 barcode library, representative LC chromatographs and MS spectra of the synthetic barcoded CPPs, the LDH release and cell viability profiles of the barcoded CPPs using four different cell lines after incubation for 3 h, the LDH release and cell viability profiles of the barcoded CPPs using four different cell lines after incubation for 24 h, physiochemical methods tested for cell fractionation efficiency, evaluating the cell fractionation efficiency of various physical and chemical methods, experimental conditions tested for freeze−thaw and digitonin-based cell fractionation methods, calibration curves of all the barcodes used in the cellular delivery assays, the cytotoxicity profiles and cellular delivery of barcoded TAT-P-ebola peptides, the extracted MS2 ion current signal for each barcode in each cell fraction across four different cell lines, quantitative cellular delivery of the barcoded CPPs in four different cell lines, image of the entire Western blot membrane shown in Figure 4, and image of the entire Western blot membrane shown in Figure S7 (PDF)

List of seed sequences used in the current version of PeptideBabel (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Alex Kentsis − *Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States; Tow Center for Developmental Oncology, Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States; Departments of Pediatrics, Pharmacology, and Physiology & Biophysics, Weill Cornell Medical College, Cornell University, New York, New York 10065, United States;* orcid.org/0000-0002-8063-9191; Email: kentsisresearchgroup@gmail.com

Authors

Ning Wang − *Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States*

Nicole A. Mcneer − *Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States*

Elliot Eton − *Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States;* orcid.org/0000-0002-8980-9310

Josh Fass − *Tri-I PhD Program in Computational Biology and Medicine, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York 10065, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.4c00068

Author Contributions

Conceptualization: N.W., N.A.M., and A.K.; investigation: N.W., N.A.M., E.E., and J.F.; analysis: N.W., N.A.M., and A.K.; resources: N.W., N.A.M., E.E., and J.F.; writing of the original draft: N.W. and A.K.; writing of the final draft: all authors; and funding acquisition: A.K.

Funding

This work was supported by NIH R01 CA214812, U54 CA243124, P30 CA08748, Starr Cancer Consortium, Doris Duke Charitable Foundation, and Mr. William H and Mrs. Alice Goodwin and the Commonwealth Foundation for Cancer Research the Center for Experimental Therapeutics at MSKCC. A.K. is a scholar of the Leukemia and Lymphoma Society.

Notes

The authors declare the following competing financial interest(s): Alex Kentsis is a consultant for Novartis, Rgenta, Blueprint Medicines, and Syndax.

■ ACKNOWLEDGMENTS

We acknowledge Markus Seeger and Pascal Egloff for helpful discussions and members of our lab for critical advice.

■ REFERENCES

(1) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic peptides: Current applications and future directions. *Signal Transduct. Targeted Ther.* **2022**, *7* (1), 48.

(2) Lu, H.; Zhou, Q.; He, J.; Jiang, Z.; Peng, C.; Tong, R.; Shi, J. Recent advances in the development of protein-protein interactions modulators: mechanisms and clinical trials. *Signal Transduct. Targeted Ther.* **2020**, *5* (1), 213.

(3) Muttenthaler, M.; King, G. F.; Adams, D. J.; Alewood, P. F. Trends in peptide drug discovery. *Nat. Rev. Drug Discovery* **2021**, *20* (4), 309−325.

(4) Ebrahimi, S. B.; Samanta, D. Engineering protein-based therapeutics through structural and chemical design. *Nat. Commun.* **2023**, *14* (1), 2411.

(5) Cao, L.; Coventry, B.; Goreshnik, I.; Huang, B.; Sheffler, W.; Park, J. S.; Jude, K. M.; Marković, I.; Kadam, R. U.; Verschueren, K. H.; et al. Design of protein-binding proteins from the target structure alone. *Nature* **2022**, *605* (7910), 551−560.

(6) Guidotti, G.; Brambilla, L.; Rossi, D. Cell-penetrating peptides: from basic research to clinics. *Trends Pharmacol. Sci.* **2017**, *38* (4), 406−424.

(7) Frankel, A. D.; Pabo, C. O. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* **1988**, *55* (6), 1189−1193.
(8) Green, M.; Loewenstein, P. M. Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. *Cell* **1988**, *55* (6), 1179−1188.
(9) Agrawal, P.; Bhalla, S.; Usmani, S. S.; Singh, S.; Chaudhary, K.; Raghava, G. P.; Gautam, A. CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res.* **2016**, *44* (D1), D1098−D1103.
(10) Kondo, E.; Saito, K.; Tashiro, Y.; Kamide, K.; Uno, S.; Furuya, T.; Mashita, M.; Nakajima, K.; Tsumuraya, T.; Kobayashi, N.; Nishibori, M.; Tanimoto, M.; Matsushita, M. Tumour lineage-homing cell-penetrating peptides as anticancer molecular delivery systems. *Nat. Commun.* **2012**, *3*, 951.
(11) Guha, S.; Ghimire, J.; Wu, E.; Wimley, W. C. Mechanistic landscape of membrane-permeabilizing peptides. *Chem. Rev.* **2019**, *119* (9), 6040−6085.
(12) Kauffman, W. B.; Fuselier, T.; He, J.; Wimley, W. C. Mechanism matters: a taxonomy of cell penetrating peptides. *Trends Biochem. Sci.* **2015**, *40* (12), 749−764.
(13) Hoffmann, K.; Milech, N.; Juraja, S. M.; Cunningham, P. T.; Stone, S. R.; Francis, R. W.; Anastasas, M.; Hall, C. M.; Heinrich, T.; Bogdawa, H. M.; et al. A platform for discovery of functional cell-penetrating peptides for efficient multi-cargo intracellular delivery. *Sci. Rep.* **2018**, *8* (1), 12538.
(14) Kauffman, W. B.; Guha, S.; Wimley, W. C. Synthetic molecular evolution of hybrid cell penetrating peptides. *Nat. Commun.* **2018**, *9* (1), 2568.
(15) Egloff, P.; Zimmermann, I.; Arnold, F. M.; Hutter, C. A. J.; Morger, D.; Opitz, L.; Poveda, L.; Keserue, H. A.; Panse, C.; Roschitzki, B.; Seeger, M. A. Engineered peptide barcodes for in-depth analyses of binding protein libraries. *Nat. Methods* **2019**, *16* (5), 421−428.
(16) Rössler, S. L.; Grob, N. M.; Buchwald, S. L.; Pentelute, B. L. Abiotic peptides as carriers of information for the encoding of small-molecule library synthesis. *Science* **2023**, *379* (6635), 939−945.
(17) Chang, C. D.; Meienhofer, J. Solid-Phase Peptide Synthesis Using Mild Base Cleavage of Nαfluorenylmethyloxycarbonylamino Acids, Exemplified By a Synthesis of Dihydrosomatostatin. *Int. J. Pept. Protein Res.* **1978**, *11* (3), 246−249.
(18) Behrendt, R.; White, P.; Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **2016**, *22* (1), 4−27.
(19) Subirós-Funosas, R.; Prohens, R.; Barbas, R.; El-Faham, A.; Albericio, F. Oxyma: An Efficient Additive for Peptide Synthesis to Replace the Benzotriazole-Based HOBt and HOAt with a Lower Risk of Explosion [1]. *Chem.—Eur. J.* **2009**, *15* (37), 9394−9403.
(20) Huang, Y.-C.; Chen, C.-C.; Li, S.-J.; Gao, S.; Shi, J.; Li, Y.-M. Facile synthesis of C-terminal peptide hydrazide and thioester of NY-ESO-1 (A39-A68) from an Fmoc-hydrazine 2-chlorotrityl chloride resin. *Tetrahedron* **2014**, *70* (18), 2951−2955.
(21) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266* (5186), 776−779.
(22) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013**, *8* (12), 2483−2495.
(23) Wan, Q.; Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem.* **2007**, *119* (48), 9408−9412.
(24) Cifani, P.; Kentsis, A. High sensitivity quantitative proteomics using automated multidimensional nano-flow chromatography and accumulated ion monitoring on quadrupole-Orbitrap-linear ion trap mass spectrometer. *Mol. Cell. Proteomics* **2017**, *16* (11), 2006−2016.
(25) Cifani, P.; Cifani, P.; Dhabaria, A.; Kentsis, A. Fabrication of nanoelectrospray emitters for LC-MS. *Protocol Exchange* **2015**.
(26) Schilling, B.; MacLean, B.; Held, J. M.; Sahu, A. K.; Rardin, M. J.; Sorensen, D. J.; Peters, T.; Wolfe, A. J.; Hunter, C. L.; MacCoss, M. J.; et al. Multiplexed, scheduled, high-resolution parallel reaction monitoring on a full scan QqTOF instrument with integrated data-dependent and targeted mass spectrometric workflows. *Anal. Chem.* **2015**, *87* (20), 10222−10229.
(27) Pino, L. K.; Searle, B. C.; Bollinger, J. G.; Nunn, B.; MacLean, B.; MacCoss, M. J. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* **2020**, *39* (3), 229−244.
(28) Zolg, D. P.; Wilhelm, M.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Delanghe, B.; Bailey, D. J.; Gessulat, S.; Ehrlich, H. C.; Weininger, M.; Yu, P.; Schlegl, J.; Kramer, K.; Schmidt, T.; Kusebauch, U.; Deutsch, E. W.; Aebersold, R.; Moritz, R. L.; Wenschuh, H.; Moehring, T.; Aiche, S.; Huhmer, A.; Reimer, U.; Kuster, B. Building ProteomeTools based on a complete synthetic human proteome. *Nat. Methods* **2017**, *14* (3), 259−262.
(29) Dupont, E.; Prochiantz, A.; Joliot, A. Penetratin story: an overview. *Cell-Penetrating Peptides: Methods and Protocols*; Springer New York, 2011; Vol. *683*, pp 21−29.
(30) Rizzuti, M.; Nizzardo, M.; Zanetta, C.; Ramirez, A.; Corti, S. Therapeutic applications of the cell-penetrating HIV-1 Tat peptide. *Drug Discov. Today* **2015**, *20* (1), 76−85.
(31) Kardani, K.; Bolhassani, A. CPPsite 2.0: An available database of experimentally validated cell-penetrating peptides predicting their secondary and tertiary structures. *J. Mol. Biol.* **2021**, *433* (11), 166703.
(32) Futaki, S.; Suzuki, T.; Ohashi, W.; Yagami, T.; Tanaka, S.; Ueda, K.; Sugiura, Y. Arginine-rich peptides: an abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J. Biol. Chem.* **2001**, *276* (8), 5836−5840.
(33) Hulo, C.; De Castro, E.; Masson, P.; Bougueleret, L.; Bairoch, A.; Xenarios, I.; Le Mercier, P. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.* **2011**, *39* (suppl_1), D576−D582.
(34) Du, Z.; He, Y.; Li, J.; Uversky, V. N. Deepadd: protein function prediction from k-mer embedding and additional features. *Comput. Biol. Chem.* **2020**, *89*, 107379.
(35) Jin, S.-S.; Ju, H.; Jung, H.-J. Adaptive Markov chain Monte Carlo algorithms for Bayesian inference: recent advances and comparative study. *Struct. Infrastruct. Eng.* **2019**, *15* (11), 1548−1565.
(36) Ramaker, K.; Henkel, M.; Krause, T.; Rockendorf, N.; Frey, A. Cell penetrating peptides: a comparative transport analysis for 474 sequence motifs. *Drug Deliv.* **2018**, *25* (1), 928−937.
(37) Takao, S.; Forbes, L.; Uni, M.; Cheng, S.; Pineda, J. M. B.; Tarumoto, Y.; Cifani, P.; Minuesa, G.; Chen, C.; Kharas, M. G.; et al. Convergent organization of aberrant MYB complex controls oncogenic gene expression in acute myeloid leukemia. *Elife* **2021**, *10*, No. e65905.
(38) Liszczak, G.; Muir, T. W. Nucleic Acid-Barcoding Technologies: Converting DNA Sequencing into a Broad-Spectrum Molecular Counter. *Angew. Chem., Int. Ed.* **2019**, *58* (13), 4144−4162.
(39) Matsuzaki, Y.; Aoki, W.; Miyazaki, T.; Aburaya, S.; Ohtani, Y.; Kajiwara, K.; Koike, N.; Minakuchi, H.; Miura, N.; Kadonosono, T.; et al. Peptide barcoding for one-pot evaluation of sequence-function relationships of nanobodies. *Sci. Rep.* **2021**, *11* (1), 21516.

https://doi.org/10.1021/acs.jproteome.4c00068
J. Proteome Res. XXXX, XXX, XXX−XXX