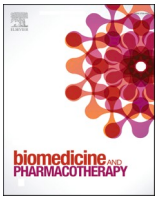
Identification of cell type-specific cell-penetrating peptides through in vivo phage display leveraged by next generation sequencing

Biomedicine & Pharmacotherapy 182 (2025) 117740

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

journal homepage: www.elsevier.com/locate/biopha



Identification of cell type-specific cell-penetrating peptides through *in vivo* phage display leveraged by next generation sequencing

Franziska Kohl [a,b], Oliver Laufkötter [c], Mike Firth [d], Luc Krimpenfort [b], Priyanka Mangla [e], Mohammadhassan Ansarizadeh [f,g,h], Gökçe Geylan [i,j], Lauri Eklund [f,g,h], Leonardo De Maria [k], Lars Jakobsson [b,1], John Wiseman [a,1,*]

[a] *Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*
[b] *Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden*
[c] *Department of Life Science Informatics, B-IT, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany*
[d] *Data Sciences and Quantitative Biology, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK*
[e] *Oligonucleotides and Targeted Delivery, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*
[f] *Oulu Center for Cell-Matrix Research, University of Oulu, Oulu, Finland*
[g] *Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland*
[h] *Biocenter Oulu, University of Oulu, Oulu, Finland*
[i] *Molecular AI, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*
[j] *Division of Systems and Synthetic Biology, Department of Life Sciences, Chalmers University of Technology, Gothenburg, Sweden*
[k] *Research and Early Development, Respiratory & Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*

ARTICLE INFO

Keywords:
Cell-penetrating peptides
In vivo phage display
Vascular anomalies
Next generation sequencing
Bioinformatic analysis

ABSTRACT

Vascular anomalies (VA) refer to abnormal blood or lymphatic vessel architecture, most often as a result of dysregulated growth. Venous malformations (VM), a subgroup of VAs, are triggered by activating mutations in the Angiopoietin/TIE2-PI3K/AKT/mTOR signaling pathway with TIE2 L914F (gene name TEK) being one of the most frequent mutations in patients with VMs. Although systemic targeting of the overactivated pathway is possible, it would be a therapeutic advantage to restrict treatment to only the affected lesions. To identify peptides with potential selective binding to TIE2 L914F lesions we applied *in vivo* phage display to TIE2 L914F-overexpressing endothelial cells (ECs) in a subcutaneous matrigel xenograft mouse model of VMs. By panning for lesion-targeting phages in combination with subcellular fractionation, a screen for cell-penetrating candidate phages was established. Employing Next Generation Sequencing (NGS) and a refined bioinformatic analysis we were able to identify many novel cell-penetrating peptides (CPPs). To pinpoint the most selective and viable CCP candidates a hierarchical clustering algorithm was utilized. This method aggregated CPPs with highly similar sequences into a small number of clusters from which consensus sequences could be derived. Selected candidate CPPs exhibited uptake in TIE2 L914F-expressing human umbilical vein endothelial cells (HUVEC) in culture and were able to deliver siRNA into these cells. In conclusion, our NGS bioinformatic-supported approach led to the identification of novel and selective CPPs capable of transporting a siRNA cargo into targeted cells.

1. Introduction

Vascular anomalies (VA), i.e. vascular malformations, often arise as a consequence of abnormal vascular patterning in early developmental phases. VAs can arise from all types of vessels (arteries, veins, capillaries, lymphatic, and combined forms) and are usually named after the affected vessel type. In patients, venous malformations (VM) manifest in dilated, vein-like vessels causing swelling, discoloration and pain. Somatic mutations in the gene *TEK*, encoding the TIE2 receptor, have been identified as a frequent cause of VMs in patients, with a change of leucine to phenylalanine in position 914 the most frequent alteration. This point mutation leads to autophosphorylation and constant activation of downstream signaling, most notably within the phosphoinositide 3-kinase (PI3K)/AKT and the mitogen-activated protein kinase (MAPK)

* Corresponding author.
E-mail address: John.W.Wiseman@astrazeneca.com (J. Wiseman).
[1] These two authors contributed equally to this work

https://doi.org/10.1016/j.biopha.2024.117740
Received 26 August 2024; Received in revised form 18 November 2024; Accepted 3 December 2024
Available online 12 December 2024

pathways in ECs [1,2]. Previous attempts to target this receptor or its downstream effectors to treat VMs have been partly successful but not curative [3–6] showing a need for improved therapy.

Cell-penetrating peptides (CPPs) are small peptides (5–30 amino acids) that can translocate from outside of the cell into the cytoplasm, either through direct membrane penetration or utilizing the endocytotic machinery. Although the precise mechanisms are not fully understood such peptides have been shown to promote delivery of conjugated molecules, such as RNA, into cells for potential modulation of their properties. To restrict uptake of CPPs to specific cell populations in pathology would open for targeted treatment, thereby reducing side effects and increasing efficacy. Vasculature-homing peptides, identified by *in vitro* and *in vivo* phage display, have been discovered binding specifically to vasculature of different organs, mapping the entire vasculature in homeostasis and targeting certain cardiovascular disease states [7]. Specific CPPs can be selected to help overcome tissue and cellular barriers, potentially improving peptide-mediated delivery of therapeutic agents. To successfully exploit the full potential of nucleotide-based formulations as novel therapeutic interventions, such as antisense oligonucleotide (ASO) [8,9], clustered regularly interspaced short palindromic repeats (CRISPR) gene editing [10] and short interfering RNA (siRNA) [11], we require strategies to effectively deliver oligonucleotides to specific tissues or cell types. Utilizing such strategies would promote treatment specificity thereby reducing side and off-target effects and lower drug doses [12]. Combining oligonucleotide therapies with a tissue specific CPP might help create an innovative way to deliver DNA and RNA based drugs precisely to a desired location [13,14].

Phage display, a commonly used technique to identify targeting peptides, was in this study applied to a xenograft mouse model of VMs for the first time. Phages are viruses that only infect and replicate in bacteria and in phage display protocols they are used as carriers of peptides [15], antibodies [16,17], cDNA [18,19] or similar modalities. Insertion of synthetic DNA sequences into the wildtype (WT) coat protein-coding gene results in phages displaying a modified surface peptide, which can in turn be assessed for specific binding abilities. Through insertion of random DNA sequences, libraries of millions of phages, each displaying a different peptide on their surface, can be generated. Multiple rounds of biopanning with such diverse libraries, against a biological target (e.g. a cell line or specific tissue), enriches for phages with strong binding or cell-penetrating properties.

Sanger sequencing, traditionally used to identify specific peptide sequences following phage display, is limited by very low throughput. Typically, at the end of each panning round 30–300 phage plaques, out of a pool of potentially millions, are hand-picked and analyzed by Sanger sequencing. This is a drawback which potentially overlooks promising candidate sequences and can lead to the introduction of bias. In comparison, NGS allows us to analyze millions of sequences, increasing the throughput 10000-fold, with near full library coverage. We applied NGS followed by bioinformatic processing and ultimately combined similar sequences into consensus peptides. A similar approach of combining phage display and NGS for probing polypropylene surfaces successfully identified sequences which could increase adhesive strength [20].

In our study, we aimed to develop a novel analysis method for *in vivo* phage display to increase the probability of identifying large numbers of high quality CPPs. With this approach we identified peptides that can penetrate targeted cell lines and at the same time deliver oligonucleotide-based therapeutics into these cells.

2. Results

2.1. In vivo phage display on xenograft mouse model of VM

TIE2 L914F point mutations in ECs have been found to cause VMs in the form of focal vessel enlargement. Towards the goal of identifying potential lesion-specific binders, we implanted matrigel plugs, containing either wild type or TIE2 L914F mutated HUVECs, under the skin of immunocompromised mice (Fig. 1A). After one week of growth, implants of TIE2 L914F HUVECs were dark red in color (blood filled) while those containing WT HUVECs were pale (Fig. 1B). This is in accordance with previously published data [4]. To analyze the xenografts in more detail, samples were immunostained for the EC marker CD31, made transparent, and imaged by light sheet fluorescence microscopy (LSFM). This exposed three-dimensional (3D) vessel networks within the plugs. To distinguish between HUVEC-derived vessels and potential invading host-derived vessels, we applied species-specific antibodies to stain extracted matrigel plugs (Suppl. Video S1). To promote visualization of endothelial structures in 3D we applied the filament tracer module and machine learning-guided algorithms in the Imaris 10 software (Suppl. Video S2). Images revealed major contributions by HUVECs, irrespective of genotype, to the vascular networks of matrigel plugs. These observations indicate that the host circulation, at least partly, connects to the vascular networks made of HUVECs in matrigels. These connections would facilitate panning of injected phages for binding to vascular structures derived from manipulated HUVECs within the plugs. To this end, one week after implantation, 2 mice per group were intravenously injected with one of our phage libraries, either PhD-C7C, PhD-12, TriCo-16, or TriCo-20. Mice were terminated 30 minutes after injection, allowing time for phages to bind. Libraries differ in insert length (7, 12, 16 or 20 amino acids), and structure (linear or cyclic) and for each library three rounds of biopanning were performed using the enriched phage pool from the xenograft of the previous panning round (Fig. 1C).

Supplementary material related to this article can be found online at doi:10.1016/j.biopha.2024.117740.

2.2. NGS and subsequent bioinformatic data analysis led to identification of putative CPPs

To analyze phage biopannings in greater depth than afforded by Sanger sequencing, we sequenced phage pools after each panning round with NGS and developed a downstream bioinformatic analysis pipeline to identify TIE2 L914F HUVEC-specific peptides. In order to build up an initial data set from which we could develop a bioinformatic analysis script, we sequenced all three panning rounds from xenografts, control organs, and the naive library that we initially injected into the animals. This dataset contained 1.73×10^7 total sequences, from which we eliminated sequences with premature stop codons, cross contaminants from other libraries, and combined sequences translating the same peptide sequence. Following these filtering steps 2×10^6 sequences remained, of which 1.7×10^5 were unique.

Analysis of all peptides identified from phages in xenografts in the first round, showed that 80 % were also found in the liver (Fig. 2A). The liver retained more than 10 times the number of unique peptide sequences found in xenografts, highlighting the anticipated non-specific binding of phages in this organ [21]. However, the number of sequences identified in the liver following panning rounds 2 and 3 decreased dramatically, confirming the power of biopanning in minimizing non-specific binding. Additionally, we saw that peptide count tended to decrease over the three selection rounds (Fig. 2A), a sign of both decreased non-specific binding and peptide enrichment. Analysis of the contribution of each library to unique peptides found in TIE2 L914F HUVEC xenografts showed that TriCo-20 specific sequences were most frequent, approximately 50 % in each tissue (Fig. 2B). To determine which peptides were specific to the TIE2 L914F mutated xenograft, we developed a python script to eliminate peptides found in TIE2 WT xenograft, kidney and liver with more than 0.01 % occurrence (Fig. 2C).

We also wanted to eliminate peptides common to both TIE2 point mutation and WT xenografts, to ensure that the peptides from the panning rounds on TIE2 L914F HUVECs are specific to the mutated cells and do not bind to ECs in general. To control for this, we implanted during the third selection round a TIE2 WT HUVEC xenograft in the

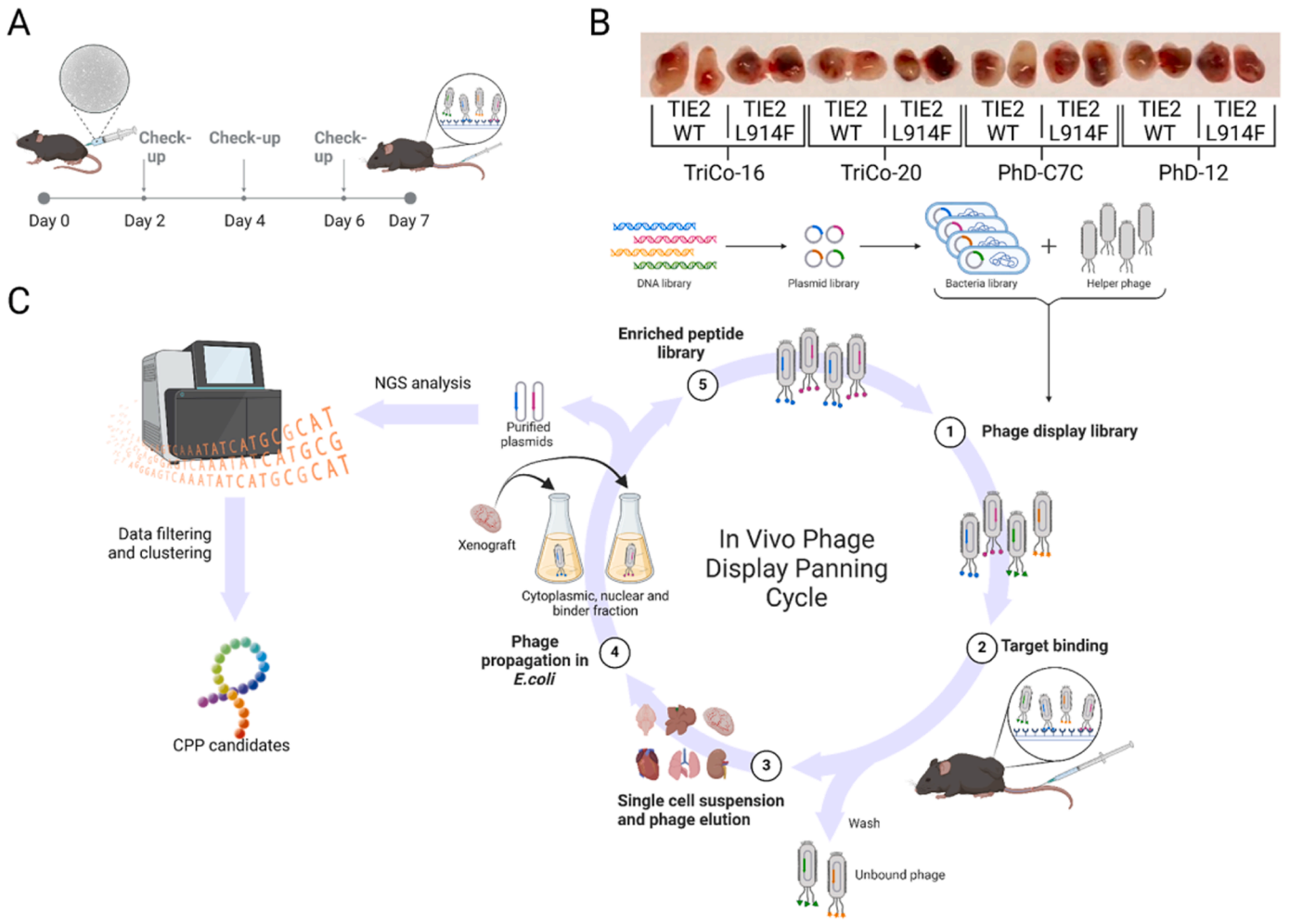


Fig. 1. Xenograft model for VMs and *in vivo* phage display. (A) Schematic overview of xenograft mouse model. (B) Ex vivo xenografts of TIE2 WT and TIE2 L914F HUVECs. (C) *In vivo* phage display scheme. Mice carrying a xenograft for 7 days were administered phage libraries intravenously. After perfusion to remove non-specific phages, target and control organs are dissected and specific phages recovered, amplified and purified for further enrichment. A total of 3 selection rounds were performed with phages enriched from TIE2 L914F HUVEC xenografts as input for each round.

same mouse on the opposing flank. A total of 759 peptides were identified as generally EC-specific, accounting for approximately 18 % of TIE2 L914F HUVEC peptides. In addition, we excluded all peptides with an occurrence of greater than 0.01 % in liver and kidney. In both organs combined, a total of 5656 peptides were detected. This step eliminated only 63 peptides accounting for 0.019 % of the HUVEC TIE2 L914F peptides, leaving 3292 unique peptides for further analysis. Thus, demonstrating minimal overlap between xenograft derived and control tissue derived peptides. Providing further evidence that non-specific binding was greatly reduced over three rounds of phage display and that peptides selected at this point of the analysis were enriched for the xenografts.

To better understand sequence diversity of the remaining TIE2 L914F HUVEC specific peptides, we performed a hierarchical clustering analysis. By comparing two peptides, a similarity score was calculated and applying this to all possible pairings led to the formation of a matrix, containing the similarity scores of all combinations. Various algorithms can be used to calculate the similarity score between two peptide sequences but due to the size of the dataset, a computationally efficient formula was required. First, we aligned each pair of peptides followed by a residue comparison at each position in the alignment. For example, if the first pair of amino acids were the same they were given a score of 1.5, but if they were different a score of 0. Additionally, if a gap opening or gap extension occurred due to the alignment, negative scores of −0.5 and −0.1 were given respectively. Finally, these scores were added together to give a total similarity score between the two sequences. This calculation is repeated for all peptide pairs of each library. Since the peptides of each library had differing lengths, similarity scores could not be compared and therefore needed normalization. Scores were normalized by dividing each similarity score by the highest similarity score present in each respective library and subsequently converted to a distance score by taking the inverse of each value (1 - sim score). This resulted in the final distance matrix that we required for input into the hierarchical clustering algorithm.

Using the hierarchical clustering data, dendrograms were generated to visualize peptide sequence diversity and distribution (Fig. 3A and Suppl. fig. S1A-C). To determine the quality of the clustering the cophenetic correlation coefficient (CP) was calculated. The CP values for the TIE2 L914F clusterings ranged from 0.89 to 0.98 which corresponds to high clustering quality. Following analysis of the dendrograms a general rule was devised to determine a distance cut off value to form distinct clusters for each of the four peptide libraries. This rule was set to be $\frac{max distance}{6}$, where max distance refers to the maximum distance between any two clusters. Using this formula peptides were divided into distinct clusters, typically consisting of up to 100 peptides with some exceptions climbing into the hundreds and a few containing only one peptide. Given that our selected peptides both came from the TriCo-20 library, the associated dendrogram is shown in Fig. 3A. Here, the

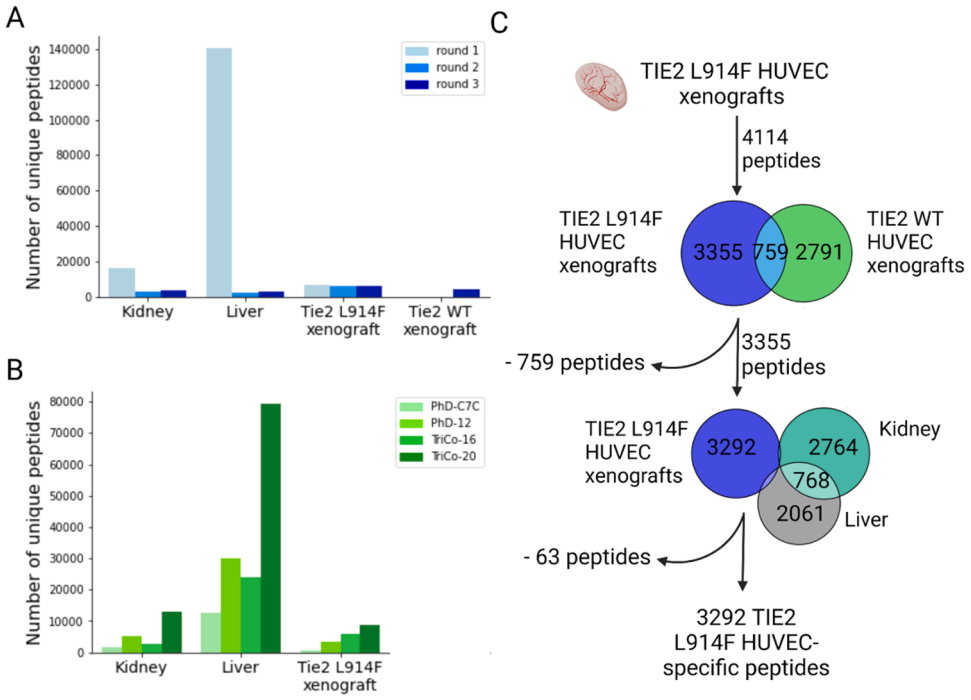


Fig. 2. Bioinformatic analysis. (A) Bar graphs representing the number of individual peptides over three rounds of selection on TIE2 L914F mutated HUVECs. Only in the third selection round, a TIE2 WT HUVEC xenograft was grown in the opposite flank of the animals to determine how specific the phage pool was for the TIE2 L914F-mutated HUVECs. (B) Bar graphs representing total identified number of peptides derived from NGS analysis of all three selection rounds on TIE2 L914F mutated HUVECs related to origin from different libraries. (C) Schematic overview of the strategy to identify CPPs specific to TIE2 L914F-mutated HUVEC xenografts. After three rounds of panning, peptides were eliminated from further analysis if found in HUVEC TIE2 WT xenografts. Subsequently peptides which overlapped with peptides found in kidney and liver were excluded, leaving only TIE2 L914F HUVEC-specific peptides.

dashed black line refers to the predetermined cut off value mentioned above. Each intersection of this line with a vertical line of the dendrogram corresponds to one peptide cluster which contains all connected leaves (peptides) below this point. In general, we extracted 53 clusters from this biopanning. We considered clusters with less than 1000 reads to be unreliable hits and in total 10 reliable clusters were identified (Table 1a). Considering the contributions from different libraries, we realized a trend that the cyclic peptides from library PhD-C7C were found most frequently, next was linear peptides from library TriCo-20, followed by TriCo-16 and least frequent was PhD-12. Even though library PhD-C7C contributed 4 out of 10 hits, we decided to focus on the peptides with the highest read numbers, which came from library TriCo-16 and TriCo-20.

The clusters contained hundreds of similar peptide sequences, too many to test in comprehensive cell culture studies, hence we generated consensus sequences from each cluster by employing probability logos (pLogos) (Fig. 3B). We implemented a weighted approach to consider that different sequences had different numbers of reads. The size of the amino acid letters in the pLogos reflect the frequency of this amino acid at a certain position in the peptide sequence. We then determined the promiscuity of the consensus sequences by cross referencing the pLogos with in house data from biopannings against cell lines where the same libraries had been tested. This analysis identified the largest cluster from the TriCo-16 library as a promiscuous peptide (TX1), given that it appeared in several different in-vitro phage display experiments on VA-model cell lines (data not shown), and was therefore excluded from further analysis.

The second and third top-ranked CPP candidates TX2 and TX3, both from library TriCo-20, showed no promiscuous sequences. Analysis of library TriCo-20 identified 13 clusters, 3 of which had more than 1000 reads (Table 1b). The number of unique peptides, all without promiscuous sequences, within the 13 clusters ranged between 1 and 1110 with the largest cluster having a total of approximately 14000 reads. This indicates that the majority, i.e. 1110 out of 1995 unique peptides, from this biopanning share a high degree of similarity. We then checked whether the consensus sequence existed in the pool of sequences used to generate the consensus and realized for TX2 that this was not the case. Thus, we decided to use the sequence with the highest similarity and highest occurrence rate (reads) in the respective cluster. Finally, after all selection steps we chose two peptides from reliable clusters for further testing. Both peptides originated from the TriCo-20 library and are referred to as TX2 and TX3.

Analysis of the TIE2 L914F HUVEC specific peptide count over three selection rounds revealed a reduction in non-specific CPPs evidenced by a decrease in unique peptide count as biopanning progressed. 30 % of peptides were common across all biopannings (Fig. 3C) but these peptides accounted for 98 % of the reads (Fig. 3D), suggesting minimal significance to the peptides found only in rounds one or two. In the third round of selection, we detected an increase in the number of reads, indicating an enrichment of peptides in the last round of biopanning. To confirm, we checked the two most frequently identified peptides and found that both were enriched, approximately 75 % for peptide TX2 and 50 % for peptide TX3 (Fig. 3E).

2.3. *Examination of physicochemical properties and 3D structures of CPPs reveal similarity between TX2 and penetratin*

To what extent a peptide can penetrate the cell membrane is mainly defined by its physicochemical properties. Based on charge and hydrophobicity CPPs can be classified as cationic, amphipathic or hydrophobic peptides. Well characterized examples of CPPs, such as transactivating transcriptional activator (TAT) and penetratin, are cationic CPPs, characterized by an excess of positively charged amino acids [22–24].

We determined five physicochemical parameters for the best 10 candidates that were derived from bioinformatic analysis and the same

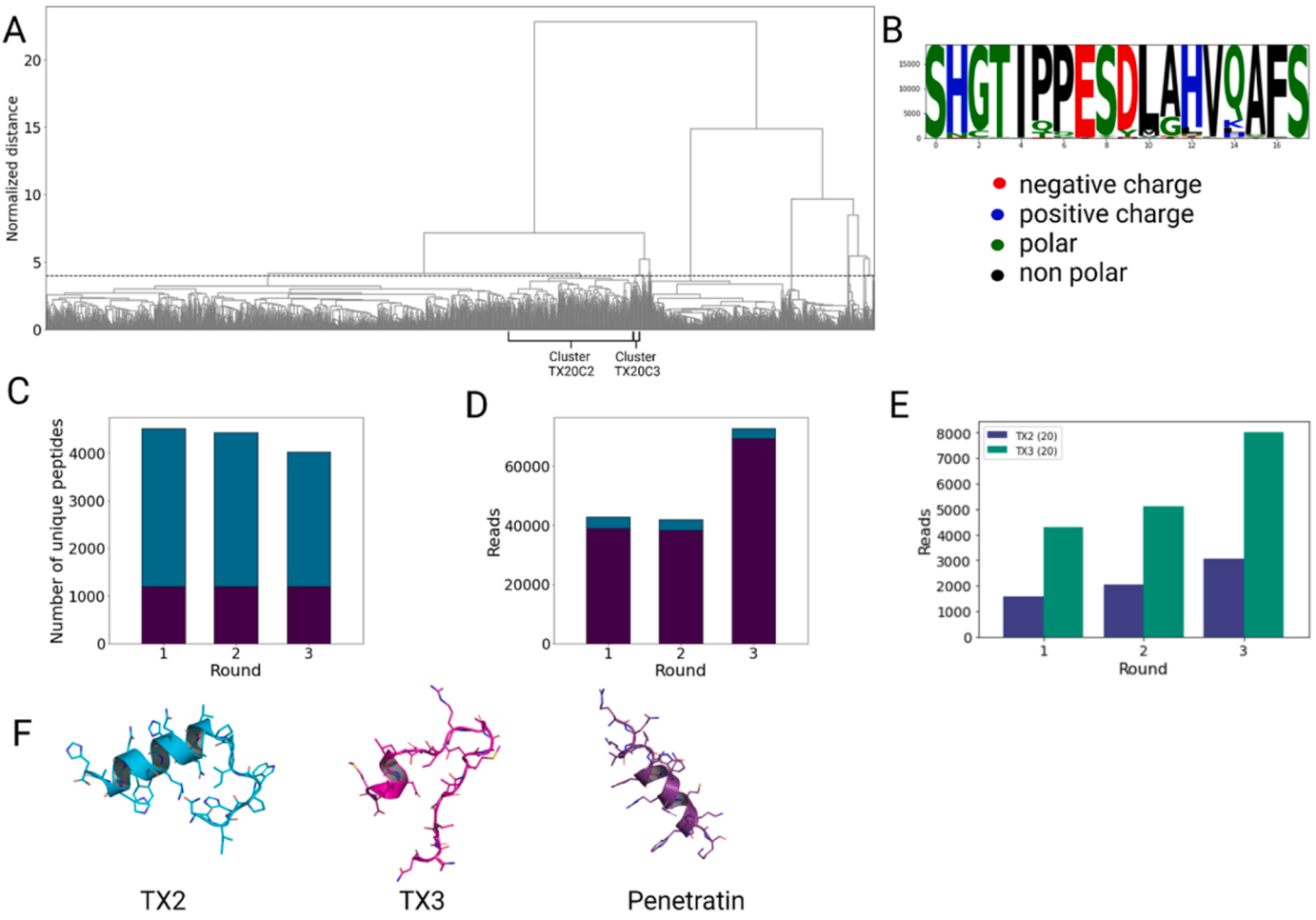


Fig. 3. Hierarchical clustering of TriCo-20 peptide sequences and enrichment analysis. (A) Dendrogram depicting the results of the hierarchical clustering analysis of peptides from library TriCo-20 which yielded the two top candidates. The distance cut off used to form the clusters is indicated by the horizontal dashed black line and the position of clusters TX20C2 and TX20C3 is outlined by black boxes. (B) Representative consensus of all sequences within a cluster were created using pLogos, with TX1 pLogo shown and amino acids color coded for negative charge (red), positive charge (blue), polar residues (green), and non-polar amino acids (black). (C) Enrichment analysis of the number of total peptides over three rounds (green) and peptides that were found in all three rounds (blue). (D) Stacked bar graph comparing enrichment of unique peptide reads found in all three rounds (blue) to reads of the peptides found in all three rounds. (E) Enrichment of selected peptides from TX2 and TX3 by following read numbers over three consecutive rounds of selection. (F) AlphaFold2 prediction of peptide tertiary structures for TX2, TX3, and penetratin.

Table 1a
Top 10 ranking of CPP candidates of all libraries based on their number of reads.

Peptide acronym	Library	Cluster ID	Number of reads	Number of unique peptides	Consensus sequence promiscuity
TX1	TriCo−16	TX16C12	18994	759	yes
TX2	TriCo−20	TX20C1	13927	1110	no
TX3	TriCo−20	TX20C7	10667	321	no
TX4	PhD-C7C	TX7C1	8365	36	yes
TX5	PhD−12	TX12C2	6250	325	yes
TX6	PhD-C7C	TX7C3	3379	23	yes
TX7	TriCo−20	TX20C8	3011	153	no
TX8	PhD-C7C	TX7C4	1452	19	yes
TX9	PhD-C7C	TX7C5	1449	14	yes
TX10	TriCo−16	TX16C2	1191	61	yes

Table 1b
Top 10 ranking of CPP candidates of library TriCo-20 based on their number of reads.

Peptide acronym	Cluster ID	Number of reads	Number of unique peptides	Consensus sequence promiscuity
TX2	TX20C1	13927	1110	no
TX3	TX20C7	10667	321	no
TX7	TX20C8	3011	153	no
TX14	TX20C2	339	304	no
TX19	TX20C10	145	19	no
TX25	TX20C9	110	38	no
TX33	TX20C3	23	23	no
TX38	TX20C4	15	15	no
TX40	TX20C11	10	4	no
TX46	TX20C5	3	3	no

parameters for our positive control penetratin (Table 2). We checked the peptides´ isoelectric point (pI), which is the pH at which a molecule has no electrical charge [25]. Here, penetratin stood out with a pI at pH12. Our top 10 peptides were divided in two groups with half having a pI at acidic pH and half at basic pH. TX2 had a rather high pI with 9.78 and interestingly TX3 was at the other end of the spectrum with a pI at 4.37. In correlation with that the net charge at neutral pH7 [26] was high for penetratin with a positive charge of 6.75, indicating that the peptide contains an excess of positively charged amino acids. TX2 and TX3 followed the trend from the pI with TX2 being positively charged and TX3 negatively charged at pH7.

Table 2
Physicochemical properties and secondary structure information of the top 10 peptide candidates.

Peptide acronym	Isoelectric point (pI)	Charge (pH7)	GRAVY index	Aliphatic index	Boman index	Secondary Structure Content (%)	Secondary Structure Type
TX1	5.14	−2.36	−0.26	70.56	1.21	62.50	α-helical
TX2	9.78	1.11	−1.00	78.00	1.86	60.00	α-helical
TX3	4.37	−1.54	−0.23	80.00	1.50	25.00	α-helical
TX4	6.46	−0.47	0.04	32.67	0.73	0.00	none
TX5	8.46	0.46	−0.65	51.76	1.24	0.00	none
TX6	8.89	1.44	−0.37	32.67	0.94	0.00	none
TX7	6.78	−0.28	−1.23	48.64	2.42	70.00	α-helical
TX8	7.81	0.53	−0.47	13.33	1.47	0.00	none
TX9	5.55	−0.56	−0.98	6.67	2.67	0.00	none
TX10	9.99	1.46	−0.70	53.89	2.11	75.00	α-helical
Penetratin	12.00	6.75	−1.73	48.75	4.05	93.80	α-helical

The GRAVY (Grand Average of Hydropathy) index [27] is a measure of hydrophobicity/hydrophilicity of peptide sequences. It is calculated by dividing the sum of all hydrophobicity values of the peptide through the number of amino acids. The more negative the GRAVY index, the less hydrophobic the corresponding peptide. For example, penetratin and TX2 have negative GRAVY indices with −1.73 and −1.0, respectively, and are not considered hydrophobic peptides. While TX3 with an index of −0.23 can be considered a hydrophobic peptide. The aliphatic index from Ikai determines the relative volume of a peptide that is occupied by aliphatic side chains. Considered as aliphatic side chains are alanine, valine, isoleucine, and leucine [28]. Here, TX2 and TX3 had very similar aliphatic values around 80, and penetratin had a rather low value of approximately 49. This is the only parameter where TX2 did not show similar values as penetratin. We then examined the Boman index [29] of the CPPs, indicating the potential of a peptide to interact with proteins, e.g. receptors, or membranes. High interaction potential can be expected for values over 2.48. Penetratin showed a very high likelihood of protein interaction with a Boman index of 4.05. TX2 and TX3 did not exceed the threshold value, implying a lower potential for protein interaction.

In addition, AlphaFold2 was used to predict secondary structures for our peptides with a focus on the presence of α-helical structures or β-sheets [30]. Approximately half of the peptides possessed α-helical structures, but no β-sheets could be identified (Fig. 3F and Suppl. fig. S1D). AlphaFold2 predicted that 93.8 % of penetratin was involved in an α-helix and showed that TX2 had great similarity with penetratin, with approximately 60 % of TX2 amino acids involved in forming an α-helix, whereas TX3 had only 25 % α-helical content (Table 2).

Generally, the physicochemical properties and secondary structure elements of TX2 were in line with penetratin's characteristics. Both are cationic peptides with few hydrophobic amino acids and mainly α-helical secondary structure.

2.4. Database searches uncover novelty of CPP candidates

BLASTp [31] searches were performed to infer functionality of TX2 and TX3 peptides by relating to annotated, functional information of similar sequences but no hits were obtained, highlighting the novelty of these sequences. In addition, we ran searches on UniProtKB and Propedia to find potential binding partners. UniProtKB [32], with 226 million sequences, is the world's leading high quality, comprehensive and freely accessible resource of protein sequence and function information and searches with TX2 and TX3 as reference did not yield any matches. Propedia v2.3 [33] is a peptide protein interaction database hosting more than 49,000 complexes and we found no known interaction partner for TX2 or TX3. To determine if TX2 and TX3 had similarities to previously identified CPPs, we searched using CPPSite 2.0 [34], a database consisting of 1855 CPP entries. This search yielded no correlation with other CPPs in the database, again stressing the novelty of our peptides.

Furthermore, we explored TX2 and TX3 with phage display-interrogating tools. Scanner And Reporter Of Target-Unrelated Peptides (SAROTUP) is an online resource that contains a collection of different tools to identify target-unrelated binders [35,36]. We used TUPScan, a tool based on already known unrelated binder motifs, to exclude that TX2 and TX3 contain such motifs. In combination with PhD7Faster 2.0 tool, we gained more insight as to whether TX2 and TX3 have been artificially enriched through growth advantages. Some peptide inserts give phages a growth advantage over other phages leading to a higher copy number in the phage pool. PhD7Faster 2.0 tool rates peptide insert sequences for their phage growth kinetics in bacteria and categorizes peptides with a probability over 50 % as fast-growing peptides. The probability for TX2 and TX3 in the PhD7Faster 2.0 tool was 9.8 % for TX2 and 12.1 % for TX3 respectively, suggesting that they do not introduce a growth advantage (Suppl. table S2). Our analyses revealed that TX2 and TX3 contain no previously identified binding motifs and are unlikely to be enriched through bacterial growth advantage and gives us confidence that these peptide sequences were enriched through target-related binding.

2.5. Candidate peptides exhibit cell-penetrating properties towards HUVECs utilizing a dynamin-dependent endocytosis mechanism

Peptides TX2 and TX3 were labeled N-terminally with a tetramethylrhodamine (TAMRA) fluorophore, and their ability to penetrate cells used in our *in vivo* biopanning studies was assessed. Penetratin, a well characterized CPP derived from Antennapedia homeodomain of Drosophila homeoprotein [37], was labeled with TAMRA and used as a positive control. In addition, WM9, a peptide shown from previous internal studies to lack cell-penetrating properties, was labeled with TAMRA and applied as negative control. CellMask Deep Red, a dye which stains the plasma membrane, was used to determine peptide localization within the cell. Uptake assays on TIE2 L914F HUVECs at 10 μM revealed membrane penetration of TX2 and TX3 peptides (Fig. 4). The negative control WM9 and TAMRA alone did not yield any uptake while the positive control penetratin exhibited high uptake in all cell compartments. Interestingly, the subcellular distribution appeared different for both tested CPPs. TX3 was found in the cytosol, often in proximity to the plasma membrane with a granular distribution pattern, whereas TX2 showed enrichment in the nucleus (Fig. 4). Precise targeting to the nucleus is of particular importance for targeted transport of therapeutic modalities which have their site of action in the nucleus, e.g. CRISPR-associated (Cas) enzymes, guide RNAs (gRNAs), and ASOs.

To determine cell type specificity of TX2 and TX3, we tested uptake in the following cell lines: TIE2 WT HUVECs, Neuro-2a, and BJ fibroblasts (Suppl. Fig. S2A-D). Penetratin uptake was, as expected, not cell type specific and exhibited similar fluorescence intensity in all cell lines tested. At a concentration of 10 μM TX2, we saw strong uptake in TIE2 WT HUVECs and no fluorescence uptake in Neuro-2a cells and BJ fibroblasts. For TX3 at the same concentration, we detected only weak uptake into TIE2 WT HUVECs and again no uptake into Neuro-2a cells and BJ fibroblasts, indicating that TX2 and TX3 are HUVEC-specific CPPs that localize to the nucleus (TX2) or cytoplasm (TX3).

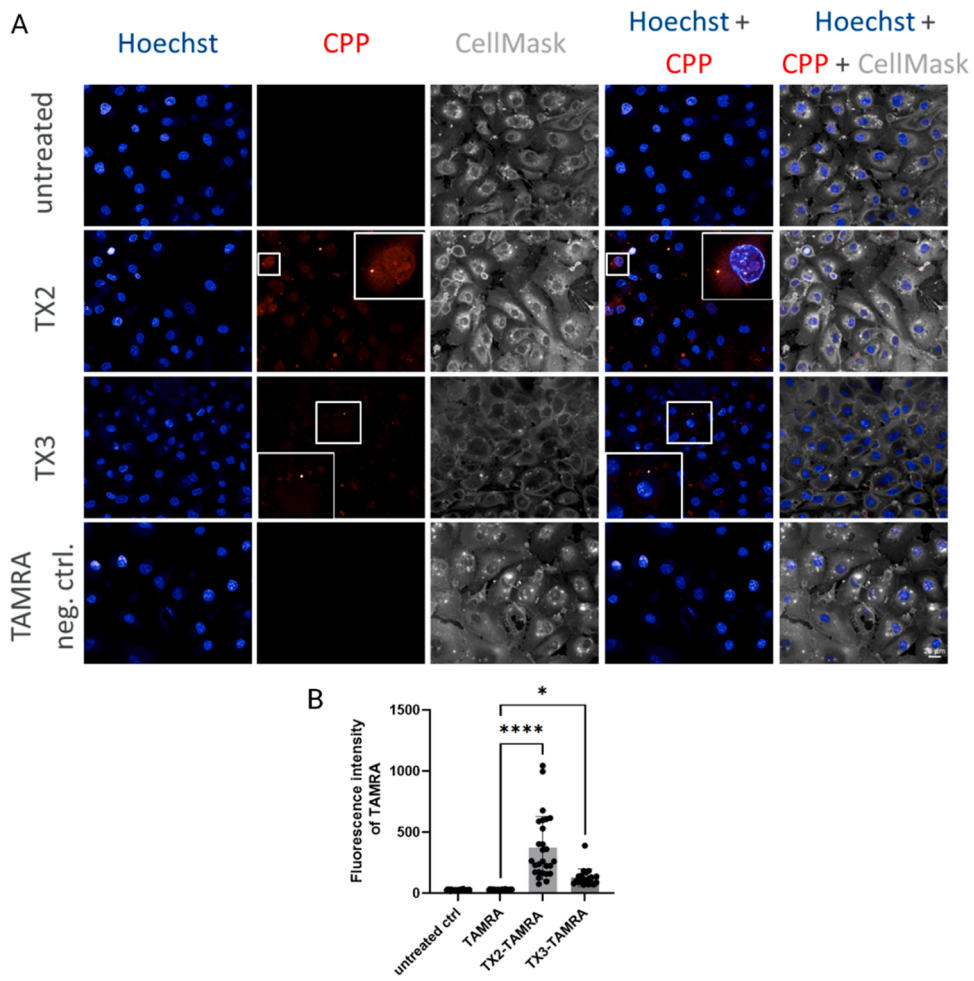


Fig. 4. Cellular uptake of peptides TX2 and TX3. (A) Confocal microscopy images depicting cellular uptake of TX2 and TX3 (red) at 10 μM into TIE2 L914F HUVECs at 37 °C to determine if the peptides have been taken up. To demonstrate that the peptides (red) were localized within cells, the cell membrane was visualized using Cell Mask deep red (gray). The nucleus was stained with Hoechst 33342 (blue). (B) Quantification of CPP uptake assay with n = 3, *p < 0.05, ****p < 0.0001.

Furthermore, we tested the EC specificity with two other EC cell lines of murine and human origin, i.e. bEND3 cells and HDLECs, confirming uptake of TX2 and TX3 (Suppl. Fig. S2E-G).

To investigate the potential uptake mechanism of both peptides, we performed a co-incubation of CPPs with Dextran 10 kDa to understand if the peptides are taken up by direct penetration or if endocytosis mechanisms are utilized. For TIE2 L914F HUVECs this co-incubation revealed that TX2 and TX3 are co-localizing with Dextran 10 kDa, which means that endocytosis is involved in the CPPs´uptake (Suppl. Fig. S3A). We then again used Cell Mask Deep Red to assess localization of peptides and in addition the endocytosis inhibitor Dynasore, which blocks dynamin-dependent uptake mechanisms [38]. Dynasore pre-treatment of cells resulted in a reduction of fluorescence signal for TX2 and TX3 (Suppl. fig. S3B, C) Furthermore, we tested Pitstop 2, a clathrin inhibitor, and EIPA, a macropinocytosis inhibitor, to potentially identify one of the different endocytosis pathways as main uptake mechanism. The uptake mechanism of penetratin is known to be complex, but at higher concentrations mainly facilitated by receptor-mediated endocytosis [39] and, in accordance with this, we found penetratin uptake strongly affected by the Dynasore pre-treatment (Suppl. fig. S3B, C). Uptake of TX2 and TX3 were decreased by all three different endocytosis inhibitors ((Suppl. fig. S3B-D). Considering these findings, we concluded that TX2 and TX3 utilize multiple endocytosis pathways for entering cells.

2.6. Conjugation of CPP TX2 and TX3 to siRNA reveals functional delivery of siRNA into HUVECs

Achieving the functional delivery of cargo molecules into the cell by novel CPPs is important to demonstrate, particularly to specific cell compartments such as the cytoplasm or nucleus. Towards this goal we conjugated TAMRA-labeled TX2 and TX3 to superoxide dismutase 1 (SOD1) siRNA via click chemistry. SOD1, an enzyme that breaks down free superoxide radicals and is therefore involved in oxidative stress [40], is ubiquitously expressed and furthermore plays a role in apoptosis in HUVECs [41]. We incubated TIE2 WT and TIE2 L914F HUVECs with 1 μM of the SOD1 siRNA-CPP-TAMRA conjugates for 18 h and evaluated uptake with confocal microscopy (Fig. 5A, B). We were able to detect uptake of the conjugates with a similar distribution pattern as seen with unconjugated CPPs. TX2 conjugates were found in the nucleus with additional vesicle-like structures in the cytosol and TX3 conjugates were mainly located in the cytosol within similar vesicle-like structures. Both TX2 and TX3 conjugates showed high levels of uptake into TIE2 WT and TIE2 L914F HUVECs.

To determine functionality, we measured SOD1 expression for siRNA mediated knockdown via quantitative PCR (qPCR) after 48 h of incubation with TX2 and TX3 conjugates (Fig. 5C). The TX2-conjugate showed significant knockdown of SOD1 in both HUVEC cell lines with more than 40 % reduction in WT cells and about 20 % reduction in mutated HUVECs and a much less pronounced reduction of expression in BJ fibroblasts. The TX3-conjugate on the other hand showed significant

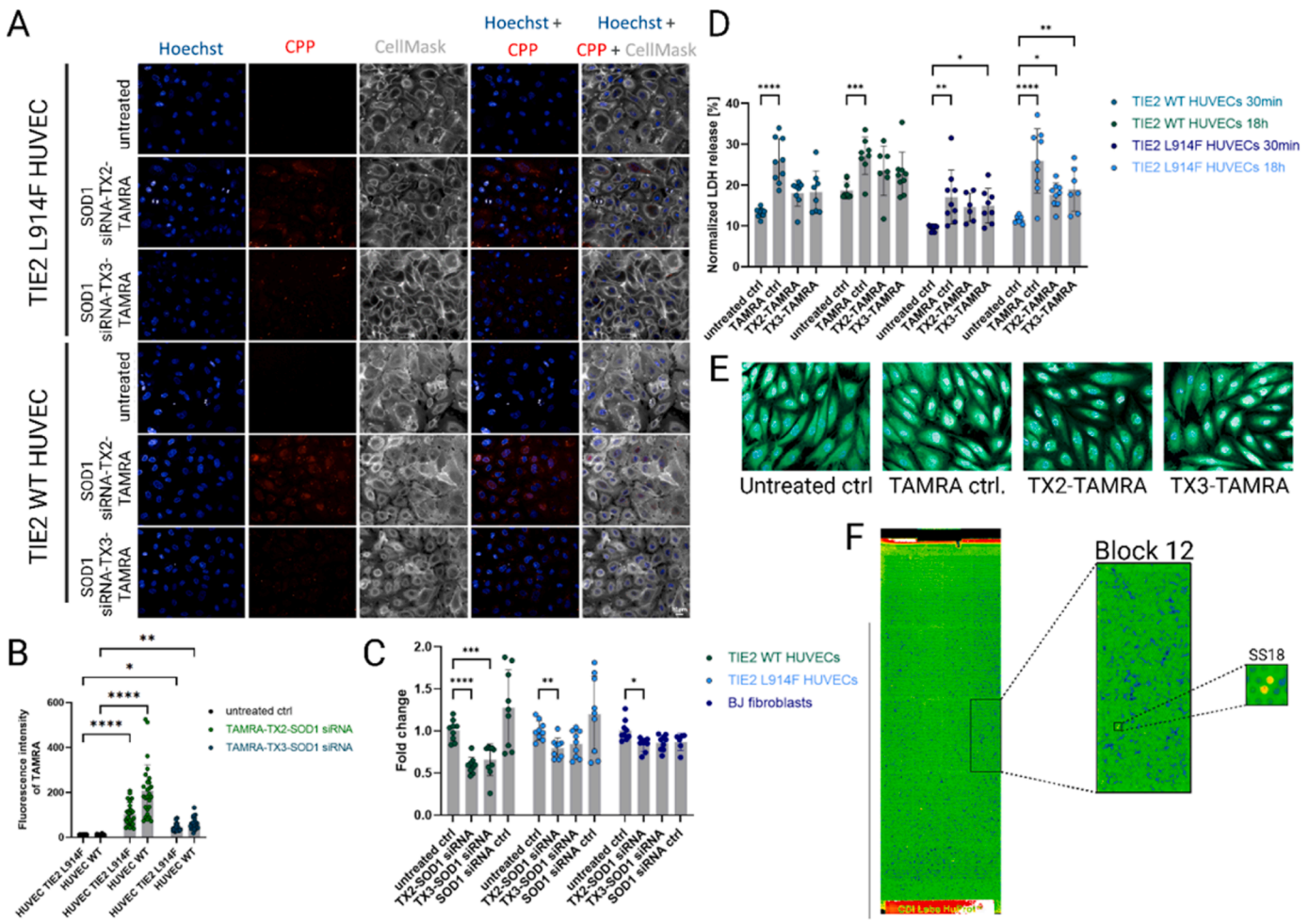


Fig. 5. Conjugation of siRNA to CPPs and functional delivery. (A) Immunofluorescence of SOD1 siRNA-CPP-TAMRA conjugates uptake assay (1 μM) displaying conjugates (red), plasma membrane (gray) and nucleus (blue). (B) Quantification of CPP-siRNA conjugate uptake assay with n = 3, *p < 0.05, **p < 0.01, ****p < 0.0001. (C) Functional delivery of siRNA-peptide-TAMRA conjugates (1 μM) was quantified using qPCR to determine SOD1 gene expression levels. n = 3, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. (D) LDH assay to test cytotoxicity of CPPs (10 μM) on TIE2 WT and TIE2 L914F HUVECs after 30 min and 18 h. n = 3, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.(E) Calcein staining (green; 1:10.000) to check for cell viability of TIE2 L914F HUVECs 30 min after treatment with CPPs. The nucleus is stained with Hoechst 33342 (blue). (F) Protein microarray to identify binding partners of TX2 and TX3 (10 μM). The array is depicted in green and proteins tested for interaction are highlighted. Binding intensities are color-coded as follows: black = low, blue-green-yellow-orange-red = increasing, white=saturated.

reduction of SOD1 expression only in TIE2 WT HUVECs, with a decrease in gene expression of 35 % detected. In addition, the TX3-conjugate exhibited a trend towards reduced SOD1 levels in TIE2 L914F HUVECs. Quantitative PCR thus confirming that TX2 and TX3 can transport siRNA through the plasma membrane leading to knockdown of SOD1 gene expression.

We were encouraged by these results and further extended our studies to determine whether the CPPs or their corresponding SOD1 siRNA conjugates have a cytotoxic effect on cells by testing CPPs and conjugates at a concentration of 10 μM or 1 μM, respectively, and at the same time points that we used for the uptake assays. To measure toxicity, we applied a colorimetric assay to measure lactate dehydrogenase (LDH) concentration in the medium. LDH is a cytosolic enzyme that is usually not released from cells and if found in medium is indicative of membrane disintegrity and cytotoxicity. LDH was measured in the medium of TIE2 WT and TIE2 L914F HUVECs after 30 min and 18 h to explore the short- and medium-term effect of adding TAMRA labeled TX2 and TX3 CPPs to the media (Fig. 5D). Results showed a slight tendency towards elevated LDH levels compared to untreated cells in the medium of TIE2 WT HUVECs after 30 min and 18 h with a similar trend seen for both peptides on TIE2 L914F HUVECs. In the study we included a TAMRA only control and found that the TAMRA label on its own has a moderate short- and medium-term toxic effect. This finding may explain the increases we see with TX2 and TX3 in this study as they were labeled with TAMRA. Calcein staining on both HUVEC lines showed that the majority of cells look healthy and the cytotoxic effect of the CPPs was minimal (Fig. 5E and Suppl. fig. S4A).

When we repeated the LDH assay with SOD1 siRNA-CPP-TAMRA conjugates, we focused on examining their medium- and long-term cytotoxicity (Suppl. figure S4B). We chose 18 h and 48 h for LDH assay as these were the same time points used in the study of functional delivery of SOD1 siRNA-CPP-TAMRA conjugates. Similar to the results with unconjugated CPPs, we saw moderate cytotoxicity on TIE2 L914F HUVECs for the SOD1 siRNA-TX3-TAMRA conjugate after 18 h of incubation. However, no long-term toxicity was detected for either conjugate. The half-life of LDH is 9 h and interestingly we detected rather low LDH release after 48 hours leading us to conclude that the addition of conjugates led to a mild shock for the cells which triggered minor medium-term cytotoxicity (likely as a result of the TAMRA label), but in long term conjugates were tolerated well and showed no cytotoxicity.

2.7. Identification of potential binding partners for TX2 and TX3 emphasize nuclear and cytoplasmic localization

Potential protein interaction partners for TX2 and TX3 were interrogated by protein microarray screening. Interaction profiling was performed at 10 μM with TX2 and TX3 peptide connected N-terminally to a TAMRA label on HuProt™ v4.0 - Human Proteome Microarrays, which contain approximately 20,000 proteins, covering approximately 16,000 human protein coding genes, representing 80 % of the human protein coding genome. Some genes are represented multiple times on the microarray with recombinant proteins from different clones present. A strong interaction between peptide and protein should be detected across more than one of these clones.

The analysis of TX2 was challenging due to high background staining of the microarray chip (Fig. 5F). Nevertheless, data processing revealed 23 hits with moderate binding activity (Table 3 and full ranked list in Suppl. Table S3). To normalize the fluorescence signals detected for all protein spots, a z score was calculated. This score determines how many standard deviations one particular data point is away from the mean of all data points. A z score over 2 was taken as threshold for a significant interaction and yielded 13 interactions for TX2 (Suppl. Table S3). Four of these hits, i.e. CALCOCO2, COL10A1, SS18 and HDAC6, showed similar interaction with TX2 with two different clones at different locations of the microarray (Table 3). This internal validation of the hits implies an increased likelihood that these are significant binding interactions. Proteins detected in the microarray screening have a variety of different structures and functions and are all located in the cytosol and nucleus (Suppl. Table S3). This confirms the results from the immunofluorescence imaging of uptake assays. We did not identify any interactions with cell surface receptors.

For TX3 only weak interactions were detectable at 10 μM, so we repeated the array at 40 μM. Interactions were still at a lower level, but 9 proteins, i.e. RBFOX2, NAA10, GAPDH, SCCPDH, ACOX1, METTL16, KCNAB1, PFKP, and NOSTRIN, were detected at either TX3 concentration confirming that these are valid but weak interactions.

3. Discussion and outlook

Blood vessel malfunction and malformation have a major impact on the development, progression and treatment of multiple human diseases such as cancer, psoriasis, stroke, diabetes complications and VAs and new treatments are urgently required. However, physically delivering therapeutic modalities to the desired tissues has proven challenging and novel targeting strategies are required.

In this study we present a new method to analyze *in vivo* phage display data and identify novel and effective CPPs, which can facilitate targeting and uptake of otherwise non-cell permeable therapeutic modalities into ECs. We established a model for VMs in immunocompromised mice and developed a more refined way to analyze phage display samples by introducing NGS and a downstream bioinformatic analyses platform. Two peptides were selected from a candidate list obtained from this platform and tested for their cellular uptake and cargo-transporting properties *in vitro*.

Previously, phage display has been used extensively to map the vasculature of different organs and to detect cell-homing and -penetrating peptides in models of cardiovascular diseases [7], but to our knowledge this study is the first to use phage display in a mouse model for VAs. We chose to perform our experiments in a mouse model of VMs instead of *in vitro*, as mutations that lead to VMs lack marker molecules and phage display on isolated molecules was not possible. Additionally, using a xenograft model introduces increased relevance to the phage display experiments, as it is known that some surface antigens maintain their physiological conformation only in intact cells or 3D models. Furthermore, *in vivo* phage display introduces the potential for the depletion of phages through binding to plasma components or ubiquitous surface proteins and thus provided us with a complex and challenging model to investigate the so-far intricate targeting of VMs.

Traditional analysis methods for phage display, such as phage plaque assay and Sanger sequencing of single hand-picked phage plaques were not followed in this study. Rather, we applied a NGS approach previously used in material science and protein science phage display studies [20,42]. The power of using NGS was exemplified by our finding that in one study a significant fraction of sequences (1.6×10^7) arose from insertless phages. In this case traditional analysis of 30–300 representative phage plaques would not reflect the true diversity of the phage pool and introduces bias in the decisions taken towards which peptides to test further. Here, one can appreciate the power of NGS compared to Sanger sequencing, we have interrogated millions of potential sequences as opposed to several hundred. Phage display studies that first introduced NGS were pioneering and innovative, but it became clear that complementing the approach with bioinformatic analysis is a powerful and crucial step towards identifying true and relevant hits. In this study, we used python and associated packages to build up a bioinformatic workflow that identified the peptide inserts in the NGS data, filtered out mismatches (e.g. sequences with premature stop codon), clustered the remaining peptides using a hierarchical algorithm and generated pLogos and consensus sequences. Our analytical focus was aimed at ensuring enrichment of the CPPs over the selection rounds and identifying promiscuity of the consensus sequences. Checking for promiscuity was performed on two levels: firstly, we examined how many peptides identified in the TIE2 L914F HUVEC xenograft were also identified in TIE2 WT xenografts and organs (e.g. kidney and liver which are well-known to have high levels of non-specific binding); secondly, we used in-house data sets and online databases to ensure that our newly identified CPPs were indeed novel.

Examining the physicochemical properties of the CPPs showed that our identified peptides were rather diverse in their features. Surprisingly, the two CPPs that we investigated experimentally had dissimilar properties, with TX2 more resembling the classic CPP penetratin. Interestingly, differences in physicochemical properties between TX2 and TX3 did not influence uptake mechanism as both were taken up by endocytosis, as was penetratin. Dissecting the exact uptake mechanism of CPPs is rather difficult, since these peptides usually do not adhere to a specific pathway and furthermore endocytosis inhibitors often block one of the pathways mainly, but have slight inhibitory effect on other endocytosis pathways at the same time. For example, Pitstop 2 is mainly an inhibitor for clathrin-dependent endocytosis, but is also known to inhibit clathrin-independent pathways in lower levels.

The aim of this study was to identify novel CPPs that we could use as cell targeting moieties, coupled to therapeutic modalities, for targeted drug delivery. Therefore, we wanted to show that our novel CPPs could firstly cross cell membranes and secondly deliver a functional, therapeutically relevant molecule. To assess membrane permeability we coupled TX2 and TX3 CPPs to a TAMRA fluorophore, choosing N-

Table 3
Alphabetically ordered interaction partners detected for sample TX2 (10 μM).

Gene name	Full name	z score	No. of clones
CALCOCO2	Calcium binding and coiled-coil domain 2	3.2 & 2.5	2
COL10A1	Collagen type X alpha 1 chain	2.7 & 2.4	2
HDAC6	Histone deacetylase 6	2.3 & 2.1	2
IBSP	Integrin binding sialoprotein	2.7	1
KHSRP	KH-type splicing regulatory protein	2.9	1
ODF2	Outer dense fiber of sperm tails 2	3.4	1
PHOSPHO1	Phosphoethanolamine/phosphocholine phosphatase	2.9	1
SOX9	SRY-box 9	2.3	1
SS18	SS18 subunit of BAF chromatin remodeling complex	2.5 & 2.2	2

terminus conjugation for both CPPs. To investigate functional delivery we conjugated a SOD1 specific siRNA, at the C-terminus in addition to the TAMRA label at the N-terminus. Finding SOD1 knockdown in HUVEC cells only, highlights the ability of cell-specific targeting of these peptides compared to penetratin. The unspecific uptake of penetratin, might be beneficial in certain research settings, but is a disadvantage when it comes to targeting. The siRNA delivery approach highlights furthermore the utility of our CPPs as toolbox molecules, where the CPP stays constant as the backbone and N- and C-termini can be connected to a variety of therapeutic modalities or tool molecules. In our study, CPPs were connected directly to siRNA, but the CPPs could theoretically be coupled to other modalities or even delivery agents such as lipid nanoparticles (LNPs), which in turn could be loaded with siRNA. Systemic delivery of LNPs is known to have high non-specific accumulation in the liver and therefore coupling to a targeting peptide might facilitate specific delivery to an alternate, more relevant tissue. If, due to positional effects, a CPP would inhibit the function of the cargo, a variety of linker compositions and lengths could be considered and if post-targeting removal of the CPP from the cargo is desirable (self-)cleaving systems based on enzymatic activity or pH sensitivity could be an option.

Uptake assays revealed that TX2 accumulated in the nucleus, implying that it can potentially facilitate delivery specifically to the nucleus of HUVECs. This important when considering the delivery of genome editing modalities such as CRISPR or ASOs, which would require nuclear targeting. In particular, for genetic diseases caused by point mutations such as VMs, the possibility to selectively deliver therapeutic modalities aimed at gene correction within an affected cell population would be reliant on nuclear targeting.

Analysis of TIE2 WT and TIE2 L914F-overexpressing HUVECs revealed 80 genes that were differentially expressed [1]. Gene set enrichment analysis showed that dysregulated genes are involved in vascular development, angiogenesis, cell migration and ECM processing, but how this influences VM formation is still not fully understood. By using TIE2 WT and TIE2 L914F-overexpressing HUVECs in a xenograft mouse model we attempted to differentiate these two cell lines from a new perspective, not previously covered by transcriptomic analysis. We showed that TX2 and TX3 CPPs were capable of HUVEC-specific penetration but did not differentiate between TIE2 WT HUVECs and TIE2 L914F HUVECs. Our hope was to be able to find a CPP that could discriminate between the two cell lines but we were encouraged to see that both TX2 and TX3 showed specificity for HUVECs compared to the other cell lines we tested, e.g. BJ fibroblasts, and also encouraging to see that both were able to deliver a functionally active siRNA into HUVECs. HUVECs are known to be difficult to transfect and here the CPPs could help to tackle this issue as a research tool. Given that we tested only two peptides from our bioinformatic analysis pipeline, there is the possibility that potential candidates, specific for TIE2 L914F HUVECs, can be found. The lack of TX2 and TX3 selectivity for mutated TIE2 L914F over TIE2 WT HUVECs in culture, despite the previous *in vivo* phage enrichment, may reflect the significant and obvious differences between *in vivo* and *in vitro* systems. The tolerability of novel CPPs is important to assess, particularly if one's goal is to introduce them into the clinic. We examined TX2 and TX3 CPPs for short-, medium-, and long-term cytotoxicity in cell culture found in general that they were well tolerated. Slightly elevated LDH levels were seen but this was likely due to the presence of a fluorophore on the CPPs. Building upon this it will be important to evaluate the peptides *in vivo* with a range of safety and toxicity assays, including blood analysis for inflammation markers.

One key question we wanted to address was the mechanism of cellular penetration for TX2 and TX3 CPPs and we explored the possibility that a common receptor expressed in HUVEC lines was responsible for binding and entry. We ran protein microarrays to identify protein interaction partners but only proteins with a cytoplasmic or nuclear localization were identified as having a specific binding interaction. Thus, further studies will be required to interrogate the cell surface interactions related to TX2 and TX3 binding and entry.

In summary, we identified novel CPPs via *in vivo* phage display combined with a bioinformatic-supported NGS approach and demonstrated the CPPs' ability to functionally deliver a siRNA cargo specifically into HUVECs.

4. Materials and methods

4.1. Cell culture

HUVECs were cultivated as described previously [43]. In brief, cells were grown in M200 medium supplied with 10 % fetal calf serum (FCS), 10 ml Low Serum Growth Supplement and penicillin/streptomycin (10.000 units/ml of penicillin and 10.000 μg/ml of streptomycin) at 37 °C and 5 % CO_2. Cell culture flasks were coated with Attachment Factor Solution (AFS) at 37 °C 30 min prior to cell seeding according to manufacturer's instructions. Cells were only used between passage 4–10 and seeding density was 1×10^4 cells/cm^2. ECs were cultivated to a confluency of 80–90 % before implantation in xenografts. TIE2 WT and TIE2 L914F HUVECs were chosen for this study, because this point mutation is well known from VA patients. Both cell lines were generated by retroviral transduction as previously described [2].

Neuro-2a cells were purchased from Jackson Laboratory and cultured in DMEM containing 10 % FCS at 37 °C and 5 % CO_2.

BJ fibroblasts were sourced from ATCC (CRL-2522) and cultured in DMEM containing 10 % FCS and 1 x NEAA at 37 °C and 5 % CO_2.

4.2. Animals

All experiments were carried out with permission by the Gothenburg Ethics Committee for Experimental Animals (license numbers 3074–2020 and 4123–2022). 6–8-week-old female SCID Beige mice (CB17.Cg-*Prkdc*^{scid}*Lyst*$^{bg\text{-}J}$/Crl) were purchased from Charles River (Germany) and acclimatized for at least 1 week upon arrival. Mice were group housed, were given access to water and fed ad libitum. They were exposed to a 12 h light/ 12 h dark cycle with fixed temperature (21 °C) and humidity (45–55 %).

4.3. Xenograft mouse model for VMs

To minimize the potential of unwanted immune reactions when implanting human cells into mice we used an immunocompromised SCID mouse strain. On the day of the experiment, HUVECs expressing TIE2 WT or TIE2 L914F were detached, counted and diluted in M200 medium to a cell number of 2×10^6 cells per injection. Cells were stored on ice until further use. Mice were anesthetized with isoflurane and the flanks were shaved for subcutaneous injection. Immediately before each injection, cells were mixed with growth factor reduced Matrigel Matrix (354230), 40 μl fibrinogen (diluted as instructed by manufacturer), 1 μg/ml VEGF-A (165), 1 μg/ml FGF-2 and 1 U thrombin. A final volume of 450 μl cell/Matrigel/growth factor suspension was loaded into a syringe and injected subcutaneously in the animals' flanks. All plastic equipment and the syringe needles were cooled on ice to avoid polymerization of matrigel in the syringe. Animals were checked daily and xenograft size was measured and animals were weighed every second day. *In vivo* phage display was performed 7 days after xenograft implementation.

4.4. Optical clearing

Xenografts were dissected 7 days after implantation and fixated in 4 % paraformaldehyde (PFA) overnight. Samples were washed 3 times in PBS for 5 min prior to the discoloration procedure. For this, xenograft tissues were cleared with the CUBIC method [44], followed by immunolabeling and dehydration in line with the iDISCO+ protocol [45]. In short, xenografts were incubated with 50 % CUBIC-1 solution (25 g quadrol, 25 g urea, 15 g Triton X-100, and 35 g H_2O) for 3 h followed by

100 % CUBIC-1 solution for 4 days. Decolorization was finalized by overnight incubation in 3 % H_2O_2 at 4 °C. Immunolabeling was started with two washing steps in PTx.2 solution (0.2 % Triton X-100 in PBS) for 1 h each at RT and xenografts were subsequently permeabilized for 2 days at 37 °C in permeabilization solution (400 ml PTx.2, 100 ml DMSO, 11.5 g glycine). After blocking for 2 days in blocking solution (42 ml PTx.2, 3 ml donkey serum, 5 ml DMSO) at 37 °C, xenografts were incubated with primary antibodies diluted in PTwH (0.2 % Tween-20, 0.01 mg/ml heparin in PBS) with 5 % DMSO and 3 % donkey serum for 4 days at 37 °C. Secondary antibodies were diluted in PTwH and 3 % donkey serum and incubated for 4 days at 37 °C. Dehydration of the samples was achieved by 1 h incubation steps in the following series of methanol/H_2O solutions at RT: 20 %, 40 %, 60 %, 80 %, 100 %, and 100 %. Finally, xenografts were shaken in a mix of 66 % DCM and 33 % methanol for 3 h at RT followed by incubating twice in 100 % DCM for 15 min each and DBE incubation overnight.

4.5. Light sheet fluorescence microscopy

Image stacks of immunostained xenografts were acquired with the LaVision Biotec Ultramicroscope II utilizing an Olympus MVX10 zoom body, containing an internal 2 x objective lens and an Andor Zyla 5.5 sCMOS camera. High resolution images were obtained using a 6.3 x magnification, setting the light sheet width at 10 %, applying chromatic aberration correction, and using a single laser from the left. 3D image stacks in TIFF format were loaded into ImageJ and merged into one TIFF per channel acquisition for analysis. Imaris and the build-in filament tracer module (Oxford Industries, version 10) were used for visualization of vessel networks.

4.6. In vivo phage display

Animals carrying xenografts for 7 days were anesthetized and each phage library was injected intravenously into 2 animals. Four different libraries from two companies were used in these experiments. TriCo-16 and TriCo-20 were purchased from Creative Biolabs with a capacity of 2.6×10^{10} and 1.1×10^{10}, respectively. PhD-C7C and PhD-12 were supplied from New England Biolabs and had a titer of 2×10^{13} or 1×10^{13}, respectively. 30 min after phage library injection, animals were anesthetized and perfused with 10 ml HBSS to minimize non-specific binding of phages. Thereafter, xenografts and control organs (kidney and liver) were dissected and collected in PBS. Single cell suspensions were prepared by mashing tissues and subsequent incubation with dispase II in RPMI medium for 40 min at 37 °C. In case tissue clumps were left, these were homogenized by pipetting. The cell suspension was then transferred through a 100 μm mesh to exclude cell clumps and digestion was stopped by addition of RPMI medium containing 10 % FCS. Cells were washed once with RPMI medium with 10 % FCS and passed through a 30 μm cell strainer.

4.7. Preparation of subcellular fractions

Samples were split in half preparing nuclear fractions from one part and cytosolic and membrane fractions were derived from the other half [46]. To prepare the nuclear fraction, cells were resuspended in 4 ml lysis buffer I (10 mM Tris HCl, pH 7.9, 10 mM NaCl, 3 mM $MgCl_2$ and 0.1 % NP-40) and incubated on ice for 1 h. The lysed cells were then layered onto 4 ml of lysis buffer I containing 30 % (w/v) sucrose and centrifuged at 2500 g for 15 min. The pelleted nuclei were resuspended in 50 μl PBS and lysed through addition of 200 μl lysis buffer II (10 mM Tris HCl, pH 7.9, 10 mM EDTA, 0.1 % SDS and 0.2 % NP-40). In parallel to the nuclear fraction, the cytosolic fraction was prepared by resuspending cells in 400 μl homogenization buffer (PBS containing 0.1 % NP-40, 1 % BSA, 0.5 mM PMSF, 1 mg/ml aprotinin, 1 mg/ml leupeptin and 1 mg/ml pepstatin) and homogenizing the samples with a Dounce homogenizer. Lysed cells were centrifuged to separate cytosol from membrane fragments. To also capture phages that interact with membranes, the spun down fraction was resuspended in 400 μl 0.2 M glycine HCl (pH 2.2) for 10 min with shaking at 200 rpm. Finally, pH was neutralized by addition of 50 μl 1 M Tris (pH 9.1) and the volume of all subcellular fractions was adjusted to 1 ml with PBS.

4.8. Phage purification, recovery and preparation for next selection round

To purify phages from the subcellular fraction preparations, phages were precipitated with 200 μl PEG/NaCl (5 x) on ice for 1 h and pelleted by centrifugation at 12000 g for 15 min at 4 °C. Phage pellets were dissolved in 25 μl PBS and added to a log phase culture of ER2738 *E.coli* in LB medium containing tetracycline. The bacteria inoculated with phages were incubated for 30 min shaking at 225 rpm, before LB medium containing tetracycline and glucose was added for incubation overnight. On the following day, bacteria were removed by centrifugation at 4600 rpm for 15 min at 4 °C and the bacteria pellet was resuspended in LB medium containing tetracycline and 15 % glycerol and stored at 80 °C. Phages were precipitated from LB medium by incubation with PEG/NaCl on ice for 1 h, pelleted by centrifugation, and resuspended in sterile PBS. Phage titer was determined by measuring absorption on a Nanodrop One spectrophotometer (Thermo Fisher Scientific).

4.9. Sample preparation and NGS

Bacteria, from glycerol stocks, were grown in tetracycline containing LB medium at 37 °C with shaking at 225 rpm. Phage plasmid dsDNA was purified by using the QIAprep Spin Miniprep Kit. A PCR primer pair was used (forward: 5'-TTGTCGTCTTTCCAGACGTT-3'; reverse: 5'-GCAAGCTGATAAACCGATACA-3') [20], that covers peptide inserts in pIII. Depending on the library used, amplicons varied between 234 and 255 bp. Phage insert primers (0.5 μM), Phusion Flash High Fidelity PCR Master Mix, nuclease free water and purified phage plasmid were amplified in a final volume of 10 μl. The following PCR cycling conditions were used: 98 °C for 2 min, then for 25 cycles 98 °C for 5 sec, 64 °C for 6 sec, 72 °C for 5 sec and as a last step 1 min at 72 °C. Amplicons were purified by using the Agencourt AmPure XP purification system. To verify high sample purity and determine concentration for sequencing input, 2 μl of the samples were analyzed on a 5200 Fragment Analyzer system (Agilent Technologies) using the dsDNA 915 Reagent kit (35–5000 bp). DNA concentration was adjusted to an input concentration of 0.067 ng/μl using QuBit dsDNA assay. Subsequently, standard paired end 2×150 bp read sequencing on an Illumina NextSeq500 machine was performed. Samples were mapped on the pIII coat protein of the M13 bacteriophage using the Burrows Wheeler aligner (BWA) [47] and quality controlled using the phred score.

4.10. Bioinformatic analysis

Data curation and analysis was carried out using python3 and jupyter notebooks. Figures were constructed using matplotlib. To begin with, datasets were filtered to identify relevant sequences. The fastq reads were scanned for the sequences expected to flank the region of interest (CCTTTCTATTCTCAC-GCCGAAACTGTTGAA) with 2 mismatches allowed per flank and the intervening DNA sequences extracted, counted and translated to protein. Non-coding sequences with no corresponding peptide and sequences containing stop codons were removed. Next, the sequence flanks corresponding to each library were determined and sequences filtered to match the flank criteria and sequence length (Table 4).

Data sets were separated into fractions (nucleus, cytosol, membrane) in order to determine if peptides were common or unique to each fraction. The matplotlib_venn package was used to generate venn diagrams to identify overlapping sequences.

Only the peptides found in the third round of NGS were considered

Table 4
Phage libraries and their specific flank sequences used for the panning experiments.

Phage library	No. of displayed amino acids	No. of inserted amino acids	Flank sequences
PhD-C7C	7	15	$SACX_7CGGGS$
PhD−12	12	17	$SX_{12}GGGS$
TriCo−16	16	18	$SX_{16}S$
TriCo−20	20	22	$SX_{20}S$

for the final analysis as these were considered the enriched peptide fraction. In order to identify which peptides were selective within each cell line, each peptide was cross referenced with its respective control group. Peptides were only labeled as selective if they had an occurrence frequency in the control groups of 0.01 % or lower.

In preparation for the clustering a distance matrix describing the similarity between each of the peptides was required. To do this, we performed a pairwise similarity calculation using the pairwise2 package of biopython [48]. A global alignment between each sequence pair was performed and matching residues were given a score of 1.5, while mismatches were given a score of 0. Gap openings and gap extensions were penalized with scores of −0.5 and −0.1 respectively. From these calculations a similarity matrix was generated, which was subsequently converted to a distance matrix by using the inverse of the normalized similarity scores (see equation).

$$Dist(X) = 1 - \frac{Sim(X)}{MaxSim},$$

where dist(X) is the distance being calculated for the pairwise comparison X, Sim(X) is the previously calculated similarity score for the pairwise comparison X, and MaxSim is the maximum value in the similarity matrix. This results in a normalized distance matrix where two sequences with high similarity have a score closer to 0, while sequences with little similarity have a score closer to 1. The distance matrix was then used to perform the clustering analysis. Hierarchical clustering was performed using the SciPy [49] hierarchy.linkage module. The distance between peptide clusters was calculated using an averaging method and measured as the euclidean distance. To measure the quality of the clustering the CP was calculated. A CP value close to 1.0 indicates a good preservation of the original values from the distance matrix. The SciPy hierarchy.dendrogram module was then used to visualize the clustering results. Dendrograms were visually inspected to determine an appropriate distance cut off to form the final clusters. The final distance cut off was determined to have a phylogenetic distance of $\frac{\max(dist)}{6}$ or the nearest value that resulted in a maximum of 20 clusters.

Consensus sequences for each cluster were calculated and pLogo graphs were generated using the Logomaker.logo module [50]. The consensus sequence was determined by calculating the occurrence frequency of any residue at each position within the peptide. The residue with the highest occurrence rate was set for each position within the sequence. The total number of reads for each peptide was considered when calculating the consensus sequence. For selection of the final consensus sequences for testing in cell culture, we checked whether the same consensus sequence was also found in clustering results from other cell lines within our in-house datasets. Thus, sequences with a higher selectivity and in consequence lower promiscuity could be identified and selected for testing.

Enrichment of selected peptides was investigated by plotting the number of reads of each peptide present in each of the three rounds. Before ordering peptides labeled with a fluorophore, we checked if the consensus sequence occurred in the samples. If the consensus sequence could not be detected, we took a closer look at the sequences building up this cluster and picked the one with the highest reads for this cluster, which usually had high similarity with the consensus sequence.

4.11. Calculation of physicochemical properties and 3D peptide structures

Physicochemical properties were calculated using two different python packages. The pI, charge (pH 7), and GRAVY index (hydrophobicity using KyteDoolittle parameters) were calculated using the bioPython package [48]. The aliphatic index (using Ikai parameters) and Boman index (potential peptide interaction score) were calculated using the peptides.py package [51]. The resulting data is shown in Table 2.

AlphaFold, version 2.2.0 [30], was used to predict the structure of penetratin and the top 10 peptides. For penetratin we used the RCSB PDB 1kz0. However, several structures of penetratin are available in RCSB PDB (1kz0 and 1omq), on which calculations gave 37.5 % and 43.8 % α-helical content. We chose to use the AlphaFold model instead of already published PDB files for our comparisons to guarantee comparability with the peptides derived from phage display in this study. Secondary structure calculations were performed with DSSP, version 2.0.4, on all the predicted AlphaFold models. Eight types of secondary structure elements can be predicted by DSSP [52,53]. If present, only H (α-helical) or E (β-sheet) were considered. All the others were grouped into “none”.

4.12. Database searches

BLAST+ sequence searches [31], with task settings tuned for short sequences (blastp short), were performed with the peptide hit sequences. The searches were conducted against the following peptide and protein databases: UniprotKB v.2022_012 [32], Propedia v2.33 [33], CPPsite2.04 [34], TUPScan and PhD7Faster [35,36]. These databases were downloaded and filtered to obtain sequences with greater than 3 amino acids to avoid non-specific matches as well as the redundant sequences.

4.13. CPP uptake assay

HUVECs, Neuro-2a and BJ fibroblasts were seeded at a density of 4×10^4 cells/cm^2, 5×10^4 cells/cm^2 or 3×10^4 cells/cm^2, respectively, in black 96 well plates. Prior to the experiment, peptides were centrifuged at 12000 g for 5 min at 4 °C. When cells reached 80 % confluency, they were incubated with 10 μM peptide for 30 min at 37 °C. Subsequently, cells were washed with 0.2 M glycine-HCl buffer (pH 2.8) for 3 min, fixed with 4 % PFA for 20 min, and incubated with primary antibody anti-human VEGFR2 (1:200) at 4 °C overnight. The following day, cells were washed 3 times using PBS-T for 5 min each. A secondary antibody, Alexa Fluor 488 donkey anti-mouse IgG (H+L) (1:500), was incubated for 1 h at RT, followed by washing 3 times for 5 min each with PBS-T. Hoechst 33342 and CellMask Deep Red were added to PBS$^{-/-}$ (1:1000 both) to stain the nucleus and the cell membrane after the third wash for 20 min at RT. Imaging was carried out on the CellVoyager CV7000 microscope (Yokogawa) with 60 x magnification and analyzed with Columbus image analysis system (PerkinElmer).

4.14. CPP uptake mechanism assay

To test whether CPPs were taken up by endocytosis, we co-incubated cells with Dextran 10 kDa and CPP for 2 h at 37 °C. To have a closer look at the different endocytosis pathways, we pre-treated the cells with different endocytosis inhibitors. To inhibit dynamin-dependent endocytosis, cells were pre-incubated for 30 min with 30 μM Dynasore. Pitstop 2, a clathrin-dependent endocytosis inhibitor, was used at 25 μM for 30 min and macropinocytosis was blocked by pre-incubating cells for 1 h with 25 μM EIPA. Then, we followed the uptake assay protocol described in the previous paragraph with corresponding endocytosis inhibitor present during the uptake assay.

4.15. Cytotoxicity assay and viability staining

The toxicity of CPPs and CPP conjugates was assessed by using the CytoTox 96® Non-Radioactive Cytotoxicity Assay from Promega. Cells were seeded at a density of 4×10^4 cells/cm^2 for HUVECs, 5×10^4 cells/cm^2 for Neuro-2a and 3×10^4 cells/cm^2 for BJ fibroblasts in 96 well plates. To measure the CPPs´long- and short-term toxicity, cells were incubated with CPPs at 10 μM for 30 min and 18 h at 37 °C. Similarly, the CPP conjugates were incubated using 1 μM for 18 h and 48 h at 37 °C. Maximum LDH control wells were treated with 10 μl 10 x Lysis Solution/100 μl medium 45 minutes before assay finish. After respective incubation times, 50 μl medium of each well was transferred to a fresh 96 well plate and 50 μl of CytoTox 96® Reagent was added. The reaction was incubated for 30 min at RT protected from light. Then, 50 μl Stop Solution was added and the absorbance at 490 nm was determined on a Victor X4 plate reader (PerkinElmer).

The cell plates from the LDH assay were reused for viability staining with Calcein-AM and live imaging. Once all medium was removed, the cells were incubated with Calcein-AM (1:10.000 in PBS) and Hoechst 33342 (1:1000 in PBS) for 20 min at 37 °C. Staining solution was removed, and cells washed 3 times for 5 min each with PBS. Then, the cells were imaged on a CellVoyager CV7000 microscope (Yokogawa) with 60 x magnification and pictures were generated with Columbus image analysis system (PerkinElmer).

4.16. Click chemistry conjugation of CPPs to SOD1siRNA

Peptides TX2 and TX3 used for conjugation were equipped with a N-terminal TAMRA label and an azide group on the C-terminus (short TAMRA-CPP-azide) and were purchased from Hefei KS-V Peptide Biological Technology. SOD1-sense RNA strand with a bicyclo[6.1.0]non-4-yne (BCN)-amine group on the RNA 3´end (short SOD1-sense-BCN) was provided by the oligochemistry department in AstraZeneca. Firstly, the TAMRA-CPP-azide was conjugated to the sense strand of the SOD1 siRNA sequence via copper-free click chemistry and then the antisense strand of SOD1 siRNA was annealed.

4.17. Preparation of TAMRA-TX2-triazole-SOD1 sense conjugate

To the solution of SOD1-sense-BCN (2.86 mmol/l) in 1:5 PBS/water mixture, TAMRA-TX2-azide (7.2 mmol/l) dissolved in DMF was added. The reaction was shaken at RT for 24 h but was not complete after this time. The temperature was increased to 40 °C for a few hours to push the reaction forward and then brought back to RT and shaken overnight.

CPP-oligonucleotide conjugate was precipitated with sodium acetate (0.3 M) and cold absolute ethanol, and allowed to precipitate furthermore overnight at −20 °C.

The next day, the vial was centrifuged, decanted and dried under vacuum. The CPP-oligonucleotide conjugate was dissolved in Milli-Q water and lyophilised overnight. The dried product was used for annealing the antisense DNA strand without further purification.

4.18. Preparation of TAMRA-TX3-triazole-SOD1 sense conjugate

To the solution of SOD1-sense-BCN (2.23 mmol/l) in 1:6 PBS/water mixture, TAMRA-TX3-azide (6.6 mmol/l) dissolved in DMF was added. The reaction was shaken at RT overnight.

The CPP-oligonucleotide conjugate was precipitated and prepared in the same way as the first conjugate for annealing.

4.19. Annealing of TAMRA-CPP-triazole-SOD1 sense conjugates to antisense DNA strand

SOD1-sense-TX2-TAMRA (0.82 mM) or SOD1-sense-TX3-TAMRA (0.64 mM) in PBS was added to solution of SOD1-antisense strand (3.09 mM) in PBS and brought to a final volume of 592 μL or 568 μL with PBS, respectively. The solution was shaken at 100 rpm at 95 °C for 5 min and then cooled down to RT over a period of an hour. Gel electrophoresis was performed to analyze the annealing.

4.20. qPCR

TIE2 WT and TIE2 L914F HUVECs as well as BJ fibroblasts were seeded at a density of 1.5×10^4 cells/cm^2 in 6 well plates and cultivated for 24 h. For HUVECs, the plates were as usual coated with AFS for 30 min prior to seeding. One day later, medium was replaced with medium containing 1 μM SOD1 siRNA-CPP-TAMRA conjugate or unconjugated SOD1 siRNA and incubated for 48 h. Then, medium was discarded, cells detached with TripLE for 5 min and the cell pellet was washed once with PBS before cell lysis in RLT buffer. RNA was extracted with RNeasy Plus Mini Kit following the manufacturer's instructions. RNA concentration was quantified with Nanodrop One spectrophotometer (Thermo Fisher Scientific) and cDNA was prepared with High Capacity cDNA Reverse Transcription kit. Quantitative PCR was performed using TaqMan Fast Advanced Master Mix and TaqMan expression assays for human SOD1, HPRT1, GAPDH and TBP on QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific). Expression of SOD1 was normalized to the expression of housekeeping genes HPRT1, GAPDH and TBP.

4.21. Protein array

The HuProt™ v4 Human Protein Microarray was used to identify binding partners of TAMRA-labeled TX2 and TX3 peptides. This work was performed by Cambridge Protein Arrays according to the protocol previously described here [54]. TAMRA fluorescence signal was measured using 532 nm as excitation wavelength.

4.22. Statistical analysis

Since *in vivo* phage display is an explorative experiment, biopanning was performed in 2 mice per library. Each library was panned 3 consecutive times, enriching specific CPPs in each round.

LDH assay and uptake experiments were performed in triplicate. LDH assay and qPCR data were analyzed with two way analysis of variance (ANOVA). Data are presented as mean ± SD and p values < 0.05 were applied as criteria for statistical significance.

Author contributions

Conceptualization, L.J., J.W. and F.K.; methodology, F.K., J.W.; formal analysis, F.K.; O.L.; investigation, F.K., O.L., L.K., P.M., M.A., G. G., L.E., L.D.M.; data curation, M.F, O.L, F.K.; writing original draft preparation, F.K., L.J., J.W.; writing, review and editing, F.K., L.D.M., L. J., J.W.; visualization, F.K.; L.K., L.D.M.; software, F.K, O.L., L.K., L.D. M.; supervision, L.J., J.W.; project administration, L.J., J.W.; funding acquisition, L.J., J.W.; All authors have read and agreed to the published version of the manuscript.

Funding

F.K. and M.A. received funding from the European Union's Horizon 2020 research and innovation programme under the Skłodowska Curie grant agreement No 814316 (to L.J, J.W. and L.E). L.J. received support from the Swedish Research Council (2021 01210), the Swedish Heart and Lung Foundation (20220354), the Swedish Brain Foundation (FO2022–0055), the Swedish Cancer Society (22 2440 Pj) and Karolinska Institutet. L.E. received funding from the Sigrid Jusélius Foundation and M.A. from the Finnish Cultural Foundation.

CRediT authorship contribution statement

Mohammadhassan Ansarizadeh: Investigation. **Priyanka Mangla:** Investigation. **Luc Krimpenfort:** Visualization, Software, Investigation. **Mike Firth:** Data curation. **Oliver Laufkötter:** Software, Investigation, Formal analysis, Data curation. **Franziska Kohl:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **John Wiseman:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Lars Jakobsson:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. **Leonardo De Maria:** Writing – review & editing, Visualization, Software, Investigation. **Lauri Eklund:** Investigation. **Gökçe Geylan:** Investigation.

Declaration of Competing Interest

F.K., O.L., M.P., M.F., P.M., G.G., L.D.M., J.W. are current or former employees of AstraZeneca Pharmaceuticals.

Acknowledgements

We thank Marta Potapo for her support with amplicon sequencing and Antonio Queiro Palou for helpful discussions about light sheet microscopy. We are grateful for fruitful discussions with Maria Volkova and Anders Dahlén about conjugation of CPPs to siRNA and with Michael Munson about endocytosis pathways. We furthermore thank Soniya Savant for sharing her confocal microscopy expertise and Antje Rottner for sharing her qPCR expertise.

Institutional review board statement

The animal study protocol was approved by the Gothenburg Ethics Committee for Experimental Animals (license number 4123 2022, 26th of January 2022).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.biopha.2024.117740.

Data availability

All data has been incuded in the manuscript.

References

[1] M. Uebelhoer, M. Nätynki, J. Kangas, A. Mendola, H.L. Nguyen, J. Soblet, C. Godfraind, L.M. Boon, L. Eklund, N. Limaye, M. Vikkula, Venous malformation-causative TIE2 mutations mediate an AKT-dependent decrease in PDGFB, Hum. Mol. Genet. 22 (2013) 3438–3448, https://doi.org/10.1093/hmg/ddt198.

[2] M. Nätynki, J. Kangas, I. Miinalainen, R. Sormunen, R. Pietilä, J. Soblet, L.M. Boon, M. Vikkula, N. Limaye, L. Eklund, Common and specific effects of TIE2 mutations causing venous malformations, Hum. Mol. Genet. 24 (2015) 6374–6389, https://doi.org/10.1093/hmg/ddv349.

[3] A.M. Hammill, M. Wentzel, A. Gupta, S. Nelson, A. Lucky, R. Elluru, R. Dasgupta, R.G. Azizkhan, D.M. Adams, Sirolimus for the treatment of complicated vascular anomalies in children, Pediatr. Blood Cancer 57 (2011) 1018–1024, https://doi.org/10.1002/pbc.23124.

[4] E. Boscolo, N. Limaye, L. Huang, K.T. Kang, J. Soblet, M. Uebelhoer, A. Mendola, M. Natynki, E. Seront, S. Dupont, J. Hammer, C. Legrand, C. Brugnara, L. Eklund, M. Vikkula, J. Bischoff, L.M. Boon, Rapamycin improves TIE2-mutated venous malformation in murine model and human subjects, J. Clin. Invest. 125 (2015) 3491–3504, https://doi.org/10.1172/JCI76004.

[5] D.M. Adams, C.C. Trenor 3rd, A.M. Hammill, A.A. Vinks, M.N. Patel, G. Chaudry, M.S. Wentzel, P.S. Mobberley-Schuman, L.M. Campbell, C. Brookbank, A. Gupta, C. Chute, J. Eile, J. McKenna, A.C. Merrow, L. Fei, L. Hornung, M. Seid, A. R. Dasgupta, B.H. Dickie, R.G. Elluru, A.W. Lucky, B. Weiss, R.G. Azizkhan, Efficacy and safety of sirolimus in the treatment of complicated vascular anomalies, Pediatrics 137 (2016) e20153257, https://doi.org/10.1542/peds.2015-3257.

[6] X. Li, Y. Cai, J. Goines, P. Pastura, L. Brichta, A. Lane, T.D. Le Cras, E. Boscolo, Ponatinib combined with rapamycin causes regression of murine venous malformation, Arterioscler. Thromb. Vasc. Biol. 39 (2019) 496–512, https://doi.org/10.1161/ATVBAHA.118.312315.

[7] J. Bábíčková, Ľ. Tóthová, P. Boor, P. Celec, *In vivo* phage display-a discovery tool in molecular biomedicine, Biotechnol. Adv. 31 (2013) 1247–1259, https://doi.org/10.1016/j.biotechadv.2013.04.004.

[8] C. Rinaldi, M.J.A. Wood, Antisense oligonucleotides: the next frontier for treatment of neurological disorders, Nat. Rev. Neurol. 14 (2018) 9–21, https://doi.org/10.1038/nrneurol.2017.148.

[9] C.F. Bennett, Therapeutic antisense oligonucleotides are coming of age, Ann. Rev. Med. 70 (2019) 307–321, https://doi.org/10.1146/annurev-med-041217-010829.

[10] M. Bjursell, M.J. Porritt, E. Ericson, A. Taheri-Ghahfarokhi, M. Clausen, L. Magnusson, T. Admyre, R. Nitsch, L. Mayr, L. Aasehaug, F. Seeliger, M. Maresca, M. Bohlooly-Y, J. Wiseman, Therapeutic genome editing with CRISPR/Cas9 in a humanized mouse model ameliorates α1-antitrypsin deficiency phenotype, EBioMedicine 29 (2018) 104–111, https://doi.org/10.1016/j.ebiom.2018.02.015.

[11] H. Dana, G.M. Chalbatani, H. Mahmoodzadeh, R. Karimloo, O. Rezaiean, A. Moradzadeh, N. Mehmandoost, F. Moazzen, A. Mazraeh, V. Marmari, M. Ebrahimi, M.M. Rashno, S.J. Abadi, E. Gharagouzlo, Molecular mechanisms and biological functions of siRNA, Int. J. Biomed. Sci. 13 (2017) 48–57.

[12] J. Winkler, Oligonucleotide conjugates for therapeutic applications, Ther. Deliv. 4 (2013) 791–809, https://doi.org/10.4155/tde.13.47.

[13] P.P. Seth, M. Tanowitz, C.F. Bennett, Selective tissue targeting of synthetic nucleic acid drugs, J. Clin. Invest. 129 (2019) 915–925, https://doi.org/10.1172/JCI125228.

[14] C. Ämmälä, W.J. Drury 3rd, L. Knerr, I. Ahlstedt, P. Stillemark-Billton, C. Wennberg-Huldt, E.M. Andersson, E. Valeur, R. Jansson-Löfmark, D. Janzén, L. Sundström, J. Meuller, J. Claesson, P. Andersson, C. Johansson, R.G. Lee, T. P. Prakash, P.P. Seth, B.P. Monia, S. Andersson, Targeted delivery of antisense oligonucleotides to pancreatic β-cells, Sci. Adv. 4 (2018) eaat3386, https://doi.org/10.1126/sciadv.aat3386.3.

[15] S.S.K. Dasa, R. Suzuki, M. Gutknecht, L.T. Brinton, Y. Tian, E. Michaelsson, L. Lindfors, A.L. Klibanov, B.A. French, K.A. Kelly, Development of target-specific liposomes for delivering small molecule drugs after reperfused myocardial infarction, J. Control Release 220 (2015) 556–567, https://doi.org/10.1016/j.jconrel.2015.06.017.

[16] M. Hust, A. Frenzel, T. Meyer, T. Schirrmann, S. Dubel, Construction of human naive antibody gene libraries, Methods Mol. Biol. 907 (2012) 85–107, https://doi.org/10.1007/978-1-61779-974-7_5.

[17] B. Parhami-Seren, M. Viswanathan, M.N. Margolies, Selection of high affinity p-azophenyarsonate Fabs from heavy-chain CDR2 insertion libraries, J. Immunol. Methods 259 (2002) 43–53, https://doi.org/10.1016/s0022-1759(01)00488-4.

[18] R. Crameri, M. Suter, Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production, Gene 137 (1993) 69–75, https://doi.org/10.1016/0378-1119(95)00190-h.

[19] K. Hertveldt, J. Robben, G. Volckaert, Whole genome phage display selects for proline-rich Boi polypeptides against Bem1p, Biotechnol. Lett. 28 (2006) 1233–1239, https://doi.org/10.1007/s10529-006-9082-y.

[20] C. Juds, J. Schmidt, M.G. Weller, T. Lange, U. Beck, T. Conrad, H.G. Börner, Combining phage display and next-generation sequencing for materials sciences: a case study on probing polypropylene surfaces, J. Am. Chem. Soc. 142 (2020) 10624–10628, https://doi.org/10.1021/jacs.0c03482.

[21] C.I. Øie, D.L. Wolfson, T. Yasunori, G. Dumitriu, K.K. Sørensen, P.A. McCourt, B. S. Ahluwalia, B. Smedsrød, Liver sinusoidal endothelial cells contribute to the uptake and degradation of entero bacterial viruses, Sci. Rep. 10 (2020) 898–906, https://doi.org/10.1038/s41598-020-57652-0.

[22] M. Green, P.M. Loewenstein, Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein, Cell 55 (1988) 1179–1188, https://doi.org/10.1016/0092-8674(88)90262-0.

[23] A.D. Frankel, C.O. Pabo, Cellular uptake of the tat protein from human immunodeficiency virus, Cell 55 (1988) 1189–1193, https://doi.org/10.1016/0092-8674(88)90263-2.

[24] A. Joliot, C. Pernelle, H. Deagostini-Bazin, A. Prochiantz, Antennapedia homeobox peptide regulates neural morphogenesis, Proc. Natl. Acad. Sci. USA 88 (1991) 1864–1868, https://doi.org/10.1073/pnas.88.5.1864.

[25] A. Sillero, J.M. Ribeiro, Isoelectric points of proteins: theoretical determination, Anal. Biochem. 179 (1989) 319–325, https://doi.org/10.1016/0003-2697(89)90136-x.

[26] D.S. Moore, Amino acid and peptide net charges: a simple calculational procedure, Biochem. Educ. 13 (1985) 10–11, https://doi.org/10.1016/0307-4412(85)90114-1.

[27] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, J. Mol. Biol. 157 (1) (1982 May 5) 105–132, https://doi.org/10.1016/0022-2836(82)90515-0.

[28] A. Ikai, Thermostability and aliphatic index of globular proteins, J. Biochem. 88 (1980) 1895–1898.

[29] H.G. Boman, Antibacterial peptides: basic facts and emerging concepts, *J. Intern. Med.* 254 (2003) 197–215, https://doi.org/10.1046/j.1365-2796.2003.01228.x.

[30] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction

with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[31] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+ : architecture and applications, BMC Bioinforma. 10 (2009) 421–429, https://doi.org/10.1186/1471-2105-10-421.
[32] UniProt Consortium. UniProt: the universal protein knowledgebase in 2023, Nucleic Acids Res 51 (2023) D523–D531, https://doi.org/10.1093/nar/gkac1052.
[33] P. Martins, D. Mariano, F.C. Carvalho, L.L. Bastos, L. Moraes, V. Paixão, R. Cardoso de Melo-Minardi, Propedia v2.3: a novel representation approach for the peptide-protein interaction database using graph-based structural signatures, Front Bioinform 3 (2023) 1103103, https://doi.org/10.3389/fbinf.2023.1103103.
[34] P. Agrawal, S. Bhalla, S.S. Usmani, S. Singh, K. Chaudhary, G.P. Raghava, A. Gautam, CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides, Nucleic Acids Res 44 (2016) D1098–1103, https://doi.org/10.1093/nar/gkv1266.
[35] B. He, H. Chen, N. Li, J. Huang, SAROTUP: a suite of tools for finding potential target-unrelated peptides from phage display data, Int. J. Biol. Sci. 15 (2019) 1452–1459, https://doi.org/10.7150/ijbs.31957.
[36] J. Huang, B. Ru, S. Li, H. Lin, F.B. Guo, SAROTUP: scanner and reporter of target-unrelated peptides, J. Biomed. Biotechnol. 2010 (2010) 101932, https://doi.org/10.1155/2010/101932.
[37] D. Derossi, A.H. Joliot, G. Chassaing, A. Prochiantz, The third helix of the Antennapedia homeodomain translocates through biological membranes, J. Biol. Chem. 269 (1994) 10444–10450.
[38] T. Kirchhausen, E. Macia, H.E. Pelish, Use of dynasore, the small molecule inhibitor of dynamin, in the regulation of endocytosis, Methods Enzym. 438 (2008) 77–93, https://doi.org/10.1016/S0076-6879(07)38006-3.
[39] I. Ruseska, A. Zimmer, Internalization mechanisms of cell-penetrating peptides, Beilstein J. Nanotechnol. 11 (2020) 101–123, https://doi.org/10.3762/bjnano.11.10.
[40] E.C.A. Eleutherio, R.S. Silva Magalhães, A. de Araújo Brasil, J.R. Monteiro Neto, L. de Holanda Paranhos, SOD1, more than just an antioxidant, Arch. Biochem. Biophys. 697 (2021) 108701, https://doi.org/10.1016/j.abb.2020.108701.
[41] H. Peng, S. Zhang, Z. Zhang, X. Wang, X. Tian, L. Zhang, J. Du, Y. Huang, H. Jin, Nitric oxide inhibits endothelial cell apoptosis by inhibiting cysteine-dependent SOD1 monomerization, FEBS Open Bio 12 (2022) 538–548, https://doi.org/10.1002/2211-5463.13362.
[42] R. Braun, N. Schönberger, S. Vinke, F. Lederer, J. Kalinowski, K. Pollmann, Application of next generation sequencing (NGS) in phage displayed peptide selection to support the identification of arsenic-binding motifs, Viruses 12 (2020) 1360–1392, https://doi.org/10.3390/v12121360.
[43] H. Elamaa, M. Kihlstrom, E. Kapiainen, M. Kaakinen, I. Miinalainen, S. Ragauskas, M. Cerrada-Gimenez, S. Mering, M. Natynki, L. Eklund, Angiopoietin-4-dependent venous maturation and fluid drainage in the peripheral retina, Elife 7 (2018) e37776, https://doi.org/10.7554/eLife.37776.
[44] K. Matsumoto, T.T. Mitani, S.A. Horiguchi, J. Kaneshiro, T.C. Murakami, T. Mano, H. Fujishima, A. Konno, T.M. Watanabe, H. Hirai, H.R. Ueda, Advanced CUBIC tissue clearing for whole-organ cell profiling, Nat. Protoc. 14 (2019) 3506–3537, https://doi.org/10.1038/s41596-019-0240-9.
[45] N. Renier, E.L. Adams, C. Kirst, Z. Wu, R. Azevedo, J. Kohl, A.E. Autry, L. Kadiri, K. Umadevi Venkataraju, Y. Zhou, V.X. Wang, C.Y. Tang, O. Olsen, C. Dulac, P. Osten, M. Tessier-Lavigne, Mapping of brain activity by automated volume analysis of immediate early genes, Cell 165 (2016) 1789–1802, https://doi.org/10.1016/j.cell.2016.05.007.
[46] J.W. Wiseman, E.S. Scott, P.A. Shaw, W.H. Colledge, Enhancement of gene delivery to human airway epithelial cells *in vitro* using a peptide from the polyoma virus protein VP1, J. Gene Med. 7 (2005) 759–770, https://doi.org/10.1002/jgm.721.
[47] H. Li, R. Durbin, Fast and accurate short read alignment with burrows-wheeler transform, Bioinformatics 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.
[48] P.J. Cock, T. Antao, J.T. Chang, B.A. Chapman, C.J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, M.J. de Hoon, Biopython: freely available python tools for computational molecular biology and bioinformatics, Bioinformatics 25 (2009) 1422–1423, https://doi.org/10.1093/bioinformatics/btp163.
[49] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in python, Nat. Methods 17 (2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[50] A. Tareen, J.B. Kinney, Logomaker: beautiful sequence logos in python, Bioinformatics 36 (2020) 2272–2274, https://doi.org/10.1093/bioinformatics/btz921.
[51] ⟨https://github.com/althonos/peptides.py⟩.
[52] R.P. Joosten, T.A. te Beek, E. Krieger, M.L. Hekkelman, R.W. Hooft, R. Schneider, C. Sander, G. Vriend, A series of PDB related databases for everyday needs, Nucleic Acids Res 39 (2011) D411–419, https://doi.org/10.1093/nar/gkq1105.
[53] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637, https://doi.org/10.1002/bip.360221211.
[54] T. Zhao, Y. Bao, X. Gan, J. Wang, Q. Chen, Z. Dai, B. Liu, A. Wang, S. Sun, F. Yang, L. Wang, DNA methylation-regulated QPCT promotes sunitinib resistance by increasing HRAS stability in renal cell carcinoma, Theranostics 9 (2019) 6175–6190, https://doi.org/10.7150/thno.35572.