UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
INSTITUTE OF MOLECULAR AND CELL BIOLOGY
FACULTY OF MEDICINE
INSTITUTE OF BIOMEDICINE AND TRANSLATIONAL MEDICINE
LABORATORY OF PRECISION AND NANOMEDICINE

Development of tissue-specific homing peptides using probabilistic machine learning

Bachelor's thesis
12 EAP

Jasper August Tootsi

Supervisors
Prof Tambet Teesalu,
Prof Juhan Sedman,
Dr Karlis Pleiko

TARTU 2025

Abstracts

Tõenäosuslike masinõppe meetodite rakendamine kullerpeptiidide arendamiseks

Kompleksete haiguste ravi probleemiks on ravimite tugevad kõrvalmõjud ning madal efektiivsus. Üheks võimaluseks sellise ravi terapeutilise efektiivsuse suurendamiseks on selle kombineerimine koespetsiifiliste kullerpeptiididega, suurendades sellega ravimi hulka haiguskoldes ja vähendades kõrvalmõjusid. Käesolevas töös implementeeriti ning testiti integreeritud *in silico* ja *in vitro* arendusprotsessi, mis potentsiaalselt kiirendab retseptor-spetsiifiliste kullerpeptiidide avastamist ja optimeerimist. Töö käigus genereeriti NRP-1 ja p32 seonduvaid kandidaatpeptiide. Kandidaatide selektiivsust ja biodistributsiooni hinnati kõrge ja madala koopiaarvuga faagidispleiga nii rakuvabas süsteemis kui ka *in vivo* katsetega hiirtemudelitel. Katsete tulemused näitavad, et multivalentsus mängib olulist rolli koe-spetsiifiliste peptiidide seondumisel. Samuti kinnitab antud töö tõenäosuslike masinõppemudelite potentsiaali retseptor-spetsiifiliste peptiidide avastamisel.

Märksõnad: generatiivsed mudelid, faagidisplei, kullerpeptiidid, täppis-nanomeditsiin
CERCS kood: T490 Biotehnoloogia

Development of tissue-specific homing peptides using probabilistic machine learning

A central challenge for curing complex diseases is inefficient drug delivery. One potential solution is to use homing peptides that target unique tissue markers for selective drug delivery, increasing therapeutic effectiveness and decreasing side effects. This thesis presents an integrated *in silico* and *in vitro* workflow that accelerates the discovery and optimization of receptor-specific homing peptides from months and years to weeks. High-copy T7-phage display proved the importance of multivalency, showing higher binding in all cases. These results show that generative modelling can learn receptor-specific pharmacophores and potentially yield functional ligands, offering a hand to time-consuming receptor-aware peptide design.

Keywords: generative models, phage display, homing peptides, targeted nanomedicine
CERCS code: T490 Biotechnology

Table of Contents

1. Abbreviations

ADCP	AutoDock CrankPep
ANOVA	analysis of variance
BSA	bovine serum albumin
CendR	C-end Rule
ECM	extracellular matrix
FDA	Food and Drug Administration
iRGD	internalizing RGD, tumor-penetrating peptide, sequence CRGDKGPDC
IV	intravenous
linTT1	linear TT1, p32-binding homing peptide, sequence AKRGARSTA
Lyp-1	p32-homing peptide, sequence CGNKRTRGC
NRP-1	neuropilin-1
RES	reticuloendothelial system
RPAR	prototypic CendR peptide, NRP-1 binding homing peptide, sequence RPARPAR
SEM	standard error of the mean
VEGF	vascular endothelial growth factor

2. Introduction

Targeted drug delivery remains one of the central obstacles in treating cancer and other complex diseases, as most available therapeutics act systematically, requiring high concentrations with reduced efficacy and off-target toxicity (Anchordoquy et al., 2024; Ruoslahti, 2012). Short, receptor-specific homing peptides offer an elegant workaround because of their low molecular weight offers rapid extravasation into deep tissue while clearing just as quickly from healthy tissue (Lingasamy & Teesalu, 2021). Their modular chemistry allows attachment of a wide variety of imaging agents, cytotoxins, or nanoparticles for theranostic use (Ruoslahti, 2012; Xiao et al., 2025). Over the past three decades, phage display selections in live animals have mapped dozens of vascular ZIP codes, from brain to kidney, to tumor stroma, establishing Neuropilin-1 (NRP-1) and the mitochondrial protein p32 as clinically relevant target proteins (Teesalu et al., 2012).

Yet classical *in vivo* phage display is resource intensive. A single mouse can receive only ca 10^{10} phage particles per injection, meaning that even with a CX_7C (where X corresponds to any amino acid) constrained library, only around 15% of sequence space is covered (Põšnograjeva et al., 2025; Sloth et al., 2022; Teesalu et al., 2012). In this case, achieving meaningful enrichment and convergence requires multiple rounds with several animals in each of the rounds and weeks of wet-lab labor (Pleiko et al., 2021; Ruoslahti, 2017). These physical constraints incentivize a hybrid approach where machine learning can explore the enormous combinatorial sequence landscape *in silico* before a wet-lab experiment is warranted (Zhou & Huang, 2024).

Recent probabilistic machine learning models learn not only which amino-acid motifs prefer a given receptor but also how the backbone torsion angles accommodate the target receptor pocket, returning quantitative uncertainty along with each prediction (Ben-Hur & Noble, 2005; Ho et al., 2020; Vázquez Torres et al., 2025; Watson et al., 2023). Diffusion architectures drive this further by generating backbones from random noise a single atom at a time (Ho et al., 2020; Lin et al., 2024; Watson et al., 2023). PPFlow, implementing conditional flow-matching as the next iteration of this denoising approach, steer the generation towards a shaped binding

pocket, reducing post-processing and preserving stereochemistry for the resulting peptides (Lin et al., 2024).

This thesis explores embedding a target-aware generative model into an *in silico* phage-display loop to speed up homing peptide discovery. Specifically, we 1) re-implemented and fine-tuned PPFlow on the University of Tartu HPC cluster, 2) built an automated pipeline from target receptor establishment to candidate peptide sequences, 3) used the setup to identify candidate peptides for NRP-1 and p32, and 4) validated their specificity through cell-free receptor binding assays and *in vivo* play-off.

This work was completed in the Laboratory of Precision and Nanomedicine, Institute of Biomedicine and Translational Medicine, University of Tartu, Estonia.

3. Literature Review

3.1. Affinity ligands

Affinity ligands are either natural or synthetic molecules that bind reversibly to a target with high specificity and affinity, providing a method to selectively recognize, capture, or modify the target (Lei et al., 2022). Affinity-based targeting of blood vessels, lymphatic systems, or even tumor vasculature-specific receptors can be done with the help of antibodies, aptamers, small molecules, or peptides (Laakkonen et al., 2002; Ruoslahti, 2002a, 2004; Teesalu et al., 2012).

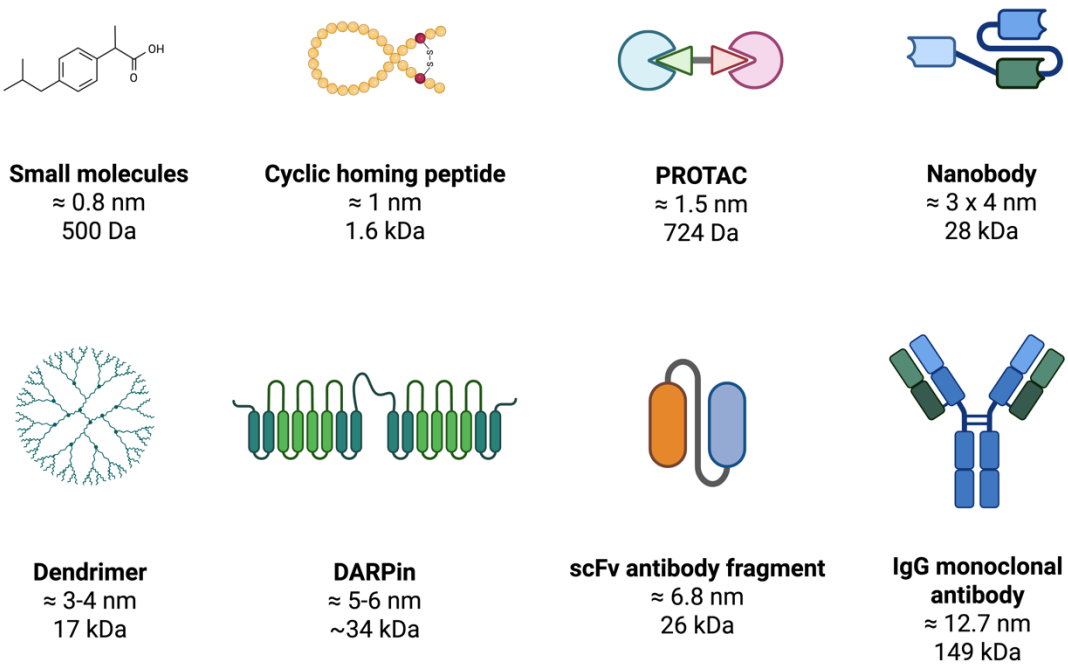

Figure 1. A selection of different affinity ligands with their weights and sizes. Created with Biorender.com.

Antibodies serve as affinity ligands and are currently very popular, and for good reason: they provide picomolar binding with good specificity (Lei et al., 2022). These Y-shaped, heavy (approximately 150 kDa) proteins are 10 nm (size comparison in Fig. 1) in length (Reth, 2013). Out of 50 molecules approved by the FDA in 2024 (Fig. 2), 16 were antibodies, providing relief

to patients with a wide range of diseases like esophageal squamous cell carcinoma, hemophilia A, and pancreatic adenocarcinoma (Mullard, 2025). At the same time, because of their large size, antibodies come with several challenges. They are expensive to manufacture, clear relatively slowly systemically, and are unable to penetrate deep (malignant or healthy) tissues or reliably cross the blood-brain barrier (Blanco et al., 2015; Jovčevska & Muyldermans, 2020; Sugahara et al., 2009; Trédan et al., 2007).

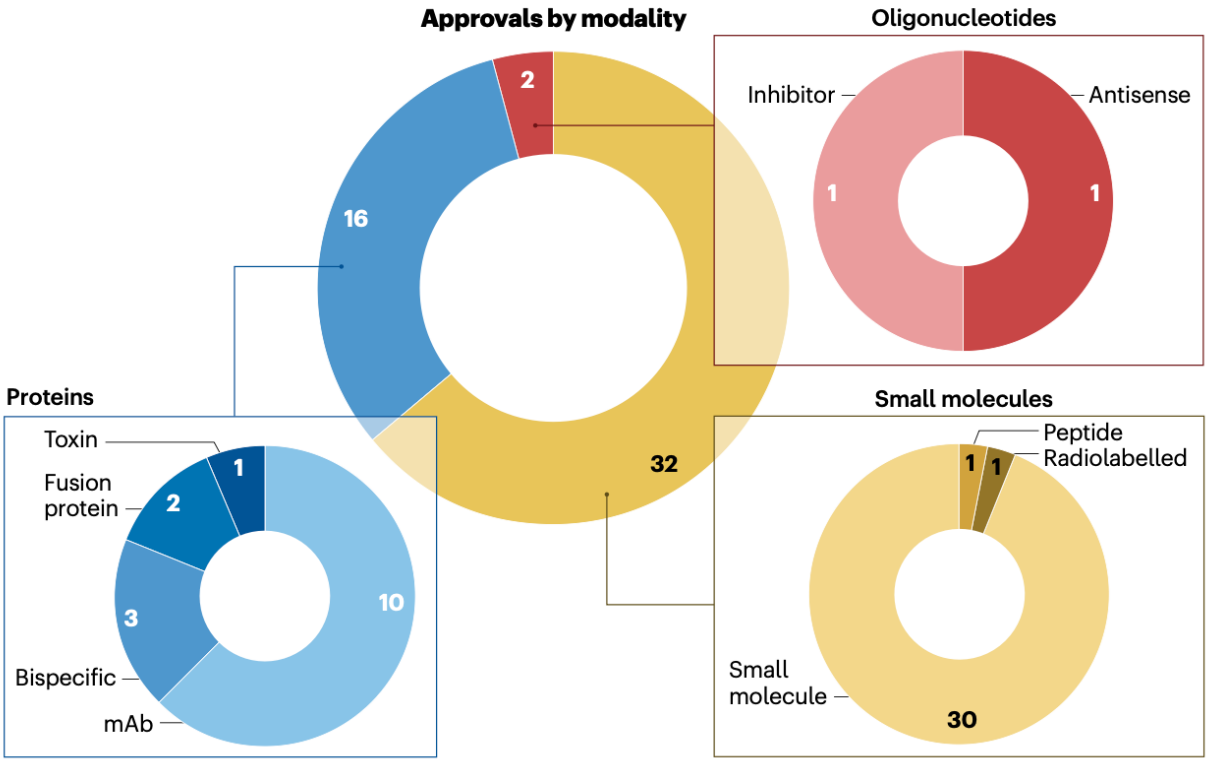

Figure 2. Breakdown of molecules approved in 2024 by the FDA Center for Drug Evaluation and Research. From Asher Mullard 2025 report; https://doi.org/10.1038/d41573-025-00001-5

3.2. Homing Peptides

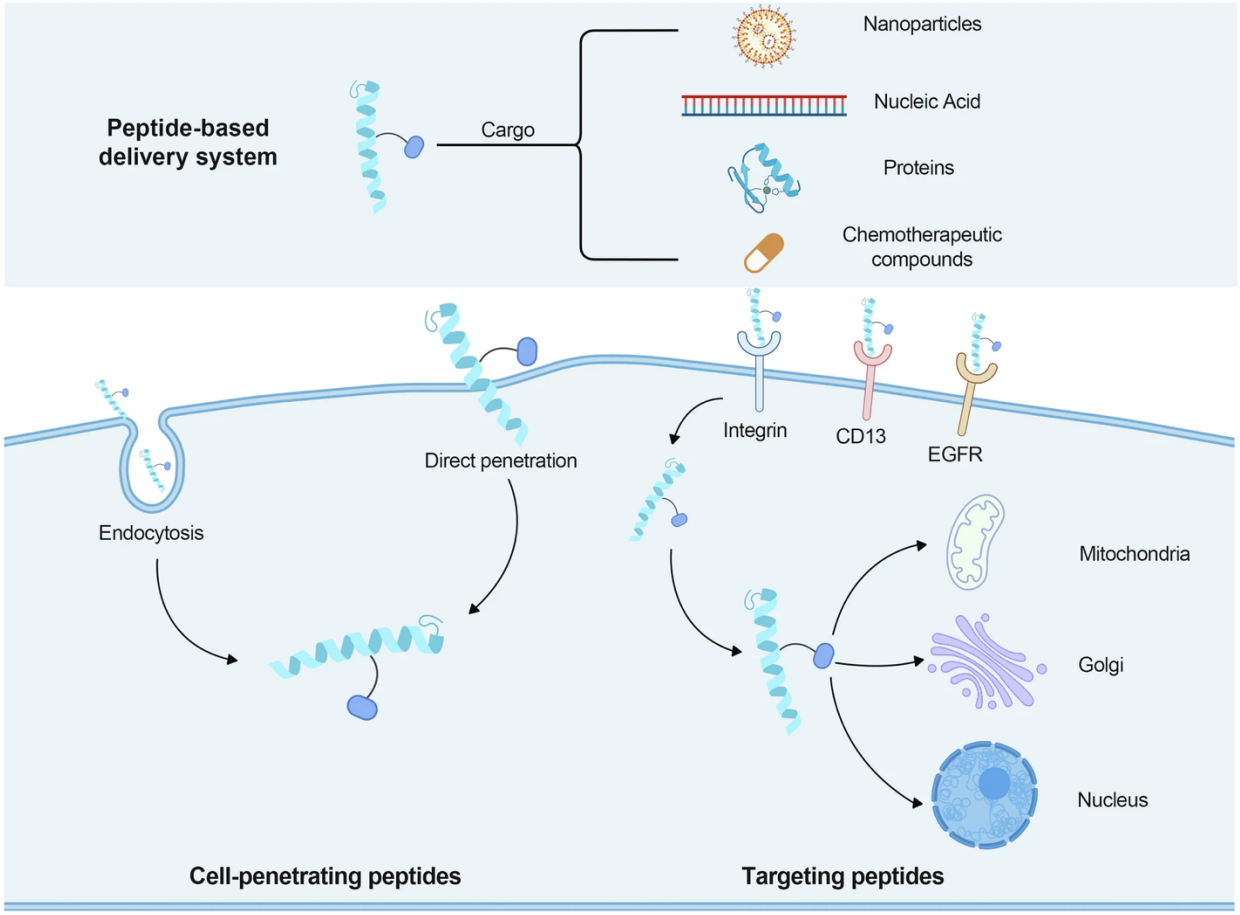

Figure 3. Tumor homing and cell-penetrating peptides with multiple types of carriers and possible cargo. Illustration from Xiao et al., 2025.

As another class of affinity ligands, homing or targeting peptides are short amino acid sequences (5-30 aa) that accumulate selectively in a tissue, cell type, or extracellular matrix (ECM) (Bodanszky, 1988). Their homing and binding are achieved through protein-to-protein interactions between specific amino acid sequence motifs and cell or extracellular matrix (EC) receptors, like integrins, Neuropilin-1 (NRP-1), and p32 (also known as gC1qR) (Laakkonen et al., 2002; Lingasamy & Teesalu, 2021; Ruoslahti, 2017; Teesalu et al., 2009). Because of their small size, they enable guided delivery payloads to tissues that antibodies cannot reach because of biological barriers (Ruoslahti, 2017; Xian et al., 2025). Together with methods to enhance their stability and *in vivo* degradation decreasing strategies, such as cyclization, conjugation to nanoparticles, synthetic amino acids, they offer a well-rounded technology alternative against

other affinity ligands (Kue et al., 2016; Lingasamy & Teesalu, 2021; Xian et al., 2025). Peptides can be used as detection or targeting agents for cancer microenvironment or deep tissue (example mechanisms with payloads and carriers provided in Fig. 3), offering a more affordable alternative to antibodies (Laakkonen et al., 2002; Ladner et al., 2004; Sugahara et al., 2009; Vadevoo et al., 2019; Yu et al., 2022).

Table 1. A list of blockbuster peptide drugs based on 2022 retail sales. Table is from Chang et al., (2024)

Peptide	Brand name	Length	Year of approval	Disease	Retail sales (billion USD)
Insulin	Humulin, NovoRapid	51	1982	Diabetes	15.12
Semaglutide	Ozempic, Wegovy	31	2012	Diabetes	11.287
Dulaglutide	Trulicity	31	2014	Diabetes	7.439
Liraglutide	Saxenda	30	2010	Nutritional deficiency	1.556
Carfilzomib	Kyprolis	4	2012	Oncology	1.328
Lanreotide	Somatuline	8	2014	Oncology	1.306
Octreotide	Sandostatin	8	1989	Oncology	1.238
Fam-trastuzumab Deruxtecan-nxki	Enhertu	4	2019	Oncology	1.096
Linaclotide	Linzess	14	2012	Gastrointestinal diseases	1.088
Goserelin	Zoladex	10	1989	Oncology	0.927
Leuprorelin Acetate	Leuplin	9	1995	Oncology	0.843
Ixazomib	Ninlaro	2	2015	Oncology	0.743
Glatiramer	Copaxone	4	2014	Neurology	0.691
Cyclosporine	Restasis	11	1983	Ophthalmology	0.666
Etelcalcetide	Parsabiv	8	2017	Endocrinology	0.447

Note: The color code classifies the first row across three categories (orally → red, subcutaneously → yellow, intravenously → cyan).

Peptides are an excellent modality for multiple clinically relevant theranostic directions (as seen in Table 1 outlining peptide-based molecule retail sales), because of their ability to penetrate diseased tissue, such as tumors, or damaged vasculature, and their quick extravasation caused by their low molecular weight (Chang et al., 2024; Mullard, 2025). At the same time, their small size also increases their systemic clearance, reducing off-target toxicity in non-target organs (Anchordoquy et al., 2024; Srinivasarao & Low, 2017).

The primary method for discovering homing peptides has historically been and often still is experimental screening via combinatorial peptide phage display libraries (Xiao et al., 2025).

3.3. Phage Display

Phage display is an advancement in molecular biology that allows the display of peptides on the surface of bacteriophages. This technique facilitates the identification of peptides with high specificity and affinity for a given target, rendering it invaluable for therapeutic, diagnostic, and research applications. Over the years, it has progressed from small volumes in tubes with specific targets known beforehand to protocols where small animals like rats or mice are used for organ- or disease-tropic peptide selection (Pasqualini & Ruoslahti, 1996).

The initial concept of phage display was first introduced by George P. Smith in 1985, who demonstrated that foreign peptides could be fused to the minor coat protein (pIII) of filamentous M13 bacteriophages, which results in their display on the phage surface (Smith, 1985). His innovative method established a direct link between the display phenotype and the encoding genotype within the same virus particle and became so profoundly important that George P. Smith and Greg Winter won the 2018 Nobel Prize in chemistry for developing the antibody and peptide phage display (*All Nobel Prizes 2018*, n.d.; Smith, 1985). This allowed the selection of peptides with desired binding properties from vast combinatorial libraries.

Parmley and Smith improved the method by constructing random peptide libraries displayed on bacteriophages, which significantly enhanced the utility of this technology for identifying peptide ligands to antibodies, receptors, and other proteins (Parmley & Smith, 1988). Since then, phage display has been improved by many to use various bacteriophages, coat proteins, and display systems, further enhancing the usefulness of the technology for drug and diagnostic system development (Kong et al., 2020; Krumpe et al., 2006; Ladner et al., 2004; Pasqualini & Ruoslahti, 1996).

3.4. T7 Phage Display System

The *E. coli* infecting T7 lytic phage has several advantages over the initial M13-based phage display system, as it has fewer amino acid biases, greater stability, and decreased mutation rate, and a compact 60 nm capsid diameter that resembles many synthetic nanoparticles (Danner & Belasco, 2001; Dunn et al., 1983; Krumpe et al., 2006). Another factor benefiting the strictly lytic

T7 system is the release of progeny by host-cell lysis, completely bypassing the M13 constraining periplasmic export pathway (Yue et al., 2022). This cytoplasmic pathway allows efficient display of highly charged, cysteine-rich, or folded protein domains, greatly broadening the chemical and structural diversity accessible for the display library (Krumpe et al., 2006).

The T7 phage display consists of six key steps, illustrated in Fig. 4. The T7 display system involves fusing foreign peptide sequences to the C-terminus of the 10B major capsid protein or 10A minor capsid protein (Krumpe et al., 2006). This enables the peptide to be encoded and then displayed on the phage surface. This recombinant phage is then packaged using a T7 packaging extract and then amplified in an exponential-growth phase *E. coli* culture (Teesalu et al., 2012).

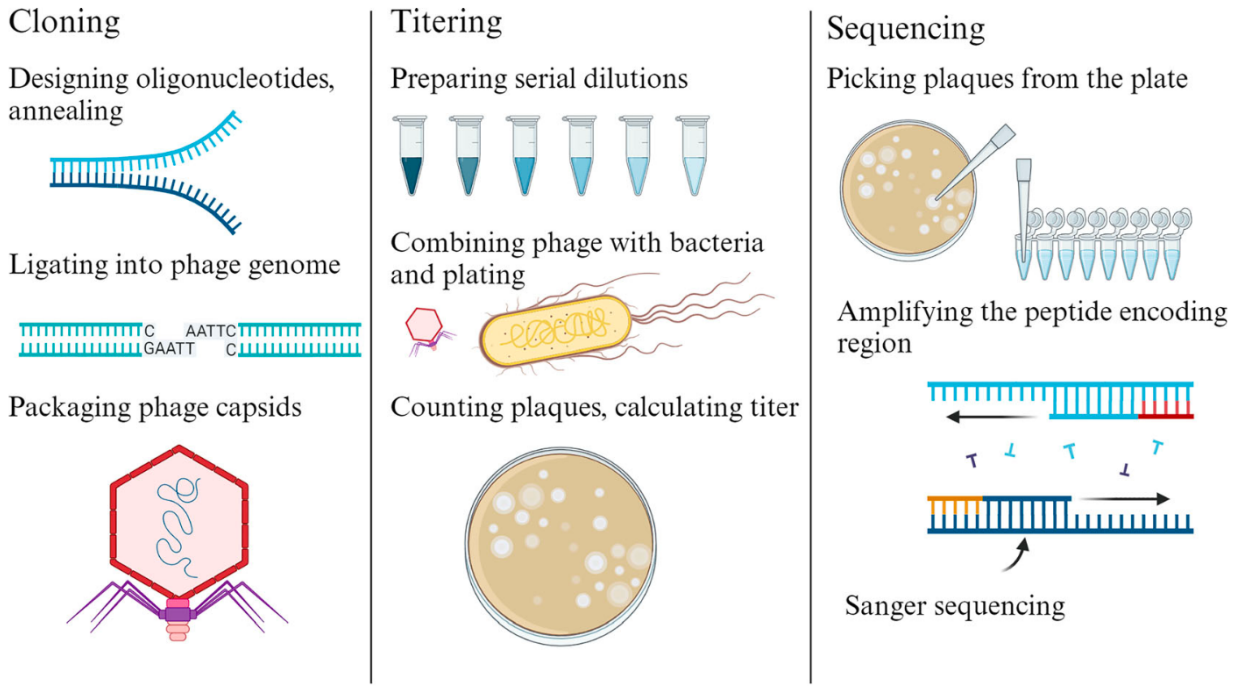

Figure 4. Schematic representation of basic manipulation of T7 phages. The main steps of cloning, titering, and sequencing of the T7 phage display. Illustration from Põšnograjeva et al., 2025.

Next, a cDNA phage display library is developed (Marshall & Noireaux, 2018). Depending on the peptide's length, libraries often incorporate millions of independent, unique clones to maximize

diversity and enhance the likelihood of identifying high-affinity target-binding peptides (Pleiko et al., 2021). The source of the peptide sequences may come from a specific tissue or be completely random, to ensure extensive coverage of the protein landscape (Pasqualini & Ruoslahti, 1996). (Danner & Belasco, 2001)

Typically, phage biopanning is performed on purified target proteins or RNAs to be synthesized *in vitro* and immobilized on a solid support such as magnetic beads. The phage library is incubated with this molecular bait to enrich for phage clones displaying peptides that selectively bind the target (Laakkonen et al., 2002). The bound phages are then re-amplified for further biopanning rounds (Danner & Belasco, 2001; Kim et al., 2004).

These cycles of iterative selection and enrichment are crucial to isolate good homing peptide-displaying phages (Ruoslahti, 2017). Each cycle involves binding, capture, washing, and phage amplification, gradually narrowing the diversity of the unique phage pool while enriching it with peptides showing high affinities toward the target molecule (Lingasamy & Teesalu, 2021). Enrichment after each selection round can be monitored by phage titration, while shifts in the peptide landscape are detectable through high-throughput DNA sequencing of the peptide-encoding region of the phage genome (Pleiko et al., 2021).

To further refine and increase stringency, competitive molecules can be introduced during the selection. As an example, a non-biotinylated RNA might bind high-affinity target molecules, allowing amplification of lower-affinity peptides that would otherwise be outcompeted (Danner & Belasco, 2001; Kolonin et al., 2006). (Danner & Belasco, 2001)

Finally, the select few phage clones undergo characterization, validation, and optimization. Sequencing, structural analysis, and affinity tests confirm the identity and specificity of these exposed targeting peptides, ensuring reliable and robust experimental results. (Danner & Belasco, 2001; Pasqualini & Ruoslahti, 1996; Pleiko et al., 2021; Sloth et al., 2022)

3.5. Vasculature ZIP Codes

In higher organisms, the circulatory and lymphatic systems form an extensive network of vessels that support nutrient delivery and waste removal. These vessels are uniformly lined by a single layer of endothelial cells, which play key roles in vascular function, homeostasis, and tissue-specific regulation. Blood vessels exhibit distinct structural and molecular characteristics shaped by their tissue microenvironment, including unique morphologies and expression profiles (Fig. 5) (Gomez-Salinero et al., 2025; Potente & Mäkinen, 2017).

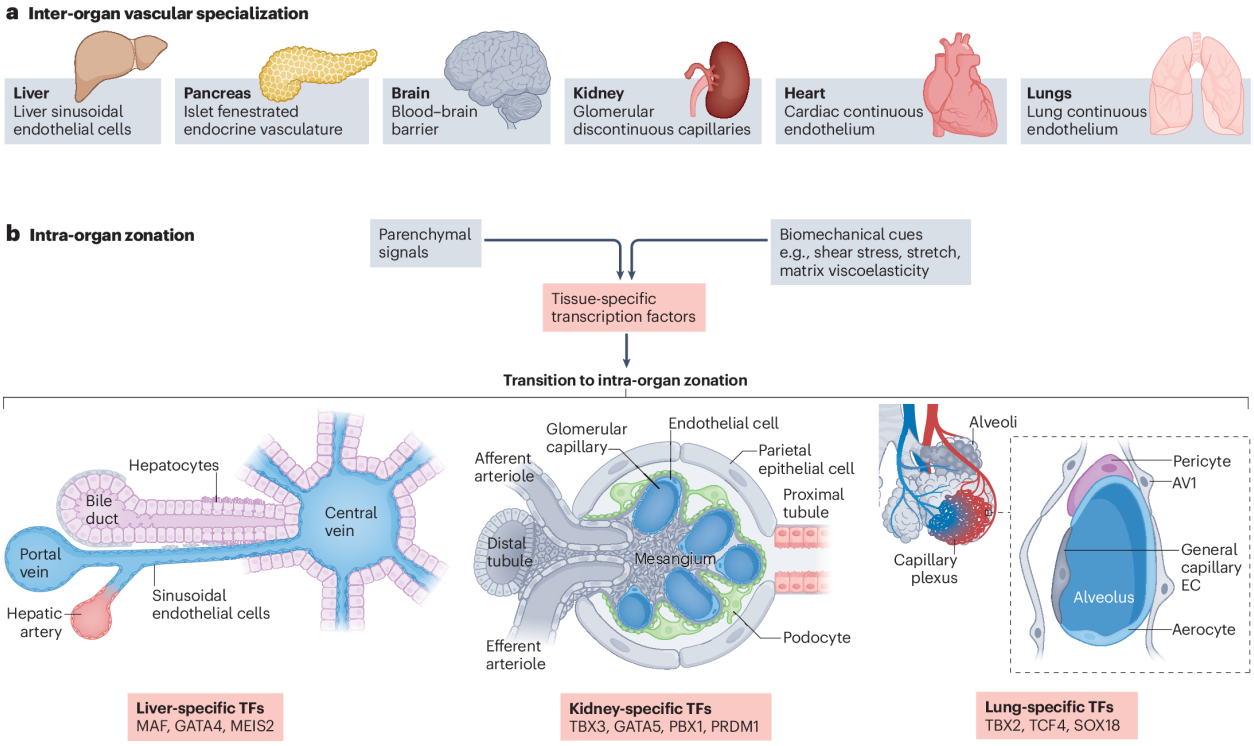

Figure 5. Organ vascularization varies wildly based on heterogenous induction of tissue-specific endothelial cell transcription factors. Illustration from Gomez-Salinero et al., (2025)

This unique tissue receptor expression patterning provides a way to target these different tissue zones or "vasculature ZIP codes" with affinity-based binding molecules (e.g. homing peptides in Fig. 6) and even trigger biological pathways (Ruoslahti, 2004; Teesalu et al., 2012).

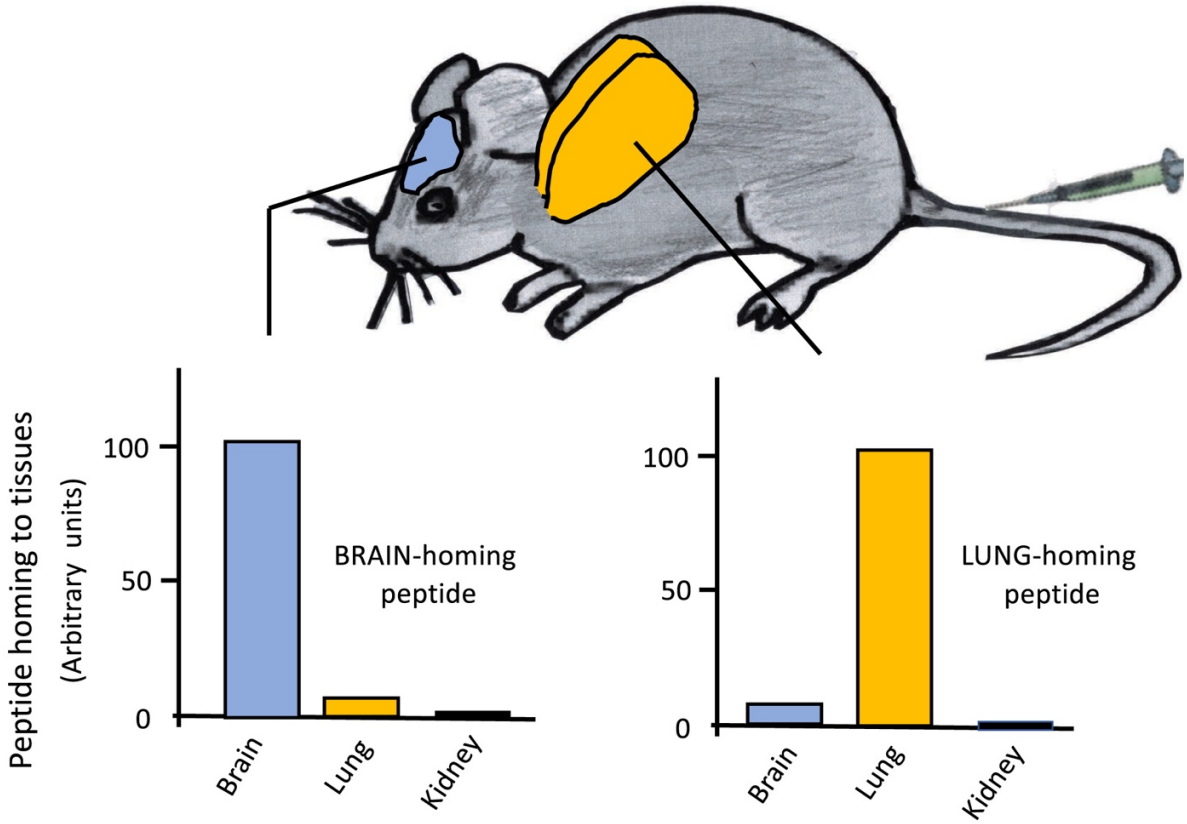

Figure 6. Illustrative example of brain and lung-specific homing peptides. Illustration from Ruoslahti, (2022)

In vivo phage display was introduced by Erkki Ruoslahti and collaborators in their pioneering work in 1996. *In vivo* phage display entails injecting living animals (often a mouse) with a random peptide library expressed on the surface of phages as previously described in "Phage Display". After circulation and organ isolation, phages are recovered from each organ separately, re-amplified, and re-injected for multiple rounds. This approach, for the first time, showed selective enrichment (exemplified in Fig. 6) of different peptide motifs in different organs' blood vessels, aligning with the theory that each organ's blood vessels express unique receptors that are accessible from the luminal side. These first findings mapped two initial "zip codes" with two different peptides (SRL, CLPAVP), one peptide for the kidney and one peptide for the brain blood vessels (Pasqualini & Ruoslahti, 1996).

After initial reporting on this breakthrough technology, multiple studies reconfirmed and expanded on the *in vivo* phage display to other healthy and diseased tissues. Ruoslahti's lab and

others identified homing peptides for lungs, skin, pancreas, liver, and more (Koivunen et al., n.d.; Rajotte & Ruoslahti, 1999). Through the injection of tumor-bearing mice and the harvesting of tumors, peptides that selectively bound to tumor blood vessels were discovered. One of such peptide motifs was Arg-Gly-Asp (RGD), found to bind to $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins, targeting several tumor tissues in a highly specific and selective manner (Laakkonen et al., 2002; Ruoslahti, 1996). The second tumor homing motif, Asn-Gly-Arg (NGR), targeted CD13 receptors overexpressed on tumor cells (Graziadio et al., 2016; Ruoslahti, 1996). These two peptides were the first generation of tumor-homing peptides that were also shown to have the ability to transport conjugated drugs to tumors in animal models (Arap et al., 1998).

In 2002, Laakkonen et al., building on previous work, reported a new cyclic peptide CGNKRTRGC (named Lyp-1) that penetrated tumor tissue, homing to tumor lymphatic vessels and tumor macrophages (Laakkonen et al., 2002). Its target receptor was later identified as p32 (also known as gC1qR or HABP1), a protein selectively expressed in tumor lymph cells (Fogal et al., 2008). This expanded the "vasculature ZIP code" idea to mapping vascular ZIP codes beyond blood vessels to lymphatic vessels or markers in extravascular tumor stroma.

3.6. Advantages and limitations of *In Vivo* Phage Display

As early *in vivo* screenings that injected highly diverse peptide libraries into mice demonstrated, *in vivo* phage display can uncover organ-specific "vascular ZIP codes" without any previous knowledge of target receptors, allowing for an unbiased discovery of new tissue-targeting peptides (Ruoslahti, 2002b). This gives researchers the ability to do random peptide library selection under fully perfused conditions, where receptors are properly oriented, glycosylated, and in their native active state, to find homing peptides with better homing in clinical applications (Arap et al., 2002). Furthermore, the combination of biopanning and high-throughput sequencing (HTS) enables quantification and tracking of each peptide's biodistribution across rounds with high resolution (Pleiko et al., 2021).

In addition, the whole organism context provides constant negative selection pressure to all injected peptide-phages caused by non-specific phage binding at off-target vascular beds,

removing them from selection (Arap et al., 1998; Teesalu et al., 2012). The whole-body context, therefore, enriches biologically relevant binders while simultaneously depleting pan-cell binding phage from circulation.

One of the limitations of *in vivo* phage display (illustrated in Fig. 7) in general is the practical limits of the quantity that could be injected and the diversity of the initial naive screening library (Koivunen et al., n.d.; Sloth et al., 2022). Considering that most of the studies are done in mice with a tolerated intravenous bolus volume of the phage pool being ~200 µl depending on mouse size, we see that the with the assumption of 5×10^{10} PFU/ml phage pool concentration we will have 1×10^{10} phages and for meaningful results, if we have 50 copies of each unique peptide display bacteriophage particles, we will have 200 million (2×10^{8}) different display peptides (Healing & Smith, 2000; Sloth et al., 2022; Teesalu et al., 2012). Conducting an experiment with a classical CX_7C (X– random amino acid), where two cysteines flank both sides of a 7-amino acid sequence peptide library, the theoretical sequence coverage of the whole peptide space is $20^7 \approx 1.28 \times 10^9$, meaning that a single mouse experiment cover only around 15% of the possible variants; of course, the sampling coverage becomes orders of magnitude worse for libraries with an even longer random region (Kong et al., 2020; Teesalu et al., 2012).

Another factor influencing *in* vivo phage selection specifically is the extremely rapid reticulo-endothelial (reticuloendothelial system, RES) clearance of T7 virions (Huh et al., 2019). It has been shown by systemic pharmacokinetic studies on mice that injected T7 particles fall to ~1 % of the input dose within 1 hour of tail-vein injection (Srivastava et al., 2004). Confirming the liver as a phage sink, quantitative biodistribution studies locate >90 % of the sequestered injected phages in the liver by 20–30 min after injection, through adsorption to Kupffer cells and then transferring the particles to hepatocytes (Ludtke et al., 2007). This liver-centric filtration forces an intrinsic negative selection during *in vivo* biopanning as phages that fail to engage vascular targets are swept into the hepatic RES and destroyed, whereas clones that bind to tissue receptors survive and are preferentially recovered. This results in higher enrichment in non-liver tissues but also biases the library towards peptides that evade this sink and should be taken into account during selection schedule design (Huh et al., 2019; Ludtke et al., 2007).

These constraints bring out why *in vivo* biopanning typically requires multiple rounds, *ex vivo* experiments, or pooling across multiple animals to achieve comprehensive exploration of the peptide variant landscape. One of the ways to mitigate these physical limitations is to combine the random library-based directed evolution approach with modern computational methods. (Teesalu et al., 2012)

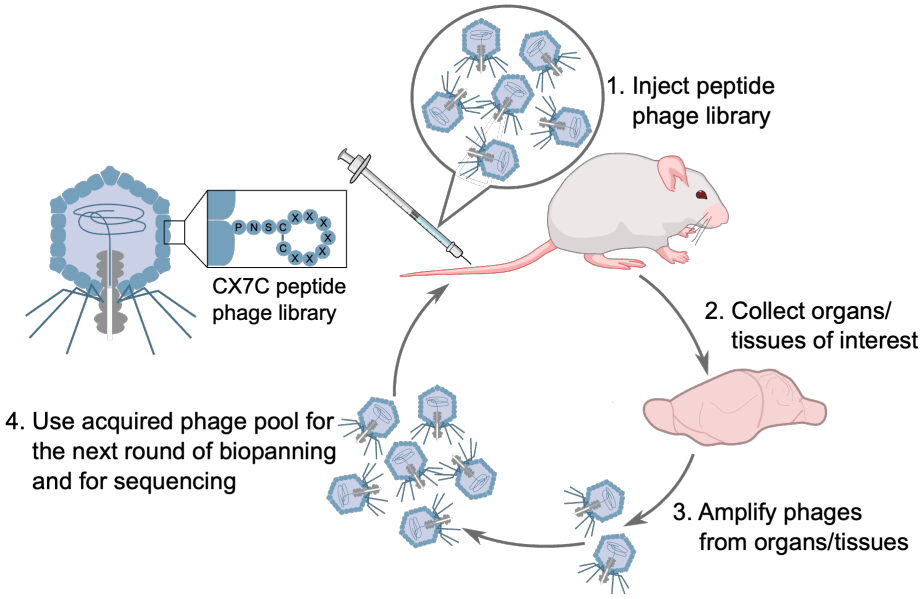

Figure 7. In vivo phage display schematic. Illustration made by Kristina Põšnograjeva

3.7. Targets for Homing Peptide Generation

3.7.1. NRP-1 as a homing peptide target for Drug Delivery

Neuropilin-1 (NRP-1, Uniprot No: O14786) is a 923-residue transmembrane glycoprotein that has five extracellular domains (a1, a2 CUB; b1, b2 coagulation-factor V/VIII; c/MAM), a single-pass helix, and a 44-residue cytoplasmic tail (Nakamura & Goshima, 2013).

Beyond its vital roles during development in axon guidance and angiogenesis, NRP-1 coordinates vascular permeability, immune synapse stability, and, recently recognized, even SARS-CoV-2 entry, acting in cooperation with other receptors (Cantuti-Castelvetri et al., 2020; Dragoni et al., 2024). Pathologically, NRP-1 is upregulated on tip endothelial cells during

angiogenesis in a wide spectrum of solid-tumor cells (Jubb et al., 2012; Varanasi et al., 2025). It is also highly expressed on immunosuppressive regulatory T cells ($CD4^+$), amplifying VEGF and Semaphorin signals to drive new vessel sprouting, invasion, and immune evasion (Ding et al., 2019; Dragoni et al., 2024; Ruoslahti, 2004; Varanasi et al., 2025).

C-end rule or CendR peptides with an R/KXXR/K recognition motif use the NRP-1 dependent pathway to induce tissue permeability (Sugahara et al., 2009, 2010). It has been shown with crystal and cryo-EM structures that the b1 domain forms a deep pocket for a peptide with a C-terminal arginine (CendR) motif from physiological ligands like vascular endothelial cell growth factor (VEGF) (as seen in Fig. 8) and Semaphorin-3A (Varanasi et al., 2025). Peptides such as iRGD (CRGDKGPDC, LSTA1) exploit this accessibility to trigger deep tumor penetration and extravasation, reaching stage II clinical trials in pancreatic cancer in combination with other drugs (Sugahara et al., 2009, Clinical trial ID: NCT05712356). In combination, the disease-triggered over-expression, a well-known binding cavity exposed extracellularly, and a demonstrated clinical translational effect make NRP-1 a great target for rational peptide design.

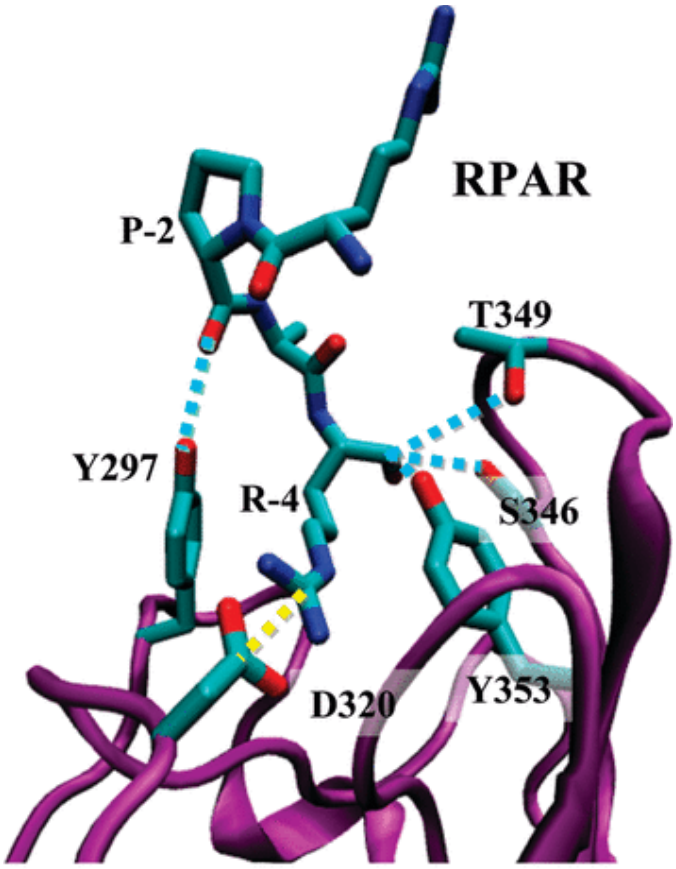

Figure 8. Ribbon representation of the complex of NRP-1 b1 domain and RPAR peptide with important interactions illustrated between them; hydrogen bonds in blue and salt bridges in yellow dashed lines. Modified illustration from Haspel et al., (2011).

3.7.2. P32 as a homing peptide target for Tumor drug delivery

P32 is a doughnut-shaped acidic trimer receptor (crystal structure Fig. 9) that orchestrates complement activation, coagulation, and metabolic stress pathways that sustain tumor survival (J. Jiang,Y. Zhang,A.R. Krainer,& R. Xu, n.d.). The 282-residue acidic protein has a molecular weight of ~96 kDa (J. Jiang,Y. Zhang,A.R. Krainer,& R. Xu, n.d.). P32, normally a mitochondrial matrix protein, is trafficked to and accumulated on the plasma membrane of hypoxic tumor cells, tumor-associated macrophages, and tumor lymphatic endothelial cells, while remaining infrequent on the surface of healthy adult tissue cells (Fogal et al., 2008). Its density on the cell surface correlates with poor prognosis across multiple solid tumors, like breast, colorectal, and

melanoma cancers (Fogal et al., 2008; Gotoh et al., 2018; Lei et al., 2023). Cell-surface p32 targeting homing peptide conjugated nanoparticles loaded with paclitaxel showed > 5-fold higher intratumoral accumulation and tumor growth delay in orthotopic breast and pancreatic tumor models compared to untargeted controls, reconfirming the target as clinically relevant (Paasonen et al., 2016; Sharma et al., 2017).

The identification of p32 as the peptide-binding receptor and multiple generations of affinity ligands thereafter perfectly illustrate the toilsome but powerful multi-step process of *in vivo* phage display. As the frontline discovery tool, Laakkonen and colleagues used *in vivo* phage display to initially identify LyP-1 (CGNKRTRGC), a peptide selectively targeting breast tumor lymphatic and malignant cells (Laakkonen et al., 2002). Afterwards, Fogal et al. identified the p32 to be the target receptor for LyP-1 with affinity chromatography (Fogal et al., 2008). Advancing on these findings, cell-free phage display was used to identify linTT1 (AKRGARSTA) as the second-generation p32-binding peptide and vast small-molecule libraries were screened for low molecular weight p32 ligands (Paasonen et al., 2016).

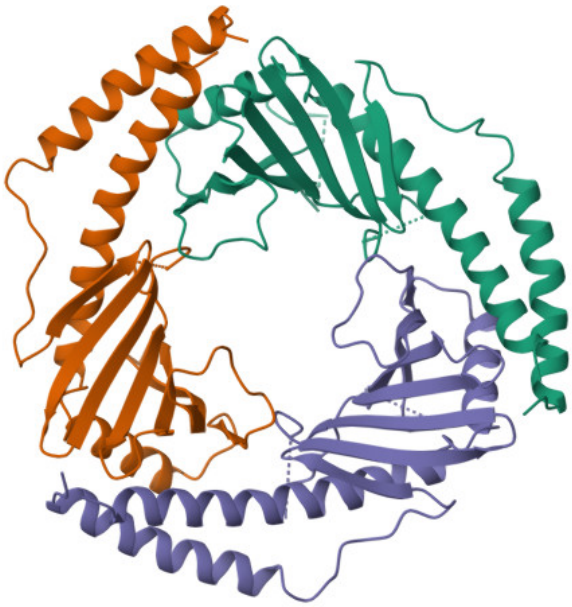

Figure 9. Crystal structure of doughnut-shaped trimeric human P32, a mitochondrial matrix protein. Structure from Jiang et al., (1999)

3.8. Machine Learning for Biology

As the number of known affinity ligands and targeting motifs has grown, researchers have turned to artificial intelligence and machine learning (ML) tooling to aid the design and subsequent optimization of new targeting molecules (Basith et al., 2020). Machine learning is a subset of artificial intelligence and consists of algorithms that learn from data to make predictions or decisions, helping to handle the complexity and scale of biological data (Chang et al., 2024; Giguère et al., 2015; Puchakayala et al., 2023; Zhou & Huang, 2024). While traditional computational approaches have supported experimental workflows through physics-based docking and heuristic scoring, modern ML models learn from the empirical data itself, potentially achieving better performance than hand-crafted models in prediction accuracy and generative abilities (Zhou & Huang, 2024). In biology, ML is particularly useful for analyzing genomic sequences, protein or peptide structures, and gene expression profiles, where traditional statistical methods often struggle to generalize and capture hidden patterns (Ching et al., 2018).

For example, researchers have combined the vasculature mapping through experimental biopanning with *in silico* approaches, where large datasets generated from previous phage display experiments are integrated into machine learning (ML) models to model peptide motif-function relationships and predict peptide-receptor interactions (Shoombuatong et al., 2019). ML-guided protein design has been especially successful where groups like the professor's. David Baker's lab and Genentech’s research accelerator have designed neutralizing proteins for snake venom, highly selective protein inhibitors for well-known cancer therapy targets, and achieved antibody 100x binding affinity maturation with an *in silico*-guided approach (Makowski et al., 2022; Roy et al., 2023; Vázquez Torres et al., 2025). A selection of deep learning models with their advantages and limitations is presented in Table 2 below.

Table 2. A constrained selection of example deep learning models used in peptide structure prediction, generation, and research. Table adapted from Wu et al., 2024.

Model name	Model type	Advantages	Limitations
AlphaFold3	Deep learning combined with residue adversarial network	High accuracy, faster than traditional protocols; requires only sequence information as input.	High demand for computing resources. Sensitivity to large-scale mutations is
RoseTTAFold	CNN combined with a multi-head self-attention	Structural prediction at the full atomic level; high computing efficiency.	Performance is inferior to AlphaFold2; data dependency.
ESMFold	Transformer	Does not require constructing MSA; inference faster than AlphaFold2.	Orphan sequence problem
iNNterfaceDesign	NN-based models	One-sided design	Depends on the known structure
AfDesign	AlphaFold1+ deep network hallucination	Solubility control; design peptide sequences with higher binding affinity.	Training process complicate
RFdiffusion	RoseTTAFold+ Diffusion	Design of diverse functional proteins(peptides) from simple molecular specifications.	Computationally expensive when dealing with large amounts of protein.
Evobind	An *in silico* directed-evolution platform based on	Design binders towards any interface conditioned on if AF can predict these.	Not possible to predict the affinity of binders.
Evoplay	A self-play RL framework based on the single-player version of AlphaZero	Efficiently design high-quality peptides for diverse applications, including PPIs, enzyme design, and drug discovery.	Depends on prior knowledge and settings.

3.8.1. Probabilistic Machine Learning

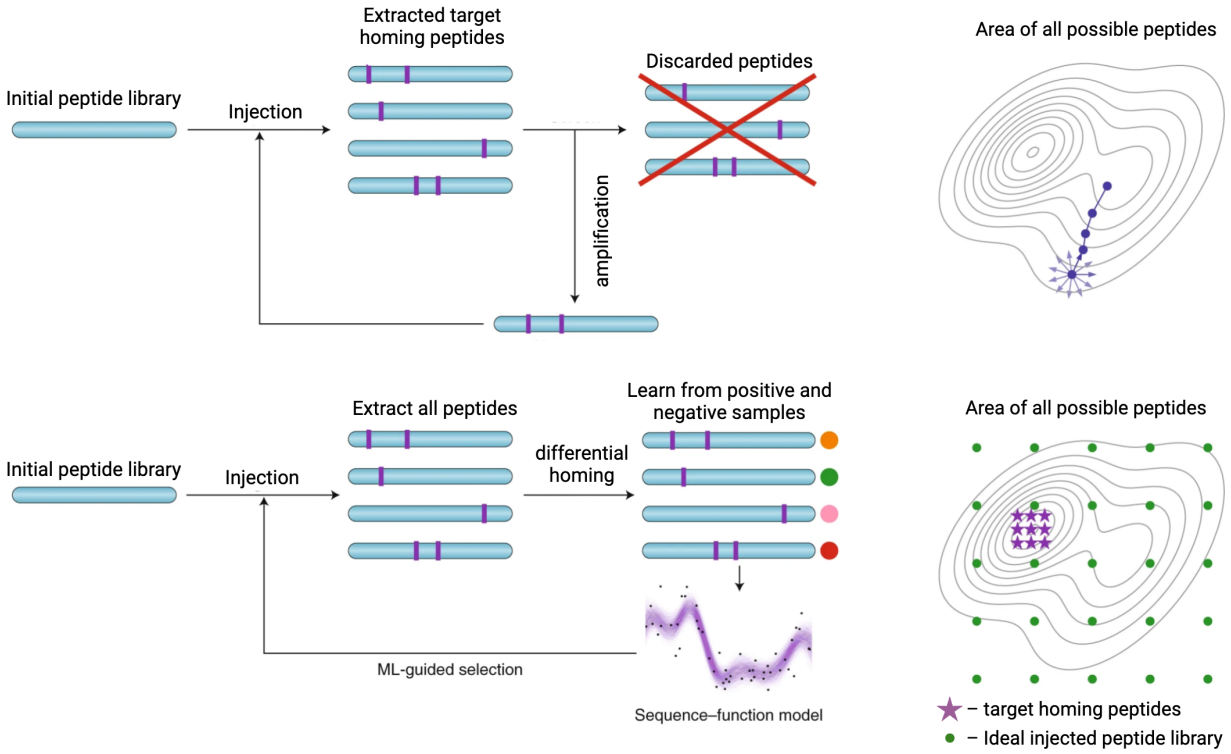

Figure 10. Comparison of classic *in vivo* peptide phage random library selection (above) versus probabilistic machine-learning guided selection learning on a *peptide* fitness landscape (below). On the right is the theoretical fitness landscape for all possible peptides and their fitness based on their homing ability. Modified from Yang et al., (2019)

Probabilistic models have a long history of use in biology, especially in sequence analysis like position weight matrices, hidden Markov models (HMMs), and Bayesian networks (Chang et al., 2024). Probabilistic machine learning is particularly valuable in biology, where data is noisy or even missing, as it also provides a quantified uncertainty metric together with the predictions. In the context of homing peptides, probabilistic machine learning can be used to capture the sequence patterns, or "rules," that underlie peptide binding to a specific target receptor (Chen et al., 2024). For example, if a number of peptides are known to bind to a specific protein, one can train a model on those same peptide sequences to learn which amino acids are favored at each position (Shoombuatong et al., 2019). This model can then generate new peptides with a high likelihood of binding. An example of this process can be seen in Fig. 10. At its core, these methods train models on experimental datasets (e.g., peptide amino acid sequences paired with measured *in vitro* target receptor binding affinities) and then apply chosen frameworks,

e.g., Gaussian processes or Bayesian neural networks, to estimate a function that predicts the bioactivity of a peptide in the presence of a target protein (Chen et al., 2024; Giguère et al., 2015; Puchakayala et al., 2023). (Bell & Chen, 2021; Ching et al., 2018; Giguère et al., 2015)

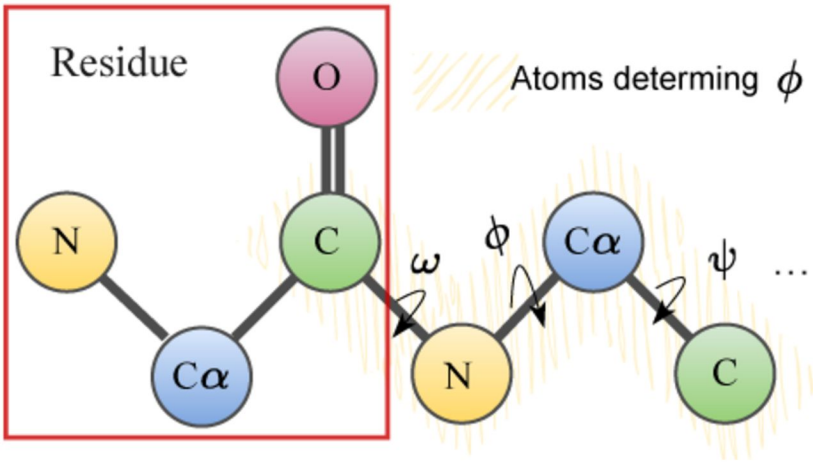

Figure 11. Torsion angles for each amino acid in the sequence as used when modelling peptides with PPFLOW. Illustration from Lin et al., 2024.

3.8.2. Diffusion ML Models

As the field of probabilistic ML in biology has progressed from only looking at the amino acid sequence of the proteins to full-atom generating models, each approach has focused more on the ability to model the geometry of the molecules in greater detail (Watson et al., 2023). This classic approach scores existing or marginally mutated sequences, telling which ones are *likely* to bind, saving wet lab time. The issue with this approach is that every candidate protein or peptide sequence still must come from somewhere, either from a random library screen or a completely human-created library of sequences. To go beyond into "generating" novel sequences, researchers like Watson et al. have turned to generative models which sample brand-new sequences or protein structures from a learned distribution (Roy et al., 2023; Watson et al., 2023). Diffusion models begin with pure random noise and then teach the ML model network to "rewind" back to the initial realistic object (in our case, a peptide), eliminating the noise in a step-by-step manner (Ho et al., 2020; Sohl-Dickstein et al., 2015). In these models, most peptide problems live in torsion-angle space (ϕ, ψ, χ) (Fig. 11), not raw XYZ

coordinates, as Cartesian coordinates waste effort on rigid-body translations and sometimes yield backbones whose internal angles violate stereochemistry, requiring cumbersome post-processing fixes. For proteins, RFdiffusion applies this idea to 3-D atomic coordinates and can output whole backbones that fold and function (Watson et al., 2023).

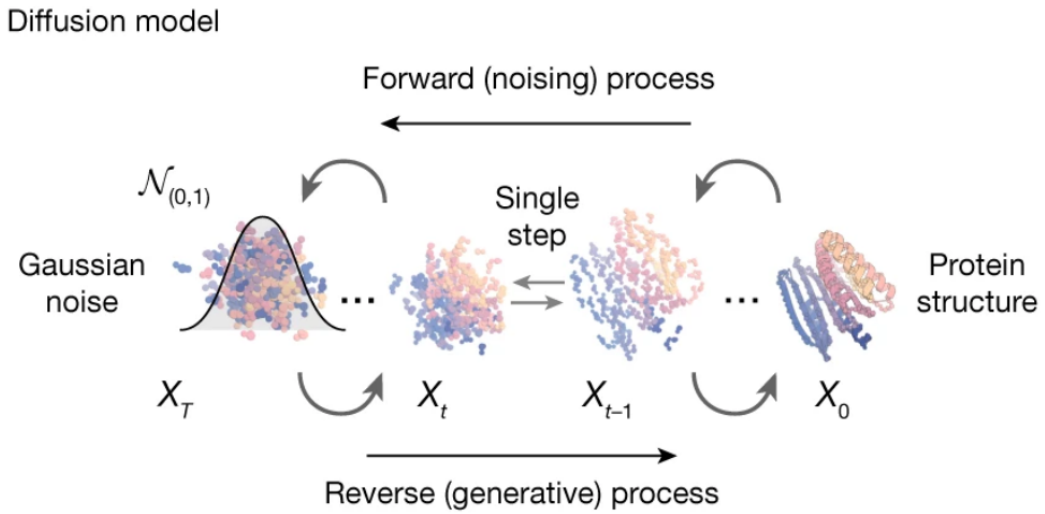

Figure 12. A single step schematic of the denoising process of diffusion models during their learning step. Modified from Watson et al., (2023)

3.8.3. PPFLOW: Target-Aware Conditional Flow Matching Model

Flow matching is another step forward from diffusion. As illustrated on Fig. 12, the denoising is done in steps and in a random manner. Flow matching keeps the trajectory approach but does that in field vector corresponding to each point on the torus manifold describing each backbone angle between 0°-360° (Lin et al., 2024; Lipman et al., 2023). In conditional flow matching, introducing protein binding pockets as limitations to this mathematical 3-dimensional space of all possible peptide backbone angles gives the possibility to assign a vector to each of the points (representing all three angles of the backbone) that shows the researchers which direction and in which magnitude the angles should be changed (Lin et al., 2024; Lipman et al., 2023). The combination of the torus describing all the backbone torsion angles with the vector field in the 3D space in the case of PPFlow model is illustrated in Fig. 13.

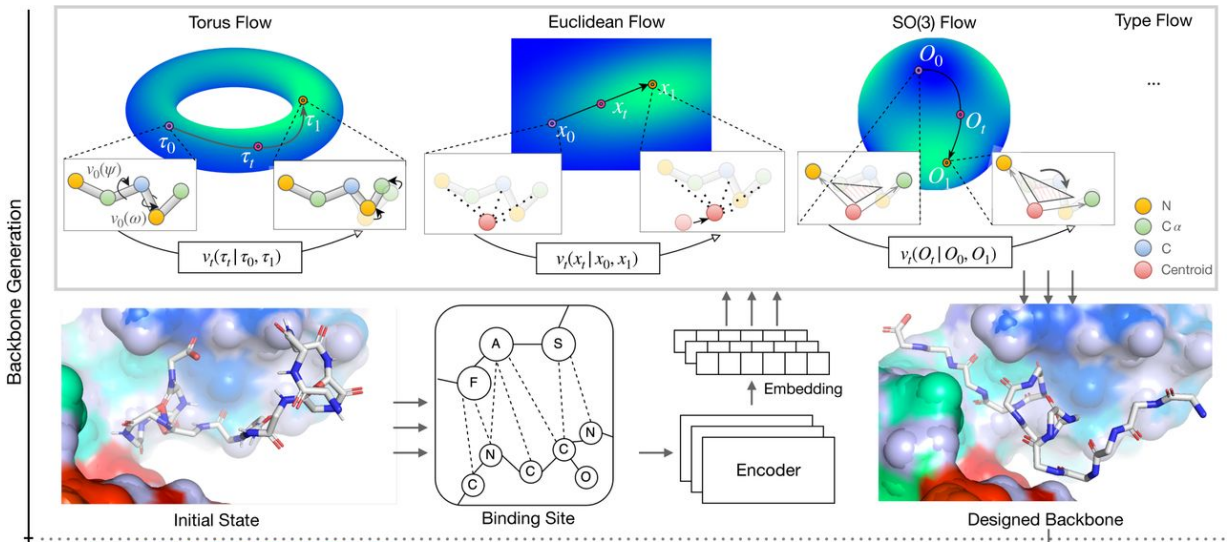

Figure 13. Workflows of PPFlow in target-aware peptide drug generation task. Illustration from Lin et al., (2024)

In conclusion, conditional flow matching approaches such as PPFlow, may offer an alternative to noisy trial-and-error clean-up models. The advantage is that they enable a guided and efficient transformation of random noise into realistic, target-shaped peptides, delivered rapidly and accompanied by an intrinsic measure of confidence.

4. Aims of the Study

For many life-threatening diseases, one of the main challenges to effective treatment is the targeted delivery of the therapeutics. A promising strategy uses systemic homing peptides that selectively bind to diseased tissues. The goal of this bachelor's thesis is to enhance the discovery of such peptides by integrating current *in vivo* phage display experiments with machine learning approaches based on *in silico* methods.

Specific aims:

1. Reimplement and improve the PPFlow conditional flow matching model in the University of Tartu High Performance Computing Centre (UTHPC) using the experimental binding data.
2. Develop an end-to-end *in silico*, *in vitro*, and *in vivo* pipeline to go from a known target receptor to validated peptide candidates for further optimization.

3. Use the pipeline to generate candidate peptides that bind Neuropilin-1 and p32 protein receptors.
4. Validate candidates in wet-lab experiments.

5. Materials and Methods

5.1. Materials

Phosphate-buffered saline (PBS) was purchased from Lonza (Verviers, Belgium). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Sigma-Aldrich (Munich, Germany). Neuropilin-1 B1 domain and p32 recombinant protein were expressed, purified, and validated in-house by Rasmus Tenno and Prakash Lingasamy.
All Oligonucleotides for phage cloning were designed using my own software (available at https://phages.streamlit.app/oligonucleotide-generator) and ordered from IDT (Integrated DNA Technologies, Belgium)

5.2. AI-Assisted *in Silico* Peptide Generation and Verification Pipeline

5.2.1. Target Data and Structure Preparation

As the general pipeline and PPFlow method itself is target-aware, protein target structures must be prepared. The human neuropilin-1 b1 domain crystal structure and human p32 protein were retrieved from the Protein Data Bank (PDB ID: 7jjc and 1p32) and loaded to UCSF ChimeraX (v1.8), a molecular visualization and analysis software, directly from this local file as specified in the configuration settings (Pettersen et al., 2021). Initial preparation involved extracting key structural data, atomic coordinates, residue identities, chain identifiers, and metadata such as crystallographic resolution, using ChimeraX's built-in tooling.

To prepare the protein structure for downstream PPFlow generation and analysis, a cleaning procedure was performed within ChimeraX. This step involved removing water molecules, free

ions, and crystallographic artifacts (ligands not relevant to this study, additional conformations). Afterwards, missing heavy atoms or incomplete side chains were repaired with PDBFixer (Eastman et al., 2017). This module reconstructed missing components by predicting their positions based on standard geometry and energy minimization principles. Alternative conformations were removed to retain a single, representative conformation that is suitable for further *in silico* steps. To ensure physiological relevance, charges of the structures were adjusted to reflect *in vivo* pH (7.4 pH) conditions.

The modified structure was saved as a PDB file which served as input for all the downstream modules.

5.2.2. Peptide Candidate Generation with PPFlow

Peptide sequences designed to target human neuropilin-1 b1 (NRP1-b1) domain and human p32 mitochondrial matrix protein tetramer (p32, also known as gC1qR) were generated using PPFlow target-aware peptide design model (Lin et al., 2024). The retraining process and modifications to PPFlow are fully documented and accessible in the associated GitHub repository (https://github.com/JasperAugust/ppflow.git). The model output candidates were exported in two formats: comma-separated values (CSV) for sequence data and PDB for structural coordinates for future analysis and validation.

Each generated peptide candidate was subjected to a rigorous validation pipeline. First, the folding free energy (ΔG) of each peptide conformation was calculated using FoldX (version 5.0, code included in github tooling section) to assess their thermodynamic stability. Peptides with invalid structural conformations or $\Delta G = 2 >$ kcal/mole conformations were discarded (Delgado et al., 2019). This processing confirmed that generated peptides were in plausible confirmations and passed both stability and structural checks before advancing to the docking stage.

5.2.3. Binding Evaluation and Docking

Stable peptide candidates were paired with the previously prepared protein structures (3.2.1) to form receptor-ligand complexes for docking analysis, which was conducted with AutoDock

Crankpep (ADCP) and AlphaFold3 (local version) (Abramson et al., 2024; Zhang & Sanner, 2019). The docking parameters were as follows:

- Docking grid center: The center of the docking grid was calculated as the arithmetic mean of the target residue coordinates within the ligand structures, using Cα atoms as positional references.
- Docking box size: The dimensions of the docking box were calculated based on the ligands maximum absolute deviation from the grid center, which was then subsequently scaled by a factor of 1.5 with an additional 10 angstroms (Å) of padding.

5.3. T7 Phage Peptide Display Construction

5.3.1. Bacteriophage Titering

Single bacteriophage clone or phage mixture titers were measured using 10x serial dilutions of the sample in LB media. Dilutions were made in the range of 10^2 to 10^{10} plaque-forming units per mL (PFU/mL). 500 µl *E. coli* BLT5615 bacterial culture (with an optical density/OD of 0.5) was pipetted to a 15 mL Falcon tube together with 100 µl of the previously completed phage dilution. Finally, 5 mL of top agar (0.7% agar solution in LB media) with Isopropyl β-D-1-thiogalactopyranoside (IPTG, final conc 2 mM) was added, and the whole mixture was plated on an LB agar plate. The plates were incubated at room temperature for 16 hours or 4 hours at 37°C. Then, visible phage plaques were counted, and the sample titer was calculated.

5.3.2. Oligonucleotide Design

The cloning was performed using complementary oligonucleotide sequences.
All Oligonucleotides for phage cloning were designed using my own software (available at https://phages.streamlit.app/oligonucleotide-generator) according to our laboratory handbook by Põšnograjeva et al (2025).

5.3.3. T7 Phage Cloning

Complementary oligonucleotides were diluted in milli-Q (mQ) water at 100 nM, heated to 95°C for 5 minutes, and cooled over 2 hours to room temperature for annealing.

The phage cloning was performed according to Novagen's protocol (*T7Select® Packaging Kit | 70014*, n.d.). High copy phages with up to 415 copies of the fused peptide displayed were cloned using the T7Select® 415-1b Cloning Kit (Millipore, 70015-3), and low copy phages were cloned with the T7Select® 10-3 Cloning Kit (Millipore, 70550-M).
As the next step, the ligation reaction was mixed on ice; details of it are brought out in Table 3. Finally, previously annealed oligonucleotides (oligonucleotides used for the experiment in Table 5 and 6) were diluted to 100 nM, and then 5 µl of the dilution was added to the ligation mixture. Incubation was done at room temperature for 1 h.

Table 3. Ligation mix.

Reagent	per 1 reaction
T7 vector 415-1b vector arms (Novagen) (undiluted 0.05 µg/µl) **OR** T7 vector 10-3 vector arms (Novagen) (undiluted 0.05 µg/µl)	1 µl
10X T4 DNA Ligase Buffer (Novagen)	1 µl
T4 DNA Ligase enzyme (Novagen)	1 µl
milli-Q water	2 µl

After ligation, 2.5 µl of packaging mix was carefully mixed with 0.5 µl of ligation reaction. This phage packaging reaction was done for 2 hours at room temperature.

The reaction mixture volume was first increased 200 µl LB media with carbenicillin and then 10x dilutions were made and plated on soft-agar plates with bacteria for easy plaque picking. For each phage clone, up to 8 plaques were picked the next day with a small pipette tip and placed into 30 µl 1 × PBS in PCR tubes. For each clone, a polymerase chain reaction (PCR) was ran to prepare for Sanger Sequencing. The mixture for the PCR was prepared as shown in Table 4 and the oligonucleotides used for each clone in Table 5 and Table 6.

Table 4. PCR reagents

Reagent	**per 1 reaction**

5 × HOT FIREPOL Blend Master Mix (Solis BioDyne, Estonia)	3 µl
Forward T7 upstream primer 5'- TAATACGACTCACTATAGGG-3' (Integrated DNA Technologies, Belgium)	0.075 µl
Reverse T7 downstream primer 5'- GCTAGTTATTGCTCAGCGG-3' (Integrated DNA Technologies, Belgium)	0.075 µl
milli-Q water	10.85 µl
phage DNA in PBS	1 µl

PCR was run with the program described below on a thermocycler (Applied Biosystems SimpliAmp Thermal Cycler) and steps 2.-4. were repeated for 25 cycles.

1. 95°C 12 min
2. 25 cycles
3. 95°C 20 sec
4. 62°C 20 sec
5. 72°C 10 sec
6. 72°C 10 min
7. 4°C Stop and Hold

Sanger sequencing was done in the University of Tartu Institute of Genetics Core Facility of Genomics.

Table 5. Candidate peptide sequences constructed by the Phage Oligonucleotide Generator against the NRP-1b1 protein. Peptide sequences themselves were produced by the *in silico* pipeline consisting of pre-trained and finetuned PPFlow, AutoDock CrankPep, FoldX, TM-Score, and AlphaFold3. Oligonucleotides are in the 5'-3' direction.

Peptide sequence	**Sense oligonucleotide**	**Antisense oligonucleotide**
EYNMGKDK	AATTCTGAATATAACATGGGCAAAGATAAATA	AGCTTTTTATCTTTGCCCATGTTATATTCAG
QRLMASVR	AATTCTCAGCGCCTGATGGCGTCTGTGCGCTA	AGCTTGCGCACAGACGCCATCAGGCGCTGAG
IPSPGVFR	AATTCTATTCCGTCTCCGGGCGTGTTTCGCTA	AGCTTGCGAAACACGCCCGGAGACGGAATAG

DGGHIRNR	AATTCTGATGGCGGCCATATTCGCAACCGCTA	AGCTTGCGGTTGCGAATATGGCCGCCATCAG
PLFIRSGR	AATTCTCCGCTGTTTATTCGCTCTGGCCGCTA	AGCTTGCGGCCAGAGCGAATAAACAGCGGAG
RRTCRQQV	AATTCTCGCCGCACCTGCCGCCAGCAGGTGTA	AGCTTCACCTGCTGGCGGCAGGTGCGGCGAG

Table 6. Candidate peptides sequences constructed by Phage Oligonucleotide Generator against p32 protein. Peptide sequences themselves were produced by the in-silico pipeline consisting of pre-trained and finetuned PPFlow, AutoDock CrankPep, FoldX, TM-Score, and AlphaFold3. Oligonucleotides are in 5'-3' direction.

Peptide Sequence	**Sense oligonucleotide**	**Antisense oligonucleotide**
STGKLLRVG	AATTCTTCTACCGGCAAACTGCTGCGCGTGGGCTA	AGCTTAGCCCACGCGCAGCAGTTTGCCGGTAGAAG
SALLCGGPG	AATTCTTCTGCGCTGCTGTGCGGCGGCCCGGGCTA	AGCTTAGCCCGGGCCGCCGCACAGCAGCGCAGAAG
SHLRSHLVL	AATTCTTCTCATCTGCGCTCTCATCTGGTGCTGTA	AGCTTACAGCACCAGATGAGAGCGCAGATGAGAAG
PLAVLILQV	AATTCTCCGCTGGCGGTGCTGATTCTGCAGGTGTA	AGCTTACACCTGCAGAATCAGCACCGCCAGCGGAG
PYMLEDLGR	AATTCTCCGTATATGCTGGAAGATCTGGGCCGCTA	AGCTTAGCGGCCCAGATCTTCCAGCATATACGGAG
SKLAGAAQL	AATTCTTCTAAACTGGCGGGCGCGGCGCAGCTGTA	AGCTTACAGCTGCGCCGCGCCCGCCAGTTTAGAAG
DGLNCQKGA	AATTCTGATGGCCTGAACTGCCAGAAAGGCGCGTA	AGCTTACGCGCCTTTCTGGCAGTTCAGGCCATCAG
AELLLLWI	AATTCTGCGGAACTGCTGCTGCTGTGGATTTA	AGCTTAAATCCACAGCAGCAGCAGTTCCGCAG
SGLWNPLLL	AATTCTTCTGGCCTGTGGAACCCGCTGCTGCTGTA	AGCTTACAGCAGCAGCGGGTTCCACAGGCCAGAAG
SDSGFYIYI	AATTCTTCTGATTCTGGCTTTTATATTTATATTTA	AGCTTAAATATAAATATAAAAGCCAGAATCAGAAG
GAATEWAKR	AATTCTGGCGCGGCGACCGAATGGGCGAAACGCTA	AGCTTAGCGTTTCGCCCATTCGGTCGCCGCGCCAG
CDGLNGQKGAC	AATTCTTGCGATGGCCTGAACGGCCAGAAAGGCGCGTGCTA	AGCTTAGCACGCGCCTTTCTGGCCGTTCAGGCCATCGCAAG
CSKLAGAAQLC	AATTCTTGCTCTAAACTGGCGGGCGCGGCGCAGCTGTGCTA	AGCTTAGCACAGCTGCGCCGCGCCCGCCAGTTTAGAGCAAG
CDPRKTSVRLC	AATTCTTGCGATCCGCGCAAAACCTCTGTGCGCCTGTGCTA	AGCTTAGCACAGGCGCACAGAGGTTTTGCGCGGATCGCAAG

5.4. Phage Amplification and Purification

Single-peptide amplification and purification were done based on our University of Tartu Laboratory of Precision and Nanomedicine handbook (Põšnograjeva et al., 2025).

E. coli BLT5403 strain (Novagen, USA) was cultured on ampicillin (50 µg/ml) infused LB agar plate and incubated overnight at 37°C. From the lab, a single colony was picked to inoculate 10 ml of liquid LB media with carbenicillin (100 µg/ml, LB-carb), and the resulting bacterial culture was grown in a 100 ml Erlenmeyer flask in a bacterial incubator at 37°C and shaking at 220 rotations/min (RPM). The next morning, fresh LB-carb media was inoculated with the overnight bacterial broth with a dilution ratio of 1:100. This diluted inoculum was incubated until an optical density (OD) OD_{600}= 0.5 was reached, measured with a spectrophotometer (BioPhotometer Plus, Eppendorf, USA).

For bacteriophage amplification, the liquid bacterial broth was infected with a single phage with a final multiplicity of infection (MOI) of 0.01 and a final volume of 26 ml in a 50 ml Falcon tube and incubated at 37°C with 220 RPM shaking until the mixture cleared and bacterial lysate was visible. This lysate was cooled on ice for 30 minutes. Afterwards, 3 ml of 5M NaCl was added and the mixture vortexed. Lysate was then cleaned of bacterial debris by centrifuging it at 12,000 × g at 4°C for 10 minutes (Eppendorf 5810R, rotor F-34-6-38). Supernatant was poured into a new 50 ml Falcon tube and combined with 9 ml of 50% polyethylene glycol in PBS (PEG8000), vortexed, and left on ice for at least 30 min. Phages were precipitated by centrifugation at 8,000 × g at 4°C for 10 minutes. Supernatant was carefully discarded, and tubes were left to dry upside-down on paper towels for 15 min. Then, the precipitated phage was resuspended in 1.5 ml PBS.

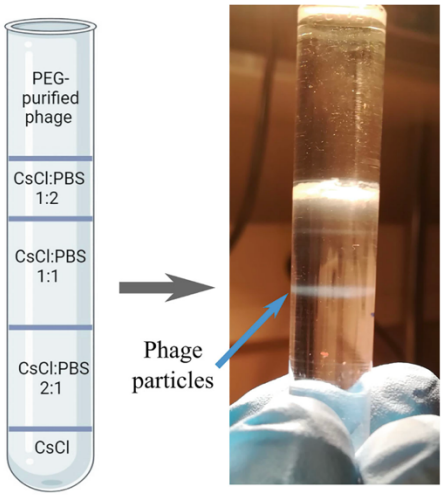

Figure 14. CsCl density gradient in the ultracentrifugation tube. After ultracentrifugation, phage particles are visible between the third and second layer and can be isolated through the tube wall with a 21G syringe. Fig. is from Põšnograjeva et al. (2025)

Final phage purification was performed using a CsCl density gradient ultracentrifugation. A 62.5% filtered CsCl solution was utilized to create CsCl:PBS mixtures in the ratios of 2:1, 1:1, and 1:2, which were then pipetted into an ultracentrifugation tube. Layers of CsCl:PBS were carefully pipetted on top of one another in this sequence:

1. CsCl (62.5%) 0.25 mL,
2. CsCl:PBS 2:1 1.2 mL
3. CsCl:PBS 1:1 1.2 mL
4. CsCl:PBS 1:2 0.25 mL

The previously precipitated phage sample was then pipetted on top of the gradient and the gradient, and the phage particle line is visible in Fig. 14. Tubes were ultracentrifuged for 45 minutes at 40,000 RPM with a Beckman Coulter SW60TI swing-rotor at room temperature, and the phages were collected with a syringe. Purified phages were dialyzed in Thermo Fisher Scientific (USA) dialysis cassettes against room temperature PBS for 1 h, PBS replaced, and then dialysis repeated.

5.5. *In vitro* cell-free binding assays

Candidate peptide binding capabilities were tested in Costar 96-Well ELISA plate (#3590, Corning Life Sciences, Tewksbury, MA, USA) against nickel-NTA (Ni-NTA) Magnetic Agarose Beads (QIAGEN, Hilden, Germany) coated with hexahistidine (His-6X) tagged NRP-1b1 and P32 (15 µg of protein per 5 µl of beads) at room temperature for 1h in 200ul BWB3 (50 mM Tris, 0.05% Igepal CA-630, 5 mM Imidazole, pH 7.4, 1M NaCl 1000; Thermo Scientific Inc.). The functionalized beads were washed 3 times with 1 ml BWB3-BSA washing buffer, which is BWB3 wash buffer with additional 0.1% bovine serum albumin (BSA, GE Healthcare, Little Chalfont, UK), incubated with 5×10^8 PFU of phages in 100 µl of phosphate-buffered saline (PBS) cloned, amplified, and purified as mentioned in section 3.2 above. After 1h of room temperature co-incubation, unbound phages were removed by rinsing 6 times with 200 µl BWB3-BSA washing buffer. During each wash, samples were transferred to the next row of wells with a 12-channel pipette and pipette tips were changed to avoid contamination. After the 6th wash, the bound phages were eluted with 200µl of PBS containing 400 mM Imidazole and 0.1% Igepal CA-630 (Thermo Scientific Inc.). RPARPAR phage binding to NRP-1b1-coated beads and Lintt-1 phage

binding to P32-coated beads were used as a positive control. In both cases, phage clones displaying heptaglycine peptide (GGGGGGG, G7) were used as a negative control.

The eluted phages were titered according to section 3.2.1 above. After each experiment, 4 of each unique phage clones were subjected to Sanger sequencing of peptide-encoding phage DNA (Teesalu et al., 2012). AB1 files received from sequencing were analyzed using T7 Phage peptide Sequence Analyzer.

5.6. *In vivo* phage play-off experiment

The thesis project *in vivo* experiments were done according to Animal Experiment License No. 159, and experiments were conducted in the University of Tartu Institute of Biomedicine and Translational Medicine Laboratory Animal Centre. During the project, 6 female Balb/C albino mice aged 8–10 weeks were used. Animals were housed in a 12-hour light/dark cycle environment with free access to food and water.

As a preparation for the experiment, phages were cloned, amplified, purified, precipitated, and dialyzed in PBS based on section 3.2 above and then combined into a single solution in equimolar concentration.

For peptide homing, after 2 min of tail warming in warm water, mice were injected into the tail vein with 100 µl purified phage mix (5×10^9 PFU, 5×10^{10} PFU/ml) to the tail vein was done and left to circulate for 60 minutes. Mice were anaesthetized with an intraperitoneal (IP) injection of 400 µl dexmedetomidine (0.1 mg/kg) and ketamine (75 mg/kg) and then fixated to a polystyrene holder. The abdominal cavity and diaphragm were cut open, and a small cut was made into the right atrium. Mice were then perfused intracardially through the left ventricle with a perfusion needle with PBS until all major organs were discolored. Lung, heart, and liver were isolated, separately weighed, and placed into 1 ml LB + 1% Igepal CA-630 solution (Sigma-Aldrich Co, USA). Organs were homogenized mechanically with a tissue homogenizer (OMNI International, USA) and titered as previously mentioned in section 3.2.1. After titering, rest of the tissue homogenate was plate amplified according to Põšnograjeva et al. (2025) and sent to be sequenced using Ion Torrent Next Generation Sequencing (Section Ion Torrent Sequencing).

5.7. Ion Torrent Sequencing

Biodistribution of *in vivo* play-off peptide-phages was sequenced using high-throughput sequencing, Ion Torrent (HTS; Thermo Fisher Scientific Inc., USA). All tissue samples, as well as the injected phage mixture, were sequenced for later analysis.
Sample preparation and Ion Torrent sequencing were done by the University of Tartu Nanomedicine lab according to the protocol described in Pleiko et al. (2021), section "Determination of phage displayed peptide sequences".

5.8. Homing analysis

Homing analysis was done using the code included in https://github.com/JasperAugust/ppflow.git under NGS based on methods shown in Põšnograjeva et al., (2025). After file preprocessing, the graphs were generated with GraphPad Prism Software, Boston, Massachusetts USA, www.graphpad.com.

6. Results and Discussion

This bachelor's thesis aims to enhance the discovery of targeting peptides by combining in vivo phage display studies with machine learning-driven *in silico* methods. To develop the discovery pipeline, we initially focused on identifying new peptides targeting well-characterized homing peptide receptors. Binding peptides engaging with NRP-1 (PDB: 7JJC) and P32 (PDB: 1P32) were generated using a retrained version of PPFlow (conditional flow matching-based machine learning model) that samples peptides directly from a target-binding pocket constrained torsion-angle space.

Each peptide went through an *in silico* evaluation and docking pipeline. The structure of candidate peptides (9.1 and 9.2) was evaluated, and unstable or poorly folded sequences were filtered out with FoldX. Candidates were ranked based on their ADCP docking energy (ΔG), isoelectric point, charge, and hydrophobicity (GRAVY index). Top candidate peptides were cloned into T7 bacteriophage vectors, allowing either high copy (415-1b with up to 415 peptide

copies/phage) or low copy (10-3b) peptide display to allow evaluation of the peptide valency on the target engagement. Both high- and low-copy peptide-phages were tested using Ni-NTA magnetic bead binding assays (detailed in the section 5.5). For the NRP-1 peptide candidates, *in vivo* homing experiments were also performed.

6.1. NRP-1 and P32 Protein Targeting Peptide Generation

Peptides generated by the *in silico* pipeline, docked and evaluated as described in section 1.11, are listed in Appendices 6.1 and 6.2. A total of 42 candidate peptides for NRP-1b1 and 63 peptide sequences for P32 were docked with ADCP (two docking results are shown in Fig. 15). Positive control peptides, RPAR for NRP-1 and Lintt1 for P32, were chosen as previously well-known and characterized homing peptides (Haspel et al., 2011; Paasonen et al., 2016; Sharma et al., 2017). For phage cloning and binding experiments, sequences were selected based on diversity of sequence motifs, low docking energy, no more than three identical residues in tandem, and compatibility with the T7 display system (Table 7 and Table 8). In addition, sequences with higher ADCP docking clusters (e.g., QRLMASVR with a cluster size of 52) were preferred as the increased cluster size points towards a stable binding position (Bender et al., 2021).

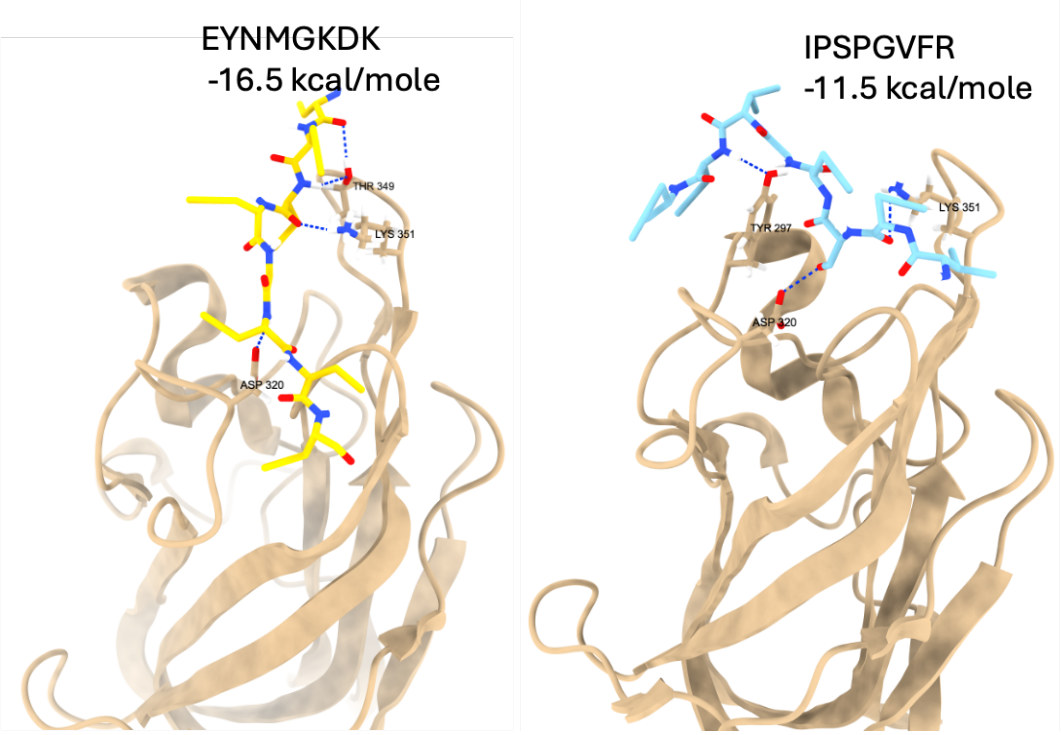

Figure 15. Structural models of two candidate NRP-1-binding peptides, EYNMGKDK and IPSPGVFR, in complex with the NRP-1 b1 domain. Peptides were generated with the PPFlow pipeline and docked with AutoDock Crankpep. Binding affinity is calculated by Autodock's algorithm and is displayed below the peptide sequence. Protein is in beige, and blue dotted lines indicate hydrogen bonds.

Table 7. Candidate NRP-1b1 targeting peptides. CendR peptides chosen for further in vitro and in vivo experiments. High copy number indicates phages with up to 415 copies displayed on the phage surface and low copy number indicates phage particles with 5-15 peptides exposed on the surface.

Peptide ID	Amino-acid sequence	Copy number
RPAR	RPARPAR	High
RPAR	RPARPAR	Low
15	QRLMASVR	High
15	QRLMASVR	Low
42	DGGHIRNR	High
42	DGGHIRNR	Low
2	IPSPGVFR	High
2	IPSPGVFR	Low
27	PLFIRSGR	High
2	PLFIRSGR	Low

For the P32 receptor, three peptides—DPRKTSVKL, DGLNCQKGA, and SKLAGAAQL—were engineered with N- and C-terminal cysteines to enable cyclization and expressed on T7 phages at different valencies, as described above. Peptide cyclization can enhance stability, improve binding affinity by constraining the conformation, and increase resistance to proteolytic degradation. Our past studies showed that both linTT1 (AKRGARSTA) and its cyclic version (CKRGARSTC) bind to p32 receptors.

Table 8. Candidate p32 targeting peptides. Peptides with two flanking cysteines are marked with "B" to distinguish them from the originally generated sequences. As the *71B sequence failed to clone into the phage genome and consistently had a single-point mutation of K->R at the 9th amino acid, the mutated version was used in experiments. High copy number indicates phages with up to 415 copies displayed on the phage surface and low copy number indicates phage particles with 5-15 peptides exposed on the surface.

Peptide ID	Amino-acid sequence	Copy number
linTT1	AKRGARSTA	High
32	CSKLAGAAQLC	High
32	CSKLAGAAQLC	Low
83B	CDGLNGQKAC	High
83B	CDGLNGQKAC	Low
83	DGLNCQKGA	High
83	DGLNCQKGA	Low
72	SALLCGPGC	High
72	SALLCGPGC	Low
71B*	CDPRKTSVRLC	High
71B*	CDPRKTSVRLC	Low

6.2. *In Vitro* Binding on NRP-1 b1 Domain-Functionalized Ni-NTA beads

Multiple experiments assessed binding of high and low valency T7 phages displaying candidate peptides to recombinant NRP-1 protein b1 domain (Fig. 16). Enrichments is reported as fold increase over the heptaglycine (G7) displaying phage on a log10 scale. The established NRP-1b1 ligand RPARPAR (positive control, RPAR) served as the benchmark in a one-way ANOVA followed by Holm-Šidak's multiple comparison test ($\alpha = 0.05$, $n = 3$), and none of the novel peptides differed significantly from the positive control. "Not significant" does not imply equivalence. With only three replicates, the test has limited power, and the possibility that true differences exist should not be excluded.

All candidates showed good binding only in their high-copy-number versions. A similar lower performance was evident with the positive control low-copy phage, binding threefold lower than its high-copy version (312 vs 100-fold over G7). This high avidity caused by multivalency is

a recognized effect (Ruoslahti, 2012) and balances the homing peptides' otherwise relatively low affinity compared to other affinity ligands (Lei et al., 2022; Xian et al., 2025). As a consequence, peptides are particularly well suited for multivalent targeting of nanoparticles.

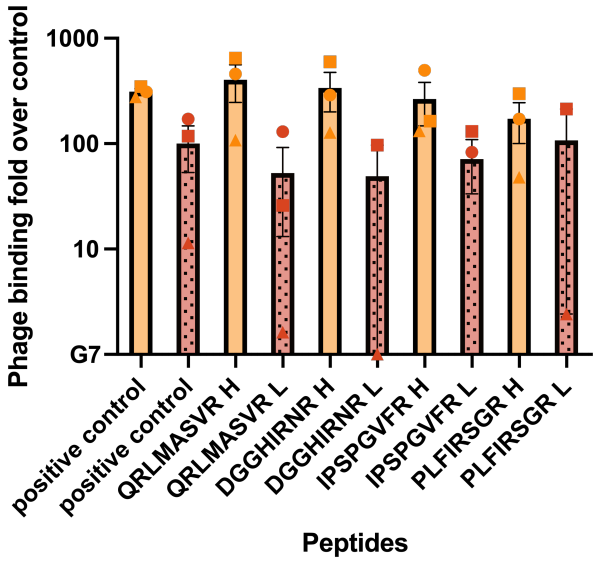

Figure 16. Cell-free binding of peptide-displaying T7 phages to the recombinant NRP-1 b1 domain. Recombinant human NRP-1 b1 domain was immobilized on Ni-NTA agarose beads and separately incubated with each high-copy peptide (solid bars) and low-copy peptide (dotted bars) phage clones. Bars show mean ± SEM. Bound phage titers are expressed as fold-enrichment over the untargeted G7 control phage, baseline 1, $\log_{10}$ scale; Data were analyzed by one-way ANOVA followed by Holm-Šidák's multiple-comparison test ($\alpha = 0.05$).

The *in vitro* experiment can be strengthened in several ways. First, increasing both the number and diversity of replicates —ideally contributed by multiple researchers—would improve reproducibility (Holman et al., 2015). Additionally, incorporating blind protocols into the peptide generation pipeline would help reduce experimenter bias. For example, initial phage cloning and sample coding could be performed by an independent researcher, or peptide sequences could be masked as labels during analysis.

6.3. *In Vivo* Phage Homing Play-Off with NRP-1 Peptides

In vitro binding assays, while informative, do not account for complex biological conditions such as protein conformation differences on cells, biodistribution, proteases, or competing serum proteins (NDong et al., 2015). To assess *in vivo* homing ability, an equimolar mixture of peptide-

displaying phage clones was intravenously injected into mice and circulated for 1 hour. Subsequently, mice were perfused with PBS to remove non-binding phages, and organs (lung, liver, heart) were collected, homogenized, and phage titers from each tissue were determined.

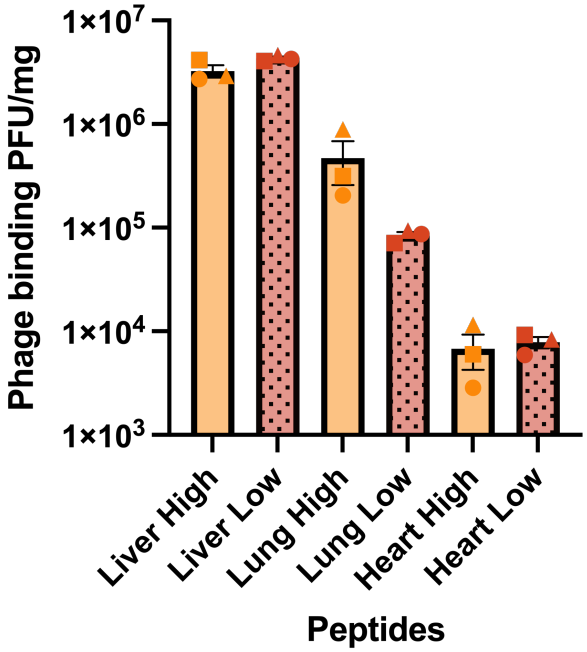

Figure 17. *In vivo* biodistribution of the whole peptide-phage pool after 1 hour of circulation. Balb/c mice (n = 3 per group) were injected intravenously with a total of 5 x 10^9 plaque-forming units (PFU), consisting of an equimolar mixture of each peptide-displaying phage in either high-copy (solid color bars) or low-copy (dotted bars) version. The animals were euthanized, perfused with PBS, and the liver, lung, and heart were isolated and homogenized. Recovered phages from each tissue were amplified and sequenced. $\log_{10}$ scale; Bars denote mean ± SEM; individual symbols show biological replicates.

Upon systemic administration, RPARPAR peptide-displaying phages are known to exhibit robust lung accumulation in a neuropilin-1 (NRP-1)-dependent manner (Teesalu et al. 2009). *In vivo* tissue titers of NRP-1-targeting peptide phages revealed selective accumulation in lungs (Fig. 17). Phage accumulation in pulmonary tissue was valency-dependent: we observed the highest recovered phage titer (~4.7 x 10^5) for high-copy clones, compared to low-copy phage titers that were ~5.6 times lower (~8.4 x 10^4). Liver tissue, the blood's main filtration organ, showed the highest recovered phage titers with only a small difference between the high-copy clones (~3.3 x 10^6) and low-copy clones (~4.3 x 10^6). In contrast, heart tissue showed the lowest phage

recovery for both formats (~6.7 x 10^3 and ~7.8 x 10^3). These data are in line with previously described lung specificity of NRP-1 targeting peptides (Teesalu et al., 2009).

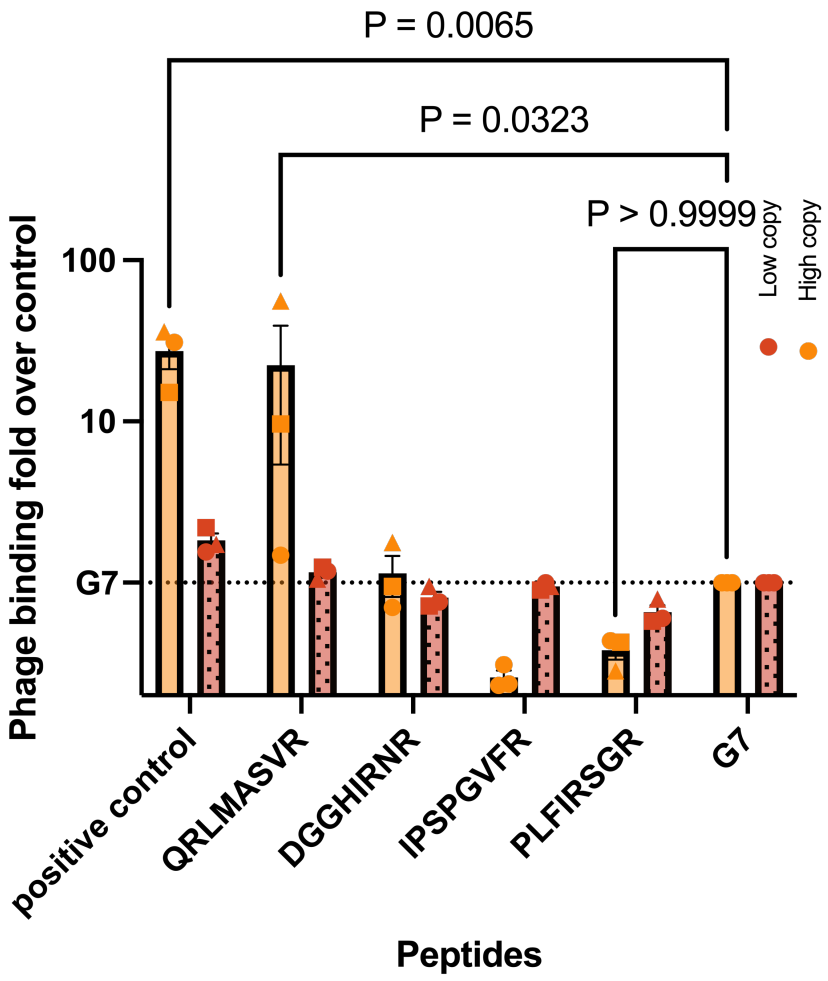

Figure 18. *In vivo* lung homing of individual candidate peptide-phages after 1 hour circulation. Balb/c mice (n = 3 per group) were injected *i.v.* with a total of 5 x 10^9 plaque-forming units (PFU), consisting of an equimolar mixture of each peptide-displaying phage in either high-copy (solid color bars) or low-copy (dotted bars) version. The animals were euthanized, perfused with PBS, and the liver, lung, and heart were homogenized, and phage titers in homogenates were determined. Titers are expressed as fold change over the untargeted G7 control (dotted line set to 1; $\log_{10}$ scale). Bars denote mean ± SEM; individual symbols show biological replicates. Bars marked with a p-value are significantly above the negative-control G7 (one-way ANOVA, Dunnett post-hoc). Only positive control reached significance ($P < 0.0001$); all other comparisons were non-significant ($P > 0.05$).

Individual peptide data recovered from the lung graph (Fig. 18) revealed two peptides, our positive control, and QRLMASVR, one of our candidate peptides, with the highest enrichment difference between high- and low copy phage versions and the highest lung-liver ratio of around

34. Expressing the GRPARPAR and QRLMASVR peptides at high copy resulted in 43.4-fold and 22.3-fold enrichment above the untargeted heptaglycine (G7) control peptide, while the same sequences at low copy performed significantly worse, achieving 5.7-fold and 1.05-fold enrichment over baseline, highlighting the importance of not just the CendR motif but also the peptide copy number (Haspel et al., 2011; Teesalu et al., 2009). At the same time, data from heart homing indicate that peptide-phage particle accumulation in lung tissue is not due to promiscuous binding (Fig. 19 and 20). Interestingly, compared to other peptides, IPSPGVFR and PLFIRSGR peptide-phages show an opposite valency-dependent effect, as they reach tested tissues in a similar manner to the control peptide as low-copy phages, but the high-copy phages show almost no binding at all. This can point to valency-dependent quick *in vivo* degradation, phage frameshift mutations, or that these single peptide mixtures were degraded before the experiment, requiring further study for an answer (Livesey & and Nicholls, 2012; Mirzaei et al., 2020; Pang et al., 2014; Werle & Bernkop-Schnürch, 2006).

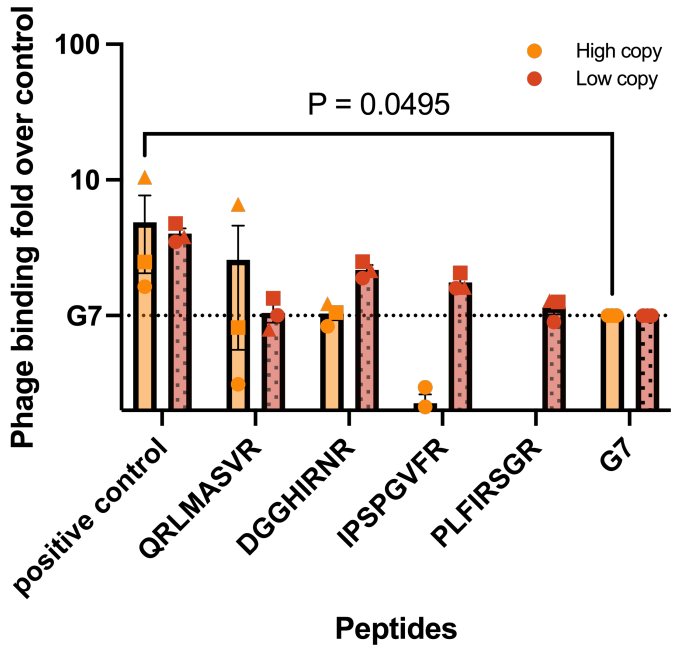

Figure 19. In vivo heart homing of individual candidate peptide-phages after 1 hour circulation. Balb/c mice (n = 3 per group) were injected i.v. with a total of 5 x 10^9 plaque-forming units (PFU), consisting of an equimolar mixture of each peptide-displaying phage in either high-copy (solid color bars) or low-copy (dotted bars) version. The animals were euthanized, perfused with PBS, and the liver, lung, and heart were homogenized, and phage titers in homogenates were determined. Titers are expressed as fold change over the untargeted G7 control (dotted line set to 1; $\log_{10}$ scale). Bars denote mean ± SEM; individual

symbols show biological replicates. Bar marked with a p-value is significantly above the negative-control G7 (one-way ANOVA, Dunnett post-hoc). Only high-valency GRPARPAR reached significance (P = 0.0495); all other comparisons were non-significant (P > 0.05).

In contrast to the lung, in the cardiac tissue, high- and low-copy phages show less than 4-fold differences in titer. These data support that the observed lung selectivity of GRPARPAR (positive control) and QRLMASVR at high copy is genuine and not just increased systemic stickiness (Balmert et al., 2015; Werle & Bernkop-Schnürch, 2006).

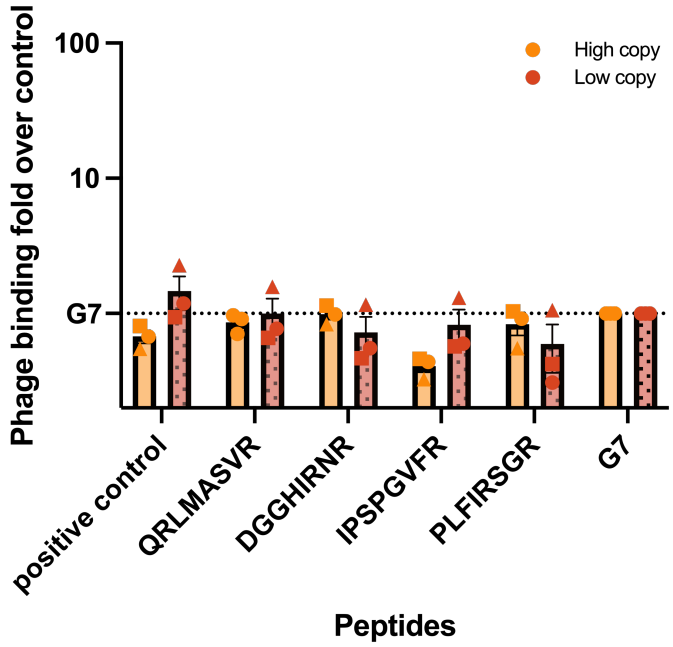

Figure 20. In vivo liver homing of individual candidate peptide-phages after 1 hour circulation. Balb/c mice (n = 3 per group) were injected i.v. with a total of 5 x 10^9 plaque-forming units (PFU), consisting of an equimolar mixture of each peptide-displaying phage in either high-copy (solid color bars) or low-copy (dotted bars) version. The animals were euthanized, perfused with PBS, and the liver, lung, and heart were homogenized, and phage titers in homogenates were determined. Titers are expressed as fold change over the untargeted G7 control (dotted line set to 1; $\log_{10}$ scale). Bar marked with a p-value is significantly above the negative-control G7 (one-way ANOVA, Dunnett post-hoc). Only low-valency GRPARPAR reached significance (P < 0.0001); all other comparisons were non-significant (P > 0.05).

All peptide-phages were recovered from the liver at less than 5-fold above baseline with no significant valency effects on clearance.

During the preclinical search for targeting molecules (peptides or otherwise), a key goal is not just high homing to the target tissue/disease, but the ability of the targeting technology to minimize uptake in nontarget tissues (Hoppenz et al., 2020; Lingasamy & Teesalu, 2021). When

comparing lung over liver homing ratios, QRLMASVR came out as the candidate with the highest selectivity (lung-liver ratio of 36.7, compared to the positive control's 40) as it has similar liver and heart homing to baseline while still greatly enriching the lung tissue.

In summary, displaying the high copy number of peptides on phage particles significantly increases their lung homing ability by one or two orders of magnitude, without increasing hepatic or cardiac tissue background.

6.4. *In Vitro* Binding on P32 Functionalized Ni-NTA Beads

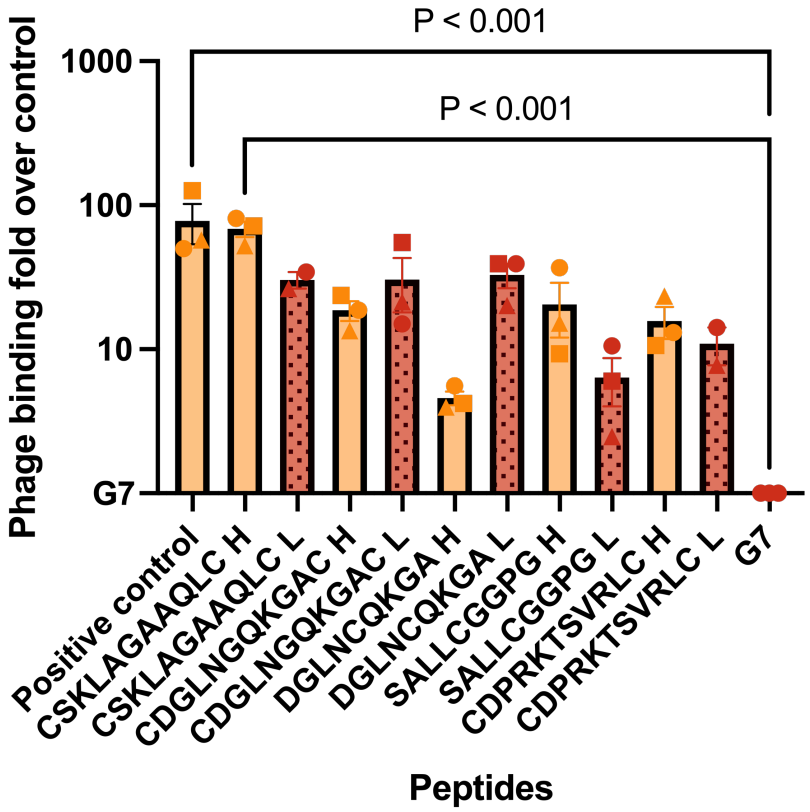

Figure 21. Cell-free binding of candidate peptide phages to recombinant p32. Human P32 was immobilized on Ni-NTA agarose beads and incubated with high-copy peptide (solid bars) and low-copy peptide (dotted bars) phage clones. Lintt1 (AKRGARSTA) peptide phage serves as a positive control. Bars show mean ± SEM. Recovered phage data are expressed as fold enrichment over the untargeted G7 control phage, baseline 1, $\log_{10}$ scale. Bars marked with a p-value are significantly above the negative-control G7 (one-way ANOVA, Dunnett post-hoc).

From the candidates, only the CSKLAGAAQLC high copy variant reached significance ($P < 0.001$); all other comparisons were non-significant ($P > 0.05$).

The same Ni-NTA pull-down strategy used for NRP-1 (section 2.2) was applied to assess *in vitro* binding of six tumor-homing peptides for recombinant p32 (Fig. 21). Every clone shown, regardless of valency, exceeded the control peptide-phage (G7) background by an order of magnitude. However, only high-valency CSKLAGAAQLC was significantly above the negative control. It's been generally shown (Ruoslahti et al., 2010) that nanoparticles, but especially nanoparticles functionalized by peptides, benefit from enhanced avidity brought on by multivalent binding (Ruoslahti, 2012; Werle & Bernkop-Schnürch, 2006).

Comparing this cell-free phage binding experiment to the neuropilin-1 peptide candidate experiment, a magnitude difference can be seen in phage binding fold over control between neuropilin-1 and p32 targeting peptides. This binding difference could be explained by the NRP-1 receptors' inherently high electronegativity that allows the C-terminal arginine to bind with high affinity (Haspel et al., 2011; Tymecka et al., 2024; Zanuy et al., 2013). At the same time, p32 is very acidic (pI of 4) with an asymmetric charge distribution, only having negative charge on one side of the protein and inside the doughnut-shaped structure itself, severely constraining the available space and conformation of the binding ligand (Balmert et al., 2015; Gotoh et al., 2018; J. Jiang,Y. Zhang,A.R. Krainer,& R. Xu, n.d.; Yenugonda et al., 2017). Timur et al. have shown that indeed, the p32 binding pocket is inside with restricted access (Timur et al., 2018). Further investigation to clarify the involvement of hindered access to the binding pocket is required.

7. Conclusion

This thesis set out to develop an *in silico* pipeline, which includes a probabilistic generative machine learning model with a validation step combined with a classical phage-display workflow, to accelerate the discovery of tissue-homing peptides for known target receptors. By re-implementing and fine-tuning PPFlow on the University of Tartu HPC cluster, we generated tens of novel NRP-1 and p32 binder candidates in torsion-angle space, filtered them, and shortlisted them for wet-lab experimental work. This integrated approach demonstrated the

potential of combining computational modeling with experimental workflows to streamline and enhance the discovery of receptor-specific homing peptides.

Building on this computational foundation, we validated candidate peptides using high- and low-valency T7 phage display to assess their binding performance under cell-free conditions. High-copy (<415 peptides/phage) display consistently outperformed low-copy (5-15 peptides/phage) for NRP-1, and every candidate bound at least an order of magnitude better than the G7 (heptaglycine) control, only with the high-copy display version. High-copy QRLMASVR was the top de novo hit (404-fold over G7, $p=0.022$) and matched the prototypic CendR-motif peptide GRPARPAR. For p32, multivalency again drove the binding assay, as only the high-copy version of the CSKLAGAAQLC peptide was significantly above the negative control ($p<0.001$), whereas its low-copy version and the non-cyclic peptides performed modestly. All of this underscored the clear avidity requirement in studies on purified target proteins.

To further confirm these findings in a physiological context, we proceeded with in vivo experiments to evaluate the targeting efficiency of the peptide-displaying phages. During the *in vivo* experiments, equimolar pools of the NRP-1 phages were injected into mice for a 60-minute circulation. High-copy phages accumulated preferentially in the lung, 5.6-fold above their low-copy variants and two orders of magnitude over heart tissue. When titers were normalized to G7, QRLMASVR almost matched the positive control with a lung/liver enrichment ratio of 36.7 vs 40 while remaining at baseline in control tissues. Therefore, PPFlow not only learned the known CendR pharmacophore but also proposed a novel sequence that reproduces its biodistribution.

Despite these promising results, several limitations were identified that highlight areas for improvement in future studies. Replicates were limited, lacked blinded experiments, and p32 candidates were not backed by *in vivo* validation. Future work will address these issues, expand the general cohort size for all steps in the pipeline, and further validate valency effects. Additionally, integrating AlphaFold 3-level docking replacement could replace the current slow brute-force docking approach, enabling more efficient computational screening earlier in the process.

However, these limitations aside, the key findings underscore the substantial progress made by this integrated approach. The discovery of a novel lung-selective ligand, QRLMASVR, along with a family of p32 binders, demonstrates that machine learning guided phage display can already

approach the classical directed evolution performance. By combining *in silico* generation, bench assays, and biodistribution studies within a single feedback loop, this work advances targeting nanomedicine toward programmable, receptor-aware delivery systems, enabling rapid improvement of vascular ZIP code peptides against known receptors.

Tõenäosuslike masinõppemeetodite rakendamine kullerpeptiidide arendamiseks

Jasper August Tootsi

Resümee

Vähi ja teiste kompleksete haiguste ravi probleemiks on olemasolevate ravimite suured kõrvalmõjud ning madal efektiivsus. Üheks võimaluseks sellise ravi terapeutilise efektiivsuse suurendamiseks on nende kombineerimine koespetsiifiliste kullerpeptiididega. Käesoleva bakalaureustöö eesmärgiks oli retseptor-spetsiifiliste kullerpeptiide genereeriva hübriidse masinõppemudelil põhineva arendusprotsessi väljatöötamine ja valideerimine. Töö käigus genereeriti tuntud vähi sihtmärkide NRP-1 ja p32 seonduvaid kandidaatpeptiide. Kandidaatide selektiivsust ja biodistributsiooni hinnati kõrge ja madala koopiaarvuga faagidispleiga nii rakuvabas süsteemis kui ka *in vivo* katsetega hiirtemudelitel. Katsete tulemused näitavad, et multivalentsus mängib olulist rolli koe-spetsiifiliste peptiidide seondumisel. Neuropiliin-1-ga seonduvate kandidaatide seast osutus parimaks kõrge koopiaarvuga ekspresseeritud peptiid, QRLMASVR, mille rakuvaba seondumiskatse ja intravenooselt süstitud võistluslikul faagidispleil põhineva eksperimendi biodistributsioon sarnanesid varem põhjalikult kirjeldatud prototüüpilise CendR motiiviga kullerpeptiidi RPAR-ga (RPARPAR). Kokkuvõtvalt näitlikustab see töö masinõppemudelite suurt potentsiaali retseptor-spetsiifiliste peptiidide avastamisel ja viib täppis-nanomeditsiini ühe sammu võrra lähemale kliiniliselt relevantsete peptiidide laiapõhjalisele kaardistamisele.

8. References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

All Nobel Prizes 2018. (n.d.). NobelPrize.Org. Retrieved May 7, 2025, from https://www.nobelprize.org/all-nobel-prizes-2018/

Anchordoquy, T., Artzi, N., Balyasnikova, I. V., Barenholz, Y., La-Beck, N. M., Brenner, J. S., Chan, W. C. W., Decuzzi, P., Exner, A. A., Gabizon, A., Godin, B., Lai, S. K., Lammers, T., Mitchell, M. J., Moghimi, S. M., Muzykantov, V. R., Peer, D., Nguyen, J., Popovtzer, R., … Simberg, D. (2024). Mechanisms and Barriers in Nanomedicine: Progress in the Field and Future Directions. *ACS Nano*, *18*(22), 13983–13999. https://doi.org/10.1021/acsnano.4c00182

Arap, W., Kolonin, M. G., Trepel, M., Lahdenranta, J., Cardó-Vila, M., Giordano, R. J., Mintz, P. J., Ardelt, P. U., Yao, V. J., Vidal, C. I., Chen, L., Flamm, A., Valtanen, H., Weavind, L. M., Hicks, M. E., Pollock, R. E., Botz, G. H., Bucana, C. D., Koivunen, E., … Pasqualini, R. (2002). Steps toward mapping the human vasculature by phage display. *Nature Medicine*, *8*(2), 121–127. https://doi.org/10.1038/nm0202-121

Arap, W., Pasqualini, R., & Ruoslahti, E. (1998). Cancer Treatment by Targeted Drug Delivery to Tumor Vasculature in a Mouse Model. *Science*, *279*(5349), 377–380. https://doi.org/10.1126/science.279.5349.377

Balmert, S. C., Zmolek, A. C., Glowacki, A. J., Knab, T. D., Rothstein, S. N., Wokpetah, J. M., Fedorchak, M. V., & Little, S. R. (2015). Positive Charge of "Sticky" Peptides and Proteins Impedes Release From Negatively Charged PLGA Matrices. *Journal of Materials Chemistry. B, Materials for Biology and Medicine*, *3*(23), 4723–4734. https://doi.org/10.1039/C5TB00515A

Basith, S., Manavalan, B., Hwan Shin, T., & Lee, G. (2020). Machine intelligence in peptide therapeutics: A next-generation tool for rapid disease screening. *Medicinal Research Reviews*, *40*(4), 1276–1314. https://doi.org/10.1002/med.21658

Bell, D. R., & Chen, S. H. (2021). Toward Guided Mutagenesis: Gaussian Process Regression Predicts MHC Class II Antigen Mutant Binding. *Journal of Chemical Information and Modeling*, *61*(10), 4857–4867. https://doi.org/10.1021/acs.jcim.1c00458

Bender, B. J., Gahbauer, S., Luttens, A., Lyu, J., Webb, C. M., Stein, R. M., Fink, E. A., Balius, T. E., Carlsson, J., Irwin, J. J., & Shoichet, B. K. (2021). A practical guide to large-scale docking. *Nature Protocols*, *16*(10), 4799–4832. https://doi.org/10.1038/s41596-021-00597-z

Ben-Hur, A., & Noble, W. S. (2005). Kernel methods for predicting protein–protein interactions. *Bioinformatics*, *21*(suppl_1), i38–i46. https://doi.org/10.1093/bioinformatics/bti1016

Blanco, E., Shen, H., & Ferrari, M. (2015). Principles of nanoparticle design for overcoming biological barriers to drug delivery. *Nature Biotechnology*, *33*(9), 941–951. https://doi.org/10.1038/nbt.3330

Bodanszky, M. (1988). Amino Acid Analysis. In M. Bodanszky (Ed.), *Peptide Chemistry: A Practical Textbook* (pp. 13–15). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-97886-9_2

Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., van der Meer, F., Kallio, K., Kaya, T., Anastasina, M., Smura, T., Levanov, L., Szirovicza, L., Tobi, A., Kallio-Kokko, H., Österlund, P., Joensuu, M., Meunier, F. A., Butcher, S. J., … Simons, M. (2020). Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science (New York, N.Y.)*, *370*(6518), 856–860. https://doi.org/10.1126/science.abd2985

Chang, L., Mondal, A., Singh, B., Martínez-Noa, Y., & Perez, A. (2024). Revolutionizing peptide-based drug discovery: Advances in the post-AlphaFold era. *WIREs Computational Molecular Science*, *14*(1), e1693. https://doi.org/10.1002/wcms.1693

Chen, S., Lin, T., Basu, R., Ritchey, J., Wang, S., Luo, Y., Li, X., Pei, D., Kara, L. B., & Cheng, X. (2024). Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations. *Nature Communications*, *15*(1), 1611. https://doi.org/10.1038/s41467-024-45766-2

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P.-M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., … Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, *15*(141), 20170387. https://doi.org/10.1098/rsif.2017.0387

Danner, S., & Belasco, J. G. (2001). T7 phage display: A novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proceedings of the National Academy of Sciences*, *98*(23), 12954–12959. https://doi.org/10.1073/pnas.211439598

Delgado, J., Radusky, L. G., Cianferoni, D., & Serrano, L. (2019). FoldX 5.0: Working with RNA, small molecules and a new graphical interface. *Bioinformatics*, *35*(20), 4168–4169. https://doi.org/10.1093/bioinformatics/btz184

Ding, Z., Du, W., Lei, Z., Zhang, Y., Zhu, J., Zeng, Y., Wang, S., Zheng, Y., Liu, Z., & Huang, J.-A. (2019). Neuropilin 1 modulates TGF-β1-induced epithelial-mesenchymal transition in non-small cell lung cancer. *International Journal of Oncology*, *56*(2), 531–543. https://doi.org/10.3892/ijo.2019.4938

Dragoni, S., Brash, J. T., Fantin, A., Burridge, C., Denti, L., Turowski, P., & Ruhrberg, C. (2024). Targeting NRP1 with EG00229 induces neurovascular permeability. In *bioRxiv*. https://doi.org/10.1101/2023.11.07.564946

Dunn, J. J., Studier, F. W., & Gottesman, M. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology*, *166*(4), 477–535. https://doi.org/10.1016/S0022-2836(83)80282-4

Eastman, P., Swails, J., Chodera, J. D., McGibbon, R. T., Zhao, Y., Beauchamp, K. A., Wang, L.-P., Simmonett, A. C., Harrigan, M. P., Stern, C. D., Wiewiora, R. P., Brooks, B. R., & Pande, V. S. (2017). OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, *13*(7), e1005659. https://doi.org/10.1371/journal.pcbi.1005659

Fogal, V., Zhang, L., Krajewski, S., & Ruoslahti, E. (2008). Mitochondrial/Cell-Surface Protein p32/gC1qR as a Molecular Target in Tumor Cells and Tumor Stroma. *Cancer Research*, *68*(17), 7210–7218. https://doi.org/10.1158/0008-5472.CAN-07-6752

Giguère, S., Laviolette, F., Marchand, M., Tremblay, D., Moineau, S., Liang, X., Biron, É., & Corbeil, J. (2015). Machine Learning Assisted Design of Highly Active Peptides for Drug Discovery. *PLOS Computational Biology*, *11*(4), e1004074. https://doi.org/10.1371/journal.pcbi.1004074

Gomez-Salinero, J. M., Redmond, D., & Rafii, S. (2025). Microenvironmental determinants of endothelial cell heterogeneity. *Nature Reviews Molecular Cell Biology*, 1–20. https://doi.org/10.1038/s41580-024-00825-w

Gotoh, K., Morisaki, T., Setoyama, D., Sasaki, K., Yagi, M., Igami, K., Mizuguchi, S., Uchiumi, T., Fukui, Y., & Kang, D. (2018). Mitochondrial p32/C1qbp Is a Critical Regulator of Dendritic Cell Metabolism and Maturation. *Cell Reports*, *25*(7), 1800-1815.e4. https://doi.org/10.1016/j.celrep.2018.10.057

Graziadio, A., Zanda, M., Frau, S., Fleming, I., Musolino, M., Dall'Angelo, S., Baldassarre, M., & Piras, M. (2016). NGR Tumor-Homing Peptides: Structural Requirements for Effective APN (CD13) Targeting. In *Bioconjugate chemistry: Vol. 27 5* (pp. 1332–1340). https://doi.org/10.1021/acs.bioconjchem.6b00136

Haspel, N., Zanuy, D., Nussinov, R., Teesalu, T., Ruoslahti, E., & Aleman, C. (2011). Binding of a C-End Rule Peptide to the Neuropilin-1 Receptor: A Molecular Modeling Approach. *Biochemistry*, *50*(10), 1755–1762. https://doi.org/10.1021/bi101662j

Healing, G., & Smith, D. (Eds.). (2000). *Handbook of pre-clinical continuous intravenous infusion*. Taylor & Francis.

Ho, J., Jain, A., & Abbeel, P. (2020). *Denoising Diffusion Probabilistic Models* (arXiv:2006.11239). arXiv. https://doi.org/10.48550/arXiv.2006.11239

Holman, L., Head, M. L., Lanfear, R., & Jennions, M. D. (2015). Evidence of Experimental Bias in the Life Sciences: Why We Need Blind Data Recording. *PLOS Biology*, *13*(7), e1002190. https://doi.org/10.1371/journal.pbio.1002190

Hoppenz, P., Els-Heindl, S., & Beck-Sickinger, A. G. (2020). Peptide-Drug Conjugates and Their Targets in Advanced Cancer Therapies. *Frontiers in Chemistry*, *8*. https://doi.org/10.3389/fchem.2020.00571

Huh, H., Wong, S., St. Jean, J., & Slavcev, R. (2019). Bacteriophage interactions with mammalian tissue: Therapeutic applications. *Advanced Drug Delivery Reviews*, *145*, 4–17. https://doi.org/10.1016/j.addr.2019.01.003

J. Jiang,Y. Zhang,A.R. Krainer,& R. Xu,. (n.d.). *Crystal structure of human p32, a doughnut-shaped acidic mitochondrial matrix protein*. https://doi.org/10.1073/pnas.96.7.3572

Jiang, J., Zhang, Y., Krainer, A. R., & Xu, R.-M. (1999). Crystal structure of human p32, a doughnut-shaped acidic mitochondrial matrix protein. *Proceedings of the National Academy of Sciences*, *96*(7), 3572–3577. https://doi.org/10.1073/pnas.96.7.3572

Jovčevska, I., & Muyldermans, S. (2020). The Therapeutic Potential of Nanobodies. *Biodrugs*, *34*(1), 11–26. https://doi.org/10.1007/s40259-019-00392-z

Jubb, A., Strickland, L. A., Liu, S. D., Mak, J., Schmidt, M., & Koeppen, H. (2012). Neuropilin-1 expression in cancer and development. In *The Journal of Pathology* (Vol. 226). https://doi.org/10.1002/path.2989

Kim, S., Lee, S. H., & Park, J. W. (2004). Selection and characterization of cyclic peptides targeting transferrin receptor by phage display. *Molecules and Cells*, *18*(1), 47–52.

Koivunen, E., Arap, W., Rajotte, D., Lahdenranta, J., & Pasqualini, R. (n.d.). *Identification of Receptor Ligands with Phage Display Peptide Libraries*.

Kolonin, M. G., Sun, J., Do, K. A., Vidal, C. I., Ji, Y., Baggerly, K. A., & Pasqualini, R. (2006). Synchronous selection of homing peptides for multiple tissues by in vivo phage display. *The FASEB Journal*, *20*(7), 979–981.

Kong, X.-D., Carle, V., Díaz-Perlas, C., Butler, K., & Heinis, C. (2020). Generation of a Large Peptide Phage Display Library by Self-Ligation of Whole-Plasmid PCR Product. *ACS Chemical Biology*, *15*(11), 2907–2915. https://doi.org/10.1021/acschembio.0c00497

Krumpe, L. R. H., Atkinson, A. J., Smythers, G. W., Kandel, A., Schumacher, K. M., McMahon, J. B., Makowski, L., & Mori, T. (2006). T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries. *PROTEOMICS*, *6*(15), 4210–4222. https://doi.org/10.1002/pmic.200500606

Kue, C. S., Kamkaew, A., Voon, S. H., Kiew, L. V., Chung, L. Y., Burgess, K., & Lee, H. B. (2016). Tropomyosin Receptor Kinase C Targeted Delivery of a Peptidomimetic Ligand-Photosensitizer Conjugate Induces Antitumor Immune Responses Following Photodynamic Therapy. *Scientific Reports*, *6*(1), 37209. https://doi.org/10.1038/srep37209

Laakkonen, P., Porkka, K., Hoffman, J. A., & Ruoslahti, E. (2002). A tumor-homing peptide with a targeting specificity related to lymphatic vessels. *Nature Medicine*, *8*(7), 751–755. https://doi.org/10.1038/nm720

Ladner, R. C., Sato, A. K., Gorzelany, J., & Souza, M. (2004). Phage display-derived peptides as therapeutic alternatives to antibodies. *Drug Discovery Today*, *9*(12), 525–529.

Lei, Y., Li, X., Qin, D., Zhang, Y., & Wang, Y. (2023). gC1qR: A New Target for Cancer Immunotherapy. *Frontiers in Immunology*, *14*. https://doi.org/10.3389/fimmu.2023.1095943

Lei, Y., Shen, Y., Zuo, C., Lu, L., Crommen, J., Wang, Q., & Jiang, Z. (2022). Emerging affinity ligands and support materials for the enrichment of monoclonal antibodies. *TrAC Trends in Analytical Chemistry*, *157*, 116744. https://doi.org/10.1016/j.trac.2022.116744

Lin, H., Zhang, O., Zhao, H., Jiang, D., Wu, L., Liu, Z., Huang, Y., & Li, S. Z. (2024). *PPFlow: Target-Aware Peptide Design with Torsional Flow Matching* (p. 2024.03.07.583831). bioRxiv. https://doi.org/10.1101/2024.03.07.583831

Lingasamy, P., & Teesalu, T. (2021). Homing Peptides for Cancer Therapy. In F. Fontana & H. A. Santos (Eds.), *Bio-Nanomedicine for Cancer Therapy* (pp. 29–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-58174-9_2

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2023). *Flow Matching for Generative Modeling* (arXiv:2210.02747). arXiv. https://doi.org/10.48550/arXiv.2210.02747

Livesey, J. H., & and Nicholls, M. G. (2012). Stickiness to Hydrophobic Surfaces Varies Widely Among Peptides and Proteins. *Journal of Immunoassay and Immunochemistry*, *33*(3), 302–313. https://doi.org/10.1080/15321819.2011.647188

Ludtke, J. J., Sololoff ,Alex V., Wong ,So C., Zhang ,Guofeng, & and Wolff, J. A. (2007). In Vivo Selection and Validation of Liver-Specific Ligands Using a New T7 Phage Peptide Display System. *Drug Delivery*, *14*(6), 357–369. https://doi.org/10.1080/10717540601098765

Makowski, E. K., Kinnunen, P. C., Huang, J., Wu, L., Smith, M. D., Wang, T., Desai, A. A., Streu, C. N., Zhang, Y., Zupancic, J. M., Schardt, J. S., Linderman, J. J., & Tessier, P. M. (2022). Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. *Nature Communications*, *13*(1), 3788. https://doi.org/10.1038/s41467-022-31457-3

Marshall, R., & Noireaux, V. (2018). Synthetic Biology with an All E. coli TXTL System: Quantitative Characterization of Regulatory Elements and Gene Circuits. In J. C. Braman (Ed.), *Synthetic Biology: Methods and Protocols* (pp. 61–93). Springer New York. https://doi.org/10.1007/978-1-4939-7795-6_4

Mirzaei, M., Mirdamadi, S., Safavi, M., & Soleymanzadeh, N. (2020). The stability of antioxidant and ACE-inhibitory peptides as influenced by peptide sequences. *LWT*, *130*, 109710. https://doi.org/10.1016/j.lwt.2020.109710

Mullard, A. (2025). 2024 FDA approvals. *Nature Reviews Drug Discovery*, *24*(2), 75–82. https://doi.org/10.1038/d41573-025-00001-5

Nakamura, F., & Goshima, Y. (2013). Structural and Functional Relation of Neuropilins. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience. https://www.ncbi.nlm.nih.gov/books/NBK6408/

NDong, C., Tate, J. A., Kett, W. C., Batra, J., Demidenko, E., Lewis, L. D., Hoopes, P. J., Gerngross, T. U., & Griswold, K. E. (2015). Tumor Cell Targeting by Iron Oxide Nanoparticles Is Dominated by Different Factors In Vitro versus In Vivo. *PLOS ONE*, *10*(2), e0115636. https://doi.org/10.1371/journal.pone.0115636

Paasonen, L., Sharma, S., Braun, G. B., Kotamraju, V. R., Chung, T. D. Y., She, Z.-G., Sugahara, K. N., Yliperttula, M., Wu, B., Pellecchia, M., Ruoslahti, E., & Teesalu, T. (2016). New p32/gC1qR Ligands for Targeted Tumor Drug Delivery. *ChemBioChem*, *17*(7), 570–575. https://doi.org/10.1002/cbic.201500564

Pang, H.-B., Braun, G. B., She, Z.-G., Kotamraju, V. R., Sugahara, K. N., Teesalu, T., & Ruoslahti, E. (2014). A free cysteine prolongs the half-life of a homing peptide and improves its tumor-penetrating activity. *Journal of Controlled Release : Official Journal of the Controlled Release Society*, *175*, 48–53. https://doi.org/10.1016/j.jconrel.2013.12.006

Parmley, S. F., & Smith, G. P. (1988). Antibody-selectable filamentous fd phage vectors: Affinity purification of target genes. *Gene*, *73*(2), 305–318. https://doi.org/10.1016/0378-1119(88)90495-7

Pasqualini, R., & Ruoslahti, E. (1996). Organ targeting in vivo using phage display peptide libraries. *Nature*, *380*(6572), 364–366.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators,

and developers. *Protein Science : A Publication of the Protein Society*, *30*(1), 70–82. https://doi.org/10.1002/pro.3943

Pleiko, K., Põšnograjeva, K., Haugas, M., Paiste, P., Tobi, A., Kurm, K., Riekstina, U., & Teesalu, T. (2021). In vivo phage display: Identification of organ-specific peptides using deep sequencing and differential profiling across tissues. *Nucleic Acids Research*, *49*(7), e38. https://doi.org/10.1093/nar/gkaa1279

Põšnograjeva, K., Pleiko, K., & Teesalu, T. (2025). Identification of Organ- and Disease-Specific Homing Peptides Using In Vivo Peptide-Phage Display. In Ü. Langel (Ed.), *Cancer-Targeted Drug Delivery* (pp. 177–212). Springer US. https://doi.org/10.1007/978-1-0716-4374-7_9

Potente, M., & Mäkinen, T. (2017). Vascular heterogeneity and specialization in development and disease. *Nature Reviews Molecular Cell Biology*, *18*(8), 477–494. https://doi.org/10.1038/nrm.2017.36

Puchakayala, H. C., Bhatnagar, P., Nambiar, P., Dutta, A., & Mitra, D. (2023). Design of a machine learning-aided screening framework for antibiofilm peptides. *Digital Chemical Engineering*, *8*, 100107. https://doi.org/10.1016/j.dche.2023.100107

Rajotte, D., & Ruoslahti, E. (1999). Membrane Dipeptidase Is the Receptor for a Lung-targeting Peptide Identified by in Vivo Phage Display. *Journal of Biological Chemistry*, *274*(17), 11593–11598. https://doi.org/10.1074/jbc.274.17.11593

Reth, M. (2013). Matching cellular dimensions with molecular sizes. *Nature Immunology*, *14*(8), 765–767. https://doi.org/10.1038/ni.2621

Roy, A., Shi, L., Chang, A., Dong, X., Fernandez, A., Kraft, J. C., Li, J., Le, V. Q., Winegar, R. V., Cherf, G. M., Slocum, D., Poulson, P. D., Casper, G. E., Vallecillo-Zúniga, M. L., Valdoz, J. C., Miranda, M. C., Bai, H., Kipnis, Y., Olshefsky, A., … Baker, D. (2023). De novo design of highly selective miniprotein inhibitors of integrins αvβ6 and αvβ8. *Nature Communications*, *14*(1), 5660. https://doi.org/10.1038/s41467-023-41272-z

Ruoslahti, E. (1996). RGD AND OTHER RECOGNITION SEQUENCES FOR INTEGRINS. *Annual Review of Cell and Developmental Biology*, *12*(1), 697–715. https://doi.org/10.1146/annurev.cellbio.12.1.697

Ruoslahti, E. (2002a). Drug targeting to specific vascular sites. *Drug Discovery Today*, *7*(22), 1138–1143. https://doi.org/10.1016/S1359-6446(02)02501-1

Ruoslahti, E. (2002b). Specialization of tumour vasculature. *Nature Reviews Cancer*, *2*(2), 83–90. https://doi.org/10.1038/nrc724

Ruoslahti, E. (2004). Vascular zip codes in angiogenesis and metastasis. *Biochemical Society Transactions*, *32*(3), 397–402. https://doi.org/10.1042/bst0320397

Ruoslahti, E. (2012). Peptides as Targeting Elements and Tissue Penetration Devices for Nanoparticles. *Advanced Materials*, *24*(28), 3747–3756. https://doi.org/10.1002/adma.201200454

Ruoslahti, E. (2017). Tumor penetrating peptides for improved drug delivery. *Advanced Drug Delivery Reviews*, *110–111*, 3–12. https://doi.org/10.1016/j.addr.2016.03.008

Ruoslahti, E. (2022). Molecular ZIP codes in targeted drug delivery. *Proceedings of the National Academy of Sciences*, *119*(28), e2200183119. https://doi.org/10.1073/pnas.2200183119

Ruoslahti, E., Bhatia, S. N., & Sailor, M. J. (2010). Targeting of drugs and nanoparticles to tumors. *The Journal of Cell Biology*, *188*(6), 759–768. https://doi.org/10.1083/jcb.200910104

Sharma, S., Kotamraju, V. R., Mölder, T., Tobi, A., Teesalu, T., & Ruoslahti, E. (2017). Tumor-Penetrating Nanosystem Strongly Suppresses Breast Tumor Growth. *Nano Letters*, *17*(3), 1356–1364. https://doi.org/10.1021/acs.nanolett.6b03815

Shoombuatong, W., Schaduangrat, N., Pratiwi, R., & Nantasenamat, C. (2019). THPep: A machine learning-based approach for predicting tumor homing peptides. *Computational Biology and Chemistry*, *80*, 441–451. https://doi.org/10.1016/j.compbiolchem.2019.05.008

Sloth, A. B., Bakhshinejad, B., Jensen, M., Stavnsbjerg, C., Liisberg, M. B., Rossing, M., & Kjaer, A. (2022). Analysis of Compositional Bias in a Commercial Phage Display Peptide Library by Next-Generation Sequencing. *Viruses*, *14*(11), 2402. https://doi.org/10.3390/v14112402

Smith, G. P. (1985). Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science*, *228*(4705), 1315–1317. https://doi.org/10.1126/science.4001944

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). *Deep Unsupervised Learning using Nonequilibrium Thermodynamics* (arXiv:1503.03585). arXiv. https://doi.org/10.48550/arXiv.1503.03585

Srinivasarao, M., & Low, P. S. (2017). Ligand-Targeted Drug Delivery. *Chemical Reviews*, *117*(19), 12133–12164. https://doi.org/10.1021/acs.chemrev.7b00013

Sugahara, K. N., Teesalu, T., Karmali, P. P., Kotamraju, V. R., Agemy, L., Girard, O. M., Hanahan, D., Mattrey, R. F., & Ruoslahti, E. (2009). Tissue-Penetrating Delivery of Compounds and Nanoparticles into Tumors. *Cancer Cell*, *16*(6), 510–520. https://doi.org/10.1016/j.ccr.2009.10.013

Sugahara, K. N., Teesalu, T., Karmali, P. P., Kotamraju, V. R., Agemy, L., Greenwald, D. R., & Ruoslahti, E. (2010). Coadministration of a Tumor-Penetrating Peptide Enhances the Efficacy of Cancer Drugs. *Science*, *328*(5981), 1031–1035. https://doi.org/10.1126/science.1183057

T7Select® Packaging Kit | 70014. (n.d.). Retrieved May 6, 2025, from https://www.merckmillipore.com/INTL/en/product/T7Select-Packaging-Kit,EMD_BIO-70014#anchor_USP

Teesalu, T., Sugahara, K., Kotamraju, V. R., & Ruoslahti, E. (2009). C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. In *Proceedings of the National Academy of Sciences* (Vol. 106, pp. 16157–16162). https://doi.org/10.1073/pnas.0908201106

Teesalu, T., Sugahara, K. N., & Ruoslahti, E. (2012). Chapter two—Mapping of Vascular ZIP Codes by Phage Display. In K. D. Wittrup & G. L. Verdine (Eds.), *Methods in Enzymology* (Vol. 503, pp. 35–56). Academic Press. https://doi.org/10.1016/B978-0-12-396962-0.00002-1

Timur, S. S., Yalçın ,Gözde, Çevik ,Özge, Andaç ,Cenk, & and Gürsoy, R. N. (2018). Molecular dynamics, thermodynamic, and mutational binding studies for tumor-specific LyP-1 in complex with p32. *Journal of Biomolecular Structure and Dynamics*, *36*(5), 1134–1144. https://doi.org/10.1080/07391102.2017.1313779

Trédan, O., Galmarini, C. M., Patel, K., & Tannock, I. F. (2007). Drug Resistance and the Solid Tumor Microenvironment. *JNCI: Journal of the National Cancer Institute*, *99*(19), 1441–1454. https://doi.org/10.1093/jnci/djm135

Tymecka, D., Redkiewicz, P., Lipiński, P. F. J., & Misicka, A. (2024). Peptidomimetic inhibitors of the VEGF-A165/NRP-1 complex obtained by modification of the C-terminal arginine. In *Amino Acids* (Vol. 56). https://doi.org/10.1007/s00726-024-03411-8

Vadevoo, S. M. P., Gurung, S., Khan, F., Haque, Md. E., Gunassekaran, G. R., Chi, L., Permpoon, U., & Lee, B. (2019). Peptide-based targeted therapeutics and apoptosis imaging probes for cancer therapy. *Archives of Pharmacal Research*, *42*(2), 150–158. https://doi.org/10.1007/s12272-019-01125-0

Varanasi, S. M., Gulani, Y., Rachamala, H. K., Mukhopadhyay, D., & Angom, R. S. (2025). Neuropilin-1: A Multifaceted Target for Cancer Therapy. *Current Oncology*, *32*(4), Article 4. https://doi.org/10.3390/curroncol32040203

Vázquez Torres, S., Benard Valle, M., Mackessy, S. P., Menzies, S. K., Casewell, N. R., Ahmadi, S., Burlet, N. J., Muratspahić, E., Sappington, I., Overath, M. D., Rivera-de-Torre, E., Ledergerber, J., Laustsen, A. H., Boddum, K., Bera, A. K., Kang, A., Brackenbrough, E., Cardoso, I. A., Crittenden, E. P., … Baker, D. (2025). De novo designed proteins neutralize lethal snake venom toxins. *Nature*, *639*(8053), 225–231. https://doi.org/10.1038/s41586-024-08393-x

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., Ahern, W., Borst, A. J., Ragotte, R. J., Milles, L. F., Wicky, B. I. M., Hanikel, N., Pellock, S. J., Courbet, A., Sheffler, W., Wang, J., Venkatesh, P., Sappington, I., Torres, S. V., … Baker, D. (2023). De novo design of protein structure and function with RFdiffusion. *Nature*, *620*(7976), 1089–1100. https://doi.org/10.1038/s41586-023-06415-8

Werle, M., & Bernkop-Schnürch, A. (2006). Strategies to improve plasma half life time of peptide and protein drugs. *Amino Acids*, *30*(4), 351–367. https://doi.org/10.1007/s00726-005-0289-3

Wu, X., Lin, H., Bai, R., & Duan, H. (2024). Deep learning for advancing peptide drug development: Tools and methods in structure prediction and design. *European Journal of Medicinal Chemistry*, *268*, 116262. https://doi.org/10.1016/j.ejmech.2024.116262

Xian, X., Ren, Q., Du, H., Qi, Y., & Yan, J. (2025). Advances in Homing Peptide Targeted Therapies. *International Journal of Peptide Research and Therapeutics*, *31*(3), 36. https://doi.org/10.1007/s10989-024-10682-y

Xiao, W., Jiang, W., Chen, Z., Huang, Y., Mao, J., Zheng, W., Hu, Y., & Shi, J. (2025). Advance in peptide-based drug development: Delivery platforms, therapeutics and vaccines. *Signal Transduction and Targeted Therapy*, *10*(1), 1–56. https://doi.org/10.1038/s41392-024-02107-5

Yenugonda, V., Nomura, N., Kouznetsova, V., Tsigelny, I., Fogal, V., Nurmemmedov, E., Kesari, S., & Babic, I. (2017). A novel small molecule inhibitor of p32 mitochondrial protein overexpressed in glioma. *Journal of Translational Medicine*, *15*, 210. https://doi.org/10.1186/s12967-017-1312-7

Yu, B., Su, H., Zhao, L., Yang, J., Zhu, M., & Zhao, J. (2022). 99mTc-labeled iRGD for single-positron emission computed tomography imaging of triple-negative breast cancer. In *Frontiers in Bioengineering and Biotechnology* (Vol. 10). https://doi.org/10.3389/fbioe.2022.1001899

Yue, H., Li, Y., Yang, M., & Mao, C. (2022). T7 Phage as an Emerging Nanobiomaterial with Genetically Tunable Target Specificity. *Advanced Science*, *9*(4), 2103645. https://doi.org/10.1002/advs.202103645

Zanuy, D., Kotla, R., Nussinov, R., Teesalu, T., Sugahara, K. N., Alemán, C., & Haspel, N. (2013). Sequence dependence of *C-end rule* peptides in binding and activation of neuropilin-1 receptor. *Journal of Structural Biology*, *182*(2), 78–86. https://doi.org/10.1016/j.jsb.2013.02.006

Zhang, Y., & Sanner, M. F. (2019). AutoDock CrankPep: Combining folding and docking to predict protein–peptide complexes. *Bioinformatics*, *35*(24), 5121–5127. https://doi.org/10.1093/bioinformatics/btz459

Zhou, J., & Huang, M. (2024). Navigating the landscape of enzyme design: From molecular simulations to machine learning. *Chemical Society Reviews*, *53*(16), 8202–8239. https://doi.org/10.1039/D4CS00196F

9. Addendums

9.1. NRP-1 Generated Peptides

Peptide	Sequence	Binding Affinity (kcal/mol)	Cluster Size	CendR motif	Charge at pH 7.4	Lipophilic C-terminus	Isoelectric Point	Hydrophobicity (GRAVY)
25	EYNMGKDK	-16.5	27	FALSE	-0.34	FALSE	6.17	-2.26
23	GDHGELWY	-15.6	25	FALSE	-2.41	TRUE	4.35	-1.18
36	GRLDYLGR	-15.2	33	TRUE	0.56	FALSE	8.75	-0.88
39	WYGLRGYG	-14.7	20	FALSE	0.55	FALSE	8.59	-0.68
12	VMQEDQGG	-14.2	64	FALSE	-2.48	FALSE	4.05	-1.09
7	SGIDYGGS	-14	85	FALSE	-1.75	FALSE	4.05	-0.39
49	GGKRGYDS	-13.6	34	FALSE	0.55	FALSE	8.59	-1.90
27	PLFIRSGR	-13.5	12	TRUE	1.90	FALSE	12.00	-0.09
1	NGYSRAQC	-13.3	25	FALSE	0.53	FALSE	8.22	-1.21
41	QEESGAQS	-13.3	27	FALSE	-2.44	FALSE	4.05	-1.78
15	QRLMASVR	-13.1	52	TRUE	1.56	FALSE	12.00	-0.20
45	GQFGGMGG	-13.1	44	FALSE	-0.44	FALSE	5.53	-0.10
22	GLLFMGAG	-13.1	52	FALSE	-0.44	FALSE	5.53	1.61
26	RRTCRQQV	-13	40	FALSE	2.53	TRUE	11.70	-1.81
42	DGGHIRNR	-12.9	11	TRUE	0.59	FALSE	9.61	-1.94
28	KRKRRPGG	-12.9	11	FALSE	4.55	FALSE	12.00	-2.96
37	KGHRGPGG	-12.7	16	FALSE	1.59	FALSE	11.00	-1.85
35	KRVGLPHG	-12.6	36	FALSE	1.59	FALSE	11.00	-0.75
30	GMDQEPGW	-12.6	26	FALSE	-2.44	TRUE	4.05	-1.49

19	GGQLSEYS	-12.5	3	FALSE	-1.44	FALSE	4.05	-0.86
31	CGAKGGRG	-12.5	32	FALSE	1.53	FALSE	9.51	-0.71
8	GGHGLHLR	-12.2	21	TRUE	0.63	FALSE	9.76	-0.56
18	DYPGNPGR	-12.2	9	TRUE	-0.44	FALSE	5.84	-2.10
16	IGPGQSQE	-11.9	26	FALSE	-1.44	FALSE	4.60	-1.15
48	GPGGRALG	-11.9	51	FALSE	0.56	FALSE	9.75	-0.26
33	GTNGPRCD	-11.7	19	FALSE	-0.47	FALSE	6.08	-1.51
2	IPSPGVFR	-11.5	23	TRUE	0.56	FALSE	9.75	0.33
43	GQPGGMRS	-11.4	13	FALSE	0.56	FALSE	9.75	-1.21
17	DKPPINGG	-11.4	16	FALSE	-0.44	FALSE	5.84	-1.30
24	GGGPVRTF	-11.3	14	FALSE	0.56	TRUE	9.75	-0.13
29	GGPPKKTG	-11.1	79	FALSE	1.55	FALSE	10.00	-1.61
13	PCPVKLPG	-11	26	FALSE	0.87	FALSE	8.64	0.18
10	RGAGLAGS	-11	23	FALSE	0.56	FALSE	9.75	0.11
14	YPAAGEAG	-10.9	93	FALSE	-1.44	FALSE	4.05	-0.23
34	SKPGAVGL	-10.8	10	FALSE	0.25	TRUE	8.47	0.34
44	GGGGEGAL	-10.7	19	FALSE	-1.44	TRUE	4.05	0.01
6	GVLAPGGA	-10.6	50	FALSE	-0.44	TRUE	5.53	1.10
32	PPCAGSTG	-10.4	19	FALSE	-0.12	FALSE	5.91	-0.15
47	PGPSRPLR	-10.3	49	TRUE	1.90	FALSE	12.00	-1.40
38	GRSPGPTE	-10	39	FALSE	-0.44	FALSE	6.22	-1.69
0	CGGGAPPG	-9.8	60	FALSE	-0.47	FALSE	5.52	-0.06
reference	GRPARPAR	-9.7	4	TRUE	2.56	FALSE	12.00	-1.69

5	GGGPGPAG	-8.7	36	FALSE	-0.44	FALSE	5.53	-0.43

9.2. P32 Generated Peptides

Peptide	Sequence	Binding Affinity (kcal/mol)	Cluster Size	Cend R motif	Charge at pH 7.4	Lipophilic C-terminus	Isoelectric Point	Hydrophobicity (GRAVY)
49	RLHLLNLRL	-19.9	15	FALSE	1.59	TRUE	12.00	0.37
26	QKHLYLLKL	-19.7	26	FALSE	1.59	TRUE	9.70	-0.07
41	KALLQLAKL	-19.6	24	FALSE	1.55	TRUE	10.00	0.83
62	RKTSLLAAC	-18.5	14	FALSE	1.53	FALSE	9.51	0.42
64	DMLMYARLK	-18.5	9	FALSE	0.55	FALSE	8.59	0.00
8	GGWLSKLWP	-18.5	12	FALSE	0.55	FALSE	8.75	-0.14
45	STGKLLRVG	-18.5	43	FALSE	1.25	FALSE	11.00	0.12
reference	CGNKRTRGC	-18.3	20	FALSE	2.51	FALSE	9.70	-1.43
89	HLLLLRLLY	-18.2	29	FALSE	0.59	TRUE	8.75	1.53
87	GFRLSNPTK	-18.1	30	FALSE	1.55	FALSE	11.00	-0.98
31	GAATEWAKR	-18	46	TRUE	0.56	FALSE	8.75	-0.94
76	KHCSNKLTL	-17.6	25	FALSE	1.56	TRUE	9.31	-0.66
58	PLHFWVLWV	-17.4	28	FALSE	-0.06	TRUE	7.17	1.36
75	SVLRWILML	-17.2	28	FALSE	0.25	TRUE	9.47	1.76
25	KGADALWLN	-17.2	26	FALSE	-0.44	FALSE	5.84	-0.11
90	WTHCLGDNW	-17.1	43	FALSE	-1.43	TRUE	5.08	-0.76
66	SIRLHMQKT	-17.1	14	FALSE	1.29	FALSE	11.00	-0.71
6	FQGSYADLK	-17.1	14	FALSE	-0.45	FALSE	5.83	-0.56
71	DPRKTSVKL	-17	25	FALSE	1.55	TRUE	9.99	-1.21

73	RALGLYWSQ	-17	9	FALSE	0.55	FALSE	8.75	-0.22
44	SGLWNPLLL	-17	53	FALSE	-0.75	TRUE	5.24	0.89
74	RYLALSFAR	-16.9	42	TRUE	1.55	FALSE	10.84	0.32
47	SNLVLFLNY	-16.8	33	FALSE	-0.75	TRUE	5.24	1.03
48	AELRLLLWI	-16.8	54	FALSE	-0.39	TRUE	6.05	1.40
38	YLLLQQRGQ	-16.8	28	FALSE	0.55	FALSE	8.75	-0.59
67	SMALGLNTK	-16.7	43	FALSE	0.25	FALSE	8.47	0.22
69	MYFDSNKFH	-16.7	20	FALSE	-0.68	FALSE	6.49	-0.97
13	SDSGFYIYI	-16.6	50	FALSE	-1.75	TRUE	4.05	0.41
18	FEGQLLLRS	-16.4	24	FALSE	-0.44	FALSE	6.00	0.17
37	MMRCGLVRL	-16.3	19	FALSE	1.26	TRUE	10.35	0.97
81	KLCLILLLM	-16.3	20	FALSE	0.53	TRUE	8.22	2.67
68	KRMLLMSLP	-16.2	33	FALSE	1.55	FALSE	11.00	0.49
60	HGGLLLLWC	-16.2	37	FALSE	-0.43	FALSE	6.73	1.42
27	HLRLNLVPL	-15.9	42	FALSE	0.59	TRUE	9.76	0.73
29	SHLRSHLVL	-15.8	60	FALSE	0.33	TRUE	9.51	0.34
52	GLFKYLLCA	-15.7	16	FALSE	0.53	TRUE	8.20	1.43
51	MSVLLQVLY	-15.6	15	FALSE	-0.72	TRUE	5.27	1.79
78	PYMLEDLGR	-15.6	53	TRUE	-1.10	FALSE	4.37	-0.59
63	GRPDLGELE	-15.6	37	FALSE	-2.44	FALSE	4.43	-1.09
82	CLRWDMLLT	-15.5	26	FALSE	-0.47	FALSE	5.83	0.69

77	GGLAKNRSA	-15.3	21	FALSE	1.55	TRUE	11.00	-0.68
24	LELQDKGIA	-15.2	24	FALSE	-1.44	TRUE	4.37	-0.10
83	DGLNCQKGA	-15	81	FALSE	-0.47	TRUE	5.83	-0.79
55	KDLHALSGC	-14.9	20	FALSE	-0.43	FALSE	6.73	0.01
80	SQLLSALLL	-14.7	36	FALSE	-0.75	TRUE	5.24	1.74
5	GQCIAHEQS	-14.6	50	FALSE	-1.43	FALSE	5.24	-0.68
32	SKLAGAAQL	-14.5	49	FALSE	0.25	TRUE	8.47	0.49
33	VMPDIACWL	-14.5	36	FALSE	-1.50	TRUE	4.05	1.41
42	PLAVLILQV	-14.4	45	FALSE	-0.10	TRUE	5.95	2.33
46	CLRAAHSPG	-14.4	42	FALSE	0.57	FALSE	8.26	-0.07
79	NASLSWAQL	-14.2	12	FALSE	-0.44	TRUE	5.53	0.19
21	TIIDRLSGG	-14.2	20	FALSE	-0.79	FALSE	5.50	0.28
70	SIDSLFELC	-13.9	13	FALSE	-2.77	FALSE	4.05	0.98
35	STLMLVCLL	-13.7	21	FALSE	-0.77	TRUE	5.24	2.48
65	VVGLLADMS	-13.7	10	FALSE	-1.48	FALSE	4.05	1.67
86	SDNELLTLL	-13.6	43	FALSE	-2.75	TRUE	4.05	0.36
2	GRANPILCD	-13.3	49	FALSE	-0.47	FALSE	6.08	-0.10
61	LDCEPFDLM	-13	41	FALSE	-3.47	TRUE	4.05	0.30
72	SALLCGGPG	-12.9	83	FALSE	-0.77	FALSE	5.24	0.92
57	CLLAADGLA	-12.9	38	FALSE	-1.47	TRUE	4.05	1.71
43	FLSAPASIL	-12.9	6	FALSE	-0.44	TRUE	5.53	1.70
84	DNPAGLLLP	-12.6	43	FALSE	-1.44	FALSE	4.05	0.29
53	TPYEGPLLP	-11.4	20	FALSE	-1.79	FALSE	4.05	-0.34

39	ALHDPALAA	-10.6	34	FALSE	-1.36	TRUE	5.08	0.72

Non-exclusive licence to reproduce the thesis and make the thesis public

I, Jasper August Tootsi ,

1. grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the digital archives of the University of Tartu until the expiry of the term of copyright, my thesis

   Development of tissue-specific homing peptides using probabilistic machine learning ,

   supervised by Tambet Teesalu and Karlis Pleiko and Juhan Sedman ;

2. grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright;

3. am aware of the fact that the author retains the rights specified in points 1 and 2;

4. confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Jasper August Tootsi

20/05/2025