Compilation and Analysis of Tumor-Homing Peptides

A Project Report

submitted by

DIKSHA KASHYAP

in partial fulfilment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

Department of Computational Biology
INDRAPRASTHA INSTITUTE OF INFORMATION TECHNOLOGY DELHI

NEW DELHI- 110020

June, 2025

THESIS CERTIFICATE

This is to certify that the thesis titled **"Compilation and Analysis of Tumor-Homing Peptides"**, submitted by **Diksha Kashyap**, to the Indraprastha Institute of Information Technology, Delhi, for the award of the degree of Master of Technology, is a bona fide record of the research work done by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Prof. Gajendra Pal Singh Raghava
Thesis Supervisor
Professor & Head of Department
Dept. of Computational Biology
IIIT Delhi, 110020

Place: New Delhi

Date: June 2025

ACKNOWLEDGEMENTS

I would like to genuinely express my gratitude and recognition to **Prof. Gajendra P. S. Raghava** , my M.Tech thesis supervisor at the Indraprastha Institute of Information Technology, Delhi, for permitting me to conduct research under his mentorship, guiding me from the initial stages of the topic to fostering my development as a researcher. His patience and dedication inspired me to persist throughout my research journey.

In addition to my supervisor, I wish to convey my heartfelt thanks to Ph.D. scholar Naman Mehta for their constant support and encouragement, as well as for being a key player in the entire research endeavor. Finally, I want to extend my appreciation to my family and friends for their unwavering support and encouragement, which allowed me to pursue my research.

Diksha Kashyap
MT23228
M.Tech(Computational Biology)

ABSTRACT

TumorHoPe2 is a comprehensive bioinformatics resource focused on tumor homing peptides (THPs), designed to support cancer research and targeted therapy development. Leveraging the base of a previous platform created in 2012, this new system much more expansively and deeply offers data. The new version includes 1847 carefully curated records, which constitute 1297 distinct peptide sequences, effectively doubling the dataset of its successor. Each peptide entry is annotated with essential attributes such as amino acid sequence, terminal or chemical modifications, associated tumor types, and corresponding cancer cell lines. This information is sourced from two key experimental methodologies: phage display libraries and synthetic peptides.

The dataset includes 594 chemically modified peptides, detailing 255 N-terminal modifications and 195 C-terminal modifications. These peptides have been studied across 172 distinct cancer cell lines, targeting 37 different tumor types. Advanced features like search, filtering, and visualization tools, as well as a RESTful API are offered by the database to ensure seamless accessibility and usefulness for researchers. The resource is freely accessible at: https://webs.iiitd.edu.in/raghava/tumorhope2

KEYWORDS: Tumor Homing Peptides (THPs); Bioinformatics Database; Phage Display; Synthetic Peptides; Cancer Cell Lines

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

THP	Tumor Homing Peptide
EPR	Enhanced Permeability and Retention
TME	Tumor Microenvironment
LAMP	Linux, Apache, MySQL, PHP
HTTP	HyperText Transfer Protocol
MySQL	My Structured Query Language
PDB	Protein Data Bank
DSSP	Dictionary of Secondary Structure of Proteins
BLAST	Basic Local Alignment Search Tool
URL	Uniform Resource Locator
CURL	Client URL
JSON	JavaScript Object Notation
AAC	Amino Acid Composition

CHAPTER 1

Introduction

1.1 Background on Cancer and Limitations of Conventional Therapy

Cancer is known to be one of the leading causes of illness and mortality across the globe. It encompasses a broad spectrum of diseases, all marked by the abnormal and uncontrolled growth of cells, which can infiltrate neighboring tissues and, in many cases, metastasize to distant parts of the body through the circulatory or lymphatic systems. As reported by international cancer surveillance organizations, over 19 million people were newly diagnosed with some form of cancer in 2020, and close to 10 million individuals lost their lives to the disease in the same year [1]. These numbers portray the enormous burden cancer places not only on patients and their families but also on public health worldwide.

Even though we've made great progress in understanding cancer and coming up with new treatment options, the most common methods doctors still rely on are surgery, chemotherapy, and radiation. These have been around for decades and continue to help, especially when cancer is caught early. They've saved lives and improved survival for countless individuals. But the problem arises when the cancer begins to spread or lose its response to these therapies. In those situations, the old-fashioned approaches usually are not sufficient to prevent the cancer from growing or significantly improving the patients' outcomes in the long run.

1.1.1 Systemic Toxicity and the Narrow Therapeutic Window

One of the biggest issues with conventional chemotherapy is that it does not only attack cancer cells, but also healthy cells. Because these medications are formulated to kill rapidly growing cells, they accidentally damage quickly dividing normal cells, such as those in the bone marrow, gastrointestinal tract, and hair follicles. That is why nausea,

tiredness, compromised immune function, oral sores, and hair loss occur so frequently. The range between a toxic dose and an effective dose of a therapy is usually narrow. Doctors are constantly trying to find that delicate balance too little might not stop the cancer, while too much can seriously hurt the patient.

1.1.2 Tumor Microenvironment and Challenges in Drug Delivery

Solid tumors don't develop by themselves, they emerge and progress inside a multifaceted and commonly aggressive environment called the tumor microenvironment (TME). Anything but a mere background player, the TME is instead actively involved in cancer development. It's composed of several non-tumor cells such as immune cells, fibroblasts, and endothelial cells, all housed in a highly dense, irregular extracellular matrix. One of the defining characteristics of this environment is its poorly organized blood vessel network, which is incapable of providing the tumor with sufficient oxygen and nutrients. This results in areas of low oxygen levels (hypoxia) and non-uniform drug delivery. In addition to these challenges, tumors also have acidic pH, high interstitial pressure, and rigid surrounding tissue conditions that not only assist in cancer cells' survival and evasion of the immune system but also disrupt how efficiently treatments can gain access to and impact their target. [2,3]. These obstacles frequently translate to even effective medicines being unable to work properly, providing some cancer cells with a chance to survive and be resistant to therapy.

1.1.3 Cancer Heterogeneity and Resistance

Perhaps the greatest challenge of cancer to treat is its built-in diversity. In a single tumor, there are potentially numerous different cell types, each with its own genetic characteristics and behavior. This diversity, which scientists call tumor heterogeneity, implies that all of the cancer cells will not respond equally well to treatment. While others may be eliminated by therapy, still others may resist or learn to adapt and become predominant. This tends to ultimately result in the recurrence of the cancer, sometimes more aggressively and/or harder to treat.

1.1.4 Limitations of Radiation and Surgical Approaches

While radiation therapy is capable of targeting tumors with decent precision, its use is limited by the potential to harm surrounding healthy tissues, especially in delicate areas like the brain or spinal cord. Surgery tends to work well when tumors are localized and easy to reach, but not all cancers are. Deep-seated, inaccessible, or widely spread tumors are often beyond what can be safely removed. Even if a surgery appears successful, there's always the chance that microscopic cancer cells are left behind. If not properly managed with follow-up treatments like chemo or radiation, these leftover cells can grow back and sometimes the cancer that comes back is more aggressive than the one initially treated.

1.2 Rise of Tumor-Targeting and Tumor-Homing Peptides in Cancer Therapeutics

Even after decades of progress in cancer research, conventional treatment methods such as chemotherapy, radiotherapy, and surgery continue to face major challenges. A lot of these treatments are seen to be associated with significant side effects, and even delivering therapeutic agents effectively into tumor tissues remains difficult. Along with this, cancer cells often become resistant to treatment after a while, which causes curtailing long-term success. This points out the great need to identify some new strategies that are able to target tumor cells specifically with minimal damage to healthy tissues. Among these options, one is the application of tumor-homing peptides (THPs), which possess the unique capacity to bind selectively to tumors or to the environment surrounding them. This form of targeted treatment not only has the potential to enhance the outcome of treatment but can also prove useful in minimizing the damage that is typically encountered with conventional cancer treatments.

Tumor-homing peptides are brief peptides of amino acids, usually ranging from 5 to 30 residues long, engineered or identified to specifically recognize and bind molecular markers expressed preferably or exclusively in tumor cells, tumor vasculature, or the tumor microenvironment (TME) [4]. Such specific binding enables THPs to act as

"guided missiles" for targeting therapeutic drugs into cancerous tissue, thus enhancing treatment efficacy and decreasing systemic side effects.

The idea of using peptides for targeted delivery to tumors started gaining popularity in the late 1990s, driven by improvements in phage display technology and combinatorial peptide libraries. These methods allowed the targeting of peptides that target tumors by binding to overexpressed receptors or distinctive characteristics of the TME [5, 6]. In contrast to antibodies or bigger targeting constructs, THPs are advantageous due to their small size, allowing them to penetrate more deeply into the tumor and be cleared more quickly from non-target tissues, reducing undesired toxicity [7]. THPs achieve their tumor selectivity through a number of mechanisms:

- **Receptor-Mediated Targeting:** Several THPs are engineered to bind with high specificity to those receptors overexpressed on tumor vasculature or tumor cells. For instance, integrins (e.g., $\alpha_v\beta_3$ and $\alpha_v\beta_5$), aminopeptidase N (CD13), and nucleolin are expressed richly in the majority of solid tumors[8, 9].
- **Tumor Microenvironment Sensitivity:** Certain THPs get activated or cleaved selectively by the specific biochemical conditions in tumors, e.g., acidic pH or increased protease activity[10].
- **Leveraging the Enhanced Permeability and Retention (EPR) Effect:** THPs are able to passively accumulate in tumors, increasing the local concentration of therapeutic drugs, by exploiting the EPR effect [11].

The advantages of THPs over traditional therapies are substantial. Their ability to selectively localize within tumors allows for:

- Increased concentration of drugs at the disease site while reducing exposure to healthy tissues, improving the therapeutic index.
- Enhanced penetration into dense tumor tissues, overcoming one of the major obstacles in solid tumor treatment.
- Flexibility in conjugation with a wide array of therapeutic and diagnostic agents including chemotherapy drugs, radionuclides, imaging probes, nanoparticles, and nucleic acid-based therapeutics making them versatile tools for theranostics and personalized medicine [12].

One of the pioneering successes in the field is the peptide iRGD (internalizing RGD), which targets integrins on tumor endothelial cells and facilitates deeper tumor penetration by engaging neuropilin-1 receptors. Co-administration of iRGD with chemotherapeutic agents has been shown in animal models to significantly improve drug delivery and tumor suppression, while reducing systemic toxicity [13, 14].

1.3 General Characteristics, Mechanisms of Action, and Applications of Tumor-Homing Peptides

Besides their selective targeting characteristics, tumor-homing peptides (THPs) have a number of structural and biochemical properties that enable their development as clinically effective medicines. These peptides can be obtained from natural sequences or synthetically engineered to maximize therapeutic potential. Structural modifications such as cyclization, N-terminal acetylation, C-terminal amidation, and the incorporation of D-amino acids are frequently employed to enhance resistance to proteolytic degradation and prolong circulatory half-life without compromising affinity for tumor targets [14,15].

One of the earliest established methods for THP identification involves phage display, a technology that allows for high-throughput screening of peptide libraries for binding to tumor-specific molecular markers. In this method, bacteriophages are designed to have their surfaces decorated with many different peptides, and those with affinity for specific tumor cells or the tumor microenvironment (TME) are recovered through iterative rounds of biopanning. In vivo phage display, specifically, has been crucial in the identification of peptides that target tumors under physiological conditions. Peptides like LyP-1 and F3 have been identified using it and localized to tumor vasculature and nuclei, respectively [16,17].
Apart from phage display, improvements in computational modeling and high-throughput peptide screening have also made the rational design and optimization of THPs possible. The techniques enable the prediction of peptide-receptor binding affinity and conformational stability, greatly minimizing experimental costs and speeding up the discovery pipeline [18].

THPs serve as modular platforms that can be conjugated to a wide spectrum of diagnostic and therapeutic agents depending on the intended clinical application. Notable examples include:

- **Drug conjugates:** THPs are conjugated with chemotherapeutics like doxorubicin or paclitaxel, which enables targeted delivery and minimized off-target toxicity [19].
- **Imaging agents:** Fluorophore- or radioisotope-labeled peptides are applied to improve tumor imaging in modalities like fluorescence imaging and PET/CT for

early diagnosis and intraoperative navigation [20].

- **Nanoparticle targeting:** THPs can be usually grafted onto nanocarriers such as liposomes, micelles, or dendrimers in order to enable active targeting and enhance the pharmacokinetics of nanomedicines [21].
- **Gene delivery vectors:** Some peptides have been modified to deliver nucleic acids like siRNA, mRNA, or plasmids to facilitate localized modulation of genes within tumors [22].

Mechanisms of Tumor Localization

THPs exploit a variety of biological mechanisms to selectively accumulate in tumors:

- **Receptor-mediated targeting:** Many THPs bind to receptors overexpressed in tumor tissue. For instance, RGD motifs target $\alpha_v\beta_3$ and $\alpha_v\beta_5$ integrins involved in angiogenesis, while NGR peptides bind aminopeptidase N (CD13), a protease frequently found in tumor vasculature [23,24].
- **Microenvironment-responsive activation:** The unique conditions of the tumor microenvironment such as acidic pH, hypoxia, and elevated protease activity are exploited by pH-sensitive or protease-activated peptides to ensure selective activation within the tumor site [25].
- **CendR pathway:** Certain peptides, notably iRGD, first bind integrins and are subsequently cleaved to expose a C-terminal "C-end Rule" (CendR) motif, which binds neuropilin-1 (NRP-1). This interaction triggers a transport mechanism that enhances extravasation and deep tissue penetration [6].
- **Enhanced Permeability and Retention (EPR) effect:** Tumors often exhibit leaky vasculature and impaired lymphatic drainage, allowing macromolecules and conjugated peptides to accumulate passively. While the EPR effect alone is often insufficient for precise targeting, it can be synergistically exploited by actively targeted THPs [26].

1.4 Motivation for Database Development: TumorHoPe and Its Advancement

Tumor-homing peptides (THPs) have been the focus of greater attention in the past few years because of their significance in targeted cancer diagnosis and therapy. TumorHoPe is now the sole dedicated database specifically dealing with THPs, aggregating scientifically proven data on their structures, sequences, and biological functions. The initial TumorHoPe database set the platform for this topic by collecting information up to 2012, serving researchers as a valuable reference point at the moment.

Given the significant expansion of THP-related research over the past decade, this thesis presents TumorHoPe2, a substantially updated and extended version of the database. This new release captures developments up to 2024, incorporating data from peer-reviewed publications, patents, and newly characterized peptides. In addition to a broader and more current dataset, TumorHoPe2 provides enhanced details on peptide modifications, structural attributes, binding targets, and therapeutic applications. For researchers in oncology, peptide engineering, and drug delivery systems, the platform is now an even stronger and more usable resource due to its new interface and increased browsing and search capabilities.

CHAPTER 2

Database Design and Compilation

2.1 Evolution from TumorHoPe to TumorHoPe2

The shift from the original TumorHoPe database to TumorHoPe2 marks a significant step forward in both the depth and breadth of tumor-homing peptide (THP) data. While the first version mainly served as a well-curated collection of cancer-targeting peptides gathered from scattered sources, TumorHoPe2 builds on that foundation and takes things much further. It not only includes more peptide entries but also provides richer, more varied information, making it a more powerful and accessible tool for researchers.

One of the key improvements in TumorHoPe2 is the addition of over a thousand newly identified peptides. These sequences were manually curated from two major information repositories: synthetic peptides and phage display libraries. Six hundred fifty-two peptides were sourced from patents, while 457 were sourced from scientific studies, with a minor overlap. Together with the 744 peptides from the previous database(TumorHoPe), the total number of entries is 1,847, comprising 1,297 non-redundant sequences.

An analysis of the database shows that specific peptide lengths are more common than others. Short peptides, especially those with nine amino acids, appear most frequently. In contrast, more extended sequences, those with more than 20 residues are relatively rare. This pattern likely reflects practical choices made during peptide design and synthesis, as shorter peptides are generally easier to produce and handle in tumor-targeting research.

The peptide set in TumorHoPe2 covers a broad spectrum of cancer types. Breast cancer remains the most represented, followed by lung, prostate, melanoma, and brain cancers. Many entries are also associated with other solid tumors such as pancreatic, colon, and gastric cancers. Notably, the updated dataset also includes peptides reported for relatively less studied cancers like neuroblastoma and osteosarcoma, thereby enhancing the biological scope and relevance of the platform.

There has been an increase in the Cell line diversity. In TumorHoPe2, peptides have been validated across 172 cancer cell lines, as compared to 83 in the previous version. Specific cell lines, including MDA-MB-435, 4T1, B16F1, PC-3, and PPC1, appear in a notably high number of entries.

Another notable improvement in TumorHoPe2 is the thorough annotation of structural properties and chemical modifications for each peptide. Peptides are systematically classified as either linear or cyclic, and all relevant modifications—whether at the terminal regions or within the core sequence—are carefully recorded to provide a complete understanding of their design and functionality. Most peptides remain linear in structure, although cyclic forms and hybrid entries are also present. Additional structural analysis shows that most peptides are rich in flexible coil-like conformations and lack pronounced secondary structure elements. This is consistent with their short length and adaptability in receptor binding.

The interface has been redesigned to increase usability and include various search options, such as keyword-based queries, peptide-specific searches, and advanced filtering capabilities. Prominent motifs like RGD, NGR, and CendR remain among the most frequently observed. The number of unique sequences has almost doubled, and chemically modified peptides have more than tripled. The database now supports information across various cancer types and includes data from a broader array of biological models. A key advancement is the development of the MAP (Modification and Annotation in Proteins) format, which makes it possible to embed annotation tags at the residue level.

2.2 Data collection methodology (manual curation, literature mining)

To construct the TumorHoPe2 dataset, a structured manual curation strategy was adopted. Data were gathered from two primary sources: published research articles accessed through PubMed and patent filings available via Patent Lens. The curated dataset ultimately included 1,847 entries, with 457 peptides from PubMed, 652 from patent documents, and 744 inherited from the previous version. Duplicate entries (6 in total) across sources were carefully reconciled during the integration process.

Relevant studies were identified using keyword searches, such as "tumor-homing peptides" and "tumor-targeting peptides". The review covered literature published between 2012 and 2024. Only peptides with experimental validation were considered. Each peptide entry was carefully checked by hand to gather important details like its amino acid sequence, references to the original studies, and how it was tested or validated. In addition to this core information, we also noted other valuable features, such as its structural traits, the types of amino acids it contains, and its basic physicochemical properties. This multi-tiered annotation approach offered essential insights and a contextual understanding of each peptide's biological function.

2.3 Inclusion criteria for peptides

Peptides included in TumorHoPe2 were selected using strict criteria to ensure data reliability and relevance. Only experimentally validated peptides through in vitro, in vivo, or confirmed binding studies were accepted, ensuring the database represents biologically meaningful THPs and excludes unverified or predicted sequences.
To ensure quality and relevance, peptides were included in TumorHoPe2 database, only if they met the following four criteria:

- **Experimental Evidence:** Validated through lab-based methods such as phage display, binding assays, or functional studies in cancer models.
- **Peer-reviewed Source:** Sourced from PubMed-indexed journals or verified patents (2012–2024); studies based solely on computational predictions were excluded.
- **Tumor Specificity:** Demonstrated selective binding to tumor tissues, cells, or vasculature. Peptides with only general activity were not considered.
- **Minimum Information:** Core metadata is required, including sequence, source (PubMed ID or patent), target type, and validation method.

2.4 Database schema, backend architecture and front-end interface

In order to continue with the increasing THP information and ensure ease of use on the platform, TumorHoPe2 was rewritten with a newer, more dynamic web framework.

The new platform combines a responsive front-end with a strong, scalable data storage system that allows for efficient data storage and retrieval. Data integrity was given priority, performance optimized, and a friendly interface was offered to enable free access.

2.4.1 Backend Architecture

The TumorHoPe2 backend is based on the LAMP stack, consisting of the Apache HTTP Server, MySQL database engine, and PHP for server-side scripting. This setup provides a robust platform-independent base that facilitates dynamic content generation and seamless data transactions. MySQL acts as the data repository, handling large-scale peptide-related entries with the help of relational schema design. This facilitates complex queries, optimized index creation, and storage scalability. The backend was modularly expandable, allowing integration with external scripts and APIs in order to be upgraded in the future.

2.4.2 Database Schema and Organization

The schema of the TumorHoPe2 database is structured in a way that facilitates storage, retrieval, and analysis of tumour-homing peptide data. The database is implemented using MySQL. The database consists of six tables. The heart of TumorHoPe2 is the main table, which contains the primary peptide entries. Each record in this table represents a unique tumor-homing peptide and includes comprehensive metadata across various experimental and biological dimensions:

- **Peptide Identification:** id, sequence, name_source, pmid, title, year
- **Target Information:** target_cell, target_tumor, receptors_biomarker
- **Experimental Validation:** invivo, invitro, phage_display
- **Structural/Functional Tags:** motif, n_term, c_term, homing, conjugate, patent
- **Miscellaneous:** comments, ref, source

This table serves as the central node, linking to auxiliary tables via the id field. To ensure modularity and scalability, secondary information is distributed across designated tables, which are as follows:

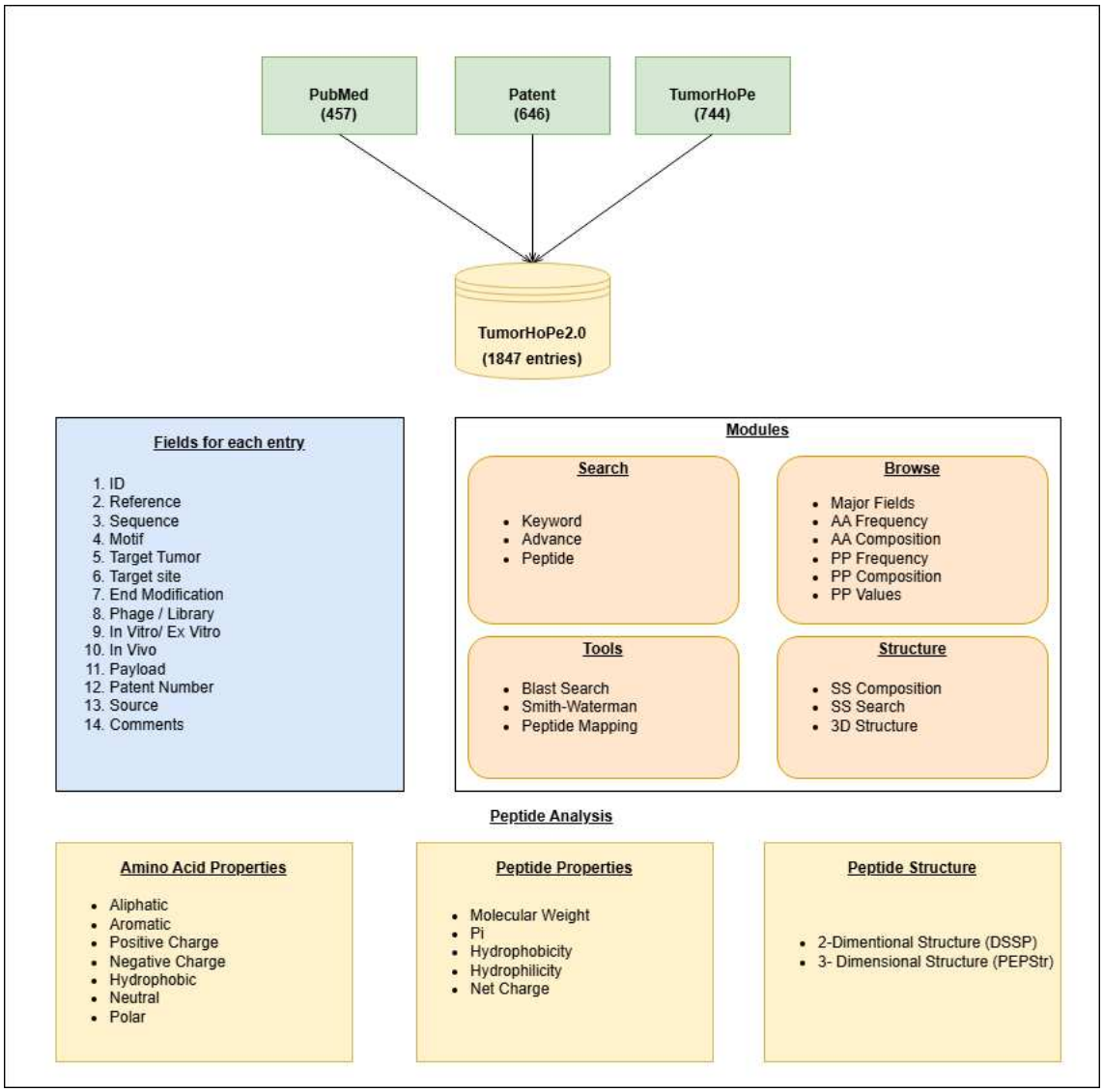

Figure 2.1: Schematic representation of TumorHoPe2 Database

- **aaccomposition, aafreq, aaproperty:** Store detailed information about amino acid-level characteristics, such as but not limited to, composition percentages and physicochemical properties (e.g., hydrophobicity, charge).
- **dssp:** This table contains secondary structure information (Helix, Strand, Coil) derived from DSSP annotations, supporting structure-based retrieval and browsing tools.
- **property:** Includes overall physicochemical parameters like isoelectric point, molecular weight, and net charge, calculated from the peptide sequence.

Each of these auxiliary tables establishes a foreign key relationship with the id attribute in the main table, thereby maintaining referential integrity across the schema. This relational design facilitates normalized data storage and enables efficient JOIN operations to implement advanced query, search, and analysis functionalities.

2.4.3 Front-End Interface

The front-end interface is built to provide a responsive, interactive, and user-friendly experience. Enhanced with client-side scripting, the interface offers real-time feedback, dynamic filtering, and seamless page transitions. The web portal is structured around five primary modules: keyword search, advanced search, peptide-based search, browsing tools, and structure-based retrieval.

Users can perform basic searches using free text inputs or refine results through the advanced search module, which supports Boolean operators and field-specific filters. The peptide search tool allows both exact match and subsequence-based querying, enhancing functional discovery. In addition, the browsing interface supports exploration based on target tumor type, cell line, year of publication, amino acid composition, and physicochemical properties.

2.4.4 Visualization and Analytical Tools

TumorHoPe2 includes a structure-based search module that utilizes predicted peptide conformations. The secondary structure browser allows filtering based on DSSP annotations, while a built-in 3D viewer (powered by NGL) supports interactive visualization of tertiary structures.

For computational tasks, BLAST and Smith-Waterman sequence alignment algorithms

are incorporated to facilitate similarity searches. Local alignment is carried out utilizing a sliding window technique for finding conserved regions. Furthermore, users can enter custom sequences in FASTA format for calculating properties such as hydrophobicity, isoelectric point, net charge, and molecular weight.

2.4.5 REST API Integration

A RESTful API interface was used to simplify data access via APIs and support automated workflows. This interface supports users to query the database via standard HTTP requests, with URL, CURL, or Wget formats supported, and JSON formats returned. The API endpoints support selective searches across given fields, such as peptide source, target cell type, and tumor category, to support integration of data into bioinformatics pipelines and web applications.

CHAPTER 3

Results and Analysis

3.1 Sequence and Structural Characteristics of THPs

To gain a better insight into the functional diversity and structural principles of tumor-homing peptides (THPs), we compared their sequence-dependent and structural features. This comprised the assessment of peptide size distributions to detect typical size patterns, the occurrence of established tumor-targeting motifs like RGD and NGR, categorization of peptides according to structural alterations and topology, and secondary structure prediction with DSSP. These analyses unveil important information on the discriminatory features of THPs and their prospects in selective cancer targeting.

3.1.1 Peptide Length Distribution

A detailed analysis of the length and the number of peptide distributions, helped reveal that short peptides are highly prevalent in the dataset. The majority of THPs fall within the 0–10 amino acid range (773 peptides), followed by the 10–20 range (453 peptides). Peptides of exactly 9 amino acids were found to be the most dominant, highlighting a preference for short sequences in tumor-targeting applications. Peptides longer than 20 amino acids are relatively rare, with only 53 peptides in the 20–30 range, and fewer than 20 peptides beyond 30 residues.

3.1.2 Motif Analysis

In TumorHoPe2, the presence of known tumor-targeting motifs was analyzed, revealing that several well-established motifs were commonly found. The RGD motif appeared most frequently, found in 89 peptides, followed by NGR in 54 peptides and CendR in 51 peptides. Apart from these, a few less common motifs like RXD/QXR (8 peptides), LTVXPW (3 peptides), and bombesin (2 peptides) were identified.

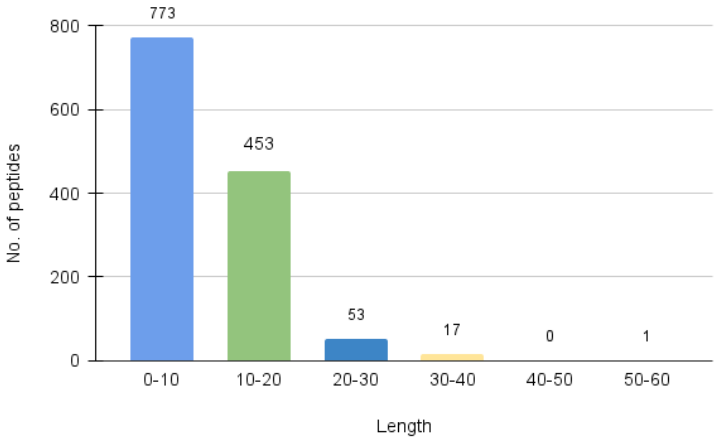

Figure 3.1: Distribution of unique tumor-homing peptides based on their sequence length.

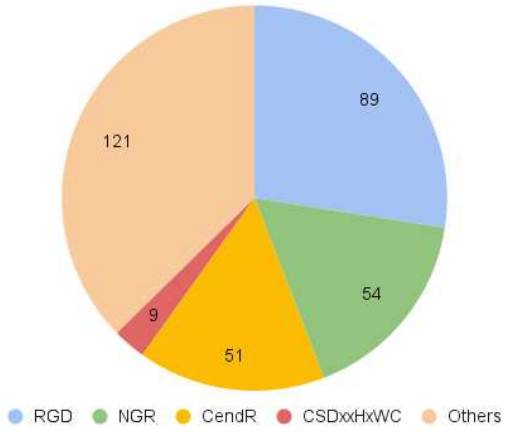

Figure 3.2: Distribution of unique tumor-homing peptides based on their sequence length

3.1.3 Structural Classification

Most peptides in the database are linear, with 1,789 entries, while 58 peptides are cyclic, and just one was noted to exist in both forms. In addition to the basic structure, many peptides in the database are chemically modified. These include changes at the N-terminal, C-terminal, or within the core sequence. There were a total of 255 peptides with modification at the N-terminal, while 195 peptides with modification at the C-terminal.

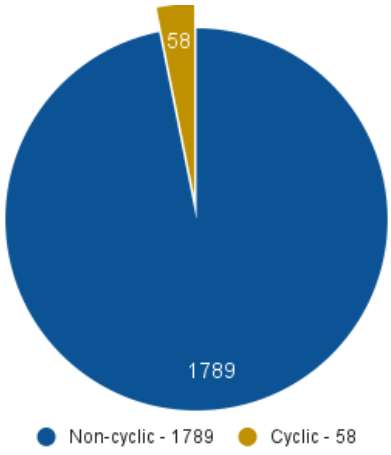

Figure 3.3: Distribution of entries based on cyclic and non-cyclic peptide types

3.1.4 Secondary and Tertiary Structure Analysis

To further explore the structure of THPs, we predicted the secondary structures and tertiary structure of the peptides, using DSSP and PEPstr respectively. Out of all entries, a large majority, 1,608 peptides exhibited low helical content (0–20%), and 1,790 peptides had minimal β-strand content in the same range. A similar trend was seen for turn structures, with 709 peptides showing low occurrence. Only a small number of peptides showed moderate to high secondary structure content, such as 136 peptides with 20–40% helicity and 85 peptides with 40–60% helicity. Notably, just 14 peptides had significant helical regions above 60%, and none had prominent β-strands in that range. Interestingly, 495 peptides were almost entirely composed of unstructured or coil regions, indicating a high degree of conformational flexibility.

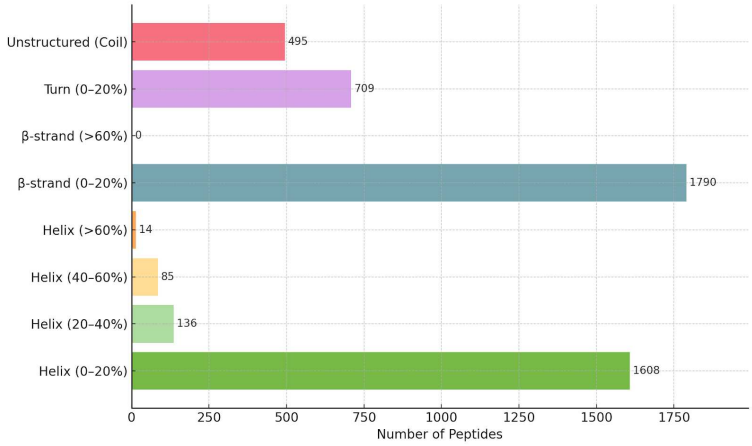

Figure 3.4: Distribution of entries based on cyclic and non-cyclic peptide types

3.2 Tumor and Targeting Spectrum

3.2.1 Tumor Type Coverage

TumorHoPe2 includes THPs targeting a wide range of cancer types. Among the entries, breast cancer accounts for the largest share, with 511 peptides associated with this tumor type. This is followed by lung (151 peptides), prostate (146), melanoma (135), and brain cancer (134). The dataset also highlights peptides tested against pancreatic (94), colon (58), gastric (54), and medullary thyroid (33) cancers. Interestingly, even rare or underrepresented cancers such as osteosarcoma, neuroblastoma, and esophageal/gastroesophageal cancer are included, albeit with fewer entries.

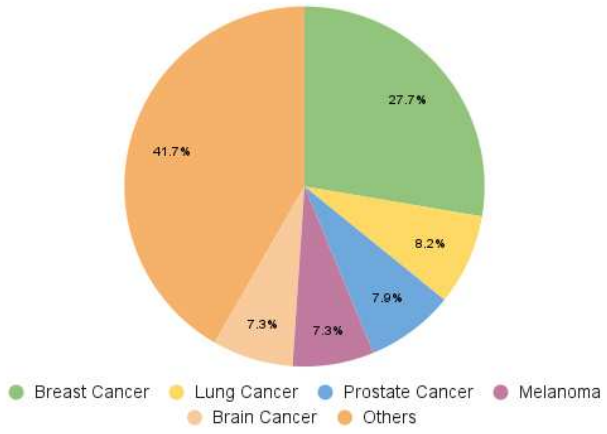

Figure 3.5: Distribution of peptide entries across various cancer types

3.2.2 Target Cell and Receptor Distribution

TumorHoPe2 exhibits a broad range of target cell specificity, highlighting the complex cellular heterogeneity of the tumor microenvironment. While tumor cells remain the predominant target, many peptides also home to tumor vasculature, endothelial cells, angiogenic vessels, tumor lymphatics, and tumor stem cells, among others.

3.3 Experimental Validation Models

3.3.1 Cell Line Diversity

The peptides in TumorHoPe2 have been validated across an extensive range of cancer cell lines. Peptides were most commonly validated in MDA-MB-435 (292 entries), 4T1 (233), B16F1 (214), PC-3 (209), and PPC1 (206) cell lines. Other frequently used cell lines include MCF-7 (146), HeLa (71), and A549 (55). These targeting preferences are largely influenced by the overexpression of tumor-associated receptors such as integrins (e.g. $\alpha_v\beta_3$) and neuropilin-1, which are frequently recognized by motifs like RGD and CendR, respectively [27, 28].

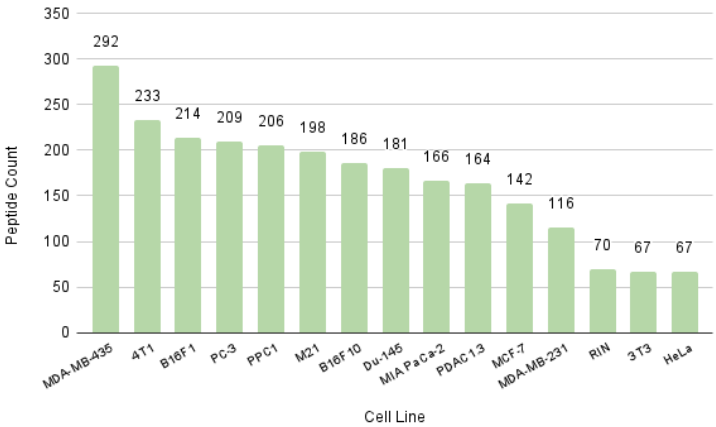

Figure 3.6: Distribution of peptides validated across different cancer cell lines

3.3.2 In Vivo Validation

The In vivo validations were performed on a wide range of murine models, including immunocompetent (e.g., BALB/c, C57BL/6, ICR) and immunodeficient strains (e.g., athymic nude, SCID, NOD/SCID, NSG).

3.4 Molecular Annotation and Chemical Modification Features in TumorHoPe2

Chemical modifications are critical in peptide engineering, as they influence peptide behavior in biological environments modulating serum stability, resistance to enzymatic degradation, binding affinity, and cell penetration. TumorHoPe2 provides detailed annotations of such modifications, making it the first THP-specific resource to comprehensively catalog these molecular alterations.

A notable advancement in TumorHoPe2 is the adoption of the MAP (Modification and Annotation in Proteins) format, which facilitates detailed, residue-level annotation of peptides. While traditional FASTA formats are limited to linear amino acid sequences and they do not provide any contextual biological information, the MAP format extends this structure to include embedded annotations that describe chemical modifications, non-canonical residues, cleavage sites, and functional motifs directly within the sequence.

The MAP format [29] includes bracketed tags to peptide sequences that are human as well as machine readable. The method is capable of annotating sequence properties.

The MAP format utilizes standardized tags to annotate structural and chemical modifications directly within peptide sequences. Key examples include:

- **Cyclization:** {cyc:N-C} for head-to-tail cyclization or {cyc:2-4} for disulfide bridges.
- **Terminal Modifications:** {nt:Amid} (N-terminal amidation), {ct:Acet} (C-terminal acetylation).
- **Non-Canonical Features:** {d} for D-amino acids, {nnm:PEG} for PEGylation, and ptm:Glyc for glycosylation.

Incorporating the MAP format into TumorHoPe2 has rendered annotation far more accurate. By incorporating tags into peptide sequences, we were able to capture such par-

ticular modifications such as PEGylation, cyclization, or D-amino acids at the residue level. This simplified sorting, querying, and analyzing peptides according to their chemical characteristics.

In data curation, the MAP format facilitated quicker identification of altered peptides and facilitated direct input into motif analysis or structure prediction tools without the requirement for external annotation files. Since the format is machine-parsable as well as human-readable, it also facilitates inclusion into computational pipelines, and hence it is a useful option to be employed both in wet-lab as well as in silico approaches.

CHAPTER 4

Discussion

This chapter critically assesses the order, form, and function of tumor-homing peptides (THPs) listed in the TumorHoPe2 database from a biological interpretation and therapeutic applicability perspective, with consideration for how these observations can inform design paradigms for future cancer-targeting molecules. It also considers the important methodological advances that set TumorHoPe2 apart from the initial TumorHoPe platform.

4.1 Interpreting Tumor-Homing Peptide Characteristics

One of the consistent trends that we have seen throughout the database is the prevalence of short peptides, with most sequences being 9 amino acids in size. This trend is probably influenced both by practical and biological factors. Technically, shorter peptides are less expensive to produce and are generally more stable in biological systems [30]. Their compact size also allows for effective tissue penetration without increasing the chance of eliciting immune responses [31], also making them perfect for drug development. Also, it is observed that shorter peptides can have better access to the cell surface receptors through the extracellular matrix, enhancing their targeting efficiency.

Another striking observation is the consistent occurrence of well-documented motifs like RGD, NGR, and CendR. The RGD motif is well known to interact with integrins, specifically $\alpha_v\beta_3$, which are generally upregulated in tumor vasculature and are implicated in angiogenesis [32]. NGR motifs bind to CD13 (aminopeptidase N), a protein that is present in high levels on tumor blood vessels [33]. CendR motifs that bind neuropilin-1 enable more penetration into deeper tissues and enhance intracellular delivery by the C-end Rule pathway [34]. Their repetition in a wide range of tumor models reinforces their ongoing applicability in cancer-targeting design. Further, unusual motifs like RXD, QXR, and LTVXPW, and bioactive peptides like bombesin, indicate that the area may be worth exploring further in terms of lesser-appreciated targeting mechanisms [34].

The structural examination of peptides in TumorHoPe2 demonstrates a sharp preference for disordered or coil-dominant conformations. Although this may at first glance seem to be a weakness, such structural freedom can also be beneficial. Flexible peptides can more readily bind to various receptor conformations in an induced-fit mechanism[35]. This plasticity also favors efficient membrane interaction and enables peptides to operate in the highly dynamic environment of the tumor microenvironment [36]. In contrast to stiff proteins, these peptides are able to conform to the heterogeneous and frequently capricious morphology of tumor cells and adjacent tissues.

Chemical modifications and peptide topology also contribute importantly to the tuning of biological activity. Cyclic peptides, for instance, tend to be more stable towards enzymatic cleavage and frequently have greater receptor affinity as a result of their constrained structure [37].

Similarly, terminal modifications like N-terminal acetylation or C-terminal amidation can enhance serum stability and minimize immune recognition [38]. Chemical modifications such as PEGylation or incorporation of D-amino acids become more common to prolong circulation and optimize further pharmacokinetics [39]. TumorHoPe2 documents these properties at the residue level so that users can select peptides according to structural characteristics. It thus qualifies as a useful tool for rational peptide design and preclinical drug development.

4.2 Broader Implications for Therapeutic Design and Research

TumorHoPe2's broad coverage of tumor types, including underrepresented tumors like neuroblastoma and osteosarcoma, significantly increases its therapeutic utility. By expanding beyond common cancers such as breast and lung, the database supports a broader research community and highlights neglected cancer forms in need of focused delivery tactics [40]. The variety of target cell types ranging from tumor cells to tumor vasculature, lymphatics, and stem cells reflects a growing realization that effective therapy must address not just the tumor itself but also the surrounding milieu [41]. This variation in target spectrum shows THPs' versatility and potential in multi-modal treatment regimens.

The cell lines and animal models used in experimental validation further lend credibility to the curated peptides. Frequent use of established cancer lines like MDA-MB-435 and 4T1, as well as mouse models like BALB/c and athymic nude mice, reflects standardization in peptide evaluation across the field [42]. This consistency enables comparative analysis and provides a strong foundation for reproducibility. Moreover, the fact that these peptides have been validated experimentally not just computationally adds significant translational value, particularly for pre-clinical pipeline development [43].

4.3 Comparison with Previous Version

The upgraded version of the database is a considerable improvement over the TumorHoPe database in that it includes more content and increased capabilities for cancer-targeting research. TumorHope2 nearly doubles the number of peptide entries, each carefully curated from experimentally validated sources, including both peer-reviewed literature and patent databases. This broader dataset covers a wider range of cancer types, receptor targets, and cell models, improving the representativeness and utility of the resource.

Several key additions set TumorHoPe2 apart. The adoption of the MAP (Modification and Annotation in Proteins) format enables residue-level annotation of post-translational modifications, cyclization, and non-canonical residues, facilitating advanced structural and functional analysis. A 3D structure viewer has also been included, allowing users to visually explore peptide conformations and spatial characteristics. This is extremely valuable for structure-activity studies. Additionally, a rich search interface now enables users to query the database with detailed biochemical, structural, or validation parameters, significantly enhancing user engagement and accuracy.

The additions improve TumorHoPe2 from a static archive to a dynamic analysis tool. Although the initial TumorHoPe was a valuable foundation for the field, TumorHoPe2 expands its scope and utility and thus is a valuable and useful tool for peptide-based cancer studies.

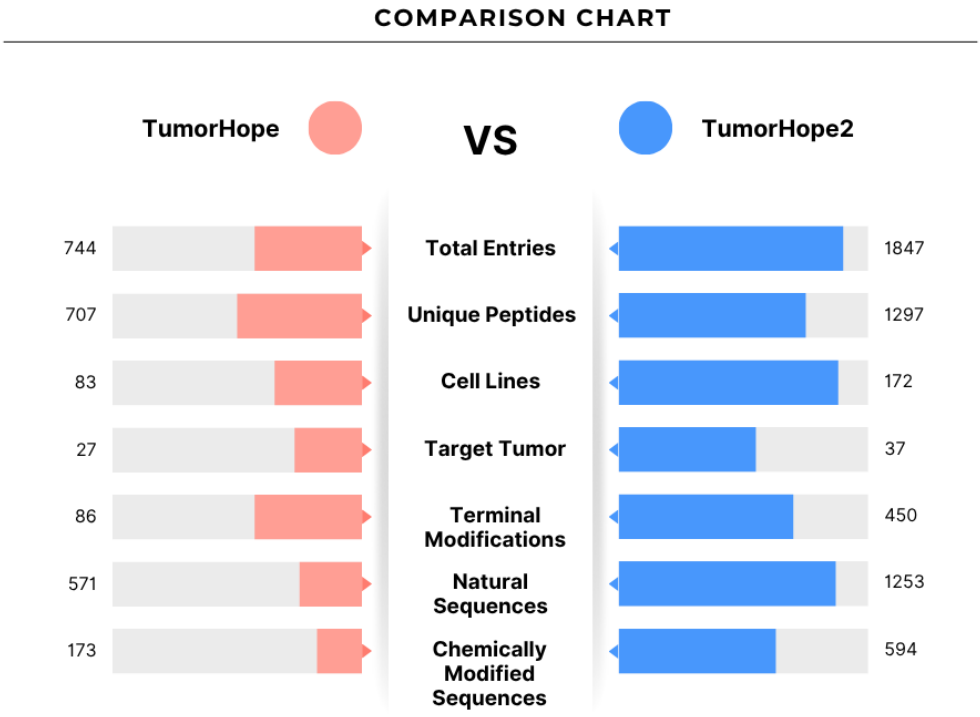

Figure 4.1: Visual overview showcasing the key advancements in TumorHoPe2

CHAPTER 5

Conclusion and Future Directions

In conclusion, TumorHoPe2 is a major milestone in tumor-targeting peptide research. The database has emerged as a powerful platform for both exploratory analysis and focused peptide design via the inclusion of experimentally validated data, an extension of curated entries, and the incorporation of advanced features such as structure visualization, MAP-based residue annotations, and a refined search interface. Recurrent patterns of peptide size, sequence motifs, and structural characteristics validate known biological principles and further indicate new lines of research, for example unexplored motifs and chemical modifications with biomedical potential. Future directions involve continued work on the construction of a specialized prediction server for tumor-homing peptides. Looking ahead, ongoing efforts are focused on the development of a dedicated prediction server for tumor-homing peptides.

This tool will allow users to computationally search and forecast peptides with tumor-targeting activity based on established motifs, physicochemical characteristics, and structural features. Deployed, it will augment TumorHoPe2 by offering a design-oriented, forward-looking extension to the current repository. This is a significant milestone toward establishing an end-to-end ecosystem for data-driven discovery as well as rational design of peptide-based therapeutics for cancer research.

REFERENCES

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

[2] Holohan, C., Van Schaeybroeck, S., Longley, D. B., & Johnston, P. G. (2013). Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer*, *13*(10), 714–726. https://doi.org/10.1038/nrc3599

[3] Jain, R. K. (2001). Delivery of molecular and cellular medicine to solid tumors. *Advanced Drug Delivery Reviews*, *46*(1-3), 149–168. https://doi.org/10.1016/s0169-409x(00)00131-9

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

[5] Teesalu, T., Sugahara, K. N., & Ruoslahti, E. (2013). Tumor-penetrating peptides. *Frontiers in Oncology*, *3*, 216. https://doi.org/10.3389/fonc.2013.00216

[6] 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

[7] Zhang, M., & Xu, H. (2023). Peptide-assembled nanoparticles targeting tumor cells and tumor microenvironment for cancer therapy. *Frontiers in Chemistry*, *11*, 1115495. https://doi.org/10.3389/fchem.2023.1115495

[8] 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

[9] Svensen, N., Walton, J. G., & Bradley, M. (2012). Peptides for cell-selective drug delivery. *Trends in Pharmacological Sciences*, *33*(4), 186–192. https://doi.org/10.1016/j.tips.2012.02.002

[10] Boohaker, R. J., Lee, M. W., Vishnubhotla, P., Perez, J. M., & Khaled, A. R. (2012). The use of therapeutic peptides to target and to kill cancer cells. *Current Medicinal Chemistry*, *19*(22), 3794–3804. https://doi.org/10.2174/092986712801661004

[11] 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

[12] Yin, H., Yang, J., Zhang, Q., Yang, J., Wang, H., Xu, J., & Zheng, J. (2017). iRGD as a tumor-penetrating peptide for cancer therapy (Review). *Molecular Medicine Reports, 15*(5), 2925–2930. https://doi.org/10.3892/mmr.2017.6419

[13] Gaurav, I., Wang, X., Thakur, A., Iyaswamy, A., Thakur, S., Chen, X., Kumar, G., Li, M., & Yang, Z. (2021). Peptide-Conjugated Nano Delivery Systems for Therapy and Diagnosis of Cancer. *Pharmaceutics, 13*(9), 1433. https://doi.org/10.3390/pharmaceutics13091433

[14] Lu, L., Zhang, Q., Wang, Z., Gao, L., & Shen, J. (2021). Peptide-Modified Nanoparticles for Tumor Targeting and Molecular Imaging. *Current Medicinal Chemistry, 28*(31), 6411–6436. https://doi.org/10.2174/0929867327666201022122131

[15] Fosgerau, K., & Hoffmann, T. (2015). Peptide therapeutics: current status and future directions. *Drug Discovery Today, 20*(1), 122–128. https://doi.org/10.1016/j.drudis.2014.10.003

[16] Vlieghe, P., Lisowski, V., Martinez, J., & Khrestchatisky, M. (2010). Synthetic therapeutic peptides: science and market. *Drug Discovery Today, 15*(1-2), 40–56. https://doi.org/10.1016/j.drudis.2009.10.009

[17] Pasqualini, R., & Ruoslahti, E. (1996). Organ targeting In vivo using phage display peptide libraries. *Nature, 380*, 364–366. https://doi.org/10.1038/380364a0

[18] Porkka, K., Laakkonen, P., Hoffman, J. A., Bernasconi, M., & Ruoslahti, E. (2002). A fragment of the HMGN2 protein homes to the nuclei of tumor cells and tumor endothelial cells in vivo. *PNAS, 99*(11), 7444–7449. https://doi.org/10.1073/pnas.062189599

[19] Conibear, A. C., Schmid, A., Kamalov, M., Becker, C. F. W., & Bello, C. (2020). Recent Advances in Peptide-Based Approaches for Cancer Treatment. *Current Medicinal Chemistry, 27*(8), 1174–1205. https://doi.org/10.2174/0929867325666171123204851

[20] Zwicke, G. L., Mansoori, G. A., & Jeffery, C. J. (2012). Utilizing the folate receptor for active targeting of cancer nanotherapeutics. *Nano Reviews, 3*, 18496. https://doi.org/10.3402/nano.v3i0.18496

[21] Chen, K., & Chen, X. (2010). Design and development of molecular imaging probes. *Current Topics in Medicinal Chemistry, 10*(12), 1227–1236. https://doi.org/10.2174/156802610791384225

[22] Bertrand, N., Wu, J., Xu, X., Kamaly, N., & Farokhzad, O. C. (2014). Cancer nanotechnology: The impact of passive and active targeting in the era of modern cancer biology. *Advanced Drug Delivery Reviews, 66*, 2–25. https://doi.org/10.1016/j.addr.2013.11.009

[23] Koivunen, E., Arap, W., Rajotte, D., Lahdenranta, J., & Pasqualini, R. (1999). Identification of receptor ligands with phage display peptide libraries. *Journal of Nuclear Medicine, 40*(5), 883–888.

[24] Christian, S., Pilch, J., Akerman, M. E., Porkka, K., Laakkonen, P., & Ruoslahti, E. (2003). Nucleolin expressed at the cell surface is a marker of endothelial cells in angiogenic blood vessels. *The Journal of Cell Biology, 163*(4), 871–878. https://doi.org/10.1083/jcb.200304132

[25] Gialeli, C., Theocharis, A. D., & Karamanos, N. K. (2011). Roles of matrix metalloproteinases in cancer progression and their pharmacological target-

ing. *The FEBS Journal, 278*(1), 16–27. https://doi.org/10.1111/j.1742-4658.2010.07919.x

[26] Maeda, H., Wu, J., Sawa, T., Matsumura, Y., & Hori, K. (2000). Tumor vascular permeability and the EPR effect in macromolecular therapeutics: a review. *Journal of Controlled Release, 65*(1-2), 271–284. https://doi.org/10.1016/s0168-3659(99)00248-5

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

[28] Teesalu, T., Sugahara, K. N., Kotamraju, V. R., & Ruoslahti, E. (2009). C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. *PNAS, 106*(38), 16157–16162. https://doi.org/10.1073/pnas.0908201106

[29] Shendre, A., Mehta, N. K., Rathore, A. S., Kumar, N., Patiyal, S., & Raghava, G. P. S. (2025). MAP format for representing chemical modifications, annotations, and mutations in protein sequences: An extension of the FASTA format (arXiv No. 2505.03403). *arXiv*. https://doi.org/10.48550/arXiv.2505.03403

[30] Fosgerau, K., & Hoffmann, T. (2015). Peptide therapeutics: current status and future directions.*Drug discovery today, 20*(1), 122–128. https://doi.org/10.1016/j.drudis.2014.10.003

[31] Scodeller, P., & Asciutto, E. K. (2020). Targeting Tumors Using Peptides. *Molecules, 25*(4), 808. https://doi.org/10.3390/molecules25040808

[32] Desgrosellier, J. S., & Cheresh, D. A. (2010). Integrins in cancer: biological implications and therapeutic opportunities. *Nature Reviews Cancer, 10*(1), 9–22. https://doi.org/10.1038/nrc2748

[33] Pasqualini, R., Koivunen, E., & Ruoslahti, E. (1997). Alpha v integrins as receptors for tumor targeting by circulating ligands. *Nature Biotechnology, 15*(6), 542–546. https://doi.org/10.1038/nbt0697-542

[34] Tan, T., Wang, Y., Wang, J., Wang, Z., Wang, H., Cao, H., Li, J., Li, Y., Zhang, Z., & Wang, S. (2020). Targeting peptide-decorated biomimetic lipoproteins improve deep penetration and cancer cells accessibility in solid tumor. *Acta Pharmaceutica Sinica B, 10*(3), 529–545. https://doi.org/10.1016/j.apsb.2019.05.006

[35] Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology, 6*(3), 197–208. https://doi.org/10.1038/nrm1589

[36] Disfani, F. M., Hsu, W. L., Mizianty, M. J., Oldfield, C. J., Xue, B., Dunker, A. K., Uversky, V. N., & Kurgan, L. (2012). MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics (Oxford, England), 28*(12), i75–i83. https://doi.org/10.1093/bioinformatics/bts209

[37] Thundimadathil, J. (2012). Cancer treatment using peptides: current therapies and future prospects. *Journal of Amino Acids, 2012*, 967347. https://doi.org/10.1155/2012/967347

[38] Lau, J. L., & Dunn, M. K. (2018). Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorganic & Medicinal Chemistry, 26*(10), 2700–2707. https://doi.org/10.1016/j.bmc.2017.06.052

[39] Uhlig, T., Kyprianou, T., Martinelli, F. G., Oppici, C. A., Heiligers, D., Hills, D., Ribes Calvo, X., & Verhaert, P. (2014). The emergence of peptides in the pharmaceutical business: From exploration to exploitation. *EuPA Open Proteomics, 4*, 58–69. https://doi.org/10.1016/j.euprot.2014.05.003

[40] Tao, Z., Shi, A., Lu, C., Song, T., Zhang, Z., & Zhao, J. (2015). Breast Cancer: Epidemiology and Etiology. *Cell Biochemistry and Biophysics, 72*(2), 333–338. https://doi.org/10.1007/s12013-014-0459-6

[41] Joyce, J. A., & Pollard, J. W. (2009). Microenvironmental regulation of metastasis. *Nature Reviews Cancer, 9*(4), 239–252. https://doi.org/10.1038/nrc2618

[42] Teicher, B. A. (2006). Tumor models for efficacy determination. *Molecular Cancer Therapeutics, 5*(10), 2435–2443. https://doi.org/10.1158/1535-7163.MCT-06-0391

[43] Krall, N., Pretto, F., Decurtins, W., Bernardes, G. J., Supuran, C. T., & Neri, D. (2014). A small-molecule drug conjugate for the treatment of carbonic anhydrase IX expressing tumors. *Angewandte Chemie International Edition, 53*(16), 4231–4235. https://doi.org/10.1002/anie.201310709