SEMMELWEIS EGYETEM
DOKTORI ISKOLA

Ph.D. értekezések

3086.

IVÁNCZI MÁRTON

A gyógyszerészeti tudományok korszerű kutatási irányai

című program

Programvezető: Dr. Antal István, egyetemi tanár

Témavezető: Dr. Mándity István, egyetemi docens

Dr. Balogh Balázs, egyetemi adjunktus

DATABASE OF DRUG-CONJUGATED PEPTIDES AND MOLECULAR DYNAMICS SIMULATIONS OF CELL-PENETRATING PEPTIDES

PhD thesis

Márton Ivánczi

Semmelweis University Doctoral School

Pharmaceutical Sciences Division

Supervisor: István Mándity, PhD
Balázs Balogh, PhD

Official reviewers: Arash Mirzahosseini, PhD
Csaba Hetényi, PhD

Head of the Complex Examination Committee: Romána Zelkó, DSc

Members of the Complex Examination Committee: Zoltán Mucsi, PhD
Zoltán Gáspári, PhD

Budapest
2024

Table of Contents

List of Abbreviations

AMBER	Assisted Model Building with Energy Refinement
AMOEBA	Atomic Multipole Optimized Energetics for Biomolecular Applications
AMP	antimicrobial peptide
CAS RN	Chemical Abstracts registration number
CG	coarse-grained
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CPP	cell-penetrating peptide
ELBA	electrostatics-based
FF	force field
GPU	graphical processor unit
GROMOS	Groningen Molecular Simulation
GUI	graphical user interface
HIV	humane immune deficiency virus
HTML	Hypertext Markup Language
HTTPS	Hypertext Transfer Protocol Secure
JSME	JavaScript Molecular Editor
LIME	lipid intermediate resolution model
MAO B	monoamine oxydase B
MCoTI-II	*Momordica comhinchinensis* trypsin inhibitor II
MD	molecular dynamics
MM	molecular mechanics
NPT	constant substance, pressure, and temperature
NVT	constant substance, volume, and temperature
OPLS	Optimised Potential for Liquid Simulations
OPM	Orientations of Proteins in Membranes
PHP	Personal Home Page (hypertext preprocessor)
PNA	peptide nucleic acid
POPC	palmitoyl-oleoyl-phosphatidylcholine
POPE	palmitoyl-oleoyl-phosphatidylethanolamine
pVEC	penetrating vascular endothelial-cadherin

RCSB PDB	Research Collaboratory for Structural Bioinformatics Protein Database
RDBMS	relational database management system
RMSD	root mean square deviation
SDK	Shinoda-DeVane-Klein
SFTI-1	sunflower trypsin inhibitor 1
SIRAH	South American Initiative for a Rapid and Accurate Hamiltonian
Slipids	Stockholm Lipids
SPC	single-point charge
SQL	Structured Query Language

1. Introduction

1.1. Cell-penetrating peptides

1.1.1. Definition and features

Membrane-active peptides are oligopeptides and short polypeptides, usually containing 8 to 30 amino acid residues, which can translocate through biological membranes. There are two main groups: antimicrobial peptides (AMPs) and cell-penetrating peptides (CPPs). AMPs are part of the innate immune response as potent, broad-spectrum antibiotics acting through the destabilisation of the plasma membrane of the pathogens. This study presents the other group, the CPPs. They can translocate into live cells and cell organelles and cause lasting damage only in extremely high concentrations so that they can deliver several compounds, for example, drug molecules, to intracellular targets or across the blood-brain barrier or epithelia. [1] [2] [3]

The first peptide identified as a CPP is the Tat, a part of a regulatory protein isolated from the human immune deficiency virus (HIV). [4] The next one was the penetratin. [5] [6]

The interest in CPPs has increased in the last two decades because of their versatility. [7] The first CPP database was published in 2012 and became obsolete in a few years. [8] The CPPsite 2.0 (https://webs.iiitd.edu.in/raghava/cppsite/index.html) contains 1699 unique peptides described as experimentally validated CPPs, and new structures are still being uploaded. [9]

These databases are essential source of information regarding CPPs because they not only contain information on sequence, but also subcellular localisation, physicochemical properties and uptake efficiency. The above mentioned revision of the database has completed it with further including knowledge of the model system, cargo information, chemical modifications and predicted tertiary structure. The CPPsite 2.0 is now considered to be the most complete dataset and as such, it is frequently used for building predictive models, especially in case of machine learning methods. [10]

CPP databases were also proved to be useful in planning preclinical experiments and in some cases, also clinical trials by selecting peptides those were appropriately identified previously. The most significant limitation of CPPsite 2.0 is the lack of information about

relevant pharmacology related data for example plasma half-life, bio-distribution and toxicity. [11]

1.1.2. Classification of CPPs

CPPs can be classified depending on the chain type, the origin, the configuration and the physicochemical properties. Most of them are linear. Nevertheless, some cyclic peptides have a peptide bond between the N and C terminals of the backbone (Figure 1. a). [9] [12] One of the main advantages of cyclic peptides is the much slower biodegradation. [13]

More than half of the CPPs are synthetic, so the sequence is artificial and often repetitive. The second largest group is the natural; they are parts of proteins. The chimeric CPPs, a minor group, are connected by two natural peptide sequences. Another small group is the design peptides: they are of natural origin but modified artificially (Figure 1. b). [14]

Most CPPs are built of natural α-L amino acids, but some synthetic peptides contain D residues, even D and L mixed or other modified amino acids (Figure 1. c). The modified amino acids can be α-L with unnatural side chains, β amino acids even with two side chains, or anything else that can form peptide bonds. [9]

The two most important features of the amino acid sequence are polarity and charge. The primary amphipathic peptides have hydrophilic and hydrophobic residues separated in the series. The secondary amphipathic peptides have residues with different polarity mixed; usually, they have a polar and an apolar side in a helical conformation. The third important group is cationic; these peptides contain more positively charged side chains. [7] [15]
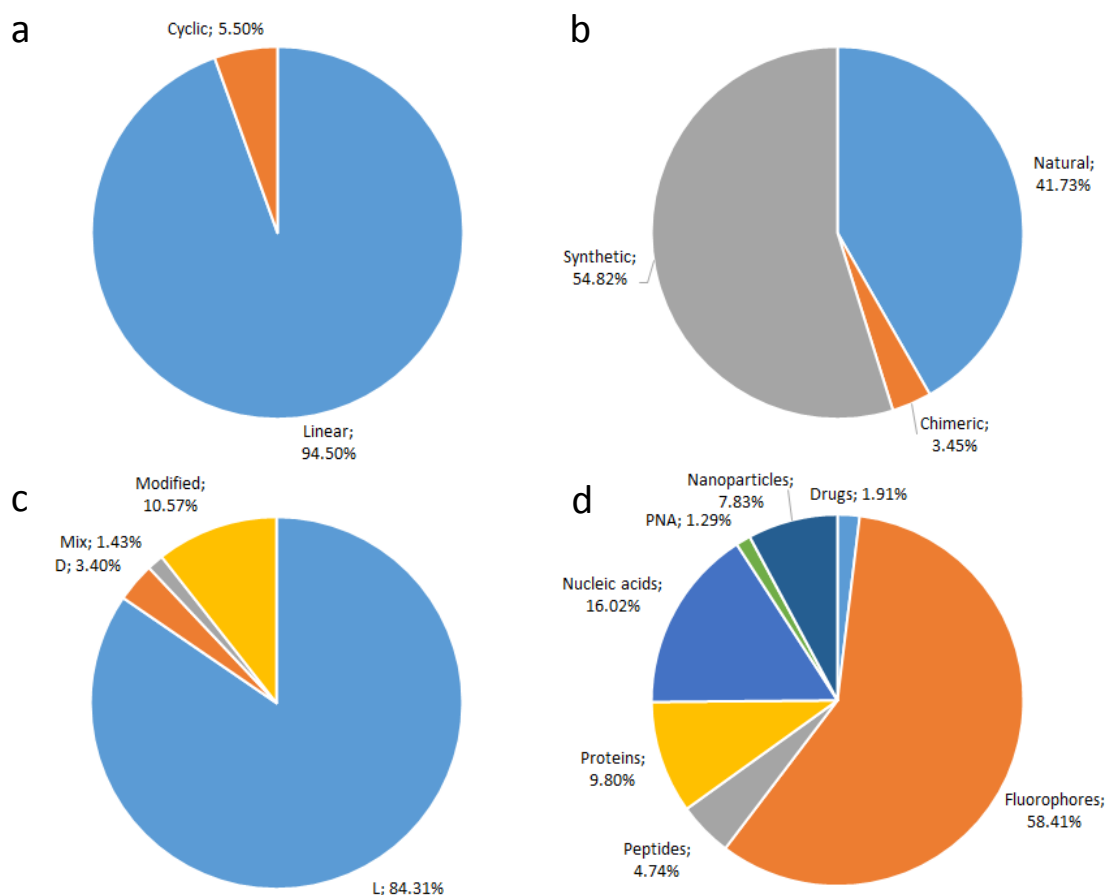
Figure 1. Statistics about CPPs: backbone type (a); origin of the peptide sequence (b); configuration of the amino acids (c); transported compound as a conjugate (see below) (d)

1.1.3. Mechanisms of the penetration

Several possible mechanisms of penetration are described in the literature. CPPs may enter the cells via all usual types of endocytosis: phagocytosis, macropinocytosis, clathrin-mediated endocytosis, caveolae or constitutive endocytosis (Figure 2. a-e). The energy-independent mechanisms are direct translocation, membrane thinning, the two kinds of pore formation, the inverted micelles and the carpet model. [16] The composition of the membrane probably modifies the mechanism: the same peptide can react differently with animal cells, bacteria or other membranes, even with different cells of the same organism. [17] In *in vitro* experiments, specific endocytosis inhibitors decrease the internalisation but usually do not prohibit it altogether. Low temperature, when the energy-independent mechanisms can not process, also only drops the speed of penetration. [18]

The less complicated mechanism is direct translocation: passive diffusion through the cell membrane. Linear peptides usually go in a perpendicular position during significant conformational changes. Hydrophobic or at least amphipathic character is needed for this process (Figure 2. f). [19] [20]

The membrane thinning is not only a single mechanism; the other energy-independent processes can also initialise it. When a cationic peptide approaches the membrane, it attracts the negative groups (phosphate) and represses the membrane lipids' positive groups (for example, choline). This makes an electrochemical bilayer that spreads to the membrane's other face and pulls it closer (Figure 2. g). [21]

Two different types of pore formation are described. In the first case, called toroidal pore, a usually helical peptide is in the middle of the small pore pre-alone while the membrane lipids turn around the peptide (Figure 2. h). In the second case, called barrel pore, more peptides form a circle closely next to each other in the membrane, and there is a much broader pore in this circle (Figure 2. i). [22]

Inverted micelles are in the middle of the membrane and contain one peptide or sometimes a few peptides covered by membrane lipids. The hydrophilic part of the phospholipids orientates to the peptide (Figure 2. j). [5] [23]

The carpet model occurs in a high CPP concentration and causes severe membrane damage, leading to cell death. The peptides are adsorbed on the surface of the membrane and make it burst, while the lipids form toroid aggregates stabilised by the amphipathic peptides (Figure 2. k). [17] [24]

The mechanisms of endocytosis need specific features of the peptide, too. Some peptides can bind to receptors, but receptor-mediated endocytosis is not exclusive. The peptide has to be adsorbed on the membrane surface to initialise endocytosis as an energy-independent mechanism. [19] The CPP may sometimes indicate vesicular uptake spontaneously after interaction with the membrane lipids. [25]

Endocytosis is usually combined with an energy-independent mechanism as an endosomal escape. Specific domains in the peptide sequence with triptofanes are essential for this process. [15] Early endosomal escape is vital to avoid lysosomal degradation. [26] In another mechanism investigated *in vitro* tests on live cells, CPPs could reach the lumen

of cell organelles via vesicular transport following the endocytosis, especially when the vesicles are coated with clathrin molecules. [27]
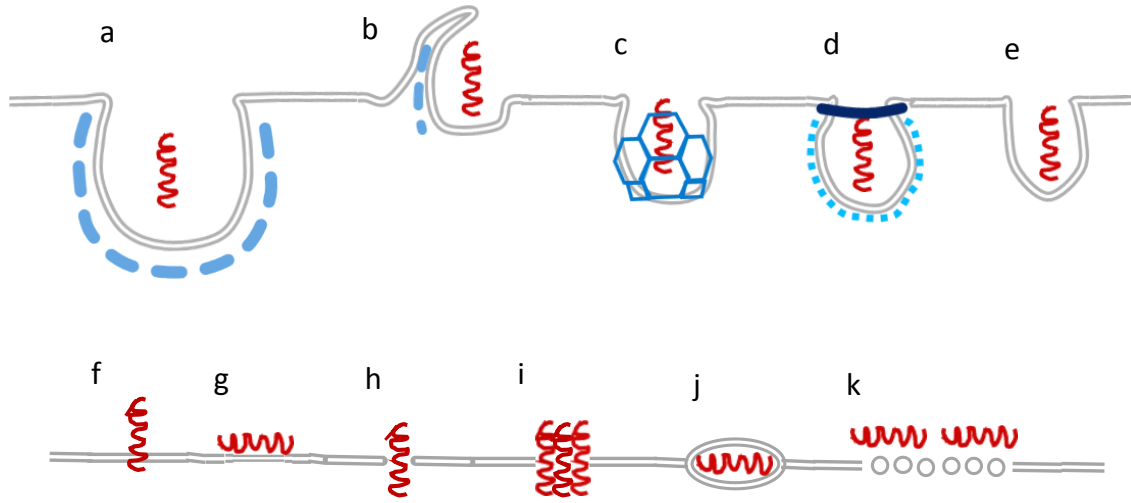

Figure 2. Active (a-e) and passive (f-k) mechanisms of the CPP penetration: phagocytosis (a); macropinocytosis (b); clathrin-mediated endocytosis (c); caveola (d); constitutive endocytosis (e); direct translocation (f); membrane thinning (g); toroidal pore (h); barrel pore (i); inverted micelle (j); carpet model (k) [28]

1.2. Penetratin and cyclic CPPs

Six peptides were selected in this study (Table 1). Three are linear: penetratin, 6,14-Phe-penetratin and dodeca-penetratin, while the other three are Kalata B1, SFTI-1 (Sunflower trypsin inhibitor 1), and MCoTI-II (*Momordica comhinchinensis* trypsin inhibitor II) are cyclic.

Table 1. Amino acid sequences of the investigated peptides [28]

Peptide	PDB ID	Sequence	Reference
Penetratin	1KZ0	RQIKIWFQNRRMKWKK	[29]
6,14-Phe-penetratin	1KZ2	RQIKIFFQNRRMKFKK	[29]
Dodeca-penetratin	1KZ5	RQIKIWFRKWKK	[29]
Kalata B1	1NB1	[CGETCVGGTCNTPGCTCSWPVCTRNGLPV]	[30]
SFTI-1	1JBL	[GRCTKSIPPICFPD]	[31]
MCoTI-II	1HA9	[SGSDGGVCPKILKKCRRDSDCPGACICRGNGYCG]	[32]

Penetratin (Figure 3. a) is the helix III homeodomain fragment of the Antennapedia homeoprotein isolated from *Drosophila melanogaster*. It is one of the most frequently investigated and the earliest known CPPs. It consists of 16 amino acids, 7 of which are cationic (three arginines, four lysins), so it has seven positive charge units because there are no anionic residues (Figure 4. a). The three aromatic side chains are essential: the Trp 6 and the Phe 7 forms a hydrophobic core with π-π stacking, and there can be an intramolecular π-cation interaction between Trp 14 and Lys 15. The other amino acids

are aliphatic hydrophobic or hydrophilic neutral. Its conformation is helical primarily; it can be α-helix or 3_{10} helices depending on the conditions, and short β-turn parts can be observed close to both terminals. [29] [33]

6,14-Phe-penetratin (Figure 3. b) is a modified version of penetratin: both triptofanes are replaced with phenylalanines, so the hydrophobic core and the intramolecular π-cation interaction are missing (Figure 4. b). Its charge in physiological pH is +7, like the native penetratin. Its secondary structure is mostly α-helical, with a β-turn at the C-terminal. According to *in vitro* experiments, its biological activity is significantly weaker than penetratin. [29] [33]

Dodeca-penetratin (Figure 3. c) is another penetratin analogue; four amino acids, the original Gln 8, Asn 9, Arg 11 and Met 12, are excluded (Figure 4. c). Despite the relevant residues' existence, there is no π-π stacking or π-cation interaction between the side chains. The conformation of this peptide is mostly 3_{10} helical. Its penetrating ability is as good as the penetratin's because the essential amino acids are included. [29] [33]

Kalata B1 (Figure 3. d) is a natural cyclic peptide, a self-defending material of *Oldenalia affinis*. It has antimicrobial and hemolytic effects and is stable against proteolytic mechanisms. It consists of 29 amino acids, and the six cysteines are the most important to determine the structure: three disulphide bonds between Cys 1 and 5, Cys 10 and 15, Cys 17 and 22 (Figure 4. d). The secondary structure of this peptide is dominated by a β-sheet and a β-turn; there are no helices. It contains a hydrophobic core and loops, which can form van der Waals binding to the fatty acid part of the membrane lipids. [30]

SFTI-1 (Figure 3. e) was isolated from *Helianthus annuus and* may have antitumor and anti-inflammatory abilities. It comprises 14 amino acids with a cyclic backbone and a disulphide bridge between Cys 3 and 11 (Figure 4. e). Several hydrogen bonds stabilise the secondary structure of two short β-sheet parts. It is a potent trypsin inhibitor and proteolytic stable. [31]

MCoTI-II (Figure 3. f) is a cyclotide isolated from *Momordica comhinchinensis*. It consists of 34 residues, one of the largest known CPPs. Like other cyclic CPPs, cysteines are critical: disulphide bridges between Cys 8 and 25, Cys 15 and 27, Cys 21 and 33 form the cyclic cysteine knot motif (Figure 4. f). This knot is very rigid and has a hydrophobic

core inside. A three-line antiparallel β-sheet dominates the conformation of this peptide, and there are two β-turns and a short α-helical part too. [32]
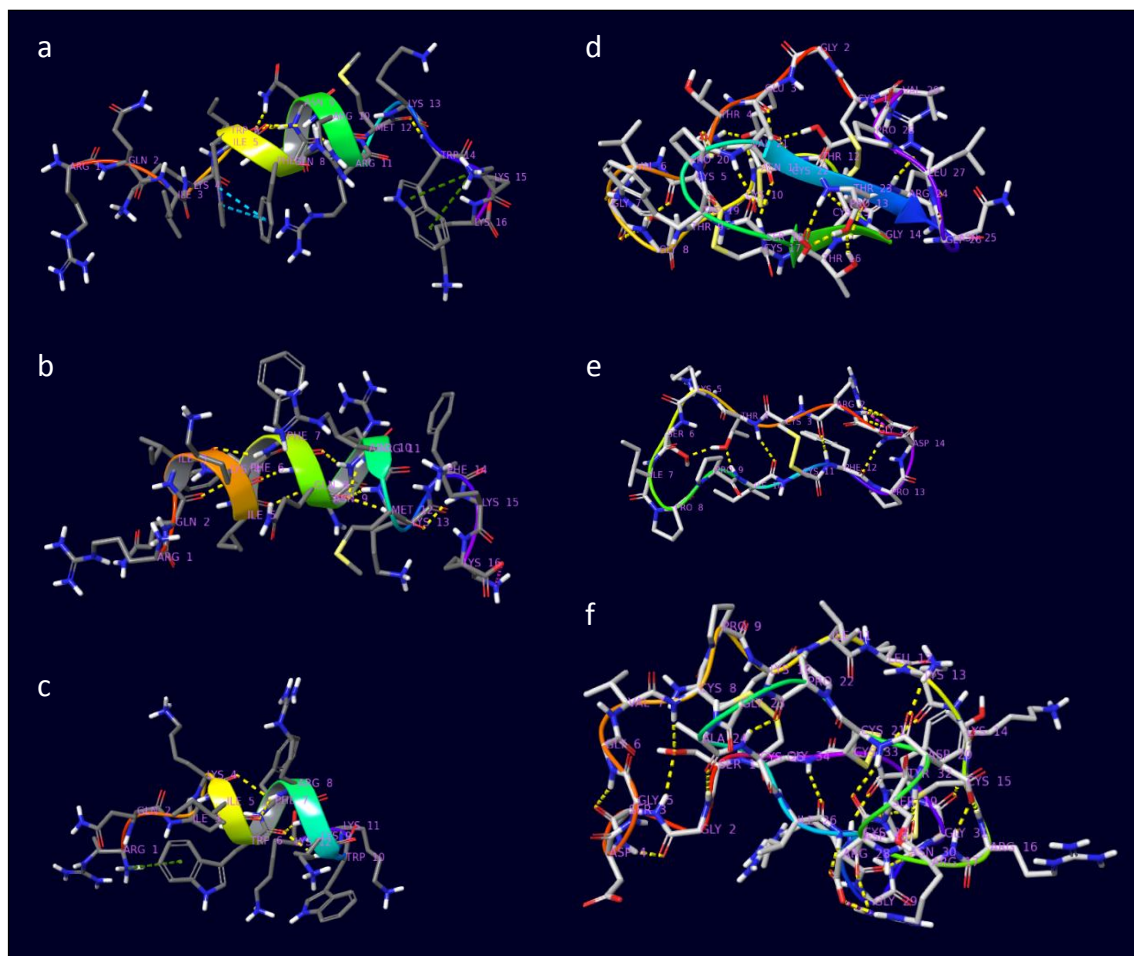

Figure 3. The 3D structures of the peptides gained by NMR from the Protein Data Base by the following PDB IDs: 1KZ0 - penetratin (a); 1KZ2 – 6,14-Phe-penetratin (b); 1KZ5 – dodeca-penetratin (c); 1NB1 – Kalata B1 (d); 1JBL – SFTI-1 (e); 1HA9 – MCoTI-II (f). The secondary structures are represented either as ribbons (for the α-helixes and β-sheets coloured bes (unsettled) and coloured by residue positions (from red to violet). Intramolecular interactions are represented as dashed lines: H-bond – yellow; π-π stacking – turquoise; π-cationic – dark green; salt bridge – purple. [28]
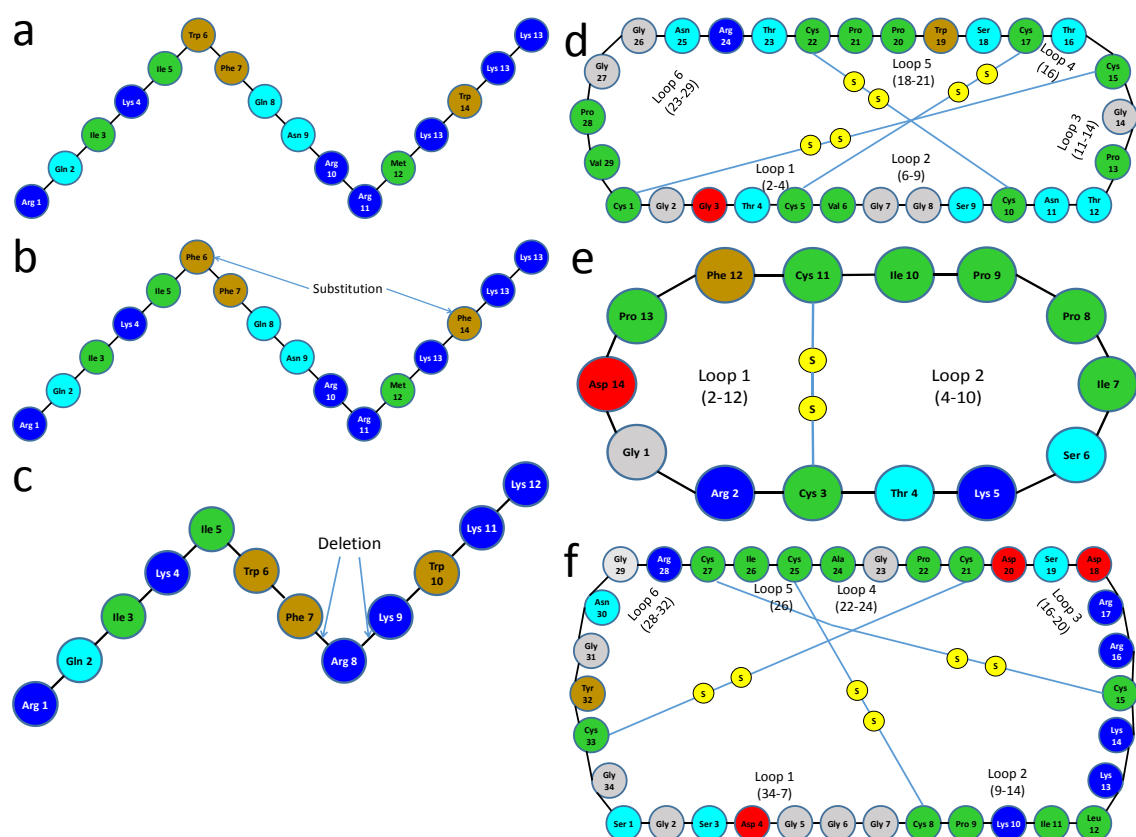
Figure 4. Amino acid sequence of the peptides: penetratin (a), 6,14-Phe-penetratin (b), dodeca-penetratin (c), Kalata B1 (d), SFTI-1 (e), MCoTI-II (f). The colours represent the type of amino acids: glycine–light grey, aliphatic hydrophobic–green, hydrophilic neutral–turquoise, cationic–dark blue, anionic–red, and aromatic–brown. The modifications in penetratin analogues and disulphide bridges with cyclic peptide loops are also marked.

1.3. Peptide-drug conjugates

The covalent binding of a small organic drug molecule to a peptide provides the peptide–drug conjugate. The small organic drug molecule generally has a prominent, strong pharmacological action. However, for these compounds, conjugation as a strategy is helpful because these compounds often suffer from some distinct drawbacks, for example, high toxicity or low tissue penetration. Moreover, their site-targeted or site-specific application might also offer substantial benefits. [34] [35] [36]

Careful design principles should be followed for constructing peptide–drug conjugates. First, the drug molecule should be carefully selected. It should have a functional group that can be attached to the peptide or the linker without losing bioactivity. Additionally, the linker has to be selected cautiously. Thus, the length, stability, release mechanism, and solubility must be considered. There are stimuli-

responsive and biodegradable linkers, which release the active drug molecule into the environment of, for example, the cancerous tissue with which the general toxicity of an anticancer agent can be lowered. [34] [35]

CPPs are capable of transporting into cells and even cell organelles and nucleus or transporting transcellular nearly all types of cargo: small molecules (drug compounds, diagnostics), macromolecules (DNA, RNA, PNA, proteins) and nanoparticles. [37] [24] Fluorophores are the most frequent. [9] In gene technology, CPPs are one of the most effective methods for inserting plasmids into cells. [14]

This study is about drug-conjugated CPPs, which our research group has already investigated. One of the first applications, conjugating paclitaxel to penetratin, helped reach the nucleus, thereby increasing the anticancer activity of the active ingredient. [38] Delivering isoniazid into mycobacteria with penetratin is under investigation. [39] One of the most recent applications of CPPs (but not small molecule conjugate) is orally active insulin; the effective preclinical development was published in 2021. [40]

Different strategies were published in the literature for the conjugation of the cargo, while specific macromolecules could be attached in a non-covalent manner and, for example, positively charged peptides with nucleic acids via charge-dependent complex formation with the CPP. [41] Proteins can be physically complex and covalently conjugated, too. [3]

However, small molecules mainly were bound covalently. [42] The chemistry to attach the small organic drug molecule to the peptides is a challenging task. The chemical bond between the two molecules should be orthogonal to other functional groups of the conjugate, ensuring the formation of only a single compound. The conjugation can be created by either the terminals of the peptide or through a side chain with the application of an appropriate functional group. A linker is frequently used to increase the distance between the peptide and the active ingredient or provide reversibility (capability for detachment under proper conditions). Almost in three-quarters of the cases, the connecting group is an amide, but nearby, all of the most known organic functional groups (ester, ether, sulphide, disulphide, sulphonamide, carbamate, amidine, guanidine, triazole, carbon-to-carbon) are represented. [43] [44] [45] [46] [17]

The current study selected three different types of drugs with known penetration issues. All three were known for their peptide conjugates in the literature, but only one was investigated as a complex before. [28]

Doxorubicin is a topoisomerase-2 inhibitor cytotoxic anticancer drug. Like most members of this pharmacological group, it causes severe side effects. Due to its hydrophilicity (Table 2), its excretion through the kidneys is relatively rapid. To make its pharmacokinetics even worse, some cancer cells have developed an active efflux system as protection. In animal tests, the peptide-conjugated form of doxorubicin has been excreted much slower; therefore, a much lower blood concentration was needed to have an equal therapeutic effect. [45] [47] [48]

Rasagiline is a specific, irreversible MAO-B (monoamine oxidase B) inhibitor used to treat Parkinson's disease; it has to get through the blood–brain barrier. Experimentally confirmed that the drug attached to penetratin was more effective than its unconjugated form. [46] [49]

Zidovudine, a reverse transcriptase inhibitor antiviral drug, has been developed to cure HIV infection. The peptide conjugation can increase its specificity towards the infected cells, which reduces the side effects of this hydrophilic molecule (Table 2). [44]

Table 2. Partitioning of the investigated drugs [28]

Drug	Molar weight (g/mol)	Predicted logP*	Experimental logP
Doxorubicin	543.52	-0.184	0.32 [50]
Zidovudine	267.24	-0.038	0.04 [51]
Rasagiline	171.24	2.447	2.462 [52]

* **Schrödinger Release 2023-1**: QikProp, Schrödinger, LLC, New York, NY, 2021.

1.4. Molecular dynamics

Molecular modelling, including molecular dynamics (MD), is essential in modern drug development and pharmaceutical research to design active compounds for targets and optimise biopharmacy for original and supergeneric medicines.

Molecular dynamics is a computational method for modelling the time course of molecular processes. It uses the tools of molecular mechanics (MM), consequently operates with force fields, and applies classical physics rules with Newton's and Coulomb's laws at the molecular level. Mass points represent the atoms, the covalent

bonds by springs; the ions appear as charge points and the polar bonds as dipoles. So, it needs much less computational power than quantum mechanics. The covalent bonds and formal charges do not change during the simulation, so MD cannot model reactions. Delocalisation can be modelled very limitedly because it does not calculate with the exact electron structure.

1.4.1. Force fields

The Optimised Potential for Liquid Simulations (OPLS) is the most commonly used MM force field (FF) and is also suitable for MD. In this model, the nonpolar hydrogen atoms belong to one unit with the carbon atom to which they are bonded. The OPLS3e version can also handle the nonbonding electron pairs and the halogen bonds. The medium can be an implicit or explicit solvent model. The implicit model includes dielectric constant and density. Explicit models, what we used, contain individual molecules of the water, membrane and any other components. [53] [54] [55] [56]

Another popular (FF) is CHARMM, developed directly for MD simulations. It has particular parametrisation for small molecules, biopolymers and membranes focusing on biophysical properties. AMBER is a widely used FF for MD simulations, too; it has parametrisation for druglike molecules and macromolecules but for a limited set of lipids; nevertheless, it can handle complex lipid membranes. Slipids has taken over many parameters from CHARMM and AMBER and has the most extensive collection of steroids. GROMOS is developed for lipid-protein simulations and has parametrisation for more unique (for example, bacterial) lipids. [57]

Polarisable models are a recent innovation; they modify the partial charges during the simulation due to the changing sterical and environmental effects. The most important types are the Drude oscillator model, the FlexQ method and AMOEBA. They have been tested on a limited lipid set and need more computation than standard FFs. [57]

1.4.2. MD methods

Various unique methods of MD are developed: atomistic MD, metadynamics, umbrella sampling, replica exchange methods and coarse-grained models. Atomistic MD is the standard method, which calculates the complete molecular structure without additions or simplifications, so it must calculate with many interactions while investigating contacts and clashes and their changes over time. However, the drastic

improvement in computational capacities, especially using graphical processor units (GPU), allowed us to apply all-atom calculations on hundred nanoseconds or even microseconds time scales. [58] [59]

Metadynamics operates with additional forces to turn over energetic barriers for simulating membrane penetration, folding or peptide aggregation. It lets us observe the system's unstable arrangements, but choosing the appropriate energy bias is challenging.

Umbrella sampling contains multiple parallel simulations on the same system with different restraints to generate collective variables. The collective variable describes the importance of system parameters, for example, distances, angles and dihedrals.

The replica exchange method runs parallel simulations of the same system in different physical conditions (especially temperatures) and compares them. The collective variables need not be given manually.

Coarse-grained (CG) models reduce the complexity of the molecules; more atoms are represented by one particle, so it needs less computation than atomistic MD and is even suitable for high-throughput processes. One CG bead usually contains three to six heavy atoms, so a typical lipid molecule is represented by eight to fourteen beads. The parametrisation should represent atomistic interactions (structure-based, bottom-up strategy) and thermodynamic properties (thermodynamic-based, top-down strategy). One of the most important limitations is the lack of some common interactions, especially hydrogen bonds. The Martini model is the most commonly used CG method, has parametrisation for most biomolecules, and is compatible with the most important modelling programs. SDK model has been developed for membrane simulations, making proteins less accurate. SIRAH force field has been developed to model proteins and nucleic acids. ELBA model is the most accurate in electrostatic interactions and lipid-water interactions. Solvent-free models, for example, Dry Martini, PLUM and LIME, exclude the solvent interactions and incorporate them into the potentials between the CG beads. Supra-CG modes simplify the molecules even more for long-timescale simulations of extensive systems. [57]

1.4.3. Examples

Several studies have aimed at modelling the penetration mechanism by applying molecular dynamics. The following section will address a few examples concerning CPP penetration with MD. [60] [61] [62]

Lensink *et al.* took one of the first studies on simulating penetratin in 2005. The peptide did not translocate during the simulation utilising Gromacs; several interactions were observed between the peptide and the membrane lipids. [63]

Herce and Garcia published a critical study about MD simulations of CPPs in 2007. They found that the HIV Tat peptide translocates spontaneously, mainly via transient pores, while the positively charged side chains interact with the phosphate groups of the membrane lipids. [64] Their next article in 2009 is about pore formation by arginine-rich peptides. [65]

In a study published in 2015 by He *et al.*, the role of the membrane tension was confirmed in simulations with the coarse-grained MD method. The results showed that polyarginine in low concentrations was only adsorbed on the membrane surface, and translocation in higher concentrations was completed in less than 100 ns. The mentioned mechanism was direct translocation after membrane thinning. [66]

In a study by Bennett 2016, the CM15 antimicrobial peptide did not translocate; it only entered the membrane and reached its equilibrium point inside the lipid bilayer. [67]

Another study by Alaybeyoglu *et al.* 2016 successfully simulated the direct translocation of pVEC (an amphipathic CPP) using the steered MD method. The trajectory also revealed that the penetration was led by the N-terminal amino acid of the peptide, while the cationic side chains interacted with the phosphatide groups, enhancing the adsorption on the membrane. [68]

The importance of transmembrane electric potential *in silico* was also demonstrated in MARTINI coarse-grained MD and metadynamics simulations, in which the translocation of arginine-rich design peptides was successfully promoted by the introduction of the electrostatic gradient in a study published in 2018. [69]

Ulmschneider, in 2017 and 2018, investigated the mechanism of antimicrobial peptides using molecular dynamics simulations. The mechanism could be direct

translocation, transient pore-forming or carpet model, depending on the membrane composition and the physical conditions. The complete process needed microseconds while ions translocated together; the peptides and lipids did a flip-flop. [70] [20]

The energetic aspect of the transmembrane penetration of peptides was studied by Yao *et al.* in 2019. The formation of salt bridges and parallel hydrogen bonds between guanidium and phosphate groups ensures more than half of the energy needed for the membrane translocation; the bonds between ammonium and phosphate groups give less energy. Additionally, the hydration of the peptide inhibits the penetration. [71]

Several works have been published in MD simulations on CPPs in recent years, summarised in a review by Reid *et al.* in 2019. This study categorises all unique methods and tools used, such as atomistic MD, metadynamics, umbrella sampling, and replica exchange methods, to name a few. However, despite continuous developments, no universally applicable method has been developed, and many technical problems still need to be assessed. [58]

Tran *et al.* 2021 used MD simulations to validate and prioritise the penetration of CPPs generated by artificial intelligence. As a result, a novel CPP sequence named Pep-MD was *de novo* identified by computational methods and then synthesised, and its penetration potential into living cells was demonstrated by *in vitro* experiments. [72]

In a recent study, Gimenez-Dejoz and Numata demonstrated that the penetration mechanism for CPPs containing unnatural amino acids depends on the lipid composition of the membrane utilising steered MD simulations. The side chains of α-aminoisobutyric acid facilitated hydrophobic interactions inside the membrane, while the lysine side chains formed electrostatic interactions with the phosphatide groups in the outer layers. [73]

2. Objectives

2.1. Peptide conjugate database

The covalent conjugation of a drug compounds with a peptide can improve its pharmacokinetic properties and cell specificity. There was a number of successful examples on this field in the past and there has been a significal increase in their numbers recently, partially because of the relative ease of their chemical synthesis. Considering their importance in biomedical and pharmaceutical applications, constructing a comprehensive peptide–drug conjugate database is particularly important.

However, according to our knowledge no such public database has existed before. It should be noted that there has been a database available for cell-penetrating peptides that contains some information on peptide–drug conjugates. Nonetheless, this source focuses on data collected only for cell-penetrating peptides.

We proposed to collect the synthesised peptide-drug conjugates from the literature and organise them into a new database. We intended to create a comprehensive collection including drugs and other relevant organic compounds connected to any peptides, not only CPPs. We applied only a few but still strict restrictions: the connection must be covalent, and the peptide contains less than 100 residues mostly focusing on natural amino acids, artificial peptide analogues were excluded.

The database should include a primary identifier for all molecules and properties, like the name and pharmacological class of the drug, the sequence and name of the peptide (if aviable), the exact mode of conjugation. We intended to share the database for all researchers who may be interested in the topic. This source of information freely accessible through a web-based interface with multiple search options.

2.2. MD simulations

We aimed this study to ascertain the effect of the covalently conjugated drug molecules on the peptide. We ran atomistic MD simulations to model the changes of secondary and tertiary structure of three linear (penetratin, 6,14-Phe-penetratin, dodeca-penetratin) and three cyclic (SFTI-1, Kalata B1, MCoTI-II) CPPs and also their conjugation with doxorubicin, rasagiline and zidovudine. [28]

With these simulations, we intended to explore how these peptides behave near the membrane surface and inside the membrane and how this is affected by certain modifications to the amino acid sequence and the conjugation. We have also attempted to investigate whether such a relatively simple explicit system can be used to model membrane penetration.

The target of doxorubicin is located in the nucleus, and zidovudine has to enter the infected cells. Rasagiline has to translocate across the blood-brain barrier, which certain peptides could facilitate. Applying CPPs could improve the delivery of these active ingredients in lower doses with fewer side effects.

We utilised explicit solvent and membrane models during the MD simulations to study the behaviour of the conjugates in water, inside the membrane, and on the surface of the membrane. We were looking for similarities and differences between the unconjugated peptides and their conjugated forms. We intended to determine if the conjugated cargo helps or inhibits the penetrating ability of the peptides.

Despite the high number of studies about CPPs, not to many results with MD simulation of drug-conjugated CPPs have been published yet, and comprehensive studies with more different peptides and small molecules are rare. [28]

The original aim was to model the internalisation process, but despite our effots it did not occur spontaneously. Therefore, we focused on some aspects that might be involved in the penetration leading to the internalisation, such as the formation of intermolecular interactions, orientation and conformational changes of the penetratin and analogues, and the influence of the conjugated small molecule on the process.

Despite the lack of penetration, based on these interactions and changes we may be able to extrapolate to the penetrating capability of these molecules.

3. Methods

3.1. Database building

3.1.1. Data collection and processing

The entries in the ConjuPepDB (https://conjupepdb.ttk.hu/) result from a rigorous curation process. We painstakingly collect and annotate covalent peptide–drug conjugates from the literature, ensuring the utmost data quality. Our search in the CAS database (https://www.cas.org/) via the SciFinder[n] tool (https://scifinder-n.cas.org/) was conducted using various keywords such as 'peptide conjugate', 'peptide conjugated' and 'peptide conjugations' combined with 'drug'. A conjugate was included in our database only if it contained a peptide and a small molecule connected through a covalent bond. In most cases, the latter component was a drug or imaging compound, but other pharmacologically and biologically relevant conjugates were also discovered. Each literature entry was meticulously evaluated, and only those with one or more relevant substances assigned were included. Peptide conjugates and peptides with protecting groups were not included. Several articles were omitted because no valid peptide conjugate structures were set in the CAS database (small molecules only peptide only or no structures at all) despite indication in their title or abstract. Other sources, such as Elsevier's Reaxys database (https://www.reaxys.com/), were also searched. However, the number of articles with assigned peptide conjugates was significantly lower (this is the reason why no cross-references were included in our database). [43]

Conjugate structures are referred to by their CAS registration numbers and the small molecule parts. The amino acid sequences of the peptide part are also included. These were taken from the CAS database (if they were specified) or added manually recognised from the 2D structures. The conjugated small molecules were identified either by name, as referred to within the articles, or manually by their 2D structure. The conjugate structures and the corresponding small molecules were exported from the CAS database in MDL mol format. [43]

According to the type of the covalent bond, each selected conjugate was classified into one of the following groups: amide, urea, triazole, sulfonamide, sulfide, guanidine, ether, ester, disulfide, carbamate, carbon-to-carbon (C–C), azide, ammonium, amidine. Based on the indication of the drug conjugates, they were also grouped into types of biomedical

application: anticancer, anti-inflammatory, neurological, and antimicrobial, while non-drug conjugates were categorised as 'other'. It's important to note that these non-drug conjugates, despite not being the primary focus, have significant practical applications concerning organic synthetic strategy and planning for organic chemists, thereby adding to the overall value and relevance of the ConjuPepDB. [43]

3.1.2. Database design and implementation

Information about the peptide–drug conjugates is stored in several relational tables connected via parent-child relationship entity. Apache HTTPS server 2.4 with MySQL server 5.7 in the back end was used to host the ConjuPepDB database. A web graphical interface allows users to view and interact with that data. This dynamic interface uses PHP 7.2, HTML5, CSS, and JavaScript technologies. Bootstrap3 and jQuery libraries are also used to make a responsive, mobile-first front end (Figure 5.). [43]

JpGraph library (https://jpgraph.net/) was used to plot charts. Jmol (http://www.jmol.org/) is used for rendering the 3D model of the peptide–drug conjugates, whereas JSME Molecular Editor v2017-02-26 was employed for depicting the 2D model of small molecules present in a conjugate. [74] Chemical search is implemented using RDKit, an Open-source toolkit for cheminformatics (http://www.rdkit.org). ConjuPepDB is built upon relational database management system (RDBMS) technology for easy retrieval and scalability. [43]

3.1.3. Structure search

ConjuPepDB allows users to perform different structure searches on the small peptide molecule–drug conjugates. Users can draw the query structures using JSME, a free molecule editor written in JavaScript. [74] RDKit toolkit was used to generate the chemical fingerprints and perform different types of chemical searches in the database (Figure 5). [43]

Three types of chemical search were implemented. An exact match returns whether or not two molecules are the same. Substructure search returns whether or not the query molecule is a substructure of the target molecule. If the molecules are represented as a 2D graph, atoms are indices, and their bonds are vertices. Then, substructure search can be approached as a subgraph isomorphism problem, where the task is to find out whether, of two given graphs, A and B, A contains a subgraph that is isomorphic to B. Similarity

search returns whether or not the Tanimoto similarity between two molecules (fingerprints) exceeds the cut-off value. [43]
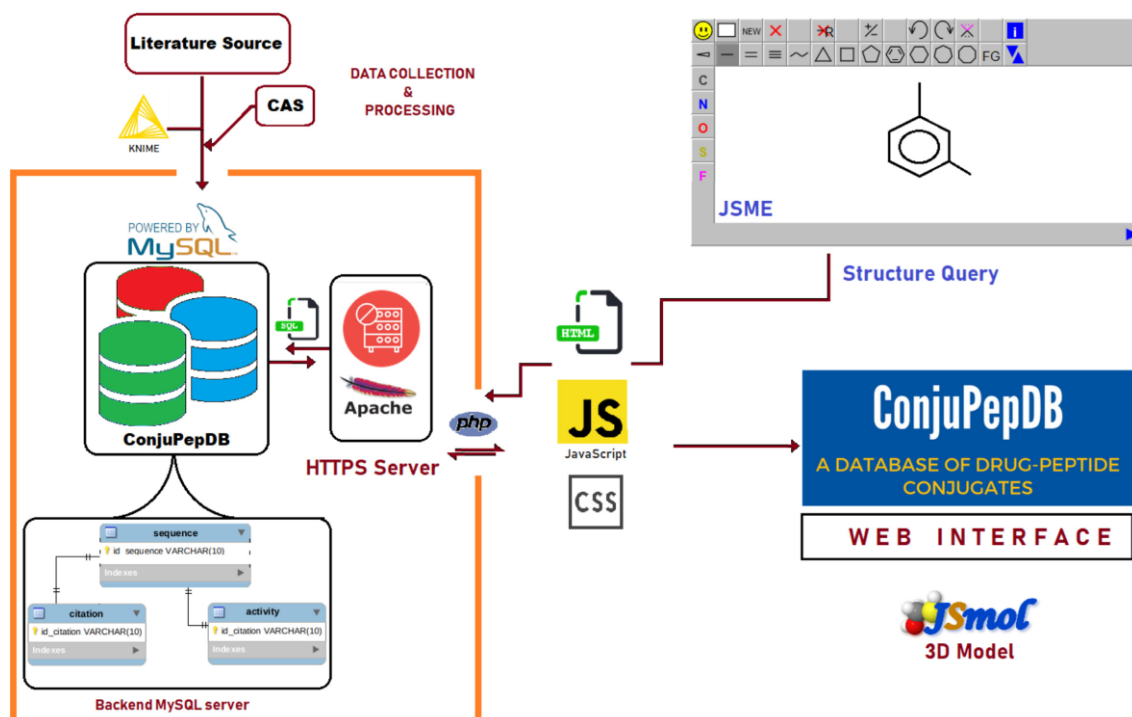
Figure 5. The schematic diagram of data collection and processing workflow and layout of information retrieval in ConjuPepDB. Taken from ConjuPepDB: a database of peptide–drug conjugates [43]

3.2. MD simulations of peptide conjugates

The Maestro graphical user interface (GUI) of the Schrödinger molecular modelling package was used in this study (**Schrödinger Release 2022-3**: Maestro, Schrödinger, LLC, New York, NY, 2022). The entire workflow is summarised in Figure 6. [28]
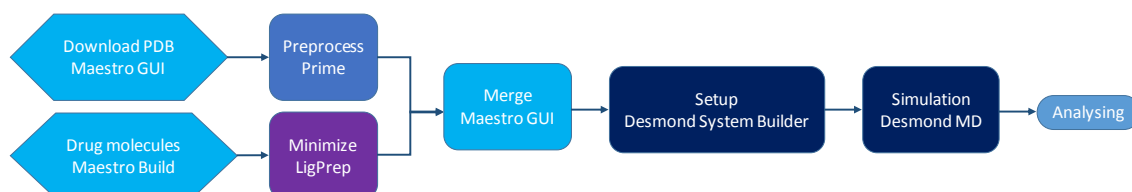
Figure 6. Schematic chart of the workflow of modelling the peptide conjugates [28]
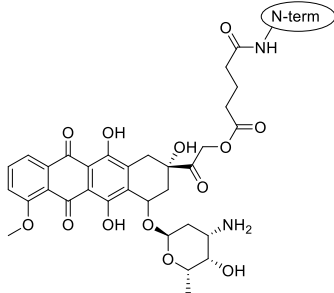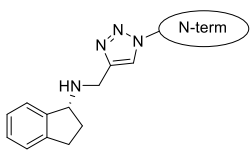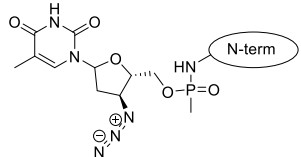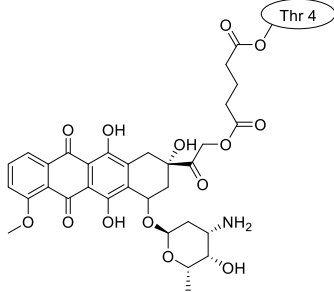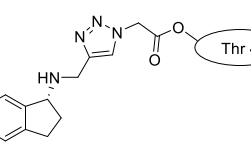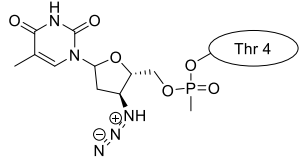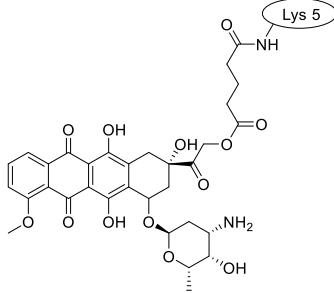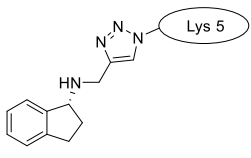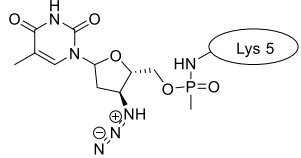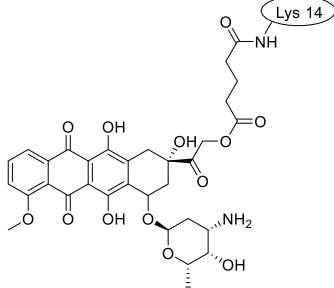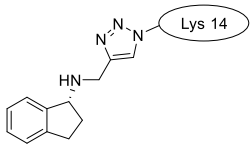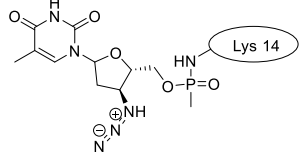
3.2.1. Preprocessing of peptides and conjugates

The peptide structures were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Database (RCSB PDB, http://rcsb.org) based on the identifying code (PDB ID), see Table 1 and imported into the project. [75] All entries were derived from NMR spectroscopy with multiple structures always working with the

first member of the most significant cluster. Each structure was analysed with Schrödinger Protein Preparation Wizard, and the preprocess option was used to cap the termini of the linear peptides (the N-terminal was acetylated, and the C-terminal was transformed into an N-methyl-amide group). [76] (**Schrödinger Release 2022-3:** blood–brain reparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2022; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2022). [28]

Drug molecules were drawn by the 2D sketcher module of the Maestro GUI and minimised by the LigPrep module (**Schrödinger Release 2022-3**: LigPrep, Schrödinger, LLC, New York, NY, 2022). The conjugates were made by merging the optimised (most probable protonation, tautomerisation and conformation) small molecule and selected peptide structures, and the linkers were added by the 3D Builder application within the interface: a glutaryl group for doxorubicin, a triazole ring for rasagiline, and phosphoric amide for zidovudine. In the case of the three linear peptides, the conjugations were formed through their N-terminal, while an appropriate amino acid side chain was utilised in the case of the cyclic peptides (Lys 5 of SFTI-1, Lys 14 for MCoTI-II, and Thr 4 for Kalata B1, table 3). [44-46] The simulations included the six unconjugated peptides, all possible combinations of the three drugs, and the six peptides, so in all 24 different peptides and conjugates were built. [28]

Table 3. Schematic representation of the investigated conjugates. [28]

	doxorubicin	rasagiline	zidovudine
penetratin and analogues			
Kalata B1			
SFTI-1			
MCoTI-II			

3.2.2. Setup and building of the systems

All MD simulations in this study were completed with the Desmond Molecular Dynamic software under Schrödinger (**Schrödinger Release 2022-3**: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2022. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2022). The run setup was assembled with the Desmond System Builder application under Maestro. Every simulation box contained one conjugate or peptide compound. Initially, we intended to study the conformational changes of the peptides and their complexes in water. Water

box simulations with the explicit single-point charge (SPC) water model were utilised to complete this. In these simulations, the peptides and conjugates were placed into an orthorhombic simulation box using the buffer method, where the medium spread of 10 Å was in every direction from the compound. [77] Sodium and chloride ions were automatically placed to statistically reach the isotonic (0.15 M) concentration within the system, and additional counter ions were added to neutralise the charge of the peptide or conjugate if needed. That is, the net charge of the system was reduced to zero.

In the case of runs with membrane, the unimolecular lecithin (palmitoyl-oleoyl-phosphatidylcholine, POPC) model was utilised for all simulations. Like the water box runs, the orthorhombic box with a 10 Å buffer was used to build the setup and assembles were automatically completed with water boxes on both the membrane's top and bottom. The SPC water model and ions were also added, as described before. The membrane components were added automatically, and the software OPLS3e was used for the final positioning of the molecules to avoid steric clashes. The OPLS-AA is an all-atom force field parameter for proteins and many general classes of organic molecules; therefore, no further parametrisation for the drug molecules is necessary. [53] [54] [56] According to the literature, this system is suitable for modelling peptides with membranes. [78] [28]

Initially, the automatic placement option within the "Set Up Membrane" dialogue box was used for the default position and orientation of the peptides. Most of our peptides can be found in the OPM (Orientations of Proteins in Membranes, https://opm.phar.umich.edu/) database; hence, the software associated with the membrane was placed using the conventions. The "Adjust membrane position" option was used if a membrane orientation change was needed.

We intended to study the initial steps of the penetration, so in the case of the "top" simulations, we expected to see spontaneous penetration, which was not observed. In this series of simulations designated as "top", peptides were manually positioned outside the membrane at approximately 10 Å distance from the surface, using the "Adjust membrane position" function. In the third set of simulations, we expected the peptides/conjugates to exit the intramembrane space, so in the runs assigned with the label "intra", the peptide/conjugate was positioned inside the membrane. Because of the relatively small size of the peptides/conjugates, all were fully submerged into the membrane.

Approximately equal distances were kept from the membrane's inner and outer surfaces. In the case of penetratin and its analogues, simulations were started with parallel if „top" and perpendicular if „intra" orientations of the chains relative to the membrane plane. (Figure 7.)

All 24 peptides and conjugates mentioned above were placed into three simulation boxes (water box, with POPC membrane, starting from the intra- and top positions) and an additional paralell intra-membrane system, so 73 systems were included in the study.
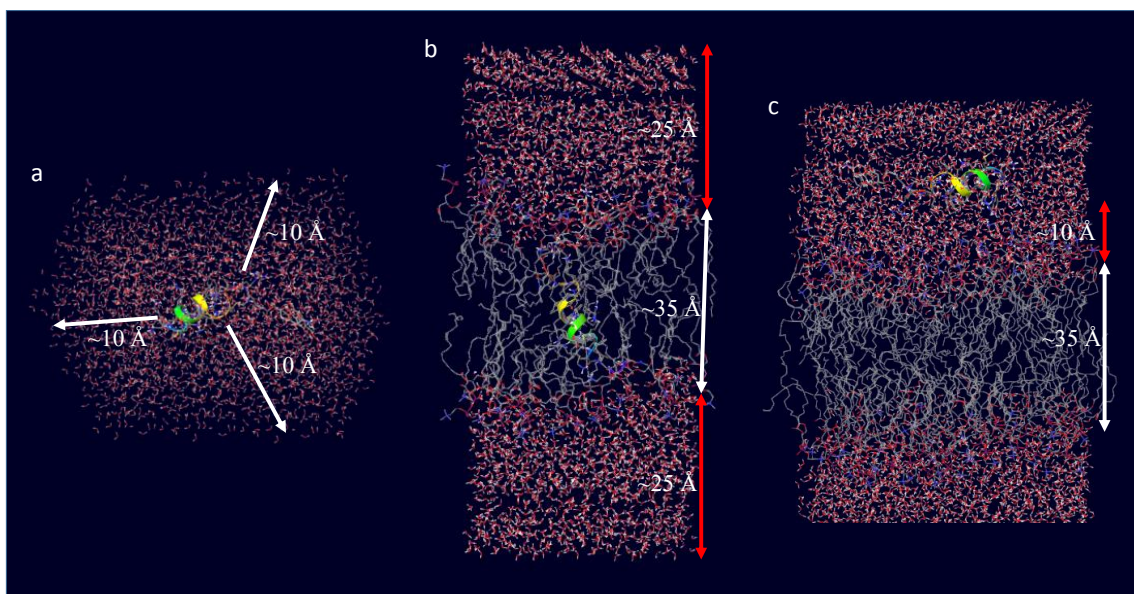
Figure 7. The arrangements of the MD simulation systems: water box (a), starting from inside of the phosphatidylcholine (POPC) membrane bilayer (b) and starting from the top of the phosphatidylcholine (POPC) membrane bilayer (c).

3.2.3. MD simulations

The completed setups were loaded into Desmond's Molecular Dynamics interface, where the simulations were initialised. Although the Desmond software supports several variants of the Amber, CHARMM and OPLS-AA force fields, Schrödinger only supports running Desmond MD simulations with the OPLS family of force fields. [56] All MD simulations began with the standard relaxation protocol, which also includes equilibration utilising the default settings: starting with a 12-ps-long NVT (constant substance, volume, and temperature) ensemble simulation at 10 K temperature, followed by two 12-ps-long NPT (constant substance, temperature, and pressure) ensemble simulation at 10 K temperature and 1.01325 bar pressure and, finally, a 24-ps-long NPT ensemble simulation at 300 K temperature and 1.01325 bar pressure. The program re-randomizes the velocity

of the atoms at the beginning of each stage, based on statistical thermodynamics. [79] [28]

Following the relaxation, all MD simulations were always carried out with NPT settings, where the pressure was 1.01325 bar, and the temperature was 300.0 K. [80] [28]

Simulations were completed on hardware with nVidia® GeForce GTX 1070 Ti 1683 MHz x2432 GPU under Linux Ubuntu. The 100-ns simulations took approximately 7–8 hours, while the 1000-ns simulations took a maximum of 70–80 hours. The recording interval of the trajectory was set to 100 ps in the case of the 100-ns simulations and 1000 ps for the 1000-ns simulations; that is, all trajectories contained 1001 frames (including before the first and after the last step).

All water box and intra-membrane systems simulations were run for 100 ns except two additional 1000 ns simulations, while the membrane-top systems were run for 1000 ns, so we completed 74 simulations.

3.2.4. Structure analysis

The Structure Analysis application of Schrödinger was used to evaluate the trajectories (**Schrödinger Release 2022-3**: Prime, Schrödinger, LLC, New York, NY, 2022). The most critical data is the amino acid alpha carbon atoms' root mean square deviation (RMSD), depending on the running time. All the RMSD values were calculated compared to the 0 ns geometry of the trajectories (after the relaxation and equilibration). If the RMSD does not change, the conformation is stable. In contrast, if the RMSD increases or decreases, the atoms move, and the system is not in equilibrium. In addition, if the included atoms are near their reference position, the RMSD is lower and high numbers show significant motion. [81] [82] [28]

The evaluation also included monitoring the change in the number of intramolecular hydrogen bonds within the peptides over time, which correlates with the changes in the secondary structure. Less intramolecular hydrogen bonds may indicate irregular, less stable conformations, while a higher number usually means a more organised and energetically stable folded structure. [28]

Comparing the Ramachandran plots of the peptides at different simulation times can also show structural differences. In the case of minor conformational movements during

simulation, the Φ and Ψ dihedrals were quite similar at the initial and final frames. However, the coordinates of the dominant conformations changed when significant changes were observed in the Φ and Ψ dihedrals on the Ramachandran plot.

Because of the software's limitations, we could not track all the interactions between the peptides-small molecules and the membrane during the simulations. Therefore, we manually counted the hydrogen bonds, salt bridges and π-cation interactions marked by the graphical interface in the final frame (at 1000 ns) within the trajectories of membrane-top simulations. The hydrogen bridge can span a maximum distance of 2.8 Å. The angle between the donor and the bridge is at least 120°, while the angle between the acceptor and the bridge is at least 90°. The maximum distance between salt bridges is 5.0 Å. Additionally, the maximum distance for π-cation interaction is 6.6 Å, with a maximum angle of 30°.More interactions mean more robust anchoring to the membrane; nevertheless, if the peptide reaches the interior of the membrane, it can form fewer interactions among the apolar fatty acid chains.

4. Results

4.1. Setting up ConjuPepDB, a database for peptide conjugates

Our research group has built and recently published a database of peptide-drug conjugates called ConjuPepDB. It contains over 1640 covalent conjugates of various conjugates of peptides (not only CPPs) and small molecules collected from 230 research articles. [43] It provides comprehensive information about a peptide conjugate, for example, 2D and 3D models, molecular properties, compound identifier, structural information including peptide sequence, covalent binding mode, conjugated small molecule structure, bibliography and other information. [43]

Here, we introduce ConjuPepDB, a database containing information on conjugates formed by a covalent bond between a peptide and a small organic drug molecule. The database includes details such as the Chemical Abstracts Service registration number (CAS RN) of the conjugate, the names of the peptide and the small organic drug molecule, the biomedical application, and the type of conjugation. Additionally, a chemical structure search for the small organic drug molecule is available, making it easier to explore the various conjugates (Figure 8. a-b). ConjuPepDB hopes to further increase the progress of this intensively developing field of drug research by providing a quick overview for interested non-specialists. More importantly, detailed information collected interests researchers from various fields, such as chemists, pharmacists, clinicians, or biologists. [43]

ConjuPepDB is crafted to be intuitive and user-friendly, offering several navigation options. Users can browse through conjugates, articles, structures, and biological activities. The interface provides robust querying capabilities, enabling easy retrieval of specific conjugates from the database. Users can perform simple searches as well as complex, chemical fingerprint-based substructure searches. Additionally, the responsive web interface ensures compatibility with devices of various screen sizes (Figure 8. c-e). [43]

Figure 8. Overview of different ConjuPepDB webpages. The layout of the browse conjugates table along with the filters (a), structures and related data of a conjugate entry (b), search options in ConjuPepDB (c), browse article table (d), single article page (e). [43]

4.2. 100 ns simulations in explicit water and membrane models

As mentioned, the penetratin structure and its analogues were much more flexible than the cyclic peptides.

The unconjugated penetratin and 6,14-Phe-penetratin remained α-helical primarily in the water model, but the shorter dodeca-penetratin could not maintain the helical structure to the end of the simulations. In membrane models, both penetratin and 6,14-Phe-penetratin mostly sustained the initial helical structures during the simulations, but dodeca-penetratin was uncoiled and became disordered. The unconjugated and conjugated forms of penetratin and its analogues did not show any fundamental differences in the water box or POPC membrane during the 100-ns simulations.

The conformationally constrained cyclic peptides in both conjugated and unconjugated forms did not show any noticeable changes during the 100-ns simulations in water or with the POPC membrane model because of their rigid structure (Figure 9.). The most significant movements were observed with the zidovudine conjugate of the MCoTI-II.
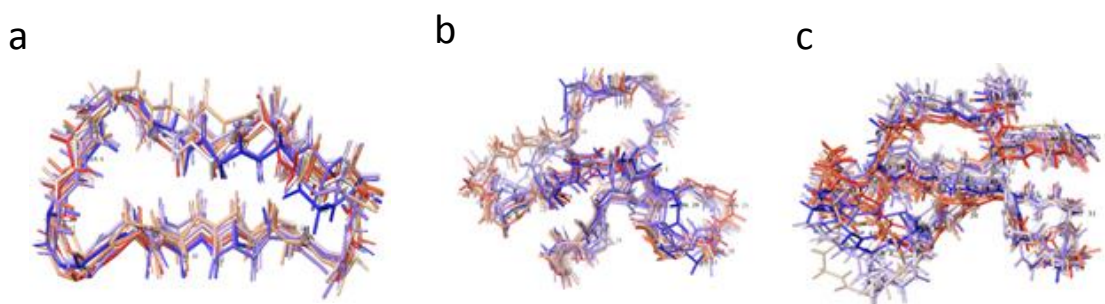
Figure 9. The displays of multiple frames of superimposed structures of the same peptide from the trajectories (snapshots), colour-coded by simulation time using a blue-red colour gradient. These 100-ns simulations were run in water boxes with the three cyclic CCPs SFTI-1 (a), Kalata B1 (b) and MCoTI-II (c).

4.3. Unconjugated penetratin and analogues

Penetratin moved into a perpendicular position during 1000 ns even if its starting position was parallel in the membrane. (Figure 10.)
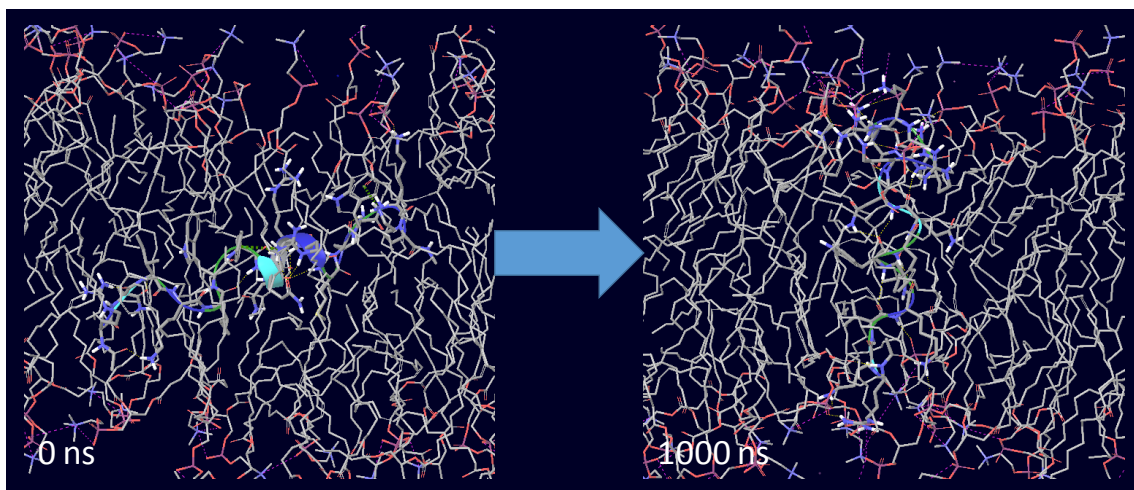
Figure 10. The positional change of penetratin in the 1000-ns MD simulation with POPC membrane model – the peptide moves from inside the bilayer, and the longest axis becomes approximately parallel with the membrane plane.

A complete structural rearrangement was observed with penetratin in the proximity of the surface of the POPC membrane model. At first, the helix was uncoiled and ceased to exist entirely and then slowly transformed into two-strand antiparallel β-sheets connected by a β-turn, laid to the surface of the membrane. The analysis of the last frame of the trajectory also revealed that four salt bridges and eleven hydrogen bonds were formed between the peptide and the membrane molecules, and no π-cation interactions were observed (Figure 11. a.). The 6,14-Phe-penetratin and the dodeca-penetratin remained helical primarily, with their N-terminal partially sinking into the POPC membrane (Figure 11. b-c). [28]
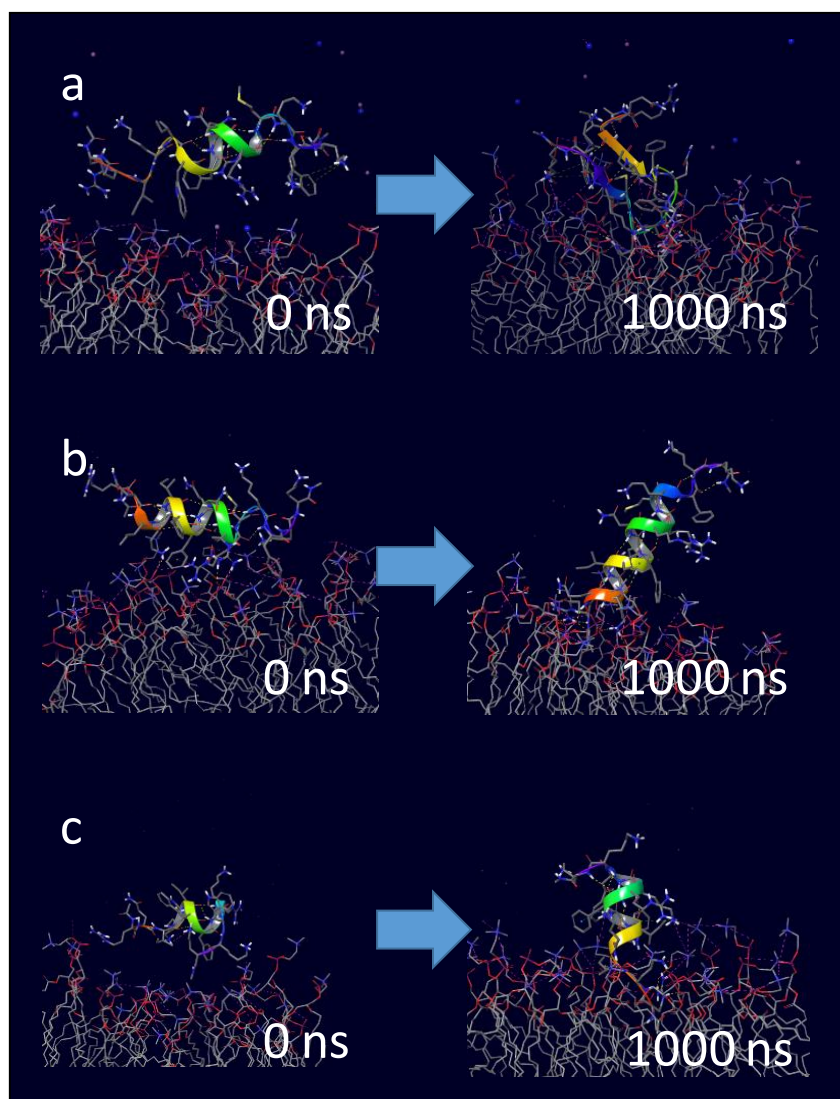
Figure 11. Comparison of the initial (left) and final (right) positions of unconjugated CCPs during the 1000-ns membrane simulations: penetratin with POPC (a); 6,14-Phe-penetratin with POPC (b); dodeca-penetratin with POPC (c) – all peptides starting from the surface of the bilayer. [28]

During the simulation of penetratin on the POPC membrane model, the RMSD fluctuated in the first half of the time, increased rapidly before 600 ns and then stayed in a close interval until the final frame (Figure 12. a). The number of hydrogen bonds decreased in the simulation's middle, quickly increasing at 600 ns and then staying on that level, indicating the new stable conformation (Figure 12. b).
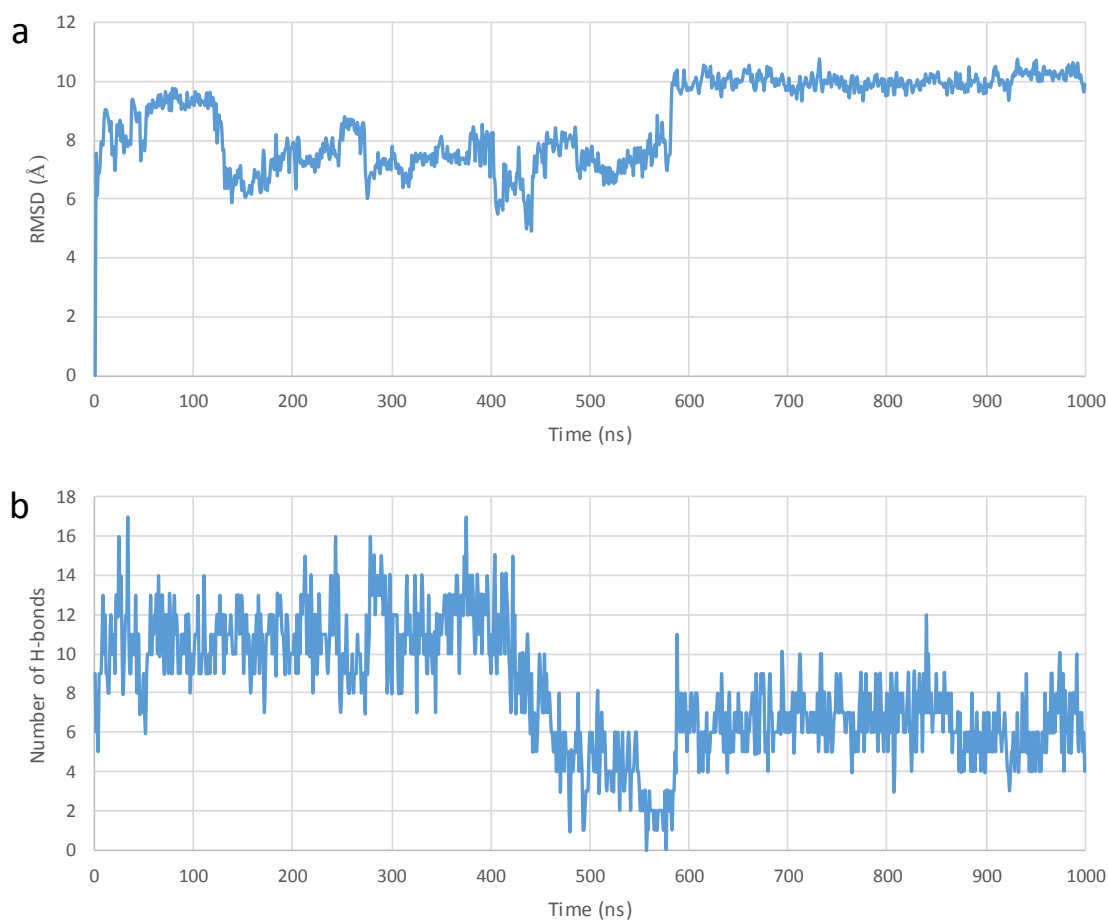
Figure 12. RMSD diagram of the α-carbon atoms of the peptides (a) and the total number of intramolecular hydrogen bonds (b) plotted against simulation time during the 1000-ns MD simulation of the unconjugated penetratin peptide – starting from the top of the membrane bilayer. [28]

In the Ramachandran plot, most of the dihedrals were in the range of the area belonging to the helical conformation at the beginning of the simulation of unconjugated penetratin on the POPC membrane surface (Figure 13. a). After 1000 ns, most dihedrals moved to the β-sheet and β-turn areas (Figure 13. b).
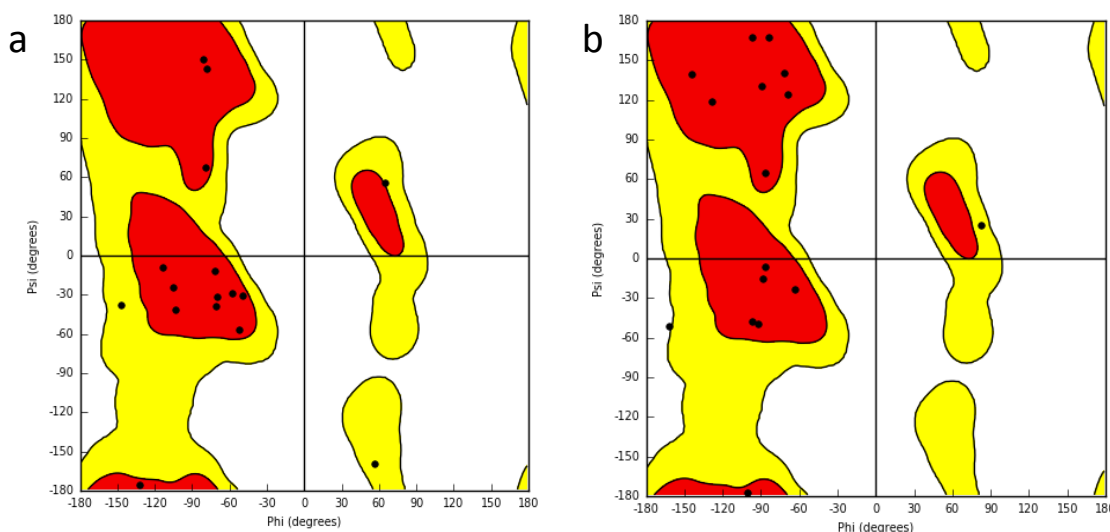
Figure 13. Comparison of the Ramachandran plots of the unconjugated penetratin peptide at the beginning (a) and end (b) of the 1000-ns simulation with the POPC membrane model – starting from the top of the membrane bilayer. [28]

4.4. Conjugated penetratin and analogues

During 1000-ns simulations, however, we observed that the cargo molecules significantly affected the peptide's structure and the conjugate's position relative to the membrane. In contrast to native penetratin (which entirely decoiled), the helical structure remained intact in all conjugated forms. Those positioned with one of the terminals towards the membrane, often partially sank into it, with the axis of the helix either perpendicular or closing angle with the membrane's plane.

All doxorubicin conjugates were positioned differently: the penetratin conjugate moved apart from the membrane, and no interactions were formed (Figure 14. a). The 6,14-Phe-penetratin conjugate anchored to the membrane through its N-terminal with the doxorubicin tightly bound on the surface (Figure 14. b). In contrast, the dodeca-penetrarin conjugate positioned with its C-terminal and doxorubicin in the water box (Figure 14. c). [28]

The rasagiline conjugates of all three penetratin analogues were always positioned with their N-terminals toward the membrane, and the conjugated compound sank into the bilayer (Figure 14. d-f). [28]

The simulations with zidovudine conjugates showed that among the three drugs, this was the least transportable, and its conjugates had significantly less interaction with the membrane: the penetratin conjugate was anchored with its C-terminal to the membrane,

while the N-terminal with the cargo moved away (Figure 14. g), the 6,14-Phe-penetratin conjugate entirely moved apart from the membrane (Figure 14. h), while only the dodeca-penetratin conjugate turned towards it with the N-terminal (Figure 14. i). [28]
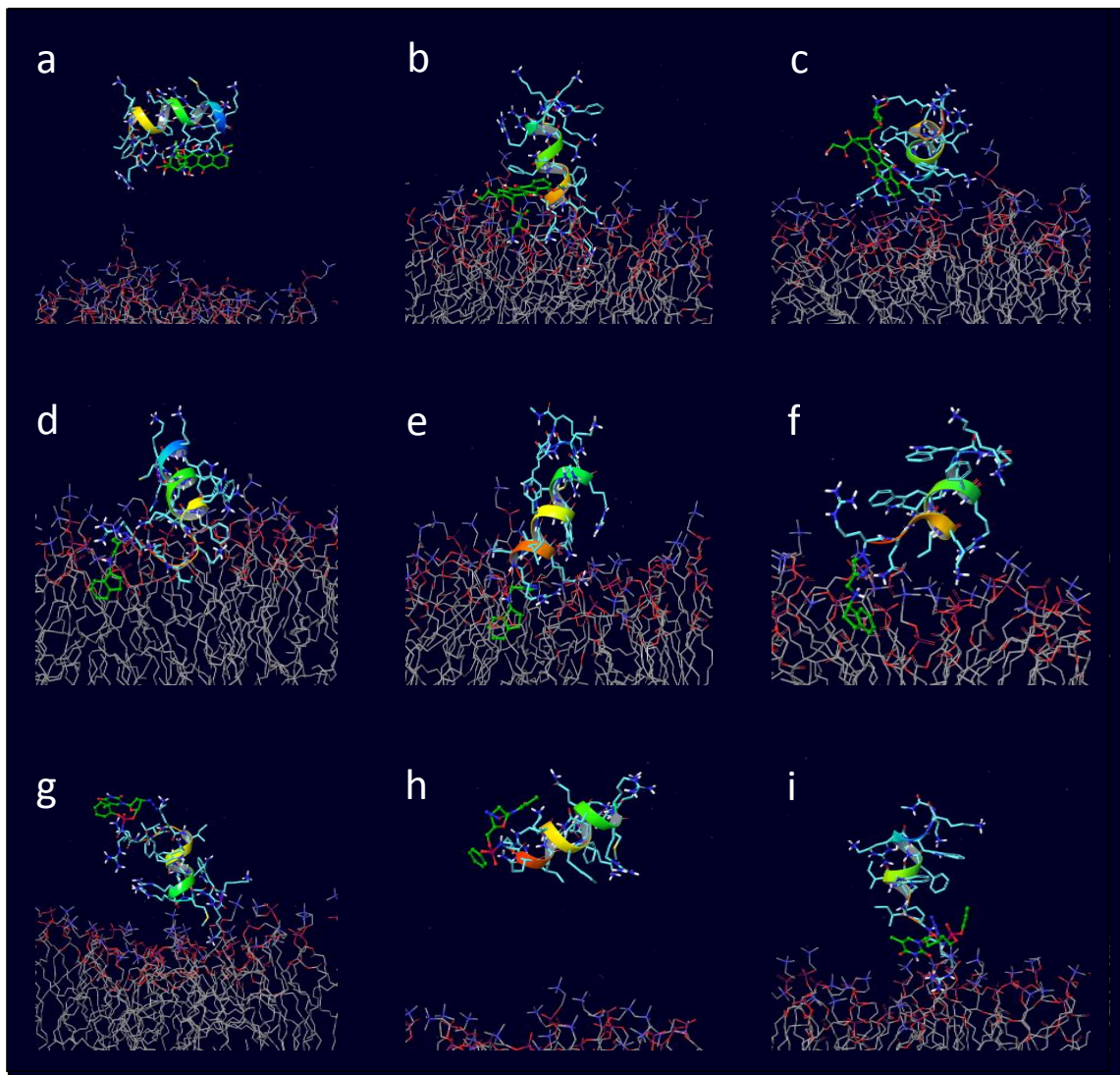
Figure 14. The final positions of the 1000-ns POPC membrane simulations with different CCP-conjugates: penetratin–doxorubicin (a); 6,14-Phe-penetratin–doxorubicin (b); dodeca-penetratin–doxorubicin (c); penetratin–rasagiline (d); 6,14-Phe-penetratin–rasagiline (e); dodeca-penetratin–rasagiline (f); penetratin–zidovudine (g); 6,14-Phe-penetratin–zidovudine (h); dodeca-penetratin–zidovudine (i) – all conjugates were started from the top of the membrane bilayer. [28]

The penetratin-doxorubicin conjugate remained helical; the RMSD fluctuated within a very small interval during the entire simulation (Figure 15. a). The number of hydrogen bonds never dropped significantly, indicating no significant conformational changes (Figure 15. b).
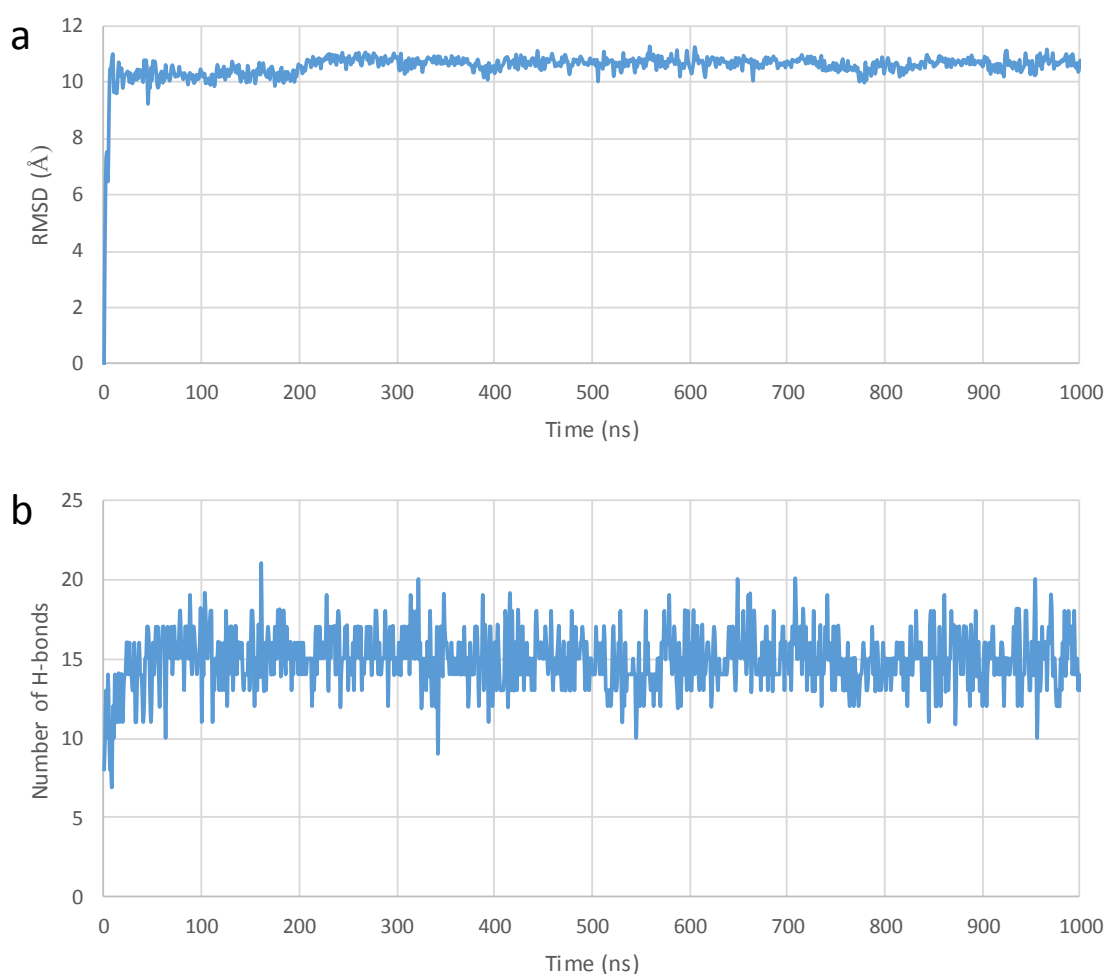
Figure 15. The RMSD diagram of the α-carbon atoms of the peptides (a) and the total number of intramolecular hydrogen bonds (b) plotted against simulation time during the 1000-ns MD simulation of penetratin–doxorubicin conjugate – starting from the top of the membrane bilayer. [28]

4.5. Unconjugated and conjugated cyclic peptides

We observed that the unconjugated cyclic peptide structures did not change during the 1000 ns simulations, regardless of the medium. During the POPC runs, they all positioned toward the membrane and then tightly anchored to its surface with a minimal sinking into the bilayer. Several mostly uncharged hydrogen bonds were observed between the peptide heteroatoms (as donors) and the heteroatoms of the membrane (as acceptors) because these cyclic peptides – except MCoTI-II – have only a few amino acids with polar side chains, which limits their capability for the formation of ionic interactions (Figure 16. a-c,). [28]

The impact of conjugation in the case of the cyclic CPPs seemed to be less significant, indicating that their capability for adherence was less hindered. Namely, the positions of all conjugates were very similar to those of their unconjugated forms with the POPC membrane model after 1000 ns. The doxorubicin conjugate of all three cyclic peptides sank into the membrane (Figure 16. d-f). Rasagiline sank into the membrane when it was conjugated with the MCoTI-II, but it remained on the surface of the membrane when conjugated with the other two cyclic peptides (Figure 16 g-i). Similarly, zidovudine sank deeply in the bilayer when conjugated with MCoTI-II but stayed on the surface with Kalata B1 and SFTI-1 (Figure 16. j-l). Also, it must be pointed out that the simulation with MCoTI-II-zidovudine was the only one in which a considerable membrane distortion was observed, although no full penetration nor perforation occoured (Figure 16. l).
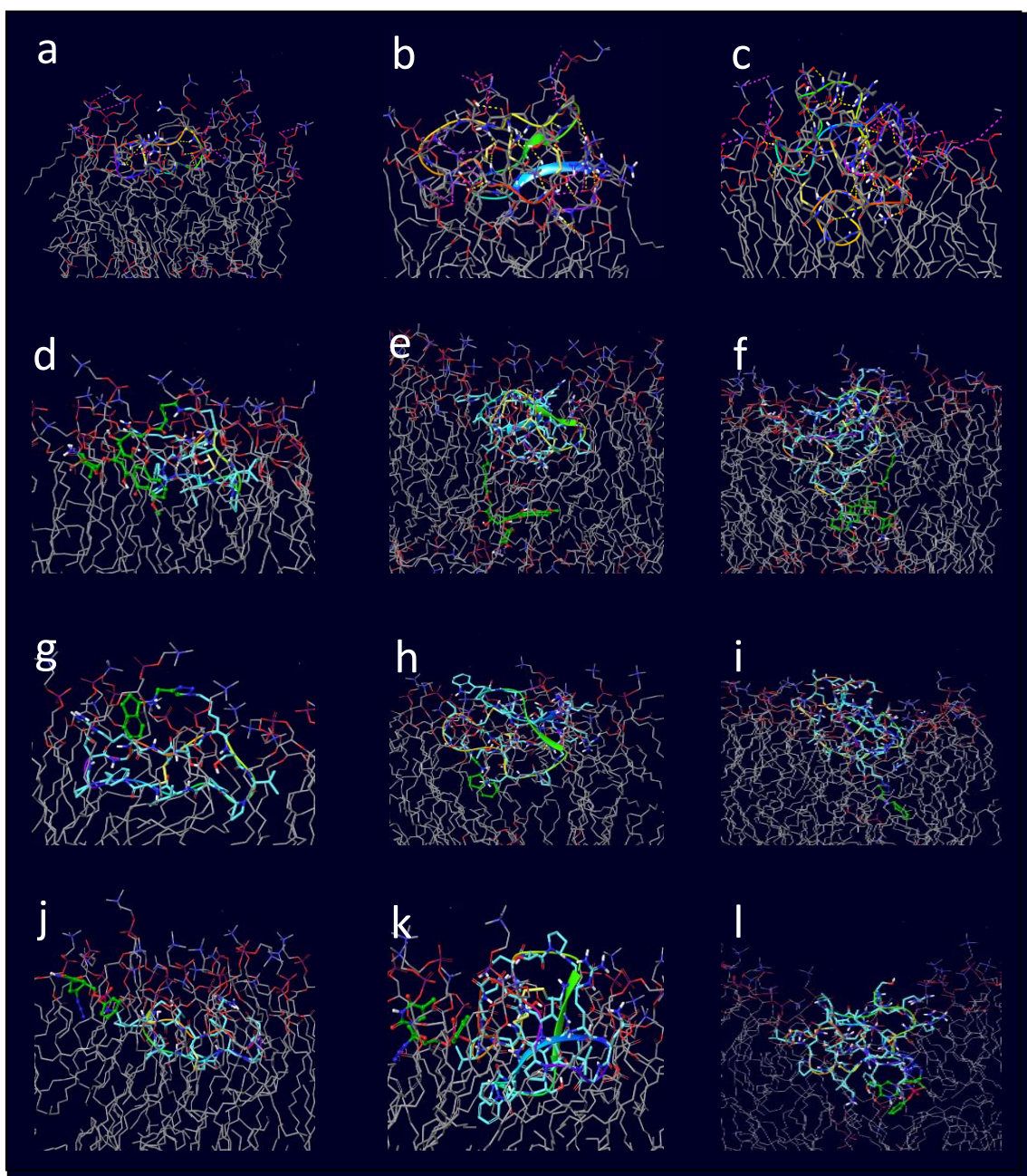
Figure 16. The final positions of the 1000-ns POPC membrane simulations with cyclic CCPs and their conjugates: unconjugated SFTI-1 (a); unconjugated Kalata B1 (b); unconjugated MCoTI-II (c); SFTI-1-doxorubicin (d); Kalata B1-doxorubicin (e); MCoTI-II-doxorubicin (f); SFTI-1-rasagiline (g); Kalata B1-rasagiline (h); MCoTI-II-rasagiline (i); SFTI-1-zidovudine (j); Kalata B1-zidovudine (k); MCoTI-II-zidovudine (l) – all conjugates were started from the top of the membrane bilayer. [28]

The RMSD of the unconjugated SFTI-1 has an uncertain increasing tendency without any outlier values (Figure 17. a). The number of hydrogen bonds fluctuated but remained stable, so the peptide conformation did not rearrange (Figure 17. b).

a

b
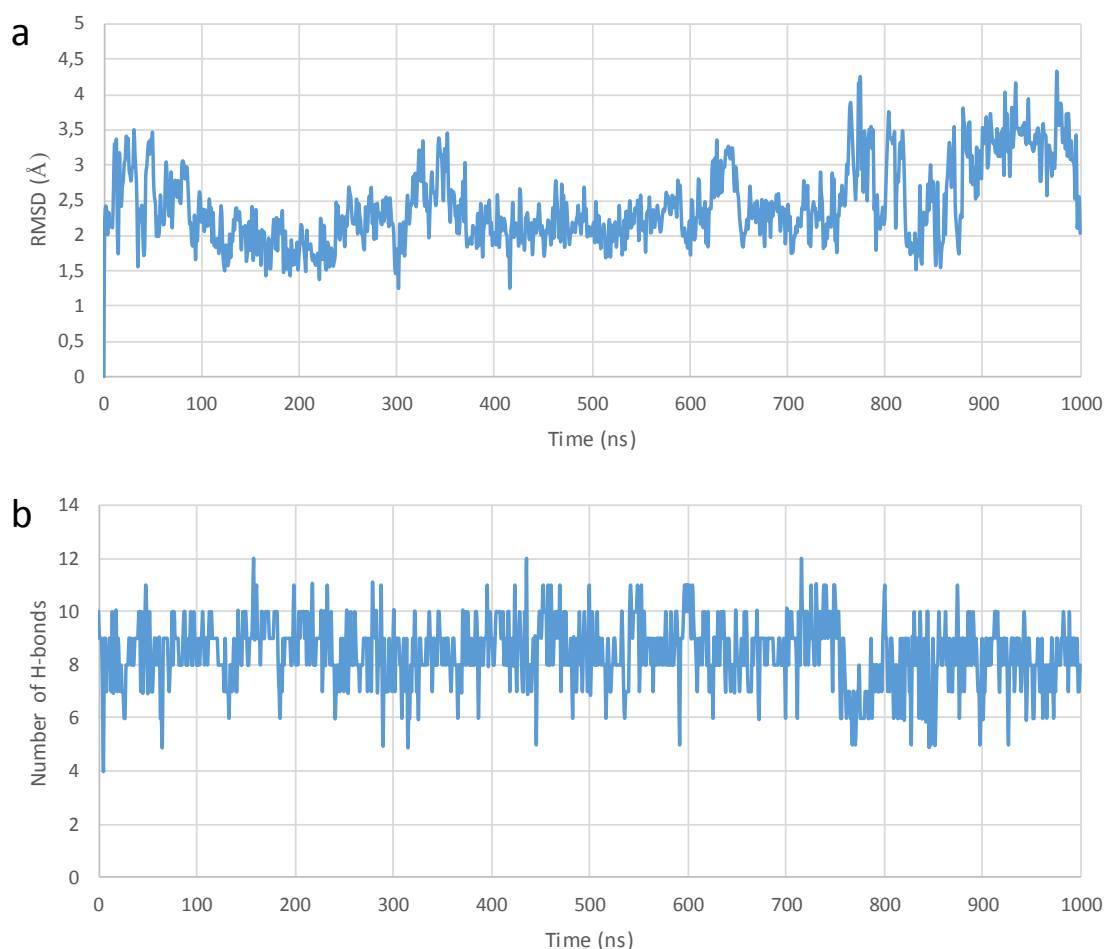
Figure 17. The RMSD diagram of the α-carbon atoms of the peptides (a) and the total number of intramolecular hydrogen bonds (b) plotted against simulation time during the 1000-ns MD run of the unconjugated SFTI-1 peptide – starting from the top of the membrane bilayer. [28]

4.6. Summary of the interactions

The number of interactions between membrane-peptide and membrane-conjugate based on the last frame of the trajectories was summarised in Table 4. Despite our expectations, the unconjugated penetratin did not form any π-cation interactions with the membrane lipids. Interestingly, in the case of the unconjugated 6,14-Phe-penetratin although the contact area was limited (compared to penetratin), five salt bridges, seven hydrogen bonds and one π-cation interaction were formed (Table 4 Entry 5.). The unconjugated dodeca-penetratin connected even more loosely to the membrane surface with only one salt bridge and six hydrogen bonds (Table 4 Entry 9.). [28]

Despite the perpendicular position, the 6,14-Phe-penetratin-doxorubicin formed five salt bridges, seven hydrogen bonds and one π-cation interaction with membrane

molecules (Table 4 Entry 6.). The dodeca-penetratin-doxorubicin connected even more loosely (compared to its unconjugated form) to the membrane surface with only one salt bridge and six hydrogen bonds (Table 4 Entry 10.). The position of the rasagiline conjugate of all three penetratin analogues was stabilised by forming several hydrogen bonds and salt bridges between the peptide and membrane molecules (Table 4 Entries 3, 7, 11.). The zidovudine conjugate of penetratin and its analogues formed the fewest interactions on average (Table 4. Entries 4, 8, 12.).

The cyclic peptides and conjugates generally had fewer interactions with the membrane than the penetratin and analogues because the more significant portions of these peptides were located within the membrane and only slightly above. The lack of aromatic side chains also excluded the appearance of π-cation interactions with the positively charged choline groups of the POPC membrane. There were some exceptions, though, like the MCoTI-II-doxorubicin and MCoTI-II-zidovudine conjugates, with many hydrogen bonds salt bridges (Table 4. Entries 22, 24.).

Table 4. The number of observed interactions between the peptides/conjugates and the POPC membrane molecules at the end of the 1000 ns simulations. [28]

Entry	Peptide	Conjugate	H-bond	π-cation	salt bridge
1	penetratin (1KZ0)	unconjugated	11	0	4
2	penetratin (1KZ0)	doxorubicin	0	0	0
3	penetratin (1KZ0)	rasagiline	12	1	7
4	penetratin (1KZ0)	zidovudine	1	0	2
5	6,14-Phe-penetratin (1KZ2)	unconjugated	7	1	5
6	6,14-Phe-penetratin (1KZ2)	doxorubicin	8	0	3
7	6,14-Phe-penetratin (1KZ2)	rasagiline	10	0	4
8	6,14-Phe-penetratin (1KZ2)	zidovudine	0	0	0
9	dodeca-penetratin (1KZ5)	unconjugated	6	0	1
10	dodeca-penetratin (1KZ5)	doxorubicin	2	1	3
11	dodeca-penetratin (1KZ5)	rasagiline	5	0	4
12	dodeca-penetratin (1KZ5)	zidovudine	4	0	2
13	SFTI-1 (1NB1)	unconjugated	0	0	0
14	SFTI-1 (1NB1)	doxorubicin	10	0	2
15	SFTI-1 (1NB1)	rasagiline	6	0	2
16	SFTI-1 (1NB1)	zidovudine	2	0	2
17	Kalata B1 (1JBL)	unconjugated	2	0	0

18	Kalata B1 (1JBL)	doxorubicin	6	0	3
19	Kalata B1 (1JBL)	rasagiline	4	0	3
20	Kalata B1 (1JBL)	zidovudine	3	0	1
21	MCoTI-II (1HA9)	unconjugated	0	0	0
22	MCoTI-II (1HA9)	doxorubicin	8	0	11
23	MCoTI-II (1HA9)	rasagiline	6	0	8
24	MCoTI-II (1HA9)	zidovudine	16	0	6

5. Discussion

5.1 Peptide conjugate database

Anticancer (372) and antimicrobial (262) utilisations are the most common application types of peptide conjugates (Figure 18). Some chemotherapeutic agents are poorly soluble in water or are toxic. These drawbacks can be mitigated by conjugating them with a small peptide. Other conjugates are designed as 'prodrugs' activated by an enzyme the tumour cells produce by cleaving the peptide part. Tumor-specific peptides made the targeted therapy available: the Human epidermal growth factor (HER2) receptor and the gonadotropin-releasing hormone (GnRH) receptors are typical examples of these. Adding a CPP can more effectively cumulate the anticancer drug in the tumour cells.

The antimicrobial (containing antibacterial, antiviral, antifungal, and antiprotozoal) conjugates (110) are the second most frequent conjugates analysed (Figure 18). In most instances, these conjugates enhance the selectivity or effectiveness of the compounds (or reduce their resistance) compared to their unconjugated forms. For example, the antimycobacterial activity of peptide-conjugated pyridopyrimidine derivatives was more effective and less toxic in a series of in vivo and in vitro testing than the original drug compound. Antiprotozoal (23 conjugates, mainly antimalarial), antiviral (114 conjugates, for example. HIV-1, Hepatitis-C and Herpes simplex virus), and antifungal (15 conjugates) also were found in the literature in a smaller number.

A significant number of anti-inflammatory drug conjugates (122 in total) have been identified (Figure 18), including those used for treating conditions like rheumatoid arthritis. Additionally, conjugates targeting specific tissues, such as bone-targeting ibuprofen, have also been explored.

Central nervous system diseases, primarily Alzheimer's (Figure 18), also appear to be a promising research area, with 71 conjugates identified.

Our literature search primarily focused on drug conjugates with biomedical applications. Several other conjugates (749 in total) were identified without distinct pharmacological properties (Figure 18). These primarily involved various synthetic endeavors aimed at inventing and refining conjugation techniques, often without accompanying biological testing. Various rare biomedical applications were also

published, including facilitated blood−brain barrier transport biotechnological applications like DNA cleavage, radiolabeling, imaging, or chelationIn the context of the database, these entries are valuable as they highlight the broader applicability of peptide conjugates and offer valuable insights into synthetic solutions for organic chemists interested in this field.
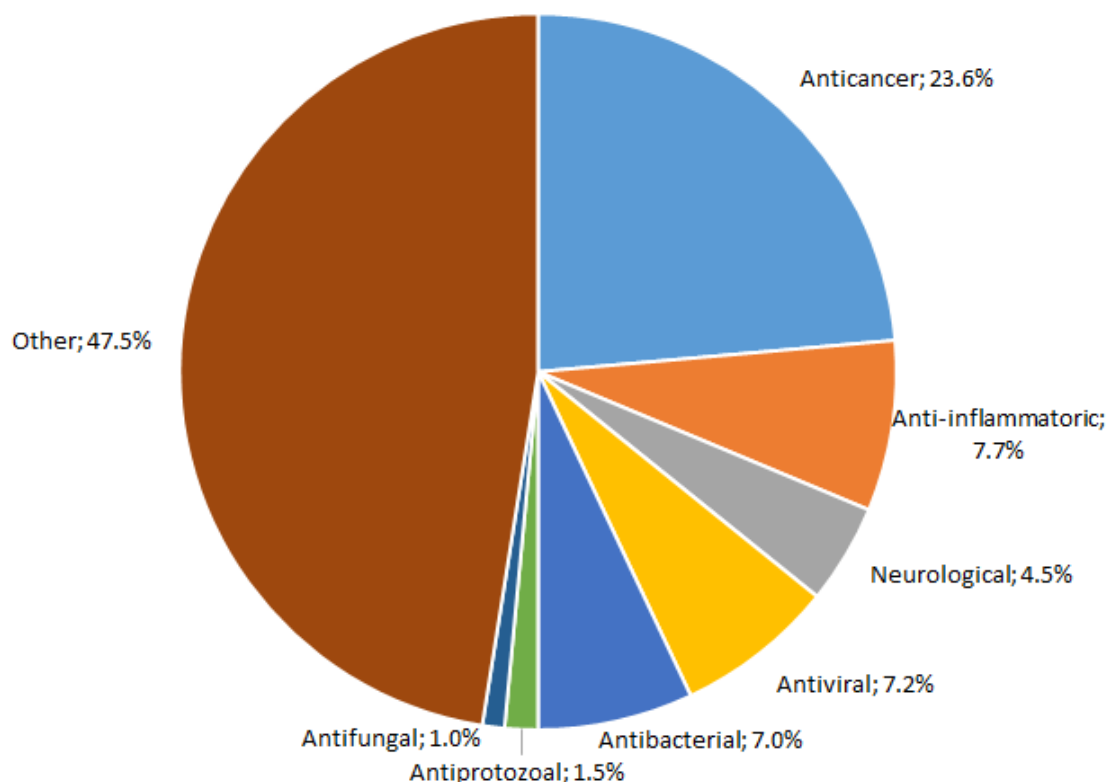
Figure 18. Distribution of applications found for the database entries. [43]

5.2. Differences between water box, intra-membrane and membrane top

In the water box and the POPC membrane model, we ran 100 ns MD simulations to observe the conformation of the peptides and their conjugates. The 1000 ns membrane-top simulations should attempt to penetrate.

Compared to the structures obtained from the initial NMR structures in the water model, both penetratin (1KZ0) and 6,14-Phe-penetratin (1KZ2) showed an increase in helical section size over 100 ns. This is likely due to the correct basic to polar neutral side chain ratio. Dodeca-penetratin (1KZ5), on the other hand, adopted an irregular

conformation at the end of the run, presumably because it contains too few amino acids to form a stable helix.

When running MD simulations in a membrane model, only relatively minor changes from baseline were observed for the penetratin and 6,14-Phe analogue structures, with a typical decrease in the dihedral angle dispersion and some modification of the peptide position within the membrane. dodeca-penetratin, on the other hand, was not found to be more stable in the membrane than in the water model.

At the end of 1000 ns run-time simulations from the inside of the POPC membrane model, started parallel and perpendicular to the membrane plane, the penetratin axis was in a position almost vertical to the membrane, independent of the initial position; it can be assumed that this may be the most stable position in the membrane.

Complete membrane penetration was generally not observed for any of the molecules investigated. In the case of linear CPPs, penetratin showed a significant conformational change during the simulation, creating several salt bridges and hydrogen bonds with the negatively charged lipid surface, forming two strands connected by a ß-turn. Phe-modified penetratin showed a different behaviour: it maintains its original helical conformation, and the N-terminal of the helix slightly merges into the lipid bilayer. The dodeca-penetratin derivative behaves similarly but combines slightly more deeply into the lipid bilayer. Fewer H-bonds and salt bridges were formed than the penetratin, but additional π-cationic interactions were also observed.

The cyclic CPPs behave in a radically different manner. All three investigated compounds extruded water between themselves and the membrane, and they were attached to the lipid bilayer, forming direct interactions. Entropic reasons, a leading cause of water exclusion, can explain the peptide's anchoring to the membrane.

When calculations were started from the lipid bilayer, linear peptides stayed within the membrane but were oriented perpendicular to the surface of the membrane. However, cyclic peptides behaved differently, migrating between the two lipid bilayers.

As we mentioned, all three cyclic peptides showed strong conformational stability. The disulfide bridges of the knot motif kept the tight structure, and the intramolecular hydrogen bonds increased the rigidity, too. In addition, the included cyclic CPPs contain

proline residues that stabilise their region's secondary structure despite glycines. The penetratin and its analogues have neither glycines nor prolines, so their conformation is much more variable overall. The linear backbone and the lack of cysteines give them even more structural freedom.

5.3. Affect of the conjugated molecules

The conjugated compounds generally did not interact with the CPPs during the simulation. However, in all cases, the drug molecule's conjugation significantly influenced the molecules' membrane-interacting behaviour.

Doxorubicin is an amphiphilic molecule possessing a hydrophobic anthraquinone ring substituted with a hydrophilic aminosugar derivative. When doxorubicin was attached to penetration, the conjugate moved apart from the membrane. In the case of Phe-derivative–doxorubicin conjugate, the molecule behaved differently. The N-terminus carrying the conjugate merged slightly into the membrane. The dodeca-penetratin derivative–doxorubicin conjugate joined barely into the membrane with its C-terminal. The hydrophilic part of the molecule can form hydrogen bonds with the phospholipids. These facts suggest that the amphipathic nature of doxorubicin influences the behaviour of the conjugate.

The following drug investigated is zidovudine, a more hydrophilic molecule than doxorubicin. When it was conjugated to penetratins, no interaction was found between the two parts of the conjugates. The C-terminal of the penetratin conjugate slightly merged into the membrane, while the N-terminal equipped with the zidovudine remained in the water. In the case of the Phe-derivative–zidovudine conjugate, the assembly diverged from the membrane and persisted in the water phase. In the case of the 12-AA-long dodeca-penetratin derivative conjugate, the N-terminus slightly merged into the membrane. The phosphate group of the zidovudine could form a salt bridge to the choline part of a POPC molecule.

Rasagiline, a hydrophobic compound, was also tested and conjugated to the three penetratins. All three conjugates behaved similarly; the peptidic part retained its helical conformation and slightly merged into the membrane with its N-terminal region. However, the rasagiline part sincerely joined the bilayer because of its highly apolar nature. The aromatic ring can form π-cation interaction with the choline part of lecithin.

When calculations were started from the lipid bilayer, penetratin and its two derivatives remained in their original helical conformation and oriented perpendicular to the lipid bilayer. Under these conditions, the conjugates of these peptides behaved in a very similar way.

For the cyclic CPPs, the polarity of the small organic molecule dominantly influenced the behaviour of the conjugate. The peptide part was attached to the membrane in all cases, and water was extruded. As a conjugate, Doxorubicin slightly merged into the membrane, forming a hydrogen bond with the oxygen atom bound to the phosphorous atom in the lipid's head part. Zidovudine diverged from the membrane because of its hydrophilic nature and formed hydrogen bonds with water molecules. The most hydrophobic rasagiline deeply merged into the membrane as long as its linker allowed.

When cyclic CPP conjugates were positioned in the middle of the membrane and used as starting geometry, MD calculations revealed that the peptide inserts between the two lipid bilayers and the conjugates followed the CPP.

Penetratin suffered the most significant effect on its conformation: no one of its conjugates went through the structural change as observed in the unconjugated peptide case. The two penetratin analogues' conjugates are arranged into a similar conformation as unconjugated. As expected, the cyclic peptides maintained prominent conformational stability after the conjugation.

Secondary interactions between the conjugated compound and the peptide were not decisive, but we observed some—a double π-π stacking formed between the doxorubicin and the indole group of tryptophane in the penetratin. The phosphate group of zidovudine connected to the first backbone amide group of penetratin. In the case of dodeca-penetratin, π-π stacking formed between the ring system of doxorubicin and the phenylalanine.

6. Conclusions

In brief, we developed [ConjuPepDB](#), a freely accessible database containing comprehensive information on peptide-drug conjugates. Through manual curation, we compiled and summarized crucial details from this rapidly advancing field of potential drug candidates. Each entry includes essential information such as the conjugate's structure, CAS RN, peptide and small organic drug molecule names, along with details on its biomedical application and the type of chemical linkage used for coupling. ConjuPepDB encompasses 1,645 entries sourced from 238 research papers across 95 scientific journals, establishing it as the first comprehensive database of its kind. We anticipate that this resource will underscore the significance of these organic compounds and their substantial biomedical potential for future industrial applications.

In the following study, we examined the behaviour of CPPs with covalently conjugated drug molecules, applying atomistic MD simulations. Although a complete membrane penetration was not observed, some interesting conformational and positional changes were supervised during the 1000-ns simulation time.

We found that only the unconjugated penetratin underwent significant conformational rearrangement, while less flexible 6,14-Phe-penetratin and dodeca-penetratin kept their mainly helical structure. Penetratin and its analogues were more affected by the polarity of the conjugated small molecule: the hydrophilic zidovudine ostensibly inhibited the interaction between the peptide and the membrane, the more hydrophobic rasagiline led the entire conjugate in between the membrane bilayer, during the impact of the amphiphilic doxorubicin depended on the peptide.

The three cyclic peptides (SFTI-1, Kalata B1 and MCoTI-II) behaved similarly during the simulations. Because their structure is so stable, only minimal conformational changes were observed, and their position compared to the surface of the lipid bilayer was less variable. The effect of the conjugates on the penetration also seemed less significant, but the conjugated small molecules were oriented according to their polarity.

The final position of the native penetratin could be the first step of a mechanism of passive penetration (direct translocation, pore formation, inverted micelles or carpet model) as described in the literature, but some form of endocytosis could also follow it.

The anchoring on the membrane surface of the other peptides and conjugates may also indicate the possibility of penetration.

The failure of direct penetration might result from the model's relative simplicity. Although we used atomistic MD, a simple monomolecular membrane model was applied, and other possible membrane components and the membrane potential needed to be adequately implemented. Our system contained only a single CPP conjugate; therefore, more complex multi-molecular mechanisms could not be examined, such as complexation with other proteins or pore formation.

It is also conceivable that a more complex system, a more detailed, multi-component membrane model, might be needed to successfully model penetration, including, for example, sphingomyelin and ceramide, but most importantly, cholesterol, in addition to POPC and POPE (palmitoyl-oleoyl-phosphatidylethanolamine); all these components mixed in proportions approximating the composition of biological membranes, even in different proportions in the two layers. The membrane potential of the cell and the other extra- and intracellular concentrations of ions may also be essential to consider.

A larger scale system with multiple peptides or conjugates should also be considered, as in most models described in the literature, various CPP molecules and their interaction and cooperation with each other may be required simultaneously. A sufficiently large system is suitable for modelling pore formation, inverse micelles, and carpet models with long enough run times.

We sincerely hope that we can build a much larger simulation box including not only a single peptide in sufficiently high concentration; spontaneous penetration may be observed, even if such a system is unsuitable for modelling most types of endocytosis.

7. Summary

Peptide-drug conjugates are compounds consisting of a small drug molecule and a peptide linked together. Despite their significance, there was no comprehensive database for these compounds. To address this, we developed ConjuPepDB, a free, curated database providing detailed annotations of peptide-drug conjugates.

ConjuPepDB includes fundamental details such as the CAS number and components of each conjugate, their biomedical applications, and specific chemical conjugation methods. The database features over 1600 conjugates from nearly 230 publications.

The user-friendly web interface is accessible across devices and offers multiple search criteria, including chemical structures, with a dedicated help page. ConjuPepDB supports researchers in related fields.

Cell-penetrating peptides (CPPs) can cross biological membranes, transporting conjugated cargo into cells. Our computational study models several peptides and their conjugates. We chose three linear CPPs: penetratin, and two analogs (6,14-Phe-penetratin and dodeca-penetratin), and three cyclic CPPs: Kalata B1, SFTI-1, and MCoTI-II. Each was conjugated with doxorubicin, zidovudine, and rasagiline.

Molecular dynamics (MD) simulations examined the behavior of conjugated and unconjugated peptides in water and membrane models. Analysis showed penetratin's interaction with the membrane caused significant structural rearrangements, while cyclic peptides remained stable due to their conformational stability.

Membrane–peptide and membrane–conjugate interactions were identified and compared. Covalent conjugation influenced penetratin and its analogs more than the cyclic CPPs. Simulations underscored the effectiveness of computational methods for CPP complexes, enhancing our understanding of penetration mechanisms and advancing drug delivery to intracellular targets. Promising conjugates included rasagiline-penetratin and its analogs, Kalata B1-rasagiline, MCoTI-II-rasagiline, Kalata B1-doxorubicin, MCoTI-II-doxorubicin, and MCoTI-II-zidovudine.

8. References

1. Bolhassani, A., B.S. Jafarzade, and G. Mardani, *In vitro and in vivo delivery of therapeutic proteins using cell penetrating peptides.* Peptides, 2017. **87**: p. 50-63.

2. Heitz, F., M.C. Morris, and G. Divita, *Twenty years of cell-penetrating peptides: from molecular mechanisms to therapeutics.* Br. J. Pharmacol., 2009. **157**(2): p. 195-206.

3. Kristensen, M., D. Birch, and H.M. Nielsen, *Applications and challenges for use of cell-penetrating peptides as delivery vectors for peptide and protein cargos.* Int. J. Mol. Sci., 2016. **17**(2): p. 185/1-185/17.

4. Frankel, A.D. and C.O. Pabo, *Cellular uptake of the tat protein from human immunodeficiency virus.* Cell, 1988. **55**(6): p. 1189-1193.

5. Derossi, D., G. Chassaing, and A. Prochiantz, *Trojan peptides: the penetratin system for intracellular delivery.* Trends Cell. Biol., 1998. **8**(2): p. 84-87.

6. Lindberg, M., H. Biverstahl, A. Graeslund, and L. Maeler, *Structure and positioning comparison of two variants of penetratin in two different membrane mimicking systems by NMR.* Eur. J. Biochem., 2003. **270**(14): p. 3055-3063.

7. Avci, F.G., B.S. Akbulut, and E. Ozkirimli, *Membrane Active Peptides and Their Biophysical Characterization.* Biomolecules, 2018. **8**(3): p. 77.

8. Gautam, A., H. Singh, A. Tyagi, K. Chaudhary, R. Kumar, P. Kapoor, and G.P. Raghava, *CPPsite: a curated database of cell penetrating peptides.* Database (Oxford), 2012. **2012**: p. bas015.

9. Agrawal, P., S. Bhalla, S.S. Usmani, S. Singh, K. Chaudhary, G.P. Raghava, and A. Gautam, *CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides.* Nucleic. Acids. Res., 2016. **44**(D1): p. D1098-1103.

10. Su, R., J. Hu, Q. Zou, B. Manavalan, and L. Wei, *Empirical comparison and analysis of web-based cell-penetrating peptide prediction tools.* Brief Bioinform., 2020. **21**(2): p. 408-420.

11. Kardani, K. and A. Bolhassani, *Cppsite 2.0: An Available Database of Experimentally Validated Cell-Penetrating Peptides Predicting their Secondary and Tertiary Structures.* J. Mol. Biol., 2021. **433**(11): p. 166703.

12. Cascales, L., S.T. Henriques, M.C. Kerr, Y.H. Huang, M.J. Sweet, N.L. Daly, and D.J. Craik, *Identification and characterization of a new family of cell-penetrating peptides: cyclic cell-penetrating peptides.* J. Biol. Chem., 2011. **286**(42): p. 36932-36943.

13. Jafari, S., S.M. Dizaj, and K. Adibkia, *Cell-penetrating peptides and their analogues as novel nanocarriers for drug delivery.* BioImpacts, 2015. **5**(2): p. 103-111.

14. Sawant, R.R., N.R. Patel, and V.P. Torchilin, *Therapeutic delivery using cell-penetrating peptides.* Eur. J. Nanomed., 2013. **5**(3): p. 141-158.

15. Kalafatovic, D. and E. Giralt, *Cell-Penetrating Peptides: Design Strategies beyond Primary Structure and Amphipathicity.* Molecules, 2017. **22**(11): p. 1729.

16. Madani, F., S. Lindberg, U. Langel, S. Futaki, and A. Graeslund, *Mechanisms of cellular uptake of cell-penetrating peptides.* J. Biophys., 2011: p. 414729, 10 pp.

17. Dufourc, E.J., S. Buchoux, J. Toupe, M.A. Sani, F. Jean-Francois, L. Khemtemourian, A. Grelard, C. Loudet-Courreges, M. Laguerre, J. Elezgaray, B. Desbat, and B. Odaert, *Membrane interacting peptides: from killers to helpers.* Curr. Protein. Pept. Sci., 2012. **13**(7): p. 620-631.

18. Huang, Y.W., H.J. Lee, L.M. Tolliver, and R.S. Aronstam, *Delivery of nucleic acids and nanomaterials by cell-penetrating peptides: opportunities and challenges.* Biomed. Res. Int., 2015. **2015**: p. 834079.

19. Gao, X., S. Hong, Z. Liu, T. Yue, J. Dobnikar, and X. Zhang, *Membrane potential drives direct translocation of cell-penetrating peptides.* Nanoscale, 2019. **11**(4): p. 1949-1958.

20. Ulmschneider, J.P. and M.B. Ulmschneider, *Molecular Dynamics Simulations Are Redefining Our View of Peptides Interacting with Biological Membranes.* Acc. Chem. Res., 2018. **51**(5): p. 1106-1116.

21. Cardenas, A.E., R. Shrestha, L.J. Webb, and R. Elber, *Membrane Permeation of a Peptide: It Is Better to be Positive.* J. Phys. Chem. B, 2015. **119**(21): p. 6412-6420.

22. Herce, H.D. and A.E. Garcia, *Cell penetrating peptides: how do they do it?* J. Biol. Phys., 2007. **33**(5-6): p. 345-356.

23. Kawamoto, S., T. Miyakawa, M. Takasu, R. Morikawa, T. Oda, H. Saito, S. Futaki, and H. Nagao, *Cell-penetrating peptide induces various deformations of lipid bilayer membrane: Inverted micelle, double bilayer, and transmembrane.* Int. J. Quantum Chem., 2012. **112**(1): p. 178-183.

24. Durzynska, J., L. Przysiecka, R. Nawrot, J. Barylski, G. Nowicki, A. Warowicka, O. Musidlak, and A. Gondzicka-Jozefiak, *Viral and other cell-penetrating peptides as vectors of therapeutic agents in medicine.* J. Pharmacol. Exp. Ther., 2015. **354**(1): p. 32-42.

25. Antunes, E., N.G. Azoia, T. Matama, A.C. Gomes, and A. Cavaco-Paulo, *The activity of LE10 peptide on biological membranes using molecular dynamics, in vitro and in vivo studies.* Colloids Surf. B Biointerfaces, 2013. **106**: p. 240-247.

26. Qian, Z., J.R. LaRochelle, B. Jiang, W. Lian, R.L. Hard, N.G. Selner, R. Luechapanichkul, A.M. Barrios, and D. Pei, *Early endosomal escape of a cyclic cell-penetrating peptide allows effective cytosolic cargo delivery.* Biochemistry, 2014. **53**(24): p. 4034-4046.

27. Zhang, X., Y. Jin, M.R. Plummer, S. Pooyan, S. Gunaseelan, and P.J. Sinko, *Endocytosis and membrane potential are required for HeLa cell uptake of R.I.-CKTat9, a retro-inverso Tat cell penetrating peptide.* Mol. Pharm., 2009. **6**(3): p. 836-848.

28. Ivánczi, M., B. Balogh, L. Kis, and I. Mándity, *Molecular Dynamics Simulations of Drug-Conjugated Cell-Penetrating Peptides.* Pharmaceuticals, 2023. **16**(9).

29. Czajlik, A., E. Mesko, B. Penke, and A. Perczel, *Investigation of penetratin peptides. Part 1. The environment dependent conformational properties of penetratin and two of its derivatives.* J. Pept. Sci., 2002. **8**(4): p. 151-171.

30. Rosengren, K.J., N.L. Daly, M.R. Plan, C. Waine, and D.J. Craik, *Twists, knots, and rings in proteins. Structural definition of the cyclotide framework.* J. Biol. Chem., 2003. **278**(10): p. 8606-8616.

31. Korsinczky, M.L., H.J. Schirra, K.J. Rosengren, J. West, B.A. Condie, L. Otvos, M.A. Anderson, and D.J. Craik, *Solution structures by 1H NMR of the novel cyclic trypsin inhibitor SFTI-1 from sunflower seeds and an acyclic permutant.* J. Mol. Biol., 2001. **311**(3): p. 579-591.

32. Heitz, A., J.F. Hernandez, J. Gagnon, T.T. Hong, T.T. Pham, T.M. Nguyen, D. Le-Nguyen, and L. Chiche, *Solution structure of the squash trypsin inhibitor MCoTI-II. A new family for cyclic knottins.* Biochemistry, 2001. **40**(27): p. 7973-7983.

33. Letoha, T., S. Gaal, C. Somlai, Z. Venkei, H. Glavinas, E. Kusz, E. Duda, A. Czajlik, F. Petak, and B. Penke, *Investigation of penetratin peptides. Part 2. In vitro uptake of penetratin and two of its derivatives.* J. Pept. Sci., 2005. **11**(12): p. 805-811.

34. Vrettos, E.I., G. Mezo, and A.G. Tzakos, *On the design principles of peptide-drug conjugates for targeted drug delivery to the malignant tumor site.* Beilstein J. Org. Chem., 2018. **14**: p. 930-954.

35. He, R., B. Finan, J.P. Mayer, and R.D. DiMarchi, *Peptide Conjugates with Small Molecules Designed to Enhance Efficacy and Safety.* Molecules, 2019. **24**(10): p. 1855.

36. Wang, M., K.P. Rakesh, J. Leng, W.Y. Fang, L. Ravindar, D. Channe Gowda, and H.L. Qin, *Amino acids/peptides conjugated heterocycles: A tool for the recent development of novel therapeutic agents.* Bioorg. Chem., 2018. **76**: p. 113-129.

37. Copolovici, D.M., K. Langel, E. Eriste, and U. Langel, *Cell-Penetrating Peptides: Design, Synthesis, and Applications.* ACS Nano, 2014. **8**(3): p. 1972-1994.

38. Wang, S., N.Z. Zhelev, S. Duff, and P.M. Fischer, *Synthesis and biological activity of conjugates between paclitaxel and the cell delivery vector penetratin.* Bioorg. Med. Chem. Lett., 2006. **16**(10): p. 2628-2631.

39. Pari, E., K. Horvati, S. Bosze, B. Biri-Kovacs, B. Szeder, F. Zsila, and E. Kiss, *Drug conjugation induced modulation of structural and membrane interaction features of cationic cell-permeable peptides.* Int. J. Mol. Sci., 2020. **21**(6): p. 2197.

40. Diedrichsen, R.G., S. Harloff-Helleberg, U. Werner, M. Besenius, E. Leberer, M. Kristensen, and H.M. Nielsen, *Revealing the importance of carrier-cargo association in delivery of insulin and lipidated insulin.* J. Control. Release., 2021. **338**: p. 8-21.

41. Zaro, J.L. and W.-C. Shen, *Cationic and amphipathic cell-penetrating peptides (CPPs): Their structures and in vivo studies in drug delivery.* Front. Chem. Sci. Eng., 2015. **9**(4): p. 407-427.

42. Bashyal, S., G. Noh, T. Keum, Y.W. Choi, and S. Lee, *Cell penetrating peptides as an innovative approach for drug delivery; then, present and the future.* J. Pharm. Invest., 2016. **46**(3): p. 205-220.

43. Balogh, B., M. Ivanczi, B. Nizami, T. Beke-Somfai, and I.M. Mandity, *ConjuPepDB: a database of peptide-drug conjugates.* Nucleic. Acids. Res., 2021. **49**(D1): p. D1102-D1112.

44. Liotard, J.F., M. Mehiri, A. Di Giorgio, N. Boggetto, M. Reboud-Ravaux, A.M. Aubertin, R. Condom, and N. Patino, *AZT and AZT-monophosphate prodrugs incorporating HIV-protease substrate fragment: synthesis and evaluation as specific drug delivery systems.* Antivir. Chem. Chemother., 2006. **17**(4): p. 193-213.

45. Nasrolahi Shirazi, A., R. Tiwari, B.S. Chhikara, D. Mandal, and K. Parang, *Design and biological evaluation of cell-penetrating peptide-doxorubicin conjugates as prodrugs.* Mol. Pharm., 2013. **10**(2): p. 488-499.

46. Vale, N., C. Alves, V. Sharma, D.F. Lazaro, S. Silva, P. Gomes, and T.F. Outeiro, *A new MAP-Rasagiline conjugate reduces alpha-synuclein inclusion formation in a cell model.* Pharmacol. Rep., 2020. **72**(2): p. 456-464.

47. Polgar, L., E. Lajko, P. Soos, O. Lang, M. Manea, B. Merkely, G. Mezo, and L. Kohidai, *Drug targeting to decrease cardiotoxicity - determination of the cytotoxic effect of GnRH-based conjugates containing doxorubicin, daunorubicin and methotrexate on human cardiomyocytes and endothelial cells.* Beilstein J. Org. Chem., 2018. **14**: p. 1583-1594.

48. Lajko, E., S. Spring, R. Hegedus, B. Biri-Kovacs, S. Ingebrandt, G. Mezo, and L. Kohidai, *Comparative cell biological study of in vitro antitumor and antimetastatic activity on melanoma cells of GnRH-III-containing conjugates modified with short-chain fatty acids.* Beilstein. J. Org. Chem., 2018. **14**: p. 2495-2509.

49. Bera, S., R.K. Kar, S. Mondal, K. Pahan, and A. Bhunia, *Structural Elucidation of the Cell-Penetrating Penetratin Peptide in Model Membranes at the Atomic*

Level: Probing Hydrophobic Interactions in the Blood-Brain Barrier. Biochemistry, 2016. **55**(35): p. 4982-4996.

50.   Verma, R.P. and C. Hansch, *Cytotoxicity of organic compounds against ovarian cancer cells: a quantitative structure-activity relationship study.* Mol Pharm, 2006. **3**(4): p. 441-50.

51.   Machatha, S.G., P. Bustamante, and S.H. Yalkowsky, *Deviation from linearity of drug solubility in ethanol/water mixtures.* Int J Pharm, 2004. **283**(1-2): p. 83-8.

52.   Kong, Z., D. Sun, Y. Jiang, and Y. Hu, *Design, synthesis, and evaluation of 1, 4-benzodioxan-substituted chalcones as selective and reversible inhibitors of human monoamine oxidase B.* J Enzyme Inhib Med Chem, 2020. **35**(1): p. 1513-1523.

53.   Jorgensen, W.L. and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.* J. Am. Chem. Soc., 1988. **110**(6): p. 1657-1666.

54.   Jorgensen, W.L., D.S. Maxwell, and J. Tirado-Rives, *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids.* J. Am. Chem. Soc., 1996. **118**(45): p. 11225-11236.

55.   Shivakumar, D., J. Williams, Y. Wu, W. Damm, J. Shelley, and W. Sherman, *Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field.* J. Chem. Theory Comput., 2010. **6**(5): p. 1509-1519.

56.   Harder, E., W. Damm, J. Maple, C. Wu, M. Reboul, J.Y. Xiang, L. Wang, D. Lupyan, M.K. Dahlgren, J.L. Knight, J.W. Kaus, D.S. Cerutti, G. Krilov, W.L. Jorgensen, R. Abel, and R.A. Friesner, *OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins.* J. Chem. Theory. Comput., 2016. **12**(1): p. 281-296.

57.   Marrink, S.J., V. Corradi, P.C.T. Souza, H.I. Ingolfsson, D.P. Tieleman, and M.S.P. Sansom, *Computational Modeling of Realistic Cell Membranes.* Chem. Rev., 2019. **119**(9): p. 6184-6226.

58.   Reid, L.M., C.S. Verma, and J.W. Essex, *The role of molecular simulations in understanding the mechanisms of cell-penetrating peptides.* Drug Discov. Today, 2019. **24**(9): p. 1821-1835.

59. Kabelka, I., R. Brozek, and R. Vacha, *Selecting Collective Variables and Free-Energy Methods for Peptide Translocation across Membranes.* J. Chem. Inf. Model., 2021. **61**(2): p. 819-830.

60. de Oliveira, E.C.L., K.S. da Costa, P.S. Taube, A.H. Lima, and C.d.S.d.S. Junior, *Biological Membrane-Penetrating Peptides: Computational Prediction and Applications.* Front. Cell. Infect. Microbiol., 2022. **12**: p. 838259.

61. Jobin, M.-L., L. Vamparys, R. Deniau, A. Grelard, C.D. Mackereth, P.F.J. Fuchs, and I.D. Alves, *Biophysical insight on the membrane insertion of an arginine-rich cell-penetrating peptide.* Int. J. Mol. Sci., 2019. **20**(18): p. 4441.

62. Huang, K. and A.E. Garcia, *Free energy of translocating an arginine-rich cell-penetrating peptide across a lipid bilayer suggests pore formation.* Biophys. J., 2013. **104**(2): p. 412-420.

63. Lensink, M.F., B. Christiaens, J. Vandekerckhove, A. Prochiantz, and M. Rosseneu, *Penetratin-membrane association: W48/R52/W56 shield the peptide from the aqueous phase.* Biophys. J., 2005. **88**(2): p. 939-952.

64. Herce, H.D. and A.E. Garcia, *Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes.* Proc. Natl. Acad. Sci. USA, 2007. **104**(52): p. 20805-20810.

65. Herce, H.D., A.E. Garcia, J. Litt, R.S. Kane, P. Martin, N. Enrique, A. Rebolledo, and V. Milesi, *Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides.* Biophys. J., 2009. **97**(7): p. 1917-1925.

66. He, X., M. Lin, B. Sha, S. Feng, X. Shi, Z. Qu, and F. Xu, *Coarse-grained molecular dynamics studies of the translocation mechanism of polyarginines across asymmetric membrane under tension.* Sci. Rep., 2015. **5**: p. 12808.

67. Bennett, W.F., C.K. Hong, Y. Wang, and D.P. Tieleman, *Antimicrobial Peptide Simulations and the Influence of Force Field on the Free Energy for Pore Formation in Lipid Bilayers.* J. Chem. Theory Comput., 2016. **12**(9): p. 4524-4533.

68. Alaybeyoglu, B., B. Sariyar Akbulut, and E. Ozkirimli, *Insights into membrane translocation of the cell-penetrating peptide pVEC from molecular dynamics calculations.* J. Biomol. Struct. Dyn., 2016. **34**(11): p. 2387-2398.

69. Via, M.A., J. Klug, N. Wilke, L.S. Mayorga, and M.G. Del Popolo, *The interfacial electrostatic potential modulates the insertion of cell-penetrating peptides into lipid bilayers.* Phys. Chem. Chem. Phys., 2018. **20**(7): p. 5180-5189.

70. Ulmschneider, J.P., *Charged Antimicrobial Peptides Can Translocate across Membranes without Forming Channel-like Pores.* Biophys. J., 2017. **113**(1): p. 73-81.

71. Yao, C., Z. Kang, B. Yu, Q. Chen, Y. Liu, and Q. Wang, *All-Factor Analysis and Correlations on the Transmembrane Process for Arginine-Rich Cell-Penetrating Peptides.* Langmuir, 2019. **35**(28): p. 9286-9296.

72. Tran, D.P., S. Tada, A. Yumoto, A. Kitao, Y. Ito, T. Uzawa, and K. Tsuda, *Using molecular dynamics simulations to prioritize and understand AI-generated cell penetrating peptides.* Sci. Rep., 2021. **11**(1): p. 10630.

73. Gimenez-Dejoz, J. and K. Numata, *Molecular dynamics study of the internalization of cell-penetrating peptides containing unnatural amino acids across membranes.* Nanoscale Adv., 2022. **4**(2): p. 397-407.

74. Bienfait, B. and P. Ertl, *JSME: a free molecule editor in JavaScript.* J. Cheminform., 2013. **5**: p. 24.

75. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.* Nucleic. Acids. Res., 2000. **28**(1): p. 235-242.

76. Sastry, G.M., M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.* J. Comput. Aided. Mol. Des., 2013. **27**(3): p. 221-234.

77. Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, *Comparison of simple potential functions for simulating liquid water.* J. Chem. Phys., 1983. **79**(2): p. 926-935.

78. Kurki, M., A. Poso, P. Bartos, and M.S. Miettinen, *Structure of POPC Lipid Bilayers in OPLS3e Force Field.* J. Chem. Inf. Model., 2022. **62**(24): p. 6462-6474.

79. Bowers, K.J., E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, J.L. Klepeis, I. Kolossvary, M.A. Moraes, F.D. Sacerdoti, J.K. Salmon, Y. Shan, and D.E. Shaw, *Scalable Algorithms for Molecular Dynamics Simulations on*

Commodity Clusters, in *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*. 2006: Tampa, Florida.

80.   Ivanova, N. and A. Ivanova, *Testing the limits of model membrane simulations-bilayer composition and pressure scaling.* J. Comput. Chem., 2018. **39**(8): p. 387-396.

81.   Jacobson, M.P., R.A. Friesner, Z. Xiang, and B. Honig, *On the Role of the Crystal Environment in Determining Protein Side-chain Conformations.* J. Mol. Biol., 2002. **320**(3): p. 597-608.

82.   Jacobson, M.P., D.L. Pincus, C.S. Rapp, T.J.F. Day, B. Honig, D.E. Shaw, and R.A. Friesner, *A hierarchical approach to all-atom protein loop prediction.* Proteins: Struct., Funct., Bioinf., 2004. **55**(2): p. 351-367.

9. Bibliography of the candidate's publications

I.   Balázs Balogh, Márton Ivánczi, Bilal Nizami, Tamás Beke-Somfai, István M. Mándity:
      ConjuPepDB: a database of peptide-drug conjugates.
      Nucleic Acids Res. **2021.** *49*(D1): D1102-D1112.   IF.:19,16

II.   Márton Ivánczi, Balázs Balogh, Loretta Kis, István M. Mándity:
      Molecular Dynamics Simulations of Drug-Conjugated Cell-Penetrating Peptides.
      Pharmaceuticals **2023**. *16*(9).   IF.:4,6

10. Acknowledgments

I am grateful to my supervisor, Dr. István Mándity, Director of the Department of Organic Chemistry, Semmelweis University for his guidance of my work, his constructive criticism, his helpful advice and his inspiring ideas.

My special thanks are also due to my co-supervisor, Dr. Balázs Balogh, for his continuous support, his unrestricted supervision of my work and his constructive criticism as well.

I am also grateful to Prof. István Kolossváry for helpful discussions.