ELSEVIER

Check for updates

Feature

CPPsite3: An updated large repository of experimentally validated cell-penetrating peptides

Nisha Bajiya [1,3], Sohana Najrin [2,3], Pankaj Kumar [1,3], Shubham Choudhury [1], Ritu Tomer [1], Gajendra P.S. Raghava [1,*]

[1] Department of Computational Biology, Indraprastha Institute of Information Technology, Okhla Phase 3, New Delhi 110020, India
[2] Jamia Hamdard University, Block D, Hamdard Nagar, New Delhi 110062, India

Over the past two decades, cell-penetrating peptides (CPPs) have emerged as key intracellular drug delivery vehicles. This review provides a systematic overview of advancements in CPP research, encompassing both experimental techniques and computational resources crucial for their discovery. Our primary focus is on CPPsite3, a meticulously curated repository that currently houses 6788 entries, representing 4285 unique CPPs. This updated version builds upon CPPsite and CPPsite2, both of which have been extensively utilized by the scientific community for designing CPPs and developing robust prediction methods. Furthermore, we explore the clinical applications of CPPs, specifically highlighting those currently undergoing clinical use. This comprehensive review aims to be a vital resource for the drug delivery research community.

Keywords: cell-penetrating peptides; therapeutic agents; delivery carriers; CPPsite3 database

Introduction

Many therapeutic compounds, especially large molecules like DNA, have difficulty with intracellular delivery due to membrane barriers and lysosomal degradation after endocytosis, limiting their *in vivo* bioavailability despite promising *in vitro* results.[(p1),(p2)] Advancing the precision drug industry requires an effective delivery agent to cross the lipophilic membrane and transport it to the targeted action location. Over the past 30 years, CPPs have primarily been used in basic and preclinical research as a transporting carrier. As depicted in Figure 1, CPPs are short peptides, fewer than 30 amino acids, that are mainly positively charged and derived from both natural and non-natural sources that enter via different mechanisms. Their ability to enter cells and facilitate the passage of drugs or CPP–cargo complexes across the plasma membrane makes them potentially useful for disease diagnosis and treatment.[(p3)] They have been widely used as a transport vehicle for delivering various impermeable molecules with different properties, such as drugs, imaging agents, nanoparticles, oligonucleotides, proteins and other functional peptides[(p4),(p5),(p6),(p7)] not only in animals but also in plants.[(p8)] These applications include conditions related to the central nervous system, diabetes, immunotherapies,[(p9)] otologic and ocular disorders, cancer,[(p10)] and inflammation.[(p11)] Since the initial discoveries of these

[1] Authors contributed equally.

1359-6446/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
https://doi.org/10.1016/j.drudis.2025.104421

cell-penetrating abilities in the late 1980s and early 1990s, the field has significantly advanced, moving from understanding basic uptake to sophisticated applications in targeted drug and gene delivery, with current efforts focusing on rational design, 'smart' CPPs, artificial intelligence (AI)-driven optimization, and increasing clinical translation for enhanced delivery, reduced toxicity, and targeted therapies.

Experimental techniques for CPP characterization

A number of experimental techniques that confirm a peptide's cell-penetrating capabilities are available for the identification of CPP attributes.[p12] They can be classified into three categories: *in vitro* techniques, *in vivo* techniques and mechanistic assays. The first category includes methods like qualitative identification such as confocal microscopy and other types of microscopic techniques for tracking the subcellular localization,[p13],[p14],[p15] and fluorescence labeling [e.g. fluorescein isothiocyanate (FITC), rhodamine] to track the pathway of CPP–cargo complexes and quantitative analysis such as flow cytometry,[p16],[p17] fluorescence-activated cell sorting (FACS) for quantifying any intracellular biochemical entity,[p18] mass spectrometry, and HPLC for assessing purity and chemical validation,[p19],[p20] and matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS).[p21] The second category includes various animal models (e.g. mice, zebrafish) to evaluate *in vivo* bioavailability, targeting efficiency, and toxicity; and biodistribution assays using radiolabeled CPPs to help assess organ/tissue-specific delivery.[p22],[p23] The third category includes mechanistic assays that indirectly determine CPP properties such as activity assays [e.g. luciferase activity, Cre recombinase, chloroalkane penetration assay (CAPA)], which are only indirect estimations of the amount of cargo internalized by CPPs[p24],[p25] and mutagenesis studies for analyzing the contribution of specific residues to uptake efficiency. Similarly, their mode of internalization into the cells involves various endocytotic (clathrin-mediated, caveolin-mediated, macropinocytosis), non-endocytotic and direct translocation pathways, determined experimentally using various pathway inhibitors to interpret the uptake route, explained in detail in several review papers.[p26],[p27]

Computational resources for CPPs

The experimental techniques for the identification of CPPs are the gold standard, which identify CPPs with high precision. These wet lab experiments are challenging due to their high cost, time-consuming nature, and labor-intensive requirements. These limitations have encouraged researchers to explore alternative discovery strategies, particularly computational methods for predicting CPPs within proteins. In the era of AI, machine learning (ML) techniques have emerged as powerful tools for prediction and classification. Novel CPPs have been successfully identified by ML-driven predictions in recent investigations, and they are currently being sought for experimental evaluation.[p28] However, training robust ML models necessitates comprehensive datasets of existing CPPs, as models are built upon established knowledge. Historically, the CPPs discovered through experimental means were scattered across numerous research papers, precluding their availability from a single, centralized source. This fragmentation presented a major hurdle for researchers aiming to compile datasets for ML model development. Thus, to create a dataset for building ML-based models, one needs to compile these CPPs.

To address this challenge, Gautam *et al.*[p29] made a seminal systematic attempt to compile CPPs from literature, databases, and other resources, culminating in the creation of CPPsite. This initial database provided detailed information on 843 experimentally validated CPPs and has since been widely adopted by the scientific community. Researchers extensively utilized CPPsite data to derive datasets for predicting, designing, and scanning for CPPs within proteins. Recognizing the need for up-to-date information, CPPsite was subsequently updated in 2015 to CPPsite2,[p30] expanding its coverage to include over 1850 CPPs, more than doubling from its initial version. Beyond CPPsite and CPPsite2, a few other databases contain related information based on different purposes like POSEIDON,[p31] which includes quantitative uptake values of CPP; and CycPeptMPDB,[p32] which contains information about the membrane permeability of cyclic peptides. Table 1 provides a summary of these available repositories and the specific types of information they contain.

In the past, a number of prediction tools have been developed to predict CPP efficiently and accurately using various machine and deep-learning methods. Table 2 depicts the key computational tools that aid in CPP identification and characterization, varying in the algorithm utilized from support vector machines to ensemble learning and the recently rising

TABLE 1

List of available CPP databases.

Database	Year	No. of entries	Information type	Web link
CPPsite	2012	843	Includes detailed information of CPPs like chirality, origin, nature of peptide, subcellular localization, uptake efficiency, uptake mechanism, hydrophobicity, amino acid frequency, and composition	https://crdd.osdd.net/raghava/cppsite1/
CPPsite 2.0	2016	1855	Includes detailed information of CPPs as CPPsite, and additional information about the structure of peptides	https://crdd.osdd.net/raghava/cppsite/
CycPeptMPDB	2023	7334	Database of cyclic peptide membrane permeability	https://cycpeptmpdb.com
POSEIDON	2024	2300	Provides information about experimental quantitative uptake values and physicochemical properties of peptides	https://moreiralab.com/resources/poseidon/

TABLE 2

List of computational tools employed in the field of CPP prediction.

Tools	Year	ML employed	Key features	Web server availability	Refs
CellPPD	2013	SVM	Predicts and designs CPPs	https://crdd.osdd.net/raghava/cellppd/	(p33)
CPPpred	2013	N-to-1 neural networks (N1-NNs)	Predicts CPPs	https://bioware.ucd.ie/cpppred	(p34)
C2Pred	2016	SVM	Predicts CPPs and employs feature selection	https://lin-group.cn/server/C2Pred	(p35)
SkipCPP-Pred	2017	RF	Predicts CPPs using adaptive k-skip-n-gram	https://server.malab.cn/SkipCPP-Pred/Index.html (NW)	(p36)
CPPred-RF	2017	RF	Predicts CPPs and uptake efficiency	https://server.malab.cn/CPPred-RF (NW)	(p37)
CellPPD-Mod	2018	RF	Predict CPPs from the tertiary structure of natural and modified peptides	https://webs.iiitd.edu.in/raghava/cellppdmod/	(p38)
MLCPP	2018	RF, SVM, ERT, and *k*-NN	Two-layer prediction framework (predicts CPPs and uptake efficiency)	https://www.thegleelab.org/MLCPP (NW)	(p39)
KELM-CPPpred	2018	Kernel ELM	Predicts CPPs	https://sairam.people.iitgn.ac.in/KELM-CPPpred.html	(p40)
G-DipC	2020	XGBoost	Distinguish the category of cargo carried by CPPs using the general dipeptide composition (G-DipC)	NA	(p41)
CPPred-FL	2020	RF	Predicts CPPs using 19 probabilistic features	https://server.malab.cn/CPPred-FL (NW)	(p42)
StackCPPred	2020	Stacking-based ML (Base – XGBoost, LightGBM, SVM, *k*-NN and RF; and meta – SVM)	Predicts CPPs and uptake efficiency using RECM-composition, PseRECM and RECM–DWT	NA	(p43)
TargetCPP	2020	Gradient boost decision tree	Predicts CPPs using features CPSR, CTD, SAAC, and ITF	NA	(p44)
BChemRF-CPPred	2021	ANN, SVM, and a Gaussian process classifier	Predict synthetic and natural CPP structures	https://comptools.linc.ufpa.br/BChemRF-CPPred (NW)	(p45)
DeepCPPred	2022	Cascade deep-forest (CDF) classifier	Discrimination of CPPs and their uptake efficiencies	NA	(p46)
MLCPP 2.0	2022	119 baseline models (17 encodings × 7 ML classifiers)	Two-layer prediction framework (Predicts CPPs and uptake efficiency)	https://balalab-skku.org/mlcpp2/	(p47)
SiameseCPP	2023	Siamese neural network	Predicts CPPs	NA	(p48)
AiCPP	2023	Ensemble learning (LSTM)	Predicts CPPs	NA	(p49)
PractiCPP	2024	deep-learning framework	Predicts CPPs using imbalanced data	NA	(p50)
pLM4CPPs	2025	Protein language models (PLMs)	Predicts CPPs	https://ry2acnp6ep.us-east-1.awsapprunner.com/	(p51)
CPPCGM	2025	PLMs	Identify and generate novel CPPs	NA	(p52)

ANN, artificial neural network; CPSR, composite protein sequence representation; CTD, composition, transition and distribution; DWT, discrete wavelet transform; ELM, extreme learning machine; ERT, ensemble representation learning; ITF, intelligent talent finder; LSTM, long short-term memory; NA, not available; NW, not working; PseRECM, pseudo residue pairwise energy content matrix; RF, random forest; SAAC, split amino acid composition; SVM, support vector machine.

models employing natural language processing techniques. All of these tools heavily depend on databases for creating datasets, and the majority of them have derived their training and validation data from CPPsite and CPPsite2, which contain information only up to 2015. Relying on outdated data can limit the reliability of these tools, and the growth of CPPs in the last 9 years has been tremendous, as shown in Figure 3. Therefore, the availability of an up-to-date and comprehensive database is critical. Such a database would accelerate the discovery of novel CPPs, facilitate the development of more sophisticated prediction models, support targeted drug deliveries, and serve as a valuable educational resource to the scientific community.

CPPsite3: an updated repository of CPPs

Despite being updated only in 2015, CPPsite2 is still actively used by the scientific community. However, significant technical advancements and the growing importance of CPPs have concurrently led to the discovery of numerous additional experimentally validated CPPs. To overcome this challenge, the group updated CPPsite2 to incorporate updated information. The new version is called CPPsite3, and contains updated information up to 2024. This resource offers integrated tools for effective data browsing, searching, and analysis, enhancing its utility for researchers. The responsive web-server design ensures seamless compatibility across all devices, including smartphones and tablets, maximizing user convenience. In this review, we will provide an extensive overview of this database.

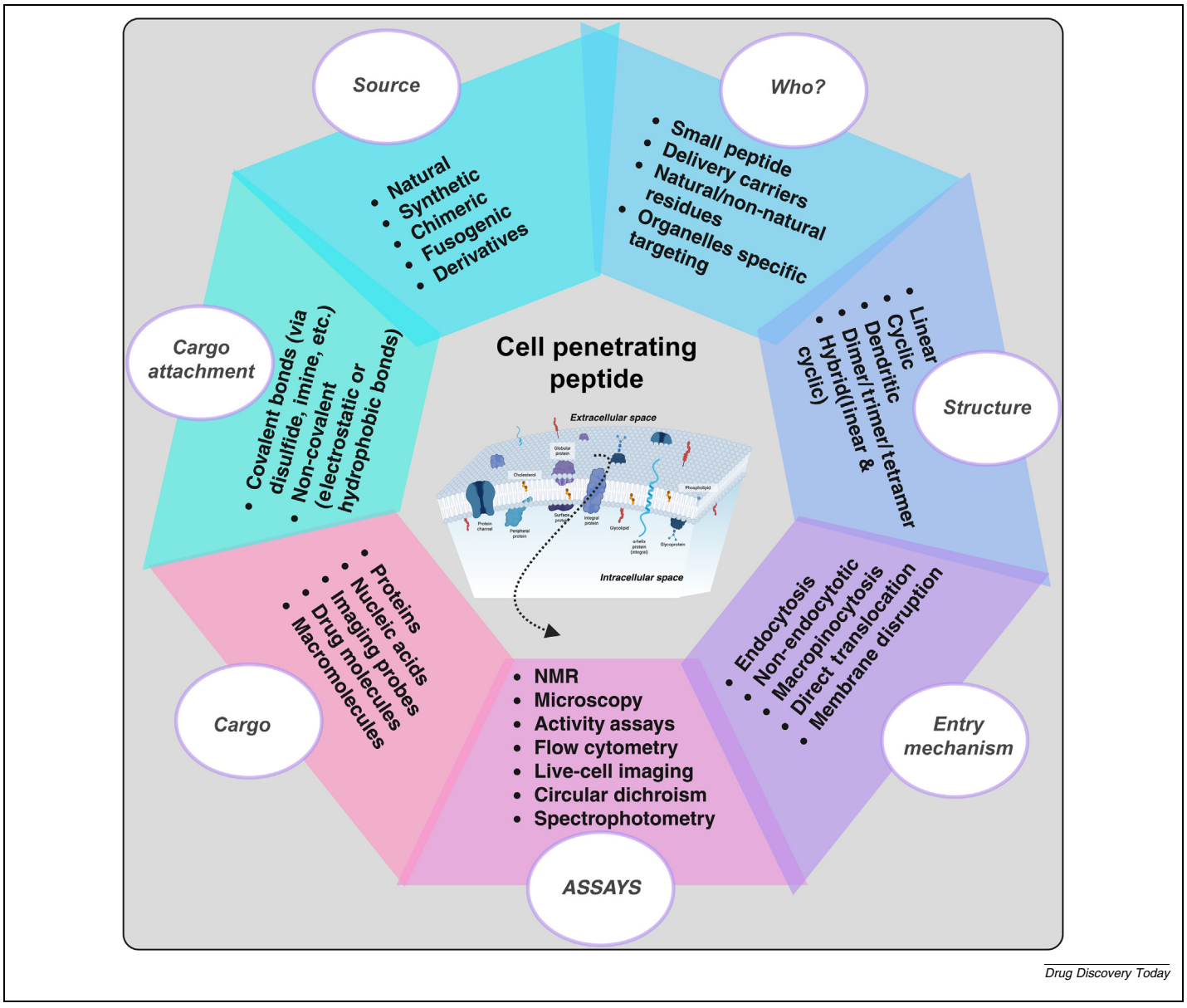

FIGURE 1

A comprehensive theoretical framework of cell penetrating peptides outlining their origin, structural diversity, cellular uptake pathways, identification assays, and the nature of delivered cargos. Figure created with BioRender.

Architecture and organization of CPPsite3

The CPPsite3 database architecture uses an Apache HTTP server (version 2.4.63) to process web requests, and MySQL (version 8.0.41) is the stable backend system for effective data management. The user-friendly and adaptable front end is built with the latest web technologies that are open-source and platform-independent, such as HTML (version 5), CSS (bootstrap@5.3.5), PHP (version 5.6.40), and JavaScript (version 12.22.9). This ensures smooth compatibility across a range of devices, including desktops, mobile phones, and tablets, and a unified interface enables user interaction by utilizing the capabilities of the PHP scripting language. The overall framework of CPPsite3, as illustrated in Figure 2, is built upon this stable and widely adopted technology stack.

Data curation approach

To search CPPs from the literature, the search term 'cell penetrating peptides' was used to search PubMed from July 2015 to April 2024 to extract relevant research publications. Initially, 2361 papers were obtained, which were filtered by removing review articles, prediction methodologies, and book chapters. Remaining research papers were thoroughly scanned and manually examined, and relevant CPP details were manually extracted. Only studies providing experimental validation of CPPs and their analogues were selected for curation. To further broaden the scope, the references in relevant review articles were also explored, and the relevant articles were included. This rigorous process culminated

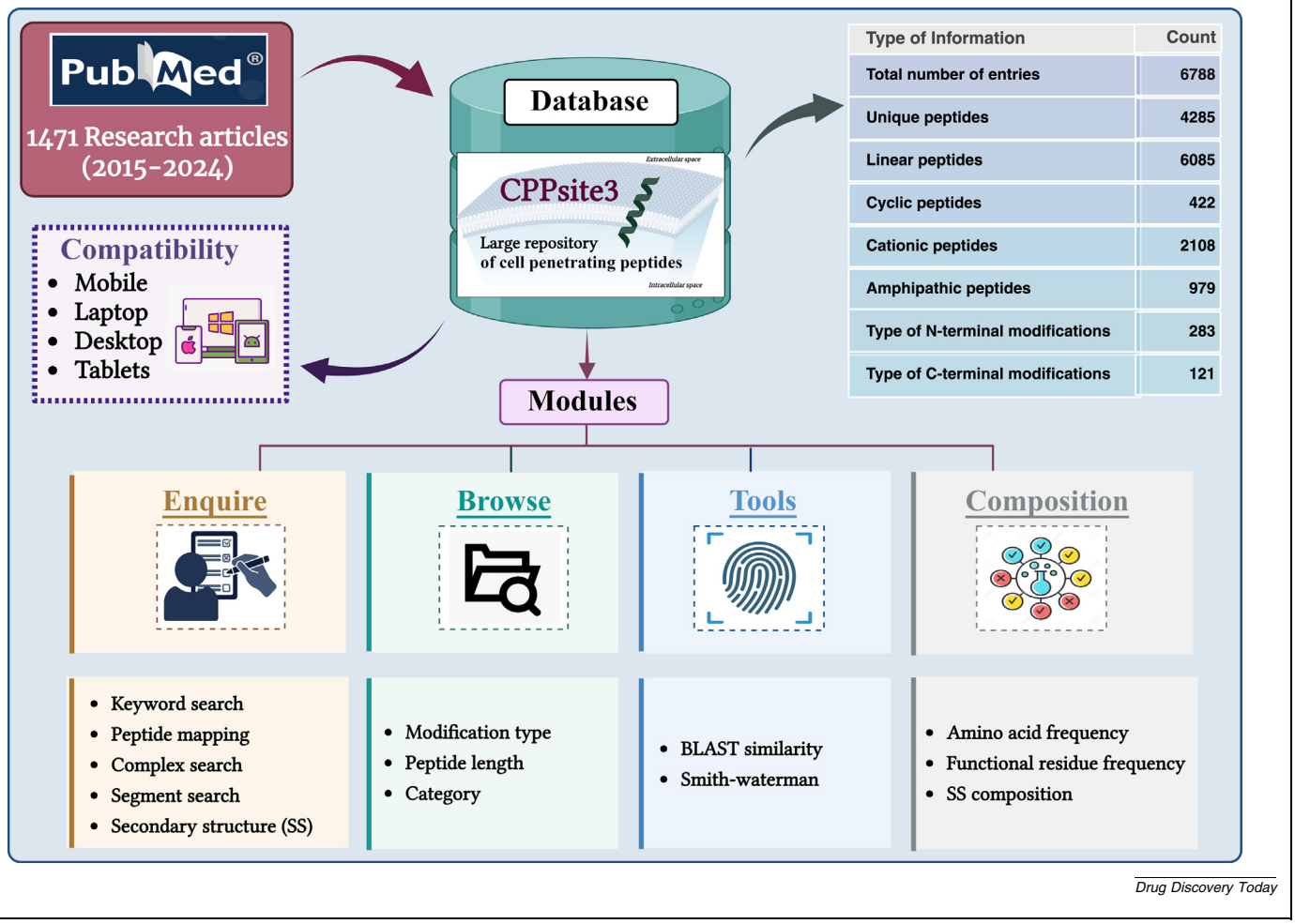

FIGURE 2
Architectural framework of the CPPsite3 database. Figure created with BioRender.

in identifying 1471 unique PubMed IDs, the data from which form the core of CPPsite3. The retrieved data included key components, including the nature, chirality, subcellular localization, end modifications, cell penetration effectiveness, and features of the *in vitro* or *in vivo* model systems (e.g. particularly cell lines or animal models). A thorough manual reading method was used to collect all of the data, which were then tabulated for ease of understanding and access for further studies. We included each instance of a similar CPP, even if it was evaluated differently in separate studies, to present all of the diverse findings that had been reported and to make sure the database was up to date and thorough. Consequently, CPPsite3 comprises a total of 6788 entries, integrating 1855 records from earlier versions with 4933 novel entries extracted from recent publications. This substantial update introduces 4143 novel CPP sequences, considerably enriching the database and enhancing its value for CPP researchers and the broader biomedical science community.

Data content of CPPsite3

Table 3 provides a detailed overview of the primary data fields incorporated into the CPPsite3. Each entry within the database is directly linked to its corresponding PubMed ID (PMID), enabling users to easily access the original research publication from which the information was extracted.

Statistics of CPPsite3

The CPPsite3 contains 7688 entries, of which 1855 have been taken from CPPsite3, and the remaining 4933 entries are manually curated from 1471 research articles on PubMed from the year 2015 to 2024. Figure 3 depicts the number of entries added from each year in the database. The increase in entries from the previous versions was more than threefold. To be precise, the number of entries increased by approximately 3.14 times. In Figure 4a, the distribution of linear peptides constitutes the majority (approximately 90.8%), suggesting their prevalence in research or natural systems, whereas cyclic (approximately 6.3%), stapled (approximately 1.6%), hybrid (linear and cyclic, approximately 0.7%), and bicyclic (approximately 0.5%) peptides represent greater structural complexity and often enhanced stability, conformation, and ultimately, cellular uptake. These nonlinear architectures likely enhance resistance to proteolysis and can preorganize the peptide into a more favorable conformation for membrane interaction.[p53),(p54)]

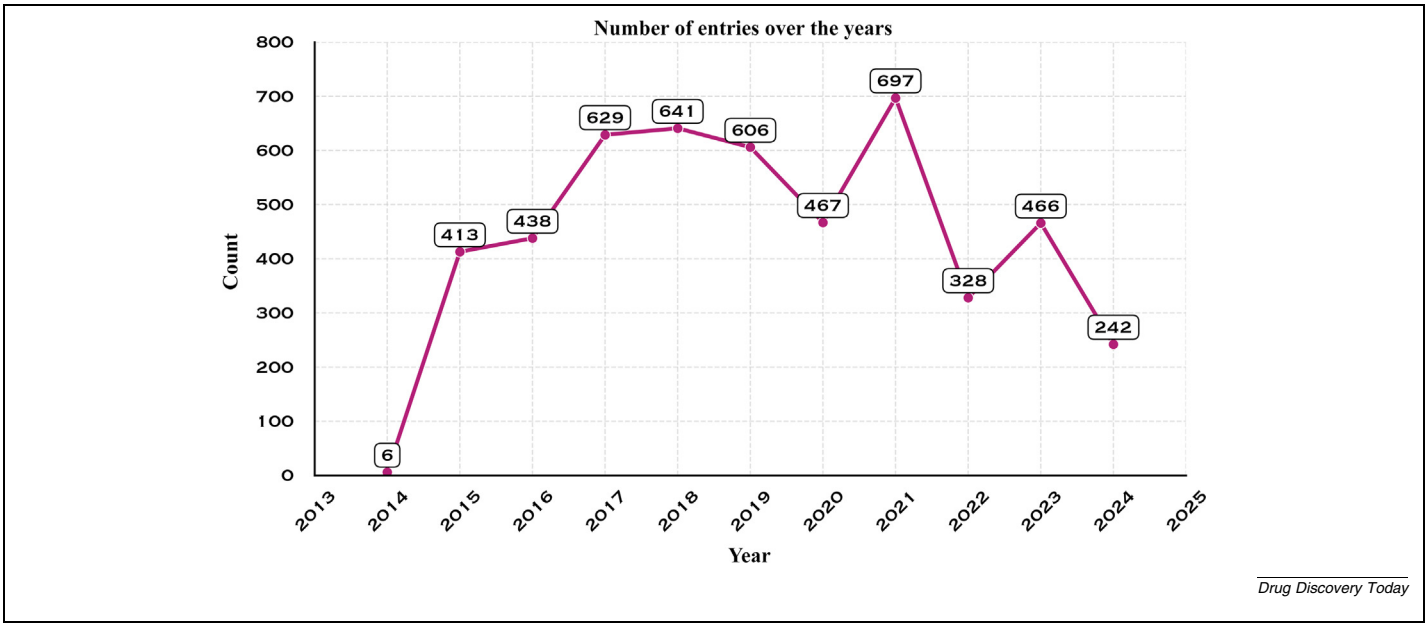

FIGURE 3
Year-wise trend of database entries. Figure created with BioRender.

TABLE 3

Primary data fields in CPPsite 3.0 and their descriptions.

S. No.	Categories	Description	Example
1	Name	Provides the name of the peptide referenced in the literature.	Penetratin
2	Sequence	Representing the peptide's amino acid sequence determines both its chemical makeup and its biological action.	RRRGADFASDLF
3	Chirality	Represents the chirality of the peptide, influencing its stability and cellular interactions	D/L/mix
4	Source	Reveals the source to which it belongs	Protein-derived, synthetic, chimeric
5	Linearity	Determining the linearity of the peptide enhances its membrane-crossing efficiency and resistance to breakdown.	Linear/cyclic/hybrid
6	Nature	It indicates the nature of the peptide, important in determining how it interacts with negatively charged cell membranes and influences its penetration behavior.	Cationic/amphipathic/anionic/ hydrophilic, etc.
7	Subcellular localization	Provides information about where the peptide ends up inside the cell, which is key to its functional impact	Nucleus/cytoplasm/mitochondria, etc.
8	N-terminal modification	Contains information about modification at the start of the peptide, which can boost stability or uptake.	Acetylation, etc.
9	C-terminal modification	Contains information about the modification of the end of the peptide.	Amidation, etc.
10	Chemical modification	Provides information about any chemical modification within the sequence of the peptide, added to improve performance or specificity	PEGylation, non-natural amino acid insertion, etc.
11	Uptake efficacy	Represents the relative efficiency of peptides in cell penetration compared to other peptides in experimental settings	Twofold/high/lower than, etc.
12	Uptake mechanism	Provides information about the mechanism of CPP uptake used by the peptide.	Endocytosis/direct translocation, etc.
13	Experimental type	Provides context for the experimental conditions/model system under which CPP properties were tested	*In vivo/in vitro/ex vivo*
14	Cell line used	Provides information on the specific cell line type used for validation of the peptide's effect	HeLa/Chinese hamster ovary, etc.
15	Animal model	Describes the animal model used to test CPP and validate its real-world potential.	BALB/c mice/zebrafish, etc.
16	Cargo types	Defines what the peptide can transport as a cargo in various model systems	Nucleic acids/proteins/drugs, etc.

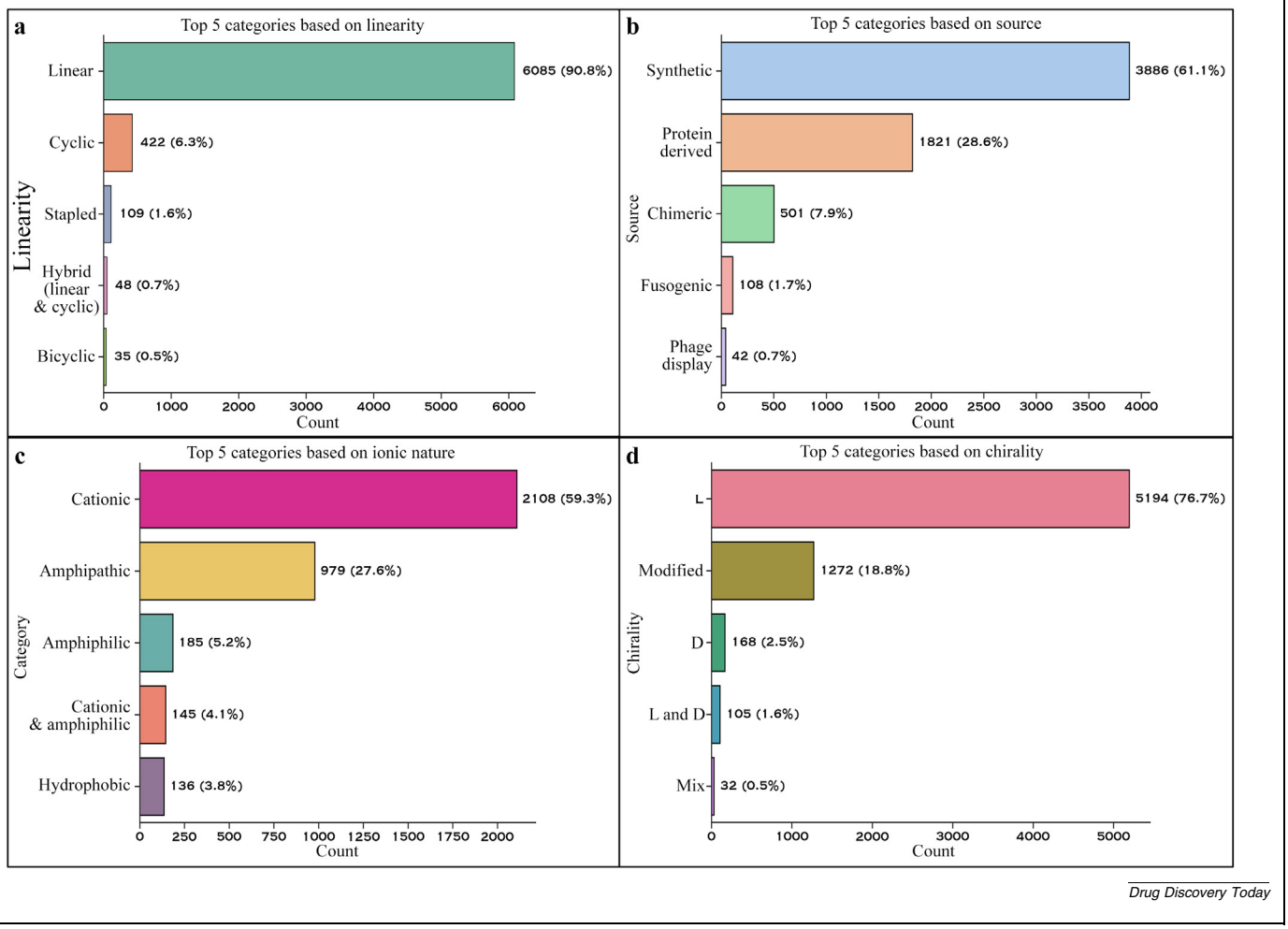

FIGURE 4

Statistics of top five categories based on **(a)** linearity; **(b)** source, **(c)** ionic nature; and **(d)** chirality in CPPsite3 based on the number of entries. Figure created with BioRender.

Similarly, from Figure 4b, most CPPs are synthetically sourced (61.1%), reflecting the importance of *de novo* design and optimizing CPP sequences for specific applications. However, protein-derived peptides (28.6%) also form a substantial portion, indicating the continued relevance of naturally occurring sequences and their bioactivity. The presence of chimeric (7.9%), fusogenic (1.7%), and phage display-derived (0.7%) CPPs showcases the diverse methodologies employed in CPP discovery and engineering. In terms of ionic nature, a significant majority of the CPPs in the database are cationic (approximately 59.3%), as shown in Figure 4c. This observation suggests that they are linked to their interactions with negatively charged biological membranes, facilitating cell penetration and antimicrobial activity. The distribution of peptides based on chirality in Figure 4d reveals a strong bias towards L-amino acid-based CPPs (76.7%), reflecting the natural abundance of L-amino acids in biological systems. However, the presence of modified and D-peptides (18.8% and 2.5%, respectively) points to efforts to explore alternative chiralities and chemical modifications for enhanced stability, resistance to enzymatic degradation, and novel functions. Moreover, the database also holds a plethora of modification types internally or at the terminal of the peptides, most of which include biotinylation, amidation, and unnatural residue addition. The mode of uptake of the CPP includes the majorly endocytosis pathway (approximately 1105 entries), including different types such as caveolin-mediated, clathrin-mediated, and lipid rafts-mediated energy-dependent processes. Other uptake pathways used by CPPs to enter the cell include direct translocation, exocytosis, and receptor-mediated processes. In addition, it contains predicted tertiary structures of CPPs, which were predicted using PEPstrMOD.[p55]

Web interface

To facilitate data searching, browsing, and analysis with unparalleled simplicity and efficiency, a suite of robust tools and an intuitive, open-source, and adaptable online interface were developed for the benefit of the scientific community. The 'Enquire', 'Browse On', 'Composition', and 'Similarity' modules are the four main components of CPPsite3 that provide effortless data searching. It is available for users from the URL https://webs.iiitd.edu.in/raghava/cppsite3/.

(i) Enquire: The user may effectively search the database using five different search options: keyword search,

complex search, segment search, peptide mapping, and secondary structure search.

(ii) Browse On: It enables data browsing by (i) chemical modifications, which shows all kinds of alterations, such as N- and C-terminal changes and sequence modifications like the incorporation of non-natural residues; (ii) category, which groups CPPs according to the type, class, nature, and chirality of the peptide; and (iii) length, which returns entries within a user-selected length range.

(iii) Composition: In addition to residue compositions, several physicochemical characteristics of the peptides have been computed, including charge, hydrophobicity, amphipathicity, and isoelectric point.

(iv) Similarity: Two similarity tools have been incorporated: the Basic Local Alignment Search Tool (BLAST)[p56] and Smith–Waterman search algorithms[p57] to enable accurate peptide matching, in addition to a peptide alignment tool for comparing query peptides with database peptides.

Comparison with the previous version

CPPsite1 and CPPsite2 were developed in 2012 and 2015, respectively, comprising different information fields containing 1855 peptide entries. Since 2015, there has been no update, and no other database has been developed. The growing interest in penetrating peptides to deliver therapeutic molecules for precise targets have become prominent, therefore to fill this gap we have created an updated version named CPPsite3, in which we have included 4933 peptide entries compared to 1855 in the previous version with their names, sequences, nature, uptake efficiencies, modifications and experimental conditions. All the entries have been linked with their PubMed IDs to provide users with maximum accessibility to the source. Comparative statistics are shown in Figure 5.

Concluding remarks on CPPsite3

With the addition of more than 4000 new entries, CPPsite3 represents a significant expansion over its original release in 2012 and the subsequent update in 2015. It remains the only publicly accessible database offering comprehensive information on CPPs. Given the growing interest in precision medicine and targeted intracellular delivery of therapeutic agents, we believe this updated version will be highly valuable to the scientific community. However, due to limited available data, this release does not include certain key aspects of therapeutic peptides, such as toxicity, plasma half-life, and biodistribution. Another limitation is the exclusion of data from patents. We hope this database will be updated at regular intervals and will incorporate the missing information to provide a more complete and up-to-date resource.

Clinical applications of CPPs

From preclinical models to early clinical research, CPPs have advanced impressively, providing versatile methods for targeted delivery of diagnostics and treatments across a range of diseases and immunotherapy.[p53],[p58],[p59] In oncology, CPPs make it easier to precisely administer anticancer agents and imaging probes. For example, by activating the neuropilin 1-mediated CendR pathway, internalized arginylglycylaspartic acid cyclic peptide (iRGD), a cyclic, tumor-penetrating CPP, improves tumor accumulation and the effectiveness of chemother-

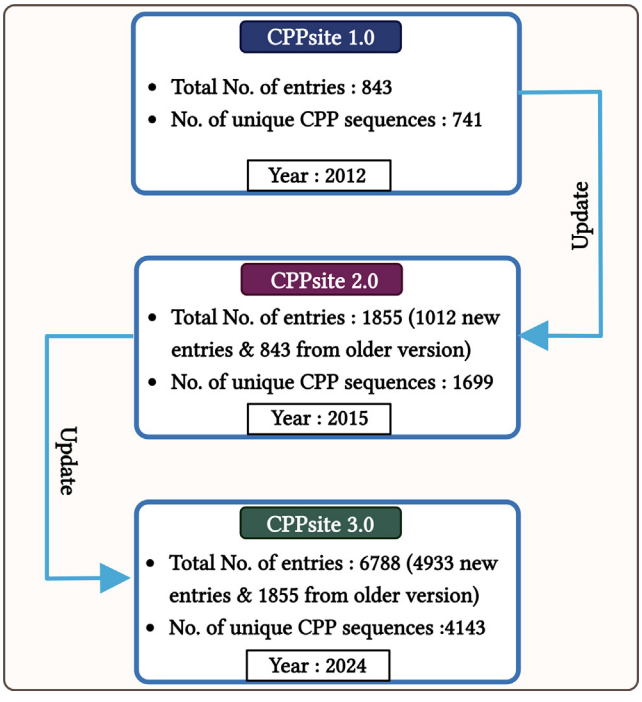

FIGURE 5

Comparison of data in three versions of the CPPsite database. Figure created with BioRender.

apeutics when co-administered. Its derivative certepide has already been designated as an FDA orphan drug and is undergoing Phase I/II trials in solid tumors (30647857).[p60] In macromolecular drug delivery, the chemotherapeutic prodrugs, DTS-108 (an SN38 conjugate) or CPP-fused botulinum toxin, have entered clinical trials, showing enhanced intracellular delivery and decreased systemic toxicity.[p61] Through covalent or non-covalent methods, CPPs have been linked with small interfering RNA (siRNA), antisense oligonucleotides, plasmid DNA, and more recently, mRNA. In preclinical models, cationic peptide–siRNA conjugates targeted HIV and hepatitis C *in vitro* and *in vivo*, whereas oligoarginine CPPs (e.g. R15) achieved ~80% tumor growth suppression via siHER 2 in ovarian cancer xenografts.[p62] In immunotherapy, CPPs help to enhance T-cell infiltration, antigen cross-presentation, and vaccine efficacy. CPP-modified peptides, for instance, are being assessed in therapeutic trials for glioblastoma, facilitating more effective dendritic cell absorption of antigens and enhancing humoral and cell-mediated immunity. Additionally, CPP-based systems have been evaluated for the delivery of siRNA, checkpoint inhibitors, and cytokines in autoimmune and inflammatory diseases.[p63] In diagnostic imaging and disease monitoring, CPP conjugates containing imaging probes demonstrate enhanced detection sensitivity by accumulating in tumors or inflammatory tissues.[p64] Because of their small size and affinity, these agents exhibit great tumor selectivity and quick clearance from non-diseased tissue. Lastly, in ocular therapeutics, in early clinical settings, TAT-linked c-Jun N-terminal kinase (JNK) inhibitors (e.g. XG-102/AM-111/D-JNKI-1) have shown promise in treating inner eye inflammation and sudden hearing loss; similarly, other applications in ocular treatments using CPPs have been reported in a review.[p65] Overall, despite the fact that no CPP-based medications have received full FDA clearance, CPPs have tremendous therapeutic promise. A number of Phase I/II clinical trials are now underway to target cancer, neurological disorders, and vaccines, signaling a significant shift from laboratory to patient.

Authors' contributions

The data were collected by **N.B.**, **S.N.**, **P.K.**, and **R.T**. **N.B.** curated and analyzed the data. The web server's front end was developed by **N.B**. **S.C.** developed the web server's backend. The manuscript was prepared by **N.B.**, **P.K.**, and **G.P.S.R**. **G. P.S.R.** conceived the idea and coordinated the project. All authors have reviewed and approved the final paper.

Declarations of interest

No interests are declared.

Acknowledgments

The authors are thankful to the Council of Scientific and Industrial Research (CSIR), Department of Biotechnology (DBT), Indraprastha Institute of Information Technology (IIITD), Department of Science and Technology (DST-INSPIRE) for fellowships and financial support and the Department of Computational Biology, IIITD, New Delhi, for infrastructure and facilities. We thank the DBT for funding the institute's infrastructure.

Funding

The grant (BT/PR40158/BTIS/137/24/2021) from the Department of Biotechnology (DBT), Government of India, India, supported this work.

Data availability

CPPsite3 is available for free at https://webs.iiitd.edu.in/raghava/cppsite3/ and is compatible with desktops, tablets, and smartphones. For downloading the database sequences and their PDB structure, available at https://webs.iiitd.edu.in/raghava/cppsite3/downloads.php. For the convenience of users of previous versions, CPPsite and CPPsite3 will continue to be available at https://webs.iiitd.edu.in/raghava/cppsite1/ and https://webs.iiitd.edu.in/raghava/cppsite/, respectively, so that access to these resources is not lost.

References

1. Craik DJ, Fairlie DP, Liras S, Price D. The future of peptide-based drugs. *Chem Biol Drug Des*. 2013;81:136–147. https://doi.org/10.1111/cbdd.12055.
2. Lee ACL, Harris JL, Khanna KK, Hong JH. A comprehensive review on current advances in peptide drug development and design. *Int J Mol Sci*. 2019;20:2383. https://doi.org/10.3390/ijms20102383.
3. Kardani K, Milani A, Shabani SH, Bolhassani A. Cell penetrating peptides: the potent multi-cargo intracellular carriers. *Expert Opin Drug Deliv*. 2019;16:1227–1258. https://doi.org/10.1080/17425247.2019.1676720.
4. Hejtmánková A, Váňová J, Španielová H. Cell-penetrating peptides in the intracellular delivery of viral nanoparticles. *Vitam Horm*. 2021;117:47–76. https://doi.org/10.1016/bs.vh.2021.06.010.
5. Hammond SM et al. Delivery of oligonucleotide-based therapeutics: challenges and opportunities. *EMBO Mol Med*. 2021;13, e13243. https://doi.org/10.15252/emmm.202013243.
6. Gori A, Lodigiani G, Colombarolli SG, Bergamaschi A, Vitali A. Cell penetrating peptides: classification, mechanisms, methods of study, and applications. *ChemMedChem*. 2023;18, e202300236. https://doi.org/10.1002/cmdc.202300236.
7. Falato L, Gestin M, Langel Ü. Cell-penetrating peptides delivering siRNAs: an overview. *Methods Mol Biol*. 2021;2282:329–352. https://doi.org/10.1007/978-1-0716-1298-9_18.
8. Kumar V, Chugh A. Cell-penetrating peptide for targeted macromolecule delivery into plant chloroplasts. *Appl Microbiol Biotechnol*. 2022;106:5249–5259. https://doi.org/10.1007/s00253-022-12053-3.
9. Du JJ et al. Applications of cell penetrating peptide-based drug delivery system in immunotherapy. *Front Immunol*. 2025;16, 1540192. https://doi.org/10.3389/fimmu.2025.1540192.
10. Bottens RA, Yamada T. Cell-penetrating peptides (CPPs) as therapeutic and diagnostic agents for cancer. *Cancers (Basel)*. 2022;14:5546. https://doi.org/10.3390/cancers14225546.
11. Xie J et al. Cell-penetrating peptides in diagnosis and treatment of human diseases: from preclinical research to clinical application. *Front Pharmacol*. 2020;11, 541971. https://doi.org/10.3389/fphar.2020.00697.
12. Peraro L, Kritzer JA. Emerging methods and design principles for cell-penetrant peptides. *Angew Chem Int Ed Engl*. 2018;57:11868–11881. https://doi.org/10.1002/anie.201801361.
13. Ochocki JD, Mullen DG, Wattenberg EV, Distefano MD. Evaluation of a cell penetrating prenylated peptide lacking an intrinsic fluorophore via in situ click reaction. *Bioorg Med Chem Lett*. 2011;21:4998–5001. https://doi.org/10.1016/j.bmcl.2011.04.138.
14. Chen Y et al. Stabilization of peptides against proteolysis through disulfide-bridged conjugation with synthetic aromatics. *Org Biomol Chem*. 2017;15:1921–1929. https://doi.org/10.1039/c6ob02786e.
15. Futaki S et al. Arginine-rich peptides. An abundant source of membrane-permeable peptides having potential as carriers for intracellular protein delivery. *J Biol Chem*. 2001;276:5836–5840. https://doi.org/10.1074/jbc.M007540200.
16. Mai JC, Shen H, Watkins SC, Cheng T, Robbins PD. Efficiency of protein transduction is cell type-dependent and is enhanced by dextran sulfate. *J Biol Chem*. 2002;277:30208–30218. https://doi.org/10.1074/jbc.M204202200.

17. Manceur A, Wu A, Audet J. Flow cytometric screening of cell-penetrating peptides for their uptake into embryonic and adult stem cells. *Anal Biochem*. 2007;364:51–59. https://doi.org/10.1016/j.ab.2007.02.015.
18. Rezgui R, Blumer K, Yeoh-Tan G, Trexler AJ, Magzoub M. Precise quantification of cellular uptake of cell-penetrating peptides using fluorescence-activated cell sorting and fluorescence correlation spectroscopy. *Biochim Biophys Acta*. 2016;1858:1499–1506. https://doi.org/10.1016/j.bbamem.2016.03.023.
19. Aussedat B, Sagan S, Chassaing G, Bolbach G, Burlina F. Quantification of the efficiency of cargo delivery by peptidic and pseudo-peptidic Trojan carriers using MALDI-TOF mass spectrometry. *Biochim Biophys Acta*. 2006;1758:375–383. https://doi.org/10.1016/j.bbamem.2006.01.012.
20. Palm C, Netzereab S, Hällbrink M. Quantitatively determined uptake of cell-penetrating peptides in non-mammalian cells with an evaluation of degradation and antimicrobial effects. *Peptides*. 2006;27:1710–1716. https://doi.org/10.1016/j.peptides.2006.01.006.
21. Burlina F, Sagan S, Bolbach G, Chassaing G. A direct approach to quantification of the cellular uptake of cell-penetrating peptides using MALDI-TOF mass spectrometry. *Nat Protoc*. 2006;1:200–205. https://doi.org/10.1038/nprot.2006.30.
22. Reveret L, Leclerc M, Morin F, Émond V, Calon F. Pharmacokinetics, biodistribution and toxicology of novel cell-penetrating peptides. *Sci Rep*. 2023;13:11081. https://doi.org/10.1038/s41598-023-37280-0.
23. Järver P, Mäger I, Langel Ü. In vivo biodistribution and efficacy of peptide mediated delivery. *Trends Pharmacol Sci*. 2010;31:528–535. https://doi.org/10.1016/j.tips.2010.07.006.
24. Yu P, Liu B, Kodadek T. A high-throughput assay for assessing the cell permeability of combinatorial libraries. *Nat Biotechnol*. 2005;23:746–751. https://doi.org/10.1038/nbt1099.
25. Ballister ER, Aonbangkhen C, Mayo AM, Lampson DM, Chenoweth DM. Localized light-induced protein dimerization in living cells using a photocaged dimerizer. *Nat Commun*. 2014;5:5475. https://doi.org/10.1038/ncomms6475.
26. Guidotti G, Brambilla L, Rossi D. Cell-penetrating peptides: from basic research to clinics. *Trends Pharmacol Sci*. 2017;38:406–424. https://doi.org/10.1016/j.tips.2017.01.003.
27. Patel LN, Zaro JL, Shen WC. Cell penetrating peptides: intracellular pathways and pharmaceutical perspectives. *Pharm Res*. 2007;24:1977–1992. https://doi.org/10.1007/s11095-007-9303-7.
28. Guerrero-Vázquez K, Del Rio G, Brizuela CA. Cell-penetrating peptides predictors: a comparative analysis of methods and datasets. *Mol Inform*. 2023;42, e202300104. https://doi.org/10.1002/minf.202300104.
29. Gautam A et al. CPPsite: a curated database of cell penetrating peptides. *Database (Oxford)*. 2012;2012, bas015. https://doi.org/10.1093/database/bas015.
30. Agrawal P et al. CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res*. 2016;44:D1098–D1103. https://doi.org/10.1093/nar/gkv1266.
31. Preto AJ et al. POSEIDON: Peptidic Objects SEquence-based Interaction with cellular DOmaiNs: a new database and predictor. *J Cheminform*. 2024;16:18. https://doi.org/10.1186/s13321-024-00810-7.
32. Li J, Yanagisawa K, Sugita M, Fujie T, Ohue M, Akiyama Y. CycPeptMPDB: a comprehensive database of membrane permeability of cyclic peptides. *J Chem Inf Model*. 2023;63:2240–2250. https://doi.org/10.1021/acs.jcim.2c01573.
33. Gautam A et al. *In silico* approaches for designing highly effective cell penetrating peptides. *J Transl Med*. 2013;11:74. https://doi.org/10.1186/1479-5876-11-74.
34. Holton TA, Pollastri G, Shields DC, Mooney C. CPPpred: prediction of cell penetrating peptides. *Bioinformatics*. 2013;29:3094–3096. https://doi.org/10.1093/bioinformatics/btt518.
35. Tang H, Su ZD, Wei HH, Chen W, Lin H. Prediction of cell-penetrating peptides with feature selection techniques. *Biochem Biophys Res Commun*. 2016;477:150–154. https://doi.org/10.1016/j.bbrc.2016.06.035.
36. Wei L, Tang J, Zou Q. SkipCPP-Pred: an improved and promising sequence-based predictor for predicting cell-penetrating peptides. *BMC Genomics*. 2017;18:742. https://doi.org/10.1186/s12864-017-4128-1.
37. Wei L, Xing P, Su R, Shi G, Ma ZS, Zou Q. CPPred-RF: a sequence-based predictor for identifying cell-penetrating peptides and their uptake efficiency. *J Proteome Res*. 2017;16:2044–2053. https://doi.org/10.1021/acs.jproteome.7b00019.
38. Kumar V et al. Prediction of cell-penetrating potential of modified peptides containing natural and chemically modified residues. *Front Microbiol*. 2018;9:725. https://doi.org/10.3389/fmicb.2018.00725.
39. Manavalan B, Subramaniyam S, Shin TH, Kim MO, Lee G. Machine-learning-based prediction of cell-penetrating peptides and their uptake efficiency with improved accuracy. *J Proteome Res*. 2018;17:2715–2726. https://doi.org/10.1021/acs.jproteome.8b00148.
40. Pandey P, Patel V, George NV, Mallajosyula SS. KELM-CPPpred: kernel extreme learning machine based prediction model for cell-penetrating peptides. *J Proteome Res*. 2018;17:3214–3222. https://doi.org/10.1021/acs.jproteome.8b00322.
41. Wang S, Cao Z, Li M, Yue Y. G-DipC: an improved feature representation method for short sequences to predict the type of cargo in cell-penetrating peptides. *IEEE/ACM Trans Comput Biol Bioinform*. 2020;17:739–747. https://doi.org/10.1109/TCBB.2019.2930993.
42. Qiang X, Zhou C, Ye X, Du PF, Su R, Wei L. CPPred-FL: a sequence-based predictor for large-scale identification of cell-penetrating peptides by feature representation learning. *Brief Bioinform*. 2020;21:11–23. https://doi.org/10.1093/bib/bby091.
43. Fu X, Cai L, Zeng X, Zou Q. StackCPPred: a stacking and pairwise energy content-based prediction of cell-penetrating peptides and their uptake efficiency. *Bioinformatics*. 2020;36:3028–3034. https://doi.org/10.1093/bioinformatics/btaa131.
44. Arif M, Ahmad S, Ali F, Fang G, Li M, Yu DJ. TargetCPP: accurate prediction of cell-penetrating peptides from optimized multi-scale features using gradient boost decision tree. *J Comput Aided Mol Des*. 2020;34:841–856. https://doi.org/10.1007/s10822-020-00307-z.
45. de Oliveira ECL, Santana K, Josino L, Lima e Lima C, de Souza de Sales Júnior C. Predicting cell-penetrating peptides using machine learning algorithms and navigating in their chemical space. *Sci Rep*. 2021;11:7628. https://doi.org/10.1038/s41598-021-87134-w.
46. Arif M et al. DeepCPPred: a deep learning framework for the discrimination of cell-penetrating peptides and their uptake efficiencies. *IEEE/ACM Trans Comput Biol Bioinform*. 2022;19:2749–2759. https://doi.org/10.1109/TCBB.2021.3102133.
47. Manavalan B, Patra MC. MLCPP 2.0: an updated cell-penetrating peptides and their uptake efficiency predictor. *J Mol Biol*. 2022;434, 167604. https://doi.org/10.1016/j.jmb.2022.167604.
48. Zhang X et al. SiameseCPP: a sequence-based Siamese network to predict cell-penetrating peptides by contrastive learning. *Brief Bioinform*. 2023;24, bbac545. https://doi.org/10.1093/bib/bbac545.
49. Park H, Park JH, Kim MS, Cho K, Shin JM. In silico screening and optimization of cell-penetrating peptides using deep learning methods. *Biomolecules*. 2023;13:522. https://doi.org/10.3390/biom13030522.
50. Shi K et al. PractiCPP: a deep learning approach tailored for extremely imbalanced datasets in cell-penetrating peptide prediction. *Bioinformatics*. 2024;40, btae058. https://doi.org/10.1093/bioinformatics/btae058.
51. Kumar N, Du Z, Li Y. pLM4CPPs: protein language model-based predictor for cell penetrating peptides. *J Chem Inf Model*. 2025;65:1128–1139. https://doi.org/10.1021/acs.jcim.4c01338.
52. Chen Q, Zhang Y, Gao J, Zhang J. CPPCGM: a highly efficient sequence-based tool for simultaneously identifying and generating cell-penetrating peptides. *J Chem Inf Model*. 2025;65:3357–3369. https://doi.org/10.1021/acs.jcim.5c00199.
53. Zhang H et al. Recent advances of cell-penetrating peptides and their application as vectors for delivery of peptide and protein-based cargo molecules. *Pharmaceutics*. 2023;15:2093. https://doi.org/10.3390/pharmaceutics15082093.
54. Dougherty PG, Sahni A, Pei D. Understanding cell penetration of cyclic peptides. *Chem Rev*. 2019;119:10241–10287. https://doi.org/10.1021/acs.chemrev.9b00008.
55. Singh S et al. PEPstrMOD: structure prediction of peptides containing natural, non-natural and modified residues. *Biol Direct*. 2015;10:73. https://doi.org/10.1186/s13062-015-0103-4.
56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
57. Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol*. 1981;147:195–197. https://doi.org/10.1016/0022-2836(81)90087-5.
58. Khairkhah N, Namvar A, Bolhassani A. Application of cell penetrating peptides as a promising drug carrier to combat viral infections. *Mol Biotechnol*. 2023;65:1387–1402. https://doi.org/10.1007/s12033-023-00679-1.
59. Desale K, Kuche K, Jain S. Cell-penetrating peptides (CPPs): an overview of applications for improving the potential of nanotherapeutics. *Biomater Sci*.

2021;9:1153–1188. https://doi.org/10.1039/d0bm01755h.
60. Tripathi PP, Arami H, Banga I, Gupta J, Gandhi S. Cell penetrating peptides in preclinical and clinical cancer diagnosis and therapy. *Oncotarget*. 2018;9:37252–37267. https://doi.org/10.18632/oncotarget.26442.
61. Jankovic J et al. Injectable daxibotulinumtoxinA in cervical dystonia: a phase 2 dose-escalation multicenter study. *Mov Disord Clin Pract*. 2018;5:273–282. https://doi.org/10.1002/mdc3.12613.
62. Cummings JC, Zhang H, Jakymiw A. Peptide carriers to the rescue: overcoming the barriers to siRNA delivery for cancer treatment. *Transl Res*. 2019;214:92–104. https://doi.org/10.1016/j.trsl.2019.07.010.
63. Koo JH, Kim GR, Nam KH, Choi JM. Unleashing cell-penetrating peptide applications for immunotherapy. *Trends Mol Med*. 2022;28:482–496. https://doi.org/10.1016/j.molmed.2022.03.010.
64. Yang D, Liu B, Sha H. Advances and prospects of cell-penetrating peptides in tumor immunotherapy. *Sci Rep*. 2025;15:3392. https://doi.org/10.1038/s41598-025-86130-8.
65. Nhàn NTT, Maidana DE, Yamada KH. Ocular delivery of therapeutic agents by cell-penetrating peptides. *Cells*. 2023;12:1071. https://doi.org/10.3390/cells12071071.

Nisha Bajiya [1,3], Sohana Najrin [2,3], Pankaj Kumar [1,3], Shubham Choudhury [1], Ritu Tomer [1], Gajendra P.S. Raghava [1,*]

[1] Department of Computational Biology, Indraprastha Institute of Information Technology, Okhla Phase 3, New Delhi 110020, India
[2] Jamia Hamdard University, Block D, Hamdard Nagar, New Delhi 110062, India

* *Corresponding author.*
raghava@iiitd.ac.in (G.P.S. Raghava),

[1] Authors contributed equally.