



OPEN

# DCTPep, the data of cancer therapy peptides

DATA DESCRIPTOR

Xin Sun, Yanchao Liu, Tianyue Ma, Ning Zhu, Xingzhen Lao &amp; Heng Zheng

With the discovery of the therapeutic activity of peptides, they have emerged as a promising class of anti-cancer agents due to their specific targeting, low toxicity, and potential for high selectivity. In particular, as peptide-drug conjugates enter clinical, the coupling of targeted peptides with traditional chemotherapy drugs or cytotoxic agents will become a new direction in cancer treatment. To facilitate the drug development of cancer therapy peptides, we have constructed DCTPep, a novel, open, and comprehensive database for cancer therapy peptides. In addition to traditional anticancer peptides (ACPs), the peptide library also includes peptides related to cancer therapy. These data were collected manually from published research articles, patents, and other protein or peptide databases. Data on drug library include clinically investigated and/or approved peptide drugs related to cancer therapy, which mainly come from the portal websites of drug regulatory authorities and organisations in different countries and regions. DCTPep has a total of 6214 entries, we believe that DCTPep will contribute to the design and screening of future cancer therapy peptides.

## Background & Summary

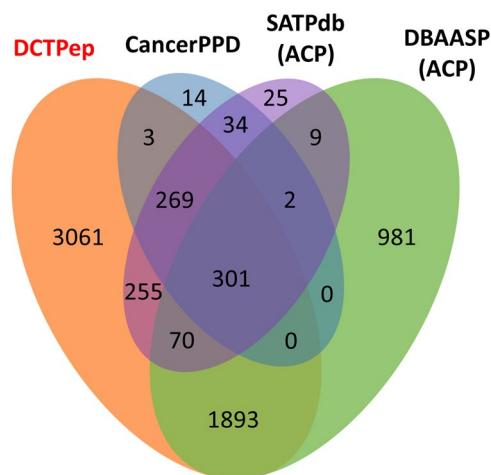
Cancer is a leading cause of death and a significant barrier to increasing life expectancy worldwide<sup>1</sup>. Cancer treatment has improved in the past few decades, but chemotherapy remains the mainstay of cancer treatment. Multidrug resistance is a major problem associated with anticancer chemotherapy<sup>2,3</sup>. Data show that 90% of cancer deaths can be attributed to multidrug resistance<sup>4</sup>. Due to structural differences with small-molecule compounds, bioactive peptides have received much attention and are believed to be alternative candidates for multidrug-resistant cancer therapy<sup>5-7</sup>.

Anticancer peptides (ACPs) are biologically active peptides with antitumor activities that exist widely in a variety of organisms, including mammals, amphibians, insects, plants, and microorganisms. ACPs have many advantages in the treatment of tumours, such as low molecular weight (compared to protein-based therapy), simple structure, high anticancer activity, high selectivity, fewer side effects, easy modifications, and less possibility to cause resistance<sup>8,9</sup>. Although ACPs have been extensively studied, its mechanism of action is not fully understood. At present, the known mechanisms of ACPs mainly include inhibition of tumour cell proliferation or migration<sup>10,11</sup>, inhibition of tumour blood vessel formation<sup>12</sup>, causing cancer cell lysis<sup>13</sup>, and induction of cancer cell apoptosis<sup>14</sup>. In addition, peptides can also serve as a targeted therapeutic agent that can target and directly bind specific cancer cells or cancer related biomarkers, and can also serve as a peptide carrier linked to traditional anticancer drugs<sup>15,16</sup>.

Although the enormous potential of peptides in cancer therapeutics, there is a relative scarcity of dedicated databases specifically storing cancer therapy peptides information. Most of the ACPs information is dispersed in bioactive peptide databases, such as DRAMP<sup>17</sup>, APD<sup>18</sup>, DBAASP<sup>19</sup>, HORDB<sup>20</sup>, CPPsite<sup>21</sup>, and SATPdb<sup>22</sup>, which mainly focus on antimicrobial peptides or hormones. The CancerPPD<sup>23</sup> database is a known database for annotating ACPs and anticancer proteins; however, its data have not been updated since 2015. Many antimicrobial peptide databases also store information about the anticancer activity of some antimicrobial peptides, but it does not contain detail annotation of ACPs. For example, they did not fully provide information on cancer cells or molecular targets of ACPs, nor do they include peptide drugs. Therefore, we constructed an open, comprehensive database of cancer therapy peptides, DCTPep, that not only includes traditional ACPs, but also peptides with targeted effects on cancer therapeutics. DCTPep can be freely accessed and downloaded from <http://dct-pep.cpu-bioinform.org/>.

Developing targeted therapies that selectively act on cancer cells has always been an ideal approach for cancer treatment. A promising targeted therapy is drug conjugates, which involve linking targeting carriers

School of Life Science and Technology, China Pharmaceutical University, 24 Tongjiaxiang, Nanjing, 210009, P. R. China. ✉e-mail: [lao@cpu.edu.cn](mailto:lao@cpu.edu.cn); [zhengh18@hotmail.com](mailto:zhengh18@hotmail.com)



**Fig. 1** Venn diagram illustrating the numbers of overlapping and non-overlapping peptide sequences related to cancer therapy from the DCTPep, CancerPPD, SATPdb and DBAASP.

Features	DCTPep	CancerPPD	DRAMP 3.0	APD3	DBAASP v3	SATPdb
Entries related to cancer therapy peptide	6106	624	455	276	3,409	1,099
ACP drugs	108	NA	NA	NA	NA	NA
Cancer cell line	>400	249	NA	NA	NA	NA
Targets	>60	NA	NA	NA	NA	NA
predicted structures of ACPs	1646 (Based on AlphaFold)	617 (Based on molecular dynamics)	NA	NA	1372 (Based on molecular dynamics)	NA
Download dataset	Yes	Yes	Yes	Yes	Yes	Yes

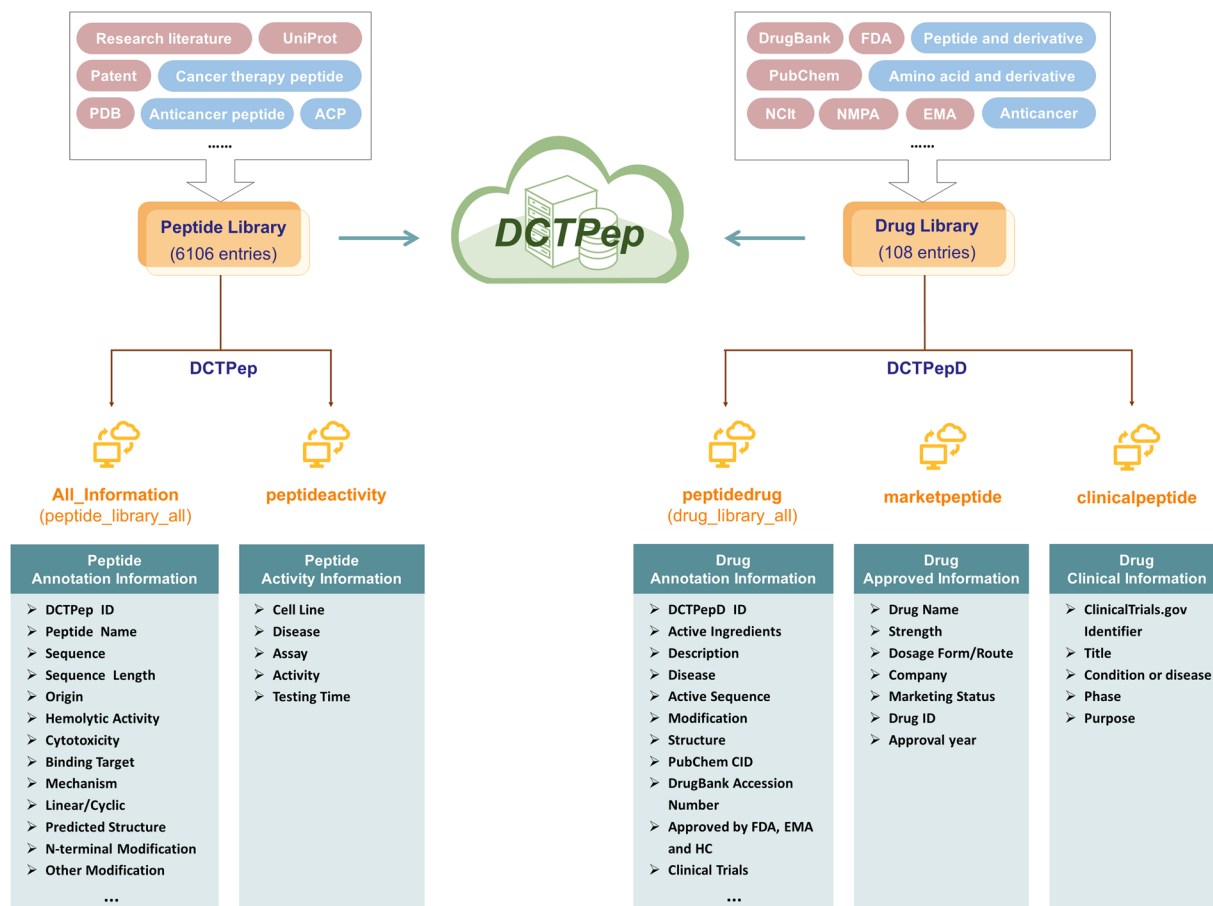
**Table 1.** Comparison of peptides related to cancer therapy in DCTPep with other peptide databases (data as of 2023.12.20).

with chemotherapy drugs or cytotoxic agents through a linker, such as antibody-drug conjugates (ADCs) and peptide-drug conjugates (PDCs)<sup>24</sup>. Currently, the most common drug conjugates used in cancer treatment in clinical practice are ADCs. However, with the increasing presence of peptides in clinical, PDCs has also emerged. PDCs have the potential to overcome the limitations of ADCs, such as smaller molecular weight and ease of synthesis<sup>25</sup>. Nowadays, only two PDCs, <sup>177</sup>Lu-dotatate (DCTPepD0013) and Melflufen (DCTPepD0108), have been approved for clinical cancer treatment, of which Melflufen being withdrawn from the market by the FDA. However, there are still many PDCs in cancer clinical development or about to enter clinical trials. The potential of PDCs cannot be ignored. Peptides play a crucial role as carriers in PDCs. Therefore, DCTPep not only focuses on collecting ACPs but also emphasizes the collection of cancer targeted peptides. The carrier peptides in PDCs include cell-penetrating peptides (CPP) and cell-targeting peptides (CTP)<sup>26</sup>. The classification field in the database also follows a similar category, including cell-penetrating peptides, cancer-targeting peptides, and targeted peptide conjugates.

Figure 1 and Table 1 presents the comparative results of DCTPep datasets with ACP datasets in other peptide databases. Compared to DBAASP, CancerPPD and SATPdb, DCTPep possesses over 3000 unique entries. DCTPep provides a vast amount of cancer therapy peptide data, including clinically relevant peptide drugs curated in the drug library, filling the gaps in existing data and offering assistance in the design and screening of novel cancer therapeutic peptides. Particularly, the targeted peptide data will offer more options for PDC design. In order to better understand the mechanism of action of cancer therapy peptides, we have added target annotations and collected over 60 targets for these peptides that are not included in other ACPs databases. The dataset is freely available to all via the web without the need to login or registration and is not password protected. We believe that DCTPep will become a valuable resource for the development of novel bioactive peptides, particularly in the field of cancer therapeutics.

## Methods

**Data collection and compilation.** In order to develop DCTPep, extensive searches were conducted on published articles, patents, and public databases. The data of DCTPep was stored in two sub libraries: peptide library and drug library. The inclusion criteria for the peptide library in the DCTPep were as follows: 1. The sequence of amino acids is reported; 2. Mature peptide sequences without precursor and signal regions; 3. The length of the sequence does not exceed 100 amino acids; 4. Peptides that exhibit anticancer/antitumor activity or target specific molecules/biomarkers overexpressed in cancer cells; 5. Cell-penetrating peptides that can



**Fig. 2** Architecture of the datasets in DCTPep.

enhance the delivery of drugs into cancer cells. The inclusion criteria for drug library were similar to those for peptide library: 1. Peptides and their derivatives or amino acid derivatives related to cancer treatment; 2. Entered clinical research or approved by FDA, EMA or HC.

To collect peptide data, keywords were used to search in academic search engines such as Google Scholar, Web of Science, PubMed, and Google Patents. The keywords included “ACP”, “antiangiogenic peptides”, “cancer therapy peptide”, “cancer targeted peptide”, and “peptide conjugates”. After collecting research papers, patents, and clinical research literature, data were manually extracted. In addition to manually extracting information of cancer therapy peptide from literature, also included other information related to peptides (such as three-dimensional structures) in UniProt<sup>27</sup>, PDB<sup>28</sup>, and other databases. The physicochemical information of peptides is calculated using ExPASy ProtParam server (<https://web.expasy.org/protparam/>, accessed on March 2024) and SciDBMaker<sup>29</sup>.

The data of drug library mainly originated from the portal websites of drug regulatory authorities and organisations in several countries and regions. In addition, it was supplemented by the drug databases DrugBank<sup>30</sup>, PubChem<sup>31</sup>, NCI Thesaurus<sup>32</sup> and Global Substance Registration System (GSRS)<sup>33</sup>. By entering keywords such as “peptides and their derivatives”, “amino acids and their derivatives”, and “anticancer” into the aforementioned website or database, relevant information can be found.

**Structural prediction and evaluation.** Due to the difficulties in experimental determination of peptide and protein structures, most of the peptides lack experimental determined structures. AlphaFold<sup>34</sup> was used to predict the potential 3D structures of DCTPep peptides. Default structure parameters for AlphaFold prediction were used: peptide was modeled as a monomer; Multiple sequence alignment (MSA) information databases: full\_dbs (all gene databases)<sup>34</sup>. Each peptide generates 5 structures, and the structure with the highest score is selected based on predicted local distance difference test (pLDDT)<sup>34</sup>. To evaluate the reliability of AlphaFold predicted peptide structures, 30 peptides with experimental determined structures were selected and their structures were predicted by AlphaFold. The differences between predicted structure and experimentally determined structure were calculated by Root-Mean-Square Deviation (RMSD)<sup>35</sup>. Given two conformations,  $\alpha$  and  $\beta$  of  $N$  residues, let  $r_\alpha$  and  $r_\beta$  be the respective coordinates of their residues at position  $i$ , for  $1, \dots, N$ . RMSD between  $\alpha$  and  $\beta$  as Eq. (1):

Fields	Description	
General information	DCTPep ID	The identifier of peptides in DCTPep.
	Peptide Name	Name of each peptide in DCTPep.
	Sequence	Single letter sequence of peptides.
	Sequence Length	Number of residues in the peptide sequence.
	Uniprot ID	Corresponding Uniprot link ID.
	Source	Organism source.
	Type	Native peptide or synthetic peptide.
	Classification	Various possible classifications. Including ACP, Cancer targeted peptides, Membrane-targeted mechanism...
Activity information	Anticancer activity	Anticancer activity verified by experiment (Data from literature).
	Hemolytic Activity	Hemolytic activity information against red blood cells (RBCs).
	Cytotoxicity	Cytotoxicity information against normal (non-cancerous) cell line.
	Target	The action site of peptides against cancer cell.
	Affinity	Binding affinity between peptides and targets.
	Mechanism	Mechanisms by peptides exert anticancer agents.
Structure information	PDB ID	Corresponding PDB link ID.
	Predicted Structure	Structure predicted by AlphaFold.
	Helicity	Percentage of $\alpha$ -helix.
	Linear/Cyclic	Linear or cyclic of peptides.
	Disulfide/Other Bond	Disulfide bond (DSB) or other bond, such as N-C termini peptide bond (NCB).
	N/C-terminal Modification	The modifications of N/C-terminal according to the references.
	Other Modification	Special amino acids (out of 20 common amino acids).
Chiral	The L/D amino acid composition of peptides.	
Physicochemical Information	Formula, mass, pI, Net charge and other information, calculated by ProtParam and SciDBMaker.	
Literature Information	The information of peptides come from all kinds of papers or patents, and the section provides the way to find the full text.	
Link	Corresponding link to other peptide databases.	

**Table 2.** Peptide library data annotation field list.

Fields	Description	
General information	DCTPepD ID	Identification code for DCTPepD drug library, the field provides the unique accessing number linking to the corresponding DCTPepD entry.
	Active Ingredients	Active pharmaceutical ingredient. Substance in which the drug actually works.
	Description	Drug description. Derived from descriptions in NCI or literature sources.
	Synonyms	Other names of drug.
	Disease	Applicable diseases.
	Classification	Drug Categories.
Structure information	Molecular Formula, Molecular Weight, Active Sequence, Sequence Length, Modification, and other structure information.	
External Codes	External identification code, also provides the accessing link to PubChem, DrugBank, NCI Thesaurus and GRSR.	
Drug approval	Drug indication	Stemmed from DrugBank or clinical trials.
	Approved information	Approved drug formulation information, sourced from Drugs@FDA, European Medicines Agency (EMA), and Health Canada.
	Clinical information	Information sourced from ClinicalTrials.gov.

**Table 3.** Drug library data annotation field list.

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_{\alpha,i} - Qr_{\beta,i})^2} \quad (1)$$

Where Q is the unitary rotation matrix that optimally aligns the vectors. Disulfide bonds are also considered to see if AlphaFold can correctly predict the disulfide bonds. Whatcheck<sup>36</sup> and Procheck<sup>37</sup> are used to assess

DCTPep_ID	PDB_ID	pLDDT	C $\alpha$ RMSD(Å)	Disulfide bond position*
DCTPep00004	1D9L	81.43	2.314	/
DCTPep00007	1VM5	92.59	0.629	/
DCTPep00015	7OVZ	85.33	2.510	/
DCTPep00073	2KCG	89.84	0.758	TRUE
DCTPep00155	2FBS	87.92	0.419	/
DCTPep00177	2MAG	84.56	0.632	/
DCTPep00194	1LFC	76.00	2.060	TRUE
DCTPep00267	1PG1	84.8	2.145	TRUE
DCTPep00287	2LAM	84.35	2.345	TRUE
DCTPep00334	1XC0	75.86	3.583	/
DCTPep00433	2G9P	83.30	3.740	/
DCTPep00468	1BH1	79.11	3.435	/
DCTPep00494	2K6O	83.08	2.148	/
DCTPep00514	1D7N	82.51	1.927	/
DCTPep00623	1HA9	84.38	1.029	TRUE
DCTPep00759	1KFP	88.10	1.632	TRUE
DCTPep00823	2N9A	92.31	0.972	/
DCTPep00852	6PIP	91.06	0.618	TRUE
DCTPep00857	6PI2	91.14	0.688	TRUE
DCTPep00916	6PIN	90.74	0.880	TRUE
DCTPep01108	1VM4	91.98	0.598	/
DCTPep01112	1T52	82.43	0.916	/
DCTPep01139	2IGR	71.96	3.413	/
DCTPep01237	1KZ0	79.60	2.538	/
DCTPep01305	1SMZ	82.25	1.276	/
DCTPep01728	2JMY	83.34	1.985	/
DCTPep02004	1WQK	86.35	1.281	TURE
DCTPep02324	6VPN	97.15	0.878	TURE
DCTPep03586	2MFS	88.20	0.689	TURE
DCTPep04747	3C8P	95.60	0.579	TURE
Average	—	—	1.621	—

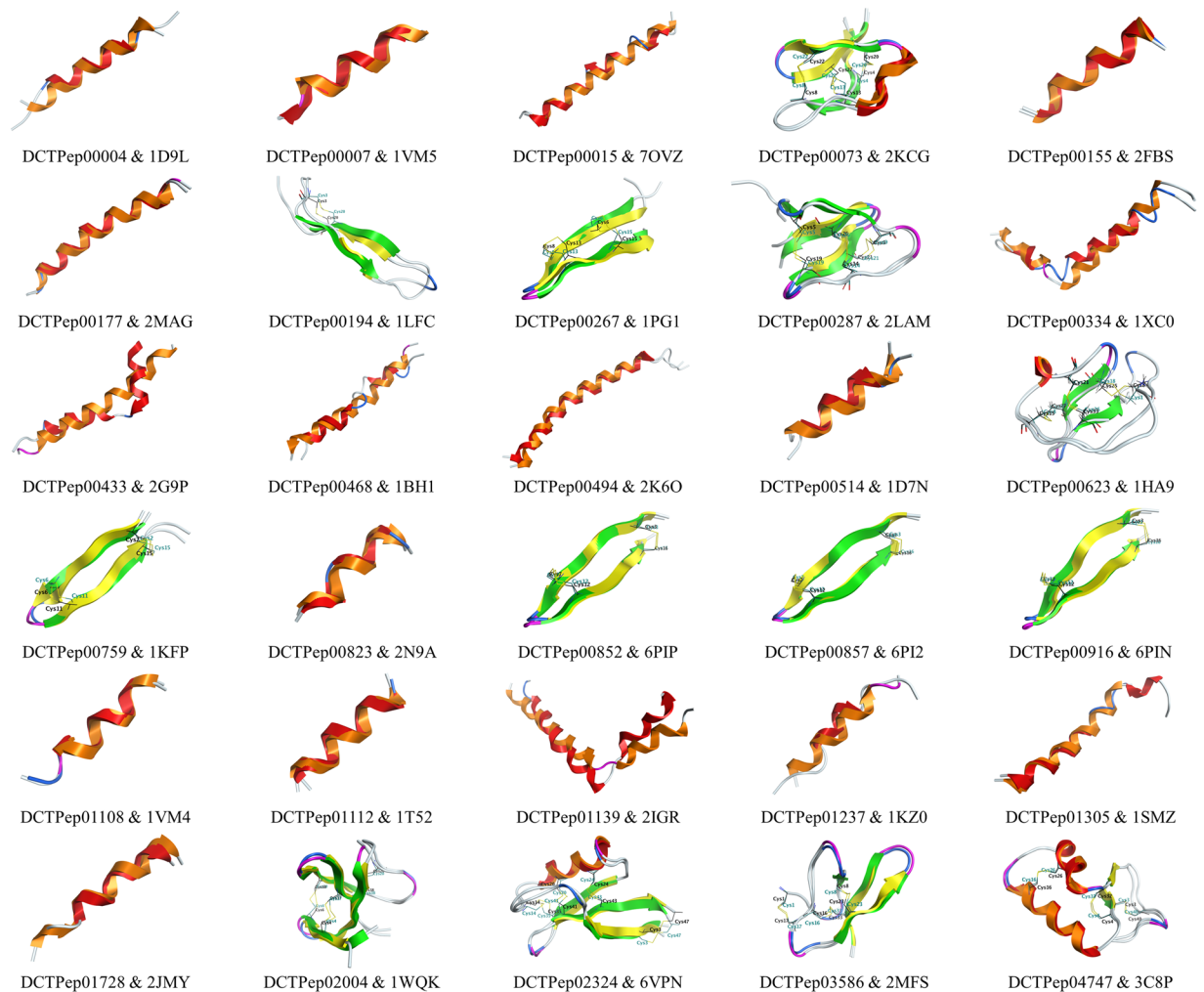
**Table 4.** Comparison between predicted structures and experimental structures. \*“TURE” indicates consistent disulfide bond formation positions; “FALSE” indicates inconsistent disulfide bond formation positions; “/” indicates no disulfide bond formation.

the quality of the predicted structures. Whatcheck<sup>36</sup> evaluates multiple parameters such as bond lengths, bond angles, and torsion angles of the input structure. Procheck<sup>37</sup> assesses the stereochemical quality of the input structure and provides various graphical outputs. Ramachandran plot<sup>38</sup> is used to evaluate the rationality of the structure, where peptide bond dihedral angles  $\Psi$ (psi) and  $\Phi$ (phi) combinations are expected to be located in most favored regions and allowed regions (core regions) in the plot. Ideally, a protein structure should have over 90% dihedral angles  $\Phi$ - $\Psi$  of residues in these core regions<sup>37</sup>.

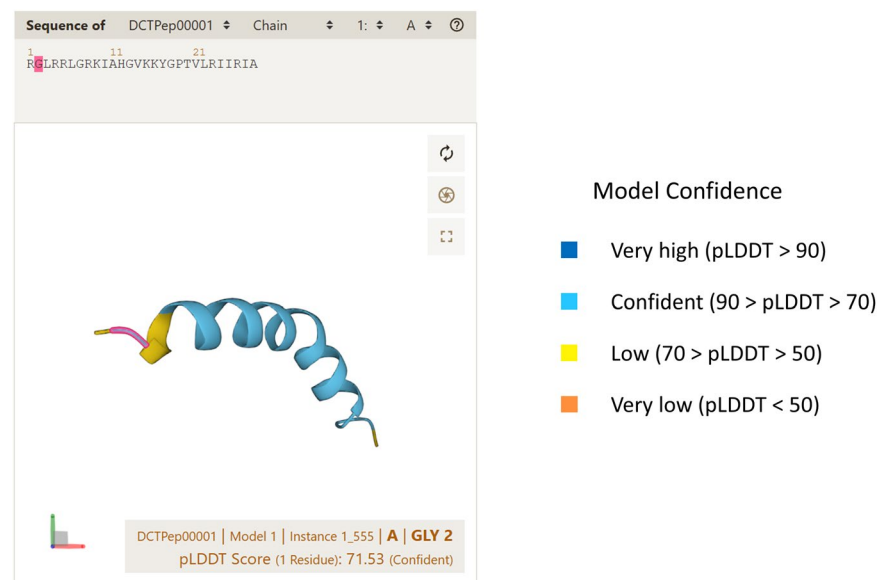
### Data Records

The datasets of DCTPep are available at Figshare<sup>39</sup> and contains the following files: All\_information (annotation information table for storing peptide library entries), peptideactivity (activity information annotation of peptide library entries), peptidedrug (annotation information table for storing active Ingredients of drug library entries), marketpeptide (approved drug preparations information annotation of drug library entries), clinicalpeptide (clinical peptide information annotation of drug library entries), peptide\_library\_all (peptide library data stored in Fasta format) and prediction\_pdb (compressed packets for storing predicted structures). The architecture of the DCTPep is shown in Fig. 2. DCTPep contains a total of 6214 peptide entries, of which 6106 are stored in the peptide library and 108 are stored in the drug library (DCTPepD), involving over 60 targets and over 380 cancer cell lines.

Table 2 displays detailed annotation information of the data in the peptide library. Each entry in the peptide library consists of the following sections: general information, activity information, structural information, physicochemical information, literature information, and links. The peptides in the peptide library included cancer therapeutic peptides such as traditional ACP and cancer targeted peptides. Low cytotoxicity and hemolytic activity are also important criteria for developing peptide-based drugs. Therefore, in addition to anticancer activity and targets, activity information also includes cytotoxicity and hemolytic activity. All annotation information is manually extracted from the literature, and corresponding paper or patent source information is provided. The physicochemical information is calculated by ProtParam and SciDBMaker<sup>29</sup>. For the same



**Fig. 3** Alignment and superimposition plot of predicted structures and experimental structures. Predicted structures: helix-orange, strand-green, turn-magenta, Cys-dark cyan; Experimental structures: helix-red, strand-yellow, turn-blue, Cys-black.



**Fig. 4** Example of the predicted structure of DCTPep00001 showing by Mol\* Viewer.

DCTPep_ID	Whatcheck			Procheck		
	Total metrics	Error	Error rate	Core regions	Additional/generous allowed regions	Disallowed regions
DCTPep00004	40	4	10.00%	100.00%	0.00%	0.00%
DCTPep00007	40	4	10.00%	100.00%	0.00%	0.00%
DCTPep00015	42	4	9.52%	100.00%	0.00%	0.00%
DCTPep00073	42	5	11.90%	87.50%	12.50%	0.00%
DCTPep00155	42	4	9.52%	100.00%	0.00%	0.00%
DCTPep00177	42	4	9.52%	100.00%	0.00%	0.00%
DCTPep00194	43	6	13.95%	100.00%	0.00%	0.00%
DCTPep00267	40	5	12.50%	92.30%	0.00%	7.70%
DCTPep00287	44	5	11.36%	81.80%	18.20%	0.00%
DCTPep00334	43	4	9.30%	96.20%	3.80%	0.00%
DCTPep00433	43	4	9.30%	100.00%	0.00%	0.00%
DCTPep00468	42	4	9.52%	100.00%	0.00%	0.00%
DCTPep00494	45	5	11.11%	90.60%	9.40%	0.00%
DCTPep00514	38	4	10.53%	100.00%	0.00%	0.00%
DCTPep00623	42	6	14.29%	69.20%	30.80%	0.00%
DCTPep00759	41	5	12.20%	86.70%	13.30%	0.00%
DCTPep00823	39	4	10.26%	100.00%	0.00%	0.00%
DCTPep00852	40	6	15.00%	92.90%	7.10%	0.00%
DCTPep00857	40	5	12.50%	92.90%	7.10%	0.00%
DCTPep00916	40	4	10.00%	92.90%	7.10%	0.00%
DCTPep01108	40	4	10.00%	100.00%	0.00%	0.00%
DCTPep01112	38	4	10.53%	100.00%	0.00%	0.00%
DCTPep01139	40	6	15.00%	100.00%	0.00%	0.00%
DCTPep01237	39	4	10.26%	100.00%	0.00%	0.00%
DCTPep01305	39	4	10.26%	100.00%	0.00%	0.00%
DCTPep01728	39	5	12.82%	100.00%	0.00%	0.00%
DCTPep02004	45	6	13.33%	90.30%	9.70%	0.00%
DCTPep02324	45	6	13.33%	95.10%	4.90%	0.00%
DCTPep03586	43	7	16.28%	87.50%	8.30%	0.00%
DCTPep04747	43	5	11.63%	97.40%	2.60%	0.00%
Average	—	—	11.52%	95.11%	4.49%	0.26%

**Table 5.** The results of predicted structures in Whatcheck and Procheck.

peptide, the emphasis of the information recorded in different databases may vary. Therefore, DCTPep provides corresponding peptide entry codes in other peptide databases.

The data in the drug library includes peptide drugs that have been approved or are in clinical research stage. Table 3 shows detailed annotation information for drug library data. Each entry consists of four sections: general information, structural information, external codes, and drug approval. The external codes provide identification codes for drug entries in other public databases, allowing users to obtain more comprehensive information on related entries from other sources. Approved drug formulations and clinical information can be found in the drug approval section. A total of 28 approved anticancer peptide drugs and 80 peptides in various clinical trial stages are included in the drug library.

### Technical Validation

AlphaFold demonstrated unprecedented accuracy in 14th Critical Assessment of protein Structure Prediction (CASP14)<sup>34</sup>. The study conducted by McDonald *et al.*<sup>40</sup> also indicated that AlphaFold can accurately predict peptides with  $\alpha$ -helices,  $\beta$ -sheets, and rich in disulfide bonds. To evaluate the accuracy of AlphaFold, 30 ACPs with experimentally determined structures were predicted by AlphaFold.

Table 4 and Fig. 3 displays the comparison results between predicted structures and experimental structures, including RMSD and disulfide bond positions. The results indicate that the predicted structures have high accuracy. The deviations between the predicted and experimental structures are small, with an average of C $\alpha$  ( $\alpha$ -carbon atom) RMSD value is 1.621 Å. For structures containing disulfide bonds, AlphaFold can accurately predict the positions of the disulfide bonds. Some of the predicted structures of peptides can be directly obtained from the AlphaFold Protein Structure Database<sup>41</sup>, for example, AF-P82393-F1 (DCTPep00006) and AF-P80400-F1 (DCTPep00097).

pLDDT is an important parameter for assessing the confidence of predictions<sup>34</sup>. While using pLDDT alone to define the accuracy of predicted peptide structures may not be entirely accurate, it can still reflect their accuracy

to some extent. DCTPep integrates the Mol\* Viewer<sup>42</sup> to display the predicted structures, where the pLDDT of each residue can be visualized in the structure<sup>43</sup> (Fig. 4).

The quality assessment of the predicted structures was performed using Whatcheck<sup>36</sup> and Procheck<sup>37</sup> (Table 5), and the results indicate that the predicted structures are reliable. The average error rate of Whatcheck is 11.52%, which is at a relatively low level. In the Ramachandran plot generated by Procheck, the average core regions occupancy rate is 95.11%, only the DCTPep00623 has a low occupancy rate of core regions. The average disallowed regions occupancy rate is 0.26%, only DCTPep00267 has one residue present in the disallowed regions. These errors are within an acceptable range.

### Code availability

DCTPep can be freely accessed at <http://dctpep.cpu-bioinform.org/>. The data are stored in the Figshare repository available at <https://doi.org/10.6084/m9.figshare.25796353.v1>.

Received: 19 July 2023; Accepted: 20 May 2024;

Published online: 25 May 2024

### References

- Bray, F., Laversanne, M., Weiderpass, E. & Soerjomataram, I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. *Cancer*. **127**, 3029–3030 (2021).
- Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S. & Baradaran, B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv. Pharm. Bull.* **7**, 339–348 (2017).
- Alfarouk, K. O. *et al.* Resistance to cancer chemotherapy: failure in drug response from ADME to P-gp. *Cancer Cell Int.* **15**, 71 (2015).
- Bukowski, K., Kciuk, M. & Kontek, R. Mechanisms of Multidrug Resistance in Cancer Chemotherapy. *Int. J. Mol. Sci.* **21**, 3233 (2020).
- Arumugam, V., Venkatesan, M., Ramachandran, S. & Sundaresan, U. Bioactive Peptides from Marine Ascidians and Future Drug Development—A Review. *Int. J. Pept. Res. Ther.* **24**, 13–18 (2018).
- O'Brien-Simpson, N. M., Hoffmann, R., Chia, C. S. B. & Wade, J. D. Editorial: Antimicrobial and Anticancer Peptides. *Front. Chem.* **6**, 13 (2018).
- Kardani, K. & Bolhassani, A. Antimicrobial/anticancer peptides: bioactive molecules and therapeutic agents. *Immunotherapy*. **13**, 669–684 (2021).
- Xie, M., Liu, D. & Yang, Y. Anti-cancer peptides: classification, mechanism of action, reconstruction and modification. *Open Biol.* **10**, 200004 (2020).
- Chiangjong, W., Chutipongtanate, S. & Hongeng, S. Anticancer peptide: Physicochemical property, functional aspect and trend in clinical application (Review). *Int J Oncol.* **57**, 678–696 (2020).
- Kannan, A., Hettiarachchy, N. S., Marshall, M., Raghavan, S. & Kristinsson, H. Shrimp shell peptide hydrolysates inhibit human cancer cell proliferation: Shrimp shell peptide hydrolysates inhibit cancer cell proliferation. *J. Sci. Food Agric.* **91**, 1920–1924 (2011).
- Khamessi, O., Ben Mabrouk, H., ElFessi-Magouri, R. & Kharrat, R. RK1, the first very short peptide from *Buthus occitanus tunetanus* inhibits tumor cell migration, proliferation and angiogenesis. *Biochem. Biophys. Res. Commun.* **499**, 1–7 (2018).
- Gong, F. *et al.* A Novel Peptide from Abalone (*Haliotis discus hannai*) to Suppress Metastasis and Vasculogenic Mimicry of Tumor Cells and Enhance Anti-Tumor Effect *In Vitro*. *Mar. Drugs*. **17**, 244 (2019).
- Gajski, G. & Garaj-Vrhovac, V. Melittin: A lytic peptide with anticancer properties. *Environ. Toxicol. Pharmacol.* **36**, 697–705 (2013).
- Ren, S. X. *et al.* FK-16 Derived from the Anticancer Peptide LL-37 Induces Caspase-Independent Apoptosis and Autophagic Cell Death in Colon Cancer Cells. *PLoS One*. **8**, e63641 (2013).
- Peyressatre, M., Prevel, C., Pellerano, M. & Morris, M. C. Targeting cyclin-dependent kinases in human cancers: from small molecules to Peptide inhibitors. *Cancers (Basel)*. **7**, 179–237 (2015).
- Zhang, Y., Wang, C., Zhang, W. & Li, X. Bioactive peptides for anticancer therapies. *Biomater Transl.* **4**, 5–17 (2023).
- Shi, G. *et al.* DRAMP 3.0: an enhanced comprehensive data repository of antimicrobial peptides. *Nucleic Acids Res.* **50**, D488–D496 (2022).
- Wang, G., Li, X. & Wang, Z. APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **44**, D1087–1093 (2016).
- Pirtskhalava, M. *et al.* DBAASP v3: database of antimicrobial/cytotoxic activity and structure of peptides as a resource for development of new therapeutics. *Nucleic Acids Res.* **49**, D288–D297 (2021).
- Zhu, N., Dong, F., Shi, G., Lao, X. & Zheng, H. HORDB a comprehensive database of peptide hormones. *Sci. Data*. **9**, 187 (2022).
- Agrawal, P. *et al.* CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res.* **44**, D1098–1103 (2016).
- Singh, S. *et al.* SATPdb: a database of structurally annotated therapeutic peptides. *Nucleic Acids Res.* **44**, D1119–1126 (2016).
- Tyagi, A. *et al.* CancerPPD: a database of anticancer peptides and proteins. *Nucleic Acids Research*. **43**, D837–D843 (2015).
- Alas, M., Saghaeidehkordi, A. & Kaur, K. Peptide-Drug Conjugates with Different Linkers for Cancer Therapy. *J Med Chem.* **64**, 216–232 (2021).
- Lindberg, J., Nilvebrant, J., Nygren, P. A. & Lehmann, F. Progress and Future Directions with Peptide-Drug Conjugates for Targeted Cancer Therapy. *Molecules*. **26** (2021).
- Hoppenz, P., Els-Heindl, S. & Beck-Sickinger, A. G. Peptide-Drug Conjugates and Their Targets in Advanced Cancer Therapies. *Front Chem.* **8**, 571 (2020).
- Consortium, U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
- Bittrich, S. *et al.* RCSB Protein Data Bank: improved annotation, search and visualization of membrane protein structures archived in the PDB. *Bioinformatics*. **38**, 1452–1454 (2022).
- Hammami, R., Zouhir, A., Naghmouchi, K., Ben Hamida, J. & Fliss, I. SciDBMaker: new software for computer-aided design of specialized biological databases. *BMC Bioinf.* **9**, 121 (2008).
- Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
- Kim, S. *et al.* PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* **49**, D1388–D1395 (2021).
- Sioutos, N. *et al.* NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J Biomed Inform.* **40**, 30–43 (2007).
- Peryea, T. *et al.* Global Substance Registration System: consistent scientific descriptions for substances related to health. *Nucleic Acids Res.* **49**, D1179–D1185 (2021).
- Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*. **596**, 583–589 (2021).
- Betancourt, M. R. & Skolnick, J. Universal similarity measure for comparing protein structures. *Biopolymers*. **59**, 305–309 (2001).
- Hoof, R. W., Vriend, G., Sander, C. & Abola, E. E. Errors in protein structures. *Nature*. **381**, 272 (1996).



37. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
38. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology.* **7**, 95–99 (1963).
39. Sun, X. DCTPep, the data of cancer therapy peptides. *Figshare* <https://doi.org/10.6084/m9.figshare.25796353.v1> (2024).
40. McDonald, E. F., Jones, T., Plate, L., Meiler, J. & Gulsevin, A. Benchmarking AlphaFold2 on peptide structure prediction. *Structure.* **31**, 111–119.e112 (2023).
41. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
42. Sehnal, D. *et al.* Mol\* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res.* **49**, W431–W437 (2021).
43. Ruff, K. M. & Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *J Mol Biol.* **433**, 167208 (2021).

## Acknowledgements

This work was funded by the National Natural Science Foundation of China (grant number 82273834) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD2014–65). The authors would like to thank for the computational support from the High-Performance Computing Centre at China Pharmaceutical University.

## Author contributions

H.Z. conceived the project; X.S. and N.Z. designed and developed the database; X.S., Y.L. and T.M. extracted and organized data; X.S. and Y.L. tested the database; X.S. and T.M. proofread the database; X.S. written the manuscript; H.Z. reviewed the manuscript; X.L. provided constructive suggestions for the project; H.Z. guide and supervise the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to X.L. or H.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024