Database tool

DPL: a comprehensive database on sequences, structures, sources and functions of peptide ligands

Fangyu Wang[1,2,*,#], Ning Li[2,#], Chunfeng Wang[3], Guangxu Xing[1], Shuai Cao[2], Qian Xu[1], Yunshang Zhang[1], Man Hu[1] and Gaiping Zhang[1,2,*]

[1]Henan Key Laboratory of Animal Immunology, Henan Academy of Agricultural Sciences, 116# Huayuan Road, Zhengzhou, Henan Province, 450002, China, [2]College of Food Science and Technology, Henan Agricultural University, 63#Agricultural Road, Zhengzhou, Henan Province, 450000, PR China and [3]Department of Gastroenterology, The First Affiliated Hospital of Zhengzhou University, 1# Mianfang Street, Zhengzhou, Henan Province, 450052, China

*Corresponding author: Email: sprinkle.w@126.com, Tel(fax): +86 0371 65723528
Correspondence may also be addressed to Gaiping Zhang. Email: zhanggaiping2003@163.com

#These authors contributed equally to this work.

Abstract

DPL (http://www.peptide-ligand.cn/) is a comprehensive database of peptide ligand (DPL). DPL1.0 holds 1044 peptide ligand entries and provides references for the study of the polypeptide platform. The data were collected from PubMed-NCBI, PDB, APD3, CAMPR3, etc. The lengths of the base sequences are varied from 3 to78. DPL database has 923 linear peptides and 88 cyclic peptides. The functions of peptides collected by DPL are very wide. It includes 540 entries of antiviral peptides (including SARS-CoV-2), 55 entries of signal peptides, 48 entries of protease inhibitors, 45 entries of anti-hypertension, 37 entries of anticancer peptides, etc. There are 270 different kinds of peptide targets. All peptides in DPL have clear binding targets. Most of the peptides and receptors have 3D structures experimentally verified or predicted by CYCLOPS, I-TASSER and SWISS-MODEL. With the rapid development of the COVID-2019 epidemic, this database also collects the research progress of peptides against coronavirus. In conclusion, DPL is a unique resource, which allows users easily to explore the targets, different structures as well as properties of peptides.

Small peptide ligands have been highlighted over the past few decades on account of their particular advantages, such as less cost, little immunogenic responses and more stable physicochemical properties (1, 2). Especially, peptide ligands' chemical structures are highly compatible with those of the target proteins (3).

Peptide–protein interactions are ubiquitous in living cells and are an important part of the entire protein–protein interaction network. These interactions have attracted increasing attention due to their role in signaling and regulation and are therefore attractive targets for computational structure modeling. Peptide-mediated interactions are a major target for drug design because they are primarily present in signaling and regulatory networks. A reliable data set of non-redundant protein–peptide complexes is an indispensable basis for modeling and design, but current data sets of protein–peptide interactions tend to be biased towards specific types of interactions or limited to interactions with small ligands (4).

Peptide–protein interactions can happen in a lot of interaction networks and only need a small interface (5). As a result, these small molecules and inhibitory peptides are attractive drug targets (6, 7). This means that the synthetic peptides can be designed to change the specific interaction of disease or other signal pathways (8). Besides the peptide structure stored in the Protein Data Bank (PDB) (9), there are about 20 new items to show the interaction of small peptides each month (10). As the new and interesting structure of the protein–peptide complex is growing, our understanding of the interaction mechanism between protein and peptide also should be improved. Peptides tend to bind at the largest pocket available on the protein surface (11). To understand and analyze the interaction mechanism of protein and peptide, establishing a reliable database of peptide ligands is necessary. There are many protein–peptide interaction database based on sequences, such as Phospho.ELM (12), DOMINO (13), Pep Bank (14), SCANSITE (15), APD (16), BIOPEP (17) and ASPD (18). However, the database of peptide ligand (DPL) is a set of 1044 peptides for non-redundant protein–peptide complexes based on different binding targets.

Previous studies have reported combined with multiple peptide or protein with the heterogeneity of the structures of the domain (e.g. there are at least 13 different types of peptides was reported to SH3 domain structure (19)). For a detailed analysis of similar proteins and the interaction between different peptides, it needs a lot of data on the structure and ligand of protein–peptide complexes. To solve this problem, we created a DPL.

This DPL project has built a clear target of the peptide ligands database through the literature summary, including specific peptide information 1044, which provides a reference for the study of the polypeptide platform. All the peptides have a clear binding target, have to be experimentally verified and collect the 3D structures of all ligands and receptors. DPL is a unique resource, which allows users easily to explore the different structures as well as properties of peptides.

Construction and content

Data collection

Database of peptide ligand in DPL was collected from PubMed-NCBI, PDB, the 3rd version of the APD (APD3) and Collection of Anti-Microbial Peptides(CAMPR3) by using keywords such as 'antibacterial peptide', 'antiviral peptide', 'antifungal peptide', 'antiparasitic peptide', 'anticancer peptide', 'insecticidal peptide', 'anti-inflammatory peptides', 'anti-toxic peptides', 'protease inhibitors peptides', 'antioxidant peptides' 'anti-hypertension peptides' or 'signal peptides' (12–18).

Utility

The main web page of DPL contains the following aspects: Home, Database search, Tools, News, Links, Publications and Our team.

Home page

The use and main criteria of DPL are introduced on the Home page briefly. DPL is a specialized database for the collection of targeted binding polypeptides. There are three main criteria for data collection in this database: the peptides have a clear binding target; these peptides have to be experimentally verified, and this database strives to collect the 3D structure of all ligands and receptors. The prediction method of the structure of peptide using the web tools as CYCLOPS, I-TASSER or SWISS-MODEL (20–22).

Search page

A quick search was constructed on the search page through some appropriate keywords, such as peptide name, ID, sequence, function or receptor name, function. The search will ambiguously match any residue in the peptide name, ID, sequence, function or receptor name, function. To get more accurate results, please try to enter more detailed search terms.

Results and discussion

Sequences

Figure 1 summarizes the basic amino acid distribution. As shown, alanine, lysine, leucine and valine make up the predominant composition in peptides (See Figure 1).

The length of the base sequence varied from 3 to78. As the length of the peptide chain is different, the proportion is different. The most proportion of peptide is in the length of 11–20 (53%), followed by 1–10 (27%); 51–80 proportion is at least (1%) (See Figure 2).
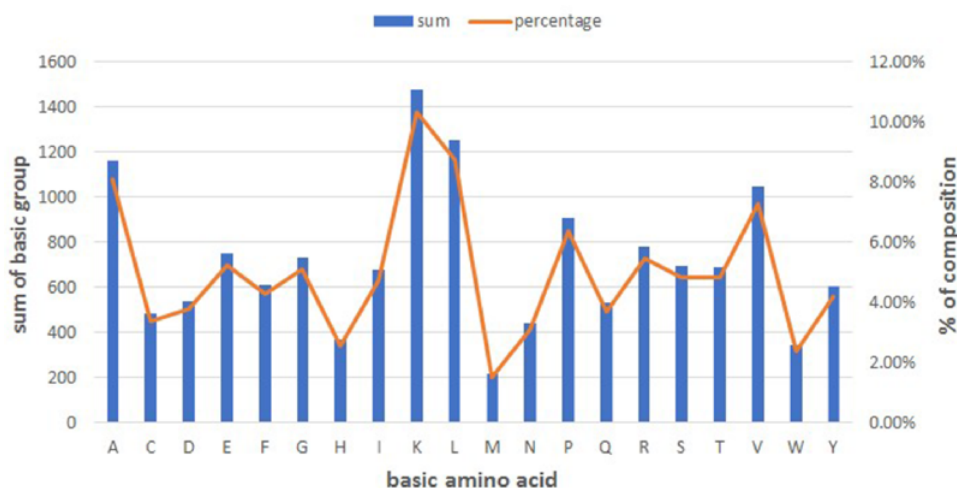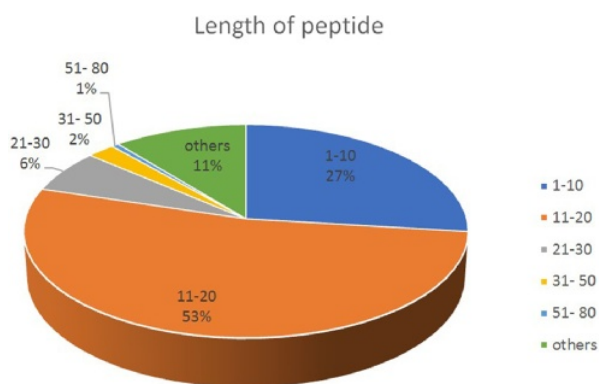
Figure 1. Basic amino acid distribution in DPL.

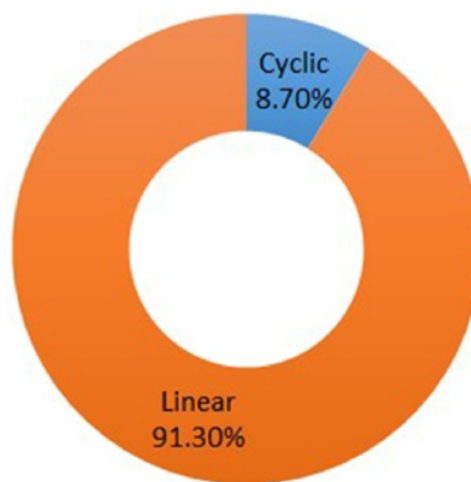**Figure 2.** The proportion of different length of peptide in DPL.

The type of peptide

The DPL database has 1044 entries in total. There are two kinds of peptide structures, such as linear peptide and cyclic peptide, respectively, in which linear peptide has 923 entries, accounts for 91.30%; cyclic peptide has 88 entries, accounted for 8.70% (See Figure 3).

Function

The function of peptide collected by DPL database is very wide, such as 'antibacterial peptide', 'antiviral peptide', 'antifungal peptide', 'antiparasitic peptide', 'anti-cancer peptide', 'insecticidal peptide', 'anti-inflammatory peptides', 'anti-toxic peptides', 'protease inhibitors peptides', 'antioxidant peptides', 'anti-hypertension peptides', 'signal peptides', etc.

A total of 540 kinds of antiviral peptides account for 53.14%; 267 kinds of others accounted for 26.41%; 55 kinds of signal peptides accounted for 5.44%; 48 kinds of protease inhibitors accounted for 4.75%; 45 kinds of anti-hypertension account for 4.45%; 37 kinds of anticancer

Figure 3. The proportion of two kinds of peptide structure in DPL.

peptides accounted for 3.66%; 5 kinds of antioxidant peptides accounted for 0.49%; 5 kinds of anti-parasitic peptides accounted for 0.49%; 3 kinds of antibacterial peptides accounted for 0.30%; 4 kinds of anti-inflammatory peptides accounted for 0.40%; 2 kinds of anti-toxin peptides accounting for 0.20% (See Figure 4).

The source of peptide structures

The structure of peptide has different sources, such as CYCLOPS (846 entries), I-TASSER (1 entry), SWISS-MODEL (25 entries), RCSB Protein Data Bank (50 entries), no structure (67 entries), etc. The proportion of different sources of the structure is shown in Figure 5.
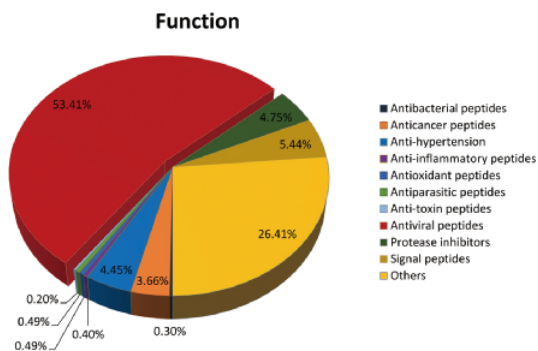
Function

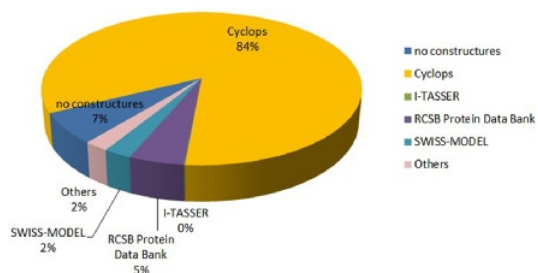**Figure 4.** The functions of peptides in DPL.

Figure 5. The sources of peptide structure in DPL.

Targets of peptide

There are 270 different kinds of peptide targets.

Peptide ligands for COVID-2019

With the rapid development of the COVID-2019 epidemic, this database also collects and organizes the research progress of peptides against coronavirus (Table 1). Detailed information such as peptide sequences, targets and research literature are recorded in this database.

The most influential databases in this field are PDB, APD3, CAMP3, etc. This resource is powered by the PDB archive-information about the 3D shapes of proteins, nucleic acids and complex assemblies that help students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease (23). APD3 reported 2619 peptides. New web pages for FAQs, interesting AMP discovery timeline, classification, nomenclature, AMP facts, My tools, Sequence download and APD News have been created (16). A unified peptide classification system has been proposed and introduced in APD. Besides, the prediction interface has been improved and additional peptide properties can be calculated in

APD. CAMPR3 has been created to expand and accelerate antimicrobial peptide family based studies. Antimicrobial peptides have family specific sequence composition which can be mined to discover and design novel AMPs (24). In a word, each database has its advantages and disadvantages.

Peptide ligands can simulate protein–protein interactions and have large binding interfaces with receptors; thus, they possess much higher binding affinity and specificity than small-molecule ligands. Peptides offer a potent resource for targeted drug delivery. Compared to protein ligands, peptides have many advantages, including better penetration, ease of synthesis and lower immunogenicity and cost. Large-scale synthesis of peptides presents a convenient and economical option for drug use; also, due to the abundant chemical groups in peptides, they are suitable for manipulation.

However, this article briefly introduces the DPL database to collect many peptide ligands for users. This DPL database has built a clear target of the peptide ligands database through the literature summary, including specific peptide information 1044, which provides a reference for the study of the polypeptide platform. All the information of peptides and receptors collected in DPL provides material for molecular docking and virtual screening in future. DPL database will build the virtual peptide library through the computer program and set up a molecular docking platform and analyze the differences between global and local molecular docking results in next version. All the peptides and targets have a clear binding target. 16 items anti-coronavirus peptides also were added in DPL, it provides technical support for the target screening and research of vaccines and drugs. DPL database is a unique resource and still being updated, which allows users easily to explore the different structures as well as properties of peptides.

Conclusion

With 1044 entries, DPL is an open-access, manually curated database with a clear binding target, be experimentally verified, and collect the 3D structure of all peptide ligands and receptors. To the best of authors' knowledge, DPL is the only database available to the public, which provides comprehensive information on DPL, especially provides structures of all peptides. User-friendly interfaces have been established to facilitate peptides searching, browsing and alignment. DPL should help promote our understanding of peptide ligands and should provide a valuable resource for the development of peptide application. We believe that the DPL will be very useful for scientists in peptide research.

Table 1. Detailed information of some peptides against coronavirus collected in DPL

DPL_ID	Sequence	Source	Virus	IC50	Model
1028	GYHLMSFPQAAPHGVVFLHVTW	S2	SARS	~2 μM	VeroE6, L2
1030	GYFVQDDGEWKFTGSSYYY	S2	MHV	4 μM	VeroE6, L2
1031	LTQINTTLLQDLTYEMLSLQQVVKALNESYIDLKEL	HR2	MERS	~3.013 μM	293T
1032	SIPNFGSLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELGNY	HR2	MERS	~0.5 μM	293T/EGFP + HUH-7
1033	SLTQINTTLLDLTYEMLSLQQVVKALNESYIDLKELY	HR2	MERS	~0.97 ± 0.15 μM	293T/EGFP + HUH-7
				~0.6 μM	Vero
				~0.6 μM	Calu-3
				13.9 μM	HFL
1034	SLTQINTTLLDLEYEMRSLQQVVKALNESYIDLKEL	HR2	MERS	~0.85 ± 0.08 μM	293T/EGFP + HUH-7
1035	SLTQINTTLLDLEYEMKKLEEVVKKLEESYIDLKEL	HR2	MERS	~0.55 ± 0.04 μM	293T + HUH-7
1036	NGAICWGPCPTAFRQIGNCGHRKVRCCKIR	β-4	MERS	5 μM	Mice
1037	FGGASCCLYCRCHIDHPNPKGFCDLKGKY	Nsp10	SARS	160 μM	*E. coli*
1038	GGASCCLYCRCH	Nsp10	SARS	160 μM	*E. coli*
1039	LFRLIKSLIKRLVSAFK	AMP	SARS	7.15 μg/ml	MDCK
1040	HVTTTFAPPPPR	pAMN	TGEV	11 μg/ml	ST
1041	SVVPSKATWGFA	pAMN	TGEV	15 μg/ml	ST
1042	YKYRYL	RBD	SARS	KD = 46 μM	VeroE6
1043	PSSKRFQPFQQFGRDVSDFT	S	SARS		293T
1044	CANLLLQYGSFCTQLNRALSGIA	S	SARS		293T

Acknowledgements

Availability

DPL can be freely accessed at the URL: http://www.peptide-ligand.cn/. Users are invited to submit their suggestions and peptide sequences by subscription to the site. The database will be expanded and improved by the authors in the course of time.

Authors' contributions

F.W. and G.Z. conceived of the idea and supervised the study. N.L. constructed and maintained the database and web interface. F.W. wrote the program. C.W. and G.X. designed the system architecture. S.C., Q.X. and Y.Z. participated in the data acquisition. M.H. participated in the statistical analysis. F.W. and N.L. wrote the manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Conflicts of Interest The authors declare that there are no conflicts of interest.

References

1. Fang, Y., Lin, D. and Yao, S. (2018) Review on biomimetic affinity chromatography with short peptide ligands and its application to protein purification. *J. Chromatogr. A*, **1571**, 1–15.

2. Mishra, A., Bansal, R., Sreenivasan, S. *et al.* (2020) Structure-based design of small peptide ligands to inhibit early-stage protein aggregation nucleation. *J. Chem. Inf. Model.*, **60**, 3304–3314.

3. Vanhee, P., Der Sloot, A.M.V., Verschueren, E. *et al.* (2011) Computational design of peptide ligands. *Trends Biotechnol.*, **29**, 231–239.

4. Vanhee, P., Reumers, J., Stricher, F. *et al.* (2010) PepX: a structural database of non-redundant protein–peptide complexes. *Nucleic Acids Res.*, **38**, 545–551.

5. Das, A.A., Sharma, O.P., Kumar, M.S. *et al.* (2013) PepBind: a comprehensive database and computational tool for analysis of protein-peptide interactions. *Genomics Proteomics Bioinf.*, **11**, 241–246.

6. Jiang, Z., Guan, J., Qian, J. *et al.* (2019) Peptide ligand-mediated targeted drug delivery of nanomedicines. *Biomater. Sci.*, **7**, 461–471.

7. Ruff, L.E., Pfeilsticker, J.A., Johnsen, N.E. *et al.* (2018) Identification of peptide mimotope ligands for natalizumab. *Sci. Rep.*, **8**, 14473–14473.

8. Hruby, V.J. (2019) Multivalent peptide and peptidomimetic ligands for the treatment of pain without toxicities and addiction. *Peptides*, **116**, 63–67.

9. Berman, H.M., Kleywegt, G.J., Nakamura, H. *et al.* (2012) The Protein Data Bank at 40: reflecting on the past to prepare for the future. *Structure*, **20**, 391–396.

10. Das, A.A., Sharma, O.P., Kumar, M.S. *et al.* (2013) PepBind: a comprehensive database and computational tool for analysis of protein-peptide interactions. *Genomics Proteomics Bioinf.*, **11**, 241–246.

11. London, N., Raveh, B. and Schueler-Furman, O. (2012) Modeling peptide-protein interactions. *Methods Mol. Biol. (N.Y.)*, **857**, 375–398.

12. Dinkel, H., Chica, C., Via, A. *et al.* (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.*, **39**, 261–267.

13. Ceol, A., Chatraryamontri, A., Santonico, E. *et al.* (2007) DOMINO: a database of domain–peptide interactions. *Nucleic Acids Res.*, **35**, 557–560.

14. Shtatland, T., Guettler, D., Kossodo, M. *et al.* (2007) PepBank—a database of peptides based on sequence text mining and public peptide data sources. *BMC Bioinf.*, **8**, 280–280.

15. Obenauer, J.C., Cantley, L.C. and Yaffe, M.B. (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **31**, 3635–3641.

16. Wang, G., Li, X. and Wang, Z. (2016) APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.*, **44**, 1087–1093.

17. Minkiewicz, P., Dziuba, J., Iwaniak, A. *et al.* (2008) BIOPEP database and other programs for processing bioactive peptide sequences. *J. AOAC Int.*, **91**, 965–980.

18. Valuev, V.P., Afonnikov, D.A., Ponomarenko, M.P. *et al.* (2002) ASPD (Artificially Selected Proteins/Peptides Database): a database of proteins and peptides evolved in vitro. *Nucleic Acids Res.*, **30**, 200–202.

19. Li, S.S.C. (2005) Specificity and versatility of SH3 and other proline-recognition domains: structural basis and implications for cellular signal transduction. *Biochem. J.*, **390**, 641–653.

20. Anafi, R.C., Francey, L.J., Hogenesch, J.B. *et al.* (2017) CYCLOPS reveals human transcriptional rhythms in health and disease. *Proc. Nat. Acad. Sci. U S A.*, **114**, 5312–5317.

21. Yang, J., Yan, R., Roy, A. *et al.* (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods*, **12**, 7–8.

22. Waterhouse, A., Bertoni, M., Bienert, S. *et al.* (2016) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.*, **46**, W296-W303.

23. Whitmore, L. and Wallace, B.A. (2004) The Peptaibol Database: a database for sequences and structures of naturally occurring peptaibols. *Nucleic Acids Res.*, **32**, 593–594.

24. Waghu, F.H., Barai, R.S., Gurung, P. *et al.* (2016) CAMPR3: a database on sequences, structures and signatures of antimicrobial peptides. *Nucleic Acids Res.*, **44**, 1094–1097.