CyclicPepedia (V1.3.1) Tutorial

Table of Contents

Overvi	ew3
Browsi	ng interface4
1)	Cyclic peptide4
2)	Source
3)	Function/ Family/ Target6
4)	Reference
5)	DrugBank9
Cyclic]	peptide details page10
Search	tools
1)	Advanced search18
2)	Structure search21
3)	Sequence search23
i.	Local alignment
ii.	Graph alignment
Cyclic]	peptide tools
1)	Structure to sequence
2)	Sequence to structure
	SociEtro for accortial amine acida

ii.	Seq2Struc for non-essential amino acids	34
3)	Peptide Property Prediction	.36
i.	Chemical and Physical Property Prediction (CPPP)	36
ii.	Peptide Sequence Property Prediction (PSPP)	38
4)	Structure format transformation	.40
5)	Sequence format transformation	.43
Table 1	. List of the chemical and physical properties	.46
Table 2	. List of the peptide sequence properties	.47
Table 3	. Sequence formats	.49

Overview

CyclicPepedia is a pioneering database that encompasses a large amount of known cyclic peptides. This repository houses data on 8744 cyclic peptides, with a wealth of information regarding cyclic peptide sources, categorizations, structural characteristics, pharmacokinetic profiles, physicochemical attributes, patented drug applications, and a collection of relevant publications. Therefore, CyclicPepedia forms a comprehensive knowledge network of cyclic peptides, thereby facilitating advancements in the early stage of cyclic peptide drug development.



Browsing interface

CyclicPepedia provides multiple browsing interfaces dedicated to cyclic peptides, sources, functions, families, targets, references, and DrugBank. Click the **Browse** in the top navigation bar to enter different browsing interfaces.



1) Cyclic peptide

The cyclic peptide browsing interface consists of two parts: a bar plot displaying the distribution of cyclic peptide sequence length and a table of all cyclic peptides.



- **a.** The bar plot can be downloaded through the button in the top-right corner.
- **b.** Click the **Export as Excel** button to download the search results.
- **c.** The number of cyclic peptide entries (i.e., 10, 25, 50, and 100) per page can be adjusted through the drop-down list.
- **d.** Use the **Search** box to filter cyclic peptides by ID, name, sequence, families, sources, functions, and targets.
- e. The Sequence, Family, Source, Function, and Target columns provide filtering functionality for the search results.
- f. Click the CyclicPepedia ID to enter the corresponding cyclic peptide details page.

2) Source

The **Source** browsing interface displays the top 20 biological sources of cyclic peptides, and a full list of sources can be accessed by clicking the **More sources** button.



Click the **source name** to access the details page. The number of cyclic peptides included in this source classification is shown in the blue box. The Taxonomy ID, a general description, and a link to Wikipedia are listed.



The details page contains a statistical plot of cyclic peptide sequence length and a list of cyclic peptides derived from this source.



3) Function/ Family/ Target

The layout and functionality of the Function, Family, and Target browsing interfaces are similar to those of the Source browsing interface, please refer to the description above.

Functions

Functional annotations classify the biological activities and pharmacological characteristics of cyclic peptides.



Family

Family annotations provide a classification of cyclic peptides, which may based on the source, biological activities, pharmacological characteristics, or physicochemical properties.

		es				=		
				Orbitide	Inhibitse-Like	Bracelet Subtamily	Detensin-Like	. Chi constada
	Constants	Cyclotide			Beyanah Tryysin lehikitlar		karangala Intern	
	Conotoxin					Quinquine		
		Alpha conotoxin			Cyclic Depelopytos			
Ø	Conotoxin 2360 Toxin > Conotoxin> Unknown	$\langle \! \rangle$	Cycloti Cyclotid	ide 1025 e				
	Alpha conotoxin 672 Toxin > Conotoxin> Alpha conotoxin		Orbitid Orbitide	e 197				
	Frog Skin Active Peptide 187 Frog Skin Active Peptide	$\langle \rangle$	Brevini Frog Ski	n Subfamily n Active Peptie	167 de > Brevi	nin Subfamily		

Targets

Cyclic peptides have a larger area of contact with targets than small-molecule drugs. Therefore, they have higher specificity, target-binding affinity, and fewer side effects.



4) Reference

The reference browsing interface displays a full list of references with links to PubMed. Click the **arrow** to show abstracts.

ference											
PUBMED_ID	TITLE 🕐	O IOU	JOURNAL								
10022120	Tyrosine phosphorylation	10.1038/sj.onc.1202411.	Oncogene								
Tyrosine phosphorylatic Abstract • Cbl-b, a mammalian ho region containing multip	n and complex formation of Cbl molog of Cbl, consists of an N-termin ple proline-rich stretches and potent	-b upon T cell receptor stim nal region (CbI-b-N) highly homo ial tyrosine phosphorylation sites	ulation logous to oncogenic v-Cbl, . In the present study, we d	a Ring finger, and a C-terminal emonstrate that upon							
engagement of the Toc phosphorylated on tyro Syk associates and pho tyrosine phosphorylatic Grb2 and becomes ass Crk-L-binding sites wer implicate that Cbl-b is i	engagement of the T cell receptor (TCR), endogenous Cbl-b becomes rapidly tyrosine-phosphorylated. In heterogeneous COS-1 cells, Cbl-b was phosphorylated on tyrosine residues by both Syk- (Syk/Zap-70) and Src- (Fyn/Lck) family kinases, with Syk kinase inducing the most prominent effect. Syk associates and phosphorylates Cbl-b in Jurkat T cells. A Tyr-316 Cbl-binding site in Syk was required for the association with and for the maximal tyrosine phosphorylation of Cbl-b. Mutation at a loss-of-function site (Gly-298) in Cbl-b-N disrupts its interaction with Syk. Cbl-b constitutively binds Grb2 and becomes associated with Crk-L upon TCR stimulation. The Grb2- and the Crk-L-binding regions were mapped to the C-terminus of Cbl-b. The Crk-L-binding sites were further determined to be Y655DVP and Y709KIP, with the latter being the primary binding site. Taken together, these results implicate that Cbl-b is involved in TCR-mediated intracellular signaling pathways.										
10026169	Identification of Grb4/Nck	10.1074/jbc.274.9.5542.	J Biol Chem	•							
10048485	Prediction of the coding s	10.1093/dnares/5.6.355	DNA Res.	•							
10051406	The gene structure of the	10.1006/geno.1998.5692.	Genomics	•							

5) DrugBank

The DrugBank data is listed in the DrugBank browsing interface. Click the **tab name (a)** to show drugs according to their categories. Click the **DrugBank ID (b)** or the **arrow (c)** to show the **drug summary (d)**.

DrugBank	[15235]							
All drugs	Cyclic Peptide	Small Molecule	Peptide	Protein Ant	tibody	а		
DRUGBA	INK ID	NAME	ТҮРЕ	GROUPS		BIOCLASS		
воо	001	Lepirudin	Biotech	Approved, With	hdrawn	Protein Based Therapies	C 💌	
Categories				Summary				٦d
 Amino Av Antichoro Antithror Antithror Blood an Cardiova Enzyme Fibrin Mu Hematol Peptides Protease Protease Proteins Serpins Thrombia 	cids, Peptides, and Pro- ulants mbin Proteins mbins d Blood Forming Orga scular Agents Inhibitors odulating Agents ogic Agents Inhibitors rotease Inhibitors n Inhibitors	ns		Lepirudin is anticoagula	s a protein-ba	ased direct thrombin s with heparin-induc	n inhibitor used as an ed thrombocytopenia.	
DB00	002	Cetuximab	Biotech	Approve	d	Protein Based Therapies	•	

Cyclic peptide details page

The cyclic peptide details page contains basic information, structure, sequence, biologic determination, chemical and physical properties, binding target, manufacturers, forecasting tools, information sources, and references. These data can be quickly accessed through the **navigation bar (a)** on the left.

		CyclicPepedia Know	ledge Base
 ➢ Hor A Structure Sequence Biologic Determinat Kr. Chemical and Physical ⓒ Binding Target ※ Manufacturers ※ Forecasting tools information Source Reference 	ne ﷺ Browse~ Q Sea Maculosin Basic informa СРКВ ID IUPAC Name © Synonyms	arch ~ ★ Tools ~ Y Statistics DataSource ⑦ ation _ _ _ _ _ _ _ _ _	Help Download Q Search Iropyrrolo[1,2-a]pyrazine-1,4-dione (35,8As)-3-(4-Hydroxybenzyl)Hexahydropyrrolo[1,2- A]Pyrazine-1,4-Dione (35,8As)-Hexahydro-3-[(4- Hydroxyphenyl)Methyl]Pyrrolo[1,2-A]Pyrazine-1,4-Dione Chebi:6631

Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060

1) Basic information: This section displays CyclicPepedia ID, IUPAC name, synonyms, source, family, function, description information, and a knowledge network.

Maculosin Basic information blacktrianCPKB ID CP00060 (3S,8aS)-3-[(4-hydroxyphenyl)methyl]-2,3,6,7,8,8a-hexahydropyrrolo[1,2-a]pyrazine-1,4-dione IUPAC Name Synonyms (35,8Ar)-3-(4-Hydroxybenzyl)Hexahydropyrrolo[1,2-A]Pyrazine-1,4- (35,8As)-3-(4-Hydroxybenzyl)Hexahydropyrrolo[1,2-A]Pyrazine-1,4-Dione Dione (3S,8As)-3-[(4-Hydroxyphenyl)Methyl]-2,3,6,7,8,8A-(3S,8As)-Hexahydro-3-[(4-Hydroxyphenyl)Methyl]Pyrrolo[1,2-Hexahydropyrrolo[1,2-A]Pyrazine-1,4-Dione A]Pyrazine-1,4-Dione Chebi:6631 1W1Y Source Bacillus cereus [Division : Bacteria] Taxonomy :1396 (Unassigned-Bacillota-Bacillales-Bacilli-Bacillaceae Bacillus) Wikipedia: Bacillus cereus Streptomyces [Division : Bacteria] Taxonomy :1883 (Unassigned-Actinomycetota-Kitasatosporales-Actinomycetes-Streptomycetaceae Unassigned) Wikipedia: Streptomyces PubChem Function DrugBank Chitinases, antagonists & inhibitors Information Maculosin is a homodetic cyclic peptide that is a dipeptide composed of L-proline and L-tyrosine joined by peptide linkages. It has a role as a metabolite. It is a dipeptide, a homodetic cyclic peptide, a pyrrolopyrazine and a member of phenols. It is functionally related to a L-proline and a Ltyrosine. Mauritine A is a cyclic peptide. It has a role as a metabolite DrugBanklDB04520 PubCheml119404

Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060

Click the **network node** to enter the cyclic peptide details page or the corresponding external database.



Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060

 Structure: This section provides data on cyclic peptide molecular formula, molecular weight, SMILES, InchI, InChI Key, and two-dimensional and three-dimensional structures. Click the **similarity structure (a)** button to search for cyclic peptides with similarity scores > 0.9. Structure files can be downloaded by clicking the **download (b)** button.

Structure ⊥	Q similarity structure	🛃 download ~ b
Molecular Formula	C14H16N2O3	DOWNLOAD STRUCTURE 2D IMG
Molecular Weight	260.1160924 g/mol	300x300 pixels 100x100 pixels
SMILES 🖒	O=C1N[C@@H](Cc2ccc(0)cc2)C(=0)N2CCC[C@@H]12	500x500 pixels
RUN SEA Predictions		FILE 2D_Mol 산 3D_Mol 산 3D_PDB 산
InChi 🗋	InChI=1S/C14H16N2O3/c17-10-5-3-9(4-6-10)8-11-14(19)16-7-1-2-12(16)13(18)15-11/h3-6, 8H2,(H,15,18)/t11-,12-/m0/s1	11-12,17H,1-2,7-
InChiKey 🗋	LSGOTAXPWMCUCK-MRLYJYMGNA-N	

Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060#Structure

The sources of structural data and the types of 3D structures (e.g., complex) are listed on the right.



Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060#Structure



Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060#Structure



Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060#Structure

3) Sequence: This section presents different sequence formats, for example, one-letter code, IUPAC condensed, amino acid chain, graph representation, and SVG image, as well as a plot of amino acid composition and a report of the Structure-to-Sequence (Struc2Seq) transformation.



Gramicidin S: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00038#Sequence

Refer to the Tools|Structure to Sequence for a description of the Struc2Seq report.

Structure 2 Sequence Report

Structure-to-Sequence (s2s) is a computing process based on <u>RDkit</u> and the characteristics of cyclic peptide sequences, which can convert cyclic peptide SMILES into sequence information. This process mainly relies on the completeness of the <u>monomer reference library</u>. You can access our default monomer reference library through <u>download link</u>. The details of s2s are available on <u>dfwlab/cyclicpepedia</u> on Github. And you can use this tool online on the <u>cyclicpepedia</u>.

Version : 1.0.1 (2023-12-26)

Load SMILES :

 $\label{eq:static_stat$

SMILES is corrected!

Identify peptide skeleton and renumber atoms

Gramicidin S: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00038#Sequence

Click the Local alignment or Graph alignment button to query similar sequences. Refer to the Search Sequence Search for details.

Sequence ⊥	Q Local alignment Q Graph alignment
------------	-------------------------------------

4) Biologic Determination: This section lists bioassay results related to the cyclic peptide.

BIO	Assay Results						
Sho	w 4 entrie	es				Search:	
		VALUE(UM) 🗘	OPERATION 🛇	TARGET NAME \Diamond	BIOASSAY NAME 🗘	BIOASSAY AID 🛇	SUBSTANCE SID \Diamond
	Unspecified	37.71	Equal to	Spodoptera exigua (beet armyworm)	IC50	1092124	103456411
	Unspecified	42.48	Equal to		IC50	1092125	103456411
	Unspecified	47.79	Equal to		IC50	1092126	103456411
				Chain A, Chitinase B	IC50		

Maculosin: https://www.biosino.org/iMAC/cyclicpepedia/detail?id=CP00060#Biologic%20Determination

- **5)** Chemical and Physical Properties: It contains two parts—the structural properties and the sequence properties computed by CyclicPepedia.
- 6) Binding Target: Data on associated targets are presented in this section. Click the arrow button to expand/collapse the information tab. Click the Detail button to access the target's details page.

Chitinase B		^
Uniprot: P11797 C		Detail
Kind: Protein>Chitinase		
Organism: Serratia marce	escens	
Evidevce: DrugBank		
Sequence: MSTRKAVIGY NPSLRIMFSIGGWYYSNDL ALPYQLTIAGAGGAFFLSR` DAAVQQHLMMEGVPSAKI YQRLWNDKTKTPYLYHAQ VGPGNLPIMTAPAYVPGTT	YFIPTNQINNYTETDTSVVPFPVSNITPAKAKQLTHINFSFLDINSNLECAV LGVSHANYVNAVKTPAARTKFAQSCVRIMKDYGFDGVDIDWEYPQAAE\ YYSKLAQIVAPLDYINLMTYDLAGPWEKITNHQAALFGDAAGPTFYNALF IVMGVPFYGRAFKGVSGGNGGQYSSHSTPGEDPYPNADYWLVGCDEC\ NGLFVTYDDAESFKYKAKYIKQQQLGGVMFWHLGQDNRNGDLLAALD TYAQGALVSYQGYVWQTKWGYITSAPGSDSAWLKVGRLA	WDPATNDAKARDVVNRLTALKAH VDGFIAALQEIRTLLNQQTIADGRQ REANLGWSWEELTRAFPSPFSLTV VRDKDPRIASYRQLEQMLQGNYG RYFNAADYDDSQLDMGTGLRYTG
Conoral Eurotion:		
General Function.		

- 7) Manufacturers: CyclicPepedia provides links to popular manufacturers such as Merck, Baxter Healthcare Corp, and Upsher-Smith laboratories.
- Forecasting tools: It presents connections to CyclicPepedia tools such as Structure-to-Sequence conversion, Sequence-to-Structure conversion, Structure Properties, and Sequence Properties computation, as well as several external predictive tools.
- 9) Information Source: This section lists links to external information sources.
- **10) Reference:** This section lists associated literature with links to PubMed. Click the **arrow** to show abstracts.

PUBMED_ID	TITLE 🕐	DOI ()	JOURNAL	
2859510	New specific radiolig	10.1016/0024- 3205(85)90155-9.	Life Sci	•
New specific radio	ligand for one subpopulatio	n of brain somatostatin re	eceptors	
Abstract				
Abstract • Cyclic octapeptid	e analogues of somatostatin (S	S) like SMS 201-995 H-(D) Pr	ne-Cys-Phe-(D) Trp	-Lys-Thr-Cys-Thr(ol) or its

Search tools

CyclicPepedia offers four search methods: quick full-text search (a) on the homepage,

advanced search, structure search, and sequence search. Click the Search (b) to select search methods.



1) Advanced search

Click Search Advanced Search to enter the Advanced Search page.



Advanced search provides users with multiple criteria, for example, **sequence and structure information**, **physiochemical properties**, **sequence properties**, and **biological annotation information**, to create custom search queries. Click the **arrow button** to expand the **filtering criteria**.

Advanced search

Basic Information	Basic Information								
Name or Synonyms	Enter name or synonyms								
Sequence and Structure Information A									
Has Sequence	All	Sequence Length	0	-	999				
Has Structure	All	Exact Mass	0	-	9999				
Physicochemical I	Properties 😣								
Sequence Propert	ties ≽								
Biological Annota	tion Information st								

Enter the cyclic peptide name and/or select filtering criteria, and click the **Search** button to get search results. Click the **Clear input** button to clear all inputs.

Clear input

Search

Example: We want to search for anti-bacterial cyclic peptides that have sequence and structure information and have amino acid sequence length > 5. The filtering criteria are: "Has Sequence = Yes" & "Sequence Length > 5" & "Has Structure = Yes" & "Function = Anti-Bacterial (743)"

Click the **Search** button.

Sequence and St	ructure Information 😞					
Has Sequence	Yes	Seq	uence Length	5	-	999
Has Structure	Yes	E	Exact Mass	0	-	9999
Physicochemical	Properties ≽					
Sequence Proper	ties ≽					
Biological Annota	ition Information 😞					
Source	All		F	unction	✓ All	
					Anti-Microbial (1005)	
Family	All			Target	Anti-Bacterial (743)	
					Anti-Gram- (599)	
					Anti-Fungal (409)	
					Anti-Biotics (277)	
					Toxin (256)	
						*

The **Search result table** is presented at the bottom of the page.

- a. Click the Export as Excel button to download the search results.
- **b.** The number of cyclic peptide entries (i.e., 10, 25, 50, and 100) per page can be adjusted through the drop-down list.
- **c.** Use the **Search** box to filter search results by ID, name, sequence, families, sources, functions, and targets.
- **d.** The **Sequence**, **Family**, **Source**, **Function**, and **Target** columns provide filtering functionality for the search results.
- e. Click the CyclicPepedia ID to enter the corresponding cyclic peptide details page.

Cycl	lic peptides				a 🛃	oport as Excel
Show 10 🗸	entries b				C Search:	
ID	Name	d Sequence ?	Family 🕜 All 🗸	Source 🕐	Function 😢	Target 🕜 All 🗸
СР00142	Enniatin B		Enniatin	Halosarpheia;Pinus sylvestris	Toxin;Enzyme inhib itor;Anti-Microbial;I mmunomodulator y;Anti-Fungal;Anti- Bacterial	
CP00346	Valinomicin		Valinomycin	Streptomyces	Anti-Viral;Anti-Infe ctive;Anti-Microbia I;Anti-Bacterial	
CP00423	Fusafungine		Cyclic Depsipeptide	Fusarium tricinctu m	Anti-Biotics;Anti-In fective;Anti-Bacteri al	
CP00709	Circulin B	GVIPCGESCVFIPCI STLLGCSCKNKVC YRN	Cyclotide	Chassalia parvifolia	Hemolytic;Anti-Hi v;Anti-Bacterial	
CP00788	Bacitracin A	CLEIKAIFHDN	Antimicrobial Peptide	Bacillus licheniform is	Anti-Microbial;Anti -Bacterial	C55-isoprenyl pyrophosphate; Alpha-2-macro globulin;Insulin- degrading enzy me

2) Structure search

Click Search|Structure Search to enter the Structure Search page.

A		CyclicPepedia Knowledge Base	Visitor Visitor
🟠 Home	8≣ Browse √ 0	Q Search ∨ 💥 Tools ∨ 🕎 Statistics 🖯 DataSource ⑦ Help ∨ 介 Download	Q Search
	Search fo	Advanced Search Structure Search Information	
	Cyclic pepu	Sequence Search ae search	Q
	Example: N receptor	ame :Aureobasidin E ; Family :Frog Skin Active Peptide ; Source :Conus ; Function :Anti-Bacterial ; Target :	Somatostatin

Tip! Structure search may take a long time, depending on input molecules and parameter selection.

Step 1. Upload structure files in PDB/SDF formats or paste your SMILES into the left panel.

Example:

```
CC1C(C(=0)NC(C(=0)N2CCC2C(=0)N(CC(=0)N(C(C(=0)O1)C(C)C)C)C)C(C)C)NC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)NC6C(OC(=0)C(N(C(=0)C)C)C)C)C)C)C)C)C(C)C)N)C
```

Step 2. Select search type, for example, exact match, substructure search, and similarity search.

Step 3. For **similarity search**, users can choose the **molecular fingerprint types** and the **similarity metrics**. The fingerprint types include **RDKit Fingerprint**, **MACCS Keys** (Molecular ACCess System), and **Morgan Fingerprint**; The similarity metrics include **Tanimoto similarity** and **Dice similarity**.

Step 4. Adjust the slider to select the Similarity threshold range.





The Search result table is presented at the bottom of the page.

- a. Click the Export as Excel button to download the search results.
- **b.** The number of cyclic peptide entries (i.e., 10, 25, 50, and 100) per page can be adjusted through the drop-down list.
- **c.** Use the **Search** box to filter search results by ID, name, molecular formula, and similarity score.
- d. Click the CyclicPepedia ID to enter the corresponding cyclic peptide details page.

Exact mate	E	Export as Excel			
Query SMILES : CC1C(C(=0)NC(C(=0)N2CCC22C(=0)N(CC(=0)O1)C(C)C)C)C(C)C)NC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)C)AC6C(OC(=0)C(N(C(=0)C)C)C)C(C)C)C)C(C)C)NC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)C)AC6C(OC(=0)C)A(C(=0)C)C)C(C)C)C(C)C)NC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)C)AC6C(OC(=0)C)A(C(=0)C)C)C(C)C)C(C)C)C(C)C)NC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)C)AC6C(OC(=0)C)A(C(=0)C)A(C)C)C)C(C)C)C(C)C)NC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)C)A(C)C)C(C)C)C(C)C)C(C)C)AC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C(=C(C5=N4)C(=0)C)A(C)C)C(C)C)C(C)C)A(C)C)C(C)C)AC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C)AC6C(OC(=0)C)A(C(=0)C)A(C)C)C)C(C)C)C(C)C)A(C)C)C(C)C)AC(=0)C3=C4C(=C(C=C3)C)OC5=C(C(=0)C)A(C)C)C(C)C)C(C)C)A(C)C)C(C)C)A(C)C)C(C)C)AC(=0)C3=C4C(=CC(=C3)C)OC5=C(C(=0)C)A(C)C)C(C)C)C(C)C)A(C)C)C(C)C)A(C)C)C(C)C)AC(=0)C3=C4C(=CC(=C3)C)OC5=C(C(=0)C)A(C)C)C(C)C)C(C)C)A(C)C)C(C)C)A(C)C)C(C)C)A(C)C)AC(=0)C3C(C)C)A(C)C)AC(=0)CACACACACACACACACACACACACACACACACACACA					
Structure	CPID	Name	Formula	Score 😮	
¢ j ©	d CP00001	Actinomycin D	C62HB6N12O16	1.0	
¢,ô	CP00315	Dactinomycin	C62H86N12O16	1.0	
Ç	CP00904	2-Amino-4,6-Dimethyl-3-Oxo-1-N,9-N-Bis[(3R)-7,11,14-Trimethyl-2,5,9,12,15	C62HB6N12O16	1.0	

3) Sequence search

Click Search|Sequence Search to enter the Sequence Search page. The sequence search is divided into local alignment and graph alignment.

		CyclicPepedia Knowledge Base	Visitor Visitor
🟠 Home	B≣ Browse ∨	$\bigcirc Search \lor \hspace{0.2cm} \not{\hspace{0.2cm} X} \hspace{0.2cm} {\rm Tools} \lor \hspace{0.2cm} \underline{\mathbb{V}} \hspace{0.2cm} {\rm Statistics} \hspace{0.2cm} \boxminus \hspace{0.2cm} {\rm DataSource} \hspace{0.2cm} \textcircled{O} \hspace{0.2cm} {\rm Help} \lor \hspace{0.2cm} \textcircled{O} \hspace{0.2cm} {\rm Download}$	Q Search
		Advanced Search	
	Search fo	Structure Search information	
	Cyclic pep	Sequence Search	Q
	Example: I receptor	Name :Aureobasidin E ; Family :Frog Skin Active Peptide ; Source :Conus ; Function :Anti-Bacterial ; Target	:Somatostatin

i. Local alignment

The Smith-Waterman algorithm is utilized for the **Local alignment** of peptide sequences (Biopython: pairwise2.align.localxx, https://biopython.org/). This method allows for parameter adjustment to set penalties for matches, mismatches, and gaps, ensuring tailored alignment based on specific sequence characteristics.

Tip! This method is only applicable to peptides represented by one-letter amino acid codes.

Step 1. Enter or paste your peptide sequence into the panel.

Example: FLPAVIRVAANVLPTAFCAISKKC

Step 2. Adjust the algorithm parameters, for example, match score, mismatch score, open gap score, and extend gap score, to set penalties for matches, mismatches, and gaps; The e-value threshold is used to filter the alignment results.

Step 3. Click the Search button to perform the local alignment.

Local alignment	Graph alignment
Local alignment (Smith-Waterman) essential amino acids single-letter amino acid code linear peptide	
The Smith-Waterman algorithm is utilized for the local alignment of peptide sequences (Biopyth matches, mismatches, and gaps, ensuring tailored alignment based on specific sequence chare single-letter amino acid codes.	In: pairwise2.align.locaixy). This method allows for parameter adjustment to set penalties for teristics. It should be noted that this method is only applicable to peptides represented by Step 2
FLPAVIRVAANVLPTAFCAISKKC Step 1	Parameters (2) match score 1.0
	mismatch score -1.0 open gap score -0.5
	extend gap score -0.1
Example: FLPAVIRVAANVLPTAFCAISKKC	Step 3 Search

The Search result table is presented as a five-column table.

ID: CyclicPepedia ID.

Name: cyclic peptide name.

Sequence: peptide sequence.

Score: the alignment score.

E-value: the expected value. A measure of the significance of the match.

Cyclic peptides			a 🛃	Export as Excel
Show 10 🗸 e	ntries b		C Search:	
ID	Name	Sequence d	Score 💡	E-value 💡
CP03412	e Brevinin-1-Oa12	FLPAVIRVNVLPTAFCAISKKC	21.4	6.684e-6
CP03445	Brevinin-1Jda	FLPAVIRVNVLPTVFCAISKKC	19.4	3.885e-5
CP03387	Brevinin-1Hsa	FLPAVLRVKIVPTVFCAISKKC	14.2	3.773e-3
CP03388	Brevinin-1Hsb (Brevinin-1Jdb)	FLPAVLRVQVVPTVFCAISKKC	14.2	3.773e-3
CP03421	Brevinin-1-Or10	FLPAVLLVATHVLPTVFCAITRKC	12.4	2.006e-2
CP03446	Brevinin-1Jdc	FLPAVLRVKVVPTVFCLISKKC	12.2	2.193e-2
CP03386	Brevinin-1Chb	FLPVIAGLKVLPKLFCAITKKC	11.4	4.434e-2
CP03454	Brevinin-1Ba	FLPFIAGMKFLPKIFCAISKKC	10.6	8.965e-2
CP03447	Brevinin-1Pb	FLPIIAGIKVFPKIFCAISKKC	10.6	8.965e-2
CP03385	Brevinin-1Cha	FLPIIAGVKVLPKLFCAITKKC	10.2	1.275e-1
Showing 1 to 10) of 16 entries		Previous	1 2 Next

- a. Click the Export as Excel button to download the search results.
- **b.** The number of cyclic peptide entries (i.e., 10, 25, 50, and 100) per page can be adjusted through the drop-down list.
- **c.** Use the **Search** box to filter search results by ID, name, sequence, alignment score, and E-value.
- d. Click the column name to sort the results.
- e. Click the CyclicPepedia ID to enter the corresponding cyclic peptide details page.

ii. Graph alignment

To leverage the cyclization information in cyclic peptide sequences, we developed a graph alignment algorithm based on NetworkX. The graph alignment can convert cyclic peptides into graphical structures and measure the similarity by Graph Isomorphism. The input sequence format can be **IUPAC condensed**, **amino acid chain**, and **graph presentation** (refer to **Sequence format transformation** for sequence examples).

Tip! Users can adjust parameters to reduce search space (e.g., set min AA <= sequence length <= max AA and set a higher value for AA comp threshold).



Step 1. Enter or paste your sequence into the panel.

Example: Cys(1)--Tyr--Trp--Lys--Val--Cys(1)

Step 2. Adjust the algorithm parameters.

min AA: minimum amino acid sequence length.

max AA: maximum amino acid sequence length.

AA comp threshold: threshold for amino acid composition similarity.

Similarity threshold: threshold for filtering alignment results.

Step 3. Click the Search button to perform the graph alignment.

vs(1)TyrTrpLysValCys(1) Step 1	Parameters 👔
	min AA 2
	max AA 9
	AA comp threshold 0.25
	similarity threshold 0.25

The Search result table is presented as a five-column table.

ID: CyclicPepedia ID.

Name: cyclic peptide name.

Sequence: peptide sequence.

Matched: the number of matched graph nodes and edges.

Similarity: graph similarity, is the degree of similarity between nodes or edges in a network.

Cycl	ic peptides	a 🕹 Exp	port as Excel	
Show 10 🗸	entries b	C	Search:	
ID	Name	Sequence	Matched ?	Similarity ?
CP00468	H-D-Phe-Cys(1)-Tyr-Trp-Lys-Val-Cys(1)-Trp-Nh2	H-D-PheCys(1)TyrTrpLysValCys(1)- -Trp-NH2	12	0.75
CP01334	Cid 45588099	H-D-PheCys(1)TyrTrpLysValCys(1)- -D-Trp-NH2	12	0.75
CP00576	Somatuline Depot	H-D-2NalCys(1)TyrD-TrpLysValCys (1)Thr-NH2	9	0.562

- **a.** Click the **Export as Excel** button to download the search results.
- **b.** The number of cyclic peptide entries (i.e., 10, 25, 50, and 100) per page can be adjusted through the drop-down list.
- **c.** Use the **Search** box to filter search results by ID, name, sequence, matched, and similarity.
- d. Click the column name to sort the results.
- e. Click the CyclicPepedia ID to enter the corresponding cyclic peptide details page.

Cyclic peptide tools

1) Structure to sequence

Click the **Tools**|Structure to Sequence to enter the Structure-to-Sequence conversion page.

Cycli	cPepedia Knowledge Base	Visitor Visitor
$\textcircled{$\widehat{\]}$Home} \textcircled{$\widehat{\]}$ Browse \lor Q Search \lor ${${${${${${$}{$}$}}}{${${${${$}$}}$ Tools $\lor $$\underline{${${${$}$}}$ Statistics}}}$	E DataSource ⑦ Help - ۞ Download	Q Search
Structure to Sequence		
Search for cyc Sequence to Structure	'n	
Peptide Property Prediction		0
Cyclic peptide se Structure format transformation	n	Q
Example: Name Sequence format transformation	g Skin Active Peptide; Source:Conus; Function:Anti-Bacterial;	Target
:Somatostatin receptor		

Structure-to-Sequence (Struc2Seq) converter is a computing process based on RDKit (http://www.rdkit.org) and the characteristics of cyclic peptide sequences. It can extract amino acid units from the cyclic peptide skeleton and match them with the monomer reference library, thereby transforming cyclic peptide SMILES into sequence information. This process mainly relies on the completeness of the monomer reference library. You can access our default monomer reference library on our website. The details of Struc2Seq are available at dfwlab/cyclicpepedia on GitHub.



convert cyclic peptide SMILES into sequence information. This process mainly relies on the completeness of the monomer reference library. You can access our default monomer reference library through the download link. The details of Struc2Seq are available at dfwlab/cyclicpepedia on GitHub.

Step 1. Enter your SMILES into the text box.

Example:

CC[C@H](C)[C@H]1C(=O)NCC(=O)N[C@H]2C[S@@](=O)C3=C(C[C@@H](C(=O)NCC(=O)N1)NC(=O)[C@@H](NC(=O)[C@@H]4C[C@H](CN4C(=O)[C@@H](NC2=O)CC(=O)N)O)[C@@H](C)[C@H](CO)O) C5=C(N3)C=C(C=C5)O Step 2. Click the Transform button to perform Struct2Seq conversion.



A detailed Struc2Seq report is provided. It contains the results of each step of Struc2Seq.

a. Check the accuracy of the SMILES.

Load SMILES :

```
\begin{aligned} \text{SMILES}: \text{CC}[\text{C}@\text{H}](\text{C})[\text{C}@\text{H}]1\text{C}(=\text{O})\text{NCC}(=\text{O})\text{N}[\text{C}@\text{H}]2\text{C}[\text{S}@@](=\text{O})\text{C}3=\text{C}(\text{C}[\text{C}@\text{Q}\text{H}](\text{C}(=\text{O})\text{NCC}(=\text{O})\text{N})\text{NC}(=\text{O})[\text{C}@\text{Q}\text{H}](\text{NC}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H})](\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}\text{H}))(\text{C}(=\text{O})(\text{C}@\text{Q}))(\text{C}(=\text{O})(\text{C}@\text{Q}))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=\text{O})((\text{C}@\text{Q})))(\text{C}(=(\text{O})((\text{C})))((\text{C}@\text{Q})))
```

SMILES is corrected!

b. Convert SMILES into atomic structure, identify cyclic peptide skeleton, and renumber atoms.



c. Identify amino acid units.



d. Extract individual amino acid units.



e. Map the atomic structures of amino acid units to the monomer reference library.

> Chain 1 :
Amino acid sequence : Gly(1)--Cys(2)--Asn--4OH-Pro--lle--Trp(2)--Gly--lle(1)



2) Sequence to structure

Click the **Tools**|Sequence to Structure to enter the Sequence-to-Structure conversion page.

Cycl	cPepedia Knowledge Base	Visitor Visitor
$\textcircled{$\square$} \ \ \ \ \ \ \ \ \ \ \ \ \ $	🖯 DataSource 🗇 Help 🗸 🗘 Download	Q Search
Structure to Sequence		
Search for cyc Sequence to Structure	n	
Peptide Property Prediction		0
Cyclic peptide se Structure format transformati	n	Q
Example: Name Sequence format transformat :Somatostatin receptor	on g Skin Active Peptide ; Source :Conus ; Function :Anti-Bacterial ;	Target

Sequence-to-Structure (Seq2Struc) is a computing process based on RDKit. It can create cyclic peptide sequences and convert sequences to structural information. The details of Seq2Struc are available at **dfwlab/cyclicpepedia** on GitHub.

Tip! Seq2Struc posits a **head-to-tail** cyclization, which may not always be the case for all cyclic peptides. Thus, we provide an **online editing interface** whereby users can refine predicted structures using additional structural information they possess.



i. Seq2Struc for essential amino acids

Step 1. Enter your sequence into the text box. This tool accepts amino acid sequences with **one-letter code** and **three-letter code**.

Example: Ala--Ala--Cys--Asp

Step 2. Select the Cyclic parameter.

Step 3. Click the Transform button to perform the Seq2Struc transformation.



The transformation result will be presented. Click the **Open with editor (a)** button to enter the **online editing interface**. The result can be downloaded by clicking the **Download report (b)** button.

	a Open with editor Download report b
Transform completed!	
Your input sequence : Ala-Ala-Cys-Asp	
SMILES : C[C@@H]1NC(=O)[C@H](C)NC(=O)[C@H](CC(=O)O)NC(=O)[C@H](CS)NC1=O	

Users can refine the predicted structure by using the **online editing interface**.

Tip! After editing the structure, remember to click the Save changes (a) button, otherwise it will not be updated in the final result.



ii. Seq2Struc for non-essential amino acids

The process of **Seq2Struc for non-essential amino acids** is similar to that of **Seq2Struc for essential amino acids**. It provides an additional monomer library for users to choose from.

Click the Select from library (a) button to enter the Monomer library.



Over 500 amino acid structural units are provided in our monomer library. Click the **monomer name (a)** to show its **SMILES** and **atomic structure (b)**. Add the monomer to your sequence by clicking the **Add this monomer to peptide (c)** button. The selected monomers are shown in the **sequence box (d)**. Click the **Create peptide (e)** button to generate a new peptide sequence for **Seq2Struc** transformation.

Monomer library

Select monomer from table :

72	Chrl	isopyoverdin chromophore	['Chromophores', 'Chromophores']	Cys SMILES :NC(CS)C(=0)0
73	ChrD	5,6-dihydropyoverdin chromophore	['Chromophores', 'Chromophores']	
74	Cit	Citrulline	['Cit*', 'Cit*']	Q
75	D-Cit	D-Citrulline	['Cit*', 'Cit*']	SH
76	СМА	coronamic acid	['Cma*', 'Cma*']	
77	norCMA	norcoronamic acid	['Cma*', 'Cma*']	
78	NMe-Cys	N-methylcysteine	['Cys*', 'Cys*']	
79 79	Cys	Cysteine	['Cys*', 'Cys*']	с
80	diMe-Cys	N,S-dimethylcysteine	['Cys*', 'Cys*']	Add this monomer to peptide
Seque Aha	nce: ad ×	d		
				e Create peptide

The subsequent steps are the same as Seq2Struc for essential amino acids.

3) Peptide Property Prediction

Click the **Tools**|**Peptide Property Prediction** to enter the **Peptide Property Prediction** page.



i. Chemical and Physical Property Prediction (CPPP)

Chemical and physical properties are computed using RDKit, involving topological polar surface area, complexity, Log(P), hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, drug-likeness, and fingerprints. The details of CPPP are available at **dfwlab/cyclicpepedia** on GitHub.

Step 1. Enter your SMILES into the text box.

Example:

CC[C@H](C)[C@H]1C(=O)NCC(=O)N[C@H]2C[S@@](=O)C3=C(C[C@@H](C(=O)NCC(=O)N1)NC(=O)[C@@H](NC(=O)[C@@H]4C[C@H](CN4C(=O)[C@@H](NC2=O)CC(=O)N)O)[C@@H](C)[C@H](CO)O) C5=C(N3)C=C(C=C5)O

Step 2. Click the Predict button to perform CPPP.

Chemical ar	nd Physical Property Prediction		Peptide Sequence Property F	Prediction
Chemical and Phy	sical Property Predicti	on		
SMILES				
Chemical and Physical Proper Topological Polar Surface Ar and Fingerprints. The details	ty Prediction (CPPP) is an algorithm ea, Complexity, Log(P), Hydrogen I of CPPP are available at dfwlab/cycl	based on RDKit to pred 3ond Donor Count, Hy icpepedia on GitHub.	dict chemical and physical properties of cyclic drogen Bond Acceptor Count, Rotatable Bor	: peptides, such as nd Count, Drug likeness,
Input your SMILES Ste	ep 1			
CC[C@H](C)[C@H]10 (NC(=O)[C@@H]4C[0 (CO)O)C5=C(N3)C=0	C(=O)NCC(=O)N[C@H]2C C@H](CN4C(=O)[C@@H] C(C=C5)O	[S@@] (=0)C3=C [NC2=0)CC(=0)	:(C[C@@H](C(=O)NCC(=O)N1)N(N)O)[C@@H](C)[C@H]	C(=O)[C@@H]
Example: Ctopa(CP00005)	alpha-Amanitine(CP01656) Gramic	idin S(CP00038)		Step 2 Predict

A **downloadable report** will be provided. Please see **Table 1** for a complete list of

properties.

Chemical and Physical Property Prediction

Property	Value
Molecule	
Number of Atoms	64
Number of Rings	5
Exact Mass	918.3541674080002
Topological Polar Surface Area 📀	380.8800000000005
Complexity ?	0.875
Crippen Log(P) ?	-5.9202000000009
Heavy Atom Count ?	64

Click the **Download report** button to download the prediction results.

MACCS Keys 🕢	00000000000000000000000000000000000000	
		Download report

ii. Peptide Sequence Property Prediction (PSPP)

Peptide sequence properties are predicted by the "Peptides" package of R

(https://github.com/dosorio/Peptides/). It predicts more than 100 indices, such as the Boman index, charge, aliphatic index, instability index, and amino acid composition. The details of PSPP are available at dfwlab/cyclicpepedia on GitHub.

Step 1. Enter your sequences into the text box. This tool can only be used for peptide sequences in one-letter amino acid code.

Example: ATPTTT

Step 2. Click the Predict button to perform PSPP.

Tip! Peptide Sequence Property Prediction may take some time.

Chemical and Physical Property Prediction	Peptide Sequence Property Prediction
Peptide Sequence Property Prediction	
essential amino acids one-letter amino acid code	
Peptide Sequence Properties are predicted by the "Peptides" package of R (Osorio, D., Kondo antimicrobial peptides. The R Journal. 7(1), 4–14 (2015).). It predicts more than 100 indices, su and Amino acid composition. The details of sequence property prediction are available at dfw	n-Villarreal, P. & Iorres, R. Peptides: A package for data mining of Jch as the Boman index, Charge, Aliphatic index, Instability inde vlab/cyclicpepedia on GitHub.
Peptide Sequence Property Prediction may take some time.	
nput your sequence Step 1	
ΔΤΡΤΤΤ	

A **downloadable report** will be provided. Please see <u>Table 2</u> for a complete list of properties.

Peptide Sequence Properties					
Property	Value	Value			
Sequence	ATPTTT	ATPTTT			
Length	6	6			
Boman index	1.411666	1.41166666666666664			
Charge ?	-0.0020	-0.0020157006072527572			
Aliphatic index ?	16.666666666666664				
Instability index 🕐	8.333333333333334				
Amino acid		PROPERTY	RESIDUES	NUMBER	MOLE%
composition	0	Tiny	(A+C+G+S+T)	5.0	83.333
	1	Small	(A+B+C+D+G+N+P+S+T+V)	6.0	100.000
	2	Aliphatic	(A+I+L+V)	1.0	16.667
	3	Aromatic	(F+H+W+Y)	0.0	0.000



Click the **Download report** button to download the prediction results.

4) Structure format transformation

Click the **Tools**|**Structure format transformation** to enter the **Structure format transformation** page.

		Cyclic	Pepedia Knowledge Base	Visitor Visitor
☆ Home	$\mathbb{Q}_{=}^{=}$ Browse \sim \mathbb{Q} Search \sim	⊁ Tools -> 🕑 Statistics	🖯 DataSource 🕜 Help 🗸 🗘 Download	Q Search
		Structure to Sequence		
	Search for cyc	Sequence to Structure	n	
		Peptide Property Prediction		0
	Cyclic peptide se	Structure format transformation		Q
	Example: Name :Somatostatin rec	Sequence format transformation	g Skin Active Peptide ; Source :Conus ; Function :Anti-Bacterial ;	Target

Structure format transformation is a computing process based on RDKit to transform molecules between SMILES, InChI, InChIKey, Mol block, and PDB block formats. The details of this algorithm are available at **dfwlab/cyclicpepedia** on GitHub.

Step 1. Enter or paste your structural information into the text box. This tool accepts peptide structural data in **SMILES**, **InChI**, **InChIKey**, **Mol block**, and **PDB block** formats.

Example: InChI=1S/C50H67N11O11S2/c1-26(62)39(42(53)65)59-49(72)41-50(3,4)74-73-25-38(58-43(66)33(52)21-28-11-6-5-7-12-28)47(70)56-36(22-29-16-18-31(64)19-17-29)45(68)57-37(23-30-24-54-34-14-9-8-13-32(30)34)46(69)55-35(15-10-20-51)44(67)60-40(27(2)63)48(71)61-41/h5-9,11-14,16-19,24,26-27,33,35-41,54,62-64H,10,15,20-23,25,51-52H2,1-4H3,(H2,53,65)(H,55,69)(H,56,70)(H,57,68)(H,58,66)(H,59,72)(H,60,67)(H,61,71)

Step 2. Select the parameters.

3D conformation: Whether to generate a 3D conformation. The 3D conformation is generated by minimizing the energy of Universal Force Field (UFF), which is suitable for small molecule.

Optimize: Whether to perform an optimization.

Tip! It will take some time to create 3D conformations. The generation of PDB block requires additional stereochemical information. The larger the molecule, the longer the optimization time.

Step 3. Click the Transform button to perform the structure format transformation.

Structure format transform molecules between SMILES, Incl	nl, InchIKey, Mol block, ai	nd PDB bloc	ck
ormats. The details of this algorithm are available at dfwlab/cyclicpepedia on GitHub.			
It will take some time for creating 3D conformations. The generation of PDB block requires additional stereochemic the longer the optimization time.	al information. The large	r the molect	ule,
nput your structure Step 1	Step 2		
InChI=1S/C50H67N11011S2/c1-26(62)39(42(53)65)59-49(72)41-50(3,4)74-73-25-	Parameters ?		
38(58-43(66)33(52)21-28-11-6-5-7-12-28)47(70)56-36(22-29-16-18-31(64)19-17-	3D conformation	Yes	~
29)45(68)57-37(23-30-24-54-34-14-9-8-13-32(30)34)46(69)55-35(15-10-20-			
51)44(67)60-40(27(2)63)48(71)61-41/h5-9,11-14,16-19,24,26-27,33,35-41,54,62-	Optimize	Yes	\sim
64H,10,15,20-23,25,51-52H2,1-4H3,(H2,53,65)(H,55,69)(H,56,70)(H,57,68)(H,58,66)			
(H,59,72)(H,60,67)(H,61,71)			
		0.4	

The transformation results will be presented on the webpage. The input structural information will be converted into multiple structure formats, such as **SVG**, **SMILES**, **InChI**, **InChIKey**, **Mol Block**, **PDB Block**, and **3D conformation** (if selected). And formats with a **download** button can be downloaded separately.



Click the **Download report** button to download the complete report.



5) Sequence format transformation

Click the **Tools**|Sequence format transformation to enter the Sequence format transformation page.

	CyclicPepedia Knowledge Base	Visitor Visitor
☆ Home = Browse → Q Search → ★ Tools	\checkmark $\underline{\mathbb{V}}$ Statistics \bigcirc DataSource \bigcirc Help \checkmark \bigcirc Download	Q Search
Structure	o Sequence	
Search for cyc Sequence	to Structure In	
Peptide P	operty Prediction	0
Cyclic peptide se Structure	ormat transformation	Q
Example: Name Sequence	format transformation g Skin Active Peptide ; Source :Conus ; Function :Anti-Bacterial	; Target
:Somatostatin receptor		

Sequence formats such as one-letter code, IUPAC condensed, amino acid chain, graph representation, and sequence graph formats can be inter-converted through the **Sequence format transformation** tool. A description of these formats is available online (a) and in **Table 3**. The details of this algorithm are available at **dfwlab/cyclicpepedia** on GitHub.

Step 1. Paste or enter your sequence in one-letter code, IUPAC condensed, amino acid chain, graph representation, or sequence graph formats.

Example: cyclo[DL-Cys(1)-DL-Cys-DL-Val(1)-DL-Leu-DL-xiIle]

Step 2. Click the Transform button to perform the Sequence format transformation.

Sequence format transformation

essential amino acids non-essential amino acids one-letter amino acid code three-letter amino acid code

Sequence format transformation is a computing process to transform peptide sequences between one-letter codes, IUPAC condensed, amino acid chain, graph representation, and sequence graph formats. The details of this algorithm are available at dfwlab/cyclicpepedia on GitHub.

Format	Example	Detail
One letter code	FGIKPPQR	The simplest representation of peptide sequences, ignoring all loops.
IUPAC condensed	cyclo[DL-N(Me)Ala-DL-Leu-N(Me)Phe(a,b- dehydro)-Gly]	Developed by the International Union of Pure and Applied Chemistry (IUPAC). The prefix"Cyclo" indicates a head-to-tail cyclization. The sequence of amino acids is represented by standard three-letter codes, separated by '-'. Modifications to the amino acids are indicated in the sequence, such as "D" and "L" refer to the chirality of the amino acid, and ring closure bonds are represented by "(num)". It can represent multiple chains through separator '.' , for example: D-N(1)Ala-Arg(CONHMe)-N(Me)Phe-Asp(2)-OH.N(2)Asp(1)-OH.
Amino acid chain	Gly(1)Cys(2)Asn4OH-Prolle Trp(2)Glylle(1)	Define by CyclicPepdia, basically consistent with IUPAC condensed. The separator changes to "" to adapt to situations where the amino acid unit (monomer) has a "-"
Graph representation	aThr,Tyr,dhAbu,bOH- Gln,Gly,Gln,His,Dab,C13:2(t4.t6)- OH(2.3),Lyx,dhAbu @1,5 @6,10 @0,8	Inspired by the NOR format from the Norine database, monomers are divided by comma, and ring closure bonds are represented by '@idx,idy.'
Sequence graph	G(nodes=[]; edges=[])	The sequence graph is built by networkx through a list of nodes and edges. For example: nodes = [(0, 'Gly'), (1, '4OH-Pro'), (2, 'Ala')], edges = [(0, 1), (1, 2), (0, 2)]
Input your seque cyclo[DL-Cys	ence Step 1 (1)-DL-Cys-DL-Val(1)-DL-Leu-DL->	xille]
		ĥ
Example: FGIKPPQ -AsnProIleTrp	R cyclo[DL-Cys(1)-DL-Cys-DL-Val(1)-DL-Leu- p(2)GlyIle(1) aThr,Tyr,dhAbu,bOH-Gln,Gly,	-DL-xille] cyclo[DL-N(Me)Ala-DL-Leu-N(Me)Phe(a,b-dehydro)-Gly] Gly(1)Cys(2)- Step 2 Transform Gln,His,Dab,C13:2(t4.t6)-OH(2.3),Lyx,dhAbu @1,5 @6,10 @0,8

The transformation results will be presented on the webpage. The input sequence information will be converted into multiple sequence formats. And formats with a **download** button can be downloaded separately.

Input	Value IUPAC condensed
One letter peptide	None
IUPAC condensed	cyclo[DL-Cys(1)-DL-Cys-DL-Val(1)-DL-Leu-DL-xille]
Amino acid chain	DL-Cys(1)(2)DL-CysDL-Val(1)DL-LeuDL-xille(2)
Graph presentation	DL-Cys,DL-Val,DL-Leu,DL-xille @0,2 @0,4
	DL-Cys DL-Cys DL-xille

Click the **Download report** button to download the complete report.

Table 1. List of the chemical and physical properties.

Property	Description
Number of Atoms	
Number of Rings	
Exact Mass	
Topological Polar Surface Area	Measures the surface area occupied by polar atoms, often used to predict drug transport properties.
Complexity	Indicates the structural complexity of the molecule, with higher values representing more intricate structures.
Crippen Log(P)	Represents the logarithm of the partition coefficient between n-octanol and water, used to estimate the molecule's hydrophobicity.
Heavy Atom Count	Counts the number of non-hydrogen atoms in the molecule, reflecting its size and complexity.
Hydrogen Bond Donor Count	The number of atoms in the molecule that can donate hydrogen bonds, important for molecular interactions.
Hydrogen Bond Acceptor Count	The number of atoms capable of accepting hydrogen bonds, crucial for molecular recognition and binding.
Rotatable Bond Count	Counts the number of bonds that allow free rotation around themselves, affecting the molecule's flexibility.
Formal Charge	The overall electric charge of the molecule, with zero indicating a neutral molecule.
Refractivity	Measures the molecule's ability to refract light, related to polarizability and electronic properties.
Rule of Five	Indicates non-compliance with Lipinski's rule of five, suggesting potential issues with bioavailability as an oral drug.
Veber's Rule	Shows non-adherence to Veber's rules, potentially impacting oral bioavailability and permeability.
Ghose Filter	A molecular property filter used to assess the drug-likeness of a compound based on its physicochemical properties.
RDKit Fingerprint	This is an RDKit-specific fingerprint.
Daylight-like Fingerprint	This is an RDKit-specific fingerprint that is inspired by (though it differs significantly from) public descriptions of the Daylight fingerprint.
Morgan Fingerprint	The RDKit implementation uses the feature types Donor, Acceptor, Aromatic, Halogen, Basic, and Acidic.
MACCS Keys	SMARTS definitions for the publicly available MACCS keys and a MACCS fingerprinter.

Table 2. List of the peptide sequence properties.

Property	Description	
Sequence		
Length	Sequence length	
Boman index	This property computes the potential protein interaction index proposed by Boman (2003) based in the amino acid sequence of a protein. The index is equal to the sum of the solubility values for all residues in a sequence, it might give an overall estimate of the potential of a peptide to bind to membranes or other proteins as receptors, to normalize it is divided by the number of residues. A protein have high binding potential if the index value is higher than 2.48.	
Charge	This property computes the net charge of a protein sequence based on the Henderson-Hasselbalch equation described by Moore, D. S. (1985). The net charge can be calculated at defined pH using one of the 9 pKa scales availables: Bjellqvist, Dawson, EMBOSS, Lehninger, Murray, Rodwell, Sillero, Solomon or Stryer.	
Aliphatic index	This property calculates the Ikai (1980) aliphatic index of a protein. The aindex is defined as the relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins.	
Instability index	This property calculates the instability index proposed by Guruprasad (1990). This index predicts the stability of a protein based on its amino acid composition, a protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable.	
Amino acid composition	Tiny, Small, Aliphatic, Aromatic, Non-polar, Polar, Charged, Basic and Acidic based on their size and R-groups using same function implemented in EMBOSS 'pepstat'.	
BLOSUM62	BLOSUM indices were derived of physicochemical properties that have been subjected to a VARIMAX analyses and an alignment matrix of the 20 natural AAs using the BLOSUM62 matrix.	
Cruciani properties	The Cruciani properties of an amino-acids sequence is calculated using the scaled principal component scores that summarize a broad set of descriptors calculated based on the interaction of each amino acid residue with several chemical groups (or 'probes'), such as charged ions, methyl, hydroxyl groups, and so forth.	
FASGAI vectors	The FASGAI vectors (Factor Analysis Scales of Generalized Amino Acid Information) is a set of amino acid descriptors, that reflects hydrophobicity, alpha and turn propensities, bulky properties, compositional characteristics, local flexibility, and electronic properties, that can be utilized to represent the sequence structural features of peptides or protein motifs.	
Hydrophobic moment	This properties compute the hmoment based on Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984). Hydriphobic moment is a quantitative measure of the amphiphilicity perpendicular to the axis of any periodic peptide structure, such as the a-helix or b-sheet. It can be calculated for an amino acid sequence of N residues and their associated hydrophobicities Hn.	
Hydrophobicity index	This property calculates the GRAVY hydrophobicity index of an amino acids sequence using one of the 38 scales from different sources.	
Kidera factors	The Kidera Factors were originally derived by applying multivariate analysis to 188 physical properties of the 20 amino acids and using dimension reduction techniques. This function calculates the average of the ten Kidera factors for a protein sequence.	

Theoretical class	This property calculates the theoretical class of a protein sequence based on the relationship between the hydrophobic moment and hydrophobicity scale proposed by Eisenberg (1984).	
MS-WHIM scores	MS-WHIM scores were derived from 36 electrostatic potential properties derived from the three-dimensional structure of the 20 natural amino acids.	
Isoelectic point (pI)	The isoelectric point (pI), is the pH at which a particular molecule or surface carries no net electrical charge.	
protFP	The ProtFP descriptor set was constructed from a large initial selection of indices obtained from the AAindex database for all 20 naturally occurring amino acids.	
ST-scales	ST-scales were proposed by Yang et al, taking 827 properties into account which are mainly constitutional, topological, geometrical, hydrophobic, electronic, and steric properties of a total set of 167 AAs.	
T-scales	T-scales are based on 67 common topological descriptors of 135 amino acids. These topological descriptors are based on the connectivity table of amino acids alone, and to not explicitly consider 3D properties of each structure.	
VHSE-scales	VHSE-scales (principal components score Vectors of Hydrophobic, Steric, and Electronic properties), is derived from principal components analysis (PCA) on independent families of 18 hydrophobic properties, 17 steric properties, and 15 electronic properties, respectively, which are included in total 50 physicochemical variables of 20 coded amino acids.	
Z-scales	Z-scales are based on physicochemical properties of the AAs including NMR data and thin-layer chromatography (TLC) data.	
Amino acid count		

Format	Example	Description	
One letter code	FGIKPPQR	The simplest representation of peptide sequences, ignoring all loops.	
IUPAC condensed	cyclo[DL-N(Me)Ala-DL- Leu-N(Me)Phe(a,b- dehydro)-Gly]	Developed by the International Union of Pure and Applied Chemistry (IUPAC). The prefix"Cyclo" indicates a head-to-tail cyclization. The sequence of amino acids is represented by standard three-letter codes, separated by '-'. Modifications to the amino acids are indicated in the sequence, such as "D" and "L" refer to the chirality of the amino acid, and ring closure bonds are represented by "(num)". It can represent multiple chains through separator '.' , for example: D-N(1)Ala-Arg(CONHMe)- N(Me)Phe-Asp(2)-OH.N(2)Asp(1)-OH.	
Amino acid chain	Gly(1)Cys(2)Asn4OH- ProIleTrp(2)GlyIle(1)	Define by CyclicPepdia, basically consistent with IUPAC condensed. The separator changes to "" to adapt to situations where the amino acid unit (monomer) has a "-"	
Graph representation	aThr,Tyr,dhAbu,bOH- Gln,Gly,Gln,His,Dab,C13:2 (t4.t6)-OH(2.3),Lyx,dhAbu @1,5 @6,10 @0,8	Inspired by the NOR format from the Norine database, monomers are divided by comma, and ring closure bonds are represented by '@idx,idy.'	
Sequence graph	G(nodes=[]; edges=[])	The sequence graph is built by Networkx through a list of nodes and edges. For example:nodes = [(0, 'Gly'), (1, '4OH-Pro'), (2, 'Ala')], edges = [(0, 1), (1, 2), (0, 2)]	

Table 3. Sequence formats.