

钱欣博士

cqian@acs-i.org

美国化学文摘社北京代表处

SciFinderⁿ专题培训

---生物序列信息检索



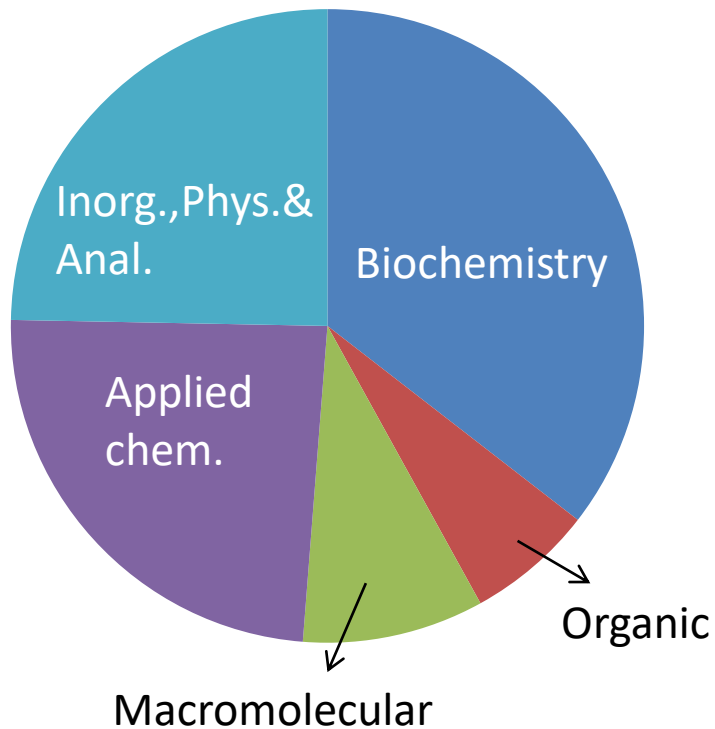
CAS[®]

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

介绍:

- SciFinder-n中生物信息覆盖
- 生物序列检索方法
 - 物质识别符检索
 - 结构检索
 - 关键词检索
 - Biosequence检索

SciFinder-n提供全面的“化学”信息——源于化学，超越化学



生物化学

有机化学各领域

大分子化学各领域

应用化学各领域

物理、无机、分析化学各领域



CAS[®]

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

SciFinder-n覆盖全面的生物学领域

- Biochemical methods
- Biochemistry
- Bioinformatics
- Biophysics
- Cellular biology
- Drug metabolism
- Enzymology
- Genetics
- Microbiology
- Molecular biology
- Pharmacogenetics
- Pharmacokinetics
- Pharmacology
- Proteomics
- Toxicology
- Amino Acids, Peptides, and Proteins
- ...



SciFinder-n覆盖生物学家每日阅读的生物医学领域的出版物

Molecular Biology and Genetics

- *Annual Review of Genetics*
- *Cell*
- *Developmental Cell*
- *Genome Research*
- *Journal of Cell Biology*
- *Molecular Genetics and Genomics*
- *Nature*
- *New England Journal of Medicine*
- *Proceedings of the National Academy of Sciences*
- *Science*

Biochemistry ACS Chemical Biology

- *ACS Synthetic Biology*
- *Annual Review of Biochemistry*
- *Biochemistry and Cell Biology*
- *Cellular Physiology and Biochemistry*
- *Journal of Biological Chemistry*
- *Journal of Cellular Biochemistry*
- *Molecular and Cellular Biochemistry*
- *Preparative Biochemistry and Biotechnology*

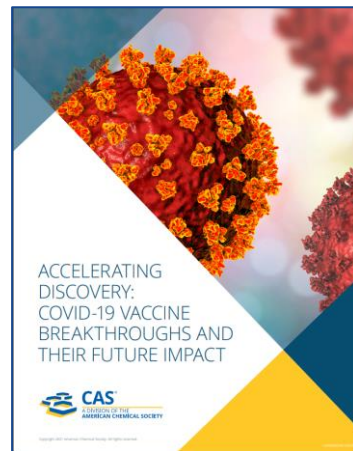
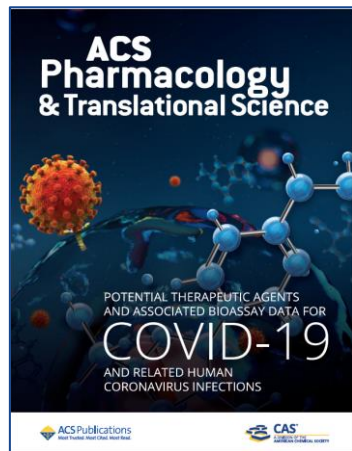
Pharmaceuticals and Medicinal Chemistry

- *Advanced Drug Delivery Reviews*
- *Annual Review of Pathology: Mechanisms of Disease*
- *Anti-Inflammatory Anti-Allergy Agents in Medicinal Chemistry*
- *Circulation Research*
- *Immunity*
- *Journal of the American Medical Association*
- *Journal of Experimental Medicine*
- *Nature Reviews Drug Discovery*
- *Trends in Immunology*

全球科技趋势报告系列——抗体药物、疫苗、融合蛋白类药物和基因与细胞治疗



2020年CAS美国化学文摘社发布关于新冠治疗研究分析报告



介绍:

- 生物序列检索方法
 - 物质识别符检索
 - 结构检索
 - 关键词检索
 - Biosequence检索

如何在SciFinderⁿ 进行序列检索? ——名称或代码检索

- Chemical name: e.g. [INSULIN](#)
- Brand name (drug name): e.g. [Argireline](#)
- Generic name: e.g. [Somavubove](#)
- CAS Registry Number: e.g. [1190939-33-3](#)
- GenBank Number: e.g. [GenBank KT001084](#)
- EC number: e.g. [EC 3.2.1.83](#)

Search

All

Substances

Reactions

References

Suppliers

Search by Substance Name, CAS RN, Patent Number, etc.

Use [Advanced Search](#) for Molecular Formula, Substance Property, or Experimental Spectra

通过名称、编码等识别符进行物质检索的结果

SciFinderⁿ
A CAS SOLUTION

Substances ▾ insulin

← Return to Home

Filter by

- Commercial Availability
 - Available (1)
- Reaction Role
 - Product (1)
 - Reactant (1)
 - Reagent (1)
 - Catalyst (1)
- Reference Role
 - Adverse Effect (1)

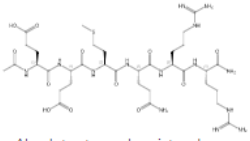
Substances (1)

References ▾ Reactions ▾

9004-10-8
View Detail
Image Not Available
Unspecified
Insulin

256K References 211 Reactions 6 Suppliers

616204-22-9
View Detail



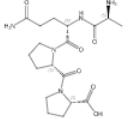
Absolute stereochemistry shown

C₃₄H₆₀N₁₄O₁₂S
Argireline

Protein/Peptide Sequence
Sequence Length: 6

568 References 99 Reactions 48 Suppliers

1190939-33-3
View Detail



Absolute stereochemistry shown

C₁₈H₂₉N₅O₆
L-Proline, L-alanyl-L-glutaminy-L-prolyl-
Protein/Peptide Sequence
Sequence Length: 4

2 References 0 Reactions 0 Suppliers

126752-39-4
View Detail
Image Not Available
Unspecified
Somavubove

Protein/Peptide Sequence
Sequence Length: 191

13 References 0 Reactions 0 Suppliers

2092642-35-6
View Detail
Image Not Available
Unspecified
GenBank KT001084

Nucleic Acid Sequence
Sequence Length: 22089

1 Reference 0 Reactions 0 Suppliers

37288-59-8
View Detail
Image Not Available
Unspecified
EC 3.2.1.83

126 References 1 Reaction 0 Suppliers

如何在SciFinderⁿ 进行序列检索？——文献检索，再由文献获得序列信息

The screenshot shows the SciFinder interface for a search on "Argireline". The top navigation bar includes the SciFinder logo, a search bar with "Argireline" entered, and icons for "Draw", search, and "star". Below the search bar, there are tabs for "References" and "Substances". The "References" tab is active, showing a list of search results. A yellow box highlights the "Substances" filter button in the left sidebar. The main content area displays a list of references, with the first one selected. The abstract of the selected reference is visible, mentioning "Argireline is a synthetic peptide that is patterned from...".

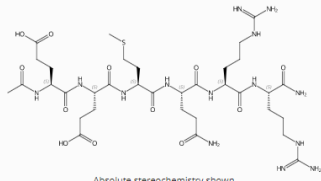
The screenshot shows the SciFinder interface for a search on "Argireline". The top navigation bar includes the SciFinder logo, a search bar with "Argireline" entered, and icons for "Draw", search, and "star". Below the search bar, there are tabs for "References" and "Substances". The "Substances" tab is active, showing a list of search results. A yellow box highlights the "Substances" filter button in the left sidebar. The main content area displays a grid of substance results, each with a chemical structure and a "View Detail" button. A yellow box highlights the "Protein/Peptide Sequence" filter option in the left sidebar. A yellow text box with the Chinese text "可以直接将物质结果集限定为蛋白质/多肽序列" (You can directly limit the substance result set to protein/polypeptide sequences) is overlaid on the substance results.

点击CAS RN获取序列信息详情

Substance Detail (1 of 2,621)

References (568) Reactions (99) Suppliers (48)

CAS Registry Number
616204-22-9



$C_{34}H_{60}N_{14}O_{12}S$
L-Argininamide, N-acetyl-L- α -glutamyl-L- α -glutamyl-L-methionyl-L-glutamyl-L-arginyl-

Key Physical Properties	Value	Condition
Molecular Weight	888.99	-
Density (Predicted)	1.54±0.1 g/cm ³	Temp: 20 °C; Press: 760 Torr
pKa (Predicted)	4.43±0.10	Most Acidic Temp: 25 °C

Protein/Peptide Sequence
Sequence Length: 6 modified

Related Sequences (85)

Other Names

Sequence Details

Sequence: linear

1	EEMQRR	-	-	-	-
---	--------	---	---	---	---

Sequence Modifications

Type	Location	Description
terminal mod.	glutamic acid-1	N-acetyl
terminal mod.	arginine-6	C-terminal amide

Predicted Properties

Predicted Spectra

Target Indicators

Regulatory Information

Additional Details

介绍:

- 生物序列检索方法
 - 物质识别符检索
 - 结构检索
 - 关键词检索
 - Biosequence检索

如何在SciFinderⁿ 进行序列检索? ——结构检索

The screenshot displays the CAS Draw software interface. The main window is titled "CAS Draw" and contains a toolbar with various drawing tools. A search bar at the top right of the main window contains the text "Enter a CAS RN, SMILES or InChI". Below the search bar is a yellow banner that reads "Select and draw structures with templates." The left sidebar contains a vertical list of drawing tools, with a template icon (a square with a diagonal line) highlighted by a yellow box. The right sidebar is titled "Templates" and contains a search bar with the text "Enter 3 or more characters...". Below the search bar is a list of templates under the heading "Amino Acid (25)". The list includes the following amino acids: Alanine, Arginine, Asparagine, Aspartic Acid, Cysteine, Dipeptide backbone, Glutamic Acid, Glutamine, and Glycine. Each amino acid is represented by its chemical structure and name. The bottom of the main window has a status bar with a zoom level of 100% and a zoom slider. At the bottom left of the main window are "OK" and "Cancel" buttons.

可以使用结构编辑器中的氨基酸模板

一次结构检索可以同时获得精确结构/亚结构/相似结构检索

Structure Match

As Drawn (1)

Substructure (585)

Similarity (938K)

Structure Precision

Conventional Results (584)

Filter by

- Commercial Availability
- Reaction Role
- Reference Role
- Stereochemistry
- Absolute Stereo Match (584)
- No Stereo in Answer Structure (1)
- Number of Components
- Substance Class
- Isotopes
- Metals
- Molecular Weight
- Regulatory Information
- Bioactivity Indicator
- Target Indicator
- Search Within Results

Substances (584)

Sort: Relevance View Partial

References Reactions Suppliers

<input type="checkbox"/> 1190939-33-3 View Detail Absolute stereochemistry shown C ₁₈ H ₂₉ N ₅ O ₈ L-Proline, L-alanyl-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 2 References 0 Reactions 0 Suppliers	<input type="checkbox"/> 1351352-05-0 View Detail Absolute stereochemistry shown C ₂₁ H ₃₂ N ₅ O ₈ L-Proline, L-leucyl-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 3 References 1 Reaction 0 Suppliers	<input type="checkbox"/> 1190939-22-0 View Detail Absolute stereochemistry shown C ₂₀ H ₃₂ N ₅ O ₇ L-Proline, L-glutaminy-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 2 References 0 Reactions 0 Suppliers
<input type="checkbox"/> 547751-69-9 View Detail Absolute stereochemistry shown C ₁₈ H ₂₉ N ₅ O ₇ L-Proline, L-seryl-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 1 Reference 0 Reactions 0 Suppliers	<input type="checkbox"/> 1537874-95-5 View Detail Absolute stereochemistry shown C ₂₂ H ₃₃ N ₅ O ₈ L-Proline, glycyl-L-glutaminy-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 5 1 Reference 0 Reactions 0 Suppliers	<input type="checkbox"/> 191922-08-4 View Detail Absolute stereochemistry shown C ₂₄ H ₃₇ N ₅ O ₉ L-Proline, N-acetylglycyl-L-glutaminy-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 5 1 Reference 1 Reaction 0 Suppliers

Structure Match

As Drawn (1)

Substructure (585)

Similarity (938K)

Filter by

- Similarity
- >=99 (1)
- 95-98 (8)
- 90-94 (195)
- 85-89 (4,006)
- 80-84 (37K)

Commercial Availability

Reaction Role

Reference Role

Stereochemistry

Number of Components

Substance Class

Isotopes

Metals

Molecular Weight

Regulatory Information

Bioactivity Indicator

Target Indicator

Search Within Results

Substances (204)

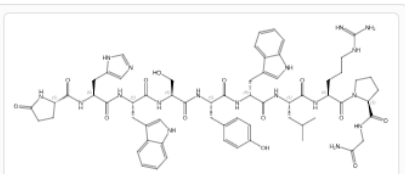
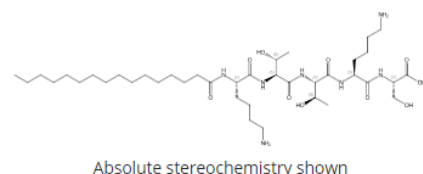
Sort: Relevance View Partial

References Reactions Suppliers

<input type="checkbox"/> 1190939-33-3 View Detail Absolute stereochemistry shown C ₁₈ H ₂₉ N ₅ O ₈ L-Proline, L-alanyl-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 2 References 0 Reactions 0 Suppliers	<input type="checkbox"/> 1351352-05-0 View Detail Absolute stereochemistry shown C ₂₁ H ₃₂ N ₅ O ₈ L-Proline, L-leucyl-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 3 References 1 Reaction 0 Suppliers	<input type="checkbox"/> 956007-88-8 View Detail Absolute stereochemistry shown C ₁₇ H ₂₈ N ₅ O ₈ L-Proline, N-acetyl-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 1 Reference 0 Reactions 0 Suppliers
<input type="checkbox"/> 1190939-22-0 View Detail Absolute stereochemistry shown C ₂₀ H ₃₂ N ₅ O ₇ L-Proline, L-glutaminy-L-glutaminy-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 2 References 0 Reactions 0 Suppliers	<input type="checkbox"/> 1190939-40-2 View Detail Absolute stereochemistry shown C ₁₈ H ₂₉ N ₅ O ₈ L-Proline, L-glutaminy-L-alanyl-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 2 References 0 Reactions 0 Suppliers	<input type="checkbox"/> 1799916-09-8 View Detail Absolute stereochemistry shown C ₂₁ H ₃₂ N ₅ O ₈ L-Proline, L-glutaminy-L-leucyl-L-prolyl- Protein/Peptide Sequence Sequence Length: 4 1 Reference 0 Reactions 0 Suppliers
<input type="checkbox"/> 160790-68-1 View Detail	<input type="checkbox"/> 547751-69-9 View Detail	<input type="checkbox"/> 1537874-47-7 View Detail

相似度选择

可以通过生物适应症/靶点等筛选序列集

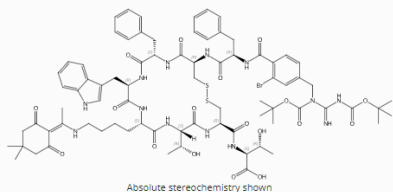
<p>^ Bioactivity Indicator</p> <ul style="list-style-type: none"><input type="checkbox"/> Anti-infective agents (159)<input type="checkbox"/> Dermatological agents (155)<input type="checkbox"/> Receptor antagonists (140)<input type="checkbox"/> Radioprotectants (135)<input type="checkbox"/> Anti-inflammatory agents (115) <p>View All</p> <hr/> <p>^ Target Indicator</p> <ul style="list-style-type: none"><input type="checkbox"/> Ligand-binding proteins (140)<input type="checkbox"/> Enzymes (109)<input type="checkbox"/> RNA formation factors (104)<input type="checkbox"/> Glycoproteins (88)<input type="checkbox"/> Membrane proteins (88) <p>View All</p> <hr/> <p>^ Search Within Results</p>	<p>Absolute stereochemistry shown</p> <p>C₉₁H₁₁₄N₁₂O₁₇S₃ L-Ornithine, <i>N</i>-[(9<i>H</i>-fluoren-9-ylmethoxy)carbonyl]-L-α-glutamyl-L-methionyl-<i>N</i>-(tr... Protein/Peptide Sequence Sequence Length: 5</p> <p><input type="checkbox"/> 1 Reference <input type="checkbox"/> 6 Reactions <input type="checkbox"/> 0 Suppliers</p>	<p>Absolute stereochemistry shown</p> <p>C₈₂H₉₉N₁₁O₁₄S₃ L-Ornithine, <i>N</i>-[(9<i>H</i>-fluoren-9-ylmethoxy)carbonyl]-L-methionyl-<i>N</i>-(triphenylmethyl... Protein/Peptide Sequence Sequence Length: 4</p> <p><input type="checkbox"/> 1 Reference <input type="checkbox"/> 5 Reactions <input type="checkbox"/> 0 Suppliers</p>
	<p><input type="checkbox"/> 140194-24-7 View Detail</p> <p> Absolute stereochemistry shown, Rotation (-)</p> <p><input type="checkbox"/> 214047-00-4 View Detail</p> <p> Absolute stereochemistry shown</p> <p>C₃₉H₇₅N₇O₁₀ Matrixyl Protein/Peptide Sequence</p>	

物质详情

Substance Detail (1 of 1) ← Prev Next →

Reference (1) Reactions (7) Suppliers (0) 📄 📧 ⭐ Save

CAS Registry Number
1000613-79-5



C₇₈H₁₀₀BrN₁₃O₁₈S₂

L-Threonine, N-[2-bromo-4-[[[(1,1-dimethylethoxy)carbonyl][[[(1,1-dimethylethoxy)carbonyl]amino][iminomethyl]amino]methyl]benzyl]amino]-L-threonine, cyclic (2→7)-disulfide

Key Physical Properties	Value	Condition
Molecular Weight	1651.74	-
Density (Predicted)	1.42±0.1 g/cm ³	Temp: 20 °C; Press: 760 Torr
pKa (Predicted)	3.18±0.10	Most Acidic Temp: 25 °C

Spectra

Protein/Peptide Sequence
Sequence Length: 8
modified (modifications unspecified)
[Related Sequences \(676\)](#)

点击Related Sequences可以获得不同修饰的相关序列

Sequence Details 序列信息详情

Sequence: linear

1 FCFWKTCT

Sequence Modifications

Type	Location	Description
bridge	cysteine-2 to cysteine-7	disulfide bridge
modification	phenylalanine-1	undetermined modification
modification	lysine-5	undetermined modification

Sequence Details

Expand All | Collapse All

[← Return to Home](#)

获得676条不同修饰的相关序列

Substances (676) Sort: Relevance ▾ View Partial ▾

[References ▾](#) [Reactions ▾](#) [Suppliers ▾](#) [Download](#) [Email](#) [Save](#)

Filter by

- Commercial Availability
- Reaction Role
- Reference Role
- Stereochemistry
- Number of Components
- Substance Class
- Isotopes
- Metals
- Molecular Weight
- Experimental Property
- Regulatory Information
- Bioactivity Indicator
- Target Indicator
- Search Within Results**

83150-76-9 [View Detail](#)

Absolute stereochemistry shown, Rotation (-)

$C_{49}H_{66}N_{10}O_{10}S_2$
Octreotide
Protein/Peptide Sequence
Sequence Length: 8

References Reactions Suppliers

96443-50-4 [View Detail](#)

Absolute stereochemistry shown

$C_{49}H_{64}N_{10}O_{11}S_2$
Octreotate
Protein/Peptide Sequence
Sequence Length: 8

References Reactions Supplier

99685-66-2 [View Detail](#)

Absolute stereochemistry shown

$C_{49}H_{65}N_{11}O_{10}S_2$
D-Phenylalanyl-L-cysteiny-L-phenylalanyl-D-tryptophyl-L-lysyl-L-threonyl-L-cyst...
Protein/Peptide Sequence
Sequence Length: 8

References Reactions Suppliers

205652-45-5 [View Detail](#)

119643-69-5 [View Detail](#)

328072-88-4 [View Detail](#)

可以用结构/元素等进一步限定结果

MARPAT结果包括可以结构化并可能检索其他可参考的短序列

The screenshot shows the SciFinder interface with the following elements:

- Header: SciFinder A CAS SOLUTION, Substances, Enter a query..., Edit, Search, Favorites, Clock, User.
- Left sidebar: Return to Home, Patent Markush Match (As Drawn (16), Substructure (443)), Filter by Patent Office (World Intellectual Property Organization (11), United States (4), Spain (1)).
- Main content: Patent Markush (16), Sort: Relevance, References, WO2017216177 (View Reference Detail), Patent claim 1 (PATENTPAK, Full Text), Patent table (English A1, French A1), WO2016210141 (View Reference Detail), Patent claim 1 (PATENTPAK, Full Text), and a note: "There are no notes to display for this structure."

WO 2017/216177

PCT/EP2017/064439

22

CLAIMS

1. Peptide comprising from 3 to 10 amino acids including at least one peptide sequence K*(Ac)GH or a peptide sequence K*(Ac)HIG and which may comprise an N-terminal and/or C terminal modification,
- 5
- Wherein:
- K* is selected from the group consisting of: lysine (Lys, K), ornithine (Orn), diaminobutyric acid (Dab), diaminopropionic acid (Dap) and a hydroxylated derivative of thereof;
 - K*(Ac) corresponds to a lysine, ornithine, diaminobutyric acid, diaminopropionic acid or a hydroxylated derivative thereof, acetylated on the amine of their lateral hydrocarbon chain;
 - Said modification at the N-terminus is -CO-R₁ or -SO₂-R₁;
 - Said C-terminal modification is selected from the group consisting of -OR₁, -NH₂, -NHR₁ and -NR₁R₂; and
- 10
- 15



CAS®

A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

介绍:

- 生物序列检索方法
 - 物质识别符检索
 - 结构检索
 - 关键词检索
 - Biosequence检索

如何在SciFinderⁿ 进行序列检索？——主题词检索

The screenshot displays the SciFinder search results page for the query "antibody-drug conjugate". The interface includes a search bar at the top with the query and a "Draw" button. Below the search bar, there are navigation options for "References" (7,313) and sorting options (Publication Year: Newest, View: Partial Abstract). A sidebar on the left offers filtering options such as Document Type, Language, Publication Year (with a histogram), Author, Organization, Publication Name, CAS Solutions, Formulation Purpose, Database, and Search Within Results. The "Concept" filter is highlighted with a yellow box. The main content area shows two search results. The first result is titled "Combination of antibody-drug conjugate and tubulin inhibitor" and includes a summary of the invention. The second result is titled "Potential drugs used in the antibody-drug conjugate (ADC) architecture for cancer therapy" and includes a review summary. Both results have buttons for "Full Text", "Substances", "Reactions", "Cited By", and "Citation Map".



标准概念词库 Concepts, 列出文献中的研究核心点

Concept

Top Count Alphanumeric Search

4 Selected

<input type="checkbox"/> Homo sapiens (2,967)	<input type="checkbox"/> Pancreatic neoplasm (448)	<input type="checkbox"/> Xenograft Model Antitumor Assays (289)
<input type="checkbox"/> Human (2,967)	<input type="checkbox"/> Immunoglobulin Fab fragments (446)	<input type="checkbox"/> Kidney neoplasm (288)
<input type="checkbox"/> Immunoconjugates (2,721)	<input type="checkbox"/> Drug delivery systems (441)	<input type="checkbox"/> Trastuzumab (288)
<input type="checkbox"/> Antitumor agents (2,277)	<input type="checkbox"/> Pharmaceutical carriers (439)	<input type="checkbox"/> Epidermal growth factor receptors (280)
<input checked="" type="checkbox"/> Antibody-drug conjugates (1,832)	<input type="checkbox"/> Prostate gland neoplasm (431)	<input type="checkbox"/> Colorectal neoplasm (272)
<input type="checkbox"/> Humans (1,778)	<input type="checkbox"/> Bispecific antibodies (430)	<input type="checkbox"/> Non-small-cell lung carcinoma (271)
<input type="checkbox"/> Antibodies and Immunoglobulins (1,438)	<input type="checkbox"/> Immunoglobulin G1 (422)	<input type="checkbox"/> Bladder neoplasm (267)
<input type="checkbox"/> Neoplasm (1,304)	<input type="checkbox"/> Pharmacokinetics (421)	<input type="checkbox"/> Drug design (266)
<input checked="" type="checkbox"/> Monoclonal antibodies (1,166)	<input type="checkbox"/> Proteins (413)	<input type="checkbox"/> Immunoglobulin Fab' fragments (264)
<input type="checkbox"/> Animals (974)	<input type="checkbox"/> Stomach neoplasm (409)	<input type="checkbox"/> Immunoglobulin Fab' fragments (264)
<input type="checkbox"/> Protein sequences (969)	<input type="checkbox"/> Immunoglobulin G (407)	<input type="checkbox"/> Leukemia (264)
<input type="checkbox"/> Mammary gland neoplasm (883)	<input type="checkbox"/> Peptides (401)	<input type="checkbox"/> Multiple myeloma (264)
<input type="checkbox"/> Immunoglobulin fragments (797)	<input type="checkbox"/> Antibodies, Monoclonal, Humanized (395)	<input type="checkbox"/> Enzymes (262)
	<input type="checkbox"/> Therapeutic antibodies (382)	<input type="checkbox"/> Toxins (257)

Apply Cancel

Concept

Top Count Alphanumeric Search

<input type="checkbox"/> Mammary gland neoplasm (883)	<input type="checkbox"/> Peptides (401)	<input type="checkbox"/> Multiple myeloma (264)
<input type="checkbox"/> Immunoglobulin fragments (797)	<input type="checkbox"/> Antibodies, Monoclonal, Humanized (395)	<input type="checkbox"/> Enzymes (262)
<input type="checkbox"/> Antineoplastic Agents (726)	<input type="checkbox"/> Therapeutic antibodies (382)	<input type="checkbox"/> Toxins (257)
<input type="checkbox"/> Humanized antibodies (700)	<input checked="" type="checkbox"/> Cytotoxic agents (379)	<input type="checkbox"/> CD30 antigens (252)
<input type="checkbox"/> Antibodies, Monoclonal (693)	<input type="checkbox"/> Melanoma (366)	<input type="checkbox"/> Maytansine (250)
<input type="checkbox"/> Passive immunotherapy (631)	<input type="checkbox"/> Drugs (356)	<input type="checkbox"/> T cell (250)
<input type="checkbox"/> Immunotoxins (597)	<input type="checkbox"/> Epitopes (350)	<input type="checkbox"/> Immunoassay (249)
<input type="checkbox"/> Single-chain antibodies (584)	<input type="checkbox"/> Neoplasms (348)	<input type="checkbox"/> Single-domain antibodies (249)
<input type="checkbox"/> Epidermal growth factor receptor HER2 (581)	<input type="checkbox"/> Drug targets (339)	<input type="checkbox"/> Immunoglobulin Fv fragments (246)
<input type="checkbox"/> Female (571)	<input type="checkbox"/> Hodgkin disease (337)	<input type="checkbox"/> Apoptosis (245)
<input type="checkbox"/> Combination chemotherapy (540)	<input type="checkbox"/> Targeted drug delivery systems (337)	<input type="checkbox"/> Immunoglobulin Fc fragments (244)
<input type="checkbox"/> Immunoglobulin heavy chains (528)	<input type="checkbox"/> Radionuclides (333)	<input type="checkbox"/> Programmed cell death protein 1 (244)
<input type="checkbox"/> Cancer immunotherapy (520)	<input type="checkbox"/> Cell proliferation (325)	<input type="checkbox"/> Liver neoplasm (240)
	<input type="checkbox"/> Labeled antibodies (320)	
	<input checked="" type="checkbox"/> Linking agents (318)	

Apply Cancel

抗体药物偶联物、单抗、细胞毒药物、连接器

Patent

Patent Information

Patent Number
WO2018126092

Publication Date
2018-07-05

Application Number
WO2017-US68872

Application Date
2017-12-29

Kind Code
A1

Assignee

Development Center for
Biotechnology, Taiwan
DCB-USA LLC, Taiwan

Source

World Intellectual Property
Organization

Database Information

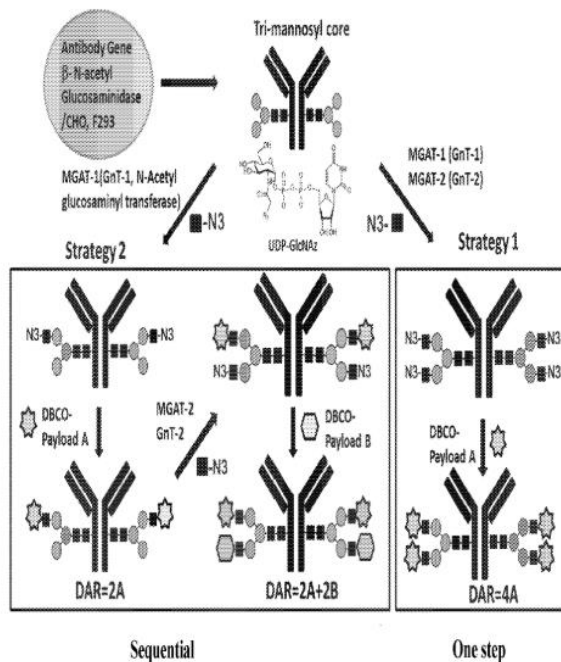
AN: 2018:1251680
CAN: 169:140742
CAplus

Language
English

Processes for preparing glycoprotein-drug conjugates

By: Tsai, Shih-Chong; Lee, Chun-Chung; Lee, Meng-Sheng; Chen, Ching-Yao; Chuang, Shih-Hsien; Chen, Yi-Jen; Wei, Win-Yin

Abstract: A process for modifying glycoproteins is provided. The invention also provides a process for producing glycoprotein-payload conjugates, as well as the conjugates produced thereby.



Patent Family

Patent	Language	Kind Code	PatentPak Options	Publication Date	Application Number	Application Date
WO2018126092	English	A1	PDF PDF+ Viewer	2018-07-05	WO2017-US68872	2017-12-29
					US2016-6262440075	2016-12-29
CA3048452	English	A1	PDF	2018-07-05	CA2017-3048452	2017-12-29
AU2017388556	English	A1	PDF	2019-07-11	AU2017-388556	2017-12-29
KR2019086029	Korean	A	PDF	2019-07-19	KR2019-7019261	2017-12-29
CN110121365	Chinese	A	PDF	2019-08-13	CN2017-80081469	2017-12-29
TW673363	Chinese	B		2019-10-01	TW2017-146600	2017-12-29
EP3568161	English	A1		2019-11-20	EP2017-889215	2017-12-29
JP2020503031	Japanese	T	PDF	2020-01-30	JP2019-534823	2017-12-29
US20200087697	English	A1	PDF	2020-03-19	US2019-16470807	2019-06-18

Expand All | Collapse All

- Concepts
- Substances
- Formulations

Glycoprotein-Payload Conjugate: Antibody-Drug Conjugate

View Formula@ Detail

Location: Claim 1, 2, 3, 4, 8, 9, 10, 11, 12
Purpose: Antibody-drug conjugates

Component	Function	Amount Reported
Group: modified glycoprotein	-	-
Group: payloads	-	-
Linking agents	-	-

Glycoprotein-Payload a/B Conjugate: Antibody-Drug Conjugate

View Formula@ Detail

Location: Claim 23, 24, 28, 29, 30, 31, 32

介绍:

- 生物序列检索方法
 - 物质识别符检索
 - 结构检索
 - 关键词检索
 - **Biosequence**检索

Biosequences: 检索生物序列及可视化分析

Biosequences: 运用Blast检索生物序列

Searching for...

- All
- Substances
- Reactions
- References
- Suppliers
- Biosequences**

Biosequences

Enter a protein or nucleotide string, or upload a .txt or .fasta file. [Learn more about Biosequence Search.](#)

BLAST CDR Motif Upload Sequence Clear Search

GSETATSGSETAGTSESATSESGAGSTAGSETSTEAGTSESATSESGAGSETATSGSETAGSETATSGSETAGTSTEASEGSASGTSTEASEGSASGTSESATSESGAGSETATSGSETAGTSTEASEGSASGSETAGTSTEAGTSESAT

Sequence Type:
 Nucleotide Protein

Search Within:
 Nucleotides Proteins

Limit Total Sequence Results to:
100

Start Biosequence Search

[Advanced Biosequence Search](#)

四种检索选择：
Protein-Protein
Protein-Nucleotides
Nucleotide-Nucleotides
Nucleotide-Proteins

Sources: https://scifinder-n.cas.org/help/#t=Searching_in_SciFinder-n%2FBiosequence_Search%2FBiosequence_Search.htm&rhsearch=biosequence&rhhlterm=biosequence&rhsyns=%20

高级检索：设置相关参数

Searching for...

- All
- Substances
- Reactions
- References
- Suppliers
- Biosequences**

Biosequences

Enter a protein or nucleotide string, or upload a .txt or .fasta file. [Learn more about Biosequence Search.](#)

BLAST | CDR | Motif |

```
GSETATSGSETAGTSESATSESGAGSTAGSETSTEAGTSESATSESGAGSETATSGSETAGSETATSGSETAGTSTEASE
GSASGTSEASEGSASGTSESATSESGAGSETATSGSETAGTSTEASEGSASGSTAGSETSTEAGTSESAT
```

Sequence Type:

Search Within: Nucleotides Proteins

Limit Total Sequence Results to:

Advanced Biosequence Search |

Alignment Identity %

Match with Gaps? Yes No

Gap Costs

Query Coverage %

Word Size

Scoring Matrix

BLAST Algorithm

E-Value

Exclude Low Complexity Regions Yes No

Query coverage = coverage/query
Sequence identity = matches/coverage
(Coverage = matches + mismatches)



BLAST检索结果

可视化地图

结果筛选

BLAST Search Details

Sequence Type: Protein
Search Within: Proteins
BLAST Algorithm: BLASTp
Sequence Identity: -
Query Coverage: 90%
E-Value: 10
Match with Gaps?: No
Gap Costs: Existence 11
Extension 1
Word Size: 3

Bioscape Analysis

Visually explore sequence similarity with a new tool.
Learn more about Bioscape.

Create Bioscape Analysis

Filter by

E-Value

0 to 10⁶

Query Coverage %

0 to 100

Subject Coverage %

0 to 100

Sequence Identity %

0 to 100

Apply Reset Filters

Biosequences (100) View: Expanded

Query Details GSETATSGSETAGTSESATSSESGAGSTAGSETSTEAGTSESATSESGAGSETATSGSETAGSETATSGSETAGTSTEASEGGSAS... View More

1 Sequence Identity: 100%

Query 1 151

Subject 1 890

Matches: 151
Mismatches: 0

View Less

Alignment Subject References

Alignment Data
BLAST Score: 711
E-Value: 2.34675e-74

```
Q 1 GSETATSGSE TAGTSESATS EGGAGSTAGS ETSTEAGTSE SATSESGAGS ETATSGSETA GSETATSGSE 70
  |||
S 1 GSETATSGSE TAGTSESATS EGGAGSTAGS ETSTEAGTSE SATSESGAGS ETATSGSETA GSETATSGSE 70
Q 71 TAGTSTEASE GSASGTSTEA SEGSASGTSE SATSESGAGS ETATSGSETA GTSTEASEGS ASGSTAGSET 140
  |||
S 71 TAGTSTEASE GSASGTSTEA SEGSASGTSE SATSESGAGS ETATSGSETA GTSTEASEGS ASGSTAGSET 140
Q 141 STEAGTSESA T 151
  |||
S 141 STEAGTSESA T 151
```

2 Sequence Identity: 100%

Query 1 151

Subject 1 935

Matches: 151
Mismatches: 0

View Less

Alignment Subject References

Alignment Data
BLAST Score: 711
E-Value: 2.34675e-74

序列匹配详情
被匹配的序列结果
披露序列的专利文献

获取SciFinder-n中披露
该序列的文献



A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

BLAST检索结果

Query Coverage% = Alignment Length / Query Length * 100%

Subject Coverage % = Alignment Length / Subject Length * 100%

Sequence identity % = No. of Matches / Alignment Length * 100%

Coverage = matches + mismatches = Alignment Length

BLAST Search Details

Sequence Type: Protein
Search Within: Proteins
BLAST Algorithm: BLASTp
Sequence Identity: -
Query Coverage: 90%
E-Value: 10
Match with Gaps?: No
Gap Costs: Existence 11
Extension 1
Word Size: 3

Bioscope Analysis

Visually explore sequence similarity with a new tool.
[Learn more about Bioscope.](#)
[Create Bioscope Analysis](#)

Filter by

^ E-Value
0 to 10⁶

^ Query Coverage %
98 to 100

^ Subject Coverage %
20 to 30

^ Sequence Identity %
58 to 59

[Apply](#) [Reset Filters](#)

Biosequences (4) View: Expanded

Query Details GSETATSGSETAGTSESATSGAGSTAGSETSTEAGTSESATSGAGSETATSGSETAGSETATSGSETAGTSTEASEGSAS... [View More](#)

1 Sequence Identity: 58.94%

Query 1 151

Subject 1 663

Matches: 89
Mismatches: 62

[View Less](#)

Alignment Subject References [References](#)

Alignment Data
BLAST Score: 394
E-Value: 3.25269e-37

Q	1	GSETATSGSE	TAGTSESATS	ESGAGSTAGS	ETSTEAGTSE	SATSESGAGS	ETATSGSETA	GSETATSGSE	70
		+ + + +							
S	26	GSSTPSGATG	SPGASPGTSS	TGSPGSPAGS	PTSTEEGTSE	SATPESGPGT	STEPSEGSAP	GSPAGSPTST	95
Q	71	TAGTSTEASE	GSASGTSTEA	SEGSASGTSE	SATSESGAGS	ETATSGSETA	GTSTEASEGS	ASGSTAGSET	140
S	96	EEGTSTEPSE	GSARGTSTEP	SEGSARGTSE	SATPESGPGS	EPATSGSETP	GSEPATSGSE	TRGSPAGSPT	165
Q	141	STEAGTSESA	T 151						
S	166	STEEGTSESA	T 176						

89/151=58.94%

2 Sequence Identity: 58.94%

Query 1 151

Subject 1 659

Matches: 89
Mismatches: 62

[View Less](#)

Alignment Subject References [References](#)

Alignment Data
BLAST Score: 394
E-Value: 3.25269e-37



CAS
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

Query Coverage%= $\text{Alignment Length} / \text{Query Length} * 100\%$

Subject Coverage %= $\text{Alignment Length} / \text{Subject Length} * 100\%$

Sequence identity %= $\text{No. of Matches} / \text{Alignment Length} * 100\%$

Coverage = matches + mismatches = Alignment Length

Bioscape——依据相似性可视化序列，获得相应专利布局信息

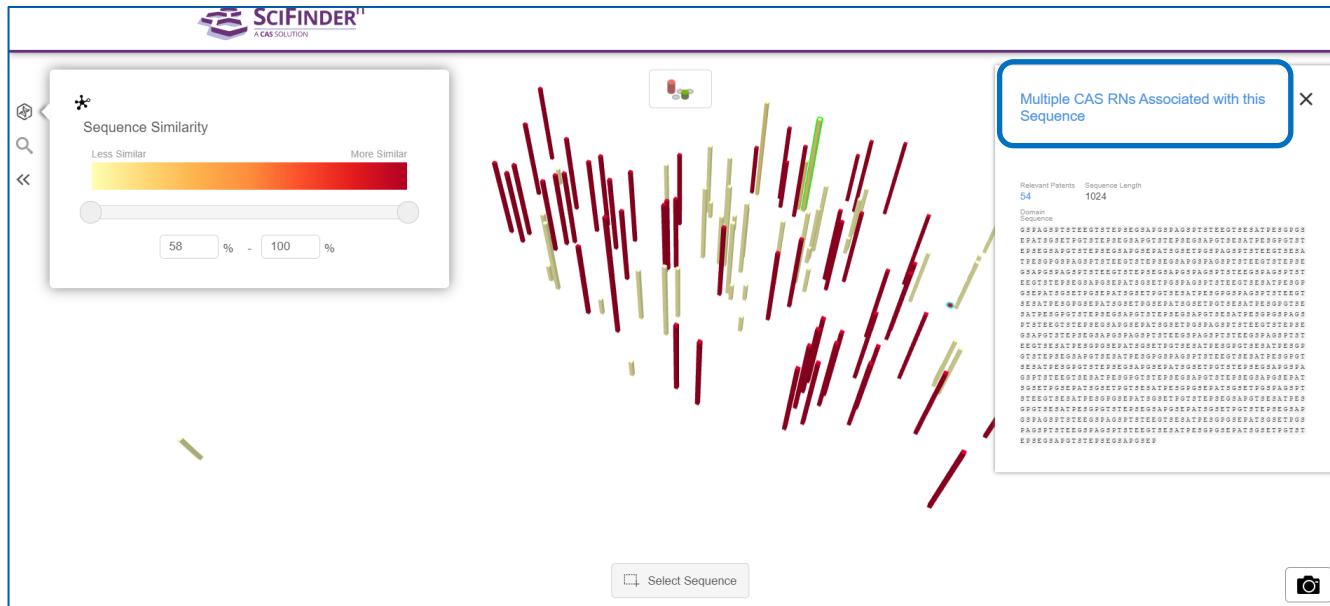
BLAST Search Details

Sequence Type: Protein
Search Within: Proteins
BLAST Algorithm: BLASTp
Sequence Identity: -
Query Coverage: 90%
E-Value: 10
Match with Gaps?: No
Gap Costs: Existence 11
Extension 1
Word Size: 3

Bioscape Analysis

Visually explore sequence
similarity with a new tool.
[Learn more about Bioscape.](#)

Create Bioscape Analysis



CAS[®]
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

输入关键词、法律状态等筛选序列

The image shows the SciFinder search interface. On the left, there is a search panel with the following fields and options:

- Search icon (magnifying glass)
- Home icon (house)
- Back icon (double left arrow)
- Field: All (input field)
- Field: Tit./Abs./Cla. (input field)
- Field: Simple legal status (dropdown menu showing "None")
- Search button (green)
- Reset button (blue)

On the right, there is a visualization of search results represented by a collection of vertical bars in red and green. A legend box is overlaid on the visualization, containing the following options:

- Undetermined
- Active
- Inactive
- Pending

At the bottom right of the visualization area, there is a button labeled "Select Sequence".

选中感兴趣的部分序列

Selected (36) X

- Sequence Result 84 54
- Sequence Result 5 42
- Sequence Result 35 25
- Sequence Result 4 23
- Sequence Result 25 23
- Sequence Result 33 23
- Sequence Result 38 23
- Sequence Result 41 23
- Sequence Result 44 23
- Sequence Result 59 23
- Sequence Result 60 23
- Sequence Result 63 23

Drag to Select/Unselect
Right Mouse Drag to Rotate

Exit

36 Selected Clear

View Patents

与目标序列相关的更多其他序列

Multiple CAS RNs Associated with this Sequence

Relevant Patents: **54** Sequence Length: **1024**

Domain Sequence

```
G S F A G S P T S T E E G T S T E P S E G S A F G S F A G S P T S T E E G T S E S A T F E S G F G S
E P A T S G S E T P G T S T E P S E G S A P G T S T E P S E G S A P G T S E S A T F E S G F G T S T
E P S E G S A P G T S T E P S E G S A P G S E P A T S G S E T P G S F A G S P T S T E E G T S E S A
T P E S G F G S F A G S P T S T E E G T S T E P S E G S A F G S F A G S P T S T E E G T S T E P S E
G S A P G S F A G S P T S T E E G T S T E P S E G S A P G S F A G S P T S T E E G S F A G S P T S T
E E G T S T E P S E G S A P G S E P A T S G S E T P G S F A G S P T S T E E G T S E S A T F E S G P
G S E P A T S G S E T P G S E P A T S G S E T P G T S E S A T F E S G F G S F A G S P T S T E E G T
S E S A T F E S G F G S E P A T S G S E T P G S E P A T S G S E T P G T S E S A T F E S G F G T S E
S A T F E S G F G T S T E P S E G S A P G T S T E P S E G S A P G T S E S A T F E S G F G S F A G S
P T S T E E G T S T E P S E G S A P G S E P A T S G S E T P G S F A G S P T S T E E G T S T E P S E
G S A F G T S T E P S E G S A P G S F A G S P T S T E E G S P A G S P T S T E E G S F A G S P T S T
E E G T S E S A T F E S G F G S E P A T S G S E T P G T S E S A T F E S G F G T S E S A T F E S G P
G T S T E P S E G S A P G T S E S A T F E S G F G S F A G S P T S T E E G T S E S A T F E S G F G T
S E S A T F E S G F G T S T E P S E G S A P G S E P A T S G S E T P G T S T E P S E G S A P G S F A
G S P T S T E E G T S E S A T F E S G F G T S T E P S E G S A P G T S T E P S E G S A P G S E P A T
S G S E T P G S E P A T S G S E T P G T S E S A T F E S G F G S E P A T S G S E T P G S F A G S P T
S T E E G T S E S A T F E S G F G S E P A T S G S E T P G T S T E P S E G S A P G T S E S A T F E S
G P G T S E S A T F E S G F G T S T E P S E G S A P G S E P A T S G S E T P G T S T E P S E G S A P
G S F A G S P T S T E E G S P A G S P T S T E E G T S E S A T F E S G F G S E P A T S G S E T P G S
F A G S P T S T E E G S P A G S P T S T E E G T S E S A T F E S G F G S E P A T S G S E T P G T S T
E P S E G S A P G T S T E P S E G S A P G S E P
```

Substances (6)

Sort: Relevance View: Partial

References Reactions Suppliers

1	2	3
1454337-54-2 Image Not Available Unspecified Protein (synthetic clone AE1236) Protein/Peptide Sequence Sequence Length: 1024 2 References 0 Reactions 0 Suppliers	1451688-30-4 Image Not Available Unspecified 98: PN: WQ2013122617 SEQID: 102 unclaimed protein Protein/Peptide Sequence Sequence Length: 1024 2 References 0 Reactions 0 Suppliers	1427477-48-2 Image Not Available Unspecified Protein (synthetic clone AE1236) Protein/Peptide Sequence Sequence Length: 1024 1 Reference 0 Reactions 0 Suppliers
4	5	6
1384912-83-7 Image Not Available Unspecified XTEN protein (synthetic clone AE1236) Protein/Peptide Sequence Sequence Length: 1024 1 Reference 0 Reactions 0 Suppliers	1338302-41-2 Image Not Available Unspecified Extended recombinant protein AE1236 (synthetic) Protein/Peptide Sequence Sequence Length: 1024 1 Reference 0 Reactions 0 Suppliers	1338291-97-6 Image Not Available Unspecified Protein/Peptide Sequence Sequence Length: 1024 1 Reference 0 Reactions 0 Suppliers

Biosequences: 运用CDR检索生物序列

Searching for...

- All
- Substances
- Reactions
- References
- Suppliers
- Biosequences**

Biosequences

Enter a protein string, or upload a .txt or .fasta file. [Learn more about Biosequence Search.](#)

BLAST **CDR** Motif [Clear Search](#)

CDR1	<input type="text" value="VFPLAPSSKS"/>	<input type="button" value="X"/>
CDR2	<input type="text" value="TSGGTAALGC"/>	<input type="button" value="X"/>
CDR3	<input type="text" value="LVKDYFPEPV"/>	<input type="button" value="X"/>

Limit Total Sequence Results to:

https://scifinder-n.cas.org/help/#t=Searching_in_SciFinder-n%2FBiosequence_Search%2FBiosequence_Search_-_CDR.htm&rhsearch=CDR&rhhlterm=CDR&rhsyns=%20

CDR检索结果

- 左侧呈现匹配到query中某一个或者多个CDR区的subject序列的数量。
- 点击圈内的数字，再点Apply即可查看匹配的序列结果。
- 点Reset segments，可重新选择查看匹配的序列结果。

Query Coverage % = coverage/query
Subject Coverage % = coverage/subject
Alignment Identity % = matches/query

The screenshot displays the Biosequences interface with the following components:

- CDR Segments:** A Venn diagram showing the overlap of CDR1, CDR2, and CDR3. The numbers in the regions are: CDR1 only (0), CDR2 only (0), CDR3 only (9), CDR1 & CDR2 (0), CDR1 & CDR3 (0), CDR2 & CDR3 (0), and all three (91). Buttons for 'Apply' and 'Reset Segments' are present.
- Bioscape Analysis:** A section for visualizing sequence similarity with a new tool, including a 'Create Bioscape Analysis' button.
- Filter by:** A section with sliders and dropdowns for 'E-Value' (0 to 10⁶), 'Query Coverage %' (0 to 100), 'Subject Coverage %' (0 to 100), and 'Alignment Identity %' (0 to 100). Buttons for 'Apply' and 'Reset Filters' are at the bottom.
- Query Details:** Shows the query sequence: CDR1 VFPLAPSSKS, CDR2 TSGGTAALGC, and CDR3 LVKDYFPEPV.
- Alignment 1:** Shows a subject sequence (445) with a CDR3 region highlighted. The alignment identity is 100%. Matches: 10, Mismatches: 0. A 'References' button is visible.
- Alignment 2:** Shows another subject sequence (444) with a CDR3 region highlighted. The alignment identity is 100%. Matches: 10, Mismatches: 0. A 'References' button is visible.
- Alignment Data:** Shows BLAST Score: 79 and E-Value: 0.00137694 for both alignments.
- Sequence Alignment:** A table showing the alignment of CDR3 between the query and subject sequences:

Query	Subject
CDR3 1 LVKDYFPEPV 18	146 LVKDYFPEPV 155

序列匹配详情
被匹配的序列结果
披露序列的专利文献

获取SciFinder-n中
披露该序列的文献



CAS[®]
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

CDR检索结果

查看query中三个CDR区都
被匹配的91个序列结果

CDR Segments

Select a segment below to view individual or intersecting CDR results.

CDR1 CDR2 CDR3

91

Apply Reset Segments

Bioscape Analysis

Visually explore sequence similarity with a new tool.
[Learn more about Bioscape.](#)

Create Bioscape Analysis

Filter by

E-Value

0 to 10^6

Query Coverage %

0 to 100

Subject Coverage %

0 to 100

Alignment Identity %

0 to 100

Biosequences (91) View: Expanded

Query Details View Less

> CDR1
VFPLAPSSKS

> CDR2
TSGGTAALGC

> CDR3
LVKDYFPEPV

1 Alignment Identity: 100%

Matches: 30
Mismatches: 0

Subject: 1 470

CDR3
CDR1
CDR2

View Less

Alignment Subject References

항 CD147 항체
Assignee: 다이이찌 산쿄 가부시키가이샤
KR1020200032185 A | Seq ID No: 98

Anti-cd147 antibody
Assignee: DAIICHI SANKYO COMPANY, LIMITED
EP3660155 A1 | Seq ID No: 98

抗CD147抗体
Assignee: 第一三共株式会社
CN111051513 A | Seq ID No: 84

Biosequences: 运用Motif检索生物序列

Searching for...

- All
- Substances
- Reactions
- References
- Suppliers
- Biosequences**

Biosequences

Enter a protein or nucleotide string. [Learn more about Biosequence Search.](#)

BLAST CDR **Motif** Clear Search

[SG]x{4}GK[DT]

Advanced Biosequence Search ^ Reset All

Query Coverage % E-Value

Sequence Type:

Nucleotide **Protein**

Limit Total Sequence Results to:

100

Start Biosequence Search

https://scifinder-n.cas.org/help/#t=Searching_in_SciFinder-n%2FBiosequence_Search%2FBiosequence_Search_-_Motif.htm&rhsearch=motif&rhhlterm=motif&rhsyns=%20

Motif检索结果

Motif Search Details

Sequence Type: **Protein**
Query Coverage: **90%**
E-Value: **10**

Bioscape Analysis

Visually explore sequence similarity with a new tool.
[Learn more about Bioscape.](#)

Create Bioscape Analysis

Filter by

E-Value
0 to 10⁶

Query Coverage %
0 to 100

Subject Coverage %
0 to 100

Alignment Identity %
0 to 100

Apply **Reset Filters**

Biosequences (100)

View: Expanded

Query Details [View More](#)

> Seq 1: 1 SXXXXGKD 8

1

Alignment Identity: 87.5%

Query 1 8

Subject 1 231

Matches: 7
Mismatches: 1

[View Less](#)

Alignment Subject References

Alignment Data
BLAST Score: 78
E-Value: 0.000435191

Q	1	SXXXXGKD	8
		+	
S	71	XXXXXGKD	78

2

Alignment Identity: 87.5%

Query 1 8

Subject 1 51

Matches: 7
Mismatches: 1

小结:

- 物质识别符: 名称、代码、CAS RN, etc.
- 结构检索: Substance、Markush
- 关键词检索: substance key words, research topic
- Biosequence检索: protein & nucleotides sequences code(BLAST/CDR/Motif)

谢谢关注!



china@acs-i.org

+86-10-62508026/27

www.cas.org

美国化学文摘社北京代表处