

核酸・ペプチド・タンパク質の検索

2021 年 3 月

* 目次 *

核酸・ペプチド・タンパク質の登録例	1
核酸・ペプチド・タンパク質の検索手順	3
化学物質名称, CAS RN [®] , GenBank 番号	3
構造検索	4
BLAST 配列検索	5
(参考) Advanced Biosequence Search 画面 (BLAST 検索のパラメータ設定)	7
CDR 配列検索	8
Motif 配列検索	9
Bioscape	10
同主鎖の配列を持つ物質の検索	12

JAICI
化学情報協会

〒113-0021 東京都文京区本駒込6-25-4 中居ビル
TEL: 0120-003-462
E-mail: support@jaici.or.jp

核酸・ペプチド・タンパク質の登録例

核酸・ペプチド・タンパク質は一つでも配列が異なれば別物質として収録される。

- 化学修飾された配列，側鎖の置換基の異なる配列，同位体で置換された配列等は，同じ配列でも別物質として収録される。
- GenBank から収録された配列は，1 GenBank 番号につき 1 物質として収録される。

■ 核酸（9 以上の塩基配列を持つ物質）のレコード例

Substance Detail (1 of 1)

Reference (1) | Reactions (0) | Suppliers (0)

CAS Registry Number: 2128329-99-5

Image Not Available

Unspecified
DNA (synthetic clone WO2017-140839 oligonucleotide) (ACI)

Nucleic Acid Sequence
Sequence Length: 115
34 a, 27 c, 27 g, 27 t
modified

Related Sequences (1)

Other Names and Identifiers
1 Other Name for this Substance
1: PN: WO2017140839 SEQID: 2 claimed DNA

Sequence Details
Sequence: DNA: linear

1	catgttcgat	gaagcacgat	agatgtacgc	tttgacatgc	gctttgacaa
51	tacttgasca	ctcgccagat	ataggatgtt	gcaactcgc	tgagtcacc
101	aaaccaaaaa	cctcg	-	-	-

Sequence Modifications

Type	Location	Description
modified base	cytidine-1	5'-phosphate

Patent Annotations
Source: Not Given
Reference: WO2017140839, SEQID 2; claimed

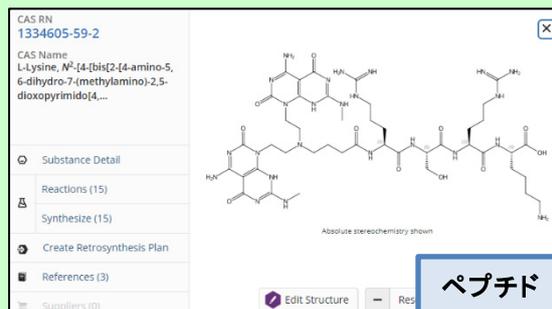
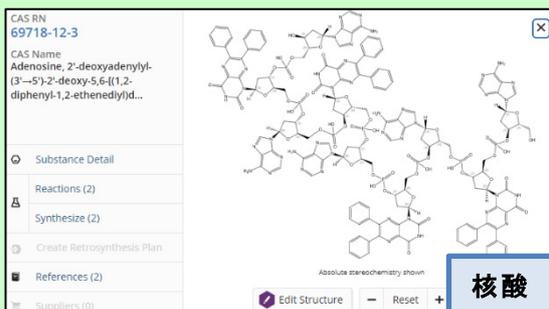
Callouts:

- 関連情報のリンク
- CAS RN®
- 化学物質名
- 配列長
- 核酸タイプと数
- 同主鎖の配列を持つ物質の検索ボタン
- 配列
- 特徴表
- 特許番号と配列の記載位置，配列番号



大部分の核酸・ペプチド・タンパク質のレコードには，構造図や分子式は収録されない。ただし，水素以外の元素数が 252 以下の核酸・タンパク質・ペプチドでは，物質情報に構造図や分子式が収録されており，構造検索や分子式検索も可能。

【構造図が収録されているレコード例】



■ ペプチド・タンパク質（4 以上のアミノ酸残基を持つ物質）のレコード例

関連情報のリンク

CAS RN®

分子式

化学物質名

配列

特徴表

**特許番号と配列の
記載位置, 配列番号**

Substance Detail (1 of 1)

References (7) Reactions (0) Suppliers (0)

CAS Registry Number: 216259-64-2

構造図
(水素以外の元素数が 252 以下の場合, 構造検索可能)

Absolute stereochemistry shown

配列長

Protein/Peptide Sequence
Sequence Length: 6 modified (modifications unspecified)

同主鎖の配列を持つ物質の検索ボタン

Other Names and Identifiers

Canonical SMILES
O=C1NC(C(=O)NC(C(=O)NC(C(=O)N2CCCC2C(=O)NC1CC3=CC=C(O)C=C3)CC=4C=CC=CC4(O)C)C5CCCC(N)CC5)CC6=CNC=7C=CC=CC76

Isomeric SMILES
C[C@H]3C(=O)N[C@@H](C[C@@H]4CC[C@H](N)CC4C(=O)N[C@@H](C[C@@H](O)C(=O)N[C@@H](C[C@@H](O)N)CC5=CC=CC=C5)C(=O)N6(C)C(=O)N[C@@H](C7=CC=C(O)C=C7)C(=O)N3)CCC6[H]][H]

2 Other Names for this Substance
70: PN: US20020042374 PAGE: 10 claimed protein
74: PN: US6268342 SEQID: 80 claimed protein

Sequence Details

Sequence: cyclic

1	ATFPY#	-	-	-	-
---	--------	---	---	---	---

Sequence Modifications

Type	Location	Description
stereo	tryptophan-6	D

Patent Annotations

Source: Not Given
Reference: US6268342, SEQID 80; claimed

Source: Not Given
Reference: US20020042374, PAGE 10; claimed

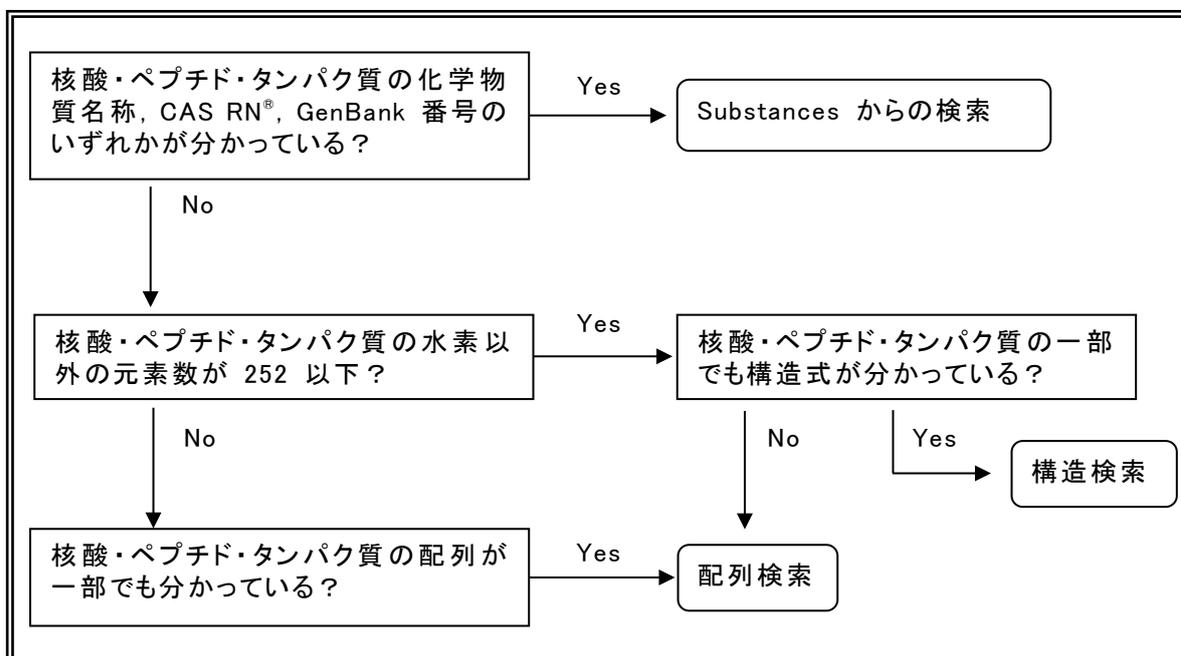
登録ルールの例外

- ・ 2002 年以降の特許・雑誌由来の配列は, 修飾基も含めて同一の配列であっても同一レコードにならず特許・雑誌ごとに別レコードとして収録される場合がある。
- ・ GenBank 由来の配列と文献・特許由来の配列が同じ配列であっても同一レコードにならず, 別レコードとして収録される場合がある。

2

核酸・ペプチド・タンパク質の検索手順

■ 核酸・ペプチド・タンパク質の検索方針



■ 化学物質名称, CAS RN®, GenBank 番号

The first screenshot shows a search for "GABAA receptor GABRA1 (human brain)" under the "Substances" category. A callout box labeled "化学物質名称" points to the search input.

The second screenshot shows a search for "1448554-62-8" under the "Substances" category. A callout box labeled "CAS 登録番号 (CAS RN)" points to the search input.

The third screenshot shows a search for "AAH30696" under the "Substances" category. A callout box labeled "GenBank 番号" points to the search input. Below the search bar, a dropdown menu is open, showing "GenBank AAH30696" selected with a red arrow and the text "この番号を選択". Other options include "GenBank AAH30696 (Translated from: GenBank BC030696)" and "GenBank AAH30196".

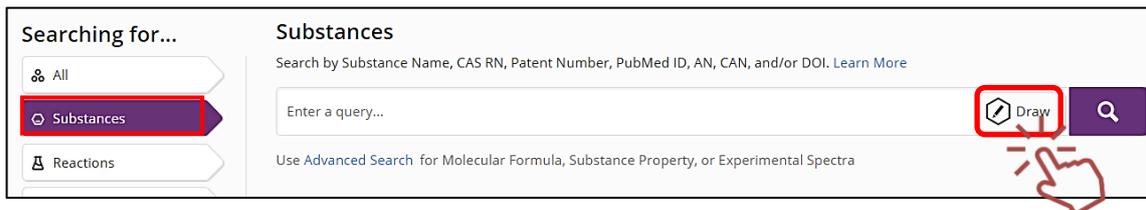


GenBank 番号の入力ルール

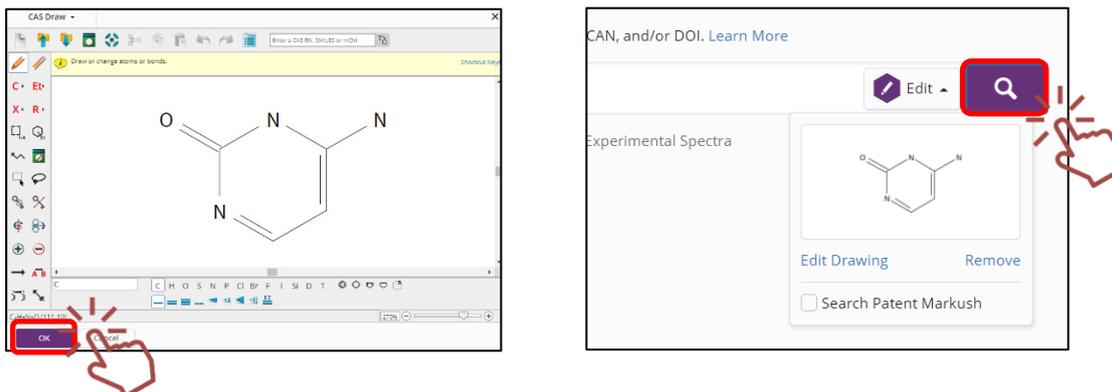
GenBank 番号から検索する場合は番号の前に「GenBank」を入力して検索します。検索ボックスに GenBank 番号を入力すると、オートサジェスト機能で「GenBank」の文字を含む候補が表示されるので、検索したい番号を選択してください。

■ 構造検索

- ① Substances から Draw をクリックし、構造を作図する。



- ② 作図が終了したら OK をクリックして、検索を実行する。



- ③ 回答が表示される。得られた回答が多かった場合は、物質の種類（Substance Class）のフィルタを用いて核酸やペプチド・タンパク質に限定すると探しやすい。

The screenshot shows the search results page with the 'Substances' tab selected. The 'Substance Class' filter menu is open, and the 'Nucleic Acid Sequence' filter is selected. A green box highlights the filter options with a label '物質の種類による限定'.

Substance ID	Chemical Name	References	Reactions	Suppliers
928769-18-0	C ₆₂ H ₁₁₀ N ₃₈ O ₂₈ Peptide nucleic acid, (Gly-C-C-T-[1'-de(6-amino-9H-purin-9-yl)-1'-[3,4,5,6,7,8-h...	3	9	0
1334605-59-2	C ₄₃ H ₇₀ N ₂₄ O ₁₁ L-Lysine, N ² -[4-bis[2-[4-amino-5,6-dihydro-7-(methylamino)-2,5-dioxopyrimido(4...	3	15	0
1334605-58-1	C ₁₂₂ H ₁₇₈ N ₂₄ O ₃₁ S ₂ L-Lysine, N ² -[4-bis[2-[4-bis[[1,1-dimethylthoxy]carbonyl]amino]-7-[[[1,1-dime...	3	14	0
1356399-08-0	C ₇₀ H ₉₉ N ₃₈ O ₁₇ Peptide nucleic acid, (H-G-G-G-[W-[3-(trimethylammonio)propyl]]IsoG-[W-[3-(trime...			
928769-23-7	C ₈₂ H ₁₁₀ N ₃₈ O ₂₈ :C ₇₀ H ₈₅ N ₃₅ O ₃ ... Components: 2			
928769-21-5	C ₈₂ H ₁₁₀ N ₃₈ O ₂₈ :C ₇₀ H ₈₅ N ₃₅ O ₃ ... Components: 2			

■ BLAST 配列検索

Biosequences から配列質問式を使って、BLAST ホモロジー検索を実行することができます。

- ① Biosequences をクリックします。
- ② 配列質問式を入力するか Upload (.txt または .fasta) で呼び出します。
- ③ 配列質問式の選択
- ④ 回答の配列の選択
- ⑤ 回答の上限数を選択
- ⑥ (任意) パラメータ設定
- ⑦ 検索の実行

The screenshot shows the Biosequences search interface. On the left, a sidebar lists navigation options: All, Substances, Reactions, References, Suppliers, and Biosequences (highlighted). The main area is titled 'Biosequences' and contains a search input field with the sequence 'gcgtttgctcttcttcttgcg' (callout 2). Above the input are tabs for 'BLAST', 'CDR', and 'Motif' (callout 1), and buttons for 'Upload Sequence' and 'Clear Search'. To the right of the input are settings for 'Sequence Type' (Nucleotide selected, callout 3), 'Search Within' (Nucleotides selected, callout 4), and 'Limit Total Sequence Results to' (100, callout 5). Below these are 'Advanced Biosequence Search' parameters (callout 6) including Sequence Identity % (80), Query Coverage % (90), Match with Gaps? (No), Word Size (11), Gap Costs (Existence 5 Extension 2), Reward for Match, Penalty for Mismatch (2, -3), BLAST Algorithm (BLASTn), E-Value (10), and Exclude Low Complexity Regions (No). A 'Start Biosequence Search' button (callout 7) is at the bottom right, with a hand icon pointing to it.

- ⑧ 配列検索には時間がかかることがあります。Search History で「Searching」と表示されている場合は検索実行中です。検索が終わると Searching の表示が Complete に変わります。回答を表示するには、View Results ボタンを押します。

The screenshot shows the Search History page. The top navigation bar includes the SCIFINDER logo, a search input field, and icons for Draw, Search, Favorites, Search History (highlighted), and User Profile. The main content area is titled 'Search History (11)'. A filter sidebar on the left shows 'Result Type' with 'Biosequences (11)' selected. The search history table shows a search from January 14, 2021 at 9:25 AM. The search details are: Biosequences, Sequence Type: Nucleotide, Search Within: Nucleotides, BLAST Algorithm: BLASTn, Sequence Identity: 80%, and Query Coverage: 90%. A 'View Results' button (callout 7) is highlighted with a hand icon, along with 'Edit Search' and 'Complete' buttons.

⑨ View Results ボタンをクリックすると、配列検索の結果が表示されます。

検索条件

絞り込み (フィルタ)

- E-Value (期待値)*
- Query Coverage %*
- Subject Coverage %*
- Sequence Identity %*

アライメントの概略図
(質問配列の 1~21 番目のコードが回答配列の一部と一致している)

エクスポート

同一性パーセント*

Subject タブ: 回答配列の詳細情報

References タブ: 出典文献情報

Alignment タブ: アライメント*・スコア値*・期待値*

記載文献の検索

記載文献
(2021年3月時点では特許のみ表示されます)

* 用語説明

用語	内容
E-Value (期待値)	データベース中の配列に対してマッチする際の統計的有意性の閾値
Query Coverage %	類似領域 ÷ 配列質問式 × 100
Subject Coverage %	類似領域 ÷ 回答配列 × 100
Sequence Identity % (同一性パーセント)	一致したコード ÷ 類似領域 × 100
Alignment (アライメント)	配列質問式と回答配列のどこが類似した領域か特定できるように並べたもの
BLAST Score (スコア値)	配列質問式と回答配列の対応するコードの一致度をスコアとして計算したもの

(参考) Advanced Biosequence Search 画面 (BLAST 検索のパラメータ設定)

- 1** 配列長が 30 以下の短い配列質問式の際に用いるパラメータ
- 2** 配列質問式に対して回答配列の類似領域中で一致する割合 (一致したコード ÷ 類似領域 × 100)
- 3** 配列質問式の配列長に対して、ヒットした回答の配列と重複している長さの割合 (類似領域 ÷ 配列質問式 × 100)

4 検索タイプ

検索タイプ	検索機能	質問式	回答
BLASTn	塩基配列の質問式に類似した塩基配列を検索	核酸	
Mega BLASTn	塩基配列の質問式により類似した塩基配列を高速で検索	核酸	
BLASTn-short	短い塩基配列の質問式に類似した塩基配列を検索	核酸	
tBLASTn-fast	データベース中の塩基配列をアミノ酸配列に翻訳した配列の中から、アミノ酸配列の質問式に類似した配列を高速で検索	タンパク質	核酸
BLASTp	アミノ酸配列の質問式に類似したアミノ酸配列を検索	タンパク質	
BLASTp-fast	アミノ酸配列の質問式に類似したアミノ酸配列を高速で検索	タンパク質	
BLASTp-short	短いアミノ酸配列の質問式に類似したアミノ酸配列を検索	タンパク質	
BLASTx-fast	塩基配列の質問式をアミノ酸配列に翻訳して、データベース中の類似したアミノ酸配列を高速で検索	核酸	タンパク質

- 5** ギャップを考慮してマッチさせる/させない
- 6** 検索用の文字列の長さ
- 7** 期待値 (デフォルトは 10)
- 8** Open Gap Cost と Extend Gap Cost の組み合わせ
- 9** マッチや ミスマッチ の場合の ペナルティ スコア
- 10** 低複雑度領域のマスクフィルタリングの使用

■ CDR 配列検索

Biosequences で CDR タブを選択すると、抗体や T 配列細胞受容体の CDR を指定した配列検索を実行することができます。

- ① Biosequences 画面で CDR タブをクリックします。
- ② 配列質問式を入力するか Upload (.txt または .fasta) で呼び出します。
- ③ 回答の上限数を選択
- ④ 検索の実行

The screenshot shows the Biosequences search interface. On the left, there's a sidebar with 'Biosequences' selected. The main area has tabs for 'BLAST', 'CDR', and 'Motif'. The 'CDR' tab is active, and three CDR regions are entered: CDR1 (RASQGI RNYLA), CDR2 (AASTLQS), and CDR3 (QRYNRAPYT). A dropdown menu for 'Limit Total Sequence Results to:' is set to 20000. A red box highlights the search input area, and a hand icon points to the 'Start Biosequence Search' button.

- ⑤ 検索が終わると検索履歴中の *Searching* の表示が *Complete* に変わります。View Results ボタンをクリックすると、配列検索の結果が表示されます。

The screenshot shows the Biosequences search results page. On the left, there's a 'CDR Segments' section with a Venn diagram showing the overlap of CDR1, CDR2, and CDR3. The 'Query Details' section shows alignment information for CDR1, CDR2, and CDR3. The 'Subject' tab is selected, showing alignment data and a diagram of the CDR regions. The 'References' tab is also visible, showing search results. Green boxes highlight specific features: 'アライメントの類似性' (Alignment Similarity), 'アライメントの概略図' (Alignment Overview Diagram), 'Subject タブ: 回答配列の詳細情報' (Subject Tab: Detailed Information of Answer Sequences), '記載文献の検索' (Search for Cited Literature), 'References タブ: 出典文献情報' (References Tab: Citation Information), and 'Alignment タブ: アライメント・スコア値・期待値' (Alignment Tab: Alignment, Score, and Expectation Value).

■ Motif 配列検索

Biosequences で Motif タブを選択すると、短いパターン配列を検索することができます。

- ① Biosequences 画面で Motif タブをクリックします。
- ② 配列質問式を入力します。角括弧を用いることにより、代替残基を含めた配列検索を行うことができます。
- ③ (任意) パラメータ設定
- ④ 配列質問式の種類
- ⑤ 回答の上限数を選択
- ⑥ 検索の実行

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

[LI][ALKV]G[FL][VI]D[AG]DG

Sequence Type:
Nucleotide Protein

Limit Total Sequence Results to:
20000

Start Biosequence Search

Advanced Biosequence Search ^
Query Coverage % 90 E-Value 10 Reset All

- ⑦ 検索が終わると検索履歴中の *Searching* の表示が *Complete* に変わります。View Results ボタンをクリックすると、配列検索の結果が表示されます。

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 (4,718) View: Collapsed ^

Query Details View More

> Seq 1: 1 LAGFVDADG 9
> Seq 2: 1 LAGFVDG 9
> Seq 3: 1 LAGFIDADG 9
> Seq 4: 1 LAGFIDG 9
> Seq 5: 1 LAGLVDADG 9
> Seq 6: 1 LAGLVDG 9
> Seq 7: 1 LAGLIDADG 9
> Seq 8: 1 LAGLIDG 9
> Seq 9: 1 LLAGFVDADG 9
> Seq 10: 1 LLAGFVDG 9
> Seq 11: 1 LLAGFIDADG 9

Alignment Identity: 100%
Matches: 9
Mismatches: 0

Alignment Data
BLAST Score: 63
E-Value: 0.42984

Q 1 LAGFVDADG 9
| | | | | | | | | |
S 204 LAGFVDADG 212

代替残基を含めた検索結果の場合は、Query Details の配列をクリックして表示されるリストから別の配列を選択することができます

アライメントの概略図

アライメントの類似性

Subject タブ: 回答配列の詳細情報

参考文献の検索

References タブ: 出典文献情報

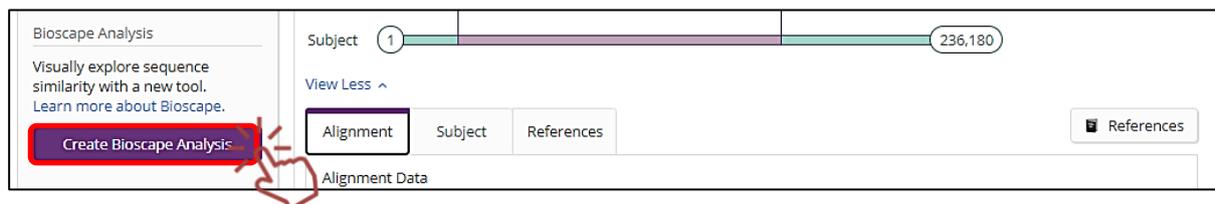
Alignment タブ: アライメント・スコア値・期待値

■ Bioscape (バイオスケープ)

Bioscape は配列検索で得られた回答集合から、配列の類似性により解析したマップを作成する機能です。マップ内には関連特許の件数が 3D で示されるため配列関連特許を視覚的に分かりやすく解析することができます。システム制限（一度に解析できる配列の件数）は 1,000 件です。

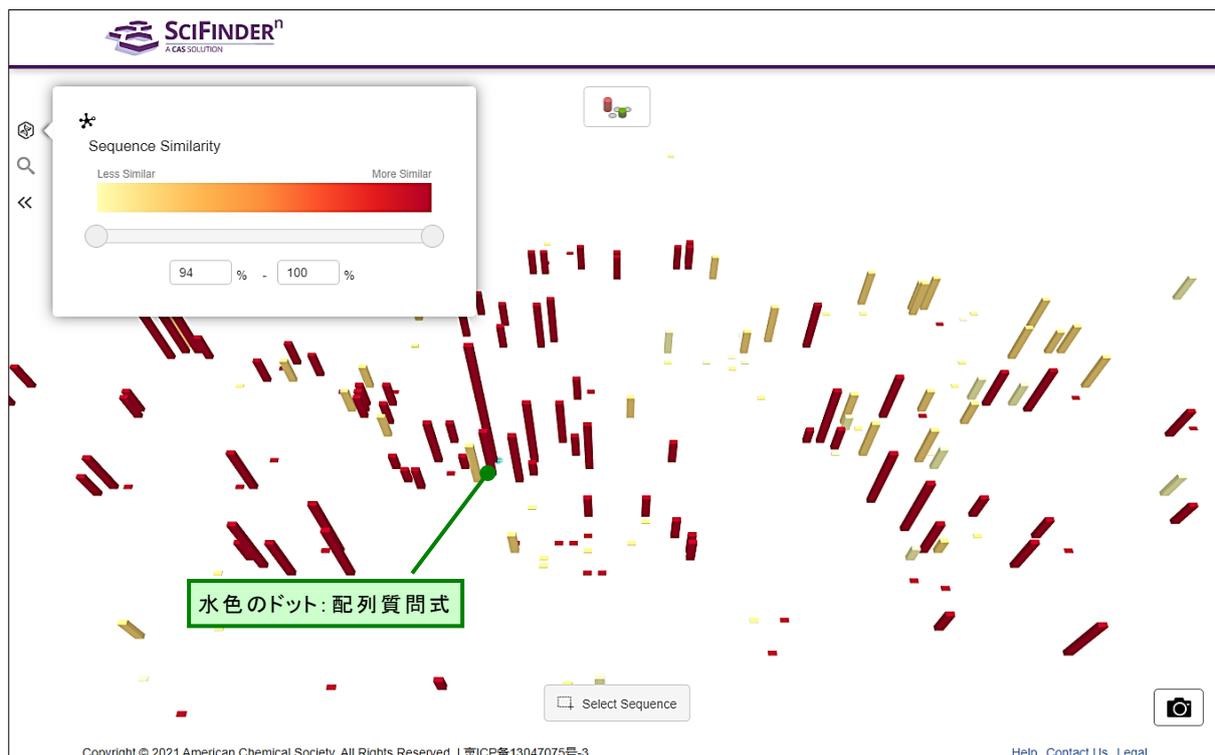
Bioscape 解析マップ作成手順

配列検索結果の画面左側に表示される「Create Bioscape Analysis」ボタンをクリックします。



別のタブに Bioscape の解析結果が表示されます。実行直後は、Sequence Similarity で解析した配列の類似性によるマップが表示されます。検索に用いた配列質問式が水色のドットです。その周囲に類似性が高い配列が赤色で、類似性が低い物質は黄色で示されます。

3D マップのバーの高さは、特許の件数を表しています。バーをクリックすると、該当する配列のモーダルウィンドウが表示され、その中の数字のリンクをクリックすると特許を表示することができます。



【コントロールパネル】画面左にはコントロールパネルが表示され、下記の操作を実行できます。

	Chart	マップに反映する類似度の設定
	Search	解析結果内の該当物質の検索

同主鎖の配列を持つ物質の検索

同じ配列であっても異なる CAS RN® が付与されたレコードが存在することがあります。

- ・ 配列の主鎖は同じであっても、化学修飾、側鎖の置換基の異なるもの、同位体で置換された配列は別配列として別レコードに収録されます。化学修飾などの情報は特徴表に記載されます。
- ・ GenBank 由来の核酸配列では、更新前と更新後の配列は、同じ配列であっても別レコードとして収録されています。
- ・ 登録ルールの例外によって、同じ配列であっても別レコードとして収録されている場合があります。

■ Related Sequences

同主鎖の配列を持つ配列を調べるには、Related Sequences 機能を使います。

核酸やタンパク質の物質の詳細情報を表示し、Related Sequences ボタンをクリックすると同主鎖の配列を持つ物質を検索できます。

同主鎖の配列レコードが 11 件存在する

Substance Detail (1 of 1)

References (2) Reactions (0) Suppliers (0)

CAS Registry Number
479879-70-4

Unspecified
Gamma-aminobutyric acid (GABA) A receptor, alpha 1 (human clone MGC:26564 IMAGE:4820972) (9CI)

Protein/Peptide Sequence
Sequence Length: 456

Related Sequences (11)

Other Names and Identifiers

Sequence Details

Sequence: linear					
1	MRKSPGLSDC	LWAWLLLLST	LTGRSYGRPS	LDDELKDNNT	VFTRILDRLL
51	DGYDNLRLRP	LGERVFEVKT	DI FVTSFGPV	SDHDMVYTD	VFFRSGWKDE
101	RLFKFGPMTV	LRLNLMASK	IWPDFFHN	GKSVVHMT	MPNKLRI TE
151	DGTLTYMRL	TVRAECPMHL	EDFPDAHAC	PLKFGSYAYT	RAEVVYEWTR
201	EPARSVYVAE	DGSRNLWYDL	LGFVDSGIV	QSSTEVYVM	TFHFILKRIKI
251	GYFVIQTYLP	CIMTVILSQV	SFWLNRESVP	ARTVFGVTV	LTMTLSIS SA
301	RNSLPKRYVA	TAMDFI AVC	YAFVSALIE	FATVNYFTR	GYRWGKSVV
351	PEKPKKWDP	LIKKNTYAP	TATSYFHLA	RGDPLATIA	KSATIEPKV
401	KPEFKPEPK	KTFMSVSKID	ALSRIFPDL	FQIFNLVYA	TYLNRPEPK
451	APTPHQ	-	-	-	-

同主鎖の配列を持つ物質

Substance

References

176229-25-7
Image Not Available
Unspecified
GABAA receptor GABRA1 (human brain)
Protein/Peptide Sequence
Sequence Length: 456

2098448-41-8
Image Not Available
Unspecified
GABAA receptor GABRA1 (human)
Protein/Peptide Sequence
Sequence Length: 456

1700675-57-5
Image Not Available
Unspecified
GABAA receptor GABRA1 (human)
Protein/Peptide Sequence
Sequence Length: 456

1700675-53-1
Image Not Available
Unspecified
GABAA receptor GABRA1 (human)
Protein/Peptide Sequence
Sequence Length: 456

1448554-62-8
Image Not Available
Unspecified
Protein/Peptide Sequence
Sequence Length: 456

1448104-12-8
Image Not Available
Unspecified
GABAA receptor GABRA1 (human clone FR298321-SEQID:23)
Protein/Peptide Sequence
Sequence Length: 456

1190908-30-5
Image Not Available
Unspecified
GABAA receptor GABRA1 (human precursor)
Protein/Peptide Sequence
Sequence Length: 456

1172644-36-8
Image Not Available
Unspecified
GABA (gammaaminobutyric acid A) receptor alpha 1 (human)
Protein/Peptide Sequence
Sequence Length: 456

1087358-21-1
Image Not Available
Unspecified
GABAA receptor GABRA1 (human)
Protein/Peptide Sequence
Sequence Length: 456

628759-25-1
Image Not Available
Unspecified
Protein (human clone US030219744-EE QID:331)
Protein/Peptide Sequence
Sequence Length: 456

528642-91-3
Image Not Available
Unspecified
Protein (human clone US669961:SEQID:331)
Protein/Peptide Sequence
Sequence Length: 456