

* 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 登録番号 (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

The screenshot shows a Substance Detail page for a DNA sequence. Annotations on the left side point to specific fields:

- 関連情報のリンク**: Points to the Reference (1) link.
- CAS RN®**: Points to the CAS Registry Number 2128329-99-5.
- 化学物質名**: Points to the name 'Unspecified' and the description 'DNA (synthetic clone WO2017-140839 oligonucleotide) (ACI)'.
- 配列長**: Points to the 'Sequence Length: 115'.
- 核酸タイプと数**: Points to the '34 a, 27 c, 27 g, 27 t modified'.
- 同主鎖の配列を持つ物質の検索ボタン**: Points to the 'Related Sequences (1)' link.
- 配列**: Points to the 'Sequence Details' section, which contains a table of the DNA sequence:

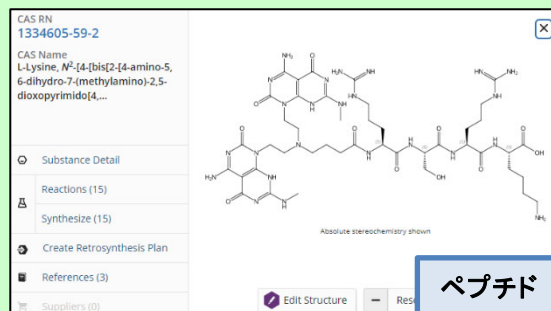
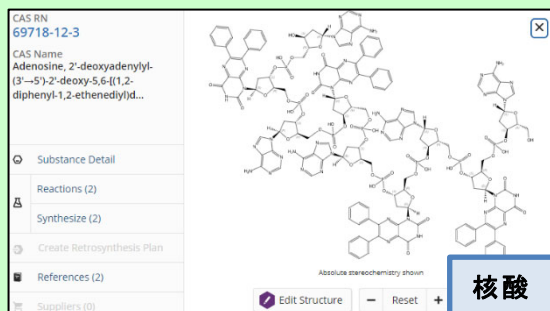
Position	1	51	101
Sequence	catgttcgat gaggcacgat agatgtacgc tttagacatc gctttgacaa	tacttgagca gtcggcagat ataggatgtt gcaagctccg tgagtcaccac	aaaccaaaaa cctcg - - -
- 特徴表**: Points to the 'Sequence Modifications' table:

Type	Location	Description
modified base	cytidine-1	5'-phosphate
- 特許番号と配列の記載位置，配列番号**: Points to the 'Patent Annotations' section, which lists 'Source: Not Given' and 'Reference: WO2017140839, SEQID 2: claimed'.



大部分の核酸・ペプチド・タンパク質のレコードには，構造図が収録されていないが，水素以外の元素数が 252 以下でさらに原文献から構造が特定できる場合は収録される。構造図が収録されている場合は，構造検索が可能（配列長が 9 以上の核酸については，構造検索できない）。

【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)N(C(=O)N(C(=O)N2CCC2C(=O)N1CC3=CC=C(O)C=C3)CC=CC=C4(C)C(C5=CCCN)CC5)CC6=CNC=7C=CC=C76

Isomeric SMILES
C[C@H]1C=2C(NC1)=CC=C2[C@H]3C(=O)N[C@@H](C[C@H]4C[C@@H](N)CC4)C(=O)N(C@@H)(C[C@@H](CO)C(=O)N)C[C@@H](CC5=CC=CC=C5)C(=O)N6[C@@H](C(=O)N)C[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	ATFPYW	-	-	-	-
---	--------	---	---	---	---

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

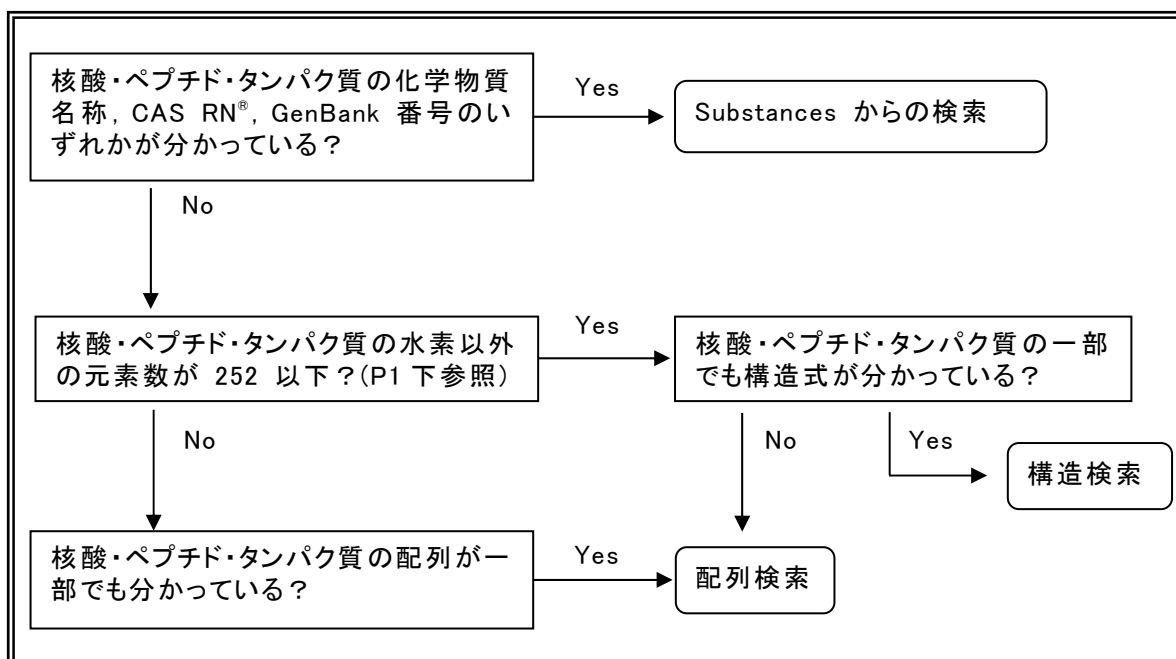


登録ルールの例外

- ・ 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。

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)" using the "Chemical Substance Name" (化学物質名称) field. The second screenshot shows a search for "1448554-62-8" using the "CAS Registration Number (CAS RN)" (CAS 登録番号 (CAS RN)) field. The third screenshot shows a search for "AAH30696" using the "GenBank Number" (GenBank 番号) field, with a dropdown menu highlighting "GenBank AAH30696" and a red arrow pointing to it with the text "この番号を選択" (Select this number).

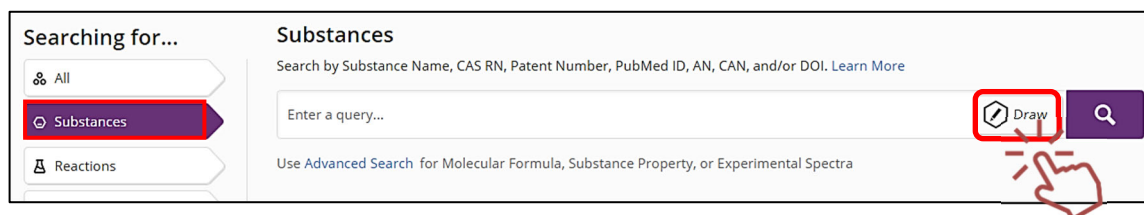


GenBank 番号の入ルール

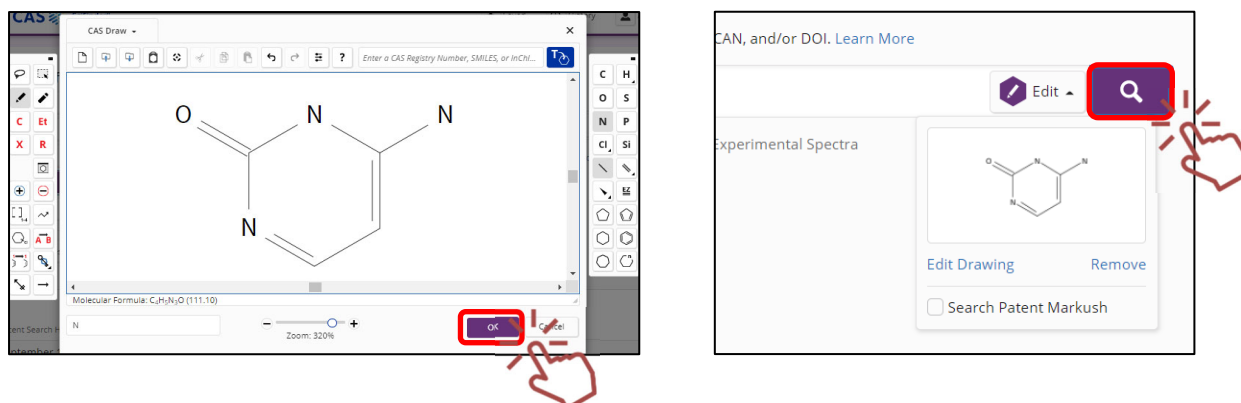
GenBank 番号から検索する場合は番号の前に「GenBank」を入力して検索します。検索ボックスに GenBank 番号を入力すると、オートサジェスト機能で「GenBank」の文字を含む候補が表示されるので、検索したい番号を選択してください。

■ 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、Filter by (選択した項目に限定) から Substance Class (物質の種類) のフィルター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。

Structure Match

- As Drawn (454)
- Substructure (169K)
- Similarity (5,778)

Analyze Structure Precision

Chemscape Analysis

Visually explore structure similarity with a powerful new tool. Learn more about Chemscape.

Create Chemscape Analysis

Filter Behavior

- Filter by
- Exclude

Commercial Availability

Reaction Role

Reference Role

Stereochemistry

Number of Components

Substance Class

- Organic/Inorganic Small Molecule (161K)
- Salt and Compound With (5,172)
- Coordination Compound (1,781)
- Incompletely Defined Substance (220)
- Polymer (124)
- Mixture (110)
- Radical Ion (56)
- Protein/Peptide Sequence (25)
- Nucleic Acid Sequence (16)
- General Derivative (11)
- Mineral (1)

Substances (41)

Sort: Relevance View: Partial

Filtering: Substance Class: 2 Selected

1 866826-92-8

2 1130595-80-0

3 391612-23-0

4 391612-22-9

5 391612-21-8

6 391612-26-3

物質の種類による限定

■ 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 配列質問式を選択します。
- ④ 回答の配列を選択します。
- ⑤ 回答の上限数を選択します。
- ⑥ (任意) パラメータを設定します。
- ⑦ 検索を実行します。

The screenshot shows the Biosequences search page. On the left, a sidebar lists search categories, with 'Biosequences' highlighted. The main area is titled 'Biosequences' and contains a search bar with the text 'gcgtttgctcttcttcttgcg'. Below the search bar are several configuration options: 'Sequence Identity %' (80), 'Match with Gaps?' (No), 'Gap Costs' (Existence 5 Extension 2), 'Query Coverage %' (90), 'Word Size' (11), 'Reward for Match, Penalty for Mismatch' (2, -3), 'BLAST Algorithm' (BLASTn), 'E-Value' (10), and 'Exclude Low Complexity Regions' (No). On the right side, there are additional settings: 'Sequence Type' (Nucleotide), 'Search Within' (Nucleotides), and 'Limit Total Sequence Results to' (100). A red dashed box highlights the search parameters section, and a red hand icon points to the 'Start Biosequence Search' button.

- 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 CAS SciFinder logo and a search bar. The main content area is titled 'Search History (375)' and shows a list of search results. The first result is dated 'September 2, 2021' and has a time of '3:01 PM'. It is categorized as 'Biosequences' and shows details: 'Sequence Type: Nucleotide', 'Search Within: Nucleotides', 'BLAST Algorithm: BLASTn', 'Alignment Identity: 80%', and 'Query Coverage: 90%'. A red hand icon points to the 'View Results' button, which is highlighted with a red box. Below it are '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。

このスクリーンショットは、Biosequences の検索結果ページを示しています。左側の「検索条件」パネルには、E-Value、Query Coverage %、Subject Coverage %、Alignment Identity % のフィルターが設定されています。中央には「アライメントの概略図」があり、質問配列と回答配列の一致部分を視覚的に示しています。右側の「References」タブには、関連する文献へのリンクが提供されています。また、「絞り込み (フィルター)」ボックスには、E-Value (期待値) * Query Coverage % * Subject Coverage % * Alignment Identity % の組み合わせが推奨されています。

Callouts (呼び出し):

- 検索条件
- 絞り込み (フィルター)
- E-Value (期待値) * Query Coverage % * Subject Coverage % * Alignment Identity %
- ヒットした回答配列の記載特許をまとめて検索
- 回答の並び替え
- エクスポート
- 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)
- アライメントの類似性*
- 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*
- 参考文献 (特許もしくは非特許) の検索
- 参考文献

* 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
Alignment Identity % (アライメントの類似性)	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します。
- ④ 検索を実行します。

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 Upload Sequence Clear Search

CDR1 RASGGIRNYLA X

CDR2 AASTLQ8 X

CDR3 QRYNRAPYT X

Limit Total Sequence Results to: 20000

Start Biosequence Search

- 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。

ヒットした回答配列の記載特許をまとめて検索

Biosequences (20,000)

Sort: Alignment Identity View: Collapsed

References

Query Details [View More](#)

151

Alignment Identity: 100%

Matches: 9 Mismatches: 0

Subject 1 107

View Less

Alignment Subject References

Alignment Data

BLAST Score: 67

E-Value: 0.0362696

CDR1 2 ASGGIRNYL 10

|||||

S 25 ASGGIRNYL 33

References

152

Alignment Identity: 100%

Matches: 27 Mismatches: 0

Subject 1 107

CDR1 11 CDR2 7 CDR3 9

CDR 配列検索では図中の特定のエリアの回答に限定できる

回答の並び替え

アライメントの類似性

アライメントの概略図

Subject タブ: 回答配列の詳細情報

References タブ: 出典文献情報

アライメント タブ: アライメント・スコア値・期待値

記載文献(特許もしくは非特許)の検索

■ 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) パラメータを設定します。
- ④ 配列質問式を選択選択します
- ⑤ 回答の上限数を選択します。
- ⑥ 検索を実行します。

The screenshot shows the Biosequences search page. On the left is a navigation menu with 'Biosequences' selected. The main area has 'Motif' selected in the search type dropdown (1). The search input field contains the motif '[L][ALKW]G[FL][VI]D[AG]DG' (2). Below the input are 'Query Coverage %' (90) and 'E-Value' (10) dropdowns (3). On the right, 'Sequence Type' is set to 'Protein' (4), and 'Limit Total Sequence Results to:' is set to '20000' (5). A 'Start Biosequence Search' button is highlighted with a hand cursor (6).

- 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。

The screenshot shows the search results for the motif. The page title is 'Biosequences (4,737)'. The search parameters are: Sequence Type: Protein, Query Coverage: 90%, E-Value: 10. The results list shows 'Seq 1: 1 LAGFVDADG 9'. An alignment view is shown with 'Query' and 'Subject' sequences. The alignment identity is 100%. The alignment data shows a BLAST Score of 63 and an E-Value of 0.429846. The alignment is:

Q	1	LAGFVDADG 9
S	13	LAGFVDADG 21

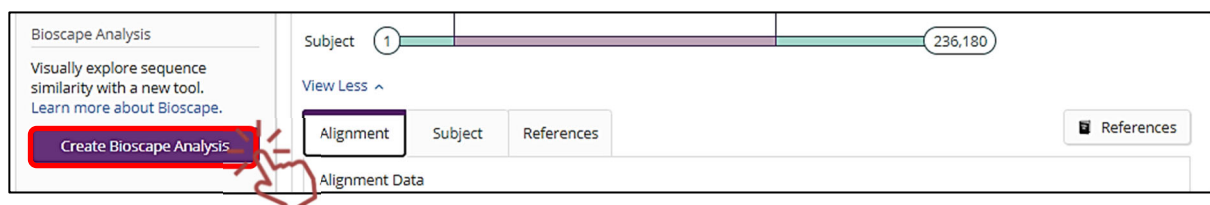
 Annotations include: 'ヒットした回答配列の記載特許をまとめて検索' (Search for patents on hit answer sequences), '回答の並び替え' (Sort answers), '代替残基を含めた検索結果の場合は、Query Details の配列をクリックして表示されるリストから別の配列を選択することができます' (If search results include alternative residues, you can click on the sequence in the list shown in Query Details to select a different sequence), 'アライメントの概略図' (Alignment overview diagram), 'アライメントの類似性' (Alignment similarity), '記載文献(特許もしくは非特許)の検索' (Search for cited literature (patent or non-patent)), 'Subject タブ: 回答配列の詳細情報' (Subject tab: Detailed information of the answer sequence), '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 expected value).

■ 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。

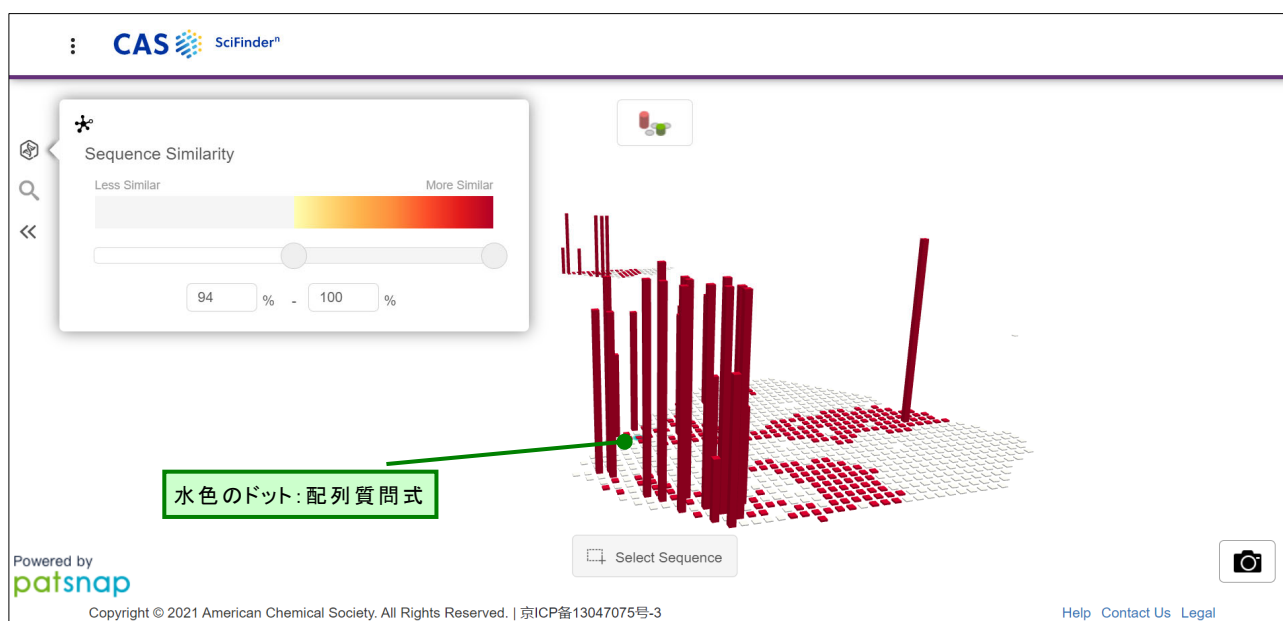
Bioscape 解析マップ作成手順

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

【範囲の選択】

中央下部にある Select Sequence をクリックすると、ポインタが範囲指定用のツールに変わり、マウスをドラッグすることでその範囲に含まれる配列をまとめて選択することができます。選択された配列と関連特許の件数が左のボックス内に表示されます。

また、画面右下には View Patent のリンクが表示されます。このリンクから、選択した配列の特許をまとめて表示することができます。

範囲選択を終了するには Exit をクリックします。

The screenshot illustrates the workflow for selecting a sequence range in CAS SciFinder. It shows the 'Selected (7)' list, the 'More Similar' slider, the 'Select Sequence' tool, and the 'View Patents' button. The 'Patent Patents Sequence Length' window displays the selected sequence length (9388) and the sequence '1258564-05-4'. The 'Substance Detail' window shows the CAS Registry Number (1258564-05-4) and the substance name 'Unspecified'. The 'References (3)' window shows the patent information for 'Recombinant vaccines comprising genetically stable more antigenic portions of cytomegalovirus protein II'.

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 12 件存在する

同主鎖の配列を持つ物質

Sequence: Linear
1 MRKSPGLSDC LKAWILLIST LTGRSYGQPS LQDELKQNTT VFTRILDRL
51 DGYDNRLRPG LGERVTEVKT DIFVTSFGPV SDHMEYITD VFFRQSHKDE
101 RLKFKGPMIV LRLNLMASK IWTPTFFHN GKSVAHNMT MFKLLRITTE
151 DGTLLYIMRL TVRAECPMHL EDFPMDAHC PLKFGSYAVT RAEVYVETR
201 EPARSYVVAE DGSRLNOYDL LGGTYDGGIV QGSTGEYVVM TTHFLKRIK1
251 GYFV10ITVLP CIMTVLISGV SFWLNRESVP ARTVFGYTTV LTMITLSISA
301 RNSLPKIVATA TAMDFIAYC YAFVFSALIE FATWVFTKR GYAWDGKSVV
351 PEKPKKVKDP LJKMNTYAP TATSYPNLA RGDPLATIA KSAATIEPKEV
401 KPETHKPEPK KTFNSYSKID RLSRIAFPLL FGFNLVYVA TYLNRPEPK
451 APTPHQ - - - -