



- ・ Substances とCAS Sequences の違い
- Substances 検索
- ・CAS Sequences 検索
- Bioscape Analysis





## SUBSTANCES と CAS SEQUENCES の違い

3 © 2025 American Chemical Society. All rights reserved.

## Substances と CAS Sequences

CAS SciFinder では様々な角度から核酸・タンパク質を検索できる

◇ All 😰 Substances 🛛 Reactions 🗊 References 沖 Suppliers	
Substance 検索 Patent Number, PubMed ID, AN, CAN, and/or DOI.	Ø Draw Q
-  Molecular Formula	
Examples: C6H6   (C8H8)x   C22H26CuN2O5.C2H3N	
+ Add Advanced Search Field CAS Sequences*	
Retrosynthetic Analysis Search CAS Lexicon Search CAS Sequences	
Make reaction plans with conditions, yields, catalysts, and experimental procedures. Build powerful searches using CAS concepts, chemical classes. and taxonomy. Query BLAST, CDR, and Motif algorithms for nucleotide and protein based sequences.	

\* CAS Sequences へのアクセス条件は、契約により異なります。詳しくはお問合せください。





#### Substances と CAS Sequences の違い

	Substances	CAS Sequences
検索可能な 項目	名称、化学構造、CAS 登録番号 (CAS RN®) 分子式、GenBank 番号 など	配列情報
	<ul> <li>CAS が独自のルールに従って収集した配列</li> <li>CAS References に収録された</li> <li>特許・非特許に記載の重要な配列</li> <li>GenBank に登録された配列</li> </ul>	<ul> <li>CAS が独自のルールに従って収集した配列</li> <li>CAS References に収録された</li> <li>特許・非特許に記載の重要な配列</li> <li>GenBank に登録された配列</li> </ul>
収録	<b>配列情報を持たないタンパク質のレコード</b> ・新規または新しい知見があったタンパク質	主要国を中心とした 73 カ国の特許発行 機関の特許から抽出した配列
	CAS RN <sup>®</sup> が付与された化学物質 ・ 有機化合物、無機化合物、タンパク質、 核酸、ポリマー、配位化合物、金属、 有機金属化合物、合金、鉱物、元素	NCBI 由来の配列
© 2025 America	n Chemical Society. All rights reserved.	CAS

## 得られる核酸・タンパク質情報の違い

CAS Sequences と Substances で得られる核酸・タンパク質の情報は異なる



## SUBSTANCES 検索

7 © 2025 American Chemical Society. All rights reserved.

#### 核酸・ペプチド・タンパク質の登録

物質情報は CAS RN<sup>®</sup>で管理、1物質に1つの CAS RN<sup>®</sup>を付与

#### 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

#### GenBank 由来の配列は、1 GenBank 番号につき1レコード(物質)として収録される

・文献・特許由来の配列と同じ場合は、一つのレコードにまとまる



CAS

# 核酸の登録例(1/2)

核酸: 塩基配列9以上



## 核酸の登録例(2/2)

#### 核酸: 塩基配列 9 以上 ^ Other Names and Identifiers 1 Other Name for this Substance 化学物質名 1: PN: WO2017140839 SEQID: 2 claimed DNA Sequence Details Sequence: DNA: linear 配列 catgttcgat agatgtacgc gctttgacaa 1 gaggcacgat tttgacatac 51 tacttgagca gtcggcagat ataggatgtt gcaagctccg tgagtcccac 101 aaaccaaaaa cctcg 特徴表 Sequence Modifications Location Description Туре modified base cytidine-1 5'-phosphate Patent Annotations Source: Not Given Reference: WO2017140839, SEQID 2; claimed 特許情報 (特許番号と記載位置) Additional Details

CASS Advision of the American Chemical Society

## ペプチド・タンパク質の登録例(1/2)

ペプチド・タンパク質:アミノ酸残基4以上



11 © 2025 American Chemical Society. All rights reserved.

# ペプチド・タンパク質の登録例(2/2)

ペプチド・タンパク質:アミノ酸残基4以上

equence Length: 6 ●	 鎖の D検索	Expand All   Collapse Al
Other Names and Identifiers		
Sequence Details		
Sequence: cyclic	•	配列
1 ATFPW	特徴	·····································
Туре	Location	Description
stereo	tryptophan-6	D
Patent Annotations		
Source: Not Given Reference: US6268342, SEQID 80; claimed	特許情報 (特許番号と記載位置)	
Source: Not Given		



CAS

## Related Sequences ボタン

同主鎖の核酸・タンパク質も検索する



13 © 2025 American Chemical Society. All rights reserved.

### 検索方法

- 化学物質名称
- CAS RN<sup>®</sup>
- Gen Bank 番号
- ・化学構造(構成元素数が252以下の場合のみ実行可能)



#### 名称からの検索-検索画面

例:Catalase について生物種を問わずに検索する

#### ・ ワイルドカード (\*)を使って前方一致検索をすると生物種を問わずに検索できる



### 名称からの検索 - 回答一覧

#### Substances には配列情報を持つレコードと持たないレコードがある



CAS

16

## タンパク質の物質レコード例

Substances には配列情報を持つレコードと持たないレコードがある



17 © 2025 American Chemical Society. All rights reserved.



(例) ヒト肝臓由来カタラーゼに関する文献の索引





## GenBank 番号からの検索

オートサジェストを使って検索する

- GenBank 番号は "GenBank" から始まる
- ・ GenBank 番号から検索する場合は番号の前に "GenBank" を入力して検索する



### 化学構造からの検索 – 検索画面

例:下記の構造を含む核酸・ペプチド・タンパク質を検索する



- ・水素以外の元素数が252を超える物質は、構造データが収録されないため、構造式から検索できない
- 9 塩基以上の核酸で、水素以外の元素数が 252 以下であっても、文献に構造が明確に 与えられていないときは構造が収録されないので、構造検索ではヒットしない このような核酸は名称検索や、配列検索など他の方法で検索する



## 化学構造からの検索 – 回答一覧、絞り込み

Substance Class(物質の種類)フィルターの活用



21 © 2025 American Chemical Society. All rights reserved.



## CAS SEQUENCES 検索

#### 配列検索とは

塩基、アミノ酸の配列を比較して類似性を見つける





2つの配列を比較するため、対応づけて並べることをアラインメントという







### CAS Sequences の検索

配列質問式を使用したBLAST、CDR、Motifの3つの検索が可能



25 © 2025 American Chemical Society. All rights reserved.

### BLAST 配列検索

配列ホモロジー検索でよく使われている NCBI のプログラム

・BLAST (Basic Local Alignment Search Tool) の名が示す通り局所的に類似した配列を検索





### BLAST 配列検索 – 検索画面

例:核酸医薬品 Fomivirsen の類似配列を検索する



#### スライド 27

ATO 他の検索のスライドも同じなのですが、 NCBI は必ず含めるようにするのが前提ですか?含めた方がよいとは思うのですが、 普通なら、NCBI 由来の配列を回答に含める場合はチェックを入れる になります。 Ayako TAZAWA, 2025-03-26T09:38:56.097

### BLAST 配列検索 - 回答一覧

例:核酸医薬品 Fornivirsen の類似配列を検索する

めて検索 BLAST Search Details	Query Details gogtttgotottottottgog View More	Sort : 回答の	並べ替え
Sequence Type: Nucleot Search Within: Nucleoti	ide 1,006 Results	Sort: Alignment Identity - View: Expanded -	
BLAST Algorithm: BLAST NCBI Included: Yes Alignment Identity: 80%	n 🗌 1	Alignment Identity: 100%	
検索条件 Watch wire Gaps?: No Gap Cott: Existence 5 Extension 2 Word Size: 11	Query ()	(21) Matches: 21 Mismatches: 0	
Bioscape Analysis	View Less V		献の検索
similarity with a new too	Alignment Subject References	References	
Create Bioscape An	Alignment Data BLAST Score: 42 E-Value: 0.0127524		
Filter by	0 1 GCGTTTGCTC TTCTTCTTGC 6 21		
✓ E-Value			
交り込み Vuery Coverage %			
イルター) Subject Coverage %			
✓ Alignment Identity 9	6 2	Alignment Identity: 100%	
✓ Sequence Length		-	
	Query (1)	(21)	

### BLAST 配列検索 - 回答一覧 (アラインメント)

例:核酸医薬品 Fomivirsen の類似配列を検索する



### BLAST 配列検索 - 回答一覧 (フィルター)

#### 例:核酸医薬品 Fomivirsen の類似配列を検索する

Filter by		E-Value (期待値)	Sequence Length
<ul> <li>E-Value</li> <li>to 10<sup>6</sup></li> <li>Query Coverage %</li> </ul>		<ul> <li>データベース中の配列に対して</li> <li>マッチする際の統計的有意性の閾値</li> </ul>	・配列長
<b></b> C		・E-Value が小さいほど、偶然には	Organisms
0 to 100	Sequence Length	起こりえないことを示している	・生物種 (NCBI 由来に限定)
0 to 100	10 to \$22207	Query Coverage %	
▲ Alignment Identity %		・類似領域 ÷配列質問式 × 100	Alignment Identity %
o	Organisms		• 一致したコード÷類
0 to 100	(406)	Subject Coverage %	似領域 × 100
	unidentified (66)	• 類似領域:回答配列 × 100	
	Homo sapiens (37)	发展模块·白白能/引入100	
	Gallus gallus (25)		

30 © 2025 American Chemical Society. All rights reserved.

View All

# (参考)フィルターの計算値

様々なフィルターによる絞り込みが可能



項目	内容	上記の例
Alignment Identity % (Sequence Identity%)	一致したコード ÷ 類似領域 × 100	$9 \div 10 \times 100 = 90\%$
Query Coverage %	類似領域 ÷ 配列質問式 × 100	$10 \div 10 \times 100 = 100\%$
Subject Coverage %	類似領域÷回答配列×100	$10 \div 11 \times 100 = 90.91\%$



CAS

# (参考) BLAST 配列検索 – パラメーター (1/3)

選択した質問式の配列、回答の配列によって自動的にパラメータが設定される



## (参考) BLAST 配列検索 – パラメーター (2/3)

選択した質問式の配列、回答の配列によって自動的にパラメータが設定される

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より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



## (参考) BLAST 配列検索 – パラメーター (3/3)

選択した質問式の配列、回答の配列によって自動的にパラメータが設定される

- 5 ギャップを考慮してマッチさせる / しない
- 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 Expand 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

\* 低複雑度領域:同じアミノ酸の連続や、少数個のアミノ酸からなる単調な配列を含む領域 低複雑度領域を含む配列で BLAST 検索を行うと非常に高いスコアになる傾向にあるが、 統計的に有意であっても、生物学的には類似度は高くない。 低複雑度領域のマスクフィルタリングを行うと、生物学的に無意味なアライメントは取り除かれる。 網羅的な検索が必要な特許調査では低複雑度領域マスクフィルタリングは使用しない方が良い。

34 © 2025 American Chemical Society. All rights reserved.

## (参考) 短い配列の検索

短い配列の検索で回答が得られない場合は、期待値を大きくし Word Size を小さくする

#### Advanced Sequence Search 🔨

Alignment Identity % 🧕	Match with Gaps?		
80	🔾 Yes 🔍 No		
Query Coverage % 🧕	Word Size 🥥		
90	11 •		
BLAST Algorithm	E-Value 🛛		
BLASTn -	10 -		

#### E-Value (期待値)

- ・短い配列質問式では回答の期待値が大きくなる傾向にある
- ・ 回答が得られない場合は期待値を大きくする

#### Word Size

- ・ 配列を固定長の断片 (Word) に区切り、Word 単位で類似する 断片を検索する
- ・ Word Size を小さくすることで類似領域の検出感度は上昇する
- \* 適切な値は検索の目的と対象データベースによって変わるため 推奨値はない



### CDR 配列検索

CDR (相補性決定領域)を指定して検索するプログラム

#### CDR 配列検索とは

- ・ BLAST をベースにしている
- 検索対象をあらかじめ抗体やT細胞受容体 といった CDR を持つ配列に限定している

#### 抗体とは

- ・抗原と特異的に結合して生体内から除去する ためのタンパク質
- 抗体は2本の軽鎖(L鎖)と重鎖(H鎖)から成る
- CDRは抗原結合部位であり、L 鎖とH 鎖に それぞれ CDR1、CDR2、CDR3 がある
- ・抗原ごとに異なる配列を持つ

36 © 2025 American Chemical Society. All rights reserved.





#### CDR 配列検索 - 検索画面

例:アダリムマブの軽鎖の CRD の類似配列を検索する



## CDR 配列検索 – 回答一覧

例:アダリムマブの軽鎖の CRD の類似配列を検索する



## CDR 配列検索 – 回答一覧 (アラインメント)

例:アダリムマブの軽鎖の CRD の類似配列を検索する 701 Alignment Identity: 92.59% **Alignment Identity** (アラインメントの類似性) 類似領域 CDR1 (1)(11) CDR3 致したコード÷類似領域×100 致:25 不 -致:2 Matches: 25 Matches: 2 Subject 1 214 References Alignment Subject References Subject タブ CAS Registry Number: 1170740-98-3 Length: 214 aa 回答配列の詳細情報 ns: synthetic construct, unidentified View Less N Sequence DIGMTGSPSS LSASVGDRVT ITCRASGGIR NYLAWYQQKP GKAPKLLIYA ASTLQSGVPS RFSGSGSGTD FTLTISSLOP Alignment Subject References 81 EDVATYYCOR YNMPPYTEGO GTKVEIKRTV AAPSVEIEPP SDEOLKSGTA SVVCLLNNEY PREAKVOWKV DNALOSGNSO 161 ESVTEODSKD STYSLSSTLT LSKADYEKHK VYACEVTHOG LSSPVTKSFN RGEC Alignment Data BLAST Score: 60 Subject Alignment E-Value: 18.140627386 TNFaに対する抗体 CDR1 1 RASOGIRNYL & 11 ゲゼルシャフト Assignee: パイエルインテレク: JP5570998 B2 | Seq ID No: 124 CDR1の質問配列1~11は、配列長214 S 24 RASQGIRNYL A 34 antihodies to tnf.alpha の回答配列の 24~34 で完全一致 1 AASTLOS 7 CDR2 Assignee: BAYER INTELLECTUA CA2710779 C | Seq ID No: 124 CDR2の質問配列1~7は回答配列の s 50 AASTLQS 56 References タブ Antibodies to TNF alpha 50~56 で完全一致 1 QRYNRAPYT 9 CDR3 Assignee: BAYER INTELLECTUAL EP2225276 A1 | Seq ID No: 124 出典文献情報 1111 111 CDR3の質問配列1~9は回答配列の CAS S 89 QRYNMPPYT 97 89~97で類似(不一致あり) © 2025 American Chemical Society. All rights reserved.

### Motif 配列検索

DNA、RNA、 タンパク質中の短いパターン配列を検索するプログラム

#### Motif 配列検索とは

- ・ BLAST をベースにしている
- •[]などの記号を利用した検索が可能で、 数パターンの配列を一度に検索できる

#### 例:[LI]AGFVD[AG]DG

- LAGFVDADG LAGFVDGDG
- IAGFVDADG
   IAGFVDGDG

#### Motif とは

・機能的に重要、立体構造と関連する 保存配列パターンを Motif という

40 © 2025 American Chemical Society. All rights reserved.





#### Motif 配列検索 - 検索画面

例:下記の代替残基を含めた Motif 配列を検索する



# **BIOSCAPE ANALYSIS**

© 2025 American Chemical Society. All rights reserved.

### Bioscape Analysis とは

配列の類似性と関連特許の件数で 3D マップを作成する機能

- ・ 検索に用いた配列質問式が水色のドットで示される
- ・ 類似性が高い配列が赤色、類似性が低い配列は黄色で示される







## **Bioscape Analysis 作成の流れ**



## Bioscape Analysis の見方 (1/2)





CAS

## Bioscape Analysis の見方 (2/2)



## Bioscape Analysis の機能

Bioscape Analysisの機能はコントロールパネルで操作する





### **Sequence Similarity**

マップに反映する類似度を設定できる



# Search (1/2)

解析結果内の該当物質を検索できる





CAS

## **Search (2/2)**

検索を実行すると物質とヒットした関連特許の数が一覧で表示される



50 © 2025 American Chemical Society. All rights reserved.

### **Select Structure**







#### JAICI ヘルプデスク

0120-003-462 (平日 9:00-17:00)

support@jaici.or.jp

© 2025 American Chemical Society. All rights reserved.

## Thank you

#### Connect with us at cas.org



in linkedin.com/company/cas







