

CAS  **STNext[®]**
A division of the
American Chemical Society

CAS Sequences 検索ガイド

2022 年 03 月

JAICI
化学情報協会

* 目次 *

STNext Biosequences Search ガイド

STNext の Biosequences Search.....	1
検索初期画面	2
BLAST 配列検索	3
BLAST 配列検索の結果	4
検索結果の出力	8
(参考) Advanced Biosequence Search.....	9
CDR 配列検索.....	10
CDR 配列検索の結果.....	11
Motif 配列検索.....	12
Motif 配列検索の結果.....	13
他ファイルへのクロスオーバー.....	15
Bioscape Analysis (配列類似性 × 特許の解析).....	17



STNext Biosequences Search, Bioscape Analysis のご利用について

- ・ 包括契約 (STN IP Protection Suite) でご利用のお客様: すぐにご利用いただけます.
- ・ STN を定額でご利用のお客様: オプション契約でご利用いただけます.
- ・ STN を従量制でご利用のお客様: すぐにご利用いただけます

ご興味のある方は, 社内の契約担当の方にご相談の上, 化学情報協会 情報事業部にお問合せください.

STNext の Biosequences Search

■ Biosequence Search は 2021 年 2 月に追加された STNext の独自の配列検索機能です。

・収録している配列情報

- CAS が独自のルールに従い収集した REGISTRY ファイルの配列
- 7 つの主要特許発行機関 (US, EP, WO, CA, KR, JP, CN) の特許から抽出した配列 **New**
- NCBI 由来の配列 **New**

・配列検索プログラム

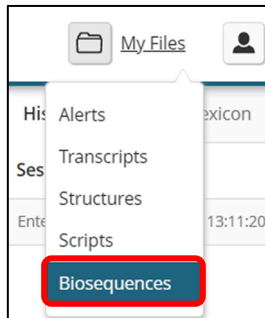
配列検索プログラム	内容
BLAST 配列検索	局所的に類似した配列を検索するプログラム
CDR 配列検索 New	抗体と T 細胞受容体の CDR を指定し、検索するプログラム
Motif 配列検索 New	DNA, RNA, タンパク質中の短いパターン配列を検索するプログラム

・特長

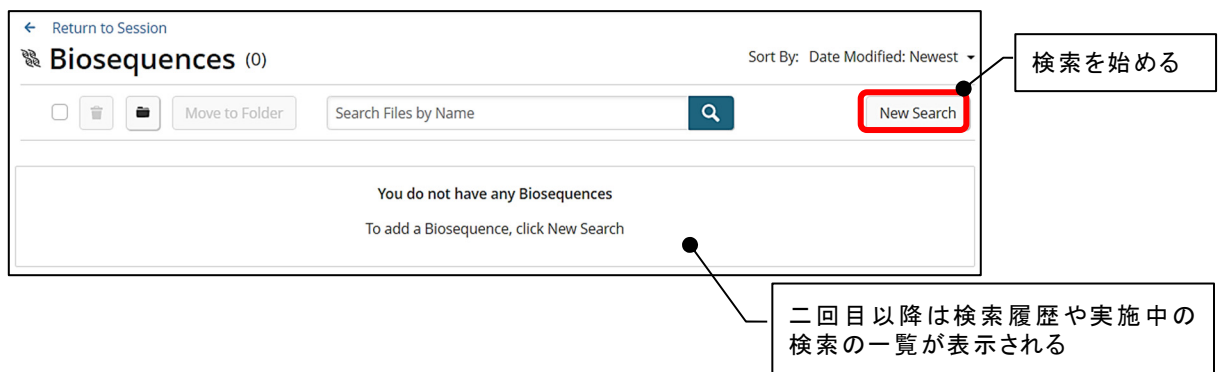
- アライメントの概要図をカラーでわかりやすく表示. Excel 形式でのダウンロードも可能.
- 絞り込みやソートが簡単にできる.
- 配列情報と共に配列の由来となった文献などの情報も表示.
- Bioscape Analysis 機能を使って類似配列と特許を解析できる.
- 配列検索用の特別なソフトウェアのインストール不要.

初期画面

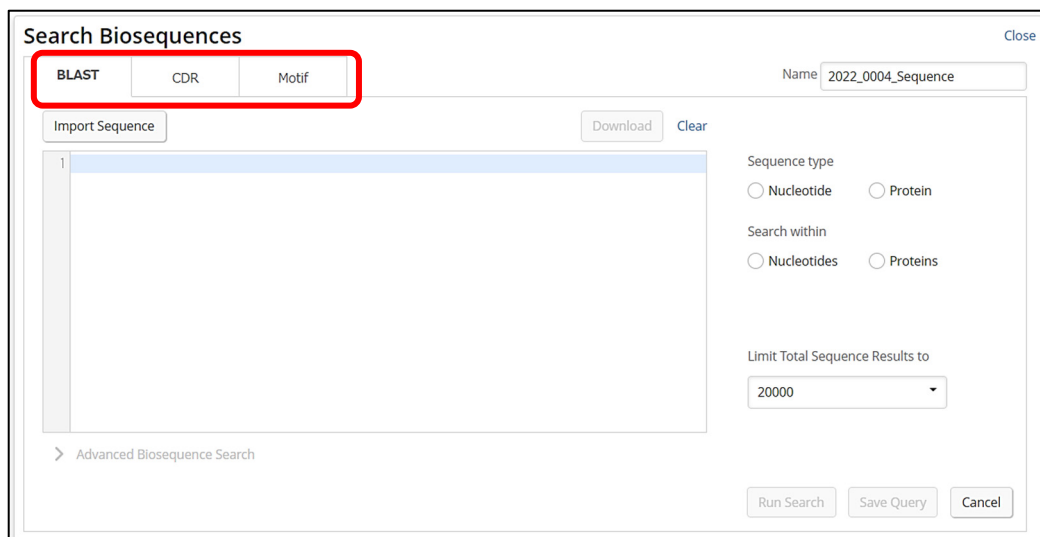
- 画面右上 My Files から Biosequences を選択します。



- 最初のアクセスでは何も表示されませんが、二回目以降は検索履歴や実施中の検索の一覧が表示されます。New Search をクリックし検索を始めます。



- 検索したい配列検索プログラムのタブを選びます。BLAST, CDR, Motif の 3 種類の配列検索プログラムを利用できます。



BLAST 配列検索

BLAST (Basic Local Alignment Search Tool) は、データベース中の配列の中から、局所的に類似した配列を高速に検索するプログラムです。

■ BLAST タブを選び、配列質問式を入力、あるいはアップロードします。次に、質問式の配列、回答の配列、回答の上限数を選択します。Run Search ボタンをクリックし、検索を実行します。

- ・ 選択した質問式・回答の配列タイプによって自動的にパラメータが設定されます。
- ・ パラメータを確認あるいは変更する場合は、Advanced Biosequence Search をクリックします。詳細は p.9 をご覧ください。

The screenshot shows the 'Search Biosequences' interface. The 'BLAST' tab is selected. A sequence is entered in the 'Import Sequence' field. The 'Sequence type' is set to 'Protein' and 'Search within' is set to 'Proteins'. The 'Limit Total Sequence Results to' is set to 20000. The 'Run Search' button is highlighted. Annotations include: '名前の変更 (任意)' pointing to the sequence name, '質問式の配列' pointing to the 'Protein' radio button, '回答の配列' pointing to the 'Proteins' radio button, '回答の上限数' pointing to the 'Limit Total Sequence Results to' dropdown, '検索の実行' pointing to the 'Run Search' button, and 'パラメータの変更 (任意)' pointing to the 'Advanced Biosequence Search' link. A text box explains: '配列を直接入力するか, Import Sequence から配列をアップロードする (.txt 形式)'.

■ Biosequences ページに戻り検索が始まります。画面右の Cancel Search 表示が、View Results 表示になったら、検索結果を確認できます。

The screenshot shows the 'Biosequences' page. The top part shows the search in progress with a 'Cancel Search' button. The bottom part shows the search completed with a 'View Results' button and '5226 results' displayed. Annotations include: '配列質問式や検索条件の確認' pointing to the 'View' button, and '検索結果の確認' pointing to the 'View Results' button.

BLAST 配列検索の結果

- ヒットした配列に対する配列質問式のアライメントの図や各項目の計算値と共に BLAST 配列検索の結果の一覧が表示されます。

- ・ アライメントは、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です。

絞り込みフィルター

ソート

Filter By

Query Identity %

Query Coverage

Subject Coverage

Subject Identity %

Organism

Apply Reset

Biosequence Search Results (5226)

Sort By: Alignment Identity %: Descending

Create Bioscape Analysis Get All Patent Numbers Show Search Details

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	260	100.00	100.00	100.00	7 View More
	231	100.00	88.85	100.00	10 View More
	232	100.00	88.85	99.57	5 View More
	230	100.00	88.46	100.00	0 View More
	226	100.00	86.92	100.00	6 View More
	695	100.00	85.77	32.09	0 View More
	1627	100.00	70.38	11.25	0 View More
	1627	100.00	70.38	11.25	0 View More
	240	100.00	58.85	63.75	2 View More
	1627	99.45	70.00	11.19	0 View More

Show 10 per page Previous 1 2 3 4 5 ... 523 Next

生物名による絞り込み (NCBI 由来の配列に収録)

ヒットした配列に対する配列質問式のアライメントの図

計算値

- ・ ヒットした配列に対する配列質問式のアライメントが Alignment Identity % により色分けされた図で表示されます。

(■ 99-100%, ■ 97-98.99%, ■ 95-96.99%, ■ 90-94.99%, ■ 80-89.99%, ■ 60-79.99%, ■ 0-59.99%)

- ・ 各項目の計算値による絞り込みやソートが可能です。

項目	内容	Filter By	Sort By
Alignment Identity %	一致したコード ÷ 類似領域 × 100	-	○ (降順がデフォルト)
Query Identity %	一致したコード ÷ 配列質問式 × 100	○	○
Query Coverage	類似領域 ÷ 配列質問式 × 100	○	○
Subject Identity %	一致したコード ÷ 回答配列 × 100	○	○
Subject Coverage	類似領域 ÷ 回答配列 × 100	○	○

■ View More ボタンをクリックすると、選択した配列について詳細を確認できます。

- ・ Alignment タブ : ヒットした配列に対する配列質問式のアライメントの詳細が表示されます。

(例) 260 配列長の配列質問式で検索してヒットしたタンパク質の回答

計算値

① $233/242 * 100 = 96.28\%$

② $233/260 * 100 = 89.62\%$

③ $233/1690 * 100 = 13.79\%$

ヒットした配列の全長
1690 配列

①

②

③

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	1690	96.28	89.62	13.79	12

[View Less ^](#)

Alignment Patents (10) Journals (2) Subject

配列に関する文献数

質問式の配列

ヒットした配列

Query: 19 260 242 aa

Subject: 1449 1690

類似していると判断された領域 242 配列

- ・ 質問式は 19~260 番目でヒット
- ・ ヒットした回答は 1449~1690 番目でヒット

完全に一致した配列は 233 配列 (9 アミノ酸が合致していない)

+ : 等価のアミノ酸でヒット

Query:	19 PLADYKDDDDKPGYLGFFLLVLHSQTDQPTCPLGMPRLWTGYSLLYLEG 68
	P+ D PGYLGFFLLVLHSQTDQPTCPLGMPRLWTGYSLLYLEG
Subject:	1449 PIGDPGPKGFGPGYLGFFLLVLHSQTDQPTCPLGMPRLWTGYSLLYLEG 1498
Query:	69 QKKAHNQDLGLAGSCLPVFSTLPPFAYCNIHQVCHYAQRNDRSYWLASAAP 118
	QKKAHNQDLGLAGSCLPVFSTLPPFAYCNIHQVCHYAQRNDRSYWLASAAP
Subject:	1499 QKKAHNQDLGLAGSCLPVFSTLPPFAYCNIHQVCHYAQRNDRSYWLASAAP 1548
Query:	119 LPMMLSEEAIRPYVSRCAVCEAPAQAVAVHSQDSIPPCPQTRSLWIG 168
	LPMMLSEEAIRPYVSRCAVCEAPAQAVAVHSQDSIPPCPQTRSLWIG
Subject:	1549 LPMMLSEEAIRPYVSRCAVCEAPAQAVAVHSQDSIPPCPQTRSLWIG 1598
Query:	169 YSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRGTCHFFANKY 218
	YSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRGTCHFFANKY
Subject:	1599 YSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRGTCHFFANKY 1648
Query:	219 SFWLTTVKADLQFSSAPADTLKESQAQRQKISRCQVCVKYS 260
	SFWLTTVKADLQFSSAPADTLKESQAQRQKISRCQVCVKYS
Subject:	1649 SFWLTTVKADLQFSSAPADTLKESQAQRQKISRCQVCVKYS 1690

- Patents タブ : ヒットした配列の由来となった特許が公報単位で表示されます。
- 下記のボタンをクリックすると、ヒットした配列の由来となった特許番号を各種特許データベースで検索できます (p.15 参照)。

Biosequence Search Results (39) Sort By: Alignment Identity %: Descending

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	1690	96.28	89.62	13.79	12

Alignment | **Patents (10)** | Journals (2) | Subject

10 patents found.

1 of 10

Collagen iv replacement

Assignees: GOLDFINCH BIO, INC.
 Patent No.: EP3171889A1
 Sequence ID: 4

The present invention provides pharmaceutical compositions, formulations and methods for treating Alport syndrome by administering recombinant human collagen IV protein to a patient in need.

2 of 10

Polymorphisms in known genes associated with human disease, methods of detection

- Journals タブ : ヒットした配列の由来となった非特許文献が表示されます。
- 下記のボタンをクリックすると、ヒットした配列の由来となった非特許文献を CPlus ファイルあるいは MEDLINE ファイルで検索できます (p.16 参照)。

Alignment | Patents (10) | **Journals (2)** | Subject

2 journals found.

1 of 2

Complete primary structure of the human alpha4(IV) chain with structure and expression of the other alpha(IV) chains

Author Name: Leinonen, Anu; Mariyama, Mariko; Mochizuki, Toshio; Tryggvason, Karl; Reeders, Stephen
 Accession Number: 1995:75583
 PubMed ID: 7523402
 Source: Journal of Biological Chemistry, (1994), 269(42), 26172

The entire sequence of the human $\alpha 4(\text{IV})$ collagen chain was determined from cDNA clones and polymerase chain reaction-amplified DNAs. The complete translation product has 1690 amino acid residues and the processed $\alpha 4(\text{IV})$ chain proper 1652 residues. There is a 38-residue putative signal peptide, a 1421-residue collagenous domain starting with a 23-residue noncollagenous sequence, and a 231-residue NC1 domain. The gly-Xaa-Yaa-repeat sequence of the collagenous domain is interrupted at 26 locations by noncollagenous sequences of 1-12 residues in length. The $\alpha 4(\text{IV})$ chain contains 31 cysteine residues of which 18 are conserved in the other type IV collagen α chains. The calculated mol. weight of the mature $\alpha 4(\text{IV})$ chain is 164,123. Anal. of the primary structure showed that the $\alpha 4(\text{IV})$ chain belongs to the $\alpha 2$ -like type IV collagen chains together with $\alpha 2(\text{IV})$ and $\alpha 6(\text{IV})$. Northern analyses with RNA from several human fetal tissues revealed quite similar expression patterns for the $\alpha 4(\text{IV})$ and $\alpha 3(\text{IV})$ chains, but there were also distinct differences in some tissues. The expression patterns of $\alpha 5(\text{IV})$ and $\alpha 5(\text{IV})$ differed extensively between each other and they also differed from those of $\alpha 3(\text{IV})$ and $\alpha 4(\text{IV})$.

2 of 2

Determination of the genomic structure of the COL4A4 gene and of novel mutations

- Subject タブ : ヒットした配列の配列長, 配列の詳細, CAS RN®, GenBank Accession 番号が表示されます.
- 配列に該当する CAS RN® や GenBank Accession 番号がある場合は, それらの番号が表示されます. 各ボタンをクリックすると, 配列に関する CAS RN® や GenBank Accession 番号を検索できます (p.16 参照).

Alignment	Patents (10)	Journals (2)	Subject
Sequence Length: 1690 aa			
CAS Registry Number®: 791139-98-5, 159348-50-2, 459495-63-7, 480647-33-4			
<div style="border: 1px solid red; display: inline-block; padding: 2px;">Get All CAS Registry Numbers</div>			
GenBank Accession No.: CAA56943, Y17397, X81053, CAA76763			
<div style="border: 1px solid red; display: inline-block; padding: 2px;">Get Genbank Accession Nos.</div>			
Organism: Homo sapiens			
Sequence:			
<pre> 1 MWSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYVYVSG KKYIGPCGGR 51 DCSVCHCVPE KGSRGPPGPP GPQGPIGPLG APGPIGLSGE KGMRGDRGPP 101 GAAGDKGDKG PTGVPGFPGL DGIPGHPGPP GPRGKPGMSG HNGSRGDPGF 151 PGGRGALGPG GPLGHPGEGK EKGNSVFILG AVKGIQGDRG DPGLPGLPGS 201 WGAGGPAGPT GYPGEPGLVG PPGQPGRPGL KGNPGVGVKG QMGDPGEVQ 251 QGSPGPTLLV EPPDFCLYKG EKGIGIPGM VGLPGPPGRK GESGIGAKGE 301 KGIPGFFGPR GDPGSYGSPP FPGLKGELGL VGDPLGLGLI GPKGDGPNRG 351 HPGPPGVVLT PPLPLKGGPG DPGFPGRYGE TGDVGGPPGPP GLLGRPGEAC 401 AGMIGPPGPQ GFFGLPGLPG EAGIPGRPDS APGKPGKPGS PGLPGAPGLQ 451 GLPGSSVIYC SVGNPGPQGI KGKVGPPGGR GPKGEGKNEG LCACEPGPMG </pre>			
View More ▾			

STNext のコマンドラインに戻り, 配列に関する CAS RN® を検索する

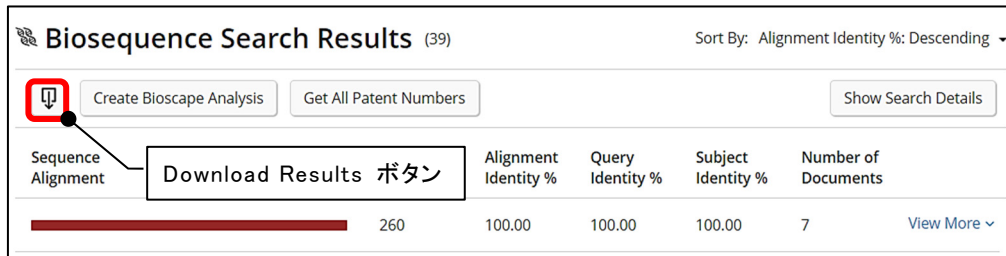
STNext のコマンドラインに戻り, 配列に関する GenBank Accession 番号を検索する

検索結果の出力

検索結果は、レポート作成や結果の共有に適した Excel 形式で出力できます。

■ Biosequence Search Results 画面の Download Results ボタンをクリックし、ダウンロードします。

・ダウンロードの上限は 1,000 件です。



■ Excel 形式ダウンロード例

・回答集合に項目に対応する情報がない場合、その項目が Excel に含まれません。



E	F	G	H	I	J	K	L	M
Number of GenBanks	GenBank Accession No.	Organism	Number of Patents	Patent No.	Sequence ID (Patent)	Number of Journals	Accession Number	Source
2	XF_034811187 XF_003821876	Fan paniscus	0				2012:941046	Nature (London, United Kingdom), (2012), 486(7404), 527-531
0			2	US46812339B1 US20070037165A1	7065 7065	0		
4	CA566943 Y17267 X11053 CAA76763	Homo sapiens	10	EP3171889A1 US6812339B1 CA2955493A1 US20070037165A1 US20070083333A1 WO2016014781A1 JP201753266A US2011018392A1 US7745391B2 US20180207240A1	4 5884 4 5884 758404 4 4 758404 758404 4		2:1995:75583 1998:785255	Journal of Biological Chemistry, (1994), 269(42), 26172 American Journal of Human Genetics (1998), 63(5), 1329-1340

A: アライメントの図
B: アライメントの配列
C: 配列長
D: CAS RN®

E: GenBbank Accession 番号の数
F: GenBbank Accession 番号
G: 生物名
H: 特許番号の数
I: 特許番号
J: Sequence ID (特許)
K: レコード番号の数 (CAplus)
L: レコード番号 (CAplus)
M: 収録源

・ Motif 配列検索 (p.12-14) で Combine Motif Results にチェックを入れた場合は、Number of Motif Child の項目が含まれます。

(参考) Advanced Biosequence Search

- 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ます. 確認あるいは変更する場合は, Advanced Biosequence Search をクリックします.

(例) blastp のパラメータ

① 検索タイプ

検索タイプ	検索機能	質問式配列	回答配列
BLASTn	塩基配列質問式に類似した塩基配列を検索	核酸	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	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		
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		
tBLASTn-fast	データベース中の塩基配列をアミノ酸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
BLASTx-fast	塩基配列質問式をアミノ酸配列に翻訳して, これに類似したアミノ酸配列を高速で検索	核酸	タンパク質

- ② Query Coverage % : 配列質問式に対する回答配列の類似領域の割合
- ③ ギャップを考慮してマッチさせる/させない
- ④ 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
- ⑤ 検索用の文字列の長さ ⑥ 低複雑度領域のマスクフィルタリングを使用する/しない
- ⑦ 置換行列マトリックス ⑧ Open Gap Cost と Extend Gap Cost の組み合わせ

■ 短い配列質問式を検索する場合のパラメータ設定

- ・ 30 より短いタンパク質の場合, BLASTp-short を選択します.
- ・ 50 より短い核酸の場合, BLASTn を選択し, 短い配列検索用のボックスにチェックします.
 - Automatically adjust parameters for short sequences (less than 50 nucleotides)
- ・ E-value の上限は 1,000 です.

CDR 配列検索

CDR (Complementarity determining region) 配列検索は、抗体と T 細胞受容体の CDR を指定し検索できるプログラムです。BLAST アルゴリズムをベースとしています。

- ・ CDR 配列検索では、検索対象を CDR 領域を持つ可能性のある配列に限定しています。

(例) アダリムマブの軽鎖の CDR

```
DIQMTQSPSSLSASVGDRTITTCRASQGIRNYLAWYQQKPKGKAPKLLIYAASSTLQSGVPSRFSGSGSGTDFLTISL  
QPEDVATYYCQRYNRAPYTFGGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNAL  
QSGNSQESVTEQDSKSTYLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
```

- CDR タブを選び、配列質問式を入力します。次に、回答の上限数を選択します。Run Search ボタンをクリックし、検索を実行します。

- ・ CDR 配列検索では、パラメータ設定はありません。

Search Biosequences

BLAST **CDR** Motif

Import Sequence Download Clear

Name 2022_0005_Sequence

Limit Total Sequence Results to 20000

> CDR 1 RASQGIRNYLA

> CDR 2 AASTLQS

> CDR 3 QRYNRAPYT

Run Search Save Query Cancel

名前の変更 (任意)

回答の上限数

検索の実行

- Biosequences ページに戻り検索が始まります。画面右の Cancel Search 表示が、View Results 表示になったら、検索結果を確認できます。

2022_0005_Sequence 2022 Mar 1 1:49 PM

> CDR 1 RASQGIRNYLA

> CDR 2 AASTLQS

> CDR 3

View

Cancel Search ...

Your search may take some time; while it is running, you may continue to use STNext.

2022_0005_Sequence 2022 Mar 1 1:50 PM

> CDR 1 RASQGIRNYLA

> CDR 2 AASTLQS

> CDR 3

View

View Results ...

20000 results

配列質問式や検索条件の確認

検索結果の確認

CDR 配列検索の結果

- CDR 配列検索の結果一覧が、ヒットした配列に対する配列質問式のアライメントの図と、各項目の計算値と共に表示されます。

- ・ CDR 配列検索の結果の場合、各項目の計算値の絞り込みだけでなく、ベン図の領域をクリックして絞り込むことが可能です。

The screenshot shows the 'Biosequence Search Results' page with 20,000 results. On the left, there are filter options for CDR Matching (Venn diagram), Query Identity %, Query Coverage, Subject Coverage, Subject Identity %, and Organism. The main table lists sequence alignments with columns for Sequence Alignment, Sequence Length, Alignment Identity %, Query Identity %, Subject Identity %, and Number of Documents. A 'Sort By: Alignment Identity %: Descending' dropdown is visible at the top right. Callouts point to the Venn diagram, numerical filters, organism list, alignment diagrams, and numerical values in the table.

ベン図による絞り込み

数値による絞り込み

生物名による絞り込み (NCBI 由来の配列に収録)

ヒットした配列に対する各配列質問式のアライメントの図

計算値

ソート

- View More ボタンをクリックすると、選択した配列について詳細を確認できます。

- ・ BLAST 配列検索の結果 (p.5-7) と同様に表示されますが、CDR の配列質問式ごとのアライメントが表示されます。

This detailed view shows the alignment for a specific sequence (214 aa). It highlights the alignment identity percentages (96.30, 96.30, 12.15) for different regions. Below, the alignment is shown for CDR 1, CDR 2, and CDR 3, with query and subject sequences and their respective positions. Callouts explain that multiple CDR regions are summed for calculation and that alignments are shown for each CDR region.

複数の CDR 配列質問式で検索した場合、すべての領域を合計して算出した計算値が表示される

ヒットした配列に対する各配列質問式のアライメントの図

配列質問式ごとにアライメントを表示

Motif 配列検索

Motif 配列検索は、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です。BLAST アルゴリズムをベースとしています。

(例)

代替残基を [] で指定した配列パターンの配列質問式 : [L][ALKV]G[FL][VI]D[AG]DG

■ Motif タブを選び、配列質問式を入力、あるいはアップロードします。次に、質問式の配列、回答の上限数を選択します。Run Search ボタンをクリックし、検索を実行します。

- ・ Motif 配列検索では、代替残基 [] などの記号を利用できます。100 質問式の組み合わせまで検索可能です。利用できる記号については、こちらの資料をご覧ください。

<https://www.cas.org/sites/default/files/documents/stnextmotif.pdf>

The screenshot shows the 'Search Biosequences' interface with the 'Motif' tab selected. Annotations include:

- 名前の変更 (任意)**: Points to the 'Name' field containing '2022_0004_Sequence'.
- 質問式の配列**: Points to the 'Sequence type' section where 'Protein' is selected.
- 回答表示のオプション**: Points to the 'Combine Motif Results' checkbox, which is checked.
- 回答の上限数**: Points to the 'Limit Total Sequence Results to' dropdown menu, set to '20000'.
- 検索の実行**: Points to the 'Run Search' button.
- パラメータの変更 (任意)**: Points to the 'Advanced Biosequence Search' section, with a note: 'Motif 配列検索のパラメータは Query coverage% と E-Value のみ変更できる'.

- ・ 回答表示のオプション Combine Motif Results のチェックを入れると、代替残基などを使った配列質問式の場合でも、全ての組み合わせの回答をまとめて表示します。チェックをはずすと、生じた配列質問式ごとに回答が表示されます。

■ Biosequences ページに戻り検索が始まります。画面右の Cancel Search 表示が、View Results 表示になったら、検索結果を確認できます。

The screenshot shows the search results page with two entries for '2022_0004_Sequence'. Annotations include:

- 配列質問式や検索条件の確認**: Points to the query sequence '[L][ALKV]G[FL][VI]D[AG]DG' and the 'View' button.
- 検索結果の確認**: Points to the 'View Results' button and the text '8320 results'.
- Combine Motif Results のチェックをはずした時の表示**: Points to a second entry showing '64 of 64 searches complete' and '30800 results'.

Motif 配列検索の結果

■ Motif 配列検索の結果一覧が、ヒットした配列に対する配列質問式のアライメントの図と、各項目の計算値と共に表示されます。

- ・ Combine Motif Results にチェックを入れた場合、代替残基などを使った配列質問式の場合でも、全ての組み合わせの回答をまとめて表示します（デフォルト）。

絞込みフィルター

絞込みフィルター

ソート

生物名による絞り込み (NCBI 由来の配列に収録)

ヒットした配列に対する各配列質問式のアライメントの図

計算値

- ・ Combine Motif Results のチェックをはずした場合、生じた配列質問式ごとに表示します。プルダウンで選択して、回答を確認します。

生じた配列質問式ごとの表示

■ View More ボタンをクリックすると、選択した配列について詳細を確認できます。

- ・ BLAST 配列検索の結果 (p.5-7) と同様に表示されますが、Combine Motif Results にチェックを入れた場合で、複数のヒットがあると、生じた配列ごとのアライメントが表示されます。

ヒットした配列質問式のパターン数

93.18 91.11 82.00 1 View Less ^

Alignment (5) Patents (1) Journals (0) Subject

Query 2 9 8 aa

Subject 3 10 10 aa

Query: 2 AGFVDGDG 9
AGFVDGDG
Subject: 3 AGFVDGDG 10

Query: 1 IAGFVDGDG 9
IAGFVD DG
Subject: 2 IAGFVDGDG 10

Query: 1 IAGFVDGDG 9
IAGFVDGDG
Subject: 2 IAGFVDGDG 10

Query: 1 IAGFVDGDG 9
IAGFVDGDG
Subject: 2 IAGFVDGDG 10

Query: 1 IAGFVDGDG 9
IAGFVDGDG
Subject: 2 IAGFVDGDG 10

複数ヒットした場合、平均の計算値が表示される

本例の質問式 ([LI][ALKV]G[FL][VI]D[AG]DG) から生じた配列質問式ごとのアライメント

他ファイルへのクロスオーバー

ヒットした配列の由来となった特許番号を STN の各種特許データベースで検索することができます。

■ Patents タブ : ヒットした配列の由来となった特許の特許番号を抽出できます。

- 抽出できる特許番号の上限は 5,000 件です。

The screenshot shows the 'Biosequence Search Results' interface. At the top, there is a 'Get All Patent Numbers' button highlighted with a red box and a callout box that says '全回答の特許番号を検索する'. Below this is a table with columns: Sequence Alignment, Sequence Length (1690), Alignment Identity % (96.28), Query Identity % (89.62), Subject Identity % (13.79), and Number of Documents (12). The 'Patents (10)' tab is selected. Below the table, there is another 'Get All Patent Numbers' button highlighted with a red box and a callout box that says '表示している配列に関する特許番号を検索する'. The first patent entry is shown: 'Collagen iv replacement', Assignees: GOLDFINCH BIO, INC., Patent No.: EP3171889A1, Sequence ID: 4. Below this entry is a 'Get Patent Number' button highlighted with a red box and a callout box that says 'この特許番号を検索する'.

- ボタンをクリックし、検索したいデータベースを選択後 Continue ボタンをクリックすると、コマンドラインに移動し検索が実行されます。

The screenshot shows the 'Get Patent Numbers from STN File' dialog box. It has a 'STN File' section with radio buttons for various databases: AUPATFULL, GBFULL, PCTFULL, CANPATFULL, HCAPLUS, PCTGEN, CAPLUS, IFIALL, USGENE, CNFULL, INFULL, USPATFULL, DEFULL, INPADOCDB, WPINDEX (selected), DGENE, INPAFAMDB, WPIX, EPFULL, JPFULL, ZCAPLUS, FRFULL, and KRFULL. There is a 'Save Script' checkbox checked, with a callout box that says 'Save Script にチェックを入れると、検索式のスク립トが My Files - Scripts に保存される'. Below this is a 'Script Name' field containing '2022_0007_Script' and a 'Continue' button highlighted with a red box. A large blue arrow points from the 'Continue' button to the STNNext interface. The STNNext interface shows a command line with the following text: 'Transcript ON 2022_0102_Transcript', 'File WPINDEX', and a list of patent numbers: '1 US20210130845A1/PN (US20210130845/PN)', '1 CN112771071A/PN (CN112771071/PN)', '1 US20210139557A1/PN (US20210139557/PN)', '1 KR20210068478A/PN (KR2021068478/PN)', '1 KR20210073520A/PN (KR2021073520/PN)', '1 JP2021516958A/PN (JP2021516958/PN)', '1 US11078247B2/PN (US11078247/PN)', '1 WO2021-181118A1/PN (WO2021181118/PN)'. A callout box points to this list with the text 'STN のコマンドラインに移動し、選択したファイルで検索が実行される'. The interface also shows a 'Submit' button and a 'Scripts' tab.

■ Journals タブ : ヒットした配列の由来となった非特許文献のレコード番号を抽出できます。

- 抽出できる番号の上限は 5,000 件です。ボタンをクリックし、検索したいデータベースを選択後 Continue ボタンをクリックすると、コマンドラインに移動し検索が実行されます。

表示している配列に関する文献を検索する

Get All Accession Numbers Get All PubMed IDs

2 Journals found.

1 of 2

Complete primary structure with structure and exons.

Author Name: Leinonen
Accession Number: 1995:7558
PubMed ID: 7523402
Source: Journal of Biological Chemistry

The entire sequence of the human alpha 1(I) procollagen DNAs. The complete translation of the alpha 1(I) procollagen 38-residue putative signal peptide residue NC1 domain. The gly-Xaa sequences of 1-12 residues in the alpha 1(I) procollagen chain belongs to the alpha 2-like type. Fetal tissues revealed quite similar sequences. The expression patterns and alpha 4(IV).

Get Accession Number Get PubMed ID

2 of 2

この文献のレコードを検索する

CAPLUS のレコード番号 (AN) を抽出し検索

Get Accession Numbers from STN File

STN File

CAPLUS HCAPLUS ZCAPLUS

Save Script
Script Name: 2022_0008_Script

Each request is limited to 5000 Accession Numbers.

Continue Cancel

MEDLINE の PubMed ID (DN) を抽出し検索

Get PubMed IDs from STN File

STN File

MEDLINE

Save Script
Script Name: 2022_0008_Script

Each request is limited to 5000 PubMed IDs.

Continue Cancel

■ Subject タブ : ヒットした配列の由来となった CAS RN®, GenBank Accession 番号を抽出できます。

- 抽出できる番号の上限は 5,000 件です。ボタンをクリックし、検索したいデータベースを選択後 Continue ボタンをクリックすると、コマンドラインに移動し検索が実行されます。

Alignment Patents (10) Journals (2) Subject

Sequence Length: 1690 aa
CAS Registry Number®: 791139-98-5, 159348-50-2, 459495-63-7, 480647-33-4, 1865810-39-4
GenBank Accession No.: CAA56943, Y17397, X81053, CAA76763
Organism: Homo sapiens

Sequence:

```

1 MMSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYYVYSG KKYIGPCGGR
51 DCSVCHCVPE KGSRGPPGPP GPQGPITGPLG APGPIGLSGE KGMRGDRGPP
101 GAAGDKGDKG PTGVPGFPLG DGIPGHGPPP GPRGKPMMSG HNGSRGDPGF
151 PGRGRGALGPG GPLGHPGCEKG EKNSVFIILG AVKGIQDGRG DPGLPLPGS
201 WGAGGAPGPT GYPGEPGLVG PPGQGRPRLG KGNPVGIVKQ QMGDPGEVQ
251 QSSPGPTLLV EPPDFCLYKG EKGIKIGIPGM VGLPGPPGRK GESGIGAKGE
301 KGIPIGFGPR GDPGSYGSPG FPLKLGELGL VGDPLFGLI GPKGDPGNRG
351 HPGPPGLVLT PPLPLKGPPLG DPGFPGRYGE TGDVPGPPGP GLLGRPGEAC
401 AQMIGPPGPQ GFPLPLPLPG EAGIPGRPDS APGKPKPKGS PGLPGAPLQ
451 GLPSSVIYC SVGNRGPQGI KGVVPPGGR GPKGKGNEG LCACEPGPMG
    
```

View More >

配列に関する CAS RN® を抽出し検索

Get CAS Registry Numbers from STN File

STN File

REGISTRY CAPLUS HCAPLUS
 ZREGISTRY

Save Script
Script Name: 2022_0009_Script

Each request is limited to 5000 CAS Registry Numbers.

Continue Cancel

配列に関する GenBank Accession 番号を抽出し検索

Get GenBank Accession Numbers from STN File

STN File

GENBANK CAPLUS HCAPLUS
 REGISTRY ZCAPLUS MEDLINE
 ZREGISTRY

Save Script
Script Name: 2022_0009_Script

Each request is limited to 5000 GenBank Accession Numbers.

Continue Cancel

Bioscape Analysis (配列類似性 × 特許の解析)

Bioscape Analysis (バイオスケープアナリシス) 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。

- Biosequence Search Results 画面の上側にある Create Bioscape Analysis ボタンをクリックします。最初の 1,000 回答について解析が行われます。

Sequence Alignment	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
1365	100.00	100.00	22.20	3

- 配列のアライメントの類似性により解析したマップが表示されます。

- ・ 配列質問式が水色のドットで表示されます (CDR 配列検索の結果解析では表示されません)
- ・ アライメントの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。
- ・ バーの高さは、特許の件数を示しています。

マップに表示するアライメントの類似度の設定 (高い→赤/低い→黄色)

配列質問式は水色のドットで表示

- ・ 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 キーワードや法的状況で配列の由来となる特許を検索

- ・ その他にも、表示の仕方の変更や、特定範囲を選択することが可能です。

- 🎨 表示の仕方を変更
- 📷 スクリーンショットをとる
- Select Sequence 特定範囲を選択
- Exit 選択の解除

■ バーを選択すると、配列に関する情報を確認できます。

- ・ 特許件数、配列長、情報があれば CAS RN®, ドメインが表示されます。
- ・ 特許件数と CAS RN® はリンクになっており、クリックすると STN のコマンドラインに移動し、選択したファイルで検索できます。

REGISTRY ファイルの検索へ (p.16 参照)

1257178-91-8
Relevant Patents: 7
Sequence Length: 1923

Multiple CAS RNs Associated with this Sequence X
Relevant Patents: 4
Sequence Length: 410

複数の CAS RN® が該当する場合の表示

特許データベースの選択へ (p.15 参照)

- ・ Select Sequence で特定の範囲を選択して、まとめて確認することも可能です。

選択を開始する

ドラッグして選択

配列に関する情報の表示 (上記参照)

特許データベースの選択へ

選択のクリア

終了する

特許データベースの選択へ

■ コントロールパネルの Search から、キーワードや法的状況で、配列の由来となる特許を検索し、限定して配列を確認することも可能です。

All

Tit./Abs./Cla. cancer

Simple legal status Active

Search

Reset

Undetermined

Active

Inactive

Pending

該当するバーが青色で表示される



ヘルプデスク



ご不明な点がありましたら、ヘルプデスクまでお問い合わせください。
機能やコンテンツの追加・改善に関するご要望もお待ちしております。

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

URL: www.jaici.or.jp

E-mail: support@jaici.or.jp