

CAS STNext®

核酸・タンパク質配列検索

化学情報協会 情報事業部
202401



本日の内容

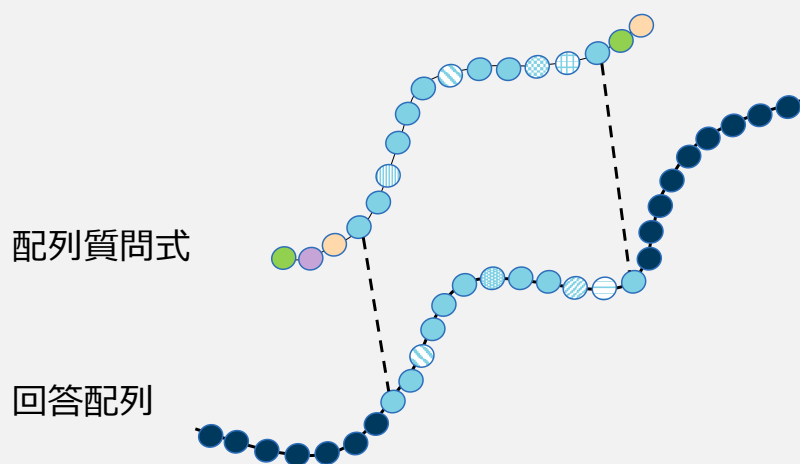
1. CAS STNext の配列検索
2. 完全配列検索・部分配列検索
3. ホモロジー検索
 - REGISTRY
 - GENESEQ
4. CAS Sequences

CAS STNNext の配列検索

配列検索とは

3

塩基コード、アミノ酸コードを用いた検索



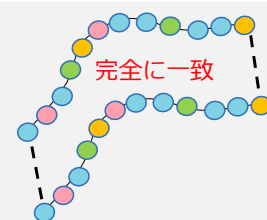
化学物質名検索や構造検索*ではヒットしない核酸・タンパク質が、**配列検索**でヒットする場合がある (*水素以外の元素数が 252 以下の場合には構造検索も可能)

① 完全配列検索

- 質問式と**完全に一致**する配列を検索

配列質問式

回答配列

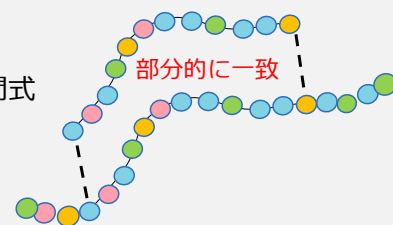


② 部分配列検索

- 質問式を**一部に含む**配列を検索

配列質問式

回答配列

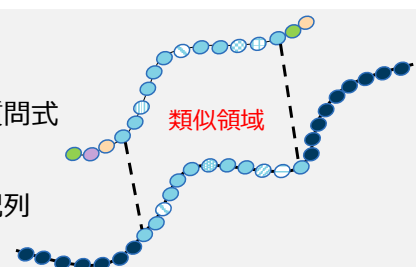


③ ホモロジー検索

- 質問式と**類似**した配列を検索

配列質問式

回答配列



© 2024 化学情報協会

配列検索の選択

5

配列質問式

SPATSGSTPE

質問の配列と
完全に一致する配列を検索

① 完全配列検索

1 SPATSGSTPE
=====

①' 完全配列 ファミリー検索*

1 SAPATGSSTE
=====

質問式の配列を
一部に含む配列を検索

② 部分配列検索

121 EGTSESATPE SGPGESPATSGSTPEGTS..
=====

②' 部分配列 ファミリー検索*

1 SAGSPGSPAG SPTSTEEGTS ESATPE...
=====

質問式の配列に
類似した配列を検索

③ ホモロジー検索

Score: 23.1 bits (47)
Expect value: 3.612e3:
Identities: 7 / 9 (77.8%)
Q: 1 SPATSGSTP 9
||||| |
S: 60 SPATSGASP 68

* ファミリー検索：等価なアミノ酸も含めた配列検索

© 2024 化学情報協会

ファイル名	内容
REGISTRY	世界の雑誌や特許から抽出した配列を収録
GENESEQ	世界の特許から抽出した配列を収録
USGENE	米国特許に記載された配列を収録
PATGENE	PCT 出願に記載された配列を収録
GENBANK	米国国立衛生研究所作成の核酸配列データベース

機能	内容
CAS Sequences	下記から配列を収録 <ul style="list-style-type: none">REGISTRY ファイル7 つの主要特許発行機関の特許NCBI 由来の配列

* 網掛は本講習会で説明するファイルや機能

検索方法・コマンドの違い

ファイル名	完全配列検索 部分配列検索	ホモロジー検索	
		BLAST	GETSIM
REGISTRY	コマンドで検索 SEARCH	独立したソフトウェアで検索 コマンド不要	—
GENESEQ USGENE PATGENE	コマンドで検索 RUN GETSEQ	コマンドで検索 RUN BLAST	コマンドで検索 RUN GETSIM
GENBANK	配列検索はできない		

機能	完全配列検索 部分配列検索	ホモロジー検索		
		BLAST		GETSIM
CAS Sequences コマンド不要	—	BLAST	CDR Motif	—

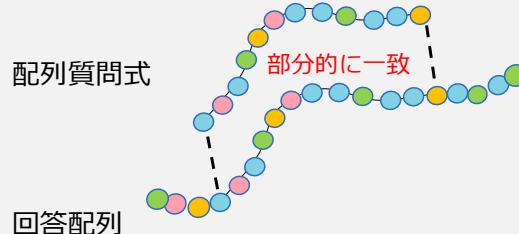
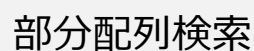
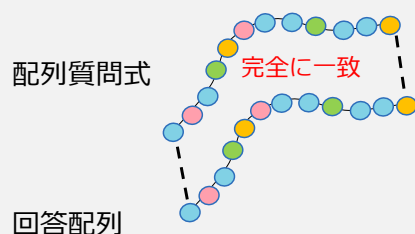
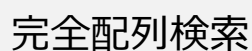
* 網掛は本講習会で説明する検索

完全配列検索・部分配列検索

完全配列検索・部分配列検索

9

質問式と**一致する配列を検索**したい場合は完全配列検索または部分配列検索を実行する



データベース中の配列の中から 一致する配列を検索する

REGISTRY では SEARCH コマンド、GENESEQ では RUN コマンドを使う

REGISTRY : => S コード/検索フィールド

GENESEQ : => RUN GETSEQ コード/検索フィールド

	検索タイプ	検索フィールド
核酸	完全配列	/SQEN
	部分配列検索	/SQSN
タンパク質	完全配列	/SQEP
	完全配列ファミリー*	/SQEFP
	部分配列	/SQSP
	部分配列ファミリー*	/SQSFP

* 等価なアミノ酸も含めた配列がヒットする © 2024 化学情報協会

核酸の検索では下記の塩基コードを利用できる

- 塩基コードは 5' 末端から 3' 末端の順で入力する
- 各ファイルで利用可能な核酸の塩基コードの詳細は、=> HELP NUC 参照。下記以外に曖昧コードも利用可能

塩基	塩基名	完全配列検索でヒット		部分配列検索でヒット	
		REGISTRY	GENESEQ	REGISTRY	GENESEQ
A	アデニン	A	A	A	A
C	シトシン	C	C	C	C
G	グアニン	G	G	G	G
T	チミン (DNA)	T	T, U	T, U	T, U
U	ウラシル (RNA)	U	T, U	T, U	T, U

タンパク質の検索では下記の 20 種類のアミノ酸コードと次のページに示すコードも利用できる

- アミノ酸コードは N 末端 (NH₂) から C 末端 (COOH) の順で入力する

1文字コード	3文字コード	アミノ酸名	完全・部分配列検索でヒット		1文字コード	3文字コード	アミノ酸名	完全・部分配列検索でヒット	
			REGISTRY	GENESEQ				REGISTRY	GENESEQ
A	ALA	アラニン		A	M	MET	メチオニン		M
C	CYS	システイン		C	N	ASN	アスパラギン		N
D	ASP	アスパラギン酸		D	P	PRO	プロリン		P
E	GLU	グルタミン酸		E	Q	GLN	グルタミン		Q
F	PHE	フェニルアラニン		F	R	ARG	アルギニン		R
G	GLY	グリシン		G	S	SER	セリン		S
H	HIS	ヒスチジン		H	T	THR	トレオニン		T
I	ILE	イソロイシン		I	V	VAL	バリン		V
K	LYS	リシン		K	W	TRP	トリプトファン		W
L	LEU	ロイシン		L	Y	TYR	チロシン		Y

© 2024 化学情報協会

アミノ酸コード (続き)

13

1文字コード	3文字コード	アミノ酸名	完全配列検索でヒット		部分配列検索でヒット	
			REGISTRY	GENESEQ	REGISTRY	GENESEQ
O	PYL	ピロリシン	O	O	O	O
U	SCY	セレノシステイン	U	U	U	U
B	ASX	アスパラギン アスパラギン酸	-	B,D,N	D, N	B,D,N
J	XLE	イソロイシン ロイシン	-	I,L	I,L	I,L
Z	GLX	グルタミン酸 グルタミン	-	Z,E,Q	E,Q	Z,E,Q
X	XXX	特殊・未定義	X	A-Y, X*	X	A-Y, X*

- 各ファイルで利用できるアミノ酸コードの詳細は、=> [HELP AAC](#) 参照
- REGISTRY ファイルでは、X (特殊・未定義のアミノ酸コード) の詳細を => [HELP AAU](#) で確認できる

* GENESEQ ファイルの X の詳細は [HELP ACC](#) を参照

コードの詳細は「REGISTRY ファイル配列検索」「GENESEQ ファイル配列検索」を参照
<https://www.jaici.or.jp/stn-ip-protection-suite/cas-stnext/documents/#07>

© 2024 化学情報協会

WTLNSAGYLLGPH を一部に含むタンパク質を REGISTRY ファイルで調査する

[検索条件]

- 配列長は 30 以下で限定する
- 特許でクレームされている配列に限定する

© 2024 化学情報協会

絞り込みに便利な検索フィールド - REGISTRY

検索項目	フィールド	入力例
配列長	/SQL	<ul style="list-style-type: none"> ● 数値検索フィールド => S L1 AND 100/SQL => S L3 AND 10-20/SQL => S L5 AND 200=<SQL
特許情報	/PNTE または /FEAT	<ul style="list-style-type: none"> ● クレームされている配列に限定 => S L2 AND CLAIM?/PNTE ● クレームされていない配列に限定 => S L3 AND UNCLAIM?/PNTE
特徴表	/NTE	<ul style="list-style-type: none"> ● クレームされている配列の形態や修飾の情報を収録 => S L1 AND CYCLIC/NTE => S L3 AND MODIFIED/NTE ● 特殊・未定義のアミノ酸コードの定義を収録

その他の配列関連検索フィールドおよび特徴表についての詳細は、<詳細版> テキスト参照
<https://www.jaici.or.jp/stn-ip-protection-suite/cas-stnext/documents/#07>

© 2024 化学情報協会

- SQIDE 表示形式などの配列情報が表示される表示形式を使用する
 - デフォルトの IDE 表示形式では配列情報は表示されない

表示形式	内容
SCAN	CA 索引名、分子式、クラス識別子、構造図、配列長 * 配列は表示されない
SQIDE	IDE 表示形式 (基本的な物質情報)、配列長 (SQL)、核酸 (NA)、特徴表 (NTE)、特許情報 (PNTE)、1 文字コードの配列データ (SEQ)
SQIDE3	SQIDE と同じ。ただし配列データ (SEQ3) は 3 文字コードで表示される

検索例 - REGISTRY

17

```

=> FILE REGISTRY          ← REGISTRY ファイルに入る
=> S WTLNSAGYLLGPH/SQSP   ← 部分配列検索を実行する
L1      344 WTLNSAGYLLGPH/SQSP

=> S L1 AND SQL=<30       ← 配列長で限定
L2      123 L1 AND SQL=<30

=> S L2 AND CLAIM?/PNTE   ← クレームで限定
L3      12 L2 AND CLAIM?/PNTE

=> D L3 1-12 SQIDE        ← 配列情報を表示する

L3  ANSWER 1 OF 12  REGISTRY  COPYRIGHT 2023 ACS on STN
RN   1381842-73-4  REGISTRY
ED   Entered STN:  06 Jul 2012
CN   L-Serine, glycyl-L-tryptophyl-L-threonyl-L-leucyl-L-asparaginy-L-seryl-L-alanylglycyl-
      :
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL  30      ●
PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source   |Reference
=====+=====
Not Given|AU2012200046
         |claimed SEQID
         |89
-----+-----
      :
SEQ      1 GWTLSAGYL LGPHAVGNHR SFSDLNGLTS
         =====
HITS AT: 2-14  ●
  
```

配列長

クレーム

ヒットしたコードには二重線 (=) が付く

ヒット位置

```
=> FILE CAPLUS          ← CPlus ファイルに入る
=> S L3                  ← REGISTRY ファイルの L3 をクロスオーバー検索する
L4          151 L3

=> D 5 BIB ABS HITSEQ    ← BIB ABS HITSEQ 表示形式で表示する
L4  ANSWER 5 OF 151  CAPLUS  COPYRIGHT 2023 ACS on STN

AN  2021:2774180  CAPLUS  Full-text
DN  178:51304
TI  Site-selective itaconation of complex peptides by photoredox catalysis
AU  Wang, Siyao; Zhou, QingQing; Zhang, Xiaoheng; Wang, Ping
    :
SO  Angewandte Chemie, International Edition (2022), 61(5), e202111388
    CODEN: ACIEF5; ISSN: 1433-7851
    :
AB  Site-selective peptide functionalization provides a straightforward and
    cost-effective access to diversify peptides for biol. studies. Among many
    :
IT  136024-41-4
    RL: RCT (Reactant); THU (Therapeutic use); BIOL (Biological study); RACT
    (Reactant or reagent); USES (Uses)
        (synthesis of itaconated peptides and their derivs. through combination
        of transamination and photoredox conditions)
RN  136024-41-4  CAPLUS
CN  L-Serine, glycyl-L-tryptophyl-L-threonyl-L-leucyl-L-asparaginy-L-seryl-L-
    alanyl-glycyl-L-tyrosyl-L-leucyl-L-leucyl-glycyl-L-prolyl-L-histidyl-L-
    alanyl-L-valyl-glycyl-L-asparaginy-L-histidyl-L-arginyl-L-seryl-L-
    phenylalanyl-L-seryl-L-α-aspartyl-L-lysyl-L-asparaginyglycyl-L-
    leucyl-L-threonyl- (CA INDEX NAME)

SEQ  1 GWTLSAGYL LGPHAVGNHR SFSDKNGLTS
```

HITSEQ

ヒットした CAS RN®,
そのロールとテキスト説
明句、CA 索引名、配列

使用できる記号、配列長の制限値

19

● ギャップ記号、特殊記号は => HELP SQQ 参照

記号の利用例は「REGISTRY ファイル配列検索」を参照

<https://www.jaici.or.jp/stn-ip-protection-suite/cas-stnext/documents/#07>

● REGISTRY ファイルのギャップ記号、特殊記号利用例)

=> S SY...RPG/SQSP → SYTYRPG AFWSYKRLRPG

=> S QS[ILM]SSW/SQSP → QSLSSW QSISWLA などがヒット

● 配列質問式の長さの制限

- REGISTRY ファイル

入力方法	核酸	タンパク質
配列コードを直接入力し一回で検索	250 コード	
QUERY コマンドで作成した検索フィールド 付きの質問式の L 番号を & 記号でつなげる	1,000 コード	・完全配列 1,000 コード ・部分配列 2,400 コード
REGISTRY BLAST ソフトウェア	50,000 コード	

- GENESEQ ファイルは => HELP QLMITS 参照

- REGISTRY ファイルでは同主鎖の配列を持っていますが、個別の CAS RN® を持つ場合がある
- SEQLINK EXACT コマンドを使用すると同主鎖であっても別のレコードになった核酸、タンパク質をまとめることができる

=> SEQ CAS RN®

=> SEQ L# ← 配列レコードを含む回答セットの L 番号

検索例 - GENESEQ

21

GACAGCCAAGCGGCAGGGCTGCCAGCAGCC と完全に一致する核酸配列を GENESEQ ファイルで調査する

[検索条件]

- 相補鎖も含める (デフォルト)

5' GACAGCCAAGCGGCAGGGCTGCCAGCAGCC (質問式) 3'
 CTGTCGGTTCGCCGTCCTCCGACGGTCGTCGG (相補鎖) 5'



5' GGCTGCTGGCAGCCCTGCCGCTTGGCTGTC (相補鎖の質問式) 3'

- 特許でクレームされている配列に限定する

● REGISTRY ファイル

- 相補鎖は自動的に検索されないので、必要に応じて相補鎖も別途検索する

=> S GCCCAAGCTGGC/SQSN ← 入力した配列コード（一本鎖）のみ検索（デフォルト）
=> S GCCAGCTTGGGC/SQSN ← 相補鎖の配列コードを 5' →3' の順で作成し検索

● GENESEQ ファイル

- デフォルトで自動的に相補鎖を含めて検索される
- 相補鎖を含めるかはオプションで変更できる

=> RUN GETSEQ GCCCAAGCTGGC/SQSN ← 相補鎖を含めて検索（デフォルト）
=> RUN GETSEQ GCCCAAGCTGGC/SQSN -S BOTH （上記と同じ検索）
=> RUN GETSEQ GCCCAAGCTGGC/SQSN -S SIN ← 入力した配列コード（一本鎖）のみ検索
=> RUN GETSEQ GCCCAAGCTGGC/SQSN -S COM ← 入力した配列コードの相補鎖のみ検索

© 2024 化学情報協会

検索項目	フィールド	入力例
配列長	/SQL	● 数値検索フィールド => S L1 AND 100/SQL => S L3 AND 10-20/SQL => S L5 AND 200=<SQL
特許中の配列の記載位置	/PSL	● クレームされている配列に限定 => S L2 AND CLAIM?/PSL ● クレームされていない配列に限定 => S L3 AND DISCLOSURE/PSL
特徴表	/FEAT	● 配列の特徴を収録 => S L1 AND (DISULFIDE(W)BOND)/FEAT => S L3 AND MRNA/FEAT ● 特殊・未定義のアミノ酸コード (X) に対応する情報を収録 => S (HSE OR HOMOSERINE)/FEAT
WPI レコード番号	/OS	● WPI ファイルのレコード番号を検索 => S 94-151326/OS

その他の配列関連検索フィールドおよび特徴表についての詳細は、<詳細版> テキスト参照
<https://www.jaici.or.jp/stn-ip-protection-suite/cas-stnext/documents/#07>

© 2024 化学情報協会

- 配列の概要やアライメント情報など、回答の適合性を確認したい時は、 TRIAL や ALIGN 表示形式を使用する
- 書誌情報や配列に関する詳細な情報を確認したい場合は、 BIB や ALL 表示形式を使用する

表示形式	内容
TRIAL	レコード番号 (AN)、標題 (TI)、分子式タイプ (MTY)、配列の説明 (DESC)、キーワード (KW)、配列長 (SQL)
ALIGN	<ul style="list-style-type: none">● 完全配列、部分配列検索結果の場合<ul style="list-style-type: none">- 全配列情報（ヒット位置には二重下線が付く）● ホモロジー検索結果の場合<ul style="list-style-type: none">- ヒットした配列と質問配列のアライメント情報（一致/不一致/ギャップ）
BIB	書誌情報、配列の説明 (DESC)、クロスレファレンス (CR)
ALL	書誌情報、抄録、配列情報およびキーワード (KW)

© 2024 化学情報協会

=> FILE GENESEQ

← GENESEQ ファイルに入る

=> RUN GETSEQ GACAGCCAAGCGGCAGGGCTGCCAGCAGCC/SQEN

●

核酸の配列検索では、デフォルトで自動的に相補鎖を含めて検索される

L1 RUN STATEMENT CREATED

L1 3 GACAGCCAAGCGGCAGGGCTGCCAGCAGCC/SQEN

=> S L1 AND CLAIM?/PSL

← クレームで限定

L2 3 L1 AND CLAIM?/PSL

=> D L2 1-3 TRIAL ALIGN

← TRIAL ALIGN 表示形式を用いて標題、配列長、ヒットしたコードを確認する

L2 ANSWER 1 OF 3 GENESEQ COPYRIGHT 2023 CLARIVATE on STN.

AN BKQ82934 GENESEQ

TI Diagnosing subject with benign, pre-malignant, or malignant

:

MTY cDNA

DESC Human orphan non-coding RNA (oncRNA), SEQ ID 6810.

KW cancer; cytostatic; diagnostic test; oncRNA; orphan noncoding RNA; rna

detection; ss; therapeutic; tumor marker

SQL 30

← 配列長

ALIGN

ALIGNMENT FROM L-NUMBER L1

Sequence Length: 30;

Strand: Plus Minus;

Hits at: 30-1

30 GACAGCCAAG CGGCAGGGCT GCCAGCAGCC

=====

●

ヒットしたコードには二重線 (=) が付く

Minus の場合は相補鎖でヒットしている

ヒット位置

© 2024 化学情報協会

L2ANSWER 2 OF 3 GENESEQ COPYRIGHT 2023 CLARIVATE on STN.
ANKQ82933 GENESEQ
TIDiagnosing subject with benign, pre-malignant, or malignant
:
DESCHuman orphan non-coding RNA (oncRNA), SEQ ID 6809.
KWcancer; cytostatic; diagnostic test; oncRNA; orphan noncoding RNA; rna
detection; ss; therapeutic; tumor marker
SQL30 ← 配列長
ALIGN
ALIGNMENT FROM L-NUMBER L1
Sequence Length: 30;
Strand: Plus / Plus;
Hits at: 1-30
1 GACAGCCAAG CGGCAGGGCT GCCAGCAGCC
=====

L2ANSWER 3 OF 3 GENESEQ COPYRIGHT 2023 CLARIVATE on STN.
ANKQ82932 GENESEQ
:
MTYRNA
DESCHuman orphan non-coding RNA (oncRNA), SEQ ID 6808.
KWcancer; cytostatic; diagnostic test; oncRNA; orphan noncoding RNA; rna
detection; ss; therapeutic; tumor marker
SQL30
ALIGN
ALIGNMENT FROM L-NUMBER L1
Sequence Length: 30;
Strand: Plus / Plus;
Hits at: 1-30
1 GACAGCCAAG CGGCAGGGCTU GCCAGCAGCC
=====

入力した質問式のコードでヒット

=> D L2 1 BIB ALIGN ← 書誌情報とヒットしたコードを組み合わせる表示する

L2ANSWER 1 OF 3 GENESEQ COPYRIGHT 2023 CLARIVATE on STN.
ANKQ82934 GENESEQ ED 20220325 UP 20220325
DED 20220324 Full-text
TIDiagnosing subject with benign, pre-malignant, or malignant
hyperproliferative cell used for detecting cancer cell in subject,
involves detecting presence, absence, and/or quantity of non-coding RNA
or its functional fragment in sample.
INGoodarzi H
PAUNIV CALIFORNIA (REGC)
LAEnglish
DTPatent
PIWO 2022040106 A2 20220224
PITWOA2 INTERNATIONAL APPLICATION PUBLISHED WITHOUT INTERNATIONAL SEARCH
REPORT or INTERNATIONAL APPLICATION PUBLISHED WITH DECLARATION UNDER
ARTICLE 17 (2) (A) [FROM 20090101 ONWARDS]
AIWO 2021-US46186 20210816
PRAI 20200816
FSNUCLEIC; NS
OS2022-27666X [020]
MTYcDNA
PSLClaim 64; SEQ ID NO 6810; 916pp
DESCHuman orphan non-coding RNA (oncRNA), SEQ ID 6810.
ALIGN
ALIGNMENT FROM L-NUMBER L1
Sequence Length: 30;
Strand: Plus / Minus;
Hits at: 30-1
30 GACAGCCAAG CGGCAGGGCT GCCAGCAGCC
=====

特許中の配列情報の記載位置

© 2024 化学情報協会

下記の配列を含むタンパク質を REGISTRY ファイルと
GENESEQ ファイルで調べる

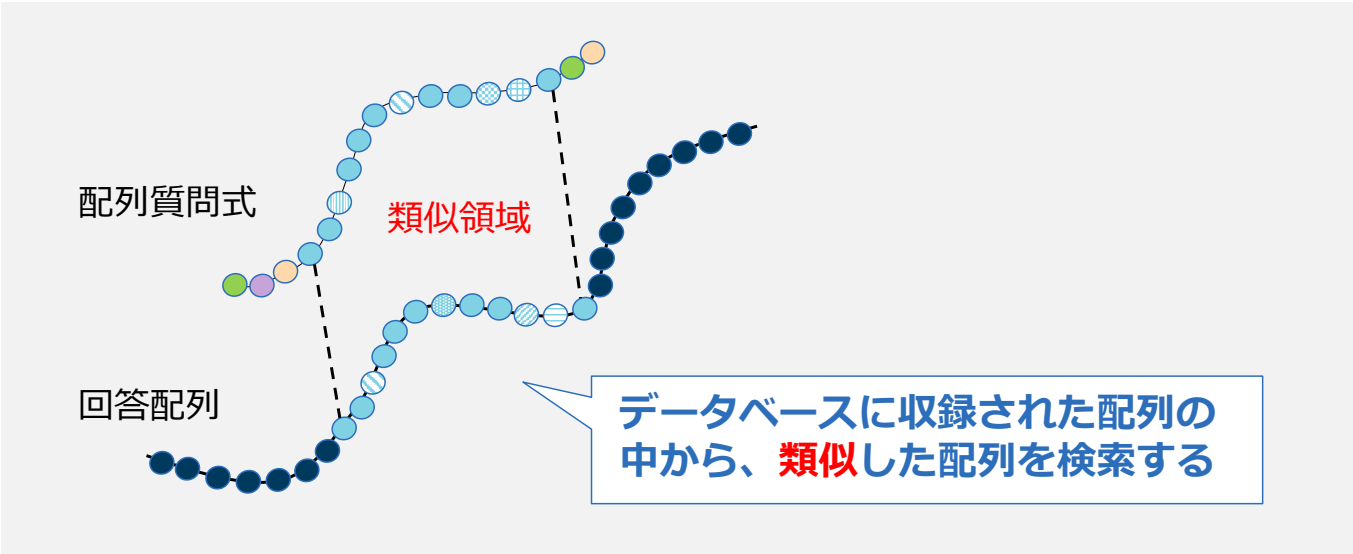
CRHKPMRTVTNFIYIANLAATDVTFLCCVPFTALLYPLPGWV
LGDFMCKFVNYI

[検索条件] 400 以下の配列長でクレームに記載された配列に
限定する

検索	REGISTRY	GENESEQ	参照スライド
タンパク質の 部分配列検索	=> S 質問式/SQSP	=> RUN GETSEQ 質問式/SQSP	10
配列長で限定	SQL 検索フィールドを利用する		15, 23
クレームに限定	CLAIM?/PNTE	CLAIM?/PSL	
表示形式	SQIDE	BIB ALIGN	16, 24

ホモロジー検索

配列質問式と類似した配列を検索したい場合はホモロジー検索を実行する



© 2024 化学情報協会

CAS STNnext では BLAST と GETSIM の 2 種類のホモロジー検索を実行できる

プログラム	BLAST	GETSIM
概要	Basic Local Alignment Search Tool 最もよく利用されている配列検索プログラム。他のプログラムに比べて高速処理できる。ギャップをあまり考慮しないため、検出感度や選択性が低いと考えられがちだが、実際には他と比べてそれほど遜色はない。	FASTA 系列のプログラム。 データベース中のすべての配列との間で忠実にアラインメントを行ってホモロジースコアを算定する。BLAST ホモロジー検索で回答が得られない場合でも、GETSIM ホモロジー検索で回答が得られることがある。
処理速度	速い	遅い
類似性の高い配列	○	○
類似性の低い配列	△	○
比較方法	短い部分配列を比較 ギャップはあまり考慮されない	配列全体を比較 ギャップも考慮される
感受性	デフォルトの設定では低い	高い
データベース	REGISTRY, GENESEQ, USGENE, PATGENE, CAS Sequences	GENESEQ, USGENE, PATGENE

© 2024 化学情報協会

ホモロジー検索 REGISTRY

REGISTRY BLAST 検索の流れ

33

準備

① ソフトウェアのインストール



BLAST 検索 (独立したソフトウェアで実施)

② 検索

- ソフトウェアを起動
- 配列質問式の入力
- 検索タイプの選択
- パラメータの設定

③ 回答表示

- 結果の確認

④ STN 移行のための準備

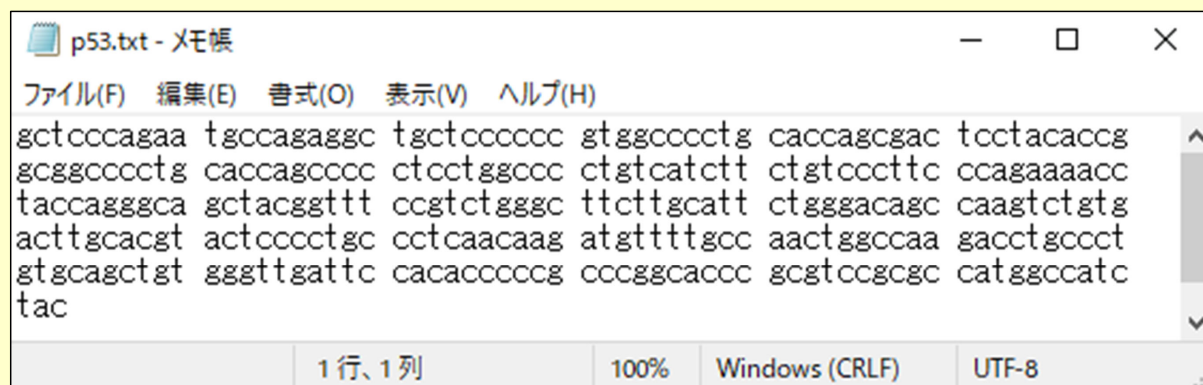
- スクリプトファイル保存
- アライメントデータ保存



⑤ CAS STNNext での検索と表示

- スクリプトを実行、回答を表示
- レポート作成 (オプション)

下記のがん抑制遺伝子 p53 に類似する配列を
REGISTRY BLAST 検索で調査する



```
p53.txt - メモ帳
ファイル(F) 編集(E) 書式(O) 表示(V) ヘルプ(H)
gctcccagaa tgccagaggc tgcctccccc gtggcccctg caccagcgac tcttacaccg
gcggcccctg caccagcccc ctcttgccc ctgtcatctt ctgtcccttc ccagaaaacc
taccagggca gctacggttt ccgtctgggc ttcttgcatl ctgggacagc caagtctgtg
acttgcacgt actcccctgc cctcaacaag atgttttggc aactggccaa gacctgcctt
gtgcagctgt gggttgattc cacacccccg cccggcaccg gcgtccgcgc catggccatc
tac
```

1行、1列 100% Windows (CRLF) UTF-8

配列長 303

© 2024 化学情報協会

① 準備：ソフトウェアのインストール

35

REGISTRY BLAST はソフトウェアで実行するため、
ソフトウェアをインストールする

1) ダウンロードサイトにアクセスして

<https://www.stn.org/stn/downloads/blast-download.html>

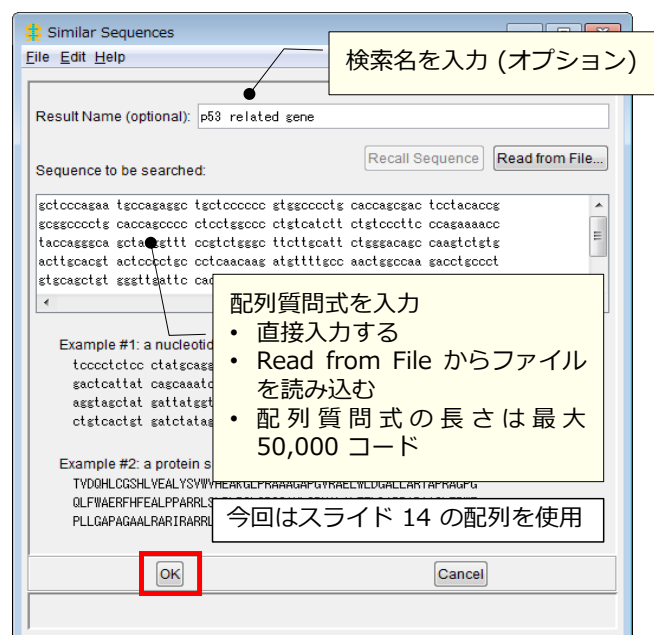
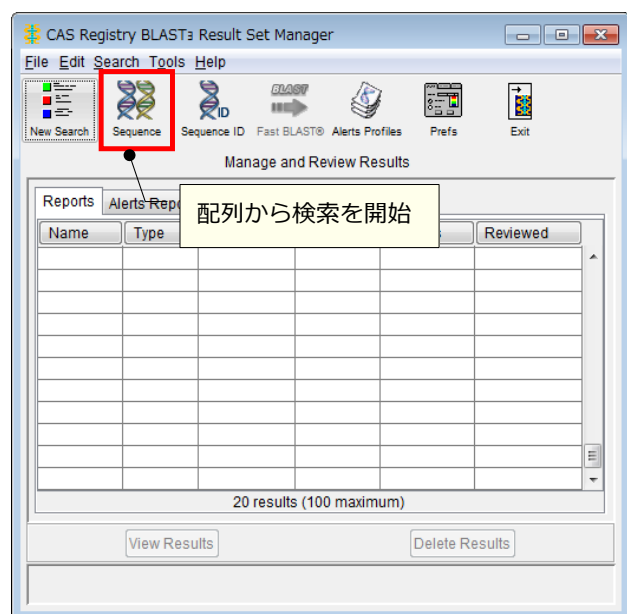
ソフトウェア (.exe) をダウンロードする



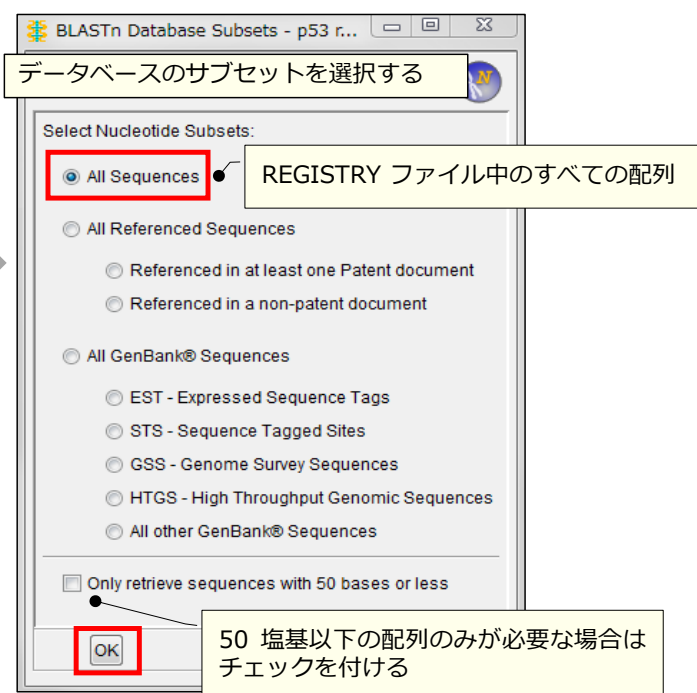
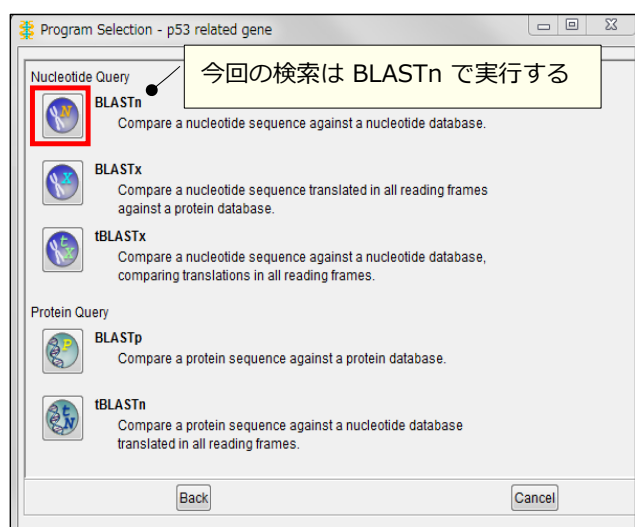
2) .exe ファイルを実行してインストールする

© 2024 化学情報協会

ソフトウェアをクリックして REGISTRY BLAST を起動する。
次に、Sequence ボタンをクリックして配列質問式を入力する。



検索タイプを選択する。



REGISTRY BLAST 検索の検索タイプ

検索タイプ	検索機能	配列質問式	回答
BLASTn	塩基配列の質問式に類似した塩基配列を検索	塩基配列	塩基配列
tBLASTn	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の質問式に類似した配列を検索	アミノ酸配列	塩基配列
tBLASTx	塩基配列の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に翻訳された塩基配列を検索	塩基配列	塩基配列
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	アミノ酸配列	アミノ酸配列
BLASTx	塩基配列の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検索	塩基配列	アミノ酸配列

- 核酸検索時は相補鎖も含めて検索される

② REGISTRY BLAST 検索

パラメータと回答件数の最大値を設定する

「Low Complexity Filtering」のデフォルトはONで低複雑度領域のマスクフィルタリングが行われ、生物学的に無意味なアライメントは取り除かれる設定になっている
特許性調査の場合はチェックをはずした方がよい

その他のパラメータは、チェックを付けると表示される

OK をクリックすると検索が実行される

回答件数の最大値は 1,000 件

検索完了後、結果を表示する

Results は、1,000件×100 セットが最大
(101個目の回答セットを作成する際は、最も古い回答セットが削除される)

検索名	検索タイプ	実行日	状況	回答件数	既読チェック
Result - 601982	BLASTp	2019-04-23 04:20 午後	Complete	1,000	✓
Result - 601980	BLASTn	2019-04-23 03:45 午後	Complete	1,000	✓
Result - 599344	BLASTn	2019-01-29 03:47 午後	Complete	1,000	✓
Result - 599340	BLASTn	2019-01-29 03:15 午後	Complete	1,000	
Result - 599336	BLASTn	2019-01-29 03:09 午後	Complete	1,000	

21 results (100 maximum)

View Results Delete Results

Status (状況) 表示

- Running : 実行中
- Complete : 完了
- Queued : 実行待ち
- Failed : 失敗

(Failed になった場合は再度検索する)

回答はスコア値の高い順に表示される

件数情報

スコア値分布

アライメントの概略

アライメントの詳細

重複する配列が存在する場合は、☐ が表示される

+ をクリックすると詳しいアライメント情報が表示される

スコア値

同一性パーセント

アライメント表示

Query : 配列質問式

Subject : 回答の配列

| は一致しているコード

REGISTRY BLAST の回答はテキスト形式で保存可能



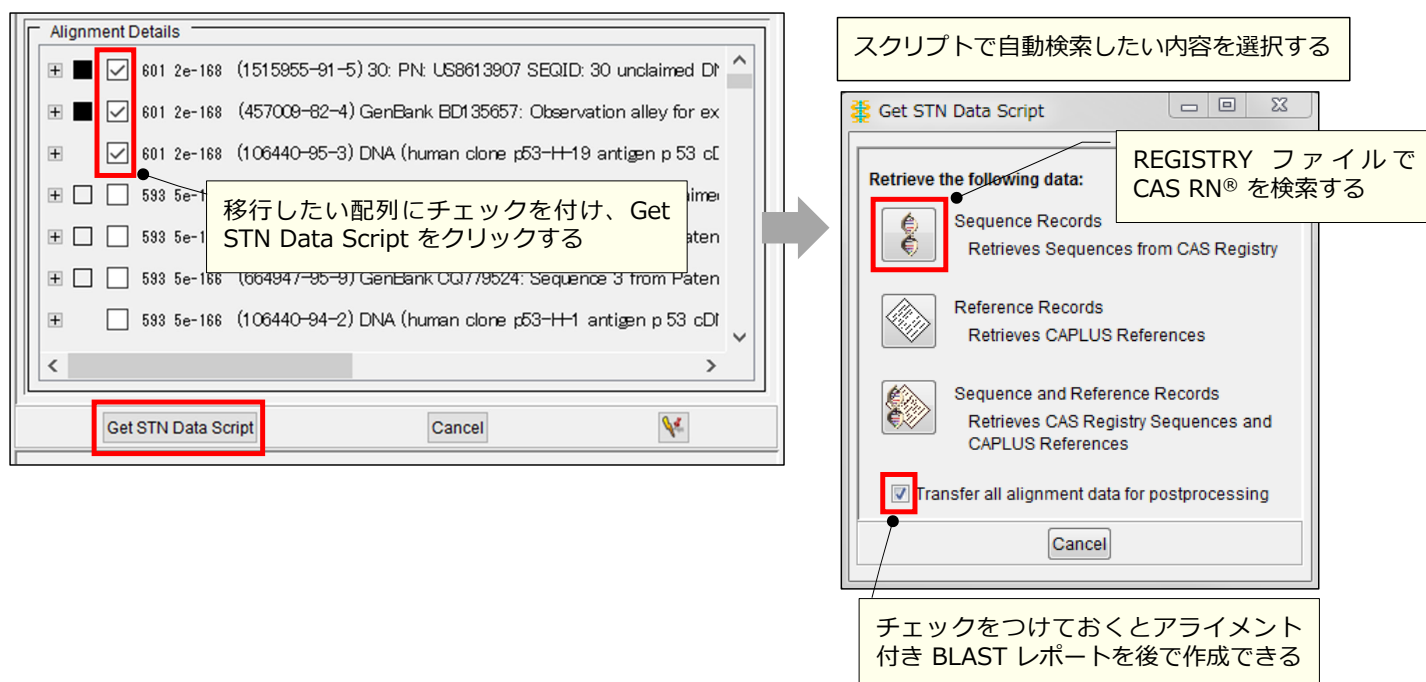
検索条件

回答表示

© 2024 化学情報協会

④ CAS STNext 移行のための準備

得られた配列を CAS STNext へ移行するために Script を作成する



スクリプトで自動検索したい内容を選択する

REGISTRY ファイルで CAS RN® を検索する

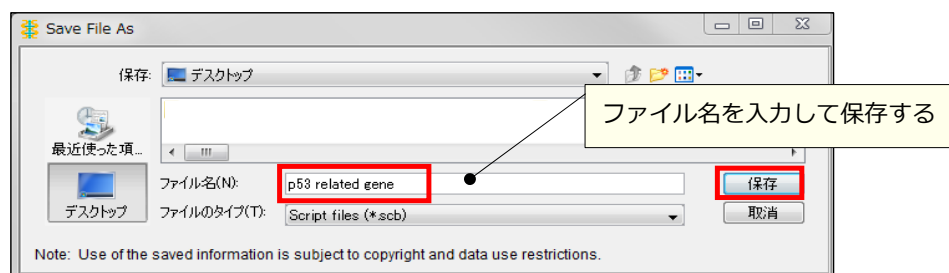
移行したい配列にチェックを付け、Get STN Data Script をクリックする

チェックをつけておくとアライメント付き BLAST レポートを後で作成できる

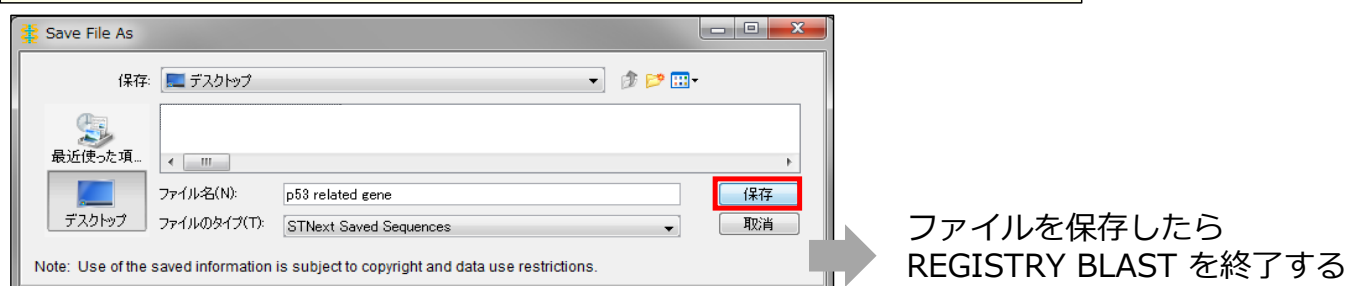
© 2024 化学情報協会

スクリプトファイルとアライメントデータを保存する

STNext へ移行するための Script (.scb) ファイルを保存する

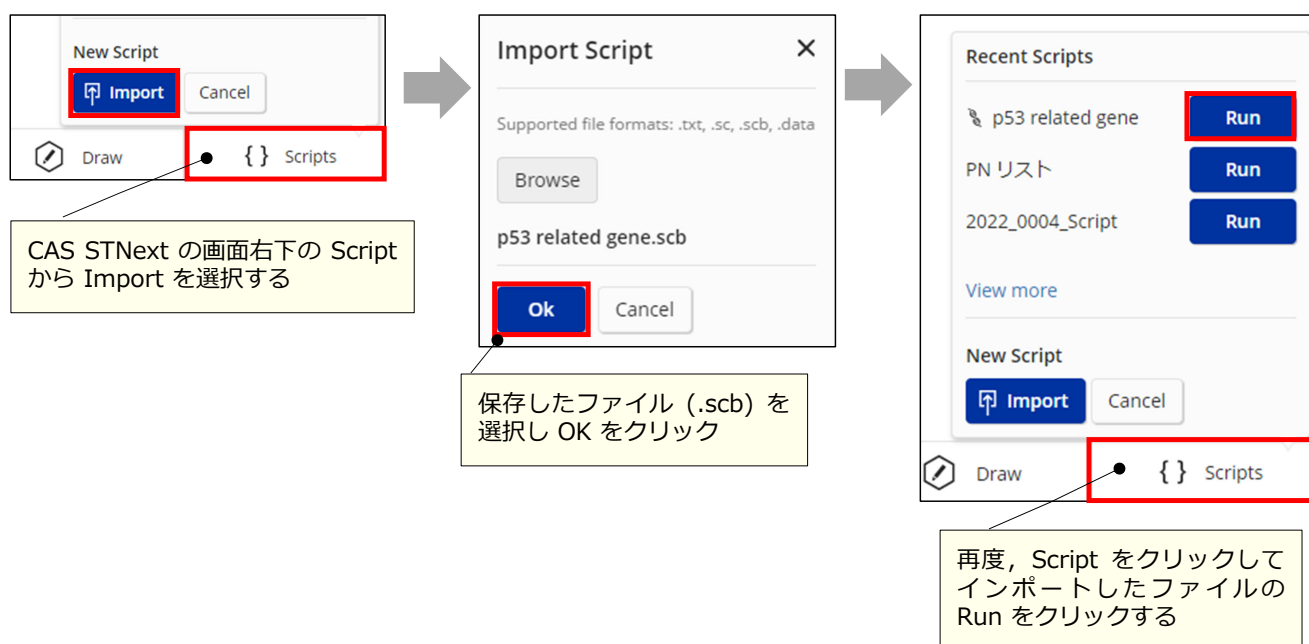


アライメント付きレポートを作成するためにアライメントデータ (.xss) を保存する (オプション)



© 2024 化学情報協会

⑤ CAS STNext での検索と表示

CAS STNext に接続し、Script を実行する
(REGISTRY ファイルで配列レコードを検索する)

© 2024 化学情報協会

自動的に REGISTRY ファイルに入り配列レコードが検索される

The screenshot shows the CAS STNext interface. The 'File REGISTRY' tab is selected. The main area displays a search script and its results. The script is a REGISTRY BLAST search. The results show a list of sequences (L1, L2, L3) and a session log on the right. A callout box points to the 'L3' result, stating: 'Script は、REGISTRY BLAST 検索結果から抽出した配列の CAS RN® の検索式'. Another callout box points to the 'L3' result, stating: '配列の回答集合が作成された'.

© 2024 化学情報協会

回答を SQIDE 表示形式などで表示する

The screenshot shows the SQIDE output of a search. The output includes sequence information, patent annotations, and the sequence itself. Callout boxes highlight specific parts of the output:

- Callout box pointing to 'ANSWER 1 OF 16': 配列長
- Callout box pointing to 'PATENT ANNOTATIONS (PNTE)': 特許情報 (特許番号, 配列の記載位置, 配列番号)
- Callout box pointing to the sequence: 配列

© 2024 化学情報協会

ホモロジー検索 GENESEQ

GENESEQ ファイルのホモロジー検索

51

GENESEQ ファイルの配列検索には RUN コマンドを使う

=> RUN BLAST コード/検索フィールド パラメータ

=> RUN GETSIM コード/検索フィールド パラメータ

- コードは 5 つ以上を入力する
 - コードは以下の順で入力する
 - ・ 核酸 : 5' 末端から 3' 末端の順
 - ・ タンパク質 : N 末端 (NH₂) から C 末端 (COOH)
- アミノ酸の 3 文字コード、ギャップ記号、特殊記号は利用できない
- 回答数の上限のデフォルトは 15,000 件。パラメータで最大 10 万件に変更可

検索 フィールド	BLAST	GETSIM	内容	質問式	回答
/SQN	○	○	塩基配列の質問式に類似した塩基配列を検索	塩基配列	塩基配列
/TSQN	○	○	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の質問式に類似した配列を検索	アミノ酸配列	塩基配列
/SQP	○	○	アミノ酸配列の質問式に類似したアミノ酸配列を検索	アミノ酸配列	アミノ酸配列
/SQM	○	×	非常に類似した (種内などの) 配列用に最適化された BLASTn (megaBLAST)	塩基配列	塩基配列
/SQDM	○	×	一部の塩基を無視し (多少のミスマッチを許容し)、より離れた (種間などの) 配列を検索するために最適化された BLASTn (discontiguous megaBLAST)	塩基配列	塩基配列
/TSQP	○	×	塩基配列の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検索 (BLASTx)	塩基配列	アミノ酸配列
/TSQNX	○	×	塩基配列の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に翻訳された塩基配列を検索 (tBLASTx)	塩基配列	塩基配列

© 2024 化学情報協会

- 核酸検索のデフォルトでは相補鎖も含めて検索される
 - 完全・部分配列検索と同様に相補鎖を含めるかどうかをパラメータにより変更できる

例 : **フィルタリングを行わず**、**入力したコードの鎖のみを検索する (相補鎖を含まない)**

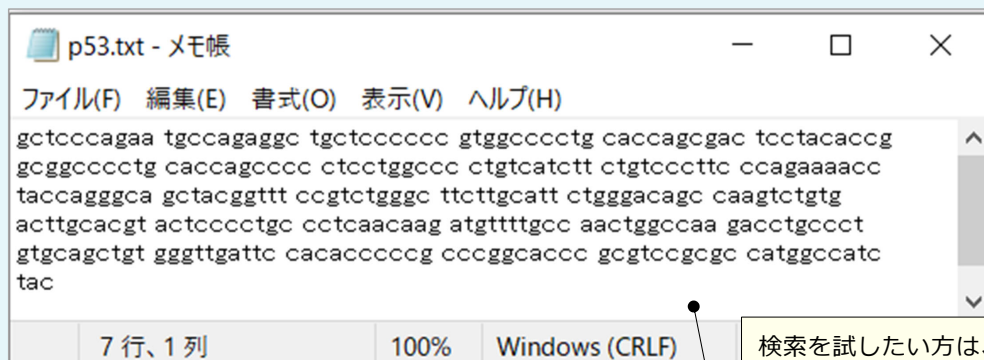
=> RUN BLAST GCTCCCAGAATGC/SQN -F F -S SIN
 - 回答上限の変更
- 例 : **フィルタリングを行わず**、**回答上限を 10 万件に変更して検索する**
- => RUN BLAST L1/SQN -F F -MAXSEQ 100000

* パラメータの詳細については、=> HELP OPTIONS 参照

© 2024 化学情報協会

がん抑制遺伝子 p53 に類似する核酸を GENESEQ ファイルの BLAST 検索で調査する

【事前準備】長い配列質問式の場合はテキストファイル (.txt) を作成する



検索を試したい方は、チャットで送ったコードをコピーし、メモ帳に貼り付け、デスクトップに「p53」と名前をつけて保存してください。


アップロードを使用した => RUN BLAST コード/SEQN の検索の上限は 30,000 コード。詳細は => HELP QLIMITS 参照

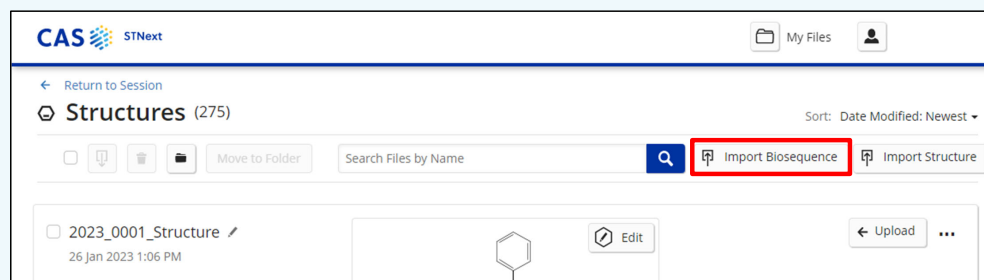
© 2024 化学情報協会

① 配列質問式のアップロード

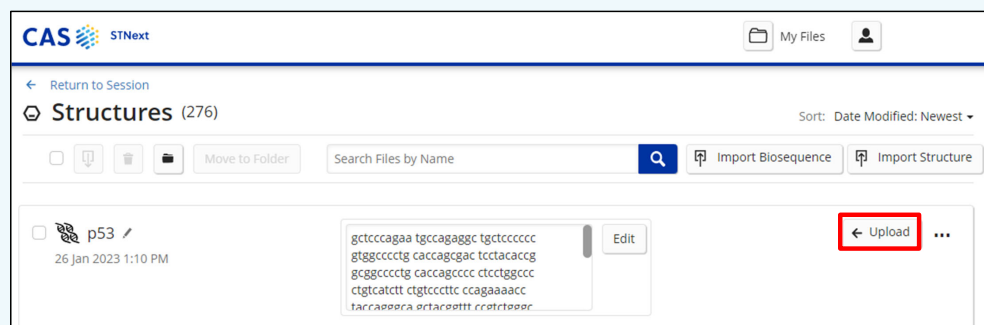
55

=> [FILE GENESEQ](#) ← GENESEQ ファイルに入る

- 1) 画面右上の My Files のプルダウンメニューから Structures を選択する（あるいは画面左の  アイコンをクリック）
- 2) Import Biosequence をクリックし、事前に作成した配列質問式のテキストファイルを読み出して、OK をクリックする。



- 3) 配列質問式がインポートされたら Upload をクリックする。



- 4) アップロードした配列質問式の L 番号が表示される。

© 2024 化学情報協会

=> RUN BLAST L1/SQN -F F

← フィルタを外して BLAST ホモロジー検索を実行する

Algorithm: BLAST - BLASTN. Version: 2.12.0+

The BLAST software is used herein-w
Center for Biotechnology Information
Medicine (NLM). See also: Zheng Zh
and Webb Miller (2000), "A greedy a
sequences", J Comput Biol 2000; 7(1-

オプション：フィルタをはずす (-F F)

Low Complexity Filtering (低複雑度領域フィルタ)

- 低複雑度領域のマスクフィルタリングを行う。
- これにより統計的に有意であっても生物学的に無意味なアラインメント（例：酸性リッチ、塩基リッチ、プロリンリッチ領域）を取り除くことができる。
- 配列質問式中にマスクされるコードは塩基配列の場合は N で、アミノ酸配列の場合は X で置き換えられ検索される。

Database: GENESEQ

Posted Date: Mar 2, 2022 06:00 AM C

Number of Hits to DB: 3430

Expectation Value: 10.0

Number of Sequences: 40160811

Length of query: 303

Length of database: 22574594370

Search space: 6840102094110

Effective search space: 5834704045646

Lambda: 0.625

Kappa: 0.41

Entropy: 0.78

配列質問式の長さ

配列質問式が回答中の配列に完全に一致した場合に得られるスコア値

Highest possible score value: 547.7

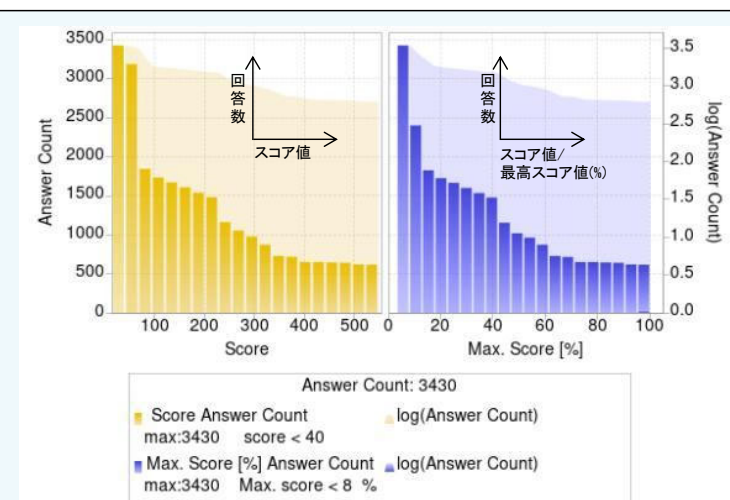
Best answer score value: 547.7

最も類似度が高い回答のスコア値

3430 ANSWERS FOUND BELOW EXPECTATION VALUE OF: 10.0

GENESEQ

回答



Query time: 708

ENTER EITHER "ALL" TO KEEP ALL ANSWERS

OR ENTER THE MINIMUM SCORE VALUE YOU WISH TO KEEP

OR ENTER THE MINIMUM PERCENT OF SCORE FOLLOWED BY "% SCORE"

OR ENTER THE MINIMUM PERCENT OF IDENTITY FOLLOWED BY "% IDENT"

OR COMBINE MINIMUM PERCENT OF SCORE AND IDENTITY AS "X% SCORE Y% IDENT"

OR ENTER "END". "END" MUST BE ENTERED TO COMPLETE THE RUN COMMAND.

ENTER (ALL) OR ? : ALL

回答のオプションを指定する

- 回答全件 (ALL と入力)
- スコア値の最低値 (数字を入力)
- スコア値パーセント (スコア値/最高スコア値) の最低値
- 同一性パーセント (一致コード数 / Alignment コード数) の最低値
- スコア値パーセントの最低値と同一性パーセントの最低値

58

© 2024 化学情報協会

59

© 2024 化学情報協会

L3 ANSWER 3000 OF 3430 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.
 AN BBA99493 GENESEQ
 TI Realizing efficient and fixed-point transgenesis by using transcription activator-like effector nuclease (TALEN) technology comprises designing TALEN target point, and co-transfecting TALEN plasmid and exogenous gene into recipient cells.
 MTY DNA
 DESC Pig P53 target DNA.
 KW dna detection; ds; p53 gene; p53 tumor suppressor protein; rna detection
 SQL 47
 ALIGN

スコア値 (類似性) が低い回答

Query Length: 303; Sequence Length: 47;
 Score: 41.9 bits (45) , 7.6% of highest possible score 547.7;
 Expect value: 1.46;
 Identities: 27 / 30 (90.0%);
 Strand: Plus / Plus; Alignment Length: 30;
 Q: 160 TCTGGGACAGCCAAGTCTGTGACTTGCACG 189
 ||||| ||||| ||||| || |||||
 S: 1 TCTGGAACAGCCAAGTCTGTAACCTGCACG 30

=> D L3 1 ALL ALIGN

← 書誌情報, 抄録, 配列情報を表示する

L3 ANSWER 1 OF 3430 GENESEQ COPYRIGHT 2022 CLARIVATE on STN.
 AN AYM36275 GENESEQ ED 20211030 UP 20211030
 DED 20110120 [Full-text](#)
 TI Evaluating a patient with acute lymphoblastic leukemia (ALL) that is characterized by the presence of Philadelphia chromosome comprises generating an expression profile of ALL biomarkers from a test biological sample.
 :
 PI WO 2010138843 A2 ● ベーシック特許の特許番号のみ収録されている
 :
 OS 2010-P75161 [82] ● WPI ファイルのレコード番号 (AN)
 MTY cDNA
 :

PSL Disclosure; SEQ ID NO 101; 43pp ● 配列情報の記載位置
 :
 AB The present invention relates to novel biomarkers and a method for evaluating the prognosis of a patient with acute lymphoblastic leukemia (ALL) that is characterized by the presence of Philadelphia chromosome (Ph+). The method involves generating an expression profile from a biological sample obtained from a ALL diagnosed patient, where the
 :
 SEQN 101
 SQL 1303 ● 配列長
 SEQK 9102e84c616917a77a5eb3093278c9ff90b18976b8511d450a081fe418cadc
 SEQ

配列独自の抄録

← 配列
 1 gtccaggagc aggtagctgc tgggctccgg ggacactttg cgttcgggct

ALL 表示形式では全配列長が確認できる

:
 1251 cctcccccca taaaaaactc atgttcaaga cagaagggcc tgactcagac
 1301 tga

NA ← 核酸の種類

Code	Count	Percent
A	292	22.4
C	403	30.9

ALIGN

Query Length: 303; Sequence Length: 1303;
 Score: 547.7 bits (606) , 100.0% of highest possible score 547.7;
 Expect value: 7.759e-153;
 Identities: 303 / 303 (100.0%);
 Strand: Plus / Plus; Alignment Length: 303;
 Q: 1 GCTCCAGAGATGCCAGAGGCTGCTCCCCCGTGGCCCTGCACCAGCGACTCCTACACCG 60
 |||||
 S: 308 GCTCCAGAGATGCCAGAGGCTGCTCCCCCGTGGCCCTGCACCAGCGACTCCTACACCG 367
 :

=> D HIS
:
L3 3430 SORT L2 1- SCORE D IDENT D

=> FILE WPIINDEX ← WPI ファイルに入る

=> TRA L3 OS /AN ●
L4 TRANSFER L3 1- OS : 866 TERMS
L5 866 L4/AN
ALL TERMS IN L4/AN RETRIEVED.

=> D L5 10 ← 書誌情報を表示する

L5 ANSWER 10 OF 866 WPIINDEX COPYRIGHT 2022 CLARIVATE ANALYTICS on STN
AN [2021-B5534W](#) [2021087] WPIINDEX [Full-text](#)
TI New recombinant Newcastle disease virus obtained by replacing F protein of
newcastle disease virus losata with the F protein of virulent strain of
Newcastle disease virus useful for preparing medicine for treating tumor e.g. liver cancer
DC B04; C06; D16
IN JIANG S; LI D; LIU T; LIU Z; WANG Z; XIAO W
PA (JIAN-N) JIANGSU KANGYUAN RUIAO BIOMEDICAL TECHNO; (JIAN-N) JIANGSU
KANIONREAL BIOMEDICAL TECHNOLOGY CO LTD
CYC 135
PI WO 2021197506 A1 20211007 (2021087)* ZH 27[13]
CN 113462658 A 20211001 (2021087) ZH ●
ADT WO 2021197506 A1 WO 2021-CN95200 20210521; CN 113462658 A CN 2020-10238162
20200330
PRAI CN 2020-10238162 20200330
IPCI A61K0035-768 [I, A]; A61K0035-768

GENESEQ ファイルの OS フィールドを抽出して WPI
ファイルの AN (レコード番号) フィールドで検索する

特許ファミリーを確認できる

WPI ファイルへクロスオーバーする際に、特許番号 (PN) を用いると GENESEQ
ファイルでヒットした配列情報を含む特許に加え、関連レコード (分割出願、継
続出願など) がヒットする可能性がある

© 2024 化学情報協会

実習 2

63

右記に類似するタンパク質を
GENESEQ ファイルの BLAST
ホモロジー検索で調べる

	実習の流れ	参照スライド
1	GENESEQ ファイルに入る	
2	配列質問式のファイルをアップロードする	55
3	=> RUN BLAST L#/SQP -F F	51-53
4	全件 (ALL) を入手する	57
5	スコア値と同一性の高い順に並べ替える	58
6	ALL ALIGN 表示形式で表示する	60

© 2024 化学情報協会

CAS Sequences

CAS STNext に搭載されている配列検索ツール

65

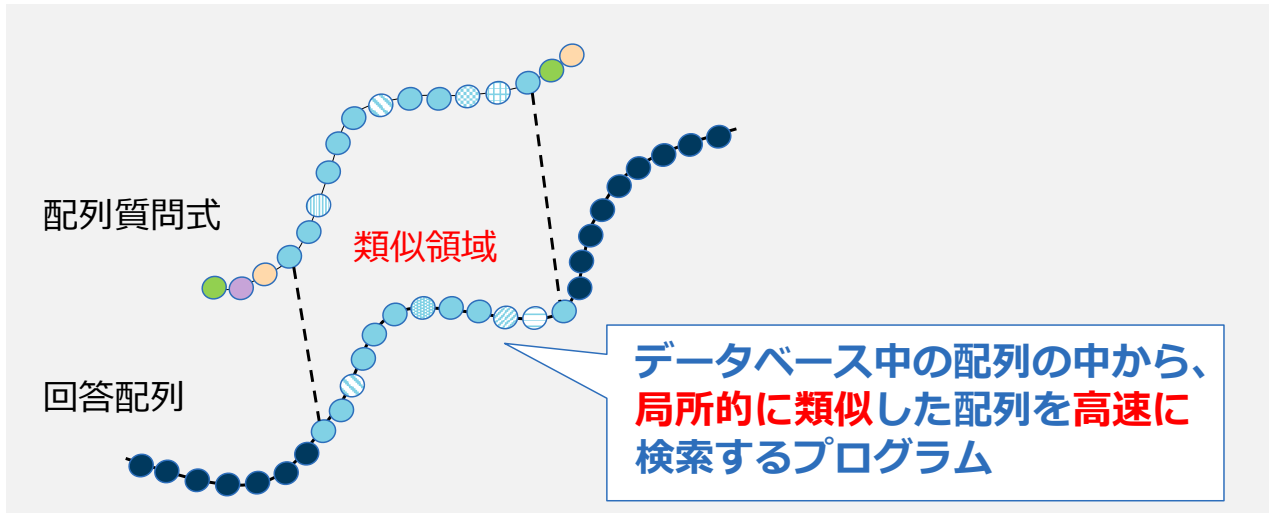
CAS Sequences 機能で膨大な配列コンテンツを検索できる

- 収録源
 - CAS が独自のルールに従い収集した REGISTRY ファイル収録の配列
 - 主に PCT 出願、欧州、米国、カナダ、日本、中国、韓国の特許から抽出した配列
 - NCBI 由来の配列
- 3 つの配列検索プログラム
 - BLAST ホモロジー検索
 - CDR 配列検索
 - Motif 配列検索

詳細は、CAS STNext CAS Sequences ガイド参照
<https://www.jaici.or.jp/stn-ip-protection-suite/cas-stnext/documents/#07>

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する



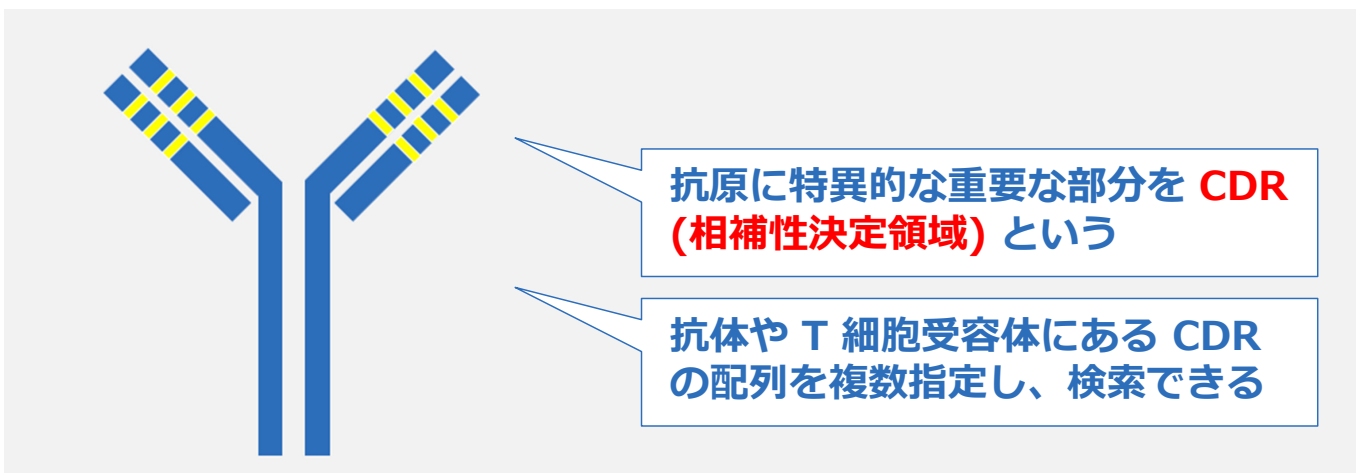
© 2024 化学情報協会

CDR 配列検索

67

CDR (相補性決定領域) を指定し検索するプログラム

- BLAST をベースにしている。CDR 配列検索は、検索対象をあらかじめ抗体や T 細胞受容体といった CDR を持つ配列を対象にしている



© 2024 化学情報協会

パターン配列を検索するプログラム

- 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。数パターンの配列を一度に検索できる

配列 A --- [Pattern A] ---

配列 B --- [Pattern B] ---

配列 C --- [Pattern C] ---

配列 D --- [Pattern D] ---

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

© 2024 化学情報協会

CAS Sequences の検索の流れ

My Files の CAS Sequences から検索をスタート

CAS STNext

My Files

Return to Session

CAS Sequences (0)

Sort By: Date Modified: Newest

Search Files by Name

New Search

You do not have any Sequences

To add a Sequence, click New Search

Show 10 per page

Alerts

Transcripts

Structures

Scripts

CAS Sequences

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る

© 2024 化学情報協会

CASSTNext

My Files

Return to Session

CAS Sequences (0)

Sort By: Date Modified: Newest

Move to Folder

Search Files by Name

New Search

Search Sequences

Close

BLAST

CDR

Motif

Import Sequence

Download

Clear

1 QAVYVTEPSLTIVSPGGTTLTORSSTGAVTTSNYANWYQOKPGOAPRGLIGGINKRAPWTPARFSGSLGGKAALTITGAAEADYYCALWYSNLWVF

2 GGGTKLTVLGGGGGGGGEVQVYSGAELKKPGASVKYSCKASGYTFDYIMKIVRQAPGGLEWIGDIIPNGATFYNGKFKGRVTITVDKSTAYME

3 LSSLRSEDATVY

Sequence type

Nucleotide

Protein

Search within

Nucleotides

Proteins

2023_0001 Sequence

配列を直接入力するか、
Import Sequence から配列
をアップロード

配列検索プログラム

質問式の配列

回答配列

回答上限は 10 万件

Advanced Sequence Search

パラメータ等の設定 (任意)

Run Search

Save Query

Cancel

フィルター

配列長

計算値

ソート

CAS Sequences (149)

Create Bioscape Analysis

Get All Patent Numbers

Show Sequences

Filter By

Sequence Length

237 to 1225

Query Identity %

No Min to No Max

Query Coverage

No Min to No Max

Subject Coverage

No Min to No Max

Subject Identity %

No Min to No Max

Organism

synthetic construct (140)

unidentified (56)

Apply

Reset

Sequence Alignment

Sequence Length

Alignment Identity %

Query Identity %

Subject Identity %

Number of Documents

View More

272

100.00

100.00

77.94

102

View More

272

100.00

100.00

72.80

10

View More

272

93.27

91.51

71.59

1

View More

271

92.92

92.92

72.69

10

View More

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

詳細情報を確認するには
View More をクリック

Alignment Identity% による色分け

低

0-59.99

60-79.99

80-89.99

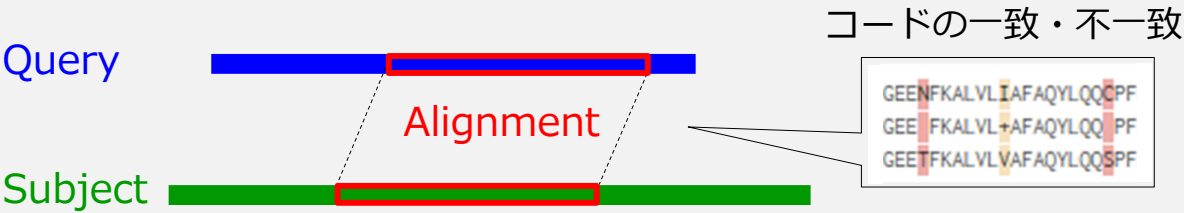
90-94.99

95-96.99

97-98.99

99-100

高

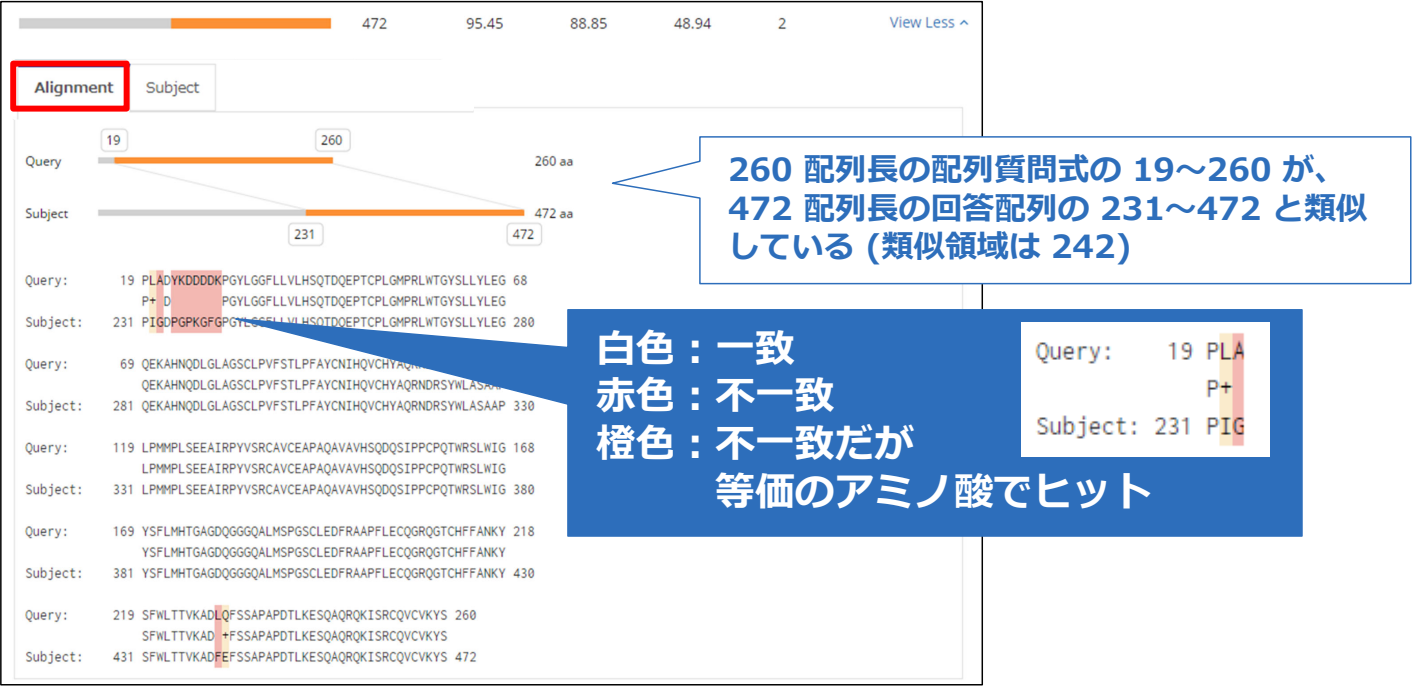


用語	内容
Alignment (類似領域)	配列質問式と回答配列を並べてどこが類似領域か示したもの
Query (配列質問式)	検索した配列質問式
Subject (回答配列)	ヒットした回答配列

検索結果詳細 – Alignment タブ



(例) タンパク質



(例) 核酸



1690100.0088.8513.6712View Less ^

AlignmentPatents (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

Get All CAS Registry Numbers

Get Genbank Accession Nos.

Organism: Homo sapiens
Sequence:

Download Sequence

1 MWSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYVYVGGG KKYIGPCGGR
51 DCSVCHCVPE KGSRGPPGPP GPQGPIGPLG APGPIGLSGE KGMRGDRGPP
1601 FLMHTGAGDQ GGGQALMSGP SCLEDFRAAP FLECQRQGT CHFFANKYSF
1651 LTTWTYKADLQ FSSAPAPDPTL KESQAQRQKI SRCQVCVKYS

View Less ^

回答配列に関連する CAS RN®

回答配列に関連する GenBank Accession 番号

表示している配列に関する CAS RN® や GenBank Accession 番号を検索する (上限 5,000)

① CAS RN® を検索

② GenBank Accession 番号を検索

169196.2889.6213.782View Less ^

AlignmentPatents (2)Journals (0)**Subject**

Sequence Length: 1691 aa
Sequence:

Download Sequence

1 MWSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYVYVGGG KKYIGPCGGR
51 DCSVCHCVPE KGSRGPPGPP GPQGPIGPLG APGPIGLSGE KGMRGDRGPP

CAS RN® の収録がない場合がある

© 2024 化学情報協会

検索結果のダウンロード

CAS Sequences (56)

↓

Create Bioscape AnalysisGet All Patents

Excel で検索結果をダウンロード (上限 1,000 件)

Alignment Image	Alignment Text	Sequence Length	CAS Registry Number®	Number of Patents	Patent No.	Sequence ID (Patent)	Number of Journals	Accession Number
<div> <div>Query</div> <div>Subject</div> </div>	<div> <div>Query: 1 MSMTFFLLCLACRAIAAFLADYVGGGKCYGLPLVLSQTQPTTC 50</div> <div>Subject: 1 MSMTFFLLCLACRAIAAFLADYVGGGKCYGLPLVLSQTQPTTC 50</div> </div>	260	1201979-17-0, 300426-61-3	6	US6498140B1 WO2009153783A1 WO2000059532A1 US9012160B2 US20110151489A1 US6440729B1	8 27 8 27 27 8	0	
<div> <div>Query</div> <div>Subject</div> </div>	<div> <div>Query: 1 MSMTFFLLCLACRAIAAFLADYVGGGKCYGLPLVLSQTQPTTC 50</div> <div>Subject: 1 MSMTFFLLCLACRAIAAFLADYVGGGKCYGLPLVLSQTQPTTC 50</div> </div>	260	273191-20-1	0			1	2000:208665
<div> <div>Query</div> <div>Subject</div> </div>	<div> <div>Query: 19 PLAVYKQDQDQPCYGLPLVLSQTQPTTCGLMGRNTCYLLYLG 68</div> <div>Subject: 977 PIGGKQDQDQPCYGLPLVLSQTQPTTCGLMGRNTCYLLYLG 1026</div> </div>	1218	791151-70-7	2	US20070037165A1 US6812339B1	7065 7065	0	
<div> <div>Query</div> <div>Subject</div> </div>	<div> <div>Query: 19 PLAVYKQDQDQPCYGLPLVLSQTQPTTCGLMGRNTCYLLYLG 68</div> <div>Subject: 1449 PIGGKQDQDQPCYGLPLVLSQTQPTTCGLMGRNTCYLLYLG 1498</div> </div>	1690	1629936-51-1	107	US10124042B2 US10137206B2 WO2018035377A1 JP2017141230A CA3072777A1 US20190241633A1	69 495, 69 495, 69 1044, 5112 3301 2420	13	1998:785255 1997:409028 1994:602449 1993:510145 1998:153085 1996:552738

アライメント

回答配列

特許

非特許

下記の配列と類似した核酸を調査する

ccacagcaca gggtagcaga gcgataacca cacaacgccc atcctctgcg
gagcccaata cagaatacac acgcacgggtg tcttcagagg cattcaggat
gtgcgacgtg tgcctggagt agccccgact cttgtacggt cggcatctgag

	実習の流れ	参照スライド
1	CAS Sequences にアクセスする	69
2	BLAST タブを開き、配列質問式を入力する	70
3	Sequence Type、Search within で Nucleotides を選択する	
4	回答の View More をクリックして回答の詳細情報を確認する	71
5	配列長で限定する	

CAS Sequences (20)

Sort By: Date Modified: Newest

Move to Folder

Search Files by Name

New Search

Search Sequences

Close

BLAST

CDR

Motif

Name2023_0003_Sequence

Import Sequence

Download

Clear

1ccacagcaca gggtagcaga gcgataacca cacaacgccc atcctctgcg
2gagcccaata cagaatacac acgcacgggtg tcttcagagg cattcaggat
3gtgcgacgtg tgcctggagt agccccgact cttgtacggt cggcatctgag

Sequence type

☒ Nucleotide

☐ Protein

Search within

☒ Nucleotides

☐ Proteins

> Advanced Sequence Search

Run Search

Save Query

Cancel

2023_0003_Sequence_01

26 Jan 2023 4:35 PM

ccacagcaca gggtacgaga gcgataacca
cacaacgcc atcctctgcg
gagccaata cagaatacac acgcacggtg
tttcagagg cattcaggat
atccacata tccctgaat acccccaat

View

Cancel Search

...

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

検索中

2023_0003_Sequence_01

26 Jan 2023 4:35 PM

ccacagcaca gggtacgaga gcgataacca
cacaacgcc atcctctgcg
gagccaata cagaatacac acgcacggtg
tttcagagg cattcaggat
atccacata tccctgaat acccccaat

View

View Results

...

1007 results

検索完了

結果を表示

2023_0003_Sequence

16 Jun 2023 4:48 PM

ccacagcaca gggtacgaga gcgataacca
cacaacgcc atcctctgcg
gagccaata cagaatacac acgcacggtg
tttcagagg cattcaggat
atccacata tccctgaat acccccaat

View

Cancel Search

...

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

© 2024 化学情報協会

実習 3（回答）： 検索結果画面

81

Filter By

Sequence Length

151 to 11680

Query Identity %

No Min to No Max

Query Coverage

No Min to No Max

Subject Coverage

No Min to No Max

Subject Identity %

No Min to No Max

Organism

☐ Homo sapiens (109)

☐ unidentified (56)

☐ synthetic construct (49)

☐ Canis lupus dingo (18)

☐ Canis lupus familiaris (15)

View All

Apply

Reset

CAS Sequences (1058)

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	
<div><div></div></div>	2213	100.00	100.00	6.82	2	View More
<div><div></div></div>	2394	100.00	100.00	6.31	2	View More
<div><div></div></div>	1596	100.00	100.00	9.46	1	View More
<div><div></div></div>	2466	100.00	100.00	6.12	31	View More
<div><div></div></div>	2308	100.00	100.00	6.54	2	View More
<div><div></div></div>	1410	100.00	100.00	10.71	28	View More
<div><div></div></div>	2106	100.00	100.00	7.17	1	View More
<div><div></div></div>	2343	100.00	100.00	6.44	2	View More
<div><div></div></div>	3653	100.00	100.00	4.13	1	View More
<div><div></div></div>	870	100.00	100.00	17.36	7	View More

Show

100

per page

Previous

1

2

3

4

5

...

11

Next

お気軽にお問い合わせください

化学情報協会 ヘルプデスク
(平日 9:00~17:00)

Tel : 0120-003-462

Mail : support@jaici.or.jp