STNext 独自の配列検索 Biosequences Search

2021.9.14 化学情報協会 情報事業部

目次

A division of the American Chemical Society

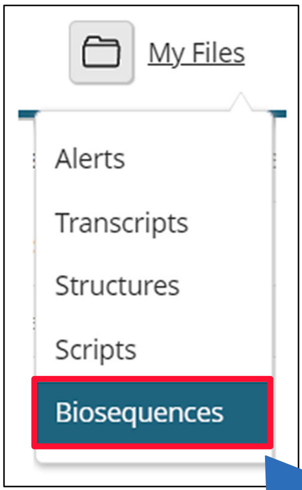
化学情報協会
PARTNER OF CAS

1. Biosequences Search 概要

Biosequences Search とは

Biosequences Search とは STNext に搭載された新たな配列検索機能

1. 膨大な配列コンテンツ
2. 3 つの配列検索プログラム
3. 便利な評価・解析機能



My Files のメニューに Biosequences を追加

1．膨大な配列コンテンツ

5 億 8,000 万件以上の配列を検索可能に！

- 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

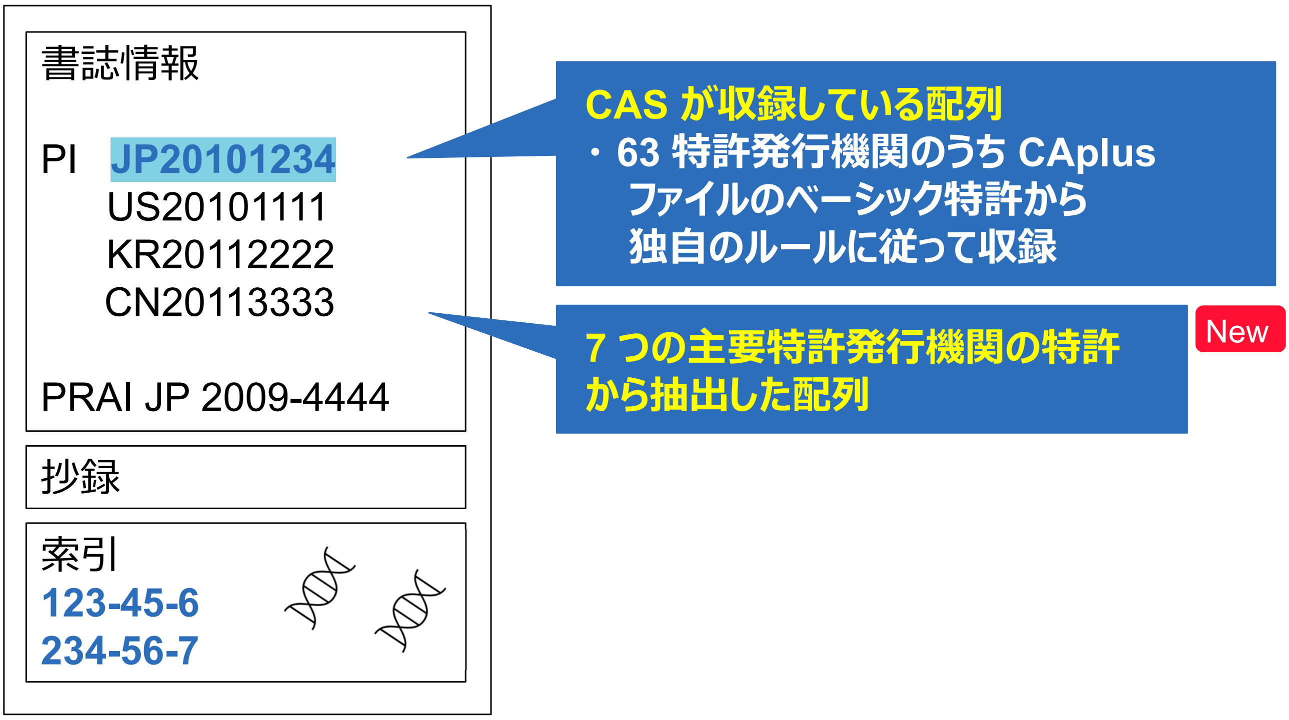
New

- 7 つの主要特許発行機関の特許から抽出した配列 (PCT 出願, 欧州, 米国, カナダ, 日本, 中国, 韓国)

特許
約 5 億 5000万以上

雑誌
約 2,300 万以上

CAplus ファイルのレコード例



書誌情報

PI JP20101234
US20101111
KR20112222
CN20113333

PRAI JP 2009-4444

抄録

索引
123-45-6
234-56-7

CAS が収録している配列
・63 特許発行機関のうち CAplus ファイルのベーシック特許から独自のルールに従って収録

7 つの主要特許発行機関の特許から抽出した配列

New

2．3 つの配列検索プログラム

BLAST, CDR, Motif の 3 つの配列検索プログラムを搭載

BLAST **ホモロジー検索**	局所的に類似した配列を検索するプログラム
New **CDR** **配列検索**	抗体と T 細胞受容体の CDR を指定し，検索するプログラム
New **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

3．便利な評価･解析機能

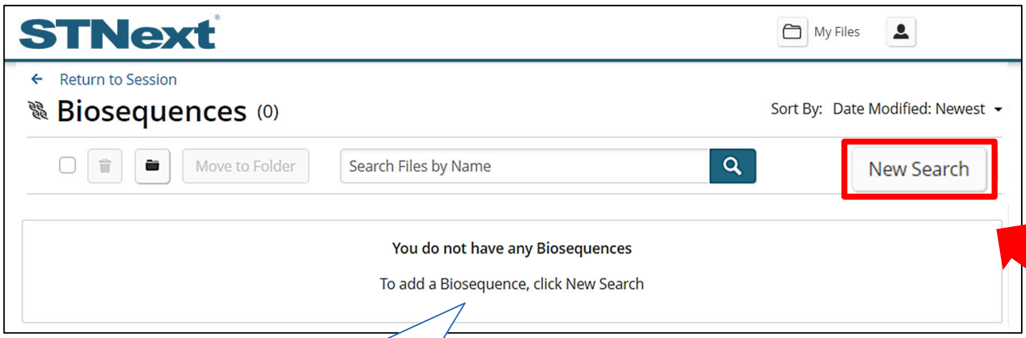
配列検索結果は，以下の機能を使って評価･解析できる

- 様々な計算値による限定やソート
- Excel 形式でダウンロード
- Bioscape Analysis による解析

2. Biosequences Search 利用の流れ
～ BLAST ホモロジー検索を例に ～

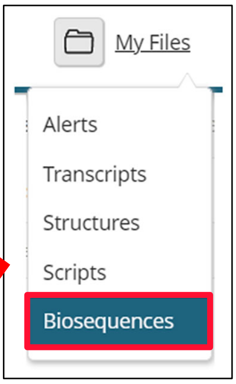
検索スタート

My Files の Biosequences から検索をスタート



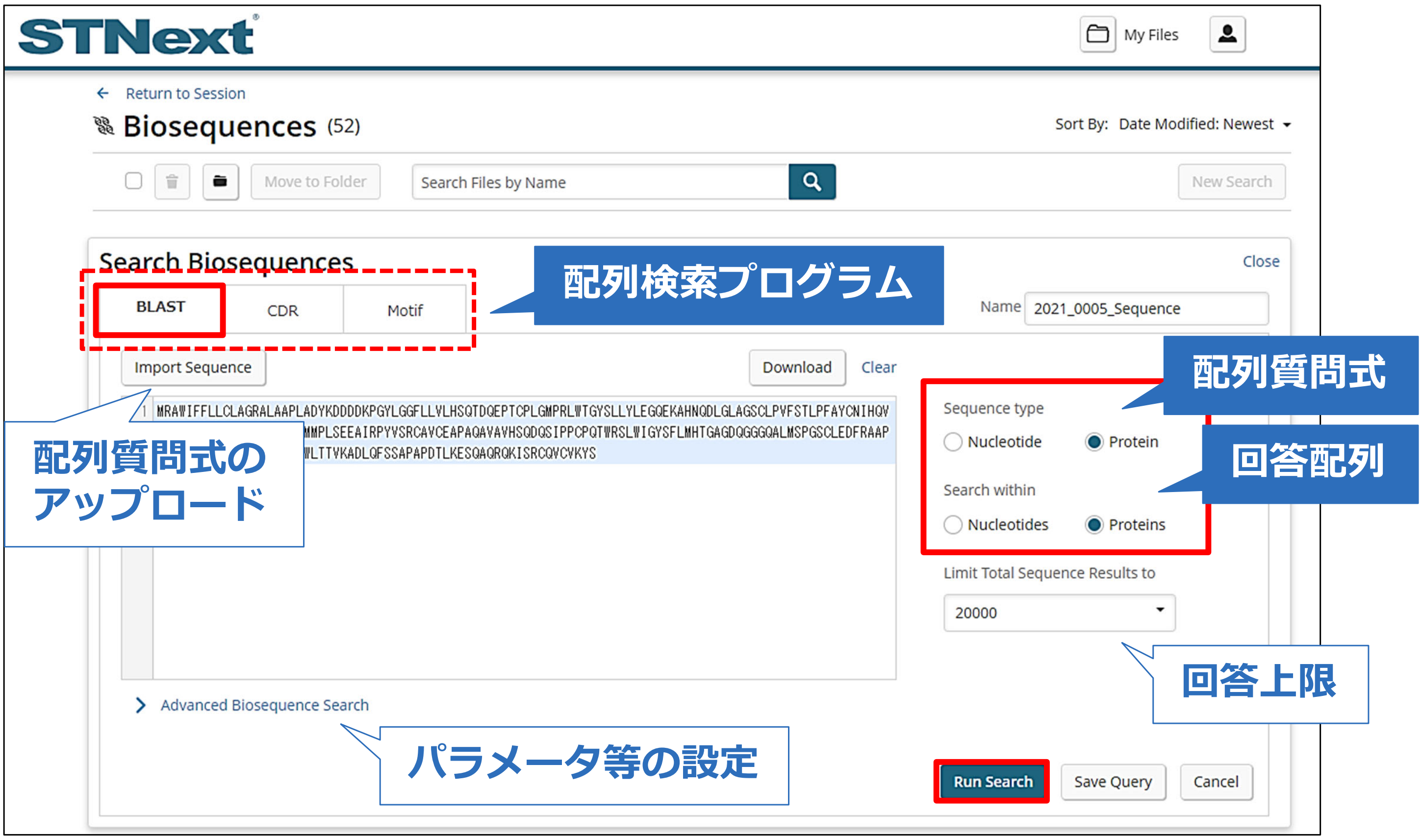
You do not have any Biosequences

To add a Biosequence, click New Search



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

初期画面 - BLAST



Name: 2021_0005_Sequence

1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV
MMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAP
WLTTVKADLQFSSAPAPDTLKESQAQRQKISRCQVCVKYS

Sequence type: ○ Nucleotide ● Protein

Search within: ○ Nucleotides ● Proteins

Limit Total Sequence Results to: 20000

配列検索プログラム

配列質問式

回答配列

配列質問式の
アップロード

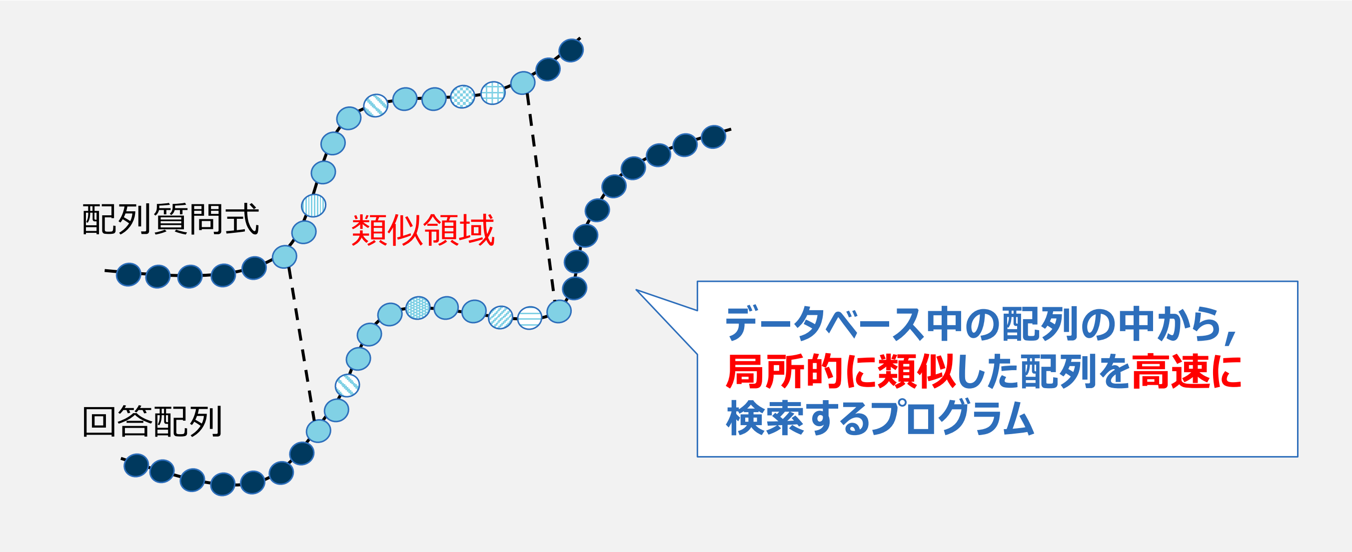
回答上限

パラメータ等の設定

BLAST ホモロジー検索

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する



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

検索タイプ

配列質問式と回答配列を選ぶと自動的に検索タイプが設定される．Advanced Biosequence Search で変更可能．

検索タイプ	内容	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		
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
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

検索履歴画面

MRAWIFFLLCLAGRALAAPLADYKDDDDKPG
YLGGFLLVLHSQTDQEPTCPLGMPRLWTGYS
LLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAY
CNIHQVCHYAQRNDRSYWLASAAPLPMMPL
SFEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPP

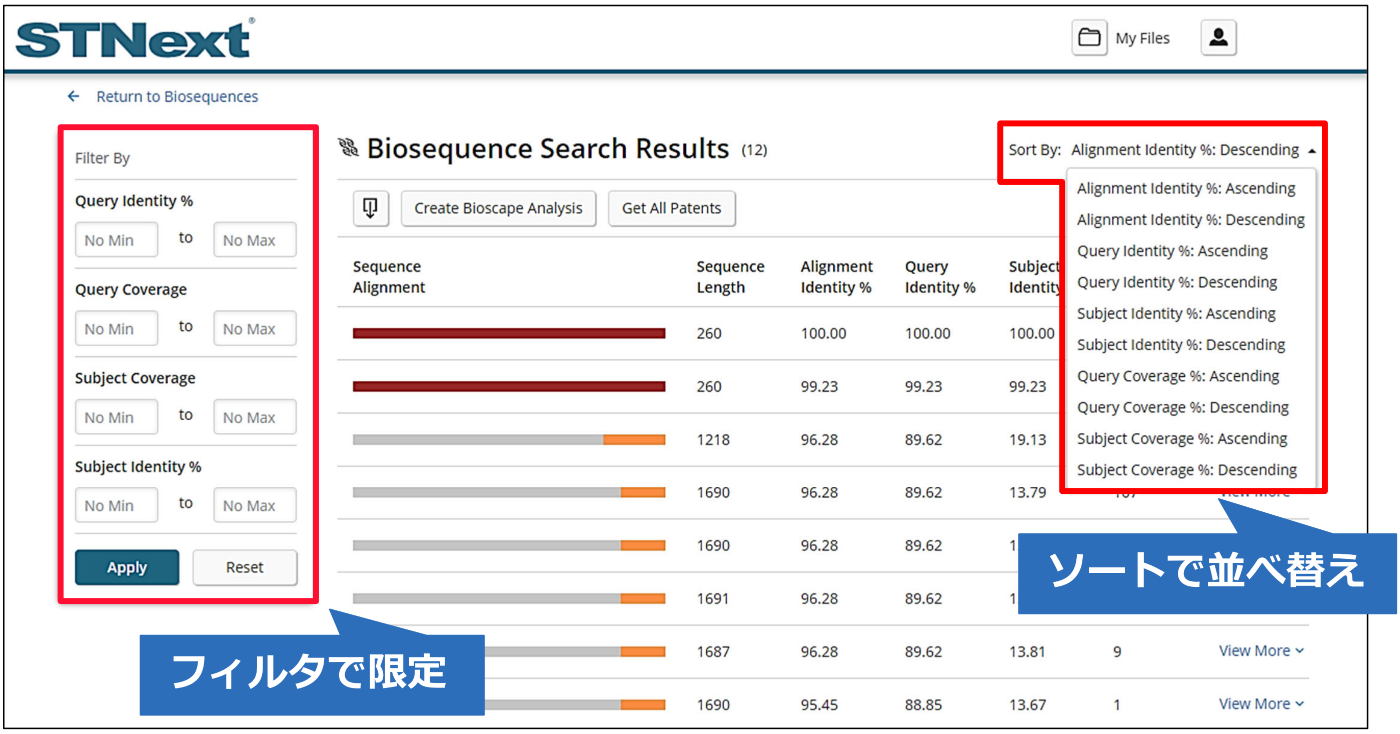
Your search may take some time; while it is running, you may continue to use STNext.

検索実行中

検索完了

検索結果の表示

検索結果一覧



Biosequence Search Results (12)

Filter By

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

Apply Reset

Sort By: Alignment Identity %: Descending

- Alignment Identity %: Ascending
- Alignment Identity %: Descending
- Query Identity %: Ascending
- Query Identity %: Descending
- Subject Identity %: Ascending
- Subject Identity %: Descending
- Query Coverage %: Ascending
- Query Coverage %: Descending
- Subject Coverage %: Ascending
- Subject Coverage %: Descending

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity		
	260	100.00	100.00	100.00		
	260	99.23	99.23	99.23		
	1218	96.28	89.62	19.13		
	1690	96.28	89.62	13.79		
	1690	96.28	89.62			
	1691	96.28	89.62			
	1687	96.28	89.62	13.81	9	View More
	1690	95.45	88.85	13.67	1	View More

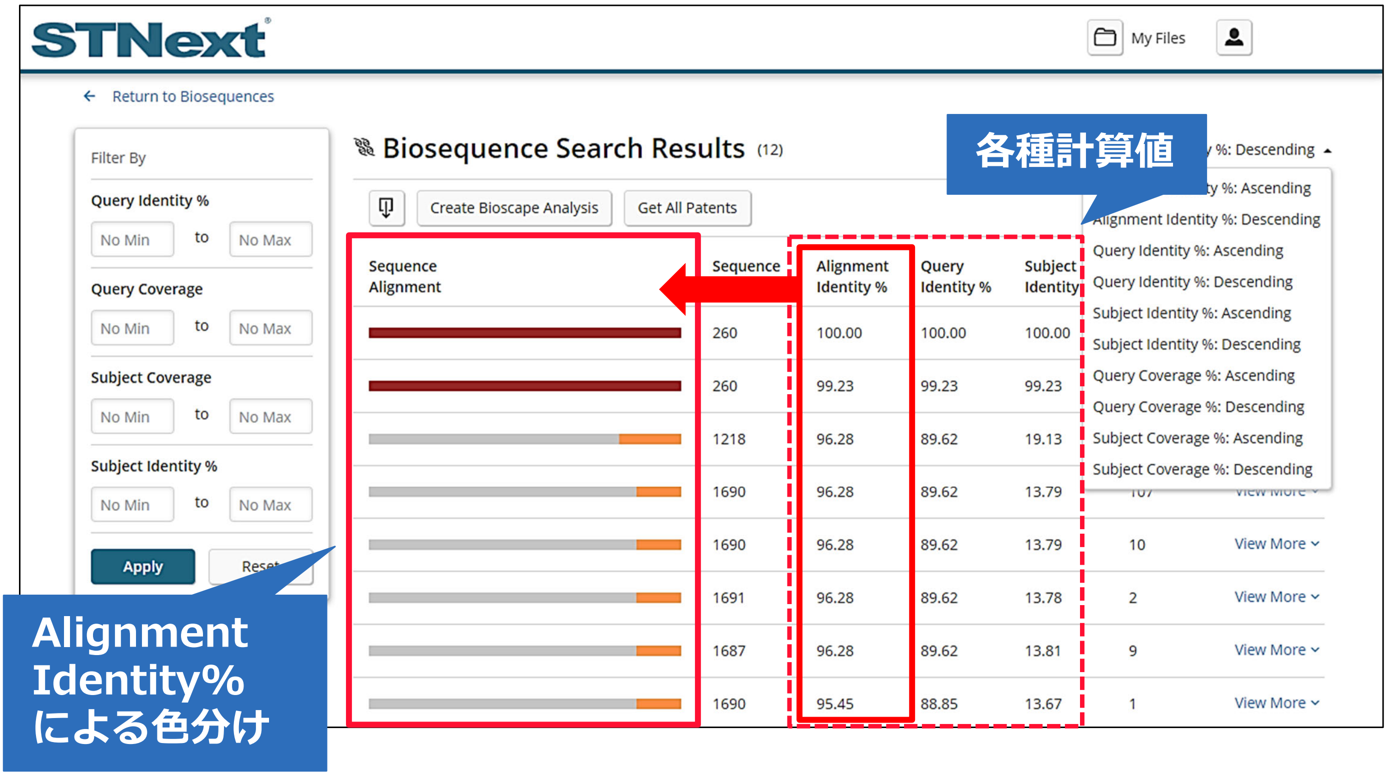
ソートで並べ替え

フィルタで限定

CAS
A division of the American Chemical Society

JAICI
化学情報協会
PARTNER OF CAS

検索結果一覧



Biosequence Search Results (12)

Filter By

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

Apply Reset

%: Descending

- %: Ascending
- Alignment Identity %: Descending
- Query Identity %: Ascending
- Query Identity %: Descending
- Subject Identity %: Ascending
- Subject Identity %: Descending
- Query Coverage %: Ascending
- Query Coverage %: Descending
- Subject Coverage %: Ascending
- Subject Coverage %: Descending

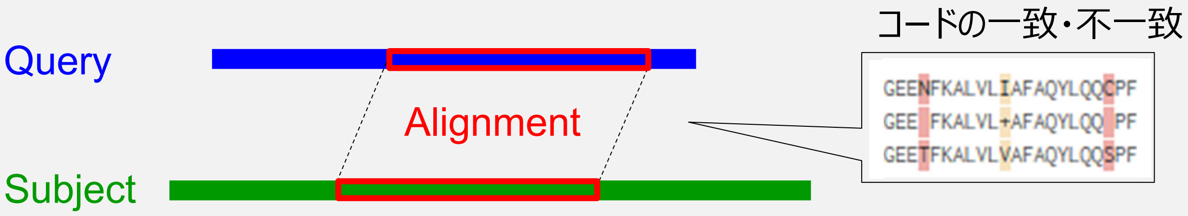
Sequence Alignment	Sequence	Alignment Identity %	Query Identity %	Subject Identity		
	260	100.00	100.00	100.00		
	260	99.23	99.23	99.23		
	1218	96.28	89.62	19.13		
	1690	96.28	89.62	13.79		
	1690	96.28	89.62	13.79	10	View More
	1691	96.28	89.62	13.78	2	View More
	1687	96.28	89.62	13.81	9	View More
	1690	95.45	88.85	13.67	1	View More

各種計算値

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

用語の説明



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

CAS A division of the American Chemical Society

JAICI 化学情報協会 PARTNER OF CAS

計算値の説明

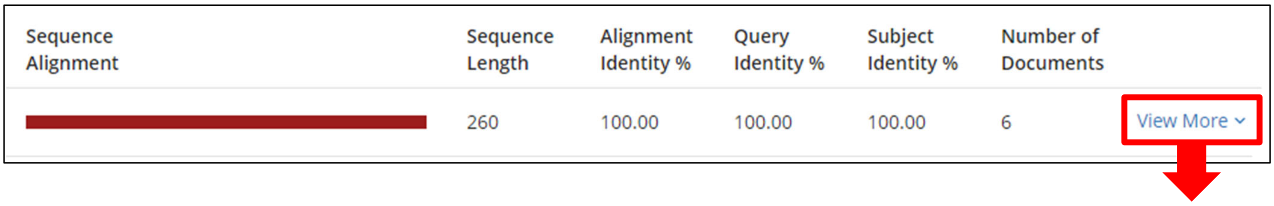
色なし：一致したコード (9)
色つき：一致しないコード (1)

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 ÷ 10 × 100 = 90%
Query Identity %	一致したコード÷配列質問式×100	9 ÷ 10 × 100 = 90%
Subject Identity %	一致したコード÷回答配列×100	9 ÷ 11 × 100 = 81.82%
Query Coverage %	類似領域÷配列質問式×100	10 ÷ 10 × 100 = 100%
Subject Coverage %	類似領域÷回答配列×100	10 ÷ 11 × 100 = 90.91%

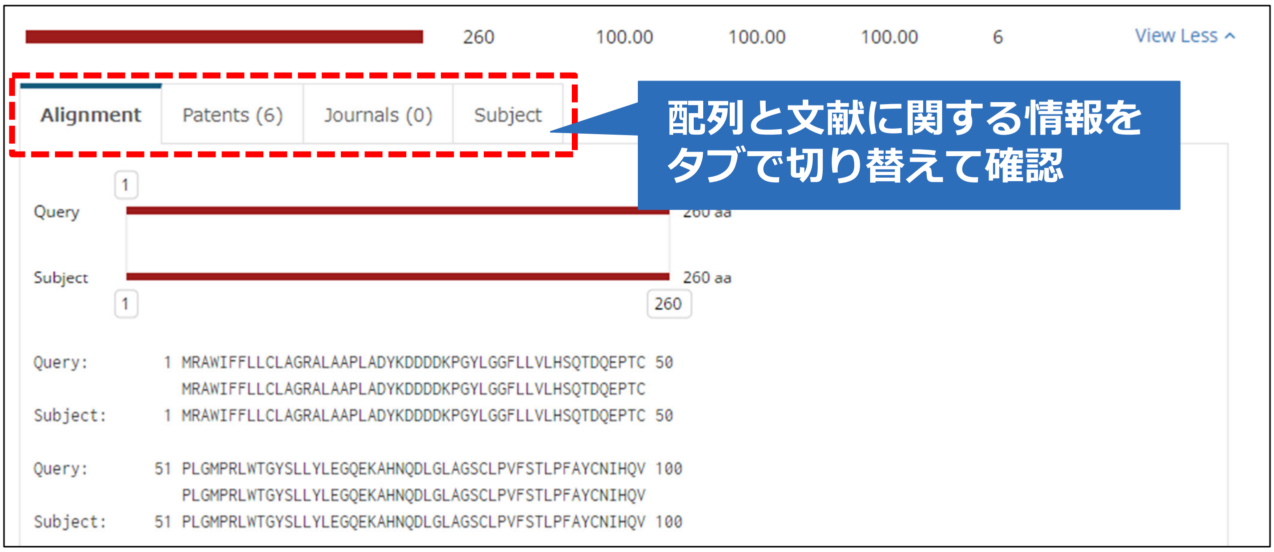
CAS A division of the American Chemical Society

JAICI 化学情報協会 PARTNER OF CAS

検索結果詳細



Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	260	100.00	100.00	100.00	6	View More ⌄

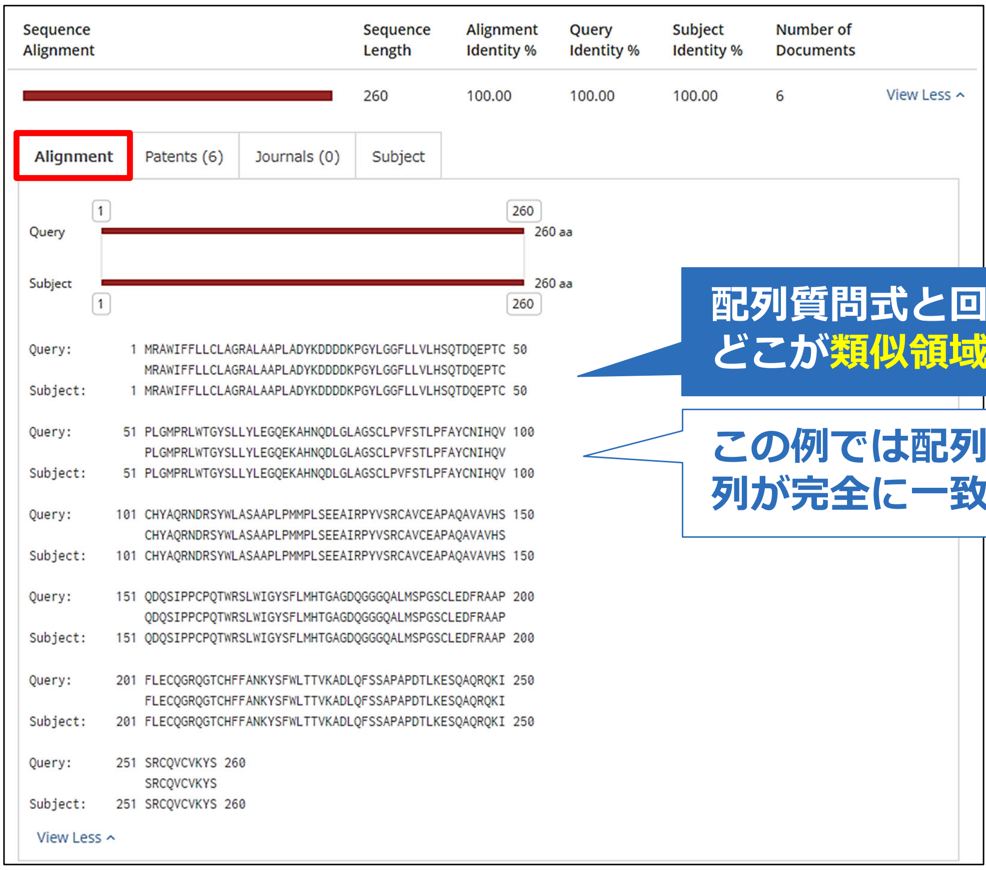


	260	100.00	100.00	100.00	6	View Less ⌃

```
Query:      1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC 50
              MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC
Subject:    1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC 50

Query:     51 PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV 100
              PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV
Subject:   51 PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV 100
```

配列と文献に関する情報を
タブで切り替えて確認

検索結果詳細 – Alignment タブ



Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	260	100.00	100.00	100.00	6	View Less ⌃

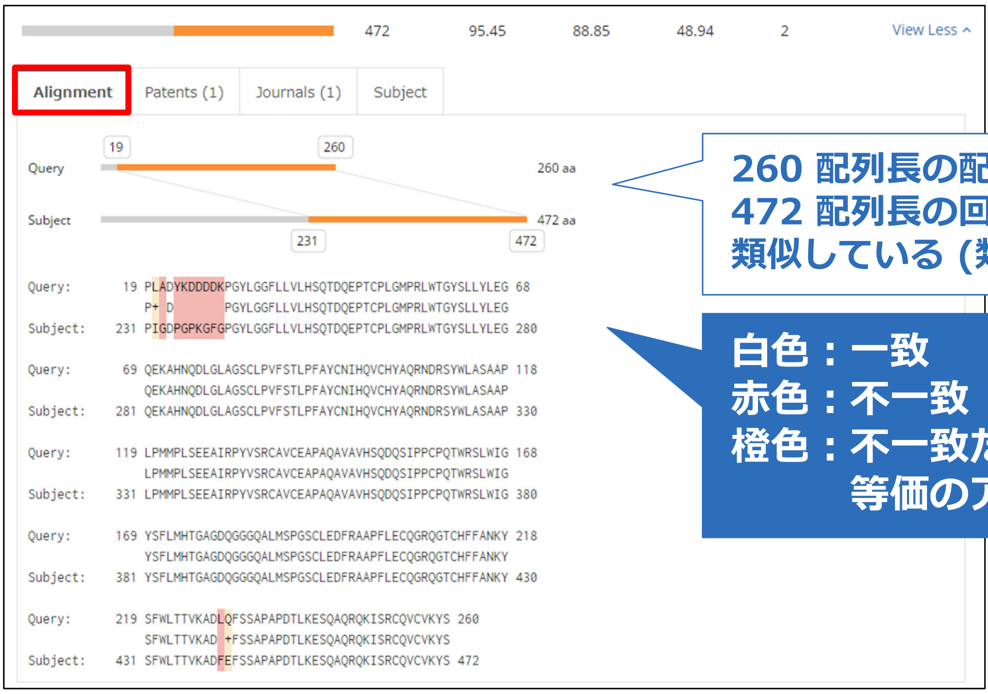
```
Query:      1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC 50
              MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC
Subject:    1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC 50

Query:     51 PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV 100
              PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV
Subject:   51 PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV 100

Query:    101 CHYAQRNDRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHS 150
              CHYAQRNDRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHS
Subject:  101 CHYAQRNDRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHS 150

Query:    151 QDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAP 200
              QDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAP
Subject:  151 QDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAP 200

Query:    201 FLECQGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRQKI 250
              FLECQGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRQKI
Subject:  201 FLECQGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRQKI 250

Query:    251 SRCQVCVKYS 260
              SRCQVCVKYS
Subject:  251 SRCQVCVKYS 260
```

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

この例では配列質問式と回答配
列が完全に一致している

Alignment 例

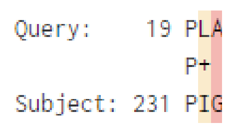
(例) タンパク質



```
Query:     19 PLADYKDDDDKPGYLGGFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEG 68
              P+ D       PGYLGGFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEG
Subject:  231 PIGDPGPKGFGPGYLGGFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEG 280

Query:     69 QEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDRSYWLASAAP 118
              QEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDRSYWLASAAP
Subject:  281 QEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCHYAQRNDRSYWLASAAP 330

Query:    119 LPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIG 168
              LPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIG
Subject:  331 LPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIG 380

Query:    169 YSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFANKY 218
              YSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFANKY
Subject:  381 YSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFLECQGRQGTCHFFANKY 430

Query:    219 SFWLTTVKADLQFSSAPAPDTLKESQAQRQKISRCQVCVKYS 260
              SFWLTTVKAD +FSSAPAPDTLKESQAQRQKISRCQVCVKYS
Subject:  431 SFWLTTVKADFEFSSAPAPDTLKESQAQRQKISRCQVCVKYS 472
```

260 配列長の配列質問式の 19～260 が, 472 配列長の回答配列の 231～472 と類似している (類似領域は 242)

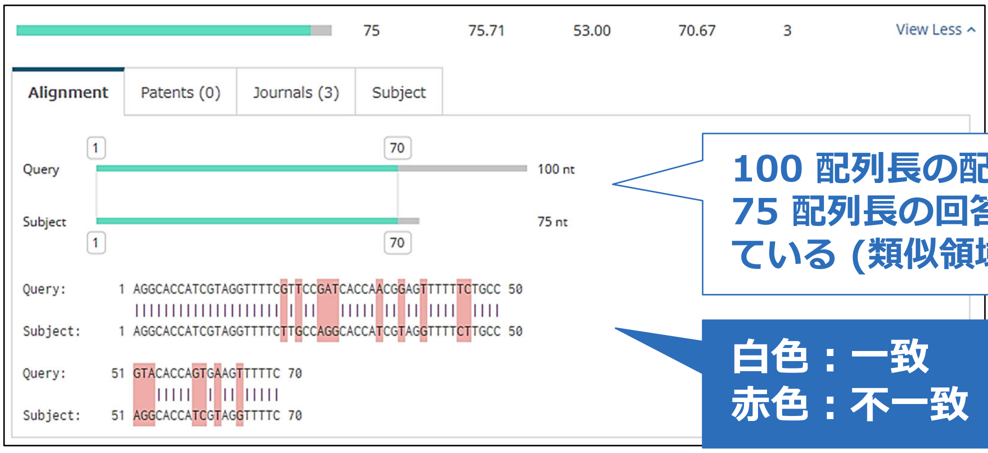
白色：一致
赤色：不一致
橙色：不一致だが
等価のアミノ酸でヒット



```
Query:     19 PLA
              P+
Subject:  231 PIG
```

Alignment 例

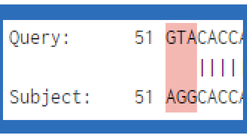
(例) 核酸



```
Query:      1 AGGCACCATCGTAGGTTTTCGTTCCGATCACCAACGGAGTTTTTTCTGCC 50
              |||||||||||||||||||| | ||   ||||| || || ||||  ||||
Subject:    1 AGGCACCATCGTAGGTTTTCTTGCCAGGCACCATCGTAGGTTTTCTTGCC 50

Query:     51 GTACACCAGTGAAGTTTTTC 70
                 |||||  | || |||||
Subject:   51 AGGCACCATCGTAGGTTTTC 70
```

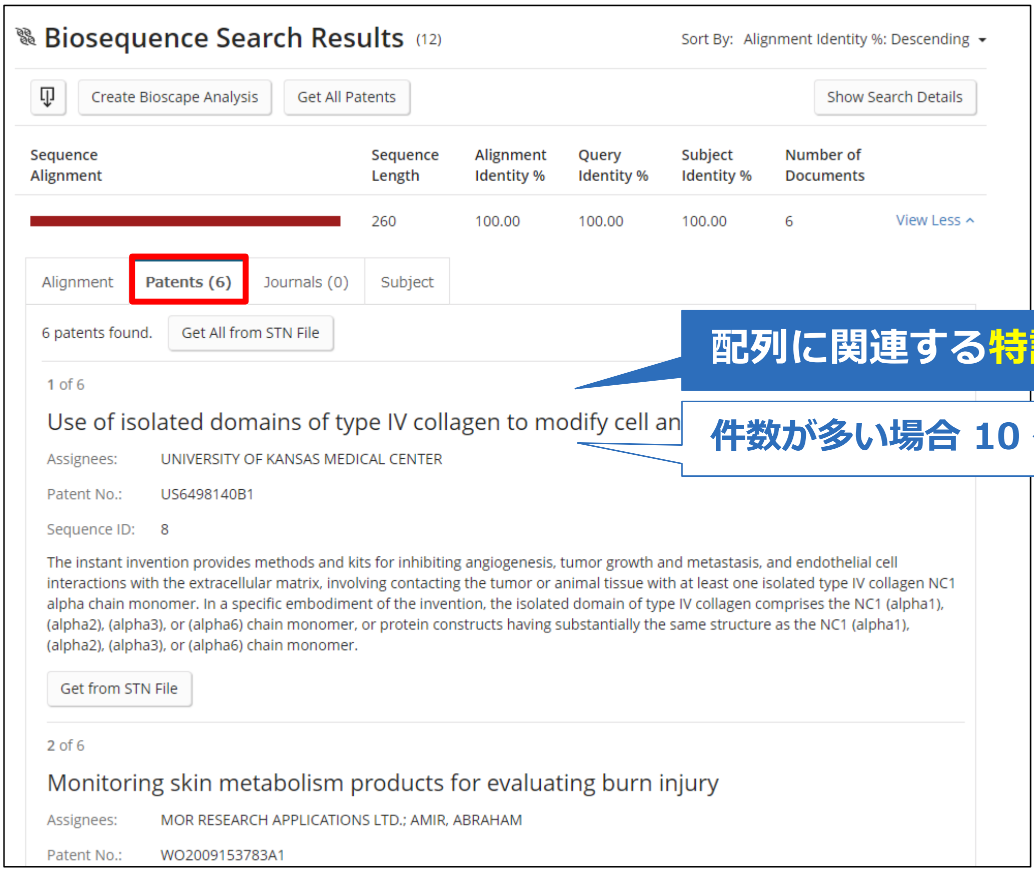
100 配列長の配列質問式の 1～70 が, 75 配列長の回答配列の 1～70 と類似している (類似領域は 70)

白色：一致
赤色：不一致



```
Query:     51 GTACCA
                 ||||
Subject:   51 AGGCACCA
```

検索結果詳細 – Patents タブ



Biosequence Search Results (12)

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	260	100.00	100.00	100.00	6	View Less ^

6 patents found.

1 of 6

Use of isolated domains of type IV collagen to modify cell an

Assignees: UNIVERSITY OF KANSAS MEDICAL CENTER

Patent No.: US6498140B1

Sequence ID: 8

The instant invention provides methods and kits for inhibiting angiogenesis, tumor growth and metastasis, and endothelial cell interactions with the extracellular matrix, involving contacting the tumor or animal tissue with at least one isolated type IV collagen NC1 alpha chain monomer. In a specific embodiment of the invention, the isolated domain of type IV collagen comprises the NC1 (alpha1), (alpha2), (alpha3), or (alpha6) chain monomer, or protein constructs having substantially the same structure as the NC1 (alpha1), (alpha2), (alpha3), or (alpha6) chain monomer.

2 of 6

Monitoring skin metabolism products for evaluating burn injury

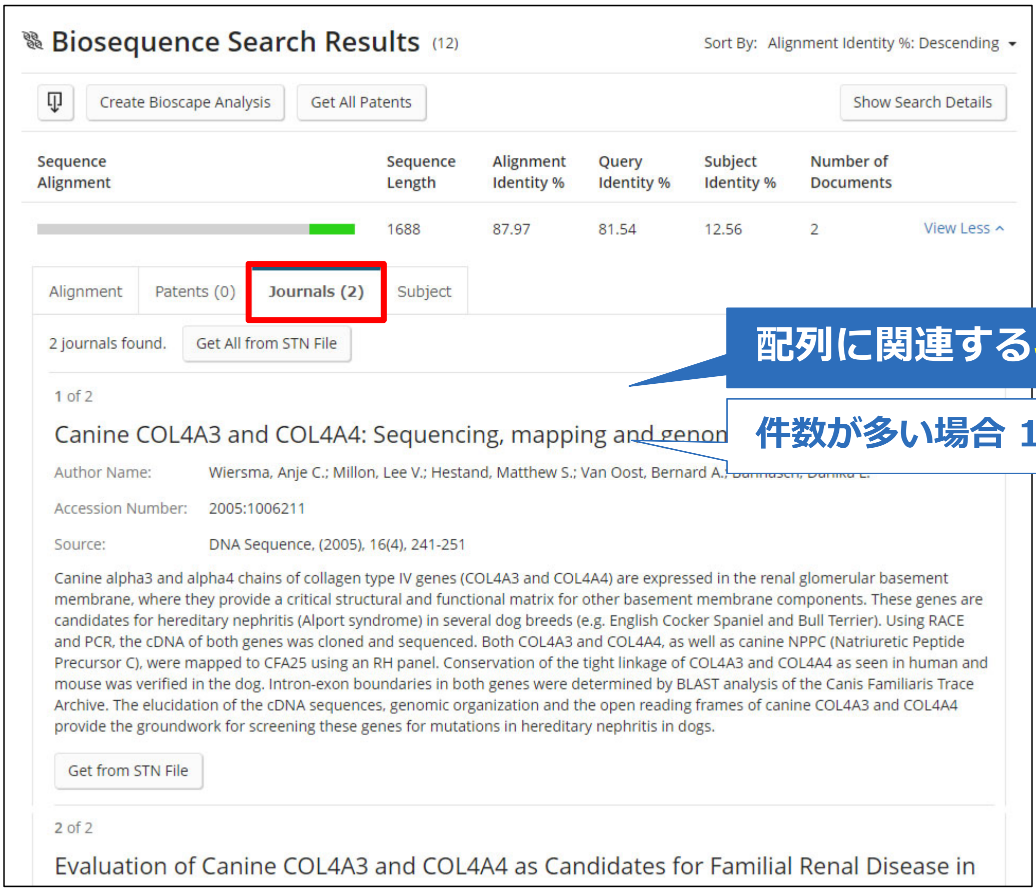
Assignees: MOR RESEARCH APPLICATIONS LTD.; AMIR, ABRAHAM

Patent No.: WO2009153783A1

配列に関連する特許を公報単位で表示

件数が多い場合 10 件のみ表示される

検索結果詳細 – Journals タブ



Biosequence Search Results (12)

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	1688	87.97	81.54	12.56	2	View Less ^

2 journals found.

1 of 2

Canine COL4A3 and COL4A4: Sequencing, mapping and genon

Author Name: Wiersma, Anje C.; Millon, Lee V.; Hestand, Matthew S.; Van Oost, Bernard A.;

Accession Number: 2005:1006211

Source: DNA Sequence, (2005), 16(4), 241-251

Canine alpha3 and alpha4 chains of collagen type IV genes (COL4A3 and COL4A4) are expressed in the renal glomerular basement membrane, where they provide a critical structural and functional matrix for other basement membrane components. These genes are candidates for hereditary nephritis (Alport syndrome) in several dog breeds (e.g. English Cocker Spaniel and Bull Terrier). Using RACE and PCR, the cDNA of both genes was cloned and sequenced. Both COL4A3 and COL4A4, as well as canine NPPC (Natriuretic Peptide Precursor C), were mapped to CFA25 using an RH panel. Conservation of the tight linkage of COL4A3 and COL4A4 as seen in human and mouse was verified in the dog. Intron-exon boundaries in both genes were determined by BLAST analysis of the Canis Familiaris Trace Archive. The elucidation of the cDNA sequences, genomic organization and the open reading frames of canine COL4A3 and COL4A4 provide the groundwork for screening these genes for mutations in hereditary nephritis in dogs.

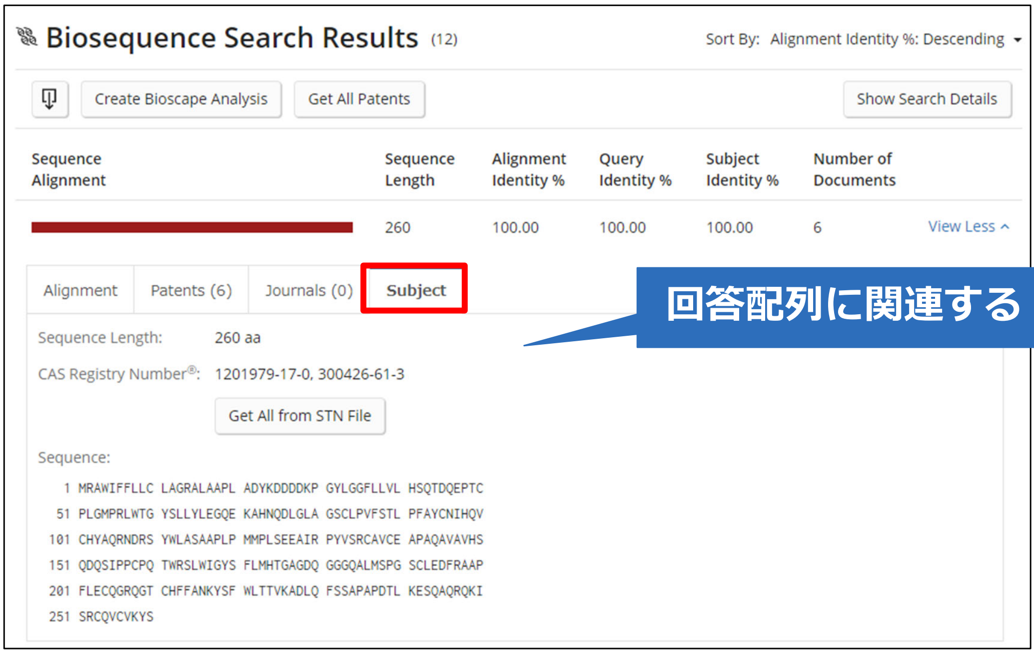
2 of 2

Evaluation of Canine COL4A3 and COL4A4 as Candidates for Familial Renal Disease in

配列に関連する非特許文献を表示

件数が多い場合 10 件のみ表示される

検索結果詳細 – Subject タブ



Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	260	100.00	100.00	100.00	6

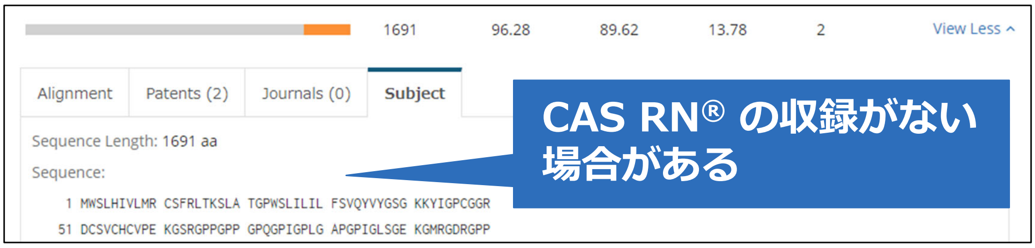
Sequence Length: 260 aa

CAS Registry Number®: 1201979-17-0, 300426-61-3

Sequence:

```
  1 MRAWIFFLLC LAGRALAAPL ADYKDDDDKP GYLGGFLLVL HSQTDQEPTC
 51 PLGMPRLWTG YSLLYLEGQE KAHNQDLGLA GSCLPVFSTL PFAYCNIHQV
101 CHYAQRNDRS YWLASAAPLP MMPLSEEAIR PYVSRCAVCE APAQAVAVHS
151 QDQSIPPCPQ TWRSLWIGYS FLMHTGAGDQ GGGQALMSPG SCLEDFRAAP
201 FLECQGRQGT CHFFANKYSF WLTTVKADLQ FSSAPAPDTL KESQAQRQKI
251 SRCQVCVKYS
```

回答配列に関連する CAS RN® を表示



1691	96.28	89.62	13.78	2

Sequence Length: 1691 aa

Sequence:

```
  1 MWSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYVYGSG KKYIGPCGGR
 51 DCSVCHCVPE KGSRGPPGPP GPQGPIGPLG APGPIGLSGE KGMRGDRGPP
```

CAS RN® の収録がない場合がある

JAICI 化学情報協会 PARTNER OF CAS

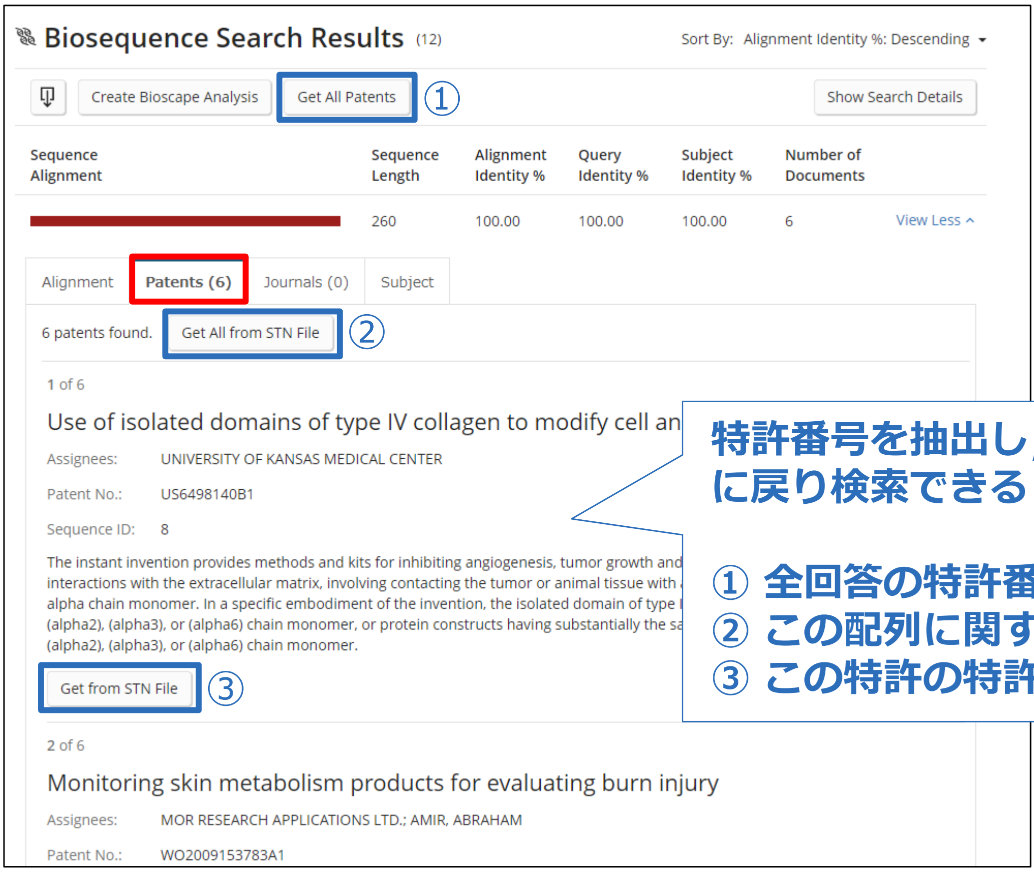
検索結果のダウンロード

Excel で検索結果をダウンロード (上限 1,000 件)

Sequence Length	CAS Registry Number®	Number of Patents	Patent No.	Sequence ID (Patent)	Number of Journals	Accession Number
260	1201979-17-0, 300426-61-3	6	US6498140B1 WO2009153783A1 WO2000059532A1 US9012160B2 US20110151489A1 US6440729B1	8 27 8 27 27 8	0	
260	273191-20-1	0			1	2000:208665
1218	791151-70-7	2	US20070037165A1 US6812339B1	7065 7065	0	
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

特許番号の抽出 - Patents



Biosequence Search Results (12)

Sort By: Alignment Identity %: Descending

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	260	100.00	100.00	100.00	6

6 patents found.

1 of 6

Use of isolated domains of type IV collagen to modify cell an

Assignees: UNIVERSITY OF KANSAS MEDICAL CENTER

Patent No.: US6498140B1

Sequence ID: 8

The instant invention provides methods and kits for inhibiting angiogenesis, tumor growth and interactions with the extracellular matrix, involving contacting the tumor or animal tissue with alpha chain monomer. In a specific embodiment of the invention, the isolated domain of type (alpha2), (alpha3), or (alpha6) chain monomer, or protein constructs having substantially the sa (alpha2), (alpha3), or (alpha6) chain monomer.

2 of 6

Monitoring skin metabolism products for evaluating burn injury

Assignees: MOR RESEARCH APPLICATIONS LTD.; AMIR, ABRAHAM

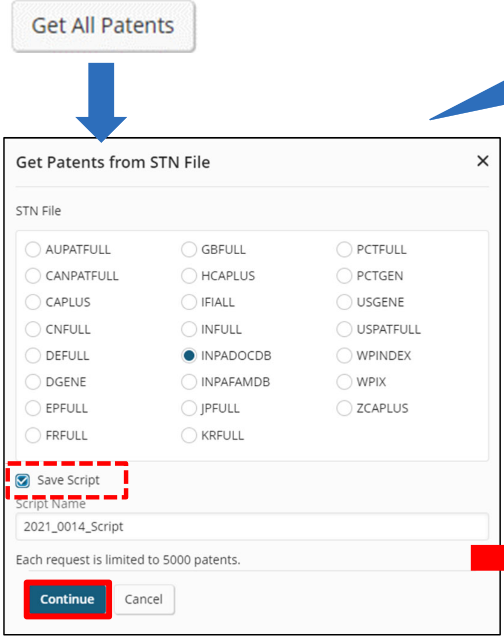
Patent No.: WO2009153783A1

特許番号を抽出し，コマンドライン に戻り検索できる (上限 5,000)

① 全回答の特許番号を抽出
② この配列に関する特許番号を抽出
③ この特許の特許番号を抽出

STN ファイルへのクロスオーバー

特許番号をクロスオーバーする ファイルを選択する (上限 5,000)



Get Patents from STN File

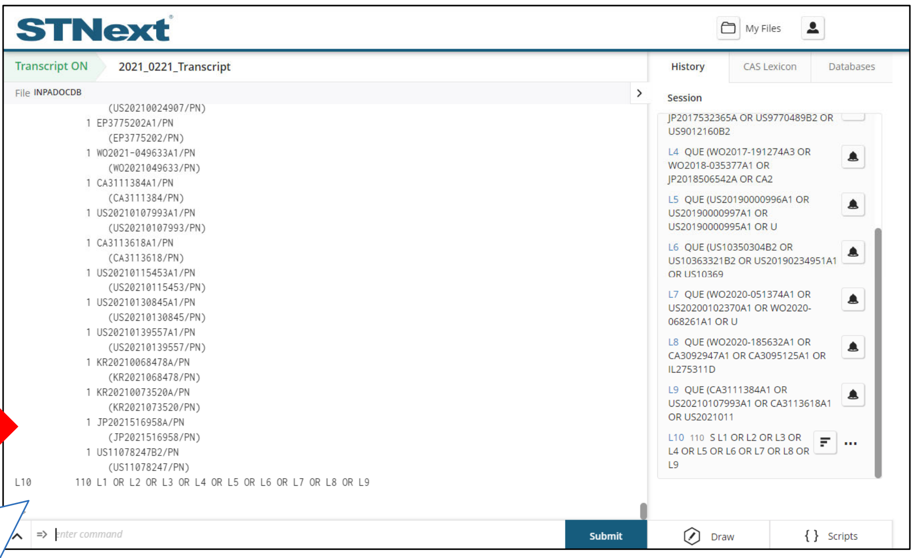
STN File

○ AUPATFULL
○ CANPATFULL
○ CAPLUS
○ CNFULL
○ DEFULL
○ DGENE
○ EPFULL
○ FRFULL
○ GBFULL
○ HCAPLUS
○ IFIALL
○ INFULL
● INPADOCDB
○ INPAFAMDB
○ JPFULL
○ KRFULL
○ PCTFULL
○ PCTGEN
○ USGENE
○ USPATFULL
○ WPINDEX
○ WPIX
○ ZCAPLUS

☑ Save Script

Script Name: 2021_0014_Script

Each request is limited to 5000 patents.



File INPADOCDB

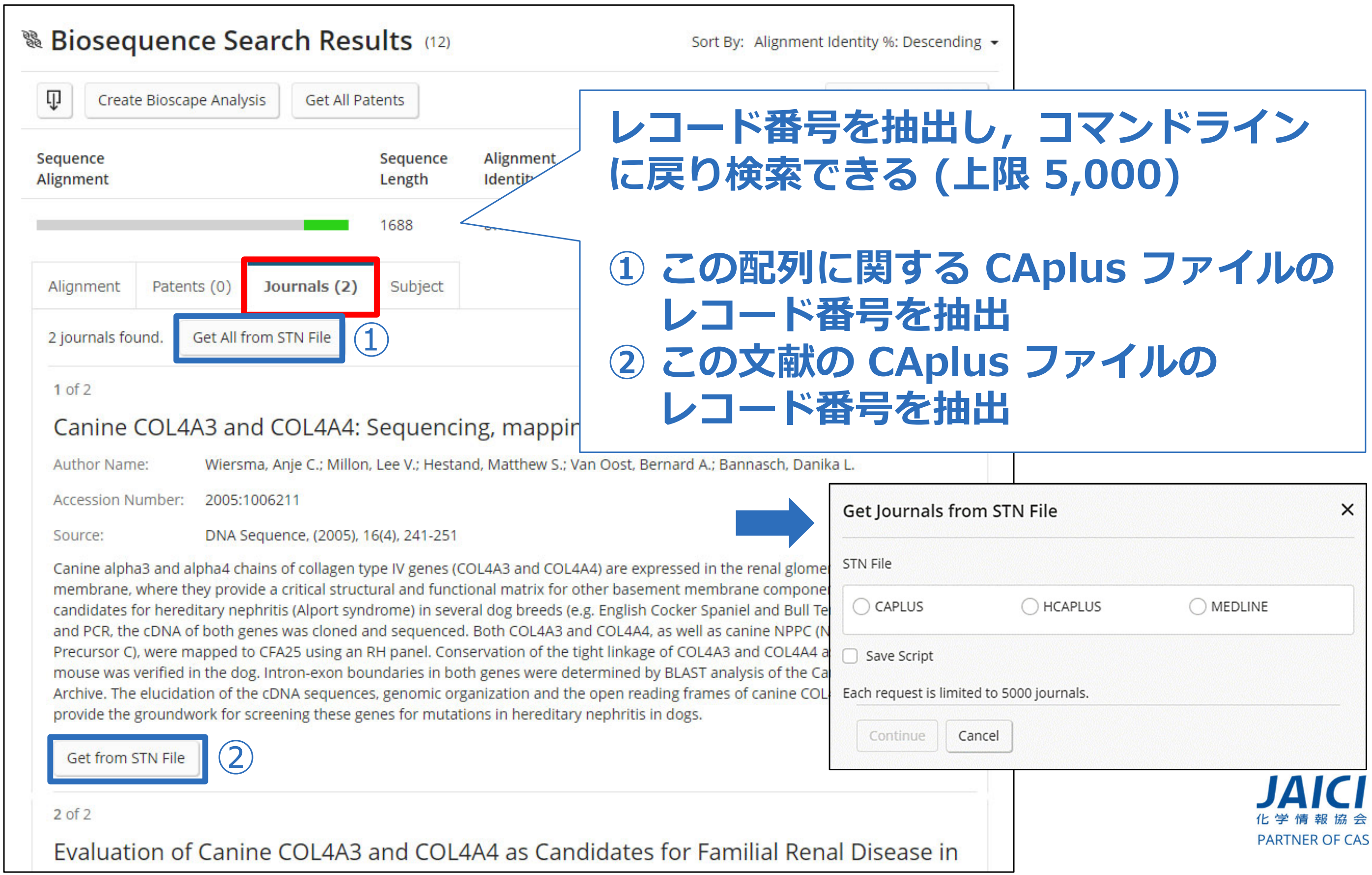
```
                    (US20210024907/PN)
                  1 EP3775202A1/PN
                    (EP3775202/PN)
                  1 WO2021-049633A1/PN
                    (WO2021049633/PN)
                  1 CA3111384A1/PN
                    (CA3111384/PN)
                  1 US20210107993A1/PN
                    (US20210107993/PN)
                  1 CA3113618A1/PN
                    (CA3113618/PN)
                  1 US20210115453A1/PN
                    (US20210115453/PN)
                  1 US20210130845A1/PN
                    (US20210130845/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)
L10           110 L1 OR L2 OR L3 OR L4 OR L5 OR L6 OR L7 OR L8 OR L9

=> enter command
```

Session

JP2017532365A OR US9770489B2 OR US9012160B2
L4 QUE (WO2017-191274A3 OR WO2018-035377A1 OR JP2018506542A OR CA2
L5 QUE (US20190000996A1 OR US20190000997A1 OR US20190000995A1 OR U
L6 QUE (US10350304B2 OR US10363321B2 OR US20190234951A1 OR US10369
L7 QUE (WO2020-051374A1 OR US20200102370A1 OR WO2020-068261A1 OR U
L8 QUE (WO2020-185632A1 OR CA3092947A1 OR CA3095125A1 OR IL275311D
L9 QUE (CA3111384A1 OR US20210107993A1 OR CA3113618A1 OR US2021011
L10 110 S L1 OR L2 OR L3 OR L4 OR L5 OR L6 OR L7 OR L8 OR L9

Save Script にチェックし，特許番号の検索式をスクリプトとして 保存することも可能 (My Files - Scripts に保存)

レコード番号の抽出 - Journals



Biosequence Search Results (12)

Sort By: Alignment Identity %: Descending

2 journals found. Get All from STN File ①

1 of 2

Canine COL4A3 and COL4A4: Sequencing, mappin

Author Name: Wiersma, Anje C.; Millon, Lee V.; Hestand, Matthew S.; Van Oost, Bernard A.; Bannasch, Danika L.

Accession Number: 2005:1006211

Source: DNA Sequence, (2005), 16(4), 241-251

Canine alpha3 and alpha4 chains of collagen type IV genes (COL4A3 and COL4A4) are expressed in the renal glome membrane, where they provide a critical structural and functional matrix for other basement membrane compone candidates for hereditary nephritis (Alport syndrome) in several dog breeds (e.g. English Cocker Spaniel and Bull Te and PCR, the cDNA of both genes was cloned and sequenced. Both COL4A3 and COL4A4, as well as canine NPPC (N Precursor C), were mapped to CFA25 using an RH panel. Conservation of the tight linkage of COL4A3 and COL4A4 a mouse was verified in the dog. Intron-exon boundaries in both genes were determined by BLAST analysis of the Ca Archive. The elucidation of the cDNA sequences, genomic organization and the open reading frames of canine COL provide the groundwork for screening these genes for mutations in hereditary nephritis in dogs.

Get from STN File ②

2 of 2

Evaluation of Canine COL4A3 and COL4A4 as Candidates for Familial Renal Disease in

レコード番号を抽出し，コマンドラインに戻り検索できる (上限 5,000)

① この配列に関する CAplus ファイルのレコード番号を抽出
② この文献の CAplus ファイルのレコード番号を抽出

Get Journals from STN File

STN File

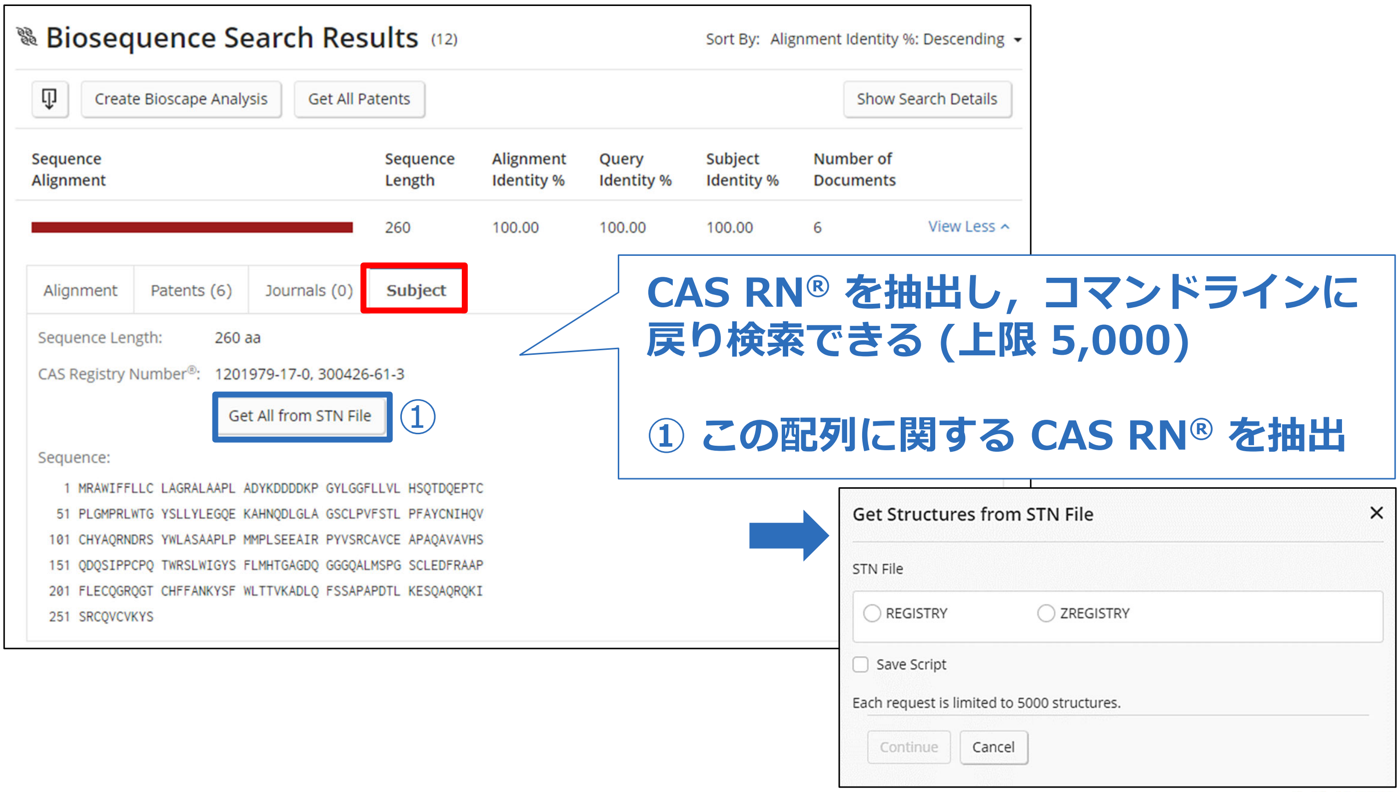
○ CAPLUS ○ HCAPLUS ○ MEDLINE

☐ Save Script

Each request is limited to 5000 journals.

JAICI 化学情報協会
PARTNER OF CAS

CAS RN® の抽出 – 回答配列



Biosequence Search Results (12)

Sort By: Alignment Identity %: Descending

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	260	100.00	100.00	100.00	6	View Less

Sequence Length: 260 aa

CAS Registry Number®: 1201979-17-0, 300426-61-3

Get All from STN File ①

Sequence:

```
  1 MRAWIFFLLC LAGRALAAPL ADYKDDDDKP GYLGGFLLVL HSQTDQEPTC
 51 PLGMPRLWTG YSLLYLEGQE KAHNQDLGLA GSCLPVFSTL PFAYCNIHQV
101 CHYAQRNDRS YWLASAAPLP MMPLSEEAIR PYVSRCAVCE APAQAVAVHS
151 QDQSIPPCPQ TWRSLWIGYS FLMHTGAGDQ GGGQALMSPG SCLEDFRAAP
201 FLECQGRQGT CHFFANKYSF WLTTVKADLQ FSSAPAPDTL KESQAQRQKI
251 SRCQVCVKYS
```

CAS RN® を抽出し，コマンドラインに戻り検索できる (上限 5,000)

① この配列に関する CAS RN® を抽出

Get Structures from STN File

STN File

○ REGISTRY ○ ZREGISTRY

☐ Save Script

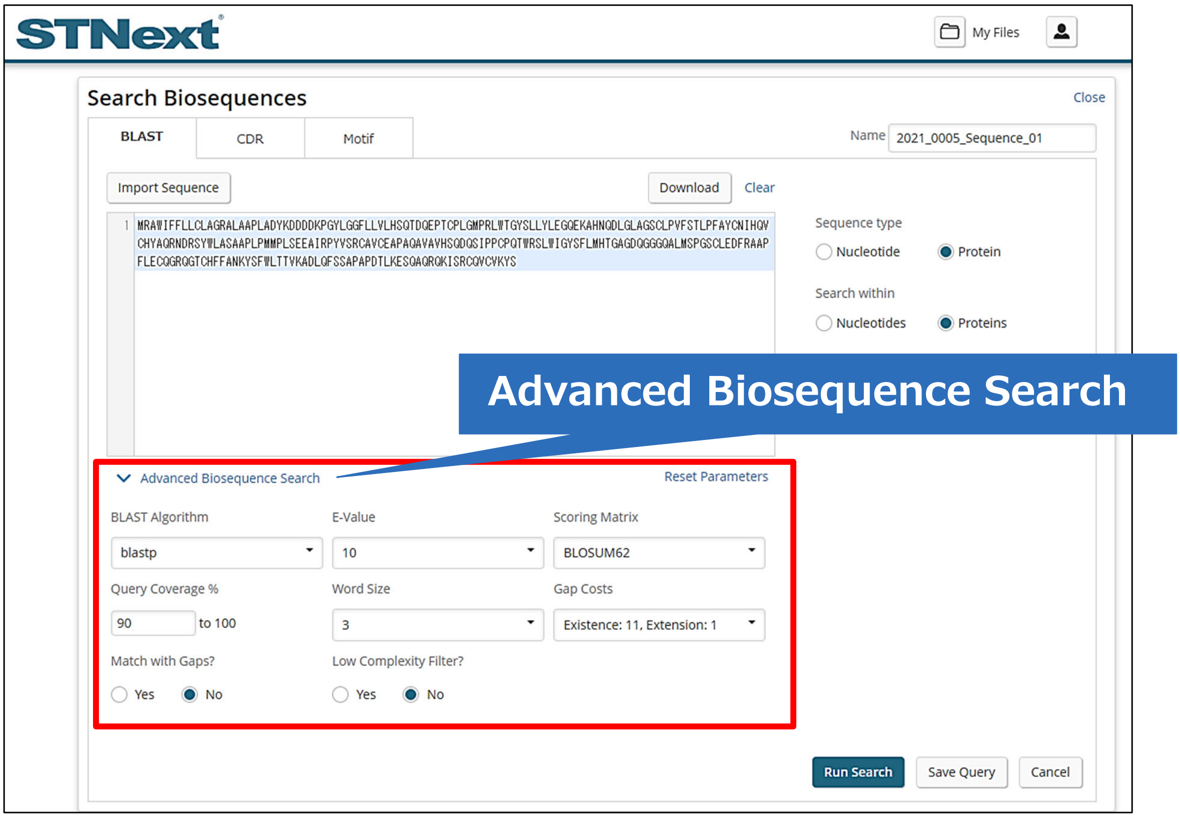
Each request is limited to 5000 structures.

JAICI 化学情報協会
PARTNER OF CAS

3. 検索のヒント

Advanced Sequence Search

パラメータや検索条件の設定を確認・変更できる



Name 2021_0005_Sequence_01

```
MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV
CHYAQRNDRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSIPPCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAP
FLECQGRQGTCHFFANKYSFWLTTVKADLQFSSAPAPDTLKESQAQRQKISRCQVCVKYS
```

Sequence type ○ Nucleotide ● Protein

Search within ○ Nucleotides ● Proteins

Advanced Biosequence Search

Advanced Biosequence Search

BLAST Algorithm: blastp

E-Value: 10

Scoring Matrix: BLOSUM62

Query Coverage %: 90 to 100

Word Size: 3

Gap Costs: Existence: 11, Extension: 1

Match with Gaps? ○ Yes ● No

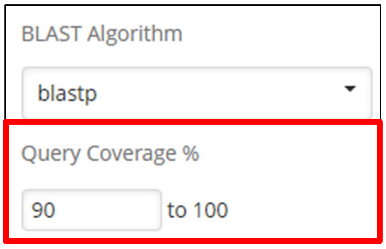
Low Complexity Filter? ○ Yes ● No

JAICI 化学情報協会
PARTNER OF CAS

1．計算値の設定

以下の計算値が低い回答も含めるか検討する

■ BLASTp デフォルト



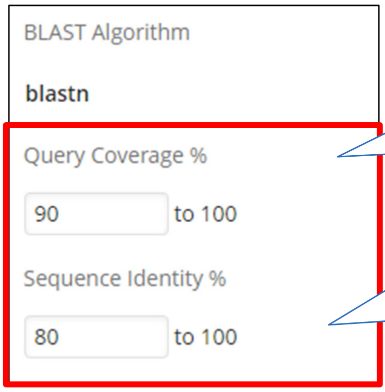
BLAST Algorithm

blastp

Query Coverage %

90 to 100

■ BLASTn デフォルト



BLAST Algorithm

blastn

Query Coverage %

90 to 100

Sequence Identity %

80 to 100

Query Coverage %
類似領域÷配列質問式×100

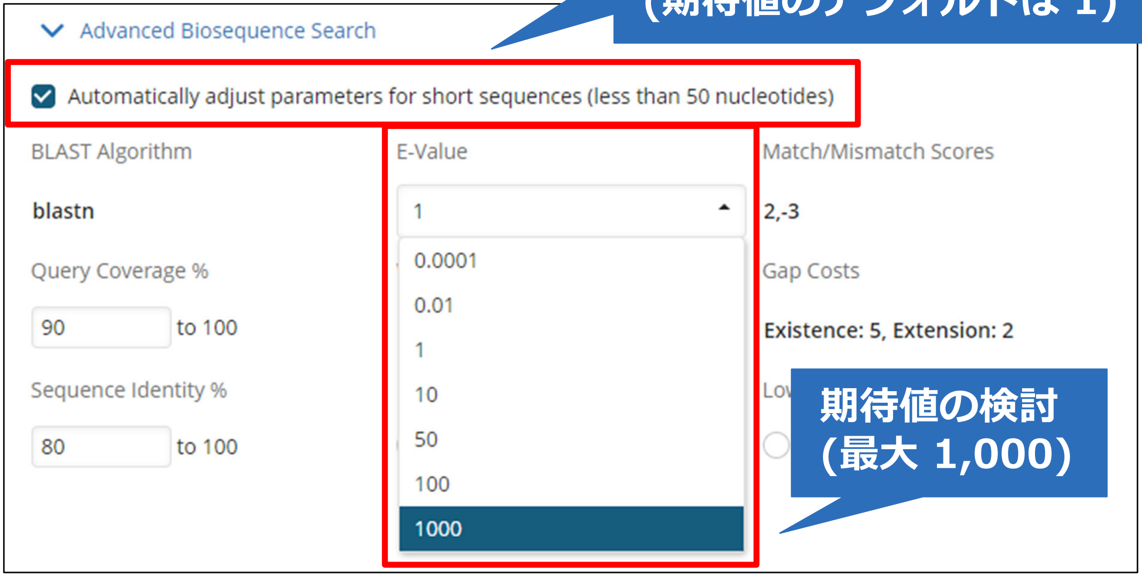
Sequence Identity %
(Alignment Identity%)
一致したコード÷類似領域×100

2．短い配列検索用の設定

短い配列質問式用のパラメータでうまく検索できない場合は，期待値を大きくする

(例) BLASTn

50 未満の核酸配列質問式の検索に適したパラメータに変更
(期待値のデフォルトは 1)



Advanced Biosequence Search

☑ Automatically adjust parameters for short sequences (less than 50 nucleotides)

BLAST Algorithm

blastn

Query Coverage %

90 to 100

Sequence Identity %

80 to 100

E-Value

1

0.0001
0.01
1
10
50
100
1000

Match/Mismatch Scores

2,-3

Gap Costs

Existence: 5, Extension: 2

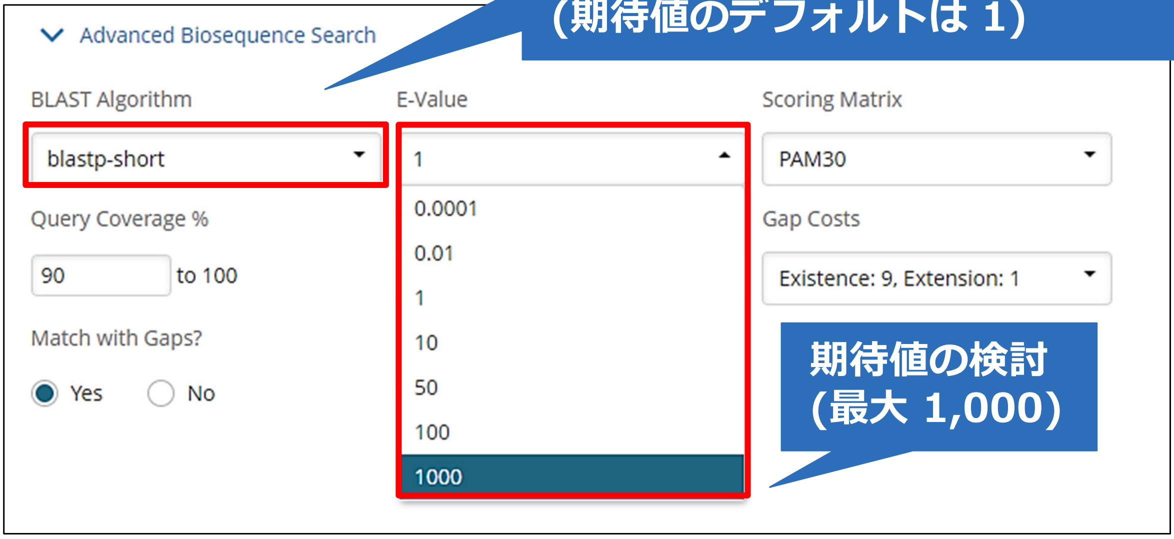
期待値の検討
(最大 1,000)

NCBI FAQs Q: What is the Expect (E) value?
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect

2．短い配列検索用の設定

短い配列質問式用のパラメータでうまく検索できない場合は，期待値を大きくする

(例) BLASTp-short



Advanced Biosequence Search

BLAST Algorithm: blastp-short

E-Value: 1

- 0.0001
- 0.01
- 1
- 10
- 50
- 100
- 1000

Scoring Matrix: PAM30

Query Coverage %: 90 to 100

Gap Costs: Existence: 9, Extension: 1

Match with Gaps? ● Yes ○ No

30 未満のタンパク質配列質問式の検索に適したパラメータに変更（期待値のデフォルトは 1）

期待値の検討（最大 1,000）

NCBI FAQs Q: What is the Expect (E) value?
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ#expect

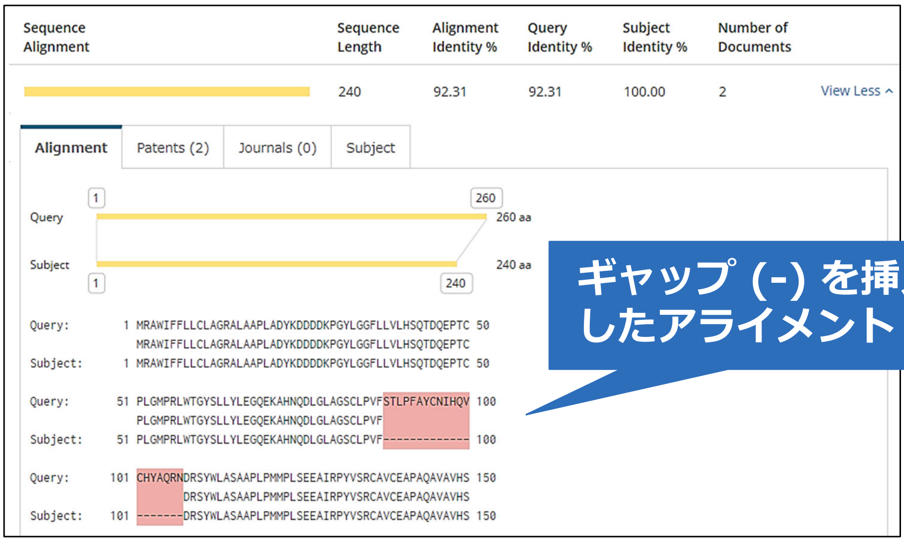
---

3．Match with Gaps?

ギャップを挿入してアライメントした回答も含める場合は Match with Gaps? を Yes にする

(例) BLASTp-short

Match with Gaps? ● Yes ○ No



Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	240	92.31	92.31	100.00	2

```
Query:       1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC 50
               MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC
Subject:     1 MRAWIFFLLCLAGRALAAPLADYKDDDDKPGYLGGFLLVLHSQTDQEPTC 50

Query:      51 PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQV 100
               PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVF
Subject:    51 PLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVF------------- 100

Query:     101 CHYAQRNDRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHS 150
                      DRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHS
Subject:   101 -------DRSYWLASAAPLPMMPLSEEAIRPYVSRCAVCEAPAQAVAVHS 150
```

ギャップ (-) を挿入したアライメント

4. CDR 配列検索，Motif 配列検索

CDR 配列検索

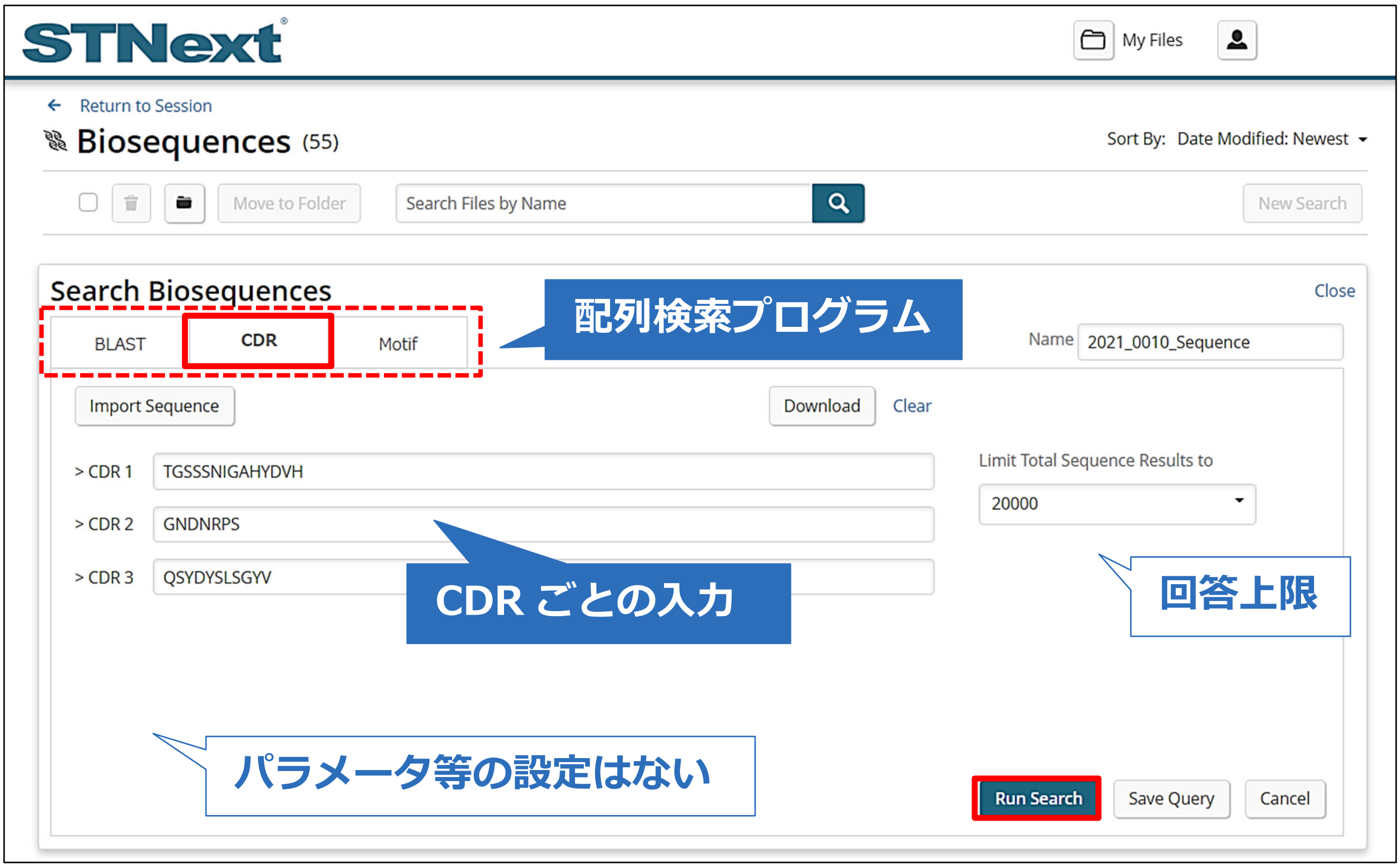
CDR (相補性決定領域) を指定し検索するプログラム

- 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

抗原に特異的な重要な部分を CDR (相補性決定領域) という

抗体や T 細胞受容体にある CDR の配列を複数指定し，検索できる

初期画面 - CDR



Name: 2021_0010_Sequence

> CDR 1 TGSSSNIGAHYDVH

> CDR 2 GNDNRPS

> CDR 3 QSYDYSLSGYV

Limit Total Sequence Results to: 20000

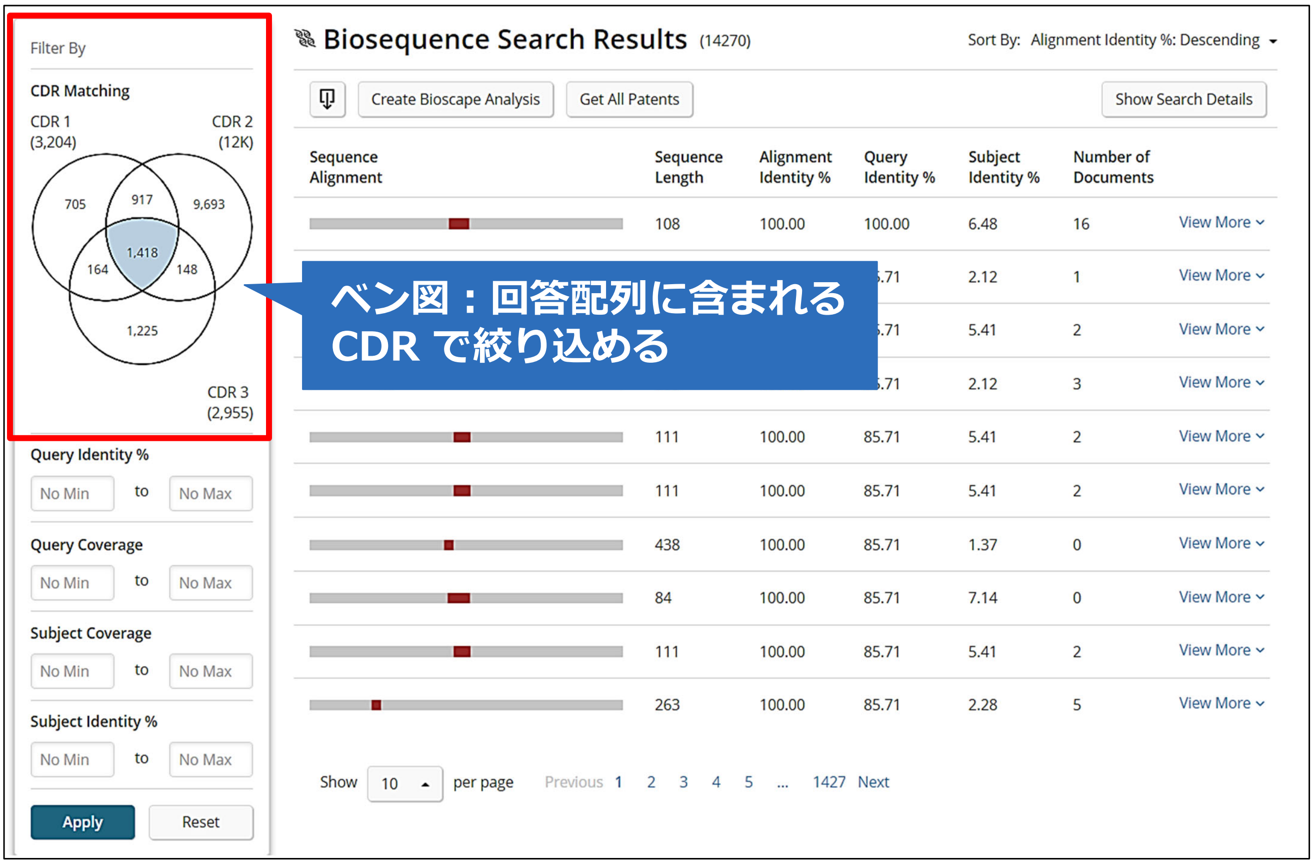
配列検索プログラム

CDR ごとの入力

回答上限

パラメータ等の設定はない

検索結果一覧 - CDR



Biosequence Search Results (14270)

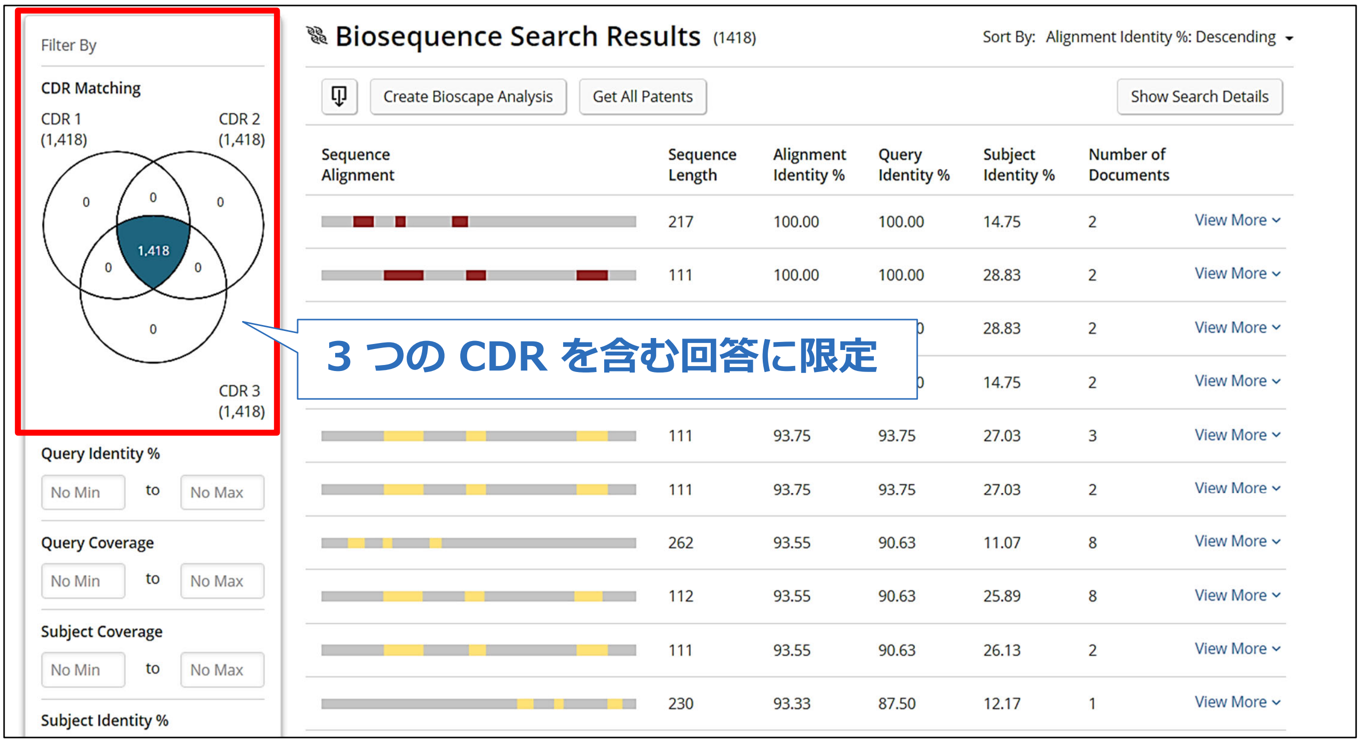
Sort By: Alignment Identity %: Descending

Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
	108	100.00	100.00	6.48	16
	[illegible]	[illegible]	.71	2.12	1
	[illegible]	[illegible]	.71	5.41	2
	[illegible]	[illegible]	.71	2.12	3
	111	100.00	85.71	5.41	2
	111	100.00	85.71	5.41	2
	438	100.00	85.71	1.37	0
	84	100.00	85.71	7.14	0
	111	100.00	85.71	5.41	2
	263	100.00	85.71	2.28	5

Show 10 per page

ベン図：回答配列に含まれる
CDR で絞り込める

検索結果一覧 - CDR



Query Identity %: No Min to No Max

Query Coverage: No Min to No Max

Subject Coverage: No Min to No Max

Subject Identity %

Biosequence Search Results (1418)

Sort By: Alignment Identity %: Descending

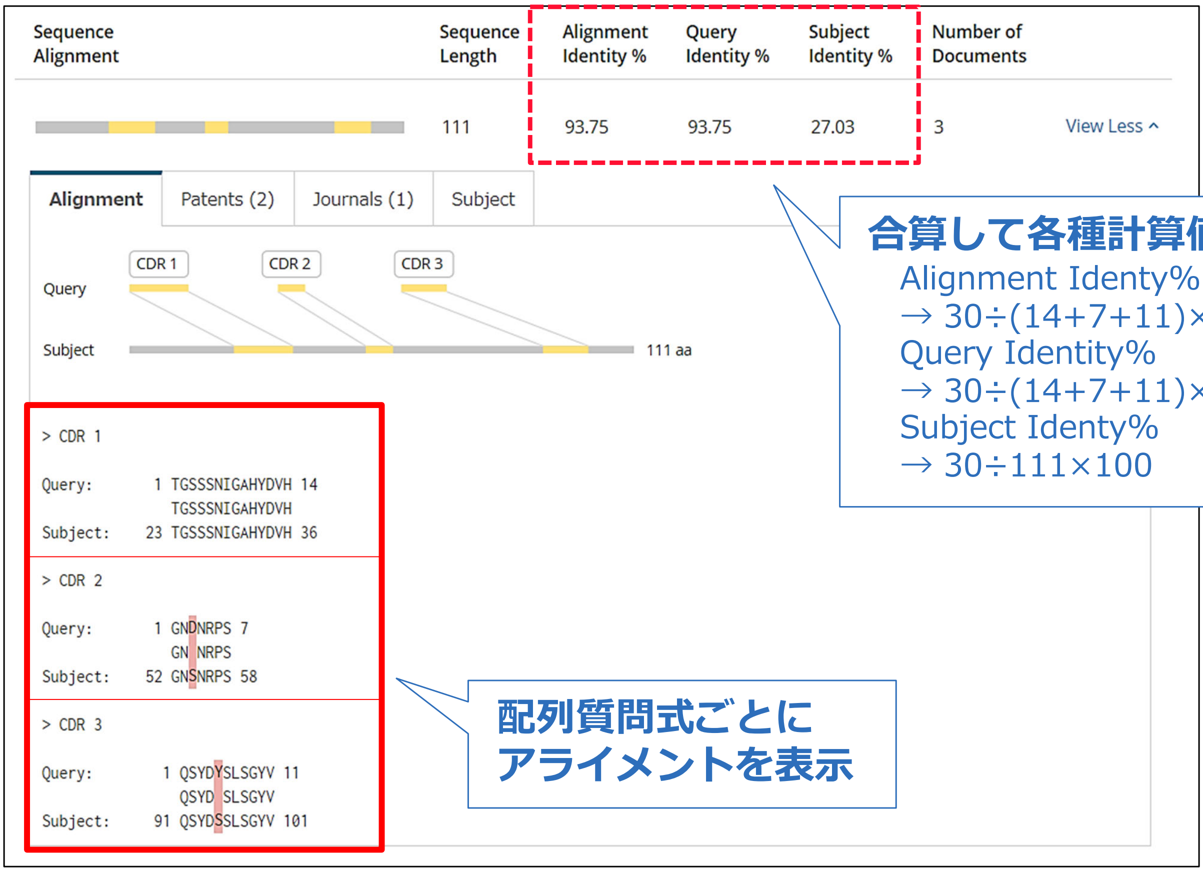
Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	217	100.00	100.00	14.75	2	View More
	111	100.00	100.00	28.83	2	View More
				28.83	2	View More
				14.75	2	View More
	111	93.75	93.75	27.03	3	View More
	111	93.75	93.75	27.03	2	View More
	262	93.55	90.63	11.07	8	View More
	112	93.55	90.63	25.89	8	View More
	111	93.55	90.63	26.13	2	View More
	230	93.33	87.50	12.17	1	View More

3 つの CDR を含む回答に限定

CAS
A division of the
American Chemical Society

JAICI
化学情報協会
PARTNER OF CAS

検索結果詳細 - CDR



Sequence Alignment	Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents	
	111	93.75	93.75	27.03	3	View Less

```
> CDR 1

Query:      1 TGSSSNIGAHYDVH 14
              TGSSSNIGAHYDVH
Subject:   23 TGSSSNIGAHYDVH 36
```

```
> CDR 2

Query:      1 GNDNRPS 7
              GN NRPS
Subject:   52 GNSNRPS 58
```

```
> CDR 3

Query:       1 QSYDYSLSGYV 11
               QSYD SLSGYV
Subject:    91 QSYDSSLSGYV 101
```

合算して各種計算値を表示

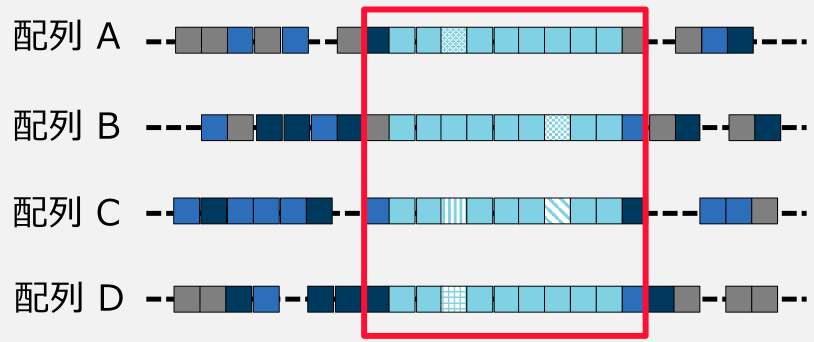
Alignment Identy%
→ 30÷(14+7+11)×100
Query Identity%
→ 30÷(14+7+11)×100
Subject Identy%
→ 30÷111×100

**配列質問式ごとに
アライメントを表示**

Motif 配列検索

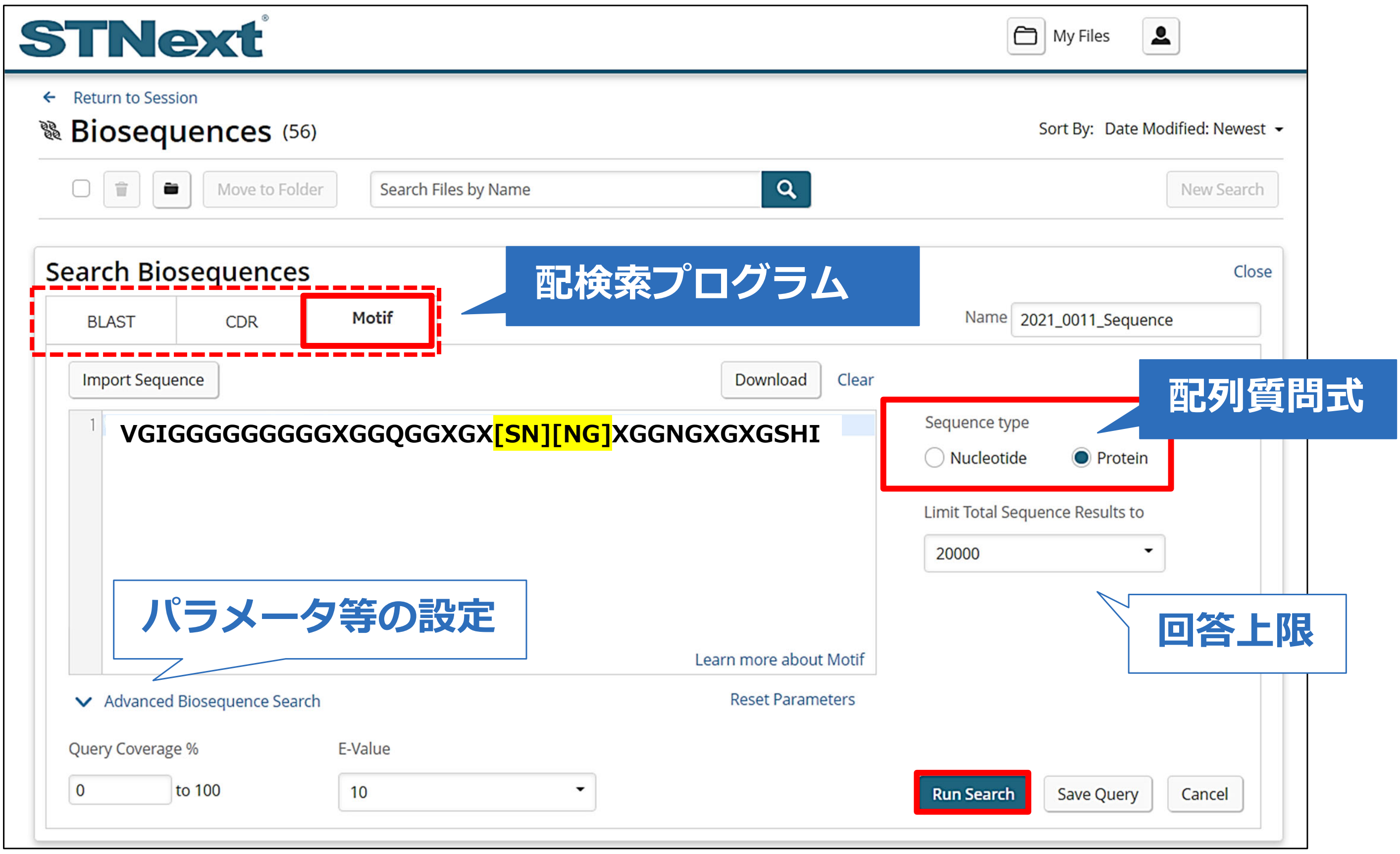
パターン配列を検索するプログラム

- 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



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

初期画面 - Motif



Name: 2021_0011_Sequence

VGIGGGGGGGGXGGQGGXGX[SN][NG]XGGNGXGXGSHI

Sequence type: ○ Nucleotide ● Protein

Limit Total Sequence Results to: 20000

Query Coverage %: 0 to 100

E-Value: 10

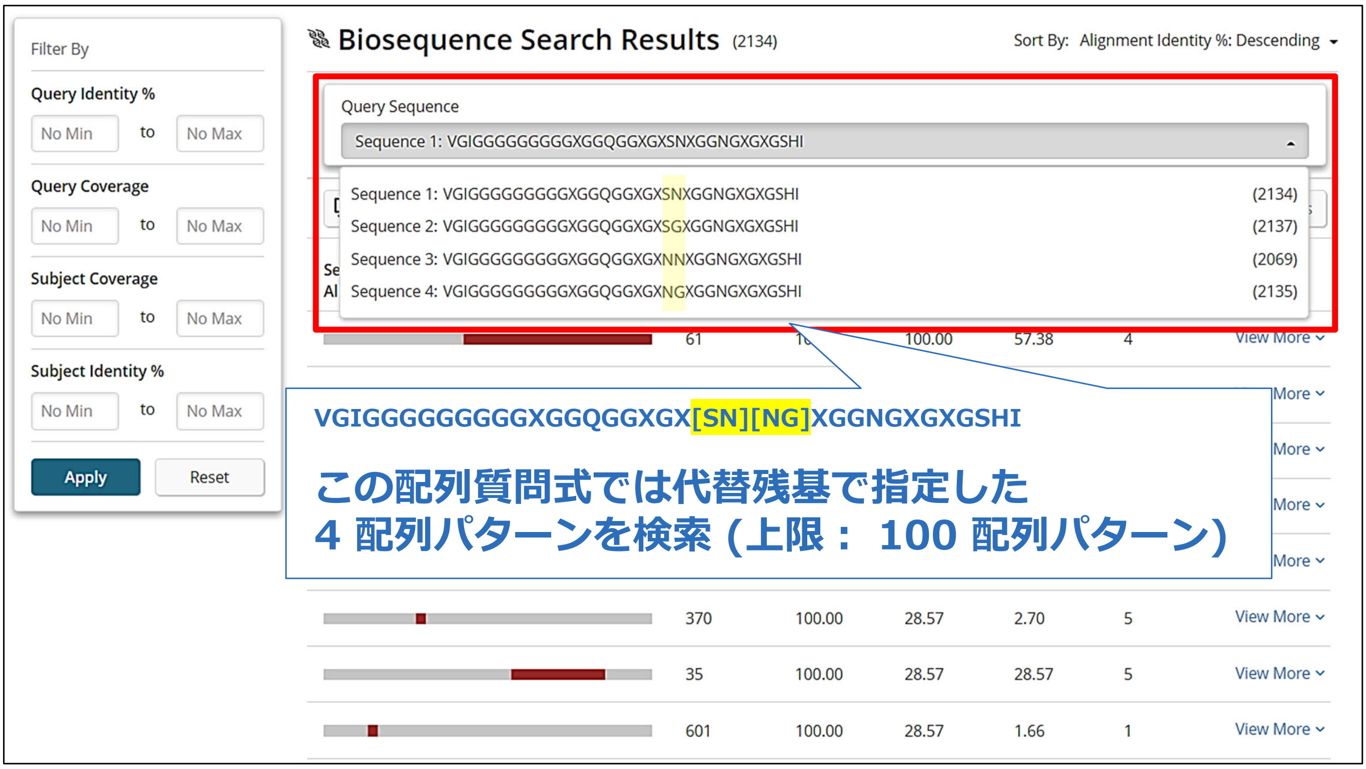
配検索プログラム

配列質問式

パラメータ等の設定

回答上限

検索結果一覧 - Motif



Sequence 1: VGIGGGGGGGGGXGGQGGXGXSNXGGNGXGXGSHI

Sequence 1: VGIGGGGGGGGGXGGQGGXGXSNXGGNGXGXGSHI	(2134)
Sequence 2: VGIGGGGGGGGGXGGQGGXGXSGXGGNGXGXGSHI	(2137)
Sequence 3: VGIGGGGGGGGGXGGQGGXGXNNXGGNGXGXGSHI	(2069)
Sequence 4: VGIGGGGGGGGGXGGQGGXGXNGXGGNGXGXGSHI	(2135)

61	10	100.00	57.38	4	View More
370	100.00	28.57	2.70	5	View More
35	100.00	28.57	28.57	5	View More
601	100.00	28.57	1.66	1	View More

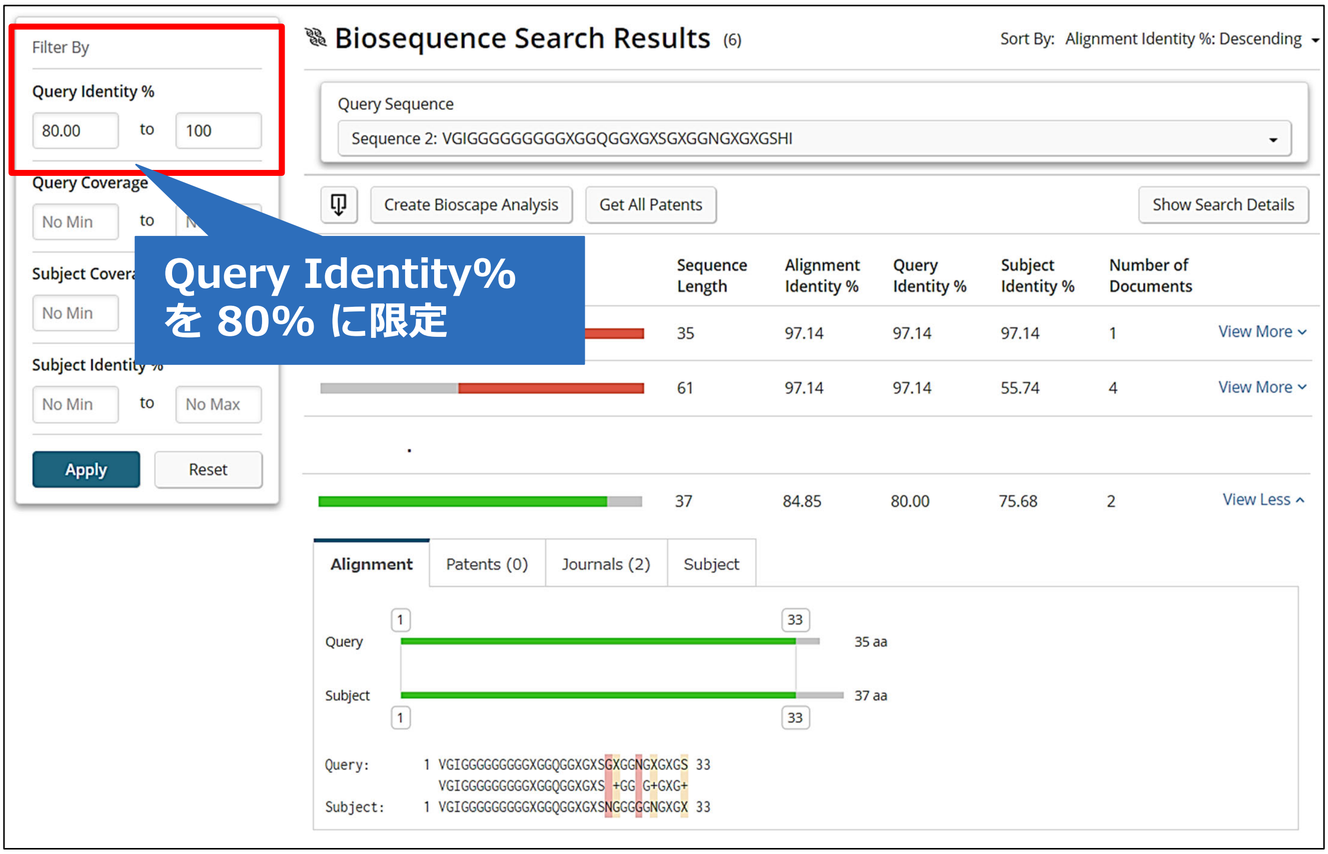
VGIGGGGGGGGGXGGQGGXGX[SN][NG]XGGNGXGXGSHI

この配列質問式では代替残基で指定した
4 配列パターンを検索 (上限： 100 配列パターン)

A division of the American Chemical Society

化学情報協会
PARTNER OF CAS

検索結果詳細 - Motif



Query Identity %: 80.00 to 100

Sequence 2: VGIGGGGGGGGGXGGQGGXGXSGXGGNGXGXGSHI

Sequence Length	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
35	97.14	97.14	97.14	1
61	97.14	97.14	55.74	4
37	84.85	80.00	75.68	2

```
Query:      1 VGIGGGGGGGGGXGGQGGXGXSGXGGNGXGXGS 33
              VGIGGGGGGGGGXGGQGGXGXS +GG G+GXG+
Subject:    1 VGIGGGGGGGGGXGGQGGXGXSNGGGGNGXGX 33
```

Query Identity%
を 80% に限定

Motif 検索で利用できる記号

記号	内容	入力例	回答例
. (ピリオド)	1 残基コードが入る ・タンパク質の場合 X に翻訳し検索 ・核酸の場合 A, G, C, T の 4 パターン検索	S....GKD	SXXXXGKD SFTSYGKD
[ ]	代替残基の指定	[SG]XXXXGKD	SXXXXGKD GXXXXGKD
^XYZ$	指定した配列に完全一致するを検索する	^SXXXXGKD$	SXXXXGKD
{m,n}	直前の配列を m～n 回繰り返す	SX{3,4}GKD	SXXXGKD SXXXXGKD
{n}	直前の配列を n 繰り返す	SX{4}GKD	SXXXXGKD

Learn more about Motif (Quick Reference Card)
https://www.cas.org/sites/default/files/documents/stnextmotif.pdf

---

Motif 検索で利用できるその他のコード

■ タンパク質

コード	内容
X	任意のアミノ酸, Uncommon amino acids もヒット
B	D or N (Aspartic acid or Asparagine)
Z	E or Q (Glutamic acid or Glutamine)
J	I or L (Isoleucine or Leucine)

■ 核酸

記号	内容
N	A or C or G or T
R	A or G
Y	C or T
M	A or C
K	G or T
S	C or G
W	A or T
H	A or C or T
B	C or G or T
V	A or C or G
D	A or G or T