

CAS Sequences

検索ガイド

2022 年 03 月



* 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

CAS Sequences の仕様が変更され、回答中の Patents/Journals タブがなくなりました。
得られた配列の参考文献情報はデータをダウンロードしてご確認ください。



STNext Biosequences Search 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
- NCBI 由来の配列

New

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	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	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

New

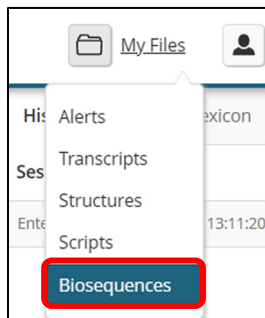
New

・特長

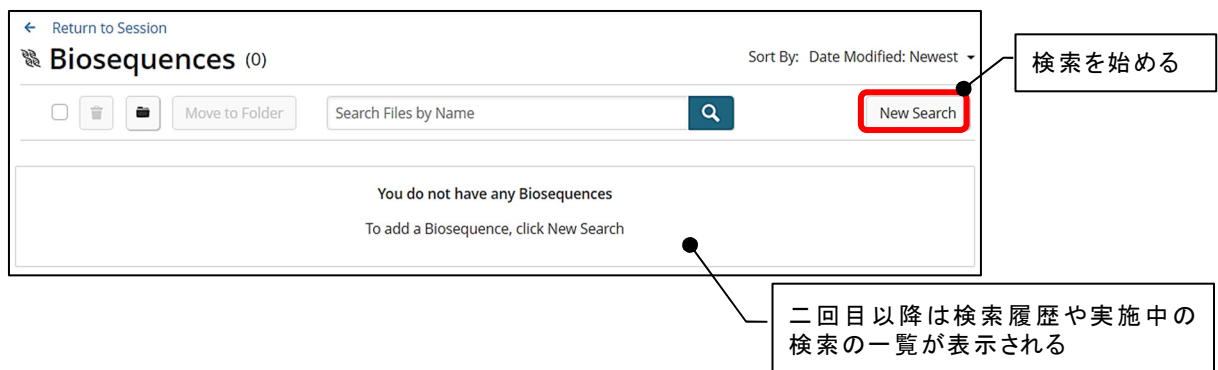
- アライメントの概要図をカラーでわかりやすく表示。Excel 形式でのダウンロードも可能。
- 絞り込みやソートが簡単にできる。
- 配列情報と共に配列の由来となった文献などの情報も表示。
- 配列検索用の特別なソフトウェアのインストール不要。

初期画面

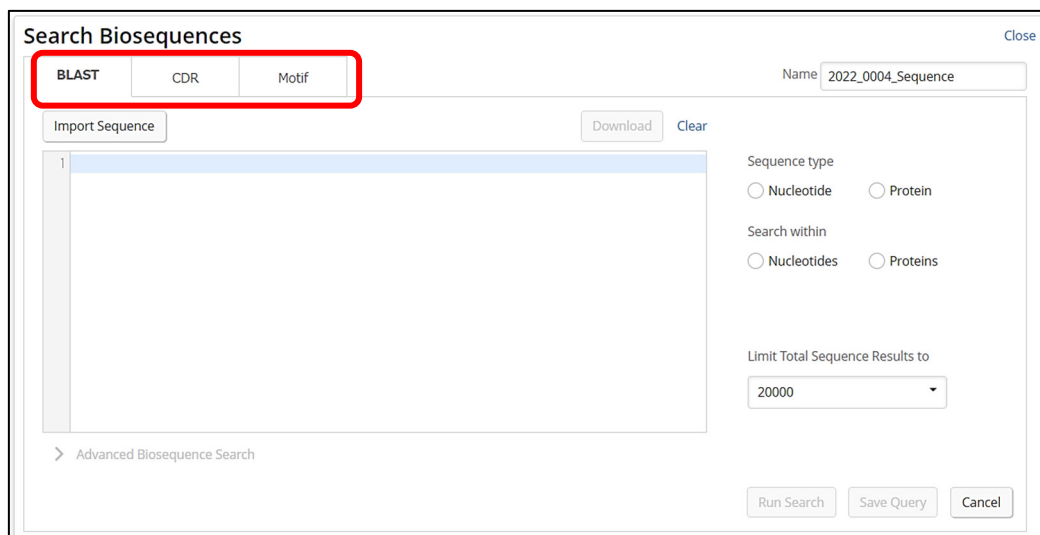
- 画面右上 My Files から Biosequences を選択します。



- 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' window. The 'BLAST' tab is selected. The 'Name' field contains '2022_0004_Sequence'. The 'Import Sequence' button is highlighted with a red box and a callout: '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)'. The 'Sequence type' section has 'Protein' selected for both 'Sequence type' and 'Search within'. The 'Limit Total Sequence Results to' is set to '20000'. The 'Run Search' button is highlighted with a red box and a callout: '検索の実行'. The 'Advanced Biosequence Search' link is highlighted with a red dashed box and a callout: 'パラメータの変更 (任意)'. Other callouts include '名前の変更 (任意)' pointing to the 'Name' field, '質問式の配列' pointing to the 'Sequence type' section, and '回答の配列' pointing to the 'Search within' section.

■ 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 section shows the search progress with a 'Cancel Search' button. The bottom section shows the search results with a 'View Results' button highlighted with a red box. The 'View Results' button has a callout: '検索結果の確認'. The 'View' button next to the sequence list has a callout: '配列質問式や検索条件の確認'. The 'View Results' button also shows '5226 results'.

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。

The screenshot shows the 'Biosequence Search Results' page for 5226 results. On the left, a 'Filter By' sidebar includes sections for 'Query Identity %', 'Query Coverage', 'Subject Coverage', 'Subject Identity %', and 'Organism'. The 'Organism' section is highlighted with a blue dashed box and labeled '絞り込みフィルター' (Filter). The main table has columns: 'Sequence Alignment', 'Sequence Length', 'Alignment Identity %', 'Query Identity %', 'Subject Identity %', and 'Number of Documents'. The 'Sequence Alignment' column is highlighted with a red box and labeled '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' (Diagram of alignment for hit sequences). The 'Alignment Identity %', 'Query Identity %', and 'Subject Identity %' columns are highlighted with a red box and labeled '計算値' (Calculation values). A 'Sort By' dropdown at the top right is set to 'Alignment Identity %: Descending' and labeled 'ソート' (Sort). At the bottom left, a list of organisms is highlighted with a blue dashed box and labeled '生物名による絞り込み (NCBI 由来の配列に収録)' (Filter by organism name (sequences from NCBI)).

- ・ 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%)

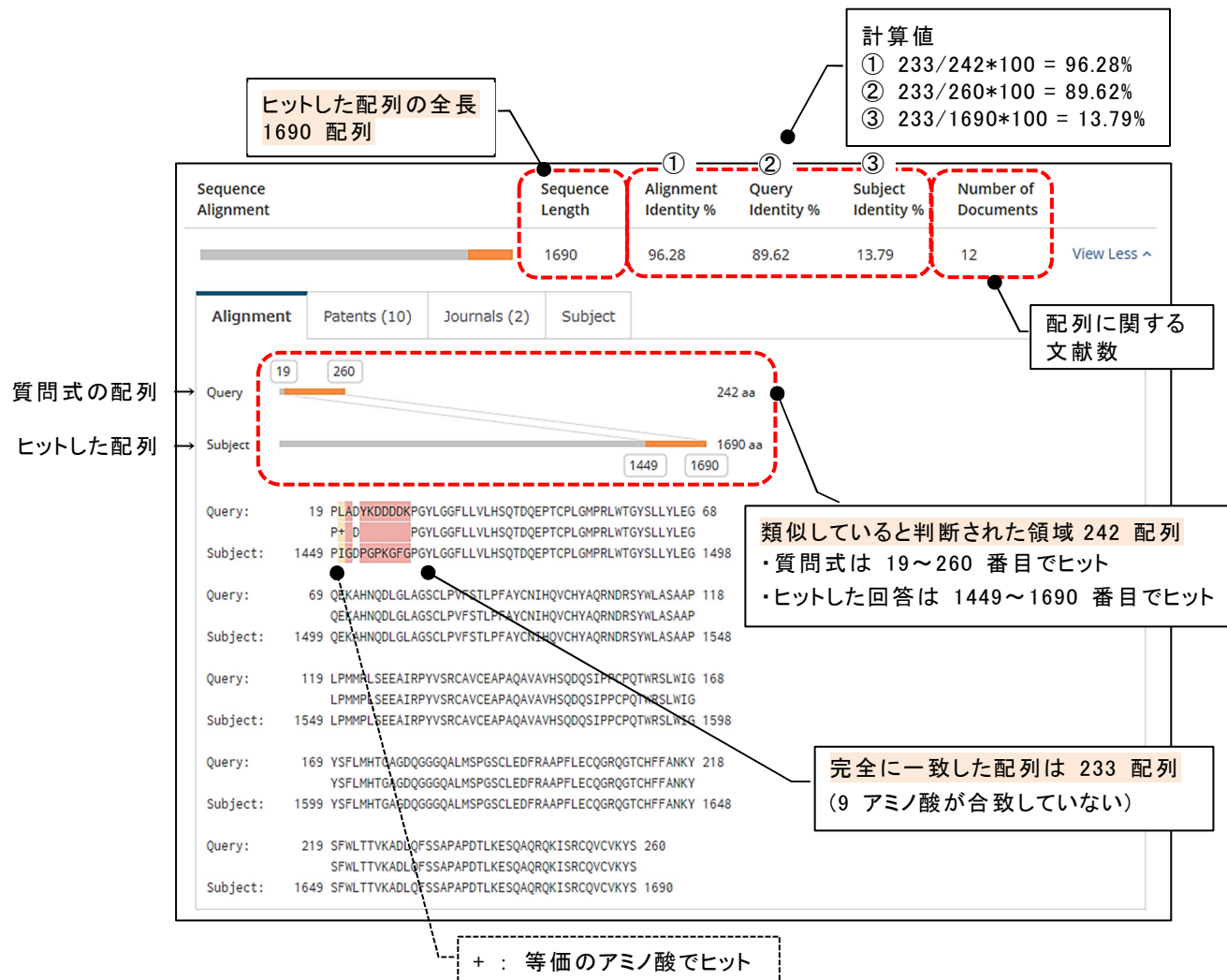
- ・ 各項目の計算値による絞り込みやソートが可能です。

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



CAS Sequences の仕様が変更され、回答中の Patents/Journals タブがなくなりました。
得られた配列の参考文献情報はデータをダウンロードしてご確認ください。

- 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 | View Less

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

10 patents found. **Get All Patent Numbers**

STNext のコマンドラインに戻り、全回答の特許番号を検索する

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.

Get Patent Number

STNext のコマンドラインに戻り、この特許番号を検索する

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 CAlplus 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. **Get All Accession Numbers** **Get All PubMed IDs**

STNext のコマンドラインに戻り、表示している配列に関する文献を検索する

1 of 2

Complete primary structure of the human alpha4(IV) collagen 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 alpha4(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 alpha4(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 alpha4(IV) chain contains 31 cysteine residues of which 18 are conserved in the other type IV collagen alpha chains. The calculated mol. weight of the mature alpha4(IV) chain is 164,123. Anal. of the primary structure showed that the alpha4(IV) chain belongs to the alpha2-like type IV collagen chains together with alpha2(IV) and alpha6(IV). Northern analyses with RNA from several human fetal tissues revealed quite similar expression patterns for the alpha4(IV) and alpha3(IV) chains, but there were also distinct differences in some tissues. The expression patterns of alpha5(IV) and alpha5(IV) differed extensively between each other and they also differed from those of alpha3(IV) and alpha4(IV).

Get Accession Number **Get PubMed ID**

STNext のコマンドラインに戻り、この文献のレコードを検索する

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			
Get All CAS Registry Numbers 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			
GenBank Accession No.: CAA56943, Y17397, X81053, CAA76763			
Get Genbank Accession Nos. 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			
Organism: Homo sapiens			
Sequence:			
<pre> 1 MWSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYVYVSG KKYIGPCGGR 51 DCSVCHCVPE KSGRGPPGPP GPQGPIGPLG APGPIGLSGE KGMRGDRGPP 101 GAAGDKGDKG PTGVPGFPGL DGIPGHPPGP GPRGKPGMSG HNGSRGDPGF 151 PGGRGALGPG GPLGHPGEGK EKGNSVFILG AVKGIQGDGR DPGLPGLPGS 201 WGAGGPAGPT GYPGEPLVG PPGQPRPGL KGNPGVGVKG QMGDPGEVGQ 251 QGSPGPTLLV EPPDFCLYKG EKGIGIPGM VGLPGPPGRK GESGIGAKGE 301 KGIPGFPGR GDPGSYGSPP FPLKGELGL VDGPLFLGLI GPKGDPGNRG 351 HPGPPGVLT PPLPLKGPPG DPGFPGRYGE TGDVGPPGPP GLLGRPGEAC 401 AGMIGPPGPQ GFFGLPGLPG EAGIPGRPDS APGKPGKPGS PGLPGAPGLQ 451 GLPGSSVIYC SVGNPGPQGI KGKVGPPGGR GPKGEKGNEG LCACEPGPMG </pre>			
View More ▾			

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。

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

Get All Patent Numbers Show Search Details

Download Results ボタン

Sequence Alignment	Alignment Identity %	Query Identity %	Subject Identity %	Number of Documents
260	100.00	100.00	100.00	7

View More

■ 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。

Alignment Image		Alignment Text		C	D
				Sequence Length	CAS Registry Number®
1	Query 19 260 242 aa	Query: 19 PLADYVDEKPPYLGGLFLVLRNQDQFCTCLGRLMTVSLVLYS 48 P: D POLYGLFLVLRNQDQFCTCLGRLMTVSLVLYS Subject: 1449 P:GDGPGKPPYLGGLFLVLRNQDQFCTCLGRLMTVSLVLYS 1498		1690	
	Subject 1449 1690 1690 aa	Query: 49 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP 118 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP Subject: 1449 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP 1548			
10	Query 19 260 242 aa	Query: 119 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG 168 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG Subject: 1549 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG 1598			
	Subject 977 1218 1218 aa	Query: 169 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY 218 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY Subject: 1899 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY 1648			
11	Query 19 260 242 aa	Query: 19 PLADYVDEKPPYLGGLFLVLRNQDQFCTCLGRLMTVSLVLYS 48 P: D POLYGLFLVLRNQDQFCTCLGRLMTVSLVLYS Subject: 977 P:GDGPGKPPYLGGLFLVLRNQDQFCTCLGRLMTVSLVLYS 1026		1218	791151-70-7
	Subject 977 1218 1218 aa	Query: 49 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP 118 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP Subject: 1027 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP 1076			
	Subject 1449 1690 1690 aa	Query: 119 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG 168 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG Subject: 1077 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG 1126			
		Query: 169 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY 218 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY Subject: 1127 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY 1176			
		Query: 19 PLADYVDEKPPYLGGLFLVLRNQDQFCTCLGRLMTVSLVLYS 48 P: D POLYGLFLVLRNQDQFCTCLGRLMTVSLVLYS Subject: 1449 P:GDGPGKPPYLGGLFLVLRNQDQFCTCLGRLMTVSLVLYS 1498		1690	791139-90-5, 155348-50-2, 459495-63-7, 480647-33-4, 1865810-39-4
		Query: 49 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP 118 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP Subject: 1449 QEFAMQGLAGSCLFVFTTFLFAVCHNQVCHTAQRWSVLAASAP 1548			
		Query: 119 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG 168 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG Subject: 1549 LRMPLESEA18PVVBCAVCAFAQAVAVRQDQIIPPCQWELMIG 1598			
		Query: 169 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY 218 YFLMTGAGDQGGQALMPGSCLEDFRPAAPFLCQGRQTCFTFANY			

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	Pan paniscus	0				12012:941046	Nature (London, United Kingdom), (2012), 486(7404), 527-531
0			2	US6812339B1 US20070037165A1	7065 7065	0		
4	CAA56943 Y17357 X01053 CAA76763	Homo sapiens	10	EP3171889A1 US6812339B1 CA2555401A1 US20070037165A1 US20070033334A1 WO2016014761A1 JP201753266A US20110183924A1 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

```
DIQMTQSPSSLSASVGDRVTITCRASQGIRNYLAWYQQKPGKAPKLLIYAASTLQSGVPSRFSGSGSGTDFLTITSSL  
QPEDVATYYCQRYNRAPYTFGGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNAL  
QSGNSQESVTEQDSKSTYLSSTLTLSKADYEKHKVYACEVTHQGLSPVTKSFNRGEC
```

- 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

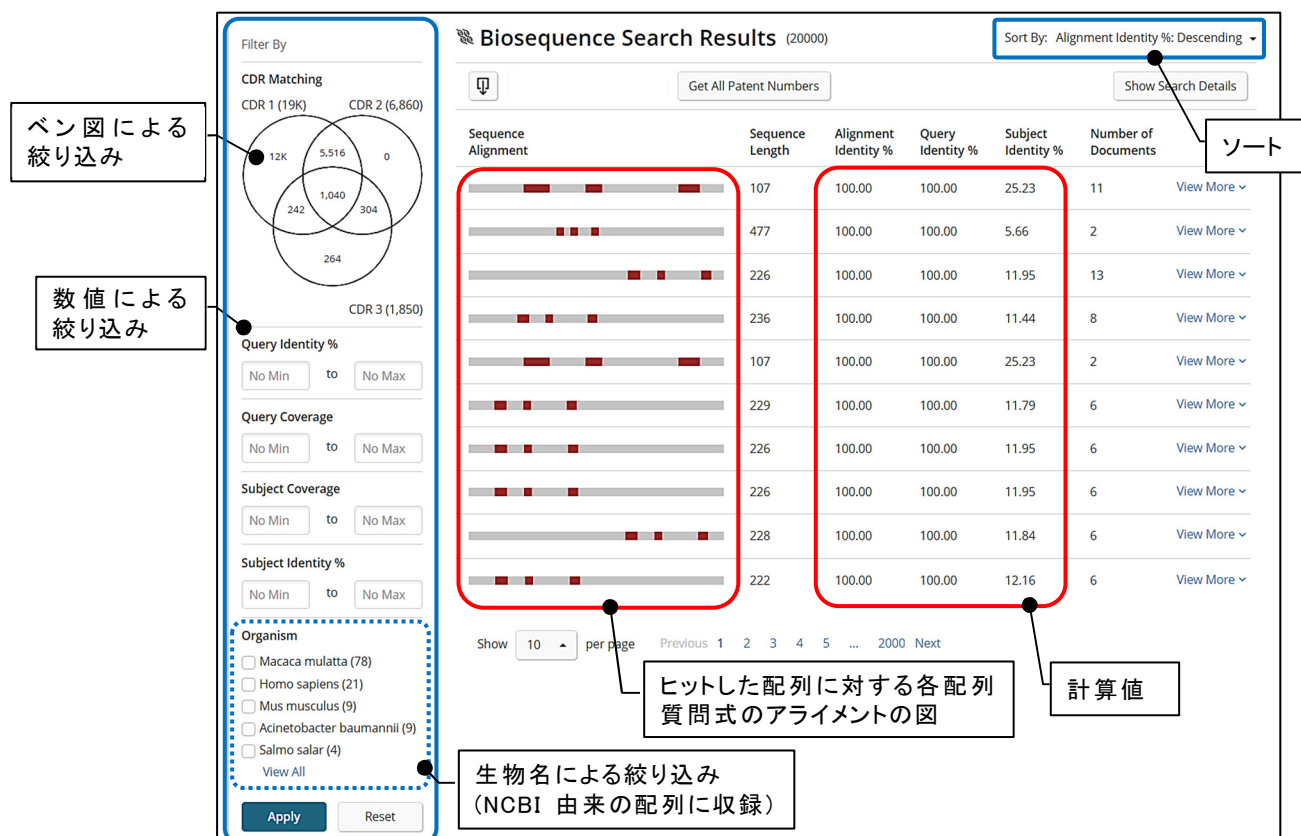
配列質問式や検索条件の確認

検索結果の確認

CDR 配列検索の結果

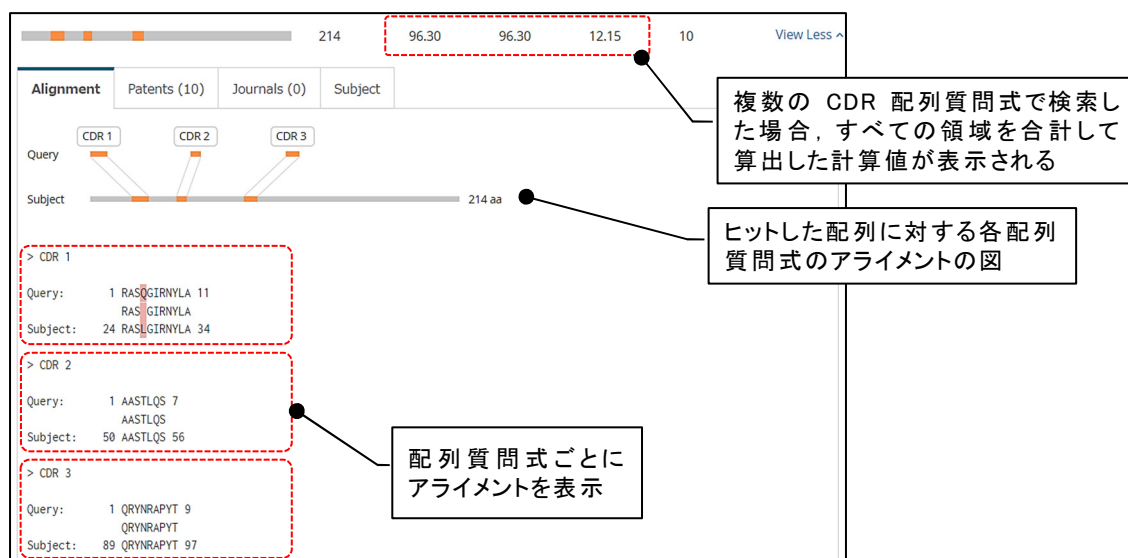
- CDR 配列検索の結果一覧が、ヒットした配列に対する配列質問式のアライメントの図と、各項目の計算値と共に表示されます。

- ・ 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。



- 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。



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][V]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>

Search Biosequences

BLAST CDR **Motif**

Name 2022_0004_Sequence

Sequence type
☐ Nucleotide ☒ Protein

☒ Combine Motif Results

Limit Total Sequence Results to
20000

Run Search Save Query Cancel

パラメータの変更 (任意)
Motif 配列検索のパラメータは Query coverage% と E-Value のみ変更できる

Advanced Biosequence Search

Query Coverage % 0 to 100 E-Value 10

- ・ 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。

2022_0004_Sequence 2022 Mar 2 11:46 AM

[L][ALKV]G[FL][V]D[AG]DG View

Cancel Search ...

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

2022_0004_Sequence 2022 Mar 2 11:46 AM

[L][ALKV]G[FL][V]D[AG]DG View

View Results ... 8320 results

Combine Motif Results のチェックをはずした時の表示

View Results ... 64 of 64 searches complete 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）。

The screenshot shows the 'Biosequence Search Results' page with 8320 results. The table includes columns for Sequence Alignment, Sequence Length, Alignment Identity %, Query Identity %, Subject Identity %, and Number of Documents. A red box highlights the first 10 rows of the table. Annotations point to various features:

- 絞り込みフィルター** (Filter By): Points to the left sidebar containing filters for Query Identity %, Query Coverage, Subject Coverage, Subject Identity %, and Organism.
- ソート** (Sort): Points to the 'Sort By: Alignment Identity %: Descending' dropdown menu.
- 生物名による絞り込み (NCBI 由来の配列に収録)** (Filter by organism (sequences from NCBI)): Points to the 'Organism' section in the sidebar.
- 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** (Alignment diagram for each query sequence against the hit sequence): Points to the 'Sequence Alignment' column.
- 計算値** (Calculation values): Points to the numerical values in the 'Alignment Identity %', 'Query Identity %', and 'Subject Identity %' columns.

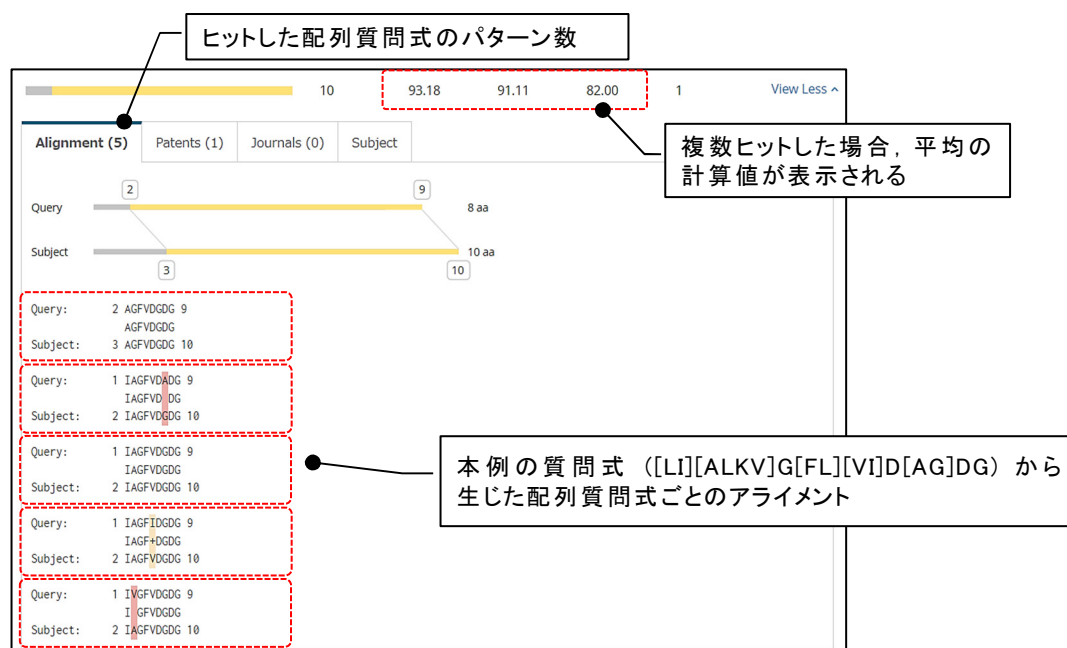
- Combine Motif Results のチェックをはずした場合、生じた配列質問式ごとに表示します。プルダウンで選択して、回答を確認します。

The screenshot shows the 'Biosequence Search Results' page with 4790 results. A red box highlights a list of generated query sequences. An annotation points to this list:

- 生じた配列質問式ごとの表示** (Display by generated query sequence): Points to the list of sequences.

■ 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。



CAS Sequences の仕様変更され、回答中の Patents/Journals タブがなくなりました。
得られた配列の参照文献情報はデータをダウンロードしてご確認ください。

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: '全回答の特許番号を検索する'. Below this, a table lists search results with columns: Sequence Alignment, Sequence Length, Alignment Identity %, Query Identity %, Subject Identity %, and Number of Documents. The 'Patents' tab is selected, showing '10 patents found.' and another 'Get All Patent Numbers' button highlighted with a red box and a callout: '表示している配列に関する特許番号を検索する'. Below the table, a specific patent entry is shown for 'Collagen iv replacement' with 'Assignees: GOLDFINCH BIO, INC.' and 'Patent No.: EP3171889A1'. A 'Get Patent Number' button is highlighted with a red box and a callout: '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, DGENE, INPAFAMDB, WPIX, EPFULL, JPFULL, ZCAPLUS, and FRFULL, KRFULL. The 'Save Script' checkbox is checked, and a callout points to it: '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'. The 'Continue' button is highlighted with a red box. Below the dialog, a large blue arrow points to the STNNext command line interface. The STNNext interface shows a list of patent numbers under the 'File WPINDEX' tab. A callout points to the list: '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 'Submit' button is highlighted with a red box.

■ 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

2 Journals found. Get All Accession Numbers Get All PubMed IDs

1 of 2

Complete primary structure with structure and exons.

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

The entire sequence of the human DNAs. The complete translation of 38-residue putative signal peptide residue NC1 domain. The gly-Xaa sequences of 1-12 residues in le collagen α chains. The calculated chain belongs to the α2(I) type fetal tissues revealed quite similar tissues. The expression patterns and α4(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 MWSLHIVLMR CSFRLTKSLA TGPWSLILIL FSVQYYVSG KKYIGPCGGR
51 DCSVCHCVPE KGSRGPPGPP GPQGPIGPLG APGPILSGE KGMRGDRGPP
101 GAAGDKGDKG PTGVPGFPLG DGIPGHPGP GPRGKPGMSG HNGSRGDPGF
151 PGGRGALGPG GPLGHPGCKG EKGNSVFILG AVKGIQDORG DPGLPLPGS
201 WGAGGPAGPT GYPGEPLVG PPGQGRPLG KGNPGVGVKG QMGDPGEVQ
251 QSSPGPTLLV EPPDFCLYKG EKGIGIPGM VGLPGPPGRK GESGIGAKGE
301 KGIPGFGPR GDGSGYSGPG FPLGKGLGL VGDPLGLI GPKGDPGNRG
351 HPGPPGLVT PPLPLKPPG DPFGPRGYE TGDVPPGPP GLLGRPEAC
401 AGMIGPPGPQ GPGLPLPLG EAGIPGRPS APGKPKPGS PGLPGAPLQ
451 GLPSSVIYC SVGNPGQGI KGVGPPGGR 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



ヘルプデスク



ご不明な点がございましたら、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