

## USGENE

USGENE (The USPTO Genetic Sequence Database) ファイルは、米国特許商標庁 (USPTO) が発行した公開特許・登録特許中の核酸およびアミノ酸の配列を収録するデータベースです。USGENE ファイルには、発明の名称 (標題), 抄録, 特許出願人や完全な発明者名, PCT の出願番号や日付などを含む豊富な書誌情報が収録されており、データは通常 USPTO から公報発行後 7 日以内に収録されます。

USGENE ファイルでは完全配列検索および部分配列検索用に GETSEQ, ホモロジー検索用に BLAST および GETSIM の RUN パッケージを提供しています。さらに、生物名, 分子タイプ, 配列長, 特徴表などの追加の生物学的データも利用できます。

## 収録内容

米国公開特許・米国登録特許由来の核酸およびアミノ酸の配列情報 (レコードは配列単位)

## 収録源

米国特許商標庁 (USPTO) が発行した特許

## ファイル内容

1982 年から現在まで  
6,300,000 件以上のレコード (2007 年 8 月現在)  
- 4,570,000 件以上の核酸配列 (2007 年 8 月現在)  
- 1,750,000 件以上のタンパク質配列 (2007 年 8 月現在)  
更新は毎週  
アラート (自動 SDI 検索) は毎週実施

## 検索補助資料

オンラインヘルプ (HELP DIRECTORY で利用できるすべてのヘルプメッセージのリストが表示できます)  
オンラインヘルプ PDF 版 ([http://www.stn-international.de/training\\_center/bioseq/usgene\\_help.pdf](http://www.stn-international.de/training_center/bioseq/usgene_help.pdf))  
STNGUIDE  
NCBI からの BLAST (R) (<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html>)

## ヨーロッパ

### STN カールスルーエ

FIZ Karlsruhe  
P.O. Box 2465  
76012 Karlsruhe  
Germany  
Phone: +49-7247-808-555  
Fax: +49-7247-808-259  
E-mail: [helpdesk@fiz-karlsruhe.de](mailto:helpdesk@fiz-karlsruhe.de)  
Internet: [www.stn-international.de](http://www.stn-international.de)

## 日本

### STN 東京

#### (社)化学情報協会

〒113-0021 東京都文京区本駒込6-25-4 中居ビル  
Phone: 0120-003-462 (Help Desk)  
: 0120-151-462 (上記以外)  
Fax: 03-5978-3600  
E-mail: [support@jaici.or.jp](mailto:support@jaici.or.jp)(Help Desk)  
[customer@jaici.or.jp](mailto:customer@jaici.or.jp) (上記以外)  
Internet: [www.jaici.or.jp](http://www.jaici.or.jp)

## 北アメリカ

### STN コロンバス

CAS  
P.O. Box 3012  
Columbus, Ohio 43210-0012 U.S.A  
CAS Customer Care:  
Phone: 800-753-4227 (North America)  
614-447-3700 (worldwide)  
Fax: 614-447-3751  
E-mail: [help@cas.org](mailto:help@cas.org)  
Internet: [www.cas.org](http://www.cas.org)

## SEARCH および DISPLAY フィールド

中間一致および後方一致検索可能なフィールドはアスタリスク (\*) で示してあります。

フィールド	SEARCH コード	SEARCH 例	DISPLAY コード
基本索引 <sup>3)*</sup> 標題 (/TI) 抄録 (/AB) 生物名 (/ORGN) 分子タイプ (/MTY) (以上からの切り出し語)	なし または /BI	S PLANT ?GENE? AND RNA S DNA DEPENDENT	TI, AB, ORGN, MTY
抄録 出願国 (WIPO コードとテキスト) 出願日 <sup>1)</sup> レコード番号 出願番号 <sup>2)</sup> 出願年 <sup>1)</sup> 資料種類 (コードとテキスト) 入力日 <sup>1)</sup> 代表クレーム 特徴表 <sup>3)*</sup> ファイルセグメント (コードとテキスト) 発明者 分子タイプ 生物名 特許出願人 <sup>4)</sup> 特許発行国 (WIPO コードとテキスト) 特許発行日 <sup>1)</sup> 特許種別 特許番号 <sup>2)</sup> 特許発行年 <sup>1)</sup> 関連出願国 (WIPO コードとテキスト) 関連出願日 <sup>1)</sup> 関連出願番号 <sup>2)</sup> 関連出願年 <sup>1)</sup> 配列の数 <sup>1)</sup> 配列同定番号 <sup>1)</sup> 配列長 <sup>1)</sup> 配列の収録源 標題	/AB /AC /AD /AN /AP /AY /DT または /TC /ED または /UP /ECLM または /MCLM /FEAT /FS /IN または /AU /MTY /ORGN /PA または /CS /PC /PD /PK /PN または /PATS /PY /RLC /RLD /RLN または /RLI /RLY /SEQC /SEQN /SQL /SSO /TI	S ANAPHYLATOXIN/AB S US/AC AND L1 S JUN 2000/AD S 4305872.2/AN S US1980-162082/AP S 2000/AY S PATENT/DT AND L1 S P/TC AND L1 S ED>=JAN 2007 AND L1 S NANOCRYSTAL/ECLM S ?COMBINAT?/FEAT S NUCLEIC/FS AND L1 S PROTEIN/FS AND L1 S MANDREKAR MICHELLE?/IN S RNA/MTY S CRICETULUS AUREUS/ORGN S AMGEN/PA S US/PC AND L1 S PD=4 MAR 2003 S USB2/PK AND L1 S US6686189/PN S 2002-2003/PY AND L1 S WO/RLC AND L1 S 20010313/RLD S WO2000-EP4786/RLN S 2000-2001/RLY S 3-4/SEQC S 337/SEQN S 150-175/SQL AND L1 S NCBI/SSO AND L1 S HYBRIDIZATION ASSAY##/TI	AB AI (AP) AI (AP) AN AI (AP) AI (AP) DT (TC) ED (UP) ECLM FEAT FS IN (AU) MTY ORGN PA (CS) PI (PN) PI (PN) PI (PN) PI (PN) PI (PN) RLI (RLN) RLI (RLN) RLI (RLN) SEQC SEQN SQL SSO TI

1) 数値演算子あるいは範囲指定による検索が可能な数値検索フィールドです。

2) STN 形式またはダウエント形式のどちらでも利用できます。

3) このフィールドでは前方一致、後方一致、および中間一致が利用できます。中間一致、後方一致の際は語幹として少なくとも 4 文字が必要です。

4) このフィールドでは (S) 演算子はスペースで代用できます。

## スーパー検索フィールド

必要な情報が含まれる一つまたは複数のフィールドを検索するときは、スーパー検索フィールドコードを利用します。スーパー検索フィールドを利用すると、クロスファイルおよびマルチファイル検索が簡単に実行できます。スーパー検索フィールドで EXPAND は利用できませんので、個々のフィールドコードで EXPAND してください。

フィールド	スーパー検索フィールドコード	検索されるフィールド	SEARCH 例	DISPLAY コード
出願番号グループ <sup>1)</sup>	/APPS	/AP, /RLN	S US2001-809524/APPS	AI, RLI

1) STN 形式またはダウエント形式のどちらでも利用できます。

## 配列検索項目

配列情報（アミノ酸および核酸配列）は、GETSEQ パッケージプログラムで利用できる多種の検索フィールドから得られます。GETSEQ で使う質問式には、QUERY コマンドで作成した L 番号を利用したり、配列検索可能な STN の他のファイル、例えば REGISTRY や DGENE ファイルで配列検索した L 番号を質問式として利用することも可能です。また、RUN GETSEQ の後に直接入力することもできます。

質問式	SEARCH 例
一般アミノ酸の 1 文字コード <sup>1)</sup> 一般アミノ酸の 3 文字コード <sup>1)</sup> コードまたはコードの列を一重引用符で囲む。 各コード間にハイフンを入力する。 核酸の1文字コード <sup>2)</sup>	QUE LAGLL/SQSP : 'MET-GLY-LEU-TRP-TRP-ARG'/SQSFP  : ACTACCTTCAAATACTAC/SQEN QUE ACTACCTTCAAATACTAC/SQSN

1) 矢印プロンプト (=>) の後で、HELP AAC と入力すると、一般アミノ酸の 1 および 3 文字コード表が表示されます。  
2) 矢印プロンプト (=>) の後で、HELP NUC と入力すると、核酸のコード表が表示されます。

## 配列検索の種類

核酸およびタンパク質配列データは、SEQ フィールド中に 1 文字コードで、またタンパク質のみ SEQ3 フィールド中に 3 文字コードで表示されます。

種類	定義	SEARCH コード	SEARCH 例
タンパク質, 完全配列	質問式に一致する配列を配列検索。質問式は完全に定義されていないなければならない。	/SQEP	QUE AKRSSKM/SQEP QUE 'ALA-LYS-ARG-SER-SER-LYS'/SQEP
タンパク質, 完全配列ファミリー	質問式に一致する配列および質問式中のアミノ酸とファミリーに相当するアミノ酸の配列 <sup>1)</sup>	/SQEFP	: SETLR/SQEFP : 'SER-GLU-THR-LEU-ARG'/SQEFP
タンパク質, 部分配列	完全に一致する配列と質問式の配列を含む配列を検索。特殊記号が利用できる。	/SQSP	: SKGYF/SQSP QUE 'SER-LYS-GLN-TYR-PHE'/SQSP
タンパク質, 部分配列ファミリー	完全に一致する配列と質問式のアミノ酸とファミリーに相当するアミノ酸の配列が含まれる。特殊記号が利用できる。 <sup>1)</sup>	/SQSFP	QUE SYVVE/SQSFP : 'SER-TYR-VAL-VAL-GLU'/SQSFP
核酸, 完全配列	質問式に完全に一致する配列を検索。非確定コードが利用できる。	/SQEN	QUE ACTACCTTCAAATACTAC/SQEN
核酸, 部分配列	完全に一致する配列と質問式の配列を含む配列を検索。非確定コードおよび特殊記号が利用できる。	/SQSN	: ACTACCTTCAAATACTAC/SQEN

(続く)

1) タンパク質の等価ファミリーは以下の通りです.

P, A, G, S, T	(弱疎水性, 中性)
Q, N, E, D, B, Z	(親水性, 酸アミン)
H, K, R	(親水性, 塩基性)
F, Y, W	(疎水性, 芳香族)
L, I, V, M	(疎水性)
C	(架橋)

部分配列検索の特殊記号 (/SQSP, /SQSFP, および /SQSN)

記号	機能	SEARCH 例
[ ]	代替残基の特定	QUE LGP[VL]/SQSP : LGP[VAL' LEU' LYS']/SQSP
[-]	特定の代替残基の除外	QUE LGP[-H]/SQSP QUE LGP[-HIS']/SQSP : LGP[-HL]/SQSP
{m}	直前の配列, 配列質問式, または L 番号を m 回繰り返す	: (FL) {2}/SQSP : L4 {2}/SQSP : (CTG) {2}/SQSN QUE TAA (TAAA) {2}/SQSN
{m, u} または {m-u}	直前の配列, 配列質問式, または L 番号を m から n 回繰り返す	QUE GG (FL) {1, 2}/SQSP : L3 {1, 3}/SQSP : (CTG) {1, 3}/SQSN
? または {0, 1} または {0-1}	直前の配列, 配列質問式, または L 番号をゼロまたは 1 回繰り返す	QUE FLRRI (RP)?K/SQSP : FLRRI (RP) {0, 1}K/SQSP : L1 {0, 1}NN/SQSP QUE CAT (CGA) {0, 1}GGAC/SQSN
* または {0, } または {0-}	直前の配列, 配列質問式, または L 番号をゼロ回以上繰り返す	QUE KLK (WD) {0, }N/SQSP : KLK (WD)*N/SQSP QUE L1 {0-}NN/SQSP : L1 {0-}NN/SQSP QUE CAT (CTG) {0, }TATT/SQSN
+ または {1, } または {1-}	直前の配列, 配列質問式, または L 番号を 1 回以上繰り返す	QUE KLK (DLE) {1, }/SQSP : KLK (DLE)+/SQSP : L2 {1-}/SQSP : L2 {1, }/SQSP QUE CAT (CTG) {1, }TATT/SQSN
&	配列表記または質問式 (L 番号) を結合する	QUE L1&L3/SQSFP : L2&L5 {1, 3}/SQSP

1) 部分配列質問式のさまざまな指定法に関する詳細は, USGENE ファイルで矢印プロンプト(=>)の後に HELP SQQ と入力すると表示できます.

上記の他に, キャレット (^), 垂直バー (|) も利用できます.

キャレット (^) は, 配列中の最初または最後の部分を指定して検索する場合に用います.

垂直バー (|) は, 代替記号です. 例えば, 代替配列質問式に利用します.

## 部分配列検索のギャップ記号 (/SQSP, /SQSFP, および /SQSN)

記号	機能	SEARCH 例
.	1 残基のギャップ	QUE SY. RPG/SQSP : SY..RPG/SQSP : AAG...TGC/SQSN
.{m} または {m.}	m 残基のギャップ	QUE SY. {2}RPG/SQSP : SY[2.]RPG/SQSP
.{m, u} または .{m-u}	m から u 残基のギャップ	: GFF. {2, 10}LSS/SQSP : GFF. {2-10}LSS/SQSP
:または .? または . {0, 1} または .{0-1}	ゼロまたは 1 残基のギャップ	QUE AAG. {2, 5}TGC/SQSN : AGA:SRI/SQSFP : AGA. ?SRI/SQSFP QUE AGA. {0, 1}SRI/SQSFP QUE AGA. {0-1}SRI/SQSFP
. * または . {0, } または .{0-}	ゼロ残基以上のギャップ	: HLC. *TYG/SQSP : HLC. {0, }TYG/SQSP : HLC. {0-}TYG/SQSP : AAGGCAGATG. *GCAA/SQSN
.+ または . {1, } または .{1-}	1 残基以上のギャップ	QUE SY. +TYG/SQSP : SY. {1, }TYG/SQSP : SY. {1-}TYG/SQSP : TCCTG. +GTGG/SQSN

## 配列ホモロジー検索 (BLAST/GETSIM)

BLAST (R) および GETSIM パッケージプログラムは、タンパク質および核酸配列データを相同性 (類似性) で検索することが可能です。BLAST は、National Library of Medicine (NLM) の National Center for Biotechnology Information (NCBI) より提供されています。GETSIM は、FIZ Karlsruhe GmbH より提供されており、FASTA アルゴリズムに基づいています。

BLAST および GETSIM で検索する際は、次の検索フィールドコードを使用します。

- /SQP — タンパク質の配列検索 (デフォルト)
- /SQN — 核酸の配列検索
- /TSQN — 核酸配列から翻訳されたタンパク質配列のデータベースを検索

BLAST および GETSIM 検索は、バッチ検索やアラートの登録が可能です。検索フィールドに /SQN または /TSQN を使用する場合は、単独 (SIN), 相補鎖 (COM), 双方 (BOTH) を指定することができます。これらのオプションは /SQN COM のように TSQN あるいは SQN の検索フィールドコードと共に使用します。BLAST では BOTH, GETSIM では SIN がデフォルトです。

核酸およびタンパク質配列はさまざまな方法で相同性検索ができます。質問式には、QUERY コマンドを使用して作成・保存した L 番号を使うことができます。また RUN BLAST/GETSIM コマンドと同じ行に直接入力したり、UPLOAD コマンドでアップロードしたテキストファイルを利用することも可能です。

検索を実行すると、質問式と得られた結果の配列の相同性を示すグラフが作成されます。Y 軸は相同性の数値を示し、X 軸は対応する回答数を表わします。すべての回答、もしくは任意の相同性以上の回答を選んで回答セットを作成することができます。最初に作成された回答セットでは、回答がレコード番号の降順に並んでいます。相同性の高い順に回答を並べ替える際は、SOR SCORE D と L 番号を入力してください。

ALIGN 表示形式を使用して、BLAST および GETSIM の検索結果を質問式の配列と対応づけて表示できます。上の列は質問式、下の列は検索でヒットした配列を示します。BLAST の ALIGN 表示形式は、NCBI の対応表示の標準規則に従います。GETSIM の ALIGN 表示形式では、同一のアミノ酸/核酸残基を 2 ドット、化学的に“等価”のマッチを 1 ドット、マッチがない場合はスペース (ドットなし) で示されます。質問式または回答配列に挿入されたギャップは、下線で示されます。

USGENE ファイルでは、SEQO 表示形式を利用するとオリジナルの配列情報を表示することができます。SEQO 表示形式では、特許出願人によって記述されたオリジナルの核酸配列に相当するタンパク質配列が表示されます。

## BLAST/GETSIMの検索タイプ

検索タイプ	SEARCH コード	SEARCH 例 <sup>4)</sup>
タンパク質ホモロジー 核酸ホモロジー 単独鎖 <sup>1)</sup> 相補鎖 双方 <sup>2)</sup> 翻訳タンパク質ホモロジー 単独鎖 <sup>1)</sup> 相補鎖 双方 <sup>2)</sup>	/SQP /SQN  /TSQN	RUN BLAST L1/SQP RUN BLAST L1/SQN RUN GETSIM L1/SQN SIN RUN GETSIM L1/SQN COM RUN BLAST L1/SQN BOTH RUN BLAST L1/TSQN RUN GETSIM L1/TSQN SIN RUN BLAST L1/TSQN COM RUN BLAST L1/TSQN BOTH
バッチ検索	/SQP BATCH /SQN BATCH /TSQN BATCH	RUN GETSIM L1/TSQN BOTH BATCH RUN BLAST L1/SQN COM BATCH
アラート検索 <sup>3)</sup>	/SQP ALERT /SQN ALERT /TSQN ALERT	RUN GETSIM L1/TSQN BOTH ALERT RUN BLAST L1 /SQN COM ALERT

- 1) GETSIM ではデフォルト  
2) BLAST ではデフォルト  
3) アラート検索はデータ更新毎、すなわち毎週実行されます。  
4) L1 は UPLOAD または QUERY を使用して生成された配列質問式です。

## BLAST の上級ユーザー向けオプション

BLAST の熟練ユーザーのために、多数のオプションが可能となっています。これらのパラメータの変更は検索回答出力に大きな効果を発揮します。FIZ Karlsruhe はこれらの機能パラメータのどれについても利用する前に、ユーザーが NCBI の文献を読み、理解されることを強く推奨します。詳細情報は次をご覧ください：

<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/information3.html>

上級ユーザーオプションはハイフン + 1 文字コード + スペース + 値の書式で入力します：

RUN BLAST L1/SQN -E 0.1.

オプション	スイッチ	値
フィルタ	-f	T (初期値), F, C T がセットされた場合、タンパク質に対しては SEG, 核酸に対しては DUST フィルタが有効となる。C 記号は “COILED COILED” フィルタ。
期待値	-e	浮動小数点数 (初期値10)
Word Size	-w	核酸: 7-23, 初期値は 11 タンパク質: 2-3, 初期値は 3
ストランド	-s	1 (SIN), 2 (COM), または 3 (BOTH) (初期値)
マトリックス	-m	BLOSUM62 (初期値), BLOSUM80, BLOSUM45, PAM30 または PAM70
ギャップ反則	-g	核酸: 5 (初期値) タンパク質: 11 (初期値)
ギャップ拡張	-x	核酸: 2 (初期値) タンパク質: 1 (初期値)
核酸のミスマッチに対する減点	-q	-3 (初期値)
核酸のマッチに対する加点	-r	1 (初期値)

## BLAST マトリックスの設定

特定のマトリックスに対して、可能性のあるギャップとギャップ拡張値の組合せは制限されています。各マトリックスに対して可能な設定値と初期値は以下の表のとおりです。これ以外の組合せはシステムによって拒絶され、警告メッセージが表示されます。

マトリックス	ギャップ	ギャップ拡張値
BLOSUM62	9	2
	8	2
	7	2
	12	1
	11	1 (デフォルト)
	10	1
BLOSUM80	8	2
	7	2
	6	2
	11	1
	10	1 (デフォルト)
	9	1
BLOSUM45	13	3
	11	3
	12	3
	9	3
	15	2 (デフォルト)
	14	2
	13	2
	12	2
	19	1
	18	1
	17	1
16	1	
PAM30	7	2
	6	2
	5	2
	10	1
	8	1
	9	1 (デフォルト)
PAM70	8	2
	7	2
	6	2
	11	1
	10	1 (デフォルト)
	9	1

## DISPLAY および PRINT 形式

回答のディスプレイとプリントには下記の表示形式を自由に組み合わせることができます。複数のコードを指定するときは、"D LI 1-5 TI PA KW" のようにスペースやカンマで区切ってください。フィールドは指定された順序で表示されます。

DED, ED, FEAT, UP を除き、すべての検索フィールドでハイライト機能が利用できます。HIT, KWIC および OCC の各形式を使うためには、検索時にハイライト機能は ON である必要があります。

形式	英語名	内容	入力例
AB	Abstract	抄録	D TI AB
AI (AP) <sup>1)</sup>	Application Information	出願情報	D AI
AIO <sup>2)</sup>	Application Information, Original	オリジナルの出願情報	D AIO
AN	Accession Number	レコード番号	D AN
CLM	Claims	クレーム	D CLM
DT (TC)	Document Type	資料種類	D TI DT
ECLM (MCLM)	Ezemplary Claim Date	代表クレーム	D BIB AB ECLM
ED (UP)	Entry Date	入力日	D ED
FEAT	Feature Table	特徴表	D FEAT
FS <sup>2)</sup>	File Segment	ファイルセグメント	D FS
IN (AU)	Inventor	発明者	D IN PA
MTY	Molecule Type	分子タイプ	D MTY
ORGN	Organism Name	生物名	D ORGN
PA (CS)	Patent Assignee	特許出願人	D IN PA
PI <sup>1)</sup>	Patent Information	特許情報	D PI
(PN, PATS)			
RLI (RLN) <sup>1)</sup>	Related Application Number	関連出願情報	D RLI
RLIO <sup>2)</sup>	Related Application Information, Original	オリジナルの関連出願情報	D RLIO
SCORE <sup>3)</sup>	Similarity Score	相同性スコア	D TI SCORE
SEQ <sup>4)</sup>	Sequence (1-letter codes)	配列 (1文字コード)	D SEQ
SEQ3 <sup>4)</sup>	Sequence (3-letter codes)	配列 (3文字コード)	D SEQ3 1-3
SEQC <sup>2)</sup>	Sequence Number Count	配列の数	D SEQC
SEQN <sup>2)</sup>	Sequence Identity Number	配列同定番号	D SEQN
SEQO <sup>2)</sup>	Original Sequence (alignment of nucleotide sequence and peptide sequence it express when given)	オリジナルの配列	D SEQO
SQL	Sequence Length	配列長	D SQL
SSO	Sequence Source	配列の収録源	D SSO
TI	Title	標題	D TI IN PA
ALIGN <sup>3)</sup>	質問式と相同性検索の結果の回答を対応させて表示 (RUN GETSIM または RUN BLAST)		D ALIGN 1, 10, 100
ALL <sup>1)</sup>	AN, MTY, TI, IN, PA, PI, AI, RLI, ED, DT, AB, CLM, SSO, ORGN, SQL, SEQ, FEAT		DIS ALL
APPS	AI, PRAI		D APPS
BIB (STD) <sup>1)</sup>	AN, MTY, TI, IN, PA, PI, AI, RLI, DT (デフォルト)		D BIB ALIGN
BRIEF <sup>1)</sup>	AN, MTY, TI, IN, PA, PI, AI, RLI, ED, DT, AB		D BRIEF
IALL <sup>1)</sup>	フィールド名付きインデント型 ALL 形式		D IALL 1-3 L4
IBIB <sup>1)</sup>	フィールド名付きインデント型 BIB 形式		D IBIB KWIC
IBRIEF <sup>1)</sup>	フィールド名付きインデント型 BRIEF 形式		D IBRIEF KWIC
SCAN <sup>5)</sup>	TI (回答番号なしのランダム表示)		D SCAN
SQIDE	AN, SQL, SEQ, FEAT		D SQIDE AB
SQ3IDE	AN, SQL, SEQ3, FEAT		D SQ3IDE
TRIAL (TRI, SAM)	AN, MTY, SQL		D TRI
HIT	ヒットタームを含むフィールド		D HIT
KWIC	ヒットタームの前後 50 語を表示 (Keyword-In-Context)		D KWIC NOH
OCC	ヒットタームの出現頻度をフィールドごとに表示		D OCC

1) デフォルトでは、特許番号、出願番号、および優先権出願番号は STN 形式で表示されます。ダウエント形式に変更するときには、矢印プロンプト (=>) の後で SET PATENT DERWENT と入力します。STN 形式に戻す場合は、SET PATENT STN と入力します。

2) カスタム形式のみの表示です。

- 3) RUN GETSIM あるいは RUN BLAST の回答である必要があります。詳しくは、配列ホモロジー検索の項を参照してください。
- 4) USGENE ファイルの配列は、WIPO の WST. 25 に従って表記されます。
- 5) この表示形式のオンライン・ディスプレイ料金は無料です。SCAN は、コマンドに続けて入力します。例: D SCAN または DISPLAY SCAN

### SELECT, ANALYZE および SORT フィールド

SELECT コマンドは、回答セットの指定したフィールドから抽出した語句に E 番号を付与します。  
 ANALYZE コマンドは、回答セットの指定したフィールドから抽出した語句に L 番号を付与します。  
 SORT コマンドは、検索結果を指定したフィールドのアルファベット順または数値順に並べ替えます。  
 (該当項目は Y, 該当しないものは N で表示されています)

フィールド	フィールドコード	ANALYZE/SELECT <sup>1)</sup>	SORT
抄録	AB	Y	Y
出願国	AC	Y	N
出願日	AD	Y	N
出願情報	AI (AP)	Y	Y
オリジナルの出願情報	AIO	Y	N
レコード番号	AN	Y	Y
出願年	AY	Y	N
クレーム	CLM	Y	Y
資料種類	DT	Y	Y
代表クレーム	ECLM (MCLM)	Y	Y
入力日	ED (UP)	Y	Y
特徴表	FEAT	Y	N
ファイルセグメント	FS	Y	Y
発明者 (著者名)	IN (AU)	Y	Y
分子タイプ	MTY	Y	Y
生物名	ORGN	Y	Y
特許出願人	PA	Y	Y
特許発行国	PC	Y	Y
特許情報	PI (PN)	Y	Y
特許種別	PK	Y	Y
特許発行年	PY	Y	Y
関連出願国	RLC	Y	N
関連特許出願日	RLD	Y	N
オリジナルの関連出願情報	RLIO	Y	N
関連出願番号	RLN (RLI)	Y	Y
関連出願年	RLY	Y	N
相同性スコア	SCORE	N	Y <sup>2)</sup>
配列 (1 文字コード)	SEQ	Y	N
配列 (3 文字コード)	SEQ3	Y	N
配列の数	SEQC	Y	N
配列同定番号	SEQN	Y	Y
配列長	SQL	Y	Y
配列の収録源	SSO	Y	Y
標題	TI	Y (デフォルト)	Y

1) ヒットタームだけを抽出させるには、HIT を使います。例: SEL HIT IN

2) RUN GETSIM あるいは RUN BLAST の回答である必要があります。

## サンプルレコード

## IAL形式での表示

L1 ANSWER 1 OF 1 USGENE COPYRIGHT 2007 SEQUENCEBASE CORP on STN  
 AN 6881821.13 DNA USGENE  
 TI Hepatitis-C virus type 4, 5, and 6 (Patent)  
 IN Simmonds Peter (Edinburgh, GB); Yap Peng Lee (Edinburgh, GB); Pike Ian Hugo (Bromley, GB)  
 PA Common Services Agency(Edinburgh GB); Murex Diagnostics International Inc (Bridgetown BB)  
 PI US 6881821 B2 20050419  
 US 20050032047 A1 20050210  
 WO 1994025602 A 19941110  
 AI US 1995-537802 19951221  
 RLI WO 1994-GB957 19940505  
 ED 20070328  
 DT Patent  
 AB Newly elucidated sequences of hepatitis C virus type 4 and type 5 are described, together with those of a newly discovered type 6. Unique type-specific sequences in the NS4, NS5 and core regions enable HCV detection and genotyping into types 1 to 6. Antigenic peptides and immunoassays are described.

## CLM

US6881821 B2: What is claimed is:1. An isolated peptide having an antigenic sequence selected from the following:a) QPAVIPDREVLVYQQFDEN (SEQ ID NO:32); and,b) ECSKHLPLVEHGLQLAEQF (SEQ ID NO:46).

2. A peptide according to claim 1 which is bound to a multiple antigen peptide core.

3. A peptide according to claim 2 having a sequence selected from the following:a) [H.sub.2 N-QPAVIPDREVLVYQQFDEN].sub.8 K.sub.4 K.sub.2 K-COOH (SEQ ID NO:32); and,b) [H.sub.2 N-ECSKHLPLVEHGLQLAEQF].sub.8 K.sub.4 K.sub.2 K-COOH (SEQ ID NO:46);where K.sub.4 K.sub.2 K is the multiple antigen peptide core.

4. A peptide according to claim 1 which is fused to another peptide to form a fusion peptide.

5. A peptide according to claim 4 fused to another peptide selected from the group consisting of  $\beta$ -galactosidase, glutathione-S-transferase, trpE and polyhedrin coding sequence.

6. A peptide according to claim 1, wherein said peptide is labelled.

7. An immunoassay device which comprises a solid substrate having immobilized thereon a peptide according to claim 1.

8. A device according to claim 7 wherein a mixture of antigenic peptides of HCV type 4, type 5 or type 6 is immobilized on the solid substrate.

## IAL形式での表示 (続き)

9. A device according to claim 7 wherein a mixture of antigenic peptides of HCV types 4, 5 and 6 is immobilized on the solid substrate.
10. A device according to claim 7 wherein the mixture further comprises one or more antigenic NS4 peptides of HCV types 1 to 3.
11. A device according to claim 10 wherein the mixture is a mixture of HCV type 1, 2, 3, 4, 5, and 6 antigenic peptides.
12. A device according to claim 7 for HCV typing which comprises a solid substrate containing HCV-4, HCV-5 and HCV-6 antigenic peptides.
13. A device according to claim 7 for HCV typing, further comprising a mixture of non-immobilized heterologous-type blocking HCV peptides, wherein said mixture excludes the peptide of the HCV type being detected.
14. An immunoassay kit which comprises an immunoassay device according to claim 7 for HCV typing, together with a series of solutions, each solution comprising a mixture of heterologous-type blocking HCV peptides, wherein each solution excludes the peptide of the HCV type being detected.
15. An immunoassay kit as claimed in claim 14, wherein the immunoassay device comprises a solid substrate having having immobilized thereon a mixture of antigenic peptides of HCV types 1, 2, 3, 4, 5, and 6; together with a series of six competing solutions, each solution containing a mixture of different antigenic peptides of HCV types 1, 2, 3, 4, 5, and 6.
16. A method of in vitro screening a sample for HCV antibodies which comprises:a) obtaining said sample;b) contacting said sample with a peptide of claim 1 and,c) detecting any antibody-antigen complex produced.
17. A method according to claim 16 wherein the peptide is immobilized on a solid substrate.
18. A method according to claim 17 wherein a mixture of peptides is immobilized on the solid substrate.
19. A method according to claim 18 wherein the mixture is a mixture of HCV type 1, 2, 3, 4, 5 and 6 antigenic peptides.
20. A method according to claim 17 wherein the sample and a mixture of heterologous-type blocking HCV peptides are applied to the peptide immobilized on the solid substrate.
21. A method according to claim 16, wherein HCV antibodies present in the sample are captured on a solid substrate, wherein said peptide is labeled, and wherein said peptide is applied to said HCV antibodies captured on said substrate for

## IALI形式での表示 (続き)

detection of any captured HCV antibodies.

SSO NUCLEIC; NCBI; GRANTED

ORGN Unknown

SQL 222

SEQ

```

1 gtctatcagt gttgtaacct ggagcccgaa gctcgcaagg ctattactgc
51 cctcacagaa agactctacg tggcgggccc catgcacaac agcaaggag
101 acctttgtgg gtatcggaga tgcgggcaa gcggagtctt tacgaccagc
151 ttcggaaca cgctgacgtg ctacctaaaa gccacggccg ctattagac
201 ggcggggctg agagactgca ct

```

## FEATURE TABLE:

Key |Location|

```

=====+=====+=====
source |1..222 |

```

## 検索例 (続き)

=> UPLOAD

IS THIS DATA A QUERY, OR FOR A RUN PACKAGE? Q/R/(END) :R

ENTER NAME OF RUN PACKAGE, END OR (?) :BLAST

START LOCAL KERMIT TRANSMIT PROCESS

UPLOAD SUCCESSFULLY COMPLETED

L1 GENERATED

=> D L1 LQUE

L1 ANSWER 1 USGENE COPYRIGHT 2007 SEQUENCEBASE CORP on STN

```

LQUE gtatatataa cgtgatgagc gtacgggtgc ggagacgcac cggagcgcctcggccagccgc
cgctccaagc cctgaggtt tccggggacc acaatgaacaagttgctgtg ctgcgcgctc
gtgtttctgg acatctccat taagtggaccaccaggaaa cgtttcctcc aaagtacctt
cattatgacg aagaaacctctcatcagctg ttgtgtgaca aatgtcctcc tggtaacctac
ctaaaacaacactgtacagc aaagtggaag accgtgtgcg cccottgccc
tgaccactactacacagaca gctggcacac cagtacgag tgtctatact
gcagccccgtgtgcaaggag ctgcagtacg tcaagcagga gtgcaatcgc
accacaaccgcgtgtgcga atgcaaggaa gggcgctacc ttgagataga
gtttctgttgaacatagga gctgccctcc tggatttga gttgtgcaag
ctggaacccagagcgaat acagtttgca aaagatgtcc agatgggttc
ttctcaaatgagacgtcacc taaagcacc ttagaaaac acacaaattg
cagtgtctttggtctcctgc taactcagaa aggaaatgca acacagaca
acatatgttcggaaacagt gaatcaactc aaaaatgtgg aatagatgtt
acctgtgtgaggaggcatt cttcaggttt gctgttccta caaagtttac
gcctaactggottagtgtct tggtagaaa tttgcctggc accaaaagtaa
acgcagagagtagtagagg ataaaacggc aacacagctc acaagaacag
actttccagctgctgaagtt atggaacat caaaacaag cccaagatat

```

```

agtcaagaagatcatccaag atattgacct ctgtgaaaac agcgtgcago
ggcacattggacatgctaac ctcaccttog agcagcttgc tagcttgatg
gaaagcttaccgggaaagaa agtggggagca gaagacattg aaaaaacaat
aaaggcatgcaaaccagtg accagatcct gaagctgctc agtttgtggo
gaataaaaaatggcgaccaa gacaccttga agggcctaata gcacgcacta
aagcactcaaagactacca ctttccaaa actgtcactc agagtctaaa
gaagaccatcaggttccttc acagcttcac aatgtacaaa ttgtatcaga
agttatTTTTtagaaatgata ggtaaccagg tccaatcagt aaaaataagc
tgcttataactggaaatggc cattgagctg tttcctcaca attggcgaga toccatggatgataa
    
```

=> RUN BLAST L1/SQN

BLAST Version 2.2

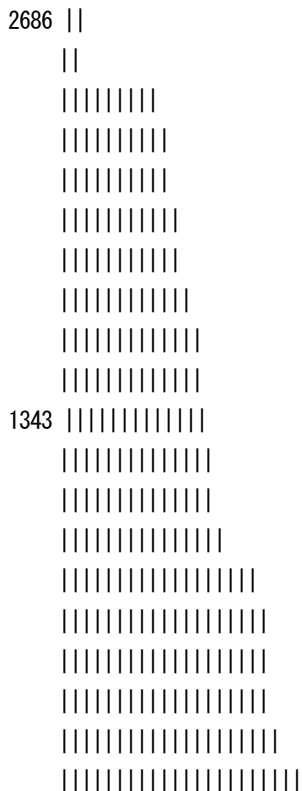
The BLAST software is used herein with permission of the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM). See also, Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res.* 25:3389-3402

. . . . .

480 ANSWERS FOUND BELOW EXPECTATION VALUE OF 10.0

Similarity

Score



Answer Count    100    200    300    400    500

HOW MANY ANSWERS WOULD YOU LIKE TO KEEP ? (ALL) OR ? :ALL

L2 RUN STATEMENT CREATED

```
L2      480 GTATATATAACGTGATGAGCGTACGGGTGCGGAGACGCACCGGAGCGCTC
      GCCCAGCCGCCGCTCCAAGCCCTGAGGTTTCCGGGGACCACAATGAACA
      AGTTGCTGTGCTGCGCGCTCGTGTCTGGACATCTCCATTAAGTGGACC
      ACCCAGGAAACGTTTCTCCAAAGTACCTTATTATGACGAAGAAACCTC
      TCATCAGCTGTTGTGTGACAAATGTCCTCCTGGTACCTACCTAAAACAAC
      ACTGTACAGCAAAGTGAAGACCGTGTGCGCCCTTGCCCTGACCACTAC
      TACACAGACAGCTGGCACACCAAGTACGAGTGTCTATACTGCAGCCCGT
      GTGCAAGGAGCTGCAGTACGTCAAGCAGGAGTGAATCGCACCCACAACC
      GCGTGTGCGAATGCAAGGAAGGGCGCTACCTTGAGATAGATTCTGCTTG
      AAACATAGGAGCTGCCCTCCTGGATTTGGAGTGGTGAAGCTGGAACCCC
      AGAGCGAAATACAGTTTGCAAAGATGTCCAGATGGGTTCTTCTCAAATG
      AGACGTCTCTAAAGCACCCCTGTAGAAAACACACAAATTGCAGTGTCTTT
      GGTCTCCTGCTAACTCAGAAAGGAAATGCAACACACGACAACATATGTTT
      CGGAAACAGTGAATCAACTCAAAAATGTGAATAGATGTTACCCTGTGTG
      AGGAGGCATTCTTCAGGTTTGTCTTCTACAAAGTTTACGCCTAACTGG
      CTTAGTGTCTTGGTAGACAATTTGCCTGGCACCAAAGTAAACGCAGAGAG
      TGTAGAGAGGATAAAACGGCAACACAGCTCACAAGAACAGACTTCCAGC
      TGCTGAAGTTATGGAACATCAAAACAAGCCCAAGATATAGTCAAGAAG
      ATCATCCAAGATATTGACCTCTGTGAAAACAGCGTGCAGCGGCACATTGG
      ACATGCTAACCTCACCTTCGAGCAGCTTCGTAGCTTGTGAAAAGCTTAC
      CGGGAAAGAAAGTGGGAGCAGAAGACATTGAAAAACAATAAAGGCATGC
      AAACCCAGTGACCAGATCCTGAAGCTGCTCAGTTTGTGGCGAATAAAAAA
      TGGCGACCAAGACACCTTGAAGGGCTAATGCACGCACTAAAGCACTCAA
      AGACGTACCACCTTCCAAAAGTGTCACTCAGAGTCTAAGAAGACCATC
      AGGTTCTTACAGCTTACAATGTACAAATTGTATCAGAAGTATTTTT
      AGAAATGATAGGTAACCAGGTCCAATCAGTAAAAATAAGCTGCTTATAAC
      TGGAAATGGCCATTGAGCTGTTTCTCACAATTGGCGAGATCCCATGGAT
      GATAA/SQN. -E 10.0
```

Answer set arranged by accession number; to sort by descending similarity score, enter at an arrow prompt (=>) "sor score d".

=> SOR SCORE D

PROCESSING COMPLETED FOR L2

L3 480 SOR L2 SCORE D

=> D TRIAL ALIGN 1 480

L3 ANSWER 1 OF 480 USGENE COPYRIGHT 2007 SEQUENCEBASE CORP on STN

TI Compositions and methods for the prevention and treatment of  
cardiovascular diseases (Patent)

MTY mRNA

SQL 1355

BLASTALIGN

Query = 1355 letters  
 Length = 1355  
 Score = 2686 bits (1355), Expect = 0.0  
 Identities = 1355/1355 (100%)  
 Strand = Plus / Plus

```

Query: 1      gtatatataacgtgatgagcgtacgggtgoggagacgcaccggagcgtcgcccagccgc
              |||
Sbjct: 1      gtatatataacgtgatgagcgtacgggtgoggagacgcaccggagcgtcgcccagccgc
Query: 61     cgctccaagcccctgaggtttccggggaccacaatgaacaagttgctgtgctgcgcgctc
              |||
Sbjct: 61     cgctccaagcccctgaggtttccggggaccacaatgaacaagttgctgtgctgcgcgctc
Query: 121    gtgtttctggacatctcattaagtggaccaccaggaaacgtttcctccaaagtacctt
              |||
Sbjct: 121    gtgtttctggacatctcattaagtggaccaccaggaaacgtttcctccaaagtacctt
.
.
.
.
.
Query: 1261   ggtaaccaggccaatcagtaaaaataagctgcttataactggaaatggccattgagctg
              |||
Sbjct: 1261   ggtaaccaggccaatcagtaaaaataagctgcttataactggaaatggccattgagctg
Query: 1321   tttcctcacaattggogagatcccatggatgataa 1355
              |||
Sbjct: 1321   tttcctcacaattggogagatcccatggatgataa 1355
    
```

L3 ANSWER 480 OF 480 USGENE COPYRIGHT 2007 SEQUENCEBASE CORP on STN  
 TI Osteoprotegerin (Patent)  
 MTY DNA  
 SQL 24  
 BLASTALIGN

Query = 1355 letters  
 Length = 24  
 Score = 40.1 bits (20), Expect = 2e-08  
 Identities = 20/20 (100%)  
 Strand = Plus / Plus

```

Query: 1108 caagacacottgaaggcct 1127
              |||
Sbjct: 1    caagacacottgaaggcct 20
    
```

=> S L3 AND ((TNF## OR TUMOR NECROSIS FACTOR?) (2A) (RECEPTOR? OR BINDING PROTEIN?))/BI, ECLM AND  
 AY>1990 AND GRANTED/SSO

L6 480 S L3  
 8124 TNF##/BI  
 5134 TNF##/ECLM  
 178934 TUMOR/BI

55956 TUMORS/BI  
 215867 TUMOR/BI  
     ((TUMOR OR TUMORS)/BI)  
 6270 NECROSIS/BI  
 254502 FACTOR?/BI  
 5965 TUMOR NECROSIS FACTOR?/BI  
     ((TUMOR (W) NECROSIS (W) FACTOR?)/BI)  
 43712 TUMOR/ECLM  
 2778 NECROSIS/ECLM  
 117817 FACTOR?/ECLM  
 2596 TUMOR NECROSIS FACTOR?/ECLM  
     ((TUMOR (W) NECROSIS (W) FACTOR?)/ECLM)  
 297291 RECEPTOR?/BI  
 142895 RECEPTOR?/ECLM  
 260220 BINDING/BI  
 3052258 PROTEIN?/BI  
 24305 BINDING PROTEIN?/BI  
     ((BINDING (W) PROTEIN?)/BI)  
 192272 BINDING/ECLM  
 872693 PROTEIN?/ECLM  
 16941 BINDING PROTEIN?/ECLM  
     ((BINDING (W) PROTEIN?)/ECLM)  
 4363 ((TNF## OR TUMOR NECROSIS FACTOR?) (2A) (RECEPTOR? OR BINDING PROT  
     EIN?))/BI, ECLM  
 6320174 AY>1990  
     (AY>1990)  
 2073916 GRANTED/SSO  
 L7       120 L6 AND ((TNF## OR TUMOR NECROSIS FACTOR?) (2A) (RECEPTOR? OR BINDI  
           NG PROTEIN?))/BI, ECLM AND AY>1990 AND GRANTED/SSO

=> SOR SCORE D

PROCESSING COMPLETED FOR L7

L8       120 SOR L7 SCORE D

=> D BIB AB ECLM ALIGN 1

L8   ANSWER 1 OF 120 USGENE COPYRIGHT 2007 SEQUENCEBASE CORP on STN  
 AN   6284740.5 cDNA       USGENE  
 TI   Osteoprotegerin (Patent)  
 IN   Boyle William J. (Moorpark, CA); Lacey David L. (Thousand Oaks, CA);  
       Calzone Frank J. (Westlake Village, CA); Chang Ming-Shi (Newbury Park,  
       CA)  
 PA   Amgen Inc (Thousand Oaks CA)  
 PI   US 6284740           B1   20010904  
 AI   US 1997-974186       19971118  
 DT   Patent  
 AB   The present invention discloses a novel secreted polypeptide, termed  
       Osteoprotegerin, which is a member of the tumor necrosis factor receptor

superfamily and is involved in the regulation of bone metabolism. Also disclosed are nucleic acids encoding Osteoprotegerin, polypeptides, recombinant vectors and host cells for expression, antibodies which bind Osteoprotegerin, and pharmaceutical compositions. The polypeptides are used to treat bone diseases characterized by increased resorption such as osteoporosis.

## ECLM

US6284740 B1: What is claimed is:1. A method of increasing levels of osteoprotegerin in a mammal comprising administering to the mammal a nucleic acid encoding osteoprotegerin, wherein the administration results in an increase in the level of osteoprotegerin and wherein the increase in the level of osteoprotegerin in the mammal results in increased bone density.

## BLASTALIGN

Query = 1355 letters  
 Length = 1355  
 Score = 2686 bits (1355), Expect = 0.0  
 Identities = 1355/1355 (100%)  
 Strand = Plus / Plus

```

Query: 1      gtatatataacgtgatgagcgtacgggtgoggagacgcaccggagcgctgcccagccgc
              |||
Sbjct: 1      gtatatataacgtgatgagcgtacgggtgoggagacgcaccggagcgctgcccagccgc
Query: 61     cgctccaagcccctgaggtttcggggaccacaatgaacaagttgctgtgctgcgcgctc
              |||
Sbjct: 61     cgctccaagcccctgaggtttcggggaccacaatgaacaagttgctgtgctgcgcgctc
              .
              .
              .
Query: 1261   ggtaaccagggtccaatcagtaaaaataagctgottataactggaaatggccattgagctg
              |||
Sbjct: 1261   ggtaaccagggtccaatcagtaaaaataagctgottataactggaaatggccattgagctg
Query: 1321   tttcctcacaattggcgagatcccatggatgataa 1355
              |||
Sbjct: 1321   tttcctcacaattggcgagatcccatggatgataa 1355
  
```