PATGENE (Patent Genetic Sequence Database)

Subject Coverage	Nucleotide and amino acid sequence data as submitted by patent applicants to the World Intellectual Property Organization (WIPO).
File Type	Bibliographic, sequence
Features	For direct code match or similarity (homology) sequence searching, FIZ Karlsruhe provides three specialized RUN package options, GETSEQ, GETSIM and BLAST®. Alerts (SDIs): Weekly or monthly (weekly is the default) CAS Registry Number® Identifiers ☐ SLART ☑ Keep & Share ☑ Structures ☐
Record Content	• Records contain sequence and patent information as given by the patent applicant. • Each record includes the actual sequence and additional information on the sequence, e.g., molecule type and organism, and patent information, e.g., publication, application and priority data and patent assignees.
File Size	• More than 54.1 million records (08/2025) • More than 38.7 million nucleic acid sequences (08/2025) • More than 15.4 million protein sequences (08/2025)
Coverage	1999 - present
Updates	Weekly
Language	English
Database Producer	FIZ Karlsruhe - Leibniz-Institute for Information Infrastructure Hermann-von-Helmholtz-Platz 1 76344 Eggenstein-Leopoldshafen Phone: +49 7247 808-0 Email: helpdesk@fiz-karlsruhe.de Copyright Holder
Sources	Sequence listings submitted by patent applicants as a formal part of WIPO/PCT applications.
User Aids	• Online Helps (HELP DIRECTORY lists all help messages available) • STNGUIDE
Clusters	• ALLBIB • BIOSCIENCE • CORPSOURCE • HPATENTS • MEDICINE • PATENTS • PHARMACOLOGY STN Database Cluster information

Search and Display Field Codes

Fields that allow left truncation are indicated by an asterisk (*).

General Search Fields

Search Field Name	Search Code	Search Examples	Display Codes
Basic Index* (contains single words from the title (TI), organism species (ORGN), molecule type (MTY) and feature table (FEAT) fields)	None or /BI	S ANAPHYLATOXIN S PLANT GENE# AND RNA	TI, ORGN, MTY, FEAT
Accession Number	/AN	S 2002060924.37/AN	AN
Amino Acid	/AA	S (T OR M)/AA	AA
Amino Acid Count **(1)**	/AA.CNT	S (T OR M OR F OR H)/AA (S) 50-100/AA.CNT	AA
Amino Acid Percentage **(1)**	/AA.PER	S (T OR M OR F OR H)/AA (S) 25-30/AA.PER	AA
Application Country	/AC	S US/AC	AI
Application Date **(1)**	/AD	S 20011129/AD	AI
Application Number **(2)**	/AP	S WO2020-CA51103/AP	AI
Application Number, Original	/APO	S WOAU2004/000622/APO	APO
Application Year **(1)**	/AY	S 2002/AY	AI
Cross Reference	/CR	S GENBANK/NM 002650/CR	CR
Data Entry Date **(1)**	/DED	S 20190228/DED	DED
Data Update Date **(1)**	/DUPD	S 20190228/DUPD	DUPD
Document Type (code and text)	/DT (or /TC)	S PATENT/DT	DT
Entry Date **(1)**	/ED	S 20210528/ED	ED
Field Availability	/FA	S AI/FA	FA
Feature Table*	/FEAT	S (RNA AND BINDING)/FEAT S ?COMBINAT?/FEAT	FEAT
File Segment (code and text)	/FS	S PROTEIN/FS S NS/FS	FS
Inventor	/IN	S MOORE/IN	IN
Molecule Type	/MTY	S RNA/MTY	MTY
Nucleic Acid**)**	/NA	S (G OR C)/NA	NA
Nucleic Acid Count **(1)**	/NA.CNT	S (G OR C)/NA (S) 50-100/NA.CNT	NA
Nucleic Acid Percentage **(1)**	/NA.PER	S (G OR C)/NA (S) 60-70/NA.PER	NA
Organism Name* **(3)**	/ORGN	S CRASSOSTREA GIGAS/ORGN	ORGN
Patent Assignee **(3)**	/PA (or /CS)	S MOLECULAR DYNAMICS/PA	PA
Patent Country (code and text)	/PC	S WO/PC	PI
Patent Number **(2)**	/PN	S WO 2002074961/PN	PI
Patent Number, Original	/PNO	S WO2020-000099/PNO	PNO
Patent Number Group **(2)**	/PATS	S WO 2002074961/PATS	PI
Patent Sequence Location	/PSL	S 6/PSL	PSL
Publication Date **(1)**	/PD	S 20030130/PD	PI
Publication Year **(1)**	/PY	S 2003/PY	PI
Priority Country	/PRC	S FR/PRD	PRAI
Priority Date **(1)**	/PRD	S 20150606/PRD	PRAI
Priority Date, First	/PRDF	S 20150608/PRDF	PRAI
Priority Number **(2)**	/PRN	S EP2001-102050/PRN	PRAI
Priority Number, Original	/PRNO	S DE10 200$ 051 727.5/PRNO	PRNO
Priority Year **(1)**	/PRY	S 2000-2001/PRY	PRAI
Priority Year, First	/PRYF	S 2015/PRYF	PRAI
Sequence Count	/SEQC	S 15/SEQC S A0000030BD19782FC1774AF58E4CFFEE7	SEQC
Sequence Key	/SEQK	F0E30588CBA14DCD38C/SEQK	SEQK
Sequence Identity Number **(1)**	/SEQN	S 337/SEQN	SEQN
Sequence Length **(1)**	/SQL	S 150-175/SQL	SQL
Title*	/TI	S HYBRIDIZATION ASSAY#/TI	TI
Update Date **(1)**	/UP	S 20210528/UP	UP

(1) Numeric search field that may be searched using numeric operators or ranges.
(2) Either STN or Derwent format may be used.
(3) Search with implied (S) proximity is available in this field.

Super Search Fields

Enter a super search code to execute a search in one or more fields that may contain the desired information. Super search fields facilitate cross-file and multi-file searching. EXPAND may not be used with super search fields. Use EXPAND with the individual field codes instead.

Search Field Name	Search Code	Fields Searched	Search Examples	Display Codes
Application Number Group	/APPS	/AP, /PRN	S US2001-809003/APPS	AI, PRAI

DISPLAY and PRINT Formats

Any combination of formats may be used to display or print answers. Multiple codes must be separated by spaces or commas, e.g., D L1 1-5 TI AU. The fields are displayed or printed in the order requested.

Hit-term highlighting is available for all fields. Highlighting must be ON during SEARCH to use the HIT, KWIC, and OCC formats.

Format	Content	Examples
AA	Amino Acid table	D AA
AI (AP) **(1)**	Application Information	D 21 2 AI
AN	Accession Number	D AN TI
APO (AIO)	Application Number, Original	D APO
CR	Cross Reference	D CR
DED	Data Entry Date	D DED
DUPD	Data Update Date	D DUPD
DT (TC)	Document Type	D DT
ED	Entry Date	D AN ED
FASTA	Sequence (FASTA format)	D FASTA
FEAT	Feature Table	D 1 5 10 FEAT
FS **(2)**	File Segment	D FS
IDENT **(2,3)**	Percent Identity	D IDENT
IN	Inventor	D IN
MTY	Molecule Type	DIS L5 1-10 MTY
ORGN	Organism Name	D ORGN
PA (CS)	Patent Assignee	D 1-25 PA
PI (PN) **(1)**	Patent Information	D 1-15 PA PI
PNO	Patent Number, Original	D PNO
PRAI	Priority Information	D PRAI
PRNO	Priority Number, Original	D PRNO
PSL	Patent Sequence Location	D PSL
SCORE **(2,4)**	Similarity Score	D TI SCORE
SEQ **(5)**	Sequence (one-letter codes)	D 1-3 TI SEQ
SEQ3 **(5)**	Sequence (three-letter codes)	D 1 5 10 TI SEQ3
SEQC	Sequence Count	D SEQC
SEQK	Sequence Key	D SEQK
SEQN	Sequence Identity Number	D SEQN
SQL	Sequence Length	D 1-20 SQL
TI	Title	D L7 1-25 TI
UP	Update Date	D AN TI UP

(1) By default, patent numbers, application and priority numbers are displayed in STN format. To display them in Derwent format, enter SET PATENT DERWENT at an arrow prompt. To reset display to STN format, enter SET PATENT STN.
(2) Custom display only.
(3) Use RUN BLAST first. See page 7, Similarity Search.
(4) Use RUN GETSIM or RUN BLAST first. See page 7, Similarity Search.
(5) Sequences in PATGENE are given according to ST.25 of the WIPO for records published before 2023, and mostly according to ST.26 since 2023.

Predefined Display and Print Formats

Format	Content	Examples
ALIGN **(1)**	Alignment as text between query and retrieved sequence in a similarity search (RUN GETSIM, RUN BLAST, or RUN GETSEQ)	D ALIGN
ALIGNG **(1)**	Alignment as image between query and retrieved sequence in a similarity search (RUN GETSIM, RUN BLAST, or RUN GETSEQ)	D ALIGNG
ALL	AN, ED, UP, DED, DUPD, TI, IN, PA, DT, PI, AI, PRAI, FS, CR, MTY, PSL, ORGN, SEQC, SEQN, SQL, SEQK, SEQ, AA or NA, FEAT	D ALL
APPS	AI, PRAI	D APPS
BIB	AN, ED, UP, DED, DUPD, TI, IN, PA, DT, PI, AI, PRAI, FS, CR, MTY, PSL (BIB is the default)	D BIB
IBIB	BIB, indented with text labels	D IBIB ALIGN
FASTA	FASTA format	D FASTA
IALL	ALL, indented with text labels	D L2 1-5 IALL
SCAN	ED, UP, DED, DUPD, TI (random display without answer numbers)	D SCAN
SQIDE	ED, UP, DED, DUPD, MTY, ORGN, SEQC, SEQN, SQL, SEQK, SEQ, AA or NA, FEAT	D SQIDE
SQ3IDE	ED, UP, DED, DUPD, MTY, ORGN, SEQC, SEQN, SQL, SEQK, SEQ3, AA or NA, FEAT	D SQ3IDE
TRIAL (TRI, SAM, SAMPLE, FREE)	TI, MTY, SQL	D 1-20 TRI
HIT	Hit term(s) and field(s)	D HIT
KWIC	Up to 50 words before and after hit term(s) (KeyWord-In-Context)	D KWIC
OCC	Number of occurrences of hit term(s) and field(s) in which they occur	

(1) Use RUN GETSIM, RUN BLAST or RUN GETSEQ first.

SELECT, ANALYZE, and SORT Fields

The SELECT command is used to create E-numbers containing terms taken from the specified field in an answer set.

The ANALYZE command is used to create an L-number containing terms taken from the specified field in an answer set.

The SORT command is used to rearrange the search results in either alphabetic or numeric order of the specified field(s).

Field Name	Field Code	ANALYZE/ SELECT (1)	SORT
Accession Number	AN	N	Y
Amino Acid	AA	Y	N
Amino Acid, Count	AA.CNT	Y	N
Amino Acid, Percentage	AA.PER	Y	N
Application Country	AC	Y	Y
Application Date	AD	Y	Y
Application Number	AP (AI)	Y	Y
Application Number, Original	APO (AIO)	Y	Y
Application Number and Related Application Number	APPS	Y	N
Application Year	AY	Y	Y
Cross Reference	CR	Y	Y
Data Entry Date	DED	Y	Y
Data Update Date	DUPD	Y	Y

SELECT, ANALYZE, and SORT Fields (cont'd)

Field Name	Field Code	ANALYZE/ SELECT (1)	SORT
Document Type	DT (TC)	Y	Y
Entry Date	ED	Y	Y
Feature Table	FEAT	Y	N
File Segment	FS	Y	Y
Inventor	IN	Y	Y
Molecule Type	MTY	Y	Y
Nucleic Acid	NA	Y	N
Nucleic Acid, Count	NA.CNT	Y	N
Nucleic Acid, Percentage	NA.PER	Y	N
Organism Name	ORGN	Y	Y
Patent Assignee	PA	Y	Y
Patent Country	PC	Y	Y
Patent Number	PN (PI)	Y	Y
Patent Number Group	PATS	Y	Y
Percent Identity	IDENT	N	Y
Priority Country	PRC	Y	Y
Priority Date	PRD	Y	Y
Priority Date, First	PRDF	Y **(2)**	Y
Priority Number	PRN	Y	Y
Priority Number, Original	PNRO	Y	Y
Priority Year	PRY	Y	Y
Priority Year, First	PRYF	Y **(2)**	Y
Patent Sequence Location	PSL	Y	Y
Publication Date	PD	Y	Y
Publication Year	PY	Y	Y
Sequence Count	SEQC	Y	Y
Sequence Identity Number	SEQN	Y	Y
Sequence Key	SEQK	Y	Y
Sequence Length	SQL	Y	Y
Similarity Score	SCORE **(3)**	N	Y
Title	TI	Y (default)	Y
Update Date	UP	Y	Y

(1) HIT may be used to restrict terms extracted to terms that match the search expression used to create the answer set, e.g., SEL HIT PA.
(2) SELECT HIT and ANALYZE HIT are not valid with this field.
(3) Used with a L-number created with BLAST and GETSIM.

Sequence Similarity Searching (BLAST/GETSIM)

The GETSIM and BLAST® run packages are available to search the PATGENE database for protein and nucleotide sequence data by similarity (homology). BLAST is provided in PATGENE with the permission of the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM). GETSIM is using the FASTA algorithm.

Nucleotide and protein sequences can be subjected to a similarity search as a query entered directly on the command line using RUN GETSIM/BLAST or they may be uploaded via the “Structures” page. See details here. The uploaded sequence can be displayed with D LQUE.

To initiate a BLAST or GETSIM search with the command RUN BLAST or RUN GETSIM the following search codes have to be specified:

- /SQP for searching peptide sequences
- /SQN for nucleotide sequences
- /TSQN for searching peptide sequences translated from PATGENE nucleotide sequences.

For the BLAST package four additional search codes are available:

- /SQM (megaBLAST) for searching highly similar nucleotide sequences
- /SQDM (discontiguous megaBLAST) for searching similar nucleotide sequences allowing more mismatches
- /TSQP for searching nucleotide sequences translated from PATGENE protein sequences
- /TSQNX for searching translated nucleotides form PATGENE protein sequences

It is recommended to use the search codes /SQM or /SQDM rather than /SQN when searching longer sequences as the response time is much faster. The commands /TSQN, /TSQP and /TSQNX are more time consuming compared to the other commands.

When using the /SQN, /SQM, /SQDM, or /TSQNX option, it is possible to specify whether single (SIN), complementary (COM), or BOTH strands should be searched. The options can be specified with the search code, e.g., /SQN -S COM. If no search option is given, BOTH (both) will be used by BLAST and GETSIM. Note that for the /TSQN option generally both strands will be searched.

GETSIM / BLAST: Types of Searches

Description	Search Code	Search Examples (1)
Peptide homology	/SQP	RUN BLAST L1 /SQP RUN GETSIM L1/SQP
Nucleotide homology	/SQN	RUN BLAST L1 /SQN RUN GETSIM L1/SQN
	/SQM **(2)**	RUN BLAST L1 /SQM
	/SQDM **(2)**	RUN BLAST L1 /SQDM
Translated peptide homology	/TSQN	RUN BLAST L1 /TSQN RUN GETSIM L1 /TSQN
Translated peptide homology from translated peptide	/TSQNX **(2)**	RUN BLAST L1/TSQNX
Translated nucleotide homology	/TSQP **(2)**	RUN BLAST L1 /TSQP

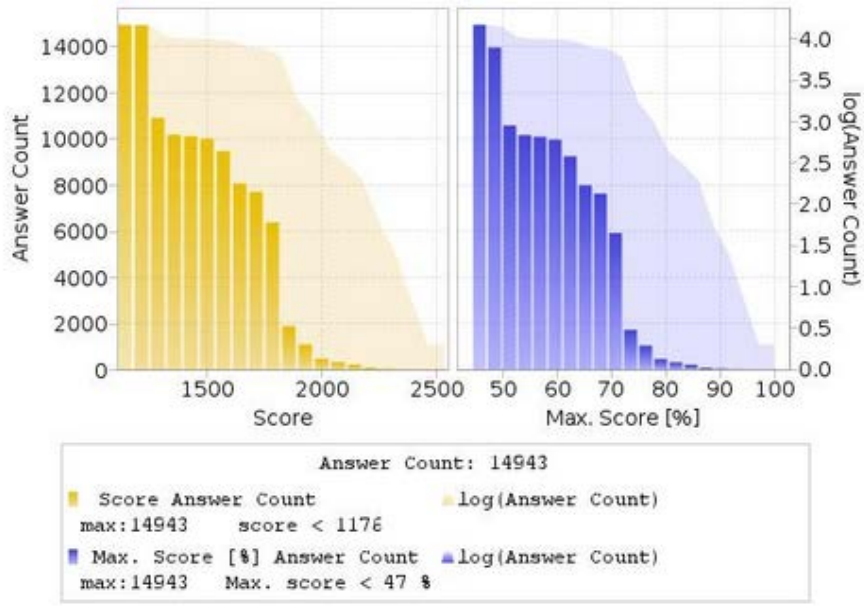
(1) Where L1 is a sequence query generated using the “Structure” page. **(2)** BLAST only

The maximum number of hits is by default 15,000 records. The parameter "-maxseq" allows to increase the maximum number of hits to 100,000 records, e.g.,=>RUN BLAST L1/SQN -F F -MAXSEQ 100000.
The number of additional results and their relevance in terms of high score and/or high identity values depend on the length of the query sequence and the number of subject sequences in the database.
In general, searching a short sequence with -maxseq 100000 may retrieve additional documents with high score and high identity values while searching a longer sequence with -maxseq 100000 may retrieve only additional documents with high identity values.

After a search with BLAST or GETSIM the number of retrieved sequences for the different score values are displayed in two diagrams. The y-axis of these diagrams represents the number of answers (absolute values are displayed as bars, logarithmic values are shaded) and the x-axis the score as the specific degree of similarity for this search. In the left diagram the score values are displayed, in the right diagram the percentage values of the maximum score.

In addition, two score values are given, the highest possible score value defining the maximum score when the query is aligned to itself, and the score of the best answer of the retrieved answer set. Both values are the same, if the query and at least one retrieved sequence are identical.

```
Highest possible score value: 2533.2
Best answer score value: 2533.2
```



Multiple answer sets (L-numbers) can be created with different cut off values for the score and the percentage identity. Five options are available:

1) Select a part of the answer set using the score value from the left histogram. The generated L-number contains all records with a score above the entered value.

```
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 ? :2300

L16   RUN STATEMENT CREATED
L16         10 ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGT
```

2) Select a part of the answer set using the percentage score value from the right histogram, e.g., "85%" or "85% SCORE". The generated L-number contains all records with a percentage score above the entered value.

```
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 ? :85% SCORE

L17   RUN STATEMENT CREATED
L17        143 ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGT
```

3) Select a part of the answer set using the percentage identity value, e.g., "100% IDENT". The generated L-number contains all records with a percentage identity above the entered value.

```
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 ? :100% IDENT

L18   RUN STATEMENT CREATED
L18         51 ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGT
```

4) Select a part of the answer set combining the percentage score and the percentage identity value, e.g., "85% SCORE 100% IDENT". The generated L-number contains all records which have a percentage score and percentage identity above the entered value.

```
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 ? :85% SCORE 100% IDENT

L19   RUN STATEMENT CREATED
L19          2 ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGT
```

5) Keep the complete answer set with ALL.

```
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

L20   RUN STATEMENT CREATED
L20      14943 ATGGGATGGAGCTGTATCATCCTCTTCTTGGTAGCAACAGCTACAGGTGT
```

The percentage score or identity cut off value can be set to two digits after the period. Small differences in the cut off value can have a tremendous effect on the number of results like in this example:

```
ENTER (ALL) OR ?:99.45%
L28              1292 MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP…
ENTER (ALL) OR ?:99.46%
L29                79 MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP…
```

In order to complete the RUN BLAST or the RUN GETSIM command, END must be entered.

```
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 ? :END
```

An L-number is generated for each selection, which contains all answers of the specified subset. Each L-number can be used for further processing. As the initial L-number is sorted by descending accession number, the selected L-number may be re-arranged by descending similarity score (SORT SCORE D L1) or descending percent identity (SORT IDENT D L1).

The alignment between the retrieved sequence and the query sequence can be displayed as text with the display format ALIGN or as an image with ALIGNG. The top line is the query sequence and the bottom line the hit sequence. Above each alignment the percentage of the BLAST and GETSIM score compared to the query self-score value and the percentage of identity is given. Both values can also be displayed as well with D SCORE and D IDENT. Both BLAST and GETSIM ALIGN format follows the standard convention for NCBI alignment displays. See further details in HELP ALIGNMENT.

```
ALIGNG
```

Query Length: 303; **Sequence Length:** 591;
Score: 277.2 bits (306), 50.6% of highest possible score 547.7;
Expect value: 1.877e-71;
Identities: 158 / 160 (98.8%);
Query Identity: 52.1%; **Query Coverage:** 52.8%;
Subject Identity: 26.7%; **Subject Coverage:** 27.1%;
Strand: Plus / Plus; **Alignment Length:** 160;

```
Q: 144 TCTGGGCTTCTTGCATTCTGGGACAGCCAAGTCTGTGACTTGCACGTACTCCCCTGCCCT 203
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S: 1   TCTGGGCTTCTTGCATTCTGGGACAGCCAAGTCTGTGACTTGCACGTACTCCCCTGCCCT 60
Q: 204 CAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGATTCCAC 263
       |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
S: 61  CAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGCGCAGCTGTGGGTTGATTCCA- 119
Q: 264 ACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTAC                     303
       |||||||||||||||||||||||||||||||||||||||
S: 120 ACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTAC                     159
```

Advanced User Options for BLAST and GETSIM

For the experienced user of BLAST® and GETSIM a variety of options are available via the STN command line. Altering these parameters will have a profound effect on the outcome of the search. FIZ Karlsruhe strongly recommends that users are completely familiar with NCBI documentation before embarking on customizing any of these settings. For further information see the information on the NCBI website.

The advanced user options are specified with a single letter code preceded by a hyphen and followed by a blank and the required value, e.g., RUN BLAST L1/SQN -F F or RUN BLAST L1/SQP -E 0.1 -M PAM30.

Advanced User Options

Option	Switch	Values
1. Filter	-f	T (True), F (False), Default value is T. If T is set, for peptides the SEG, and for nucleotides the DUST filter is employed.
2. Expectation Value	-e	Floating point number. (Default is 10)
3. Word Size	-w	11 (default) or 7-23 for nucleotides 3 (default) or 2 for peptides
4. Strand for nucleotides only	-s	1 (SIN), 2 (COM) or 3 (BOTH) default value is 3
5. Matrix for peptides only	-m	BLAST: BLOSUM62 (default), BLOSUM80, BLOSUM45, PAM30, PAM70 GETSIM: BL50 (default), BL62, BL80, MD10, MD20, MD40, OPT5, P120, P250, VT160
6. Gap Penalty	-g	Peptides (default): BLAST 11; GETSIM 12 Nucleotides (default): BLAST 5; GETSIM 12
7. Gap Extension	-x	Peptides (default): BLAST 1; GETSIM 2 Nucleotides (default): BLAST 2; GETSIM 4
8. Penalty for nucleotide mismatch	-q	BLAST: -3 (default); GETIM: -2 (default)
9. Reward for nucleotide match	-r	BLAST: 1 (default); GETSIM: 3 (default)

BLAST Matrix settings (for option 5. Matrix)

Please note that for a certain matrix only a restricted set of possible gap and gap extension values are possible. The settings available to each matrix are summarised in the table below. Default settings are indicated in the table. Any different combinations will be rejected by the system and a warning message issued.

Matrix	Gap	Gap Extension
BLOSUM62	9	2
	8	2
	7	2
	12	1
	11	1 (default)
	10	1
BLOSUM80	8	2
	7	2
	6	2
	11	1
	10	1 (default)
	9	1
BLOSUM45	13	3
	11	3
	12	3
	9	3
	15	2 (default)
	14	2
	13	2
	12	2
	19	1
	18	1
	17	1
	16	1
BLOSUM50	32767	32767
	13	3
	12	3
	11	3
	10	3
	9	3
	16	2
	15	2
	14	2
	13	2 (default)
	12	2
	19	1
	18	1
	17	1
	16	1
	15	1

Matrix	Gap	Gap Extension
BLOSUM90	32767	32767
	9	2
	8	2
	7	2
	6	2
	11	1
	10	1 (default)
PAM30	9	1
	7	2
	6	2
	5	2
	10	1
	8	1
	9	1 (default)
PAM70	8	2
	7	2
	6	2
	11	1
	10	1 (default)
	9	1
PAM250	32767	32767
	15	3
	14	3
	13	3
	12	3
	11	3
	17	2
	16	2
	15	2
	14	2 (default)
	13	2
	21	1
	20	1
	19	1
	18	1
	17	1

Searching Sequence Data with the GETSEQ RUN Package

The GETSEQ run package is a tool to search the PATGENE database for a direct sequence code match of peptide and nucleic acid sequences. This method is ideal for short and/or highly conserved sequence queries where similarity (homology) searching is not required. The maximum number of hits is 250,000 records.

Nucleotide and protein sequences can be subjected to a GETSEQ search as a query entered directly on the command line using RUN GETSEQ or the query may be created with the QUERY command, and subsequently searched through the GETSEQ run package specifying the query L-number (e.g., RUN GETSEQ L1, if L1 represents the sequence query).

```
=> RUN GETSEQ MCLHFLVLVICIL/SQSP

RUN GETSEQ AT 08:57:25 ON 2021-10-11
COPYRIGHT (C) 2021 FIZ KARLSRUHE on STN

GetSeq motif search by FIZ Karlsruhe; Version: 1.0.0

Query time:          115
L13   RUN STATEMENT CREATED
L13          8 MCLHFLVLVICIL/SQSP
```

Long sequences may be uploaded via the “Structures” page; see details here. The L-number may also derive from a previous sequence search in another STN database with bio sequence search capabilities, e.g., the CAS REGISTRY[SM] file.

Any L-numbered sequence answer set from RUN GETSEQ may be combined with any search field in the PATGENE file, for example => S L1 AND ARTIFICIAL SEQUENCE/ORGN where L1 represents the answer set from a RUN GETSEQ operation.

Hits of the retrieved sequence can be displayed in context of the whole sequences as text with the display format ALIGN or as an image with ALIGNG.

```
=> RUN GETSEQ CAGGGCGCGGACAAGCCGCGCCGTCGCCACTCGACCGCCGGCGCCCACAT/SQSN
L3        2594 CAGGGCGCGGACAAGCCGCGCCGTCGCCACTCGACCGCCGGCGCCCACAT/SQSN

=> D ALIGN
L3     ANSWER 1 OF 2594 GENESEQ COPYRIGHT 2024 CLARIVATE on STN.
ALIGN
  Sequence Length: 10446;
  Strand: Plus / Minus;
  Hits at: 6632-6583
      6726 CGGTCGCTGC GCTCCCTACG CCCCGCCGCT TCGCGTCGGC CTATCGCGGC CGCTGGCCGC
      6666 TCAAAAATGG CTGGCCTACG GCCAGGCAAT CTACCAGGGC GCGGACAAGC CGCGCCGTCG
                                                ====== ========== ==========
      6606 CCACTCGACC GCCGGCGCCC ACATCAAGGC ACCCTGCCTC GCGCGTTTCG GTGATGACGG
           ========== ========== ====
      6546 TGAAAACCTC TGACACATGC AGCTCCCGGA GACGGTCACA GCTTGTCTGT AAGCGGATGC
      6486 CGGGAGCAGA CAAGCCCGTC AGGGCGCGTC AGCGGGTGTT GGCGGGTGTC GGGGCGCAGC
```

The HIT display format contains only the part of the hit sequence with the matching residues which are highlighted with double underlining. In addition, the information “Hits at” displays the residue numbers of the start and end point of the matching part of the hit sequence.

```
=> D ALIGN
L4     ANSWER 52 OF 942 GENESEQ COPYRIGHT 2024 CLARIVATE on STN.
ALIGN
  Sequence Length: 5362;

  Hits at: 412-420 3432-3440
      301 LARTLRNCGV QPDTLVAILA DRSLEMIVSI IAVWKAGGAY VPLDPEYPKE RLQYLLHDAD
      361 ADVLLVQHHL KNSLAFDGPV IDLNDEASYH ADCSLLSPVA GHSHLAYVIY TSGTTGKPKG
                                                           =========
      421 VMVEHGGIVN SLQWKKAFFK HSPADRVLVL YPYVFDAFIL NFFGPLISGA TLHLLPNEEN
      481 KETFAIQNAI KQERITHFST SPRLLKTMIE QMNREDFIHV QHVVVGGEQL ETDTVEKLHS

     3361 LPIDPDSPSE RIRYILNDSS ISVLLYCGKL QDDIGFSGTC IDLMEEHFYH EKDSSLALSY
     3421 QSSQLAYAIY TSGTTGKPKG TLIEHRQVIH LIEGLSRQVY SAYDAELNIA MLAPYYFDAS
                      =========
     3481 VQQMYASLLS GHTLFIVPKE IVSDGAALCR YYRQHSIDIT DGTPAHLKLL IAAGDLQGVT
     3541 LQHLLIGGEA LSKTIVNKLK QLFGEHGAAP GITNVYGPTE TCVDASLFNI ECSSDAWARS
```

Sequence Search Terms

Amino acid and nucleic acid sequences may be searched with the one-letter code, amino acids also with the three-letter codes for common amino acids. Enter HELP AAC for a table of the one- and three-letter codes of the common amino acids and HELP NUC for a table of the codes for nucleic acids.

Uncommon amino acids are represented in the sequence by an 'X' (or 'Xaa'). ‘X’ is used also as an unspecified amino acid since July 2022 with standard ST.26. If you want to search specifically for an 'X' in the sequence, it has to be placed in square brackets, e.g., =>RUN GETSEQ TF[X]C[X]T/SQSP

Terms	Search Examples
One-letter codes for common amino acids	LAGLL/SQSP
Three-letter codes for common amino acids Enclose strings of codes in single quotes and use dashes to separate codes in strings.	'HIS-LEU-TYR-LEU-GLN-TYR-ILE-ARG-LYS-LEU'/SQSFP 'HIS-LEU-TYR-LEU-GLN-TYR-ILE-ARG-LYS-LEU' /SQEP
One-letter codes for nucleic acids	ATGAAN/SQEN CATCTGTATT/SQSN

Types of Sequence Searches

In the GETSEQ run package four options are available for searching polypeptide sequences using amino acid codes and two options for searching nucleic acid sequences.

Sequence data for nucleic acid and protein sequences are displayed in the SEQ field with one-letter codes and the SEQ3 field with three-letter codes for proteins only.

Type	Definition	Search Code	Query Examples
Sequence Exact Protein	Search for sequences that match the query.	/SQEP	GAPGEK/SQEP 'ASP-HIS-ALA-ILE-HIS' /SQEP
Sequence Exact Family, Protein	Search for sequences that match the query and those in which family-equivalent substitution of the query amino acids occur.	/SQEFP	YGGFL/SQEFP 'TYR-GLY-GLY-PHE-LEU'/SQEFP
Subsequence, Protein	Search for exact answers plus sequences in which the query sequence is embedded.	/SQSP	LAGLL/SQSP 'ASP-HIS-ALA'/SQSP
Subsequence Family, Protein	Search for exact sequences, subsequences, and answers in which family-equivalent substitution of the query amino acids occurs.	/SQSFP	ATCXAWV/SQSFP 'THR-ASP-SER-GLU-SER-SER-HIS' /SQSFP
Sequence Exact, Nucleic Acid	Search for sequences that match the query. Ambiguity codes for nucleic acids are allowed.	/SQEN	ATGAAN/SQEN
Subsequence, Nucleic Acid	Search for exact answers, plus sequences in which the query sequence is embedded. Ambiguity codes for nucleic acids are allowed.	/SQSN	TGGAGAAGGC/SQSN

The families of amino acid equivalents retrieved in the polypeptide family searches SQEFP and SQSFP are:

P, A, G, S, T	(neutral, weakly hydrophobic)
Q, N, E, D, B, Z	(hydrophilic, acid amine)
H, K, R	(hydrophilic, basic)
F, Y, W	(hydrophobic, aromatic)
L, I, V, M	(hydrophobic)
C	(cross-link forming)

Variability Symbols for Sequence Code Match Searches

Variability symbols are allowed in all GETSEQ search options. For more information on specifying variability in sequence code match queries, enter HELP SQQ.

Symbol(s)	Function	Query Examples
[ ]	to specify alternate residues	NGSLLAGAYAIST[LV]I/SQSP LGP['VAL-LEU-LYS']/SQSP
[-]	to exclude a specific residue or alternate residues	LGP[-H]/SQSP LGP[-'HIS']/SQSFP LGP[-HL]/SQSP
{m}	to repeat the preceding sequence m times	(FL){2}/SQSP (CTGA){3}/SQSN TAA(TAAA){2}/SQSN
{m, u} or {m-u}	to repeat the preceding sequence m to u times	GG(FL){1,2}/SQSP (CTGA){2,4}/SQSN
? or {0,1} or {0-1}	to repeat the preceding sequence zero or one time	FLRRI(RP)?K/SQSP FLRRI(RP){0,1}K/SQSP CATG(CGTA){0,1}GGAC/SQSN
* or {0,} or {0-}	to repeat the preceding sequence zero or more times	KLK(WD){0,}N/SQSP KLK(WD)*N/SQSP CATAA(CTG){0,}TATT/SQSN
+ or {1,} or {1-}	to repeat the preceding sequence one or more times	KLK(DLE){1,}/SQSP KLK(DLE)+/SQSP CATA(CTG){1,}TATT/SQSN
^ (Caret)	search at the beginning or end of a sequence	^MCGIL/SQS VCDS^/SQSP
|	specifies alternate residues	ACDS|KLMP/SQSP
&	to join together sequence expressions or queries (L#s)	

SPECIFYING GAPS IN GETSEQ SEQUENCE QUERIES

A gap may be specified in a sequence expression using the period (.) for one residue, the colon (:) for zero or one residue or the period (.) followed by an appropriate repeat expression. The following table summarizes all the options for specifying gaps in GETSEQ sequence searches**.**

Symbol(s)	Function	Query Examples
.	a gap of one residue	SY.RPG/SQSP SY..RPG/SQSP AAG...TGC/SQSN
{m} or [m.]	a gap of m residues	SY.{2}RPG/SQSP SY[2.]RPG/SQSP
{m,u} or .{m-u}	a gap of m to u residues	GFF.{2,10}LSS/SQSP GFF.{2-10}LSS/SQSP AAG.{2,5}TGC/SQSN
: or .? or .{0,1} or .{0-1}	a gap of zero or one residues	AGA:SRI/SQSFP AGA.?SRI/SQSFP AGA.{0,1}SRI/SQSFP AGA.{0-1}SRI/SQSFP
* or {0,} or .{0-}	a gap of zero or more residues	HLC.*TYG/SQSP HLC.{0,}TYG/SQSP HLC.{0-}TYG/SQSP AAGGCAGATG.*GCAA/SQSN
+ or {1,} or {1-}	a gap of one or more residues	SY.+TH/SQSP SY.{1,}TH/SQSP SY.{1-}TH/SQSP TCCTG.+GTGG/SQSN

Display ALIGNC for combined alignments

The search results originated from a BLAST- or FASTA (GETSIM)-search and a GETSEQ-search may be combined in one L-number. With the command **D ALIGNC** a combined alignment display is provided in which supports the quick evaluation of short sequence stretches within a longer sequence.

1) Run the BLAST-search

```
=> RUN BLAST L1/SQP -F F
Algorithm: BLAST - BLASTP. Version: 2.12.0+
...
14454 ANSWERS FOUND BELOW EXPECTATION VALUE OF: 10.0

ENTER (ALL) OR ?:80%

L2     RUN STATEMENT CREATED
L2           68 IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSG
                KTLTIQVKEFGDAGQYTCHKGGEVLSHSLLLLHKKEDGIWSTDILKDQKE
                ...
                GGNNIGTKSVHWYQQKPGQAPVLVVYADSDRPSGIPERVSGSNSGNTATL
                TISRVEAGDEADYYCQVWDSRSDHLWVFGGGTKLTVLG/SQP -F F
```

2) Run the GETSEQ-search

```
=>RUN GETSEQ QVWDSRSDHLWV/SQSP
L4           49 QVWDSRSDHLWV/SQSP
```

3) Combine the two results with AND

```
=>S L3 AND L4
L5            7 L3 AND L4
```

4) Use D ALIGNC for the display

Displayed are the usual values from the BLAST-search and within the BLAST-alignment the GETSEQ alignment with the double underlining and the L-number of the GETSEQ-search.

```
=> D ALIGNC
L7     ANSWER 1 OF 7 GENESEQ COPYRIGHT 2024 CLARIVATE on STN.
ALIGN
  ALIGNMENT FROM L-NUMBER L6
  Query Length: 788; Sequence Length: 788;
  Score: 1614.0 bits (4178), 100.0% of highest possible score 1614;
  Expect value: 0.0e0;
  Identities: 788 / 788 (100.0%); Positives: 788 / 788 (100.0%);
  Query Identity: 100.0%; Query Coverage: 100.0%;
  Subject Identity: 100.0%; Subject Coverage: 100.0%;
  Alignment Length: 788;
  Q:    1 IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  S:    1 IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEF 60

  Q:   61 GDAGQYTCHKGGEVLSHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  S:   61 GDAGQYTCHKGGEVLSHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTC 120

  …

  Q: 661 GGGGSGGGGSGGGGSVHSSYVLTQPPSVSVAPGQTATITCGGNNIGTKSVHWYQQKPGQA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  S: 661 GGGGSGGGGSGGGGSVHSSYVLTQPPSVSVAPGQTATITCGGNNIGTKSVHWYQQKPGQA 720

  Q: 721 PVLVVYADSDRPSGIPERVSGSNSGNTATLTISRVEAGDEADYYCQVWDSRSDHLWVFGG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  L3                                                  ============
  S: 721 PVLVVYADSDRPSGIPERVSGSNSGNTATLTISRVEAGDEADYYCQVWDSRSDHLWVFGG 780

  Q: 781 GTKLTVLG                                                    788
          ||||||||
  S: 781 GTKLTVLG                                                    788
```

ALIGNC-display for the evaluation of combined search for CDR- and antibody sequences

The ALIGNC display may be used for the evaluation of the alignments after a search for Complementarity-determining regions (CDR) with GETSEQ and a search for antibody sequences using the BLAST- or FASTA- (GETSIM) algorithm which have been combined in one L-number.

1) Run the GETSEQ-searches for each CDR

```
=> RUN GETSEQ SYIMM/SQSP
L1         1134 SYIMM/SQSP

=> RUN GETSEQ IKLGTVTTV[DEN]Y/SQSP
L2         1218 IKLGTVTTV[DEN]Y/SQSP

=> RUN GETSEQ SIYPSGGITFYAD..../SQSP
L3         1202 SIYPSGGITFYAD..../SQSP
```

and concatenate the results with AND

```
=> S L1 AND L2 AND L3
L4          891 L1 AND L2 AND L3
```

2) Run the BLAST-search with the maximum of results using the -MAXSEQ parameter

```
=>RUN BLAST L5/SQP -F F MAXSEQ 100000
Algorithm: BLAST - BLASTP. Version: 2.12.0+
…
97257 ANSWERS FOUND BELOW EXPECTATION VALUE OF: 10.0
```

3) Use high Score and Ident values

```
OR ENTER "END". "END" MUST BE ENTERED TO COMPLETE THE RUN COMMAND.
ENTER (ALL) OR ?:80% SCORE 90% IDENT
L6       47093 EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYIMMWVRQAPGKGLEWVSS
               IYPSGGITFYADTVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARIK
               …
               VYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPV
               LDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
               /SQP -F F -MAXSEQ 100000
```

4) Combining the GETSEQ- and BLAST-search with AND results in one L-number

```
=>S L4 AND l6
L7          553 L4 AND L6

=> D ALIGNC
L7     ANSWER 1 OF 553 GENESEQ COPYRIGHT 2024 CLARIVATE on STN.
ALIGN
  ALIGNMENT FROM L-NUMBER L6
  Query Length: 450; Sequence Length: 450;
  Score: 922.2 bits (2382), 99.6% of highest possible score 925.6;
  Expect value: 1.38e-266;
  Identities: 447 / 450 (99.3%); Positives: 450 / 450 (100.0%);
  Query Identity: 99.3%; Query Coverage: 100.0%;
  Subject Identity: 99.3%; Subject Coverage: 100.0%;
  Alignment Length: 450;
```

In the combined displayed after D ALIGNC each retrieved CDR is represented by the respective Lnumber and the short alignment from the GETSEQ-search within the alignment from the BLASTsearch.

```
Q:    1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYIMMWVRQAPGKGLEWVSSIYPSGGITFY 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L1                                  =====
L3                                                      ===========
S:    1 EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYIMMWVRQAPGKGLEWVSSIYPSGGITFY 60

Q:   61 ADTVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARIKLGTVTTVDYWGQGTLVTVSS 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L2                                          ===========
L3      ======
S:   61 ADTVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARIKLGTVTTVDYWGQGTLVTVSS 120

Q:  121 ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S:  121 ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS 180

Q:  181 GLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGG 240
        ||||||||||||||||||||||||||||||||||+||||||||||||||||||||||||||
S:  181 GLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGG 240

Q:  241 PSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYN 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S:  241 PSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYN 300

Q:  301 STYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREE 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||+|
S:  301 STYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDE 360

Q:  361 MTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW 420
        +||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
S:  361 LTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW 420

Q:  421 QQGNVFSCSVMHEALHNHYTQKSLSLSPGK                                450
        ||||||||||||||||||||||||||||||
S:  421 QQGNVFSCSVMHEALHNHYTQKSLSLSPGK                                450
```

Sample Records

DISPLAY TRIAL

```
L5      ANSWER 1 OF 22155 PATGENE COPYRIGHT 2021 FIZ KARLSRUHE on STN.
TI      A novel method for determining nucleic acids by use of a labeled
        nucleotide. [File created by using OCR software]
MTY     DNA
SQL     118
```

DISPLAY SQIDE

```
L5      ANSWER 2300 OF 22155 PATGENE COPYRIGHT 2021 FIZ KARLSRUHE on STN.
AN      2021046655.55   PATGENE  ED 20211004 UP 20211004
        DED 20210318 DUPD 20210318
MTY     DNA
ORGN    human papillomavirus
SEQC    1064
SEQN    55
SQL     120
SEQK    58ce30650d8deb81e47fab322725b621a3d7bf479800d6aa38e3f4376d0881d9

SEQ
               1 actgatttgc atttattgca caaatgaatt aacaacagca gaagtgctgt
              51 cctttgcatg gaaggagctg tgtattaagt gggaccacga actgccctac
             101 ggagcgtgtg cacagtgtct
NA
        Code    Count  Percent
           A       33     27.5
           C       24     20.0
           G       31     25.8
           U        0      0.0
           T       32     26.7
           B        0      0.0
           D        0      0.0
           H        0      0.0
           I        0      0.0
           K        0      0.0
           M        0      0.0
           R        0      0.0
           S        0      0.0
           V        0      0.0
           W        0      0.0
           X        0      0.0
      Others        0      0.0
FEATURE TABLE:
Key       |Location|
==========+========+========================================
          |        |gi/5059324.E6/lcl/HPV83_E6.1/ Human papi
          |        |llomavirus 83 (HPV83), E6 gene_1
          |        |
```

DISPLAY IALL

```
L35    ANSWER 1000 OF 17907 PATGENE COPYRIGHT 2021 FIZ KARLSRUHE on STN.
AN     2018098446.512   PATGENE  ED 20211004 UP 20211004
       DED 20180531 DUPD 20180531 Full-text
TI     Methods for Modulating RNA Splicing
PA     PTC Therapeutics, Inc.
DT     Patent
PI     WO 2018098446              20180531
AI     WO 2017-US63323           20171127
PRAI   US 2016-62426619          20161128 (62)
FS     NUCLEIC; NS
MTY    RNA
PSL    SEQ ID NO 512
ORGN   Artificial Sequence
SEQC   3866
SEQN   512
SQL    10
SEQK   f7468d232250a6ab16c0d9235c8bda8420d8491ad475be611b6f231e516f54a6

SEQ
               1 cagagurugu
NA
       Code    Count  Percent
          A        2     20.0
          C        1     10.0
          G        3     30.0
          U        3     30.0
          T        0      0.0
          B        0      0.0
          D        0      0.0
          H        0      0.0
          I        0      0.0
          K        0      0.0
          M        0      0.0
          R        1     10.0
          S        0      0.0
          V        0      0.0
          W        0      0.0
          X        0      0.0
     Others        0      0.0


FEATURE TABLE:
Key       |Location|
==========+========+========================================
          |        |intronic recognition element for splicin
          |        |g modifier (REMS)
          |        |
```

DISPLAY FASTA

```
L1      ANSWER 1 OF 3714 PATGENE COPYRIGHT 2021 FIZ KARLSRUHE on STN.

FASTA
        >PATGENE|2021189053.1866|protein|sequence 1866 from WO2021189053
        spkqekmkmdchkdekgtiydyeaialnkneyvsfkqyvgkhilfvnvatycgltaqypelnalqeelkpyg
        lvvlgfpcnqfgkqepgdnkeilpglkyvrpgggfvpsfqlfekgdvngekeqkvfsflkhscphpseilgt
        fksiswdpvkvhdirwnfekflvgpdgipvmrwshratvssvktdilaylkqfktk
```

DISPLAY SEQ3

```
L1      ANSWER 1 OF 3714 PATGENE COPYRIGHT 2021 FIZ KARLSRUHE on STN.

SEQ3
                  1 Ser-Pro-Lys-Gln-Glu-Lys-Met-Lys-Met-Asp-
                 11 Cys-His-Lys-Asp-Glu-Lys-Gly-Thr-Ile-Tyr-
                 21 Asp-Tyr-Glu-Ala-Ile-Ala-Leu-Asn-Lys-Asn-
                 31 Glu-Tyr-Val-Ser-Phe-Lys-Gln-Tyr-Val-Gly-
                 41 Lys-His-Ile-Leu-Phe-Val-Asn-Val-Ala-Thr-
                 51 Tyr-Cys-Gly-Leu-Thr-Ala-Gln-Tyr-Pro-Glu-
                 61 Leu-Asn-Ala-Leu-Gln-Glu-Glu-Leu-Lys-Pro-
                 71 Tyr-Gly-Leu-Val-Val-Leu-Gly-Phe-Pro-Cys-
                 81 Asn-Gln-Phe-Gly-Lys-Gln-Glu-Pro-Gly-Asp-
                 91 Asn-Lys-Glu-Ile-Leu-Pro-Gly-Leu-Lys-Tyr-
                101 Val-Arg-Pro-Gly-Gly-Gly-Phe-Val-Pro-Ser-
                111 Phe-Gln-Leu-Phe-Glu-Lys-Gly-Asp-Val-Asn-
                121 Gly-Glu-Lys-Glu-Gln-Lys-Val-Phe-Ser-Phe-
                131 Leu-Lys-His-Ser-Cys-Pro-His-Pro-Ser-Glu-
                141 Ile-Leu-Gly-Thr-Phe-Lys-Ser-Ile-Ser-Trp-
                151 Asp-Pro-Val-Lys-Val-His-Asp-Ile-Arg-Trp-
                161 Asn-Phe-Glu-Lys-Phe-Leu-Val-Gly-Pro-Asp-
                171 Gly-Ile-Pro-Val-Met-Arg-Trp-Ser-His-Arg-
                181 Ala-Thr-Val-Ser-Ser-Val-Lys-Thr-Asp-Ile-
                191 Leu-Ala-Tyr-Leu-Lys-Gln-Phe-Lys-Thr-Lys
```