

WIPO National Workshop on Search and Examination  
of Inventions related to Genetic Resources  
Manila, Philippines, May 20 to 24, 2019

# Hands-on Exercises

## <STN>

Yoo Lim Kam, Ph.D.  
Patent Examiner  
Korean Intellectual Property Office

# Exercises

Examples are from **“Sequence Motif Searches in CAS REGISTRY<sup>SM</sup> ”**  
([https://www.stninternational.org/uploads/tx\\_ptgsarelatedfiles/20130730\\_Sequence\\_Motif\\_Searches\\_in\\_REGISTRY.pdf](https://www.stninternational.org/uploads/tx_ptgsarelatedfiles/20130730_Sequence_Motif_Searches_in_REGISTRY.pdf))

# Exercise 1

Claim 1: A peptide having the sequence

$X_1 - X_2 - \text{PRO} - \text{THR} - \text{ALA} - X_3 - (X_4)_n$

wherein

$X_1$  is an amino acid other than proline,

$X_2$  is Lys, His, Trp, Asp, Leu, Pro, or Tyr,

$X_3$  is an amino acid other than proline,

$X_4$  is any amino acid, and n repeats 0-6 times.

# Exercise 1

Select the sequence symbols to use in the query

- Use single letter codes to represent the residues
- Use square brackets to represent possible residues at a specific location
- Use square brackets and minus sign to exclude residues
- Use a period (.) to represent any amino acid
- Use curly braces { } to represent repeats

# Exercise 1

## Construct the sequence query

Recall the sequence claim:

$X_1 - X_2 - \text{PRO} - \text{THR} - \text{ALA} - X_3 - (X_4)_n$   
wherein

$X_1$  = an amino acid other than proline

$X_2$  = Lys, His, Trp, Asp, Leu, Pro, or Tyr

$X_3$  = any amino acid except proline

$X_4$  = any amino acid

$n$  = repeat between 0-6 times

$\{0-6\}$

$[-P]$

$[KHWDLPY]$

Repeat symbols can  
not be combined with  
excluding symbols.

Claim 1: A peptide having the sequence

$X_1 - X_2 - \text{PRO} - \text{THR} - \text{ALA} - X_3 - (X_4)_n$

wherein

$X_1$  is an amino acid other than proline,

$X_2$  is Lys, His, Trp, Asp, Leu, Pro, or Tyr,

$X_3$  is an amino acid other than proline,

$X_4$  is any amino acid, and  $n$  repeats 0-6 times.

## Query

$[-P][KHWDLPY]PTA[-P].\{0-6\}/SQSP$

## Select the sequence symbols to use in the query

- Use single letter codes to represent the residues
- Use square brackets to represent possible residues at a specific location
- Use square brackets and minus sign to exclude residues
- Use a period (.) to represent any amino acid
- Use curly braces { } to represent repeats

# Exercise 1

## Query

**[-P][KHWDLPY]PTA[-P].{0-6}/SQSP**

To search for...	Use this field code	Retrieves	Example
Exact Sequence	/SQEP	Exact match; same length	=> <b>S FCFWKTCT/SQEP</b>
Subsequence	/SQSP	Sequences in which the query sequence may or may not be embedded	=> <b>S LAGLL/SQSP</b>
Exact Family	/SQEFP	Functionally similar amino acids; same length	=> <b>S YGGFL/SQEFP</b>
Subsequence Family	/SQSFP	Functionally similar amino acids; may or may not be embedded	=> <b>S ATCXAWV/SQSFP</b>
Sequence Length	/SQL	Sequences of a certain length	=> <b>S SQL&lt;=10</b>
Annotation	/NTE	Sequences with the search term in the NTE field	=> <b>S MULTICHAIN/NTE</b>

# Exercise 1

## Query

`[-P][KHWDLPY]PTA[-P].{0-6}/SQSP`

### Run SQSP sequence query in REGISTRY

`=> FILE REGISTRY`

`=> S [-P][KHWDLPY]PTA[-P].{0-6}/SQSP`

`L1 101893 [-P][KHWDLPY]PTA[-P].{0-6}/SQSP`

`=> S L1 AND 6-12/SQL`

`L2 156 L1 AND 6-12/SQL`

`=> D L2 SQIDE`

Gaps, wildcards, and brackets are invalid for a SQEP search.

Refine the answers with sequence length (SQL).

Type D SQIDE to see sequence identification information, including the hit sequence.

To search for...	Use this field code	Retrieves
Exact Sequence	/SQEP	Exact match; same length

#### Some display formats

Format	Content
ALL	All available fields, including sequence data and the 10 most recent CA references
SQD	Sequence data, one-letter codes
SQD3	Sequence data, three-letter codes
SQIDE	Sequence data, CN, MF, SR, LC, DT.CA, RL, REF
HIT	All fields containing hit terms
KWIC	All hit terms plus 20 words on either side

# Exercise 1

## Query

`[-P][KHWDLPHY]PTA[-P].[0-6]/SQSP`

### Example answer

```

RN  141785-31-1  REGISTRY
CN  L-Norleucinamide, L-lysyl-L-alanyl-L-arginyl-L-valyl-L-tyrosyl-
     L-prolyl-L- threonyl-L-alanyl- (9CI)  (CA INDEX NAME)
FS  PROTEIN SEQUENCE; STEREOSEARCH
SQL  9
NTE  modified
  
```

type	location		description
terminal mod.	N1e-9	-	C-terminal amide
uncommon	N1e-9	-	-

```

SEQ      1  KARVYPTAX
         |  |  |
         |  |  |
         X1 X2 X3
  
```

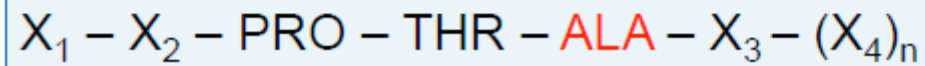
HITS AT: 4-9

This sequence does not have an X4 residue (zero repeats) .



# Exercise 2

Dependent Claim 2 requires a modification on the alanine residue in the sequence of Claim 1:



# Exercise 2

Run SQSP sequence query Modifications are searched in the NTE field

```
=> FILE REGISTRY
```

```
=> S [-P][KHWDLPY]PTA[-P].{0-6}
L1      101893 [-P][KHWDLPY]PTA
```

```
=> S L1 AND 6-12/SQL
```

```
L2      156 L1 AND 6-12/SQL
```

```
=> D L2 SQIDE
```

```
=> S ALA/NTE
```

```
L3      60197 ALA/NTE
```

```
=> S (ALA (P) MOD?)/NTE
```

```
L4      38099 (ALA (P) MOD?)/NTE
```

```
=> S (ALA (P) STEREO?)/NTE
```

```
L5      19233 (ALA (P) STEREO?)/NTE
```

```
=> S (ALA (P) BRIDG?)/NTE
```

```
L6      3437 (ALA (P) BRIDG?)/NTE
```

```
=> S (ALA (P) REPLAC?)/NTE
```

```
L7      391 (ALA (P) REPLAC?)/NTE
```

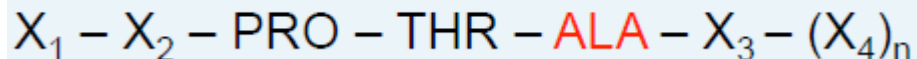
```
=> S L2 AND (L4 OR L5 OR L6 OR L7)
```

```
L8      3 L2 AND (L4 OR L5 OR L6 OR L7)
```

Use the (P) operator to ensure that alanine and the modification are listed on the same row within the Notes table.

To search for...	Use this field code	Retrieves	Example
Annotation	/NTE	Sequences with the search term in the NTE field	=> S MULTICHAIN/NTE

# Exercise 2



Information is listed in the NTE table

=> D SQIDE

RN 365254-57-5 REGISTRY

CN Glycine, L-alanyl-L-cysteinyl-L-valyl-L-tryptophyl-L-prolyl-L-threonyl-3-cyclohexyl-L-alanyl-L-tryptophyl-L-asparaginyl-L-cysteinyl- (9CI) (CA INDEX NAME)

OTHER NAMES:

CN 42: PN: WO0172771 FIGURE: 8-2 claimed protein

FS PROTEIN SEQUENCE; STEREOSEARCH

SQL 11

NTE modified (modifications unspecified)

type	location	description
modification	Ala-7	cyclohexyl<Chx>

Specific modifications can be searched in REGISTRY. See Quick Reference Guide.

Alanine is modified by the addition of cyclohexyl.

# Exercise 2



https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins



U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

BLAST® >> blastp suite

BLAST® >> blastp suite

blastn blastp blastx tblastn tblastx

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>pep1
vwptawd
```

```
>pep1
vwptawd
```

Or, upload file

Or, upload file

파일 선택 선택된 파일 없음

파일 선택 선택된 파일 없음

Job Title

Job Title

pep1

pep1

Enter a descriptive title for your BLAST search

Enter a descriptive title for your BLAST search

Align two or more sequences

Align two or more sequences

Choose Search Set

Choose Search Set

Database

Non-redundant protein sequences (nr)

Database

Non-redundant protein sequences (nr)

Organism

Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude

Models (XM/XP)  Non-redundant RefSeq proteins (WP)  Uncultured/er

Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
  - blastp (protein-protein BLAST)
  - PSI-BLAST (Position-Specific Iterated BLAST)
  - PHI-BLAST (Pattern Hit Initiated BLAST)
  - DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
- Choose a BLAST algorithm

BLAST

Search database nr using Blastp (protein-protein BLAST)

Show results in a new window

Algorithm parameters

# Exercise 2

← → ↻ https://blast.ncbi.nlm.nih.gov/Blast.cgi ☆ ⓘ ⋮

**NIH** U.S. National Library of Medicine **NCBI** National Center for Biotechnology Information Sign in to NCBI

**BLAST** » **blastp suite** » RID-DW8DEV7H015 Home Recent Results Saved Strategies Help

Format Request Status

[Formatting options]

Job

← → ↻ https://blast.ncbi.nlm.nih.gov/Blast.cgi

**NIH** U.S. National Library of Medicine **NCBI** National Center for Biotechnology Information

**BLAST** » **blastp suite** » RID-DW8DEV7H015

Format Request Status

BLAS

[Formatting options]

NCBI

Natio

8600

Polici

**Job Title: pep1**

**ⓘ Your search parameters were adjusted to search for a short input sequence.**

Request ID	DW8DEV7H015
Status	Searching
Submitted at	Thu May 16 22:51:26 2019
Current time	Thu May 16 22:51:38 2019
Time since submission	00:00:11

This page will be automatically updated in 2 seconds

# Exercise 2

BLAST » blastp suite » RID-DW8G465C015

Home Recent Results Saved Strategies Help

## BLAST Results

Your search parameters were adjusted to search for a short input sequence.

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

[YouTube](#) [How to read this page](#) [Blast report description](#) **NEW** [Click here to see the new BLAST results page](#)

Job title: pep1

RID [DW8G465C015](#) (Expires on 05-18 10:52 am)  
Query ID [Icd|Query\\_122009](#)  
Description pep1  
Molecule type amino acid  
Query Length 7

Database Name nr  
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
Program BLASTP 2.9.0+ [Citation](#)

Other reports: [Search Summary](#) [[Taxonomy reports](#)] [[Distance tree of results](#)] [[Multiple alignment](#)] [[MSA viewer](#)]

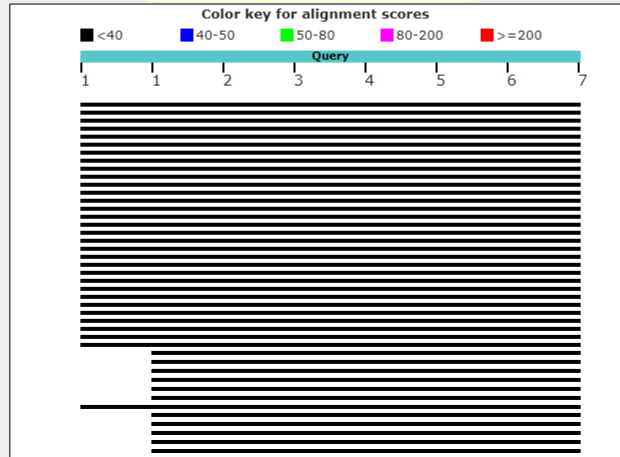
### Graphic Summary

Show Conserved Domains

No putative conserved domains have been detected

Distribution of the top 101 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



Questions/comment

# Exercise 2

← → ↻ https://blast.ncbi.nlm.nih.gov/Blast.cgi

## Descriptions

### Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/> DUF4982 domain-containing protein [Solitalea canadensis]	29.5	29.5	100%	22	100.00%	<a href="#">WP_014679911.1</a>
<input type="checkbox"/> hypothetical protein DMF86_22110 [Acidobacteria bacterium]	29.5	29.5	100%	22	100.00%	<a href="#">PYR72903.1</a>
<input type="checkbox"/> hypothetical protein DMF85_06420 [Acidobacteria bacterium]	29.5	29.5	100%	22	100.00%	<a href="#">PYR59962.1</a>
<input type="checkbox"/> Riboflavin transporter [Maliponia aquimaris]	29.5	29.5	100%	22	100.00%	<a href="#">SMX43657.1</a>
<input type="checkbox"/> DMT family transporter [Maliponia aquimaris]	29.5	29.5	100%	22	100.00%	<a href="#">WP_094021791.1</a>
<input type="checkbox"/> glycoside hydrolase family 2 protein [Pedobacter sp. PACM 27299]	27.4	27.4	100%	124	85.71%	<a href="#">WP_082461532.1</a>
<input type="checkbox"/> glycoside hydrolase [Pedobacter sp. PACM 27299]	27.4	27.4	100%	124	85.71%	<a href="#">ALL08785.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Solitalea longa]	27.4	27.4	100%	124	85.71%	<a href="#">WP_103798368.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Pedobacter antarcticus]	27.4	27.4	100%	124	85.71%	<a href="#">WP_074587852.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Pedobacter antarcticus]	27.4	27.4	100%	124	85.71%	<a href="#">WP_037437960.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Pedobacter antarcticus]	27.4	27.4	100%	124	85.71%	<a href="#">WP_088775952.1</a>
<input type="checkbox"/> beta-galactosidase [Pedobacter antarcticus]	27.4	27.4	100%	124	85.71%	<a href="#">SDL76401.1</a>
<input type="checkbox"/> ERG4/ERG24 ergosterol biosynthesis protein [Hessellinella vesiculosus]	27.4	45.8	100%	125	85.71%	<a href="#">ORX49430.1</a>
<input type="checkbox"/> MULTISPECIES_YnfA family protein [Variororax]	27.4	27.4	100%	131	85.71%	<a href="#">WP_125950686.1</a>
<input type="checkbox"/> DNA internalization-related competence protein ComEC/Rec2 [Robbsia andropogonis]	26.9	26.9	100%	175	85.71%	<a href="#">WP_024902369.1</a>
<input type="checkbox"/> DNA internalization-related competence protein ComEC/Rec2 [Robbsia andropogonis]	26.9	26.9	100%	175	85.71%	<a href="#">WP_052738378.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Mucilaginibacter sp. HYN0043]	26.9	26.9	100%	176	85.71%	<a href="#">WP_119411025.1</a>
<input type="checkbox"/> glycoside hydrolase family 2 protein [Mucilaginibacter oryzae]	26.9	26.9	100%	176	85.71%	<a href="#">WP_109609233.1</a>
<input type="checkbox"/> glycoside hydrolase family 2 protein [Cytophagales bacterium B6]	26.9	26.9	100%	176	85.71%	<a href="#">WP_031267789.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Mucilaginibacter sp. RS1]	26.9	26.9	100%	176	85.71%	<a href="#">WP_114940216.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Mucilaginibacter kameinonensis]	26.9	26.9	100%	176	85.71%	<a href="#">WP_112574562.1</a>
<input type="checkbox"/> DUF4982 domain-containing protein [Mucilaginibacter kameinonensis]	26.9	26.9	100%	176	85.71%	<a href="#">WP_121809041.1</a>
<input type="checkbox"/> alpha-galactosidase [Ammibacterium kyongjense]	26.9	26.9	100%	176	85.71%	<a href="#">WP_133765948.1</a>
<input type="checkbox"/> hypothetical protein AURDEDRAFT_166132 [Auricularia subglabra TFB-10046 S55]	26.9	26.9	100%	178	85.71%	<a href="#">EJD44945.1</a>
<input type="checkbox"/> hypothetical protein CANARDRAFT_23100 [Candida] arabinofermentans NRRL YB-2248]	26.9	26.9	100%	178	85.71%	<a href="#">QDV85572.1</a>
<input type="checkbox"/> DMT family transporter [Svnechococcus sp. CB0101]	26.9	26.9	100%	178	85.71%	<a href="#">WP_010307974.1</a>
<input type="checkbox"/> DMT family transporter [Tropicibacter sp. LIMIT003]	26.9	26.9	100%	178	85.71%	<a href="#">WP_121631660.1</a>
<input type="checkbox"/> RDD family protein [Aclinochaetozetes chiangmaiensis]	26.9	26.9	100%	179	85.71%	<a href="#">WP_018333550.1</a>
<input type="checkbox"/> hypothetical protein B9255_022877 [Caenorhabditis nigoni]	26.9	26.9	100%	187	85.71%	<a href="#">PIC16193.1</a>
<input type="checkbox"/> hypothetical protein [Phycisphaeraeaceae bacterium]	26.9	26.9	100%	188	85.71%	

Questions/comments

# Exercise 2

## Alignments

Download GenPept Graphics

DUF4982 domain-containing protein [Solitalea canadensis]

Sequence ID: [WP\\_014679911.1](#) Length: 807 Number of Matches: 1

[▶ See 1 more title\(s\)](#)

Range 1: 520 to 526 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 520 [VWPTAWD](#) 526

Download GenPept Graphics

hypothetical protein DMF86\_22110 [Acidobacteria bacterium]

Sequence ID: [PYR72903.1](#) Length: 629 Number of Matches: 1

Range 1: 565 to 571 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 565 [VWPTAWD](#) 571

Download GenPept Graphics

hypothetical protein DMF85\_06420 [Acidobacteria bacterium]

Sequence ID: [PYR59962.1](#) Length: 317 Number of Matches: 1

Range 1: 253 to 259 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 253 [VWPTAWD](#) 259

Download GenPept Graphics

Riboflavin transporter [Maliponia aquimaris]

Sequence ID: [SMX43657.1](#) Length: 299 Number of Matches: 1

Range 1: 202 to 208 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 202 [VWPTAWD](#) 208

Download GenPept Graphics

DMT family transporter [Maliponia aquimaris]

## Alignments

Download GenPept Graphics

DUF4982 domain-containing protein [Solitalea canadensis]

Sequence ID: [WP\\_014679911.1](#) Length: 807 Number of Matches: 1

[▶ See 1 more title\(s\)](#)

Range 1: 520 to 526 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 520 [VWPTAWD](#) 526

Download GenPept Graphics

hypothetical protein DMF86\_22110 [Acidobacteria bacterium]

Sequence ID: [PYR72903.1](#) Length: 629 Number of Matches: 1

Range 1: 565 to 571 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 565 [VWPTAWD](#) 571

Download GenPept Graphics

hypothetical protein DMF85\_06420 [Acidobacteria bacterium]

Sequence ID: [PYR59962.1](#) Length: 317 Number of Matches: 1

Range 1: 253 to 259 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 253 [VWPTAWD](#) 259

Download GenPept Graphics

Riboflavin transporter [Maliponia aquimaris]

Sequence ID: [SMX43657.1](#) Length: 299 Number of Matches: 1

Range 1: 202 to 208 [GenPept](#) [Graphics](#)

Next Match Previous Match

Score	Expect	Identities	Positives	Gaps
29.5 bits(62)	22	7/7(100%)	7/7(100%)	0/7(0%)

Query 1 [VWPTAWD](#) ?  
[VWPTAWD](#)  
 Sbjct 202 [VWPTAWD](#) 208

Download GenPept Graphics

DMT family transporter [Maliponia aquimaris]



# Exercise 3

Claim 1. Peptides or proteins having the following sequence:



wherein  $X_1$  is an uncommon amino acid, and  $X_2$  is 2-aminobutanoic acid.

# Exercise 3

## Common amino acids

1-Letter Code	3-Letter Code	Name
A	Ala	Alanine
B	Asx	Aspartic acid or Asparagine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
J	Xle	Isoleucine or Leucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
O	Pyl	Pyrrolysine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
U	Scy	Selenocysteine
V	Val	Valine
W	Trp	Tryptophan
X	Xxx	Uncommon or Unspecified
Y	Tyr	Tyrosine
Z	Glx	Glutamic acid or Glutamine

## Uncommon amino acids

3-Letter Code	Name
Aaa	$\alpha$ -amino acid
Aad	2-aminoadipic acid (2-aminohexanedioic acid)
Aan	$\alpha$ -asparagine
Abu	2-aminobutanoic acid
Aca	2-aminocaproic acid (2-aminodecanoic acid)
Agn	$\alpha$ -glutamine
Aib	$\alpha$ -aminoisobutyric acid ( $\alpha$ -methylalanine)
Apm	2-aminopimelic acid (2-aminoheptanedioic acid)

**Note:** The codes B, J, and Z may be used only in subsequence searches (/SQSP and /QSFP).

# Exercise 3

Claim 1. Peptides or proteins having the following sequence:



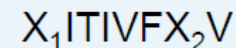
wherein  $X_1$  is an uncommon amino acid, and  $X_2$  is 2-aminobutanoic acid.

Single letter codes and three letter codes can be combined in a query

- Use the code X to search uncommon amino acids in REGISTRY
- Place single quotation marks around the three-letter code
  - Search 2-aminobutanoic acid as 'ABU'
- Claimed sequence:  $X_1ITIVFX_2V$
- Query sequence:  $XITIV'ABU'V/SQSP$

# Exercise 3

Claim 1. Peptides or proteins having the following sequence:



wherein  $X_1$  is an uncommon amino acid, and  $X_2$  is 2-aminobutanoic acid.

## Search sequence in REGISTRY

```
=> S XITIVF'ABU'V/SQSP
L1      185 XITIVF'ABU'V/SQSP

=> D SQIDE
RN      1350829-44-5  REGISTRY
CN      INDEX NAME NOT YET ASSIGNED
FS      PROTEIN SEQUENCE; STEREOSEARCH
SQL     13
NTE     modified (modifications unspecified)
```

Modifications that are not clearly stated within a document will be indexed as "undetermined modification."

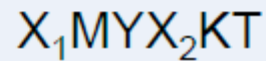
type	location	description
bridge	Thr-8 - Val-13	lactone
uncommon	Orn-6	-
uncommon	Abu-12	-
modification	Val-1	undetermined modification
modification	Orn-6	undetermined modification

```
SEQ      1 VTVVPXITIV FXV
           ↑      ↑
           6      12
```

Uncommon amino acids and modified will display as X in 1-letter codes.

# Exercise 4

Claim 1. A composition comprising a peptide having the following sequence



wherein  $X_1$  is D-phenylalanine, and  $X_2$  is D-tryptophan.

Query

FMYWKT/SQSP

# Exercise 4

## Query

FMYWKT/SQSP

D-amino acids are included in the Notes table

```
=> FILE REGISTRY
```

```
=> S FMYWKT/SQSP
```

```
L1          5 FMYWKT/SQSP
```

```
=> S L1 AND ((PHE (P) D) AND (TRP (P) D))/NTE
```

```
L2          1 L1 AND ((PHE (P) D) AND (TRP (P) D))/NTE
```

```
=> D NTE SEQ
```

```
L2  ANSWER 1 OF 1  REGISTRY
```

```
NTE  modified (modifications unspecified)
```

```
-----
```

type	-----	location	-----	description
bridge		Phe-1	- Lys-9	covalent bridge
stereo		Phe-1	-	D
stereo		Trp-4	-	D

```
-----
```

```
SEQ      1 FMYWKTMTK
```

```
HITS AT:  1-6
```

Search in the NTE field with the (P) operator.

D-amino acids are listed as *STEREO*.

# Exercise 5

Claim 1. A peptide comprising a carboxy terminal sequence

KTDSX<sub>1</sub>VCDS

wherein X<sub>1</sub> is a gap of zero to ten residues, or functional equivalents thereof.

Query

KTDS.{0-10}VCDS<sup>^</sup>/SQSFP

Possible family substitutions for KTDS\_VCDS:

<u>K</u>	<u>T</u>	<u>D</u>	<u>S</u>	<u>V</u>	<u>C</u>	<u>D</u>	<u>S</u>
H	A	B	A	I	C	B	A
R	G	E	G	M		E	G
	P	N	P	L		N	P
	S	Q	T			Q	T

Subsequence  
Family

/SQSFP

Functionally similar  
amino acids; may or  
may not be embedded

# Exercise 5

## Query

KTDS.{0-10}VCDS^/SQSFP

Possible family substitutions for KTDS\_VCDS:

<b>K</b>	<b>T</b>	<b>D</b>	<b>S</b>	<b>V</b>	<b>C</b>	<b>D</b>	<b>S</b>
H	A	B	A	I	C	B	A
R	G	E	G	M		E	G
	P	N	P	L		N	P
	S	Q	T			Q	T

Claim 1. A peptide comprising a carboxy terminal sequence

KTDSX<sub>1</sub>VCDS

wherein X<sub>1</sub> is a gap of zero to ten residues, or functional equivalents thereof.

Property	Functionally Similar Amino Acids
Neutral-Weakly Hydrophobic	Ala,Gly,Pro,Ser,Thr (A, G, P, S, T)
Hydrophilic-Acid Amine	Asn,Asp,Gln,Glu (N, D, Q, E)
Hydrophilic-Basic	Arg,His,Lys (R, H, K)
Hydrophobic	Ile,Met,Leu,Val (I, M, L, V)
Hydrophobic-Aromatic	Phe,Trp,Tyr (F, W, Y)
Cross-linking	Cys (C)

Use this symbol...	To...	Example
^	Require the base occur at the beginning or the end of the sequence	=> S ^GGAAGGG/SQSN => S CCTC^/SQSN



# Exercise 5

## Query

KTDS.{0-10}VCDS^/SQSFP

Use family searches to find functional equivalents

```
=> FILE REGISTRY
```

*sequences using  
one-letter codes (SEQ)*

```
=> S KTDS.{0-10}VCDS^/SQSFP
```

```
L1      38 S KTDS.{0-10}VCDS^/SQSFP
```

Alternate display formats  
can be used. SEQ3  
displays the sequence in  
3-letter codes.

```
=> D IDE SEQ SEQ3
```

```
RN 1282622-50-7  REGISTRY
```

```
CN L-Alanine, L-seryl-L-isoleucyl-L-seryl-L-arginyl-L-threonyl-L-  
.alpha.-glutamyl-L-alanyl-L-alanyl-L-.alpha.-aspartyl-L-leucyl-  
L-cysteinyl-L-glutamyl- (CA INDEX NAME)
```

```
SEQ      1 SISRTEAADL CQA  
=====
```

```
HITS AT:  4-13
```

```
SEQ3     1 Ser-Ile-Ser-Arg-Thr-Glu-Ala-Ala-Asp-Leu-  
          ===  ===  ===  ===  ===  ===  ===  
          11 Cys-Gln-Ala  
          ===  ===  ===
```

```
HITS AT:  4-13
```

Possible family substitutions for KTDS\_VCDS:

K	T	D	S	V	C	D	S
H	A	B	A	I	C	B	A
R	G	E	G	M		E	G
	P	N	P	L		N	P
	S	Q	T			Q	T

**KTDS \_ \_ VCDS**  
**RTEA A D LCQA**

**Thank You !**