**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**

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

Exercises

Examples are from "Sequence Motif Searches in CAS REGISTRY^{SM} " (https://www.stninternational.org/uploads/tx_ptgsarelatedfiles/20130730_Sequence_Motif_Searches_in_REGISTRY.pdf)

Exercise 1

Claim 1: A peptide having the sequence

$X_1 - X_2 - PRO - THR - ALA - X_3 - (X_4)_n$

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

Claim 1: A peptide having the sequence

$X_1 - X_2 - PRO - THR - ALA - X_3 - (X_4)_n$

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.

Recall the sequence claim:

$X_1 - X_2 - PRO - THR - ALA - X_3 - (X_4)n$

wherein

X_1 = an amino acid other than proline   [-P]

X_2 = Lys, His, Trp, Asp, Leu, Pro, or Tyr   [KHWDLPY]

X_3 = any amino acid except proline [-P]

X_4 = any amino acid

n = repeat between 0-6 times   .{0-6}

Repeat symbols can not be combined with excluding symbols.

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

Exercise 1

Query

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

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

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.

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

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

7

Exercise 1

Query

[-P][KHWDLPY]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.      Nle-9              -            C-terminal amide
uncommon           Nle-9              -            -
------------------------------------------------------------------


SEQ        1 KARVYPTAX

HITS AT:     4-9
```

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

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

Exercise 2

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

$$X_1 - X_2 - PRO - THR - ALA - X_3 - (X_4)_n$$
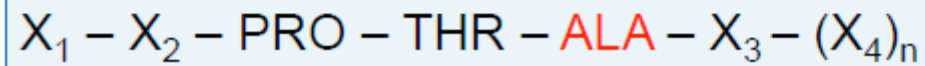
Exercise 2

Run SQSP sequence qu 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

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

10

Exercise 2

$$X_1 - X_2 - PRO - THR - ALA - X_3 - (X_4)_n$$
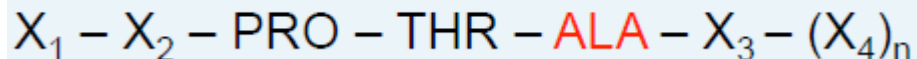
Information is listed in the NTE table

```
=> D SQIDE

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

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>
--------------------------------------------------------------------

                        Alanine is modified by the
                        addition of cyclohexyl.
```

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

Exercise 2

Exercise 2

Exercise 2

Exercise 2

Exercise 2

Exercise 3

Claim 1. Peptides or proteins having the following sequence:

$$X_1ITIVFX_2V$$

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	α-amino acid
Aad	2-aminoadipic acid (2-aminohexanedioic acid)
Aan	α-asparagine
Abu	2-aminobutanoic acid
Aca	2-aminocapric acid (2-aminodecanoic acid)
Agn	α-glutamine
Aib	α-aminoisobutyric acid (α-methylalanine)
Apm	2-aminopimelic acid (2-aminoheptanedioic acid)

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

Exercise 3

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: XITIVF'ABU'V/SQSP

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

Exercise 3

Claim 1. Peptides or proteins having the following sequence:

$$X_1 ITIVFX_2 V$$
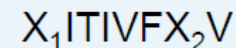
wherein X_1 is an uncommon amino acid, and X_2 is 2-aminobutanoic acid.

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)
------------------------------------------------------------------
 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
```

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

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

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

20

Exercise 4

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

$$X_1MYX_2KT$$
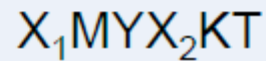
wherein X_1 is D-phenylalanine, and

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*.

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

22

Exercise 5

Claim 1. A peptide comprising a carboxy terminal sequence

$$KTDSX_1VCDS$$

wherein X_1 is a gap of zero to ten residues, or functional equivalents thereof.

Query
KTDS.{0-10}VCDS^/SQSFP

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

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:

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

Claim 1. A peptide comprising a carboxy terminal sequence

KTDSX$_1$VCDS

wherein X$_1$ 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

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


=> 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-glutaminyl-   (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
```

sequences using one-letter codes (SEQ)

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

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

KIPO
KOREAN INTELLECTUAL PROPERTY OFFICE

Thank You !