



Search Strategies for Oligonucleotides and Proteins

Greg Roland, Senior Manager
Global Legal Information Science Team (GLIST)
PIUG Biotech Meeting February 12, 2007
Cambridge, MA





New therapeutics create market opportunities



- Protein therapeutics market has increased from \$25 billion in 2001 to \$51 billion in 2005. The Protein Therapeutic Market: The Science and Business of a Growing Sector. Kalorama Information.
- The monoclonal antibody market reached \$14 billion in 2005 representing a 37% increase over the previous year. Monoclonal Antibody Therapies: Evolving into a \$30 billion market. Datamonitor. April 15, 2005
- DNA/RNA Therapies: Translating the Genome into a \$1.2 billion Market by 2010. Datamonitor February 2005
- New therapeutic approaches
 - ✓ Aptamers, siRNA, micro RNA, spiegelmers, short peptides, catalytic ribosymes, and antibodies



Who wrote this and how do I search it?

1. An Apo A-I agonist compound comprising: (i) a 22 to 29-residue D-enantiomeric peptide or peptide analogue which forms an amphipathic α -helix in the presence of lipids and which comprises formula (I): Z1-X1-X2-X3-X4-X5-X6-X7 -X8-X9-X10-X11-X12-X 13-X14-X15-X16 -X17-X 18-X19-X20-X21-X22-X23-Z2 or a pharmaceutically acceptable salt thereof, wherein: X, is D-Ala (a), Gly (G), D-Gln (q), D-Asn (n) D-Asp (d) or D-Pro (p); **X2 is a D-enantiomeric aliphatic residue;** X3 is D-Leu (l) or D-Phe (f); X4 is D-Glu (e); X5 is a D-enantiomeric aliphatic residue; X6 is D-Leu (l) or D-Phe (f); X7 is D-Glu (e) or D-Leu (l); X8 is D-Asn (n) or D-Gln (q); X9 is D-Leu (l); X10 is D-Leu (l), D-Trp (w) or Gly (G); X11 is a D-enantiomeric acidic residue; X12 is D-Arg (r); X13 is D-Leu (l) or Gly (G); **X14 is D-Leu (l), D-Phe (f) or GLy (G);** X15 is D-Asp (d); X16 is D-Ala(a); X17 is D-Leu (l); X15 is D-Asn (a) or D-Gln (q); X19 is a D-enantiomeric basic residue; X20 is a D-enantiomeric basic residue; X21 is D-Leu (l); X22 is a D-enantiomeric basic residue; X23 is absent or a D-enantiomeric basic residue; Z1 is R2N--or RC(O)NR--; Z2 is --C(O)NRR,--C(O)OR or --C(O)OH or a salt thereof; each R is independently--H, (C1-C6) alkyl, (C1-C6) alkenyl, (C1-C6) alkynyl, (C5-C20) aryl, (C6-C26) alkaryl, 5-20 membered heteroaryl, 6-26 membered alkheteroaryl or a 1 to 7-residue peptide or peptide analogue in which one or more bonds between residues 1 through 7 are independently a substituted amide, an isostere of an amide or an amide mimetic; and each "-" between residues X1, through X23 independently designates an amide linkage, a substituted amide linkage, an isostere of an amide or an amide mimetic; or (ii) a 14 to 28-residue deleted D-enantiomeric peptide or peptide analogue formula (I) **in which at least one and up to eight of residues X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X 11, X12, X13, X14, X15, X 16, X17, X18, X19, X20, X 21, X22, and X23 are optionally deleted;** or (iii) a 22 to 29-residue altered D-enantiomeric peptide or peptide analogue according to formula (I) in which at least one of residues X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X 11, X12, X13, X14, X15, X 16, X17, X18, X19, X20, X 21, X22, and X23 **is conservatively substituted with another D-enantiomeric residue.**



So many dials and levers to adjust!





Indexing policies and System limitations



- **Registry**
 - ✓ As early as 1957
 - ✓ Expanded coverage of sequences beginning in 1988
 - ✓ Peptides
 - Register single residue
 - Sequence searchable at 4 residues using exact
 - Text or structure search di- and tri-peptides
 - Help AAU for Uncommon residues
 - “.” as filler or {1}
 - ✓ NA
 - Sequence searchable at 9 residues
 - Sequences <9 residues searchable by name, molecular formula and structure
 - ✓ Starting March 2005 no longer registering sequences from mega-patents (>4,000 sequences).
 - ✓ Only claimed sequences indexed before 1999
 - ✓ The same sequence may have more than one Registry number
- **DGENE or GeneSeq**
 - ✓ 1981 to present
 - ✓ Peptides
 - Sequence searchable >4 residues
 - Structure search sequences less than 4 residues
 - Use X for uncommon residues
 - ✓ NA
 - 9 bp
 - Use N for uncommon or filler
 - ✓ For Getseq use “.” as filler



Search options



- **What do you want to accomplish?**
 - ✓ **A sequence search has four components to consider: query, database, program, and search purpose/goal.**
- **Search method**
 - ✓ **Text**
 - ✓ **EXACT**
 - ✓ **BLAST (Altschul)**
 - **Faster than FASTA without significant loss of ability to find similar sequences**
 - **BLAST and FASTA equivalent for highly similar sequences**
 - **Comparison of shorter sequence parts**
 - ✓ **FASTA or GETSIM (Pearson and Lipman)**
 - **FASTA better for less similar sequences**
 - **Comparison of entire sequence**
 - **More sensitive, misses fewer homologs**



Peptide coverage in Registry and DGENE



▪ Registry

=> s sql=2 and protein/fs

2518 SQL=2

5928229 Protein/FS

L2 2265 L1 AND PROTEIN/FS

=> s sql=3 and protein/fs

2600 SQL=3

5928229 PROTEIN/FS

L3 2290 SQL=3 AND PROTEIN/FS

=> s sql=4 and protein/fs

72920 SQL=4

5928229 PROTEIN/FS

L4 72445 SQL=4 AND PROTEIN/FS

▪ DGENE

=> s sql=2 and protein/fs

218 SQL=2

2789840 PROTEIN/FS

L5 215 SQL=2 AND PROTEIN/FS

=> s sql=3 and protein/fs

1386 SQL=3

2789840 PROTEIN/FS

L6 1367 SQL=3 AND PROTEIN/FS

=> s sql=4 and protein/fs

29319 SQL=4

2789840 PROTEIN/FS

L7 29258 SQL=4 AND PROTEIN/FS



Text searching in Registry



=> e dalda/cn

E1 1 DALCOTIDINE/CN
 E2 1 DALCRIODAIN/CN
 E3 2 --> DALDA/CN
 L1 2 DALDA/CN

L1 ANSWER 1 OF 2 REGISTRY COPYRIGHT 2007 ACS on STN
 RN 118476-85-0 REGISTRY
 ED Entered STN: 20 Jan 1989

CN L-Lysinamide, L-tyrosyl-D-arginyl-L-phenylalanyl- (9CI) (CA INDEX NAME)

OTHER NAMES:

CN 1: PN: WO205748 SEQID: 4 claimed protein
 CN 1: PN: WO2005001023 PAGE: 51 claimed protein

Note: Unclaimed sequences added beginning in 1999

CN DALDA

FS PROTEIN SEQUENCE; STEREOSEARCH

SQL 4

PATENT ANNOTATIONS (PNTE):

Sequence |Patent
 Source |Reference

=====+=====

Not Given|WO2002005748

|claimed
 |SEQID 4

SEQ 1 YRFK



Text searching in DGENE



=> e dalda

E1 1 DALBAHEPTIDE/BI

E2 1 DALCE/BI

E3 15 --> DALDA/BI

L5 ANSWER 2 OF 15 DGENE COPYRIGHT 2007 The Thomson Corp on STN

AN ADV77950 peptide DGENE

TI Delivering molecule such as antibiotic, antioxidant to cell, by contacting cell with carrier complex having molecule and aromatic cationic peptide comprising net positive charge, of three to ten amino acids.

DESC Aromatic cationic peptide DALDA.

KW Drug delivery; membrane; cell permeabilization.

SQL 4

SEQ

1 yrk

PA (CORR) CORNELL RES FOUND INC.

PI WO 2005001023 A2 20050106 77

PSL Claim 15; Page 51



Exact tetrapeptide searching in Registry



```

=> s yrfk/sqep
    43 YRFK/SQEP
    72965 SQL=4
L8   43 YRFK/SQEP
     (YRFK/SQEP AND SQL=4)
=> s s8 and p/dt
    3695 S8
    22286150 P/DT
L9   123 S8 AND P/DT
=> s l8 and p/dt
    22286150 P/DT
L10  14 L8 AND P/DT
L10  ANSWER 1 OF 14 REGISTRY COPYRIGHT 2007 ACS on STN
RN   881848-38-0 REGISTRY
ED   Entered STN: 25 Apr 2006
CN   L-Lysine, 2,6-dimethyl-L-tyrosyl-D-arginyl-L-phenylalanyl- (9CI) (CA
     INDEX NAME)
FS   PROTEIN SEQUENCE; STEREOSEARCH
MF   C32 H48 N8 O6
SR   CA
LC   STN Files: CA, CAPLUS, TOXCENTER, USPATFULL
SEQ  1 YRFK
=====

```

Note: a combination of 1 and 3 letter codes can be used; 3 letter codes need to be set off in single quotes and separated by a dash; also, the Expand command can be used here.

```

HITS AT: 1-4
**RELATED SEQUENCES AVAILABLE WITH SEQLINK**
NTE modified (modifications unspecified)

```

type	location	description
modification	Tyr-1 -	methyl<2; Me>



Exact tetrapeptide searching in DGENE



```
=> run getseq
PLEASE ENTER SEQUENCE PATTERN OR ?:yrfk
TYPE OF SEARCH ? (SQSP):sqep
```

```
RUN GETSEQ AT 21:15:05 ON 08 FEB 2007
COPYRIGHT (C) 2007 FIZ KARLSRUHE GMBH
L1      25 YRFK/SQEP
```

```
=> d l1 tri seq align feat
```

TRIAL or TRI AN, MTY, TI, DESC, and KW

```
L1 ANSWER 1 OF 25 DGENE COPYRIGHT 2007 The Thomson Corp on STN
AN AEJ09241 peptide DGENE
TI New composition comprising a polypeptide and a pharmaceutical, cosmetic,
agricultural or veterinary excipient, useful in treating a neoplastic
condition, pain or Alzheimer's disease in an animal.
DESC Synthetic dermorphin therapeutic peptide, SEQ ID NO:34.
KW pain; cell signaling; pharmaceutical; cosmetics; veterinary; antibody
therapy; temperature disorder; antipyretic; heat tolerance; cold
tolerance; mitochondria; inflammation; antiinflammatory; neoplasm;
cytostatic; pain; analgesic; neurological disease; Alzheimers disease;
neuroprotective; nootropic; degeneration; drug screening; bacterial
infection; antibacterial; infection; fungal infection; fungicide; viral
infection; virucide; crop improvement; dermorphin.
```

11 more hits than in Registry

```
SQL 4
SEQ
1 yrfk
====
```

HITS AT: 1-4

FEATURE TABLE:

Key	Location Qualifier
Misc-difference 2	note "D-form residue"



Exact tetrapeptide searching in DGENE



Need at least four characters in order to use left truncation

=> s dmt/feat
L2 64 DMT/FEAT

=> s l1 and l2
L3 2 L1 AND L2

=> d l3 tri seq feat

L3 ANSWER 1 OF 2 DGENE COPYRIGHT 2007 The Thomson Corp on STN

AN ADR20498 peptide DGENE

TI Reducing or preventing mitochondrial permeability transitioning, useful for treating ischemia, reperfusion, hypoxia, or neurodegenerative disease, comprises administering to the mammal an aromatic-cationic peptide.

DESC Aromatic-cationic peptide that can activate Mu-opioid receptors #1.

KW mitochondrial permeability transition; MPT; ischaemia; neurodegenerate disease; gene therapy; stroke; reperfusion; hypoxia; drug-induced MPT; Parkinson's disease; Alzheimer's; Huntington's; Amyotrophic Lateral Sclerosis; ALS; vasotropic; antiparkinsonian; neuroprotective; nootropic; anticonvulsant.

SQL 4

SEQ

1 yrfk

====

HITS AT: 1-4

FEATURE TABLE:

Key |Location|Qualifier|

Key	Location	Qualifier
Modified-site	1	note "Optionally Tyr can be 2, 6 dimethyltyrosine (2,6 Dmt), 3,5 Dmt, 2-methyltyrosine,



Exact search for a 8 bp NA in DGENE

```
=> run getseq cgcgtctc/sqen
```

```
RUN GETSEQ AT 21:43:36 ON 06 FEB 2007  
COPYRIGHT (C) 2007 FIZ KARLSRUHE GMBH
```

```
GETSEQ CONTINUING...  
GETSEQ CONTINUING...  
GETSEQ CONTINUING...  
GETSEQ CONTINUING...  
GETSEQ CONTINUING...  
L4 RUN STATEMENT CREATED  
L4 6 CGCGTCTC/SQEN
```

```
=> d l4 tri seq align
```

```
L4 ANSWER 1 OF 6 DGENE COPYRIGHT 2007 The Thomson Corp on STN  
AN AEM03321 DNA DGENE  
TI Isolated soluble receptor useful for treating e.g. psoriasis comprises  
two polypeptide units linked together through a disulfide bond, where one  
unit fused to constant region of light chain and other fused to heavy  
chain of immunoglobulin.  
DESC Human IgG gamma 1 heavy chain amplifying PCR primer, SEQ ID: 45.  
KW Therapeutic; respiratory distress syndrome; respiratory-gen.; septic  
shock; antibacterial; immunosuppressive; multiple organ failure;  
pulmonary disease; asthma; antiasthmatic; bronchitis; antiinflammatory;  
bacterial pneumonia; antibacterial; psoriasis; antipsoriatic; eczema;  
dermatological; atopic dermatitis; contact dermatitis; inflammatory bowel  
disease; gastrointestinal-gen.; ulcerative colitis; antiulcer; Crohns  
disease; immunoglobulin; IgG; antibody; heavy chain; PCR; primer; ss.
```

```
SQL 8
```

```
SEQ
```

```
1 cgcgtctc  
=====
```



Use of other fields in DGENE



L6 117860 SQL=24 AND NUCLEIC/FS

L6 ANSWER 1 OF 117860 DGENE COPYRIGHT 2007 The Thomson Corp on STN

AN AEM25969 DNA DGENE

TI Use of an oligonucleotide for manufacturing a medicament used for stimulating the proliferation of pluripotent mesenchymal stem cells.

IN Lopez R A

PA (IMMU-N) IMMUNOTECH SA.

PI EP 1728514 A1 20061206 110

AI EP 2005-104854 20050603

PRAI EP 2005-104854 20050603

PSL Claim 11; SEQ ID NO 117

DED 25 JAN 2007 (first entry)

DT Patent

LA English

OS 2007-027558 [04]

the invention.

NA 0 A; 2 C; 0 G; 22 T; 0 U; 0 Other

SQL 24

SEQ

1 tttttcttt ttctttttt tttt

=> s sql=24 and 22 t/na

128891 SQL=24

56532 22/NA\$

5494847 T/NA\$

13716 22 T/NA

(22/NA\$ (S) T/NA\$)

L13 12 SQL=24 AND 22 T/NA

Need to confirm location of nucleotides. Use Align



How BLAST works



- BLAST first searches for exact matches over the set Word (W) length.

AGTTACT
ACTTAGT

- Next, BLAST tries to extend the match in both directions using the W match as the anchor

AGTTACT
ACTTAGT

- Finally, BLAST performs a gapped alignment between the query sequence and the database sequence using a variation of the Smith-Waterman algorithm. Statistically significant alignments are then displayed to the user
- Short sequences searches need to have the Expectation (E) value increased. A short query is more likely to occur by chance in the database. Therefore, even a perfect match can have low statistical significance and may not be reported. Increasing the E value allows you to look farther down in the hit list and see matches that would normally be discarded because of low statistical significance.



Short peptide parameters



Query Length	Matrix	Word size	E value
Default (>85)	Blosum 62	3	10
<35 ≤10	PAM 30	2	20,000 50,000



BLAST a tetrapeptide in DGENE

```
=> run blast yrfk/sqp
QUERY TOO SHORT (MINIMUM STEM LENGTH OF 5)

=> run blast yrfkx/sqp
=> '000.000' : Blast: Query must be at least twice wordsize for two hit
mode

=> run blast yrfkx/sqp -m pam30 -w 2 -e 50000 -f f

HOW MANY ANSWERS WOULD YOU LIKE TO KEEP ? (ALL) OR ?:all
L18 RUN STATEMENT CREATED
L18 1014 YRFKX/SQP.-M PAM30 -W 2 -E 50000 -F F
=> s l18 and 4/sql
29319 4/SQL
L19 25 L18 AND 4/SQL
L19 ANSWER 1 OF 25 DGENE COPYRIGHT 2007 The Thomson Corp on
STN
AN AEJ09241 peptide DGENE
TI New composition comprising a polypeptide and a pharmaceutical,
cosmetic,
agricultural or veterinary excipient, useful in treating a neoplastic
condition, pain or Alzheimer's disease in an animal.
DESC Synthetic dermorphin therapeutic peptide, SEQ ID NO:34.
KW pain; cell signaling; pharmaceutical; cosmetics; veterinary; antibody
therapy; temperature disorder; antipyretic; heat tolerance; cold
tolerance; mitochondria; inflammation; antiinflammatory; neoplasm;
cytostatic; pain; analgesic; neurological disease; Alzheimers disease;
neuroprotective; nootropic; degeneration; drug screening; bacterial
infection; antibacterial; infection; fungal infection; fungicide; viral
infection; virucide; crop improvement; dermorphin.
```

```
SQL 4
SEQ
1 yrfk

BLASTALIGN
Query = 5 letters
Length = 4
Score = 17.6 bits (34), Expect = 2e-04
Identities = 4/4 (100%), Positives = 4/4 (100%)
Query: 1 YRFK 4
YRFK
Sbjct: 1 YRFK 4
```



BLAST a 17 residue peptide in DGENE

```
=> run blast DFKAFYDKVAEKFKKEAF/sqp
L11 546 DFKAFYDKVAEKFKKEAF/SQP. -E 10.0
=> sor sql d
=> s l11 and sql=17
    211252 SQL=17
L20 5 L11 AND SQL=17
```

Sort by sequence length in descending or ascending order

```
=> d tri seq align 1-5
L20 ANSWER 1 OF 5 DGENE COPYRIGHT 2007 The Thomson Corp on STN
AN AEM34848 peptide DGENE
TI New peptide comprising specific amino acid sequence or specific retro
    amino acid sequence of a specific peptide useful for the treatment of a
    vascular condition e.g. atherosclerotic plaque.
DESC Therapeutic helical peptide apoA-I mimetic, SEQ ID:1184.
SQL 17
SEQ
```

```
1 dwfkafydkv aekfkea
BLASTALIGN
Query = 17 letters
Length = 17
Score = 34.3 bits (77), Expect = 2e-08
Identities = 15/15 (100%), Positives = 15/15 (100%)
Query: 2 FKA FYDKVA EKFK EA 16
      FKA FYDKVA EKFK EA
Sbjct: 3 FKA FYDKVA EKFK EA 17
```

ALIGN display only gives region of similarity plus it is free



BLAST a 17 residue peptide in DGENE

⇒ `run blast DFKAFYDKVAEKFKKEAF/sqp -m pam30 -w 2 -e 50000 -f f`

9973 ANSWERS FOUND BELOW EXPECTATION VALUE OF 50000.0

Significantly more answers found by increasing E value

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

By answering "End" no BLAST fee is charged but you find out how many answers were retrieved.
Good approach for testing parameters



Parameters for short NA



Query Length	Filter Setting	Word size*	E value
Standard BLASTN	On	11	10
<100 bases	Off	7	1000
≤10	Off	7	50000

*Minimum Word size is 7



BLAST a 21 bp NA in DGENE

```
=> run blast GACCUGCCUCCUCAUCGUCTT/sqn
L18 35 GACCUGCCUCCUCAUCGUCTT/SQN. -E 10.0
=> s l18 and sql=21
    236420 SQL=21
L19 2 L18 AND SQL=21

L19 ANSWER 2 OF 2 DGENE COPYRIGHT 2007 The Thomson Corp on STN
AN AEE39625 RNA DGENE
TI New isolated polynucleotide, which is able to inhibit RhoA or RhoC
expression by transcriptional interference, useful for treating a solid
tumor cancer, in particular aggressive breast cancer, and for targeting
metastasis in cancer.
DESC Anti-RhoC siRNA (sense strand).
KW RhoC; Ras gene; oncogene; GTPase; RNA interference; RNAi; small
interfering RNA; siRNA; gene silencing; DNA-RNA hybrid; drug delivery;
breast tumor; neoplasm; cytostatic; endocrine disease; gynecology and
obstetrics; metastasis; hyperproliferation; ds.
SQL 21
SEQ
 1 gaccugccuc cucaucguct t
BLASTALIGN
Query = 21 letters
Length = 21
Score = 42.1 bits (21), Expect = 7e-11
Identities = 21/21 (100%)
Strand = Plus / Plus

Query: 1 gacctgcctcctcatcgtctt 21
      |||
Sbjct: 1 gacctgcctcctcatcgtctt 21
```



BLAST a 21 bp NA in DGENE

=> run blast GACCUGCCUCCUCAUCGUCTT/sqn -f f -w 7 -e 1000
HOW MANY ANSWERS WOULD YOU LIKE TO KEEP ? (ALL) OR ?:all

L14 9961 GACCUGCCUCCUCAUCGUCTT/SQN.-F F -W 7 -E 1000

=> s l15 and 2 T/NA
873026 2/NA\$
5494847 T/NA\$
227951 2 T/NA

(2/NA\$ (S) T/NA\$)
L17 2 L15 AND 2 T/NA

=> d tri seq align

L17 ANSWER 1 OF 2 DGENE COPYRIGHT 2007 The Thomson Corp on STN

AN AEE39636 RNA DGENE

TI New isolated polynucleotide, which is able to inhibit RhoA or RhoC expression by transcriptional interference, useful for treating a solid tumor cancer, in particular aggressive breast cancer, and for targeting metastasis in cancer.

SQL 21

SEQ

1 gacgaugagg aggcagguct t

BLASTALIGN

Query = 21 letters

Length = 21

Score = 38.2 bits (19), Expect = 1e-09

Identities = 19/19 (100%)

Strand = Plus / Minus

100% identity over 19 bp

Query: 1 gacctgcctcctcatcgtc 19

|||||

Sbjct: 19 gacctgcctcctcatcgtc 1

Turning filter off and changing W size to 7 returned significantly more hits!



BLAST a 8 bp NA in DGENE



=> run blast cgcgtctc/sqn -w 7 -e 50000 -f f

9827 ANSWERS FOUND BELOW EXPECTATION VALUE OF 50000.0

NO ANSWERS FOUND at E VALUE <50,000



Cost comparison



- Registry
 - ✓ Connect Hour Fee (per hour) \$39.00
 - ✓ Exact Sequence \$7.25
 - ✓ BLAST Search \$29.35

- DGENE
 - ✓ Connect Hour Fee (per hour). \$120.00
 - ✓ RUN GETSEQ \$19.78
 - ✓ RUN BLAST \$26.08

- PCTGEN
 - ✓ Connect Hour Fee (per hour) \$90
 - ✓ Sequence Search per RUN GETSEQ \$11.38
 - ✓ Homology search per RUN BLAST \$14.22

*PCTGEN follows Standard ST.25 found at <http://www.wipo.int/scit/en/standards/pdf/03-25-01.pdf>

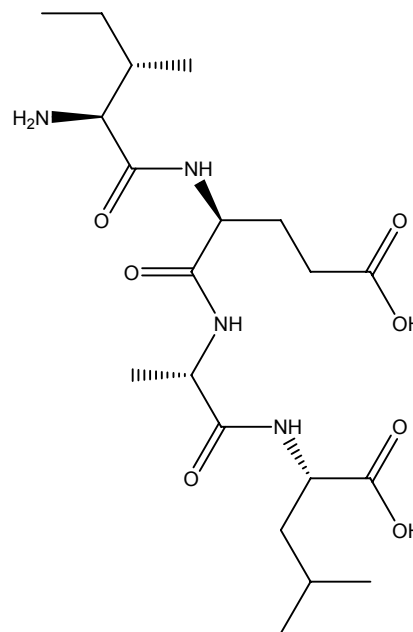
Four or more unbranched amino acids and 10 nucleotides



Structure search options



- **ChemDraw Ultra 8.0**
 - ✓ Enter "HlleGluAlaLeuValOH".
 - ✓ Select the text using Lasso or Marquee tool from the main tool bar.
 - ✓ Select "Expand Label" from the "Structure" menu to get the structure
 - ✓ Search in Reg or DCR of WPIX
 - ✓ Locate name in Reg CN field
 - ✓ Select "Convert name to structure and paste in
 - ✓ Consult chemistry searcher
 - ✓ MSS on Q-O has 3 letter AA codes linked by X bond covering sections Derwent sections B, C, and E





Summary



- Use both Registry and DGENE for comprehensive coverage
- Exact sequence searching is possible with 4 or fewer residues
- Gaps, Wildcards, and Brackets are invalid for Exact search
- Modified residues can be found using unmodified sequence query
- Use SEQLINK in Registry to find related sequences; It's free
- BLAST requires at least 5 residues
- Use "." or {1} as place holder
- Help AAU for a list of uncommon residues in Registry
- Use X for uncommon residues in DGENE
- Adjust the Word score to 2 and E value to 50,000 for peptides less than 10
- Adjust the Word score to 7 and use E value 50,000 for NA less 10



Who has questions?

