Searching in Applications Containing Bio-Sequences

Ram R. Shukla

Supervisory Patent Examiner Art Unit 1634

571 272 0735

ram.shukla@uspto.gov

Sequence Structure

> A polynucleotide sequence that encodes a polypeptide (cDNA/ Genomic)
>
> Oligomers
>
> Probes/Primers
>
> Fragments

Function

Antisense/Complements

RNAi/Ribozymes/Triplex

Aptamers

Amino Acid Binding Domains

Immunostimulatory CpG Sequences

Transgene

Regulatory Sequences

Other

    Accession Number

    Single Polynucleotide Polymorphism

    rs (Reference SNP) number

    Biological Deposit

Types of Molecules Claimed: Amino Acid Sequences

Structure

    An amino acid sequence

    Oligopeptide

    Specifically Identified Fragments

    Accession Number

    A polypeptide encoded by a polynucleotide sequence

Function

    Nucleic Acid Binding Domains

    Antibody

    Dominant Negative Mutant

Types of Sequences Claimed: Sequence Disclosure and Compliance

IF an application discloses a nucleotide or an amino acid sequence and

(Sequences may be anywhere in the application including specification, drawings, abstract)

the nucleic acid sequence is a specific unbranched sequence of 10 or more nucleotides; and/or

the amino acid sequence is a specific unbranched sequence of 4 or more amino acids,

THEN the application must be analyzed for compliance with the sequence rules (37 CFR 1.821-1.825).

See MPEP 2422 for Sequence Compliance Requirements
*www.cabic.com/bcp/060408/R**Wax**_SRCPAI.ppt*

Types of Sequences Claimed: Accession Number

If a sequence is claimed by an Accession No, the specific sequence has to be disclosed in the specification.

If the specification does not disclose the sequence of the Accession No, the office may object to the specification.

If the sequence is added to the Specification

It must be determined if the sequence has been properly incorporated by reference and adds no new matter.

The sequence must be uniquely identified.

For discussion of incorporation by reference of a sequence, see the BCP presentations by Jean Witz at the Sept 2008 BCP meeting (http://www.cabic.com/bcp/090908/JWitz_IBR.ppt) and Julie Burke at the June 2008 BCP meeting (http://www.cabic.com/bcp/060408/JBurke_SREI.ppt)

The sequence recited in the Claim is used as the search query.

The interpretation of a claim requires a sequence to be present and used as a query for a search of sequence databases.

Claim interpretation

Complementary/Antisense sequences

Reverse Transcription/Translation

RNA reverse transcribed to DNA

Protein back translated to DNA

cDNA, genomic DNA

Oligomer/primers/probes

Smith-Waterman

Finds an optimal local alignment between two protein (p2p) or two nucleic (n2n) sequences.

Uses a two-dimensional matrix to look for the highest scoring alignment

Similarity score is calculated based on:

Comparison matrix: **provides probability scores for all substitutions between pairs of residues**

Gap penalties: **cost of inserting or deleting residues in the alignment**

There are two gap penalty models:

Non-affine: **a single gap penalty value is applied to any unmatched residue**

Affine: **a penalty for a gap is calculated as** *gapop+gapext*l*, **where** *gapop* **is the penalty for opening a gap,** *gapext* **is the penalty for extending the gap, and** *l* **is the length of the gap.**

*Smith and Waterman, Advances in Applied Mathematics, 2:482-489 (1981)*

Query Match   1.8%; Score 132.8; DB 8; Length 403;
Best Local Similarity   77.4%; Pred. No. 8.5e-26;
Matches 161; Conservative   0; Mismatches   47; Indels   0; Gaps   0;

```
Qy      7280 GACTCTTAGGGACCCCATGGACTGCAGCATGCCAAGTTTCCCTGTCCTTCACTATCACTT 7339
             |||||||| | |||||||||    ||||| |||| |   |||||||| |||| | | |
Db       213 GACTCTTTGCGACCCCATGAATCACAGCACGCCAGGCCTCCCTGTCCATCACCAACTCCC 154

Qy      7340 GGAGTTTGCTCAAACTCATGTCCATTGAGTCTGTGATGCCATTCAACCACCTCATCCTCT 7399
             || |||  |||| ||||||||||||| ||||| |||| ||||| || ||| ||||||||||
Db       153 GGGGTTCACTCAGACTCATGTCCATAGAGTCAGTGACGCCATCCAGCCATCTCATCCTCT 94

Qy      7400 GTGGCCTCCTTGTCCTCCTGCCGTCAGTCTTTCCCAGCGTAAGGGTCTTTTCCAGTGAGT 7459
             || | | |||| |||||||||||  || ||  ||||||| | || |||||||||| |||||
Db        93 GTTGTCCCCTTCTCCTCCTGCCCCCAATCCCTCCCAGCATCAGAGTCTTTTCCAATGAGT 34

Qy      7460 CAGCTGTTTGCAGCAGTTGGCTAAAGAA 7487
             || || || |||  || |||| |||| |
Db        33 CAACTCTTCGCATGAGGTGGCCAAAGGA 6
```

12

Search Request Considerations:
Does the claim allow for gaps?
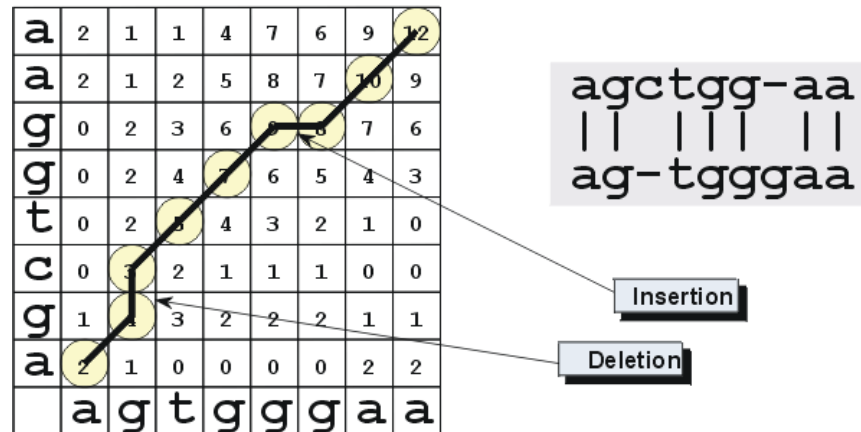
In each cell, the algorithm stores the highest score of all possible paths leading to the cell

Each path can be described as a traversal of an *automaton* consisting of three states:

Match: **two residues are matched**

Insert: **the query matches a gap to a database residue**

Delete: **the query matches a residue with a gap in the database sequence**

The path leading to the highest score can be of any length, and, by definition of a *local alignment*, doesn't have to start at the beginning or end at the end of both sequences.
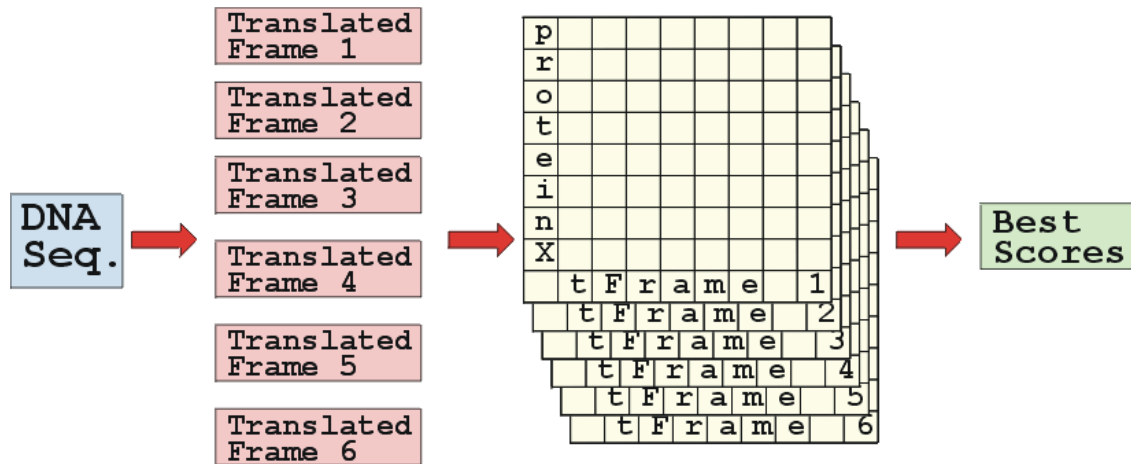
First translate the nucleic sequence into three or six reading frames, then align each frame independently to the protein sequence; in the results, indicate the frame that produced each high-scoring hit.

Types of Nucleotide Sequence Searching: Standard Search

Query using the full length of the SEQ ID NO (up to 10 Kb in size)
useful for finding full length hits.

hit size could be limited to a size range by requesting a "length-limited" search (range provided by the examiner).

the search parameters are the default parameters-Gap Opening Penalty 10 & Gap Extension Penalty of 1.

Interpretation of the search results is needed to find fragments and genomic sequences.

Fragments are buried in the hit list.

The presence of introns in the database sequence results in low scores.

For a large sequence, 10 kb or greater, multiple large subsections of the sequence are used as a query to search the databases.

For a genomic sequence,

> If exons and their boundaries are known, several exons are searched.

> If exons are not known, multiple large subsections of the sequence are used as a query to search the database.

Impact of Sequence Identity and Length

Adjustment of search parameters (e.g. Smith-Waterman Gap values) influences % Query Match value.

- % Query Match value approximates overall identity
- Mismatches
    - Varying Degrees of Percent Identity
- Gaps
    - Insertion or Deletions
    - Gap Extensions
- Wild Cards
- Complements/Matches

Standard Oligomer:

Prioritizes the longest uninterrupted hits.

Accomplished by significantly increasing the gap penalty.

The hit size could be limited to a size range by requesting a "length-limited" search (range provided by the examiner).

Not optimal for finding small sequences that are 100% identical or complementary.

Score Over Length Searching

Optimal for finding small sequences that are 100% identical or complementary.

Calculated by dividing the hit score by the hit length

hit "score" represents the number of perfect matches between query and hit.

The number of perfect matches relative to a hit's length is calculated (Score/Length).

hits then sorted by Score/Length value.

Hits with a Score/Length value closer to "1" are prioritized.

GenEMBL

N_Genseq

Issued_Patents_NA

EST

Published_Applications_NA

A-Geneseq

UniProt

PIR

Published_Applications_AA

Issued_AA

Published_Applications_NA

Issued_NA

Pending_Applications_NA

Published_Applications_AA

Issued_AA

Pending_Applications_AA

```
                              GenCore version 5.1.6
1                 Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:              November 11, 2004, 05:19:18 ; Search time 3263 Seconds
                                            (without alignments)
                                            9710.113 Million cell updates/se

Title:               US-XX-XXX-XXXC-15
Perfect score:       670                                          2
Sequence:            1 ctacctggagcgaaatggc..........gcgccagtagacaattagcc 670

Scoring table:       IDENTITY_NUC
                     Gapop 10.0 , Gapext 1.0          3

Searched:            4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters:        9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :           GenEmbl:*
                 1:   gb_ba:*
                 2:   gb_htg:*
                 3:   gb_in:*
                 4:   gb_om:*
                 5:   gb_ov:*
                 6:   gb_pat:*
                 7:   gb_ph:*
                 8:   gb_pl:*
                 9:   gb_pr:*
                10:   gb_ro:*
                11:   gb_sts:*
                12:   gb_sy:*
                13:   gb_un:*
                14:   gb_vi:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```
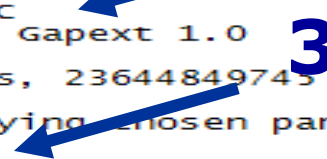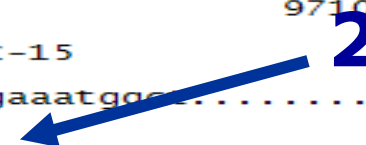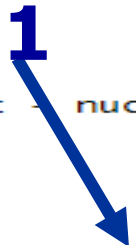
1: Query length/Perfect Score;
2: Gap Parameters;
3: Minimum and Maximum Length Limitations

Search Results

SUMMARIES

```
            %
Result    Query
  No.  Score  Match  Length  DB   ID                     Description
-------------------------------------------------------------------------
   1    670  100.0    670    6   E59319                 E59319 Vector expr
   2    670  100.0    670    6   AR207433               AR207433 Sequence
   3  650.8   97.1   1154    5   AY334554               AY334554 Misgurnus
   4    534   79.7   1158    5   HNPGH                  X60473 H.nobilis (
   5    534   79.7   1170    5   AY170124               AY170124 Megalobra
   6  530.8   79.2   1170    5   HMPGH                  X60475 H.mulitrix
   7  530.8   79.2   1171    5   CIPGH                  ... dellus (
   8  530.8   79.2   1171    5   CTEGRH                 ... opharyn
   9  530.2   79.1   1279    5   AY375302               AY375302 Ictiobus
  10  529.4   79.0   1158    5   CCGH                   X13670 Cyprinius c
  11    526   78.5   1179    5   AF069399               AF069399 Carassius
  12    519   77.5   1279    5   AY375301               AY375301 Ictiobus
  13  516.2   77.0    633    5   AY616661               AY616661 Ctenophar
  14  508.2   75.9    633    5   AF389238               AF389238 Mylophary
  15  506.8   75.6   1164    5   CYIGH                  M27000 Cyprinus ca
  16  505.2   75.4   1195    6   AR222492               AR222492 Sequence
  17  502.8   75.0   1180    5   AY265352               AY265352 Carassius
  18  501.8   74.9    633    5   AY160977               AY160977 Carassius
  19  496.8   74.1   1117    5   AF069398               AF069398 Carassius
  20  496.2   74.1   1083    6   AR222493               AR222493 Sequence
  21    495   73.9   1150    5   AF140281               AF140281 Cirrhinus
  22  493.8   73.7    633    5   AF389237               AF389237 Carassius
  23  488.6   72.9    894    5   AY053361               AY053361 Catla cat
  24  479.6   71.6   1065    5   AF337033               AF337033 Megalobra
  25  476.6   71.1   1065    5   AF332563               AF332563 Megalobra
  26  473.2   70.6   1104    5   AF401273               AF401273 Carassius
  27  472.4   70.5   1156    5   AF140282               AF140282 Catla cat
  28    468   69.9    885    5   AF416490               AF416490 Labeo roh
  29    468   69.9   1162    5   AF134200               AF134200 Labeo roh
  30    460   68.7    567    5   AF458105               AF458105 Cirrhinus
  31  456.6   68.1    567    5   CAU28388               U28388 Carassius a
  32  453.4   67.7    567    5   AF332594               AF332594 Cyprinus
  33  452.8   67.6    819    6   BD096861               BD096861 Method fo
  34  451.6   67.4   1047    5   AF401272               AF401272 Carassius
  35  376.8   56.2    476    5   AY286447               AY286447 Danio rer
```

**1:Score;
2:% Query Match**

26

```
Qy     1 CTACCTGGAGCGAAATGGCTAAAGCTTTAGTGCTGCTGTCTTTGGTCCTGGTCAGTGTTT  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTACCTGGAGCGAAATGGCTAAAGCTTTAGTGCTGCTGTCTTTGGTCCTGGTCAGTGTTT  60

Qy    61 TTGTGAATAATGGGACTGCCTCAGAAAACCAGAGGCTCTTCAACAACGCAGTCATCCGTG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TTGTGAATAATGGGACTGCCTCAGAAAACCAGAGGCTCTTCAACAACGCAGTCATCCGTG 120

Qy   121 TACAACACCTGCACCAGCTGGCTGCAAAAATGATCAATGACTTTGAGGACAGCCTGTTAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TACAACACCTGCACCAGCTGGCTGCAAAAATGATCAATGACTTTGAGGACAGCCTGTTAC 180

Qy   181 CTGAGGAACGCAGGCAGCTGAGTAAAATCTTCCCATTGTCCTTCTGCAACTCTGACTCTA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGAGGAACGCAGGCAGCTGAGTAAAATCTTCCCATTGTCCTTCTGCAACTCTGACTCTA 240

Qy   241 TAGAGGCTCCCACTGGCAAAGATGAAACGCAGAAAAGCTCTGTGCTGAAGCTGCTTCGCA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TAGAGGCTCCCACTGGCAAAGATGAAACGCAGAAAAGCTCTGTGCTGAAGCTGCTTCGCA 300

Qy   301 TCTCCTTCCGCCTCATTGAGTCTTGGGAGTATCCCAGCCAGACCCTTAGTGGAACCATCT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TCTCCTTCCGCCTCATTGAGTCTTGGGAGTATCCCAGCCAGACCCTTAGTGGAACCATCT 360

Qy   361 CAAACAGCCTGACCATCGGCAACCCCAGCCAGATCACAGAGAAGCTGGCCGATCTGAAAG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAAACAGCCTGACCATCGGCAACCCCAGCCAGATCACAGAGAAGCTGGCCGATCTGAAAG 420

Qy   421 TGGGCATCAGCGTGCTCATAAAGGGATGTCTTGATGGACAGCCAAACATGGACGATAATG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TGGGCATCAGCGTGCTCATAAAGGGATGTCTTGATGGACAGCCAAACATGGACGATAATG 480

Qy   481 ACTCCCTGCCATTGCCTTTTTGAGGATTTCTACTTGACTTTGGGGGAGAATAACCTCAGAG 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACTCCCTGCCATTGCCTTTTTGAGGATTTCTACTTGACTTTGGGGGAGAATAACCTCAGAG 540

Qy   541 AGAGCTTTCGTCTGCTGGCCTGCTTTAAGAAAGACATGCACAAGGTTGAAACCTACCTGA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGAGCTTTCGTCTGCTGGCCTGCTTTAAGAAAGACATGCACAAGGTTGAAACCTACCTGA 600

Qy   601 GGGTTGCGAACTGCAGGCGATCCCTCGATTCCAACTGTACCCTGTAGAGGGCGCCAGTAG 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGGTTGCGAACTGCAGGCGATCCCTCGATTCCAACTGTACCCTGTAGAGGGCGCCAGTAG 660

Qy   661 ACAATTAGCC  670
         ||||||||||
Db   661 ACAATTAGCC  670
```

27

```
                       GenCore version 5.1.6
              Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:           November 11, 2004, 03:54:36 ; Search time 3262 Seconds
                                                (without alignments)
                                                9713.090 Million cell updates/sec

Title:            US-XX-XXX-XXXC-15
Perfect score:    670
Sequence:         1 ctacctggagcgaaatggct..........gcgccagtagacaattagcc 670

Scoring table:    OLIGO_not
                  Gapop 60.0 , Gapext 60.0

Searched:         4526/29 seqs, 23044849745 residues

Word size :       0

Total number of hits satisfying chosen parameters:        9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :        GenEmbl:*
                  1:  gb_ba:*
                  2:  gb_htg:*
                  3:  gb_in:*
                  4:  gb_om:*
                  5:  gb_ov:*
                  6:  gb_pat:*
                  7:  gb_ph:*
                  8:  gb_pl:*
                  9:  gb_pr:*
                  10:  gb_ro:*
                  11:  gb_sts:*
                  12:  gb_sy:*
                  13:  gb_un:*
                  14:  gb_vi:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

Gap Penalties

28

Oligomer Search GenBank Hit Table Against cDNA

```
                               SUMMARIES
         I        %
Result        Query
  No.   Score Match Length DB   ID                              Description
--------------------------------------------------------------------------------
    1    670  100.0    670  6    E59319                         E59319 Vector expr
    2    670  100.0    670  6    AR207433                       AR207433 Sequence
    3    321   47.9   1154  5    AY334554                       AY334554 Misgurnus
    4    229   34.2   5099  5    AF133815                       AF133815 Misgurnus
    5    229   34.2   5108  6    E59318                         E59318 Vector expr
    6    229   34.2   5108  6    AR207432                       AR207432 Sequence
    7     38    5.7    567  5    AF332594                       AF332594 Cyprinus
    8     38    5.7    567  5    AF458105                       AF458105 Cirrhinus
    9     38    5.7    819  6    BD096861                       BD096861 Method fo
   10     38    5.7    885  5    AF416490                       AF416490 Labeo roh
   11     38    5.7    894  5    AY053361                       AY053361 Catla cat
   12     38    5.7   1150  5    AF140281                       AF140281 Cirrhinus
   13     38    5.7   1156  5    AF140282                       AF140282 Catla cat
   14     38    5.7   1162  5    AF134200                       AF134200 Labeo roh
   15     38    5.7   1164  5    CYIGH                          M27000 Cyprinus ca
   16     38    5.7   2718  5    AF418921                       AF418921 Labeo roh
   17     38    5.7   2838  5    CCGHG                          X51969 Cyprinus ca
   18     38    5.7  11576  5    AY053360                       AY053360 Catla cat
   19     35    5.2   1279  5    AY375301                       AY375301 Ictiobus
   20     32    4.8    633  5    AF389238                       AF389238 Mylophary
   21     32    4.8    633  5    AY616661                       AY616661 Ctenophar
   22     32    4.8    893  5    AY148493                       AY148493 Anguilla
   23     32    4.8   1065  5    AF332563                       AF332563 Megalobra
   24     32    4.8   1065  5    AF337033                       AF337033 Megalobra
   25     32    4.8   1158  5    HNPGH                          X60473 H.nobilis (
   26     32    4.8   1170  5    AY170124                       AY170124 Megalobra
   27     32    4.8   1170  5    HMPGH                          X60475 H.mulitrix
   28     32    4.8   1171  5    CIPGH                          X60474 C.idellus (
   29     32    4.8   1171  5    CTEGRH                         M27094 Ctenopharyn
   30     32    4.8   1279  5    AY375302                       AY375302 Ictiobus
   31     32    4.8   2393  5    AY616666                       AY616666 Anguilla
   32     32    4.8   2484  5    HYPSCGH                        M94348 Hypophthalm
   33     32    4.8   3627  5    CIGCGH                         X60988 C.idella gc
   34     32    4.8   3975  5    CIGH                           X60419 C.idellus g
   35     32    4.8   6260  5    AF463498                       AF463498 Megalobra
   36     29    4.3    567  5    CAU28388                       U28388 Carassius a
   37     29    4.3    633  5    AY160977                       AY160977 Carassius
   38     29    4.3   1104  5    AF401273                       AF401273 Carassius
   39     29    4.3   1158  5    CCGH                           X13670 Cyprinius c
   40     29    4.3   1179  5    AF069399                       AF069399 Carassius
   41     28    4.2    603  5    PNAGH                          M63713 Pangasius p
   42     28    4.2   1083  6    AR222493                       AR222493 Sequence
   43     28    4.2   1132  5    AF147792                       AF147792 Heteropne
   44     28    4.2   1176  5    PGSGHA                         L27835 Pangasianod
   45     28    4.2   1195  6    AR222492                       AR222492 Sequence
   46     28    4.2 150738  2    BX908793                       BX908793 Danio rer
c  47     28    4.2 150738  2    BX908793                       BX908793 Danio_rer
```

29

```
RESULT 10
AF416490
AF416490
LOCUS          AF416490                    885 bp      RNA        linear    VRT 02-MAR-2002
DEFINITION     Labeo rohita growth hormone protein precursor RNA, complete cds.
ACCESSION      AF416490
VERSION        AF416490.1   GI:19071231
KEYWORDS       .
SOURCE         Labeo rohita (rohu)
  ORGANISM     Labeo rohita
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
               Cypriniformes; Cyprinidae; Labeo.
REFERENCE      1   (bases 1 to 885)
  AUTHORS      Rajesh,R.  and Majumdar,K.C.
  TITLE        Labeo rohita growth hormone gene mRNA sequence, complete CDS
  JOURNAL      Unpublished
REFERENCE      2   (bases 1 to 885)
  AUTHORS      Rajesh,R.  and Majumdar,K.C.
  TITLE        Direct Submission
  JOURNAL      Submitted (05-SEP-2001) Centre for Cellular and Molecular Biology,
               Uppal Road, Hyderabad, AP 500 007, India
FEATURES             Location/Qualifiers
     source          1. .885
                     /organism="Labeo rohita"
                     /mol_type="pre-RNA"
                     /db_xref="taxon:84645"
     prim_transcript 1. .885
     5'UTR           1. .47
     CDS             48. .671
                     /codon_start=1
                     /product="growth hormone protein"
                     /protein_id="AAL84169.1"
                     /db_xref="GI:19071232"
                     /translation="MARALVLLSVVLVSLLVNQGRASDNQRLFNNVVVRVQHLHQLAA
                     KMINDFDDNLLPEDRRLLSKTIPMSFCISDYIEAPTGKDEAQRSSMLKLLRISFRLIE
                     SWELASQILSRTVSNSLTANQINEKLADLKMGISVLIKGCLDGQPNMDDNDSLPLPFE
                     DFYLTTEDNDLTKNFRLLACFKKDMHKVETYLRVANCRRSLDSNCTL"
     3'UTR           672. .885
     polyA_signal    879. .885
ORIGIN

  Query Match              5.7%;   Score 38;   DB 5;   Length 885;
  Best Local Similarity   100.0%;   Pred. No. 8.8e-09;
  Matches    38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        117 CGTGTACAACACCTGCACCAGCTGGCTGCAAAAATGAT 154
              ||||||||||||||||||||||||||||||||||||||
Db        150 CGTGTACAACACCTGCACCAGCTGGCTGCAAAAATGAT 187
```

Length-Limited (8 to 20) Oligomer Search
GenBank Hit Table cDNA

```
                        GenCore version 5.1.6
                Copyright (c) 1993 - 2004  Compugen Ltd.
     I

OM nucleic - nucleic search, using sw model

Run on:           November 11, 2004, 07:42:54 ; Search time 3264 Seconds
                                    (without alignments)
                                    9707.138 Million cell updates/sec

Title:            US-XX-XXX-XXXC-15
Perfect score:    670
Sequence:         1 ctacctggagcgaaatggct..........gcgccagtagacaattagcc 670

Scoring table:    OLIGO_NUC
                  Gapop 60.0 , Gapext 60.0

Searched:         4526729 seqs, 23644849745 residues

word size :       0

Total number of hits satisfying chosen parameters:      773916

Minimum DB seq length:  8
Maximum DB seq length: 20
Post-processing: Listing first 1000 summaries

Database :        GenEmbl:*
              1:    gb_ba:*
              2:    gb_htg:*
              3:    gb_in:*
              4:    gb_om:*
              5:    gb_ov:*
              6:    gb_pat:*
              7:    gb_ph:*
              8:    gb_pl:*
              9:    gb_pr:*
              10:   gb_ro:*
              11:   gb_sts:*
              12:   gb_sy:*
              13:   gb_un:*
              14:   gb_vi:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

Gap Penalties

Minimum & Maximum Length Limitations

```
RESULT 2
A10411/c
LOCUS           A10411                          17 bp    DNA        linear    PAT 23-NOV-1993
DEFINITION      Oligonucleotide.
ACCESSION       A10411
VERSION         A10411.1  GI:490708
KEYWORDS        .
SOURCE          synthetic construct
  ORGANISM      synthetic construct
                artificial sequences.
REFERENCE       1  (bases 1 to 17)
  AUTHORS       Harris,T.J.R., Docherty,A.J.P., Reynolds,J.J. and Murphy,G.
  TITLE         Process for the production of a protein
  JOURNAL       Patent: EP 0188312-A 10 23-JUL-1986;
                CELLTECH LIMITED
FEATURES                 Location/Qualifiers
     source              1. .17
                         /organism="synthetic construct"
                         /mol_type="unassigned DNA"
                         /db_xref="taxon:32630"

ORIGIN

  Query Match                   2.2%;  Score 15;  DB 6;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 5.1e+04;
  Matches    15; Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          222 TTCTGCAACTCTGAC 236
                |||||||||||||||
Db           17 TTCTGCAACTCTGAC 3
```

Claim 1

Claim:

An isolated polynucleotide comprising SEQ ID NO:1.

Claim Interpretation :

Comprising: must have all of SEQ ID NO:1, may include any flanking sequences, as in the claim above.

Consisting of: limited to only SEQ ID NO:1, with **No** flanking sequences.

Search Strategy:

A standard search looking for full length hits is performed.

Claim 2

Claim:

An isolated polypeptide comprising SEQ ID NO: 2.

Claim Interpretation:

Comprising: must have all of SEQ ID NO:2, may include any flanking sequences, as in the claim above.

Consisting of: limited to only SEQ ID NO:2, with **No** flanking sequences.

Search Strategy:

A standard search looking for full length hits is performed in all the amino acid databases.

Claim 3

Claim:

An isolated polynucleotide comprising a nucleotide sequence of SEQ ID NO:1

Claim Interpretation:

This claim embraces any fragment of SEQ ID NO:1 due to the language "--**a** nucleotide sequence of--."

This could be obviated by amending to read "--**the** nucleotide sequence of--."

Search Strategy:

A standard nucleotide sequence search as well as a standard oligomer search is performed using SEQ ID NO:1 as a query.

Claim 4

Claim:

> An isolated polynucleotide comprising a polynucleotide with at least 90% identity over its entire length to SEQ ID NO:1.

Claim Interpretation:

> This claim encompasses any sequence that has 90% or higher sequence identity over its entire length to SEQ ID NO:1.

Search Strategy

> A standard search looking for full length hits is performed.

> Hits having at least 90% identity will appear in the results.

Claim 5

Claim:

An isolated polynucleotide comprising a polynucleotide encoding the amino acid sequence of SEQ ID NO:2.

Claim Interpretation:

The claim encompasses any polynucleotide that encodes the polypeptide of SEQ ID NO:2.

Search Strategy:

SEQ ID NO:2 is "back translated" into a nucleic acid sequence, which is used as a query to search the nucleic acid databases .

Claim 6

Claim:

An isolated polynucleotide comprising a polynucleotide which hybridizes under stringent conditions to SEQ ID NO: 1.

Claim Interpretation:

Claim is interpreted as embracing any sequence with less than 100% complementarity or identity to SEQ ID NO:1.

Search Strategy

A standard oligomer search as well as a standard search is perfomed.

Claim 7

Claim:

An isolated polynucleotide comprising at least 15 <u>contiguous</u> nucleotides of SEQ ID NO:1.

Claim Interpretation:

The claim embraces any fragment of 15 nucleotides or greater of SEQ ID NO:1.

Search Strategy:

A standard oligomer search is performed with a length of 15 nucleotides set as the lower limit for a hit.

Claim:

An isolated polypepide comprising at least 15 <u>contiguous</u> amino acids of SEQ ID No:2.

Claim Interpretation:

The claim embraces any fragment of 15 amino acids or greater of SEQ ID NO:2.

Search Strategy:

A standard oligomer search is performed with a length of 15 amino acids set as the lower limit for a hit.

Claim 9

Claim:

An oligonucleotide consisting of 8 to 20 nucleotides which specifically hybridizes to the nucleic acid sequence of SEQ ID NO:1.

Claim Interpretation:

The specification teaches that oligonucleotides which specifically hybridize need not have 100% sequence correspondence.

Search Strategy

A Score/Length search is performed with 8 and 20 as lower and upper limits respectively.

Claim 10: Searching a SNP

Claim:

A nucleic acid comprising SEQ ID NO:1 where the nucleotide at position 101 is a T.

Claim Interpretation:

The Claim encompasses any sequence comprising SEQ ID NO:1, with a T at position 101.

Search Strategy:

A standard nucleotide sequence search is performed for SEQ ID NO:1. The examiner manually searches for any changes at position 101.

Thanks!

James (Doug) Schultz
Supervisory Patent Examiner
Art Unit 1635
571-272-0763
James.Schultz@uspto.gov

Barb O'Bryen
Scientific and Technical Information Center (STIC)

Questions?

**Ram R. Shukla
Supervisory Patent Examiner Art Unit
1634
571 272 0735
ram.shukla@uspto.gov**