WIPO

PCT/MIA/V/2
ORIGINAL: English
DATE: November 25, 1994

WORLD INTELLECTUAL PROPERTY ORGANIZATION
GENEVA

INTERNATIONAL PATENT COOPERATION UNION
(PCT UNION)

MEETING OF INTERNATIONAL AUTHORITIES UNDER THE PCT

Fifth Session
Geneva, November 28 to December 1, 1994

THE TRILATERAL PATENT-EJU-SEQUENCE DATABASE:
POSSIBILITIES TO USE THAT DATABASE BY
DESIGNATED AND/OR ELECTED OFFICES

prepared by the European Patent Office

The Trilateral Patent–EJU–Sequence Database : possibilities to use that Database by designated and/or Elected Offices

---

The present document is intended to inform the representatives of the international authorities not being parties to the EPO–JPO-USPTO Cooperation of the ongoing activities relating to the Trilateral Patent Sequence Database.

I. The Trilateral Patent Sequence Database

1. In the framework of the Trilateral Cooperation, the three Offices have decided to capture all nucleotide and amino acid sequences from all published patent documents. In order to capture these sequence data, each Office is collaborating with a contractor. The JPO is collaborating with JAPIO, the USPTO with the NCBI (National Center for Biotechnology Information) and the EPO with the EMBL Data Library in Heidelberg, Germany (to be changed in the European Bioinformatics Institute, Hinxton, United Kingdon in September 1994).

2. The data to be captured comprise two parts, a backfile and a frontfile. The backfile consists of sequence-containing patent documents since 1960 for which the sequences are not available in electronic form. The frontfile consists of sequence-containing documents (published after August 1, 1993) for which the sequence information has been provided by the applicant in electronic form.

3. For the backf ile the sequences are selected by annotators, entered twice by data entry staff, merged with the patent document frontfile information and added to the database. For the frontpage information (applicant, publication date etc.), the sequences are, after publication of the patent document, merged with the frontpage information of the patent document and added to the database. It is estimated that the EPO part of the backfile will be completed by the end of this year. The capture of the frontfile data will start shortly.

4. Each office is capturing the sequences from those patent documents for which that office has acted as priority country. For first filings in the EPO Contracting States, the EPO captures, in principle, the second EP and PCT applications and for those first filings not leading to a EP or PCT application, it captures on the basis of the national applications which are watched for that purpose. This in addition to EP-applications being first filings.

The three parts of the database so created are exchanged by the three offices and an entire database is generated which is made available to all interested parties at marginal costs.

5. The database will be made available via the existing media provided by, amongst others, the EMBL Data Library and will therefore be searchable with the commonly used search tools.

II. Possibility to use the Trilateral Patent Sequence Database by the Designated, and/or Elected Offices

6. In order to avoid that Designated and/or Elected Offices have to ask the applicants for a sequence listing in computer-readable form it is proposed that these offices make use of the Trilateral Patent Sequence Database in case they would like to access the sequences from a given sequence listing. The following aspects should be considered:

   a) Do the Elected Offices and/or the Designated Offices need access to the computer-readable form of the sequence listing as filed or do they need access to the searchable sequences of the sequence listing?

   b) How are the Elected and/or Designated Offices going to access the database?

7. In order to have access to the Trilateral Patent Sequence Database, the designated/elected Office would need to have either an on-line connection to an appropriate host (e.g. European Patent Office) or a dedicated computer system equipped with the necessary software and databases .

8. If a designated or elected Office wishes to carry out a supplementary search (e.g. the EPO, where the international search has been carried out by the USPTO or the JPO) , that designated or elected Office may access the database as described under 7 on the basis of the WO publication number and the hardcopy of the sequence listing.

   The proposed procedure implies that the sequence from all WO documents are present in the database.

9. An example of a database entry created from a WO document is included (see Annex).

10. The question of costs for accessibility of the designated/elected Offices to the Trilateral Patent Sequence Listing Database is not addressed in the present document.

[Annex follows]

ANNEX

EXTRACT FROM A DATABASE ENTRY CREATED FROM A WIPO DOCUMENT

```
AC    A00142;
XX
DT    11-FEB-1993 (Rel. 34, Created)
DT    11-FEB-1993 (Rel. 34, Last updated, Version 1)
XX
DE    H. sapiens LAG–2 gene encoding lymphokine LAG–2
XX
KW    LAG–2 gene;  lymphokine.
XX
OS    Homo sapiens (human)
OC    Eukaryota;  Animalia;  Metazoa;  Chordata;  Vertebrata;  Mammalia;
OC    Theria;  Eutheria;  Primates;  Haplorhini;  Catarrhini;  Hominidae.
XX
PN    WO9003394–A/1
PD    05-APR-1990
PF    26-SEP-1989 WO89FR00491
PR    26-SEP-1988 FR880012538
PA    Roussel UCLAF.
PI    Hercend T.;
PT    "NEW LIMPHOKINES, DMA SEQUENCES CODING FOR SAID LIMPHOKINES
PT    AND PHARMACEUTICAL COMPOSITIONS CONTAINING SAID LIMPHOKINES";
PC    ;
XX
FH    Key                 Location/Qualifiers
FH
FT    source              1..634
FT
FT                        /organism="Homo sapiens"
FT    CDS                 25..462
FT                        /gene="LAG–2"
FT                        /product= "lympholine LAG-2"
XX
SQ    Sequence 634 BP;  131 A;  203 C;  169 G;  131 T;  0 other;
      cggcatctca  gcggctgccc  caccatggct  acctgggccc  tcctgctcct  tgcagccatg   60
      ctcctgggca  acccaggtct  ggtcttctct  cgtetgagcc  ctgagtacta  cgacctggca  120
      agagcccacc  tgcgtgatga  ggagaaatcc  tgcccgtgcc  tggcccagga  gggcccccag  180
      ggtgacctgt  tgaccaaaac  acaggagctg  ggccgtgact  acaggacctg  tctgacgata  240
      gtccaaaaac  tgaagaagat  ggtggataag  cccacccaga  gaagtgtttc  caatgctgcg  300
      acccgggtgt  gtaggacggg  gaggtcacga  tggcgcgacg  tctgcagaaa  tttcatgagg  340
      aggtatcagt  ctagagttat  ccagggcctc  gtggccggag  aaactgccca  gcagatctgt  420
      gaggacctca  ggttgtgtat  accttctaca  ggtccctct  gagccctctc  accttgtcct  480
      gtggaagaag  cacaggctcc  tgtcctcaga  tcccgggaac  gtcagcaacc  tctgccggct  540
      cctcgcttcc  tcgatccaga  atccactctc  cagtctccct  cccctgactc  cctctgctgt  600
      cctcccctct  caggagaata  aagtgtcaag  caag                              634
//
ID    A00144          standard;  DNA;  PRI;  705 BP.
XX    A00144;
AC
XX
DT    11-FEB-1993 (Rel. 34, Created)
DT    11-FEB-1993 (Rel. 34, Last updated, Version 1)
XX
DE    H. sapiens LAG-2 gene promoter region
XX
KW
XX
OS    Homo sapiens (human)
OC    Eukaryota;  Animalia;  Metazoa;  Chordata;  Vertebrata;  Mammalia;
OC    Theria;  Eutheria;  Primates;  Haplorhini;  Catarrhini;  Eominidae.
XX
PN    WO9003394-A/3
```

[End of Annex and of document]