Resources for Intellectual Property
Sequence searching on STN®

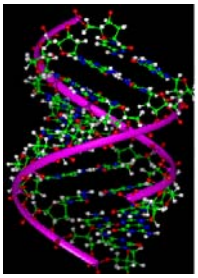
Robert Austin – FIZ Karlsruhe

Agenda

- Sequences in patent publications
- STN patent sequence databases
- Comparison to web based resources
- Results of a comparative search example
- Summary and resources

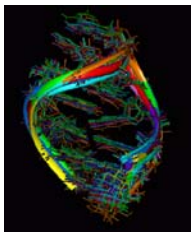
A completely new subject for many….

3



DNA

PROTEIN



RNA

PNA

Bluff Your Way in Genetics!!
http://www.stn-international.com/training_center/bioseq/bluff.pdf

STN FIZ Karlsruhe

Sequences are often described in patents, and usually via a sequence identity number

4

L1 ANSWER 1 OF 1 USGENE COPYRIGHT 2008 SEQUENCEBASE CORP on STN

CLM US6831165 B1: What is claimed is:
1. An isolated nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, or the full complement thereof.

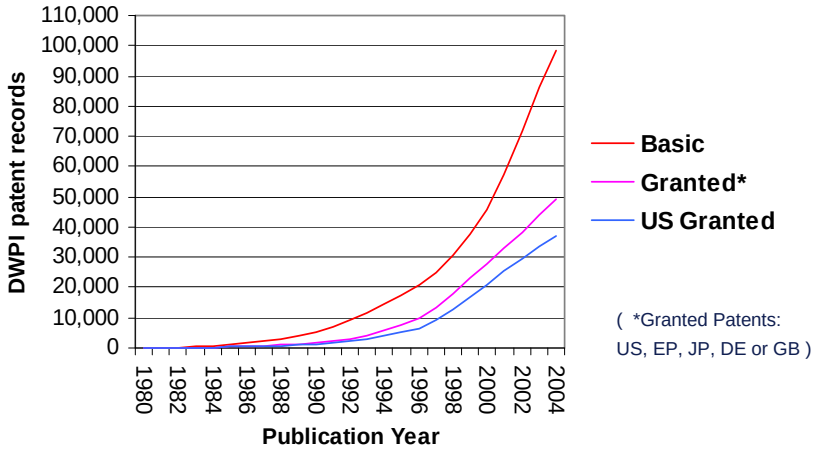
2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.

3. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to the complement of a nucleic acid molecule consisting of SEQ ID NO:1 in 6.times.SSC at 45.degree. C., followed by one or more washes in 0.2.times.SSC, 0.1% SDS at 50-65.degree. C., and wherein said nucleic acid molecule encodes a sulfate adenylate transferase subunit 2.
. . . .

STN FIZ Karlsruhe

Cumulative growth of patent publications with sequences

5



(*Granted Patents:
US, EP, JP, DE or GB)

STN

FIZ Karlsruhe

Sequences referred to by SEQ ID NO typically appear in a Sequence Listing

6

```
<210> SEQ ID NO 1
<211> LENGTH: 314
<212> TYPE: DNA
<213> ORGANISM: Corynebacterium glutamic
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(291)
<223> OTHER INFORMATION: RXA02548
<400> SEQUENCE: 1
      cca ccg atc tac ttc tcc cac gac cgc gaa gtt ttc gag cgc gac ggc   48
      Pro Pro Ile Tyr Phe Ser His Asp Arg Glu Val Phe Glu Arg Asp Gly
        1               5                  10                  15
      atg tgg ctg acc gca ggc gag tgg ggt gga cca aag aag ggc gag gag   96
      Met Trp Leu Thr Ala Gly Glu Trp Gly Gly Pro Lys Lys Gly Glu Glu
                   20                  25                  30
      atc gtc acc aag act gtc cgc tac cgc acc gtc ggc gat atg tcc tgc   144
      Ile Val Thr
               35
      acc ggt gct
. . . .
```

STN databases, DGENE, USGENE, PCTGEN and REGISTRY take patent sequence data like this and provide it in a suitably searchable format.

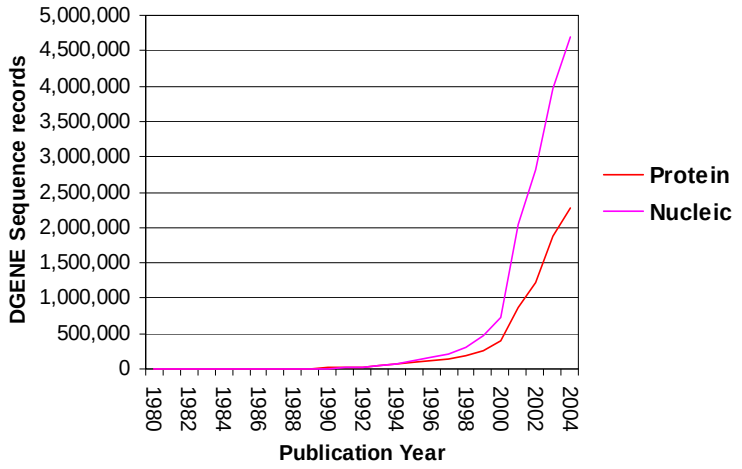
For further details about WIPO ST.25 format, visit:
http://www.wipo.int/scit/en/standards/pdf/03-25-01.pdf

STN

FIZ Karlsruhe

Cumulative growth of sequences published in patent literature

7



STN FIZ Karlsruhe

Agenda

8

- Sequences in patent publications
- STN patent sequence databases
- Comparison to web based resources
- Results of a comparative search example
- Summary and resources

STN FIZ Karlsruhe

STN sequence searchable databases

9

- DGENE
  - Thomson Reuters GENESEQ™
  - Value-added patent sequence data from around the globe
- USGENE®
  - The USPTO Genetic Sequence Database
  - A new and unique access point to USPTO sequence data
- PCTGEN
  - WIPO/PCT Patent Application Biosequences
  - The complete collection of e-published sequences from WIPO
- CAS REGISTRY℠
  - Chemical Abstracts Service (CAS) REGISTRY File
  - Worldwide value-added patent and non-patent sequence data

STN FIZ Karlsruhe

Thomson Reuters GENESEQ (DGENE)

10

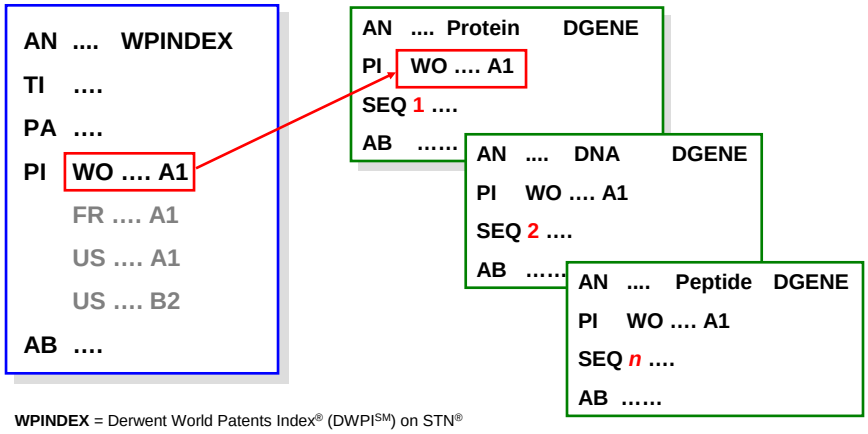
- Largest value-added patent sequence database
- Used routinely by all major patent offices*
- Sequences from the basic patents of the 40 authorities of the *Derwent World Patents Index*®
- Bibliography, enhanced title, abstract, indexing and patent location provided for each sequence
- Patent Family and Legal Status display
- Updated every two weeks
- 1981 - present

* See page 11: http://www.trilateral.net/projects/biotechnology/search_guidebook_vers_1.pdf

STN FIZ Karlsruhe

Relationship between DWPI patent family and DGENE sequence database

11



WPINDEX = Derwent World Patents Index® (DWPI℠) on STN®

DGENE = GENESEQ™ on STN®

STN FIZ Karlsruhe

What exactly is the "value-add" in DGENE?

12

- DWPI patent title, concise sequence description, abstract and keyword indexing *per sequence*
  - Illuminate the context of *each sequence* within the invention
  - Superior text based refinement of sequence searches
  - Efficient scanning and review of search results for relevance
- Feature tables for sequence modifications/annotations
  - Extensive detailed annotations are provided by Indexers
- Patent sequence location (claim, example, etc)
  - Assigned manually by Thomson Indexers
  - Ability to filter searches to those described in the claims
- Sequences intellectually derived by Indexers
  - Provides unique sequence hits not disclosed in formal listings

STN FIZ Karlsruhe

Some editorial insights regarding WIPO/PCT sequences indexed in DGENE

13

- On average 120 WIPO/PCT basic patents have sequences indexed into DGENE each week
- Of these, about 15-20 may have electronic listings available – the rest are keyed manually
  - Sequences are independently double-keyed with a guaranteed accuracy of 99.995% (1 in 20,000)
- About 15% of PCTs with electronic listings have extra sequences indexed from the specification
- Typically 1 or 2 documents per week will also have intellectually derived sequences indexed, based upon the wording of the patent claims

Source: Colin Williams, GENESEQ Editorial & Content Manager, Thomson Reuters (12/2006)

STN FIZ Karlsruhe

Derived sequences are intellectually created by Indexers from wording in the patent text

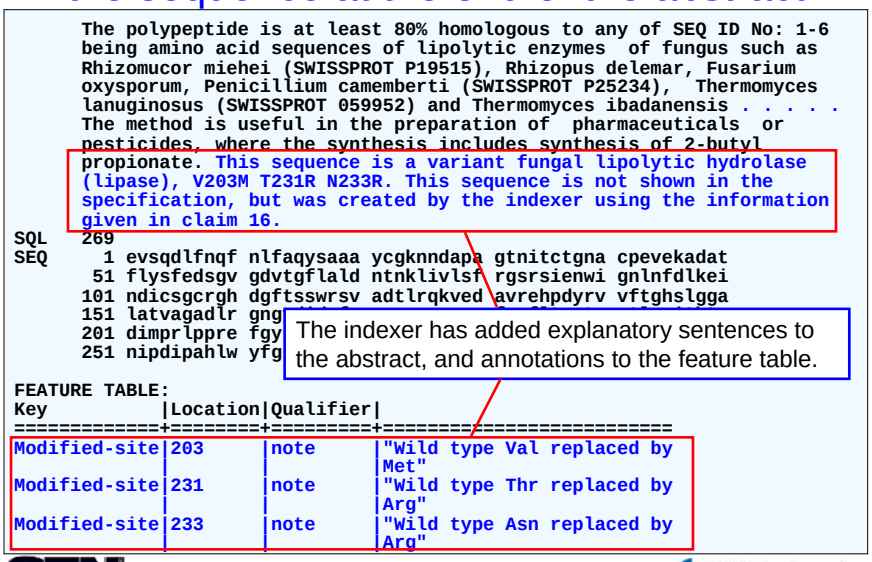
14

```
AN    AEJ92622  protein       DGENE
TI    Hydrolyzing/synthesizing carboxylic acid ester/amide from
      chiral/prochiral reactants for preparing e.g. pharmaceuticals,
      comprises  contacting reactants with a polypeptide having hydrolytic
      activity.
IN    Svendsen A; Vind J; Brask J; De Maria L
PA    (NOVO)      NOVOZYMES AS.
PI    WO 2006084470   A2 20060817               17
AI    WO 2006-DK76          20060210
PRAI  EP 2005-388012        20050210
PSL   Claim 16
DED   19 OCT 2006  (first entry)
LA    English
OS    2006-560037 [57]
DESC  Variant fungal lipolytic hydrolase #2.
KW    hydrolysis; lipase; pharmaceutical; pesticide; enzyme; mutein.
ORGN  Thermomyces lanuginosus. Synthetic.
AB    The new invention relates to a enzymatic method of hydrolyzing or
      synthesizing carboxylic acid ester or amide from chiral or prochiral
      reactants, by providing reactants for hydrolysis or synthesis, and
      contacting the reactants with a polypeptide which has hydrolytic
      activity on ester or amide, and a sequence 50% homologous to
      Thermomyces  lanuginosus lipase. Also described is a polypeptide,
      which has hydrolase  activity on an ester or amide substrate, and has
      an amino acid sequence  that has at least 80% identity to SEQ ID No:
      5 and compared to SEQ ID No: 5 comprises a substitution corresponding
      to I90Q, N92TD, F95Y, F113Y, I202M, V203GM, L269T and 270F. . . . .
```

In this example, the indexer has intellectually derived this sequence from the wild-type lipolytic hydrolase.

STN FIZ Karlsruhe

Indexers explain exactly how they derived the sequence at the end of the abstract



STN FIZ Karlsruhe

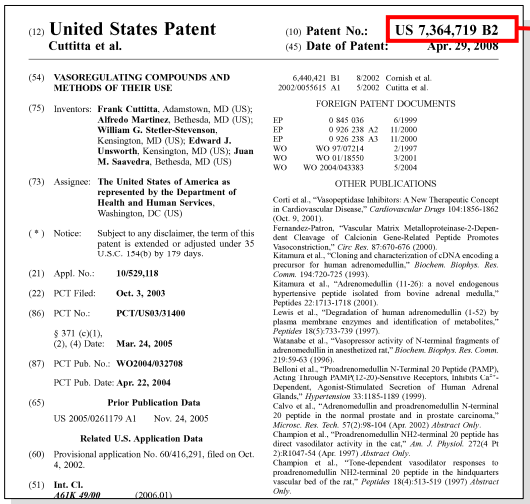
USGENE is the USPTO Genetic Sequence Database

- Sequences from all relevant USPTO published patent applications and granted (issued) patents
- Assignee and full inventor names; publication, application and parent case PCT numbers and dates; original publication **title**, **abstract**, and **claims**
- Organism name, sequence length, Molecule Type, SEQ ID, and feature tables for features/annotations
- Produced by the SequenceBase Corporation
- Updated weekly – within **3 days** of publication
- 1982 – present

STN FIZ Karlsruhe

An individual publication is represented by one or more USGENE sequence records

17



(12) **United States Patent**
Cuttitta et al.

(10) **Patent No.: US 7,364,719 B2**
(45) **Date of Patent: Apr. 29, 2008**

(54) **VASOREGULATING COMPOUNDS AND METHODS OF THEIR USE**

(75) Inventors: **Frank Cuttitta**, Adamstown, MD (US); **Alfredo Martinez**, Bethesda, MD (US); **William G. Stetler-Stevenson**, Kensington, MD (US); **Edward J. Unsworth**, Kensington, MD (US); **Juan M. Saavedra**, Bethesda, MD (US)

(73) Assignee: **The United States of America as represented by the Department of Health and Human Services**, Washington, DC (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 179 days.

(21) Appl. No.: **10/529,118**

(22) PCT Filed: **Oct. 3, 2003**

(86) PCT No.: **PCT/US03/31400**

§ 371 (c)(1),
(2), (4) Date: **Mar. 24, 2005**

(87) PCT Pub. No.: **WO2004/032708**

PCT Pub. Date: **Apr. 22, 2004**

(65) **Prior Publication Data**

US 2005/0261179 A1 Nov. 24, 2005

Related U.S. Application Data

(60) Provisional application No. 60/416,291, filed on Oct. 4, 2002.

(51) **Int. Cl.**
A61K 49/00 (2006.01)

6,440,421 B1 8/2002 Cornish et al.
2002/0055615 A1 5/2002 Cutitta et al.

FOREIGN PATENT DOCUMENTS

EP 0 845 036 6/1999
EP 0 926 238 A2 11/2000
EP 0 926 238 A3 11/2000
WO WO 97/07214 2/1997
WO WO 01/18550 3/2001
WO WO 2004/043383 5/2004

OTHER PUBLICATIONS

Corti et al., "Vasopeptidase Inhibitors: A New Therapeutic Concept in Cardiovascular Disease," *Cardiovascular Drugs* 104:1856-1862 (Oct. 9, 2001).
Fernandez-Patron, "Vascular Matrix Metalloproteinase-2-Dependent Cleavage of Calcitonin Gene-Related Peptide Promotes Vasoconstriction," *Circ Res.* 87:670-676 (2000).
Kitamura et al., "Cloning and characterization of cDNA encoding a precursor for human adrenomedullin," *Biochem. Biophys. Res. Comm.* 194:720-725 (1993).
Kitamura et al., "Adrenomedullin (11-26): a novel endogenous hypertensive peptide isolated from bovine adrenal medulla," Peptides 22:1713-1718 (2001).
Lewis et al., "Degradation of human adrenomedullin (1-52) by plasma membrane enzymes and identification of metabolites," *Peptides* 18(5):733-739 (1997).
Watanabe et al., "Vasopressor activity of N-terminal fragments of adrenomedullin in anesthetized rat," *Biochem. Biophys. Res. Comm.* 219:59-63 (1996).
Belloni et al., "Proadrenomedullin N-Terminal 20 Peptide (PAMP), Acting Through PAMP(12-20)-Sensitive Receptors, Inhibits Ca^{2+}-Dependent, Agonist-Stimulated Secretion of Human Adrenal Glands," *Hypertension* 33:1185-1189 (1999).
Calvo et al., "Adrenomedullin and proadrenomedullin N-terminal 20 peptide in the normal prostate and in prostate carcinoma," *Microsc. Res. Tech.* 57(2):98-104 (Apr. 2002) *Abstract Only*.
Champion et al., "Proadrenomedullin NH2-terminal 20 peptide has direct vasodilator activity in the cat," *Am. J. Physiol.* 272(4 Pt 2):R1047-54 (Apr. 1997) *Abstract Only*.
Champion et al., "Tone-dependent vasodilator responses to proadrenomedullin NH2-terminal 20 peptide in the hindquarters vascular bed of the rat," *Peptides* 18(4):513-519 (1997) *Abstract Only*.

AN .... Protein USGENE
PI US B2
SEQ 1 ....

AN .... DNA USGENE
PI US B2
SEQ 2 ....

AN .... cDNA USGENE
PI US B2
SEQ *n* ....

STN FIZ Karlsruhe

USGENE consolidates unique USPTO sequence data from different sources

18

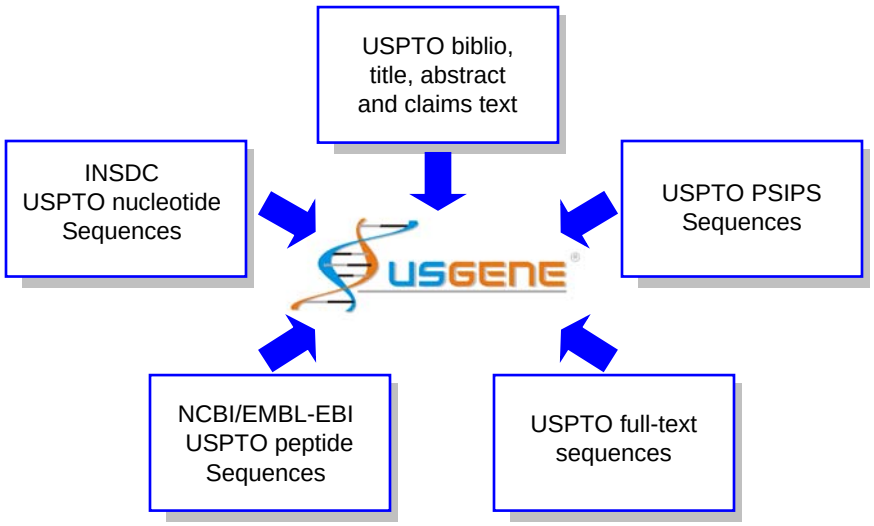
- USPTO Publication Site for Issued and Published Sequences (PSIPS)
  - The official mega-publication download site, 2001-date
- International Nucleotide Sequence Database Collaboration (INSDC) (NCBI/EMBL/DDBJ, Genbank)
  - U.S. granted patent nucleotide sequences, 1982-date
- USPTO Protein Database (NCBI/EMBL)
  - U.S. granted patent protein/peptide sequences, 1982-date
- USPTO Published Applications and Patents Full-Text
  - Filling in omissions, coverage gaps and to enhance timeliness

The USGENE Sequence Source (/SSO) field indicates which source any given USGENE sequence record was derived from.

STN FIZ Karlsruhe

USGENE combines these sequences with bibliographic data and claims text

19



STN FIZ Karlsruhe

USGENE is an essential additional tool for tackling business critical searches

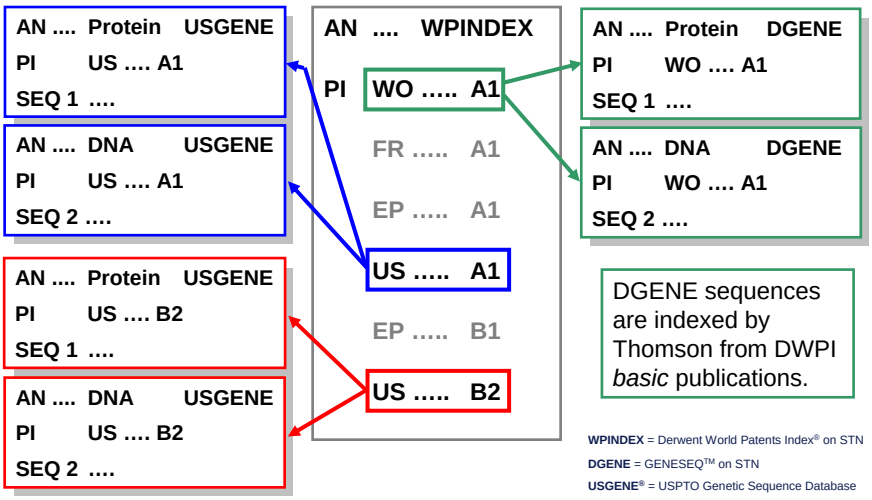
20

- DGENE provides curated and indexed patent sequence data from the DWPI *basic* publication
  - 61% of *basics* are WIPO/PCT published applications
  - Updated biweekly, typically 65 days from publication
- USGENE provides all available sequence data from the USPTO as a single merged resource
  - Both **U.S. patents** and **U.S. published applications**
  - Updated weekly, within **3 days** of USPTO publication
- Sequence listing variation often occurs between PCT and U.S. granted patent publication stages
  - Especially important, e.g. for freedom-to-operate

STN FIZ Karlsruhe

USGENE and DGENE capture sequence data from different patent family members

21



DGENE sequences are indexed by Thomson from DWPI *basic* publications.

WPINDEX = Derwent World Patents Index® on STN
DGENE = GENESEQ™ on STN
USGENE® = USPTO Genetic Sequence Database

STN

FIZ Karlsruhe

Sequence listing variation often occurs between PCT and U.S. granted patent stage

22

```
L1   ANSWER 1 OF 1  WPINDEX COPYRIGHT 2008       THOMSON REUTERS on STN
AN   1994-358278 [44]   WPINDEX
TI   New polynucleotide(s) specific for hepatitis C virus types 4, 5 and 6  -
     and related antigenic peptide(s) and antibodies, useful in vaccines,
     diagnosis, HCV typing and treatment
DC   B04; D16; S03
IN   PIKE I H; SIMMONDS P; YAP P L
PA   (COMM-N) COMMON SERVICES AGENCY; (MURE-N) MUREX DIAGNOSTICS INT INC; . . .
PI   WO 9425602      A1 19941110 (199444)* EN  70[5]
     AU 9465797      A  1994
     FI 9505224      A  1995
     EP 698101       A1 1996
     JP 09500009     W  1997
     AU 695259       B  1998
     EP 698101       B1 2004
     DE 69434116     E  2004
     US 20050032047  A1 20050210 (200512)  EN
     US 6881821      B2 20050419 (200527)  EN
     . . . . .
ADT  WO 9425602 A1 WO 1994-GB957 19940505 . . . .
PRAI GB 1994-263 19940107
     GB 1993-9237 19930505
```

In this example the patent family has:

- 9 sequences from WO9425602 in DGENE
- 50 sequences from US20050032047 in USGENE
- 58 sequences from US6881821 in USGENE

STN

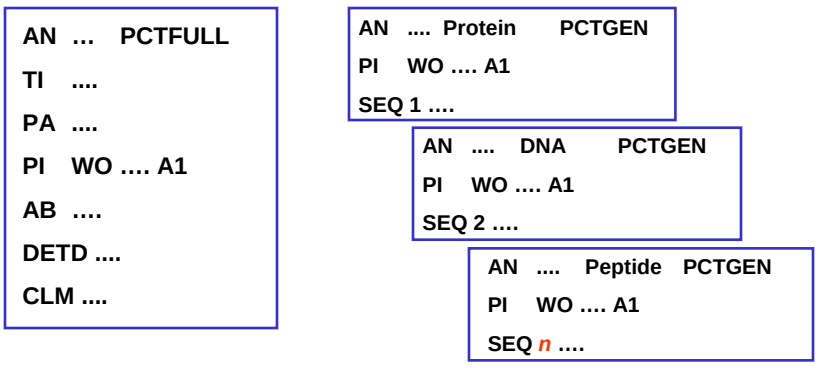
FIZ Karlsruhe

PCTGEN is the World Patent Application Biosequences database on STN

- Produced by FIZ Karlsruhe and WIPO
- Sequences submitted & published electronically as a formal part of PCT patent applications
- Publication number and date, patent applicant name(s) and the original publication title are provided for each sequence
- Sequence length, SEQ ID, organism name and molecule type are included for each sequence
- Updated weekly – within 24 hours of publication
- August 2001 – present

STN FIZ Karlsruhe

Relationship between PCTFULL and PCTGEN databases



PCTFULL = WIPO/PCT patent applications full-text

PCTGEN = WIPO/PCT patent application biosequences

STN FIZ Karlsruhe

Chemical Abstracts Service (CAS) REGISTRY database

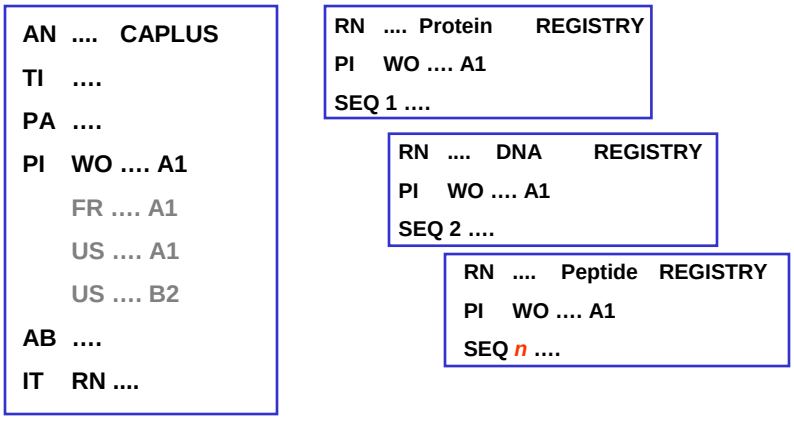
25

- Sequences from >3000 life science journals
- Sequences from the basic patents of the 50 patent authorities of the CAplus file on STN
- Patent number, location and standardized nomenclature are provided for each sequence
- Updated daily
- 1907 - present

STN FIZ Karlsruhe

Relationship between CAplus patent family and CAS Registry databases

26



STN FIZ Karlsruhe

Agenda

- Sequences in patent publications
- STN patent sequence databases
- Comparison to web based resources
- Results of a comparative search example
- Summary and resources

STN FIZ Karlsruhe

There are three main web resources which provide searchable patent sequence data

- National Center for Biotechnology Information (NCBI) of the U.S. National Library of Medicine
  - www.ncbi.nlm.nih.gov
- European Bioinformatics Institute* (EBI)
  - www.ebi.ac.uk
- DNA DataBank of Japan (DDBJ)
  - www.ddbj.nig.ac.jp
- The USPTO, EPO and JPO rely on the NCBI, EBI and DDBJ, respectively, to provide an interface for searching patent sequence data

(* The EBI is the U.K. based outstation of the European Molecular Biology Laboratory – EMBL)

STN FIZ Karlsruhe

STN databases have more patent sequences than web resources

29

Database	Peptide	Nucleotide
REGISTRY	2.50	9.39
DGENE	2.71	5.47
PCTGEN	0.54	4.01
NCBI	0.57	3.43

(Statistics in millions: October 8th, 2006)

STN FIZ Karlsruhe

Why does STN have more patent sequences than the NCBI?

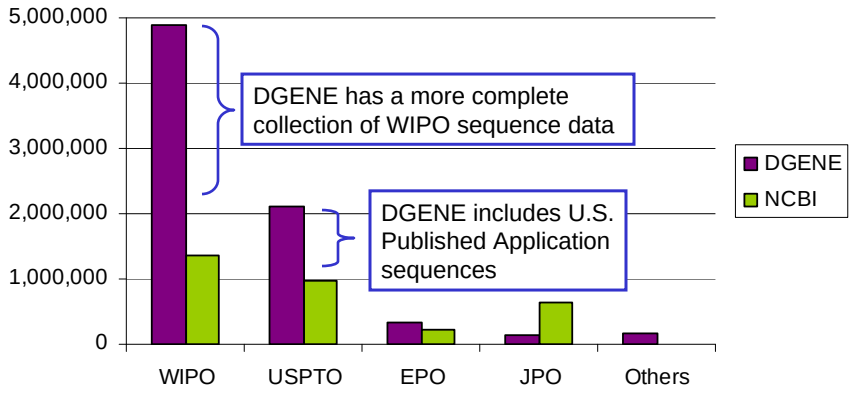
30

1. NCBI has a substantially incomplete collection of WIPO/PCT sequence data
2. NCBI does not cover any USPTO Published Application sequence data
3. DGENE and REGISTRY cover many more patent authorities than NCBI
4. DGENE, USGENE, PCTGEN and REGISTRY are typically much more timely than NCBI
5. NCBI does not cover all of the USPTO granted patent sequence data available in USGENE

STN FIZ Karlsruhe

Total sequence count comparison by patent authority in DGENE versus NCBI

31



(Statistics: January 21st , 2006)

STN FIZ Karlsruhe

In addition, NCBI/EMBL/DDBJ patent records only have minimal bibliographic and text data

32

General Information	
Accession #	AAA00521
SRS Entry ID	USPO_PRT:AAA00521
Molecule Type	PRT
Sequence Length	40
Entry Data Class	STANDARD
Sequence Version	AAA00521.1
Creation Date	21-MAY-1993
UniParc	UPI0000035113

Reminder: NCBI/EMBL/DDBJ cover sequences from U.S. granted patents – sequences from U.S. published applications are **not** covered (see slide 18).

Description	
Description	Sequence 1 from Patent US 4563352.
Organism	Unknown

References

1. Rivier; J.E.F.; Spiess; J. and Vale; W.W. Jr.;
Human pancreatic GRF
Patent number US4563352-A/1 07-JAN-1986; The Salk Institute For Biological Studies; San Diego, CA
Position 1-40

Features

Key	Location	Qualifier	Value
source	1..40		

Sequence

Characteristics: **Length:** 40 AA

Sequence:
```
>uspo_prt|AAA00521|AAA00521 Sequence 1 from Patent US 4563352.
YADAIFTNSYRKVLGQLSARKLLQDIMSRQQGESNQERGA
```

STN FIZ Karlsruhe

Agenda

- Sequences in patent publications
- STN patent sequence databases
- Comparison to web based resources
- Results of a comparative search example
- Summary and resources

STN FIZ Karlsruhe

A simple BLAST example shows the importance of using STN databases

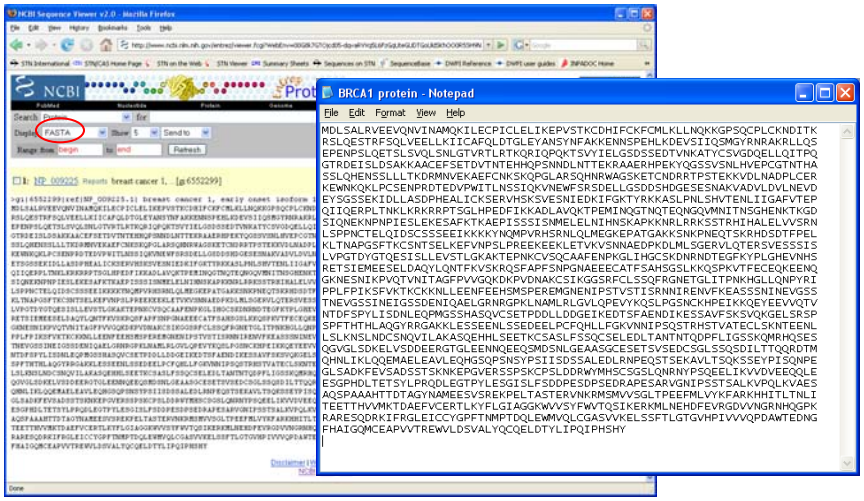
Search Question:
Find patent all references to Breast Cancer 1, early onset isoform 1 (NCBI: NP_009225), or other very similar proteins (i.e. >80% match)

(Search conducted on June 12th, 2008)

STN FIZ Karlsruhe

Homo sapiens Breast Cancer 1, early onset isoform 1 protein (NCBI: NP_009225)

35



STN FIZ Karlsruhe

The best answer from DGENE…

36

```
=> D BIB AB PSL ORGN SCORE ALIGN
L2    ANSWER 1 OF 114  DGENE  COPYRIGHT 2008 THOMSO
AN    AFU71404  protein       DGENE
TI    Diagnosing breast cancer or a susceptibility to breast cancer
      comprises  detecting BRCA2 999del5 and BARD1 Cys557Ser.
IN    Stacey S N; Sulem P; Thorsteinsdottir U; Kong A
PA    (STAC-I)    STACEY S N. . . . .
PI    US 2007092900   A1   20070426
AI    US 2006-515368       20060831
PRAI  US 2005-730703P      20051026
OS    2007-556498 [54]
CR    N-PSDB: AFU71403
DESC  Human breast cancer 1 (BRCA1) protein.
PSL   Disclosure; SEQ ID NO 2
AB    The present invention relates to a methods for diagnosing and
      characterizing breast cancer or a susceptibility to breast cancer in
      an individual comprising detecting breast cancer 2 (BRCA2) 999del5
      and BRCA1 associated RING domain 1 (BARD1) Cys557Ser. . . . .
ORGN  Homo sapiens.
SCORE 3722      100% of query self score 3722
BLASTALIGN
   Query  = 1863 letters
   Length = 1863
   Score  = 3722 bits (9652), Expect = 0.0
   Identities = 1863/1863 (100%), Positives = 1863/1863 (100%)
  Query: 1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
               MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
  Sbjct: 1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
```

114 sequence hits were found in DGENE.

DGENE records feature extensive Thomson value added content.

STN FIZ Karlsruhe

The best answer from USGENE…


```
=> D BIB AB ECLM ORGN SCORE ALIGN

L4    ANSWER 1 OF 70  USGENE COPYRIGHT 2008 SEQUENCE
AN    20070083334.873670  Protein       USGENE
TI    Methods and systems for annotating biomolecular sequences
      (PublishedApplication)
IN    Mintz Liat (Kendall-Park, NJ); Xie Hanqing (Plainsboro, NJ); . . . .
PA    Compugen Ltd
PI    US 20070083334      A1   20070412
AI    US 2006-443428           20060531
DT    Patent
AB    A method of annotating biomolecular sequences. The method comprises
      (a) computationally clustering the biomolecular sequences according
      to a progressive homology range, to thereby generate a plurality of
      clusters each being of a predetermined homology . . . .
ECLM  US20070083334 A1: 1-98. (canceled) 99. An isolated protein, having a
      sequence according to the amino acid sequence of SEQ ID NO: 836875.
ORGN  Homo Sapiens
SCORE 3722      100% of query self score 3722
BLASTALIGN
   Query  = 1863 letters
   Length = 1863
   Score  = 3722 bits (9652), Expect = 0.0
   Identities = 1863/1863 (100%), Positives = 1863/1863 (100%)
  Query: 1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
               MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
  Sbjct: 1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
```

70 sequence hits were found in USGENE.

Note that this is a different best answer from that found in DGENE.

STN FIZ Karlsruhe

The best answer from PCTGEN…


```
=> D BIB ORGN SCORE ALIGN

L6    ANSWER 1 OF 7  PCTGEN COPYRIGHT 2008 WIPO on S
AN    2007047796.8180  PRT       PCTGEN
TI    TISSUE- AND SERUM-DERIVED GLYCOPROTEINSAND METHODS OF THEIR USE
PA    Institute for Systems Biology
      Zhang, Hui
      Aebersold, Rudolf H.
PI    WO 2007047796     20070426
RLI   US 2005-728044P 20051017
ED    20070427
DT    Patent
ORGN  Homo sapiens
SCORE 3722      100% of query self score 3722
BLASTALIGN
   Query  = 1863 letters
   Length = 1863
   Score  = 3722 bits (9652), Expect = 0.0
   Identities = 1863/1863 (100%), Positives = 1863/1863 (100%)
  Query: 1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
               MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
  Sbjct: 1    MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ
```

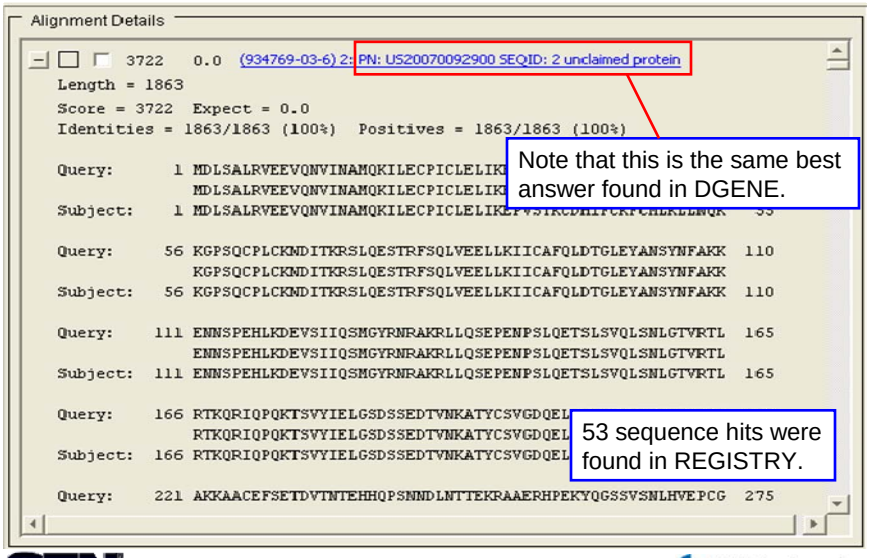
7 sequence hits were found in PCTGEN.

Note that this is a different best answer from those found in DGENE and USGENE (previous slides).

STN FIZ Karlsruhe

The best patent answer from REGISTRY…

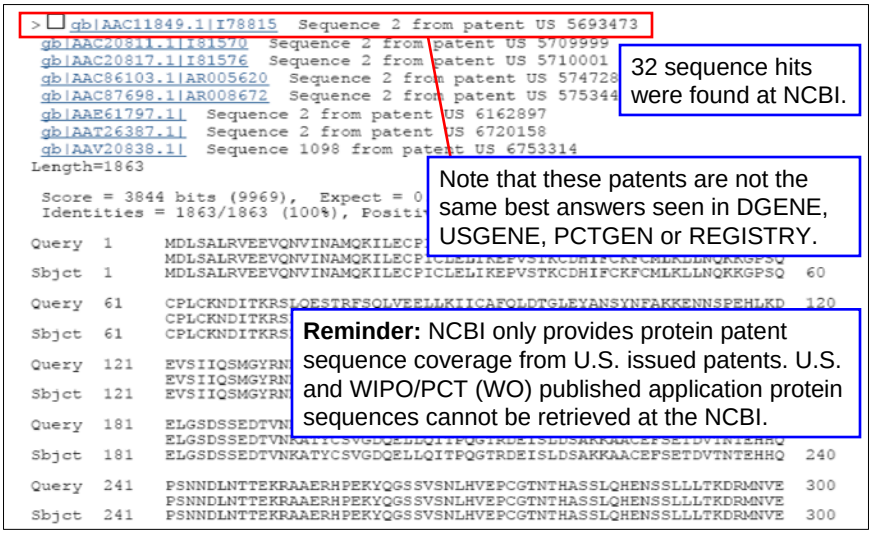
39



STN FIZ Karlsruhe

The best patent answer from NCBI…

40



STN FIZ Karlsruhe

Summary of results for Breast Cancer 1, early onset isoform 1 protein (NCBI: NP_009225)

41

	SEQs > 80%	PNs	Patent Families
DGENE	114	36	26
USGENE	70	38	22
PCTGEN	7	4	4
REGISTRY	52	30	23
NCBI*	32	21	9
TOTAL	-	-	34

(* All of the NCBI patent hits were for U.S. patents, and none of them were unique compared to USGENE.)

STN FIZ Karlsruhe

Agenda

42

- Sequences in patent publications
- STN patent sequence databases
- Comparison to web based resources
- Results of a comparative search example
- Summary and resources

STN FIZ Karlsruhe

Comparing databases…

43

- DGENE
  - The most comprehensive global patent sequence database
  - Implemented in-house at major patent offices*
- USGENE
  - Sequences from equivalent USPTO applications and patents
  - Incorporates all U.S. patent sequence data available at NCBI
- PCTGEN
  - Sequences from equivalent WIPO/PCT publications
  - The complete collection of e-published sequences from WIPO
- REGISTRY
  - Complementary value-added indexing to DGENE
  - Unique non-patent literature coverage

* See page 11: http://www.trilateral.net/projects/biotechnology/search_guidebook_vers_1.pdf

STN FIZ Karlsruhe

Database timeliness is an important factor in understanding comprehensiveness

44

	Update Frequency	Typical Timeliness	Value added
PCTGEN	Weekly	24 hours	
USGENE	Weekly	3 days	
REGISTRY	Daily	27 days	
DGENE	Biweekly	65 days	
NCBI/EMBL	Daily	1-6 months	

STN FIZ Karlsruhe

Resources for sequence searching on STN

- *Sequence Searching on STN* modular workshop
  http://www.fiz-k.com/bostonsequenceworkshop
  - Sequence Code Match (SCM) searching
  - DGENE, USGENE, PCTGEN content and searching
  - CAS REGISTRY and REGISTRY BLAST
  - Multifile searching using USGENE and DGENE
- USGENE resources, reference materials and FAQ
  http://www.sequencebase.com
- CAS REGISTRY sequence coverage and resources
  http://www.cas.org/support/stngen/stndoc/sequences.html

STN FIZ Karlsruhe

STN®

Resources for Intellectual Property
Sequence searching on STN

www.stn-international.com