

STN[®]

STN[®] sequence searchable databases

Robert Austin – FIZ Karlsruhe

Agenda

- STN sequence searchable databases
- Comparison to NCBI and EMBL-EBI
 - Focus on patent sequence coverage

STN sequence searchable databases

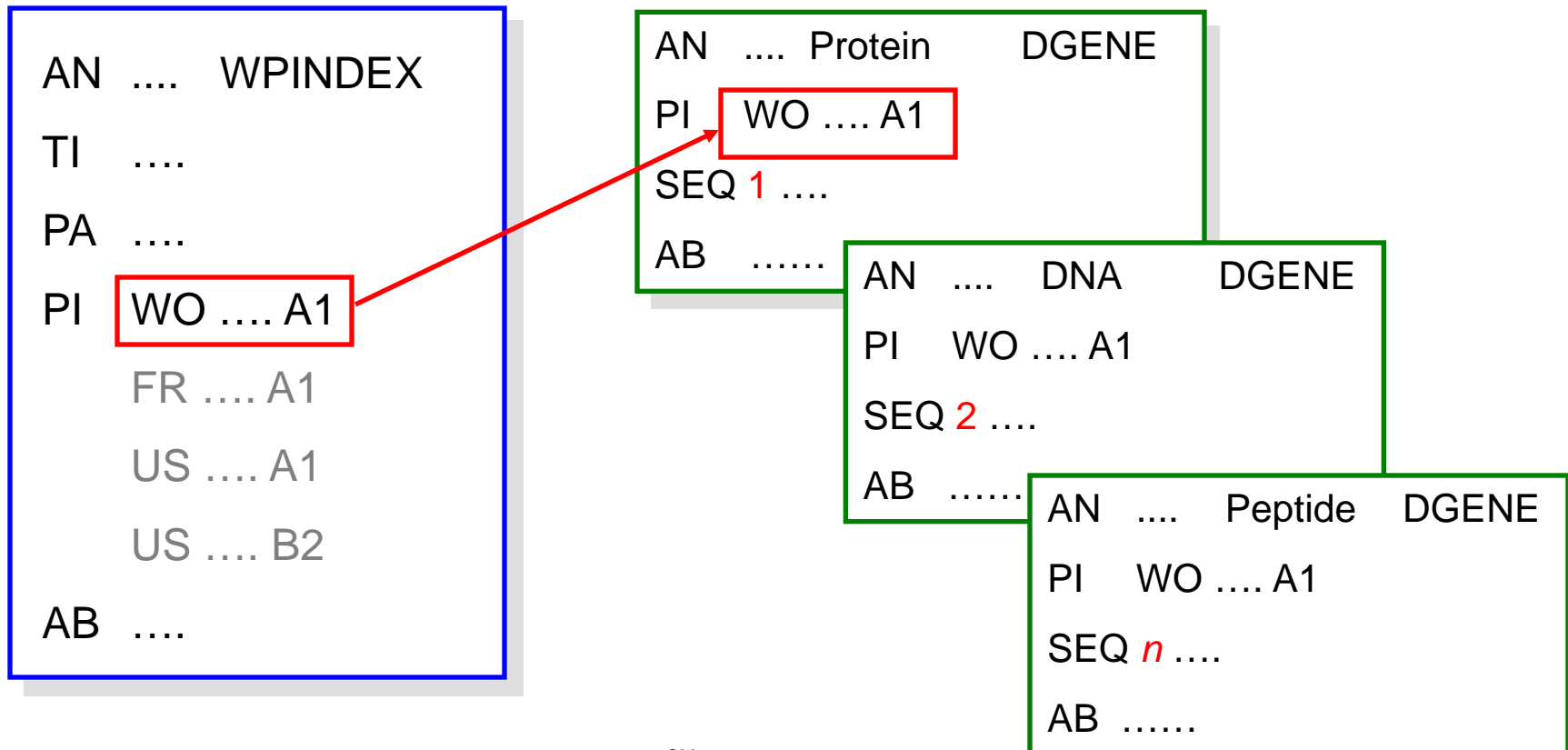
- **DGENE**
 - Thomson Reuters GENESEQ™
 - Value-added patent sequence data from around the globe
- **USGENE®**
 - The USPTO Genetic Sequence Database
 - All available sequence data from the USPTO
- **PCTGEN**
 - WIPO/PCT Patent Application Biosequences
 - All available e-published sequence data from WIPO
- **CAS REGISTRYSM**
 - Chemical Abstracts Service (CAS) REGISTRY
 - Worldwide value-added **non-patent** and **patent** sequences

Thomson Reuters GENESEQ (DGENE)

- Largest value-added patent sequence database
- Used routinely by all major patent offices*
- Sequences from the basic patents of the 48 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 10: <http://www.trilateral.net/projects/biotechnology/guide2.pdf>

Relationship between DWPI patent family and DGENE sequence database



WPINDEX = Derwent World Patents Index (DWPISM) on STN

DGENE = GENESEQ on STN

What exactly is the “value-add” in DGENE?

- DWPI patent title, concise sequence description, abstract, and keyword indexing *per sequence*
 - Context of *each sequence* illuminated 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 provided by indexers
- Patent sequence location (claim, example, etc.)
 - Assigned manually by Thomson Reuters indexers
 - Flexible filtering of searches to those described in the claims
- Sequences intellectually identified or derived by indexers
 - Unique sequence hits not disclosed in formal listings

Example: derived sequence record

L1 ANSWER 1 OF 1 DGENE COPYRIGHT 2013 THOMSON REUTERS on STN
AN AEJ92622 protein DGENE [Full-text](#)
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.

Example: derived sequence record (cont.)

The polypeptide is at least 80% homologous to any of SEQ ID No: 1-6 being amino acid sequences of lipolytic enzymes of fungus such as Rhizomucor miehei (SWISSPROT P19515), Rhizopus delemar, Fusarium oxysporum, Penicillium camemberti (SWISSPROT P25234), Thermomyces lanuginosus (SWISSPROT 059952) and Thermomyces ibadanensis The method is useful in the preparation of pharmaceuticals or pesticides, where the synthesis includes synthesis of 2-butyl propionate. This sequence is a variant fungal lipolytic hydrolase (lipase), V203M T231R N233R. This sequence is not shown in the specification, but was created by the indexer using the information given in claim 16.

```
SQL 269
SEQ 1 evsqdlfnqf nlfagysaaa ycgknndapa gtnitctgna cpevekadat
51 flysfedsgv gdvtgflald ntnklivlsf rgsrsienwi glnlnfdlkei
101 ndicsgcrgh dgftsswrsv adtlrqqkved avrehpdyrv vftghslgga
151 latvagadlr gng
201 dimprlppre fgy
251 nipdipahlw yfg
```

The indexer has added explanatory sentences to the abstract and annotations to the feature table.

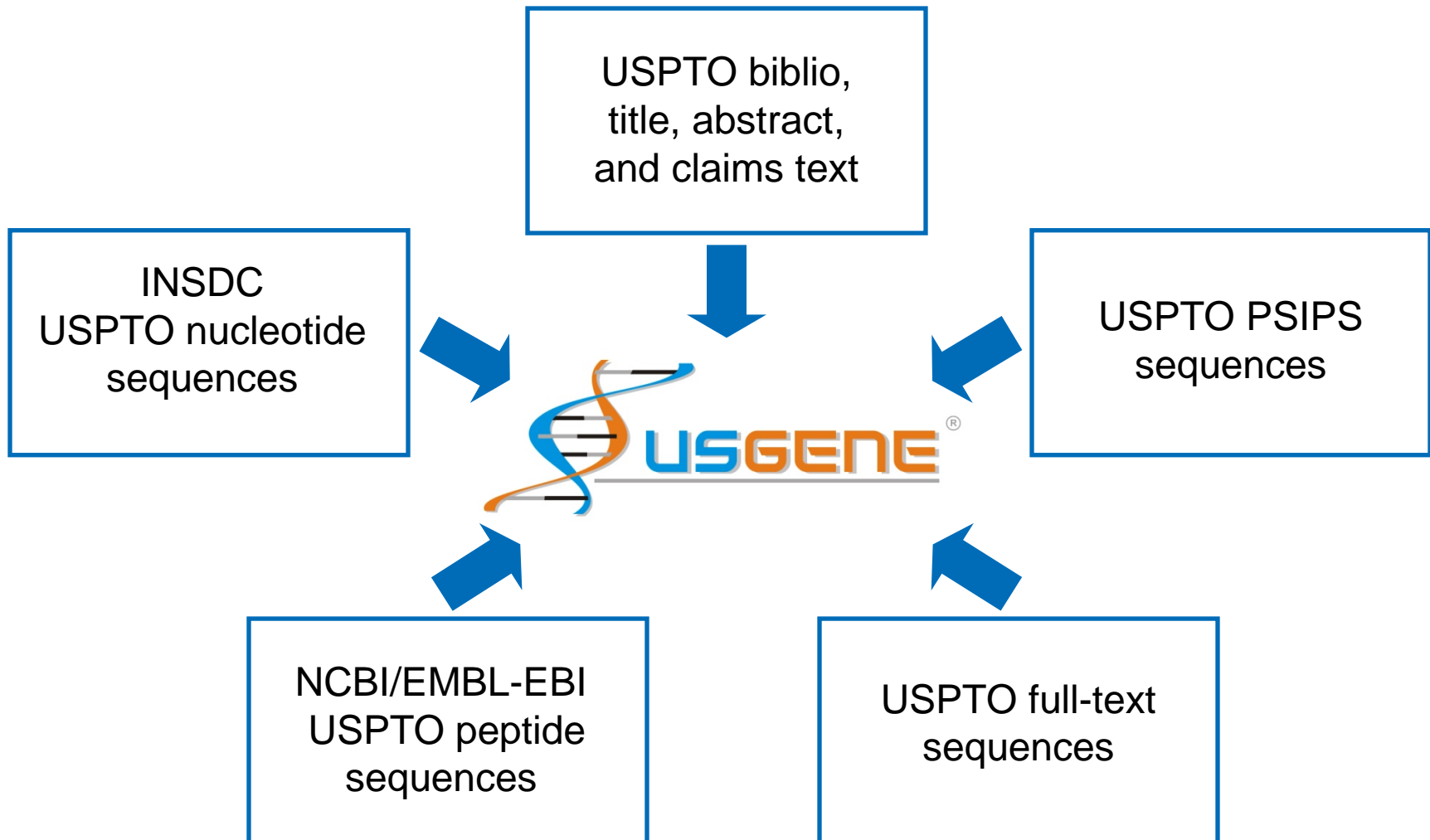
FEATURE TABLE:

Key	Location	Qualifier	
Modified-site	203	note	"Wild type Val replaced by Met"
Modified-site	231	note	"Wild type Thr replaced by Arg"
Modified-site	233	note	"Wild type Asn replaced by Arg"

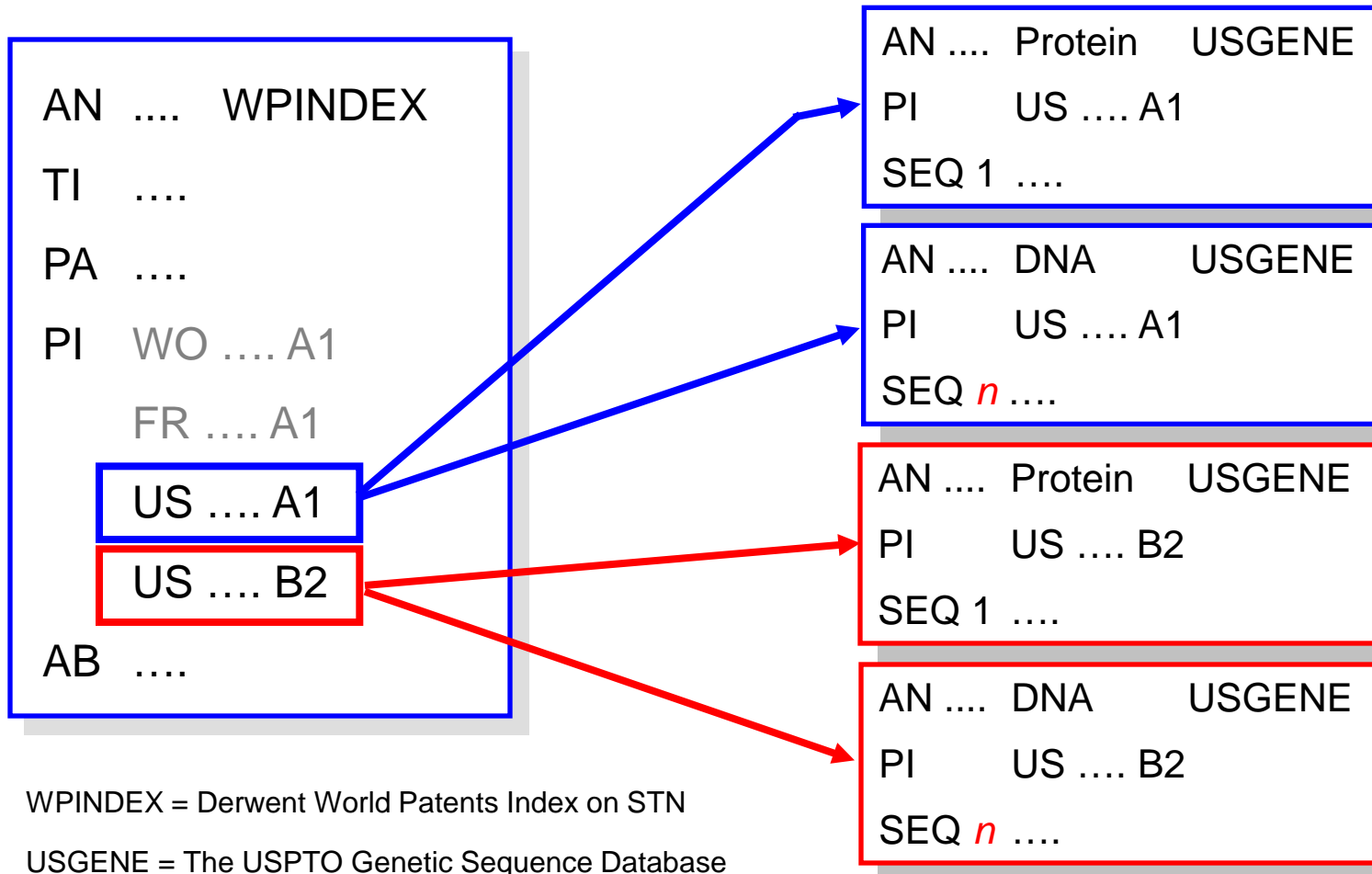
SequenceBase USGENE

- Sequences from USPTO published patent applications and granted (issued) patents
- Original publication title, abstract and claims
- Publication, application, related application and priority numbers and dates; assignees and inventors
- Organism, Sequence Length, SEQ ID NO, Molecule Type, Features, and Patent Sequence Location
- Calculated patent expiration date
- Patent Family and Legal Status display
- Updated weekly
- 1981 – present

USGENE combines sequences with bibliographic data and claims text



Relationship between DWPI patent family and USGENE sequence databases



Example: published application record

L1 ANSWER 1 OF 1 USGENE COPYRIGHT 2013 SEQUENCEBASE CORP on STN
AN **20100273154.6** DNA USGENE [Full-text](#)
TI NOVEL GLYCOSYLTRANSFERASES AND POLYNUCLEOTIDES ENCODING THE SAME
(PublishedApplication)
IN Ochiai Misa (Osaka, JP); Fukami Harukazu (Osaka, JP); Nakao Masahiro
(Osaka, JP); Noguchi Akio (Osaka, JP)
PA SUNTORY HOLDINGS LIMITED (Osaka JP)
PI US 20100273154 A1 20101028
AI US 2008-523068 20080118
RLI WO 2008-JP50618 20080118
PRAI JP 2007-10759 20070119
PSL Claim 1; SEQ ID NO 6
DESC Trichoderma viride DNA; Strain IAM 5141; sequence 6 of 35
DT Patent
AB The present invention provides novel glycosyltransferase proteins
produced by ascomycetous filamentous
belonging to the genus Trichoderma,
viride), as well as genes encoding t
proteins provided by the present inv
an enzyme protein obtained from the culture supernatant of
Trichoderma viride strain IAM5141. The novel enzymes of the present
invention allow glycosylation of flavonoid compounds to thereby
improve their water solubility.

AN 20100273154.6 is SEQ ID
NO: 6 from US20100273154.

Patent Sequence Location (PSL)
indicates if the SEQ ID NO is
referred to in the claims.

Example: published application record (cont.)

ECLM US20100273154 A1: 1. A polynucleotide comprising any one of (A) to (K) shown below: (A) a polynucleotide which consists of all or part of the nucleotide sequence shown in SEQ ID NO: 6; (B) a polynucleotide which is hybridizable under stringent conditions with a polynucleotide consisting of a nucleotide sequence complementary to the nucleotide sequence of the polynucleotide shown in (A) and which encodes a protein having glycosylation activity on a flavonoid compound; (C) a polynucleotide which consists of a nucleotide

SSO NUCLEIC; USPTO; APPLICATION
ORGN Trichoderma viride
SQL 2633
SEQ

```
1 tcatacaaag ctatttcgaa gaccaatatt ctacc
51 agagccctat ggtaatgggc cgatggggtg cttat
101 ccggggcatg cgtgtttgac atggtgaagc ttgggcagat ctgggatcgg
. . . .
2451 caaatagcct ttatagtccc gtgtatgctc gcaagtcttg tccgtcgctc
2501 ttagcctgta ttattcacat gcgtgtgtca tcgatagcct tgtgtttctc
2551 gtgtcgcacg atttgctctt gatattgacc acattccttc ggtacaatga
2601 gcaaatatat ataccgttgt ggtgggtaat gag
```

AN 20100273154.6 is displayed here in BRIEF format, which includes the Exemplary Claim (ECLM).

FEATURE TABLE:

Key	Location
	Strain IAM 5141

Example: issued patent record

L1 ANSWER 1 OF 1 USGENE COPYRIGHT 2013 SEQUENCEBASE CORP on STN
AN **8278088.6** DNA USGENE [Full-text](#)
TI Polynucleotide encoding a polypeptide having glycosylation activity
on a flavonoid (Patent)
IN Ochiai Misa (Osaka, JP); Fukami Harukazu (Osaka, JP); Nakao Masahiro
(Osaka, JP); Noguchi Akio (Osaka, JP)
PA Suntory Holdings Limited (Osaka JP)
PI US 8278088 B2 20121002
US 20100273154 A1 20101028
WO 2008088046 A 20080724
AI US 2008-523068 20080118
RLI WO 2008-JP50618 20080118
PRAI JP 2007-10759 20070119
XPD 20280118 (calculated)
PSL Claim 1; SEQ ID NO 6
DESC Trichoderma viride DNA; Strain IAM 5141; sequence 6 of 33
DT Patent
AB The present invention provides novel glycosyltransferase proteins
produced by ascomycetous filamentous
belonging to the genus Trichoderma,
viride), as well as genes encoding t
proteins provided by the present inv
an enzyme protein obtained from the
Trichoderma viride strain IAM5141. The novel enzymes of the present
invention allow glycosylation of flavonoid compounds to thereby
improve their water solubility.

AN 8278088.6 is SEQ ID NO: 6
from US8278088.

Patent Sequence Location (PSL)
indicates if the SEQ ID NO is
referred to in the claims.

Example: issued patent record (cont.)

ECLM US8278088 B2: 1. An isolated polynucleotide comprising a polynucleotide selected from the group consisting of: (A) a polynucleotide consisting of the nucleotide sequence of SEQ ID NO:6; (B) a polynucleotide consisting of the nucleotide sequence of SEQ ID NO:6, except 1 to 9 nucleotides are substituted, deleted, inserted, and/or added, wherein the polynucleotide encodes a protein having flavonoid glycosylation activity;(C) a polynucleotide

SSO NUCLEIC; USPTO; GRANTED
ORGN Trichoderma viride
SQL 2633
SEQ

```
1 tcatacaaag ctatttcgaa gaccaatatt ctac
51 agagccctat ggtaatgggc cgatggggtg ctt
101 ccggggcatg cgtgtttgac atgttgaagc ttgggcagac ctgggacgg
. . . .
2451 caaatagcct ttatagtccc gtgtatgctc gcaagtcttg tccgtcgtcg
2501 ttagcctgta ttattcacat gcgtgtgtca tcgatagcct tgtgtttctc
2551 gtgtcgcacg atttgctctt gatattgacc acattccttc ggtacaatga
2601 gcaaatatat ataccgttgt ggtgggtaat gag
```

AN 8278088.6 is displayed here in BRIEF format, which includes the Exemplary Claim (ECLM).

FEATURE TABLE:

Key	Location	
USGENE	1..2633	http://www.sequencebase.com/usgene.php?d=8278088.6
other_info	1..2633	Strain IAM 5141

WIPO/FIZ Karlsruhe PCTGEN

- Produced from sequence listings published electronically by WIPO
- Image and text based sequence listings are processed by FIZ Karlsruhe editorial
- Publication number and date, patent applicant name(s) and the original publication title are provided for each sequence
- Sequence length, SEQ ID NO, organism, and molecule type are included for each sequence
- Updated weekly – within **24 hours** of publication
- August 2001 – present

Example: sequence record from PCTGEN

L1 ANSWER 1 OF 1 PCTGEN COPYRIGHT 2013 WIPO on STN

AN **2012131673.6** RNA PCTGEN [Full-text](#)

TI CCAT-1 SILENCING NUCLEIC ACID AGENTS FOR TREATING CANCER [File created by using OCR software]

PA Hadasit Medical Research Services a
MOR Research Applications LTD.

PI WO 2012131673 **20121004**

RLI US 2011-469849P 20110331

ED **20121005**

DT Patent

ORGN Artificial

SQL 25

SEQ

1 cacauaguag cuaaugccua cucnn

AN 2012131673.6 is SEQ ID NO:
6 from WO2012131673 .

Sequences are typically added to
PCTGEN within 24 hours of
publication by WIPO.

FEATURE TABLE:

Key	Location	
		synthetic Polynucleotide
misc_feature	(24)..(25)	Deoxycytidine

CAS REGISTRY

- Sequences from >3000 life science journals and the basic patents of the 62 authorities of CAplusSM
- Patent number, location, and standardized nomenclature provided for each sequence
- Updated daily
- 1907 - present

Example: patent sequence record

L1 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2013 ACS on STN
ED Entered STN: 16 Jan 2007
RN 917531-59-0 REGISTRY
CN 20: PN: WO2006137596 SEQI
FS PROTEIN SEQUENCE
SQL 128

WIPO and other patent sequences typically enter REGISTRY within 27 days of publication – in this example only 20 days.

PATENT ANNOTATIONS (PNTE):

Sequence	Patent
Source	Reference
=====+=====	
Not Given	WO2006137596
	unclaimed
	SEQID 21

Since October 1999, CAS REGISTRY patent sequence records include the publication number, SEQ ID number, and an intellectually assigned claimed/unclaimed notation.

SEQ 1 KCDLALDPDL ARIMAHSR
51 HAARLNGFHD AGQQQREAYE DSDINSQLTE LWATLAPLYR ELHAYVRRHL
101 VQRYGPERVR PDGPMPAHLG GNMWSRAN

MF Unspecified

CI MAN

SR CA

LC STN Files: CA, CAPLUS

DT.CA Cplus document type: Patent

RL.P Roles from patents: PRP (Properties)

Text describing a patent sequence is typically provided in the Cplus indexing (next slide).

Example: corresponding CPlus record

L1 ANSWER 1 OF 1 CAPLUS COPYRIGHT © 2010 STN INTL
AN 2006:1357195 CAPLUS
DN 146:95066
TI Screening for effectors of insect dipeptidyl carboxypeptidase A as insecticides
IN Shimokawatoko, Yasutaka; Craen, Marc Van De; Nooren, Irene; Turconi, Sandra; Naudet, Yann; Nys, Guy; Debaveye, Jurgen
PA Sumitomo Chemical Company, Limited, Japan
FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 2006137596	A2	20061228	WO 2006-JP313039	20060623
	JP 2007000060	A	20070111	JP 2005-183031	20050623
PRAI	JP 2005-183031	A	20050623		

AB Methods of screening for effector of the dipeptidyl carboxypeptidase A (I) of insects for use as insecticides is described.
IC ICM A01N
CC 5-4 (Agrochemical Bioregulation) Section cross-reference(s)
IT 917531-51-2 917531-53-4 917531-55-6 917531-57-8 917531-59-0
RL: PRP (Properties)
(unclaimed protein sequence; screening for effectors of insect dipeptidyl carboxypeptidase A as insecticides)

CAS provides value-added abstracting and indexing for the CPlus basic publication.

The sequence is linked to the CPlus patent family record by its CAS Registry Number®.

917531-59-0

Agenda

- STN sequence searchable databases
- Comparison to NCBI and EMBL-EBI
 - Focus on patent sequence coverage

There are three main web resources that 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 Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI)
 - www.ebi.ac.uk
- DNA DataBank of Japan (DDBJ)
 - www.ddbj.nig.ac.jp

Note: The USPTO, EPO and JPO/KIPO rely on the NCBI, EMBL-EBI and DDBJ, respectively, to provide an interface for searching patent sequence data.

Why does STN have more patent sequences than NCBI, EBI and DDBJ?

1. NCBI/EBI/DDBJ have a significantly incomplete collection of WIPO/PCT sequence data
2. NCBI/EBI/DDBJ do not provide any USPTO published patent application sequence data
3. NCBI/EBI/DDBJ have an incomplete collection of USPTO granted patent sequence data
4. DGENE and REGISTRY cover many more patent authorities than NCBI/EBI/DDBJ
5. DGENE, USGENE, PCTGEN and REGISTRY are typically more timely than NCBI/EBI/DDBJ

A simple BLAST® example shows the importance of using STN databases

Search Question:

Find all patent references to Breast Cancer 1, early onset isoform 1 [Homo sapiens] protein, or other very similar proteins (i.e. >80% match).

(Search conducted on 28th October 2011.)

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

The image shows a screenshot of the NCBI Protein database interface. The search results for "breast cancer 1, early onset isoform 1 [Homo sapiens]" are displayed. The "FASTA" format is selected and circled in red. A Notepad window titled "BRCA1 protein - Notepad" is open, showing the amino acid sequence of the protein. The sequence is as follows:

```
>gi|6552299|ref|NP_009225.1| breast cancer 1, early onset isoform 1 [Homo sapiens]
MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITK
RSLQESTRFSQLVEELLKIIICAFQLDTGLEANSYNFAKKENNSPEHLKDEVSIQSMGYRNRARLLQS
EPENPSLQETSLSVQLSNLGTVRTLRKQRIQPQKTSVYIELGSDSSEDTVNKATYCSVGDQELLQITPQ
GTRDEISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSVSNLHVEPCGNTHA
SSLQHENSLLLLTKDRMNVEKAEFCNKSKQPGLARSQHNRWAGSKETCNDRRTPSTTEKKVDLNADPLCER
KEWNKQKLPCCSENPRDTEDPWITLNSSIQKVNWFVSRDELGSDSDHDGESESNKAVADVLDVLEVD
EYSGSSEKIDLLASDPHEALICKSERVHKSVESENIEDKIFGKTYRKKASLPNLSHVTEENLIIGAFVTEP
QIIQERPLTNLKRKRPTSGLHPEDFIKKADLAVQKTPEMINQGTNQTQNGQVMNITNSGHENKTKGD
SIQNEKNPNPIESLEKESAFKTKAEPISSEISNMELELNHNKAPKKNRRLRRKSSSTRHIIHALELVSRN
LSPPNCTELQIDSCSSSEIIEKKNYQMPVHRNQLQMEGKEPATGAKKSNKPNQTSKRHSDTFPEL
KLTNAPGSGFTKCSNTSELKEFVNPSLPREEKEKLETVKVSNNAEKPKDMLSGERVLQTERSVESISSI
LVPGTDYGTQESISLLEVSTLGGAKTEPNKCVSQCAAFENPKGLIHGCSKDNDRNDTEGFKYPLGHEVNH
RETSIEMEESELDQAQLQNTFKVSKRQSFAPFSPNGNAEEECATFSAHSGSLKKQSPKVTFECEQKEENQ
GKNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRI
PPLFPKISFVTKKCKNLLLENFEEHSMSPEREMGNENIPSTVSTISRNNIRENVFKEASSNI NEVGSS
TNEVGSSINEIGSSDENIQAE LGRNRGPKLNAMLRLGLVQPEVYKQSLPGSNCKHPEIKKQYEEVVQTV
NTDFSPYLI SDNLEQPMGSSHASQVCSSETPDDLDDGEEKEDTSFAENDIKESSAVFSKSVQKGLSRSP
SPFTHTHLAQGYRRGAKKLESSEENLSEDEELPCFQHLLFGKVNNI PSQSTRHSTVATECLSKNTEENL
LSLKNLSLNDCSNQVILAKASQEHHLSEETKCSASLFSQQSELEDLTANTNTQDPFLIGSSKQMRHQSES
QGVGLSDKELVSDDEERTGLEENNQEEQSMDSNLGEAASGCESETSVSEDCSGLSSQSDILTTQQRDTM
QHNLIKQEMAELEAVLEQHGSPNSYPSIISDSSALEDLRNPQSTSEKAVLTSQKSSEYPI SQNPE
GLSADKFEVSADSSSTSKNKEPGVERSSPKCPSLDDRWMHSCSGSLQNRNYPQEEELIKVVDVEEQLE
ESGPHDLTETSYLPRQDLEGTPLYESGILSFDSDPESDPSEDRAPE SARVGNIPSSSALKVQPQLKVAES
AQSPAAAHTTDTAGYNAMESVSRKPELTASTERVNRKMSMVVSGLTPEEFMLVYKFAKHHITLTNLI
TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVVSYFVWTQSIKERKMLNEHDFEVRGDVVNGRNHQGP
RARESQRKIFRGLIEICCYGPFNTMPTDQLEWVQLCGASVVKELSSFTLGTGVHPIVVVQPDAWHTDNG
FHAIGQMCEAPVVTREWVLDVALYQCQELDTYLIPIPHSHY
```

Example: The best answer from DGENE

=> D L2 BIB SCORE ALIGN 1

L2 ANSWER **1 OF 140** DGENE COPYRIGHT 2011 THOMSON

AN AZI42431 protein DGENE [Full-text](#)

TI Determining whether tumor sample from breast cancer patient has breast cancer-associated gene (BRCA) deficiency, somatic BRCA deficiency and germline BRCA deficiency, useful for treating cancer e.g. breast and ovarian cancer.

IN Timms K; Potter J; Lanchbury J
PA (MYRI) MYRIAD GENETICS INC.

PI WO 2011057125 A2 20110512

AI WO 2010-US55708 20101105

PRAI US 2009-258504P 20091105

PSL Disclosure; SEQ ID NO 2

DT Patent

LA English

OS 2011-F19431 [34]

CR N-PSDB: AZI42430

PC-NCBI: gi6552299

PC-SWISSPROT: P38398

PC-BIND: 151619; 12718; 12717; 149281

DESC **Human breast cancer 1 susceptibility protein, SEQ ID:2.**

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%)

. . . .

140 patent sequence hits were found in DGENE.

DGENE records feature extensive Thomson Reuters value-added content.

Example: The best patent answer from NCBI

```
>  gb|AAC11849.1|I78815 Sequence 2 from patent US 5693473
gb|AAC20811.1|I81570 Sequence 2 from patent US 5709999
gb|AAC20817.1|I81576 Sequence 2 from patent US 5710001
▶7 more sequence titles
Length=1863
```

39 patent sequence hits were found at NCBI.

```
Score = 3722 bits (9652), Expect = 0.0
Identities = 1863/1863 (100%), Positives = 1863/1863 (100%), Gaps = 0/1863 (0%)
```

```
Query 1 MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ 60
Sbjct 1 MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQ 60

Query 61 CPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEAYANSYNFAKKENNSPEHLKD 120
Sbjct 61 CPLCKNDITKRSLQESTRFSQLVEELLKIICAFQLDTGLEAYANSYNFAKKENNSPEHLKD 120

Query 121 EVSIIQSMGYRNR
Sbjct 121 EVSIIQSMGYRNR
```

Reminder: NCBI does not provide coverage of U.S. published patent application sequences, and only has partial coverage of WIPO/PCT sequences.

Summary of results: Breast Cancer 1, early onset isoform 1 [Homo sapiens] protein

	Sequences > 80%	Patent Families
DGENE	140	40
USGENE	138	33 (5)*
PCTGEN	13	9
REGISTRY	65	33 (1)*
NCBI	39	13
Total Unique	-	46

* Indicated in (red) are the number of additional families retrieved which contribute to the **Total Unique**.

Comparing databases...

- **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 10: www.trilateral.net/projects/biotechnology/guide2.pdf

Database timeliness is an important factor in understanding comprehensiveness

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

Summary

- STN sequence searchable databases
 - DGENE, USGENE, PCTGEN and REGISTRY
- Comparison to NCBI and EMBL-EBI
 - Focus on patent sequence coverage

Resources for sequence searching on STN

- Recorded e-Seminars (audio/video)
http://www.stn-international.com/recorded_events.html
 - Sequence Basics (all databases)
 - Multifile patent sequence searching (all databases)
- DGENE Workshop Manual
http://www.stn-international.com/dgene_wm.html
- USGENE Workshop Manual
http://www.stn-international.com/usgene_wm.html

STN[®]

For more information ...

CAS

E-mail: help@cas.org

Support and Training:

www.cas.org

FIZ Karlsruhe

helpdesk@fiz-karlsruhe.de

Support and Training:

www.stn-international.de