

STN[®]

Searching for antibody information on STN[®]

Robert Austin – FIZ Karlsruhe

Agenda

- Introduction to antibodies and immunoglobulins
- Understanding and searching antibody indexing in DGENE, REGISTRY and CAplusSM
- Sequence searching for antibodies using DGENE, USGENE[®], PCTGEN and REGISTRY
 - Complementarity Determining Regions (CDRs) using BLAST[®] and Sequence Code Match (SCM) searching
 - Multi-file search and post-processing

See also: Sequence Basics e-Seminar (May 2011):

http://www.stn-international.com/Sequence_Basics_Seminar.html

Why use STN to search for antibodies?

- Sequence information on STN is comprehensive
 - Four sequence databases allow users to achieve a comprehensive search of sequences from both patent and non-patent documents
- Sequence databases on STN are timely
 - Allows you to keep up-to-date with the most current information
- Sequence indexing is unique
 - Allows you to retrieve sequences containing uncommon residues or chemical modifications that are difficult to find

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 REGISTRY**
 - Chemical Abstracts Service (CAS) REGISTRY
 - Worldwide value-added patent and non-patent sequences

Antibodies are produced as a defence against foreign substances (antigens)

- Antibodies (Ab) are specialised glycoproteins, which differ in size, charge, carbohydrate content and amino acid sequence composition
- They are also known as immunoglobulins (Ig)
- Antibodies are found in blood and other bodily fluids of mammals and some other vertebrates
- They are a central part of the humoral immune response (HIR) and are synthesised by B-cells

Antibodies are useful because of their biological properties and high specificity

- There are different classes of antibodies, depending on their structure
 - Mammals have five classes of antibodies
 - α (IgA), δ (IgD), ϵ (IgE), γ (IgG), μ (IgM)
 - Each class has different biological properties
- Antibodies are highly specific to antigens
 - Able to locate one molecule of a protein antigen out of more than 10^8 similar molecules
 - Useful in targeted therapy and as diagnostic tools

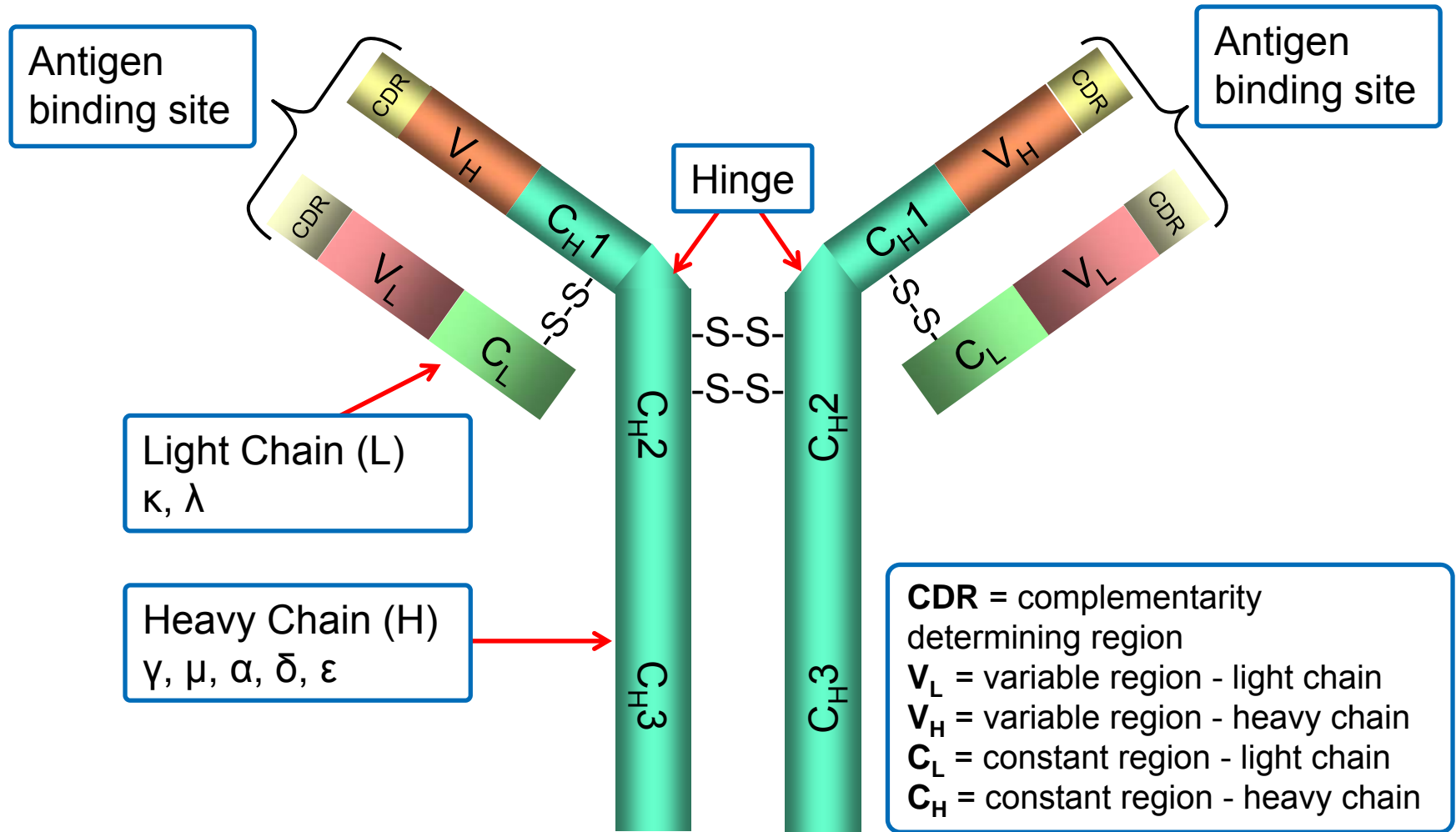
Mammalian antibodies are Y-shaped and composed of heavy and light chains

- Antibodies are composed of four polypeptides
 - Two identical light chain (L)
 - Two identical heavy chain (H)
- Both light and heavy chains consist of constant (C) region domains with little variability, and variable (V) region domains with high variability
- The four chains are held together by a several disulphide bonds and form a Y-shaped molecule
- The antigen binding sites (CDRs) are located at the tips of the Y-shaped arms

Additional nomenclature describes variable region domains

- Light chains exist in two forms
 - kappa (κ), lambda (λ)
- Heavy chains exist in five forms
 - α (IgA), δ (IgD), ε (IgE), γ (IgG), μ (IgM)
 - Variation in heavy chains gives rise to various antibody subclasses: IgG1, IgG2, IgA1, etc.

Mammalian antibodies are Y-shaped and composed of heavy and light chains



Humanization of antibodies is an important process for therapeutic usage

- Immunotherapy (or biotherapy)
 - Uses certain parts of the immune system to fight diseases such as cancer
 - Treatments are less toxic and potentially more effective than chemical drugs
 - Types of antibodies used for therapy
 - Monoclonal antibodies
 - Chimeric antibodies
 - CDR-Grafted antibodies
 - Phage Display antibodies

Antibodies are also used as diagnostic tools

- Antibody tools reduce assay time without compromising sensitivity
 - Flow cytometric analysis
 - Analysis of morphological complexity of the cells, DNA content (cell cycle analysis), cell sorting
 - Microarray technology
 - Proteomics: Protein characterization and analysis of diseased vs. healthy patients
 - Immunoblotting (or western blot)
 - Detection of a specific protein from a tissue or cell sample
 - Immunohistochemistry
 - Localization of protein(s) in cells or tissue sections using antibodies

Agenda

- Introduction to antibodies and immunoglobulins
- Understanding and searching antibody indexing in DGENE, REGISTRY, and CAplusSM
- Sequence searching for antibodies using DGENE, USGENE[®], PCTGEN and REGISTRY
 - Complementarity Determining Regions (CDRs) using BLAST[®] and Sequence Code Match (SCM) searching
 - Multi-file search and post-processing

Antibody sequences are indexed in GENESEQ™ on STN (DGENE)

- Description (/DESC)
 - Concise one-line description of the sequence
 - E.g. Mouse anti-protein X antibody VL region
- Keyword (/KW) indexing for antibody sequences
 - **Type**, e.g. humanized, monoclonal
 - **Region**, e.g. light chain constant region
 - **Activity**, e.g. antibody therapy, immune stimulation
 - **Target**, e.g. protein X
 - **Disease**, e.g. immune disorder, autoimmune disease
 - **Technology**, e.g. antibody engineering, antibody array
- Abstract (AB)
 - Includes the use of the antibody within the invention
- Features Table (/FEAT)
 - Details about Domain, Region, Disulphide-bonds, etc

Each DGENE record has a concise one-line description of the antibody sequence

```
L1      ANSWER 1 OF 1  DGENE  COPYRIGHT 2010 THOMSON REUTERS on STN
ACCESSION NUMBER: AEN02775  protein          DGENE
TITLE:      New antibody useful for treating e.g. cancer
            competitively inhibiting binding of competitor antibody
            having complementarity determining region of specific
            amino acid sequences as given in the specification to G-
            protein coupled receptor.
INVENTOR:   Howard M; Schall T
PATENT ASSIGNEE: (CHEM-N)CHEMOCENTRYX INC.
PATENT INFO:   WO 2006116319  A2  20061102          60
APPLICATION INFO: WO 2006-US15492          20060419
PRIORITY INFO:  US 2005-674140P          20050421
PAT. SEQ. LOC:  Disclosure; SEQ ID NO 22
DATA ENTRY DATE: 22 FEB 2007 (first entry)
DOCUMENT TYPE:  Patent
LANGUAGE:      English
OTHER SOURCE:  2007-110158 [11]
DESCRIPTION:  anti-CCX-CK2-antibody 11G8 VL region SEQ ID NO 22.
```

DGENE records also include the DWPI patent title (/TI).

Each DGENE record has keyword indexing for the antibody sequence

KEYWORD: cytostatic; neuroprotective; nootropic; nephrotropic; antirheumatic; antiarthritic; cardiant; antiarteriosclerotic; antiasthmatic; dermatological; antiinflammatory; gastrointestinal -gen.; antipsoriatic; vasotropic; immunosuppressive; antiulcer; ophthalmological; antidiabetic; vulnerary; hepatotropic; anorectic; respiratory-gen.; gynecological; hemostatic; cardiovascular-gen.; contraceptive; protein interaction; **antibody; angiogenesis inhibition; cell proliferation; protein detection; antibody therapy;** arthritis; Alzheimers disease; multiple sclerosis; renal failure; rheumatoid arthritis; transplant rejection; asthma; glomerulonephritis; contact dermatitis; inflammatory bowel disease; colitis; psoriasis; reperfusion injury; ocular disease; diabetic retinopathy; retinopathy of prematurity; macular degeneration; graft rejection; neovascular glaucoma; rubeosis; Osier-Webber Syndrome; telangiectasis; angiofibroma; Crohns disease; eczema; . . . wound healing; osteopathic; fractures; burns; inflammation; ischemia; peripheral vascular disease; pre-eclampsia; **cardiovascular disease; 11G8; light chain variable region.**

ORGANISM: Mus sp.

Each DGENE abstract describes the use of the antibody sequence within the invention

ABSTRACT:

The invention describes an antibody (A1) that competitively inhibits binding of a competitor antibody (a1) to CCX-CKR2 (G-protein coupled receptor), where the competitor antibody comprises the complementarity determining region (CDR) of specific amino acid sequences as given in the specification. The antibody is useful in a pharmaceutical composition for inhibiting angiogenesis or proliferation of a cancer cell in an individual such as other than human having or pre-disposed to have arthritis; for detecting a cell expressing CCX-CKR2 in a biological sample; for treating Alzheimer's disease; multiple sclerosis; kidney dysfunction; rheumatoid arthritis; cardiac allograft rejection; atherosclerosis; asthma; glomerulonephritis; contact dermatitis; inflammatory bowel disease; colitis; psoriasis; reperfusion injury; ocular angiogenic diseases, for example, ••• joints (e.g. arthritis and hemophilic joints), healing of wounds, fractures, and burns, inflammatory diseases, ischemic heart, and peripheral vascular diseases; preclampsia and cardiovascular disease; for birth control. The antibody competitively inhibits binding of a competitor antibody to CCX-CKR2, and potently inhibits angiogenesis. This is the amino acid sequence of anti-CCX-CK2-antibody 11G8 light chain variable region.

The DGENE Feature Table includes detailed annotations for the antibody sequence

AMINO ACID COUNTS: 2 A; 4 R; 2 N; 6 D; 0 B; 2 C; 5 Q; 3 E; 0 Z; 9 G; 3 H;
5 I; 10 L; 4 K; 1 M; 4 F; 6 P; 15 S; 5 T; 1 W; 6 Y; 7 V;
0 Others

SEQUENCE LENGTH: 100

SEQUENCE

1 dvlmtqtpls lpvslgdqas iscrsshiv hsdgntylew ylqkpggspk
51 lliykvsnrf sgvpdrfsgs gsgtdftlki srveaedlgi yycfqgshvp

FEATURE TABLE:

Key	Location	Qualifier	
Region	1..23	note	"framework region 1"
Region	24..39	note	"CDR1"
Region	40..54	note	"framework region 2"
Region	55..61	note	"CDR2"
Region	62..93	note	"framework region 3"

Antibody light and heavy chain sequences are indexed in separate DGENE records

L1 ANSWER 1 OF 2 DGENE COPYRIGHT 2010 THOMSON REUTERS on STN
AN AWO22746 protein DGENE
TI Purifying an antibody from a composition comprises loading the . . .
DESC Anti-VEGF bevacizumab humanized antibody **light chain sequence,**
SEQ ID 12.
KW protein purification; VEGF ligand; cation-exchange; chromatography;
light chain; humanized antibody; protein purification; vascular
endothelial growth factor.
SQL **214**

L1 ANSWER 2 OF 2 DGENE COPYRIGHT 2010 THOMSON REUTERS ON STN
AN AWO22745 protein DGENE
TI Purifying an antibody from a composition comprises loading the . . .
DESC Anti-VEGF bevacizumab humanized antibody **heavy chain sequence,**
SEQ ID 11.
KW protein purification; VEGF ligand; cation-exchange; chromatography;
heavy chain; humanized antibody; protein purification; vascular
endothelial growth factor.
SQL **453**

Thomson Reuters indexing makes clear which one is which.

Note: This example comes from WO2009058812.

Antibody light and heavy chain sequences are indexed in separate USGENE records

=> D AN TRIAL 1-2

L2 ANSWER 1 OF 2 USGENE COPYRIGHT 2010
AN 20090148435.12 Protein USGENE
TI ANTIBODY PURIFICATION BY CATION EXCHANGE CHROMATOGRAPHY
(PublishedApplication)
DESC Artificial Protein; Sequence is synthesized; sequence 12 of 20
MTY Protein
SQL 214

Note: This example comes from US20090148435 A1, which is equivalent to WO2009058812 A1.

L2 ANSWER 2 OF 2 USGENE COPYRIGHT 2010 SEQUE
AN 20090148435.11
TI ANTIBODY PURIFICATION BY CATION EXCHANGE CHROMATOGRAPHY
(PublishedApplication)
DESC Artificial Protein; Sequence is synthesized; sequence 11 of 20
MTY Protein
SQL 453

Patent applicants often do not provide a clear description.

Antibodies are indexed as substances in CAS REGISTRY

- Antibodies are indexed as sequences if the sequence(s) is provided by the author(s)
 - The Note (/NTE) field contains additional information about the sequence (i.e. chemically modified, linkages, uncommon amino acids, etc.)
 - Separate records for the full multi-chain antibody, light chain and heavy chain sequences may be created
- Antibody sequences are also indexed with
 - Index Names
 - Trade names
 - Generic names
 - Lab names

CA Index Names provide information, such as sequence type, organisms, and strain/cell/tissue types

L1 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2010 ACS on STN
RN 216974-75-3 REGISTRY

CN Immunoglobulin G1, anti-(human vascular endothelial growth factor)(human-mouse monoclonal rhuMAb-VEGF g1-chain), disulfide with human-mouse monoclonal rhuMAb-VEGF light chain, dimer (9CI) (CA INDEX NAME)

OTHER NAMES:

CN Anti-VEGF monoclonal antibody

CN Avastatin

CN Avastin

CN Bevacizumab

CN rhuMAb-VEGF

FS PROTEIN SEQUENCE

SQL 1334,453,453,214,214

The CA Index Name for Avastin contains additional information, such as the isotype (G1), the antigen (VEGF), monoclonal antibody, etc.

Brand/generic names, and lab names are listed under "Other names".

Avastin is registered as a multichain sequence (two heavy and two light chains).

Modifications and/or linkages between chains are listed in the /NTE field

NTE **multichain**

type	-----	location	-----	description
bridge	Cys-22	-	Cys-96	disulfide bridge
bridge	Cys-150	-	Cys-206	disulfide bridge
bridge	Cys-226	-	Cys-214''	disulfide bridge
bridge	Cys-232	-	Cys-232'	disulfide bridge
bridge	Cys-235	-	Cys-235'	disulfide bridge
bridge	Cys-267	-	Cys-327	disulfide bridge
bridge	Cys-373	-	Cys-431	disulfide bridge
bridge	Cys-22'	-	Cys-96'	disulfide bridge
bridge	Cys-150'	-	Cys-206'	disulfide bridge
bridge	Cys-226'	-	Cys-214''''	disulfide bridge
bridge	Cys-267'	-	Cys-327'	disulfide bridge
bridge	Cys-373'	-	Cys-431'	disulfide bridge
bridge	Cys-23''	-	Cys-88''	disulfide bridge
bridge	Cys-134'''	-	Cys-194'''	disulfide bridge

In REGISTRY, the specific residue(s) position(s) are listed in the /NTE field.

CAS will index both light and heavy chains for antibodies

SEQ

Heavy Chain

```
1 EVQLVESGGG LVQPGGSLRL SCAASGYTFT NYGMNWVRQA PGKGLEWVGW
51 INTYTGEPTY AADFKRRFTF SLDTSKSTAY LQMNSLRAED TAVYYCAKYP
101 HYYGSSHWFY DVWGQGTLLV VSSASTKGPS VFPLAPSSKS TSGGTAALGC
. . .
401 PPVLDSGDSF FLYSKLTVDK SRWQQGNVFS CSVMHEALHN HYTQKLSLSL
451 PGK
```

SEQ

Heavy Chain

```
1 EVQLVESGGG LVQPGGSLRL SCAASGYTFT NYGMNWVRQA PGKGLEWVGW
51 INTYTGEPTY AADFKRRFTF SLDTSKSTAY LQMNSLRAED TAVYYCAKYP
101 HYYGSSHWFY DVWGQGTLLV VSSASTKGPS VFPLAPSSKS TSGGTAALGC
. . .
401 PPVLDSGDSF FLYSKLTVDK SRWQQGNVFS CSVMHEALHN HYTQKLSLSL
451 PGK
```

SEQ

Light Chain

```
1 DIQMTQSPSS LSASVGDRVT ITCSASQDIS NYLNWYQQKP GKAPKVLIIYF
51 TSSLHSGVPS RFSGSGSGTD FTLTISLQP EDFATYYCQQ YSTVPWTFGQ
. . .
201 LSSPVTKSFN RGEC
```

SEQ

Light Chain

```
1 DIQMTQSPSS LSASVGDRVT ITCSASQDIS NYLNWYQQKP GKAPKVLIIYF
51 TSSLHSGVPS RFSGSGSGTD FTLTISLQP EDFATYYCQQ YSTVPWTFGQ
. . .
201 LSSPVTKSFN RGEC
```

Light and heavy chain sequences may also be indexed in separate REGISTRY records

=> D SQIDE 1-2

```
L4 ANSWER 1 OF 2 REGISTRY COPYRIGHT 2010 ACS on STN
RN 1150802-75-7 REGISTRY
CN 8: PN: WO2009058812 SEQID: 12 unclaimed protein (CA INDEX
NAME)
FS PROTEIN SEQUENCE
SQL 214
. . .
```

Note: REGISTRY sequence records from patents, often do not include a description of the sequence.

```
L4 ANSWER 2 OF 2 REGISTRY COPYRIGHT 2010 ACS on STN
RN 1150802-74-6 REGISTRY
CN 7: PN: WO2009058812 SEQID: 11 unclaimed protein (CA INDEX
NAME)
FS PROTEIN SEQUENCE
SQL 453
. . .
```

Tip: Compare to DGENE (slide 14) and to USGENE (slide 19).

The antibody REGISTRY numbers are indexed in CAplus bibliographic records

```

L26 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2010 ACS on STN
AN 2009:553192 HCAPLUS
DN 150:513022
TI Antibody purification by cation exchange chromatography using a high
pH wash step to remove of contaminants prior to eluting in a buffer
with increased conductivity
IN Lebreton, Benedicte Andree; O'Connor, Deborah Ann; Safta, Aurelia;
Sharma, Mandakini
PA Genentech, Inc., USA
SO PCT Int. Appl., 53pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 1
PATENT NO. KIND DATE APPLICATION NO. DATE
-----
PI WO 2009058812 A1 20090507 WO 2008-US81516 20081029
. . . .
AB A method for purifying an antibody by cation exchange chromatog. is
described in which a high pH wash step is used to remove of
contaminants prior to eluting the desired antibody using an elution
buffer with increased conductivity Preferably the antibody binds
human CD20, such as rituximab, or binds human vascular endothelial
growth factor (VEGF), such as bevacizumab. . . .
    
```

CAplus records provide title, patent family and abstract.

The antibody Registry Numbers are linked to detailed roles and index terms in CAplus

CC 15-3 (Immunochemistry)
Section cross-reference(s): 9
.

IT 174722-31-7P, Rituximab **216974-75-3P, Bevacizumab**
RL: ARG (Analytical reagent use); BPN (Biosynthetic preparation); PUR (Purification or recovery); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); PREP (Preparation); USES (Uses)
(antibody purification by cation exchange chromatog. using high pH wash step to remove of contaminants prior to eluting in buffer with increased conductivity)
.

IT 192433-87-7 214551-08-3 214551-09-4 214551-11-8 214551-12-9
214551-13-0 444104-00-1 556112-97-1 556112-98-2 556112-99-3
556113-00-9 1150529-46-6 **1150802-74-6 1150802-75-7** 1150802-76-8
1150802-77-9 1150802-78-0 1150802-79-1 1150802-80-4 1150802-81-5
RL: PRP (Properties)
(unclaimed protein sequence; antibody purification by cation exchange chromatog. using a wash step to remove of contaminants prior to eluting in buffer with increased conductivity)
.

This CAplus record has indexing for both the multi-chain antibody, and separately for the light and heavy chain sequences.

Antibodies are indexed in CAplus

- Covered by controlled terms in CAplus
 - Consult the Lexicon for old and new terms
 - Refine with additional concepts contained within the same index term by using the (L) operator
- CAS Registry Numbers (CAS RNs) are available to supplement a CAplus search, if retrieval of specific antibody substances is required

How are antibodies indexed in CPlus?

- Antibodies are indexed to the most specific level disclosed in a document
 - Light chains
 - κ (kappa), λ (lambda)
 - Heavy chains
 - α (IgA), δ (IgD), ϵ (IgE), γ (IgG), μ (IgM)
 - Subclasses: IgG1, IgG2, IgA1, etc.
- Descriptors (limiters) can provide additional information
 - Examples: bispecific, catalytic, labeled, neutralizing, humanized, chimeric, etc.

Antibody controlled term indexing has changed over time in CPlus

Controlled Term	Years of Coverage
Antibodies and Immunoglobulins/CT	2002 to present
Amboceptors/CT	1907-1946
Antibodies/CT	1907-2001
Globulins, immune/CT	1967-1976
Immunoglobulins/CT	1962-2001

Immunoglobulin (Ig) nomenclature can be used to focus on specific forms of interest

Search Question:

Find records covering lambda light chain immunoglobulins. Are any of these IgG immunoglobulins?

Identify uses of these substances.

Search Strategy

To find references on immunoglobulins...

- 1) Search appropriate Ig terms
- 2) Refine with class, subclass, or chain nomenclature
- 3) Evaluate using **D SCAN HIT**
- 4) (Optional) Refine with additional concepts or CAS Roles
- 5) Display results

Tips for searching in CPlus

- Use BOTH single word immunoglobulin and antibody terms in the /IT field, especially to include records prior to 2002
- Use (L) proximity to add modifying terms
- Specific subclass may be used but allow for generic class for comprehensive results
 - For example, IgG2a is of most interest but IgG2 encompasses it and should be included
- Supplement CAS Roles with text terms in the modifier

Use SET commands to automatically add plurals and abbreviations

```
=> SET PLU ON; SET ABB ON; SET SPE ON
```

```
SET COMMAND COMPLETED
```

```
=> S (IMMUNOGLOBULIN(L)LAMBDA)/IT
```

```
11999 IMMUNOGLOBULIN/IT
```

```
143185 IMMUNOGLOBULINS/IT
```

```
150880 IMMUNOGLOBULIN/IT
```

```
((IMMUNOGLOBULIN OR IMMUNOGLOBULINS)/IT)
```

```
20421 IG/IT
```

```
6316 IGS/IT
```

```
25278 IG/IT
```

```
((IG OR IGS)/IT)
```

```
156119 IMMUNOGLOBULIN/IT
```

```
((IMMUNOGLOBULIN OR IG)/IT)
```

```
178328 LAMBDA
```

```
68 LAMBDA S
```

```
178342 LAMBDA
```

```
(LAMBDA OR LAMBDA S)
```

```
L1 1213 IMMUNOGLOBULIN/IT (L) LAMBDA
```

Adding antibody term retrieves additional answers

```
=> S ANTIBODY/IT (L) LAMBDA
      70328 ANTIBODY/IT
      214281 ANTIBODIES/IT
      223724 ANTIBODY/IT
            ((ANTIBODY OR ANTIBODIES)/IT)
      178328 LAMBDA
            68 LAMBDAS
      178342 LAMBDA
            (LAMBDA OR LAMBDAS)
L2      568 ANTIBODY/IT (L) LAMBDA

=> S L1 OR L2
L3      1374 L1 OR L2
```

Alternative search using **BOTH** terms

```
=> S (ANTIBODY OR IMMUNOGLOBULIN)/IT (L) LAMBDA
L4 1374 (ANTIBODY OR IMMUNOGLOBULIN)/IT (L) LAMBDA
```

```
=> S L4 (L) IGG?
      80343 IGG?
L5      58 L4 (L) IGG?
```

Both Lambda and Kappa forms are indexed.

```
=> D HIT SCAN
```

```
L5 58 ANSWERS CAPLUS COPYRIGHT 2010 ACS on STN
IT Antibodies and Immunoglobulins
```

```
RL: ADV (Adverse effect, including toxicity); BSU (Biological study, unclassified); DGN (Diagnostic use); BIOL (Biological study); USES (Uses)
      (IgG, .lambda., .kappa.; gammopathy detected by serum protein electrophoresis for predicting and managing therapy of lymphoproliferative disorder in liver transplant recipients)
```

```
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):END
```

Remove the kappa light chain entries from the index term search

```
=> S L5 (NOTL) KAPPA
      70588 KAPPA
          11 KAPPAS
      70594 KAPPA
          (KAPPA OR KAPPAS)
L6          32 L5 (NOTL) KAPPA
```

Exclude Kappa from the index term search with **(NOTL)**.

```
=> D SCAN HIT
```

```
L6      32 ANSWERS      CAPLUS      COPYRIGHT 2010 ACS on STN
IT      Antibodies and Immunoglobulins
      RL: BSU (Biological study, unclassified); BIOL (Biological
      study)
          (light chain, .lambda.; of IgG autoantibodies to
          TSH receptors in Graves disease of humans)
```

```
HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):2
```

Index terms describe therapeutic applications

L6 32 ANSWERS CAPLUS COPYRIGHT 2010 ACS on STN

IT Lymphoma

(B-cell, monoclonal **antibodies** to **IgG1**
.lambda. paraprotein variable region epitopes in
diagnosis of human)

IT **Antibodies**

RL: BIOL (Biological study)

(monoclonal, **IgG1.lambda.** paraprotein variable
region epitopes recognition by, of human, lymphoid
disease diagnosis in relation to)

L6 32 ANSWERS CAPLUS COPYRIGHT 2010 ACS on STN

IT **Immunoglobulins**

RL: PRP (Properties)

(mapping of **.lambda.** light chain epitopes for human
lupus **IgG** autoantibodies)

IT Protein sequences

(of **Ig .lambda.** light chains of humans in relation
to lupus **IgG** autoantibody binding)

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0

Agenda

- Introduction to antibodies and immunoglobulins
- Understanding and searching antibody indexing in DGENE, REGISTRY, and CAplusSM
- Sequence searching for antibodies using DGENE, USGENE[®], PCTGEN and REGISTRY
 - Complementarity Determining Regions (CDRs) using BLAST[®] and Sequence Code Match (SCM) searching
 - Multi-file search and post-processing

See also: Multi-file BLAST searching (June 2010):
http://www.stn-international.com/usgene_wm_mfs.html

Sequence searching for CDRs

- Within the variable domains of the heavy and light chains, there are three hyperactivity regions called Complementarity Determining Region (CDR)
- Each chain contains three CDR regions
- CDR is the region that recognizes the different antigens
 - CDR is important when developing antibodies

BLAST and/or Sequence Code Match (SCM) can be used to retrieve CDRs

Search Question:

The epithelial cell adhesion molecule (EpCAM) is a cell surface protein that is expressed by a variety of tumor cells. We have identified CDR sequence **DMGWGSGWRP** **YYYGMDV** in our laboratory. Find all patent publications that disclose this sequence or similar sequences.

Search Strategy for DGENE, USGENE, PCTGEN and REGISTRY/CAplus

- 1) RUN BLAST in DGENE, USGENE and PCTGEN using offline BATCH mode
- 2) Repeat the search using CAS REGISTRY BLAST
- 3) Retrieve, merge, organize by patent family, and display DGENE, USGENE and PCTGEN results
- 4) Retrieve, identify and display unique CAplus references from the REGISTRY BLAST search
- 5) Post-process DGENE, USGENE and PCTGEN results using the STN Express Table Tool
- 6) Post-process unique REGISTRY BLAST results using the BLAST Report Tool

NCBI recommended settings* for searching small sequence queries

Peptide sequences

- E-value: 20,000
- Word size: 2
- Matrix: PAM-30
- Gap cost: 9 and 1

Nucleotide sequences

- E-value: 1,000
- Word size: 7
- Matrix: Leave as is
- Gap cost: n/a

* <http://www.ncbi.nlm.nih.gov/blast/Why.shtml>

RUN BLAST searches in DGENE, USGENE and PCTGEN in offline BATCH mode

```
=> FILE DGENE; RUN BLAST DMGWGSGWRPYYYYGMDV/SQP -F F -E 20000  
-W 2 -M PAM30 BATCH
```

```
PLEASE ENTER BATCH IDENTIFIER (MAX. 8 CHARS):EPCAMCDR
```

```
BATCH PROCESSING STARTED FOR EPCAMCDR
```

Add **BATCH** to the end of the command.

```
=> FILE USGENE; RUN BLAST DMGWGSGWRPYYYYGMDV/SQP -F F -E 20000  
-W 2 -M PAM30 BATCH
```

```
PLEASE ENTER BATCH IDENTIFIER (MAX. 8 CHARS):EPCAMCDR
```

```
BATCH PROCESSING STARTED FOR EPCAMCDR
```

Use the NCBI's recommended BLAST settings.

```
=> FILE PCTGEN; RUN BLAST DMGWGSGWRPYYYYGMDV/SQP -F F -E 20000  
-W 2 -M PAM30 BATCH
```

```
PLEASE ENTER BATCH IDENTIFIER (MAX. 8 CHARS):EPCAMCDR
```

```
BATCH PROCESSING STARTED FOR EPCAMCDR
```

Tip: BATCH mode BLAST searches can be run *concurrently* in each database.

Repeat using CAS REGISTRY BLAST

- a) Launch BLAST
- b) Search the sequence
- c) Examine and evaluate alignment/relevance of sequence answers
- d) Retrieve REGISTRY records on STN

See also: Sequence Basics e-Seminar (May 2011):

http://www.stn-international.com/Sequence_Basics_Seminar.html

Adjust the REGISTRY BLAST settings for optimal search retrieval

BLASTp Settings - Additional Options - EPCAMCDR

BLASTp Settings - Additional Options - EPCAMCDR

Additional Option Presets

Search Sensitivity

Fewer Answers → More Answers

Show Additional Options

Basic Options

Low Complexity Filtering

Query Genetic Code: Standard(1)

Max No. of Answers: 1,000

Additional Options

Expectation Value: 20000

Word Size: 2

Weight Matrix: PAM-30

Gap Cost: Open: 9 Extend: 1

Penalty for Mismatch: -1

Reward for Match: 1

Reset to Defaults

OK Back Cancel

Increase the maximum number of answers to 1,000.

Use the NCBI's recommended BLAST settings.

Retrieve references for sequences

CAS Registry BLAST® Report - EPCAMCDR

File Edit View Search Tools Help

Unique Sequences: 1,000 Redundant: 326 Selected Results: 36

Alignment Scores

<40 40-50 50-80 80-200 >=200

Alignment Summary

Alignment Details

<input checked="" type="checkbox"/>	69.8	1e-11	(1065745-06-3) Immunoglobulin IgG1, anti-(human tumor antigen 17-1A) (human HD69 y1-chain V region)
<input checked="" type="checkbox"/>	69.8	1e-11	(1065745-11-0) Immunoglobulin IgG1, anti-(human tumor antigen 17-1A) (human HD69 Vy1-VL region sing
<input checked="" type="checkbox"/>	69.8	1e-11	(1065745-13-2) Immunoglobulin IgG1, anti-(human tumor antigen 17-1A) (human HD69 Vy1-VL region sing
<input checked="" type="checkbox"/>	69.8	1e-11	(1065745-15-4) Immunoglobulin IgG1, anti-(human tumor antigen 17-1A) (human HD69 Vy1-VL region sing
<input checked="" type="checkbox"/>	69.8	1e-11	(1065745-17-6) Immunoglobulin IgG1, anti-(human tumor antigen 17-1A) (human HD69 Vy1-VL region sing
<input checked="" type="checkbox"/>	69.8	1e-11	(942165-85-7) Immunoglobulin G2a, anti-(human tumor antigen 17-1A) (human-mouse y2a-chain)
<input checked="" type="checkbox"/>	69.8	1e-11	(945905-53-3) L-Valine, L-α-aspartyl-L-methionylglycyl-L-tryptophylglycyl-L-serylglycyl-L-tryptophyl-L-argi
<input checked="" type="checkbox"/>	69.8	1e-11	(945940-54-5) Immunoglobulin G1, anti-(tumor antigen 17-1A) (human y1-chain)
<input checked="" type="checkbox"/>	69.8	1e-11	(487483-96-5) GenBank CAB69288
<input checked="" type="checkbox"/>	69.8	1e-11	(481333-75-9) GenBank CAC08835

Get STN Data Cancel

Note: in this example, BLAST sequences with scores of 42 or more (60% match or more) are selected.

Retrieve REGISTRY BLAST results

Transfer BLAST sequences with scores of 42 or more (60% match or more).

```
=> FIL REGISTRY
=> QUE (1065745-06-3 OR 1067695-29-7 OR 1065745-11-0 OR . . . )/RN
L1   QUE (1065745-06-3 OR 1067695-29-7 OR 1065745-11-0 OR . . . )/RN
=> QUE (862861-57-2 OR 487483-96-5 OR 215027-98-8 OR . . . )/RN
L2   QUE (862861-57-2 OR 487483-96-5 OR 215027-98-8 OR . . . )/RN
=> QUE (1089240-08-3 OR 1074003-68-1 OR 960549-36-4 OR . . . )/RN
L3   QUE (1089240-08-3 OR 1074003-68-1 OR 960549-36-4 OR . . . )/RN
=> S L1 OR L2 OR L3
L4           36 L1 OR L2 OR L3
```

Commands within the dotted box are automatic commands.

36 similar sequences (L4) with BLAST scores of 42 or more.

Tip: BLAST is better than SCM for searching short sequences for less than 100% match

=> **FIL REG**

=> **S DMGWGSGWRPYYYYGMDV/SQSFP**

L5 20 DMGWGSGWRPYYYYGMDV/SQSFP

=> **S L5 NOT L4**

L6 0 L5 NOT L4

=> **DEL L5-L6 Y**

Subsequence family protein search (/SQSFP) (L5), allows for amino acid family substitution.

REGISTRY BLAST (L4) retrieved 16 extra sequences with 60% or higher match by score that were not retrieved with /SQSFP (L5).

Retrieve the DGENE, USGENE and PCTGEN in BLAST search results

```
=> FILE DGENE; RUN GETBATCH EPCAMCDR
```

```
. . . .
```

```
ENTER (ALL) OR ? :60%
```

```
L5 RUN STATEMENT CREATED
```

```
L5 32 DMGWGSGWRPYYYYGMDV/SQP.-F F -E 20000 -W 2 -M PAM30
```

Use RUN GETBATCH to retrieve completed BATCH search results.

```
=> FILE USGENE; RUN GETBATCH EPCAMCDR
```

```
. . . .
```

```
ENTER (ALL) OR ? :60%
```

```
L6 RUN STATEMENT CREATED
```

```
L6 17 DMGWGSGWRPYYYYGMDV/SQP.-F F -E 20000 -W 2 -M PAM30
```

```
=> FILE PCTGEN; RUN GETBATCH EPCAMCDR
```

```
. . . .
```

```
ENTER (ALL) OR ? :60%
```

```
L7 RUN STATEMENT CREATED
```

```
L7 12 DMGWGSGWRPYYYYGMDV/SQP.-F F -E 20000 -W 2 -M PAM30
```

In this example, 60% of the *Query Self Score* is used to select results of interest.

Merge and review the DGENE, USGENE and PCTGEN in BLAST search results

```
=> SET DUPORDER FILE
SET COMMAND COMPLETED
```

```
=> DUP IDE L5 L6 L7
```

```
FILE 'DGENE' ENTERED AT . . .
COPYRIGHT (C) 2010 THOMSON REUTERS
```

```
FILE 'USGENE' ENTERED AT . . .
COPYRIGHT (C) 2010 SEQUENCEBASE CORP
```

```
FILE 'PCTGEN' ENTERED AT . . .
COPYRIGHT (C) 2010 WIPO
```

```
L8      61 DUP IDE L5 L6 L7 (INCLUDES 0 SETS OF DUPLICATES)
        ANSWERS '1-32' FROM FILE DGENE
        ANSWERS '33-49' FROM FILE USGENE
        ANSWERS '50-61' FROM FILE PCTGEN
```

```
=> SOR SCORE D AN D
```

```
PROCESSING COMPLETED FOR L8
```

```
L9      61 SOR L8 SCORE D AN D
```

SET DUPORDER FILE ensures that multifile records merged using **DUP IDE** are organized by database (file).

DUPLICATE IDENTIFY (DUP IDE) is used here to create a single multifile L-number (**L8**).

The multifile L-number (**L8**) can be sorted by BLAST **SCORE**, or Percent Identity (**IDENT**).

Review multifile answers with a free-of-charge format including alignment

=> D L9 TRIAL SCORE ALIGN 1-61; FILE STNGUIDE

```
L9      ANSWER 1 OF 61  DGENE  COPYRIGHT 2010 THOMSON REUTERS on STN
AN      ATS30301  protein  DGENE
TI      New polypeptide comprises a binding domain capable of binding to an
        epitope of human and non-chimpanzee primate CD3 epsilon chain, useful
        for preventing, treating, or ameliorating a proliferative, tumor, or
        immunological disorder.
DESC    Anti-CD3/anti-EpCAM cross-species single chain Ab protein, SEQ: 592.
KW      single chain antibody; CD3E; T-cell CD3 glycoprotein epsilon chain;
        TACSTD1; EpCAM; protein production; protein therapy; therapeutic;
        prophylactic to disease; protein detection; immune disorder;
        immunomodulator; cancer; cytostatic; hyperproliferation; Fusion
        protein.
SQL     504
SCORE   70      100% of query self score 70
BLASTALIGN
        Query = 18 letters
        Length = 504
        Score = 69.8 bits (157), Expect = 5e-18
        Identities = 18/18 (100%), Positives = 18/18 (100%)
Query:  1  DMGWGSGWRPYYYYGMDV 18
        DMGWGSGWRPYYYYGMDV
Sbjct:  99 DMGWGSGWRPYYYYGMDV 116
```

Query Self Score
and percentage.

Review answers with a free-of-charge format including alignment (cont.)

```
L9 ANSWER 21 OF 61 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN
TI Method of identifying binding site domains that retain the capacity
of binding to an epitope (Patent)
DESC Homo Sapiens Protein; sequence 54 of 77
MTY Protein
SQL 127
SCORE 70 100% of query self score 70
BLASTALIGN
  Query = 18 letters
  Length = 127
  Score = 69.8 bits (157), Expect = 2e-18
  Identities = 18/18 (100%) Positives = 18/18 (100%)
Query: 1 DMGWGSGWRPYYYYGMDV 18
        DMGWGSGWRPYYYYGMDV
Sbjct: 99 DMGWGSGWRPYYYYGMDV 116
. . .
```

BLAST Percent Identity (IDENT).

```
L9 ANSWER 25 OF 61 PCTGEN COPYRIGHT 2010 WIPO on STN
TI CROSS-SPECIES-SPECIFIC BINDING DOMAIN
MTY PRT
SQL 504
SCORE 70 100% of query self score 70
BLASTALIGN
  Query = 18 letters
  Length = 504
  Score = 69.8 bits (157), Expect = 5e-18
  Identities = 18/18 (100%), Positives = 18/18 (100%)
Query: 1 DMGWGSGWRPYYYYGMDV 18
        DMGWGSGWRPYYYYGMDV
Sbjct: 99 DMGWGSGWRPYYYYGMDV 116
```

Use the STN Express 8.4 Patent Family Manager wizard display the results

Access the patent family manager wizard from the *Discover!* Menu.

Set Loginid Parameters...

Select a Database ▶

Search by ▶

Upload Sequence or Structure to STN...

Display by ▶

Analyze ▶

Create CAS RN® and Role Report... ▶

STN AnaVist ▶

Patent Family Manager...

Evaluate with STN Viewer... ▶

Get Related ▶

SDI ▶

Save ▶

Display Session Costs Ctrl+M

Logoff Your STN Session ▶

Discover! | Transcript: C:\Docume | STNGUIDE

Patent Family Manager Wizard

STN Express with Discover!

61 answers have been retrieved.

Extract the first member from each patent family (limit of 5,000 answers)

Include non-patent answers in result set.

Customize display of patent family results

Display format for first member of each patent family:
BIB PSL SQL SCORE ALIGN

Examples: bib abs
ti an

Display format for additional members of each patent family:
TRIAL SCORE ALIGN

Examples: bib abs
ti an

Insert a page break between each patent family display

< Back **Display** Finish

Choose a bibliographic display format with alignment for the first (best) hit, and a free-of-charge format with alignment for the rest of the sequences in each patent family group.

The patent family manager begins by organising the results using FSORT...

=> **FSORT L9**

. . . .

```
L10          61 FSO L9

              11 Multi-record Families      Answers 1-60
                Family 1                    Answers 1-18
                Family 2                    Answers 19-20
                Family 3                    Answers 21-22
                Family 4                    Answers 23-25
                Family 5                    Answers 26-27
                Family 6                    Answers 28-29
                Family 7                    Answers 30-34
                Family 8                    Answers 35-42
                Family 9                    Answers 43-56
                Family 10                   Answers 57-58
                Family 11                   Answers 59-60
                1 Individual Record         Answer 61
                0 Non-patent Records
```

Commands in **RED** are those issued automatically by the STN Express Patent Family Manager.

FSORT organizes the patent sequence records by Publication, Application, Related, and Priority numbers.

In this example, 12 patent family groups (i.e. 11 + 1) are retrieved.

...and then continues by displaying the family groups in the specified formats

=> DIS L10 PFAM=4 1 BIB,PSL,SQL,SCORE,ALIGN

```
L10 ANSWER 23 OF 61 DGENE COPYRIGHT 2010 THOMSON REUTERS on STN FAMILY4
AN AEC16143 protein DGENE
TI Treating tumorous disease, such as breast, colon, prostate, liver,
skin, ovarian, cervical and lung cancer, by administering a human
immunoglobulin specifically binding to the human epithelial cell
adhesion molecule (EpCAM) antigen.
IN Peters M; Locher M; Prang N; Quadt C
PA (MICR-N) MICROMET AG.
PI US 20050180979 A1 20050818
AI US 2004-778915 20040213
PRAI US 2004-778915 20040213
LA English
OS 2005-590351 [60]
DESC Human anti-EpCAM immunoglobulin heavy chain, SEQ ID NO: 1.
PSL Claim 12; SEQ ID NO 1
SQL 457
SCORE 70 100% of query self score 70
BLASTALIGN
Query = 18 letters
Length = 457
Score = 69.8 bits (157), Expect = 5e-18
Identities = 18/18 (100%), Positives = 18/18 (100%)
Query: 1 DMGWGSGWRPYYYYGMDV 18
DMGWGSGWRPYYYYGMDV
Sbjct: 99 DMGWGSGWRPYYYYGMDV 116
```

Commands in **RED** are those issued automatically by the STN Express Patent Family Manager.

...and then continues by displaying selected results in the specified formats (cont.)

=> DIS L10 PFAM=4 2-TOT TRIAL,SCORE,ALIGN

```
L10 ANSWER 24 OF 61 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN FAMILY4
TI Anti-EpCam Immunoglobulins (PublishedApplication)
DESC Artificial Protein; Anti-EpCAM Heavy Chain; sequence 1 of 2
MTY Protein
SQL 457
SCORE 70 100% of query self score 70
BLASTALIGN
  Query = 18 letters
  Length = 457
  Score = 69.8 bits (157), Expect = 5e-18
  Identities = 18/18 (100%), Positives = 18/18
Query: 1 DMGWGSGWRPYYYYGMDV 18
        DMGWGSGWRPYYYYGMDV
Sbjct: 99 DMGWGSGWRPYYYYGMDV 116
```

These two USGENE hits are in the same family as the DGENE record on the previous slide (**FAMILY 4**).

```
L10 ANSWER 25 OF 61 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN FAMILY4
TI Anti-EpCAM immunoglobulins (PublishedApplication)
MTY Protein
SQL 457
SCORE 70 100% of query self score 70
BLASTALIGN
  Query = 18 letters
  Length = 457
  Score = 69.8 bits (157), Expect = 5e-18
  Identities = 18/18 (100%), Positives = 18/18 (100%)
Query: 1 DMGWGSGWRPYYYYGMDV 18
        DMGWGSGWRPYYYYGMDV
Sbjct: 99 DMGWGSGWRPYYYYGMDV 116
```

...and then continues by displaying selected results in the specified formats (cont.)

=> DIS L10 61 BIB,PSL,SQL,SCORE,ALIGN

```
L10 ANSWER 61 OF 61 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN
AN 20070081993.104 Peptide USGENE
TI Pharmaceutical composition comprising a bispecific antibody for epcam
(PublishedApplication)
IN Kufer Peter (Moosburg, DE); Berry Meera (Ulm, DE); Offner Sonja
(Munich, DE); Brischwein Klaus (Munich, DE); Wolf Andreas (Planegg,
DE)
PA NO ASSIGNEE AT PUBLICATION
PI US 20070081993 A1 20070412
AI US 2004-554851 20040526
RLI WO 2004-EP5687 20040526
DT Patent
SQL 18
SCORE 70 100% of query self score 70
BLASTALIGN
Query = 18 letters
Length = 18
Score = 69.8 bits (157), Expect = 3e-19
Identities = 18/18 (100%), Positives = 18/18 (100%)
Query: 1 DMGWGSGWRPYYYYGMDV 18
DMGWGSGWRPYYYYGMDV
Sbjct: 1 DMGWGSGWRPYYYYGMDV 18
```

This USGENE record is the “individual record” in the FSORT answer set (L10).

Display additional CPlus answers including the HITRN for alignment post-processing

```
=> FILE HCAPLUS
```

```
=> TRA L10 PN
```

```
L11          TRANSFER L10 1- PN :          22 TERMS  
L12          16 L11  
ALL TERMS IN L11 RETRIEVED.
```

Transfer Publication Numbers (PN) from DGENE, USGENE and PCTGEN (L10) to find corresponding HCAplus records (L12).

```
=> S L4
```

```
L13          14 L4
```

The 36 REGISTRY records (L4) correspond to 14 HCAplus records (L13).

```
=> S L13 NOT L12
```

```
L14          1 L13 NOT L12
```

In this example, one additional relevant reference has been found by including a REGISTRY/CPlus search (L14).

```
=> D L14 BIB ABS HITRN
```

Display additional CPlus answers including the HITRN for alignment post-processing (cont.)

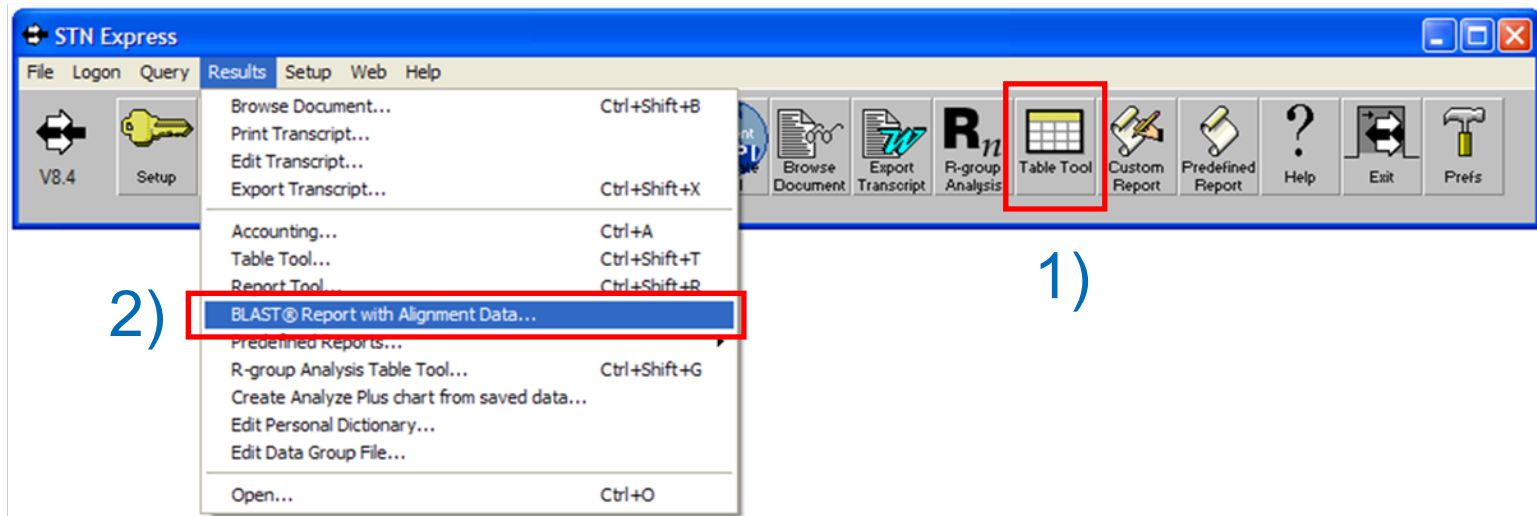
```

L14 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2010 ACS on STN
AN 2008:1210508 HCAPLUS
DN 149:446284
TI Cross-species bispecific single-chain antibodies to human . . . .
IN Ebert, Evelyn; Meier, Petra; Sriskandarajah, Mirnaalini; . . . .
PA Micromet A.-G., Germany
SO PCT Int. Appl., 397pp.
CODEN: PIXXD2
DT Patent
LA English
FAN.CNT 3
PATENT NO. KIND DATE APPLICATION NO. DATE
-----
PI WO 2008119565 A2 20081009 WO 2008-EP2662 20080403
. . . .
AB The present invention provides polypeptides comprising an antibody
binding domain capable of binding to an epitope of human . . . .
IT 1067695-29-7 1067695-33-3 1067695-35-5
1067695-37-7 1067695-39-9
RL: PRP (Properties); THU (Therapeutic
USES (Uses)
(amino acid sequence; cross-species
antibodies to human or non-chimpanz
antigen for treating tumorous, prol
and cancer)

```

Note: HITRN must be included, so that the CAS REGISTRY BLAST alignments can be merged into the BLAST Report.

STN Express post-processing tools provide the finishing touches to the multi-file search



- 1) DGENE, USGENE and PCTGEN results (L10) can be conveniently tabulated using the STN Express Table Tool and exported to a Microsoft Excel worksheet
- 2) The REGISTRY “BLAST Report with Alignment data” tool merges BLAST alignments with corresponding unique CPlus records (L14) to form a single RTF file

1) DGENE, USGENE and PCTGEN results can be tabulated and exported to Excel

Table Tool

All transcripts listed in the box below will be merged into your report or table. Select transcripts to add to this list by clicking the 'Browse to Add' button and locating the transcripts of interest.

Selected Transcripts :

Austin\My Documents\STN Express 8.4\Trnsript\EPCAMCDR2.trn

Table Tool

Available Templates in C:\Documents and Settings\Rob Austin\My Documents\STN Express 8.4\Trnsript

Report Table All

blast v8.4.prf
sequence table.prf

Description

FIELDS:
Family Number : Arial : 10
Patent Title : Arial : 10
Sequence Description : Arial : 10
BLAST Alignment : Courier New : 8
BLAST Score : Arial : 10
Sequence Length : Arial : 10
Sequence Location : Arial : 10
Inventor : Arial : 10
Patent Assignee : Arial : 10
Individual PI : Arial : 10
Accession Number : Arial : 10

Preferred fields, fonts, labels, etc, can be saved as a *Template* for repeated re-use.

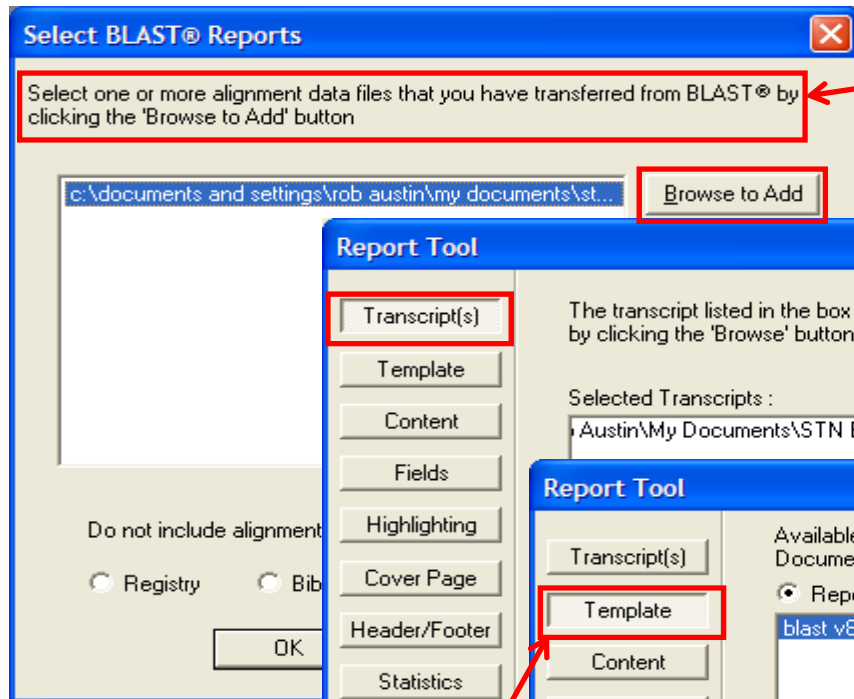
Once in Excel there are various options to sort, filter and review the multi-file results

A	B	C	D	E	F	G
Family Number	Patent Title	Sequence Description	BLAST Alignment	BLAST Score	Sequence Length	Sequence Location
1	FAMILY 4	Treating tumorous disease, such as breast, colon, prostate, liver, skin, ovarian, cervical and lung cancer, by administering a human immunoglobulin specifically binding to the human epithelial cell adhesion molecule (EpCAM) antigen.	Human anti-EpCAM immunoglobulin heavy chain, SEQ ID NO: 1.	70 100% of query self score 70	457	Claim 12; SEQ ID N
24	FAMILY 4	Anti-EpCam Immunoglobulins (PublishedApplication)	Artificial Protein; Anti-EpCAM Heavy Chain; sequence 1 of 2.	70 100% of query self score 70	457	
25	FAMILY 4	Anti-EpCAM immunoglobulins (PublishedApplication)				
26	FAMILY 5	New multifunctional compounds useful for preventing and/or treating malignant cell growth and for detection and diagnosis -	HD70scFv-Ck 2.			Example 10; Fig 55
27	FAMILY 5	New multifunctional compounds useful for	HD70scFv-CH1-GM-CSF chain.	70 100% of query self score 70	523	

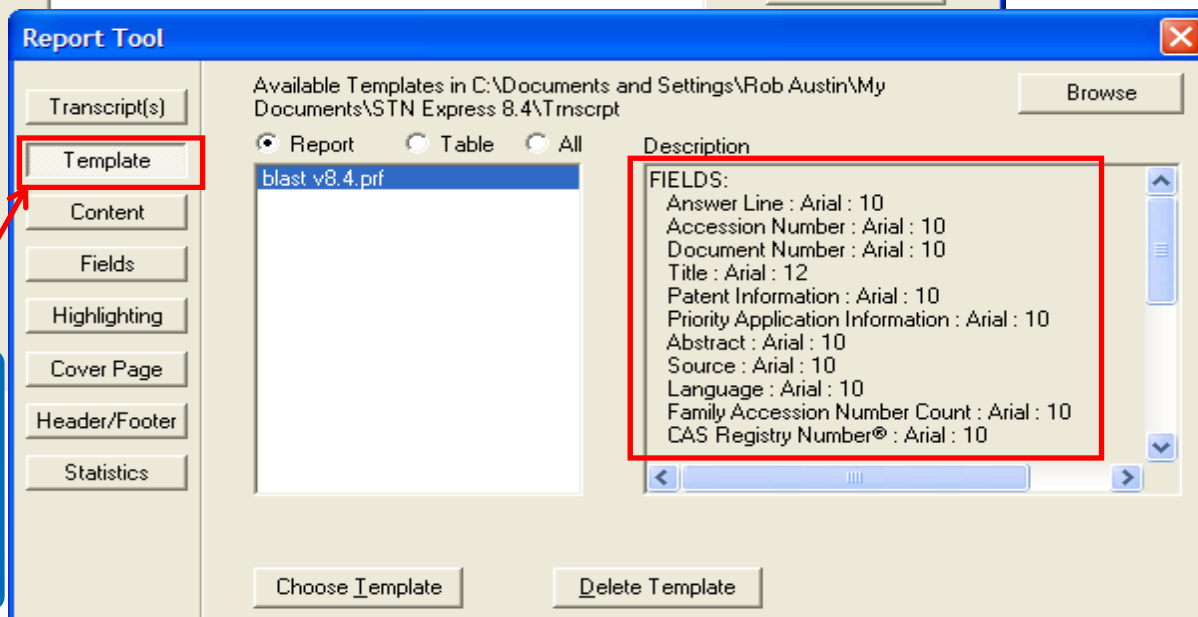
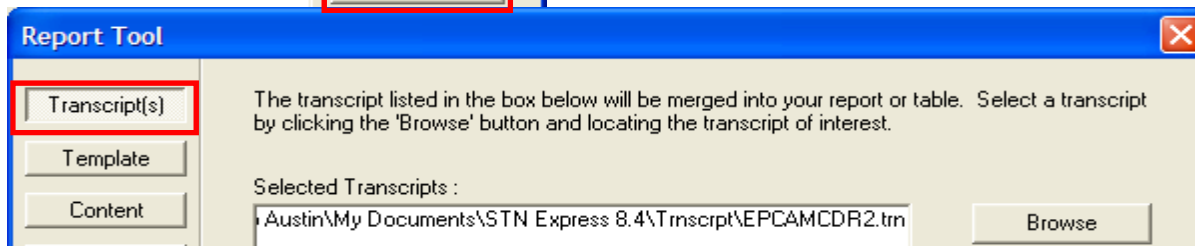
Some tips for Microsoft Excel:

- Resize columns and rows as desired – especially the BLAST alignment column to approx 77
- View, Freeze panes – holds the top row fixed when scrolling down
- Add Filters – provides a great way to navigate results – for example by BLAST score (above)

2) REGISTRY BLAST Alignments can be merged with corresponding CPlus records



REGISTRY BLAST alignments are merged with the CPlus records in the transcript.



Preferred fields, fonts, labels, etc, can be saved as a *Template* for repeated re-use.

The REGISTRY “BLAST Report with Alignment data” tool provides an RTF file

L14 ANSWER 1 OF 1 HCAPLUS C

Accession Number
2008:1210508 HCAPLUS Ful

Document Number
149:446284

Title
**Cross-species bispecific
non-chimpanzee primate
tumorous, proliferative, o**

Patent Information

PATENT NO.	KIND	DATE
WO 2008119565	A2	20081009
WO 2008119565	A3	20090108

Priority Application Information

Application Number	Kind	Date
EP 2007-6988	A	20070403
EP 2007-6990	A	20070403
US 2007-913668P	P	20070424
EP 2008-4741	A	20080313

Abstract
The present invention provides p
of binding to an epitope of huma
surface antigen selected from ep
receptor variant III, melanoma c
CD30, CD33, CD44 variant 6, Ep

Family Accession Number Count
3

Controlled or Index Terms and BLAST® Data
1067695-29-7 1067695-33-3 1067695-35-5
1067695-37-7 1067695-39-9
RL: PRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(amino acid sequence; cross-species bispecific single-chain antibodies to
human
or non-chimpanzee primate CD3e and surface antigen for treating tumorous,
proliferative, or immunol. disease and cancer)
1067695-29-7

Length = 127 Score = 69.8 Expect = e-11

Score = 69.8 Expect = 1e-11
Identities = 18/18 (100%) Positives = 18/18 (100%)

Query: 1 DMGWGSGWRPYYYYGMDV 18
DMGWGSGWRPYYYYGMDV
Subject: 99 DMGWGSGWRPYYYYGMDV 116

1067695-33-3

Length = 249 Score = 69.8 Expect = e-11

Score = 69.8 Expect = 1e-11
Identities = 18/18 (100%) Positives = 18/18 (100%)

Query: 1 DMGWGSGWRPYYYYGMDV 18
DMGWGSGWRPYYYYGMDV
Subject: 99 DMGWGSGWRPYYYYGMDV 116

Download the post-processing template (.PRF) files used in this seminar:
http://www.stn-international.com/stn_biosequence_searching_mfs.html

Patent classifications for antibody topics

<u>USCL</u> Class	Subclass	<u>IPC8</u> Subclass	Group	Topic
530	387.1+	C07K	16/00+	Immunoglobulins, Antibodies
530	388.1+	C12P	21/08+	Monoclonal Antibodies
530	391.1+	C07K	16/00+	Conjugated antibodies
424	1.49	A61K	49/00+	Compositions containing radio-labeled antibodies
424	130.1	A61K	39/395+	Compositions for body treatment containing antibodies (therapeutics, vaccines)
435	7.1+	G01N	33/53+	Immunoassays using antibodies
435	188	C12N	09/96+	Antibodies conjugated to enzymes
435	188.5	C07K	16/00+	Catalytic antibodies
435	325+	C12N	5/00+	Cells that express antibodies (fused, recombinant)
525	54.1	A61K	47/48+	Antibodies bound to resins

Derwent World Patents Index[®] (DWPISM) Manual Codes (/MC) for antibody topics

The screenshot shows a web browser window titled "Manual Code Lookup - Science - Thomson Reuters". The address bar contains the URL <http://scientific.thomson.com/cgi-bin/mc/search.cgi>. The page header reads "THOMSON REUTERS - DWPI MANUAL CODE LOOKUP".

Search details:
Enter terms to search:
Select boolean search type:
Number of codes matched: 121

Search results table:

CODE <i>Status</i>	TITLE
B04-B04C <i>Current</i>	Antigens, general antibody (pre-94)
B04-B04C3 <i>Retired</i>	Microbial antibody*
B04-B04C4 <i>Retired</i>	Anticancer antibody*
B04-B04C5 <i>Retired</i>	Monoclonal antibody*
B04-B04C6 <i>Retired</i>	Other antibody including immunoglobulin and haemagglutinin*
B04-B04C7 <i>Current</i>	Haptens
B04-G <i>Current</i>	Antibody defined in terms of antigen
B04-G01	Antibody defined in terms of antigen general and other

Details for B04-B04C7:

Enter code to look up:

- | B04 : Natural products (or genetically engineered), polymers
- | - B04-B : Animal, microbiological and general extracts
- | -- B04-B04 : Animal extract [general]*
- | --- B04-B04C : Antigens, general antibody (pre-94)

CODE	B04-B04C7
TITLE	Haptens
STATUS	Current
DATE	(1994-)
RELATED CODES	B04-B04C
SCOPE NOTES	A substance which can combine with antibody but cannot itself initiate an immune response unless it is attached to a carrier molecule.
SEARCH TERMS	

More than 120 Manual Codes (/MC) are available for antibody searching.

Summary

- DGENE provides detailed annotations and indexing for text searching for antibody technologies
- REGISTRY provides extensive annotations, and common, trade, generic, and lab antibody names
- CAS controlled and index terms are all useful for retrieving antibody information in CAplus
 - Use text terms to search for types of antibodies in CAplus
 - Class (α (IgA), δ (IgD), ε (IgE), γ (IgG), μ (IgM))
- BLAST and SCM searches are available for antibody sequence searching in DGENE, USGENE, PCTGEN and REGISTRY on STN

Resources for sequence searching on STN

- CAS REGISTRY sequence searching resources

<http://www.cas.org/support/stngen/stndoc/sequences.html>

- DGENE Workshop Manual

http://www.stn-international.com/dgene_wm.html

- USGENE Workshop Manual

http://www.stn-international.com/usgene_wm.html

- Multifile BLAST searching (step-by-step guide)

http://www.stn-international.com/usgene_wm_mfs.html

Recorded STN e-Seminars are available to watch at your own pace....

- FIZ Karlsruhe recorded e-Seminars:

http://www.stn-international.com/recorded_events.html

- Sequence Basics (all databases)
- Multifile patent sequence searching (all databases)

- CAS recorded e-Seminars:

<http://www.cas.org/support/stngen/stntraining/recorded.html>

- Sequence motif searching (all databases)
- Processing sequence data (REGISTRY)
- Unmasking the World of Antibodies (REGISTRY)

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