



STN[®] WORKSHOP MANUAL

© FIZ Karlsruhe, April 2010

Robert Austin

FIZ Karlsruhe
Hermann-von-Helmholtz-Platz 1
76344 Eggenstein-Leopoldshafen
Germany
Phone: +49 7247 808 555
Fax: +49 7247 808 259

www.fiz-karlsruhe.de

USGENE[®] is a registered trademark of [SequenceBase Corporation](#).

BLAST[®] is a registered trademark of the [U.S. National Library of Medicine](#).

MULTI-FILE BLAST SEARCHING

In this supplement, you will learn about

- STN sequence searchable databases
- Multi-file BLAST searching in USGENE, DGENE and PCTGEN
- Extending a multi-file search to include CAS REGISTRY BLAST
- Post-processing USGENE, DGENE and PCTGEN results
- Using the CAS REGISTRY BLAST Report Tool

Multi-file BLAST searching

The purpose of this document

Multi-file BLAST searching is a supplement to the [USGENE Workshop Manual](#). This document focuses on explaining how to extend a [USGENE](#) BLAST search to other sequence databases on STN, for the purposes of completing a comprehensive international patent sequence search.

note

USGENE STN Workshop Manual:

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

STN sequence searchable databases

There are four databases on STN which support BLAST sequence searching. Each of these databases provides unique content, and all should be used for a comprehensive sequence search.

DGENE Thomson Reuters GENESEQ™ value-added worldwide patent sequence data

USGENE® United States Patent and Trademark Office (USPTO) sequence data

PCTGEN World Intellectual Property Organization (WIPO) e-published sequence data

REGISTRY CAS REGISTRYSM value-added worldwide patent and non-patent sequence data

The three patent sequence databases, **DGENE**, **USGENE** and **PCTGEN**, are all searchable using the same command line based system, either using [STN Express](#) or [STN on the Web](#). Guidance on effective command line BLAST searching may be found at the [STN International website](#).

In contrast, the **REGISTRY** file does not support any form of command line BLAST searching. Instead, an independent, standalone, graphic-user-interface platform is available, known as **CAS REGISTRY BLAST**. This BLAST search interface, is accessible either from the main window of STN Express, or from the Sequence Search Assistant in STN on the Web. Details of how to use CAS REGISTRY BLAST are provided in [Chapter 10 of the STN Express User Guide](#)¹.

Helpful HINT

Sequence Searching on STN Workshop:

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

¹ STN Express User Guide: http://www.stn-international.com/user_guide.html

Protein sequence BLAST search example

Search Question: Find all patents that disclose the following Epithelial cell adhesion molecule (EpCAM) antibody CDR sequence, or similar sequences

DMGWGSGWRPYYYYGMDV

Search Strategy

To conduct a multi-file search for sequences using BLAST...

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

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

Initiate BLAST protein (/SQP) offline BATCH search in USGENE

```
=> FILE USGENE
FILE 'USGENE' ENTERED AT 11:47:21 ON 16 APR 2010
COPYRIGHT (C) 2010 SEQUENCEBASE CORP

=> RUN BLAST DMGWGSGWRPYYYYGMDV/SQP -F F -E 1000 -W 2 -M PAM30 BATCH

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

BLAST Version 2.2
. . . .
BATCH PROCESSING STARTED FOR EPCAMCDR
```

Add **BATCH** to the end of a RUN BLAST command to search in offline batch search mode.

Initiate BLAST protein (/SQP) offline BATCH search in DGENE

```
=> FILE DGENE
FILE 'DGENE' ENTERED AT 11:47:41 ON 16 APR 2010
COPYRIGHT (C) 2010 THOMSON REUTERS

=> RUN BLAST DMGWGSGWRPYYYYGMDV/SQP -F F -E 1000 -W 2 -M PAM30 BATCH

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

BLAST Version 2.2
. . . .

BATCH PROCESSING STARTED FOR EPCAMCDR
```

Each BLAST search is adjusted with the NCBI recommended advanced options for short queries.

Initiate BLAST protein (/SQP) offline BATCH search in PCTGEN

```
=> FILE PCTGEN
FILE 'PCTGEN' ENTERED AT 11:48:02 ON 16 APR 2010
COPYRIGHT (C) 2010 WIPO

=> RUN BLAST DMGWGSGWRPYYYYGMDV/SQP -F F -E 1000 -W 2 -M PAM30 BATCH

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

BLAST Version 2.2
. . . .

BATCH PROCESSING STARTED FOR EPCAMCDR
```

Each offline BATCH search runs in parallel.

“Park” in STNGUIDE to wait for each of the BATCH searches to finish

```
=> FILE STNGUIDE
FILE 'STNGUIDE' ENTERED AT 11:48:03 ON 16 APR 2010
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 2010 AMERICAN CHEMICAL SOCIETY (ACS)

FILE CONTAINS CURRENT INFORMATION.
LAST RELOADED: Apr 16, 2010 (20100416/UP).
```

Helpful HINT

There are a few alternatives to simply waiting in STNGUIDE, for example:

- (a) LOGOFF and log back in again later to retrieve the results.*
 - (b) Begin the CAS REGISTRY BLAST search in parallel (page xvi).*
-

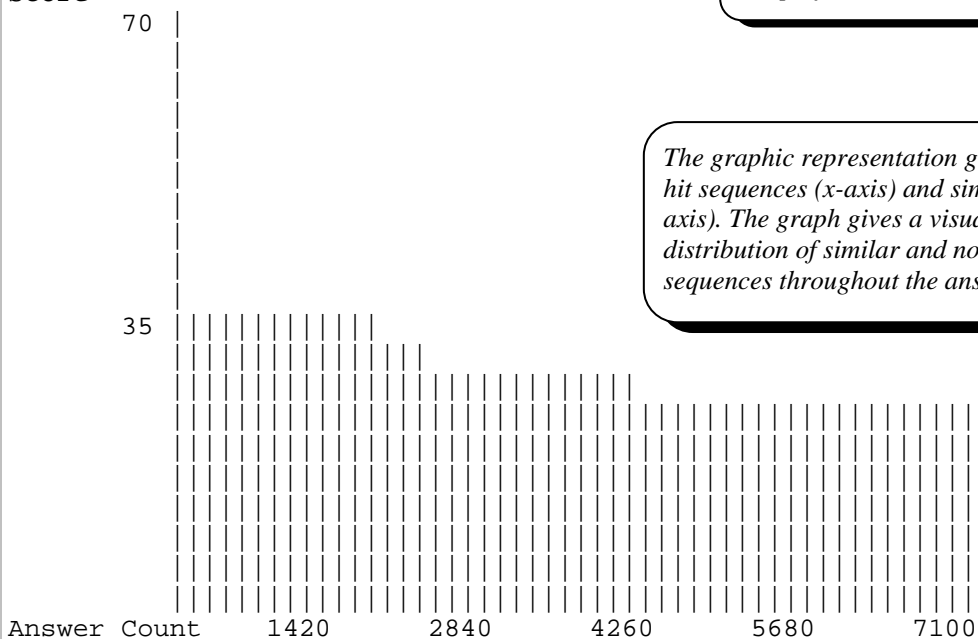
Retrieve the USGENE offline BATCH search results=> **FILE USGENE**FILE 'USGENE' ENTERED AT 11:57:27 ON 16 APR 2010
COPYRIGHT (C) 2010 SEQUENCEBASE CORPUse **RUN GETBATCH**
to retrieve the results.=> **RUN GETBATCH EPCAMCDR**Please enter your batch identifier
or enter # for batch id list
or enter * for batch id at top of list
or enter - before batch id to delete
or enter . for (end)

Database USGENE AA

Posted date: Jan 15, 2010 7:01 PM

.

7054 ANSWERS FOUND BELOW EXPECTATION VALUE

QUERY SELF SCORE VALUE IS 70
BEST ANSWER SCORE VALUE IS 70The **Query Self Score** value is the ideal
score for a perfect answer match.The **Best Answer Score** value is also
given. In this example, there is at least
one perfect answer match.Similarity
ScoreThe graphic representation gives a count of
hit sequences (x-axis) and similarity score (y-
axis). The graph gives a visual clue about the
distribution of similar and not so similar
sequences throughout the answer set.ENTER EITHER THE NUMBER OF ANSWERS YOU WISH TO KEEP
OR ENTER MINIMUM PERCENT OF SELF SCORE FOLLOWED BY %
(BEST ANSWER PERCENTAGE OF SELF SCORE IS 100%)

ENTER (ALL) OR ? :60%

L1 RUN STATEMENT CREATED
L1 20 DMGWGSGWRPYYIn this example, 60% of the **Query Self Score** is used to
select out just the most relevant results (**L1**).Answer set arranged by accession number; to sort by descending
similarity score, enter at an arrow prompt (=>) "sor score d".

Retrieve the DGENE offline BATCH search results

```
=> FILE DGENE
FILE 'DGENE' ENTERED AT 11:57:42 ON 16 APR 2010
COPYRIGHT (C) 2010 THOMSON REUTERS

=> RUN GETBATCH EPCAMCDR

. . . . .

ENTER EITHER THE NUMBER OF ANSWERS YOU WISH TO KEEP
OR ENTER MINIMUM PERCENT OF SELF SCORE FOLLOWED BY %
(BEST ANSWER PERCENTAGE OF SELF SCORE IS 100%)
ENTER (ALL) OR ? :60%
L2 RUN STATEMENT CREATED
L2 32 DMGWGSGWRPYYYYGMDV/SQP.-F F -E 1000 -W 2 -M PAM30

Answer set arranged by accession number; to sort by descending
similarity score, enter at an arrow prompt (=>) "sor score d".
```

Use RUN GETBATCH to retrieve the results.

Retrieve the PCTGEN offline BATCH search results

```
=> FILE PCTGEN
FILE 'PCTGEN' ENTERED AT 11:58:03 ON 16 APR 2010
COPYRIGHT (C) 2010 WIPO

=> RUN GETBATCH EPCAMCDR

. . . . .

ENTER EITHER THE NUMBER OF ANSWERS YOU WISH TO KEEP
OR ENTER MINIMUM PERCENT OF SELF SCORE FOLLOWED BY %
(BEST ANSWER PERCENTAGE OF SELF SCORE IS 100%)
ENTER (ALL) OR ? :60%
L3 RUN STATEMENT CREATED
L3 12 DMGWGSGWRPYYYYGMDV/SQP.-F F -E 1000 -W 2 -M PAM30

Answer set arranged by accession number; to sort by descending
similarity score, enter at an arrow prompt (=>) "sor score d".

=>
```

Use RUN GETBATCH to retrieve the results.

note



- After being retrieved, BATCH results remain available for 8 days.
- During this period a BATCH result may be retrieved repeatedly.
- After 8 days BATCH results are deleted from the system.

Merge, organize by patent family, and display USGENE, DGENE and PCTGEN results

Merge the USGENE, DGENE and PCTGEN results together into a single L-number

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

*SET DUPORDER FILE PERM ensures that multifile records merged into a single L-number using **DUP IDE**, are organized by database (file). Add **PERM** to make this a permanent setting.*

```
=> DUP IDE L1 L2 L3
```

```
DUPLICATE IS NOT AVAILABLE IN 'USGENE, DGENE, PCTGEN'.
ANSWERS FROM THESE FILES WILL BE CONSIDERED UNIQUE
. . . . .
```

*DUPLICATE IDENTIFY (DUP IDE) creates a single L-number (**L4**) by merging all the results together. This is useful for any subsequent patent family sort (**FSORT**) and Table Tool post-processing steps.*

```
L4          64 DUP IDE L1 L2 L3 (INCLUDES 0 SETS OF DUPLICATES)
           ANSWERS '1-20' FROM FILE USGENE
           ANSWERS '21-52' FROM FILE DGENE
           ANSWERS '53-64' FROM FILE PCTGEN
```

Sort the combined USGENE, DGENE and PCTGEN results by descending identity percentage

```
=> SOR L4 IDENT D
```

```
PROCESSING COMPLETED FOR L4
L5          64 SOR L4 IDENT D
```

*The multi-file, merged answer set (**L4**) may be sorted like any standard single-file L-number, including by **IDENT** or **SCORE**.*

Prior to display ensure Capture Session is on to record a transcript for use in post-processing

The screenshot shows the STN Online and Results interface. The main window displays a transcript of commands and their outputs. A 'Capture Session' dialog box is open, allowing the user to save the current session. The dialog box shows the file name 'EPCAM CDR MULTIFILE SEARCH 2' and the 'Files of type' set to 'Transcript Files (*.tm)'. The 'Capture retrospectively' checkbox is checked. A 'Select Discover! Wizard' window is also visible in the background, showing a search history table.

Search history	Count	Details
L1	20	# DMGWGSGWRPYYYYGMDV/:
L2	32	# DMGWGSGWRPYYYYGMDV/:
L3	12	# DMGWGSGWRPYYYYGMDV/:
L4	64	DUP IDE L1 L2 L3
L5	64	SOR IDENT D

Note: check the *Capture Retrospectively* box, to capture the STN session so far, as well as the session from this point forwards.

Display multi-file answers retrieved in a free-of-charge format, including alignment (ALIGN)**=> D TRI SCORE IDENT ALIGN 1-64**

L5 ANSWER 1 OF 64 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN
TI Antibodies that bind human 17-A1/EpCAM tumor antigen (Patent)
DESC Artificial Protein; Synthetic Peptide; sequence 144 of 169
MTY Protein
SQL 138
SCORE 70 100% of query self score 70
IDENT 100%
BLASTALIGN

Query = 18 letters
Length = 138
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
.

L5 ANSWER 13 OF 64 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
IDENT 100%
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
.

L5 ANSWER 33 OF 64 PCTGEN COPYRIGHT 2010 WIPO on STN
TI CROSS-SPECIES-SPECIFIC BINDING DOMAIN
MTY PRT
SQL 504
SCORE 70 100% of query self score 70
IDENT 100%
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
.

```

L5    ANSWER 49 OF 64  PCTGEN COPYRIGHT 2010 WIPO on STN
TI    ANTAGONISTS TO IL-17A, IL-17F, ANDIL-23P19 AND METHODS OF USE
MTY   PRT
SQL   125
SCORE 49          70% of query self score 70
IDENT 92%
BLASTALIGN
  Query = 18 letters
  Length = 125
  Score = 48.6 bits (107), Expect = 4e-12
  Identities = 13/14 (92%), Positives = 13/14 (92%)
Query: 5   GSGWRPYYYYGMDV 18
          GSGWR YYYYYGMDV
Sbjct: 101 GSGWRRYYYYGMDV 114

L5    ANSWER 50 OF 64  USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN
TI    IL-18 binding proteins (PublishedApplication)
MTY   Protein
SQL   127
SCORE 44          62% of query self score 70
IDENT 86%
BLASTALIGN
  Query = 18 letters
  Length = 127
  Score = 44.3 bits (97), Expect = 8e-11
  Identities = 13/15 (86%), Positives = 13/15 (86%), Gaps = 1/15 (6%)
Query: 5   GSGWRP-YYYYGMDV 18
          GSGW P YYYYYGMDV
Sbjct: 102 GSGWPPFYYYYGMDV 116

. . . . .

```

Option: after reviewing the BLAST results choose answers to keep for further processing

```

=> SOR 1-49 IDENT D
PROCESSING COMPLETED FOR L5
L6          49 SOR L5 1-49 IDENT D

```

To keep selected records (**L6**) from the BLAST answer set (**L5**), repeat the SORT command on the chosen records (**1 - 49**).

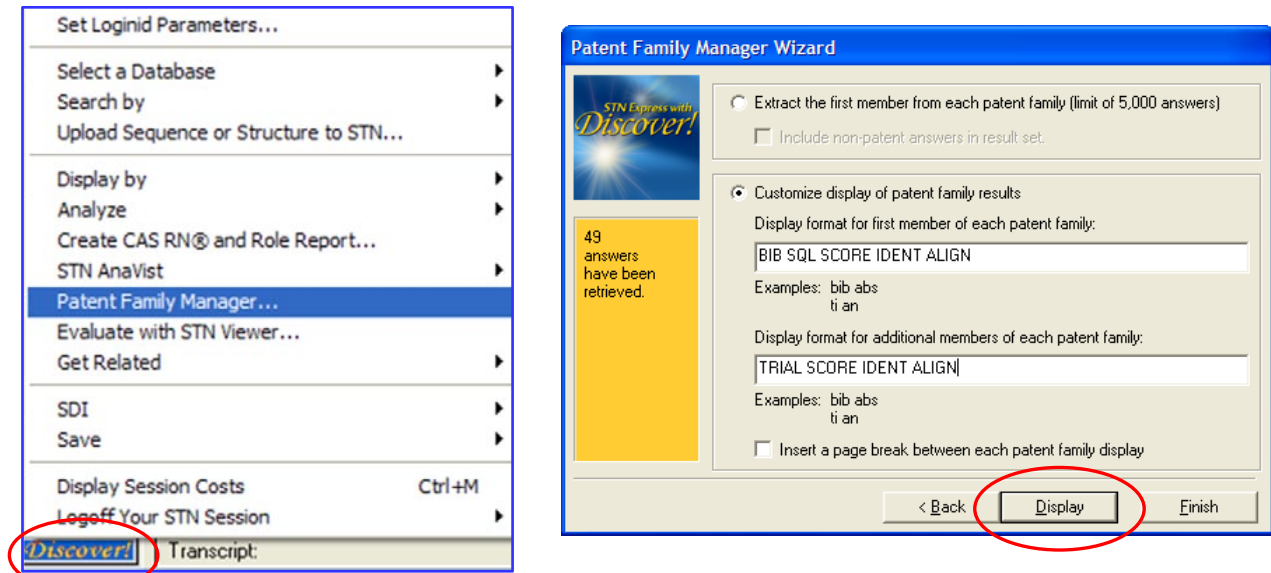
Helpful HINT

Learn more about sorting BLAST results by percent identity:
http://www.stn-international.com/percent_identity_sorting.html

Option: Use the Patent Family Manager to group and display multi-file BLAST search results

As described detail on page 56 of the USGENE Workshop Manual, the Patent Family Manager may be used to automatically issue appropriate FSORT and D PFAM commands on STN.

1. Click on the *Discover!* button and select the *Patent Family Manager*.
2. Select the second option, and enter preferred formats for the first and the additional members of the family.



Review the commands issued and results displayed by the Patent Family Manager

```

=> FIL USGENE, DGENE, PCTGEN

FILE 'USGENE' ENTERED AT 12:02:06 ON 16 APR 2010
COPYRIGHT (C) 2010 SEQUENCEBASE CORP

FILE 'DGENE' ENTERED AT 12:02:06 ON 16 APR 2010
COPYRIGHT (C) 2010 THOMSON REUTERS

FILE 'PCTGEN' ENTERED AT 12:02:06 ON 16 APR 2010
COPYRIGHT (C) 2010 WIPO

=> FSORT L6
. . . . .

L7          49 FSO L6

          9 Multi-record Families   Answers 1-48
            Family 1                 Answers 1-6
            Family 2                 Answers 7-10
            Family 3                 Answers 11-12
            Family 4                 Answers 13-14
            Family 5                 Answers 15-17
            Family 6                 Answers 18-35
            Family 7                 Answers 36-37
            Family 8                 Answers 38-44
            Family 9                 Answers 45-48
          1 Individual Record       Answer 49
          0 Non-patent Records

. . . . .
    
```

The STN Express **Patent Family Manager** issues an **FSORT** command, and a series of **PFAM** display commands to display the requested information automatically.

FSORT (L7) of the 49 BLAST answers (**L6**) reveals 10 patent family groups: 9 with multiple sequence answers for the search, and 1 with just one sequence hit.

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

=> DIS L7 PFAM=1 1 BIB,SQL,SCORE,IDENT,ALIGN

L7 ANSWER 1 OF 49 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN **FAMILY 1**
AN 7632925.144 Protein USGENE [Full-text](#)
TI Antibodies that bind human 17-A1/EpCAM tumor antigen (Patent)
IN Kufer Peter (Moosburg, DE); Raum Tobias (Munich, DE)
PA Micromet AG (Munich DE)
PI US 7632925 B2 20091215
US 20090081191 A1 20090326
AI US 2007-860242 20070924
RLI US 2002-325694 20021219
US 1998-403107 19980414
WO 1998-EP2180 19980414
PRAI EP 1997-106109 19970414
XPD 20180414 (calculated)
NTE Subject to any Disclaimer, the term of this patent is extended or
adjusted under 35 USC 154(b) by 165 days.
PSL SEQ ID NO 144
DESC Artificial Protein; Synthetic Peptide; sequence 144 of 169
DT Patent
SQL 138
SCORE 70 100% of query self score 70
IDENT 100%
BLASTALIGN
Query = 18 letters
Length = 138
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

U.S. issued patent (USB2) sequence records typically include preceding publication information (PI).

=> DIS L7 PFAM=1 2-TOT TRIAL,SCORE,IDENT,ALIGN

L7 ANSWER 2 OF 49 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN **FAMILY 1**
TI Antibodies that bind human 17-A1/EpCAM tumor antigen
(PublishedApplication)
DESC Artificial Protein; Synthetic Peptide; sequence 144 of 169
MTY Protein
SQL 138
SCORE 70 100% of query self score 70
IDENT 100%
BLASTALIGN
Query = 18 letters
Length = 138
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

. . . .

=> DIS L7 PFAM=4 1 BIB,SQL,SCORE,IDENT,ALIGN

L7 ANSWER 13 OF 49 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN **FAMILY4**
AN 7435549.54 Protein USGENE [Full-text](#)
TI Method of identifying binding site domains that retain the capacity of
binding to an epitope (Patent)
IN Kufer Peter (Moosburg, DE); Raum Tobias (Munich, DE); Borschert Katrin
(Munich, DE); Zettl Florian (Kempten, DE); Lutterbuse Ralf (Munich, DE)
PA Micromet AG (Munich DE)
PI US 7435549 B1 20081014
WO 9925818 A 19990527
AI US 1998-554465 19981116
RLI WO 1998-EP7313 19981116
PRAI EP 1997-120096 19971117
XPD 20181116 (calculated)
PSL SEQ ID NO 54
DESC Homo Sapiens Protein; sequence 54 of 77
DT Patent
SQL 127
SCORE 70 100% of query self score 70
IDENT 100%
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

=> DIS L7 PFAM=4 2-TOT TRIAL,SCORE,IDENT,ALIGN

L7 ANSWER 14 OF 49 DGENE COPYRIGHT 2010 THOMSON REUTERS on STN **FAMILY 4**
AN AAY17954 Protein DGENE
TI Phage display system for identification of binding site domains retaining
capacity to bind an epitope
DESC Human D4.5 heavy chain variable region.
KW Binding site domain; BSD; epitope; fusion protein; therapeutic; cancer;
autoimmune disease; scFv-antibody; single-chain Fv.
SQL 127
SCORE 70 100% of query self score 70
IDENT 100%
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

.

=> **DIS L7 49 BIB,SQL,SCORE,IDENT,ALIGN**

```
L7 ANSWER 49 OF 49 USGENE COPYRIGHT 2010 SEQUENCEBASE CORP on STN
AN 20070081993.104 Peptide USGENE Full-text
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
IDENT 100%
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
```

The BLAST local identity percentage (IDENT) may be included in the display, for use as a column by the Table Tool.

“Park” in STNGUIDE before moving on to the CAS REGISTRY BLAST search (next page...)

=> **FILE STNGUIDE**

```
FILE 'STNGUIDE' ENTERED AT 12:09:55 ON 16 APR 2010
USE IS SUBJECT TO THE TERMS OF YOUR CUSTOMER AGREEMENT
COPYRIGHT (C) 2010 AMERICAN CHEMICAL SOCIETY (ACS)
```

```
FILE CONTAINS CURRENT INFORMATION.
LAST RELOADED: Apr 16, 2010 (20100416/UP).
```

=>

Helpful HINT

Learn more about the STN Express Patent Family Manager:
http://www.stn-international.com/stn_express_pat_fam_manage.html

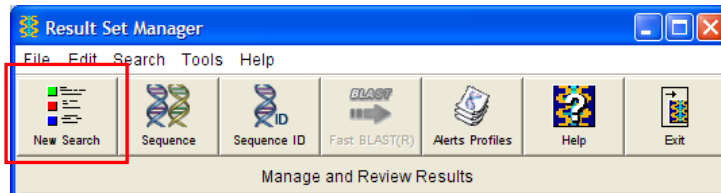
Repeat the search using CAS REGISTRY BLAST

Run the CAS REGISTRY BLAST search

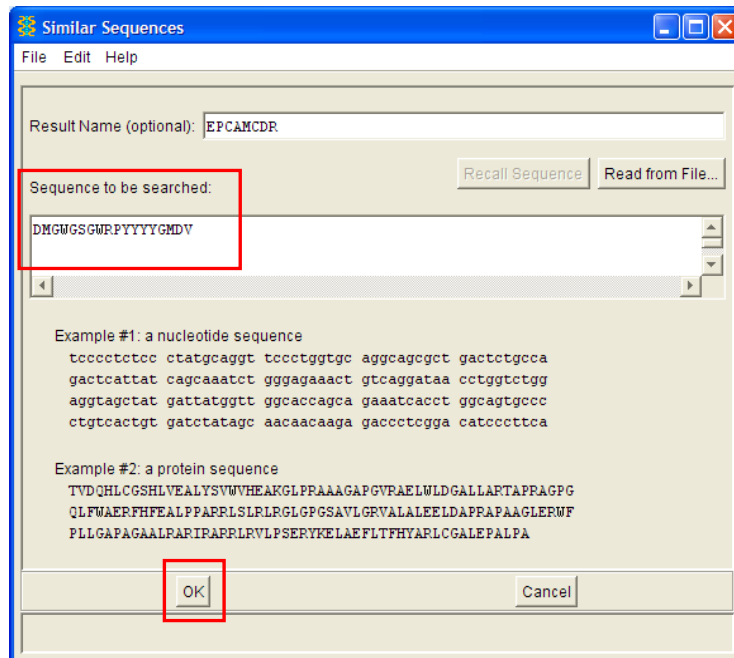
- Launch and log in to REGISTRY BLAST



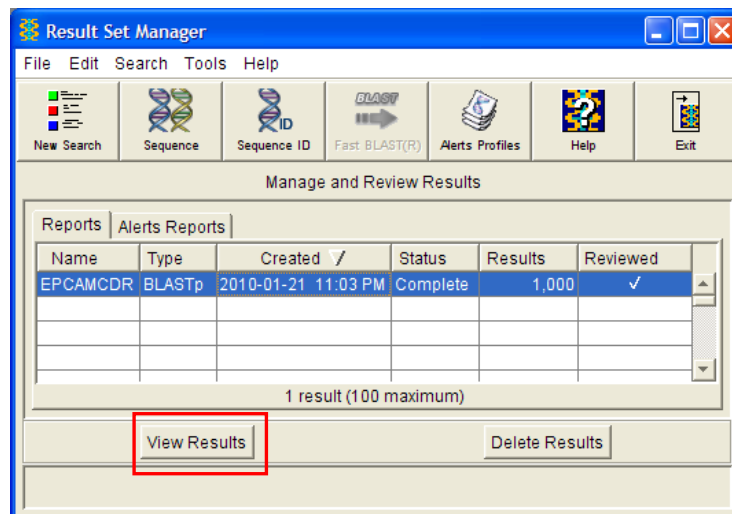
- To run a BLAST search *Select New Search* and then *Similar Sequences*



- Submit the query sequence and Provide a result name (optional)



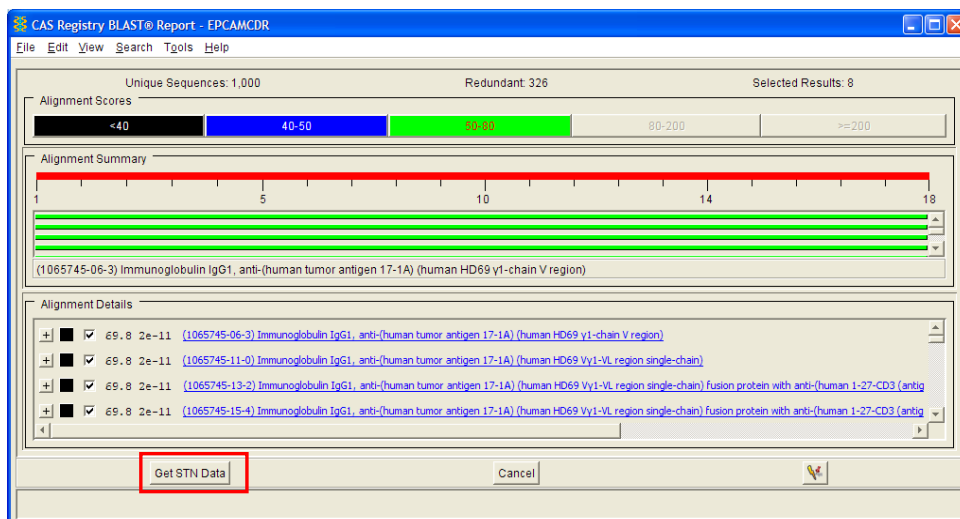
- Select the results name and click *View Results*



- (5) Review and select answers of interest

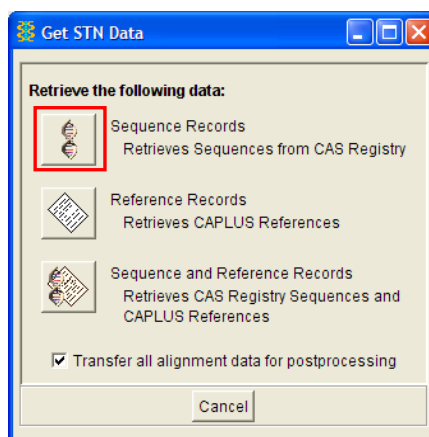
In this example, all records with 70% or more of *Query Self Score* are selected

When ready, click on *Get STN Data*, to transfer the results to STN



- (6) Choose the *Sequence Records* option to retrieve REGISTRY records on STN.

When prompted, save the BLAST alignment data (.XSS file)



- (7) The REGISTRY BLAST wizard retrieves the REGISTRY records on STN.

```

=> FIL REGISTRY
FILE 'REGISTRY' ENTERED AT 15:43:37 ON 22 JAN 2010
COPYRIGHT (C) 2010 American Chemical Society (ACS)

=> QUE (1065745-06-3 OR 1067695-29-7 OR . . . )/RN

```

Helpful HINT

Learn more about CAS REGISTRY BLAST:

<http://www.cas.org/support/stnexp/seehow/expressblast.html>

CAS REGISTRY BLAST automatically retrieves the selected REGISTRY records on STN=> **FIL REGISTRY**FILE 'REGISTRY' ENTERED AT 13:23:01 ON 16 APR 2010
COPYRIGHT (C) 2010 American Chemical Society (ACS)*Commands in **RED** are those issued automatically by STN Express.*=> **QUE (1065745-06-3 OR 1067695-29-7 OR 1065745-11-0 OR 1067695-33-3 OR 1065745-13-2 OR 1067695-35-5 OR 1065745-15-4 OR 1067695-37-7 OR 1065745-17-6 OR 1067695-39-9 OR 942165-85-7 OR 945905-53-3 OR 945940-54-5)/RN**

L8 QUE (1065745-06-3 OR 1067695-29-7 OR 1065745-11-0 OR 1067695-33-3 OR 1065745-13-2 OR 1067695-35-5 OR 1065745-15-4 OR 1067695-37-7 OR 1065745-17-6 OR 1067695-39-9 OR 942165-85-7 OR 945905-53-3 OR 945940-54-5)/RN

=> **QUE (862861-57-2 OR 487483-96-5 OR 215027-98-8 OR 481333-75-9 OR 258495-00-0 OR 481333-76-0 OR 258495-02-2 OR 1163341-05-6 OR 1089241-79-1 OR 1074004-53-7 OR 960551-07-9 OR 960523-29-9)/RN**

L9 QUE (862861-57-2 OR 487483-96-5 OR 215027-98-8 OR 481333-75-9 OR 258495-00-0 OR 481333-76-0 OR 258495-02-2 OR 1163341-05-6 OR 1089241-79-1 OR 1074004-53-7 OR 960551-07-9 OR 960523-29-9)/RN

=> **S L8 OR L9**

L10 25 L8 OR L9

=> Enter display format:d 1-25 **END**

Enter Data Here

END

Enter

*Tip: type **END** to exit the STN Express Wizard.***Retrieve, identify and display unique CAS REGISTRY BLAST CAplus records****Retrieve the corresponding CAplus patent records**=> **FILE HCAPLUS**FILE 'HCAPLUS' ENTERED AT 13:23:58 ON 16 APR 2010
COPYRIGHT (C) 2010 AMERICAN CHEMICAL SOCIETY (ACS)*The 25 REGISTRY sequence records (**L10**) correspond to 12 CAplus patent records (**L11**).*=> **S L10 AND PATENT/DT**

L11 12 L10 AND PATENT/DT

Identify the unique CAplus records for post-processing=> **TRANSFER L7 PN 1-**L12 TRANSFER L7 1- PN : 22 TERMS
L13 14 L12
ALL TERMS IN L12 RETRIEVED.*The 49 USGENE, DGENE and PCTGEN records (**L7**) correspond to 14 CAplus patent records (**L13**).*=> **S L11 NOT L13**

L14 1 L11 NOT L13

*One (**L14**) of the 12 CAplus records (**L11**) retrieved using CAS REGISTRY BLAST is unique, compared to USGENE, DGENE and PCTGEN (**L13**).*

Display additional CPlus answers including Hit Registry Numbers (HITRN)=> **D BIB HITRN**

L14 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2010 ACS on STN

AN 2008:1210508 HCAPLUS [Full-text](#)

DN 149:446284

TI Cross-species bispecific single-chain antibodies to human or non-chimpanzee primate CD3ε and surface antigen for treating tumorous, proliferative, or immunological disease and cancer

IN Ebert, Evelyn; Meier, Petra; Sriskandarajah, Mirnaalini; Burghart, Elke; Wissing, Sandra; Klinger, Matthias; Bluemel, Claudia; Raum, Tobias; Rau, Doris; Mangold, Susanne; Kvesic, Majk; Kischel, Roman; Steiger, Carola; Lutterbuese, Ralf; Mayer, Petra; Schaller, Evelyne; Hoffmann, Patrick; Strasser, Susanne; Cierpka, Ronny; Kufer, Peter; Hausmann, Susanne; Riethmueller, Gert

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
	WO 2008119565	A3	20090108		

W: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

RW: AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG, BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM, AP, EA, EP, OA

PRAI EP 2007-6988 A 20070403
 EP 2007-6990 A 20070403
 US 2007-913668P P 20070424
 EP 2008-4741 A 20080313

IT 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 CD3ε and surface antigen for treating tumorous, proliferative, or immunol. disease and cancer)

=> **LOG H**

SESSION WILL BE HELD FOR 120 MINUTES

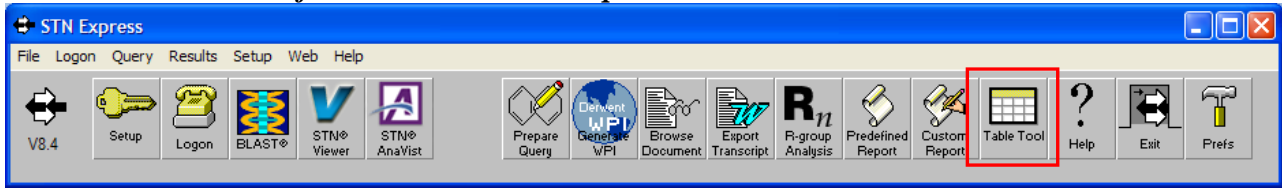
STN INTERNATIONAL SESSION SUSPENDED AT 13:26:52 ON 16 APR 2010

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

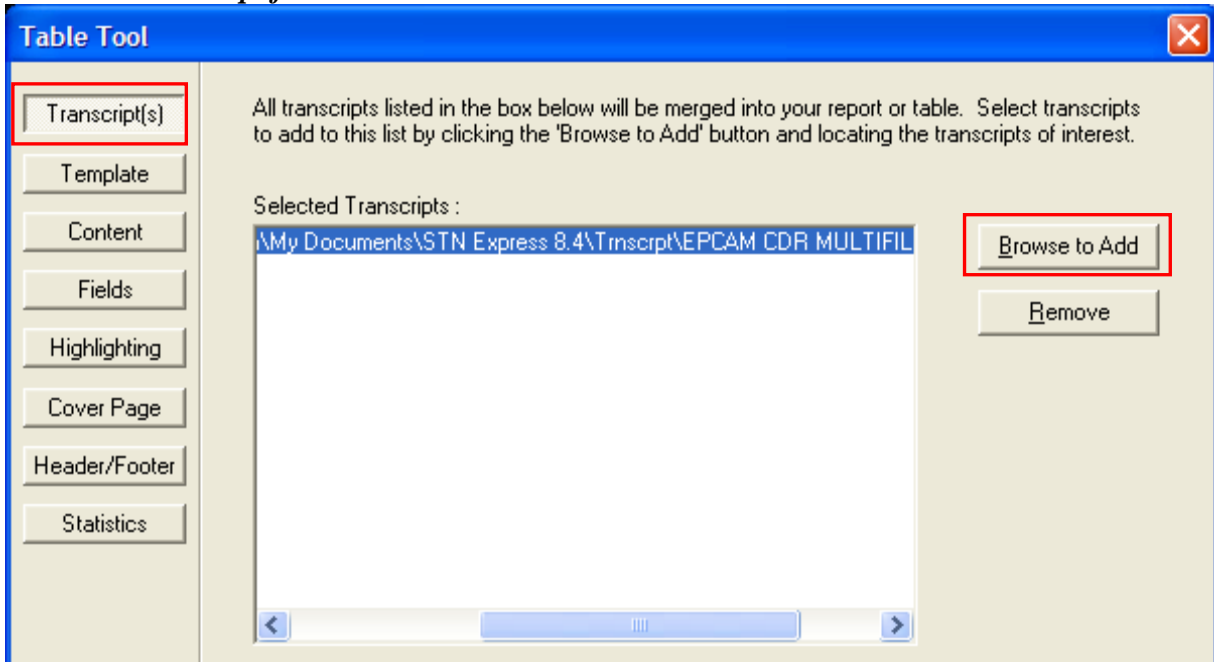
Tip: use **LOGOFF HOLD (LOG H)** to be able to return to this STN session within two hours.

Post-process USGENE, DGENE and PCTGEN results (L7) using the Table Tool

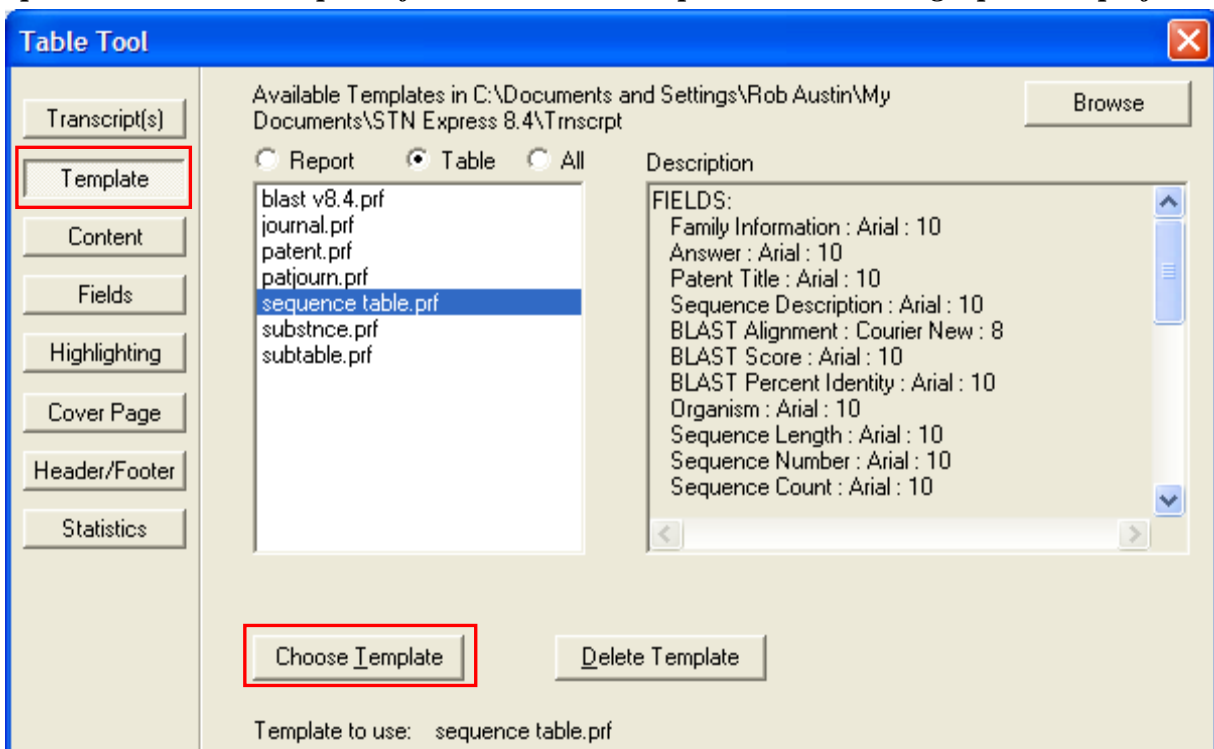
Access the Table Tool from the main STN Express window



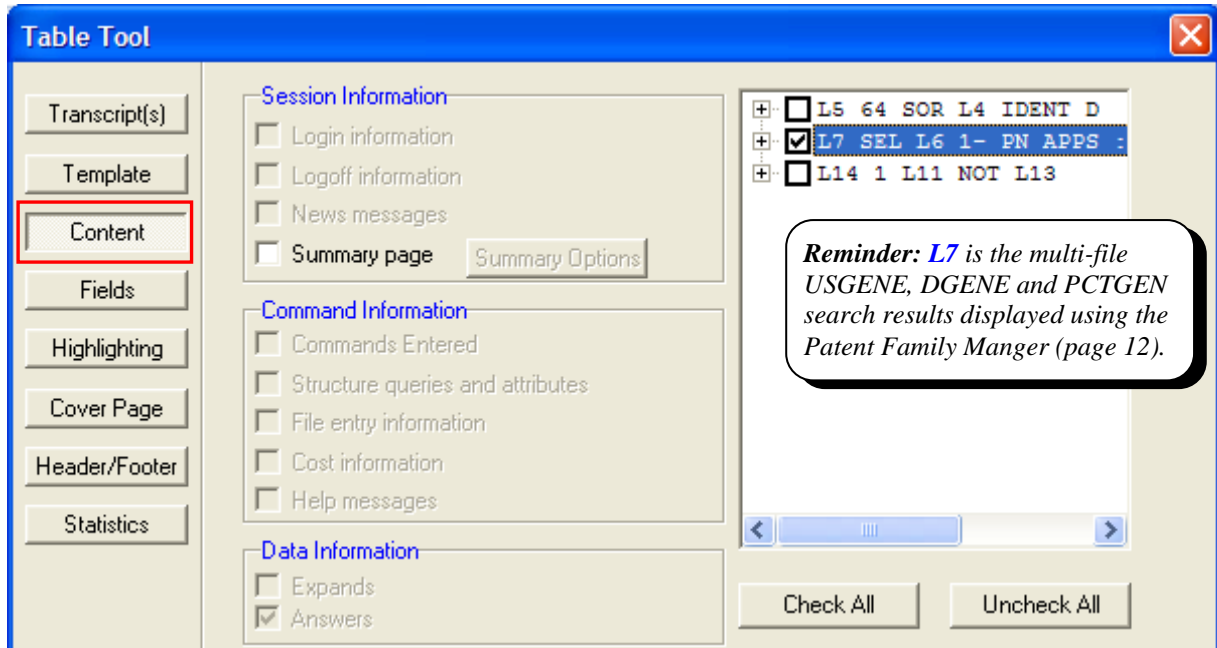
Select the transcript file



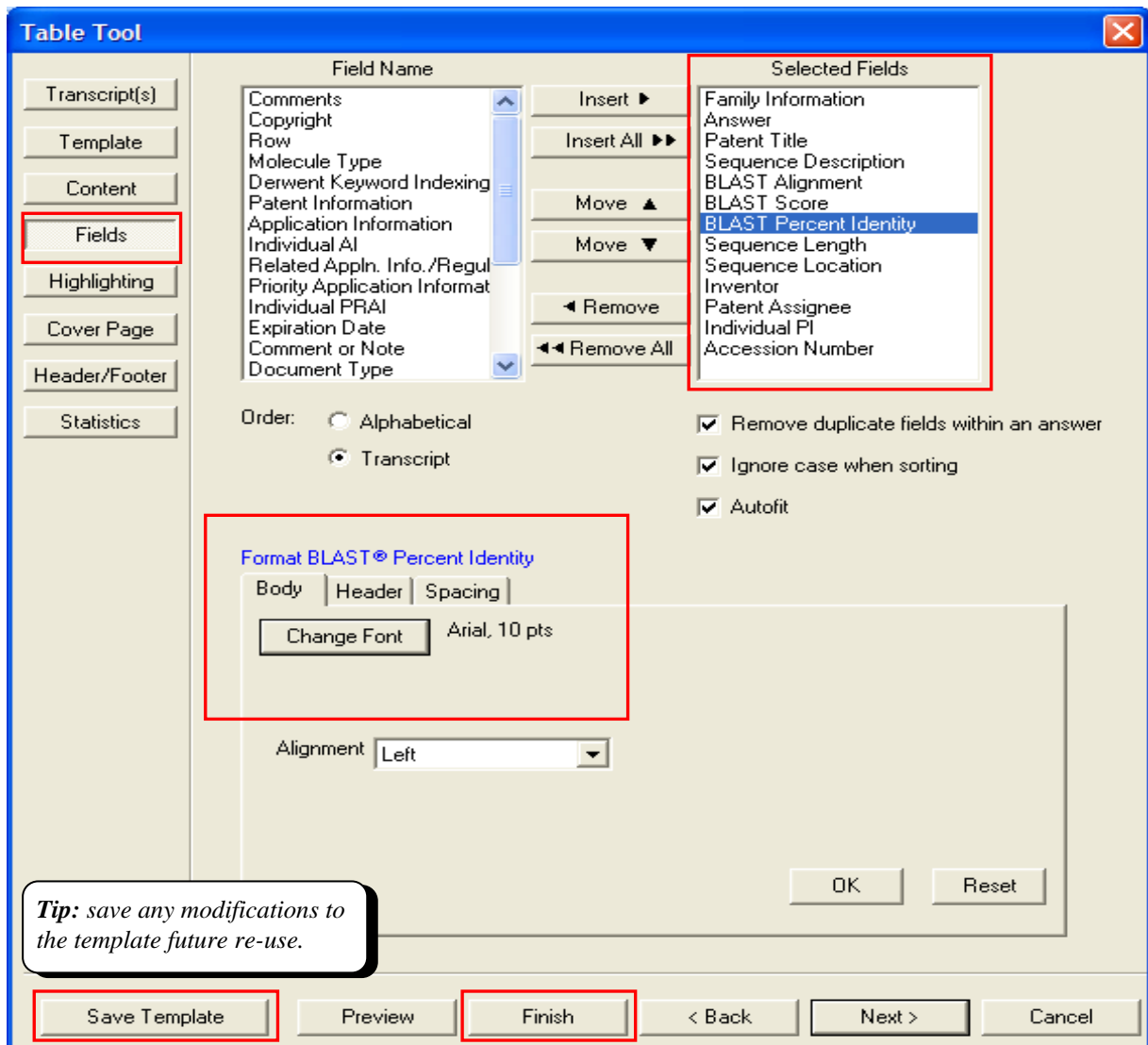
Option: select a custom pre-defined Table Tool template created during a previous project



Select the L-number to be used for the content of the table



Select fields, column order, column headings, fonts and spacing for the table



Review the STN Express Table and adjust the content and layout if needed

The screenshot shows the STN Online and Results application window. The main window displays a table with columns: Family, Patent Title, Sequence Description, BLAST Alignment, and BLAST Score. A context menu is open over the 'Edit Table' option, listing options like Transcript, Template, Content, Highlighting, Cover Page, Header/Footer, Fields, and Statistics. A red box highlights the 'FAMILY 1' column header in the table. A callout box contains the following text:

Note: The STN Express Table Tool provides “family information” – the FSORT family number – as a column choice for the table.

Family	Patent Title	Sequence Description	BLAST Alignment	BLAST Score
FAMILY 1	Antibodies that bind human 17-A1/EpCAM tumor antigen (Patent)	Artificial Protein: Synthetic Peptide; sequence 144 of 169	Query: 18 letters Length = 138 Score = 69.8 bits (157), Expect = 2e-18 Identities = 18/18 (100%), Positives = 18/18 (100%) Query: 1 DMGWGSGWRPYYYYGMDV18 DMGWGSGWRPYYYYGMDV Sbjct: 99 DMGWGSGWRPYYYYGMDV116	70 100% of query self score 70

Export the STN Express Table into a Microsoft Excel spreadsheet

The screenshot shows the STN Online and Results application window. The 'File' menu is open, and the 'Save As...' option is highlighted. A file dialog box is open, showing the file name 'EPCAM CDR MULTIFILE SEARCH 2' and the file type 'Excel Files (*.xls)'. The dialog box also shows the file location 'Tmscript' and a list of files in the directory.

Family	Patent Title	Sequence Description	BLAST Alignment	BLAST Score	BLAST Percent Ident
FAMILY 1	Antibodies that bind human 17-A1/EpCAM tumor antigen (Patent)	Artificial Protein: Synthetic Peptide; sequence 144 of 169	Query: 18 letters Length = 138 Score = 69.8 bits (157), Expect = 2e-18 Identities = 18/18 (100%), Positives = 18/18 (100%) Query: 1 DMGWGSGWRPYYYYGMDV18 DMGWGSGWRPYYYYGMDV Sbjct: 99 DMGWGSGWRPYYYYGMDV116	70 100% of query self score 70	100%

Explore the USGENE BLAST results further using Sort & Filter tools in Microsoft Excel

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 percent identity (above)

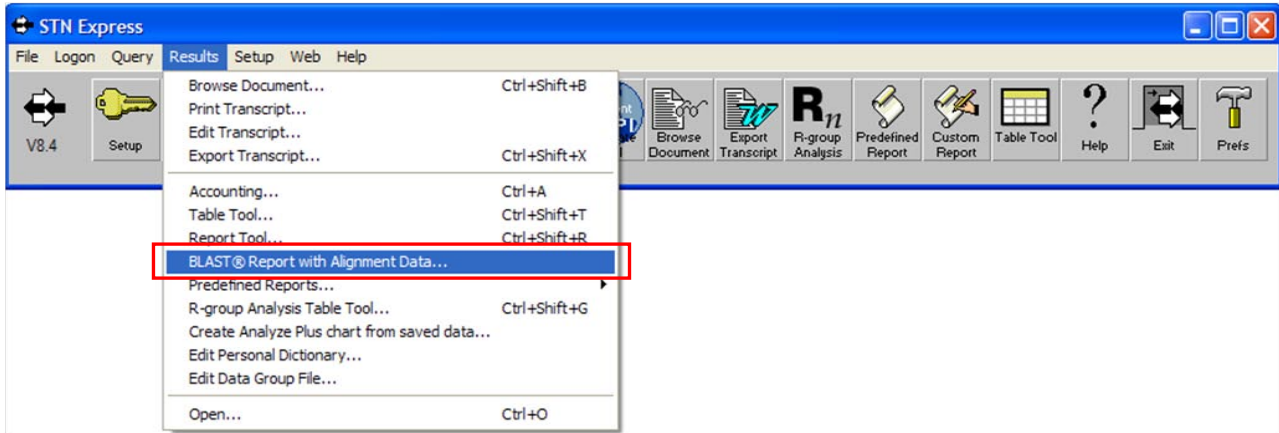
A	B	C	D	E	F	G	H
Family Informatic	Answer	Patent Title	Sequence Description	BLAST Alignment	BLAST Score	BLAST Percent Identity	Sequence Length
FAMILY 2	L7 7 OF 49 USGENE	TREATMENT OF METASTATIC BREAST CANCER (PublishedApplication)	Artificial Protein; Synthetic peptide; sequence 5 of 8	Query = 18 letters Length = 18 Score = 69.8 bits (157), Expect = 3e-19 Identities = 18/18 (100%), Positives = 18/18 (100%) Query: 1 DMGWSGWRPFYYIGMDV 18 DMGWSGWRPFYYIGMDV Sbjct: 1 DMGWSGWRPFYYIGMDV 18	70 100% of query self score 70	100%	18
FAMILY 2	L7 8 OF 49 USGENE	TREATMENT OF METASTATIC BREAST CANCER (PublishedApplication)	Artificial Protein; Anti-EpCAM Heavy Chain; sequence 1 of 8	Query = 18 letters Length = 457 Score = 69.8 bits (157), Expect = 5e-18 Identities = 18/18 (100%), Positives = 18/18 (100%) Query: 1 DMGWSGWRPFYYIGMDV 18 DMGWSGWRPFYYIGMDV Sbjct: 99 DMGWSGWRPFYYIGMDV 116	70 100% of query self score 70	100%	457
FAMILY 2	L7 9 OF 49 DGENE	Use of an anti-EpCAM antibody comprising SEQ ID NOs 3, 4, 5, 6, 7 and/or 8 for manufacturing a medicament for treating human metastatic breast	Human anti-EpCAM heavy chain CDR region, SEQ ID 5		70 100% of query self score 70	100%	18
FAMILY 2	L7 10 OF 49 DGENE	Use of an anti-EpCAM antibody comprising SEQ ID NOs 3, 4, 5, 6, 7 and/or 8 for manufacturing a medicament for treating human metastatic breast	Human anti-EpCAM heavy chain region ID 1.		70 100% of query self score 70	100%	457
FAMILY 3	L7 11 OF 49 USGENE	DOMAIN-GRAFTED ANTIBODIES (PublishedApplication)	Artificial Protein; EpCAM heavy chain (amino acid); seq 2 of 11		70 100% of query self score 70	100%	456
FAMILY 3	L7 12 OF 49 DGENE	New domain-grafted antibody, which specifically binds a human cell-surface molecule, useful for evaluating the functional	Domain grafted-Anti-EpCAM heavy chain.	Query = 18 letters Length = 456 Score = 69.8 bits (157), Expect = 5e-18 Identities = 18/18 (100%), Positives = 18/18 (100%) Query: 1 DMGWSGWRPFYYIGMDV 18 DMGWSGWRPFYYIGMDV Sbjct: 99 DMGWSGWRPFYYIGMDV 116	70 100% of query self score 70	100%	456

Reminder: In this example, one sequence record per family was displayed in bibliographic format, and the others in a free-of-charge review format, by using the Patent Family Manager (page 12).

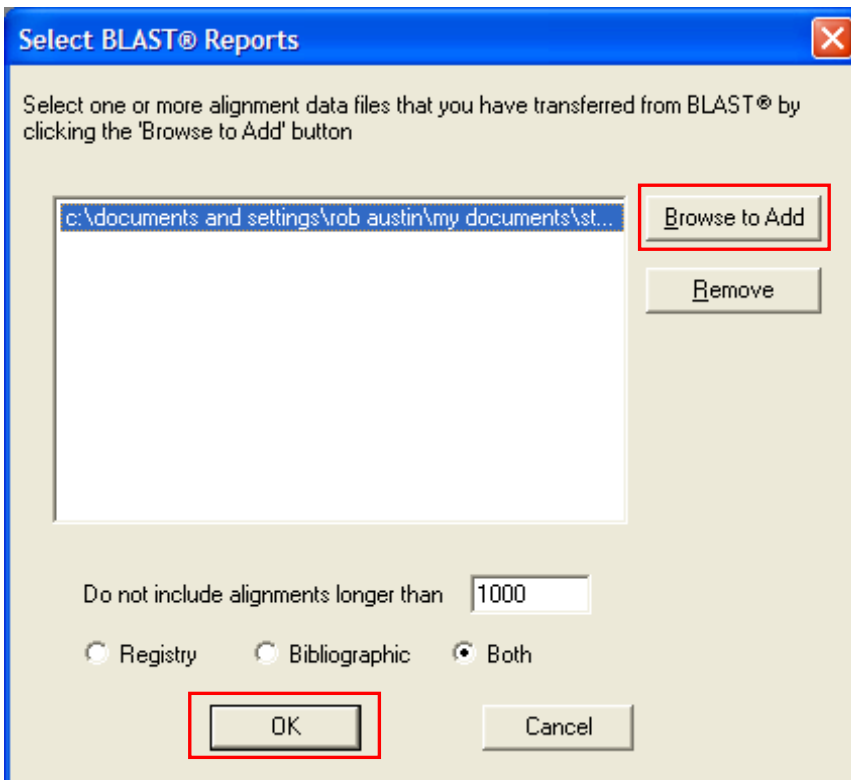
A	H	I	J	K	L	M	N	O	P
Family Informatic	Sequence Length	Sequence Location	Inventor	Patent Assignee	Patent Number	Kind Code	Patent Publication Date	Accession Number	Full-Text
FAMILY 2	18	Claim 11; SEQ ID NO 5	Reinhardt Carsten (Munchen, DE); Saller Robert (Munchen, DE)	MICROMET AG (Munchen DE)	US 20090304716	A1	20091210	20090304716.5 Protein USGENE	Full-Text
FAMILY 2	457								
FAMILY 2	18							AJF32207 peptide DGENE	
FAMILY 2	457							AJF32203 protein DGENE	
FAMILY 3	456	SEQ ID NO 2	Kufer Peter (Moosburg, DE); Raum Tobias (Muenchen, DE); Lutterbuesse Petra (Muenchen, DE); Schliereth Bernd (Germering, DE);	MICROMET AG (Munchen DE)	US 20090241202	A1	20090924	20090241202.2 Protein USGENE	Full-Text
FAMILY 3	456							AGD67246 protein DGENE	

Post-process unique REGISTRY BLAST results (L14) using the BLAST Report Tool

Access the BLAST Report Tool from the main STN Express window

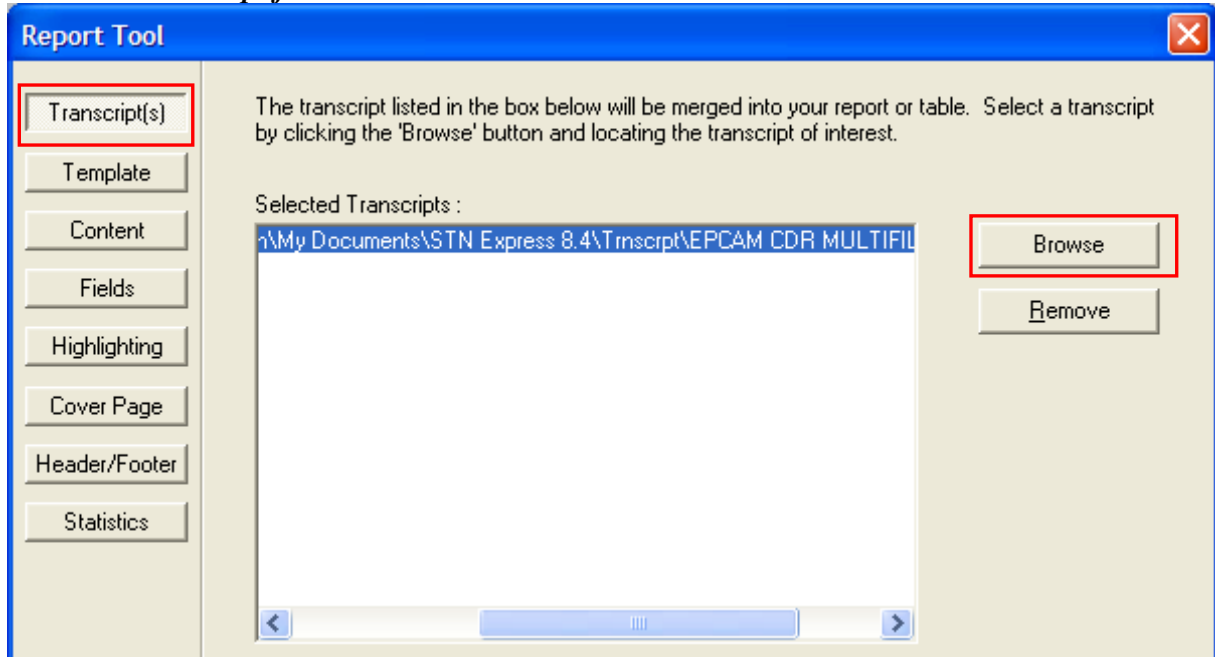
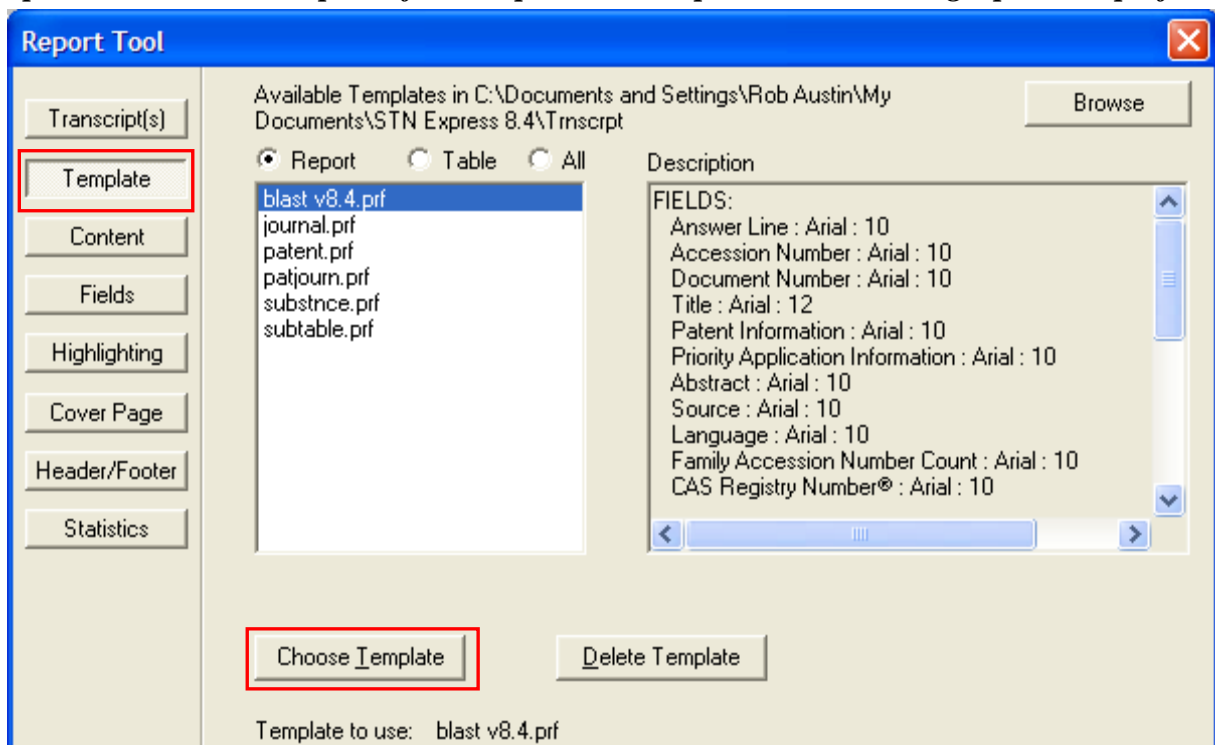


Select the CAS REGISTRY BLAST alignment (.XSS) data file



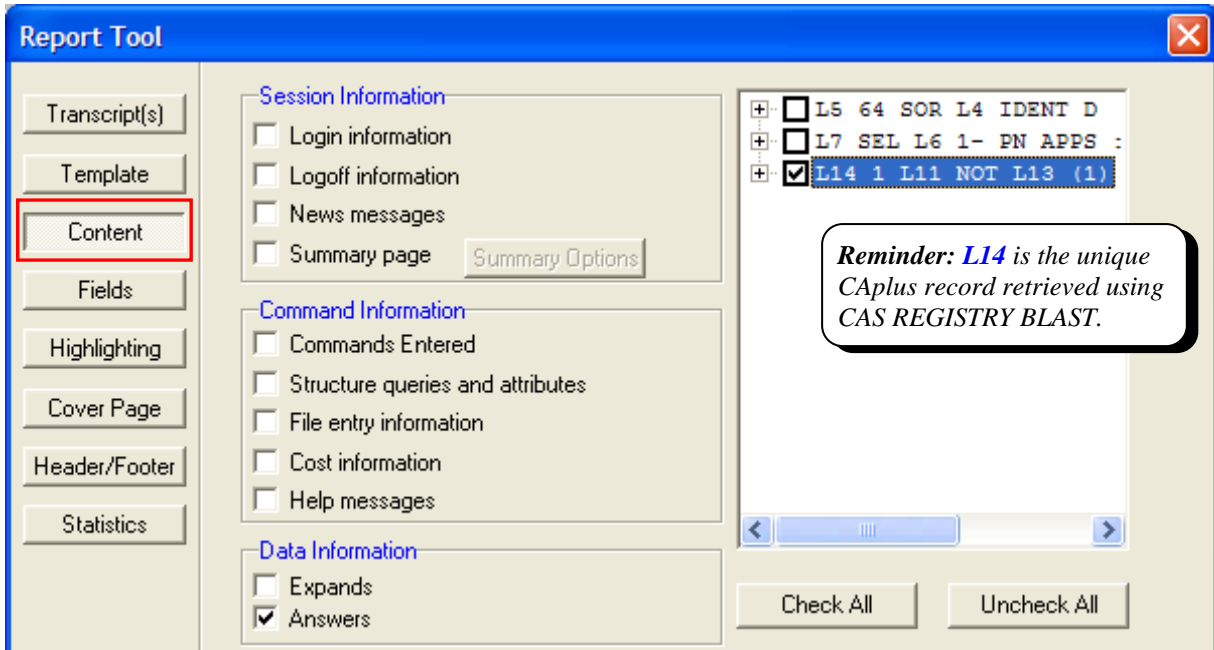
Helpful HINT

Learn more about the CAS REGISTRY BLAST Report Tool:
<http://www.cas.org/support/stnexp/seehow/expressblast.html>

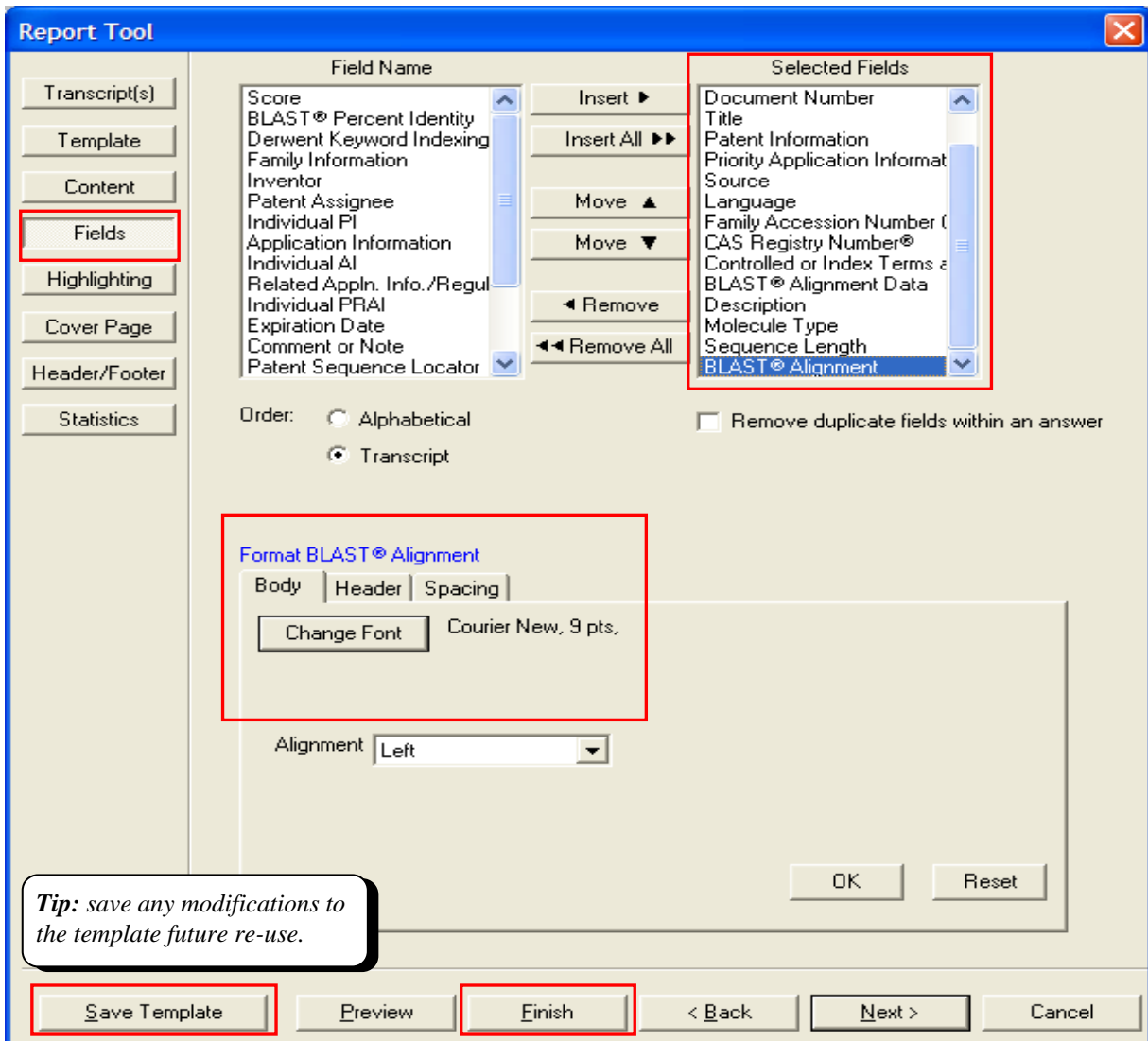
Select the transcript file**Option: select a custom pre-defined Report Tool template created during a previous project****Helpful HINT**

The STN Express template (.PRF) files used in this example are available here: http://www.stn-international.com/stn_biosequence_searching_mfs.html

Select the L-number to be used for the content of BLAST Report



Select fields, display order, header name, fonts and spacing for the report



Review the BLAST Report and adjust the content and layout if needed

STN Online and Results - [Report Output - EPCAM CDR MULTIFILE SEARCH 2.rep]

File Edit Preferences! Web Window Help

L14 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2010 ACS on STN [Add Comments](#)

Accession Number
2008.1210508 HCAPLUS [Full-text](#)

Document Number
149-446284

Title
Cross-species bispecific single-chain antibodies to human or non-chimpanzee primate CD3 ϵ and surface antigen for treating tumorous, proliferative, or immunological disease and cancer

Patent Information

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2008119565	A2	20081009	WO 2008-EP2662	20080403
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

Source
PCT Int. Appl., 397pp. CODEN: PIXXD2

Language
English

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: FRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(amino acid sequence; cross-species bispecific single-chain antibodies to human or non-chimpanzee primate CD3 ϵ and surface antigen for treating tumorous, proliferative, or immunol. disease and cancer)

1067695-29-7
Length = 127 Score = 69.8 Expect = 2e-11
Score = 69.8 Expect = 2e-11
Identities = 18/18 (100%) Positives = 18/18 (100%)
Query: 1 DMGWGSGWRPYYYYGMDV 18

Export the BLAST Report into an RTF file for Microsoft Word

STN Online and Results

File Edit Preferences! Web Window Help

New (Logon)... Ctrl+N
CAS Registry BLAST@
Open... Ctrl+O
Close
Table Tool... Ctrl+Shift+T
Report Tool... Ctrl+Shift+R
BLAST® Report with Alignment Data... Ctrl+Shift+B
R-group Analysis Table Tool... Ctrl+Shift+G
Create Analyze Plus chart from saved data...
Edit Personal Dictionary...
Edit Data Group File...
Predefined Reports...
Save... Ctrl+S
Save As... Ctrl+Shift+S
Save Marked Answers...
Delete Marked Answers
Save Selected...
Edit Text File...
Edit Transcript...
Page Setup...
Print Selected...
Print... Ctrl+P
Printer Setup...
Print Preview...
View TIFF image...
Transcript Filters...
Combine Transcripts...
Create Query Summary File...
Exit STN Online and Results

Source
PCT Int. Appl., 397pp. CODEN: PIXXD2

Language
English

Family Accession Number Count
3

Controlled or Index Terms and BLAST® Data

Enter Filename

Save in: Tmsrpt

ACETYL CYSTEINE
AGENTS MULTIFILE 2
ALLSTR EXAMPLE
another multifile example
autodoc test
BEILSTEIN EXAMPLE
beilstein unique
BENZOTHAZOLE
BIG ANSWERS
BLAST ADVANCED EXAMPLE 1
BLAST BBTV
BLAST EXAMPLE 1

File name: EPCAM CDR MULTIFILE SEARCH 2 Save
Save as type: RTF Files (*.rtf) Cancel
 Include Comments

STN Service Centers

FIZ Karlsruhe

STN Europe
P.O. Box 2465
76012 Karlsruhe
Germany

Phone: +49 7247 808 555
Fax: +49 7247 808 259
E-mail: helpdesk@fiz-karlsruhe.de
Internet: www.stn-international.de

CAS

STN North America
P.O. Box 3012
Columbus, Ohio 43210-0012

CAS Customer Care:
Phone: 800-753-4227 (North America)
614-447-3700 (worldwide)
Fax: 614-447-3751
E-mail: help@cas.org
Internet: www.cas.org

Japan Association for International Chemical Information (JAICI)

STN Japan
Nakai Building
6-25-4 Honkomagome, Bunkyo-ku
Tokyo 113-0021, Japan

Phone: +81-3-5978-3601 (Technical Service)
+81-3-5978-3621 (Customer Service)
Fax: +81-3-5978-3600
E-mail: helpdesk@jaici.or.jp (Technical Service)
cas-stn@jaici.or.jp (Customer Service)
Internet: www.jaici.or.jp