

CAS REGISTRYSM: BLAST[®] similarity searching via STN Express[®]

November 2008

Table of Contents

Overview of content	3
Accessing CAS REGISTRY BLAST	4
Searching similar sequences	5
/iewing ČAS REGISTRY BLAST results	9
Searching by sequence identifier	10
Creating alerts	11
Getting STN data for BLAST results	12
Creating a report with BLAST alignment data	14
For more information	. 16

Overview of content

You can conduct similarity searching of peptides and nucleotides in the CAS REGISTRY database by using the BLAST[®] (Basic Local Alignment Search Tool) algorithm. CAS REGISTRY BLAST is available only for commercial accounts.

Unique content

- Search sequences for peptides and nucleotides from GenBank[®] as well as from patents and journals indexed by CAS
- Find patents, journals, and other technical literature on sequences in many of the sci-tech databases available on STN, such as BIOSIS[®], CAplusSM, USPATFULL and AGRICOLA

Coverage

- Naturally occurring and synthetic proteins and nucleic acid molecules
- · Regulatory elements and other defined genetic elements
- Genetically engineered proteins or genes
- Protein sequences deduced from gene translation
- Oligonucleotide probes and PCR primers
- · Chemically modified peptides, proteins, and nucleotides
- · Sequences with uncommon amino acids or uncommon nucleotides
- · Fusion proteins, genes, and nucleic acids
- Multi component protein complexes
- · Cyclic peptides, peptide metal complexes, and peptide nucleic acid (PNA) sequences

Unparalleled update frequency

• REGISTRY is updated daily

Flexibility

- Import or cut and paste large sequences (up to 50,000 characters) from an electronic file
- Use the common sequence file formats (ASCII, GCG, and FASTA)

Security

• 128-bit encryption

1 Click the **BLAST** button on the STN Express toolbar.



2 Select the desired setup.

The CAS REGISTRY BLAST splash screen is displayed to indicate that you are connected.

Accessing CAS REGISTRY BLAST



Searching similar sequences

 In the Results Set Manager window, you can select a report from a previous search or click New Search to conduct a new BLAST search.

ŝ	Result Se	et Manager								
F	ile <u>E</u> dit	<u>S</u> earch T <u>o</u> o	ls <u>I</u>	<u>H</u> elp						
	New Search	Sequence	Sequ	Jence ID	ELAST Fast BLAST	Aerts Profiles	Help	→ Exit		
	Manage and Review Results									
Γ	Reports	Alerts Repor	ts							
l	1	Name		Туре	Creat	ted 🗸	Status	Results	Reviewed	
l	BLAST se	earch for Dr. S	mith	BLASTn	2004-10-0	6-10:14 AM	Complete	200	~	
l										
l	<u> </u>									
l										
l										
l	L									_
l	<u> </u>									
l	<u> </u>									
l										-
	1 result (100 maximum)									
	View Results Delete Results									
Γ										

2 Click Similar Sequences.

🐉 New Sea	rch
Select Or	ie:
X	Similar Sequences Identify similar sequences using BLAST®.
	Similar Sequences using Sequence Identifier Use Sequence ID to specify query in BLAST search.
61467 •••••	Fast BLAST Search a new sequence using previous settings.
-	Existing Alerts Profiles Update your existing Alerts profiles.
	Cancel

3 Enter a name to track and save the search results and the searched sequence. You can enter the sequence by typing it, copying/pasting it from another application, or "opening", "inserting", or "reading" a sequence from an existing file. The limit for entering sequences is 50,000 characters. Click **OK**.

Similar Sequences	
ile <u>E</u> dit <u>H</u> elp	
Result Name (optional): coagulation facto	r
Sequence to be searched:	Recall Sequence Read from File
atggctcactccccggtgcagtcgggcctgcccggca	atgcagaacctaaaggcagacccagaagag 📥
cctttggagaggtgttcaaaggcattgacaatcggad	ctcagaaagtggttgccataaagatcattg
tcaacaagaaatcacagtgctgagtcagtgtgacagt	ccatatgtaaccaaatattatggatccta
tatettggtggaggeteegaatagatetattagaaa	cctggcccattagatgaaacccagatcgct
acceccacceggagaagaaaacceacagagacaccaa	
<u></u>	
Example #1: a nucleotide sequence	
teccetetee ctatgeaggt tecctggtge	aggcagcgct gactctgcca
gactcattat cagcaaatct gggagaaact	gtcaggataa cctggtctgg
aggtagctat gattatggtt ggcaccagca	gaaatcacct ggcagtgccc
ctgtcactgt gatctatagc aacaacaaga	gaccetegga catecettea
Example #2: a protein sequence	
TVDQHLCGSHLVEALYSVWVHEAKGLPRAAAGA	APGVRAELWLDGALLARTAPRAGPG
QLFWAERFHFEALPPARRLSLRLRGLGPGSAVI	GRVALALEELDAPRAPAAGLERWF
PLIGAPAGAALRARIRARRLRVLPSERYKELAP	IFLTFHYARLUGALEPALPA
ок	Cancel

4 Select a BLAST program.

Selection - coagulation factor <u>- 🗆 ×</u> Nucleotide Query BLASTn Compare a nucleotide sequence against a nucleotide database. BLASTX 8x Compare a nucleotide sequence translated in all reading frames against a protein database. tBLASTx Compare a nucleotide sequence against a nucleotide database, comparing translations in all reading frames. Protein Query BLASTp E Compare a protein sequence against a protein database. tBLASTn En Compare a protein sequence against a nucleotide database translated in all reading frames. Back Cancel

5 Select the entire database or a database subset.



6 Specify the BLAST settings. Enter the maximum number of answers to be retrieved or accept the default. Click **OK**.

For more information on BLAST settings, refer to the NCBI BLAST web page:

http://blast.ncbi.nlm.nih.go v/Blast.cgi

BLASTn Settings - coagulation factor			<u> </u>					
BLASTn Settings - coagulation factor								
Additional Option Presets Search Sensitivity I I I I I I Fewer Answers → More Answers Show Additional Options	Basic Options	Filtering Standard(1)	250					
ок	Back	Cancel						

The Result Set Manager window automatically shows you the status of your search. If you close this window before the search is complete, the search continues to run in the background. You can view the results after **Complete** is displayed in the Status column.

3	Resul	t Set	Manager								_ 🗆 🗵
Ē	<u>F</u> ile <u>E</u> dit <u>S</u> earch T <u>o</u> ols <u>H</u> elp										
	New Sean	ch	Sequence	Seq	Jence ID	Fast BLAST	Aerts Profiles	Help	→ Exit		
					1	Manage and	Review Re:	sults			
	Repor	ts ,	Alerts Rep	orts							
		N	ame		Туре	Creat	ed 🗸	Status	Results	Reviewed	
н	coagu	latio	n factor		BLASTn	2008-06-11	1 10:18 AM	Complete	250		
н	BLAST	r sea	rch for Dr.	. Smith	BLASTn	2004-10-08	6-10:14 AM	Complete	200	√	
н											
н											
н											
н	<u> </u>										
н	L										_
н	L										II
н	L										— II
н	L										— <u> </u>
н	<u> </u>										
						2 results (*	100 maximu	m)			
	View Results Delete Results										

Results are kept in the table until you have 100 sets of results. When you reach 100 sets of results, you have to delete one set before you can conduct another search.

1 In the Result Set Manager window, right-click on the set of results and select View Results.

Result Set Manager <u>File Edit Search Tools H</u>elp ŻŻ 2 GLAST ١ 2 New Search Sequence Sequence ID Alerts Profiles Help Exit Manage and Review Results Reports Alerts Reports Туре Created 🗸 Status Results Reviewed Name coagulation factor BLASTN 2008-06-11 10:18 AM Complete BLAST search for Dr. Smith BLASTN 2004-10-06 10:14 AM Complete View Results Examine <u>Q</u>uery <u>R</u>ename Result Set Up <u>A</u>lert Delete Results Delete -2 results (100 maximum) View Results Delete Results

The BLAST results report shows:

- Number of sequences in the results
- Alignment scores
- Alignment summary with each line representing an individual answer, and the color of each line corresponding to the score range
- Alignment details

<u>F</u> ile	<u>E</u> dit	View	<u>S</u> ea	rch T	<u>o</u> ols	<u>H</u> elp											
	Un	ique S	Seque	nces:	144			Red	undant	82			Sele	cted F	Result	is: 0	
	Alignm	ent Sc 40	ores		0-60			50-8	10		80	-200			>=2	00	
Ľ																	
	Allynm	ent st	umma	ry													
	1			3	31		1	1	1			91		1		121	
																<u> </u>	
	Alianm	ent De	etails														4
Шг		_															
	<u>+</u> 🗆		240	1e-60	0 (44	5369-3	(7-9) 1	: PN: VA	002060	<u>951 S</u>	EQID: 1	unclain	ned DN	A			
	± 🗆		240	1e-60	0 (40)5354-3	<u>35-0) 1</u>	: PN: V	002247	23 SE	EQID: 1 u	Inclaim	ed DN/	<u>≜</u>			
	+		240	le-60	0 (38	1999-0	<u>)1-5) (</u>	SenBan	k AR071	306							
	± 🗆		240	1e-60	0 (38	4412-5	<u>8-2) (</u>	SenBan	k A0532	8							
	± 🗆		240	le-60	0 (38	6074-1	5-3) 0	GenBan	k AR138	377							
	<u>+</u> 🗆		240	le-6	0 (38	6074-1	7-5) 0	SenBan	k AR138	378							
	+ 🗆		240	le-6	0 (38	9189-1	8-8) 0	GenBan	<u>k M1411</u>	3							
	⊥ □		240	10 61	n 790		e 45 c	`-n¤-n	6 0 D440	040						_	
IL	<u> </u>																
		Get	STN	Data					Cano	el					\$/		
Res	Result complete.																

Viewing CAS REGISTRY BLAST results

Searching by sequence identifier





2 Enter the sequence identifier to be searched and a name for the results.



Follow the same steps as for searching similar sequences:

- 1. Select a BLAST program in the Program Selection window.
- 2. Select the database or a database subset.
- 3. Specify the BLAST settings.
- 4. Run the search.

Creating alerts

After you conduct a CAS REGISTRY BLAST search, follow these steps to create a CAS REGISTRY BLAST alert:

3	Result Sel	t Manager								_	
<u>F</u> ile <u>E</u> dit <u>S</u> earch T <u>o</u> ols <u>H</u> elp											
Ne	ew Search	Sequence	Sequ	ence ID	Fast BLAST	Aerts Profiles	Help		* Exit		
	Manage and Review Results										
	Reports	Alerts Repor	tsl								
Iг	N	iame	<u> </u>	Туре	Creat	ted 🗸	Status	Results	R	eviewed	
	coagulatio	n factor		BLASTn	2008-06-1	1 10:18 AM	Complete		View	, Reculte	
E	BLAST sea	arch for Dr. S	mith	BLASTn	2004-10-0	6 10:14 AM	Complete		<u>v</u> iew	rtesuits	
Iŀ			_						Exam	iine <u>Q</u> uery	
									<u>R</u> ena	me Result	
									Set U	p <u>A</u> lert	
Uŀ.									<u>D</u> elet	e Results	Delete
											-
					2 results (100 maximu	im)				
		View	Resu	ilts			D	elete Res	sults		
381	Create Pr	ofile								-	
Elle	e <u>E</u> dit	<u>H</u> elp								_	
			_								_
11'	Databa	ne. ji see: REGIS	TRY	T RI	in Frequer	or Maakt		Evniratio	n:	2-2009	ΞIJ
	Inclu	de previous	ly retr	rieved se	quences a	s they are u	pdated	Expiratio	Set E-	mail Addre	ISS
	elect the	etene vou w	antto	have as	nart of the	profile (the	firet etan i	e alwave	include	ad):	
	cicci inc i	51005 900 111		nave as	purrorare	prome (me	inst step it	5 diwdy5	Terada		
	💌 Seque	nce searc	h wi	th the	followin	g paramet	ers:				
			Resi	ilt Name	e: coagul	ation fac	ctor				-
				Subsets	a: BLASTr s: Patent	1 .s					
					Non-pe EST	tents					
					STS						
					HTGS	ConRon's					
	Re	eturn Sequ	ience	e Length	n: All	Gennanke					
	Low	Complexit	y Fi	ltering	7: On						

Word Size: 1. Open Gap Cost: 5 Extend Gap Cost: 2 Penalty for Mismatch: -Reward for Match: 1

1 In the Result Set Manager window, right-click on the set of results of interest and select **Set Up Alert**.

> Define the options for the alert in the Create Profile window. Click **OK**. Click **Set E-mail Address** to specify an e-mail address for sending the alert results and for notifying about expiration dates.

Cancel

CAS REGISTRY BLAST alerts are set to expire 1 year from the date when they are created. If you provide an e-mail address, you will be notified via e-mail that your alert will soon expire.

- Select the sequences for which you want to get STN data.
 - Choose Edit > Select All to select all sequences.
 - In the Alignment Scores section, click the bar for the desired range to select the entire range.
 - Click the selection boxes for individual sequences or their CAS Registry Numbers[®] to select individual sequences.
- 2 Click Get STN Data.
- 3 Click the desired option. The Transfer all alignment data for postprocessing box is checked by default so that your BLAST alignment data is saved and can be combined with your STN transcript, if you choose to create a report.

Getting STN data for BLAST results





4 Enter a file name in which to save the BLAST alignment sequence data. The data are saved in an STN Express Saved Sequences (.xss) file.

Save File As							_ 🗆 ×
Save įn:	🛅 Trnscrpt			•	1		
My Recent Documents Desktop My Documents My Computer PC88973							
Mu Notwork	File <u>n</u> ame:	coagulation factor				<u></u> a	ve
Places	Files of type:	STN Express Saved	Sequences		-	Car	ncel
Note: Use of the s	saved informat	ion is subject to co	pyright and data us	e restri	ctions.		

The standard STN logon procedure is initiated. Once an online session is initiated, STN Express automatically enters appropriate databases and conducts searches to obtain the results that you requested in the Get STN Data window.

Note: You can create a report from the BLAST alignment data and records from a transcript. For the data from an STN record to be included in the transcript, it must first be displayed in STN. Be sure to display any data that you intend to include in a report.

- Use a display format in REGISTRY that includes CAS Registry Numbers[®] if you want the BLAST alignment data included within the REGISTRY records in your transcript.
- Conduct a search on the REGISTRY L-number from BLAST results in CAplus and display records in the desired format in order to include bibliographic data from the transcript in your report. For example, if you want to include bibliographic information and abstracts for CAplus records, display answers in the BIB ABS format.

Creating a report with BLAST alignment data

1 Select the data to retrieve.

To create a report that merges transcript results from your STN online session with BLAST alignment data, make sure that the box **Transfer all alignment data for postprocessing** is checked in the Get STN Data window of the BLAST search.

2 Log off STN.

3 Select Results > BLAST Report with Alignment Data.

Get STN	Data 🗕	
Retrieve th	e following data:	
¢	Sequence Records Retrieves Sequences from CAS Regis	try
	Reference Records Retrieves CAPLUS References	
	Sequence and Reference Records Retrieves CAS Registry Sequences an CAPLUS References	d
🔽 Trans	fer all alignment data for postprocessing	
	Cancel	

File Logon Query Results Setup Web Help

Browse Document	Ctrl+Shift+B	
Print Transcript		
Edit Transcript		
Export Transcript	Ctrl+Shift+X	
Accounting	Ctrl+A	
<u>T</u> able Tool	Ctrl+Shift+T	
<u>R</u> eport Tool	Ctrl+Shift+R	
BLAST® Report with Alignment Data		
Predefined Reports		•
R-group Analysis Table Tool	Ctrl+Shift+G	
Create Analyze Plus chart from saved data		
Edit Personal Dictionary		
Edit Data Group File		
Open	Ctrl+O	

4 Select one or more alignment data files in the Select BLAST Reports window. Use **Browse to Add** to locate and add files. Note that the default is set to exclude alignments longer than 1000 residues. You may also choose to include only the REGISTRY data, only the bibliographic data, or both. Click **OK**.

5 The first screen of the Report Tool window shows the Transcript portion of the window with the last transcript that was created. Use the displayed transcript or click **Browse** to locate the transcript you want to use.

Se	lect BLAST® Reports	x
Si cl	elect one or more alignment data files that you have transferred icking the 'Browse to Add' button	l from BLAST® by
		Browse to Add
		Remove
	Do not include alignments longer than 1000	
	🔿 Registry 🔿 Bibliographic 💿 Both	
	OK Cancel	1

Report Tool		×
Transcript(s) Template	The transcript listed in the box below will be merged into your report or table. Select a transcript by clicking the 'Browse' button and locating the transcript of interest. Selected Transcripts :	
Content	Settings\ekh87\My Documents\STN Express 8.3\Trnscrpt\blast.tm Browse	
Highlighting	Remove	
Cover Page		
Header/Footer		
Fields		
Statistics		
	•	
<u>S</u> ave Temp	late	

Select from the buttons on the left side of the Report Tool window.

Use this button…	То
Transcript(s)	Select one or more transcripts for the report
Template	Select a template for the report
Content	Define the content of the report
Highlighting	Define highlighting in the report
Cover Page	Define the content and format for the optional cover page
Header/Footer	Define the content and format for the optional header or footer
Fields	Select and format the fields from answers
Statistics	Define charts for the report based on data from fields in the transcript(s)

For more information

Refer to STN Express Training at <u>www.cas.org</u>.



 CAS Customer Center

 Phone:
 800-753-4227 (North America) 614-447-3700 (worldwide)

 Fax:
 614-447-3751

 E-mail:
 help@cas.org

 Internet:
 www.cas.org

November 2008 CAS2535-1108