SRS USERS MANUAL

Version 6.0.7

Copyright (c)2000 LION Bioscience AG (LION). All rights reserved.

LION Bioscience SRS6.07 Documentation

This manual and the accompanying software are furnished under the terms of the SRS license and may only be used or copied, or in any other way reproduced under the terms and conditions set out in that license. The information in this manual is furnished for information purposes only and is subject to change without notice. The material herein should not be construed as a commitment by LION. LION assumes no responsibility or liability for any errors, inaccuracies, or omissions that may appear in this book.

Please be sure to obtain permission before you make copies of any images or other artwork in this book.

Comments are welcome to documentation@lionbio.co.uk

For customer support issues please e-mail support@lionbio.co.uk

Overview of SRS

A Quick Tour of SKS4	uick Tour of SRS
----------------------	------------------

SRS Sessions

Introducing Sessions	. 18
Starting your Session	. 18
Permanent Sessions	. 21
Using the Session Manager	. 21
Temporary Sessions	. 28
Managing Temporary Sessions	. 28

Querying With SRS

The basics of Querying	2
Picking a Search Method	2
Constructing Search Terms	б
Querying	0
Doing a Quick Search	1
Browsing Databank Indices	2
Using the Standard Query Form	5
Using the Extended Query Form	б
Doing an Expression Query4	8

Linking

What is Linking?	51
What is a Cross-Reference	
Picking a Linking Option	
Initiate A Linking Operation	53
Link Whole Queries	53
Multiple Entry Linking	55
Linking a Single Entry	56
Expression Linking	57
Linking Operators.	58
Linking Examples	59

Views

Overview of Views	. 62
Using Views	. 62
How to Apply Views	. 63
Creating Views	. 70
Creating Views with the Query Forms	. 71
Creating Views with the "View Manager"	. 72

Applications

Introducing Applications	. 78
Using Applications	. 78

Command Line SRS

getz	. 84
Important Options	. 84

The SRS Query Language

General Syntax	
List of Operands	
List of Operators	
Searching in Indices	
Using Logical Operators	
Entries and Subentries	
Storing Intermediate Results in Sets	

СНАРТЕК

OVERVIEW OF SRS

This documentation assumes that you have a working knowledge of computers including the operating system and the hardware. It also assumes that you have access to an SRS server and an HTML 4.x or higher web browser.

1.1 A Quick Tour of SRS

To help you realize the potential of SRS the following guided tour has been provided. It will demonstrate many of the features that SRS is famous for such as querying, linking, creating and applying predefined and custom views and saving queries in a systematic way.

Whenever an SRS concept or important topic is introduced a margin note will be included that will tell you where to go for more information.

SRS Sessions - Chapter 21.We begin at the "Start" page.
Ask your SRS administrator for the address to your server.



FIGURE 1. The Start Page.

Querying SRS - Chapter 3

 Click the "Start" link. This takes you to the "Top" page where you begin working with SRS.

Navigation T	abs — 🗩	TOP PAGE QUERY RESULTS SESSIONS VIEWS DATABANKS
	Reset	Quick Search All Entries
Options	Query forms	show all
	Standard Extended	EMBL EMBLNEW GENBANK SWISSPROT SWISSNEW D PIR SPTREMBL REMTREMBL GENPEPT SWALL
		$\Box \underline{\text{NRL3D}} \Box \underline{\text{IMGT}} \Box \underline{\text{AAGENESEQ}} \Box \underline{\text{NAGENESEQ}} \Box \underline{\text{OGLYC}}$
Data	bookmark this link to return to	 ✓ SeqRelated ✓ TransFac
) your session	User Owned Databanks

FIGURE 2. The "Top" page.

- 3. Choose the databanks that you want to search.
 - Select the checkbox to the left of the EMBL hyperlink.
- 4. Click the "Standard" button.

This will bring up the Standard query form for the EMBL databank.

e	OP PAGE QUERY	RESULTS SESSION	5 VIEWS DATABANK	S HELP
Reset search EME	<u>3L</u>	Info ab	out field AllText	-
Subnit Query	separate multi AllText AllText AllText	ple values by & (and), search string	(or),	

FIGURE 3. The Standard query form for the EMBL databank.

5. Enter the search phrase: Change the first drop-down data field to "Description" and enter "ascorbate oxidase".

Standard

separate multiple values by & (and), (or), ! (and not)				
Description	ascorbate oxidase			
AllText				
AllText				
AllText				

FIGURE 4. Search for a specific entry.

6. Click the "Submit Query" button. This will bring up the "Query Results" page. If you want to look at the complete entry, click on the hyperlink.

TOP PAGE	QUERY	RESULTS	SESSIONS	VIEWS	DATABANKS	HELP
Reset		<i>Quer</i> ascor	v "[embl-Desc bate oxidase*	cription: asco] " found 18	rbate* & oxid entries	ase*
Perform operation on all but selected Link Save View *Names only* V		MBL:AA840 MBL:ATTS4 MBL:ATTS4 MBL:CMAA MBL:CMAA MBL:CSAS0 MBL:CSAS0	724 167 945 1798 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			
Launch OligoBlast show entries in chunks o 30 Printer Friendly		MBL:MTY1 MBL:NTASC MBL:NTTAC MBL:OSAB4 MBL:E10589 MBL:E12719 MBL:E12720	2295 200XRP 20P 4799 2			

FIGURE 5. The results of the query.

Linking - Chapter 4

7. Click the "Link" button. This will bring up the "Linking" page.



FIGURE 6. Set linking operation.

- 8. Link to SWISS-PROT: Select the checkbox to the left of the SWISSPROT hyperlink.
- Click the "Submit Link" button. The results of the linking operation will be a list of SWISS-PROT entries, displayed in the "Query Results" page, that are related to the EMBL entries we started with.
- 10. Click the "Views" tab to create a new view. This brings up the "View Manager" page.

Views - Chapter 5

a	TOP PAGE	QUERY	RESULTS	SESSIONS	VIEWS	DATABANKS	(HELP)
Reset				V	iews Ta	b	
View name: myTestView Display view as C list © t In table headers to © full field nat C short field n Select fields from C all fields in 1 © just common	able ise mes iames ibraries m fields	Delete View Edit View Feate New Vi Select d to define MBL_featurest MBLNEW MBLNEW MBLNEW_fe MBLTEST_fe ENBANK_fee WISSPROT	Select Nan iew atabanks a view for atures atures atures	Select intervention Select Imked EMBL	r delete: t databank t o displayi eatures W W_features ST_features VK K_features ROT	s to be ed entry	

FIGURE 7. Create your own views (Steps 11, 12, 13, 14).

- 11. In the "View name" textbox type "myTestView".
- 12. In the "Select databanks to define view for" menu select EMBL.
- 13. In the "Select databanks to be linked to displayed entries" select SWISSPROT.
- 14. Click the "Create New View" button. This brings up the "Data Field Selection" page.



FIGURE 8. Select the "Description" fields for both databanks.

15. Choose the data fields that you want displayed when the view is applied to an entry or list of entries.
Select the "Description" field for both detabanks

Select the "Description" field for both databanks.

16. Click the "Save View" button.

The new view is available with EMBL entries.

17. Click the "Results" tab.

This brings up the "Results List" page. See figure 9.

Тор ра	AGE Q	UERY	RE	SULTS SESSI	ONS	S VIEWS DATABANKS	HELP
Reset				Expression	4		×
Perform operation on					;	Successful Queries	
selected queries Save	Name	Туре	N Total	From Library	N	Query Expression	Comment
Delete	□ Q2	link	3	SWISSPROT	3	(([embl-Description:	
View myTestView ▼ *Names only* *Complete entries* SequenceSimple SwissView FastaSeqs proteinChart tfProteinChart SqSimLdr SqSimLdr SqSimLdr SqSimLdr SqSimLdr	₩ Q1	query	18	EMBL	18	[embl-Description: a	

FIGURE 9. View with your own view (Steps 17, 18, 19).

- 18. Click the checkbox to the left of the "Q1" list item to select it.
- 19. Change the view option to "myTestView". Click the drop-down menu below the "View" button and scroll through until you find the view you want.
- 20. Click the "View" button. This will display the query with the "myTestView" view.

TOP P/	AGE QUERY	RESULTS	SESSIONS	VIEWS	DATABANKS	(HELP) 4167 ATTS4945
Perform operation	EMBL:C	BL MAAO	Description	ima ASC	VISSPROT	Description L-ASCORBATE
 on all but selected on selected Link Save Vieu 			AAO gene for ascorbate oxid complete cds.	ase,		OXIDASE PRECURSOR (EC <u>1.10.3.3)</u> (ASCORBASE).
myTestView	EMBL:C	<u>SASCO</u>	Cucumber ascorbate oxid (aso) mRNA, complete cds.	ase ASC	CUCSA	L-ASCORBATE OXIDASE PRECURSOR (EC <u>1.10.3.3</u>) (ASCORBASE) (ASO).
show entries in chunks o 30 Y Printer Friendly	f 🗆 <u>EMBL:C</u>	<u>SASOX</u>	Pumpkin mRN for ascorbate oxidase	A ASC	CUCMA	L-ASCORBATE OXIDASE PRECURSOR (EC <u>1.10.3.3)</u> (ASCORBASE).

FIGURE 10. The view shows linked entries.

- 21. Click the "Results" tab to go back to the "Results" page.
- 22. Select "Q2".
- 23. Click "View".

The default view "Names only" will be fine.

Applications - Chapter 6

24. Click the drop-down menu below the "Launch" button. The drop-down menu shows all the available applications that are valid for the current entries.



FIGURE 11. Select the application (Steps 23, 24, 25).

- 25. Select "BlastP".
- 26. Click the "Launch" button. This will bring up the "Application Launch" page.

BlastP	Name of job: temp	Database to search: SW	/ISSNEW 🔽		
select view to display	SWISSPROTASO begin 1 1 MLQNGKAI 61 FPGPTIR. 121 YNFTVDNI 4	CUCMA 11 21 REPNFLILFFGLILAFGISS 71 81 ANAGDTVVVELINKLHTEGVV 131 141 PGTFFYHGHLGMQRSAGLYGS	31 41 SEGSQIRHYKWEVEYMFWA 91 101 YIHWHGILQRGTPWADGTA 151 161 SLIVDPPQGKKEPFHYDGE	51 PDCNENIVMGINGQ 111 SISQCAINPGETFF 171 INLLLSDWWHQSIH	▲
* Complete entries * 💌 show results automatically 🔽	Output (Number of alignments to si	Options how 250	Search Param Filter query sequence	e 🗖	
Reset	Number of best hits from a Number of one-line descri	a region to keep 100 ptions 500	Scoring matrix BLO: The E value 10.0000 word size Default Perfom gapped align	SUM62 🔽	
			Cost to open a gap Cost to extend a gap	Default 💌	

FIGURE 12. Default settings for "Application Launch" page.

27. Click the "Launch" button.

The application will run and the "Application Invocation" page will be displayed.

TOP PAGE	QUERY RESULTS SESSIONS VIEWS DATABANKS HEL	2
	STOP!!	
	Don't go BACK	
G	BlastP is currently running. Please be patient The results will appear here shortly. running /data/appl/bin/blastall -p blast	

FIGURE 13. The Application is running.

When the "Application Invocation" page finishes loading the results will be displayed automatically. See figure 14.

Reset	Query "[BlastP-JobName.temp_job1]" found 6 entries		
Perform operation	□ <u>BLASTP:temp_job1_Report</u>		
• on all but selected	BL&STP 2.0.8 [Jan-05-1999]		
Link Save View	Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Sch Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database s programs", Nucleic Acids Res. 25:3389-3402.	naffer, , search	
Names only *Complete entries* BlastAlignment	Query= <u>SWISSPROT:ASO_CUCMA</u> (579 letters)		
BlastPSwissProsEnz BlastPFamilies BlastPFamilies2	Database: /data/flatfiles/blast/swissnew 245 sequences; 60,098 total letters		
BlastTest BlastPApp Printer Friendly	Searching	done	
	Sequences producing significant alignments:	Score (bits)	E Value
	SWISSNEW 052393 RECA_CHLTE_RECA PROTEIN. SWISSNEW 018917 B3A3_RABIT_ANION_EXCHANGE_PROTEIN_3 (NEURONAL B SWISSNEW 065888 NLTP_STRHE_NONSPECIFIC_LIPID-TRANSFER_PROTEIN_P SWISSNEW P73709 YI19_SYNY3_HYPOTHETICAL_35.3 KD_PROTEIN_SLR1819.	23 22 22 22	3.1 5.4 5.4 9.2

FIGURE 14. Application Results.

28. Select the drop-down menu below the "View" button.

29. Select the "BlastPFamilies" view.

30. Click "View".

This displays the results of the application using the "BlastPFamilies" views. See figure 15.

Reset Query "[BlastP-JobName:temp_job1]" found 6 entries							
Perform operation	BLASTP	<u>tsc</u>	SearchDB	des	<u>sl</u>	PROSITE	BLASTP
on all but selected on selected Link Save Vieu	□ <u>BLASTP.temp_job1_Report</u>		ASO CUCMA	L-ASCORBATE OXIDASE PRECURSOR (EC <u>1.10.3.3</u>) (ASCORBASE).	579		1
BlastPFamilies	PLASTDtemp.ioh1 migner PECA_CHITE		RECA CHITE	RECA PROTEIN	346		
Launch BlastP Show entries in chunks of 30 Printer Friendly	BLASTP.temp_job1_swissnew_B3A3_RABIT		B3A3 RABIT	ANION EXCHANGE PROTEIN 3 (NEURONAL BAND 3-LIKE PROTEIN) (ANION EXCHANGER 3 BRAIN ISOFORM).	1233		

FIGURE 15. One of the Pre-packaged SRS Views.

Saving Queries

Chapter 2 "SRS Sessions".

31. Click the "Results" tab.

This brings up the "Results" page where you can pick the queries to save.

32. Select the queries to save.





FIGURE 16. Saving a Query.

33. Click the "Save" button.

This brings up the "Download Options" page.

34. Choose the options you want.

Use the "Download Options" page to specify the way the output should look and to pick a format for the output file.

Reset	Query "[BlastP-JobName:temp_job1]" found 8 entries
Save	sequence format not sequence format
	use view [★] Complete entries [★] ▼ Save table as ASCII text/table with
	Column Separator
	Record Separator \n
	use mime types: binary(to file)
	binary(to file) chemical/x-pdb k

FIGURE 17. Set the Download Options.

35. Click the "Save" button.

The browsers "File Download" dialog (figure 18) appears.

File Download		×
	You have chosen to download a file from this location. wgetz1cdaf0af from www.lionbio.co.uk What would you like to do with this file? © Open this file from its current location © Save this file to disk May ask before opening this type of file	
	OK Cancel <u>M</u> ore Info	

FIGURE 18. The "File Download" Dialog.

36. Click the "OK" button.

This will bring up the "Save As" dialog.

Save As			? ×
Save jn: 🕎 rol	o on 'PCSERVER' (Z:)	•	<u> </u>
📄 .desktop-picasso	🚞 apache	🚞 printfiles	🚞 uploads
🚞 .ebtpriv	🚞 apache_1.3.2	🚞 public_html	🚞 usefullmag
📄 .netscape	🚞 dumpster	🧰 rob	🚞 user_mif
📄 .ntprofile	🚞 icabkup	🚞 srs5.1	🚞 works
🚞 .web	🚞 nsmail	🚞 srs6users	🚞 workshops
🚞 .webcache	🚞 picSRS	🚞 tempexec	
•			•
File <u>name</u> : wgetz	1cdaf0af		<u>S</u> ave
Save as type: Docu	iment	•	Cancel

FIGURE 19. Set the Save location.

37. Click the "Save" button.

The file will be saved to the location you specified.

That's it! Now you have seen most of SRS and are ready to start using it at work. Well, almost. There are other tutorials in this manual so take a few hours, read the manual and you will have fewer problems in the future.

CHAPTER



SRS SESSIONS

You can do your work in SRS using either temporary or permanent sessions.

- **Permanent Session:** If you want to return to your session from time to time, if you think you may need to move work from one session to another, if you want a record of all your sessions in one place and the ability to recall any of your previous sessions or begin new sessions, permanent sessions are the way to go.
- **Temporary Session:** Use temporary sessions for simple lookups, for example, looking up something once in a while or to run an occasional Blast search.

You will learn about both types of sessions in this chapter including:

- How to begin or return to a session.
- Benefits of using either session type.
- Additional features made available to the permanent session users.
- Making temporary session data available in other sessions.
- Connecting temporary sessions to permanent sessions.

2.1 Introducing Sessions

All the queries you do and views you create are stored in the session history and hence are available to use again. If you have accessed SRS with the permanent sessions option, all your work will be available when you return to that session. If you are using temporary sessions your work may still be available. Making a bookmark with the browser allows you to return to a temporary session for a limited amount of time.

Starting your Session

You begin working with SRS in the "Start" page. See figure 1.





There are four links on the "Start" page, they are:

Start

Begin a new temporary session.

Permanent Session

Begin or return to a permanent session.

Databanks

Check the status of databanks available on your SRS server.

Information

Online documentation such as the HTML and PDF version of this manual. Other online documentation includes the Administrators manual, Developers manual, Classes and Commands quick references, the context sensitive help, and the SRS FAQ's databank.

Start a Permanent Session

To begin or return to a permanent session:

- 1. Click the "Permanent Session" link.
- 2. Complete the login information.

The web browser will prompt you to give a user name and password. You need to ask the SRS administrator for these

because they may not be the same as your system account ID and password.

Enter Net	work Passwo	rd	? ×
?	Please type yo	ur user name and password.	
- ∛	Site:	www.lionbio.co.uk	
	Realm	session	
	<u>U</u> ser Name		
	<u>P</u> assword		
	□ <u>S</u> ave this p	bassword in your password list	
		OK Can	cel
	User Name Password	Dassword in your password list	cel

FIGURE 2. SRS user account access.

Note: Dialogs shown throughout this chapter are generated in the Win32 version of IE5. The dialog you see may look different, it depends on your system and browser.

3. Click "OK".

This will bring up the "Session Manager" page. From here you can either continue working with the most recently used session, switch to another session, or begin a new session. These options will be explained in the "Permanent Sessions" section below.

Skip ahead in this chapter to the "Permanent Sessions" section if you are following along on-screen and would like to continue. Otherwise read on to find out how to begin a temporary session.

Starting a Temporary Session

To begin a temporary session from the "Start" page:

1. Click the "Start" link.

This takes you to the "Top" page where you can immediately begin working with SRS.

Note: You can also begin a new temporary session by clicking the "Databanks" link. This will start a new temporary session in the "Databank Information" page.

To continue working with temporary sessions skip-ahead now to section 3, "Temporary Sessions."

2.2 Permanent Sessions

To keep your work in a reliable location use the permanent sessions option. You can swap between sessions and move work from one session to another easily. In addition, your work is secure because no one can access it without first going through an authentication process.

Note: Check with your SRS administrator about your site policy for permanent sessions. They may need to create an account for you.

Using the Session Manager

If you followed the steps outlined in "Starting a Permanent Session", you will see the "Session Manager" page. See figure 3.

The "Session Manager" page provides you with the tools necessary to manage sessions. You can move queries or views between sessions, save and delete them, and upload or download complete sessions. Additionally, the "Session Manager" page lets you choose the session that you want to work in. These features are all explained throughout the rest of this section.



FIGURE 3. Session Manager Page.

Resume the last Session

The last session that you worked with is displayed when you login as a permanent session user. See figure 3. To continue working on that session:

1. Click the "Resume" button.

The "Top" page is displayed allowing you to begin a new query. Alternatively, you can go to the "Results" page to continue working with the queries you have already done. You can continue to manage the current session (as explained in this chapter) or otherwise continue with your work as usual.

Switch to Another Session

Near the top of the "Session Manager" page you will see a button labelled "Show" with a drop-down menu to the right containing a list of your sessions. This is the permanent session history list.

To return to a session in the history list:

- 1. Select the drop-down menu.
- 2. Scroll down the list till you find the session you want.
- 3. Click the session name indicating your selection.



Resune

 Click the "Show" button to switch to the selected session. Details for the session you have selected are now displayed in the "Session Manager" page. You can work with the session as usual.

Begin a New Session

1. Click the "New Session" button, it as at the top of the options area for the page, see "The "Top" page." on page 5. This will begin a new session. The "Top" page is displayed so that you can start querying databanks right away.

Managing Sessions

There are five session management features. They are:

Сору

New Session

Allows you to copy selected queries or views from one session to another.

Rename

Change the session name.

Delete

Remove a session.

Upload

Upload a session.

Download

Download a session.

Copy

Once a view is made or a query done you can copy them between sessions. This way you can make best use of your work. You do not have to re-invent them in each session.



FIGURE 4. Sharing Queries/Views Between Sessions.

To share a view or query between sessions:

1. Make the session with the queries or views you want to share the current session.

See the "Switch to Another Session" section earlier in this chapter.

- Select the queries and/or views that you want to copy. Click the checkbox to the left of the query or view to indicate your selection.
- 3. Pick the target session.

Near the top of the page you will see a section with "Copy selected views/queries to".

- 1 Scroll through the drop-down menu until you find the session that you want to copy the queries/views to (See figure 4).
- 2 Click the session name to select it.
- 4. Click the "go" button.

The selected queries and views will be copied to the target session.

Rename

The default naming convention (*session1*, *session2*, *sessionN*), is fine when you will only have a few sessions in the history list. If you have several sessions in the history list and are finding it difficult to remember which session has your, e.g., "dog" data and which one has "cat" data, you could give them personalized names.



FIGURE 5. Personalized Session names.

To rename a session:

- 1. Type the new name in the textbox labelled "Rename:".
- Click the "go" button, which is located near the top right of the page under the navigation bar.
 The page will be refreshed with the new name set.

Delete

If you no longer need a particular session you can delete it.

1. Make the session that you want to delete the current session.

See "Switch to Another Session".

2. Click the "Delete" button.

Download

If you want to share a session with other users or if you want to move your temporary session work to your permanent session list, or to another temporary session, you will first need to download a file containing the session work to your local disk.

To download a session:

Download

- Make it the current session. See "Switch to Another Session".
- 2. Click the "Download" button.
- 3. Select "Save this file to disk" when prompted.

File Download	×
	You have chosen to download a file from this location. wgetzf46baacb from www.lionbio.co.uk What would you like to do with this file? O Open this file from its current location O Save this file to disk Always ask before opening this type of file
	OK Cancel More Info

FIGURE 6. "File Download" Dialog Box.

- 4. Click "OK".
- 5. Give the session file a name.

To use the default name provided by SRS skip to the next step.

Save As			? ×				
Savejn: 🗐	[C:]	- 🗈 🖻					
🚞 \$0EM\$	🚞 DELL	🚞 INDEX	🚞 Pro				
🚞 Acrobat3	🚞 DOS	🚞 Multimedia Files	🚞 prc				
adobeapp 📄	🚞 DRVLIB	🚞 Outlook 98 Setup	🚞 PS				
E BACKUP	🧰 ерр	🚞 pdfusers	🚞 Re				
🚞 bioSCOUT	🧰 gifs	🚞 Perl	🚞 Ro				
COLLWIN	🚞 I386	🚞 pkware	🚞 SP				
•			Þ				
File name: wgetz54083d13							
Save as type: Document Cancel							

FIGURE 7. "Save As" Dialog Box.

6. Click "Save".

The file is now saved and ready to upload to another session.

Upload

Upload Session Browse... Work can be shared between different users using the Upload and Download features. You can also use these features with a temporary session to move work from a temporary session into your permanent session history list or vice versa.

To upload a session from your local disk:

- 1. Type the file name into the textbox below the "Upload Session" button.
- 2. Click the "Upload Session" button.

If you do not know where the session file is:

- 1. Click the "Browse" button.
- 2. Locate the file to upload.



FIGURE 8. Browsing for the Session File.

3. Click "Open".

This will put the file name and path into the textbox.

 Click "Upload Session". This will load the session. You can start working with the session as usual.

2.3 Temporary Sessions

If you followed the "Starting a Temporary Session" instructions, the "Top" page will be displayed (see figure 9). You can begin using SRS as usual.

	*
	TOP PAGE QUERY RESULTS SESSIONS VIEWS DATABANKS CHEP
Reset	Quick Search All Entries
Query forms Standard Extended bookmark this link to return to your session If you find problems or have suggestions please mail the SRS administrator	show all collapse all Sequence SWISSPROT PIR EMBLNEW PIR SPTREMBL REMTREMBL GENPEPT NRL3D MGT AGENESEQ NAGENESEQ SeqRelated TransFac User Owned Databanks Application Results Protein3DStruct Genome Mapping Mutations Mutations Metabolic Pathways Others System
	SRS 6.1 <u>feedback</u>

FIGURE 9. The Top Page.

You shouldn't use temporary sessions for anything other than temporary work. The session data will be removed when the system administrator performs a routine cleanup operation.

Managing Temporary Sessions

If you click the "Sessions" tab on the navigation bar, the session manager will be displayed. The "Session Manager" for temporary sessions is different from the "Session Manager" for permanent sessions in many ways.

 Op page
 QUERY
 RESULTS
 SESSION
 Diews
 Diews

 Download Session
 Q4 Q1 & Q2
 Q3 [embl-ALLTEXT: liver*]
 Q2 [embl-ALLTEXT: dog*]

 Q1 [spider-Description: jump*]
 Views
 NiewEmVi

For example, you can't rename, delete, switch between, or share queries and views like you can with permanent sessions.

FIGURE 10. Temporary Session Manager.

You can still upload and download sessions. After completing your session work you may want to save it to disk because the next time you access SRS, the temporary session work might not be available from the server. This is also useful for moving temporary work to your permanent session history list or permanent session work to a temporary session or to share session work with another user.

For upload instructions please see the "*Upload*" entry in the "Permanent Sessions" section of this chapter.

For download instructions please see the "*Download*" entry in the "Permanent Sessions" section of this chapter.

SRS Users Manual

CHAPTER



QUERYING WITH SRS

3.1 The basics of Querying

You have a question, you want to know how many entries there are in one or more databanks, on a particular subject. As you will see throughout this chapter, using SRS to ask that question means that you only need to supply some of the details that you already know. SRS will examine all the entries in that databank and return a list of entries that contain your search term.

For example, using the term "Aldehyde reductase" as the basis for a search against ENZYME, will return a list of entries in the databank that are related to this enzyme. The query method you use will make a difference in the amount of and type of information included in the results list. Throughout this chapter you will learn about the different query methods, the ways to formulate a query, and how to use the various query methods.

Picking a Search Method

There are several ways to search a databank with SRS. They are listed here:

- Quick Search
- Browse Index
- Standard Query
- Extended Query
- Expression Query

Deciding the query type that is best suited to the task at hand can seem somewhat intimidating considering the options mentioned. A brief description is given below for each search option to help you choose the best one. You will learn how to use these in the next section.

Quick Search

The "Quick Search" option on the "Top" page (shown in figure 1), is useful when you want to get a general idea about the type of and approximate amount of information available for a subject. It works by searching all data fields of type "text" for the selected databanks. This is the fastest way to generate query results; there are only three steps from selecting the databanks to viewing the results. There is a possibility with using this search method that the results will have only a cursory relationship with the subject of interest.

Quick Search string

FIGURE 1. The Quick Search Query.

To learn how to use the "Quick Search" method, please skip ahead now to "Doing a Quick Search" on page 41.

Browse Index

Browse specific data field indices when you want to find out how many entries exist for a subject that are available through that data field. Figure 2 shows the "Browse Index" form. The "Keyword" field, for example, is available for many databanks; this field is used to make clear the relationship between an entry and a particular subject.



FIGURE 2. Browse an Index.

If you are interested in the proteins involved in a subject you can search the "Keyword" field of SWISS-PROT for that subject. You might not get all the entries that are related to the subject, but this will give you a starting point. To learn how to browse the indices of a databank with SRS, please skip ahead now to "Browsing Databank Indices" on page 42.

Standard Query

The more details you have about a subject, the better your chances of finding only the entries that are interesting. The "Standard" query form allows you to enter up to four separate search terms against up to four different data fields at once, as you can see in figure 3. Using the "Standard" query form you can search for, a phrase or expression in the "Description" field, a specific "Organism", a "Date" range, and a particular "Author".

If you know of a scientist who published data about "breast cancer" in "humans" at some time over a two year period, enter the details in the query to get a list of related entries.

*	TOP PAGE QUERY	RESULTS	SESSIONS	VIEWS	DATABANKS	HELP
Reset search <u>E</u>	MBL		Info about	field AllTe	xt	_
Subnit Query append wildcards to words I combine searches with AND I display entries	separate mi AliText AliText AliText AliText	ultiple values by	v & (and), (o ng retrieve entries	r), (and n : of type E	ot)	
in chunks of 30	Use predefined vi	iew	*Names only	* 🔻		

FIGURE 3. The Standard Query Form.

To learn how to query a databank with the "Standard" query form, please skip ahead now to the "Using the Standard Query Form" section.
Extended Query

A variation on the "Standard" query form is the "Extended" query form. It lists all the common data fields and allows you to enter a search term for as many fields as you want, you must use at least one field.

The "Extended" query form, shown in figure 4, is similar to the "Standard" query in that they are both forms that can use several search terms for the query.

TOP PAGE	QUERY RE	SULTS SESSIONS VIEWS DATABANKS	Đ
Reset search EM	<u>BL</u>	Info about field AllText	T
Submit Query	56	eparate multiple values by & (and), (or), ! (and not)	
append wildcards to words	Use view Na	mes only* 💌 sequence format embl 💌	
combine searches 	Field Name	Query	Include in List
AND I display entries in	<u>AllText</u>		
chunks of	Ð		
Standard	T	□ est □ fun □ gss □ htg □ hum □ inv	_
query form	Division	□ mam □ org □ phg □ phn □ pro □ rod □ sts □ syn □ unc □ vrl □ vrt	
Make default query page 🗖	<u>AccNumber</u>		
	SeqVersion	>= ▼	

FIGURE 4. The Extended Query Form.

They are different in a many ways. For example, the "Extended" query allows you to enter a search term for more than four data fields. This means that you can further limit the results of a query.

The data fields in the "Standard" query form can be arranged in any order that you want. The "Extended" query form lists the data fields that are available in a static way. The "Extended" form makes numeric range queries easier. The different ways of searching for numeric range data are all displayed in a drop-down menu. You only need to pick the one you want and enter the range details.

Finally, if a field has a limited number of search terms available, like the "Division" field of EMBL. They will all be displayed in a group in the "Extended" form, (see figure 4).

Expression Query

Using the syntax described for the SRS Query Language you can write your query in the "Expression" box, found in the "Results" page, (see figure 5).

Expression	Q1 < Q2	*
	4	

FIGURE 5. The Expression Query.

To learn how to query a databank using the "Expression" box please skip ahead now to the "Doing an Expression Query" section. To learn more about the SRS query language see the "SRS Query Language" chapter.

Constructing Search Terms

You can enter a query in the "Quick Search" input field of the "Top" page or you can use one of the query forms. The query forms make searching a databank very easy because they set out in an ordered way the data fields that you can search against. The "Standard" query form lets you mix and match data fields to create your own query.

Completing an "Extended" query form is as easy as completing a questionnaire where you answer only the questions that you want. Regardless of the query option you have selected, you will still have to make a choice about the way to construct your search term.

There are several ways to construct a search in SRS, they are listed here:

- Single-Word Search
- Multi-Word Search
- Number
- Regular Expressions

Single-Word Search

When you search a databank for a single word in a single field you will get, as the results of your query, a list of entries that match that word in the selected data field. To increase the number of entries that are returned you can search for the word in more fields.

For example, when you look for "reductase" in the "AltName" (alternate name) field of ENZYME you will get all the entries that include "reductase" as part of the name. If you also search for "reductase" in the "Description" field, you will get a listing of entries that have the word "reductase" in the "description" as well as those that have it in the "AltName" field. The "combine search words with 'OR'" option must be used. There is more information about query operators later. See figure 6.

AltName	• [eductase
Description	 	eductase

FIGURE 6. Single-Word, Multi-Field Phrase.

The "Quick Search" option combines all the fields with a data type of text using the "OR" combination operator.

There will be some overlap because some of the entries that include the word in the "AltName" field will also include it in the "Description" field (or some other field). Regardless of the number of times the query hits an entry, SRS will include it in the results listing only once.

Note: Ranking of hits is not dependent on the number of hits for an entry.

Multiple-Word Phrases

You can search for a phrase such as "Aldehyde reductase" in many ways, depending on the search method and query operators you use. For example, if you type "Aldehyde reductase", with quotes, SRS will search for the complete string "Aldehyde reductase". But without the quotes, SRS will search for entries that have both "Aldehyde" and "reductase" or the complete string "Aldehyde reductase". You can make explicit the relationship between the words by including an operator in the string, for example, "Aldehyde & reductase" (AND), or "Aldehyde | reductase" (OR), or "Aldehyde ! reductase" (BUTNOT).

TOP P/	AGE QUERY	RESULTS SESSIONS VIEWS DATABANKS	HELP
Reset search EN	IZYME	Info about field AllText	•
Submit Query	sepa Use view *Nar	arate multiple values by & (and), (or), (and not) nes only*	
combine searches with AND	Field Name	Query	Include in List
display entries in chunks of 30 🔽	<u>AllText</u> ID	Aldehyde reductase]

FIGURE 7. Multi-Word, Multi-Field Phrase.

If you are using one of the query forms and split the phrase into two textboxes, the relationship needs to be made explicit by using one of the options in the "combine searches with" drop down menu. It is set to "AND" by default.



FIGURE 8. "Combine Searches with" Drop-down menu.

Numbers and Regular Expressions

You are not limited to words and phrases when querying a databank with SRS. You can search for numeric data like dates and you can search a databank using a regular expression.

Numbers. You can look for entries that have, for example, a sequence length "SeqLength" (see figure 9), ":n" (less or equal), ":!n" (less), "!n:" (greater), "n:" (greater or equal), to something. Here is how to use it.

```
12:15
Greater than or equal to 12 less than or equal to 15.
12:
Greater than or equal to 12.
!12:
Greater than 12 (but not 12).
:12
Less than or equal to 12.
:!12
Less than 12 (but not 12).
```

You can write a numeric query in the "Expression" box like this:

'[embl-s1#12:15]'

it tells SRS to look for all entries in EMBL that have a "SeqLength" that is greater than or equal to 12 and less than or equal to 15.

The numeric range query is written using the SRS query language syntax. You can also write this same query using the "Standard" or "Extended" query form. See figure 9. When done in the "Extended" query form you do not need to use the query language syntax.

The difference between doing the query with the "Extended" query form and using any other query option is that the "Extended" query form lets you select, from a drop-down menu, the significance of the values.



FIGURE 9. Numeric Data Query.

Regular Expression. If you are unsure of the spelling of a search word you could use some combination of characters along with regular expression characters and get a list of matching entries as your result. For example, "/ca.*r/" will include "cancer" and "carter", etc., in the results. You can also apply controls to the regular expression that will limit the type of search it performs, thus saving a lot of time for the query. You need to include the "/" character at the start and end of the regular expression string.

Wildcards. You can apply the SRS query language wildcard to the search word if you would rather not use regular expression syntax. Using the query language the search term above becomes "ca*r". Read more about the query language in chapter 8, "SRS Query Language".

3.2 Querying

This section contains a tutorial allowing you to build on what you've already learned. The first thing you need to do before continuing is

begin an SRS session. A temporary session will be sufficient for the purposes of this chapter. Click the "Start" link from the "Start" page.

Note: Go to the "Starting an SRS Session" chapter now for information on the session options available.

Doing a Quick Search

The "Quick Search" was defined earlier in this chapter; it allows you to search the selected databanks using the "AllText" data field. The following image shows the top part of the "Top" page. Near the centre of this page you will see a button labelled "Quick Search" with an input field to the right of the button.



FIGURE 10. Quick Search Query.

To do a "Quick Search":

- 1. Select the databanks that you want to search. Click the checkbox to the left of the databank name.
- 2. Enter the search term in the textbox. Use a word or a regular expression, as discussed earlier.
- 3. Click the "Quick Search" button.

For this example we will search the SWISS-PROT databank for entries that have the word "cancer".



FIGURE 11. Quick Search Query.

Steps:

- 1. Select the checkbox to the left of the SWISSPROT hyperlink.
- 2. Type "cancer" into the "Quick Search" query textbox.
- 3. Click "Quick Search". See figure 11.

The system will search the indices for the fields that belong to the "AllText" virtual field in SWISS-PROT and return a list of matches.

Browsing Databank Indices

You can browse indices from the "Data Field Information" page.

To get to a "Data Field Information" page:

From the Standard Query Form

There is a drop-down list on the "Standard" query form of all data fields available for a particular databank, select the data field that you wish to browse and click the "Info" button.

Info about field AllText

FIGURE 12. Info Button and Drop-down menu.

From the Extended Query Form

Each of the data fields on the "Extended" query form are highlighted indicating that they are hyperlinks. Click on the name of the field that you want to browse to go to that fields information page.

Ð	
<u>AccNumber</u>	

FIGURE 13. Go to the Information Page.

From the Databank Information Page

You will see a section devoted to data fields for that databank. Select one of those data field hyperlinks to go to its "Data Field Information" page.

Data-fields in SRS	Name	Short Name	Туре	No of Keys	No of Entry References	Indexing Date	Status
	<u>AllText</u>	<u>all</u>	group	0	0		not indexed
	Ð	id	id	80000	80000	03-Nov-1999	ok
	AccNumber	acc	index	88425	88498	03-Nov-1999	ok
	Description	<u>des</u>	index	45468	450863	03-Nov-1999	ok
	<u>GeneName</u>	gen	index	50276	92360	03-Nov-1999	ok
	<u>Keywords</u>	<u>key</u>	index	825	269082	03-Nov-1999	ok
	<u>Date</u>	<u>dat</u>	num	38	80000	03-Nov-1999	ok
	<u>Organism</u>	org	index	14257	805167	03-Nov-1999	ok

FIGURE 14. The Databank Information Page.

For the example below we will search the SWISS-PROT databank for "hist*" in the "Description" data field.



FIGURE 15. The "Data Field Information" page.

Steps:

- Select the SWISSPROT link from the Databanks page or the Top page. This will bring up the "Databank Information" page for SWISS-PROT.
- Select the data field that you want to browse.
 This will bring up the "Data Field Information" page.
- 3. Enter the search term, "hist*", in the input field.

Note: Wild cards are allowed when browsing the data field but they are not automatically appended to the search term like with other query methods.

4. Click the "List Values" button.

This will take you to the "Browse Index" page.

List Values

4	TOP PAGE QUERY	RESULTS SESSIONS	
Reset			
More Values	S Value Values in SWIS	No of Entries ISPROT esis 174	
Make Query	🗖 <u>histidine metaboli</u> :	<u>sm</u> 20	

FIGURE 16. Browse the Index.

 Select the terms that best match your needs and click the "Make Query" button. This returns entries that match the selected terms.

Using the Standard Query Form

Query EMBL for entries about "insulin" that have been added to the databank before "January 1, 1990" using the "Standard" query form.

NEED	∜ NE		e query	RESULTS	SESSIONS	VIEWS	DATABANKS	HELP	
	Reset	search	<u>EMBL</u>			Info about	field AllText		•
	Submit	; Query	sepa	ırate multipi	le values by &	'e (and), (01), (and not)		
	app) end	AllText	•	insulin				
	to woi	rds 🗹	Date AllText		1-Jan-1998				

FIGURE 17. The "Standard" query form.



- 4. Enter the search phrase. In the main body of the "Standard" query form are four textboxes where you enter the query string. To the left of each textbox is a data field selector. Specify the query string as follows:
 - 1 Enter the word "insulin" in the first input field.
 - 2 The age constraint goes in the next input field. This will be formatted according to the SRS query language rules. Type ":1-Jan-1990" and change the data field selection to "Date".
- 5. Ensure that the "Combine searches with" option is set to "AND".
- 6. Click the "Submit Query" button.

In addition to specifying the number of entries SRS should display, you can define a view for the results. See the "Views" chapter to learn more.

Note: The databanks to search were chosen in the "Top" page. If you want to change the list of databanks to search go back there now to change your selections.

Using the Extended Query Form

Extended

In an almost interview format the "Extended" query form allows you to enter search constraints or criteria about any of the data fields that are available. You just pick the databank or databanks that you want to search and start filling in the blanks. This time we will search the "Description", "Organism", "Date", and "Keyword" fields. (Use more if you want).

Steps:

- 1. Click the "Top" page tab.
- 2. Select the databank EMBL.

 Click the "Extended" button. This takes you to the "Extended" query page. See figure 18.

TOP PAGE	QUERY RE	SULTS SESSIONS VIEWS DATABANKS	HELP
Reset search EM	<u>BL</u>	Info about field AllText	•
Submit Query	S	eparate multiple values by & (and), (or), ! (and no	t)
append wildcards to words 🔽	Use view 🕅 Ne	ames only* 💌 sequence format embl 💌	
combine searches with	Field Name	Query	Include in List
display entries in	<u>AllText</u>		
chunks of 30 💌	Ð		
Standard query form	Division	□ est □ fun □ gss □ htg □ hum □ inv □ mam □ org □ phg □ ph □ pro □ rod □ sts □ syn □ unc □ vrl □ vrt	
Make default query page 🗖	<u>AccNumber</u>		
	SeqVersion	>= ▼	
	NID		

FIGURE 18. Extended query form.

- 4. In the "Description" field enter "histamine".
- 5. In the "Keywords" field enter "histamine".
- 6. In the "Organism" field enter "human".
- 7. In the "Date" field pick the "<=" option. Enter "1-Jan-1990".

Description	histamine	
<u>Keywords</u>	histamine	
<u>Organism</u>	human	
Authors		
Title		
Citation		
Date	>= 💌 1-Jan-1998 <= 💌	W IMAGE
<u>SeqLength</u>	>= v	

FIGURE 19. Completed fields.

8. Click the "Submit Query" button.

Doing an Expression Query

You can use the "Expression" query box to combine, link, or refine the results of existing queries.



FIGURE 20. Expression query box.

Steps:

- In the "Expression" query box, type "Q1 & Q2". This will filter through the results of "Q1" and "Q2" and return a new listing, "Q3", that has all entries that are in both "Q1" and "Q2".
- 2. Type "Q3 < SWISSPROT", to begin a linking operation.



FIGURE 21. Linking with an Expression

This will create Q4 which contains a listing of all the entries in Q3 that are also linked to the SWISS-PROT databank.

СНАРТЕК



LINKING

After querying a databank with SRS you may want to know more information about some of the entries. For example, a query of EMBL may return several DNA sequences about a subject, but you may only be interested in those sequences that use a specific protein. To refine your results you could link them to SWISS-PROT or link them with a set of SWISS-PROT entries from another query. The best way to understand a new concept is to dive in and start doing it.

 Query EMBL for accession number "M18415". The EMBL entry will be displayed in the "Entry List" page. See figure 1.



FIGURE 1. The selected entry will be linked.

Linked data in the entry:	2. 3.	Click on the hyperlink to view the entire entry. Scroll through the entry until you find the SWISS-PROT hyperlink with the unique identifier (the accession number) "P08913". This hyperlink represents an SRS link.
	4.	Click the hyperlink to view the entry.
	5.	Click your browsers "Back" button to return to the "Entry List" page.
Using the "Link" page:	6.	Click the "Link" Button . This will bring up the "Link" page. See figure 2.



FIGURE 2. The SRS Linking Page.

- 7. Choose SWISSPROT. Select the checkbox to the left of the SWISSPROT hyperlink.
- Click the "Submit Link" button. The default settings are used for this example.

You're done! You have just successfully completed a linking operation. There is a lot more to linking, however, and this little example only served to scratch the surface.

4.1 What is Linking?

SRS finds entries in other databanks that have a cross-reference relationship with entries in your query or vice versa. SRS finds these cross-references and either returns the set of entries in the linked databanks, or returns a subset of the original entries that have (or that specifically do not have any) links with the selected databanks.

What is a Cross-Reference

Two kinds of cross-references exist in a databank entry, explicit and implicit. They can both be found in many places in the entry.

The explicit cross-reference might be defined in the form of "databank; unique identifier", where the unique identifier is often the accession number. For example, you looked through the EMBL entry in the previous example and found "swissprot; P08913;" with the accession number as a hyperlink to the actual entry.

An explicit cross-reference exists because someone, either the originator of the entry or the databank maintainer, has recognized the relationship that the entry has with entries in other databanks and added it.

Implicit cross-references may be derived from organism or small compound names. Unfortunately, implicit references are harder to deal with due to the lack of standards for identities such as organism names.

4.2 Picking a Linking Option

There are three options for linking sets of entries and one option for linking a single entry.

The options for linking sets of data entries are shown on the "Link" page under the heading "Find all Entries". They are explained here.

In the selected databanks that are linked to the current query.

Returns entries from other databanks that have a cross-reference with entries in the current query.

In the current query that are linked to all selected databanks.

You are refining the original query, limiting it so that it includes only the entries from the original query that are also linked to the databanks selected.

In the current query that are not linked to any of the selected databanks. This is another limiting operation. Only entries that do not link to the specified databanks will be included with the results.

You used the first option in the linking operation at the beginning of this chapter. The results of your linking operation were a set of entries from SWISS-PROT.

4.3 Initiate A Linking Operation

It is possible to initiate a linking operation from many pages, for example:

- the "Results" page
- the "Entry List" page
- the "Entry" page

are just a few.

Link Whole Queries

You can link an entire set from the "Entry List" page or from the "Results" page. To link from the "Entry List" page: do your query as normal, when the "Entry List" page appears click the "Link" button to bring up the "Link" page. Finish the linking operation as you did in the previous example.

There are two ways to link your queries from the "Results" Page. The most common is to select the checkbox that corresponds to a query set and click the "Link" button. This takes you to the "Link" page where you specify the type of link to perform. The second option is to use the "Expression" box near the top of the "Results" page. This option is discussed later in the "Expression Linking" section.

Using the "Link" Page:

 Select the query to link. In the "Results" page you will see a listing of all the queries performed during the current session. See figure 3.

Т	OP PAGE	QUERY	RES	SULTS SESSIO	NS	VIEWS DATABANKS	(HELP)
Reset				Expression	4		× V V
Perform operation on					Suco	cessful Queries	
selected queries	Name	Туре	N Total	From Library	N	Query Expression	Comment
Delete Link	□ Q3	link	1	SWISSPROT	1	(([EMBL-ID:HSADRA])	<u>т</u>
View *Names only*	₩ Q2	select	1	EMBL	1	[EMBL-ID:HSADRA]	7
Combine	🗆 Q1	query	1	EMBL	1	[embl-AccNumber: m18	(A)

FIGURE 3. Select the Query to Link.

- Click the "Link" button. This will bring up the "Link" page.
- 3. Pick the type of linking operation to perform. See "Picking a Linking Option" earlier in this chapter.
- 4. Choose the databanks to run the linking operation against. Select the checkbox to the left of the databank name.
- 5. Click the "Submit Link" button.

Multiple Entry Linking

Link several entries from a query by selecting only the entries that you want to perform the linking operation on (or selecting the entries to exclude from the linking operation).

Linking "selected" entries from the "Entry List":



- 1. Select all the checkboxes for the entries that you want to link. The "Perform Operation" option is set to "On Selected" automatically when you begin selecting entries to link.
- 2. Click the "Link" button. This will bring up the "Link" page.
- 3. Choose a linking operation (or use the default option). See the "Picking a Linking Option" section earlier in this chapter.
- 4. Select the databanks.
- 5. Click the "Submit Link" button.

Linking "all but selected" entries from the "Entry List":

You can also link most of the entries in a set, just select the ones that you do not want included in the linking operation and change the "Perform Operation" selection to "on all but selected".

- 1. Select all the checkboxes for the entries that you do not want included in the linking operation.
- 2. Set the "Perform Operation" option to "on all but Selected".
- 3. Click the "Link" Button.
- 4. Pick the type of linking operation.
- 5. Select the databanks.

6. Click the "Submit Link" button.

Linking a Single Entry

Before you can link a single entry you need to bring up the "Entry" page for that entry.

- 1. Click the "Results" tab.
- 2. Choose the set that contains the entry you want to link.
- 3. Select the checkbox to the left of the query set.
- Click the "View" button. This will bring up the "Entry List" page for the query set.
- 5. Click the hyperlink for the entry that you want to link. This will bring up the "Entry" page. See figure 4.



FIGURE 4. Link the entry.

To link a single entry from the Entry Page (once the "Entry" page is displayed):

 Click the "Link" button. This brings up the "Single Entry Linking" page. See figure 5.



FIGURE 5. Linking a single entry.

- 2. Select the databanks to perform the linking operation on.
- 3. Click the "Submit Link" button.

You do not have the options with a link from the "Entry" page that you have when you link a single entry from a list of entries or when you link a complete set of entries. You are limited to linking the entry to other databanks.

Expression Linking

If you want to link between two sets of results, a set of DNA sequences in "Q1" and a set of protein sequences in "Q2", for example, you would use the "Expression" box.

The question you are asking might be, 'are there any links between' 01 and 02

If there are links between entries in the queries you have the choice of having the entries in "Q1" that are linked to "Q2" returned as the results or of having the entries in "Q2" that are linked to "Q1" returned as the results.

The "Expression" box is shown in figure 6. Enter your expression in the textarea as is indicated in step 4 below.



FIGURE 6. Expression Linking.

To link complete query sets using the "Expression" box:

- Key in the name of the query set to link.
 Q1
- 2. Specify the type of query (See example below).
- 3. Specify the databank or second query that the set should link to. Q^2
- 4. Click the "Query Expression" button.

Q1 < Q2will list all the entries in "Q1" that are linked to "Q2".

Q1 > Q2will list all the entries in "Q2" that are linked to "Q1".

Linking Operators

The linking operators are explained here:

<

Entries to the left of the operator are returned in the "Results List" if they have a link to any entries in the set or databank to the right of the operator.

>

Entries to the right of the operator are returned in the "Results List" if they have a link to any entries in the set or databank to the left of the operator.

You can also use the query operators, as is shown in the following table, with a linking operation. For example, "Q1 < Q2 & Q3" will return a set of entries from "Q1" that have links to "Q2" and "Q3".

Operators	Example	Returns Entries in:
<	"Q1 < Q2"	"Q1" that have link to "Q2".
&	"Q1 < Q2 & Q3"	"Q1" that link to "Q2" and "Q3".
	"Q1 < Q2 Q3"	"Q1" that link to "Q2" or "Q3".
!	"Q1 < Q2 ! Q3"	"Q1" that link to "Q2" but not "Q3".
>	"Q1 > Q2"	"Q2" that have link to "Q1".

TABLE 4.1Linking operations.

Linking Examples

Linking through the "Expression" box is a powerful alternative to the standard linking methods discussed earlier. In addition to linking between queries you can use the "Expression" box to check for databank links without going through the "Link" page.

If you have a set of EMBL entries in "Q1" that you want to link with SWISS-PROT, type:

Q1 < SWISSPROT

To see the EMBL entries from the original query that have links to SWISS-PROT.

If you would rather see the SWISS-PROT entries that the above operation returned turn the link around, like this:

Q1 > SWISSPROT

Which returns the SWISS-PROT entries that have links with the entries in the "Q1" set.

SRS Users Manual

СНАРТЕК



VIEWS

5.1 Overview of Views

Views allow you to see only the data that you are interested in. A view can be created to work for only one databank, or for multiple databanks. It can contain data from the databank it is applied to, and to data from linked entries in other databanks.

This chapter is split into "Using Views" and "Creating Views". You learn how to use views with data in the first section, this includes the predefined views and the custom views. The second section shows you how to create custom views.

5.2 Using Views

SRS comes pre-configured with certain default views. The SRS Administrator can edit these default views or create new views in the default set. Furthermore, users can create their own views. User created views are saved as part of the session file, enabling the user to save and reuse their view in the future.

Views can be applied from any page that creates or holds a query, such as:

- The "Standard" query form.
- The "Extended" query form.
- The "Entry List" page. (The results of an individual query).
- An "Entry" page.
- The "Results" page.
- The "Link" page.
- The "Application Launch" page.

How to Apply Views

Following are instructions for applying views to your query results. Each is given as a mini tutorial. In the first two you will begin by querying a databank and setting the query options for views so that the query results are displayed with your chosen view. The others show how to apply to an already existing set.

From the "Standard" query form

To apply a view from the "Standard" query form:

- 1. Select the EMBL databank checkbox from the "Top" page and click the "Standard" button. See chapter 3, "Querying" if necessary.
- 2. Enter "oxidase" in the input box. The default "AllText" data field is fine.
- 3. Select the view you would like from the pull-down menu beside the heading "Use predefined view". See figure 1.

Use predefined view	*Names only*
Create your own view	*Names only* *Complete entries*
Display in O List © Table	SequenceSimple FastaSeqs k SeqVersion NID Molecule Description ▼
sequence format embl 🔽	

FIGURE 1. Choose a predefined view for your "Standard" query.

4. Click on Submit Query.

This will bring up the "Entry List" page with the results of your query displayed using the selected view. See figure 2.

Reset Query "[embl-AllText: ascorbate* & oxidase* ascorbate oxidase*] " found 31 entries					
Perform operation	RootLibs	acc	des	<u>sl</u>	
• on all but selected • on selected Link Save	□ <u>EMBL:AA840724</u>	AA840724	CFR4 Fruit cDNA library of Hot pepper Capsicum annuum cDNA clone CFR4 5' similar to ascorbate oxidase homolog (X81376).	396	
	EMBL:ATTS4167	Z37702	A. thaliana transcribed sequence; clone VBVKB03; 5' end; Similar to L- ASCORBATE OXIDASE PRECURSOR HOMOLOG ; Brassica napus.	413	
View SequenceSimple	EMBL:ATTS4945	F14023	A. thaliana transcribed sequence; clone VBVZB07; 5' end; similar to L- ASCORBATE OXIDASE PRECURSOR; Brassica napus.	363	
Launch	EMBL:ATAB4798	AB004798	Arabidopsis thaliana mRNA for ascorbate oxidase, partial cds.	2113	
OligoBlast 💌	EMBL:BNBPDNA	X64257	B.napus gene Bp10	3562	
show entries in chunks of	EMBL:CMAAO	D55677	Cucurbita maxima AAO gene for ascorbate oxidase, complete cds.	4267	
30 T	EMBL:CMAAOP	D45066	Cucurbita maxima mRNA for AOBP (ascorbate oxidase promoter-binding protein), complete cds.	1485	
Transa Trandig	EMBL:CMA01	Y10224	C.melo ao 1 gene, partial	3921	

FIGURE 2. The "Entry List" page using the selected view.

From the "Extended" query form

To apply a view for an "Extended" query:

- 1. Select the EMBL databank checkbox from the "Top" page and click the "Extended" button. See chapter 3, "Querying" if necessary.
- 2. Enter "oxidase" in the "AllText" input box.
- 3. Select the view that you want to apply to your query results from the drop-down menu beside "Use view". See figure 3.



FIGURE 3. Choose a view for your "Extended" query.

4. Click the "Submit Query" button.

This will bring up the "Entry List" page with the results of your query displayed using the selected view.

From the "Entry List" page

Querying, linking, and launching applications always returns the "Entry List" page. It contains a list of the entries found that meet the requirements specified in the operation. If you do not select a view for the results when the operation is run the default "Names only" view is used. You have already seen how to override the default view for the "Standard" and "Extended" queries, later you will see how to override the default view for the "Linking" and "Application Launch". Now, you will see how to apply a different view to the "Entry List" page when you want to change views.

To change views in the "Entry List" page:

In the left side-margin you will see a drop-down menu below the "View" button. See figure 4. The drop-down menu displays the name of the view currently applied.

Reset	
Perform operation	□ <u>EMBL:AA840724</u> □ EMBL:ATTS4167
• on all but selected • on selected Link Save View	EMBL:ATTS4945 EMBL:ATAB4798 EMBL:BNBPDNA EMBL:CMAAOD EMBL:CMAAOD
Names only	EMBL:CMA01
Names only *Complete entries*	EMBL:CMAO2
SequenceSimple	EMBL:CMAO3
FastaSeqs	EMBL:CSASCO

FIGURE 4. Changing views in the "Entry List" page.

- 1. Select the drop-down menu.
- 2. Scroll through the list until you find the view you want.
- 3. Click the view name to select it.
- 4. Click the "View" button.

The "Entry List" page will be replaced with a new "Entry List" that displays the results using the selected view. See figure 5.

Reset	<i>Query</i> "[embl-AllText: ascorbate* & oxidase* ascorbate oxidase*] " found 31 entries
Perform operation	EMBL:AA840724
 on all but selected on selected Link Save View FastaSeqs	>AA840724 gctcgtgcgtcttagggtttaacgtttgcttaactcctcgatatcttcttcttcagagc aattactatgggcaagtgctatcccactgtgagcgaggaatacctcaaggctgttgacaa atgtaaaaggaaactcagaggatcattgctgagaagaattgtgctcctcttatgctccg tcttgcatggcactcggtggtacctatgatgtgtgtgctccaaaactggaggtcctttgg taccatgaggttcaaaactgagcaattcacggagcaaacaatggtattgacattgctc caggctcttggagcattaggggcagtttcccctctcctatgctgattctacaa ttggctggtgttgttgctgtggtgtgtggggga

FIGURE 5. The new view in the "Entry List" page.

From the "Entry" page

To apply a view from the "Entry" page:

1. You need to be on an "Entry" page so start with the "Entry List" page from the previous example and click the hyperlinked entry name for one of the entries. See figure 6.

□ <u>EMBL:AA840724</u>

FIGURE 6. Click the entry hyperlink.

- 2. Select the drop-down menu that is above the entry and beside the "View" button.
- 3. Scroll through the list until you find the view that you want to use for the entry. See figure 7.

Reset	Vieu * Complete entries * 💌	
This entry is from: <u>EMBL</u>	ID AA84072 * Complete entries * EST; 396 BP. XX SequenceSimple AC AA84072 FastaSeqs	
Save	SV AA840724.1 XX	
Launch	NI g2924163 XX	
OligoBlast	DT 10-MAR-1998 (Rel. 54, Created) DT 10-MAR-1998 (Rel. 54, Last updated, Version 1) XX	

FIGURE 7. Choosing a view.

- 4. Click the view name to select it.
- 5. Click the "View" button.

The "Entry" page is replaced with an "Entry" page that uses the selected view. See figure 8.

Reset	View SequenceSimple				
This entry is from:	RootLibs	acc	des	<u>sl</u>	
EMBL Save	EMBL:AA840724	AA840724	CFR4 Fruit cDNA library of Hot pepper Capsicum annuum cDNA clone CFR4 5' similar to ascorbate oxidase homolog (X81376).	396	
Link Launch OligoBlast	,	<u> </u>			

FIGURE 8. The new view.

From the "Results" page

To apply a view from the results page:

- 1. Select the set that you want to apply the view to. Click the checkbox to the left of "Q1".
- 2. Select the view from the pull-down in the left margin. See figure 9.

Reset			l	Expression	 ◀]	×
Perform operation on selected queries Save	Name	Туре	N Total	From Library	N	Successful Queries Query Expression	Comment
Delete Link View * Names only * * Complete entries * SequenceSimple FastaSeqs	₩ Q1	query	31	EMBL	31	[embl-AllText: ascor	

FIGURE 9. Select the view for your set.

3. Click the "View" button.

From the "Link" page

To apply a view for the results of a linking operation from the "Linking" page:

1. Click the "Results" tab.

This will bring up the "Results" page.

- 2. Select the checkbox to the left of "Q1".
- Click the "Link" button.
 This will bring up the "Link" page.

4. Choose the databank to look for links in.

Select the checkbox to the left of the SWISSPROT hyperlink. See figure 10.

Reset Current qu	ery: "[embl-AllText: ascorbate* & oxidase* ascorbate oxidase*] "
S	Find all Entries Image: Set Db Find all Entries Image: Set Db Find all Entries Image: Set Db in the selected databanks which are linked to the current query Image: Set Db in the current query which are linked to all selected databanks Image: Set Db in the current query which are not linked to any of the selected databanks
Submit Link	show all 🔹 collapse all 🗖
select chunk size 30 view with *Names only* *Complete entries* SequenceSimple SwissView FastaSeqs proteinChart tProteinChart BlastAlignment BlastAlignment BlastPSwissProsEnz BlastPFamilies	 Sequence EMBL EMBLNEW EMBLTEST GENBANK SWISSPR SWISSNEW PIR SPTREMBL REMTREMBL GENPEPT REMPIR NRL3D IMGT AAGENESEQ NAGENES OGLYC SeqRelated TransFac User Owned Databanks Application Results Protein3DStruct Mapping Mutations Metabolic Pathways Others System

FIGURE 10. Pick a view for links.

- 5. Choose the view that you want to apply to the results of the linking operation, click the drop-down menu in the left-side margin under the "view result with" heading.
- 6. Click the "Submit Link" button.

From the "Application Launch" page

To choose a view for the results of an application other than the default view:

1. From the "Results" page pick the application to run and click the "Launch" button.

See chapter 6, "Applications" for more information on launching applications.

- 2. You will see a drop-down menu near the heading "select view to display results". Select this drop-down menu.
- 3. Scroll through the list until you find the view that you want. See figure 11.



FIGURE 11. Choose a view for the application results.

4. Launch the application as usual.

The results of the application will be displayed in the "Entry List" page using the view you selected.

5.3 Creating Views

The predefined views may not show the information that you actually want to see. You can create your own views using the "View Manager" or from within the query forms.
Creating Views with the Query Forms

Both the "Standard" and "Extended" query forms provide a mechanism for specifying the data fields that should be displayed in the "Entry List" that is returned by the query.

Creating Views from the "Standard" Query Form

The bottom half of the "Standard" query form has a section called "Create your own view", that lets you specify how the results of the query should be displayed. You can either use the default view, one of the predefined views (as was explained in the previous section), or you can design your own view. See figure 12.





To design your own view:

1. Select the data fields in the list that you want displayed.

Note: Refer to your browsers documentation for help selecting multiple fields if required.

- 2. Choose the layout that you want from the "Display in" section. Your choices are "List" or "Table".
- Complete the query as normal. The results will be displayed in the "Entry List" page as specified.

Creating Views from the "Extended" Query Form

To the right of the input field, (see figure 13) is a checkbox that is used to tell SRS to include a field in the results. You do not have to enter a search term in a field to have it included in the results.

*	TOP PAGE QUI	ERY RESULTS SESSIONS VIEWS DATAB	ANKS HELP
Depet	/RI	TEER 1	
Reset search		about held All ext	
Submit Query	s	reparate multiple values by & (and), (or), ! (and no	ot)
append wildcards to words 🔽	Use view Ne	ames only* 💌 sequence format embl 💌	
combine searches with	Field Name	Query	Include in List
AND display entries in	AllText		
chunks of 30 💌	₽		v
Standard		\Box est \Box fun \Box gss \Box htg \Box hum \Box inv	
query form	<u>Division</u>	□ mam □ org □ phg □ phn □ pro □ rod □ sts □ svn □ unc □ vrl □ vrt	
Make default query page 🗖	<u>AccNumber</u>		

FIGURE 13. Select the fields to include in the "Entry List".

To create a view in the "Extended" query from:

- 1. Enter your search term as usual. See chapter 3. "Querying" for details.
- 2. Select the checkbox to the right of the data fields for any fields that you want displayed in the results.
- 3. Continue with your query as usual.

Creating Views with the "View Manager"

You can create new views easily if the built-in views do not do what you need. Click the "Views" tab to open the "View Manager" page. You can control the following elements of views that you create:

- A list of databanks the view is applied to (and therefore available for). These are called "Root Libraries".
- A list of databanks that may have a link with the entries in the "Entry List".
- A list of fields from the "Root Libraries" and the linked databanks that should be displayed.
- The format of some fields such as "Sequences".
- Whether the view is displayed as a list or a table.
- The name of the view.

In selecting the "Root Libraries" for a view, you specify two important things. First, the view will only be available for queries and result sets of a subset of these databanks. Secondly, the available fields from the "Root Libraries" will be those that are common across all of the "Root Library" selections. Therefore, it is useful to select only databanks that are close to each other when you are creating a new view.

To Create a new View

- 1. Click on the "Views" tab.
- 2. Select the databanks for which the view will be available in the left column of databanks. These are the "Root Libraries" for the view.
- 3. Select the link databanks.
- Name the view.
 Enter the name in the "View name:" input field.
- 5. Click on the "Create New View" button. See figure 14.

С Р	AGE QUERY RESULTS	SESSIONS VIEWS DATABANKS	HELP
Reset			
View name: Display view as C list C table In table headers use C full field names C short field names	Delete View Select w Edit View Name Create New View Select databanks to define a view for	iew to edit or delete: as only* Select databanks to be linked to displayed entry	
Select fields from all fields in libraries just common fields	EMBL EMBL_features EMBLNEW EMBLNEW_features EMBLTEST EMBLTEST_features GENBANK_features SWISSPROT SWISSPROT_reference SWISSPROT_comment	EMBL_features EMBLNEW EMBLNEW_features EMBLTEST EMBLTEST_features GENBANK_features SWISSPROT_reference SWISSPROT_comment	

FIGURE 14. Create your own view with the View Manager

You will now have a page to select the specific fields for your view. The top box contains the common fields among *all* of the "Root Libraries" you selected in the previous page. There are additional boxes below this for each linked database you selected on the previous page. From these boxes, select the fields that you would like to see in your view by checking the appropriate checkbox.

On this page you can also control the following:

- The view layout, it can be displayed as a "Table" or "List".
- The view name.
- If "full" or "short" field names are used in the table headings.
- 6. Click on Save View.



FIGURE 15. Select the data fields to include in your view

Finished! Your new view is ready to be used on any of the databanks indicated in the "Root Libraries" from step 2 above. Try your new view by performing a new query on one or more of the "Root Libraries" in your new view.

SRS Users Manual

CHAPTER



APPLICATIONS

6.1 Introducing Applications

You can launch an application on data you found in one of the databanks or with your own data. Application results are stored as user databanks. You can query the application results just like any other databank, and if the application was a similarity search against one of the databanks available in SRS, the results may be linked to those databanks.

Additionally, the results of an application can be used by another application. "ClustalW" takes the results of other applications, "BlastP", for example, and performs an alignment on those results. The "Hmm" family of applications takes results from other applications or raw data and searches that data.

For example, you could begin with "BlastP", run the results through "ClustalW", and run a Hidden Markov Marker ("Hmm") on those results.

6.2 Using Applications

This section contains a tutorial for working with applications in SRS.

1. Query EMBL for the entry with the accession number "M18415".

If you need help with this go to the "Querying" chapter.

2. Select the applications drop-down menu from the "Entry List" page.

The list of applications for DNA sequences is shown in figure 1.



FIGURE 1. Select an Application to run.

- Click an application name.
 For this example click "BlastN".
- 4. Click the "Launch" button.

This brings up the "Application Launch" page. See figure 2.

BlastN	Name of job: temp	Database to search:	MBLNEW 🔽		
Launch select chunk size for viewing results 30 select view to display results	EMBL·HSADRA begin 1 1 cccgcct 61 gggctcc 121 cggcgcc 4	11 21 tcatcttccgccaggaggc 71 81 ctgcagccggacgcgggca 131 141 cgggccacccttactccc	31 caaggccgttggo 91 acgcgagctggas 151 tgcaggtgacgct	41 51 ccgagggcagettt 101 11 acgggaccgaggcg 161 17 cgacgetggtgtge	gcgcccat 1 ccgggggg 1 ctggccgg ▼
show results	Output (Options		Search Paramete	ers
automatically 🔽	Number of alignments to s Number of best hits from a 100 Number of one-line descri	thow 250 a region to keep iptions 500	Filter query Penalty for Reward for The E value word size Perfom gap	7 sequence a nucleotide mism r a nucleotide mat e 10.000000 Default oped alignment	natch -3 ch 1

FIGURE 2. Application Launch Page.

5. Click the "Launch" button.

Use the default settings for the application invocation. This will bring up the "Application Invocation" page, shown in figure 3.

STOP!!	
Don't go BACK	
BlastP is currently running. Please be patient The results will appear here shortly.	
running /data/appl/bin/blastall -p blast	

FIGURE 3. Application Invocation Page.

When the "Application Invocation" page finishes loading the "Entry List" page will be displayed. See figure 4.

Reset	<i>Query</i> "[BlastN-JobName:temp_job1]" found 6 entries
Perform operation on all but selected on selected Link Save Vieu *Names only* Launch OligoBlast show entries in chunks of 30 Printer Friendly	 BLASTN.temp job1 Report BLASTN.temp job1 emblnew HS439F8 BLASTN.temp job1 emblnew HS362J20 BLASTN.temp job1 emblnew HS1158B12 BLASTN.temp job1 emblnew CEY51H4 BLASTN.temp job1 Statistics

FIGURE 4. Application Results.

Click the hyperlink for one of the entries to examine the results of the application or apply one of the applications views.

Reset	vieu Complete entries * 💌
This entry is from: <u>BLASTN</u>	> <u>emblnew AL021392 H5439F8</u> Human DNA sequence from clone 439F8 on chromosome 22q13.31-33 Length = 120206
Save Link Launch	Score = 36.2 bits (18), Expect = 0.12 Identities = 18/18 (100%) Strand = Plus / Minus
OligoBlast 👤	Query: 582 agggcggcggcggcggcggc 599 Sbjct: 111124 agggcggcggcggcggccggcc 111107
	Score = 32.2 bits (16), Expect = 1.8 Identities = 16/16 (100%) Strand = Plus / Minus

FIGURE 5. Application Results.

Note: If you need help with the options in the "Application Launch" page click the option name hyperlink. You will see a description for the option as well as valid fields when appropriate.

SRS Users Manual

CHAPTER



COMMAND LINE SRS

7.1 getz

The SRS command line interface is made available to you through a unix shell window. You will need to contact your SRS administrator to get access.

The program used for command line SRS is called "getz". Once you have access to "getz" you can query a databank from the command line.

The "SRS Query Language" chapter will help you to write out your queries, here are a few examples:

getz "[swissprot-des:azurin]"

gets a list of entries from the swissprot databank that have "azurin" mentioned in the description.

getz -e "[swissprot-des:azurin]"

same as above but returns the complete entries.

getz -f seq -sf fasta "[swissprot-des:azurin]"

gets the sequence for the entries in the fasta format.

Important Options

The following table lists the options available for command line SRS.

You can use it to help improve your getz queries. You can learn more about getz with the command:

getz -help

TABLE 7.1 Important getz Options

Option	Default	What it does
-help		Help with getz.
-е	FALSE	prints entire entry.
-f <string></string>		include fields in entry list.
-vf <string></string>		List of fields that will be placed into a table view.
-W	FALSE	appends a wildcard to each search word.
-lb <n></n>	0	number of first entry in set to be viewed.
-ll <n></n>	0	number of entries to be viewed in one go.
-lv	FALSE	List all values that match the query.
-lvf	FALSE	List all values after the first match of.
-lmin <n></n>	0	List only values that occur at least specified number of times. Use together with "lv".
-c	FALSE	Report only number of entries found.
-info	FALSE	prints info about specified library.
-libs	FALSE	prints a list of all active libraries.
-view <string></string>		Name of view to be used when displaying entries.
-rs <string></string>		String of one or more characters to separate records in view.
-cs <string></string>		String of one or more characters to separate columns in view.
-sf <string></string>		format of sequence output file.
-af <string></string>		format of sequence alignment output file.
-html	FALSE	Select HTML format for output.
-off	FALSE	Accesses the off-line version of a databank.

SRS Users Manual





THE SRS QUERY LANGUAGE

In SRS any retrieval command, logical operations with sets that were obtained by previous queries, links between sets of different databanks, or a combination of all can be expressed by the SRS query language. This chapter describes the language and gives examples of its use.

8.1 General Syntax

A query is an expression with operands and operators. In this chapter queries are sometimes shown as equations in the form:

queryName = expression

The query name will be associated with the resultant set of entries. Expressions can specify index searches, logical operations between sets and link operations or a combination of all.

8.1.1 List of Operands

Name of a Databank

In SRS only one name for each databank is used (eg, "EMBL").

Set Name

Each query must be given a name (eg, "Q1") which can be later used to operate on the set that results from that query.

Index Search

A command to search in one or more indices of one or more libraries.

Expression

An expression is treated as an operand if it occurs within parentheses, for example, "(Q1&Q2)>EMBL". Parentheses can be nested to any degree.

Parent

Special operand that allows the conversion of a set of subentries, "Sequence" features, for example, into a set of entries, "Sequence" entries in this case. This is achieved by linking the set with the subentries to "parent" (eg, "Q1 > parent"). Used in the special case where the entry has subentries.

8.1.2 List of Operators

The following table shows a list of SRS query language operators.

Operator	Meaning
}	Logical OR
&	Logical AND
!	Logical AND NOT (BUT NOT) (operator may need to be escaped "/!" in unix.)
>	Link left

 TABLE 8.1 SRS Query Language Operators

Operator	Meaning
<	Link right
>^	Get subtree defined by left operand (hierarchical links)
>_	Get leaf entries of the subtree defined by left operand. (hierarchical links)

 TABLE 8.1 SRS Query Language Operators

The link operators have precedence over logical operators.

8.1.3 Searching in Indices

This command specifies a search in a single index or a group of indices of one or more databanks. An index search must specify within square brackets the databank or databank group name, the index or index group name, and a search expression. The two names must be separated by a hyphen ('-') and be separated from the search expression by a ':' (string search) or a '#' (range search). Both the field name (eg, "Description") and its abbreviation ("Des") can be used as index name. All strings, including the search words, are treated case insensitive.

[pir-des:elastase]

searches for the string "elastase" in the "Description" field of the protein databank PIR. "des" is the short name for the same field.

As indicated above two different types of searches, the string search and the range search exist. A string search expression starts with a colon (':') after the index name whereas the numeric range search starts with a hash ('#'). Range searches can be performed only in indices of the types "num" and "real".

Searching Strings

A search expression is a single search word or several words separated by a logical operator. Parentheses may be used for grouping.

EXAMPLE 8.1 - Search strings

```
[embl-key:insulin]
[embl-des:(acetylchol*&receptor)!muscarinic]
[swissprot-aut:sanger,f*!coulson,a*]
```

Wildcards are useful for searching a group of words or whenever it is unclear how a word is spelt in the databank. You can use two different types of wildcards:

TABLE 8.2 Wild Cards in SRS

*	Matches one or more characters of any value
?	Matches a single character of any value

Any number of wildcards can be placed anywhere in the search word. If you place one at the beginning you might have a slightly longer response time since then **all** words of the index have to be searched. But don't worry too much about that - it is still quite fast!

Regular Expressions

Since search words with wildcards are translated into regular expressions, it is of course possible to enter regular expressions directly. They must appear within forward slashes ('/'). Some characters (" $^{...}$ have a special meaning. They must be prefixed with a backslash ('\') if to be matched literally.

TABLE 8.3 Regular Expression Syntax

^	marks the begin of a string, eg, "/^phos/" will find all
	words beginning with "phos", eg, "phosphate"

\$	marks the end of a string, eg, "/ase\$" will find all words ending with "ase", eg, "kinase"
•	means any character
[]	indicates a set of characters to match, eg "[()]" matches an opening or closing parenthesis. Character ranges can be specified by using a hyphen ('-'), eg "[0-9]" matches a digit. A caret ('^') in front of the character set (after the opening square bracket) negates the character set, eg, "[^0-9]" matches any non-digit character.
()	Groups a series of pattern elements to a single element
*	The preceding group may be repeated zero or more times
+	The preceding group may be repeated one or more time
?	The preceding group may or may not occur

 TABLE 8.3 Regular Expression Syntax

TABLE 8.4 Example Usage of Regular Expressions

/^j\$/	finds all 3 character strings with a j at the beginning
/^5[0-9][0-9][0-9]\$/	finds all 4 digit numbers with a five at the beginning
/^nif[a-e]\$/	finds the gene names "nifa", "nifb", "nifc", "nifd", "nife"
/^mue?ller\$/	finds both "muller" and "mueller"

Note that searches with regular expressions can be slow since **all** words in the index have to be searched.

Searching Numeric Ranges

In a numeric index (integers, reals) it is possible to search numeric ranges. The number index is only applicable where there is a one to

one relationship of entry and value (eg, sequence length, creation date, resolution).

A range can be specified by a single value or two values separated by a colon (':') where the left value must be smaller than the right value. To exclude a boundary value from the range put a '!' in front. A missing number on the left means the minimum, and on the right the maximum value in the index.

The following are queries in an index of the sequence length.

 TABLE 8.5
 Numeric Range Examples

400	selects all sequences with the length of 400
400:500	selects all sequences from 400 to 500 residues
400:	selects all sequences longer than 400 residues
:500	selects all sequences up to 500 residues
400:!500	selects all sequences between 400 and 500 residues excluding 500
:	is also valid and retrievesall sequences

Note that ranges can be combined by logical operators.

```
300:!500 | !600:700
or
300:700 ! 500:600
```

retrieve the same set of sequences, namely all sequences from 300 to 500, excluding 500, plus all sequences from 600 to 700 where the 600 is excluded.

Searching Dates

It is possible to search dates using one of the two special formats that are recognized by the SRS query language. They are: "YYYYMMDD" or "DD-MMM-YYYY". Examples are:

```
"19990515"
or
"13-DEC-1999"
```

Dates can be used within ranges as normal numbers:

```
[swissprot-date#19910101:19951231]
[swissprot-date#13-DEC-1998:13-DEC-1999]
```

Searching Multiple Databanks

Instead of the databank name in a search expression, a list of names may be used. The list must be enclosed within curly braces and the names must be separated by spaces. The query:

```
[{swissprot swissnew sptrembl}-des:kinase]
```

searches the word "kinase" in the Description index of SWISS-PROT, SWISSNEW and SPTREMBL. If many index queries over multiple databanks must be combined in a single query, it is convenient to define a group of databanks to be used later in the query instead of a databank name. The example creates the group "dbs" with the 3 databanks SWISS-PROT, SWISSNEW and SPTREMBL within the first search command and uses the group name "dbs" in the second search name.

```
[dbs={swissprot swissnew sptrembl}-des:kinase] &[dbs-org:human]
```

8.1.4 Using Logical Operators

The three logical operators OR ('|'), AND ('&') and BUTNOT ('!') can be used to combine search words in an index search, or sets in a query expressions.

The following figure illustrates the effect of the three operators in an expression of the form "A operator B".



FIGURE 1. Logical Operators in SRS.

Logical operations can be only performed between two sets of the same type. It is not possible to combine a set of entries and a set of subentries (see below). In those cases an additional link operation must be specified.

Using Link Operators

The powerful **link** operators are unique in the SRS query language. They resemble the **join** in relational databank systems to some extent. The two link operators '<' and '>' combine two sets from different databanks.

The figure shows two databanks "A" and "B" where some entries in "A" have cross references to entries in "B" as indicated by dark lines. These cross references are processed to build link indices which provide the basis for the link operation.



FIGURE 2. Cross-Referenced Data.

A > B

gives entries in "B" that are referenced by entries in "A".

A < B

gives those entries in "A" that reference entries in "B".

References are by nature not bidirectional, that is, there is no guarantee that entries in "B" that are referenced by entries in "A" have references back to "A". The link indices in SRS, however, can be always used bidirectionally. The two link expression can now be seen as:

$\mathbf{A} > \mathbf{B}$

gives those entries in "B" that **are linked** to entries in "A".

A < B

gives those entries in "A" that are linked to entries in "B".

In this context it is irrelevant which databank contributed the information for building the link index.

Link queries to refine search results

[swissprot-des:kinase] > PDB

The index search retrieves all kinase sequences from the SWISS-PROT protein sequence databank which are then linked to the PDB databank of solved tertiary protein structures. The result is all the PDB entries with atomic coordinates for all kinases for which the tertiary structure has been determined.

In SRS all the pairwise links are combined to generate a network of databanks. Within this network a link can be performed from any databank to any other. If two databanks are not directly connected by a link then a series of links is performed. The following rules are applied to resolve links:

- The same pairwise link index is used for both directions.
- SRS tries to find the shortest possible way for linking entries from two libraries. Ideally this is the direct link as, eg, between EMBL and SWISS-PROT; if no direct link is available (eg, EMBL > PDB) then automatically the optimal or cheapest succession of links is performed (eg, EMBL > SWISSPROT > PDB).
- If a set contains entries from different databanks (eg, from EMBL and SWISS-PROT) then the subsets of all libraries found in the set are linked independently.

Advanced queries using links:

enzyme < pdb

The Enzyme databank contains all known reactions catalyzed by enzymes. The number of entries retrieved by the query is the number of different reactions catalyzed by proteins with known structure.

[swissprot-id:acha_human] > prosite > swissprot

The index search retrieves the SWISS-PROT entry "ACHA_HUMAN". This entry is then linked to the Prosite entry(s) that documents the protein family where "ACHA_HUMAN" is a member, in this case the family of neuronal acetylcholine receptors. The next link retrieves all SWISS-PROT entries that belong to that family. In effect, the entry "ACHA_HUMAN" is amplified to all members of the protein family or families it belongs to. The example shows that it is possible to navigate over the databank network with an explicit succession of links that is evaluated from left to right.

[swissprot-id:gshr_caeel] > prodom > pdb

This example shows how the amplification by homologous entries in the same databank can be used to find related information in another databank to which the single entry itself is not linked. The query retrieves a DNA polymerase from SWISS-PROT, expands it by related proteins as documented in Prodom (databank of protein domains, another possibility would be Prosite or HSSP) and links these entries to PDB. The result is the set of PDB tertiary protein structures that are homologous to the SWISS-PROT entry "GSHR_CAEEL".

(q = [{swissprot swissnew}-des:kinase])!(q<swissnew)</pre>

Finds all entries for "kinase" in all the databanks and excludes any that are replaced by an entry in SWISSNEW. The trick here is that SWISSNEW is not linked to SWISSNEW. It will only return the entries that are truly the most recent.

Most protein or DNA databanks overlap to a great extent which creates a lot of redundancy. The annotation of equivalent entries in different databanks can be quite different which can be very useful for string searching since the probability of finding a certain enzyme name is much greater if you search all sequence databanks. After the search the links come in handy to remove the overlap. The first query in the example searches "kinase" in all sequence databanks (eg, EMBL, PIR, SWISS-PROT, GenBank). The second query removes the overlap. The goal is to have SWISS-PROT entries plus those in other databanks that do not have equivalents in SWISS-PROT. This is achieved by subtracting all entries from "q" that are linked to SWISS-PROT and won't be removed!. Just change "SWISS-PROT" to "PIR" if you would rather have PIR entries.

8.1.5 Entries and Subentries

Sets originating from the same databank may have different set types. Consider the two queries:

```
[swissprot-keywords:transmembrane]
[swissprot-ftkey:transmem]
```

The first query retrieves all SWISS-PROT entries that have transmembrane segments, the second finds all transmembrane **features** contained in SWISS-PROT entries.

The second query will retrieve many more entries since most transmembrane proteins have more than one membrane spanning segment. If you requested the sequences for entries in the second set you would get the transmembrane segments and not the parent entry's sequence.

The first query returns a set of **entries** whereas the second returns a set of **subentries**. The "ftkey" index has a special type: searches in that index will for all sequence databanks result in sets of subentries. Sets of entries and subentries can not be combined with logical operators. Only the link operators may be used between them, ie, it is always possible to link subentries to their respective parent entries.

Performing links with sub-entries

[swissprot-org:human] > [swissprot-ftkey:transmem]

returns all transmembrane segments found in human proteins.

[swissprot-org:human] < [swissprot-ftkey:transmem]

returns all human proteins that have transmembrane segments.

Sometimes it is necessary to do an explicit conversion from subentries to entries. This can be done with a link to "parent".

Advanced links with sub-entries

```
[swissprot-ftkey:transmem] > parent |
[swissprot-key:transmembrane]
```

returns all entries that have the "transmembrane" keyword or "transmem" sequence features. This may be necessary to ensure all entries with a certain property are retrieved - the annotation is often not complete!

[swissprot-ftkey:transmem] > parent > pdb

finds all PDB entries of proteins that have transmembrane segments (even though the known structure may not yet include the transmembrane region).

8.1.6 Storing Intermediate Results in Sets

If a query becomes very complicated it may be convenient to store an intermediate result within a query into a set with a name that can be used in later parts of the query. An example is an index query in EMBL that then is linked to both SWISS-PROT and SWISSNEW.

```
[embl-org:escherichia coli] > SWISSPROT |
[embl-org:escherichia coli]>SWISSNEW
```

The index query in EMBL has to be specified twice. It is possible to store the result of the index query in a set (called "temp" in the example but can be called anything). The assignment operation must be within parentheses.

(temp=[embl-org:escherichia coli])>PROSITE | temp > SWISSNEW