SRS USER GUIDE

Version 7.0

LION Bioscience SRS 7.0 Documentation

Comments about the documentation are welcome at: documentation@uk.lionbioscience.com

For customer support issues please e-mail support@uk.lionbioscience.com

TABLE OF CONTENTS

CHAPTER 1: SRS QUICK TOUR 1

1.1 Introduction 2

1.2 Starting an SRS Project 2

1.3 SRS Queries 6

Making a Query 6

1.4 Sorting Results 9

1.5 Links 9

Linking to Related Information 10

1.6 Using Views 12

Creating a View 12

Applying a View 16

Deleting a View 19

1.7 Using Applications 20

Running a Batch Queued Application 23

Running an Interactive Application 25

Viewing Application Results 28

1.8 Saving your Results 30

1.9 Review 32

CHAPTER 2: SRS PROJECTS 33

2.1 Introducing Projects 34

Temporary Projects 34

Permanent Projects 34

Starting a Project 35

2.2 Temporary Projects 37

Starting a Temporary Project 37
Managing Temporary Projects 38
2.3 Permanent Projects 40
Starting a Permanent Project 40
Managing Permanent Projects 42
2.4 Using the Project Manager 43
Creating a New Project 44
Switching to Another Project 44
Copying Project Information 45
Saving a Project 46
Opening a Saved Project 48
Deleting a Project 49
Renaming a Project 49

CHAPTER 3: QUERYING WITH SRS 51
3.1 Search Terms 52
Single Word Searches 52
Multiple-Word Searches 53
Numbers and Dates 53
Regular Expressions 55
Wildcards. 55
3.2 Quick Search 55
About Quick Search 55
Using Quick Search 56
3.3 Query Forms 57
About Query Forms 57

Using the Standard Query Form 57
Using the Extended Query Form. 60
3.4 Subentries 62
About Subentries. 62
Use of Subentries 62
Querying Using Subentries 63
3.5 Expression Queries 73
About Expression Queries. 73
Using Expression Queries 74
3.6 Sorting Results. 76
What is Sorting? 76
Sorting a Set of Results 76
3.7 Browse Index. 80
About Browsing Indices 80
Browsing Indices. 80
Getting to the Field Information Page. 81

CHAPTER 4: LINKS TO ADDITIONAL DATA. 85
4.1 What is a Link? 86
4.2 Hypertext Links 86
4.3 Index Links 89
Searching for Data using Index Links. 89
Find All Entries Options 92
4.4 Getting to the LINK Page 93
Finding Links from the Manage your Query Results Page. 93
Finding Links from the Query Result Page. 94

Finding Links from the Entry Page 97

4.5 Expression Linking 98

Expression Linking Procedure 98

Linking Operators 99

Expression Linking Examples. 100

CHAPTER 5: VIEWS. 103

5.1 What is a View?. 104

5.2 Using Views. 104

How to Apply Views. 105

5.3 Creating Views. 118

Creating Views from the Query Forms 118

Creating Views using the View Manager Pages 124

CHAPTER 6: APPLICATIONS 131

6.1 Introducing Applications 132

6.2 Accessing Applications 132

6.3 Launch Page. 133

6.4 Interactive and Batch Applications 134

Introduction 134

Interactive Applications 134

Batch Applications 136

CHAPTER 7: COMMAND LINE SRS 141

7.1 getz. 142

getz Usage 143

CHAPTER 8: SRS QUERY LANGUAGE 145

8.1 Introduction 146
8.2 Searching in Indices. 146
Introduction 146
General Syntax 147
Search Strings 147
Searching Using Numerical Ranges 151
Searching for Dates 152
Searching Multiple Databanks 153
8.3 Combining Searches 153
Introduction 153
General Syntax 154
Operands 155
Operators 156
Use of Operators to Combine Search Items 159
8.4 Entries and Subentries 161
Links with Sets Containing Subentries 162
8.5 Storing Intermediate Results in Sets 164

INDEX **145**

CHAPTER

1

SRS QUICK TOUR

SRS is the world’s premier data integration, analysis and display tool for bioinformatic, genomic and related data.

This chapter contains a guided tour of SRS which will help you become familiar with the SRS web interface. During the guided tour, you will learn:

- How to start an SRS project.
- How to perform a query using SRS.
- How to link your query results to other databanks.
- How to change the way in which your results are displayed by creating your own view.
- How to run an application.
- How to save your working project.

1.1 Introduction

This chapter is intended to introduce new users to SRS. It will take you through some basic procedures that you can expect to do with SRS. It is intended as an introduction and will get you started, but it is not intended as a complete description of the software.

If you want more detail of how to do a certain task refer to the relevant chapter of this manual. The notes in the margins tell you which chapter of this guide is relevant to each set of steps.

If you want to know more about a particular SRS page, then you should refer to the relevant page in the SRS Online manual. You can get to the Online manual by clicking the HELP button on your SRS page at any time. This will take you to the information relevant for the page you were using when you pressed the HELP button.

1.2 Starting an SRS Project

See "SRS Projects" - Chapter 2 for more information.

1. When you access SRS the **Start** page is usually shown first. From here you can start a new project, or access the online help files.

Note: Your SRS Administrator should be able to give you the web address of your local server, or of an external site to which you have access.

Want to know more about the type of data you can find using SRS?
- Go to the **Databanks Information** pages.

Want to learn more about using SRS?
- Look in the **SRS Documentation**

SRS Version 7.0 Copyright © 1997-2002 LION Bioscience. All Rights Reserved. Terms of Use

Figure 1.1 The **Start** page.

2. Moving the mouse cursor over one of the options will give more details of what that option does. For this tour, choose a temporary project. Move your mouse cursor over the **Start a Temporary Project** option (without clicking) to find out more about this option.

This starts a temporary project in SRS.

The results of any work you do will be stored in this project and are available to you until you close SRS. You can choose to save your results to your desktop, or you can use a permanent project.

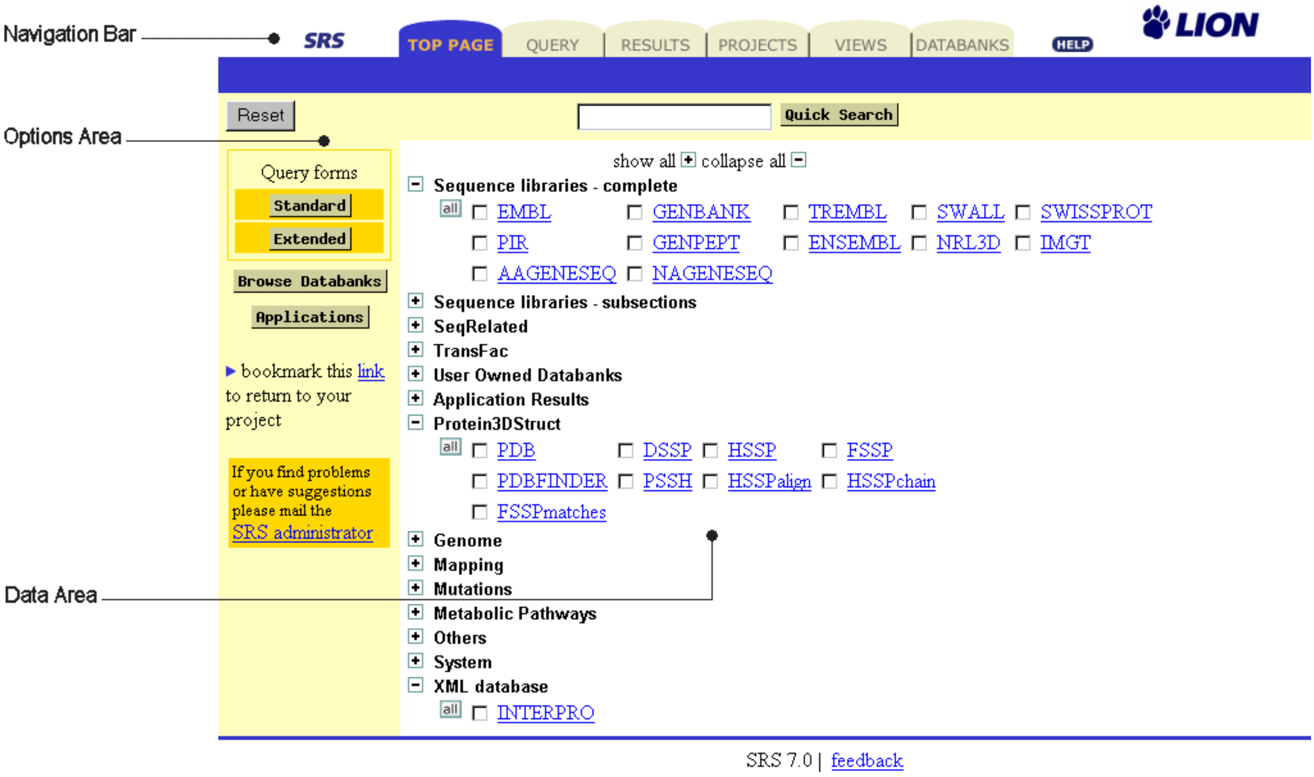
Want to know more about the type of data you can find using SRS?
- Go to the **Databanks Information** pages.

Want to learn more about using SRS?
- Look in the **SRS Documentation**

SRS Version 7.0 Copyright © 1997-2002 LION Bioscience. All Rights Reserved. Terms of Use

Figure 1.2 The **Start** page, showing details for the **Start a Temporary Project** option.

3. Now click on the **Start a Temporary Project** option. This takes you to the **Select Databanks To Search** page which is where you begin to work with SRS.



▸ bookmark this link to return to your project

If you find problems or have suggestions please mail the SRS administrator

Figure 1.3 The **Select Databanks to Search** page (**Top** page).

Note: This page can be reached, at any time from within SRS, by clicking on the **Top Page** tab.

The next section will take you through making a query.

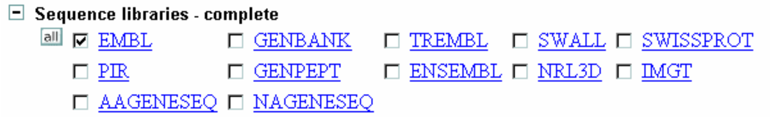
1.3 SRS Queries

1.3.1 Making a Query

See “Querying with SRS” - Chapter 3 for more information.

This example will take you through making a query, which uses the SRS **Standard Query Form** to search for the word **kinase** in the Description field of the EMBL databank.

1. Before you start your query you must choose the databanks in which you wish to search. For this example, select the EMBL databank by ticking the check box to the left of the **EMBL** hyperlink. (EMBL is the European International Sequence Database.)



Sequence libraries - **complete**

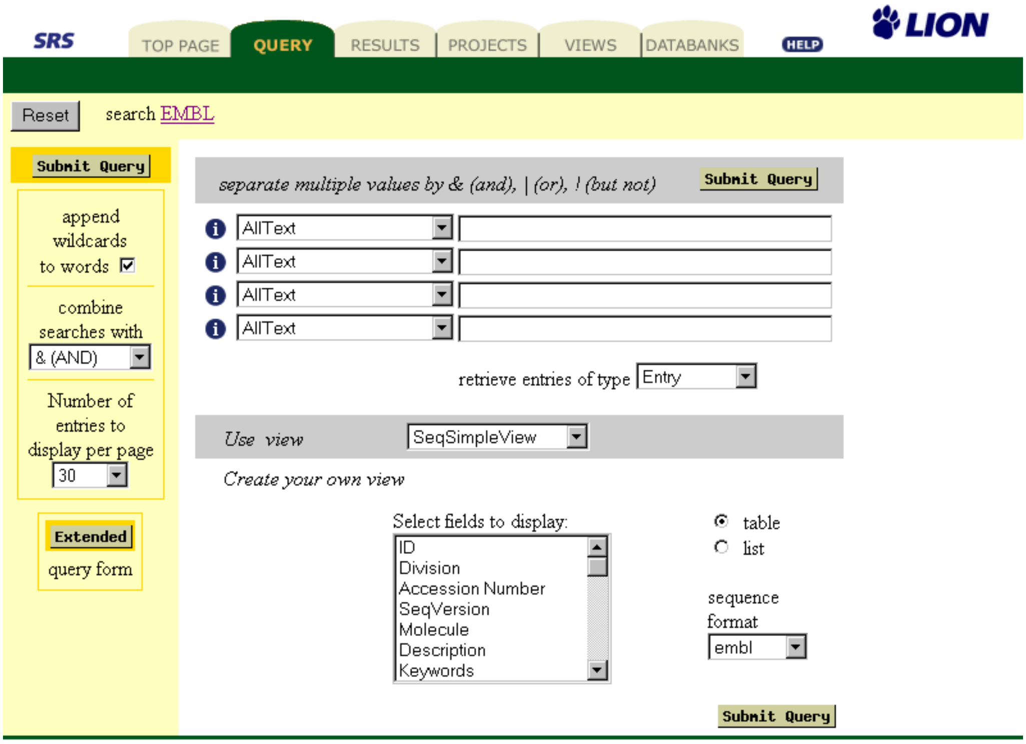
all ☑ EMBL ☐ GENBANK ☐ TREMBL ☐ SWALL ☐ SWISSPROT

☐ PIR ☐ GENPEPT ☐ ENSEMBL ☐ NRL3D ☐ IMGT

☐ AAGENESEQ ☐ NAGENESEQ

Figure 1.4 Selecting the EMBL databank on the **Select Databanks to Search** page.

2. Click the Standard button to display the **Standard Query Form**. From here you can search databanks in many different ways.



search EMBL

separate multiple values by & (and), | (or), ! (but not)

AllText
AllText
AllText
AllText

append wildcards to words ☑

combine searches with & (AND)

Number of entries to display per page 30

retrieve entries of type Entry

Use view SeqSimpleView

Create your own view

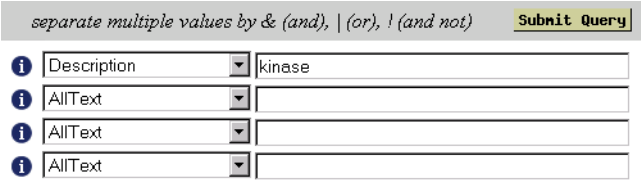
Select fields to display:

● table
○ list

sequence format embl

Figure 1.5 Standard Query Form for the EMBL databank.

3. In the first field box, select the **Description** field using the drop-down list.
4. Type **kinase** in the text box beside it.



separate multiple values by & (and), | (or), ! (and not)

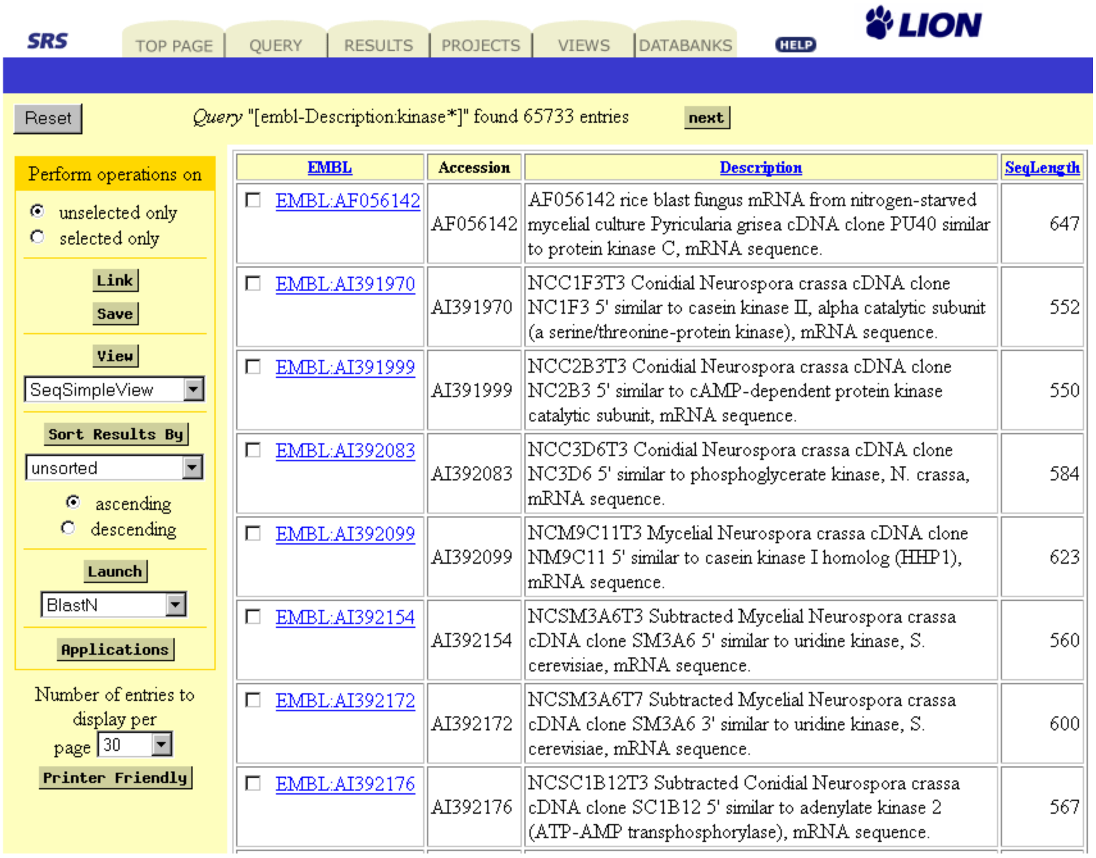
Description kinase
AllText
AllText
AllText

Figure 1.6 Search the **Description** field for **kinase**.

Note: If you want to know more about the chosen field, click on the ℹ icon beside it. This will display the Field Information page (see also section 7.3 “Field Information Page” in the *Online Documentation*). Use the **Back** button on your browser to return to the **Query Form**.

---

5. Click one of the `Submit Query` buttons. This runs the query and then displays the **Query Result** page, showing your results.



Query "[embl-Description:kinase*]" found 65733 entries

Perform operations on

● unselected only
○ selected only

View: SeqSimpleView

Sort Results By: unsorted

● ascending
○ descending

Launch: BlastN

Number of entries to display per page 30

	EMBL	Accession	Description	SeqLength
☐	EMBL:AF056142	AF056142	AF056142 rice blast fungus mRNA from nitrogen-starved mycelial culture Pyricularia grisea cDNA clone PU40 similar to protein kinase C, mRNA sequence.	647
☐	EMBL:AI391970	AI391970	NCC1F3T3 Conidial Neurospora crassa cDNA clone NC1F3 5' similar to casein kinase II, alpha catalytic subunit (a serine/threonine-protein kinase), mRNA sequence.	552
☐	EMBL:AI391999	AI391999	NCC2B3T3 Conidial Neurospora crassa cDNA clone NC2B3 5' similar to cAMP-dependent protein kinase catalytic subunit, mRNA sequence.	550
☐	EMBL:AI392083	AI392083	NCC3D6T3 Conidial Neurospora crassa cDNA clone NC3D6 5' similar to phosphoglycerate kinase, N. crassa, mRNA sequence.	584
☐	EMBL:AI392099	AI392099	NCM9C11T3 Mycelial Neurospora crassa cDNA clone NM9C11 5' similar to casein kinase I homolog (HHP1), mRNA sequence.	623
☐	EMBL:AI392154	AI392154	NCSM3A6T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 5' similar to uridine kinase, S. cerevisiae, mRNA sequence.	560
☐	EMBL:AI392172	AI392172	NCSM3A6T7 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 3' similar to uridine kinase, S. cerevisiae, mRNA sequence.	600
☐	EMBL:AI392176	AI392176	NCSC1B12T3 Subtracted Conidial Neurospora crassa cDNA clone SC1B12 5' similar to adenylate kinase 2 (ATP-AMP transphosphorylase), mRNA sequence.	567

Figure 1.7 The **Query Result** page for your query.

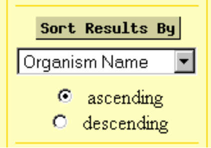
Note: If you want to look at a complete entry, click on its hyperlink.

1.4 Sorting Results

See section 3.6 “Sorting Results” (p. 76) for more information.

Results are listed, by default, in the order in which they are stored in the databank (**unsorted**). To make it easier for you to find the results you want, you can use sorting to reorganize them. Sorting is particularly useful if you have a large number of results.

1. To sort a set of results, choose the required sort order from the drop-down list under the **Sort Results By** button, e.g. **Organism Name**, **Sequence Length**. Then use the option buttons to choose whether to sort in **ascending** or **descending** order.



Organism Name

● ascending

○ descending

Figure 1.8 Sort by **Organism Name** in an **ascending** order.

2. When you have set up your sorting requirements, press the **Sort Results By** button to sort the results.

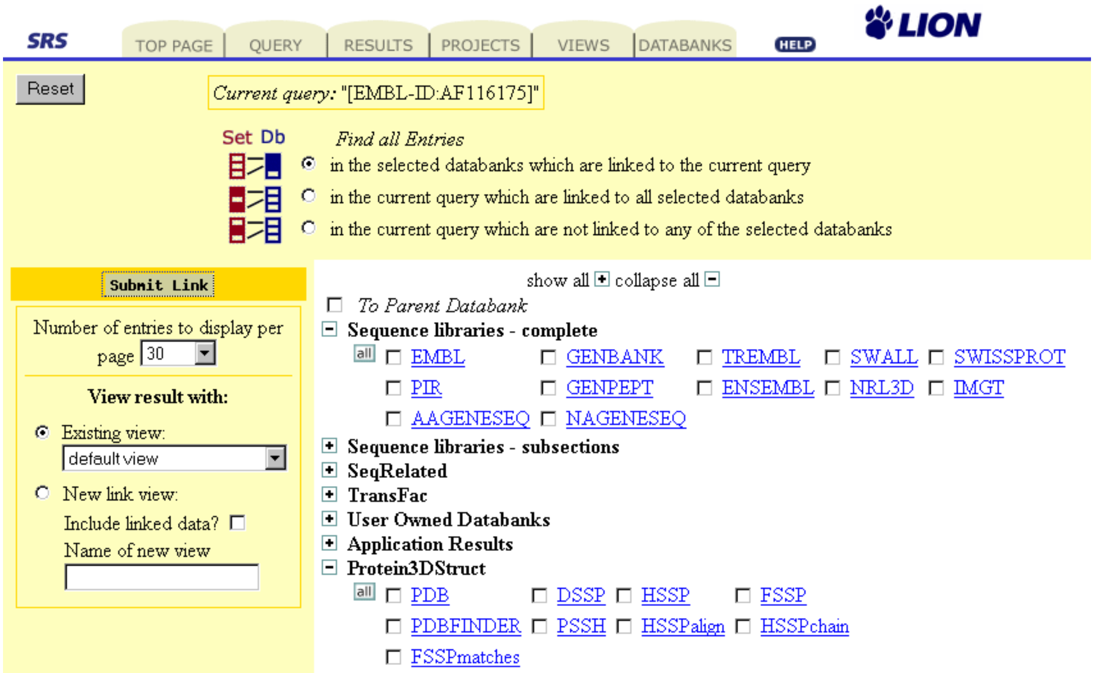
1.5 Links

See “Links to Additional Data” - Chapter 4 for more information.

One of the many features of SRS is its ability to search for links between your current results and related information in other databanks. In this example you will search for links from one of your EMBL results to related entries in SWISS-PROT.

1.5.1 Linking to Related Information

1. On the **Query Result** page, tick the check box next to the entry for which you want to find related items.
2. Click the Link button to display the **LINK** page.



Current query: "[EMBL-ID:AF116175]"

Find all Entries

● in the selected databanks which are linked to the current query

○ in the current query which are linked to all selected databanks

○ in the current query which are not linked to any of the selected databanks

Number of entries to display per page 30

View result with:

● Existing view: default view

○ New link view:

Include linked data? ☐

Name of new view

☐ *To Parent Databank*

⊟ **Sequence libraries - complete**

all ☐ EMBL ☐ GENBANK ☐ TREMBL ☐ SWALL ☐ SWISSPROT
☐ PIR ☐ GENPEPT ☐ ENSEMBL ☐ NRL3D ☐ IMGT
☐ AAGENESEQ ☐ NAGENESEQ

⊞ **Sequence libraries - subsections**

⊞ **SeqRelated**

⊞ **TransFac**

⊞ **User Owned Databanks**

⊞ **Application Results**

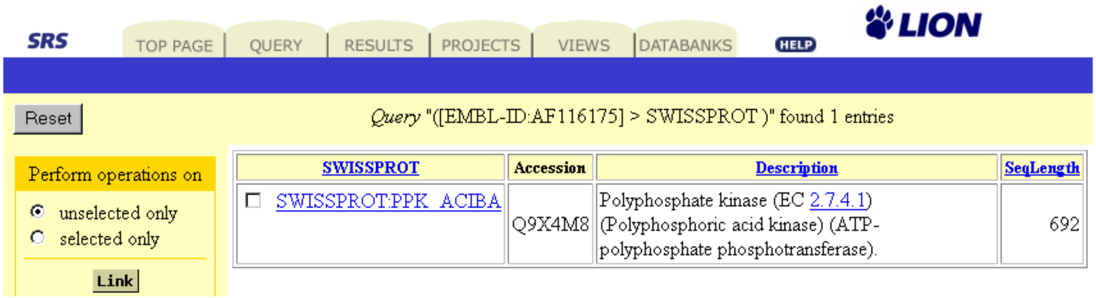
⊟ **Protein3DStruct**

all ☐ PDB ☐ DSSP ☐ HSSP ☐ FSSP
☐ PDBFINDER ☐ PSSH ☐ HSSPalign ☐ HSSPchain
☐ FSSPmatches

Figure 1.9 LINK page.

3. Tick the check box to the left of the databank in which you wish to find links, e.g. **SWISSPROT**.
4. Click the Submit Link button to search for the related results.

The result will be a list of all the SWISS-PROT entries that are related to the EMBL entry (or entries) with which you started. These will be displayed on the **Query Result** page.



Query "([EMBL-ID:AF116175] > SWISSPROT)" found 1 entries

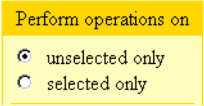
Perform operations on
● unselected only
○ selected only

SWISSPROT	Accession	Description	SeqLength
☐ SWISSPROT:PPK_ACIBA	Q9X4M8	Polyphosphate kinase (EC 2.7.4.1) (Polyphosphoric acid kinase) (ATP-polyphosphate phosphotransferase).	692

Figure 1.10 Query Result page showing a SWISS-PROT entry that has a link to the EMBL entry: AF116175, from the above query for **kinase**.

If there are no items linked to your selection, then go back to the original **Query Result** page, choose a different selection, and try again.

If you want to search all your results for links to SWISSPROT, from a **Query Result** page, click the **unselected only** option button, ensure that the check box beside each of the entries is unticked, and repeat the search.



Perform operations on
● unselected only
○ selected only

Figure 1.11 Perform operations on option buttons.

Similarly, by ticking several entries in your list, you can search all those you have selected, or all those that are not selected using the option buttons.

Note: Searching a large number of results for links might take some time when large databanks are involved.

1.6 Using Views

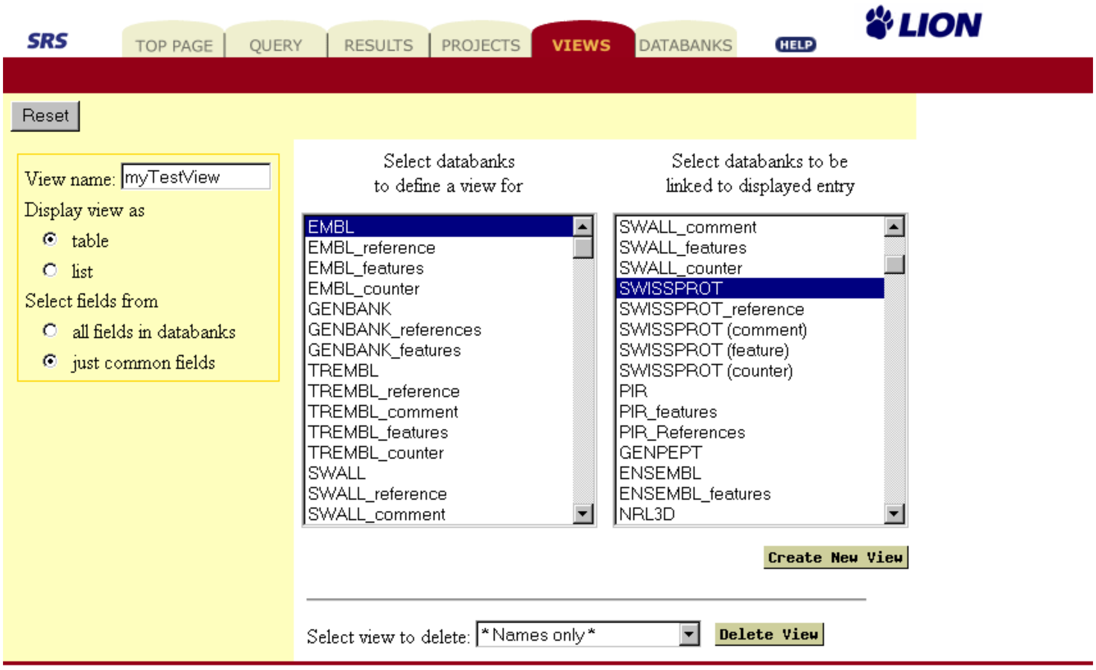
See "Views" - Chapter 5 for more information.

SRS allows you to customize the way in which you display data. This is usually done using the **View Manager** pages, although simple view creation is also possible from both of the **Query Forms** (see section 5.3.1 "Creating Views from the Query Forms", p. 118).

1.6.1 Creating a View

This example explains how to create a table view that can be used for EMBL, and for which links to SWISS-PROT should be shown.

1. Click the **Views** tab to display the **View Manager** page 1.
   On this page you can select databanks and choose a name for your customized view. See Figure 1.12.



View name: myTestView

Display view as

● table

○ list

Select fields from

○ all fields in databanks

● just common fields

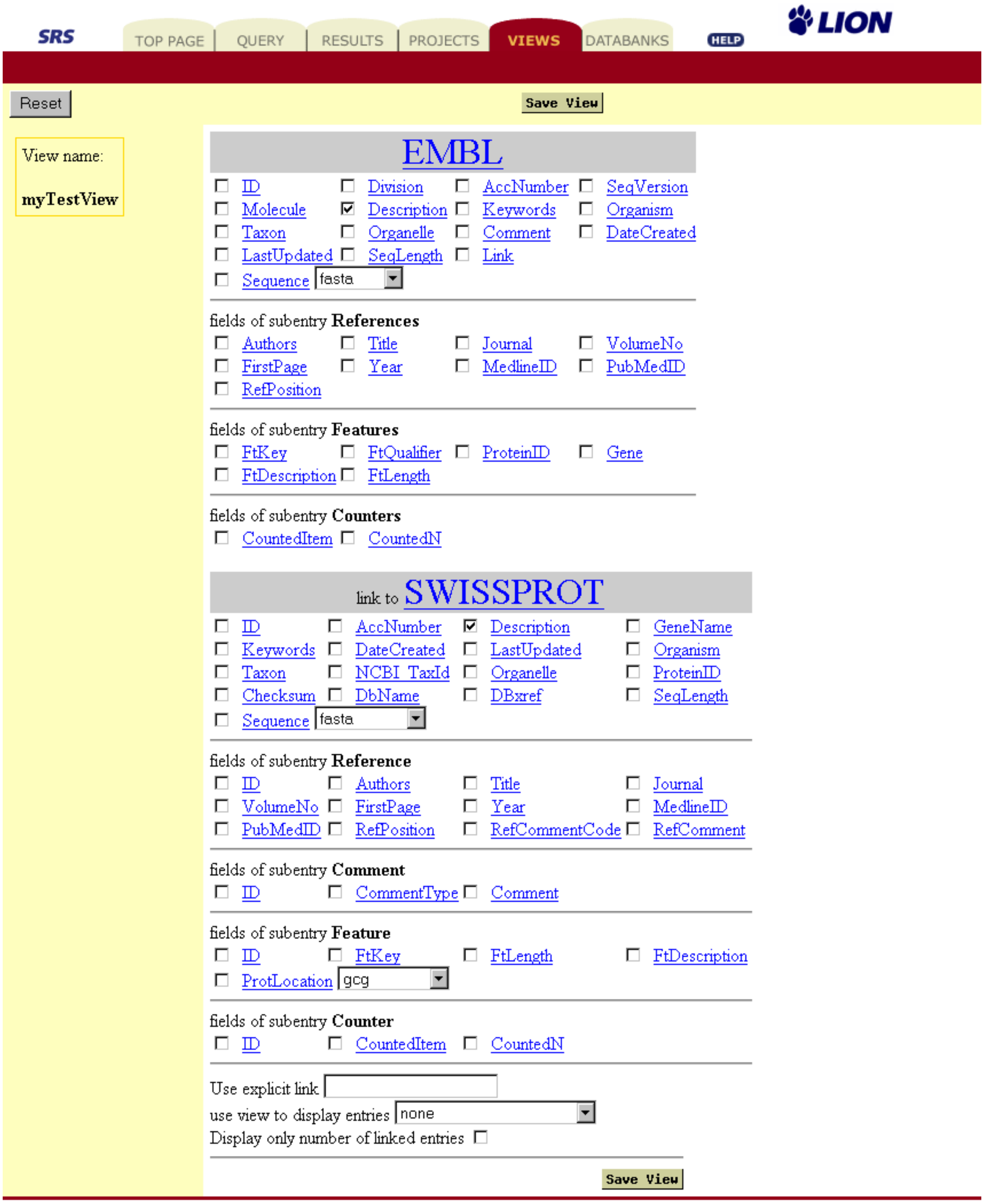
Select view to delete: *Names only*

Figure 1.12 Create your own views (Steps 2 - 7).

2. Give your view a name by typing it in the **View name** box. This example uses the name “myTestView”.
3. Use the **Display view as** option buttons to indicate whether you want to create a table or list view.
4. Using the **Select fields from** option buttons, indicate whether you want to be able to select fields from all of those available for your chosen databanks, or only common fields.
5. In the list under **Select databanks to define a view for**, click **EMBL** to select it.
6. In the list under **Select databanks to be linked to displayed entry** click **SWISSPROT**.

7. Click the Create New View button to display the **View Manager** page 2 from which you can select the fields for your views (see Figure 1.13).
8. Choose the fields you want to view by ticking the check boxes beside them. For the purpose of this example, select the **Description** field for both databanks.

Note: The field names are also hyperlinks to the **Field Information** page (see also section 7.3 “Field Information page for the Description field.” in the *Online Documentation*), which contains information about each field.



View name:

myTestView

EMBL

☐ ID ☐ Division ☐ AccNumber ☐ SeqVersion
☐ Molecule ☑ Description ☐ Keywords ☐ Organism
☐ Taxon ☐ Organelle ☐ Comment ☐ DateCreated
☐ LastUpdated ☐ SeqLength ☐ Link
☐ Sequence fasta

fields of subentry **References**
☐ Authors ☐ Title ☐ Journal ☐ VolumeNo
☐ FirstPage ☐ Year ☐ MedlineID ☐ PubMedID
☐ RefPosition

fields of subentry **Features**
☐ FtKey ☐ FtQualifier ☐ ProteinID ☐ Gene
☐ FtDescription ☐ FtLength

fields of subentry **Counters**
☐ CountedItem ☐ CountedN

link to SWISSPROT

☐ ID ☐ AccNumber ☑ Description ☐ GeneName
☐ Keywords ☐ DateCreated ☐ LastUpdated ☐ Organism
☐ Taxon ☐ NCBI_TaxId ☐ Organelle ☐ ProteinID
☐ Checksum ☐ DbName ☐ DBxref ☐ SeqLength
☐ Sequence fasta

fields of subentry **Reference**
☐ ID ☐ Authors ☐ Title ☐ Journal
☐ VolumeNo ☐ FirstPage ☐ Year ☐ MedlineID
☐ PubMedID ☐ RefPosition ☐ RefCommentCode ☐ RefComment

fields of subentry **Comment**
☐ ID ☐ CommentType ☐ Comment

fields of subentry **Feature**
☐ ID ☐ FtKey ☐ FtLength ☐ FtDescription
☐ ProtLocation gcg

fields of subentry **Counter**
☐ ID ☐ CountedItem ☐ CountedN

Use explicit link
use view to display entries none
Display only number of linked entries ☐

Figure 1.13 View Manager page 2.

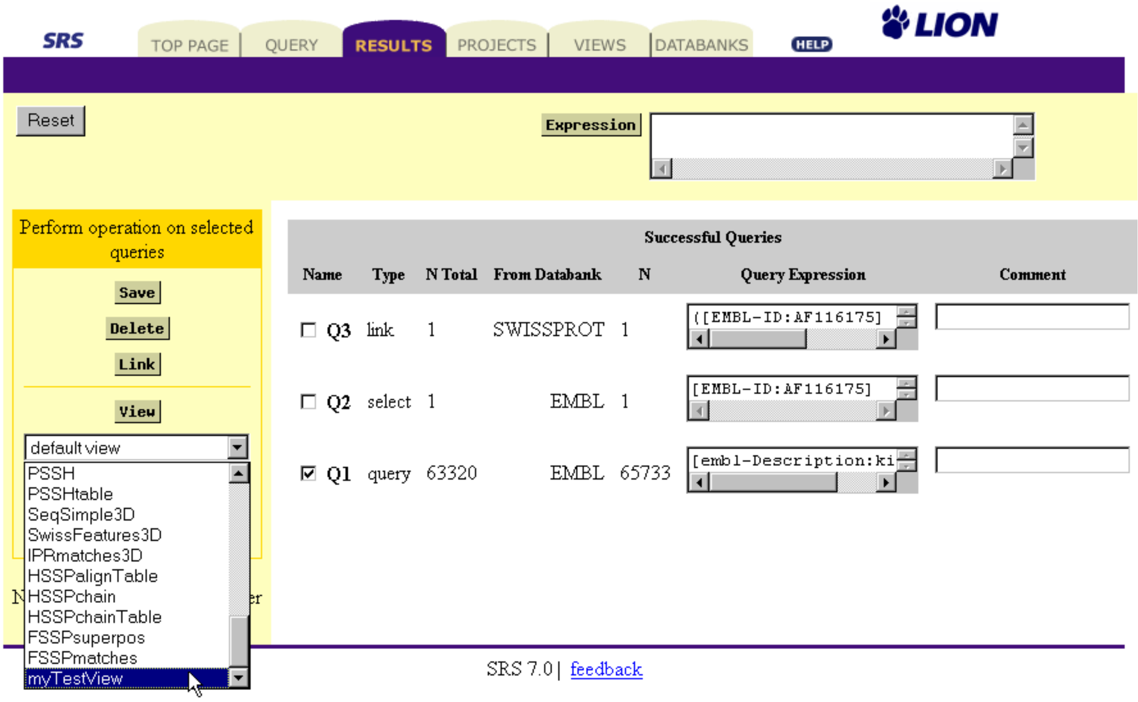
9. Click the Save View button to save your view.
   Your view will be saved and you will be returned to the **View Manager** page 1, where you may create more views.

Your saved views will be added to the drop-down list below the View button. It can then be applied in the same way that any other view is applied (see section 1.6.2 “Applying a View”, p. 16).

1.6.2 Applying a View

This section demonstrates how to apply a view to display your results. The example uses the view created in section 1.6.1 “Creating a View”, p. 12, but can equally refer to any other view in the drop-down list.

1. Choose a set of results that you wish to view, e.g. by ticking the check box beside them on the **Manage your Query Results** page. (Click the **Results** tab to get to the **Manage your Query Results** page.) See Figure 1.14.
2. Tick the check box beside the query labelled **Q1**. This was the first query you made (hence **Q1**).
3. Select **myTestView** from the drop-down list below the View button.

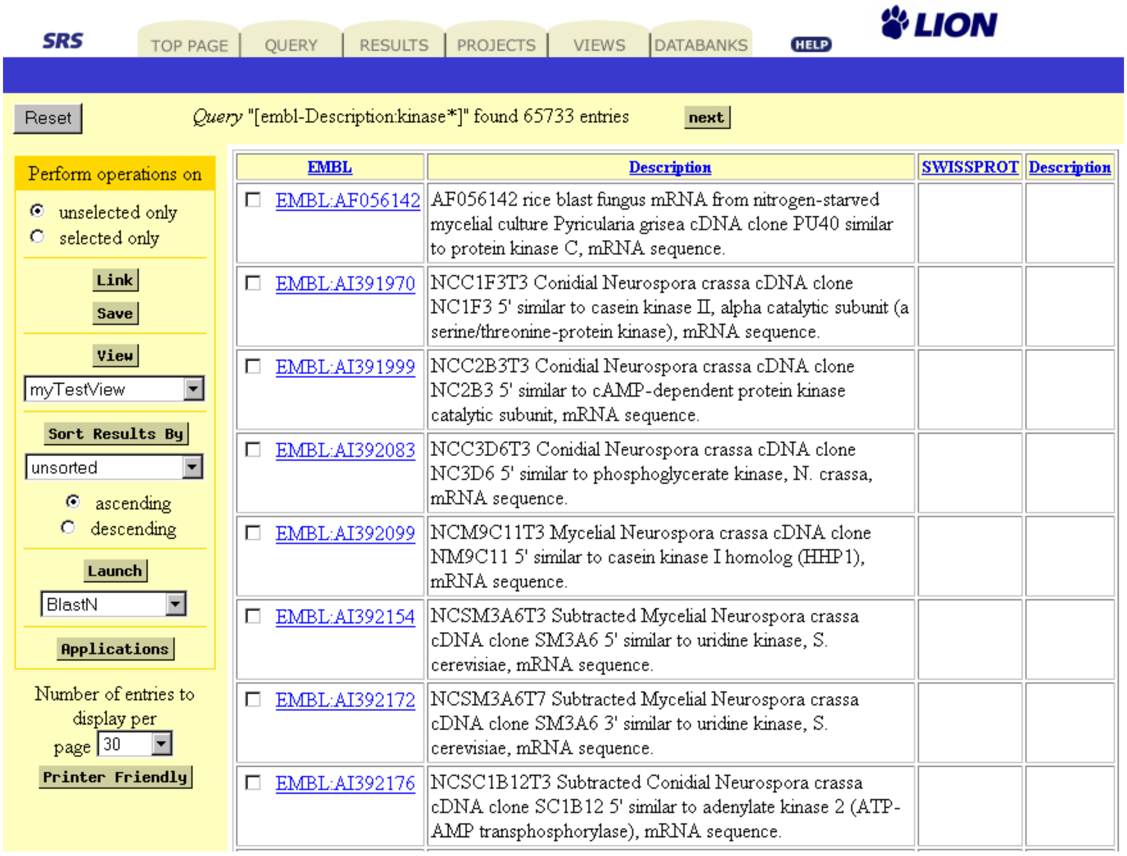


Successful Queries

Name	Type	N Total	From Databank	N	Query Expression	Comment
☐ Q3	link	1	SWISSPROT	1	([EMBL-ID:AF116175]	
☐ Q2	select	1	EMBL	1	[EMBL-ID:AF116175]	
☑ Q1	query	63320	EMBL	65733	[embl-Description:ki	

Figure 1.14 View a set of results using your own view.

4. Click the View button to display the query result set, **Q1**, using **myTestView**. The entries in **Q1** will be displayed using the view you selected.



Query "[embl-Description:kinase*]" found 65733 entries

Perform operations on
● unselected only
○ selected only

myTestView

unsorted
● ascending
○ descending

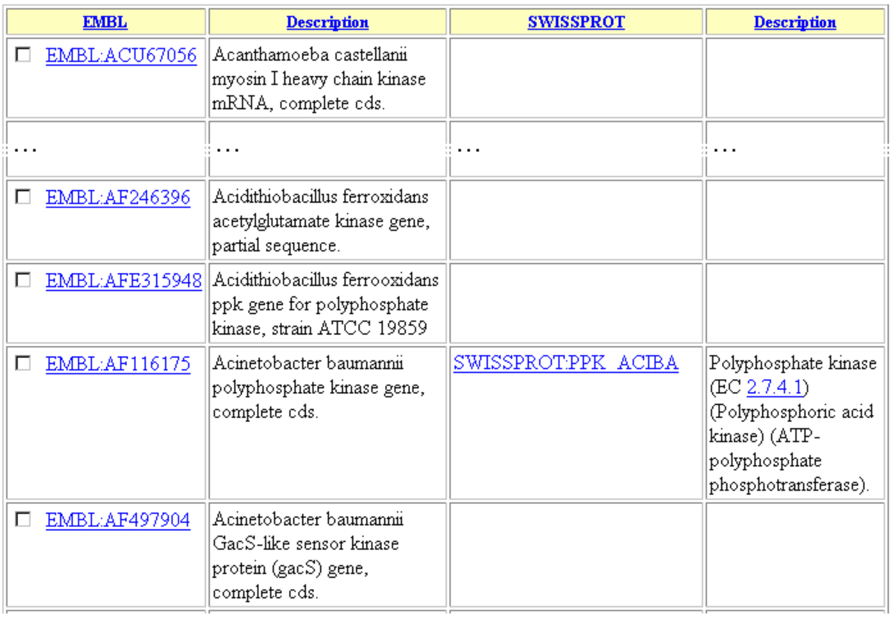
BlastN

Number of entries to display per page 30

EMBL	Description	SWISSPROT	Description
☐ EMBL:AF056142	AF056142 rice blast fungus mRNA from nitrogen-starved mycelial culture Pyricularia grisea cDNA clone PU40 similar to protein kinase C, mRNA sequence.		
☐ EMBL:AI391970	NCC1F3T3 Conidial Neurospora crassa cDNA clone NC1F3 5' similar to casein kinase II, alpha catalytic subunit (a serine/threonine-protein kinase), mRNA sequence.		
☐ EMBL:AI391999	NCC2B3T3 Conidial Neurospora crassa cDNA clone NC2B3 5' similar to cAMP-dependent protein kinase catalytic subunit, mRNA sequence.		
☐ EMBL:AI392083	NCC3D6T3 Conidial Neurospora crassa cDNA clone NC3D6 5' similar to phosphoglycerate kinase, N. crassa, mRNA sequence.		
☐ EMBL:AI392099	NCM9C11T3 Mycelial Neurospora crassa cDNA clone NM9C11 5' similar to casein kinase I homolog (HHP1), mRNA sequence.		
☐ EMBL:AI392154	NCSM3A6T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 5' similar to uridine kinase, S. cerevisiae, mRNA sequence.		
☐ EMBL:AI392172	NCSM3A6T7 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 3' similar to uridine kinase, S. cerevisiae, mRNA sequence.		
☐ EMBL:AI392176	NCSC1B12T3 Subtracted Conidial Neurospora crassa cDNA clone SC1B12 5' similar to adenylate kinase 2 (ATP-AMP transphosphorylase), mRNA sequence.		

Figure 1.15 Query Result page showing typical EMBL entries and any corresponding SWISSPROT entries, shown using **myTestView**.

If you cannot see any entries with links, try looking for the entry for which you found the original link. The entry (EMBL:AF116175) used to generate the link images above was in the lower part of the first page of results, when the results were sorted by **Organism Name**. By sorting the results again, an entry with a link was found. (Try relocating your own linked entry by repeating the process you used to find a linked entry previously. If you had to try linking to a databank other that SWISS-PROT, you may want to create a view that will show EMBL and your chosen databank, rather than EMBL and SWISS-PROT. Do not worry, at this stage, if you cannot find your result.)



EMBL	Description	SWISSPROT	Description
☐ EMBL:ACU67056	Acanthamoeba castellanii myosin I heavy chain kinase mRNA, complete cds.		
…	…	…	…
☐ EMBL:AF246396	Acidithiobacillus ferroxidans acetylglutamate kinase gene, partial sequence.		
☐ EMBL:AFE315948	Acidithiobacillus ferrooxidans ppk gene for polyphosphate kinase, strain ATCC 19859		
☐ EMBL:AF116175	Acinetobacter baumannii polyphosphate kinase gene, complete cds.	SWISSPROT:PPK_ACIBA	Polyphosphate kinase (EC 2.7.4.1) (Polyphosphoric acid kinase) (ATP-polyphosphate phosphotransferase).
☐ EMBL:AF497904	Acinetobacter baumannii GacS-like sensor kinase protein (gacS) gene, complete cds.		

Figure 1.16 Part of a **Query Result** page showing the EMBL entry AF116175 and its corresponding SWISSPROT entry, shown using **myTestView**.

1.6.3 Deleting a View

If you no longer need a view, you can delete it.

1. From the **View Manager** page 1 (use the **VIEWS** tab to get there) select the view you want to delete, e.g. **myTestView**.

Select view to delete: myTestView

Figure 1.17 Selecting a view to delete.

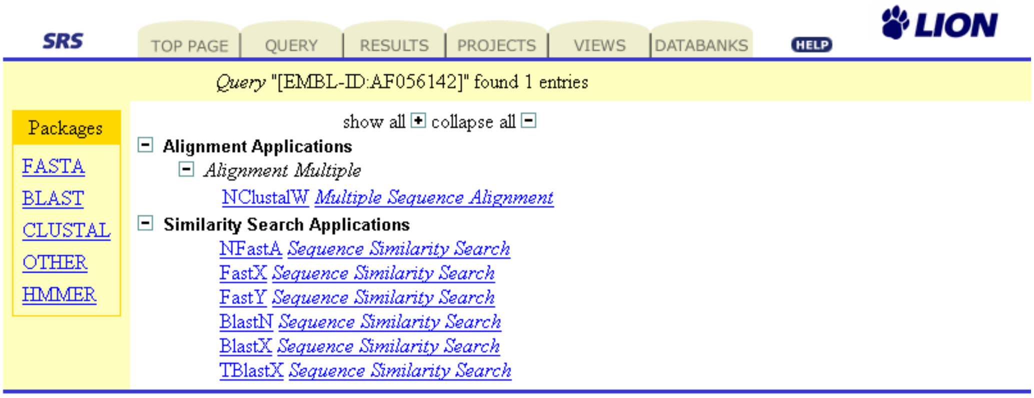
2. Click the `Delete View` button.

1.7 Using Applications

See “Applications” - Chapter 6 for more information.

SRS is able to analyze the results of your search using many bioinformatics analysis tools or applications. This enables you to seek out further information that may be relevant to your initial search. This section will take you through running an application on a set of results.

1. Go to a **Query Result** page.
   (If you can’t remember how to get to the **Query Result** page, use the **Results** tab to go to the **Manage your Query Results** page, tick the check box beside a set of results, set the view to something appropriate (e.g. **default view**), and click the View button to take you to the **Query Result** page for that set of results.)
2. Select an entry from the list of the results, by ticking the check box beside it. It will be the information in this entry that is then used to run the application.
3. Click the Applications button. This will take you to the **Application Select** page which shows all the applications that can be used on the current entry. You may need to open up the lists of applications by clicking on the ⊞ next to a group.



Query "[EMBL-ID:AF056142]" found 1 entries

show all ⊞ collapse all ⊟

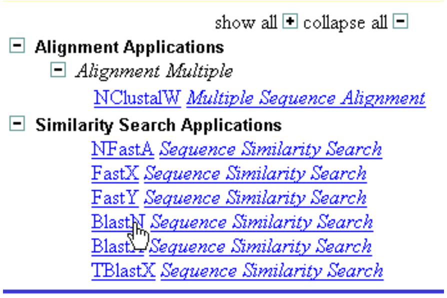
⊟ **Alignment Applications**
  ⊟ *Alignment Multiple*
    NClustalW *Multiple Sequence Alignment*
⊟ **Similarity Search Applications**
    NFastA *Sequence Similarity Search*
    FastX *Sequence Similarity Search*
    FastY *Sequence Similarity Search*
    BlastN *Sequence Similarity Search*
    BlastX *Sequence Similarity Search*
    TBlastX *Sequence Similarity Search*

Figure 1.18 **Application Select** page.

Note: You can also access many of the most commonly used applications using the drop-down list below the **Launch** button. Selecting the application from the drop-down list and pressing the **Launch** button takes you directly to the **Launch** page, see step 4.

---

4. If you have selected a query for which **BlastN** is available, then click on the hyperlink for **BlastN**.

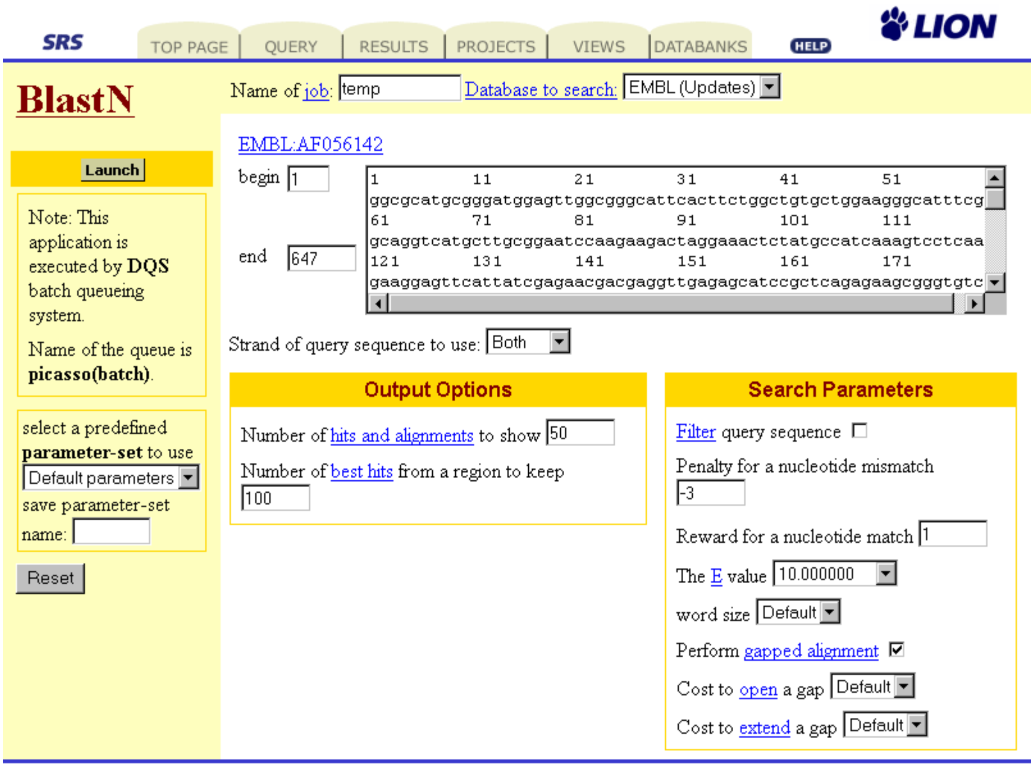


show all ⊞ collapse all ⊟

⊟ **Alignment Applications**
  ⊟ *Alignment Multiple*
    NClustalW *Multiple Sequence Alignment*
⊟ **Similarity Search Applications**
    NFastA *Sequence Similarity Search*
    FastX *Sequence Similarity Search*
    FastY *Sequence Similarity Search*
    BlastN *Sequence Similarity Search*
    BlastX *Sequence Similarity Search*
    TBlastX *Sequence Similarity Search*

Figure 1.19 Selecting **BlastN**.

5. This will take you to the **Launch** page.
   If **BlastN** is not available, then try again with another query, or select one of the other applications that is available, noting that the images below refer to BLASTN, and will be slightly different for your case.



BlastN

Name of job: temp Database to search: EMBL (Updates)

EMBL:AF056142

begin 1

end 647

```
1         11        21        31        41        51
ggcgcatgcgggatggagttggcgggcattcacttctggctgtgctggaagggcatttcg
61        71        81        91        101       111
gcaggtcatgcttgcggaatccaagaagactaggaaactctatgccatcaaagtcctcaa
121       131       141       151       161       171
gaaggagttcattatcgagaacgacgaggttgagagcatccgctcagagaagcgggtgtc
```

Strand of query sequence to use: Both

Note: This application is executed by **DQS** batch queueing system.

Name of the queue is **picasso(batch)**.

select a predefined **parameter-set** to use
Default parameters
save parameter-set name:

Output Options

Number of hits and alignments to show 50

Number of best hits from a region to keep 100

Search Parameters

Filter query sequence ☐

Penalty for a nucleotide mismatch -3

Reward for a nucleotide match 1

The E value 10.000000

word size Default

Perform gapped alignment ☑

Cost to open a gap Default

Cost to extend a gap Default

Figure 1.20 Default settings for the **Launch** page.

Note: If you want to know more about any of the applications, click on the short description hyperlink beside any of the applications listed on the **Application Select** page. (For example, to find out more about BLASTN, click on *Sequence Similarity Search*)

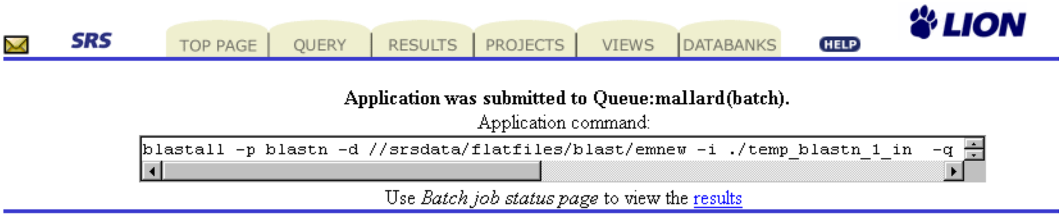
---

6. When you run an application, you can set values for the run-time options on the **Launch** page. For this example, use the **Default parameters**. More information about application options is available in chapter 6, “Applications”, (p. 131).
   In addition, for applications that will be sent to a batch queue, there is a drop-down list that allows you to choose which batch queue, where there is more than one batch queue available to you. This is not shown in Figure 1.20 because there is only one batch queue available on that installation.

Note: Depending on how your installation is set up, the application may be run as a batch job, using a batch queueing system, or it may be run interactively. Figure 1.20 shows that BLASTN will be run as a batch job for this SRS installation, but this may not be the case on your installation. Section 1.7.1 “Running a Batch Queued Application” (p. 23) demonstrates how to run a batch application, while section 1.7.2 “Running an Interactive Application” (p. 25) describes how to run an interactive application.

1.7.1 Running a Batch Queued Application

1. Click the Launch button to start the application running (see Figure 1.21).



Application was submitted to Queue:mallard(batch).

Application command:

```
blastall -p blastn -d //srsdata/flatfiles/blast/emnew -i ./temp_blastn_1_in  -q
```

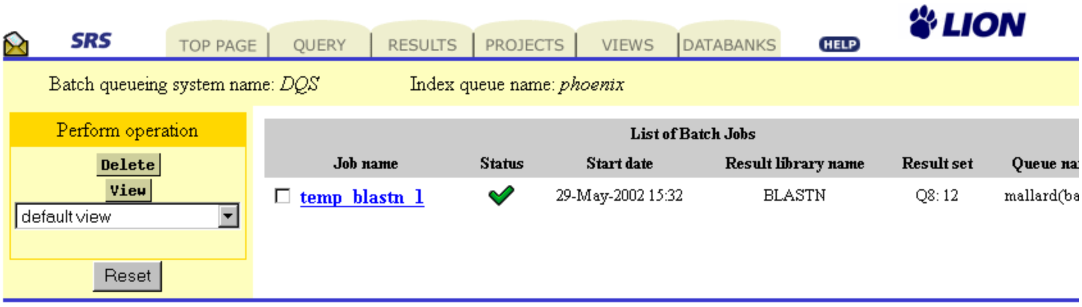
Use *Batch job status page* to view the results

Figure 1.21 Application Invocation page.

Note: The ✉ icon that appears at the top left of your SRS window, indicates that you have a batch job running. (See section 6.4.3 “Batch Applications”, p. 136 for more information on this icon and others used when running batch applications.)

---

2. Click on the ✉, at the top left corner, to see the **Job Status** page, which will show the status of any batch queue jobs you have run.



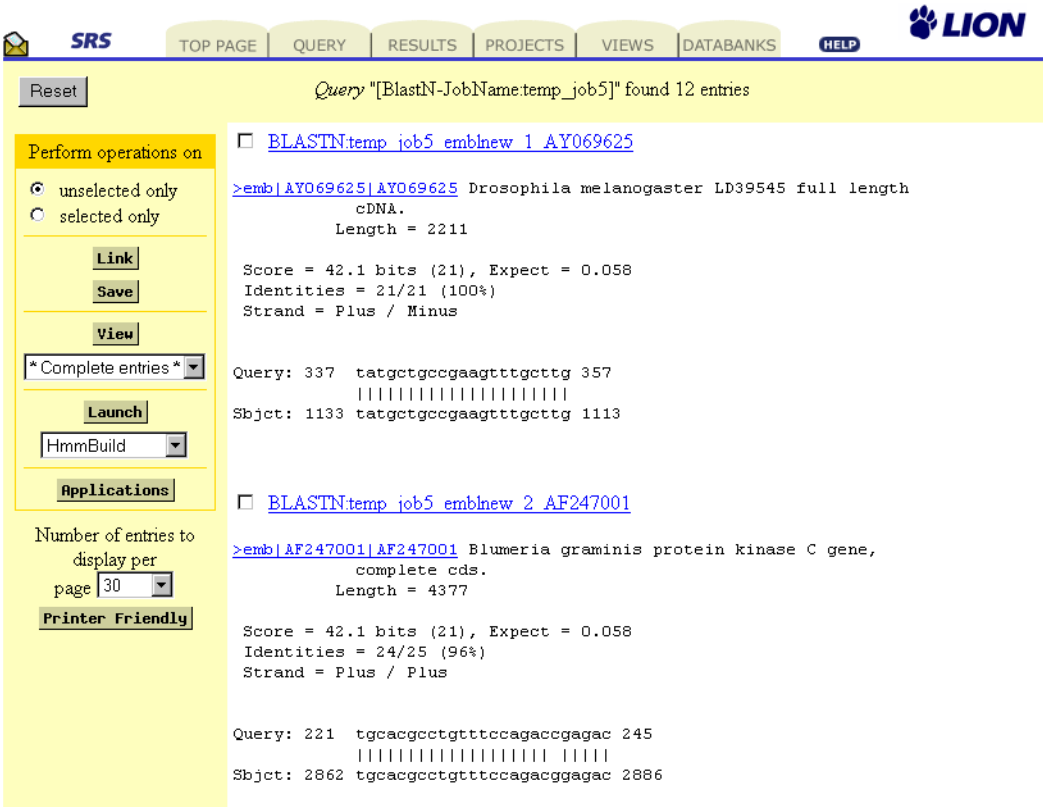
Batch queueing system name: *DQS* Index queue name: *phoenix*

List of Batch Jobs

	Job name	Status	Start date	Result library name	Result set	Queue na
☐	temp_blastn_1	✓	29-May-2002 15:32	BLASTN	Q8: 12	mallard(ba

Figure 1.22 Job Status page.

3. The ✓ icon shows that the application has finished. Click on the hyperlink **temp_blastn_1** to access the results.



Query "[BlastN-JobName:temp_job5]" found 12 entries

☐ BLASTN:temp_job5_emblnew_1_AY069625

```
>emb|AY069625|AY069625 Drosophila melanogaster LD39545 full length
            cDNA.
          Length = 2211

 Score = 42.1 bits (21), Expect = 0.058
 Identities = 21/21 (100%)
 Strand = Plus / Minus


Query: 337  tatgctgccgaagtttgcttg 357
            |||||||||||||||||||||
Sbjct: 1133 tatgctgccgaagtttgcttg 1113
```

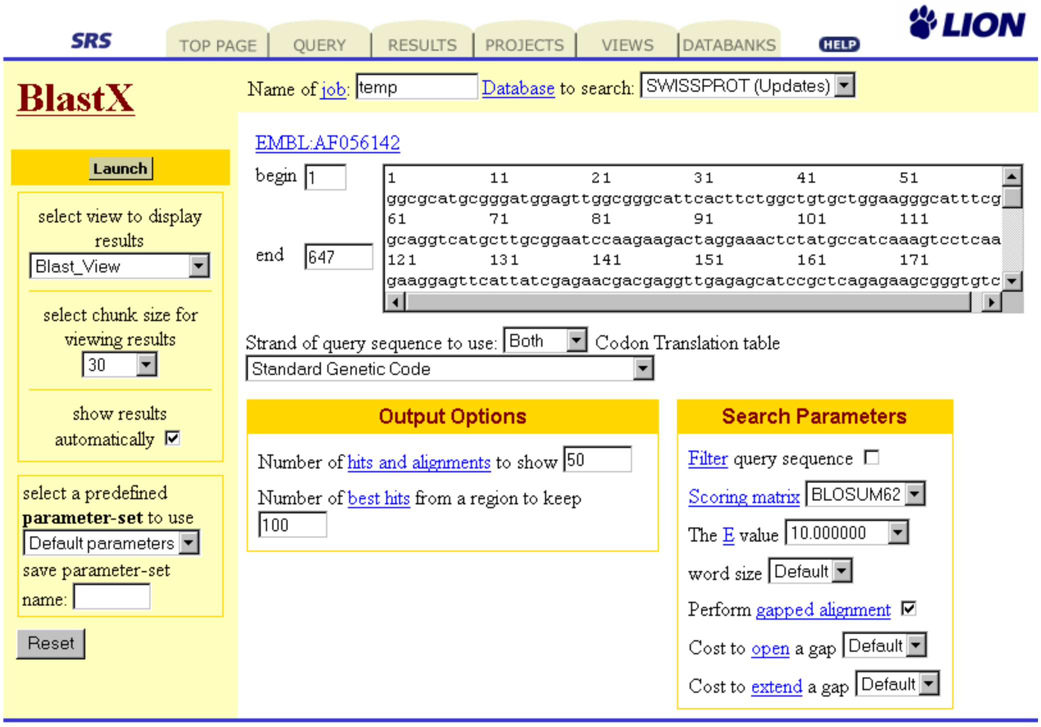
☐ BLASTN:temp_job5_emblnew_2_AF247001

```
>emb|AF247001|AF247001 Blumeria graminis protein kinase C gene,
            complete cds.
          Length = 4377

 Score = 42.1 bits (21), Expect = 0.058
 Identities = 24/25 (96%)
 Strand = Plus / Plus


Query: 221  tgcacgcctgtttccagaccgagac 245
            ||||||||||||||||||| |||||
Sbjct: 2862 tgcacgcctgtttccagacggagac 2886
```

Figure 1.23 Entry page showing the application results.

1.7.2 Running an Interactive Application

1. If your application is not run via a batch queueing system (i.e. it is run interactively) then the **Launch** page should look something like that shown in Figure 1.24.



BlastX

Launch

select view to display results: Blast_View

select chunk size for viewing results: 30

show results automatically ☑

select a predefined **parameter-set** to use: Default parameters

save parameter-set name:

Reset

Name of job: temp Database to search: SWISSPROT (Updates)

EMBL:AF056142

begin 1

end 647

```
1         11        21        31        41        51
ggcgcatgcgggatggagttggcgggcattcacttctggctgtgctggaagggcatttcg
61        71        81        91        101       111
gcaggtcatgcttgcggaatccaagaagactaggaaactctatgccatcaaagtcctcaa
121       131       141       151       161       171
gaaggagttcattatcgagaacgacgaggttgagagcatccgctcagagaagcgggtgtc
```

Strand of query sequence to use: Both Codon Translation table: Standard Genetic Code

Output Options

Number of hits and alignments to show 50

Number of best hits from a region to keep 100

Search Parameters

Filter query sequence ☐

Scoring matrix BLOSUM62

The E value 10.000000

word size Default

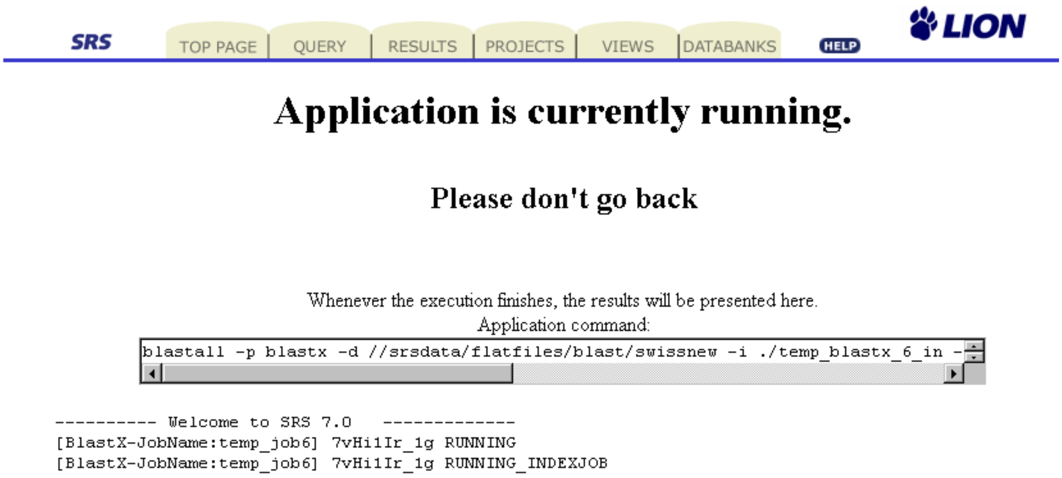
Perform gapped alignment ☑

Cost to open a gap Default

Cost to extend a gap Default

Figure 1.24 Launch page for an interactive application.

2. Use the default parameters (see step 6, above).



Application is currently running.

Please don't go back

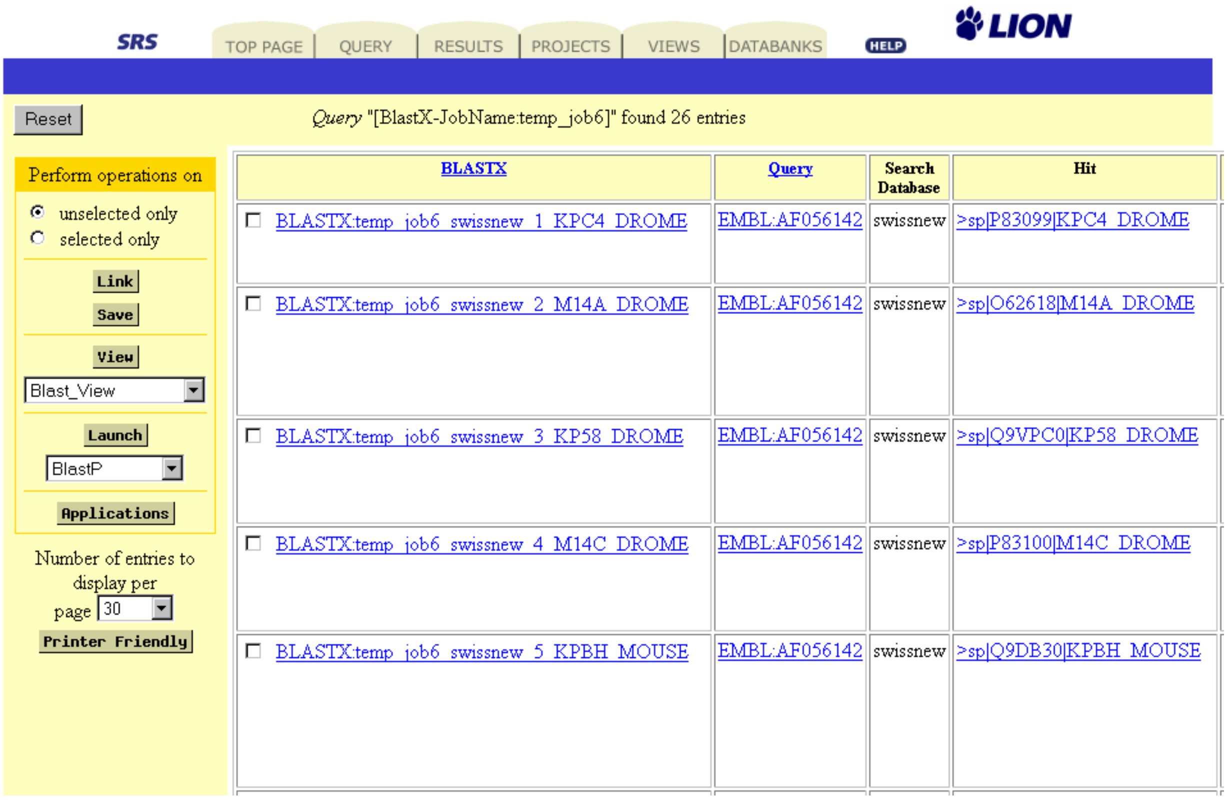
Whenever the execution finishes, the results will be presented here.
Application command:

```
blastall -p blastx -d //srsdata/flatfiles/blast/swissnew -i ./temp_blastx_6_in -
```

```
---------- Welcome to SRS 7.0   -------------
[BlastX-JobName:temp_job6]  7vHi1Ir_1g RUNNING
[BlastX-JobName:temp_job6]  7vHi1Ir_1g RUNNING_INDEXJOB
```

Figure 1.25 Application Invocation page for an interactive application.

3. When the application finishes running, the results will be displayed automatically. See Figure 1.26.



Query "[BlastX-JobName:temp_job6]" found 26 entries

Number of entries to display per page 30

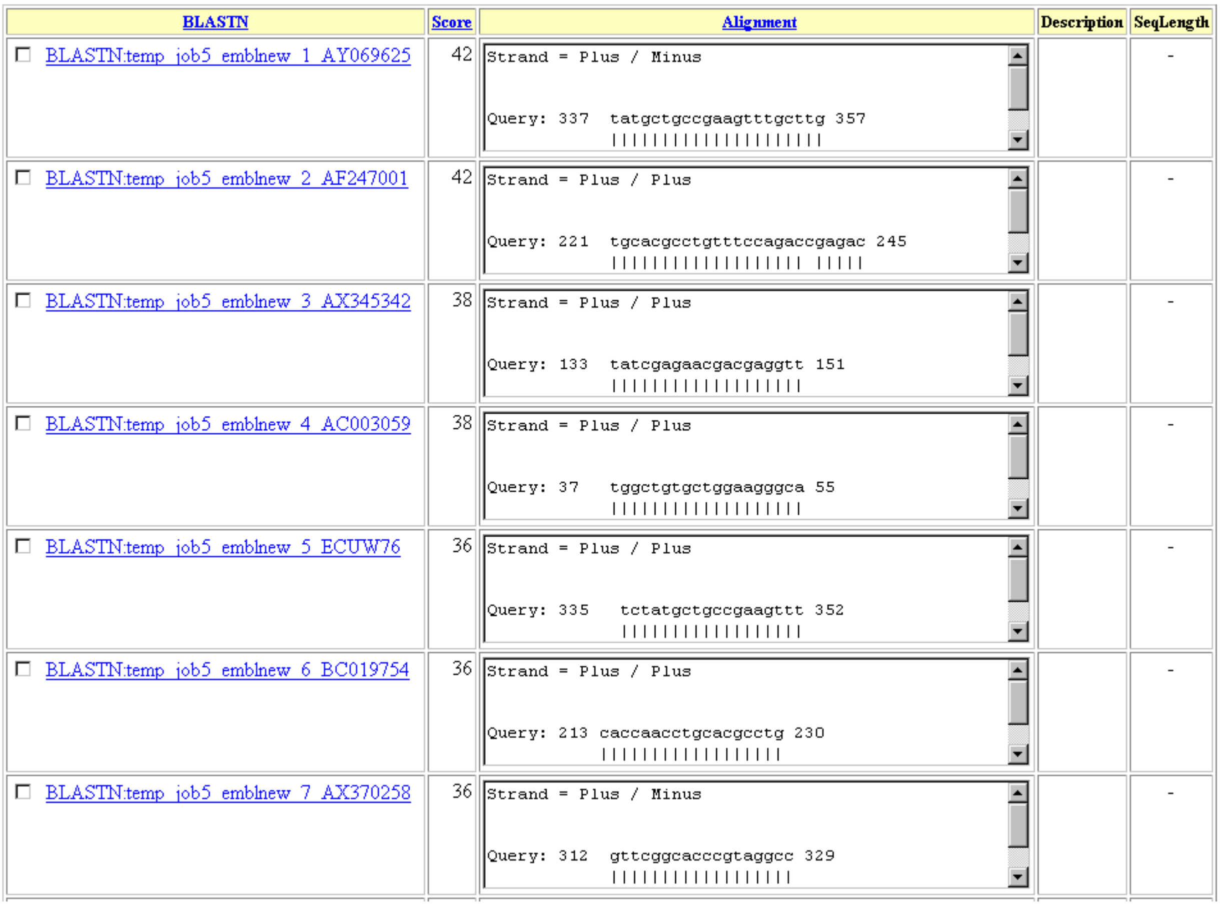
	BLASTX	Query	Search Database	Hit
☐	BLASTX:temp_job6_swissnew_1_KPC4_DROME	EMBL:AF056142	swissnew	>sp|P83099|KPC4_DROME
☐	BLASTX:temp_job6_swissnew_2_M14A_DROME	EMBL:AF056142	swissnew	>sp|O62618|M14A_DROME
☐	BLASTX:temp_job6_swissnew_3_KP58_DROME	EMBL:AF056142	swissnew	>sp|Q9VPC0|KP58_DROME
☐	BLASTX:temp_job6_swissnew_4_M14C_DROME	EMBL:AF056142	swissnew	>sp|P83100|M14C_DROME
☐	BLASTX:temp_job6_swissnew_5_KPBH_MOUSE	EMBL:AF056142	swissnew	>sp|Q9DB30|KPBH_MOUSE

Figure 1.26 Query Result page showing the application results.

1.7.3 Viewing Application Results

1. SRS provides many ways to view application results.
2. Using the drop-down list below the **View** button, select **BlastAlignment**. If this option is not available, choose another view type, noting that the figures presented below may refer to **BlastAlignment** and were generated for the BlastN results created above.

3. Click the **View** button. This displays the results of the application using the **BlastAlignment** view. See Figure 1.27.



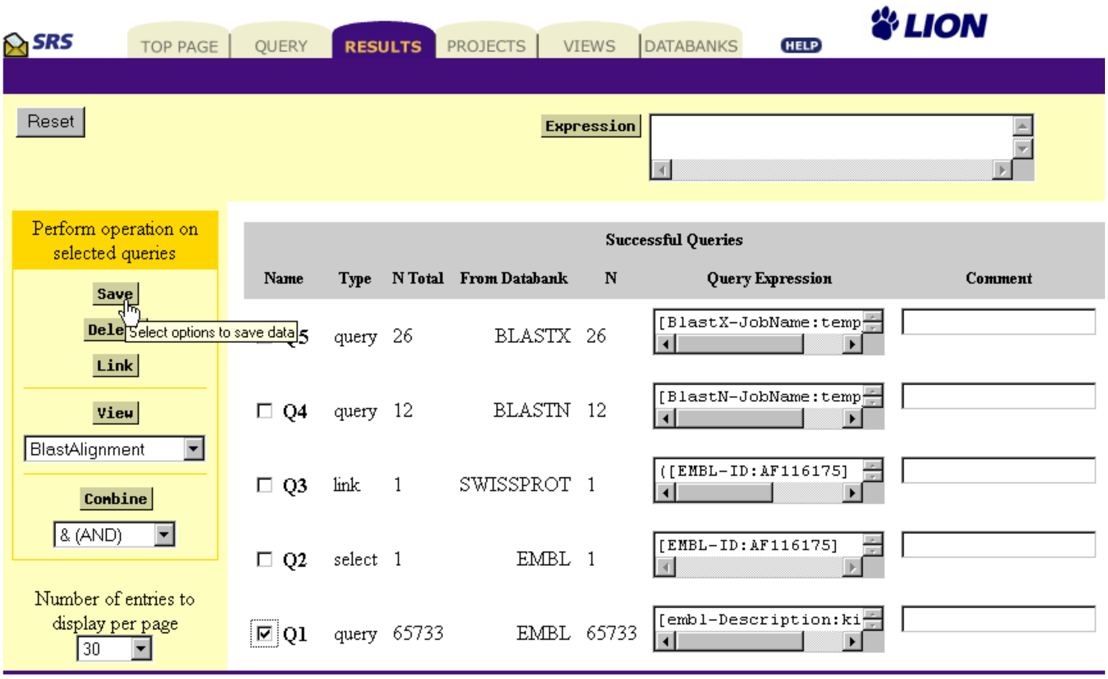
BLASTN	Score	Alignment	Description	SeqLength
☐ BLASTN:temp job5 emblnew 1 AY069625	42	Strand = Plus / Minus Query: 337 tatgctgccgaagtttgcttg 357 |||||||||||||||||||||		-
☐ BLASTN:temp job5 emblnew 2 AF247001	42	Strand = Plus / Plus Query: 221 tgcacgcctgtttccagaccgagac 245 ||||||||||||||||||| |||||		-
☐ BLASTN:temp job5 emblnew 3 AX345342	38	Strand = Plus / Plus Query: 133 tatcgagaacgacgaggtt 151 |||||||||||||||||||		-
☐ BLASTN:temp job5 emblnew 4 AC003059	38	Strand = Plus / Plus Query: 37 tggctgtgctggaagggca 55 |||||||||||||||||||		-
☐ BLASTN:temp job5 emblnew 5 ECUW76	36	Strand = Plus / Plus Query: 335 tctatgctgccgaagttt 352 ||||||||||||||||||		-
☐ BLASTN:temp job5 emblnew 6 BC019754	36	Strand = Plus / Plus Query: 213 caccaacctgcacgcctg 230 ||||||||||||||||||		-
☐ BLASTN:temp job5 emblnew 7 AX370258	36	Strand = Plus / Minus Query: 312 gttcggcacccgtaggcc 329 ||||||||||||||||||		-

Figure 1.27 Part of a **Query Result** page showing one of the predefined SRS views.

1.8 Saving your Results

See section 2.2 “Temporary Projects” (p. 37) for more information.

1. SRS allows you to save queries so you can use them later. Click the **Results** tab to take you to the **Manage your Query Results** page. Here you can choose which queries you want to save.
2. Select a query that you want to save by ticking the check box beside it.

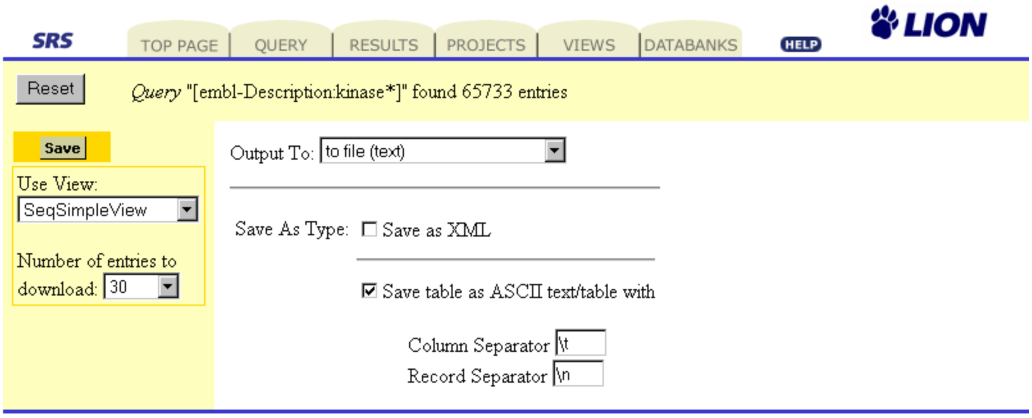


Successful Queries

Name	Type	N Total	From Databank	N	Query Expression	Comment
Q5	query	26	BLASTX	26	[BlastX-JobName:temp	
☐ Q4	query	12	BLASTN	12	[BlastN-JobName:temp	
☐ Q3	link	1	SWISSPROT	1	([EMBL-ID:AF116175]	
☐ Q2	select	1	EMBL	1	[EMBL-ID:AF116175]	
☑ Q1	query	65733	EMBL	65733	[embl-Description:ki	

Figure 1.28 Select a query to save.

3. Click the **Save** button to display the **DownLoad Options** page.
4. Use the **DownLoad Options** page to specify what is saved, the way the output is saved and where it goes (to a text file, or to the screen (**Output To:** drop-down list).



Query "[embl-Description:kinase*]" found 65733 entries

Output To: to file (text)

Use View: SeqSimpleView

Number of entries to download: 30

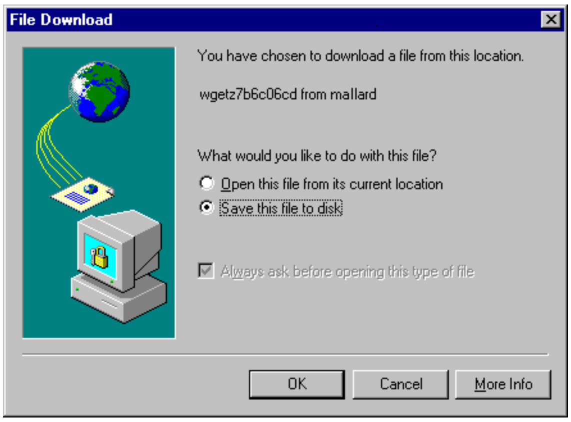
Save As Type: ☐ Save as XML

☑ Save table as ASCII text/table with

Column Separator \t
Record Separator \n

Figure 1.29 Set the download options on the **DownLoad Options** page.

5. Click the Save button. Your browser's **File Download** dialog box (Figure 1.30) will appear.



You have chosen to download a file from this location.

wgetz7b6c06cd from mallard

What would you like to do with this file?
○ Open this file from its current location
● Save this file to disk

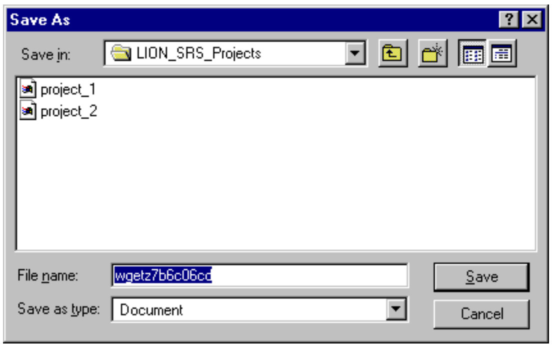
☑ Always ask before opening this type of file

Figure 1.30 The **File Download** dialog box.

Note: Dialog boxes shown throughout this User Guide were generated in the WindowsNT version of Internet Explorer 5. The dialog box you see may look different depending on your system and web browser.

---

6. Make sure that **Save this file to disk** is selected and click the OK button. This will display the **Save As** dialog box.



Save in: LION_SRS_Projects

File name: wgetz7b6c06cd

Save as type: Document

Figure 1.31 The **Save As** dialog box.

7. Select a location for the file, choose a suitable name and then click the Save button to save the file.

1.9 Review

You have now completed a brief overview of the key features of SRS and should be ready to start using it on your own. You should now know how to start an SRS project, be able to perform a query using SRS, link your query results to other databanks, change the way in which your results are displayed by creating your own view, run an application, and save your working project. If you want more information on any particular subject, refer to the relevant chapter.

CHAPTER

2

SRS PROJECTS

All the work you do using SRS will be within projects. These are simply a way of keeping related work together, and any queries, views, etc. that are created in a project will be stored in the project history. This chapter introduces SRS projects. By the end of this chapter you will have learned more about:

- The benefits of using temporary and permanent projects.
- How to start or return to a project.
- Features that are available in permanent projects.
- How to make temporary project data available in other projects.
- How to move temporary project work to a permanent project.

2.1 Introducing Projects

There are two types of project that can be used with SRS: temporary and permanent. These are described below.

2.1.1 Temporary Projects

When you use a temporary project, your queries and views are stored in a temporary location. They may remain available for a time after you have finished working, but you should not rely on this. If you bookmark the page in your web browser, you should be able to return to the project until the System Administrator clears the files.

You should use a temporary project for:

- Simple searches. For example, a temporary project is useful if you want to look something up quickly, or run an occasional BLAST search.

To find out more about running a temporary project see section 2.2 “Temporary Projects” (p. 37).

2.1.2 Permanent Projects

Permanent projects are used within a permanent SRS user account. There may be one or more (up to 99) projects within any such account. All your queries and views are stored in a project or projects. The fact that the projects are part of a user account, means that they remain available for you to use in the future, whenever you choose to return to that user account. User accounts (and the permanent projects within them) can also be password protected, allowing you to restrict access.

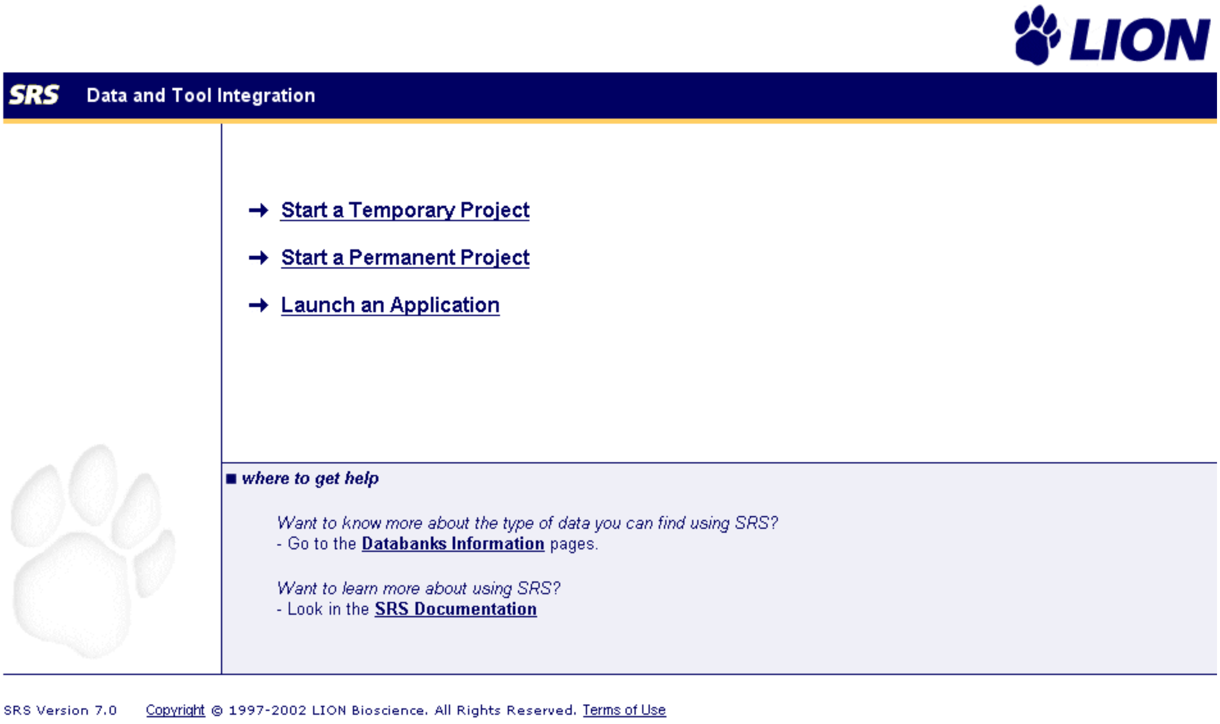
You should work in a permanent project if any of the following apply:

- You or a colleague will want to return to a project at a later time.
- You want to be able to move your work from one project to another.
- You want to keep a safe record of all your projects.
- You want to recall any of your previous permanent or saved projects.

To find out more about running a permanent project see section 2.3 “Permanent Projects” (p. 40).

2.1.3 Starting a Project

Whether you decide to use a temporary project or a permanent project, you will usually begin your project from the **Start** page (see Figure 2.1).



SRS Data and Tool Integration

→ Start a Temporary Project

→ Start a Permanent Project

→ Launch an Application

■ *where to get help*

Want to know more about the type of data you can find using SRS?
- Go to the **Databanks Information** pages.

Want to learn more about using SRS?
- Look in the **SRS Documentation**

SRS Version 7.0 Copyright © 1997-2002 LION Bioscience. All Rights Reserved. Terms of Use

Figure 2.1 The **Start** page.

There are various links on the **Start** page. Holding the mouse cursor over any of the main options will display details about what that option does (see also section 1.2 “Starting an SRS Project”, p. 2). The main links allow you to:

- Start a temporary project.
- Start a permanent project (which allows you to return to projects in an existing account or open a new account).
- Launch an application.

There are also links which allow you to:

- Find out more information about the databanks available from your server.

- Look up the SRS on-line documentation that is available on your server.

The SRS on-line documentation consists of HTML and PDF versions of the available documentation. This includes this User Guide and online page by page SRS help (the SRS OnLine Help) as well as things designed to help SRS Administrators, e.g. the SRS Administrator’s Guide, the Icarus Guide, Icarus Quick Reference - Classes & Commands, etc.

2.2 Temporary Projects

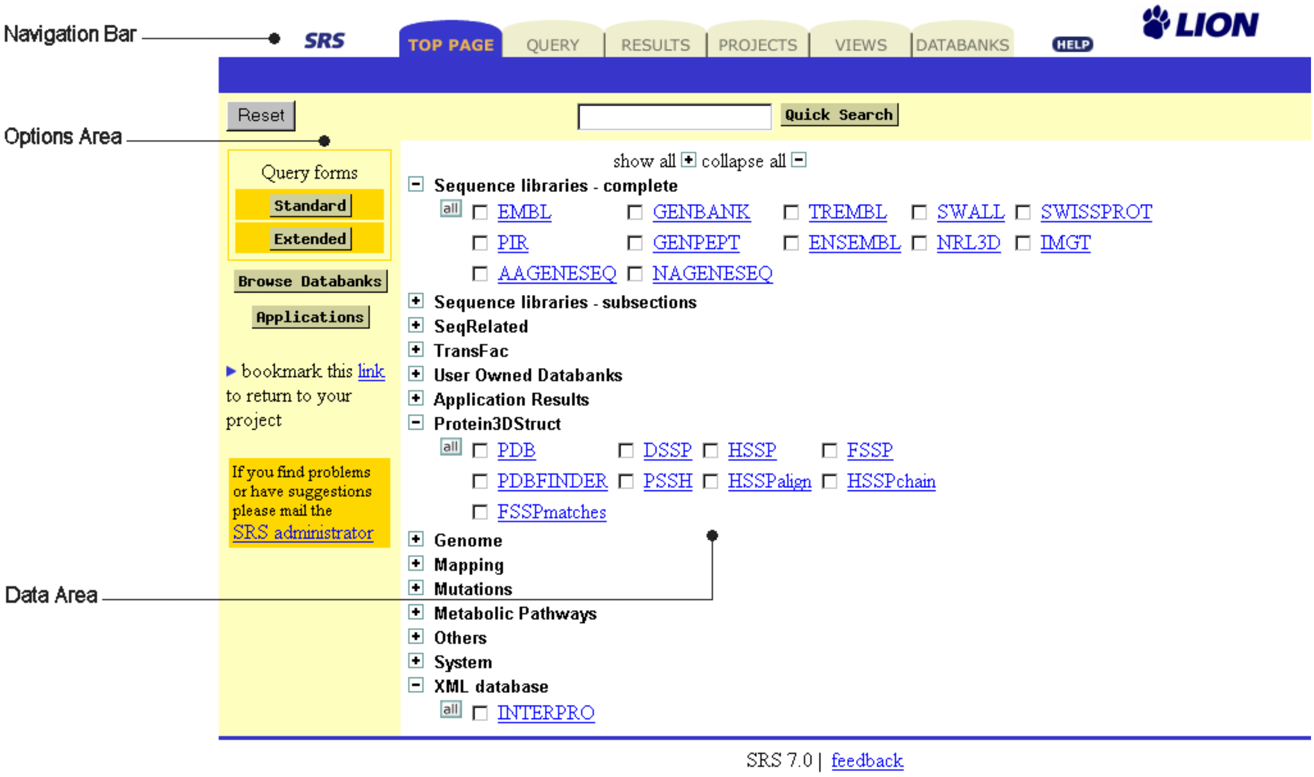
Temporary projects should only be used for temporary work. Temporary project data will be removed whenever your System Administrator performs a routine clean-up operation.

2.2.1 Starting a Temporary Project

A temporary project is started from the **Start** page (see section 2.1.3 “Starting a Project” (p. 35) for more details):

1. Click the **Start a Temporary Project** link.
   This takes you to the **Select Databanks To Search** page (**Top** page) where you can begin working with SRS.

Note: Clicking on the **Databanks Information page** or **SRS Documentation** links will also start a temporary project on the **List of Databanks** or **SRS Documentation** pages, respectively.



▸ bookmark this link to return to your project

If you find problems or have suggestions please mail the SRS administrator

Figure 2.2 The **Select Databanks To Search** page.

2.2.2 Managing Temporary Projects

If you click the **Projects** tab on the navigation bar, the **Project Manager** page will be displayed (see Figure 2.3).

This temporary project contains:

Queries

name	*query*
Q6	([swissprot-Organism:homo sapiens] & (([swissprot-Description:kinase*] & [swissprot-Description:inhibitor*]) | [swissprot-Description:kinase inhibitor*]))
Q5	[libs={embl swissprot} -Organism:toad
Q4	[swissprot-ID:rat*]
Q3	[swissprot-Description:cellulase*]
Q2	[embl-Description:cellulase*]
Q1	[embl-Description:kinase*]

Views

Swiss_Embl_View
Embl_View
Swiss_View

Figure 2.3 Project Manager for temporary projects.

The **Project Manager** page provides you with the necessary tools to manage your SRS projects. Here you can save projects and open previously saved projects. After completing your project you may want to save your work to disk. (The next time you access SRS, unsaved temporary project work is unlikely to be available from the server.) Saving is also useful for moving temporary work to your permanent project history list, or permanent project work to a temporary project, or if you want to share project work with another user.

Unlike permanent projects, you cannot rename, delete, or switch between temporary projects, nor can you share queries and views. This means that any queries or views created for one project will usually have to be recreated from scratch if you wish to use them with another (temporary) project.

The **Project Manager** page is described in section 2.4 “Using the Project Manager” (p. 43).

2.3 Permanent Projects

Permanent projects exist within SRS user accounts. There are two types of account: secure and non-secure.

- A secure user account gives you an httpd password. This ensures that the account and the permanent projects within it are available only to those authorized to access it.
- A non-secure user account also uses a login procedure to identify the account that you wish to open, but does not enforce any kind of access control. Anyone who knows the account name can view the account and the permanent projects within it.

Aermanent project, regardless of the type of account used, stores all your project data in a single location from where it can be recalled.

Within your user account you can swap between projects and move work from one project to another, as well as retrieving saved projects, e.g. from other accounts to which you have access or from temporary projects that have been saved.

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

2.3.1 Starting a Permanent Project

This section describes how to open a permanent project by entering your user account. SRS will automatically create a new permanent project within your account if you are entering it for the first time. If you are re-entering an existing account, then the projects within that account will be available to you.

To start or return to a permanent project:

1. Click the **Start a Permanent Project** link on the **Start** page.

Secure or Non-Secure?

When you start a permanent project, your system will determine whether it is set up for secure or non-secure accounts. This was determined when SRS was installed, and you cannot choose at this point whether to use a secure or a non-secure account.

The system will display a **Log in** dialog, allowing you to log into your account. Typical dialogs for secure and non-secure accounts are given in Figure 2.4 and Figure 2.5 respectively.

Note: The dialog boxes shown throughout this manual were generated in the Windows32 version of Internet Explorer 5. The dialog boxes you see may look different, depending on your system and web browser.

Please type your user name and password.

Site: mallard.lionbio.co.uk

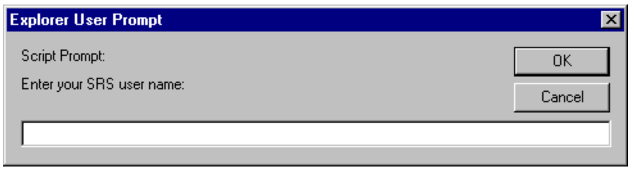
Realm session

User Name user1

Password

☐ Save this password in your password list

Figure 2.4 SRS secure (password protected) account **Log in** dialog box.


Explorer User Prompt

Script Prompt:

Enter your SRS user name:

Figure 2.5 SRS non-secure account **Log in** dialog box.

2. For secure account access complete the login information.
   The web browser will prompt you to give a **User Name** and **Password**. You may need to ask your SRS Administrator for these because they may not be the same as your system account ID and password.
   For non-secure account access enter your **user name** at the prompt.
3. Click OK.

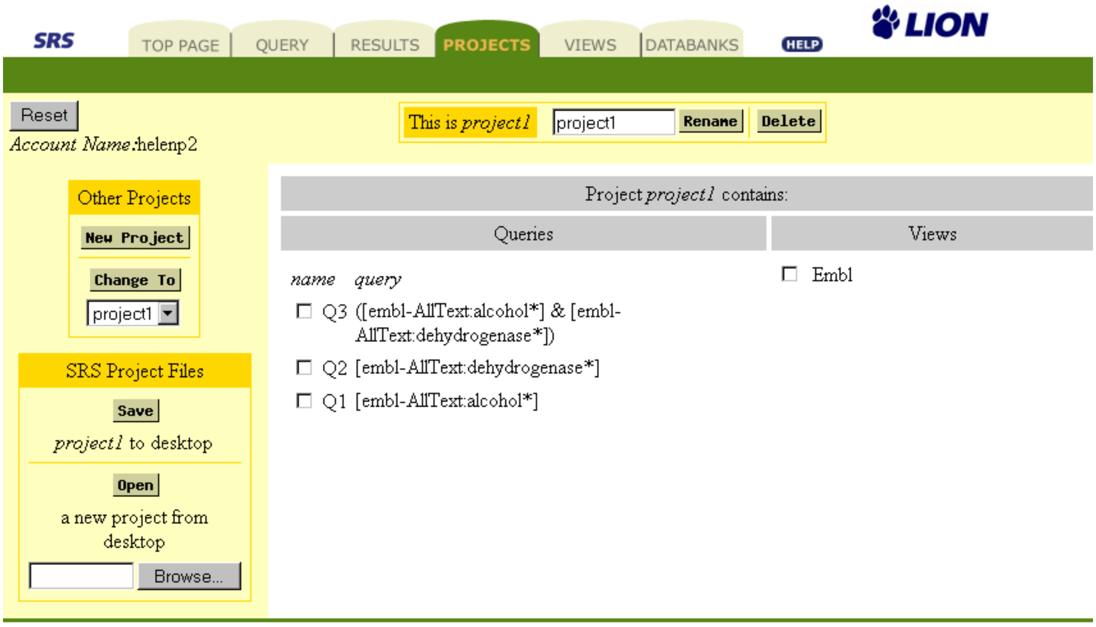
In either case the **Project Manager** page is displayed. From here you can continue working with the most recently used project, switch to another project, or create a new project. This is described in section 2.3.2 “Managing Permanent Projects” (p. 42) and section 2.4 “Using the Project Manager” (p. 43).

2.3.2 Managing Permanent Projects

If you followed the steps outlined in section 2.3.1 “Starting a Permanent Project” (p. 40), the **Project Manager** page will be displayed (see Figure 2.6).

The **Project Manager** page provides you with the necessary tools to manage your SRS projects. It allows you to move queries or views between projects, save and open projects, delete unwanted projects, start new projects, and swap to any other project currently on your account.

The **Project Manager** page is described in section 2.4 “Using the Project Manager” (p. 43).



Account Name:helenp2

project1

project1 to desktop

a new project from desktop

Project *project1* contains:

Queries		Views
name	*query*	☐ Embl
☐ Q3	([embl-AllText:alcohol*] & [embl-AllText:dehydrogenase*])	
☐ Q2	[embl-AllText:dehydrogenase*]	
☐ Q1	[embl-AllText:alcohol*]	

Figure 2.6 SRS **Project Manager** page for permanent projects.

2.4 Using the Project Manager

The Project Manager page is used for project administration. From this page you can create a new project, switch to another project, copy information between projects, open saved projects, save projects, rename projects and delete projects.

These functions are described below.

Note: Some functions, namely creating a new project, switching to other projects, deleting and renaming projects, are not available for temporary projects.

2.4.1 Creating a New Project

The **Project Manager** page allows you to create new permanent projects within an account and move between them (see also section 2.4.2 “Switching to Another Project”, p. 44).

Note: This feature is not available for temporary projects.

To create a new project:

1. Click the New Project button.

This will start a new project and display the **Select Databanks To Search** page.

2.4.2 Switching to Another Project

The **Project Manager** page allows you to move between projects that have been saved in your account.

Note: This feature is not available for temporary projects.

To return to a project that you have already saved:

1. Select the project you want to resume from the drop-down list, under the Change To button.
2. Click the Change To button.
   The project you have just selected is now displayed on the **Project Manager** page. You can work with this project as usual.

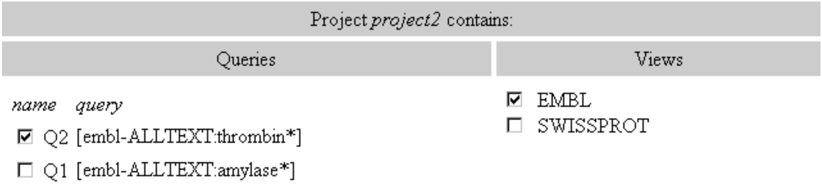
2.4.3 Copying Project Information

You can copy information such as views and queries from one project to another.

Note: This option will only be available if you have more than one project in your account. This option is not available for temporary projects.

To copy project information:

1. Make sure you are in the project from which you want to copy items (see section 2.4.2 “Switching to Another Project”, p. 44). Select the items you wish to copy by ticking the check boxes beside those items.



Project *project2* contains:

Queries

name *query*

☑ Q2 [embl-ALLTEXT:thrombin*]

☐ Q1 [embl-ALLTEXT:amylase*]

Views

☑ EMBL

☐ SWISSPROT

Figure 2.7 Choosing items to be copied.

2. Choose the project to which you want the items to be copied:

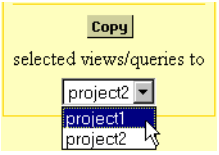


Figure 2.8 Choosing a project to which to copy items.

3. Click on the **Copy** button to copy the items.

This will copy the selected queries and views to the chosen project, and switch to using that project.

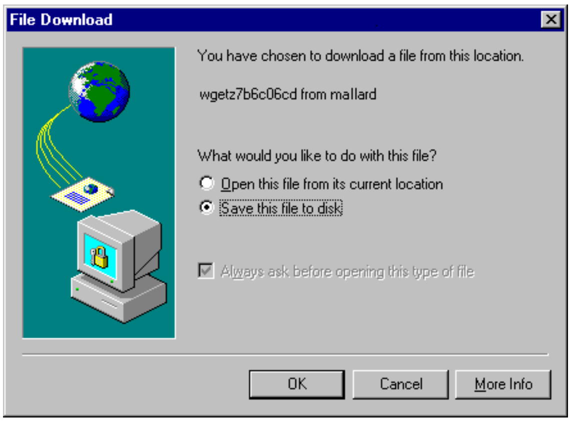
2.4.4 Saving a Project

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

To save a project:

1. Make it the current project.
   (See section 2.4.2 “Switching to Another Project”, p. 44.)
2. Click the **Save** button.

3. Select **Save this file to disk** when prompted.



You have chosen to download a file from this location.

wgetz7b6c06cd from mallard

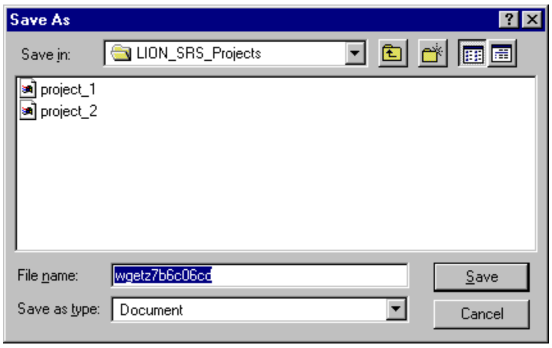
What would you like to do with this file?

○ Open this file from its current location
● Save this file to disk

☑ Always ask before opening this type of file

Figure 2.9 File Download dialog.

4. Click OK.
5. Give the project file a name.



Save in: LION_SRS_Projects

project_1
project_2

File name: wgetz7b6c06cd

Save as type: Document

Figure 2.10 Save As dialog.

6. Click Save.
   The file is now saved and the sysem is ready to save another project, if you wish.

2.4.5 Opening a Saved Project

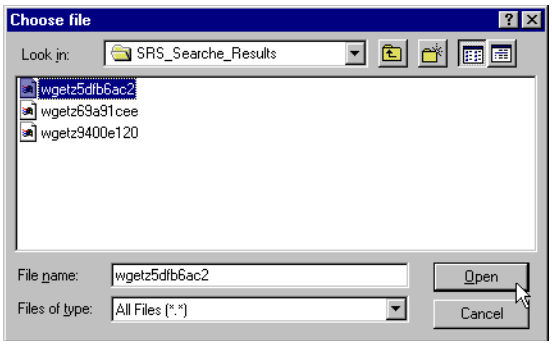
You can open a saved project (e.g. from a different account (to which you have access), or a temporary project that was saved) using the **Open** button on the **Project Manager** page.

To open a project from your local disk:

1. Type the file name in the text box below the **Open** button.
2. Click the **Open** button.

If you do not know where the project file is:

- Click the Browse... button.
- Locate the file to be opened.



Look in: SRS_Searche_Results

wgetz5dfb6ac2
wgetz69a91cee
wgetz9400e120

File name: wgetz5dfb6ac2

Files of type: All Files (*.*)

Figure 2.11 Browsing for a project file.

3. Click Open.
   This will put the file name and its directory path into the text box.
4. Click **Open**.
   This will open the project. You can start working with the project as usual.

2.4.6 Deleting a Project

When you no longer need a project you might want to delete it. This may become necessary if the number of projects for a particular permanent project account nears 99. (The upper limit for the number of projects available to a permanent project account is 99.)

Note: This feature is not available for temporary projects.

To delete a project:

1. Set the project that you want to delete as the current project. (See section 2.4.2 "Switching to Another Project", p. 44.)
2. Click the `Delete` button.

2.4.7 Renaming a Project

Default project names (***project1***, ***project2***, ***projectN***) can be personalized to make it easier to keep track of the work you have performed in each project.

Note: This feature is not available for temporary projects.

To rename a project:

1. Type a new name in the text box labelled **This is *projectN***.

2. Click the **Rename** button.
   The page will be refreshed showing the new project name.

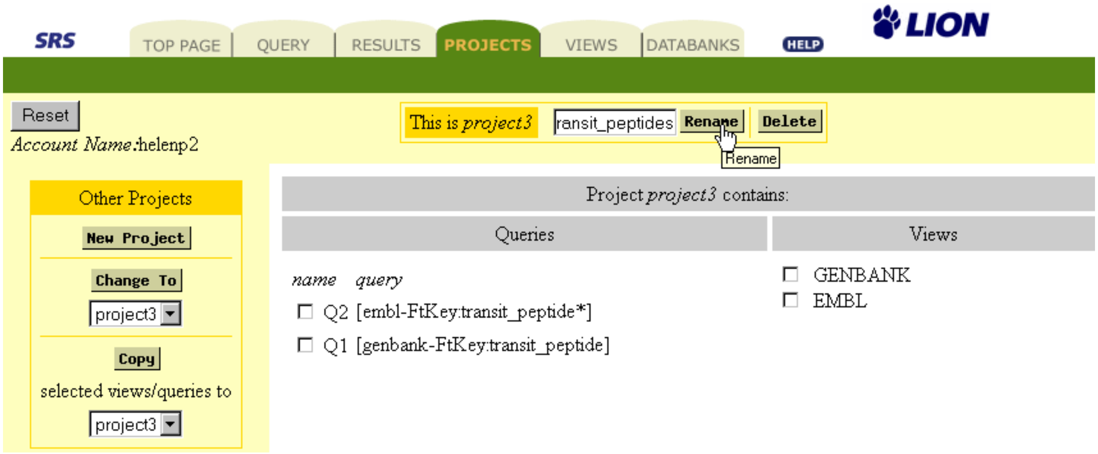


Figure 2.12 Personalized project names.

CHAPTER

3

QUERYING WITH SRS

One of the greatest advantages of using SRS is its ability to search databanks, and the range of methods available to do so. This chapter introduces you to the ways in which you can perform a query using SRS.

By the end of this chapter you will have learned:

- What a search term is and how to construct one.
- How to perform a **Quick Search**.
- How to use the **Standard Query Form** and **Extended Query Form**.
- What subentries are and the advantages of using them in queries.
- How to search for entries that relate to a specific expression or phrase (**Expression Query**).
- How to sort your query results.
- How to browse the index.

3.1 Search Terms

Whichever search method you use, you will need to use some sort of search term. Search terms used in SRS can be categorized as follows:

- Single-word search (see section 3.1.1 “Single Word Searches”, p. 52).
- Multiple-word phrases (see section 3.1.2 “Multiple-Word Searches”, p. 53).
- Numbers and dates (see section 3.1.3 “Numbers and Dates”, p. 53).
- Regular expressions (see section 3.1.4 “Regular Expressions”, p. 55).
- Wildcards (see section 3.1.5 “Wildcards”, p. 55).

The rest of this section explains how search terms are constructed. The individual search methods are detailed later in this chapter (see section 3.2 “Quick Search”, p. 55, section 3.3 “Query Forms”, p. 57, section 3.5 “Expression Queries”, p. 73 and section 3.7 “Browse Index”, p. 80).

3.1.1 Single Word Searches

When you search a databank for a single word, such as “reductase”, in a single field, the results of your query will be a list of entries that included that word in the index for the selected datafield.

3.1.2 Multiple-Word Searches

You can search for a phrase having more than one word, such as **aldehyde reductase**, in several ways. For example, if the phrase is enclosed in quotation marks, SRS will search for the complete string, only returning exact matches. If the phrase is not enclosed in quotation marks (**“aldehyde reductase”**), SRS will search for each word separately — combining the results in some fashion. The default is to require an entry to contain each of the words in the phrase so that a search for **aldehyde reductase** (without quotation marks) would find entries that have either **aldehyde** and **reductase** or the complete string **aldehyde reductase**.

You can make the relationship between the words explicit by including an operator in the string (see section 8.3.4 “Operators”, p. 156). For example, you could search for **aldehyde & reductase** (AND), or **aldehyde | reductase** (OR), or **aldehyde ! reductase** (BUTNOT).

3.1.3 Numbers and Dates

SRS uses numerical entries to allow you to search for information such as dates, or entries with a specified length.

Numerical entries can be combined into expressions using various operators. This allows you to search within particular ranges. Typically these operators are: less than; less than or equal; greater than; and greater than or equal. These are created by combining two simpler operators, namely the colon, `:`, and the exclamation mark, `!`. The colon means greater than or less than depending on which side of your number it lies. The exclamation mark indicates that the number to the right of the exclamation mark is to be excluded. The exclamation mark can be regarded as “not” or “not equal to”. It is probably easier to demonstrate this with some examples:

12:15

Greater than or equal to 12, but less than or equal to 15.

12:

Greater than or equal to 12, with no specific upper limit.

!12:

Greater than but not equal to 12, with no specific upper limit.

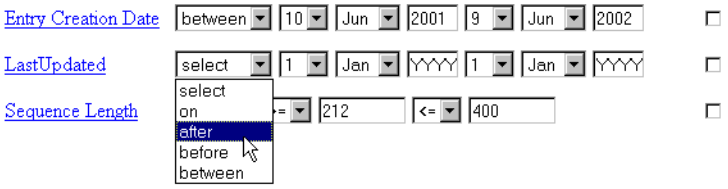
:12

Less than or equal to 12, with no specific lower limit.

:!12

Less than but not equal to 12, with no specific lower limit.

When using the **Extended Query Form** (see section 3.3.3 “Using the Extended Query Form”, p. 60) such a range search is often simplified for you, so that you do not need to use the query language syntax (see Figure 3.1). Using this method allows you to select the operators from a drop-down list.



Entry Creation Date	between	10	Jun	2001	9	Jun	2002	☐
LastUpdated	select	1	Jan	YYYY	1	Jan	YYYY	☐
Sequence Length	=	212	<=	400				☐

Figure 3.1 Extended Query Form showing a numerical data query being set up, using the drop-down lists.

See chapter 8, “SRS Query Language”, (p. 145) for more details on the syntax for numerical searches.

3.1.4 Regular Expressions

Regular expression searches are useful when you wish to search for alternative spellings, or words with the same root, but different suffixes, etc. Regular expressions allow you to use a combination of characters, along with regular expression characters, and get a list of matching entries as your result. You always need to include the forward slash (/) character at the start and end of the regular expression string. For example, “/^phos/” will find all words beginning with “phos” (e.g., phosphate, phosphorylase), “/ase$/” will find words ending in “ase” (e.g., kinase, phosphatase). (See section “Regular Expressions” (p. 149) for an explanation.)

3.1.5 Wildcards

The SRS query language also uses the familiar “*” and “?” wildcards. This is usually much simpler than using regular expressions for basic searches. For example, “cell*ase” would find all words starting with “cell” and ending in “ase” (e.g., cellobiase, cellobiohydrolase, cellulase).

3.2 Quick Search

3.2.1 About Quick Search

Quick Search works by searching all datafields of type **text** in the selected databanks, and is the quickest way to generate query results.

If you do not get the results you expect, then a slight change to the search phrase may help to target your search and thus may yield improved results. Alternatively, try one of the other search methods, as these allow you to specify your area of interest in more detail.

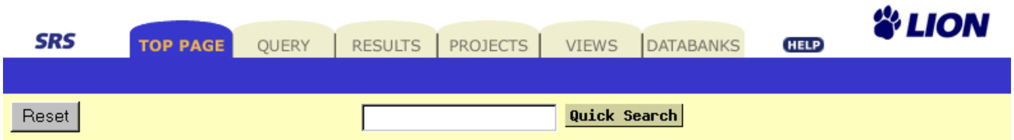


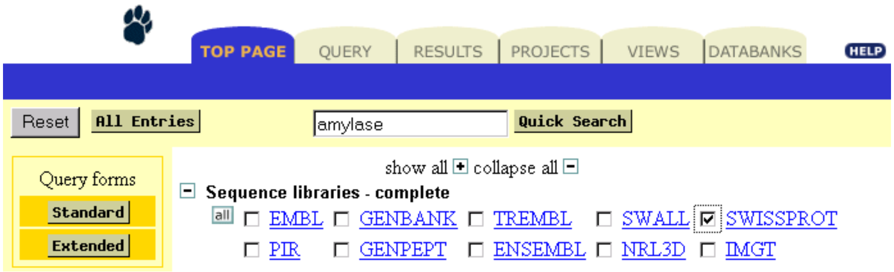
Figure 3.2 The location of **Quick Search** on the **Select Databanks To Search** page.

Note: The **Quick Search** option combines all the fields with a data type of **Text** using the **OR** operator.

3.2.2 Using Quick Search

Quick Search allows you to search with a minimum number of steps from selecting the databanks to viewing the results.

1. From the **Select Databanks To Search** page (**Top** page), select the databank(s) you wish to search.
   Tick the check box to the left of a suitable databank name.
2. Enter the search term in the text box beside the `Quick Search` button.
   Use a suitable word or expression (see section 3.1 "Search Terms", p. 52).
3. Click the `Quick Search` button.



amylase

Sequence libraries - complete

☐ EMBL ☐ GENBANK ☐ TREMBL ☐ SWALL ☑ SWISSPROT
☐ PIR ☐ GENPEPT ☐ ENSEMBL ☐ NRL3D ☐ IMGT

Figure 3.3 Quick Search of the SWISS-PROT databank for “amylase”.

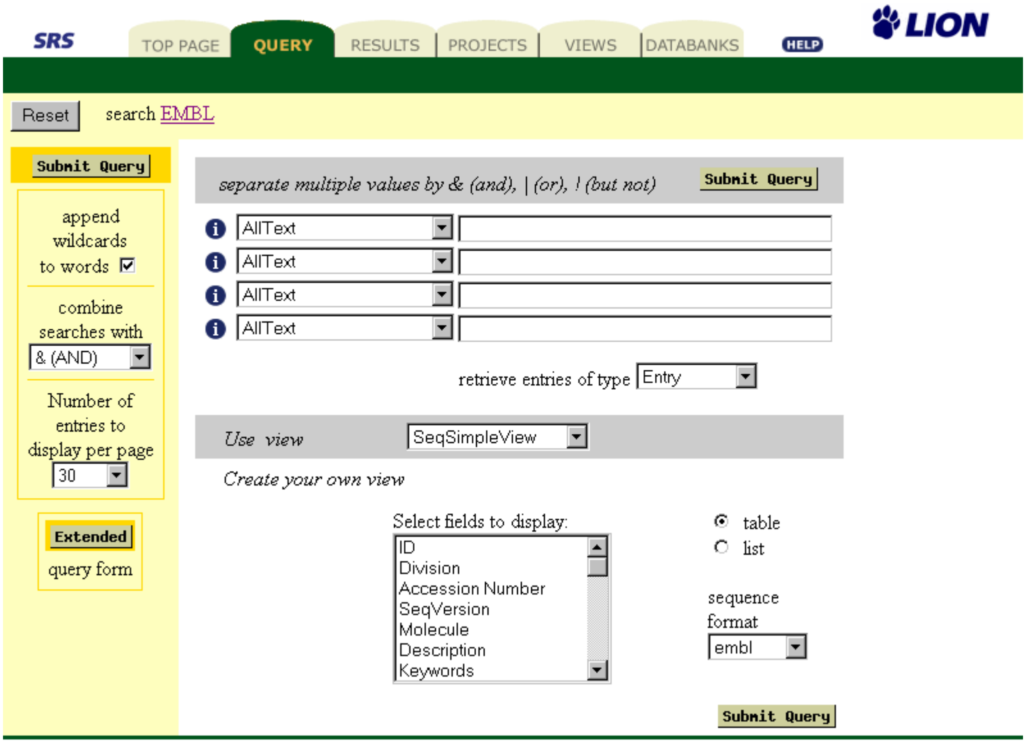
3.3 Query Forms

3.3.1 About Query Forms

Results of queries can usually be refined by adding to the information used for the search. If you are able to supply a larger range of information about your area of interest, then SRS can target your query more precisely. The query forms allow you to enter specified information about your subject, in various fields. There are two types of query form: the **Standard Query Form** and **Extended Query Form**. These are described in the remainder of this section.

3.3.2 Using the Standard Query Form

The **Standard Query Form** allows you to enter up to four separate search terms and search against up to four different datafields simultaneously.



search EMBL

append wildcards to words ☑

combine searches with & (AND)

Number of entries to display per page 30

separate multiple values by & (and), | (or), ! (but not)

AllText

AllText

AllText

AllText

retrieve entries of type Entry

Use view SeqSimpleView

Create your own view

Select fields to display:
ID
Division
Accession Number
SeqVersion
Molecule
Description
Keywords

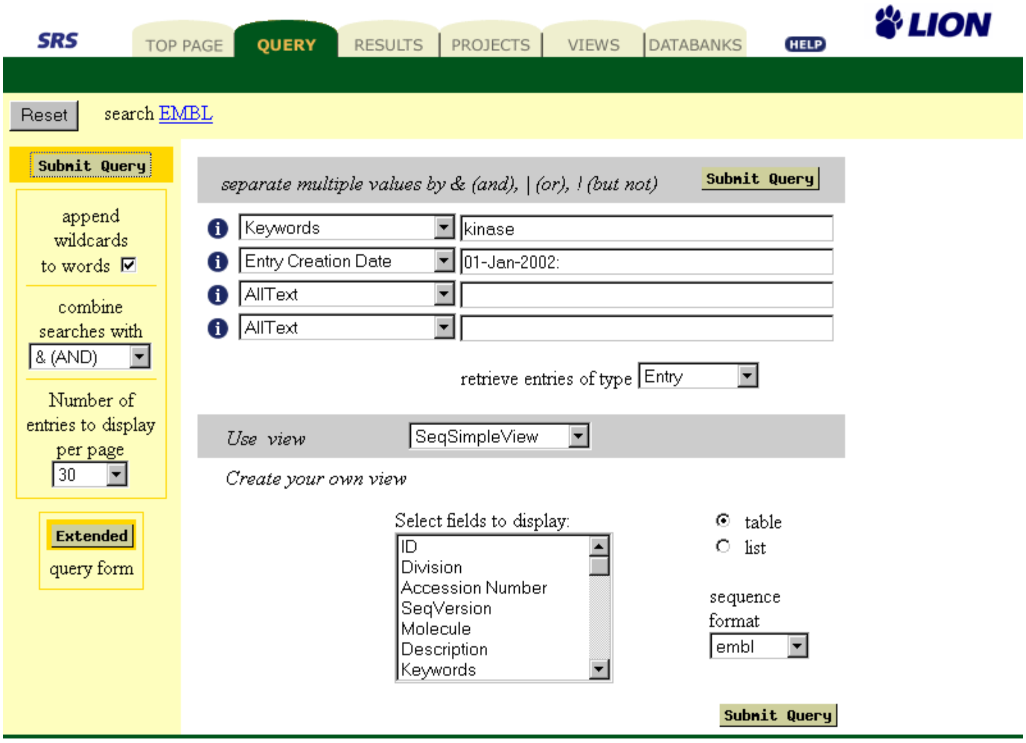
● table
○ list

sequence format embl

Figure 3.4 Standard Query Form.

1. From the **Select Databanks To Search** page (**Top** page), select the databank you want to search, by ticking the check box to the left of a databank name.
2. Click the **Standard** button from the **Query forms** box.
   This will display the **Standard Query Form**.
3. Enter the search phrase(s). See section 3.1 “Search Terms” (p. 52).
   In each of the four search rows, specify the field to be searched using the left-hand drop-down list and enter the search word or phrase in the corresponding right-hand text box. You do not have to use all four rows, but greater detail may help refine your search.

4. Set **combine searches with** to the required combination method (**AND**, **OR** or **BUTNOT**).
5. Specify the **Number of entries to display per page**.
6. Choose the type of entries you wish to retrieve, using the **retrieve entries of type** drop-down list.
   Leave it set at **Entry** for this example.
7. Choose a view using the **Use view** drop-down list. (See chapter 5, “Views”, p. 103, for information on defining a view.)
8. Click the `Submit Query` button.



search EMBL

append wildcards to words ☑

combine searches with & (AND)

Number of entries to display per page 30

separate multiple values by & (and), | (or), ! (but not)

Keywords: kinase

Entry Creation Date: 01-Jan-2002:

AllText:

AllText:

retrieve entries of type Entry

Use view SeqSimpleView

Create your own view

Select fields to display: ID, Division, Accession Number, SeqVersion, Molecule, Description, Keywords

● table
○ list

sequence format embl

Figure 3.5 Standard Query Form showing a query for EMBL for entries with the keyword, “kinase”, that were added to the databank after January 1, 2000.

Note: If you want to know more about any of the fields, choose that field from the drop-down list, and click the ℹ beside it.

3.3.3 Using the Extended Query Form

The **Extended Query Form** lists all the common datafields and allows you to enter search terms for as many of these fields as you want.

search EMBL

append wildcards to words ☑

combine searches with & (AND)

Number of entries to display per page 30

Standard query form

Make default query page ☐

separate multiple values by & (and), | (or), ! (but not)

Use view SeqSimpleView sequence format embl

Field Name	Query	Include in table view
AllText		
ID		☐
Division	☐ est ☐ fun ☐ gss ☐ htc ☐ htg ☐ hum ☐ inv ☐ mam ☐ mus ☐ org ☐ phg ☐ pln ☐ pro ☐ rod ☐ sts ☐ syn ☐ unc ☐ vrl ☐ vrt	☐
Accession Number		☐
SeqVersion	>= <=	☐

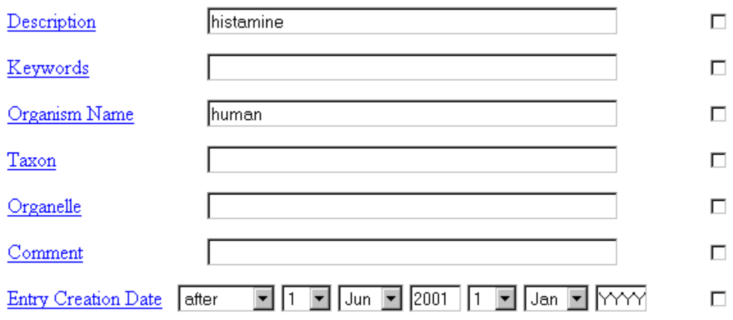
Figure 3.6 Extended Query Form.

1. From the **Select Databanks To Search** page (use the **Top Page** tab), select the databank you wish to search by ticking the check box to the left of the databank name.
2. Select **Extended** button from the **Query forms** box. This will display the **Extended Query Form**.
3. Enter the search phrase(s) (see section 3.1 “Search Terms”, p. 52) in the text boxes beside the relevant fields.

Note: You do not have to complete all the fields, but the greater the detail, the more refined your search will be.

---

4. Set **combine searches with** to the required combination method (**AND**, **OR** or **BUTNOT**).
5. Specify the **Number of entries to display per page**.
6. Choose a view using the **Use view** drop-down list. (See chapter 5, “Views”, p. 103, for information on defining a view.)
7. Click the **Submit Query** button.



Description	histamine	☐
Keywords		☐
Organism Name	human	☐
Taxon		☐
Organelle		☐
Comment		☐
Entry Creation Date	after ▾ 1 ▾ Jun ▾ 2001 1 ▾ Jan ▾ YYYY	☐

Figure 3.7 Extended Query Form querying EMBL for entries about “histamine” in humans, that were added to the databank after 1st June, 2001. The search uses the **Description**, **Organism** and **Entry Creation Date** fields.

3.4 Subentries

3.4.1 About Subentries

Subentries are parts of data entries that have an internal structure, which needs to be conserved when searching a database. A typical example of this might be a databank whose entries contain multiple literature references, or one that contains features (e.g. within a sequence) in which the relationship between the features is important.

3.4.2 Use of Subentries

Subentries are particularly useful for queries where a relationship between items is important.

Suppose a user wants to search for references written jointly by authors Smith and Jones, but does not want to find papers written only by one of those authors.

In a case like this it becomes important that the user can search for data where Smith and Jones occur within the same literature reference, whilst excluding entries that only contain papers written by each author separately, with no papers authored jointly by Smith & Jones.

A search of entries, even where it is specified that both Smith and Jones must appear within an entry will reveal all entries containing references by Smith & Jones jointly, and those which contain separate papers by Smith and by Jones.

A search using subentries, on the other hand will allow the user to specify a relationship between the two authors, i.e. to specify that there must be at least one reference that is written jointly.

Using subentries allows each literature reference to become a specific subentry: in effect creating a mini-entry within an entry.

The next section will take you through a typical subentry query.

3.4.3 Querying Using Subentries

This worked example will take you through using subentries to make queries using both of the **Query Forms**.

Searching for Entries which Reference Papers that include those by Smith and by Jones using the Standard Query Form

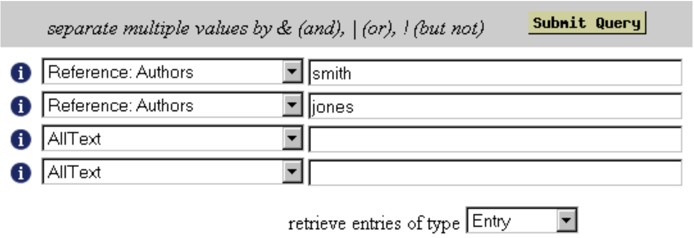
In this example entries will be retrieved that reference papers by Smith and Jones. For an entry to be retrieved there must be papers whose authors include Smith and Jones, but Smith and Jones need not be co-authors of the same paper.

1. From the **Select Databanks To Search** page (use the **Top Page** tab), select the databank you wish to search by ticking the check box to the left of the databank name, e.g. SWISS-PROT.
2. Select Standard button from the **Query forms** box. This will display the **Standard Query Form**.
3. Use the first drop-down datafield list to select an appropriate subentry field, e.g. **Reference:Authors**.

Note: Datafields for subentry fields are shown in the drop-down lists using a special format to identify them. In many cases, they are the same as for SWISS-PROT, i.e. a subentry name (e.g. **Reference**) followed by a field within that subentry (e.g. **Authors**); the two parts are separated by a colon (**:**). Hence **Reference:Authors** will search for authors fields within subentries.

---

4. Enter **smith** in the first text box, beside where you selected **Reference:Authors**. The search is case insensitive, so it does not matter whether smith is written with a lower or upper case S.
5. Repeat the process for the next line of information, using the name **jones** rather than smith.



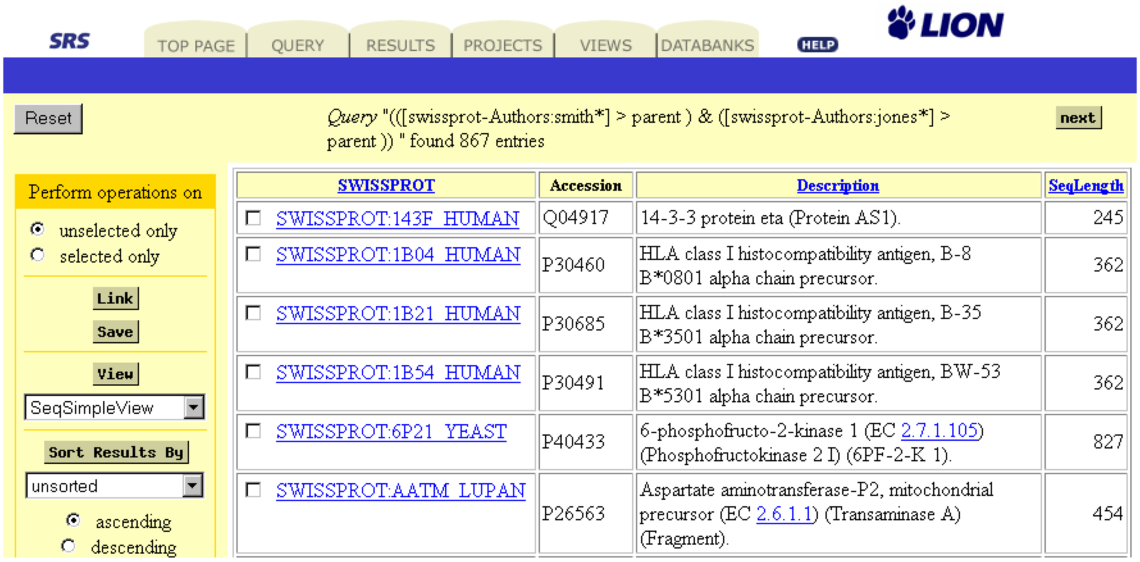
separate multiple values by & (and), | (or), ! (but not)

Reference: Authors smith
Reference: Authors jones
AllText
AllText

retrieve entries of type Entry

Figure 3.8 Standard Query Form showing a query for entries, using the **References:Authors** subentry fields.

6. Leave **combine searches with** set to **& (AND)**.
7. Leave **retrieve entries of type** set to **Entry**. This will cause entire entries to be retrieved. Setting this option to one of the other (subentry) fields, e.g. **Reference** will tell SRS to retrieve each reference (which fulfils the search criteria) as a separate entity.
8. Click the **Submit Query** button.



Query "(([swissprot-Authors:smith*] > parent) & ([swissprot-Authors:jones*] > parent)) " found 867 entries

Perform operations on
● unselected only
○ selected only

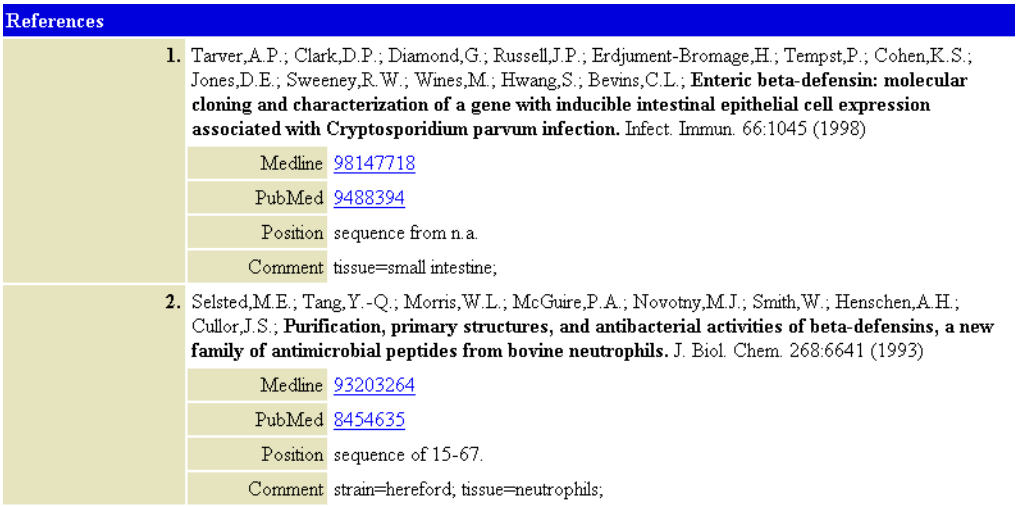
View: SeqSimpleView

Sort Results By: unsorted
● ascending
○ descending

SWISSPROT	Accession	Description	SeqLength
☐ SWISSPROT:143F_HUMAN	Q04917	14-3-3 protein eta (Protein AS1).	245
☐ SWISSPROT:1B04_HUMAN	P30460	HLA class I histocompatibility antigen, B-8 B*0801 alpha chain precursor.	362
☐ SWISSPROT:1B21_HUMAN	P30685	HLA class I histocompatibility antigen, B-35 B*3501 alpha chain precursor.	362
☐ SWISSPROT:1B54_HUMAN	P30491	HLA class I histocompatibility antigen, BW-53 B*5301 alpha chain precursor.	362
☐ SWISSPROT:6P21_YEAST	P40433	6-phosphofructo-2-kinase 1 (EC 2.7.1.105) (Phosphofructokinase 2 I) (6PF-2-K 1).	827
☐ SWISSPROT:AATM_LUPAN	P26563	Aspartate aminotransferase-P2, mitochondrial precursor (EC 2.6.1.1) (Transaminase A) (Fragment).	454

Figure 3.9 Query Result page showing the results of a query for entries, using the **References:Authors** subentry fields, using two separate text boxes to search on the **Standard Query Form**.

Using two separate text boxes for the subentries search retrieves all entries containing references by Smith, and by Jones, including those where there are no papers which were co-authored by Smith and Jones. Figure 3.10 shows an entry from the above search.



References

1. Tarver,A.P.; Clark,D.P.; Diamond,G.; Russell,J.P.; Erdjument-Bromage,H.; Tempst,P.; Cohen,K.S.; Jones,D.E.; Sweeney,R.W.; Wines,M.; Hwang,S.; Bevins,C.L.; **Enteric beta-defensin: molecular cloning and characterization of a gene with inducible intestinal epithelial cell expression associated with Cryptosporidium parvum infection.** Infect. Immun. 66:1045 (1998)

Medline	98147718
PubMed	9488394
Position	sequence from n.a.
Comment	tissue=small intestine;

2. Selsted,M.E.; Tang,Y.-Q.; Morris,W.L.; McGuire,P.A.; Novotny,M.J.; Smith,W.; Henschen,A.H.; Cullor,J.S.; **Purification, primary structures, and antibacterial activities of beta-defensins, a new family of antimicrobial peptides from bovine neutrophils.** J. Biol. Chem. 268:6641 (1993)

Medline	93203264
PubMed	8454635
Position	sequence of 15-67.
Comment	strain=hereford; tissue=neutrophils;

Figure 3.10 Entry page showing that some entries reference only papers that are not co-authored by Smith & Jones. (SWISS-PROT accession number P46161.)

In effect, this search looks for reference subentries that contain smith or jones. It then takes the entries for each of the references matching the criteria, to create two lists of entries: one for smith and one for jones. It then combines those, seeking only those which are in both lists.

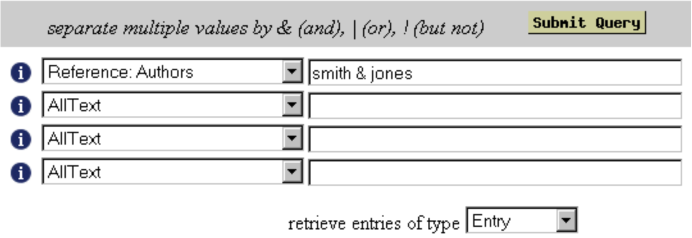
Searching for Entries which Reference Papers that are Co-authored by Smith & Jones using the Standard Query Form

In the above example, entries were retrieved that referenced papers by Smith and Jones, but without concern for whether Smith & Jones were co-authors of any papers.

If you wish to search for entries that reference papers that are always co-authored by Smith & Jones, you should modify your search, so that rather than searching in two separate text boxes on the **Standard Query Form**, you search for both authors in a single text box.

1. From the **Select Databanks To Search** page, select a databank to search, e.g. SWISS-PROT and go to the **Standard Query Form**.
2. Use the first drop-down datafield list to select an appropriate subentry field, e.g. **Reference:Authors**.
3. In the text box beside the datafield you have used, type **smith & jones**.

Note: Make sure you use the & sign which tells SRS that you want to look for both authors combined using the boolean operator **&** (AND), rather than | (OR), or ! (NOT).



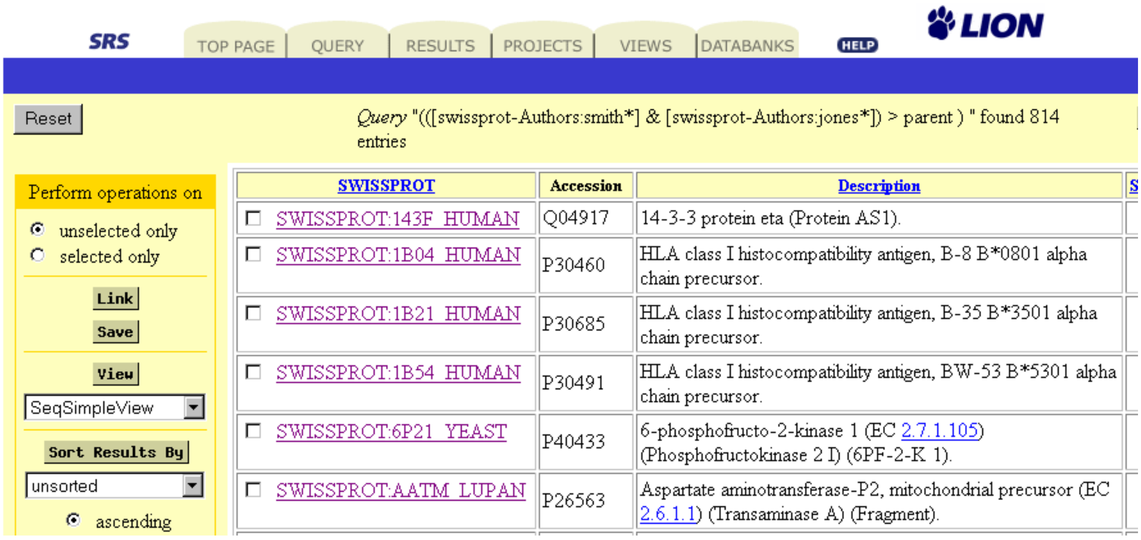
separate multiple values by & (and), | (or), ! (but not)

Reference: Authors	smith & jones
AllText	
AllText	
AllText	

retrieve entries of type Entry

Figure 3.11 Standard Query Form showing a query for entries, using the **References:Authors** subentry fields, searching for papers which are co-authored by Smith & Jones.

4. Leave **combine searches with** set to **& (AND)**, and **retrieve entries of type** set to **Entry**.
5. Click the **Submit Query** button.



Query "(([swissprot-Authors:smith*] & [swissprot-Authors:jones*]) > parent) " found 814 entries

SWISSPROT	Accession	Description	S
☐ SWISSPROT:143F_HUMAN	Q04917	14-3-3 protein eta (Protein AS1).	
☐ SWISSPROT:1B04_HUMAN	P30460	HLA class I histocompatibility antigen, B-8 B*0801 alpha chain precursor.	
☐ SWISSPROT:1B21_HUMAN	P30685	HLA class I histocompatibility antigen, B-35 B*3501 alpha chain precursor.	
☐ SWISSPROT:1B54_HUMAN	P30491	HLA class I histocompatibility antigen, BW-53 B*5301 alpha chain precursor.	
☐ SWISSPROT:6P21_YEAST	P40433	6-phosphofructo-2-kinase 1 (EC 2.7.1.105) (Phosphofructokinase 2 I) (6PF-2-K 1).	
☐ SWISSPROT:AATM_LUPAN	P26563	Aspartate aminotransferase-P2, mitochondrial precursor (EC 2.6.1.1) (Transaminase A) (Fragment).	

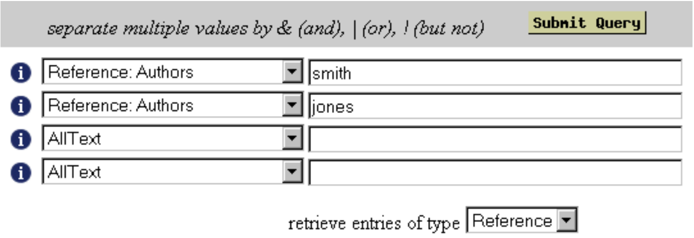
Figure 3.12 Query Result page showing the results of a query for entries, using the **References:Authors** subentry fields, using a single text box with two search terms, to search on the **Standard Query Form**.

In effect, this search looks for reference subentries that contain smith or jones. It then combines these to create a list of those that contain both Smith and Jones in a single reference subentry. It then takes the entries for each of the references matching the criteria. This creates a single list of entries, which contain references to papers which are co-authored by Smith and Jones.

Searching for Subentries which Reference Papers that are Co-authored by Smith & Jones using the Standard Query Form

If you wish to retrieve the subentry fields that are found, rather than the parent entries, the searches can be repeated with the **retrieve entries of type** set to **Reference**.

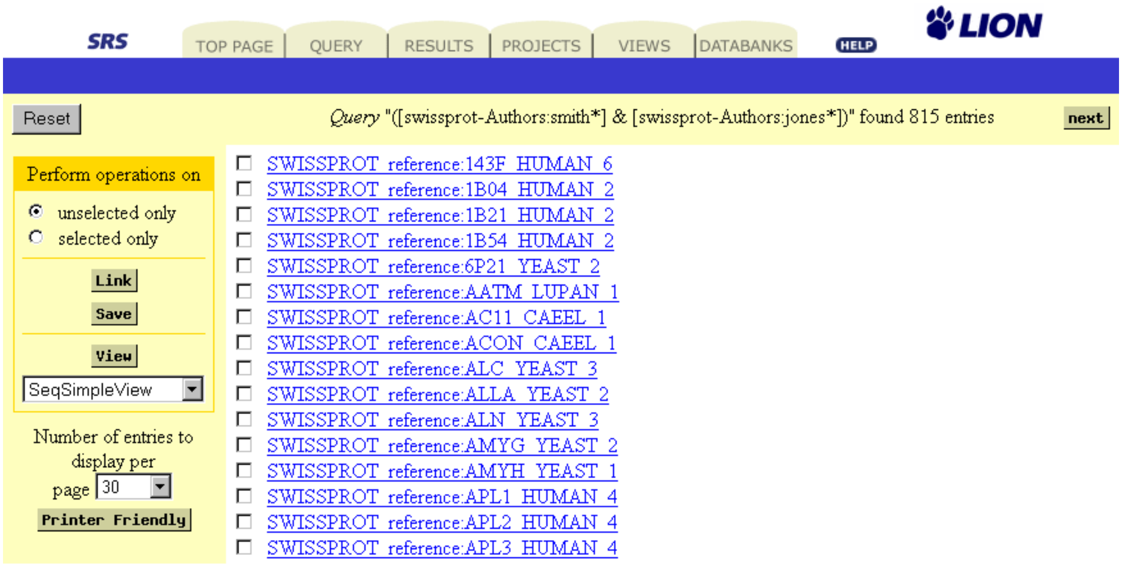
1. Repeat one of the above searches, but set **retrieve entries of typ**e to **Reference**.



Reference: Authors smith
Reference: Authors jones
AllText
AllText
retrieve entries of type Reference

Figure 3.13 Standard Query Form showing a query for subentries, using the **References:Authors** fields.

This search will retrieve the subentries that match the criteria rather than the parent entries. Otherwise, the search behaves similarly to the above, so that searching for the two terms in a single text box, combined with & will produce reference subentries that contain only co-authored papers, whilst searching for the two terms using separate text boxes will produce reference subentries which contain papers by both authors, but will not check whether they are co-authors.



Query "([swissprot-Authors:smith*] & [swissprot-Authors:jones*])" found 815 entries

- ☐ SWISSPROT_reference:143F_HUMAN_6
- ☐ SWISSPROT_reference:1B04_HUMAN_2
- ☐ SWISSPROT_reference:1B21_HUMAN_2
- ☐ SWISSPROT_reference:1B54_HUMAN_2
- ☐ SWISSPROT_reference:6P21_YEAST_2
- ☐ SWISSPROT_reference:AATM_LUPAN_1
- ☐ SWISSPROT_reference:AC11_CAEEL_1
- ☐ SWISSPROT_reference:ACON_CAEEL_1
- ☐ SWISSPROT_reference:ALC_YEAST_3
- ☐ SWISSPROT_reference:ALLA_YEAST_2
- ☐ SWISSPROT_reference:ALN_YEAST_3
- ☐ SWISSPROT_reference:AMYG_YEAST_2
- ☐ SWISSPROT_reference:AMYH_YEAST_1
- ☐ SWISSPROT_reference:APL1_HUMAN_4
- ☐ SWISSPROT_reference:APL2_HUMAN_4
- ☐ SWISSPROT_reference:APL3_HUMAN_4

Figure 3.14 Query Result page showing the results of a query for subentries, using the **References:Authors** subentry fields, using the **Standard Query Form**.

Searching for Subentries which Reference Papers that are Co-authored by Smith & Jones using the Extended Query Form

Querying subentry fields on the **Extended Query Form** is similar to that on the **Standard Query Form**. However, because the **Extended Query Form** does not allow you to search the same field using two separate text boxes, it is not possible to repeat the type of search done in "Searching for Entries which Reference Papers that include those by Smith and by Jones using the Standard Query Form" (p. 63).

1. From the **Select Databanks To Search** page, select a databank to search, e.g. SWISS-PROT and go to the **Extended Query Form**.

2. Scroll down the **Extended Query Form** until you find the section marked, **fields of subentry: Reference**. Enter **smith & jones** into the **Authors** text box.

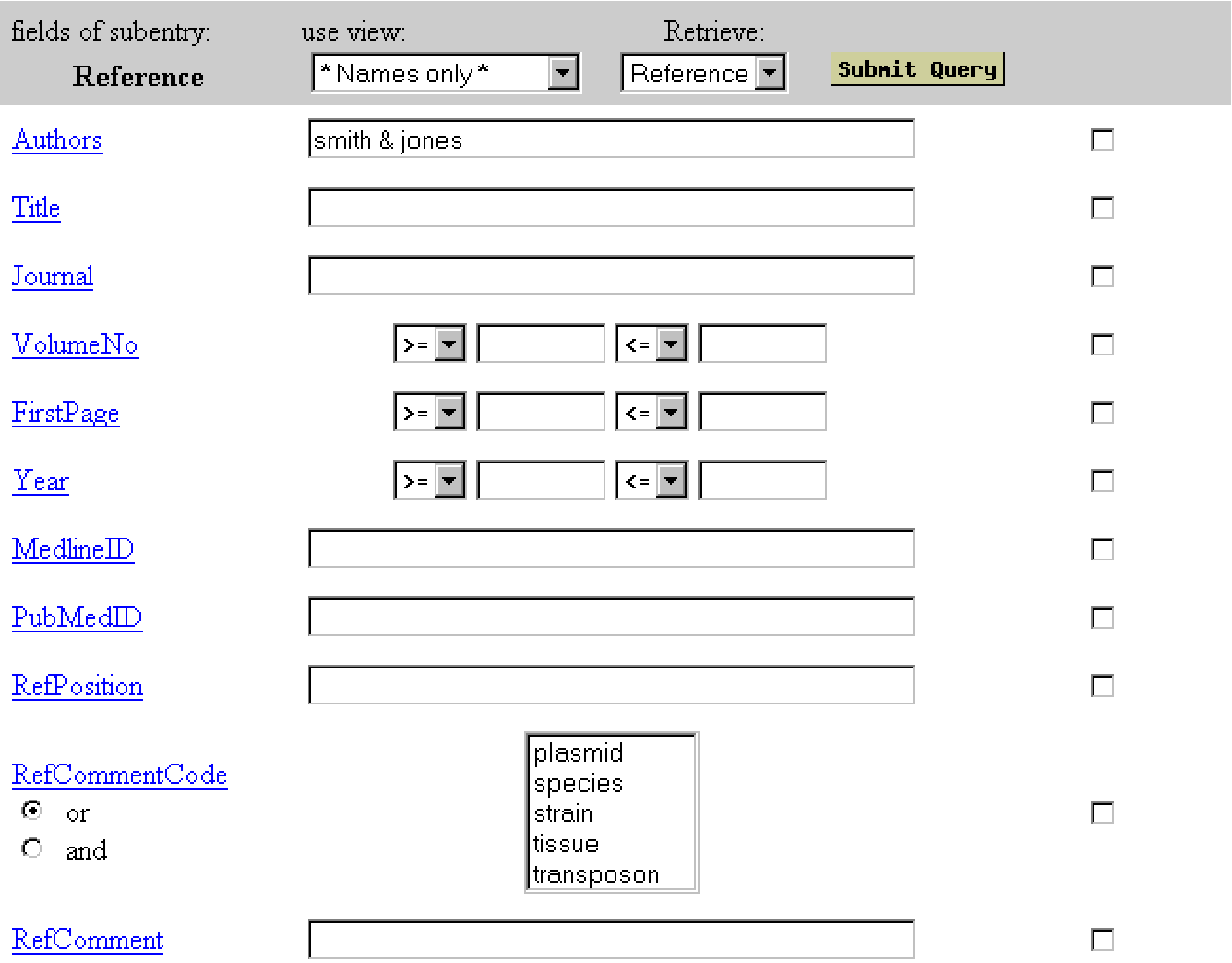
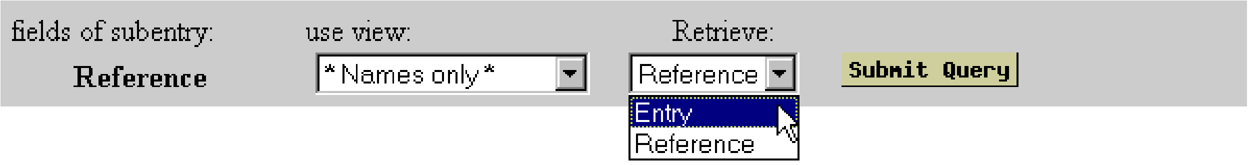

fields of subentry: Reference
use view: * Names only *
Retrieve: Reference
Authors: smith & jones
RefCommentCode: ● or ○ and

Figure 3.15 Part of the **Extended Query Form** showing a query for subentries, using the **References:Authors** fields.

3. Choose whether you wish to retrieve entire entries, or the **Reference** subentries.



fields of subentry: Reference

use view: * Names only *

Retrieve: Reference

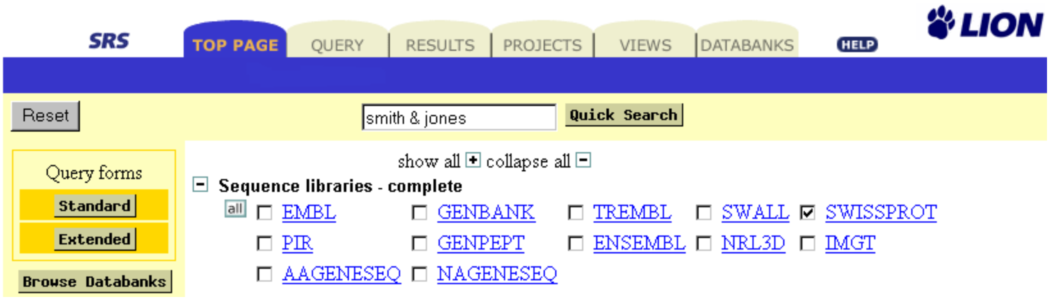
Figure 3.16 Choosing whether to retrieve entire entries or the **Reference** subentries.

4. Submit the query as usual.

Searching for Subentries which Reference Papers that are Co-authored by Smith & Jones using Quick Search

The quick search option will always return complete entries.You can, nevertheless, search for entries using the subentries as search items, however, as Quick Search simply searches all the text fields in an entry, you cannot control the field within which the search text occurs.

1. From the **Select Databanks To Search** page, select a databank to search, e.g. SWISS-PROT.
2. Enter **smith & jones** in the **Quick Search** text box.



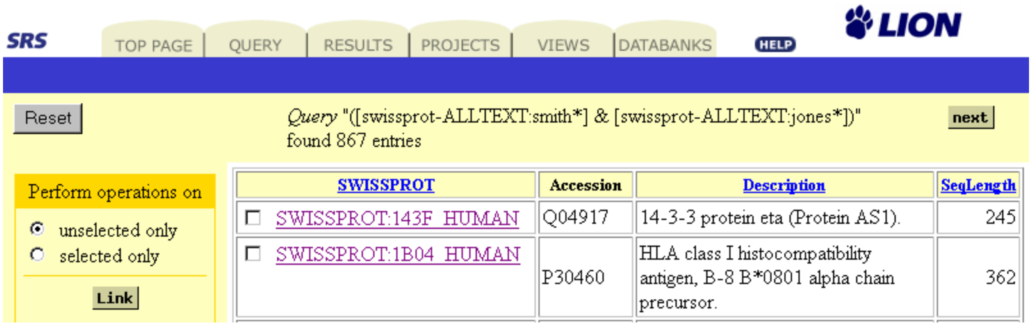
Quick Search: smith & jones

Sequence libraries - complete

☐ EMBL ☐ GENBANK ☐ TREMBL ☐ SWALL ☑ SWISSPROT
☐ PIR ☐ GENPEPT ☐ ENSEMBL ☐ NRL3D ☐ IMGT
☐ AAGENESEQ ☐ NAGENESEQ

Figure 3.17 **Quick Search** of SWISS-PROT for smith & jones.

In this case, you cannot tell SRS that you want only to search the subentries, and you have no way of indicating that you wish to search only for papers that are co-authored by Smith & Jones. The results will be entries which contain papers by both authors, without any control over whether the papers are co-authored. The same results could be achieved using the Alltext fields in either the **Standard** or **Extended Query Forms**.



Query "([swissprot-ALLTEXT:smith*] & [swissprot-ALLTEXT:jones*])"
found 867 entries

	SWISSPROT	Accession	Description	SeqLength
☐	SWISSPROT:143F_HUMAN	Q04917	14-3-3 protein eta (Protein AS1).	245
☐	SWISSPROT:1B04_HUMAN	P30460	HLA class I histocompatibility antigen, B-8 B*0801 alpha chain precursor.	362

Figure 3.18 Quick Search of SWISS-PROT for smith & jones.

3.5 Expression Queries

3.5.1 About Expression Queries

You can make queries using the text box beside the **Expression** button (Figure 3.19) which you will find on the **Manage your Query Results** page. You can also use this method to combine, link, or refine the results of existing queries.

Q1 < Q2

Figure 3.19 Expression button and text box.

3.5.2 Using Expression Queries

1. On the **Manage your Query Results** page, enter your query into the text box to the right of the Expression button.
2. Click on the Expression button.

For example, to search for all entries that satisfy two existing queries, **Q1** and **Q2**, you could type:

Q1 & Q2

This will return a new listing, **Q3**, that has all entries that are in both **Q1** and **Q2**.

If you wanted to search for the entries in query, **Q3**, that had links to the SWISS-PROT databank, you would type:

Q3 < swissprot

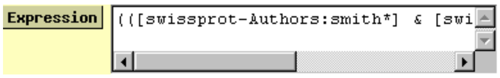
Typical operators allow you to combine searches using the standard logical functions (AND, OR, BUT NOT), and to look for links between result sets. For more help on linking using expressions, see section 4.5 “Expression Linking” (p. 98). See Table 8.5, page 156 for a list of the SRS query language operators that may be used.

This method of searching is very powerful because many detailed queries can be created using the SRS Query Language. (See chapter 8, “SRS Query Language”, p. 145 for more information.)

Using Expression Queries to Search Subentries

You can use Expression Queries to search for subentries. This requires a knowledge of the correct SRS query language. However, until you are confident enough to create your own queries, you can find out the correct query syntax by looking back at the **Query Result** page that you have previously generated. For example, use the text generated in “Searching for Entries which Reference Papers that are Co-authored by Smith & Jones using the Standard Query Form” (p. 66).

1. From the **Select Databanks To Search** page, select a databank to search, e.g. SWISS-PROT.
2. Click the Results tab to display the Manage your Query Results page. Enter the following text in the **Expression Query** text box:
   (([swissprot-Authors:smith*] & [swissprot-Authors:jones*]) > parent)
   The text should all be on one line, and the text box will scroll as you type. You can scroll back to the left to check what you have written, but it is often difficult to see the whole query. You could try typing the query in a suitable text editor, so that you can see it, and then using copy and paste to insert it into the **Expression Query** text box.



Expression (([swissprot-Authors:smith*] & [swi

3.6 Sorting Results

3.6.1 What is Sorting?

Many queries will yield a large number of results. Sometimes, it is useful to be able to sort these to help you select the most relevant results. You could sort by almost any of the available fields. However, the fields which are available for sorting will be limited to those selected for the installation of SRS that you are using. In addition, a few fields are not appropriate for sorting. Some typical sortable fields are shown in Figure 3.20.

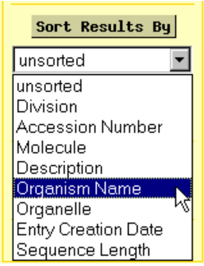
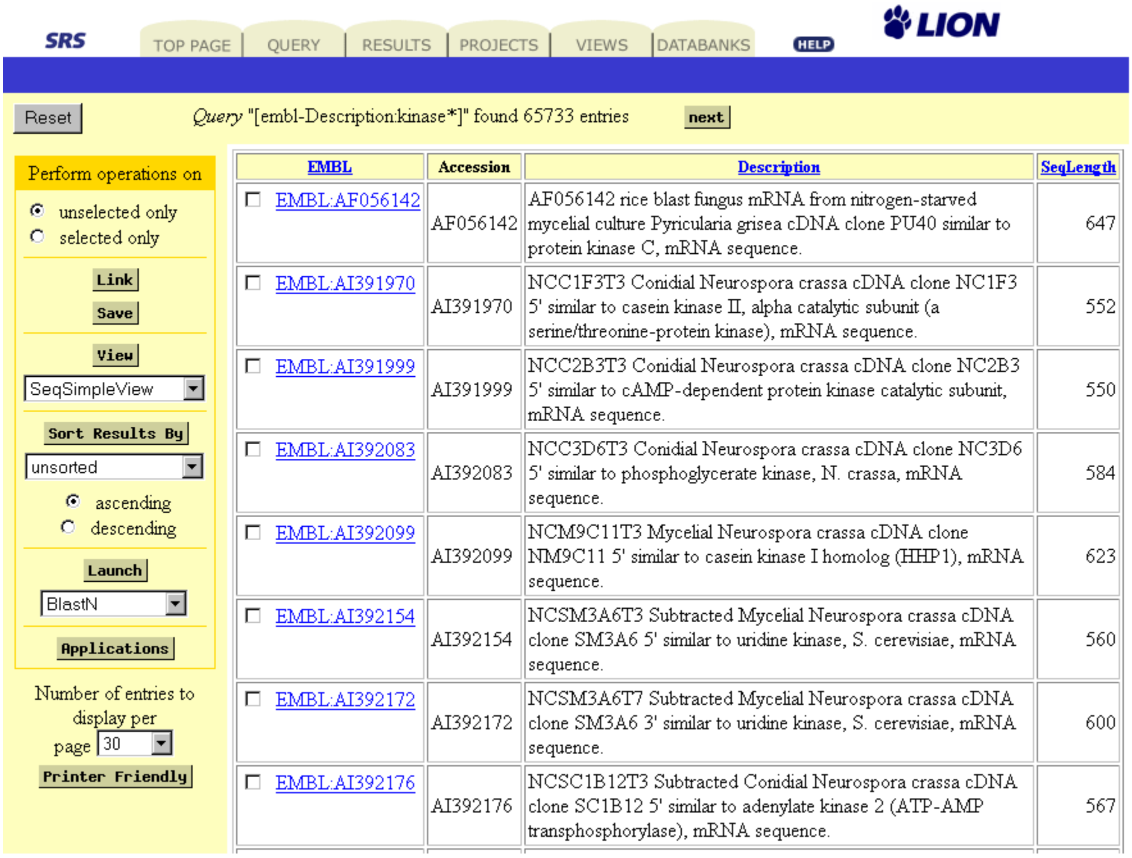


Figure 3.20 Drop-down list, showing sortable fields on a **Query Result** page.

3.6.2 Sorting a Set of Results

1. First choose a set of results, e.g. by making a search, and view them on a **Query Result** page.



Query "[embl-Description:kinase*]" found 65733 entries

Perform operations on
● unselected only
○ selected only

View: SeqSimpleView

Sort Results By: unsorted
● ascending
○ descending

Launch: BlastN

Number of entries to display per page 30

	EMBL	Accession	Description	SeqLength
☐	EMBL:AF056142	AF056142	AF056142 rice blast fungus mRNA from nitrogen-starved mycelial culture Pyricularia grisea cDNA clone PU40 similar to protein kinase C, mRNA sequence.	647
☐	EMBL:AI391970	AI391970	NCC1F3T3 Conidial Neurospora crassa cDNA clone NC1F3 5' similar to casein kinase II, alpha catalytic subunit (a serine/threonine-protein kinase), mRNA sequence.	552
☐	EMBL:AI391999	AI391999	NCC2B3T3 Conidial Neurospora crassa cDNA clone NC2B3 5' similar to cAMP-dependent protein kinase catalytic subunit, mRNA sequence.	550
☐	EMBL:AI392083	AI392083	NCC3D6T3 Conidial Neurospora crassa cDNA clone NC3D6 5' similar to phosphoglycerate kinase, N. crassa, mRNA sequence.	584
☐	EMBL:AI392099	AI392099	NCM9C11T3 Mycelial Neurospora crassa cDNA clone NM9C11 5' similar to casein kinase I homolog (HHP1), mRNA sequence.	623
☐	EMBL:AI392154	AI392154	NCSM3A6T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 5' similar to uridine kinase, S. cerevisiae, mRNA sequence.	560
☐	EMBL:AI392172	AI392172	NCSM3A6T7 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 3' similar to uridine kinase, S. cerevisiae, mRNA sequence.	600
☐	EMBL:AI392176	AI392176	NCSC1B12T3 Subtracted Conidial Neurospora crassa cDNA clone SC1B12 5' similar to adenylate kinase 2 (ATP-AMP transphosphorylase), mRNA sequence.	567

Figure 3.21 Query Result page, showing the results from a search of the EMBL **Description** field, for the word **kinase**.

2. Choose the **Organism Name** field from the drop-down list below the `Sort Results By` button.

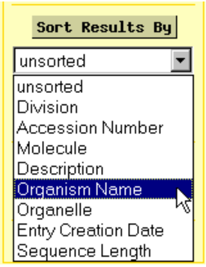
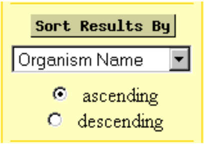


Figure 3.22 Drop-down list.

Note: Entries for which the chosen sort field is missing or null will be listed at the bottom of a sorted list whether sorting was in ascending or descending order.

---

3. For this example, leave **ascending** order selected.



Organism Name

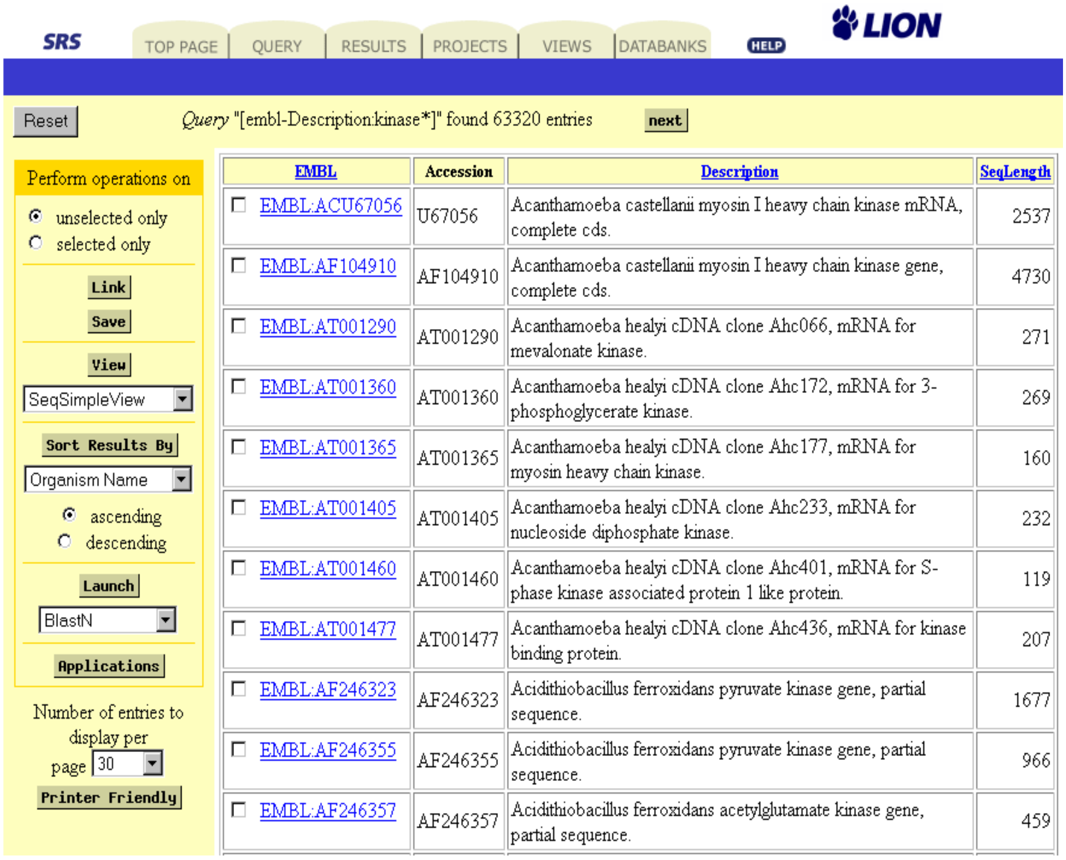
● ascending
○ descending

Figure 3.23 Choose order for sorting.

Note: Unsorted results will be listed in **ascending** order. It is not possible to reverse this order for unsorted results.

---

4. Initiate sorting by clicking on the `Sort Results By` button. The result of the above sort is shown in Figure 3.24.



Query "[embl-Description:kinase*]" found 63320 entries

Perform operations on
● unselected only
○ selected only

SeqSimpleView

Organism Name
● ascending
○ descending

BlastN

Number of entries to display per page 30

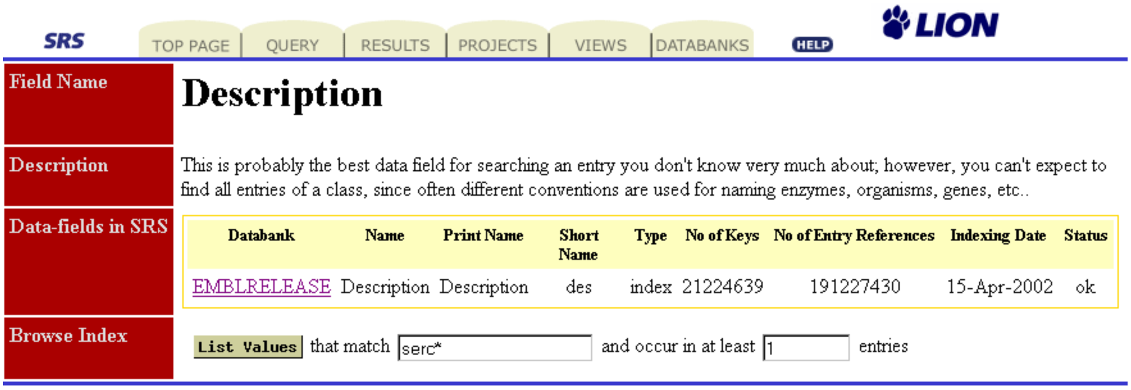
EMBL	Accession	Description	SeqLength
☐ EMBL:ACU67056	U67056	Acanthamoeba castellanii myosin I heavy chain kinase mRNA, complete cds.	2537
☐ EMBL:AF104910	AF104910	Acanthamoeba castellanii myosin I heavy chain kinase gene, complete cds.	4730
☐ EMBL:AT001290	AT001290	Acanthamoeba healyi cDNA clone Ahc066, mRNA for mevalonate kinase.	271
☐ EMBL:AT001360	AT001360	Acanthamoeba healyi cDNA clone Ahc172, mRNA for 3-phosphoglycerate kinase.	269
☐ EMBL:AT001365	AT001365	Acanthamoeba healyi cDNA clone Ahc177, mRNA for myosin heavy chain kinase.	160
☐ EMBL:AT001405	AT001405	Acanthamoeba healyi cDNA clone Ahc233, mRNA for nucleoside diphosphate kinase.	232
☐ EMBL:AT001460	AT001460	Acanthamoeba healyi cDNA clone Ahc401, mRNA for S-phase kinase associated protein 1 like protein.	119
☐ EMBL:AT001477	AT001477	Acanthamoeba healyi cDNA clone Ahc436, mRNA for kinase binding protein.	207
☐ EMBL:AF246323	AF246323	Acidithiobacillus ferroxidans pyruvate kinase gene, partial sequence.	1677
☐ EMBL:AF246355	AF246355	Acidithiobacillus ferroxidans pyruvate kinase gene, partial sequence.	966
☐ EMBL:AF246357	AF246357	Acidithiobacillus ferroxidans acetylglutamate kinase gene, partial sequence.	459

Figure 3.24 Query Result page, showing the results from a search of the EMBL **Description** field, for the word "kinase", sorted according to the **Organism Name** field.

3.7 Browse Index

3.7.1 About Browsing Indices

In addition to the other search methods, you can also browse the indices for a search term.



Field Name: Description

Description: This is probably the best data field for searching an entry you don't know very much about; however, you can't expect to find all entries of a class, since often different conventions are used for naming enzymes, organisms, genes, etc..

Data-fields in SRS:

Databank	Name	Print Name	Short Name	Type	No of Keys	No of Entry References	Indexing Date	Status
EMBLRELEASE	Description	Description	des	index	21224639	191227430	15-Apr-2002	ok

Browse Index: List Values that match serc* and occur in at least 1 entries

Figure 3.25 Field Information page for the **Description** field index.

3.7.2 Browsing Indices

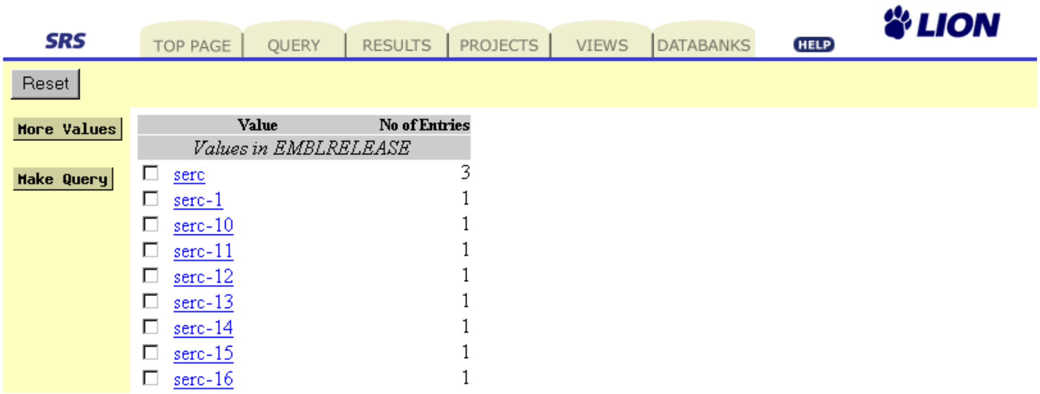
You can browse indices from the **Field Information** page for the datafield type (e.g., **Description**, **SeqLength**) that interests you. See section 3.7.3 “Getting to the Field Information Page” (p. 81) for more details of how to get to this page.

1. On the **Field Information** page for the appropriate databank and datafield, enter your search term, using wildcards as appropriate.

Note: In contrast to other query methods, implicit wildcards are not automatically appended to the search term when browsing indices, and must be specified explicitly.

---

2. Click the `List Values` button.
   This will take you to the **Browse Index** page.



	Value	No of Entries
	Values in EMBLRELEASE	
☐	serc	3
☐	serc-1	1
☐	serc-10	1
☐	serc-11	1
☐	serc-12	1
☐	serc-13	1
☐	serc-14	1
☐	serc-15	1
☐	serc-16	1

Figure 3.26 Browsing the **Description** field index on the **Browse Index** page.

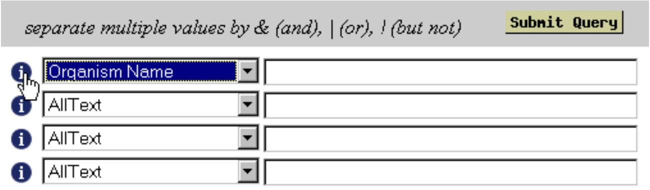
3. Tick the check boxes beside the resultant terms that best match your needs and click the `Make Query` button.
   This returns entries that match the selected terms.

3.7.3 Getting to the Field Information Page

There are a number of routes you can use to get to the **Field Information** page. The common routes are detailed here.

From the Query Forms

On the **Standard Query Form** there is an ⓘ icon beside each field box. Choose the field of interest from the drop-down list, and click on the ⓘ icon.

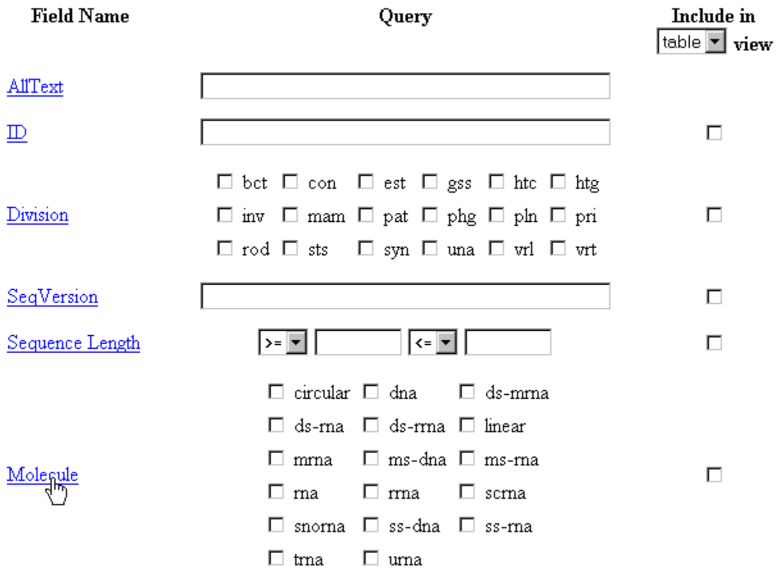


separate multiple values by & (and), | (or), ! (but not)

ⓘ Organism Name
ⓘ AllText
ⓘ AllText
ⓘ AllText

Figure 3.27 Standard Query Form showing use of the ⓘ icon to access the **Field Information** page.

On the **Extended Query Form** each of the datafields is a hyperlink to the **Field Information** page. Simply click on the hyperlink for the field of interest.

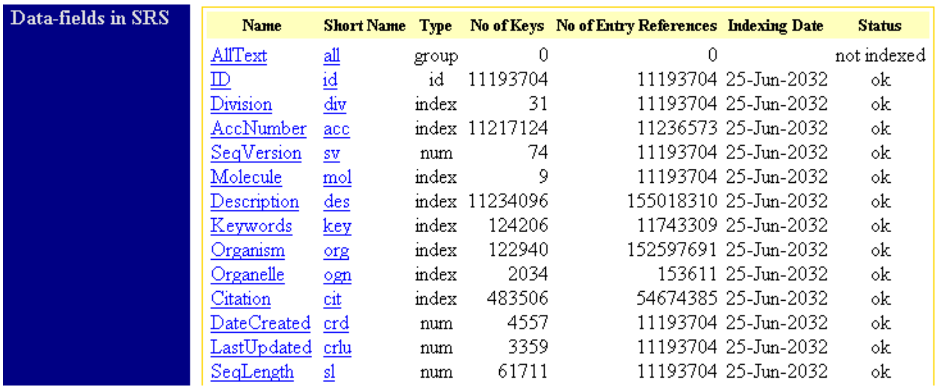


Field Name	Query	Include in table ▾ view
AllText		
ID		☐
Division	☐ bct ☐ con ☐ est ☐ gss ☐ htc ☐ htg ☐ inv ☐ mam ☐ pat ☐ phg ☐ pln ☐ pri ☐ rod ☐ sts ☐ syn ☐ una ☐ vrl ☐ vrt	☐
SeqVersion		☐
Sequence Length	>= ▾ <= ▾	☐
Molecule	☐ circular ☐ dna ☐ ds-mrna ☐ ds-rna ☐ ds-rrna ☐ linear ☐ mrna ☐ ms-dna ☐ ms-rna ☐ rna ☐ rrna ☐ scrna ☐ snorna ☐ ss-dna ☐ ss-rna ☐ trna ☐ urna	☐

Figure 3.28 Extended Query Form showing hyperlinks for datafields.

From the List Of Databanks Page

The **Databanks** tab takes you to the **List of Databanks** page. Here you will see a list of the available databanks. Clicking on the hyperlink for a databank will take you to its **Databank Information** page. This contains a section listing the datafields for the relevant databank.



Data-fields in SRS

Name	Short Name	Type	No of Keys	No of Entry References	Indexing Date	Status
AllText	all	group	0	0		not indexed
ID	id	id	11193704	11193704	25-Jun-2032	ok
Division	div	index	31	11193704	25-Jun-2032	ok
AccNumber	acc	index	11217124	11236573	25-Jun-2032	ok
SeqVersion	sv	num	74	11193704	25-Jun-2032	ok
Molecule	mol	index	9	11193704	25-Jun-2032	ok
Description	des	index	11234096	155018310	25-Jun-2032	ok
Keywords	key	index	124206	11743309	25-Jun-2032	ok
Organism	org	index	122940	152597691	25-Jun-2032	ok
Organelle	ogn	index	2034	153611	25-Jun-2032	ok
Citation	cit	index	483506	54674385	25-Jun-2032	ok
DateCreated	crd	num	4557	11193704	25-Jun-2032	ok
LastUpdated	crlu	num	3359	11193704	25-Jun-2032	ok
SeqLength	sl	num	61711	11193704	25-Jun-2032	ok

Figure 3.29 The **Databank Information** page showing datafield hyperlinks.

Click on one of the datafield hyperlinks (e.g., **Description**) to go to its **Field Information** page (see Figure 3.25).

CHAPTER

4

LINKS TO ADDITIONAL DATA

After you have queried a databank using SRS, you may find that you want to refine your search further. For example, you may want to find further information on the entries returned by your initial search, e.g., by looking for related entries in another databank. Alternatively you may have queried a databank that returned several entries relating to your query, but you only want to know about the entries that relate to a specific protein. During this chapter you will learn:

- How to use SRS links to find entries which are related to each other in more than one databank.

4.1 What is a Link?

A databank entry may contain references to other databanks, and vice versa. In SRS these relationships are known as links and can be used to extend a query across multiple databanks. Thus you can obtain all the entries in one databank that are linked to an entry (or entries) in another databank.

From a user perspective there are two types of link: hypertext links and index links (query links).

Hypertext links are links between entries which are displayed as hypertext. These are hardcoded into SRS and you can use them whenever you wish. They are useful for examining entries that are referenced directly from entries.

Index links are built into the SRS indices at the same time as databanks are added. They allow you to construct queries using relationships between databanks. They require SRS to search through entries or indices in other databanks, looking for matches.

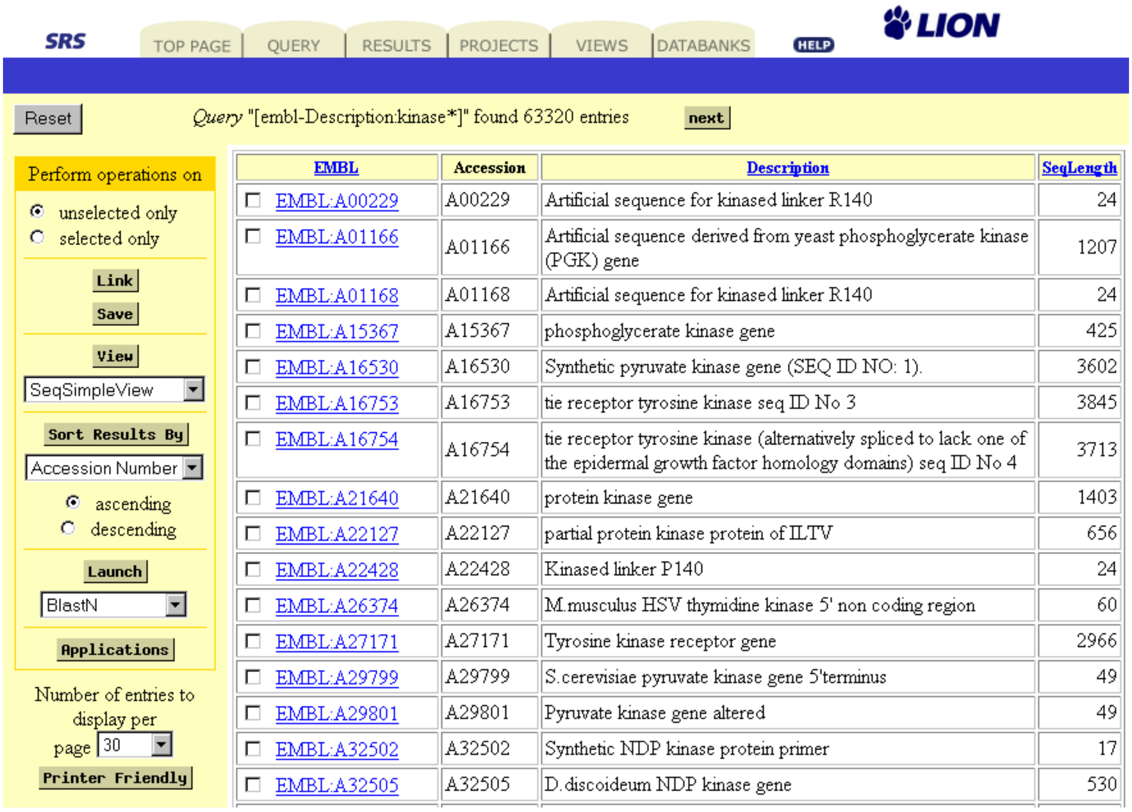
It is assumed that you are already familiar with hypertext links, so only a limited demonstration of them is given in this chapter (see section 4.2 “Hypertext Links”, p. 86). The remainder of this chapter is devoted to explaining index links.

4.2 Hypertext Links

Some entries have hypertext links which allow you to access linked entries directly. Example 4.1 shows such an entry.

Example 4.1 Accessing a hypertext linked entry.

1. Query EMBL for entries whose description field contains the word “kinase”.
   EMBL entries will be displayed on the **Query Result** page.



Query "[embl-Description:kinase*]" found 63320 entries

Perform operations on
● unselected only
○ selected only

SeqSimpleView

Accession Number
● ascending
○ descending

BlastN

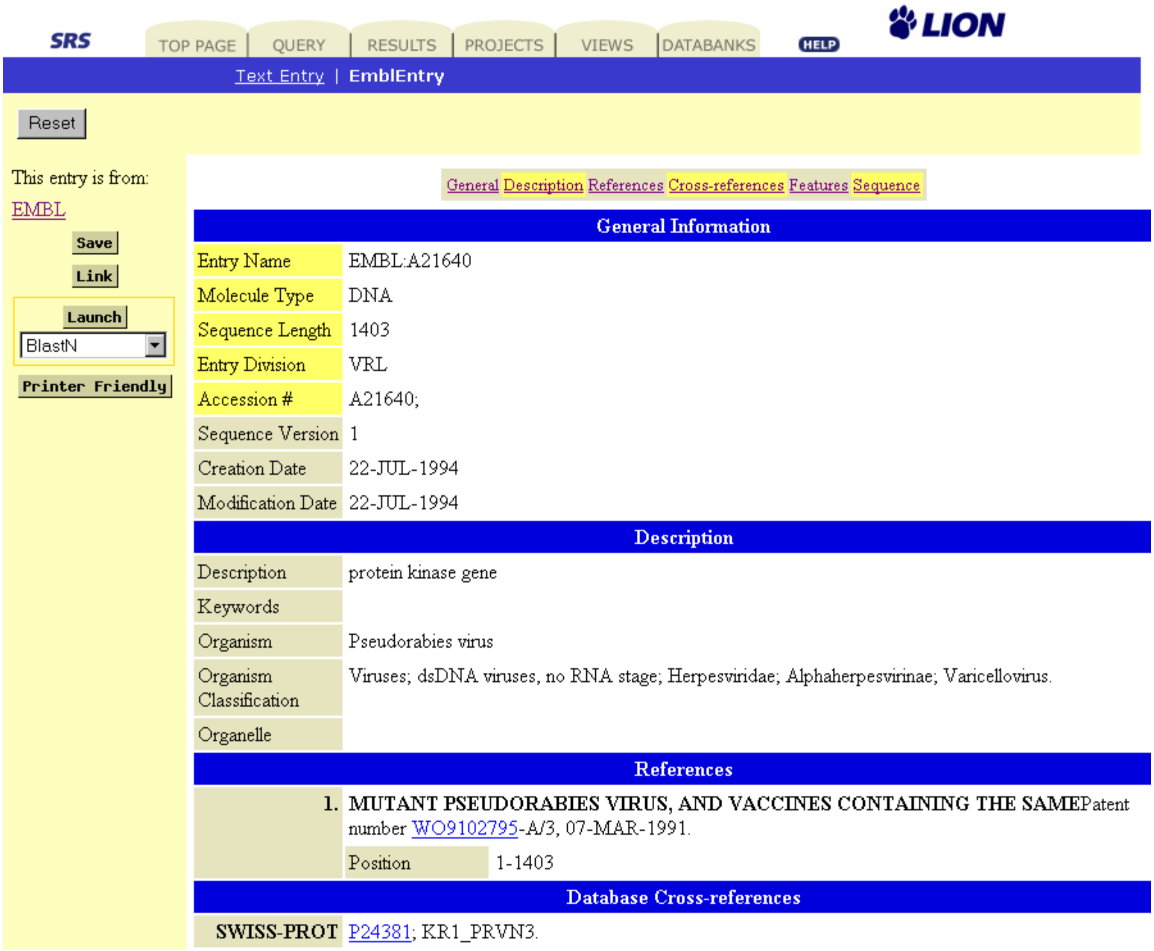
Number of entries to display per page 30

	EMBL	Accession	Description	SeqLength
☐	EMBL:A00229	A00229	Artificial sequence for kinased linker R140	24
☐	EMBL:A01166	A01166	Artificial sequence derived from yeast phosphoglycerate kinase (PGK) gene	1207
☐	EMBL:A01168	A01168	Artificial sequence for kinased linker R140	24
☐	EMBL:A15367	A15367	phosphoglycerate kinase gene	425
☐	EMBL:A16530	A16530	Synthetic pyruvate kinase gene (SEQ ID NO: 1).	3602
☐	EMBL:A16753	A16753	tie receptor tyrosine kinase seq ID No 3	3845
☐	EMBL:A16754	A16754	tie receptor tyrosine kinase (alternatively spliced to lack one of the epidermal growth factor homology domains) seq ID No 4	3713
☐	EMBL:A21640	A21640	protein kinase gene	1403
☐	EMBL:A22127	A22127	partial protein kinase protein of ILTV	656
☐	EMBL:A22428	A22428	Kinased linker P140	24
☐	EMBL:A26374	A26374	M.musculus HSV thymidine kinase 5' non coding region	60
☐	EMBL:A27171	A27171	Tyrosine kinase receptor gene	2966
☐	EMBL:A29799	A29799	S.cerevisiae pyruvate kinase gene 5'terminus	49
☐	EMBL:A29801	A29801	Pyruvate kinase gene altered	49
☐	EMBL:A32502	A32502	Synthetic NDP kinase protein primer	17
☐	EMBL:A32505	A32505	D.discoideum NDP kinase gene	530

Figure 4.1 The **Query Result** page, with results sorted by Accession Number (see section 3.6 “Sorting Results”, p. 76).

Linked data in the entry:

2. Click on the hypertext link beside an entry to view the entire entry.



This entry is from: EMBL

General Information

Entry Name	EMBL:A21640
Molecule Type	DNA
Sequence Length	1403
Entry Division	VRL
Accession #	A21640;
Sequence Version	1
Creation Date	22-JUL-1994
Modification Date	22-JUL-1994

Description

Description	protein kinase gene
Keywords	
Organism	Pseudorabies virus
Organism Classification	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae; Varicellovirus.
Organelle	

References

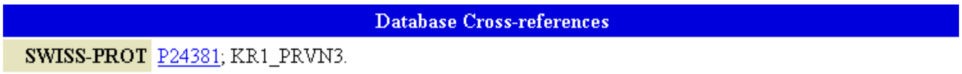
1. **MUTANT PSEUDORABIES VIRUS, AND VACCINES CONTAINING THE SAME**Patent number WO9102795-A/3, 07-MAR-1991.
Position 1-1403

Database Cross-references

SWISS-PROT P24381; KR1_PRVN3.

Figure 4.2 The **Entry** page for the EMBL entry with accession number A21640.

3. Scroll through the entry until you find a hyperlink (e.g. in the entry shown in Figure 4.2 there is a hypertext link to **SWISS-PROT**). This will have a unique identifier (accession number) — “P24381” in figures 4.2 and 4.3.



Database Cross-references

SWISS-PROT P24381; KR1_PRVN3.

Figure 4.3 Database Cross-references entry, showing a hypertext link to **SWISS-PROT**.

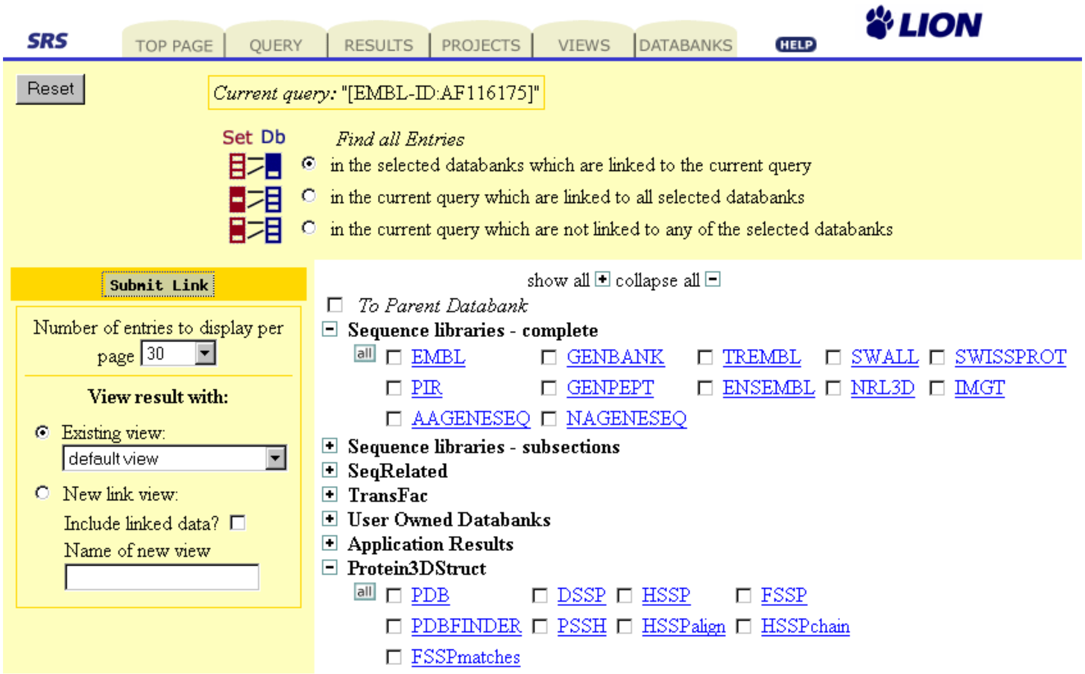
If you do not find a link in your chosen entry, try choosing a different entry from the **Query Result** page.

4. Click the hypertext link to view the **Entry** page for the SWISS-PROT entry.

4.3 Index Links

4.3.1 Searching for Data using Index Links

It is assumed that you have arrived at the **LINK** page (Figure 4.4). Use section 4.4 “Getting to the LINK Page” (p. 93) to help you navigate here if necessary.



Current query: "[EMBL-ID:AF116175]"

Find all Entries

● in the selected databanks which are linked to the current query

○ in the current query which are linked to all selected databanks

○ in the current query which are not linked to any of the selected databanks

Number of entries to display per page 30

View result with:

● Existing view: default view

○ New link view:

Include linked data? ☐

Name of new view

show all ⊞ collapse all ⊟

☐ *To Parent Databank*

⊟ **Sequence libraries - complete**

all ☐ EMBL ☐ GENBANK ☐ TREMBL ☐ SWALL ☐ SWISSPROT

☐ PIR ☐ GENPEPT ☐ ENSEMBL ☐ NRL3D ☐ IMGT

☐ AAGENESEQ ☐ NAGENESEQ

⊞ **Sequence libraries - subsections**

⊞ **SeqRelated**

⊞ **TransFac**

⊞ **User Owned Databanks**

⊞ **Application Results**

⊟ **Protein3DStruct**

all ☐ PDB ☐ DSSP ☐ HSSP ☐ FSSP

☐ PDBFINDER ☐ PSSH ☐ HSSPalign ☐ HSSPchain

☐ FSSPmatches

Figure 4.4 The **LINK** page for linking initiated from the **Manage your Query Results** and **Query Result** pages.

Note: There are two forms of the **LINK** page. If you have initiated the links from either the **Manage your Query Results** or **Query Result** pages then you will see the **LINK** page as shown in Figure 4.4. If you have initiated linking from the **Entry** page then the **Find all Entries** options are not available (see Figure 4.5).

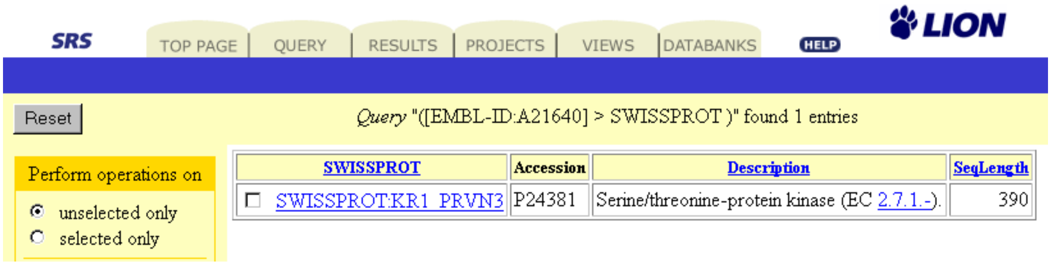
Link the entry: "SWISSPROT:DPY2_RAT" to other databanks

Number of entries to display per page 30

● Existing view: default view

Figure 4.5 The **LINK** page for linking initiated from the **Entry** page.

1. Choose the databank in which you wish to search for links, by ticking the check box to the left of it.
2. Set the **Find all Entries** option you require (see section 4.3.2 “Find All Entries Options” p. 92).
3. Click the `Submit Link` button to display the results.



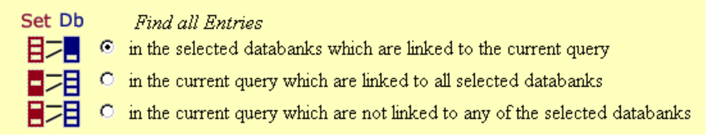
Query "([EMBL-ID:A21640] > SWISSPROT)" found 1 entries

SWISSPROT	Accession	Description	SeqLength
☐ SWISSPROT:KR1_PRVN3	P24381	Serine/threonine-protein kinase (EC 2.7.1.-).	390

Figure 4.6 Typical **Query Result** page for a linking request.

4.3.2 Find All Entries Options

Unless you have initiated your search for links from the **Entry** page there are several options for choosing the set of entries for which you wish to search. These are listed on the **LINK** page under the heading **Find all Entries**



Find all Entries

● in the selected databanks which are linked to the current query

○ in the current query which are linked to all selected databanks

○ in the current query which are not linked to any of the selected databanks

Figure 4.7 Find all Entries options.

The meanings for these options are as follows:

in the selected databanks which are linked to the current query.

This returns entries from other databanks which have links with entries in the current query.

in the current query which are linked to all selected databanks.

This limits the query so that it includes only the entries (from the original query) which are linked to all of the selected databanks.

in the current query which are not linked to any of the selected databanks.

This limits the query so that it includes only the entries (from the original query) which do not have links to the specified databanks.

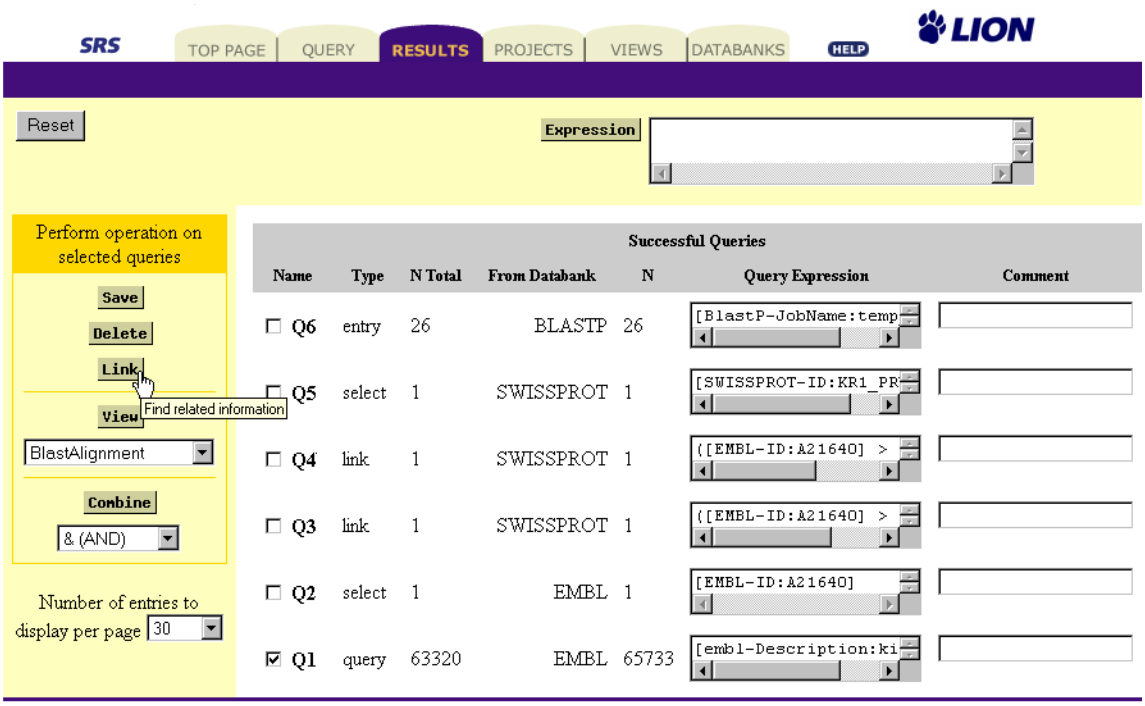
4.4 Getting to the LINK Page

It is possible to initiate a linking operation from many pages. These include the **Manage your Query Results** page (see section 4.4.1 “Finding Links from the Manage your Query Results Page” p. 93), the **Query Result** page (see section 4.4.2 “Finding Links from the Query Result Page” p. 94) and the **Entry** page (see section 4.4.3 “Finding Links from the Entry Page” p. 97).

4.4.1 Finding Links from the Manage your Query Results Page

This method is useful for finding links for complete queries. You can search for links from a single query or from multiple queries.

There are two ways to find links for your queries from the **Manage your Query Results** page. The first and most common method is to tick the check box that corresponds to a query set and click the Link button.



Successful Queries

Name	Type	N Total	From Databank	N	Query Expression	Comment
☐ Q6	entry	26	BLASTP	26	[BlastP-JobName:temp	
☐ Q5	select	1	SWISSPROT	1	[SWISSPROT-ID:KR1_PR	
☐ Q4	link	1	SWISSPROT	1	([EMBL-ID:A21640] >	
☐ Q3	link	1	SWISSPROT	1	([EMBL-ID:A21640] >	
☐ Q2	select	1	EMBL	1	[EMBL-ID:A21640]	
☑ Q1	query	63320	EMBL	65733	[embl-Description:ki	

Figure 4.8 The **Manage your Query Results** page.

This takes you to the **LINK** page (see section 4.3 “Index Links” p. 89) from where you can complete your search for linked data.

The second option is to use the text box beside the **Expression** button on the **Manage your Query Results** page. This option is discussed later (see section 4.5 “Expression Linking” p. 98).

4.4.2 Finding Links from the Query Result Page

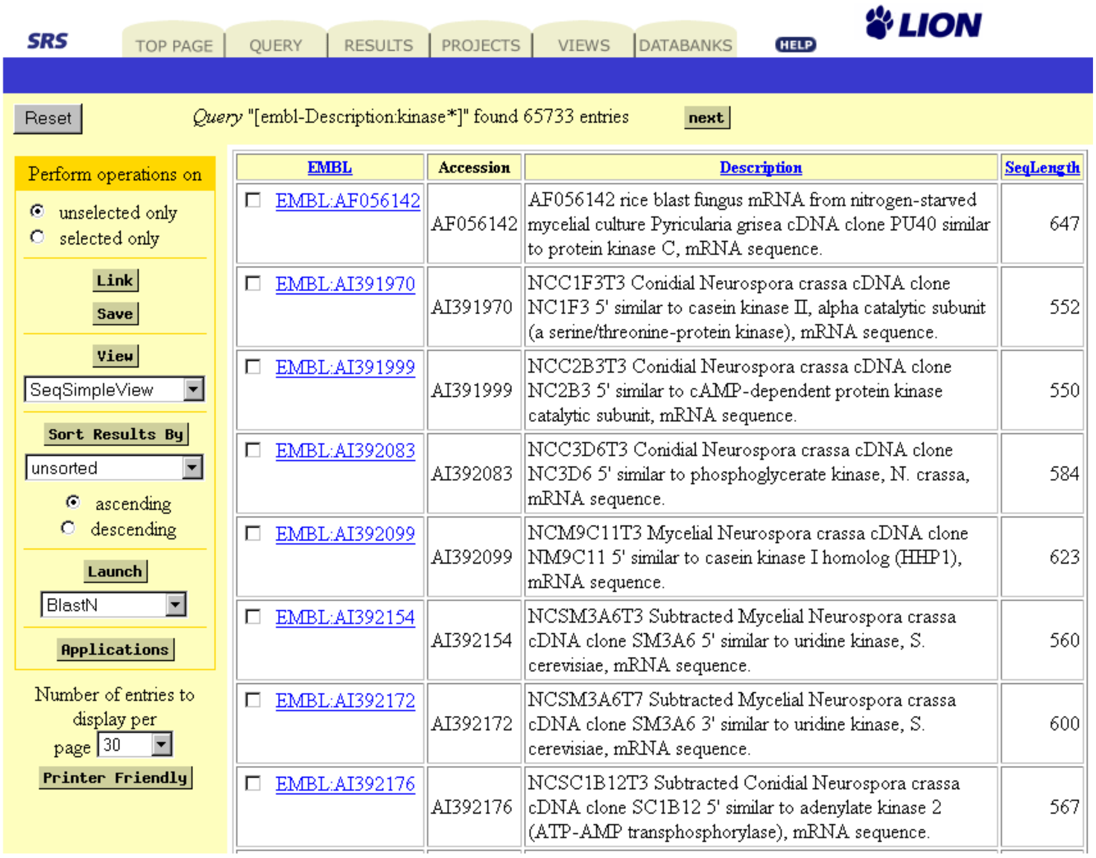
This method is useful for searching for links from individual entries or groups of entries.

It is assumed that you have made your query as normal. From the **Query Result** page, choose those entries which are of interest and click the Link button to display the **LINK** page (see section 4.3 “Index Links” p. 89).

Note: You can choose whether to search for links to your results based only on the selected entries/results (**selected only**) or based only on the entries/results that are not selected (**unselected only**) items, by setting the **Perform operations on** option. This is particularly useful if you wish to find links for a large proportion of your results. The **Perform operations on** option is usually set to **unselected only** when you first visit this page, but will change to **selected only** automatically when you start selecting items. Change it back if you wish to use **unselected only**.

Example 4.2 Initiating a link from the Query Result page.

1. Query EMBL for entries whose description field contains the word “kinase”.
   EMBL entries will be displayed on the **Query Result** page.



Query "[embl-Description:kinase*]" found 65733 entries

Perform operations on
● unselected only
○ selected only

SeqSimpleView

unsorted
● ascending
○ descending

BlastN

Number of entries to display per page 30

EMBL	Accession	Description	SeqLength
☐ EMBL:AF056142	AF056142	AF056142 rice blast fungus mRNA from nitrogen-starved mycelial culture Pyricularia grisea cDNA clone PU40 similar to protein kinase C, mRNA sequence.	647
☐ EMBL:AI391970	AI391970	NCC1F3T3 Conidial Neurospora crassa cDNA clone NC1F3 5' similar to casein kinase II, alpha catalytic subunit (a serine/threonine-protein kinase), mRNA sequence.	552
☐ EMBL:AI391999	AI391999	NCC2B3T3 Conidial Neurospora crassa cDNA clone NC2B3 5' similar to cAMP-dependent protein kinase catalytic subunit, mRNA sequence.	550
☐ EMBL:AI392083	AI392083	NCC3D6T3 Conidial Neurospora crassa cDNA clone NC3D6 5' similar to phosphoglycerate kinase, N. crassa, mRNA sequence.	584
☐ EMBL:AI392099	AI392099	NCM9C11T3 Mycelial Neurospora crassa cDNA clone NM9C11 5' similar to casein kinase I homolog (HHP1), mRNA sequence.	623
☐ EMBL:AI392154	AI392154	NCSM3A6T3 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 5' similar to uridine kinase, S. cerevisiae, mRNA sequence.	560
☐ EMBL:AI392172	AI392172	NCSM3A6T7 Subtracted Mycelial Neurospora crassa cDNA clone SM3A6 3' similar to uridine kinase, S. cerevisiae, mRNA sequence.	600
☐ EMBL:AI392176	AI392176	NCSC1B12T3 Subtracted Conidial Neurospora crassa cDNA clone SC1B12 5' similar to adenylate kinase 2 (ATP-AMP transphosphorylase), mRNA sequence.	567

Figure 4.9 The **Query Result** page.

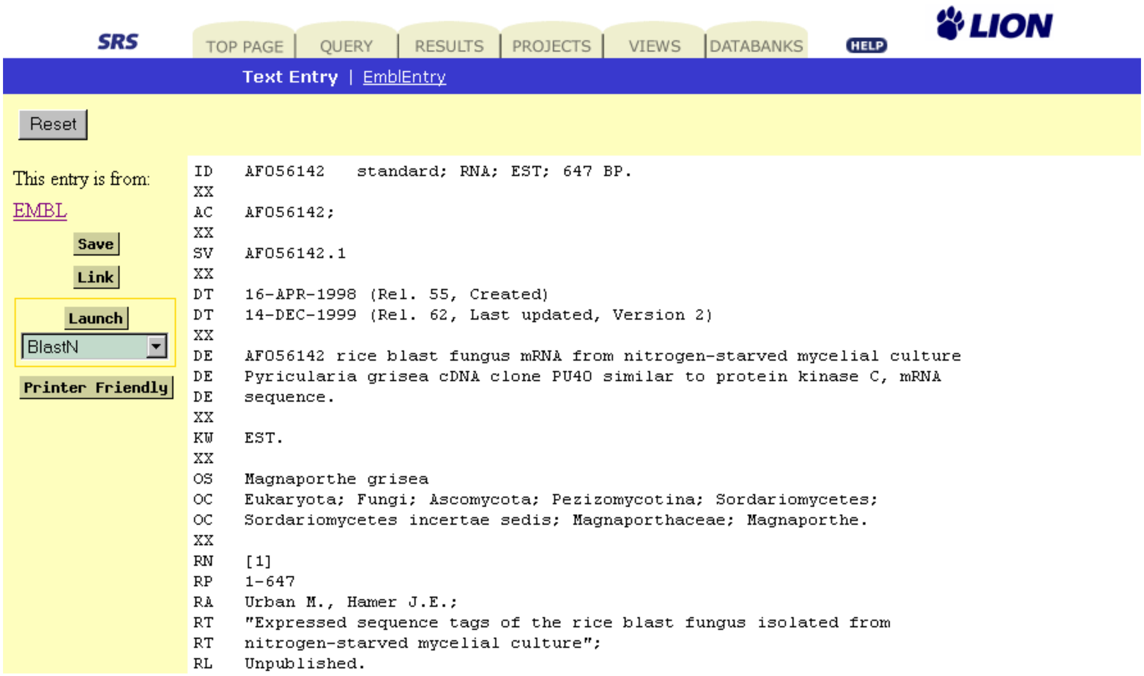
2. Select the entries of interest, by ticking the check boxes beside them.

3. Click on the **Link** button. This will display the **LINK** page. This will display the **LINK** page (see section 4.3 “Index Links” p. 89).

4.4.3 Finding Links from the Entry Page

This method is useful for linking from individual entries.

It is assumed that you have made your query as normal. When the **Query Result** page appears, choose one of the entries and click on its hyperlink to display the **Entry** page for that entry. Note that you can usually view such a page using **Text Entry** format or in a databank-specific format, e.g., **SwissEntry** format.



```
ID   AF056142   standard; RNA; EST; 647 BP.
XX
AC   AF056142;
XX
SV   AF056142.1
XX
DT   16-APR-1998 (Rel. 55, Created)
DT   14-DEC-1999 (Rel. 62, Last updated, Version 2)
XX
DE   AF056142 rice blast fungus mRNA from nitrogen-starved mycelial culture
DE   Pyricularia grisea cDNA clone PU40 similar to protein kinase C, mRNA
DE   sequence.
XX
KW   EST.
XX
OS   Magnaporthe grisea
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
XX
RN   [1]
RP   1-647
RA   Urban M., Hamer J.E.;
RT   "Expressed sequence tags of the rice blast fungus isolated from
RT   nitrogen-starved mycelial culture";
RL   Unpublished.
```

Figure 4.10 A typical **Entry** page shown using **Text Entry** format.

Click the **Link** button. This will display the **LINK** page for single entry linking (see section 4.3 “Index Links” p. 89).

4.5 Expression Linking

The text box beside the `Expression` button (e.g., on the **Manage your Query Results** page) is a useful alternative to the linking methods discussed earlier in this chapter. It can be used to search for a link between two or more sets of results or between a set of results and a databank.

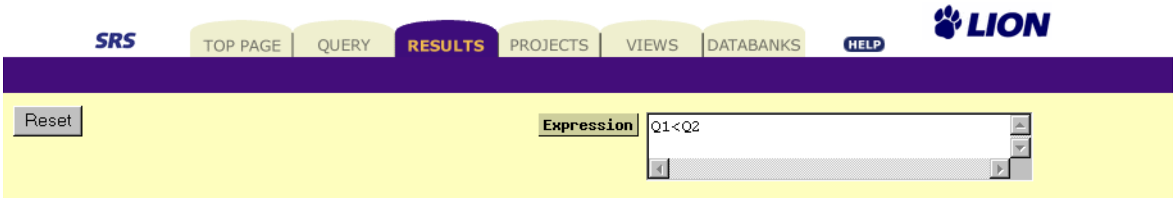
Note: Using the text box beside the `Expression` button allows you to search for links without going through the **LINK** page.

4.5.1 Expression Linking Procedure

You might be interested in the entries from a set of DNA sequences (in **Q1**) that are linked to a set of protein sequences (in **Q2**).

If there are links between entries in the queries, you have the choice of returning the entries in **Q1** that have links with **Q2** (which is what is wanted here) or of returning the entries in **Q2** that have links with **Q1**. (See also the section 4.5.2 “Linking Operators” p. 99.)

The `Expression` button and text box are shown in Figure 4.11. You should enter your expression (e.g., “Q1 < Q2”) as shown.



Q1<Q2

Figure 4.11 Expression linking.

To link complete query sets using expression linking:

1. Type in the name of the query set to be linked (“Q1”), the type of query (e.g., “<” or “>”, see section 4.5.2, “Linking Operators”, p. 99, for more details) and the databank or second query (“Q2”) to which the set should link. For instance:

Q1 < Q2

will list all the entries in **Q1** that are linked to **Q2**.

Q1 > Q2

will list all the entries in **Q2** that are linked to **Q1**. See the section 4.5.2 “Linking Operators” (p. 99).

2. Click the **Expression** button.

4.5.2 Linking Operators

Expression linking uses linking operators to describe the nature of the links between the specified queries or databanks. Typical operators are the < and > which indicate a link:

<

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

>

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

You can combine linking (< and >) and logical (&, | and !) operators to build up more complex queries (see the examples in Table 4.1).

Table 4.1 Linking operations.

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

4.5.3 Expression Linking Examples

If you have a set of EMBL entries in a query, **Q3**, which you wish to search for links with the SWISS-PROT databank, type:

```
Q3 < swissprot
```

This will show the EMBL entries from the original query that have links to the SWISS-PROT databank.

If you would rather see the SWISS-PROT entries that the above operation returned, turn the linking operator around so that it points towards SWISS-PROT:

```
Q3 > swissprot
```

This returns the SWISS-PROT entries that have links with the entries in **Q3**.

CHAPTER

5

VIEWS

SRS includes a number of predefined views that you can use to change the way in which your results are displayed. In addition, you can create your own views. During this chapter you will learn:

- About the predefined views available in SRS.
- How to work with views.
- How to create and use your own views.

5.1 What is a View?

SRS uses views to determine which data types are shown. A view may include only a few data types, or it may display all available types of data.

In addition to the predefined views which are part of SRS you can create your own views. A view can be defined to work with a specific databank, or with several databanks. The data in a view could be from one or more databanks.

For convenience, views can be categorized as “predefined views” and “project views” (see section 5.3 “Creating Views”, p. 118). Predefined views are the default views that are available to everybody using an SRS server. Some of these are designed to work with specific databanks and will only be available when they are appropriate. Project views are views that you define, using the **View Manager** page (see section 5.3.2 “Creating Views using the View Manager Pages”, p. 124).

5.2 Using Views

Views can be applied from any page that creates or holds a query. The following list identifies the pages where views can be created or applied to data sets.

- The **Standard Query Form**.
- The **Extended Query Form**.
- The **Query Result** page (which contains the results of a query).
- The **Entry** page (which contains the details of a single entry).
- The **Manage your Query Results** page.

- The **LINK** page.
- The **Launch** page (from which you can launch applications).

The views drop-down list (available on pages such as the **Manage your Query Results** page) contains the views that are currently available (see Figure 5.1).

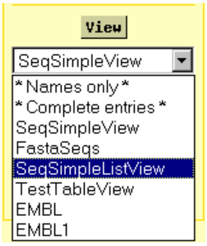


Figure 5.1 Views drop-down list.

5.2.1 How to Apply Views

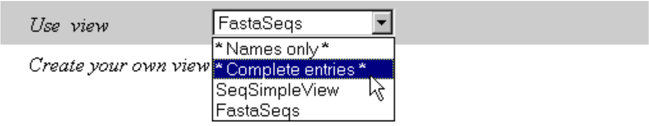
This section gives instructions for applying views to your query results. In the first two, you will begin by querying a databank and setting the query options for views so that the query results are displayed with your chosen view. The others show how to apply your view to an already existing set.

Applying a View From the Standard Query Form

When you query a databank with either of the query forms you are given an opportunity to choose (from a drop-down list) a view to be used for the query results.

To apply a view to an EMBL search from the **Standard Query Form**:

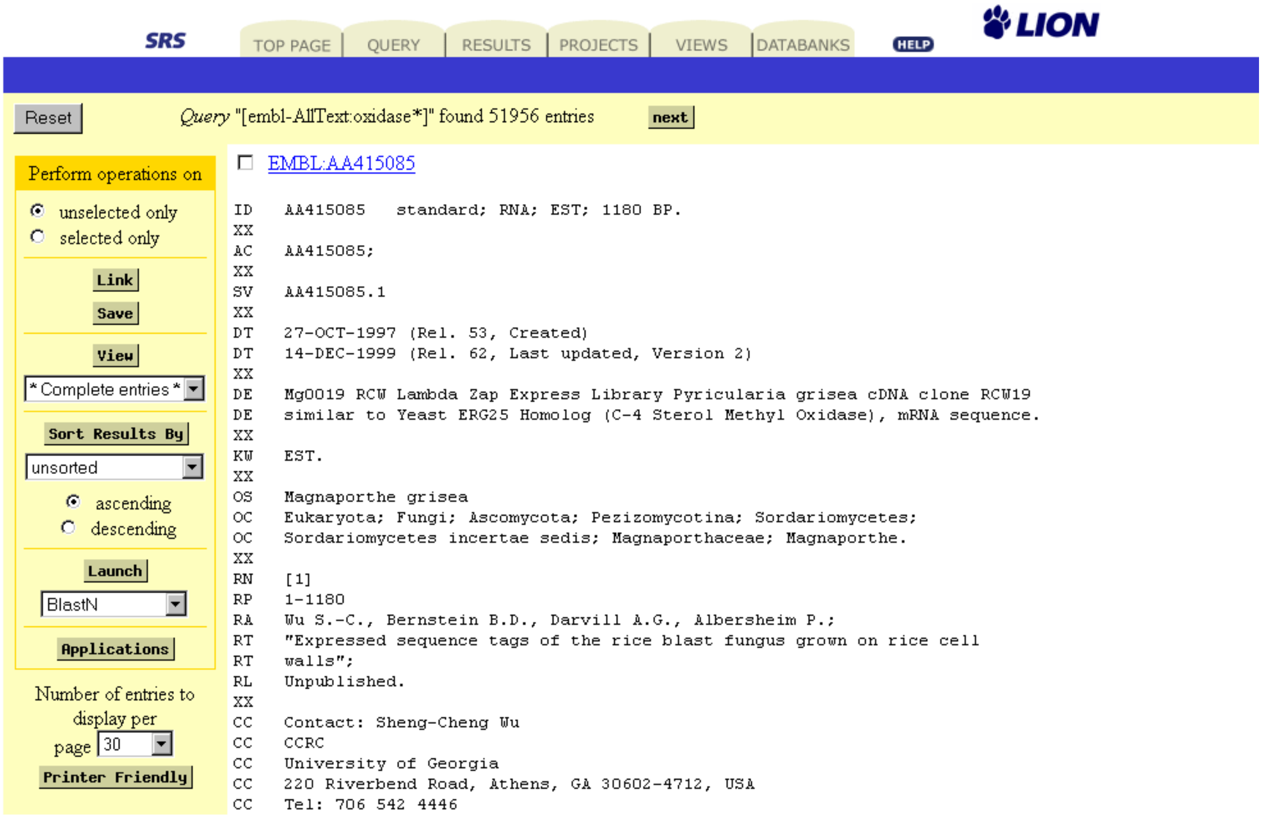
1. On the **Select Databanks To Search** page, select the EMBL databank using the check box beside it and click the `Standard` button. (See chapter 3, “Querying with SRS”, p. 51, for more information.)
2. Enter “oxidase” in the input box. The default **AllText** datafield is fine.
3. Select a view from the drop-down list beside the heading **Use view** (see Figure 5.2).



Use view FastaSeqs

Figure 5.2 Choosing a predefined view for results from the **Standard Query Form**.

4. Click on the `Submit Query` button.
   This will display the **Query Result** page with the results of your query displayed using the selected view (see Figure 5.3).



Query "[embl-AllText:oxidase*]" found 51956 entries

☐ EMBL:AA415085

```
ID   AA415085   standard; RNA; EST; 1180 BP.
XX
AC   AA415085;
XX
SV   AA415085.1
XX
DT   27-OCT-1997 (Rel. 53, Created)
DT   14-DEC-1999 (Rel. 62, Last updated, Version 2)
XX
DE   Mg0019 RCW Lambda Zap Express Library Pyricularia grisea cDNA clone RCW19
DE   similar to Yeast ERG25 Homolog (C-4 Sterol Methyl Oxidase), mRNA sequence.
XX
KW   EST.
XX
OS   Magnaporthe grisea
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
XX
RN   [1]
RP   1-1180
RA   Wu S.-C., Bernstein B.D., Darvill A.G., Albersheim P.;
RT   "Expressed sequence tags of the rice blast fungus grown on rice cell
RT   walls";
RL   Unpublished.
XX
CC   Contact: Sheng-Cheng Wu
CC   CCRC
CC   University of Georgia
CC   220 Riverbend Road, Athens, GA 30602-4712, USA
CC   Tel: 706 542 4446
```

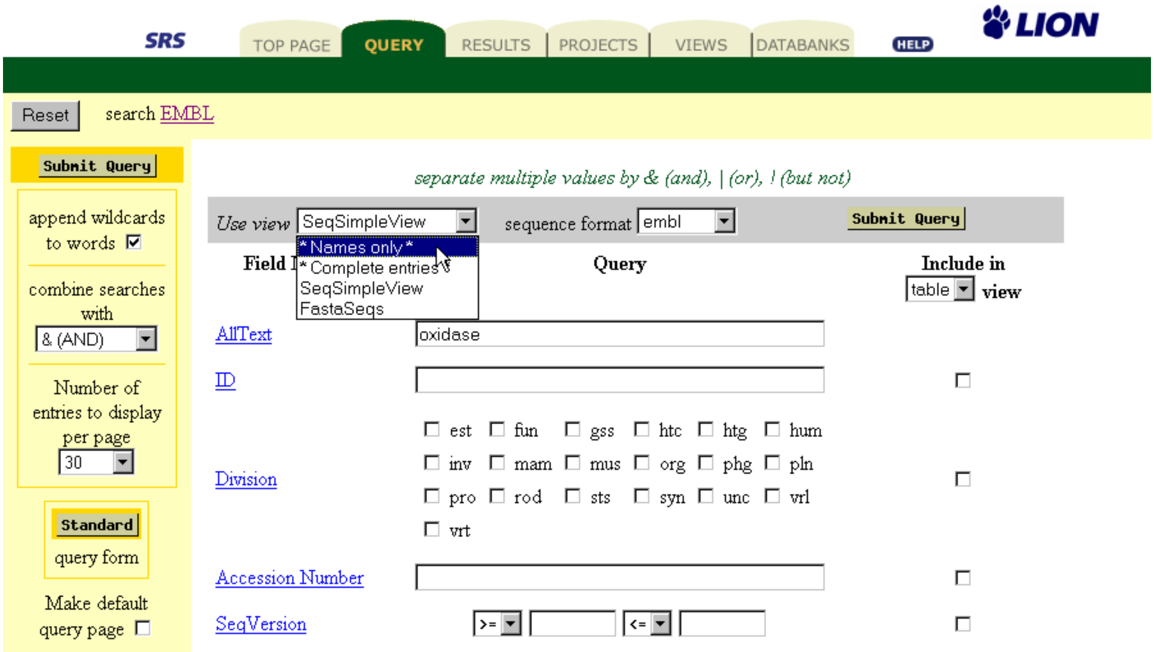
Figure 5.3 The **Query Result** page using the selected view.

Applying a View From the Extended Query Form

The previous section demonstrated the steps involved in applying a view to the results of a query at the same time as the query is performed from the **Standard Query Form**. On the **Extended Query Form** the position of the drop-down views menu and the look of the query form are different but the basic steps are similar.

To apply a view to an EMBL query from the **Extended Query Form**:

1. On the **Select Databanks To Search** page, select the EMBL databank using its check box and click the **Extended** button. (See Chapter 3, "Querying" for more information.)
2. Enter “oxidase” in the **AllText** text box.
3. Select the view for your query results from the drop-down list beside the heading **Use view** (see Figure 5.4).



append wildcards to words ☑

combine searches with & (AND)

Number of entries to display per page 30

Use view SeqSimpleView

sequence format embl

AllText oxidase

Make default query page ☐

Figure 5.4 Choosing a view for your query from the **Extended Query Form**.

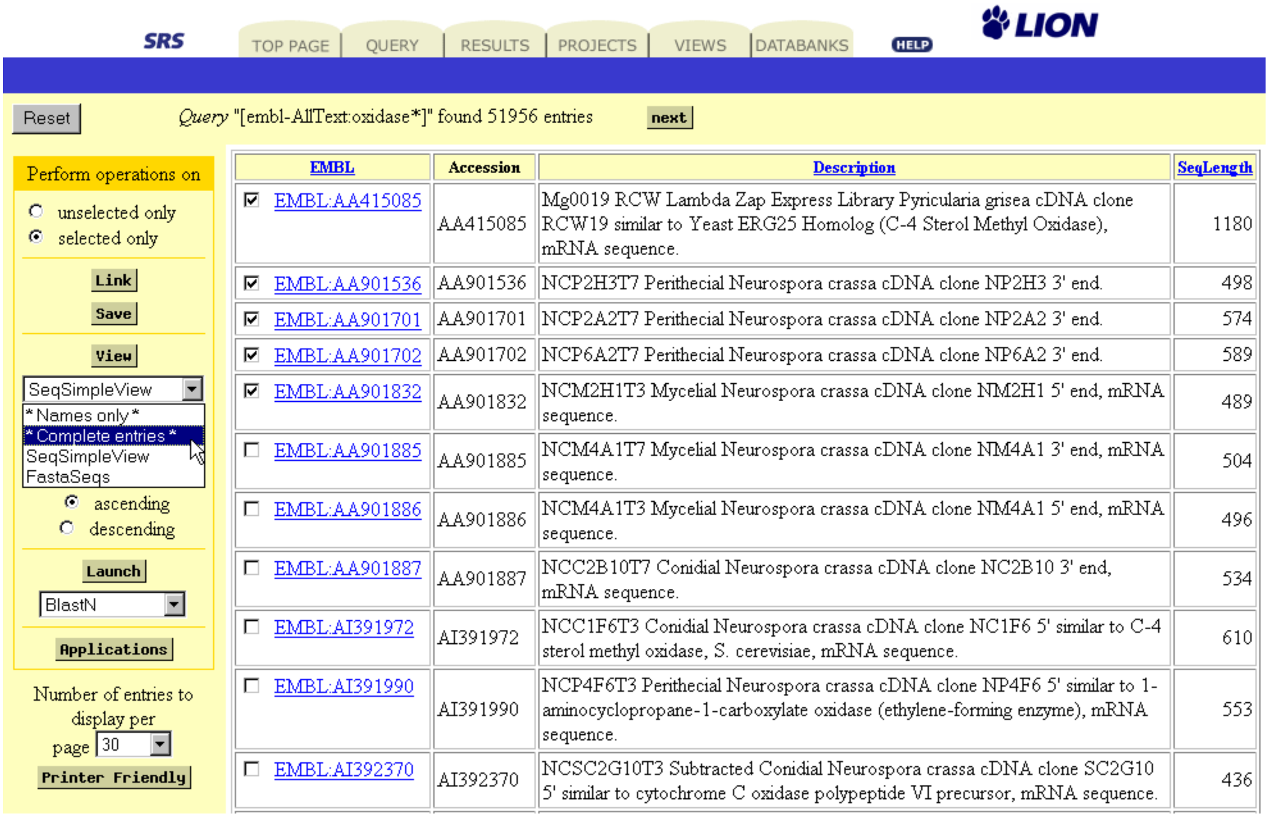
4. Click the **Submit Query** button.
This will display the **Query Result** page, with the results of your query, displayed using the selected view. See, for example, Figure 5.3.

From the Query Result Page

Querying, linking, and launching applications always returns the **Query Result** page. This page contains a list of the entries that meet the requirements specified. If you do not select a view for the results, when the operation is run, a default view is used.

To change the view on the **Query Result** page:

In the options area (see Figure 1.3 on page 5) you will see a drop-down list below the **View** button, (see Figure 5.5). The drop-down list displays the name of the view currently applied.



Query "[embl-AllText:oxidase*]" found 51956 entries

Perform operations on
○ unselected only
● selected only

SeqSimpleView
Names only
Complete entries
SeqSimpleView
FastaSeqs

● ascending
○ descending

BlastN

Number of entries to display per page 30

	EMBL	Accession	Description	SeqLength
☑	EMBL:AA415085	AA415085	Mg0019 RCW Lambda Zap Express Library Pyricularia grisea cDNA clone RCW19 similar to Yeast ERG25 Homolog (C-4 Sterol Methyl Oxidase), mRNA sequence.	1180
☑	EMBL:AA901536	AA901536	NCP2H3T7 Perithecial Neurospora crassa cDNA clone NP2H3 3' end.	498
☑	EMBL:AA901701	AA901701	NCP2A2T7 Perithecial Neurospora crassa cDNA clone NP2A2 3' end.	574
☑	EMBL:AA901702	AA901702	NCP6A2T7 Perithecial Neurospora crassa cDNA clone NP6A2 3' end.	589
☑	EMBL:AA901832	AA901832	NCM2H1T3 Mycelial Neurospora crassa cDNA clone NM2H1 5' end, mRNA sequence.	489
☐	EMBL:AA901885	AA901885	NCM4A1T7 Mycelial Neurospora crassa cDNA clone NM4A1 3' end, mRNA sequence.	504
☐	EMBL:AA901886	AA901886	NCM4A1T3 Mycelial Neurospora crassa cDNA clone NM4A1 5' end, mRNA sequence.	496
☐	EMBL:AA901887	AA901887	NCC2B10T7 Conidial Neurospora crassa cDNA clone NC2B10 3' end, mRNA sequence.	534
☐	EMBL:AI391972	AI391972	NCC1F6T3 Conidial Neurospora crassa cDNA clone NC1F6 5' similar to C-4 sterol methyl oxidase, S. cerevisiae, mRNA sequence.	610
☐	EMBL:AI391990	AI391990	NCP4F6T3 Perithecial Neurospora crassa cDNA clone NP4F6 5' similar to 1-aminocyclopropane-1-carboxylate oxidase (ethylene-forming enzyme), mRNA sequence.	553
☐	EMBL:AI392370	AI392370	NCSC2G10T3 Subtracted Conidial Neurospora crassa cDNA clone SC2G10 5' similar to cytochrome C oxidase polypeptide VI precursor, mRNA sequence.	436

Figure 5.5 Selecting the view on the **Query Result** page.

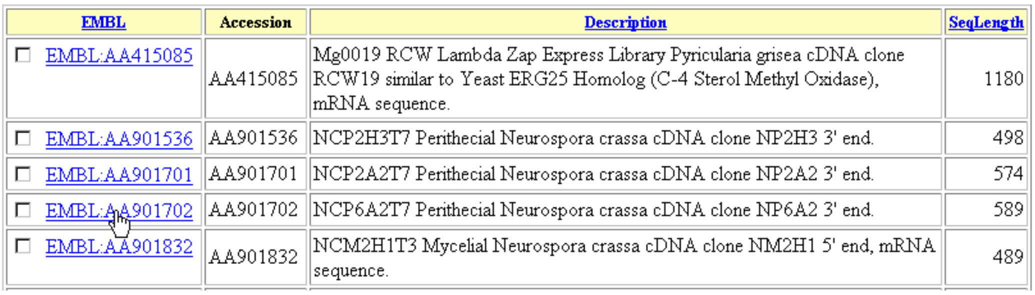
1. Select the view you want to use from the drop-down list.
2. Click the View button.
   The **Query Result** page will be refreshed so that it displays the results using the selected view (see Figure 5.3).

Note: If you select a number of entries prior to clicking the View button then only those entries will be shown when you click View.

Applying a View From the Entry Page

There are two ways to apply a view from the **Entry** page, although only one method is usually available for a given databank. You will either have a simple choice of viewing in a text format or a databank-specific format; or of using a view from the drop-down views menu. The use of the drop-down view menu is the same as on other SRS pages (see, for example the section "From the Query Result Page", p. 109). The example below describes only the first method.

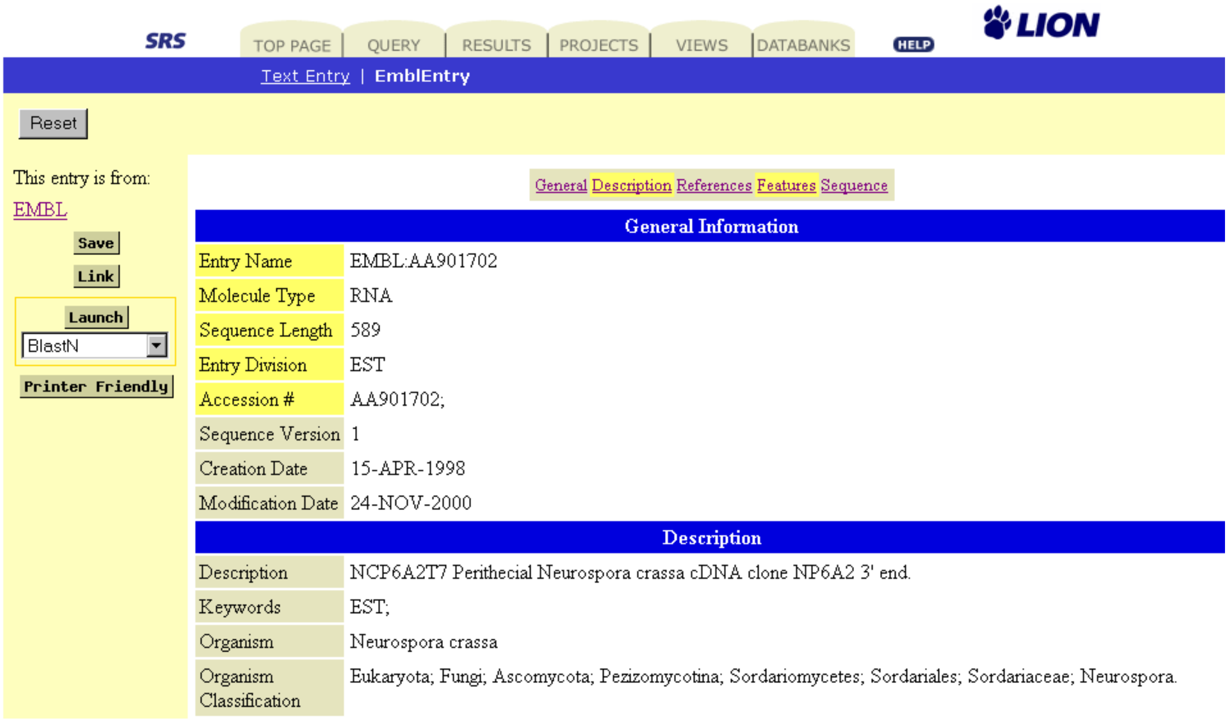
1. You need to be on an **Entry** page, so start with a **Query Result** page, and click the entry name hyperlink for one of the entries (see Figure 5.6).



EMBL	Accession	Description	SeqLength
☐ EMBL:AA415085	AA415085	Mg0019 RCW Lambda Zap Express Library Pyricularia grisea cDNA clone RCW19 similar to Yeast ERG25 Homolog (C-4 Sterol Methyl Oxidase), mRNA sequence.	1180
☐ EMBL:AA901536	AA901536	NCP2H3T7 Perithecial Neurospora crassa cDNA clone NP2H3 3' end.	498
☐ EMBL:AA901701	AA901701	NCP2A2T7 Perithecial Neurospora crassa cDNA clone NP2A2 3' end.	574
☐ EMBL:AA901702	AA901702	NCP6A2T7 Perithecial Neurospora crassa cDNA clone NP6A2 3' end.	589
☐ EMBL:AA901832	AA901832	NCM2H1T3 Mycelial Neurospora crassa cDNA clone NM2H1 5' end, mRNA sequence.	489

Figure 5.6 Click the hyperlink for an entry.

2. Assuming that the current databank uses databank-specific and text formats, the **Entry** page will display the entry using the databank-specific format (see Figure 5.7).



This entry is from: EMBL

General Information

Entry Name	EMBL:AA901702
Molecule Type	RNA
Sequence Length	589
Entry Division	EST
Accession #	AA901702;
Sequence Version	1
Creation Date	15-APR-1998
Modification Date	24-NOV-2000

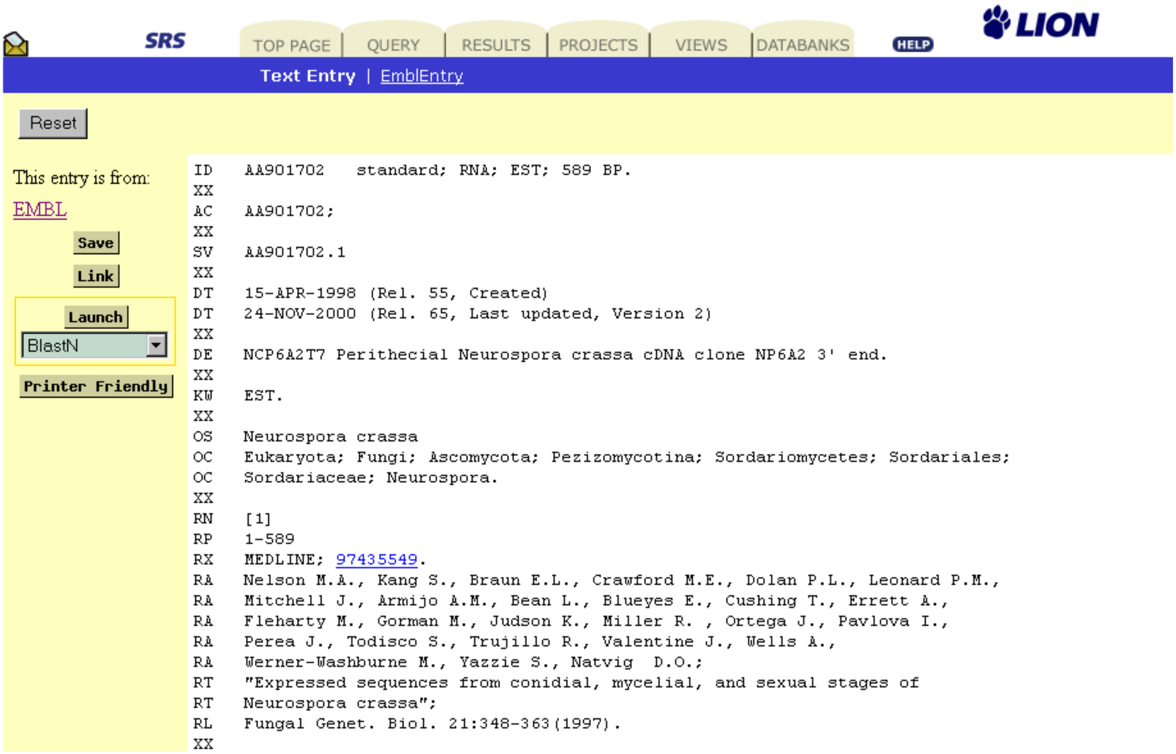
Description

Description	NCP6A2T7 Perithecial Neurospora crassa cDNA clone NP6A2 3' end.
Keywords	EST;
Organism	Neurospora crassa
Organism Classification	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales; Sordariaceae; Neurospora.

Figure 5.7 **Entry** page using **EmblEntry** format.

3. The available viewing options are shown across the top of the page:

4. The option currently being displayed will be shown in bold, with the alternative entry shown as a hyperlink.
   Click on the **Text Entry** hyperlink to view the entry using a text only format (see Figure 5.8).



```
ID   AA901702   standard; RNA; EST; 589 BP.
XX
AC   AA901702;
XX
SV   AA901702.1
XX
DT   15-APR-1998 (Rel. 55, Created)
DT   24-NOV-2000 (Rel. 65, Last updated, Version 2)
XX
DE   NCP6A2T7 Perithecial Neurospora crassa cDNA clone NP6A2 3' end.
XX
KW   EST.
XX
OS   Neurospora crassa
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariales;
OC   Sordariaceae; Neurospora.
XX
RN   [1]
RP   1-589
RX   MEDLINE; 97435549.
RA   Nelson M.A., Kang S., Braun E.L., Crawford M.E., Dolan P.L., Leonard P.M.,
RA   Mitchell J., Armijo A.M., Bean L., Blueyes E., Cushing T., Errett A.,
RA   Fleharty M., Gorman M., Judson K., Miller R. , Ortega J., Pavlova I.,
RA   Perea J., Todisco S., Trujillo R., Valentine J., Wells A.,
RA   Werner-Washburne M., Yazzie S., Natvig  D.O.;
RT   "Expressed sequences from conidial, mycelial, and sexual stages of
RT   Neurospora crassa";
RL   Fungal Genet. Biol. 21:348-363(1997).
XX
```

Figure 5.8 Entry page using **Text Entry** format.

Applying a View From the Manage your Query Results Page

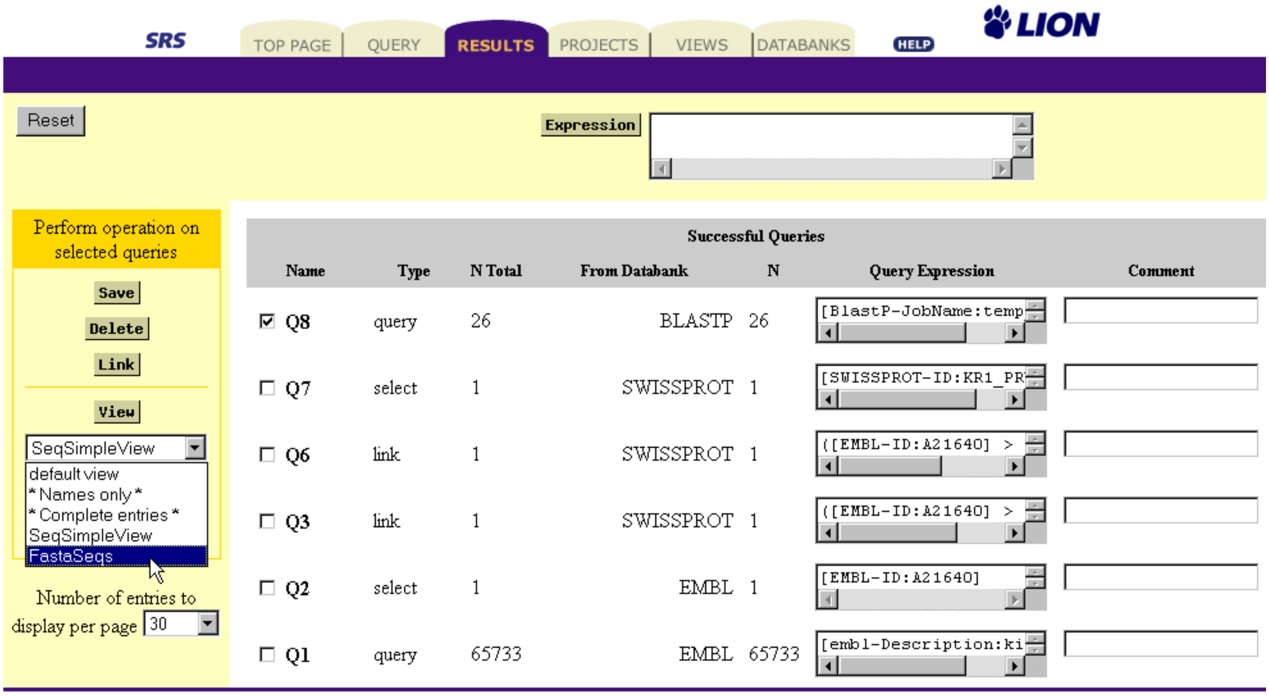
To apply a view from the **Manage your Query Results** page:

1. Select the query to which you want to apply the view, by ticking the check box beside it.

Note: Only one query can be displayed at a time. If more than one query is chosen, only one will be displayed.

---

2. Select the required view from the drop-down list in the options area (see Figure 5.9).



Successful Queries

	Name	Type	N Total	From Databank	N	Query Expression	Comment
☑	Q8	query	26	BLASTP	26	[BlastP-JobName:temp	
☐	Q7	select	1	SWISSPROT	1	[SWISSPROT-ID:KR1_PR	
☐	Q6	link	1	SWISSPROT	1	([EMBL-ID:A21640] >	
☐	Q3	link	1	SWISSPROT	1	([EMBL-ID:A21640] >	
☐	Q2	select	1	EMBL	1	[EMBL-ID:A21640]	
☐	Q1	query	65733	EMBL	65733	[embl-Description:ki	

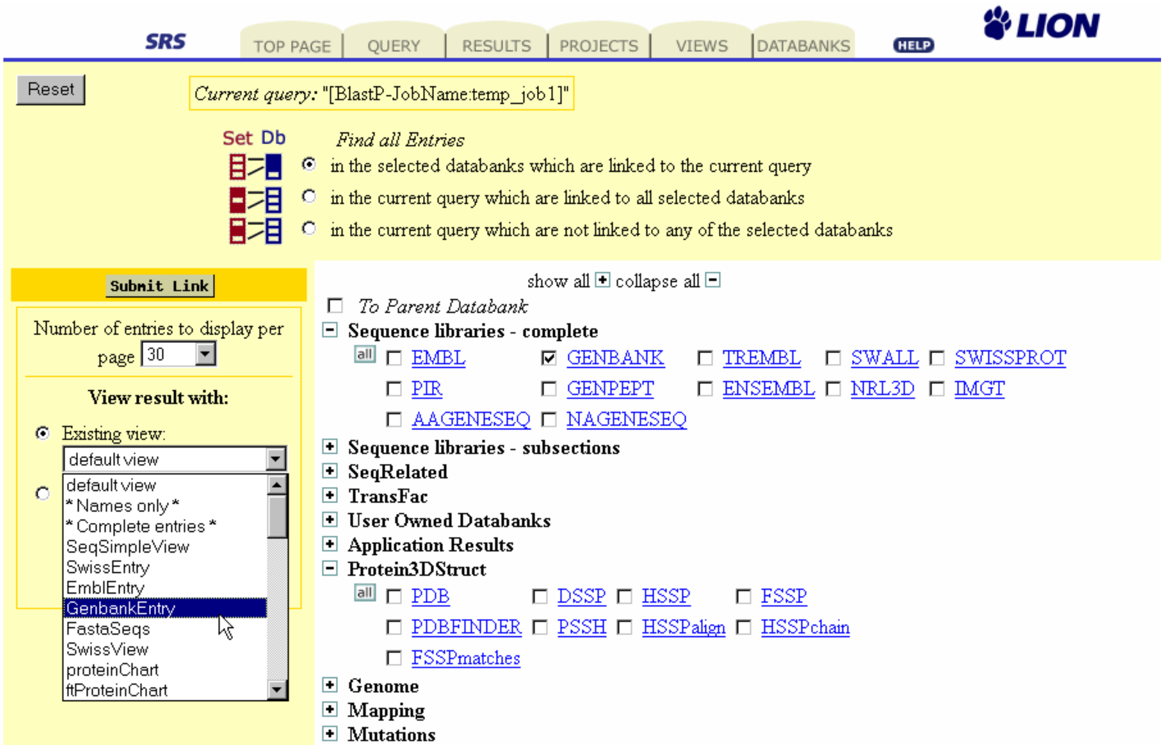
Figure 5.9 Select the view for your query.

3. Click the `View` button.

From the LINK Page

To apply a view for the results of a linking operation from the **LINK** page:

1. Click the **Results** tab.
   This will display the **Manage your Query Results** page.
2. Select a query for which you want to search for links by ticking the check box to the left of it.
3. Click the Link button.
   This will display the **LINK** page.
4. Choose the databank in which you wish to look for links by ticking the check box beside it (see Figure 5.10).
5. Choose the view that you want to apply to the results of the linking operation, using the drop-down list in the options area (under the **View result with** heading) and make sure that **Existing view** is selected, using the option button beside it.



Current query: "[BlastP-JobName:temp_job1]"

Find all Entries

● in the selected databanks which are linked to the current query
○ in the current query which are linked to all selected databanks
○ in the current query which are not linked to any of the selected databanks

Number of entries to display per page 30

View result with:

● Existing view: default view

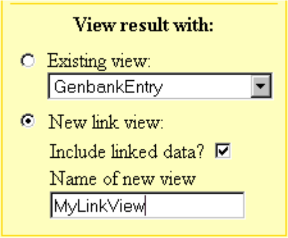
default view
Names only
Complete entries
SeqSimpleView
SwissEntry
EmblEntry
GenbankEntry
FastaSeqs
SwissView
proteinChart
ftProteinChart

☐ *To Parent Databank*
⊟ **Sequence libraries - complete**
☐ EMBL ☑ GENBANK ☐ TREMBL ☐ SWALL ☐ SWISSPROT
☐ PIR ☐ GENPEPT ☐ ENSEMBL ☐ NRL3D ☐ IMGT
☐ AAGENESEQ ☐ NAGENESEQ
⊞ **Sequence libraries - subsections**
⊞ **SeqRelated**
⊞ **TransFac**
⊞ **User Owned Databanks**
⊞ **Application Results**
⊟ **Protein3DStruct**
☐ PDB ☐ DSSP ☐ HSSP ☐ FSSP
☐ PDBFINDER ☐ PSSH ☐ HSSPalign ☐ HSSPchain
☐ FSSPmatches
⊞ **Genome**
⊞ **Mapping**
⊞ **Mutations**

Figure 5.10 Choosing a view for links.

6. Click the **Submit Link** button.

Alternatively, at step 5 you can choose to use a **New link view**. Complete steps 1 to 4, above, then select **New link view** using the option button beside that option, tick the check box to **Include linked data**, if required, and enter a name for your view.

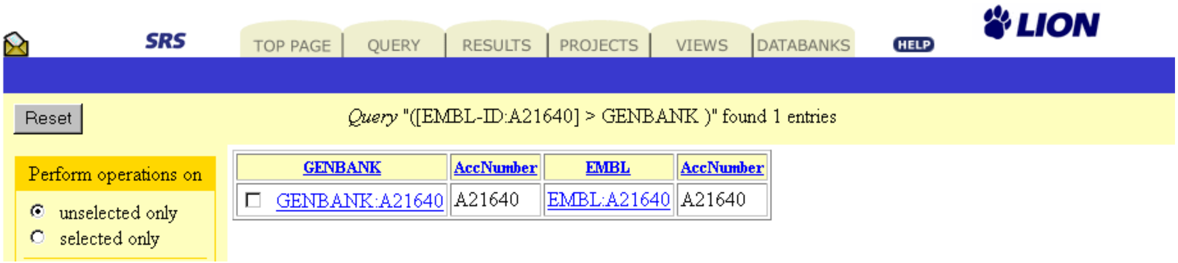


View result with:

○ Existing view:
GenbankEntry

● New link view:
Include linked data? ☑
Name of new view
MyLinkView

Figure 5.11 Setup to use a link view.



Query "([EMBL-ID:A21640] > GENBANK)" found 1 entries

Perform operations on
● unselected only
○ selected only

GENBANK	AccNumber	EMBL	AccNumber
☐ GENBANK:A21640	A21640	EMBL:A21640	A21640

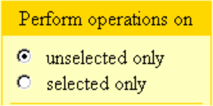
Figure 5.12 The entries that have links specified in your query, shown using a link view.

From the Launch Page for Interactive Applications

Note: For applications, run via a batch queue, the option to set a view is not available on the **Launch** page.

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

1. From the **Query Result** page choose the entries against which you want to run the application, or choose **unselected only**, using the option buttons.



Perform operations on

● unselected only

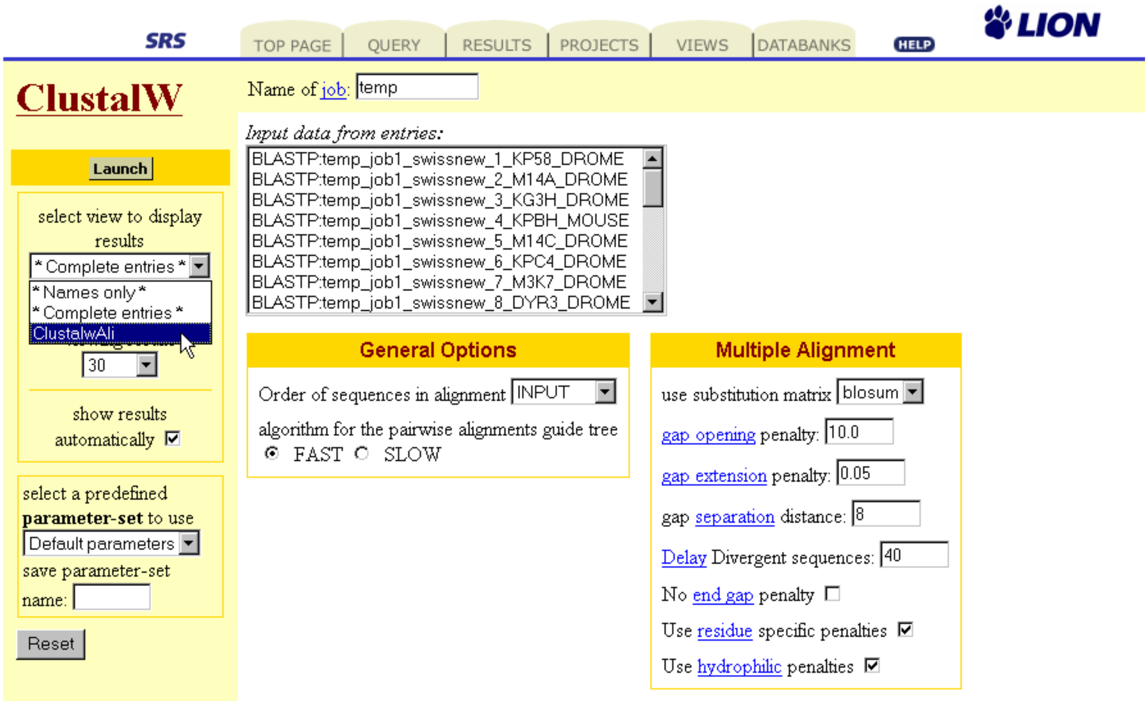
○ selected only

Figure 5.13 Option buttons to choose **unselected only**.

2. Select an application from the menu below the Launch button.

   Click the Launch button to display the **Launch** page.
   See chapter 6, “Applications”, (p. 131) for more information on launching applications.

3. You will see a drop-down list near the heading **select view to display results**. Using this drop-down list, scroll through the list until you find the view you want (see Figure 5.14).



Name of job: temp

Input data from entries:

BLASTP:temp_job1_swissnew_1_KP58_DROME
BLASTP:temp_job1_swissnew_2_M14A_DROME
BLASTP:temp_job1_swissnew_3_KG3H_DROME
BLASTP:temp_job1_swissnew_4_KPBH_MOUSE
BLASTP:temp_job1_swissnew_5_M14C_DROME
BLASTP:temp_job1_swissnew_6_KPC4_DROME
BLASTP:temp_job1_swissnew_7_M3K7_DROME
BLASTP:temp_job1_swissnew_8_DYR3_DROME

select view to display results: * Complete entries *
* Names only *
* Complete entries *
ClustalwAli

30

show results automatically ☑

select a predefined **parameter-set** to use: Default parameters

save parameter-set name:

Order of sequences in alignment: INPUT

algorithm for the pairwise alignments guide tree: ● FAST ○ SLOW

use substitution matrix: blosum

gap opening penalty: 10.0

gap extension penalty: 0.05

gap separation distance: 8

Delay Divergent sequences: 40

No end gap penalty ☐

Use residue specific penalties ☑

Use hydrophilic penalties ☑

Figure 5.14 Choosing a view for the application results.

4. Launch the application as usual (see section 1.7 “Using Applications”, p. 20). The results of the application will be displayed on the **Query Result** page using the selected view.

5.3 Creating Views

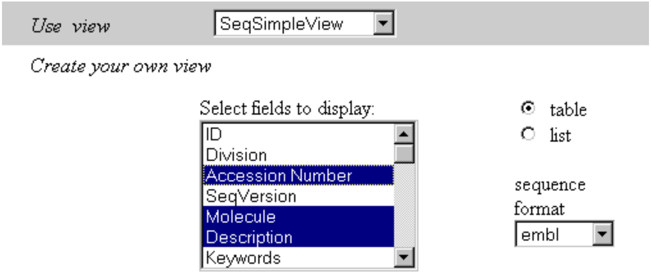
Sometimes the predefined SRS views may not show the information that you want to see. To allow you to view the particular information you require, you can create your own views using the **View Manager** page or from the **Standard Query Form** or **Extended Query Form**. These are referred to as “project views”.

5.3.1 Creating Views from the Query Forms

Both the **Standard Query Form** and the **Extended Query Form** provide a mechanism for specifying the datafields to be displayed on the **Query Result** page that results from the current query. This section describes how this is done.

Creating Views from the Standard Query Form

The bottom half of the **Standard Query Form** has a section entitled **Create your own view** which allows you to specify the datafields that are to be used to display the results of your query (see Figure 5.15). You can also choose whether to display the results using a table or a list view, and the sequence format that is used.



Use view SeqSimpleView

Create your own view

Select fields to display:
ID
Division
Accession Number
SeqVersion
Molecule
Description
Keywords

● table
○ list

sequence format embl

Figure 5.15 Specifying the datafields to be used to display results on the **Standard Query Form**.

To design your own view:

1. Select the datafields you want to display from the **Select fields to display** drop-down list (see Figure 5.15).

Note: You can usually use the **Shift** or **Control** keys to select multiple fields, but refer to your browser's documentation for help, if required.

---

2. Choose whether you wish to display the results using a table or a list view, using the option buttons.

● table
○ list

Figure 5.16 Table and list view option buttons.

3. Choose the sequence format using the drop-down list.

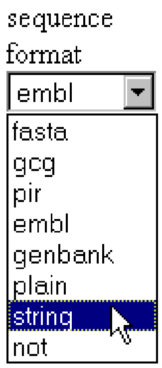
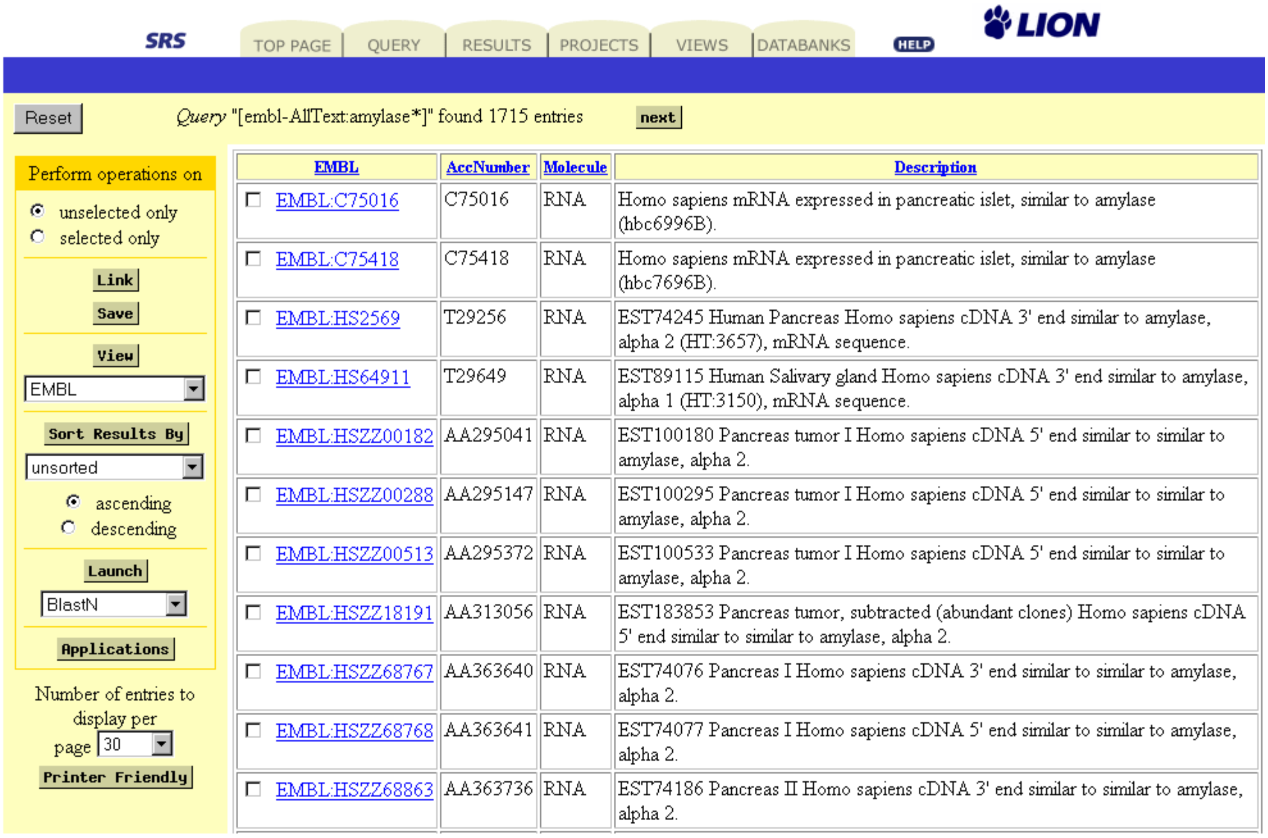


Figure 5.17 Sequence format drop-down list.

4. Complete the query as normal.
   The results will be displayed on the **Query Result** page as specified.



Query "[embl-AllText:amylase*]" found 1715 entries

Perform operations on
● unselected only
○ selected only

View: EMBL

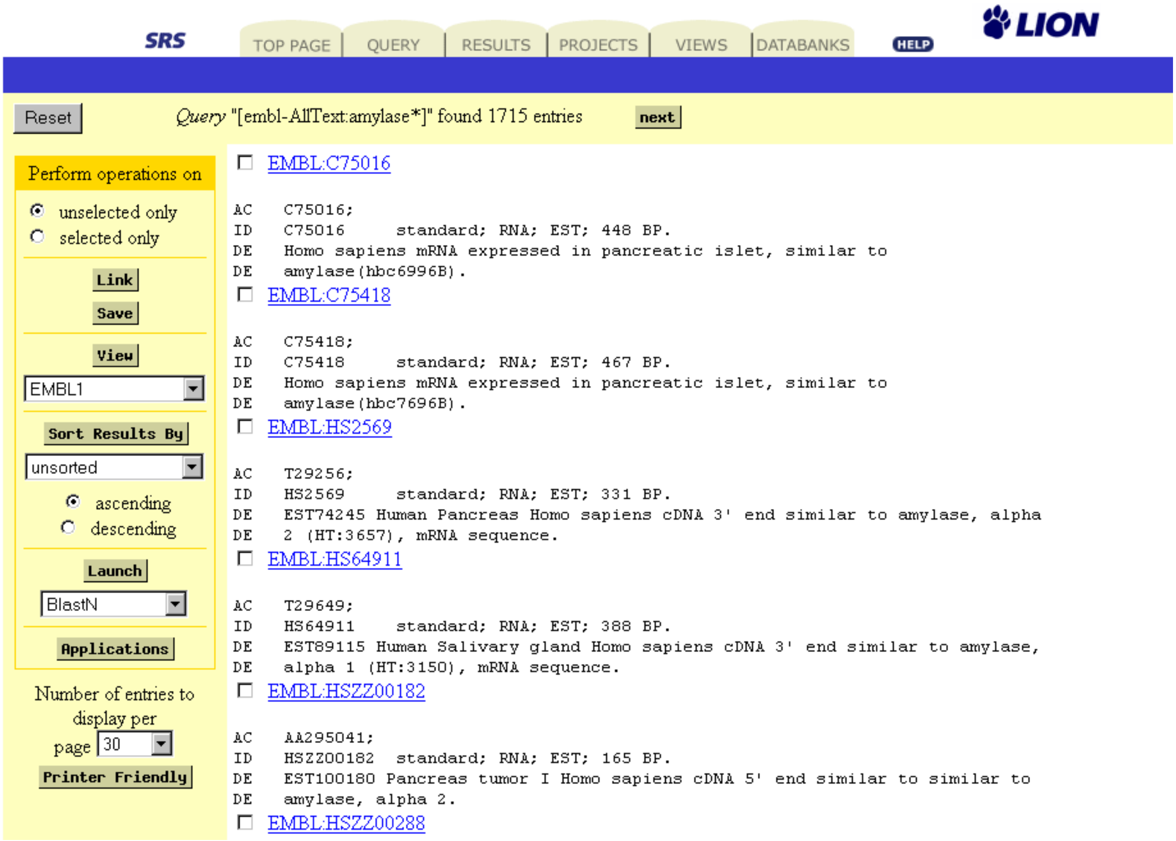
Sort Results By: unsorted
● ascending
○ descending

Launch: BlastN

Number of entries to display per page 30

	EMBL	AccNumber	Molecule	Description
☐	EMBL:C75016	C75016	RNA	Homo sapiens mRNA expressed in pancreatic islet, similar to amylase (hbc6996B).
☐	EMBL:C75418	C75418	RNA	Homo sapiens mRNA expressed in pancreatic islet, similar to amylase (hbc7696B).
☐	EMBL:HS2569	T29256	RNA	EST74245 Human Pancreas Homo sapiens cDNA 3' end similar to amylase, alpha 2 (HT:3657), mRNA sequence.
☐	EMBL:HS64911	T29649	RNA	EST89115 Human Salivary gland Homo sapiens cDNA 3' end similar to amylase, alpha 1 (HT:3150), mRNA sequence.
☐	EMBL:HSZZ00182	AA295041	RNA	EST100180 Pancreas tumor I Homo sapiens cDNA 5' end similar to similar to amylase, alpha 2.
☐	EMBL:HSZZ00288	AA295147	RNA	EST100295 Pancreas tumor I Homo sapiens cDNA 5' end similar to similar to amylase, alpha 2.
☐	EMBL:HSZZ00513	AA295372	RNA	EST100533 Pancreas tumor I Homo sapiens cDNA 5' end similar to similar to amylase, alpha 2.
☐	EMBL:HSZZ18191	AA313056	RNA	EST183853 Pancreas tumor, subtracted (abundant clones) Homo sapiens cDNA 5' end similar to similar to amylase, alpha 2.
☐	EMBL:HSZZ68767	AA363640	RNA	EST74076 Pancreas I Homo sapiens cDNA 3' end similar to similar to amylase, alpha 2.
☐	EMBL:HSZZ68768	AA363641	RNA	EST74077 Pancreas I Homo sapiens cDNA 5' end similar to similar to amylase, alpha 2.
☐	EMBL:HSZZ68863	AA363736	RNA	EST74186 Pancreas II Homo sapiens cDNA 3' end similar to similar to amylase, alpha 2.

Figure 5.18 Query Result page, showing a table view.



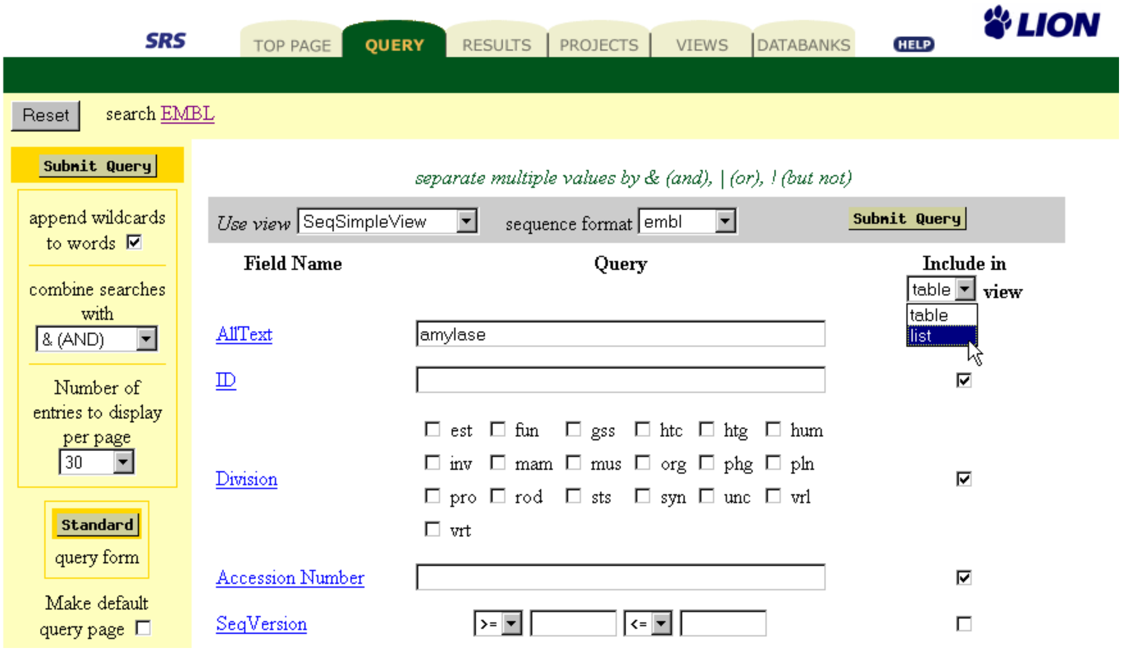
Query "[embl-AllText:amylase*]" found 1715 entries

```
☐ EMBL:C75016

AC   C75016;
ID   C75016     standard; RNA; EST; 448 BP.
DE   Homo sapiens mRNA expressed in pancreatic islet, similar to
DE   amylase(hbc6996B).
☐ EMBL:C75418

AC   C75418;
ID   C75418     standard; RNA; EST; 467 BP.
DE   Homo sapiens mRNA expressed in pancreatic islet, similar to
DE   amylase(hbc7696B).
☐ EMBL:HS2569

AC   T29256;
ID   HS2569     standard; RNA; EST; 331 BP.
DE   EST74245 Human Pancreas Homo sapiens cDNA 3' end similar to amylase, alpha
DE   2 (HT:3657), mRNA sequence.
☐ EMBL:HS64911

AC   T29649;
ID   HS64911    standard; RNA; EST; 388 BP.
DE   EST89115 Human Salivary gland Homo sapiens cDNA 3' end similar to amylase,
DE   alpha 1 (HT:3150), mRNA sequence.
☐ EMBL:HSZZ00182

AC   AA295041;
ID   HSZZ00182  standard; RNA; EST; 165 BP.
DE   EST100180 Pancreas tumor I Homo sapiens cDNA 5' end similar to similar to
DE   amylase, alpha 2.
☐ EMBL:HSZZ00288
```

Figure 5.19 Query Result page, showing a list view.

Creating Views from the Extended Query Form

On the right of the Data Area is a column entitled **Include in ... View** (see Figure 5.20). This contains check boxes (one for each datafield) which are used to tell SRS to include a datafield in the results. You do not have to enter a search term in a field to have it included in the results. The drop-down list at the top of this column allows you to choose whether to create a table or list view.



append wildcards to words ☑

combine searches with & (AND)

Number of entries to display per page 30

Make default query page ☐

Use view SeqSimpleView

sequence format embl

Include in table view (table / list)

AllText: amylase

ID ☑

Division: ☐ est ☐ fun ☐ gss ☐ htc ☐ htg ☐ hum ☐ inv ☐ mam ☐ mus ☐ org ☐ phg ☐ pln ☐ pro ☐ rod ☐ sts ☐ syn ☐ unc ☐ vrl ☐ vrt ☑

Accession Number ☑

SeqVersion >= <= ☐

Figure 5.20 Select the fields to include on the **Extended Query Form**.

To create a view from the **Extended Query Form**:

1. Enter your search term(s) as usual.
   See chapter 3, "Querying with SRS", (p. 51), for details.
2. Use the drop-down list to choose whether the view will display as a table or list view.
3. Tick the check box to the right of any fields that you want to display in the results.
4. Continue with your query as usual.

Typical results for table and list views are shown in Figures 5.18 and 5.19 respectively.

5.3.2 Creating Views using the View Manager Pages

If the built-in views do not do what you need and the options on the **Query Forms** are insufficient, you can use the **View Manager** pages to create views.

Note: There are two **View Manager** pages, referred to here as **View Manager** page 1 and **View Manager** page 2.

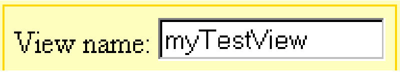
Click the **Views** tab to go to the **View Manager** pages. You can control the following elements of any views that you create:

- The name of the view (**View Manager** page 1).
- Whether the view displays as a table or list format (**View Manager** page 1).
- Whether the fields show on **View Manager** page 2 will include all available fields, or just the common ones (**View Manager** page 1).
- The list of databanks for which the view is defined (and for which the view is therefore available). These are referred to as “root databanks” (**View Manager** page 1).
- The list of databanks that may have a link with the entries (**View Manager** page 1).
- The list of fields from the root databanks and the linked databanks that should be displayed (**View Manager** page 2).
- The format of some fields such as **Sequences** (**View Manager** page 2).

In selecting the root databanks for a view, you specify two important things. Firstly, the view will only be available for queries and result sets based on these databanks. Secondly, the available fields from the root databanks will be those that are common across all of the root databank selections. Therefore, it is sensible to select databanks that have similar fields when you are creating a new view.

To Create a View

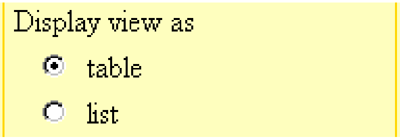
1. Click on the **Views** tab. This will take you to the **View Manager** page 1.
2. Name the view by entering a suitable name in the **View name** text box.



View name: myTestView

Figure 5.21 Naming a view.

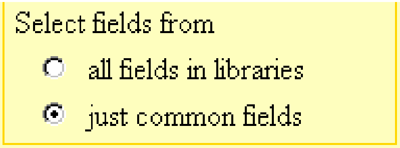
3. Choose whether the new view will be displayed using a table or list format.



Display view as
● table
○ list

Figure 5.22 Table and list view option buttons.

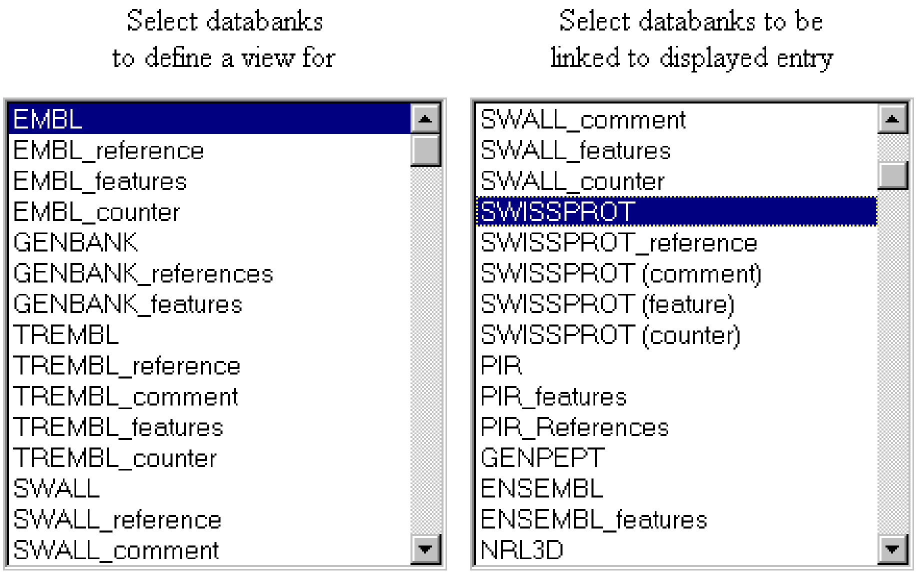
4. Choose whether you wish to choose from all available fields in the selected databanks, or just the common ones.



Select fields from
○ all fields in libraries
● just common fields

Figure 5.23 Selecting fields.

5. Select the databanks for which the view will be available, using the left-hand column of databanks, headed **Select databanks to define a view for**. These are the root databanks for the view.
6. Select the databanks to which you wish to link, using the right-hand column of databanks, headed **Select databanks to be linked to displayed entry**.



Select databanks to define a view for

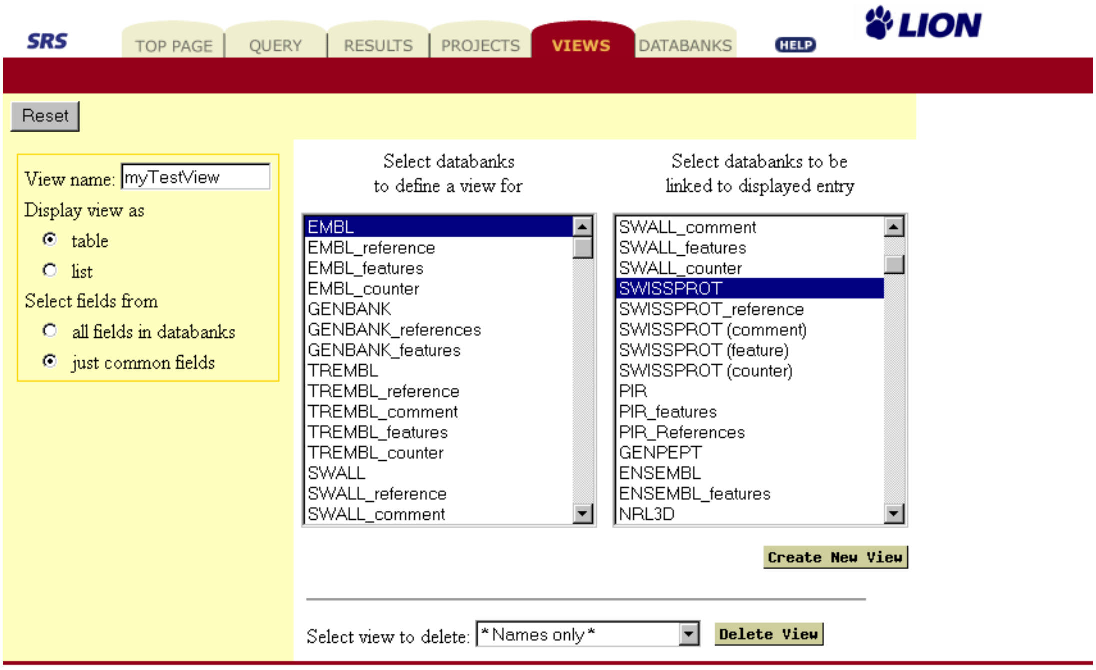
- EMBL
- EMBL_reference
- EMBL_features
- EMBL_counter
- GENBANK
- GENBANK_references
- GENBANK_features
- TREMBL
- TREMBL_reference
- TREMBL_comment
- TREMBL_features
- TREMBL_counter
- SWALL
- SWALL_reference
- SWALL_comment

Select databanks to be linked to displayed entry

- SWALL_comment
- SWALL_features
- SWALL_counter
- SWISSPROT
- SWISSPROT_reference
- SWISSPROT (comment)
- SWISSPROT (feature)
- SWISSPROT (counter)
- PIR
- PIR_features
- PIR_References
- GENPEPT
- ENSEMBL
- ENSEMBL_features
- NRL3D

Figure 5.24 Selected databanks.

7. Click on the **Create New View** button (see Figure 5.25).



View name: myTestView

Display view as

● table

○ list

Select fields from

○ all fields in databanks

● just common fields

Select databanks to define a view for

EMBL
EMBL_reference
EMBL_features
EMBL_counter
GENBANK
GENBANK_references
GENBANK_features
TREMBL
TREMBL_reference
TREMBL_comment
TREMBL_features
TREMBL_counter
SWALL
SWALL_reference
SWALL_comment

Select databanks to be linked to displayed entry

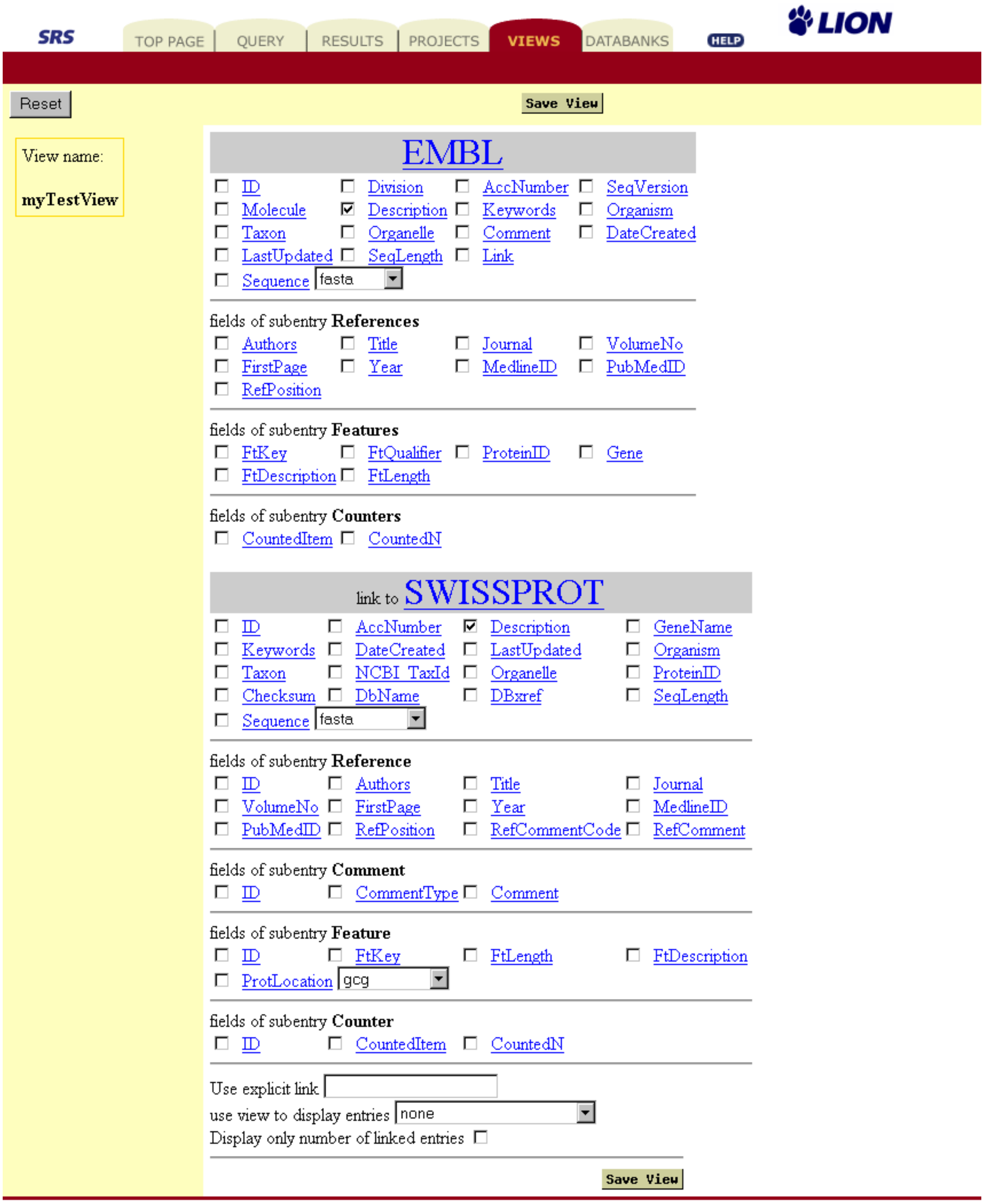
SWALL_comment
SWALL_features
SWALL_counter
SWISSPROT
SWISSPROT_reference
SWISSPROT (comment)
SWISSPROT (feature)
SWISSPROT (counter)
PIR
PIR_features
PIR_References
GENPEPT
ENSEMBL
ENSEMBL_features
NRL3D

Select view to delete: *Names only*

Figure 5.25 View Manager page 1.

SRS will display the **View Manager** page 2 on which you can select the specific fields for your view. By default (it can be changed on **View Manager** page 1) the top box contains the fields which are common to all of the root databanks selected on the previous page. There are additional boxes below this for each linked databank you chose. Using these check boxes, select the fields that you would like to see in your view.

8. Select the fields you would like to display in your view by ticking the check boxes beside the relevant fields.



View name: **myTestView**

EMBL

☐ ID ☐ Division ☐ AccNumber ☐ SeqVersion
☐ Molecule ☑ Description ☐ Keywords ☐ Organism
☐ Taxon ☐ Organelle ☐ Comment ☐ DateCreated
☐ LastUpdated ☐ SeqLength ☐ Link
☐ Sequence fasta

fields of subentry **References**
☐ Authors ☐ Title ☐ Journal ☐ VolumeNo
☐ FirstPage ☐ Year ☐ MedlineID ☐ PubMedID
☐ RefPosition

fields of subentry **Features**
☐ FtKey ☐ FtQualifier ☐ ProteinID ☐ Gene
☐ FtDescription ☐ FtLength

fields of subentry **Counters**
☐ CountedItem ☐ CountedN

link to SWISSPROT

☐ ID ☐ AccNumber ☑ Description ☐ GeneName
☐ Keywords ☐ DateCreated ☐ LastUpdated ☐ Organism
☐ Taxon ☐ NCBI_TaxId ☐ Organelle ☐ ProteinID
☐ Checksum ☐ DbName ☐ DBxref ☐ SeqLength
☐ Sequence fasta

fields of subentry **Reference**
☐ ID ☐ Authors ☐ Title ☐ Journal
☐ VolumeNo ☐ FirstPage ☐ Year ☐ MedlineID
☐ PubMedID ☐ RefPosition ☐ RefCommentCode ☐ RefComment

fields of subentry **Comment**
☐ ID ☐ CommentType ☐ Comment

fields of subentry **Feature**
☐ ID ☐ FtKey ☐ FtLength ☐ FtDescription
☐ ProtLocation gcg

fields of subentry **Counter**
☐ ID ☐ CountedItem ☐ CountedN

Use explicit link
use view to display entries none
Display only number of linked entries ☐

Figure 5.26 Select the datafields to be included in your view.

Note: Each field name is a hypertext link to its **Field Information** page (see section 7.3 “Field Information page for the Description field.”, p. 106). So if you want to know more about a particular field, use these hypertext links to find out.

---

9. Click on the Save View button.

Your new view can now be used for any of the databanks indicated in the root databanks (see step 5, above). Try your new view for a query on one or more of the root databanks.

CHAPTER

6

APPLICATIONS

Various applications can be run on queries within SRS. This chapter describes how to:

- Use the applications that are available within SRS.

6.1 Introducing Applications

SRS applications are bioinformatics programs that use a databank query as input. The output file from the application is indexed in the same way as any other databank. This enables users to store and query their application results.

The results from one application can be used by another. Hits in a sequence databank from a sequence similarity search (e.g., one of the BLAST programs) can be selected and aligned. The resulting alignment can then be used for further analysis.

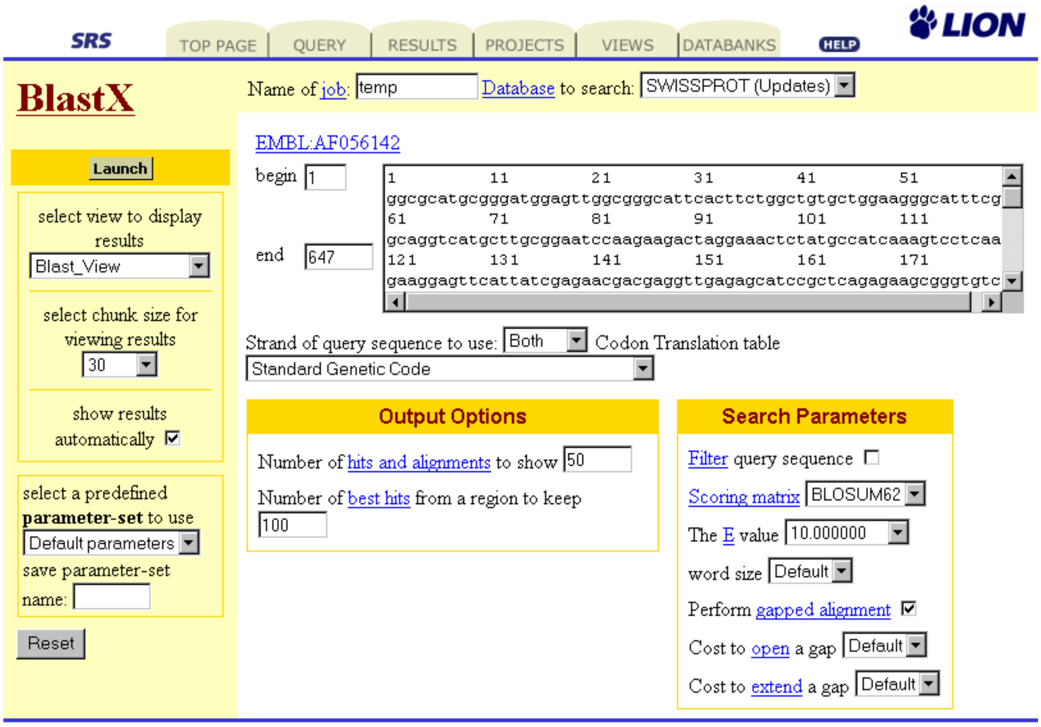
6.2 Accessing Applications

All application runs start with a databank query. The **Query Result** page includes a drop-down list (below the Launch button) which lists the most commonly used available applications. This can be used with the Launch button to access the applications in the drop-down list. Alternatively, the Applications button will take you to the **Application Select** page. This lists all the applications that can use the query data as input. You can also get to the **Application Select** page from the **Select Databanks To Search** page, using the Applications button there. This page will then list all of the applications that are available on your SRS installation. However, you will have to enter your own sequence data on the **Launch** page.

Section 1.7 “Using Applications” (p. 20) explains how to start an application. The remainder of this chapter gives further information on the use of applications, and further details of the differing behavior of interactive and batch applications.

6.3 Launch Page

Application Parameters



Name of job: temp
Database to search: SWISSPROT (Updates)

EMBL:AF056142

begin 1
end 647

```
1         11        21        31        41        51
ggcgcatgcgggatggagttggcgggcattcacttctggctgtgctggaagggcatttcg
61        71        81        91        101       111
gcaggtcatgcttgcggaatccaagaagactaggaaactctatgccatcaaagtcctcaa
121       131       141       151       161       171
gaaggagttcattatcgagaacgacgaggttgagagcatccgctcagagaagcgggtgtc
```

Strand of query sequence to use: Both
Codon Translation table: Standard Genetic Code

select view to display results: Blast_View
select chunk size for viewing results: 30
show results automatically ☑
select a predefined **parameter-set** to use: Default parameters
save parameter-set name:

Output Options
Number of hits and alignments to show: 50
Number of best hits from a region to keep: 100

Search Parameters
Filter query sequence ☐
Scoring matrix: BLOSUM62
The E value: 10.000000
word size: Default
Perform gapped alignment ☑
Cost to open a gap: Default
Cost to extend a gap: Default

Figure 6.1 Launch page for an application which will be run interactively (see also section 6.4 “Interactive and Batch Applications”, p. 134).

In many cases the values on the **Launch** page will not need to be changed. For sequence similarity searches, there is usually a choice of databanks to be searched. Other values can be changed freely, and SRS will display an error message if an invalid value is given.

The full sequence is displayed (providing the application has been accessed from a valid query, rather than direct from the **Select Databanks To Search** page). This means that you can edit it, usually by changing the begin or end position. Editing the contents of the sequence itself is also possible.

6.4 Interactive and Batch Applications

6.4.1 Introduction

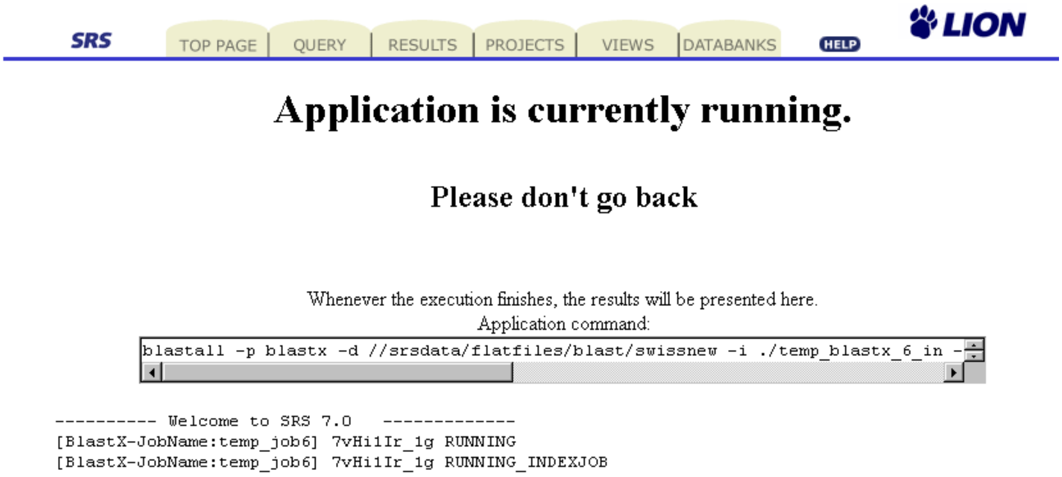
When you click on the Launch button, the application will either be run interactively (you wait for the results to be displayed) or submitted to a batch queue. The choice depends on how SRS has been set up.

The **Launch** page will indicate whether or not an application will be run interactively, or submitted to a batch queue. Figure 6.1 shows the **Launch** page for an interactive application. Figure 6.4 (below) shows a typical batch queue message from a **Launch** page for a batch queued application.

The next sections describe these methods (see sections 6.4.2 “Interactive Applications”, p. 134 and 6.4.3 “Batch Applications”, p. 136).

6.4.2 Interactive Applications

Applications that do not take long to run are usually run interactively. This means they will be started immediately. The **Application Invocation** page (Figure 6.2) will be displayed while the application is running.



Application is currently running.

Please don't go back

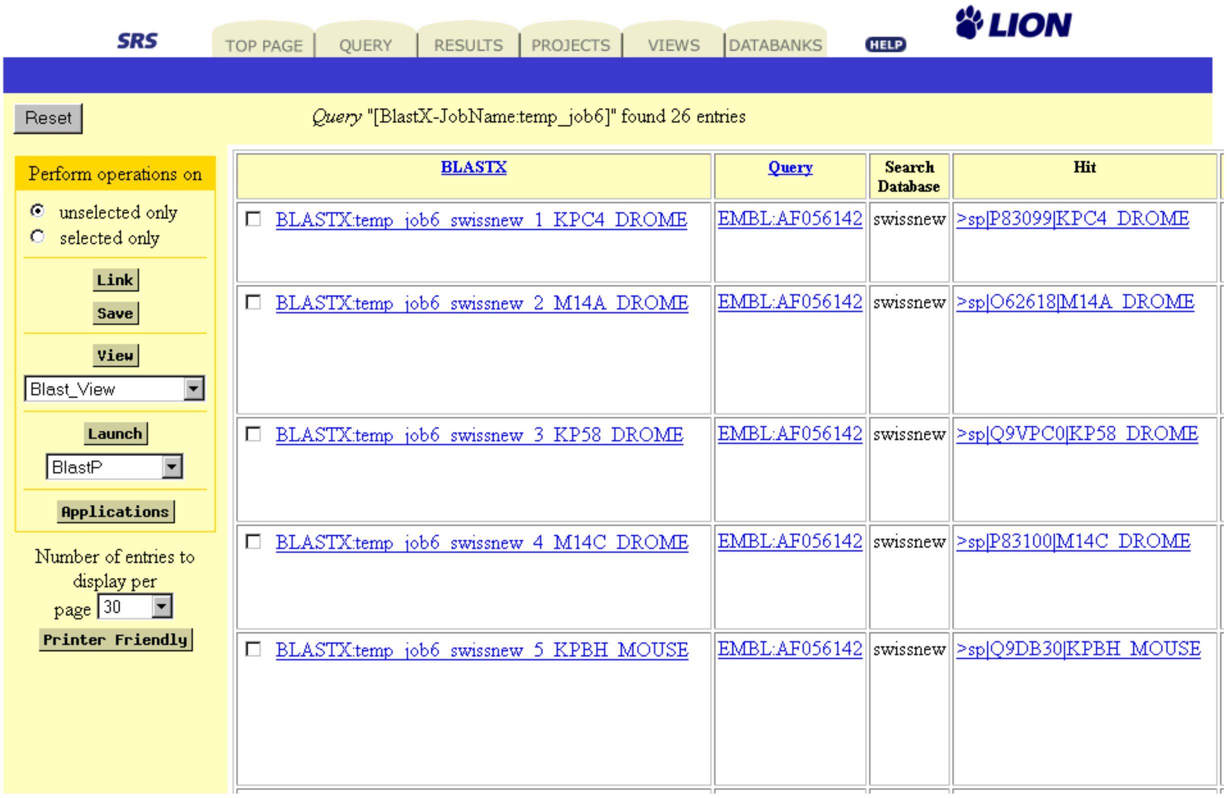
Whenever the execution finishes, the results will be presented here.
Application command:

```
blastall -p blastx -d //srsdata/flatfiles/blast/swissnew -i ./temp_blastx_6_in -
```

```
---------- Welcome to SRS 7.0   -------------
[BlastX-JobName:temp_job6] 7vHi1Ir_1g RUNNING
[BlastX-JobName:temp_job6] 7vHi1Ir_1g RUNNING_INDEXJOB
```

Figure 6.2 Application Invocation page.

The **Query Result** page (Figure 6.3) will appear automatically on completion unless the **show results automatically** check box has been unticked.



Query "[BlastX-JobName:temp_job6]" found 26 entries

Perform operations on
● unselected only
○ selected only

Blast_View

BlastP

Number of entries to display per page 30

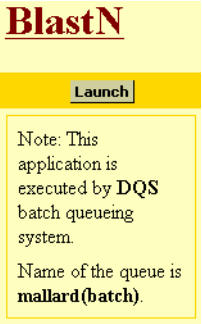
BLASTX	Query	Search Database	Hit
☐ BLASTX:temp_job6_swissnew_1_KPC4_DROME	EMBL:AF056142	swissnew	>sp|P83099|KPC4_DROME
☐ BLASTX:temp_job6_swissnew_2_M14A_DROME	EMBL:AF056142	swissnew	>sp|O62618|M14A_DROME
☐ BLASTX:temp_job6_swissnew_3_KP58_DROME	EMBL:AF056142	swissnew	>sp|Q9VPC0|KP58_DROME
☐ BLASTX:temp_job6_swissnew_4_M14C_DROME	EMBL:AF056142	swissnew	>sp|P83100|M14C_DROME
☐ BLASTX:temp_job6_swissnew_5_KPBH_MOUSE	EMBL:AF056142	swissnew	>sp|Q9DB30|KPBH_MOUSE

Figure 6.3 **Query Result** page showing application results for BLASTX.

6.4.3 Batch Applications

Introduction to Batch Applications

Applications which are likely to take a while to run are usually set to run as batch jobs. SRS does not wait for batch jobs. If a job will be run from the batch queue this is indicated prior to launching (Figure 6.4).



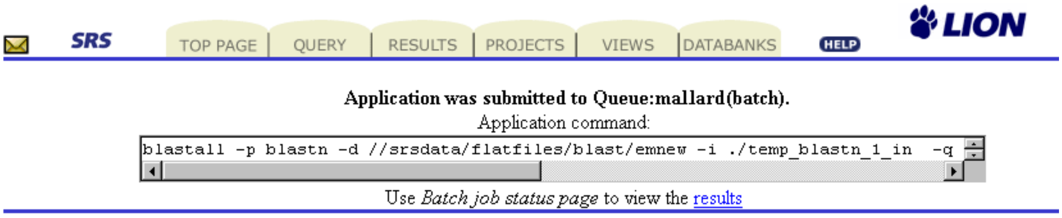
Note: This application is executed by **DQS** batch queueing system.

Name of the queue is **mallard(batch)**.

Figure 6.4 SRS batch queuing message for BLASTN.

Application Invocation Page for Batched Applications

After the **Launch** button is clicked, the **Application Invocation** page will be shown.



Application was submitted to Queue:mallard(batch).

Application command:

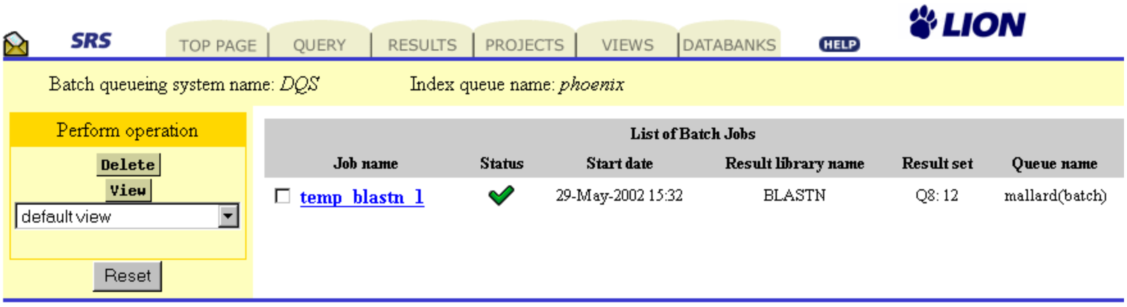
```
blastall -p blastn -d //srsdata/flatfiles/blast/emnew -i ./temp_blastn_1_in  -q
```

Use *Batch job status page* to view the results

Figure 6.5 **Application Invocation** page for batch processing.

Accessing the Results

The results from an application can be reached from the **Job Status** page (see Figure 6.6). This page is reached by clicking on the results hyperlink or on the ✉ icon on the **Application Invocation** page.



Batch queueing system name: *DQS* Index queue name: *phoenix*

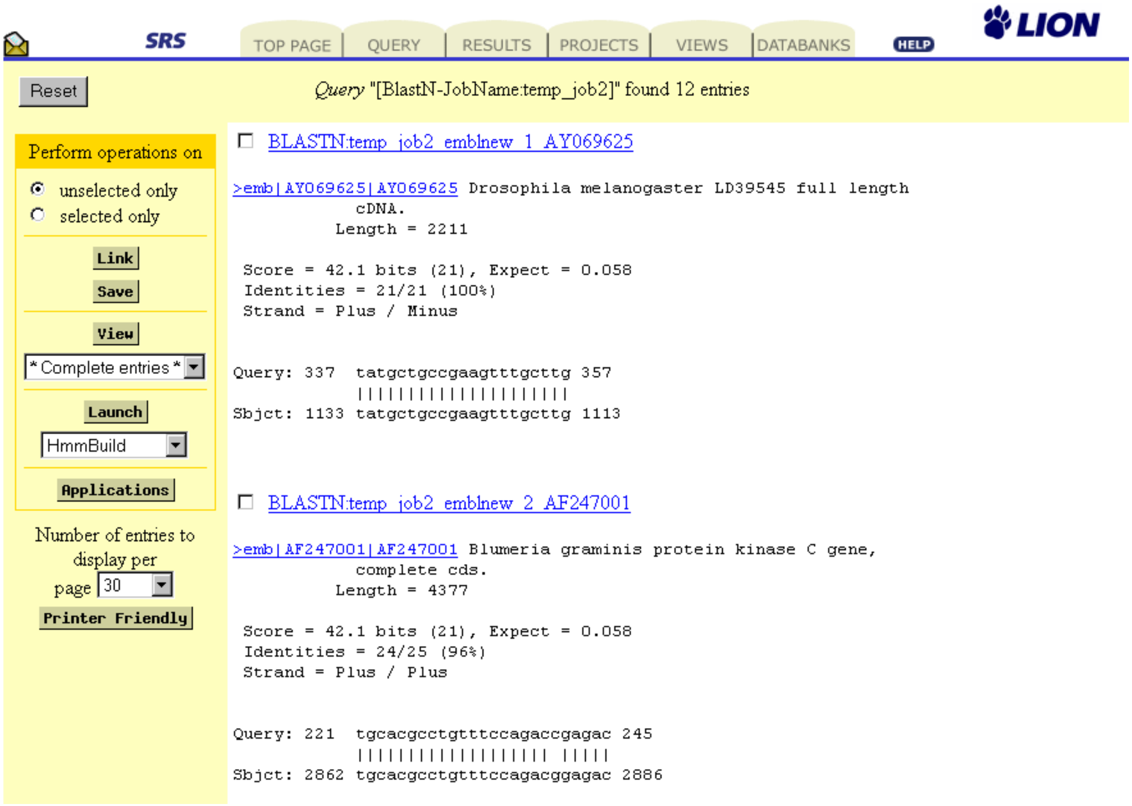
Perform operation

List of Batch Jobs

	Job name	Status	Start date	Result library name	Result set	Queue name
☐	temp_blastn_1	✓	29-May-2002 15:32	BLASTN	Q8: 12	mallard(batch)

Figure 6.6 Job Status page.

When a job has completed, the job name becomes a link to the application results, and the closed ✉ becomes an open ✉ . Clicking on the hyperlink for the job that has just run, e.g. temp_blastn_1 will display the **Entry** page containing the results.



Query "[BlastN-JobName:temp_job2]" found 12 entries

☐ BLASTN:temp_job2_emblnew_1_AY069625

```
>emb|AY069625|AY069625 Drosophila melanogaster LD39545 full length
            cDNA.
          Length = 2211

 Score = 42.1 bits (21), Expect = 0.058
 Identities = 21/21 (100%)
 Strand = Plus / Minus


Query: 337  tatgctgccgaagtttgcttg 357
            |||||||||||||||||||||
Sbjct: 1133 tatgctgccgaagtttgcttg 1113
```

☐ BLASTN:temp_job2_emblnew_2_AF247001

```
>emb|AF247001|AF247001 Blumeria graminis protein kinase C gene,
            complete cds.
          Length = 4377

 Score = 42.1 bits (21), Expect = 0.058
 Identities = 24/25 (96%)
 Strand = Plus / Plus


Query: 221  tgcacgcctgtttccagaccgagac 245
            ||||||||||||||||||| |||||
Sbjct: 2862 tgcacgcctgtttccagacggagac 2886
```

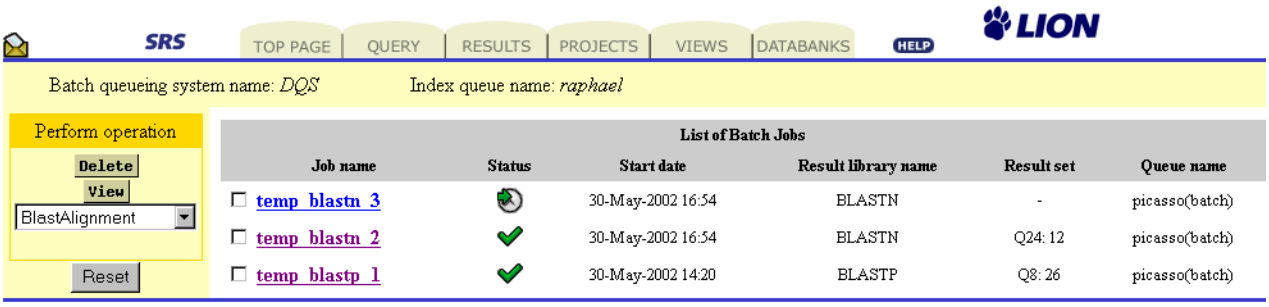
Figure 6.7 Entry page showing open ✉.

Alternatively, it is possible to select a job by ticking the check box beside it, selecting a view and clicking the **View** button. This will display the results on a **Query Result** page.

Job Status

The **Job Status** page (see Figure 6.8) lists all batch jobs in the current project and shows their current status. (The most recent job is at the top of the list.) Completed jobs are shown with a ✓ icon.

Jobs still running are shown with a ⟳ icon. Stopped jobs (i.e. those which are waiting to be run) are shown with a 🚦 icon. Jobs which have produced no meaningful results (e.g. those for which the search sequence produced no appropriate results or those which have crashed) are shown with a ⚠ icon.



Batch queueing system name: *DQS*   Index queue name: *raphael*

List of Batch Jobs

	Job name	Status	Start date	Result library name	Result set	Queue name
☐	temp_blastn_3	⟳	30-May-2002 16:54	BLASTN	-	picasso(batch)
☐	temp_blastn_2	✓	30-May-2002 16:54	BLASTN	Q24: 12	picasso(batch)
☐	temp_blastp_1	✓	30-May-2002 14:20	BLASTP	Q8: 26	picasso(batch)

Figure 6.8 Job Status page showing multiple jobs.

Accessing the Job Status Page

The ✉ icon appears on SRS pages when batch jobs exist,and can be used to access the **Job Status** page. It is closed, ✉, when there is a job waiting to be run and/or running. It changes to open, 📨, when all jobs have completed.

CHAPTER

7

COMMAND LINE SRS

The SRS command line interface is called getz. This chapter:

- Introduces getz.
- Explains the available getz options.

7.1 getz

The SRS command line is available through a UNIX shell window, and uses a program called getz. You may need to contact your SRS Administrator to set this up.

Using getz you can query databanks from the command line. For further information on creating getz queries, see section 7.1.1 “getz Usage” (p. 143) and chapter 8, “SRS Query Language”, (p. 145).

Here are a few examples:

To retrieve a list of entries from the SWISS-PROT databank that have “azurin” mentioned in the description, you might enter:

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

If you want to perform the same search as above, but retrieve the complete entries, use the `-e` argument:

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

To retrieve the sequence for the entries using the FASTA format, type:

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

Section 7.1.1 “getz Usage”, p. 143 explains more about the arguments.

7.1.1 getz Usage

The following table lists the options available for command line SRS. These options can be used to refine your getz queries. Alternatively, to find out more about getz, use the command:

```
getz -help
```

Table 7.1 Common getz options.

Option	Default	Function
-help		Help with getz.
-e	FALSE	Prints entire entry.
-f <string>	""	Include fields in entry list.
-vf <string>	""	List of fields that will be placed into a table view.
-w	FALSE	Appends a wildcard to each search word.
-lb <n>	0	Number of first entry in set to be viewed.
-ll <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 that match the query, plus the number of entries for each match.
-lmin <n>	0	List only values that occur at least the specified number of times. Use together with `-lv`.

Table 7.1 Common getz options. (Continued)

Option	Default	Function
-c	FALSE	Report the number of entries that were found, but not the entries themselves.
-info	FALSE	Prints info about the specified databank.
-libs	FALSE	Prints a list of all active databanks.
-view <string>	""	Name of view to be used when displaying entries.
-rs <string>	""	String of one or more characters to separate records in view.
-cs <string>	""	String of one or more characters to separate columns in view.
-sf <string>	""	Format of sequence output file.
-af <string>	""	Format of sequence alignment output file.
-html	FALSE	Select HTML format for output.
-off	FALSE	Accesses the off-line version of a databank.

CHAPTER

8

SRS QUERY LANGUAGE

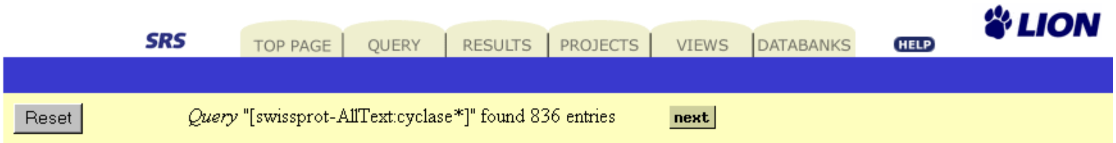
In SRS many actions, e.g., retrieval commands, logical operations with sets obtained from previous queries, links between entries from different databanks, or a combination, can be expressed using the SRS query language. This chapter describes the language and gives examples of its use. During this chapter you will learn:

- About searching in indices.
- The general syntax of the SRS query language.
- About using logical and link operators.
- About entries and subentries.
- About storing intermediate results in sets.

8.1 Introduction

In earlier chapters you have seen how to use the SRS **Query Forms**, **LINK** page, etc., to create a query. In addition to this method, you can type your query using the SRS query language. There are various places where this can be done, including, the **Quick Search** text box on the **Select Databanks To Search** page, for example.

The SRS query language syntax is also used internally by SRS whenever you make any query (e.g., using the query forms, linking, etc.), so you will see examples of them on SRS pages that contain the results of any such queries, (e.g., at the top of the **Query Result** page).



Query "[swissprot-AllText:cyclase*]" found 836 entries

Figure 8.1 Query Result page, showing a query of the **AllText** field of the SWISS-PROT databank, for “cyclase”.

In addition to simple searches of single databanks you can search multiple databanks, link searches, etc.

8.2 Searching in Indices

8.2.1 Introduction

Probably the simplest form of the SRS query language syntax is that used for simple searches in indices. Index searches include searches for simple strings, searches for numbers and ranges of numbers, as well as searches for dates. This section covers the various forms of index search.

8.2.2 General Syntax

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, either a colon (:), (string search, see section 8.2.3 “Search Strings”, p. 147), or a hash (#), (range search). Range searches can be performed only in indices of the types num and real (see section 8.2.4 “Searching Using Numerical Ranges”, p. 151 and section 8.2.5 “Searching for Dates”, p. 152).

Either the field name (e.g., `description`) or its abbreviation (`des`) can be used as the index name. All strings, including the search words, are case-insensitive. For example:

```
[pir-des:elastase]
```

searches for the string “elastase” in the `des` (`description`) field of the protein databank, PIR.

8.2.3 Search Strings

A search string may be a single search word, or several words separated by logical operator(s) (see section 8.3.4 “Operators”, p. 156). Parentheses may be used to create a group which will be treated as a single operand (see Example 8.1 Search strings).

Wildcards and regular expressions may also be used (see section “Wildcards”, p. 148 and section “Regular Expressions”, p. 149).

Example 8.1 Search strings

To search the `keywords` field of the EMBL databank for “insulin” you might enter:

```
[embl-key:insulin]
```

To search the `description` field of the EMBL databank for entries which include “acetylchol*” and “receptor”, but remove any entries that contain “muscarinic” you might enter:

```
[embl-des:(acetylchol*&receptor)!muscarinic]
```

To search the `authors` index field of the SWISS-PROT databank to look for entries containing “sanger,f*” but not “coulson,a*”, you might use a query like:

```
[swissprot-aut:sanger,f*!coulson,a*]
```

Wildcards

Wildcards are useful if, for example, you wish to search for a group of words (e.g., all words starting with “cell” and ending with “ase”) or if it is unclear how a word is spelt in a databank.

SRS uses two types of wildcard:

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

Any number of wildcards can be placed anywhere in a search word.

Note: Placing a wildcard at the start of a word or string may increase the response time because all words in the index have to be checked against your string.

Regular Expressions

In addition to the use of wildcards, it is also possible to enter regular expressions directly. Regular expressions must appear within forward slashes (/).

Some characters (^$.[]()*+?) have a special meaning — these must be prefixed with a backslash (\) to indicate that the specified character is to be matched literally.

Tables 8.1 and 8.2 respectively, give a list typical regular operands and examples of their use.

Table 8.1 Regular expression operators.

Operator	Meaning
`^`	A caret (`^`) is used to mark the start of a string. For example, `/^phos/` will find all words beginning with phos (e.g., phosphate).
`$`	A dollar sign (`$`) is used to mark the end of a string. For example, `/ase$` will find all words ending with “ase” (e.g., kinase).
`.`	A dot indicates any single character.

Table 8.1 Regular expression operators. (Continued)

Operator	Meaning
`[...]`	Characters enclosed in square brackets are regarded as a set, any of which can be matched. For instance, `[()]` matches an opening or closing parenthesis. Character ranges can be specified using a hyphen (`-`), e.g., `[0-9]` matches any single digit. A caret (`^`) in front of the character set (after the opening square bracket) negates the character set: e.g., `[^0-9]` matches any non-digit character.
`(...)`	A series of pattern elements enclosed in parentheses is regarded as comprising a single element.
`*`	An asterisk indicates that the preceding group may be repeated zero or more times.
`+`	A plus sign indicates that the preceding group may be repeated one or more times.
`?`	A question mark indicates that the preceding character or group of characters occurs one or zero times.

Table 8.2 Examples of regular expressions.

Expression	Meaning
`/^j..$/`	This expression finds all three-character strings that start with j.
`/^5[0-9][0-9][0-9]$/`	This expression finds all four-digit numbers that start with 5.
`/^nif[a-e]$/`	This expression finds the gene names nifa, nifb, nifc, nifd, nife.
`/^mue?ller$/`	This expression finds both muller and mueller.

Note: Searches with regular expressions are sometimes slow since all the words in the index have to be searched.

8.2.4 Searching Using Numerical Ranges

In a numerical index (whether it contains integers or reals) it is possible to search numerical ranges. A numerical index is only possible where there is a one-to-one relationship between entry and value (e.g., sequence length, creation date, resolution).

A range can be specified using a single value or by two values separated by a colon (:). The value on the left must be smaller than the value on the right. To exclude a value from the range, put an exclamation mark (!), in front of it. The absence of a number on the left indicates that the search should start at the minimum value in the index. Similarly, an absent value on the right indicates that the search should include values up to the maximum for that index.

Table 8.3 Examples of queries on an index of the sequence length.

Written Range	Meaning
400	All sequences with a length of exactly 400.
400:500	All sequences with lengths between 400 and 500.
400:	All sequences with lengths greater than 400.
:500	All sequences with lengths less than 500.
400:!500	All sequences with lengths between 400 and 500, excluding 500.
:	A range from the minimum value to the maximum value, i.e., all sequences.

Combining Ranges

Ranges can be combined using logical operators. For instance either:

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

or

```
300:700 ! 500:600
```

would retrieve the same set of sequences, i.e., all sequences from 300 to 500, excluding 500, and all sequences from 600 to 700, excluding 600.

8.2.5 Searching for Dates

Searches for dates can be made using one of the two special formats recognized by the SRS query language. These are: YYYYMMDD or DD-MMM-YYYY. For example,

```
20020619
19-Jun-2002
```

Dates can be used within ranges in the same way as other numbers. For example,

```
[swissprot-date#20010415:20020414]
[swissprot-date#15-APR-2001:14-APR-2002]
```

8.2.6 Searching Multiple Databanks

As well as allowing you to search a field of a single databank, the SRS query language allows you to search multiple databanks in a single query expression. This is done using a list of databank names, enclosed in curly brackets, to replace the single databank name seen in earlier examples. The names in the list must be separated by spaces. For example:

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

searches for the word, “kinase”, in the **Description** index of the SWISS-PROT, SWISSNEW and SPtrEMBL databanks.

It is often convenient to give a name to a group of databanks so that that name can be used later in the query rather than repeating the list of names. For instance:

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

creates the group, `dbs`, which combines the three databanks SWISS-PROT, SWISSNEW and SPtrEMBL, and then uses the group name, `dbs`, to replace the search name in the second part of the search.

8.3 Combining Searches

8.3.1 Introduction

The earlier parts of this chapter dealt with simple index searches. The SRS query language can also be used to create more complex queries. These take the form of expressions and are constructed using operands (see section 8.3.3 “Operands”, p. 155) and operators (see section 8.3.4 “Operators”, p. 156).

8.3.2 General Syntax

Queries using the SRS query language take the general form:

```
operand operator operand ...
```

(See section 8.3.3 “Operands”, p. 155 and section 8.3.4 “Operators”, p. 156 for more information on operands and operators, respectively.)

For example:

```
enzyme > pdb
```

where `enzyme` and `pdb` are operands specifying the databanks, Enzyme and PDB respectively, and > is an operator telling SRS to search for links between the databanks, and keep only those entries which belong to the databank on the right. The result of this query is a list of entries in the PDB databank that have links to the Enzyme databank.

Combinations may also include index searches. For instance:

```
[swissprot-des:kinase] > pdb
```

will create a list of all the entries in the PDB databank that have links to the results of the index search:

```
[swissprot-des:kinase]
```

See section 8.2 “Searching in Indices” (p. 146), for an explanation of this sub-search.

The above examples are fairly trivial, but the SRS query language allows you to build up more complex queries using the various operators. In addition expressions can be grouped using parentheses so that they are treated as a single entity. You will see examples of searches and ways of combining them, throughout this chapter.

8.3.3 Operands

Operands are the items upon which the expression is performed, e.g., the name of a databank, or a set, etc.

Table 8.4 Typical SRS operands.

Operand	**Example**	**Meaning**
Databank name	EMBL	Each databank has a unique name.
Set name	Q1	SRS gives each query a name which can be used when you want to perform an operation on the set of results from that query.
Index search	[embl-des: kinase]	This command initiates a search in one or more indices, of one or more databanks.

Table 8.4 Typical SRS operands. (Continued)

Operand	Example	Meaning
Expression	(Q1&Q2)	If an expression is enclosed in parentheses, it is treated as a single operand. Parentheses can be nested to any degree.
Parent	parent	This is a special operand that allows the conversion of a set of subentries into a set of entries (see section 8.4 "Entries and Subentries", p. 161).

8.3.4 Operators

Operators tell SRS what to do with the operands, e.g., search for links between two operands. Table 8.5 shows a list of available operators. For more information on the types of operator see section "Logical Operators" (p. 157) and section "Link Operators" (p. 158). For information on the use of operators, see section 8.3.5 "Use of Operators to Combine Search Items" (p. 159).

Table 8.5 SRS query language operators.

Operator	Type	Meaning
|	Logical	OR.
&	Logical	AND.
!	Logical	BUT NOT. This operator may need to be escaped in UNIX, using "/!".
>	Link	Link, keeping items in the set to the right of the >.

Table 8.5 SRS query language operators. (Continued)

Operator	Type	Meaning
<	Link	Link, keeping items in the set to the left of the >.
>^	Link	Get subtree defined by left operand (hierarchical links).
>_	Link	Get leaf entries of the subtree defined by left operand (hierarchical links).

Logical Operators

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

The following figure illustrates the effects of the three operators in an expression of the form, `A operator B`.

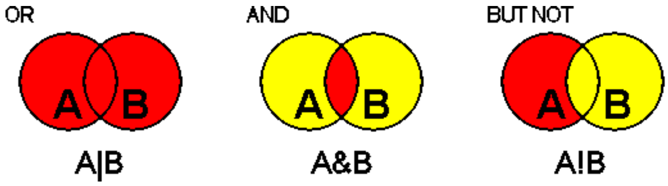


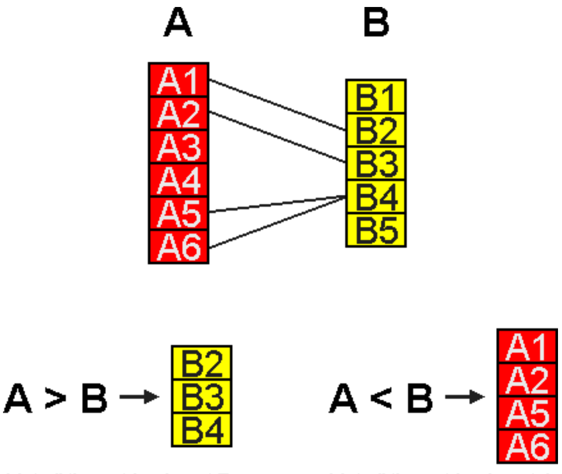
Figure 8.2 Logical operators in SRS.

Logical operations can only be performed between sets of the same type. It is not possible, for instance, to combine a set of entries and a set of subentries (see section 8.4 “Entries and Subentries”, p. 161) using logical operators. In such cases, an additional link operation must be specified (see section 8.4.1 “Links with Sets Containing Subentries”, p. 162).

Link Operators

Link operators are unique to the SRS query language. The two link operators, < and >, allow sets of data from different databanks to be combined.

Figure 8.3 shows two databanks, A and B, in which some entries in A have links to entries in B. These links are processed to build link indices that provide the basis for the link operation. The figure shows the results of two searches for links between sets A and B, using the operators, < and >.



List all the entries in set B that have links with set A.

List all the entries in set A that have links with set B.

Figure 8.3 Linked Data.

Links are not usually bidirectional; however, the link indices in SRS are used bidirectionally. For instance:

A > B

This retrieves those entries in B that are linked to entries in A.

A < B

This retrieves those entries in A that are linked to entries in B.

8.3.5 Use of Operators to Combine Search Items

Combining queries allows you to refine your search results. This can be done using logical operators (OR, AND, BUTNOT, see section “Logical Operators”, p. 157) or link operators (see section “Link Operators”, p. 158). Note that link operators take precedence over logical operators.

Example 8.2 Simple linking

This example searches for links to a specified databank in the results of an index search:

```
[swissprot-des:kinase] > pdb
```

The result will be a list of all the entries in the PDB databank that have links to the results of the index search:

```
[swissprot-des:kinase]
```

See section 8.2 “Searching in Indices” (p. 146), for an explanation of index searches.

Example 8.3 Multiple linking 1

It is possible to combine several linking queries. For example:

```
[swissprot-id:acha_human] > prosite > swissprot
```

The search first retrieves the SWISS-PROT entry “acha_human”. This entry will have links to PROSITE entries that document the protein family of which “acha_human” is a member. (In this case it is the family of neuronal acetylcholine receptors.) These items in the PROSITE databank are retrieved. The next link retrieves SWISS-PROT entries that are linked to the PROSITE entries, i.e., belong to the neuronal acetylcholine receptor family. In effect, the entry “acha_human” is being amplified to retrieve all the entries in SWISS-PROT which document members of the protein family or families to which it belongs.

Example 8.4 Multiple linking 2

A similar technique to that used in example 8.3 Multiple linking 1, can be used to find related information in another databank to which the initial entry is not linked.

```
[swissprot-id:gshr_caeel] > prodom > pdb
```

The query retrieves a probable glutathione reductase (whose **ID** is “gshr_caeel”) from SWISS-PROT, searches for entries in ProDom which document related proteins, and then looks for links to PDB in these entries. The result is a set of PDB protein structures that are homologous to the SWISS-PROT entry “gshr_caeel”.

Example 8.5 Complex linking

This example can be thought of as being composed of two parts:

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

The first part of the query searches the `description` fields of the SWISS-PROT and SWISSNEW databanks, looking for “kinase”. The second part of the search excludes any entries that have links to the SWISSNEW databank. Using this technique, it is possible to retrieve the entries for “kinase”, but exclude any that are replaced by (more up to date) entries in SWISSNEW.

The distinction being made here is that entries in SWISSNEW will not be linked to themselves (or other entries in SWISSNEW), so all the SWISSNEW entries will be kept. However, any entry in SWISS-PROT that has been replaced by an entry (to which that entry is linked) in SWISSNEW will be picked up and rejected. In this way, out-of-date entries are excluded.

Note: The first part of the query also defines a group, `q`, which contains the SWISS-PROT and SWISSNEW databanks. This is used in the second part of the query rather than listing the databanks explicitly (see section 8.2.6 "Searching Multiple Databanks", p. 153).

Example 8.6 Searching multiple databanks and screening for overlaps

Many protein or DNA databanks overlap to a great extent, which creates a lot of redundancy; however, the annotation of equivalent entries in different databanks can be quite varied. This can be useful for string searching because the probability of finding a certain enzyme name is greater if you can search all sequence databanks. After the search, links can be used to remove any overlaps.

See Example 8.5 Complex linking, for how this might be done.

8.4 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 include "transmembrane" in the `keywords` index. The `ftkey` index, however, has a special type, allowing it to find features of a given type within entries. Thus the second query retrieves a set of subentries within SWISS-PROT entries. Each subentry is a transmembrane feature.

Note: The second query above will retrieve many more entries than the first because most transmembrane proteins have more than one membrane spanning segment. If you requested the sequences for the entries in the second set, you would get the transmembrane segments rather than the parent entry's sequence.

8.4.1 Links with Sets Containing Subentries

Simple Links

It is not possible to combine sets of entries with sets of subentries using the logical operators; however, link operators may be used between sets of entries and sets of subentries. For example:

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

returns a set of transmembrane segment subentries found in human proteins, whereas

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

returns all human proteins that have transmembrane segments.

Parent Links

Sometimes it is necessary to do an explicit conversion from subentries to entries. This can be done using the operand, `parent`. This method looks for links from the subentries to their respective parent entries and retrieves a set containing parent entries. For example:

```
[swissprot-ftkey:transmem] > parent
```

retrieves the parent entries (the entries to which the subentries belong) for the set of subentries from SWISS-PROT that have transmembrane sequence features. Logical operators can then be used to combine the set of parent entries with another set of entries. Furthermore, as with other SRS query language commands, it is possible to combine the link to the parent entries and the subsequent operations into the same command, see section "More Complex Links using the Parent Operand" (p. 163).

More Complex Links using the Parent Operand

The command:

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

returns all entries that have the keyword, "transmembrane", or that have "transmem" sequence features. The index search `[swissprot-ftkey:transmem]` results in a set of subentries whereas the index search `[swissprot-key:transmembrane]` returns a set of entries. These sets cannot be combined directly using a logical operator. Instead, an extra step must be added, which finds the set of parent entries to which the subentries belong. The resultant set of parent entries can then be combined with the set from the second index search using a logical operator.

This type of search may be necessary to ensure all entries with a certain property are retrieved because the annotation is often incomplete.

8.5 Storing Intermediate Results in Sets

If a query becomes very complicated, it may be convenient to store an intermediate result, into a set with a name which can be used later in the query. This is particularly useful to save typing out the query several times, but does not actually save a file. This example is an index query in EMBL that is then linked to both SWISS-PROT and SWISSNEW:

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

The EMBL index query is specified twice, however, it is possible to store the result of the index query in a set, e.g., `temp`, which saves duplicate typing of the index query. The assignment operation must be within parentheses, as shown:

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

INDEX

- (index search) 147

Symbols

(range search) 147
* (wildcard) 55
: (string search) 147
? (wildcard) 55

A

administration
  projects 43
application invocation page 24, 27, 28
application select page 20
applications 131, 132
  batch 136
  interactive 134
  invocation page 24, 27, 28
  parameters 23
  select page 20
  using 132
apply
  views 105

B

batch
  applications 136
applications
  batch queued 23
  not batch queued 25
browse
  index 80

C

command line 141
copy
  project information 45
creating
  views 118

D

databanks
  list of databanks page 37
  search multiple 153
dates
  in indices, search for 53
  query language, search for 152
delete 49
dialog
  file download 31
  save as 32
documentation
  SRS documentation page 37
download
  file download dialog 31
download options page 30

E

entries 161
  find all entries, linking 92
  linking 162
entry page
  linking from 97
examples
  expression linking 100
  getz 142
  linking 159, 160, 161
  regular expressions 150
expression
  expression linking 98
    examples 100
    procedure 98
  expression query 73
    using 74
expressions
  regular expressions 55
    examples 150
    query language 149
extended query form 60

F
field information page 81
file download dialog 31
forms
  extended query 60
  query 57
  standard query 6, 57

G
getz 141, 142
  arguments 143
  examples 142

H
how to
  expression linking 98
  linking 89

I
index
  browse 80
  search
    - 147
    query language 146
    syntax 147
information
  field information page 81
  list of databanks page 37
  SRS documentation page 37
interactive
  applications 134
intermediate results
  storing 164

J
job status
  page 139

L
language
  query language 145
link page 10
linking 85, 86
  entries and subentries 162
  examples 159, 160, 161
  expression linking 98
    examples 100
    procedure 98
  find all entries options 92
  from entry page 97
  from query manager page 93
  from query result page 94
  link page, finding 93
  operators 99
    query language 158
  parent 163
  procedure 89
list of databanks 37
list view 118, 119, 122
logical operators
  query language 157

M
manage
  manage your query results page 30
managing
  project manager page 43
  projects
    permanent 42
    temporary 38
multiple
  databanks, searching 153

N
non-secure permanent projects 40
  starting 41
numbers
  in indices, search for 53
numerical ranges
  combining 152
  searching, query language 151

O
open 48
operands 155
operators
  linking 99, 158
  logical 157
  query language 156
  regular expressions 149
  using 159
order
  sorting results 76
    overview 9

P
page
  application invocation 24, 27, 28
  application select 20
  download options 30
  entry 97
  extended query form 60
  field information 81
  job status 139
  link 10
  list of databanks 37
  manage your query results 30
  query manager 16, 19, 93
  query result 8, 94
  SRS documentation 37
  standard query form 57
  start 2
  top 4
  view manager 1 12
  view manager 2 14
parameters
  applications 23
parent
  links 163
permanent projects 34
  managing 42
  non-secure 40
  secure 40
  starting 41
project
  starting 2
projects
  administration 43
  permanent 34
    managing 42
    non-secure 40
    secure 40
    starting 41
  project manager page 43
  starting 35
  switching 44
  temporary 34, 37
    managing 38
    starting 37

Q
query 51
  expression query 73
  manage your query results page 30
  sorting results 76
    overview 9
  using expression queries 74
query forms 57
  extended 60
  standard 6, 57
query language 145
  operands 155
  operators 156
    linking 158
    logical 157
  regular expressions 149
  search
    combining 153
    dates 152
    index 146
    multiple databanks 153
    numerical ranges 151

        syntax 154
    what is it 146
    wildcards 148
query manager page 16, 19
    linking from 93
query result page 8
    linking from 94
quick search 55
    using 56

R
range search
    # 147
ranges
    combining 152
    numbers and dates 53
regular expressions 55
    examples 150
    query language 149
rename 49
results
    sorting 76
        overview 9

S
save 31, 46
save as dialog 32
search
    dates, query language 152
    index, query language 146
    numerical ranges, query language 151
    query language
        dates 152
        index 146
        numerical ranges 151
    quick search 55
        using 56
    strings
        query language 147
    terms 52
        dates 53
        multiple word 53
        numbers 53
        regular expressions 55
        single word 52
        wildcards 55
secure permanent projects 40
    starting 41
select
    application select page 20
sets 164
single word
    search terms 52
sort
    results 76
        overview 9
SRS documentation page 37
standard query form 6, 57
start
    permanent projects 41
    project 35
    start page 35
    temporary projects 37
start page 2
string
    search, query language 147
string search
    : 147
subentries 161
    linking 162
switch
    projects 44
syntax
    index search 147

T
table view 118, 119, 122
temporary projects 34, 37
    managing 38

  starting 37
top page 4

U
UNIX 142
upload 48
using
  applications 132
  getz 143
  operators 159
  views 104

V
view
  list view 118, 119, 122
  table view 118, 119, 122
view manager page 1 12
view manager page 2 14
views 103
  applying 28, 105
  creating 118
  using 104

W
what is
  application 132
  linking 86
  query language 146
  view 104
wildcards 55
  * 55
  ? 55
  query language 148
word
  multiple word search terms 53
  single word search terms 52