

# BONDplus

Version 1.0

## USER GUIDE

## COPYRIGHT

BONDPLUS USER GUIDE  
Version 1.0 (November, 2007)

Copyright © 2007 The Thomson Corporation.

All rights reserved. Reproduction or redistribution of Thomson content, including by framing or similar means, is prohibited without the prior written consent of Thomson. 'Thomson' and the Thomson logo are registered trademarks and trademarks of The Thomson Corporation and its affiliated companies. All other product and service names cited are trademarks of their respective companies.

# Contents

---

<b>1</b>	<b>Welcome.....</b>	<b>1</b>
	Contacting Support .....	1
	Contact Information.....	1
	Support Checklist .....	1
<b>2</b>	<b>About BONDplus .....</b>	<b>3</b>
	BONDplus Sequences.....	3
	BINDplus on BONDplus .....	4
	Interactions .....	4
	Complexes .....	4
	GENESEQ on BONDplus .....	5
	Common Use Cases .....	6
<b>3</b>	<b>The BONDplus Interface .....</b>	<b>7</b>
	BONDplus Login and Logout .....	7
	Using the Interface.....	8
<b>4</b>	<b>Searching for Data.....</b>	<b>9</b>
	Searching From the Home Page .....	9
	Browse Search .....	11
	Performing a Browse Search .....	11
	Text Search.....	13
	Performing a Text Search .....	13
	Identifier Search.....	15
	Performing an Identifier Search.....	15
	Advanced Field Specific Search .....	19
	Performing an Advanced Field Specific Search.....	19
	BLAST Search .....	24
	Performing a BLAST Search.....	24
	Constructing Queries .....	28
	Apache Lucene Query Language.....	28
	Viewing Results .....	30
	Saving Search Results .....	31
<b>5</b>	<b>Sequences.....</b>	<b>37</b>
	Performing a Query for Sequences .....	37
	Using the Sequences Filters.....	38
	Sequences Summary Record View .....	40

Sequences Summary Record View Types .....	41
Sequences Single Record View .....	43
Sequences Single Record View Types .....	44
Exporting Sequence Search Data .....	53
Exporting Data From a Sequence Summary or Single Record View .....	53
Sequence Export Formats .....	53
<b>6 Interactions and Complexes.....</b>	<b>55</b>
Interactions .....	55
Complexes .....	55
Performing a Query for Interactions .....	55
Using the Interactions and Complexes Filters .....	56
Interactions Summary Record View .....	60
Interactions Summary Record View Types.....	62
Visualize Results .....	68
Interactions Single Record View .....	69
Interactions Single Record View Types .....	69
Exporting Interactions Search Data .....	77
Exporting Data From an Interactions Summary or Single Record View .....	78
Interactions Export Formats .....	78
Complexes Summary Record View .....	79
Complexes Summary Record View Types.....	81
Visualize Results .....	86
Complexes Single Record View .....	86
Complexes Single Record View Types .....	87
Exporting Complexes Search Data .....	93
Exporting Data From a Complexes Summary or Single Record View .....	93
Complexes Export Formats .....	94
<b>7 GENESEQ.....</b>	<b>97</b>
Performing a Query for GENESEQ.....	97
Using the GENESEQ Filters.....	98
GENESEQ Summary Record View.....	99
View Types.....	100
GENESEQ Single Record View .....	102
GENESEQ Single Record View Types .....	103
Exporting GENESEQ Search Data.....	110
Exporting Data From a GENESEQ Summary or Single Record View.....	110
GENESEQ Export Formats .....	111
<b>Appendix A About Ontoglyphs .....</b>	<b>113</b>
Detailed Ontoglyph Descriptions by Category .....	113
Function .....	114
Sub-Cellular Localization .....	124
Binding .....	130



Detailed Ontoglyph Example.....	134
---------------------------------	-----

## **Appendix B About Proteoglyphs ..... 137**

What is a Proteoglyph Summary?.....	139
-------------------------------------	-----

## **Appendix C Frequently Asked Questions ..... 141**

BONDplus FAQ.....	141
-------------------	-----

BINDplus FAQ.....	144
-------------------	-----

GENESEQ FAQ.....	146
------------------	-----

Browser FAQ.....	148
------------------	-----

## **Index ..... 149**



# Chapter 1

## Welcome

---

Welcome to the BONDplus User Guide. This documentation includes overviews, procedures, and examples which will assist you in performing certain tasks and navigating throughout the application.

### Contacting Support

Click the Contact Us menu in BONDplus and choose Report A Problem to let Customer Support know of any problems you have encountered. If a problem persists then contact a Customer Support representative.

Customer Support hours are:

Monday to Thursday	07:30 AM to 7:00 PM EST
Friday	07:30 AM to 6:00 PM EST

If you phone Customer Support outside of these hours, leave a detailed message and you will be contacted during regular business hours. Please provide detailed information (as outlined in the [Support Checklist](#)).

### Contact Information

<b>Phone:</b>	<b>Americas:</b> +1-800-336-4474 ext. 1591 or +1 215 386 0100 <b>Europe, Middle East and Africa:</b> +44 20 7433 4999 <b>Japan:</b> 0800 888 8855 or +81 3 5218 6500 <b>Asia Pacific:</b> +65-6879-4118 <b>China:</b> +86 10 8286 2099 <b>Korea:</b> 080 010 8100 or +82 2 2076 8100 <b>Australia:</b> 1800 007 214 <b>New Zealand:</b> 0800 443 162
<b>Web:</b>	<a href="http://scientific.thomson.com/support/techsupport/">http://scientific.thomson.com/support/techsupport/</a>

### Support Checklist

When contacting customer support, please provide the following information:

- Your name, company name, email address, and phone number.

- Error messages and the circumstances of their occurrence.
- A full description of the problem:
  - What happened? What were the sequence of events that preceded the problem?
  - In which screen or window did the problem occur?
  - Was the problem the result of pressing a key?
  - Did the screen freeze? What functions of the application are affected?

## Chapter 2

### About BONDplus

---

BONDplus (Biomolecular Network Databank plus) is a life-science data management platform that leverages public and proprietary data to help drive discovery across your organization.

Integrating GENESEQ™ (see “GENESEQ on BONDplus” on page 5) and BINDplus (see “BINDplus on BONDplus” on page 4), BONDplus takes you beyond the genome to take full account of all the relevant biological and IP relationships surrounding your molecule of interest.

BONDplus gathers foundational public data and makes it relevant to you. No other resource integrates sequence, interaction, and patent data in such a way to leverage new insight.

BONDplus delivers a powerful range of data to your desktop, including sequence, interaction, patent, taxonomy, publication, annotation, domain, and cross-reference data from fifteen assembled information sources.

#### **What you can do**

- Search GENESEQ™, BINDplus and public life science data on a single platform
- Integrate external and internal information silos
- Accelerate your work with pre-computed datasets and cross-references
- Access data via a rich web interface
- Enable your biological applications and high performance computing resources with robust data retrieval API.

For more information about BONDplus, see “BONDplus FAQ” on page 141.

### **BONDplus Sequences**

BONDplus sequence information is derived from a biological sequence retrieval system developed by Entrez, called SeqHound. The sequence and structure database content of SeqHound is derived from GenBank and is updated on a daily basis. Unlike Genbank, SeqHound goes further by allowing you to query by way of a well-established API.

BONDplus also stores annotation links for sequences, including links to Genome Ontology descriptions, MedLine abstracts, taxon identifiers, associated structures, redundant sequences, sequence neighbours, conserved domains, data base cross-references, Online Mendelian Inheritance in Man identifiers, LocusLink identifiers and complete genomes.

For information on searching and viewing sequence-related data, see “Sequences” on page 37.

## BINDplus on BONDplus

BINDplus contains historic as well as the latest published low-throughput (LTP) and high-throughput (HTP) interactions documented according to rigorous data curation standards.

The growing data in BINDplus contain more than 60,000 unique Gene Identifiers, over 1,500 unique organisms, and 7,555 Gene Ontology terms derived from peer-reviewed scientific data extracted from over 23,800 journal articles and over 9000 corresponding authors.

To give users the most current information, BINDplus releases data in a ‘real-time’ system, with database updates available through BONDplus (Biomolecular Object Network Databank Plus) immediately upon public release. The BINDplus flat-file is published via FTP every week.

### What you can do

- Access more than 200,000 documented biomolecular interactions and complexes
- Identify biomolecules and corresponding binding partners
- Explore relevant biological relationships surrounding your molecule of interest
- Use accurate, manually captured biological information
- Locate required information without exhaustive literature searches

BINDplus is an interaction database with classifications for molecular associations: molecules that associate with each other to form [Interactions](#) and molecular [Complexes](#) that are formed from one or more interactions.

For more information, see “BINDplus FAQ” on page 144.

## Interactions

A BINDplus record represents an interaction between two or more objects that is believed to occur in a living organism. A biological object can be a protein, DNA, RNA, ligand, molecular complex, gene, photon or an unclassified biological entity.

BINDplus records are created for interactions which have been shown experimentally and published in at least one peer-reviewed journal. A record also references any papers with experimental evidence that support or dispute the associated interaction.

Interactions are the basic units of BINDplus and can be linked together to form molecular complexes.

For more information, see “Interactions and Complexes” on page 55.

## Complexes

A molecular complex is a collection of two or more molecules that associate to form a functional unit in a living organism. In BINDplus, complexes are represented as molecular complex objects, formed by linking two or more interaction records.

Molecular complex records are supplemented with additional information such as complex topology and the number of subunits (BINDplus objects) involved in the interaction.

For more information, see “Interactions and Complexes” on page 55.

## GENESEQ on BONDplus

GENESEQ Sequence Databases are produced by Thomson Scientific's qualified molecular biology experts, who assess each patent before producing individual records for each sequence within that patent. This team of qualified geneticists and molecular biologists receive patents from Thomson Scientific's Patent Administration team following patent publication, and then check through each page to find nucleic acid and amino acid sequences in claims, examples and other parts of the disclosure. The team includes individuals with foreign language skills who translate the required information to produce English language records.

### GENESEQ

Established in 1990, GENESEQ has rapidly grown to become the world's largest database dedicated to collection and annotation of sequence information from patents, dating back to 1980. Thomson Scientific scans patents from 41 patent-issuing authorities for those which contain nucleic acid and amino acid sequences. These sequences are annotated and then added to the database which is updated every two weeks. Thomson Scientific's molecular biology experts assess every patent to ensure that GENESEQ contains the most comprehensive, global coverage of sequence information. Every GENESEQ record is prepared by Thomson Scientific's team of experts to bring you:

- specially-written titles to cover the main developments and content of the patent
- full bibliographic data (for the fields included in each record)
- concise summaries detailing the pertinent points about the sequence
- specially written English language abstracts of patents presented originally in a broad range of nonEnglish languages.
- feature tables which highlight specific areas within the sequence such as promoter regions, CAAT box, coding regions, and so on.
- keywords are included which are derived from the patent and increase searchability of the database. Since 2005 the field was broken down into three keyword themes:
  - patent specific keywords
  - disease and activity terms
  - sequence specific keywords

These are applied consistently between patents and records from an in-house thesaurus which is also consistent with other Thomson Scientific products. These three themes cannot be distinguished from each other within the Keyword list.

- searchable sequences.

### GENESEQ databases

All new nucleic acids (10 or more bases in length) and amino acids (four or more residues in length) disclosed in patents are included in the database.

Current database stats for GENESEQ are more than 12 million sequences from 125,000 patents.

**NOTE:** Another approximately 35 million sequences from approximately 80 patents are not in GENESEQ due to large file sizes. Thomson Scientific is working to add these as quickly as possible with plans to increase to over 47 million sequences. A full list of outstanding sequences and patents is available on request.

Of the total records, 20 percent are foreign language patents which have been translated into English. The foreign language patents which are included in the database include patents from Japan, Germany, France, Latin America, Russia, India, Korea, China, Hungary and many more countries.

For more information, see “GENESEQ FAQ” on page 146 and [An Introduction to GENESEQ](#).

## Common Use Cases

You can use BONDplus to retrieve relevant content from Thomson Scientific’s public sequence, patented sequence (GENESEQ), and biomolecular interaction (BINDplus) databases.

To retrieve records based on sequence homology, use BONDplus’ BLAST functionality to retrieve related records from GENESEQ, BINDplus, or other public sequence databases (see “BLAST Search” on page 24).

To retrieve records based on other classifications, use BONDplus’ text search engine to retrieve arbitrary content as required.

To retrieve all sequences, interactions, and GENESEQ records relating a molecule of interest, you can perform a simple search on the molecule name or molecule alias across all databases.

To retrieve all public sequences by molecular function, you can search all gene products by GO identifier and rapidly restrict your results by organism or source public database (see “Performing a Query for Sequences” on page 37)

To retrieve BINDplus interaction records relating to a peer-reviewed publication, biologists can search on the PubMed Identifier or by the MeSH terms assigned to related publications (see “Performing a Query for Interactions” on page 55). You can view the associated publication abstract for BINDplus interaction records where available (see “Interactions Single Record View” on page 69).

To view the coverage of a therapeutic area, you can search on the GENESEQ keywords assigned by the editorial team. To understand the patent landscape claimed by a specific company, IP professionals can search GENESEQ by patentee, inventor, and claim type (see “Performing a Query for GENESEQ” on page 97).

To gain a high level understanding of all BONDplus content relating to a patented sequence of interest, you can use the GENESEQ Sequence Summary view to access a summary of all GENESEQ Accessions, NCBI GIs, Patents, and Similar Molecules contained in BONDplus (see “GENESEQ Summary Record View” on page 99).



## Chapter 3

### The BONDplus Interface

---

The BONDplus application uses a standard browser such as Microsoft Internet Explorer or Mozilla Firefox for its user interface. Check with your Thomson representative for which browser versions are compatible for use with BONDplus.

#### **BONDplus Login and Logout**

To search for data within BONDplus you must first log in to the application with the login user name and password assigned and given to you by your Thompson representative. It is recommended that you logout of the BONDplus application when you are finished using it.

##### **To log in to BONDplus**

1. Open your browser.
2. In the browser's address field, enter this URL:  
[www.bond-plus.com](http://www.bond-plus.com)

3. The BONDplus login page appears.

Thomson.com

**THOMSON**

**BONDplus**

ABOUT PRODUCTS NEWS

BONDplus

HOME SEARCH MyBOND HELP CONTACT US

BOND Remote Login December 10, 2007

An account is required to access BONDplus.

Login:

Password:

LOGIN

[More information about BONDplus](#)

Thomson Scientific | Copyright Notices | Terms of Use | Privacy Statement

4. In the Login section, enter your user name in the Login field and your password in the Password field. Click LOGIN.

Login:

Password:

LOGIN

5. The BONDplus home page appears. You can now begin using BONDplus. To information on how to begin searching for data see “Searching for Data” on page 9.

#### To log out of BONDplus

When you have finished all tasks, click LOGOUT on the menu bar.

ABOUT PRODUCTS NEWS **LOGOUT**

BONDplus

## Using the Interface

The BONDplus application uses a standard browser such as Microsoft Internet Explorer or Mozilla Firefox for its user interface. The general usage of a browser is outside the scope of this documentation – please refer to your browser’s documentation on usage and navigation.

## Chapter 4

### Searching for Data

---

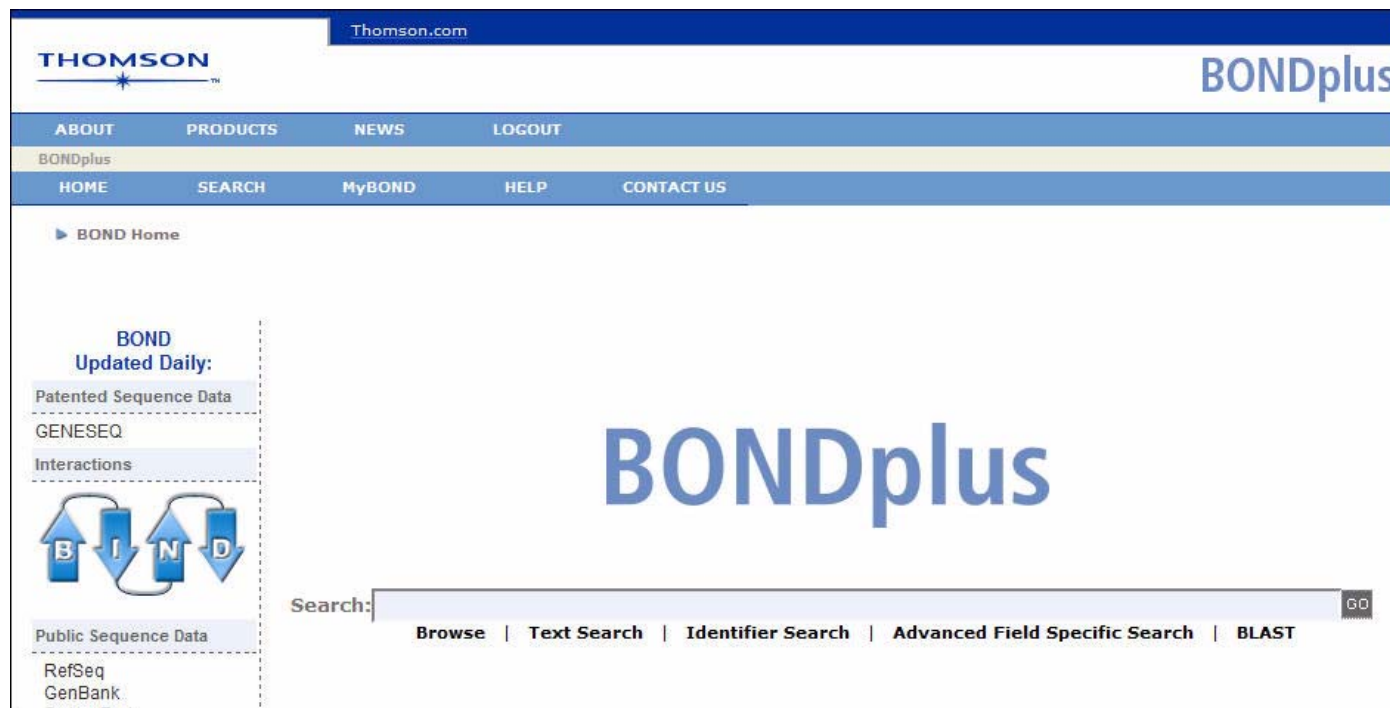
This chapter describes the various ways to search for data in BONDplus by performing cross-database queries for available sequences, interactions and complexes.

After you log in to BONDplus (see “BONDplus Login and Logout” on page 7) you begin your search on the Home page (see “Searching From the Home Page” on page 9).

**NOTE:** You can retrieve previously saved search results by clicking MyBOND on the BONDplus menu (for more information, see “Saving Search Results” on page 31).

#### Searching From the Home Page

After you log in to BONDplus (see “BONDplus Login and Logout” on page 7), the home page appears, presenting you with a search field and links to begin your search for data.



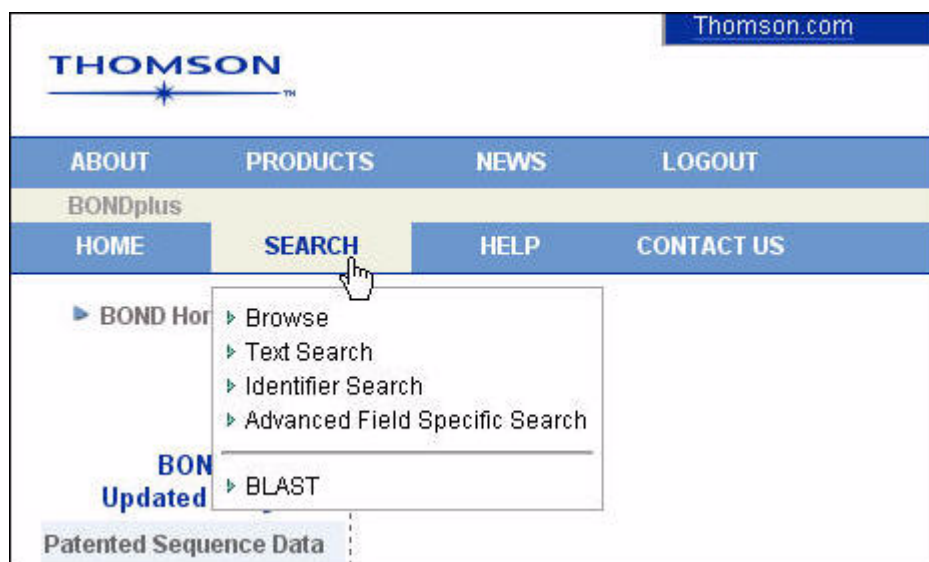
You can perform a specific search or a quick text search.

### Performing a specific search

1. To begin a specific search, click one of the links below the Search field.



**NOTE:** You can also access the search links from the Search menu.



2. A search specific page corresponding to the link you clicked appears (for example, if you clicked the Identifier Search link, the Identifier Search page appears)
3. Continue your search:
  - Browse (see “Browse Search” on page 11)
  - Text (see “Text Search” on page 13)
  - Identifier (see “Identifier Search” on page 15)
  - Advanced Field Specific (see “Advanced Field Specific Search” on page 19).
  - BLAST (see “BLAST Search” on page 24).

### Performing a quick Text search

1. To perform a quick text search, enter any text (for example, “brca1” or “linker for activation of T cells”) or query string in the Search field. What you enter is used to search within all fields in the BONDplus databases. For more information, see “Constructing Queries” on page 28.
2. You can use these boolean operators:
  - NOT
  - AND
  - OR

(for example: brca1 AND “homo sapiens”).

You can also use these shorthand boolean operators:

  - +
  - -

- &&
- ||
- !

(for example: +brca1 + “homo sapiens”).

You can use these wild card modifiers:

- ? for a single character (for example, brca?)
- \* for multiple characters (for example, lat\*).



Search:

**Browse | Text Search | Identifier Search | Advanced Field Specific Search | BLAST**

3. Click GO.
4. The Text Search page appears, providing a summary of your search results (see “Viewing Results” on page 30).
5. You can begin to work with your results or perform another text search (see “Text Search” on page 13). You can also save your search results for later retrieval (see “Saving Search Results” on page 31).

## Browse Search

The Browse Search option allows you to retrieve all records available in BONDplus.

### Performing a Browse Search

1. On the BONDplus Home Page, click the Browse link (see “Searching From the Home Page” on page 9).

- The Browse page appears with your search results displayed on the Summary tab and the number of records displayed on each respective Results tab (Sequences, Interactions, Complexes, and GENESEQ tabs).

The screenshot shows the 'Browse' page interface. At the top left, there is a 'BROWSE' button. Below it, a 'Records per page:' dropdown menu is set to '20'. A 'BROWSE' button is also present below the dropdown. Below these elements is a row of six tabs: 'Summary', 'Sequences', 'Interactions', 'Complexes', 'Pathways', and 'GENESEQ'. Each tab has a red number in brackets indicating the count of records: Summary [102606902], Sequences [89640465], Interactions [217484], Complexes [3739], Pathways [8], and GENESEQ [12745206]. Below the tabs, the text 'Your search returned...' is followed by a list of results: 89640465 Public Sequences, 217484 Interactions, 3739 Complexes, 8 Pathways, and 12745206 Patented Sequences (GENESEQ).

- If desired, adjust the number of records that will be displayed when any of the Results tabs are clicked. Click the Records per page list and select a number.

The screenshot shows the 'Records per page:' dropdown menu open. The menu lists options: 10, 20, 30, 40, and 50. The '30' option is highlighted, and a mouse cursor is pointing at it. The 'BROWSE' button is visible below the dropdown menu.

- Click a Results tab to view the records retrieved for that category. You can also save your search results for later retrieval (see "Saving Search Results" on page 31).
- Go to any of these chapters for information on viewing and using data on the Results tabs:
  - "Sequences" on page 37
  - "Interactions and Complexes" on page 55
  - "GENESEQ" on page 97.

**NOTE:** This search method results in a large amount of records being retrieved, however, data filters for each category (Sequences, Interactions, Complexes, and GENESEQ) are available on their respective Results tabs, allowing you to narrow your search results.

## Text Search

The Text Search option allows you to perform a full text query against the BONDplus database by entering query terms of your choice. The query language supported in BONDplus is the Apache Lucene query language. For more information, see “Constructing Queries” on page 28.

### Performing a Text Search

1. On the BONDplus Home Page, click the Text Search link (see “Searching From the Home Page” on page 9).
2. The Text Search page appears.

3. Enter any text (for example, “brca1” or “linker for activation of T cells”) or a query string in the Text Query field. What you enter is used to search within all fields in the BONDplus databases.

You can use these boolean operators:

- NOT
- AND
- OR

(for example: brca1 AND “homo sapiens”).

You can also use these shorthand boolean operators:

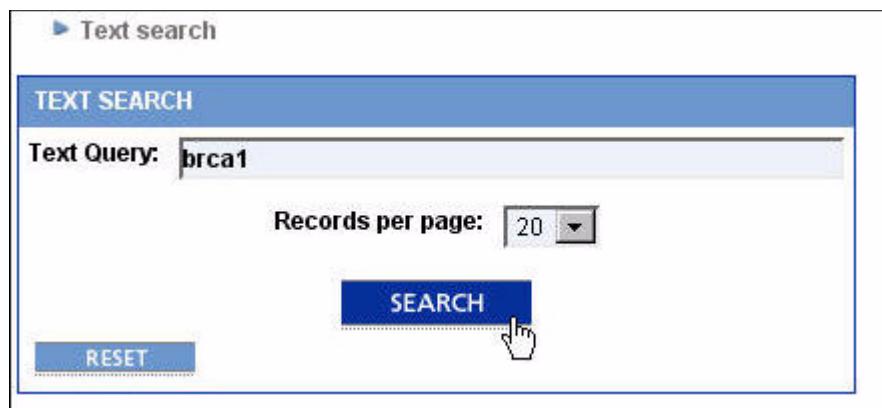
- +
- -
- &&
- ||
- !

(for example: +brca1 + “homo sapiens”).

You can use these wild card modifiers:

- ? for a single character (for example, brca?)

- \* for multiple characters (for example, lat\*).



Text search

**TEXT SEARCH**

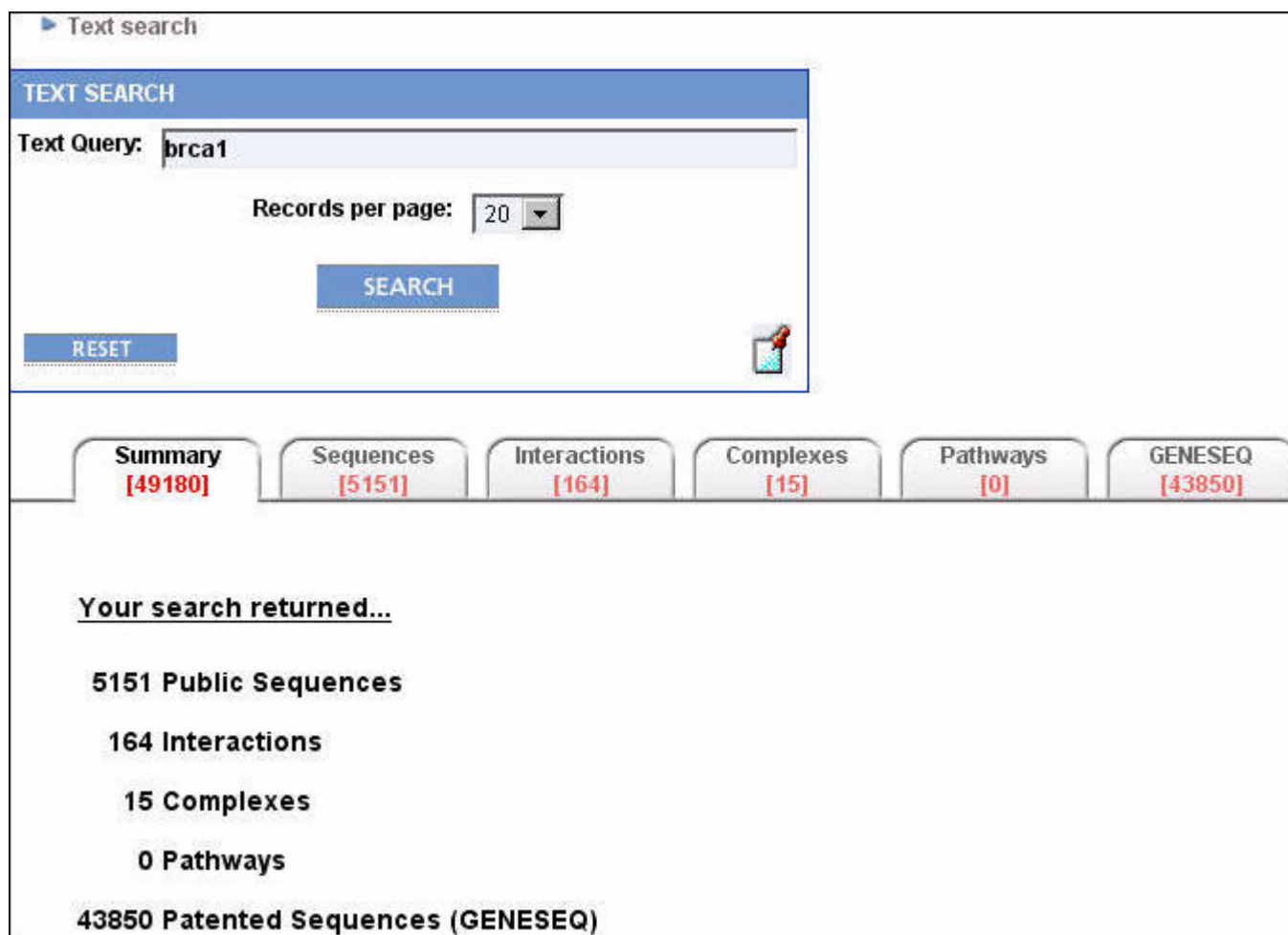
Text Query:

Records per page:

**SEARCH**

**RESET**

4. Click Search.
5. The Text page refreshes, with your search results displayed on the Summary tab and the number of records retrieved for each category displayed on each respective Results tab.



Text search

**TEXT SEARCH**

Text Query:

Records per page:

**SEARCH**

**RESET**

**Summary [49180]** **Sequences [5151]** **Interactions [164]** **Complexes [15]** **Pathways [0]** **GENESEQ [43850]**

**Your search returned...**

**5151 Public Sequences**

**164 Interactions**

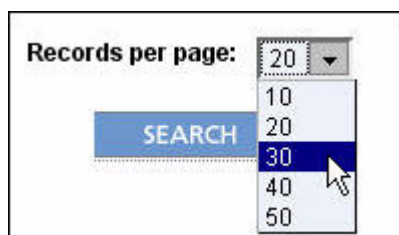
**15 Complexes**

**0 Pathways**


**43850 Patented Sequences (GENESEQ)**

6. If desired, adjust the number of records that will be displayed when any of the Results tabs are clicked. Click the Records per page list and select a number.





7. Click a Results tab to view the records retrieved for that category.
8. Go to any of these chapters for information on viewing and using data on the Results tabs:
  - “Sequences” on page 37
  - “Interactions and Complexes” on page 55
  - “GENESEQ” on page 97.

**NOTE:** The search history for your current session can be viewed by clicking the  icon in the Text Search section. Move the pointer over the icon to see a list of text searches you have performed and click an item in the list to perform the search again.

You can also save your search results for later retrieval (see “Saving Search Results” on page 31).

## Identifier Search

This Identifier Search options allows you to search BONDplus by using a common identifier.

Identifier searches can also be constructed using the Apache Lucene query language within the Advanced Field Specific Search interface. For more information, see “Constructing Queries” on page 28.

### Performing an Identifier Search

1. On the BONDplus Home Page, click the Identifier Search link (see “Searching From the Home Page” on page 9).
2. The Identifier Search page appears.

Identifier search

IDENTIFIER SEARCH

BIND Id

Records per page: 20

SEARCH

RESET

Summary

Sequences

Interactions

Complexes

Pathways

GENESEQ

3. Click on the list box and select the identifier category to which the identifier used in your search belongs.

Identifier search

IDENTIFIER SEARCH

BIND Id

GENESEQ Accession  
Patent Number  
NCBI GI  
NCBI Accession  
BIND Id  
Entrez Gene ID  
PubMed Id  
Taxon Id  
Molecular Object Database Id  
-----  
AtCS  
ASAP  
CCDS  
CDD  
COG  
coliBase  
dbEST  
DDBJ  
dictyBase  
E.C. Number

page: 20

ARCH

ces

Interactions

Complexes

Pathways

GENESEQ

4. Enter the identifier number you want to use in your search.

► Identifier search

IDENTIFIER SEARCH

BIND Id


▼

1313

Records per page: 20 ▼

SEARCH

RESET



Summary

Sequences

Interactions

Complexes

Pathways

GENESEQ

5. Click Search.
6. The Identifier Search page refreshes, with your search results displayed on the Summary tab and the number of records retrieved for each category displayed on each respective Results tab.

► Identifier search

IDENTIFIER SEARCH


BIND Id

1313

Records per page: 20

SEARCH

RESET



Summary

Sequences [0]

Interactions [1]

Complexes [0]

Pathways [0]

GENESEQ [0]

**Your search returned...**

**0 Public Sequences**

**1 Interactions**

**0 Complexes**

**0 Pathways**

**0 Patented Sequences (GENESEQ)**

7. If desired, adjust the number of records that will be displayed when any of the Results tabs are clicked. Click the Records per page list and select a number.

Records per page:

20

10


20

30


40

50

SEARCH



8. Click a Results tab to view the records retrieved for that category.
9. Go to any of these chapters for information on viewing and using data on the Results tabs:
- “Sequences” on page 37
  - “Interactions and Complexes” on page 55
  - “GENESEQ” on page 97.

**NOTE:** The search history for your current session can be viewed by clicking the  icon in the Identifier Search section. Move the pointer over the icon to see a list of search terms you have used in previous searches and click an item in the list to perform the search again.

You can also save your search results for later retrieval (see “Saving Search Results” on page 31).

## Advanced Field Specific Search

The Advanced Field Specific Search option provides a field-specific query builder that allows you to build complex field-specific queries without having to know the Apache Lucene query language. The query language supported in BONDplus is the Apache Lucene query language. For more information, see “Constructing Queries” on page 28.

When performing a query, you can specify conditions you want included in your search (displayed in blue) and conditions that are excluded in your search (displayed in red).

### Performing an Advanced Field Specific Search

1. On the BONDplus Home Page, click the Advanced Field Specific Search link (see “Searching From the Home Page” on page 9).
2. The Advanced Field Specific Search page appears.

Field-specific search

**ADVANCED FIELD SPECIFIC SEARCH**

Find all records where **at least one** of the following conditions are met:

1. The **field** **any** contains **at least one of the following words**:

**ADD CONDITION** **ADD SUB QUERY** **ADD EXCLUSION**

Records per page: **20**

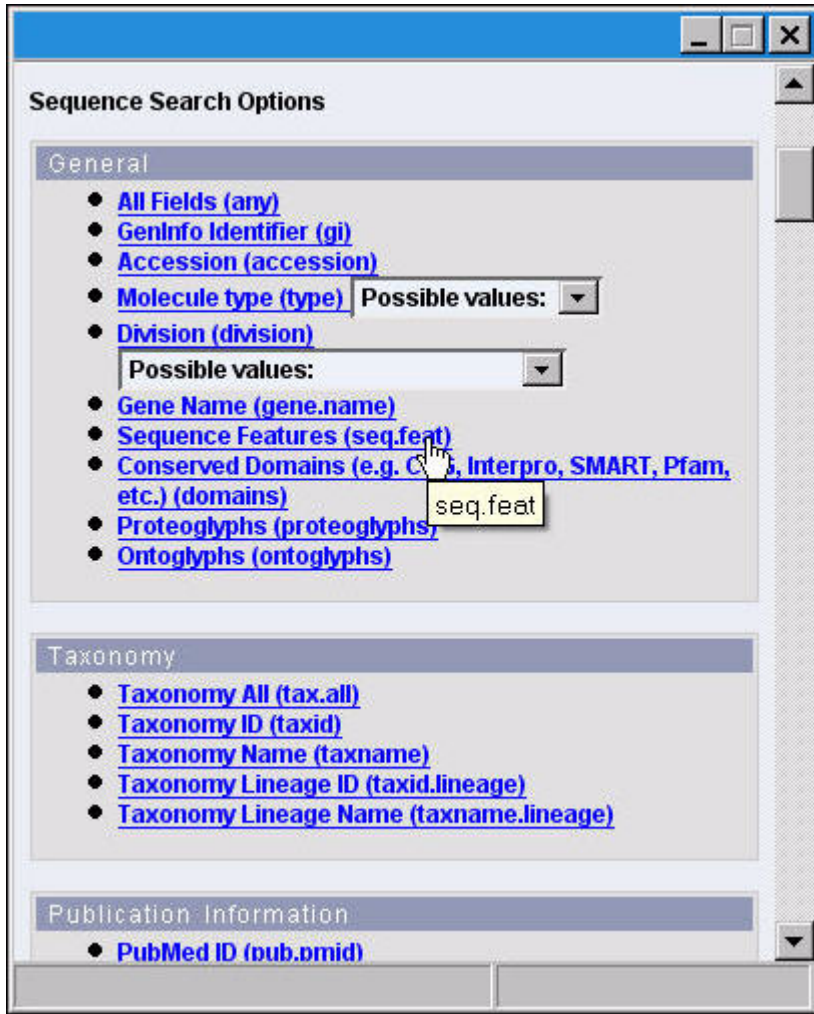
**SEARCH** **RESET**

**Summary** **Sequences** **Interactions** **Complexes** **Pathways** **GENESEQ**

3. Enter a specific field you want the condition to use in the search of the database records or leave the default “any” to include all fields.

1. The **field** **any**

**NOTE:** Click the field link and a pop-up window appears in which you can choose from a number of available fields.



4. Enter as a search term any text (for example, “brca1” or “linker for activation of T cells”) or a query string.

are met:

the following words  : brca1

5. Use the contains list box to specify that your search must contain at least one of words in the search terms you have specified, all of the words in your search term, or the entire phrase.

f the following conditions are met:

contains  : brca1

QUERY

at least one of the following words

all of the following words

the phrase

6. Add additional conditions if desired. Click Add Condition. Another default condition appears.

**ADVANCED FIELD SPECIFIC SEARCH**

Find all records where  of the following conditions are met:

1. The  contains :

2. The  contains :

7. Click ADD SUB QUERY to group conditions. This displays an indented default condition indicating that all additional conditions or exclusions are applicable only to that group.

**ADVANCED FIELD SPECIFIC SEARCH**

Find all records where  of the following conditions are met:

1. The  contains :

2. The  contains :

3.  of the following conditions are met:

1. The  contains :

**Example:** if you search for all human protein records and all yeast DNA records that do not contain the words "GTP" or "hypothetical" in their annotation, the resulting query is:

((taxid:9606 AND type:aa) OR (taxid:4932 AND type:dna)) NOT (GTP OR hypothetical)

OR

((+(taxid:9606) +(type:aa)) +(taxid:4932) +(type:dna)) !(GTP hypothetical)

8. Click EXCLUSION to create a conditions or multiple conditions that will be excluded from your search.



### ADVANCED FIELD SPECIFIC SEARCH

Find all records where at least one of the following conditions are met:

1. The field any contains at least one of the following words: brca1

2. The field any contains at least one of the following words:


ADD CONDITION
ADD SUB QUERY

... but exclude records where at least one of the following conditions are met:

1. The field any contains at least one of the following words:

2. The field any contains at least one of the following words:

ADD CONDITION
ADD SUB QUERY

**NOTE:** Click the  icon adjacent to a condition or exclusion to remove that condition or exclusion from those you have created.

- Specify if you want the search to match all the conditions you have created or at least one of the conditions.

Find all records where at least one of the

1. The field any all con

ADD CONDITION
ADD SUB QUE

- Click SEARCH. The Advanced Field Specific Search page refreshes, with your search results displayed on the Summary tab and the number of records retrieved for each category displayed on each respective Results tab.



► Field-specific search

**ADVANCED FIELD SPECIFIC SEARCH**

Find all records where  of the following conditions are met:

1. The [field](#)  contains :

Records per page:



Summary	Sequences	Interactions	Complexes	Pathways	GENESEQ
[49180]	[5151]	[164]	[15]	[0]	[43850]

**Your search returned...**

**5151 Public Sequences**

**164 Interactions**

**15 Complexes**


**0 Pathways**

**43850 Patented Sequences (GENESEQ)**

11. If desired, adjust the number of records that will be displayed when any of the Results tabs are clicked. Click the Records per page list and select a number.

Records per page:

12. Click a Results tab to view the records retrieved for that category.
13. Go to any of these chapters for information on viewing and using data on the Results tabs:
- “Sequences” on page 37
  - “Interactions and Complexes” on page 55
  - “GENESEQ” on page 97.

**NOTE:** The search history for your current session can be viewed by clicking the  icon in the Advanced Field Specific Search section. Move the pointer over the icon to see a list of search terms you have used in previous searches and click an item in the list to perform the search again.

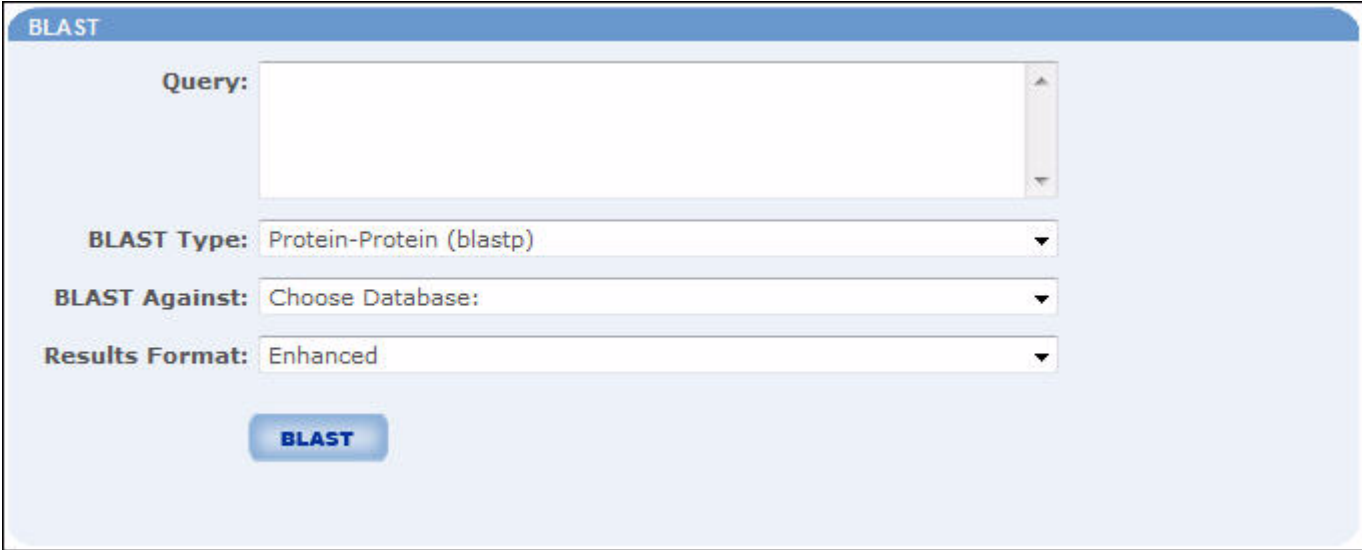
You can also save your search results for later retrieval (see “Saving Search Results” on page 31).

## BLAST Search

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. It compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

### Performing a BLAST Search

1. On the BONDplus Home Page, click the BLAST link (see “Searching From the Home Page” on page 9).
2. The BLAST page appears.



The screenshot shows the BLAST search interface. It has a blue header bar with the word "BLAST" in white. Below the header, there are four input fields with labels to their left: "Query:" followed by a large text input box; "BLAST Type:" followed by a dropdown menu showing "Protein-Protein (blastp)"; "BLAST Against:" followed by a dropdown menu showing "Choose Database:"; and "Results Format:" followed by a dropdown menu showing "Enhanced". Below these fields is a blue button with the word "BLAST" in white.

3. Enter BLAST information in the Query, BLAST type, BLAST against, and Results Format fields.

4. Enter any optional BLAST information in the Sub-sequence, Filters, Expect Cut-Off, Matrix, and Gap Cost fields in the BLAST options section.

**BLAST OPTIONS**

**Sub-sequence:**

From:

To:

**Filters:**

☐ Low Complexity Filter

☐ Human Repeats

☐ Mask For Lookup Table Only

☐ Mask Lower Case

**Expect Cut-Off:**

**Matrix:**

**Gap Cost:**

**BLAST**

5. Click BLAST. The BLAST page refreshes, with your BLAST results displayed.

**NOTE:** The display of results will vary based on what is selected in the Results Format field.

Field Name	Description
Query	The query sequence in FASTA format.
BLAST type	<p>The BLAST program to use:</p> <ul style="list-style-type: none"> <li>• <b>blastp</b>: Takes protein (AA) sequences and compares them against a protein database of your choosing</li> <li>• <b>blastn</b>: Takes nucleotides sequences and compares them against a nucleotide database of your choosing.</li> <li>• <b>blastx</b>: Use the BLAST algorithm to compare the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.</li> <li>• <b>tblastn</b>: Compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands) using the BLAST algorithm.</li> <li>• <b>tblastx</b>: Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database using the BLAST algorithm.</li> </ul>

Field Name	Description
BLAST against	<p>The database of sequences to BLAST against. The list options that are displayed will vary based on whether you are searching against a nucleotide or protein database, and the data you have access to.</p> <p><b>PROTEIN DBS</b></p> <ul style="list-style-type: none"> <li>• <b>Bindall_p.fsa:</b> All protein sequences involved in a BIND Interaction</li> <li>• <b>Geneseq_aa.fsa:</b> Non-redundant database of protein sequences from GEN- ESEQ records</li> <li>• <b>igSeqProt:</b> Immunoglobulin protein sequences</li> <li>• <b>month.aa:</b> All new or revised GenBank CDS translation+PDB+Swis- sProt+PIR+PRF released in the last 30 days</li> <li>• <b>nr:</b> All non-redundant GenBank CDS translations + RefSeq Proteins + PDB + SwissProt + PIR + PRF pataa - Proteins from the Patent division of GenBank</li> <li>• <b>pdbaa:</b> Protein sequences derived from 3-dimensional structure records contained in the Protein Data Bank</li> <li>• <b>swissprot:</b> Last major release of the SWISS-PROT protein sequence data- base.</li> </ul> <p><b>NUCLEOTIDE DBS</b></p> <ul style="list-style-type: none"> <li>• <b>bindall_n.fsa:</b> All nucleotide sequences involved in a BIND Interaction</li> <li>• <b>geneseq_na:</b> Non-redundant database of nucleotide sequences from GEN- ESEQ records</li> <li>• <b>est_human:</b> Human subset of the Expressed Sequence Tag (EST) Gen- Bank division</li> <li>• <b>est_mouse:</b> Mouse subset of the EST division</li> <li>• <b>est:</b> Database of GenBank + EMBL + DDBJ sequences from the EST divi- sion</li> <li>• <b>est_others:</b> Non-Mouse, non-Human subset of the EST division</li> <li>• <b>igSeqNt:</b> Immunoglobulin nucleotide sequences</li> <li>• <b>month.nt:</b> All new or revised GenBank + EMBL + DDBJ + PDB sequences released in the last 30 days.</li> <li>• <b>nt:</b> Redundant database of all GenBank + RefSeq Nucleotides + EMBL + DDBJ + PDB sequences (excluding HTGS0,1,2, EST, GSS, STS, PAT, WGS)</li> <li>• <b>patnt:</b> Nucleotides from the Patent division of GenBank</li> <li>• <b>pdbnt:</b> Nucleotide sequences derived from 3-dimensional structure records contained in the Protein Data Bank</li> <li>• <b>vector:</b> Vector sequences from the Synthetic (syn) division of GenBank</li> <li>• <b>wgs:</b> Whole Genome Shotgun (WGS) sequence database.</li> </ul>

Field Name	Description
Results Format	<p><b>Enhanced:</b> presents BLAST results in an easy to navigate, compact format, with links back to BONDplus for each BLAST hit. For GENESEQ BLAST results, the link takes you to a Sequence Summary page which captures all known GENESEQ Accessions, Patents or NCBI GenInfo IDs that contain the target sequence.</p> <p><b>Text:</b> Standard text output from the NCBI BLAST program. Displays alignments with no links.</p> <p><b>HTML:</b> Enhanced view of the standard text output from the NCBI BLAST program, including links from the summary table to alignment, and direct links to the NCBI for sequence records.</p> <p><b>XML:</b> XML output from the NCBI BLAST program. Intended for users who have development capabilities. The full specification for this format can be found at:  <a href="http://www.ncbi.nlm.nih.gov/dtd/NCBI_BlastOutput.mod.dtd">http://www.ncbi.nlm.nih.gov/dtd/NCBI_BlastOutput.mod.dtd</a></p>
Sub-sequence	<p>A segment of the query sequences can be used in BLAST searching. You can enter the range in the <b>From</b> and <b>To</b> fields to specify the position of this segment. For example, to limit matches to the region from 24 to 200 of a query sequence, enter 24 in the <b>From</b> field and 200 in the <b>To</b> field. If one of the limits you enter is out of range, the intersection of the [From,To] and [1,length] intervals will be searched, where length is the length of the whole query sequence.</p>
Filters: Low Complexity	<p>Select this function to mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton and Federhen (Computers and Chemistry, 1993) or, for BLASTN, by the DUST program of Tatusov and Lipman. Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (for example, hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.</p> <p>Filtering is only applied to the query sequence (or its translation products), not to database sequences. Default filtering is DUST for BLASTN, SEG for other programs.</p> <p>It is common for nothing to be masked by SEG, when applied to sequences in SWISS-PROT or refseq, so filtering should not be expected to always yield an effect. Furthermore, in some cases, PM sequences are masked in their entirety, indicating that the statistical significance of any matches reported against the unfiltered query sequence should be suspect. This will also lead to a search error when default setting is used.</p>
Filters: Human Repeats	<p>Select this option to mask Human repeats (LINE's, SINE's, plus retroviral repeats) and is useful for human sequences that may contain these repeats. Filtering for repeats can increase the speed of a search especially with very long sequences (&gt;100 kb) and against databases which contain large number of repeats (htgs). This filter should be checked for genomic queries to prevent potential problems that may arise from the numerous and often spurious matches to those repeat elements.</p>

Field Name	Description
Filters: Mask for lookup table only	BLAST searches consist of two phases: finding hits based upon a lookup table and then extending them. Select this option to mask only for purposes of constructing the lookup table used by BLAST so that no hits are found based upon low-complexity sequence or repeats (if the repeat filter is checked). The BLAST extensions are performed without masking and so they can be extended through low-complexity sequence.
Filters: Mask lower case	Select this option to cut and paste a FASTA sequence in upper case characters and denote areas you would like filtered with lower case. This allows you to customize what is filtered from the sequence during the comparison to the BLAST databases.
Expect Cut-Off	This setting specifies the statistical significance threshold for reporting matches against database sequences. The default value (10) means that 10 such matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported.
Matrix	A key element in evaluating the quality of a pairwise sequence alignment is the substitution matrix which assigns a score for aligning any possible pair of residues. The matrix used in a BLAST search can be changed depending on the type of sequences with which you are searching.
Gap Cost	The pull down menu shows the Gap Costs for the chosen Matrix. There can only be a limited number of options for these parameters. Increasing the Gap Costs results in alignments which decrease the number of Gaps introduced.

## Constructing Queries

There are two methods to construct a text query. The first is using the Apache Lucene query language in conjunction with the Advanced Field Specific Search option (see “Advanced Field Specific Search” on page 19). The second method is by using the Advanced Field Specific Search option to visually build your query.

### Apache Lucene Query Language

Lucene is a high-performance, full-featured text search engine provided by Apache as part of the Jakarta Project. BONDplus implements the Lucene libraries for its text indexing and searching capabilities. More information about Lucene can be found at: <http://jakarta.apache.org/lucene/docs/index.html>.

#### Special Characters

As with other query languages, the Lucene query language has a set of reserved special characters as follows:

+ - && || ! ( ) { } [ ] ^ " ~ \* ? : \

In order to use any of these special characters within a query, they must be escaped with the ' \' character. For example a query to search for the GO term:

5525[IEA]

should be escaped to:

5525\[IEA\]

## Simple Queries

A simple query can be any word (term). For example, using the term GTP, searches are not case-sensitive. The terms GTP, gtp, and gTP will all return the same results. If more than one term is given as a simple query, the terms are logically OR'ed together (see “Boolean Queries” on page 29).

**Example:** the search terms GTP ATP will return all records containing the word ‘GTP’ OR the word ‘ATP’ in its annotation.

## Phrase Queries

A phrase query provides the ability to search for several words together, denoted by wrapping several search terms within double quotes.

**Example:** an example of a phrase query is “Activation of T Cells”. The results of this query will return all records containing the phrase ‘Activation of T Cells’ within its annotation. If double quotes are not used, the query will return all records containing the word ‘Activation’ OR the word ‘Cell’. As with simple queries, phrase queries are not case-sensitive.

## Boolean Queries

Boolean queries contain one or more terms that alternate with boolean operators (AND, OR, NOT). Boolean operators are case-sensitive and are used to add logical expressions to search terms. Note that a Boolean query can not start with a Boolean operator and must start with a term. The following Boolean query gtp NOT hypothetical would bring back records that contain the word ‘GTP’ in its annotation, but does NOT contain the word ‘hypothetical’. A more complex example that uses a phrase in conjunction with a Boolean operator is “activation of t cells” AND hypothetical. The results of this query would be records that contain the phrase ‘activation of T cells’ AND the word ‘hypothetical’ in its annotation.

Boolean queries can also be written using these modifiers:

Modifier	Description
&&	synonymous with the 'AND' operator
	synonymous with the 'OR' operator
!	synonymous with the 'NOT' operator
+	indicates that the term following must be within the search results
-	indicates that the term following must not be within the search results

**Example:** the query +gtp !hypothetical would return the same results as the query gtp NOT hypothetical.

## Wild Card Queries

A wild card query is made up of search terms that can contain one or more variable characters. The method by which a variable character is indicated is by using a wild card modifier. Valid wild card modifiers are:

Modifier	Description
?	A single variable character
*	Multiple variable characters of unknown length

One restriction of the Lucene query engine is that a wild card modifier can not be the first character of a term.

**Example:** if you use a single variable character modifier in the query A?P, the results could have the word(s) 'ATP', or 'ADP', or 'AMP', or 'ASP', and so on, in the record annotation. If you change the type of wild card modifier used in the previous search, so that the query now becomes A\*P, the resulting records could have the word(s) 'ATP', or 'trAnscriPtion', and so on, in its annotation.

## Viewing Results

When you perform a search, a new page appears displaying all search results on a Summary tab, from which you can then access specific information by clicking on an appropriate tab.

The screenshot shows a search results page with a horizontal tab bar at the top. The tabs are labeled: Summary [102601428], Sequences [89640465], Interactions [217484], Complexes [3739], Pathways [8], and GENESEQ [12739732]. The Summary tab is selected. Below the tabs, the text 'Your search returned...' is followed by a list of results: 89640465 Public Sequences, 217484 Interactions, 3739 Complexes, 8 Pathways, and 12739732 Patented Sequences (GENESEQ).

Search results within a tab are displayed in one of two ways: Summary Record View for multiple search results, and Single Record View when a single search result is returned. Each of these two views can be displayed in a variety of ways.

Data-specific filter options allow you to narrow search results. Query results can also be exported in a number of different formats.

To view and use results, see the following chapters:


- “Sequences” on page 37
- “Interactions and Complexes” on page 55
- “GENESEQ” on page 97.



## Saving Search Results

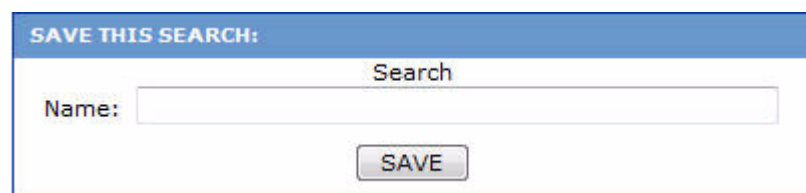
You can save any search and its results for retrieval in any subsequent session. All query text, views, and filters are preserved. You can save up to 50 search results.

### To save your search results

1. After performing a search and viewing your results (see “Viewing Results” on page 30), click the save icon  in the Search box.



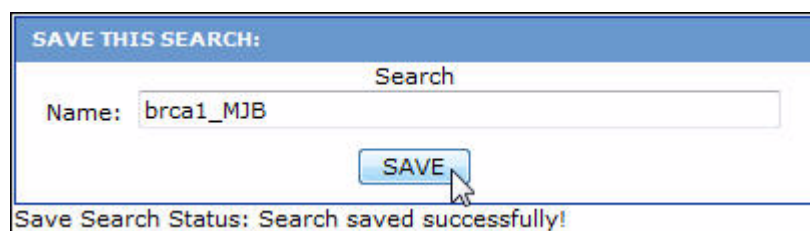
2. The Save This Search box appears.



3. Enter any text in the Name field that describes your search.

**NOTE:** Save your search with an appropriate name that will make it easily distinguishable from other saved searches.

4. Click Save.
5. A notification is given indicating if the saving of the search results was successful or not.

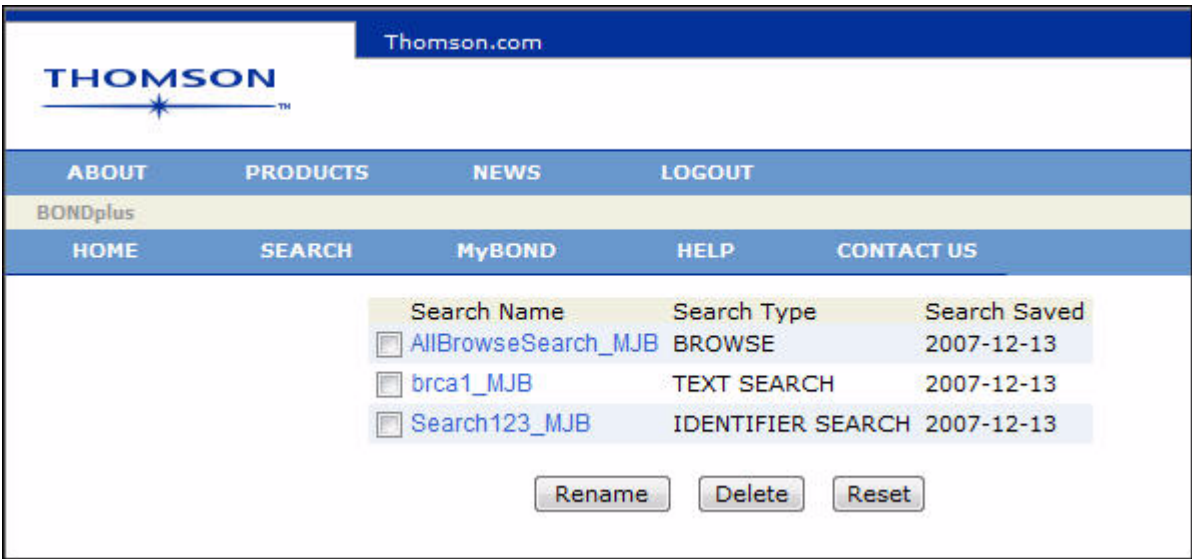


To retrieve previously saved search results

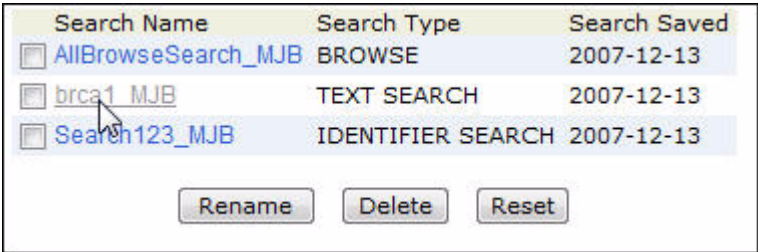
- 1. Click MyBond on the BONDplus menu and choose Saved Searches.



- 2. A list of your previously saved searches appears.



- 3. Click on a Search Name to retrieve its previously saved results.



- 4. Your search results are displayed (see “Viewing Results” on page 30).

### To rename a previously saved search

1. To rename a search, mark the box adjacent to the Search Name you want to change and click Rename.

Search Name	Search Type	Search Saved
<input type="checkbox"/> AllBrowseSearch_MJB	BROWSE	2007-12-13
<input checked="" type="checkbox"/> brca1_MJB	TEXT SEARCH	2007-12-13
<input type="checkbox"/> Search123_MJB	IDENTIFIER SEARCH	2007-12-13

2. A pop-up window appears (you may be asked first by your browser to confirm if you want a scripted window to appear. Follow the browser's directions to allow scripted windows to appear). Enter the new name for the search and click OK.

Explorer User Prompt

Script Prompt:

Please enter the new name for selected search item:

brca1\_MJB\_priority

OK Cancel

3. The saved searches list refreshes, displaying the new name for your search.

Search Name	Search Type	Search Saved
<input type="checkbox"/> AllBrowseSearch_MJB	BROWSE	2007-12-13
<input type="checkbox"/> brca1_MJB_priority	TEXT SEARCH	2007-12-13
<input type="checkbox"/> Search123_MJB	IDENTIFIER SEARCH	2007-12-13

### To delete one or more saved searches

1. To delete one or more saved searches, mark the box adjacent to the Search Name you want to delete and click Delete. **NOTE:** To clear all marked boxes, click Reset.

Search Name	Search Type	Search Saved
<input checked="" type="checkbox"/> AllBrowseSearch_MJB	BROWSE	2007-12-13
<input type="checkbox"/> brca1_MJB_priority	TEXT SEARCH	2007-12-13
<input checked="" type="checkbox"/> Search123_MJB	IDENTIFIER SEARCH	2007-12-13

2. A confirmation box appears, asking if you are sure you want to delete the saved searches you have selected. Click OK.

3. The saved searches list refreshes, with the deleted saved searches removed.

Search Name	Search Type	Search Saved
<input type="checkbox"/> brca1_MJB_priority	TEXT SEARCH	2007-12-13
<div>RenameDeleteReset</div>		





## Chapter 5

### Sequences

---

BONDplus sequence information is primarily derived from biological sequence data deposited at the National Center for Biotechnology Information (NCBI). Supplementary data is also collected from UniProt and related NCBI partners. Databases include GenBank, RefSeq, UniProt/SWISS-PROT, UniProt/TrEMBL, and all other sources of NCBI sequence content.

BONDplus also stores annotation for sequences, including Gene Ontology annotation, taxonomy identifiers, associated structures from the Protein Data Bank, redundant sequences, precomputed RPS-BLAST and BLAST results, database cross-references, Online Mendelian Inheritance in Man identifiers, Entrez Gene locus information, and complete genomes.

BONDplus public sequence content serves as a foundation for all other sequence-derived database including BINDplus and GENESEQ annotation.

### Performing a Query for Sequences

The text search engine supporting BONDplus allows for both broad and narrow queries.

To retrieve all sequence records relating to a specific database accession or identifier – NCBI GI, UniProt/SWISS-PROT ID or NCBI Entrez Gene ID – use the Identifier Search interface with the identifier of interest (see “Identifier Search” on page 15).

To perform a broad search for all BONDplus records referencing ‘brca1’:

`+brca1`

To retrieve all sequence records by NCBI Gene Name ‘brca1’:

`+gene.name:brca1`

You may not know a specific identifier associated with an internal ontology or taxonomy. For example, to restrict a query by taxonomy, you can use either the ‘taxname’ field to search by name or ‘taxid’ to search by taxonomy identifier. Therefore, you can perform the search for ‘*Saccharomyces cerevisiae*’ records which reference the GO ID for RNA polymerase activity several different ways:

`+taxname:"Saccharomyces cerevisiae" +go.all:"RNA polymerase"`

OR

`+taxid:4732 +go:3899`

Use the Sequence Filters to complement complex searches (see “Using the Sequences Filters” on page 38). To retrieve all RefSeq records referencing to human brca1, select RefSeq in the “Filter by Database Source” Sequence Filter and use the following text query:

`+brca1 +taxid:9606`


To assist you in constructing advanced queries of your own design, remember to use the Advanced Field Specific Search query builder tool (see “Advanced Field Specific Search” on page 19).

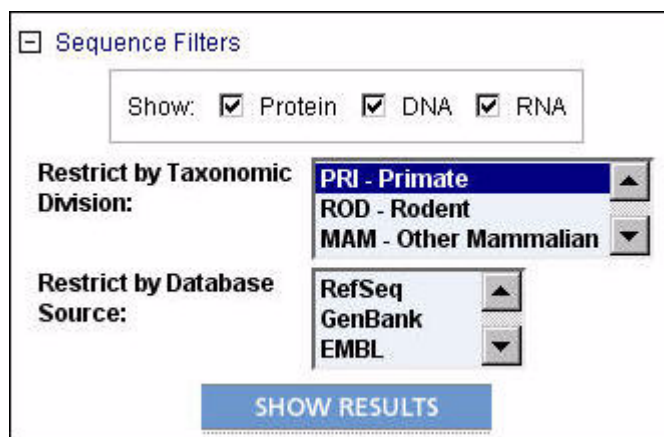
## Using the Sequences Filters

Data-specific filters are available under the Sequences tab. They allow you to further narrow your search.

These filters within the Sequence tab allow you to restrict the contents being browsed by the type of molecule (Protein, DNA, RNA), by the National Center for Biotechnology Information (NCBI) assigned division (Primate, Rodent, Other Mammalian, and so on), and by the originating database source as defined by the NCBI (RefSeq, GenBank, EMBL, and so on).

### How to use the Sequence Filters

1. Click the  icon. The filter options are displayed.



☐ Sequence Filters

Show: ☒ Protein ☒ DNA ☒ RNA

Restrict by Taxonomic Division:

- PRI - Primate
- ROD - Rodent
- MAM - Other Mammalian

Restrict by Database Source:

- RefSeq
- GenBank
- EMBL

**SHOW RESULTS**

2. Mark or clear each check box in the Show section as desired.
3. Click one or more items in the Restrict by Taxonomic Division list to include those items in your search or do not select any items to include all in your search.
4. Click one or more items in the Restrict by Database Source list to include those items in your search or do not select any items to include all in your search.
5. Click SHOW RESULTS. Your search results are updated.



## Sequence Filter Field Descriptions

Field Name	Description
Restrict by Database Source	<p>This field allows you to narrow your search by selecting specific originating database sources as defined by the National Center for Biotechnology Information (NCBI). Select one or more of the following:</p> <p><b>RefSeq</b> – The Reference Sequence (RefSeq) database stores a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products. .</p> <p><b>GenBank</b> – An annotated collection of all publicly available nucleotide and amino acid sequences.</p> <p><b>EMBL</b> – The European Molecular Biology Laboratory (EMBL) Nucleotide Sequence Database constitutes Europe's primary nucleotide sequence resource.</p> <p><b>DDBJ</b> – The DNA Data Bank of Japan (DDBJ) is the sole DNA data bank in Japan, which is officially certified to collect DNA sequences from researchers and to issue the internationally recognized accession number to data submitters.</p> <p><b>PDB</b> – The Protein Data Bank.</p> <p><b>SWISS-PROT</b> – A curated protein sequence database which provides a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, and so on).</p> <p><b>PIR</b> – The Protein Information Resource (PIR) is a comprehensive, non-redundant, and expertly annotated protein sequence database</p> <p><b>PRF</b> – The Protein Resource Foundation.</p>
Restrict by Taxonomic Division	<p>This field allows you to narrow your search by selecting specific taxonomic divisions assigned by the National Center for Biotechnology Information (NCBI). Select one or more of the following:</p> <p><b>PRI - Primate</b></p> <p><b>ROD - Rodent</b></p> <p><b>MAM - Other Mammalian</b></p> <p><b>VRT - Other Vertebrate</b></p> <p><b>INV - Invertebrate</b></p> <p><b>PLN - Plant, Fungal, Algae</b></p> <p><b>BCT - Bacterial</b></p> <p><b>VRL - Viral</b></p> <p><b>PHG - Bacteriophage</b></p> <p><b>SYN - Synthetic</b></p>
Show	<p>This field allows you to include or exclude protein, DNA, or RNA molecules in or from your search. Mark each check box as desired.</p>

## Sequences Summary Record View

After performing a sequences query (see “Performing a Query for Sequences” on page 37) or a search (see “Searching for Data” on page 9), the Sequences Summary Record View presents you with a list of sequence-related records under the Sequence tab.

<div> <div>Summary [65]</div> <div><b>Sequences [40]</b></div> <div>Interactions [21]</div> <div>Complexes [0]</div> <div>Pathways [0]</div> <div>GENESEQ [4]</div> <div>65 records fetched in 0.46 second(s)</div> </div>			
+ Sequence Filters		View: Single Line	Export Results: Select an Export Format
<div> <div>1 2</div> <div>Skip To Page: 1 GO</div> </div>			
GI	Accession	Description	Organism
30721612	AY277623	No description provided	Rhizopus stolonifer
21744316	AY114331	Uncultured candidate division BRC1 bacterium clone LD1-PB16 16S ribosomal RNA gene, partial sequence.	uncultured candidate division BRC1 bacterium
21744303	AY114318	Uncultured candidate division BRC1 bacterium clone LD1-PA27 16S ribosomal RNA gene, partial sequence.	uncultured candidate division BRC1 bacterium
21744300	AY114315	Uncultured candidate division BRC1 bacterium clone LD1-PA21 16S ribosomal RNA gene, partial sequence.	uncultured candidate division BRC1 bacterium
11611252	AJ390438	Uncultured soil bacterium PBS-II-1 partial 16S rRNA gene	uncultured soil bacterium PBS-II-1
11611253	AJ390439	Uncultured soil bacterium PBS-II-2 partial 16S rRNA gene	uncultured soil bacterium PBS-II-2
11611254	AJ390440	Uncultured soil bacterium PBS-II-4 partial 16S rRNA gene	uncultured soil bacterium PBS-II-4
11611258	AJ390444	Uncultured soil bacterium PBS-II-15 partial 16S rRNA gene	uncultured soil bacterium PBS-II-15
11611259	AJ390445	Uncultured soil bacterium PBS-II-34 partial 16S rRNA gene	uncultured soil bacterium PBS-II-34
11611263	AJ390449	Uncultured soil bacterium PBS-III-4 partial 16S rRNA gene	uncultured soil bacterium PBS-III-4
11611265	AJ390451	Uncultured soil bacterium PBS-III-16 partial 16S rRNA gene	uncultured soil bacterium PBS-III-16
11611266	AJ390452	Uncultured soil bacterium PBS-III-18 partial 16S rRNA gene	uncultured soil bacterium PBS-III-18
11611268	AJ390454	Uncultured soil bacterium PBS-III-24 partial 16S rRNA gene	uncultured soil bacterium PBS-III-24
11611271	AJ390457	Uncultured soil bacterium PBS-III-29 partial 16S rRNA gene	uncultured soil bacterium PBS-III-29
11611283	AJ390469	Uncultured soil bacterium PBS-74 partial 16S rRNA gene	uncultured soil bacterium PBS-74
11611295	AJ390481	Uncultured soil bacterium PRR-11 partial 16S rRNA gene	uncultured soil bacterium PRR-11
156079844	EU089719	Xanthomonas campestris pv. campestris strain BrC1 16S ribosomal RNA gene, partial sequence.	Xanthomonas campestris pv. campestris
68009934	NM_001021081	Schizosaccharomyces pombe 972h- hypothetical protein (SPBC582.05c), partial mRNA	Schizosaccharomyces pombe 972h-
83835046	DQ330729	Uncultured candidate division BRC1 bacterium clone 07N26B 16S ribosomal RNA gene, partial sequence.	uncultured candidate division BRC1 bacterium
83835043	DQ330726	Uncultured candidate division BRC1 bacterium clone 09D2Y74 16S ribosomal RNA gene, partial sequence.	uncultured candidate division BRC1 bacterium
<div> <div>1 2</div> <div>Skip To Page: 1 GO</div> </div>			

The information displayed for each record in this view is not the full annotation available for the record. For every type of Sequence Summary Record View, a link (typically, the NCBI GI displayed in bold red text) is provided to access various levels of annotation detail for that record in the Sequences Single Record View format (see “Sequences Single Record View” on page 43).

## Sequences Summary Record View Types

There are different types of Sequences Summary Record Views available within the Sequence Tab: Single Line, Ontoglyph, and Proteoglyph.

### Single Line

Provides basic details of a record (GI, Accession, Description, Organism). This is the default view when the Sequences tab is displayed. To display the Single Line view, select **Single Line** in the **View** list on the **Sequences** tab.



The Single Line view is displayed:

GI	Accession	Description	Organism
<b>30721612</b>	AY277623	No description provided	Rhizopus stolonifer
<b>21744316</b>	AY114331	Uncultured candidate division BRC1 bacterium clone LD1-PB16 16S ribosomal RNA gene, partial sequence.	uncultured candidate bacterium
<b>21744303</b>	AY114318	Uncultured candidate division BRC1 bacterium clone LD1-PA27 16S ribosomal RNA gene, partial sequence.	uncultured candidate bacterium

Column	Description
GI	A unique identifier assigned to all sequence depositions to the NCBI. Integer-based.
Accession	A unique identifier assigned to a biological polymer sequence submitted to a sequence database.
Description	A brief description of the sequence record.
Organism	The name of the source species for this sequence.

### Ontoglyph

Provides basic details of a record (GI, Accession, Description), as well as the ontoglyphs associated with a record. To display the Ontoglyph view, select Ontoglyph in the View list on the Sequences tab.

Summary  
[12803452]

Sequences  
[72963]

Interactions  
[217484]

Complexes  
[3739]

Pathways  
[8]

Sequence Filters

View:

Single Line

Single Line

Ontoglyph

Proteoglyph

Export Results

1 -- 8 9 10 11 1

Skip To Page:

11

GO

The Ontoglyph view is displayed:

GI	Accession	Ontoglyphs	Description
58331120	NP_001009923		hypothetical protein LOC29058 isoform 1 [Homo sapiens]
123173777	NP_066384		keratin associated protein 5-8 [Homo sapiens]
57114137	NP_001009138		semenogelin II [Pan troglodytes]
55663220	XP_521094		PREDICTED: similar to Hepatocellular carcinoma-associated antigen 127 [
55663160	XP_521086		PREDICTED: Kruppel-like factor 8 [Pan troglodytes]
89111943	NP_009181		Kruppel-like factor 8 [Homo sapiens]
55663298	XP_521102		PREDICTED: similar to hypothetical protein MGC21416 [Pan troglodytes]

Column	Description
GI	A unique identifier assigned to all sequence depositions to the NCBI. Integer-based.
Accession	A unique identifier assigned to a biological polymer sequence submitted to a sequence database.
Ontoglyphs	The ontoglyphs associated with the record. Ontoglyphs provide a visually condensed pictorial representation of gene ontology information associated with that record (for more information see “Appendix A About Ontoglyphs” on page 113.
Description	A brief description of the sequence record.

Proteoglyph

Provides basic details of a record (GI, Accession, Description), as well as the Proteoglyphs associated with a record. To display the Proteoglyph view, select Proteoglyph in the View list on the Sequences tab.



The Proteoglyph view is displayed:

GI	Accession	Proteoglyphs	Description
58331120	NP_001009923		hypothetical protein LOC29058 isoform 1 [Homo sapiens]
123173777	NP_066384		keratin associated protein 5-8 [Homo sapiens]
57114137	NP_001009138		semenogelin II [Pan troglodytes]
55663220	XP_521094		PREDICTED: similar to Hepatocellular carcinoma-associated antigen 127 [Pan troglodytes]
55663160	XP_521086		PREDICTED: Kruppel-like factor 8 [Pan troglodytes]

Column	Description
GI	A unique identifier assigned to all sequence depositions to the NCBI. Integer-based.
Accession	A unique identifier assigned to a biological polymer sequence submitted to a sequence database.
Proteoglyphs	The proteoglyphs associated with the record. Proteoglyphs provide a visually condensed pictorial representation of domain annotation associated with that record. For more information see “Appendix B About Proteoglyphs” on page 137.
Description	A brief description of the sequence record.

## Sequences Single Record View

The Sequences Single Record View provides annotation details for a single record. Levels of annotation detail depend on the type of Sequences Single Record View selected.

To display a Single Record View for a Sequence, click a record in the GI column of a Sequences Summary Record View (see “Sequences Summary Record View” on page 40) on the **Sequences** tab.

GI	Accession	
30721612	AY277623	No
<u>21744316</u>	AY114331	Un
21744303	AY114318	Un
21744300	AY114315	Un

The Sequences Single Record View is displayed for the item. The Identifier Search section displays the identifier values for the record you have opened. The default view type is HTML. For information on the different view types for a single record, see “Sequences Single Record View Types” on page 44.

## Sequences Single Record View Types

There are different types of Sequences Single Record Views available:

- [HTML](#)
- [FASTA Sequence](#)
- [GenBank](#)
- [ASN.1 BioSeq](#)
- [XML BioSeq](#)

### HTML

Provides an integrated view of all annotation associated with a record in BONDplus. To display the HTML view, select **HTML** in the **View** list when viewing a record in a Sequences Single Record View.



The HTML view is displayed:

Identifier search - Seqhound

IDENTIFIER SEARCH

NCBI GI

Records per page:

SEARCH

RESET

Summary [1]

Sequences [1]

Interactions [0]

Complexes [0]

Pathways [0]


GENESEQ [0]


+ Sequence Filters View:  Export Results:

[Expand all] [Collapse all]

Record Details

**Protein: BRCA1/BRCA2-containing complex, subunit 3 isoform 2 [Homo sapiens]**





Accession: RefSeq: [NP\\_001018065](#)

GI: **64762484**

+ Synonyms: 9 synonym(s) found

Organism: [human](#) - TaxID 9606

Division: PRI - Primate

+ Literature: 3 publication(s) found

Sequence Length: 291

+ Sequence: FASTA format [BLAST this sequence...](#)

+ Gene Ontology: 6 GO reference(s) found.


Encoding Nucleotide:	GI	DB	Accession	
<b>64762483</b>	<input type="text" value="Find this molecule in..."/>	RefSeq	<a href="#">NM_001018055</a>	Homo sapiens containing cc (BRCC3), tra

+ Conserved Domains: 3 conserved domain reference(s) found.

+ Redundant Group: 3 identical protein(s) found.

+ Similar Proteins: 83 similar protein(s) found. (evalue = 0.01)

+ Database Cross-references: 3 cross-reference(s) found

By default, the HTML Sequences Single Record View shows a collapsed form of the information, with summary descriptions of each collapsed section, and an expander icon  to show the annotation of that particular section.

By clicking Expand All or Collapse All, you can expand and collapse all annotation for the given record.

[Expand all] [Collapse all]

**Record Details**

**DNA: Uncultured candidate division BRC**

**Accession:** GenBank Accession: [AY114331](#)

**GI:** **21744316** [Find this molecule in...](#)

**Organism:** [uncultured candidate division BRC1 bacterium](#) - TaxID 200297

**Division:** BCT - Bacterial

**Literature:** • PMID: [12620817](#)

**Sequence Length:** 1376

**Sequence:**

1	TGGCGAACTG	GTGAGTAACG	GGTGGGTTAT	GTGCCTCTTA	GTGG
61	ACAGAGGCTA	ATACCGCATA	CGGCCCTTCG	GGGTTAAAGA	CTTC
121	GCCCGCTCCC	GATTAGCTTG	TTGGTGGGGT	AACGGCCTAC	CAAG
181	GGTCTGAGAG	GATGGTCACC	CACACTGGGA	CTGAGATACG	GCCC
241	CAGCAGTCTA	GAGGCTTCCA	CAATGGACGA	AAGTCTGATG	GAGC

Hyperlinks are provided for quick access to the original source of annotation.

**Record Details**

**DNA: Uncultured candidat**

**Accession:** GenBank Accession: [AY114331](#)

**GI:** **21744316** [Find this molecule in...](#)

**Organism:** [uncultured candidate division BRC1 bacterium](#) - TaxID 200297

**Division:** [BCT - Bacterial](#)

Field	Description
GI	A unique identifier assigned to all sequence depositions to the NCBI. Integer-based.
Accession	A unique identifier assigned to a biological polymer sequence submitted to a sequence database.
Synonyms	Alternate names for the sequence.
Organism	The name of the source species for this sequence.
Division	The broad NCBI division to which this sequence is assigned.
Literature	The literature references which apply to this sequence, displayed using the references' NCBI PubMed ID number.
Sequence Length	The length of the sequence.



Field	Description
Sequence	See “To BLAST a Sequence” on page 47.
Gene Ontology	Aspect GO Id Description Source Evidence Code
Encoding Nucleotide	GI: A unique identifier assigned to all sequence depositions to the NCBI. Integer-based. DB: Accession: A unique identifier assigned to a biological polymer sequence submitted to a sequence database. Title:
Conserved Domains	CDD Id Domain Id Identify Query Seq. Hit Seq. Alignment Length E-value Score Bit Score Coverage
Redundant Group	GI: A unique identifier assigned to all sequence depositions to the NCBI. Integer-based. DB Accession: A unique identifier assigned to a biological polymer sequence submitted to a sequence database. Organism: The name of the source species for this sequence. Title
Similar Proteins	GI: A unique identifier assigned to all sequence depositions to the NCBI. Integer-based. DB Accession: A unique identifier assigned to a biological polymer sequence submitted to a sequence database. Organism: The name of the source species for this sequence. Title
Database Cross-references	Database Identifier

### To BLAST a Sequence

1. Click BLAST this sequence in the Sequence field of an HTML Sequences Single Record View.

Record Details	
<b>DNA: Uncultured cand</b>	
<b>Accession:</b>	GenBank Accession: <a href="#">AY114331</a>
<b>GI:</b>	<b>21744316</b> <a href="#">Find this molecule in...</a>
<b>Organism:</b>	uncultured candidate division BRC1 bacterium - TaxID 2002
<b>Division:</b>	BCT - Bacterial
<b>Literature:</b>	1 publication(s) found
<b>Sequence Length:</b>	1376
<b>Sequence:</b>	FASTA format <a href="#">BLAST this sequence...</a>

2. The BONDplus BLAST page appears.

HOME	SEARCH	HELP	CONTACT US
▶ BOND BLAST			
<b>Query:</b>	<div> <div>&gt;gi 21744316 gb AY114331.1  Uncultured candidate division BRC1 bacterium clone LD1-PB16 16S ribosomal RNA gene, partial sequence TGGCGAACTGGTGAGTAACGGGTGGGTTATGTGCCTCTTAGTGGGGGATAACCTCTGGAA ACAGAGGCTAATACCGCATACGGCCCTTCGGGGTTAAAGACTTCGGTCGCTATGAGAGCA GCCCCGCTCCCGATTAGCTTGTTGGTGGGGTAACGGCCTACCAAGGCGATGATCGGTAGGT GGTCTGAGAGGATGGTCACCCACACTGGGACTGAGATACGGCCAGACTCCTACGGGGGG ACCAAGGCGATGATCGGTAGGTGGTCTGAGAGGATGGTCACCCACACTGGGACTGAGATACGGCCAGACTCCTACGGGGGG</div> </div>		
<b>BLAST type:</b>	Protein-Protein (blastp)		
<b>BLAST against:</b>	nr		
<b>Results Format:</b>	Text		
+ BLAST Options			
<a href="#">BLAST</a>			

### FASTA Sequence

Displays the sequence in FASTA format.

To display the FASTA Sequence view, select FASTA Sequence in the View list when viewing a record in a Sequences Single Record View.

<b>View:</b>	<div> <div>HTML</div> <div>HTML</div> <div>FASTA Sequence</div> <div>Genbank</div> <div>ASN.1 Bioseq</div> <div>XML Bioseq</div> </div>
--------------	---

The FASTA sequence view is displayed:

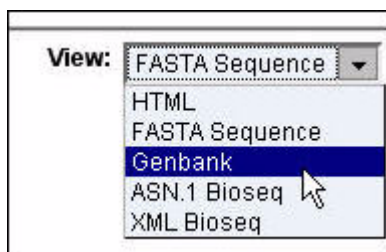
The screenshot shows a web interface with a top navigation bar containing six tabs: Summary [1], Sequences [1], Interactions [0], Complexes [0], Pathways [0], and GENESEQ [0]. The 'Sequences' tab is active. Below the tabs is a '+ Sequence Filters' link. The main content area displays a FASTA sequence entry for 'BRCA1/BRCA2-containing complex, subunit 3 isoform 2 [Homo sapiens]' with the following sequence:

```
>gi|64762484|ref|NP_001018065.1| BRCA1/BRCA2-containing complex, subunit 3 isoform 2 [Homo sapiens]
MAVQVVQAVQAVHLESDAFLVCLNHALSTEKEEVMGLCIGELNDDTRSDSKFAYTGTEMR
TVAEKVDAVRIVHIHVSIIILRRSDKRKDRVEISPEQLSAASTEAEERLAELTGRPMRVVGW
YHSHPHITVWPSHVDVRTQAMYQMMDQGFVGLIFSCFIEDKNTKTGRVLYTCFQSIQAQK
SSEYERIEIPIHIVPHVTIGKVCLESAVELPKILCQEEQDAYRRIHSLTHLDSVTKIHNH
SVFTKNLCSQMSAVSGPLLQWLEDRLQEQNQQHLQELQKEEELMQELSSLE
```

## GenBank

Displays the original GenBank Flat File format.

To display the Genbank view, select Genbank in the View list when viewing a record in a Sequences Single Record View.



The Genbank view is displayed:

Summary [1]		Sequences [1]		Interactions [0]		Complexes [0]		Pathways [0]		GENESEQ [0]	
+ Sequence Filters											
LOCUS	NP_001018065	291 aa	linear	PRI 17-OCT-2006							
DEFINITION	BRCA1/BRCA2-containing complex, subunit 3 isoform 2 [Homo sapiens].										
ACCESSION	NP_001018065										
VERSION	NP_001018065.1 GI:64762484										
DBSOURCE	REFSEQ: accession NM_001018055.1										
KEYWORDS	.										
SOURCE	Homo sapiens (human)										
ORGANISM	Homo sapiens										
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										
	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;										
	Catarrhini; Hominidae; Homo.										
REFERENCE	1 (residues 1 to 291)										
AUTHORS	Dong,Y., Hakimi,M.A., Chen,X., Kumaraswamy,E., Cooch,N.S.,										
	Godwin,A.K. and Shiekhattar,R.										
TITLE	Regulation of BRCC, a holoenzyme complex containing BRCA1 and										
	BRCA2, by a signalosome-like subunit and its role in DNA repair										
JOURNAL	Mol. Cell 12 (5), 1087-1099 (2003)										
PUBMED	14636569										
REFERENCE	2 (residues 1 to 291)										
AUTHORS	Fisch,P., Forster,A., Sherrington,P.D., Dyer,M.J. and Rabbitts,T.H.										
TITLE	The chromosomal translocation t(X;14)(q28;q11) in T-cell										
	pro-lymphocytic leukaemia breaks within one gene and activates										
	another										
JOURNAL	Oncogene 8 (12), 3271-3276 (1993)										
PUBMED	8247530										
REFERENCE	3 (residues 1 to 291)										
AUTHORS	Verma,S., Levinson,S., Taylor,S., Shapiro,A. and Strachan,T.										

## ASN.1 BioSeq

Displays the NCBI BioSeq structure in Abstract Syntax Notation One (ASN.1) ASCII format. A BioSeq is the data structure used at NCBI to store sequence annotation.

For more information on Abstract Syntax Notation One, please visit:

<http://asn1.elibel.tm.fr/en/>.

For more information on ASN.1 and its usage at NCBI, please visit:

<http://www.ncbi.nlm.nih.gov/Sitemap/Summary/asn1.html>.

For more information on the BioSeq data structure, please visit:

<http://ncbi.nih.gov/IEB/ToolBox/SDKDOCS/BIOSEQ.HTML>.

To display the ASN.1 BioSeq view, select ASN.1 BioSeq in the View list when viewing a record in a Sequences Single Record View.



The ASN.1 BioSeq view is displayed:

 A screenshot of the 'Sequences' tab in the 'Sequences Single Record View'. The 'Sequences' tab is active, showing a count of [1]. Below the tabs, there is a 'Sequence Filters' section with a plus icon. The main content area displays the ASN.1 BioSeq structure for a specific sequence, showing fields like 'id', 'other', 'accession', 'version', 'gi', 'descr', 'molinfo', 'title', 'create-date', 'std', 'year', 'month', 'day', 'inst', 'repr', 'mol', and 'length'.
 

```

Bioseq ::= {
  id {
    other {
      accession "NP_001018065" ,
      version 1 } ,
      gi 64762484 } ,
  descr {
    molinfo {
      biomol peptide } ,
      title "BRCA1/BRCA2-containing complex, subunit 3 isoform 2 [Homo sapiens]" ,
      create-date
      std {
        year 2005 ,
        month 5 ,
        day 13 } } ,
  inst {
    repr raw ,
    mol aa ,
    length 291 ,
  }
}
  
```

### XML BioSeq

Displays the NCBI BioSeq structure in Extensible Markup Language (XML) format. A BioSeq is the data structure used at NCBI to store sequence annotation.

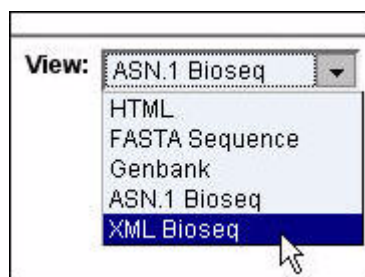
For more information on XML please visit:

<http://www.w3.org/XML/>.

For more information on the BioSeq data structure, please visit:

<http://ncbi.nih.gov/IEB/ToolBox/SDKDOCS/BIOSEQ.HTML>.

To display the XML BioSeq view, select XML Bioseq in the View list when viewing a record in a Sequences Single Record View.



The XML BioSeq view is displayed:

Summary [1]

**Sequences [1]**

Interactions [0]

Complexes [0]

Pathways [0]

GENESEQ [0]

Sequence Filters

<Bioseq>

<Bioseq\_id>

<Seq-id>

<Seq-id\_other>

<Textseq-id>

<Textseq-id\_accession>NP\_001018065</Textseq-id\_accession>

<Textseq-id\_version>1</Textseq-id\_version>

</Textseq-id>

</Seq-id\_other>

</Seq-id>

<Seq-id>

<Seq-id\_gi>64762484</Seq-id\_gi>

</Seq-id>

</Bioseq\_id>

<Bioseq\_descr>

<Seq-descr>

<Seqdesc>

<Seqdesc\_molinfo>

<MolInfo>

<MolInfo\_biomol value="peptide">8</MolInfo\_biomol>

</MolInfo>

</Seqdesc\_molinfo>

</Seqdesc>

<Seqdesc>

<Seqdesc\_title>BRCA1/BRCA2-containing complex, subunit 3 isoform 2 [Homo sapiens]</Seqdesc\_title>

</Seqdesc>

<Seqdesc>

<Seqdesc\_create-date>

<Date>

<Date\_std>

<Date-std>

<Date-std\_year>2005</Date-std\_year>

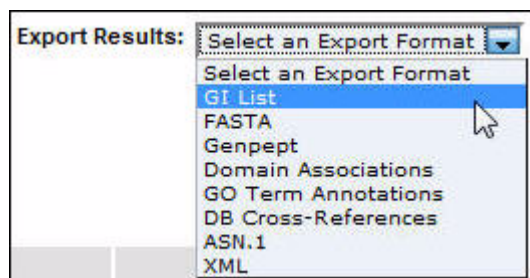


## Exporting Sequence Search Data

You can export your Sequence search results data from either a Sequence Summary or Single Record View.

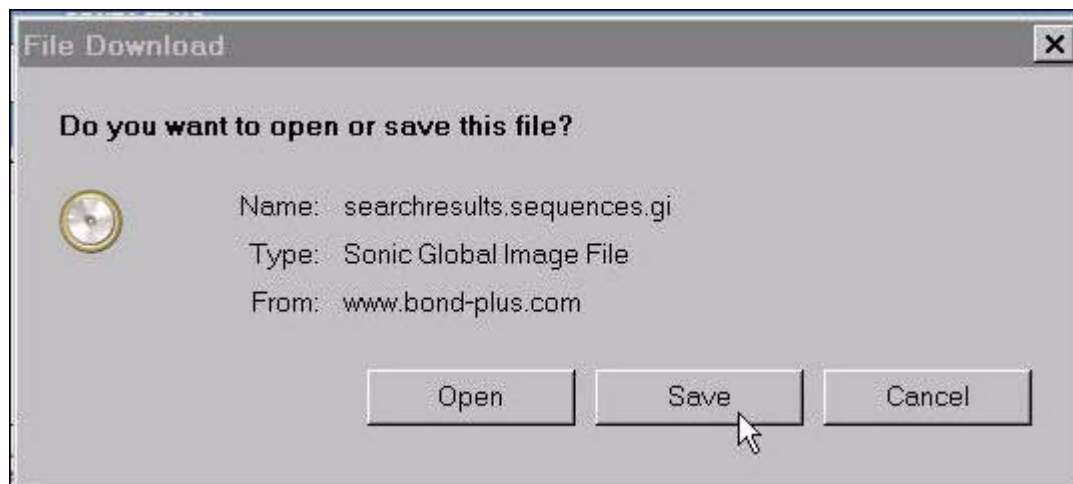
### Exporting Data From a Sequence Summary or Single Record View

1. To export search results data displayed in a Sequence Summary or Single Record View, on the Sequence tab choose a format from the Export Results list.



**NOTE:** For information about the formats available, see “Sequence Export Formats” on page 53.

2. A confirmation box appears asking if you want to open or save the file.



**NOTE:** The confirmation box may appear different than that shown above, depending on the browser you are using (for example, Internet Explorer or Mozilla Firefox). Also, your browser might attempt to block the download. Check your browser’s settings to correct this (see “Browser FAQ” on page 148).

3. Choose the Save or Save As option.
4. Depending on the browser you are using, the file is downloaded into your browser’s default download location or a dialog box appears allowing you to choose the folder to which you want to download the file.

### Sequence Export Formats

The following export formats are available in BONDplus:

Format	Description
GI List	A list of all NCBI GenInfo IDs returned. One entry per line.
FASTA	The sequence data contained by all component records returned in FASTA format. For more information, see <a href="http://www.ncbi.nlm.nih.gov/blast/fasta.shtml">http://www.ncbi.nlm.nih.gov/blast/fasta.shtml</a> .
Genpept	The records in GenBank Flat File Format. For more information, see <a href="http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html">http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html</a> .
Domain Associations	Pre-computed Protein Domain Associations for all Proteins returned per query. Format is GI, Domain ID 1, Domain Name 1, Domain ID 2, Domain Name 2... Comma Separated Values. One GI per line.
GO Term Annotations	Pre-computed Gene Ontology Associations for all Protein GIs returned per query. Format is GI, GO Term Number, GTN... Comma Separated Values. One GI per line.
DB Cross-References	All known Database Cross References per GI. Format is GI, Database 1: Database Identifier 1, Database 2: Database Identifier 2... Comma Separated Values. One GI per line.
ASN.1	A Bioseq-set in NCBI ASN.1 format containing all sequences returned by your query.
XML	A Bioseq-set in NCBI Sequence XML format, containing all sequences returned by your query.



## Chapter 6

# Interactions and Complexes

---

The Biomolecular Interaction Network Database plus (BINDplus) is a collection of records documenting molecular interactions. BINDplus is a component database of BONDplus. BINDplus is exclusively curated from peer-reviewed scientific literature.

BINDplus is an interaction database with three classifications for molecular associations: molecules that associate with each other to form interactions, molecular complexes that are formed from one or more interaction(s) and pathways that are defined by a specific sequence of two or more interactions.

### Interactions

A BINDplus record represents an interaction between two objects that is believed to occur in a living organism. A biological object can be a protein, DNA, RNA, ligand, molecular complex, gene, photon or an unclassified biological entity.

BINDplus records are created for interactions which have been shown experimentally and published in at least one peer-reviewed journal. A record also references any papers with experimental evidence that support or dispute the associated interaction.

Interactions are the basic units of BINDplus and can be linked together to form molecular complexes or pathways.

### Complexes

A molecular complex is a collection of two or more molecules that associate to form a functional unit in a living organism. In BINDplus, complexes are represented as molecular complex objects, formed by linking one or more interaction records.

Molecular complex records are supplemented with additional information such as complex topology and the number of subunits (BINDplus objects) making up the complex.

### Performing a Query for Interactions

The text search engine supporting BONDplus allows for both broad and narrow queries.

To retrieve all interaction records relating to a specific database accession or identifier – BIND ID, NCBI Entrez Gene ID or UniProt/SWISS-PROT ID – use the Identifier Search interface with the identifier of interest (see “Identifier Search” on page 15).

To perform a broad search for all BONDplus records referencing ‘brca1’:

+brca1

To retrieve all interaction records which reference 'brca1' in the short label or interaction description:

shortlabel:brca1 interaction.description:brca1

As BINDplus interactions are derived from peer reviewed journals, you may wish to retrieve records based on the associated publication. To retrieve all interaction records derived from PMID @@@@:

+pub.pmid:10783165

To retrieve all interaction records from the journal 'Genes and Development', search on the Medline title:

+pub.journal.title:"GENES DEV"

To retrieve all interaction records from publications with the MeSH term 'diabetes':

+pub.meshterm:diabetes

Sometimes, you may not know a specific identifier associated with an internal ontology or taxonomy. For example, to restrict a query by taxonomy, you can use either the 'taxname' field to search by name or 'taxid' to search by the broader taxonomy identifier. Therefore, you can perform the search for 'Saccharomyces cerevisiae' records which reference the GO ID for RNA polymerase activity several different ways:

+taxname:"Saccharomyces cerevisiae" +GO:"RNA polymerase"

OR

+taxid:4732 +GO:3899

Use the Interaction Filters to complement complex searches (see "Using the Interactions and Complexes Filters" on page 56). To retrieve all Low Throughput protein-protein interaction records referencing to human brca1, leave only the "LTP" checkbox selected under "Include", and the "interactions", and "protein-protein interactions" checkboxes selected under "Show" in the Interaction Filter and use the following text query:

+brca1 +taxid:9606

To further restrict the query above further to the subset of records with evidence deriving from Immunoprecipitation, use the following text query:

+brca1 +taxid:9606 +interaction.exp\_system:"Immunoprecipitation"

To assist you in constructing advanced queries of your own design, remember to use the Advanced Field Specific Search query builder tool (see "Advanced Field Specific Search" on page 19).

## Using the Interactions and Complexes Filters

Data-specific filters are available under the Interactions and Complexes tabs. They allow you to further narrow your search.

These filters allow you to restrict your search according to interactor type (Protein-protein, Genetic, and so on), record source (Low Throughput Curation, High Throughput Curation, and so on), BINDplus Division, and BINDplus Picks confidence score. The GO term and domain filtering of interaction results is also available through the Ontoglyph Analysis and Proteoglyph Analysis buttons.



## Interactions and Complexes Filters Field Descriptions

Field Name	Description
Include	<p><b>LTP Records:</b> BINDplus records hand-curated by manual extraction of data from published articles. Data generated by standard wet-lab research and is off high value.</p> <p><b>HTP Records:</b> BINDplus records created by computer scripts which are written to automatically create records and import into BINDplus. Data is generated from high-throughput, automated experiments, such as yeast two-hybrid, which generate large scale datasets.</p> <p><b>Database Imports:</b> Interaction data from object or other interaction databases imported into BINDplus through computer scripting methods. Records may deviate significantly from BINDplus standard text and data representation and may or may not have been derived from peer-reviewed published research.</p> <p><b>Provisional Records:</b> BINDplus records imported either from other databases or from high-throughput research with only elemental or preliminary information. Records are updated and released as completed as fully-annotated BINDplus records.</p>
PICKS cut-off	<p>Protein Interaction Confidence Kernel Scores (PICKS) represents a numerical confidence measure for <i>Saccharomyces cerevisiae</i> protein-protein interactions in BINDplus. It is recommended that PICKS interaction scores less than one constitute true interactions suitable for further investigation and analysis.</p> <p>Enter a numeric value.</p>
Show	<p><b>interactions:</b> choose this selection to display all BINDplus interaction records returned by the inputted search criteria. Interaction records contain detailed information on two interacting biomolecules, including complete annotation of the interactors, experimental evidence describing the interaction, binding site sequence ranges, chemical action and chemical states as well as reference to the publication demonstrating the interaction.</p> <p><b>complexes:</b> choose this selection to display all BINDplus complex records returned by the inputted search criteria. Complex records contain information on the interaction records making up the complex, the number of subunits involved, a brief description of the components of the complex and a reference to the publication demonstrating the interaction.</p> <p><b>protein-protein interactions:</b> choose this selection to display all BINDplus interaction records where both interacting partners are proteins.</p> <p><b>nucleic acid interactions:</b> choose this selection to display all BINDplus interaction records where one or both of the interacting partners are nucleic acids (DNA, RNA).</p> <p><b>genetic interactions:</b> choose this selection to display all BINDplus interaction records which are genetic interactions, filtering out all physical interactions.</p> <p><b>small molecule interactions:</b> choose this selection to display all BINDplus interaction records where one or both of the interacting partners are small molecules.</p>

### How to Create an Ontoglyph Filter

1. Click ONTOGLYPH ANALYSIS (or EDIT ONTOGLYPHS if a previous filter was performed) when using the data-specific filters available under the Interactions and Complexes tabs (see “Using the Interactions and Complexes Filters” on page 56). The ontoglyphs that are present in your initial search are shown according to location, binding, and function (for information about ontoglyphs see “Appendix A About Ontoglyphs” on page 113).

### Create Ontoglyph Filter

- These tables show how frequently each of the following glyphs occur in your search results.
- **Filter** your search results by clicking on a glyph:

**One click = include this glyph in my filter.**

**Two clicks = include this glyph in my filter only if it co-occurs in Molecule A and B.**

**Three clicks = exclude this glyph in my filter.**

[What are Ontoglyphs?](#)

**Location:**

3	2	2	2	1	1	1	1	1

**Binding:**

4	4	2	1	1	1

**Function:**

6	3	3	3	3	2	2	2	2	2	2
2	1	1	1							

FILTER

RESET

2. Choose the ontoglyphs you want to include or exclude in your search.

**NOTE:** Click RESET to clear all the ontoglyphs that have been selected.

3. Click FILTER. Your search results are updated.

### How to Create a Proteoglyph Filter

Click PROTEOGLYPH ANALYSIS (or EDIT PROTEOGLYPHS if a previous filter was performed) while using the data-specific filters available under the Interactions and Complexes tabs (see “Using the Interactions and Complexes Filters” on page 56). The proteoglyphs that are present in your initial search are shown (for information about proteoglyphs see “Appendix B About Proteoglyphs” on page 137).

### Create Proteoglyph Filter

- These tables show how frequently each of the following glyphs occur in your search results.
  - Filter** your search results by clicking on a glyph:

✓









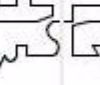







One click = *include* this glyph in my filter.

✓

Two clicks = *include* this glyph in my filter only if it co-occurs in Molecule A and B.

✗

Three clicks = *exclude* this glyph in my filter.

											
21	21	4	4	2	2	2	2	2	2	2	2
											
2	2	2	2								

FILTER
RESET

4. Choose the proteoglyphs you want to include or exclude in your search.

**NOTE:** Click RESET to clear all the ontoglyphs that have been selected.

5. Click FILTER. Your search results are updated.

## Interactions Summary Record View

After performing an interactions query (see “Performing a Query for Interactions” on page 55) or a search (see “Searching for Data” on page 9), the Interactions Summary Record View presents you with a list of interaction-related records under the Interactions tab.



Summary [49180]		Sequences [5151]		Interactions [164]		Complexes [15]		Pathways [0]		GENESEQ [43850]	
+ Interaction/Complex/Pathway Filters						View: <div>Single Line</div>		Export Results: <div>Select an Ex</div>		Visualize Results: <div>Select a visualization opt</div>	
						1 2 3 4 5 6 7 8 9					
						Skip To Page: <div>1</div>		GO			
Identifier		Mol A Short Label		Mol B Short Label		Experimental Evidence					
Interaction 12725		BRCA1		LMO4		two hybrid test, immunoprecipitation, affinity chromatography					
Interaction 12597		BRCA1		c-Jun		immunoprecipitation, affinity chromatography					
Interaction 12641		BRCA1		RbAp46		other, immunostaining, immunoprecipitation, affinity chromatograph					
Interaction 12718		BRCA1		p300		immunostaining, immunoprecipitation					
Interaction 12596		BRCA1		JnB		two hybrid test, immunoprecipitation, affinity chromatography					
Interaction 12640		BRCA1		Rb		other, immunostaining, immunoprecipitation, affinity chromatograph					
Interaction 12613		BRCA1		MYC		two hybrid test, immunoprecipitation, affinity chromatography					
Interaction 12642		BRCA1		RbAp48		immunostaining, immunoprecipitation, affinity chromatography					
Interaction 12682		BRCA1		Oct-1		immunoprecipitation, affinity chromatography					
Interaction 12598		BRCA1		JunD		two hybrid test, immunoprecipitation, affinity chromatography					
Interaction 12729		LMO4, BRCA1, and CtIP		LDB1		immunoprecipitation					
Interaction 295569		TOP2A		BRCA1		immunoprecipitation, affinity chromatography					
Interaction 12600		BRCA1		FANCA		two hybrid test, immunoprecipitation					
Interaction 12603		BRCA1		ATR		immunoprecipitation					
Interaction 12744		BRCA1		ACCA		immunoprecipitation, affinity chromatography					
Interaction 12683		BRCA1		NF-YA		immunoprecipitation, affinity chromatography					
Interaction 12590		BRCA1		RAD50		immunostaining, two hybrid test, sucrose gradient sedimentation, n specified, immunoprecipitation, affinity chromatography					
Interaction 12636		BRCA1		E2F-1		other, affinity chromatography					
Interaction 12599		BRCA1		CK2beta		other, two hybrid test, affinity chromatography					

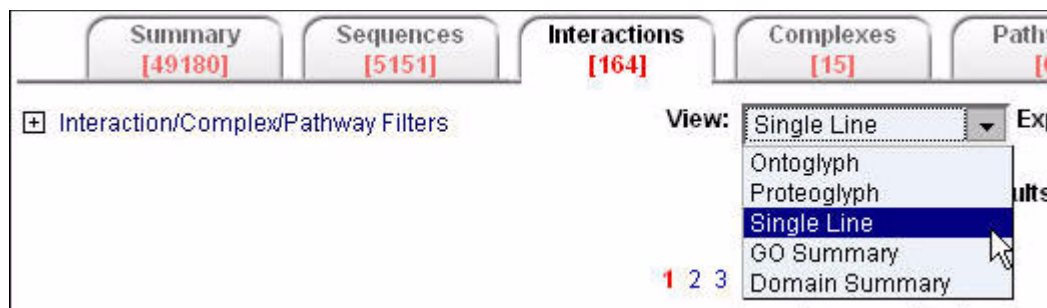
The information displayed for each record in this view is not the full annotation available for the record. For every type of Interaction Summary Record View, a link (typically, located in the Identifier column and displayed in bold red text) is provided to access various levels of annotation detail for that record in the Sequences Single Record View format (see “Interactions Single Record View” on page 69).

## Interactions Summary Record View Types

There are different types of Interactions Summary Record Views available within the Interactions Tab: Single Line, Ontoglyph, Proteoglyph, GO Summary, and Domain Summary.

### Single Line

Provides basic details of a record. This is the default view when the Interactions tab is displayed. To display the Single Line view, select Single Line in the View list on the Interactions tab.



The Single Line view is displayed.

Identifier	Mol A Short Label	Mol B Short Label	Experimental Evidence
<b>Interaction 12725</b>	BRCA1	LMO4	two hybrid test, immunoprecipitation, affinity chromatograph
<b>Interaction 12597</b>	BRCA1	c-Jun	immunoprecipitation, affinity chromatography
<b>Interaction 12641</b>	BRCA1	RbAp46	other, immunostaining, immunoprecipitation, affinity chrom
<b>Interaction 12718</b>	BRCA1	p300	immunostaining, immunoprecipitation

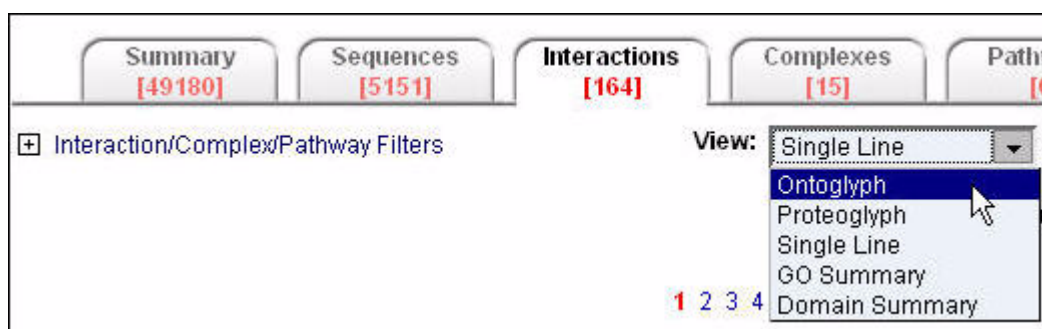
Column	Description
Identifier	BINDplus integer identifier unique to each BINDplus interaction or complex record.
Mol A Short Label	Abbreviation by which Mol A is known and recognized. This will either be the name used by the authors in the publication or the short label for the molecule at Entrez.



Column	Description
Mol B Short Label	Abbreviation by which Mol B is known and recognized. This will either be the name used by the authors in the publication or the short label for the molecule at Entrez.
Experimental Evidence	Indicates which experimental methods or systems were used in the publication to demonstrate the interaction between MolA and MolB.
Taxonomy	Describes the taxon to which the interacting molecules belong.

## Ontoglyph

Provides basic details of a record as well as the ontoglyphs associated with a record. To display the Ontoglyph view, select Ontoglyph in the View list on the Interactions tab.



The Ontoglyph view is displayed:

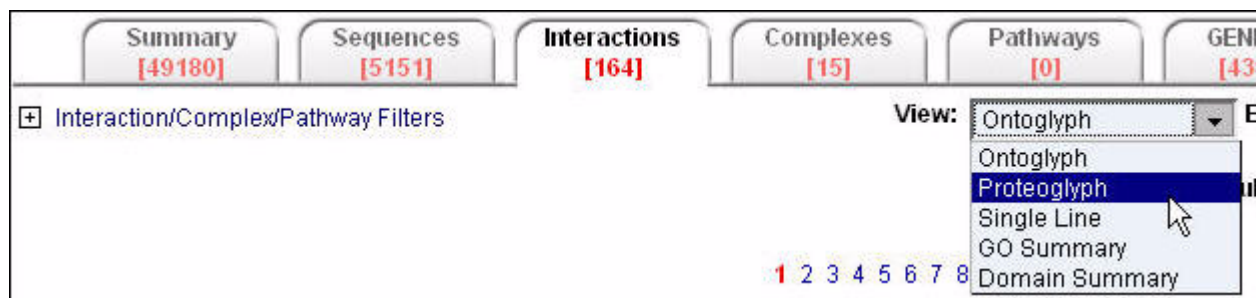
Identifier	Molecule A Short Label	Molecule A Ontoglyphs	Molecule B Short Label	Molecule B Ontoglyphs
Interaction 12725	BRCA1		LMO4	
Interaction 12597	BRCA1		c-Jun	
Interaction 12641	BRCA1		RbAp46	
Interaction 12718	BRCA1		p300	

Column	Description
Identifier	BINDplus integer identifier unique to each BINDplus interaction or complex record.
Mol A Short Label	Abbreviation by which Mol A is known and recognized. This will either be the name used by the authors in the publication or the short label for the molecule at Entrez.

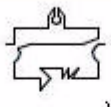
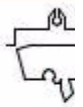




Column	Description
Molecule A Ontoglyphs	Displays Ontoglyphs associated with Molecule A. Ontoglyphs provide a visually condensed pictorial representation of gene ontology information associated with that record (for more information see “Appendix A About Ontoglyphs” on page 113.
Mol B Short Label	Abbreviation by which Mol B is known and recognized. This will either be the name used by the authors in the publication or the short label for the molecule at Entrez.
Molecule B Ontoglyphs	Mol B Ontoglyphs: Displays Ontoglyphs associated with Molecule B. The ontoglyphs associated with the record. Ontoglyphs provide a visually condensed pictorial representation of gene ontology information associated with that record (for more information see “Appendix A About Ontoglyphs” on page 113.
Taxonomy	Describes the taxon to which the interacting molecules belong.

### Proteoglyph

Provides basic details of a record as well as the Proteoglyphs associated with a record. To display the Proteoglyph view, select Proteoglyph in the View list on the Interactions tab.



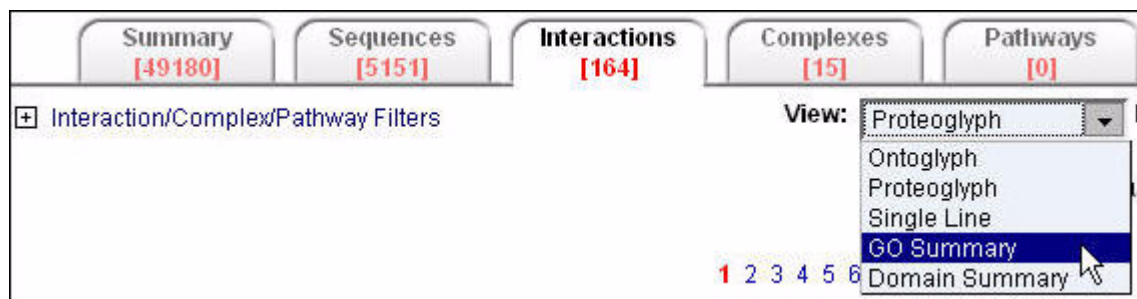
The Proteoglyph view is displayed:

Identifier	Molecule A Short Label	Molecule A Proteoglyphs	Molecule B Short Label	Molecule B Proteoglyphs
Interaction 12725	BRCA1	 x2	LMO4	
Interaction 12597	BRCA1	 x2	c-Jun	
Interaction 12641	BRCA1	 x2	RbAp46	

Column	Description
Identifier	BINDplus integer identifier unique to each BINDplus interaction or complex record.
Mol A Short Label	Abbreviation by which Mol A is known and recognized. This will either be the name used by the authors in the publication or the short label for the molecule at Entrez.
Molecule A Proteoglyphs	Displays Proteoglyphs associated with Molecule A. Proteoglyphs provide a visually condensed pictorial representation of domain annotation associated with that record. For more information see "Appendix B About Proteoglyphs" on page 137
Mol B Short Label	Abbreviation by which Mol B is known and recognized. This will either be the name used by the authors in the publication or the short label for the molecule at Entrez.
Molecule B Proteoglyphs	Displays Proteoglyphs associated with Molecule B. Proteoglyphs provide a visually condensed pictorial representation of domain annotation associated with that record. For more information see "Appendix B About Proteoglyphs" on page 137.
Taxonomy	Describes the taxon to which the interacting molecules belong.

### GO Summary

Provides basic details of a record. To display the GO Summary view, select GO Summary in the View list on the Interactions tab.



The GO Summary view is displayed:

Interaction 12725		Homo sapiens Full BIND Record		Launch Viewer: <span>Select Below ▼</span>	
Molecule	Description	Molecular Function	Cellular Component	Biological Process	Expe
BRCA1 2 alias(es)	Breast Cancer 1, early onset, breast and ovarian cancer susceptibility protein 1; nuclear phosphoprotein; mutated in familial breast and ovarian cancers; tumour suppressor that participates in DNA repair and/or DNA damage-responsive pathways. Contains an NH2-terminal RING finger domain, nuclear localization signals (NLS) and the C-terminal acidic transcriptional activation domain with two C-terminal BRCA1 COOH-terminal (BRCT) repeats; OMIM ID: 113705	<ul style="list-style-type: none"> <li>DNA binding</li> <li>damaged DNA binding</li> <li>DNA repair protein</li> <li>transcription coactivator activity</li> <li>ubiquitin-protein ligase activity</li> <li>protein binding</li> <li>tumor suppressor</li> <li>zinc ion binding</li> <li>tubulin binding</li> <li>apoptosis regulator activity</li> <li>transcriptional activator activity</li> </ul>	<ul style="list-style-type: none"> <li>ubiquitin ligase complex</li> <li>extracellular space</li> <li>intracellular</li> <li>nucleus</li> <li>transcription factor complex</li> <li>gamma-tubulin ring complex</li> </ul>	<ul style="list-style-type: none"> <li>cell cycle checkpoint</li> <li>DNA repair</li> <li>regulation of DNA repair</li> <li>regulation of transcription from RNA polymerase II promoter</li> <li>regulation of transcription from RNA polymerase III promoter</li> <li>apoptosis</li> <li>response to DNA damage stimulus</li> <li>DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator</li> <li>protein ubiquitination</li> <li>regulation of cell proliferation</li> <li>regulation of apoptosis</li> </ul>	<ul style="list-style-type: none"> <li>two</li> <li>immu</li> <li>affin</li> <li>chro</li> </ul>

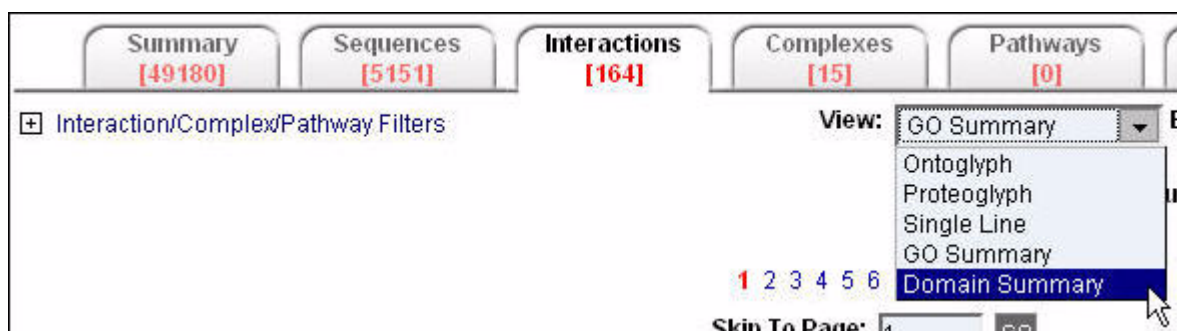
Column	Description
Molecule	Provides the short label and all short label aliases of the two interacting molecules. Short labels and aliases are derived from the publication, Entrez and species-specific databases (for example, OMIM, MGI, RGD, SGD, Flybase).
Description	Provides a list of the expanded (long-form) names of each of the interacting molecules. Expanded names are derived from the publication, Entrez and species-specific databases (for example, OMIM, MGI, RGD, SGD, Flybase). A short description of the function and characteristics of the molecule may also be provided.



Column	Description
Molecular Function	Lists the GO Molecular Function terms associated with each of the interacting molecules. GO's Molecular Function Ontology describes a series of activities, such as DNA binding or transcription factor activity, occurring at the molecular level.
Cellular Component	Lists the GO Cellular Component terms associated with each of the interacting molecules. GO's Cellular Component Ontology describes compartments and physical locations and structures within a cell or organism.
Biological Process	Lists the GO Biological Process terms associated with each of the interacting molecules. GO's Biological Process Ontology describes a series of events accomplished by one or more ordered assemblies of molecular functions. Examples include cell cycle checkpoint and DNA repair.
Experiment(s)	Indicates which experimental methods or systems were used in the publication to demonstrate the interaction between Mol A and Mol B.
Links	<p>Selection of link displays all records and data associated with the GenBank Identifier (GI) in BINDplus, SeqHound or NCBI.</p> <ul style="list-style-type: none"> <li>• <b>Taxon- BIND:</b> this link displays all records in BINDplus that contain that taxon group.</li> <li>• <b>Pubmed:</b> this link displays the publication abstract at Pubmed.</li> <li>• <b>BIND Publication Search:</b> returns all BINDplus records containing that publication.</li> </ul>

### Domain Summary

Provides basic details of a record. To display the Domain Summary view, select Domain Summary in the View list on the Interactions tab.



The Domain Summary view is displayed:

<b>Interaction 12725</b>		Homo sapiens	Full BIND Record	Launch Viewer: Select Below
Molecule	Description			
BRCA1 + 2 alias(es)	Breast Cancer 1, early onset, breast and ovarian cancer susceptibility protein 1; nuclear phosphoprotein; mutated in familial breast and ovarian cancers; tumour suppressor that participates in DNA repair and/or DNA damage-responsive pathways. Contains an NH2-terminal RING finger domain, nuclear localization signals (NLS) and the C-terminal acidic transcriptional activation domain with two C-terminal BRCA1 COOH-terminal (BRCT) repeats; OMIM ID: 113705			
	ILIM domain only 4 is a cysteine-rich protein containing two LIM domains. Its			

Column	Description
Molecule	Provides the short label and all short label aliases of the two interacting molecules. Short labels and aliases are derived from the publication, Entrez and species-specific databases (for example, OMIM, MGI, RGD, SGD, Flybase).
Description	Provides a list of the expanded (long-form) names of each of the interacting molecules. Expanded names are derived from the publication, Entrez and species-specific databases (for example, OMIM, MGI, RGD, SGD, Flybase). A short description of the function and characteristics of the molecule may also be provided.
Pfam	This section provides information on any Pfam domains within the molecule. The Protein Family Database (Pfam) is a collection of multiple sequence alignments and profile hidden Markov models (HMMs). Each Pfam HMM represents a protein family or domain. Information is derived from NCBI's CDD database.
SMART	This section provides information on any SMART domains within the molecule. Simple Modular Architecture Research Tool (SMART) allows the identification and annotation of domains and the analysis of domain architectures. Information is derived from NCBI's CDD database.
CDD	This section provides information on any conserved domains within the molecule. The Conserved Domain Database is derived from Pfam and SMART databases and contains information on conserved domains.
COG	This section provides information on any COG domain within the molecule. Clusters of Orthologous Groups of proteins (COGs) Information is a phylogenetic classification of the proteins encoded complete genomes of bacteria, archaea and eukaryotes. COG is derived from NCBI's CDD database.
Experiment(s)	Indicates which experimental methods or systems were used in the publication to demonstrate the interaction between Mol A and Mol B.
Links	<p>Selection of link displays all records and data associated with the GenBank Identifier (GI) in BINDplus, SeqHound or NCBI.</p> <ul style="list-style-type: none"> <li>• <b>Taxon- BIND:</b> this link displays all records in BINDplus that contain that taxon group.</li> <li>• <b>Pubmed:</b> this link displays the publication abstract at Pubmed.</li> <li>• <b>BIND Publication Search:</b> returns all BINDplus records containing that publication.</li> </ul>

## Visualize Results

Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data. For more information on how to use this feature, go to <http://www.cytoscape.org>.

## Interactions Single Record View

The Interactions Single Record View provides annotation details for a single record. Levels of annotation detail depend on the type of Interactions Single Record View selected.

To display a Single Record View for an Interaction, click a record in the **Identifier** column of an Interactions Summary Record View (see “Interactions Summary Record View” on page 60) on the **Interactions** tab.

Identifier	Mol A Short Label
<a href="#">Interaction 12725</a>	BRCA1
<a href="#">Interaction 12597</a>	BRCA1
<a href="#">Interaction 12641</a>	BRCA1
<a href="#">Interaction 12718</a>	BRCA1
<a href="#">Interaction 12586</a>	BRCA1

The Interactions Single Record View is displayed for the item. The Identifier Search section displays the identifier values for the record you have opened. The default view type is HTML. For information on the different view types for a single record, see “Interactions Single Record View Types” on page 69.

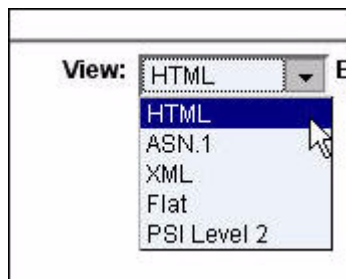
## Interactions Single Record View Types

There are different types of Interactions Single Record Views available:

- [HTML](#)
- [ASN.1](#)
- [XML](#)
- [Flat](#)
- [PSI Level 2](#)

### HTML

Provides an integrated view of all annotation associated with a record in BONDplus. To display the HTML view, select HTML in the View list when viewing a record in a Interactions Single Record View.



The HTML view is displayed:

**Identifier search - BIND**

---


### IDENTIFIER SEARCH

BIND Id

Records per page:

**SEARCH**

**RESET**




---

Summary Sequences [0] Interactions [1] Complexes [0] Pathways [0] GENESECS [0]

+ Interaction/Complex/Pathway Filters View:  Export Results:

[Expand all] [Collapse all]

### BIND Interaction

**BIND ID:** 12725 [Visualize using PDB](#)

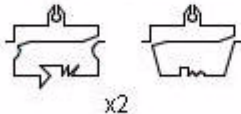
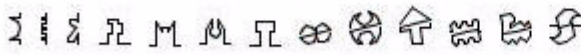

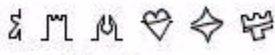
**Interaction Description:** The LIM domain protein LMO4 interacts with BRCA1

**DIVISION:** BIND Metazoa


+ **Publications:** 1 [View all publications \[NCBI\]](#)

**Record References:** 1 [Complex\(es\)](#)

**Date Last Released:** April 22, 2005

Molecule A	Molecule B
<p><b>Protein : BRCA1</b></p>  <p>x2</p>  <p><a href="#">View Sources</a></p> <p><b>Description:</b> Breast Cancer 1, early onset, breast and ovarian cancer susceptibility protein 1; nuclear phosphoprotein; mutated in familial breast and ovarian cancers; tumour suppressor that participates in DNA repair and/or DNA damage-responsive pathways. Contains an NH2-terminal RING finger domain, nuclear localization signals (NLS) and the C-terminal acidic transcriptional activation domain with two C-terminal BRCA1</p>	<p><b>Protein : LMO4</b></p>  <p>x2</p>  <p><a href="#">View Sources</a></p> <p><b>Description:</b> LIM domain only 4 is a protein containing two LIM domains. It is characterized by a GC-rich region with multiple ATTT motifs in the variant transcript missing region has been identified and confirmed because of the region. Member of the subfamily of LIM domain transcription factors. 100 mouse and human LMO</p>



By default, the HTML Interactions Single Record View shows a collapsed form of the information, with summary descriptions of each collapsed section, and an expander icon  to show the annotation of that particular section.

By clicking Expand All or Collapse All, you can expand and collapse all annotation for the given record.

Expand all
Collapse all

## BIND Interaction

**BIND Id:** 12725

**Interaction Description:** The LIM domain protein LMO4 interacts with BRCA1

**Division:** BIND Metazoa

**Publications:** 1.

**PMID:** 11751867

**Opinion:** Support

**Journal:** J Biol Chem 2002/3/7;277(10):7849-7856.

**Article Title:** The LIM domain protein LMO4 interacts with the cofactor Ct suppressor BRCA1 and inhibits BRCA1 activity.

**Authors:**

- Sum EY
- Peng B
- Yu X
- Chen J
- Byrne J

Hyperlinks are provided for quick access to the original source of annotation.

RNF53

Automatically Retrieved Annotation

Cross References:

- GenBank: AAA73985
- GenBank: AAC37594
- GenBank: AAP12647
- RefSeq: NP\_009225
- RefSeq: NP\_009226
- RefSeq: NP\_009227
- Swiss-Prot: P38398 [BRC1\_HUMAN]
- undefined: 672 []

GO Terms:

Molecular Function	Cellular Component
<ul style="list-style-type: none"> <li>DNA binding</li> <li>damaged DNA binding</li> <li>DNA repair protein</li> <li>transcription</li> <li>coactivator activity</li> </ul>	<ul style="list-style-type: none"> <li>ubiquitin ligase complex</li> <li>extracellular space</li> <li>intracellular</li> <li>nucleus</li> <li>transcription factor</li> </ul>

Field	Description
BIND Id	BINDplus integer identifier unique to each BINDplus interaction or complex record.
Interaction Description	Provides a brief description of the interaction, indicating the short labels of the two molecules shown to interact.
Division	A categorization assigned to a record based on the origin of the organism of the interacting molecules or of the interaction data. Examples include divisions based on database imports and taxonomy guidelines.
Publications	<ul style="list-style-type: none"> <li>• <b>PMID:</b> Pubmed Identifier. An unique integer identifier designated by NCBI's Pubmed Database given to a publication.</li> <li>• <b>Opinion:</b> Indicates whether the experimental evidence within the publication supports or disputes the occurrence of the interaction in vivo.</li> <li>• <b>Journal:</b> The journal in which the publication appeared. Information is derived from NCBI's Pubmed Database.</li> <li>• <b>Article Title:</b> The complete text of the article's title. Information is derived from NCBI's Pubmed Database.</li> <li>• <b>Authors:</b> List of publishing authors. Information is derived from NCBI's Pubmed Database.</li> <li>• <b>Affiliation:</b> List of affiliated institute of the last author listed on the publication. Information is derived from NCBI's Pubmed Database.</li> <li>• <b>Abstract:</b> Publication abstract or summary. Information is derived from NCBI's Pubmed Database.</li> </ul>
Record References	Indicates if the records contains link out reference to other BINDplus interaction or complex records. For example, if the displayed interaction record has a BINDplus complex listed as an interacting molecule.
Date Last Released	The most recent date the record was updated and publicly released.
Molecule A and Molecule B	<ul style="list-style-type: none"> <li>• <b>Description:</b> Provides the short label and all short label aliases of the two interacting molecules. Short labels and aliases are derived from the publication, Entrez and species-specific databases (for example, OMIM, MGI, RGD, SGD, Flybase).</li> <li>• <b>NCBI GenInfo ID:</b> NCBI's integer identifier assigned to each unique sequence listed within the NCBI sequence database.</li> <li>• <b>NCBI Entrez Gene ID:</b> NCBI's integer identifier associated with a unique gene locus of a given organism.</li> <li>• <b>Origin:</b> Indicates the organism or species from which the molecule is derived.</li> <li>• <b>Aliases:</b> Lists short labels by which the molecule is known and recognized. These are derived from Entrez Gene and by the short labels listed in the publication.</li> </ul>

Field	Description
Automatically Retrieved Annotation	<ul style="list-style-type: none"> <li>• <b>Cross References:</b> Contains and links to accession numbers of sequences redundant to the listed GI contained within sequence databases such as NCBI's RefSeq, SwissProt, EMBL and GenBank.</li> <li>• <b>GO Terms:</b> Contain information and links on any GO Terms associated with the molecule. Developed by the Gene Ontology (GO) Consortium, GO terms are generated as a part of a controlled vocabulary, or ontology, to describe gene or gene products of an organism. The ontology is listed under the areas of Cellular Component, Biological Processes and Molecular Function.</li> <li>• <b>Domains:</b> Contains information and links to any conserved domains within the molecule.</li> </ul>
Experimental Evidence	<ul style="list-style-type: none"> <li>• <b>General Experimental Evidence:</b> Describes the general nature of the experimental evidence, in vivo or in vitro.</li> <li>• <b>Experimental System Used:</b> Describes the experimental method or system that used to demonstrate the interaction as detailed in the publication.</li> <li>• <b>General Description:</b> A text description providing more specific details of the experiment performed in the publication which demonstrated the interaction. Information will include the details of the experiment, any mutants used and figure references.</li> <li>• <b>Bait Condition:</b> Specifies which molecule was used as bait in a bait versus prey experiment.</li> <li>• <b>Publications:</b> References the publication in which the experiment demonstrating the interaction was performed.</li> <li>• <b>Experimental Forms:</b> Describes any modified molecule used to perform the experiment. Examples include GST-tagged, truncation mutants, radio labelled molecules, or wild type molecules from a different species.</li> </ul>
Binding Sites	Describes the sequence range, domain information and mutant effect related to an experimentally-demonstrated binding site on molecules A and B. The binding site field also references the experimental evidence which demonstrated the binding site as well as includes a figure reference.

## ASN.1

Displays the interaction in Abstract Syntax Notation One (ASN.1) ASCII format.

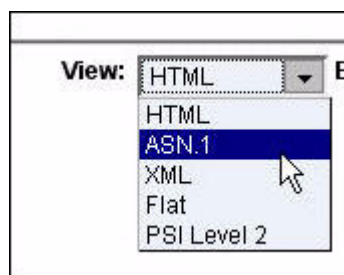
For more information on Abstract Syntax Notation One, please visit:

<http://asn1.elibel.tm.fr/en/>.

For more information on ASN.1 and its usage at NCBI, please visit:

<http://www.ncbi.nlm.nih.gov/Sitemap/Summary/asn1.html>.

To display the ASN.1 view, select ASN.1 in the View list when viewing a record in an Interactions Single Record View.



The ASN.1 view is displayed:

```
BIND-Interaction ::= {  
  date std {  
    year 2005,  
    month 4,  
    day 22  
  },  
  updates {  
    {  
      date std {  
        year 2003,  
        month 10,  
        day 28  
      },  
      descr "Public Release"  
    },  
    {  
      date std {  
        year 2004,  
        month 9,  
        day 7  
      },  
      descr "Public Release"  
    }  
  },  
  iid 12725,  
  a {  
    short-label "BRCA1",  
    other-names {  
      "PSCP",  
      "RNF53"  
    },  
    id protein {  
      gi 8552303,  
      db "P"
  }
```

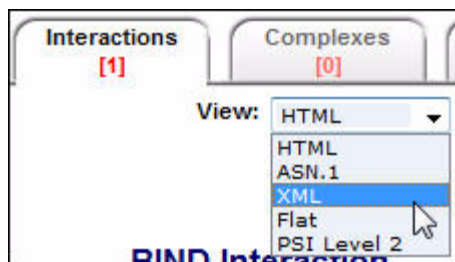
## XML

Displays the interaction in Extensible Markup Language (XML) format.

For more information on XML please visit:

<http://www.w3.org/XML/>.

To display the XML interaction view, select XML in the View list when viewing a record in an Interactions Single Record View.



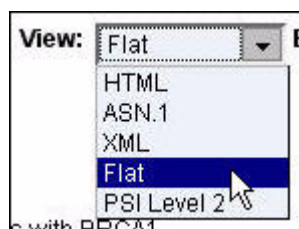
The XML view is displayed:

```
<BIND-Interaction>
  <BIND-Interaction_date>
    <Date>
      <Date_std>
        <Date-std>
          <Date-std_year>2005</Date-std_year>
          <Date-std_month>4</Date-std_month>
          <Date-std_day>22</Date-std_day>
        </Date-std>
      </Date_std>
    </Date>
  </BIND-Interaction_date>
  <BIND-Interaction_updates>
    <BIND-update-object>
      <BIND-update-object_date>
        <Date>
          <Date_std>
            <Date-std>
              <Date-std_year>2003</Date-std_year>
              <Date-std_month>10</Date-std_month>
              <Date-std_day>28</Date-std_day>
            </Date-std>
          </Date_std>
        </Date>
      </BIND-update-object_date>
      <BIND-update-object_descr>Public Release</BIND-update-object_descr>
    </BIND-update-object>
    <BIND-update-object>
      <BIND-update-object_date>
        <Date>
          <Date_std>
            <Date-std>
              <Date-std_year>2003</Date-std_year>
              <Date-std_month>10</Date-std_month>
              <Date-std_day>28</Date-std_day>
            </Date-std>
          </Date_std>
        </Date>
      </BIND-update-object_date>
      <BIND-update-object_descr>Public Release</BIND-update-object_descr>
    </BIND-update-object>
  </BIND-Interaction_updates>
</BIND-Interaction>
```

## Flat

Displays the interaction in Flat File format.

To display the Flat File interaction view, select Flat in the View list when viewing a record in an Interactions Single Record View.



The Flat File view is displayed:

BINDID	12725
DATE	2005/4/22
DESCRIPTION	The LIM domain protein LMO4 interacts with BRCA1
DIVISION	BIND Metazoa
REFERENCES	
TITLE	The LIM domain protein LMO4 interacts with the cofactor CtIP and the tumor suppressor BRCA1 activity.
PMID	11751867
OPINION	support
MOLECULEA	protein
AGI	6552303
ASHLABEL	BRCA1
AOTHNAMES	PSCP, RNF53
ADESCR	Breast Cancer 1, early onset, breast and ovarian cancer susceptibility protein 1; nuclear phosphoprotein in familial breast and ovarian cancers; tumour suppressor that participates in DNA repair and DNA damage-responsive pathways. Contains an NH2-terminal RING finger domain, nuclear localization signal, and the C-terminal acidic transcriptional activation domain with two C-terminal BRCA1 COC repeats; OMIM ID: 113705
ASOURCE	
ORGANISM	Homo sapiens
MOLECULEB	protein
BGI	5803072
BSHLABEL	LMO4
BDESCR	LIM domain only 4 is a cysteine-rich protein containing two LIM domains. Its mRNA is characterized by multiple ATTT motifs in the 3' region. A variant transcript missing a portion of 1 kb was identified but cannot be confirmed because of the GC-rich nature of the region. Member of the LIM domain-containing transcription factors. 100% identity between mouse and human. OMIM:603129
BSOURCE	
ORGANISM	Homo sapiens
FEATURES	

## PSI Level 2

Displays the interaction in PSI Level 2 format.

To display the PSI Level 2 interaction view, select PSI Level 2 in the View list when viewing a record in an Interactions Single Record View.





The PSI Level 2 interaction view is displayed:

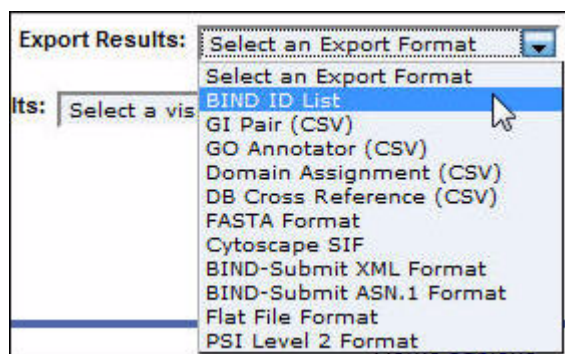
```
<entrySet level="2" version="0">
  <entry>
    <source>
      <bibref>
        <xref>
          <primaryRef id="11751867" db="PubMed" secondary="support"/>
        </xref>
      </bibref>
    </source>
    <availabilityList>
      <availability id="12725">BIND data is free for both commercial and academic use. If you use BIND please cite:
      Hogue CW. (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 31(1):248-50 PMID: 125
      distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERC
      FITNESS FOR A PARTICULAR PURPOSE. WARNING: The BIND specification does not have a one to one field mappir
      specification. As such there is considerable data loss when exporting BIND interaction and complex data in PSI format
      accurate BIND record(s), as entered by our curation team, please visit http://search.bind.ca</availability>
    </availabilityList>
    <interactionList>
      <interaction id="12725">
        <names>
          <shortLabel>The LIM domain protein LMO4 interacts with BRCA1</shortLabel>
        </names>
        <xref>
          <primaryRef id="12725" db="BIND" version="3.2"/>
        </xref>
        <availabilityRef ref="12725"/>
        <experimentList>
          <experimentDescription id="0">
            <names>
              <shortLabel>An interaction between BRCA1 and LMO4 was demonstrated by co-immunoprecipitation.
              immunoprecipitated from extracts of 293T cells transfected with BRCA1-myc and LMO4-FLAG using anti-myc and prec
              resolved by SDS-PAGE. LMO4-FLAG was detected by western blot using anti-FLAG. Fig. 4A.</shortLabel>
            </names>
            <bibref>
              <xref>
                <primaryRef id="11751867" db="PubMed" secondary="support"/>
              </xref>
            </bibref>
          </experimentDescription>
        </experimentList>
      </interaction>
    </interactionList>
  </entry>
</entrySet>
```

## Exporting Interactions Search Data

You can export your Interactions search results data from either an Interactions Summary or Single Record View.

## Exporting Data From an Interactions Summary or Single Record View

1. To export search results data displayed in an Interactions Summary or Single Record View, on the Interactions tab choose a format from the Export Results list.



**NOTE:** For information about the formats available, see “Interactions Export Formats” on page 78.

2. A confirmation box appears asking if you want to open or save the file.



**NOTE:** The confirmation box may appear different than that shown above, depending on the browser you are using (for example, Internet Explorer or Mozilla Firefox). Also, your browser might attempt to block the download. Check your browser’s settings to correct this (see “Appendix C Frequently Asked Questions” on page 141).

3. Choose the Save or Save As option.
4. Depending on the browser you are using, the file is downloaded into your browser’s default download location or a dialog box appears allowing you to choose the folder to which you want to download the file.

## Interactions Export Formats

The following Interactions export formats are available in BONDplus:



Format	Description
BIND ID List	All BIND IDs returned, one ID per line.
GI Pair (CSV)	Exported search results from a BINDplus search. Format is BIND ID, Molecule Name A, Molecule Type A, GI A, Entrez Gene A, Taxonomy Name A, Molecule Name B, Molecule Type B, GI B, Entrez Gene B, Taxonomy Name B, Supporting PMIDs. Comma Separated Values. One BIND ID per line.
GO Annotator (CSV)	Gene Ontology annotation for all Protein sequences. Format is BIND ID, GI, GO Identifier, GO Name. Comma Separated Values.
Domain Assignment (CSV)	BINDplus. Pre-computed Protein Domain Associations for all Proteins returned per query. BIND ID, GI, Domain ID, Domain Name. Comma Separated Values.
DB Cross Reference (CSV)	All known Database Cross References per GI. BIND ID, GI, Database Identifier, Database Name. Comma Separated Values.
FASTA Format	The sequence data contained by all component records returned in FASTA format. For more information, see <a href="http://www.ncbi.nlm.nih.gov/blast/fasta.shtml">http://www.ncbi.nlm.nih.gov/blast/fasta.shtml</a> .
Cytoscape SIF	Export interaction data as a Simple Interaction File (SIF) suitable for use with Cytoscape. For more information, see <a href="http://www.cytoscape.org">http://www.cytoscape.org</a> .
BIND-Submit XML Format	Export sets of BINDplus records in XML format.
BIND XML Format	Export individual BINDplus records in XML format.
BIND-Submit ASN.1 Format	Export sets of BINDplus records in ASN.1 format.
BIND ASN.1 Format	Export individual BINDplus records in ASN.1 format.
Flat File Format	Export GENESEQ or BINDplus records in an EMBL-like Flat File format.
PDF	The current record summarized in Adobe Portable Document Format (PDF).
PSI Level 2 Format	All interaction records returned from your query in HUPO Protein Standards Initiative - Molecular Interaction (PSI-MI) Level 2 format.

## Complexes Summary Record View

After performing a complexes query or a search (see “Searching for Data” on page 9), the Complexes Summary Record View presents you with a list of complex-related records under the Complexes tab.

Text search - BIND

**TEXT SEARCH**

Text Query:

Records per page:

**SEARCH**

**RESET**

**Summary** [49180]

**Sequences** [5151]

**Interactions** [164]

**Complexes** [15]

**Pathways** [0]

**GENESEQ** [43850]

☒ Interaction/Complex/Pathway Filters      **View:**       **Export Results:**

Identifier	Description	Number of Subunits	Taxi
<b>Complex 45573</b>	Nmi homodimerizes to form a complex	2 subunits	<a href="#">Homo</a>
<b>Complex 12728</b>	LMO4, BRCA1, and CtIP form a complex.	4 subunits	<a href="#">Homo</a>
<b>Complex 45575</b>	BRCA1, Nmi and Myc form a complex	4 subunits	<a href="#">Homo</a> <a href="#">Spec</a>
<b>Complex 24870</b>	The binding domain of tumor suppressor protein p53 and a zinc ion form a complex.	2 subunits	<a href="#">Homo</a>
<b>Complex 164736</b>	This molecular complex record represents a BRCA1-IRIS-interacting complex including BRCA1-IRIS, ORC1, CDC6 and Gemini...	4 subunits	<a href="#">Homo</a>
<b>Complex 164739</b>	This molecular complex record represents a complex between the rDNA core DNA replication initiation site and a BRCA1-...	2 subunits	<a href="#">Homo</a>
<b>Complex 164740</b>	This molecular complex record represents a complex between the Beta-globin core DNA replication initiation site and a...	2 subunits	<a href="#">Homo</a>
<b>Complex 258589</b>	The ETS2 (Ets-2)-Smarca4 (Brg-1)-Smarca1 (Ini1) complex was isolated by co-immunoprecipitation. The topology of the c...	3 subunits	<a href="#">Homo</a>
<b>Complex 164737</b>	This molecular complex record represents a complex between ORC5 and ORC1, as demonstrated by co-immunoprecipitation. ...	2 subunits	<a href="#">Homo</a>
<b>Complex 164741</b>	This molecular complex record represents a complex between the rDNA core DNA replication initiation site and MCM2, as...	2 subunits	<a href="#">Homo</a>
<b>Complex 164742</b>	This molecular complex record represents a complex between the Beta-globin core DNA replication initiation site and M...	2 subunits	<a href="#">Homo</a>
<b>Complex 258590</b>	The ETS2 (Ets-2)-Smarca4 (Brg-1)-Smarca1 (baf57)-Smarca1 (Ini1) complex was isolated by co-immunoprecipitation. The t...	4 subunits	<a href="#">Mus</a>
<b>Complex 164738</b>	This molecular complex record represents a complex between MCM2, MCM6 and MCM7, as determined by co-immunoprecipitati...	3 subunits	<a href="#">Homo</a>

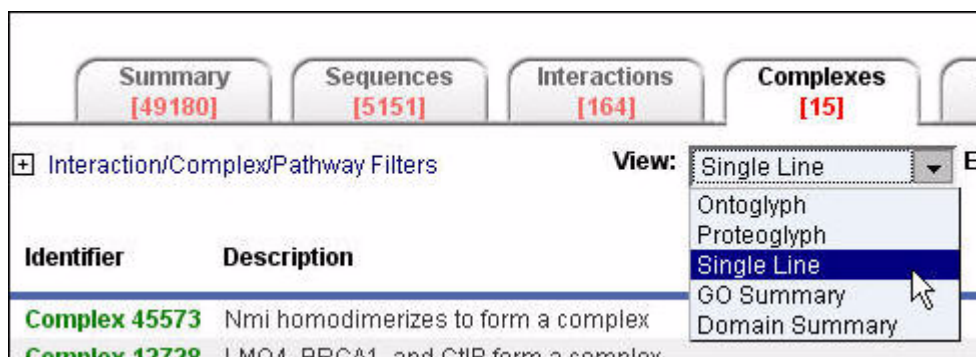
The information displayed for each record in this view is not the full annotation available for the record. For every type of Complexes Summary Record View, a link (typically, the identifier displayed in bold green text) is provided to access various levels of annotation detail for that record in the Complexes Single Record View format (see “Complexes Single Record View” on page 86).

## Complexes Summary Record View Types

There are different types of Complexes Summary Record Views available within the Complexes Tab: Single Line, Ontoglyph, Proteoglyph, GO Summary, and Domain Summary.

### Single Line

Provides basic details of a record. This is the default view when the Complexes tab is displayed. To display the Single Line view, select **Single Line** in the **View** list on the **Complexes** tab.



The Single Line view is displayed.

Identifier	Description	Number of Subunits	T
<b>Complex 45573</b>	Nmi homodimerizes to form a complex	2 subunits	H
<b>Complex 12728</b>	LMO4, BRCA1, and CtIP form a complex.	4 subunits	H
<b>Complex 45575</b>	BRCA1, Nmi and Myc form a complex	4 subunits	H S
<b>Complex 24870</b>	The binding domain of tumor suppressor protein p53 and a zinc ion form a complex.	2 subunits	H
<b>Complex 164736</b>	This molecular complex record represents a BRCA1-IRIS-interacting complex including BRCA1-IRIS, ORC1, CDC6 and Gemini...	4 subunits	H

Column	Description
Identifier	BINDplus integer identifier unique to each BINDplus interaction or complex record.
Description	Provides a brief description of the complex, indicating the function and subunits of the complex, the experimental system used to demonstrate the complex and whether or not the topology of the complex is known.
Number of Subunits	Details the number of subunits present in the complex. Subunits can be any BIND object; protein, DNA, RNA, complex, small molecule or an unspecified molecule.
Taxonomy	Describes the taxon to which the interacting molecules belong.





Summary [49180] Sequences [5151] Interactions [164] Complexes [15]

+ Interaction/Complex/Pathway Filters View: Single Line

- Ontoglyph
- Proteoglyph**
- Single Line
- GO Summary
- Domain Summary

Identifier	Description
Complex 45573	Nmi homodimerizes to form a complex
Complex 12728	LMQ4, BRCA1, and CtIP form a complex

The Proteoglyph view is displayed:

Identifier	Number of Subunits	Proteoglyphs
Complex 164737	2 subunits	
Complex 164736	4 subunits	

Column	Description
Identifier	BINDplus integer identifier unique to each BINDplus interaction or complex record.
Number of Subunits	Details the number of subunits present in the complex. Subunits can be any BIND object; protein, DNA, RNA, complex, small molecule or an unspecified molecule.
Proteoglyphs	The proteoglyphs associated with the record. Proteoglyphs provide a visually condensed pictorial representation of domain annotation associated with that record. For more information see “Appendix B About Proteoglyphs” on page 137
Taxonomy	Describes the taxon to which the interacting molecules belong.

## GO Summary

Provides basic details of a record. To display the GO Summary view, select GO Summary in the View list on the Complexes tab.

Summary [49180] Sequences [5151] Interactions [164] Complexes [15]

+ Interaction/Complex/Pathway Filters View: Single Line

- Ontoglyph
- Proteoglyph
- Single Line
- GO Summary
- Domain Summary

Identifier	Description
Complex 45573	Nmi homodimerizes to form a complex

The GO Summary view is displayed:

Complex 45573		Homo sapiens	Full BIND Record	Launch Viewer:	Select Below
Subunits	Description	Molecular Function	Cellular Component	Biological Process	
Nmi	N-myc and STAT interactor; N-Myc Interactor (Nmi); also interacts with C-myc. OMIM: 603525.	transcription cofactor activity	cytoplasm	<ul style="list-style-type: none"> <li>transcription from RNA polymerase II promoter</li> <li>inflammatory response</li> <li>JAK-STAT cascade</li> </ul>	
Nmi	N-myc and STAT interactor; N-Myc Interactor (Nmi); also interacts with C-myc. OMIM: 603525.	transcription cofactor activity	cytoplasm	<ul style="list-style-type: none"> <li>transcription from RNA polymerase II promoter</li> <li>inflammatory response</li> <li>JAK-STAT cascade</li> </ul>	

Column	Description
Subunits	Lists the subunits contained within the complex. Annotation is consistent with all BIND object annotation; short label and aliases, identifiers where appropriate, a molecule description and expanded names, species information and other database identifiers.
Description	Provides a list of the expanded (long-form) names of each of the interacting molecules. Expanded names are derived from the publication, Entrez and species-specific databases (eg. OMIM, MGI, RGD, SGD, Flybase). A short description of the function and characteristics of the molecule may also be provided.
Molecular Function	Lists the GO Molecular Function terms associated with each of the interacting molecules. GO's Molecular Function Ontology describes a series of activities, such as DNA binding or transcription factor activity, occurring at the molecular level.

Column	Description
Cellular Component	Lists the GO Cellular Component terms associated with each of the interacting molecules. GO's Cellular Component Ontology describes compartments and physical locations and structures within a cell or organism.
Biological Process	Lists the GO Biological Process terms associated with each of the interacting molecules. GO's Biological Process Ontology describes a series of events accomplished by one or more ordered assemblies of molecular functions. Examples include cell cycle checkpoint and DNA repair.
Links	Selection of link displays all records and data associated with the GenBank Identifier (GI) in BINDplus, SeqHound or NCBI. <ul style="list-style-type: none"> <li>• <b>Taxon- BIND:</b> this link displays all records in BINDplus that contain that taxon group.</li> <li>• <b>Pubmed:</b> this link displays the publication abstract at Pubmed.</li> <li>• <b>BIND Publication Search:</b> returns all BINDplus records containing that publication.</li> </ul>

### Domain Summary

Provides basic details of a record. To display the Domain Summary view, select Domain Summary in the View list on the Complexes tab.

The screenshot shows the 'Complexes' tab with a count of [15]. Below the tabs is a 'View:' dropdown menu. The menu is open, showing options: Single Line, Ontoglyph, Proteoglyph, Single Line, GO Summary, and Domain Summary. 'Domain Summary' is highlighted with a mouse cursor. Below the menu, a table with columns 'Identifier' and 'Description' is partially visible, showing 'Complex 45573' and 'Complex 12728'.

The Domain Summary view is displayed:

Complex 45573 Homo sapiens Full BIND Record Launch Viewer: Select Below				
Subunits	Description	Pfam	SMART	COG
Nmi	N-myc and STAT interactor; N-Myc Interactor (Nmi); also interacts with C-myc. OMIM: 603525.	• NID	No Domain Annotation	No Domain Annotation
Nmi	N-myc and STAT interactor; N-Myc Interactor (Nmi); also interacts with C-myc. OMIM: 603525.	• NID	No Domain Annotation	No Domain Annotation

Column	Description
Subunits	Lists the subunits contained within the complex. Annotation is consistent with all BIND object annotation; short label and aliases, identifiers where appropriate, a molecule description and expanded names, species information and other database identifiers.
Description	Provides a list of the expanded (long-form) names of each of the interacting molecules. Expanded names are derived from the publication, Entrez and species-specific databases (for example, OMIM, MGI, RGD, SGD, Flybase). A short description of the function and characteristics of the molecule may also be provided.
Pfam	This section provides information on any Pfam domains within the molecule. The Protein Family Database (Pfam) is a collection of multiple sequence alignments and profile hidden Markov models (HMMs). Each Pfam HMM represents a protein family or domain. Information is derived from NCBI's CDD database.
SMART	This section provides information on any SMART domains within the molecule. Simple Modular Architecture Research Tool (SMART) allows the identification and annotation of domains and the analysis of domain architectures. Information is derived from NCBI's CDD database.
COG	This section provides information on any COG domain within the molecule. Clusters of Orthologous Groups of proteins (COGs) Information is a phylogenetic classification of the proteins encoded complete genomes of bacteria, archaea and eukaryotes. COG is derived from NCBI's CDD database.
Links	<p>Selection of link displays all records and data associated with the GenBank Identifier (GI) in BINDplus, SeqHound or NCBI.</p> <ul style="list-style-type: none"> <li>• <b>Taxon- BIND:</b> this link displays all records in BINDplus that contain that taxon group.</li> <li>• <b>Pubmed:</b> this link displays the publication abstract at Pubmed.</li> <li>• <b>BIND Publication Search:</b> returns all BINDplus records containing that publication.</li> </ul>

## Visualize Results

Cytoscape is an open source bioinformatics software platform for visualizing molecular interaction networks and integrating these interactions with gene expression profiles and other state data. For more information on how to use this feature, go to <http://www.cytoscape.org>.

## Complexes Single Record View

The Complexes Single Record View provides annotation details for a single record. Levels of annotation detail depend on the type of Complexes Single Record View selected.



To display a Single Record View for a Complex, click a record in the Identifier column of an Complexes Summary Record View (see “Complexes Summary Record View” on page 79) on the Interactions tab.

Identifier	Description
<a href="#">Complex 45573</a>	Nmi homodimerizes to form a complex
<a href="#">Complex 12728</a>	LMO4, BRCA1, and CtIP form a complex
<a href="#">Complex 45575</a>	BRCA1, Nmi and Myc form a complex

The Complexes Single Record View is displayed for the item. The Identifier Search section displays the identifier values for the record you have opened. The default view type is HTML. For information on the different view types for a single record, see “Complexes Single Record View Types” on page 87.

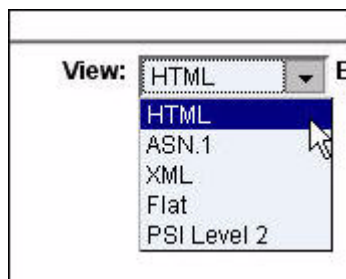
## Complexes Single Record View Types

There are different types of Complexes Single Record Views available:

- [HTML](#)
- [ASN.1](#)
- [XML](#)
- [Flat](#)
- [PSI Level 2](#)

### HTML

Provides an integrated view of all annotation associated with a record in BONDplus. To display the HTML view, select HTML in the View list when viewing a record in a Complexes Single Record View.



The HTML view is displayed:

[Expand all] [Collapse all]

## BIND Molecular Complex

Rev

Visualize using

**BIND Id:** 45573**Complex Description:** Nmi homodimerizes to form a complex**Division:** BIND Metazoa+ **Publications:** 1 [View all publications \[NCBI\]](#)**Record References:** 1 [Interaction\(s\)](#)**Date Last Released:** September 8, 2004

### + Sub units

2 Subunits.

### Interaction List

By default, the HTML Complexes Single Record View shows a collapsed form of the information, with summary descriptions of each collapsed section, and an expander icon  to show the annotation of that particular section.

By clicking Expand All or Collapse All, you can expand and collapse all annotation for the given record.

[Expand all] [Collapse all]

## BIND Molecular Complex

**BIND Id:** 45573**Complex Description:** Nmi homodimerizes to form a complex**Division:** BIND Metazoa- **Publications:** 1. -**PMID:** 11916966**Opinion:** Support**Journal:** [J Biol Chem](#) 2002/6/7;277(23):20965-20973.

**Article Title:** A novel tricomplex of BRCA1, Nmi, and c-Myc inhibits c-Myc-induced human telomerase reverse transcriptase gene (hTERT) promoter activity in breast cancer.

- **Authors:**

- Li H
- Lee TH
- Avraham H

**Affiliation:** Division of Experimental Medicine, Beth Israel Deaconess Medical Center, Harvard Institutes of Medicine, Boston, Massachusetts 02115, USA.

- **Abstract:** Germ-line mutations in BRCA1 predispose individuals to breast and ovarian cancers. We observed a novel endogenous association of BRCA1 with Nmi (N-Myc-interacting protein) in breast cancer cells. Nmi was found to interact specifically with



Field	Description
Subunit #1 and #2	<p>Description: Details a subunit present in the complex. Subunits can be any BIND object; protein, DNA, RNA, complex, small molecule or an unspecified molecule.</p> <p>NCBI GenInfo ID: NCBI's integer identifier assigned to each unique sequence listed within the NCBI sequence database.</p> <p>Origin: Indicates the organism or species from which the complex is derived.</p>
Automatically Retrieved Annotation	<p><b>Cross References:</b> Contains and links to accession numbers of sequences redundant to the listed GI contained within sequence databases such as NCBI's RefSeq, SwissProt, EMBL and GenBank.</p> <p><b>GO Terms:</b> Contain information and links on any GO Terms associated with the molecule. Developed by the Gene Ontology (GO) Consortium, GO terms are generated as a part of a controlled vocabulary, or ontology, to describe gene or gene products of an organism. The ontology is listed under the areas of Cellular Component, Biological Processes and Molecular Function.</p> <p><b>Domains:</b> Contains information and links to any conserved domains within the molecule.</p>
Interaction List	<p><b>Unordered:</b> Indicates that subunit assembly does not occur in the order in which they are shown in the complex record</p>

## ASN.1

Displays the complex in Abstract Syntax Notation One (ASN.1) ASCII format.

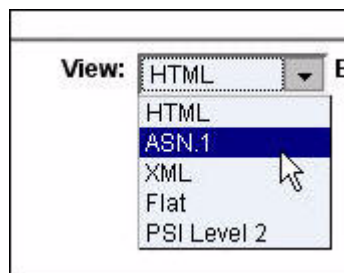
For more information on Abstract Syntax Notation One, please visit:

<http://asn1.elibel.tm.fr/en/>.

For more information on ASN.1 and its usage at NCBI, please visit:

<http://www.ncbi.nlm.nih.gov/Sitemap/Summary/asn1.html>.

To display the ASN.1 view, select ASN.1 in the View list when viewing a record in a Complexes Single Record View.



The ASN.1 view is displayed:

```

BIND-Molecular-Complex ::= {
  date std {
    year 2003,
    month 11,
    day 1
  },
  updates {
    {
      date std {
        year 2003,
        month 11,
        day 28
      },
      descr "Public Release"
    },
    {
      date std {
        year 2004,
        month 9,
        day 8
      },
      descr "Public Release"
    }
  },
  mcid 45573,
  descr "Nmi homodimerizes to form a complex",
  sub-num num 2,
  sub-units {
    {
      hmoicid 0,
      sub-unit 1
    }
  }
}

```

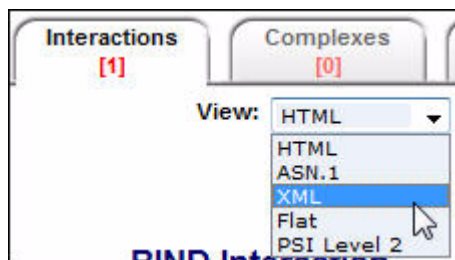
## XML

Displays the complex in Extensible Markup Language (XML) format.

For more information on XML please visit:

<http://www.w3.org/XML/>.

To display the XML complex view, select XML in the View list when viewing a record in a Complexes Single Record View.



The XML view is displayed:

```
<BIND-Molecular-Complex>
  <BIND-Molecular-Complex_date>
    <Date>
      <Date_std>
        <Date-std>
          <Date-std_year>2003</Date-std_year>
          <Date-std_month>11</Date-std_month>
          <Date-std_day>1</Date-std_day>
        </Date-std>
      </Date_std>
    </Date>
  </BIND-Molecular-Complex_date>
  <BIND-Molecular-Complex_updates>
    <BIND-update-object>
      <BIND-update-object_date>
        <Date>
```

## Flat

Displays the complex in Flat File format.

To display the Flat File complex view, select Flat in the View list when viewing a record in a Complexes Single Record View.



The Flat File view is displayed:

```
BINDID      45573
DATE        2003/11/1
DESCRIPTION  Nmi homodimerizes to form a complex
DIVISION    BIND Metazoa
REFERENCES
  TITLE     A novel tricomplex of BRCA1, Nmi, and c-Myc inhibits c-Myc-induced human telomerase reverse trans
  PMID      11916966
  OPINION    support
```

## PSI Level 2

Displays the complex in PSI Level 2 format.

To display the PSI Level 2 complexes view, select PSI Level 2 in the View list when viewing a record in a Complexes Single Record View.





The PSI Level 2 interaction view is displayed:

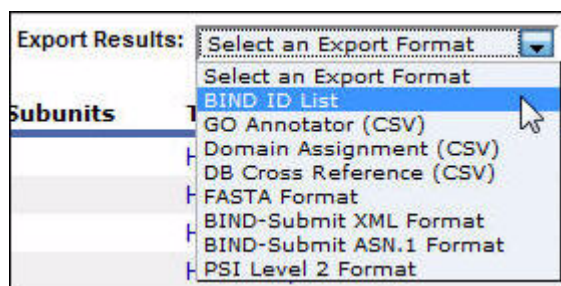
```
<entrySet level="2" version="0">
  <entry>
    <source>
      <bibref>
        <xref>
          <primaryRef id="11916966" db="PubMed" secondary="support"/>
        </xref>
      </bibref>
    </source>
    <availabilityList>
      <availability id="45573">BIND data is free for both commercial and academic use. If you use BIND please cite: E
      Hogue CW. (2003) BIND: the Biomolecular Interaction Network Database. Nucleic Acids Res. 31(1):248-50 PMID: 1251
      distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCH
      FITNESS FOR A PARTICULAR PURPOSE. WARNING: The BIND specification does not have a one to one field mapping
      accurate BIND record(s), as entered by our curation team, please visit http://search.bind.ca</availability>
    </availabilityList>
    <interactionList>
      <interaction id="45573">
        <xref>
          <primaryRef id="45573" db="BIND" version="3.2"/>
        </xref>
        <availabilityRef ref="45573"/>
        <experimentList>
          <experimentDescription id="0">
            <names>
              <shortLabel>Homodimerization of Nmi was demonstrated by co-immunoprecipitation. FLAG-Nmi was i
              from extracts of 293T cells transfected with FLAG-Nmi and HA-Nmi using anti-FLAG and precipitates were resolved by E
              was detected by western blot using anti-HA. Fig. 3A. Anti-HA co-immunoprecipitated FLAG-Nmi and HA-Nmi as detected
              3A.</shortLabel>
            </names>
            <bibref>
              <xref>
                <primaryRef id="11916966" db="PubMed" secondary="support"/>
              </xref>
            </bibref>
          </experimentDescription>
        </experimentList>
      </interaction>
    </interactionList>
  </entry>
</entrySet>
```

## Exporting Complexes Search Data

You can export your Complexes search results data from either a Complexes Summary or Single Record View.

### Exporting Data From a Complexes Summary or Single Record View

1. To export search results data displayed in a Complexes Summary or Single Record View, on the Complexes tab choose a format from the Export Results list.



**NOTE:** For information about the formats available, see “Complexes Export Formats” on page 94.

2. A confirmation box appears asking if you want to open or save the file.



**NOTE:** The confirmation box may appear different than that shown above, depending on the browser you are using (for example, Internet Explorer or Mozilla Firefox). Also, your browser might attempt to block the download. Check your browser’s settings to correct this (see “Appendix C Frequently Asked Questions” on page 141).

3. Choose the Save or Save As option.
4. Depending on the browser you are using, the file is downloaded into your browser’s default download location or a dialog box appears allowing you to choose the folder to which you want to download the file.

## Complexes Export Formats

The following Complexes export formats are available in BONDplus:

Format	Description
BIND ID List	All BIND IDs returned, one ID per line.
GO Annotator (CSV)	Gene Ontology annotation for all Protein sequences. Format is BIND ID, GI, GO Identifier, GO Name. Comma Separated Values.



Format	Description
Domain Assignment (CSV)	BINDplus. Pre-computed Protein Domain Associations for all Proteins returned per query. BIND ID, GI, Domain ID, Domain Name. Comma Separated Values.
DB Cross Reference (CSV)	All known Database Cross References per GI. BIND ID, GI, Database Identifier, Database Name. Comma Separated Values.
FASTA Format	The sequence data contained by all component records returned in FASTA format. For more information, see <a href="http://www.ncbi.nlm.nih.gov/blast/fasta.shtml">http://www.ncbi.nlm.nih.gov/blast/fasta.shtml</a> .
BIND-Submit XML Format	Export sets of BINDplus records in XML format.
BIND XML Format	Export individual BINDplus records in XML format.
BIND-Submit ASN.1 Format	Export sets of BINDplus records in ASN.1 format.
BIND ASN.1 Format	Export individual BINDplus records in ASN.1 format.
PSI Level 2 Format	All interaction records returned from your query in HUPO Protein Standards Initiative - Molecular Interaction (PSI-MI) Level 2 format.
PDF	The current record summarized in Adobe Portable Document Format (PDF).



## Chapter 7

### GENESEQ

---

GENESEQ™ contains high quality, easy-to-understand annotated biological sequences from worldwide patents – an essential tool for any organization that needs to understand how their biological research can be patented, monitor competitor activities across the globe and retrieve significant information about a sequence of interest which is difficult to retrieve from patent source.

GENESEQ is the largest manually indexed patent sequence database. It is a comprehensive and continually growing database of patented biological sequences covering genetic discovery back to 1981. As many as 40,000 new records are added every two weeks, with each record annotated by highly-trained editorial staff.

GENESEQ enables you to search sequence information from around the world in a single database. 41 patent authorities are covered, including WO, US, EP, JP, DE, IN and CN. Non-English language patents compose over 20 percent of the database – and English language titles are provided for each non-English patent.

For more information, see “GENESEQ FAQ” on page 146 and [An Introduction to GENESEQ](#).

### Performing a Query for GENESEQ

The text search engine supporting BONDplus allows for both broad and narrow queries.

To retrieve all interaction records relating to a specific database accession or identifier – GENESEQ Accession or Patent Number – use the Identifier Search interface with the identifier of interest (see “Identifier Search” on page 15).

To perform a broad search for all BONDplus records referencing ‘brca1’:

+brca1

To perform a search for all patented sequence records with ‘brca1’ as a GENESEQ record keyword:

+geneseq.keyword:brca1

To further restrict this search to records released in 1996:

+geneseq.keyword:brca1 +patent.year:1996

Sometimes, you may not know a specific identifier associated with an internal ontology or taxonomy. For example, to restrict a query by taxonomy, you can use either the 'geneseq.organism' field to search by name or 'taxid' to search by the broader taxonomy identifier. Therefore, you can perform the search for human protein records which have a match to a record in our public sequence databases:

+geneseq.organism:"Homo sapiens" +geneseq.moldesc:AA +geneseq.pc:true

OR

+taxid:9606 +geneseq.moldesc:AA +geneseq.pc:true

Use the GENESEQ Filters to complement complex searches (see "Using the GENESEQ Filters" on page 98). To restrict this query to all human protein sequences deriving from a patent claim by Pfizer, Ctrl-click to select "AA – peptide" and "AA-protein" from "Molecule Type" and select "Claim" from the "Patent Sequence Claim Type" in the GENESEQ filters, and use the following text query:

+patent.patentee:Pfizer\* +taxid:9606

To change your view from the list of GENESEQ Accessions to a list of the associated patents, select "Patent" from the "View" dropdown to enter Multi-patent view.

All filters can be expressed use the text query language as well. This query can be performed without the use of GENESEQ Filters as follows:

+patent.patentee:Pfizer\* +taxid:9606 +geneseq.moldesc:AA +geneseq.claimtype:Claim


To assist you in constructing advanced queries of your own design, remember to use the Advanced Field Specific Search query builder tool (see "Advanced Field Specific Search" on page 19).

## Using the GENESEQ Filters

Data-specific filters are available under the GENESEQ tab. They allow you to further narrow your search.

GENESEQ-specific filters within the GENESEQ tab allow you to restrict the contents being browsed by the type of molecule, by the Year the GENESEQ Record was issued, and by patent office and patent sequence claim type.

### How to use the GENESEQ Filters

1. Click the  icon. The filter options are displayed.

**Restrict records returned by:**

*CTRL-Click to multiple select or de-select items*

**Molecule Type:**

- AA - peptide
- AA - protein
- NA - cDNA
- NA - cDNA to rRNA
- NA - cDNA to mRNA
- NA - cDNA to scRNA
- NA - cDNA to tRNA
- NA - DNA
- NA - mRNA
- NA - preRNA

**Records with 100% identity to a public sequence:**

Has Match  
No Match

**Year GENESEQ Record Issued:**

1975  
1976  
1977

**Patent Office:**

AT  
AU  
BE

**Patent Sequence Claim Type :**

Claim  
Example  
Disclosure

**Released in GENESEQ Flat File:**

Released  
Not Released

**SHOW RESULTS**

- Click one or more items in the lists to include those items in your search. If no items in a list are selected then all items in that list are included in your search.
- Click SHOW RESULTS. Your search results are updated.

### GENESEQ Filter Field Descriptions

Field Name	Description
Molecule Type	The type of amino acid (AA) or nucleic acid (NA) molecule.
Records with 100% identity to a public sequence	Sequences which match exactly in length and residue identity.
Year GENESEQ Record issued	The date the record was issued in GENESEQ.
Patent Office	The two-letter WIPO code for the patent-issuing authority. For example, US for United States, EP for Europe, JP for Japan and WO for WIPO.
Patent Sequence Claim Type	Sequence is mentioned in the 'Claim' section of the patent; if not, then in the 'Example' section; if not, then in the remainder of the patent 'Disclosure'.
Released in GENESEQ Flat File	The record may appear in BONDplus while it has not yet been published in GENESEQ.

## GENESEQ Summary Record View

After performing a GENESEQ query (see “Performing a Query for GENESEQ” on page 97) or a search (see “Searching for Data” on page 9), the GENESEQ Summary Record View presents you with a list of sequence-related records under the Sequence tab.

Summary

Sequences

Interactions

Complexes

Pathways

GENESEQ

[49187]

[5158]

[164]

[15]

[0]

[43850]

+

GENESEQ Filters

View:

Single Line

Export Res

1

2

3

4

--

2193

▶

Skip To Page:

1

GO

Accession	Type	Claim Type	Description	Organism
AAB68015	protein	Example 1	Modified cytosine-specific DNA methyltransferase MspI	Synthetic
AAB68895	protein	Disclosure	Human ATM protein	Homo sapiens
AAB68896	protein	Disclosure	Human ATM protein	Homo sapiens
AAB82972	protein	Claim 1	Human G protein associated molecule GPAM-1 protein 4620239CD1	Homo sapiens
AAC60793	cDNA	Example 2	Human BRCA1 (omi1) nucleotide sequence SEQ ID NO:1	Homo sapiens
AAC60794	cDNA	Claim 1	Human BRCA1 (omi2) nucleotide sequence SEQ ID NO:3	Homo sapiens
AAC60795	cDNA	Example 4	Human BRCA1 (omi3) nucleotide sequence SEQ ID NO:5	Homo sapiens
AAC60796	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:7	Homo sapiens
AAC60797	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:8	Homo sapiens
AAC60798	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:9	Homo sapiens
AAC60799	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:10	Homo sapiens
AAC60800	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:11	Homo sapiens
AAC60801	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:12	Homo sapiens
AAC60802	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:13	Homo sapiens
AAC60803	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:14	Homo sapiens
AAC60804	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:15	Homo sapiens
AAC60805	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:16	Homo sapiens
AAC60806	DNA	Example 1	Human BRCA1 PCR primer SEQ ID NO:17	Homo sapiens

The information displayed for each record in this view is not the full annotation available for the record. For every type of GENESEQ Summary Record View, a link (typically, the Accession or Patent Numbers displayed in bold blue text) is provided to access various levels of annotation detail for that record in the GENESEQ Single Record View format (see “GENESEQ Single Record View” on page 102).

## View Types

There are different types of Sequences Summary Record Views available within the Sequence Tab: Single Line and Patent.

### Single Line

Provides basic details of a record (Accession, Description, Organism, Patent Number and so on). This is the default view when the GENESEQ tab is displayed. To display the Single Line view, select Single Line in the View list on the GENESEQ tab.



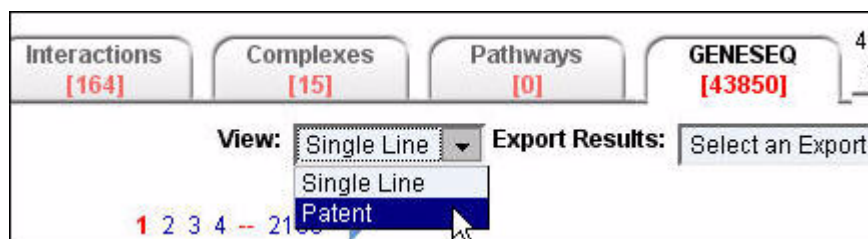
The Single Line view is displayed:

Accession	Type	Claim Type	Description	Organism	Date Issued	Patent
<a href="#">AAB68015</a>	protein	Example 1	Modified cytosine-specific DNA methyltransferase MspI	Synthetic	29 Jun 2001	<a href="#">WO2</a>
<a href="#">AAB68895</a>	protein	Disclosure	Human ATM protein	Homo sapiens	27 Apr 2001	<a href="#">WO2</a>
<a href="#">AAB68896</a>	protein	Disclosure	Human ATM protein	Homo sapiens	27 Apr 2001	<a href="#">WO2</a>

Column	Description
Accession	A unique identifier assigned to a biological polymer sequence submitted to a sequence database.
Type	The type of amino acid or nucleic acid molecule.
Claim Type	The location of the sequence in the 'Claim', 'Example' or the remainder of the patent 'Disclosure'.
Description	A single line description of the sequence.
Organism	The Latin name of the source species for this sequence.
Date Issued	The date the record was issued in GENESEQ.
Patent Number	The publication number of the patent document, composed of the two-letter country code of the issuing patenting authority, followed by a variable length serial number, followed by a 'Kind' code (for example: A1, A2, B1) indicating the status of the document.

## Patent

Provides basic details of a record (Accession, Description, Organism, Patent Number and so on). To display the Patent view, select Patent in the View list on the GENESEQ tab.





The Patent view is displayed:

Patent	Title	Patentee	Publication Date	# of sequences
<b>WO200125451-A2</b>	One step cloning vector for quicker cloning of nucleic acid fragments comprises both a positive selection gene and a gene which protects against a selective agent	TRANSGENOMIC INC	12 April 2001	11
<b>WO200107660-A1</b>	Detecting a mutation or polymorphism in human ataxia telangiectasia gene or polyexonic eukaryotic gene, involves using mega-single stranded conformation polymorphism analysis	UNIV CALIFORNIA	01 Feb 2001	14

Column	Description
Patent	The publication number of the patent document, composed of the two-letter country code of the issuing patenting authority, followed by a variable length serial number, followed by a 'Kind' code (for example: A1, A2, B1) indicating the status of the document.
Title	The WPI-enhanced title of the patent document.
Patentee	The company, institution, or person to whom the patent has been assigned.
Publication Date	The publication date of the patent.
# of sequence	The number of sequences from this patent which are present in GENESEQ.
# of claimed sequence	The number of sequences present in GENESEQ which are mentioned in the 'Claim' section of the patent.

## GENESEQ Single Record View

The GENESEQ Single Record View provides annotation details for a single record. Levels of annotation detail depend on the type of GENESEQ Single Record View selected.

To display a Single Record View for GENESEQ, click a record in the Accession column of a GENESEQ Summary Record View (see “GENESEQ Summary Record View” on page 99) on the GENESEQ tab.

Accession	Type	Claim Type	Description	Orga
<b>AAB68015</b>	protein	Example 1	Modified cytosine-specific DNA methyltransferase MspI	Synth
<b>AAB68895</b>	protein	Disclosure	Human ATM protein	Homo sapien

The GENESEQ Single Record View is displayed for the item. The Identifier Search section displays the identifier values for the record you have opened. The default view type is Accession. For information on the different view types for a single record, see “GENESEQ Single Record View Types” on page 103.

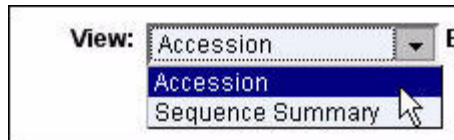
## GENESEQ Single Record View Types

There are different types of GENESEQ Single Record Views available:

- [Accession](#)
- [Sequence Summary](#)

### Accession

Provides an integrated view of all annotation associated with a record in BONDplus. To display the Accession view, select Accession in the View list when viewing a record in a GENESEQ Single Record View.



The Accession view is displayed:

Identifier search - GENESEQ


### IDENTIFIER SEARCH

GENESEQ Accession

Records per page:

**SEARCH**

**RESET**

 [+\(brca1\)](#)

<b>Summary</b> [1]	<b>Sequences</b> [0]	<b>Interactions</b> [0]	<b>Complexes</b> [0]	<b>Pathways</b> [0]	<b>GENESEQ</b> [1]
-----------------------	-------------------------	----------------------------	-------------------------	------------------------	-----------------------

☐ GENESEQ Filters **View:**  **Export Re**

[\[Expand all\]](#) [\[Collapse all\]](#)

#### Record Details

**DNA: Nucleotide sequence of plasmid pMTet1**

**Accession:** AAF84674

**Record Issued:** 29 June 2001

**Record Last Updated:** 29 June 2001


☐ **Patent Information:** [WO200125451-A2](#): One step cloning vector for quicker cloning of nucleic acid fragments comp gene and a gene which protects against a selective agent. (29 June 2001)

☐ **Comments:** The present sequence represents plasmid pMTet1, which is used in the method of the inventic one-step, non-plating method.....

**Sequence Length:** 5086 NA

**Sequence Location:** Claim; SEQ ID NO 2

☐ **Sequence:** FASTA format [BLAST this sequence...](#)

By default, the Accession GENESEQ Single Record View shows a collapsed form of the information, with summary descriptions of each collapsed section, and an expander icon  to show the annotation of that particular section.

By clicking Expand All or Collapse All, you can expand and collapse all annotation for the given record.

[\[Expand all\]](#) [\[Collapse all\]](#)

## Record Details

## DNA: Nucleotide sequence of plasmid pMTet1

**Accession:** AAF84674**Record Issued:** 29 June 2001**Record Last Updated:** 29 June 2001
☐ **Patent Information:**
**Source Patent:** [WO200125451-A2](#) **WPI:** 2001-273580/28**Patent Title:** One step cloning vector for quicker cloning of nucleic acid fragments c  
selection gene and a gene which protects against a selective agent**GENESEQ records:** 11 associated sequences.**Patent Assignee:** TRANSGENOMIC INC**Patent Inventors:** HORNBY D P J , MATIN M M**Publication Date:** 12 April 2001**No. of pages :** 49pp**Language:** English**Patent Filing:**

Patent Number	Date
	22
2000WO-GB003655	Sep
	2000

**Priority:**

Patent Number	Date
	01
99US-0157072P	Oct
	1999
	25
2000GB-00001716	Jan
	2000
	07
2000US-00656503	Sep
	2000

**Patent Family:**

Patent Number
<a href="#">AU-200075343-A</a>
<a href="#">EP-1218522-A2</a>

Hyperlinks are provided for quick access to the original source of annotation.

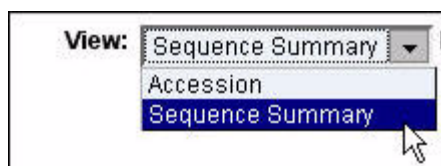
Record Details	
<b>DNA: Nucleotide sequence of plasmid pMTet1</b>	
<b>Accession:</b>	AAF84674
<b>Record Issued:</b>	29 June 2001
<b>Record Last Updated:</b>	29 June 2001
<input type="checkbox"/> <b>Patent Information:</b>	<p><b>Source Patent:</b> <a href="#">WO200125451-A2</a> <input type="button" value="Find this Patent in..."/></p> <p><b>WPI:</b> 2001-2580/28</p> <p><b>Patent Title:</b> One step cloning vector for quicker cloning of nucleic acid fragments selection gene and a gene which protects against a selective agent</p> <p><b>GENESEQ records:</b> 11 associated sequences.</p> <p><b>Patent Assignee:</b> TRANSGENOMIC INC</p>

Field	Description
Accession	A unique identifier assigned to a biological polymer sequence submitted to a sequence database.
Record Issued	The date the record was issued in GENESEQ.
Record Last Updated	If revised, the date the revised record was issued in GENESEQ.
Patent Information	<p><b>Source Patent:</b> The publication number of the patent document from which the sequence is taken, composed of the two-letter country code of the issuing patenting authority, followed by a variable length serial number, followed by a 'Kind' code (for example: A1, A2, B1) indicating the status of the document.</p> <p><b>WPI:</b> The unique identifier assigned to the patent in World Patent Index.</p> <p><b>Patent Title:</b> The WPI-enhanced patent title.</p> <p><b>GENESEQ records:</b> The number of sequences from this patent which are present in GENESEQ.</p> <p><b>Patent ASSIGNEE:</b> The company, institution or person to whom the patent has been assigned.</p> <p><b>Patent Inventors:</b> The names of the inventors given in the patent document.</p> <p><b>Publication Date:</b> The publication date of the patent.</p> <p><b>No. of pages:</b> The number of pages in the printed patent document.</p> <p><b>Language:</b> The language in which the patent document is printed.</p> <p><b>Patent Filing:</b> A number assigned to the patent application by the issuing patenting authority at the time of filing the application with that patenting authority.</p> <p><b>Priority:</b> One or more numbers assigned to the patent application by the first patenting authority to which the patent application is submitted. These numbers stay with the patent when it is subsequently filed with other patenting authorities.</p> <p><b>Patent Family:</b> Other published patent documents carrying the same priority number(s) and thus having the same content. Sometimes called Equivalent patents.</p>

Field	Description
Comments	GENESEQ abstract which highlights important aspects of the sequence in relation to the invention.
Sequence Length	The number of amino acid (AA) or nucleic acid (NA) residues in the sequence.
Sequence Location	Where the sequence can be found in the patent document.
Sequence	The sequence itself in FASTA format.
Organism	The Latin name of the source species for this sequence.
Keywords/Aliases	A set of keywords or phrases describing the sequence and patent content.
Similar Nucleotides	<p><b>GI:</b> A unique identifier assigned to all sequence depositions to the NCBI. Integer-based.DB:</p> <p><b>Accession:</b> A unique identifier assigned to a biological polymer sequence submitted to a sequence database.</p> <p><b>Organism:</b> The name of the source species for this sequence.</p>

### Sequence Summary

Provides an integrated view of all annotation associated with a record in BONDplus. To display the Sequence Summary view, select Sequence Summary in the View list when viewing a record in a GENESEQ Single Record View.



The Sequence Summary view is displayed:



► Identifier search - GENESEQ


IDENTIFIER SEARCH

GENESEQ Accession ▼ AAF84674

Records per page: 20 ▼

SEARCH

RESET



Summary [1]

Sequences [0]

Interactions [0]

Complexes [0]

Pathways [0]

GENESEQ [1]

☐ GENESEQ Filters
 View: Sequence Summary ▼ Export Results: Select an E

[\[Expand all\]](#) [\[Collapse all\]](#)

This report summarizes known annotation on the 5086 NA DNA sequence contained in GENESEQ Accession AAF84  
[Accession View](#)

Sequence Summary

Sequence Type: DNA


Sequence Length: 5086 NA

☐ Sequence: FASTA format [BLAST this sequence...](#)

☐ GENESEQ Records: This sequence shared by 1 GENESEQ Accession(s).

☐ Patents: This sequence contained in 1 Patent(s).

☐ Keywords/Aliases: 5 curated Keywords/aliases found.

By default, the Sequence Summary GENESEQ Single Record View shows a collapsed form of the information, with summary descriptions of each collapsed section, and an expander icon  to show the annotation of that particular section.

By clicking Expand All or Collapse All, you can expand and collapse all annotation for the given record.



[\[Expand all\]](#) [\[Collapse all\]](#)

This report summarizes known annotation on the 5086 NA DNA sequence contained in GENESEQ Accession A/

### Sequence Summary

**Sequence Type:** DNA

**Sequence Length:** 5086 NA

☐ **Sequence:** FASTA format [BLAST this sequence...](#)

1	AATTCTCATG	TTTGACAGCT	TATCATCGAT	AAGCTAGCTT	TAATGCGC
61	GTAAATTGC	TAACGCAGTC	AGGCACCGTG	TATGAAATCT	AACAATGC
121	CCTCGGCACC	GTCACCCTGG	ATGCTGTAGG	CATAGGCTTG	GTTATGCG
181	CCTCTTGCGG	GATATAGTCC	ATTCCGACAG	CATCGCCAGT	CACTATGC
241	GCTATATGCG	TTGATGCAAT	TTCTATGCGC	ACCCGTTCTC	GGAGCAC
301	TGGCCGCCGC	CCAGTCCTGC	TCGCTTCGCT	ACTTGGAGCC	ACTATCGA

Hyperlinks are provided for quick access to the original source of annotation.

☐

Patents:

Patent	Title	Patentee	Pub
<div> <div>WO200125451-A2</div> <div>Find this patent in...</div> </div>	One step cloning vector for quicker cloning of nucleic acid fragments comprises both a positive selection gene and a gene which protects against a selective agent	TRANSGENOMIC INC	12 / 200

☐

Keywords/Aliases:

Curated Keywords/Aliases

- nmTet1

Field	Description
Sequence Type	The specific kind of amino acid or nucleic acid sequence (for example: protein, peptide, DNA, RNA, and so on).
Sequence Length	The number of amino acid (AA) or nucleic acid (NA) residues in the sequence.
Sequence	The sequence itself in FASTA format.
GENESEQ Records	<p><b>Accession:</b> A unique identifier assigned to a biological polymer sequence submitted to a sequence database.</p> <p><b>Title:</b> A single line description of the sequence.</p> <p><b>Sequence Location:</b> Where the sequence can be found in the patent document.</p> <p><b>Organism:</b> The Latin name of the source species for this sequence.</p> <p><b>Date:</b> The date the record was issued in GENESEQ.</p>

Field	Description
Patents	<p><b>Patent:</b> The publication number of the patent document, composed of the two-letter country code of the issuing patenting authority, followed by a variable length serial number, followed by a 'Kind' code (for example: A1, A2, B1) indicating the status of the document.</p> <p><b>Title:</b> The WPI-enhanced title of the patent document.</p> <p><b>Patentee:</b> The company, institution or person to whom the patent has been assigned.</p> <p><b>Publication Date:</b> The publication date of the patent document.</p> <p><b># of sequence:</b> The number of sequences from this patent which are present in GENESEQ.</p> <p><b>Accession:</b> A unique identifier assigned to a biological polymer sequence submitted to a sequence database.</p>
Keywords/Aliases	A set of keywords or phrases describing the sequence and patent content.

## Exporting GENESEQ Search Data

You can export your GENESEQ search results data from either a GENESEQ Summary or Single Record View.

### Exporting Data From a GENESEQ Summary or Single Record View

1. To export search results data displayed in a GENESEQ Summary or Single Record View, on the GENESEQ tab choose a format from the Export Results list.



**NOTE:** For information about the formats available, see “GENESEQ Export Formats” on page 111.

2. A confirmation box appears asking if you want to open or save the file.



**NOTE:** The confirmation box may appear different than that shown above, depending on the browser you are using (for example, Internet Explorer or Mozilla Firefox). Also, your browser might attempt to block the download. Check your browser's settings to correct this (see "Appendix C Frequently Asked Questions" on page 141).

3. Choose the Save or Save As option.
4. Depending on the browser you are using, the file is downloaded into your browser's default download location or a dialog box appears allowing you to choose the folder to which you want to download the file.

## GENESEQ Export Formats

The following GENESEQ export formats are available in BONDplus:

Format	Description
Accession List	Export search results as a list of GENESEQ accessions, one per line.
Patent List	Export all patents which contain the GENESEQ Accessions returned by your query, one per line.
Redundant Accessions	Export search results as a list of GENESEQ accessions, including all accessions redundant to that accession. One redundant accession set per line.
FASTA Format	The sequence data contained by all component records returned in FASTA format. For more information, see <a href="http://www.ncbi.nlm.nih.gov/blast/fasta.shtml">http://www.ncbi.nlm.nih.gov/blast/fasta.shtml</a> .
Flat File Format	Export GENESEQ records in an EMBL-like Flat File format.
PDF	The current record summarized in Adobe Portable Document Format (PDF).



# Appendix A

## About Ontoglyphs

---





Ontoglyphs symbolize the biological attributes of molecules. The term is derived from the concepts of Gene Ontology (Onto) terms and symbols or pictorial representations (Glyphs).



In total, there are 83 symbols that represent function, binding, and cellular localization. Ontoglyphs are derived from a combination of the NCBI's Cluster of Orthologous Groups (COGs) functional categories and GO terms, and are based on grouping the nearly 18,000 GO terms into the categories used most frequently by biologists in describing gene and protein function.

### Detailed Ontoglyph Descriptions by Category





This section describes each ontoglyph according to category.





## Function

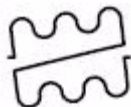


Category	Glyph
<p><b>Alcohol &amp; aromatic compound metabolism/transport</b></p> <p>The chemical reactions and physical changes involving alcohol (alkyl compound containing a hydroxyl group) or aromatic compound or the directed movement of any of alcohol or aromatic compound, into, out of, within or between cells.</p> <ul style="list-style-type: none"> <li>• GO:0004026 alcohol O-acetyltransferase activity</li> <li>• GO:0006066 alcohol metabolism</li> <li>• GO:0006725 aromatic compound metabolism</li> <li>• GO:0008414 CDP-alcohol phosphotransferase activity</li> <li>• GO:0015665 alcohol transporter activity</li> <li>• GO:0015850 organic alcohol transport</li> <li>• GO:0016614 oxidoreductase activity, acting on CH-OH group of donors</li> <li>• GO:0016773 phosphotransferase activity, alcohol group as acceptor</li> <li>• GO:0017169 CDP-alcohol phosphatidyltransferase activity</li> <li>• Links: Alcohol Metabolism Alcohol Metabolism Effects</li> </ul>	
<p><b>Behavior</b></p> <p>Specific actions/reactions of a cell or an organism, induced by stimuli from the environment. Patterned activity of a whole organism in a manner dependent upon some combination of that organism's internal state and external conditions.</p> <ul style="list-style-type: none"> <li>• GO:0007610 behavior</li> <li>• Links: Sensory Systems</li> </ul>	
<p><b>Binding activity</b></p> <p>The selective and stoichiometric interaction of a molecule with one or more specific sites on another molecule.</p> <ul style="list-style-type: none"> <li>• GO:0005488 binding</li> <li>• Links: Binding activity</li> </ul>	
<p><b>Biomembranes: organization and basic functions</b></p> <p>Any biological process which play a role in the structural organization of phospholipids and integral proteins in all biological membranes as well as the functions of the plasma membrane in prokaryotes and eukaryotes.</p> <ul style="list-style-type: none"> <li>• GO:0006944 membrane fusion</li> <li>• GO:0007016 cytoskeletal anchoring</li> <li>• GO:0009935 nutrient uptake</li> <li>• GO:0016044 membrane organization and biogenesis</li> <li>• Links: Biomembranes: Structural Organization and Basic Functions Common Features of Biological Membranes</li> </ul>	




Category	Glyph
<p><b>Carbohydrate metabolism/transport</b></p> <p>The chemical reactions and physical changes involving carbohydrates or any cellular process involved in the directed movement of carbohydrates into, out of, within or between cells.</p> <ul style="list-style-type: none"> <li>• GO:0004133 glycogen debranching enzyme activity</li> <li>• GO:0005975 carbohydrate metabolism</li> <li>• GO:0008643 carbohydrate transport</li> <li>• GO:0015066 alpha-amylase inhibitor activity</li> <li>• GO:0015144 carbohydrate transporter activity</li> <li>• GO:0015712 hexose phosphate transport</li> <li>• GO:0015920 lipopolysaccharide transport</li> <li>• GO:0016228 aldehyde-lyase activity</li> <li>• GO:0016744 transferase activity, transferring aldehyde or ketonic groups</li> <li>• GO:0016757 transferase activity, transferring glycosyl groups</li> <li>• GO:0016798 hydrolase activity, acting on glycosyl bonds</li> <li>• GO:0016832 aldehyde-lyase activity</li> <li>• GO:0016857 racemase and epimerase activity, acting on carbohydrates and derivatives</li> <li>• GO:0016861 intramolecular oxidoreductase activity, interconverting aldoses and ketoses</li> <li>• GO:0016862 intramolecular oxidoreductase activity, interconverting keto- and enol-groups</li> <li>• GO:0019200 carbohydrate kinase activity</li> <li>• Links: Overview of Carbohydrate Metabolism Carbohydrates</li> </ul>	
<p><b>Cell communication</b></p> <p>Any process other than signal transduction and response to extra-cellular stimulus that mediates interactions between a cell and its surroundings such as cell-cell signalling (interaction of cells with one another), cell adhesion (attachment of a cell, either to another cell or to the extra-cellular matrix), cell invasion and host-pathogen interaction.</p> <ul style="list-style-type: none"> <li>• GO:0007154 cell communication</li> <li>• Links: Cell-to-Cell Signaling: Hormones and Receptors</li> </ul>	



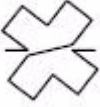





Category	Glyph
<b>Cell motility and structural activity</b> Any process involving the controlled movement of a cell or the action of a molecule that contributes to the regulation of cytoskeletal structures or to the structural integrity of a complex or assembly within or outside a cell. <ul style="list-style-type: none"> <li>• GO:0003774 motor activity</li> <li>• GO:0005198 structural molecule activity</li> <li>• GO:0006928 cell motility</li> <li>• GO:0008580 cytoskeletal regulator activity</li> <li>• Links: Cell Motility and Shape II: Microtubules and Intermediate Filaments Molecular Motors Cell Motility and Shape I: Microfilaments</li> </ul>	
<b>Cell multiplication</b> Any process which results in an increase in the number of cells by division. <ul style="list-style-type: none"> <li>• GO:0008283 cell proliferation</li> <li>• Links: Regulation of the Eukaryotic Cell Cycle Cell proliferation The Life Cycle of Cells</li> </ul>	
<b>Cell physiology</b> Any biological process other than cell motility, cell proliferation or cell biogenesis, required for the survival and growth of a cell. This encompasses actions such as cell growth (increase in cytoplasmic volume of an individual cell), homeostasis (maintenance of an internal equilibrium within a cell), cellularization (the enveloping of each nucleus in a syncytium by a plasma membrane), autophagy (self-digestion of cytoplasmic components), budding (form of asexual reproduction) and sporulation (the formation of spores). <ul style="list-style-type: none"> <li>• GO:0001775 cell activation</li> <li>• GO:0019725 cell homeostasis</li> <li>• GO:0050794 regulation of cellular process</li> <li>• GO:0050875 cellular physiological process</li> <li>• Links: The Dynamic Cell</li> </ul>	
<b>Cellular organization/biogenesis</b> Any cellular process involved in the structural and functional differentiation as well as the organization of components within a cell. <ul style="list-style-type: none"> <li>• GO:0016043 cell organization and biogenesis</li> <li>• Links: The Architecture of Cells</li> </ul>	





Category	Glyph
<p><b>Coenzymes vitamin &amp; hormone metabolism/transport</b></p> <p>The chemical reactions and physical changes involving coenzymes, prosthetic groups, vitamins and hormones or the directed movement of coenzymes, prosthetic groups, vitamins and hormones, into, out of, within or between cells.</p> <ul style="list-style-type: none"> <li>• GO:0004788 thiamin diphosphokinase activity</li> <li>• GO:0005472 flavin adenine dinucleotide carrier activity</li> <li>• GO:0005476 carnitine\;acyl carnitine carrier activity</li> <li>• GO:0006731 coenzyme and prosthetic group metabolism</li> <li>• GO:0006766 vitamin metabolism</li> <li>• GO:0008518 reduced folate carrier activity</li> <li>• GO:0009030 thiamin-phosphate kinase activity</li> <li>• GO:0015223 vitamin or cofactor transporter activity</li> <li>• GO:0015349 thyroid hormone transporter activity</li> <li>• GO:0015403 thiamin uptake transporter activity</li> <li>• GO:0015875 vitamin or cofactor transport</li> <li>• GO:0042445 hormone metabolism</li> <li>• Links: Vitamins Are Often Precursors to Coenzymes</li> </ul>	
<p><b>Death</b></p> <p>A permanent cessation of all vital functions: the end of life; can be applied to a whole organism or to a part of an organism.</p> <ul style="list-style-type: none"> <li>• GO:0008219 cell death</li> <li>• GO:0016265 death</li> <li>• GO:0016329 apoptosis regulator activity</li> <li>• Links: Apoptosis Cell Death and Its Regulation</li> </ul>	
<p><b>Defense/immune response</b></p> <p>Any process/activity that plays a role in or triggers a reaction in the defence/immune response of an organism against infection and disease.</p> <ul style="list-style-type: none"> <li>• GO:0003793 defense/immunity protein activity</li> <li>• GO:0006952 defense response</li> <li>• Links: The Immune System</li> </ul>	
<p><b>Development</b></p> <p>Biological processes specifically aimed at the progression of a cell or an organism from a lower or simpler to a higher or more complex form.</p> <ul style="list-style-type: none"> <li>• GO:0007275 development</li> <li>• GO:0030154 cell differentiation</li> <li>• Links: Organogenesis, Histogenesis, and New Growth Gene Control in Development Cell Interactions in Development</li> </ul>	

Category	Glyph
<b>DNA metabolism</b> The chemical reactions and physical changes involving deoxyribonucleic acid (DNA), one of the two main types of nucleic acid, consisting of a long, unbranched macromolecule formed from one, or more commonly, two, strands polydeoxyribonucleotides. <ul style="list-style-type: none"> <li>• GO:0006259 DNA metabolism</li> <li>• Links: DNA Replication, Repair, and Recombination Summary</li> </ul>	
<b>Energy production/conversion</b> Metabolic processes by which cells make energy (ATP) from either light or by oxidation of organic and reduced inorganic compounds; or metabolic processes by which cells release energy from ATP to either drive the production of organic molecules or generate forces for others cellular mechanisms. <ul style="list-style-type: none"> <li>• GO:0005489 electron transporter activity</li> <li>• GO:0006091 energy pathways</li> <li>• GO:0006118 electron transport</li> <li>• GO:0015979 photosynthesis</li> <li>• GO:0016652 oxidoreductase activity, acting on NADH or NADPH, NAD or NADP as acceptor</li> <li>• GO:0016655 oxidoreductase activity, acting on NADH or NADPH, quinone or similar compound as acceptor</li> <li>• GO:0016676 oxidoreductase activity, acting on heme group of donors, oxygen as acceptor</li> <li>• GO:0016681 oxidoreductase activity, acting on diphenols and related substances as donors, cytochrome as acceptor</li> <li>• Links: Transducing and Storing Energy Cellular Energetics: Glycolysis, Aerobic Oxidation, and Photosynthesis</li> </ul>	
<b>General metabolism</b> The chemical reactions and physical processes undergone by substances other than amines, aromatic compounds, amino acids & derivatives, carbohydrates, coenzymes, lipids, proteins, vitamins, hormones, ions, nucleobases, nucleosides, nucleotides & nucleic acids, organic acids & alcohols. <ul style="list-style-type: none"> <li>• GO:0003824 catalytic activity</li> <li>• GO:0008152 metabolism</li> <li>• GO:0030234 enzyme regulator activity</li> <li>• Links: Metabolism: Basic Concepts and Design Review Metabolism</li> </ul>	



Category	Glyph
<p><b>General physiological processes</b></p> <p>Any process specifically pertinent to the functioning of integrated living units other than the following processes: cell growth and maintenance, death, metabolism, pathogenesis, photosynthesis, response to stress, endogenous or external stimulus.</p> <ul style="list-style-type: none"> <li>• GO:0007582 physiological process</li> <li>• GO:0008435 anticoagulant activity</li> <li>• GO:0016015 morphogen activity</li> <li>• GO:0016172 antifreeze activity</li> <li>• GO:0016209 antioxidant activity</li> <li>• GO:0045735 nutrient reservoir activity</li> <li>• Links: Integrating Cells into Tissues</li> </ul>	
<p><b>General transport</b></p> <p>Any biological process involved in or any molecular function which enables the directed movement of substances other than amines, aromatic compounds, amino acids &amp; derivatives, carbohydrates, coenzymes, lipids, proteins, vitamins, hormones, ions, nucleobases, nucleosides, nucleotides &amp; nucleic acids, organic acids &amp; alcohols.</p> <ul style="list-style-type: none"> <li>• GO:0005215 transporter activity</li> <li>• GO:0006810 transport</li> <li>• Links: Transport across Cell Membranes</li> </ul>	
<p><b>Ion transport</b></p> <p>The directed movement of charged atoms or small charged molecules other than carboxylic anions into, out of, within or between cells.</p> <ul style="list-style-type: none"> <li>• GO:0006811 ion transport</li> <li>• GO:0015075 ion transporter activity</li> <li>• GO:0015296 anion\;cation symporter activity</li> <li>• GO:0015301 anion\;anion antiporter activity</li> <li>• GO:0015491 cation\;cation antiporter activity</li> <li>• Links: Intracellular Ion Environment and Membrane Electric Potential</li> </ul>	



Category	Glyph
<b>Lipid metabolism/transport</b> The chemical reactions and physical changes involving lipids or any cellular process involved in the directed movement of lipids into, out of, within or between cells. <ul style="list-style-type: none"> <li>• GO:0001727 lipid kinase activity</li> <li>• GO:0005319 lipid transporter activity</li> <li>• GO:0006629 lipid metabolism</li> <li>• GO:0006869 lipid transport</li> <li>• GO:0016298 lipase activity</li> <li>• GO:0016627 oxidoreductase activity, acting on the CH-CH group of donors</li> <li>• Links: Fatty Acid Metabolism Overview of Lipid Function</li> </ul>	
<b>Organic &amp; amino acids &amp; derivatives &amp; amine metabolism/transport</b> The chemical reactions and physical changes involving organic acids or any organic compound containing an amino or a substituted amino group and their derivatives or the directed movement of organic acids and organic compounds containing one or more amino groups, into, out of, within or between cells. <ul style="list-style-type: none"> <li>• GO:0005275 amine transporter activity</li> <li>• GO:0005342 organic acid transporter activity</li> <li>• GO:0006082 organic acid metabolism</li> <li>• GO:0006519 amino acid and derivative metabolism</li> <li>• GO:0009308 amine metabolism</li> <li>• GO:0015197 peptide transporter activity</li> <li>• GO:0015557 arginine targeting transporter activity</li> <li>• GO:0015837 amine transport</li> <li>• GO:0015849 organic acid transport</li> <li>• GO:0019202 amino acid kinase activity</li> <li>• Links: The Biosynthesis of Amino Acids</li> </ul>	
<b>Pathogenesis &amp; toxin activity</b> Specific processes that generate the ability of an organism to cause disease in another or the function of some compounds to cause injury to other living organisms. <ul style="list-style-type: none"> <li>• GO:0009404 toxin metabolism</li> <li>• GO:0009405 pathogenesis</li> <li>• GO:0015070 toxin activity</li> <li>• GO:0019534 toxin transporter activity</li> <li>• Links: Bacterial Pathogenesis Retroviral Pathogenesis Persistent Viral Infections</li> </ul>	

Category	Glyph
<b>Phosphorus metabolism</b> The chemical reactions and physical changes involving the non-metallic element phosphorus or compounds that contain phosphorus. <ul style="list-style-type: none"> <li>• GO:0006793 phosphorus metabolism</li> <li>• GO:0015405 P-P-bond-hydrolysis-driven transporter activity</li> <li>• GO:0016772 transferase activity, transferring phosphorus-containing groups</li> <li>• GO:0016793 triphosphoric monoester hydrolase activity</li> <li>• GO:0016794 diphosphoric monoester hydrolase activity</li> <li>• GO:0042578 phosphoric ester hydrolase activity</li> <li>• Links: Oxidative Phosphorylation</li> </ul>	
<b>Protein degradation</b> The breakdown into simpler components of a protein by the destruction of the native, active configuration, with or without the hydrolysis of peptide bonds. <ul style="list-style-type: none"> <li>• GO:0008233 peptidase activity</li> <li>• GO:0008539 proteasome inhibitor activity</li> <li>• GO:0017171 serine hydrolase activity</li> <li>• GO:0030163 protein catabolism</li> <li>• GO:0030414 protease inhibitor activity</li> <li>• Links: Protein Turnover and Amino Acid Catabolism</li> </ul>	
<b>Protein metabolism</b> Any process involved in protein folding (pathways to transform proteins from the unfolded state to the compact or native state), protein repair (pathways to restore a protein to its original state after damage) and protein complex assembly. <ul style="list-style-type: none"> <li>• GO:0003754 chaperone activity</li> <li>• GO:0005483 soluble NSF attachment protein activity</li> <li>• GO:0017028 protein stabilization activity</li> <li>• GO:0019538 protein metabolism</li> <li>• GO:0030188 chaperone regulator activity</li> <li>• Links: Overview of Protein Metabolism Protein Structure and Function</li> </ul>	


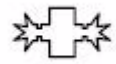
Category	Glyph
<b>Protein modification</b> Any process resulting in the co-translational or post-translational modification of one or more amino acid residues of a protein by either addition of other groups or by isomerization of residues. <ul style="list-style-type: none"> <li>GO:0004672 protein kinase activity</li> <li>GO:0006464 protein modification</li> <li>GO:0008639 small protein conjugating enzyme activity</li> <li>GO:0008641 small protein activating enzyme activity</li> <li>GO:0045182 translation regulator activity</li> <li>Links: Common covalent modifications of protein activity Covalent Modification Is a Means of Regulating Enzyme Activity</li> </ul>	
<b>Protein Synthesis, Processing, and Regulation</b> Process of translation, wherein an amino acid chain forming the basic protein (polypeptide), is constructed using mRNA as a template according to the rules specified by the genetic code. The polypeptide is further processed (folded) to produce its active form which may be subjected to mechanisms that control the activities within cells. <ul style="list-style-type: none"> <li>GO:0006412 protein biosynthesis</li> <li>GO:0017149 protein biosynthesis inhibitor activity</li> <li>GO:0030533 triplet codon-amino acid adaptor activity</li> <li>Links: The Three Roles of RNA in Protein Synthesis The Control of Gene Expression Protein Synthesis</li> </ul>	
<b>Protein transport</b> Any cellular process involved in the directed movement of proteins into, out of, within or between cells. This will also include mechanisms by which proteins are transported to the appropriate organelle for secretion or membrane insertion. <ul style="list-style-type: none"> <li>GO:0008104 protein localization</li> <li>GO:0008565 protein transporter activity</li> <li>GO:0015031 protein transport</li> <li>Links: Protein Sorting: Organelle Biogenesis and Protein Secretion</li> </ul>	
<b>Response to stress endogenous or exogenous stimulus</b> A change in state or activity of a cell or an organism that occurs as a response to stress, to endogenous or exogenous stimulus. <ul style="list-style-type: none"> <li>GO:0000156 two-component response regulator activity</li> <li>GO:0006950 response to stress</li> <li>GO:0009370 quorum sensing response regulator activity</li> <li>GO:0009605 response to external stimulus</li> <li>GO:0009719 response to endogenous stimulus</li> <li>Links: Cells Can Respond to Changes in Their Environments</li> </ul>	

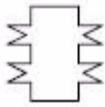






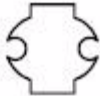


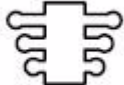

Category	Glyph
<p><b>RNA &amp; other nucleic acids metabolism and transport</b></p> <p>The chemical reactions and physical changes involving RNA (ribonucleic acid) and other nucleic acids (nucleobases, nucleosides, nucleotides and nucleic acids or the directed movement of nucleobases, nucleosides, nucleotides) into, out of, within or between cells.</p> <ul style="list-style-type: none"> <li>• GO:0004112 cyclic-nucleotide phosphodiesterase activity</li> <li>• GO:0004748 ribonucleoside-diphosphate reductase activity</li> <li>• GO:0005471 ATP\ADP antiporter activity</li> <li>• GO:0006139 nucleobase, nucleoside, nucleotide and nucleic acid metabolism</li> <li>• GO:0008998 ribonucleoside-triphosphate reductase activity</li> <li>• GO:0015931 nucleobase, nucleoside, nucleotide and nucleic acid transport</li> <li>• GO:0015932 nucleobase, nucleoside, nucleotide and nucleic acid transporter activity</li> <li>• GO:0016779 nucleotidyltransferase activity</li> <li>• GO:0017110 nucleoside-diphosphatase activity</li> <li>• GO:0019205 nucleobase, nucleoside, nucleotide kinase activity</li> <li>• GO:0030249 guanylate cyclase regulator activity</li> <li>• GO:0030695 GTPase regulator activity</li> <li>• Links: RNA Synthesis and Processing RNA Synthesis and Splicing Nucleotide Biosynthesis RNA Processing, Nuclear Transport, and Post-Transcriptional Control Nucleic Acids, the Genetic Code, and the Synthesis of Macromolecules</li> </ul>	
<p><b>Signal transduction</b></p> <p>The cascade of processes by which an extra-cellular signal interacting with a molecule at the cell surface, causes a change in the intra-cellular level of a mediator and/or second messenger and ultimately effects an alteration in the cellular activity and change in the program of genes expressed within the responding cells.</p> <ul style="list-style-type: none"> <li>• GO:0004872 receptor activity</li> <li>• GO:0005057 receptor signaling protein activity</li> <li>• GO:0007165 signal transduction</li> <li>• GO:0009927 histidine phosphotransfer kinase activity</li> <li>• GO:0017106 activin inhibitor activity</li> <li>• Links: Signal-Transduction Pathways</li> </ul>	



Category	Glyph
<b>Transcription</b> The synthesis of either RNA on a template of DNA or DNA on a template of RNA. <ul style="list-style-type: none"> <li>GO:0003969 RNA editase activity</li> <li>GO:0004386 helicase activity</li> <li>GO:0004407 histone deacetylase activity</li> <li>GO:0006350 transcription</li> <li>GO:0030528 transcription regulator activity</li> <li>Links: Basic Transcriptional Mechanism Regulation of Transcription Initiation</li> </ul>	
<b>Viral life cycle</b> Any process involved in the infection of a host cell by a virus, replication of the viral genome, and assembly of progeny particles. <ul style="list-style-type: none"> <li>GO:0016032 viral life cycle</li> <li>Links: Viruses: Structure, Function, and Uses Retroviral Virions and Genomes</li> </ul>	




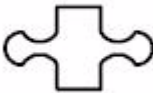
## Sub-Cellular Localization





Category	Glyph
<b>Actin cytoskeleton</b> The part of the dynamic network of protein fibers filling the cytoplasm of eukaryotic cells composed of actin filaments, which are stranded helical polymer of the protein actin. <ul style="list-style-type: none"> <li>GO:0015629 actin cytoskeleton</li> <li>Links: The Actin Cytoskeleton Yeast morphogenesis</li> </ul>	
<b>Axon or dendrite</b> The long usually single nerve-cell protrusion that conducts impulses away from the cell body or one of the threadlike extensions of the cytoplasm of a neuron that carries electrical signals from synapses to the cell body. <ul style="list-style-type: none"> <li>GO:0030424 axon</li> <li>GO:0030425 dendrite</li> <li>GO:0045202 synapse</li> <li>Links: Overview of Neuron Structure Schematic diagram of a typical vertebrate neuron</li> </ul>	

Category	Glyph
<p><b>Biological membrane</b></p> <p>A lipid bilayer of amphiphilic molecules embedded with proteins, which acts as a barrier within or around a cell. Such membranes typically define enclosed spaces or organelles inside the cells and are selectively permeable.</p> <ul style="list-style-type: none"> <li>GO:0016020 membrane</li> <li>Links: Faces of cellular membranes Common Features of Biological Membranes</li> </ul>	
<p><b>Cell periphery</b></p> <p>The structure that surrounds and separates a cell from its external environment, encloses the cytoplasm and forms a selective permeability barrier. It includes the cell membrane, the periplasm and any external encapsulating structure, and may consist of lipids, proteins and some carbohydrates.</p> <ul style="list-style-type: none"> <li>GO:0005886 plasma membrane</li> <li>GO:0005938 cell cortex</li> <li>GO:0030312 external encapsulating structure</li> <li>GO:0030428 septum</li> <li>GO:0042597 periplasmic space</li> <li>Links: Plasma membrane Cell wall Structure of the secondary cell wall</li> </ul>	
<p><b>Cytoplasm</b></p> <p>The viscous, translucent colloid that surrounds all other cell organelles outside the nucleus and within the plasma membrane.</p> <ul style="list-style-type: none"> <li>GO:0005737 cytoplasm</li> <li>Links: Cytoplasm</li> </ul>	
<p><b>Cytoplasmic vesicle</b></p> <p>Small enclosed (self-contained) compartment found in the cytoplasm of the cell.</p> <ul style="list-style-type: none"> <li>GO:0016023 cytoplasmic vesicle</li> <li>Links: Model for the formation of a clathrin-coated vesicle Model for formation of COP I-coated vesicles Cytoplasmic vesicle Cytoplasmic micro and macrovesicles</li> </ul>	
<p><b>Endoplasmic reticulum</b></p> <p>The network of folded membrane channels with or without ribosomes that occur in the cytoplasm of many eukaryotic cells and transport molecules between and within cells.</p> <ul style="list-style-type: none"> <li>GO:0005783 endoplasmic reticulum</li> <li>GO:0005789 endoplasmic reticulum membrane</li> <li>Links: Cell organelles</li> </ul>	



Category	Glyph
<b>Endosome</b> Nonlysosomal cytoplasmic vesicles either derived from the plasma membrane or formed when coated vesicles shed their clathrin coat. Endosomes internalise macromolecules bound by receptors on the cell surface. <ul style="list-style-type: none"> <li>• GO:0005768 endosome</li> <li>• GO:0030139 endocytic vesicle</li> <li>• Links: Transport from the Plasma Membrane via Endosomes Endosome</li> </ul>	
<b>Extracellular /cell surface</b> Extracellular region is the space external to the outermost structure of a cell or virus and cell surface is the external part of the cell wall and/or cell membrane. <ul style="list-style-type: none"> <li>• GO:0005576 extracellular</li> <li>• GO:0009986 cell surface</li> <li>• Links: Schematic overview of the extracellular matrix</li> </ul>	
<b>Flagellum /cilium</b> Flagellum and cilium are protrusions from cells made up of microtubules and providing cellular locomotion or moving fluid/mucous. <ul style="list-style-type: none"> <li>• GO:0005929 cilium</li> <li>• GO:0005930 axoneme</li> <li>• GO:0009289 fimbria</li> <li>• GO:0019861 flagellum</li> <li>• Links: Structure of ciliary and flagellar axonemes Cilium Flagellum Cilia and Flagella: Structure and Movement</li> </ul>	
<b>Golgi apparatus</b> Stacks of membrane-bounded cisternae located between the endoplasmic reticulum and the cell surface that function in processing, packaging, and trafficking of macromolecules. The Golgi apparatus is mainly devoted to processing the proteins synthesized in the endoplasmic reticulum (ER). <ul style="list-style-type: none"> <li>• GO:0005794 Golgi apparatus</li> <li>• GO:0017122 UDP-N-acetylglucosamine-peptide N-acetylglucosaminyl-transferase complex</li> <li>• Links: Golgi complex Three-dimensional model of the Golgi complex</li> </ul>	
<b>Lipid particle</b> Any particle of coalesced lipids in the cytoplasm of a cell. <ul style="list-style-type: none"> <li>• GO:0005811 lipid particle</li> <li>• Links: Lipid particles of the yeast</li> </ul>	

Category	Glyph
<p><b>Microtubule cytoskeleton</b></p> <p>The part of the dynamic network of protein fibers filling the cytoplasm of eukaryotic cells that is composed of microtubules, which are long and generally straight hollow tubes of proteins.</p> <ul style="list-style-type: none"> <li>• GO:0005815 microtubule organizing center</li> <li>• GO:0005819 spindle</li> <li>• GO:0015630 microtubule cytoskeleton</li> <li>• GO:0045111 intermediate filament cytoskeleton</li> <li>• Links: Microtubule Microtubule structure Intermediate Filaments</li> </ul>	
<p><b>Mitochondrion</b></p> <p>The modified double unit membrane organelle present in high number to long branched structures. Mitochondria play an important role in the regulation of cell death and the major function is to generate energy for cellular function.</p> <ul style="list-style-type: none"> <li>• GO:0000262 mitochondrial chromosome</li> <li>• GO:0005739 mitochondrion</li> <li>• GO:0005740 mitochondrial membrane</li> <li>• GO:0016469 proton-transporting two-sector ATPase complex</li> <li>• GO:0016472 sodium-transporting two-sector ATPase complex</li> <li>• GO:0017071 intracellular cyclic nucleotide activated cation channel complex</li> <li>• GO:0042645 mitochondrial nucleoid</li> <li>• GO:0045251 electron transfer flavoprotein complex</li> <li>• GO:0045252 oxoglutarate dehydrogenase complex</li> <li>• GO:0045253 pyruvate dehydrogenase (lipoamide) phosphatase complex</li> <li>• GO:0045254 pyruvate dehydrogenase complex</li> <li>• GO:0045257 succinate dehydrogenase complex (ubiquinone)</li> <li>• GO:0045259 proton-transporting ATP synthase complex</li> <li>• GO:0045271 respiratory chain complex I</li> <li>• GO:0045273 respiratory chain complex II</li> <li>• GO:0045275 respiratory chain complex III</li> <li>• GO:0045277 respiratory chain complex IV</li> <li>• Links: Mitochondrion</li> </ul>	




Category	Glyph
<b>Nuclear periphery</b> The nuclear envelope that surrounds the nucleus of eukaryotic cells and is composed of the nuclear membrane, the nuclear pore complexes, and the nuclear lamina. <ul style="list-style-type: none"> <li>• GO:0005635 nuclear membrane</li> <li>• GO:0005640 nuclear outer membrane</li> <li>• GO:0005643 nuclear pore</li> <li>• GO:0005653 nuclear membrane lumen</li> <li>• GO:0016363 nuclear matrix</li> <li>• Links: Nuclear periphery</li> </ul>	
<b>Nucleolus</b> A small dense body (sub organelle) within the nucleus of eukaryotic cells, not delimited by a membrane in which ribosomal RNA are synthesised and assembled into ribonucleoprotein subunits of ribosomes. <ul style="list-style-type: none"> <li>• GO:0005730 nucleolus</li> <li>• Links: Nucleolus</li> </ul>	
<b>Nucleus</b> The membrane-bounded spherical or ovoid compartment of a cell, which contains the chromosomes and associated molecules that control the characteristics and growth of the cell. <ul style="list-style-type: none"> <li>• GO:0000228 nuclear chromosome</li> <li>• GO:0000307 cyclin-dependent protein kinase holoenzyme complex</li> <li>• GO:0005634 nucleus</li> <li>• Links: Nucleus</li> </ul>	
<b>Peroxisome</b> A small, membrane-bounded organelle that uses molecular oxygen to oxidize organic molecules; contains some enzymes that produce and others that degrade hydrogen peroxide. <ul style="list-style-type: none"> <li>• GO:0005777 peroxisome</li> <li>• Links: Peroxisome</li> </ul>	




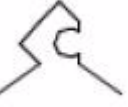




Category	Glyph
<p><b>Plastid</b></p> <p>Any member of a family of partially autonomous organelles found in the cytoplasm of plants and some protists, surrounded by a double membrane, containing some DNA, RNA and ribosomes, and reproducing by binary fission.</p> <ul style="list-style-type: none"> <li>• GO:0009508 plastid chromosome</li> <li>• GO:0009527 plastid outer membrane</li> <li>• GO:0009528 plastid inner membrane</li> <li>• GO:0009536 plastid</li> <li>• GO:0009579 thylakoid</li> <li>• GO:0046858 chlorosome</li> <li>• Links: Structure of a leaf and chloroplast</li> </ul>	
<p><b>Protoplasm</b></p> <p>The viscid and granular living contents of a cell; contained within (but not including) the plasma membrane. This category includes matters that cannot be easily classified either under nucleus or cytoplasm but does not include nucleus and cytoplasm.</p> <ul style="list-style-type: none"> <li>• GO:0005622 intracellular</li> <li>• Links: Protoplasm</li> </ul>	
<p><b>Ribonucleoprotein complex</b></p> <p>A macromolecular complex containing both protein and RNA molecules.</p> <ul style="list-style-type: none"> <li>• GO:0000178 exosome (RNase complex)</li> <li>• GO:0005971 ribonucleoside-diphosphate reductase complex</li> <li>• GO:0030529 ribonucleoprotein complex</li> <li>• Links: Stepwise Formation of Proteins on Ribosomes</li> </ul>	
<p><b>Site of polarized growth</b></p> <p>Any part of a cell where non-isotropic growth takes place. This includes the bud neck (the constriction between a daughter cell (bud) and the mother cell in an organism that reproduces by budding), the bud tip (the end of a bud distal to the site of attachment to the mother cell), the shmoo tip (the narrow end of the pear-shaped cell or shmoo formed in response to mating pheromone by cells of <i>Saccharomyces</i> and other similar fungi) and the portion of the budding yeast plasma membrane where a daughter cell will emerge.</p> <ul style="list-style-type: none"> <li>• GO:0005933 bud</li> <li>• GO:0030427 site of polarized growth</li> <li>• Links: Polarized growth of hypha The budding yeast <i>S. cerevisiae</i></li> </ul>	



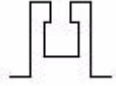

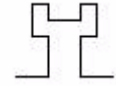


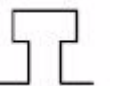
Category	Glyph
<b>Vacuole</b> Large membrane-bound compartments within some eukaryotic cells that can serve different purposes, such as capturing food materials or unwanted structural debris surrounding the cell, sequestering materials that might be toxic to cells, maintaining fluid balance within the cell, exporting unwanted substances from the cell, and even determining relative cell size. In plant cells, they provide both storage and space filling and in protozoa function in osmoregulation. <ul style="list-style-type: none"> <li>• GO:0005773 vacuole</li> <li>• Links: The plant cell vacuole</li> </ul>	
<b>Virion provirus</b> The complete fully infectious extracellular virus particle or retrovirus that has integrated itself into the DNA of a host cell. <ul style="list-style-type: none"> <li>• GO:0019012 virion</li> <li>• GO:0019038 provirus</li> <li>• GO:0046816 virion transport vesicle</li> <li>• Links: Models for retroviral structure Assembly of bacteriophage 1 virions</li> </ul>	

## Binding

Category	Glyph
<b>Adenyl nucleotide binding</b> Interacting selectively with adenyl nucleotides, any compound consisting of adenosine esterified with (ortho)phosphate. <ul style="list-style-type: none"> <li>• GO:0030554 adenyl nucleotide binding</li> <li>• Links: Adenyl nucleotide binding</li> </ul>	
<b>Amine binding</b> Interacting selectively with any organic compound that is weakly basic in character and contains an amino or a substituted amino group. <ul style="list-style-type: none"> <li>• GO:0043176 amine binding</li> </ul>	
<b>Antigen binding</b> Interacting selectively with an antigen, any substance which is capable of inducing a specific immune response and of reacting with the products of that response, the specific antibody or specifically sensitized T-lymphocytes, or both. Binding may counteract the biological activity of the antigen. <ul style="list-style-type: none"> <li>• GO:0003823 antigen binding</li> <li>• Links: Antigen binding</li> </ul>	

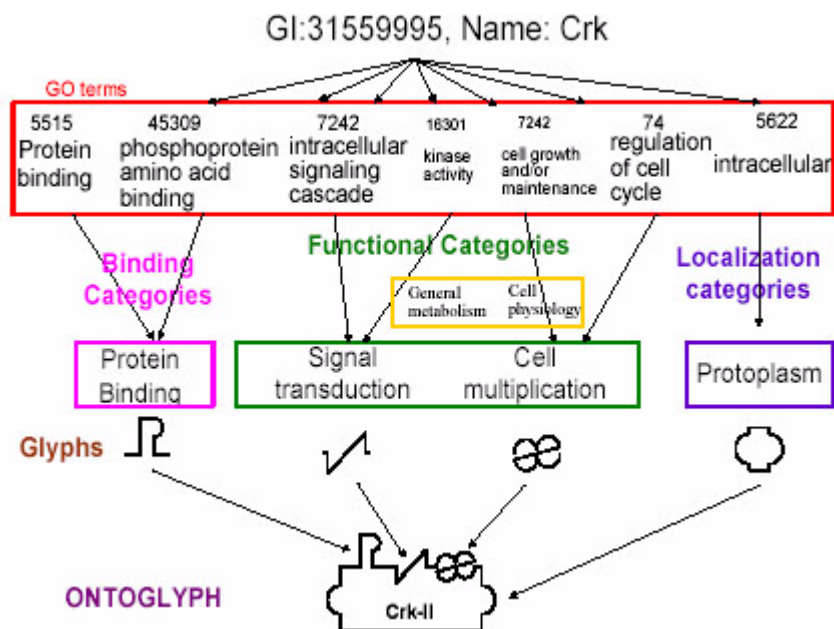
Category	Glyph
<b>ATP binding</b> Interacting selectively with ATP, adenosine 5'-triphosphate, a universally important coenzyme and enzyme regulator. <ul style="list-style-type: none"> <li>GO:0005524 ATP binding</li> <li>Links: ATP binding</li> </ul>	
<b>Calmodulin binding</b> Interacting selectively with calmodulin, a calcium-binding protein with many roles, both in the calcium-bound and calcium-free states. <ul style="list-style-type: none"> <li>GO:0005516 calmodulin binding</li> <li>Links: Calmodulin binding</li> </ul>	
<b>Carbohydrate binding</b> Interacting selectively with any carbohydrate. <ul style="list-style-type: none"> <li>GO:0030246 carbohydrate binding</li> <li>Links: Carbohydrate binding</li> </ul>	
<b>Coenzyme binding</b> Interacting selectively with a coenzyme, any nonprotein organic cofactor that is required (in addition to an enzyme and a substrate) for a reaction to proceed. <ul style="list-style-type: none"> <li>GO:0050662 coenzyme binding</li> <li>Links: Coenzyme binding</li> </ul>	
<b>Cytokine binding</b> Interacting selectively with a cytokine, any of a group of proteins that function to control the survival, growth and differentiation of tissues and cells, and which have autocrine and paracrine activity. <ul style="list-style-type: none"> <li>GO:0019955 cytokine binding</li> <li>Links: Cytokine binding</li> </ul>	
<b>Cytoskeletal protein binding</b> Interacting selectively with any protein component of any cytoskeleton (actin, microtubule, or intermediate filament cytoskeleton). <ul style="list-style-type: none"> <li>GO:0008092 cytoskeletal protein binding</li> <li>Links: Cytoskeletal protein binding</li> </ul>	
<b>Damage DNA binding</b> Interacting selectively with damaged DNA (deoxyribonucleic acid). <ul style="list-style-type: none"> <li>GO:0003684 damaged DNA binding</li> <li>Links: Damage DNA binding</li> </ul>	
<b>DNA binding</b> Interacting selectively with DNA (deoxyribonucleic acid). <ul style="list-style-type: none"> <li>GO:0003677 DNA binding</li> <li>Links: DNA binding</li> </ul>	

Category	Glyph
<b>Double stranded DNA binding</b> Interacting selectively with double-stranded DNA (deoxyribonucleic acid). <ul style="list-style-type: none"> <li>GO:0003690 double-stranded DNA binding</li> <li>Links: Double stranded DNA binding</li> </ul>	
<b>Guanyl nucleotide binding</b> Interacting selectively with guanyl nucleotides, any compound consisting of guanosine esterified with (ortho)phosphate. <ul style="list-style-type: none"> <li>GO:0019001 guanyl nucleotide binding</li> <li>Links: Guanyl nucleotide binding</li> </ul>	
<b>Ion binding</b> Interacting selectively with ions, charged atoms or small charged molecules. <ul style="list-style-type: none"> <li>GO:0043167 ion binding</li> <li>Links: Ion binding</li> </ul>	
<b>Lipid binding</b> Interacting selectively with a lipid. <ul style="list-style-type: none"> <li>GO:0008289 lipid binding</li> <li>Links: Lipid binding</li> </ul>	
<b>Metal ion binding</b> Interacting selectively with any metal ion. <ul style="list-style-type: none"> <li>GO:0046872 metal ion binding</li> <li>Links: Metal ion binding</li> </ul>	
<b>mRNA binding</b> Interacting selectively with pre-messenger RNA (pre-mRNA) or messenger RNA (mRNA). <ul style="list-style-type: none"> <li>GO:0003729 mRNA binding</li> <li>Links: mRNA binding</li> </ul>	
<b>Nucleic acid binding</b> Interacting selectively with any nucleic acid. <ul style="list-style-type: none"> <li>GO:0003676 nucleic acid binding</li> <li>Links: Nucleic acid binding</li> </ul>	
<b>Nucleotide binding</b> Interacting selectively with a nucleotide, any compound consisting of a nucleoside that is esterified with (ortho)phosphate or an oligophosphate at any hydroxyl group on the glucose moiety. <ul style="list-style-type: none"> <li>GO:0000166 nucleotide binding</li> <li>Links: Nucleotide binding</li> </ul>	

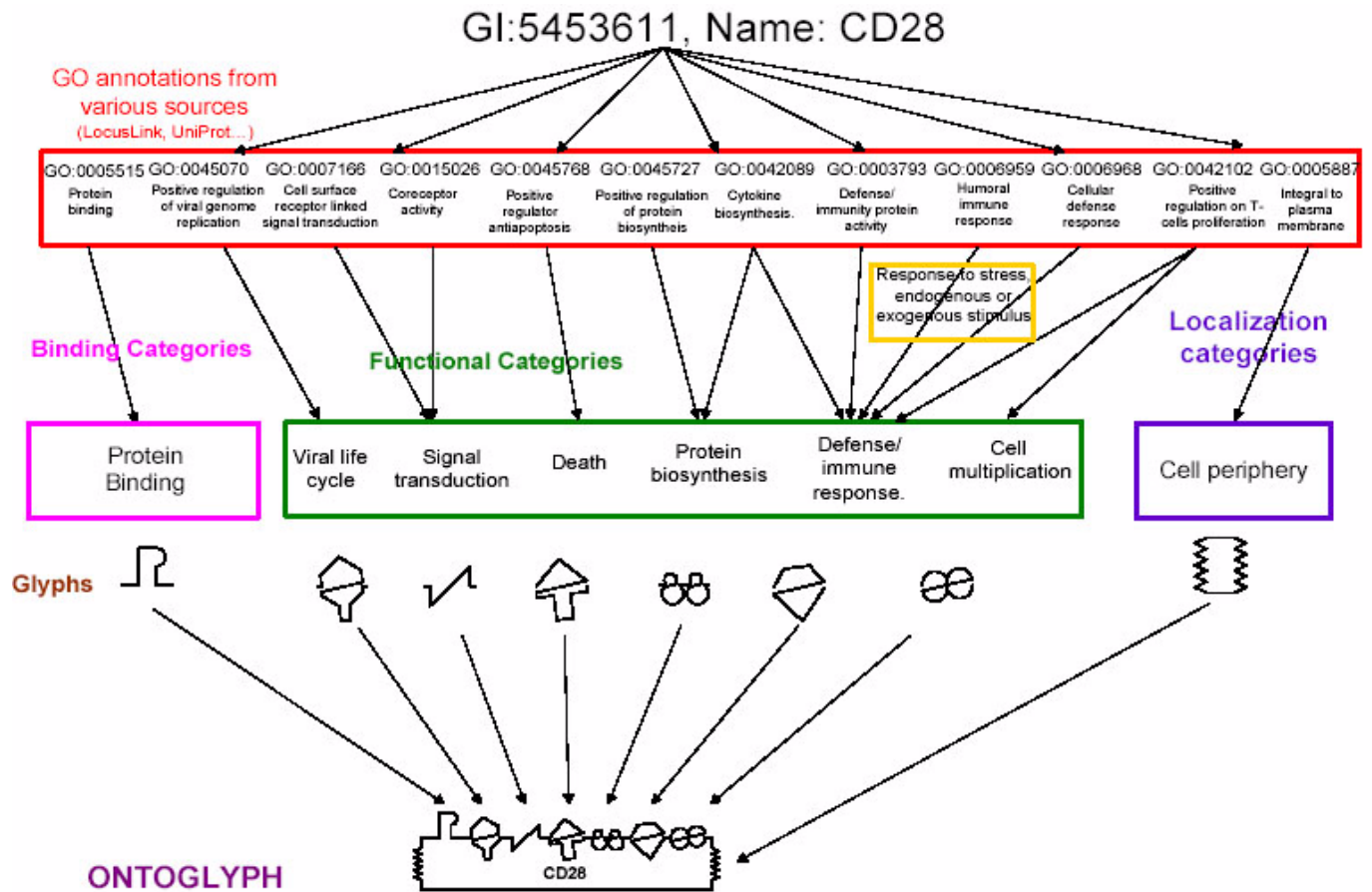
Category	Glyph
<b>Oxygen binding</b> Interacting selectively with oxygen (O <sub>2</sub> ). <ul style="list-style-type: none"> <li>GO:0019825 oxygen binding</li> <li>Links: Oxygen binding</li> </ul>	
<b>Protein binding</b> Interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules). <ul style="list-style-type: none"> <li>GO:0005515 protein binding</li> <li>Links: Protein binding</li> </ul>	
<b>Receptor binding</b> Interacting selectively with one or more specific sites on a receptor molecule, a macromolecule that undergoes combination with a hormone, neurotransmitter, drug or intracellular messenger to initiate a change in cell function. <ul style="list-style-type: none"> <li>GO:0005102 receptor binding</li> <li>Links: Receptor binding</li> </ul>	
<b>RNA binding</b> Interacting selectively with an RNA molecule or a portion thereof. <ul style="list-style-type: none"> <li>GO:0003723 RNA binding</li> <li>Links: RNA binding</li> </ul>	
<b>Single stranded DNA binding</b> Interacting selectively with single-stranded DNA (deoxyribonucleic acid). <ul style="list-style-type: none"> <li>GO:0003697 single-stranded DNA binding</li> <li>Links: Single stranded DNA binding</li> </ul>	
<b>Transcription factor binding</b> Interacting selectively with a transcription factor, any protein required to initiate or regulate transcription. <ul style="list-style-type: none"> <li>GO:0008134 transcription factor binding</li> <li>Links: Transcription factor binding</li> </ul>	

## Detailed Ontoglyph Example

This is an ontoglyph presented in detail describing the protein CRK:



This is a more complex example for the protein CD28:



Since glyphs are derived from GO terms rather than proteins. They can illustrate any protein's known biological attributes. Glyphs can be combined and integrated "on-the-fly" into broader illustrations of molecular interactions, complexes and pathways.





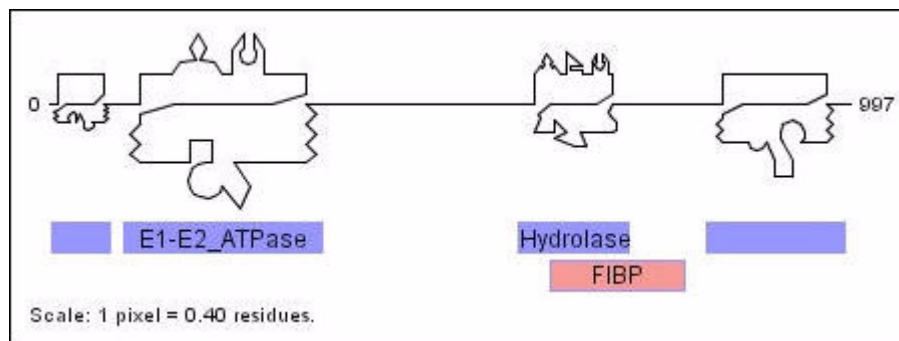
## Appendix B

### About Proteoglyphs

---

Proteoglyphs are graphical symbols used to represent the structural and binding properties of proteins at the level of conserved domains. The protein is diagrammed as a straight horizontal line, into which glyphs are inserted to represent conserved domains. Each glyph is inserted so that it represents the relative position and length of its alignment in the protein sequence.

An example proteoglyph (G:1081), showing four conserved domains:



The image above contains the proteoglyph, as well as other information about the protein sequence and the conserved domains it contains. For example, the sequence length is 997 (residues), and each pixel is 0.4 residues wide.

The proteoglyph is composed of the top-scoring conserved domains that fit on a horizontal line without overlapping. These domains are represented in the proteoglyph with their respective glyphs, and also with blue rectangles shown below the proteoglyph, in which the common label for the domain is given (if it fits within the rectangle). The red rectangles show conserved domains that also align with the sequence, though the proteoglyph does not contain their respective glyphs because they overlap other domains that have a better E-value.

2-75:pfam00690:Cation\_ATPase\_N:3.6E-18

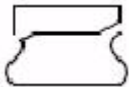


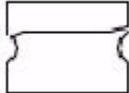

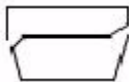
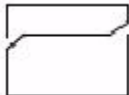
In the above image, when the mouse is positioned over a domain's glyph or rectangle, a tooltip should appear with information about the domain.

The text in the tooltip contains the following information:

- 2-75 indicates the alignment starts at residue 2 and ends at residue 75
- pfam00690 is a database-specific identifier for the domain (in effect, PFAM)
- Cation\_ATPase\_N is a commonly used label for the domain
- 3.6E-18 is the E-Value of the alignment

Clicking on a domain glyph or rectangle will open a new browser window with the Domain Info page, showing more information about the domain. However please note that this linking from various image regions and tool tip functionality is supported only when the jpg format is used.

A domain glyph is constructed based on the (known) structural and binding characteristics of the domain. The structural characteristics of a domain are given in the form of a SCOP class. The class of the domain determines the geometry of the base glyph used to construct its glyph. The following table lists the classes and shows the respective base glyphs:

All Alpha	
All Beta	
Alpha+Beta	
Alpha/Beta	
Membrane	
Small Domains	
Structure Unknown	

The base glyphs were designed to make the structural class easy to discern. The geometry for the Membrane class is squiggly, resembling how the membrane is usually drawn, the geometry for the All Alpha class uses curved segments, resembling the curving alpha-helices, and the geometry for the All Beta class uses straight segments, resembling the straight beta-sheets. Classes combining Alpha and Beta properties use both straight and curved elements: in the Alpha+Beta class are domains with segregated alpha and beta secondary structures, and hence straight and curved segments are on opposite sides of the glyph, and in the Alpha/Beta class are domains with interspersed alpha and beta structures, and hence both straight and curved segments appear on both sides. Small domains have a narrower bottom segment in their glyphs.

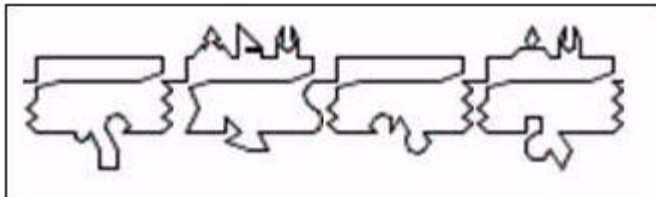
The binding characteristics (or specificity) of a domain are illustrated using Binding glyphs. The Binding categories for each domain are computed from GO annotations (for example, from the Interpro database), in the same way as for Ontoglyphs. The Binding glyphs are placed along the top horizontal segment of the base glyph. Clicking on a domain will open a new browser window, showing more information about the domain, including the binding glyphs it contains and the binding categories they represent.

Lastly, each domain glyph contains a unique geometry, which is placed along the bottom horizontal segment of the base glyph. Each conserved domain in our database (of about 8000 domains), is assigned a visually distinct geometry. This geometry has two roles. Firstly, it distinguishes different conserved domains in a proteoglyph. For example, even if the domains have the same binding characteristics and structural class, this unique part allows us to visually distinguish them as different domains. For example, in the proteoglyph above, the first and last domains are identical except for the unique geometry, which tells us they are different domains. Secondly, we think that becoming familiar with particular domains, some users may come to distinguish them based on this unique geometry.

## What is a Proteoglyph Summary?

A proteoglyph summary, used on the GI Information page as well as in the BIND interface, shows the domain glyphs that appear in a proteoglyph, however they are shown as independent icons. Every domain is shown only once, with a count of how many times it appears in the Proteoglyph if it appears more than once (for example, x8 to the right of the glyph signifies the domain appears 8 times). In this representation the domain glyphs are not scaled to the relative size of the alignment length of the domain in the sequence. Each individual domain glyph image in the summary view is linked to the Domain Information page.

Here is the Proteoglyph Summary for GI 1081:





## Appendix C

### Frequently Asked Questions

---

This chaptersection contains answers to frequently asked questions (FAQs) about BONDplus, BINDplus, and GENESEQ and other topics.

#### BONDplus FAQ

This section answers frequently asked questions about BONDplus.

##### **What is BONDplus?**

BONDplus (Biomolecular Object Network Databank Plus) is a life science data platform that combines public and proprietary sequence, interaction and related interactome information to deliver powerful, integrated search results across vast pools of life science data.

BONDplus serves the interests of the developing global interactome effort by encompassing the genomic, proteomic and metabolomic research communities on one comprehensive platform.

Through the integration of GENESEQ™ patented sequence data, BONDplus also provides valuable insight for IP professionals performing prior art or infringement searches.

##### **What kind of search methods are used on BONDplus?**

All databases are fully searchable via free text, identifier, or BLAST search. The interface supports full text queries and even field specific queries, allowing users to search across all data in BONDplus in a matter of seconds.

##### **How can I refine my search?**

BONDplus provides a full featured query language and powerful filters for users to refine their search parameters and narrow down their results.

##### **How do I view my search results?**

Every search result can be viewed using multiple layouts, and the result set can be exported in a variety of different common file formats, including PDF, XML, FASTA, and flat-file format.

##### **Does BONDplus have BLAST capability?**

Yes, BONDplus includes a comprehensive BLAST service.

**Does BONDplus contain BINDplus?**

Yes. BINDplus is continuously updated on BONDplus, so you always have access to the highest quality interaction information available.

**How do I incorporate other data sources not already included in BONDplus?**

Thomson Scientific encourages customers to work with us to develop customization and integration options specific to your research initiatives. Customization options can range from supporting and integrating an in-house database to the development and provision of additional external data sources, or custom user interface modifications. If you would like to discuss customization options with one of our specialists, please contact your regional Thomson Scientific office.

**Is training available?**

Yes, training is available. We can also design training programs that specifically meet your organization's requirements. If you would like to discuss training with one of our specialists, please contact your regional Thomson Scientific office.

**What data sources does BONDplus contain and how often is the data updated?**

BONDplus contains public and proprietary sequence, interaction, taxonomy, publication, annotation, domain, and cross-reference data from 15 assembled information sources, including:

- Over 80 million public domain sequences
- GENESEQ, containing over 9 million patented sequences
- BINDplus (Biomolecular Interaction Network Database Plus), the largest, most comprehensive interaction database available containing over 200,000 peer-reviewed interactions

BONDplus data is updated continuously from the originating host data organization. The update schedule is detailed below:

Name	Description	Organization	Frequency
Patented Sequences			
GENESEQ	Database with the highest quality collection of annotated genetic sequences from worldwide patents	Thomson Scientific	Daily
Interactions			
BINDplus	Expanded suite of interaction data, based on the benchmark setting Biomolecular Interaction Network Database (BIND).	Thomson Scientific	Daily
Interactions			
RefSeq	All Reference sequences made available by NCBI.	NCBI	Daily
GenBank	All protein, DNA and RNA sequences made available by NCBI.	NCBI	Daily
Swiss-Prot	All protein sequences made available by EBI.	ExPASy	Daily

Name	Description	Organization	Frequency
Complete Genomes	All complete genome information and annotation made available by NCBI.	NCBI	Weekly
Redundant Groups	Sequence redundancy annotation and ranking as determined by NCBI.	NCBI	Daily
Genes			
Entrez Gene (Locus Link)	All gene annotation and sequence associations made available by NCBI.	NCBI	Monthly
Homologene	Eukaryotic gene homologs made available by NCBI.	NCBI	Daily
Structure			
Protein Data Bank	All non-theoretical PDB structures made available through NCBI's MMDB.	NCBI (RCSB)	Daily
Taxonomy			
Sequence Taxonomy Annotation	Taxonomic information mapped to protein, DNA, and RNA sequences.	NCBI	Daily
Taxonomy Tree	Taxonomic information such as the taxonomy nodes, names, division, genetic codes, deleted nodes and merged nodes.	NCBI	Daily
Literature			
PubMed Reference	Association between the medline/pubmed article(s) found in a sequence record indicating who first published the sequence.	NCBI	Daily
Annotation			
GO Sequence Annotation	Association of sequences to controlled vocabularies for the description of the molecular function, biological process and cellular component of gene products.	Multiple (compiled from multiple sources)	Monthly
GO Tree	The entire Gene Ontology tree as provided by Gene Ontology Consortium.	GO Consortium	Monthly
Cross References			
DbXRef	Sequence identifier cross reference for 18 major sequence databases, including model organism databases.	Multiple (compiled from multiple sources)	Monthly
Domains			
Pfam	Pfam domain information as obtained from NCBI's CDD database.	NCBI	(Sanger Institute)
SMART	SMART domain information as obtained from NCBI's CDD database.	NCBI (EMBL)	Daily
COG	COG domain information as obtained from NCBI's CDD database.	NCBI	Daily



Name	Description	Organization	Frequency
Visualization			
Ontoglyphs	Pictorial representations of gene ontology information providing a visual approach for automated retrieval and representation of gene product annotation from various biological datasets.	Thomson Scientific	Monthly
Pre-computed Data Sets			
BLASTP	BLAST results for all protein sequences available in NCBI's NR database.	Thomson Scientific	Daily
RPS-BLAST	RPS-BLAST results for all protein sequences available in NCBI's NR database.	Thomson Scientific	Daily

## BINDplus FAQ

This section answers frequently asked questions about BINDplus.

### What is BINDplus?

BINDplus (Biomolecular Interaction Network Database Plus) is a proprietary and growing collection of over 200,000 records documenting molecular interactions observed in peer-reviewed, published literature. The contents of BINDplus include high-throughput data submissions and hand-curated information. BINDplus is the commercial version of the public BIND database, with BINDplus containing up-to-date interaction information from 120 peer-reviewed publications.

### What happened to BIND?

BIND is still publicly available from Thomson Scientific. All data in BIND was generated from public funding which remains available as a component database of BOND, our open access data warehouse. Currently BIND is a static resource of 200,000 interactions and complexes. All the great features familiar to you in BIND have been improved in BOND, with the added benefit that these features are easier to use and deliver more insightful results.

### Why the mix of open access and commercial products?

Thomson Scientific takes seriously its responsibility to provide resources to assist biodiversity in the global life science community, which includes offering open access and commercial products to the bench biologist. The product suite is offered as follows:

	Open Access	Commercial
Data Warehouse	BOND	BONDplus
Interaction Database	BIND	BINDplus

**What is an interaction?**

BINDplus is an interaction database with two classifications for molecular associations: molecules that associate with each other to form interactions and molecular complexes that are formed from one or more interactions. Each interaction record may contain information on experimental systems demonstrating the interaction, cellular localization, chemical states and actions, and kinetic parameters.

**How is BINDplus different from other interaction databases?**

BINDplus content is the most comprehensive interaction resource available to life science researchers. With more than twice the interaction data of the nearest public access interaction database, BINDplus is the definitive resource for biomolecular interaction information. Unlike other databases, BIND offers data derived only from observed interactions published in peer-reviewed journals.

Other interaction databases focus on species-specific information or particular types of interactions, eg. protein-protein only. BINDplus contains information on over 1,500 species representing over 60,000 unique Gene Identifiers (GIs) and captures information on all biomolecules such as protein, DNA, RNA, small molecules and complexes.

Some commercial interaction databases use automated techniques to extract limited interaction information (“facts”) from a given publication, missing valuable information in figures and supporting documentation. Automated techniques also extract incorrect or misleading information in the race to bulk up the database with content. This unreliable data creates more work and delays further your discovery research efforts. The quality data found in BINDplus can aid successful Biodiscovery.

**Are BINDplus curation standards equivalent to BIND standards?**

Yes. BINDplus is curated to the same rigorous standards as BIND. BINDplus curators are wet-lab scientists trained to a postgraduate level and are trained to read, extract and enter interaction information from peer-reviewed publications in a controlled and consistent format. Curation standards are supported by comprehensive standard operating procedures, case studies, an established submission system, and a BIND Curation Reference Manual.

**What is the difference between BINDplus and BONDplus?**

BONDplus (Biomolecular Object Network Databank Plus) is a life science data platform that combines sequence, interaction and related interactome information to deliver powerful, integrated search results. BINDplus is a component database delivered via the BONDplus platform.

**What is the difference between BINDplus flat-file and BINDplus on BONDplus?**

BINDplus on BONDplus includes the most recent data with an interactive interface, Lucene query search engine, record export options, and links out to other databases. The BINDplus flat-file contains the most recent interaction data available for internal customer use.

**In what formats are the BINDplus flat-file available?**

The BINDplus flat-file is available as a set of multiple computable formats consisting of XML, ASN.1, PSI level II, and tab-delimited.

**How do I incorporate other data sources in my BINDplus flat-file and BONDplus product?**

Thomson Scientific encourages customers to explore the boundaries with data. We can develop customization and integration options to drive your research initiatives. Customization options can range from supporting an in-house database to the development and provision of additional external data sources, or custom user interface modifications.

**How often is BINDplus flat-file content updated?**

The BINDplus flat-file content is updated on a daily basis. Daily update files are released nightly enabling customers to retrieve the very latest data.

**How often is BINDplus content on BONDplus updated?**

BINDplus on BONDplus is updated every 15 minutes.

## **GENESEQ FAQ**

This section answers frequently asked questions about GENESEQ.

**What does GENESEQ™ contain and when did coverage start?**

GENESEQ contains biological sequence information taken from patent applications and granted basic patents from 1981 onwards. All nucleic acids which are 10 or more bases in length, amino acids which are four or more residues in length, plus all PCR primers and probes of any length are included in the database.

**What are the differences between GENESEQ and other sequence databases?**

GENESEQ covers only sequences that appear in patents. GENESEQ records cover both the biological relevance and the IP aspect of the sequence; other databases tend to concentrate just on the biological relevance.

**How does GENESEQ content differ from other patented sequence databases?**

GENESEQ emphasizes complete coverage of patented sequences. GENESEQ covers content from 41 patent authorities (including WO, US, EP, JP, DE, IN and CN). Foreign patents compose over 20% of the database, with English language titles provided for each patent. GENESEQ editorial staff manually review each patent to extract all referenced sequences contained in the patent, going beyond those the patentee and patent offices provide in sequence listings. As a result, the comprehensive nature of GENESEQ ensures that your analysis will never miss a sequence.

**What enhancements have been made to GENESEQ in 2007?**

GENESEQ records are now annotated with links to NCBI, providing enhanced biological context for a subset of GENESEQ records. This includes links to identical records from NCBI, SWISSPROT, BINDplus (Biomolecular Interaction Network Database Plus), Gene Ontology (GO) annotation, and sequence aliases.

**How are the new links to NCBI and SWISSPROT computed?**

GENESEQ records are compared with the canonical public biological sequence databases provided by the National Center for Biotechnology Information. Where 100% identity is found among a set of redundant Gene Identifiers (GIs), the most biologically-relevant GI is chosen.

Sequence aliases, GO Annotation, and database cross-references to BINDplus (Biomolecular Interaction Network Database Plus) and SWISSPROT are all computed based on the redundant GI match.

#### **How does GENESEQ on BONDplus differ from GENESEQ flat file?**

The content of the GENESEQ database is identical on all platforms and with all the various software packages. BONDplus (Biomolecular Object Network Databank Plus) offers additional biological context through integration of GENESEQ records with publicly available sequence, taxonomy, and interaction information. GENESEQ on BONDplus improves the searchability of patented sequence records through comprehensive indexing, advanced text search algorithms, BLAST searching and pre-computed homology analysis.

#### **How does GENESEQ on STN (DGENE) differ from GENESEQ flat file?**

The content of the GENESEQ database is identical on all platforms and with all the various software packages. The only differentiating characteristics are the software packages used to search the GENESEQ database. However, some of the formats are different. In particular, sequence Accession numbers in DGENE are preceded with YYYYN for nucleotides, and YYYYYP for polypeptides/peptides, where YYYY is the year of the WPIL accession number for the patent.

#### **How do I search for D-amino acid residues?**

D-amino acid residues are represented as the parent amino acid in the sequence and the FeatureTable will show the position of any D-modified amino acids (annotated by “Misc-difference”).

#### **What are the selection rules for GENESEQ?**

All sequences appearing in the patents claims section are included in GENESEQ. All other sequences not specifically stated to be known or to have been published are included. Specific examples of generically claimed peptides are included.

**NOTE:** GENESEQ covers only basic patents. A basic patent is one in which Thomson Scientific has seen the details of this invention published for the first time.

#### **How do I search cyclic peptides in GENESEQ?**

All cyclic peptides have the word “cyclic” appearing in the Keyword field. Cyclic peptides can only be searched by carrying out a series of searches with each amino acid residue acting as the first residue in the query sequence.

#### **Does GENESEQ cover single- or double-stranded DNA?**

GENESEQ covers both single-stranded and double-stranded DNA. The two types are distinguished by the terms “ss” and “ds” in the keyword field. However, in the case of double-stranded DNA, only the 5’ to 3’ strand is input. This follows the USPTO requirements for all nucleotide sequences in the sequence listing section, which have to be presented on diskette, as single stranded, 5’ to 3’, left to right.

## Browser FAQ

### **Internet Explorer is blocking my downloads. How do I correct this?**

Open Internet Explorer's Help file. In the Search field, enter "downloading files." A number of related topics will appear. Browse through the topics and follow the directions for file downloads.

# Index

---

## A

Abstract Syntax Notation One (ASN.1) ASCII  
format 50  
accession 103  
advanced field specific search, performing a 19  
Apache Lucene Query Language 19, 28  
ASN.1 50

## B

base glyphs 138  
Basic Local Alignment Search Tool (BLAST) 24  
BIND 55  
binding 130  
BINDplus 4  
defined 144  
FAQ 141  
flat-file 4  
objects 4  
records 4  
Biomolecular Interaction Network Database  
(BIND) 55  
Biomolecular Network Databank plus  
(BONDplus) 3  
BLAST 24  
a sequence 47  
enhanced results 27  
search, performing a 24  
BONDplus  
defined 141  
FAQ 141  
log in to 7  
to log out of 8  
boolean  
operators 10  
queries 29  
browse search, performing a 11  
browsers 7, 8  
FAQ 141

## C

categories  
ontoglyph 113  
category, identifier 16  
common use cases 6  
complexes

export formats 94  
exporting search data 93  
filter 56  
molecular 4, 55  
single record view 86  
summary record view 79  
contacting support 1  
cross-database queries 9  
customer support 1  
hours 1  
Cytoscape 68, 86

## D

domain glyph 138  
Domain Summary 81

## E

enhanced BLAST 27  
export  
formats 53  
export formats  
complexes 94  
GENESEQ 111  
interactions 78  
exporting  
complexes search data 93  
GENESEQ search data 110  
interactions search data 77  
sequence data 53  
Extensible Markup Language (XML) 51

## F

FAQ  
BINDplus 141  
BONDplus 141  
browsers 141  
GENESEQ 141  
FASTA  
format 48  
sequence 48  
field-specific query builder 19  
filter  
complexes 56  
GENESEQ 98  
interactions 56

- ontoglyph 59
- proteoglyph 59
- sequences 38
- Firefox 7
- flat-file, BINDplus 4
- format
  - Abstract Syntax Notation One (ASN.1)
    - ASCII 50
  - FASTA 48
- frequently asked questions 141
- function 114

## G

- GenBank 3, 37, 49
- gene ontology 113
- GENESEQ 3, 5, 97
  - export formats 111
  - exporting search data 110
  - FAQ 141
  - filters 98
  - performing a query for 97
  - single record view 102
  - summary record view 99
- Genome Ontology descriptions 3
- GO
  - Summary 81
  - terms 113

## H

- Home Page, searching from the 9
- hours, customer support 1

## I

- identifier
  - category 16
  - number 16
  - search, performing a 15
  - taxon 3
- interactions 4, 55
  - export formats 78
  - exporting search data 77
  - filter 56
  - performing a query for 55
  - single record view 69
  - summary record view 60
- interface, using the BONDplus 7
- Internet Explorer 7

## L

- LocusLink 3
- log in to BONDplus 7
- log out of BONDplus 8

## M

- MedLine abstracts 3
- Microsoft Internet Explorer 7
- modifiers, wild card 11
- molecular complexes 4, 55
- Mozilla Firefox 7
- MyBOND 9

## N

- National Center for Biotechnology Information (NCBI) 37

## O

- objects, BINDplus 4
- ontoglyph 41, 113
  - categories 113
  - detailed 134
  - filter 59
- operators, boolean 10

## P

- password 8
- pathways 55
- performing
  - a BLAST search 24
  - a browse search 11
  - a query for GENESEQ 97
  - a query for interactions 55
  - a query for sequences 37
  - a quick text search 10
  - a specific search 10
  - a text search 13
  - an advanced field specific search 19
  - an identifier search 15
- phrase queries 29
- PICKS 58
- problem, reporting a 1
- Protein Interaction Confidence Kernel Scores (PICKS) 58
- proteoglyph 42, 137
  - filter 59
  - summary 139
- PSI Level 2 76



**Q**

## queries

- boolean 29
- cross-database 9
- phrase 29
- simple 29
- wild card 29

## query builder 19

## quick text search 9

- performing a 10

**R**

## records

- BINDplus 4
- per page 12

## RefSeq 37

## reporting a problem 1

## results

- tab 12
- viewing search 30

## retrieve previously saved search results 32

**S**

## saving search results 31

## search

- quick text 9
- results
  - retrieve previously saved 32
  - results, saving 31

## searching from the Home Page 9

## SeqHound 3

## sequence 3

- BLAST a 47
- filters 38
- query, performing a 37
- summary record view 40

## simple queries 29

## single record view 30

- complexes 86
- GENESEQ 102
- interactions 69

## special characters 28

## specific search, performing a 10

## sub-cellular localization 124

## summary

- proteoglyph 139
- record view 30
  - complexes 79
  - GENESEQ 99

## interactions 60

## sequences 40

## tab 12

## support

- contacting 1
- customer 1

**T**

## taxon identifiers 3

## text search, performing a 13

**U**

## UniProt 37

## use cases, common 6

**V**

## viewing search results 30

## visualize results 68

**W**

## wild card

- modifiers 11
- queries 29

**X**

## XML 51

