GoView and GoGet User Manual

Macalester College GO Exploration Tool Project

goproject@macalester.edu

Professor Libby Shoop, Mathematics and Computer Science Department

Contents

1.0 Introduction

PLEASE NOTE: Some of the examples in this manual were made from the May, 2004 version of the GO database. When you try the tools with a later version of the database, things will look different, because some of the term names and the structure of the ontologies will have changed. The general way that the tools work has not changed.

GoGet and *GoView* are two new tools built as part of an extensible web application system based on Java 2 Enterprise Edition technology. GoGet has a user interface that enables users to ask biologically interesting questions such as: 1) "What are the DNA binding proteins involved in DNA repair, but not in DNA replication?" and 2)"`Of the terms containing the word triphosphatase, which have associated gene products from mouse, but not fruit fly?" Results of such queries can be viewed in a collapsed tabular format that eases the burden of getting through large tables of data. GoView enables users to explore the large directed acyclic graph structure of the ontologies in the GO database. The two tools are coordinated, so that results from queries in GoGet can be visualized in GoView in the ontology in which they appear, and explorations started from GoView can request details of gene product associations to appear in a result table in GoGet.

The goal of this project was to provide powerful new ways for biologists and biology students to explore the Gene Ontology (GO) data that is being developed by the GO consortium. The GO Consortium was formed to build a common annotation to describe genes and their products in any organism (Ashburner00,GO2001,Ashburner02). As stated in (GO2001), one of the consortium's goals was to create ``a shared language of biology'' by compiling ``a comprehensive structured vocabulary of terms describing different elements of molecular biology that are shared among life forms.'' The structured vocabularies are organized as ontologies, which are a diagrammatic and notational tool for describing knowledge about a domain (Gruber93,Gruber95).

The consortium has built and continues to revise three ontologies: molecular function, biological process, and cellular component. These are defined in (GO2001):

"Briefly, molecular function describes what a gene product does at the biochemical level. Biological process describes a broad biological objective. Cellular component describes the location of a gene product, within cellular structures and within macromolecular complexes."

In addition to developing the ontologies, the consortium members are assigning gene products from 26 model organism and sequence databases to one or more of the terms in each ontology. The assignment is based on a particular kind of evidence, including inference by direct assay, by sequence similarity, or by traceable author statement, to name a few. The gene products can be associated to more than one term in all three ontologies.

More information about the Gene Ontology Databases is available at the Gene Ontology Consortium's website: www.geneontology.org

1 .1 GoView

GoView is used to illustrate the Gene Ontology Database graphically though graphs of different subsections of the three different ontologies. It is navigated by selecting parts of the graph and observing the structure of portions of the ontologies. The terms found in GoView can then be directly accessed in GoGet for more information about their gene products.

1.2 GoGet

GoGet searches for terms and gene products using a number of search criteria. It then returns the Term Name, Definition, Ontology, Term Synonyms, Gene Products' Information, and # of gene products for the term(s) found in the search. It also gives links to the original databases' web page for more information about specific gene products.

2.0 Getting the Programs

To get the GoView and GoGet software go to: http://db.math.macalester.edu/goproject

To use GoGet click on the link on the left labeled "*Search the Gene Ontology with GoGet*".

To get GoView, click on the link labeled "*Explore the Gene Ontology with GoView*". This will download an executable Java application, which should start up, if you have Java installed on your computer. So if you have a Macintosh, this will work without doing anything. If you have a Windows or linux machine and have not installed the Java Runtime Environment (JRE), we have a link for you to use to install it. It is a relatively easy procedure.

3.0 Tutorial

In this section of the manual, we will walk through several examples of how to use GoGet and GoView. We will start with the querying capabilities of GoGet, and then show the ways to explore the ontologies using GoView. Lastly, we will show how the two tools can be used together.

3.1 Example 1: Simple Searches in GoGet

From http://db.math.macalester.edu, click the 'Search the Gene Ontology with GoGet' link, and GoGet will come up in its own window. We make a lot of use of cascading style sheets in our HTML, some of which is not always displayed properly on all browsers. We have had the best luck with Mozilla and Mozilla Firefox on all platforms, and fairly good luck with Safari on Mac OS X. We are still working on how to improve the layout and functionality under Internet Explorer for Windows, so if this is your only browser option, some things may not look just like the figures in this document.

As an initial introduction to the GoGet searching tool, we will show you a relatively simple example and describe the features of the interface. The tool initially looks like Figure 1.
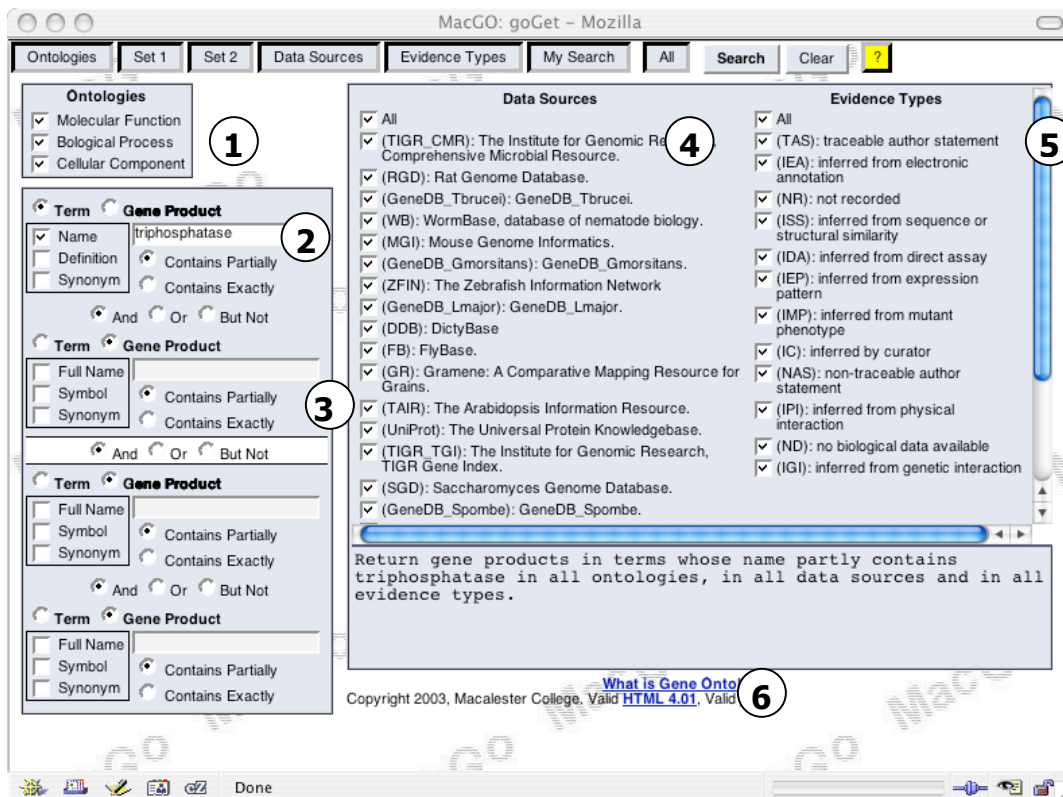
Figure 1.   Initial State of the GoGet search tool when you first access it

The tool has several different sections where you can indicate what you would like to search for.

Ontology Choice The box indicated by 1 in Figure 1 let's you choose one of the three GO ontologies. The GO database has defined *terms* in three different ontologies. An ontology is a classification system for knowledge about a domain. In this case, the GO consortium members have developed three different ways to classify genes, which are defined in (GO2001) as follows:

> "Briefly, molecular function describes what a gene product does at the biochemical level. Biological process describes a broad biological objective. Cellular component describes the location of a gene product, within cellular structures and within macromolecular complexes."

Terminology Aside

Each ontology is a 'graph' structure with a 'root' that is the starting point. From the root *term*, each term below it is a subcategory. For example, some of the subcategories under 'molecular function' are antioxidant activity, binding, catalytic activity, and transporter activity. Each of the subcategories may be broken down into additional subcategories. In graph terminology, the subcategories of a term are called the *child terms* and the category above the child is called the *parent*. The GO ontologies are graphs because any child could have more than one parent. The paths through the graph are *directed*, because when you explore the graph, you think about going from the top root and traveling down through each subcategory. The graphs are *acyclic*, because there is not a path from a child cycling back to any of its ancestors.

Set to Return   The box indicated by 2 in Figure 1 enables you to choose which set of *terms* or *associated gene products* you would like to have returned. The terms may have gene products from various species *associated* with them. For example, the molecular function term called 'nucleoside-triphosphatase activity' has one gene product from the Mouse Genome Informatics database associated with it whose full name in that database is 'nucleotriphosphatase' and whose gene symbol is 'Ntp'. In GoGet, we refer to the box indicated by 2 in Figure 1 as the Set 1 box. If you choose Term, as shown in Figure 1, then what will be returned to you is GO terms, along with information about which gene products are associated with those terms. Exactly which terms are returned depends on whether you choose 'Contains partially' or 'Contains exactly'. The example startup query shown will return any terms that contain the phrase 'triphosphatase' anywhere in it, either in a larger word or as its own word. The 'Contains exactly' choice will usually return a smaller set of terms where the exact phrase matches, and the very next character is a space to break to the next word.

You also have the choice of what information about a term you want to search: its name, its definition, or any of its synonyms. Typically, it is best to start with its name, and use the others if you are unsuccessful.

Important Searching Tips

When choosing to return a set of terms, you need to carefully consider choosing to search through name, definition, and synonyms all at once by selecting all three. The definitions can be long and contain many words, so they might get you off track. For example, the phrase 'catalyzes the reaction' appears many, many times in various term definitions, so using 'catalyze' in a search where you chose definition would return too many results. The synonyms, however, might be very useful for terms that might likely have other common names that you are familiar with. For example, the term called '8-oxo-7,8-dihydroguanine triphosphatase activity' has a synonym '8-oxo-GTPase'.

Beware: Sometimes the text index on the names does not work as you might expect. For the preceding example, choosing to search contains partially for '8-oxo' should return the '8-oxo-7,8-dihydroguanine triphosphatase activity', but it currently does not. We are working on making this work properly by switching to a new version of the database (Oracle 9i will replace Oracle 8i in our implementation eventually, and the text indices will improve).

As an alternative to searching for terms, you can also to search for particular gene products. This will return the gene products that match your text search and the terms that they are associated with. If you choose one or two of the ontologies, fewer results will be returned, because gene products associated with terms that are in an ontology that you left out will be removed from the results. You may need to do this kind of filtering if you are searching using words that might be quite prevalent in the names of gene products. This happens because there are only so many results that can be displayed in a web page without breaking the browser, so we have to check for the number of results coming back and refuse to display them if there are too many.

You can combine a term and gene product search together—we will show examples of this later. First, you can get a sense for what is returned from a search by simply using the defaults that appear when the tool comes up. In Figure 1, notice the area of the tool indicated with a '6'. This is an English version of the query that you are formulating. This is there to help you envision what will be returned using your choices. Read it and make sure that it seems to make sense to you (we will explain data sources and evidence types in a while.) Click on the button labeled 'Search' near the upper right corner. (When you do subsequent searches, it also works to hit

return when your cursor is in the text box.) The result should look like Figure 2. Notice that this results page has the English version of your query shown at the top.

Figure 2.   Results from a search for terms partially containing 'triphosphatase'

Notice that in the second column from the left are the list of terms from that search. Each of these terms has a definition, which you can see by choosing the 'display' button in the Definition column. That button is a 'toggle', so that you can display and then hide the definition again. To use database vernacular, providing those display buttons are our way of letting you 'drill down' to see more details about each term.

Notice on the far right we give you an indication of how many gene products are associated with each term. If you choose to display those, you get quite a bit of extra information about each gene product, including a link back to its original data source. For example, try choosing to display the gene products' information for the third term down from your result, then choose to display the data source information for the first gene product. Notice that the table increases in size (width and height) as you 'drill down'. The result should look like Figure 3.

Figure 3. **Result of 'drilling down' to obtain more information about gene products associated to a term**

If you click the link under 'URL', you will get a new window to the original database for this gene product. Notice that FlyBase (the home of *Drosophila melanogaster* data) calls this gene NTPase, and other databases have named the gene product 'Nucleoside triphosphatase', and some don't even have a name yet. This is why the GO consortium got together— to try to place these gene products together in one category (in this case based on their molecular function).

Refining your search

There are some other choices that you can use to refine your search further. The buttons at the top of the results page enable you to bring back the boxes that you used when you formulated your search initially. Choose the 'All' button to bring back all of the search options, as shown in Figure 4. (Each of the buttons to the left of the 'All' button will pop up the individual filtering options and search 'set' options.) Notice that these are toggles, and if you click 'All' again the choices go away, and you can still see the results from your last query.
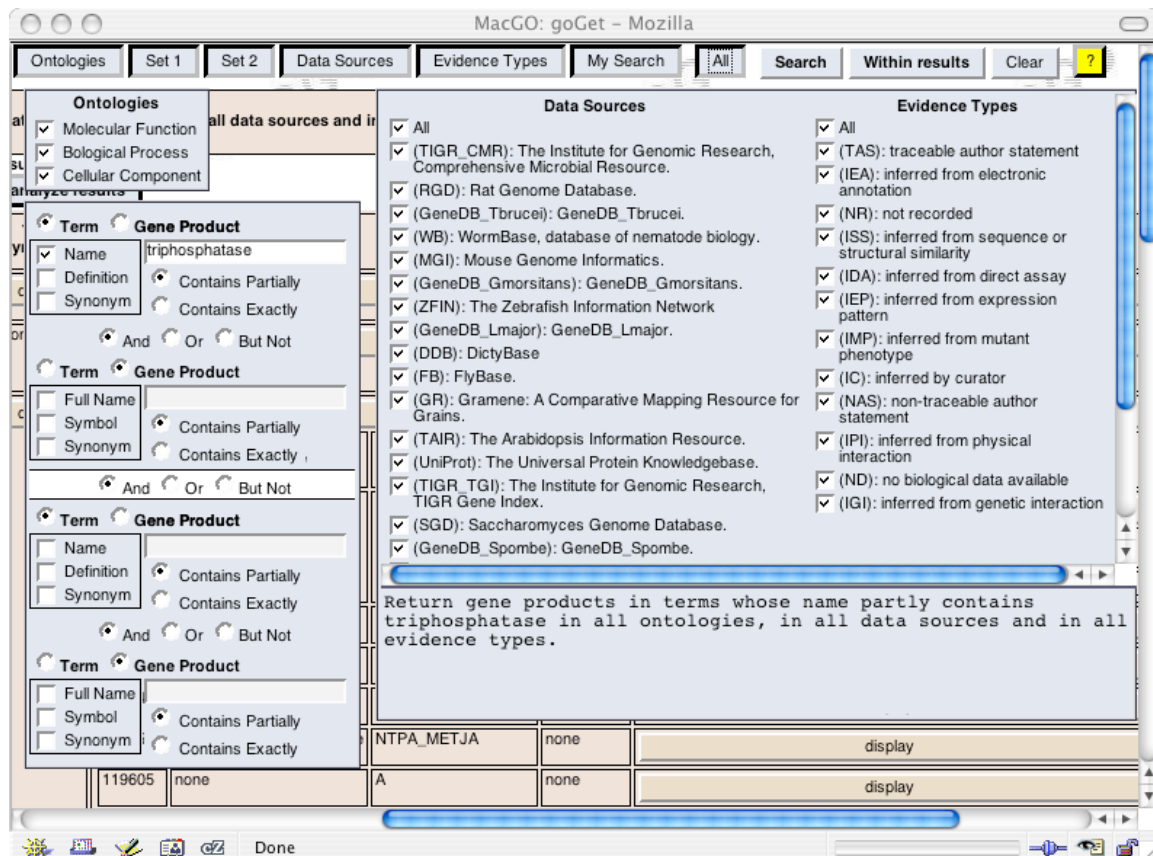
Figure 4.   Bringing the search criteria back up

You can refine your search to only show gene products from certain data sources. Click the 'All' checkbox under data sources, and all of the data sources become unchecked (this too is a toggle, so clicking it again will make them all checked). Suppose you want just the gene products from Flybase. Pick that data source, and choose 'Within results' (in this case, it also works to choose 'Search' again). This should return 1 gene product from 1 term (for the version of GO from May, 2004). If you want to go back and expand the search, you can add data sources, but then you should choose 'Search' instead. If you had chosen more data sources that produced more results, you could then refine your search by executing 'Within results' again.

You can also try some other was to narrow your search. For example, go back to 'All' choices, pick 'All' data sources, and then in the 'Set 1' box (labeled with a 1 in Figure 5), choose 'Full Name' under Gene Product and type in NTP in the box. It should look like Figure 5. Choose 'Search'. This will return fewer terms and fewer gene products, this time only those gene products that contain NTP in their full name. Notice this happens because the 'radio button' labeled 'And' is the default. This means that you want gene products with 'NTP' in their full name *and* are associated with terms that have 'triphosphatase' in their name. Choosing 'Or' or 'But Not' will have different results (be careful). For example, if you choose 'But Not" you will get all terms and their associated gene products that contain 'triphosphatase' in their term names, but not any gene products that have 'NTP' in their full names. Try this and see what happens. If you try 'Or' between the term and Gene Product search choices, you would get all terms with 'triphosphatase' in their names, and in addition all terms whose associated gene products happen to have 'NTP' in their name. This combination, using 'Or' might not always be what you're looking for. It is there mainly for when you might choose to search for terms with one phrase or another in it (you pick Term in each case). Please try this on your own.
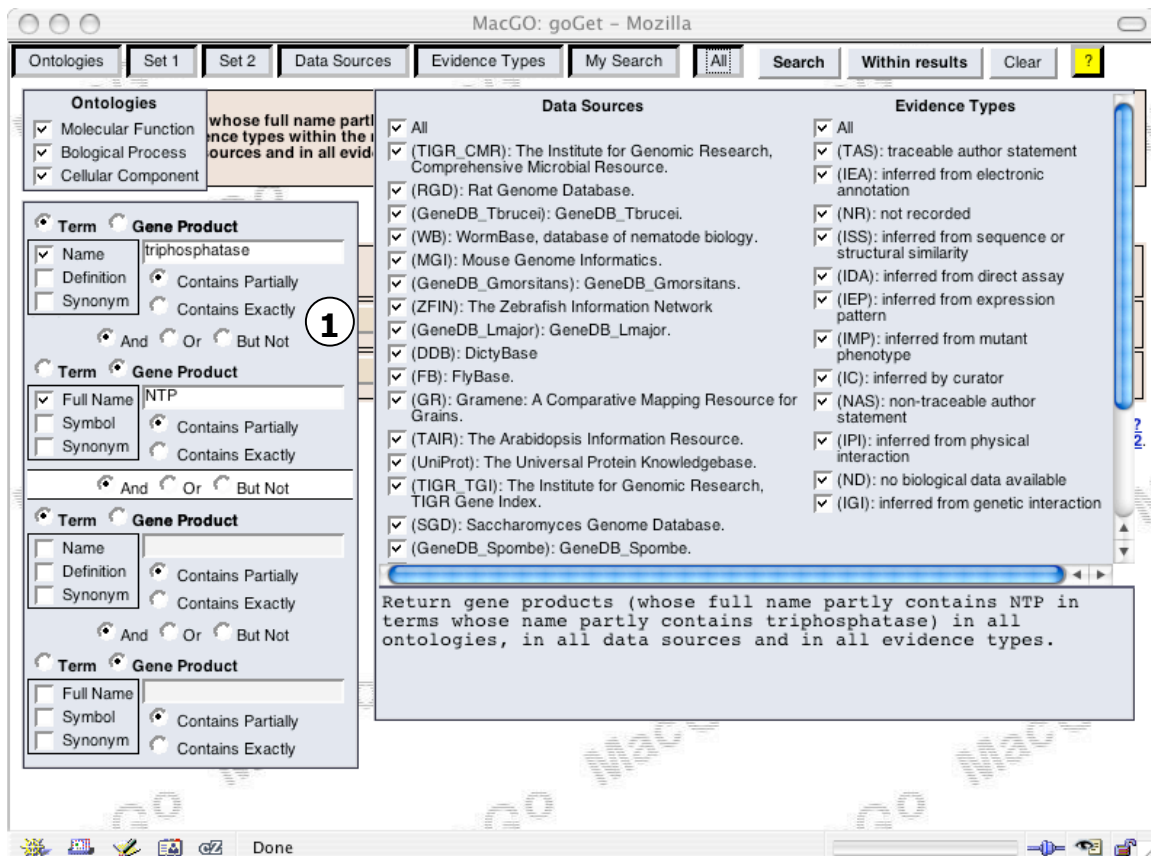
Figure 5.   A refined search to return certain gene products

Bug Warning

There are some cases when searching within results will return odd results. Most notably, if the result should return nothing, you might find that it tells you that the result is too large instead. We are working on repairing this anomaly.

3.2 Example 2: Another search using filtering

Click on the *All*[1] button on the top Menu. This will bring up all of the search options again.

Suppose we wish to find out what type of evidence there is that connects the gene product Pax6 to eye morphogenesis in mice.

We want to search all three ontologies so make sure that the boxes to the left under the bold heading **Ontologies** are checked.

Click on the Term radio box on the top of set 1.

---

[1] If you hadn't caught this, this button works like a "start over" button, or a 'refine' button to enable you to try again.

Next click on the "*Name"* box on the left. This will now let us type in our query.

In the text field 'eye morphogenesis.'

It is the default but check to see if the *and* button is selected under this text field if it not select it.

Click on the "**Gene Product**" radio box above the second text field and the check the

"**Symbol**" box.

In this next text field type 'Pax6.' In this case, Pax6 is the symbol for the gene, as opposed to its full name. Be sure the Full Name and Synonym boxes are unchecked. If you type in "Pax 6" with a space, you will get no results.

The search is not caps sensitive so typing 'pax6' will return the same results.

For now we will do a simple search so we can leave the lower search box blank.

We only want to find genes in mice so we need to select fewer databases. Click the *all* checkbox to unselect all the data sources and then click the (mgi) Mouse Genome Informatics box.

We want to search all evidence types because we are looking to see what type of evidence there is. If the "*All"* radio box under evidence is not checked click it.
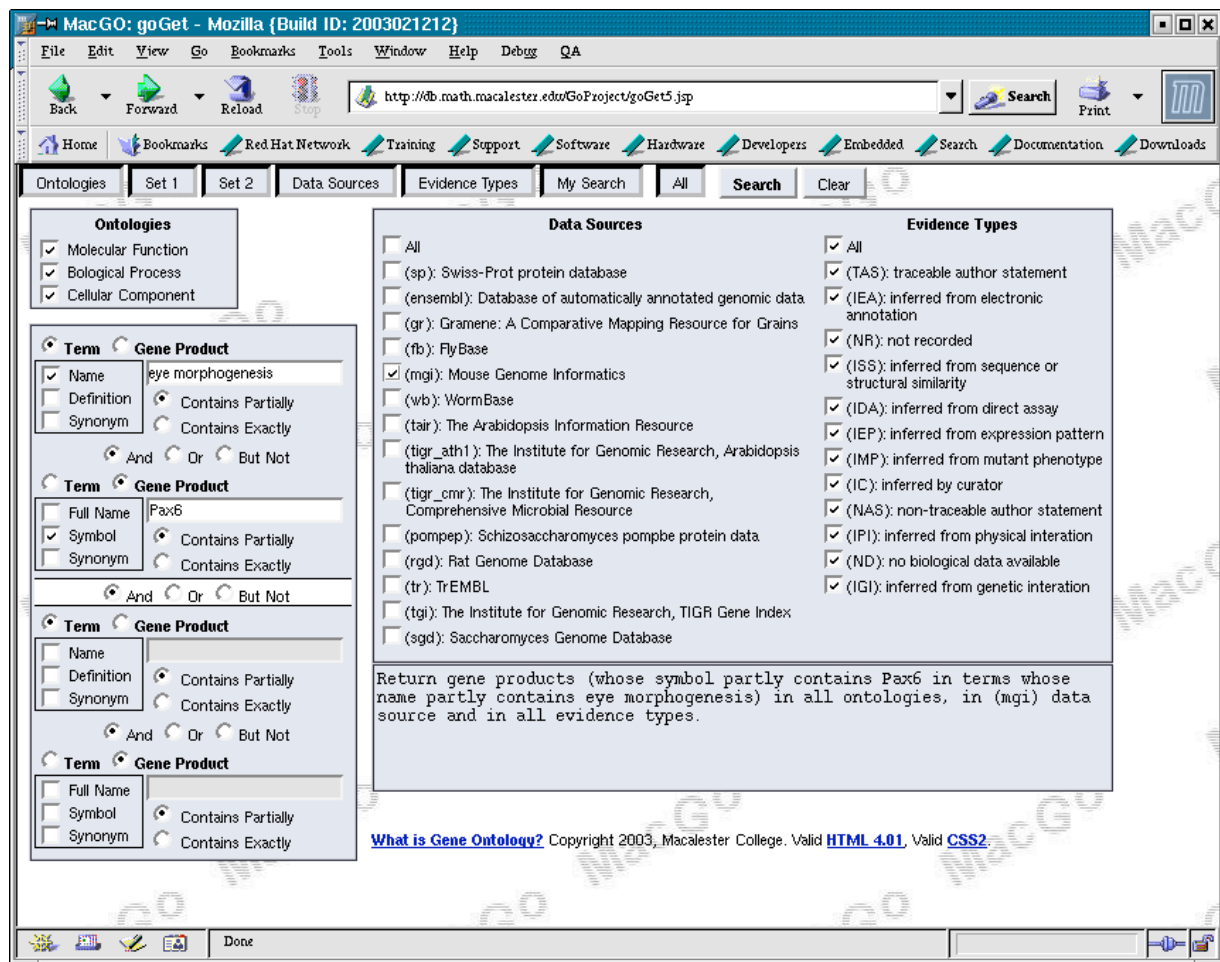
Figure 6.   Combining term and gene product search and data source filtering

To initiate the search click on the **Search** button on the top menu.

This should bring up the terms that have gene products with 'Pax6' in their symbol from the mouse genome database associated to them.

To find out what evidence type there is we will need to click on the display button under the gene products information and then again under evidence information. The result is in Figure 7, from which we can see that the connection is inferred by both mutant phenotype and genetic interaction evidence. To find out more details about each of the types of evidence, please see:

http://www.geneontology.org/GO.evidence.html

The evidence codes are another choice of filter that you have, and it works like the data sources.

Figure 7. **The types of evidence that tie pax6 to eye morphogenesis in mice.**

A Similar type of search: what can go wrong

Another question you might have would be along these lines:

What membrane gene products in C. elegans have 'unc' in their symbol?

A possible way to set up this query is shown in Figure 8. The problem that you will run into is the prevalence of the word 'membrane' in the definition and synonyms for terms in all of the ontologies. As it is shown, this query returns too many results. The way to find what you are really looking for is to only choose to look for membrane in the name of the term. Similarly, the letters 'unc' occur in far too many words in the full name and/or synonyms of gene products in the GO database, so you need to pick only the gene product symbol, as shown in Figure 8.
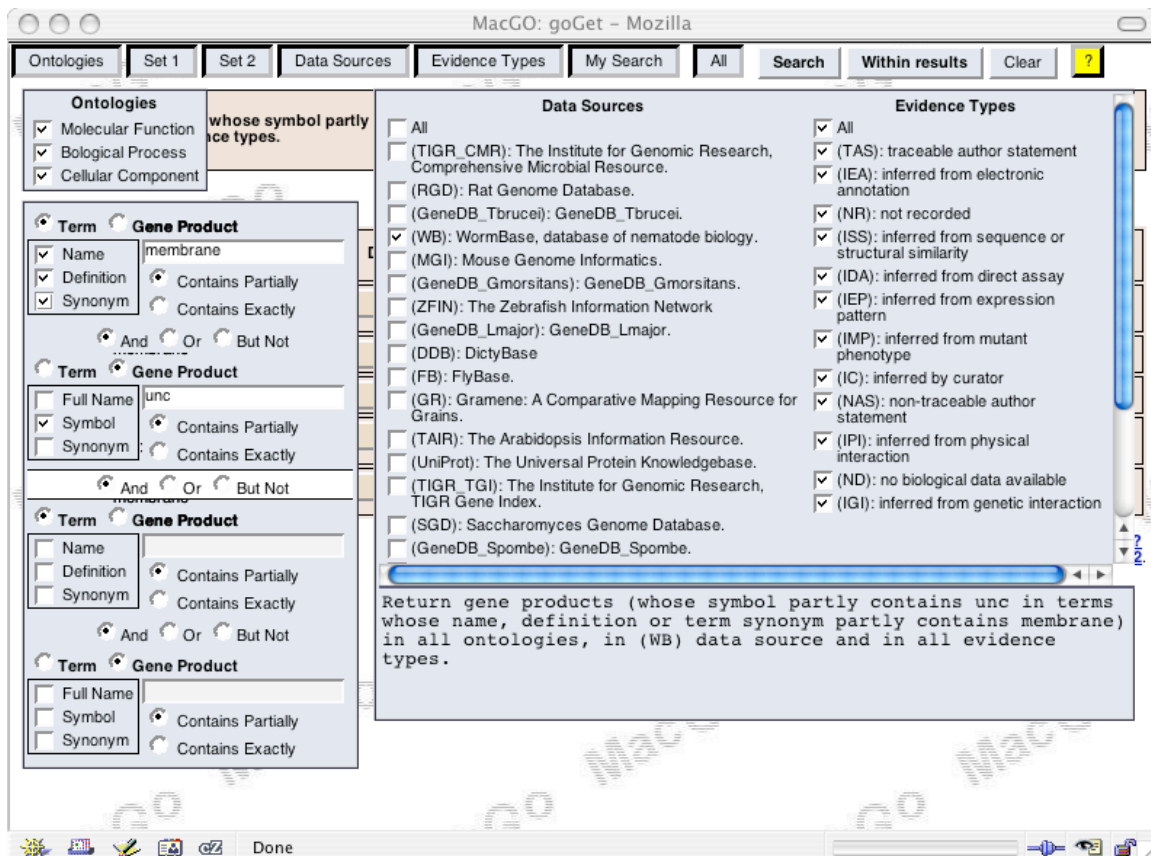
Figure 8. **A query that will return too many results (name, definition, and synonym should not all be picked)**

3.3 Example 3: A More Advanced Search in GoGet

Now let's try a more complicated search. In addition to using the 'Set 1' box that you have used in previous searches, we will also use the second 'Set 2' box below it. Notice that 'set 1' and 'set 2' have light blue backgrounds and a white background strip between them with a choice of 'And', 'Or', and 'But Not' between set 1 and set 2. A way to distinguish these is to toggle the 'Set 2' button on and off and see the set 2 box disappear and reappear.

Lets find "*DNA binding proteins that are involved in DNA repair but not involved in DNA replication*".

One way to think about doing this is to consider the two parts: 1) DNA binding proteins involved in repair, and 2) DNA binding proteins involved in replication. These two sets overlap, since some proteins (gene products) might have both repair and replication functions. Thus, subtracting the second set (2) from the first set (1) should give us what we are looking for.

Steps for this query:

Without closing the results window, click "All" and then "Clear". Then select the "Term" radio box on the top of Set 1. Check the *name* box and type in '*DNA binding*' in the text field.

Next make sure the "*And*" radio button is selected inside of set 1.

Select the "*Gene Product*" radio box on the bottom of set 1. Click the '*full name*' radio box and type out '*repair*' in the text field.

Next select the '*And*' radio box in the middle of set 2.

Then select '*Gene Product*' in the lower field and then select the '*full name*' checkbox. Type *replication* in to the text field.

Now select the '*But not*' radio box between the two sets. Your query should look like Figure 9.

Click **search.** Because this is a slightly more complicated search, it will take about a minute to return the results.
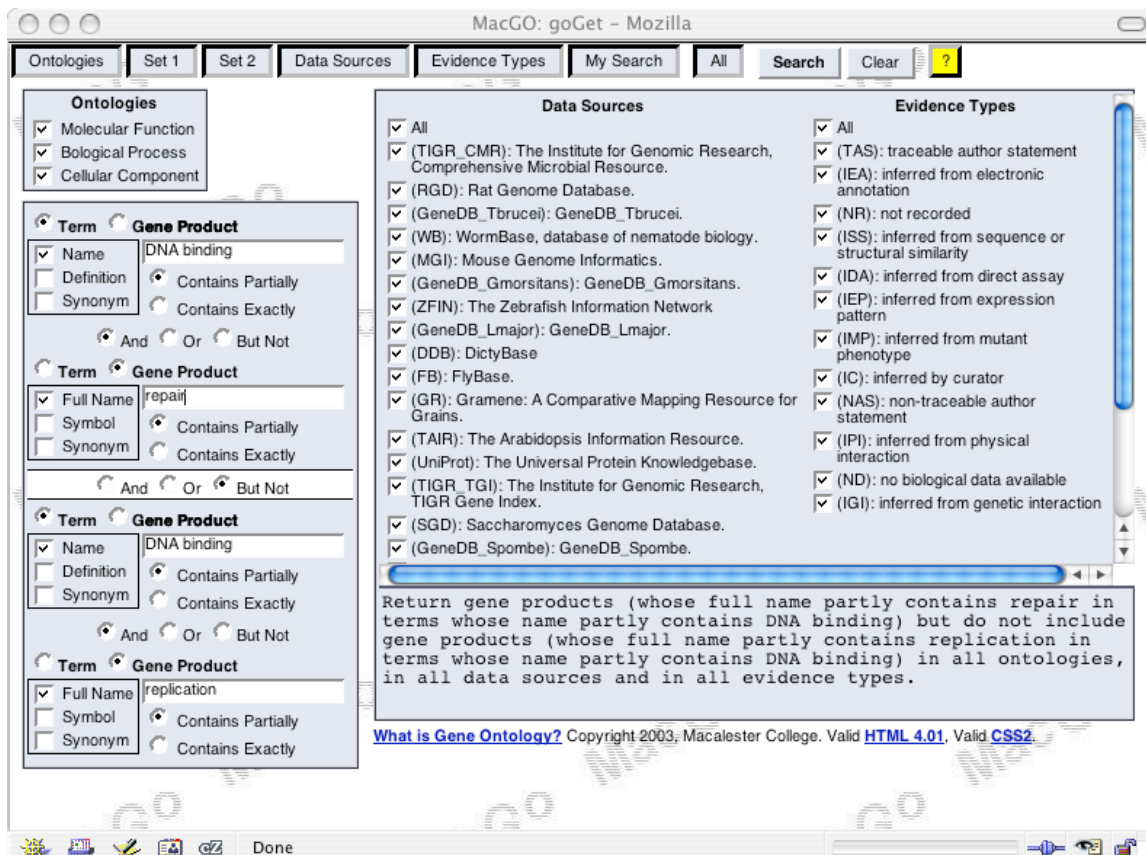
Figure 9.   Completed example using set 1 and set 2

Searching Tips

This seems like a search that could be worded any number of ways, but this is not the case— this specific setup is necessary for it to give the results that are most likely desired from the original English query "*DNA binding proteins that are involved in DNA repair but not involved in DNA replication*". The set 1 part of the query gives us a list of gene products that involve *repair* that are associated with *DNA binding* terms (technically, it is a set of gene products whose name contains the text *repair*). We then subtract any *replication* gene products that also have associations to DNA binding terms. As a check for what is going on, you can try just searching for DNA binding terms 'and' repair gene products in the set 1 box. (As of the May, 2004 version

of GO, this returns 5 terms with 1204 total gene product associations.) Then do a second search of DNA binding terms 'and' replication gene products within those terms, again using the set 1 box. This returns 4 terms with 887 gene products. The result of doing the 'but not' (a minus of the replication gene products from the repair gene products) returns 5 terms with 1067 gene products.

Also, the choice of using just the word 'repair' and just the word 'replication', as opposed to 'DNA repair' and 'DNA replication' are important. The reason is illustrated in Figure 10, which shows the results of finding DNA binding terms with associated gene products whose name contains 'replication'. Note that the full name of some of the gene products contains the phrase 'DNA replication and repair protein'. If we had chosen 'DNA repair' for our first set, the text search index that we have set up in our Oracle database would not have found these gene products. Similarly, there are some named 'DNA repair and replication protein', which also would not be returned if we had chosen 'DNA replication' in set 2, as opposed to 'replication'.

Figure 10.   Result for DNA binding terms with each associated gene product that contains the word 'replication' in its full name.

The order of the first two fields in set 1 is significant in terms of what the resulting output looks like. Because the first member of both set 1 and set 2 is a term, the *but not* between the two sets pertains to gene products. This means any <u>**gene products** </u>**returned by set 2 are removed from set 1. So if a gene product is associated to a term that contains DNA replication it would be removed.**

BEWARE: If we had chosen to switch DNA binding and repair in set 1 and DNA binding and replication in set 2 (listing the gene product first) the *but not* would have been sensitive to terms. This means terms that had DNA replication genes would have removed from the set 1 terms. This is most likely not what you would want from a query like this one.

The only change we could have made that would still return the same results are excluding DNA binding from the second. Having DNA binding in the second set is unnecessary but it is left in because the 'but not' operation is slow and narrowing the field that is being subtracted will speed it up.

3.4 Example 4: Searching Within Results

 Once you have a set of results, you can search within those to refine your results further. For example, after you have obtained the results from the query in the previous section, you could try finding only the gene products that contain the word 'excision' by using the Set 1 box and clearing out the Set 2 box (do this by unchecking all choices in Set 2).

3.5 Example 5: Another Analysis Tool: Gene Products Specific to Terms by Species

In certain situations, you may want to know whether one species has a particular function that another does not, or whether two species share a function or functions. We have created some special analysis features to help you ask these questions.

Let's walk through an example using the molecular function terms that contain the word 'triphosphatase'.

Start by pressing '**clear**' to start a new search.

Select terms in the top of set 1, check the *name* box and type in '*triphosphatase*' in the text field.

Click S**earch**.

Select all of the terms returned by the search for analysis by checking the all box under term name.

Bring up the analysis tool menu again by clicking the **Click here to analyze results** button. Your browser application should look like Figure 11.

Figure 11.   Results before analysis

This time click the **Gene Products Specific by Terms** button. You will get a screen like Figure 12.

Figure 12.   **Choosing terms that have mouse gene products but not friutfly gene products**

Pick '(MGI) Mouse Genome Informatics' from the left-hand list, 'but not' between the two lists, and '(FB) Flybase' from the right-hand list. What you are asking for by doing this is those terms from the ones you selected that have mouse gene products associated to them but do not have friutfly gene products associated to them. (Note that the latest version of the GO database now has species data for each gene product, in addition to the data source. We plan to use that in the future to help you formulate this question and get more accurate results.) Pick the **Search** button at the bottom of the blue box where you chose the two data sources. This type of database query takes a fair amount of time to execute, so you will need to wait a couple of minutes. What is returned is two of the terms and the gene products associated with them (as of the May, 2004 version of the GO database).

Using this same analysis box, you can also ask for those terms which have both mouse and fruitfly gene products associated to them (AND choice), or you can ask for those terms that have either mouse or fruitfly gene products associated with them (OR choice). You may also ask simply for those terms with mouse gene products associated to them, by choosing the 'None' choice (the right-hand column is ignored). Note that to do any of these you have to start your initial query for triphosphatase terms again.

The second analysis tool, finding gene products that are shared by a set of terms, is currently under development.

3.6 Starting your exploration in GoView

Now we will show you how to use the GoView tool. If you want to just explore around a bit and see what kinds of terms are in each ontology and see what the graphs look like, this is the place to start. Start out by opening up GoView. In Section 2 we provide download information. GoView looks like Figure 13 when it opens.

You will need to do this once: Choose 'Properties' from the 'Edit' menu located in the menu bar at the top of the tool. The only property you need to set is which browser you will use with GoView to display additional information. If you are on a Mac, you can use the 'Select' button and choose your browser. The GoGet tool displays best in the Firefox or Mozilla browser, but works pretty well in Safari on Mac OS 10.3. On Windows, Mozilla is again better than Internet Explorer. On linux, Mozilla is also the best choice, and you will need to type in the path to your Mozilla executable. *You must choose a browser this way in GoView, otherwise the following tutorial will not work.*

Each time you use GoView: you will need to make sure the browser you have chosen using Edit-Properties **is not open and running**. GoView needs to start the browser itself in order to establish a connection between itself and the browser to send information back and forth.
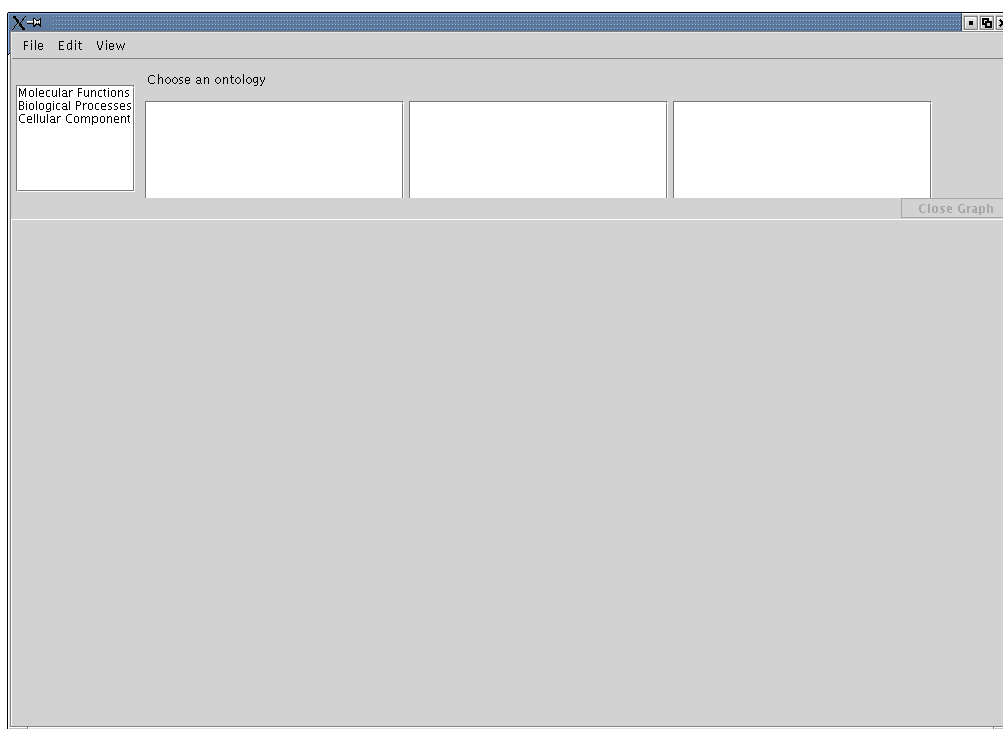

Figure 13.   GoView at startup

The GoView tool is a starting point for exploring the ontologies and their terms. To use GoView you will need to start by selecting one of the three ontologies that you would like to explore.

For this tutorial we shall look at the following example:

What are the gene products that are associated with cell growth regulation, in rats?

To answer this question, try the following:

- select *Biological processes* ontology from the list on the upper left hand corner of the screen. This will open up a list of graphs of biological processes.

- Next select *cellular processes (*)²,* in the next list. The asterisk following *cellular process* just symbolizes that the graph is very large and so another list will open to show you the terms that are 'children' of cellular process.

- From the resulting list select *cellular physiological processes (*),* and then from the next list select *cell growth and maintenance (*),* and finally select *cell growth* from the last list.

So the overall sequence is:

Biological processes → *cellular processes (*)* → *cellular physiological processes (*)* → *cell growth and maintenance (*)* → *cell growth*

This will bring up the graph of cell growth and its related terms. There will be two graphs displayed: a readable one on the right and a small one to the left, called the context graph, because it provides the overall context of what this graph looks like. This context graph is especially useful for the larger graphs in GoView (there are many- you can try opening up other graphs if you'd like- they come up in separate 'tab' windows in GoView). We've chosen this smaller graph to make it easier to show you how to pick terms and get more information about the gene producst in them.

NOTE: *The first time the graph is displayed, the context view on the left may not appear. Just resize your window a bit, and it will show up. We're trying to fix this problem.*

You can move around the graph by using the scroll bars or by clicking on the context graph on the left where you would like to be centered. You can also move the red box in the context graph and the readable graph on the right will move accordingly to match its location on the context graph.

Now going back to our example, select the node/term called *regulation of cell growth* on the readable graph by clicking on it, and it should get highlighted in red (as shown in the screen shot in Figure 14).

---

²Due to the size of some term names, each list panel is programmed to become wider when the mouse is over it. We did this to enable complete visualization of the longer terms, but it might take some getting used to.
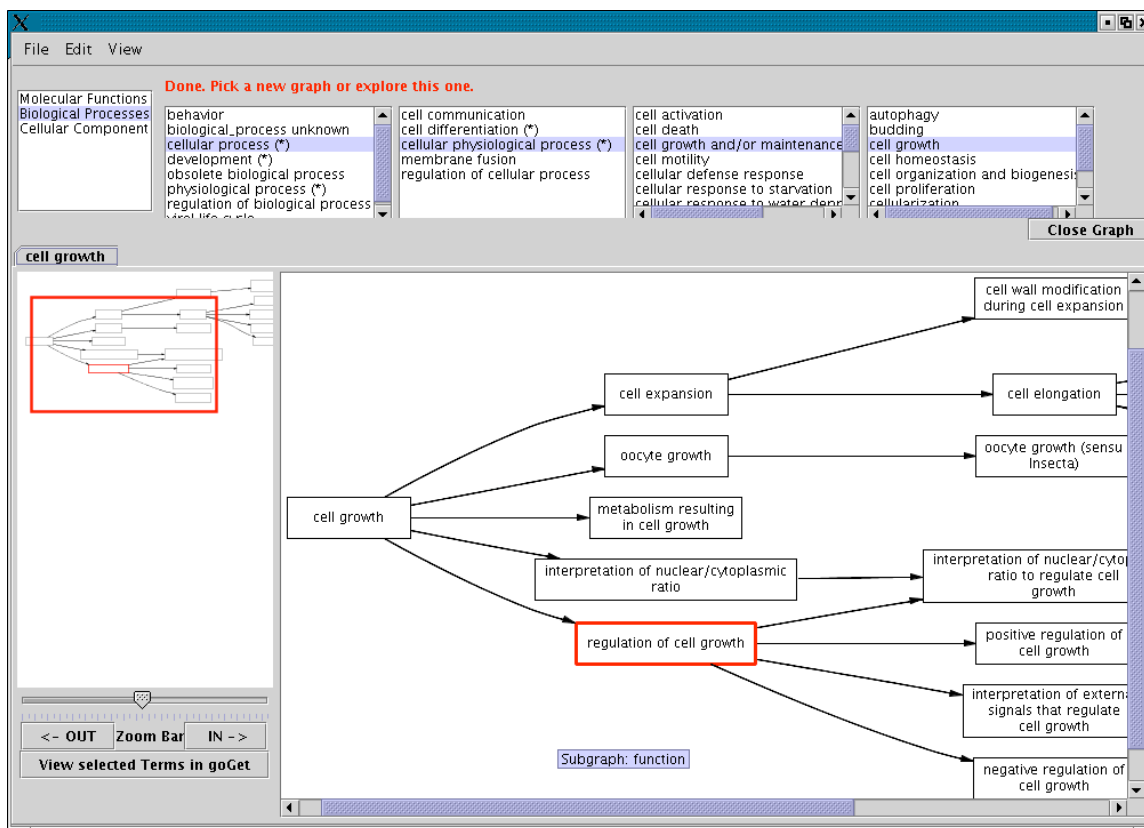
Figure 14.   A snapshot of GoView displaying the cell growth subgraph.

Once you have selected *regulation of cell growth*, click on the button "**View Selected Terms in goGet**[3]*"* situated on the lower left corner.

This will open up GoGet in a web browser window with more information on regulation of cell Growth, as shown in Figure 15.

---

[3] Before clicking make sure that you have selected a browser. The browser that work best with this is mozilla. Before selecting a browser, make sure that all of your mozilla or safari browsers are closed. After closing your browser, click on "*edit"* (at the top of the page) then select "*properties"*. A small window should open. Then select *'select'* and it should open the applications page, which contains mozilla. If you are using linux operating systems, find mozilla's path and copy and paste in into the text field by 'browser'.

Figure 15.   GoGet when called from GoView

Notice that it does not display the Definition or the Gene Products' Information. To view these click on the "**display**" buttons under "**Definition**", and "**Gene Products' Information**". The tables should unfold and the information should be displayed as shown in Figure 16.

Figure 16.   Definition and Gene Products' Information

As we can see from the #of gene products column, *regulation of cell growth* has 504 gene products.

From here we can narrow our result to be specific to one or more organism. Choose the 'Set 1' button, pick 'Term', then pick only 'name' and type in 'regulation of cell growth' in place of the default 'triphosphatase'. (This is a step that we will eventually remove, as we realize that it's a bit awkward and doesn't work well when you've chosen multiple terms to send to GoGet.) Choose the 'Data Sources' button, select 'All' to toggle off all of the data sources, then pick the 'Rat Genome Database.' You should get a smaller number of gene products, as shown in Figure 17.

Figure 17.   Search results after narrowing by database.

Now that we have narrowed down our gene products lets get a little more information about them.

Just to make sure that the gene products are only from rat genome, click on the tab "**Gene Products' information**", this should collapse and display a table with information on the gene products. From here click on the "**display**" tabs in the column labeled "**Data Source information**" and they should all display a rat genome database website[4] URL. Click on the three display boxes under Product Synonyms, Data Source Information and Evidence Information in the row of the first gene product. This will bring up more specific information including synonyms, the type of evidence that links this gene product to the term and a link to a website with more information. Figure 18 shows the data source information displayed for the first gene.

---

[4] As of now, the URL http://rgd.mcw.edu contains only rat genome database.

Figure 18.   Gene product data source information displayed

Now click on the link under the URL column. This will open up the web page of the database this gene product is from (Figure 19). *NOTE: these website links are fragile- we try to keep up, but sometimes they change and may not work.*

Figure 19.   Rat genome database web page for a specific gene product

3.7 Getting multiple nodes from GoView and The GoView Application Connection.

Go back into GoView[5]

If you wish, you can close the cell growth graph by hitting the **close graph** button on the upper right, above the readable graph, or by choosing the 'X' in the tab of the graph, to the left of the graph's name..

Click on the *cellular component* ontology on the left, and then on *cell(*)*, then on *external encapsulating structure* to bring up the graph (Figure 20). Each time you choose an ontology, a subgraph list like cell(*), or a graph itself, you have to way a few seconds for the tool to find the next level list or the graph itself and display it. Some information about the status is shown in read print above the place where the graph will eventually be displayed.

---

[5] You do not have to download it again, if you have saved it.
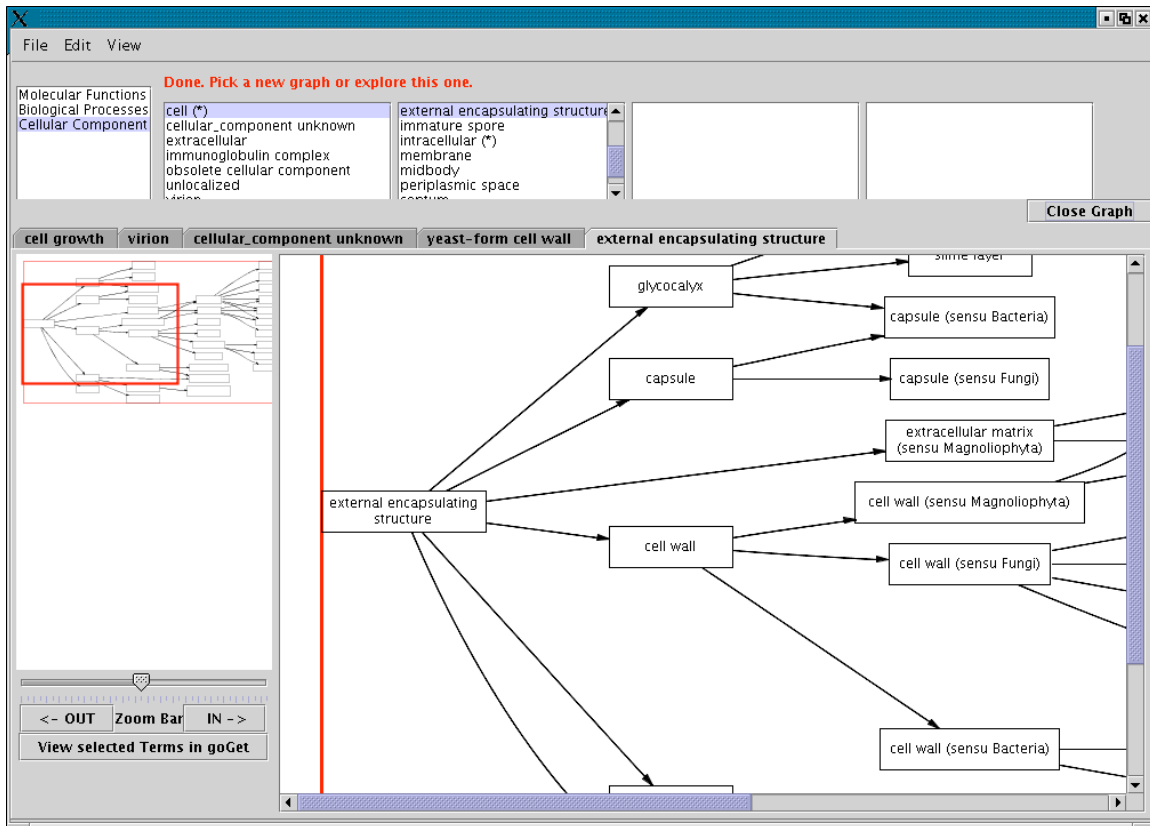
Figure 20.   External encapsulating structure graph

Now lets play with the zoom a bit.

Using either the **<-Out** button or the zoom slider bar to zoom all the way in on the graph (Full zoom out is shown in Figure 21).
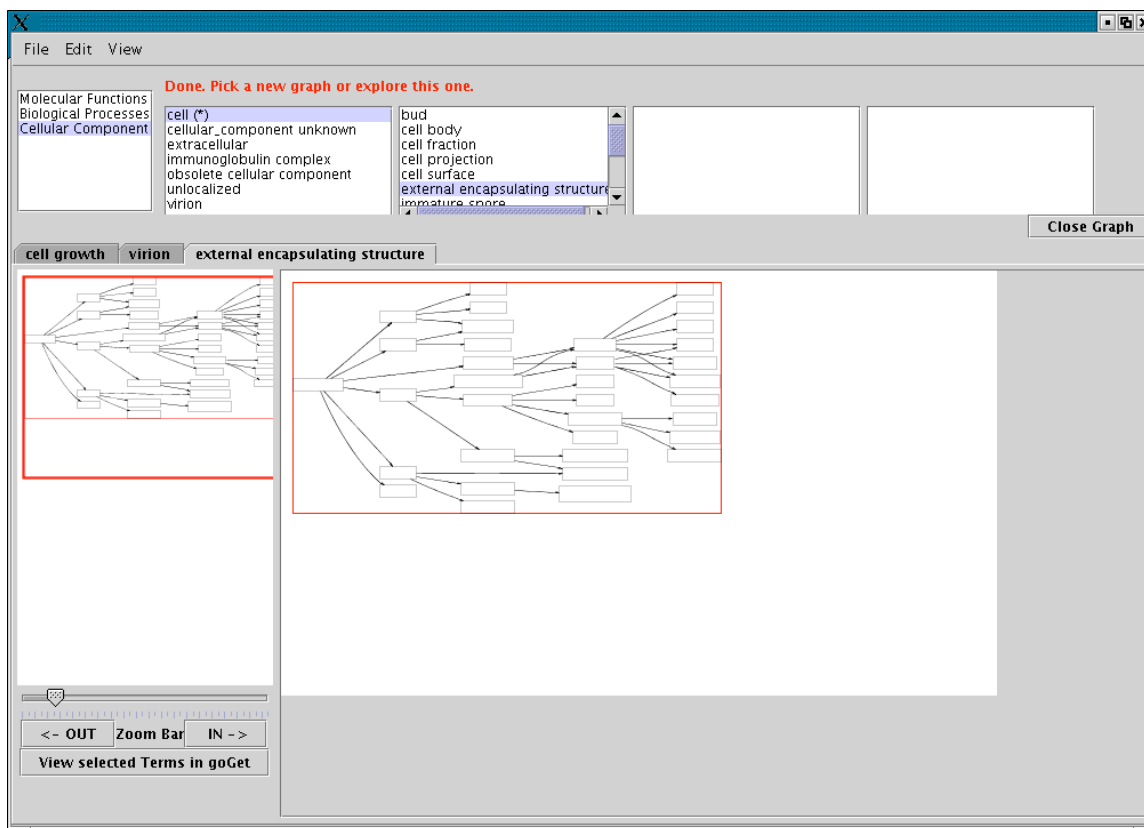
Figure 21.   Zoom out

Zoom back in a bit and select all the terms in the last two columns of nodes in the graph, as shown in Figure 22.
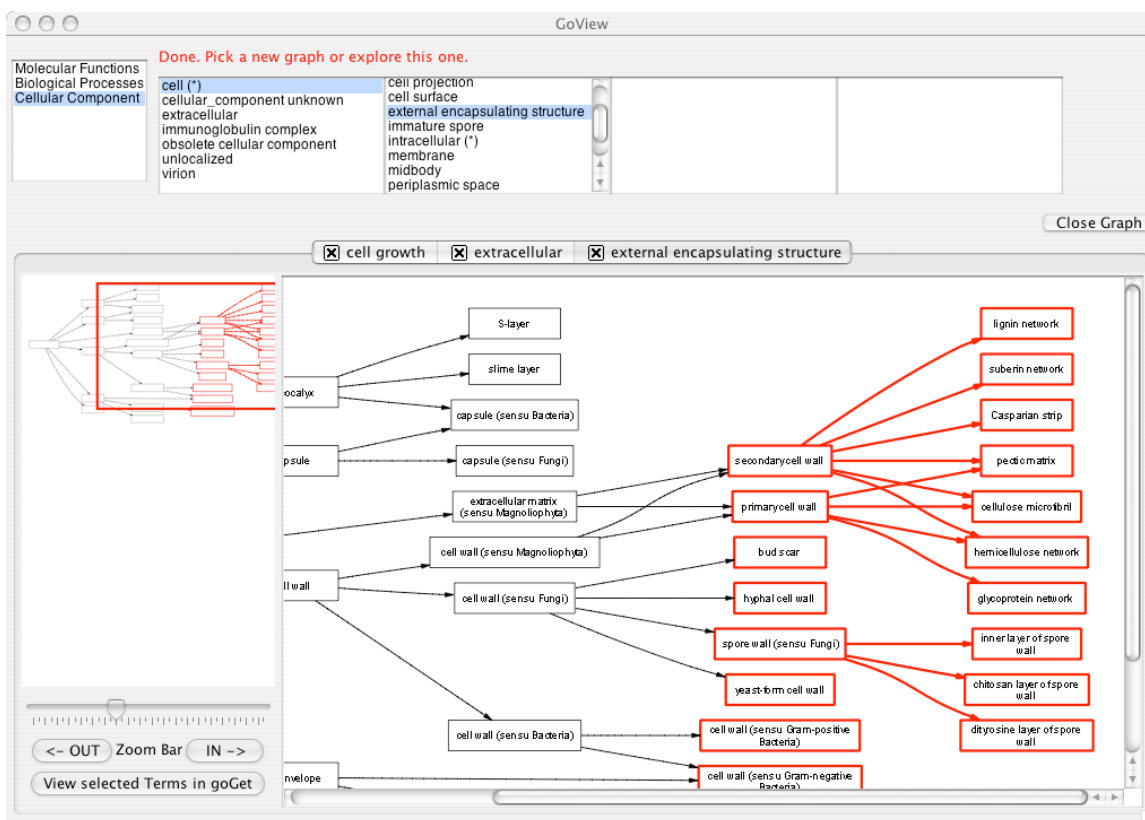
Figure 22.   Choosing multiple nodes in GoView

Now that you have selected those nodes, click **View selected Terms in GoGet**.

The GoGet browser window will display all of the terms and their gene products in a few seconds. We now have a big list of terms in GoGet. You can browse them in GoGet and see what gene products are associated with those terms.

Re-analysis example

The terms on this graph might be in other cellular component graphs. This is a byproduct of the way we have to spilt each very large ontology into smaller subgraphs. We can see what other graphs these terms are in by sending them back to GoView and getting a list of all graphs those terms are in. In GoGet, select all of the terms by checking the all box above the column of term names.

Bring up the analysis tool menu again in GoGet by clicking the **Click here to analyze results** button. The result is shown in Figure 23.

This time click on **GoView Application Connection.** A popup window will come up and give you an English version of what you are asking to be sent back to GoView. Hit the search button in that popup window (Figure 23). In a moment, this will bring up a little window of the GoView application with the names of all of these terms on the right and the graph that each is in on the left (Figure 24).

Figure 23.   GoGet popup window to send terms back to GoView

Figure 24.   GoView's graph list window that pops up when you send terms to it from GoGet

Minimize the GoGet application or move it out of the way and look at the GoView popup list shown in Figure 24. Notice that the top line shows that the 'Casparian strip' term is also in a graph called 'extracellular'. Pick that by clicking once on it in the list. GoView will take a few seconds to draw the new graph in a new tab and center the graph at the 'Casparian strip' node. Notice that other terms are also in the extracellular graph. If you chose one of those from the list, the graph will center on that term. Try closing the graph by clicking on the **Close Graph** button

or the X in the tab and than clicking on a member of that graph. Notice that the graph comes back up again with the term you chose centered.

NOTE: The first time you bring up a graph with a term, it may not center on that term. Just click it again in the list, and it will. Sometimes Java graphical user interface programming can be painful and produce odd results that are difficult to do much about ☹.

We've presented a few features of each of the two tools and shown how they can be used together. We hope that you have fun with them and we are always open to suggestions, which can be sent to goproject@macalester.edu.

4.0 Reference

This section should be used as a reference for the specific functions of aspects of the programs after going through the tutorial.

The information in GoGet can be accessed in two ways: Finding terms on a GoView graph or searching using GoGet search tools.

4.1 Exploring GoView

4.1.1 Opening a graph

Once GoView is open chose between Molecular Functions, Biological Processes and Cellular Components in the upper left corner of the screen. Clicking on the one of these three will bring up a list of graphs from this Ontology.

To see the graph of a member of this list click on it. Components, Functions, or Processes with an asterisk will have more subgraphs. Once a graph is selected it will appear in the large window and on the left in the context graph.

To close the graph click on the **Close Graph** button on the right.

More than one graph may be open at a time, to open additional graphs repeat the above process.

To switch between open graphs click on the tabs above the graphs.

To move to a specific place in the graph click where you would like to be centered on the context graph on the left. To scroll on the graph use the scroll bars on the larger graph.

To zoom in or out on the large graph drag the Zoom Bar in the lower right corner or hit the **IN->** or **<-OUT** buttons to zoom in or out.

4.1.2 Getting more information in GoGet

To find more information on a component highlight it in red by clicking on it. To select multiple nodes at once click and drag a box completely around the desired nodes. This can be done on the large graph or the small graph on the left.

Once you have selected the term(s) you want click **View Selected Terms in GoGet** in the lower left corner. This will open up GoGet in an Internet browser if it is not already open.

GoGet will come up with all the same information about that term as it would have if you searched for it with GoGet.

4.2 Searching in GoGet

To open GoGet without looking for a specific term from GoView go to the *view* menu in GoView and click *open goGet*.

In GoGet there will be 8 buttons along the top of the screen with the following functions:

Ontologies: Show/Hide Ontology search criteria.

Set 1: Show/Hide Set 1.

Set 2: Show/Hide Set 2.

Data Sources: Show/Hide Data Source search criteria

Evidence Types: Show/Hide Evidence Type search criteria.

My Search: Show/Hide English summery of search.

All: Show/Hide all search options.

Search: Begin search.

Within Results: Begin search of of results

Clear: Reset and clear the search fields and select all Ontologies,

Data Sources and Evidence Types.

To begin a search click on the **All** button if the search criteria are not already on the screen.

The starting point of GoGet is seen in Figure 25.
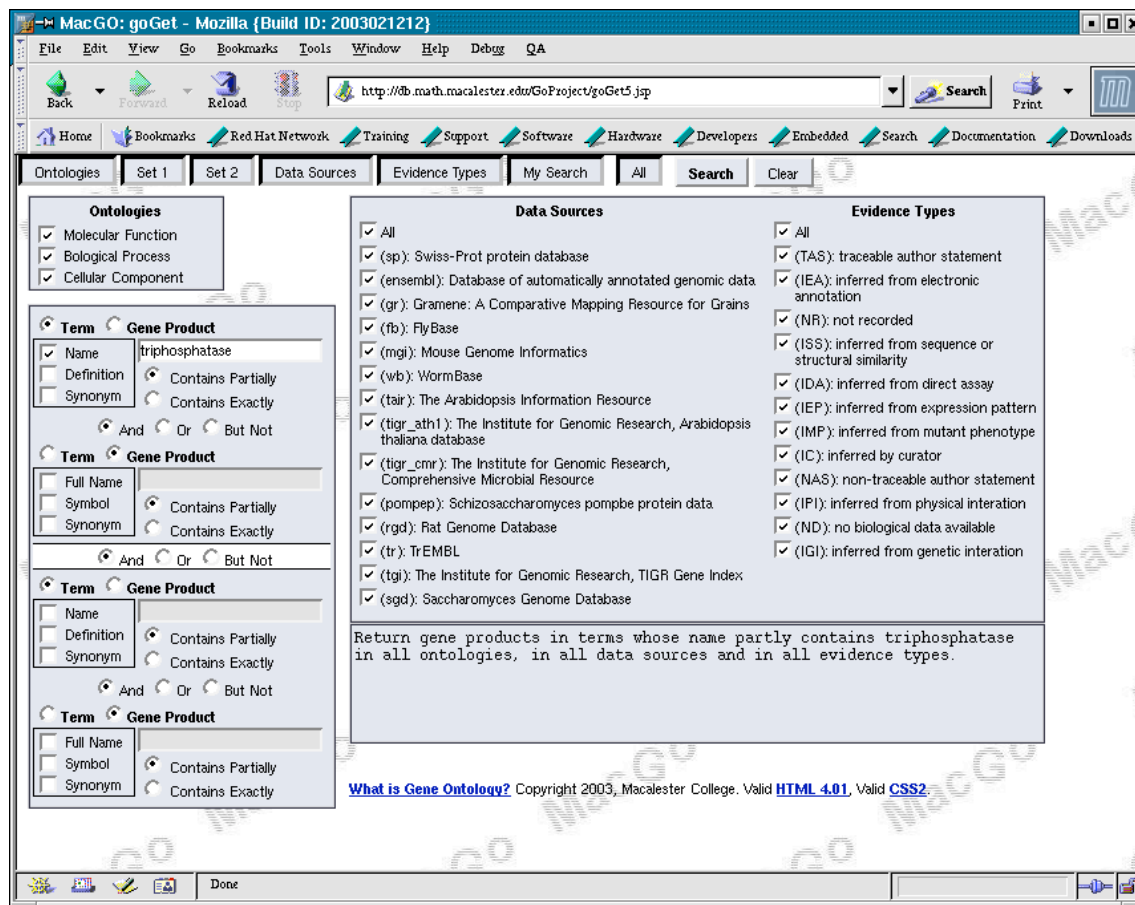
Figure 25.   Opening view of GoGet

4.2.1 Ontologies

In the Ontologies field check which of the three ontologies to search. Select one or more ontology by clicking the check boxes on the left.

4.2.2 Using Set 1

Inside a set there are a maximum of two text fields you can search. You can choose to search terms or gene products for each of these fields.

Check the term radio button if you wish to search the terms. Select if you would like to search the *term names*, their *definitions* and/or their *synonym*s by checking one or more check boxes. Type in the word or phrase you would like to search for.

Check the gene product radio box if you wish to search by the gene products. Select if you would like to search the gene products *full name*, their *symbol* and/or their *synonyms* by checking one or more check boxes. Type in the word or phrase you would like to search for.

If you use both fields select *And*, *Or* or *But Not* by marking the radio box.

And will only return gene product/term pairs the meet the criteria on both sides of the and. If you searched "DNA binding" in terms *and* "DNA repair" in gene products it will return only gene products that have "DNA repair" in the selected gene product fields and are associated with terms that have "DNA binding" in the selected term fields.

Or will return all gene product term pairs that meet either criteria. If your searched "DNA binding" in terms *or* "DNA repair" in gene products it will return all gene products that have "DNA repair" in the selected gene product fields with all terms associated with them and will also return all gene products that are associated with terms that have "DNA binding" in the selected term fields.

But not will return all the terms pairs meet the first criteria but do not meet the second. So

if you searched "DNA binding" in terms *but not* "DNA repair" in gene products it will return all terms that have "DNA binding" in the specified fields with their gene products but it will not include gene products that have "DNA repair" in specified fields.

4.2.3 Using both sets

Start by filling out both sets as if you were using them alone.

There are three queries options to choose for using both sets: and, or and but not. These have the same names as the searches inside the sets but have different functions. They are more general than their counterparts in the sets and have different functions depending on what is in set 1.

What determines their function in what is selected in the top of set 1: term or gene product. If the very top radio box is gene product than the search will be term sensitive. This means if the *and* radio box is selected the pairs from the top will be compared to the terms from the second box and any pairs from the first set that do not have one of those terms from the second set will be eliminated. So the "and" is sensitive to the terms alone not the pairs.

But not works the same way. Any pairs from the first set that have a term from the second box will be removed.

Or is different this will just combine the two list together removing any repeat pairs.

When the top radio box is term than the search will be sensitive to gene products. This means if you use "and" it will look at the list of pairs from the first set and remove all pairs that do not have a gene product that is a member of the second list.

But not works the same way. Any pair from the first set that has a gene product that is a member of the second set result will be removed.

Or will do the same thing it did before and combine the two lists not repeating pairs it is still sensitive to entire pairs not just gene products.

4.2.4 Data Sources

Select which of the databases to search. Select one or more databases by clicking the check boxes to the left of the name of the database. Select or deselect all databases by clicking the *All* check box.

You may wish to do this if you are searching for gene products in a specific organism. For example if you are looking for gene products in ***drosophila*** you want to search Flybase(fb).

4.2.5 Evidence

Select which of the 12 types of evidence types that will be allowed in the search. Select one or more of the evidence types by clicking the check boxes to the left of the type of evidence. Select or unselect all evidence types by clicking the *All* check box.

The data in the Gene Ontology Database is of variable reliable because of what types of evidence it is based on. Depending on how confident you want to be in the information you can choose what types of evidence you want to accept.

For more information about the evidence types, go to:

> *http://www.geneontology.org/doc/GO.evidence.html*

4.2.6 Search

Once these fields are filled out click search along the top bar to begin the search. If it does not begin to search one of the fields may not be complete.

If this is the case a box will pop up informing you that your search is incomplete check that you have written something in the text field. This may also happen if you have not selected any ontologies, data sources, or evidence types.

If the search was successful results will be displayed and the search tools will be hidden. If there were no results the search tools will remain open and a box will pop up that says "no match found".

4.2.7 Searching within results.

This button works the same as the search button but the search is done only on the results of the previous search and not on the entire database.

4.2.8 Results

Depending on what how the data in GoGet was searched for it will bring it up data in one of three ways.

If GoGet was brought up by a search of only terms or through GoView it will have a list of terms and each term will have a list of associated gene products

ACC	Term Name ☐ All	Definition	Ontology	Term Synonyms	Gene Products' information	# of gene products
8413	☐ 7,8-dihydro-8-oxoguanine-triphosphatase	none	function	none	display	2
8094	☐ DNA dependent adenosinetriphosphatase	none	function	display	display	130
8186	☐ RNA dependent adenosinetriphosphatase	none	function	display	display	15
4002	☐ adenosinetriphosphatase	none	function	none	display	350
17111	☐ nucleoside triphosphatase	none	function	none	display	1

A term results table

If GoGet was brought up by a search of only gene products it will list the gene products and their information and then list the associated terms under each gene product with the term information.

Data Source ID	Full Name	Symbol	Ontology	Product Synonyms	Terms' information ☐ All	# of Terms
26653	7,8-dihydro-8-oxoguanine triphosphatase	8ODP_HUMAN	function	display	display	5
27879	Bis(5'-adenosyl)-triphosphatase	FHIT_HUMAN	process	display	display	5
51537	deoxyuridine triphosphatase	Dutp	function	display	display	1
67671	RNA 5'-triphosphatase, mRNA capping enzyme beta subunit (80 kDa)	CET1	function	display	display	1
67898	RNA triphosphatase	CTL1	process	display	display	1

`A gene products results table`

If GoGet was brought up by a search of both gene products and terms it will have a separate column for each gene product term pair with all term and gene product information.

ACC	Term Name ☐ All	Defintion	Term Synonyms	Ontology	Data Source ID	Full Name	Symbol	Product Synonyms	Data Source information	Evidence information
8094	☐ DNA dependent adenosinetriphosphatase	none	display	function	46251	H-ATPase	OSA1	none	display	display
8094	☐ DNA dependent adenosinetriphosphatase	none	display	function	70482	DNA dependent ATPase, human Cockayne syndrome B gene ERCC6 homolog	RAD26	display	display	display
8186	☐ RNA dependent adenosinetriphosphatase	none	display	function	46251	H-ATPase	OSA1	none	display	display
4002	☐ adenosinetriphosphatase	none	none	function	19667	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	ATA1_HUMAN	display	display	display

`A full results table`

Term Information:

> *ACC#:* Accession number of the term in the database

> *Term Name:* Full name of the term

> *Definition:* A brief definition of the term

> *Term Synonyms:* Other names for this term

Gene Products' Information:

> *Data Source ID*: ID number of the data source

> *Full Name:* The full name of the gene product

> *Symbol:* Symbol of this gene product

> *Product Synonyms:* Other names for this product

These categories will always be listed along with the Term information if you searched by terms or along with the gene products if you searched by gene products:

> *Ontology:* Function, process or component.

> *Gene Products/Terms Information:* More specific information on the genes and gene products involved in this process, function or component, or more Information about the terms associated with this gene product.

> *# of gene products/# of Terms:* The number of gene products in the database for this Component Process or Function or the number of terms associated with this gene product.

This information will always be listed with the terms associated with a gene product or with the gene products of that term:

Data Source Information: The name of the database this product came from, comments and a link to the databases' specific information on this gene product.

Evidence information: which of the twelve evidence types this information was gotten from.

Some parts of the table will not be listed and you will need to click the **display** button in the corresponding space to view them.

To hide this information click the **hide** button in the same space.

4.2.9 Analysis

To analyze the gene products click the **click here to analyze results** button.

To hide the analysis tools click the **click here to analyze results** button again.

The **click here to analyze results** button will bring three options: Gene Products Specific to Terms, Gene Products Shared by Terms and GoView Application Connection.

To use any of these tools you will need to select one or more terms to analyze. If you got your terms through a search by terms or through GoView click on the check boxes to the left of the terms. If you got your results by searching for a gene product you will need to click on the display button under Terms' Information to bring up the check boxes for the terms.

Gene Products Specific to Terms

This will narrow the gene products by which database they came from.

Select which term(s) gene products to look for.

Choose one or two databases and specify if you want to see the products that are in one, both, either, neither or one but not the other.

Once you have selected the database(s) click search in to find the gene products. Terms that have no products in the search will no longer be displayed.

To close without searching click **cancel**.

This will return results in the same fashion as a search by term name.

Gene Products Shared by Terms

This will find out which products are shared by one or more term(s). This feature is currently under development.

Select which terms to check for shared products in.

Once the terms are selected click **search** to search.

To close without searching click **cancel**.

This search returns data the same way as if you searched by gene products.

To further analyze these results you can check the boxes of whatever terms you would like more information on and use the analysis tools on them.

They will all start out being selected, to change this you will need to press the details button to see their check boxes.

4.2.10 GoView Application Connection

This will center the GoView graph window on specific terms. You needed to have started GoGet by launching the browser from GoView, as described in section 3.

Select which term(s) you would like to be able to center the GoView graph on.

Once the terms are selected click the **search** button. This will open up a new popup window that is part of the GoView program. It will be a list of the selected terms.

To cancel and not open this window click **close**.

To center the GoView graph on a term click on that term in the new popup window.

Terms will not be removed from this window until the program is closed. More terms can be added by repeating this process.

If you only open up one term in the GoView Application Connection you will only be able to center the graph on that term once but if you add more terms you will be able to do it indefinitely.

If you close this window it can be opened again by going to View-> Open Graph List.

This tool is also useful for finding terms on the graphs that we located in a GoGet search. If we know a term and we want to see how it fits in the graph we can search for its name in GoGet and then use this tool to find it on the graphs.

5.0 Contacting Us and Acknowledgments

This is a new project, and any questions, comments and suggestions are welcome.

Contact us at: goproject@macalester.edu

This project was primarily the work of the following Macalester students: Paulo Casaes, Getiria Onsongo, Lisa Lesnett, Dennis Tkach, Erla Osk Petursdottir, Edward Kofi Yeboah Donkor, Michael Cosimini, Nyalleng Moorosi, Daniel Feldman, and Elizabeth Gabriel.

6.0 Bibliography

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J., Richardson, J., Ringwald, M., Rubin, G., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. the Gene Ontology Consortium. *Nat Genet*, 25(1):25–29.

Ashburner, M. and Lewis, S. (2002). On ontologies for biologists: the gene ontology–untangling the web. In *Novartis Found Symposium*, volume 247, pages 66–80.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6):907–928.

The Gene Ontology Consortium (2001). Creating the gene ontology resource: design and implementation. *Genome Res*, 11(8):1425–33.