Bioinformatics Database Task

Peter Hebden

16 April 2019

Abstract

What are the gene names for all variants of the protein synaptotagmin (Syt)? MySQL Workbench and the GO Ontology Data was useful for this task. What proteins are necessary for neurotransmitter vesicle docking and fusion? Quick GO was very useful for this task.

1 Introduction

At the moment, my computational neuroscience research focus is on spike sorting electrophysiological data recorded mostly from rat hippocampus, a brain area that is essential for spatial navigation. But I am also interested in the mechanisms that support synaptic transmission between pyramidal neurons, e.g. at the CA3–CA1 synapse in the hippocampus. Synaptic transmission of the neurotransmitter glutamate occurs after vesicles dock with the active zone in the presynaptic terminal. Docking requires a complex of proteins that bind to the vesicle and the terminal's plasma membrane and to each other, known as v-SNAREs and t-SNAREs respectively. When an action potential arrives at the presynaptic terminal, voltage gated calcium channels (VGCCs) open, calcium ions rush in and bind to ∼5 sites on the protein synaptotagmin, aka "the calcium sensor". When calcium binds to synaptotagmin it triggers a conformational change that causes the vesicle to fuse with the plasma membrane and release neurotransmitter into the synaptic cleft. In addition to synaptotagmin, there are other proteins in the SNARE complex.

2 Methods and Results

2.1 Introduction

The protein synaptotagmin 1 (gene name SYT1) is known as the calcium sensor that triggers vesicle fusion with the pre-synaptic plama membrane and the release of neurotransmitter into the synaptic cleft. This protein has several variants and is part of the SNARE complex which is assembled from a number of proteins.

To get started I searched Google for anything that might be useful for building SQL queries. The Google search terms "synaptotagmin rat uniprot" led me straight to the UniProt entry Synaptotagmin-1: `https://www.uniprot.org/uniprot/P21707`. Syt1 is the name of the protein coding gene in Rat.

Scrolling down to GO Biological Process, there is an entry for "synaptic vesicle exocytosis" and the link `https://www.ebi.ac.uk/QuickGO/term/GO:0016079`; and an entry for vesicle docking and `https://www.ebi.ac.uk/QuickGO/term/GO:0048278`.

2.2 MySQL WorkBench

GO database: the central tables are **term** and **term2term**. Annotations are stored in the go-association module; the main tables are **association** and **gene_product**.

Using MySQL Workbench, I queried the GO database with common names, or with GO accession numbers obtained from UniProt. In Figure 1 each simple SQL query is followed the returned entry with column headings: id, name, term_type, accession number. Each term name is classified as either a cellular component, a biological process, or a molecular function. Given a term name, I found its GO number and vice versa.

```
SELECT * FROM term WHERE name='synapse';
20690 synapse cellular_component GO:0045202

SELECT * FROM term WHERE name='exocytosis';
5206 exocytosis biological_process GO:0006887

SELECT * FROM term WHERE name='vesicle';
15219 vesicle cellular_component GO:0031982

SELECT * FROM term WHERE name like 'SNARE%';
165 SNARE binding molecular_function GO:0000149
14449 SNARE complex cellular_component GO:0031201

SELECT * FROM term WHERE acc='GO:0016079';
9836 synaptic vesicle exocytosis biological_process GO:0016079

SELECT * FROM term WHERE acc='GO:0048278';
23590 vesicle docking biological_process GO:0048278 0 0
```

Figure 1: Simple SQL queries to find relevant GO numbers and terms.

Finding gene products in the GO database with a gene symbol that starts with Syt and genus Rattus.

```
SELECT
  *
FROM
  gene_product
  INNER JOIN dbxref ON (gene_product.dbxref_id=dbxref.id)
  INNER JOIN species ON (gene_product.species_id=species.id)
WHERE
  symbol like 'Syt%'
AND
  genus='Rattus';
```

Figure 2: SQL query for all entries from gene_product where gene symbol is starts with 'Syt' and organism genus is 'Rattus'.

The above query returned 31 rows. The first row is shown in Table 1. Most rows cross referenced the Rat Genome Database (RGD), while three rows cross referenced UniProtKB. The gene symbols ranged from Syt1 to Syt17, symbols for synaptotagmin-like variants including Sytl1, Sytl3, Sytl4, Sytl5, and others.

id	**symbol**	**dbxref_id**	**species_id**	**type_id**	**full_name**	**id**
4053459	Syt1	4176890	424525	27190	synaptotagmin I	4176890

xref_dbname	**xref_key**	**id**	**ncbi_taxa_id**	**common_name**	**genus**	**species**
RGD	3803	424525	10116	Norway rat	Rattus	norvegicus

parent_id	**left_value**	**right_value**
137516	843588	843589

Table 1

However, the query for symbol Syt1 `AND 'UniProtKB';`

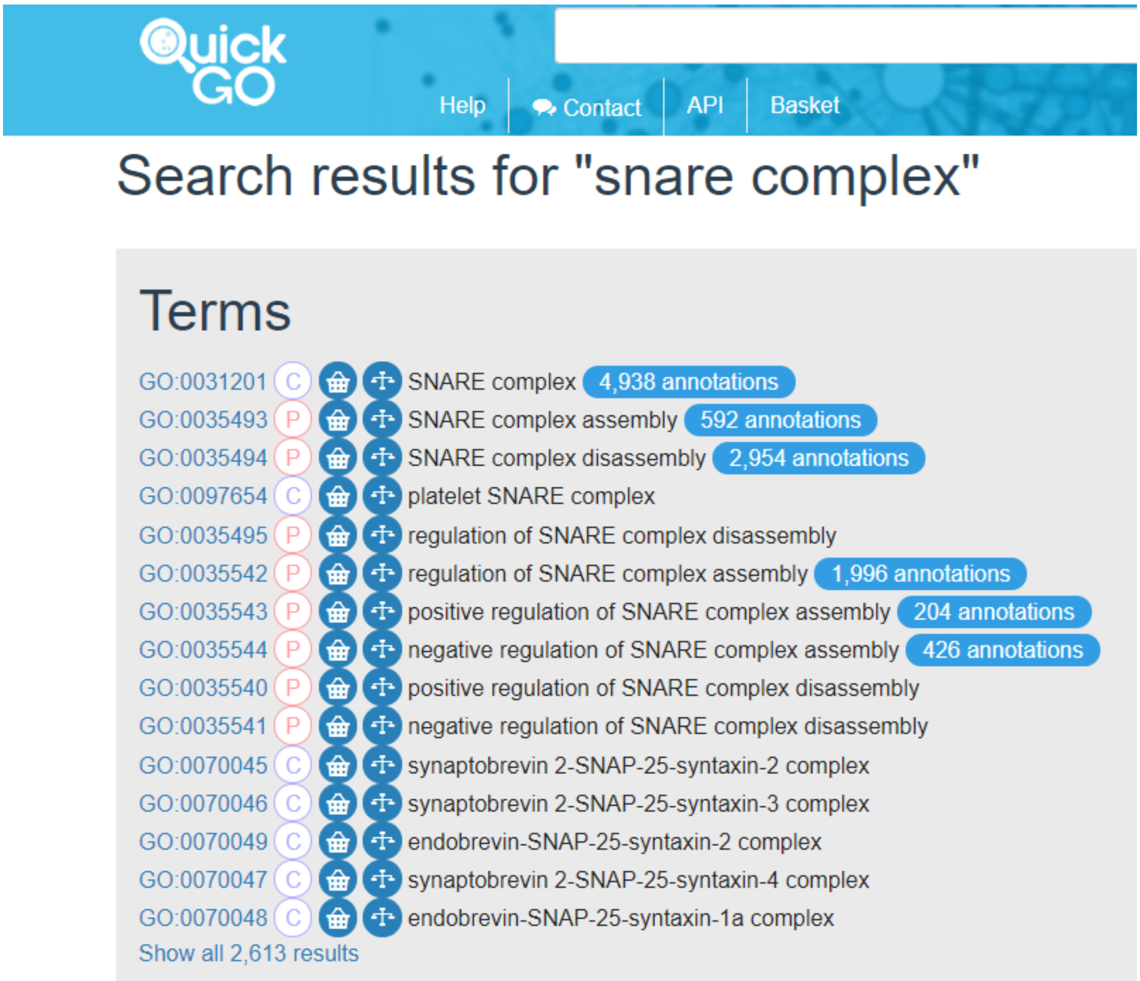
```
...
WHERE
  symbol = 'Syt1'
  AND
  xref_dbname = 'UniProtKB';
```

returned results for common names chicken, cattle and human, but not for rat even though that entry exists at UniProt.

UniProt provided a link to Quick GO: `https://www.ebi.ac.uk/QuickGO/`.

2.3 Quick GO

I queried Quick GO with "SNARE complex" and it returned the list of terms shown in Figure 3. I clicked on the first entry, GO:0031201, a cellular component with 4,938 annotations. This returned the definition "A protein complex involved in membrane fusion; a stable ternary complex consisting of a four-helix bundle, ... ", and a tree graph with SNARE complex at the bottom, the most specific term in the tree, but I needed a more specific term (a child term).



Search results for "snare complex"

Terms

GO:0031201 C SNARE complex 4,938 annotations
GO:0035493 P SNARE complex assembly 592 annotations
GO:0035494 P SNARE complex disassembly 2,954 annotations
GO:0097654 C platelet SNARE complex
GO:0035495 P regulation of SNARE complex disassembly
GO:0035542 P regulation of SNARE complex assembly 1,996 annotations
GO:0035543 P positive regulation of SNARE complex assembly 204 annotations
GO:0035544 P negative regulation of SNARE complex assembly 426 annotations
GO:0035540 P positive regulation of SNARE complex disassembly
GO:0035541 P negative regulation of SNARE complex disassembly
GO:0070045 C synaptobrevin 2-SNAP-25-syntaxin-2 complex
GO:0070046 C synaptobrevin 2-SNAP-25-syntaxin-3 complex
GO:0070049 C endobrevin-SNAP-25-syntaxin-2 complex
GO:0070047 C synaptobrevin 2-SNAP-25-syntaxin-4 complex
GO:0070048 C endobrevin-SNAP-25-syntaxin-1a complex
Show all 2,613 results

Figure 3: Quick GO query. Image: `https://www.ebi.ac.uk/QuickGO/search/snarecomplex`.

Clicking on Child Term GO:0070768 listed below the tree graph (synaptotagmin-synaptobrevin 2-SNAP-25-syntaxin-1a-syntaxin-1b-Rab3a complex), which is_a SNARE complex, returned the list of proteins that I was searching for, plus a list of synonyms, and

- Definition: A SNARE complex that contains synaptotagmin, synaptobrevin 2 (VAMP2), SNAP-25, syntaxin 1a, syntaxin1b, and Unc13b (or orthologs thereof). PMID:8999968.
- Synonyms:
  - Snap25-Syt1-Unc13b-Vamp2-Stx1b2-Stx1a complex.
  - SNARE complex (Snap25, Syt1, Unc13b, Vamp2, Stx1b2, Stx1a).

and the graph shown in Figure 4, which added the child term of interest to the bottom just below ‘SNARE complex’.

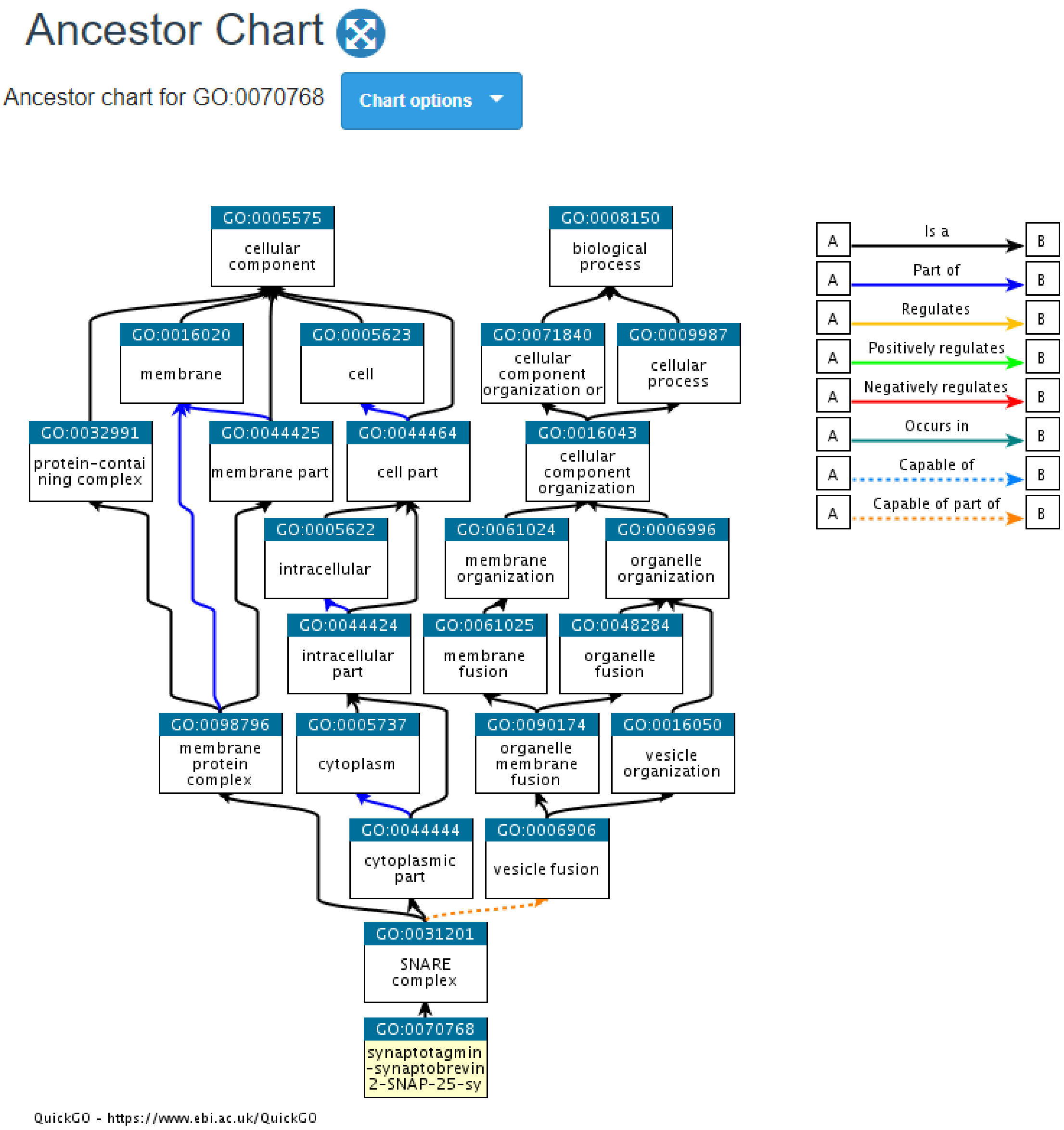


Figure 4: Quick GO tree with a SNARE complex as the child node at the bottom and ancestors above. Image: `https://www.ebi.ac.uk/QuickGO/term/GO:0070768`

Quick GO also pointed to The Rat Genome Database (RGD): `https://rgd.mcw.edu/`. The RGD is a potentially useful site for my research, but outside the scope of this report. In general, querying the GO Ontology Database does not appear to be best way to find the type of information that I usually look for.