Binding terms working group

- Mike Cherry
- Peter D'Eustachio
- Emily Dimmer
- David Hill
- Ben Hitz
- Jim Hu
- Ruth Lovering
- Debby Siegele

http://www.cardiovasculargeneontology.com

Objectives

1. What binding activities should be included in GO

2. The application of binding term usage in conjunction with column 16

3. The transfer of 'binding' term annotations via ISS/ISO

Action today

1. **To present the most recent proposed Guidelines to the GOC**
   - **Working group has agreed to these Guidelines**
   - **These Guidelines do not cover any controversial issues**
   - **Looking to GOC to ratify the current Guidelines, recognizing that further work is still required to include more controversial aspects**
2. **To discuss aspects of guidelines which have not been agreed by working group**
   - **GOC guidance is required on these issues**
     - Separate substrate binding from effector binding
     - Use of IC (inferred by curator) in binding term annotations
     - ISS/ISO and 'binding' terms
     - How much information to include in GO?
     - Boundary between binding and catalysis

Annotating cardiovascular genes

http://www.cardiovasculargeneontology.com

Proposed Guidelines (page 1 of 2)

1. Avoid Redundant Binding Relationships For Substrates/Products.
2. The purpose of the binding term guidelines is to minimize redundancy and duplication of GO term information.
3. An enzyme MUST bind all of the substrates and products of the reaction it catalyzes.
4. Similarly, a transporter MUST bind the molecules it transports.
5. Therefore, binding is implied by the molecular function GO term describing the activity of an enzyme or transporter.
6. Consequently, it is redundant to annotate an enzyme or transporter with GO binding terms for each of its substrate/products, and curators should avoid making such redundant annotations.
7. There will be some cases, however, where it is appropriate to annotate a binding relationship.
8. For example, published experiments may show that a gene product binds a non-hydrolyzable ATP analog, without demonstrating that it has ATPase activity.
9. In such a case, it would be appropriate to annotate to GO:0005524 ATP binding using an IDA evidence code.

Annotating cardiovascular genes

http://www.cardiovasculargeneontology.com

Proposed Guidelines (page 2 of 2)

10. The GO is committed to ‘annotating to the experiment’.
11. Therefore the curator should try to capture the specifics as much as feasible; use the binding term if the experiment shows binding directly, don’t use the binding term if the experiment shows catalysis, but not the specific binding activity.
12. Curators should use their judgment about when to associate an enzyme or transporter with a binding term for its substrates/products and also use their judgment to decide whether the interaction is physiologically relevant.
13. Curators should recognize that GO annotations should capture information relevant to the in vivo situation, not artificial substrates.

Other aspects of binding terms guidelines

- Separate substrate binding from effector binding
- Use of IC (inferred by curator) in binding term annotations
- ISS/ISO and 'binding' terms
- How much information to include in GO?
- Boundary between binding and catalysis

Substrate binding / effector binding

Should we distinguish substrate binding from effector binding?

- This would be hard to organize within the ontology
- Annotation to terms organized in this way would be difficult
  - would often rely on previous knowledge to decide whether a binding reaction was substrate or effector
  - newly characterized proteins
- Would we retrofit the current annotations to this new format?
- General feeling within the working group was not to do this.

Use of IC in binding term annotations

Should IC (inferred by curator) be used to annotate to a binding term where catalytic activity has been shown, but no binding assays were performed? As described in the guidelines:

- Binding is implied from catalytic/transport activity
- Curators should annotate to the experiment

If we annotate sometimes to binding but not always will the 'binding term' gene groups be ineffective because of inconsistent annotation?

- IEAs provide many annotations to binding terms
- IC could support the IEA statements, or could be used only when IEA evidence not available

The use of column 16 may influence this decision.

Will column 16 include substrates?

Would it be more appropriate to have IC to a binding term and then link to column 16 substrate list?

Or would it be better to have substrate list linked to enzyme, etc, function?

Can process terms have substrates in column 16?

Annotating cardiovascular genes

http://www.cardiovasculargeneontology.com

ISS/ISO and binding terms

What, if any, 'binding' term annotations can be transferred via ISS/ISO?

Is it of use to transfer binding term annotations to orthologous proteins or similar proteins?

GO terms such as ‘SH2 binding’ are informative even without an associated ‘with’ protein ID.

The GO term ‘protein binding’ is informative when associated ‘with’ a protein ID using IPI, however, when GO terms are transferred via ISS to similar proteins the original ‘with’ ID is lost. Therefore an ISS of ‘protein binding’ is of little value.

Cannot ISS an IC code, instead have to ISS via experimental evidence code and then IC. Example annotation: Adaptor protein IC based on ‘protein binding’. In this case ISS of ‘protein binding’ would enable IC of ‘adaptor protein’ to related sequence.

http://www.cardiovasculargeneontology.com

How much information included GO?

Should GO be used to accommodate catalogues of specific molecules and their behaviors?

If not by a core group of GO annotators then by collaborating groups?

What if there were 40 (or 100) distinct substrates identified, all physiologically relevant in some instance (or more likely, all tested in vitro and possibly physiologically relevant) is this full list going to be added to column 16?

As we accumulate more and more high-throughput data we are going to need a much better way of dealing with this.

For example, can we develop a way to annotate the "process" relationships with the various "molecular functions" and maybe cell types?
Eg: ‘protein A’ has ‘function B’ with ‘column 16 substrate’ when it is involved with ‘process C’.

Annotating cardiovascular genes

http://www.cardiovasculargeneontology.com

Boundary between binding and catalysis

Can we guide curator judgement on the interpretation of the boundary between binding and catalysis or is there a legitimate hybrid boundary region?

Would examples within the guidelines help with this?

Examples

Example 1: Rehemtulla et al. PMID: 8218226 describes the cleavage of pro-von Willebrand factor to mature von Willebrand factor by PCSK6, proprotein convertase subtilisin/kexin type 6.
PCSK6 annotated as:

- GO:0004252 serine-type endopeptidase activity
- GO:0051605 protein maturation by peptide bond cleavage
- GO:0070678 preprotein binding

With which GO terms would we add the protein ID for von Willebrand factor/VWF/P04275 in column 16?

Example 2: PMID: 17916063 describes the cleavage of synthetic peptides by SENP1, SUMO1/sentrin specific peptidase 1. The peptide sequences were derived from several different SUMO sequences.
SENP1 annotated as:
GO:0032183 SUMO binding
GO:0070139 SUMO-specific endopeptidase activity
How specific should GO annotations be?
Column 16 could clarify this with protein IDs for SUMO1 and SUMO2

Annotating cardiovascular genes

http://www.cardiovasculargeneontology.com