Discussions Monday Afternoon

--Notes to PRO Annual Meeting--

Judith Blake

April 26, 2010

Protein Ontology Meeting

Discussion Topics 1. ‘sites’ for PRO_MOD

- Sites: ‘coordinates’….might want to name as actually entities…how to define them, ‘motif-based’, does this belong in PRO? PRO needs it
- Need to separate ‘domains-PFAM” and ‘sites’. PRO needs to define sites that are modified. In both cases, you are naming ‘pieces of proteins’.
- PhosphoSite…keeps flanking region of the site..
- Keep PFAM site + interaction site to describe site information. “Site HMM, says Cathy, can cover the information.
- Another way: ‘corresponding to this particular site on this particular sequence’. There is currently represented a human and mouse ones. Right now, not a way to do this logically. Sites can be represented as parts.
- What representation has most clarity and information content for biologists. Visual vs. logical work. Logically, want to know where the same ‘parts’. For this need a ‘site’ thing.
- Question for where where does PFAM go? It’s heterogeneous…. InterPro is an integration service…relies on Panther, PFAM and other HMM entities…to map all these things
- InterPRO has also ProSITE information that deals with motifs.
- SO- tentatively this information belongs in SO.

Discussion topic 1a -

- Only propagate catalytic information in PRO
- UniProt propagates phosphorylation by inference, but this is indicated by ‘similarity’

General reaction from Drug/Pharma community

- Missing?
  - ‘Group of proteins represented by activity’, they aren’t necessarily in the same complex…so getting all ‘phosphodiesterases’ for example, they wouldn’t come up by protein family aggregations
- Drugs tested against ‘proteins’ as well as ‘complexes’. ChEBI has ‘classes of activities’ or ‘protein target’ classifications…’Protein Target’
- Action Item: GO to use PRO IDs instead of UniProt IDs? PRO-based cross-products. Also, column 17 allows a PRO-

ROLE resources

- ChEBI has ‘drug role’, other roles….needs a review…submit to OBO_Foundry for scrubbing.
- Need to understand relationship between ChEBI roles and
- Need to better understand ‘function’. Need good understanding of “‘function of the ‘organ’”
- Two kinds of function. Natural-evolved, Perturbed…defined…two kinds in terms of origins.

Needs for PRO

- PRO needs better intersection with GO_Panther Ref Genome Sets…
- Need to leverage sources….work together
- Same with Reactome…work together
- ‘Protein A’ contains ‘A’, A’’…can do this if UniProtKB doesn’t have isoform.

For this group, what aspects of PRO most important

- Microbial focus, Plant focus, missing
- Phenotype focus (microbial phenotypes?)
- BioLOG has assayed ~2,000 anologs…
- What would PRO provide?
  - IDs for proteins?
- From plant perspective, would use TAIR.
- Bring in PRO IDs for protein gp2protein sets for reference genomes
- What is minimum set for microbial groups…
  - Compare Quest set of 81 to Eisen set in recent publ.

Phenotypes

- ‘Phenotypes’ = ‘qualities’….Entity::Quality
- Alternative: Performance – 50 to 100 parameters. Particular genotype::phenotype under particular conditions. Measure of ‘performance’….growth rates, metabolic production, consumption of AA, metabolic studies
- DOE metabolic engineering…PRO not address microbial world – want to incorporate PRO IDs for relevant microbial species
- PRO captures ‘natural variation’…

Classifier sets…

- Often case, people want want define classes of proteins by something other than ‘evolution’. PRO should ‘accept’ those terms on request….
- Cathy…prefers not to create new PRO terms based on ‘other’ classifications. Better to define property by ‘attribute’, but this interfers with structure of the ontology.
- Classifier puts in the right place.
- Example: Kinase…’class’ = any protein/complex that has this function…so have superclass…
- Sounds very ‘mod-ish’; how is this different? While your editing…put the cross-product term and defined classes will be generated…have necessary and sufficient conditions…reasoner will define path to root.

‘Concept Buckets’

- So many ways to do this, may not be useful to ‘give id and classify’. Need instead tools that allow ‘ad hoc’ create groupings.

Of most interest..

- PRO-Form…most fundamental…
- Pro-Evo…not so useful to some as they use their own evolutionary classifiers to their purposes

Comments on Complexes

- Boundaries of complexes not addressed
- Current ‘complex’ vs ‘protein’ ok