BAO Reveal:

Assay Analysis Using the BioAssay Ontology and Open PHACTS

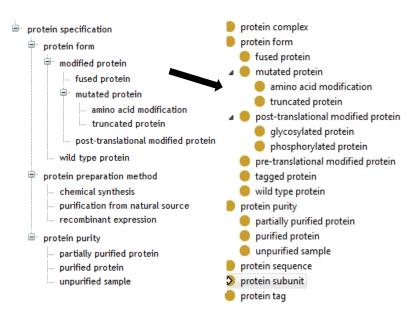
Introduction: Biochemical assay data is complex and highly variable. It contains attributes and terms that are not applicable to all assays, and uses controlled vocabularies inconsistently. Sometimes relevant data is entered in the wrong attribute, or appears in a different document altogether. The complexity and inconsistency of assay data creates challenges for those who wish to guery assay data for a given set of attributes, such as the technology chosen to conduct a given assay. We needed to develop and implement a system to address these challenges.

Methods

We have converted over 400 AstraZeneca primary highthroughput screening (HTS) assays into BAO format in order to evaluate whether this common model can improve project success analyses based on assay technologies, help us to better understand the impact of technology artifacts such as frequent hitters, and improve our ability to employ data mining methodologies against assay data. We have created static visualizations that combine our internal data with the annotated PubChem assays. Most recently, this project has created a dynamic interface, "BAO Reveal," for querying and visualizing BAO data.

Evaluation and Modification of BAO

Does the protein origin, such as the post-translational modification effect the technology?

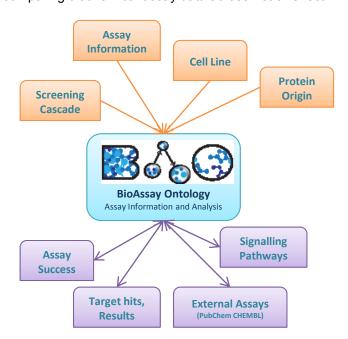


The BioAssay Ontology

The BioAssay Ontology (BAO) is an ontology developed by the University of Miami and extended by the Open PHACTS consortium, itself part of the European Union's Innovative Medicine Initiative (IMI). The purpose of the BAO is to standardize how we represent assay data. It incorporates many standard public ontologies, importing sections of the NCBI taxonomy, Uniprot, the Unit Ontology, the Ontology of Biomedical Investigation and the Gene Ontology, among others. More than 900 PubChem assays have been annotated according to BAO.

For more information, see: http://bioassayontology.org and PLoS One. 2012;7(11):e49198. doi: 10.1371/journal.pone.0049198. Epub 2012 Nov 14.

We have chosen to adopt the BAO format as the basis for comparing biochemical assay data across AstraZeneca.



Open PHACTS and BAO



The Open PHACTS project is an Innovative Medicines Initiative joint undertaking designed to reduce the barriers to drug discovery in industry, academia and for small businesses by creating an Open Pharmacological Space (OPS). It provides an API (http://dev.openphacts.org) and has fostered a number of (http://www.openphactsfoundation.org/apps.html). The Open PHACTS Foundation has been created to sustain efforts in the project (http://www.openphactsfoundation.org).

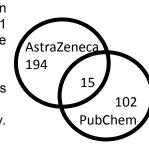
The BioAssay Ontology is one of a number of linked data standards adopted within OPS and is used to structure assay data from the CHEMBL for the framework.

Comparison Between AZ and PubChem HTS Assays

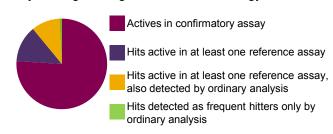
412 in-house HTS assays since 2005 have been annotated according to the BioAssay Ontology. The assay design and technology of the annotated assays were analyzed together with 239 primary assays from PubChem. The analyzed PubChem assays are biochemical assays, assays detected by luminescence and/or assays using GPCR targets.

From the annotated assays, 515 assays were using human targets and combined 311 different human targets were represented in the study.

15 of the in-house targets were also screened in at least one PubChem assay. Eight of these were GPCR targets.

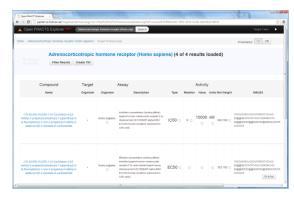


Frequent Hitter Analysis Enzyme target using fluorescent technology



Substances confirmed in concentration response assay

Frequent hitter analysis using three BioAssay ontology annotated reference assays identified almost twice as many frequent hitters as the ordinary frequent hitter analysis without the granularity of BAO



Biological Assays in Explorer 2

Simon Rakov, IS Solutions Engineer¹ Linda Zander Balderud, Postgraduate Fellow² Tom Plasterer, Semantic Framework Lead²

Contact: Simon Rakov

Simon.Rakov@astrazeneca.com

- 1 Research & Development Informatics, AstraZeneca
- 2 Chemistry Innovation Centre, Discovery Sciences, AstraZeneca

Results

Our frequent hitter analysis methodology has found twice as many frequently-hitting assays when assay data is structured using BAO than with previous methods that did not have the granularity of BAO. This has suggested improvements to the data capture process from these assays. The dynamic faceting features and linked biochemical information in BAO Reveal provide researchers ways to investigate the underlying causes of broad assay patterns. This will allow us to focus assay development efforts on the most promising approaches.

BAO Reveal facilitates identification of screening technologies used for similar targets and helps analyze the robustness of a specific assay technology for a biological target. It can identify screening data to confirm assay reproducibility, and also assist frequent hitter analysis. As a linked data application built on Open PHACTS methodologies and other semantic web standards, BAO Reveal is well positioned for exploitation in multiple directions by multiple communities.



Conclusions

BAO structured assays enabled Frequent Hitter Analysis of combined internal and external assay data annotated according to BioAssay Ontology.

BAO structured assays enabled exploration of internal and external assays side-by-side, which can be used for decision support during assay development and screening cascade design. This approach provides:

- Detailed information of annotated HTS assays and phenotypic screens according to BioAssay ontology
- Project cascade information including annotated AZ
- Analysis with external assays annotated according to BioAssay Ontology from PubChem, Chembl and GOstar (GVKBio)
- Compound activity data for tool compound identification from both internal and external data
- Customized analysis of a subset of in-house assays defined by BioAssay Ontology classes jointly with compound activity data from PubChem for statistical frequent hitter analysis.

