

BioAssay Research DataBase (BARD): A Transformative Platform for Knowledge Sharing



Andrea DeSouza¹, David Lahr¹, Joshua A. Bittker¹, Simon Chatwin¹, Steven Brudz¹, Noel Southall², Rajarshi Guha², Tudor Oprea³, Jeremy Yang³, Anna Waller³, Stephan Schürer⁴, Uma Vempati⁴, Mark Southern⁵, Eric Dawson⁶, Paul A. Clemons¹ and Thomas D.Y. Chung⁷

¹Center for the Science of Therapeutics, The Broad Institute, ²National Center for Advancing Translational Sciences, ³Center for Molecular Discovery, University of New Mexico Medical Center, ⁴Center for Computational Science, University of Miami, ⁵The Translational Research Institute, Scripps Research Institute, ⁶The Vanderbilt Specialized Chemistry Center for Accelerated Probe Development, ⁷Conrad Prebys Center for Chemical Genomics, Sanford Burnham Medical Research Institute

BARD: Chemical biology platform for generating novel hypotheses

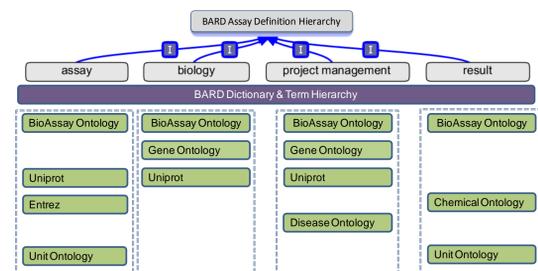
As the bio-pharmaceutical industry and academia partner to address challenges with therapeutics discovery and development through innovative translational approaches, an unmet need for a next-generation technology platform for small molecule science has emerged. The Broad Institute, the National Center for Advancing Translational Sciences, the University of New Mexico, Scripps Research Institute, Sanford Burnham Medical Research Institute, University of Miami, and Vanderbilt University through an NIH funded Molecular Libraries Program are collaborating to a) develop open source tools that include a practical standard that can serve as the currency for exchanging assay data b) provide high quality datasets with a focus on MLP data c) integrate novel and existing computational methods for data analyses and d) support complex workflows that facilitate collaboration. This project leverages other public resources including BioAssay Ontology, Gene Ontology, Disease Ontology, PubChem, Panther and others to advance small molecule science. As this project runs through its second year we plan to enhance existing BARD capabilities in cooperation with a team of academic and industry partners.

Community-wide resource of high quality chemical biology data



The data repository from publicly funded chemical-biology research efforts has been curated through a hierarchical term dictionary that integrates with:

- PubChem
- BioAssay Ontology
- Panther
- CARS
- Gene Ontology (GO)
- OMIM
- Unit Ontology
- UniProt
- Entrez (gene & protein)
- ATCC



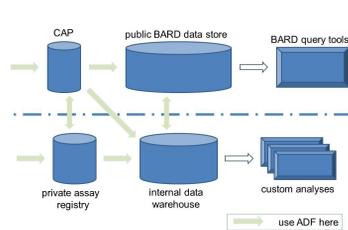
The use of authoritative, structured language enables queries and data mining not possible with unstructured, free-text assay descriptions or ambiguous terminology.

PubChem	BARD
Missing or fuzzy assay definitions, experiments and project concepts	Introduce assay definitions, experiments and projects
'Column header' centric with concentration details embedded	Result types and concentrations as experimental variables
Extensive use of unstructured text	Transition to structured use of common language

Assay Definition Standard (ADS)

The ADS requires :

- All assays annotated to a minimum standard
- Integration and extension of ontologies
- End-user enabled assay registration
- Assay, result and experiments represented using ADS
- Information in ADS is exchanged via the Assay Definition File

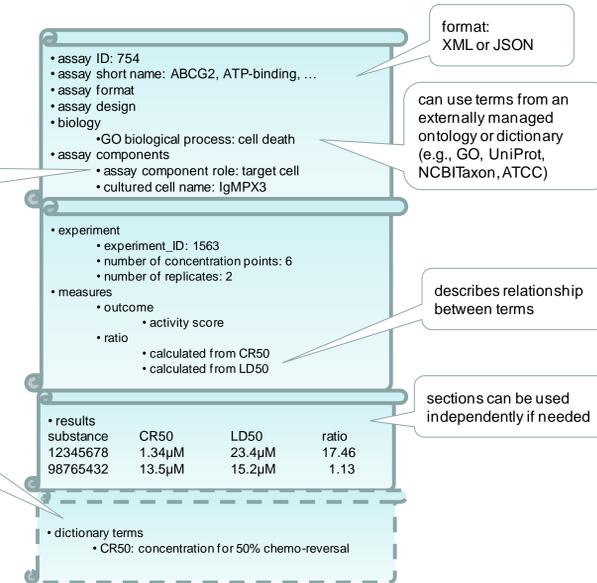


Benefits

- Tested – used for BARD interactions
- Extensible – uses industry-standard format
- Consistent meanings – controlled vocabulary for 90% of the data

terms from a controlled vocabulary

optional section to add new custom terms to the recipient's dictionary



format: XML or JSON

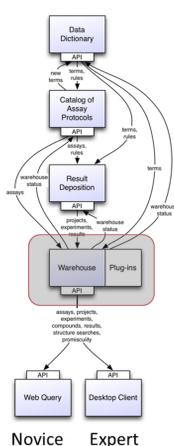
can use terms from an externally managed ontology or dictionary (e.g., GO, UniProt, NCBITaxon, ATCC)

describes relationship between terms

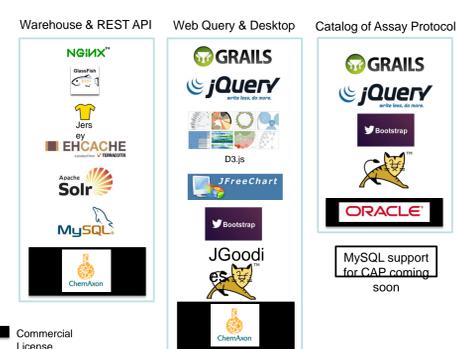
sections can be used independently if needed

Architecture, tools and components

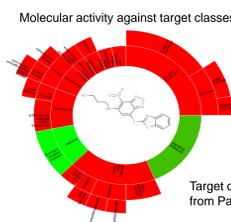
BARD architecture and data flows



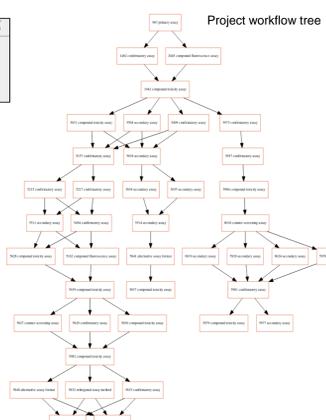
Open source components used when possible



Enable Hypothesis Generation



Target classifications from PantherDB



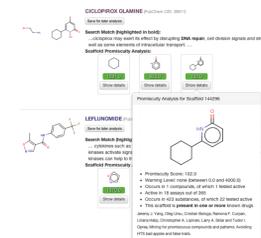
CAP and query tools support scientific project reporting

The 4 BARD modules, designed to be as Open Source as possible will be extended by the community

- Hierarchical dictionary of terms – manages the vocabulary
- Catalog of Assay Protocols –uses the dictionary to register assays and upload result data
- Warehouse – persistent storage of result data in a form that is fast and simple to query. Relies on the controlled terms from the dictionary for effective and accurate searching. Links to data from GO and other sources
- Query tools, Web Query and Desktop client – provide methods for novice and experienced users to browse and find the information they need

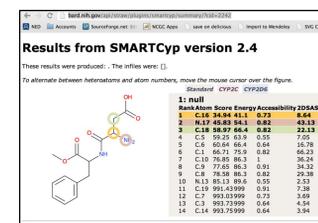
Analysis methods & plug-ins

BioActivity Data Associative Promiscuity Pattern Learning Engine (BADAPPLE)



Associations via scaffolds for chemical space navigation (University of New Mexico) Promiscuity analysis is dynamic based on up-to-date datasets in the system.

Predicting CYP450 Isoforms



Predicts CYP450 isoforms metabolism sites with 2D structures; Released under LGPL (Patrik Rydberg et al.)

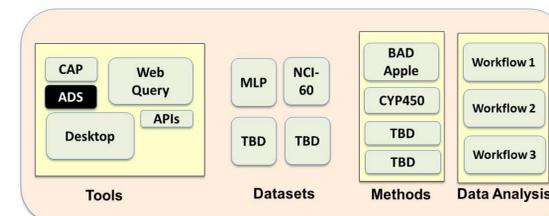
A next generation knowledge sharing ecosystem

BARD will enable novice and expert scientists to query, mine, and interpret public and private datasets. The future of BARD depends on sustaining the core components to:

- Manage the vocabulary and its accepted meaning
- Enable community contribution of plug-ins to enhance analytical capabilities
- Extend CAP and Warehouse for new data and assay types (e.g. images, gene expression profiles)
- Integrate/link to other data sources, both public and private

These capabilities will enable the research community to effectively generate new hypotheses and knowledge from high-quality chemical biology datasets. The future of BARD depends on the extension of this collaboration, and will benefit from private sector funding to sustain the ecosystem.

BARD as a Platform



Sustained Community Engagement

Apologies to the Bard¹

Friends, Romans, countrymen, lend me your data For I come to analyze the assays, not to praise them. The knowledge that men create lives after them, the data is oft interred with their bones



1. William Shakespeare (c1599) *Julius Caesar Act III: Sc 2*