Pushing Chemical Biology Data Through the Pipes

Architecting and Extending the BARD API

Rajarshi Guha, John Braisted, Ajit Jadhav, Dac-Trung Nguyen, Tyler Peryea, Noel Southall

September 9, 2014

Indianapolis, IN

Outline

Motivations for the BARD

Interacting with the BARD

Extensibility & Community

The BioAssay Research Database

- Originated from the NIH Molecular Libraries Program
- Motivated to make the bioassay data generated by the MLP more accessible and amenable to exploration and hypothesis generation
- Joint effort between NCGC, Broad, UNM, UMiami, Vanderbilt, Scripps, Burnham

Goals of the BARD

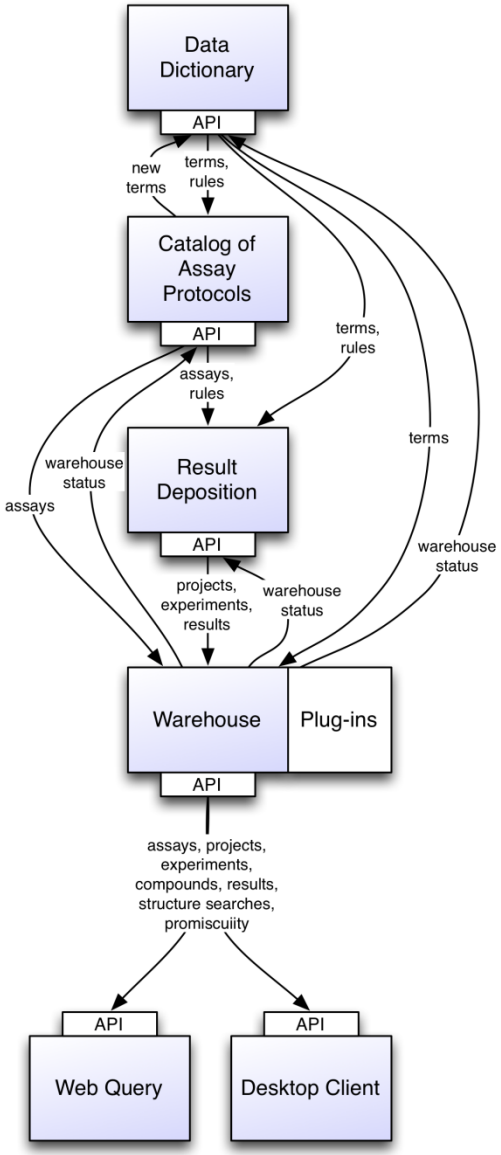
BARD's mission is to enable novice and expert scientists to effectively utilize MLP data to generate new hypotheses

- Developed as an **open-source**, industrial-strength platform to support **public** translational research
- Foster new methods to **interpret & analyze** chemical biology data
- Develop and adopt an **Assay Data Standard**
- **Enable co-location** of data and methods

Components of the BARD Platform

CAP, Data Dictionary, and Results Deposition Data model created & populated

CAP UI with View and basic editing

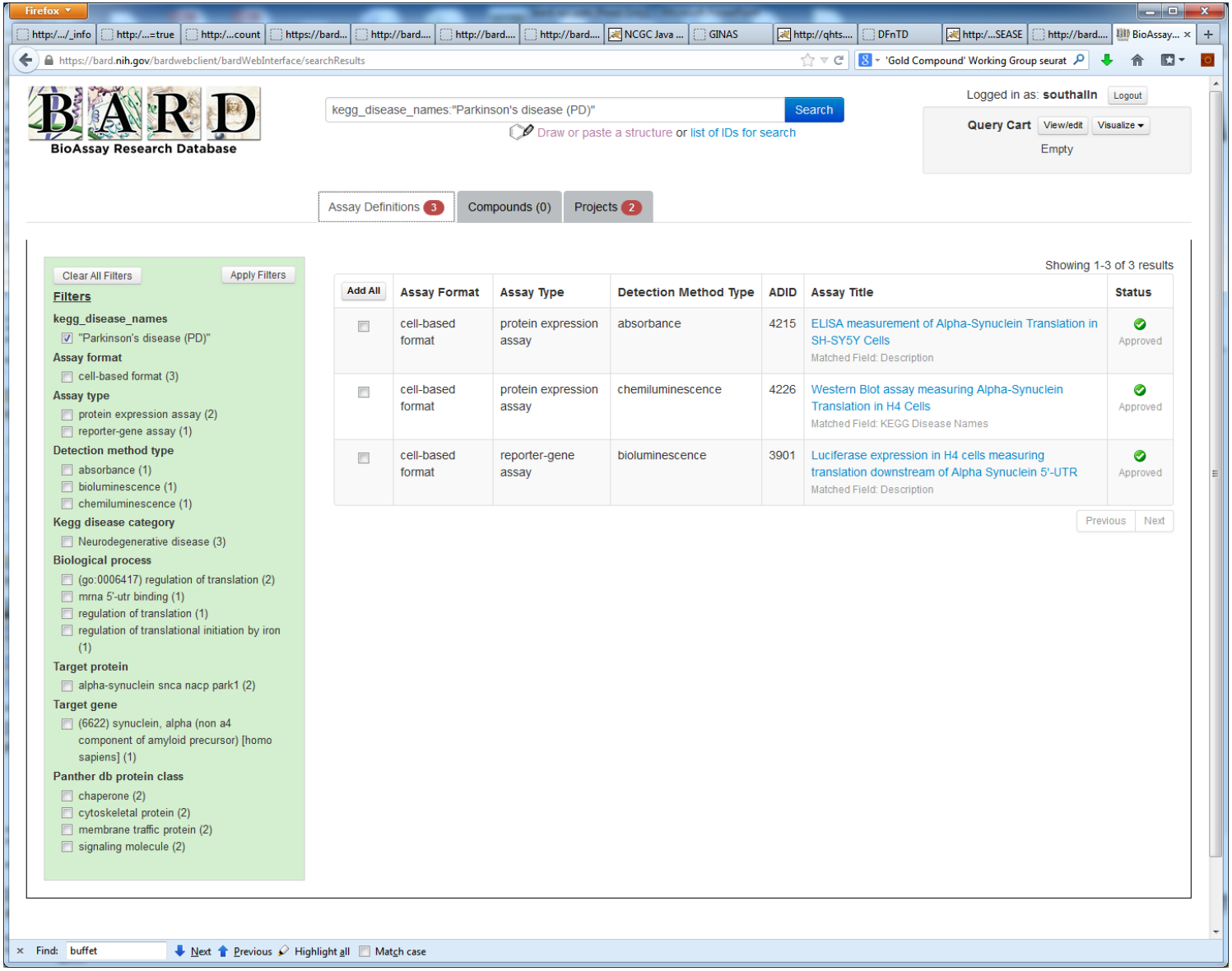


Dictionary defined as OWL using Protégé

Warehouse loaded with all PubChem AIDs and results

Warehouse loaded with GO terms, KEGG terms, and DrugBank annotations

Interacting with the BARD



Searching the BARD

- Full text search via Lucene/Solr
  - ***Key entry point for new users***
  - Search code runs queries in parallel
  - Fast faceting, auto suggest, filters
- Entities are linked manually via Solr schema
  - Allows us to pick up entities related to the query
  - Supply matching context
- Custom NCATS code for fast structure searches

The BARD API

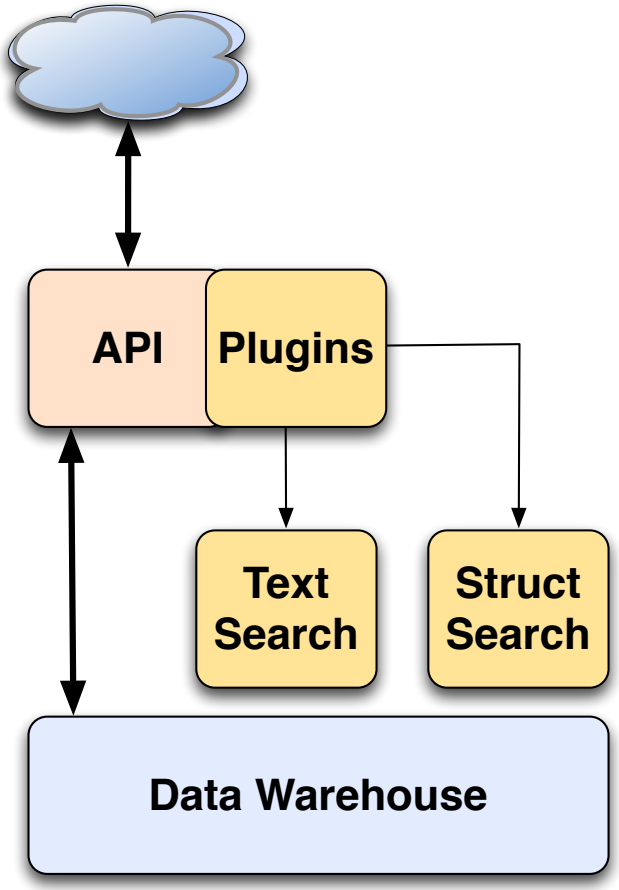
A RESTful application programming interface

- Access to individual & collections of entities
- Documented via the wiki
- Versioned
- Hit a URL, get a JSON response
  - Every language supports parsing JSON documents
  - Easy to inspect via the browser or REST clients

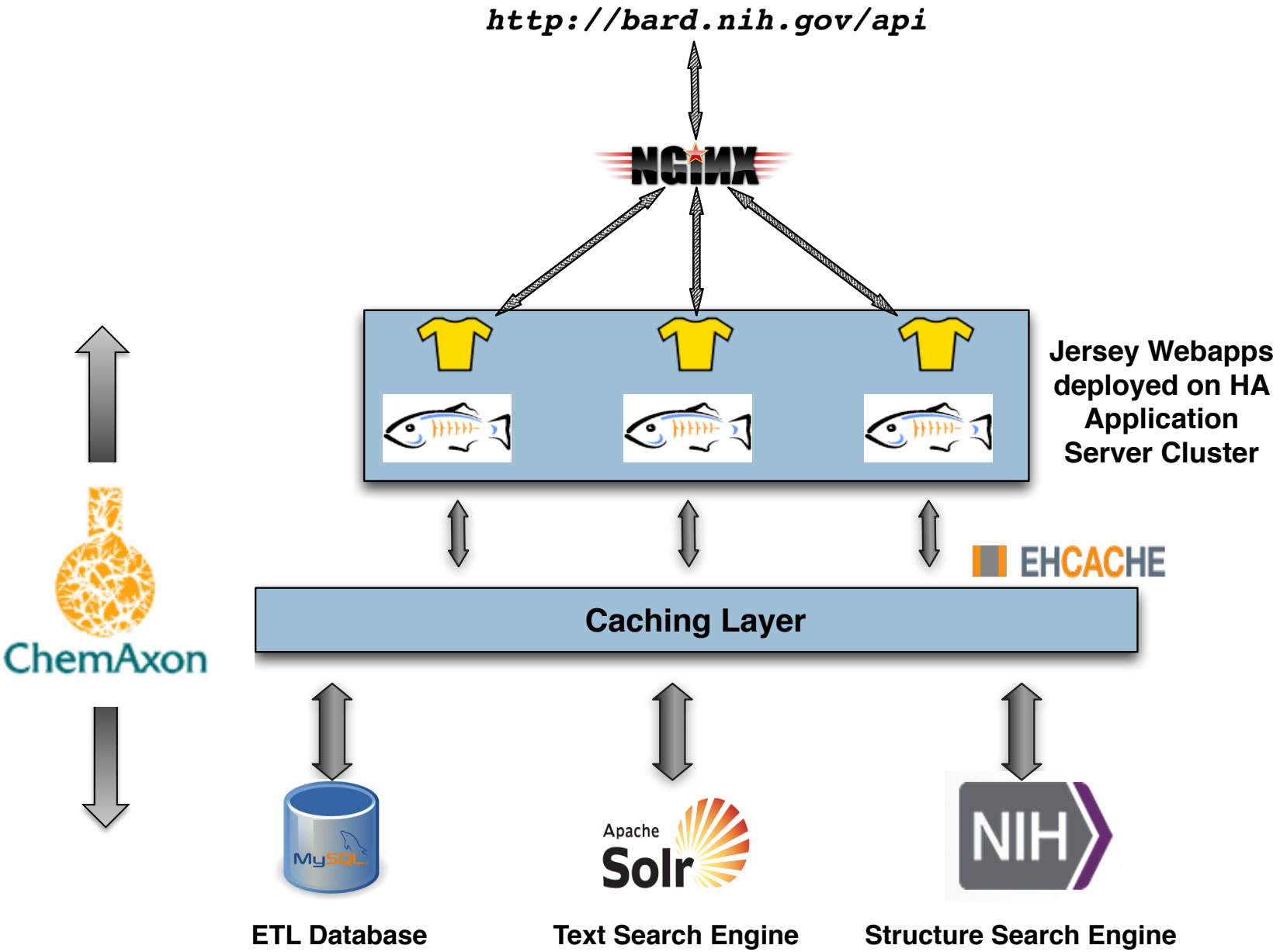
http://bard.nih.gov/api/v18/

API Architecture

- Java, read-only, deployed on Glassfish cluster
- Different functionality hosted in different containers
  - Maintenance, security
  - Stability
  - Performance



Open Source as Far as Possible



API Resources

- Covers many data types
- Each resource supports a variety of sub-resources
  - Usually linked to other resources
- Use `/_info` to see what sub-resources are available


```
{
  - collection: [
        "/assays",
        "/biology",
        "/cap",
        "/compounds",
        "/documents",
        "/etag",
        "/experiments",
        "/exptdata",
        "/projects",
        "/search",
        "/substances",
        "/target"
    ],
    link: null
}
```


```
Returns assay information

Available resources:
GET /assays/_count
GET /assays/{aid}/annotations
GET /assays/{aid}
POST /assays/
GET /assays/_info
GET /assays/etag/{etag}/facets
GET /assays/etag/{etag}
GET /assays/{aid}/targets
GET /assays/{aid}/documents
GET /assays/{aid}/projects
GET /assays/{aid}/experiments
GET /assays/{aid}/compounds
GET /assays/{aid}/substances
GET /assays/{aid}/experiments/{eid}
POST /assays/annotations
GET /assays/_schema
GET /assays/etag
POST /assays/etag
PUT /assays/etag/{etag}
GET /assays/etag/{etag}/_info
GET /assays/recent/{n}
/v1/assays/?filter=query_string[field]&
```

/assays/_info

API Resources

Entity	Count
/assays	990
/biology	1080
/cap	1942
/compounds	42,572,799
/documents	10,499
/experiments	1314
/exptdata	32,754,242
/projects	144
/substances	113,751,456

Data Warning

We're still in the process of data curation and QC so while you can access all BARD data, keep in mind that much of it will be undergoing review and curation and so may change

Summary Resources

- A number of entities have a `/summary` sub-resource (Compounds, Projects)
- Aggregates information, suitable for dashboards


```
{
    assay_count: 12,
  + experiments: […],
    description: "The primary qHTS data of human PKM2
    protocol for the PK isoforms M1, liver (L) and re
    assay that determined the activity of PK by coupl
    potency of the synthesized analog. While general
    further studied for potential cancer therapeutic
    cmpd_synthesis_count: -12,
    name: "Discovery of Lead Compounds which Modulate
  + probe_reports: […],
    cmpd_purchase_count: -12,
    depositor: null,
    experiment_count: 12,
  + targets: […],
  + probes: […]
}
```

JSON Responses

- All responses are currently JSON
- Entities can include other entities (recursively)
- JSON Schema is available via `/_schema`


```
{
    type: "object",
  - properties: {
      - bardAssayId: {
            type: "number"
        },
      - capAssayId: {
            type: "number"
        },
      - category: {
            type: "integer"
        },
      - summary: {
            type: "integer"
        },
      - assays: {
            type: "integer"
        },
      - classification: {
            type: "integer"
        },
```


/assays/_schema

Extending the API

- Concept of plugins
  - Expands the resource hierarchy
  - Has to be written in Java
- What can a plugin accept?
  - Anything
- What can a plugin provide?
  - Anything – plain text, XML, JSON, HTML, Flash
- Plugin manifest – describe available resources, argument types

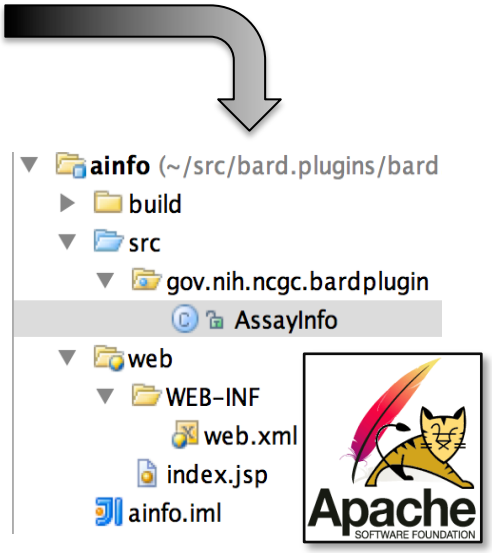
Plugin Development Workflow

```java
@Path("/ainfo")
public class AssayInfo implements IPlugin {
    private static final String VERSION = "1.1";

    /**
     * Get a description of the plugin.
     *
     * @return a description of the plugin.
     */
    @GET
    @Produces("text/plain")
    @Path("/_info")
    public String getDescription() {
        return "description of the plugin, including inputs, outputs a
    }

    @GET
    @Path("/{aid}/title")
    public Response getAssayTitle(@PathParam("aid") Long aid) {
        DBUtils db = new DBUtils();
        try {
            Assay assay = db.getAssayByAid(aid);
            db.closeConnection();
            return Response.ok(assay.getName(), MediaType.TEXT_PLAIN).
        } catch (SQLException e) {
            throw new WebApplicationException(500);
        }
    }
```

```yaml
{
    title: "BARD Assay Information",
    description: "A brief description will go here",
    version: "1.1",
    author: "Rajarshi Guha",
    authorEmail: "guhar@mail.nih.gov",
    maintainer: "Rajarshi Guha",
    maintainerEmail: "guhar@mail.nih.gov",
  - resources: [
      - {
            path: "/{aid}/title",
            mimetype: "text/plain",
            method: "GET",
          - args: [
              - {
                    arg: "aid",
                    argtype: "integer"
                }
            ]
        },
```



Plugins have to be deployable on the JVM

What Can a Plugin Expect?

- Direct access to the database via JDBC
- Faster access to REST API via co-localization
- No local storage in the BARD warehouse
  - But the plugin can use its own storage (such as an embedded database)
- Plugins have access to system JARs (e.g., XOM, JChem) but should bundle their own required dependencies

Plugin Validation

- Run a series of checks on a plugin before deployment
  - Catches manifest/resource errors
  - Doesn't check for correctness (not our job)
  - Examines plugin Java class
  - Examines final plugin package
- Run as a command line tool or from your code
- Could be made into an Ant/Maven plugin

What Plugins are Available?

- The plugin registry provides a list of deployed plugins
  - Path to the plugin
  - Version
  - Availability
- Long term goal is to have a plugin store

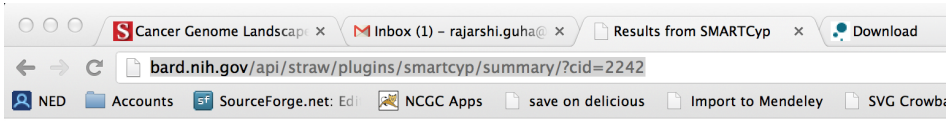

```
[
  - {
      path: "/plugins/smartcyp",
      title: "SMARTCyp",
      version: "1.1",
      available: true
    },
  - {
      path: "/plugins/whichcyp",
      title: "WhichCyp",
      version: "1.0",
      available: true
    },
  - {
      path: "/plugins/badapple",
      title: "BADAPPLE evidence-based promiscuity scores",
      version: "0.9beta",
      available: true
    },
  - {
      path: "/plugins/ssearch",
      title: "Structure Search Plugin",
      version: "1.1",
      available: true
    },
  - {
      path: "/plugins/ainfo",
      title: "BARD Assay Information",
      version: "1.1",
      available: true
    },
  - {
      path: "/plugins/csls",
      title: "Chemical Structure Lookup Service Wrapper",
      version: "1.1",
      available: true
    }
]
```


/plugins/registry/list

Exemplar Plugins

- Look at the plugin repository on Github
- Ranges from
  - trivial calls to external service
  - Incorporate external tools/programs
- Current set of plugins highlight plugin development
- Plugins let us move non-essential functionality out of the core API

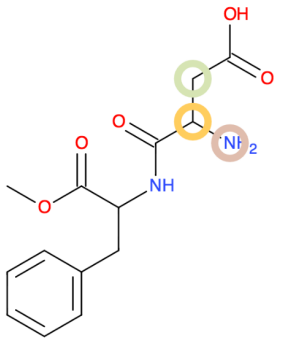
BARD- SMARTCyp



Results from SMARTCyp version 2.4

These results were produced: . The infiles were: [].

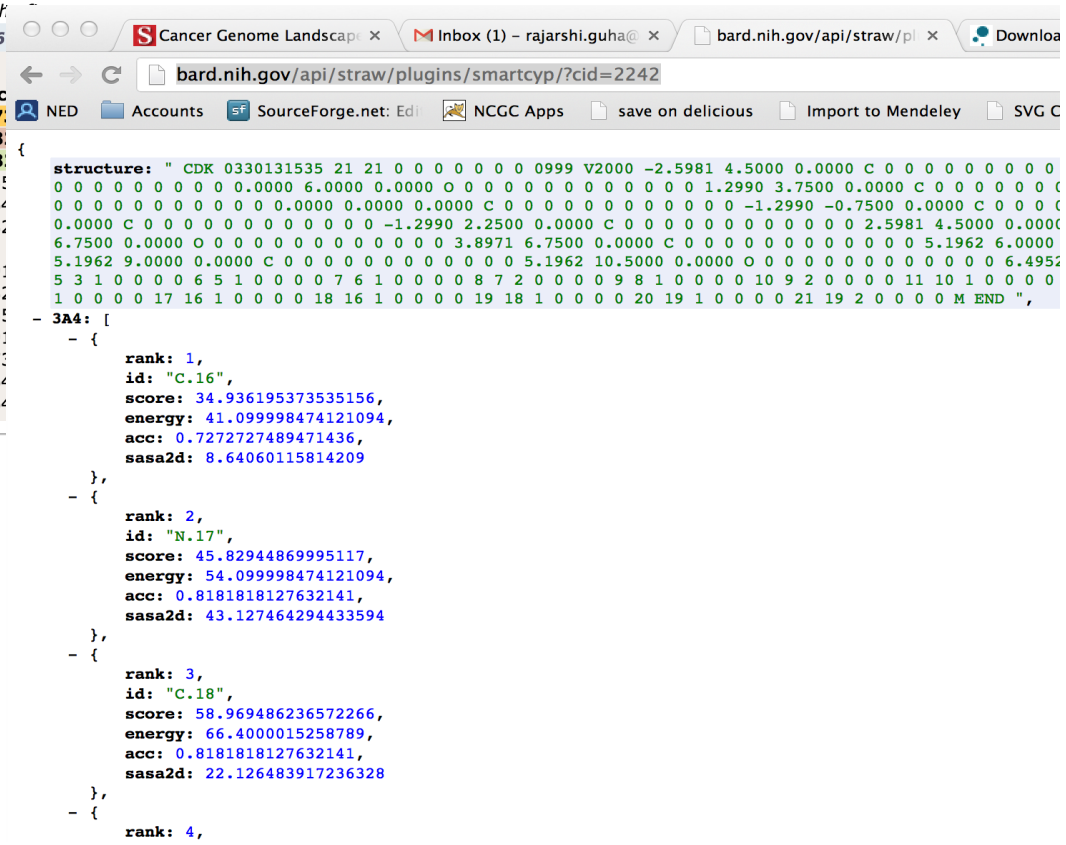
To alternate between heteroatoms and atom numbers, move the mouse cursor over th



Standard | CYP2C | CYP2D6

1: null

Rank	Atom	Score	Energy	Acc
1	C.16	34.94	41.1	0.7
2	N.17	45.83	54.1	0.8
3	C.18	58.97	66.4	0.8
4	C.5	59.25	63.9	0.5
5	C.6	60.64	66.4	0.6
6	C.1	66.71	75.9	0.8
7	C.10	76.85	86.3	1
8	C.9	77.65	86.3	0.9
9	C.8	78.58	86.3	0.8
10	N.13	85.13	89.6	0.5
11	C.19	991.43	999	0.9
12	C.7	993.03	999	0.7
13	C.3	993.73	999	0.6
14	C.14	993.75	999	0.6




```
{
   structure: " CDK 0330131535 21 21 0 0 0 0 0 0 0 0999 V2000 -2.5981 4.5000 0.0000 C 0 0 0 0 0 0 0 0 0
   0 0 0 0 0 0 0 0 0 0.0000 6.0000 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1.2990 3.7500 0.0000 C 0 0 0 0 0 0 0
   0 0 0 0 0 0 0 0 0 0 0.0000 0.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 -1.2990 -0.7500 0.0000 C 0 0 0 0
   0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 -1.2990 2.2500 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 2.5981 4.5000 0.0000
   6.7500 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 3.8971 6.7500 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 5.1962 6.0000
   5.1962 9.0000 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 5.1962 10.5000 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 6.4952
   5 3 1 0 0 0 0 6 5 1 0 0 0 0 7 6 1 0 0 0 0 8 7 2 0 0 0 0 9 8 1 0 0 0 0 10 9 2 0 0 0 0 11 10 1 0 0 0 0
   1 0 0 0 0 17 16 1 0 0 0 0 18 16 1 0 0 0 0 19 18 1 0 0 0 0 20 19 1 0 0 0 0 21 19 2 0 0 0 0 M END ",
- 3A4: [
   - {
         rank: 1,
         id: "C.16",
         score: 34.936195373535156,
         energy: 41.099998474121094,
         acc: 0.7272727489471436,
         sasa2d: 8.64060115814209
      },
   - {
         rank: 2,
         id: "N.17",
         score: 45.82944869995117,
         energy: 54.099998474121094,
         acc: 0.8181818127632141,
         sasa2d: 43.127464294433594
      },
   - {
         rank: 3,
         id: "C.18",
         score: 58.969486236572266,
         energy: 66.4000015258789,
         acc: 0.8181818127632141,
         sasa2d: 22.126483917236328
      },
   - {
         rank: 4,
```


P. Rydberg et al, http://www.farma.ku.dk/smartcyp/

More Than Just Data

BARD is not just a data store – it's a platform

- Seamlessly interact with users' preferred tools
- Allows the community to tailor it to their needs
- Serve as a meeting ground for experimental and computational methods
- Enhance collaboration opportunities

Links

http://bard.nih.gov

@AskTheBARD

API Source Code Repository

Plugin Source Code Repository

Development Wiki