



Microsoft® Research

# FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

The Path to Open Science with Illustrations  
from Computational Biology

Philip E. Bourne, University of California San Diego

# My Message

“Research is the insurance policy for the future”

Rick Rashid

*Open Research is the way to maximize  
the return on that policy*

# My Message

*I come to you as a domain scientist compelled by the belief that open science is critical to maximizing the return on the research investment and the means to embrace the maximum number of scientists worldwide*

*I commend the efforts of Microsoft to embrace open science within their business*

# My Hope

Some of you will see the unmet needs in achieving open science today and will be motivated to contribute to meet those needs

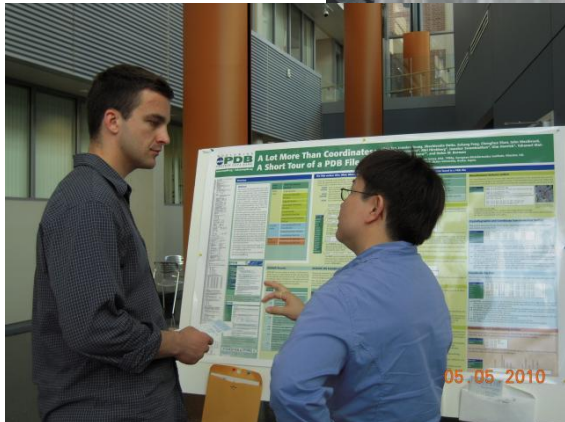
# Disclaimer

This is one persons biased view...

- Computational biologist
- Maintainer of a *well funded* major biological resource – The Protein Data Bank
- Co-founder of an *open access* journal – PLoS Computational Biology
- Contributor to *open source* archives
- Firm believer there must be a *business model*
- Co-founder of a for-profit science *dissemination* company



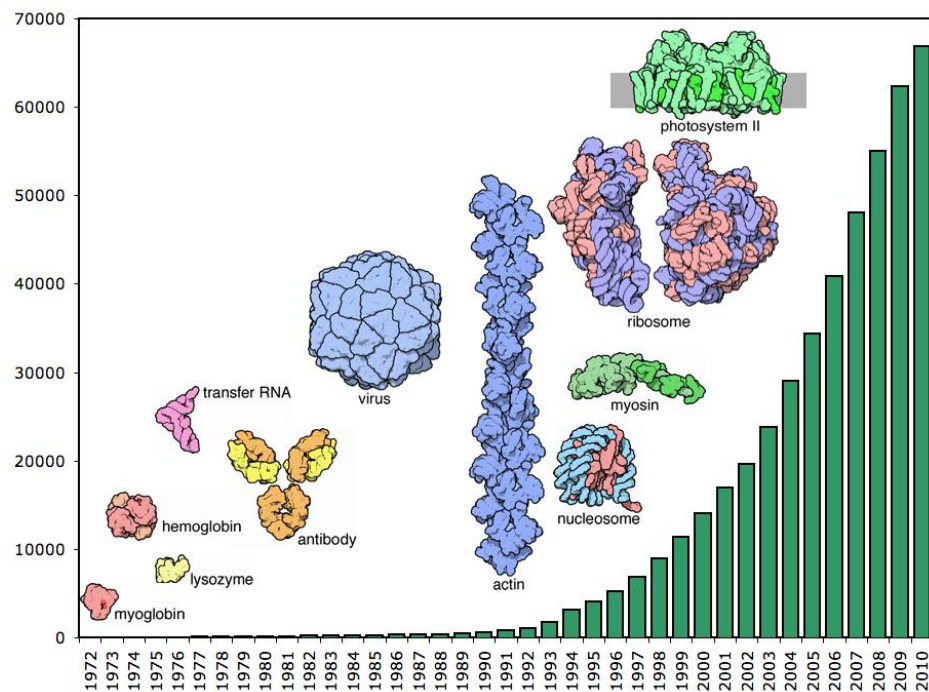
# What is the Protein Data Bank (PDB)?



- The single *community owned* worldwide repository containing structures of publically accessible biological macromolecules
- A resource used by ~ 200,000 individuals per month
- A resource distributing *worldwide* the equivalent to  $\frac{1}{4}$  the National Library of Congress each month
- A bicoastal resource
- 1TB

# What Does the PDB Tell Me About Open Science?

- *Is a biological database really different that a biological journal?*
- User base is broadening – outreach more important
- Constant demand for better performance
- Increasing use of Web services (SOAP and now RESTful)
- Uptake on the use of widgets has been slow
- Mobile use increasing
- Web 2.0 communications are in demand



# So What Am I Thinking We Need Going Forward?

- Science is increasingly digital – whether it be driven by observation or hypothesis
- There remain too many analog steps that make no sense these can be removed with human and machine consensus
- Doing science is but a complicated workflow
- Our current workflow tools are inadequate
- How we communicate science remains very much in the 16<sup>th</sup> century
- We already have many of the components we just need to put it together in a 21<sup>st</sup> century printing press



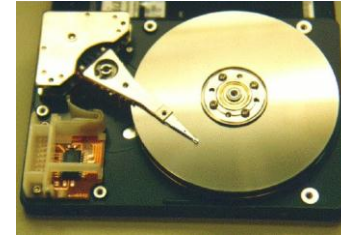
# The Game is Afoot – Pressure for Change is Growing



# Its Chaos Out There – Not so Much the Information But the Filtering



- PubMed contains ~21M entries (May 2011)
- ~100,000 papers indexed per month
- In Feb 2009:
  - 67,406,898 interactive searches were done
  - 92,216,786 entries were viewed



- 1330 databases reported in NAR 2011
- MetaBase <http://biodatabase.org> reports 2,651 entries edited 12,587 times

# Drivers of Change: The Scientific Publishing Process is Too Slow to Respond to a Crisis – Either Global or Personal



*By the time the paper is published  
we could all be dead*

 PLOS Currents | **Influenza**  
<http://currents.plos.org/>

## PLOS Currents: Influenza

**A moderated collection for rapid and open sharing of useful new scientific data, analyses, and ideas.**

The continuing public health burden due to influenza highlights the need for the rapid exchange of scientific results and ideas. PLOS Currents: Influenza aims to enable this exchange by providing an open-access online resource for immediate, open communication and discussion of new scientific data, analyses, and ideas in the field of influenza. All content is reviewed by a group of expert researchers.

[Link](#) [Citation](#) [Email](#) [Print](#) [Favorite](#) [Collect this page](#)

[Aims and Scope](#) | [Board of Reviewers](#) | [Author Guide](#) | [Reviewer Guide](#) | [FAQs](#) | [Creating a Google Account](#)

[Home](#) | [Help](#) | [Sign in](#)

Search Toolkit

Search this collection  Search all knols

[Edit this collection](#)

[Write a knol](#)



**PLOS**

The Public Library of Science (PLOS) is a non-profit open-access publisher, committed to making scientific and medical research literature a freely accessible public resource.

Article rating: ★★★★★

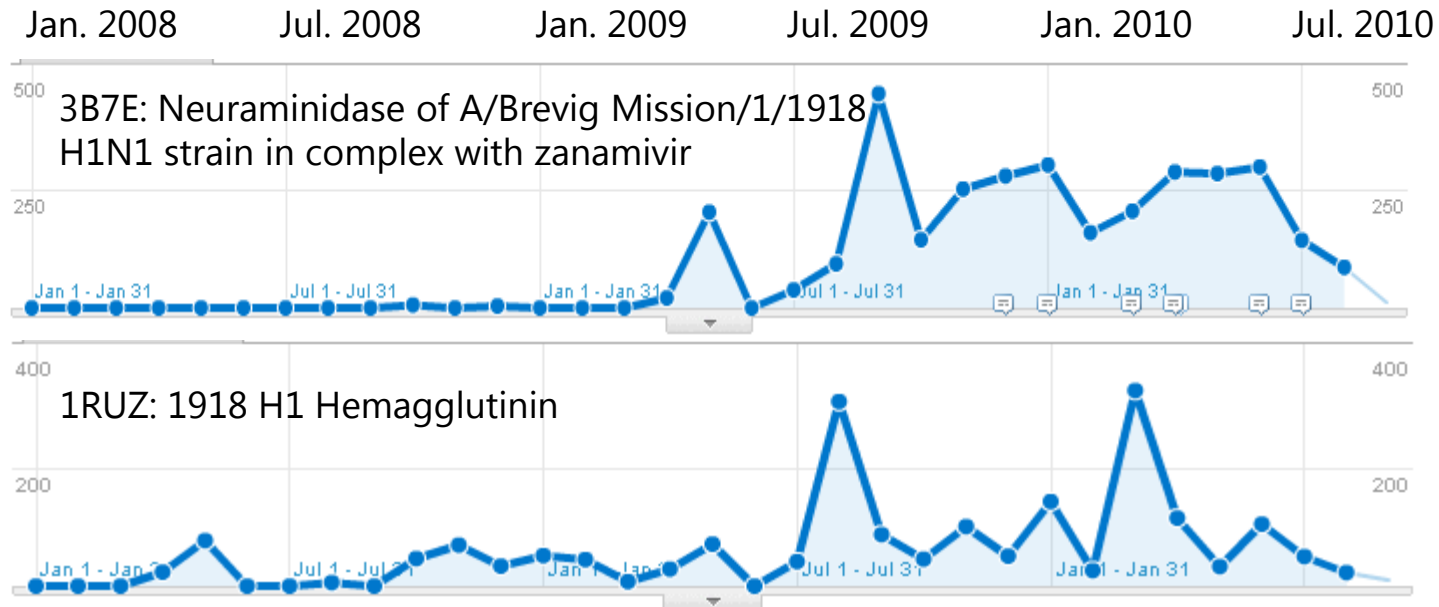
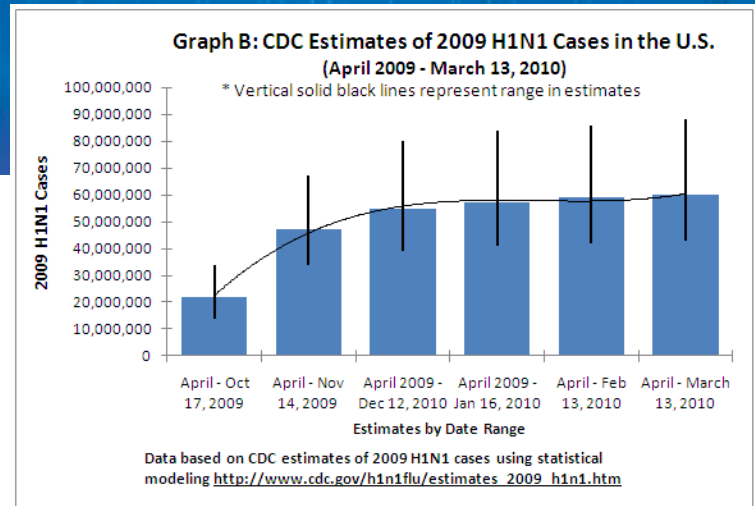
Your rating:

<http://knol.google.com/k/plos-currents-influenza#>

Drivers of Change: In a time of crisis the need for fast access to accurate data and any knowledge of that data are paramount



Structure Summary page activity for H1N1 Influenza related structures



\* [http://www.cdc.gov/h1n1flu/estimates/April\\_March\\_13.htm](http://www.cdc.gov/h1n1flu/estimates/April_March_13.htm)

# If that is not enough...

- For some people the scientific process may be too slow to save their life



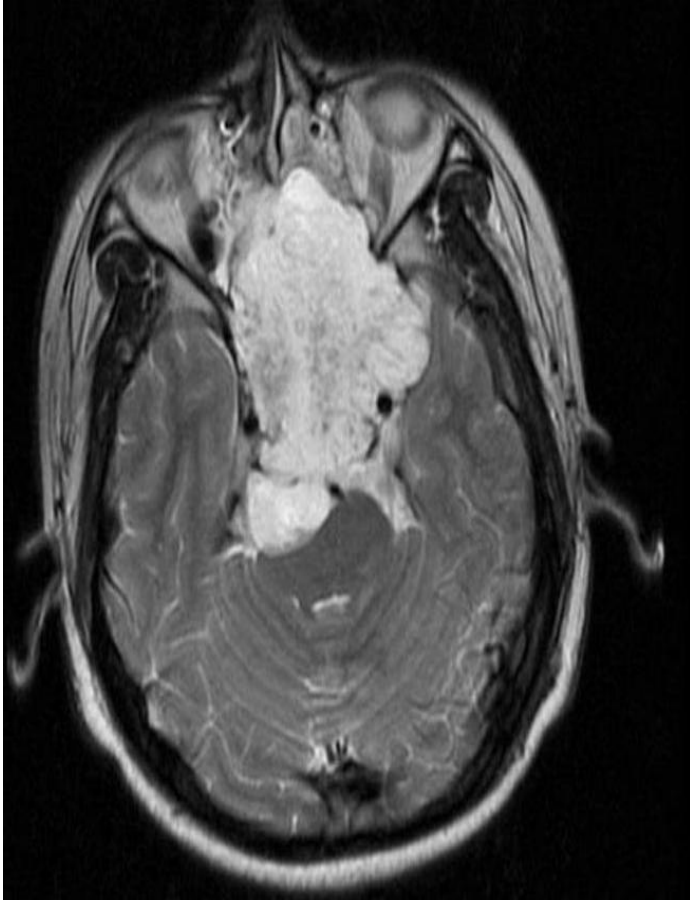
# Josh Sommer – A Remarkable Young Man Co-founder & Executive Director the Chordoma Foundation

## The Race Begins



**Diagnosed with chordoma in 2006 during freshman year at Duke University**

# Chordoma

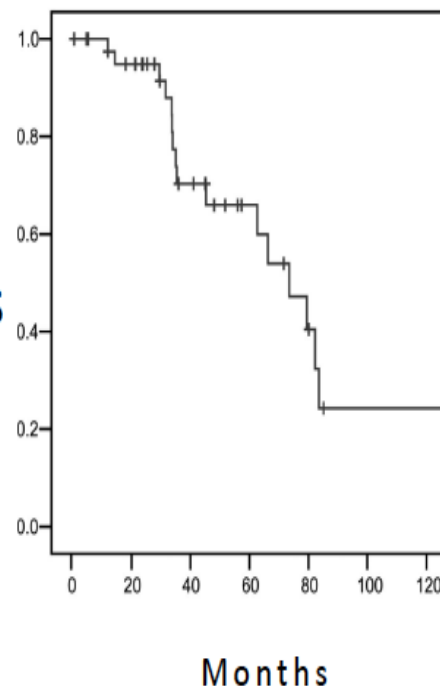


- A rare form of brain cancer
- No known drugs
- Treatment – surgical resection followed by intense radiation therapy

<http://upload.wikimedia.org/wikipedia/commons/2/2b/Chordoma.JPG>

## Unacceptable Statistics

- High rate of recurrence and numerous surgeries
- No effective chemotherapy
- Significant morbidity -Patients often live for years with disease and disability
- **7 year** average survival
- **20-30%** cure rate



# Into the Lab



**Dr. Michael Kelley was the only NIH-funded chordoma researcher... and he happened to be an oncologist at Duke**



# Assembling a Team & Developing a Plan



## Sponsors



National Human  
Genome Research  
Institute



Office of  
Rare  
Diseases



NATIONAL INSTITUTE OF  
NEUROLOGICAL  
DISORDERS AND STROKE



UPMC

University of Pittsburgh  
Medical Center



THE CENTER FOR  
CRANIAL BASE  
SURGERY



MASSACHUSETTS  
GENERAL HOSPITAL  
Department of Orthopaedic Surgery



**SYNTHES**

Spine

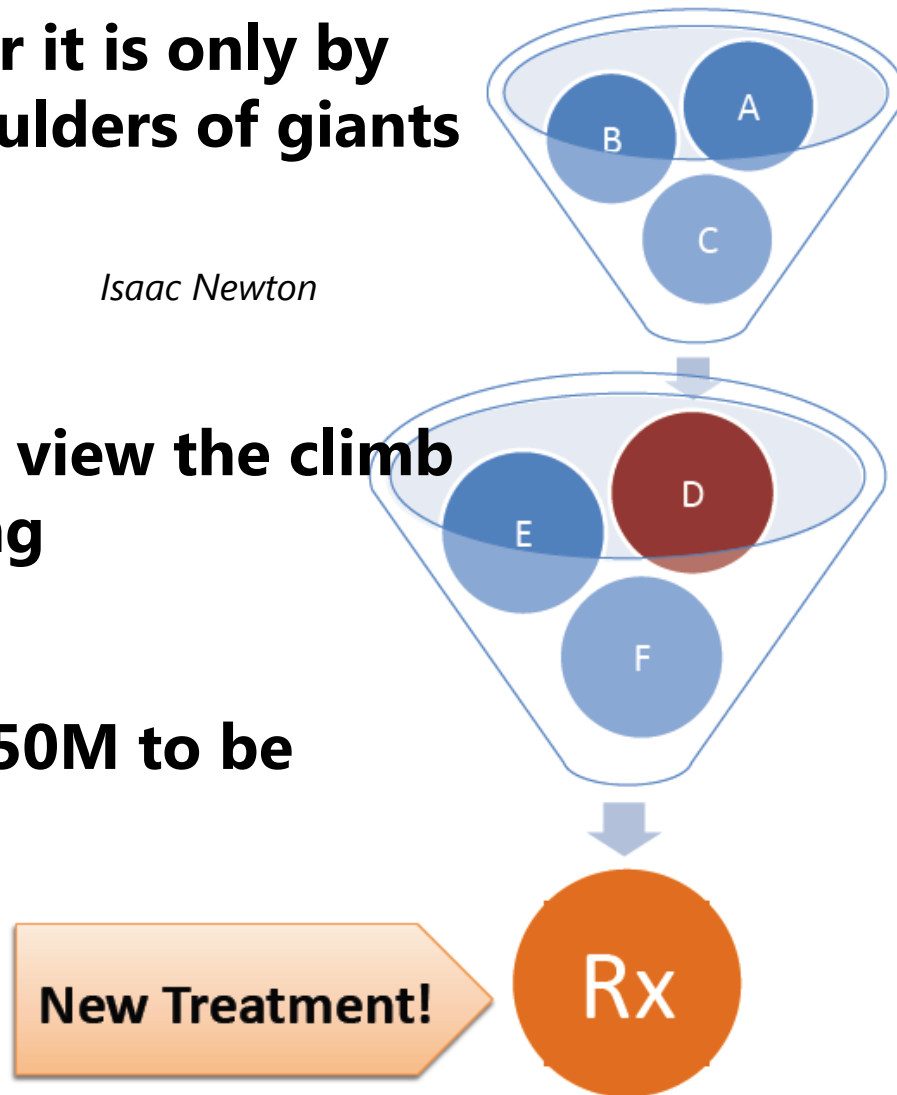


**If I have seen further it is only by standing on the shoulders of giants**

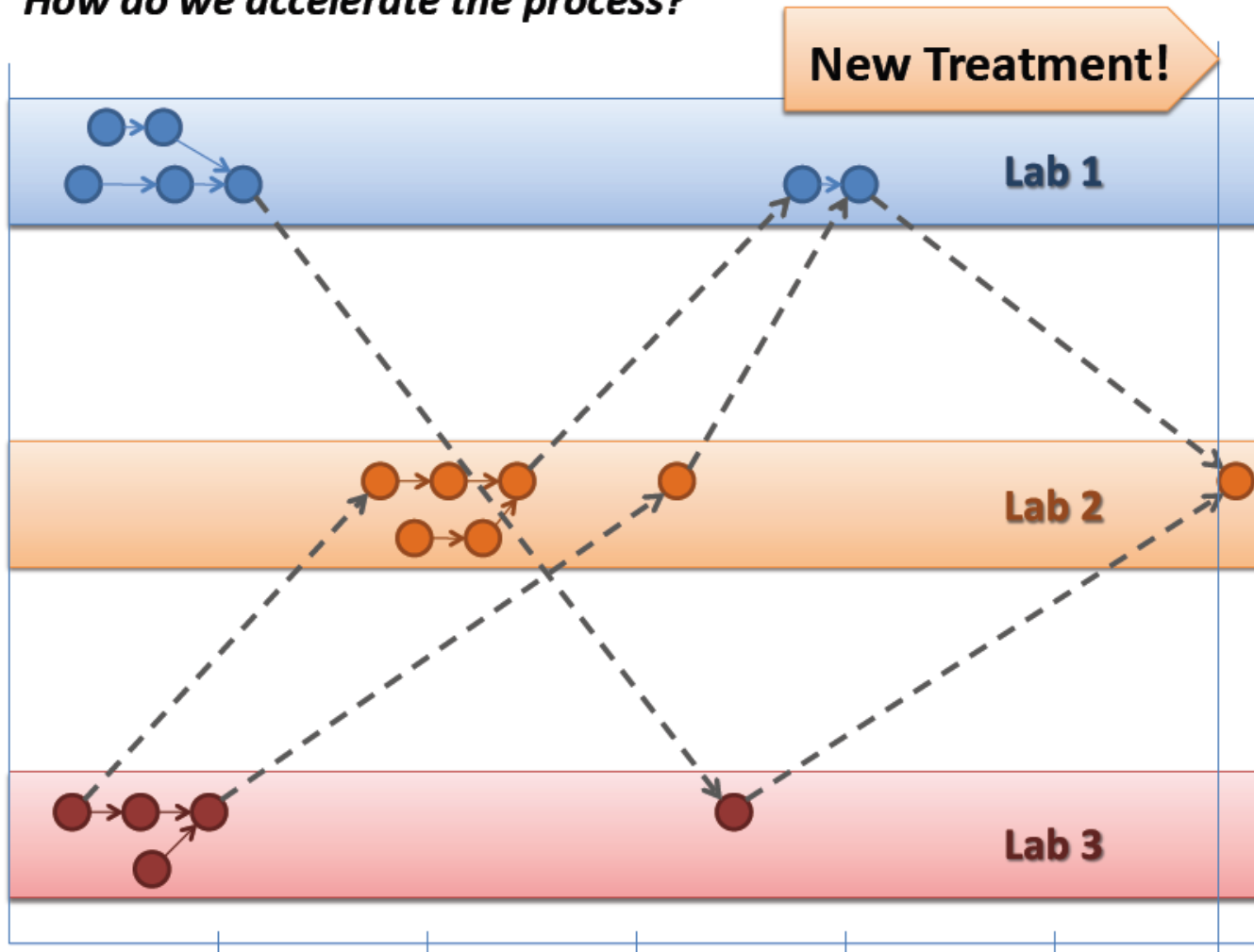
*Isaac Newton*

**From Josh's point of view the climb up just takes too long**

**> 15 years and > \$850M to be more precise**

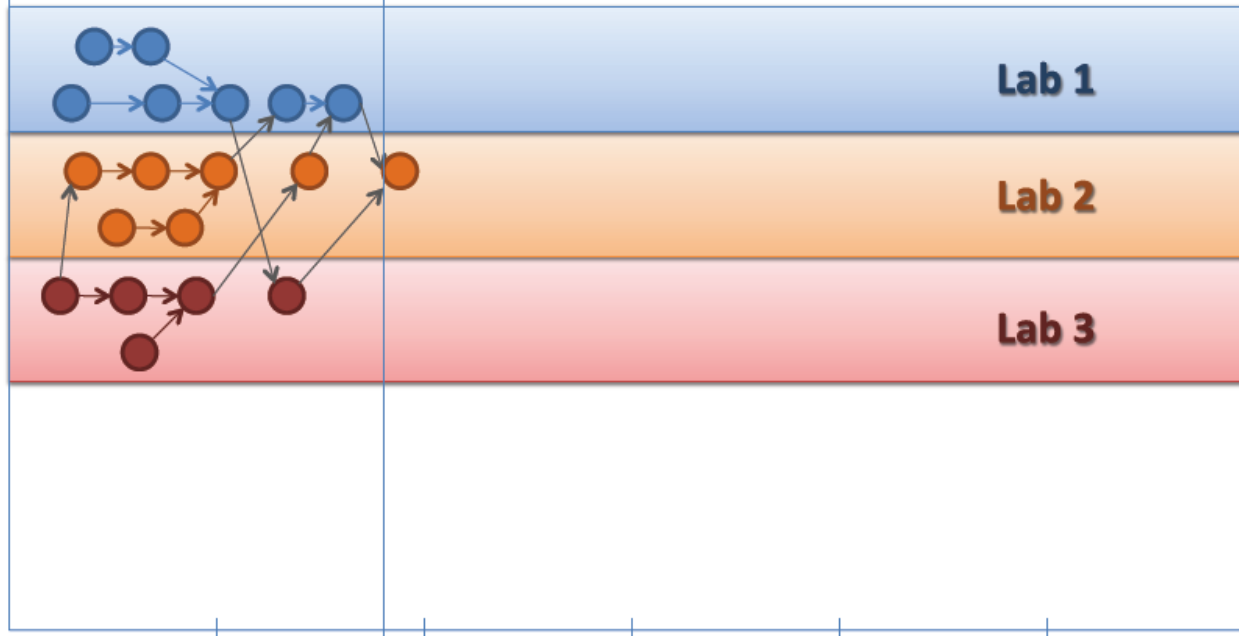


**How do we accelerate the process?**



**New Treatment!**

**The Power of the Commons:**  
To dramatically decrease the length  
of inter-lab knowledge turns and  
optimize the pace of discovery



# Time is of the Essence



[http://fora.tv/2010/04/23/Sage\\_Commons\\_Josh\\_Sommer\\_Chordoma\\_Foundation](http://fora.tv/2010/04/23/Sage_Commons_Josh_Sommer_Chordoma_Foundation)

Now we are all hopefully motivated let us break this down to what actually needs to be done in my opinion

Here are a few *big* things ...

.. and a few very little things we are contributing by way of example...



# A Big Thing: The Academic Reward System Must Change

**The Right Thing To Do**

**Reward**

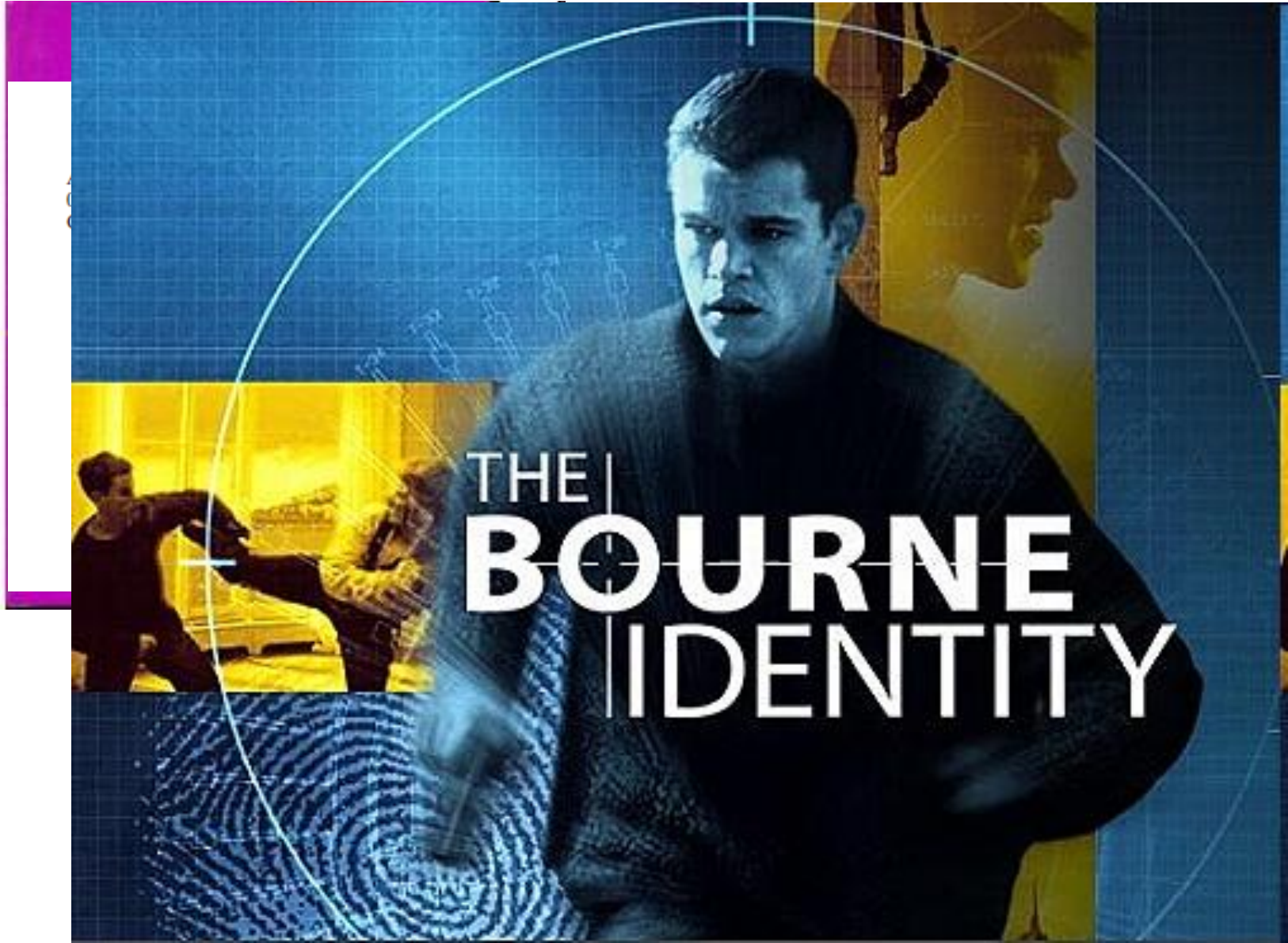


Papers  
Grants

Data availability  
Reviews  
Provision of metadata  
Open access  
Curation  
Alternative forms of dissemination  
New tools



# The Reward System Must Change: Prerequisite



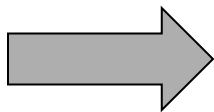
# Prerequisite: Ability to Attribute



- ORCHID - It is DOI's for people
- Some scientists will resist



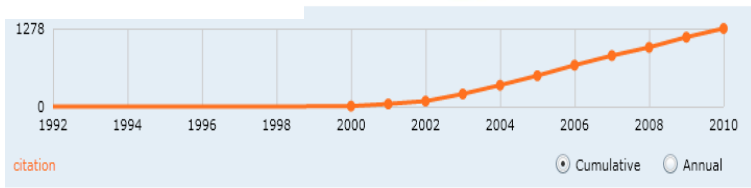
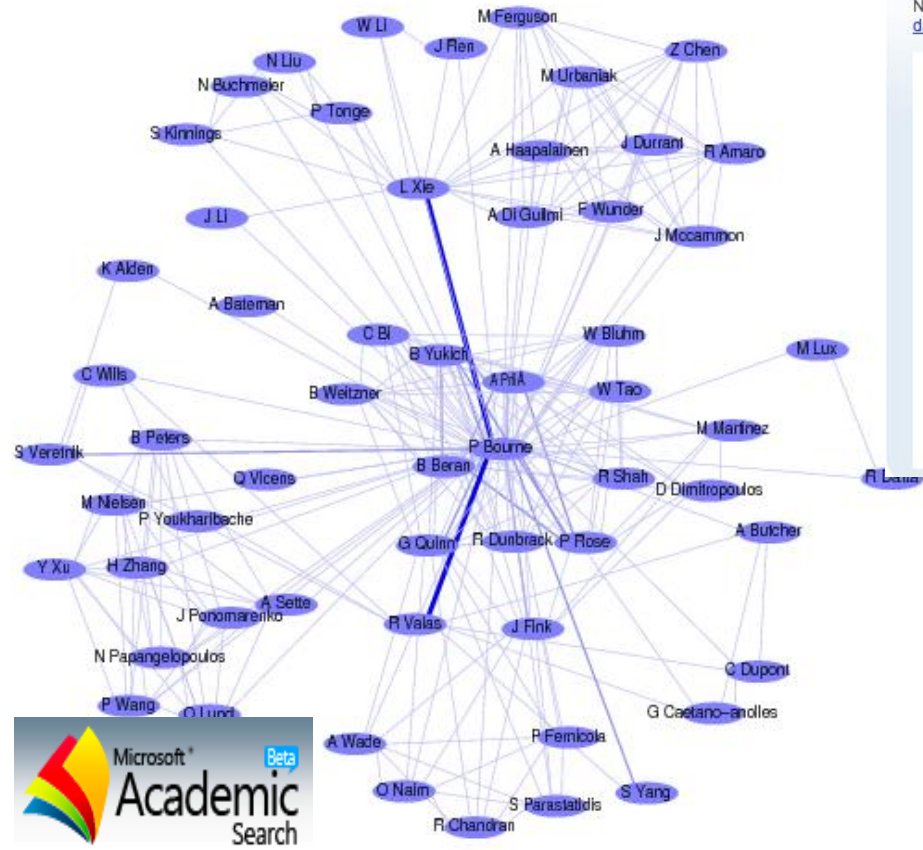
## Three "Layers" Of Licenses





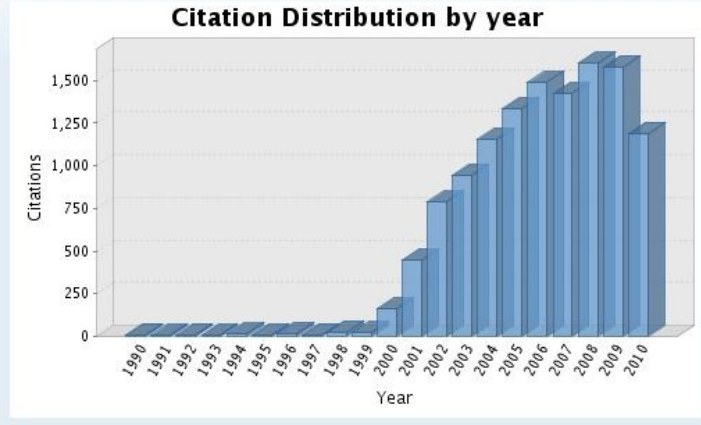
# With Attribution and Licensing *Accurate* Tools Emerge

<http://pubnet.gersteinlab.org/>



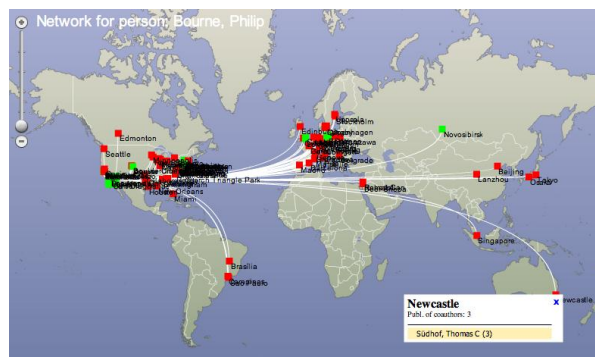
## My Publications: Citation Metrics

This graph shows the number of times the articles on the publication list have been cited in each of the last 20 years. Note: Only articles from ISI Web of Knowledge with citation data are included in the calculations. [More information about these data.](#)



Total Articles in Publication List: **153**  
 Articles With Citation Data: **152**  
 Sum of the Times Cited: **12281**  
 Average Citations per Article: **80.80**  
 h-index: **30**  
 Last Updated: **10/11/2010 14:17 GMT**

<http://www.researcherid.com/>



<http://www.biomedexperts.com>

# From Accurate Tools Come New Metrics for Success

PERSPECTIVE

OPEN ACCESS

## I Am Not a Scientist, I Am a Number

Download: PDF | Citation | XML

Print article

Download Article

Article Metrics Related Content

We define for each scientist a Scholar Factor (SF) as follows:

Philip E. Bourne\*, J. Lynn Fink

Skaggs School of Pharmacy and Pharmaceutical Sciences  
San Diego, La Jolla, California, United States of America

**Citation:** Bourne PE, Fink JL (2008) I Am Not a Scientist, I Am a Number. PLoS Comput Biol 4(12): e100247  
doi:10.1371/journal.pcbi.1000247

**Editor:** Barbara Bryant, Constellation Pharmaceuticals

**Published:** December 26, 2008

**Copyright:** © 2008 Bourne, Fink. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** None received.

**Competing interests:** PEB is a cofounder of Veritas and Veritas is a subsidiary of Veritas and is currently in negotiations with Thomson Reuters as part of a collaborative venture, which at this point in time has no relationship to ResearcherID.

\* E-mail: [bourne@sdsc.edu](mailto:bourne@sdsc.edu)

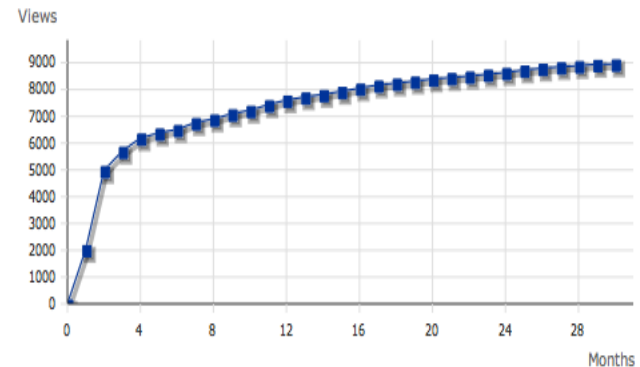
$$\begin{aligned} SF = & (\text{H Factor}) + \\ & (\text{Grant/Manuscript Review Factor}/20) + \\ & (\text{Annotations/Software/Datasets Factor}/5) \\ & + (\text{Web Factor}/50) \end{aligned}$$

[Google Scholar](#)  
[PubMed](#)

More

Share this Article

Cumulative Views from Dec 26, 2008 (publication date) - May 16, 2011\*



# Accuracy Must Appeal to Scientists Surely?

- So why do we persist with the journal impact factor?
- Why do we not educate those that review us since only 1-2 of a committee of 6 or more actually know what we do and the rest fall back on false metrics

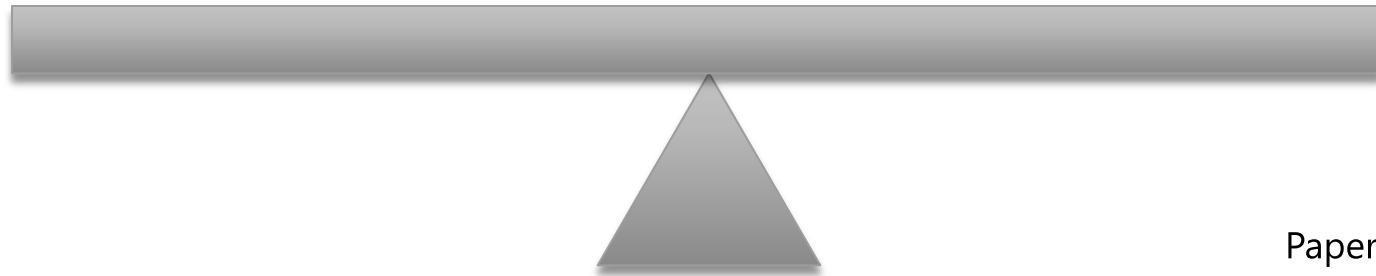
P.E. Bourne 2011 Ten Simple Rules for Getting Ahead  
as a Computational Biologist in Academia. *PLoS Comp. Biol.* 7(1) e1002001.



# The Reward System Must Change

**The Right Thing To Do**

**Reward**



Papers  
Grants

**Data availability**

- Reviews
- Provision of metadata
- Open access
- Curation
- Alternative forms of dissemination
- New tools



# Measure Data Contributions – A Step Towards Realizing the 4<sup>th</sup> Paradigm

- Data resources have an obligation to unify metadata availability to provide provenance information
- How can this happen?
- Scientists within one domain agree on a way

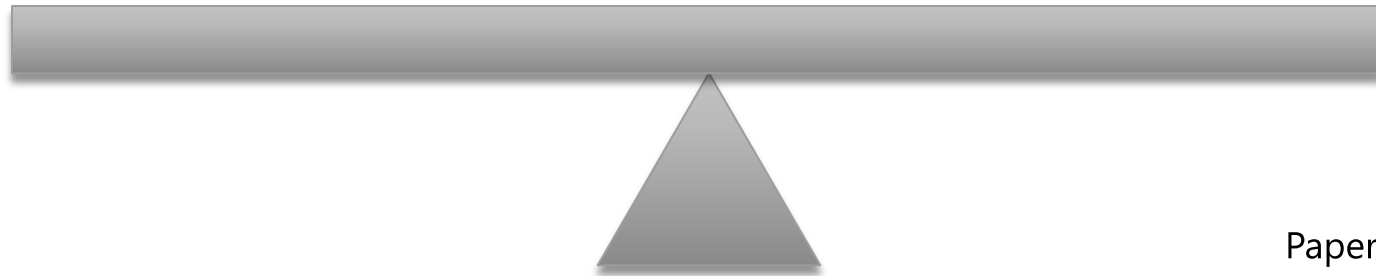
The screenshot shows the RCSB PDB website interface. At the top, it displays the RCSB PDB logo and the text 'A MEMBER OF THE CPDB'. Below this, it states 'As of Tuesday May 10, 2011 at 5 PM PDT there are 73009 Structures' and provides links for 'PDB Statistics'. The main navigation bar includes 'Contact Us | Print' and a search input field with a dropdown menu for 'PDB ID or Text'. A 'PDB ID lookup or Text search of the complete structure' button is also visible. The page is divided into several sections: 'MyPDB' (login/register), 'Help' (launch help system, display settings, video tutorials, glossary, FAQ), 'Tools' (download, compare, FTP, file formats, RESTful, SOAP, widgets), 'Search' (advanced search, latest release, new structure, papers, sequence search, chemical, components, unreleased entries, browse database, histograms), and 'PDB-101' (structural view of biology, understanding PDB data, molecule of the month). The 'Featured Molecules' section highlights 'Cytochrome bc1' as the 'Molecule of the Month', describing its role in energy production. Below this, it features 'AZA Adenosine Receptor' as a 'Protein Structure Initiative Featured Molecule'. The 'Latest Structures' section is partially visible at the bottom. On the right side, there are several sidebar widgets: 'New Features', 'RCSB PDB News' (with a 'PDB-101' logo circled in red), 'wwPDB News', and 'Special Symposium Celebrating the 40th Anniversary of the DRR'.

We are committed to ORCID  
(we already support DOIs) and  
to adopting such a standard

# The Reward System Must Change

**The Right Thing To Do**

**Reward**



Papers  
Grants

Reviews  
Provision of metadata  
Open access  
Curation  
Alternative forms of dissemination  
New tools



Data availability

# Scientists: Open Access (Biomedical Sciences Only)

"I just submitted this paper yesterday, if anyone is interested in a copy email me afterwards"

*Hot-shot young assistant professor*

Why?

- Web 2.0 / Social media?
- Stuff is happening quickly?
- Generational?
- Influence of other fields?
- Government pressure?



- PDB 10 years ago / today
- Release immed 10/50
  - Hold till pub 60/45
  - Hold 1 year 30/5

# Publishers: Open Access (Biomedical Sciences Only)

- In private most publishers would admit it is a done deal – only the crème will be able to sustain a pay wall
- Why?
  - Funds available
  - Seems the right thing
  - Agencies mandating it in any case
  - Public awareness of content

*The new model is leveraging the content*



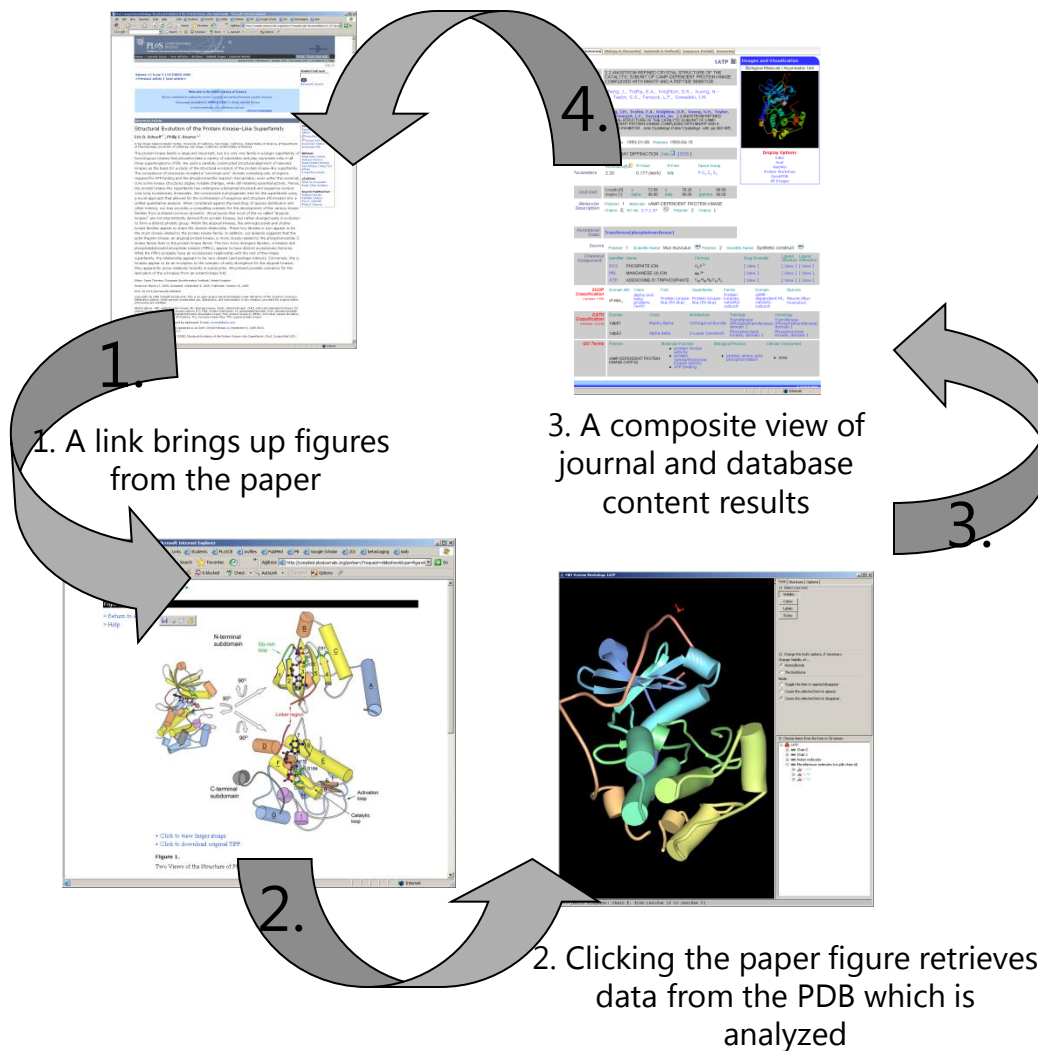
# The Knowledge and Data Cycle

0. Full text of PLoS papers stored in a database

4. The composite view has links to pertinent blocks of literature text and back to the PDB

## We Need Data and Knowledge About That Data to Interoperate

1. User clicks on *content*
2. Metadata and webservice to data provide an *interactive* view that can be *annotated*
3. Selecting features provides a data/knowledge *mashup*
4. Analysis leads to new content I can *share*



# We Have a Long Way to Go, But ...

- Crowd annotation works under selective circumstances – again related to reward
- We have some instances of data knowledge interoperability
- We have some instances of interactive papers
- We have a very active community refining the electronic printing press

# The Protein Data Bank – A Best Case Scenario

**RCSB PDB** PROTEIN DATA BANK

As of Tuesday May 10, 2011 at 5 PM PDT there are 73009 Structures | PDB Statistics

Search: PDB ID or Text | PDB ID lookup or Text search of the complete structure | Advanced Search

### A Resource for Studying Biological Macromolecules

The PDB archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. As a member of the wwPDB, the RCSB PDB curates and annotates PDB data according to agreed upon standards.

The RCSB PDB also provides a variety of tools and resources. Users can perform simple and advanced searches based on annotations relating to sequence, structure and function. These molecules are visualized, downloaded, and analyzed by users who range from students to specialized scientists.

#### Featured Molecules

**Molecule of the Month: Cytochrome bc1**

Cells are masters at squeezing every drop of energy out of their food. They disassemble the molecules in food atom by atom, driving a variety of unusual energy transformations in the process. At the end, all of the hydrogen atoms have been separated from the food molecules and are used to turn the rotary motor of **ATP synthase**. To do this, the electrons are stripped from these hydrogen atoms and used to power huge protein pumps that transport protons across a membrane. These protons then power the rotation of ATP synthase as they return to their original positions.

**Protein Structure Initiative Featured Molecule: AZA Adenosine Receptor**

The structure of AZA adenosine receptor with a powerful agonist reveals a GPCR in action.

**Full Article | PSI Featured Molecule Archive | PSI Structural Biology Knowledgebase**

**Latest Structures**

- Paper not published unless data are deposited – strong data to literature correspondence
- Highly structured data conforming to extensive ontologies
- DOI's assigned to every structure

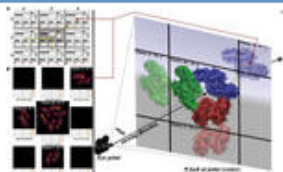
# Example Interoperability: The Database View

## Related Citations in PDB Entry (REMARK 1)

[Show](#)

Information provided by **BioLit**:

## PubMedCentral articles found to contain 1TIM ?



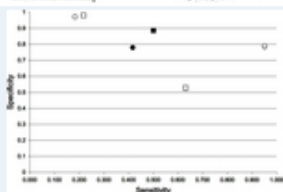
### PMG: online generation of high-quality molecular pictures and storyboarded animations

Ludovic Autin, Pierre Tufféry

*Nucleic Acids Research* 2007; 35(Web Server issue):W483-W488ff.

PubMedCentral: [1933120](#) PubMed: [17478496](#) DOI: [10.1093/nar/gkm277](#)

[Abstract](#) [Copyright](#)



### Computer-Based Screening of Functional Conformers of Proteins

Héctor Marlosti Montiel Molina, César Millán-Pacheco, Nina Pastor, Gabriel del Rio

*PLoS Computational Biology* 2008; 4(2):e1000009ff.

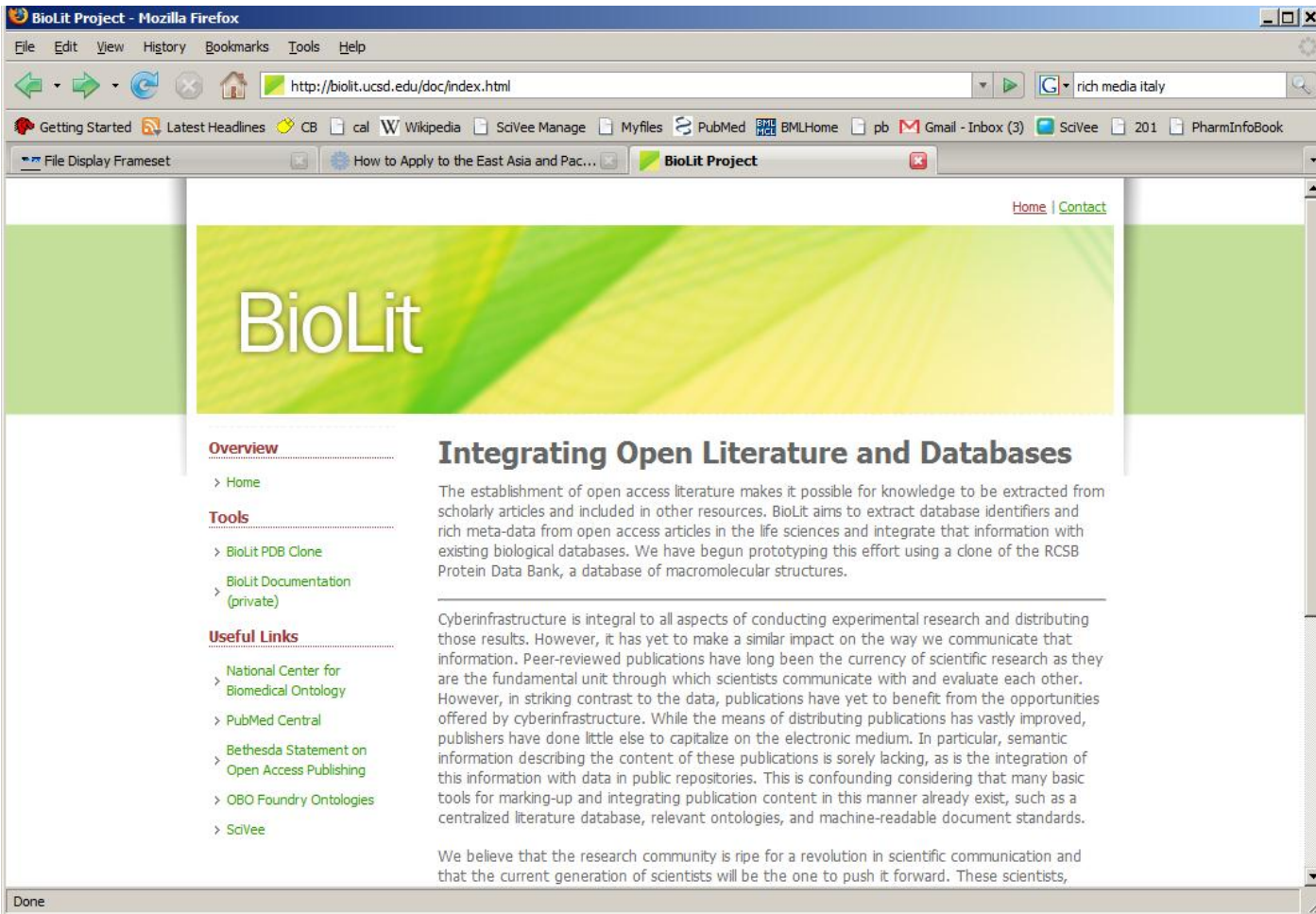
PubMedCentral: [2265533](#) PubMed:  DOI: [10.1371/journal.pcbi.1000009](#)

[Abstract](#) [Copyright](#)

## Other PDB IDs found in the above articles ?

Image	PDB ID	Structure Title	Sequence Similarity
	<b>1AON</b>	CRYSTAL STRUCTURE OF THE ASYMMETRIC CHAPERONIN COMPLEX GROEL/GROES/(ADP)7	-
	<b>1CRN</b>	WATER STRUCTURE OF A HYDROPHOBIC PROTEIN AT ATOMIC RESOLUTION. PENTAGON RINGS OF WATER MOLECULES IN CRYSTALS OF CRAMBIN	-
	<b>1A30</b>	HIV-1 PROTEASE COMPLEXED WITH A TRIPEPTIDE INHIBITOR	-

# Example Interoperability: The Literature View



The screenshot shows a Mozilla Firefox browser window displaying the BioLit Project website. The address bar shows the URL <http://biolit.ucsd.edu/doc/index.html>. The browser's menu bar includes File, Edit, View, History, Bookmarks, Tools, and Help. The toolbar contains navigation buttons and a search bar with the text "rich media italy". The browser's tab bar shows several open tabs, including "Getting Started", "Latest Headlines", "CB", "cal", "Wikipedia", "SciVee Manage", "Myfiles", "PubMed", "BMLHome", "pb", "Gmail - Inbox (3)", "SciVee", "201", and "PharmInfoBook". The active tab is "BioLit Project".

The website content features a large green and yellow banner with the text "BioLit". In the top right corner of the banner area, there are links for "Home" and "Contact". Below the banner, the page is divided into two main sections. On the left, there is a navigation menu with the following sections:

- Overview**
  - > Home
- Tools**
  - > BioLit PDB Clone
  - > BioLit Documentation (private)
- Useful Links**
  - > National Center for Biomedical Ontology
  - > PubMed Central
  - > Bethesda Statement on Open Access Publishing
  - > OBO Foundry Ontologies
  - > SciVee

The main content area on the right is titled "Integrating Open Literature and Databases". It contains two paragraphs of text:

The establishment of open access literature makes it possible for knowledge to be extracted from scholarly articles and included in other resources. BioLit aims to extract database identifiers and rich meta-data from open access articles in the life sciences and integrate that information with existing biological databases. We have begun prototyping this effort using a clone of the RCSB Protein Data Bank, a database of macromolecular structures.

Cyberinfrastructure is integral to all aspects of conducting experimental research and distributing those results. However, it has yet to make a similar impact on the way we communicate that information. Peer-reviewed publications have long been the currency of scientific research as they are the fundamental unit through which scientists communicate with and evaluate each other. However, in striking contrast to the data, publications have yet to benefit from the opportunities offered by cyberinfrastructure. While the means of distributing publications has vastly improved, publishers have done little else to capitalize on the electronic medium. In particular, semantic information describing the content of these publications is sorely lacking, as is the integration of this information with data in public repositories. This is confounding considering that many basic tools for marking-up and integrating publication content in this manner already exist, such as a centralized literature database, relevant ontologies, and machine-readable document standards.

We believe that the research community is ripe for a revolution in scientific communication and that the current generation of scientists will be the one to push it forward. These scientists,

<http://biolit.ucsd.edu>



# BioLit

- Tools ▶
- Views ▶
- Related Articles ▶
- Information ▶

Article Search:

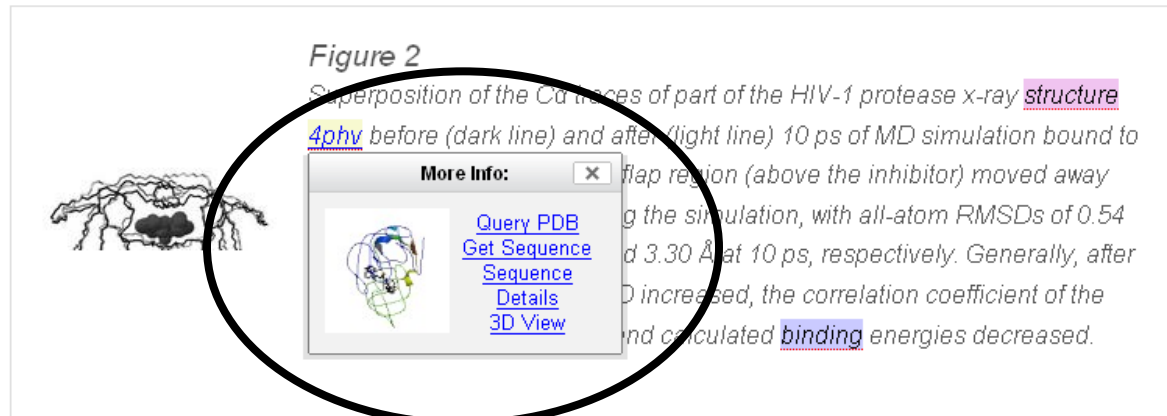
Go

- Found 70 terms.
- PDB Id ▲
    - [1HVJ](#) (1)
    - [1HVJ](#) (1)
    - [1HVK](#) (1)
    - [1HVL](#) (1)
    - [4PHV](#) (2)
  - Gene Ontology ▼
  - Infectious Disease Ontology ▼
  - Pathogen Transmission Ontology ▼
  - Physico-chemical Methods and Properties Ontology ▼
  - Physico-chemical Process Ontology ▼
  - BRENDA Tissue/Enzyme Source Ontology ▼
  - C. elegans Phenotype Ontology ▼
  - Environment Ontology ▼
  - Mouse Pathology Ontology ▼
  - Plant Structure Ontology ▼

## Influence of the protease flap movement on calculated binding energy

The beta-strand flap is the most flexible region in the HIV-1 protease. It is normally 7 Å RMSD from the active site and is in an open conformation in the native state [22,23]. The protease undergoes significant structural changes on binding to an inhibitor. The two flaps fold over the inhibitor to form a tunnel-shaped active site and are held in this close position by hydrogen bonding from Ile50 and Ile50' NH groups of the enzyme to a water molecule, which in turn is hydrogen bonded to the P2 and P1' CO groups of the inhibitor [24]. The bonding stabilizes the flaps in a closed position and inhibits the activities of the enzyme.

MD simulation has been used to study the movement of the flap region of HIV-1 protease with a ligand [25-30]. The flaps initially opened to an all-atom RMSD of 25 Å within 200 ps and became completely open at the end of a 10 ns simulation. In this study (Figure 2), the flaps opened up and moved away from the x-ray structure from 0.54 Å at 0.1 ps to 3.30 Å RMSD at 10 ps (the flap RMSD was calculated from residue 40 to 60 of each protein chain). These movements, after 0.1 ps of simulation, are inversely correlated with the quality of the binding energy prediction. As shown in Table 1, the correlation coefficient significantly decreased from 0.87 at 0.1 ps to 0.74 at 10 ps as the all-atom flap RMSD increased from 0.54 to 3.30 at 0.1 and 10 ps, respectively.



Complementarity between the ligand and the binding site is the basic concept behind ligand binding. This is manifest as steric complementarity, i.e. the shape of the ligand is mirrored in the shape of the binding site, allowing molecular interactions between

# Semantic Tagging & Widgets are Powerful Tools to Integrate Data and Knowledge of that Data, But as Yet Not Used Much

**RCSB PDB Comparison Tool**


Compare the following two proteins:

PDB1:  Chain1:

PDB2:  Chain2:

--- Select Comparison Method --- ▾

**RCSB PDB Molecule of the Month**  
July 2002



**p53 Tumor Suppressor**

© David S Goodsell and RCSB PDB

**PSI | nature**  
StructuralBiology Knowledgebase

Latest PSI SBKB Recent Articles  
Structural Function & Form

This Month's Highlights  
Industry

**FOX CHASE**  
CANCER CENTER

Enter PDB code here:

**RCSB PDB**  
PROTEIN DATA BANK

Will Widgets and Semantic Tagging Change Computational Biology?  
*PLoS Comp. Biol.* 6(2) e1000673

# Semantic Tagging of Database Content in The Literature or Elsewhere

## PDB Tag Library

*In a Nobel-prize-winning study, Wendell Stanley coaxed the virus to form crystals, and discovered that it was composed primarily of protein. Others quickly discovered that there was also RNA in the virus. Then, many prominent structural researchers (including J. D. Bernal, Rosalind Franklin, Ken Holmes, Aaron Klug, Don Caspar, and Gerald Stubbs) used X-ray diffraction and electron microscopy to probe the structure of the virus. Several structures of the whole tobacco mosaic virus are available in the PDB, including the ground-breaking one solved by X-ray fiber diffraction (shown here from PDB entry [2tmv](#)), and a recent structure solved by analysis of many electron micrographs (PDB entry [2om3](#)). The virus is composed of one strand of RNA (shown in red) wrapped inside a sheath of protein (shown in blue). The protein coat is composed of about 2130 copies of a small protein, which stack like bricks in a cylindrical chimney. The RNA strand encodes four proteins, which together orchestrate the life cycle of the virus. These include two proteins that replicate the viral RNA, a protein that transports the RNA from cell to cell, spreading the infection, and the capsid protein seen in the PDB structures.*

The PDB Tag Library is a rich markup widget that allows you to easily tag PDB ID's and keywords on your own site and automatically provide enhanced functionality that links back to us. Check out the example on the left by rolling over any of the underlined words. We currently provide 4 different tags: `rccb_id_tag` that provides an image and link to the structure explorer page, `rccb_menu_tag` that provides a menu with useful links, `rccb_author_tag` that provides a link to an exact author query and `rccb_keyword_tag` that provides a link to a keyword query results page. Note that the author tag searches both primary and citation authors.

### Options:

Simple ID Tag (`rccb_id_tag`):

```
class="rccb_id_tag"
rel="PDBID" - to override the content of the tag, you can specify the PDB ID
title="Any String" - You can add your own title
```

Menu Tag (`rccb_menu_tag`):

```
class="rccb_menu_tag"
rel="PDBID" - to override the content of the tag, you can specify the PDB ID
```

Author Tag (`rccb_author_tag`):

```
class="rccb_author_tag"
rel="AUTHOR" - to override the content of the tag you can specify the author name to search.
NOTE: This is an exact author search and you have to pass an exact name to get a result. (ex. "Lawson, C.")
```

Keyword Tag (`rccb_keyword_tag`):

```
class="rccb_keyword_tag"
```

```
<script type="text/javascript" src="http://www.rcsb.org/pdb/widgets/pdbTagLib_min.js"></script>

...study, <span class="rccb_author_tag" rel="Stanley, W.A.">Wendell Stanley</span>...
...PDB entry <span class="rccb_menu_tag">2tmv</span>), and...
...entry <span class="rccb_id_tag" title="Tobacco Mosaic Virus">2om3</span>...
...chimney. The <span class="rccb_keyword_tag">RNA</span> strand...
...and the <span class="rccb_id_tag" rel="3BWQ">capsid protein</span> seen...
```

BioLit: Article Explorer - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://biolit.ucsd.edu/biolit/explore.jsp?uid=54411

Beta BofA Cal Gmail jira More Myfiles PB PLoSCB PubMed WebCT Wikipedia SciVee SciVee PM Profdev ISI 201 Book BioLit

# BioLit

- Tools
- Views
- Related Articles
- Information

Article Search:

Go

Found 0 terms.

- PDB Id
- Gene Ontology
- Infectious Disease Ontology
- Pathogen Transmission Ontology
- Physico-chemical Methods and Properties Ontology
- Physico-chemical Process Ontology
- BRENDA Tissue/Enzyme Source Ontology
- C. elegans Phenotype Ontology
- Environment Ontology
- Mouse Pathology Ontology
- Plant Structure Ontology
- Other Terms

## Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations

Ekachai Jenwitheesuk, Ram Samudrala  
*BMC Structural Biology* (BMC Struct Biol. 2003; 3:2) [[PubMed central](#)]

### Background

The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology. This study describes the improvement of protein-ligand [binding](#) energy prediction by incorporating protein flexibility through the use of molecular dynamics (MD) simulations.

### Results

Docking experiments were undertaken using the program AutoDock for twenty-five HIV-1 protease-inhibitor complexes determined by x-ray [crystallography](#). Protein-rigid docking without any dynamics produced a low correlation of 0.38 between the experimental and calculated [binding](#) energies. Correlations improved significantly for all time scales of MD simulations of the receptor-ligand complex. The highest correlation coefficient of 0.87 between the experimental and calculated energies was obtained after 0.1 picoseconds of dynamics simulation.

### Conclusion

Our results indicate that [relaxation](#) of [protein complexes](#) by MD simulation is useful and necessary to obtain [binding](#) energies that are representative of the experimentally determined values.

### Background

The human immunodeficiency virus type 1 aspartic protease (HIV-1 PR) is an important enzyme due to its key [role](#) in viral maturation. Inactivation of the enzyme causes the production of immature, noninfectious viral particles. The enzyme therefore is an attractive target in anti-AIDS drug design, and the effect of [binding](#) various inhibitors on the protease [structure](#) is currently the focus of intensive research [[1-3](#)].

To obtain information about the position and energy of [binding](#) between an inhibitor and the corresponding protein, several



# The Publishers are Starting to Do It

## App Integration in Science Direct

Result list | previous < 4 of 12 > next

PDF (339 K) Export Citation E-mail Article

Font Size: - +

**Article** Figures/Tables (4)

**Journal of Molecular Biology**  
Volume 356, Issue 4, 3 March 2006, Pages 920-927

doi:10.1016/j.jmb.2005.11.094 | How to cite this article  
Copyright © 2005 Elsevier Ltd. All rights reserved.  
Permissions & Reprints

### Structural Insights in HIV-1 Protease NL4-3 in Complex with Inhibitor, TL-3

Holly Heasley<sup>a</sup>, Victoria Kutilek<sup>b</sup>, Torbett<sup>b</sup> and C. David Stout<sup>a</sup>

<sup>a</sup>Department of Molecular Biology, University of California, San Diego, La Jolla, CA 92037, USA  
<sup>b</sup>Department of Molecular and Experimental Medicine, University of California, San Diego, La Jolla, CA 92037, USA

#### Abstract

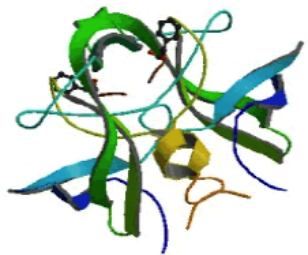
The development of resistance to HIV-1 protease inhibitors is a major problem in the treatment of HIV-1. We have also obtained the crystal structures of three mutant forms of NL4-3 protease containing one (V82A), three (V82A, M46I, F53L) and six (V82A, M46I, F53L, V77I, L24I, L63P) point mutations in complex with TL-3. The three protease mutants arose sequentially under ex vivo selective pressure in the presence of TL-3, and exhibit fourfold, 11-fold, and 30-fold resistance to TL-3, respectively. This series of protease crystal structures offers insights into the biochemical and structural mechanisms by which the enzyme can overcome inhibition by TL-3 while recovering some of its native catalytic activity.

**Keywords:** HIV-1 protease; drug resistance; viral evolution; crystal structure; mutation

**Abbreviations:** HIV-1, human immunodeficiency virus type 1

**Article Outline**

**PDB Structure Viewer**



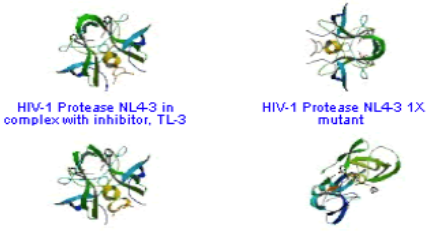
HIV-1 Protease NL4-3 in complex with inhibitor, TL-3

Release Date: 28-Feb-2006  
Exp. Method: X-RAY DIFFRACTION  
Hydrolase/hydrolase Inhibitor  
Molecule: PROTEASE RETROPEPSIN  
Polymer: 1 Type: polypeptide(L)  
Chains: A  
EC#: 3.4.23.16  
Molecule: TL-3 [[PHENYLMETHYLOXY-CARBONYL]-ALANINYL]-VALINYL-[PHENYL-1-HYDROXYPROP-2-YL]-AMINE  
Polymer: 2 Type: polypeptide(L)  
Chains: I  
Fragment: Half of TL-3 molecule in the asymmetric unit

**My Applications**

+ Add Show all apps ?

**PDB Structure Viewer**



HIV-1 Protease NL4-3 in complex with inhibitor, TL-3  
HIV-1 Protease NL4-3 1X mutant  
HIV-1 Protease NL4-3 3X mutant in complex with inhibitor, TL-3  
HIV-1 Protease NL4-3 6X mutant

About PDB Structure Viewer

**Microsoft Author Network Visualizer**

Holly Heasley John H. Elder  
Victoria Kutilek Bruce E. Torbett  
Garrett M. Morris C. David Stout  
Ying-Chuan Lin

About Microsoft Author Network Visualizer

**Net Base Analyzer**

**Related Articles**

- Can extra-dimensional effects replace dark matter?  
*Physics Letters B*
- A generic test of modified gravity models which emulate...  
*Physics Letters B*
- Dark matter as a geometric effect in f(R) gravity  
*Astroparticle Physics*
- Planet-bound dark matter and the internal heat of Uranus...  
*Physics Letters B*
- The influence of dark matter on the motion of planets a...  
*Physics Letters A*



# Others are Doing It Very Successfully



Overview Download Quick Start Publications Contact Us

## Comment. Explore. Interact.



**Utopia Documents** allows new ways of thinking about your documents: it brings your papers to life by linking to live resources on the web and turning static data into live interactive content.

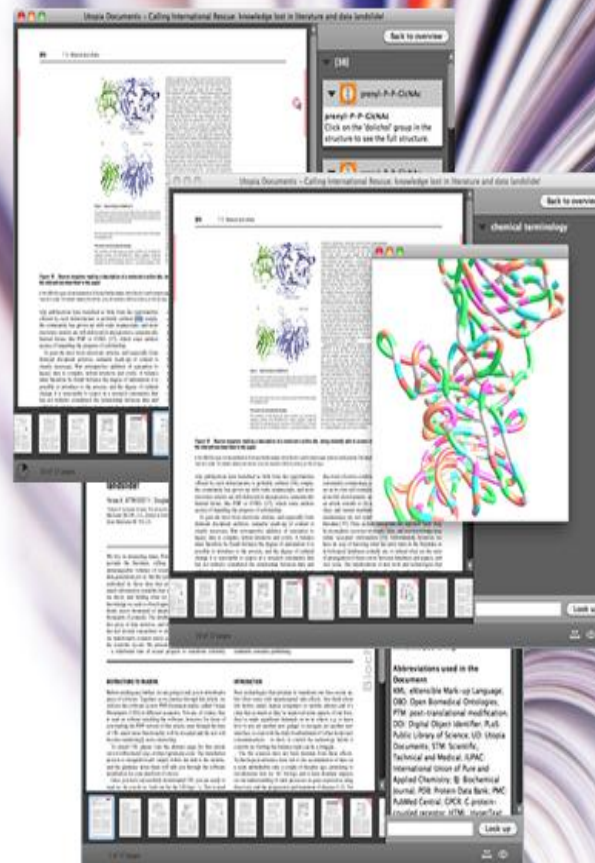
Utopia Documents is completely free to download and use.

[Download v1.1.2 now](#)

## News

25th Mar 2011 – The latest version of Utopia Documents has been released (v1.1.2). [Download now](#)

11th Feb 2011 – A [video of Steve's talk](#) from [Beyond The PDF](#) is now available to watch on YouTube.



# This is Literature Post-processing Better to Get the Authors Involved

- Authors are the absolute experts on the content
- More effective distribution of labor
- Add metadata before the article enters the publishing process

# Word Add-in for authors

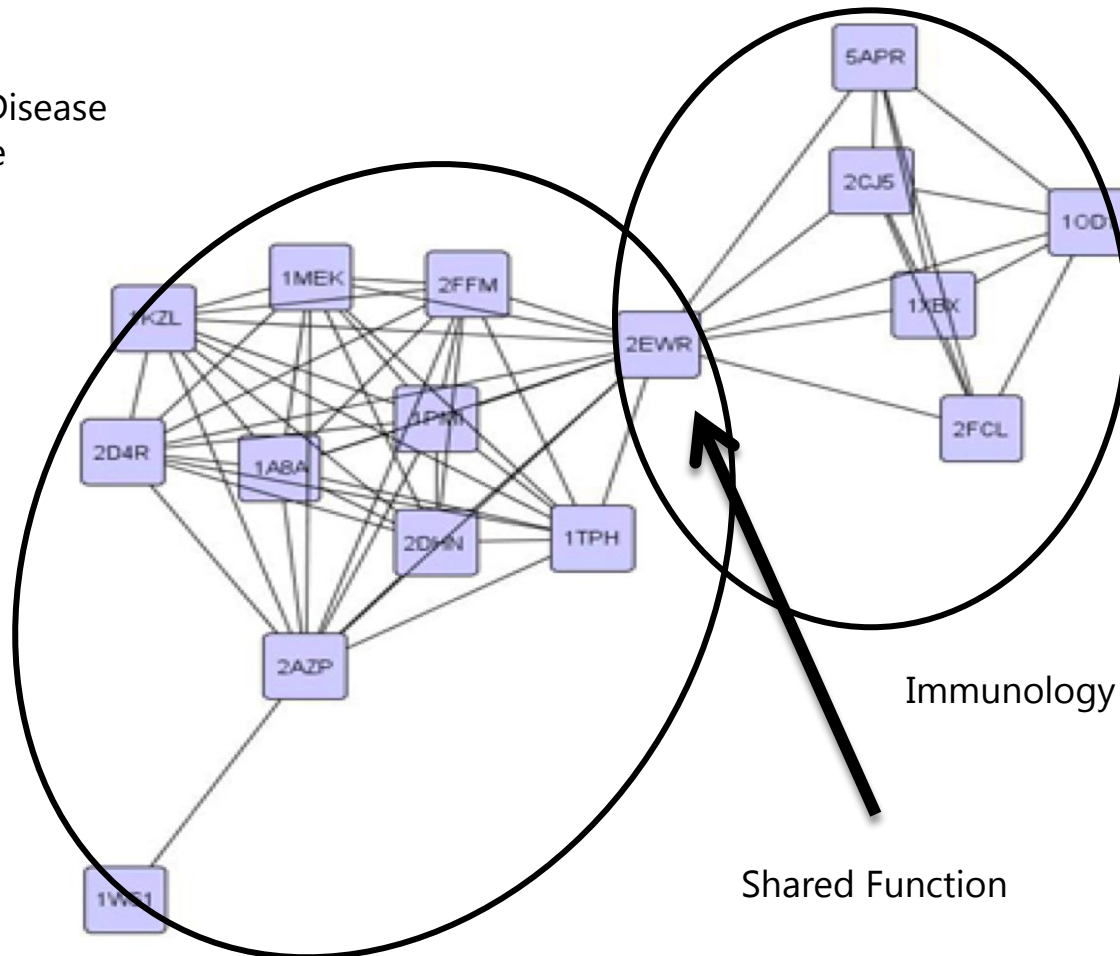
- Allows authors to add metadata as they write, before they submit the manuscript
- Authors are assisted by automated term recognition
  - OBO ontologies
  - Database IDs
- Metadata are embedded directly into the manuscript document via XML tags, OOXML format
  - Open
  - Machine-readable
- Open source, Microsoft Public License

# Challenges

- Authors
  - *Carrot* IF one or more publishers fast tracked a paper that had semantic markup it might catch on
- Publishers
  - *Carrot* Competitive advantage

# The Promise

Cardiac Disease  
Literature

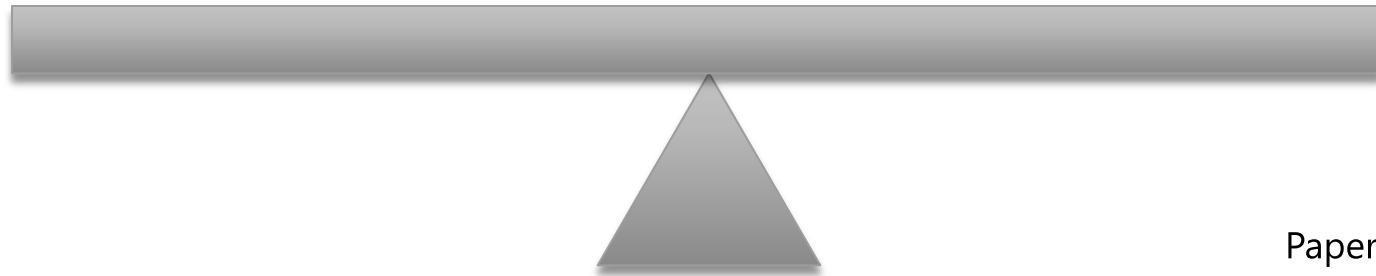




# The Reward System Must Change

**The Right Thing To Do**

**Reward**



Papers  
Grants

Reviews  
Provision of metadata

Curation  
Alternative forms of dissemination  
New tools



Data availability

Open access

# Yes YouTube Can Increase the Rate of Discovery



Making science visible

# The Lab Experiment Paper+Rich Media



- My students enjoyed the experience
- The shyest student was actually the most bold in front of the camera
- "We will become a generation of "science castors"
- They liked the exposure for the most part – rather than the PI it puts them out in front

# Three Years Later - Organic Growth

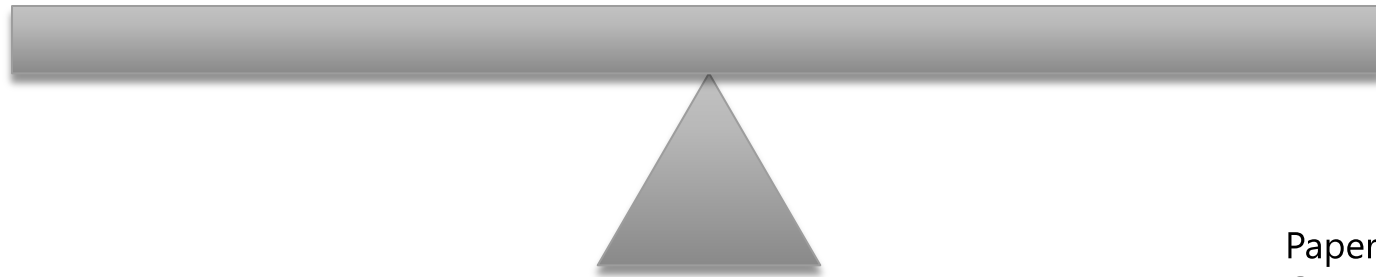


- Some of their work viewed 20,000+ times
- Global audience of researchers, educators and academic/research institutions
  - 60,000 unique visitors & 2M pageviews/month
  - 16,000 registered users & 600 communities
  - 5,000 uploads of video content (about journal articles, conferences, research news and classes)
  - Growing 4-5% monthly
- Sustainability - evolving a business model supporting journals and conferences

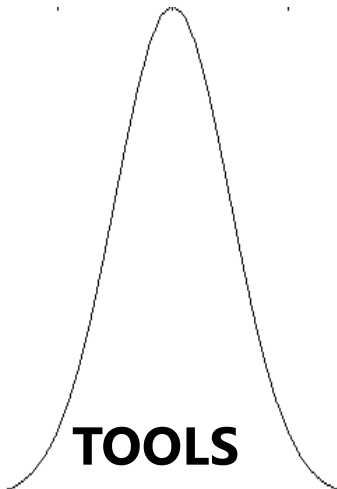
# Are We There Yet?

**The Right Thing To Do**

**Reward**



Papers  
Grants



- Data availability
- Review
- Provision of metadata
- Open access
- Curation
- Alternative forms of dissemination
- New tools





# Pipeline Assembly



Scientist

Tools for idea management just appearing in the cloud



Idea

LIMS system too limited



Experiment

Institutional repositories – Do they really know what they are doing? Metrics of success interoperability etc.



Data



Conclusions



Publish



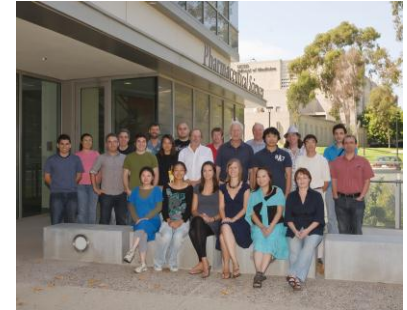
uzar.wordpress.com

It would not be the first wall to come down

# Acknowledgements

- BioLit Team
  - Lynn Fink
  - Parker Williams
  - Marco Martinez
  - Rahul Chandran
  - Greg Quinn
- Microsoft Scholarly Communications
  - Pablo Fernicola
  - Lee Dirks
  - Savas Parastitidas
  - Alex Wade
  - Tony Hey

- wwPDB team
  - Andreas Prilc
  - Dimitris Dimitropoulos
- SciVee Team
  - Apryl Bailey
  - Leo Chalupa
  - Lynn Fink
  - Marc Friedman (CEO)
  - Ken Liu
  - Alex Ramos
  - Willy Suwanto
  - Ben Yukich



<http://www.scivee.tv>



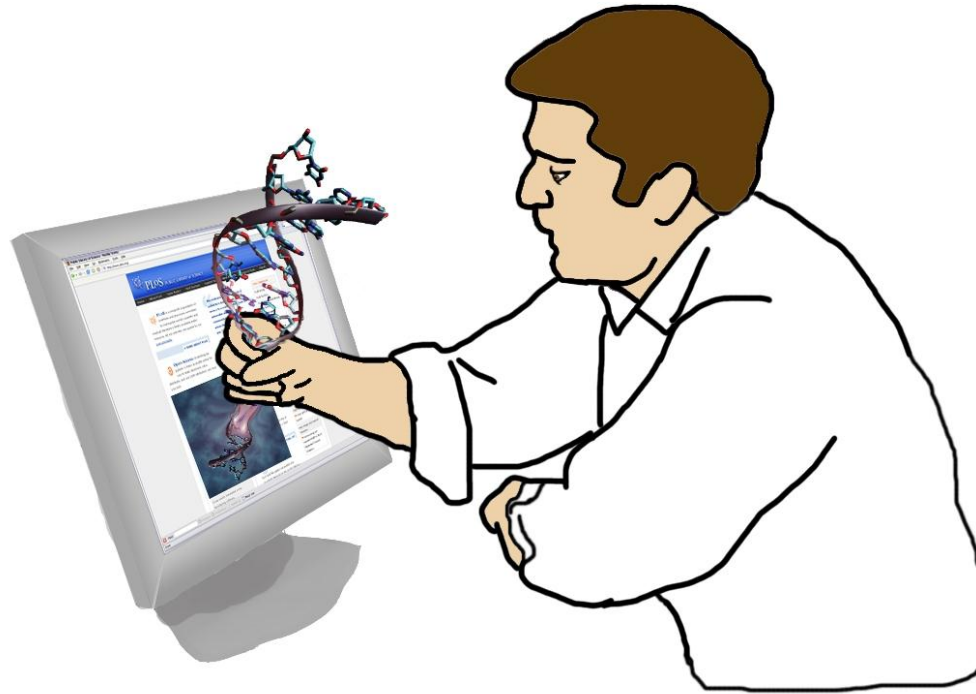
<http://biolit.ucsd.edu>

<http://www.pdb.org>

<http://www.codeplex.com/ucsdbiolit>



[pbourne@ucsd.edu](mailto:pbourne@ucsd.edu)



Questions?