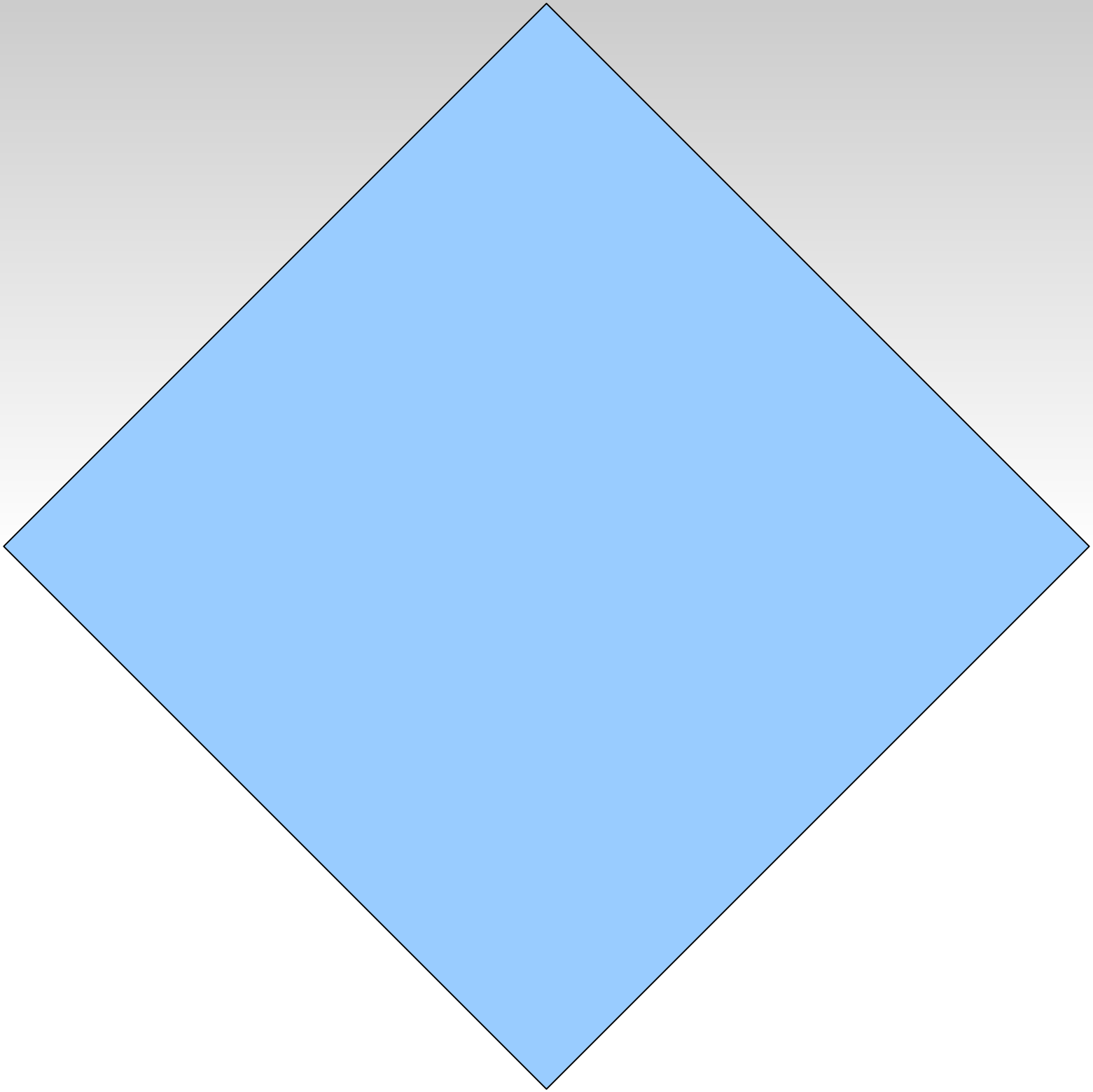
PDBWiki: success or failure?

Factors for successful community annotation projects

Dan Bolser
(dan.bolser@gmail.com)

NETTAB 2010,
Naples, Italy



Motivation for this work

- Opportunity to look at the strengths and weaknesses of the PDBWiki project
  - What did we learn?
    - Successes
    - Failures
  - How can we improve?

General principles for community annotation?

Rules for success

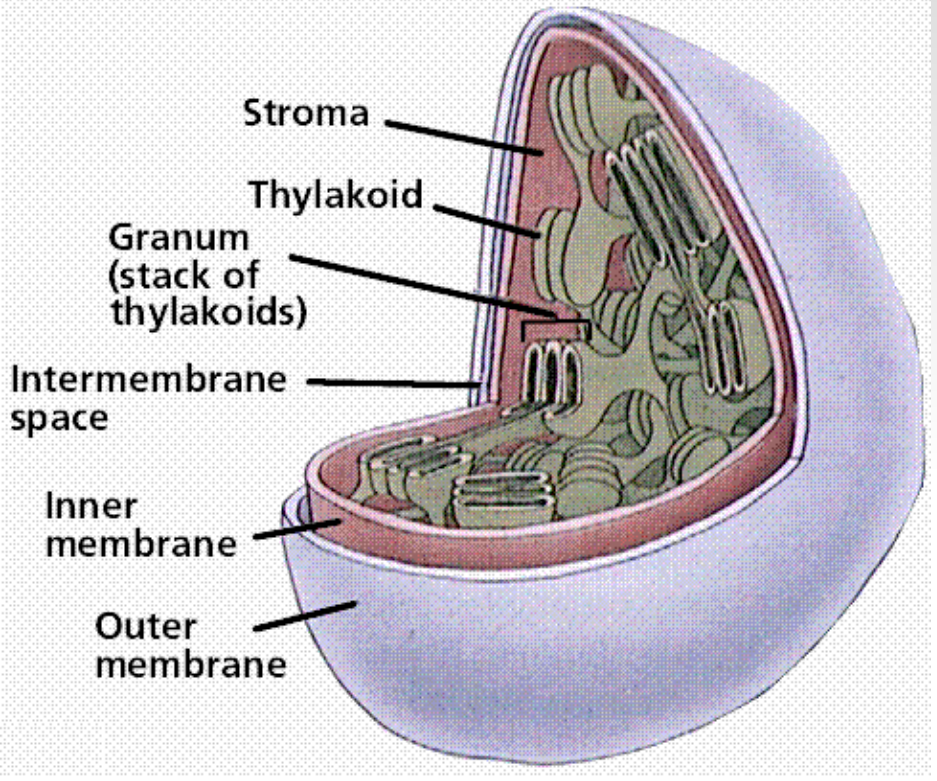
1) Useful content
2) Benefit to contributors
3) Recognition for contribution
4) Fun

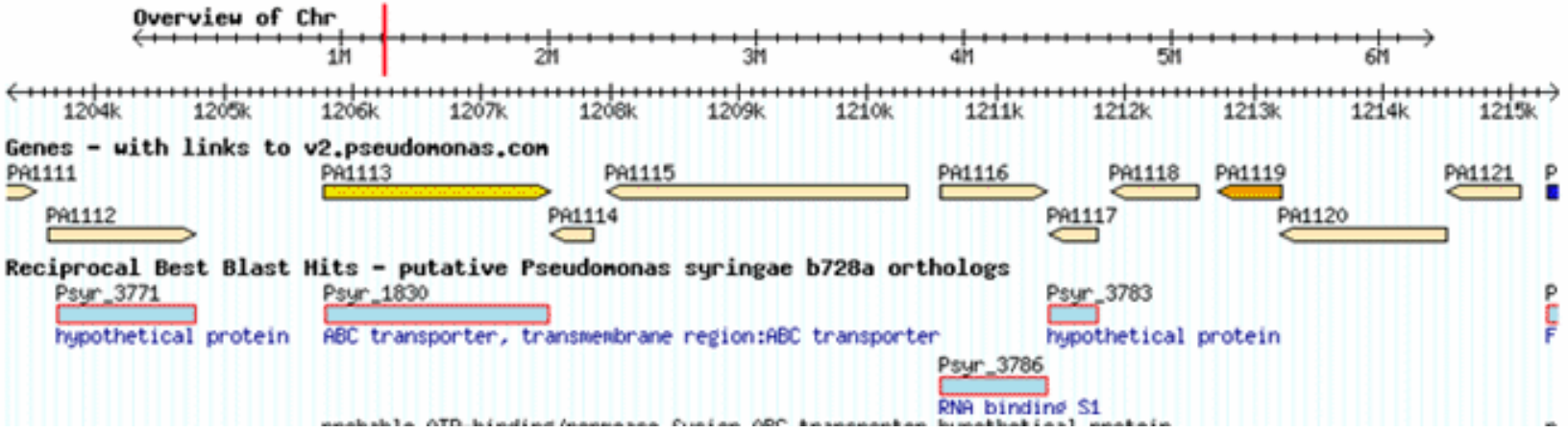
Presentation overview

- Community annotation
  - Why is it necessary?
- BioWikis:
  - The Wiki Wiki Web!
- When does it work (or not)?

Community annotation

Community annotation

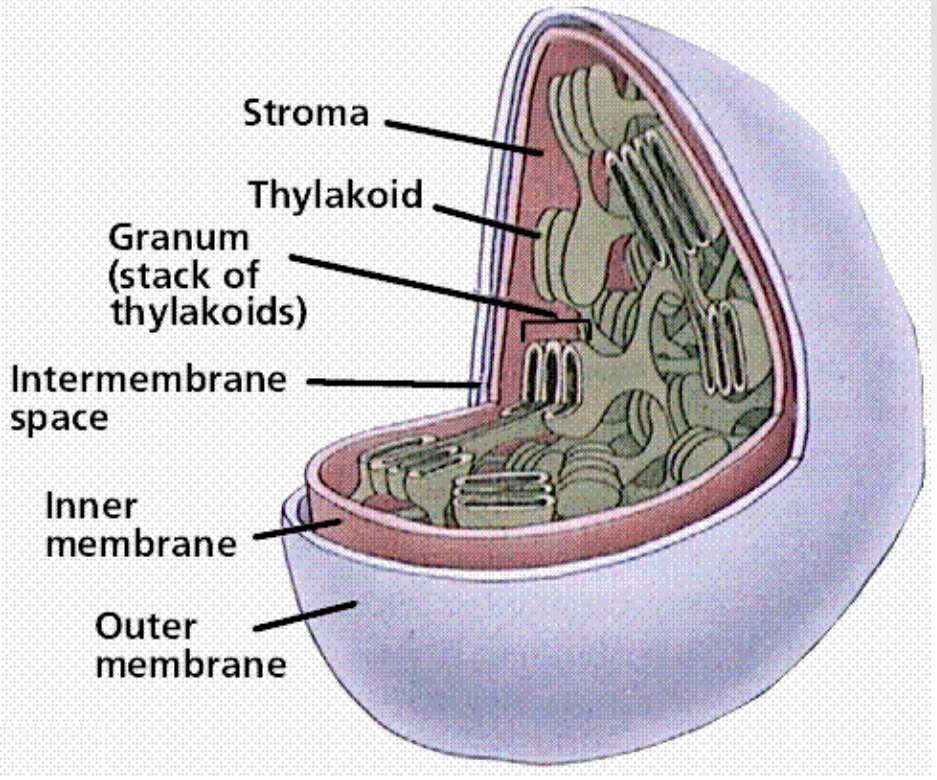


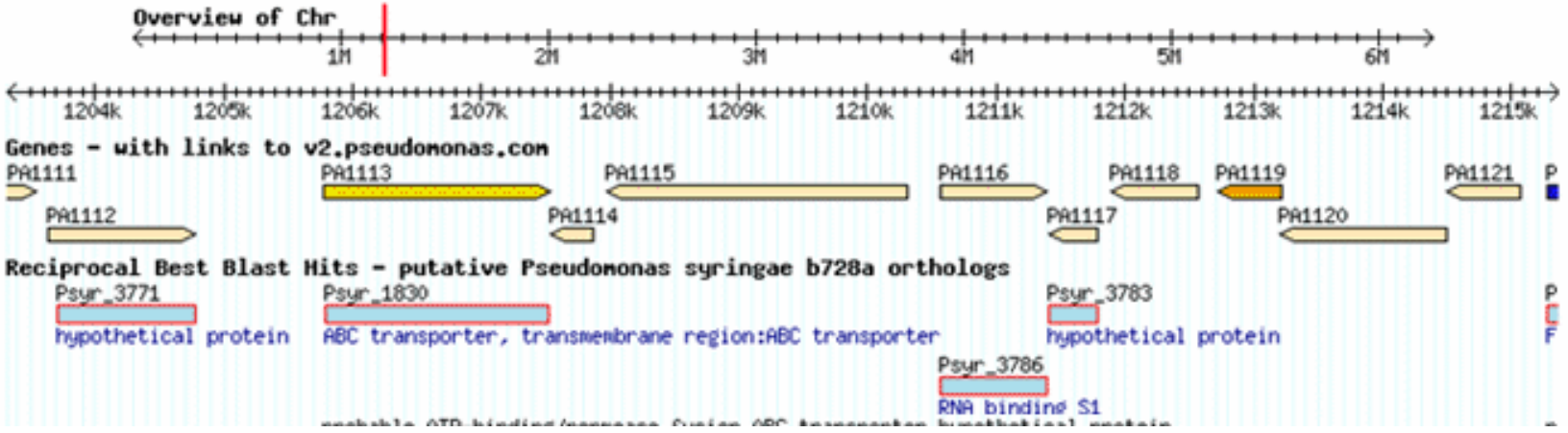


Community annotation

Has been driven by two key factors:

- The vast increase in biological data
- The clear success of Wikipedia



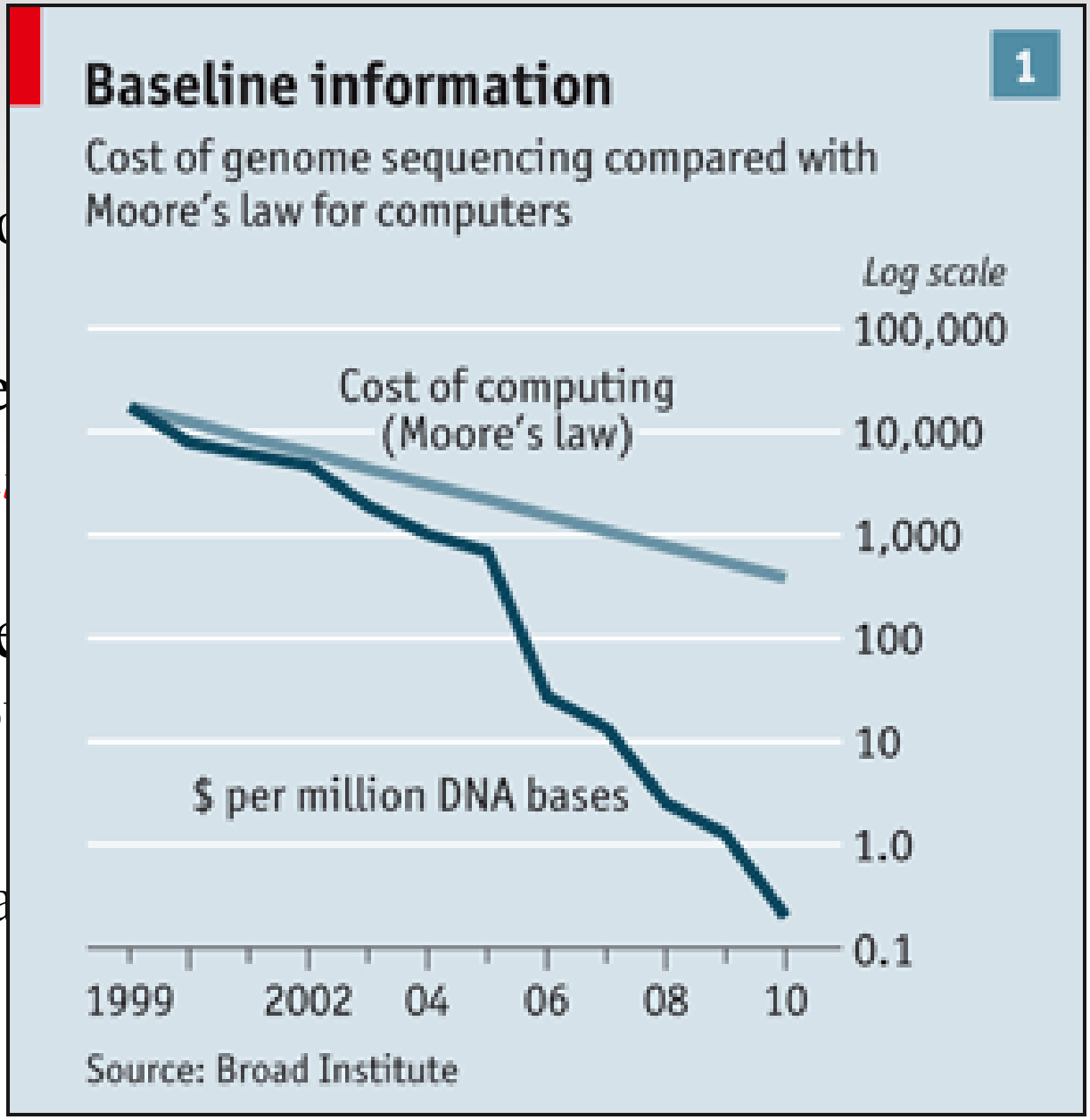


BioMoore's Law

- Over time:
  - Cost per unit of information can be decreased by orders of magnitude.
  - Throughput is increased by orders of magnitude.
    - Fan et al. **2006**. *Nat Rev Genet*.
- Comprehensive disease studies that might require ~1bn genotypes would now cost only a few million dollars.
  - Revolution in human genetics.

BioMoore's Law

- Over time:
  - Cost per unit of info
    of magnitude.
  - Throughput is incre
    - Fan et al. **2006**. *Na*
- Comprehensive dise
  ~1bn genotypes wo
  dollars.
  - Revolution in huma

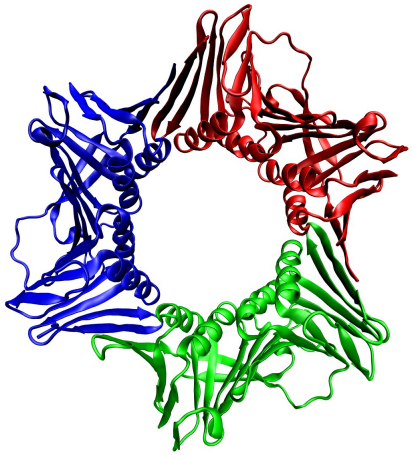


Community annotation

- Centralised databases can't cope with annotating the influx of data.
- Less investment in more specialised data.
  - Fewer people with a stake.
  - Specialists more disparate.
    - Communities are smaller and more focused.
- Do wikis hold the answer?
  - Wikipedia as a model…

The success of Wikipedia

- Wikipedia is consistently among one of the top 10 websites in the world (http://www.alexa.com).
  - Google > Facebook > YouTube > Yahoo! > Windows Live > Baidu > Wikipedia > ...
    - 200k edits per day.
    - 100k active users per month.
- WikiProject
  - Molecular and Cellular Biology



Wikipedia:WikiProject Molecular and Cellular Biology

From Wikipedia, the free encyclopedia

The MCB project needs your help!

By focusing our efforts to produce articles of very high quality, the **MCB Collaboration of the Month** serves a vital role both to the WikiProject and to the greater Wikipedia project as a whole. Unfortunately, the contributions to the CotM have fallen off greatly in the past few months. If possible, please help!

About us [edit]

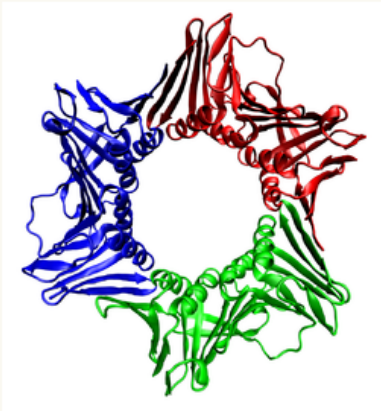
This project aims to better organize information in articles related to molecular and cell biology on Wikipedia. A major problem facing biology is organizing the vast amount of information that has been and continues to be collected about genomes, proteomes and cell function. A primary goal of this project is to collect and organize the totality of this information and make it accessible to researchers and laypeople alike by providing an entry point to the wealth of biological data that is currently hidden in obscure databases and journal articles.

If you would like to help, feel free to add yourself to the list of participants, or just look over the How you can help section below. Also of interest is the Molecular and Cellular Biology Portal associated with this project.

Concrete goals [edit]

- Create and perfect articles on the fundamental topics in molecular and cellular biology, with a particular focus on subjects that are discussed at the primary and secondary school level. The perfect article is complete, but accessible to a secondary school student.
- Create a standard system for presenting the information about the proteins and cells that are discussed in Wikipedia articles.

WikiProject Molecular and Cellular Biology



Featured Molecular and Cellular Biology Illustration. The assembled human DNA clamp, a trimer of the protein PCNA, which is a processivity factor that increases the speed and efficiency of DNA replication. In the assembled replication complex, the clamp encircles the DNA, which passes through the central pore.
Image rendered by Opabinia regalis

Project standards

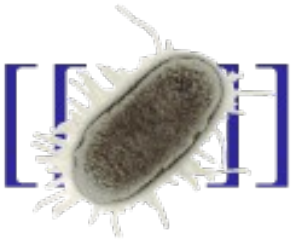
Portal: *MCB Portal*

But Wikipedia isn't always the answer ...

- Wikipedia is an educational resource.
  - All articles are encyclopaedic in style.
  - Explicitly forbids data from 'original research':
    - http://wikipedia.org/wiki/Wikipedia:No_original_research
      - **"Wikipedia does not publish original research"**.
  - No tools for the specific analysis, presentation, or collection of 'biological' data.
- BioWikis!

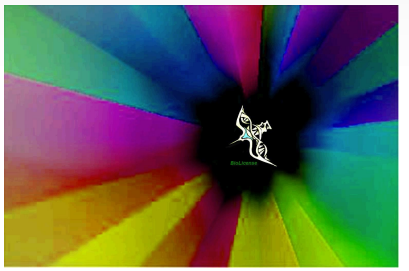
BioWikis

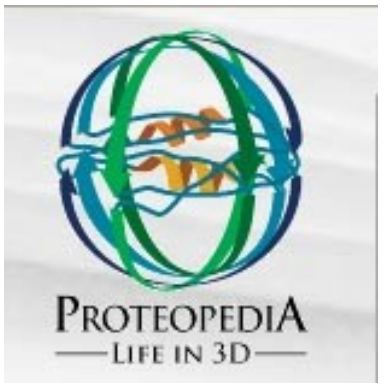
Wikis with a biological subject matter, customized for analysis, presentation and collection of specific biological data and biological data types:



Bioinformatics Organization

OpenWetWare
Share your science.





MB

SNPedia

What is PDBWiki?

- Allows the protein structures in the PDB to be tagged with specific annotations.
  - Functions as a bug tracker for users of the PDB.
  - Stehr H, Duarte JM, Lappe M, Bhak J, Bolser DM. (2010) **PDBWiki: added value through community annotation of the Protein Data Bank.** *Database*. baq009
  - http://pdbwiki.org

page | discussion | view source | history

Welcome to PDBWiki - A community annotated knowledge base of biological molecular structures

About • Help • Credits • Index

PDBWiki 1.2

PDBWiki was last updated on **2010-11-29**. It currently contains data from **69510** remediated PDB entries and **131** user annotations.

Contact: **info@pdbwiki.org**

Learn about the PDB

- What is the PDB?
- What is a PDB code?
- What is a biological unit?
- PDB FAQ

Explore protein space

- Browse structures by category
- Browse structures by user annotation
- Create custom reports
- Other access methods

How to contribute

- Help annotating structures
- Add links to databases and tools
- Ask or answer PDB related questions
- Help with PDBWiki development

New structures

- 2010-11-29 3kns
- 2010-11-29 3n4l
- 2010-11-29 3nmj
- 2010-11-29 3pb3

Recent user annotations

- 2010-11-08 3tat
- 2010-10-28 1qzw
- 2010-10-26 3hbx

Most popular structures

- 2pdt
- 1nsc
- 1hss

Most popular articles

- List more latest discussions
- PDB FAQ


1knb
CRYSTAL STRUCTURE OF THE RECEPTOR-BINDING DOMAIN OF ADENOVIRUS TYPE 5 FIBER PROTEIN AT 1.7 ANGSTROMS RESOLUTION
Authors Xia, D. Henry, L.J. Gerard, R.D. Deisenhofer, J.
Citation Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 A resolution.
Release date 1995-03-31
Exp. Method X-RAY DIFFRACTION
Resolution 1.7 Å
Classification CELL RECEPTOR RECOGNITION
Sequence
Chain(s) A (196 residues): Blast Uniprot
A
User comments
Errors in PDB file
Source:
In this structure, all the strands are anti-parallel. But in its PDB files:
SHEET 1 V 6 THR 400 TRP 402 0
SHEET 2 V 6 ASP 418 LYS 427 1
SHEET 3 V 6 SER 430 ALA 440 1
SHEET 4 V 6 ASN 479 ASN 482 1
SHEET 5 V 6 LEU 485 THR 486 1
SHEET 6 V 6 TYR 573 TYR 577 1
SHEET 1 R 4 SER 454 PHE 461 0
SHEET 2 R 4 ASN 515 TYR 521 1
SHEET 3 R 4 LYS 528 THR 535 1
SHEET 4 R 4 TYR 550 TRP 556 1
B
The senses of the strands are all incorrectly labelled as parallel as indicated in col 39-40. They should be labelled as "-1".
Name (required):
John Smith
Comment title:
Similar cases
Comment:
See also PDB [[1amr]]
Category for this comment: PDB annotation error
Submit
Links
Search for 1knb in:
Edit me!
Databases
Protein Data Bank
PDBsum
SCOP
CATH
Pfam
jenaLib
OCA
Proteopedia
TOPSAN
Electron Density Server
add database
Visualization tools
RCSB Jmol viewer
RCSB KING viewer
FirstGlance (Jmol)
AstexViewer
Polyview 3D (Rendered images)
PDBCartoon (secondary structure)
FaceViewer
Blue Star Sting
Surface topography (CastP)
Domain information (XDom)
add visualization tool
Analysis tools
Structural Neighbours (Dali Database)
Likely biochemical function (ProFunc)
Functionally important residues (FireDB)
Predicted functional sites (WebFeature)
Catalytic sites (CSA)
WHAT IF quality report
ProSA-web quality analysis
MolProbity quality analysis
NQ-Flipper
Contact Analysis (CSU)
add analysis tool
Quarternary structure
PDB (Download coordinates)
PQS
PistBuD
PiQSi
3DComplex
PISA interfaces
PISA monomers
PISA assemblies
add quarternary structure link
C
Categories: PDB annotation error | Annotated PDB entry | KN | PDB entry | X-RAY DIFFRACTION | CELL RECEPTOR RECOGNITION | HUMAN ADENOVIRUS 5 | 1995

Links

Search for 1knb in:

Databases

- Protein Data Bank
- PDBsum
- SCOP
- CATH
- Pfam
- JenaLib
- OCA
- Proteopedia
- TOPSAN
- Electron Density Server
- add database

Links

Search for **1knb** in:

	Visualization tools
▪ Pro	▪ RCSB Jmol viewer
▪ PDI	▪ RCSB KiNG viewer
▪ SC(	▪ FirstGlance (Jmol)
▪ CA1	▪ AstexViewer
▪ Pfa	▪ Polyview 3D (Rendered images)
▪ Jen	▪ PDBCartoon (secondary structure)
▪ OC/	▪ FaceView
▪ Pro	▪ Blue Star Sting
▪ TOI	▪ Surface topography (CastP)
▪ Ele	▪ Domain information (XDom)
▪ add	▪ add visualization tool

Links

Search for 1knb in:

		Analysis tools
▪ Pro	▪ RCS	▪ Structural Neighbours (Dali Database)
▪ PDt	▪ RCS	▪ Likely biochemical function (ProFunc)
▪ SC(	▪ First	▪ Functionally important residues (FireDB)
▪ CA	▪ Aste	▪ Predicted functional sites (WebFeature)
▪ Pfa	▪ Poly	▪ Catalytic sites (CSA)
▪ Jen	▪ PDB	▪ WHAT IF quality report
▪ OC	▪ Face	▪ ProSA-web quality analysis
▪ Pro	▪ Blue	▪ MolProbity quality analysis
▪ TOI	▪ Surf	▪ NQ-Flipper
▪ Ele	▪ Dom	▪ Contact Analysis (CSU)
▪ add	▪ add	▪ add analysis tool

Links

Search for **1knb** in:

Edit me!

			Quarternary structure
• Pro	• RCS	• Stru	• PDB (Download coordinates)
• PDt	• RCS	• Like	• PQS
• SC(	• First	• Fun	• ProtBuD
• CA1	• Aste	• Pre	• PiQSi
• Pfa	• Poly	• Cat	• 3DComplex
• Jen	• PDB	• WH	• PISA interfaces
• OC/	• Face	• Pro	• PISA monomers
• Pro	• Blue	• Mol	• PISA assemblies
• TOF	• Surf	• NQ-	• add quarternary structure link
• Ele	• Dom	• Contact Analysis (CSU)	
• add	• add	• add analysis tool	

When does it work?


If you build it,
they will come.
ICANHASCHEEZBURGER.COM

Rules for success(?)

١) Must provide useful content in a convenient way

Focused, unique, organised, query-able data

٢) Contributions should provide a direct benefit

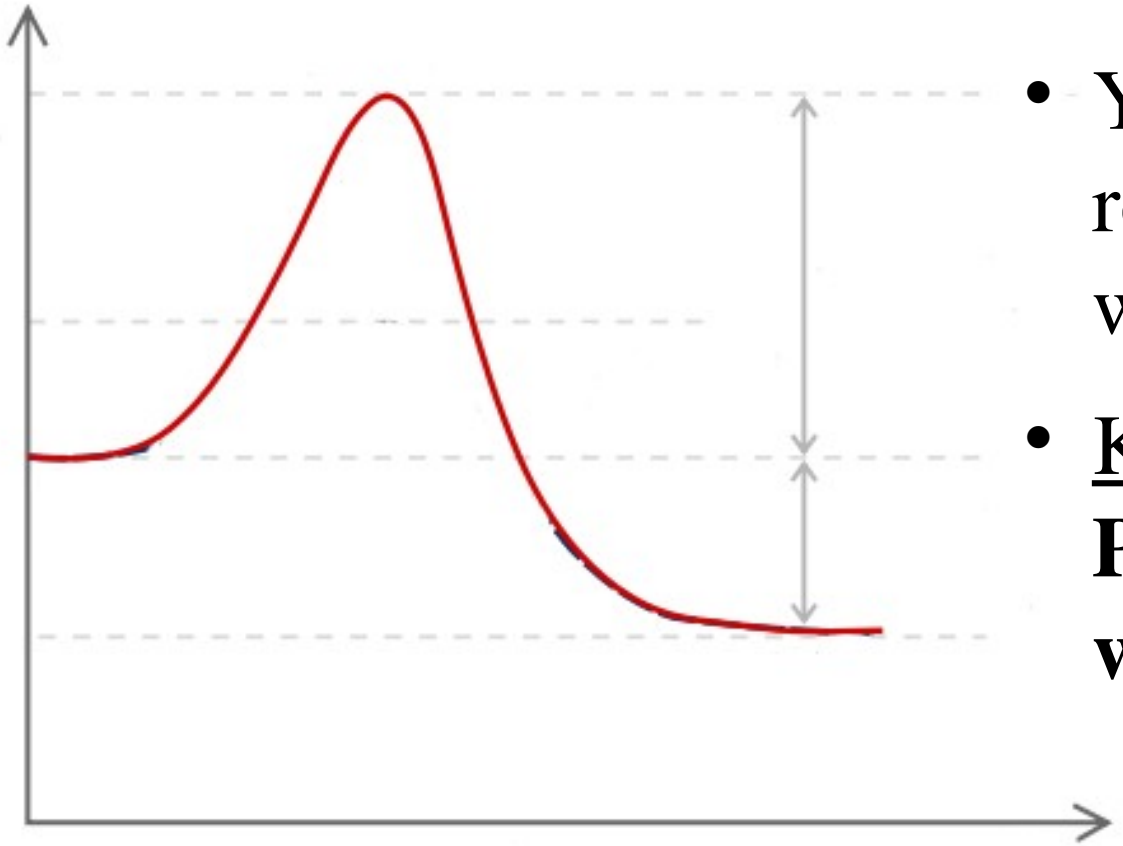
Self promotion / Functionality / Recognition

٣) Contributors should be formally 'recognized'

Visibility

These factors often depend on COMMUNITY

Building a community…



- Activation energy!
- You have to build up a resource before users will contribute!
- Kittur et. al. (2007) **Power of the few vs. wisdom of the crowd.**

http://www.parc.com/publication/1749/power-of-the-few-vs-wisdom-of-the-crowd.html

Recognition

- People work for recognition.
  - In science, this typically comes from publication of per-reviewed papers.
  - Why contribute to a wiki?
    - Perhaps this **will** get you a publication?
- Peer review is not just about papers.
  - Contributors to Wikipedia are recognised among their peers!

Recognition

- Alternative models of recognition.
  - Wiki edits are unlikely to impress anyone on a CV, however…
  - Community mailing lists are a great way to network.
    - http://biodatabase.org/index.php/List_of_mailing_lists_for_biologists
  - Recognition can come from contribution to community projects!

Game mechanics? (Fun)

- Crowd sourcing
  - Using ‘the crowd’ to do useful work
- Game mechanics
  - Applying Game Mechanics to Functional Software
  - http://www.youtube.com/watch?v=ihUt-163gZI

- Ease of use, robust infrastructure, and recognition of user contributions are encapsulated by the simple idea of making the site ‘fun’.

PDBWiki is a success(?)

1) Must provide useful content in a convenient way

   Success: Met our need for a shared 'computational kill list' for the PDB.

   Fail: These feature can be made more convenient.

2) Contributions should provide a direct benefit

   Success: We collected mostly annotations of this type, and edits to the 'links' section were especially popular.

3) Contributors should be formally 'recognized'

   Fail: We didn't do a good job of clearly acknowledging our contributors.

Conclusions

- The wiki concept is a simple improvement on the original concept of the web.
  - Sharing data.
- BioWikis must be fun and attractive for users.
- Structured wikis promise to change our idea of a ‘web database’.
  - Read only databases will be hard to imagine.

Acknowledgements

- Henning Stehr and Jose Duarte for PDBWiki
- All the contributors to http://PDBWiki.org
- Jong Bhak for his BioWiki concept
- NETTAB organisers
  - Paolo, Angelo, Claudia, and others.
- Linus Torvalds for Linux, Rasmus Lerdorf for PHP, and all scientists who pursue their work with honesty and integrity.

irc://irc.freenode.net/
#semantic-mediawiki
#bioinformatics

References

- Wikinomics: http://www.ncbi.nlm.nih.gov/pubmed/18769412
  - EcoliWiki / Gene Wiki / OpenWetWare / PDBWiki / Proteopedia / WikiGenes / WikiPathways / …
  - http://biodatabase.org/index.php/BioWiki
- Bioinformatics.Org wiki: http://bifx.org/wiki
- The SEQanswers wiki: http://SEQwiki.org
- MCB: http://wikipedia.org/wiki/Wikipedia:Project_MCB
- BiO Sites: http://BiO.CC

References

- See references within:
  - http://www.ncbi.nlm.nih.gov/pubmed/20624717
  - http://www.ncbi.nlm.nih.gov/pubmed/20193066
  - http://www.ncbi.nlm.nih.gov/pubmed/18613750

- Semantic MediaWiki:
  - http://semantic-mediawiki.org
  - irc://irc.freenode.net/#semantic-mediawiki