

Datasets – the role of the Publisher

June 2011 Philippe Terheggen Senior Vice President, Journals, ELSEVIER



Early innovation



- 1638 Elzevir published Galileo Galilei's greatest work
- Galileo published his "*Discorsi e dimostrazioni matematiche, intorno a due nuove scienze*" with Elzevir despite being banned by the Inquisition.

"An elsevier" was a trademark for any book with a certain portable format.

Why data are relevant



The screenshot shows two overlapping windows. The left window is an Adobe Reader displaying a document titled "HBR Next Scientific Revolution.pdf". The document content includes the title "The Four Paradigms of Science" and four sections: "Experimentation", "Theory", "Computation and Simulation", and "Data Mining". The right window is a web browser displaying a page titled "The Next Scientific Revolution" by Tony Hey. The page has a sub-header "THE BIG IDEA" and a paragraph discussing data mining and its relevance.

The Four Paradigms of Science

Experimentation
Beginning in ancient Greece and China, people tried to explain their world through natural laws instead of supernatural causes.

Theory
By the 17th century, scientists like Isaac Newton tried to make predictions for new phenomena and would verify hypotheses by conducting experiments.

Computation and Simulation
The advent of high-performance computers in the latter half of the 20th century allowed scientists to explore regimes inaccessible to experiment and theory, such as climate modeling or galaxy formation, by numerically solving systems of equations on a large scale and in fine detail.

Data Mining
Using more-powerful computers, scientists begin with the data and direct programs to mine enormous databases for relationships. In essence, they use computers to discover the rules by studying the data.

THE BIG IDEA

The Next Scientific Revolution

by Tony Hey

How data mashups can help save the world.

worked with in New Mexico discovered \$10 million in underpayments within the first six months of using such data-mining tools.

The relevance of the old joke “only half of all advertising dollars are successful—we just don’t know which half” will be imperiled by the new analytical tools. An electronic entertainment company in the Philippines is using Microsoft data-mining technology to customize its sales pitches to individual customers, based on extensive analysis of such factors as past buying patterns, age, gender, financial profile, and location. Almost immediately after

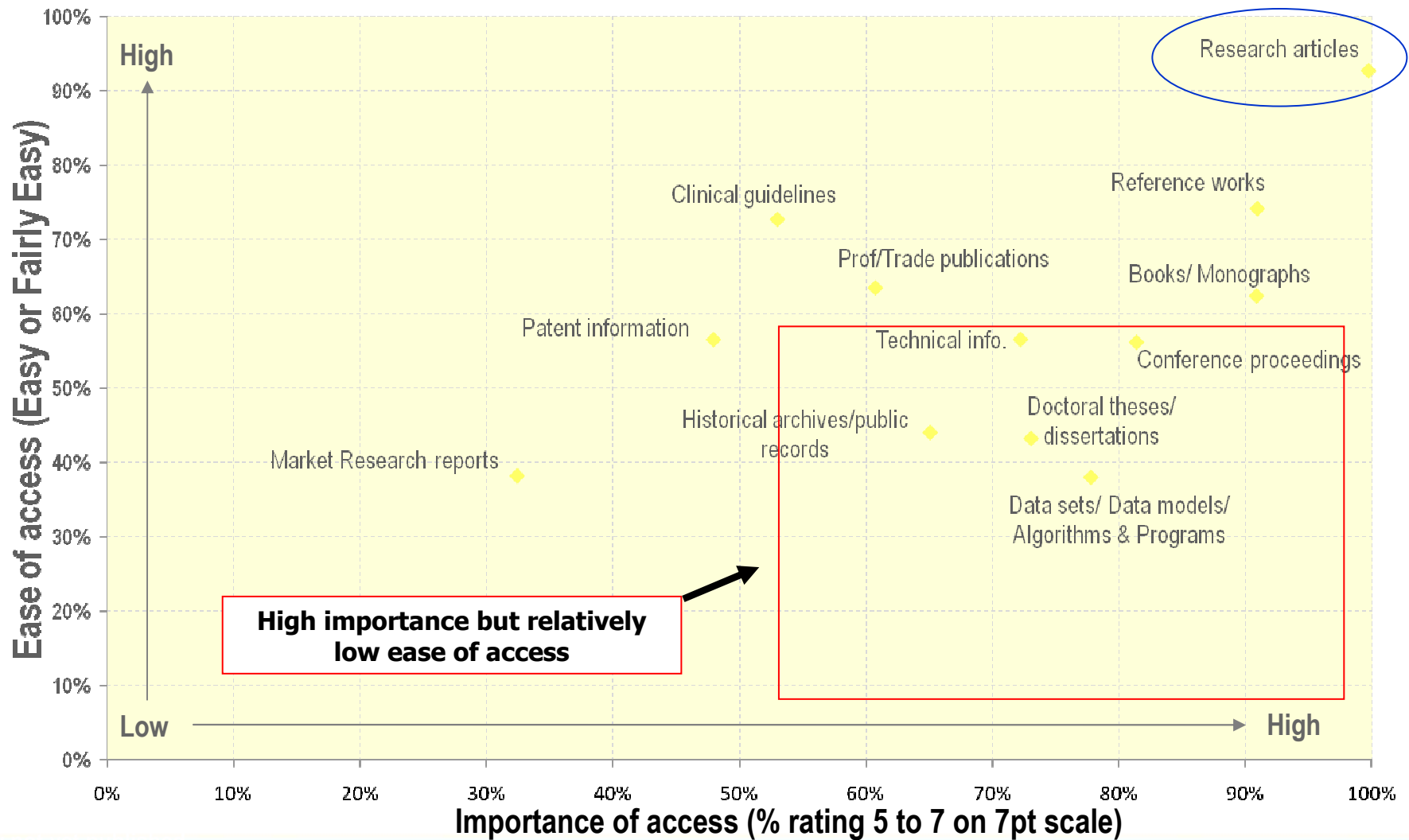
Harvard Business Reviews. Toney Hey, 2010

Access vs. Importance

(4109 respondents)



All Respondents



High importance but relatively low ease of access

Research articles

Topics



1. Overview of existing collaborations.
2. Our perspective on datasets
3. Our open invitation to collaborate further

Data enrichment in publications



1. Supplemental data

2. Article linking

- PANGAEA, CCDC, EarthChem, ...

3. Entity linking

- Genes, Proteins, CCDC Structures, Arabidopsis, ...

4. Embedded applications

- Reaxys, PANGAEA, Protein Viewer, Google Maps, ...

Our efforts on article linking are part of much wider initiative on article innovation...



1. Supplemental data – current situation

- Elsevier receives more and more supplemental materials in different formats to publish next to articles (40% 4-yr CAGR)
- The types of supplemental material vary widely per discipline (e.g. methods/protocols, datasets, executable codes, videos of experiments, etc) and vary also in importance to support the argument made in the article

	2005	2006	2007	2008	2009	CAGR
Applications*	7,499	12,670	17,217	23,492	29,349	41%
Audio	20	76	96	51	137	62%
Images	3,009	5,154	6,339	9,579	13,331	45%
Video	1,890	2,513	3,400	3,908	4,749	26%
Total	12,418	20,413	27,052	37,030	47,566	40%

* Applications: .pdf, .csv, .txt, .tar, .zip, .doc, .rtf, .xls, .ppt

2. Article Linking



yl; NHC ligand

- Nickel Complexes with Carbonyl, Isocyanide, and Carbene...
Comprehensive Organometallic Chemistry III
- Platinum Complexes with Carbonyl, Isocyanide, and Carbe...
Comprehensive Organometallic Chemistry III
- HERBICIDES
Encyclopedia of Analytical Science
- Functions Containing Doubly Bonded P, As, Sb, Bi, Si, G...
Comprehensive Organic Functional Group Transformations ...

▶ More related reference work articles

Relevant terms from this article NEW
Click for **Data Correlations**, **Clinical Trials** and more

19 Compounds

hexane	toluene
imidazole	ethylene

[View more...](#)

Powered by NextBio [What is this?](#)

Crystallographic Data

le (10)
lazole (12)

- Author posted
- Automatic connection between Elsevier and repository



CCDC Structure Request

CCDC Structure Summary for 734053

J. Fluorine Chem., (2009), 130, 966, doi: [10.1016/j.jfluchem.2009.07.015](https://doi.org/10.1016/j.jfluchem.2009.07.015)

From this page you can either access entries in [WebCSD](#) or request the deposited CIF data for selected structures.

If CIF data are not available, your request will be forwarded to a scientific editor at the CCDC

Select	CCDC No	a	b	c	Space Group	CIF available	View in WebCSD*
<input checked="" type="checkbox"/>	734053	10.468	11.9179	14.0694	P-1	yes	TUJZEE

[Request selected CIFs](#)

[View in WebCSD](#) *If your organisation has an [unlimited site licence for the CSD System](#) you will be able to access these entries in WebCSD

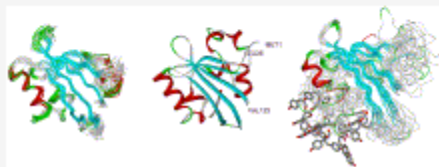
Served by '10.0.0.8'. © [Cambridge Crystallographic Data Centre](#) 2006-2009



3. Entity Linking



Upon irradiation at 446 nm, the chromophore switches to the cis form, and the protein partially unfolds [33] and [34] (Fig. 1). In the dark, the trans folded state is regained in seconds to minutes, depending on experimental conditions.³⁵ Several studies have shown that the N-terminal section (residues 1–25) of PYP undergoes an almost complete unfolding upon irradiation, exposing a hydrophobic patch on the protein.^{[36], [37], [38], [39] and [40]} The N-terminally truncated protein (Δ 1–25) is still fully functional, albeit with a slower photocycle.³³



Full-size image (72K)
High-quality image (755K)

Fig. 1.

Structures of PYP under various conditions. (a) NMR solution structure of dark-adapted Δ 25PYP (PDB code 1XFN).³³ (b) X-ray crystal structure of dark-adapted full-length wild-type PYP (the N-terminal ca. 25 residues are missing) (PDB code 1NWZ).³² (c) NMR solution structure of light-adapted Δ 25PYP (PDB code 1XFQ).³³

Our approach to using PYP as a photoswitch aims to use the folded state of PYP to sterically prevent interaction, which is then allowed to occur in the flexible light-adapted form. It has been suggested that such a mechanism may explain how PYP interacts with its putative partner protein *in vivo*.³⁸ We wish to explore the possibility of designing a genetically encoded photoswitchable DNA binding protein by fusing PYP to the prototypical leucine-zipper-type DNA binding protein GCN4-bZIP. The bZIP domain is

- Automatic or author-tagged
- Automatic linked to data:
 - External
 - Selectable
 - Personal preference

4. Embedded Applications: PDB



- Author-tagged
- Data from PDB
- 3D Visualisation
 - Select
 - Zoom
 - Rotate
- All inside article

the metal ions of the catalytic center. This is in contrast to a proposal for *Desulfovibrio* sp. hydrogenases that show a di-oxo species in this position for the Ni-A state. The additional metal site located in the large subunit appears to be a Mg^{2+} ion. Three iron-sulfur clusters were found in the small subunit that forms the electron transfer chain connecting the catalytic site with the molecular surface. The calculated anomalous Fourier map indicates a distorted proximal iron-sulfur cluster in part of the crystals. This altered proximal cluster is supposed to be paramagnetic and is exchange coupled to the Ni^{3+} ion and the medial $[Fe_3S_4]^+$ cluster that are both EPR active ($S = 1/2$ species). This finding of a modified proximal cluster in the $[NiFe]$ hydrogenase might explain the observation of split EPR signals that are occasionally detected in the oxidized state of membrane-bound $[NiFe]$ hydrogenases as from *A. vinosum*.

Protein Viewer

Keywords: $[NiFe]$ hydrogenase; *Allochroa*; *vinosum*; photosynthetic purple sulfur bacterium; iron-sulfur cluster; Ni-A state

optical atomic emission spectrometry; MAD, multiwavelength anomalous dispersion; D. *vulgaris* H, *Desulfovibrio vulgaris* Hilgenberg; D. *vulgaris* MF, *Desulfovibrio vulgaris* Miyazaki F; PDB, Protein Data Bank

This article has not yet been cited.

Related reference work articles e.g. encyclopedias

- Photosystem I, FX, FA, and FB Iron-Sulfur Clusters
Encyclopedia of Biological Chemistry
- Hydrogen Activation
Comprehensive Coordination Chemistry II
- Radical Reactions Featuring Lysine 2,3-Aminomutase
Comprehensive Natural Products Chemistry
- Complexes of Titanium in Oxidation State III
Comprehensive Organometallic Chemistry III
- Respiratory Chain Complex I
Encyclopedia of Biological Chemistry

More related reference work articles

Relevant terms from this article NEW

Click for [Data Correlations](#), [Clinical Trials](#) and more

10 Genes & Proteins

calc	Mcl1
Mcl4	Mcl2

[View more...](#)

11 Compounds

oxygen	amino acid
mg2+	ni2+

[View more...](#)

3 Tissues & Cells

distal cleavage	shoulder
---------------------------------	--------------------------

5 Biogroups

catalytic activity	cytochrome b
cytochrome c	metabolic processes

[View more...](#)

10 Organisms

desulfovibrio	bacteria
diatomic	desulfovibrio vulgaris

[View more...](#)

Powered by NextBio [What is this?](#)

Contact Us | Print

PDB ID or Text

Search

Advanced Search

MyPDB Hide

Login to your Account
Register a New Account

Home Hide

News & Publications
Usage/Reference Policies
Deposition Policies
Website FAQ
Deposition FAQ
Contact Us
About Us
Careers
External Links
Sitemap
New Website Features

Deposition Hide

All Deposit Services
Electron Microscopy
X-ray | NMR
Validation Server
BioSync Beamline
Related Tools

Search Hide

Advanced Search
Latest Release
New Structure Papers
Sequence Search

Summary Sequence Derived Data Seq. Similarity 3D Similarity Literature Biol. & Chem. Methods Geometry Links

structure of the blue shifted intermediate state of the photoactive yellow protein lacking the N-terminal part

1XFQ

- Display Files
- Download Files
- Print this Page
- Share this Page

DOI:10.2210/pdb1xfq/pdb

Primary Citation

The solution structure of a transient photoreceptor intermediate: Delta25 photoactive yellow protein.

Bernard, C., Houben, K., Derix, N.M., Marks, D., van der Horst, M.A., Hellingwerf, K.J., Boelens, R., Kaptein, R., van Nuland, N.A.

Journal: (2005) STRUCTURE 13: 953-962

PubMed: 16004868

DOI: 10.1016/j.str.2005.04.017

Search Related Articles in PubMed

PubMed Abstract:

The N-terminally truncated variant of photoactive yellow protein (Delta25-PYP) undergoes a very similar photocycle as the corresponding wild-type protein (WT-PYP), although the lifetime of its light-illuminated (pB) state is much longer. This has allowed determination of the structure of both...

[Read More & Search PubMed Abstracts]

Structure Image



More Images...

View in Jmol

SimpleViewer

Embedded Application in: PANGAEA



- Author-posted
- PANGAEA-hosted
- Real-time match
- Full Google Maps functionality
- Links to PANGAEA
- All inside article

Thumbnails | Full-Size images

s nannofossil

istocene time interval has been tested
tude deep-sea successions. The
nstructed in critical intervals, and the
G); last occurrence of *Calcidiscus*
e *Gephyrocapsa* spp. (t1G); first
pp. (reemG) and last occurrence of
zons were used to generate a detailed
by means of correlation to isotope
orizons was evaluated. Biohorizons t1G
ectively, and hmG and blG are
are confirmed as diachronous on the

PANGAEA™ – Supplementary Data
Revision of the early-middle Pleistocene calcareous nannofossil biochronology

Hybrid

Raffi, I (2002): (Table 4, 5) Occurrence and age estimations of nanofossil biohorizons of ODP Site 154-926 in the western equatorial Atlantic. doi:10.1594/PANGAEA.690443

North America
South America
Africa

Imagery ©2010 NASA - Terms of Use

Related Articles

- Glacial-interglacial ocean climate variability from pla...
Palaeogeography, Palaeoclimatology, Palaeoecology
- Comparative molecular biomarker assessment of phytopl...
Organic Geochemistry
- Calcareous nannofossil bioevents and environmental cont...
Marine Micropaleontology
- Habitats, abundance patterns and isotopic signals of mo...
Marine Micropaleontology
- Pleistocene calcareous nannofossil stratigraphy for ODP...
Marine Micropaleontology

▶ View more related articles



Embedded Application: BrainLink

- Automatically generated
- Text-mining and linking to structures
- Lists and displays all structures:
 - Occurrences
 - 2D, 3D images
 - Rat, mouse, monkey, human views
- Again inside article
- Links to more data at BrainNavigator

BrainLink Powered by Brain Navigator BETA

Information about brain structures in this article

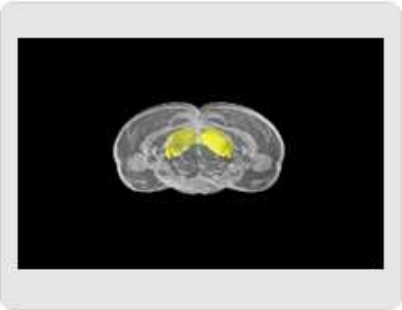
inferior colliculus (4) Back to List

[Search in ScienceDirect](#) | [View in Brain Navigator](#)
[2D Images](#) | [3D Images](#) | [Mentioned In Article \(4\)](#)

[Mouse](#) | [Rat](#) | [Monkey](#) | [Human](#)

The Rat Brain in Stereotaxic Coordinates by George Paxinos and Charles Watson

Coronal



Sagittal

Related Articles

- [Subsets of acetylcholine-stimulated 88Rb+ efflux and \[1... Neuropharmacology](#)
- [Motor activity and alcohol: Genetic analysis in the mou... Physiology & Behavior](#)
- [Genetic analysis of hypothalamic retention of \[3H\]corti... Brain Research](#)
- [A genetic analysis of morphine responses in mice](#)

Genetic analysis of nicotinic binding

Las Encinas 3370,

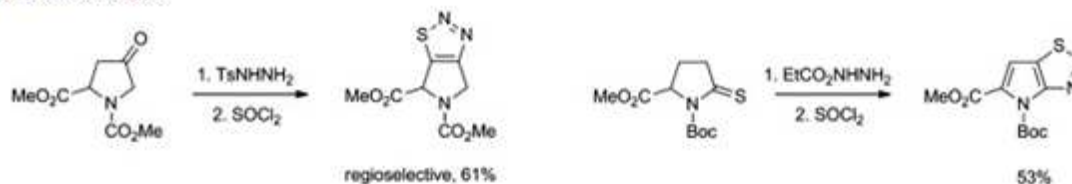
9.

acetylcholine receptors
 position 3 of the pyridone ring
 expressed in the brain.
 using wild type, $\beta 2$, and $\beta 4$ -
 tyrosine (asterisk indicates the
 line (3-IC) induced considerable
 lamina and partially inhibited
 mice. $\beta 2$ -null mice were


Embedded Application: chemistry

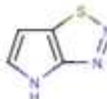
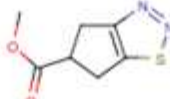




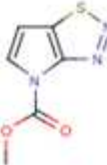
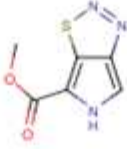






the more than 1000. The regioselectivity of the cyclization has been studied and some were established to predict the cyclization direction to afford bicyclic 1,2,3-thiadiazoles. Effects promoting and disfavoring the reaction have also been investigated to guide the synthesis of scaffolds of this type.

Graphical abstract



Author's Key Structures in this Article

powered by 

	C ₄ H ₃ N ₃ S (125.154)		C ₇ H ₉ N ₂ O ₂ S (184.219)
Search in Reaxys	Search in Reaxys	Search in Reaxys	Search in Reaxys
InChIKey Search	InChIKey Search	InChIKey Search	InChIKey Search
			
	C ₅ H ₅ N ₃ O ₂ S (183.191)		C ₆ H ₅ N ₃ O ₂ S (183.191)
Search in Reaxys	Search in Reaxys	Search in Reaxys	Search in Reaxys
InChIKey Search	InChIKey Search	InChIKey Search	InChIKey Search
			
	C ₅ H ₇ N ₃ O ₄ S (241.227)		C ₆ H ₅ N ₃ O ₂ S (183.191)

Keywords: Annelated heterocycle; Cyclization; Protecting group; Regioselectivity; Systemic acquired resistance
 BKEJQHMFJQCUTKL-UHFFFAOYSA-N; TYPJSLXIPNEPES-UHFFFAOYSA-N; LSLFJEGXOZLMIB-UHFFFAOYSA-N;



- We carry a general vision to increase discoverability of science – by universal access and integration to data and tools.
- We invite authors to publisher supplemental material together with their article.
- We invite research organisation/database owners to collaborate on the linking of research data to articles, or to relevant entities in published articles, and, from there the creation of relevant applications.
- We express a “call for applications” in all areas of science and technology to use our database of publications and data for making innovative information products and workflow solutions.