CrossRef, DOIs and Data: A Perfect Combination

Ed Pentz, Executive Director, CrossRef

CODATA '06
Session K4
October 25, 2006

Research is changing

- Researcher nirvana: all scientific data from formal published literature, theses and dissertations, datasets and grey literature all digitized, cross-linked and full-text searchable creating new opportunities for discovery in research across disciplines and content types

Or, to put it more eloquently…

*The opportunities are truly stunning. They point towards entirely new ways to think about the scholarly literature (and the underlying evidence that supports scholarship) as **an active, computationally enabled representation of knowledge that lives, grows and interacts with its contributors rather than as a passive archive or record. They suggest ways in which information technology can accelerate the rate of scientific discovery and the growth of scholarship…***
- Clifford Lynch 2006

Scholarly Publishing Trends

- Everything is online – if it's not online, it doesn't exist
- Everything is interlinked – if it's not linked it doesn't exist
- Journals are becoming more like databases
  - Articles are standard, structured entries
  - List chemical entities, genetic sequences
- Databases are becoming more like journals
  - Different levels of review, citations
- Interlinking is essential

CrossRef's Role

- Non-profit association of publishers that:
  - Assigns unique IDs – DOI numbers – to content to journal articles, book chapters, conference proceedings, standards and databases
  - Collects metadata in a central database
  - Enables reference linking among thousands of journals and publishers
  - Disseminates metadata to many different organizations

Why DOIs?

- Persistent IDs
  - http://dx.doi.org/ resolves all DOIs
  - DOI has associated URL and URL can be easily updated
  - Other IDs can be use as DOI – LSID, InCHI
- Scalable infrastruce – very flexible
- Social Infrastructure
  - Agreements in place, patent policy, obligations to support persistence

CrossRef keys to success

- Collaboration: 400 publishers, 14,000 journals, 23 million content items – critical mass
- Standards: use of DOI numbers, XML, OpenURL, OAI-PMH
- Level playing field for publishers: small and large, commercial and non-profit, open access and traditional and everything in between

56. Kruger, D.H., Kupper, D., Meisel, A., Reuter, M., Schroeder, C. (1995) The significance of distance and orientation of restriction endonuclease recognition sites in viral DNA genomes *FEMS Microbiol. Rev.*, **17**, 177–184[CrossRef][ISI][Medline] .

57. Saha, S. and Rao, D.N. (1997) Mutations in the Res subunit of the EcoPI restriction enzyme that affect ATP-dependent reactions *J. Mol. Biol.*, **269**, 342–354[CrossRef][ISI][Medline] .

Article Views

› Abstract

› Full Text (HTML)

› Full Text (PDF)

Article Tools

Manage This Article

› Submit an E-Letter

› Add to Personal Folders

› Download to Citation

< Prev | Table of Contents | Next >

REPORTS

Kinetic Measurement of the Step Size of DNA Unwinding by *Escherichia coli* UvrD Helicase

Janid A. Ali and Timothy M. Lohman [*]

The kinetic mechanism by which the DNA repair helicase UvrD of *Escherichia coli* unwinds duplex DNA was examined with the use of a series of oligodeoxynucleotides with duplex regions ranging from 10 to 40 base pairs.

Alice

Mar

DNA

Brist

Beve

"To w

928 82

65. Francino, M.P. and Ochman, H. (1997) Strand asymmetries in DNA evolution *Trends Genet.*, **13**, 240–245[CrossRef][ISI][Medline] .

66. Frank, A.C. and Lobry, J.R. (1999) Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms *Gene*, **238**, 65–77[CrossRef][ISI][Medline] .

67. Beletskii, A., Grigoriev, A., Joyce, S., Bhagwat, A.S. (2000) Mutations induced by bacteriophage T7 RNA polymerase and their effects on the composition of the T7 genome *J. Mol. Biol.*, **300**, 1057–1065[CrossRef][ISI][Medline] .

68. Studier, F.W. (1975) Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host *J. Mol. Biol*, **94**, 283–295[CrossRef][ISI][Medline] .

Journal home
Advance online publication
ᴸ About AOP
Current issue
Archive

Supplements
Web focuses
Multimedia
About the journal
For authors and referees
Online submission
Reprints
Conferences
Help

Gateways
Asia gateway
German gateway
Omics gateway

NPG Journals
by Subject Area

Chemistry
Chemistry
Drug discovery
Biotechnology

Clinical Practice & Research
Cancer
Cardiovascular medicine
Dentistry
Endocrinology - NEW
Gastroenterology & Hepatology

Advance online publication

The latest research papers brought to you ahead of print publication by *Nature* AOP. Advance online publication papers are listed below and grouped by their publication date online.

About advance online publication

24 May 2006

Distribution and three-dimensional structure of AIDS virus envelope spikes
Ping Zhu, Jun Liu, Julian Bess, Jr, Elena Chertova, Jeffrey D. Lifson, Henry Grisé, Gilad A. Ofek, Kenneth A. Taylor and Kenneth H. Roux

doi:10.1038/nature04817

Abstract | Full Text | PDF (1,874K) | Supplementary information

21 May 2006

Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF
Haitao Li, Serge Ilin, Wooikoon Wang, Elizabeth M. Duncan, Joanna Wysocka, C. David Allis and Dinshaw J. Patel

doi:10.1038/nature04802

First paragraph | Full Text | PDF (1,929K) | Supplementary information

Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2
Pedro V. Peña, Foteini Davrazou, Xiaobing Shi, Kay L. Walter, Vladislav V. Verkhusha, Or Gozani, Rui Zhao and Tatiana G. Kutateladze

doi:10.1038/nature04814

First paragraph | Full Text | PDF (1,061K) | Supplementary information

Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch
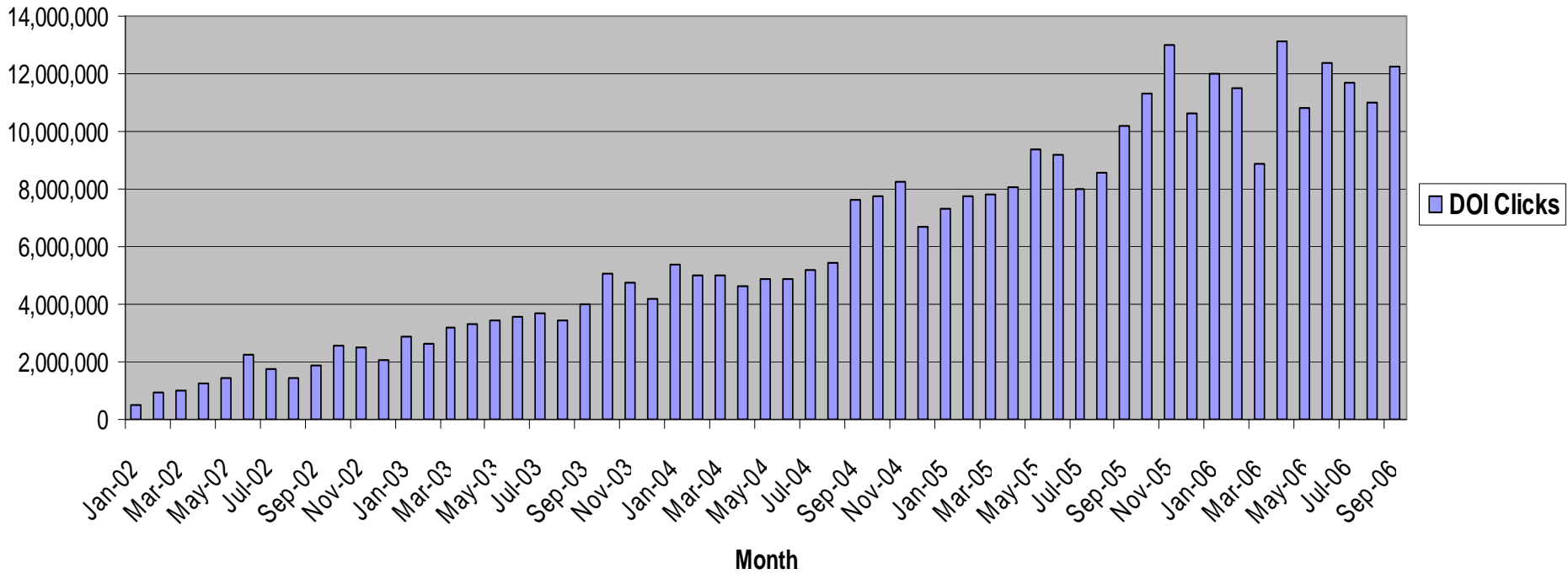
End Users are clicking!

DOI Clicks

Apply CrossRef model

- German National Library of Science and Technology (TIB)
  - May 2002 – ICSU Report "Concept of Citing Primary Data"
  - CODATA support
  - Project Website - http://www.std-doi.de/
  - TIB now DOI Registration Agency
- Make data available and discoverable
- Make data citeable like articles

Scientific Drilling Database
Data from Deep Earth Sampling and Monitoring

Dataset Description

Citation: Kamm, H; Machon, L; Donner, S; (2004): Gas mass spectrometry of gas samples from the KTB Main Hole HB1. *Scientific Drilling Database*. doi:10.1594/GFZ.SDDB.1071

[Download Citation (EndNote)]

DOI: 10.1594/GFZ.SDDB.1071

Title: Gas mass spectrometry of gas samples from the KTB Main Hole HB1

Abstract: The main objective of this drilling fluid analysis was the detection of inflows of formation fluids. Therefore different gases dissolved in the drilling mud were measured continuously and automatically at drill site with three different methods (Fig.: KTB-Report 92-2 page C13). The operation principles of the mass spectrometer and the gaschromatograph have been explained by STROH et al. (1988) and FIGGEMEIER et al. (1991). The principle of radon determination is published by ERZINGER et al. (1992). In the complete KTB-VB and in in the KTB-HB down to a depth of 3003 m the gas phase was released and collected by twirl degassers attached in front of the mud shakers. This open system led to gas losses as well as air contamination. Therefore results obtained down to this depth have only qualitative character. After casing the KTB-HB to a depth of 3003 m a bypass system was installed at the BOP (blow-out preventer) 50 cm below the flow line.

[Show in Google Earth]

Activities: **KTB-HB**

Latidude:	49.588
Longitude:	12.192
Elevation:	
Date/Time:	1990-10-06 00:00:00
Program:	Bundesministerium für Bildung und Forschung
Expedition:	KTB
Platform:	Land based
Gear:	gear not specified

Datapoints: 44648

Parameter(s):

	Parameter [Unit]	Principal Investigator	Method
☐	N2 [%]	H Kamm	gas chromatography

GFZ POTSDAM   icdp

Kam   Lab),
GFZ

AfCS nature

the**signaling**gateway

Protein A003687

Overview

Author-entered Data
V1.0, Peer Reviewed
Published 18 Oct 2006

Summary
Network Map
States
Transitions
Functions
Protein Classes

Automated Data
Not Reviewed
As At Publication

Protein Records
Interaction DBs
Domains & Motifs
Protein Structure
Gene Info
Orthologs
Blast Data

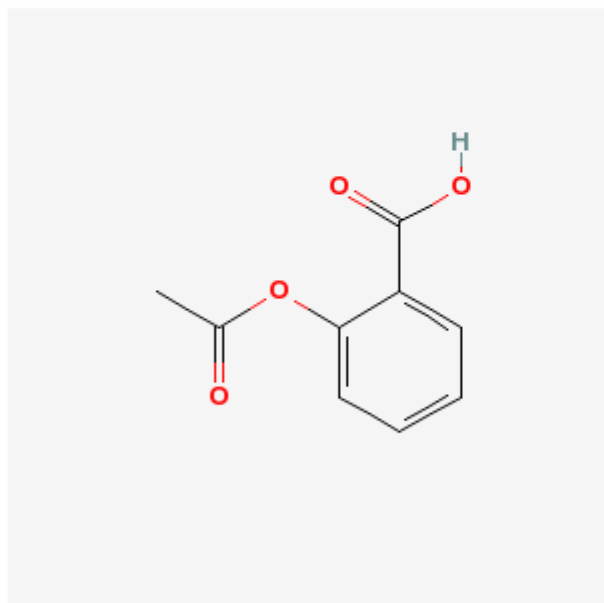# Endo180

Molecule Page Overview
Version 1.0, Peer Reviewed And Published 18 Oct 2006

doi:10.1038/mp.a003687.01   How to cite this Molecule Page

AfCS ID	A003687
AfCS Name	Endo180
All Names	Endo 180; Endo180; Mannose receptor, C type 2; Mrc2; Novel lectin; uPARAP
Functional Category	Receptor, misc.
Primary Symbol	Mrc2
Molecule Page Version	Version 1.0, Peer Reviewed And Published 18 Oct 2006
Corresponding Author	**Clare M Isacke clare.isacke@icr.ac.uk**
Authors	Dirk Wienke, Justin Sturge, Niina J Pirinen, Luke A Henry, Gareth C Davies, Clare M Isacke
Editorial Board Member	Patrick J Casey casey006@mc.duke.edu (editorial board member list)

Species	Mouse
Sequence	1 MVPIRPALAP WPRHLLRCVL LLGGLRLGHP ADSAAALLEP DVFLIFSQGM 51 QGCLEAQGVQ VRVTPVCNAS LPAQRWKWVS RNRLFNLGAT QCLGTGWPVT 101 NTTVSLGMYE CDREALSLRW QCRTLGDQLS LLLGARASNA SKPGTLERGD 151 QTRSGHWNIY GSEEDLCARP YYEVYTIQGN SHGKPCTIPF KYDNQWFHGC 201 TSTGREDGHL WCATTQDYGK DERWGFCPIK SNDCETFWDK DQLTDSCYQF 251 NFQSTLSWRE AWASCEQQGA DLLSITEIHE QTYINGLLTG YSSTLWIGLN 301 DLDTSGGWQW SDNSPLKYLN WESDQPDNPG EENCGVIRTE SSGGWQNHDC 351 SIALPYVCKK KPNATVEPIQ PDRWTNVKVE CDPSWQPFQG HCYRLQAEKR 401 SWQESKRACL RGGGDLLSIH SMAELEFITK QIKQEVEELW IGLNDLKLQM 451 NFEWSDGSLV SFTHWHPFEP NNFRDSLEDC VTIWGPEGRW NDSPCNQSLP 501 SICKKAGRLS QGAAEEDHGC RKGWTWHSPS CYWLGEDQVI YSDARRLCTD 551 HGSQLVTITN RFEQAFVSSL IYNWEGEYFW TALQDLNSTG SFRWLSGDEV 601 IYTHWNRDQP GYRRGGCVAL ATGSAMGLWE VKNCTSFRAR YICRQSLGTP 651 VTPELPGPDP TPSLTGSCPQ GWVSDPKLRH CYKVFSSERL QEKKSWIQAL 701 GVCRELGAQL LSLASYEEEH FVAHMLNKIF GESEPESHEQ HWFWIGLNRR 751 DPREGHSWRW SDGLGFSYHN FARSRHDDDD IRGCAVLDLA SLQWVPMQCQ

NCBI

Pub☐hem

National Library of Medicine NLM

Search | PubChem Compound | | GO

Compound Summary:

CID: 2244 ⍰

BioActivity: Summary ⍰

All: 83 Links
Inactive: 79 Links
Inconclusive: 3 Links

Protein Structures: 3 Links ⍰

Protein Sequences: 51 Links ⍰

NLM Toxicology: Link ⍰

Substances: ⍰

All: 120 Links
Same: 43 Links
Mixture: 77 Links

Related Compounds: ⍰

Same, Connectivity: 2 Links

Similar Compounds: 354 Links ⍰

Structure Search ⍰

| Mesh | Synonyms | Properties | Descriptors | Category | Exports |

What needs to happen?

- Databases need to assign DOI numbers to entries

- Databases need to assign standardized journal-like metadata

- Journals and databases need to interlink – they don't at the moment!

Conclusion

- Towards truly dynamic linking and a better online reading environment for scholars, with
  - Robust linking through journals
  - Robust linking between journals and databases
  - Databases to use DOI numbers and standard metadata
  - Databases to be citeable
- Goal? Researcher nirvana

epentz@crossref.org