



Federated Repositories of X-ray Diffraction Images

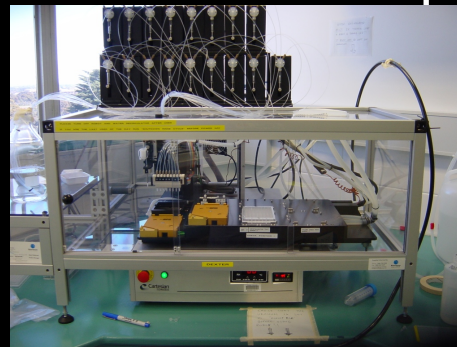
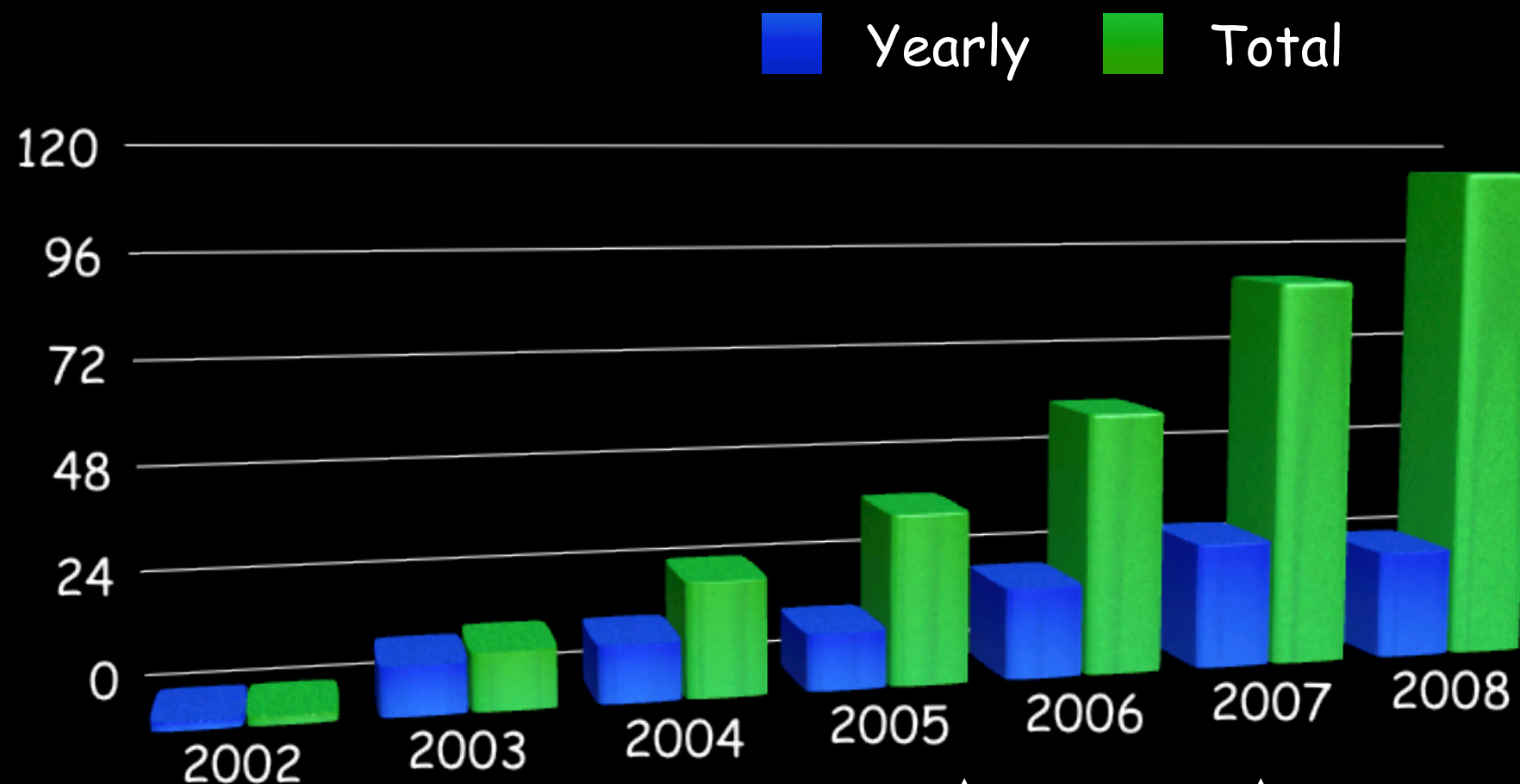
Ashley Buckle

Dept of Biochemistry & Molecular Biology
Monash University, Melbourne, Australia

IUCR, Osaka, August 30th

Protein Crystallography at Monash

Number of protein crystal structures determined

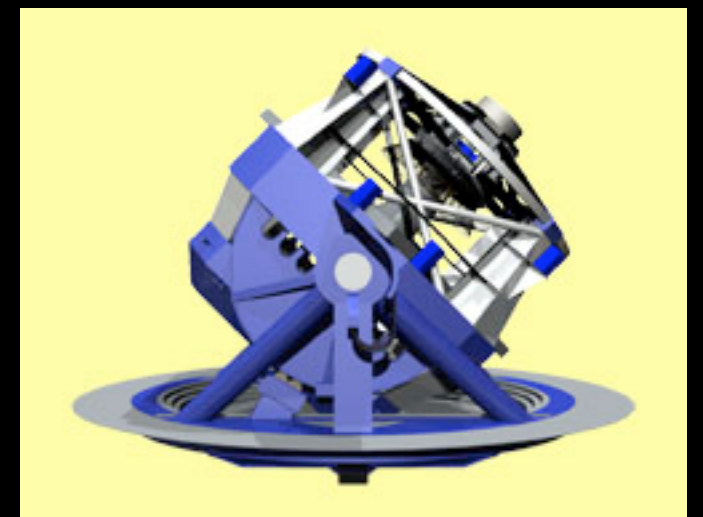


Too much raw data?

Astronomers dont think so...

"....Using these solutions as first steps, the ultimate goal is to meet the needs of the Large Synoptic Survey Telescope (LSST); when LSST begins operation in 2013 it will generate an estimated 15TB of raw data and more than 100TB of processed data every night, 365 days a year. "

source: <http://www.gridtoday.com/grid/803617.html>



Who stores the raw data?

Laboratory



Pros:

- simple
- cheap

Cons:

- media failure
- staff turnover
- lost disks
- lack of persistence

Institutional libraries



Pros:

- already storing research data
- new role of library
- federated
- economical and scalable
- persistence

Cons:

- data formats - standards?
- data discovery

wwPDB



Pros:

- centralised
- global solution
- standardisation

Cons:

- large data size
- cost



Fedora Commons provides sustainable technologies to **create, manage, publish, share and preserve digital content** as a basis for intellectual, organizational, scientific and cultural heritage by bringing two communities together.

Communities of practice that include scholars, artists, educators, Web innovators, publishers, scientists, librarians, archivists, records managers, and museum curators who present, access, or preserve digital content.

Software developers who work on the cutting edge of open source Web, semantic and enterprise content technologies to ensure that collaboratively created knowledge is available now and in the future.

Open Source Technology

Storage Repository technologies to ensure longevity and integrity for any



[Download Fedora 3.0](#)
[Download Fedora 2.2.3](#)



[View the Fedora Commons Video](#) First In a Series

Top Ten Reasons to Choose Fedora

Click on any item to learn more.

1. [Store whatever you want](#)
2. [Easy access](#)
3. [Express relationships](#)
4. [Free open source software](#)
5. [Enables permanence](#)

HatCheck Newsletter

[\[html\]](#) [\[pdf\]](#)

News



[New Fedora Commons Wiki](#)

28 August 2008, 2:24 pm

[RepoCamp at the Library of Congress](#)

20 August 2008, 3:58 pm

[NOW AVAILABLE: Fedora 2.2.3](#)

Why Fedora?

- Used as a library repository worldwide
 - 148 known as of August 2008
- Examples:
 - eSciDoc Project (key repository for the Max Planck Society)
 - National Science Digital Library (US National Science Federation)
- standards compliant
- open source

SEARCH

Project ingested into
repository with packaged,
annotated datasets, ready to
be discovered by TARDIS

Create Project

Project Descriptor

Description File Output Directory

PDB ID:

Project Title:

Author's Names (separate by ','):

Project Description:

An incomplete project for example purposes only.

Lysins are peptidoglycan hydrolases that are produced by bacteriophage and act to lyse the bacterial host cell wall during progeny phage release. Here, we describe the structure and function of a novel bacteriophage-derived lysin, PlyB, which displays potent lytic activity against the *Bacillus anthracis*-like strain ATCC 4342. This molecule comprises an N-terminal catalytic domain (PlyB(cat)) and a C-terminal bacterial SH3-like domain, SH3b. It is shown that both domains are required for effective catalytic activity against ATCC 4342. Further, PlyB has specific activity comparable to the phage lysin PlyG, an amidase being developed as a therapeutic against anthrax. In contrast to PlyG, however, the 1.6 Å X-ray crystal structure of PlyB(cat) reveals that the catalytic domain adopts the glycosyl hydrolase (GH)-25, rather than phage T7 lysozyme-like fold. PlyB therefore represents a new class of anthrax lysin and a new defensive tool in the armament against anthrax-mediated bioterrorism.

Primary Citation:

Package Data

Dataset Packager

Dataset Image Directory Path

Packaged Dataset Output Directory

Dataset Name (eg. Native 2.0 Angstroms):

Crystal Name:

Resolution Limit:

Diffraction Type:

Xray Source:

Mosaic Spread (leave blank if none) Chi Angle (leave blank if none)

Split Dataset Maximum Size mb

Note: Contact your repository administrator for an appropriate size. If unknown, try 1500 mb

Upload Data

Project Depositor

Repository Address:

Username: Password:

Project Description File

Project Upload Path

Reminder: Make sure your Project Upload Path contains all data to be uploaded, including mtz/sca files.

Data in Library repository

Monash University ARROW Repository | Home

http://arrow.monash.edu.au/vital/access/manager/Index?query=dehalogenase

Monash University ARROW Repository

Home | Browse | Contact | Help | Sign In

Search

Advanced Search Show All 3301

Home

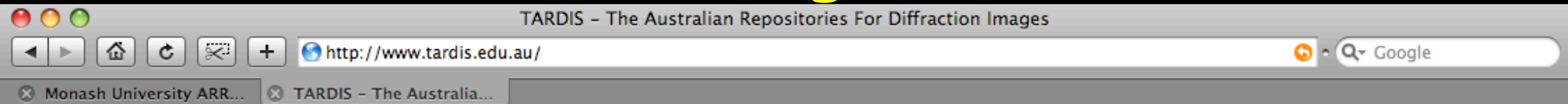
Browse	Collections	Highlights
Title	Centre for Gippsland Studies Historical Pictures	Latest Additions
Creator	Research Data holdings and Datasets	Author Highlights
Subject	PhD eTheses	
Date	Business and Economics Working Papers	
Type	Applied GIS Journal	
	Australian Archive of Jewish Music	

Disclaimer | Privacy | Copyright | Back To Top

Copyright © 2008 Monash University ABN 12 377 614 012 - [Caution](#) - [Privacy](#) - [CRICOS Provider Number: 00008C](#)
Last updated: 20 Febuary 2007 - Maintained by libweb@lib.monash.edu.au - [Accessibility information](#)

arrow powered by VITAL

Search using TARDIS



TARDIS

THE AUSTRALIAN REPOSITORIES FOR DIFFRACTION IMAGES
(Androulakis et al (2008) Acta Cryst. D) [PubMed]

Size of data indexed by TARDIS

Compressed: 20.60 GB Uncompressed: ~80 GB Average Dataset Size (Compressed): 1.48 GB

[HOME](#)[NEWS](#)[ABOUT](#)[DATA](#)[TOOLS](#)[PARTNERS](#) [SEARCH](#)

Welcome

TARDIS is a multi-institutional collaborative venture that aims to facilitate the archiving and sharing of raw X-ray diffraction images (collectively known as a 'dataset') from the Australian protein crystallography community. Whereas the model coordinates and (less often) the structure factors (processed experimental data) are stored in the [PDB](#) the raw diffraction data is often not available. Addressing this shortcoming is the central aim of TARDIS. There are several reasons why we feel this is important (see [about](#)), which can be summarised as:

- The availability of raw data is extremely useful for the development of improved methods of image analysis and data processing.
- Fostering the archival of raw data at an institutional level is one the best ways of ensuring that this data is not lost (laboratory archives are typically volatile)

We have currently finished development on a suite of [deposition tools](#) that allow X-ray diffraction images to be deposited in an open source Fedora repository, and are about to provide discovery services to index the resulting data. Much of this site is under development, however, you can find some basic deposited data [here](#).

+ RECENT NEWS

- TARDIS Portal Out Of Beta
- TARDIS Beta Portal
- Acta Cryst. Paper Online
- Dataset Tools Released
- TARDIS Mentioned In Acta Cryst
- Site Revamp
- Deposition Tools Coming Soon!
- Site Update

Advanced Search

TARDIS

THE AUSTRALIAN REPOSITORIES FOR DIFFRACTION IMAGES
(Androulakis et al (2008) Acta Cryst. D) [PubMed]

[HOME](#)[NEWS](#)[ABOUT](#)[DATA](#)[TOOLS](#)[PARTNERS](#) [SEARCH](#)

Search

Experiment Title / PDB ID

Experiment Description

Institution Name

Author's Name

Detector Type

X-ray Source

Collection Date

From:

To:

Crystal Name

High Resolution Limit

Å...

X-ray Wavelength

Between and Å...

+ TOOLS

- [Search](#)
- [Index](#)
- [Dataset Unpackaging Guide](#)



Data - Index

1SNG - Structure of a Thermophilic Serpin in the Native State

2GPZ - Transthyretin-like protein from Salmonella dublin

2H4R - Crystal structure of wildtype MENT in the native conformation

2NO4 - Crystal structure of the Haloacid Dehalogenase DehIVa from Burkholderia cepacia MBA4

2NW0 - Crystal structure of a lysin

2QP2 - Structure of a MACPF/Perforin-like protein

2R9Y - Structure of antiplasmin

3B3H - The structure of Mycobacterial bacterioferritin

3BJX - Structure of a Group I haloacid dehalogenase from Pseudomonas putida strain PP3

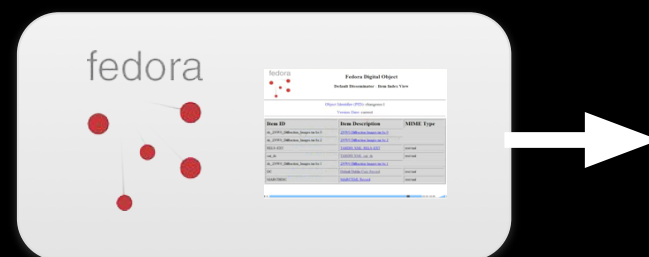
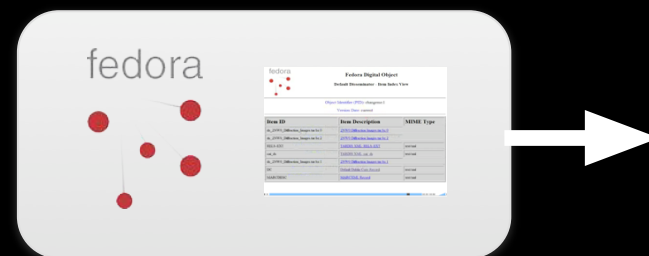
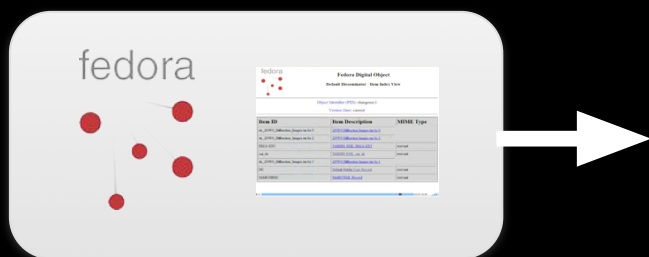
3EB# - Structure of a malarial protease

+ TOOLS

- [Search](#)
- [Index](#)
- [Dataset Unpackaging Guide](#)

Federated approach

Institutional Repositories



TARDIS

TARDIS

THE AUSTRALIAN REPOSITORIES FOR DIFFRACTION IMAGES
(Androulakis et al (2008) Acta Cryst. D) [PubMed]

[HOME](#)[NEWS](#)[ABOUT](#)[DATA](#)[TOOLS](#)[PARTNERS](#) [SEARCH](#)

2QP2 - Structure of a MACPF/Perforin-like protein

Datasets downloaded from this page can be unpackaged by following the [Dataset Unpackaging Guide](#)

MONASH UNIVERSITY

Authors: Buckle Ashley M, Dunstone Michelle A., Law Ruby H. P., Whisstock James C.

Experiment Description:

Protein crystallography raw diffraction images and unmerged reflection intensities Collection size: 36.1 GB Number of datasets: 5 Citation: Rosado et. al. (2007) A common fold mediates vertebrate defense and bacterial attack. Science. In Press.

Proteins containing membrane attack complex/perforin (MACPF) domains play important roles in vertebrate immunity, embryonic development, and neural-cell migration. In vertebrates, the ninth component of complement and perforin form oligomeric pores that lyse bacteria and kill virus-infected cells, respectively. However, the mechanism of MACPF function is unknown. We determined the crystal structure of a bacterial MACPF protein, Plu-MACPF from *Photobacterium luminescens*, to 2.0 angstrom resolution. The MACPF domain reveals structural similarity with pore-forming cholesterol-dependent cytolysins (CDCs) from Gram-positive bacteria. This suggests that lytic MACPF proteins may use a CDC-like mechanism to form pores and disrupt cell membranes. Sequence similarity between bacterial and vertebrate MACPF domains suggests that the fold of the CDCs, a family of proteins important for bacterial pathogenesis, is probably used by vertebrates for defense against infection.

Primary Citation: <http://www.sciencemag.org/cgi/content/abstract/1144706>

+ DOWNLOAD

Dataset 1

- 050107_Native_inhouse.tar.bz.0
- 050107_Native_inhouse.tar.bz.1
- 050107_Unmerged reflection intensities.mtz

Dataset 2

- 080207_EMSA.tar.bz.0
- 080207_EMSA.tar.bz.1

Dataset 3

- 120107_KPtCl.tar.bz.0
- 120107_KPtCl.tar.bz.1
- 120107_Unmerged reflection intensities.mtz

Dataset 4

- 230107_Kl.tar.bz.0
- 230107_Unmerged reflection intensities.mtz

Dataset 5

- 230307_Native_APS.tar.bz.0
- 230307_Unmerged reflection intensities.sca

Ancillary Files

- 080207_Unmerged_reflection_inten

+ TOOLS

- Search

substitution
on (30). Our
n a reversed
lized specific
n to tolerate
ferred a new
erved a more
epistasis, by
another resi-
changes the
s well known
ve effects on
, 33), but we
ifically docu-
s via confor-
ue to the lack
evolutionary
esidues to be
is may be an
on, playing a
ne functions
ions.

l. 15, 355

ion 54, 1079

6).
curr. Opin. Chem.

<http://www.sciencemag.org/cgi/content/full/317/5844/1548>

14 SEPTEMBER 2007 VOL 317 SCIENCE www.sciencemag.org

A Common Fold Mediates Vertebrate Defense and Bacterial Attack

Carlos J. Rosado,^{1,2*} Ashley M. Buckle,^{1*} Ruby H. P. Law,^{1*} Rebecca E. Butcher,^{1,3} Wan-Ting Kan,^{1,2} Catherina H. Bird,¹ Kheng Ung,¹ Kylie A. Browne,⁴ Katherine Baran,⁴ Tanya A. Bashtannyk-Puhalovich,¹ Noel G. Faux,¹ Wilson Wong,^{1,2} Corrine J. Porter,^{1,2} Robert N. Pike,¹ Andrew M. Ellisdon,¹ Mary C. Pearce,¹ Stephen P. Bottomley,¹ Jonas Emsley,⁵ A. Ian Smith,^{1,2} Jamie Rossjohn,^{1,2} Elizabeth L. Hartland,⁶ Ilia Voskoboinik,^{4,7} Joseph A. Trapani,^{4,8} Phillip I. Bird,¹ Michelle A. Dunstone,^{1,6†} James C. Whisstock^{1,2†}

Proteins containing membrane attack complex/perforin (MACPF) domains play important roles in vertebrate immunity, embryonic development, and neural-cell migration. In vertebrates, the ninth component of complement and perforin form oligomeric pores that lyse bacteria and kill virus-infected cells, respectively. However, the mechanism of MACPF function is unknown. We determined the crystal structure of a bacterial MACPF protein, Plu-MACPF from *Photorhabdus luminescens*, to 2.0 angstrom resolution. The MACPF domain reveals structural similarity with pore-forming cholesterol-dependent cytolysins (CDCs) from Gram-positive bacteria. This suggests that lytic MACPF proteins may use a CDC-like mechanism to form pores and disrupt cell membranes. Sequence similarity between bacterial and vertebrate MACPF domains suggests that the fold of the CDCs, a family of proteins important for bacterial pathogenesis, is probably used by vertebrates for defense against infection.

The membrane attack complex/perforin (MACPF) domain was originally identified and named as being common to five complement proteins (C6, C7, C8a, C8b, and C9) and perforin (1–3) (fig. S1). These molecules perform critical functions in innate and

Publication of raw data alongside publication of paper and coordinates

MONASH University
ARROW Repository

<http://arrow.monash.edu/hdl/1959.1/5863>

Home | Browse | Contact | Help |

Home > List Of Titles > 2QP2 - Structure of a MACPF/Perforin-like protein

Buckle, Ashley M, Law, Ruby H. P., Whisstock, James C., Dunstone, Michelle A., - 2QP2 - Structure of a MACPF/Perforin-like protein. 2007.

Please use this identifier to cite or link to this item: <http://arrow.monash.edu.au/hdl/1959.1/5863>

Title **2QP2 - Structure of a MACPF/Perforin-like protein**

Creator Buckle, Ashley M

Creator Law, Ruby H. P.

Creator Whisstock, James C.

Creator Dunstone, Michelle A.

Date 2007

Type dataset

Description Protein crystallography raw diffraction images and unmerged reflection intensities Collection size: 36.1 GB Number of datasets: 5 Citation: Rosado et. al. (2007) A common fold mediates vertebrate defense and bacterial attack. Science. In Press.

Description DATASET_1: Native 2.0 Angstroms Collection date: 2007-03-23 Type of Instrument: Synchrotron (IMCA-CAT beamline, APS, Chicago, USA) Detector: ADSC Oscillation angle: 0.5 deg Exposure time: 15.0 s X-ray wavelength: 1.000000 Angstroms Crystal to detector distance: 210 mm No. of images: 214 Total rotation range: 107 deg Dataset size: 1.7 GB

Description DATASET_2: Native 2.5 Angstroms Collection date: 2007-01-05 Type of Instrument: Rotating-anode (CuKa) Detector: RAXIS Oscillation angle: 0.5 deg Exposure time: 4.0 mins X-ray wavelength: 1.541800 Angstroms Crystal to detector distance: 230 mm No. of images: 509 Total rotation range: 254.5 deg Dataset size: 8.5 GB

Description DATASET_3: Derivative (EMSA) in-house Collection date: 2007-02-08 Type of Instrument: Rotating-anode (CuKa) Detector: RAXIS Oscillation angle: 0.5 deg Exposure time: 10.033334 mins X-ray wavelength: 1.541800 Angstroms Crystal to detector distance: 280 mm No. of images: 468 Total rotation range: 234 deg Dataset size: 7.8GB

Description DATASET_4: Derivative (KI) in-house Collection date: 2007-01-24 Type of Instrument: Rotating-anode (CuKa) Detector: RAXIS Oscillation angle: 0.5 deg Exposure time: 10.033334 mins X-ray wavelength: 1.541800 Angstroms Crystal to detector distance: 260 mm No. of images: 386 Total rotation range: 193 deg Dataset size: 6.5GB

Description DATASET_5: Derivative (KPIC) in-house Collection date: 2007-01-12 Type of Instrument: Rotating-anode (CuKa) Detector: RAXIS Oscillation angle: 0.5 deg Exposure time: 5.016667 mins X-ray wavelength: 1.5418 Angstroms Crystal to detector distance: 300.0 mm No. of images: 687 Total rotation range: 343.5 deg Dataset size: 11.5GB

Identifier <http://arrowprod.lib.monash.edu.au/hdl/1959.1/25920>

Identifier monash:5863

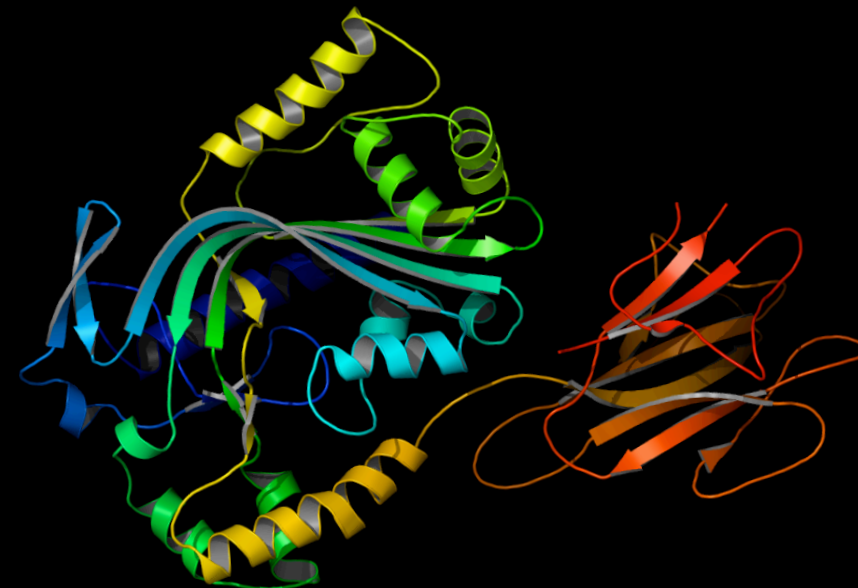
Identifier <http://arrow.monash.edu.au/hdl/1959.1/5863>



Relation <http://www.sciencemag.org/cgi/content/abstract/1144706>

Source dataset

Full Text




File	Description	Size	Format	
DS10	DATASET_4: Derivative (KPIC) in-house Part 1	1 kB	application/x-tar.bz	View/Open
DS11	DATASET_4: Derivative (KPIC) in-house Part 2	1 kB	application/x-tar.bz	View/Open
DS12	DATASET_5: Derivative (KI) in-house	1 kB	application/x-tar.bz	View/Open
DS13	DATASET_1: Unmerged reflection intensities	1 kB	text/sca	View/Open
DS14	DATASET_2: Unmerged reflection intensities	1 kB	text/mtz	View/Open
DS15	DATASET_3: Unmerged reflection intensities	1 kB	text/mtz	View/Open
DS16	DATASET_4: Unmerged reflection intensities	1 kB	text/mtz	View/Open
DS17	DATASET_5: Unmerged reflection intensities	1 kB	text/mtz	View/Open
DS18	Data extraction readme	17.2 kB	Adobe Acrobat PDF	View/Open
DS2	VITAL Metadata	166 B	XML Document	View/Open
DS5	DATASET_1: Native 2.0 Angstroms	1 kB	application/x-tar.bz	View/Open
DS6	DATASET_2: Native 2.5 Angstroms Part 1	1 kB	application/x-tar.bz	View/Open



2qp2  

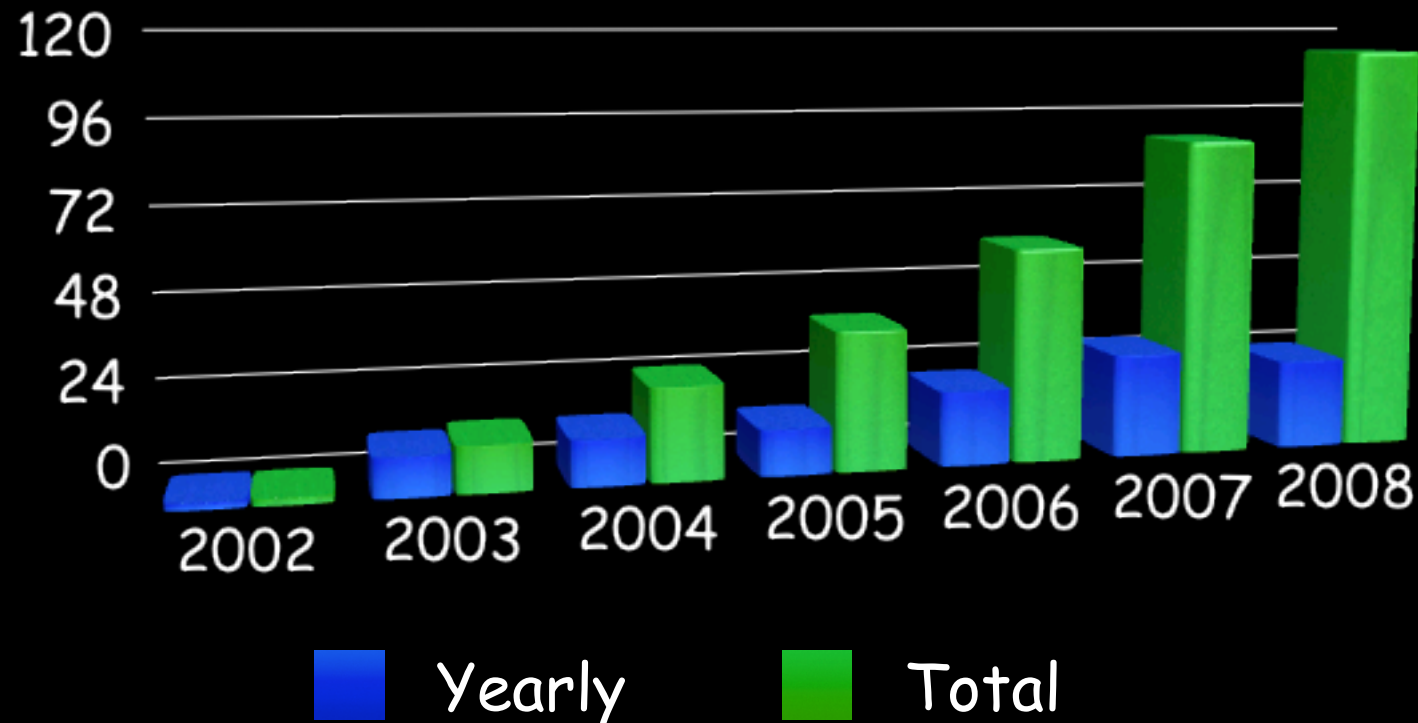
Red - Derived Information

DOI 10.2210/pdb2qp2/pdb

Title	Structure of a MACPF/perforin-like protein
Authors	Rosado, C.J., Buckle, A.M., Law, R.H.P., Butcher, R.E., Kan, W.T., Bird, C.H., Ung, K., Browne, K.A., Baran, K., Bashtannyk-Puhalovich, T.A., Faux, N.G., Wong, W., Porter, C.J., Pike, R.N., Ellisdon, A.M., Pearce, M.C., Bottomley, S.P., Emsley, J., Smith, A.I., Rossjohn, J., Hartland, E.L., Voskoboinik, I., Trapani, J.A., Bird, P.I., Dunstone, M.A., Whisstock, J.C.
Primary Citation	Rosado, C.J., Buckle, A.M., Law, R.H., Butcher, R.E., Kan, W.T., Bird, C.H., Ung, K., Browne, K.A., Baran, K., Bashtannyk-Puhalovich, T.A., Faux, N.G., Wong, W., Porter, C.J., Pike, R.N., Ellisdon, A.M., Pearce, M.C., Bottomley, S.P., Emsley, J., Smith, A.I., Rossjohn, J., Hartland, E.L., Voskoboinik, I., Trapani, J.A., Bird, P.I., Dunstone, M.A., Whisstock, J.C. A common fold mediates vertebrate defense and bacterial attack <i>Science</i> v317 pp. 1548-1551, 2007 [Abstract] 
History	Deposition 2007-07-22 Release 2007-09-04
Experimental Method	Type X-RAY DIFFRACTION Data 
Parameters	Resolution[A]  R-Value R-Free Space Group 2.00 0.182 (obs.) 0.204 P 2 ₁ 2 ₁ 2

How much data?

Number of protein structures



+ STATS

Size of data indexed by TARDIS

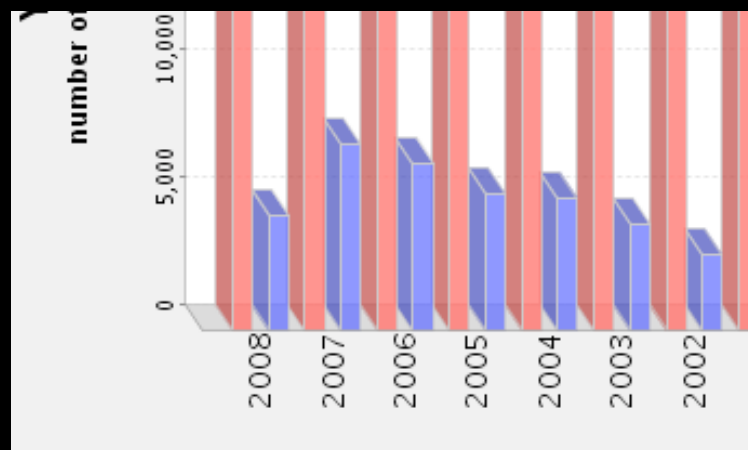
- Compressed: 20.60 GB
- Uncompressed: ~80 GB
- Average Dataset Size (Compressed): 1.48 GB

15 datasets

= 21 Gb

20 structures/year growth

= 30 Gb / year



if PDB growth = 10000 / year
= 15 Tb / year

3302 records in the Monash repository
(PhD thesis, images etc)

= 140 Gb

58% is crystallographic raw data
(uncompressed)

Benefits to making raw data available

- Data is stored safely !
- Experiment reproducibility/validation
- Methods development
- Re-refinement of published structures with new methods
- Some data cannot be processed!

Future

- Engagement
 - community
 - wwPDB
 - Journals
 - IUCR
- Data format/standardisation - eg
imgCIF/CBF

Acknowledgements



Buckle Lab

Steve Androulakis
Jason Schmidberger
Mark Bate

Whisstock Lab

James Whisstock
Ruby Law
Sheena McGowan
Corrine Porter
Khalid Mahmood

MeRC

David Abramson
Nick McPhee
Wojtek Goscinski
Paul Bonnington

ARROW

Neil Dickson
David Groenewegen

ARCHER

Anthony Beitz
Edwina Harvison
Bob Cameron
Robyn Polan
Nick Nicholas
Cyrus Keong
Andrew Treloar (ANDS)

Library

Suzanne Clarke
Andrew Harrison
Joan Gray

UQ

Jenny Martin
Bostjan Kobe
Jane Hunter

St Vincents

Michael Parker

Bio21/Melbourne Uni

Renwick Dobson
Lyle Winton
Eve Young

JCU

Ian Atkinson
Frank Eilert

U Sydney

Mitchell Guss
Peter Turner

Southern University

Simon Coles



arrow
enabling research
replication online
to the world

<http://www.tardis.edu.au>

ashley.buckle@med.monash.edu.au



MONASH University



Australian Government
Australian Research Council