The Australian Store.Synchrotron Data Management Service for Macromolecular Crystallography

Grischa R Meyer, Steve Androulakis, Philip Bertling, Ashley Buckle, Wojtek James Goscinski, David Groenewegen, Chris Hines, Anitha Kannan, Sheena McGowan, Stevan Quenette, Jason Rigby, Patrick Splawa-Neyman, James M Wettenhall (Monash University, Clayton, Australia), David Aragao, Tom Caradoc-Davies, Nathan Mudie (Synchrotron Light Source Australia, Clayton), Charles Bond (University of Western Australia, Crawley, Australia)

Store.Synchrotron is a service for management and publication of diffraction data from the macromolecular crystallography (MX) beamlines of the Australian Synchrotron. Since the start of the development, in 2013, the service has handled over 79 TB of raw data (~ 5.8 million files). Raw data is made available securely via the web and SFTP so experimenters can sync it to their labs for further analysis. With the goal of becoming a large public repository of raw diffraction data, a guided publishing workflow which optionally captures discipline specific information was built. The MX-specific workflow links

PDB coordinates from the PDB to raw data. An optionally embargoed DOI is created for convenient citation. This repository will be a valuable tool for crystallography software developers. To support complex projects, integration of other instruments such as microscopes is underway. We developed an application that captures data from instrument computers, enabling centralised data management without the need for custom ingestion workflows. The next step is to integrate the hosted data with automatic and interactive processing and analysis tools on virtual desktops.

Web access to live data SFTP access Data publishing Link to PDB DOI minting

Datasets

A	В
The Operation of Store.Synchrotron	
	Experiments
Collector	

A: Schematic overview of the Store.Synchrotron service

B: Storage schema and abstraction. Experiments are beamline alloca-



ine data access is facilitated by this data portal. The service is maintained by Monash University, but owned by the

sed under a Creative Commons BY license: Whenever a work is copied or redistributed under a Creative

s BY licence, the original creator (and any other nominated parties) must be credited and the source linked to.

1 of 4

What do you need to do?

G

	Disk			
Files	actua	data	Таре	
			Squash	S
)				
• • < > III Store, Synchrotron Home	About My Data Public Data	store.synchrotron.org.au/dataset/ Stats Help	Ċ	0 🗄
lataset				
rom the experiment: Experiment				Cit Dataset Des
elected diffraction image	es from your dataset		720 Files	Add
Internet in			Download Dataset: 🛃	TAR
ALC: NOT THE OWNER OF	100 H H		« 1 2 3	4 5 6 7 8
Sec. 1	500 B	1 mar 1		
A COL INC.	4.000	4.404	Search: Enter par	t or all of a filename, then
1_001.img	_1_241.mg	+o r.mg	Select: All / None	
atafila Information			_uur.img	(0.0 MID) 📥 +
Datafile Mota	lata			(8.0 MB) 📥 🕂
For the file: 1_001.img (3.0 MB)		± + □ 03.img	(8.0 MB) 📥 +
TRD datafile			✓ Edit 04.img	3 (8.0 MB) 🛓 +
Detector Distance	234.996 mm		0 05 im	(8 0 MB)
Detector S/N	457			
Direct Beam X-Pos	105.229 mm		06.img	(8.0 MB)
Exposure Time	1.0 s		07.img	(8.0 MB) 🛃 +
ocsillationRangeStart	0.000			
Preview Image			08.img	J (8.0 MB)
	O ≡ store	synchrotron.org.au/experiment/view/208/#me	tadata Č	0 1 0
tore.Synchrotron Home Ab	out My Data Public Data Stats H	lelp		J Log In
periment Definina the intera	action of perforin	with calcium and t	he phospholipid mer	nbrane
aouda A.K. Traore, James C V	/hisstock			
				- Pablic
Hide De	scription	1 Dataset		
Following its secretion from cytotoxic ly perforin binds to target cell membranes	mphocytes into the immune synapse, through its Ca2 + -dependent C2	▲ Download Selected		
uomain. membrane-bound perforin ther	t cell. In the present study, structural bindingtriggersaconformationslebager	Just start typing to filter data	sets based on descriptions	
and biochemical studiesrevealthatCan	vdrophobic residues to interact with	smax_r	ng_D4_3	
and biochemical studiesrevealthatCa2+ in the C2 domain that permits four key h the plasma membrane. However, in con	trast with previous suggestions, these			
pro-approprice granzymes into the target and biochemical studiesrevealthatCa2+ in the C2 domain that permits four key I the plasma membrane. However, in com movements and membrane binding do changes in the pore-forming MACPF (m domain, indicating that subsequent mor membrane surface are required for <i>perfe</i>	rast with previous suggestions, these not trigger irreversible conformational embrane attack complex/perforin-like) iomer-monomer interactions at the rin pore formation.	Synchro	tron MX Data	■ 220 3.9 GB
pro-apopuopic grazymes into the targe and biochemical studiesrevealthatCa2+ in the C2 domain that permits four key 1 the plasma membrane. However, in con movements and membrane binding do changes in the pore-forming MACPF (m domain, indicating that subsequent mor membrane surface are required for perfor Publication: Blochem J.	Trast with previous suggestions, these tot trigger inversible conformational embrane attack complex/perform. like) omer-monomer interactions at the prin pore formation.	Synchro	tron MX Data	■ 220 3.9 GB

Η

tions, datasets combine files for each crystal. The files can be stored in any number of locations and accessed transparently via Web or SFTP

- C: Experiment page right after collection. SquashFS archive added after Experiment completion, auto-processing coming soon.
- D: Dataset view, shows previews for image files, metadata for dataset (hidden here), a list of files and rich metadata for each file.
- E: Publication form start page. Three dialogs present streamlined query for information on data, PDB ID, authors, license, embargo.
- F: Public experiment view, shows title, detailed description, metadata and a list of datasets. Some metadata and related publications (hidden here) auto-harvested from PDB website.
- G: Latest statistics of Store.Synchrotron site.



Protein Data	a Bank
PDB ID	3W56
URL	http://pdb.org/pdb/search/structidSearch.do?structureId=3W56
Organism	Scophthalmus maximus
Expression System	Escherichia coli
Sequence	>3W56:A PDBID CHAIN SEQUENCE MRGSHHHHHHENLYFQGQNSHMQLRLYNLRVRGLPSDLMGITDGYVKV NDILRLEVHDEDTFFDDLLGVCQRQIKVGTHEHDCYLKEGGTLHYMYTLS
Resolution	1.6 Å
R-Value	0.214 (obs.)

H: MyData desktop application. Used for data upload from microscopes and many other instruments.

	0	store.synchrotron.org.au/stats/	Ċ	0 1 +
Store.Synchrotron Home	About My Data Public	Data Stats Help		•) Log In
Stats				
Experiments stored: 962				
Datasets stored: 83940				
Files stored: 5861588				
Data stored (at least): 79.2 TB				
Hosted at Monash University on the Ne	CTAR Cloud			Powered by MyTardis

lay 🛷 Private

1 50.7 GB

B 6



	MyData - Test Microsco	pe				×
🕒 💁 🍓 🕥 🌮 🚱				Q	Search	⊗
Folders Users Verifications Uploads Log						
ld Folder Subdirectory	Filename	File Size	Status	Progress	Message	^
¹ Amorphophallus titanum SEM	01_AM1-1_Sep2011-2 (1).jpg	6 MB	V		Upload complete!	
² Amorphophallus titanum SEM	03_AM2-1_Sep2011-1.jpg	6 MB	V		Upload complete!	_
³ Amorphophallus titanum SEM	05_AM2-1_Sep2011-3.jpg	4 MB	V		Upload complete!	
4 Amorphophallus titanum SEM	04_AM2-1_Sep2011-2.jpg	5 MB	65		22 % uploaded	_
⁵ Amorphophallus titanum SEM	06_AM2-1_Sep2011-4.jpg					
6 A 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	00 4140 4 6 2014 0.1			,,		×
Cancel Selected Upload(s) Cancel All Remaining Upload(s)						
Connected to http://127.0.0.1:8000						1

