Australian Structural Biology Computing

Community Meeting

Public document

Purpose of these meetings

To collectively answer / begin addressing these questions:

- Are there computational problems that the community of computational structural biologists need to address?
- Can we use national infrastructure to solve these problems? If so, how?
- It is difficult to set up and use AI/ML approaches. Are these useful to you?
- How do we help non-structural biologists access computational structural approaches?

Zoom link

Wednesday 12th of November 2025

11:00 am (AWST) Perth, 12:30 pm (ACDT) Adelaide, 13:00 pm (AEDT) Sydney, Melbourne, Brisbane, Hobart

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- Thomas Litfin (UNSW/BioCommons)
- Kate Michie (UNSW)
- Michael Healy
- Pavel Misiun
- Joel Haywood
- Kristina Gagalova
- Tiffanie Nelson
- YC Lin
- Farak Khan
- Josh Hardy
- Emma Rath
- Meihan Liu
- Zhengtan Zheng
- Fiona Whelan

Apologies

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	If you have feedback or items for discussion, please let Kate or Johan know. You can also add these directly to the agenda **Slack link** Mailing list: join here https://australian-structural-biology-computing.github.io/
Notifications	**Roadmap** https://doi.org/10.5281/zenodo.15786982 **Protein Design webinars** WEBINAR SERIES: Leveraging deep learning to design custom protein-binding proteins — Australian BioCommons **EMBL-EBI training collaboration has started** **Link to design competition: https://foundry.adaptyvbio.com/competition** **Link to low-memory Boltz: https://www.biorxiv.org/content/10.1101/2025.10.29.684571v2** **Link to ProteinDJ (RFdiffusion workflow):** **https://www.biorxiv.org/content/10.1101/2025.09.24.678028v1** **Protein design workshop type event combined with competition above?** Kate Michie Joel Haywood Josh Michael Healy Fiona Whelan Kristina Gagalova

Other community activities and active projects	**ABACBS 2025 workshop**
	
	Lorne Protein 2027 workshop
	● Useful to have a workshop?
	● Are there other ideas from the community?
	● Comments:
	  ○ Plan B: Satellite meeting in Melbourne - additive.
	  ○ Cost/Sponsorship?

Website	https://australian-structural-biology-computing.github.io ● Add in contribution mechanisms to the website, as they are available ● Working on workflows contribution mechanism ● i.e. making use of WorkflowHub

Presentation	**“The prokaryotic origins of the COMMD protein family involved in eukaryotic membrane trafficking”**
	Dr Michael Healy (Institute for Molecular Bioscience, UQ)
	
	Questions
	● Software suggestions
	  ○ https://github.com/gamcil/cblaster
	  ○ https://github.com/gamcil/clinker

	• How many sequences from Archaea? 110 sequences found
	  ○ Odd to go from 16 in protists to 1 in Archaea
	  ○ 300 Archaea genomes added recently, might help?
	• Proteins weren't in model systems - difficult to work with
	  ○ Means assessing function is also difficult
	• What are you using at scale? Bunya HPC and Alphafold2
	• Can you speculate as to whether both octameric/decameric assemblies are functionally relevant? Why are there two oligomeric states?
	  ○ Unknown: see function comment above

Comments & suggestions from the community & AOB •	

Action(s)
• Johan:
  ○ Meeting for protein design workshop type event combined with competition
  ○ Coordinate Lorne Protein 2027 workshop planning	

Next meeting is in 2026 ! **Community meeting / invited speaker series (5 x a year, every 2 months)** **Technical working group** **Academic panel meeting (2 x a year)**	

Wednesday 27th of August 2025

11:00 am (AWST) Perth, 12:30 pm (ACST) Adelaide, 13:00 pm (AEST) Sydney, Melbourne, Brisbane, Hobart

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- Kate Michie
- Kristina Gagalova (Curtin University)
- Farrah Blades
- Michael Healy
- James Lingford
- Matthew Downton
- Pavel Misiun
- Bostjan Kobe
- Emily Furlong
- Nathan Glades
- Thomas Litfin
- Mitchell O'Brien

- Emma Rath
- Josh Hardy
- Ruohua Gao
- Jeff Christiansen
- Joshua Caley
- Andrew Perry

Apologies

- Richard Birkinshaw
- Sarah Beecroft

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	If you have feedback or items for discussion, please let Kate or Johan know. You can also add these directly to the agenda **Slack link** Mailing list: join here https://australian-structural-biology-computing.github.io/
Notifications	**Roadmap** Michie, K. A., Litfin, T., Beecroft, S. J., Collins, B., Czabotar, P., Downton, M., Doyle, M. T., Ghosal, D., Grinter, R., Knott, G. J., Samaha, G., Christiansen, J. H., & Gustafsson, O. J. R. (2025). Australian Structural Biology Deep-Learning Infrastructure Roadmap. Zenodo. https://doi.org/10.5281/zenodo.15786982 **Protein Design seminars** WEBINAR SERIES: Leveraging deep learning to design custom protein-binding proteins WEBINAR SERIES: Leveraging deep learning to design custom protein-binding proteins — Australian BioCommons Please register if you would like to attend 🙂 **Recording for webinar #2 now available** Dr. Cyntia Taveneau (Monash University) AIcrs: AI-Designed Anti-CRISPRs as Programmable CRISPR Inhibitors **Contributors from the Australian community to EMBL-EBI training collaboration?** Developing a *Foundations of Protein Structure* module to begin with
Planning for meetings in 2026	**Community meeting / invited speaker series** (5 x a year, every 2 months) **Technical working group** **Academic panel meeting** (2 x a year)
Other community activities and active projects	**CASP competition community entry?** • Validate models deployed on Aus infrastructure. • Participation covered in tech meeting (to be set up) •

Website	https://australian-structural-biology-computing.github.io
	
	● Updates to website, according to feedback
	  ○ Consistent naming of the community i.e. ASBC
	  ○ Umbrella is ASBC, AI/DL is current content, to allow for future cryo-EM methods in another section
	  ○ Grouping resources into themes / workflows, rather than type of resource
	● Add in contribution mechanisms to the website, as they are available
	  ○ Working on workflows contribution mechanism
	● Merge website changes to incorporate new activities pages
	  ○ DONE! See https://australian-structural-biology-computing.github.io/activities

Technical update	**“Pawsey update”** **Dr Sarah Beecroft (Pawsey Supercomputing Research Centre, Perth)**

Presentation	**“Desktop Style Cryo-EM Data Processing on Bunya HPC”**
	Dr Farrah Blades (Institute for Molecular Bioscience, UQ)
	
	As structural biology embraces an era of big data, GPU-intensive workflows and heavy AI integration, traditional data processing methods are no longer keeping pace, at the University of Queensland (UQ) we have managed to solve this issue at scale for our researchers utilising Open OnDemand and clever integration to Bunya HPC.
	
	This new environment integrates key software such as CryoSPARC, Relion, ModelAngelo, ChimeraX, Coot, and more, within a single virtual desktop interface.
	
	Users can process large datasets, visualize structures, and leverage A100, H100 and L40 GPUs with minimal technical overhead, they no longer need to learn command line, how to write scripts or how to submit jobs manually to a HPC. Allowing researchers to focus on discovery, not infrastructure.
	
	In this presentation, I share how our web-based platform works with some use case examples from myself and others in the Cryo-EM community at UQ.
	
	Questions
	
	● Storage used? Can it be scaled up, or proposed to other systems?
	  ○ Research Data Management (RDM) system, with multiple levels
	● Optimisation for each experiment sent to Bunya? Are some more difficult and require optimisation?
	  ○ When particles from cryo-EM are really big
	  ○ Limitations of what GPU can process
	  ○ Ends up being compressed
	  ○ Can not be uncompressed - data lost
	● Using multiple GPUs (parallelisation) can be difficult to set up on HPC
	  ○ Running on one big GPU, to circumvent parallelisation
	  ○ Can fragment a single GPU
	● Cryosparc - initially a central deployment, now using a user-specific instance model
	  ○ How did you manage the program?
	  ○ Each user has their own instance
	  ○ Need to install cryosparc in their instance

Comments & suggestions from the community & AOB ●	

Action(s)
•
Next meeting is on 12th November 2025

Wednesday 18th of June 2025

11:00 am (AWST) Perth, 12:30 pm (ACST) Adelaide, 13:00 pm (AEST) Sydney, Melbourne, Brisbane, Hobart

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- Kate Michie (UNSW)
- Tom Litfin (Biocommons/UNSW)
- Joshua Caley (UNSW)
- Keiran Rowell (UNSW)
- Kapil Mcinerney (PMCC)
- Emily Furlong (ANU)
- Josh Hardy (WEHI)
- Jeff Christianen (Australian BioCommons)
- Andrew Perry (Monash/APDI/MGBP)
- Carlos Riveros (UoNewcastle)
- Tracy Nero (Bio21/UniMelb)
- Farah Zaib Khan (Australian BioCommons)
- Farrah Blades (UQ)
- Matthew Downton (NCI)

Apologies

- Emma Rath
- Nadia Zatsepin (SUT)

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	If you have feedback or items for discussion, please let Kate or Johan know. You can also add these directly to the agenda **Slack link** Mailing list: join here https://australian-structural-biology-computing.github.io/website/
Roadmap	• **Computational Structural Biology Roadmap v5**

	• Reviewed by international reviewers • Publishing online on June 25th • Please forward • If you would like to indicate support for the roadmap, please email Johan before the 25th (include names and affiliations) • Names will be added to page 2 of the published roadmap
Roadmap response	• **Short presentation** (JG)
Community Projects	• **Updates** (TL): PyRosetta and protein design on NCI
Protein Design seminars	• **WEBINAR SERIES:** Leveraging deep learning to design custom protein-binding proteins https://www.biocommons.org.au/events/protein-binder-series • Please register if you would like to attend 🙂

Community website	• **Walkthrough of v2.0** https://supernord.github.io/website/index
	• Send feedback to Johan or Kate
	• Once feedback in incorporated, changes will be merged with main community site
	• Feedback
	  ○ Consistent naming of the community
	  ○ ASBC
	  ○ Umbrella is ASBC, AI/DL is current content, to allow for future cryo-EM methods in another section
	  ○ Grouping resources into **themes** / workflows, rather than type of resource

Other community activities and active projects	**International connections (KM, JG)** • Creating How-to Guides together with ELIXIR community and PDBe / EBI • Please let Kate or Johan know if you would like to contribute • Proposed current guide: protein binder design **Business Council of Australia AI Agenda released on June 2nd** • Noted that increased computing resources specifically for structural biology (Alphafold) are important for Australian Business development https://aiagenda.bca.com.au/
Comments & suggestions from the community •	

AOB
• **AlphaFold Education Summit** https://www.ebi.ac.uk/training/materials/alphafold-education-summit-materials/
• JC - From Cath Brooksbank, Head of EMBL-EBI Training: We've been doing some very interesting work with the Alphafold DB team and Google Deepmind; we ran an Alphafold Education summit back in January and it's nucleated a highly motivated community of scientists to deliver Alphafold training in their local context. I'm wondering whether there would be interest in related activities with the Australian bioinformatics community? If so, happy to arrange a call.
 	
• Google contact "Would this be something Biocommons would want to collaborate with? Perhaps being the local experts (co-presenting) and organisers (providing rooms).
  ○ Any thoughts on what would be most helpful for "the community"?
  ■ Like "train the trainer" sessions?
  ■ Technical sessions like how to install AF on your facilities/cloud?
  ■ Intro to AlphaFold 101?
  ■ Advanced AlphaFold workflows?
  ■ How to deal with the AlphaFold3 License?
  ■ All of the above etc etc?"	

Action(s)

- Updates to website, according to feedback
- Add in contribution mechanisms to the website, as they are available
- Merge website changes to incorporate new activities pages
- Continue pursuing collaborations with EBI and Google

Next meeting is on 27th August 2025

Wednesday 9th of April 2025

11:00 am (AWST) Perth, 12:30 pm (ACST) Adelaide, 13:00 pm (AEST) Sydney, Melbourne, Brisbane, Hobart

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- Kate Michie (UNSW)
- Tom Litfin (Biocommons/UNSW)
- Craig Morton (CSIRO)
- Magdalena Antczak (QCIF/QUT)
- Revathy Raveendranath (UQ)
- Daniel Luque (UNSW)
- Andrew Perry (Monash/APDI/MGBP)
- Fiona Whelan (UoA)
- Rhys Grinter (UoM)
- Joshua Caley (UNSW)
- Emily Furlong (ANU
- Farah Zaib Khan (Australian BioCommons)
- George Bouras (UoA)
- Susie Grigson (Flinders)
- Nadia Zatsepin (SUT)
- Matthew Downton (NCI)
- Josh Hardy (WEHI)
- Dylan Silke (WEHI)
- Thomas Ve (Griffith)
- Keiran Rowell (UNSW)
- Emily Furlong
- Emma Rath (VCCRI)
- Kristina Gagalova
- Sarah Beecroft
- Benjamin Goudey (BioCommons)
- Gavin Knott
- Hannelore Longin
- Paul Rohde (VCCRI)
- Farrah Blades
- Brett Collins

- Michelle Christie
- Sarah Vreugde
- Joe Kaczmarski
- Thomas Ve
- Brady Johnston
- Spencer Passmore
- Joel Mackay
- James Lingford

Apologies

- David Jacques (UNSW)
-

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	If you have feedback or items for discussion, please let Kate or Johan know. You can also add these directly to the agenda **Slack link** Mailing list: join here https://australian-structural-biology-computing.github.io/website/
Previous actions	ALL: Shared document for feedback on computational structural biology website …
Presentation	**"Ultra-fast and highly sensitive protein structure alignment with segment-level representations and block-sparse optimization"** Dr. Thomas Litfin, UNSW **Questions** ● Ben Goudey: Do you find examples where all 3 methods are arriving at different answers? Does this reveal what is happening under-the-hood? ○ Cases where subtle reordering of visually similar structures will impact final metrics ● Thomas can you post the DOI and github links to the chat or here in the Agenda please! ○ https://doi.org/10.1101/2025.03.14.643159 ○ github.com/tlitfin/spfast ● Using segment level comparison of two proteins - quick check on overall segment, that we can compute efficiently ○ If nothing appears there, it allows a decision to not work on that system ● Pymol plugin - how does it perform relative to pymol alignment tools that are already there? ○ Didn't show alignment accuracy, but more detail is in the paper ○ Better matching between residues than default, and is faster ○ Only supports pair-wise alignment at the moment ○ Future: make it more of an integrated search, as this is where the benefit is for high throughput search
Presentation	

	"Filling the gaps: Massive-scale protein structure prediction beyond the AlphaFold database (with the help of Pawsey)" George Bouras, Adelaide University, Basil Hetzel Institute ● **Pharokka** https://github.com/gbouras13/pharokka ● **phold** - Phage Annotation using Protein Structures https://github.com/gbouras13/phold ● *AllTheBacteria - all bacterial genomes assembled, available and searchable* https://doi.org/10.1101/2024.03.08.584059 **Questions** ● What are the protein size ranges involved? ○ Generally smaller ○ Median ~ 200 aa ○ Can be over 3000 aa ○ Max sequence length for Pawsey - 3000 aa at the moment ○ For large sequences, these were split ● Archaea? Not at this stage ● Calculating MSAs in batches of 250k within 24 hours? ○ Used high mem queue (~ 1 TB of memory) ○ Colabfold DB ○ Expanded environmental DB ○ Shorter size of phage proteins helps, but generally this runs in less than 24 ○ Generates lots of files ● Running other tools like boltz? ○ Already computed phage structures in Dec, just when boltz was released ● Structure based annotation - function can be more variable than sequence ○ Is there a way to narrow down structural match vs functional relationship ○ Phage DB is well annotated - you can trust the annotations and therefore the hits ● Differences between *AllTheBacteria* and previous efforts (e.g. RefSeq)? ○ Focus of work presented is to improve annotations ○ RefSeq relies on assemblies submitted, and so quality is variable ■ Re-assembly could improve this
Roadmap	Computational Structural Biology Roadmap v4 Now open for comment If you wish to indicate your support for the roadmap document, as a community member, please contact either Kate or Johan via email and include your affiliation. Names and affiliations will be included in an additional supporting file as part of the final Zenodo entry for the roadmap.
Other community activities and active projects	**Grant update (KM)** ● Pause during caretaker mode ● Other efforts in progress (research infrastructure focus) **International connections (KM, JG)** ● ELIXIR 3D-BioInfo https://elixir-europe.org/communities/3d-bioinfo, INSTRUCT https://instruct-eric.org/, and PDBe https://www.ebi.ac.uk/pdbe/

	• ELIXIR BioHackathon 2025 project proposal: creating How-to Guides together with ELIXIR community and PDBe / EBI **Community Projects** • Protein Design seminars
Comments & suggestions from the community •	
AOB? •	
Action(s) •	
Next meeting is on 18th June 2025	

2024-10-22

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- James Lingford
- Kate Michie
- Melissa Burke
- Emily Furlong
- Matthew Downton
- Matthew Thomas Doyle
- Nadia Zatsepin
- Richard Birkinshaw
- Gavin Knott
- Lisa Phippard
- Giorgia Mori
- Rhys Grinter
- Sarah Beecroft
- Fiona Whelan
- Joel Mackay
- Bostjan Kobe
- Mark Larance
- Steven Manos
- Daniel Luque
- Seo-Kyung Chung
- David Teran

Apologies

- Craig Morton

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	If you have feedback or items for discussion, please let Kate or Johan know, or you can also add these directly to the agenda Zoom links! 🙁 **Slack link** Mailing list: join here https://australian-structural-biology-computing.github.io/website/
Update presentation Matthew Downton, NCI	**RFdiffusion @ NCI** • Testing RFDiffusion on ARE • Channel #rfdiffusion-nci on Computational Structural Biology slack • Iterate on container based on feedback • New templates can be added easily • NCI interested in learning more about user requirements **Questions** • Does the implementation include the AF and proteinMPNN modules? No, but should be added • Need these other software packages, so outputs can be seen by users (i.e. structures) • Jupyter notebook ok? ○ Should work, but not being used currently • MPNN AF2 implementation we use with RFdiffusion https://github.com/nrbennet/dl_binder_design • Building in alternate ways of validating outputs

Active projects	**RFdiffusion**
	• RFdiffusion @ NCI
	• Wrapping RFdiffusion for Galaxy
	  ○ Technical challenges to wrapping this tool
	  ○ Jupyter notebook on NCI would be useful - there *is* a technical barrier, but this may be overcome
	
	Scaling Alphafold2 at national facilities
	• Update
	  ○ AF2 on AMD GPUs - XLA defect - on NVIDIA you can oversubscribe memory
	  ○ A bit over 3000 aa for one GPU currently
	  ○ Scaling beyond that requires fixing the defect
	  ○ Brought in vendor HPE to assist
	  ○ Is 3000 aa or less useful? Or wait for full function?
	  ■ Yes!
	  ■ 3000 is reasonable
	  ■ 9000 would be amazing
	• Link to any of these pairwise AF2 multimer papers?
	  ○ https://www.biorxiv.org/content/10.1101/2024.10.01.615885v1

	○ Website set up ○ 18 k complexes with relative confidence ○ CPU or GPU? ■ GPU ■ RosettaFold ● Should RosettaFold be on Pawsey? ○ No, results may be non-physiological ● Trialling matrix AF2 ○ Potentially rolling this out soon ○ Then available on Australian AF service ○ Colabs may not work as well as full versions - Galaxy working better ■ May not be the case for well known / characterised proteins ○ Galaxy feedback ■ The output files and the packaging of those output from Galaxy AF2 are a bit difficult to handle and keep track of. Likewise, some of the file extensions tend to get broken which means we have to correct them manually. Would be nicer if a single run just gets packaged all up in one file. Error outputs should also be standard on ranther than having to click it on. ● Note: https://github.com/chaidiscovery/chai-lab?tab=readme-ov-file ○ We've had a lot of success with this and it's an open book compared to AF3 ● Onboarding to tier 1 and 2 compute infrastructures ○ Most people in structural biology wouldn't know what notebooks are, or what tier 1 is, for example ○ Agreement on this point from across the group **Structural Biology platform project** ● **Increasing access to computational structural biology:** user friendly landing platform that enables non-coders to run experiments without coding skills or writing proposals at HPC facilities. *Aspiration: the platform can send off compute to suitable resources wherever they may be* ● Update ● Feedback will be critical for this project **Website & How-to Guides** ● **Welcome to the Community for Structural Biology Computing in Australia** ● Feedback? ○ **Action Johan: Shared document for feedback** **Shared document for feedback on computational structural biolo** ● Community suggestions for next 'how-to' ● Overall lay-out of the web page - do we want to broaden it to cryo EM and X-ray? If so, what should it look like and who else can help? ○ Yes, combine resources ○ Possibly difficult to manage ○ Help with cryo-EM (DL) ○ Good to have experimental techniques included as well - noting some resources exist in these communities, which could be linked to e.g. myscope for cryoEM.
Other community activities	**Roadmap draft being written (JG)** ● First draft almost completed ● **Ready to share with co-authors** **MRFF grant update (KM)** ● NDRI is current focus / priority

	International connections ● ELIXIR 3D-BioInfo https://elixir-europe.org/communities/3d-bioinfo & INSTRUCT https://instruct-eric.org/ ○ Organising meeting with leads for Q1 2025 ● EMBL-EBI - Ewan Birney (Director) - https://www.ebi.ac.uk/people/person/ewan-birney/ ○ Interested in all the Australian use cases ○ Invitation to continue building relationship with EBI ○ Intention to maintain collaboration ● NIH ● NIH Beowulf have incredible tutorials on their software ● Supercomputing facilities

Comments & suggestions from the community

-

AOB?

- Other codes to focus on? Bindcraft workgroup?
  - https://www.biorxiv.org/content/10.1101/2024.09.30.615802v1
  - https://github.com/martinpacesa/BindCraft
  - Predicted binders seem more plausible biologically
  - -current users? James Lingford, Brett Collins, Rhys Grinter, UNSW-SBF

Action(s)

- Actions Johan:
  - Shared document for feedback
    Shared document for feedback on computational structural biology website and How-to …
  - Share roadmap with co-authors

Next meeting is in 2025!

Suggestion for next meeting to be a RFdiffusion forum and discussion: *"Let's talk about RFdiffusion"*
Would anyone from the community like to present or lead a discussion?

2024-08-06

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- Kate Michie
- James Lingford
- Matt Doyle 🙂
- Craig Morton - Happy to Roadmap review
- Steven Manos
- Michael Healy
- Emily Furlong
- Georgie Samaha
- Wenjing Xue

- Richard Birkinshaw
- Begoña Heras
- Rhys Grinter
- Samitha Amarapathy
- Gavin Knott
- Mark Larance
- Thomas Huber
- Fiona Whelan
- Joel Mackay
- Gordon McDonald
- Ziad Al-Bkhetan (Australian BioCommons)

Apologies

- Brett Collins
- Daniel Luque

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	If you have feedback or items for discussion, please let Kate or Johan know, or you can also add these directly to the agenda **Welcome to our new members** - including the Australian Synchrotron **Events** ● Biomolecular Horizons 2024 https://www.bmh2024.com/ ● SCANZ conference: https://www.crystal35.org/ ● 10th International Congress on Electron Tomography. https://www.electron-tomography.org ● WORKSHOP: Online data analysis for biologists ● WORKSHOP: Hello Nextflow! **Slack link** **SLIDES:** 2024-08-06 Computational structural biology

Roadmap draft being written	**Feedback from group on describing the problem space**
	● Data management - retention / maintenance
	  ○ Equal challenge to GPU availability
	● New users / new facility managers - wrangling data
	  ○ Timelines for processing large data
	  ○ Getting to grips with refining
	  ○ These timelines are short
	  ○ How do you make the most of what you have?
	  ○ Timescales scale to expertise level
	● Ability to run software solutions on local infrastructure
	  ○ And interface between local and national solutions
	  ○ Hybrid solutions (local, national, cloud)
	● Privacy - method of data release impacts IP
	● Data quality, reproducibility, validation
	● Software dev is very fast - education or notification about changes - keeping updates relevant and accessible

Current community activities	**Community set up / growing the team** ● (SM) UNSW-BioCommons co-funded role 'structural biology informatician' ○ Hiring soon ○ Combination of coding and structural biology domain knowledge ● 2x new coders at UNSW ○ Working on software incorporated into Nextflow workflows **Scaling Alphafold2 at national facilities** ● Pawsey AMD porting of AF2 (Georgie Samaha, SIH) ● Part of BioCommons Platforms Project: BioCLI ○ Alphafold2 and Colabfold ○ Increasing access by porting to other GPU types ○ Scale limitation for Alphafold2 on AMD - 3000 amino acids (upper limit) ○ Working with AMD to address the problem right now ● RFdiffusion - testing on Gadi (NCI) starting soon **Structural Biology platform project** ● **Increasing access to computational structural biology:** user friendly landing platform that enables non-coders to run experiments without coding skills or writing proposals at HPC facilities. *Aspiration: the platform can send off compute to suitable resources wherever they may be* ● Nextflow workflow being developed (Ziad, BioCommons) ○ A separate version of Proteinfold workflow ○ Successfully run on NCI - Alphafold and ESMfold + visualisation component ○ User interface as well to show predicted structures ○ Working to add Foldseek ○ Add relevant code to workflows, point workflows at diverse infrastructures using Seqera platform, and add a UI to make it easier to use the workflows ● Project comes in here: to resolve remaining questions, including what does the platform look like, and what does it do? ● Need training as well, on how to use the platform, and interpret its outputs ● Matt suggestions for implementation: ○ large scale AF2 (eg. predict my two-protein complex across every bacterial species in which it is present) ○ FoldSeek to assist above? ○ Large scale "AlphaPulldown" (predict all protein-protein interactions for a target protein within a species encoded proteins) **Wrapping RFdiffusion for Galaxy** ● To continue as the team grows **How-to Guides** ● 'Help' pages/articles. Steps forward already with James Lingford - What should we prepare? ● i.e. ○ An organised activity where a set of structural biology guides are created, followed by a publication describing the work, ○ If relevant, this could include work with international projects as well (ELIXIR community), ● GitHub: https://github.com/Australian-Structural-Biology-Computing ●
International connections	**EU** ● ELIXIR 3D-BioInfo https://elixir-europe.org/communities/3d-bioinfo ● INSTRUCT https://instruct-eric.org/

	• EMBL-EBI - Ewan Birney (Director) - https://www.ebi.ac.uk/people/person/ewan-birney/ **Who else should we connect to?** • US? Oak Ridge NL • Finland? •
MRFF grant (KM)	**Grant** • Platform that is easy to use • Includes authentication, infrastructure etc. • Worked medical cases that use computational methods that have accelerated research
AOB? • Question: Do people here rate Rosetta fold? ○	
Action(s) • **Michael Healy and James Iingford to start on the first 'How tos'.** • **Kate to finish and distribute the roadmap with todays input included.** • **Reach out to SB Grid and Instruct.**	
Next meeting is October 29th 2024	

2024-05-14

Chair: Kate Michie
Minutes: Collectively recorded

Attending

- Johan Gustafsson (Australian BioCommons)
- Kate Michie
- Bostjan Kobe
- Rhys Grinter
- Georgina Samaha
- Debnath Ghosal
- Emily Furlong
- Brett Collins
- Gordon McDonald
- Gavin Knott
- Fasseli Coulibaly
- Matthew Downton
- Daniel Luque
- Matthew Thomas Doyle
- Charlie Bond

- David Teran

Apologies

- Thomas Huber

Agenda + Minutes

Item	Item / Description
Welcome to meeting & intros Kate Michie Johan Gustafsson	• If you have feedback or items for discussion, please let Kate or Johan know, or add these directly to the agenda • Welcome to our new members - including the Australian Synchrotron.

Current activities	**Scaling AF2 at national facilities**
	Efforts at Pawsey, NCI, UNSW, Galaxy
	• Scaling batch runs of AF2
	• AF3 out - not scalable for batches currently
	• AMD porting of AF2 @ Pawsey - successes in this space
	• Testing at NCI on Sapphires - scaling potential TBD
	• Group interested in batch runs? (Please add your name ☺)
	  ○ Gavin Knott
	  ○ Georgie Samaha/Gordon McDonald
	  Rhys Grinter
	  Debnath Ghosal
	  Fasseli Coulibaly
	  Matt Doyle
	• RFdiffusion running at UNSW
	• Matt: We are using AF2 on Galaxy all the time on very large complexes. Colab don't cut it. RFDiffusion access on galaxy could be incredibly useful for Australia's building synthetic biology community in addition to inhibitor discovery
	• Tom: is there interest in abstraction layers such as kokkos/pykokkos to make code run on various gpu platforms?
	
	Wrapping RFdiffusion for Galaxy
	UNSW team is working with Galaxy Australia to wrap RF diffusion
	• Useful?
	  ○ Rosettafold All atom being wrapped - as it includes everything needed
	  ○ Matt Doyle: I'm highly support this. Happy to provide test cases and positive controls.
	
	Community pages
	GitHub: https://github.com/Australian-Structural-Biology-Computing
	Landing page including 'Help' pages/articles
	• + email list
	• Input from community on what this page looks like and contains (Please add your comments / questions here ☺)
	  ○ Use same repository in the GitHub organisation for a GitHub pages site
	  ○
	
	Events
	• Biomolecular Horizons 2024 https://www.bmh2024.com/
	  ○ Please contribute resources to the talk at this conference
	• SCANZ conference: https://www.crystal35.org/
	• 10th International Congress on Electron Tomography. https://www.electron-tomography.org

	Roadmap draft being written Critical feedback from the community? Contributors: ● Email Kate / Johan: 3 most important things to achieve, most important things for year 1, 5 and 10 ● Rhys ● Gavin ● Tom ● Georgie ● Matthew Thomas Doyle ● Gordon ● Fasseli ● Debnath **Collaboration agreements?** Setting expectations, removing roadblocks In parallel to roadmap dev **To discuss** What should be included? How should this be curated? Would anyone like to contribute? How do we want to link to data from other infrastructures?
Democratising access to computational structural biology	User friendly landing platform that enables non-coders to run experiments without coding skills or writing proposals at HPC facilities. *Aspiration: the platform can send off compute to suitable resources wherever they may be* **To discuss:** could Galaxy be used for this purpose? Have a look at neurosnap. Would this be a useful approach?(https://neurosnap.ai/) ● Gen-ai assistance like jupyter-ai, open interpreter, gorilla-cli can be a quick stop-gap (but not the ideal solution - friendly user interfaces are better but take more effort to build) ● Time / support for install and use can be an issue ● Some members of the community are able to access local support ● Good use of time to make methods accessible? ○ Useful for anyone who works with structural biology ○ Matt: Absolutely a great use of your time Kate to help democratise RF Diffusion AA on Galaxy. ● Georgie: BioCommons project for command line, scaling and access ○ AMD porting of AF2 happening through BioCommons BioCLI ○ Georgie will update the group on progress ● Kate: There will always be cmd line. Trying to address the needs of non-experts trying to access computational structural methods ● Gordon: Also such a plethora of tools becoming available - perhaps a guide or chatbot explaining which tool will get you to where you want to go? We use cogniti.ai to build chatbots quickly from text docs (USyd built tool accessible through AAF). Chatbot has the advantage that it can answer your question directly so you don't have to read about less immediately relevant things. Guide has the advantage that common use cases are covered off. Each gets better from the other. ● Tom: what does success look like? ○ easy for applied researcher to come in and do most common things by themselves? ● Matt: Use of a third party company to funnel through jobs to galaxy. Potential problems regarding IP disclosure. IP is also something that needs to be considered during the democratisation. If IP protection is not possible when

	using these easily accessible systems then a disclaimer simply need to be added. ● Gordon: In terms of web-apps that are HPC compatible we do a lot of nextflow and love what NCI has done with ARE/Open on demand ●
Feedback from the group	● Does anyone use SB Grid? ○ Rhys - useful from sysadmin perspective ○ Debnath - very useful ○ Fasseli - Similar here. We use SBgrid for standard tasks. Works well. ○ Anyone open to partnering on this work? ○ European equivalent we might also link to? ● Is anyone using nucleotide transformer?/DNABert/… ○ Models not packaged in a tool yet ○ To pay attention to ● Request feedback for other initiatives/code ○ ● Request if there are other problems we can address? ○ Debnath - online courses and resources linked to from the landing page? Agreement from group ○ ● Suggest presentations for future meetings? i.e. 10 minute case studies ○ If someone solves a particular challenge, the group is aware, and people can approach them with questions ● Slack channel?
Questions and open forum discussion time	● Training / fine tuning foundational models ○ Is there a critical data set (that we have and others don't) that could be used for this? In this community? ○ PDB - AF2 addressed this ○ Maybe for a subtask - not in general ○ Gavin - building databases of microbial resistance, including collabs to build LLMs ○ Compute for training not easily accessible to individual researchers but we can find it between us. ○ But should be guided by a key niche that isn't yet filled ○ NCI has 16 A100s
Please add your item(s) / topic(s) here!	● What resources can we provide to help non structural biologists? ● Has anyone heard about CCP4 having potential problems with funding? ○ Letter of support from this group for this resource? ○ Kate will email a letter to the group and ask for approval
AOB? ● Data management and policies for GitHub ● Standard software licence for contributions? ● Contributor agreement you need to sign before pull requests are merged? ●	
Action(s) ●	
Next meeting is August 6th 2024	

2024-02-20

Perth, Australia AWST (UTC +8)	Tue, 20 Feb 2024	10:30 am
Brisbane, Australia AEST (UTC +10) 2 hour(s) ahead	Tue, 20 Feb 2024	12:30 pm
Adelaide, Australia * ACDT (UTC +10:30) 2:30 hour(s) ahead	Tue, 20 Feb 2024	1:00 pm
Melbourne, Australia * AEDT (UTC +11) 3 hour(s) ahead	Tue, 20 Feb 2024	1:30 pm

Chair:

Minutes: Collectively recorded

Attending

Name	Affiliation	What's your focus?
Charlie Bond	UWA	Protein-RNA interactions. Whacky structural biology.
Brett Collins	IMB, University of Queensland	Structural biology of cell trafficking. Using X-ray crystallography, cryoEM, AlphaFold, RFDiffusion and interested in protein-protein complexes.
Peter Czabotar	WEHI	Protein Structure, Drug discovery
Fasséli Coulibaly	Monash University	Protein structure and engineering. Structural virology. AlphaFold, cryoEM, X-ray crystallography.
Daniel Luque	UNSW	cryoEM/cryoET. Structural virology.
Rhys Grinter	University of Melbourne	Structural biology (cryoEM, crystallography, AI guided structure prediction and design). Bacteria, Protein Complexes, Membrane proteins
Gavin Knott	Monash University	Structural Biology (AI Prediction, AI Design, CryoEM, Crystallography)
Matt Doyle	USYD	Bacterial cell surfaces, protein folding machines, protein-protein interactions, antibiotics

Emily Furlong	ANU	Structural biology, including cryo-EM, X-ray crystallography, AlphaFold
Matthew Downton	NCI	Software platforms, workflows
Richard Birkinshaw	WEHI	Structural Biology, protein design, RF diffusion, Alphafold, X-ray crystallography, CryoEM
Thomas Ve	Griffith University	Structural Biology, X-ray crystallography, CryoEM, Alphafold, RFdiffusion
Johan Gustafsson	Australian BioCommons	
Thomas Huber	ANU	Computational structural biology, protein design
Nadia Zatsepin	Swinburne	XFEL and synchrotron serial crystallography
Kate Michie	UNSW	Structural Biology, protein design, RF diffusion, Alphafold, X-ray crystallography, CryoEM
David Teran	University of Melbourne / Bio21	Structural Biology (CryoEM, crystallography, Alphafold, docking)

Agenda + Minutes

Item	Item / Description	Action(s)
Welcome to meeting & intros Kate Michie Johan Gustafsson	• Context • If you have feedback, please let Kate or Johan know • If you have items for discussion, let Kate or Johan know, or add these directly to the agenda • Introductory slides	
Question and discussion time	Mention the aims that are included below Ask for input from the group • MASSIVE mentioned - different capabilities involved / collaborating depending on needs / requirements • Galaxy Australia Alphafold performing	

	better than NIH solution at the moment ● CryoEM software availability ● Landing page for community needed? ● National voice for computational structural biology ○ Collaborating with Academy committees, or other societies? ○ Complicated space ● Microscopy Australia in cryoEM space - looking for feedback at the moment ○ Sharing knowledge with this capability ● Data movement and automating workflows between infrastructures? Is this of interest? ○ Problem for cryoEM - data heavy ○ Alphafold input and output not as large ● For DL, is it inference, or also training? ○ Mostly inference at the moment ○ Inference first, then training! ● Including people from Synchrotron ●	
Aims **(open discussion)**	Who are we? Are there others that should be here? What are the challenges? What is feasible to address in 2024? What shouldn't be included? What is in scope? ● Protein structure ● Nucleic acids ● Biological macromolecules ● Including representatives from molecular dynamics (MD) ● Avoiding confusion between broad access to software, vs bleeding edge exploration Compute infrastructure access is an issue: ABLeS provided as an example of a BioCommons service Galaxy Alphafold / RFdiffusion as an option	

	Being proactive and pursuing model training that is tailored to the requirements of this community Benchmarking and sharing knowledge of interest?	
What are the next steps?	What do we focus on? ● Troubleshooting national access to Alphafold capability ● Working on documentation / web page / central resource ● Batch runs ● Alphafold against whole genomes ○ Running locally and on MASSIVE ○ Compiling resources for how to do this in practice ● RFdiffusion on Galaxy ○ Scale of available compute needs to be appropriate ● Sharing existing knowledge of what works and what doesn't ● + existing resources available Is there software / code to triage and scale? ●	
Add your item(s) here!		
AOB? ●		
Next meeting is May 14th 2024		