UNSW SYDNEY

Serving next generation Structural Biology with GPUs

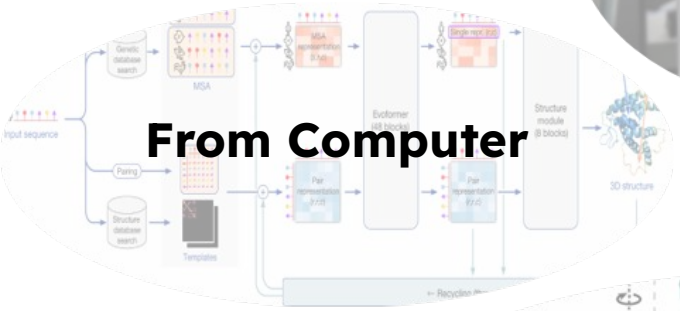
Structural Biology Facility

SBF Team (L→R): Nathan Glades, Keiran Rowell, Kate Michie, Josh Caley

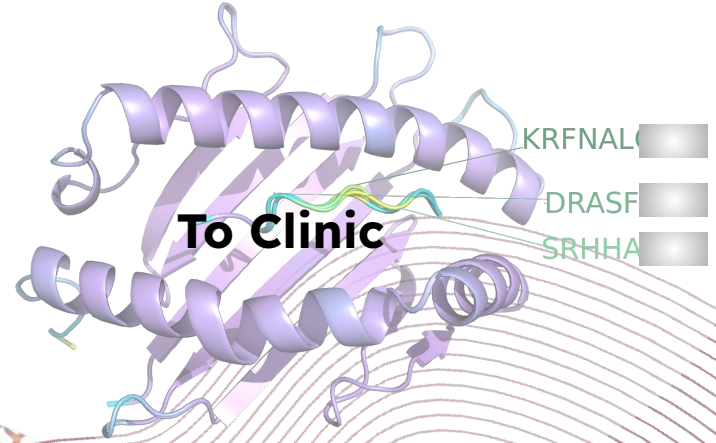
CRICOS Provider Code 00098G

The revolution in Computational Structural Biology

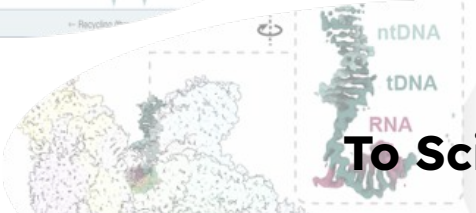
From Instrument

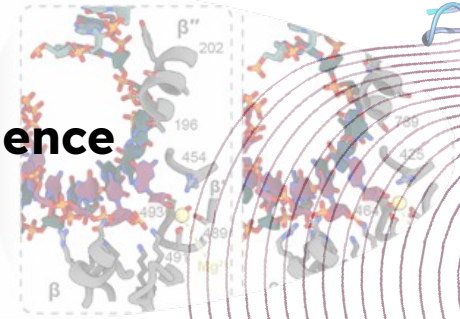


From Computer



To Clinic





To Science

CRICOS Provider Code 00098G

Deep Learning Biomolecular structures – Uses

Archaeal evolution – Michie/Burns Labs - ESMFold + FoldSeek

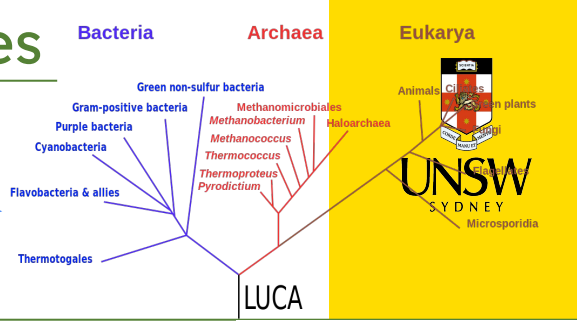
UNSW Ramaciotti Centre for Genomics

Sequence whole genome

4622 proteins

Structural

AFDB-search

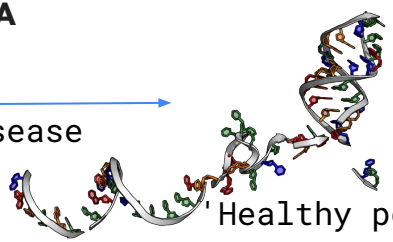


Medical genomics – Oates Lab – RFoldNA

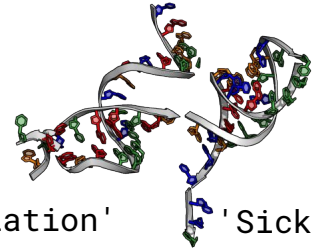
gnomAD Genome Aggregation Database

172 variants

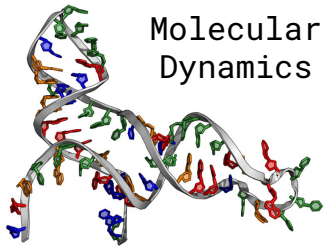
Neurological disease (genetic)



'Healthy population'



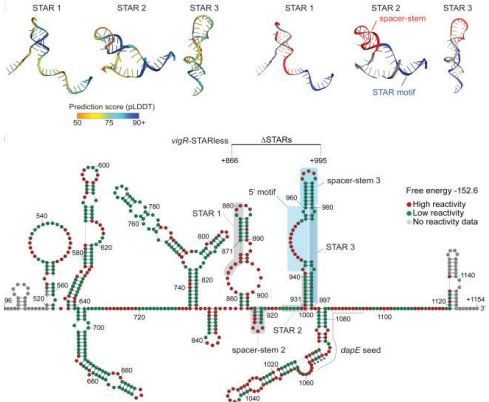
'Sick kid'



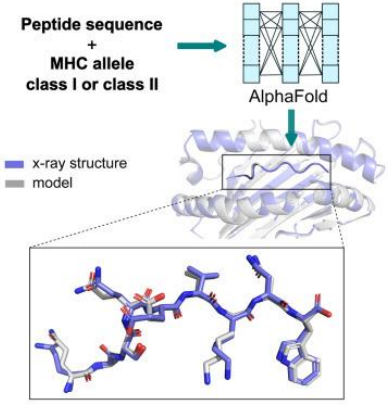
Molecular Dynamics

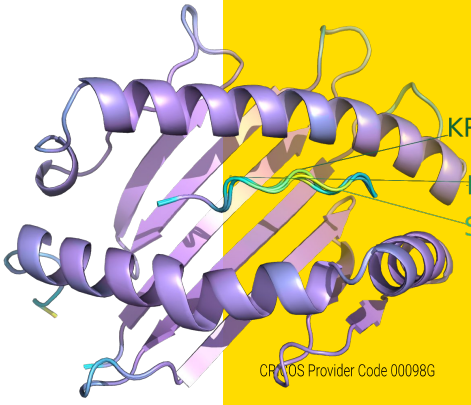
Antibiotic Resistance – Tree Lab – RFoldNA

'Golden Staph' vancomycin tolerance *via* RNA structures



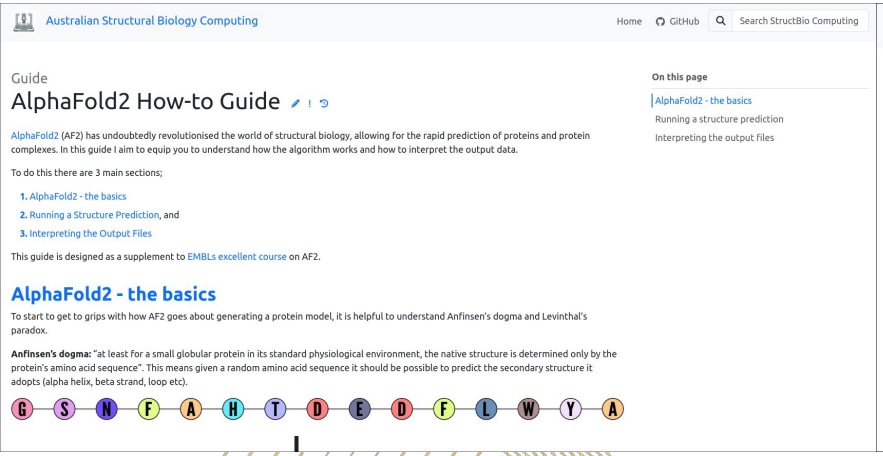
Immune response – Tedla Lab – T(Cell)Fold

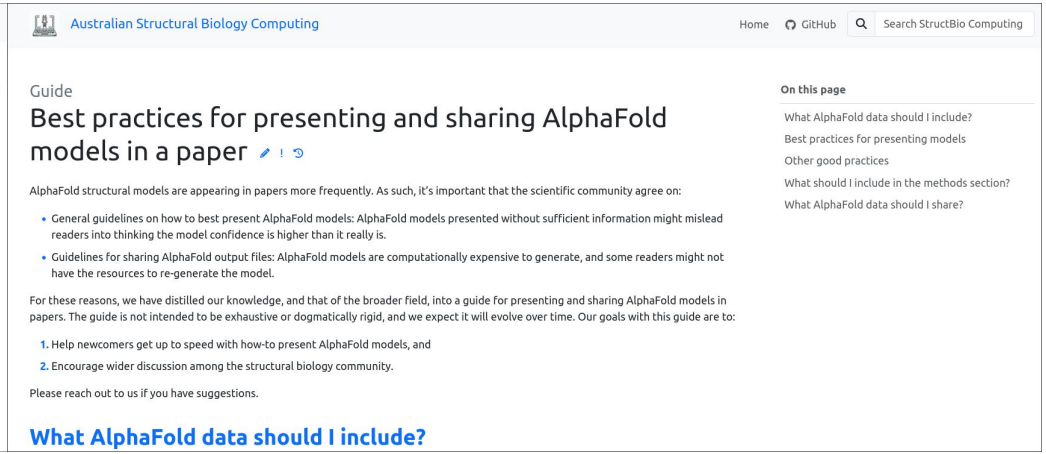




Building the Australian Computational Structural Biology Community

australian-structural-biology-computing.github.io/website/





Contributors

The projects included here represent a joint effort by the following people at multiple Australian institutions.

James Lingford

Research Officer, Greening Lab, Monash University

Johan Gustafsson

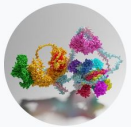
Engagement, Australian BioCommons, University of Melbourne

Kate Michie

Chief Scientist, Structural Biology Facility UNSW

Keiran Rowell

Computational Scientist, Structural Biology Facility UNSW



Michael Healy

Postdoctoral Research Fellow, University of Queensland

CRICOS Provider Code 00098G

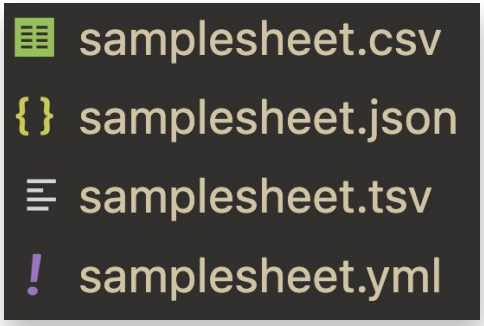
UNSW SYDNEY

Problems Facing Computational Protein Folding Adoption

Command line

Command line can be very finicky. Researchers shouldn't have to touch it.

File Formats

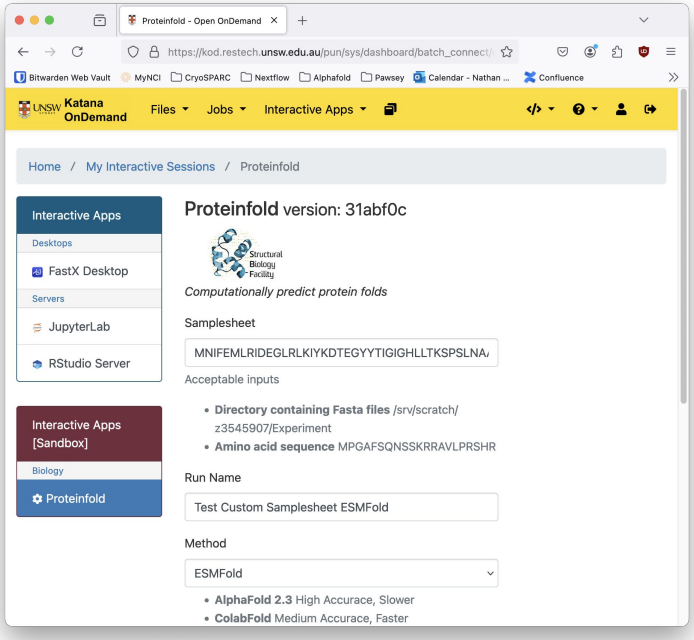


Different folding classes want different file formats and different parameters.

Resource Allocation

What resources should you allocate to your job? How much do you need?

Solutions: Command Line

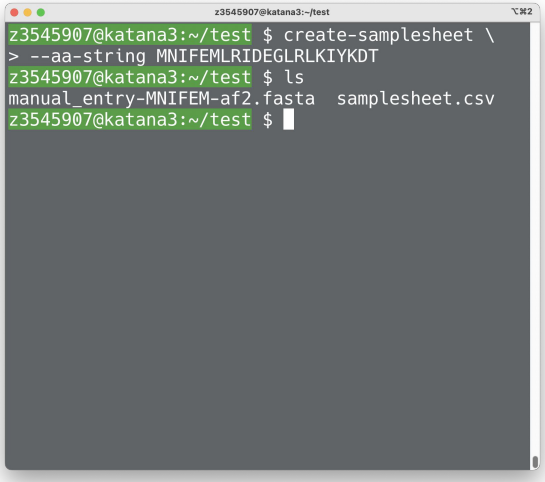


OnDemand

- Provides a nice frontend for users
- Easily manage previous and currently running jobs
- Integrates with local cluster schedulers
- Easily add, remove and modify fields

CRICOS Provider Code 00098G

Solutions: Samplesheet



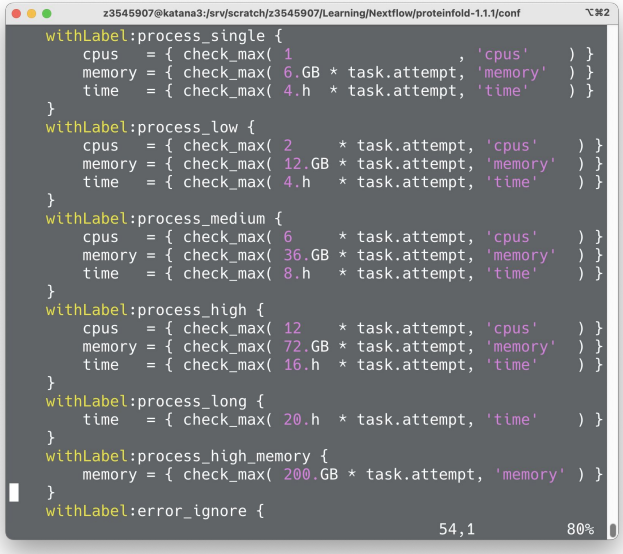
samplesheet-utils

- Easily create samplesheets from directories containing FASTA files
- Create a samplesheet from manual input sequence
- Translate samplesheets to different formats
- Easily integratable into Nextflow pipelines

https://github.com/Australian-Structural-Biology-Computing/create-samplesheet

CRICOS Provider Code 00098G

Solutions: Resource Allocation



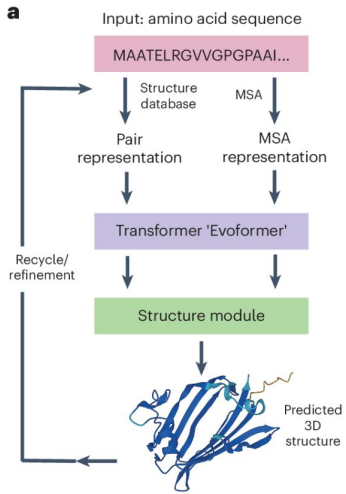
Nextflow

- Can dynamically choose how many resources to allocate depending on the task
- Automatically parallelise jobs
- Automatically retry with more resources if jobs run out of memory, etc.

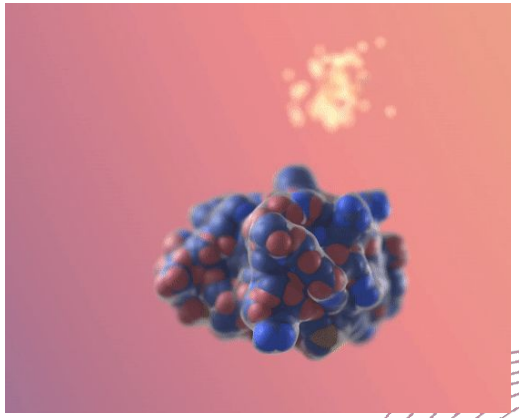
CRICOS Provider Code 00098G

Advancements in Structural Software: Predicting, Generating, and Aligning Biomolecular Structures

Structure Prediction



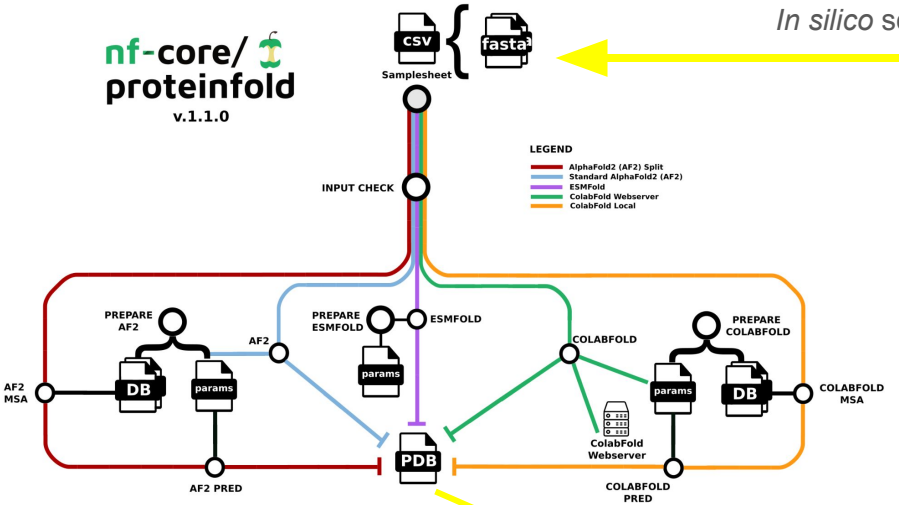
Diffusion Models



Structure Alignment

CRICOS Provider Code 00098G

Future Directions in Binder Design and Structural Biology



In silico screen

Sequence design

ProteinMPNN

Diffusion Models

RFDiffusion

BindCraft

++

RoseTTAFold-All-Atom

HelixFold3

UNSW SYDNEY

CRICOS Provider Code 00098G