Linking text mining and structure validation for reviewer support

Melanie Vollmar

ARISE/EMBL/Marie Curie Fellow

PDBe Team, EMBL-EBI

ARISE project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945405.

EMBL

Content

- About me
- Aim of my fellowship
- Challenge statement: Current setup for paper reviewing
- Progress of long-secondment project
- Next steps
- Acknowledgements

Melanie Vollmar

- University of Regensburg, Germany: undergraduate studies in biology with a diploma in biochemistry;

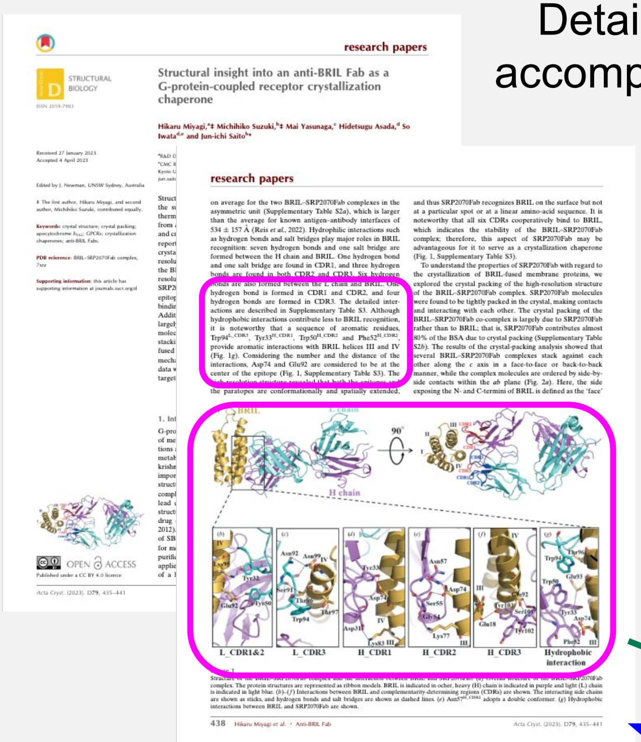
- University of Duesseldorf, Germany: PhD studies; structural biology; FoF1 ATP synthase in spinach chloroplasts

- 3 postdoc positions
  - Structural Genomics Consortium, University of Oxford, UK
  - University of Manchester, UK
  - Diamond Light Source Ltd, UK

- EMBL-EBI, UK: ARISE fellowship; NLP tools for text mining of scientific literature to extract residue-level, functional annotations for enrichment of protein structures in PDBe

Postdoc at Diamond Light Source enabled career change from structural biology to software development and machine learning

Aim of my fellowship - accelerating biocuration in the Life Sciences


research papers

STRUCTURAL BIOLOGY

ISSN 2059-7983

Structural insight into an anti-BRIL Fab as a G-protein-coupled receptor crystallization chaperone

Hikaru Miyagi,[a]‡ Michihiko Suzuki,[b]‡ Mai Yasunaga,[c] Hidetsugu Asada,[d] So Iwata[d,e] and Jun-ichi Saito[b]*

Received 27 January 2023
Accepted 4 April 2023

Edited by J. Newman, UNSW Sydney, Australia

‡ The first author, Hikaru Miyagi, and second author, Michihiko Suzuki, contributed equally.

research papers

Detailed descriptions to accompany and support the figures

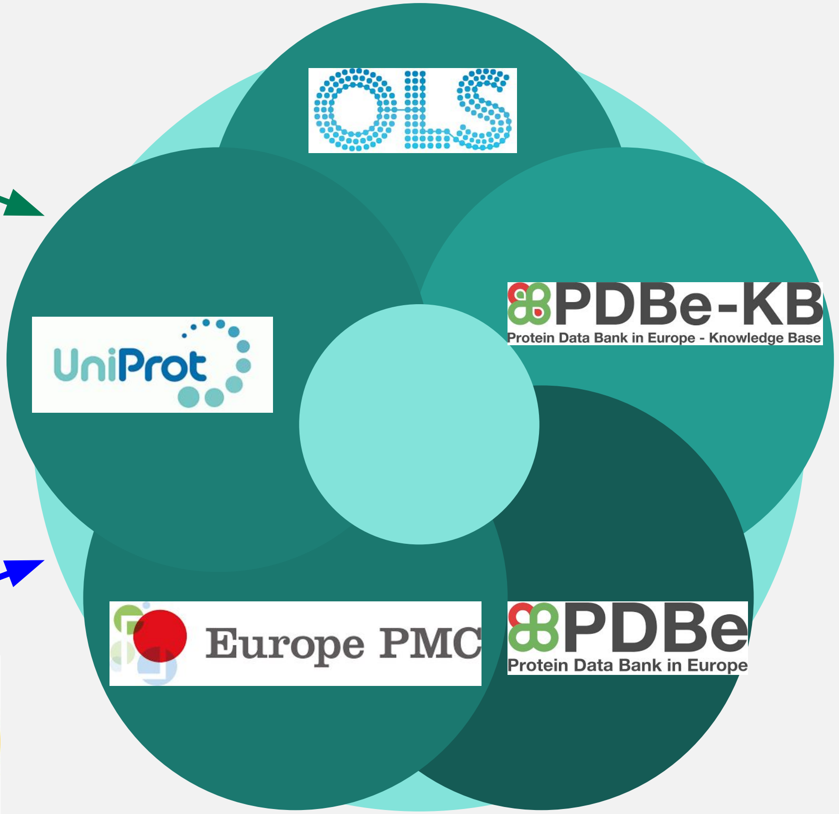
Beautiful figures of residue level functions, interactions and processes in proteins

Manual biocuration

Assisted biocuration

Source: Craiyon.com/EE Times

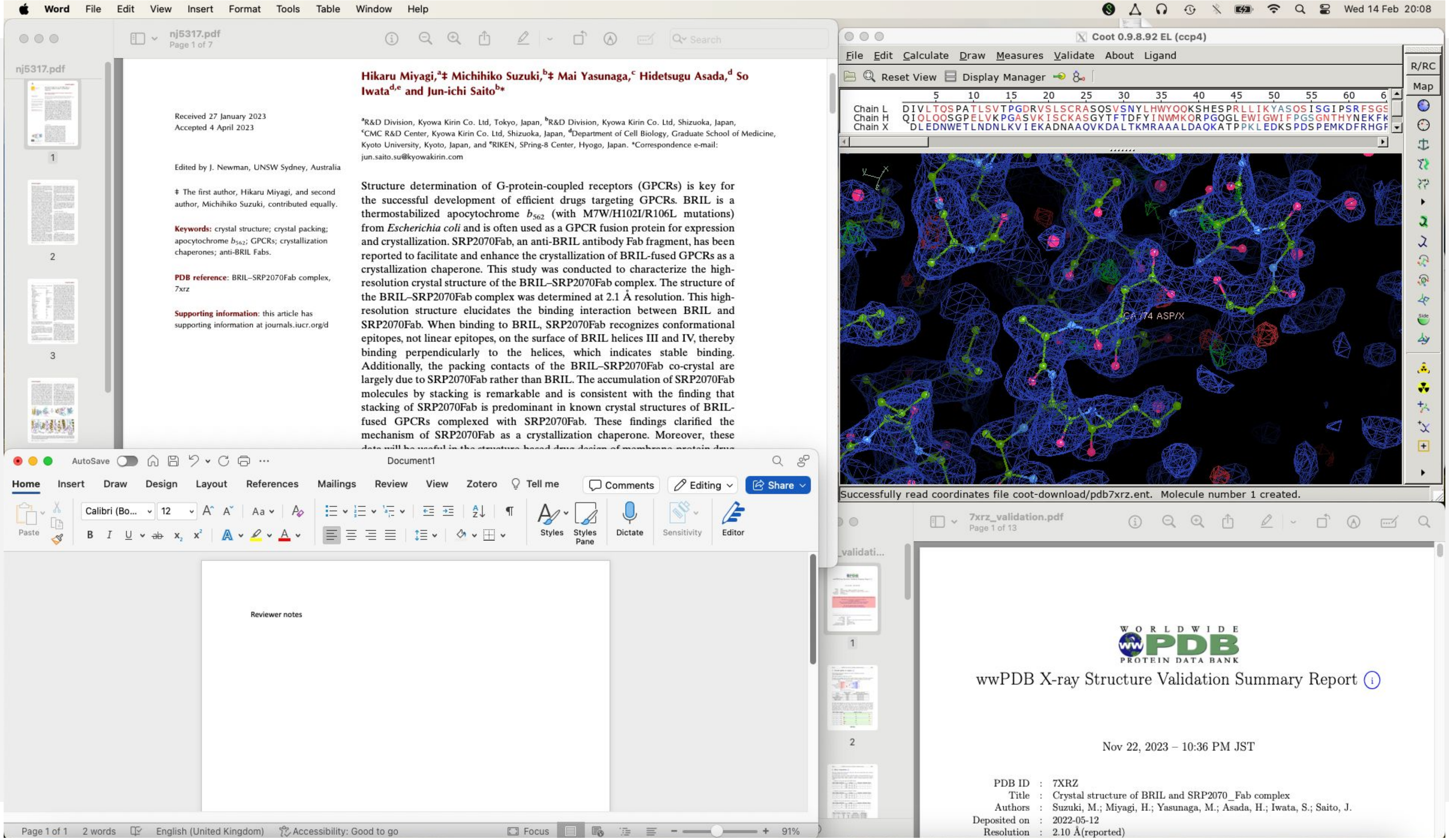
Public, open source databases



ARISE

EMBL

Challenge statement: Current setup for paper reviewing



Hikaru Miyagi,[a]‡ Michihiko Suzuki,[b]‡ Mai Yasunaga,[c] Hidetsugu Asada,[d] So Iwata[d,e] and Jun-ichi Saito[b]*

[a]R&D Division, Kyowa Kirin Co. Ltd, Tokyo, Japan, [b]R&D Division, Kyowa Kirin Co. Ltd, Shizuoka, Japan, [c]CMC R&D Center, Kyowa Kirin Co. Ltd, Shizuoka, Japan, [d]Department of Cell Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, and [e]RIKEN, SPring-8 Center, Hyogo, Japan. *Correspondence e-mail: jun.saito.su@kyowakirin.com

Received 27 January 2023
Accepted 4 April 2023

Edited by J. Newman, UNSW Sydney, Australia

‡ The first author, Hikaru Miyagi, and second author, Michihiko Suzuki, contributed equally.

Keywords: crystal structure; crystal packing; apocytochrome b_{562}; GPCRs; crystallization chaperones; anti-BRIL Fabs.

PDB reference: BRIL–SRP2070Fab complex, 7xrz

Supporting information: this article has supporting information at journals.iucr.org/d

Structure determination of G-protein-coupled receptors (GPCRs) is key for the successful development of efficient drugs targeting GPCRs. BRIL is a thermostabilized apocytochrome b_{562} (with M7W/H102I/R106L mutations) from *Escherichia coli* and is often used as a GPCR fusion protein for expression and crystallization. SRP2070Fab, an anti-BRIL antibody Fab fragment, has been reported to facilitate and enhance the crystallization of BRIL-fused GPCRs as a crystallization chaperone. This study was conducted to characterize the high-resolution crystal structure of the BRIL–SRP2070Fab complex. The structure of the BRIL–SRP2070Fab complex was determined at 2.1 Å resolution. This high-resolution structure elucidates the binding interaction between BRIL and SRP2070Fab. When binding to BRIL, SRP2070Fab recognizes conformational epitopes, not linear epitopes, on the surface of BRIL helices III and IV, thereby binding perpendicularly to the helices, which indicates stable binding. Additionally, the packing contacts of the BRIL–SRP2070Fab co-crystal are largely due to SRP2070Fab rather than BRIL. The accumulation of SRP2070Fab molecules by stacking is remarkable and is consistent with the finding that stacking of SRP2070Fab is predominant in known crystal structures of BRIL-fused GPCRs complexed with SRP2070Fab. These findings clarified the mechanism of SRP2070Fab as a crystallization chaperone. Moreover, these

Successfully read coordinates file coot-download/pdb7xrz.ent. Molecule number 1 created.

Reviewer notes

WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report

Nov 22, 2023 – 10:36 PM JST

PDB ID : 7XRZ
Title : Crystal structure of BRIL and SRP2070_Fab complex
Authors : Suzuki, M.; Miyagi, H.; Yasunaga, M.; Asada, H.; Iwata, S.; Saito, J.
Deposited on : 2022-05-12
Resolution : 2.10 Å(reported)

Aim of the long-secondment project at the IUCr

research papers

STRUCTURAL BIOLOGY

ISSN 2059-7983

Structural insight into an anti-BRIL Fab as a G-protein-coupled receptor crystallization chaperone

Received 27 January 2023
Accepted 4 April 2023

Edited by J. Newman, UNSW Sydney, Austra

‡ The first author, Hikaru Miyagi, and secon author, Michihiko Suzuki, contributed equall

Keywords: crystal structure; crystal packing; apocytochrome b_{562}; GPCRs; crystallization chaperones; anti-BRIL Fabs.

PDB reference: BRIL–SRP2070Fab complex, 7xrz

Supporting information: this article has supporting information at journals.iucr.org/d

OPEN ACCESS
Published under a CC BY 4.0 licence

Acta Cryst. (2023). D79, 435–441

IUCrJ MATERIALS | COMPUTATION research papers

The predictive power of data processing statistics

Authors

Melanie Vollmar[a]*, James M Parkhurst[a,b], Dominic Jaques[a], Arnaud Baslé[c], Garib N Murshudov[b], David Waterman[d] and Gwyndaf Evans[a]*

[a] Diamond Light Source Ltd, Harwell Science & Innovation Campus, Didcot, OX11 0DE, United Kingdom

[b] MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge, CB2 0QH, United Kingdom

[c] Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 1HH, United Kingdom

[d] Science Technology and Facilities Council, Rutherford Appleton Laboratory, Didcot, OX11 0FA, United Kingdom

Correspondence email: melanie.vollmar@diamond.ac.uk; gwyndaf.evans@diamond.ac.uk

Funding information Biotechnology and Biological Sciences Research Council (grant No. BB/L007398/1).

Synopsis One or two sentences suitable for the Journal contents listing (style: IUCr synopsis).

Abstract We present a method to estimate the likelihood of success in determining a macromolecular structure by X-ray crystallography and experimental SAD or MAD phasing based on initial data processing statistics and sample crystal properties. Such a predictive tool can rapidly assess the usefulness of data and guide the collection of an optimal dataset. The increase in data rates from modern macromolecular crystallography beamlines, together with a demand from users for real-time feedback, has led to pressure on computational resources and a need for smarter data handling. We have applied statistical and machine learning methods to construct a classifier that displays 95% accuracy for training and test datasets compiled from 434 solved structures. Applying this classifier to new data achieved 79% accuracy. These scores already provide clear guidance as to the effective use of computing resources and offer a starting point for a personalised data collection assistant.

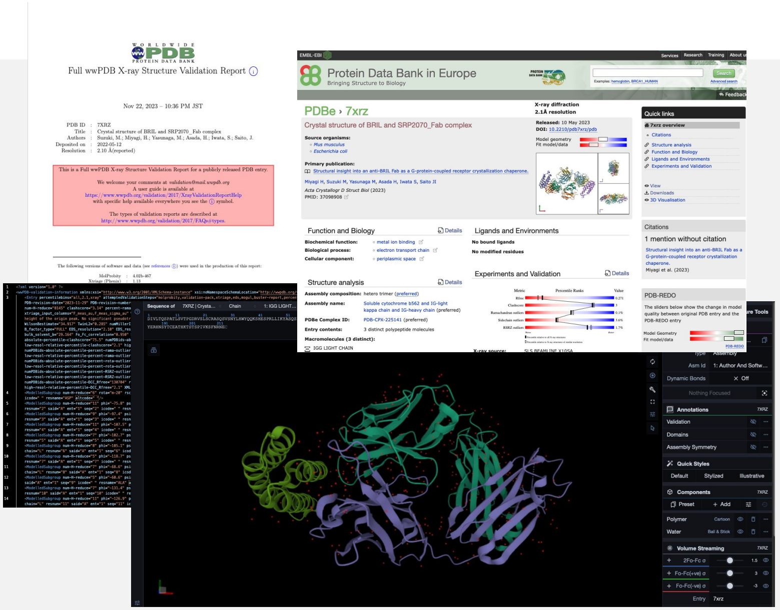
Keywords: Three or four key words/phrases separated by semi-colons [style: IUCr keywords].

1. Introduction

IMPORTANT: this document contains embedded data - to preserve data integrity, please ensure where possible that the IUCr Word tools (available from http://journals.iucr.org/services/doctemplate/) are installed when editing this document.

Publication submission and hosting

AI magic black box



WORLDWIDE PDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report

Nov 22, 2023 – 10:36 PM JST

PDB ID : 7XRZ
Title : Crystal structure of BRIL and SRP2070_Fab complex
Authors : Suzuki, M.; Miyagi, H.; Yasunaga, M.; Asada, H.; Iwata, S.; Saito, J.
Deposited on : 2022-05-12
Resolution : 2.10 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13

Protein Data Bank in Europe
Bringing Structure to Biology

PDBe › 7xrz

Crystal structure of BRIL and SRP2070_Fab complex

X-ray diffraction
2.1Å resolution

Released: 10 May 2023
DOI: 10.2210/pdb7xrz/pdb

Source organisms:
- *Mus musculus*
- *Escherichia coli*

Primary publication:
Structural insight into an anti-BRIL Fab as a G-protein-coupled receptor crystallization chaperone.
Miyagi H, Suzuki M, Yasunaga M, Asada H, Iwata S, Saito JI
Acta Crystallogr D Struct Biol (2023)
PMID: 37098908

Function and Biology
Biochemical function: metal ion binding
Biological process: electron transport chain
Cellular component: periplasmic space

Ligands and Environments
No bound ligands
No modified residues

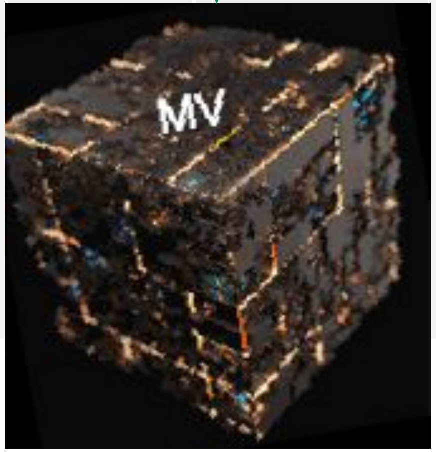
Structure analysis
Assembly composition: hetero trimer (preferred)
Assembly name: Soluble cytochrome b562 and IG-light kappa chain and IG-heavy chain (preferred)
PDBe Complex ID: PDB-CPX-225141 (preferred)
Entry contents: 3 distinct polypeptide molecules
Macromolecules (3 distinct):
IGG LIGHT CHAIN

Experiments and Validation

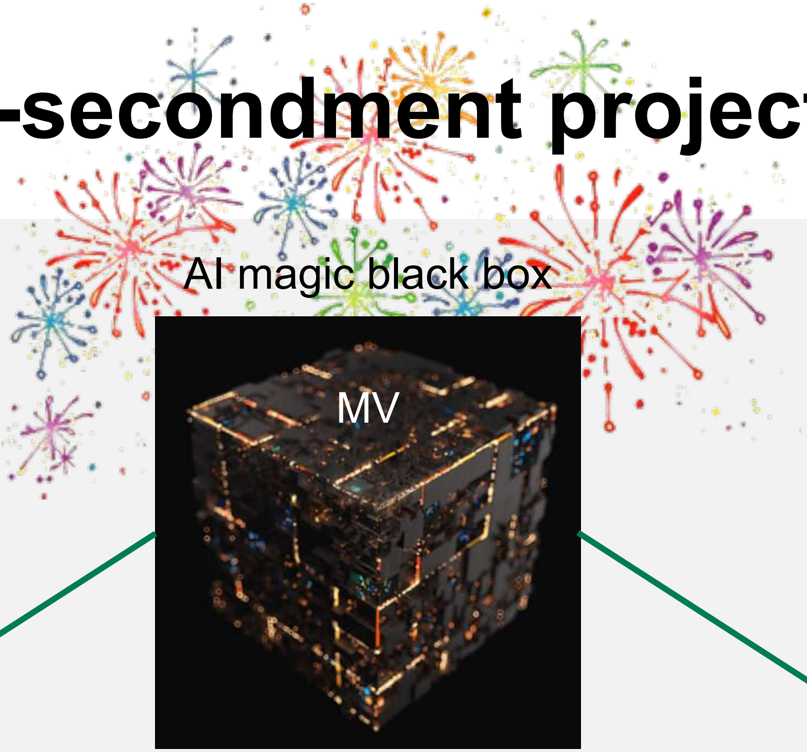
Citations
1 mention without citation
Structural insight into an anti-BRIL Fab as a G-protein-coupled receptor crystallization chaperone.
Miyagi et al. (2023)

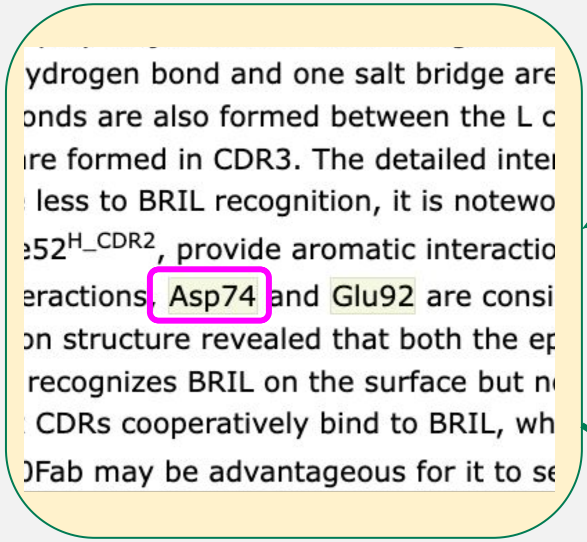
PDB-REDO
The sliders below show the change in model quality between original PDB entry and the PDB-REDO entry

Structure submission, validation and hosting



Aim of the long-secondment project at the IUCr


AI magic black box

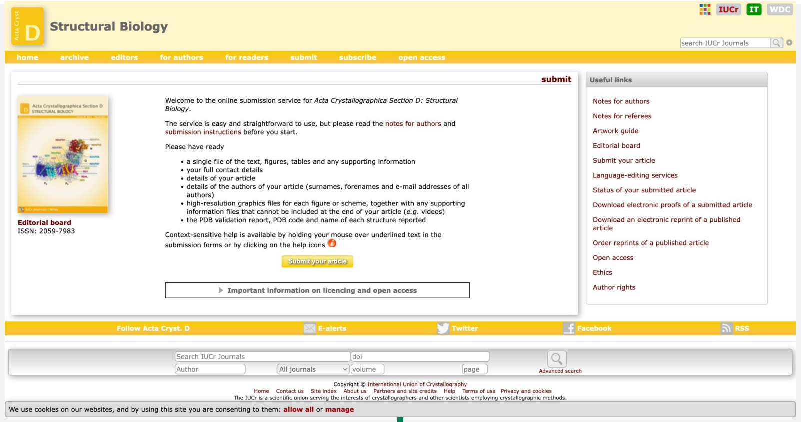

ydrogen bond and one salt bridge are
onds are also formed between the L c
re formed in CDR3. The detailed inte
less to BRIL recognition, it is notewo
52$^{\text{H\_CDR2}}$, provide aromatic interactio
eractions. Asp74 and Glu92 are consi
on structure revealed that both the ep
recognizes BRIL on the surface but n
CDRs cooperatively bind to BRIL, wh
Fab may be advantageous for it to se

PDB	7XRZ	7XRZ
chain	X	Y
PDB residue name	ASP	ASP
PDB residue number	74	74
PDB residue	ASP74	ASP74
Ramachandran score	Favored	Favored
RSCC	0.914	0.955
Alternative conformation		
clashes		
Wildtype PDB residue		
UniProt ID	P0ABE7	P0ABE7
UniProt name	C562_ECOLX	C562_ECOLX
UniProt residue	ASP96	ASP96
UniProt URI	https://www.uniprot.org/uniprot/C562_ECOLX	https://www.uniprot.org/uniprot/C562_ECOLX

ARISE

EMBL

Next steps



Structural Biology

submit

Welcome to the online submission service for *Acta Crystallographica Section D: Structural Biology*.

The service is easy and straightforward to use, but please read the notes for authors and submission instructions before you start.

Please have ready

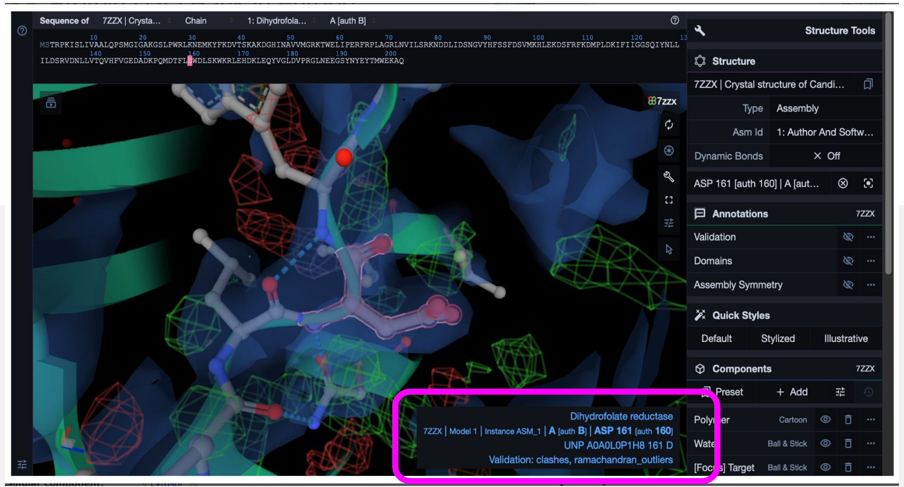
- a single file of the text, figures, tables and any supporting information
- your full contact details
- details of your article
- details of the authors of your article (surnames, forenames and e-mail addresses of all authors)
- high-resolution graphics files for each figure or scheme, together with any supporting information files that cannot be included at the end of your article (*e.g.* videos)
- the PDB validation report, PDB code and name of each structure reported

Context-sensitive help is available by holding your mouse over underlined text in the submission forms or by clicking on the help icons

Editorial board
ISSN: 2059-7983

Copyright © International Union of Crystallography
Home Contact us Site index About us Partners and site credits Help Terms of use Privacy and cookies
The IUCr is a scientific union serving the interests of crystallographers and other scientists employing crystallographic methods.

We use cookies on our websites, and by using this site you are consenting to them: allow all or manage



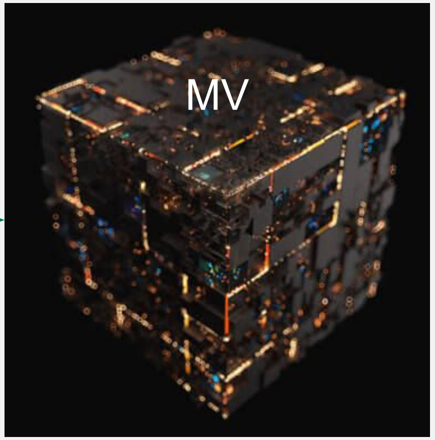
Sequence of 7ZZX | Crysta… Chain 1: Dihydrofola… A [auth B]

```
          10        20        30        40        50        60        70        80        90        100       110       120
MSTRPKISLIVAALQPSMGIGAKGSLPWRLKNEMKYFKDVTSKAKDGHINAVVMGRKTWELIPERFRPLAGRLNVILSRKNDDLIDSNGVYHFSSFDSVMKHLEKDSFRFKDMPLDKIFIIGGSQIYNLL
     140       150       160       170       180       190       200
ILDSRVDNLLVTQVHFVGEDADKPQMDTFLDWDLSKWKRLEHDKLEQYVGLDVPRGLNEEGSYNYEYTMWEKAQ
```

7ZZX | Crystal structure of Candi…
Type: Assembly
Asm Id: 1: Author And Softw…
Dynamic Bonds: Off
ASP 161 [auth 160] | A [aut…

Dihydrofolate reductase
7ZZX | Model 1 | Instance ASM_1 | A [auth B] | ASP 161 [auth 160]
UNP A0A0L0P1H8 161 D
Validation: clashes, ramachandran_outliers

AI magic black box

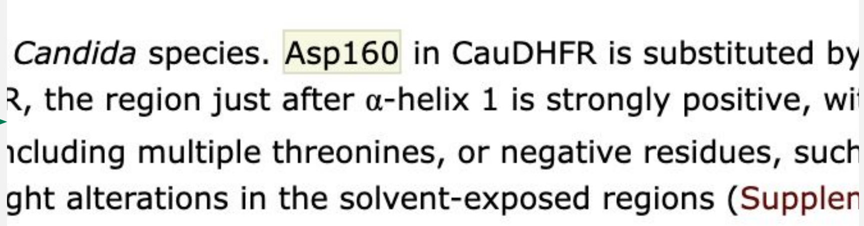


PDF

JATS XML

BioC XML

HTML



Candida species. Asp160 in CauDHFR is substituted by
R, the region just after α-helix 1 is strongly positive, wi
ıcluding multiple threonines, or negative residues, such
ght alterations in the solvent-exposed regions (Supplen

Acknowledgements


Deborah Harrus,
PDBe


Genevieve Evans,
PDBe


Deepti Gupta,
PDBe


Romana Gaborova,
PDBe


David Armstrong,
PDBe


Marcelo Querino,
PDBe


Sameer Velankar,
PDBe


Santosh Tirunagari,
EuropePMC

IUCr

Peter Strickland
Louise Jones
Simon Westrip
Koh Song Sang

arise@embl.org

ARISE - Career Accelerator for Research Infrastructure Scientists

ARISE project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945405.