

Integrative structural modelling of fission yeast contractile ring and analysis of actin-protein interactions

A thesis

Submitted towards the partial fulfilment of
BS-MS dual degree programme
by

Omkar Manojkumar Golatkar



DATE: 27/03/2024

under the guidance of

Dr. Gayathri Pananghat

Department of Biology

Indian Institute of Science Education and Research Pune

from May 2023 to Mar 2024

Certificate

This is to certify that this dissertation entitled '**Integrative structural modelling of fission yeast contractile ring and analysis of actin-protein interactions**' submitted towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by **Omkar Manojkumar Golatkar** at Indian Institute of Science Education and Research Pune, under the supervision of **Dr. Gayathri Pananghat**, Associate Professor, Department of Biology, during academic year May 2023 to May 2024.



Dr. Gayathri Pananghat

Committee:

Supervisor:

Dr. Gayathri Pananghat

Associate Professor

Department of Biology, IISER Pune

Expert:

Dr. Richa Rikhy

Professor

Department of Biology, IISER Pune

Declaration

I, hereby declare that the matter embodied in the report titled '**Integrative structural modelling of fission yeast contractile ring and analysis of actin-protein interactions**' are the results of the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research Pune under the supervision of **Dr. Gayathri Pananghat** and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgement of collaborative research and discussions.



Omkar Golatkar

20191223

BS-MS

IISER Pune

Date: 27/03/2024

Acknowledgements

I would like to thank all the individuals who directly or indirectly helped me throughout the project.

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Gayathri Pananghat. While she was the most understanding when I was exploring and trying out new things during the project, she also guided me to the right path when I was going astray. Without her constant support and guidance, it would not have been possible to take this project to its current state.

I would like to thank my TAC member, Dr. Richa Rikhy, for the evaluation and feedback that helped shape the project in a better manner.

I'd like to acknowledge my previous lab mates, Prajakta and Ranjana, for introducing and encouraging me to take on such a challenging project. I would also like to thank all the current lab members for their valuable input.

Special thanks to all my friends here at IISER, with whom I shared ups and downs during the last five years. Thanks to Science Media Center team at IISER Pune, I had great time working on the science animations. Special mention to Rahul, Neha, Vivek and Dr. Shalini Sharma for their guidance during the Student Engagement Program at SMC. Thanks to IISER Pune for providing a peaceful and nurturing study environment. A deserved mention to all the inanimate objects, including my laptop and the lab computer; without them, the project ceases to exist.

Lastly, I would like to thank my parents and all my family members for their love, care and wishes for my well-being

Contributions

Contributor name	Contributor role
Omkar Golatkar	Methodology
Dr. Gayathri Pananghat	Conceptualization
Omkar Golatkar	Software
Dr. Gayathri Pananghat	Validation
Omkar Golatkar	Formal analysis
Omkar Golatkar	Investigation
Omkar Golatkar	Data Curation
Omkar Golatkar	Writing - original draft preparation
Dr. Gayathri Pananghat	Writing - review and editing
Omkar Golatkar	Visualization
Gayathri Pananghat	Supervision
Gayathri Pananghat	Project administration

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy¹

¹<https://journals.biologists.com/jcs/pages/author-contributions>

Abstract

Cytokinesis in eukaryotes includes assembly, constriction, and disassembly of the contractile ring (CR). Failure in any of these can lead to polyploidy and tumour formation. Fission yeast has been widely used as a model organism to study the process of cytokinesis; much about the protein composition and spatial distribution in the fission yeast CR is known. However, the lack of structural information about these proteins has limited the mechanistic understanding of the process. We used a recently developed structure prediction tool, AlphaFold, to predict the structures of protein complexes in fission yeast CR. Additionally, we used MSA manipulation strategies to model the actin crosslinkers, α -actinin (Ain1) and fimbrin (Fim1) in multiple conformations. Based on the analysis, we hypothesise that the Fim1 might be a more flexible crosslinker than the Ain1.

Despite single actin being present in fission yeast, its involvement in various cytoskeletal networks and many interactors prompted us to look at the actin structures in the PDB to identify the structure-function correlation between actin and actin-binding proteins. Using heatmap visualisation of the actin-protein interfaces coupled with agglomerative clustering methods, we observed that cofilin-bound actin filaments show altered actin-actin interfaces and myosin-bound actin filaments occupy two different clusters on the interface heatmap. By calculating helical parameters of actin filaments from the PDB, outliers were observed - *Plasmodium falciparum* actin-1 and cofilin-bound actin, in terms of helical twist; human β and γ -actin isoforms along with a phalloidin-bound rabbit actin filament, in terms of helical rise; pointed-end of actin having bent filament axis.

This project provides a first step in modelling the fission yeast contractile ring and a workflow for comprehensive analysis of actin structures in the PDB.

Contents

1	Introduction	7
1.1	Cytokinesis	7
1.1.1	Contractile actomyosin ring	7
1.1.2	Cytokinesis in fission yeast	9
1.2	Actin and actin-binding proteins	13
1.2.1	Actin filament helicity	13
1.3	Protein structure prediction and modelling	14
1.3.1	AlphaFold	15
1.3.2	Integrative Structural Modelling	17
1.4	Objectives	18
2	Methods	19
2.1	Dataset of actin structures	19
2.1.1	Application Programming Interface (API)	19
2.1.2	Methodology of actin dataset generation	20
2.2	Calculation of helical parameters of actin filaments and outlier detection	23
2.3	Interface matrix generation and visualisation	27
2.3.1	Heatmap and similarity matrix visualisation of actin-protein interfaces	29
2.4	Computational structure prediction	30
2.4.1	Prediction of multiple conformations	31
2.4.2	Flexibility analysis of F-actin crosslinkers	32
2.4.3	Prediction of protein complexes by delineating the input sequences	33
2.4.4	Assessment of predictions	33
2.4.5	Map of protein-protein interactions	35

2.5	Resource Table	36
3	Results and Discussion	37
3.1	Analysing flexibility of actin crosslinking proteins in fission yeast	37
3.1.1	Multiple conformations of Ain1	37
3.1.2	Multiple conformations of Fim1	40
3.1.3	Flexibility of Ain1 and Fim1	43
3.2	Protein complexes in fission yeast contractile ring	44
3.2.1	Actin-ABP interactions in CAR	47
3.2.2	Coiled-coil prediction of fission yeast myosin tails	49
3.2.3	Homodimeric CAR proteins	50
3.3	Actin structures dataset	50
3.4	Actin filament helicity	54
3.4.1	<i>Plasmodium falciparum</i> actin-1 show increased helical twist compared to the actin-2	56
3.4.2	Cofilin-bound actin filaments show a significant decrease in the helical twist from the rest of the filaments	56
3.4.3	Outlier actin filaments in terms of helical rise	57
3.4.4	Bent actin filaments and pointed-end of F-actin show local variation in helical rise	58
3.5	Actin and Actin-binding proteins	60
3.5.1	Interfaces of actin in actin filaments	62
4	Conclusions and Future Perspectives	64
	References	67
A	Additional data	76

List of Figures

1.1	Network of fission yeast contractile ring proteins	8
1.2	Cytokinesis in fission yeast	9
1.3	Models of protein organization in fission yeast contractile ring	11
1.4	Search, capture, pull and release (SCPR) model	12
1.5	Use of co-evolutionary information by AlphaFold	15
2.1	Structural organisation in the PDB	22
2.2	Schematic of dataset generation	23
2.3	Calculation of actin filament helical parameters	25
2.4	Schematic of actin-protein interface matrix generation	28
2.5	Features of actin-protein interface heatmap and similarity matrix	30
2.6	AlphaFold2 architecture and prediction strategies	31
2.7	Flowchart of flexibility analysis of F-actin crosslinkers	32
2.8	Assessment of predicted interface using PAE	34
2.9	Visualisation of predicted interface	35
3.1	Domain organisation and dimer prediction of Ain1	38
3.2	Multiple conformations of Ain1 using MSA manipulation strategies	39
3.3	Domain organisation of Fim1	40
3.4	Multiple conformations of Fim1 using MSA manipulation strategies	41
3.5	Prediction of fimbrin (Fim1) without MSA (template-based)	42
3.6	Flexibility analysis of F-actin crosslinkers, Ain1 and Fim1	44
3.7	Subnetwork of proteins in CAR	45
3.8	Act1-CH domain interface prediction	48
3.9	Highly accurate prediction of Act1-ABP interface	48

3.10 Coiled-coil prediction of tail-regions of fission yeast myosins (Myo2 and Myp2)	49
3.11 Prediction of homodimers in CAR	50
3.12 Number of actin structures with varying resolution cutoff	51
3.13 Number of actin structures with varying numbers of non-actin interacting proteins of actin	51
3.14 Number of actin filaments with different small molecules bound	53
3.15 Helical rise vs twist of the actin filaments	54
3.16 Helical rise vs twist of actin filaments with marked nucleotide states and bound small molecule	55
3.17 Helical rise vs twist for actin filaments complexed with only one non-actin protein	55
3.18 Local variation in the helical rise of barbed and pointed-end of actin filaments	59
3.19 Actin filaments that show local variation in helical rise or twist	59
3.20 Sequence logo of consensus sequence of actin entities in the dataset	60
3.21 Heatmap and row similarity matrix of all actin-protein interfaces after clustering	61
3.22 Actin-actin interfaces mapped onto an actin chain	62
3.23 Row-wise similarity matrix of actin-protein interfaces in actin filaments	63
4.1 Proposed model of Ain1 and Fim1 crosslinking of actin filaments	65

List of Tables

1.1	List of tools based on AlphaFold2 architecture	16
2.1	List of APIs used for the actin analysis	20
2.2	List of actin structures removed from the filament dataset	21
2.3	Model confidence with varying pLDDT score	33
2.4	Resource list of software and packages used	36
3.1	AlphaFold predictions of protein complexes in fission yeast CAR	46
3.2	Average rise and twist values of actin filaments bound to small molecules	53
3.3	Helical parameters of <i>Plasmodium falciparum</i> actin filaments	56
3.4	Helical parameters of cofilin-bound actin filaments	57
3.5	Actin filaments having higher helical rise compared to the others	57
3.6	Average helical rise and twist values of outlier actin filaments	58

Abbreviations

ABP Actin-binding protein

AF2 AlphaFold2

API Application Programming Interface

CAR Contractile Actomyosin Ring

IDR Intrinsically Disordered Region

IQR Inter-quartile range

ISM Integrative Structural Modelling

MSA Multiple Sequence Alignment

PAE Predicted Aligned Error

PDB Protein Data Bank

pLDDT predicted Local Distance Difference Test

SCPR Search Capture Pull and Release model

SIFTS Structure Integration with Function Taxonomy and Sequence

SPB Spindle Pole Body

Chapter 1

Introduction

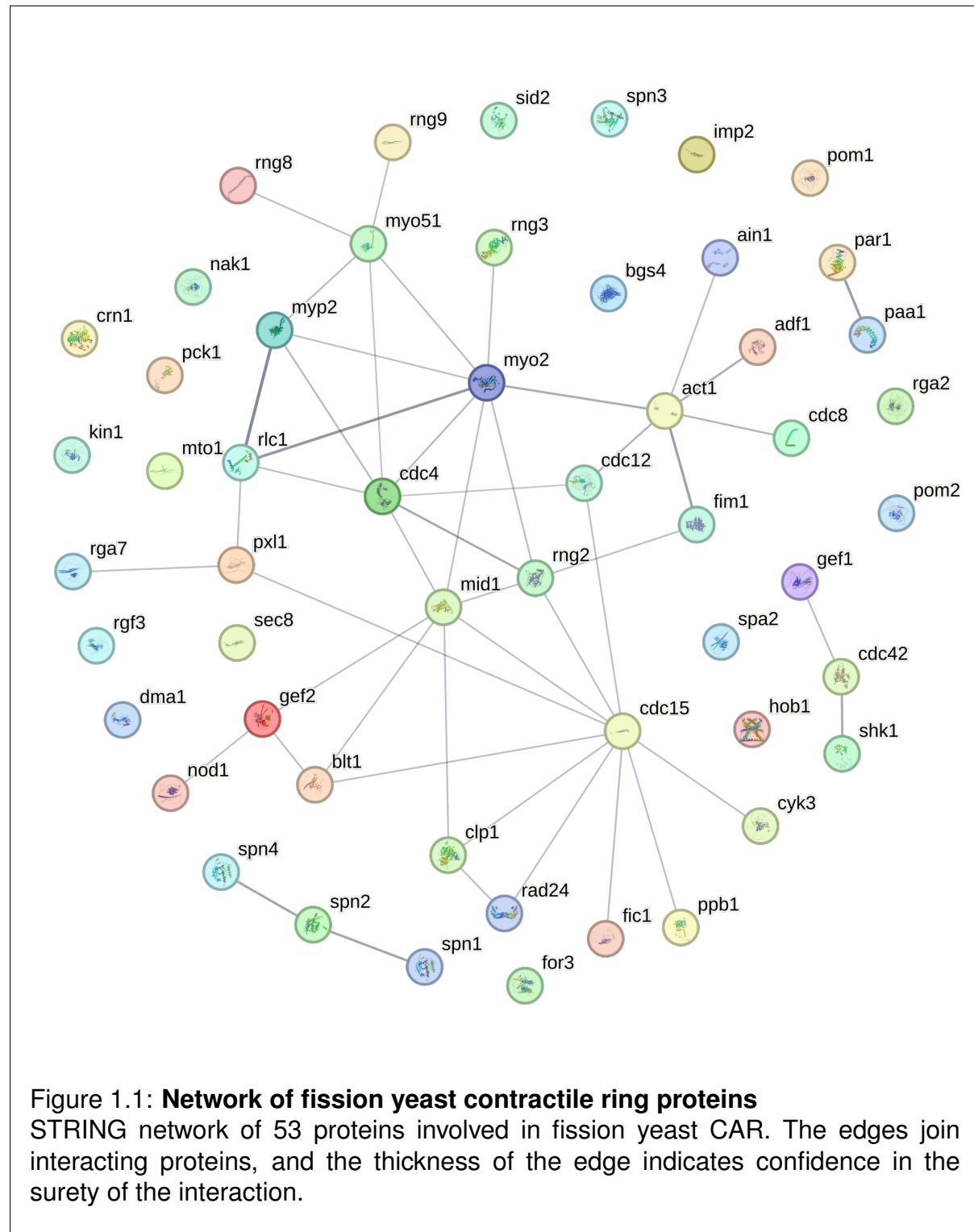
1.1 Cytokinesis

Cytokinesis is the final step in cell division, at the end of which two daughter cells are formed from one parent cell (Guertin *et al.*, 2002). In animal cells, it starts immediately after the segregation of chromosomes and ends with the abscission to physically separate the daughter cells (D'Avino *et al.*, 2015). Successful cytokinesis ensures proper segregation of genetic and cytoplasmic components into the two daughter cells (D'Avino *et al.*, 2015), and its failure has been shown to be responsible for pathological polyploidy (Fujiwara *et al.*, 2005) or aneuploidy (Gisselsson *et al.*, 2010), which are considered hallmarks of different cancer types. Apart from cancer, cytokinesis defects have also been linked to anaemia and female infertility (Lacroix and Maddox, 2012). Owing to its significance, a variety of model organisms have been used to understand the process of cytokinesis in eukaryotes such as *Schizosaccharomyces pombe* (Pollard and Wu, 2010), *Schizosaccharomyces cerevisiae* (Bhavsar-Jog and Bi, 2017), *Caenorhabditis elegans* (Pintard and Bowerman, 2019), and echinoderm zygotes (von Dassow *et al.*, 2009).

1.1.1 Contractile actomyosin ring

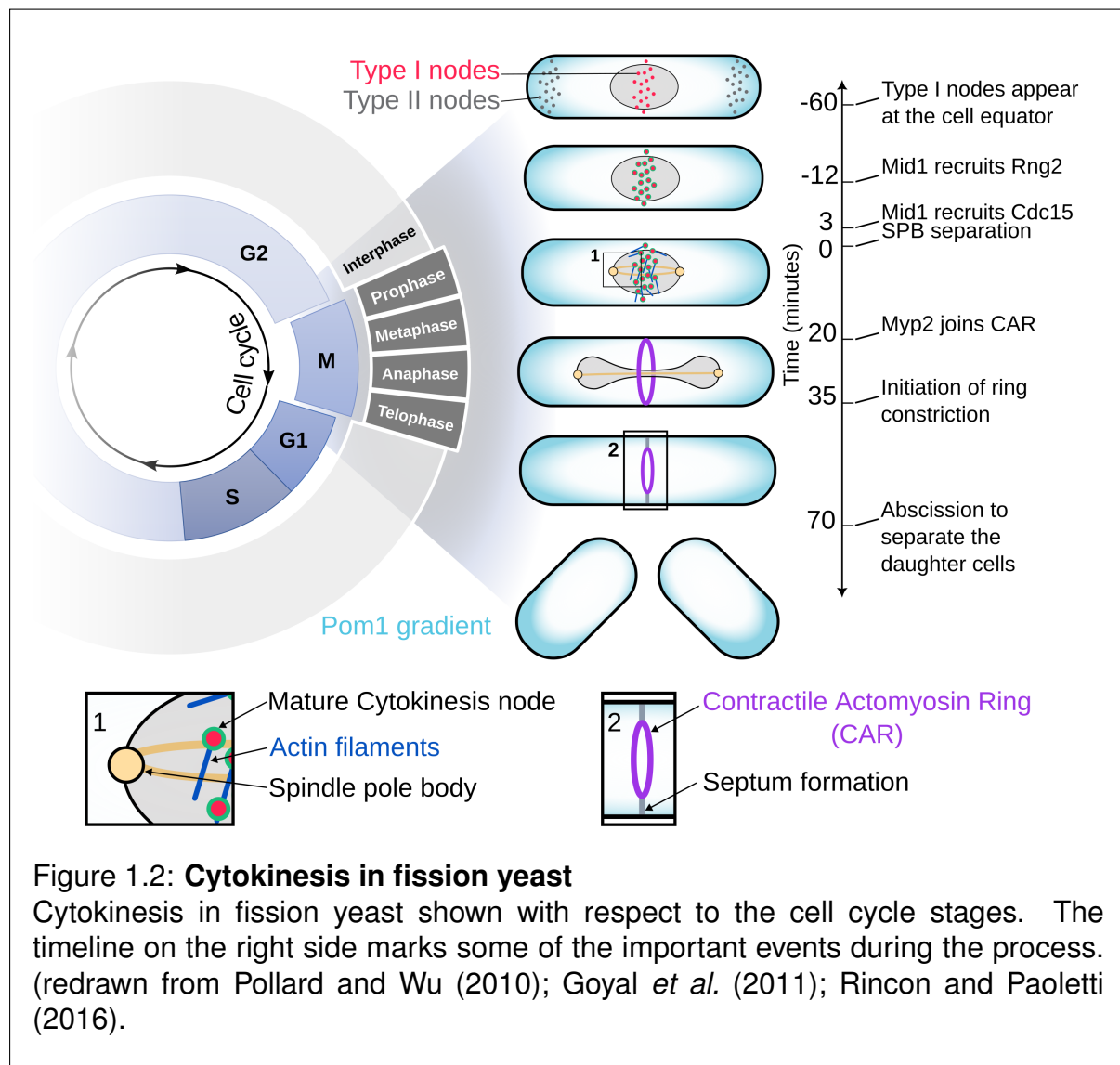
In most eukaryotic cells, cytokinesis occurs via the formation, maturation, constriction, and disassembly of a ring-like structure called the **Contractile Actomyosin Ring (CAR)** (Guertin *et al.*, 2002). The ring consists of two major

proteins, actin and myosin, and other associated proteins forming cytokinesis nodes. A STRING network (Szklarczyk *et al.*, 2015) of proteins involved in the *Schizosaccharomyces pombe* contractile ring (53, according to Pombase (Wood *et al.*, 2012)) is shown in Figure 1.1.



1.1.2 Cytokinesis in fission yeast

Fission yeast and animal cells share most of the genes involved in cytokinesis. Nearly 130 genes in fission yeast out of a total of 4900 are involved in cytokinesis. Conditional mutants are available for many of them (Pollard and Wu, 2010). Most of these 130 genes are non-redundant in terms of their functions. Owing to these advantages offered by fission yeast as a model organism, cytokinesis is a well-studied process in the fission yeast, and it provides insights into eukaryotic cytokinesis.



The main steps involved in the fission yeast cytokinesis are cytokinesis node formation and maturation, nucleation of actin filaments from the nodes, assembly and

constriction of the contractile ring and septum formation (as shown in Figure 1.2). The cytokinesis starts by defining the division plane at the cell equator by precursor nodes or type I nodes. Mainly, the negative signal from the Pom1 gradient restricts the type I nodes to the geometrical centre of the cell (Rincon and Paoletti, 2016). Unlike type I nodes, the type II nodes are remnants of the preceding cell cycle cytokinesis, they fuse with the type I nodes to form the cytokinesis nodes. The central positioning of the division plane is further reinforced by the anillin-like protein - Mid1, which binds to the plasma membrane at the cell equator after getting exported from the nucleus (Rincon and Paoletti, 2016). During the node maturation, Mid1 recruits Rng2 (IQGAP) 12 minutes prior to Spindle Pole Body (SPB) separation. This promotes the recruitment of the Myo2 and myosin light chains, Cdc4 and Rlc1, that bind to the IQ motifs in Myo2 and Rng2. This is one functional module of the cytokinesis node. Independently, Mid1 also recruits Cdc15 (F-BAR protein) 3 minutes before SPB separation, which also binds to the plasma membrane. Cdc15 recruits the formin Cdc12 (Willet *et al.*, 2015). Actin filaments nucleate from formin (Cdc12) in a profilin (Cdc3)-dependent manner (Kovar *et al.*, 2003).

Kamasaki *et al.* counted the number of actin filaments to be around 1000-2000 in the *S. pombe* CAR at different cell cycle stages. However, they could not reveal the structure of any other protein involved in the CAR. Quantitative fluorescence studies were able to quantify the number of several such proteins (Wu and Pollard, 2005; Laporte *et al.*, 2011). Superresolution microscopy experiments by Laplante *et al.* (2016) and McDonald *et al.* (2017) also shed light on the spatial organisation of these proteins relative to the plasma membrane (Figure 1.3).

The ring maturation starts 11 minutes after the SPB separation and continues for 25 minutes (Figure 1.2). Although the contractile ring does not change in size during this time, its protein composition changes (Wu and Pollard, 2005). ~20 minutes after the SPB separation, the unconventional myosin, Myp2, joins the CAR, but Mid1 leaves the ring. Imp2, Fic1, and Pxl1 stabilise the ring structure, although very little is known about the structure and function of these proteins.

A model called Search, Capture, Pull, and Release (SCPR) has been proposed to explain the formation of a tight band of the contractile ring from the loose F-actin network during maturation (Figure 1.4). As per the model, Myo2 protruding from one

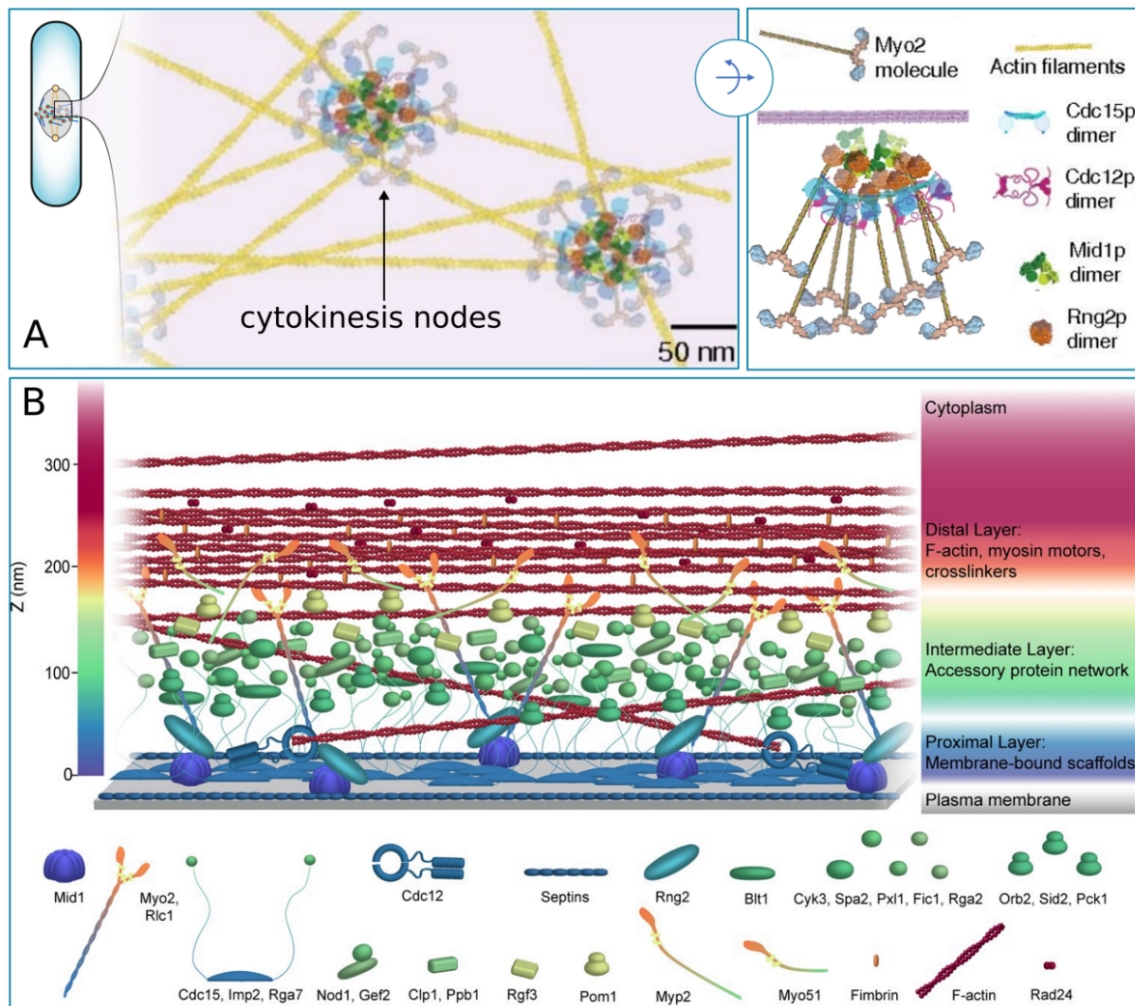
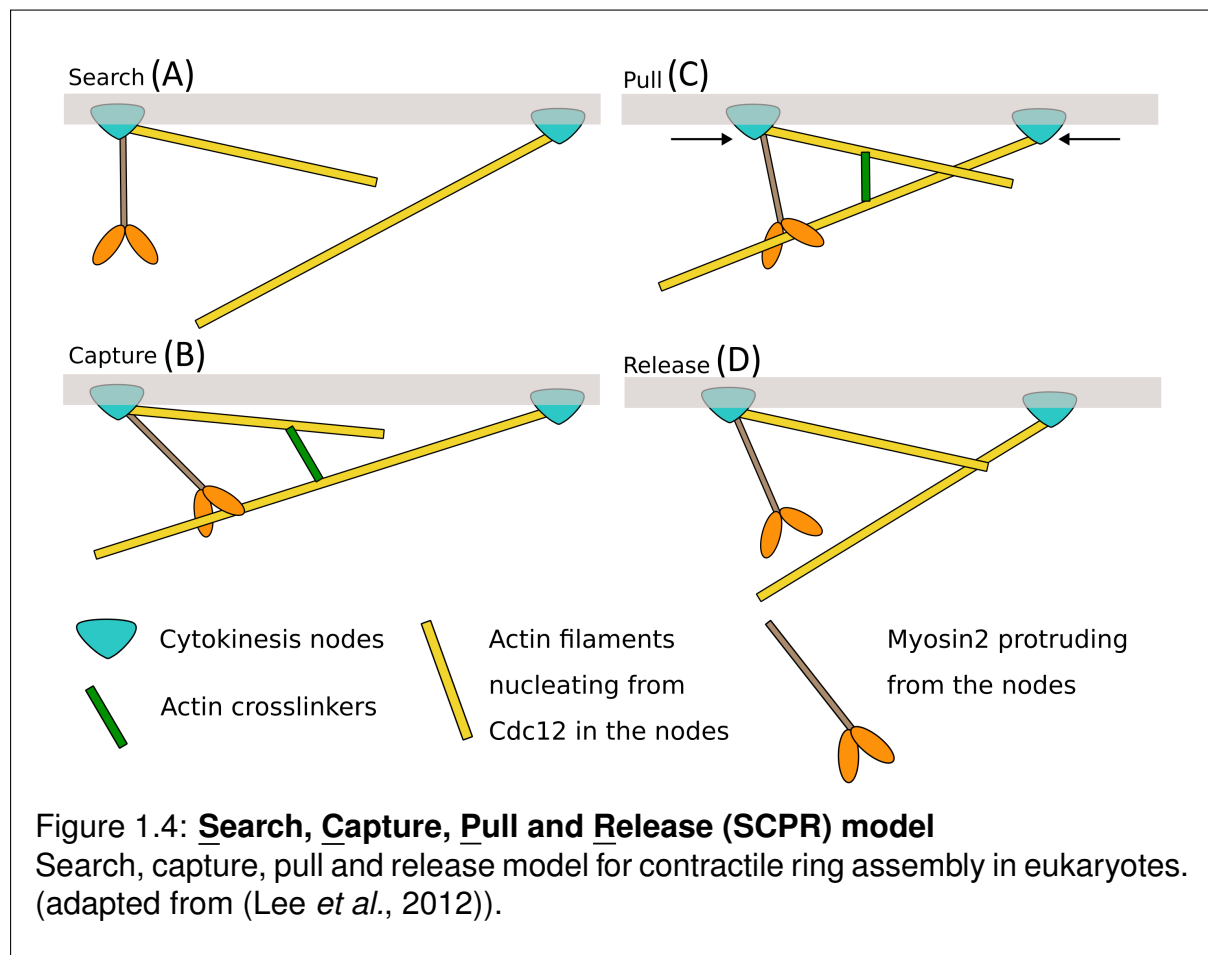


Figure 1.3: Models of protein organization in fission yeast contractile ring
 (A) Top and side-view of the proposed model of cytokinesis nodes, based on superresolution microscopy experiment (adapted from (Laplante *et al.*, 2016)) (B) Spatial arrangement of proteins in *S. pombe* contractile ring at the nanoscale (adapted from (McDonald *et al.*, 2017)).

node searches for binding sites on F-actin nucleating from Cdc12 of another node (Figure 1.4 A). The Myo2 head domain transiently captures the F-actin and pulls the two nodes closer to each other (Figure 1.4 B and C). Then, the head domain releases F-actin and searches for other actin filaments (Figure 1.4 D). The transient interactions might be controlled by the F-actin severing protein, Adf1. Such recurring events eventually lead to a tight contractile ring. To mediate the Act1-Myo2 interaction, crosslinkers (Fim1, Ain1) and actin-bundlers (Cdc12, Rng2) are proposed to help align actin filaments horizontally (Wu *et al.*, 2001; Rincon and Paoletti, 2016). Approximately 35 minutes after SPB separation, the ring starts to constrict at the rate of 5 nms^{-1} . However, not much is known about the ring constriction and disassembly.



Despite knowing the spatial and temporal occurrence of contractile ring proteins within the ring, understanding mechanistic details about cytokinesis in fission yeast has been hampered due to a lack of structural information at the atomic level. The highly dynamic nature of the proteins involved in the ring and the ring itself makes it

difficult to perform structure determination experiments. Many important proteins involved in the contractile ring are multidomain (Mid1, Rng2, Cdc12, Cdc15) with Intrinsically Disordered Region (IDR) in between, making it difficult to determine full-length protein structure. Out of 53 total proteins involved in CAR as per Pombase (Wood *et al.*, 2012), only eight have experimentally determined structures deposited in the PDB, and only two are full-length, but no structures of protein complexes are available. Hence, to enrich our understanding of proteins and assemblies in the fission yeast contractile ring in terms of structural information, we decided to use recently developed structure prediction method, AlphaFold (Jumper *et al.*, 2021).

1.2 Actin and actin-binding proteins

Actin and myosin are two of the major proteins in eukaryotic cytokinesis. Both interact with a variety of proteins in CAR (Figure 1.1). Fission yeast forms diverse actin cytoskeletal assemblies (actin cables, patches, short filaments in the contractile ring) despite having only one isoform of actin (Act1) (Kovar *et al.*, 2011). Actin-binding protein (ABP)s regulate the actin dynamics inside the cell and control actin dynamics spatially. Since the first deposited actin structure in 1991 (1ATN), many actin structures (actin filaments, ABP-bound actin) have been added to the Protein Data Bank (PDB). We decided to use this information to understand how the interaction between actin and ABP correlates with the functions of ABPs and identify patterns in these interactions. For this purpose, we created a heatmap representation of actin-protein interfaces for all the actin structures and used agglomerative clustering methods to find the correlations.

1.2.1 Actin filament helicity

A single actin filament can either be considered to be made up of two protofilaments wound around each other to form a double helix or a single left-handed helix. If the centroids of the actin subunits are projected onto the symmetry axis of the filament, the per subunit **helical rise** can be defined as the distance between the consecutive projected points, and the **helical twist** can be defined as the angle between the vectors

formed by the centroids of actin subunits and their corresponding projections. A typical actin filament has a helical rise of 27 Å and a helical twist of 167° per subunit (Jegou and Romet-Lemonne, 2020). As of March 2024, 192 actin filament structures are deposited in the PDB (see section 3.3). Many of them are bound to one or more ABPs. We wanted to know if there is a correlation between the bound ABP and the helical parameters of these filaments. Hence, we calculated the helical rise and twist of all the actin filaments (resolution < 4 Å) and identified the outliers in the data to infer about the aforementioned correlation.

Analysis of actin filament helicity was carried out previously in the lab; however, it did not involve calculating the helical parameters (rise and twist) from the filament but relied on the annotated values of these parameters in the headers of the structure. These annotations are unavailable for some actin filaments; thus, only a subset (~65 out of 192) of actin filaments were part of the analysis. Furthermore, the analysis involved tedious manual collection and curation of the actin dataset from the PDB. In this project, I automated the data acquisition process and the calculation of helical parameters, making it easier to incorporate the new actin structures added to the PDB.

1.3 Protein structure prediction and modelling

There is a significant (more than 1000 fold) gap between the experimentally determined structures (214107 till 2023 as per PDB¹ and the known sequences of proteins (250322721 till January 2024 according to UniProtKB database². To help fill this gap, various computational structure prediction methods have been developed since the 1960s. Broadly, these methods can be classified into two categories - template-based modelling (TBM) and template-free modelling (TFM) (Wuyun *et al.*, 2024). Homology modelling methods such as SWISS-model (Bordoli *et al.*, 2009) and threading methods such as I-TASSER (Yang and Zhang, 2015) belong to TBM as both use information from the template structure for structure prediction. The former uses the template structure of a homologous protein, whereas the latter uses template structures for fold recognition for the query protein sequence. Ab initio

¹<https://www.rcsb.org/stats/>

²<https://www.uniprot.org/uniprotkb/statistics>

modelling methods belong to TFM as they do not use any template structure but are based on physical principles.

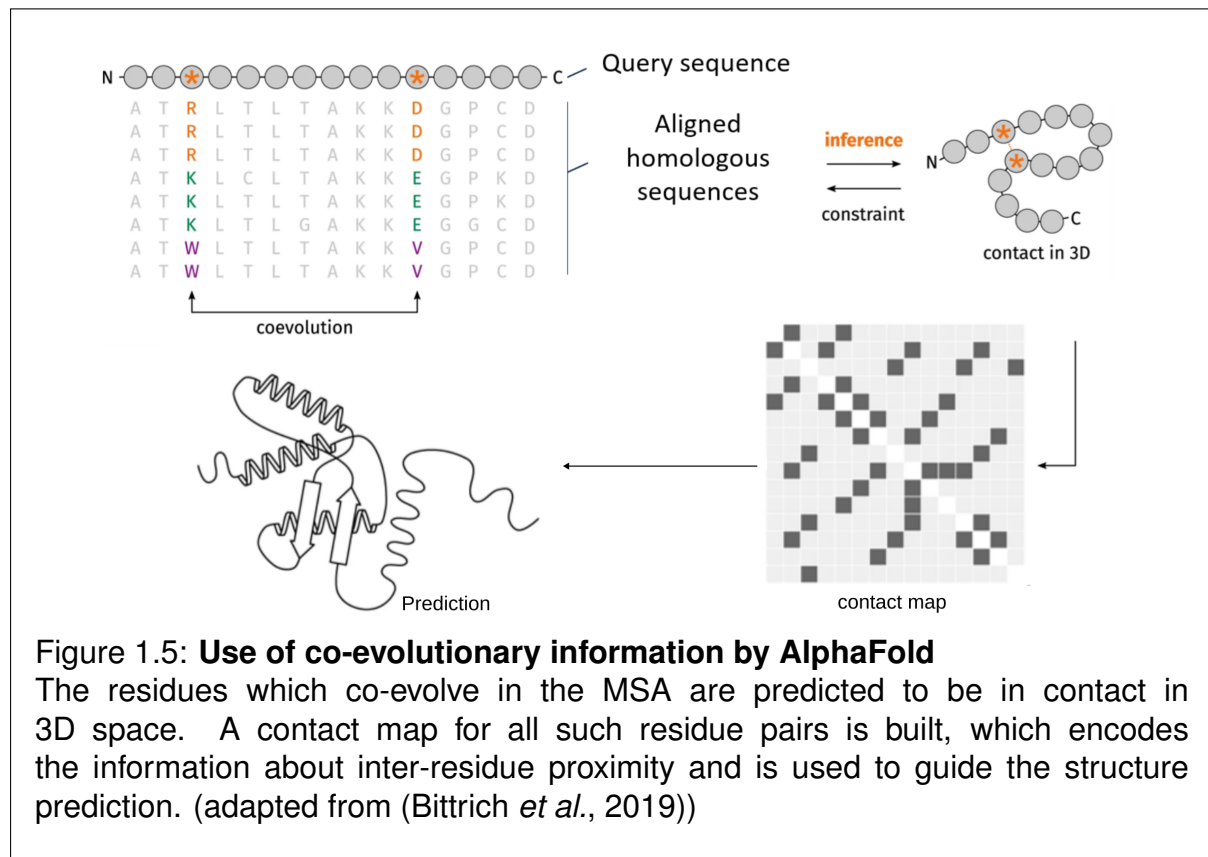


Figure 1.5: Use of co-evolutionary information by AlphaFold

The residues which co-evolve in the MSA are predicted to be in contact in 3D space. A contact map for all such residue pairs is built, which encodes the information about inter-residue proximity and is used to guide the structure prediction. (adapted from (Bittrich *et al.*, 2019))

In the past few years, new structure prediction methods have been developed (Baek *et al.*, 2021; Jumper *et al.*, 2021; Evans *et al.*, 2022; Lin *et al.*, 2023). The methods based on co-evolutionary information from the Multiple Sequence Alignment (MSA), such as AlphaFold2 (Jumper *et al.*, 2021) developed by DeepMind, rely on the observation that the residues that are closer together in the 3D space in the protein structure co-evolve (mutate together) during evolution (Figure 1.5). On the other hand, protein language model-based methods such as ESMfold (Lin *et al.*, 2023) developed by MetaAI use large-scale language models and do not rely on MSA.

1.3.1 AlphaFold

In CASP 14, AlphaFold2 (AF2) showed near experimental accuracy in predicting protein structure (Pereira *et al.*, 2021). As shown in Figure 1.5 and Figure 2.6, AF2 relies on co-evolutionary information extracted from the MSA. Figure 2.6 in the

methods section shows the internal architecture of AF2.

Due to computational limitations, AF2 clusters the sequences in MSA and takes cluster representatives (up to 1024) from each cluster in the MSA representation with additional unclustered sequences (up to 2048). It has been argued that the information about the alternate/ multiple conformations is encoded in the MSA and providing AF2 with a randomly subsampled (del Alamo *et al.*, 2022) or clustered MSA (Wayment-Steele *et al.*, 2024) or masking a part of the MSA (Stein and Mchaourab, 2022) would result in predictions corresponding to different conformations of the protein. We use this MSA ablation approach to predict multiple conformations of actin crosslinkers, Ain1 and Fim1, and analyse their flexibility.

Since its release, many other groups have used and modified the architecture to their advantage (Mirdita *et al.*, 2022; Li *et al.*, 2022; Wu *et al.*, 2022; Ahdriz *et al.*, 2023). Table 1.1 below shows some of the structure prediction methods based on AF2 architecture.

Table 1.1: **List of tools based on AlphaFold2 architecture**

Method/ Tool	Source publication	Description
AlphaFold	(Jumper et al., 2021)	Original AlphaFold notebook by DeepMind using jackhmmer for MSA generation
ColabFold	(Mirdita et al., 2022)	Modified AlphaFold notebook using mmseqs2 for MSA generation
CombFold	(Shor and Schneidman-Duhovny, 2024)	For predicting protein complex structures using a combinatorial approach from pairwise predictions
OpenFold	(Ahdriz et al., 2023)	AlphaFold reproduced using PyTorch and retrained from scratch
UniFold	(Li et al., 2022)	AlphaFold-Multimer reproduced using PyTorch

For this project, we have used the local version of ColabFold (Mirdita *et al.*, 2022) to predict the structures.

1.3.2 Integrative Structural Modelling

Biological systems involve macromolecular protein assemblies, and it is important to know their structure, preferably in the native state, and interprotein interactions to understand their function. None of the experimental methods stands triumphant over the others in determining the structure of macromolecular protein assemblies, especially of dynamic states. For X-ray crystallography, proteins need to be crystallised, which is a challenging task in the case of very large protein assemblies. On the other hand, electron microscopy, which is suitable for large assemblies, usually does not achieve atomic resolution structure for dynamic assemblies, with transiently interacting proteins. NMR spectroscopy, while helpful in studying an ensemble of structures in solution, is limited by the size of the protein. There are several other methods, such as crosslinking mass spectrometry (XL-MS), hydrogen-deuterium exchange (HDX), small-angle X-ray scattering (SAXS), and Förster resonance energy transfer (FRET), that provide valuable structural information about the protein assembly/ structure but cannot be stand-alone structure determination methods.

Integrative Structural Modelling (ISM) tries to solve this problem by integrating experimental information from multiple sources and using it to model the macromolecular assembly. Some notable ISM methods are - Integrative Modeling Platform (IMP) (Webb *et al.*, 2018), Assemblin (Rantos *et al.*, 2022), HADDOCK (Dominguez *et al.*, 2003), M3 (Karaca *et al.*, 2017), and PyRy3D³. There are four stages common to the ISM methods -

- Gather input information from various experimental methods (CryoEM, XL-MS, HDX, SAXS) and available structures from the PDB or predictions
- Decide model representation (atomic/coarse-grained/multiscale) and scoring function based on restraints (distance/volume/connectivity)
- Sample the conformational space and get good-scoring models
- Analysis and inference

³<https://genesilico.pl/pyry3d/>

1.4 Objectives

For computational structure prediction tools like AlphaFold, large protein/ complex predictions take a lot of time due to non-linear time complexity. However, as mentioned previously about the binding of Cdc4 to the IQ motifs of Rng2, Myo2, and Myp2, various studies have narrowed down the interprotein interactions in the CAR to a domain or range of residues or binding sites on the associated proteins (Almonacid *et al.*, 2011; Morita *et al.*, 2017; Snider *et al.*, 2020, 2022). We decided to use this information to map protein-protein interactions and use them to guide the protein complex structure prediction by delineating input sequences to speed up structure prediction. The complex predictions would then be used as input to ISM tools.

In essence, I use artificial intelligence (AI)-based protein structure prediction methods to get insights into the three-dimensional structural organisation of the protein assemblies in the fission yeast contractile ring, and I use heatmap visualisation along with agglomerative clustering to find the structure-function relationship of ABPs. Following were the broad objectives of the project.

- To map protein-protein interactions from the literature
- To use ColabFold to predict the structures of individual proteins (actin crosslinkers) and protein complexes (Table 3.1).
- To use the Integrative Modeling Platform (IMP) for making structural models of protein assemblies.
- To analyse of the predicted complexes to arrive at hypotheses on the mechanism of ring constriction, which have to be further validated through experimental approaches
- To create a dataset of actin-containing structures from the PDB and classify them into filament assemblies of actin and structures of globular actin
- To calculate the helical rise and twist of the actin filaments and find their correlation with the bound protein/ nucleotide/ small molecule
- To create a heatmap of interface residues, mapped onto actin, in actin-protein interactions and correlate the clusters of interactions with type of ABP

Chapter 2

Methods

This chapter is divided into four sections. The first section discusses the methods and tools used to gather information about actin structures in the PDB and build an actin dataset. The second section explains the steps used to calculate the helical parameters of actin filaments and detect outliers. The third section discusses visualising the actin interfaces using a heatmap and (row) similarity matrix. The last section describes strategies for generating and assessing structural models of proteins and protein complexes in fission yeast contractile ring using ColabFold.

2.1 Dataset of actin structures

Protein Data Bank (PDB) relies on various Application Programming Interfaces (APIs) on the back end to show the protein structure data on the website. We relied on these APIs to access this structure data programmatically. The following section discusses the APIs that were used.

2.1.1 Application Programming Interface (API)

Different APIs provided by RCSB PDB and PDB in Europe (PDBe) were used to programmatically access the protein structure and annotation data. Table (Table 2.1) denotes the APIs used.

The Search API was accessed from a Python interface provided by `rcsbsearchapi` (Rose *et al.*, 2021). It provides all the functionalities available through

Table 2.1: **List of APIs used for the actin analysis**

API	Source	Use
RCSB PDB Search API	https://search.rcsb.org/#search-api	Accessing PDB IDs based on the search query
RCSB PDB Data API	https://data.rcsb.org/#data-api	Accessing data for the IDs obtained from Search API
PDBe REST API	https://www.ebi.ac.uk/pdbe/api/	PDBePISA, ClustalO, SIFTS mappings
InterPro API	https://www.ebi.ac.uk/interpro/api/	Protein annotations

the advanced search on the RCSB PDB website¹. The Data API was also accessed from a Python interface, `python-graphql-client`, a Python package for making requests to GraphQL server. The Data API provides access to the annotations and the structure data on the structure summary page in the PDB. PDBe REST API² offers programmatic access to PDBe’s tools, services, and knowledge base. Out of these, we mainly used PDBePISA for obtaining interface information (Krissinel and Henrick, 2007), ClustalO (Madeira *et al.*, 2022) for MSA, and SIFTS mappings (Dana *et al.*, 2019) to convert the interface residues from the PDB numbering scheme to UniProt numbering. InterPro API (Paysan-Lafosse *et al.*, 2023) was also used to access the protein annotations. The advantage of using APIs is that they provide a way to automate the data collection process; so as the new actin structures get deposited in the PDB, they can be automatically added to the dataset and further analysed using the same algorithms for the updated database.

2.1.2 Methodology of actin dataset generation

RCSB PDB Search API was used to find all the PDB structures containing actin as one of the entities. The following query was used as input to the Search API.

¹<https://www.rcsb.org/search/advanced>

²<https://www.ebi.ac.uk/pdbe/pdbe-rest-api>

- Query sequence: Amino acid sequence of human cytoplasmic actin (UniProt-P60709)
- Sequence similarity cutoff: 54% (The sequence similarity cutoff of 54% was decided by trial and error. A cutoff of less than 54% resulted in actin-related proteins (ARPs) getting picked up in the structures rather than actin.)
- Initial release date: structures before 01/03/2024

For the subsequent analysis, we needed to access the actin structures from different levels of the hierarchy of the structures (entity/assembly/entry as shown in Figure 2.1); hence, using the above query, three lists were obtained corresponding to the entity IDs, assembly IDs and entry IDs (for more information about structure hierarchy in the PDB, please visit - Structure organization in PDB³).

To segregate actin filaments from these structures, we used the following two criteria:

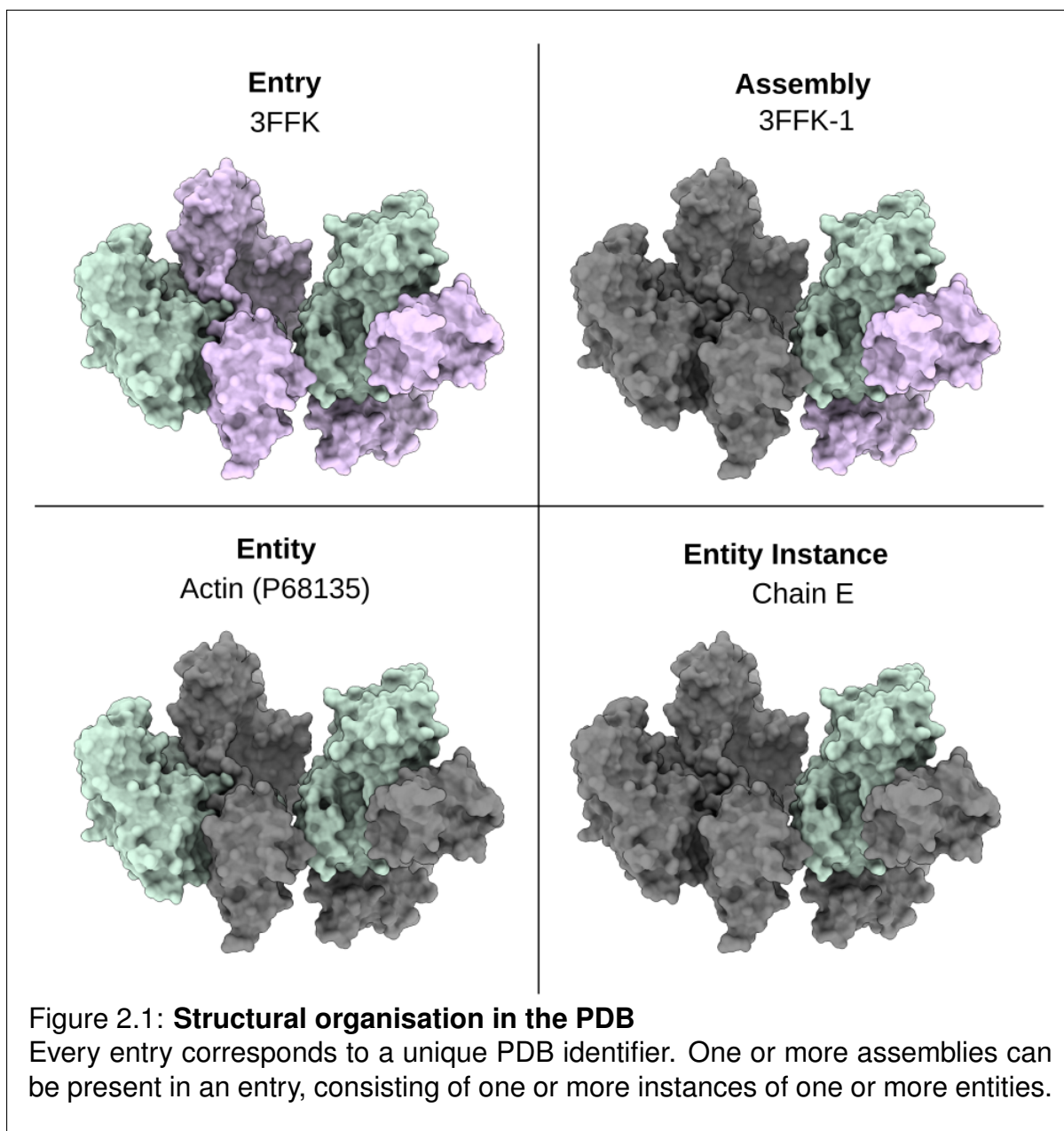
- (1) The structure should have assembly symmetry type = “**Helical**”
- (2) The number of actin chains that formed the assembly ≥ 3

The information about assembly symmetry type and number of actin chains in the structure was collected at the assembly level with assembly IDs as input to the Data API, followed by manual verification. The final dataset of actin filaments was confirmed after removing the following structures during the verification (Table 2.2).

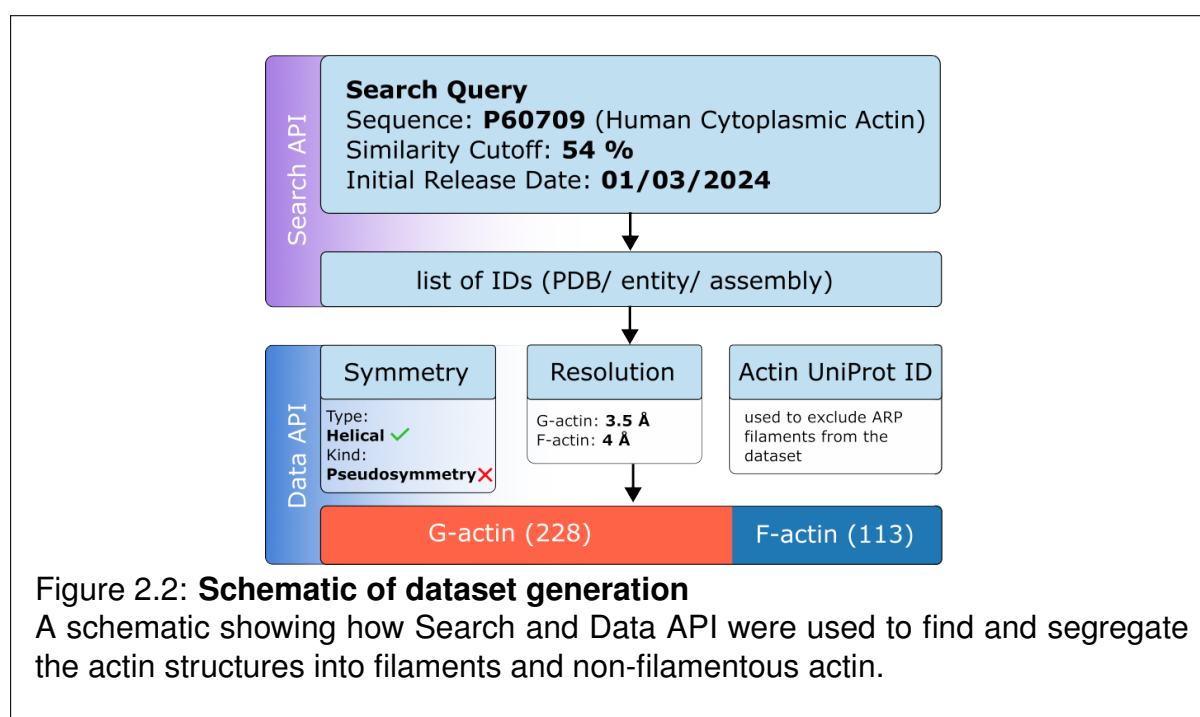
Table 2.2: **List of actin filament structures removed from the filament dataset**

PDB ID	Reason for removal
3TPQ	Five G-actins are bound to RPEL domains, resulting in helical symmetry and not the filament arrangement of actin subunits
2YJF	Three G-actins are bound to RPEL domains, resulting in helical symmetry and not the filament arrangement of actin subunits
4M63	The arrangement of actin chains is due to their binding to the protein VopL, which, according to the authors, suggests the actin arrangement during nucleation.

³<https://www.rcsb.org/docs/general-help/organization-of-3d-structures-in-the-protein-data-bank>



Using Data API, we gathered additional information about the actin structures, such as UniProt accessions of actin entities and their interactor proteins, combined resolution of the structures, chain IDs of actin entity instances, type of assembly in the structure, name and PDB identifiers of small molecules bound to actin. This information was used in the subsequent analysis. From the dataset, the actin entity in the structure 4PL7 is a chimeric protein of human thymosin beta-4 (P62328) and *Komagataella* actin (Q9P4D1). We only used the sequence fragment corresponding to the latter (Q9P4D1). Finally, the resolution cutoff of 4 Å was applied for the actin filaments and 3.5 Å for the rest of the actin structures (Figure 2.2).



The actin filament in the structure is defined at an assembly level. So, assemblies of all the actin filament structures were downloaded from the PDB to be used to calculate helical parameters.

2.2 Calculation of helical parameters of actin filaments and outlier detection

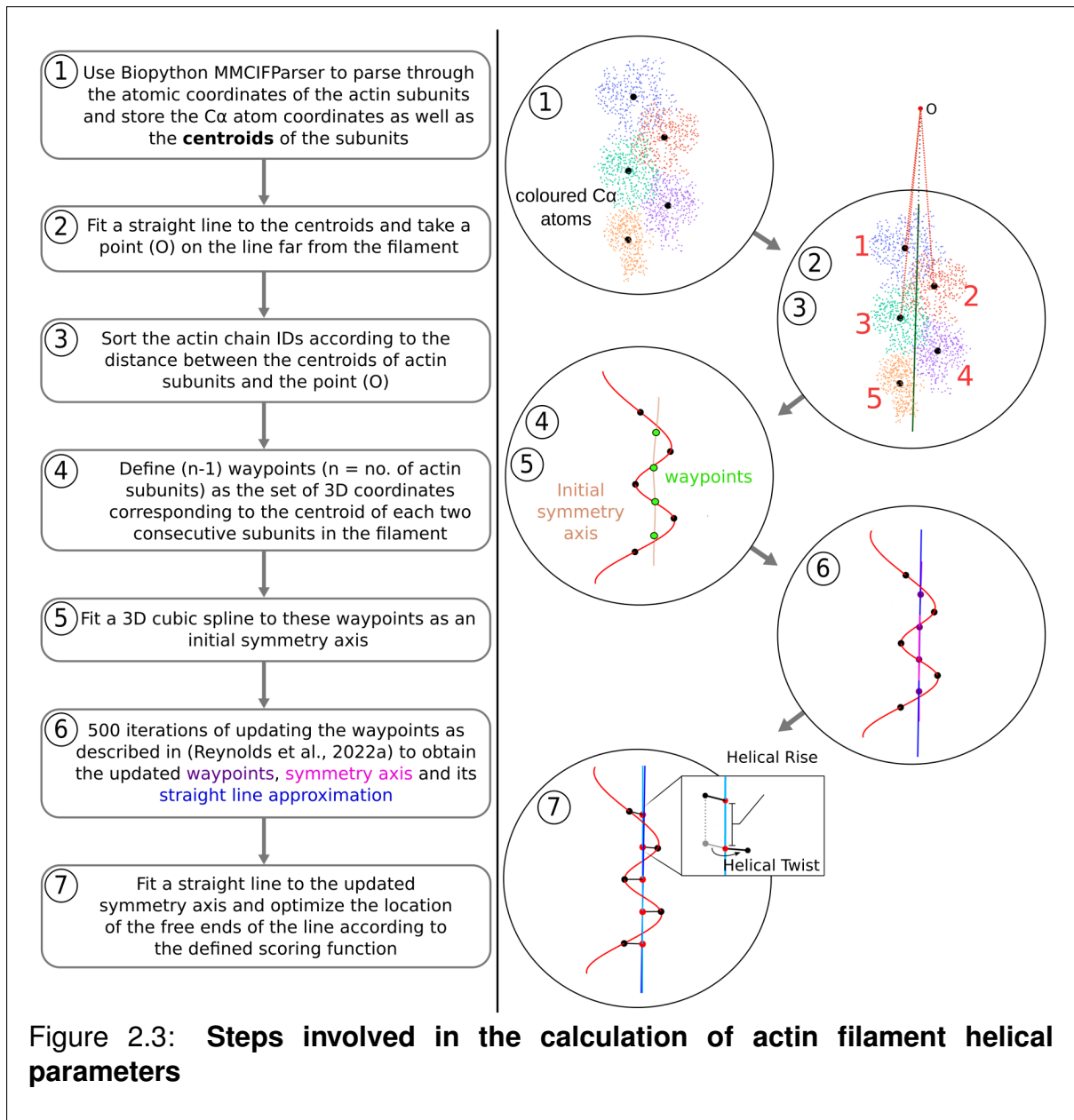
We used a custom Python script based on the method described in Reynolds *et al.* (2022) to calculate the helical parameters - rise and twist of the actin filaments.

However, the filament structures need curation, and the following steps were followed:

- Some of the assemblies (8UXX, 8UXW, 8E9B, 7AQK, 8P94, 7SX8, 7SX9, and 7TPT) contained more than one actin filament; such structures were modified to include only one actin filament.
- The calculation of helical parameters depends on the initial identification of the symmetry axis by fitting a straight line to the centroids of the actin subunits (step 2 in Figure 2.3). However, for the filaments with three actin subunits, the fitted line approximates the symmetry axis so poorly that the subsequent calculations fail to optimise the position of the symmetry axis. Hence, we extended the filaments to contain four actin subunits based on superposition as described by Reynolds *et al.* (2022).

ChimeraX (Meng *et al.*, 2023) was used for all the modifications.

To calculate the rise and twist values, we needed to know the order in which the actin subunits are arranged along the long axis of the actin filament. We also needed an accurate prediction of the symmetry axis and the projections of the centroids of the actin subunits onto this axis. Once all this information was obtained, we could calculate the helical rise by measuring the distance between the consecutive projections and the helical twist by measuring the angle between the two vectors formed by the centroids and their respective projections (step 7 in Figure 2.3).



To accurately predict the symmetry axis, Reynolds *et al.* (2022) developed their own method based on 3D spline fitting and its optimisation. The method works well if the filament length is more than 30 protomers, but in the case of smaller actin filaments, the edge effects hamper the correct calculation of the symmetry axis and, in turn, the helical parameters. Reynolds *et al.* (2022) solved this problem by extending the short filaments by superposition, but, in our case, it is difficult to do so systematically as the number of subunits and their chain IDs vary and the high computational cost due to large number of structures. Moreover, it would not be possible to identify the local aperiodic variations in the helical rise and twist with this method. So, we used a custom

scoring function on top of the spline fitting and optimisation instead of extending the filament length. With this function, we tried to minimise the edge effects by optimising the location of the ends of the symmetry axis such that the variation in all the rise and twist values and the distance of actin subunits from the symmetry axis lies close to zero. However, since a straight line is fit to approximate the symmetry axis of the filament, the underlying assumption is that the filament is straight. So, the method would not give accurate results in the case of bent actin filaments; however, it was able to capture them as outliers.

The flowchart in Figure 2.3 shows the steps for calculating the actin filament's helical rise and twist.

To detect the outliers in terms of helical rise and twist of the filaments, we used the Inter-quartile range (IQR). For both the parameters, a lower and upper bound was set as follows:

$$\text{Lower bound} = Q_1 - 1.5 \times IQR \quad (2.1)$$

$$\text{Upper bound} = Q_3 + 1.5 \times IQR \quad (2.2)$$

Where,

Q_1 = First quartile (25% of the data lies below Q_1)

Q_3 = Third quartile (75% of the data lies below Q_3)

IQR = Inter-quartile range ($IQR = Q_3 - Q_1$)

The structure was considered an outlier if it was outside the upper or lower bounds of helical rise or twist. Figure 3.15 shows the corresponding bounds as a rectangular shaded region. We used the Z-score metric to determine the outliers in terms of local variations in either twist or rise. The Z-score was calculated for the standard deviations in the helical rise and twist of each filament as follows.

$$Z - score_i = \frac{|x_i - \bar{x}|}{\sigma} \quad (2.3)$$

Where,

- x_i = standard deviation in the calculation of helical rise or twist for i^{th} structure
- $\underline{x}$ = mean of standard deviations in helical rise or twist
- σ = standard deviation of standard deviations in helical rise or twist

The filaments having a Z-score ≥ 3 were considered to be outliers.

All the analyses are hosted as a Jupyter Notebooks on Google Colaboratory.

2.3 Interface matrix generation and visualisation

We decided to use a heatmap and similarity matrix combined with agglomerative clustering to represent and analyse the actin-protein interaction. The inspiration is taken from the heatmap visualisation of gene expression data (Grant *et al.*, 2007). In our case, we wanted to portray the actin sequence on the horizontal axis and unique actin-protein interactions on the vertical axis. Since the actin dataset involved multiple actin entities corresponding to different sequences, we used the consensus sequence obtained from the MSA on the horizontal axis (Figure 3.20). The subsequent agglomerative clustering step clustered similar actin-protein interfaces together. To correlate the structural similarities of the interfaces with the functions of the actin-binding proteins, we used domain/family/superfamily annotations of the interacting proteins from InterPro (Paysan-Lafosse *et al.*, 2023).

We used PDBePISA API to access the information on the actin-protein interface. The instance-level interface information is available from the assembly entry points and was downloaded from the API using the assembly IDs from the actin dataset. Each interface can be identified by the interacting proteins' asymmetric IDs (chain IDs). From all the interfaces from an assembly, we separated the actin-protein interface by checking whether the interface had at least one of the chains as actin. The actin chain IDs were obtained from the actin dataset. The interface information contains the information about buried surface area at each residue and the residues having non-zero buried surface area are considered as interface residues (for details, please visit - PISA FAQ⁴). In this manner, we obtained a dictionary of actin-protein interfaces in which the key represented the chain ID of actin and its interacting partner, along with

⁴https://www.ebi.ac.uk/pdbe/pisa/pi_tips.html

the assembly ID. The value represented a list of interface residues of actin according to the PDB numbering scheme. To only take the unique actin-protein interactions from each assembly, a similarity threshold of 0.8 was used to remove the redundancy (Figure 2.4).

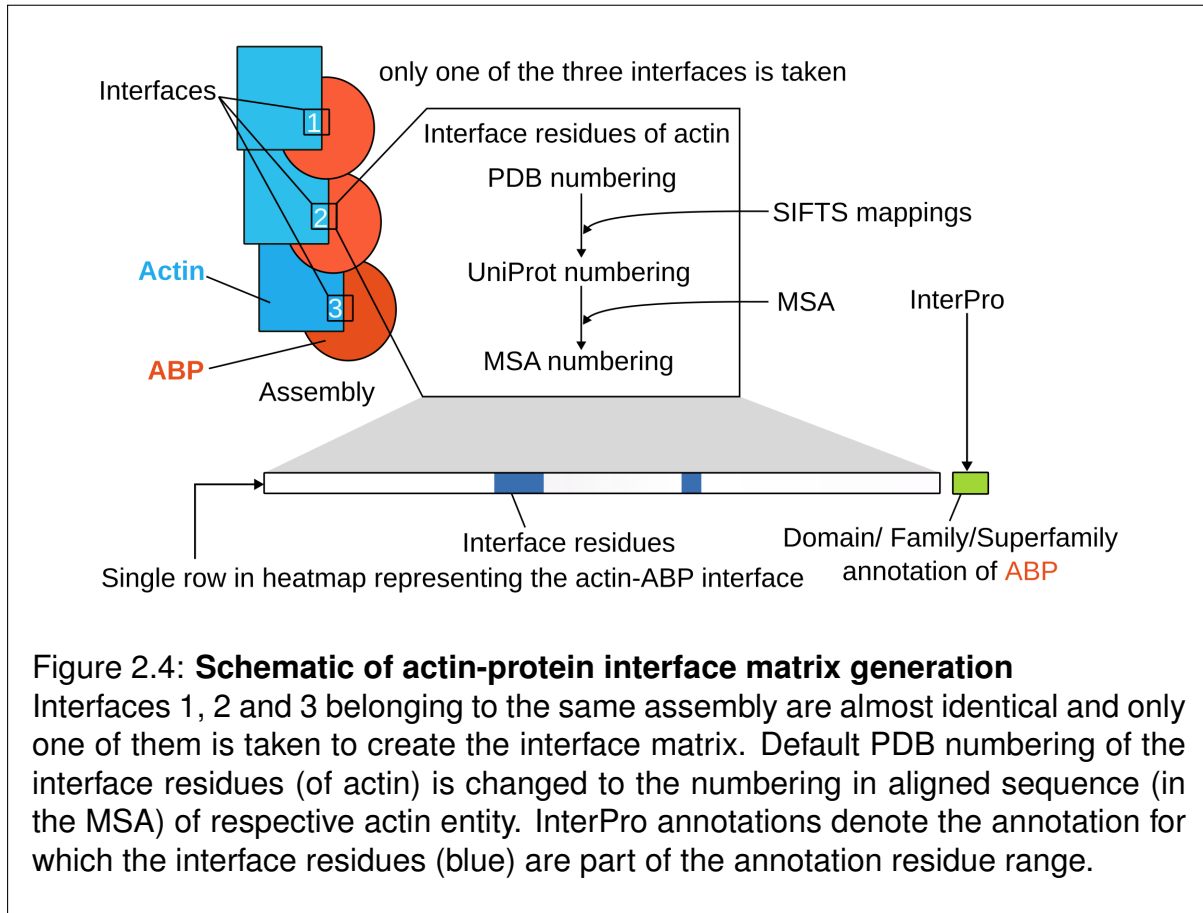


Figure 2.4: **Schematic of actin-protein interface matrix generation**

Interfaces 1, 2 and 3 belonging to the same assembly are almost identical and only one of them is taken to create the interface matrix. Default PDB numbering of the interface residues (of actin) is changed to the numbering in aligned sequence (in the MSA) of respective actin entity. InterPro annotations denote the annotation for which the interface residues (blue) are part of the annotation residue range.

The assemblies in a PDB structure can be author-defined or software-defined. In the crystal structures, often, the software-defined assemblies correspond to the crystal contacts rather than the physiologically relevant protein assembly. Hence, we removed the interfaces from the dictionary that belonged to the software-defined assemblies. We only used software-defined assemblies if it is the only one present for that particular structure.

Next, we added the UniProt IDs of actin and its interacting partner to the dictionary keys using PDB Data API. Based on them, the interface residue lists in the dictionary are converted from the PDB numbering scheme to UniProt numbering using SIFTS mapping (Figure 2.4).

To get the consensus sequence representing all the actin entities present in the

dataset, we performed Multiple Sequence Alignment (MSA) using ClustalO API. We got an aligned sequence of each actin entity from the MSA, which was used to update the UniProt numbering of interface residues with the MSA numbering (Figure 2.4). The final dictionary was then used to create an interface matrix where rows correspond to unique actin-protein interfaces from all the assemblies, and columns correspond to the residue number of the respective actin (mapped on the aligned sequence). Cells of the matrix take values 1 or 0, depending on whether the residue belongs to the interface residue list in the dictionary or not (interface matrix visualised as a heatmap in Figure 2.5).

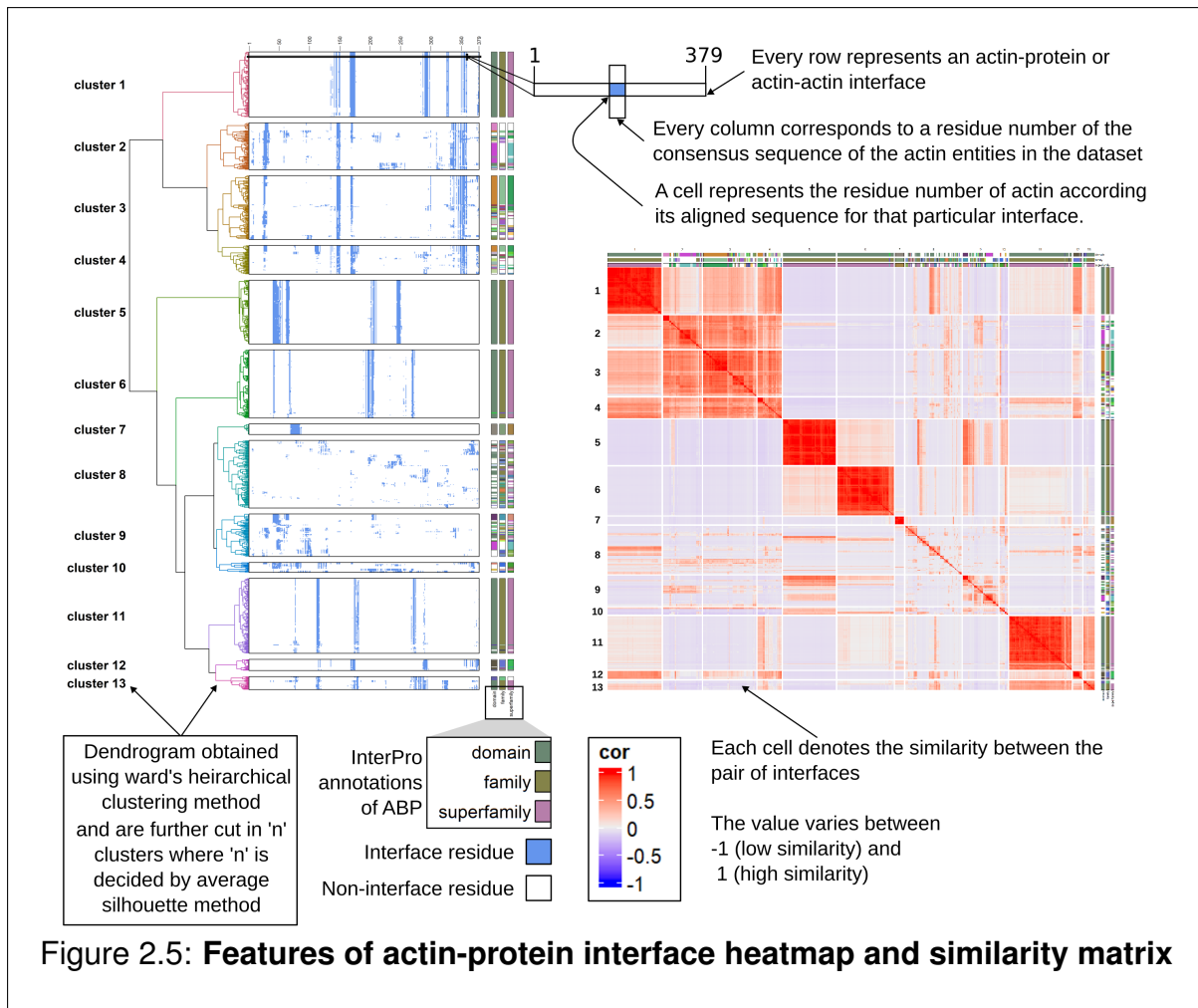
We used InterPro API (Paysan-Lafosse *et al.*, 2023) to download the protein annotations (domain, family, superfamily) for all the actin-binding proteins in the dataset. Only those annotations were kept for which at least one interface residue belonged to the residue range of the annotation.

2.3.1 Heatmap and similarity matrix visualisation of actin-protein interfaces

We used `ComplexHeatmap` and `InteractiveComplexHeatmap` packages by Gu (2022) to visualise the interface matrix from the previous section. Figure 2.5 describes the main features of the heatmap and the row similarity matrix, respectively. The dendrogram in Figure 2.5 is created using "Ward's hierarchical (agglomerative) clustering" method (Murtagh and Legendre, 2014) and is split into 13 clusters. The optimal number of clusters was decided using the "average silhouette method" in the `fviz_nbclust` function⁵ in R for all the heatmaps. The architecture and main features of the heatmap and row similarity matrix are shown in Figure 2.5.

Sequence logo of consensus actin sequence (Figure 3.20) obtained from the actin entities in the dataset was generated using a web server WebLogo (Crooks *et al.*, 2004).

⁵https://search.r-project.org/CRAN/refmans/factoextra/html/fviz_nbclust.html



2.4 Computational structure prediction

ColabFold was downloaded to the local system using the LocalColabFold installer script and used for the structure prediction of individual proteins and protein complexes in the fission yeast contractile ring. `alphafold2_ptm` and `alphafold2_multimer_v3` were used from the ColabFold for monomer and complex prediction, respectively, whereas the MSA was generated by the MMseqs2 web server as incorporated in the ColabFold. We chose the local version of ColabFold over the original AlphaFold by DeepMind for the following reasons:

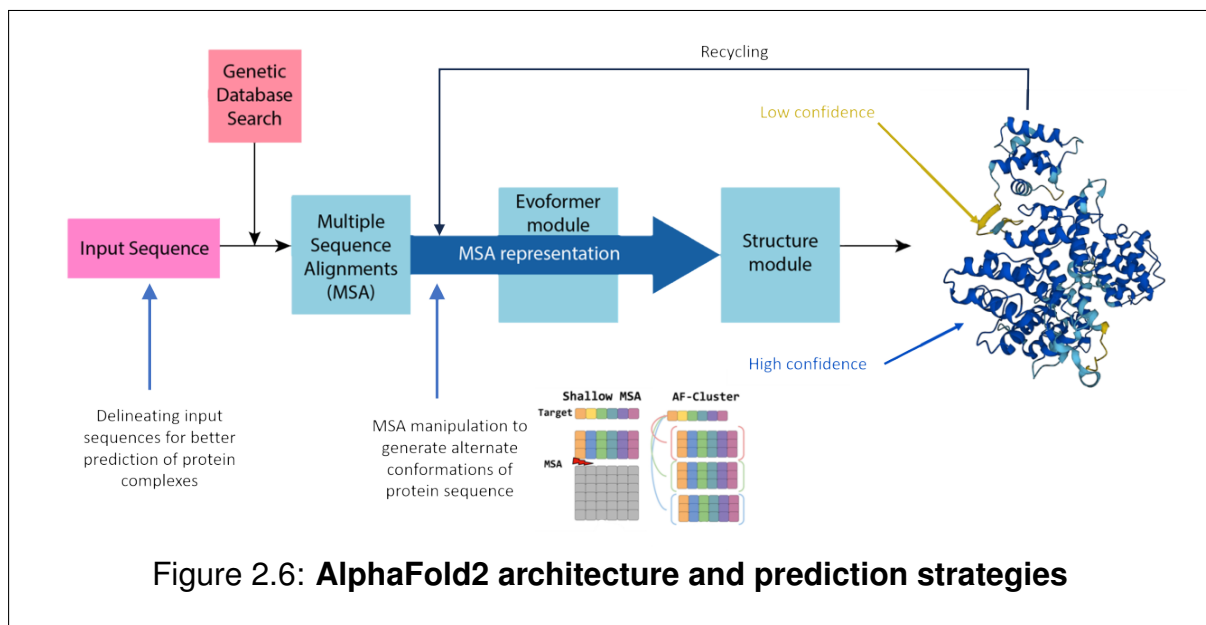
- The MSA generation step takes a significantly longer time in AlphaFold (DeepMind)
- Run-time errors on Google Colaboratory for large proteins

- For the structure prediction pipeline to run locally, the large sequence databases required for MSA generation need to be downloaded on the local system

2.4.1 Prediction of multiple conformations

To predict the actin-crosslinkers, Ain1 and Fim1 in multiple conformations, we used two different strategies as follows:

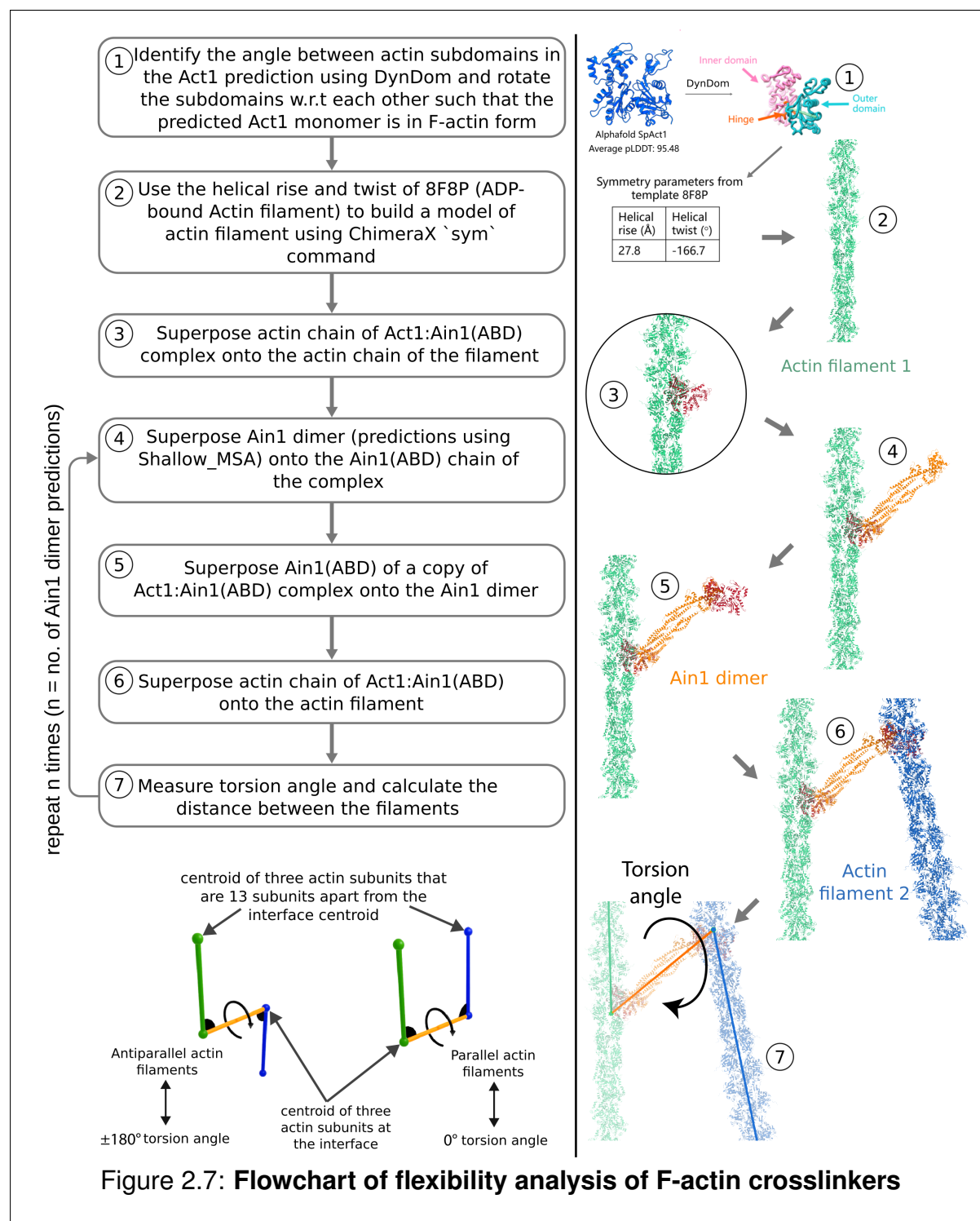
- MSA subsampling: In this strategy by del Alamo *et al.* (2022), only a subset of the generated MSA is given as an input. This is done by varying the `max_msa` parameter in the ColabFold. (Shallow_MSA in Figure 2.6)
- MSA clustering: In this strategy, the generated MSA is clustered, and the clusters are provided as input to the AlphaFold. This is done using the script provided by Wayment-Steele *et al.* (2024). (AF-Cluster in Figure 2.6)



The rationale behind both methods is that a sufficiently deep MSA encodes all the information required to predict the structures of alternate conformations of the protein, and the subsets of this MSA (randomly subsampled or clustered) correspond to distinct co-evolutionary signals being provided as an input that encode distinct conformations.

2.4.2 Flexibility analysis of F-actin crosslinkers

ChimeraX was used along with custom Python scripts to measure the torsion angles and distance between the crosslinked actin filaments. The following flowchart (Figure 2.7) shows the steps followed.



2.4.3 Prediction of protein complexes by delineating the input sequences

In the case of large proteins, the prediction of protein complexes becomes a difficult task due to high computational cost. Also, when provided with full-length sequences of the interacting proteins, AlphaFold-Multimer has a 40% success rate in predicting the correct interface structure. However, Bret *et al.* have recently shown that by delineating the input sequences, the success rate increases by up to 90%. We incorporate this strategy to make complex structure predictions.

The brute force way would be to delineate all the amino acid sequences of the proteins in the CAR into fragments of 100-500 amino acid length and perform pair-wise structure prediction for all of them. But, this is computationally heavy. So, we decided to use the information already available about the inter-protein interactions from the literature to guide the complex structure prediction (Figure 2.6).

2.4.4 Assessment of predictions

The predictions were assessed using two metrics provided by the AlphaFold - pLDDT and PAE. The predicted Local Distance Difference Test (pLDDT) metric⁶ assesses the local accuracy in the structure, and its value ranges from 0 (low accuracy) to 100 (high accuracy). The following table indicates the confidence in the predicted model with respect to the pLDDT score.

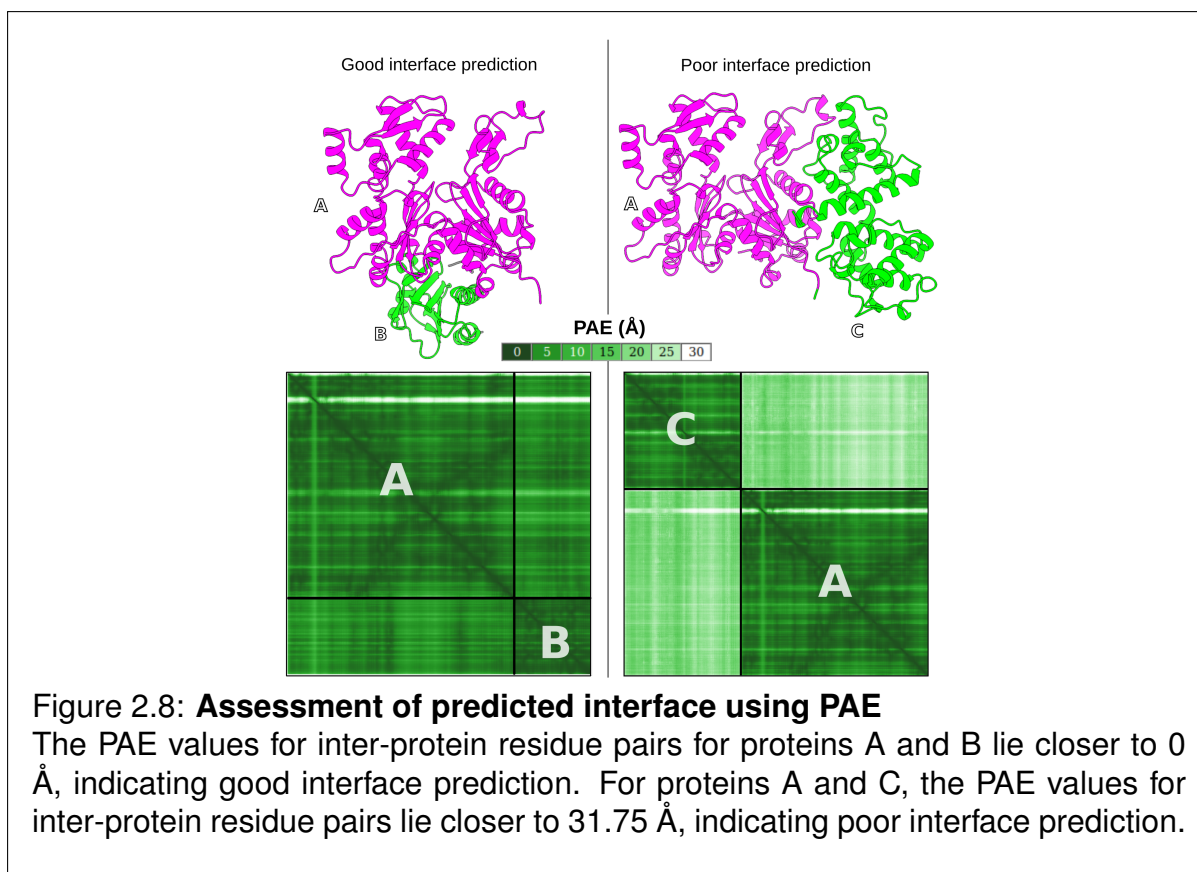
Table 2.3: **Model confidence with varying pLDDT score**

pLDDT score	Confidence
>90	Very high accuracy (both backbone and side chains are accurately predicted)
70 to 90	Good backbone prediction
50 to 70	Low confidence (should be treated with caution)
<50	Usually, a predictor of disorder in the structure

Predicted Aligned Error (PAE) assesses the variability in the pair-wise distance for all the residue pairs (Figure 2.8). Its value ranges from 0 to 31.75 Å.

Protein complex predictions were assessed using `alphafold interfaces` function

⁶<https://alphafold.ebi.ac.uk/faq>



in ChimeraX Daily Build (date: 21/03/2024) with minor modifications to the function's script. The function assesses the complex prediction based on the inter-chain PAE values. For most of the predictions, we set the number of recycles to low (`num_recycle = 2` in ColabFold), hence we set the `maxPae` cutoff to 10 Å and `distance = 8 Å` with `minConfPairs = 5`. As a result, a reliable (good) complex prediction would have at least 5 pairs of inter-chain residue-residue contacts in which any of the atoms in those residue pairs would be less than or equal to 8 Å apart in space and have $\text{PAE} \leq 10 \text{ Å}$. The interface residues⁷ are visualised on the structures using `alphafold contacts` function by pseudobonds on the structures coloured according to the PAE values (Figure 2.9)

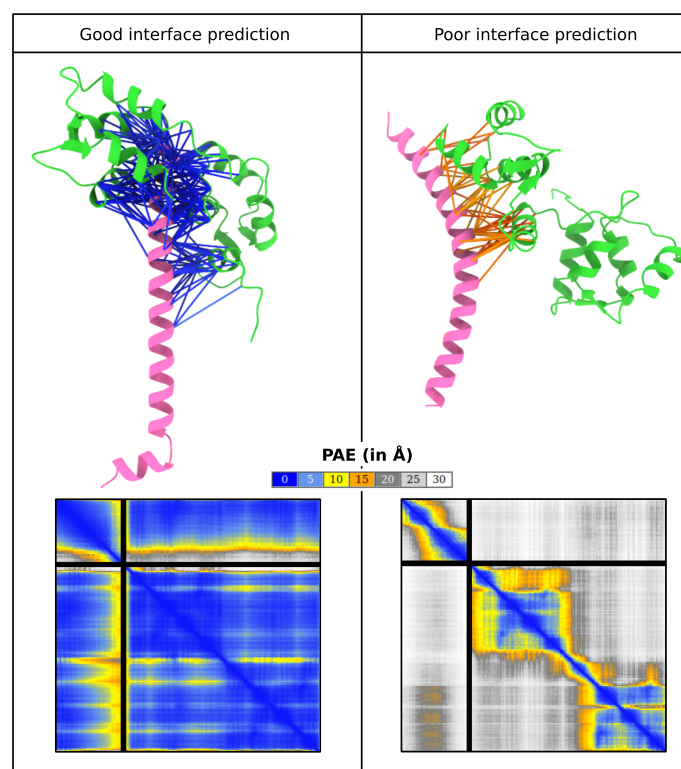


Figure 2.9: Visualisation of predicted interface

The residue pairs which are 8 Å apart connected by the pseudobonds coloured according to PAE metric. The left panel has almost all the pseudobonds with PAE near 0 Å indicating good interface prediction whereas in the right panel, the pseudobonds have PAE ≥ 15 Å indicating poor prediction. The same can be interpreted by the inter-chain regions in the PAE map as discussed in Figure 2.8

2.4.5 Map of protein-protein interactions

We used xiVIEW (Graham *et al.*, 2019), a web-based visualisation tool to visualise the predicted inter-protein interactions. We viewed the interactions as a 2D map as well as a circular chord diagram as implemented in xiVIEW. As seen in Figure 3.7 B and C, the nodes indicate the proteins whereas the edges link the interface residues⁷. The green edges link the residues in two different proteins whereas the purple and orange edges link the residues within the same protein (an indication of homodimer).

⁷The interface residues are defined as per `alphafold interfaces` function in ChimeraX i.e. residues for which any of their atoms is at most 8 Å apart in space with PAE ≤ 10 Å

2.5 Resource Table

Table 2.4: Resource list of software and packages used

Type	Name	Version
Software	ChimeraX	1.6.1, 1.7.1, 1.8.1 (daily build)
Software	Inkscape	1.3.2
Software	ColabFold	1.5.2
Software	R Studio	4.3.2
R package	Complex Heatmap	2.15.4
R package	Interactive Complex Heatmap	1.9.2
Python package	Plotly	5.19.0
Python package	Matplotlib	3.8.3
Service	Google Colaboratory	-

Chapter 3

Results and Discussion

This chapter is divided into three parts. The first two parts discuss the computational structure modelling of fission yeast CAR components. In particular, part 1 shows the results of strategies used for generating multiple conformations - Shallow_MSA and AF_cluster (for details, see section 2.4.1) and part 2 discusses the structure prediction of protein complexes in fission yeast CAR. The third part explores the dataset of actin structures from the PDB to correlate actin filament helicity and interesting patterns in the protein-binding interface of actin to the functions of actin-binding proteins.

3.1 Analysing flexibility of actin crosslinking proteins in fission yeast

The two F-actin crosslinkers, α -actinin (Ain1) and fimbrin (Fim1), are proposed to mediate the binding of Myo2 heads to the actin filament by aligning the actin filaments through their dynamic crosslinking activity (Laporte *et al.*, 2012). We used MSA subsampling strategies to predict the crosslinkers in multiple conformations and analyse their flexibility.

3.1.1 Multiple conformations of Ain1

Ain1 monomer has only one actin-binding domain (ABD) (Figure 3.1 A); it crosslinks actin filaments only in its dimeric form where the spectrin repeats (SR) form the dimerisation interface. So, we predicted its structure in the dimeric form (Figure 3.1 B

and C). The average pLDDT score of 88.4 and the ipTM score of 0.661, along with the low PAE values in the inter-chain region, indicate a good prediction of Ain1 dimer.

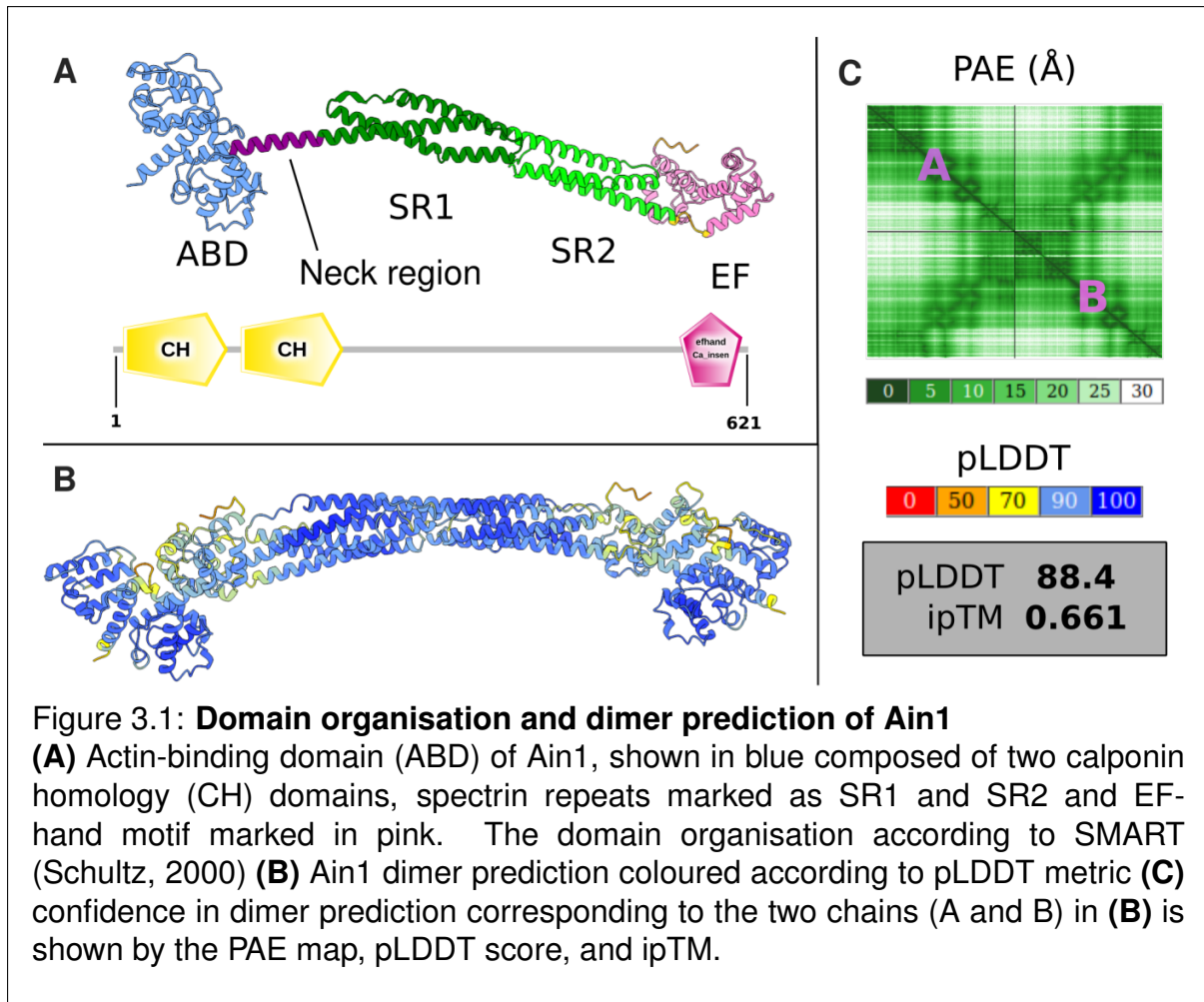


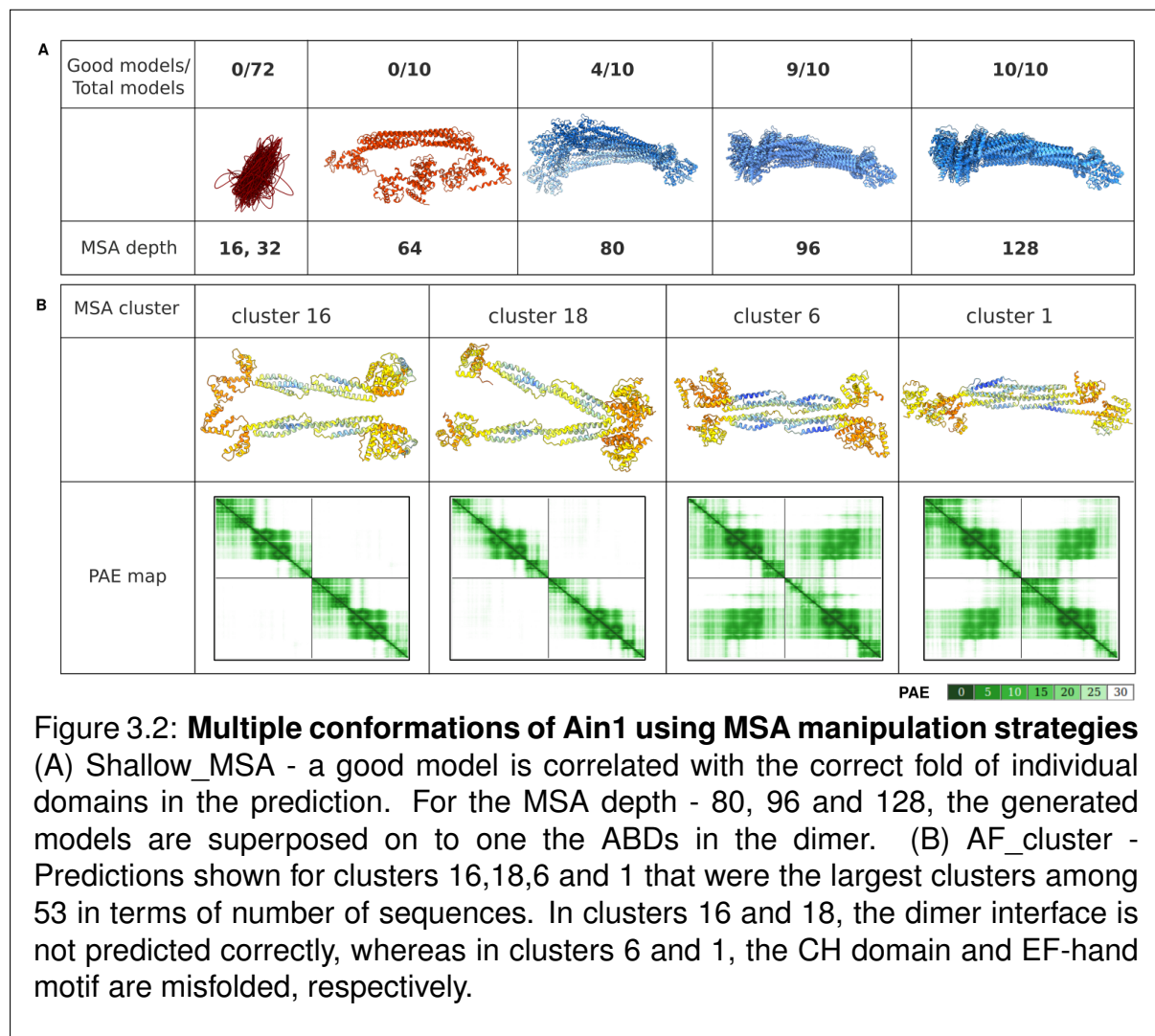
Figure 3.1: **Domain organisation and dimer prediction of Ain1**

(A) Actin-binding domain (ABD) of Ain1, shown in blue composed of two calponin homology (CH) domains, spectrin repeats marked as SR1 and SR2 and EF-hand motif marked in pink. The domain organisation according to SMART (Schultz, 2000) **(B)** Ain1 dimer prediction coloured according to pLDDT metric **(C)** confidence in dimer prediction corresponding to the two chains (A and B) in **(B)** is shown by the PAE map, pLDDT score, and ipTM.

Based on the experimental structures of *Entamoeba* and human α -actinin 2 (6SL3 and 4D1E, respectively), the proposed models of actin-crosslinking suggest that the flexibility in actinin arises from the flexible neck region connecting the ABD and SRs modulated by Ca^{2+} binding to the EF-hand motif (Ribeiro *et al.*, 2014; Pinotsis *et al.*, 2020). Although the EF-hand motif of Ain1 is Ca^{2+} insensitive, the non-binding of Ca^{2+} to the EF-hand in *Entamoeba* α -actinin is supposed to increase its flexibility. So, we used two MSA manipulation methods (see section 2.4.1), Shallow_MSA and AF_cluster, to generate multiple conformations of fission yeast α -actinin to see if the flexibility is indeed observed in the neck region.

While using Shallow_MSA, we varied the depth of MSA, starting from as low as 16. As seen in Figure 3.2, for MSA depths 16, 32 and 64, all the models were extremely

misfolded. The seemingly correctly folded structures (based on the preserved local structure of individual domains) were predicted when the depth of MSA was increased to 80. Upon further increase in MSA depth, all the predicted structures were folded correctly. Contrary to our expectations, no obvious conformational change in the neck region was observed. However, the accumulation of small-scale structural changes led to an overall heterogeneity in the predictions, as seen in Figure 3.2 A (MSA depth = 80).

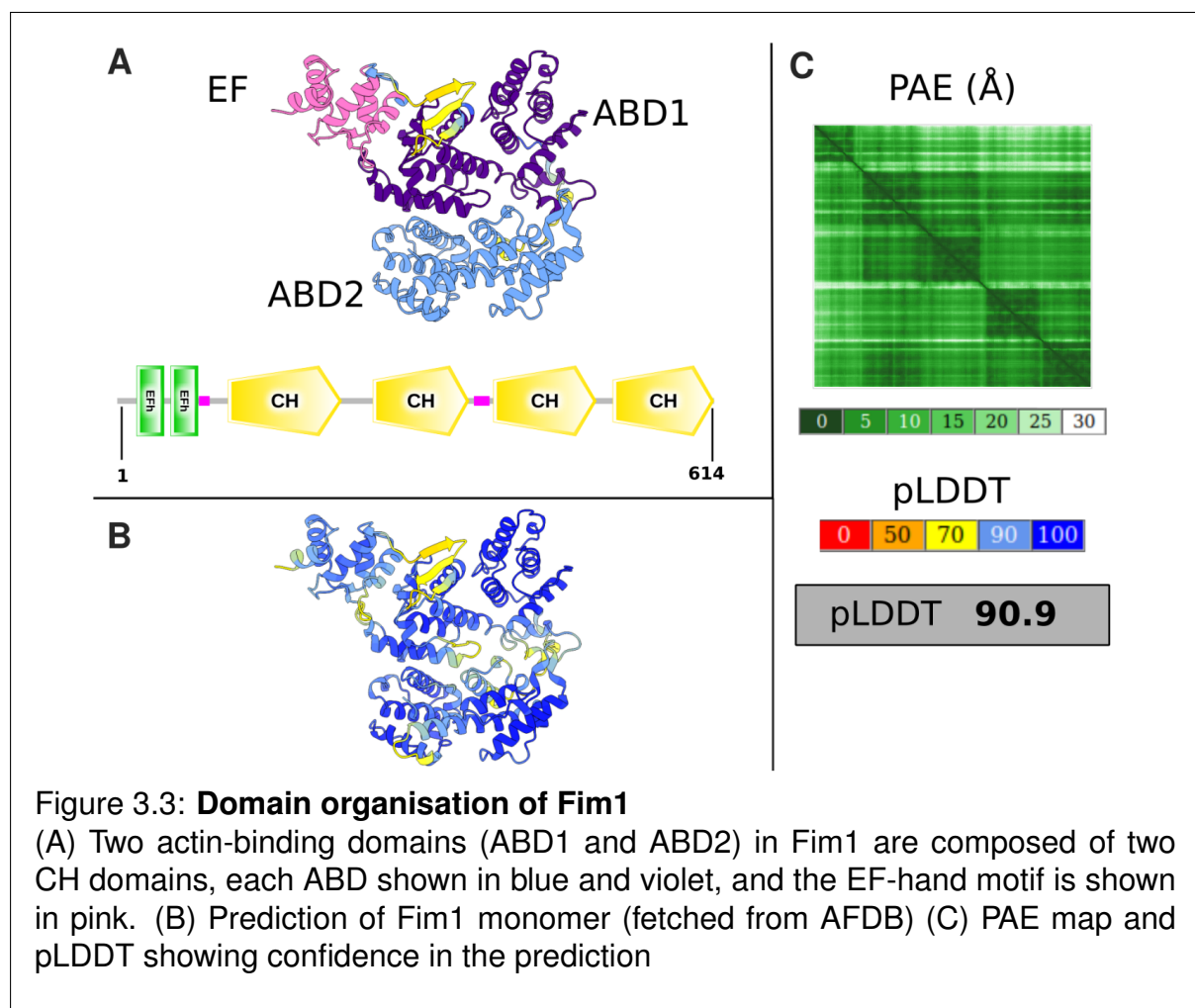


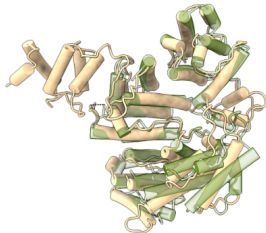
In the second method, AF_cluster, we obtained 53 clusters from the MSA sequences, out of which we used the top four clusters in terms of the number of sequences to predict the Ain1 dimer. Two of the predictions (clusters 1 and 6) showed a good prediction of the dimerisation interface at the spectrin repeats; however, the EF-hand domain was misfolded in cluster 6, whereas the ABD was misfolded in

cluster 1 (Figure 3.2 B). We proceeded with the models generated using method 1 for further analysis.

3.1.2 Multiple conformations of Fim1

Unlike Ain1, Fim1 has two ABDs (Figure 3.3) and can crosslink actin filaments as a monomer. The predicted Fim1 structure superposed well with the experimentally determined structure (1RT8) . However, Fim1 can not crosslink actin filaments in this compact state as the actin-binding site (ABS) on ABD1 is occluded. Mei *et al.* (2022) have determined the structure of T-plastin (Fim1 homolog in *H. sapiens*) in the crosslinked (parallel and antiparallel actin filaments) and pre-bundled (bound to single actin filament) state. We hypothesised that the Fim1 might undergo similar conformational changes. We used the same MSA manipulation strategies to generate multiple conformations of Fim1.



A	Fim1 (O59945)		CH1	RMSD	CH2	RMSD
				8.942 0.837		2.558 0.533
			CH3		CH4	
				10.932 0.707		17.581 1.083
	Fim1 (rank 1 model for MSA depth = 16)		Fim1 (rank 10 model for MSA depth = 16)			

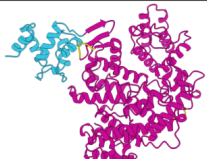
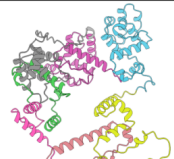
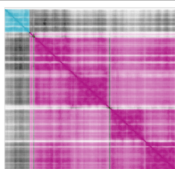
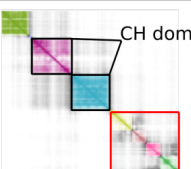
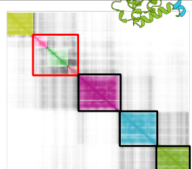
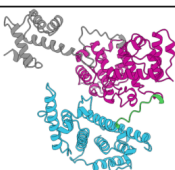
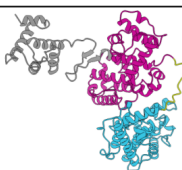
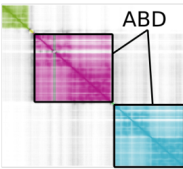
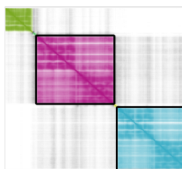
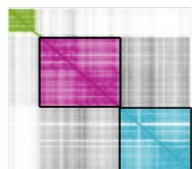
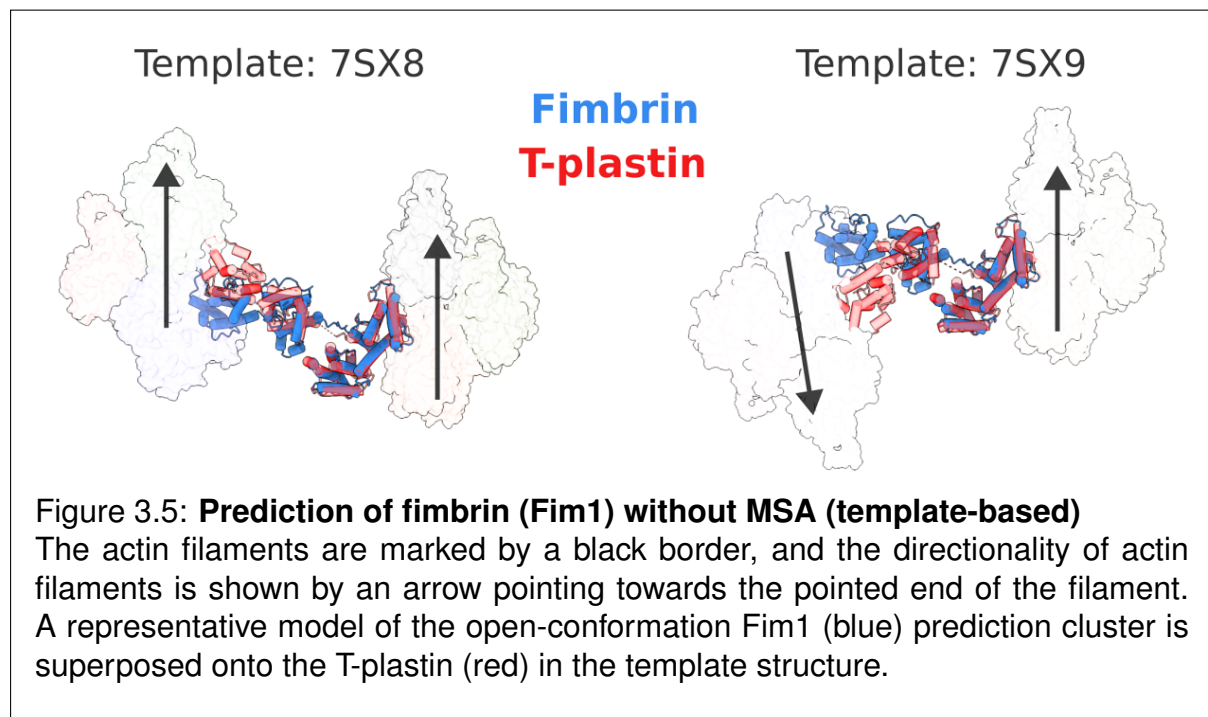
B	Cluster	Fim1 (O59945)	cluster 10	cluster 12
				
	Predicted domains			
		cluster 36	cluster 22	cluster 28
				
	Predicted domains			

Figure 3.4: Multiple conformations of Fim1 using MSA manipulation strategies

(A) Shallow_MSA - For MSA depth 16, (superposed structures on the left) the rank 1 model corresponds to a compact state similar to prediction without MSA subsampling, and (on the right) rank 10 model has misfolded CH domains indicated by high RMSD values (in Å). (B) AF_cluster - Clusters 10 and 12 detect two CH domains (CH1 and CH2) as separate but have misfolded CH3 and CH4. Clusters 36, 22 and 28 detect two ABDs as separate domains however, the ABS on ABD1 is still occluded. The predicted domains are according to the method by Tristan Croll (https://github.com/tristanic/pae_to_domains) as implemented in ChimeraX.

Using the Shallow_MSA method, reducing MSA depth to 16 resulted in Fim1 predictions that were either misfolded (misfolded CH domains as marked by high RMSD values in Figure 3.4 A, right) or highly similar to the compact state of Fim1 as observed in the Fim1 prediction without subsampling the MSA (Figure 3.4 A, left).

Using the AF_cluster method, we obtained 86 clusters from the MSA sequences, and we used the largest five clusters to predict Fim1. As seen in Figure 3.4 B, the PAE plot of Fim1 prediction fetched from AFDB detected the two Actin-binding domains (ABDs) as a single domain (magenta) corresponding to the compact state of Fim1, but the PAE maps of Fim1 predictions using AF_cluster, the two ABDs were detected as separate domains (magenta and cyan in cluster 36, 22, and 28). Moreover, in two of the cases (clusters 10 and 12), individual calponin homology (CH) domains were detected (magenta and cyan in cluster 10; magenta, cyan, and green in cluster 12). However, none of the Fim1 models was suitable for further analysis as the actin-binding site on the ABD1 was still occluded in the predictions of clusters 36, 22 and 28, and the ABD is misfolded in clusters 10 and 12 (marked by red in Figure 3.4).



As mentioned earlier, human T-plastin has been captured in three different conformations. We used them as templates and predicted the Fim1 without providing any MSA information. Figure 3.5 shows the representative predictions for the two

templates (7SX8 and 7SX9 in the antiparallel state). When the pre-bundled state (7SXA) was used as a template, all the predictions were in a state in which the crosslinked actin filaments clashed with one another. Hence, we only proceeded with the models from the first two templates (7SX8 and 7SX9) for further analysis.

3.1.3 Flexibility of Ain1 and Fim1

We analysed the crosslinkers' flexibility when crosslinked to actin filaments using the generated models (see Methods). The plots in Figure 3.6 show the distance and torsion angle between the actin filaments crosslinked by Fim1 (blue) or Ain1 (orange). The range of torsion angles spanned by Fim1 crosslinked actin filaments (Interquartile range of $\approx 211^\circ$) is much larger than that of the Ain1 crosslinked actin filaments (Interquartile range of $\approx 25^\circ$) (Figure 3.6 B). Moreover, the torsion angle in the case of Ain1 lies in the approximate range of -159 to -134° , indicating an antiparallel arrangement of actin filaments. The interfilament distance between the crosslinked filaments is smaller for Fim1 (≈ 111 Å) than for Ain1 (≈ 283 Å) (Figure 3.6 A). In this case, the Fim1 crosslinked filament models match well in terms of interfilament distance as the peak distance between the nearest neighbour filaments is shown to be ≈ 120 Å by Swulius *et al.* (2018). Ain1 crosslinked actin filament models also lie in the range shown in the distribution of interfilament distances by Swulius *et al.* (2018). The validation of the analysis and biological plausibility of the generated models remains to be tested experimentally.

Although we predict the Ain1 dimer to crosslink the actin filaments predominantly in antiparallel arrangement, Li *et al.* (2016) have observed F-actin bundles of mixed polarity in vitro using TIRF microscopy. However, they propose that the flexibility comes from the opening of CH domains based on the R216E mutation in Ain1. However, in all-atom MD simulations of Ain1 (Morita *et al.*, 2020), such conformational change was not observed. So, it remains unclear how Ain1 crosslinks the actin filaments until an experimentally determined structure is obtained.

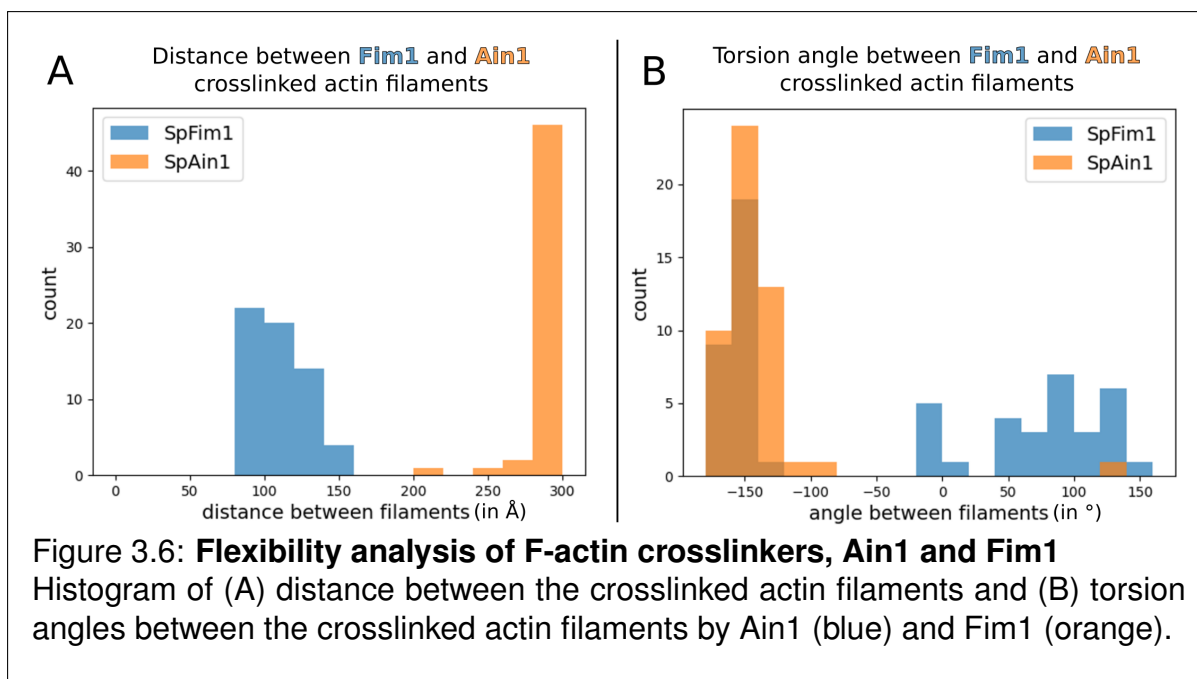


Figure 3.6: Flexibility analysis of F-actin crosslinkers, Ain1 and Fim1
Histogram of (A) distance between the crosslinked actin filaments and (B) torsion angles between the crosslinked actin filaments by Ain1 (blue) and Fim1 (orange).

3.2 Protein complexes in fission yeast contractile ring

As shown by the network of proteins in Figure 1.1, the fission yeast CAR involves a large number of protein-protein interactions. We decided to focus on the two modules described in the introduction involving the proteins - Mid1, Cdc15, Rng2, Cdc12, Act1, Cdc4, Myo2, Myp2, Cdc3, Fim1 and Ain1 (Figure 3.7 A).

Some of these proteins (Cdc12, Rng2, Mid1, Myo2) are multidomain proteins with large intrinsically disordered regions (IDRs) in between. So, it is a difficult task to predict the complexes of full-length proteins. Also, it would not be efficient to do so when the interaction regions span only a few amino acids, and the interaction is independent of the structure of the rest of the protein. However, if the structures of only the interface regions of the complexes are known, then they could be used as an input in the Integrative Modeling Platform (IMP) (Webb *et al.*, 2018) or Assemblin (Rantos *et al.*, 2022) to model the whole complexes, provided sufficient additional restraints from the experimental data. For the fission yeast CAR, we predicted the following inter-protein interactions¹ (Table 3.1 and Figure 3.7). 19 of the 27 predictions in Table 3.1 show above 0.55 ipTM score, indicating the reliability of the prediction.

¹ All the 27 reliable predictions of inter-protein interactions along with the table of interacting residues can be seen in the Appendix A

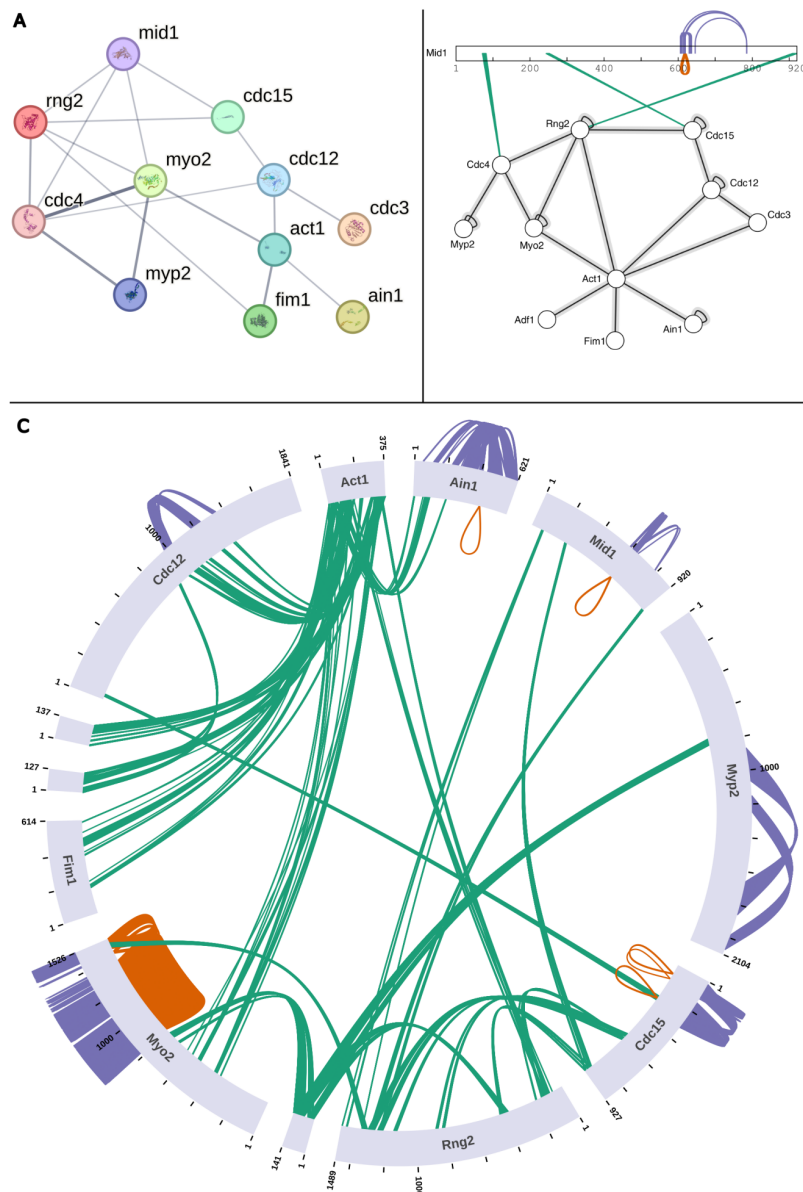


Figure 3.7: Subnetwork of proteins in CAR

(A) The subnetwork of the STRING network from Figure 1.1 shows the interactions among the proteins involved in the two modules described in the introduction. (B) 2D map and (C) chord diagram of residue-residue interactions among the reliably predicted protein complexes. Each node can be expanded to see the full-length protein sequence. The green lines indicate residue level inter-protein interactions obtained from the predictions. The purple and orange lines indicate the interactions in homodimeric structures.

Table 3.1: AlphaFold predictions of protein complexes in fission yeast CAR

Protein 1	Residue range of protein 1	Protein 2	Residue range of protein 2	ipTM score (rank 1 model)	pLDDT	Used templates for predictions (yes/no)
act1	1 - 375	fim1	1-383	0.754	90	yes
act1	1 - 375	fim1	383-614	0.817	91.9	yes
act1	1-375	myo2	1 - 745	0.806	86.5	no
act1	1-375	adf1	1-137	0.943	95.1	no
act1	1-375	cdc12	980-1434	0.88	88.8	no
act1	1-376	ain1	1-240	0.591	92.2	no
act1	1-377	rng2	30-190	0.564	92.8	yes
cdc12	20-40	cdc15 (dimer)	210-270	0.682	85.4	no
cdc4	1-141	rng2	359-684	0.897	90.5	no
cdc4	1-141	myp2	765-820	0.896	85.4	no
cdc4	1-141	myo2	765 - 813	0.875	91.9	no
mid1	1-100	rng2	1306-1489	0.247	66.9	no
myo2	816-1526	myo2	816-1526	0.286	64.4	no
ain1	1-621	ain1	1-621	0.661	88.4	no
myp2	828-1287	myp2	1610-2104	0.287	63.5	no
act1	1-375	cdc3	1-127	0.917	94.7	no
cdc12	900-917	cdc3	1-127	0.704	94.1	yes
cdc15	1-320	cdc15	1-320	0.783	86.6	no
mid1	581-920	mid1	581-920	0.423	68.9	no
cdc12	980-1434	cdc12	980-1434	0.65	84.8	no
cdc15	313-866	rng2	1-190	0.43	37.5	no
cdc15	867-927	rng2	191-712	0.707	54.2	no
mid1	1-580	cdc15	867-927	0.904	33.7	no
myo2	1390-1526	rng2	1188-1489	0.323	67.5	no
rng2	1188-1489	cdc15	313-866	0.513	44.4	no
rng2	1188-1489	mid1	788-920	0.338	70.7	no
rng2	713-1187	cdc15	313-866	0.58	44.6	no

The amino acid sequence given as an input for the pair of proteins is denoted by residue range columns. ipTM and average pLDDT scores of the rank 1 models show the prediction accuracy of the interfaces.

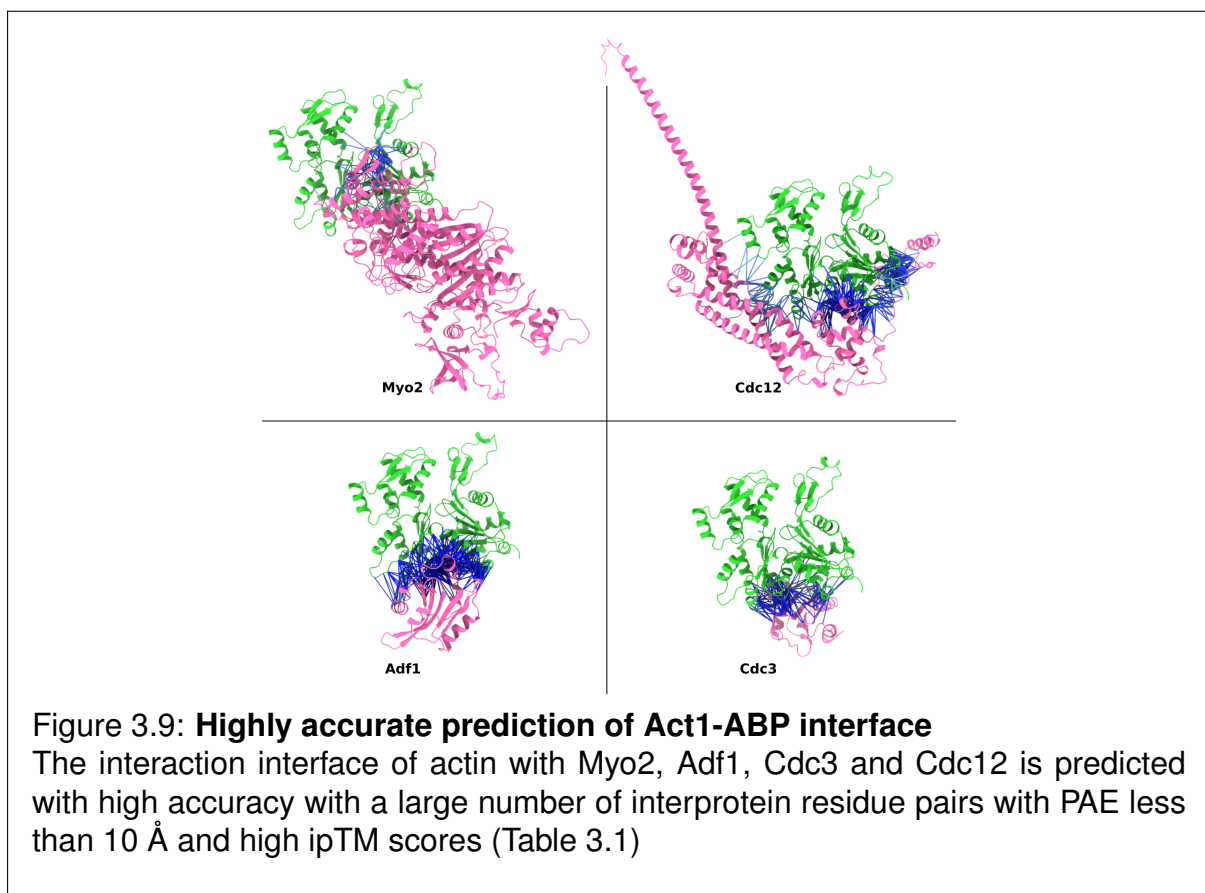
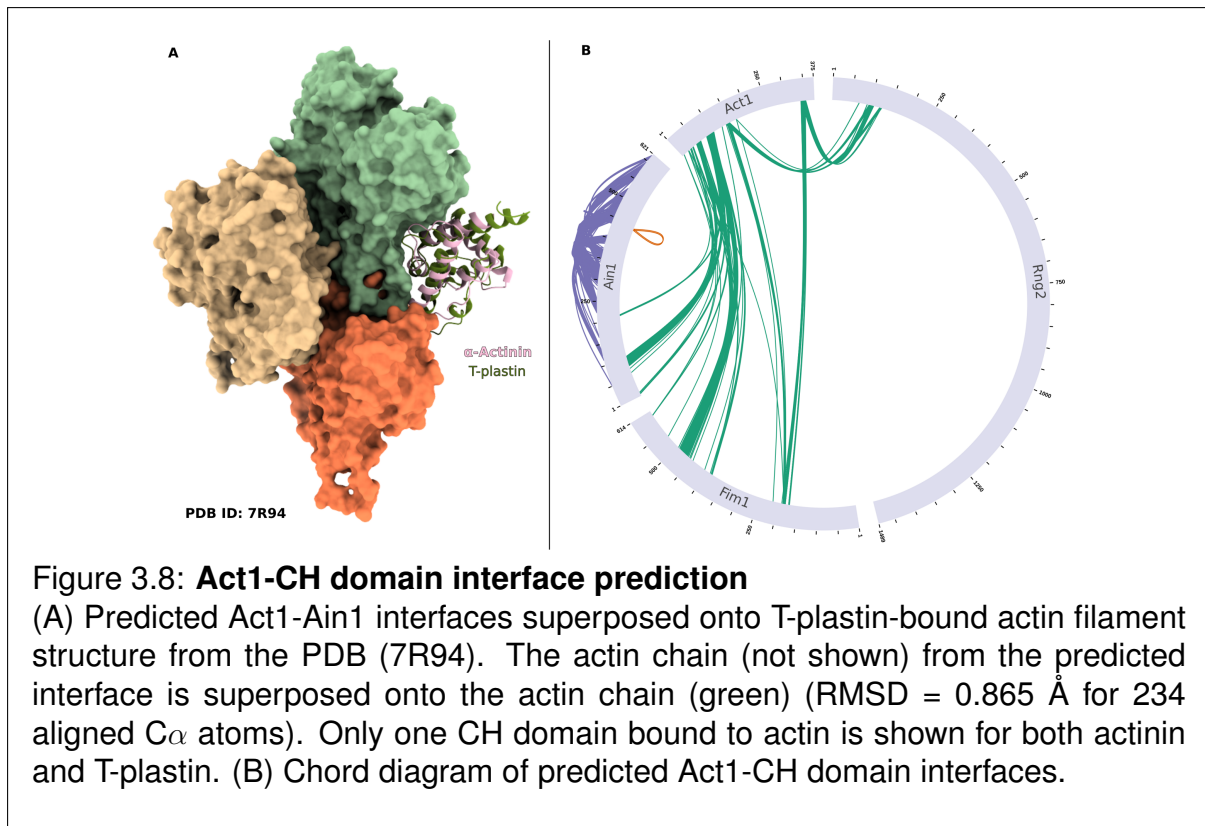
In the case of relatively smaller proteins, such as Act1, Cdc3, and Adf1, their full-length sequence was used for the prediction. For the rest of the predictions (except Ain1 dimer), the delineated sequences were provided. For the first 20 predictions in Table 3.1, we relied upon the existing information from the literature. However, such information is limited for the inter-protein interactions in the nodes. For such cases, we delineated the protein sequences based on domain boundaries and predicted the structure for all possible combinations. The reliability of all the predictions was checked as described in the section 2.4.4. Figure 3.7 B and C shows the 2D map and chord diagram² of these predictions as implemented in xiVIEW (Graham *et al.*, 2019).

3.2.1 Actin-ABP interactions in CAR

We predicted the structure of 8 distinct actin-ABP interactions (Figure 3.7 and Table 3.1). Four of them (Rng2, Ain1, Fim1-ABD1 and Fim1-ABD2) involve conserved actin-CH domain interactions. However, only the Act1-Ain1 interface was predicted with ipTM > 0.55 without using templates and superposed well with the existing CH domain-bound actin filament structures (Figure 3.8 A). The remaining proteins (Rng2, Fim1) required templates for reliable prediction. Although the input for predictions included only a monomer of Act1, their superposition onto the actin filament bound CH-domain structures indicate that AF2 predicts the binding of ABPs as if they are bound to an actin filament.

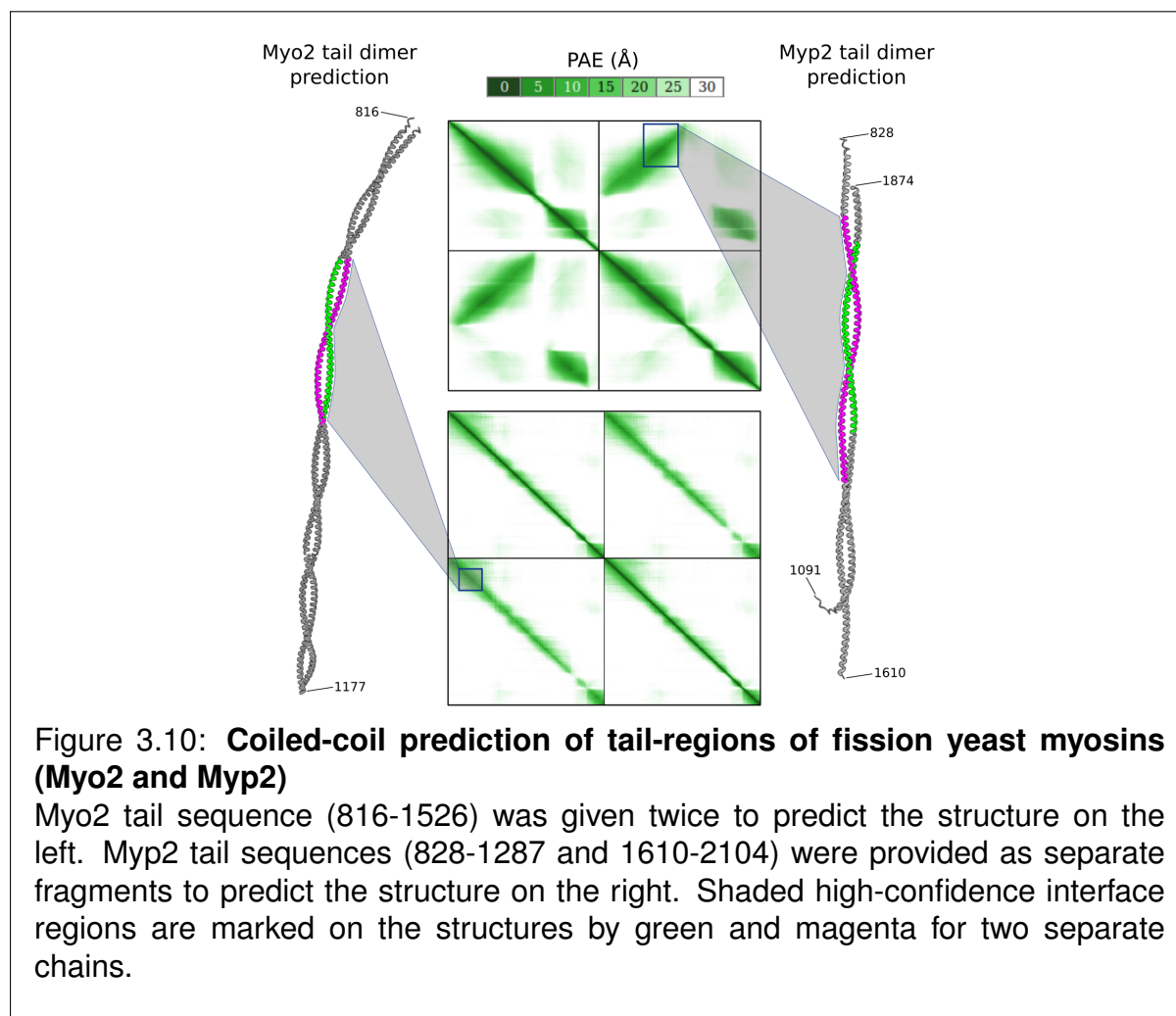
The rest four predictions (Act1 with Myo2, Cdc12, Adf1, Cdc3) have high (>0.8) ipTM values, indicating high confidence in the predicted interfaces (Figure 3.9).

²Interactive chord diagram of protein complex predictions



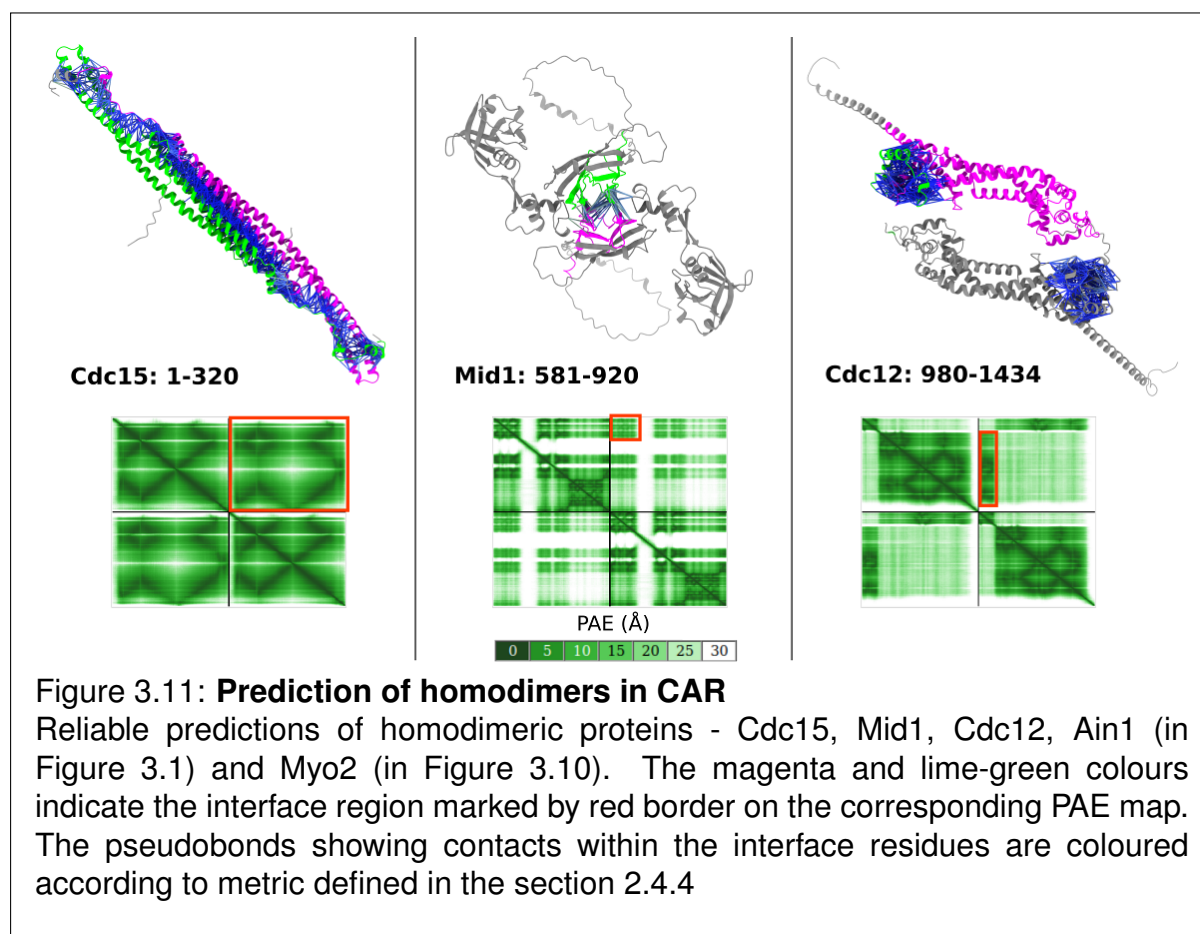
3.2.2 Coiled-coil prediction of fission yeast myosin tails

Among the models predicted with low ipTM scores, the coiled-coil prediction of two myosins of fission yeast - Myo2 and Myp2 have low interchain PAE values (Figure 3.10), indicating good confidence in the relative position of the two chains. It is known that the ipTM values are affected by the disordered or low pLDDT regions (EMBL-EBI), which might be the reason for the low ipTM score in this case (Table 3.1). The direction of off-diagonal patterns in the PAE maps indicates the parallel and antiparallel chain arrangements in the coiled-coil predictions of Myo2 and Myp2, respectively. This supports the proposition by Bezanilla *et al.* (1997) that Myo2 exists as a dimer, and the tails of each chain form a coiled-coil structure, whereas Myp2 exists as a monomer, and its tail folds on itself to form the coiled-coil structure. It is also noteworthy that Myo2 tails formed coiled-coil structures in all the five predicted models as opposed to just one model in the case of Myp2.



3.2.3 Homodimeric CAR proteins

The proteins Mid1, Cdc12, Rng2, Myo2, Ain1, and Cdc15 are known to form homodimers in vivo. However, the experimental structure of the dimer interface is present for Mid1 and Cdc15 only. We obtained reliable homodimer predictions for all of them except Rng2, for which we did not observe any interaction among the Rng2 chains even after increasing the number of recycles to 20 (Figure 2.6). Figure 3.11 shows the homodimer predictions from Table 3.1.



3.3 Actin structures dataset

A total of 468 entries were obtained from the PDB with the input sequence of human cytoplasmic actin (P60709) and a sequence similarity cutoff of 54%. The query ensured that the output structures have at least one actin entity, and each structure is released on or before 01/03/2024. 192 out of the 468 structures were actin filament

structures.

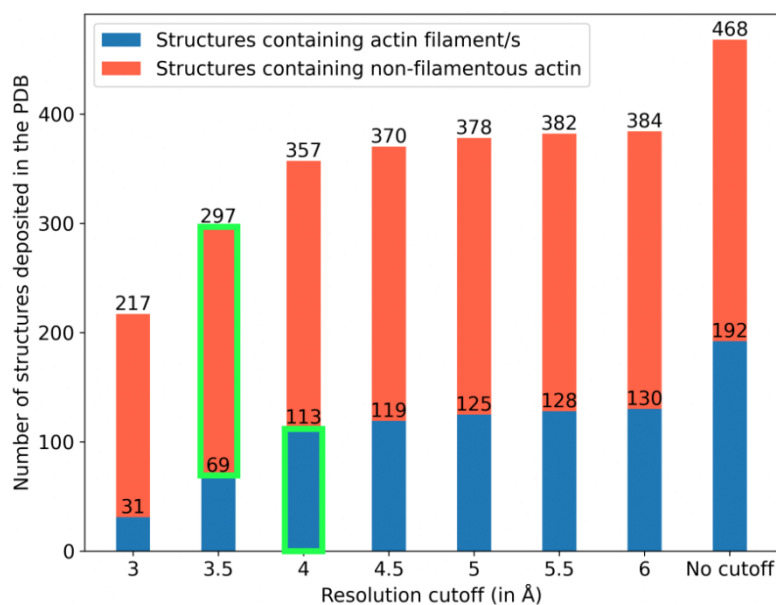


Figure 3.12: Number of actin structures with varying resolution cutoff
The structures chosen after the resolution cutoff of 4 Å for actin filaments and 3.5 Å for the rest of the actin structures are indicated by a green border.

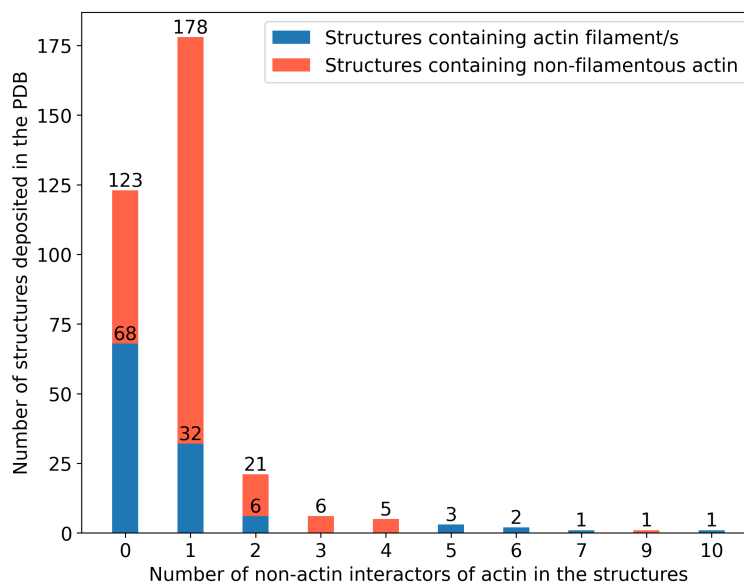


Figure 3.13: Number of actin structures with varying numbers of non-actin interacting proteins of actin
The X-axis indicates the number of distinct proteins (according to UniProt accession) actin interacts with (Phalloidin and Lifeact were not counted).

Figure 3.12 shows the number of actin structures (filamentous and non-filamentous

actin) with varying resolution cutoffs. After the resolution cutoff of 4 Å for the actin filaments and 3.5 Å for the rest of the structures, we ended up with a total of 341 actin structures (113 actin filaments and 228 non-filamentous actin structures).

Figure 3.13 shows the classification of actin structures according to the number of interacting partners (proteins other than actin) of actin. Oligopeptides such as Phalloidin and Lifeact were not considered actin-interactors in this list but were included in small molecules bound to actin filaments (see Figure 3.14). 123 structures have bare-actin entities, and 68 of them are actin filaments. In 178 structures, actin is complexed with one interacting protein, and 32 of them are actin filaments with 20 unique (by UniProt ID) interacting partners (Figure 3.13 and 3.17).

Apart from the protein interactors, there are 15 unique small molecules (13 ligands + 2 oligopeptides) bound to the actin filament structures (Figure 3.14). There are 125 Adenosine diphosphate (ADP)-bound actin filaments in the PDB. None of the actin filaments was in an Adenosine triphosphate (ATP)-bound state; however, 23 had the released phosphate (PO₄) present in the nucleotide-binding pocket alongside ADP. Only two of the structures (5OOE, 6DJM) have actin filament in the AMP-PNP (ANP as per PDB identifier)-bound state.

The numbers on the top of bars in the Figure 3.13 indicate the number of structures after applying the resolution cutoffs, whereas in Figures 3.12 and 3.14, they indicate the number of structures before applying the resolution cutoff. There is no obvious correlation observed between the small molecule bound to the actin filament and helical parameters; however, phalloidin-bound actin filaments (marked by green in Figure 3.16) are slightly on the higher side in terms of the helical twist compared to the actin filaments not bound to any small molecule except the nucleotide-binding pocket (Table 3.2).

Table 3.2: **Average rise and twist of actin filaments bound to small molecules**

Small molecule bound to F-actin	Number of structures	Average Rise (Å)	Average twist (°)
ADP	89	27.81 ± 0.42	166.66 ± 0.96
ADP, PO ₄	22	27.51 ± 0.26	166.70 ± 0.24
AMP-PNP	2	27.43 ± 0.01	166.73 ± 0.07
No small molecule except in the nucleotide-binding pocket	79	27.74 ± 0.38	166.51 ± 0.94
Jasplakinolide	13	27.73 ± 0.37	167.12 ± 0.64
Lifeact	1	27.42	167.04
Phalloidin	17	27.73 ± 0.51	167.07 ± 0.2

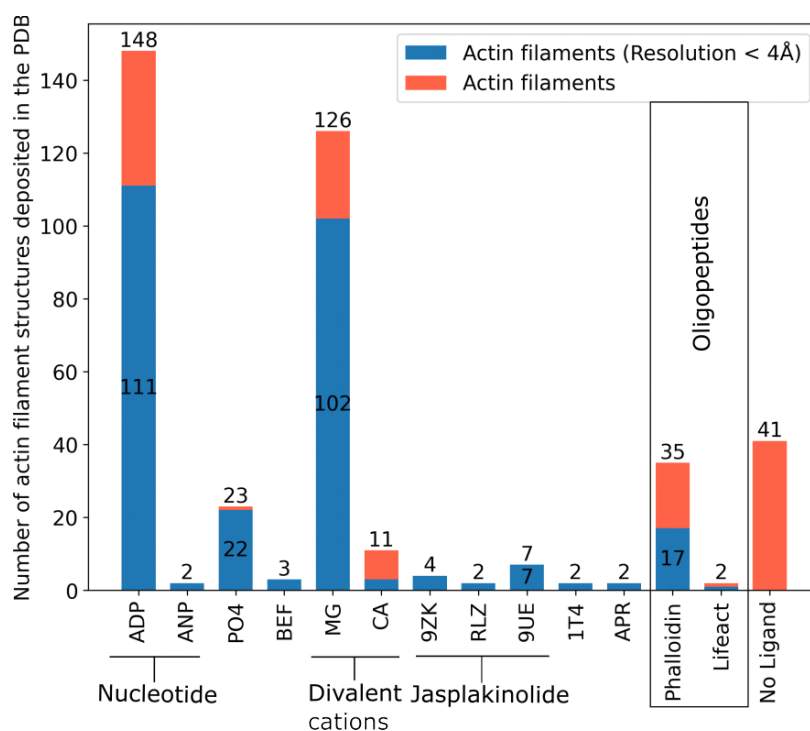
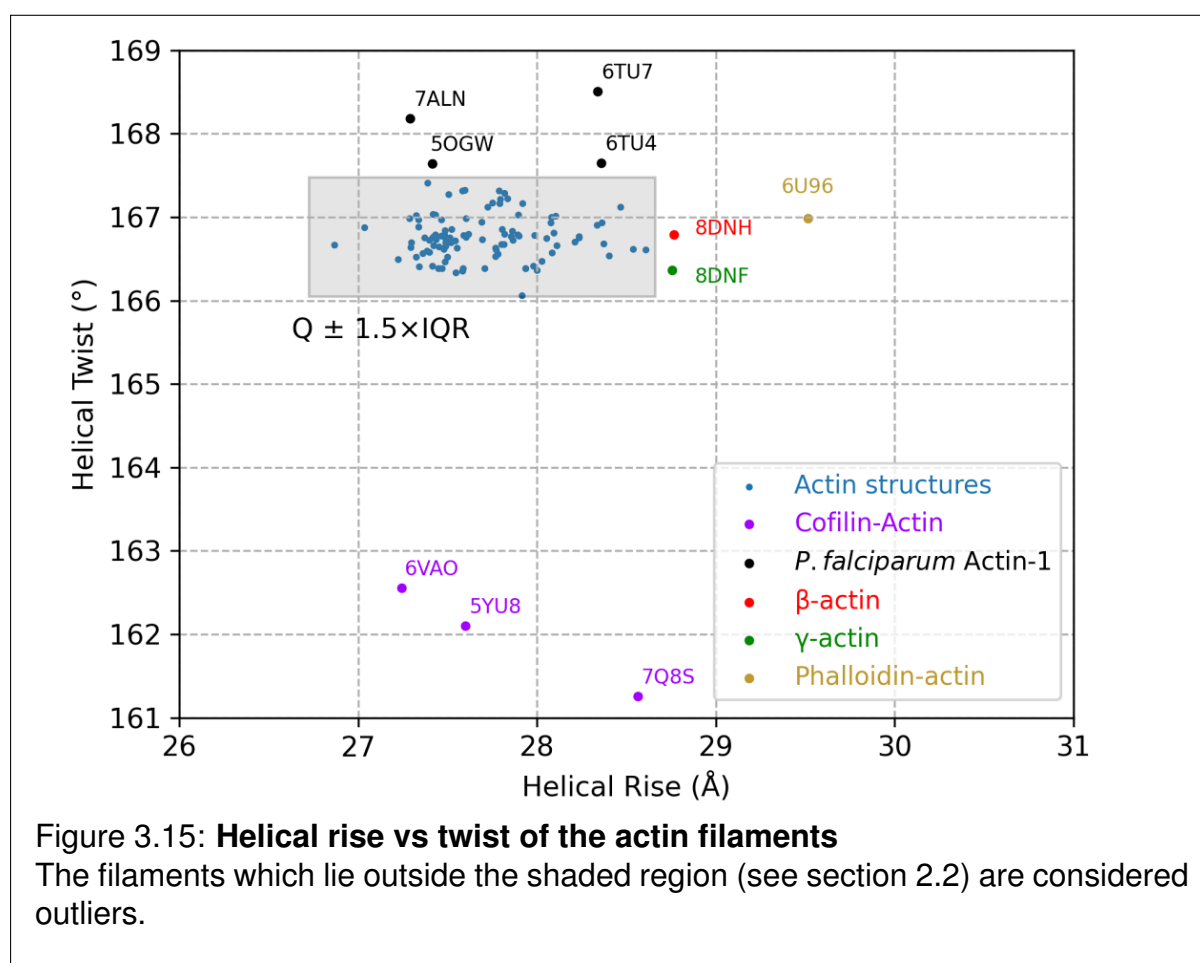


Figure 3.14: **Number of actin filaments with different small molecules bound**
The small molecules are denoted on the X-axis as per their PDB identifier. The numbers in the blue bars indicate the number of actin filaments after applying the resolution cutoff of 4 Å, whereas the number on the top of the bars indicates the total number of actin filaments without any resolution cutoff.

3.4 Actin filament helicity

The average calculated helical rise and twist of all the actin filaments was 27.74 ± 0.41 Å and $166.67 \pm 0.86^\circ$ respectively. Except for the outliers, the rise and twist values clustered in the narrow ranges (26.73 to 28.66 Å) and (166.05 to 167.48°), respectively (shaded region in Figure 3.15). While most of the filaments show negligible local variation in rise and twist, some of them vary significantly (Figure 3.15). Such outlier filaments are discussed in detail in the following subsections.



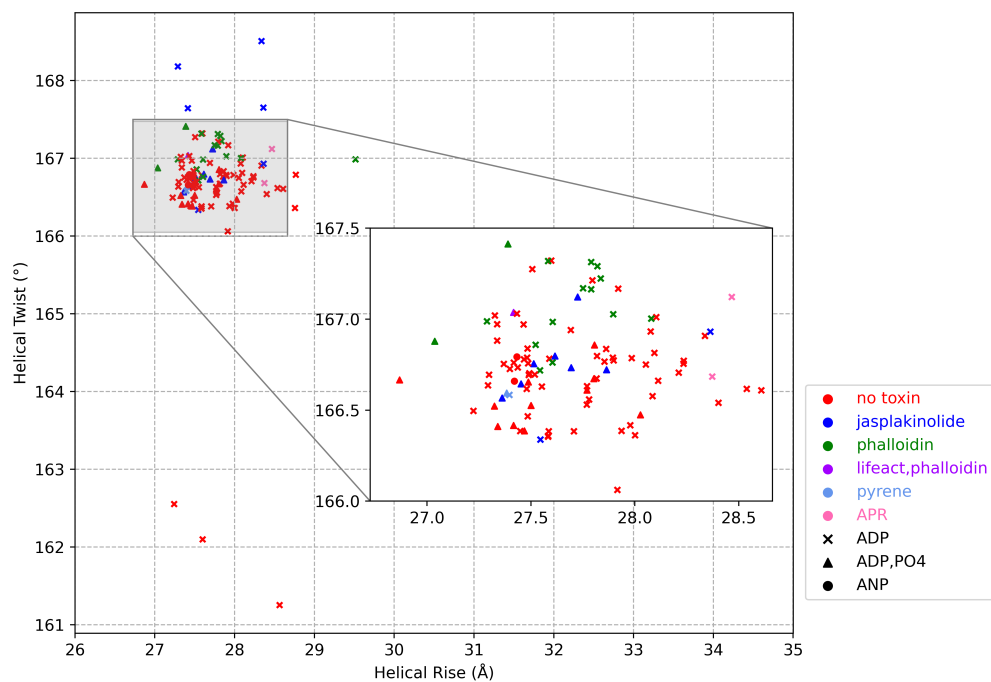


Figure 3.16: Helical rise vs twist of actin filaments with marked nucleotide states and bound small molecule

The nucleotide states are indicated by marker shape and small molecules bound to the actin filaments are indicated by marker colour.

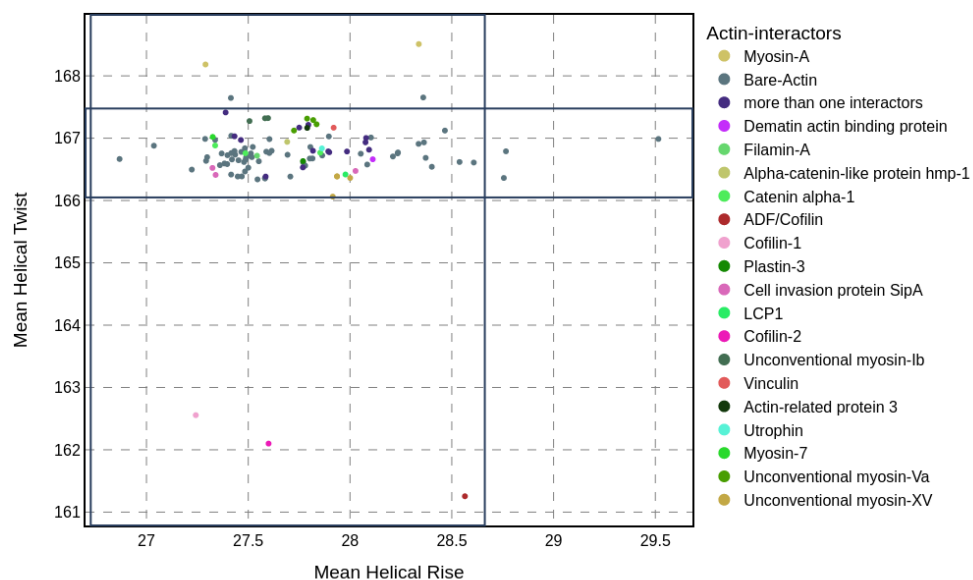


Figure 3.17: Helical rise vs twist for actin filaments complexed with only one non-actin protein

The actin interactors are marked by marker colour.

3.4.1 *Plasmodium falciparum* actin-1 show increased helical twist compared to the actin-2

Four (7ALN, 6TU7, 6TU4, 5OGW) out of all the actin filaments show more than 1° deviation (table 3.6) in the helical twist values as compared to the cluster (shaded region in Figure 3.15). The average twist of these four filaments is $168.00 \pm 0.37^\circ$. The individual rise and twist values are mentioned in Table 3.3, and very low standard deviations for all of them indicate the accuracy of capturing the symmetry axis using our method. All four of them contain *Plasmodium falciparum* actin. Out of a total of six actin filament structures available in the PDB from three different strains of *Plasmodium falciparum*, the aforementioned outliers are all Actin-1 (Q8I4X0), whereas the remaining two structures (8CCO, 8CCN), which lie within the bounds, are of Actin-2 (Q8ILW9). All six filaments are in the ADP-bound state, and five of them (7ALN, 6TU7, 6TU4, 5OGW, 8CCO) are bound to Jasplakinolide. Two of them (6TU7, 7ALN) are Myosin-A-bound actin filaments.

Table 3.3: Helical parameters of *Plasmodium falciparum* actin filaments

PDB ID	Ligand	Interacting Protein	Average Rise (Å)	Average Twist (°)
5OGW	ADP, MG, 9UE	-	27.415 ± 0.007	167.645 ± 0.050
6TU4	ADP, MG, 9UE	-	28.360 ± 0.000	167.653 ± 0.000
6TU7	ADP, MG, 9UE	Myosin-A	28.338 ± 0.001	168.510 ± 0.000
7ALN	ADP, MG, 9UE	Myosin-A	27.290 ± 0.004	168.183 ± 0.026
8CCN	ADP, MG	-	28.337 ± 0.006	166.909 ± 0.012
8CCO	ADP, MG, 9UE	-	28.363 ± 0.001	166.932 ± 0.003

The ligand column indicates the PDB identifiers of the ligands bound to the corresponding actin filament (ADP - Adenosine diphosphate, MG - Mg^{2+} , 9UE - Jasplakinolide). The proteins that interact with the filament are denoted, as well as the average values of rise and twist with the standard deviation.

3.4.2 Cofilin-bound actin filaments show a significant decrease in the helical twist from the rest of the filaments

There are three Cofilin-bound F-actin structures in the PDB (5YU8, 6VAO, 7Q8S) below the resolution cutoff of 4 Å. All three structures show a significant decrease

(more than 4°) in the helical twist compared to the other actin filaments (Figure 3.15). The average helical twist of these three filaments is $161.97 \pm 0.54^\circ$.

Table 3.4: **Helical parameters of cofilin-bound actin filaments**

PDB ID	Ligand	Interacting Protein	Average Rise (Å)	Average Twist (°)
5YU8	ADP, MG	Cofilin-2	27.600 ± 0.000	162.100 ± 0.000
6VAO	ADP, MG	Cofilin-1	27.242 ± 0.028	162.556 ± 0.085
7Q8S	ADP, MG	ADF/Cofilin	28.565 ± 0.000	161.255 ± 0.001

3.4.3 Outlier actin filaments in terms of helical rise

All the outliers discussed above showed deviation in terms of helical twist but not helical rise. There were three actin filaments in the dataset that deviated significantly from the average helical rise. Two of them (8DNH and 8DNF) are human β and γ actin isoforms, respectively, and the remaining structure (6U96) is a Phalloidin-bound rabbit actin filament. Out of 17 structures in the dataset, Phalloidin-bound actin filaments. The helical parameters of these outliers are indicated in the following table.

Table 3.5: **Actin filaments having higher helical rise compared to the others**

	Source Organism	PDB ID	Ligand	Average Rise (Å)	Average Twist (°)
β -actin	<i>Homo sapiens</i>	8DNF	ADP, MG	28.756 ± 0.005	166.362 ± 0.000
γ -actin	<i>Homo sapiens</i>	8DNH	ADP, MG	28.766 ± 0.045	166.789 ± 0.000
Skeletal muscle actin	<i>Oryctolagus cuniculus</i>	6U96	ADP	29.515 ± 0.011	166.988 ± 0.043

Table 3.6: **Average helical rise and twist values of outlier actin filaments**

	Average Rise (Å)	Average twist (°)
All actin filaments	27.74 ± 0.41	166.67 ± 0.86
Cluster (shaded region)	27.70 ± 0.33	166.76 ± 0.26
Plasmodium F-actin	27.85 ± 0.50	168.00 ± 0.37
Cofilin-bound F-actin	27.80 ± 0.56	161.97 ± 0.54
Helical rise outliers	29.01 ± 0.36	166.71 ± 0.26

3.4.4 Bent actin filaments and pointed-end of F-actin show local variation in helical rise

As we calculated the per subunit helical rise and twist in our method, We were able to see the local differences in these parameters in a single actin filament. As shown in Figure 3.19, based on the Z-scores corresponding to the deviation in the helical rise, three of all the structures ($Z\text{-score} > \pm 3$) fall in the outliers. Two of them (8F8S, 8F8T), with the highest Z-scores, are of pointed end of actin filaments. We compared the local variation of helical rise and twist of barbed and pointed-end actin filaments (Figure 3.18). The pointed end-actin filaments show local variation of more than 1 Å. Upon closer inspection, it can be seen that the pointed end filaments are not exactly straight but slightly curved (magenta-coloured spline, approximating the true symmetry axis of the filaments in Figure 3.18, showing poor overlap with the fitted straight line(green)). On the other hand, the barbed end filaments (8F8Q, 8F8R) are more or less straight, as seen by the good overlap of the spline and its straight-line approximation. The remaining structure (7TPT), which has a $Z\text{-score} > 3$, is an actin filament in the branched-actin junction and similar to the pointed end, it has a slight bent at one end of the filament, accounting for the local variation in the helical twist. However, the same is not observed in the case of other similar actin filaments at the branched-actin junction (8UXX, 8UXW).

Similar local variation, based on the Z-scores corresponding to the deviation in the helical twist, can be seen in the helical twist. Two of the structures (8D15 and 8D16 with $z\text{-score} > \pm 3$) fall in the outliers (Figure 3.19). These two structures are known to be bent actin filaments (Reynolds *et al.*, 2022) and are shown to have local cyclic

variations in the helical rise and twist.

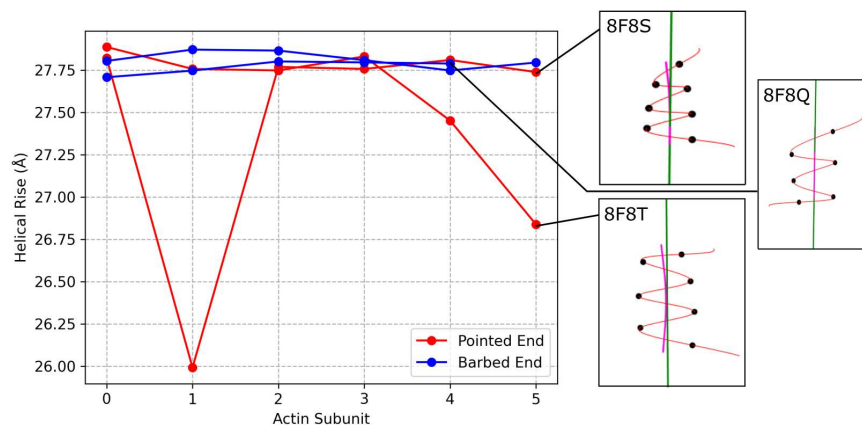


Figure 3.18: Local variation in the helical rise of barbed and pointed-end of actin filaments

Barbed end structures (8F8Q, 8F8R) shown in blue and pointed-end structures (8F8S, 8F8T) shown in red. The insets on the left indicate the bent nature of the pointed end of actin filaments (indicated by the magenta-coloured spline approximating the filament axis) as opposed to the straight barbed end.

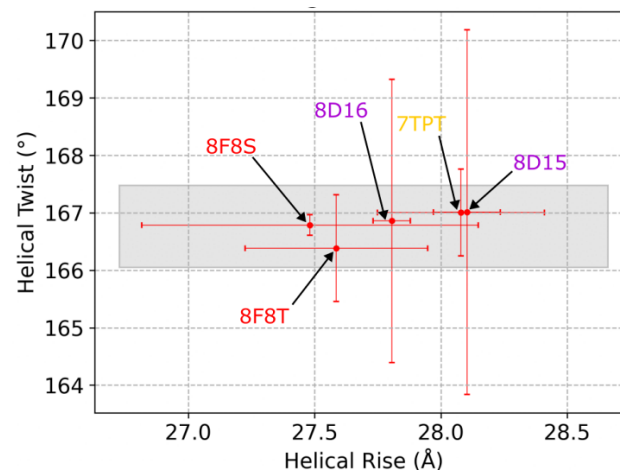


Figure 3.19: Actin filaments that show local variation in helical rise or twist
7TPT, 8F8S and 8F8T have Z-score (of standard deviation in the helical rise) > 3 and 8D15 and 8D16 have Z-score (of standard deviation in the helical twist) > 3. The error bars indicate the standard deviation in the calculation, pointing to the local variation.

3.5 Actin and Actin-binding proteins

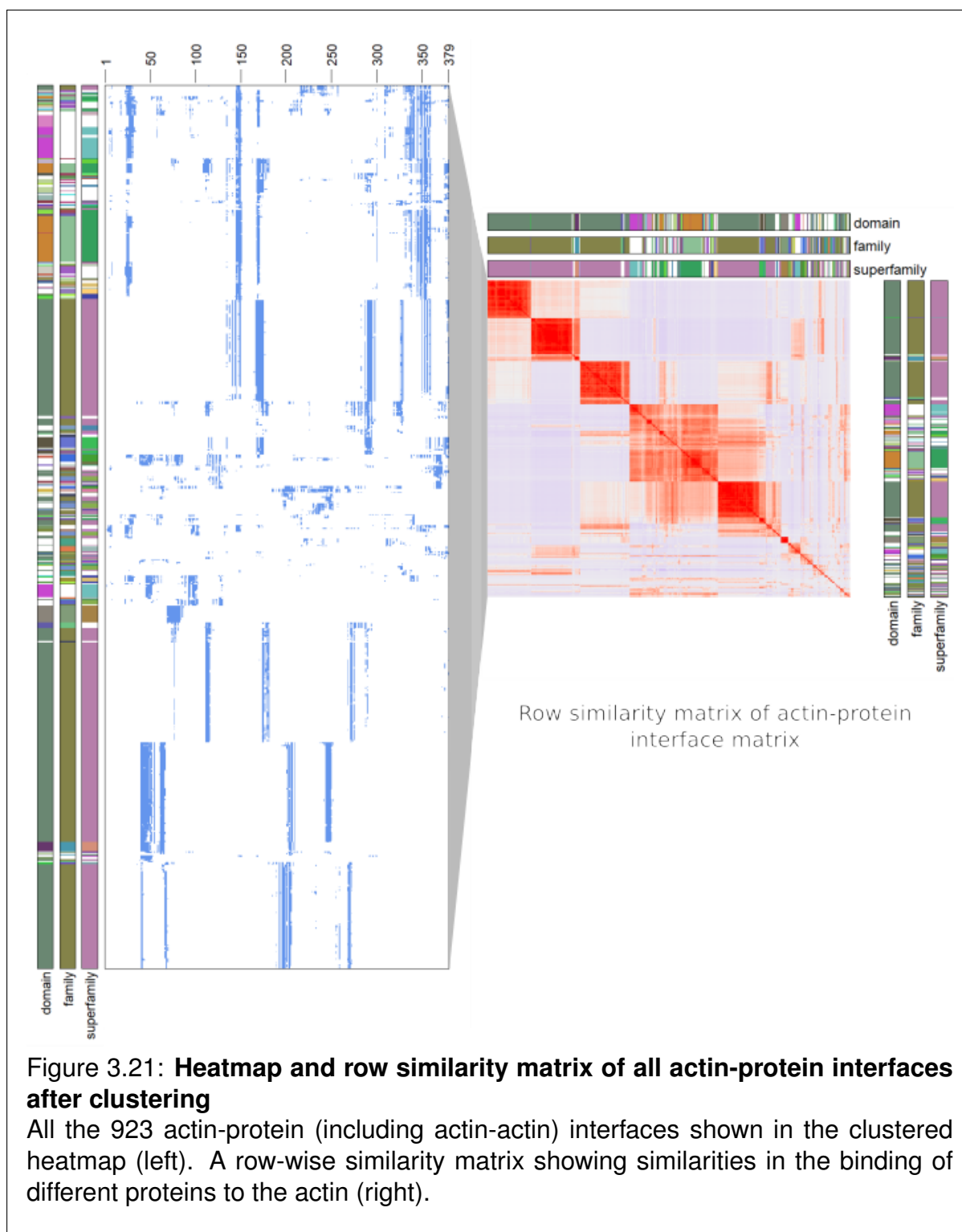
A total of 923 unique interfaces were identified from the PDB for the prepared dataset in which at least one of the proteins was actin, and the interacting partner of actin had an InterPro signature (domain/family/superfamily annotation). We further divided the interfaces into actin-actin (451) and actin-(non-actin) protein (472) interactions. Unless stated otherwise, actin-protein interfaces refer to the interaction of actin with a protein other than actin itself. As described in the section 2.3, we used a heatmap representation to visualise the interfaces. Figure 3.21 shows the clustered heatmap of 923 interfaces, and a similarity matrix. Figure 3.20 shows the sequence logo of 30 actin entities from the dataset, indicating very high sequence similarity in the actin entities.



Figure 3.20: **Sequence logo of consensus sequence of actin entities in the dataset**

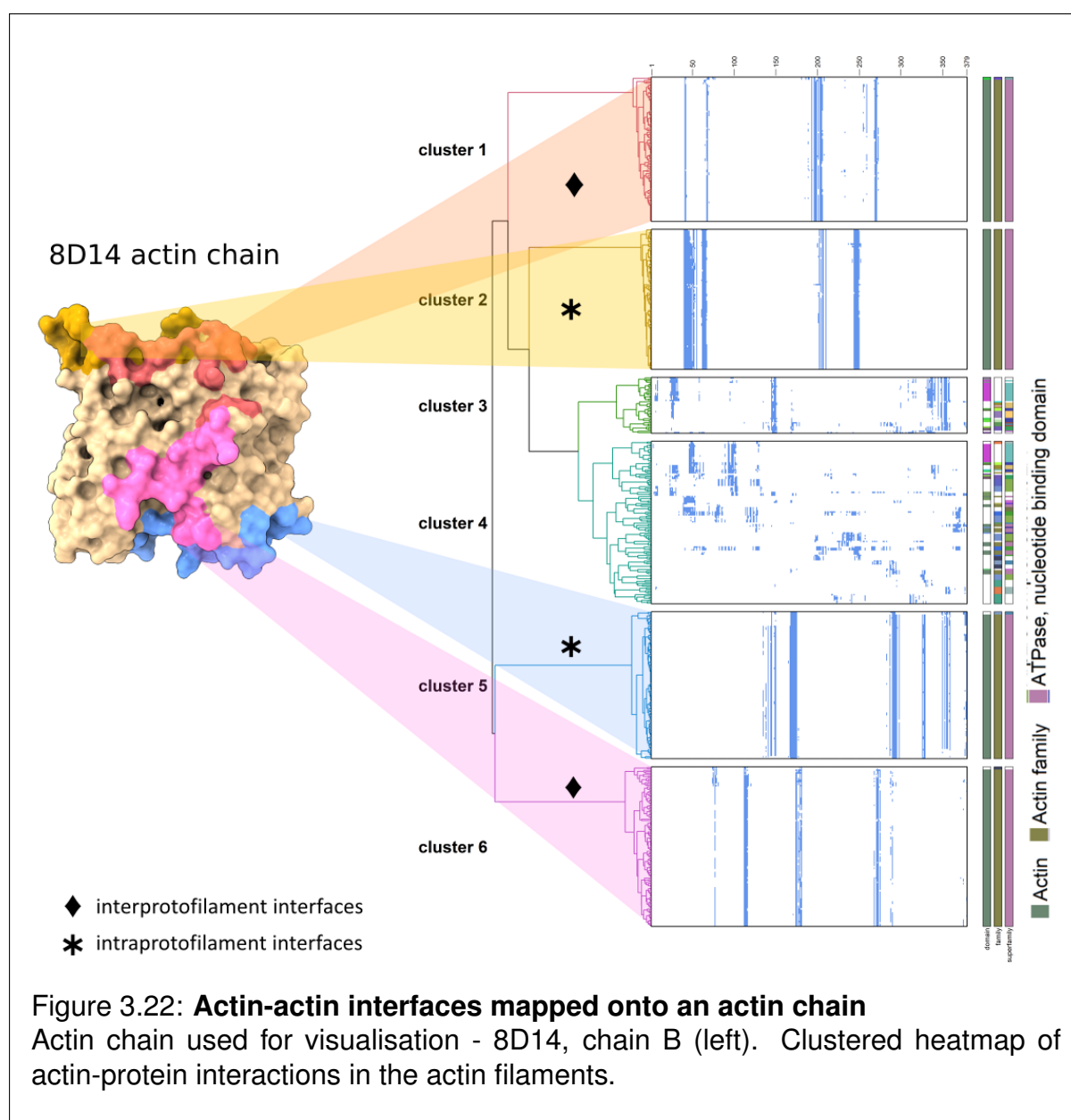
379 amino acid long sequence logo of 30 actin entities from the dataset. The sequence logo is generated from web server WebLogo (Crooks *et al.*, 2004)

For ease of analysis, we looked at the heatmap and row-similarity matrix of subsets of the whole interface matrix.



3.5.1 Interfaces of actin in actin filaments

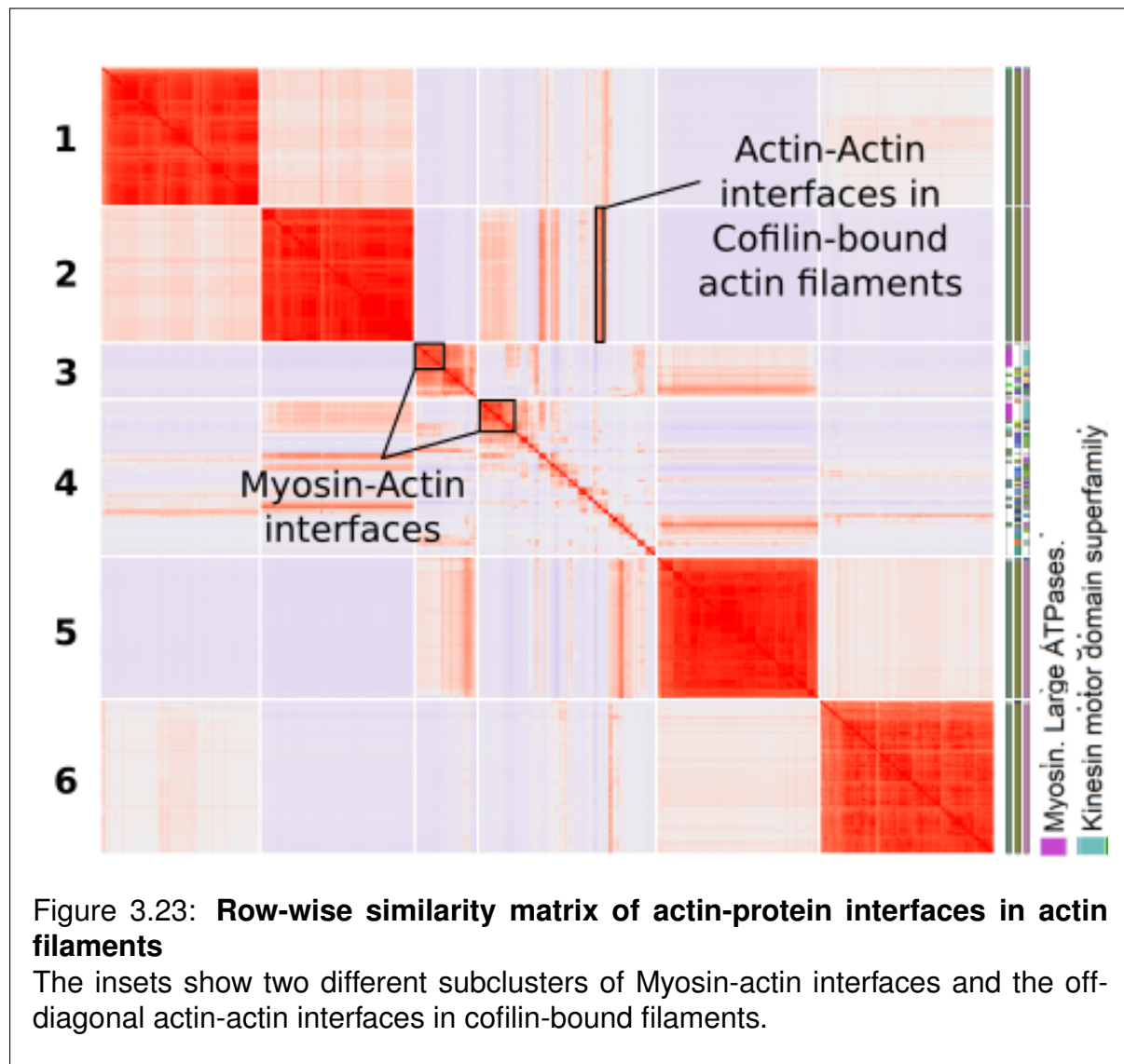
Four distinct clusters (clusters 1, 2, 5, 6) were observed in the row similarity matrices (Figure 3.21 and 3.23) as well as in the heatmap (Figure 3.22). The domain annotations for almost all of the interfaces in these clusters were “Actin”. These clusters correspond to the actin-actin interfaces that are formed in the filament. The clusters are mapped onto an actin chain of one of the representative actin structures (8D14) in Figure 3.22.



Interestingly, actin-actin interfaces of cofilin-bound actin filaments do not belong to these clusters and are present at the off-diagonal regions of the row-similarity matrix

(Figure 3.23).

Other than the four major clusters of actin-actin interfaces, two sub-clusters of Actin-myosin interfaces are observed, one in cluster 3 and the other in 4. Further analysis needs to be done to see why Myosin-actin interfaces are segregated into two different clusters instead single one.

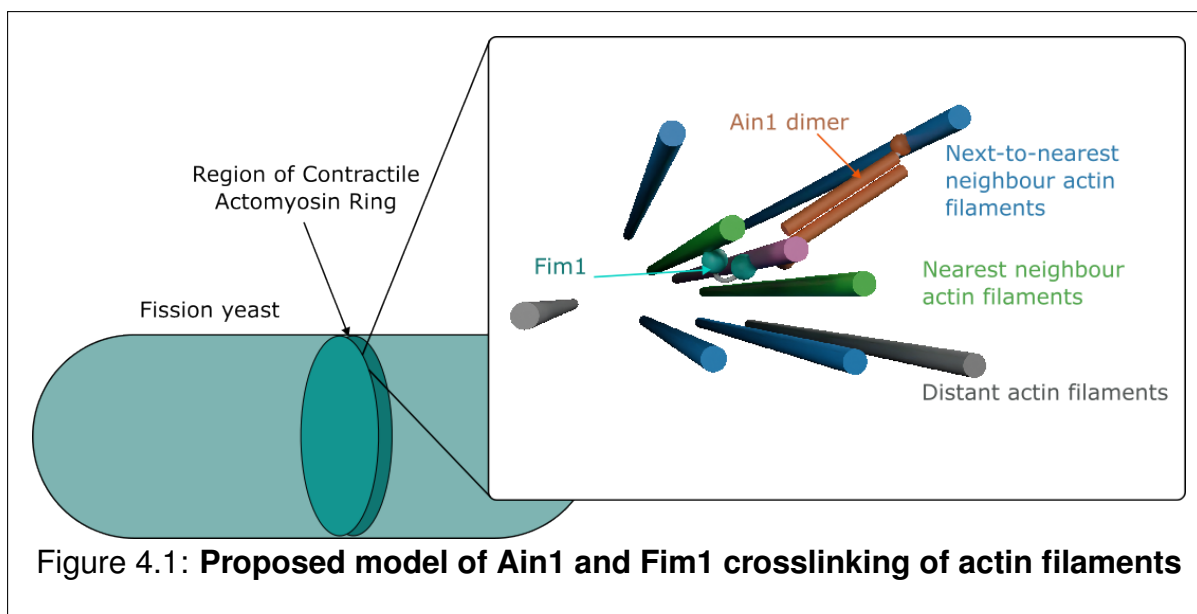


Chapter 4

Conclusions and Future Perspectives

By using the newly developed AlphaFold-based structure prediction tools, we modelled interprotein interactions in the fission yeast contractile ring. To tackle the problem of large computation time, we used delineated sequences of the interacting proteins guided by the existing literature to model the interfaces of the protein complexes. We assessed the predictions using PAE, ipTM and pLDDT metrics and created a map of the reliable interactions (Figure 3.7 B and C). Using MSA manipulation strategies described in (del Alamo *et al.*, 2022; Wayment-Steele *et al.*, 2024), we modelled actin filament crosslinkers, Ain1 and Fim1 in multiple conformations. Based on the range of torsion angles spanned by the crosslinked filaments, we hypothesize that Fim1 is a more flexible F-actin crosslinker than Ain1. Additionally, the peak interfilament distance observed by Swulius *et al.* (2018) matches with the interfilament distance in the Fim1 crosslinked actin filament models and Ain1 crosslinked actin filament models also lie within the distance distribution. Based on this, Fim1 is probably involved in crosslinking actin filament with its nearest neighbours whereas Ain1 might be crosslinking it with the next to nearest neighbours (Figure 4.1).

While we rely upon the information from the literature to model the interfaces of inter-protein interactions, the process of finding such information is cumbersome and time-consuming. One possible strategy to get around this problem is performing a large-scale screening of interprotein interactions by computational means which can be used to first identify the probable interactors and then predict the interprotein interfaces. While we tried this on a small scale with the combinatorial predictions of Cdc15, Rng2, Myo2 and Mid1, for large scale, a well-designed workflow is necessary



to analyse the results of the predictions in an efficient way. The interface predictions from the study can serve as rigid body input structures to various integrative modelling tools. However, these structures alone cannot suffice to model the multi-protein assemblies in the contractile ring and should be complemented by the restraints provided from the experiments such as Crosslinking mass-spectrometry, Hydrogen-Deuterium Exchange, Small angle X-ray scattering and volume restraints from Cryo-electron microscopy. The MSA manipulation strategies can predict proteins with conformational heterogeneity; however, the accuracy of such predictions cannot be assessed using the pLDDT or PAE metric. In vivo, whether the proteins exist in these predicted conformations or not can only be verified experimentally. There are new methods in development, such as AlphaFlow and ESMflow (Jing *et al.*, 2024), which are proposed to sample the conformational space of proteins more accurately than MSA subsampling. Such methods can be applied to model the F-actin crosslinkers Ain1 and Fim1.

We calculated helical rise and twist of actin filaments using a custom method based on the script provided by Reynolds *et al.* (2022). We observed that *Plasmodium falciparum* actin-1 filaments show increased helical twist whereas cofilin-bound actin filaments show significantly decreased value of helical twist. Also, the human β and γ actin isoforms along with a phalloidin-bound actin filament show unusually high helical rise. Our method for calculating helical parameters of actin

filaments works well for straight actin filaments without the need for extending the filament. However, in our method, as we approximate a straight line for the symmetry axis of the filament, it does not capture the rise and twist values of the bent actin filaments correctly. However, such structures are identified based on our defined Z-score metric.

By analysing the actin-protein interfaces across the PDB, we find that cofilin-bound actin filaments have altered actin-actin interface compared to other actin-actin interfaces and Myosin-actin interfaces form two distinct clusters indicating that there are two distinct interaction types between myosin and actin. Further analysis is required to comment on the other off-diagonal patterns observed in the similarity matrix. Our heatmap and similarity matrix visualisation of the actin-protein interfaces combined with the agglomerative clustering method provide an easy way of comparing multiple actin-protein interactions at once. We think our workflow for this analysis can be extended to for more candidate proteins for which sufficient structures are available in the PDB to capture the structure-function patterns.

References

- Ahdritz G, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, O'Donnell TJ, Berenberg D, Fisk I, Zanichelli N, Zhang B, Nowaczynski A, Wang B, Stepniewska-Dziubinska MM, Zhang S, Ojewole A, Guney ME, Biderman S, Watkins AM, Ra S, Lorenzo PR, Nivon L, Weitzner B, Ban YEA, Sorger PK, Mostaque E, Zhang Z, Bonneau R, AlQuraishi M (2023). OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Pages: 2022.11.20.517210 Section: New Results.
- Almonacid M, Celton-Morizur S, Jakubowski JL, Dingli F, Loew D, Mayeux A, Chen JS, Gould KL, Clifford DM, Paoletti A (2011). Temporal control of contractile ring assembly by Plo1 regulation of Myosin II recruitment by Mid1/Anillin. *Current biology* : CB 21(6), 473–479.
- Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Opperman DJ, Sagmeister T, Buhlheller C, Pavkov-Keller T, Rathinaswamy MK, Dalwadi U, Yip CK, Burke JE, Garcia KC, Grishin NV, Adams PD, Read RJ, Baker D (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373(6557), 871–876. Publisher: American Association for the Advancement of Science.
- Bezanilla M, Forsburg SL, Pollard TD (1997). Identification of a Second Myosin-II in *Schizosaccharomyces pombe*:: Myp2p Is Conditionally Required for Cytokinesis. *Molecular Biology of the Cell* 8(12), 2693. Publisher: American Society for Cell Biology.

- Bhavsar-Jog YP, Bi E (2017). Mechanics and regulation of cytokinesis in budding yeast. *Seminars in Cell & Developmental Biology* 66, 107–118.
- Bittrich S, Schroeder M, Labudde D (2019). StructureDistiller: Structural relevance scoring identifies the most informative entries of a contact map. *Scientific Reports* 9(1), 18517. Publisher: Nature Publishing Group.
- Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009). Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols* 4(1), 1–13. Publisher: Nature Publishing Group.
- Bret H, Gao J, Zea DJ, Andreani J, Guerois R (2024). From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. *Nature Communications* 15(1), 597. Publisher: Nature Publishing Group.
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004). WebLogo: a sequence logo generator. *Genome Research* 14(6), 1188–1190.
- Dana JM, Gutmanas A, Tyagi N, Qi G, O'Donovan C, Martin M, Velankar S (2019). SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins. *Nucleic Acids Research* 47(D1), D482–D489.
- del Alamo D, Sala D, Mchaourab HS, Meiler J (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* 11.
- Dominguez C, Boelens R, Bonvin AMJJ (2003). HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *Journal of the American Chemical Society* 125(7), 1731–1737. Publisher: American Chemical Society.
- D'Avino PP, Giansanti MG, Petronczki M (2015). Cytokinesis in Animal Cells. *Cold Spring Harbor Perspectives in Biology* 7(4), a015834.
- Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J,

- Hassabis D (2022). Protein complex prediction with AlphaFold-Multimer. Pages: 2021.10.04.463034 Section: New Results.
- Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D (2005). Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. *Nature* 437(7061), 1043–1047. Number: 7061 Publisher: Nature Publishing Group.
- Gisselsson D, Jin Y, Lindgren D, Persson J, Gisselsson L, Hanks S, Sehic D, Mengelbier LH, Øra I, Rahman N, Mertens F, Mitelman F, Mandahl N (2010). Generation of trisomies in cancer cells by multipolar mitosis and incomplete cytokinesis. *Proceedings of the National Academy of Sciences of the United States of America* 107(47), 20489–20493.
- Goyal A, Takaine M, Simanis V, Nakano K (2011). Dividing the spoils of growth and the cell cycle: The fission yeast as a model for the study of cytokinesis. *Cytoskeleton* 68(2), 69–88.
- Graham M, Combe C, Kolbowski L, Rappsilber J (2019). xiView: A common platform for the downstream analysis of Crosslinking Mass Spectrometry data. Pages: 561829 Section: New Results.
- Grant GR, Manduchi E, Stoeckert Jr CJ (2007). Analysis and Management of Microarray Gene Expression Data. *Current Protocols in Molecular Biology* 77(1), 19.6.1–19.6.30. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471142727.mb1906s77](https://onlinelibrary.wiley.com/doi/pdf/10.1002/0471142727.mb1906s77).
- Gu Z (2022). Complex heatmap visualization. *iMeta* 1(3), e43. [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/imt2.43](https://onlinelibrary.wiley.com/doi/pdf/10.1002/imt2.43).
- Guertin DA, Trautmann S, McCollum D (2002). Cytokinesis in Eukaryotes. *Microbiology and Molecular Biology Reviews* 66(2), 155–178.
- Jegou A, Romet-Lemonne G (2020). The many implications of actin filament helicity. *Seminars in Cell & Developmental Biology* 102, 65–72.
- Jing B, Berger B, Jaakkola T (2024). AlphaFold Meets Flow Matching for Generating Protein Ensembles. *ArXiv:2402.04845 [cs, q-bio]*.

- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873), 583–589. Publisher: Nature Publishing Group.
- Kamasaki T, Osumi M, Mabuchi I (2007). Three-dimensional arrangement of F-actin in the contractile ring of fission yeast. *Journal of Cell Biology* 178(5), 765–771.
- Karaca E, Rodrigues JPGLM, Graziadei A, Bonvin AMJJ, Carlomagno T (2017). M3: an integrative framework for structure determination of molecular machines. *Nature Methods* 14(9), 897–902. Publisher: Nature Publishing Group.
- Kovar DR, Kuhn JR, Tichy AL, Pollard TD (2003). The fission yeast cytokinesis formin Cdc12p is a barbed end actin filament capping protein gated by profilin. *Journal of Cell Biology* 161(5), 875–887.
- Kovar DR, Sirotkin V, Lord M (2011). Three's company: The fission yeast actin cytoskeleton. *Trends in cell biology* 21(3), 177–187.
- Krissinel E, Henrick K (2007). Inference of macromolecular assemblies from crystalline state. *Journal of Molecular Biology* 372(3), 774–797.
- Lacroix B, Maddox AS (2012). Cytokinesis, ploidy and aneuploidy. *The Journal of Pathology* 226(2), 338–351. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/path.3013>.
- Laplanche C, Huang F, Tebbs IR, Bewersdorf J, Pollard TD (2016). Molecular organization of cytokinesis nodes and contractile rings by super-resolution fluorescence microscopy of live fission yeast. *Proceedings of the National Academy of Sciences* 113(40), E5876–E5885. Publisher: Proceedings of the National Academy of Sciences.
- Laporte D, Coffman VC, Lee IJ, Wu JQ (2011). Assembly and architecture of precursor nodes during fission yeast cytokinesis. *Journal of Cell Biology* 192(6), 1005–1021.

- Laporte D, Ojkic N, Vavylonis D, Wu JQ (2012). α -Actinin and fimbrin cooperate with myosin II to organize actomyosin bundles during contractile-ring assembly. *Molecular Biology of the Cell* 23(16), 3094–3110.
- Lee IJ, Coffman VC, Wu JQ (2012). Contractile-ring assembly in fission yeast cytokinesis: Recent advances and new perspectives. *Cytoskeleton* 69(10), 751–763. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/cm.21052>.
- Li Y, Christensen JR, Homa KE, Hocky GM, Fok A, Sees JA, Voth GA, Kovar DR (2016). The F-actin bundler α -actinin Ain1 is tailored for ring assembly and constriction during cytokinesis in fission yeast. *Molecular Biology of the Cell* 27(11), 1821–1833.
- Li Z, Liu X, Chen W, Shen F, Bi H, Ke G, Zhang L (2022). Uni-Fold: An Open-Source Platform for Developing Protein Folding Models beyond AlphaFold. Pages: 2022.08.04.502811 Section: New Results.
- Lin Z, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, dos Santos Costa A, Fazel-Zarandi M, Sercu T, Candido S, Rives A (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 379(6637), 1123–1130. Publisher: American Association for the Advancement of Science.
- Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, Lopez R (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research* 50(W1), W276–W279.
- McDonald NA, Lind AL, Smith SE, Li R, Gould KL (2017). Nanoscale architecture of the *Schizosaccharomyces pombe* contractile ring. *eLife* 6.
- Mei L, Reynolds MJ, Garbett D, Gong R, Meyer T, Alushin GM (2022). Structural mechanism for bidirectional actin cross-linking by T-plastin. *Proceedings of the National Academy of Sciences* 119(37), e2205370119. Publisher: Proceedings of the National Academy of Sciences.
- Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE

- (2023). UCSF ChimeraX: Tools for Structure Building and Analysis. Protein Science: A Publication of the Protein Society e4792.
- Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M (2022). ColabFold: making protein folding accessible to all. Nature Methods 1–4.
- Morita R, Nakano K, Shigeta Y, Harada R (2020). Molecular Mechanism for the Actin-Binding Domain of α -Actinin Ain1 Elucidated by Molecular Dynamics Simulations and Mutagenesis Experiments. The Journal of Physical Chemistry B 124(39), 8495–8503. Publisher: American Chemical Society.
- Morita R, Takaine M, Numata O, Nakano K (2017). Molecular dissection of the actin-binding ability of the fission yeast α -actinin, Ain1, in vitro and in vivo. The Journal of Biochemistry 162(2), 93–102.
- Murtagh F, Legendre P (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification 31(3), 274–295.
- Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar G, Bileschi M, Bork P, Bridge A, Colwell L, Gough J, Haft D, Letunić I, Marchler-Bauer A, Mi H, Natale D, Orengo C, Pandurangan A, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu C, Bateman A (2023). InterPro in 2022. Nucleic Acids Research 51(D1), D418–D427.
- Pereira J, Simpkin AJ, Hartmann MD, Rigden DJ, Keegan RM, Lupas AN (2021). High-accuracy protein structure prediction in CASP14. Proteins: Structure, Function, and Bioinformatics 89(12), 1687–1699. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.26171>.
- Pinotsis N, Zielinska K, Babuta M, Arolas JL, Kostan J, Khan MB, Schreiner C, Salmazo A, Ciccarelli L, Puchinger M, Gkougkoulia EA, Ribeiro EdA, Marlovits TC, Bhattacharya A, DjinoVIC-Carugo K (2020). Calcium modulates the domain flexibility and function of an α -actinin similar to the ancestral α -actinin. Proceedings of the National Academy of Sciences 117(36), 22101–22112. Publisher: Proceedings of the National Academy of Sciences.

- Pintard L, Bowerman B (2019). Mitotic Cell Division in *Caenorhabditis elegans*. *Genetics* 211(1), 35–73.
- Pollard TD, Wu JQ (2010). Understanding cytokinesis: lessons from fission yeast. *Nature Reviews Molecular Cell Biology* 11(2), 149–155.
- Rantos V, Karius K, Kosinski J (2022). Integrative structural modeling of macromolecular complexes using Assemblin. *Nature Protocols* 17(1), 152–176. Publisher: Nature Publishing Group.
- Reynolds MJ, Hachicho C, Carl A, Gong R, Alushin GM (2022). Bending forces and nucleotide state jointly regulate F-actin structure. *Nature* 611(7935), 380–386. Publisher: Nature Portfolio.
- Ribeiro E, Pinotsis N, Ghisleni A, Salmazo A, Konarev P, Kostan J, Sjöblom B, Schreiner C, Polyansky A, Gkougkoulia E, Holt M, Aachmann F, Žagrović B, Bordignon E, Pirker K, Svergun D, Gautel M, Djinić-Carugo K (2014). The Structure and Regulation of Human Muscle α -Actinin. *Cell* 159(6), 1447–1460.
- Rincon SA, Paoletti A (2016). Molecular control of fission yeast cytokinesis. *Seminars in Cell & Developmental Biology* 53, 28–38.
- Rose Y, Duarte JM, Lowe R, Segura J, Bi C, Bhikadiya C, Chen L, Rose AS, Bittrich S, Burley SK, Westbrook JD (2021). RCSB Protein Data Bank: Architectural Advances Towards Integrated Searching and Efficient Access to Macromolecular Structure Data from the PDB Archive. *Journal of Molecular Biology* 433(11), 166704.
- Schultz J (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research* 28(1), 231–234.
- Snider CE, Bhattacharjee R, Igarashi MG, Gould KL (2022). Fission yeast paxillin contains two Cdc15 binding motifs for robust recruitment to the cytokinetic ring. *Molecular Biology of the Cell* 33(4), br4.
- Snider CE, Chandra M, McDonald NA, Willet AH, Collier SE, Ohi MD, Jackson LP, Gould KL (2020). Opposite Surfaces of the Cdc15 F-BAR Domain Create a Membrane Platform That Coordinates Cytoskeletal and Signaling Components for Cytokinesis. *Cell Reports* 33(12), 108526.

- Stein RA, Mchaourab HS (2022). SPEACH_af: Sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLOS Computational Biology* 18(8), e1010483. Publisher: Public Library of Science.
- Swulius MT, Nguyen LT, Ladinsky MS, Ortega DR, Aich S, Mishra M, Jensen GJ (2018). Structure of the fission yeast actomyosin ring during constriction. *Proceedings of the National Academy of Sciences* 115(7), E1455–E1464. Publisher: Proceedings of the National Academy of Sciences.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43(Database issue), D447–452.
- von Dassow G, Verbrugghe KJ, Miller AL, Sider JR, Bement WM (2009). Action at a distance during cytokinesis. *Journal of Cell Biology* 187(6), 831–845.
- Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Hömberger M, Ovchinnikov S, Colwell L, Kern D (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* 625(7996), 832–839. Publisher: Nature Publishing Group.
- Webb B, Viswanath S, Bonomi M, Pellarin R, Greenberg CH, Saltzberg D, Sali A (2018). Integrative structure modeling with the Integrative Modeling Platform. *Protein Science : A Publication of the Protein Society* 27(1), 245–258.
- Willet AH, McDonald NA, Bohnert KA, Baird MA, Allen JR, Davidson MW, Gould KL (2015). The F-BAR Cdc15 promotes contractile ring formation through the direct recruitment of the formin Cdc12. *Journal of Cell Biology* 208(4), 391–399.
- Wood V, Harris MA, McDowall MD, Rutherford K, Vaughan BW, Staines DM, Aslett M, Lock A, Bähler J, Kersey PJ, Oliver SG (2012). PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Research* 40(D1), D695–D699.
- Wu JQ, Bähler J, Pringle JR (2001). Roles of a Fimbrin and an α -Actinin-like Protein in Fission Yeast Cell Polarization and Cytokinesis. *Molecular Biology of the Cell* 12(4), 1061–1077. Publisher: American Society for Cell Biology (mboc).

- Wu JQ, Pollard TD (2005). Counting Cytokinesis Proteins Globally and Locally in Fission Yeast. *Science* 310(5746), 310–314. Publisher: American Association for the Advancement of Science.
- Wu R, Ding F, Wang R, Shen R, Zhang X, Luo S, Su C, Wu Z, Xie Q, Berger B, Ma J, Peng J (2022). High-resolution de novo structure prediction from primary sequence. Pages: 2022.07.21.500999 Section: New Results.
- Wuyun Q, Chen Y, Shen Y, Cao Y, Hu G, Cui W, Gao J, Zheng W (2024). Recent Progress of Protein Tertiary Structure Prediction. *Molecules* 29(4), 832. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- Yang J, Zhang Y (2015). Protein Structure and Function Prediction Using I-TASSER. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]* 52, 5.8.1–5.815.

Appendix A

Additional data

All the 27 reliable predictions can be seen here [Click here to view the image.](#)