

Structural and functional
characterization of
immune system proteins
using crystallography and
single particle cryo-EM

Thesis by
Katelyn M. Radford

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy in Biochemistry and
Molecular Biophysics



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
(Defended February 27th, 2026)

© 2026

Katelyn M Radford
ORCID: 0000-0001-8484-7580

ACKNOWLEDGEMENTS

There are so many people I need to thank for their support, love and encouragement over the last six years.

I would first like to thank Bil Clemons, for giving me the opportunity to come to Caltech in the first place in the summer of 2019. Even though there was no funding for me to be in the lab, Bil allowed me to volunteer for six months, working closely with an excellent postdoc. This was my introduction to the Caltech community, and in this time, I started forming friendships, training in cryo-EM, and singing in the two Caltech choirs.

I would also like to thank my mom, Sue Greer, and her husband DJ Byrne, for allowing me to stay with them, first during my internship with Bil and then during the first year and a half of the pandemic. Mom, I'll never forget when I was telling you about cryo-EM one day and you went, "Oh, wait, let me get something" and came back with your electron microscopy sample tweezers from your graduate work. It means so much to me to have been able to live close to you for six years. Thank you for always being ready with a hug and comfort, thank you for feeding me when I show up on your doorstep, and thank you for your endless shenanigans and can-do spirit.

Even though we live far apart, my dad, Don Radford, has been an amazing resource and support to me throughout my PhD. Dad, thank you so much for the hours of conversation about academia: I have felt so privileged to get to have your perspective when I've been applying for grants or considering next steps or worried about whatever I'm dealing with in a given moment. I also feel so lucky to be part of the same profession as you (sort of – we'll see if I can land a postdoc etc), because it has given us so much that we can connect over and talk about. This, perhaps, is one of the best gifts that doing a PhD could give me. Also, if I don't absolutely bomb, we'll both be Dr. Radford!

Tsui-Fen Chou, my advisor, changed my life in my third year. I was struggling in the lab I joined early in my PhD, and I needed a more understanding home. Tsui-Fen, thank you for being both kind and understanding, but also tough and appropriately critical, throughout the last three years. I really have appreciated your feedback on my work and your mentorship. You have done an excellent job fostering a supportive lab environment and I have really thrived here. I also appreciate so, so much your willingness to let me collaborate with the Aravin group, as well as with the Youle lab: I have

loved being able to work on these projects, and it really means a lot to me to have been able to pursue them.

I need to particularly thank Will Rosencrans, friend and mentor, who believed in me when I could not and has supported me since we started together in BMB. Will is the one who encouraged me to join the Chou lab, and he spearheaded projects that I have been involved in, including everything in Chapter 4. I have learned so much from you and benefitted so much from your kindness and willingness to connect me with other scientists, and I have truly loved getting to work on projects with you.

I also must thank Anton Kuzmenko, longtime collaborator for the RzAgo project. It has been incredible to get to work on RzAgo, with all its quirks. Thank you for your patience and your endless kindness throughout these three years as we've optimized samples, collected just one more data set again, and as I've reanalyzed and reprocessed data in the endless search for the most perfect structure I can obtain from the data we have. This project has been such a joy to get to work on, and I'm so grateful that I've gotten to be a part of it.

Caltech has been one of the most incredible communities to belong to, and that is in large part due to the student population. I have felt out of place for so much of my life, but Caltech has felt like home since the beginning. I have made life-long friends – Will Rosencrans, Noah Epstein, Scott Conn, Dan Pollack, Philip Woods, Anastasiya Ogienko, Masami Hazu, Maria Altshuller, and Georgia Squyres have been particular anchors for me over the years (and I'm sure I'm forgetting some folks). I have also been lucky enough to get to found a choral octet, OcTech, that has had so many more opportunities than could ever have been possible without Caltech and the community fostered here. We have held over a dozen shows since we started, have raised thousands of dollars for charity, and have released two albums together – and we've done it one hour-and-a-half rehearsal at a time. Here's to members of the choir, past and present: Georgia, Maria, Hannah, Magdi, Alma, Michelle, Morgan, Lauren, Annie, Noah, James, Philip, Forte, Thorgal, and Mitch. Thank you for showing up every week because singing together is medicine. I also must thank Nancy Sulahian, choir director extraordinaire, for being the most consistent mentor I've had since I got here: thank you for coming to OcTech rehearsals and shows, for training me individually in conducting, for entrusting me with the Glee Club on occasion, and for supporting me at choral workshops as I conduct alongside people actually going to school for that.

Thank you particularly to Alys Hoy. I feel so profoundly known by you, and you have supported me in ways and to a magnitude that I have rarely been supported before. Thank you for showing up for every OcTech show, for never letting me say a nasty thing about myself, for coming to the Huntington with me and taking me, in turn, for going so far above and beyond for my wedding and for continually asking how you can help. Alys, you have always done enough, and you are always enough. Thank you for making me feel like maybe I am, too.

Finally, thank you to my husband Allen Shariaty. I always hoped I would find someone like you, and here we are: a fellow scientist with a winding backstory and a music streak, who is kind to everyone and can always make me laugh. Thank you for supporting me every day, in big ways and small ways: by doing the laundry and making food for the cat, but also by loving me however I show up. I benefit so much from your conviction that I am smart and skilled, and that I can do whatever I put my mind to – you believe in me even when I don't. Thank you for endlessly supporting me as I've gone through the last couple years of my PhD, even as you were in the thick of your own. Thank you for a million mundane things, like doing the cat's voice with me, teaching me how to play racquetball, and introducing me to pottery. Thank you for taking me seriously both in things that make sense and also things that maybe make less sense. I am so grateful that I have gotten to walk this path by your side, and that whatever happens next, I get to do it with you.

Kate Radford, February 13th 2026

ABSTRACT

Immune system features are ubiquitous across living organisms, and regulate host-pathogen or host-symbiont relationships across domains of life. This thesis explores proteins mediating aspects of immunity in four different systems. First, I explore broadly neutralizing antibodies against the SARS-CoV-2 viral entry protein S. I identified the antibody 11G-12v as belonging to Class 1/4. I additionally found that it shares sequence and structural similarity to a set of previously discovered anti-SARS-CoV-2 RBD antibodies that have a long CDRH3 loop with a YYDRxG motif, which associate with and extend the central beta sheet of the SARS-CoV-2 RBD. Second, I examined the many documented activities of the p97 adapter protein Phospholipase A2 Activating protein in the context of its structure, and considered them in terms of pleiotropy, as some of its functions appear to occur without p97. Third, I solved the crystal structure of the autophagy initiation scaffold FIP200's C-terminal CLAW domain in complex with a 13-amino acid phosphopeptide derived from the TNIP1 FIP200 interacting region, or FIR. This crystal structure supported our understanding of the basis of FIR recognition by CLAW. And finally, I collaborated with the Aravin group and solved five cryo-EM structures of RzAgo, a prokaryotic argonaute protein with two unique N terminal domains. We found that RzAgo forms tetramers to assume an active state, which recapitulates structures previously observed in certain structures of homologous mammalian proteins. However, in its inactive state, RzAgo forms filaments that lock its catalytic sites in a non-catalytic organization. Together, this work demonstrates the contributions of structural biology to studying immune system proteins.

PUBLISHED CONTENT AND CONTRIBUTIONS

Radford, K. M., Rosencrans, W., & Chou, T. F. (2024). *FIP200 C-terminal CLAW domain (resid. 1490-1594) in complex with phosphorylated TNIP1 FIP200 interacting peptide: 9d34* [Dataset]. <https://doi.org/10.2210/pdb9d34/pdb>

KMR obtained the crystal structure of TNIP1 peptide in complex with FIP200 CLAW.

TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract.....	vi
Published Content and Contributions.....	vii
Table of Contents.....	viii
List of Illustrations and/or Tables.....	xi
Nomenclature.....	xiv
 Chapter 1. Structural biology: History, methods, and applications within this thesis...	 1
1. Introduction	1
2. X-ray Crystallography.....	7
2.1. Overview.....	7
2.2. Theory.....	7
2.2.1. Space groups	11
2.2.2. Crystal data collection and Bragg's Law.....	13
2.2.3. Fourier transforms and the phase problem.....	14
2.3. Advantages and Limitations.....	15
3. Cryo-Electron Microscopy (Cryo-EM).....	16
3.1. Overview.....	17
3.2. Theory.....	17
3.2.1. Electron microscope anatomy.....	17
3.2.2. The contrast transfer function and image defocus.....	19
3.2.3. Sample, sample, cold and thin.....	21
3.2.4. From particles to structures.....	22
3.3. Advantages and Limitations.....	23
3.4. Negative-stain electron microscopy.....	24
4. Chapter summaries.....	25
4.1. Chapter 2: Identification and structural characterization of a broadly neutralizing antibody in complex with the SARS-CoV-2 S protein RBD.....	26
4.2. Chapter 3: Literature review: A structural perspective on the p97 adaptor Phospholipase A2 Activating Protein (PLAA).....	26
4.3. Chapter 4: Structural characterization of the FIP200 C-terminal CLAW domain bound to an autophagy adaptor peptide, TNIP1-pFIR.....	27
4.4. Chapter 5: ATP-driven tetramerization and RNA-guided filament assembly control of a Schlafen–Argonaute antiphage defense.....	27
Bibliography.....	30
 Chapter 2. Identification and structural characterization of a broadly neutralizing antibody in complex with the SARS-CoV-2 S protein RBD.....	 35

Abstract.....	35
1. Introduction.....	36
1.1. Overview of SARS-CoV-2	36
1.2. The SARS-CoV-2 S protein.....	37
1.3. Antibodies: an overview.....	39
1.4. Antibody recognition of the SARS-CoV-2 RBD.....	41
2. Identifying a Class 1/4 antibody.....	44
2.1. Screening for mAbs that broadly neutralize SARS-CoV-2 variants of concern..	44
2.2. Antibody classification via competition ELISA	45
2.3. Cross-reactive Fab ELISA.....	46
2.4. Cross-reactive Class 1/4 IgG ELISA.....	47
3. Structural characterization of 11G-12v via cryo-EM.....	48
3.1. Sample preparation, data collection, and data processing	48
3.2. Single-particle data analysis	49
3.3. Analysis of 11G-12v Fab recognition of the S RBD.....	51
4. Discussion.....	53
Figures.....	55
Tables.....	69
Bibliography.....	72
 Chapter 3: Literature review: A structural perspective on the p97 adaptor Phospholipase A2 Activating Protein (PLAA).....	 83
Abstract.....	83
1. Introduction.....	84
2. Functional roles of PLAA	87
2.1. PLAA as a p97 adaptor protein and component of ubiquitin pathways.....	87
2.2. PLAA as a phospholipase A2 activator	89
2.3. Mitochondrial association of PLAA.....	92
2.4. PLAA shows variation in subcellular localization across studies.....	93
3. Structural features of PLAA.....	94
3.1. WD-40 domain	94
3.2. PFU domain.....	98
3.3. Intrinsically disordered region.....	101
3.4. PUL domain.....	102
4. PLAA: Clinical implications.....	105
5. Conclusions and future directions.....	108
Bibliography.....	110
 Chapter 4: Structural characterization of the FIP200 C-terminal CLAW domain bound to an autophagy adaptor peptide, TNIP1-pFIR.....	 132
Abstract.....	132
1. Introduction.....	133

2. Methods.....	141
2.1. Recombinant protein purification.....	141
2.2. Crystallization condition identification and optimization.....	142
2.3. Crystal collection and diffraction experiments.....	143
2.4. Fluorescence polarization peptide competition assays.....	144
3. Results.....	146
3.1. TNIP1 binds to a region of positive charge on an outside face of the CLAW domain.....	146
3.2. Peptide binding changes the conformation of the CLAW dimer interface.....	148
3.3. Binding kinetics of TNIP1-pFIR in comparison with other FIR peptides.....	151
3.4. Fluorescence polarization assay shows contributions from hydrophobic residues to FIR binding.....	153
4. Discussion.....	155
Bibliography.....	158
 Chapter 5: ATP-driven tetramerization and RNA-guided filament assembly control of a Schlafen–Argonaute antiphage defense.....	167
Abstract.....	167
1. Introduction.....	169
2. Results.....	171
2.1. RzAgo belongs to a clade of extra-long pAgos with a conserved Schlafen domain.....	171
2.2. RzAgo binds small RNA guides and recognizes DNA targets.....	172
2.3. Loading of guide RNA drives formation of RzAgo filaments.....	173
2.4. DNA target recognition reroutes the guide RNA and destabilizes the RzAgo filament.....	177
2.5. ATP binding by the GHKL domain drives RzAgo tetramerization.....	179
2.6. RzAgo tetramerization recapitulates mammalian SLFN domain organization and reveals a conserved catalytic pocket.....	181
2.7. RzAgo protects cells from phage infection.....	183
3. Discussion.....	187
4. Methods.....	190
4.1. Biochemical, cell biological and bioinformatic methods.....	190
4.2. Negative stain EM of RzAgo Apo, Binary and Ternary complexes.....	205
4.3. Cryo-EM structure of the RzAgo pAgo core Apo monomer.....	206
Tables.....	212
Figures.....	222
Extended Data Figures.....	236
Bibliography.....	248
 Chapter 6: Conclusions.....	253

LIST OF ILLUSTRATIONS AND TABLES

Chapter 1. Structural biology: History, methods, and applications within this thesis...	1
Figure 1. Size ranges of structural methods.....	2
Figure 2. Molecular masses associated with structural techniques.....	3
Figure 3. Phase separation diagram.....	7
Figure 4. Phase separation diagrams, temperature vs concentration, supersaturation.....	8
Table 1. Overview of crystal systems.....	10
Figure 5. Crystal system diagrams.....	11
Figure 6. X-ray diffraction pattern.....	12
Equation 1. Bragg's Law.....	13
Figure 7. Electron microscope anatomy and electromagnetic lenses.....	17
Figure 8. Diagram of transmission electron microscope features.....	18
Figure 9. Electron scattering through a TEM sample.....	19
Figure 10. The contrast transfer function.....	20
Figure 11. Grid schematic.....	21
Figure 12. Cryo-EM structures in the PDB categorized by resolution.....	24
 Chapter 2. Identification and structural characterization of a broadly neutralizing antibody in complex with the SARS-CoV-2 S protein RBD.....	 35
Figure 1. Spike protein overview.....	55
Figure 2. Antibody overview.....	56
Figure 3. RBD details.....	57
Figure 4. Neutralization data for 16 mAbs cloned in the Hahn lab.....	58
Figure 5. Competition ELISA data.....	59
Figure 6. Indirect ELISA data.....	60
Figure 7. Cross-reactive Class 1/4 IgG ELISA.....	62
Figure 8. Cryo-EM structure information.....	63
Figure 9. Map and model of S trimer in complex with 11G-12v Fab.....	65
Figure 10. Association of YYDRSG motif with RBD beta sheet.....	66
Figure 11. Comparison of footprints and approach angles of six Fabs.....	67
Table 1: Cryo-EM model refinement statistics	68
Table 2: Number of sequences per amino acid substitution for residues within the RBD contacting the YYDRSG motif in 11G-12v.....	70
 Chapter 3: Literature review: A structural perspective on the p97 adaptor Phospholipase A2 Activating Protein (PLAA).....	 83
Graphical abstract.....	83
Figure 1. Domain map of PLAA (top) with illustration of PLAA's domains (middle) and their primary known interacting partner (bottom)	94

Figure 2. PLAA WD-40 domain.....	95
Table 1. Ubiquitin-binding residues in Doa1 are conserved in human PLAA.....	97
Figure 3. NMR structure of human PLAA PFU (PDB ID: 2K8B).....	99
Figure 4. Sequence alignment (CLUSTAL Omega) of PLAA IDRs in five model organisms and humans.....	101
Figure 5. PLAA PFU features.....	103
Table 2. Pathogenic mutations in PLAA.....	107

Chapter 4: Structural characterization of the FIP200 C-terminal CLAW domain bound to an autophagy adaptor peptide, TNIP1-pFIR.....

133

Figure 1. Autophagy background.....	135
Figure 2. FIP200 CLAW domain features.....	137
Figure 3. FIP200 interacting regions and TNIP1 background.....	138
Figure 4. CLAW-TNIP1-pFIR crystals.....	144
Figure 5. Crystal structure of TNIP1-pFIR in complex with CLAW.....	148
Figure 6. Interaction of conserved hydrophobic FIR residue with CLAW aromatic cluster.....	149
Figure 7. CLAW dimer interface changes upon peptide binding.....	151
Figure 8. Effect of charge on CLAW loop 4-5 ability to bind FIR peptides.....	153
Figure 9. Fluorescence polarization assays.....	155
Table 1. TNIP1 and CCPG1 pFIR wild type and phenylalanine-isoleucine mutant sequences and dissociation constants	155
Table 2. Data collection and refinement statistics.....	158

Chapter 5: ATP-driven tetramerization and RNA-guided filament assembly control of a Schlafen–Argonaute antiphage defense.....

168

Table 1. Apo monomer structure statistics.....	213
Table 2. Apo tetramer structure statistics	215
Table 3. Binary dimer structure statistics	217
Table 4. Binary pentamer structure statistics	219
Table 5. Ternary complex structure statistics	221
Figure 1. RzAgo phylogeny and domain architecture.....	223
Figure 2. RzAgo filament formation.....	225
Figure 3. Guide binding and pAgo core interactions with adjacent protomers.....	227
Figure 4. Ternary complex structure.....	229
Figure 5. Schlafase supradomains and RzAgo Apo tetramer structure.....	231
Figure 6. RzAgo Schlafen domain comparison with hSLFN11.....	233
Figure 7. RzAgo protects host cells against phage infection.....	235
Extended Data Figure 1. GSFSC plots and cFAR plots for each structure.....	237
Extended Data Figure 2. Contacts of mobile loops.....	239
Extended data Figure 3. Contacts made between pAgo core and gRNA in the binary complex.....	240

Extended Data Figure 4. 2D class averages of the ternary complex.....	241
Extended Data Figure 5. Contacts made between pAgo core and the gRNA-target DNA heteroduplex in the ternary complex.....	242
Extended Data Figure 6. Comparison of RsAgo ternary complex with RzAgo ternary complex.....	243
Extended Data Figure 7. Docking study in Coot to examine the fit of different nucleobases in the DT21 binding pocket.....	244
Extended Data Figure 8. Interacting helices at the C-terminal base of the RzAgo SLFN domains within a dimer.....	245
Extended Data Figure 9. Coulombic potential comparison of the RzAgo SLFN dimer with mammalian Slfns, hSlfn11, hSlfn12, and rSlfn13.....	246
Extended Data Figure 10. Curvature comparison between ATPases and pAgo cores.....	247
Extended Data Figure 11. Alternate ATPase interactions seen in crosslinked Binary complex structure.....	248

NOMENCLATURE

Term	Definition
Abortive infection (Abi) <i>[Ch. 5]</i>	A prokaryotic antiviral defense strategy in which, upon phage infection, the infected cell activates a toxic response (often self-sacrifice) to prevent phage replication and protect the surrounding bacterial population.
ACE2 (angiotensin-converting enzyme 2) <i>[Ch. 2]</i>	The host cell surface receptor exploited by SARS-CoV-2 Spike for viral entry; binding of the RBD to ACE2 triggers conformational changes leading to membrane fusion and genome delivery.
Affinity maturation <i>[General]</i>	The iterative process during an adaptive immune response in which B cells with higher-affinity antigen receptors are preferentially selected and expanded after somatic hypermutation, leading to progressive improvement of antibody binding to its target.
AlphaFold <i>[Ch. 1]</i>	A deep-learning algorithm (DeepMind, 2021) that predicts three-dimensional protein structures from amino-acid sequence with high accuracy; widely used to provide initial models for molecular replacement and to guide structural interpretation.
Amino acid <i>[General]</i>	The monomer unit of proteins; 20 canonical amino acids are encoded by the genetic code, each with a unique side chain that determines its chemical properties (charged, polar, hydrophobic, aromatic, etc.) and its contribution to protein structure and function.
Ångström (Å) <i>[General]</i>	A unit of length equal to 10^{-10} m (0.1 nm); the standard unit for atomic distances in structural biology. Covalent bond lengths are ~ 1.2 – 1.5 Å, hydrogen bonds ~ 1.7 – 2.1 Å, and van der Waals radii ~ 1.4 – 2.0 Å; map resolution is also reported in Å.
Asymmetric unit <i>[Ch. 1]</i>	In crystallography, the smallest portion of a crystal's unit cell that, when subjected to all symmetry operations of the space group, generates the complete unit cell; it may contain one or more protein chains.
ATP (adenosine triphosphate) <i>[General]</i>	The primary energy currency of the cell; hydrolysis of the terminal phosphate bond ($\text{ATP} \rightarrow \text{ADP} + \text{Pi}$) releases free energy that powers cellular processes including protein conformational changes, motor activity, and signaling cascades.
Auto-TAC (autophagy-targeting chimera) <i>[Ch. 4]</i>	A bifunctional small molecule—analogue to PROTACs used for proteasomal degradation—designed to simultaneously bind a disease-relevant target and the FIP200 CLAW domain, thereby routing the target to the autophagy pathway for lysosomal degradation.

Autophagosome [Ch. 4]	The double-membrane organelle generated during autophagy that engulfs cytoplasmic material and subsequently fuses with the lysosome; its formation is nucleated and driven by the sequential action of the ULK1, PtdIns3K, and ATG conjugation complexes.
B cell [General]	A lymphocyte of the adaptive immune system that develops in the bone marrow; upon antigen exposure and T-cell help, B cells differentiate into plasma cells that secrete large quantities of antibodies, or into memory B cells that persist and respond rapidly to re-exposure.
B-factor (temperature factor / atomic displacement parameter) [General]	A parameter in crystallographic and cryo-EM models that quantifies the uncertainty in each atom's position due to thermal motion or static disorder; high B-factors indicate flexible or disordered regions; the thesis uses B-factor maps to illustrate TNIP1-pFIR flexibility at the CLAW interface.
Biological unit (biological assembly) [General]	The physiologically relevant oligomeric state of a protein, which may differ from the crystallographic asymmetric unit (e.g., a dimer in solution may crystallize with one monomer per asymmetric unit); determining the biological unit requires combining crystallographic symmetry with biochemical evidence.
Bragg's law [Ch. 1]	The relationship $\lambda = 2d \sin \theta$ that describes the condition under which X-rays scattered by parallel crystal planes (spacing d) interfere constructively at angle θ for wavelength λ ; the basis for indexing diffraction spots.
Broadly neutralizing antibody (bNAb) [Ch. 2]	An antibody capable of neutralizing multiple variants or strains of a pathogen by targeting a conserved, functionally constrained epitope; Class 1/4 antibodies against the conserved RBD core are a prime example in the SARS-CoV-2 context.
Buffer [General]	A solution containing a weak acid and its conjugate base that resists changes in pH; biochemical buffers (e.g., HEPES, Tris, phosphate) maintain physiological pH during protein purification, crystallization, and assays.
Buried surface area (BSA) [General]	The solvent-accessible surface area that becomes hidden from solvent upon complex formation; BSA is a standard measure of interface size and is correlated with binding affinity. In Chapter 4, peptide binding to the CLAW reduces the BSA of the CLAW dimer interface.
Centrifugation [General]	A technique that applies centrifugal force to separate components of a mixture by density or size; used in protein purification to remove cell debris (low-speed), pellet membranes (ultracentrifugation), or separate macromolecular complexes by sedimentation.
Chaperone [General]	A protein that assists the folding, assembly, or stabilization of other proteins without being part of the final functional complex;

molecular chaperones prevent aberrant aggregation and are critical during de novo protein synthesis and stress responses.

Chromatography

[General]

A family of techniques used to separate proteins or other biomolecules based on differential interactions with a stationary phase; common types include affinity chromatography (e.g., Ni-NTA for His-tagged proteins), ion exchange, and size exclusion chromatography.

Class 1/4 antibody

[Ch. 2]

A category of anti-SARS-CoV-2 RBD antibodies (Barnes classification) that simultaneously contacts the cryptic, conserved epitope at the base of the RBD interior face (Class 4) while sterically blocking ACE2 engagement (Class 1); often characterized by a long CDRH3 loop that extends the RBD beta-sheet.

CLAW domain

[Ch. 4]

The ~12 kDa C-terminal domain of FIP200 that forms a C2-symmetric homodimer; it presents a positively charged loop (loop 4-5) on its outer face that recognizes the FIR motifs of autophagy receptor proteins, linking cargo to the autophagy initiation machinery.

Coiled coil

[General]

A structural motif in which two or more alpha helices wrap around each other in a superhelical arrangement, stabilized by heptad repeat sequences with hydrophobic residues at positions a and d; common in dimerization and scaffolding domains.

Complementarity determining region (CDR)

[General]

One of three hypervariable loops in each of the heavy and light chain variable domains of an antibody that collectively form the antigen-binding site (paratope); CDR sequences are the primary source of antibody specificity and diversity.

Complementarity determining region H3 (CDRH3)

[Ch. 2]

The third heavy-chain hypervariable loop of an antibody's variable domain; it contributes the majority of antigen-binding contacts and, due to V(D)J recombination junctional diversity, is the most variable and often the longest CDR.

Conformation / conformational change

[General]

The three-dimensional shape of a protein determined by the folding of its polypeptide chain; a conformational change is a shift between two or more structural states, often triggered by ligand binding, phosphorylation, or changes in environment, and frequently linked to changes in function.

Contrast transfer function (CTF)

[Ch. 1]

A mathematical function describing how the objective lens of an electron microscope, combined with the degree of defocus chosen during data collection, modulates the transfer of different spatial frequencies from specimen to image; CTF correction is essential in cryo-EM data processing.

Cryo-electron microscopy (cryo-EM)

[Ch. 1]

A structural technique in which purified protein is vitrified in a thin film of amorphous ice and imaged by transmission electron microscopy; two-dimensional particle projections in many

orientations are computationally averaged to reconstruct a three-dimensional electron-density map.

Cryo-EM grid (see EM grid)

[General]

See 'Electron microscopy (EM) grid'. In cryo-EM, the grid is plunge-frozen in liquid ethane after sample application to produce vitreous ice containing randomly oriented protein particles.

Cryogenic temperature

[General]

Temperatures near or below $-150\text{ }^{\circ}\text{C}$ (123 K); cryo-EM and cryo-crystallography data are collected at cryogenic temperatures (typically in liquid nitrogen or a cold gas stream) to slow radiation damage and maintain sample in vitreous ice.

Cryoprotectant

[General]

A chemical additive (e.g., glycerol, MPD, ethylene glycol) that, when included in a crystal or cryo-EM sample solution, reduces the formation of crystalline ice during flash-freezing, preserving sample integrity for data collection at cryogenic temperatures.

Crystal

[General]

A solid material in which atoms, molecules, or proteins are arranged in a highly ordered, repeating three-dimensional lattice; protein crystals contain billions of identically oriented molecules and are required for X-ray crystallography.

Defense island

[Ch. 5]

A genomic locus in bacteria or archaea that encodes a cluster of genes involved in defense against bacteriophages and mobile genetic elements; SLFN-pAgo genes are frequently found near defense islands including restriction–modification, toxin–antitoxin, and BREX systems.

Diffraction

[General]

The scattering of waves (X-rays or electrons) by a periodic array of objects (atoms in a crystal or molecules in a sample); the resulting interference pattern encodes structural information about the scattering object.

Direct electron detector (DED)

[Ch. 1]

A camera that converts each incident electron directly into an electronic signal, enabling rapid frame-rate acquisition, movie-mode recording for beam-induced motion correction, and improved signal-to-noise ratio compared with older photographic film or CCD detectors.

DNA (deoxyribonucleic acid)

[General]

The double-stranded helical polymer of deoxyribonucleotides that encodes genetic information; individual strands carry complementary base sequences (A–T and G–C pairing) and serve as templates for replication and transcription.

Domain

[General]

A structurally and evolutionarily distinct unit of a protein that folds independently, typically 50–350 residues, and often carries a discrete biochemical function (e.g., catalysis, binding, dimerization); multi-domain proteins combine multiple such units in a single polypeptide chain.

E. coli (Escherichia coli)

[General]

A gram-negative bacterium widely used as a host for recombinant protein expression; plasmid-encoded genes under inducible

	promoters (e.g., T7/IPTG) can drive high-level production of target proteins in <i>E. coli</i> cells.
Eicosanoid [Ch. 3]	A class of bioactive lipid signaling molecules derived from arachidonic acid (released by PLA2 activity), including prostaglandins, leukotrienes, and thromboxanes; they mediate inflammation, pain, and fever.
Electron density map [Ch. 1]	A three-dimensional representation of the distribution of electrons in a molecule, derived by Fourier transformation of diffraction or EM data; atomic models are built by fitting chemical structures into this density.
Electron gun [General]	The component at the top of an electron microscope column that generates the electron beam, typically from a field-emission crystal (e.g., LaB6) or a cold/Schottky field emitter, and accelerates electrons to the operating voltage (e.g., 300 keV for a Titan Krios).
Electron microscopy (EM) grid [General]	A thin metal mesh (typically copper or gold, ~3 mm diameter) coated with a thin layer of carbon or other support film containing micrometer-scale holes; purified protein solution is applied to the grid, excess liquid is blotted away, and the remaining thin film is vitrified for cryo-EM imaging.
Enzyme-linked immunosorbent assay (ELISA) [Ch. 2]	A plate-based immunoassay that uses antibody–antigen binding coupled to an enzyme reporter (e.g., horse radish peroxidase) and a colorimetric or luminescent substrate to quantify specific molecular interactions; used here in competition and indirect formats to map epitopes and measure binding breadth.
Epitope [General]	The specific region on an antigen that is recognized and bound by an antibody or T-cell receptor; conformational epitopes depend on the native 3D fold, while linear epitopes are continuous stretches of amino acids accessible in the denatured protein.
Eukaryote [General]	An organism whose cells contain a membrane-bound nucleus and membrane-bound organelles (mitochondria, ER, Golgi, etc.); includes all animals, plants, fungi, and protists. Contrasted with prokaryotes (bacteria and archaea), which lack a nucleus.
Expression / recombinant protein expression [General]	The production of a protein of interest by introducing its gene (often codon-optimized and fused to an affinity tag) into a host cell (<i>E. coli</i> , insect cells, mammalian cells) and inducing transcription and translation of that gene.
FIP200 (RB1CC1) [Ch. 4]	A large scaffold protein (also called RB1CC1) and essential component of the ULK1 autophagy initiation complex; its N-terminal domain scaffolds ULK1 and ATG13, while its C-terminal CLAW domain directly binds autophagy receptors to recruit cargo.
FIR (FIP200-interacting region)	A short peptide motif (consensus D_D_I[F/Y/W/L/V]_X_X_L[I/V]) found in autophagy receptors that mediates direct binding to the

<i>[Ch. 4]</i>	CLAW domain of FIP200; binding is often enhanced by phosphorylation of upstream serine residues.
Fluorescence polarization (FP) assay <i>[Ch. 4]</i>	A solution-based binding assay that measures the rotational correlation time of a fluorescently labeled molecule: free (small) ligands tumble faster and emit depolarized fluorescence, while bound (large) complexes tumble slowly and maintain polarization; used here to determine dissociation constants (K_d) for FIR peptide–CLAW interactions.
Fourier transform <i>[General]</i>	A mathematical operation that converts a function expressed as a sum of spatial (or temporal) contributions into its frequency components, or vice versa; in crystallography and EM, inverse Fourier transforms convert measured diffraction or frequency-space data into real-space electron density maps.
Fragment antigen binding (Fab) <i>[Ch. 2]</i>	The arm of an IgG antibody, generated by papain cleavage, that contains the variable and first constant domains of the heavy and light chains; it retains full antigen-binding activity and is preferred for structural studies due to reduced conformational flexibility compared with intact IgG.
Genome <i>[General]</i>	The complete set of genetic information (DNA, or RNA in some viruses) of an organism, including all coding and non-coding sequences; analysis of genome sequences (genomics) is used to identify and characterize defense systems, gene clusters, and evolutionary relationships.
GHKL ATPase domain <i>[Ch. 5]</i>	A member of the GHKL (Gyrase, Hsp90, Histidine Kinase, MutL) superfamily of ATPases characterized by a Bergerat ATP-binding fold; in RzaGo, ATP binding to the GHKL domain drives tetramerization of the Schlafen/ATPase supradomain (the 'Schlatpase'), whereas ATP hydrolysis is essential for antiviral activity.
Glycan / glycosylation <i>[General]</i>	A sugar chain (oligosaccharide) covalently attached to a protein (N-linked at asparagine or O-linked at serine/threonine); glycosylation affects protein folding, stability, and immune recognition; the SARS-CoV-2 Spike protein is extensively glycosylated.
Gold-standard Fourier shell correlation (GSFSC) <i>[Ch. 5]</i>	The standard metric used in cryo-EM to estimate map resolution by computing the correlation between two independently refined half-maps as a function of spatial frequency; resolution is reported at the 0.143 threshold.
Guide RNA (gRNA) <i>[Ch. 5]</i>	In the context of RzaGo, short RNA molecules that load into the pAgo core and direct the protein to complementary single-stranded DNA targets; guide loading also triggers higher-order filament assembly that suppresses Schlafen nuclease activity.
Helical filament <i>[General]</i>	A polymer in which multiple copies of a protein (protomers) assemble end-to-end with a rotational offset to generate a repeating

helical superstructure; in R2Ago, guide-RNA loading drives formation of helical filaments that suppress Schlafen nuclease activity.

Heteroduplex
[General]

A double-stranded nucleic acid formed by base-pairing of two complementary strands from different sources; in the Argonaute field, the guide RNA–target DNA heteroduplex is the key intermediate formed upon target recognition.

His-tag (hexahistidine tag)
[General]

A short sequence of six or more consecutive histidine residues appended to the N- or C-terminus of a recombinant protein; histidines coordinate Ni^{2+} or Co^{2+} ions on immobilized metal affinity chromatography (IMAC) resins, enabling rapid purification from cell lysate.

Homolog / ortholog / paralog
[General]

Genes or proteins related by evolutionary descent; orthologs are homologs in different species derived from a single ancestral gene (e.g., human PLAA and yeast Doa1); paralogs are homologs within the same species arising from gene duplication.

Hydrogen bond
[General]

A non-covalent interaction between a hydrogen-bond donor (electronegative atom covalently bonded to H, e.g., N–H or O–H) and a hydrogen-bond acceptor (lone electron pair on N or O); hydrogen bonds stabilize protein secondary structures (helices, sheets) and mediate specific protein–protein and protein–ligand interactions.

Hydrophobic effect
[General]

The thermodynamic tendency of nonpolar (hydrophobic) groups to minimize their contact with water by clustering together; the hydrophobic effect is the dominant driving force for protein folding and the burial of hydrophobic residues in protein cores and binding interfaces.

Immunoglobulin (Ig)
[General]

A globular protein fold, composed of a beta-sandwich of two antiparallel beta sheets, that is the fundamental structural unit of antibodies; each antibody chain contains multiple Ig domains, with N-terminal variable domains mediating antigen recognition and C-terminal constant domains determining isotype and effector functions.

In vitro / in vivo
[General]

Experimental contexts: in vitro (Latin: 'in glass') describes experiments performed with purified or isolated components outside of a living organism; in vivo ('in the living organism') describes experiments in intact cells or animals.

Intrinsically disordered region (IDR)
[Ch. 3]

A polypeptide segment that lacks a stable folded structure under physiological conditions but may adopt ordered conformations upon binding a partner; IDRs often mediate regulatory protein–protein interactions and are characterized by low sequence complexity and enrichment in charged and small residues.

Kinase <i>[General]</i>	An enzyme that catalyzes the transfer of a phosphate group from ATP to a substrate (protein, lipid, or small molecule), thereby regulating the substrate's activity, localization, or interactions; serine/threonine kinases (e.g., ULK1, mTORC1) are prominent regulators in the autophagy pathway.
Kosmotrope / chaotrope <i>[Ch. 1]</i>	Reagents that modulate protein solubility in crystallization solutions: kosmotropes (e.g., ammonium sulfate, PEG) stabilize the solvent structure and precipitate protein, while chaotropes (e.g., urea, certain salts) disrupt solvent structure and increase protein solubility.
Ligand <i>[General]</i>	Any molecule that binds specifically to a target protein; ligands include substrates, inhibitors, cofactors, peptides, nucleic acids, and small-molecule drugs. In structural biology, 'ligand' often refers specifically to small molecules observed bound in a crystal or cryo-EM structure.
LIR (LC3-interacting region) <i>[Ch. 4]</i>	A short peptide motif (consensus F[W/Y]_X_X_V[L/I]) found in autophagy receptors and adaptors that binds ATG8-family proteins (e.g., LC3, GABARAP) lipidated on the phagophore membrane; FIR sequences share structural similarity to LIRs.
Lysosome <i>[General]</i>	A membrane-bound organelle containing acid hydrolases (proteases, lipases, nucleases) that degrade macromolecules delivered by autophagy, endocytosis, or phagocytosis; the terminal degradative compartment of both the autophagy and endosomal sorting pathways.
Macroautophagy (autophagy) <i>[Ch. 4]</i>	A conserved intracellular degradation pathway in which a double-membrane structure (the autophagosome) engulfs cytoplasmic cargo—organelles, protein aggregates, or pathogens—and delivers it to the lysosome for proteolytic breakdown and recycling of macromolecular components.
Macromolecule <i>[General]</i>	A large biological polymer, including proteins, nucleic acids (DNA, RNA), polysaccharides, and lipid assemblies; structural biology focuses primarily on proteins and nucleic acids and their complexes.
Map (electron density / cryo-EM map) <i>[General]</i>	A three-dimensional volumetric representation of the distribution of electron density in a molecule, obtained by Fourier transformation of crystallographic structure factors or by back-projection and refinement of cryo-EM particle images; atomic models are built into maps.
Membrane protein <i>[General]</i>	A protein that is fully or partially embedded in, or stably associated with, a lipid bilayer; integral membrane proteins often require detergents for solubilization and present special challenges for structural studies because their hydrophobic transmembrane regions are unstable in aqueous solution.
Micrograph <i>[General]</i>	An image of a specimen recorded with a microscope; in cryo-EM, a micrograph is a single frame or motion-corrected movie recorded

from a thin vitrified sample, containing many projections of individual protein particles.

Model building / model refinement
[General]

The iterative process of fitting an atomic model (amino acid sequence + 3D coordinates) into an electron-density or cryo-EM map and adjusting bond lengths, angles, and side-chain rotamers to maximize agreement with experimental data while maintaining good stereochemistry.

Model statistics (R-factor, Ramachandran, clashscore)
[General]

Quantitative metrics that assess the quality of an atomic model: the R-factor measures agreement between observed and calculated diffraction data; the Ramachandran plot evaluates backbone dihedral angles (favored regions >95% is typical); the clashscore counts steric overlaps between non-bonded atoms.

Molecular replacement
[Ch. 1]

A phasing method in which a known structural model (experimental or AlphaFold-predicted) is positioned in the unit cell to provide initial phase estimates, enabling computation of an electron-density map for the unknown structure.

Molecular weight (MW) / kilodalton (kDa)
[General]

The mass of a molecule; protein masses are typically reported in kilodaltons (kDa; 1 kDa = 1000 Da = ~1000 g/mol). As a rough guide, 100 amino acids ≈ 11 kDa. The relative size of a protein strongly influences which structural technique is most appropriate.

Monoclonal antibody (mAb)
[General]

An antibody produced by a single clonal B-cell lineage and therefore possessing a single defined antigen-binding specificity; used therapeutically and as research tools; distinguished from polyclonal antibodies, which are mixtures from multiple B-cell clones.

NDMSBA (Neurodevelopmental disorder with progressive microcephaly, spasticity, and brain anomalies)
[Ch. 3]

A rare autosomal recessive disorder caused by homozygous or compound heterozygous loss-of-function mutations in PLAA; characterized by infantile-onset neurological symptoms including microcephaly, spasticity, and severe intellectual and motor impairment.

Negative-stain electron microscopy
[Ch. 1]

An EM sample-preparation method in which a biological sample is coated with a heavy-metal salt (e.g., uranyl formate), providing high-contrast images at room temperature; resolution is limited to ~15–30 Å but the technique is useful for rapid assessment of sample quality and particle size.

NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells)
[General]

A family of transcription factors that regulate immune and inflammatory gene expression in response to cytokines, pathogen-associated patterns, and cellular stress; TNIP1 is a negative regulator of NF-κB activation pathways.

Nuclear magnetic resonance (NMR) spectroscopy

A solution-based structural technique that exploits the magnetic properties of atomic nuclei (^1H , ^{13}C , ^{15}N , ^{31}P) to determine atomic

<i>[Ch. 1]</i>	coordinates and molecular dynamics; most applicable to proteins below ~40 kDa.
Nucleic acid <i>[General]</i>	A polymer of nucleotides (DNA or RNA) that stores and transmits genetic information; in structural biology, nucleic acids are frequently co-crystallized or co-imaged with protein binding partners to reveal recognition mechanisms.
Oligomer / homooligomer <i>[General]</i>	A complex composed of a defined small number of identical (homooligomer) or non-identical (heterooligomer) subunits (protomers); examples include dimers (2 subunits), tetramers (4), hexamers (6); higher-order oligomeric states of RzAgo are central to its regulatory mechanism.
p97 / VCP / Cdc48 <i>[Ch. 3]</i>	A highly abundant, essential AAA+ ATPase that forms homo-hexameric rings and uses ATP hydrolysis to unfold or extract ubiquitinated proteins from membranes and macromolecular complexes, targeting them for proteasomal degradation or other fates; known as p97 in mammals, VCP (valosin-containing protein) in biochemistry literature, and Cdc48 in yeast.
Paratope <i>[General]</i>	The antigen-binding site of an antibody, formed by the six CDR loops (three from the heavy chain, three from the light chain) that make direct contact with the epitope; the paratope is the structural complement to the epitope.
Particle picking <i>[General]</i>	The computational step in cryo-EM processing in which individual protein particles are identified and extracted from raw micrographs, either by template matching, reference-free blob-picking, or machine-learning-based detectors; the quality of particle picking strongly influences downstream map quality.
pFIR (phosphorylated FIR) <i>[Ch. 4]</i>	A FIR peptide in which one or more upstream serine (or threonine) residues have been phosphorylated; phosphorylation dramatically enhances electrostatic complementarity with the positively charged CLAW loop 4-5 and can increase binding affinity by an order of magnitude or more.
PFU domain (PLAA family ubiquitin-binding domain) <i>[Ch. 3]</i>	A structurally unique alpha-helical domain found exclusively in PLAA orthologs that binds ubiquitin via recognition of the hydrophobic I44 patch through a conserved helical motif (FIIDNT in humans); essential for PLAA function in concert with the WD-40 domain.
Phage (bacteriophage) <i>[General]</i>	A virus that infects bacteria; bacteriophages are the primary selective pressure driving evolution of prokaryotic immune systems such as CRISPR-Cas and the Argonaute-based defense systems studied in Chapter 5.
Phagophore <i>[Ch. 4]</i>	The nascent, cup-shaped precursor membrane that elongates around autophagic cargo to form the completed double-membrane

autophagosome; LC3/ATG8 proteins are lipidated onto the phagophore membrane during this process.

Phase problem

[Ch. 1]

The fundamental challenge in X-ray crystallography: detectors record only the intensities (amplitudes squared) of diffracted X-rays, losing the phase information required for an inverse Fourier transform; phases must be recovered by methods such as molecular replacement or heavy-atom derivatization.

Phospholipase A2 (PLA2)

[Ch. 3]

A family of enzymes (EC 3.1.1.4) that hydrolyze the sn-2 ester bond of membrane phospholipids, releasing arachidonic acid (a prostaglandin and leukotriene precursor) and lysophospholipids; their activity is modulated by PLAA and plays roles in inflammation and membrane remodeling.

Phosphorylation

[General]

The addition of a phosphate group to a serine, threonine, or tyrosine residue by a kinase; phosphorylation is one of the most common post-translational modifications and can alter protein conformation, activity, binding affinity, or subcellular localization.

PIWI domain

[Ch. 5]

The catalytic domain shared by all Argonaute proteins that cleaves target nucleic acids in a guide-directed manner via a conserved catalytic tetrad (DEDX); long-B pAgos such as RzAgos lack a functional catalytic tetrad and are nuclease-inactive.

PLAA (Phospholipase A2 Activating Protein)

[Ch. 3]

A highly conserved, essential eukaryotic protein (yeast ortholog Doa1/Ufd3) composed of WD-40, PFU, IDR, and PUL domains; functions as a p97 adaptor protein linking ubiquitinated substrates to the ubiquitin-proteasome system, and also activates phospholipase A2 enzymes; mutations cause severe neurodevelopmental disorders.

Pleiotropy

[Ch. 3]

The phenomenon in which a single gene product is associated with multiple distinct cellular functions or processes; type I (moonlighting) refers to multiple molecular functions from a single polypeptide; type II refers to one biochemical activity linked to several distinct processes.

Point spread function (PSF)

[General]

A function describing how a single point source is blurred by an imaging system; in cryo-EM, a smaller PSF (achieved with thin detectors and direct electron detection) means each detected electron is assigned a more accurate position, improving image resolution.

Polypeptide / protein

[General]

A polypeptide is a linear chain of amino acids linked by peptide bonds; a protein is a polypeptide (or set of polypeptides) that has folded into a stable three-dimensional structure conferring specific biochemical function.

Post-translational modification (PTM)

[General]

A chemical modification to a protein occurring after translation, including phosphorylation, ubiquitination, glycosylation, acetylation, palmitoylation, and many others; PTMs regulate protein activity, stability, localization, and interactions.

Prokaryote <i>[General]</i>	An organism lacking a membrane-bound nucleus and other membrane-enclosed organelles; includes bacteria and archaea. Prokaryotes are the hosts and evolvers of the Argonaute-based defense system studied in Chapter 5.
Prokaryotic Argonaute (pAgo) <i>[Ch. 5]</i>	A member of the Argonaute protein superfamily found in bacteria and archaea; pAgos use short nucleic-acid guide strands (DNA or RNA) to recognize and cleave complementary target nucleic acids and can serve as antiphage immune effectors.
PROTAC (proteolysis-targeting chimera) <i>[Ch. 4]</i>	A bifunctional small molecule that simultaneously binds a target protein and an E3 ubiquitin ligase, inducing ubiquitination and subsequent proteasomal degradation of the target; the Auto-TAC concept adapts this principle to autophagy via FIP200 CLAW recruitment.
Proteasome <i>[General]</i>	A large, barrel-shaped multi-subunit protease complex that degrades ubiquitinated and misfolded proteins; p97 extracts substrates from membranes or complexes and delivers them to the proteasome for destruction.
Protein complex <i>[General]</i>	A stable or transient assembly of two or more protein subunits (or a protein with a nucleic acid); structural biology of complexes reveals how subunits interact and cooperate to carry out biological functions.
Protein Data Bank (PDB) <i>[Ch. 1]</i>	The global repository for experimentally determined three-dimensional structures of biological macromolecules; all structures discussed in this thesis were deposited in or referenced from the PDB.
Protein purification <i>[General]</i>	The sequential use of chromatographic and other separation techniques to isolate a protein of interest from the complex mixture of molecules in a cell lysate, achieving sufficient purity and concentration for structural or biochemical studies.
PUL domain (PLAP/Ufd3p/Lub1p domain) <i>[Ch. 3]</i>	The C-terminal alpha-solenoid domain of PLAA, formed from five Armadillo repeats, responsible for recognizing the C-terminal tail of p97 (specifically p97 Tyr805), thereby docking PLAA onto the D2 face of the p97 hexamer.
Radiation damage <i>[General]</i>	The structural deterioration of a biological sample caused by energy deposition from the X-ray or electron beam during data collection; minimized in crystallography by cryocooling crystals and in cryo-EM by limiting total electron dose and collecting data in a movie format.
Real space vs. reciprocal space <i>[General]</i>	Two mathematically related representations of structural data: real space describes atomic positions in physical (x, y, z) coordinates; reciprocal space (Fourier space) encodes the spatial frequencies of those positions. Diffraction data are collected in reciprocal space and converted to real space by Fourier transform.

Receptor binding domain (RBD) [Ch. 2]	The subdomain within the S1 subunit of the SARS-CoV-2 Spike protein that directly contacts the host receptor ACE2; it consists of a conserved core beta-sheet scaffold and a variable receptor-binding motif (RBM).
Receptor binding motif (RBM) [Ch. 2]	The long loop within the RBD that forms the contact interface with ACE2; it is the most variable region of the Spike protein and the target of many potentially neutralizing—but strain-specific—antibodies.
Reciprocal space [Ch. 1]	The mathematical space (Fourier space) in which diffraction data are recorded; spot positions and intensities encode spatial frequencies of the scattering object, and an inverse Fourier transform converts this information back to real-space electron density.
Refinement [General]	The iterative computational optimization of an atomic model's parameters (atomic coordinates, B-factors, occupancies) to maximize agreement with experimental data (X-ray structure factors or cryo-EM map density) while applying restraints on stereochemistry.
Resolution [Ch. 1]	A measure of the level of detail resolvable in a structural map, typically reported in ångströms (Å); lower numerical values indicate finer detail (e.g., 2 Å resolves individual atoms, while 6 Å resolves only secondary-structure elements).
RNA (ribonucleic acid) [General]	A single-stranded polymer of ribonucleotides that serves diverse cellular roles including messenger RNA (mRNA, carries genetic code), ribosomal RNA (rRNA, structural/catalytic component of ribosomes), transfer RNA (tRNA, amino acid adapter), and regulatory small RNAs (siRNA, miRNA, piRNA).
RzAgo [Ch. 5]	The extra-long prokaryotic Argonaute protein from <i>Runella zeae</i> studied in Chapter 5; it possesses an N-terminal Schlafen nuclease domain and a GHKL ATPase domain in addition to the canonical pAgo core (MID, PAZ, PIWI domains), lacks the N-terminal domain, and forms ATP-driven tetramers and guide-RNA-driven filaments.
Salt bridge [General]	A non-covalent electrostatic interaction between an oppositely charged pair of ionizable side chains (e.g., Asp/Glu and Arg/Lys) within a protein or at a protein–protein interface; salt bridges contribute to structural stability and binding specificity.
Sarbecovirus [Ch. 2]	The subgenus of betacoronaviruses that includes SARS-CoV-2, SARS-CoV, and related bat coronaviruses; broadly neutralizing vaccines and antibodies aim to protect against all four sarbecovirus clades.
SARS-CoV-2 [Ch. 2]	Severe Acute Respiratory Syndrome Coronavirus 2; the enveloped, positive-sense single-stranded RNA betacoronavirus responsible for the COVID-19 pandemic that emerged in late 2019.

Schlafen (SLFN) domain [Ch. 5]	A broadly distributed nuclease domain found in organisms across all domains of life; in mammals, Schlafen proteins restrict viral replication by cleaving tRNA or rRNA; in RzAgo, the SLFN domain functions as the catalytic effector that is activated upon tetramerization.
Schlatpase [Ch. 5]	The stable supradomain combination of a Schlafen nuclease fused to a GHKL ATPase; designated by the authors to describe this evolutionarily conserved module found in SLFN-pAgos and many other prokaryotic defense proteins.
SDS-PAGE (sodium dodecyl sulfate–polyacrylamide gel electrophoresis) [General]	A technique for separating proteins by molecular weight; proteins are denatured and coated with the negatively charged detergent SDS, then separated in a polyacrylamide gel by an electric field; used routinely to verify protein identity, purity, and size during purification.
Secondary structure [General]	Local regular structural motifs in a polypeptide chain stabilized by hydrogen bonds between backbone atoms; the two most common elements are the alpha helix (backbone H-bonds within a single chain) and the beta sheet (H-bonds between extended strands, either parallel or antiparallel).
Signal-to-noise ratio (SNR) [General]	The ratio of the true signal amplitude to background noise; low SNR is a fundamental challenge in both cryo-EM (biological molecules scatter electrons weakly) and X-ray crystallography, and is addressed by averaging many identical particles or unit cells.
Single-particle analysis (SPA) [Ch. 1]	The cryo-EM workflow in which thousands of individual particle projections, each representing a randomly oriented molecule, are aligned and averaged to produce a three-dimensional reconstruction without requiring crystalline order.
Size exclusion chromatography (SEC) [General]	A chromatographic technique that separates proteins by hydrodynamic size as they pass through a porous resin; larger molecules elute earlier (excluded from pores) and smaller molecules later; used both as a purification step and to assess oligomeric state and sample homogeneity.
Small-angle X-ray scattering (SAXS) [Ch. 1]	A solution-based diffraction method that measures the scattering of X-rays from molecules in solution to yield low-resolution (~ 10 Å) information about overall shape, size, and conformational states without requiring crystallization.
Somatic hypermutation (SHM) [Ch. 2]	The process by which activated B cells introduce point mutations throughout immunoglobulin variable regions during an adaptive immune response; mutations that improve antigen binding are selected through affinity maturation, diversifying the antibody repertoire beyond germline sequences.
Space group [Ch. 1]	The complete description of the symmetry of a crystal, combining the translational periodicity of the Bravais lattice with point-group

rotational and screw-axis symmetry operations; protein crystals occupy 65 of the 230 possible space groups (the chiral, or Sohncke, groups).

Spike (S) protein
[Ch. 2]

The trimeric class I fusion glycoprotein displayed on the SARS-CoV-2 virion surface; it comprises the receptor-binding subunit S1 and the membrane-fusion subunit S2, and is the primary target of neutralizing antibodies and vaccines.

Stoichiometry
[General]

The molar ratio of subunits or components in a protein complex; determining stoichiometry (e.g., 2:1:1 for a trimeric complex) is essential for understanding assembly and function.

Structural biology
[Ch. 1]

The scientific discipline concerned with determining the three-dimensional arrangement of atoms in biological macromolecules (proteins, nucleic acids, lipids, carbohydrates) and their assemblies, with the goal of understanding how molecular structure governs biological function.

Symmetry expansion
[General]

A cryo-EM data-processing strategy in which particles with known internal symmetry are 'expanded' so that each symmetry-related copy is treated as an independent particle, allowing sub-particle classification and reconstruction of asymmetric features within a symmetric complex.

Synchrotron
[General]

A large particle accelerator that produces highly intense, tunable X-ray beams used for crystallographic data collection; synchrotron beamlines (e.g., SSRL beamline 12-1, used in Ch. 4) provide far brighter and more collimated X-ray beams than laboratory sources, enabling data collection from small or weakly diffracting crystals.

Tertiary structure
[General]

The overall three-dimensional fold of a single polypeptide chain, resulting from packing of secondary structure elements (helices, sheets, loops) stabilized by hydrophobic interactions, hydrogen bonds, salt bridges, and disulfide bonds.

TNIP1 (TNFAIP3-interacting protein 1 / ABIN-1)
[Ch. 4]

A multidomain protein that inhibits NF- κ B activation downstream of cytokines and TLRs by scaffolding the A20 deubiquitinase; also binds ATG8 proteins and the FIP200 CLAW domain via phosphorylated FIR/LIR motifs, positioning it as a selective autophagy regulator with implications for autoimmune disease.

Transcription / translation
[General]

The two steps of gene expression: transcription is the synthesis of RNA from a DNA template by RNA polymerase; translation is the ribosome-mediated synthesis of a polypeptide chain from an mRNA template, using tRNA adaptors to decode each codon.

Transmission electron microscope (TEM)
[General]

An electron microscope in which an electron beam passes through a thin specimen and is focused by electromagnetic lenses to form a magnified image or diffraction pattern on a detector; both cryo-EM and negative-stain EM use TEMs.

Ubiquitin*[Ch. 3]*

A highly conserved 76-residue protein that is covalently conjugated to lysine residues of substrate proteins, targeting them for diverse fates including proteasomal degradation (K48-linked chains), DNA repair, endocytosis (K63-linked chains), and innate immunity depending on chain linkage type.

ULK1 complex*[Ch. 4]*

The master autophagy-initiating kinase complex composed of ULK1/2, ATG13, ATG101, and FIP200; it is activated by AMPK under nutrient-starvation and inhibited by mTORC1 under nutrient-replete conditions, serving as the switch that initiates autophagosome formation.

Unit cell*[Ch. 1]*

The smallest repeating parallelepiped unit of a crystal lattice, defined by three edge lengths (a , b , c) and three inter-edge angles (α , β , γ), whose stacking in three dimensions reconstructs the entire crystal.

Virus / virion*[General]*

A non-living infectious agent consisting of nucleic acid (DNA or RNA) enclosed in a protein coat (capsid), sometimes surrounded by a lipid envelope; a virion is the mature, extracellular, infectious particle. SARS-CoV-2 is an enveloped RNA virus.

Vitrification*[Ch. 1]*

The rapid freezing of a thin aqueous sample in liquid ethane such that water solidifies into an amorphous (glassy) state rather than crystalline ice, preserving the native hydration environment and particle orientations for cryo-EM imaging.

WD-40 domain*[Ch. 3]*

One of the most common structural folds in eukaryotes, formed from ~40–60 residue repeating units ending in a tryptophan-aspartate dipeptide that fold into a seven-to-eight-bladed beta-propeller; frequently serves as a scaffold for protein-protein, protein-peptide, or protein-nucleic acid interactions.

X-ray crystallography*[Ch. 1]*

A structural technique in which a protein crystal is exposed to a coherent X-ray beam; the resulting diffraction pattern is used to compute an electron-density map and build an atomic model of the molecule.

YYDRxG motif*[Ch. 2]*

A conserved sequence motif found in the CDRH3 loops of a subset of broadly neutralizing Class 1/4 antibodies (including 11G-12v), derived from the IGHD3-22 D-gene reading frame; the motif forms a beta-hairpin that extends the RBD core beta-sheet and positions the light chain to sterically block ACE2.

STRUCTURAL BIOLOGY: HISTORY, METHODS, AND APPLICATIONS WITHIN THIS THESIS

1. Introduction

Structural biology is the study of the three-dimensional organization of atoms in biological macromolecules—proteins, nucleic acids, lipids, carbohydrates, and assemblages thereof. Its central tenant holds that, like any chemical, biological molecules derive their functions directly from their structures, and therefore structural information is essential for understanding biological processes at their most fundamental level. From small molecules and the domains they interact with to massive, heterogeneous macromolecular complexes, structural methods provide insight across many molecular scales. This diversity is matched by an equally broad spectrum of experimental methodologies, each with its own range of interrogatable sizes and commonly achieved resolutions. The choice of technique is inseparable from the biological question. Therefore, for those who choose to specialize in structural biology and wish to answer questions spanning many biological scales, expertise across multiple structural methods is a desirable trait.

The field comprises a set of complementary experimental techniques. Nuclear magnetic resonance (NMR) spectroscopy, small-angle X-ray scattering (SAXS), and other solution-based methods are powerful for probing dynamics and conformational ensembles. X-ray crystallography and single-particle electron cryo-microscopy (SPA cryo-EM or just cryo-EM) permit the determination of static structures, and together can obtain structures across molecular masses spanning several orders of magnitude. Though these and other structural methods differ dramatically in theory, they can all be used to obtain real-space maps of atomic density. Their strengths and limitations shape the kinds of biological questions they can address, and an understanding of their underlying principles is crucial for interpreting structural results.

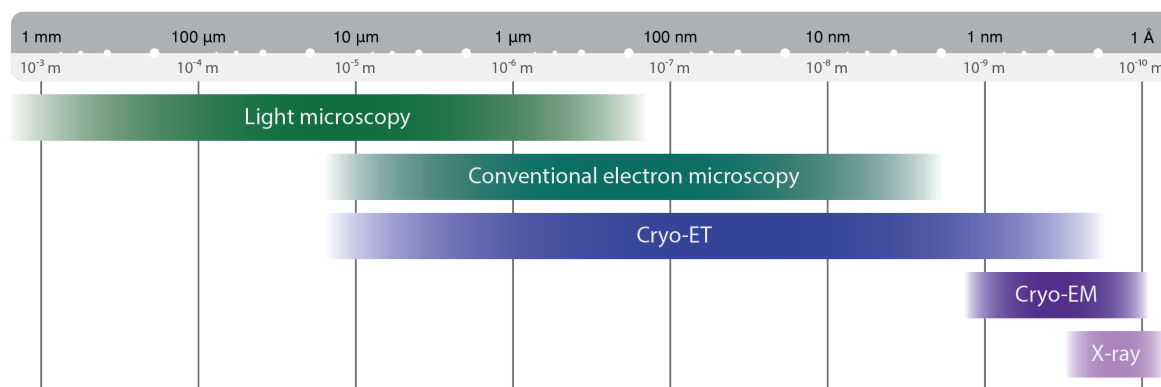


Figure 1. Size ranges of typical features explored with visualization and structural methods.

Structural biology has its origins in the early 20th century, with the discovery that the diffraction pattern from x-rays passing through a crystal could be used to determine the positions of atoms within that lattice (Bragg & Bragg, 1913). X-ray crystallographic methods were first applied to minerals and other small molecule chemical compounds, which often crystallize readily. However, in 1934, John Bernal and Dorothy Crowfoot (later Hodgkin) captured a clear X-ray diffraction pattern from pepsin crystals formed in a sealed capillary, which had protected them from dehydration (Bernal & Crowfoot, 1934). This discovery not only was foundational for protein crystallography, but also settled a long debate on the physical nature of proteins, revealing that, rather than a heterogeneous aggregate or a colloidal material, proteins generally have a set of clearly defined conformations. This observation was key to developing the field of structural biology. Crystallography remains the dominant structural technique to this day, and details on its theory and practice can be found in Section 2. Other diffraction techniques like small-angle X-ray/neutron scattering (SAXS/SANS) were also developed to measure the diffraction of molecules in solution, which does not require crystal formation but is limited to ~ 10 Å resolution for *de novo* structures (Gräwert & Svergun, 2020).

NMR would prove a higher resolution technique to obtain atomic structures of smaller proteins in solution (Figure 2). Developed through the 1930's and 40's, the first protein

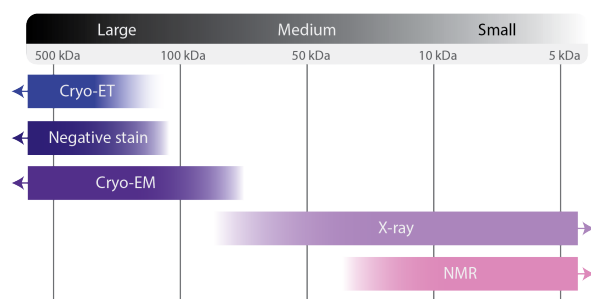


Figure 2. Molecular masses of proteinaceous features and the techniques most appropriate for capturing them.

structure was solved by NMR in 1957 (Barrett et al., 2013; Giunta & Mainz, 2020). In NMR, a magnetic field is applied to proteins in solution to align the magnetic moments of atomic nuclei, then pulses of radio-frequency electromagnetic radiation at a right angle to the external field are applied to the sample. A nucleus with

a magnetic spin, such as ^1H , ^{13}C , ^{15}N or ^{31}P , will absorb the frequency matching its spin frequency, converting its spin state. When irradiation stops, the nuclei relax back to equilibrium, and the released energy is detected. The absorbed frequencies depend not only on the isotope, but also its chemical environment, and thus these measurements provide specific information about the proximity of a given nucleus to neighboring nuclei. This information can be used to determine atomic coordinates. Solution-state NMR is a powerful tool for investigating molecular interactions and conformational dynamics, as crystal formation, often a serious bottleneck for protein crystallography, is not required. However, due in part to the need for molecules to tumble quickly in solution and to the fact that the rate of nuclear relaxation, which is inversely related to the strength of the signal, increases as molecule size increases, solution-state NMR is not as useful for proteins over 40 kDa (Liang & Tamm, 2016). Despite this utility, NMR remains a relatively uncommon method of protein structure determination, representing ~6% of total structures deposited in the Protein Data Bank (PDB) and only 235 of the 17,631 structures released in 2025 (compared to 10,087 crystal structures and 7,240 cryo-EM structures).

Electron microscopy (EM), developed in 1931 by Max Knoll and Ernst Ruska, was not initially thought to be appropriate for biological samples. First, beam damage at doses thought to be necessary for a suitable signal-to-noise ratio (SNR) or contrast led to the destruction of structural information. Second, no detector was yet sensitive enough to preserve high-resolution information. And third, biological samples have a low contrast due

to their atomic composition, and the smaller the feature, the more profound this lack of contrast becomes. Despite this, the allure of studying the intricate interiors of living systems too small for investigation by light microscopy was powerful, and in the decades closely following the advent of EM, Ladislaus Marton and others developed techniques that allowed biological samples to be imaged, such as sample cooling or the addition of heavy metal stains (negative staining). Scientists like June Almeida used these techniques to document and research biological material previously too small to see, such as viruses (Almeida & Tyrrell, 1967). In 1968, David DeRosier and Aaron Klug obtained the first 3D EM structure from 2D projections using negative stain, revealing the helically symmetrical T4 bacteriophage tail to low (~ 35 Å) resolution (De Rosier & Klug, 1968). Computer averaging techniques were developed in the 1970's, enabling the averaging of randomly oriented particles (Frank et al., 1978). For unstained samples, however, researchers continued to use crystalline protein, for which cryoprotection protocols had been developed and which also provided enough copies of the protein in question to amplify the signal, allowing maps to be obtained. In 1990, Richard Henderson published the first high-resolution cryo-EM structure, which required a crystalline sample (Henderson et al., 1990). Despite the success, Henderson envisioned improvements to the technique that could eventually permit structure determination from non-crystalline molecular assemblies of suitably large targets (> 20 or > 50 kDa, depending on the analysis) from far fewer particles than typically used in X-ray crystallography ($\sim 10^4$ rather than $\sim 10^{11+}$) (Glaeser, 1999; Henderson, 1995; Henderson et al., 1990). Simultaneously, Jacques Dubochet and his team were developing methods to vitrify samples, preserving the aqueous environment of biological particles (Adrian et al., 1984). This was critical for permitting non-crystalline samples to be analyzed, as it could capture particles suspended in random orientations, revealing 2D projections representing every angle of the molecule that could be assembled into a 3D structure. However, physical limitations in electron detectors, as well as ongoing developments in image processing to parse heterogeneous data, remained a challenge to obtaining high-resolution EM structures (Scheres, 2012a). Computational developments that enabled the handling of large datasets and the execution of complex processes such as motion correction, coupled with the introduction of a suitable direct electron detector (DED), the Gatan K2 Summit DED camera,

in 2013 enabled the “resolution revolution”, demonstrated with structures such as the TRVP1 channel at 3.4 Å (Liao et al., 2013). Thanks to these technical advances in cryo-EM, its productivity in biological structural discovery now significantly surpasses that of other methods. Since then, single-particle cryo-EM has democratized, providing a comparatively fast and accessible method to determine the structures of novel biological assemblages. An outline of the theory and practice of this technique can be found in Section 3.

Despite these technological advances, structural biology remains a challenging field. Many of these challenges arise from the physics of measurement itself. Biological macromolecules are composed primarily of weakly scattering elements, and thus both X-ray and unstained EM data often have a low SNR. To increase the apparent signal from the particle of interest, signal from tens of thousands to millions of, ideally, identical molecules are averaged to extract usable structural information. Heterogeneity, either conformational or compositional, can therefore complicate data analysis and prevent high-resolution structure determination. Preparation of the sample for certain methods can be fraught, and often extensive optimization is required. Data collection is also hazardous to the sample itself, as radiation damage by both X-rays and electrons can destroy the particle of interest and alter the medium in which it is being measured. Additionally, many biologically important proteins have features such as unstructured regions or membrane-embedded domains that complicate protein purification and preparation or defy particle averaging. Even for proteins that are tractable, the techniques themselves can introduce artefacts: crystallization can impose non-physiological lattice contacts, and cryo-EM can expose proteins to an air-water interface, causing proteins to unfold or orient non-randomly in the ice, complicating data processing. Despite these challenges, experimental structures continue to provide essential information about biological systems. Though deep learning algorithms like AlphaFold can produce high-quality hypotheses for globular domains and single proteins and can sometimes accurately model intermolecular interactions, they are still limited in their predictive power, particularly for small molecule interactor development and for investigating protein complexes (Jumper et al., 2021). Experimental structural biology therefore remains crucial to understanding the molecules of life and their organization.

2. X-ray Crystallography

2.1. *Overview*

X-ray crystallography has the longest history and remains the dominant approach for determining the atomic structures of proteins. To create crystals, purified protein is combined with a solution containing solubilization and precipitation agents and forms an ordered crystal lattice in which billions of identical molecules are arrayed in a regular three-dimensional pattern. When a crystal is illuminated with X-rays, the X-rays are scattered by the electron clouds of periodically arranged atomic centers, producing an interference pattern in reciprocal (Fourier) space. With the addition of phase information, such as, for native samples, a putative model of the crystallized protein, the spot positions and intensities can be used to generate a real-space electron density map.

Crystallography has been extraordinarily successful in producing high-resolution structures, often at data quality surpassing that achievable by cryo-EM for proteins below ~100 kDa: though there is no theoretical size limit to what can be resolved, larger proteins and complexes can be immensely challenging for crystallization, and large proteins can pose challenges to collecting interpretable diffraction patterns, requiring longer exposure times and other accommodations (Smyth & Martin, 2000). It remains the method of choice for many < 100 kDa soluble proteins, enzymes, and domains, as well as for high-throughput drug design pipelines that require atomic precision for ligand placement. It is well suited to projects where homogeneity and stability are achievable, and it provides the highest level of detail when crystals diffract well.

2.2. *Theory*

Crystals are three-dimensional repeating arrays of a given compositional unit, forming a lattice. Depending on how molecules are packed in a crystalline, the repetitive molecular units can be translated, rotated, or screwed, forming diverse symmetrical space groups. The parameters emerging from these translations and rotations determine the space group, or

geometric class, to which a crystal belongs. In protein crystallography, there are three compositional units of interest. The asymmetric unit is the smallest portion of a protein structure which, when subjected to symmetry operations, can be used to devise the unit cell. The unit cell is defined as the smallest repeating unit of the crystal, not essentially containing only one asymmetric unit. Finally, the biological unit is the biologically relevant protein organization.

Crystals form through a process of phase separation of a thermodynamically non-ideal system that is controlled primarily by kinetic parameters. Parameters that contribute substantially to thermodynamic properties of the system, such as temperature, pressure, and protein concentration, will all play a role in the crystallization of the protein of interest, but are not predictive of crystallization. Instead, protein crystal formation depends heavily on kinetic factors such as nucleation and crystal growth mechanisms, which relate to the

interaction between the protein and its crystallization solution.

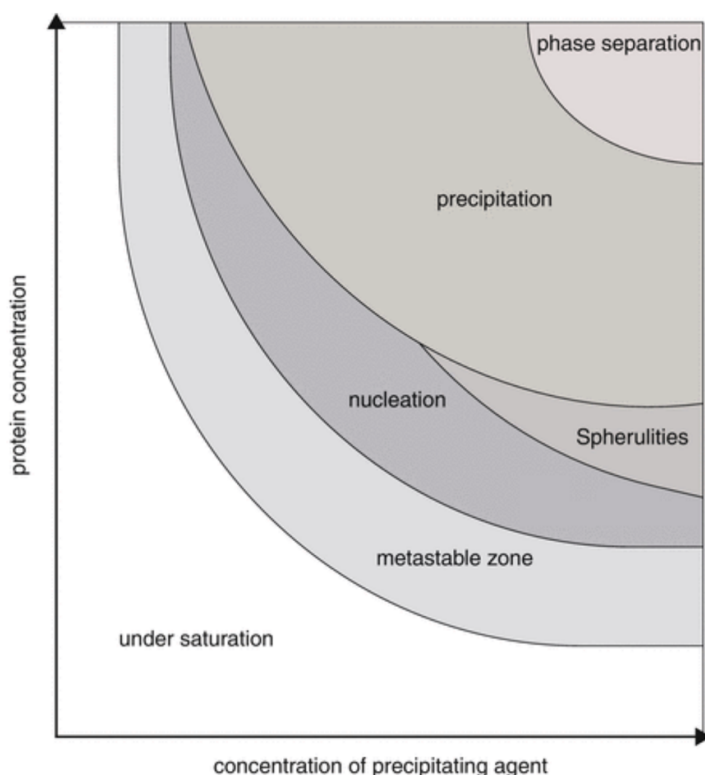


Figure 3. Phase separation diagram of proteins in a crystallization solution, from Ranatunga et al., 2014.

The phase separation diagram of proteins in a given crystallization solution is shown to the left (Figure 3, from Ranatunga et al., 2014). In this phase diagram, the two major components of the system's behavior are the protein's concentration and the concentration of precipitating agents, which make the protein less

soluble in the crystallization solution. Crystal nucleation – where crystals can start to grow – only occurs within the nucleation range. Crystal growth, however, can occur in both the nucleation and metastable zones. At low concentrations of either protein or precipitant, no phase separation will occur (undersaturation). Protein precipitation occurs when the precipitant and/or protein concentration is/are too high, and can represent either the formation of microcrystals, spherulites, and other ordered forms that could indicate a promising crystallization condition, or can be the result of protein aggregation or denaturation.

However, crystallization solutions are not made out of just protein, precipitant, and water. These solutions contain both precipitation and solubilization agents – kosmotropes and chaotropes, respectively (Cray et al., 2013). These reagents work together to modulate protein solubility and thus the phase separation process. Not all reagents are created equal, varying in the degree to which they cause the precipitation or solubilization of a given protein, and so an appropriate crystallization condition will optimize the choice of reagents.

Additionally, the system of protein with crystallization solution is not closed: the rate and extent of water evaporation from the system has a dramatic effect on crystal formation, and is crucial to the crystallization process. The system begins in an undersaturated state, as protein is generally contained in an aqueous buffer and thus the crystallization reagents are diluted upon mixing with protein solution. Evaporation therefore needs to occur before

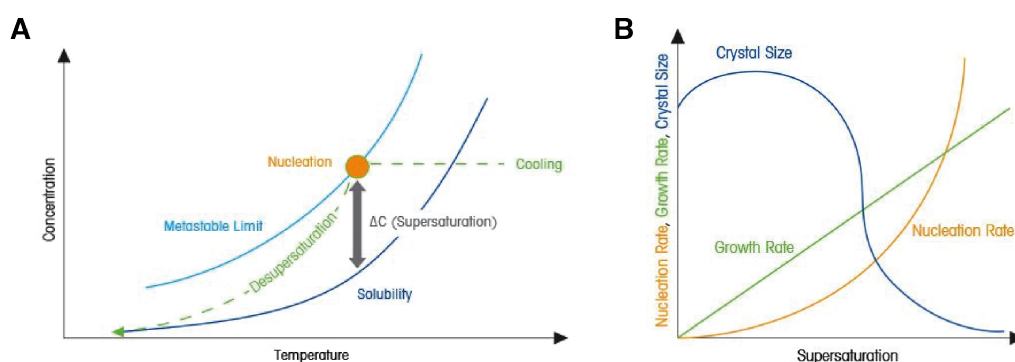


Figure 4. A. Effects of temperature and concentration on a sample in a crystallization solution. B. Effect of supersaturation on the nucleation rate, growth rate and crystal size. Adapted from Mettler Toledo.

supersaturation can be achieved. The rate of evaporation can be modulated through lowering or raising the temperature, changing the ratio of protein to crystallization condition, and changing the volume of the crystal-forming solution, among other things. Changing the rate of water evaporation changes how the protein solution progresses through the above phase diagram. Typically, slowing this process is beneficial: it can allow bigger and more well-formed crystals to grow. This is one reason it can be advantageous to grow crystals at a lower temperature (e.g., 4°C as opposed to standard room temperature (RT), around 20°C). However, this is not the case for all systems.

Temperature also affects the solubility of both the protein and reagents in your solution. The above left curve (Figure 4A) shows how changes in temperature can affect nucleation: the lower the temperature, the lower the concentration at which nucleation can occur. The curves on the right (Figure 4B) show how supersaturation effects the nucleation rate, growth rate, and crystal size of the condition. Temperature can therefore be modulated to optimize crystal growth conditions.

Thus, the number of parameters that both contribute to crystal formation and which can be modulated by the experimenter is extensive, and includes:

- Protein concentration
- Protein purification buffer components (which are themselves kosmo/chaotropes)
- Solubilization agents, both strength and concentration
- Precipitation agents, both strength and concentration
- Temperature
- pH
- Volume of both protein and crystallization condition to add

Because of the complexity of the conditions contributing to protein solubility and phase separation, and because this process depends heavily on the kinetics of the system rather than thermodynamic properties, devising crystallization conditions for a protein of interest is a

trial-and-error process. Therefore, using high-throughput techniques to screen for a new crystallization condition is advantageous. Companies sell many sets of pre-made crystallization conditions in formats like 96-well deep well plates and make specialized 96-well crystallization trays, and researchers use robots to test many crystallization conditions quickly using minimal sample protein. Even with all that, though, the process of identifying a crystallization condition can take months, be very expensive, and require vast amounts of purified protein.

2.2.1. *Space groups*

Crystal formation hinges on the organization of molecules into a regular, three-dimensional lattice. The repeating component of this lattice is the unit cell, which can be described using geometric parameters: the lengths of each length dimension of the unit cell (a, b, c), the angles where dimensions meet (α, β, γ), and sometimes an element of rotation, reflection or inversion. The relationships between these parameters, along with other operations, assign your unit cell to a given space group.

Space groups can be clustered into crystal systems, which are geometric families that share certain general features, outlined below. The seven crystal systems each represent an overarching geometric class with the same rules of side length and angle relationships. These are further grouped by translational symmetry operations (Bravais lattices) and symmetry operations around a single point (point groups, categorized by rotation, reflection, and inversion). Screw axes and glide planes, which combine rotation/reflection and translation in their symmetry operations, also can govern unit cell organization. An overview of crystal systems may be found in Table 1, and illustrations of each crystal system are show in Figure 5.

Table 1: Overview of crystal systems

<i>Crystal system</i>	<i>Symmetry</i>	<i>Bravais*</i>	<i>PG**</i>	<i>SG***</i>	<i>a</i>	<i>b</i>	<i>c</i>	α	β	γ
<i>Triclinic</i>	No symmetry axis	1	2	P:2	<i>a</i>	<i>b</i>	<i>c</i>	α	β	γ
<i>Monoclinic</i>	2-fold axis	2 (P,S)	3	P:8,C:5	<i>a</i>	<i>b</i>	<i>c</i>	90°	β	90°
<i>Orthorhombic</i>	Three mutually perpendicular 2-fold axes	4 (P,S,F,I)	3	P:30,C:10, F:5,I:9,A:4	<i>a</i>	<i>b</i>	<i>c</i>	90°	90°	90°
<i>Trigonal</i>	3-fold axis	1 (R)	5	P:18,R:6	<i>a</i>	<i>a</i>	<i>c</i>	α	β	120°
<i>Tetragonal</i>	4-fold axis	2 (P,I)	7	P:48,I:19	<i>a</i>	<i>a</i>	<i>c</i>	90°	90°	90°
<i>Hexagonal</i>	6-fold axis	1	7	P:27	<i>a</i>	<i>a</i>	<i>c</i>	90°	90°	120°
<i>Cubic</i>	Four 3-fold axes	3 (P,S,F)	5	P:15,F:11, I:10	<i>a</i>	<i>a</i>	<i>a</i>	90°	90°	90°

*Number of Bravais lattices and which ones: *P*: simple (no centering), *S*: base centering (divided into *A*, *B*, or *C*: centered on *X* face only), *F*: face centering, *I*: body centering, *R*: rhombohedral

** Number of point groups

***Number of space groups and their respective Bravais lattice classifications

Crystal system diagrams

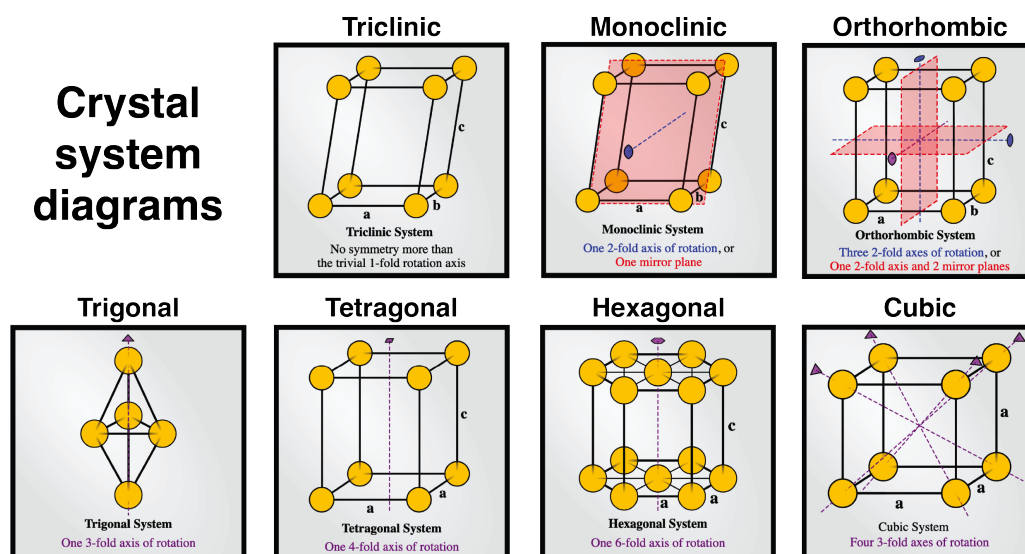


Figure 5. Diagrams of each crystal system, showing mirror planes, axes of rotation, and general geometry. Adapted from the Materials Science Engineering Student page on crystal systems by Brandon Ohl, 2024.

There are 230 total space groups ($14 \times 32 = 448 - \text{duplicates} = 230$); however, protein crystals only occupy the 65 Sohncke, or non-enantiogenic, groups, as they are chiral (reviewed in Fecher et al., 2022). They therefore do not occur in space groups that would require symmetry operations that mirror the molecule or otherwise require enantiomorphs of the

molecule in the crystal. Certain space groups are also more common than others in protein crystals: $P2_12_12_1$ (orthorhombic with a twofold screw axis in each dimension) alone represents around 30% of protein structures solved. This space group, along with other highly represented space groups like $C2$ and $P2_1$, impose fewer requirements on rigid-body packing, and can thus be formed in more ways than other space groups (Wukovitz & Yeates, 1995).

Correctly identifying a crystal's space groups is key to solving a given structure. Diffraction patterns obtained during data collection are not a one-to-one map from atom to position. Rather, spot positions derive from the crystal's space group and its parameters.

2.2.2. *Crystal data collection and Bragg's Law*

Single-crystal protein crystallographic data is generated through a process of X-ray diffraction. A beam of coherent electromagnetic radiation in the X-ray spectrum is shot at a sample crystal, which is rotated to obtain data from all angles. Rather than passing through and forming an image, like in microscopy, X-rays are diffracted by the electron clouds of atomic centers within the crystal, which produces a diffraction pattern by interfering the incident and scattering beams constructively or destructively (Figure 6, reproduced with permission from Biomolecular Crystallography, B. Rupp, 2009). This pattern, which is a 2D collection of reflections with varying intensities, represents a modulus squared of the crystal's unit cell at a given angle.

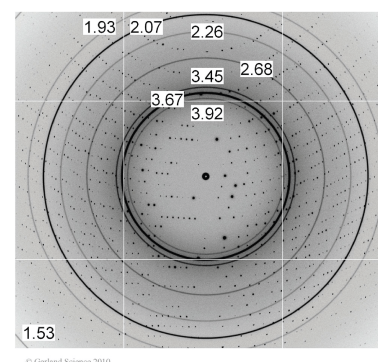


Figure 6. An x-ray diffraction pattern with resolutions labeled. From Rupp Web.

The use of X-rays relates to the length of atoms/atomic bonds and the wavelengths of electromagnetic radiation being used to collect data on a sample. Fundamentally, the resolution of the data is directly correlated to the wavelength of radiation used to collect the data: smaller wavelengths can probe higher resolution features. In the case of crystallography, Bragg diffraction, which is the specific measurement process, occurs with wavelengths comparable to inter-atomic distances. Thus, data collection is typically

performed using electromagnetic radiation with a wavelength of $\sim 1 \text{ \AA}$, which is in the X-ray range.

Diffraction of coherent X-rays through a crystal lattice can be described using Bragg's law:

$$\lambda = \frac{2d \sin \theta}{n}$$

where λ is the wavelength of the X-rays, d is the spacing between crystal planes, θ is the angle of incidence (in this case, also the angle of diffraction), and n is an integer representing the order of diffraction. The resulting pattern of spots is in what is referred to as reciprocal space, and undergo a Fourier transform to yield positions in real space. Importantly, crystal planes are sets of parallel planes that pass through all unit cell origins. The spacing of these planes can vary anywhere from less than an angstrom to orders of magnitude larger than this, as unit cells themselves can range from $<10 \text{ \AA}$ to $>1000 \text{ \AA}$ (Powell, 2021). Large crystal planes, representing large features of the crystal, reflect with a shallow angle, while small crystal planes, representing small features of the crystal, reflect with high angles. This places the spots from small crystal planes toward the outside of the diffraction pattern, while information from large features is centered in the middle. The amplitude of the rays diffracted from a given Bragg plane is the sum of the electron density of that plane, so more electrons on a given plane yields a higher intensity of spot.

2.2.3. *Fourier transforms and the phase problem*

Fourier transforms are integral transforms that allow a user to take a given function and convert it into another function yielding the contributions from frequencies present in the original function. It is the ideal process to take a complex waveform and break it down into all the single-frequency waves that sum to yield that complex waveform. When an object is measured using waves, the data can only be interpreted, initially, using the math of waves – thus, the Fourier transform is inescapable in techniques that take measurements using radiation, like X-ray crystallography and electron microscopy.

Waves have three parameters: wavelength, amplitude, and phase. Reconstructing the real-space solution from reciprocal space data requires information about all three parameters. However, the detectors, or any other physical devices, can only record the intensities of reflections in a diffraction pattern, without phases. Spot position is related to the spacing between crystal planes (which describes the crystal space group) and therefore to the wavelength of scattered radiation. Spot intensities are derived from the sum of the electrons on a given Bragg plane and therefore relate to the amplitude of these waves. However, the phase information, which relates to the direction a given wave came from, is lost.

This is referred to as the phase problem: the phase, which is absent, is needed to be supplied for the inverse Fourier transform back to real-space information. A few methods exist to solve the phase problem. For some applications, heavy atoms are added to the sample, such as through the replacement of methionine with seleniomethionine (which replaces the sulfur atom with selenium) to use the combined information of increased spot intensities and protein sequence. However, for a native crystal, molecular replacement is much more commonly used. Here, an existing structure, either a homologous experimental structure taken from the Protein Data Bank (PDB) or a computationally generated structure made using AlphaFold or similar, is used to guess and check the phase of the data (Jumper et al., 2021; wwPDB consortium et al., 2019). The quality of the real-space electron density depends on how well the phase problem has been solved. Once a suitable guess is obtained, further rounds of refinement – making small adjustments to the position, number and type of atoms within the model – will continue to increase the quality of your map. Thus, crystallographic maps improve as model accuracy increases, while maps from techniques like cryo-EM, which do not suffer from the phase problem, do not change with the model.

2.3. Advantages and Limitations

Crystallography excels at producing high-resolution structures, particularly of small to medium-sized well-behaved proteins. It is historically the backbone of structural enzymology and rational drug design. However, it is limited by its requirement for ordered single crystals. Many proteins—including membrane proteins, flexible complexes, and large

assemblies—can prove intractable to crystallization. Even those which crystallize may produce crystals that are unsuitable for diffraction experiments due to defects. Crystallization may also trap proteins in non-physiological conformations due to lattice packing or non-physiological solvent conditions. These limitations have motivated the rise of cryo-EM as a complementary approach capable of handling more heterogeneous and dynamic samples.

3. Cryo-Electron Microscopy (Cryo-EM)

3.1. Overview

The development of single-particle cryo-EM has transformed structural biology by enabling the visualization of macromolecular assemblies without the need for crystallization. In single-particle analysis (SPA), purified protein suspended in aqueous buffer is applied to a support, rapidly frozen to ensure ice vitrification rather than crystallization, and imaged using transmission electron microscopy. The resulting micrographs contain thousands of particle projections—2D “shadows” of the particle—in random orientations. These projections are averaged and assembled using computational algorithms into a 3D electron density map.

Cryo-EM is particularly powerful as it can be used to solve structures from heterogeneous samples, unlike crystallography. It is especially suitable for studying macromolecular assemblies and integral membrane proteins, whose requirement of a membrane-like environment can impede crystal formation. Cryo-EM structures, though on average lower in resolution than crystal structures, can nevertheless achieve sub-angstrom resolution, dependent on the rigidity of the particle and the quality of the data/analysis. Cryo-EM is more suitable to larger particles, a consequence of the need to align particle projections, though this can be overcome through the use of molecular scaffolds when preparing the sample and/or symmetry expansion during data processing. Single-particle cryo-EM along with other cryo-EM techniques are rapidly developing and becoming increasingly practical for groups who do not specialize in structural biology, expanding the utility of the technique and resulting in an increasing number of structures every year. Below, I will outline the basics of

the theory and methods behind cryo-EM, discuss some of its advantages and limitations, and introduce its sister technique, negative stain EM.

3.2. Theory

Ernst Abbe noted in the late 19th century that the limit of resolution for light microscopes was due not simply to the technology available at the time, but was also related to the wavelength of the radiation itself. Resolution was limited to half the wavelength of the light used, and in the case of the visible spectrum (~700-400 nm), that limited what could be seen to ~200 nm resolution. Higher energy photons outside the visible spectrum, such as X-rays, have a suitably small wavelength, but they are both highly damaging to biological samples and difficult to focus. Hence, electron microscopes were developed to address these problems. Electron beams can be effectively focused, and their picometer wavelengths do not place a limit on resolution. Two primary challenges of electron microscopy remain. First, electrons do not penetrate deeply into thick materials (~500 nm), requiring samples to be exceedingly thin. This necessitates optimization of sample thickness and composition (such as the use of glycerol in a cryo-EM sample buffer, which can reduce contrast, or the addition of heavy atom markers like gold nanoparticles, which can provide high-contrast points), which can modulate the contrast of the desired features. Second, electrons interact strongly with matter and have a large scattering section, and, per particle, electrons supply more energy, and therefore damage, to a sample, than radiation like X-rays. However, because they interact more strongly with the sample, electrons necessitate a much lower total dose than would be needed if using X-rays. Optimization of electron exposure is therefore required to minimize sample damage while maintaining suitable contrast.

3.2.1. Electron microscope anatomy

Electron microscopes are much like light microscopes, but each component is specialized for electrons rather than photons. First, the column through which the beam passes – and thus, where your sample is exposed – must be at high vacuum: 10^{-6} mbar for the sample chamber, and 10^{-5} mbar for the column and camera (and 10^{-9} mbar at the electron gun itself). Unwanted

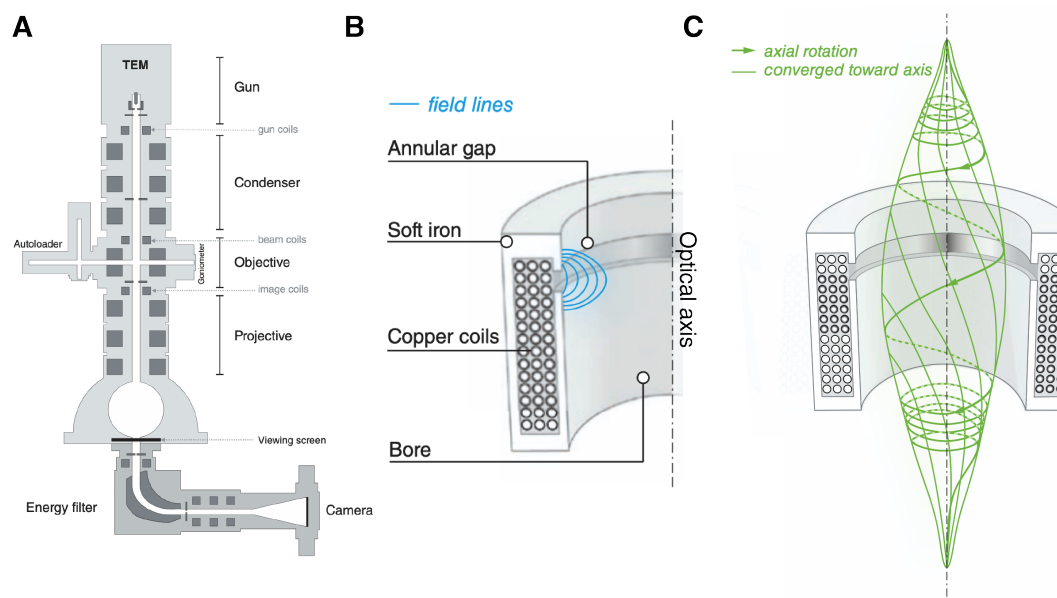


Figure 7. Electron microscope anatomy and electromagnetic lenses. A. Diagram of a transmission electron microscope. B. Anatomy of an electromagnetic lens. C. Path of an electron through the magnetic field of the lens. Figures adapted from Plitzko & Bollschweiler, 2024.

vapor in the column can not only scatter the beam, but can also corrode or damage the equipment over time. The electron beam is generated from an electron gun, such as a bent filament or a crystal of lanthanum hexaboride. Current passes through the filament, exciting electrons, and some of these high-energy electrons may escape from the point of the bend. The electron spray is focused and then accelerated toward the sample using an accelerator stack (labeled “gun coil” in Figure 7A, adapted from Plitzko & Bollschweiler, 2024), a series of electrified discs connected with resistors that form a stepwise energy gradient and can accelerate each electron to the same voltage. Because the voltage determines the wavelength and thus the focal plane for that electron, accelerating electrons to the same voltage is essential.

Electron microscopes have a series of lenses: a set of condenser lenses, an objective lens, and a projector lens (Figure 7A). Electrons are focused using an electromagnetic lens, which is a coil of wire around the optical axis of the microscope that generates a toroidal magnetic field (Figure 7B). The toroidal field controls the path of each electron, leading to a spiral trajectory

centered on the optical axis. This coiled path causes the image to rotate for each lens the electrons must pass through before they reach the sample (Figure 7C). However, this lens alone is not sufficient to focus the beam to the angstrom precision necessary for electron microscopy. Lenses have three additional components: first, the beam encounters deflectors, which redirect the incoming beam to the optical axis of the lens, which focuses the radiation; second, a set of stigmators which correct imperfections in the shape of the lens' magnetic field; and third, there is an aperture, which limits the radiation of the exiting beam to only those particles along the

optical axis (Figure 8). The two condenser lenses, C1 and C2, immediately follow the accelerator stack in the path of the beam, and are responsible for determining the spot size (C1) and intensity (C2) of the beam. These are followed by the objective lenses, which are located immediately above and below the sample, and produces a magnified image of the sample. This magnified image then passes through the projector lens system, which further magnifies that image and projects it to the detector. These lenses are all tunable by modulating the current passing through them, changing the point at which the beam is focused.

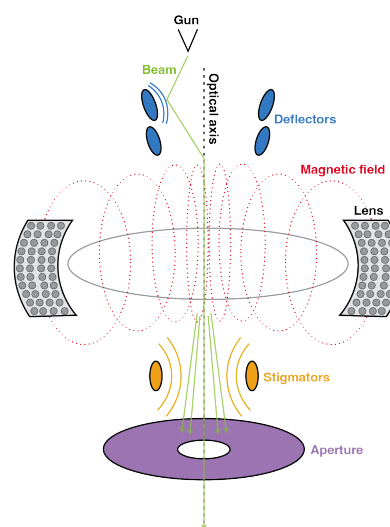


Figure 8. Diagram showing positions of deflectors, stigmators, and aperture relative to lens. Adapted from Grant Jensen's electron microscopy lectures.

When electrons interact with the sample, they may be unscattered, scattered elastically, or inelastically scattered (Figure 9). While no energy is lost in the case of unscattered or elastically scattered electrons and information from both is important for reconstructing features of the image, inelastic scattering results in the loss of energy and can contribute to noise in the image. Energy filters, which remove lower energy particles, therefore can be

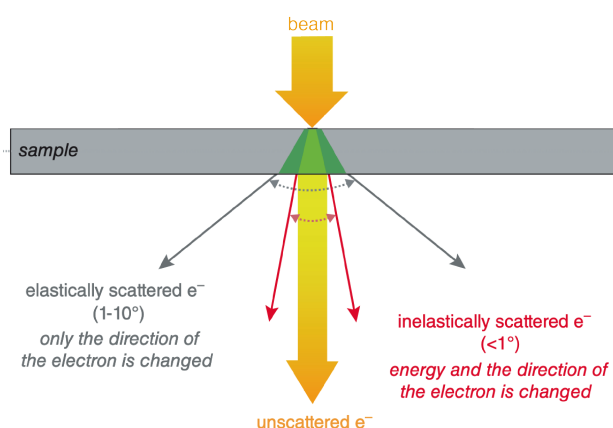


Figure 9. Electron scattering through a TEM sample. Adapted from Plitzko & Bollschweiler, 2024.

inserted ahead of the detector to improve the signal to noise ratio of the image. Only signal from electrons that pass through the filter are detected. While early detectors were photographic film, now we typically use direct electron detectors. These work by converting the signal from each incident electron to a charge and can be quite thin,

mitigating the effects of scattering on the detector and yielding a smaller point spread function, which describes the positional error of each detected electron. Images can be collected very rapidly, recording many frames per second, allowing images to be aligned to produce a higher contrast image, and enabling the exclusion of images with evidence of radiation damage from the final micrograph.

3.2.2. *The contrast transfer function and image defocus*

Contrast, or the degree to which a feature can be distinguished from background, is defined as the difference between the scattered and reference beams, and can be determined from differences in either the amplitude or phase of detected waves. Amplitude contrast results from variations in sample electron density: denser areas scatter more electrons, resulting in more loss of those electrons and thus a lower amplitude signal. Phase contrast, however, results from the interference of scattered rays as they recombine at the image plane. The

angle of electron scattering is related to the size of the feature that a given electron is scattered from, called scattering centers: large centers, which have a low spatial frequency, scatter at small angles, and small centers, with a high spatial frequency, scatter at large angles. These scattered waves travel further than unscattered waves, but maintain the same wavelength. This means that, once focused, waves that were scattered will be out of phase with the unscattered radiation (Figure 10E, adapted from MyScope). Signal from each feature, therefore, is the sum of scattered and unscattered waves that interacted with it. Out of phase components will reduce the amplitude of the wave, while in phase components will increase it.

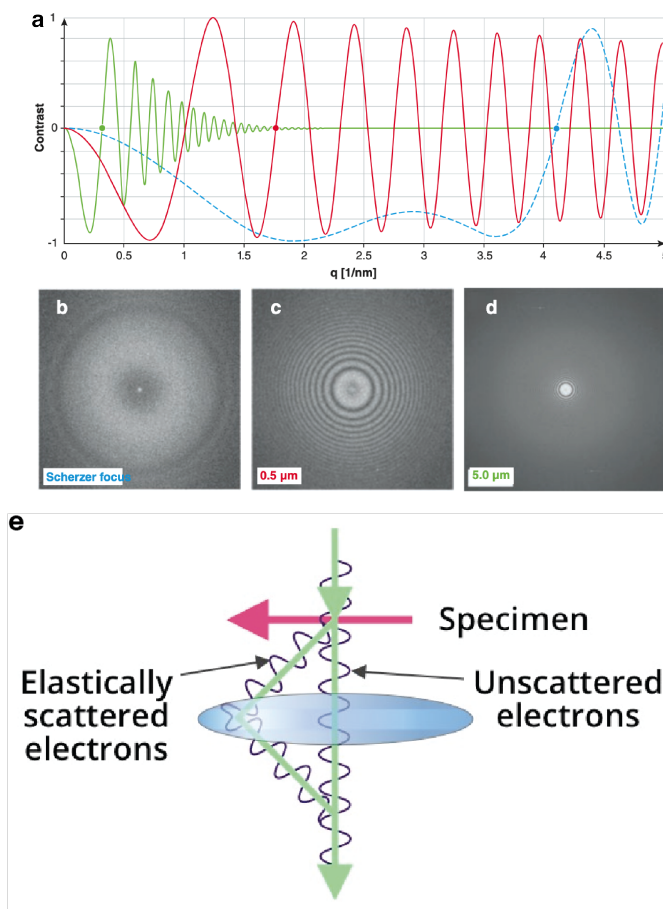


Figure 10. The contrast transfer function. A. Three lines (blue, red, and green) represent CTFs at different defocus levels. B,C,D. show the 2D representation of the CTF at each focus level. E. Phase contrast visual explanation. Adapted from Plitzko & Bollschweiler, 2024 (A-D) and MyScope (E).

The contrast transfer function (CTF) represents how imperfections in the lenses and spherical aberration impact the magnitude of each spatial frequency in the image (Figure 10A-D). Additionally, imperfections in the microscope contribute to variation in the signal transfer fidelity across spatial frequencies as well, so the CTF is further modified by the instrument. To some extent, defects in the CTF can be corrected during data processing, but CTF correction cannot fully restore lost information. However, the user can tune the shape the CTF will ultimately have at the point of data

collection by changing the focus of the microscope. Images that are close to focus bias high-resolution details, but sacrifice the low-resolution features that permit particle alignment. Therefore, users generally collect data sets across a range of defocus levels.

3.2.3. *Sample, sample, cold and thin*

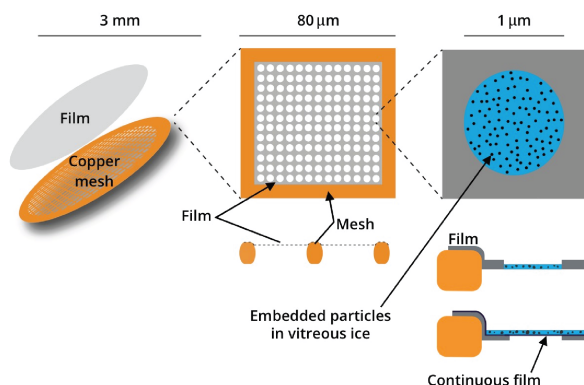


Figure 11. Grid schematic, adapted from MyScope.

One of the major innovation points in the development of electron microscopy was the ability to flash-freeze aqueous samples. Frozen samples offer several advantages over room-temperature samples: first, proteins can be in a native-like environment without as great a risk of evaporation, enabling the capture of more native structures, and

second, freezing, like crystallization, reduces both the conformational and spatial dynamics of the sample. However, it is critical that the ice in the sample is vitreous and disordered rather than crystalline. Crystalline ice is exceptionally damaging to biological samples, plus it diffracts electrons quite well and produces features of higher contrast than the sample itself. This necessitates samples to be quite thin, allowing heat to be drawn evenly and quickly from them by the vitrifying system. Purified protein samples for cryo-EM, which are suspended in aqueous buffer, are therefore loaded onto thin supports, or grids, and excess liquid is removed. The remaining suspension stretches across arrays of holes in the grid at a thickness between 200-500 nm.

While a thin sample is good both for vitrification and for the transmission of electrons through the sample, it can be remarkably hazardous to the sample itself. Proteins fold not only as a function of their sequence, but also in response to their environment, and the hydrogen bonding network of a protein's hydration sphere is crucial for maintaining that protein's proper folding. Exposure to air can rapidly denature a protein, though proteins vary

in their sensitivity. While this is not always an issue, as theoretically there is plenty of protein that does not engage with the air-water interface, some proteins might preferentially partition toward it, and some domains might be more sensitive than others, leading to non-random, or preferred, particle orientation or partial denaturation of the sample. Certain additives, like detergents, can alter the manner in which proteins partition throughout the ice, and others, like glycerol, can act as a cryoprotectant, reducing the risk of ice crystallinity. However, additives come with their own costs and interactions with the sample, and therefore no additive is a universal solution.

3.2.4. From particles to structures

Single-particle cryo-EM data sets contain many thousands of projections of the particle of interest, ideally with relatively low conformational and compositional heterogeneity, in many different equally-represented orientations, each of which yields a unique 2D projection. Joaquim Frank and others developed methods to select particles and align 2D projections, increasing the SNR of positionally identical features and allowing projections belonging to the particle of interest to be selected and reconstructed into a 3D electron density map (Frank et al., 1978). Currently, these steps are performed in open-source software such as RELION, developed by Sjors Scheres at the MRC in Cambridge, or CryoSPARC, a proprietary software developed by graduates of the University of Toronto (Punjani et al., 2017; Scheres, 2012b). A diversity of algorithms in each program enable different methods of particle picking, sorting, and 3D reconstruction and refinement. Structure post-processing tools like EMRinger, developed by James Fraser and his group at UCSF (including Yifan Cheng, art of the team to obtain the 2013 structure that heralded the resolution revolution), have also been created to further sharpen and refine maps (Barad et al., 2015).

Depending on the final resolution of a map and the needs of the researcher, models can be built in a number of different ways. For high resolution (~ 3 Å, though in practice better than 2.5 Å) structures, algorithms like ModelAngelo (also from the Scheres lab) can build a sequence into a density *de novo* (Jamali et al., 2024). Many structures can be modeled by modifying existing experimental structures that have been deposited in the PDB. When no

suitable experimental structure exists, AlphaFold structures are exceptionally useful for model building (Jumper et al., 2021). Domains often show correct or nearly correct secondary and tertiary structure, and can be refined to more accurately reflect the density using interactive software such as Coot or ISOLDE, or using real-space refinement tools in Phenix (Croll, 2018; Emsley et al., 2010; Liebschner et al., 2019).

3.3. Advantages and Limitations

As is demonstrated by the explosion of atomic-resolution cryo-EM structures published since 2013, cryo-EM has opened the door to a world of possibilities in structural biology. First, it permits structure determination directly from a purified sample – no crystallization is required, previously a massive bottleneck to obtaining a given structure. Much less protein is therefore required, both because theoretically less screening is needed to find conditions suitable for structure determination, and because each cryo-EM sample requires only $\sim 3 \mu\text{L}$ of sample at around 1 mg/mL, much less than what is needed for either crystal or NMR structures. Second, larger particles are easier to work with in cryo-EM, and though particle heterogeneity can render data processing more difficult, it is not impossible, and heterogeneous data sets can have the advantage of being able to produce structures of multiple assemblages. This has enabled structures to be determined of large protein complexes previously inaccessible to other techniques.

However, limitations still remain. Many proteins suffer damage at air-water interfaces, which is a salient hazard when producing thin samples on grids. This can result at best in structures missing domains, or can be preferentially oriented such that one view is heavily biased in the data, resulting in anisotropy that can distort data along one axis of the structure. Solutions to these problems at the point of sample preparation, like the addition of detergents, can redistribute particles on the EM grid, requiring higher concentrations, and can also in some cases reduce the overall contrast of the sample, or render the ice thinner and more susceptible to damage from the beam, and many other pitfalls. Data processing itself takes

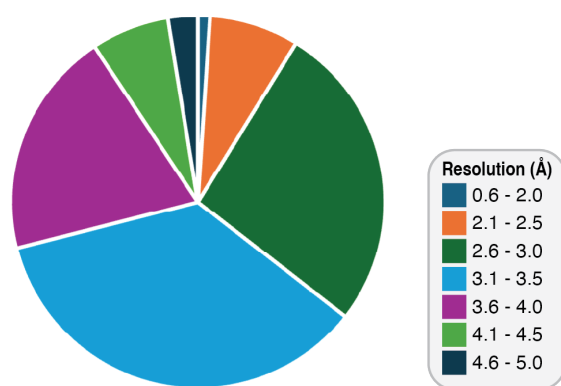


Figure 12. Proportion of cryo-EM structures deposited on the PDB per resolution bin.

an enormous amount of processing power, often requiring the user to be able to access a server with several graphical processing units (GPUs) (compared to processing crystal data, which can be performed locally on a laptop). Data sets and files from processing are also exceptionally large, requiring dozens of terabytes of storage, though maps themselves

are not usually so unwieldy. Finally, though cryo-EM maps are often solved to high resolution (Figure 12), maps solved to only moderate resolutions (~ 4 Å and worse) can be challenging for accurate model building. Additionally, models derived from low resolution maps are often refined with stringent Ramachandran and rotamer restraints, resulting in models that report “perfect” model statistics but may contain significant inaccuracies. Though this is common practice in the cryo-EM community, it can be a problem for non-structural scientists, who represent the vast majority of structure users.

3.4. Negative-stain electron microscopy

Negative-stain TEM was developed in the late 1950's, first published in 1959 by Brenner and Horne at the MRC in Cambridge, and became a widely used technique for investigating biological samples (Brenner & Horne, 1959). While cryo-EM uses unstained samples and relies on particle abundance, varying the defocus, and performing image processing to boost the sample's signal, as well as requiring the sample to remain at cryogenic temperatures, negative-stain EM samples are prepared at room temperature and stained with a heavy metal salt such as uranyl formate or uranyl acetate. The stain surrounds and supports the biological sample, allowing it to be dehydrated and stored at room temperature for long periods of time. The stain has strong contrast, as it is very electron-dense for scattering, and imaging at defocus is not necessary; however, these features reduce the resolution that can be achieved to $\sim 15\text{-}30$ Å. Though insufficient for an atomic structure, negative-stain EM can provide important information about sample heterogeneity, particle size, and other basic features. It also requires concentrations much lower than cryo-EM, around 0.02 mg/mL, though the concentrations for better visualization under EM should be screened. It is thus often used to screen samples ahead of cryo-EM. However, importantly, many heavy metal stains are quite acidic (pH $\sim 3\text{-}4$) and can denature some proteins or disrupt protein complexes, so some samples might require fixation prior to negative staining.

4. Chapter summaries

This thesis represents the application of crystallography and single-particle cryo-EM to a diversity of projects, all focused on molecules associated with immune activity. These projects span not only multiple domains of life, but also multiple labs, featuring work with the Bjorkman group from the earliest part of my PhD as well as collaborations between the Chou lab and Richard Youle's group at the NIH and Alexei Aravin's group, previously at Caltech, now at Cornell.

4.1. Chapter 2: Identification and structural characterization of a broadly neutralizing antibody in complex with the SARS-CoV-2 S protein RBD

In this chapter, representing work up to and after my candidacy, I characterize 11G-12v, an antibody that binds a conserved region of the SARS-CoV-2 receptor binding domain (RBD). With the support of staff scientist Dr. Jennifer Keefe, I performed ELISAs to assess the binding of 11G-12v with SARS-CoV-2 variants of concern, other coronavirus RBDs representing multiple clades, and in competition with previously characterized Fabs with known binding sites to identify the putative binding site and breadth of neutralization of 11G-12v. I solved a cryo-EM structure showing three unequally resolved copies of the 11G-12v Fab in complex with the SARS-CoV-2 Spike (S) trimer to 3.4 Å. Analysis revealed that the 11G-12v Fab belongs to a larger group of antibodies which feature a motif in their CRDH3 (YYDRxG). I determined that 11G-12v is a Class 1/4 antibody, a category of broadly neutralizing antibodies that simultaneously binds to a conserved but cryptic epitope on the S RBD and prevents the RBD from being able to associate with the SARS-CoV-2 entry receptor, ACE2.

4.2. Chapter 3: Literature review: A structural perspective on the p97 adaptor Phospholipase A2 Activating Protein (PLAA)

The Chou lab, which I joined halfway through my third year, researches p97, a AAA+ ATPase responsible for pulling proteins out of membranes and protein complexes and targeting them to the proteasome for degradation, and is critical for a vast network of processes in the cell. In order to do this, p97 associates with a wide array of adapter proteins, linking it to ubiquitinated targets. This chapter reviews the literature on PLAA, a p97 adapter about which little is known but which shows a tantalizing range of functions, from the perspective of multifunctional proteins such as moonlighting proteins, using clues from its structure to inform my analysis.

4.3. Chapter 4: Structural characterization of the FIP200 C-terminal CLAW domain bound to an autophagy adaptor peptide, TNIP1-pFIR

Among many collaborators, the Chou lab worked with Richard Youle's group at the NIH for several years, spearheaded by fellow BMB student, now graduate, Dr. William Rosencrans. Joint work focuses on autophagy initiation complex proteins and their many properties, interactors, and potential for drug development. In this chapter, I worked with Will and NIH postdoc Dr. David Nyenhuis to investigate the interaction between the FIP200 C-terminal CLAW domain and TNIP1, following up on research previously published by the Youle group. Using a peptide derived from the TNIP1 LIR2, I obtained a crystal structure and determined the manner in which TNIP1 binds to the CLAW, which follows a larger pattern of other CLAW-binding proteins. Together with binding analyses via fluorescent polarization assays and NMR data collected by Dr. Nyenhuis, this structure provided information for us to better understand the manner in which CLAW binds to adapter proteins like TNIP1 and CCPG1. Unfortunately, shortly after we obtained the structure but before we were able to publish, we were scooped.

4.4. Chapter 5: ATP-driven tetramerization and RNA-guided filament assembly control of a Schlafen–Argonaute antiphage defense

Finally, in 2023 I was independently approached by a postdoc in Alexei Aravin's group, Dr. Anton Kuzmenko, who was interested in obtaining structures of a prokaryotic Argonaute protein, RzAgo, which he and others had been studying since 2018. This chapter represents the culmination of that work, and is a modified version of the paper we are readying for submission. Here, Anton's biochemical, genomic and bioinformatic analyses identify RzAgo as representative of a novel clade of prokaryotic Argonaute proteins, which possess two domains N-terminal of the Argonaute machinery: an active nuclease reminiscent of

mammalian Schlafen domains, and a GHKL ATPase domain. The RzAgo Argonaute domains are not catalytically active, and lack one of the four canonical domains comprising an Argonaute protein, the N domain. Curiously, RzAgo had been observed to form higher-order oligomers when incubated with guide RNA, which would dissipate upon the addition of target DNA. This suggested a novel regulatory mechanism that had not previously been seen in prokaryotic Argonaute proteins.

I obtained five structures of RzAgo, of the apo, guide-bound and bound to the guide-target heteroduplex, with and without a non-hydrolysable ATP analogue and glutaraldehyde crosslinking. Three of those structures, without glutaraldehyde stabilization, resolved the Argonaute domains alone, and together show details of how guide and target binding are mediated by features unique to RzAgo and other members of the extra-long pAgo clade. The remaining two structures, along with negative stain data, allowed us to resolve the Schlafen and ATPase domains, which form extensive interfaces with each other, permitting the oligomerization noted biochemically. Additionally, the structures of the apo and guide-bound oligomers reveal two markedly different binding strategies. The apo protomers tetramerize, forming a dimer of dimers via interfaces formed exclusively between their Schlafen/ATPase domains. The adjoining dimers are arranged with C2 symmetry, which orients the Schlafen nuclease domains such that they recapitulate the organization observed in the active state of mammalian Schlafen proteins. This suggests that this might be an active conformation. Also of note, while in each dimer of protomers, the Argonaute domains assume a head-to-tail orientation, the tetramer places them in an opposing head-to-head position. Guide binding alters the conformation of the Argonaute domains and introduces another interfacial component, the guide itself, to produce interfaces that associate neighboring Argonaute domains to each other in exclusively a head-to-tail manner. This interface appears to drive the formation of a helical chain of guide-bound RzAgo protomers, splitting the Apo tetramer into dimers (and sometimes linear trimers). This disrupts the Schlafen tetramer, and the binary complex is therefore thought to be an inactive form of RzAgo. Curiously, the tetrameric form may not be unique to the Apo state: RzAgo bound to a guide-target heteroduplex shows a nearly identical pattern when crosslinked and investigated

by SDS-PAGE, and structural evidence from the guide-target bound RzAgo Argonaute domains shows that the heteroduplex interrupts the interface between the Argonaute domains themselves.

Anton performed experiments to investigate the activity of the Schlafen and ATPase domains. He first found that the Schlafen domain showed sequence elements that matched active Schlafen domains, and that mutations to these residues could halt or accelerate Schlafen activity in the same pattern observed for mammalian Schlafens. Additionally, alanine mutations of residues essential for guide binding both prevented the formation of the higher-order oligomers observed in the wild type, and also caused RzAgo to become toxic to the cells. This suggests that the Schlafen nuclease is indiscriminate about its targets without the Argonaute domains. He was also able to confirm that the ATPase domain is active. ATPase activity is critical to the story of RzAgo's oligomeric reorganization: many interface-forming residues are found in the loop that forms the ATP lid, a structure that is mobile in other GHKL ATPases upon ADP release. ATP lid stabilization by ATP binding, and ultimate destabilization by ADP release, is therefore postulated as a mechanism of RzAgo protomer reorganization.

The biochemical data in tandem with structural data suggest a unique mechanism to explain how RzAgo and other extra-long pAgos protect the cell against phage infection. While in the guide-bound form, RzAgo is sequestered in oligomers with no active Schlafen tetramers, unable to harm the cell but able to sense single-stranded phage DNA. Once bound to guide and target, the Schlafen domains can be reorganized into the active tetramer, allowing substrates to be cleaved. The ATPase activity further permits the dissociation and reorganization of RzAgo protomers, but the exact mechanism determining the rate of ADP release is not yet understood.

Bibliography

- Adrian, M., Dubochet, J., Lepault, J., & McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature*, 308(5954), 32–36. <https://doi.org/10.1038/308032a0>
- Almeida, J. D., & Tyrrell, D. A. (1967). The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. *The Journal of General Virology*, 1(2), 175–178. <https://doi.org/10.1099/0022-1317-1-2-175>
- Barad, B. A., Echols, N., Wang, R. Y.-R., Cheng, Y., DiMaio, F., Adams, P. D., & Fraser, J. S. (2015). EMRinger: Side chain–directed model and map validation for 3D cryo-electron microscopy. *Nature Methods*, 12(10), 943–946. <https://doi.org/10.1038/nmeth.3541>
- Barrett, P. J., Chen, J., Cho, M.-K., Kim, J.-H., Lu, Z., Mathew, S., Peng, D., Song, Y., Van Horn, W. D., Zhuang, T., Sönnichsen, F. D., & Sanders, C. R. (2013). The Quiet Renaissance of Protein NMR. *Biochemistry*, 52(8), 1303–1320. <https://doi.org/10.1021/bi4000436>
- Bernal, J. D., & Crowfoot, D. (1934). X-Ray Photographs of Crystalline Pepsin. *Nature*, 133(3369), 794–795. <https://doi.org/10.1038/133794b0>
- Bragg, W. H., & Bragg, W. L. (1913). The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 88(605), 428–438. <https://doi.org/10.1098/rspa.1913.0040>
- Brenner, S., & Horne, R. W. (1959). A negative staining method for high resolution electron microscopy of viruses. *Biochimica Et Biophysica Acta*, 34, 103–110. [https://doi.org/10.1016/0006-3002\(59\)90237-9](https://doi.org/10.1016/0006-3002(59)90237-9)

- Cray, J. A., Russell, J. T., Timson, D. J., Singhal, R. S., & Hallsworth, J. E. (2013). A universal measure of chaotropy and kosmotropy. *Environmental Microbiology*, 15(1), 287–296. <https://doi.org/10.1111/1462-2920.12018>
- Croll, T. I. (2018). ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, 74(6), 519–530. <https://doi.org/10.1107/S2059798318002425>
- De Rosier, D. J., & Klug, A. (1968). Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature*, 217(5124), 130–134. <https://doi.org/10.1038/217130a0>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), Article 4. <https://doi.org/10.1107/S0907444910007493>
- Fecher, G. H., Kübler, J., & Felser, C. (2022). Chirality in the Solid State: Chiral Crystal Structures in Chiral and Achiral Space Groups. *Materials*, 15(17), 5812. <https://doi.org/10.3390/ma15175812>
- Frank, J., Goldfarb, W., Eisenberg, D., & Baker, T. S. (1978). Reconstruction of glutamine synthetase using computer averaging. *Ultramicroscopy*, 3, 283–290. [https://doi.org/10.1016/S0304-3991\(78\)80038-2](https://doi.org/10.1016/S0304-3991(78)80038-2)
- Giunta, C. J., & Mainz, V. V. (2020). Discovery of Nuclear Magnetic Resonance: Rabi, Purcell, and Bloch. In *Pioneers of Magnetic Resonance* (Vol. 1349, pp. 3–20). American Chemical Society. <https://doi.org/10.1021/bk-2020-1349.ch001>
- Glaeser, R. M. (1999). Review: Electron crystallography: present excitement, a nod to the past, anticipating the future. *Journal of Structural Biology*, 128(1), 3–14. <https://doi.org/10.1006/jsbi.1999.4172>

- Gräwert, M., & Svergun, D. (2020). A beginner's guide to solution small-angle X-ray scattering (SAXS). *The Biochemist*, 42(1), 36–42. <https://doi.org/10.1042/BIO04201036>
- Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Quarterly Reviews of Biophysics*, 28(2), 171–193. <https://doi.org/10.1017/s003358350000305x>
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E., & Downing, K. H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of Molecular Biology*, 213(4), 899–929. [https://doi.org/10.1016/S0022-2836\(05\)80271-2](https://doi.org/10.1016/S0022-2836(05)80271-2)
- Jamali, K., Käll, L., Zhang, R., Brown, A., Kimanius, D., & Scheres, S. H. W. (2024). Automated model building and protein identification in cryo-EM maps. *Nature*, 628(8007), 450–457. <https://doi.org/10.1038/s41586-024-07215-4>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Liang, B., & Tamm, L. K. (2016). NMR as a Tool to Investigate Membrane Protein Structure, Dynamics and Function. *Nature Structural & Molecular Biology*, 23(6), 468–474. <https://doi.org/10.1038/nsmb.3226>
- Liao, M., Cao, E., Julius, D., & Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504(7478), 107–112. <https://doi.org/10.1038/nature12822>

- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., ... Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, 75(10), Article 10. <https://doi.org/10.1107/S2059798319011471>
- Plitzko, J. M., & Bollschweiler, D. (2024). Fundamentals of Instrumentation and Electron Optics for Cryo-Electron Tomography. In F. Förster & A. Briegel (Eds.), *Cryo-Electron Tomography: Structural Biology in situ* (pp. 1–52). Springer International Publishing. https://doi.org/10.1007/978-3-031-51171-4_1
- Powell, H. R. (2021). A beginner's guide to X-ray data processing. *The Biochemist*, 43(3), 46–50. https://doi.org/10.1042/bio_2021_124
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3), Article 3. <https://doi.org/10.1038/nmeth.4169>
- Ranatunga, W. K., Su, D., & Romero, M. F. (2014). Obtaining Crystals. In *Molecular Life Sciences* (pp. 1–9). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6436-5_28-2
- Scheres, S. H. W. (2012a). A Bayesian View on Cryo-EM Structure Determination. *Journal of Molecular Biology*, 415(2), 406–418. <https://doi.org/10.1016/j.jmb.2011.11.010>
- Scheres, S. H. W. (2012b). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3), 519–530. <https://doi.org/10.1016/j.jsb.2012.09.006>

- Smyth, M. S., & Martin, J. H. J. (2000). X Ray crystallography. *Molecular Pathology*, 53(1), 8–14. <https://doi.org/10.1136/mp.53.1.8>
- Wukovitz, S. W., & Yeates, T. O. (1995). Why protein crystals favour some space-groups over others. *Nature Structural Biology*, 2(12), 1062–1067. <https://doi.org/10.1038/nsb1295-1062>
- wwPDB consortium, Burley, S. K., Berman, H. M., Bhikadiya, C., Bi, C., Chen, L., Costanzo, L. D., Christie, C., Duarte, J. M., Dutta, S., Feng, Z., Ghosh, S., Goodsell, D. S., Green, R. K., Guranovic, V., Guzenko, D., Hudson, B. P., Liang, Y., Lowe, R., ... Ioannidis, Y. E. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1), D520–D528. <https://doi.org/10.1093/nar/gky949>

*Chapter 2*IDENTIFICATION AND STRUCTURAL CHARACTERIZATION OF A
BROADLY NEUTRALIZING ANTIBODY IN COMPLEX WITH THE
SARS-COV-2 S PROTEIN RBD*With acknowledgement to:**Chengcheng Fan, Jennifer Keefe, Ashwin Skelly (UPenn), and Pamela J. Björkman***Abstract**

SARS-CoV-2 continues to pose a global health threat due to ongoing viral evolution and immune escape, underscoring the need for broadly neutralizing antibodies and pan-sarbecovirus vaccine strategies. The viral Spike (S) glycoprotein, particularly the receptor binding domain (RBD), is central to host cell entry and is the primary target of neutralizing antibodies. While many potent antibodies target variable regions of the RBD, some subsets recognize conserved and/or cryptic epitopes and exhibit broad cross-reactivity. Here, we describe the identification and structural characterization of 11G-12v, a broadly neutralizing Class 1/4 antibody. Using neutralization assays, competition ELISA, and cross-reactive RBD binding studies, 11G-12v was shown to neutralize diverse SARS-CoV-2 variants and bind a range of sarbecovirus RBDs. Single-particle cryo-electron microscopy resolved the structure of 11G-12v Fab in complex with the prefusion S trimer at 3.4 Å resolution. Structural analysis reveals an elongated CDRH3 loop containing a conserved YYDRxG motif that extends the RBD core β -sheet, stabilizing the Fab position and sterically impeding ACE2 engagement. These findings support existing work into the mechanism of Class 1/4 antibody recognition and help inform rational vaccine design targeting conserved sarbecovirus epitopes.

1. Introduction

1.1. *Overview of SARS-CoV-2*

SARS-CoV-2 is a pandemic coronavirus that emerged in late 2019, causing the respiratory disease COVID-19 (P. Zhou et al., 2020). It rapidly spread around the world and is now active globally (*WHO Coronavirus (COVID-19) Dashboard*, n.d.). As of August 18th, 2022, there have been over 590 million confirmed cases, and over 6.4 million deaths have been attributed to this disease (*WHO Coronavirus (COVID-19) Dashboard*, n.d.). Despite the fact that over 12.4 billion vaccine doses have been administered and have proven effective against severe illness, the risk of mutations that cause escape from these vaccines is high, as shown with the rise of the Omicron variants across the world throughout 2022 (Cao, Wang, et al., 2022; Cao, Yisimayi, et al., 2022). Therefore, there is great interest in both developing a vaccine that provides resistance against all sarbecoviruses in a non-strain-specific manner (Cohen et al., 2021, 2022), as well as understanding the binding mechanisms of antibodies that are neutralizing across many sarbecoviruses (Barnes, Jette, et al., 2020; Barnes, West, et al., 2020; Dejnirattisai et al., 2021; Fan et al., 2022; Greaney et al., 2021; Jette et al., 2021a; Starr et al., 2021; Xiang et al., 2022).

SARS-CoV-2 is an enveloped Betacoronavirus (Z. Zhou et al., 2021). It features a 30-kilobase, single-stranded, positive-sense RNA genome (V'kovski et al., 2021), which is translated in the host cell into two polyproteins (V'kovski et al., 2021; Zhu et al., 2020) comprising 29 genes (P. Zhou et al., 2020; Zhu et al., 2020). Many of these genes are involved in replication of the viral genome, including an RNA-dependent RNA polymerase (RdRp) (Malone et al., 2022; V'kovski et al., 2021), viral helicase, exonucleases, and other replication proteins (Malone et al., 2022). Once translated, the viral polyprotein is cleaved with internal proteases, allowing the formation of the non-structural proteins involved in evasion of the innate immune system, membrane remodeling and stimulation of autophagy, viral RNA replication, and other processes that permit the production and release of virion

particles (V'kovski et al., 2021). Four additional proteins are translated outside of the polyprotein, which are structural components of the virion particle itself (V'kovski et al., 2021). These four components – the Spike glycoprotein (S), Membrane protein (M), Envelope protein (E), and Nucleocapsid protein (N) – are involved in stabilizing the viral genome within the virion particle (N) (Burbelo et al., 2020; Lu et al., 2021; Savastano et al., 2020; Scherer et al., 2022), stabilizing the viral lipid envelop (M (Dolan et al., 2022; Schoeman & Fielding, 2019; Zhu et al., 2020) and possibly E (Schoeman & Fielding, 2019; Zhu et al., 2020)), and interacting with the host receptor to enter host cells (S) (Cai et al., 2020; Duan et al., 2020; Huang et al., 2020; Yi et al., 2020; P. Zhou et al., 2020).

1.2. The SARS-CoV-2 S protein

The S protein, or Spike protein, is a class 1 fusion protein, and is critical for viral entry into the cell. It has two roles: first, to recognize and bind the host receptor, here angiotensin-converting enzyme 2 (ACE2) (Lan et al., 2020; P. Zhou et al., 2020), displayed by susceptible populations of host cells, and second, to mediate fusion of viral and host membranes (Bosch et al., 2003), permitting the release of the viral genome into cells where viral replication may then occur. Due to its integral role in viral entry, the S protein has been the target of vaccines against SARS-CoV-2 and is the main protein of interest in the study of how antibodies prevent SARS-CoV-2 infection via neutralization.

S is arrayed on the surface of the SARS-CoV-2 virion particle (Ke et al., 2020; Turoňová et al., 2020; Yao et al., 2020) (Figure A) and is formed of three protomers (Wrapp et al., 2020) (Figure B). Each protomer is around 1273 amino acids in length, depending on the variant, with a mass of roughly 140 kDa (*S - Spike Glycoprotein - Severe Acute Respiratory Syndrome Coronavirus 2 (2019-nCoV) | UniProtKB | UniProt*, n.d.). The S protomer is made of two subunits, S1 and S2 (Huang et al., 2020) (Figure C). S1, responsible for host receptor recognition, comprises the N-terminal domain (NTD) whose function is still being characterized (Meng, Datir, et al., 2022), the receptor binding domain (RBD) responsible for

binding the host receptor ACE2, and a C-terminal domain that joins with S2 (Huang et al., 2020) (Figure D). S2, which mediates the membrane fusion process between virus and cell, forms a coiled-coil stalk in the center of the prefusion S trimer and is responsible for membrane fusion following ACE2 recognition (Cai et al., 2020; Marcink et al., 2022). It is composed of two sets of heptad repeats, which are highly conserved helical elements involved in the conformational changes of the S2 subunit from the pre- to the post-fusion states, as well as a transmembrane and cytoplasmic domains which will anchor in the host cell (Cai et al., 2020). Cleavage of S1 from S2 by furin proteases is required both for ACE2 recognition and for the release of S1 for S2 function. This occurs during viral particle maturation in the infected cell, and the two subunits remain non-covalently associated until ACE2 recognition (Jackson et al., 2022; Lavie et al., 2022). Subsequent cleavage events by serine proteases like transmembrane serine protease TMPRSS2 (Iwata-Yoshikawa et al., 2022; Meng, Abdullahi, et al., 2022; Shapira et al., 2022) or lysosomal proteases like cathepsin L (Zhao et al., 2022) then occur during viral entry into the cell that permit the S2-mediated membrane fusion process.

Recognition of the host receptor marks the beginning of viral infection, and interrupting this process is therefore of great interest. The receptor binding domain (RBD) has therefore been at the forefront of investigations into SARS-CoV-2 neutralization: not only is this domain responsible for ACE2 recognition, but also it has been found that many antibodies formed in response to SARS-CoV-2 infection target the RBD (Premkumar et al., 2020). The RBD is formed of two subdomains: a core beta-sheet structure made of five antiparallel beta strands surrounded on either side by short alpha helices, and a long loop that forms the receptor binding motif (RBM), forming the top of the RBD and mediating the interaction with ACE2 (Shang et al., 2020). This may also be termed the receptor binding site or ridge (Dejnirattisai et al., 2022; Shang et al., 2020). The core subdomain is much more highly conserved than the RBM (Starr et al., 2021). Preventing RBD, particularly RBM, binding to the ACE2 receptor prevents infection, effectively neutralizing the viral particle (Barnes, Jette, et al., 2020; Barnes, West, et al., 2020).

In the prefusion Spike, RBDs sample two conformations: “down”, in which the RBD interior face is packed against the top of the S2 subunit helices (Henderson et al., 2020; Jette et al., 2021b), or “up”, where the ACE2 binding site – the RBM – is able to sample the environment and potentially bind ACE2, but is vulnerable to recognition by potentially neutralizing antibodies that directly block this interaction (Figure 1E) (Barnes, Jette, et al., 2020; Barnes, West, et al., 2020; Kirchdoerfer et al., 2016; Walls et al., 2016; Wrapp et al., 2020; Yuan et al., 2017). When “down”, the interior face of the RBD core packs against the top of the helical core of the S2 subunit. Mutations in this region could destabilize the “down” conformation, rendering the RBM more vulnerable to antibody recognition. The RBM, conversely, is more variable across SARS-CoV-2 strains and across coronaviruses likely due to selective pressure from neutralizing antibodies that prevent binding of the RBM to ACE2 (Starr et al., 2021). Mutations in this subdomain are selected through balancing antibody evasion with maintaining a high-enough affinity to ACE2 to enable viral transmission.

1.3. Antibodies: an overview

The diversity of microbial, including viral, epitopes is immense. It is also an ever-changing landscape as mutations occur that facilitate escape from immune defenses. In order to specifically counter this process, the adaptive immune system of jawed vertebrates evolved a molecular tool capable of rapid specification toward epitope recognition: antibodies. This extensively diversifiable catalogue of epitope-recognizing proteins is uniquely poised to defend an organism against a changing antigen landscape.

Antibodies are formed by B cells, immune cells born in the bone marrow. Immature B cells express gene segments that code for two polypeptide chains, the heavy chain (Hc) and the light chain (Lc) (Sompayrac, 2019). Each chain is composed of (generally) four or two immunoglobulin (Ig) domains, respectively, which are highly regular globular domains consisting of between seven and nine antiparallel beta sheets that form a two-layer beta sandwich. The N-terminal Ig domains is termed the “variable” domain and is the site of the

majority of diversity-introducing processes in antibody formation. The other Ig regions are “constant” domains, which determine the antibody isotype.

There are five major isotypes, or classes, of antibodies, which have different functions in the immune system. For this project, the antibody class of interest is immunoglobulin G (IgG), which are expressed in the bloodstream by plasma B cells (Sompayrac, 2019). IgGs are Y-shaped, and both arms recognize the same antigen (FigureB). Early research into antibodies found that IgGs could be cleaved into functional fragments: the fragment antigen binding (Fab) comprising the “arm” of the IgG, and the fragment crystallizable (Fc), formed from the IgG “stalk” (FigureC). Fabs contain the variable region of the antibody, which is generally the region of interest for antibody characterization. For structural studies, Fabs are typically preferred over full-length IgGs, in large part because the Fab is much less dynamic than the IgG and therefore produces less conformational heterogeneity.

Antibodies begin life as B-cell receptors, held in place on the cell surface by a complex of transmembrane proteins (Alberts et al., 2002). Diversity in B cell receptors is generated at the genetic level. Hc and Lc genes consist of a Variable (V) region and a Constant (C) region. V regions are themselves assembled from a catalogue of multiple Variable (V), Diversity (D) (only for Hc), and Joining (J) gene segments, allowing many different V regions to be created. A representative from each segment type is selected and incorporated into the new Hc or Lc V region and unused segments are deleted, resulting in a high degree of combinatorial diversity. This process of recombination provides the first layer of receptor diversity. Additionally, junctional diversity, or variability in the joints between these segments, is introduced by the proteins involved in the recombination process itself. Multiple C regions are encoded for both Hcs and Lcs, adding another layer of diversity to each chain, and finally different Hcs and Lcs may be combined together, producing up to 10^{11} possible receptors in naïve mature B cells (Charles A Janeway et al., 2001b).

Within the Hc and Lc V domains, a high degree of variability is concentrated in complementarity determining regions (CDRs) which form loops that mediate antigen binding. Both the Hc and Lc V domains have three CDR loops, the third of which sits at the

V(D)J junction and features the greatest degree of diversity prior to antigen exposure. However, the landscape of antigenic epitope possibilities is vast, and a given epitope may require highly tailored recognition in order for effective neutralization to occur. Thus, upon exposure to antigen as well as signals from T cells, mature B cells undergo somatic hypermutation, which introduces point mutations throughout the V regions of Hcs and Lcs. Mutations that improve B cell receptor recognition of their cognate antigen are selected through affinity maturation, and B cells that then produce higher affinity receptors are selected and proliferated through clonal expansion. B cell activation also triggers their differentiation into either memory B cells or antibody-secreting plasma B cells (Charles A Janeway et al., 2001a). The latter produce large amounts of the selected antibody as soluble protein rather than as membrane-associated receptors, providing systemic protection against the infective agent.

1.4. Antibody recognition of the SARS-CoV-2 RBD

The SARS-CoV-2 RBD is the target of a large percentage of neutralizing antibodies that prevent its association with ACE2. Many RBD-targeting antibodies have been described and characterized structurally. Research teams around the world have categorized these antibodies in a few different ways (Barnes, West, et al., 2020; Dejnirattisai et al., 2021). For the purposes of our research, we use the Barnes classification system (Barnes, West, et al., 2020). This system divides RBD-recognizing antibodies into four different classes (FigureB).

Class 1 antibodies recognize the interior face of the RBM, and as such can only recognize “up” RBDs. They are generally derived from IGVH3-53 and IGVH3-66 genes and are close to germline and have not undergone substantial somatic hypermutation. They feature a short CDRH3 loop which contacts RBM residues. Class 1 antibodies are some of the most potent antibodies described, but are also at risk of obsolescence from evasive mutations in the RBD (Barnes, West, et al., 2020).

Class 2 antibodies recognize the exterior face of the RBM and may recognize either “up” or “down” RBDs. They feature CDRH3s of varying lengths and can, in some cases, contact multiple RBDs at once (e.g., antibody C144). Class 2 antibodies are very strongly neutralizing, and like Class 1 antibodies, this class contains some of the most potent antibodies yet identified. However, like Class 1 antibodies, they may not be broadly neutralizing due to the variability in the RBM (Barnes, West, et al., 2020).

Class 3 antibodies recognize epitopes on the exterior face of the core subdomain of the RBD and recognize both “up” and “down” RBDs. The Class 3 epitope sits outside of the ACE2-binding site, and can, in some cases, be potentially neutralizing, as in the case of S309 (Pinto et al., 2020). Early characterization of these antibodies suggested that Class 3 antibodies, notably S309, which was approved for use as a monoclonal antibody for treatment of SARS-CoV-2, may neutralize by recruiting effector mechanisms, and could be paired with other monoclonals to mitigate the risk of viral escape (Pinto et al., 2020). However, later studies showed that in addition to this, the RBD N343 glycan, which is a feature of the Class 3 epitope, may be involved in mediating RBD opening by intercalating between residues F490, Y489, F456 and R457 in the RBM, pushing the RBD from the “down” to the “up” conformation (Sztain et al., 2021). Binding of this glycan by antibodies such as S309 may therefore prevent RBD opening, providing another neutralization mechanism for this class of antibodies.

Class 4 antibodies recognize a cryptic epitope buried at the base of the interior of the RBD’s core subdomain, therefore binding only “up” RBDs (Barnes, West, et al., 2020). This RBD epitope both interfaces with the S2 subunit and also is less accessible to antibodies than other parts of the RBD, and is therefore the best-conserved region of this domain (Fan et al., 2022). This makes eliciting Class 4 antibodies an attractive challenge for vaccine developers, as some of the antibodies that target this region of the RBD have been shown to be broadly neutralizing, binding RBDs across not only various strains of SARS-CoV-2, including the highly divergent Omicron variants, but also generally across all four sarbecovirus clades (Cohen et al., 2021, 2022; Fan et al., 2022; Greaney et al., 2021; Jette et al., 2021b; Starr et

al., 2021). Certain Class 4 antibodies, such as CR3022, have been shown to interrupt interprotomer S interactions: CDR loops from both the CR3022 heavy chain and light chain recognize the Class 4 epitope of the RBD, orienting the Fab constant domains such that they occupy the same space as the neighboring trimer (Wrobel et al., 2020). Similarly, class 5 antibodies, exemplified by S2H97, target an epitope at the base of the RBD's interior face; however, these antibodies uniquely interact with residues in a crevice at the center of this epitope via their CDRH3 (Starr et al., 2021). These antibodies are also broadly neutralizing.

However, not all antibodies that bind the Class 4 epitope are neutralizing, and those that are potentially neutralizing were found to share a specific structural feature: an elongated CDRH3 loop that interacts with and extends the antiparallel beta sheet forming the RBD core (Burnett et al., 2021). In the case of the Class 4 antibody AB-3467, over 90% of contacts between the heavy chain and the RBD are mediated by the long CDRH3 loop, which features multiple aromatic residues (three tyrosine and a single tryptophan) that form contacts with hydrophobic moieties in the RBD. This positions the light chain such that it sterically obstructs the RBM from recognizing ACE2 (Burnett et al., 2021). This is an entirely different mechanism of interaction from Class 4 antibodies like CR3022, and produces reliably potent and crossreactive antibodies, such as C022 and C118 (Jette et al., 2021b). Additionally, though these antibodies may be derived from a diverse set of gene segments, many, including C022, have been found to share a common YYDRSGY motif in their CDRH3 loop (L. Liu et al., 2022).

This type of antibody, designated Class 1/4 (Fan et al., 2022; Jette et al., 2021b), is therefore of particular interest. Studying and eliciting these broadly neutralizing antibodies is particularly important when considering the design of a pan-sarbecovirus vaccine. Here, I will describe my work identifying a Class 1/4 antibody and obtaining its structure at 3.4 Å in complex with the SARS-CoV-2 S 6P trimer (Hsieh et al., 2020).

2. Identifying a Class 1/4 antibody

Identifying a Class 1/4 antibody, or indeed any broadly neutralizing antibody, is a varied and multi-step process. For this study, we performed three sets of experiments. First, our collaborators screened a panel of antibodies via a neutralization assay to identify antibodies that neutralize multiple strains of SARS-CoV-2, and in some cases SARS-CoV as well. Then, we assayed test IgGs against Fabs of known binding sites to assess their binding epitope in a competition enzyme-linked immunosorbent assay (ELISA). Finally, we assayed a select group of test Fabs derived from the IgGs for binding to RBDs from different sarbecovirus clades using an indirect ELISA. These RBDs show greater variation than was tested in the neutralization assays to assess the breadth of neutralization for each antibody.

2.1. Screening for mAbs that broadly neutralize SARS-CoV-2 variants of concern

Our collaborators in the Hahn lab at the University of Pennsylvania performed neutralization assays as per Schmidt *et al.* (2020) and Liu *et al.* (2021) on a panel of antibodies collected at three timepoints from a single donor. The first collection was in March 2020, shortly after the patient recovered from infection with SARS-CoV-2. The second collection occurred several months later, but before the patient was vaccinated – these antibodies are tagged “pv” for “pre-vaccination”. The third and final collection occurred a few weeks after the patient was vaccinated, and their names feature the tag “v”.

Neutralization assays test how efficiently antibodies prevent viral entry into cells. Here, pseudotyped, replication-deficient lentivirus containing a luciferase reporter gene was constructed to express the SARS-CoV-2 S protein on its surface. Pseudoviruses were then incubated with serial dilutions of monoclonal antibodies (mAbs) to allow mAb binding to equilibrate. Subsequently, this pseudovirus-mAb mixture was added to HEK-293T cells expressing surface ACE2. If an antibody is not neutralizing and pseudoviruses can enter the

cell, cells will translate the pseudovirus backbone and express the luciferase reporter gene. The more neutralizing an antibody is, the less luciferase expression is seen. After two days of infection, cells were lysed and the lysate was incubated with a luciferase substrate. Luminescence was then read and calibrated against a positive control infected with pseudovirus treated with no mAbs, considered to represent “100% infectivity”. In Figure, IC50 values of mAbs are reported and color-coded according to their neutralization potential: red is highly neutralizing, yellow moderately neutralizing, and green or white the least neutralizing.

2.2. Antibody classification via competition ELISA

We then tested these IgGs against Fabs of known binding epitope using a sandwich ELISA. In this assay, capture Fabs of known class are bound to a plate and then incubated with RBD. IgG from the test antibodies is then added to each well, acting as a primary antibody, and a secondary antibody conjugated to horse radish peroxidase (HRP) against human Fc is used as the fluorescent probe (FigureA). If the test IgG shares a binding epitope with the capture Fab, no binding will be detected: if it has a lower binding affinity than the capture Fab, the IgG will remain in solution and will be aspirated off the plate before incubation with secondary antibody, and if its affinity is higher, the test IgG may bind the RBD and both will be lost to aspiration. Binding will only be detected when the capture Fab and test IgG can bind the RBD simultaneously. We have set the threshold for “bound” versus “unbound” to be around 200,000 relative light units (RLU). Red designates cases where IgGs do not bind RBD, signifying competition with the capture Fab.

Antibodies were tested against themselves as controls and show an almost entirely reciprocal pattern of competition (FigureB). This indicates that our controls compete with themselves for the same epitopes, which is what we would expect. Additionally, these controls recapitulate previous data demonstrating their epitopes. For example, a structure of S309(Pinto et al., 2020) showed that it is a Class 3 antibody (Barnes, West, et al., 2020), and

here it is shown that S309 competes with itself and other known Class 3 antibodies like C135 and C110 (Barnes, West, et al., 2020). This also recapitulates previous observations from structural data that may otherwise be surprising: for example, the epitope recognized by C110, a Class 3 antibody, includes residues in the Class 2 epitope, including E484 and Q493, explaining its competition with Class 2 C144.

From this panel of IgGs, we selected four antibodies to proceed with based on data from the neutralization assay and from the competition ELISA. 11G-12v was selected because it competes with both Class 1 and Class 4 Fabs, suggesting it may be a Class 1/4 antibody. Three other test IgGs (11G-13v, 11G-14v, and 11G-20v) were also selected for further investigation, but will not be discussed further in this study.

2.3. Cross-reactive Fab ELISA

We tested 11G-12v Fab against RBDs from sarbecoviruses representing two clades using an indirect ELISA. Here, RBD is immobilized on the plate, and test Fabs are added. After incubation, unbound Fab is removed via aspiration, and then incubated with secondary antibody conjugated to HRP (FigureA). Following another incubation period, excess secondary is removed and luminescence is measured. Data are presented in relative light units (RLUs) without normalization, as an “area under the curve” (FigureB). 11G-12v, our candidate Class 1/4 Fab, can then be compared to C118, which was previously characterized as Class 1/4 (Fan et al., 2022; Jette et al., 2021b). Our assay shows that C118 is broadly cross-reactive as previously found, binding to all tested RBDs. 11G-12v shows a comparable degree of cross-reactivity (FigureB). This indicates that the epitope bound by 11G-12v may be well-conserved across tested RBDs.

To better understand patterns of conservation in the sarbecovirus RBD, structures of all RBDs tested (except Pang17, whose structure is unavailable) were compared using the ConSurf server (FigureC). This tool uses sequence data to assess residue conservation in a

protein, then codes the degree of conservation to a nine-tone divergent colormap to indicate sequence conservation visually on the input structure. Darker pink indicates greater conservation across aligned RBDs, while darker blues indicate greater sequence variability. The C118 Fab binds an epitope that is well-conserved across tested RBDs (FigureD). 11G-12v, therefore, may be expected to bind a similarly well-conserved epitope. It is, however, worth mentioning that cross-reactivity can depend on a number of features other than the degree of conservation of an epitope, such as the ability of the antibody paratope to accommodate non-conserved residues or bind to altogether unrelated epitopes (Frank, 2002).

2.4 Cross-reactive Class 1/4 IgG ELISA

11G-12v IgG was compared to known Class 1/4 IgGs C118, M8a-3, and S2X259 for cross-reactivity against RBDs derived from sarbecoviruses belonging to clades 1a, 1b, 2 and 3 using an indirect ELISA as in the previous section (Figure 7). 11G-12v IgG binds more tightly to clade 1a and clade 2 RBDs than to clade 1b or clade 3 RBDs, and overall showed moderate to high affinity to all but two tested RBDs (SARS-CoV-2 BA.1, BM4831). This pattern is similar to that seen for C118 IgG, which shows better binding across the board than 11G-12v IgG, but has a higher rate of binding to members of clade 1b. S2X259 binds tightly to SARS-CoV-2 Delta, Pang17 and RaT13G RBDs but is a poor binder to other clade 1b and all tested clade 2 RBDs, while M8a-3 also shows some specificity to SARS-CoV-2 Delta RBD and clade 1a members SHC014 and LyRa3 RBDs. This shows that binding site alone does not determine cross-reactivity: this is dependent on other features of binding as well.

3. Structural characterization of 11G-12v via cryo-EM

After identifying 11G-12v as a putative Class 1/4 antibody, I then proceeded with structural characterization to describe the specific manner in which 11G-12v binds the RBD in the context of the S trimer. Due to the large size of the S trimer and the probable heterogeneity of the sample, single-particle cryo-electron microscopy (cryo-EM) was the best method for initial structural characterization of this complex. Subsequent characterization via crystallography of the Fab-RBD interaction will be necessary to obtain a map of sufficiently high resolution to describe the interaction between the Fab and the RBD. There were three stages of this process: sample preparation, data collection, and data processing. These data were then used to construct an electron density map, which can be fitted with an atomic model.

3.1. Sample preparation, data collection, and data processing

Fab and S proteins were expressed and purified as described previously (Barnes, West, et al., 2020; Cohen et al., 2022; Fan et al., 2022; Jette et al., 2021a). Briefly, Expi293 cells were transiently transfected with 1 μ g plasmid DNA (encoding the His-tagged proteins of interest) per mL cell culture and treated according to the manufacturer's instructions (ThermoFisher). Supernatant was isolated and protein purified via Ni^{2+} immobilized metal affinity chromatography (IMAC). Eluent was then purified via size exclusion chromatography. Fabs were purified over a Superdex 200 Increase 10/300 GL column (GE Healthcare), while S protein was purified first over a Superdex 200 column followed by purification using a Superose 6 Increase 10/300 GL (GE Healthcare) column to exclude aggregates. Fractions of purified Fab were combined and concentrated to between 2-15 mg/mL, while S protein fractions were stored separately and not subject to concentration. All proteins were stored at

4 °C. The sample of S on which these data were collected was purified only days prior to grid preparation.

To set up complexes for cryo-EM, Fab and S protein were diluted such that the final concentration would come to 2 mg/mL. Diluted Fab was added to Spike with 2.4-fold molar excess of Fab relative to S protomer. This was allowed to incubate at room temperature for 35 minutes to allow Fab-RBD association to achieve equilibrium. During this time, Quantifoil R1.2/1.3 copper grids were glow discharged. Immediately before adding sample to grids, fluorinated octyl maltoside (FOM) detergent (Anatrace) was added to the sample to a final concentration of 0.02%. 3 μ L of sample was deposited on grids in a Mark IV Vitrobot (Thermofisher) with a blot time of 4 seconds and a blot force of 0, and then plunge frozen in liquid ethane.

Data were collected on a Titan Krios (FEI) at 300 keV in superresolution, with a raw pixel size of 0.416 Å and 60 e/Å² total exposure. Data collection yielded 12,212 movies (Figure 8 A), which were processed using CryoSPARC. An overview of the data processing workflow is shown in Figure 8 E. Briefly, after patch motion correction, CTF alignment and selection of 11,556 movies for further processing, 2,785,760 particles were selected using blob picker (box size = 456 (228)) and extracted with a bin size of ½. 2D class averages (Figure 8 B) were computed on all particles, and then were recomputed for the subset of particles in good classes, yielding 777,793 good particles. A preliminary round of 3D *ab initio* classification was performed, requesting 5 *ab initio* models. Particles from the best *ab initio* class were then re-extracted at bin = 1 and box size = 456. These 345,917 particles were subject to further 2D class averaging, and 341,146 particles were accepted. Two *ab initio* models were requested, and the 313,607 particles comprising the selected model were subject to a further round of *ab initio* modeling. For this final *ab initio*, the maximum resolution was set to 3 Å, compared to the default of 12 Å. Non uniform refinement was performed on the resulting map using C1 symmetry, and a 3.4 Å map of the complex was obtained (Figure 8, C and D).

3.2. Single-particle data analysis

Local resolution calculations were performed using Phenix (v 1.20.1-4487-000). Half maps of the density generated in CryoSPARC were processed with a box size of 24 grid units, and the resulting data were mapped onto the density using ChimeraX (v1.5) (Figure 8 C). Due to conformational heterogeneity of the RBD in the “up” conformation, the RBD-Fab binding region has the lowest average resolution of the map, around 5 Å, while the S2 subunit, for example, has a local resolution close to 2 Å. The obtained volume may be seen in Figure 9 A, where gray indicates the S trimer, teal indicates glycan densities, dark purple indicates the heavy chain and light purple the light chain. This color map reflects colors chosen for the fitted model and was created using the “color near atoms” feature in the ChimeraX (v1.5) GUI.

A model of the S trimer in complex with Fab 11G-12v was assembled in ChimeraX (v1.9) using a combination of the experimentally derived structure of the S trimer with C118 Fab (PDB Id: 7rkv) and the Fab chains of an AlphaFold 3 model of 11G-12v in complex with the RBD (S residues 330-530). Sections of a full-length structure of the S trimer made using AlphaFold 3 were further used to build into regions of the map where 7rkv had no model. First, the experimental model 7rkv was fitted into the density using the “fitmap” command, and the C118 Fab models removed. Then, residue-by-residue refinement was performed in ISOLDE (v 1.7) to fit the remaining model into the map. Once satisfied with the fit, three copies of the AlphaFold 3 model of the RBD with 11G-12v were imported and aligned to the 7rkv RBDs. The AlphaFold RBD was then deleted, leaving behind only the Fab in approximately the correct position. Fab chains were refined residue-by-residue using ISOLDE. Then, global refinement of the structure was performed using Phenix real space refinement (v 1.21.2-5419). The target RMSD for bonds was set to 0.0001 and to 0.001 for angles to reduce geometry violations, and secondary structure restraints were employed. The resulting model was further fit to the density in ChimeraX (v 1.9). Structural features in the model that extended beyond the map at a reasonable contour were deleted in and features of the S N-terminal domain were built into unmodeled density using an AlphaFold 3 model of

the S trimer (Figure 9 B). Additionally, N-linked glycans were modeled into the appropriate unmodeled densities, such as N343 of each RBD. Chains were reconnected where necessary and the model was passed through Phenix real space refinement a final time for validation. Final model statistics may be found in Table 1.

3.3. Analysis of 11G-12v Fab recognition of the S RBD

Sequence analysis of 11G-12v using the IGBLAST server suggests that the variable region of the heavy chain is derived from genes IGHV3-21, IGHD3-22, and IGHJ4. The CDRH3 loop is 20 residues in length, which is substantially longer than the mean CDRH3 length in humans. Additionally, loop residues ⁹⁹YYDRSG_{100D} are consistent with the YYDRxG motif which has been found in common with a number of other cross-reactive RBD-targeting antibodies, including C022 and COVA1-16 (Figure 10 A). This motif is derived from the second reading frame of the D gene IGHD3-22, and the nucleotide sequence of 11G-12v shows only 3 bases different from the germline gene, of which two are silent mutations, suggesting that only limited somatic hypermutation is necessary for this sequence to arise. Together, these three features are characteristic of a group of broadly neutralizing SARS-CoV-2 antibodies (L. Liu et al., 2022).

The YYDRSG motif of 11G-12v is found at the apex of the CDRH3 loop (Figure 10 B). The peptide backbone of the 20-residue loop forms a beta hairpin, extending the five-strand beta sheet in the center of the RBD and stabilizing the Fab-RBD interaction through backbone-mediated hydrogen bonding. This corroborates what has been observed in other anti-RBD antibodies containing this motif (Burnett et al., 2021; L. Liu et al., 2022). Interactions between side chains conserved in the motif and RBD side chains add an additional layer of stability (Figure 10 B). Y99, the first tyrosine in the motif, coordinates RBD residue K378 with the assistance of CDRH3 loop residue I100G, which falls outside the conserved motif (Figure 10 C, left panel). This coordination is also seen in COVA1-16 between K378, Y103 and Q100G. R100B, which appears at the apex of the CDRH3 loop, contacts the backbone

of conserved RBD residue S371 (Figure 10 C, middle panel). This interaction is particularly interesting, as the single mutation of S371 to F/L reduces SARS-CoV-2 ACE2 binding and infectivity by 5- to 10-fold (S. Liu et al., 2025). This mutation is seen in the Omicron variant, along with its corresponding compensatory mutation, S375F, and others. 11G-12v Y100, second in the motif, contacts RBD S383, which is both a conserved residue across sarbecoviruses and shows little variation in SARS-CoV-2 variants (Figure 10 C, right panel) (Table 2, data from GISAID (Shu & McCauley, 2017)).

As stated above, a number of other Class 1/4 antibodies also include this sequence motif and bind to the RBD in a similar manner to 11G-12v, extending the internal RBD beta sheet. However, the mode of CDRH3 loop interaction is not necessarily determinative of the antibody's angle of approach, which is a major contributor to the ability of the antibody to prevent receptor recognition. The binding footprint and angle of approach were compared between 11G-12v and C118, C022, M8a3, S2X259, and COVA1-16 (Figure 11).

Fabs sharing the YYDRSG motif share a very similar, but not identical, footprint on the interior face of the RBD. Figure 11, first and second rows, show the RBD residues that form contacts with Fab residues. The heavy chains of 11G-12v, COVA1-16, and C022 Fabs bind lower on the face, slanting upward in the direction of the high point of the RBM, thus following the distribution of conserved residues (shown in dark purple on the conservation plots, Figure 11). The light chains of these Fabs associate minimally. C118, whose CDRH3 loop associates in a very similar manner to the YYDRSG Fabs, shows more contacts from the light chain, and the heavy chain footprint is rotated slightly to be more perpendicular to the RBM. M8a-3 and S2X259 use a dramatically different binding strategy, associating primarily with one side of the RBD interior face. The heavy chain footprints are less extensive than either C118 or the YYDRSG-containing Fabs, and the light chain associations are still relatively minimal, contacting only around three residues.

These footprints contribute to the angle of approach for each Fab. To see how this might affect ACE2 binding, RBDs of Fab-RBD/S trimer experimental models were aligned and the position of each Fab compared. Fabs containing the YYDRSG motif showed much greater

correlation in the positions of their constant domains, suggesting that the motif stabilizes Fabs in a relatively narrow range of approach angles (see Figure 11, bottom left panel). This is reflective of the fact that the CDRH3 loop of YYDRSG Fabs provides two beta strands to extend the RBD beta sheet, reducing the degrees of freedom of the Fab. To observe how this impacts association with ACE2, an experimental model of ACE2 binding to the RBD (PDB ID: 6VW1) was used to observe the clash between the antibody and ACE2. None of the footprints of the analyzed antibodies impinge upon the ACE2 binding ridge (Figure 11, second row). However, the bulk from the constant domains is oriented such that the bound Fab sterically blocks ACE2 access to its binding region, the feature which renders these antibodies Class 1/4. YYDRSG-containing Fabs, which are held at a more acute angle to the RBD, clash significantly with ACE2, with the most clashes shown by 11G-12v (1318), then COVA1-16 (1180), then C022 (782). The angle C118 creates with the RBD is wider (Figure 11, middle panel), moving it away from where ACE2 would be compared to YYDRSG Fabs, and showing 715 clashes with ACE2. Structures of M8a-3 and S2X259, unfortunately, do not show the constant domains of the heavy or light chains; however, even without them, the M8a-3 variable domains form 971 clashes with ACE2, while the S2X259 variable domains form 303 ACE2 clashes. These Fabs twist down and to the side relative to the YYDRSG Fabs.

4. Discussion

Together, these results place 11G-12v into this group of Class 1/4 antibodies and support the previous findings about YYDRSG motif-containing antibodies. Like other YYDRSG-containing antibodies, 11G-12v shows a long CDRH3 that extends the beta sheet at the center of the RBD, and its angle of approach matches that seen in other YYDRSG-containing antibodies COVA1-16 and C022, which clashes substantially with ACE2 binding. 11G-12v Fab and IgG also show broad cross-reactivity to RBDs from four sarbecovirus clades, though

the strength of binding is not as high, overall, as C118. Additionally, the YYDRxG motif can be formed from multiple heavy chain genes and does not require a high degree of somatic hypermutation (L. Liu et al., 2022). Thus, 11G-12v and other YYDRSG-containing antibodies demonstrate a promising group of Class 1/4 antibodies to protect against evolving strains of SARS-CoV-2. However, YYDRxG antibodies have been noted to be found at relatively low abundance in serum, possibly due to danger of self-reactivity of long CDRH3s (Kaplinsky et al., 2014; L. Liu et al., 2022; Wardemann et al., 2003; Wu et al., 1993).

Though the average resolution of the S protein-11G-12v Fab structure presented here averages 3.4 Å, the Fab-RBD interface is at a lower resolution than this. This is due to the flexibility of the position of the RBD in the “up” conformation. Obtaining a high-resolution structure of this interaction will likely require crystallographic methods, using purified RBD and 11G-12v Fab. Additionally, there may be some interest in other Fabs from this study, and some work has already been done toward those (specifically, 11G-13v, 11G-14v, and 11G-20v). The structure of the 11G-12v Fab in complex with the S trimer helps shed light on the ways in which Class 1/4 antibodies can bind to the SARS-CoV-2 RBD and help us consider how best to design vaccines that elicit these antibodies. In combination with known Class 1/4 antibodies, it may reveal features of potent, broadly neutralizing antibodies that can then be used therapeutically to treat acute infection.

Figures

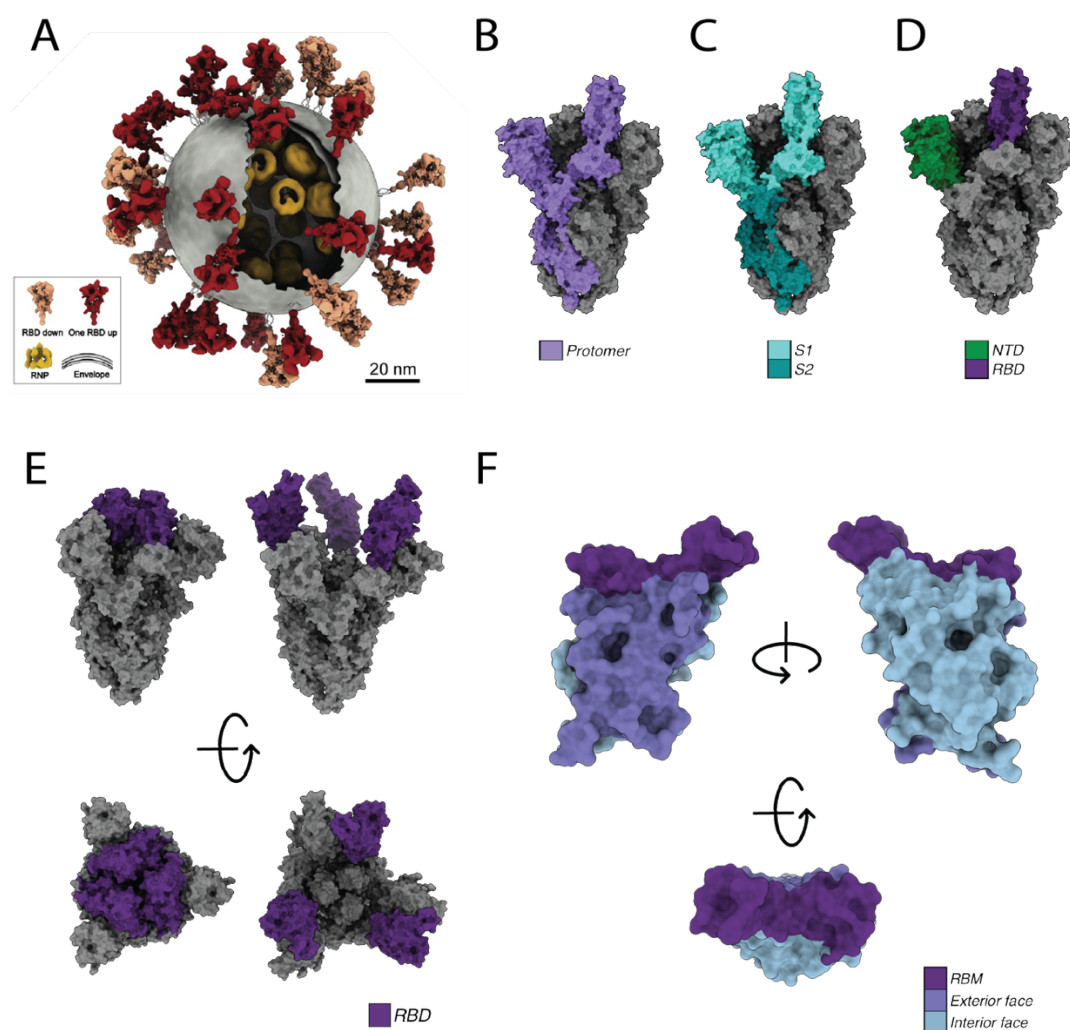


Figure 1. Spike protein overview. (A) Molecular architecture of authentic SARS-CoV-2 virion showing the S protein arrayed on the surface, from Yao et al. (2020)(Yao et al., 2020). (B) S protein, protomer highlighted in purple showing one up RBD. (C) S protein, S1 and S2 domains highlighted in lighter turquoise (S1) and darker teal (S2). (D) S protein, NTD and RBD highlighted in green and dark purple. (E) S trimer demonstrating RBD conformations, all down vs all up. RBD is in purple. (F) RBD details, highlighting the RBM (dark purple), exterior (periwinkle) and interior (sky blue).

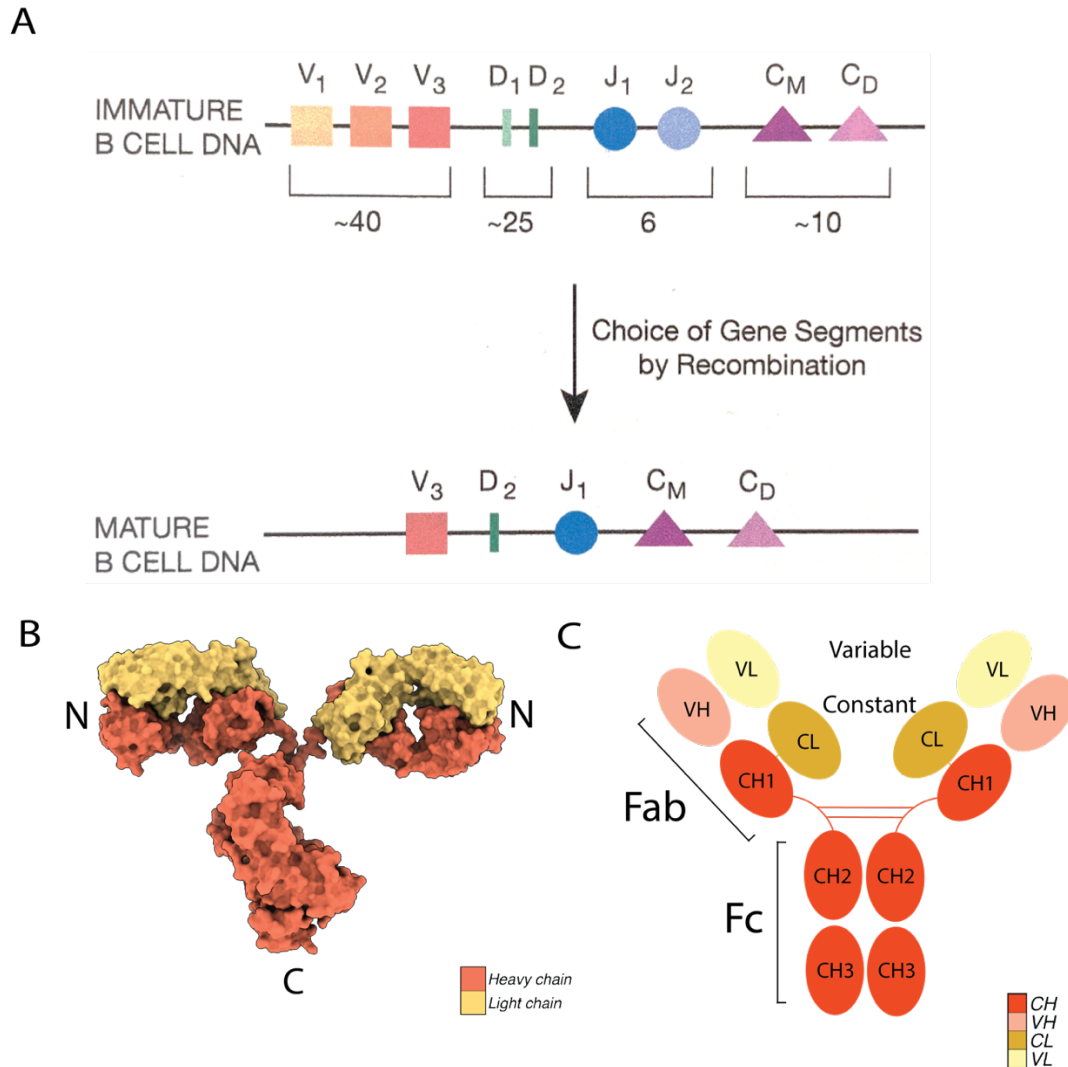


Figure 2. Antibody overview. (A) Cartoon depicting recombination process of antibody gene segments that occurs during B cell maturation. From Sompayrac, 2019(Sompayrac, 2019). (B) Antibody structure (PDB ID: 1igt). Heavy chain shown in orange, light chain in yellow. N and C termini labeled. (C) Antibody schematic. Inter-chain disulfide bonds are shown as lines; intradomain disulfide bonds not shown.

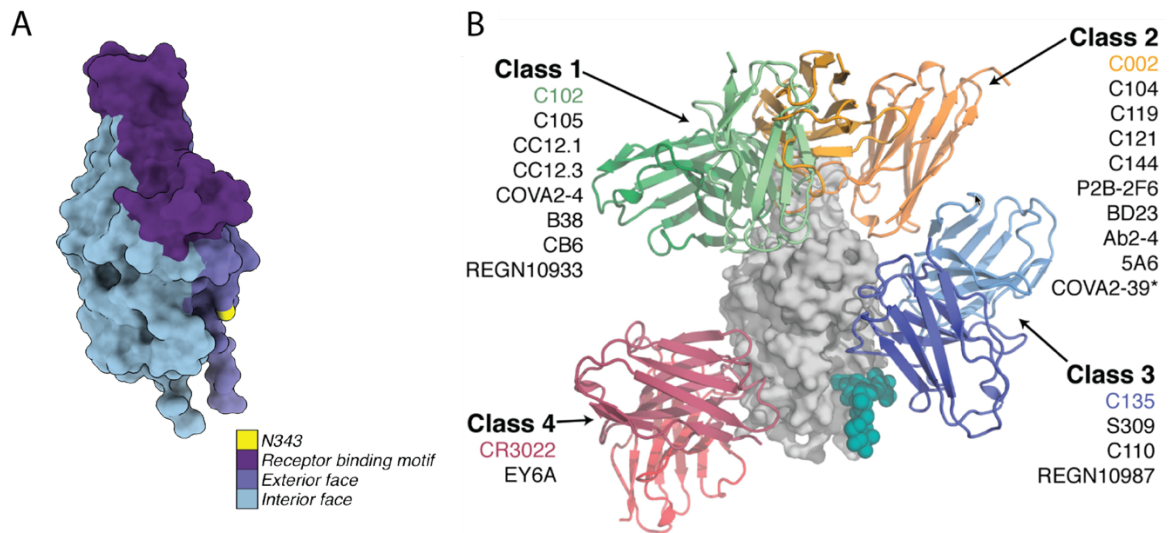


Figure 3. RBD details. (A) RBD colored as above, shown in the same orientation as in panel B, for ease of interpretation. (B) Figure from Barnes et al. (2020, Nature)(Barnes, Jette, et al., 2020) showing different classes of RBD-binding antibodies.

mAb ID	VH Gene	%SHM	D614G	Beta	Delta	Alpha	Gamma	Omicron	SARS1
11G-3v	IGHV1-2	8.68	2.454	98.31	5.183	2.096	73.94	>1000	>1000
11G-4v	IGHV3-74	12.5	4.73	490.3	1.375	0.9198	495.2	6.051	>1000
11G-8	IGHV3-9	0.69	4.665	6.723	5.481	2.435	8.033	104.3	>1000
11G-9v	IGHV3-9	3.12	2.685	43.66	8.856	3.311	22.98	187.7	>1000
11G-10v	IGHV1-2	5.9	11.41	19.89	5.402	3.011	11.38	9.454	>1000
11G-12v	IGHV3-21	7.99	43.48	34.73	71.99	28.3	25.05	173.2	7.326
11G-13v	IGHV1-2	9.03	2.239	2.949	4.058	1.89	0.5564	26.29	>1000
11G-14v	IGHV1-46	10.42	639.4	394	228.6	216	368.9	629.9	>1000
11G-16v	IGHV1-8	4.51	5.073	268.5	5.032	2.516	81.59	120.5	>1000
11G-19pv	IGHV3-66	4.91	17.39	9.224	19.31	9.136	2.274	55.05	>1000
11G-20v	IGHV3-66	6.67	18.3	3.477	19.15	62.4	1.281	5.224	>1000
11G-21	IGHV3-66	0.35	43.86	3.514	10.22	51.9		>1000	>1000
11G-25v	IGHV3-53	7.02	18.47	5.066	18.55	8.512	5.695	35.74	>1000
11G-26pv	IGHV3-53	7.02	36.95	23.12	57.31	12.71	13.24	80.86	>1000
11G-27v	IGHV3-53	6.67	22.85	4.381	20.04	11.65	4.261	18.2	>1000
11G-28pv	IGHV3-53	5.61	78.83	20.07	73.55	23.26	12.75	61.14	>1000



 Neutralizing Non-neutralizing

Figure 4. Neutralization data for 16 mAbs cloned in the Hahn lab. mAb ID is the name given to the monoclonal antibody. VH gene indicates the gene whence the antibody derives. %SHM indicates the percentage of the gene composed of somatic hypermutations. D614G, Beta, Delta, Alpha, Gamma, and Omicron are SARS-CoV-2 variants of concern from the first two years of the pandemic. SARS1 is SARS-CoV. IC50 values are shown in ng/mL. Red represents high neutralization, orange/yellow medium neutralization, and green/white indicates low to no neutralization.

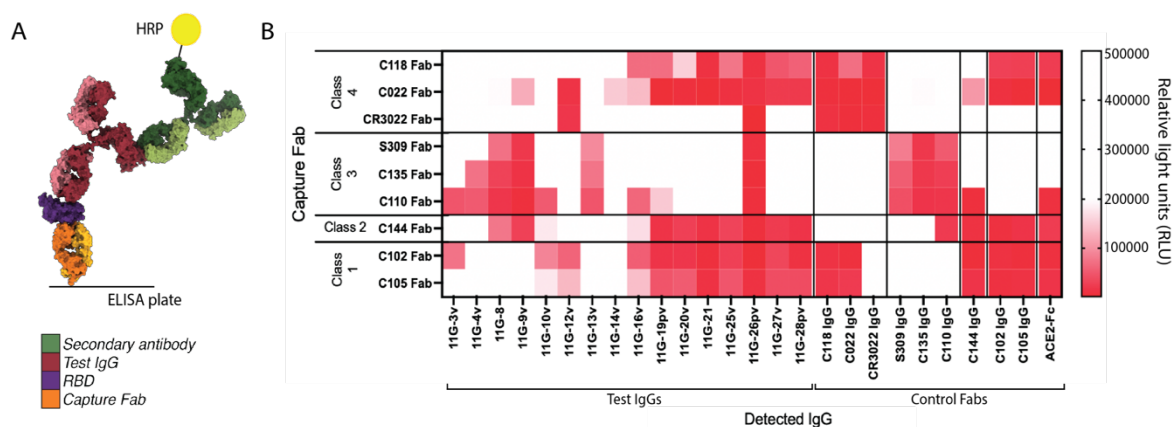


Figure 5. Competition ELISA data. (A) Representation of the competition ELISA. For Fab and antibodies, light chains are shown in lighter shades of the colors in the key. HRP is horse radish peroxidase. (B) Data from competition ELISA. Competition is shown in red. Capture Fabs from each class are represented in rows, and test IgGs and controls shown in columns. Units are Relative Light Units.

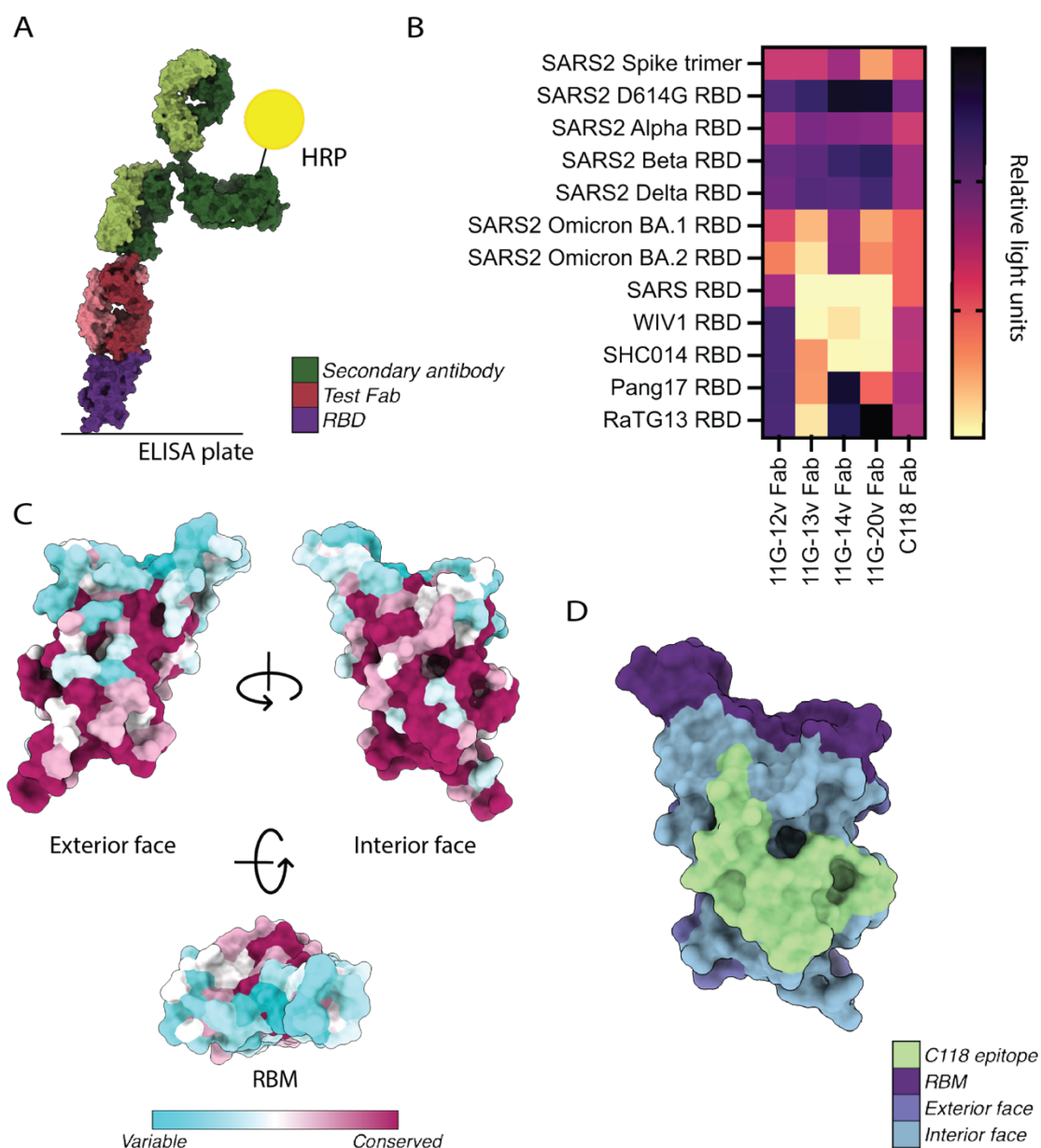


Figure 6. Indirect ELISA data. (A) Representation of the indirect ELISA. For Fab and antibodies, light chains are shown in lighter shades of the colors in the key. HRP is horse radish peroxidase. (B) Indirect ELISA data for 11G-12v Fab, other Fabs screened, and control C118, a Class 1/4 antibody. RBDs used shown in rows, Fabs shown in columns. Units are relative light units, and lighter colors depict less binding. Data are not normalized. (C) Colormap on the SARS-CoV-2 D614G RBD showing degree of conservation across

RBDs in panel B made using ConSurf server and ChimeraX (v1.3). Magenta residues are conserved, cyan residues are variable. (D) RBD colored as above, showing C118 epitope in mint green.

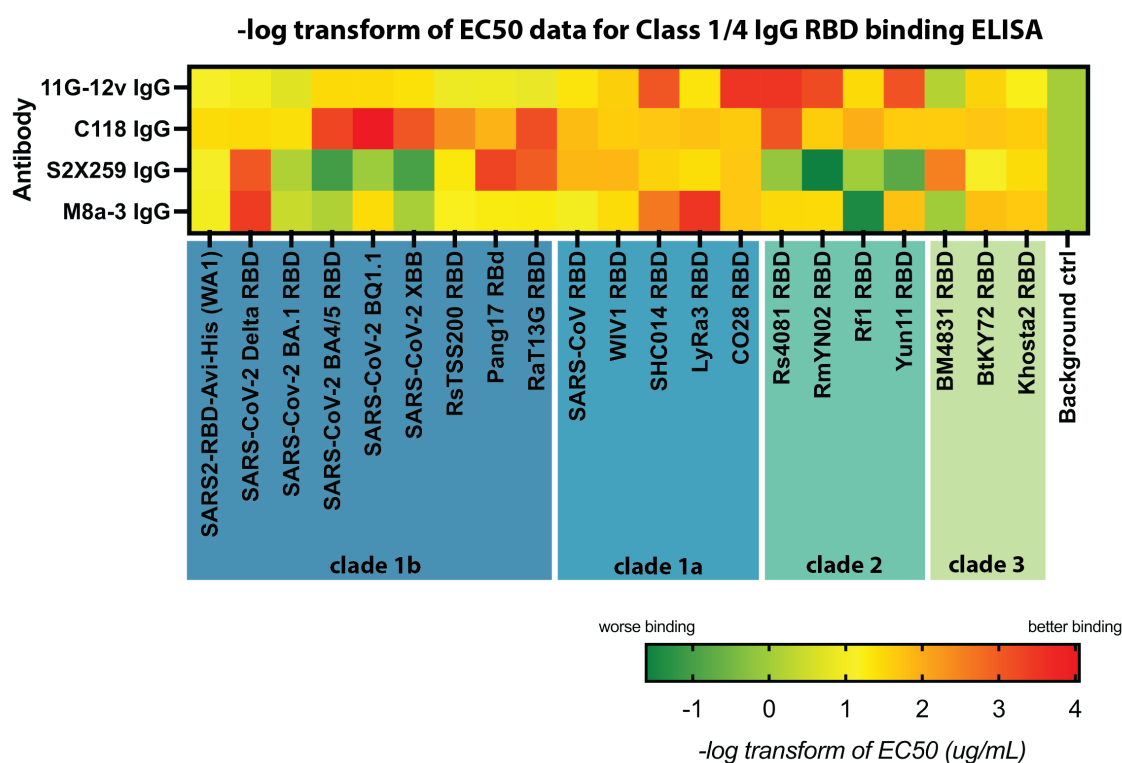


Figure 7. Cross-reactive Class 1/4 IgG ELISA. Binding of IgGs 11G-12v, C118, S2X259 and M8a-3 are compared across four clades of sarbecovirus (clade 1b, clade 1a, clade 2 and clade 3). Worse binding is depicted in green, while better binding is shown in yellow turning to red.

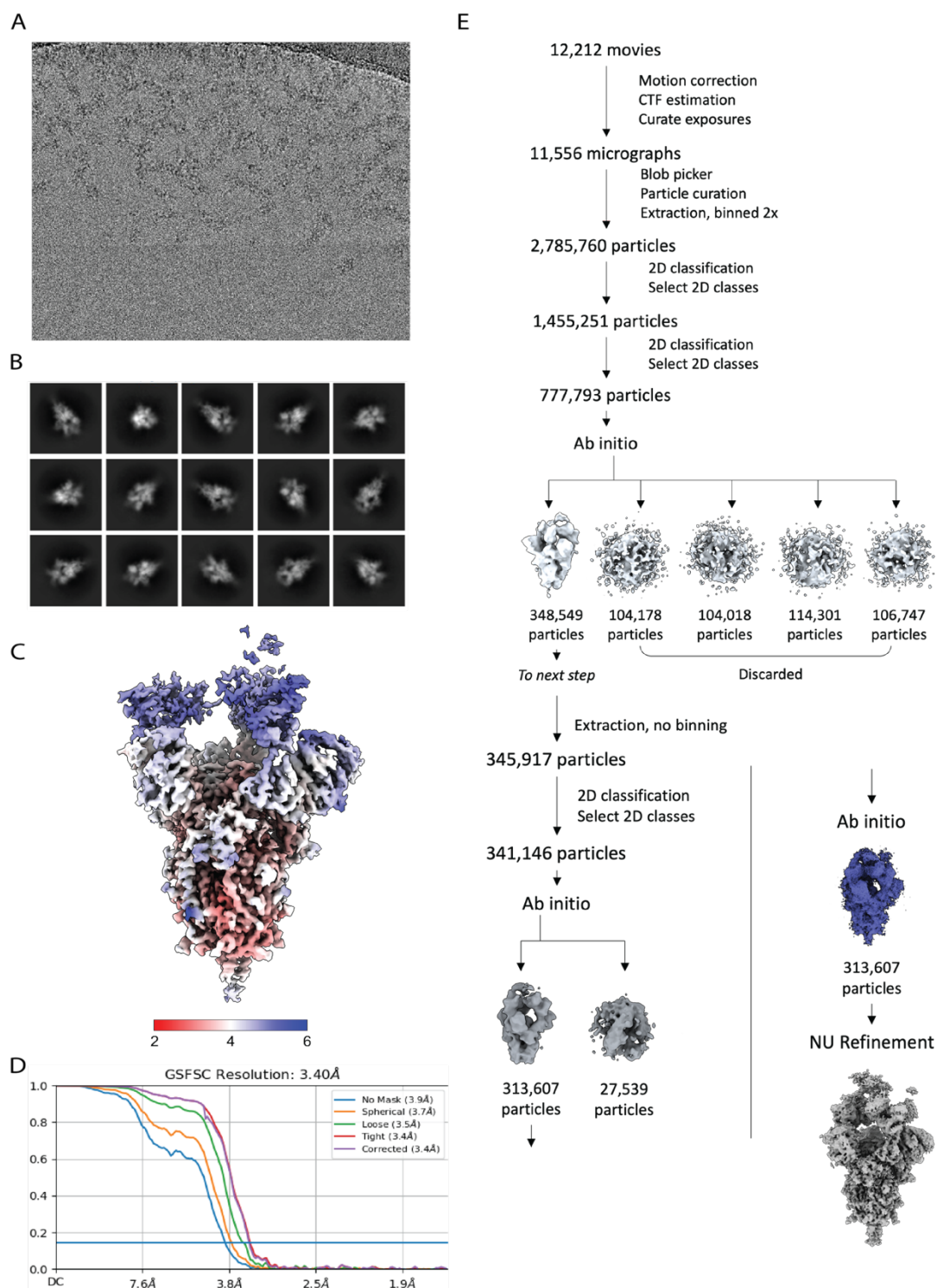


Figure 8. Cryo-EM structure information. (A) Representative micrograph. (B) Representative 2D class averages. (C) 3.4 Å structure refined with C1 symmetry. Colormap (created using Phenix local.resolution (v1.20.1-4487-0000)) shows resolution across map, with 2 Å resolution shown in red, 4 Å in white, and 6 Å in dark blue. (D) Gold-standard Fourier shell correlation (GSFSC) plot. (E) Workflow of single-particle data processing in cryoSPARC.

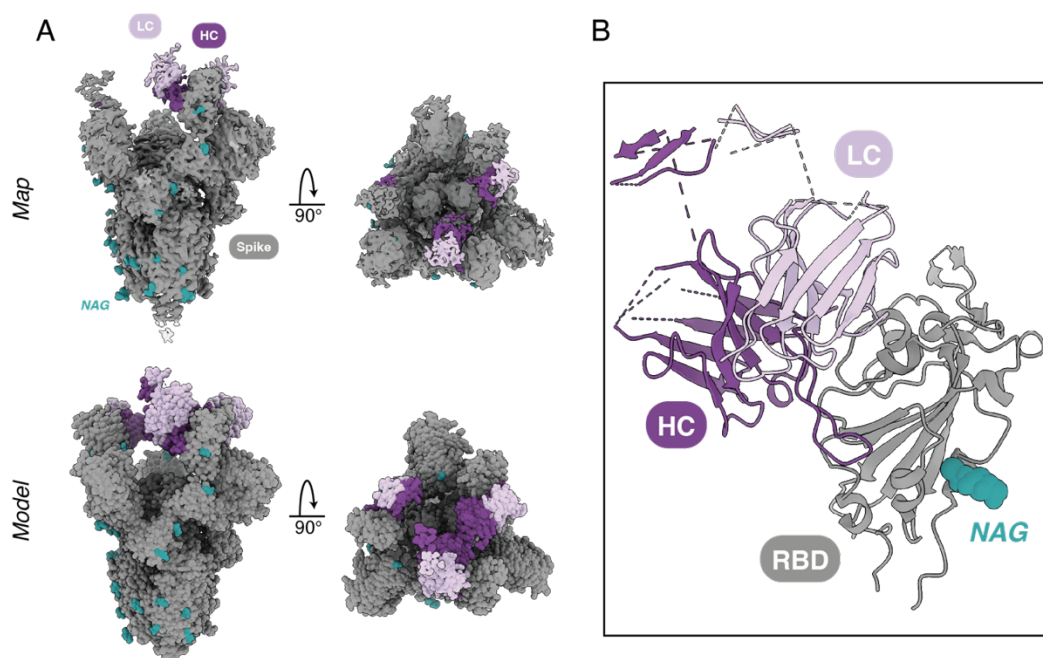
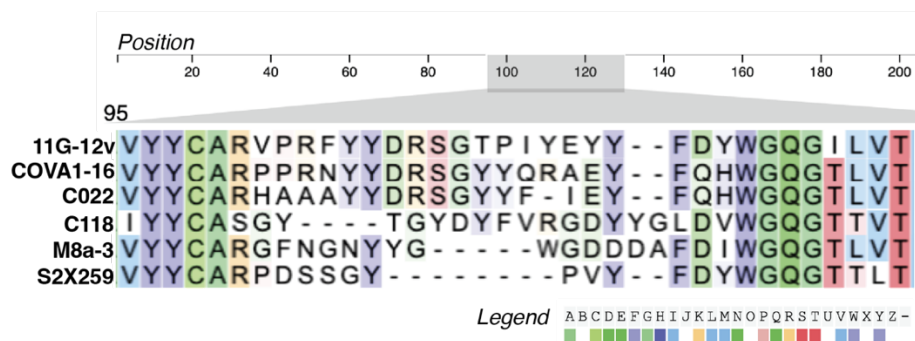
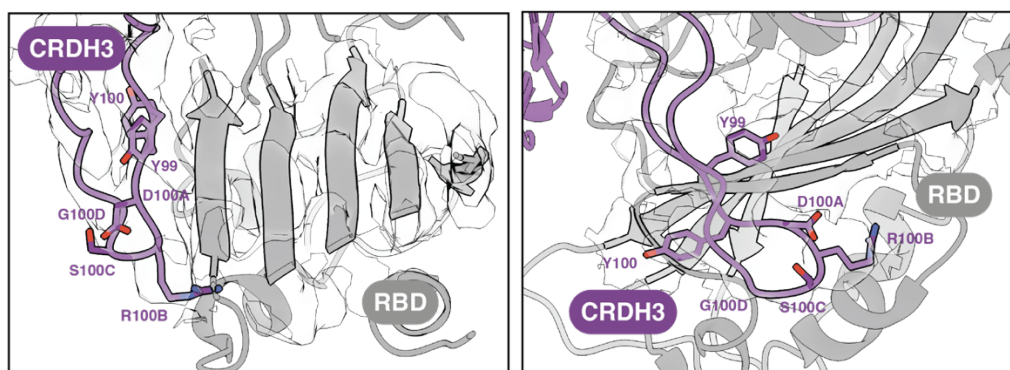


Figure 9. Map and model of S trimer in complex with 11G-12v Fab. (A) Map and model of S protein in complex with three 11G-12v Fabs. Map is shown at a high contour. Proteinaceous S protein elements are colored gray, NAG moieties are colored teal. Fabs are colored purple, with a dark heavy chain and a pale light chain. Side and top views are shown, with the map views on top and the model views on bottom. (B) Close-up of a cartoon representation of the model of the RBD bound to the best-resolved Fab. Coloring is as in (A).

A



B



C

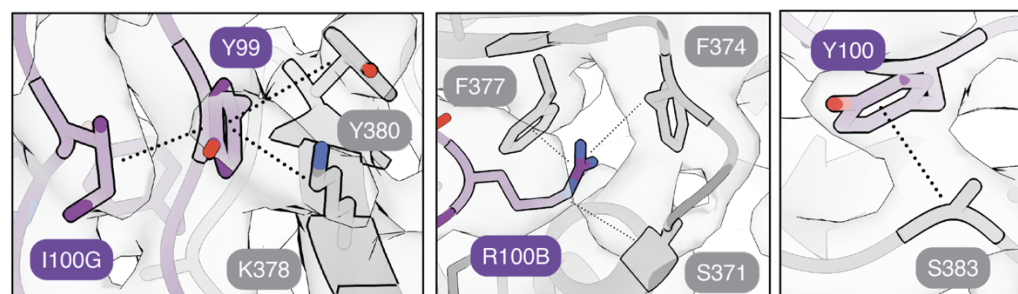


Figure 10. Association of YYDRSG motif with RBD beta sheet. (A) Sequence alignment of the CDRH3 loops of three YYDRSG-motif Fabs (11G-12v, COVA1-16, and C022), beta-strand forming C118, and two alternate strategy Fabs (M8a-3 and S2X259). Alignment created using Clustal Omega. (B) Close-up of the RBD beta sheet and 11G-12v CDRH3 extension of it, with the YYDRSG motif highlighted. Two views are shown: from the top (left) and side (right). (C) Close-up of interactions of RBD residues with YYDRSG-motif residues in 11G-12v.

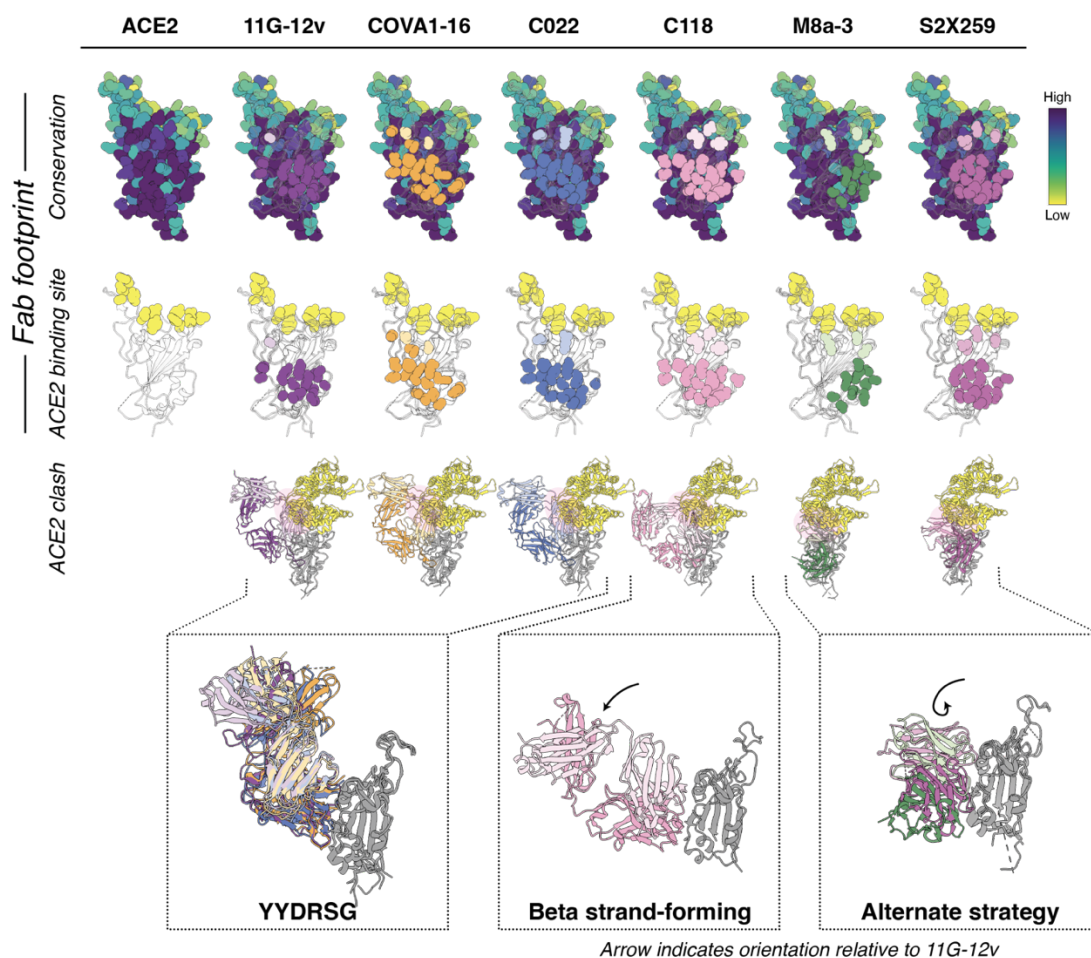


Figure 11. Comparison of footprints and approach angles of six Fabs. Top row: Fab footprints overlaid on sequence conservation plots. Second row: Fab footprints overlaid on ACE2 binding site (RBM). Third row: Angle of approach for Fabs. Bottom row: Consolidated angles of approach per category, with arrows to denote the deviation from the position of YYDRSG-containing Fabs. PDB IDs: ACE2-RBD – 6VW1; COVA1-16 – 8YUC; C022 – 7RKU; C118 – 7RKS; M8a-3 – 7UZ4; S2X259 – 7RAL.

Table 1: Cryo-EM model refinement statistics

Model	
Composition (#)	
Chains	15
Atoms	49976 (Hydrogens: 23969)
Residues	Protein: 3421 Nucleotide: 0
Water	0
Ligands	NAG: 36
Bonds (RMSD)	
Length (Å) (# > 4 σ)	0.002 (0)
Angles (°) (# > 4 σ)	0.827 (79)
MolProbity score	1.09
Clash score	1.9
Ramachandran plot (%)	
Outliers	0
Allowed	2.71
Favored	97.29
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 3288)	0.99 (0.15)
helix (N = 640)	2.51 (0.21)
sheet (N = 847)	0.64 (0.17)
loop (N = 1801)	0.04 (0.16)
Rotamer outliers (%)	0.04
C β outliers (%)	NA
Peptide plane (%)	
Cis proline/general	0.6/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	2.08
ADP (B-factors)	
Iso/Aniso (#)	26007/0
min/max/mean	
Protein	3.15/238.71/94.24
Nucleotide	---
Ligand	46.17/196.74/107.76
Water	---
Occupancy	

Mean	1	
occ = 1 (%)	100	
0 < occ < 1 (%)	0	
occ > 1 (%)	0	
Data		
Box		
Lengths (Å)	145.60, 148.93, 198.85	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	3.4	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	3.6/---/---	3.5/---/---
d model	3.5	3.5
d FSC model (0/0.143/0.5)	3.3/3.3/3.5	3.3/3.4/3.7
Map min/max/mean	-58.490566	
Model vs. Data		
CC (mask)	0.82	
CC (box)	0.73	
CC (peaks)	0.65	
CC (volume)	0.79	
Mean CC for ligands	0.63	

Table 2: Number of sequences per amino acid substitution for residues within the RBD contacting the YYDRSG motif in 11G-12v

<i>Amino acid substitution</i>	RBD residue					
	S383	S371	S373	Y369	F377	K378
<i>A</i>	20	57	122	1	1	0
<i>C</i>	0	6	0	39	4	0
<i>D</i>	0	0	0	7	0	0
<i>E</i>	0	0	0	0	0	18
<i>F</i>	33	5848267	0	47	NA	1
<i>G</i>	0	0	1	1	0	0
<i>H</i>	0	2	0	176	0	0
<i>I</i>	1	89	6	0	4	1
<i>K</i>	3	0	1	1	0	NA
<i>L</i>	1	2125763	1127	0	193	1
<i>M</i>	0	0	0	0	0	93
<i>N</i>	0	8	0	4	0	407
<i>P</i>	0	68	8047290	1	2	2
<i>Q</i>	0	0	2	5	0	11
<i>R</i>	0	0	1	0	0	153
<i>S</i>	NA	NA	NA	18	11	0
<i>T</i>	6	188	38	0	5	15
<i>V</i>	0	5	1	0	6	0
<i>W</i>	0	0	1	0	0	1
<i>Y</i>	10	1791	9	NA	14	0

Bibliography

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). B Cells and Antibodies. In *Molecular Biology of the Cell. 4th edition*. Garland Science. <https://www.ncbi.nlm.nih.gov/books/NBK26884/>
- Barnes, C. O., Jette, C. A., Abernathy, M. E., Dam, K.-M. A., Esswein, S. R., Gristick, H. B., Malyutin, A. G., Sharaf, N. G., Huey-Tubman, K. E., Lee, Y. E., Robbiani, D. F., Nussenzweig, M. C., West, A. P., & Bjorkman, P. J. (2020). SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature*, 588(7839), 682–687. <https://doi.org/10.1038/s41586-020-2852-1>
- Barnes, C. O., West, A. P., Huey-Tubman, K. E., Hoffmann, M. A. G., Sharaf, N. G., Hoffman, P. R., Koranda, N., Gristick, H. B., Gaebler, C., Muecksch, F., Lorenzi, J. C. C., Finkin, S., Hägglöf, T., Hurley, A., Millard, K. G., Weisblum, Y., Schmidt, F., Hatzioannou, T., Bieniasz, P. D., ... Bjorkman, P. J. (2020). Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell*, 182(4), 828-842.e16. <https://doi.org/10.1016/j.cell.2020.06.025>
- Bosch, B. J., van der Zee, R., de Haan, C. A. M., & Rottier, P. J. M. (2003). The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *Journal of Virology*, 77(16), 8801–8811. <https://doi.org/10.1128/JVI.77.16.8801-8811.2003>
- Burbelo, P. D., Riedo, F. X., Morishima, C., Rawlings, S., Smith, D., Das, S., Strich, J. R., Chertow, D. S., Davey, R. T., & Cohen, J. I. (2020). Detection of Nucleocapsid Antibody to SARS-CoV-2 is More Sensitive than Antibody to Spike Protein in COVID-19 Patients. *medRxiv*, 2020.04.20.20071423. <https://doi.org/10.1101/2020.04.20.20071423>

- Burnett, D. L., Jackson, K. J. L., Langley, D. B., Aggarwal, A., Stella, A. O., Johansen, M. D., Balachandran, H., Lenthall, H., Rouet, R., Walker, G., Saunders, B. M., Singh, M., Li, H., Henry, J. Y., Jackson, J., Stewart, A. G., Witthauer, F., Spence, M. A., Hansbro, N. G., ... Goodnow, C. C. (2021). Immunizations with diverse sarbecovirus receptor-binding domains elicit SARS-CoV-2 neutralizing antibodies against a conserved site of vulnerability. *Immunity*, 54(12), 2908-2921.e6. <https://doi.org/10.1016/j.immuni.2021.10.019>
- Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S. M., Walsh, R. M., Rawson, S., Rits-Volloch, S., & Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. *Science*, 369(6511), 1586–1592. <https://doi.org/10.1126/science.abd4251>
- Cao, Y., Wang, J., Jian, F., Xiao, T., Song, W., Yisimayi, A., Huang, W., Li, Q., Wang, P., An, R., Wang, J., Wang, Y., Niu, X., Yang, S., Liang, H., Sun, H., Li, T., Yu, Y., Cui, Q., ... Xie, X. S. (2022). Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature*, 602(7898), Article 7898. <https://doi.org/10.1038/s41586-021-04385-3>
- Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., Du, S., Wang, J., Li, Q., Chen, X., Yu, Y., Wang, P., Zhang, Z., Liu, P., An, R., Hao, X., Wang, Y., Wang, J., Feng, R., ... Xie, X. S. (2022). BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature*, 1–3. <https://doi.org/10.1038/s41586-022-04980-y>
- Charles A Janeway, J., Travers, P., Walport, M., & Shlomchik, M. J. (2001a). B-cell activation by armed helper T cells. In *Immunobiology: The Immune System in Health and Disease*. 5th edition. Garland Science. <https://www.ncbi.nlm.nih.gov/books/NBK27142/>
- Charles A Janeway, J., Travers, P., Walport, M., & Shlomchik, M. J. (2001b). The generation of diversity in immunoglobulins. In *Immunobiology: The Immune System in Health*

and Disease. 5th edition. Garland Science.

<https://www.ncbi.nlm.nih.gov/books/NBK27140/>

Cohen, A. A., Doremalen, N. van, Greaney, A. J., Andersen, H., Sharma, A., Starr, T. N., Keeffe, J. R., Fan, C., Schulz, J. E., Gnanapragasam, P. N. P., Kakutani, L. M., West, A. P., Saturday, G., Lee, Y. E., Gao, H., Jette, C. A., Lewis, M. G., Tan, T. K., Townsend, A. R., ... Bjorkman, P. J. (2022). *Mosaic RBD nanoparticles protect against multiple sarbecovirus challenges in animal models* (p. 2022.03.25.485875). bioRxiv. <https://doi.org/10.1101/2022.03.25.485875>

Cohen, A. A., Gnanapragasam, P. N. P., Lee, Y. E., Hoffman, P. R., Ou, S., Kakutani, L. M., Keeffe, J. R., Wu, H.-J., Howarth, M., West, A. P., Barnes, C. O., Nussenzweig, M. C., & Bjorkman, P. J. (2021). Mosaic nanoparticles elicit cross-reactive immune responses to zoonotic coronaviruses in mice. *Science*, 371(6530), 735–741. <https://doi.org/10.1126/science.abf6840>

Dejnirattisai, W., Huo, J., Zhou, D., Zahradník, J., Supasa, P., Liu, C., Duyvesteyn, H. M. E., Ginn, H. M., Mentzer, A. J., Tuekprakhon, A., Nutalai, R., Wang, B., Dijokaite, A., Khan, S., Avinoam, O., Bahar, M., Skelly, D., Adele, S., Johnson, S. A., ... Screaton, G. R. (2022). SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell*, 185(3), 467-484.e15. <https://doi.org/10.1016/j.cell.2021.12.046>

Dejnirattisai, W., Zhou, D., Ginn, H. M., Duyvesteyn, H. M. E., Supasa, P., Case, J. B., Zhao, Y., Walter, T. S., Mentzer, A. J., Liu, C., Wang, B., Paesen, G. C., Slon-Campos, J., López-Camacho, C., Kafai, N. M., Bailey, A. L., Chen, R. E., Ying, B., Thompson, C., ... Screaton, G. R. (2021). The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell*, 184(8), 2183-2200.e22. <https://doi.org/10.1016/j.cell.2021.02.032>

- Dolan, K. A., Dutta, M., Kern, D. M., Kotecha, A., Voth, G. A., & Brohawn, S. G. (2022). *Structure of SARS-CoV-2 M protein in lipid nanodiscs* (p. 2022.06.12.495841). bioRxiv. <https://doi.org/10.1101/2022.06.12.495841>
- Duan, L., Zheng, Q., Zhang, H., Niu, Y., Lou, Y., & Wang, H. (2020). The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Frontiers in Immunology*, 0. <https://doi.org/10.3389/fimmu.2020.576622>
- Fan, C., Cohen, A. A., Park, M., Hung, A. F.-H., Keeffe, J. R., Gnanapragasam, P. N. P., Lee, Y. E., Kakutani, L. M., Wu, Z., Malecek, K. E., Williams, J. C., & Bjorkman, P. J. (2022). *Neutralizing monoclonal antibodies elicited by mosaic RBD nanoparticles bind conserved sarbecovirus epitopes* (p. 2022.06.28.497989). bioRxiv. <https://doi.org/10.1101/2022.06.28.497989>
- Frank, S. A. (2002). Specificity and Cross-Reactivity. In *Immunology and Evolution of Infectious Disease*. Princeton University Press. <https://www.ncbi.nlm.nih.gov/books/NBK2396/>
- Greaney, A. J., Starr, T. N., Barnes, C. O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S., Wang, Z., Poston, D., Muecksch, F., Hatzioannou, T., Bieniasz, P. D., Robbani, D. F., Nussenzweig, M. C., Bjorkman, P. J., & Bloom, J. D. (2021). Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-24435-8>
- Henderson, R., Edwards, R. J., Mansouri, K., Janowska, K., Stalls, V., Gobeil, S. M. C., Kopp, M., Li, D., Parks, R., Hsu, A. L., Borgnia, M. J., Haynes, B. F., & Acharya, P. (2020). Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nature Structural & Molecular Biology*, 27(10), Article 10. <https://doi.org/10.1038/s41594-020-0479-4>

- Hsieh, C.-L., Goldsmith, J. A., Schaub, J. M., DiVenere, A. M., Kuo, H.-C., Javanmardi, K., Le, K. C., Wrapp, D., Lee, A. G., Liu, Y., Chou, C.-W., Byrne, P. O., Hjorth, C. K., Johnson, N. V., Ludes-Meyers, J., Nguyen, A. W., Park, J., Wang, N., Amengor, D., ... McLellan, J. S. (2020). Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science*, 369(6510), 1501–1505. <https://doi.org/10.1126/science.abd0826>
- Huang, Y., Yang, C., Xu, X., Xu, W., & Liu, S. (2020). Structural and functional properties of SARS-CoV-2 spike protein: Potential antiviral drug development for COVID-19. *Acta Pharmacologica Sinica*, 41(9), Article 9. <https://doi.org/10.1038/s41401-020-0485-4>
- Iwata-Yoshikawa, N., Kakizaki, M., Shiwa-Sudo, N., Okura, T., Tahara, M., Fukushi, S., Maeda, K., Kawase, M., Asanuma, H., Tomita, Y., Takayama, I., Matsuyama, S., Shirato, K., Suzuki, T., Nagata, N., & Takeda, M. (2022). Essential role of TMPRSS2 in SARS-CoV-2 infection in murine airways. *Nature Communications*, 13(1), 6100. <https://doi.org/10.1038/s41467-022-33911-8>
- Jackson, C. B., Farzan, M., Chen, B., & Choe, H. (2022). Mechanisms of SARS-CoV-2 entry into cells. *Nature Reviews Molecular Cell Biology*, 23(1), Article 1. <https://doi.org/10.1038/s41580-021-00418-x>
- Jette, C. A., Cohen, A. A., Gnanapragasam, P. N. P., Muecksch, F., Lee, Y. E., Huey-Tubman, K. E., Schmidt, F., Hatzioannou, T., Bieniasz, P. D., Nussenzweig, M. C., West, A. P., Keeffe, J. R., Bjorkman, P. J., & Barnes, C. O. (2021a). Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *bioRxiv*, 2021.04.23.441195. <https://doi.org/10.1101/2021.04.23.441195>
- Jette, C. A., Cohen, A. A., Gnanapragasam, P. N. P., Muecksch, F., Lee, Y. E., Huey-Tubman, K. E., Schmidt, F., Hatzioannou, T., Bieniasz, P. D., Nussenzweig, M. C.,

- West, A. P., Keeffe, J. R., Bjorkman, P. J., & Barnes, C. O. (2021b). Broad cross-reactivity across sarbecoviruses exhibited by a subset of COVID-19 donor-derived neutralizing antibodies. *Cell Reports*, 36(13). <https://doi.org/10.1016/j.celrep.2021.109760>
- Kaplinsky, J., Li, A., Sun, A., Coffre, M., Koralov, S. B., & Arnaout, R. (2014). Antibody repertoire deep sequencing reveals antigen-independent selection in maturing B cells. *Proceedings of the National Academy of Sciences*, 111(25), E2622–E2629. <https://doi.org/10.1073/pnas.1403278111>
- Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., Nakane, T., Zivanov, J., Neufeldt, C. J., Cerikan, B., Lu, J. M., Peukes, J., Xiong, X., Kräusslich, H.-G., Scheres, S. H. W., Bartenschlager, R., & Briggs, J. A. G. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*, 588(7838), Article 7838. <https://doi.org/10.1038/s41586-020-2665-2>
- Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., Corbett, K. S., Graham, B. S., McLellan, J. S., & Ward, A. B. (2016). Pre-fusion structure of a human coronavirus spike protein. *Nature*, 531(7592), 118–121. <https://doi.org/10.1038/nature17200>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), Article 7807. <https://doi.org/10.1038/s41586-020-2180-5>
- Lavie, M., Dubuisson, J., & Belouzard, S. (2022). SARS-CoV-2 Spike Furin Cleavage Site and S2' Basic Residues Modulate the Entry Process in a Host Cell-Dependent Manner. *Journal of Virology*, 96(13), e00474-22. <https://doi.org/10.1128/jvi.00474-22>
- Liu, L., Iketani, S., Guo, Y., Reddem, E. R., Casner, R. G., Nair, M. S., Yu, J., Chan, J. F.-W., Wang, M., Cerutti, G., Li, Z., Morano, N. C., Castagna, C. D., Corredor, L., Chu,

- H., Yuan, S., Poon, V. K.-M., Chan, C. C.-S., Chen, Z., ... Ho, D. D. (2022). An antibody class with a common CDRH3 motif broadly neutralizes sarbecoviruses. *Science Translational Medicine*, 14(646), eabn6859. <https://doi.org/10.1126/scitranslmed.abn6859>
- Liu, S., Liu, P., Lu, Q., Shen, Y., Zhang, L., Liang, Z., Yu, Y., Huang, W., & Wang, Y. (2025). The Compensatory Effect of S375F on S371F Is Vital for Maintaining the Infectivity of SARS-CoV-2 Omicron Variants. *Journal of Medical Virology*, 97(3), e70242. <https://doi.org/10.1002/jmv.70242>
- Lu, S., Ye, Q., Singh, D., Cao, Y., Diedrich, J. K., Yates, J. R., Villa, E., Cleveland, D. W., & Corbett, K. D. (2021). The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nature Communications*, 12(1), 502. <https://doi.org/10.1038/s41467-020-20768-y>
- Lv, Z., Deng, Y.-Q., Ye, Q., Cao, L., Sun, C.-Y., Fan, C., Huang, W., Sun, S., Sun, Y., Zhu, L., Chen, Q., Wang, N., Nie, J., Cui, Z., Zhu, D., Shaw, N., Li, X.-F., Li, Q., Xie, L., ... Wang, X. (2020). Structural basis for neutralization of SARS-CoV-2 and SARS-CoV by a potent therapeutic antibody. *Science*, 369(6510), 1505–1509. <https://doi.org/10.1126/science.abc5881>
- Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*, 23(1), Article 1. <https://doi.org/10.1038/s41580-021-00432-z>
- Marcink, T. C., Kicmal, T., Armbruster, E., Zhang, Z., Zipursky, G., Golub, K. L., Idris, M., Khao, J., Drew-Bear, J., McGill, G., Gallagher, T., Porotto, M., des Georges, A., & Moscona, A. (2022). Intermediates in SARS-CoV-2 spike–mediated cell entry. *Science Advances*, 8(33), eabo3153. <https://doi.org/10.1126/sciadv.abo3153>

- Meng, B., Abdullahi, A., Ferreira, I. A. T. M., Goonawardane, N., Saito, A., Kimura, I., Yamasoba, D., Gerber, P. P., Fatihi, S., Rathore, S., Zepeda, S. K., Papa, G., Kemp, S. A., Ikeda, T., Toyoda, M., Tan, T. S., Kuramochi, J., Mitsunaga, S., Ueno, T., ... Gupta, R. K. (2022). Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature*, 603(7902), Article 7902. <https://doi.org/10.1038/s41586-022-04474-x>
- Meng, B., Datir, R., Choi, J., Baker, S., Dougan, G., Hess, C., Kingston, N., Lehner, P. J., Lyons, P. A., Matheson, N. J., Owehand, W. H., Saunders, C., Summers, C., Thaventhiran, J. E. D., Toshner, M., Weekes, M. P., Maxwell, P., Shaw, A., Bucke, A., ... Gupta, R. K. (2022). SARS-CoV-2 spike N-terminal domain modulates TMPRSS2-dependent viral entry and fusogenicity. *Cell Reports*, 40(7), 111220. <https://doi.org/10.1016/j.celrep.2022.111220>
- Pinto, D., Park, Y.-J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., Peter, A., Guarino, B., Spreafico, R., Cameroni, E., Case, J. B., Chen, R. E., Havenar-Daughton, C., Snell, G., Telenti, A., ... Corti, D. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, 583(7815), Article 7815. <https://doi.org/10.1038/s41586-020-2349-y>
- Premkumar, L., Segovia-Chumbez, B., Jádí, R., Martínez, D. R., Raut, R., Markmann, A. J., Cornaby, C., Bartelt, L., Weiss, S., Park, Y., Edwards, C. E., Weimer, E., Scherer, E. M., Roupahel, N., Edupuganti, S., Weiskopf, D., Tse, L. V., Hou, Y. J., Margolis, D., ... de Silva, A. M. (2020). The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Science Immunology*, 5(48), eabc8413. <https://doi.org/10.1126/sciimmunol.abc8413>
- Ruffolo, J. A., Sulam, J., & Gray, J. J. (2022). Antibody structure prediction using interpretable deep learning. *Patterns*, 3(2), 100406. <https://doi.org/10.1016/j.patter.2021.100406>

S - Spike glycoprotein—Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) | UniProtKB | UniProt. (n.d.). Retrieved August 16, 2022, from <https://www.uniprot.org/uniprotkb/P0DTC2/entry>

Savastano, A., Ibáñez de Opakua, A., Rankovic, M., & Zweckstetter, M. (2020). Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-19843-1>

Scherer, K. M., Mascheroni, L., Carnell, G. W., Wunderlich, L. C. S., Makarchuk, S., Brockhoff, M., Mela, I., Fernandez-Villegas, A., Barysevich, M., Stewart, H., Suau Sans, M., George, C. L., Lamb, J. R., Kaminski-Schierle, G. S., Heeney, J. L., & Kaminski, C. F. (2022). SARS-CoV-2 nucleocapsid protein adheres to replication organelles before viral assembly at the Golgi/ERGIC and lysosome-mediated egress. *Science Advances*, 8(1), eabl4895. <https://doi.org/10.1126/sciadv.abl4895>

Schoeman, D., & Fielding, B. C. (2019). Coronavirus envelope protein: Current knowledge. *Virology Journal*, 16(1), 69. <https://doi.org/10.1186/s12985-019-1182-0>

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, 581(7807), Article 7807. <https://doi.org/10.1038/s41586-020-2179-y>

Shapira, T., Monreal, I. A., Dion, S. P., Buchholz, D. W., Imbiakha, B., Olmstead, A. D., Jager, M., Désilets, A., Gao, G., Martins, M., Vandal, T., Thompson, C. A. H., Chin, A., Rees, W. D., Steiner, T., Nabi, I. R., Marsault, E., Sahler, J., Diel, D. G., ... Jean, F. (2022). A TMPRSS2 inhibitor acts as a pan-SARS-CoV-2 prophylactic and therapeutic. *Nature*, 1–13. <https://doi.org/10.1038/s41586-022-04661-w>

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13), 30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>

Sompayrac, L. (2019). *How the immune system works* (Sixth edition). Wiley-Blackwell.

Starr, T. N., Czudnochowski, N., Liu, Z., Zatta, F., Park, Y.-J., Addetia, A., Pinto, D., Beltramello, M., Hernandez, P., Greaney, A. J., Marzi, R., Glass, W. G., Zhang, I., Dingens, A. S., Bowen, J. E., Tortorici, M. A., Walls, A. C., Wojcechowskyj, J. A., De Marco, A., ... Snell, G. (2021). SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. *Nature*, 597(7874), Article 7874. <https://doi.org/10.1038/s41586-021-03807-6>

Sztain, T., Ahn, S.-H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., Seitz, E., McCool, R. S., Kearns, F. L., Acosta-Reyes, F., Maji, S., Mashayekhi, G., McCammon, J. A., Ourmazd, A., Frank, J., McLellan, J. S., Chong, L. T., & Amaro, R. E. (2021). A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nature Chemistry*, 13(10), Article 10. <https://doi.org/10.1038/s41557-021-00758-3>

Turoňová, B., Sikora, M., Schürmann, C., Hagen, W. J. H., Welsch, S., Blanc, F. E. C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., van Zandbergen, G., Landry, J., de Azevedo, N. T. D., Mosalaganti, S., Schwarz, A., Covino, R., Mühlebach, M. D., Hummer, G., Krijnse Locker, J., & Beck, M. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science*, 370(6513), 203–208. <https://doi.org/10.1126/science.abd5223>

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: Implications for SARS-CoV-2. *Nature Reviews Microbiology*, 19(3), 155–170. <https://doi.org/10.1038/s41579-020-00468-6>

Walls, A. C., Tortorici, M. A., Bosch, B.-J., Frenz, B., Rottier, P. J. M., DiMaio, F., Rey, F. A., & Veesler, D. (2016). Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature*, 531(7592), 114–117. <https://doi.org/10.1038/nature16988>

- Wardemann, H., Yurasov, S., Schaefer, A., Young, J. W., Meffre, E., & Nussenzweig, M. C. (2003). Predominant Autoantibody Production by Early Human B Cell Precursors. *Science*, *301*(5638), 1374–1377. <https://doi.org/10.1126/science.1086907>
- WHO Coronavirus (COVID-19) Dashboard*. (n.d.). Retrieved August 18, 2022, from <https://covid19.who.int>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260–1263. <https://doi.org/10.1126/science.abb2507>
- Wrobel, A. G., Benton, D. J., Hussain, S., Harvey, R., Martin, S. R., Roustan, C., Rosenthal, P. B., Skehel, J. J., & Gamblin, S. J. (2020). Antibody-mediated disruption of the SARS-CoV-2 spike glycoprotein. *Nature Communications*, *11*(1), 5337. <https://doi.org/10.1038/s41467-020-19146-5>
- Wu, T. T., Johnson, G., & Kabat, E. A. (1993). Length distribution of CDRH3 in antibodies. *Proteins: Structure, Function, and Bioinformatics*, *16*(1), 1–7. <https://doi.org/10.1002/prot.340160102>
- Xiang, Y., Huang, W., Liu, H., Sang, Z., Nambulli, S., Tubiana, J., Williams, K. L., Duprex, W. P., Schneidman-Duhovny, D., Wilson, I. A., Taylor, D. J., & Shi, Y. (2022). Superimmunity by pan-sarbecovirus nanobodies. *Cell Reports*, *39*(13). <https://doi.org/10.1016/j.celrep.2022.111004>
- Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., Cheng, L., Shi, D., Lu, X., Lei, J., Crispin, M., Shi, Y., Li, L., & Li, S. (2020). Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, *183*(3), 730-738.e13. <https://doi.org/10.1016/j.cell.2020.09.018>

- Yi, C., Sun, X., Ye, J., Ding, L., Liu, M., Yang, Z., Lu, X., Zhang, Y., Ma, L., Gu, W., Qu, A., Xu, J., Shi, Z., Ling, Z., & Sun, B. (2020). Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cellular & Molecular Immunology*, 17(6), Article 6. <https://doi.org/10.1038/s41423-020-0458-z>
- Yuan, Y., Cao, D., Zhang, Y., Ma, J., Qi, J., Wang, Q., Lu, G., Wu, Y., Yan, J., Shi, Y., Zhang, X., & Gao, G. F. (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications*, 8(1), Article 1. <https://doi.org/10.1038/ncomms15092>
- Zhao, M.-M., Zhu, Y., Zhang, L., Zhong, G., Tai, L., Liu, S., Yin, G., Lu, J., He, Q., Li, M.-J., Zhao, R.-X., Wang, H., Huang, W., Fan, C., Shuai, L., Wen, Z., Wang, C., He, X., Chen, Q., ... Yang, J.-K. (2022). Novel cleavage sites identified in SARS-CoV-2 spike protein reveal mechanism for cathepsin L-facilitated viral infection and treatment strategies. *Cell Discovery*, 8(1), 53. <https://doi.org/10.1038/s41421-022-00419-w>
- Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., ... Shi, Z.-L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), Article 7798. <https://doi.org/10.1038/s41586-020-2012-7>
- Zhou, Z., Qiu, Y., & Ge, X. (2021). The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Animal Diseases*, 1(1), 5. <https://doi.org/10.1186/s44149-021-00005-9>
- Zhu, G., Zhu, C., Zhu, Y., & Sun, F. (2020). Minireview of progress in the structural study of SARS-CoV-2 proteins. *Current Research in Microbial Sciences*, 1, 53–61. <https://doi.org/10.1016/j.crmicr.2020.06>

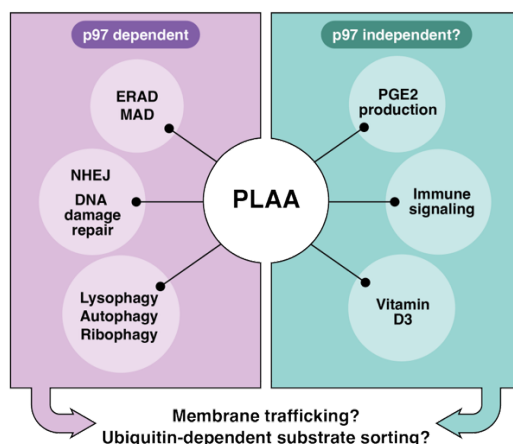
LITERATURE REVIEW: A STRUCTURAL PERSPECTIVE ON THE P97 ADAPTOR PHOSPHOLIPASE A2 ACTIVATING PROTEIN (PLAA)

To be published as:

Kate Radford, Paul W. Sternberg, Tsui-Fen Chou

Abstract

Phospholipase A2 activating protein (PLAA) is a highly conserved, essential eukaryotic protein whose cellular roles remain incompletely defined despite clear links to human disease. As the field increasingly recognizes that many proteins are pleiotropic—either “moonlighting” with multiple molecular functions or coupling a single biochemical activity to diverse processes—PLAA represents an instructive case study: it is clinically important, structurally distinctive, and sparsely annotated, yet implicated in an unusually broad set of pathways. Here, we synthesize decades of literature on PLAA and its orthologs (Doa1/Ufd3) from functional and structural perspectives, with the goal of unifying a fragmented and sometimes contradictory evidence base and identifying testable mechanistic models.



1. Introduction

Over the past several decades, our understanding of the relationship between protein sequence and function has moved away from the “one gene-one enzyme” hypothesis, as data uncover a growing number of proteins that are associated with multiple distinct cellular functions or processes. These proteins may be described as pleiotropic, either executing multiple distinct molecular functions with a single invariant gene product (type I pleiotropy, also referred to as “moonlighting”), or executing a single biochemical function which itself is associated with several distinct processes (type II pleiotropy) (Jeffery, 2019; Wagner & Zhang, 2011). Differences in the function of pleiotropic proteins depend on factors that change the physical nature of that protein’s environment: protein expression alterations across cell types and cell cycle points can modulate the availability of binding partners, changes in subcellular localization can transform enzymes to transcription factors, differences in concentration can enable oligomerization or crystallization, and many more (Jeffery, 2019). Because pleiotropic proteins are associated with system complexity, investigating their behavior, particularly of those that are highly conserved and/or essential, has the potential to enhance our basic understanding of cellular processes. However, functional characterization is biased toward a small percentage of human proteins, often neglecting those with poor annotation or ignoring the possibility of alternative functions for those already characterized (Kustatscher et al., 2022).

Phospholipase A2 Activating Protein, or PLAA, is one such protein. Though it is highly conserved across eukaryotes, possesses a structurally unique central domain, and has been found to be an essential protein with numerous clinical variants in mice/humans, its function remains poorly understood (Falik Zaccai et al., 2017; Hall et al., 2017). Research into PLAA and its orthologs has uncovered a number of functions with varying degrees of interconnectedness, suggesting that PLAA may be pleiotropic. PLAA’s yeast ortholog Ufd3/Doa1 has primarily been studied in its capacity as an adaptor to Cdc48/p97, an essential AAA ATPase with multiple roles including the ubiquitin-dependent degradation of proteinaceous substrates, but neither a specific mechanism of action nor substrate class was confirmed (Ghislain et al., 1996; Hochstrasser & Varshavsky, 1990). Mammalian PLAA was

discovered independently in its namesake capacity as an activator of certain phospholipases, and later found to also participate in other processes including as a p97 adapter and several involving ubiquitin recognition. Disentangling and evaluating these potential functions is complicated by contradictions in its relatively small body of literature, leaving many fundamental questions unanswered. PLAA is therefore an ideal subject to explore the challenges to interpretation posed by potentially pleiotropic proteins with sparse annotation.

The discovery of mammalian PLAA and the elucidation of its function began in the early 1990's. Initially thought to be a 28kDa protein, subsequent studies identified larger protein products and longer cDNA sequences until the 795-amino acid sequence now understood to be the full-length protein was identified (Beatty et al., 1999; Chopra et al., 1999; Clark et al., 1991; Ota et al., 2004; Ruiz et al., 1999). It is found across eukaryotes and is highly conserved, existing in those genomes as a single copy with no paralogs (Mullally et al., 2006).

PLAA was first identified in a screen for proteins thought to activate phospholipase A2 (PLA2) family enzymes (EC 3.1.1.4) (Clark et al., 1987). A short region of the PLAA C-terminal domain was found to have sequence similarity to the bee venom protein melittin, which stimulates PLA2 activity upon introduction to human tissues (Clark et al., 1987; Koumanov et al., 2003a; Mollay et al., 1976). In the following years, several studies identified inflammation-associated roles for PLAA, linking it to inflammation in synovial tissues and implicating it in prostaglandin E2 and cytokine production in human monocytes and neutrophils, apoptosis and mitochondrial damage (Bomalaski, Baker, et al., 1990; Bomalaski et al., 1989, 1995; Bomalaski, Fallon, et al., 1990; Peterson et al., 1996; D. A. Ribardo et al., 2001; G. Y. Sun et al., 2010; Y. Sun et al., 2012; F. Zhang, 2009; F. Zhang et al., 2008).

Concurrent but unrelated research in yeast identified the PLAA ortholog Doa1/Ufd3 as a critical component of ubiquitin-mediate protein processing. Doa1/Ufd3 was identified as an “*in vivo* ligand” of Cdc48 (p97/VCP), and was found to be essential for efficient protein

degradation and for maintaining cellular ubiquitin levels (Ghislain et al., 1996; Hochstrasser & Varshavsky, 1990; Qiu et al., 2010a). The human PLAA gene rescues Doa1-knock out phenotypes in yeast, suggesting a conserved function across PLAA genes, and also suggesting that PLAA and orthologs are cofactors of Cdc48/VCP/p97 (Mullally et al., 2006). Further research, described below, connected Doa1 and PLAA to even more ubiquitin-related functions, hinting at a complex and critical role.

This laundry list of roles drives interest in PLAA. When and where these functions are executed, and their manner of regulation, remain to be determined. Here, we synthesize the literature concerning PLAA and its orthologs from a functional and structural perspective with the dual aims of uniting a disparate body of literature and providing an example of how investigating an under-annotated protein has the potential to enhance our understanding of cellular systems beyond their immediate sphere.

2. **Functional roles of PLAA**

2.1. PLAA as a p97 adaptor protein and component of ubiquitin pathways

Cdc48/p97/VCP (hereafter p97) is an extremely abundant and essential AAA ATPase. As mentioned above, p97 is critical for ubiquitin-dependent protein degradation, extracting ubiquitinated polypeptides from macromolecular complexes and membranes and unfolds them prior to degradation in the proteasome (Gödderz et al., 2015; Jentsch & Rumpf, 2007; Olszewski et al., 2019; Peters et al., 1990; Stolz et al., 2011). Monomers of p97 comprise two consecutive AAA ATPase domains (D1 and D2), which translate ATP hydrolysis into the mechanical forces that contribute to protein unfolding. Both its N-terminal and C-terminal domains are implicated in substrate and cofactor recognition (Jentsch & Rumpf, 2007; Li et al., 2008). Monomers assemble into homohexameric rings, creating a pore through which substrate proteins are ratcheted during the unfolding process (X. Zhang et al., 2000). The ability of p97 to alter the structure of ubiquitinated proteins and, with the help of certain adaptor proteins, edit a substrate protein's ubiquitination state, provides critical regulation and flexibility to ubiquitin-mediated processes (Chapman et al., 2011; Meyer et al., 2012a). Thus, since its discovery in 1990, p97 has been implicated in an increasingly diverse set of cellular processes, including but not limited to participating in protein homeostasis and trafficking, membrane remodeling, and chromatin-associated nuclear processes, as well as governing aspects of the cellular stress response (Braxton & Southworth, 2023; Meyer et al., 2012b; Stolz et al., 2011; Yamanaka et al., 2012).

Proteins that interact with p97 (adaptor proteins) enable this startling diversity of functions. These adaptors not only target p97 to specific locations and substrates of interest, but also facilitate the direct interaction of p97 with its substrates: substrate recognition and unfolding require adaptor association (Braxton & Southworth, 2023; Buchberger et al., 2015). Most adaptor proteins belong to five structurally conserved classes that interact with the p97 N-terminal domain; however, PLAA contains one of only two protein domains known to interact with the p97 C-terminal tail, which localizes PLAA to the D2 face of p97 (Braxton & Southworth, 2023). This interaction was first characterized in yeast between the PLAA

and p97 orthologues Doa1 and Cdc48, respectively (Mullally et al., 2006). Later structural and mutational evidence showed significant contribution from the interaction of the hydroxyl group of p97 Tyr805 with PLAA Asp615, which is abolished upon Tyr805 phosphorylation (Qiu et al., 2010b; G. Zhao et al., 2007). Indeed, phosphorylation of p97 Tyr805 regulates binding of other adaptors that recognize the p97 C-terminus and is involved in a wide range of cellular activities (Li et al., 2008).

Adaptors to p97 generally possess additional domains that are not directly involved in p97 binding, but rather enable interaction with substrate proteins through ubiquitin binding, ubiquitin modification, or other posttranslational modification (Braxton & Southworth, 2023). PLAA is no exception: the WD-40 and PLAA family ubiquitin-binding (PFU) domains of the yeast orthologue Doa1 have been shown to bind ubiquitin, permitting the recruitment of ubiquitylated substrates and the formation of Doa1-Cdc48-substrate complexes (Mullally et al., 2006; Pashkova et al., 2010). Features of PLAA also appear to allow interactions with other p97 adaptors. For example, UBXD1 and YOD1, two other p97 adaptors, together with p97 and PLAA, and together assist p97 in lysophagy of K48-labeled lysosomes (Papadopoulos et al., 2017). PLAA and its orthologues may compete with other adaptors, such as Ufd2 (Rumpf & Jentsch, 2006).

The association of Doa1/Ufd3/PLAA with Cdc48/p97 plays a crucial role in certain ubiquitin pathways, including ribophagy, lysophagy, autophagy, DNA damage repair, and mitochondria-associated degradation (MAD) (Balakirev et al., 2015; Ossareh-Nazari et al., 2010; Papadopoulos et al., 2017; Wu et al., 2016). In each case, Doa1/Ufd3/PLAA is joined by other Cdc48/p97 adaptors, specifying the function of the complex.

Defects in the UFD3/DOA1 gene have been associated with unusually low abundance of free ubiquitin, and many phenotypes associated with knockouts of Doa1/Ufd3 can be rescued by overexpression of ubiquitin (Ghislain et al., 1996; Johnson et al., 1995; Lis & Romesberg, 2006; Ogiso et al., 2004; Rumpf & Jentsch, 2006; F. Zhang, 2009). Indeed, Doa1 activity in yeast is a major source of ubiquitin for the DNA damage response, funneling ubiquitin from the proteasomal degradation pathway instead to processes involved in altered DNA

replication and chromatin modification, interacting with genes involved in histone ubiquitination of histone H2B (Lis & Romesberg, 2006). Doa1/Ufd3/PLAA binds a diversity of ubiquitin chains, including K6-, K11-, K27-, K29-, K48-, and K63-linked chains, and have a particular affinity for polyubiquitin chains (Russell & Wilkinson, 2004; Tsuchiya et al., 2017; X. Zhang et al., 2017; X. Zhao et al., 2017). Its ability to recognize different ubiquitin linkage patterns may be determined by its binding partners (e.g., Npl4 and Otu1/YOD1), or may be attributed to an unknown mechanism (Tsuchiya et al., 2017). Structural, biochemical, and biological data support ubiquitin binding for both the PFU and WD-40 domains (further described in section 3).

Some ubiquitin-related activities of Doa1/Ufd3/PLAA have been described without mention of p97/VCP/Cdc48, and thus their dependence on p97 is not known. Doa1 was found to be partially responsible for non-homologous end joining DNA repair (NHEJ) (Wilson, 2002), and was found to relocalize from the nucleus to the nuclear periphery upon stress induction (Tkach et al., 2012). The Doa1 PFU domain interacts with Hse1, a component of the yeast ESCRT-0 complex with a ubiquitin-like SH3 domain that participates in processing ubiquitinated membrane proteins to be sorted into multivesicular bodies (Han et al., 2014; Ren et al., 2008). Furthermore, mutational studies of PLAA in mouse models demonstrate that PLAA is required in brain tissue for sorting membrane proteins from early to late endosome in a ubiquitin-dependent manner, thus targeting them for degradation in the lysosome and regulating both the trafficking of certain signaling receptors and the sorting of synaptic vesicle components during recycling (Hall et al., 2017). This finding corroborates clinical studies of patients with mutations in PLAA (discussed in section 4).

2.2. PLAA as a phospholipase A2 activator

Phospholipases are a class of enzyme that hydrolyze the ester bonds of membrane-bound phospholipids. They are divided into two general groups: acylhydrolases, which cleave the ester bonds within the carboxyl groups, and phosphodiesterases, which cleave phosphodiester bonds within the phosphate group (Aloulou et al., 2018). Phospholipase A2 enzymes (PLA2s) are responsible for hydrolyzing the ester bonds of phospholipids from the

sn-2 position (Bonventre, 1992). This cleavage site is enriched in arachidonic acid, the precursor of eicosanoids such as prostaglandin and leukotriene, which serve as signaling molecules in numerous physiological processes (Bonventre, 1992; Harizi et al., 2008). Critically, the presence of prostaglandin E2 (PGE2) can induce inflammation, pain sensation, and fever (Kawahara et al., 2015). Thus, PLA2 enzymes are not only tightly regulated and of great pharmacological interest, but also the target of natural toxins, such as the bee venom peptide melittin (Aloulou et al., 2018; Mollay et al., 1976). PLA2s show a high degree of structural diversity, and there is some dispute in the literature regarding whether PLAA activates secretory (low molecular weight) PLA2s (sPLA2s) or cytoplasmic (high molecular weight) PLA2s (cPLA2s) (Clark et al., 1987; Koumanov et al., 2003b; Pilane & LaBelle, 2004; Polizotto et al., 1999).

As stated above, mammalian PLAA was identified in a screen for PLA2-activating proteins (Clark et al., 1987). While an inhibitor of certain PLA2s had been identified in human cells, activators had only been identified in insect venoms, such as melittin (Mollay et al., 1976). Immunoprecipitation of mouse cell lysate with an anti-melittin antibody uniquely isolated a fragment of PLAA, which, when added back to whole cell lysate, led to an increase in PLA2 activity, indicated by the production of lysophospholipids (phospholipids with a single fatty acyl group) (Clark et al., 1986, 1987). Both this PLAA fragment and a peptide derived from it proved to be better PLA2 activators than melittin (Clark et al., 1991; Steiner et al., 1993).

Due to the role eicosanoids play in inflammatory processes, innate immune activation has also been observed upon the addition or expression of PLAA in mammalian tissues. A 28 kDa PLAA fragment was found in synovial fluid isolated from patients with rheumatoid arthritis, and PLAA expression could be induced in cells through treatment with monosodium urate, a salt that causes inflammation in synovial tissues (Bomalaski, Fallon, et al., 1990). Exogenous treatment with either the 28 kDa PLAA fragment or PLA2 both induced production of PGE2 in human monocytes and leukotriene B4 in neutrophils, and cytokines TNF- α and IL-1 β were found to induce PLAA and PLA2 expression and could themselves be induced by treatment with PLAA and PLA2 in human peripheral blood

monocytes (Bomalaski et al., 1995). Other cytokine levels also appear to be PLAA-dependent (e.g. IL-32), and expression of PLA2, COX2, and NF- κ B following induction with TNF- α is also increased in a PLAA-dependent manner (F. Zhang et al., 2008). This study also suggested a role for PLAA in apoptosis, citing higher levels of DNA fragmentation, caspase activity, PLA2 activation, and mitochondrial damage in cells expressing high levels of PLAA (F. Zhang, 2009). Moreover, PLAA was found to act as a pain-related mediator in a PLA2-dependent manner (Y. Sun et al., 2012). Other studies also showed a link between PLAA and PGE2 production. Cholera toxin, which causes rapid PGE2 production, also was associated with an increase in the abundance of PLAA mRNA and a 35 kDa band that stained with an antibody against a PLAA peptide (Peterson et al., 1996). Macrophages stimulated with PLAA similarly exhibit PGE2 production dependent on the activation not only of PLA2, but also of phospholipase C and D (D. A. Ribardo et al., 2001).

Eicosanoids can be critical signaling molecules for cell differentiation, and PLAA activation of PLA2 has thus been linked to the vitamin D3-dependent differentiation of growth plate chondrocytes in bone tissue (Doroudi et al., 2012, 2014; Doroudi, Olivares-Navarrete, et al., 2015; Doroudi, Plaisance, et al., 2015; Doroudi, Schwartz, et al., 2015; Schwartz et al., 2005). These studies also observed direct interactions of PLAA with caveolin, which is also a target of p97 in complex with its adaptor protein UBXD1, as well as with calmodulin, within this differentiation pathway.

PLA2 activity may not only release the base component of many signaling molecules, but also cause physical alterations to the lipid bilayer through the formation of lysophospholipids. Golgi tubulation and retrograde trafficking to the ER, essential cellular processes which mediate receptor trafficking and membrane protein homeostasis, require cPLA2 activity (de Figueiredo et al., 1998). However, whether tubules form as a result of lysophospholipid concentration or a downstream impact of eicosanoid-derived signaling molecules remains an open question (Brown et al., 2003; Choukroun et al., 2000; de Figueiredo et al., 1998; Polizotto et al., 1999). Both cPLA2 and iPLA2 enzymes have also

been found to play a more general role in membrane protein trafficking, and are a critical element to regulating membrane composition (Brown et al., 2003; Lee et al., 2010; Schmidt et al., 2010).

It is worth noting that, though the original fragment identified was primarily the mouse PLAA WD-40 domain, it possessed a long tail neither derived from any other PLAA domain within this reading frame nor documented in any known isoforms of PLAA, and the peptide identified to show melittin homology resided in this region (Clark et al., 1991). A later study identified an alternate melittin homology region in the PUL, PLAA's C-terminal domain, which the authors reported had a 38% sequence homology to melittin, but this claim cannot be replicated with modern sequence alignment methods (D. Ribardo et al., 2002). Limited research has been conducted that directly links PLAA to membrane changes derived from PLA2 activation, and PLAA is not mentioned in reviews describing mechanisms of PLA2 activation. Melittin exists only as a short peptide, and its amphipathic structure allows it to partition into and onto membranes (Polizotto et al., 1999; Wimley, 2018). Thus, though it was discovered through its ability to activate PLA2s, more research is needed to identify both how PLAA activates PLA2 enzymes and which PLA2 enzymes it can activate.

2.3. Mitochondrial association of PLAA

AAA-ATPases, including p97, are essential for extracting and degrading outer mitochondrial membrane (OMM) proteins (Karbowski & Youle, 2011; S. Xu et al., 2011; Ye et al., 2001). These functions are part of the ubiquitin-proteasome system, and involve the interaction of p97 with various adaptor proteins. Doa1, the yeast ortholog of PLAA, forms a functional complex with Cdc48, Ufd1 and Npl4 that mediates mitochondria-associated degradation (MAD) of ubiquitinated proteins in the OMM (Wu et al., 2016). Later experiments in mammalian cell culture (SH-SY5Y cells) observed that PLAA plays a similar role: it was observed to be partially localized to mitochondria, and a deficiency in PLAA impaired mitochondrial function and induced mitophagy (Zheng et al., 2023). This finding is particularly salient as reduced mitochondrial function is implicated in neurodegenerative and neurodevelopmental diseases, and mutations to PLAA have also been associated with

neurodevelopmental pathologies (reviewed in Anitha et al., 2023; Zuccoli & Carregari, 2022).

The ubiquitin-proteasome system in yeast is also important for regulating the import of tRNAs to the mitochondria, and Doa1 was found to be a negative effector of this process, targeting mitochondrial tRNA import factors for ubiquitin-mediated degradation (Brandina et al., 2007). However, similar evidence has not been described for mammalian PLAA.

2.4. PLAA shows variation in subcellular localization across studies

Despite its high conservation and status as an essential gene, PLAA remains sparsely annotated in large protein databases. Its subcellular localization is disputed across the literature, with some sources showing nuclear localization, cytoplasmic localization, extracellular exosome localization, or even localization to membrane-bound compartments like the ER or at the plasma membrane itself. The Human Protein Atlas reports that PLAA protein is highly expressed in the brain, testes, and bone marrow, and is annotated for its roles in development and neurogenesis, among others (Thul et al., 2017). Subcellular localization of endogenous PLAA in mouse brain (Hall, 2017) and in HEK293T cells shows PLAA present in both cytosol and nucleus. Comprehensive subcellular proteomics studies show inconsistent mapping of PLAA. A 2016 study that used SILAC labeling in HeLa cells coupled with mechanical cell fractionation identified PLAA in the “large protein complexes” fraction in the nucleus and cytosol (composed of complexes like ribosomes and proteasomes) with low confidence (Itzhak et al., 2016). Similarly, a 2019 mechanical fractionation and proteomics study of five cancer cell lines identified PLAA in the cytosolic and nuclear fractions for all lines (Orre et al., 2019). However, a 2025 study that fractionated by organelle immunocapture from HEK293T cells observed enrichment of PLAA when immunoprecipitated with ER, plasma membrane, and proteasome bait (Hein et al., 2025), aligning better with biological evidence supporting a role for PLAA in membrane protein turnover (Hall et al., 2017). These inconsistencies may be largely driven by differences in protein expression across cell types, and demonstrate the challenges associated with interpreting the behavior of PLAA.

3. Structural features of PLAA

PLAA comprises three globular domains: an N-terminal WD-40 repeat domain with seven segments (WD-40), a central PLAA family ubiquitin-binding domain (PFU) unique to PLAA and its orthologs, and a C-terminal “PLAP, Ufd3p, and Lub1p” (PUL) domain forming an alpha solenoid. The PFU and PUL domains are connected by an intrinsically disordered region, which varies in length across species. Though the WD-40 and PUL domains belong to common tertiary structural motifs, the contribution of these domains to PLAA’s function remains poorly understood. Therefore, the mechanism of action through which pathological mutations to PLAA cause disease is largely not known. Here, we describe the domains of PLAA, their common functions, and catalogue the effects of mutations on domain stability and function.

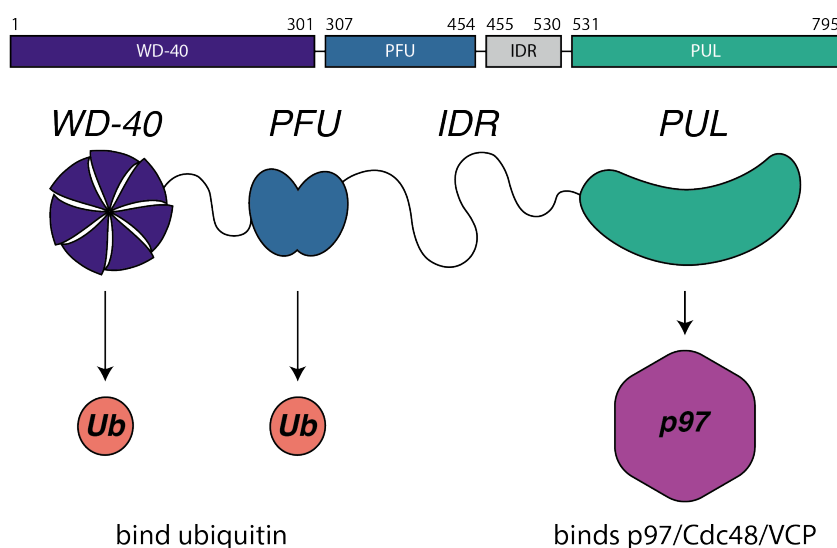


Figure 1. Domain map of PLAA (top) with illustration of PLAA’s domains (middle) and their primary known interacting partner (bottom).

3.1. *WD-40 domain*

WD-40 domains rank as one of the top ten most common structural motifs across eukaryotic proteins (Lander et al., 2001). WD repeat proteins are formed of ~40-60 residue

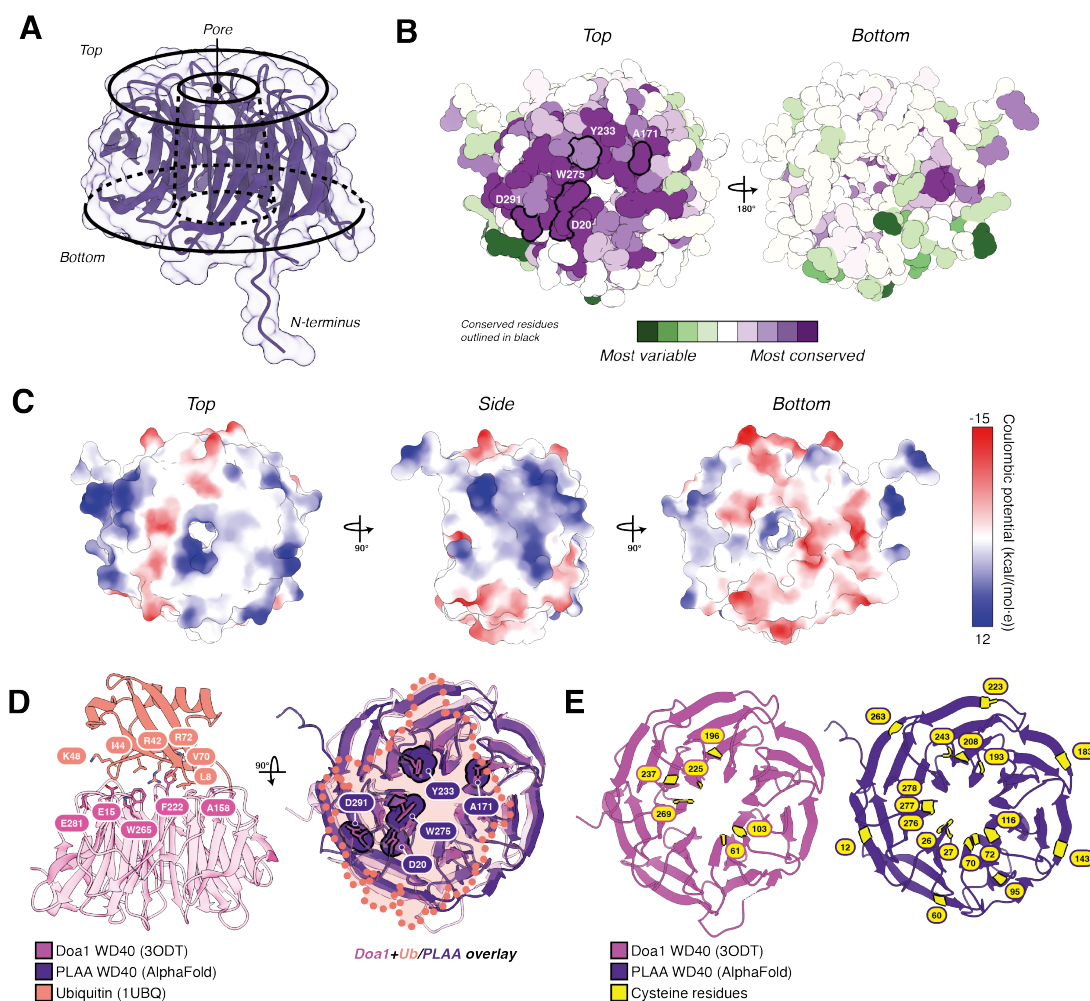


Figure 2. PLAA WD-40 domain. *A.* Human PLAA WD-40 domain topology, highlighting the top, bottom, and pore feature (structure generated with AlphaFold3). *B.* Structural conservation of PLAA WD-40 residues using divergent color map, with darker greens indicating increasing variability and darker purples indicating increased conservation. *C.* Coulombic potential of PLAA WD-40 domain surface in kcal/(mol·e), using divergent color map where darker red represents more acidic regions, and darker blue represents more basic regions. *D.* Ubiquitin docked on the Doa1 WD-40 structure per data in Pashkova et al (2010) to represent the interface observed in their work (PDB ID: 3ODT). *E.* Distribution of cysteine residues in the human PLAA WD-40 domain (AlphaFold).

repeated units terminating in a tryptophan-aspartate (WD) dipeptide. These repeats form structures of seven to eight twisted four-stranded antiparallel beta sheets arranged in a ring, forming a propeller-like structure (Figure 2A) (Smith et al., 1999). The N-terminal strand often forms part of the C-terminal sheet, helping to stabilize the domain.

WD-40 domains are associated with many functions, frequently serving as a platform for protein-protein, protein-peptide, or protein-nucleic acid interactions (Stirnemann et al., 2010). The two main interacting faces are the “top”, or narrow face and “bottom” or wider face, but the sides may also participate in interactions. The top face is more highly conserved (Figure 2B, left) and, in the case of PLAA-family WD-40 domains, based on NMR results, is responsible for interactions with ubiquitin (Figure 2D) (Pashkova et al., 2010). This face features a high proportion of arginine and large aromatic residues (tryptophan and tyrosine) around the pore, leading to a region of positive charge (Figure 2C). This charged region in the PLAA WD-40 domain may interact with negatively charged DNA, lipids, small molecules, or proteinaceous elements, as seen in other WD-40 domain containing proteins (Smith et al., 1999; Stirnemann et al., 2010; C. Xu & Min, 2011). The bottom face, however, is more evolutionarily variable (Figure 2B, right) and features amino acids with smaller side chains and more hydrophobic character and overall negative (acidic) charge (Figure 2C). The charge differential between the two faces may help orient binding partners. The sides feature a highly basic groove at the junction between the N and C termini of the domain, which is largely composed of conserved residues. This groove could act as a docking site for negatively charged (acidic) components of a binding partner. The central pore of many WD-repeat domains has been shown to interact with peptide regions of certain binding partners, and may in some cases bind small molecules with appropriate chemical features (Schapira et al., 2017). The PLAA WD-40 domain pore is positively charged (basic) in nature, and tapered, widening toward the bottom face. It might be responsible for mediating some of PLAA’s interactions.

In yeast the Doa1 WD-40 domain is responsible in part for PLAA’s interactions with ubiquitin (Pashkova et al., 2010). Though the central PFU domain has been characterized as

ubiquitin binding, the Doa1 WD-40 domain expressed as a fragment was shown to bind ubiquitin by both immunoprecipitation and NMR, and its affinity is around 220 μ M, compared to $\sim$ 1 mM reported for the PFU (Pashkova et al., 2010). Analysis of the Doa1 crystal structure in combination with NMR data revealed that ubiquitin binding occurs on the conserved top face of the Doa1 WD-40 domain, and critically the key residues of the interaction are conserved between Doa1 and human PLAA (Table 1) (Ashkenazy et al., 2016; Ben Chorin et al., 2020; Pashkova et al., 2010). Chemical shift perturbations in 15 N-Ub bound to the Doa1 WD-40 domain map this interaction to a hydrophobic patch on ubiquitin composed of L8, I44, and V70, along with basic residues K48, R42 and R72 (Figure 2D) (Pashkova et al., 2010). These residues correspond to a solvent-exposed hydrophobic patch commonly bound by ubiquitin binding domains (Dikic et al., 2009). D15S and/or F222A mutations in Doa1 WD-40 diminish ubiquitin binding, but do not appear to compromise Doa1 function; however, adding a third mutation, W265A, leads to abrogated Doa1 function (Pashkova et al., 2010).

Table 1. Ubiquitin-binding residues in Doa1 are conserved in human PLAA.

<i>Doa1</i>	<i>PLAA</i>	<i>Conservation score from Consurf (highest = 9)</i>
Asp15	Asp20	9
Phe222	Tyr233	8
Trp265	Trp275	9
Asp281	Asp291	9
Ala158	Ala171	9

Further evidence that the WD-40 domain mediates PLAA's interaction with certain ubiquitinated substrates comes from research into the autosomal recessive mutation of glycine 23 to valine, which when homozygous is embryonic or perinatal lethal in mice and is linked to severe neurodevelopmental disorder (PLAA-associated neurodevelopmental disorder, or PLAAND) in humans (Hall et al., 2017). In neurons of both humans and mice, the mutation G23V leads to the accumulation of K63-polyubiquitylated proteins as well as

synaptic membrane proteins, indicating a disruption of synaptic vesicle recycling and neurotransmission (Hall et al., 2017).

A striking feature of the human PLAA WD-40 domain is its abundance of cysteine residues (Figure 2E). Though only 300 residues in length, the human PLAA WD-40 domain contains 17 cysteine residues, accounting for 5.7% of the domain's amino acid content and double the average cysteine residue count for human proteins (2.26%) (Miseta & Csutora, 2000). Analysis based on NCBI blastP results of human WD-40 domains show an average of 3.5% cysteine residues, elevated above this average as well; however, even in its own structural family, the PLAA WD-40 domain is an outlier. The AlphaFold model of human PLAA shows cysteines in two predominant locations: clustered in the central pore and at the outside edge of each blade (Jumper et al., 2021). These cysteine residues do not appear to form disulfide bonds within a given PLAA monomer, nor do they seem to be involved in zinc chelation. We speculate that these cysteines may be the site of certain post-translational modifications, such as transient palmitoylation to assist with targeting PLAA to membranes. However, a role for these cysteines is not currently documented, and no similar cysteine-rich pore structures of other proteins can be readily identified.

3.2.PFU domain

The PFU, or “PLAA Family Ubiquitin-binding” domain, forms the second structured component of PLAA. This domain has been demonstrated to bind ubiquitin, though it shares no similarity to several known ubiquitin binding domains (Figure 3A) (Hurley et al., 2006). Mutations of certain conserved residues in the Doa1 PFU (F417D and F434D) abrogated binding to a monoubiquitin-Sepharose resin, suggesting their importance in ubiquitin binding (Mullally et al., 2006). An NMR structure (PDB ID: 2K8B) of the human PLAA PFU in complex with ubiquitin demonstrated not only a stable interaction between these proteins, but also characterized the novel fold architecture of the PFU. While the full-length PFU stretches from residues ~312-465, NMR-based structural efforts could only identify a stable structure from residues 386-465 (Fu et al., 2009b). AlphaFold predictions do identify alpha-helical secondary structure within this region, though the pLDDT (predicted Local Distance

A

Ub

PFU

2K8B

F450

D453

B

PFU

Ub

I44 patch

I36 patch

TEK box

F45

S65

N410

T66

K6

T9

M459

D445

K449

L73

W427

F450

D453

F441

I44

H68

Q441

V70

L8

Q49

N440

M442

R42

K63

K48

K29

K33

K27

K6

K11

C

F450

D453

H. sapiens 450-FIIDNT-455

S. cerevisiae 434-FILKNT-439

bound to substrates tagged for degradation and other fates (Fu et al., 2009b). NMR results suggest that the PFU recognizes the solvent-exposed hydrophobic I44 patch also bound by the WD-40 domain via a conserved helical motif (hPLAA: 450-FIIDNT-455, Doa1: 434-FILKNT-439) (Figure 3B, C) (Fu et al., 2009a; Pashkova et al., 2010). The I44 patch on the surface of ubiquitin (residues I44, L8, V70, and H68) is the most commonly recognized patch, bound by the proteasome and by most ubiquitin binding domains (Komander & Rape, 2012). Despite seemingly redundant ubiquitin recognition, deletion of this helical motif

combined with WD-40 domain mutations D15S/F222A leads to a severe loss of function in Doa1, suggesting that ubiquitin recognition by both domains is required for Doa1/PLAA to function properly (Pashkova et al., 2010).

Ubiquitin interactor affinity enrichment-mass spectrometry results captured PLAA interactions with a broad set of ubiquitin linkages. PLAA from HEK293T lysates was noted to be highly enriched on K27-linked chains (Zhao et al., 2017). K27 is the least solvent-accessible lysine on ubiquitin, and K27 linkages represent less than 1% of ubiquitin conjugates in human cells (~9% in yeast) (Castañeda et al., 2016; Dammer et al., 2011). These ubiquitin conjugates appear to be found predominantly on nuclear substrates and have been connected to DNA damage repair and innate immunity (Dammer et al., 2011; Gatti et al., 2015; Wang et al., 2014, 2017), and appear to play an epistatic role in p97-mediated nuclear substrate processing and subsequent cell proliferation (Shearer et al., 2022). Moreover, K27 linkages on the mitochondrial outer membrane protein VDAC1 are connected to initiating mitophagy, another function to which PLAA is connected through an unknown mechanism, though this may instead be attributed to K27-linked ubiquitin recognition by the autophagy adapter p62/SQSTM1 (Geisler et al., 2010; Zhou & Zhang, 2022).

PLAA was also observed to bind K11-, K29-, and K6-linked diubiquitin affinity resins in lysates from both mouse embryonic stem cell and mouse neuronal precursor cells, as well as K48- and K63-linked diubiquitin affinity resins from mouse neuronal precursor cell lysate alone (Zhao et al., 2017). While this includes ubiquitin linkages canonically associated with p97 interaction and proteasomal degradation (K11/K48), it also includes linkages connected to a wide array of processes, including the DNA damage response (K6, K27, K63), innate immune processes (K27, K63), Wnt/ β -catenin signaling (K29), and endocytosis (K63) (reviewed in Akutsu et al., 2016; Tracz & Bialek, 2021). PLAA was not identified upon diubiquitin enrichment in HeLa cell lysates, however, suggesting that PLAA shows cell-type specific ubiquitin binding (Zhang et al., 2017). Given that the WD-40 and PFU domains recognize the same hydrophobic binding surface on ubiquitin and thus at first glance appear

interchangeable, the wide array of functions associated with PLAA and the diversity of ubiquitin linkages suggest that there may be unknown components or mechanisms that permit ubiquitin-linked substrate differentiation, including the potential that PLAA recognizes branched ubiquitin chains, similar to some other p97 interacting proteins (Lange et al., 2024). However, experimental exploration is necessary before any conclusions may be drawn.

Two crystal structures have been solved of the PFU of Doa1 in complex with its own PUL (PDB IDs 3L3F and 3PSP) (Liu & Sun, 2011). Analysis with the PRODIGY-CRYSTAL webserver suggests that these contacts are likely crystal contacts rather than biological contacts (probability of 67.2% and 82.8%, respectively), and therefore further experiments are necessary before this interface can be considered biologically significant (Elez et al., 2018; Jiménez-García et al., 2019).

3.3. Intrinsically disordered region

PLAA's PFU and PUL are linked by an intrinsically disordered region (IDR). Large (>50 amino acid) unstructured regions are common in functional proteins, and may serve crucial functions such as transcription regulation, cell signaling, and as central hubs in protein interaction networks (Dyson, 2016; Dyson & Wright, 2005; Wright & Dyson, 2015). These regions show a biased amino acid distribution and low sequence complexity, favoring small,

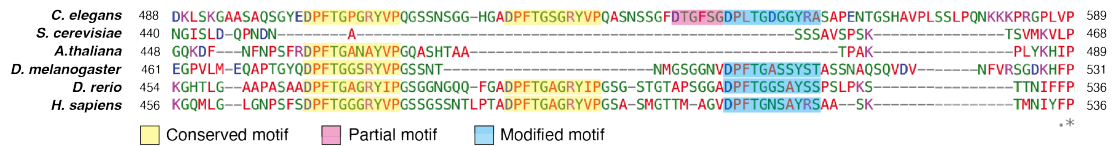


Figure 4. Sequence alignment (CLUSTAL Omega) of PLAA IDRs in five model organisms and humans. Conserved sequence features are highlighted, notably the conserved and repeating DPFTGxGRYVP motif (yellow), a partial version of that motif (pink), and a modified repeat of that motif (blue).

charged amino acids over hydrophobic and bulky residues (Uversky et al., 2000; Wright & Dyson, 2015). Upon association with a binding partner, IDRs may adopt an ordered tertiary

structure and permit a specific function; thus, IDRs may also be classed as “conditionally disordered/ordered” (Hausrath & Kingston, 2017; Struhl, 2024).

The human PLAA IDR is composed of >50% glycine, serine and threonine, 9% proline, and 9% aromatic residues. It has an isoelectric point of 9.4 (calculated with ExPASy ProtParam (Gasteiger et al., 2005)), indicating an overall net positive charge at physiological pH. AlphaFold predicts the presence of an IDR in this location in PLAA orthologues from eukaryotic model organisms in vascular plants (*A. thaliana*), fungi (*S. cerevisiae*), invertebrates (*D. melanogaster* and *C. elegans*) and vertebrates (*D. rerio* and *H. sapiens*). These vary in length from 24 amino acids (*S. cerevisiae*) to 86 (*C. elegans*). The longer IDRs in multicellular organisms contain repeats (including partial or with some sequence differences) of the conserved motif DPFTGxGRYVP (Figure 4) (Keil et al., 1996). The high degree of conservation concomitant with its presence in multicellular, but not single-cellular, eukaryotic organisms raises the possibility that this sequence may be responsible for interactions with a specific binding partner or class of binding partners relevant for intercellular activities; however, no function for this conserved region has been experimentally described. Recently, Das et al. (2025) showed that the IDR of *C. elegans* Ufd-3 is responsible for recruiting the mRNA decapping complex regulatory subunit DCAP-1 to mRNA processing bodies (P-bodies) in a ubiquitin-independent manner, demonstrating that the IDR supports Ufd-3 functions that are not mediated by ubiquitin (Das et al., 2025).

3.4.PUL domain

The PUL (named for three PLAA orthologs, PLAP, Ufd3p, and Lub1p) domain is an alpha solenoid domain formed from five Armadillo repeats (Figure 5A). This repetitive, alpha-helical structural motif is characteristic of many protein-binding proteins, including beta-catenin and F- and U-box proteins (Parmeggiani et al., 2008; Peifer et al., 1994). Domains composed of Armadillo repeats are held together through the interaction of conserved leucine residues expressed every 3-4 amino acids, yielding alpha helices with a leucine-rich face. These residues interact to form a leucine zipper between adjoining alpha helices that stabilizes the domain (Figure 5B; Hatzfeld, 1998).

The PUL is responsible for p97 recognition (Figure 5C) (Stach & Freemont, 2017). A crystal structure of the PUL of human PLAA (PDB ID: 3EBB) and a 10-amino acid peptide from p97 suggests the interaction is driven in part by hydrogen bonding between PLAA D615 on the interior face of the PUL and p97 Y805, which is highly conserved and subject to phosphorylation (Egerton & Samelson, 1994; Li et al., 2008; Qiu et al., 2010a; G. Zhao et al., 2007). This interaction with p97 is thought to be of critical importance. Because p97 does

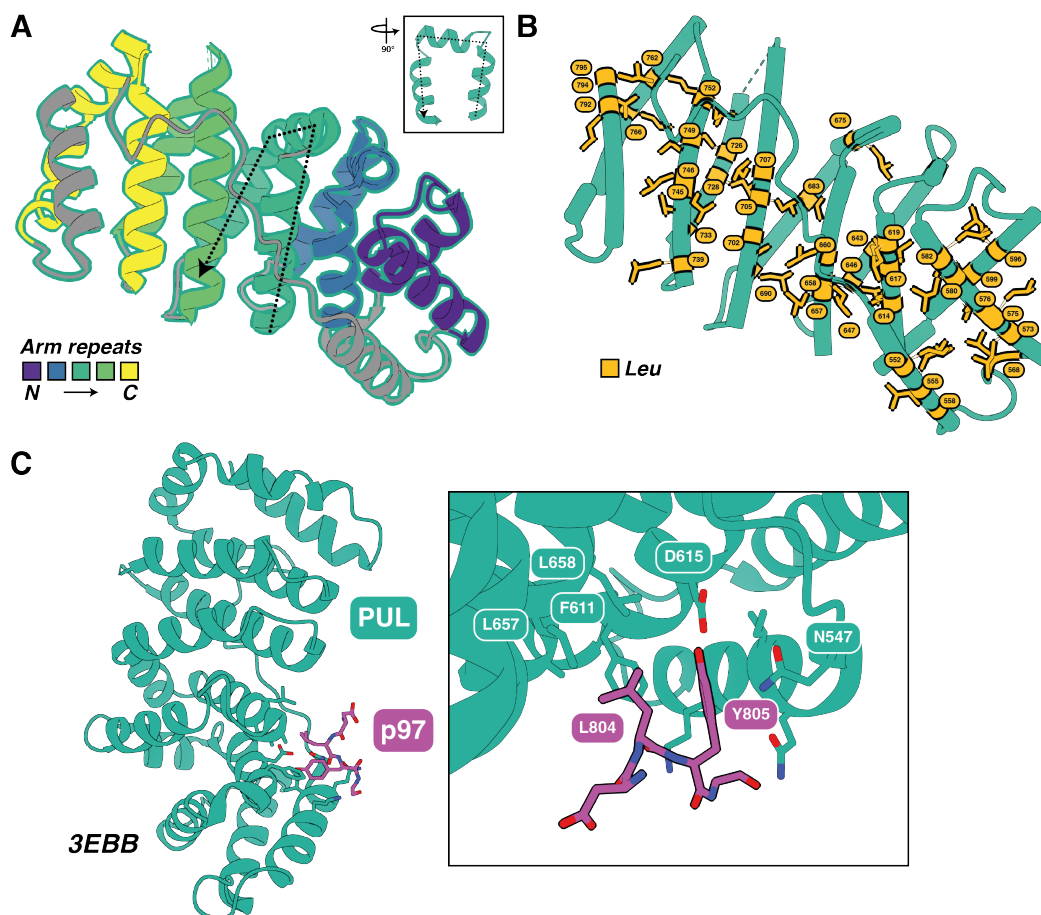


Figure 5. PLAA PFU features. A. Armadillo repeats in the PUL. Each repeat is colored individually, and the middle repeat is shown rotated 90 degrees in the upper corner to show the basic structure of the motif. B. Arrangement of leucine residues (orange) in the human PLAA PUL (PDB ID: 3EBB). C. Experimental structure of PUL (teal) in complex with p97 C-terminal peptide (magenta).

not bind substrates on its own, adaptor proteins to p97 fill this gap, possessing both p97- and substrate (ubiquitin)-recognition domains (reviewed in Braxton & Southworth, 2023). It is hypothesized that PLAA associates with ubiquitylated substrates via its WD-40 and/or PFU

domains and then, depending on the ubiquitination pattern, binds to p97 via the PUL, bringing proteins marked for degradation to the ubiquitin-proteasome system or the lysosome. While PLAA has been shown to interact with a variety of ubiquitin linkages and some of these interactions may relate to p97-independent functions, no specific data has been published to investigate this.

The PUL has also been thought to contain a region homologous with melittin, and thus to be the region responsible for PLAA's observed PLA2 activating properties (D. Ribardo et al., 2002). As discussed above, we could not replicate the percent homology claimed, using the SSEARCH2SEQ algorithm (Madeira et al., 2024). Also, the mechanism by which melittin is thought to work depends on its alpha helical peptide form (Guha et al., 2021; Jepson et al., 2022; Wimley, 2018). While evidence that PLAA can act as a PLA2 activator appears to be robust, many of the experiments were performed before PLAA was fully characterized and therefore may warrant revisiting. Further characterization is therefore required to establish which PLAA domain, if any, activates PLA2s.

4. PLAA: Clinical implications

PLAA is important for both ubiquitin-mediated trafficking and eicosanoid release, two processes integral to cell health and, if perturbed, can have far-reaching effects on the organism. Indeed, PLAA has been shown to be an essential gene in mouse, and lack of its expression causes a perinatal- or embryo-lethal phenotype in mice (Falik Zaccai et al., 2017; Hall et al., 2017). Though many of the pathological mutations to PLAA that have been identified are autosomal recessive, some heterozygous individuals may also show a disease phenotype (Landrum et al., 2014). Pathogenic mutations to PLAA have mainly been associated with neurodevelopmental disorder with progressive microencephaly, spasticity, and brain anomalies (NDMSBA), a rare disorder characterized by infantile-onset neurological symptoms including microencephaly, spasticity, and severe developmental impairment of both motor and intellectual function (see Table 2) (Falik Zaccai et al., 2017; Hall et al., 2017). PLAA's gene locus, 9p21, has also been flagged in genome-wide association studies of schizophrenia, though this result was not replicated in subsequent studies (Athanasiau et al., 2010; Sullivan et al., 2024).

To date, only three papers have been published that both describe disorders associated with PLAA mutations and use model systems and biochemical analyses to unpack the mutation's mechanistic impact. Homozygous mutations c.68G>T (G23V) and c.68dupG (L24fs) were found to cause a severe infantile neurodysfunctional disorder and premature lethality (Hall et al., 2017). Evidence from cerebellar tissue excised from $PLAA^{-/-}$, $PLAA^{G23V/G23V}$, and $PLAA^{G23V/-}$ mice as well as mouse embryonic fibroblasts (MEFs) of the same genotype grown *in vitro* demonstrated an essential role for PLAA in endolysosomal trafficking, specifically early steps of the Endosomal Sorting Complexes Required for Transport (ESCRT) pathway. Both G23V and L24fs mutations result in the dramatic reduction of PLAA expression and may truncate the protein through reduced expression of exon 1, accounting for the first 57 residues of the WD-40 domain. Less functional PLAA was shown to lead to impaired trafficking of K63-ubiquitylated substrates and their accumulation. Not

only does this impact neurotransmission, it also is critical for mammalian development (Hall et al., 2017).

Another homozygous mutation, c.2254C>T (L752F), was found in patients diagnosed with progressive leukoencephalopathy (Falik Zaccai et al., 2017). Biochemical assays using patient-derived fibroblasts showed that at rest, patient fibroblasts had half the level of PGE2 as healthy fibroblasts, but upon stimulation with LPS, PGE2 levels in healthy fibroblasts skyrocket to 5000-fold that found in patient fibroblasts, suggesting that the mutation causes a severe deficit in PGE2 production. Structurally, domains composed of Armadillo repeats, like the PUL, are held together through the interaction of conserved leucine residues. The L752F mutation is thought to reduce the stability of this leucine zipper and deform the PUL (Falik Zaccai et al., 2017; Hatzfeld, 1998).

The monoallelic mutations I609T and L795M have been linked to autism spectrum disorder (ASD) (Iacomino et al., 2024). L795 is highly conserved (GERP++ (Davydov et al., 2010), Consurf (Ashkenazy et al., 2016; Ben Chorin et al., 2020)), while I609 is ranked as highly conserved via GERP++ but as slightly variable by Consurf. These mutations appear to diminish the number of contacts made to the surrounding residues, possibly weakening the fold architecture of the PUL. Both mutations were identified in two individuals presenting with ASD, and expression in HEK293T cells showed these mutations to be associated with reduced binding of PLAA to p97. This is likely responsible for the changes in abundance of proteins associated with vesicle formation and transport, protein transport, and ERAD; however, changes in the abundance of mitochondrial respiratory chain proteins may reveal yet undiscovered roles for PLAA (Iacomino et al., 2024).

Other mutations to PLAA have been identified as causative agents for neurodevelopmental disorders, but have yet to be studied biochemically. D755G alone as well as the dual mutation of C277R and E350V together have been linked to NDMSBA (Dai et al., 2019; Dehghani et al., 2021). One might imagine that D755E, found in the PUL, may affect either interaction of PLAA with p97 or may affect PLAA's ability to activate PLA2s; however, the cooperative

action of C277R (WD-40 domain) and E350V (PFU) suggests the potential for as-yet undescribed mechanisms through which PLAA may execute its function.

Table 2. Pathogenic mutations in PLAA

MUTATION LOCATION WITHIN NM_001031689.3(PLAA)	PUBLICATION	AMINO ACID MUTATION	TYPE OF MUTATION	ZYGOCITY	PATHOLOGY	RATING
C.2383C>A (P.LEU795MET)	Iacomino 2024	L795M	SNV (missense)	Heterozygous	ASD	Pathogenic
C.2350DEL (P.LYS783_VAL784IN STER)	None	K783_V784*	Deletion (nonsense)	Recessive	NDMSBA	Likely pathogenic
C.2264A>G (P.ASP755GLY)	Dehghani 2021	D755G	SNV (missense)	Homozygous	NDMSBA	Pathogenic
C.2254C>T (P.LEU752PHE)	Falik Zaccai 2017	L752F	SNV (missense)	Recessive	NDMSBA	Pathogenic
C.1826T>C (P.ILE609THR)	Iacomino 2024	I609T	SNV (missense)	Heterozygous	ASD	Pathogenic
C.1800G>A (P.TRP600TER)	None	W577*	SNV (nonsense)	Recessive	NDMSBA	Likely pathogenic
C.1570C>T (P.ARG524TER)	Falik Zaccai 2017, Hall 2017, Dai 2019	R501*	SNV (nonsense)	Recessive	NDMSBA	Likely pathogenic
C.1049A>T (P.GLU350VAL)	Dai 2019	E350V	SNV (missense)	Heterozygous	NDMSBA	Pathogenic
C.861T>A (P.VAL287=)	None	V287V	SNV (synonymous)	Not listed	not provided	Pathogenic
C.850_860DEL (P.ASP284FS)	None	D284fs	Deletion (frameshift)	Heterozygous	NDMSBA	Likely pathogenic
C.829T>C (P.CYS277ARG)	Dai 2019	C277R	SNV (missense)	Heterozygous	NDMSBA	Pathogenic
C.802C>T (P.ARG268TER)	None	R268*	SNV (nonsense)	Not listed	Inborn genetic diseases	Pathogenic
C.240_241INSTAG (P.PRO81TER)	None	P81*	Insertion (nonsense)	Recessive	NDMSBA	Pathogenic
C.120C>G (P.ASP40GLU)	None	D40E	SNV (missense)	Recessive	NDMSBA	Pathogenic
C.68DUP (P.LEU24FS)	Hall 2017	L24fs	Duplication (frameshift)	Homozygous	NDMSBA	Pathogenic
C.68G>T (P.GLY23VAL)	Hall 2017	G23V	SNV (missense)	Homozygous	NDMSBA	Pathogenic

Data taken from the ClinVar database of the National Library of Medicine, accessed 2/26/2025, and synthesized from recent publications. Mutations classified as conflicting, benign, likely benign, or of uncertain significance have not been included in this table. Abbreviations: SNV = Single nucleotide variant, NDMSBA = Neurodevelopmental disorder with progressive microcephaly, spasticity, and brain anomalies, ASD = autism spectrum disorders.

5. Conclusions and future directions

Evidence in the literature supports PLAA as a pleiotropic protein. While it does appear that its primary roles may be in the ubiquitin pathway, where it is indispensable for the regulation of certain proteins and acts as an adapter to Cdc48/p97/VCP, it does also seem to possess some PLA2 activating functions, though none of the data yet describe a mechanism for its action or can agree on which PLA2 it may interact with. As with many proteins, the literature remains contradictory, and many experiments seem to be founded on faulty information, such as the claim that PLAA possesses homology to melittin. It is also worth noting that there is no mention of PLAA as an effector in the PLA2 literature.

While these roles may be entirely separate, it is also possible that there is synergy between them. Like p97, cPLA2 stimulation can be involved in processes like retrograde ER trafficking and other membrane protein trafficking pathways, and PLAA was observed to be associated with the ESCRT pathway in a mouse model (Hall et al., 2017). This opens the door for the possibility that PLAA is linked with both the protein and lipid components of membrane remodeling: as PLAA assists p97 in recycling ubiquitinated membrane proteins, it could also activate PLA2s to alter membrane curvature at the lipid level. Alternatively, it is possible that, as an essential player in development, PLAA's role in PLA2 activation could be related to signaling molecules generated downstream of arachidonic acid production – which may only be possible to observe in certain cell lines or tissue types.

There is also room to explore the mechanism of ubiquitin recognition by PLAA. Both the WD-40 domain and the PFU have been shown to recognize ubiquitin, and PLAA has been shown to bind a diversity of diubiquitin chains, but the mechanisms of ubiquitin linkage differentiation, critical for determining the fates of ubiquitinated proteins, have yet to be elucidated. Because the PFU is unique to PLAA orthologs and is a consistent characteristic of these proteins, it seems likely that multiple methods of ubiquitin recognition may be a feature of PLAA, perhaps allowing the protein to differentiate substrate fates based on

whether the WD-40 domain or the PFU is bound by ubiquitin. It could also be possible that PLAA recognizes branched ubiquitin chains.

Fundamentally, though, the question remains: what makes PLAA an essential gene in some eukaryotes? Though p97 itself is essential, not all p97 adaptors are, so it is not simply this relationship that renders PLAA deletion and certain mutations embryo lethal. PLAA's potential role in membrane trafficking may be the key: membrane remodeling plays a large role not only in the health of individual cells, but also in the health and survival of cellular communities, determining the efficiency and rapidity with which cells respond to each other and to changes in their environment. These processes must also be tightly controlled, and the loss of regulators can be very costly. When and where PLAA is expressed in an organism may therefore dramatically influence its expression level, subcellular localization, and ultimately its importance. Understanding the mechanisms of PLAA may therefore require investigation in cell types engaging in abundant intercellular trafficking, including neuronal cell lines and model organisms like *C. elegans* (Das et al., 2025).

With acknowledgements and thanks to Alakananda Das, Yanping Qiu, Chai Foong (Vivian) Lai, Shyue-Fang Battaglia-Hsu, and Jakub Ettrich for discussions.

Bibliography

- Akutsu, M., Dikic, I., & Bremm, A. (2016). Ubiquitin chain diversity at a glance. *Journal of Cell Science*, 129(5), 875–880. <https://doi.org/10.1242/jcs.183954>
- Aloulou, A., Rahier, R., Arhab, Y., Noiriél, A., & Abousalham, A. (2018). Phospholipases: An Overview. In G. Sandoval (Ed.), *Lipases and Phospholipases: Methods and Protocols* (pp. 69–105). Springer. https://doi.org/10.1007/978-1-4939-8672-9_3
- Anitha, A., Thanseem, I., Iype, M., & Thomas, S. V. (2023). Mitochondrial dysfunction in cognitive neurodevelopmental disorders: Cause or effect? *Mitochondrion*, 69, 18–32. <https://doi.org/10.1016/j.mito.2023.01.002>
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, 44(W1), W344–W350. <https://doi.org/10.1093/nar/gkw408>
- Athanasios, L., Mattingsdal, M., Kähler, A. K., Brown, A., Gustafsson, O., Agartz, I., Giegling, I., Muglia, P., Cichon, S., Rietschel, M., Pietiläinen, O. P. H., Peltonen, L., Bramon, E., Collier, D., Clair, D. St., Sigurdsson, E., Petursson, H., Rujescu, D., Melle, I., ... Andreassen, O. A. (2010). Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *Journal of Psychiatric Research*, 44(12), 748–753. <https://doi.org/10.1016/j.jpsychires.2010.02.002>
- Balakirev, M. Y., Mullally, J. E., Favier, A., Assard, N., Sulpice, E., Lindsey, D. F., Rulina, A. V., Gidrol, X., & Wilkinson, K. D. (2015). Wss1 metalloprotease partners with Cdc48/Doa1 in processing genotoxic SUMO conjugates. *eLife*, 4, e06763. <https://doi.org/10.7554/eLife.06763>

- Beatty, B. G., Qi, S., Pienkowska, M., Herbrick, J.-A., Scheidl, T., Zhang, Z. M., Kola, I., Scherer, S. W., & Seth, A. (1999). Chromosomal Localization of Phospholipase A2 Activating Protein, an Ets2 Target Gene, to 9p21. *Genomics*, 62(3), 529–532. <https://doi.org/10.1006/geno.1999.5999>
- Ben Chorin, A., Masrati, G., Kessel, A., Narunsky, A., Sprinzak, J., Lahav, S., Ashkenazy, H., & Ben-Tal, N. (2020). ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Science*, 29(1), 258–267. <https://doi.org/10.1002/pro.3779>
- Bomalaski, J. S., Baker, D. G., Brophy, L. M., & Clark, M. A. (1990). Monosodium urate crystals stimulate phospholipase A2 enzyme activities and the synthesis of a phospholipase A2-activating protein. *The Journal of Immunology*, 145(10), 3391–3397. <https://doi.org/10.4049/jimmunol.145.10.3391>
- Bomalaski, J. S., Baker, D. G., Brophy, L., Resurreccion, N. V., Spilberg, I., Muniain, M., & Clark, M. A. (1989). A phospholipase A2-activating protein (PLAP) stimulates human neutrophil aggregation and release of lysosomal enzymes, superoxide, and eicosanoids. *The Journal of Immunology*, 142(11), 3957–3962. <https://doi.org/10.4049/jimmunol.142.11.3957>
- Bomalaski, J. S., Fallon, M., Turner, R. A., Crooke, S. T., Meunier, P. C., & Clark, M. A. (1990). Identification and isolation of a phospholipase A2 activating protein in human rheumatoid arthritis synovial fluid: Induction of eicosanoid synthesis and an inflammatory response in joints injected in vivo. *The Journal of Laboratory and Clinical Medicine*, 116(6), 814–825.
- Bomalaski, J. S., Ford, T., Hudson, A. P., & Clark, M. A. (1995). Phospholipase A2-activating protein induces the synthesis of IL-1 and TNF in human monocytes. *The Journal of Immunology*, 154(8), 4027–4031. <https://doi.org/10.4049/jimmunol.154.8.4027>

- Bonventre, J. V. (1992). Phospholipase A2 and signal transduction. *Journal of the American Society of Nephrology*, 3(2), 128. <https://doi.org/10.1681/ASN.V32128>
- Brandina, I., Smirnov, A., Kolesnikova, O., Entelis, N., Krashennnikov, I. A., Martin, R. P., & Tarassov, I. (2007). tRNA import into yeast mitochondria is regulated by the ubiquitin-proteasome system. *FEBS Letters*, 581(22), 4248–4254. <https://doi.org/10.1016/j.febslet.2007.08.013>
- Braxton, J. R., & Southworth, D. R. (2023). Structural insights of the p97/VCP AAA+ ATPase: How adapter interactions coordinate diverse cellular functionality. *Journal of Biological Chemistry*, 299(11). <https://doi.org/10.1016/j.jbc.2023.105182>
- Brown, W. J., Chambers, K., & Doody, A. (2003). Phospholipase A2 (PLA2) Enzymes in Membrane Trafficking: Mediators of Membrane Shape and Function. *Traffic*, 4(4), 214–221. <https://doi.org/10.1034/j.1600-0854.2003.00078.x>
- Buchberger, A., Schindelin, H., & Hänzelmann, P. (2015). Control of p97 function by cofactor binding. *FEBS Letters*, 589(19, Part A), 2578–2589. <https://doi.org/10.1016/j.febslet.2015.08.028>
- Castañeda, C. A., Dixon, E. K., Walker, O., Chaturvedi, A., Nakasone, M. A., Curtis, J. E., Reed, M. R., Krueger, S., Cropp, T. A., & Fushman, D. (2016). Linkage via K27 Bestows Ubiquitin Chains with Unique Properties among Polyubiquitins. *Structure*, 24(3), 423–436. <https://doi.org/10.1016/j.str.2016.01.007>
- Chapman, E., N. Fry, A., & Kang, M. (2011). The complexities of p97 function in health and disease. *Molecular BioSystems*, 7(3), 700–710. <https://doi.org/10.1039/C0MB00176G>
- Chopra, A. K., Ribardo, D. A., Wood, T. G., Prusak, D. J., Xu, X.-J., & Peterson, J. W. (1999). Molecular characterization of cDNA for phospholipase A2-activating protein. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1444(1), 125–130. [https://doi.org/10.1016/S0167-4781\(98\)00249-8](https://doi.org/10.1016/S0167-4781(98)00249-8)

- Choukroun, G. J., Marshansky, V., Gustafson, C. E., McKee, M., Hajjar, R. J., Rosenzweig, A., Brown, D., & Bonventre, J. V. (2000). Cytosolic phospholipase A₂ regulates Golgi structure and modulates intracellular trafficking of membrane proteins. *The Journal of Clinical Investigation*, 106(8), 983–993. <https://doi.org/10.1172/JCI8914>
- Clark, M. A., Conway, T. M., Shorr, R. G., & Crooke, S. T. (1987). Identification and isolation of a mammalian protein which is antigenically and functionally related to the phospholipase A₂ stimulatory peptide melittin. *Journal of Biological Chemistry*, 262(9), 4402–4406. [https://doi.org/10.1016/S0021-9258\(18\)61362-6](https://doi.org/10.1016/S0021-9258(18)61362-6)
- Clark, M. A., Littlejohn, D., Conway, T. M., Mong, S., Steiner, S., & Crooke, S. T. (1986). Leukotriene D₄ treatment of bovine aortic endothelial cells and murine smooth muscle cells in culture results in an increase in phospholipase A₂ activity. *Journal of Biological Chemistry*, 261(23), 10713–10718. [https://doi.org/10.1016/S0021-9258\(18\)67444-7](https://doi.org/10.1016/S0021-9258(18)67444-7)
- Clark, M. A., Ozgür, L. E., Conway, T. M., Dispoto, J., Crooke, S. T., & Bomalaski, J. S. (1991). Cloning of a phospholipase A₂-activating protein. *Proceedings of the National Academy of Sciences*, 88(12), 5418–5422. <https://doi.org/10.1073/pnas.88.12.5418>
- Dai, C., Zeng, S., Tan, Z., Yang, X., Du, J., Lu, G., & Wang, J. (2019). Neurodevelopmental disorder with progressive microcephaly, spasticity, and brain anomalies in China caused by novel mutations of PLAA. *Clinical Genetics*, 96(4), 380–381. <https://doi.org/10.1111/cge.13608>
- Dammer, E. B., Na, C. H., Xu, P., Seyfried, N. T., Duong, D. M., Cheng, D., Gearing, M., Rees, H., Lah, J. J., Levey, A. I., Rush, J., & Peng, J. (2011). Polyubiquitin Linkage Profiles in Three Models of Proteolytic Stress Suggest the Etiology of Alzheimer Disease *. *Journal of Biological Chemistry*, 286(12), 10457–10465. <https://doi.org/10.1074/jbc.M110.149633>

- Das, A., Qiu, Y., Wolf, T. J., Brissett, E., Cho, J., Park, H., Chen, E. C., Chou, T.-F., & Sternberg, P. W. (2025). PLAA/UFD-3 regulates P-bodies through its intrinsic disordered domain. *Proceedings of the National Academy of Sciences*, 122(26), e2427250122. <https://doi.org/10.1073/pnas.2427250122>
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLOS Computational Biology*, 6(12), e1001025. <https://doi.org/10.1371/journal.pcbi.1001025>
- de Figueiredo, P., Drecktrah, D., Katzenellenbogen, J. A., Strang, M., & Brown, W. J. (1998). Evidence that phospholipase A2 activity is required for Golgi complex and trans Golgi network membrane tubulation. *Proceedings of the National Academy of Sciences*, 95(15), 8642–8647. <https://doi.org/10.1073/pnas.95.15.8642>
- Dehghani, A., Razmara, E., Rasoulinezhad, M., Bitarafan, F., Tavasoli, A. R., & Garshasbi, M. (2021). *Novel Imaging and Clinical Manifestations of Ndmsba Disorder Caused by a Homozygous Missense Variant of Plaa* [Preprint]. In Review. <https://doi.org/10.21203/rs.3.rs-407675/v1>
- Dikic, I., Wakatsuki, S., & Walters, K. J. (2009). Ubiquitin-binding domains—From structures to functions. *Nature Reviews Molecular Cell Biology*, 10(10), 659–671. <https://doi.org/10.1038/nrm2767>
- Doroudi, M., Boyan, B. D., & Schwartz, Z. (2014). Rapid 1 α ,25(OH) $_2$ D $_3$ membrane-mediated activation of Ca $^{2+}$ /calmodulin-dependent protein kinase II in growth plate chondrocytes requires Pdia3, PLAA and caveolae. *Connective Tissue Research*, 55(sup1), 125–128. <https://doi.org/10.3109/03008207.2014.923882>
- Doroudi, M., Olivares-Navarrete, R., Boyan, B. D., & Schwartz, Z. (2015). A review of 1 α ,25(OH) $_2$ D $_3$ dependent Pdia3 receptor complex components in Wnt5a non-

- canonical pathway signaling. *The Journal of Steroid Biochemistry and Molecular Biology*, 152, 84–88. <https://doi.org/10.1016/j.jsbmb.2015.04.002>
- Doroudi, M., Plaisance, M. C., Boyan, B. D., & Schwartz, Z. (2015). Membrane actions of $1\alpha,25(\text{OH})_2\text{D}_3$ are mediated by Ca^{2+} /calmodulin-dependent protein kinase II in bone and cartilage cells. *The Journal of Steroid Biochemistry and Molecular Biology*, 145, 65–74. <https://doi.org/10.1016/j.jsbmb.2014.09.019>
- Doroudi, M., Schwartz, Z., & Boyan, B. D. (2012). Phospholipase A2 activating protein is required for $1\alpha,25$ -dihydroxyvitamin D3 dependent rapid activation of protein kinase C via Pdia3. *The Journal of Steroid Biochemistry and Molecular Biology*, 132(1), 48–56. <https://doi.org/10.1016/j.jsbmb.2012.03.001>
- Doroudi, M., Schwartz, Z., & Boyan, B. D. (2015). Membrane-mediated actions of $1,25$ -dihydroxy vitamin D3: A review of the roles of phospholipase A2 activating protein and Ca^{2+} /calmodulin-dependent protein kinase II. *The Journal of Steroid Biochemistry and Molecular Biology*, 147, 81–84. <https://doi.org/10.1016/j.jsbmb.2014.11.002>
- Dyson, H. J. (2016). Making Sense of Intrinsically Disordered Proteins. *Biophysical Journal*, 110(5), 1013–1016. <https://doi.org/10.1016/j.bpj.2016.01.030>
- Dyson, H. J., & Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3), 197–208. <https://doi.org/10.1038/nrm1589>
- Egerton, M., & Samelson, L. E. (1994). Biochemical characterization of valosin-containing protein, a protein tyrosine kinase substrate in hematopoietic cells. *Journal of Biological Chemistry*, 269(15), 11435–11441. [https://doi.org/10.1016/S0021-9258\(19\)78142-3](https://doi.org/10.1016/S0021-9258(19)78142-3)
- Elez, K., Bonvin, A. M. J. J., & Vangone, A. (2018). Distinguishing crystallographic from biological interfaces in protein complexes: Role of intermolecular contacts and

energetics for classification. *BMC Bioinformatics*, 19(15), 438.
<https://doi.org/10.1186/s12859-018-2414-9>

Falik Zaccai, T. C., Savitzki, D., Zivony-Elboun, Y., Vilboux, T., Fitts, E. C., Shoval, Y., Kalfon, L., Samra, N., Keren, Z., Gross, B., Chasnyk, N., Straussberg, R., Mullikin, J. C., Teer, J. K., Geiger, D., Kornitzer, D., Bitterman-Deutsch, O., Samson, A. O., Wakamiya, M., ... Chopra, A. K. (2017). Phospholipase A2-activating protein is associated with a novel form of leukoencephalopathy. *Brain*, 140(2), 370–386.
<https://doi.org/10.1093/brain/aww295>

Fu, Q.-S., Zhou, C.-J., Gao, H.-C., Jiang, Y.-J., Zhou, Z.-R., Hong, J., Yao, W.-M., Song, A.-X., Lin, D.-H., & Hu, H.-Y. (2009a). Structural Basis for Ubiquitin Recognition by a Novel Domain from Human Phospholipase A2-activating Protein*. *Journal of Biological Chemistry*, 284(28), 19043–19052.
<https://doi.org/10.1074/jbc.M109.009126>

Fu, Q.-S., Zhou, C.-J., Gao, H.-C., Jiang, Y.-J., Zhou, Z.-R., Hong, J., Yao, W.-M., Song, A.-X., Lin, D.-H., & Hu, H.-Y. (2009b). Structural Basis for Ubiquitin Recognition by a Novel Domain from Human Phospholipase A2-activating Protein *. *Journal of Biological Chemistry*, 284(28), 19043–19052.
<https://doi.org/10.1074/jbc.M109.009126>

Gasteiger, E., Hoogland, C., Gattiker, A., S'everine, D., Wilkins, M.R., Appel, R.D., Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, J.M. (eds) The Proteomics Protocols Handbook. Springer Protocols Handbooks. Humana Press. <https://doi.org/10.1385/1-59259-890-0:571>

Gatti, M., Pinato, S., Maiolica, A., Rocchio, F., Prato, M. G., Aebersold, R., & Penengo, L. (2015). RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage. *Cell Reports*, 10(2), 226–238. <https://doi.org/10.1016/j.celrep.2014.12.021>

- Geisler, S., Holmström, K. M., Skujat, D., Fiesel, F. C., Rothfuss, O. C., Kahle, P. J., & Springer, W. (2010). PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biology*, 12(2), 119–131. <https://doi.org/10.1038/ncb2012>
- Ghislain, M., Dohmen, R. J., Levy, F., & Varshavsky, A. (1996). Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. *The EMBO Journal*, 15(18), 4884–4899. <https://doi.org/10.1002/j.1460-2075.1996.tb00869.x>
- Gödderz, D., Heinen, C., Marchese, F. P., Kurz, T., Acs, K., & Dantuma, N. P. (2015). Cdc48-independent proteasomal degradation coincides with a reduced need for ubiquitylation. *Scientific Reports*, 5(1), 7615. <https://doi.org/10.1038/srep07615>
- Guha, S., Ferrie, R. P., Ghimire, J., Ventura, C., Wu, E., Sun, L., Kim, S. Y., Wiedman, G. R., Hristova, K., & Wimley, W. C. (2021). Applications and evolution of melittin, the quintessential membrane active peptide. *Biochemical Pharmacology*, 193, 114769. <https://doi.org/10.1016/j.bcp.2021.114769>
- Hall, E. A., Nahorski, M. S., Murray, L. M., Shaheen, R., Perkins, E., Dissanayake, K. N., Kristaryanto, Y., Jones, R. A., Vogt, J., Rivagorda, M., Handley, M. T., Mali, G. R., Quidwai, T., Soares, D. C., Keighren, M. A., McKie, L., Mort, R. L., Gammoh, N., Garcia-Munoz, A., ... Mill, P. (2017). PLAA Mutations Cause a Lethal Infantile Epileptic Encephalopathy by Disrupting Ubiquitin-Mediated Endolysosomal Degradation of Synaptic Proteins. *The American Journal of Human Genetics*, 100(5), 706–724. <https://doi.org/10.1016/j.ajhg.2017.03.008>
- Han, S., Shin, D., Choi, H., & Lee, S. (2014). Molecular determinants of the interaction between Doa1 and Hse1 involved in endosomal sorting. *Biochemical and Biophysical Research Communications*, 446(1), 352–357. <https://doi.org/10.1016/j.bbrc.2014.02.118>

- Harizi, H., Corcuff, J.-B., & Gualde, N. (2008). Arachidonic-acid-derived eicosanoids: Roles in biology and immunopathology. *Trends in Molecular Medicine*, 14(10), 461–469. <https://doi.org/10.1016/j.molmed.2008.08.005>
- Hatzfeld, M. (1998). The Armadillo Family of Structural Proteins. In K. W. Jeon (Ed.), *International Review of Cytology* (Vol. 186, pp. 179–224). Academic Press. [https://doi.org/10.1016/S0074-7696\(08\)61054-2](https://doi.org/10.1016/S0074-7696(08)61054-2)
- Hausrath, A. C., & Kingston, R. L. (2017). Conditionally disordered proteins: Bringing the environment back into the fold. *Cellular and Molecular Life Sciences: CMLS*, 74(17), 3149–3162. <https://doi.org/10.1007/s00018-017-2558-1>
- Hein, M. Y., Peng, D., Todorova, V., McCarthy, F., Kim, K., Liu, C., Savy, L., Januel, C., Baltazar-Nunez, R., Sekhar, M., Vaid, S., Bax, S., Vangipuram, M., Burgess, J., Njoya, L., Wang, E., Ivanov, I. E., Byrum, J. R., Pradeep, S., ... Leonetti, M. D. (2025). Global organelle profiling reveals subcellular localization and remodeling at proteome scale. *Cell*, 188(4), 1137–1155.e20. <https://doi.org/10.1016/j.cell.2024.11.028>
- Hochstrasser, M., & Varshavsky, A. (1990). In vivo degradation of a transcriptional regulator: The yeast alpha 2 repressor. *Cell*, 61(4), 697–708. [https://doi.org/10.1016/0092-8674\(90\)90481-s](https://doi.org/10.1016/0092-8674(90)90481-s)
- Hurley, J. H., Lee, S., & Prag, G. (2006). Ubiquitin-binding domains. *Biochemical Journal*, 399(3), 361–372. <https://doi.org/10.1042/BJ20061138>
- Iacomino, M., Houerbi, N., Fortuna, S., Howe, J., Li, S., Scorrano, G., Riva, A., Cheng, K.-W., Steiman, M., Peltekova, I., Yusuf, A., Baldassari, S., Tamburro, S., Scudieri, P., Musante, I., Di Ludovico, A., Guerrisi, S., Balagura, G., Corsello, A., ... Salpietro, V. (2024). Allelic heterogeneity and abnormal vesicle recycling in PLAA-related neurodevelopmental disorders. *Frontiers in Molecular Neuroscience*, 17, 1268013. <https://doi.org/10.3389/fnmol.2024.1268013>

- Itzhak, D. N., Tyanova, S., Cox, J., & Borner, G. H. (2016). Global, quantitative and dynamic mapping of protein subcellular localization. *eLife*, 5, e16950. <https://doi.org/10.7554/eLife.16950>
- Jeffery, C. J. (2019). An enzyme in the test tube, and a transcription factor in the cell: Moonlighting proteins and cellular factors that affect their behavior. *Protein Science*, 28(7), 1233–1238. <https://doi.org/10.1002/pro.3645>
- Jentsch, S., & Rumpf, S. (2007). Cdc48 (p97): A ‘molecular gearbox’ in the ubiquitin pathway? *Trends in Biochemical Sciences*, 32(1), 6–11. <https://doi.org/10.1016/j.tibs.2006.11.005>
- Jepson, T. A., Hall, S. C., & Chung, J. K. (2022). Single-molecule phospholipase A2 becomes processive on melittin-induced membrane deformations. *Biophysical Journal*, 121(8), 1417–1423. <https://doi.org/10.1016/j.bpj.2022.03.019>
- Jiménez-García, B., Elez, K., Koukos, P. I., Bonvin, A. M., & Vangone, A. (2019). PRODIGY-crystal: A web-tool for classification of biological interfaces in protein complexes. *Bioinformatics*, 35(22), 4821–4823. <https://doi.org/10.1093/bioinformatics/btz437>
- Johnson, E. S., Ma, P. C. M., Ota, I. M., & Varshavsky, A. (1995). A Proteolytic Pathway That Recognizes Ubiquitin as a Degradation Signal *. *Journal of Biological Chemistry*, 270(29), 17442–17456. <https://doi.org/10.1074/jbc.270.29.17442>
- 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, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 1–11. <https://doi.org/10.1038/s41586-021-03819-2>

- Karbowski, M., & Youle, R. J. (2011). Regulating mitochondrial outer membrane proteins by ubiquitination and proteasomal degradation. *Current Opinion in Cell Biology*, 23(4), 476–482. <https://doi.org/10.1016/j.ceb.2011.05.007>
- Kawahara, K., Hohjoh, H., Inazumi, T., Tsuchiya, S., & Sugimoto, Y. (2015). Prostaglandin E2-induced inflammation: Relevance of prostaglandin E receptors. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1851(4), 414–421. <https://doi.org/10.1016/j.bbalip.2014.07.008>
- Keil, R. L., Wolfe, D., Reiner, T., Peterson, C. J., & Riley, J. L. (1996). Molecular genetic analysis of volatile-anesthetic action. *Molecular and Cellular Biology*, 16(7), 3446–3453.
- Komander, D., & Rape, M. (2012). The Ubiquitin Code. *Annual Review of Biochemistry*, 81(Volume 81, 2012), 203–229. <https://doi.org/10.1146/annurev-biochem-060310-170328>
- Koumanov, K., Momchilova, A., & Wolf, C. (2003a). Bimodal regulatory effect of melittin and phospholipase A2-activating protein on human type II secretory phospholipase A2. *Cell Biology International*, 27(10), 871–877. [https://doi.org/10.1016/S1065-6995\(03\)00176-8](https://doi.org/10.1016/S1065-6995(03)00176-8)
- Koumanov, K., Momchilova, A., & Wolf, C. (2003b). Bimodal regulatory effect of melittin and phospholipase A2-activating protein on human type II secretory phospholipase A2. *Cell Biology International*, 27(10), 871–877. [https://doi.org/10.1016/S1065-6995\(03\)00176-8](https://doi.org/10.1016/S1065-6995(03)00176-8)
- Kustatscher, G., Collins, T., Gingras, A.-C., Guo, T., Hermjakob, H., Ideker, T., Lilley, K. S., Lundberg, E., Marcotte, E. M., Ralser, M., & Rappsilber, J. (2022). Understudied proteins: Opportunities and challenges for functional proteomics. *Nature Methods*, 19(7), 774–779. <https://doi.org/10.1038/s41592-022-01454-x>

- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., ... The Wellcome Trust: (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), Article 6822. <https://doi.org/10.1038/35057062>
- Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, 42(D1), D980–D985. <https://doi.org/10.1093/nar/gkt1113>
- Lange, S. M., McFarland, M. R., Lamoliatte, F., Carroll, T., Krshnan, L., Pérez-Ràfols, A., Kwasna, D., Shen, L., Wallace, I., Cole, I., Armstrong, L. A., Knebel, A., Johnson, C., De Cesare, V., & Kulathu, Y. (2024). VCP/p97-associated proteins are binders and debranching enzymes of K48–K63-branched ubiquitin chains. *Nature Structural & Molecular Biology*, 31(12), 1872–1887. <https://doi.org/10.1038/s41594-024-01354-y>
- Lee, O. R., Kim, S. J., Kim, H. J., Hong, J. K., Ryu, S. B., Lee, S. H., Ganguly, A., & Cho, H.-T. (2010). Phospholipase A2 Is Required for PIN-FORMED Protein Trafficking to the Plasma Membrane in the Arabidopsis Root. *The Plant Cell*, 22(6), 1812–1825. <https://doi.org/10.1105/tpc.110.074211>
- Li, G., Zhao, G., Schindelin, H., & Lennarz, W. J. (2008). Tyrosine phosphorylation of ATPase p97 regulates its activity during ERAD. *Biochemical and Biophysical Research Communications*, 375(2), 247–251. <https://doi.org/10.1016/j.bbrc.2008.08.018>
- Lis, E. T., & Romesberg, F. E. (2006). Role of Doa1 in the *Saccharomyces cerevisiae* DNA Damage Response. *Molecular and Cellular Biology*, 26(11), 4122–4133. <https://doi.org/10.1128/MCB.01640-05>

- Liu, Y., & Sun, J. (2011). *RCSB PDB - 3PSP: Crystal structure of PUL and PFU domain*.
<https://www.rcsb.org/structure/3psp>
- Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., Lopez, R., & Butcher, S. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Research*, 52(W1), W521–W525.
<https://doi.org/10.1093/nar/gkae241>
- Meyer, H., Bug, M., & Bremer, S. (2012a). Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature Cell Biology*, 14(2), 117–123.
<https://doi.org/10.1038/ncb2407>
- Meyer, H., Bug, M., & Bremer, S. (2012b). Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nature Cell Biology*, 14(2), 117–123.
<https://doi.org/10.1038/ncb2407>
- Miseta, A., & Csutora, P. (2000). Relationship Between the Occurrence of Cysteine in Proteins and the Complexity of Organisms. *Molecular Biology and Evolution*, 17(8), 1232–1239. <https://doi.org/10.1093/oxfordjournals.molbev.a026406>
- Mollay, C., Kreil, G., & Berger, H. (1976). Action of phospholipases on the cytoplasmic membrane of Escherichia coli. Stimulation by melittin. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 426(2), 317–324. [https://doi.org/10.1016/0005-2736\(76\)90340-0](https://doi.org/10.1016/0005-2736(76)90340-0)
- Mullally, J. E., Chernova, T., & Wilkinson, K. D. (2006). Doa1 Is a Cdc48 Adapter That Possesses a Novel Ubiquitin Binding Domain. *Molecular and Cellular Biology*, 26(3), 822–830. <https://doi.org/10.1128/MCB.26.3.822-830.2006>
- Ogiso, Y., Sugiura, R., Kamo, T., Yanagiya, S., Lu, Y., Okazaki, K., Shuntoh, H., & Kuno, T. (2004). Lub1 Participates in Ubiquitin Homeostasis and Stress Response via

- Maintenance of Cellular Ubiquitin Contents in Fission Yeast. *Molecular and Cellular Biology*, 24(6), 2324–2331. <https://doi.org/10.1128/MCB.24.6.2324-2331.2004>
- Olszewski, M. M., Williams, C., Dong, K. C., & Martin, A. (2019). The Cdc48 unfoldase prepares well-folded protein substrates for degradation by the 26S proteasome. *Communications Biology*, 2(1), 1–8. <https://doi.org/10.1038/s42003-019-0283-z>
- Orre, L. M., Vesterlund, M., Pan, Y., Arslan, T., Zhu, Y., Woodbridge, A. F., Frings, O., Fredlund, E., & Lehtiö, J. (2019). SubCellBarCode: Proteome-wide Mapping of Protein Localization and Relocalization. *Molecular Cell*, 73(1), 166–182.e7. <https://doi.org/10.1016/j.molcel.2018.11.035>
- Ossareh-Nazari, B., Bonizec, M., Cohen, M., Dokudovskaya, S., Delalande, F., Schaeffer, C., Van Dorsselaer, A., & Dargemont, C. (2010). Cdc48 and Ufd3, new partners of the ubiquitin protease Ubp3, are required for ribophagy. *EMBO Reports*, 11(7), 548–554. <https://doi.org/10.1038/embor.2010.74>
- Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Ishii, S., Yamamoto, J., ... Sugano, S. (2004). Complete sequencing and characterization of 21,243 full-length human cDNAs. *Nature Genetics*, 36(1), 40–45. <https://doi.org/10.1038/ng1285>
- Papadopoulos, C., Kirchner, P., Bug, M., Grum, D., Koerver, L., Schulze, N., Poehler, R., Dressler, A., Fengler, S., Arhzaouy, K., Lux, V., Ehrmann, M., Wehl, C. C., & Meyer, H. (2017). VCP/p97 cooperates with YOD1, UBXD1 and PLAA to drive clearance of ruptured lysosomes by autophagy. *The EMBO Journal*, 36(2), 135–150. <https://doi.org/10.15252/embj.201695148>
- Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O., Caflisch, A., & Plückthun, A. (2008). Designed Armadillo Repeat Proteins as General

Peptide-Binding Scaffolds: Consensus Design and Computational Optimization of the Hydrophobic Core. *Journal of Molecular Biology*, 376(5), 1282–1304. <https://doi.org/10.1016/j.jmb.2007.12.014>

Pashkova, N., Gakhar, L., Winistorfer, S. C., Yu, L., Ramaswamy, S., & Piper, R. C. (2010). WD40 Repeat Propellers Define a Ubiquitin-Binding Domain that Regulates Turnover of F Box Proteins. *Molecular Cell*, 40(3), 433–443. <https://doi.org/10.1016/j.molcel.2010.10.018>

Peifer, M., Berg, S., & Reynolds, A. B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell*, 76(5), 789–791. [https://doi.org/10.1016/0092-8674\(94\)90353-0](https://doi.org/10.1016/0092-8674(94)90353-0)

Peters, J. M., Walsh, M. J., & Franke, W. W. (1990). An abundant and ubiquitous homooligomeric ring-shaped ATPase particle related to the putative vesicle fusion proteins Sec18p and NSF. *The EMBO Journal*, 9(6), 1757–1767. <https://doi.org/10.1002/j.1460-2075.1990.tb08300.x>

Peterson, J. W., Saini, S. S., Dickey, W. D., Klimpel, G. R., Bomalaski, J. S., Clark, M. A., Xu, X. J., & Chopra, A. K. (1996). Cholera toxin induces synthesis of phospholipase A2-activating protein. *Infection and Immunity*, 64(6), 2137–2143. <https://doi.org/10.1128/iai.64.6.2137-2143.1996>

Pilane, C. M., & LaBelle, E. F. (2004). cPLA2 activator peptide, PLAP, increases arachidonic acid release and apoptosis of vascular smooth muscle cells. *Journal of Cellular Physiology*, 198(1), 48–52. <https://doi.org/10.1002/jcp.10382>

Polizotto, R. S., de Figueiredo, P., & Brown, W. J. (1999). Stimulation of golgi membrane tubulation and retrograde trafficking to the ER by phospholipase A2 activating protein (PLAP) peptide. *Journal of Cellular Biochemistry*, 74(4), 670–683. [https://doi.org/10.1002/\(SICI\)1097-4644\(19990915\)74:4%253C670::AID-JCB16%253E3.0.CO;2-%2523](https://doi.org/10.1002/(SICI)1097-4644(19990915)74:4%253C670::AID-JCB16%253E3.0.CO;2-%2523)

- Qiu, L., Pashkova, N., Walker, J. R., Winistorfer, S., Allali-Hassani, A., Akutsu, M., Piper, R., & Dhe-Paganon, S. (2010a). Structure and Function of the PLAA/Ufd3-p97/Cdc48 Complex*. *Journal of Biological Chemistry*, 285(1), 365–372. <https://doi.org/10.1074/jbc.M109.044685>
- Qiu, L., Pashkova, N., Walker, J. R., Winistorfer, S., Allali-Hassani, A., Akutsu, M., Piper, R., & Dhe-Paganon, S. (2010b). Structure and Function of the PLAA/Ufd3-p97/Cdc48 Complex*. *Journal of Biological Chemistry*, 285(1), 365–372. <https://doi.org/10.1074/jbc.M109.044685>
- Ren, J., Pashkova, N., Winistorfer, S., & Piper, R. C. (2008). DOA1/UFD3 Plays a Role in Sorting Ubiquitinated Membrane Proteins into Multivesicular Bodies*. *Journal of Biological Chemistry*, 283(31), 21599–21611. <https://doi.org/10.1074/jbc.M802982200>
- Ribardo, D. A., Crowe, S. E., Kuhl, K. R., Peterson, J. W., & Chopra, A. K. (2001). Prostaglandin Levels in Stimulated Macrophages Are Controlled by Phospholipase A2-activating Protein and by Activation of Phospholipase C and D *. *Journal of Biological Chemistry*, 276(8), 5467–5475. <https://doi.org/10.1074/jbc.M006690200>
- Ribardo, D., Peterson, J., & Chopra, A. (2002). Phospholipase A2-activating protein—An important regulatory molecule in modulating cyclooxygenase-2 and tumor necrosis factor production during inflammation. *Indian Journal of Experimental Biology*. <https://www.semanticscholar.org/paper/Phospholipase-A2-activating-protein--an-important-Ribardo-Peterson/cd4b604733c118eacaa63bd29f2c940b46d51158>
- Ruiz, A., Nadal, M., Puig, S., & Estivill, X. (1999). Cloning of the human phospholipase A2 activating protein (hPLAP) gene on the chromosome 9p21 melanoma deleted region. *Gene*, 239(1), 155–161. [https://doi.org/10.1016/S0378-1119\(99\)00354-6](https://doi.org/10.1016/S0378-1119(99)00354-6)

- Rumpf, S., & Jentsch, S. (2006). Functional division of substrate processing cofactors of the ubiquitin-selective Cdc48 chaperone. *Molecular Cell*, 21(2), 261–269. <https://doi.org/10.1016/j.molcel.2005.12.014>
- Russell, N. S., & Wilkinson, K. D. (2004). Identification of a novel 29-linked polyubiquitin binding protein, Ufd3, using polyubiquitin chain analogues. *Biochemistry*, 43(16), 4844–4854. <https://doi.org/10.1021/bi035626r>
- Schapira, M., Tyers, M., Torrent, M., & Arrowsmith, C. H. (2017). WD40 repeat domain proteins: A novel target class? *Nature Reviews Drug Discovery*, 16(11), Article 11. <https://doi.org/10.1038/nrd.2017.179>
- Schmidt, J. A., Kalkofen, D. N., Donovan, K. W., & Brown, W. J. (2010). A Role for Phospholipase A2 Activity in Membrane Tubule Formation and TGN Trafficking. *Traffic*, 11(12), 1530–1536. <https://doi.org/10.1111/j.1600-0854.2010.01115.x>
- Schwartz, Z., Graham, E. j., Wang, L., Lossdörfer, S., Gay, I., Johnson-Pais, T. I., Carnes, D. I., Sylvia, V. I., & Boyan, B. d. (2005). Phospholipase A2 activating protein (PLAA) is required for $1\alpha,25(\text{OH})_2\text{D}_3$ signaling in growth plate chondrocytes. *Journal of Cellular Physiology*, 203(1), 54–70. <https://doi.org/10.1002/jcp.20212>
- Shearer, R. F., Typas, D., Coscia, F., Schovsbo, S., Kruse, T., Mund, A., & Mailand, N. (2022). K27-linked ubiquitylation promotes p97 substrate processing and is essential for cell proliferation. *The EMBO Journal*, 41(9), EMBJ2021110145. <https://doi.org/10.15252/embj.2021110145>
- Smith, T. F., Gaitatzes, C., Saxena, K., & Neer, E. J. (1999). The WD repeat: A common architecture for diverse functions. *Trends in Biochemical Sciences*, 24(5), 181–185. [https://doi.org/10.1016/S0968-0004\(99\)01384-5](https://doi.org/10.1016/S0968-0004(99)01384-5)
- Stach, L., & Freemont, P. S. (2017). The AAA+ ATPase p97, a cellular multitool. *Biochemical Journal*, 474(17), 2953–2976. <https://doi.org/10.1042/BCJ20160783>

- Steiner, M. R., Bomalaski, J. S., & Clark, M. A. (1993). Responses of purified phospholipases A2 to phospholipase A2 activating protein (PLAP) and melittin. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 1166(1), 124–130. [https://doi.org/10.1016/0005-2760\(93\)90292-H](https://doi.org/10.1016/0005-2760(93)90292-H)
- Stirnemann, C. U., Petsalaki, E., Russell, R. B., & Müller, C. W. (2010). WD40 proteins propel cellular networks. *Trends in Biochemical Sciences*, 35(10), 565–574. <https://doi.org/10.1016/j.tibs.2010.04.003>
- Stolz, A., Hilt, W., Buchberger, A., & Wolf, D. H. (2011). Cdc48: A power machine in protein degradation. *Trends in Biochemical Sciences*, 36(10), 515–523. <https://doi.org/10.1016/j.tibs.2011.06.001>
- Struhl, K. (2024). Intrinsically disordered regions (IDRs): A vague and confusing concept for protein function. *Molecular Cell*, 84(7), 1186–1187. <https://doi.org/10.1016/j.molcel.2024.02.023>
- Sullivan, P. F., Yao, S., & Hjerling-Leffler, J. (2024). Schizophrenia genomics: Genetic complexity and functional insights. *Nature Reviews Neuroscience*, 25(9), 611–624. <https://doi.org/10.1038/s41583-024-00837-7>
- Sun, G. Y., Shelat, P. B., Jensen, M. B., He, Y., Sun, A. Y., & Simonyi, A. (2010). Phospholipases A2 and Inflammatory Responses in the Central Nervous System. *Neuromolecular Medicine*, 12(2), 133–148. <https://doi.org/10.1007/s12017-009-8092-z>
- Sun, Y., Li, X.-Q., Sahbaie, P., Shi, X.-Y., Li, W.-W., Liang, D.-Y., & Clark, J. D. (2012). miR-203 Regulates Nociceptive Sensitization after Incision by Controlling Phospholipase A2 Activating Protein Expression. *Anesthesiology*, 117(3), 626–638. <https://doi.org/10.1097/ALN.0b013e31826571aa>

- Thul, P. J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L. M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L., Gnann, C., Hober, S., Hjelmare, M., Johansson, F., ... Lundberg, E. (2017). A subcellular map of the human proteome. *Science*, 356(6340), eaal3321. <https://doi.org/10.1126/science.aal3321>
- Tkach, J. M., Yimit, A., Lee, A. Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J. A., Ou, J., Moffat, J., Boone, C., Davis, T. N., Nislow, C., & Brown, G. W. (2012). Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nature Cell Biology*, 14(9), Article 9. <https://doi.org/10.1038/ncb2549>
- Tracz, M., & Bialek, W. (2021). Beyond K48 and K63: Non-canonical protein ubiquitination. *Cellular & Molecular Biology Letters*, 26(1), 1. <https://doi.org/10.1186/s11658-020-00245-6>
- Tsuchiya, H., Ohtake, F., Arai, N., Kaiho, A., Yasuda, S., Tanaka, K., & Saeki, Y. (2017). In Vivo Ubiquitin Linkage-type Analysis Reveals that the Cdc48-Rad23/Dsk2 Axis Contributes to K48-Linked Chain Specificity of the Proteasome. *Molecular Cell*, 66(4), 488-502.e7. <https://doi.org/10.1016/j.molcel.2017.04.024>
- Uversky, V. N., Gillespie, J. R., & Fink, A. L. (2000). Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Bioinformatics*, 41(3), 415–427. [https://doi.org/10.1002/1097-0134\(20001115\)41:3%253C415::AID-PROT130%253E3.0.CO;2-7](https://doi.org/10.1002/1097-0134(20001115)41:3%253C415::AID-PROT130%253E3.0.CO;2-7)
- Wagner, G. P., & Zhang, J. (2011). The pleiotropic structure of the genotype–phenotype map: The evolvability of complex organisms. *Nature Reviews Genetics*, 12(3), 204–213. <https://doi.org/10.1038/nrg2949>
- Wang, Q., Huang, L., Hong, Z., Lv, Z., Mao, Z., Tang, Y., Kong, X., Li, S., Cui, Y., Liu, H., Zhang, L., Zhang, X., Jiang, L., Wang, C., & Zhou, Q. (2017). The E3 ubiquitin ligase

- RNF185 facilitates the cGAS-mediated innate immune response. *PLOS Pathogens*, 13(3), e1006264. <https://doi.org/10.1371/journal.ppat.1006264>
- Wang, Q., Liu, X., Cui, Y., Tang, Y., Chen, W., Li, S., Yu, H., Pan, Y., & Wang, C. (2014). The E3 Ubiquitin Ligase AMFR and INSIG1 Bridge the Activation of TBK1 Kinase by Modifying the Adaptor STING. *Immunity*, 41(6), 919–933. <https://doi.org/10.1016/j.immuni.2014.11.011>
- Wilson, T. E. (2002). A genomics-based screen for yeast mutants with an altered recombination/end-joining repair ratio. *Genetics*, 162(2), 677–688. <https://doi.org/10.1093/genetics/162.2.677>
- Wimley, W. C. (2018). How Does Melittin Permeabilize Membranes? *Biophysical Journal*, 114(2), 251–253. <https://doi.org/10.1016/j.bpj.2017.11.3738>
- Wright, P. E., & Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1), Article 1. <https://doi.org/10.1038/nrm3920>
- Wu, X., Li, L., & Jiang, H. (2016). Doa1 targets ubiquitinated substrates for mitochondria-associated degradation. *The Journal of Cell Biology*, 213(1), 49–63. <https://doi.org/10.1083/jcb.201510098>
- Xu, C., & Min, J. (2011). Structure and function of WD40 domain proteins. *Protein & Cell*, 2(3), 202–214. <https://doi.org/10.1007/s13238-011-1018-1>
- Xu, S., Peng, G., Wang, Y., Fang, S., & Karbowski, M. (2011). The AAA-ATPase p97 is essential for outer mitochondrial membrane protein turnover. *Molecular Biology of the Cell*, 22(3), 291–300. <https://doi.org/10.1091/mbc.E10-09-0748>

- Yamanaka, K., Sasagawa, Y., & Ogura, T. (2012). Recent advances in p97/VCP/Cdc48 cellular functions. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1823(1), 130–137. <https://doi.org/10.1016/j.bbamcr.2011.07.001>
- Ye, Y., Meyer, H. H., & Rapoport, T. A. (2001). The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. *Nature*, 414(6864), 652–656. <https://doi.org/10.1038/414652a>
- Zhang, F. (2009). *A Missing link between lipid metabolism, inflammation and apoptosis: Phospholipase A2-activating protein (PLAA)*. <https://utmb-ir.tdl.org/handle/2152.3/39>
- Zhang, F., Sha, J., Wood, T. G., Galindo, C. L., Garner, H. R., Burkart, M. F., Suarez, G., Sierra, J. C., Agar, S. L., Peterson, J. W., & Chopra, A. K. (2008). Alteration in the activation state of new inflammation-associated targets by phospholipase A2-activating protein (PLAA). *Cellular Signalling*, 20(5), 844–861. <https://doi.org/10.1016/j.cellsig.2008.01.004>
- Zhang, X., Smits, A. H., van Tilburg, G. B. A., Jansen, P. W. T. C., Makowski, M. M., Ovaa, H., & Vermeulen, M. (2017). An Interaction Landscape of Ubiquitin Signaling. *Molecular Cell*, 65(5), 941–955.e8. <https://doi.org/10.1016/j.molcel.2017.01.004>
- Zhang, X., Shaw, A., Bates, P. A., Newman, R. H., Gowen, B., Orlova, E., Gorman, M. A., Kondo, H., Dokurno, P., Lally, J., Leonard, G., Meyer, H., Heel, M. van, & Freemont, P. S. (2000). Structure of the AAA ATPase p97. *Molecular Cell*, 6(6), 1473–1484. [https://doi.org/10.1016/S1097-2765\(00\)00143-X](https://doi.org/10.1016/S1097-2765(00)00143-X)
- Zhao, G., Zhou, X., Wang, L., Li, G., Schindelin, H., & Lennarz, W. J. (2007). Studies on peptide:N-glycanase–p97 interaction suggest that p97 phosphorylation modulates endoplasmic reticulum-associated degradation. *Proceedings of the National Academy of Sciences*, 104(21), 8785–8790. <https://doi.org/10.1073/pnas.0702966104>

- Zhao, X., Lutz, J., Höllmüller, E., Scheffner, M., Marx, A., & Stengel, F. (2017). Identification of Proteins Interacting with Ubiquitin Chains. *Angewandte Chemie International Edition*, 56(49), 15764–15768. <https://doi.org/10.1002/anie.201705898>
- Zheng, Z., Han, L., Li, Y., Chen, Z., Yang, W., Liu, C., Tao, M., Jiang, Y., Ke, X., Liu, Y., & Guo, X. (2023). Phospholipase A2-activating protein induces mitophagy through anti-apoptotic MCL1-mediated NLRX1 oligomerization. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1870(6), 119487. <https://doi.org/10.1016/j.bbamcr.2023.119487>
- Zhou, Q., & Zhang, J. (2022). K27-linked noncanonic ubiquitination in immune regulation. *Journal of Leukocyte Biology*, 111(1), 223–235. <https://doi.org/10.1002/JLB.4RU0620-397RR>
- Zuccoli, G. S., & Carregari, V. C. (2022). Mitochondrial Dysregulation and the Influence in Neurodegenerative Diseases. In V. Corasolla Carregari (Ed.), *Understanding PTMs in Neurodegenerative Diseases* (pp. 109–118). Springer International Publishing. https://doi.org/10.1007/978-3-031-05460-0_8

Chapter 4

STRUCTURAL CHARACTERIZATION OF THE FIP200 C-TERMINAL
CLAW DOMAIN BOUND TO AN AUTOPHAGY ADAPTOR PEPTIDE,
TNIP1-PFIR

With acknowledgement to:

William Rosencrans, David Nyenhuis, Ryan Lee, Jan Matthias, and Richard Youle

Abstract

Macroautophagy relies on selective cargo recognition by the ULK1 complex, in which FIP200 functions as a central scaffold linking substrates to the nascent autophagosome. The C-terminal CLAW domain of FIP200 is critical for binding autophagy substrates via adaptor proteins, and is therefore an emerging target for autophagy-directed therapeutics. However, the molecular determinants governing CLAW–adaptor recognition remain incompletely defined. Here, we report the 1.42 Å crystal structure of the phosphorylated FIP200-interacting region (pFIR) of the NF-κB regulator TNIP1 in complex with the FIP200 CLAW dimer. The TNIP1-pFIR peptide engages a basic surface on the outer face of CLAW primarily through electrostatic complementarity, while a conserved phenylalanine interacts with a network of aromatic residues, substantially stabilizing the interaction. Peptide binding induces an allosteric rearrangement of the CLAW dimer interface, leading to altered aromatic stacking interactions and reducing the buried interface area. Fluorescence polarization assays comparing TNIP1-pFIR to other FIRs demonstrate that phosphorylation markedly enhances TNIP1-FIR affinity for CLAW and explore the contribution of aromatic interactions to FIR-CLAW stability. Together, these data support the growing body of literature investigating and defining FIRs, and highlight the impact of CLAW-adaptor interactions on the development of small molecules designed to initiate selective autophagy.

1. Introduction

Macroautophagy (hereafter autophagy) is a conserved process that is essential for the degradation and turnover of cellular components, including organelles, protein aggregates, and other large intracellular structures (reviewed in Chew & Yip, 2014; Feng et al., 2014; Klionsky, 2007; Yang & Klionsky, 2009, 2010). It is essential for maintaining homeostasis under both basal and stress conditions. Cargo destined for destruction is engulfed in a double membrane structure, known as the autophagosome, which then fuses with the proteolytic enzyme-containing lysosome, permitting the degradation of autophagosome contents. Autophagy proceeds through four coordinated steps: initiation, in which cargo is recruited by the ULK-1 complex; nucleation, where membrane-forming components are recruited, including the class III phosphatidylinositol 3-kinase (PtdIns3K) complex; elongation of the phagophore around the cargo to create the classical double-membrane autophagosome; and maturation via fusion with the lysosome (Figure 1 A) (summarized in Adriaenssens et al., 2022; Parzych & Klionsky, 2014). Cargo recruitment is described in the literature as occurring either through constitutive bulk processes or as a selective process (reviewed in Lizarrondo & Wilfling, 2024). Selective autophagy is enabled by a vast repertoire of autophagy receptors (also referred to as adaptors) that recognize key components of the machinery forming the autophagosome (see e.g. Vargas et al., 2023). Autophagy adaptor recognition can be performed by members of the ATG8 family such as LC3, single-domain proteins which bear some resemblance to ubiquitin and which are lipidated with phosphatidylethanolamine in early autophagy to anchor them in the membrane of the nascent phagophore (Figure 1 B), and the C-terminal domain of FIP200, called the “CLAW” (Johansen & Lamark, 2020; Vargas et al., 2019).

FIP200 is a component of the ULK-1 complex, which acts to bind cargo for autophagy and phosphorylate downstream autophagy effectors (Hara et al., 2008a; Lin & Hurley, 2016).

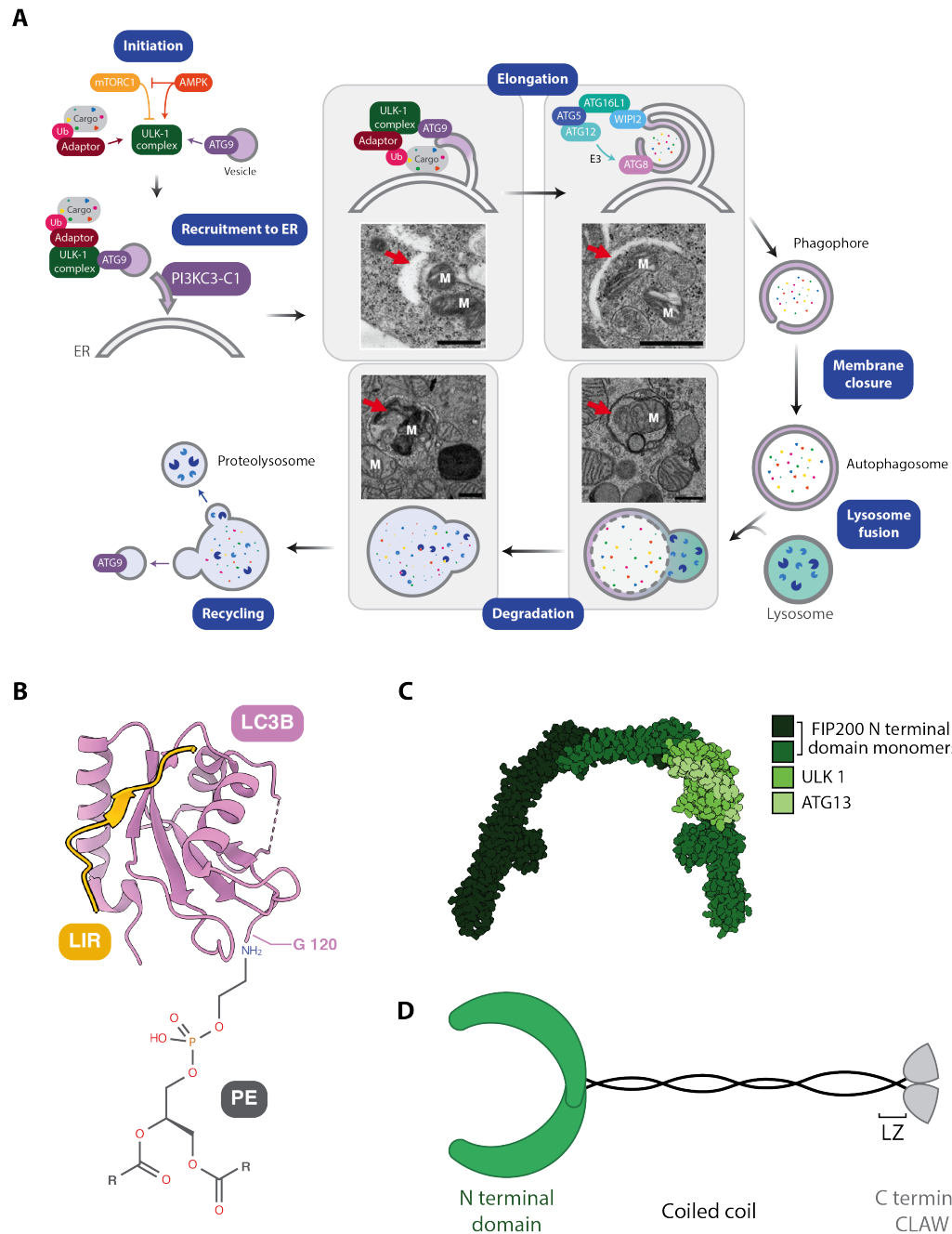


Figure 1. Autophagy background. A. Autophagy steps highlighting major players. Micrographs are taken from Jung et al. (2019). B. Cartoon representation of LC3B, an ATG8 protein, showing it associated to a LIR peptide and modified by phosphatidylethanolamine lipid (R groups indicate hydrocarbon chains). C. ULK1 2:1:1 FIP200 N-terminal domain:ULK1:ATG13 complex structure from Chen et al. (2025). D. Schematic of FIP200 dimer domain architecture. N-terminal domains are shown in green, alpha helical coiled-coil connectors are in black, and the C-terminal CLAW domain is in gray. The leucine zipper (LZ) region immediately preceding the CLAW is also indicated.

paralog, ULK-2), ATG13, ATG101, and FIP200 (also called RB1CC1) (Figure 1 C) (Lin & Hurley, 2016). Phosphorylation regulates many steps of autophagy, including its initiation: ULK1 and ATG13 both possess multiple phosphorylation sites, and may be phosphorylated by kinases both for negative and positive regulation (Bhattacharya et al., 2024; Dorsey et al., 2009). For example, mTORC1 phosphorylates ULK1/2 and ATG13 under nutrient-rich conditions and suppresses autophagy initiation, while AMPK phosphorylation of ULK-1/2 permits translocation of the ULK-1 complex to pre-autophagosomal membranes, allowing phosphorylation by ULK-1 to activate downstream autophagy processes (Hosokawa et al., 2009; J. Kim et al., 2011) (also note Park et al 2023 for AMPK rebuttal). FIP200, which forms a dimer *in vivo*, acts as a scaffold, supporting the ULK-1 complex proteins with its N-terminal domain and linking them to the cargo bound by its C-terminal domain, the CLAW (Hara et al., 2008b; Popelka & Klionsky, 2022; Turco et al., 2019). Its N and C termini are linked by an alpha helical coiled coil region with a leucine zipper at its C-terminal end, immediately before the CLAW (Figure 1 D) (Chano et al., 2002).

The FIP200 CLAW is a small, 12 kDa domain that forms a 24 kDa homodimer with C2 symmetry (Figure 2 A) (Turco et al., 2019). The interface between the two domains forms a hydrophobic pocket, making it an attractive target for small-molecule drug design (Figure 2 B). Additionally, a group of arginine and lysine residues form a basic patch, which locates to the external faces of the CLAW dimer (Figure 2 C). This basic patch is of interest, as it has been found to be involved in recognizing peptides from putative autophagy receptors, such as p62, Optineurin and CCPG1, and therefore is suspected to be critical to autophagy receptor recognition by FIP200 *in vivo* (Turco et al., 2019; Zhou et al., 2021). Vargas et al. (2019) found that autophagy could be initiated by forcing the recruitment of FIP200 to a substrate using chemically induced dimerization, suggesting that receptors could be replaced by a chimeric small molecule drug (Vargas et al., 2019). These bifunctional molecules, reminiscent of a PROTAC, or proteolysis-targeting chimera, would bind the target for

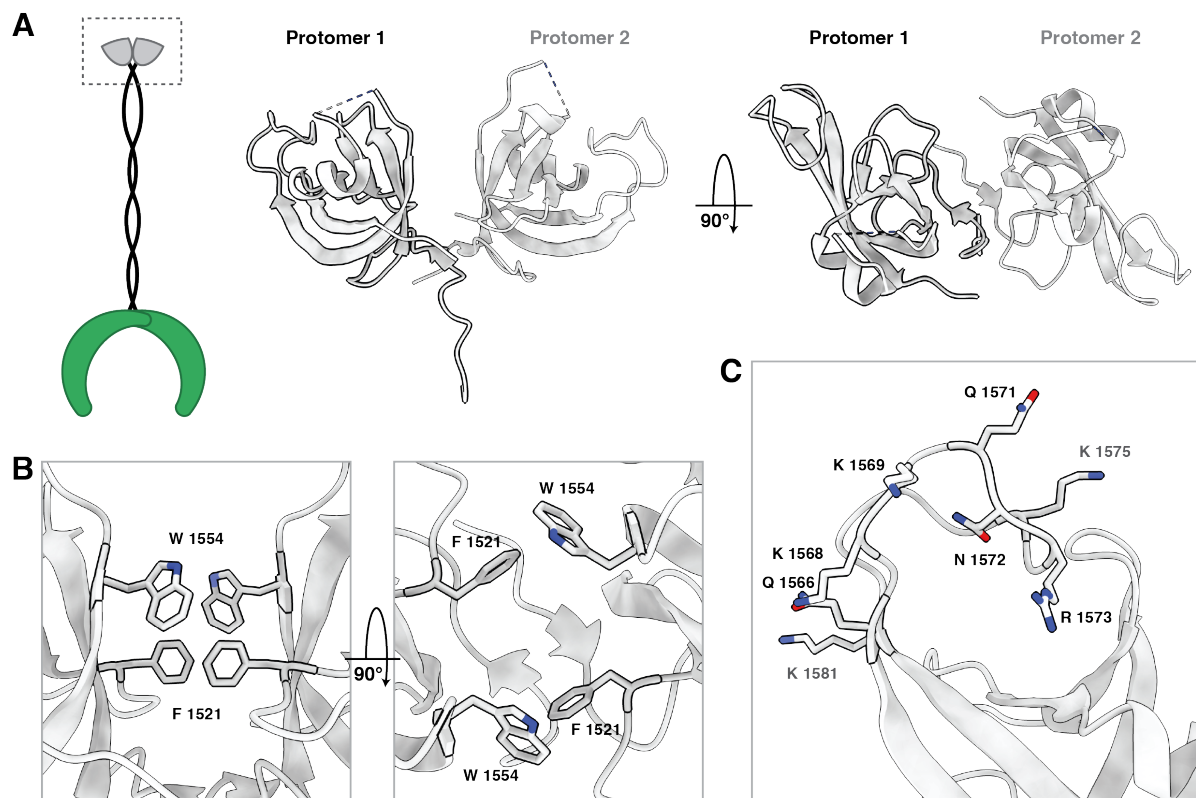


Figure 2. FIP200 CLAW domain features. A. Diagram of full-length FIP200 indicating the CLAW domain dimer and side and top views of a cartoon model of the CLAW dimer, which shows C2 symmetry. B. Side and top views of hydrophobic “basket” formed from F1521 and W1554 contributed by both dimer partners. C. Side view of basic loop between beta strands 4 and 5, highlighting charged and amidic residues (Q1566, K1568, K1569, Q1571, N1572, R1573, K1575, and K1581) contributing to the loop’s coulombic potential.

autophagy at one end and the FIP200 CLAW at the other. Development of an Auto-TAC (autophagy-targeting chimera) has therefore been underway for several years.

Bifunctionality is a critical component of autophagy receptors. These proteins must recognize both ubiquitin or other substrate-specific markers and FIP200 or ATG8 family proteins simultaneously to effectively connect the substrate to the autophagy machinery. ATG8 recognition is enabled by short, helical motifs termed “LC3 interacting regions”, or LIRs, that mediate the interaction. LIRs are found across proteins that interact with ATG8s, sharing a conserved sequence motif F(W/Y)_X_X_V(L/I) (Johansen & Lamark, 2020).

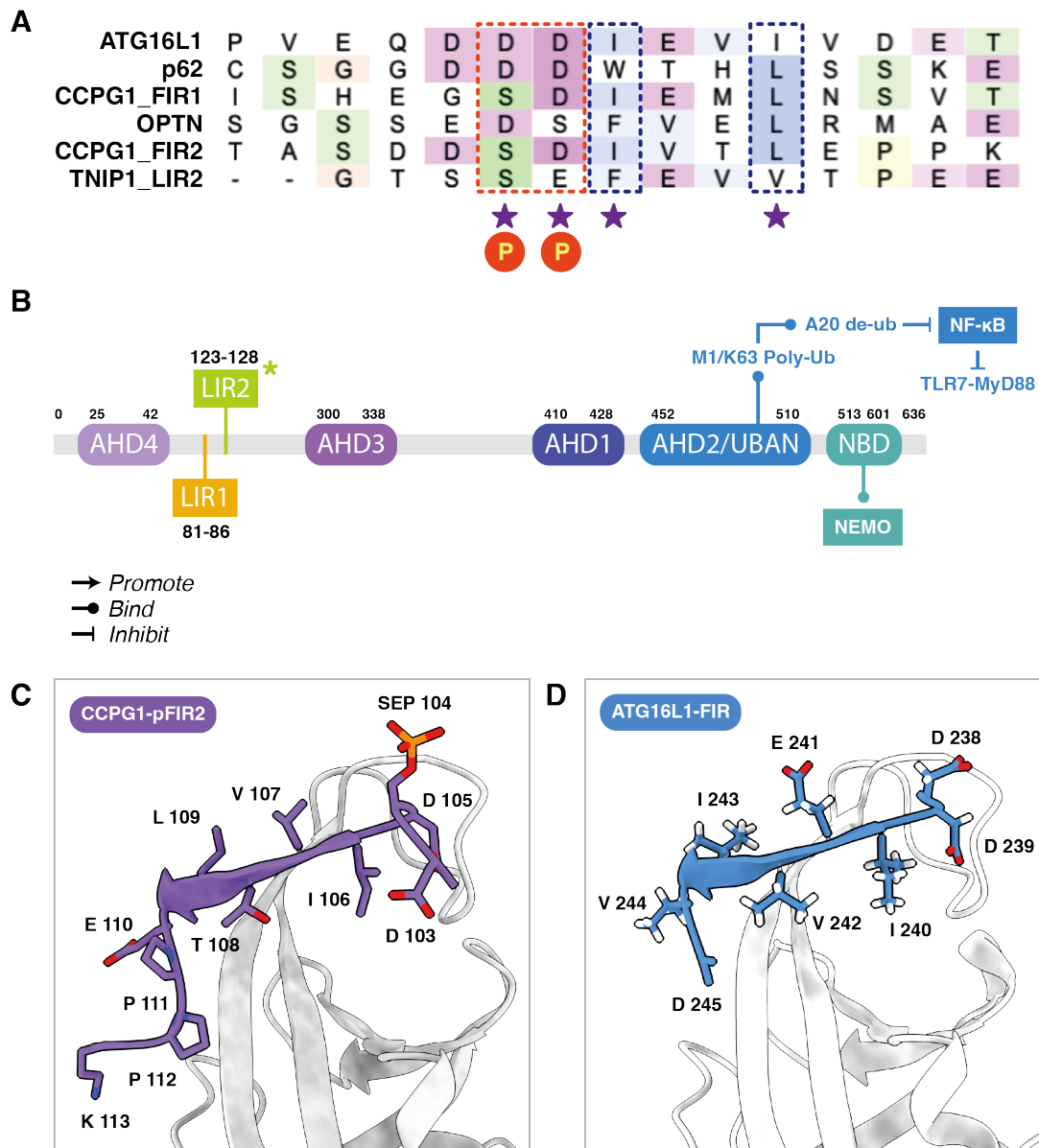


Figure 3. FIP200 interacting regions and TNIP1 background. A. Sequence alignment (CLUSTAL Omega) of FIRs with the TNIP1_LIR2 included for comparison. B. Domain architecture of TNIP1, including indicating selected interactors/pathways. Lines with dots indicate binding or activation, T-bars indicate suppression. C and D. Cartoon representations of experimental structures of CCPG1 (PDB ID: 7D0E) (C) and ATG16L1 (PDB ID: 9J54) (D) FIRs in complex with the CLAW loop 4-5. Peptide residues are labeled.

Receptor recognition of the FIP200 CLAW targets a similar FIP200-interacting region (FIR)

(Smith et al., 2018; Turco et al., 2019; Zhou et al., 2021). Many identified FIRs share sequence similarity to LIRs, with a conserved FIR peptide sequence D_D_I(F/Y/W/L/V)_X_X_L(I/V) (Figure 3 A). Novel autophagy receptors may therefore be screened for through the identification of a domain containing surface-accessible LIRs or FIRs along with a ubiquitin-recognition domain (B.-W. Kim et al., 2016).

TNIP1 (also termed ABIN-1/NAF/VAN) is a multidomain protein composed of a ubiquitin-binding domain found in many ABIN proteins (UBAN – ubiquitin binding of ABIN and NEMO), three ABIN homology domains (AHDs 1, 3 and 4), and a NEMO binding domain (Figure 3 B) (Guerroué et al., 2023). It belongs to a larger family of ABIN proteins (A20-binding inhibitors of NF- κ B) that are responsible for preventing NF- κ B activation by cytokines and lipopolysaccharide (LPS) (Verstrepen et al., 2009, 2014). It has been found to sense linear (M1) and K63 polyubiquitin chains and scaffold the A20 deubiquitinase to targets in NF- κ B pathways (Heyninck et al., 1999; Hong et al., 2021; Mauro et al., 2006; Shamilov et al., 2020; Wagner et al., 2008). Mechanistically, ABIN1/TNIP1 requires its UBAN domain to curb TLR7–MyD88 signaling, and mice with a UBAN loss-of-function mutation develop lupus-like autoimmunity, positioning TNIP1 as a negative regulator at the level of post-receptor signalosomes (Nanda et al., 2011; Shamilov & Aneskievich, 2018). Genetic studies repeatedly associate TNIP1 with autoimmune and inflammatory disorders including SLE, psoriasis, systemic sclerosis and others, suggesting that small changes in its expression level or functionality can drastically affect immune homeostasis (Adrianto et al., 2012; Allanore et al., 2011; Gateva et al., 2009; H. Gong et al., 2020; Han et al., 2009; Kawasaki et al., 2010; Liu et al., 2018; Pasula et al., 2020). Recent work additionally shows a role for TNIP1 in STING activation, specifically in the context of autophagy receptors like p62 and NDP52 (Bunker et al., 2025).

TNIP1 is of particular interest because it is a known inhibitor of selective autophagy through the formation of contacts with ATG8 proteins such as LC3 and GABARAP, autophagy receptors like TAX1BP1 and optineurin (OPTN) (both involved in xenophagy and other autophagic processes (White et al., 2023; Ying & Yue, 2016), and with the FIP200 C-

terminal CLAW domain (Guerroué et al., 2023). Overexpression of TNIP1 slows PINK1/Parkin-evoked mitophagy and delays inner-membrane COXII turnover. Additionally, forced relocalization of TNIP1 to mitochondria can induce mitophagy in otherwise mitophagy-deficient cells lacking five major autophagy receptors, but this induction is lost in the absence of either FIP200 or WIPI2, indicating that TNIP1 ultimately acts upstream, likely with the early isolation-membrane machinery (Guerroué et al., 2023).

Recognition of ATG8 proteins by TNIP1 is enabled by two LIRs that mediate its interaction with LC3 and GABARAP. The two LIRs show differential binding, with LIR1 showing higher affinity for LC3 and LIR2 preferring GABARAP (Guerroué et al., 2023). It is possible that this motif also can bind to FIP200's CLAW domain. Phosphorylation of serine residues immediately preceding the LIR2 motif (S122, S123) has been reported to enhance its affinity for the FIP200 CLAW domain (K_d improving from ~46 μ M to single digit micromolar values upon phosphorylation), enabling TNIP1 to compete with canonical autophagy receptors for CLAW docking (Guerroué et al., 2023).

This is similarly seen in studies of the cell cycle progression 1 protein (CCPG1), an ER-resident protein that, like TNIP1, can bind both ATG8s and FIP200's CLAW domain, facilitating autophagy of the ER (Ishii et al., 2023; Smith et al., 2018). The CCPG1 FIR bears close resemblance to the TNIP1 FIR, including phosphorylation sites that modulate its affinity for the CLAW (Zhou et al., 2021). A crystal structure of the CCPG1 phosphorylated FIR (pFIR) in complex with the FIP200 CLAW has been described, showing that the CCPG1 pFIR binds to a basic loop on the outer face of the CLAW (PDB ID: 7D0E) (Figure 3 C). The same team to describe this structure also confirmed that ATG16L1 possesses a functioning FIR motif (Figure 3 D) (X. Gong et al., 2025). ATG16L1 is a critical component of autophagosome formation which forms an E3-like complex with ATG12 and ATG5 that recruits ATG8s, allowing the elongation of the phagophore (Adriaenssens & Martens, 2025; Chen et al., 2025; Lystad et al., 2019). However, the ATG16L1 FIR, in alignment with the sequence similarity observed between FIRs and LIRs, appears to also serve as a noncanonical LIR. Certain mutations to the ATG16L1 FIR can preserve ATG8 recognition while

disrupting FIP200 binding. Unlike CCPG1 and TNIP1, however, the ATG16L1 FIR is not regulated through phosphorylation (X. Gong et al., 2025).

Understanding the manner in which autophagy receptors bind to the FIP200 CLAW may allow a second avenue for Auto-TAC development. Here, we report the structure of the TNIP1-pFIR in complex with the CLAW dimer.

Methods

2.1. Recombinant protein purification

BL21 DE3 Gold E. coli were transformed with a pET15T vector expressing the globular CLAW domain (FIP200 residues 1490-1594) fused to an N-terminal 6x-His tag followed by a TEV cleavage site. Cells were grown to an optical density at 600 nm of 1.0 at 37°C and induced with 1 mM IPTG. Protein was then expressed overnight at 16°C with agitation. Bacteria were harvested and the pellet was mechanically disrupted with lysis buffer (100 mM Tris pH 7.4, 500 mM KCl, 5 mM MgCl₂, 20 mM Imidazole, 5% glycerol (v/v), 2 mM b-ME, protease inhibitors) with a microfluidizer (LM-10, Microfluidics). Cell debris were removed via centrifugation at 20,000 x g for 1 hour at 4°C. The lysate was then passed through a cell strainer and either applied to NiNTA resin via gravity column or incubated with NiNTA bead slurry overnight at 4°C, stirred. Beads were subsequently washed with buffer containing 10 mM imidazole (50 mM HEPES pH 7.4, 150 mM KCl, 5 mM MgCl₂, 10 mM Imidazole) until no protein was detectable in column eluent via Bradford assay. Then, proteins were eluted off the column using a stepwise gradient of buffers containing 100, 200, 300, and 400 mM imidazole (other buffer components: 50 mM HEPES pH 7.4, 150 mM KCl, 5 mM MgCl₂). Fractions were assessed via SDS-PAGE using either stain-free gels (Bio-RAD) or Coomassie stain (30% methanol (v/v), 5% glacial acetic acid (v/v), 0.1% Coomassie Brilliant Blue R-250 (w/v)). Fractions containing crystal claw with few contaminants were pooled and diluted to ~50 mM imidazole, and His-tagged TEV protease was added to the solution. Protein was incubated overnight at room temperature without agitation to permit efficient cleavage while minimizing protein aggregation. The solution was then reapplied to NiNTA resin to remove the TEV protease. The column was washed with wash buffer and once again protein was eluted from the column using a stepwise elution as above, based on prior observations of incomplete crystal CLAW recovery at low imidazole concentrations. Eluent was once again assessed via SDS-PAGE, and crystal CLAW-containing fractions were pooled and concentrated in 10 kda-cutoff filters (Amicon) to a final volume suitable for loading onto a size exclusion column. Size exclusion was then

performed over a Superdex 200 10/300 column (Cytiva), both as a means of further purification and to buffer exchange into crystallography buffer (20 mM HEPES pH 7.4, 150 mM KCl, 1 mM MgCl₂, 5% glycerol (v/v)). Fractions were assessed via SDS-PAGE and pooled according to their purity. Pooled fractions were then concentrated as much as possible, generally ~30 mg/mL.

2.2 Crystallization condition identification and optimization

Solutions of crystal CLAW with a 1:1.5 molar excess of TNIP1 were assembled to 10 mg/mL and 15 mg/mL final CLAW concentrations, diluted with crystallography buffer (above). Solutions were incubated for 30 minutes at 4°C to permit binding. Subsequently, crystal screens were set up in 96-well Intelli-plate trays (Art Robbins 102-0001-20) with 2 shallow wells per condition, permitting both protein concentrations to be tested at once. Screens were aliquoted using an Art Robbins Instruments Gryphon robot, with sitting drops formed from 200 nL protein solution and 200 nL well solution. Six commercially available sparse matrix and PEG crystallization screens were tested. Crystals were obtained in Morpheus I condition E4 (12.5% w/v PEG 1000, 12.5% w/V PEG 3350, 12.5% w/v MPD, 0.03 M of each ethylene glycol (diethyleneglycol, triethyleneglycol, tetraethyleneglycol, and pentaethyleneglycol), and 0.1 M MES/imidazole pH = 6.5) at 7°C following three months of incubation (Figure 4 A). Crystals could be replicated in this condition overnight at 22-25°C (Figure 4 B). Gel analysis of crystals shows CLAW, while the surrounding supernatant shows a lower concentration (Figure 4 C). Controls do not show crystals (Figure 4 D). Crystals show tetragonal pyramidal morphology and were weakly birefringent, and were sparse unless obtained through seeding. Using low concentrations of seed particles (created using Seed Beads according to the manufacturer's protocol (Hampton Research)), crystals ~200 um across were obtained (Figure 4 E).

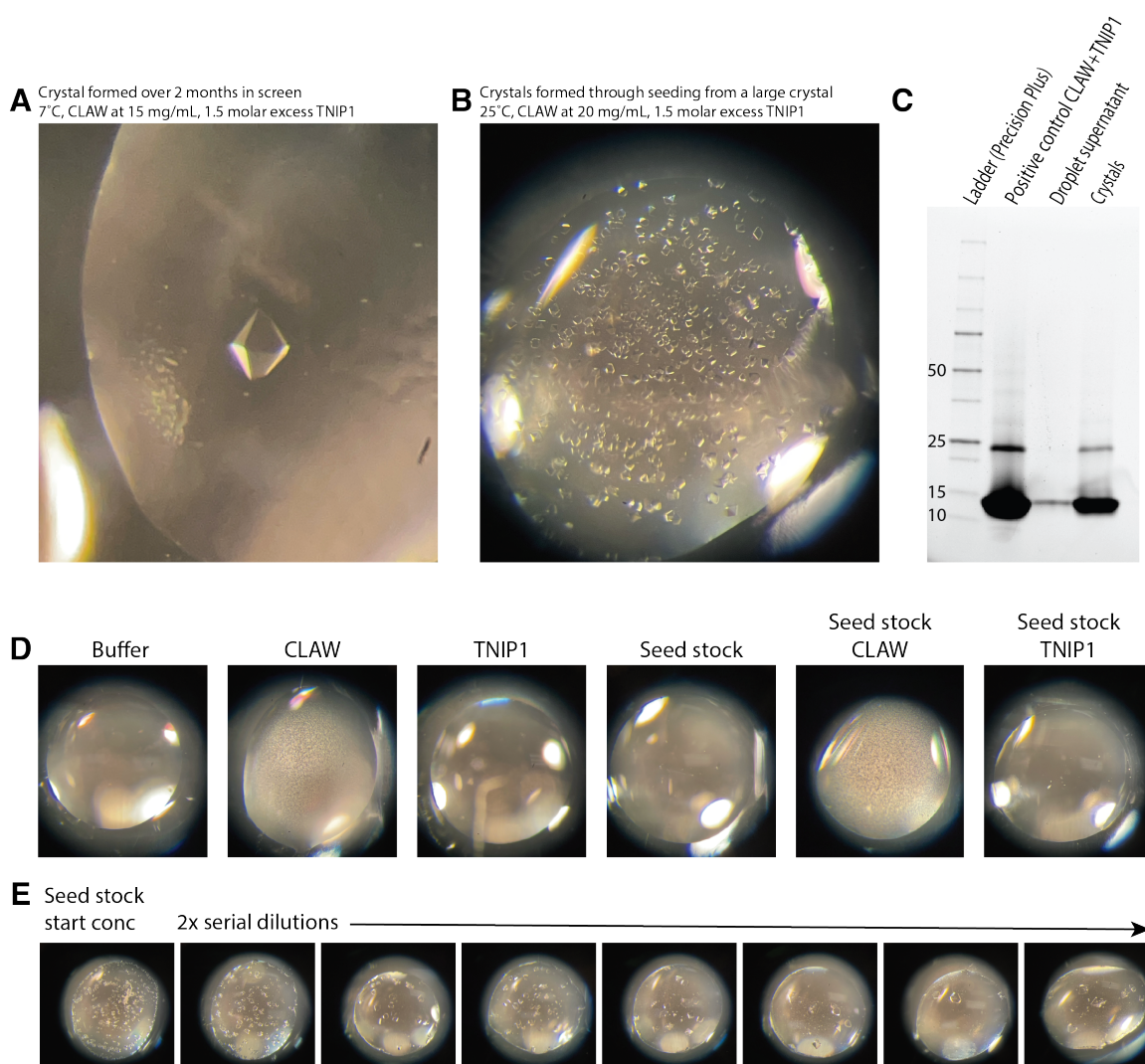


Figure 4. CLAW-TNIP1-pFIR crystals. A. Crystal identified in Morpheus 1 screen. B. Crystals formed after seeding. C. Stain-free gel (BioRad) of crystals to show that crystals contain CLAW. Lane 1: ladder (Precision Plus), lane 2: positive control containing CLAW and TNIP1-pFIR, lane 3: droplet supernatant, lane 4: crystal contents. D. Negative controls to demonstrate that crystal formation requires both CLAW and TNIP1-pFIR. Left to right: buffer alone, buffer plus CLAW alone, buffer plus TNIP1-pFIR alone, buffer plus seed stock alone, buffer plus seed stock and CLAW or buffer plus seed stock and TNIP1-pFIR. E. Effects of seed stock dilution on crystal number, morphology and size.

2.3 Crystal collection and diffraction experiments

During fishing, crystals were transferred to a droplet of well solution and re-fished to improve placement within the loop and to reduce contaminants. Crystals were plunge-frozen in liquid nitrogen with no additional cryoprotectant.

Diffraction data were collected remotely at the Stanford Synchrotron Radiation Lightsource (SSRL) at the Stanford Linear Accelerator Center (SLAC) on beamline 12-1 at 100K with a wavelength of 1.0 Å and a transmittance of 20%. All data were processed using the SSRL autoproccessing pipeline, wherein autoindexing and integration are carried out with XDS, analyzed with POINTLESS and XTRIAGE, scaled and merged with AIMLESS, and finally transformed to amplitudes using TRUNCATE. Crystals belonged to space group 92 (P 41 21 2) and the dataset chosen for publication diffracted to 1.42 Å. Phasing by molecular replacement was then performed in Phenix (version 1.21.2-5286-000) using a single monomer of PDB ID 7CZG (CLAW residues 1490-1594 solved with no peptide ligand). TNIP1 peptide was manually built into unmodelled density using the “C-alpha baton mode” modelling tool in Coot (v. 0.9.8.8). Refinements and validation were performed in Phenix and Coot. Figures were prepared with ChimeraX (v. 1.9). Data processing statistics are summarized in Table 1.

2.4 Fluorescence polarization peptide competition assays

Fluorescence polarization (FP) binding assays were performed using FITC-labeled peptides at a constant concentration of 100 nM and a two-fold serial dilution of crystal CLAW (FIP200 residues 1490-1594) with a high concentration of 100 µM and 10 nonzero dilutions, plus a CLAW-free negative control. Mixtures were assembled in a 96-well plate and divided upon transfer to the 384-well test plate to yield four technical replicates per condition. Fluorescence polarization data were collected using a BioTek Synergy Neo2 plate reader. Data were exported as a .csv and analyzed with a Python script (Script 1, made with ChatGPT). For each concentration, replicate polarization values were summarized as mean $\pm$ standard deviation after optional 3σ outlier rejection (points with $|x-\mu| > 3\sigma$ were excluded, and statistics recomputed). Binding curves were fit using nonlinear least squares to a 1:1 quadratic binding model, to account for the finite fluorescent peptide concentration. The

fraction of peptide bound was computed as $f_b = PL/P_0$, where $PL = ((P_0 + L + K_d) - \sqrt{((P_0 + L + K_d)^2 - 4P_0 L)})/2$, P_0 is the total peptide concentration (0.1 μM), and L is the total CLAW concentration. Polarization was modeled as $P(L) = P_{\text{"free"}} + (P_{\text{"bound"}} - P_{\text{"free"}})f_b$, where $P_{\text{"free"}}$, $P_{\text{"bound"}}$, and K_d were fit parameters. K_d values are reported in μM .

Results

3.1 TNIP1 binds to a region of positive charge on an outside face of the CLAW domain

The TNIP1-pFIR peptide associates with a positively charged loop formed between beta strands 4 and 5 and the solvent-exposed underside of the loop linking them together (loop 4-5) (Figure 5 A). Unlike the CCPG1-pFIR, the full length of the TNIP1-pFIR could not be resolved, leading to a total buried surface area of $\sim 454 \text{ \AA}^2$ between the TNIP1 peptide and CLAW loop 4-5. However, like the CCPG1-pFIR, TNIP1 pFIR binding is largely mediated by charge-charge interactions. Given the high B-factors associated with both the positively charged residues in the CLAW loop and the negatively charged residues and phosphate group resolved in the TNIP1-pFIR, these interactions do not appear to be specific, as no stable salt bridges are observed (Figure 5 B). Though the positive charge stems primarily from residues K1568, K1569, R1573 and R1584, these flexible residues appear to contribute instead to a region of delocalized positive Coulombic potential (Figure 5 C).

In addition to charge-charge interactions, the TNIP1 pFIR also forms a stable interaction between TNIP1 residue F125 and a network of phenylalanines within CLAW (F1529, F1574 and F1582) (Figure 6 A). These residues in CLAW interact both through offset stacking and T-stacking (as seen in pure benzene) (Headen et al., 2010). While the sidechains of F1529 and F1582 show rigidity in their positions across FIP200 CLAW experimental structures (e.g. PDB IDs 6DCE, 7CZG, 7DOE (shown in Figure 6 A)), the F1574 sidechain shows some variability. Apo structures of CLAW (6DCE, 7CZG) show F1574 participating in offset stacking with both F1529 and F1582 to some extent. Upon binding the CCPG1-pFIR, which substitutes an isoleucine (I106) in the position analogous to the TNIP1-pFIR F125, CLAW F1574 is turned to engage in offset stacking with only F1529, and both F1529 and F1574 stack edge-on with F1582. However, F125 in the TNIP1-pFIR pulls CLAW F1574 closer to itself, bringing it into an offset stacking arrangement with F1582 and enlarging this aromatic network and contributing significantly to the binding stability of the TNIP1 pFIR. This orientation change was recapitulated via NMR (Figure 6 B, NMR data from Dr. David

Nyenhuis, NIH). Further, fluorescent peptide competition assays (described below) with peptides that substitute an isoleucine for TNIP1 F125 (TNIP1 pFIR F125I) show a significantly lower K_d compared to WT.

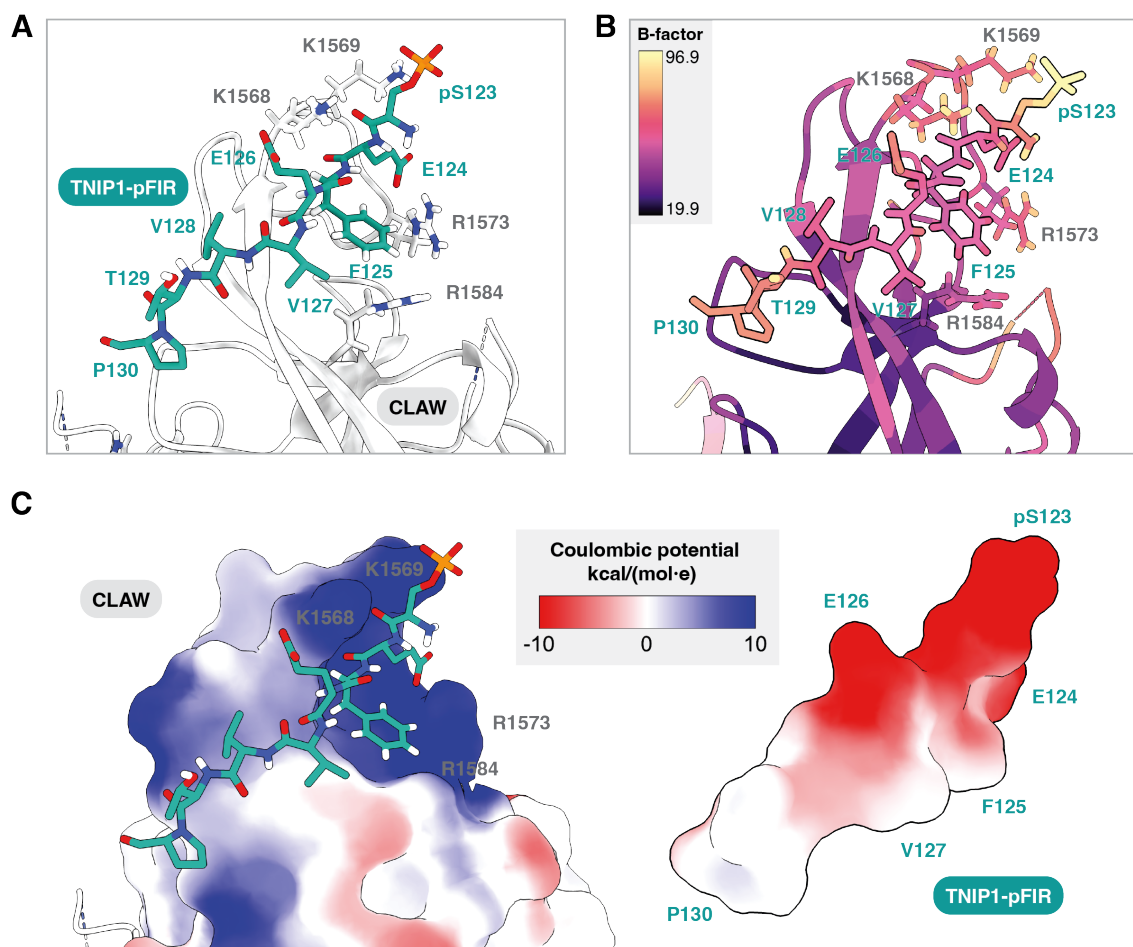


Figure 5. Crystal structure of TNIP1-pFIR in complex with CLAW. A. TNIP1-pFIR (teal) in complex with CLAW loop 4-5 (white). TNIP1-pFIR residues are labeled, plus four key interacting CLAW residues K1568, K1569, R1573 and R1584. B. View (A) colored by B-factor using Magma colormap: darker colors indicate low B-factors, lighter colors indicate high B-factors. C. View (A) colored to show surface coulombic potential of the CLAW loop 4-5 (left) and TNIP1-pFIR (right). Negative (acidic) potentials shown in red, positive (basic) potentials shown in blue. Calculations performed in ChimeraX (v. 1.9).

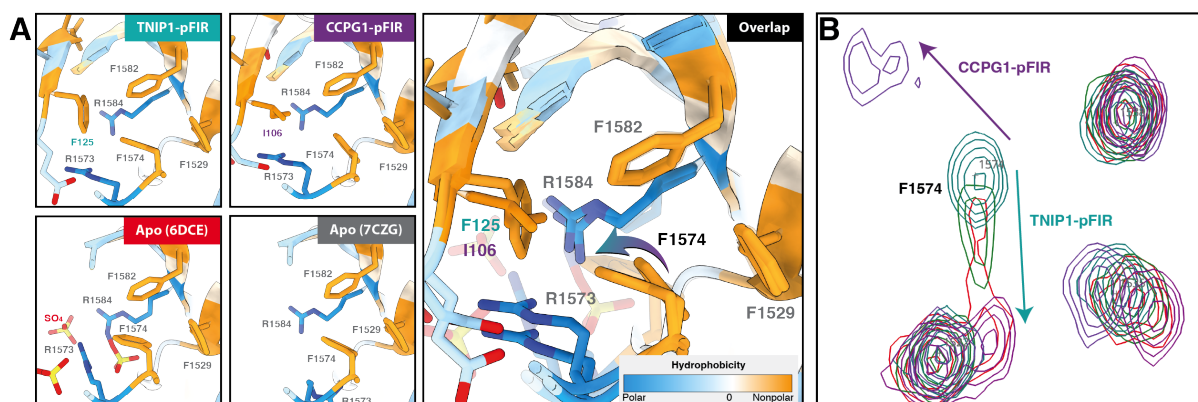


Figure 6. Interaction of conserved hydrophobic FIR residue with CLAW aromatic cluster. A. Structural comparison of CLAW aromatic cluster (F1529, F1574, F1582, plus contributions from R1573 and R1584) in two apo conformations and interacting with TNIP1-pFIR F125, CCPG1-pFIR I106. Right panel shows overlap of all structures, with purple-to-teal arrow indicating positional difference between F1574 bound to CCPG1-pFIR and TNIP1-pFIR. Structures are colored by hydrophobicity (Eisenberg), with blue indicating polar residues and orange indicating hydrophobic residues. B. NMR data comparing chemical shift of F1574 in apo state (center), when bound to the CCPG1-pFIR (upper left) and the TNIP1-pFIR (lower center).

3.2 Peptide binding changes the conformation of the CLAW dimer interface

As has been previously observed with the CCPG1-pFIR, binding to the TNIP1-pFIR induces what appears to be a shift in the dimer interface between CLAW monomers. FIP200 is known to exist in cells as a dimer (Shi et al., 2020). Dimerization of the globular CLAW construct (FIP200 residues 1490-1594) is documented by Turco et al (2019) and has been observed by NMR (Dr. David Nyenhuis) (Turco et al., 2019). Additionally, evidence from NMR supports a shift in the dimer interface upon peptide binding, but not a disruption of the dimers themselves (Figure 7 A, NMR data from Dr. David Nyenhuis, NIH). Therefore, we believe that, though the asymmetric unit of this crystal is a CLAW monomer, the biological unit is a dimer, and the demonstrated interface is biologically relevant rather than an artefact of crystal packing.

Though changes to the backbone structure of CLAW upon peptide binding are extremely subtle, monomers within a dimer undergo a rotational displacement relative to one another that dramatically alters the interface between them. The C alpha of W1554, found in the interface, shifts by nearly 12 Å between the apo and peptide-bound dimer conformations (Figure 7 B, left). Similarly, a shift of 24 Å is observed for A1570, located at the apex of loop 4-5 (Figure 7 B, right). Within the interface, this shift exchanges two T-stacking aromatic interactions at the interface, between the side chains of W1554 of one chain and F1521 of the other, to a single offset π - π interaction between each chain's F1521 side chain upon peptide binding. The energetic cost of losing a set of aromatic interactions is offset by the emergence of the interaction between W1554 with V1523 of the opposite chain.

Though these changes in the interface are visually striking, other interactions contribute more substantially to the association of the two monomers. Analysis of the unbound CLAW structure (PDB ID: 7CZG) shows that in the apo conformation, intrachain interactions between D1500 and R1491, D1505 and R1491, and D1505 and K1495 contribute much more to the stability of the dimer. Similarly, in the TNIP1-bound CLAW structure presented in this manuscript (PDB ID: 9D34), significant contributions to dimer stability are seen from the intrachain interactions between R1499 and E1494, and D1505 and K1495. The interaction of D1505 and K1495 contributes more to dimer formation of the unbound CLAW structure than TNIP1-bound CLAW, while the interaction of R1499 and E1494 shows the opposite trend. It is worth noting that the structure of Apo CLAW retains more residues at its N-terminus from FIP200 than the structure of CLAW bound to the TNIP1-pFIR, and thus there is no direct comparison available for the contributions of residues upstream of E1494.

Interestingly, the dimer interface shift observed upon peptide binding reduces the buried surface area of the CLAW dimer from 831.09 Å² to 619.49 Å². This increases the exposed area accessible to small molecule binding partners targeting the dimer interface, raising the possibility of synergy between CLAW-binding peptides and small molecules targeting dimeric CLAW. However, this depends largely on the specific manner in which a given small

molecule binds, and currently excludes the small molecules that have been tested thusfar for use as an AutoTAC.

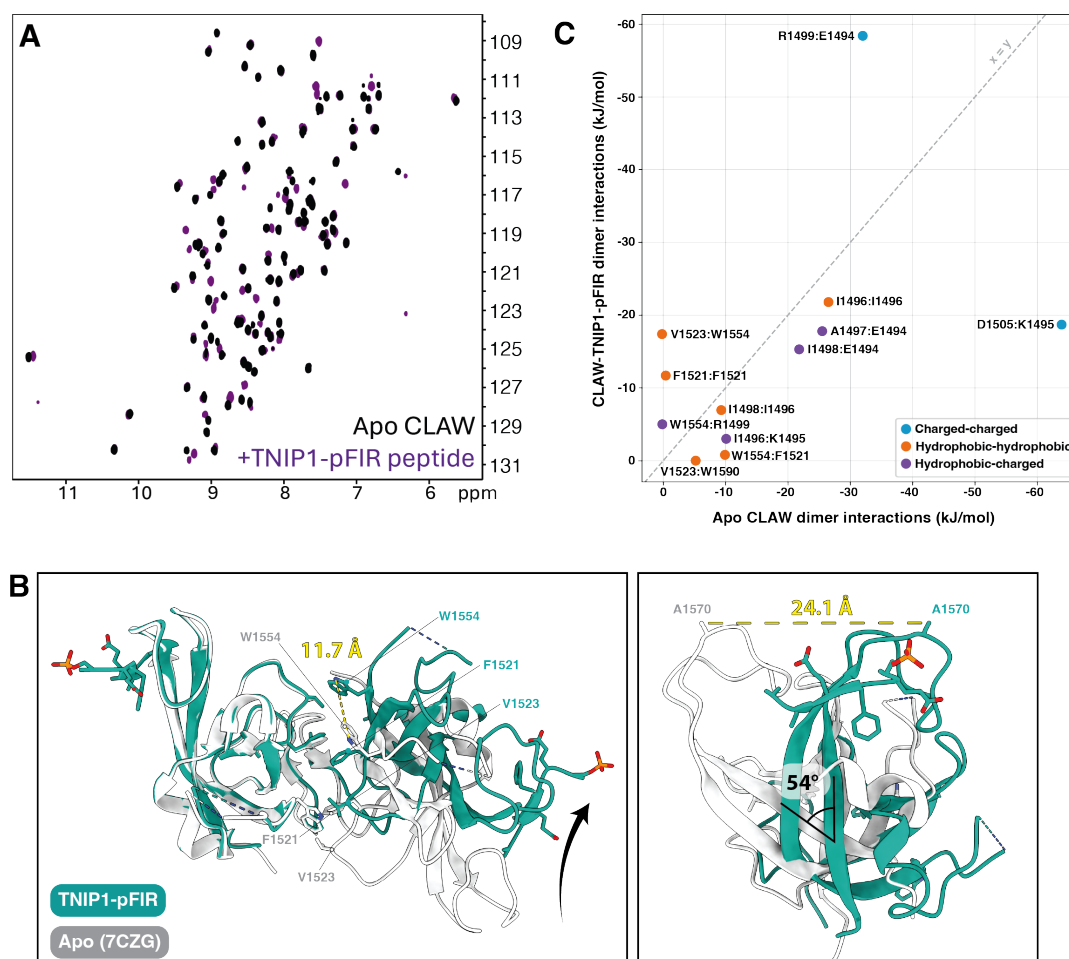


Figure 7. CLAW dimer interface changes upon peptide binding. A. NMR data to show residue positions in Apo CLAW (black) compared to CLAW with TNIP1-pFIR (purple). B. Overlay of Apo (PDB ID: 7CZG) (white) and TNIP1-pFIR-bound CLAW (teal) structures. Left: Top view with F1521, V1523 and W1554 labeled, showing distance shifted by W1554 as 11.7 Å. Arrow indicates rotation direction. Right: side view comparing Apo and TNIP1-pFIR-bound CLAW structures, labeling 54° rotation and 24.1 Å shift in A1570 position (top of loop 4-5). C. Comparison of interaction energies (kJ/mol) between interacting pairs in Apo CLAW dimer and TNIP1-pFIR-bound CLAW dimer interface. Increased negativity indicates increased stabilization. X=Y line allows visual assessment of which interactions are stronger in the apo vs. peptide bound structures.

3.3 Binding kinetics of TNIP1-pFIR in comparison with other FIR peptides

Though similar to each other, LIRs and FIRs have key sequence differences that set them apart from one another. Canonical LIRs have the sequence F(W/Y)_X_X_V(L/I), while the consensus FIR as reported adds two negatively charged residues to the LIR N terminus: D_D_I(F/Y/W/L/V)_X_X_L(I/V), where aspartate residues may be replaced by phosphoserines, phosphothreonines, or in some cases glutamate residues (Popelka & Klionsky, 2022; Zhou et al., 2021). These negatively charged residues play a crucial role in FIR recognition, supplying the largest energetic contribution to binding of the FIR peptide to the CLAW. Using the Amino Acid Interactions (INTAA) web server (v2.0) and the CHARMM 36 forcefield, D105, the second residue in the CCPG1-pFIR (PDB ID: 7D0E), is calculated to contribute -146.4 kJ/mol to peptide binding (total -566.35 kJ/mol), and E124 in the TNIP1 peptide, the analogous residue, to contribute -47.67 kJ/mol (total -138.76 kJ/mol) (Galgonek et al., 2017; Vymětal et al., 2021). Unfortunately, contributions from phosphorylated residues could not be calculated.

Phosphorylation of either S122 or S123 improves the binding affinity of the TNIP1-FIR peptide to the FIP200 CLAW. Guerroué et al. (2023) demonstrated that phosphorylation of S123 shows a greater improvement in binding affinity (data recapitulated by Will Rosencrans, Figure 8 A) (Guerroué et al., 2023). They additionally found that unphosphorylated TNIP1-FIR cannot compete with the CCPG1-pFIR (data from Will Rosencrans, Figure 8 B). Our structural data shows that, though no specific interactions to the phosphate group are formed, phosphorylation adds a large region of negative charge to the peptide, enhancing its association with the dispersed positive charge created by CLAW residues K1568, K1569, and Q1571. The effect of this phosphate group is performed in some FIRs by substitution with aspartate or glutamate, such as ATG16L1. This lack of kinase-dependent binding suggests that D-D containing FIRs like ATG16L1 have a more constitutive approach to CLAW binding than S-D containing FIRs like TNIP1 and CCPG1, possibly indicating an alternate pathway to recruit substrates and initiate autophagy.

The importance of these charge-charge interactions is further underscored by mutations of lysine or arginine residues in loop 4-5 to glutamate, swapping from a positive to a negative charge residue (data from William Rosencrans, Figure 8 C). While the TNIP1-pFIR binds to the wild-type CLAW with low-micromolar affinity, mutations K1568E/K1569E, R1573E, and to all three prevent peptide binding. This further reflects the necessity of these interactions to FIR recognition.

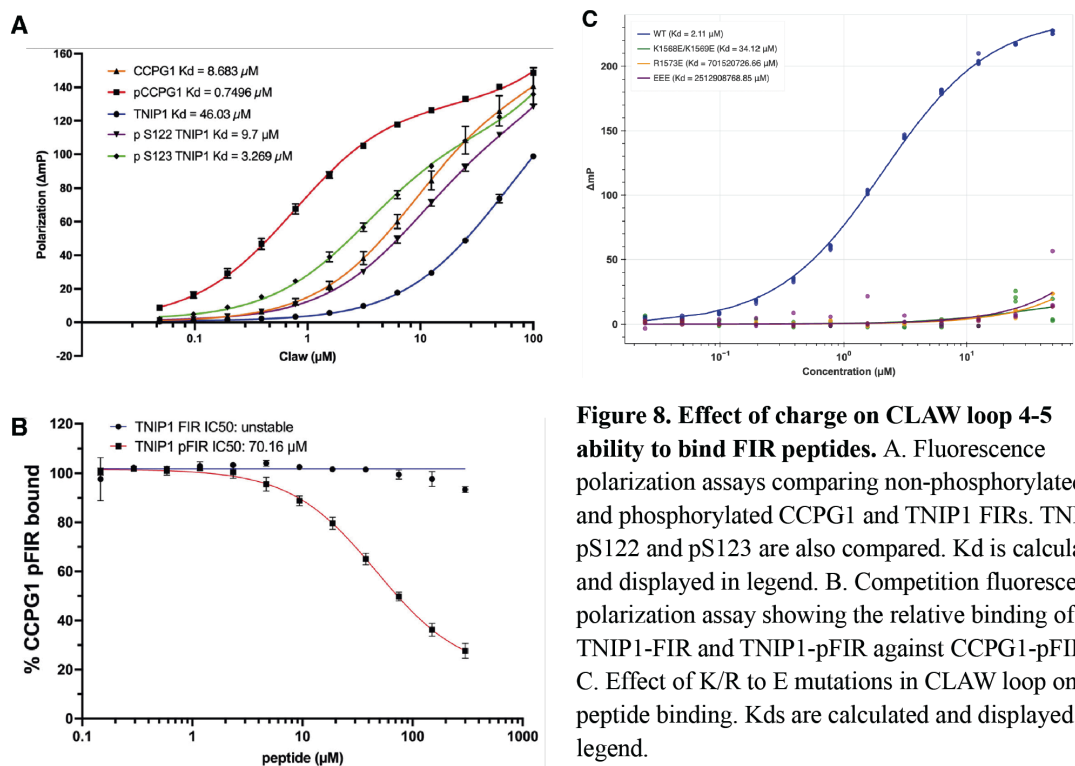


Figure 8. Effect of charge on CLAW loop 4-5 ability to bind FIR peptides. A. Fluorescence polarization assays comparing non-phosphorylated and phosphorylated CCPG1 and TNIP1 FIRs. TNIP1 pS122 and pS123 are also compared. Kd is calculated and displayed in legend. B. Competition fluorescence polarization assay showing the relative binding of TNIP1-FIR and TNIP1-pFIR against CCPG1-pFIR. C. Effect of K/R to E mutations in CLAW loop on peptide binding. Kds are calculated and displayed in legend.

While negatively charged residues are critical to FIR recognition and binding by the acidic, solvent-exposed residues of CLAW loop 4-5, the hydrophobic residues that define the FIR's similarity to a LIR also provide important contacts to the CLAW. The most N-terminal of these residues, TNIP1-pFIR residue F125, interacts with a highly aromatic pocket formed in the CLAW by F1529, F1574 and F1582, as well as with the π systems of either R1573 or R1584. Given the previously observed differences in the CCPG1-pFIR binding kinetics

compared to TNIP1's pFIR, we sought to understand the contribution from these hydrophobic residues by creating mutant peptides substituting an isoleucine for TNIP1's F125 and phenylalanine for CCPG1's I106. Calculations performed on the INTAA web server using the experimental structure of CLAW bound to the CCPG1 pFIR (PDB ID: 7D0E) show that I106 contributes 67.95 kJ/mol to the stability of the peptide-CLAW interaction, far less than the contribution from D105 (-151.59 kJ/mol). However, F125 in the TNIP1 pFIR contributes -66.31 kJ/mol (compared to -47.67 kJ/mol from E124).

3.4 Fluorescence polarization assay shows contributions from hydrophobic residues to FIR binding

A fluorescence polarization assay was designed to experimentally examine the binding kinetics of the FIR peptides. This assay uses the ratio of polarized to unpolarized light to compare the relative polarization of fluorescent objects in solution. Smaller objects – here, unbound peptide – tumble faster and therefore lack a polarized orientation. Larger objects will tumble more slowly and their orientation shows greater polarization. Analyzing the ratio of polarized to unpolarized fluorescence from FITC-labeled peptides at a low concentration combined with CLAW at a range of concentrations will allow us to assess the amount of fluorescent peptide in solution for each condition, thus enabling us to determine a binding affinity for each peptide (Figure 9 A).

Fluorescent peptide competition assay analysis of the mutant peptides shows that, in both the CCPG1-pFIR and TNIP1-pFIR peptides, phenylalanine provides more stability than isoleucine in this position (Figure 9 B). The K_d of TNIP1-pS123-F125I is higher than that of WT, suggesting worse binding, while the K_d of CCPG1-pS104-I106F is slightly lower, suggesting better binding. However, the percentage contribution of this hydrophobic interaction to binding is much lower in CCPG1 than TNIP1, leading to a much lower impact from the I106F mutation. For peptide sequences and K_d s, see table 2 below.

Table 1. *TNIP1* and *CCPG1* pFIR wild type and phenylalanine-isoleucine mutant sequences and dissociation constants

Peptide	Sequence	Kd
<i>TNIP1</i> pS123	GTS{pSER}EFEVVTPEE	4.37 μ M
<i>CCPG1</i> pS104	TASDD{pSER}DIVTLEPPK	0.911 μ M
<i>TNIP1</i> pS123 F125I	GTS{pSER}EIEVVTPEE	9.07 μ M
<i>CCPG1</i> pS104 I106F	TASDD{pSER}DFVTLEPPK	0.797 μ M

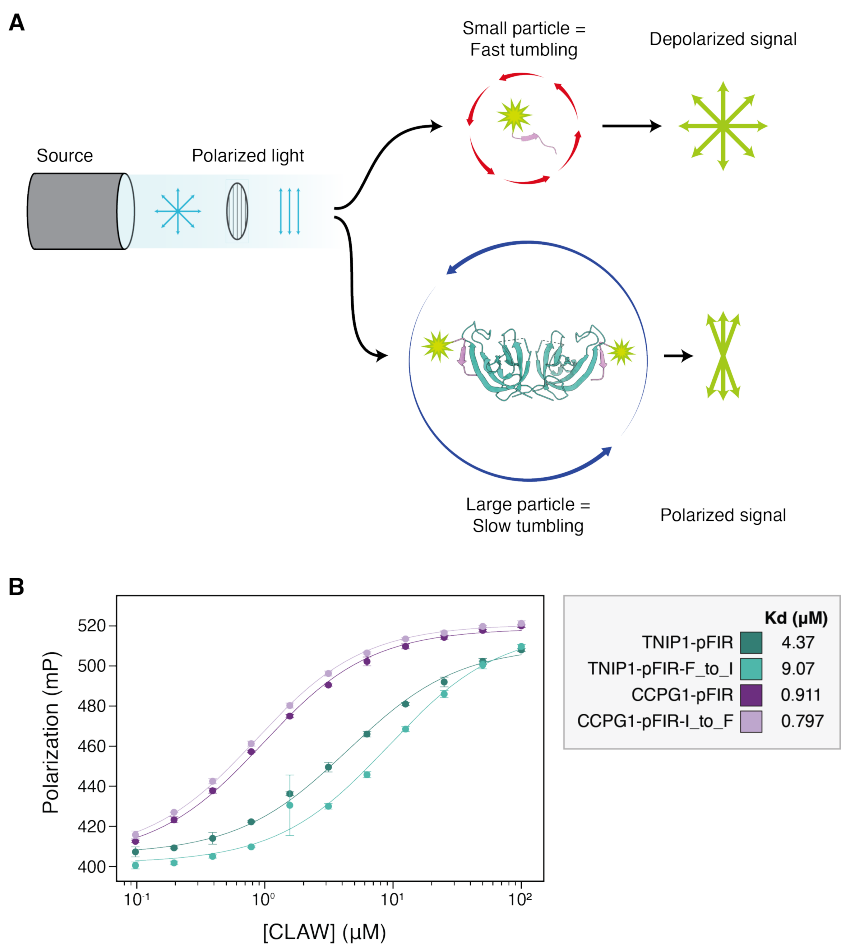


Figure 9. Fluorescence polarization assays. A. Schematic illustrating a fluorescence polarization assay. B. Fluorescence polarization assay to assess the effect of phenylalanine vs. isoleucine in the first hydrophobic position of the *TNIP1* and *CCPG1* pFIRs. Kd for each peptide is displayed in the legend.

Discussion

Selective autophagy depends on the precise recognition of cargo by the ULK1 complex, with FIP200 acting as a central scaffold linking autophagy initiation factors with adaptors and their substrates. Although the CLAW domain of FIP200 has emerged as a key receptor-binding module, the structural logic underlying this interaction has remained only partially defined. The structure of the TNIP1 pFIR–CLAW complex presented here supports our burgeoning understanding of how autophagy adaptors engage FIP200, and how phosphorylation may be a key step in regulating that process.

Like the CCPG1-pFIR, the TNIP1-pFIR binds to the CLAW at loop 4-5. This region features a set of basic residues (K1568, K1569, Q1571, and R1573), which interact with the peptide via delocalized charge-charge interactions. For peptides including the TNIP1 and CCPG1 FIRs, this interaction is enhanced by the addition of a phosphate group, which improves the electrostatic complementarity of the peptide with the positively charged CLAW loop. This distributed mode of interaction is well suited to the conformational flexibility of loop 4–5 and explains the marked affinity enhancement observed for pFIRs relative to their unmodified counterparts. More broadly, this suggests that post-translational modification may serve as a tunable switch that controls competitive access of receptors to FIP200.

Peptide binding also employs π – π interactions with residues forming a hydrophobic pocket in the CLAW (F1529, F1574, F1582). TNIP1-pFIR binding depends significantly on the presence of TNIP1-FIR residue F125, which inserts into this aromatic pocket. Substitution with isoleucine increases the K_d from 4.37 μ M to 9.07 μ M, weakening the interaction. However, the aromatic pocket appears to be less important to binding the CCPG1-pFIR, which features an isoleucine in this position natively (I106). Though mutating this residue to a phenylalanine does improve binding of the CCPG1-pFIR, the observed change is much smaller, improving from 0.911 μ M to 0.797 μ M.

We also find that peptide engagement allosterically remodels the CLAW dimer interface.

Although CLAW remains dimeric upon TNIP1-pFIR binding, the relative orientation of the monomers shifts, altering aromatic stacking interactions and reducing the buried surface area of the interface. While no biological function may yet be ascribed to this shift, it modulates the accessibility of the dimer interface to any regulatory factors or small molecules that may be targeted there. This conformational change, observed not only for TNIP1-pFIR-bound CLAW but also in the experimental structure of CLAW bound to CCPG1-pFIR, may therefore be functionally relevant in the assembly of the early autophagy machinery.

Together, these findings support our understanding of the structural and energetic basis for FIR recognition by FIP200 and provide a framework for the systematic identification of FIR-containing proteins as candidate autophagy receptors. Moreover, the definition of both electrostatic and hydrophobic hot spots on the CLAW surface offers new opportunities for the rational design of autophagy-targeting chimeras that exploit endogenous receptor-binding principles.

Table 2. Data collection and refinement statistics.

FIP200 CLAW (1490-1594) in complex with FIR peptide from TNIP1	
Wavelength	1.0 Å
Resolution range	33.57 - 1.42 (1.471 - 1.42)
Space group	P 43 21 2
Unit cell	51.339 51.339 88.28 90 90 90
Total reflections	586110 (48167)
Unique reflections	22984 (1889)
Multiplicity	25.5 (21.6)
Completeness (%)	96.30 (83.99)
Mean I/sigma(I)	21.41 (1.79)
Wilson B-factor	27.03
R-merge	0.08872 (1.496)
R-meas	0.09071 (1.532)
R-pim	0.01843 (0.3252)
CC1/2	0.996 (0.814)
CC*	0.999 (0.947)
Reflections used in refinement	22152 (1889)
Reflections used for R-free	1085 (89)
R-work	0.2017 (0.2574)
R-free	0.2334 (0.3530)
CC(work)	0.935 (0.918)
CC(free)	0.915 (0.875)
Number of non-hydrogen atoms	802
macromolecules	755
ligands	0
solvent	47
Protein residues	95
RMS(bonds)	0.011
RMS(angles)	1.26
Ramachandran favored (%)	98.86
Ramachandran allowed (%)	1.14
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	3.42
Average B-factor	38.09
macromolecules	37.76
solvent	43.48

Bibliography

- Adriaenssens, E., Ferrari, L., & Martens, S. (2022). Orchestration of selective autophagy by cargo receptors. *Current Biology*, 32(24), R1357–R1371. <https://doi.org/10.1016/j.cub.2022.11.002>
- Adriaenssens, E., & Martens, S. (2025). Mechanism of autophagy initiation by transmembrane selective autophagy receptors. *The EMBO Journal*, 1–11. <https://doi.org/10.1038/s44318-025-00615-w>
- Adrianto, I., Wang, S., Wiley, G. B., Lessard, C. J., Kelly, J. A., Adler, A. J., Glenn, S. B., Williams, A. H., Ziegler, J. T., Comeau, M. E., Marion, M. C., Wakeland, B. E., Liang, C., Kaufman, K. M., Guthridge, J. M., Alarcón-Riquelme, M. E., BIOLUPUS and GENLES Networks, Alarcón, G. S., Anaya, J.-M., ... Gaffney, P. M. (2012). Association of two independent functional risk haplotypes in TNIP1 with systemic lupus erythematosus. *Arthritis & Rheumatism*, 64(11), 3695–3705. <https://doi.org/10.1002/art.34642>
- Allanore, Y., Saad, M., Dieudé, P., Avouac, J., Distler, J. H. W., Amouyel, P., Matucci-Cerinic, M., Riemekasten, G., Airo, P., Melchers, I., Hachulla, E., Cusi, D., Wichmann, H.-E., Wipff, J., Lambert, J.-C., Hunzelmann, N., Tiev, K., Caramaschi, P., Diot, E., ... Martinez, M. (2011). Genome-Wide Scan Identifies TNIP1, PSORS1C1, and RHOB as Novel Risk Loci for Systemic Sclerosis. *PLOS Genetics*, 7(7), e1002091. <https://doi.org/10.1371/journal.pgen.1002091>
- Bhattacharya, A., Torggler, R., Reiter, W., Romanov, N., Licheva, M., Ciftci, A., Mari, M., Kolb, L., Kaiser, D., Reggiori, F., Ammerer, G., Hollenstein, D. M., & Kraft, C. (2024). Decoding the function of Atg13 phosphorylation reveals a role of Atg11 in bulk autophagy initiation. *EMBO Reports*, 25(2), 813–831. <https://doi.org/10.1038/s44319-023-00055-9>

- Bunker, E. N., Fischer, T. D., Zhu, P.-P., Guerroué, F. L., & Youle, R. J. (2025). TNIP1 and Autophagy Receptors regulate STING Signaling (p. 2025.04.21.649822). *bioRxiv*. <https://doi.org/10.1101/2025.04.21.649822>
- Chano, T., Ikegawa, S., Kontani, K., Okabe, H., Baldini, N., & Saeki, Y. (2002). Identification of RB1CC1, a novel human gene that can induce RB1 in various human cells. *Oncogene*, 21(8), 1295–1298. <https://doi.org/10.1038/sj.onc.1205178>
- Chen, M., Nguyen, T. N., Ren, X., Khuu, G., Cook, A. S. I., Zhao, Y., Yildiz, A., Lazarou, M., & Hurley, J. H. (2025). Structure and activation of the human autophagy-initiating ULK1C:PI3KC3-C1 supercomplex. *Nature Structural & Molecular Biology*, 32(9), 1596–1605. <https://doi.org/10.1038/s41594-025-01557-x>
- Chew, L. H., & Yip, C. K. (2014). Structural biology of the macroautophagy machinery. *Frontiers in Biology*, 9(1), 18–34. <https://doi.org/10.1007/s11515-014-1293-3>
- Dorsey, F. C., Rose, K. L., Coenen, S., Prater, S. M., Cavett, V., Cleveland, J. L., & Caldwell-Busby, J. (2009). Mapping the Phosphorylation Sites of Ulk1. *Journal of Proteome Research*, 8(11), 5253–5263. <https://doi.org/10.1021/pr900583m>
- Feng, Y., He, D., Yao, Z., & Klionsky, D. J. (2014). The machinery of macroautophagy. *Cell Research*, 24(1), Article 1. <https://doi.org/10.1038/cr.2013.168>
- Galgonek, J., Vymětal, J., Jakubec, D., & Vondrášek, J. (2017). Amino Acid Interaction (INTAA) web server. *Nucleic Acids Research*, 45(W1), W388–W392. <https://doi.org/10.1093/nar/gkx352>
- Gateva, V., Sandling, J. K., Hom, G., Taylor, K. E., Chung, S. A., Sun, X., Ortmann, W., Kosoy, R., Ferreira, R. C., Nordmark, G., Gunnarsson, I., Svenungsson, E., Padyukov, L., Sturfelt, G., Jönsen, A., Bengtsson, A. A., Rantapää-Dahlqvist, S., Baechler, E. C., Brown, E. E., ... Graham, R. R. (2009). A large-scale replication study identifies

- TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature Genetics*, 41(11), 1228–1233. <https://doi.org/10.1038/ng.468>
- Gong, H., Gao, S., Pu, X., Kang, X., & Wu, X. (2020). Association of rs610604 in TNFAIP3 and rs17728338 in TNIP1 gene polymorphisms with psoriasis susceptibility: A meta-analysis of case-control studies. *BMC Medical Genetics*, 21(1), 103. <https://doi.org/10.1186/s12881-020-01041-x>
- Gong, X., Zhou, Y., Wang, Y., Tang, Y., Liu, H., Zhou, X., Zhang, Y., Guo, H., Guo, Z., & Pan, L. (2025). Molecular bases of the interactions of ATG16L1 with FIP200 and ATG8 family proteins. *Nature Communications*, 16(1), 9035. <https://doi.org/10.1038/s41467-025-64097-4>
- Guerroué, F. L., Bunker, E. N., Rosencrans, W. M., Nguyen, J. T., Basar, M. A., Werner, A., Chou, T.-F., Wang, C., & Youle, R. J. (2023). TNIP1 inhibits selective autophagy via bipartite interaction with LC3/GABARAP and TAX1BP1. *Molecular Cell*, 83(6), 927–941.e8. <https://doi.org/10.1016/j.molcel.2023.02.023>
- Han, J.-W., Zheng, H.-F., Cui, Y., Sun, L.-D., Ye, D.-Q., Hu, Z., Xu, J.-H., Cai, Z.-M., Huang, W., Zhao, G.-P., Xie, H.-F., Fang, H., Lu, Q.-J., Xu, J.-H., Li, X.-P., Pan, Y.-F., Deng, D.-Q., Zeng, F.-Q., Ye, Z.-Z., ... Zhang, X.-J. (2009). Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nature Genetics*, 41(11), 1234–1237. <https://doi.org/10.1038/ng.472>
- Hara, T., Takamura, A., Kishi, C., Iemura, S., Natsume, T., Guan, J.-L., & Mizushima, N. (2008a). FIP200, a ULK-interacting protein, is required for autophagosome formation in mammalian cells. *The Journal of Cell Biology*, 181(3), 497–510. <https://doi.org/10.1083/jcb.200712064>
- Hara, T., Takamura, A., Kishi, C., Iemura, S., Natsume, T., Guan, J.-L., & Mizushima, N. (2008b). FIP200, a ULK-interacting protein, is required for autophagosome formation

in mammalian cells. *Journal of Cell Biology*, 181(3), 497–510.
<https://doi.org/10.1083/jcb.200712064>

Headen, T. F., Howard, C. A., Skipper, N. T., Wilkinson, M. A., Bowron, D. T., & Soper, A. K. (2010). Structure of π - π Interactions in Aromatic Liquids. *Journal of the American Chemical Society*, 132(16), 5735–5742. <https://doi.org/10.1021/ja909084e>

Heyninck, K., De Valck, D., Berghe, W. V., Van Crielinge, W., Contreras, R., Fiers, W., Haegeman, G., & Beyaert, R. (1999). The Zinc Finger Protein A20 Inhibits TNF-induced NF- κ B-dependent Gene Expression by Interfering with an RIP- or TRAF2-mediated Transactivation Signal and Directly Binds to a Novel NF- κ B-inhibiting Protein ABIN. *Journal of Cell Biology*, 145(7), 1471–1482.
<https://doi.org/10.1083/jcb.145.7.1471>

Hong, J.-Y., Lin, S.-C., Kuo, B.-J., & Lo, Y.-C. (2021). Structural and Biochemical Basis for Higher-Order Assembly between A20-Binding Inhibitor of NF- κ B 1 (ABIN1) and M1-Linked Ubiquitins. *Journal of Molecular Biology*, 433(18), 167116.
<https://doi.org/10.1016/j.jmb.2021.167116>

Hosokawa, N., Hara, T., Kaizuka, T., Kishi, C., Takamura, A., Miura, Y., Iemura, S., Natsume, T., Takehana, K., Yamada, N., Guan, J.-L., Oshiro, N., & Mizushima, N. (2009). Nutrient-dependent mTORC1 Association with the ULK1–Atg13–FIP200 Complex Required for Autophagy. *Molecular Biology of the Cell*, 20(7), 1981–1991.
<https://doi.org/10.1091/mbc.e08-12-1248>

Ishii, S., Chino, H., Ode, K. L., Kurikawa, Y., Ueda, H. R., Matsuura, A., Mizushima, N., & Itakura, E. (2023). CCPG1 recognizes endoplasmic reticulum luminal proteins for selective ER-phagy. *Molecular Biology of the Cell*, 34(4), ar29.
<https://doi.org/10.1091/mbc.E22-09-0432>

- Johansen, T., & Lamark, T. (2020). Selective Autophagy: ATG8 Family Proteins, LIR Motifs and Cargo Receptors. *Journal of Molecular Biology*, 432(1), 80–103. <https://doi.org/10.1016/j.jmb.2019.07.016>
- Jung, M., Choi, H., & Mun, J. Y. (2019). The autophagy research in electron microscopy. *Applied Microscopy*, 49(1), 11. <https://doi.org/10.1186/s42649-019-0012-6>
- Kawasaki, A., Ito, S., Furukawa, H., Hayashi, T., Goto, D., Matsumoto, I., Kusaoi, M., Ohashi, J., Graham, R. R., Matsuta, K., Behrens, T. W., Tohma, S., Takasaki, Y., Hashimoto, H., Sumida, T., & Tsuchiya, N. (2010). Association of TNFAIP3 interacting protein 1, TNIP1 with systemic lupus erythematosus in a Japanese population: A case-control association study. *Arthritis Research & Therapy*, 12(5), R174. <https://doi.org/10.1186/ar3134>
- Kim, B.-W., Kwon, D. H., & Song, H. K. (2016). Structure biology of selective autophagy receptors. *BMB Reports*, 49(2), 73–80. <https://doi.org/10.5483/BMBRep.2016.49.2.265>
- Kim, J., Kundu, M., Viollet, B., & Guan, K.-L. (2011). AMPK and mTOR regulate autophagy through direct phosphorylation of Ulk1. *Nature Cell Biology*, 13(2), 132–141. <https://doi.org/10.1038/ncb2152>
- Klionsky, D. J. (2007). Autophagy: From phenomenology to molecular understanding in less than a decade. *Nature Reviews Molecular Cell Biology*, 8(11), Article 11. <https://doi.org/10.1038/nrm2245>
- Lin, M. G., & Hurley, J. H. (2016). Structure and Function of the ULK1 Complex in Autophagy. *Current Opinion in Cell Biology*, 39, 61–68. <https://doi.org/10.1016/j.ceb.2016.02.010>

- Liu, X., Qin, H., Wu, J., & Xu, J. (2018). Association of TNFAIP3 and TNIP1 polymorphisms with systemic lupus erythematosus risk: A meta-analysis. *Gene*, 668, 155–165. <https://doi.org/10.1016/j.gene.2018.05.062>
- Lizarrondo, J., & Wilfling, F. (2024). Selective Autophagy of Macromolecular Complexes: What Does It Take to be Taken? *Journal of Molecular Biology*, 436(15), 168574. <https://doi.org/10.1016/j.jmb.2024.168574>
- Lystad, A. H., Carlsson, S. R., de la Ballina, L. R., Kauffman, K. J., Nag, S., Yoshimori, T., Melia, T. J., & Simonsen, A. (2019). Distinct functions of ATG16L1 isoforms in membrane binding and LC3B lipidation in autophagy-related processes. *Nature Cell Biology*, 21(3), 372–383. <https://doi.org/10.1038/s41556-019-0274-9>
- Mauro, C., Pacifico, F., Lavorgna, A., Mellone, S., Iannetti, A., Acquaviva, R., Formisano, S., Vito, P., & Leonardi, A. (2006). ABIN-1 Binds to NEMO/IKK γ and Co-operates with A20 in Inhibiting NF- κ B*. *Journal of Biological Chemistry*, 281(27), 18482–18488. <https://doi.org/10.1074/jbc.M601502200>
- Nanda, S. K., Venigalla, R. K. C., Ordureau, A., Patterson-Kane, J. C., Powell, D. W., Toth, R., C. Arthur, J. S., & Cohen, P. (2011). Polyubiquitin binding to ABIN1 is required to prevent autoimmunity. *Journal of Experimental Medicine*, 208(6), 1215–1228. <https://doi.org/10.1084/jem.20102177>
- Parzych, K. R., & Klionsky, D. J. (2014). An Overview of Autophagy: Morphology, Mechanism, and Regulation. *Antioxidants & Redox Signaling*, 20(3), 460–473. <https://doi.org/10.1089/ars.2013.5371>
- Pasula, S., Tessneer, K. L., Fu, Y., Gopalakrishnan, J., Pelikan, R. C., Kelly, J. A., Wiley, G. B., Wiley, M. M., & Gaffney, P. M. (2020). Role of Systemic Lupus Erythematosus Risk Variants With Opposing Functional Effects as a Driver of Hypomorphic Expression of 1 and Other Genes Within a Three-Dimensional Chromatin Network. *Arthritis & Rheumatology*, 72(5), 780–790. <https://doi.org/10.1002/art.41188>

- Popelka, H., & Klionsky, D. J. (2022). The RB1CC1 Claw-binding motif: A new piece in the puzzle of autophagy regulation. *Autophagy*, 18(2), 237–239. <https://doi.org/10.1080/15548627.2022.2029234>
- Shamilov, R., Ackley, T. W., & Aneskievich, B. J. (2020). Enhanced Wound Healing- and Inflammasome-Associated Gene Expression in TNFAIP3-Interacting Protein 1- (TNIP1-) Deficient HaCaT Keratinocytes Parallels Reduced Reepithelialization. *Mediators of Inflammation*, 2020(1), 5919150. <https://doi.org/10.1155/2020/5919150>
- Shamilov, R., & Aneskievich, B. J. (2018). TNIP1 in Autoimmune Diseases: Regulation of Toll-like Receptor Signaling. *Journal of Immunology Research*, 2018(1), 3491269. <https://doi.org/10.1155/2018/3491269>
- Shi, X., Yokom, A. L., Wang, C., Young, L. N., Youle, R. J., & Hurley, J. H. (2020). ULK complex organization in autophagy by a C-shaped FIP200 N-terminal domain dimer. *The Journal of Cell Biology*, 219(7), e201911047. <https://doi.org/10.1083/jcb.201911047>
- Smith, M. D., Harley, M. E., Kemp, A. J., Wills, J., Lee, M., Arends, M., Kriegsheim, A. von, Behrends, C., & Wilkinson, S. (2018). CCPG1 Is a Non-canonical Autophagy Cargo Receptor Essential for ER-Phagy and Pancreatic ER Proteostasis. *Developmental Cell*, 44(2), 217-232.e11. <https://doi.org/10.1016/j.devcel.2017.11.024>
- Turco, E., Witt, M., Abert, C., Bock-Bierbaum, T., Su, M.-Y., Trapannone, R., Sztacho, M., Danieli, A., Shi, X., Zaffagnini, G., Gamper, A., Schuschnig, M., Fracchiolla, D., Bernklau, D., Romanov, J., Hartl, M., Hurley, J. H., Daumke, O., & Martens, S. (2019). FIP200 Claw Domain Binding to p62 Promotes Autophagosome Formation at Ubiquitin Condensates. *Molecular Cell*, 74(2), 330-346.e11. <https://doi.org/10.1016/j.molcel.2019.01.035>

- Vargas, J. N. S., Hamasaki, M., Kawabata, T., Youle, R. J., & Yoshimori, T. (2023). The mechanisms and roles of selective autophagy in mammals. *Nature Reviews Molecular Cell Biology*, 24(3), 167–185. <https://doi.org/10.1038/s41580-022-00542-2>
- Vargas, J. N. S., Wang, C., Bunker, E., Hao, L., Maric, D., Schiavo, G., Randow, F., & Youle, R. J. (2019). Spatiotemporal Control of ULK1 Activation by NDP52 and TBK1 during Selective Autophagy. *Molecular Cell*, 74(2), 347–362.e6. <https://doi.org/10.1016/j.molcel.2019.02.010>
- Verstrepen, L., Carpentier, I., & Beyaert, R. (2014). The Biology of A20-Binding Inhibitors of NF- κ B Activation (ABINS). In C. Ferran (Ed.), *The Multiple Therapeutic Targets of A20* (pp. 13–31). Springer. https://doi.org/10.1007/978-1-4939-0398-6_2
- Verstrepen, L., Carpentier, I., Verhelst, K., & Beyaert, R. (2009). ABINs: A20 binding inhibitors of NF- κ B and apoptosis signaling. *Biochemical Pharmacology*, 78(2), 105–114. <https://doi.org/10.1016/j.bcp.2009.02.009>
- Vymětal, J., Jakubec, D., Galgonek, J., & Vondrášek, J. (2021). Amino Acid Interactions (INTAA) web server v2.0: A single service for computation of energetics and conservation in biomolecular 3D structures. *Nucleic Acids Research*, 49(W1), W15–W20. <https://doi.org/10.1093/nar/gkab377>
- Wagner, S., Carpentier, I., Rogov, V., Kreike, M., Ikeda, F., Löhr, F., Wu, C.-J., Ashwell, J. D., Dötsch, V., Dikic, I., & Beyaert, R. (2008). Ubiquitin binding mediates the NF- κ B inhibitory potential of ABIN proteins. *Oncogene*, 27(26), 3739–3745. <https://doi.org/10.1038/sj.onc.1211042>
- White, J., Suklabaidya, S., Vo, M. T., Choi, Y. B., & Harhaj, E. W. (2023). Multifaceted roles of TAX1BP1 in autophagy. *Autophagy*, 19(1), 44–53. <https://doi.org/10.1080/15548627.2022.2070331>

- Yang, Z., & Klionsky, D. J. (2009). An Overview of the Molecular Mechanism of Autophagy. *Current Topics in Microbiology and Immunology*, 335, 1–32. https://doi.org/10.1007/978-3-642-00302-8_1
- Yang, Z., & Klionsky, D. J. (2010). Eaten alive: A history of macroautophagy. *Nature Cell Biology*, 12(9), 814–822. <https://doi.org/10.1038/ncb0910-814>
- Ying, H., & Yue, B. Y. J. T. (2016). Optineurin: The autophagy connection. *Experimental Eye Research*, 144, 73–80. <https://doi.org/10.1016/j.exer.2015.06.029>
- Zhou, Z., Liu, J., Fu, T., Wu, P., Peng, C., Gong, X., Wang, Y., Zhang, M., Li, Y., Wang, Y., Xu, X., Li, M., & Pan, L. (2021). Phosphorylation regulates the binding of autophagy receptors to FIP200 Claw domain for selective autophagy initiation. *Nature Communications*, 12(1), 1570. <https://doi.org/10.1038/s41467-021-21874-1>

Chapter 5

ATP-DRIVEN TETRAMERIZATION AND RNA-GUIDED FILAMENT
ASSEMBLY CONTROL OF A SCHLAFEN–ARGONAUTE ANTIPHAGE
DEFENSE

Anton Kuzmenko, Kate Radford*, Anastasiya Oguienko, Tsui-Fen Chou, Andrey
Kulbachinskiy, Alexei A. Aravin*

**co-first authors*

NOTE: This chapter represents a manuscript that is actively being written and modified.

Abstract

Prokaryotic Argonaute proteins (pAgos) show remarkable architectural and functional diversity, yet how unusual domain fusions reprogram their immune activities remains largely unexplored. Here we identify a previously uncharacterized clade of extra-long pAgos fused to a Schlafen nuclease domain and a GHKL-family ATPase. These proteins, which we term SLFN-pAgos, are broadly distributed across bacteria and archaea and frequently associated with genomic defense islands. Using the SLFN-pAgo from *Runella zeae* as a model, we show that this system provides protection against bacteriophages by suppressing phage gene expression. Structural and biochemical analyses reveal that ATP binding drives tetramerization through the Schlafen–ATPase domains, whereas ATP hydrolysis is essential for antiviral activity. Unexpectedly, loading of short RNA guides induces assembly of extended helical filaments in which guide RNA bridges adjacent protomers through a non-canonical trajectory. Filament assembly enforces a domain arrangement incompatible with

the active Schlafen nuclease configuration, thereby suppressing toxicity, while target DNA binding reroutes the guide, and dismantles the filament. Together, our findings uncover a previously unknown mode of Argonaute regulation in which guide RNA and ATP cooperatively control higher-order assembly to gate effector activation during antiviral defense.

1. Introduction

Bacteria and archaea live under relentless threat from bacteriophages and other mobile genetic elements, driving the evolution of diverse defense strategies. Alongside CRISPR–Cas and restriction–modification systems, prokaryotic genomes encode many additional nucleic acid–targeting complexes whose mechanisms remain poorly understood (Anton et al., 2025; Sander & Joung, 2014). One such family is the Argonaute (Ago) proteins, which are conserved across all domains of life (Ryazansky et al., 2018; Swarts et al., 2014).

Argonautes use short nucleic acid guides to recognize complementary targets. In eukaryotes, this process underlies RNA interference, where small RNAs direct Argonautes to silence transcripts. Prokaryotic Argonautes (pAgos), however, exhibit striking diversity in both architecture and activity. Long pAgos harbor the same domains as eukaryotic Agos (eAgos), including the PIWI domain, which mediates guide-directed target cleavage, MID and PAZ domains, which bind the 5′ and 3′ ends of the guide nucleic acid, respectively, and the N-terminal domain involved in guide–target interactions (Lisitskaya et al., 2018). Long-A pAgos containing a catalytic tetrad in the PIWI domain (DEDX, where X = D, H, N or K) are active and mostly use DNA guides to cleave DNA, while long-B pAgos lack the nuclease activity and were reported to bind RNA guides and recognize DNA targets (Ryazansky et al., 2018; Song et al., 2023). Short pAgos, composed only of MID and PIWI domains, rely on APAZ-containing partners to provide structural support and harbor effector domains such as TIR or Sir2 (Koopal et al., 2023). These partnerships illustrate how Argonaute scaffolds can be reprogrammed by auxiliary modules to diversify immune outputs. More recently, catalytically inactive long-B pAgos have also been shown to cooperate with distinct effector proteins to mediate abortive infection, highlighting that accessory partnerships are not unique to short pAgo systems (Song et al., 2023). Yet despite their prevalence, the physiological roles of most pAgos—and especially the functional consequences of unusual domain associations—remain poorly understood.

A recurring theme in the diversification of bacterial immune systems is the repurposing of conserved effector modules, such as the Schlafen (SLFN) domain. The Schlafen protein family is broadly distributed across all domains of life and plays key roles in cell differentiation and antiviral defense in mammals (Jo & Pommier, 2022). First identified in mammals as regulators of thymocyte growth, SLFN proteins are now recognized as nucleases that cleave tRNA or rRNA to restrict viral replication (Schwarz et al., 1998; Yang et al., 2018). Schlafen homologs are also widespread in bacteria, archaea, and viruses, where they occur as stand-alone proteins or as fusions to diverse partners such as NACHT ATPases (Kibby et al., 2023; Taboada et al., 2025). Genomic associations with defense islands suggest potential immune functions, but their contribution to prokaryotic antiviral defense has remained largely untested (Makarova et al., 2011). These findings establish SLFN domains as ancient effectors with known roles in antiviral immunity across eukaryotes.

Here, we investigate the unexplored intersection of Argonautes and Schlafens. Through comparative genomics, structural analysis, and biochemical characterization, we define the evolutionary distribution of Schlafen-associated Argonautes and explore their molecular features using RZAgo from *Runella zeae* as a model.

2. Results

2.1. RzAgo belongs to a clade of extra-long pAgos with a conserved Schlafen domain

Our phylogenetic analysis revealed a previously uncharacterized monophyletic group of long-B pAgos which are substantially longer than previously characterized proteins. This group includes 78 unique proteins ranging from 1071 to 1187 residues, found in both complete genomes and metagenomic sequences (Figure 2.1.a). Similar to other long-B pAgos, these proteins harbor core pAgo domains and lack the catalytic tetrad in the PIWI domain. However, these proteins contain two additional domains positioned at their N-terminus, SLFN (AlbA_2, PFAM 04326) and a GHKL (Gyrase/Hsp90/Histidine Kinase/MutL) superfamily ATPase (PFAM 13749) (Figure 2.1.b). The extra-long pAgos, which we designated SLFN-pAgos, are broadly distributed across prokaryotes, spanning 16 bacterial phyla with notable enrichment in Cyanobacteria and Bacteroidota, and with a single archaeal representative from the Asgard lineage (Candidatus Heimdallarchaeota) (Figure 2.1.c). The phylogenetic tree of SLFN-pAgos does not fully mirror the host species tree and contains branches with mixed taxonomy, suggesting their distribution through horizontal gene transfer.

To investigate the evolutionary relationship between SLFN domains found in pAgos and in other prokaryotic proteins, we reconstructed their phylogenetic tree. This analysis revealed that SLFN domains from SLFN-pAgos separate into two distinct clades. Clade I domains are present exclusively in SLFN-pAgos. Clade II domains are present in SLFN-pAgos and in several members of a diverse NACHT protein family implicated in antiviral defense, indicating that Schlafen domains from this clade can exchange between these two families (Kibby et al., 2023). SLFN domains are also found in association with some short pAgos but they are only distantly related to SLFN-pAgos (Koopal et al., 2023; Ryazansky et al., 2018).

To investigate the origin of the distinct SLFN-pAgo domain architecture, comprising fused SLFN and ATPase domains, we performed extensive sequence search in annotated protein domain databases. This analysis showed that the SLFN-GHKL ATPase combination is extremely rare in eukaryotes but widespread in prokaryotes, including both clades of SLFN-pAgos and multiple additional groups but excluding short pAgos. Among a non-redundant set of 103,533 prokaryotic proteins containing a Schlafen domain, 60,692 (58.6%) also carry the GHKL ATPase domain, suggesting their evolutionary association (Fig. 2.1.d). We designated this stable domain combination (a “supra-domain”) as Schlafatpase (from Schlafen-ATPase) (Vogel et al., 2004). While some Schlafatpase-containing proteins, including SLFN-pAgos, possess additional extended domains, the majority are smaller proteins that contain no other domains or have a short helix–turn–helix (HTH) domain at the C-terminus. In some cases, observed in Poribacteria, the Schlafatpase module is fused to the substrate-recognition subunit of type I RM systems, illustrating that Schlafatpases can exchange their partner domains while retaining the SLFN effector module. Importantly, in SLFN-pAgo, short pAgo, NACHT and RM systems, the SLFN domain is consistently fused to domains responsible for target recognition, suggesting that it may play an effector role in diverse defense systems.

Prokaryotic genes encoding defense functions often cluster together forming genomic “defense islands” (Makarova et al., 2011). Our analysis revealed that SLFN-pAgos found in diverse species are frequently encoded near genes from defense systems, including restriction–modification, toxin–antitoxin and BREX modules as well as phage- and transposon-related integrases, suggesting that SLFN-pAgos are involved in genome defense. However, we observed no stable linkage with specific partners, suggesting that SLFN-pAgos might function autonomously rather than as a part of other defense systems.

2.2. RzAgo binds small RNA guides and recognizes DNA targets

To explore the structural and functional organization of SLFN-pAgos and to investigate their ability to act in defense against phages, we selected the SLFN-pAgo protein from

Runella zeae (RzAgo) for experimental validation. We expressed and purified RzAgo from *E. coli* to test its properties in a series of in vitro assays.

Unlike eukaryotic Argonautes, which exclusively use RNA guides to target complementary RNAs, prokaryotic Argonautes exhibit a broader range of guide and target specificity. To determine the guides and targets of RzAgo, we first performed in vitro binding assays using synthetic 18-nt single-stranded RNA or DNA oligonucleotides with different 5' nucleotides and purified RzAgo. RzAgo bound short RNAs but not ssDNA and showed a strong preference for adenine at the first (5') position, displaying subnanomolar affinity for 5'-A-RNA and ~10-fold weaker binding for RNAs with other 5' nucleotides. Deletion of the SLFN and ATPase domains (residues 1–442) did not significantly affect binding of short RNAs, indicating that these domains are dispensable for guide loading. We next tested the binding of RzAgo–RNA complexes to complementary nucleic acids. RzAgo/gRNA complexes preferentially bound complementary ssDNA, with negligible affinity for RNA targets. As expected, no cleavage of target DNA was detected, confirming that the RzAgo PIWI domain lacks catalytic guide-dependent nuclease activity. These results suggest that RzAgo uses small RNA guides to recognize complementary DNA.

2.3. Loading of guide RNA drives formation of RzAgo filaments

We next sought to understand the structural changes induced in RzAgo through the binding of guide RNA and target DNA. While analyzing guide RNA binding to RzAgo, we observed that the solution became opaque within minutes, indicating potential multimerization of protein-RNA complexes (Figure 2.2.a, top). This corresponded to a large drop in the amount of soluble RzAgo in the samples (Figure 2.2.a, bottom). Subsequent addition of single-stranded DNA complementary to the original RNA guide restored solution clarity, while random ssDNA did not. This suggested that target DNA induces disassembly of oligomeric structures formed by binary complexes of RzAgo.

To visualize high molecular weight gRNA-triggered multimers of RzAgo, we employed negative-stain electron microscopy (ns-EM). While apo-RzAgo appeared on ns-EM images

as small particles with a narrow size distribution (Figure 2.2.b, left), RzAgo-gRNA complexes assembled into regular right-handed helical filaments up to several hundred nanometers in length, indicating that interaction with RNA guides triggered orderly multimerization of RzAgo rather than aggregation (Figure 2.2.b, middle; Figure 2.2.c). Addition of complementary DNA to pre-assembled RzAgo-guide complexes led to disassembly of the filaments into small, evenly dispersed particles (Figure 2.2.b, right).

To understand the molecular mechanism of filament formation, we solved cryo-EM structures of RzAgo incubated with an 18-nt guide RNA, as well as a structure of the RzAgo ago core domains in the absence of guide for comparison. Two gRNA-bound structures were solved: with the addition of a non-hydrolysable ATP analogue, gamma thiol ATP (Adenosine 5'-O-(3-thiotriphosphate)) and mild glutaraldehyde crosslinking, and without either (Figure 2.2.d, e). The non-crosslinked structure was obtained to an overall resolution of 4.03 Å, while the overall resolution of the crosslinked structure was 4.33 Å (Extended data, Figure 1, C and D). In agreement with ns-EM, cryo-EM revealed that RzAgo-gRNA complexes assemble into right-handed helical filaments. A composite reconstruction of two complete turns of the helix in ChimeraX using the obtained structures shows ~12.5 subunits per turn, with a rise of 21.5 Å and a twist of 27.9° per protomer, yielding a pitch of 277 Å (Figure 2.2.f). The backbone of the filament is formed by pAgo core domains (PAZ, MID and PIWI), which are assembled in a head-to-tail arrangement with neighboring protomers linked through a combination of protein-protein and protein-gRNA interactions (Figure 2.2.g). The Argonaute domains are visible in both structures; however, the Schlafase domains are only visible with crosslinking. Five full-length monomers of the RzAgo-gRNA could be docked into the cryo-EM map obtained from our crosslinked sample, which revealed a surprising 3+2 arrangement of Schlafase domain clusters (Figure 2.2.h). Clusters alternate in their position relative to the curve of the Ago core backbone, with some orienting their Slfn domains to the outside and some to the inside of the helix. Extensive interactions are shown to occur between the ATPase domains as well as the Ago cores, while the Slfn domains are not as well resolved, suggesting that

the Slfn domains may retain some conformational plasticity in this oligomeric state (Figure 2.2.e, Extended Data Figure 1D).

The Argonaute domains of RzAgo adopt a characteristic bilobal architecture, with one notable distinction: the absence of an N domain. Consequently, the nucleic acid-binding cleft in RzAgo forms between two asymmetric lobes: a larger lobe comprising the MID and PIWI domains, and a smaller lobe composed of the L1 linker and PAZ domain (Figure 2.3.a). In the absence of guide RNA, the distance between these domains appears to be conformationally heterogeneous, which is most apparent in the lower resolution of the PAZ domain observed in our structure of the apo RzAgo Argonaute domains (Figure 2.3.b, Extended Data Figure 1A). This suggests an intrinsic flexibility of the pAgo core prior to guide loading.

Comparison of the apo and binary states of the RzAgo Ago core reveals that guide binding induces a ~ 10 Å separation between the L1/PAZ and MID/PIWI lobes, repositioning the PAZ domain and widening the inter-lobe cleft to accommodate the guide RNA (Figure 2.3.b). This transition is accompanied by conformational rearrangements in several mobile loops in the MID and PIWI domains (Extended Data Figure 2). Together, these changes provide a basis for extensive interprotomer interactions. These contacts bring together the L1 helix, PAZ domain, PIWI residues, and the 3'-end of the guide RNA belonging to one protomer with the L2 domain, hinge-forming PIWI loops, and a cluster of basic residues on the outer surface of the MID domain of the neighboring protomer. Proteinaceous components of both interfaces show an extensive network of basic residues, which, when adjacent, form a basic tunnel (blue features, Figure 2.3.c). The 3' end of the guide, itself acidic in nature, passes through this tunnel. These opposing Coulombic potentials are critical for stabilizing the interaction between pAgo cores. The 3' portion of the guide RNA thus acts as 'glue' between subunits, interacting simultaneously with two adjacent protomers of RzAgo.

As in other Ago proteins, the 5' end of the guide RNA is anchored within a conserved pocket located in the MID domain of one monomer, with the 5' phosphate group and a

bound Mg^{2+} ion coordinated by K832, Q843, M844, K867 and the terminal carboxyl group of I1111 (Figure 2.3.d). Specific contacts between the first adenine in RNA and residues Y828 and Y1086 in the MID and PIWI domains explain the strong preference of RzAgo for adenosine at the first guide position. In contrast, binding of the 3' portion of the guide RNA in RzAgo differs markedly from that observed in other characterized long pAgo binary complexes, in which the guide 3' end is bound within a PAZ pocket (Lisitskaya et al., 2018; Swarts et al., 2014). RzAgo and other SLFN-pAgos lack a canonical N domain, which directs the guide toward the PAZ domain in other pAgos (Doxzen & Doudna, 2017; Miyoshi et al., 2016). Furthermore, acidic residues in the L1 and PAZ domains of RzAgo (E472, D479, D574, D592, D594) deflect the canonical guide trajectory and participate in interface formation with the adjacent protomer (Figure 2.3.e). These features help to route the guide into the basic tunnel described above. To adopt this unusual pathway, the guide RNA makes a $\sim 90^\circ$ turn around nucleotides 9 and 10 (Figure 2.3.a) which is stabilized by a series of hydrogen bonds mediated by the PIWI and L1 domains. This non-canonical trajectory enables the guide RNA to form direct contacts with a trail of surface-exposed basic residues in the MID domain of the neighboring protomer (including Y759, K760, K722, H711) (Figure 2.3.f, Extended Data Figure 3), contributing $\sim 275 \text{ \AA}^2$ of the $\sim 1480 \text{ \AA}^2$ buried surface area at the protomer–protomer interface. Thus, the guide RNA plays an essential role in stabilizing intersubunit contacts in RzAgo filaments, explaining why filament formation is induced by guide loading. To confirm that loading of guide RNA into the MID pocket of RzAgo is important for its function, we generated a mutant variant with substitutions of two conserved residues (Y828A/K832A) involved in contacts with the guide 5'-phosphate (RzAgo-YK). In vitro binding assays with RzAgo-YK confirmed that it has a dramatically lower affinity for guide RNA (~ 100 -fold increase in Kd). Addition of RNA to RzAgo-YK did not induce turbidity in solution or multimerization detected on a gel, confirming that interactions of RNA with the guide-binding pocket are essential for guide binding and therefore filament formation.

The formation of filaments by guide-loaded RzAgo and their disassembly upon target binding suggest that this process might regulate RzAgo activity in cells, for example by

keeping RZAgo/guide complexes inactive until recognition of target DNA. To explore this possibility, we expressed RZAgo-YK in *F.johnsoniae* under the control of a tightly regulatable promoter. In contrast to wild-type RZAgo, induction of RZAgo-YK expression immediately arrested cell growth (Figure 2.3.g), indicating that failure to load guides renders the protein toxic. Importantly, combining the Y828A/K832A guide-binding mutations with substitutions in the catalytic residues of the Schlafen domain (D32A) eliminated toxicity, demonstrating that its toxicity is mediated by the Schlafen nuclease domain (Figure 2.3.g).

2.4. DNA target recognition reroutes the guide RNA and destabilizes the RZAgo filament

To gain mechanistic insight into how target DNA binding induces RZAgo filament disassembly, we assembled a ternary complex comprising RZAgo, an 18-nt 5'-phosphorylated guide RNA, and a complementary 24-nt DNA. Subsequent single-particle cryo-EM analysis resulted in a reconstruction with a nominal resolution of 3.28 Å (Figure 2.4.a, Extended Data Figure 1E). Like other maps obtained without crosslinking, the map lacked density for the Schlafase supradomain, though some 2D class averages show density suggesting incomplete destruction (Extended Data Figure 4). The RNA–DNA heteroduplex occupies the cleft between the L1/PAZ and MID/PIWI lobes and is positioned through an extensive network of hydrogen bonds formed primarily between basic residues lining this cleft and the sugar–phosphate backbones of each oligonucleotide (Figure 2.4.b). The heteroduplex is fully base-paired beginning at the second position of the guide (U2) (Extended Data Figure 5) and adopts a non-canonical geometry distinct from both A- and B-form helices—a feature previously observed in the RsAgo ternary complex (PDB: 6D8P, Extended Data Figure 6). The first nucleotide of the guide (A1) and its corresponding target base (dT21) are splayed apart and remain unpaired. Despite lacking an N domain, RZAgo positions the distal end of the heteroduplex in essentially the same location as RsAgo, indicating that the N domain is not required for heteroduplex positioning.

In the ternary complex, base-pairing of guide RNA with target DNA changes the RNA trajectory and renders interactions with the preceding protomer impossible (Figure 2.4.c). Instead of bending, the RNA–DNA duplex spans the cleft along its longest path, sterically blocking the filament-forming interface and clashing with interacting residues of the MID domain of the neighboring protomer (Figure 2.4.d). Furthermore, the target DNA strand of the ternary complex occupies the space normally taken by the neighboring PAZ domain, producing a steric clash incompatible with the filament arrangement (Figure 2.4.e). In both ternary and binary complexes, the PAZ domain maintains essentially the same position relative to the other pAgo core domains. Although its direct contribution to guide binding is minimal, this position is crucial for filament architecture, where the PAZ domain forms extensive interactions with the L2 and MID domains of the adjacent protomer.

Base dT21 of the target strand angles away from the heteroduplex axis and is instead bound by a pocket on the surface of the MID domain (Figure 2.4.f). Notably, several residues that mediated interprotomer contacts in the guide-bound filament assume an alternate conformation to contribute to this pocket in the ternary complex. Y759 flips back to interact with C2 of dT21, and W855, which previously π -stacked with F577 of the adjacent protomer, now forms edge-to-face π - π interactions with the nucleobase of dT21 (Figures 2.4.f,g). Further analysis in Coot revealed that these interactions are not strongly nucleotide-specific, as alternative nucleobases can be accommodated within this pocket with minimal steric clashes (Extended Data Figure 7) (Emsley et al., 2010).

Target binding also drives a coordinated rearrangement of the PIWI loop, which spans residues 891–903 and connects β -strands 1 and 2 in the PIWI domain. In the binary RzAgo-guide complex, the PIWI loop interacts with the guide around positions 10–12, but weak and fragmented density indicates substantial mobility, consistent with the dynamic PIWI loops observed in other Agos (Xiao et al., 2023). In the ternary complex, guide–target pairing is accompanied by a marked conformational transition in which the flexible loop refolds into a shortened β -hairpin that interrogates the duplex at positions U10–U12 across the minor groove. In this conformation, the PIWI loop engages both guide and target

strands through backbone- and base-level contacts, suggesting that it acts as a state-dependent structural element that senses duplex formation (Figure 2.4.h).

2.5. ATP binding by the GHKL domain drives RzAgo tetramerization

The histidine kinase-like ATP-binding domain (PFAM 13749) in RzAgo belongs to the broader GHKL superfamily found in diverse proteins, many of which bind and hydrolyze ATP (Dutta & Inouye, 2000). A defining conserved feature of these proteins is that they generally exist and function as dimers or as an interface-forming component of heterotetramers (Corbett & Berger, 2003; Reidy & Masison, 2025). For many GHKL members, the binding of ATP acts as a molecular switch, promoting the dimerization of the N-terminal ATPase domains to form a stable, closed conformation that allows for ATP hydrolysis (Corbett & Berger, 2005; Dutta & Inouye, 2000). Additionally, ATP hydrolysis is not the rate-limiting step for many GHKL member proteins: instead, ADP·Pi release limits the turnover rate (Ban et al., 1999; Corbett & Berger, 2005; Piscitelli & Manhart, 2025).

To test whether RzAgo binds nucleotides, we employed the differential radial capillary action of ligand assay (DRaCALA), which revealed that RzAgo binds ATP with submillimolar affinity (Figure 2.5.a). This binding was dependent on the Schlafpase supradomain, and was lost in its deletion mutant (Figure 2.5.a). Competition assays with excess unlabeled GTP, CTP, and UTP demonstrated high specificity for ATP binding (Figure 2.5.b). Thin-layer chromatography (TLC) assays showed that RzAgo hydrolyzes ATP to ADP and inorganic phosphate (Figure 2.5.c).

To assess the oligomeric state of RzAgo depending on ATP binding and/or hydrolysis, we analyzed RzAgo complexes using denaturing gel electrophoresis after crosslinking with glutaraldehyde (Figure 2.5.d, f). No oligomers of apo-RzAgo were detected in the absence of ATP. In contrast, higher-order assemblies, including dimers, trimers, and tetramers, formed in the presence of ATP or ATP-γ-S (Figure 2.5.d). In the presence of guide RNA, a large portion of RzAgo remained in gel wells, independently of the presence of ATP or

ATP- γ -S, indicating that formation of filaments by binary complexes is independent of ATP binding. Addition of complementary DNA triggered dissociation of the large oligomers but not dimers or tetramers (Figure 2.5.d, ternary lanes). At the same time, deletion of the entire Schlafase supradomain (residues 1–442) (Figure 2.5.e) abolished complex formation, whereas deletion of the SLFN domain alone (residues 1–142) had only a minor effect, consistent with the dominant interface contribution of the ATP-binding domain (Figure 2.5.e).

To gain structural insight into the role of ATP binding in RzAgo oligomerization, we determined the cryo-EM structure of Apo RzAgo in the presence of the non-hydrolyzable ATP analog ATP- γ -S. The resulting structure shows RzAgo protomers assembled into a tetrameric complex (Figure 2.5.f). The SLFN and ATP-binding domains mediate intersubunit interactions (Figure 2.5.g) and are solved to ~ 2.5 Å resolution, whereas the core pAgo domains do not form stable interactions with other components and are poorly resolved, if at all (Figure 2.5.f, Extended Data Figure 1B). The structure of RzAgo tetramers is thus drastically different from the structure of filaments, because the interfaces between pAgo cores cannot form.

The RzAgo tetramer is hierarchically organized as two dimers that further associate into a higher-order assembly with C2 symmetry (Figure 2.5.f, bottom panel schematic). Within each dimer, each protomer adopts a distinct conformation, designated A or B, which differ in the orientation of the SLFN domain relative to the ATP-binding domain (Figure 2.5.g, bottom left schematic). The dimer interface has a buried surface area of $\sim 1,910$ Å² and is formed primarily by the ATP-binding domains ($\sim 67\%$ of the interface), with additional contributions from an α -helix at the base of the SLFN domain (residues 151–161, Extended Data Figure 8). Two dimers associate through a larger interdimer interface ($\sim 2,600$ Å²), which is dominated by polar interactions contributed by both the SLFN ($\sim 60\%$) and ATP-binding domains, resulting in a compact, non-linear tetrameric architecture in which each protomer contacts its two neighbors via distinct domain interfaces.

The ATP-binding domain of RzAgo adopts a fold distinct from the P-loop NTPase domains present in mammalian Schlafens, instead closely resembling the Bergerat-type ATP-binding domain of human pyruvate dehydrogenase kinase 4 (hPDK4; PDB: 2E0A, Figure 2.5.h). It consists of an α - β sandwich with a seven-strand antiparallel β -sheet and three opposing α -helices, with a loop-helix-loop ATP lid traversing the helical face (Figure 2.5.i). The bound magnesium ion and γ -phosphate are coordinated by residues E314 and D318 (Figure 2.5.j). Mutation of the conserved E314 residue abolished ATP hydrolysis, but preserved ATP binding (Figure 2.5.c). Notably, this mutant still formed ATP-dependent oligomers, albeit with reduced efficiency (Figure 2.5.e, last group).

The ATP-binding domain plays a central role in tetramer formation by RzAgo, providing the major contribution to the dimer interface and participating in interdimer contacts. Some interprotomer interactions are mediated by the less conserved N-terminal subdomain (Figure 2.5.k), found in other prokaryotic examples of the GHKL domain family, whereas many contacts mediating oligomerization map to conserved elements of the domain, including the ATP lid (Figure 2.5.l). Interestingly, ATP hydrolysis stimulates profound conformational changes in the ATP lid in other studied GHKL-type ATPases, suggesting that it may directly modulate intersubunit contacts and tetramer formation in SLFN-pAgos (Corbett & Berger, 2005; Dutta & Inouye, 2000).

Together, these results indicate that ATP binding drives RzAgo complex formation, whereas ATP hydrolysis is dispensable for assembly but essential for antiviral defense.

2.6. RzAgo tetramerization recapitulates mammalian SLFN domain organization and reveals a conserved catalytic pocket

The SLFN domain of RzAgo adopts a two-layer α + β topology similar to the C-lobe of mammalian SLFN domains, which harbor a nuclease active site (Figure 2.6.a). The RzAgo Slfn domain contains a conserved E(x)₄(D/E)xK motif (E24, D32, K34), corresponding to catalytic residues described in human SLFN11 (E209, E214, K216) (Figure 2.6.b), and a highly conserved SAFAN motif which forms the hydrophobic core of the domain (Figure

2.6.c) (Metzner et al., 2022). The catalytic pocket in RzAgo closely resembles the corresponding catalytic pocket in hSLFN11 (Figure 2.6.b, left panels). In both cases, a triad of acidic amino acid residues coordinate a metal ion, consistent with a divalent-cation-dependent nuclease mechanism.

In mammalian SLFN proteins, the Schlafen domain has a bipartite architecture, with an N-lobe that provides a basic patch for RNA substrate binding and a C-lobe that contains the nuclease active site (Hou et al., 2023; Metzner et al., 2022). In contrast, SLFN-pAgos, including RzAgo, lack a canonical N-lobe by sequence (Figure 2.6.e). Instead, our cryo-EM structure of the RzAgo apo tetramer reveals that two SLFN domains within an RzAgo dimer assemble into a horseshoe-shaped configuration that recapitulates the bilobal organization observed in mammalian Schlafens (Figure 2.6.f). In the RzAgo SLFN dimer, the catalytic pocket of one protomer faces a basic surface patch (K46, K47) on the opposing protomer, mimicking the relative positioning of functional surfaces within a single mammalian SLFN domain.

Higher-order organization plays a role in mammalian SLFN activity. Human SLFN11 forms dimers to create a substrate binding cleft that accommodates type II tRNA (Figure 2.6.g, right). The tetrameric organization of RzAgo positions two SLFN dimers in a remarkably similar overall configuration, albeit with a wider cleft (Figure 2.6.g, left). Thus, dimerization of RzAgo effectively reconstructs a two-lobed SLFN architecture, with one protomer contributing the catalytic lobe and the other providing a basic surface. Furthermore, the structure of the SLFN tetramer formed by RzAgo bears a strong resemblance the structure of the dimeric mammalian SLFN.

Consistent with this architectural similarity, comparison of SLFN-domain dimer interfaces between RzAgo and mammalian SLFNs reveals conservation of overall interface geometry despite very low sequence identity, supporting the idea that RzAgo and mammalian SLFNs use related structural scaffolds to organize catalytic and RNA-binding surfaces (Extended Data Figure 9).

The organization of the SLFN and ATPase domains in the RzAgo apo tetramer is profoundly different relative to their arrangement in the filaments of RzAgo bound to guide RNA. RzAgo Apo tetramers are C2 symmetric, necessitating a head-to-head orientation of the central pAgo cores (Figure 2.5.f). RzAgo filaments rely on a repeating interface between the pAgo cores, necessitating their head-to-tail arrangement and interrupting that C2 symmetry, thus preventing the formation of RzAgo SLFN tetramers. Notably, the association of pAgo cores induces half the degree of curvature per protomer as ATPase association (Extended Data Figure 10) which prevents the formation of extended tessellating chains and instead induces clustering of Schlappase supradomains. Our crosslinked structure of the binary complex with ATP- γ -S is pentameric, and shows two non-linear Schlappase clusters: one composed of three (protomers 1, 2 and 4), which bends into the curve of the pAgo cores, and the other of two (protomers 3 and 5), which leans out of the pAgo core curve (Figure 2.6.g). In each cluster, the RzAgo architecture within a dimer is retained from the Apo tetramers: RzAgo SLFN still forms dimers, recapitulating a single mammalian SLFN horseshoe. However, the three-protomer cluster introduces a third Schlappase supradomain, positioned such that the SLFN domain blocks one end of the SLFN dimer cleft, potentially restricting substrate access (Extended Data Figure 11A). In addition, the ATPase domains form a new inter-cluster interface that further stabilizes a nucleotide-bound conformation of the ATP-lid (Extended Data Figure 11B, C). Thus, filament formation locks RzAgo into a structural state that is incompatible with the putative active tetrameric SLFN assembly, suggesting that filament formation provides a mechanism for suppressing SLFN nuclease activity.

Together, these structural and genetic analyses demonstrate that dimerization of the RzAgo SLFN domain recapitulates the bilobal architecture of mammalian SLFNs, with a conserved catalytic pocket and a basic surface for RNA binding, and that formation of filaments by binary complexes may regulate RzAgo activity by controlling the assembly of tetrameric complexes.

2.7. RzAgo protects cells from phage infection

Finally, we investigated the ability of RzAgo to act in defense against phages. Because *Runella* is not genetically tractable, we used a related Bacteroidota species *Flavobacterium johnsoniae* as a heterologous host, in which RzAgo was expressed from its native promoter. To minimize interference from endogenous immune systems of *F. johnsoniae*, which might hamper assessment of RzAgo activity, we engineered a strain lacking five defense loci: *pAgo*, *Septu*, *Gabija*, *Tmn*, and *CBASS*. We also sequenced and annotated two *F. johnsoniae* phages, $\phi\Omega 1$ and $\phi\Omega 29$, that had been isolated previously but not characterized at the genomic level (Pate et al., 1979). Genome analysis revealed that $\phi\Omega 1$ has a 352,212 bp genome encoding 600 protein-coding genes and 47 tRNAs, while $\phi\Omega 29$ has a 240,799 bp genome with 338 protein-coding genes and no tRNAs, classifying them as a jumbo and a near-jumbo phage, respectively.

Expression of RzAgo increased the number of $\phi\Omega 29$ particles required to form plaques by more than two orders of magnitude, indicating that RzAgo provided strong protection against phage infection (Figure 2.7.a). RzAgo also provided protection against $\phi\Omega 1$, although it was weaker than in the case of $\phi\Omega 29$.

To understand how RzAgo protects cells against phage infection, we performed high-throughput sequencing of total RNA and DNA from infected cultures at several time points post-infection (Figure 2.7.b). Analysis of phage mRNAs in the absence of RzAgo revealed that both phages expressed their genes in three temporal phases: early, middle, and late. $\phi\Omega 1$ displayed transitions from early to middle and from middle to late expression at ~40 and ~80 minutes post-infection, respectively. $\phi\Omega 29$ exhibited a more rapid early-to-middle transition at ~20 minutes post-infection. Expression of RzAgo led to a large decrease in $\phi\Omega 1$ -derived transcripts at all time points (Figure 2.7.c). Almost all phage transcripts were reduced by ~100-fold, whereas no cellular transcript changed more than ~3-fold in either direction. A similar reduction of phage transcripts was observed for $\phi\Omega 29$, although the effect was less pronounced.

DNA-seq analysis showed that $\phi\Omega 1$ replication—evidenced by a sharp increase in read coverage near the origin—started only after expression of late genes, between 60- and 80-

minutes post-infection (Figure 2.7.d, left panels). Immediately after infection, the $\phi\Omega 1$ genome contains long direct terminal repeats (DTRs) that must be recombined to form a circular DNA molecule before replication can begin. Chromosome ends remained unrecombined until at least one-hour post-infection, i.e. until the beginning of replication (Figure 2.7.d). We did not perform similar analysis for phage $\phi\Omega 29$ since its DNA was difficult to process in vitro, likely due to the presence of nucleotide modifications.

While phage transcripts were dramatically reduced in cell cultures expressing RzAgo, phage DNA was still detectable up to 20 minutes post-infection, though moderately reduced (~20–40% compared to the control, Figure 2.7.d, right panels). However, at 60 minutes post-infection, the time when phage replication normally begins, phage DNA was largely eliminated in cell cultures expressing RzAgo. Analysis of the relative amounts of phage and host DNA reads (RPKM) suggested that under conditions of our experiments only a smaller fraction of cells were infected. Thus, the observed reduction of phage DNA and mRNA and levels in the presence of RzAgo without any effects on cellular mRNA levels may reflect elimination of the phage-containing cells (possibly by abortive infection) rather than specific inhibition of phage replication and transcription in the infected cells (see Discussion).

Next, we analyzed the loading of guide RNAs into RzAgo during $\phi\Omega 29$ infection in *F. johnsoniae*, in cell cultures collected at 120 min post-infection. In agreement with the results obtained in vitro, RzAgo was loaded with guide RNAs starting with 5'-A, with no other sequence preferences. These small RNAs were mostly mapped to the *F. johnsoniae* transcriptome, with only a very small fraction mapped to the phage genome, corresponding to the strong inhibition of phage mRNA expression in the infected cultures. To get additional insight into the in vivo specificity of RzAgo, we also analyzed the spectrum of small RNAs associated with RzAgo in its native species, *R. zeae*. It was found that they universally start with 5'-A and are loaded proportionally to the transcriptome RNA levels, with some preference for noncoding sequences.

To test whether the SLFN domain contributes to phage defense, we introduced a point mutation at residue D32 from the catalytic E(x)₄(D/E)xK motif. The D32A RzAgo mutant showed a dramatic reduction in protection against $\phi\Omega 29$, suggesting that the nuclease activity of the SLFN domain is required for antiphage protection (Figure 2.7.d). Previous studies of human SLFN11 showed that mutation of a residue D252 within the nuclease pocket enhances its RNase activity in vitro. Sequence and structural analysis revealed that the corresponding residue, D72, is conserved in the RzAgo SLFN domain. Strikingly, the RzAgo D72A mutant conferred enhanced protection against phage infection relative to wild-type RzAgo (Figure 2.7.e), suggesting functional parallels between bacterial and mammalian SLFN domains. To assess whether ATP hydrolysis is required for antiviral defense, we compared the hydrolysis-deficient E314A mutant with wild-type RzAgo in plaque formation assays. The mutant displayed strongly diminished protection against phage infection, demonstrating that ATP hydrolysis is critical for RzAgo-mediated defense (Figure 2.7.f).

Together, these data indicate that RzAgo prevents phage propagation and functions as an anti-phage defense system, through its nuclease activity regulated formation of distinct oligomeric structures filament formation dependent on guide RNA and target DNA binding and its ATPase activity.

3. Discussion

Our work establishes Schlafen-containing Argonautes as a previously uncharacterized branch of the Argonaute family that integrates a nuclease-like Schlafen domain and a GHKL ATPase with the canonical pAgo scaffold. Using the Argonaute protein from *Runella zea* as a model, we reveal how these domains cooperate to assemble into different oligomeric states, block phage transcription, and ensure effective defense. Integration of additional functional modules allows SLFN-pAgos to operate as multistate immune systems that switch between ATP-driven tetramers and RNA-induced filaments regulated by the target binding. These findings expand the functional repertoire of Argonautes and illuminate unexpected mechanistic and evolutionary connections between prokaryotic and mammalian Schlafen proteins.

A central insight emerging from our study is that prokaryotic Schlafen domains may not encode a fixed, autonomous nuclease but instead rely on oligomerization to achieve a catalytically competent substrate-binding interface. In RzAgo, the spatial separation between the catalytic pocket and the opposing basic surface is larger than in mammalian Schlafens, producing a more open configuration that is unlikely to support tRNA binding in the manner described for some eukaryotic and prokaryotic family members (Metzner et al., 2022; Yang et al., 2018). The pronounced sequence divergence of bacterial Schlafens further argues against a conserved, rigid substrate preference. Instead, our structural data suggest that dimerization reconstructs a bilobal architecture in which catalytic and substrate-interacting surfaces are assembled across protomers. This raises the possibility that substrate specificity in SLFN-pAgos is dictated not only by sequence but also by oligomer geometry and by constraints imposed by the surrounding Argonaute scaffold.

The repeated recruitment of Schlafen domains into distinct defense systems underscores their versatility as effector modules. Schlafens are found both in long-B pAgos, where they are embedded within a single polypeptide together with a GHKL ATPase, and in short

pAgo systems, where they are encoded as separate SLFN–APAZ proteins that act in trans (Swarts et al., 2014). Phylogenetic analysis indicates that these configurations arose independently, suggesting convergent solutions rather than divergence from a single ancestral system. In SLFN-pAgos and in Schlafen–NACHT proteins implicated in antiviral defense, the Schlafen domain is consistently paired with a C-terminal ATPase that mediates oligomerization, whereas Schlafens associated with short pAgos form a distinct evolutionary group. Together, these observations indicate that Schlafen domains have been repeatedly wired into Argonaute immunity using different architectural strategies, yielding functionally analogous but structurally distinct systems.

Our broader analysis of Schlafase-containing proteins reveals a conserved modular logic that extends beyond Argonautes. The majority of prokaryotic Schlafens are fused to other domains, such as ATPases or small DNA-binding modules, suggesting a general design principle in which Schlafen functions as a latent effector whose activity is gated by an associated sensor. In SLFN-pAgos, this sensor role is fulfilled by the Argonaute core, which recognizes guide and target nucleic acids, while the ATPase provides an additional regulatory layer through nucleotide-dependent oligomerization. This arrangement ensures that potentially destructive nuclease activity is tightly coupled to appropriate molecular cues, minimizing collateral damage to the host.

A striking feature of SLFN-pAgos is their ability to adopt multiple oligomeric states with opposing functional consequences. ATP binding promotes tetramerization through the Schlafen–ATPase supradomain, whereas loading of guide RNA induces formation of extended helical filaments mediated predominantly by the Argonaute core. These filaments are dismantled upon binding of complementary target DNA. Comparison with other pAgo-associated immune systems highlights the diversity of regulatory roles played by oligomerization. In some systems, higher-order assembly activates effector function by juxtaposing catalytic residues from multiple subunits (Kanevskaya et al., 2025; Ni et al., 2023; Zhang et al., 2026). In contrast, our data suggest that in SLFN-pAgos, guide-induced filamentation serves an inhibitory role, spatially segregating Schlafen domains into

configurations incompatible with nuclease activity. Target binding collapses this inhibitory state by rerouting the guide RNA and disrupting multivalent intersubunit contacts, thereby permitting a return to an active configuration. This bidirectional use of oligomerization illustrates that higher-order assembly in immune systems is not intrinsically activating or suppressive, but instead that function is architecturally encoded.

The evolutionary distribution of SLFN-pAgos across bacteria and archaea, their frequent association with defense islands, and their functional parallels with mammalian Schlafens point to Schlafen domains as ancient antiviral effectors that have been repeatedly repurposed throughout evolution. Similar cross-domain conservation has been observed for other immune modules, including gasdermins, viperins, and NACHT which participate in antiviral responses in both prokaryotes and eukaryotes (Bernheim et al., 2021; Johnson et al., 2022; Kibby et al., 2023). SLFN-pAgos extend this emerging theme, illustrating how conserved effector activities can be integrated into diverse sensing and regulatory frameworks to meet distinct biological demands.

Together, our findings suggest that SLFN-pAgos represent a paradigm for immune regulation by structural state switching, in which ATP and RNA act as orthogonal signals that impose opposing assembly states. This raises several open questions. What are the physiological substrates of bacterial Schlafens, and do they differ across architectural contexts? Do Argonaute filaments form *in vivo* during phage infection, and if so, how are they spatially and temporally controlled? More broadly, is reversible filamentation a general strategy for suppressing effector activity in other Argonaute-based systems? By uncovering a regulatory logic centered on higher-order assembly, SLFN-pAgos provide a framework for understanding how immune systems balance potency with self-restraint—an organizing principle that may extend well beyond Argonautes.

4. Methods

4.1. Biochemical, cell biological and bioinformatic methods

Strains and growth conditions

Escherichia coli NEB Turbo was used for routine plasmid propagation and molecular cloning. *E. coli* LOBSTR (DE3) (Kerafast) was used for recombinant protein expression. *E. coli* BW25131 carrying pRK24 plasmid was as a donor strain for conjugation [10.1038/nprot.2014.081]. Plasmids were introduced into *E. coli* strains by chemical transformation using in-house prepared competent cells (Mix & Go *E. coli* Transformation Kit, Zymo Research) or by electroporation. *E. coli* strains were cultivated in LB medium at 37 °C or 30 °C (in case of the donor strain) supplemented with appropriate antibiotics where applicable (ampicillin, 100 µg ml⁻¹; kanamycin, 50 µg ml⁻¹; chloramphenicol, 12.5 µg ml⁻¹; tetracycline, 10 µg ml⁻¹). *Flavobacterium johnsoniae* DSM 6792 (DSMZ – German Collection of Microorganisms and Cell Cultures, Leibniz Institute) was cultivated in liquid or solid CYE medium (10 g l⁻¹ casitone, 5 g l⁻¹ yeast extract, 10 mM Tris-HCl pH 7.3, 8 mM MgSO₄) at 30 °C with appropriate antibiotics where necessary (cefoxitin, 100 µg ml⁻¹; erythromycin, 150 µg ml⁻¹). Electrocompetent cells of *F. johnsoniae* were prepared from exponentially grown cultures by 3x washing with ice-cold 10 % glycerol. Plasmid was added to the cells in 1 cm cuvettes and electroporated at 400 Ω, 25 µF and 10 kV cm⁻¹ (Gene Pulser II, Bio-Rad). *Runella zeae* DSM 19591 (DSMZ – German Collection of Microorganisms and Cell Cultures, Leibniz Institute) was grown on R2A agar plates at 30 °C.

Two bacteriophages infecting *F. johnsoniae* used in this study (φCj1 and φCj29) were a gift from Professor Mark McBride (University of Wisconsin-Madison) and were routinely propagated by the agarose overlay method [10.1007/bf02602855]. Briefly, 200 µl of an overnight culture of *F. johnsoniae* was mixed with 10 µl of phage lysate to achieve an MOI of ~0.1, and phage was allowed to adsorb for 30 minutes at room temperature without shaking. The phage–cell suspension was then mixed with 9 ml of molten 0.35% agarose in

HTC medium (2 g l⁻¹ tryptone, 0.5 g l⁻¹ beef extract, 0.5 g l⁻¹ yeast extract, 0.2 g l⁻¹ sodium acetate, final pH 7.4) and poured over a solid HTC agar plate. After solidifying, plates were incubated overnight at 25 °C to achieve confluent lysis. The top agarose layer was scraped into 5 ml of SM buffer (50 mM Tris-HCl, 100 mM NaCl, 8 mM MgSO₄, pH 7.4) and rocked gently overnight at 4 °C to allow phage particles to diffuse from the gel matrix. The lysate was centrifuged at 3,500g at 4 °C, passed through a 0.45 µm PES syringe filter, and stored at 4 °C. Phage titres were routinely verified by plaque assay, and phages were repurified by passaging through individual plaques after every three rounds of liquid propagation to prevent accumulation of unwanted mutations.

Plasmid construction

The wild-type RzAgo gene was PCR-amplified from *R. zeae* genomic DNA using Q5 2× Master Mix (NEB) or repliQa HiFi ToughMix (Quantabio) and inserted into pET28b for heterologous expression in *E. coli* as an N-terminal His8-SUMO fusion by Gibson assembly (NEB). The construct was designed such that Ulp1p SUMO-protease cleavage yields a wild-type RzAgo protein carrying a single additional N-terminal glycine residue. All amino acid substitutions were introduced using the QuikChange site-directed mutagenesis kit (Agilent). Domain deletions were constructed by amplifying the plasmid with primers annealing to regions flanking the desired deletion, followed by circularization by Gibson assembly.

A replicative plasmid for expression of RzAgo in *F. johnsoniae* (pFj3_RzAgo) was constructed by PCR-amplifying the RzAgo gene together with its native promoter from *R. zeae* genomic DNA and inserting the fragment by Gibson assembly into the pCP29 backbone [10.1128/jb.182.6.1671-1679.2000], which was linearized by PCR to remove the ErmF gene (region between the AflII and BamHI sites). The resulting construct is a shuttle vector containing a high-copy-number pUC origin of replication and AmpR for propagation in *E. coli*, and a *Flavobacterium*-specific origin of replication and CfxA for propagation in *F. johnsoniae*. Mutant variants of RzAgo were subsequently re-cloned into the pFj3 backbone from their pET28b templates by PCR and Gibson assembly.

To introduce gene deletions into *F. johnsoniae*, we constructed a suicide vector (pFj2) by inserting the *SacB* negative-selection marker under the strong *F. johnsoniae* *OmpA* promoter into the pCP11 backbone, linearized by PCR between two *PvuII* sites, by Gibson assembly. For each target gene, ~1.5 kb regions flanking the desired deletion were PCR-amplified and cloned into pFj2 by Gibson assembly. All final constructs were verified by Sanger (Laragen) or nanopore sequencing (Plasmidsaurus).

Generation of F. johnsoniae knock-out strains

To minimize the contribution of native defence systems when assessing RzAgo antiviral activity, we constructed a strain of *F. johnsoniae* carrying deletions of five known defence loci — *pAgo*, *tmn*, *Septu*, *Gabija*, and *gasdermin* — identified using PADLOC (v1.1.0)60 with default parameters. This strain is hereafter referred to as FjΔ5DL. Sequential single deletions were introduced into the wild-type background using the pop-in/pop-out method. In each round, the relevant pFj2-based deletion vector was introduced into the recipient *F. johnsoniae* by conjugation from a donor *E. coli* strain. Briefly, equal volumes of exponentially grown donor and recipient cells (matched by OD600) were washed twice with fresh CYE medium, concentrated by resuspension in a small volume of residual medium after the final centrifugation step, mixed at an approximately 1:1 ratio, and spotted onto CYE agar. Conjugation was allowed to proceed overnight at 30 °C, after which cells were scraped into fresh CYE medium, serially diluted, and plated on CYE agar containing erythromycin to select for transconjugants (first recombination event). Counterselection against the *E. coli* donor was not required, as the *ErmF* gene cannot confer erythromycin resistance in *E. coli*. Individual transconjugants were re-streaked onto fresh selective plates after 2 days to eliminate any residual donor cells. To select for the second recombination event (pop-out), transconjugants were inoculated into liquid CYE medium, grown overnight at 30 °C, serially diluted, and plated on CYE agar supplemented with 5% sucrose for *SacB*-mediated counterselection. Successful deletions were identified by colony PCR and confirmed by Sanger sequencing.

Sequencing, assembly and annotation of bacterial genomes

High-molecular-weight genomic DNA for nanopore sequencing was extracted from exponentially grown bacteria using the Monarch HMW DNA Extraction Kit for Tissue (NEB) following the manufacturer's instructions for gram-negative bacteria. Nanopore long-read libraries were prepared and sequenced at the Millard and Muriel Jacobs Genetics and Genomics Laboratory (Caltech) on a MinION flow cell using R10.4 chemistry.

Hybrid genome assembly was performed using Unicycler (v0.4.8)⁶¹ with default parameters, combining the nanopore long reads with paired-end Illumina reads downloaded from the SRA (accession SRR896071 for *R. zea* DSM 19591; SRR5188344 for *F. johnsoniae* DSM 6792) using the SRA Toolkit. Only reads with a quality score greater than 10 were used for assembly. Genome annotation was performed using the NCBI PGAP pipeline⁶² (Docker image 2022-04-14.build6021 for initial annotation; 2025-05-06.build7983 for final GenBank submission) with default parameters.

Sequencing, assembly and annotation of bacteriophage genomes

Phage genomic DNA was extracted from high-titre lysate ($\geq 10^{10}$ pfu ml⁻¹) as described previously⁶³. Briefly, 450 μ l of lysate was treated with 2 μ l Turbo DNase (Thermo Fisher) and 2 μ l RNase A (DNase- and protease-free, Thermo Fisher) for 1 hour at 37 °C to remove bacterial nucleic acids. The reaction was quenched with EDTA (20 mM final concentration), and capsids were digested with 2.5 μ l Proteinase K (molecular biology grade, NEB) for 1.5 hours at 56 °C. The lysate was mixed 1:1 (v/v) with buffer AL (DNeasy Blood & Tissue Kit, Qiagen) and incubated for 10 minutes at 70 °C, followed by addition of one volume of 95% ethanol. DNA was bound to a DNeasy Mini spin column, washed twice with buffers AW1 and AW2, and eluted in 30 μ l buffer AE.

Phage genomic DNA was sheared by sonication using a Bioruptor Plus (Diagenode) to achieve a final fragment range of 200–600 bp. Sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's instructions and sequenced on an Illumina NextSeq 2000 in 150 nt paired-end mode at the Millard and Muriel Jacobs Genetics and Genomics Laboratory (Caltech).

Paired-end reads were pre-processed using a custom adapter-trimming script to remove reads shorter than 14 nt and to filter reads with perfect alignment to the host genome. Phage genomes were assembled using SPAdes (v3.15.4)⁶⁴ with options `--careful` and `-k 21,33,55,77`. Initial genome annotation was performed with Pharokka (v1.5.1)⁶⁵ using default parameters. To improve annotation of unknown ORFs, structural information was incorporated: AlphaFold3⁶⁶ models of phage proteins were generated using a standalone AlphaFold3 installation with default parameters, and these models were provided as input to Phold (v0.2.0)⁶⁷, which uses Foldseek⁶⁸ to identify structural homologs in the phage protein database.

Protein expression and purification

Wild-type RzAgo and all mutant variants were expressed and purified following the same optimized protocol. The expression plasmid was introduced into *E. coli* LOBSTR (DE3) cells (Kerafast) by electroporation and grown on LB agar overnight at 37 °C. The following day, cells were scraped into 500 ml of autoinduction medium (AIM) and aerated at 37 °C until OD₆₀₀ reached ~1.0, after which cultures were transferred to 18 °C and grown for an additional 24 hours. Cells were harvested by centrifugation at 6,000g at 4 °C, and the pellet was resuspended in buffer A (50 mM HEPES-NaOH pH 7.6, 500 mM NaCl, 20 mM imidazole, 1 mM TCEP, 5% glycerol) at 5 ml per gram of wet cell weight. All subsequent steps were performed at 4 °C. Lysozyme was added to 1 mg ml⁻¹ and the suspension was incubated for 30 minutes with gentle agitation. Complete EDTA-free protease inhibitors (Roche) were added, and cells were lysed by three passages through a high-pressure homogenizer at 30 kpsi (Constant Systems). Insoluble material was removed by centrifugation at 35,000g for 45 minutes.

The cleared lysate was applied to a 5 ml Ni-charged IMAC column (Bio-Rad) connected to a Bio-Rad NGC chromatography system and washed with 10 column volumes of buffer A. Co-purifying nucleotides were removed by washing with 10 column volumes of buffer A supplemented with 1 M NaCl, after which the protein was eluted with buffer A containing 300 mM imidazole. The eluate was transferred into buffer B (50 mM HEPES-NaOH pH

7.6, 300 mM NaCl, 10 mM imidazole, 1 mM TCEP, 5% glycerol) by on-column desalting (Cytiva) and treated overnight with in-house-prepared Ulp1 SUMO-protease (~1:100 molar ratio). Precipitated material was removed by centrifugation at 12,000g for 15 minutes, and untagged RZAgo was isolated by reverse IMAC, passing the supernatant through a 1 ml Ni-charged HP IMAC column (Cytiva) at 0.5 ml min⁻¹; cleaved, untagged RZAgo was collected in the flow-through.

To remove any residual nucleic acids co-purifying with RZAgo, the flow-through fraction was applied to a heparin column equilibrated with buffer C (20 mM HEPES-NaOH pH 7.6, 300 mM NaCl, 1 mM TCEP, 5% glycerol), washed until A280 reached baseline, washed further with 10% buffer D (buffer C supplemented with 1 M NaCl), and eluted in a step of 60% buffer D. Fractions containing RZAgo were pooled, concentrated using an Amicon ultrafiltration unit (50 kDa MWCO), and subjected to size-exclusion chromatography on a Superdex 200 Increase 10/300 GL column (Cytiva) into buffer E (40 mM HEPES-KOH pH 7.4, 250 mM potassium acetate, 8 mM magnesium acetate). Protein purity was assessed by SDS-PAGE, and successful removal of co-purifying nucleic acids was confirmed by an A260/A280 ratio below 0.6. Purified protein was concentrated to 50 μ M, snap-frozen in single-use aliquots in liquid nitrogen, and stored at -80 °C.

Production and purification of anti-RZAgo antibodies

Purified wild-type RZAgo was submitted for custom rabbit polyclonal antibody production (Albabion). The crude antibody-containing fraction was isolated from serum by ammonium sulfate precipitation and dissolved in PBS. Affinity purification was performed by applying the resuspended fraction to a custom column containing RZAgo covalently coupled to NHS-activated Sepharose 4 Fast Flow resin (Cytiva) at 0.1 ml min⁻¹ at 4 °C. Bound antibodies were washed with 10 column volumes of PBS, followed by 5 column volumes of PBS supplemented with 500 mM NaCl, and eluted in 1 ml fractions with 100 mM glycine-HCl pH 3.0 directly into tubes containing 100 μ l of 1 M Tris-HCl pH 8.5 for immediate neutralization. Fractions containing anti-RZAgo antibody were identified by SDS-PAGE, pooled, and buffer-exchanged into PBS supplemented with 10% glycerol by

on-column desalting. The final antibody was concentrated using an Amicon ultrafiltration unit (30 kDa MWCO) to approximately 1 mg ml⁻¹ and stored in single-use aliquots at -20 °C.

In vitro nucleic acid binding assays

Binding affinities of wild-type RZAgo and the ATPase domain deletion mutant (Schlatpase Δ ; residues 1–442 deleted) towards RNA and DNA guides carrying different 5'-nucleotides were measured using a double-filter binding assay. Oligonucleotide substrates (5'-NUUAGACUUUAAGUCAAU-3' for RNA; 5'-NTTAGACTTTAAGTCAAT-3' for DNA) were 5'-radiolabeled by incubating 1 μ M oligonucleotide with 1 μ l of γ -32P-ATP (3,000 Ci mmol⁻¹, Revvity) and polynucleotide kinase (PNK, NEB) in a 20 μ l reaction. Twofold serial dilutions of protein (from 1,024 nM to 1 nM) were mixed with 50 nM radiolabeled guide in binding buffer (10 mM HEPES-NaOH pH 7.0, 100 mM NaCl, 5 mM MgCl₂, 5% glycerol, 100 μ g ml⁻¹ BSA) and incubated for 30 minutes at 30 °C. Samples were filtered through a sandwich of nitrocellulose membrane (Cytiva; pre-activated by incubation in 0.4 M KOH for 10 minutes, washed extensively with ultrapure water, and equilibrated in binding buffer) over Hybond N+ membrane (Cytiva) mounted in a dot-blot apparatus (Bio-Rad). The sandwich was washed three times with binding buffer, air-dried for 15 minutes at 60 °C, exposed to a phosphor-storage screen (Cytiva) for 2 hours, and scanned on a Typhoon fluorescent imager (Cytiva).

Target nucleic acid binding was measured analogously. Binary complexes of RZAgo loaded with a 5'-A-phosphorylated, unlabeled RNA guide were first assembled for 15 minutes at 30 °C. Twofold serial dilutions of the binary complex (from 204.8 nM to 0.2 nM) were then mixed with 100 nM radiolabeled target (5'-UUUAUCAAAAAGAGUAUUGACUUAAGUCUAACCUAUAGGAUACUUACAG-3' for RNA target; 5'-TTTATCAAAAAGAGTATTGACTTAAAGTCTAACCTATAGGATACTTACAG-3' for DNA target) and incubated for an additional 15 minutes before filtration as above. Raw signal was extracted using ImageQuant TL software (Cytiva) and apparent dissociation

constants (K_d) were determined by fitting to a single-site binding model in Prism 8 (GraphPad).

ATP binding and hydrolysis

ATP binding was measured using the Differential Radial Capillary Action of Ligand Assay (DRaCALA)⁶⁹. All binding assays were performed with apo protein in the absence of nucleic acid cofactors. Briefly, twofold serial dilutions of wild-type RzAgo or the ATPase-dead mutant (E314A) were mixed with α -³²P-ATP (3,000 Ci mmol⁻¹, Revvity) in a 20 μ l reaction volume (0.5 $\times$ buffer E, 100 μ g ml⁻¹ BSA) to a final concentration of $\sim$ 3.3 nM α -³²P-ATP, and incubated at 30 °C for 2 minutes. No difference in binding was observed between 2- and 20-minute incubations; the shorter incubation was used to minimize ATP hydrolysis during the assay. A 1.5 μ l aliquot of each reaction was spotted onto a 0.45 μ m nitrocellulose membrane and allowed to absorb and dry. The membrane was exposed to a phosphor-storage screen (Cytiva) for 2 hours and scanned on a Typhoon fluorescent imager (Cytiva). Raw pixel values were extracted using ImageQuant TL software (Cytiva) and binding isotherms were fitted using a custom Python script. For nucleotide competition assays, unlabeled ribonucleoside triphosphates were added to a final concentration of 0.5 mM after 5 minutes of initial incubation with α -³²P-ATP, and the reaction was continued for an additional 30 minutes before spotting.

ATP hydrolysis was assessed by thin-layer chromatography (TLC). All hydrolysis assays were performed with apo protein in the absence of nucleic acid cofactors. Wild-type RzAgo or the E314A mutant (500 nM) was mixed with γ -³²P-ATP (3,000 Ci mmol⁻¹, Revvity) in a 20 μ l reaction volume (0.5 $\times$ buffer E, 100 μ g ml⁻¹ BSA) to a final concentration of 4 nM γ -³²P-ATP, and incubated at 30 °C for 1 hour. Reactions (1.5 μ l) were spotted onto PEI-cellulose TLC plates (Sigma) and developed in 1.5 M KH₂PO₄ (pH 3.8). Plates were dried, exposed to a phosphor-storage screen for 2 hours, and scanned on a Typhoon fluorescent imager.

Analysis of protein complexes by crosslinking SDS-PAGE

Binary complexes of wild-type RzAgo or the indicated mutants (2 μ M) were assembled by incubation with a 5'-phosphorylated 18 nt RNA guide (5'-AUUAAGAGAUGUUGAUGA-3'; 4 μ M) in buffer E supplemented with 500 μ M ATP for 15 minutes at 37 °C. Ternary complexes were subsequently assembled by addition of a complementary 24 nt target DNA (5'-AACTCATCAACATCTCTTAATTTA-3'; 8 μ M) and incubation for a further 20 minutes at 37 °C. Apo-form control samples were prepared identically, substituting RNA and DNA additions with an equivalent volume of nuclease-free water. All samples were normalized with respect to buffer composition before crosslinking.

Glutaraldehyde was added to a final concentration of 0.025% (v/v) — the minimum concentration that preserved quaternary structure without introducing over-crosslinking artefacts, as determined in a series of preliminary experiments — and crosslinking was allowed to proceed for 10 minutes at room temperature. Reactions were quenched with 40 mM Tris-HCl (pH 7.4) for 15 minutes at room temperature. Samples were resolved on 3–8% Tris-acetate gels (Thermo Fisher) and stained overnight with One-Step Blue protein gel stain (Biotium). Gels were imaged on a Li-COR Odyssey infrared imager.

In vitro tRNA cleavage assay

To assess the nuclease activity of RzAgo in vitro, apo (no guide or target), binary (RNA guide-loaded), and ternary (RNA guide- and DNA target-loaded) complexes were assembled and challenged with total E. coli tRNAs (Fermatix). Briefly, 2 μ M RzAgo was incubated with a 1.25-fold molar excess of a 5'-phosphorylated 18 nt RNA guide for 20 minutes at 37 °C, followed by addition of a 1.5-fold molar excess (relative to protein) of either a complementary or a non-complementary 24 nt DNA target and incubation for a further 15 minutes. Apo-form control samples were prepared identically, substituting guide and target additions with equivalent volumes of nuclease-free water. All samples were normalized to identical buffer conditions (20 mM HEPES-KOH pH 7.5, 100 mM potassium acetate, 5 mM magnesium acetate, 5 mM manganese acetate, 5 mM ATP) before initiating the reaction.

Cleavage reactions were initiated by addition of total *E. coli* tRNA to a final concentration of 1 μM . At each indicated time point, an aliquot was withdrawn and quenched by addition of an equal volume of stop solution (8 M urea, 20 mM EDTA). Proteinase K was then added separately (0.5 μl of 20 mg ml^{-1} stock (NEB)) and samples were incubated at 50 °C for 15 minutes to ensure complete protein digestion before gel loading. Samples were resolved on a 15% denaturing TBE–urea polyacrylamide gel at 250 V for 55 minutes. Gels were stained with SYBR Gold, scanned on a Typhoon fluorescent imager (Cytiva), and raw images were processed in Fiji.

Isolation of RzAgo-associated small RNAs from bacteria

To isolate RzAgo-associated small RNAs from *R. zeae*, cells were spread from a frozen seed stock onto 120 × 120 mm R2A agar plates (~100 μl per plate at OD600 ~1) and grown at 30 °C for 24 hours. Cells were scraped into ice-cold immunoprecipitation (IP) buffer (25 mM HEPES-KOH pH 7.05, 150 mM potassium acetate, 5 mM magnesium acetate, 0.5% CHAPS) supplemented with RNaseOUT recombinant RNase inhibitor (Thermo Fisher) and the suspension was adjusted to OD600 ~5. All subsequent steps were performed at 4 °C. Cells were lysed by indirect sonication in a Bioruptor Plus (Diagenode) for 15 cycles (30 s on/30 s off, high amplitude), and insoluble material was removed by centrifugation at 21,000g for 15 minutes. Total protein concentration was measured using the Pierce Detergent-Compatible Bradford Assay (Thermo Fisher) and the lysate was normalized to 1 mg ml^{-1} . RzAgo–nucleic acid complexes were immunoprecipitated by adding 5 μl of anti-RzAgo rabbit polyclonal antibody (~1 mg ml^{-1}) and rocking for 4 hours, followed by 2 hours of incubation with 20 μl of protein A/G magnetic beads (Thermo Fisher). Beads were washed five times with IP buffer (5 minutes each) and resuspended in 300 μl of elution buffer (100 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% SDS). The sample was digested with 1 μl Proteinase K (NEB) for 1 hour at 56 °C. Nucleic acids were extracted using phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0, Sigma) with a phase-lock tube (MaXtract High-Density, Qiagen) and precipitated with ethanol in the presence of

GlycoBlue coprecipitant (Thermo Fisher). The pellet was dissolved in 25 µl ultrapure water.

RzAgo-associated small RNAs were size-selected by denaturing PAGE. For visualization, a 5 µl aliquot was dephosphorylated with rSAP (NEB) and 5'-radiolabeled with γ -³²P-ATP using PNK (NEB). The labeled aliquot was combined with the unlabeled sample, mixed with an equal volume of denaturing sample buffer (8 M urea, 20 mM EDTA, 0.005% bromophenol blue, 0.005% xylene cyanol), and resolved on a 19% polyacrylamide–urea gel under denaturing conditions. Gel slices corresponding to 14–25 nt RNAs were excised, crushed, and RNA was eluted overnight in 0.4 M NaCl at 21 °C with constant agitation. RNA was recovered by ethanol precipitation and resuspended in 20 µl ultrapure water.

Small RNA sequencing libraries incorporating unique molecular identifiers (UMIs) were prepared according to⁷⁰ with the following modifications. The 3'-adapter (5'-NNNgtcNNNtagNNNAGATCGGAAGAGCACACGTCT-3', UMI positions in *italic*) and 5'-adapter (5'-GTTTCAGAGTTCTACAGTCCGACGATCNNNcgaNNNtacNNN-3', UMI positions in *italic*) were redesigned for compatibility with the NEBNext small RNA sequencing workflow. cDNA synthesis was performed using the NEBNext SR RT primer (5'-AGACGTGTGCTCTTCCGATCT-3'). Final PCR indexing was performed using KAPA HiFi HotStart ReadyMix (Roche) with NEBNext small RNA indexing primers; cycle numbers were optimized empirically to avoid overamplification. Libraries were gel-purified on a 6% TBE gel (Thermo Fisher), stained with SYBR Gold, excised, and eluted overnight in 0.4 M NaCl at 21 °C with agitation. After ethanol precipitation, libraries were resuspended in 10 mM Tris-HCl pH 8.5 and quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher). Libraries were sequenced on an Illumina NextSeq 2000 in 100 nt single-end read mode at the Millard and Muriel Jacobs Genetics and Genomics Laboratory (Caltech).

Phage plaque assay

F. johnsoniae FjΔ5DL strains expressing the indicated RzAgo variants from a pFj3 plasmid under its native promoter were grown overnight at 30 °C with shaking at 200 rpm in CYE medium supplemented with 100 µg ml⁻¹ cefoxitin. For each assay, 200 µl of overnight culture was mixed with 9 ml of molten 0.35% agarose in HTC medium supplemented with cefoxitin, poured over an HTC agar plate, and allowed to solidify and dry at room temperature for 30 minutes. Tenfold serial dilutions of phage were spotted on the overlay, allowed to dry, and plates were incubated overnight at 25 °C. Plaque images were acquired using a custom photo box equipped with a Nikon camera against a black background.

Time course analysis of phage infection

F. johnsoniae FjΔ5DL carrying pFj3_RzAgoD72A or the empty vector control were grown in CYE medium to OD₆₀₀ ~1 and infected with φCj1 or φCj29 at an MOI of ~5. Infections were carried out at 25 °C without shaking. After 15 minutes, cultures were washed with fresh CYE medium to remove unabsorbed phage particles (3,200g for 1 minute) followed by resuspension in fresh CYE. The t = 0 sample was collected before addition of the phage. Duplicate 1 ml samples (one for RNA-seq and one for DNA-seq) were removed at the indicated time points, centrifuged at 6,000g for 30 seconds, and the pellets were immediately frozen in liquid nitrogen and stored at -80 °C until processing.

Total RNA was isolated using the RNAsnap method⁷¹. Briefly, pellets were resuspended in 100 µl of RNA extraction solution (18 mM EDTA, 0.025% SDS, 1% 2-mercaptoethanol, 95% formamide) and boiled at 95 °C for 7 minutes. Samples were centrifuged at 16,000g for 15 minutes, and the supernatant was transferred to a new tube. Samples were diluted 1:4 with ultrapure water and RNA was precipitated with ethanol. Crude RNA was purified using an RNA Clean and Concentrator column (Zymo Research) with on-column DNase I treatment according to the manufacturer's instructions. RNA concentration was quantified using the Qubit RNA Broad Range Assay Kit (Invitrogen). Ribosomal RNA was removed using the Pan-Prokaryote riboPOOL rRNA depletion kit as per manufacturer's guidelines. RNA-seq libraries from φCj29-infected samples were prepared using the NEBNext Ultra II

Directional RNA Library Prep Kit (NEB). RNA-seq libraries from ϕ Cj1-infected samples were prepared using the TGIRT-seq workflow as described⁷².

Total DNA was extracted from frozen pellets using the Monarch Spin gDNA Extraction Kit (NEB) following the manufacturer's instructions. DNA-seq libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB). All libraries were sequenced on an Illumina NextSeq 2000 in 100 nt single-end read mode at the Millard and Muriel Jacobs Genetics and Genomics Laboratory (Caltech).

Analysis of high-throughput sequencing data

Analysis of RzAgo-associated small RNAs.

Adapter sequences (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') and reads shorter than 14 nt were removed using cutadapt (v4.2). UMIs were extracted from trimmed reads using UMI-tools (v1.1.4)⁷³ with default parameters, and read quality was assessed using FastQC (v0.11.9). UMI-extracted reads were mapped to the *R. zeae* DSM 19591 genome in two passes using bowtie (v1.3.1): uniquely mapping reads were first identified with parameters -v 0 -m 1, and multimapping reads were subsequently realigned with parameters -a --best --strata -v 0 -m 10000. Only uniquely mapped reads were retained for downstream analysis. Aligned reads were converted to BAM format, sorted, and indexed using samtools (v1.16.1), then deduplicated based on UMI sequences using the UMI-tools dedup module with default parameters. Read length distributions were calculated and plotted using a custom R script. Nucleotide composition logos were generated from deduplicated, uniquely mapped reads trimmed to 14 nt using SeqKit (v2.12.0) and visualized using WebLogo3 (v3.9.0).

Analysis of total RNA-seq during phage infection.

For ϕ Cj29-infected samples, adapter sequences were removed and reads shorter than 16 nt were discarded using cutadapt. Reads were mapped to both the *F. johnsoniae* DSM 6792 and phage genome sequences with up to 2 mismatches allowed using bowtie, retaining only

uniquely mapped reads. For ϕ Cj1-infected samples, adapter trimming and minimum length filtering were performed identically with cutadapt, after which UMIs were extracted using UMI-tools for downstream positional deduplication tracking; reads were then mapped to the host and phage genomes using bowtie2 (`--very-sensitive-local`). For both libraries, gene-level count matrices were generated using featureCounts (v2.0.6) with default parameters and normalized using the TMM method implemented in the RANorm Python package (v2.1.0). Coverage plots and other visualizations were generated using custom Python scripts.

Analysis of total DNA-seq during phage infection.

Adapter sequences and reads shorter than 14 nt were removed using cutadapt, and read quality was assessed using FastQC. Reads were mapped to both the *F. johnsoniae* DSM 6792 and phage genome sequences with up to 1 mismatch allowed using bowtie, retaining only uniquely mapped reads. Aligned reads were sorted and indexed using samtools, and genome-wide coverage was calculated using the bamCoverage function of deepTools (v3.5.1) with RPKM normalization (`--normalizeUsing RPKM`). Coverage tracks were visualized using custom Python script.

Phylogenetic analysis

Phylogeny of prokaryotic Argonaute proteins.

Sequences of prokaryotic Argonaute (pAgo) proteins were searched in the NCBI RefSeq protein database using deltaBLAST with the conserved MID and PIWI domain sequences from five previously characterized pAgos as queries (TtAgo: WP_011229221.1, CbAgo: WP_045143632.1, KmAgo: WP_010289662.1, MpAgo: WP_014295921.1, RsAgo: WP_011910606.1). Identical sequences were removed using MMseqs2 hash-based clustering (mmseqs clusthash), and sequence diversity was further reduced by clustering at 80% identity over 80% coverage using MMseqs2. For multiple sequence alignment (MSA) generation, MID and PIWI domain sequences were extracted from cluster representatives

and were clustered further at 70% sequence identity. The MSA was generated using MAFFT with the L-INS-i strategy and trimmed using trimAl with a gap threshold of 0.3 (-gt 0.3). A maximum-likelihood phylogenetic tree was constructed from this alignment using IQ-TREE (v3.0.1) with 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT replicates (-B 1000 -alrt 1000 -bnni), with the best-fit partitioned substitution model selected automatically using ModelFinder (-m MFP+merge), yielding Q.PFAM+F+R10 for the MID domain partition and LG+F+R10 for the PIWI domain partition. The resulting tree was visualized and annotated in iTOL.

Phylogeny of Slfn-pAgo proteins.

Sequences of Slfn-pAgo fusion proteins were identified by running deltaBLAST against the NCBI non-redundant (nr) protein database. Identical sequences were removed using MMseqs2 hash-based clustering, and a full-length MSA was constructed from the remaining sequences using MAFFT with the L-INS-i strategy. The corresponding species tree was generated using GTDB-Tk. To assess the extent of horizontal gene transfer (HGT) of Slfn-pAgo loci, the protein tree and species tree were compared by co-visualization as a tanglegram using a custom R script.

Structural phylogeny of prokaryotic Alba2-containing proteins.

Prokaryotic Alba2-containing proteins were identified in the NCBI nr database using the Pfam profile PF04326. Sequence redundancy was first reduced using MMseqs2 hash-based clustering, and the remaining sequences were further clustered using Foldseek based on ProST5-derived 3Di structural alphabet representations. Cluster representatives — covering approximately 3,400 structurally diverse proteins from an initial set of ~115,000 hits — were modelled with AlphaFold3 using default parameters, and predicted structures were annotated using DPAM against the ECOD database. Alba2 domains were excised from the AlphaFold3 models and used to generate structure-derived 3Di sequences with Foldseek. These 3Di sequences were aligned using MAFFT with the local-pair iterative refinement strategy (L-INS-i) and the Foldseek 3Di substitution matrix supplied as a user-

defined amino-acid substitution matrix. The resulting 3Di alignment was trimmed using trimAl (-gt 0.3), and a maximum-likelihood structural phylogenetic tree was built using IQ-TREE (v3.0.1) with 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT replicates (-B 1000 -alrt 1000 -bnni). The tree was visualized and annotated in iTOL.

4.2. Negative stain EM of RzAgo Apo, Binary and Ternary complexes

3 mL of 1.5% (w/v) uranyl formate was prepared by dissolving in boiling ddH₂O and filtering to exclude large particles. 100 uL aliquots were prepared in PCR tubes and flash frozen in liquid nitrogen. Grids (Ted Pella no. 01843, Carbon support films, 300 mesh copper grids) were glow discharged for 1 minute at 15 mA.

Samples of RzAgo in the apo, binary complex (18 nucleotide guide RNA) and ternary complex (18 nucleotide guide RNA, 24 nucleotide target ssDNA) states were prepared in buffer containing no excess ATP or ATP analogue. Each sample was prepared at 2 uM RzAgo with excess gRNA and DNA, if used, and allowed to incubate at room temperature for 30 minutes. Samples were then fixed with 0.025% glutaraldehyde for 15 minutes at room temperature and quenched with 40 mM Tris for 15 minutes at room temperature. Samples were then diluted to a final concentration of 200 nM RzAgo monomer.

3 uL sample was added to the carbon face of the grid, followed by a 1 minute incubation. Excess was removed by blotting with filter paper (Wattman), touching paper to the edge of the grid at a 90° angle to the grid surface. Then, 3 uL uranyl formate solution was added, again with a 1-minute incubation. Grid was blotted to remove excess as above, then uranyl formate application and blotting were repeated. Grids were allowed to dry prior to microscopy, and stored at room temperature.

Samples were screened and then data were collected using the Talos Arctica (FEI) at 200 kV fitted with a Gatan 3 detector. 1,800 micrographs were collected for the Apo data set, 3,663 for the Binary complex data set, and 1,800 for the Ternary complex data set. All data were collected at 45,000x, with a total exposure dose of 65 e/Å² and a pixel size of 0.869

Å. Data were imported into cryoSPARC and patch motion corrected before particles were selected either via Blob Picker or manually to perform 2D class averaging for data interpretation.

4.2. Cryo-EM structure methods

pAgo core Apo monomer

Purified RzAgo was prepared at 12.5 μ M in buffer containing [no ATP/AGS, Mg]. Quantifoil R 1.2/1.3 Au 400 grids were glow discharged for 30 seconds at 15 mA (Pelco Easyglow). 4 μ L of sample was added to each grid in a Vitrobot Mark 4 (Thermofisher) at 25 °C and 100% humidity and blotted using Wattman filter paper with a blot time of 3 s and blot force of 0, with no wait time. Samples were plunge frozen in liquid ethane.

Data were collected on a Titan Krios (Thermofisher) with a Gatan K3 detector (Gatan) with an accelerating voltage of 300 kV, a magnification of 130,000x, and a total exposure dose of 60 e/Å². 3,256 movies were collected with 40 frames per movie.

All data processing was performed in CryoSPARC. Micrographs were motion corrected using the Patch Motion correction job and the CTF was corrected with CTFFIND4. 3,108 exposures were accepted. Particles were picked with Blob Picker, inspected, extracted, and assembled into 2-dimensional class averages iteratively until an appropriate box size was established. Ab initio 3D classification (2 classes) and heterogeneous refinement (2 classes) were performed and the particle stack from the best class was used as training data for template picking using Topaz. Five iterative rounds of Topaz particle picking, particle extraction, 2D and 3D classification, particle class rebalancing, and refinement were necessary. The final round of Topaz training yielded 578,612 unique particles, extracted with a box size of 320 pixels downsampled to 160 pixels. Particles were further sorted with heterogeneous refinement using 5 query volumes previously obtained, then non-uniform refinement of the best class and particle reextraction of the 219,569 selected particles to a

box size of 320 pixels with no downsampling. A final non-uniform refinement yielded a map at 3.58 Å resolution.

Model building was performed using an AlphaFold3 model of RzAgo. Model fitting was initially performed in ChimeraX (v. 1.9) and ISOLDE, then was iteratively refined using real space refinement in Phenix (v. 1.21.2-5419-000). The model was then modified and further refined as necessary in Coot (v. 9.8.96) and ISOLDE to reduce Ramachandran and rotamer outliers. Final model refinement statistics are reported in Table 1.

Apo tetramer

Purified RzAgo was prepared at 2 uM in buffer containing [STUFF –AGS, Mg but not Mn]. The sample was allowed to equilibrate for XXX minutes, then crosslinked with 0.025% glutaraldehyde for 15 minutes at room temperature and quenched with 40 mM Tris for 15 minutes at room temperature. Quantifoil R 1.2/1.3 Cu 300 grids were glow discharged for 30 seconds at 15 mA (Pelco Easyglow). 4 uL of sample was added to each grid in a Vitrobot Mark 4 (Thermofisher) at 25 °C and 100% humidity and blotted using Wattman filter paper with a blot time of 4.5 s and blot force of 0, with a wait time of one minute. Samples were plunge frozen in liquid ethane.

Data were collected on a Titan Krios (Thermofisher) with a Gatan K3 detector (Gatan) with an accelerating voltage of 300 kV, a magnification of 130,000x, and a total exposure dose of 40 e/Å². 6,451 movies were collected with 78 frames per movie.

All data processing was performed in CryoSPARC. Micrographs were motion corrected using the Patch Motion correction job and the CTF was corrected with CTFFIND4. 5,374 exposures were accepted. 2,909,653 were picked with Blob Picker, inspected, and extracted with a box size of 600 pixels, resampled to 300 pixels. Particles were classified into 300 2D class averages, of which 28 were selected (322,600 particles). These were classified into 3 ab initio maps with C1 symmetry, and particles from the best class were passed through a round of homogeneous refinement (109,024). This map was further refined with a

second round of 2D classification and particle selection and the final map was generated via non-uniform refinement. The final map has a reported resolution of 3.4 Å (GSFSC) and contained 64,577 particles.

Model building was performed using an AlphaFold model of RzAgo. Model fitting was initially performed in ChimeraX (v. 1.9). Iterative rounds of model refinement were then performed using real space refinement in Phenix (v. 1.21.2-5419-000) and ISOLDE. Ligand restraints were generated from a SMILES string in Phenix using eLBOW. Ligand fitting was performed manually in ChimeraX using ISOLDE and then refined in Phenix. Iterative refinement was then performed in Phenix, Coot (v. 0.9.8.96), and ISOLDE. Final model refinement statistics are reported in Table 2.

Binary complex pAgo core dimer

Purified RzAgo was prepared at 2 uM in buffer containing [STUFF – no ATP/AGS, Mg but not Mn], and Binary complex was assembled using a 1.5-fold excess of gRNA. Quantifoil R 1.2/1.3 Au 400 grids were glow discharged for 30 seconds at 15 mA (Pelco Easyglow). 4 uL of sample was added to each grid in a Vitrobot Mark 4 (Thermofisher) at 25 °C and 100% humidity and blotted using Wattman filter paper with a blot time of 3 s and blot force of 0, with 15 s wait time. Samples were plunge frozen in liquid ethane.

Data were collected on a Titan Krios (Thermofisher) with a Gatan K3 detector (Gatan) with an accelerating voltage of 300 kV, a magnification of 105,000x (pixel size of 0.42 Å²), and a total exposure dose of 60 e/Å². Data were acquired at a 30 ° tilt (2,908 movies).

All data processing was performed in CryoSPARC. Micrographs were motion corrected using the Patch Motion correction job and the CTF was corrected with CTFFIND4, and 1,873 exposures were accepted. Using a subset of 324 micrographs, 138,848 particles were picked using the Blob picker with a particle diameter range of 100-300 Å, inspected, and extracted with a box size of 800 pixels, downsampled to 200. These were passed through two rounds of 2D classification and class rebalancing, yielding 16,653 accepted particles.

These particles were used as a training data set for Topaz particle picking. Three rounds of Topaz training, opening to all accepted micrographs in the second round, followed by 2D classification, 3D ab initio map generation and heterogeneous refinement, were used to curate particles. A final stack of 85,382 particles extracted with a box size of 800 pixels was used to obtain a final structure at 4.03 Å via non-uniform refinement.

Model building was performed using an AlphaFold3 model of RzAgo with 18 nucleotide guide RNA. Model fitting was initially performed in ChimeraX (1.9). Iterative rounds of model refinement were then performed using real space refinement in Phenix (v. 1.21.2-5419-000) and ISOLDE. Final model refinement statistics are reported in Table 3.

Binary complex pentamer

Purified RzAgo was prepared at 4 µM in buffer containing [AGS, Mg, other components], and guide RNA was added. The sample was allowed to equilibrate for XXX minutes, then crosslinked with 0.025% glutaraldehyde for 15 minutes at room temperature and quenched with 40 mM Tris for 15 minutes at room temperature. Quantifoil R 1.2/1.3 Cu 300 grids were glow discharged for 30 seconds at 15 mA (Pelco Easyglow). 4 µL of sample was added to each grid in a Vitrobot Mark 4 (Thermofisher) at 25 °C and 100% humidity and blotted using Wattman filter paper with a blot time of 4.5 s and blot force of 0, with a wait time of one minute. Samples were plunge frozen in liquid ethane.

Data were collected on a Titan Krios (Thermofisher) with a Gatan K3 detector (Gatan) with an accelerating voltage of 300 kV, a magnification of 130,000x (pixel size of 0.32 Å²), and a total exposure dose of 45 e/Å². 6,489 movies were collected.

All data processing was performed in CryoSPARC. Micrographs were motion corrected using the Patch Motion correction job and the CTF was corrected with CTFFIND4. 6,165 micrographs were accepted via manual exposure curation. 2.9 million particles were initially picked using Blob picker with an expected particle diameter of 70-600 Å. Two rounds of 2D classification selected 232,476 particles, which were used for ab initio 3D

map generation (3 maps requested). Each ab initio map was then refined using heterogeneous refinement, further selecting a promising particle population. Particle selection was further refined through three rounds of Topaz model training, which entailed training, extraction, one or more rounds of 2D classification, heterogeneous refinement, and then non-uniform refinement of the best class. A final particle stack containing 111,503 particles was extracted with a box size of 1024 pixels. Non-uniform refinement of 111,503 particles resulted in a final structure at 4.33 Å resolution.

Model building was performed using a model generated with AlphaFold3, combined with the previously obtained Apo tetramer Slfn/ATPase domains and Binary pAgo dimer. Initial fit was performed manually in ChimeraX (1.9). The fit was then refined first with ISOLDE and then iteratively via rounds of real space refinement (Phenix v. 1.21.2-5419-000) followed by manual correction with ISOLDE. Ligand restraints were recycled from the Apo tetramer refinement process. Final model refinement statistics are reported in Table 4.

pAgo core Ternary complex

Purified RzAgo was prepared at 2 uM in buffer containing 1mM ATP [other buffer contents]. 18 nucleotide guide RNA and 24 nucleotide target DNA were added to final concentrations of 1.5 uM and 1 uM, respectively. Quantifoil R 1.2/1.3 Cu 300 grids were glow discharged for 30 seconds at 15 mA (Pelco Easyglow). 3.5 uL of sample was added to each grid in a Vitrobot Mark 4 (Thermofisher) at 25°C and 100% humidity and blotted using Wattman filter paper with a blot time of 3 s and blot force of 0, with a wait time of 15 seconds. Samples were plunge frozen in liquid ethane.

Data were collected on a Titan Krios (Thermofisher) with a Gatan K3 detector (Gatan) with an accelerating voltage of 300 kV, a magnification of 130,000x (pixel size of 0.338 Å²), and a total exposure dose of 35 e/Å². 12,916 movies were collected.

All data processing was performed in CryoSPARC. Micrographs were motion corrected using the Patch Motion correction job and the CTF was corrected with CTFFIND4. 11,692

micrographs were accepted via manual exposure curation. 3.2 million particles were initially picked from a subset of 2,001 micrographs using Blob picker with an expected particle diameter of 50-200 Å, and 801,701 particles were extracted with a box size of 512 pixels (downsampled to 128). 2D classification selected 77,831 particles, which were used for ab initio 3D map generation (3 maps requested). Each ab initio map was then refined using heterogeneous refinement and the best refined by homogeneous refinement (31,884 particles). Further 2D classification from this refinement selected 28,114 particles, which were rebalanced and used for Topaz training on the 2,001-micrograph data set, which entailed training, extraction, one or more rounds of 2D classification, heterogeneous refinement, and then non-uniform refinement of the best class. Three rounds of Topaz training were performed, expanding the number of micrographs to the total data set on the second round. A final particle stack containing 122,907 particles was extracted with a box size of 512 pixels. After 2D classification, and heterogeneous refinement, non-uniform refinement of 76,858 particles resulted in a final structure at 3.28 Å resolution.

Model building was performed using a model generated with AlphaFold3. Initial fit was performed manually in ChimeraX (1.9). The fit was then refined first with ISOLDE and then iteratively via rounds of real space refinement (Phenix v. 1.21.2-5419-000) followed by manual correction with ISOLDE. Final model refinement statistics are reported in Table 5.

Tables

Table 1. Apo monomer

<u>Model</u>	
Composition (#)	
Chains	1
Atoms	10258 (Hydrogens: 5137)
Residues	Protein: 633 Nucleotide: 0
Water	0
Ligands	0
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.016 (19)
Angles (°) (# > 4σ)	1.594 (22)
MolProbity score	2.66
Clash score	52.17
Ramachandran plot (%)	
Outliers	0
Allowed	7.5
Favored	92.5
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 627)	-2.01 (0.29)
helix (N = 201)	-1.41 (0.29)
sheet (N = 139)	-0.13 (0.41)
loop (N = 287)	-1.83 (0.33)
Rotamer outliers (%)	0
Cβ outliers (%)	0.17
Peptide plane (%)	
Cis proline/general	3.8/0.0
Twisted proline/general	0.0/0.3
CaBLAM outliers (%)	4.51
ADP (B-factors)	
Iso/Aniso (#)	5121/0
min/max/mean	
Protein	149.41/265.51/208.52
Nucleotide	---
Ligand	---
Water	---

Occupancy

Mean	1
occ = 1 (%)	100
0 < occ < 1 (%)	0
occ > 1 (%)	0

Data**Box**

Lengths (Å)	72.80, 80.60, 96.85	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	3.6	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	

d 99 (full/half1/half2)	4.4/---/---	4.4/---/---
d model	3.9	3.8
d FSC model (0/0.143/0.5)	3.4/3.6/6.6	3.4/3.7/6.8
Map min/max/mean	#DIV/0!	

Model vs. Data

CC (mask)	0.68
CC (box)	0.84
CC (peaks)	0.62
CC (volume)	0.67
Mean CC for ligands	---

Table 2. Apo tetramer

<u>Model</u>	
Composition (#)	
Chains	13
Atoms	50387 (Hydrogens: 24931)
Residues	Protein: 3096 Nucleotide: 0
Water	0
Ligands	MG: 8
	AGS: 4
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.002 (0)
Angles (°) (# > 4σ)	0.459 (6)
MolProbity score	1.41
Clash score	4.8
Ramachandran plot (%)	
Outliers	0.06
Allowed	2.82
Favored	97.12
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 3088)	0.67 (0.15)
helix (N = 932)	1.30 (0.18)
sheet (N = 597)	0.18 (0.22)
loop (N = 1559)	0.15 (0.16)
Rotamer outliers (%)	0
Cβ outliers (%)	NA
Peptide plane (%)	
Cis proline/general	7.3/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	1.56
ADP (B-factors)	
Iso/Aniso (#)	25456/0
min/max/mean	
Protein	0.00/202.45/99.41
Nucleotide	---
Ligand	1.00/118.80/21.19
Water	---
Occupancy	
Mean	1

occ = 1 (%)	100
0 < occ < 1 (%)	0
occ > 1 (%)	0

Data

Box

Lengths (Å)	100.10, 159.90, 206.70	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	3.4	
Resolution Estimates (Å)	Masked Unmasked	
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	3.6/---/--- --/---	3.5/-
d model	3.5	3.6
d FSC model (0/0.143/0.5)	3.2/3.3/3.6	3.3/3.3/3.8
Map min/max/mean	-58.959538	

Model vs. Data

CC (mask)	0.73
CC (box)	0.7
CC (peaks)	0.62
CC (volume)	0.72
Mean CC for ligands	0.8

Table 3. Binary dimer

<u>Model</u>	
Composition (#)	
Chains	6
Atoms	22556 (Hydrogens: 11094)
Residues	Protein: 1322 Nucleotide: 36
Water	0
Ligands	MG: 2
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.002 (0)
Angles (°) (# > 4σ)	0.527 (8)
MolProbity score	1.94
Clash score	6.61
Ramachandran plot (%)	
Outliers	0.08
Allowed	4.11
Favored	95.81
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 1314)	0.10 (0.25)
helix (N = 438)	1.91 (0.25)
sheet (N = 316)	-1.76 (0.28)
loop (N = 560)	-0.15 (0.29)
Rotamer outliers (%)	2.39
Cβ outliers (%)	NA
Peptide plane (%)	
Cis proline/general	3.8/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	2.22
ADP (B-factors)	
Iso/Aniso (#)	11462/0
min/max/mean	
Protein	188.52/350.10/247.33
Nucleotide	51.84/628.17/400.86
Ligand	96.18/98.74/97.46
Water	---
Occupancy	
Mean	1
occ = 1 (%)	100

0 < occ < 1 (%)	0	
occ > 1 (%)	0	
<u>Data</u>		
Box		
Lengths (Å)	92.35, 134.78, 99.84	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	4	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	5.3/---/---	5.1/---/---
d model	4.2	4.2
d FSC model (0/0.143/0.5)	3.9/4.1/4.9	3.9/4.2/6.5
Map min/max/mean	-27.777778	
<u>Model vs. Data</u>		
CC (mask)	0.7	
CC (box)	0.87	
CC (peaks)	0.58	
CC (volume)	0.68	
Mean CC for ligands	0.73	

Table 4. Binary pentamer

<u>Model</u>	
Composition (#)	
Chains	28
Atoms	92970 (Hydrogens: 45680)
Residues	Protein: 5545 Nucleotide: 90
Water	0
Ligands	MG: 15
	AGS: 5
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.007 (6)
Angles (°) (# > 4σ)	1.177 (56)
MolProbity score	1.97
Clash score	10.74
Ramachandran plot (%)	
Outliers	0.02
Allowed	6.34
Favored	93.64
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 5535)	-1.63 (0.10)
helix (N = 1843)	-0.85 (0.11)
sheet (N = 998)	-0.57 (0.15)
loop (N = 2694)	-1.38 (0.11)
Rotamer outliers (%)	0.02
Cβ outliers (%)	0.02
Peptide plane (%)	
Cis proline/general	6.3/0.0
Twisted proline/general	0.4/0.3
CaBLAM outliers (%)	3.19
ADP (B-factors)	
Iso/Aniso (#)	47290/0
min/max/mean	
Protein	105.99/1006.60/231.18
Nucleotide	56.15/306.98/170.22
Ligand	133.95/999.99/182.83
Water	---
Occupancy	

Mean	1
occ = 1 (%)	100
0 < occ < 1 (%)	0
occ > 1 (%)	0

Data

Box

Lengths (Å)	150.80, 184.60, 265.20	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	4	
Resolution Estimates (Å)	Masked Unmasked	
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	4.6/---/---	4.6/---/---
d model	2.0	2.0
d FSC model (0/0.143/0.5)	2.0/2.6/4.3	2.0/2.6/4.3
Map min/max/mean	-0.0235023	

Model vs. Data

CC (mask)	0.8
CC (box)	0.92
CC (peaks)	0.79
CC (volume)	0.79
Mean CC for ligands	0.71

Table 5. Ternary complex

<u>Model</u>	
Composition (#)	
Chains	4
Atoms	11845 (Hydrogens: 5753)
Residues	Protein: 661 Nucleotide: 36
Water	0
Ligands	MG: 1
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.013 (0)
Angles (°) (# > 4σ)	1.406 (17)
MolProbity score	1.16
Clash score	0.86
Ramachandran plot (%)	
Outliers	0.46
Allowed	5.33
Favored	94.22
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 657)	-1.57 (0.29)
helix (N = 164)	-1.51 (0.33)
sheet (N = 112)	-0.27 (0.47)
loop (N = 381)	-1.03 (0.30)
Rotamer outliers (%)	0
Cβ outliers (%)	NA
Peptide plane (%)	
Cis proline/general	3.8/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	2.14
ADP (B-factors)	
Iso/Aniso (#)	6092/0
min/max/mean	
Protein	64.29/173.29/100.29
Nucleotide	45.61/152.84/74.17
Ligand	113.34/113.34/113.34
Water	---
Occupancy	
Mean	1
occ = 1 (%)	100

0 < occ < 1 (%)	0	
occ > 1 (%)	0	
<u>Data</u>		
Box		
Lengths (Å)	77.00, 82.00, 100.00	
Angles (°)	90.00, 90.00, 90.00	
Supplied Resolution (Å)	3.3	
Resolution Estimates (Å)	Masked	Unmasked
d FSC (half maps; 0.143)	---	---
d 99 (full/half1/half2)	3.4/---/---	3.4/---/---
d model	1.4	1.4
d FSC model (0/0.143/0.5)	1.4/2.0/3.4	1.4/2.0/3.4
Map min/max/mean	-0.0077552	
<u>Model vs. Data</u>		
CC (mask)	0.76	
CC (box)	0.87	
CC (peaks)	0.74	
CC (volume)	0.76	
Mean CC for ligands	0.75	

Figures

Figure 1.

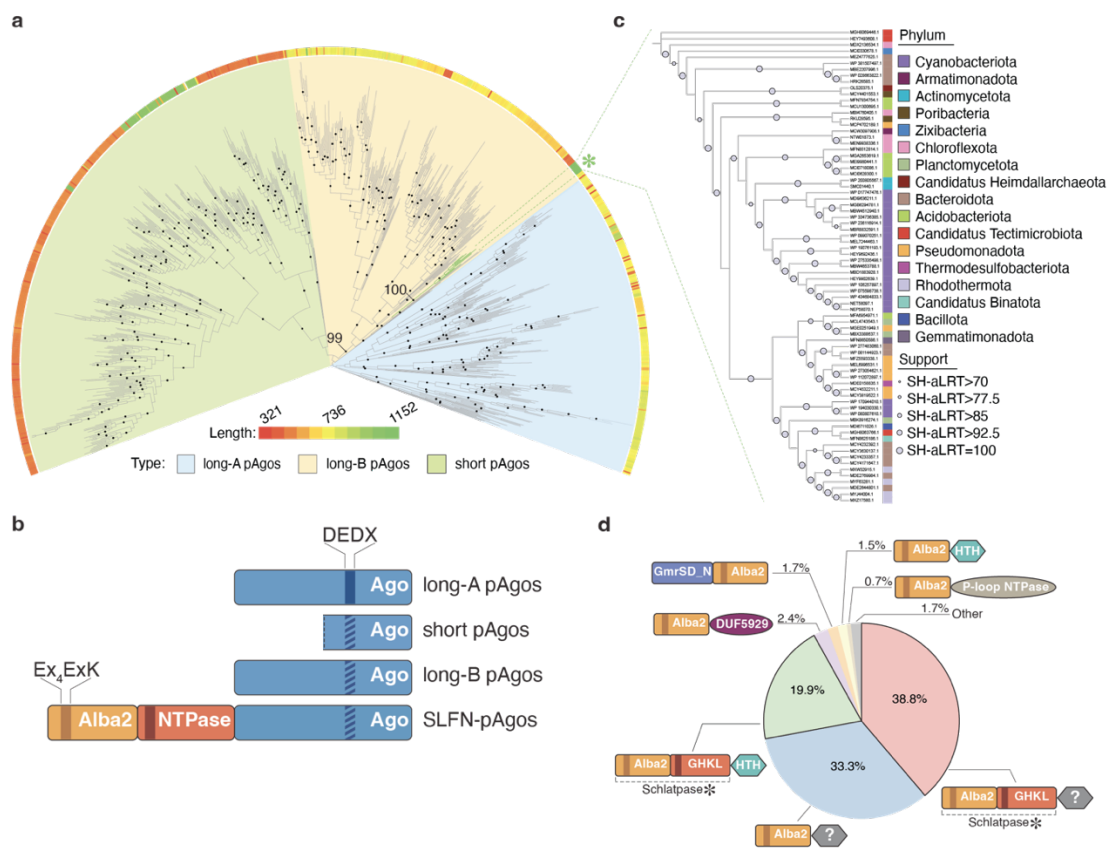


Figure 1. RzAgo phylogeny and domain architecture. **a**, Sequence-based phylogenetic tree of prokaryotic Argonautes (pAgos) inferred from a multiple sequence alignment of the MID and PIWI domains. The outer ring indicates protein length. Branches supported with both UFBoot values >95% and SH-aLRT values >80% are marked with black dots. The SLFN-pAgo clade within long-B pAgos is highlighted by a green asterisk. **b**, Domain organization of SLFN-pAgos compared with the three major pAgo groups (long-A, long-B and short). The catalytic tetrad of the PIWI domain and putative active sites within the Alba2 (Schlafen) and NTPase domains are indicated. Hatched boxes denote inactive or degenerate catalytic sites. **c**, Phylogenetic tree of SLFN-pAgos constructed from a full-length sequence alignment. Branches are colored according to prokaryotic phylum. SH-aLRT branch support values are shown where applicable. **d**, Diversity of domain architectures among prokaryotic Schlafen-containing proteins, highlighting recurrent fusions with GHKL NTPase and additional accessory domains.

Figure 2.

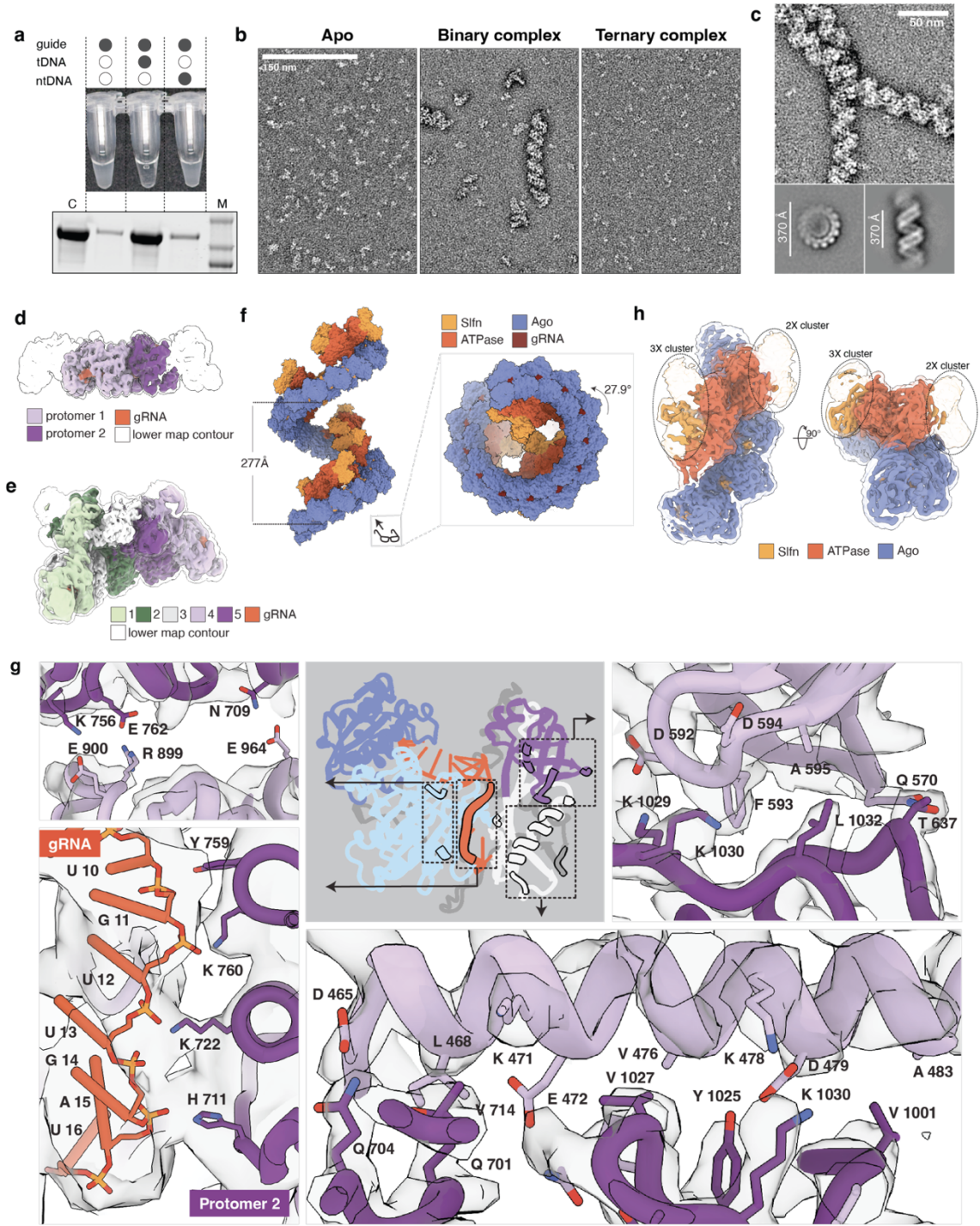


Figure 2. RzAgo filament formation. **a**, Top, visible precipitation of RzAgo induced by guide RNA binding is reversed upon addition of a fully complementary DNA target. Bottom, SDS–PAGE analysis of soluble RzAgo remaining in the supernatant after incubation with guide RNA alone, with a fully complementary DNA target, or with a non-complementary DNA control. **b**, Representative micrographs apo (left), binary (middle) and ternary (right) RzAgo, crosslinked with glutaraldehyde but no additional ATP, by negative stain. **c**, Top, representative negative-stain electron micrograph showing RzAgo filament formation in the presence of guide RNA. Bottom, representative two-dimensional class averages showing top and side views of the helical assembly. **d**, Cryo-EM structure of the RzAgo filament assembled without crosslinking, showing two consecutive protomers. Low-contour transparent density highlights adjacent protomers. Guide RNA is shown in orange. **e**, Cryo-EM structure of the RzAgo helical filament obtained after mild glutaraldehyde crosslinking in the presence of ATP- γ -S, showing five consecutive protomers. Protein domains and guide RNA are colored; low-contour transparent density highlights the positions of Schlafen domains. **f**, Composite atomic model of the RzAgo helical filament built from the cryo-EM reconstruction in **e**, with helical parameters indicated. **g**, Examples of multivalent protein–protein and RNA–protein interactions stabilizing intersubunit contacts between pAgo cores within the filament. **h**, Organization of Schlafen domains into two spatially separated clusters within the helical filament. Domains are colored as in **f**.

Figure 3.

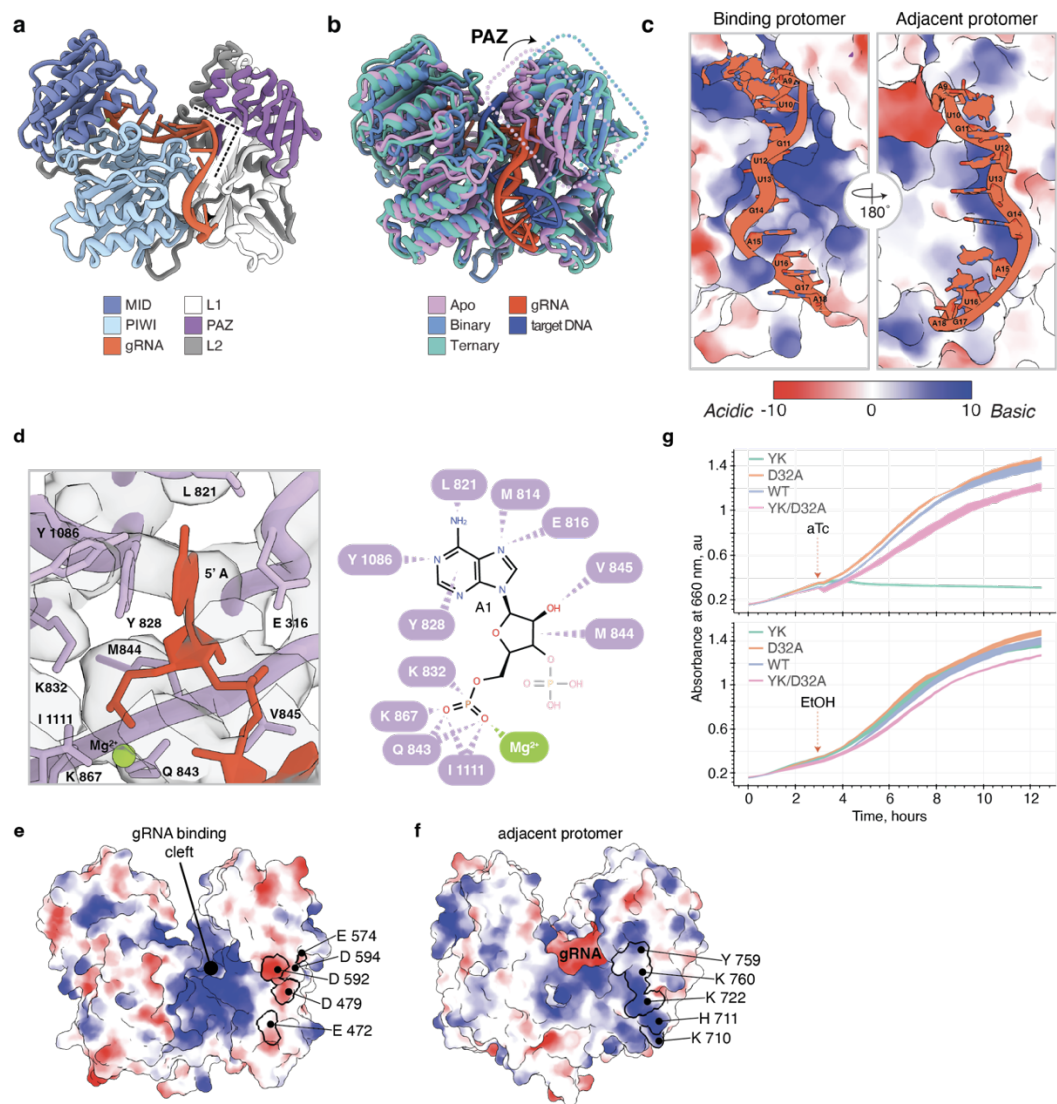


Figure 3. Guide binding and pAgo core interactions with adjacent protomers. a, Structure of an individual RzAgo monomer highlighting the absence of a canonical N-terminal domain and a pronounced kink in the bound guide RNA (dashed arrow). **b,** Overlaid ribbon structures of the apo (lilac), binary (blue) and ternary (teal) structures to show the comparative movement of the PAZ domain. **c,** Coulombic potential of the interface between pAgo cores within the filament, centered on the 3' end of the guide. **d,** Left: Binding pocket of the 5' gRNA adenine in the structure, showing the electron density in grey. Contacts are labeled in black. Right: Flattened schematic showing contacts to the 5' gRNA adenine. **e,** Coulombic potential map of one half of the surface of the pAgo core filament interface highlighting the acidic patch where the N domain would otherwise be. **f,** Coulombic potential map of one half of the surface of the pAgo core filament interface highlighting the trail of basic residues contacting the guide and forming part of the guide-binding basic tunnel. **g,** Growth curves of *F. johnsoniae* strains expressing the indicated RzAgo variants

Figure 4.

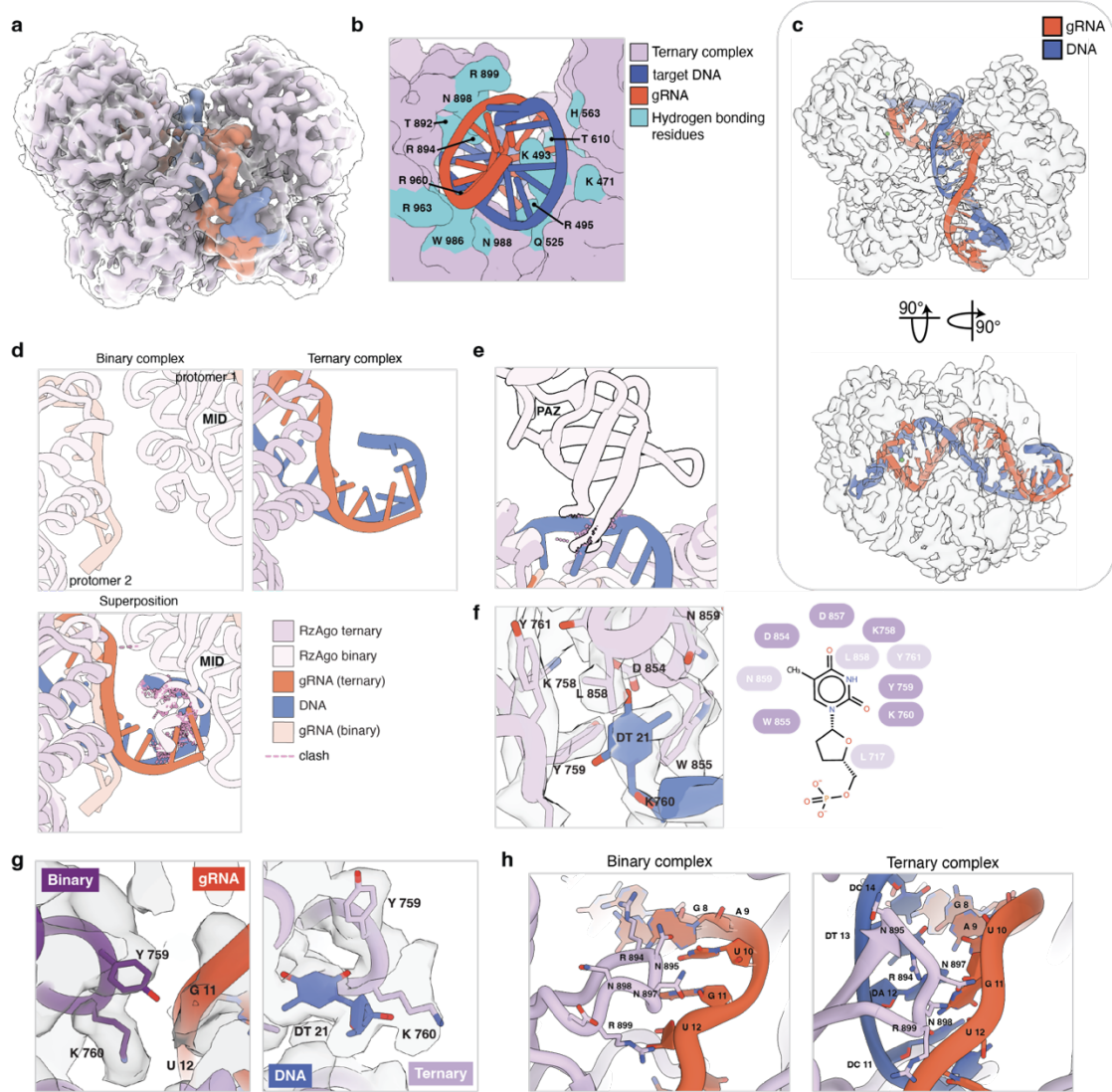


Figure 4. Ternary complex structure. **a**, Cryo-EM structure of RZAgo in complex with an 18-nt guide RNA and a fully complementary 24-nt target DNA overlaid with low-contour density map (transparent). Guide RNA is shown in red and target DNA in blue. **b**, Diagram showing residues engaging in hydrogen bonding with the gRNA-3' end of the heteroduplex. **c**, Transparent map of the ternary complex with the heteroduplex docked inside, showing the straightness of the heteroduplex. **d**, Structural superposition revealing a steric clash between the distal end of the RNA–DNA heteroduplex and a neighboring RZAgo protomer within the filament. **e**, Structural superposition showing a steric clash between the PAZ domain of the neighboring protomer and the target DNA strand upon heteroduplex formation. **f**, Left: The MID-domain pocket engages nucleotide dT21 of the target DNA strand. Right: Flat schematic showing contact network around DT21. **g**, Conformational rearrangement of residues forming the dT21-binding pocket between the binary and ternary states. **h**, PIWI loop orientation and contacts in the binary complex (left) and ternary complex (right).

Figure 5.

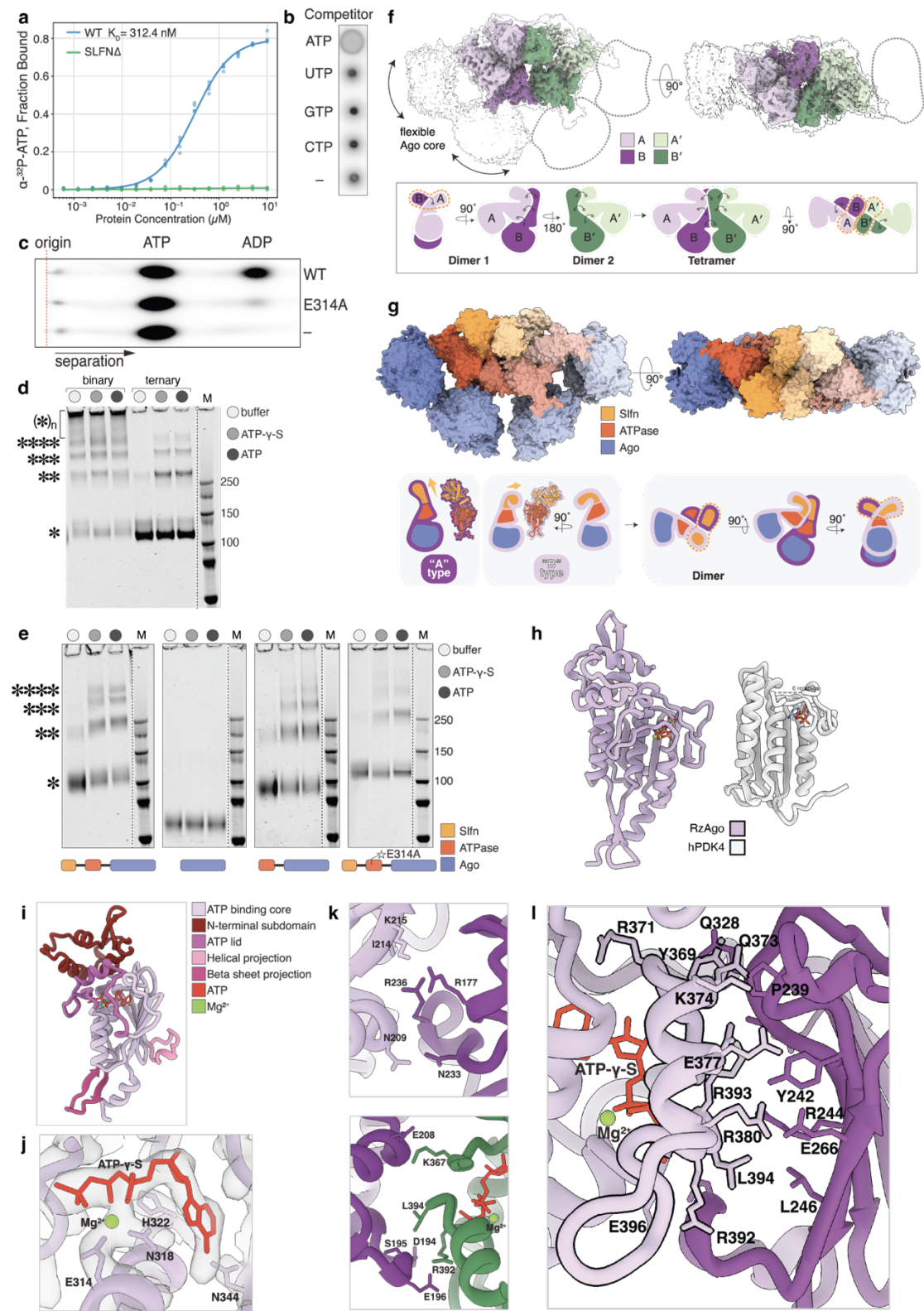


Figure 5. Schlafase supradomains and RZAgo Apo tetramer structure. **a**, ATP-binding isotherms for wild-type RZAgo and a Schlafase-deletion mutant measured by DRaCALA. **b**, Competition DRaCALA assay with unlabeled nucleotides demonstrating selective binding of ATP. **c**, Thin-layer chromatography (TLC) analysis of ATP hydrolysis by wild-type RZAgo and the ATPase-deficient E314A mutant. **d**, Crosslinking SDS–PAGE analysis of RZAgo oligomerization in the presence of guide RNA and target DNA. Positions of monomers, dimers, trimers and tetramers are indicated. **e**, Crosslinking SDS–PAGE analysis of RZAgo oligomerization upon mutation of individual domains. Positions of monomers, dimers, trimers and tetramers are indicated. **f**, Top, cryo-EM reconstruction of the RZAgo tetramer shown in side and top views, with individual protomers color-coded. A low-contour transparent density is shown to highlight conformational flexibility of the Ago core domains. Bottom, schematic model illustrating hierarchical tetramer assembly from two identical dimers arranged with pseudo-C2 symmetry. Dashed lines indicate the interface between adjacent Schlafen domains. **g**, Top, space-filling representation of the RZAgo tetramer colored by domain (side and top views as in **b**). Bottom, schematic depiction of two conformational states of RZAgo and the proposed pathway of dimer formation. **h**, Side-by-side comparison of RZAgo ATPase domain (left) with hPDK4 (right). **i**, Structure of the RZAgo GHKL ATPase domain, with major structural elements indicated. **j**, Close-up view of the ATP-binding pocket within the GHKL domain, highlighting residues involved in ATP coordination and Mg^{2+} binding. **k**, Top: Interactions between two RZAgo monomers within a dimer mediated by the N-terminal subdomain. Bottom: Inter-dimer interactions between RZAgo monomers contributing to tetramer assembly, mediated by the N-terminal subdomain. **l**, Multivalent interactions between RZAgo monomers within the dimer mediated by the ATP-lid region.

Figure 6.

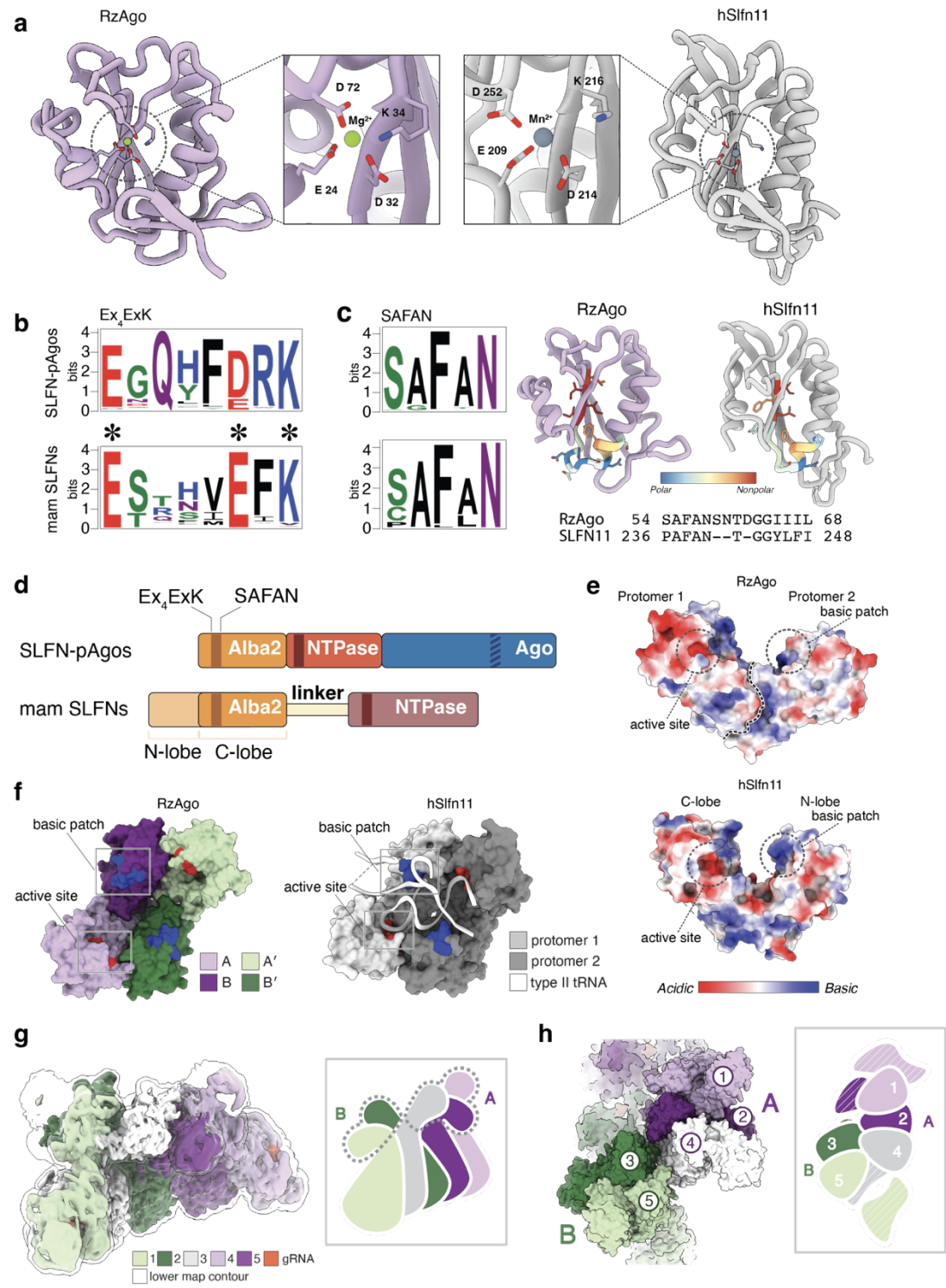


Figure 6. RzAgo Schlafen domain comparison with hSLFN11. **a**, Comparison of the Schlafen domain from RzAgo and human SLFN11 (hSLFN11), revealing a striking structural similarity. The position of the putative nuclease active site is indicated. **Insets:** Close-up views of the nuclease active sites of RzAgo and hSLFN11, showing conserved residues involved in divalent metal ion coordination. **b**, Sequence logo highlighting conservation of the nuclease active site residues. **c**, Left: Sequence logo highlighting conservation of the SAFAN motif between SLFN-pAgos and mammalian Schlafen proteins. Right: Comparison of the RzAgo Schlafen domain with the hSlfn11 Schlafen C domain, highlighting the SAFAN motifs colored by hydrophobicity (polar: blue, hydrophobic: orange). **d**, Comparison of domain organization between SLFN-pAgos and type III mammalian Schlafen proteins. **e**, Horseshoe-shaped dimer formed by two adjacent RzAgo protomers within the dimer recapitulates the overall architecture of monomeric hSLFN11. Positions of the nuclease active site and a conserved basic surface patch are indicated. **f**, Arrangement of Schlafen domains within the RzAgo tetramer resembles the dimeric organization of hSLFN11 observed in its substrate-bound state. **g**, Left: Binary pentamer map showing two contour levels and colored by protomer, Right: Schematic of (left) to show organization of pAgo cores relative to Schlafase supra-domains. **h**, Left: Surface view of binary pentamer model colored by protomer from above, Right: Schematic of (left) to show organization of pAgo cores relative to Schlafase supra-domains.

Figure 7.

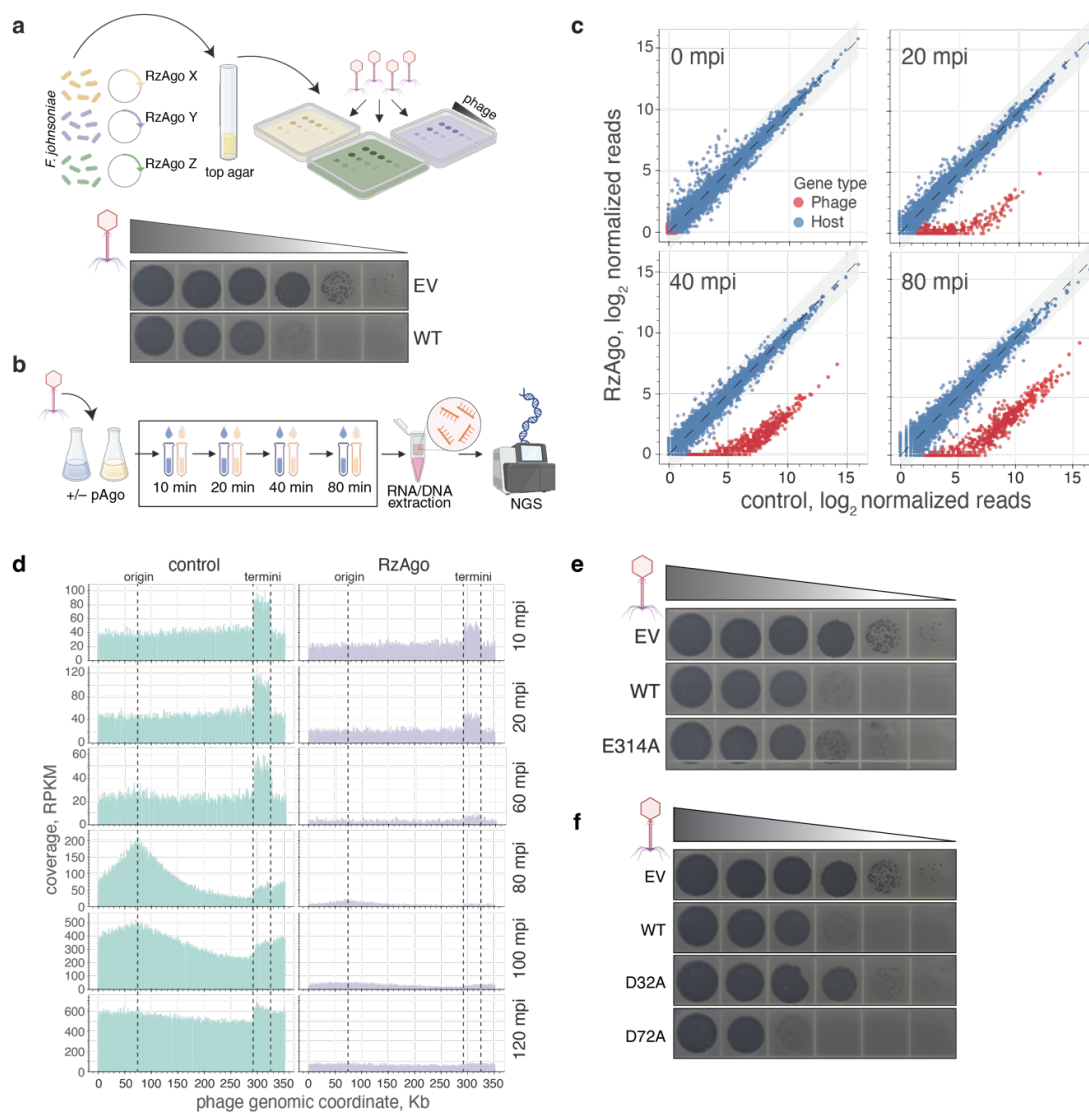
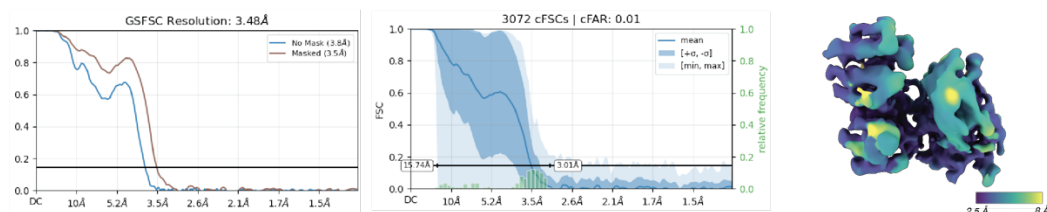


Figure 7. RzAgo protects host cells against phage infection. **a**, Top, schematic of the phage infectivity assay. Bottom, efficiency-of-plating assay of phage $\phi\Omega 29$ on *F. johnsoniae* strains carrying an empty vector (EV) or expressing wild-type RzAgo (WT). **b**, Experimental strategy to probe the molecular basis of RzAgo-mediated phage defense. Cells expressing RzAgo or an empty vector were infected with phage at a multiplicity of infection (MOI) of ~ 5 , and samples were collected at the indicated times post infection for DNA- and RNA-sequencing analyses. **c**, Scatter plot of total RNA-seq data showing gene expression changes over the course of infection. Blue dots represent host genes and red dots represent phage genes. The shaded gray region indicates a $\pm 1.5 \log_2$ fold-change window. **d**, Genome-wide distribution of sequencing reads mapped to the $\phi\Omega 1$ genome at successive time points after infection. Dashed lines denote the replication origin and long direct terminal repeats. **e**, Efficiency-of-plating assay of phage $\phi\Omega 29$ on *F. johnsoniae* strains carrying an empty vector (EV) or expressing wild-type RzAgo (WT) or the ATPase-deficient E314A mutant. **f**, Efficiency-of-plating assay of phage $\phi\Omega 29$ on *F. johnsoniae* strains carrying an empty vector (EV) or expressing wild-type RzAgo (WT), a nuclease-deficient mutant (D32A), or a hyperactive mutant (D72A).

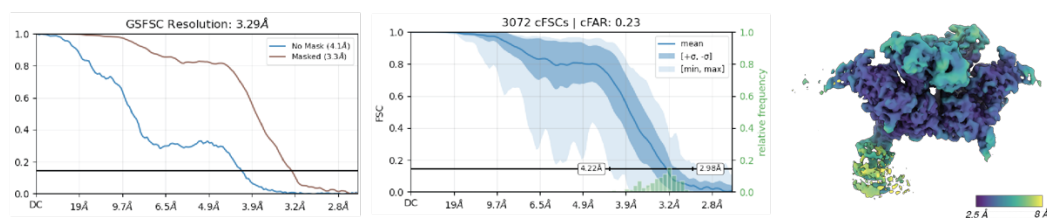
Extended Data Figures

Extended Data Figure 1.

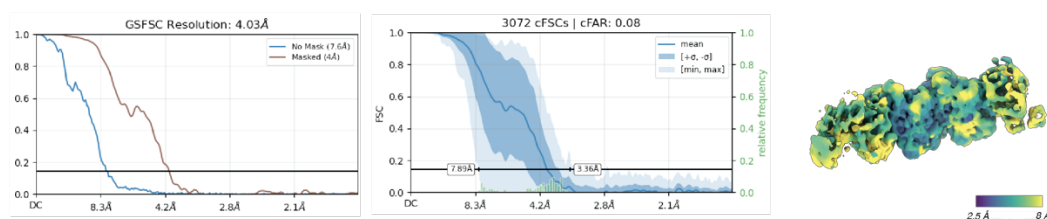
A. Apo monomer, no AGS, no glutaraldehyde



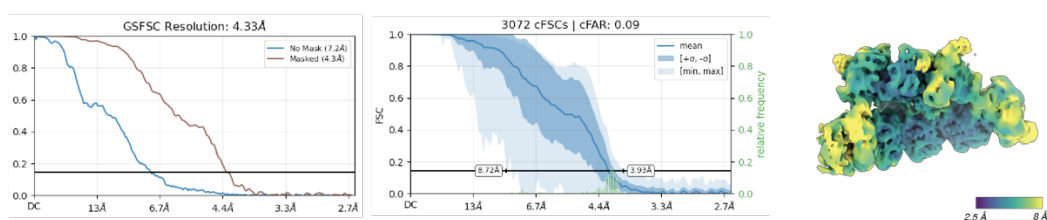
B. Apo tetramer, +AGS, +glutaraldehyde



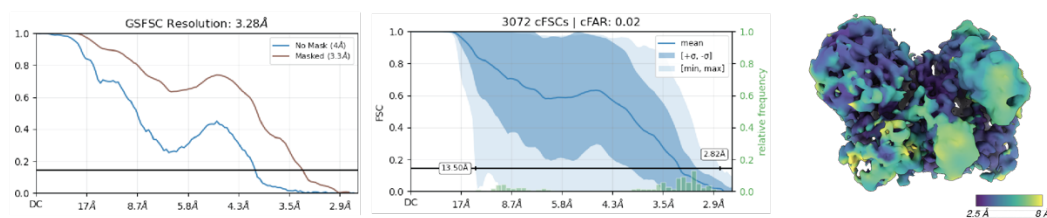
C. Binary complex, no AGS, no glutaraldehyde



D. Binary complex, +AGS, +glutaraldehyde

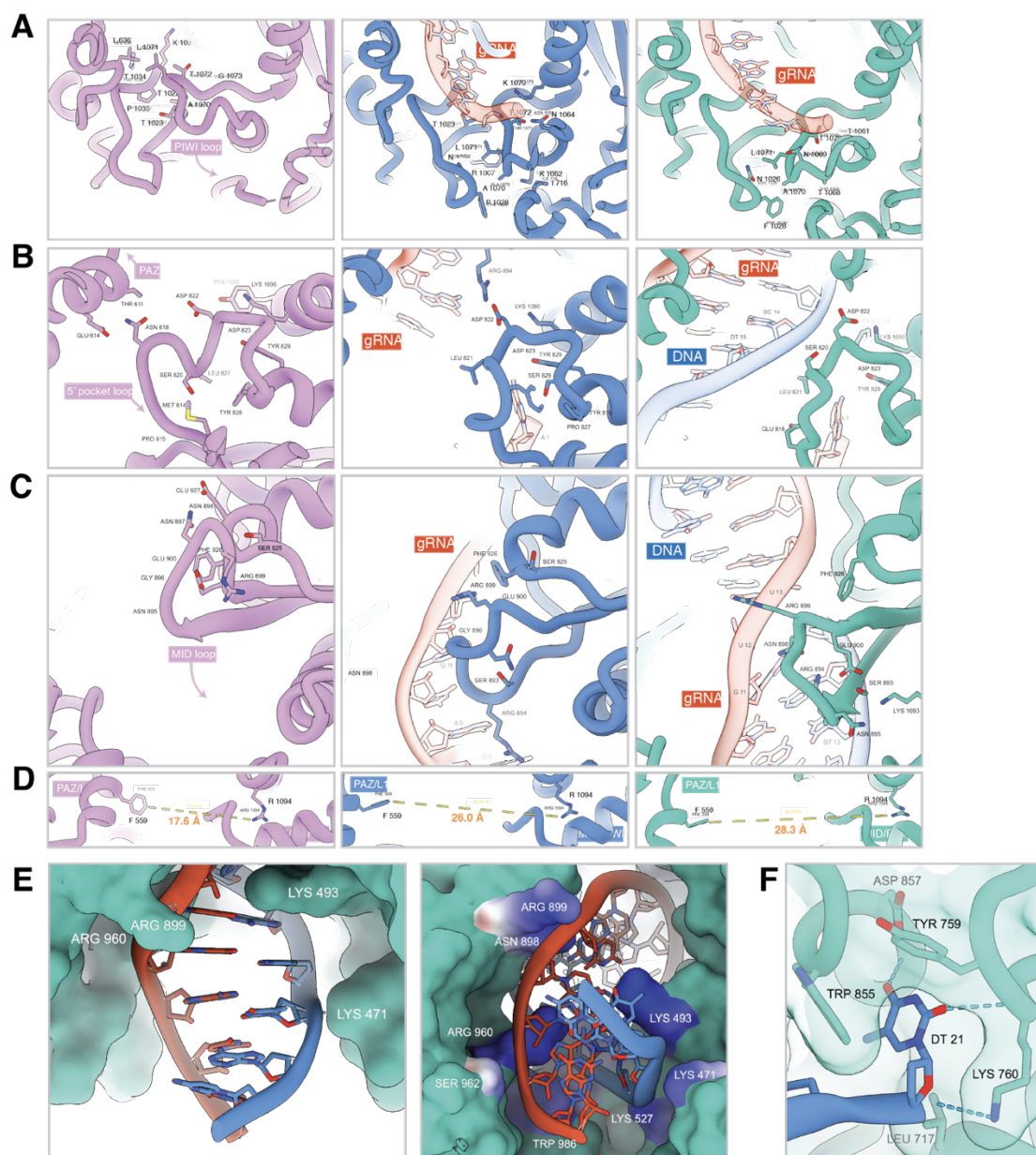


E. Ternary complex, +ATP, no glutaraldehyde



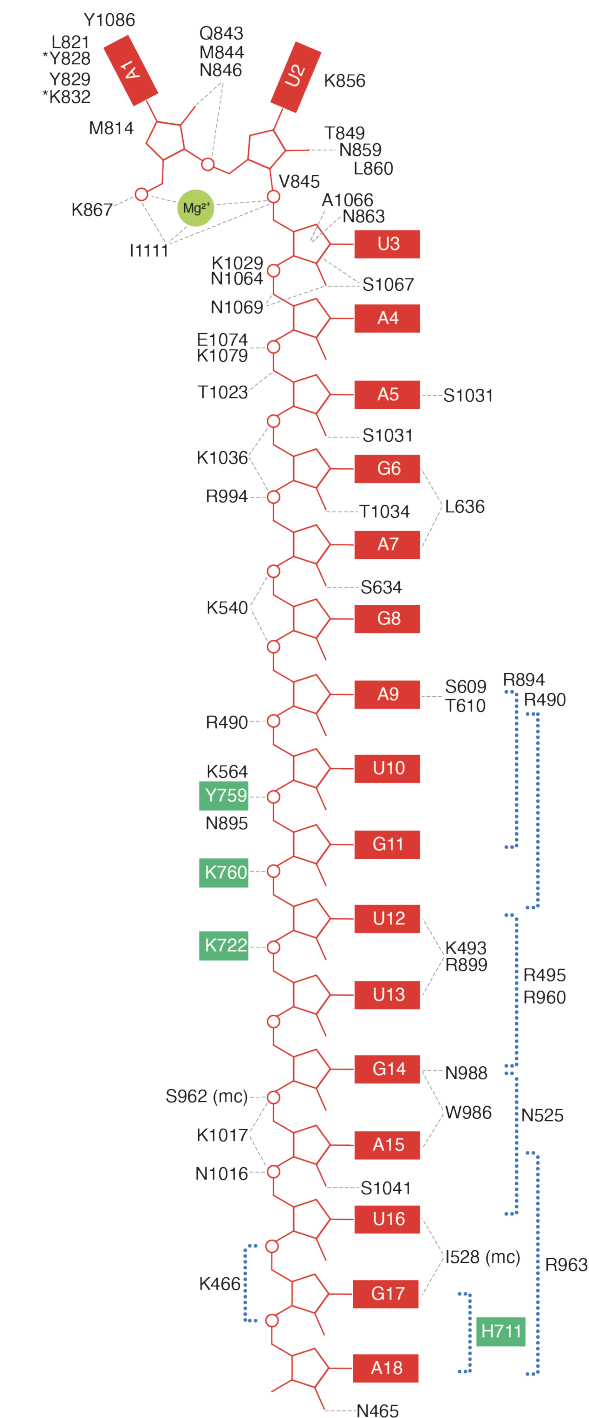
Extended Data Figure 1. GSFSC plots and cFAR plots for each structure, with each map showing the local resolution using the Viridis colormap (purple: 2.5 Å, yellow: 8 Å for all). A. Apo monomer, no ATP- γ -S, no glutaraldehyde. B. Apo tetramer, + ATP- γ -S, +glutaraldehyde. C. Binary complex, no ATP- γ -S, no glutaraldehyde. D. Binary complex, +ATP- γ -S, +glutaraldehyde. E. Ternary complex, +ATP, no glutaraldehyde.

Extended Data Figure 2.



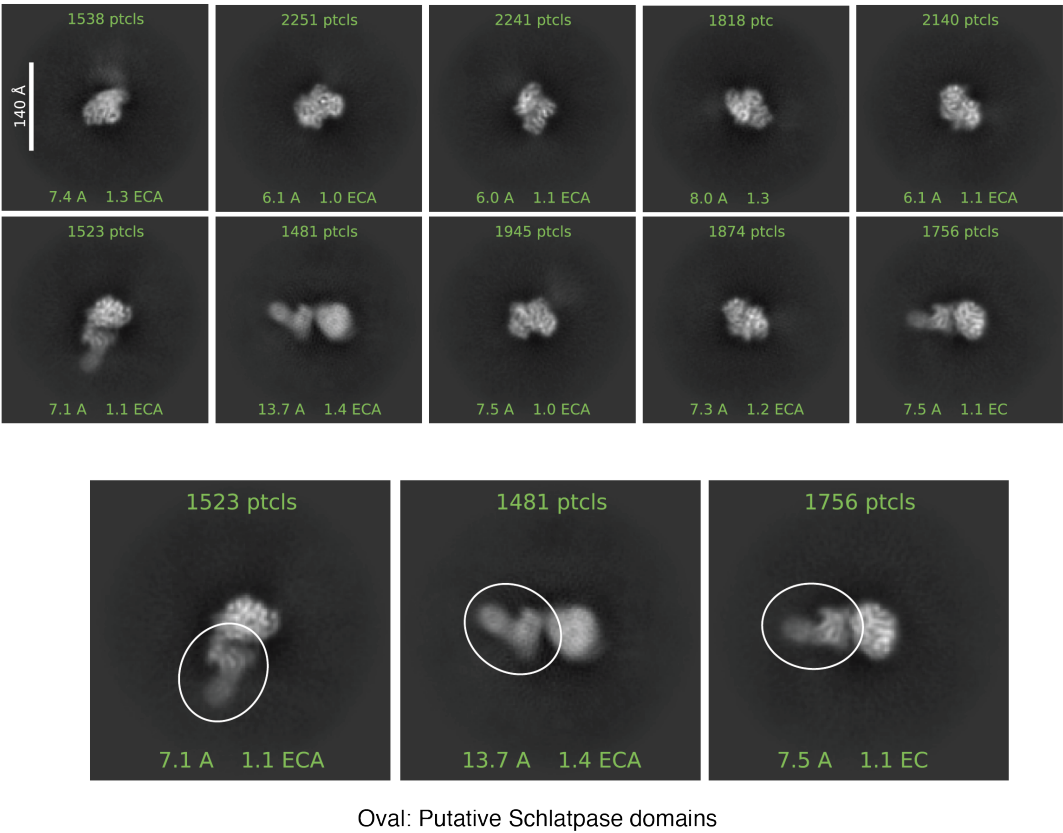
Extended Data Figure 2. A-D. Contacts of mobile loops in the apo (left, purple), binary (middle, blue) and ternary (right, teal) pAgo core structures. E. gRNA 3' end of heteroduplex in ternary complex, with contacting residues labeled in white (left) and/or colored by Coulombic potential (right). F. DT21 in binding pocket with surface overlay (translucent teal).

Extended Data Figure 3.

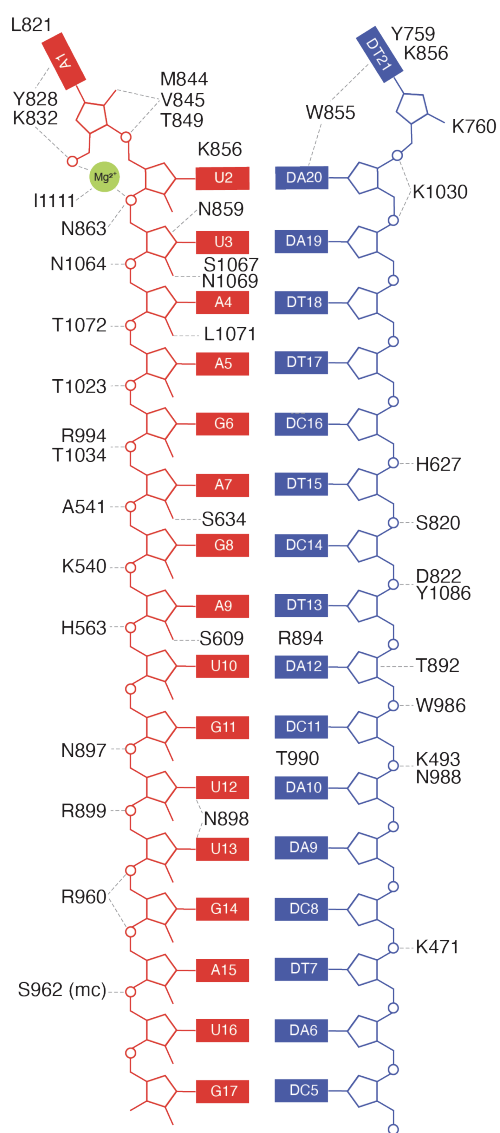


Extended data Figure 3. Contacts made between pAgo core and gRNA in the binary complex.

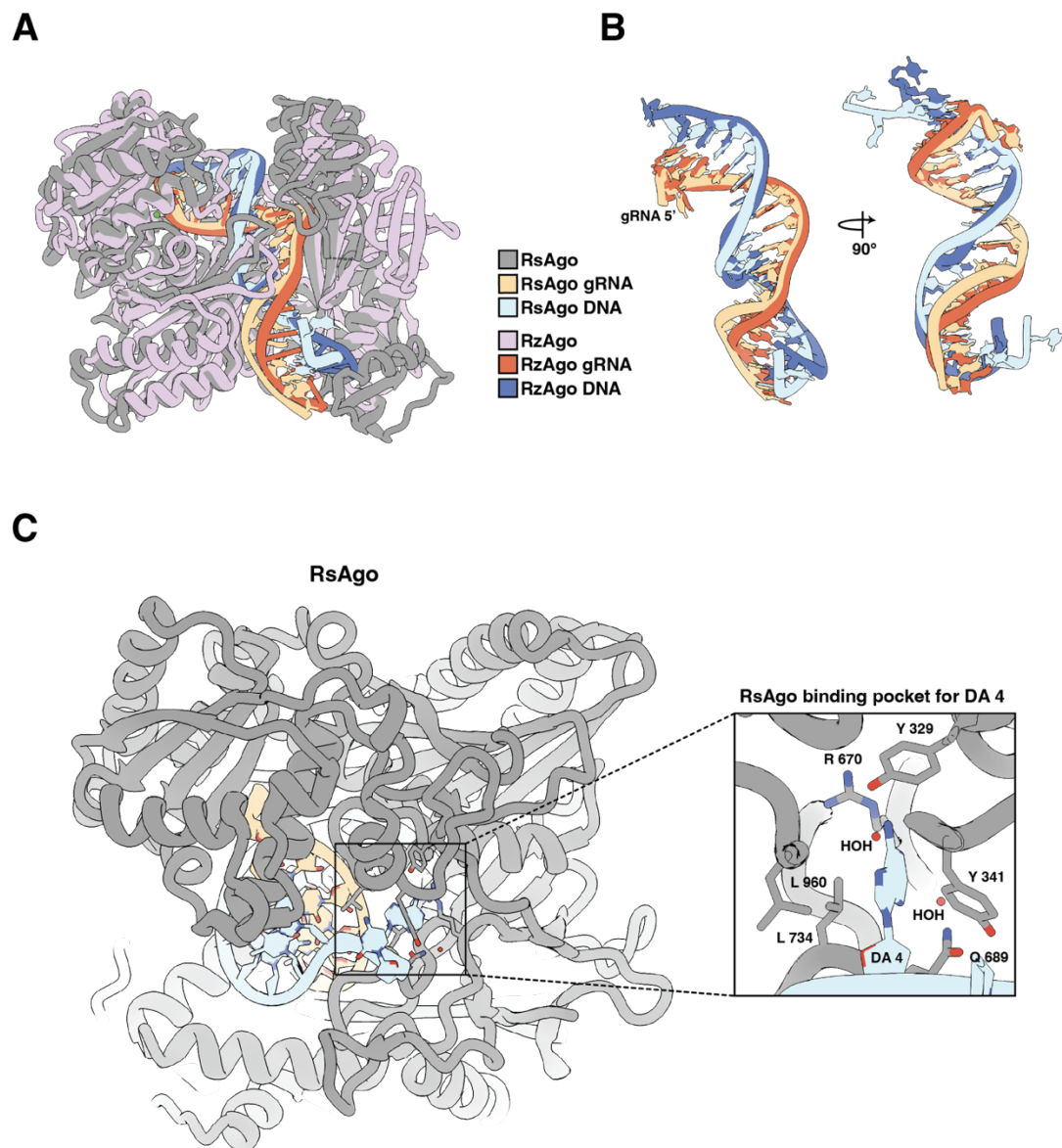
Extended Data Figure 4.



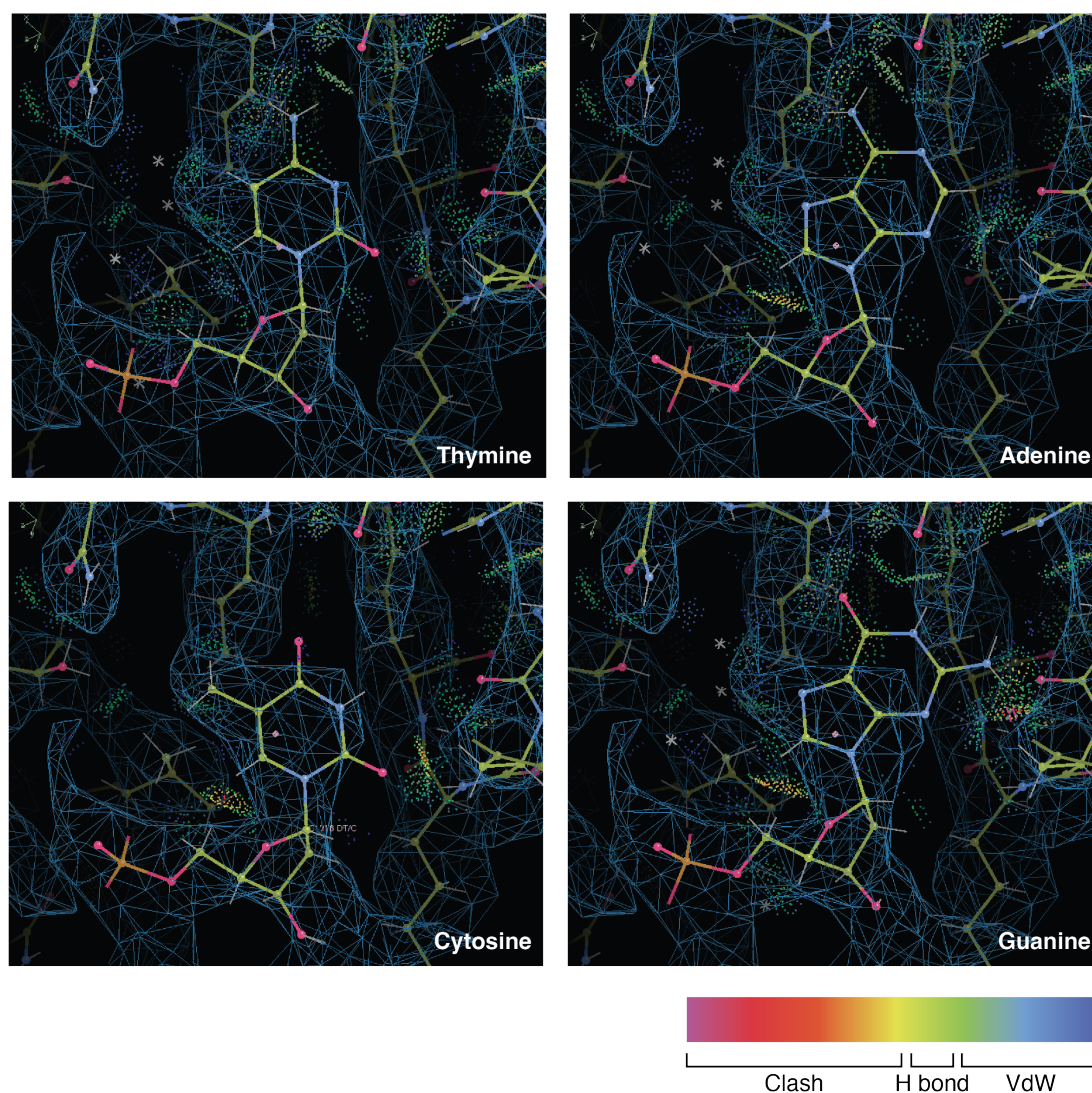
Extended Data Figure 4. 2D class averages of the ternary complex from the first round of averaging after blob picking, with three classes selected to show the putative Schlatpase domains aligned.

Extended Data Figure 5.**Extended Data Figure 5.** Contacts made between pAgo core and the gRNA-target DNA heteroduplex in the ternary complex.

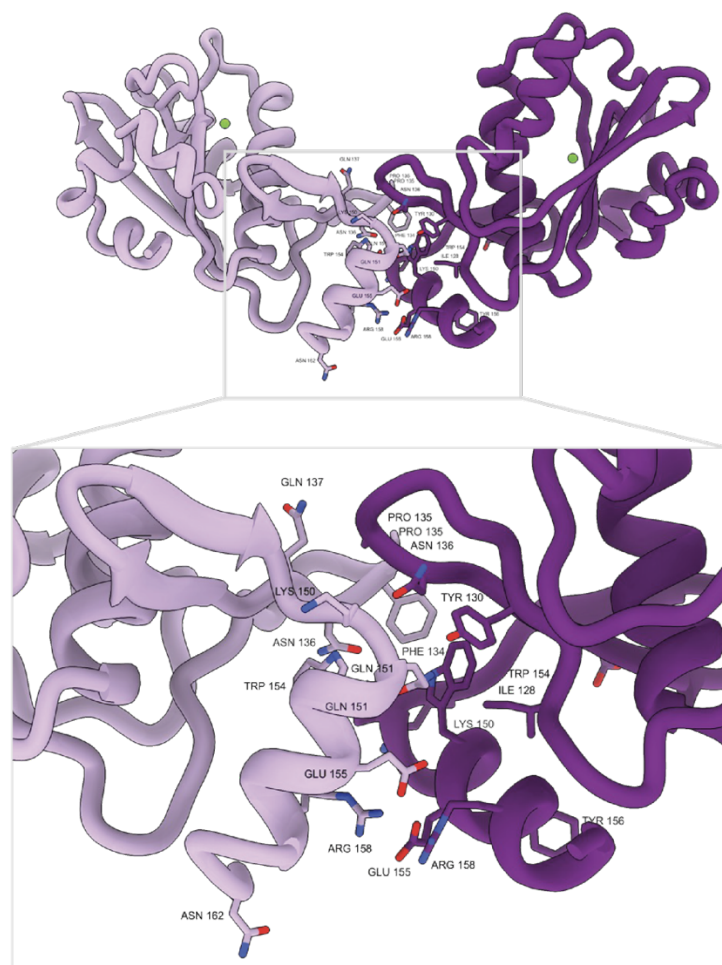
Extended Data Figure 6.

**Extended Data Figure 6.** Comparison of RsAgo ternary complex with RzAgo ternary complex.

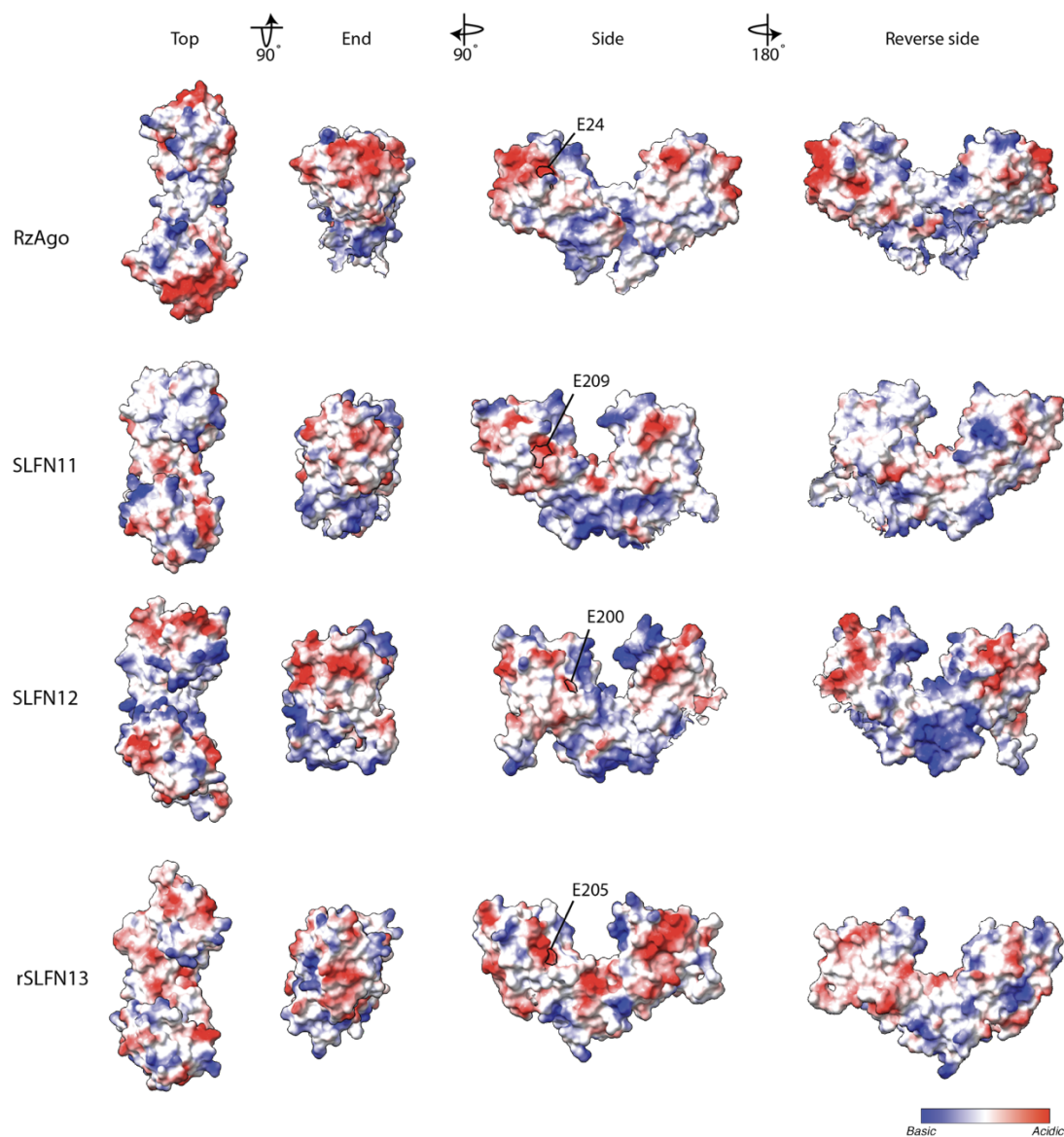
A. Overlay of RsAgo and RzAgo ternary complex ribbon diagrams. B. RsAgo (light) and RzAgo (dark) heteroduplexes overlaid to compare the curvature of each strand. C. RsAgo structure highlighting a binding pocket that mirrors the DT21 binding pocket.

Extended Data Figure 7.

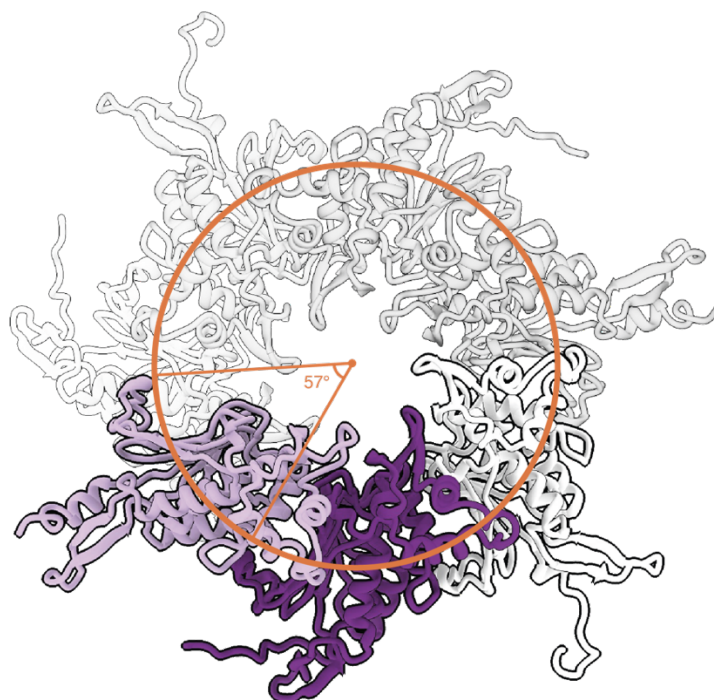
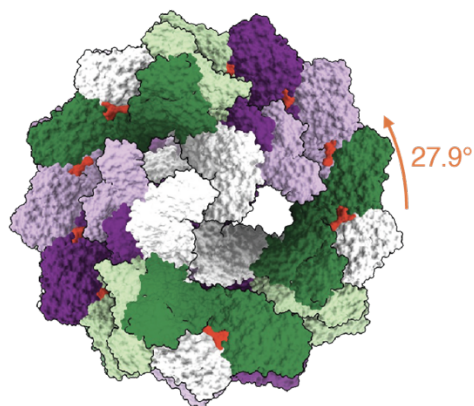
Extended Data Figure 7. Docking study in Coot to examine the fit of different nucleobases in the DT21 binding pocket, with the density overlaid. Clashes (red through yellow), hydrogen bonds (light green), and Van der Waals interactions (green through blue) are shown as dotted discs.

Extended Data Figure 8.

Extended Data Figure 8. Interacting helices at the C-terminal base of the RzAgo SLFN domains within a dimer.

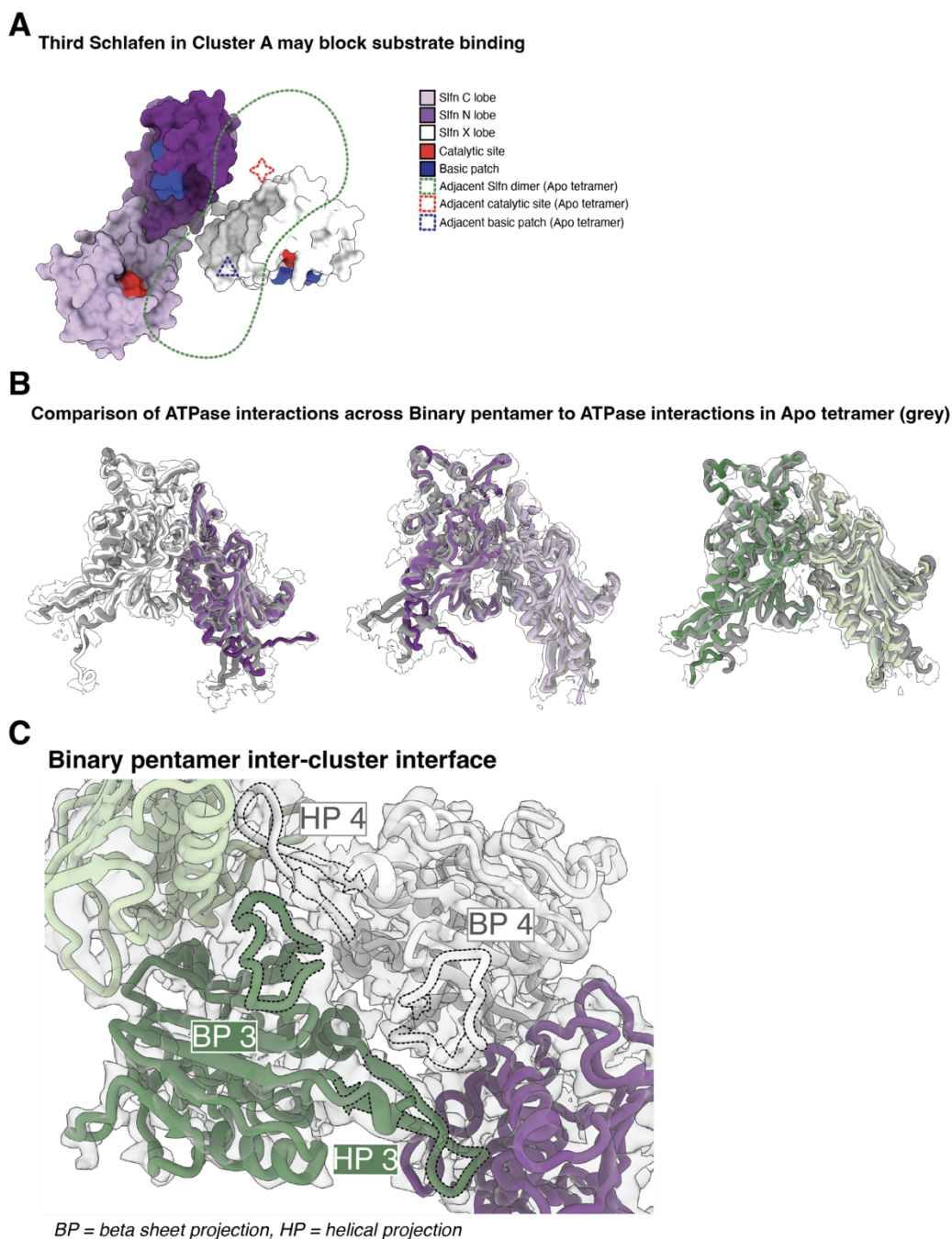
Extended Data Figure 9.

Extended Data Figure 9. Coulombic potential comparison of the RzAgo SLFN dimer with mammalian Slfns, hSlfn11, hSlfn12, and rSlfn13, from four angles. In the third column, the active glutamate in each Slfn is labeled.

Extended Data Figure 10.**A****B**

Extended Data Figure 10. Curvature comparison between ATPases and pAgo cores. A. Model to show the curvature of head-to-tail ATPase domains, with a rotation of 57 degrees between each monomer. B. Curve induced by pAgo head-to-tail interactions, which shows a rotation of 28 degrees between each monomer.

Extended Data Figure 11.



Extended Data Figure 11. Alternate ATPase interactions seen in crosslinked Binary complex structure. A. Possible clash of third SLFN domain in 3-Schlafase domain cluster. B. Comparison of three of four inter-ATPase interactions in the Binary pentamer (colors) to the same interface in the Apo tetramer (grey). C. Fourth and unique inter-ATPase interface.

Bibliography

- Anton, B. P., Blumenthal, R., Eaglesham, J. B., Mruk, I., Roberts, R. J., Xu, S., Weigele, P. R., & Raleigh, E. A. (2025). Biology of host-dependent restriction-modification in prokaryotes. *EcoSal Plus*, 13(1), eesp-0014-2022. <https://doi.org/10.1128/ecosalplus.esp-0014-2022>
- Afonine, P. V., Klaholz, B. P., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., Adams, P. D., & Urzhumtsev, A. (2018). New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallographica Section D Structural Biology*, 74(9), 814–840. <https://doi.org/10.1107/S2059798318009324>
- Ban, C., Junop, M., & Yang, W. (1999). Transformation of MutL by ATP Binding and Hydrolysis: A Switch in DNA Mismatch Repair. *Cell*, 97(1), 85–97. [https://doi.org/10.1016/S0092-8674\(00\)80717-5](https://doi.org/10.1016/S0092-8674(00)80717-5)
- Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J., & Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods*, 16(11), 1153–1160. <https://doi.org/10.1038/s41592-019-0575-8>
- Bernheim, A., Millman, A., Ofir, G., Meitav, G., Avraham, C., Shomar, H., Rosenberg, M. M., Tal, N., Melamed, S., Amitai, G., & Sorek, R. (2021). Prokaryotic viperins produce diverse antiviral molecules. *Nature*, 589(7840), 120–124. <https://doi.org/10.1038/s41586-020-2762-2>
- Bick, T., Dominiak, P. M., & Wendler, P. (2024). Exploiting the full potential of cryo-EM maps. *BBA Advances*, 5, 100113. <https://doi.org/10.1016/j.bbadv.2024.100113>
- Corbett, K. D., & Berger, J. M. (2003). Structure of the topoisomerase VI-B subunit: Implications for type II topoisomerase mechanism and evolution. *The EMBO Journal*, 22(1), 151–163. <https://doi.org/10.1093/emboj/cdg008>
- Corbett, K. D., & Berger, J. M. (2005). Structural Dissection of ATP Turnover in the Prototypical GHF ATPase TopoVI. *Structure*, 13(6), 873–882. <https://doi.org/10.1016/j.str.2005.03.013>
- Croll, T. I. (2018). ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallographica Section D: Structural Biology*, 74(6), 519–530. <https://doi.org/10.1107/S2059798318002425>

- Doxzen, K. W., & Doudna, J. A. (2017). DNA recognition by an RNA-guided bacterial Argonaute. *PLOS ONE*, 12(5), e0177097. <https://doi.org/10.1371/journal.pone.0177097>
- Dutta, R., & Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences*, 25(1), 24–28. [https://doi.org/10.1016/S0968-0004\(99\)01503-0](https://doi.org/10.1016/S0968-0004(99)01503-0)
- Emsley, P., & Crispin, M. (2018). Structural analysis of glycoproteins: Building N-linked glycans with Coot. *Acta Crystallographica Section D: Structural Biology*, 74(4), Article 4. <https://doi.org/10.1107/S2059798318005119>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), Article 4. <https://doi.org/10.1107/S0907444910007493>
- Hou, P., Hao, W., Qin, B., Li, M., Zhao, R., & Cui, S. (2023). Structural and biochemical characterization of Schlafen11 N-terminal domain. *Nucleic Acids Research*, 51(13), 7053–7070. <https://doi.org/10.1093/nar/gkad509>
- Jo, U., & Pommier, Y. (2022). Structural, molecular, and functional insights into Schlafen proteins. *Experimental & Molecular Medicine*, 54(6), 730–738. <https://doi.org/10.1038/s12276-022-00794-0>
- Johnson, A. G., Wein, T., Mayer, M. L., Duncan-Lowey, B., Yirmiya, E., Oppenheimer-Shaanan, Y., Amitai, G., Sorek, R., & Kranzusch, P. J. (2022). Bacterial gasdermins reveal an ancient mechanism of cell death. *Science*, 375(6577), 221–225. <https://doi.org/10.1126/science.abj8432>
- 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, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kanevskaya, A., Narwal, M., Lisitskaya, L., Moiseenko, A. V., Sokolova, O. S., Sluchanko, N. N., Murakami, K. S., & Kulbachinskiy, A. (2025). Argonaute-HNH filaments triggered by invader

- DNA confer bacterial immunity. *Nature Communications*, 16(1), 11389. <https://doi.org/10.1038/s41467-025-66189-7>
- Kibby, E. M., Conte, A. N., Burroughs, A. M., Nagy, T. A., Vargas, J. A., Whalen, L. A., Aravind, L., & Whiteley, A. T. (2023). Bacterial NLR-related proteins protect against phage. *Cell*, 186(11), 2410-2424.e18. <https://doi.org/10.1016/j.cell.2023.04.015>
- Koopal, B., Mutte, S. K., & Swarts, D. C. (2023). A long look at short prokaryotic Argonautes. *Trends in Cell Biology*, 33(7), 605–618. <https://doi.org/10.1016/j.tcb.2022.10.005>
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., ... Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallographica Section D: Structural Biology*, 75(10), Article 10. <https://doi.org/10.1107/S2059798319011471>
- Lisitskaya, L., Aravin, A. A., & Kulbachinskiy, A. (2018). DNA interference and beyond: Structure and functions of prokaryotic Argonaute proteins. *Nature Communications*, 9(1), 5165. <https://doi.org/10.1038/s41467-018-07449-7>
- Makarova, K. S., Wolf, Y. I., Snir, S., & Koonin, E. V. (2011). Defense Islands in Bacterial and Archaeal Genomes and Prediction of Novel Defense Systems ▽. *Journal of Bacteriology*, 193(21), 6039–6056. <https://doi.org/10.1128/JB.05535-11>
- Metzner, F. J., Wenzl, S. J., Kugler, M., Krebs, S., Hopfner, K.-P., & Lammens, K. (2022). Mechanistic understanding of human SLFN11. *Nature Communications*, 13(1), 5464. <https://doi.org/10.1038/s41467-022-33123-0>
- Miyoshi, T., Ito, K., Murakami, R., & Uchiumi, T. (2016). Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nature Communications*, 7(1), 11846. <https://doi.org/10.1038/ncomms11846>

- Ni, D., Lu, X., Stahlberg, H., & Ekundayo, B. (2023). Activation mechanism of a short argonaute-TIR prokaryotic immune system. *Science Advances*, 9(29), eadh9002. <https://doi.org/10.1126/sciadv.adh9002>
- Pate, J. L., Petzold, S. J., & Chang, L.-Y. E. (1979). Phages for the gliding bacterium *Cytophaga johnsonae* that infect only motile cells. *Current Microbiology*, 2(5), 257–262. <https://doi.org/10.1007/BF02602855>
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society*, 30(1), 70–82. <https://doi.org/10.1002/pro.3943>
- Piscitelli, J. M., & Manhart, C. M. (2025). The GHKL ATPase Family as a Paradigm for MutL Homolog Function in DNA Mismatch Repair. *International Journal of Molecular Sciences*, 26(24), 12157. <https://doi.org/10.3390/ijms262412157>
- Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods*, 14(3), Article 3. <https://doi.org/10.1038/nmeth.4169>
- Reidy, M., & Masison, D. C. (2025). ATP plays a structural role in Hsp90 function. *Nature Communications*, 16(1), 6710. <https://doi.org/10.1038/s41467-025-61962-0>
- Ryazansky, S., Kulbachinskiy, A., & Aravin, A. A. (2018). The Expanded Universe of Prokaryotic Argonaute Proteins. *mBio*, 9(6), 10.1128/mbio.01935-18. <https://doi.org/10.1128/mbio.01935-18>
- Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 32(4), 347–355. <https://doi.org/10.1038/nbt.2842>
- Schwarz, D. A., Katayama, C. D., & Hedrick, S. M. (1998). Schlafen, a new family of growth regulatory genes that affect thymocyte development. *Immunity*, 9(5), 657–668. [https://doi.org/10.1016/s1074-7613\(00\)80663-9](https://doi.org/10.1016/s1074-7613(00)80663-9)

- Song, X., Lei, S., Liu, S., Liu, Y., Fu, P., Zeng, Z., Yang, K., Chen, Y., Li, M., She, Q., & Han, W. (2023). Catalytically inactive long prokaryotic Argonaute systems employ distinct effectors to confer immunity via abortive infection. *Nature Communications*, 14(1), 6970. <https://doi.org/10.1038/s41467-023-42793-3>
- Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., Patel, D. J., & van der Oost, J. (2014). The evolutionary journey of Argonaute proteins. *Nature Structural & Molecular Biology*, 21(9), 743–753. <https://doi.org/10.1038/nsmb.2879>
- Taboada, V. P., Wu, Y., Cassidy, R., Medvedev, K. E., Loeff, L., Nemudraia, A., & Nemudryi, A. (2025). Bacterial Schlafens mediate anti-phage defense. *bioRxiv*, 2025.07.24.666596. <https://doi.org/10.1101/2025.07.24.666596>
- Vogel, C., Berzuini, C., Bashton, M., Gough, J., & Teichmann, S. A. (2004). Supra-domains: Evolutionary units larger than single protein domains. *Journal of Molecular Biology*, 336(3), 809–823. <https://doi.org/10.1016/j.jmb.2003.12.026>
- Xiao, Y., Liu, T. M., & MacRae, I. J. (2023). A tiny loop in the Argonaute PIWI domain tunes small RNA seed strength. *EMBO Reports*, 24(6), e55806. <https://doi.org/10.15252/embr.202255806>
- Yang, J.-Y., Deng, X.-Y., Li, Y.-S., Ma, X.-C., Feng, J.-X., Yu, B., Chen, Y., Luo, Y.-L., Wang, X., Chen, M.-L., Fang, Z.-X., Zheng, F.-X., Li, Y.-P., Zhong, Q., Kang, T.-B., Song, L.-B., Xu, R.-H., Zeng, M.-S., Chen, W., ... Gao, S. (2018). Structure of Schlafen13 reveals a new class of tRNA/rRNA- targeting RNase engaged in translational control. *Nature Communications*, 9(1), 1165. <https://doi.org/10.1038/s41467-018-03544-x>
- Zhang, W., Jiang, Y., Li, Y., Zhen, X., Xu, S., Xia, N., Li, S., Wang, X., Zheng, Q., & Ouyang, S. (2026). Filament assembly induced by the recognition of target DNA activates the prokaryotic Argonaute SPARDA system. *Nature Communications*, 17(1), 115. <https://doi.org/10.1038/s41467-025-68195-1>

Chapter 6

CONCLUSIONS

In this thesis, I have explored four very distinct biological systems through the lens of structural biology, and I will summarize the conclusions from each project here.

Chapter 2: Identification and structural characterization of a broadly neutralizing antibody in complex with the SARS-CoV-2 S protein RBD

In this project, I identified the antibody 11G-12v as belonging to Class 1/4. I additionally found that it shares sequence and structural similarity to a set of previously discovered anti-SARS-CoV-2 RBD antibodies that have a long CDRH3 loop with a YYDRxG motif, which associate with and extend the central beta sheet of the SARS-CoV-2 RBD. This interaction stabilizes this class of antibody such that the light chain variable domain sterically blocks interaction with ACE2. Though these antibodies arise with little somatic hypermutation, they have not been found in high abundance in patient serum. I attribute this to the fact that longer CDRH3s can be more prone to self-reactivity, and therefore their production may be reduced to prevent poor outcomes. Though my cryo-EM structure was solved to an overall resolution of 3.4 Å, the region around the Fab binding sites show poorer resolution due to flexibility of the RBDs in the “up” position, and to the Fab elbow joints. I therefore propose that a crystal structure would support a better understanding of the exact association of 11G-12v with the RBD. However, the cryo-EM structure provides unique information showing asymmetry in the positions of the RBDs and Fabs between the three protomers of the S trimer, and further show a range of angles that the RBD occupies to bind to 11G-12v, which match other antibodies bearing the YYDRxG motif like COVA1-16 and C022, but not C118, which also possesses a long CDRH3 loop and extends the RBD beta sheet. These differences, in tandem with ELISA data to compare 11G-12v to other broadly neutralizing anti-SARS-CoV-2 RBD

antibodies, contribute to our understanding of the sequence and structural features of broadly neutralizing antibodies against coronaviruses.

Chapter 3: Literature review: A structural perspective on the p97 adaptor Phospholipase A2 Activating Protein (PLAA)

In this literature review, I examined the many documented activities of PLAA in the context of its structure, and considered them in terms of pleiotropy. PLAA/Ufd-3/Doa1 is attached to many processes mediated by ubiquitin tagging, including the ubiquitin-proteasome system and association with p97, but also appears to have ubiquitin-independent functions, such as mediating the interaction between DCAP-1 and its substrates via its IDR, and perhaps its namesake role as a phospholipase A2 activator. Though these roles require further investigation, looking at them through a structural lens provides a wealth of interesting avenues of exploration. For example, what function might be linked to the high density and radial organization of cysteine residues noted in the WD-40 domain, especially in human PLAA? Additionally, what is the purpose of two ubiquitin-binding domains? Structural studies of the WD-40 domain and the PFU show that they bind to the same face of ubiquitin, eliminating one possible motivation. Is it possible that PLAA recognizes certain branched forms of ubiquitin? Finally, PLAA's interaction with the p97 C-terminal tail is well established, but the extent to which PLAA functions as a p97 adaptor, rather than independently or in complex with other proteins, is not yet understood. Overall, the data support a pleiotropic role for PLAA, and suggest that further research will not only support our understanding of PLAA, but may also help to deepen our understanding of the cellular pathways touched by PLAA.

Chapter 4: Structural characterization of the FIP200 C-terminal CLAW domain bound to an autophagy adaptor peptide, TNIP1-pFIR

In this chapter, I solved the crystal structure of the FIP200 C-terminal CLAW domain in complex with a 13-amino acid phosphopeptide derived from the TNIP1 FIP200 interacting region, or FIR. This crystal structure supported our understanding of the basis of FIR recognition by CLAW. Though many documented FIRs share the conserved sequence D_D_I(F/Y/W/L/V)_X_X_L(I/V), other sequences with chemical similarities may also be recognized by the CLAW. Recognition of TNIP1's FIR, with sequence GTS{pSER}EFEVVTPEE, occurs largely through the nonspecific charge-charge interactions of the negatively charged phosphate group with a network of positively charged residues on the surface of the CLAW. Additional contacts form between a hydrophobic pocket in the same region and the phenylalanine moiety in the TNIP1-pFIR. Together with the acidic glutamate residues, these features stabilize the interaction between the TNIP1-pFIR and the CLAW. In tandem with other data obtained but not discussed in this thesis, this basic loop of CLAW serves as a relatively indiscriminate interactor with negatively charged molecules, including lipid head groups and acidic protein surfaces. This may allow the CLAW to recruit a wide range of molecules essential for autophagy initiation, though work along this vein is still ongoing. The hydrophobic pocket in CLAW adds some specificity for potential FIR motifs, however. FIR motifs closely resemble LC3 interacting regions (LIRs), allowing autophagy adaptors like p62 to interact with both FIP200 and Atg8 proteins like LC3, though the purpose of this is not yet well understood. Together with structures of the CLAW interacting with other FIRs and with small-molecule interactors, this research supports the rational design of autophagy-targeting chimeras that follow the principles of endogenous receptors.

Chapter 5: ATP-driven tetramerization and RNA-guided filament assembly control of a Schlafen–Argonaute antiphage defense

In this chapter, I worked with the Aravin group and solved five cryo-EM structures that dramatically affected our understanding of the prokaryotic Argonaute protein (pAgo), RzAgo. RzAgo was shown to belong to a novel clade of pAgos termed “extra-long”.

Members of this clade have a catalytically inactive set of Argonaute core domains, but possess an active Schlafen nuclease and an active GHKL-type ATPase domain with similarities to mammalian homologues. In mammalian systems, these proteins are shown to engage in multimerization as an essential component of their function. Here, my structures show that the Schlafen and ATPase domains of RzAgo, along with clade-level modifications to the pAgo core, such as the absence of an N domain, not only provide critical interfaces for multimerization, but that the binding states of the pAgo core – empty (apo), loaded with guide RNA (binary complex) or with a heteroduplex of guide RNA and target ssDNA (ternary complex) – promote different arrangements of RzAgo protomers. In the apo state, RzAgo forms tetramers – dimers of dimers with C2 symmetry, centering around the organization of the Schlafen domains. This state recapitulates the active state observed for human SFLN11, which cleaves tRNAs. Alternatively, when RzAgo binds guide RNA, new interfaces form between the pAgo core domains. These interfaces permit the inter-protomer interactions seen between the Schlafen and ATPase domains in the Apo dimers, but not the interactions between two dimers within the Apo tetramer: the C2 symmetry of the tetramer is prohibited by the pAgo core interactions. Instead, guide-bound RzAgo forms helical filaments featuring a repeating arrangement of pAgo cores and clusters of Schlafen-ATPase domains. While the pAgo core interactions could be resolved without any addition of a crosslinker, glutaraldehyde crosslinking was necessary to allow us to visualize the organization of the Schlafen-ATPase domains. This showed us that the Schlafen-ATPase supradomains may partition into clusters, which due to torsional strain are angled either to the inside of the filament curve, or to the outside. Additionally, the crosslinked gRNA-bound structure showed us that the Schlafen-ATPase supradomains can partition in clusters of three, not only in groups of two as observed in the Apo tetramer. Further structural studies of RzAgo bound to the heteroduplex without crosslinking resolved the pAgo core domains bound to the heteroduplex. When compared to the binary complex, this structure showed that target DNA binding both disrupts the contacts between the guide RNA and the adjacent protomer in a filament and forms a bulky protrusion that sterically blocks the interactions between the adjacent pAgo cores. This suggests that target DNA binding might serve as the mechanism for filament disassembly. Biochemical assays with mutations to key active-site

residues, as well as domain deletion mutants, performed by my collaborator Anton Kuzmenko, provide additional support for these hypotheses and others presented in this chapter. However, many open questions remain that would be best answered with structural data. First, though biochemical evidence supports the existence of Ternary complex tetramers, the exact organization of Ternary tetramers is not known. Second, Schlafen domains in mammalian systems can bind and cleave relatively large tRNA and rRNA targets, and though Anton has been able to obtain some sparse data concerning substrates cleaved by RzAgo, the data are inconclusive. This structure could potentially be obtained through DNA digestion followed by immunoprecipitation of RzAgo from phage-infected *F. johnsoniae* cell lysates to extract substrate-bound complexes.