

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Structural and enzymatic characterization of pH-dependent chitinase activity, and contributions to Diversity, Equity, Inclusion, and Justice

Permalink

<https://escholarship.org/uc/item/0m96z59v>

Author

Diaz, Roberto Efrain

Publication Date

2023

Supplemental Material

<https://escholarship.org/uc/item/0m96z59v#supplemental>

Peer reviewed|Thesis/dissertation

Structural and enzymatic characterization of pH-dependent chitinase activity, and contributions to Diversity, Equity, Inclusion, and Justice.

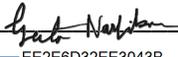
by
Roberto Efrain Diaz

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in
Biochemistry and Molecular Biology

in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

EE2E6D32EE3043B... Geeta Narlikar
Chair

DocuSigned by:

by4A3... James S. Fraser

DocuSigned by:

Richard Locksley

DocuSigned by:

8BD0E7DC6FB141A... Seemay Chou

Committee Members

Copyright 2023

by

Roberto Efraín Díaz

Dedication

This thesis is dedicated to my father, who taught me the value of an education and the importance of fighting for what's right regardless of the cost; to my mother, who constantly reminds me that the most important thing in life is the pursuit of happiness; to those fighting for a more just, equitable, queer, anti-racist world; and lastly, to myself, for not giving up even when this academic pursuit felt self-indulgent and meaningless.

Acknowledgments

There are numerous people who have been instrumental to my academic journey without whom I would not be here. I'd like to thank my various undergraduate research mentors – Michael S. Gaines (University of Miami NIH Initiative for Maximizing Student Development, 2013-2017); Kenneth J. Muller (University of Miami Lois Pope Neuroscience Summer Research, 2014); Jennifer Britton (University of Miami, 2014-2016); Steven Ullmann (University of Miami, 2014-2018); Mithu De and Roberta S. Fuller (University of Michigan Summer Research Opportunity Program, 2015); Christof Osman, Aylin Göke, Elif Karagöz, and Peter Walter (UCSF Summer Research Training Program, 2016; HHMI Exceptional Research Opportunities Program, 2017); Saloni Mehta and Vance Lemmon (University of Miami Senior Thesis in Neuroscience, 2016-2017). Without these formative research experiences, I wouldn't have discovered the fields of biochemistry or structural biology, nor would I have been accepted into graduate school. A special thanks to the administrators and mentors who told me about these research opportunities, proofread my application essays, ghostwrote my letters of recommendation, and constantly reminded me that I deserved to be there – Sean Kilpatrick (University of Miami Undergraduate Advising Services in Psychology, 2013-2014), Nicole Laviña (University of Miami Undergraduate Advising Services in Psychology, 2014-2017), Betsy Bossert Stoner (University of Miami NIH Initiative for Maximizing Student Development, 2014-2015), Gisselle Vélez Ruiz (University of Michigan Summer Research Opportunity Program, 2015), Michael Stokes (University of Miami First Year Fellows, 2015-2017), Julia Clark (UCSF Summer Research Training Program, 2016), Christine Genero, and Christy Schultz (HHMI Exceptional Research Opportunities Program, 2016-2017).

During my time at UCSF, I've had the privilege to work alongside and learn from incredibly smart, rigorous, and kind scientists. In the Fraser lab, I'd like to thank Benjamin A. Barad, my rotation mentor who introduced me to the world of chitin and Python; Iris D. Young for helping me improve (read: break) the lab website and indulging my numerous (and often nonsensical) questions about Python; Michael C. Thompson for teaching me everything I know about X-ray crystallography and for providing scientific guidance when I felt lost; Stephanie Wankowicz for providing critical scientific insight, for fostering an inclusive lab environment, and for encouraging others to do the same; and Taiasean Wu for her friendship, especially during the early days of the COVID-19 pandemic, her rigorous approach to science, and her inimitable perspective on issues of diversity, equity, inclusion, and justice (DEIJ).

For the first half of my PhD, I had the curious fortune of being a member of the R*dding lab. I'm grateful to Lucy D. Brennan, Madeline Keenen, and Luke Strauskulage for welcoming me into their lab (after they were notified that I was a lab member and not just a long-term visitor). Their unyielding curiosity, generosity, and persistence in the face of failure has always amazed me. Their friendship and faith in me propelled me forward when I felt like giving up.

On June 12th, 2016, I remember how alone I felt when I woke up in San Francisco to the news of the Pulse Shooting in Orlando, Florida. That feeling faded away when I arrived at the Castro and realized I was not alone in grieving this tragic event. I returned to UCSF the following summer and began to build a community of LGBTQ+ people. The Graduate and Postdoc Queer Alliance (GPQA) would not exist in its current iteration without support from Martin Kampmann, Shannon Noonan, Branden Barger, and Klint Jaramillo. GPQA would not have become a racially and gender diverse as well as trans-inclusive group without the

unrelenting advocacy of Iris D. Young, Adair Borges, and Zara Weinberg. Your commitment to queer and trans justice in academia and in society continues to inspire me.

I have had the great fortune to work on various DEIJ projects with numerous people. Thank you to Greyson Lewis for your contributions to the Diversity Network Initiative; to Ramiro Patiño and Melissa Spear for your collaboration on the Diversity and Allyship Breakfast panel, as well as on the [petition to SACNAS National](#); to Anna Lipkin, Jeff Nicklas, and Elizabeth Silva for conversations on transforming graduate admissions. Thank you to Jesse Holt for spending hours with me developing agendas, budgets, and more for the Full Circle symposium and being there for me when I didn't feel heard or appreciated. I want to express my deepest gratitude to Willow Coyote-Maestas for your friendship, your infectious optimism, your commitment to cultivating community for Native and Indigenous scientists, and for always being down to drink some mezcal and brainstorm how we can make academia a better place.

I don't consider myself someone who is good at networking, and yet I've found myself surrounded by a network of amazing friends. Thank you to Andreana S. Cunningham and Andrew Mudreac for your continued friendship and support over the past (almost) decade, even if we don't talk as often as we'd like to. Thank you to my original SF friend group – Sean Assanuvat, Juan Maciel, Damian Peter James Wolfe Klambauer, and Conor Carroll. Thank you to the friends I've made through Twitter, back when it was still a functioning website – Krisha Aghi, Mark Baxter, Matthew Cruz, William Clark, Lionel Rodriguez, Simon Clark, Benjamin Keisling, and Robert Nguyen Ulrich. I (probably) couldn't have made it through grad school without drinking – thank you to Chloe Miller and Kyle Greffin for making Junior a welcoming place for me to work in and always keeping me hydrated.

Tuesday – what a concept! I didn't realize how much joy, love, and friendship a weekly drag trivia show at Harvey's would inspire. Thank you to Polly Amber Ross for showcasing your art, building queer community, and always advocating for trans rights; to Myles Webb, Christian Nicoloso, Derek Boehringer, José Murillo, Jr., Dany Benitez, and Nic Candito for being a constant source of light and laughter when everything else was uncertain.

“What's the distance between those two atoms?” – Andrew K. Ecker. This simple question redefined my entire research trajectory and our relationship. I am so incredibly grateful for how much time, energy, and curiosity you've infused into my research. I would not have been able to finish this PhD without your scientific partnership in these final months.

I want to express an immense amount of gratitude to my research thesis advisor James Fraser for your willingness to confront difficult scientific and even more difficult social problems, for always encouraging me to try something out just to see what might happen, and for allowing me to pursue my passion of advancing diversity, equity, inclusion, and justice (DEIJ).

These words are insufficient to capture the impact that my unofficial thesis advisor D'Anne Duncan has had on me, but I must try anyway. Thank you for taking on a new project in 2017 when you told yourself no more news projects, for letting me express myself in an unfiltered and sometimes embarrassing way, for showing me how to approach people with care and respect without losing the essence of my message, for your constant advocacy of students and other trainees especially when no one else is advocating for us, and for your friendship throughout the years. I wouldn't have made it to this point without your support.

To my best friend Steven A. Cincotta, another person whose impact on my life cannot be captured in writing. Your childlike enthusiasm taught me that it's not only okay to be a little silly, but that life is way more enjoyable when you don't take yourself so seriously. Your

rigorous approach to science, despite all the obstacles thrown in front of you, inspires me to constantly seek new solutions to problems I'd rather leave unresolved. I can't express how much your support of my DEIJ work has meant to me. I would've lost my mind a long time ago had you not been around to anchor me down.

Mil gracias a mi madre por enseñarme que la cosa más importante en la vida es ser feliz. Si no te sientes feliz haciendo algo, no lo hagas más. Tu amor y orgullo de mi me hizo seguir adelante hasta este punto. Gracias por todo que me has hecho.

Pa, thank you for teaching me the value of an education, for reminding me that you can never be too generous, and for always believing in me. You never doubted me and for that I'm eternally grateful.

Lastly, I want to thank Boo the Cat and Timo for being the cutest cats I've ever had the pleasure of knowing and for keeping me company while I wrote this thesis. Your cuteness and affection were the motivation that I didn't know I needed.

Contributions

Chapter 1

The work in this chapter is published as Barad et al. in *Protein Science* (2020).

Chapter 2

The work in this chapter is published as Schuller, Correy, Gahbauer, Fearon, et al in *Science Advances* (2021).

Chapter 3

The work in this chapter is published as Díaz et al. in *Biorxiv* (2023).

Epigraph

“What kept me sane was knowing that things would change, and it was a question of keeping myself together until they did.”

— **Nina Simone**, *I Put a Spell on You: The Autobiography of Nina Simone*

“What we must do is commit ourselves to some future that can include each other and to work toward that future with the particular strengths of our individual identities. And in order for us to do this, we must allow each other our differences at the same time as we recognize our sameness.”

— **Audre Lorde**, *Learning from the 60s*

Structural and enzymatic characterization of pH-dependent chitinase activity, and contributions to Diversity, Equity, Inclusion, and Justice.

Roberto Efraín Díaz

Abstract

In my dissertation, I present a comprehensive investigation into chitinase enzymes and their diverse roles in biology, as well as contributions to antiviral drug development and social justice initiatives. The first chapter focuses on Acidic Mammalian Chitinase (AMCase), an enzyme responsible for degrading the resilient polysaccharide chitin in mammalian stomachs and lungs. Through the development of novel chitinase activity assays, we explore the effects of asthma-associated mutations, examine the contributions of individual enzyme domains in degrading crystalline chitin, and compare the behavior of AMCase with chitotriosidase, another chitinase in mammals. Additionally, we explore the challenges of engineering hyperactive chitinases, highlighting the limitations of traditional screening methods when assessing complex chitin substrates.

In the second chapter, I focus on the SARS-CoV-2 macrodomain (Mac1), a viral protein that antagonizes host antiviral signaling. Using computational and structural techniques, we identify numerous small molecules that bind to the active site of the macrodomain, providing a foundation for the development of potent SARS-CoV-2 macrodomain inhibitors. We validate these findings through solution binding assays, utilizing multiple biophysical techniques to confirm fragment hits.

In the third chapter, I delve into the catalytic mechanism underlying the pH-dependent activity profile of mouse Acidic Mammalian Chitinase (mAMCase). By employing a combination of biochemical, structural, and computational modeling approaches, I uncovered the ability of the mouse homolog to function effectively in both acidic and neutral environments. I determined the kinetic properties of mAMCase across a broad pH range and reveal its intriguing dual activity optima at pH 2 and 7. Through high-resolution crystal structures of mAMCase bound to chitin, I unveiled extensive conformational ligand heterogeneity, providing valuable insights into the catalytic mechanism of mAMCase. These results integrate structural, biochemical, and computational approaches to deliver a more complete understanding of the catalytic mechanism governing mAMCase activity at different pH.

In the fourth chapter, I outline my contributions to advancing diversity, equity, inclusion, and justice (DEIJ) at UCSF, other academic institutions, and beyond academia. This chapter not only recounts my experiences cultivating LGBTQ+ community, advocating for racial justice, and interrogating institutional policies regarding graduate admissions, but also renders visible otherwise neglected contributions to improving the social conditions in which we perform science.

Table of Contents

Chapter 1 Differences in the chitinolytic activity of mammalian chitinases on soluble and crystalline substrates.	1
Contributions	2
Abstract	3
Introduction	4
Results	7
<i>Engineering of hyperactive chitinases</i>	7
<i>Comparison of the activity of the catalytic domain of AMCase to the full length enzyme with new approaches</i>	8
<i>Small substrates can be misleading for engineered chitinases</i>	10
<i>Effects of human asthma-associated mutants in the mouse context</i>	10
<i>Comparison of acidic mammalian chitinase and chitotriosidase</i>	11
Discussion	13
Tables	17
Figures	22
Methods	28
<i>Protein preparation</i>	28
<i>Analysis of Kinetic Data</i>	29
<i>Continuous fluorescence measurements to quantify activity using commercial oligomeric substrates</i>	29
<i>Bulk clearance activity assay</i>	30
<i>Potassium ferricyanide reduction assay</i>	31
<i>Chito oligosaccharide oxidase coupled peroxidase assay</i>	31
<i>Random mutagenesis and screening</i>	32
References	34

Supplemental Figures	38
Acknowledgments	42
<i>Funding</i>	42
<i>Competing interests</i>	42
Chapter 2 Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking.	43
Contributions	46
Abstract	49
Introduction	50
Results	54
<i>Two crystal forms of Nsp3 Mac1 reveal differences in active site accessibility</i>	54
<i>Identifying new ligands for Nsp3 Mac1 using crystallographic fragment screening and docking</i>	55
<i>Characterization of experimental and virtual screening libraries</i>	55
<i>Hit rates and Mac1 interaction sites of fragments</i>	57
<i>Docking hits mimic the adenine recognition pattern</i>	59
<i>Analysis of key interactions between Mac1 and fragments from the crystallographic screens</i>	62
<i>Fragments binding to the adenine subsite</i>	62
<i>Fragments binding to the oxyanion subsite</i>	63
<i>Fragments binding to the catalytic and other potential allosteric sites</i>	64
<i>Fragment binding exploits protein conformational flexibility</i>	65
<i>Changes in water networks upon fragment binding</i>	67
<i>Solution binding of fragment hits</i>	68
<i>Opportunities for fragment linking and merging to optimize Mac1 inhibitors</i>	71
Discussion	72
Figures	76
Data availability	88

Methods	89
<i>Fragment libraries</i>	89
<i>C2 crystals at UCSF</i>	90
<i>Protein expression and purification</i>	90
<i>Crystallization</i>	91
<i>Crystal dehydration and fragment soaking</i>	92
<i>Lysine methylation</i>	93
<i>Crystallization of methylated Mac1</i>	94
<i>Ultra high resolution data collection, refinement and modelling</i>	94
<i>Data collection at physiological temperature, refinement and modelling</i>	96
<i>Fragment data collection, refinement and modelling</i>	97
<i>P43 crystals at UCSF</i>	99
<i>Protein expression and purification</i>	99
<i>Crystallization</i>	99
<i>Fragment and ADP-ribose soaking</i>	100
<i>Fragment data collection, processing, modelling, and refinement</i>	100
<i>P43 crystals at Oxford/XChem</i>	101
<i>Protein expression and purification</i>	101
<i>Crystallographic fragment screening</i>	102
<i>Molecular Docking Screens</i>	103
<i>Fragment linking and merging</i>	105
<i>Differential Scanning Fluorimetry (DSF)</i>	106
<i>Isothermal Titration Calorimetry (ITC)</i>	107
<i>Homogeneous Time-Resolved Fluorescence (HTRF)-based Peptide Displacement Assay</i>	107
References	109
Supplemental Information	118
<i>Purity and structure determination of fragments ZINC901381520, ZINC82473428 and ZINC89254160 from Enamine</i>	118

<i>NMR experiments for samples ZINC901391520, ZINC82473428 and ZINC89254160</i>	119
<i>Original sample ZINC901391520</i>	119
<i>Second batch sample ZINC901391520</i>	120
<i>Sample ZINC82473428</i>	120
<i>Sample ZINC89254160</i>	121
<i>Conclusions based on HPLC-MS and NMR characterization of samples ZINC901391520, ZINC82473428 and ZINC89254160</i>	122
<i>QCRG Structural Biology Consortium authorship</i>	123
Supplemental Figures	125
Acknowledgments	140
<i>General</i>	140
<i>Funding</i>	140
<i>Competing interests</i>	142
Chapter 3 Structural characterization of ligand binding and pH-specific enzymatic activity of mouse Acidic Mammalian Chitinase.	144
Contributions	146
Abstract	148
Introduction	149
Results	152
<i>New assay confirms broad pH profile for mAMCase</i>	152
<i>Characterization of mAMCase ligand occupancy and conformational heterogeneity.</i>	153
<i>Structural characterization of mAMCase catalytic triad D₁xD₂xE.</i>	156
<i>Theoretical pKa calculations of mAMCase catalytic triad D₁xD₂xE.</i>	159
<i>Molecular Dynamics</i>	161
Discussion	163
Tables	167

Figures	171
Data Availability	179
Methods	181
<i>Protein expression and purification</i>	<i>181</i>
<i>4MU-chitobioside Endpoint Assay</i>	<i>182</i>
<i>Analysis of kinetic data</i>	<i>183</i>
<i>Apo crystallization</i>	<i>184</i>
<i>Apo data collection, processing, and refinement at cryogenic temperature</i>	<i>185</i>
<i>Apo data collection, processing, and refinement at room temperature</i>	<i>186</i>
<i>Holo crystallization</i>	<i>186</i>
<i>Holo data collection, processing, and refinement at cryogenic temperature</i>	<i>187</i>
<i>Ligand modeling</i>	<i>187</i>
<i>Ringer analysis</i>	<i>188</i>
<i>pKa Analysis</i>	<i>188</i>
<i>Molecular Dynamics</i>	<i>189</i>
References	190
Supplemental Figures	196
Acknowledgments	209
<i>General</i>	<i>209</i>
<i>Funding</i>	<i>209</i>
<i>Competing interests</i>	<i>210</i>
Chapter 4 Contributions to Diversity, Equity, Inclusion, and Justice.	211
Introduction	212
Improving LGBTQ+ Inclusion in Academia	214
<i>Graduate and Postdoc Queer Alliance</i>	<i>214</i>
<i>LGBTQ+ Coffee Hour</i>	<i>214</i>

<i>Petition to SACNAS</i>	215
Improving Equity in Graduate Admissions	219
<i>Diversity Network Initiative</i>	219
Conclusion	223
References	225
Acknowledgments	226
<i>General</i>	226

List of Figures

Chapter 1 Differences in the chitinolytic activity of mammalian chitinases on soluble and crystalline substrates.

Figure 1.1 Engineering of hyperactive AMCase mutants.	22
Figure 1.2 Activity comparisons of AMCase catalytic domain and full length enzyme.	24
Figure 1.3 Engineered mutant activity with the novel chitooligosaccharide oxidase assay.	25
Figure 1.4 Comparison of activity of AMCase asthma-associated mutants.	26
Figure 1.5 Comparison of AMCase and Chitotriosidase.	27
Supplemental Figure 1.1 Data processing for 4MU assay.	38
Supplemental Figure 1.2 Data processing for colloidal chitin clearance assay.	39
Supplemental Figure 1.3 Data processing for ferricyanide reduction assay.	40
Supplemental Figure 1.4 Data processing for chitO assay.	41

Chapter 2 Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking.

Figure 2.1 Overview of the fragment discovery approach for SARS-CoV-2 Nsp3 Mac1 presented in this study.	76
Figure 2.2 Crystallographic screening identified 234 fragments bound to Mac1.	77
Figure 2.3 Docking hits confirmed by high resolution crystal structures.	79
Figure 2.4 Fragments binding to the adenine subsite.	80
Figure 2.5 Fragments binding to the oxyanion subsite.	81
Figure 2.7 Fragments targeting the catalytic and potential allosteric sites are sparsely populated compared to the adenosine site.	82

Figure 2.8 Experimentally observed conformational heterogeneity is sampled by various fragments.	84
Figure 2.9 Water networks in the active site are displaced as well as used by fragments for bridging interactions.	85
Figure 2.10 Biophysical corroboration of solution binding of crystallographic fragment hits by DSF, ITC and ADPr-peptide displacement assay.	86
Figure 2.11 Fragments bridging multiple adenosine sites provide direct merging opportunities.	87
Supplemental Figure 2.1 Ultra-high resolution features in Mac1 electron density maps.	125
Supplemental Figure 2.2 Comparison of isomorphism and DMSO tolerance of the C2 and P4 ₃ crystals.	127
Supplemental Figure 2.3 Crystal packing in Mac1 crystals determines active site accessibility.	129
Supplemental Figure 2.4 Structure and sequence comparison of Mac1 with related viral and human macrodomains.	130
Supplemental Figure 2.5 Physical properties, scaffold and chemotype analysis of screened fragment libraries.	132
Supplemental Figure 2.6 Overview of fragment binding to protomer A (white surface) and protomer B (blue surface) of the P4 ₃ crystals.	133
Supplemental Figure 2.7 Additional soaking hits from docking and adenine-N3 vs -N9-alkylated isomers.	134
Supplemental Figure 2.8 Mac1 subsites compared to the adenine binding subsite in kinases and the oxyanion binding site in carboxylesterases.	135

Supplemental Figure 2.9 Comparison of DSF, HTRF, and ITC results for compounds tested in all assays.	136
 Chapter 3 Structural characterization of ligand binding and pH-specific enzymatic activity of mouse Acidic Mammalian Chitinase.	
Figure 3.1 Kinetic properties of mAMCase catalytic domain at various pH.	172
Figure 3.2 Schematic representation of sugar-binding subsites in mAMCase.	173
Figure 3.3 Asp138 orientation correlates with ligand subsite occupancy.	174
Figure 3.4 pKa of GH18 chitinases in the D2 <i>inactive</i> and <i>active</i> conformation.	175
Figure 3.5 Distribution of distances observed every 10 ps of each simulation and their respective time courses.	176
Figure 3.6 Proposed model for ligand translocation towards the active site and ligand release post-catalysis.	178
Supplemental Figure 3.1 pH of reaction solution before and after quenching with 0.1 M Gly-NaOH pH 10.7.	196
Supplemental Figure 3.2 Kinetics of 4MU-chitobioside catalysis by mAMCase catalytic domain at various pH.	198
Supplemental Figure 3.3 96-well plate layout of crystallization conditions.	199
Supplemental Figure 3.4 pKa of apo and holo mAMCase in the D2 <i>inactive</i> and <i>active</i> conformation.	200
Supplemental Figure 3.5 Overview of key residues for mAMCase activity.	203
Supplemental Figure 3.6 Protein-ligand interactions between mAMCase and chitin.	204
Supplemental Figure 3.7 Ringer analysis of catalytic triad confirms alternative Asp138 conformations.	208

List of Tables

Chapter 1 Differences in the chitinolytic activity of mammalian chitinases on soluble and crystalline substrates.

Table 1.1 Measured rate constants for engineered mutants using 4MU-chitobioside assay.	17
Table 1.2 Calculated rate constants for AMCase catalytic domain and full length enzyme.	18
Table 1.3 Calculated rate constants for engineered mutants.	19
Table 1.4 Calculated rate constants for AMCase asthma-associated mutants.	20
Table 1.5 Calculated rate constants for AMCase and chitotriosidase.	21

Chapter 3 Structural characterization of ligand binding and pH-specific enzymatic activity of mouse Acidic Mammalian Chitinase.

Table 3.1 Data collection and refinement statistics.	167
Table 3.2 Occupancy of each ligand subsite and Asp138 in the <i>active</i> conformation.	169
Table 3.3 pKa across Asp136, Asp138, Glu140 of mAMCase structures in either Asp138 <i>inactive</i> or Asp138 <i>active</i> conformation.	170

Chapter 1 Differences in the chitinolytic activity of mammalian chitinases on soluble and crystalline substrates.

Benjamin A. Barad^{1,2}, Lin Liu¹, Roberto Efraín Díaz^{1,3}, Ralph Basilio^{1,4}, Steven J. Van Dyken⁵,
Richard M. Locksley^{6,7,8}, James S. Fraser^{1*}

Affiliations

¹ Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, USA

² Biophysics Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

³ Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

⁴ Science Education Partnership High School Intern Program, University of California, San Francisco, San Francisco, CA 94158, USA

⁵ Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

⁶ Department of Medicine, University of California, San Francisco, California 94143, USA

⁷ Department of Microbiology and Immunology, University of California, San Francisco, California 94143, USA

⁸ University of California, San Francisco, Howard Hughes Medical Institute, San Francisco, California 94143, USA

Contributions

The work in this chapter is published as Barad et al. in [*Protein Science*](#) (2020). B.A.B. designed and carried out all kinetics experiments; developed figures; and wrote and edited the manuscript. L.L. and R.B. helped with protein expression and data collection. B.A.B. and R.E.D. developed the colloidal chitin assay and performed those experiments. S.J.V.D. and R.M.L provided critical feedback and edited the manuscript. J.S.F. supervised the project and wrote and edited the manuscript.

Abstract

Chitin is an abundant polysaccharide used by many organisms for structural rigidity and water repulsion. As such, the insoluble crystalline structure of chitin poses significant challenges for enzymatic degradation. Acidic mammalian chitinase, a processive glycosyl hydrolase, is the primary enzyme involved in the degradation of environmental chitin in mammalian lungs. Mutations to acidic mammalian chitinase have been associated with asthma, and genetic deletion in mice increases morbidity and mortality with age. We initially set out to reverse this phenotype by engineering hyperactive acidic mammalian chitinase variants. Using a screening approach with commercial fluorogenic substrates, we identified mutations with consistent increases in activity. To determine whether the activity increases observed were consistent with more biologically relevant chitin substrates, we developed new assays to quantify chitinase activity with insoluble chitin, and identified a one-pot fluorogenic assay that is sufficiently sensitive to quantify changes to activity due to the addition or removal of a carbohydrate-binding domain. We show that the activity increases from our directed evolution screen were lost when insoluble substrates were used. In contrast, naturally occurring gain-of-function mutations gave similar results with oligomeric and insoluble substrates. We also show that activity differences between acidic mammalian chitinase and chitotriosidase are reduced with insoluble substrate, suggesting that previously reported activity differences with oligomeric substrates may have been driven by differential substrate specificity. These results highlight the need for assays against physiological substrates when engineering metabolic enzymes and provide a new one-pot assay that may prove to be broadly applicable to engineering glycosyl hydrolases.

Introduction

Chitin is a ubiquitous polysaccharide, comprised of β -1,4-linked N-acetylglucosamine, that is produced by fungi and arthropods for structural rigidity and water repulsion^{1,2}. With notable exceptions³, vertebrates generally do not produce chitin. However, mammals have a conserved machinery to recognize and degrade environmental chitin that is inhaled or ingested, and this machinery is tied to an innate immune response to chitin^{4,5}.

Chitin polymers assemble into water-insoluble microcrystals, which have been observed in three different crystal forms, differentiated by the parallel or antiparallel orientation of neighboring chitin strands⁶. Alpha-chitin, the most common conformation, forms antiparallel sheets that intercalate the N-acetyl groups of neighboring polymers and form tight hydrogen bonding networks⁷. Strands of chitin must be extracted from this highly crystalline structure to be degraded, and the rate limiting step of catalysis has been observed to be the processive decrystallization of additional substrate from the bulk crystal^{8,9}. This observation makes it particularly challenging to effectively associate degradation of short oligomeric analogues with true catalytic efficacy. The insolubility and recalcitrance of bulk chitin also makes it a particularly challenging substrate to quantify with high precision. Recently, several new methods have tackled this problem by using labelled chitin substrates with gel electrophoresis^{10,11} as well as enzyme-coupled assays to generate colorimetric signal from reducing ends¹². These methods have enabled new insights into chitinase behavior, but their signal-to-noise ratio and throughput limit the ability to separate total activity into binding and catalysis, as well as other components of polysaccharide catabolism such as substrate specificity and processivity.

The molecular mechanism of recognition of chitin and the signaling program generates in mammals is not well understood, but breakdown of inhaled chitin is accomplished by the

secreted enzymes acidic mammalian chitinase (AMCase) and chitotriosidase, which are conserved across mammals⁴. Both are two domain family-18 glycosyl hydrolases consisting of a catalytic TIM-barrel domain and a C-terminal carbohydrate-binding domain. In AMCase, the two domains are connected by a 25 residue glycine- and serine-rich linker that is expected to be highly glycosylated, while chitotriosidase has a shorter, proline-rich linker that has also been found to be glycosylated¹³⁻¹⁵. The roles of the linker and the C-terminal carbohydrate-binding domain in processing chitin have not been quantified.

AMCase is upregulated in response to chitin insult and is secreted into the airway lumen, where it interacts with crystalline chitin and breaks down the substrate¹⁶. Consistent with the reported role of AMCase in asthma, there are polymorphisms of human AMCase (hAMCase) that increase its activity and have been associated previously with asthma protection¹⁷. A trio of mutations found far from the active site in the catalytic domain (15-20 Å from active site inhibitor) in humans, N45D, D47N, and R61M, which change residues to the wild type identities of mouse AMCase (mAMCase), has been previously described to increase specific activity against model substrates¹⁸. Of these mutations, prior work has identified the R61M mutation as causing the largest increase in total activity, as well as the largest decrease in mice with the reverse M61R mutation¹¹. The mechanism by which these mutations alter binding and catalysis remains unclear. AMCase deficient mice accumulate chitin in their lungs and develop tissue fibrosis as an aging phenotype; external addition of recombinant chitinase to the airway reduces this phenotype¹⁹. This suggests that AMCase is predominantly responsible for clearance of chitin from airways, and further suggests that enhancing AMCase activity may reduce chitin airway levels.

In this study, we tried to evolve variants of AMCase that would have enhanced activity to test the hypothesis that enhanced clearance of chitin would reduce the potential for age-related lung fibrosis. Our directed evolution approach was based on simple fluorogenic substrates. We found mutations that dramatically increase the activity of the enzyme by both improving binding and catalysis. We developed new approaches to quantifying bulk chitin degradation and discovered that these engineered mutations did not have the same effect with bulk substrates. We used these improved methods to assay the impact of the carbohydrate-binding domain on activity and discover that it causes a minor K_M versus k_{cat} tradeoff but does not have a major effect on overall activity. We reverted the asthma-protective mutants in the mouse background and find that the dominant effect is a k_{cat} decrease from the M61R mutation. We also compared the activity of mAMCase and chitotriosidase with different small oligomeric substrates and with bulk chitin. These results highlight the need for assays against more physiological substrates when engineering complex metabolic enzymes and provide a fluorogenic one-pot reducing sugar assay that may be broadly applicable to engineering glycosyl hydrolases using realistic substrates.

Results

Engineering of hyperactive chitinases

Recent efforts have identified recombinant chitinase as a potential direct therapy to ameliorate inflammatory lung symptoms that arise when native chitinase activity is compromised¹⁹. To investigate whether we could improve the activity of mouse AMCase, we used error-prone PCR to generate libraries of mAMCase mutants (**Figure 1.1A**). Our recombinant expression approach, utilizing periplasmic secretion as described previously, also yields enzyme secreted into the media²⁰. We therefore assayed, in 96 well format, the ability of the spent media of individual mutants after protein expression to cleave 4MU-chitobioside (**Figure 1.1A**). Comparing these results to both wild-type and engineered catalytically dead mutants, we found that while most mutations resulted in either total loss of protein activity or similar activity to wild-type, a small number of mutants were much more active than the wild-type (**Figure 1.1B**). Because these assays were done directly on spent media, the measured activity for each well reports on the combination of the specific activity of the enzyme, expression level, and secretion efficiency. To determine whether our results represented improvements in activity, we isolated and purified the two most active mutants: A239T/L364Q (**Figure 1.1D**, pink) was the most active mutant identified, with a 5-fold improvement in activity, and V246A (**Figure 1.1D**, orange), which showed a 2-fold improvement in activity.

After purification, we measured the specific activity of the assay using a one-pot continuous-read fluorescent assay based on the previously developed enzyme-coupled assay¹² and replicated the improvements observed in the unpurified screening format (**Figure 1.1C**). Both mutants improved significantly in k_{cat} , while the A239T/L364Q had a nonsignificant improvement in K_m (**Table 1.1**). Structurally, the V246A mutation may have a second-shell

interaction stabilizing the active conformation, while L364Q is positioned at the binding site for chitin and may directly improve chitin hydrolysis (**Figure 1.1D**).

Comparison of the activity of the catalytic domain of AMCase to the full length enzyme with new approaches

An alternative hypothesis for the increased activity is that the engineered variants have high specificity for the fluorophore or a smaller oligomer. This motivated us to develop new assays on larger and more complex chitin material. As a first control, we first assessed the contribution of the catalytic and carbohydrate-binding domain of AMCase. Due to its small oligomeric size, hydrolysis of the 4MU substrate is likely to be driven only by local interactions in the catalytic domain and the presence of the carbohydrate-binding domain should not affect the reaction rate. In contrast, the carbohydrate-binding domain has been hypothesized to play a role in binding crystalline chitin^{21,22}.

We expressed and purified the isolated catalytic domain of AMCase, as well as the full length enzyme, using an *E. coli* periplasmic expression approach²⁰. We first measured the ability of the enzyme to catalyze the breakdown of 4-methylumbelliferone (4MU) conjugated chitobioside, using a continuous read approach at pH 7.0. The activities of the two constructs were indistinguishable, either in binding or catalysis (**Figure 1.2A, Table 1.2, $p = .3$**).

We next tested different methods of quantifying hydrolysis of insoluble chitin. We used colloidal chitin substrates, which are more uniform in size and shape and to have reduced settling times compared to other substrates such as shrimp shell chitin. We first attempted to measure colloidal chitin hydrolysis by the disappearance of scattering by solid substrate as it is converted into small oligomeric products. We could not distinguish a statistically significant difference

between the two variants with this approach, which was likely limited by the relatively small dynamic range and large amount of enzyme required to produce a measurable change in scattering (**Figure 1.2B, Table 1.2**, $p = .8$). Each hydrolysis event only minimally alters the scattering of chitin crystals, and many cuts are likely necessary to solubilize crystals.

We next attempted to quantify the production of soluble reducing ends, which we hypothesized would more sensitively report individual catalytic events. The first method we used to assay production of soluble reducing ends was a ferricyanide reduction assay²³: after incubating colloidal chitin with AMCase at 37°C for up to 18 hr, we quenched the reaction and quantified the nonenzymatic reaction of soluble reducing sugars with potassium ferricyanide, read out by the disappearance of absorbance at 420 nm. With this assay, we were not able to identify a significant difference in total activity but were able to identify that the inclusion of the carbohydrate-binding domain created a small improvement in K_M that was offset by a reduction in the k_{cat} of AMCase (**Figure 1.2C, Table 1.2**, $p = .2$). This tradeoff did not result in a large difference in activity. Moreover, the endpoint-based requirements of the assay and of the dynamic range available in measuring reduction in absorbance were limiting. We next developed a new assay based on previous work using chitooligosaccharide oxidase (chitO) in combination with horseradish peroxidase to generate signal specifically from the production of chitin reducing ends¹². To convert this assay from endpoint to continuous readout, we took advantage of fluorogenic substrates for horseradish peroxidase and carefully washed the colloidal chitin to enable signal measurement without removal of the insoluble component. This gain-of-signal fluorescent assay had much improved signal-to-noise and sensitivity, and improved quantification of the kinetic parameters of chitinase activity. Using this assay, we were able to more confidently determine the tradeoff between improved binding ($p = .02$) and loss of maximal

catalytic activity ($p = .009$) with the inclusion of the carbohydrate-binding domain, which resulted in no significant change in total activity (**Figure 1.2D**, **Table 1.2**, $p = .8$).

Small substrates can be misleading for engineered chitinases

Having this new assay in hand, we tested whether the activity increases observed with the 4MU-chitobioside mutant resulted in similar improvements to degradation of bulk chitin. Using purified protein, we measured the activity of the mutants to degrade colloidal chitin using the enzyme-coupled chitO assay, and discovered that the A239T/L364Q mutant had lost all measurable activity, while the V246A mutant was not statistically significantly more active than the wild type (**Figure 1.3**, **Table 1.3**). The loss of activity of the double mutant suggests that the improvements were driven by the L364Q mutation interacting with the 4MU fluorophore, which can be rationalized structurally (**Figure 1.1D**). The stark difference in results between the results with the 4MU and chitO assay underscores the need for assays of catabolism of bulk chitin substrate, even during the initial stages of screening.

Effects of human asthma-associated mutants in the mouse context

Motivated by the result on the engineered mutations, we wanted to test naturally occurring mutations that have previously been shown to have different activities using the 4MU assay. We focused on a trio of mutations in AMC_{ase} in humans, N45D, D47N, and R61M, that confer significantly increased activity to AMC_{ase}^{11,18}. In all three cases, the identity of the mutated residues becomes the same as the identity of the residues of the mouse wild-type protein. To better understand the mutational landscape between the mouse and human enzymes, which have 81% sequence identity and differ by 92 total polymorphisms, we made the reverse

mutations in the mouse background to quantify their effect on activity using both 4MU-chitobioside and bulk chitin. First, we measured the activity of the mutations using 4MU-chitobioside, which showed that the mouse wild-type residues were more active than the human wild-type residues. The activity difference between wild-type and the M61R mutant was caused by a decrease in k_{cat} and a small increase in K_M (**Figure 1.4A, Table 1.4, $p = .03$**). Smaller effects were observed for the individual D45N and N47D mutations, but the effects were reversed by the charge swapped D45N/N47D construct. The full triple mutant was the least active ($p = .001$). These results show strong alignment with previous results in the human background¹¹ and suggest that the different residue identities have very similar effects in the mouse and human backgrounds. To understand whether these effects observed with the oligomeric substrate are relevant to enzyme activity on bulk chitin, we assayed the activity of humanizing mutations in mAMCase using the enzyme-coupled chitO assay. The results were similar to those using the 4MU substrate, with the largest effect of any individual mutation and the majority of the effect of the triple mutation contributed by the M61R mutant (**Figure 1.4B, Table 1.4, $p = .02$**). The effects of the D45N and N47D mutations were less pronounced in the chitO assay, while the M61R mutation had a similar effect on both K_M and k_{cat} .

Comparison of acidic mammalian chitinase and chitotriosidase

Next, we wanted to compare AMCase to the other major human chitinase, Chitotriosidase. Both enzymes are expressed in lungs, but only acidic mammalian chitinase is strongly overexpressed in response to chitin insult²⁴. Previous reports using the 4MU assay have indicated activity differences and no synergistic effects²⁵, but this result is convolved with the

substrate specificity of dimer and trimer chitin oligomers. Whether AMCase and chitotriosidase have similar activity on crystalline substrates has not been previously examined.

We sought to understand how binding, substrate specificity, and hydrolytic activity differed between the two enzymes. We investigated substrate specificity by comparing the ability of each enzyme to cleave the terminal glycosidic linkage on 4MU-chitobioside and 4MU-chitotrioside, representing hydrolysis in different substrate binding poses to generate chitobiose versus chitotriose as a substrate. When assayed the 4MU-chitobioside substrate, AMCase had more than double the activity of chitotriosidase, driven by a significant difference in K_M (**Figure 1.5A, Table 1.5**, $p = .01$). In contrast, the 4MU-chitotrioside substrate led to tighter binding for both AMCase and chitotriosidase, but the difference was much larger with chitotriosidase, leading to a smaller gap in activity between the two enzymes (**Figure 1.5B, Table 1.5**). The reduction in observed k_{cat} for both enzymes was likely driven by the alternative, nonfluorogenic reaction trajectory in which the 4MU-chitotrioside is cleaved into chitobiose and 4MU-bound N-acetylglucosamine, leading to a systematic underestimate of k_{cat} . The difference in the K_M suggests that chitotriosidase benefits more from the extended binding interactions available with the larger 4MU-chitotrioside substrate. We next assayed the differences in activity with a bulk substrate using the chitO-coupled assay. The difference in activity was much smaller in this assay, with the majority of the activity difference being driven by k_{cat} differences (**Figure 1.5C, Table 1.5**, $p = .006$). These results further confirm that much of the apparent activity differences between AMCase and chitotriosidase are due to differential substrate specificity, as had been previously described²⁵, and suggest that much of this difference can be attributed to differential binding efficiency for short chitin oligomers.

Discussion

Broadly, these results demonstrate the value of quantifying chitinase kinetics with bulk substrates with the same care used with model substrates (fluorogenic oligomers). Our results suggest that the effectiveness and sensitivity of the one-pot chitooligosaccharide oxidase coupled assay makes it an ideal approach for monitoring chitinase activity. While in some cases, the results of the activity assays closely resembled the 4MU-chitobioside assays, in others, the activities were tremendously different, underscoring the need for quantitative measures of bulk chitin catabolism. This proved to be particularly true for studies of the effects of multiple domains, which necessarily cannot bind the same short oligomer the same way they could a chitin crystal, as well as for engineered variants, in which screening with short fluorogenic substrates led to artifacts that may be related to fluorophore binding. The sensitivity and throughput available with the chitO-coupled assay enables more precise and quantitative measurements of bulk chitin catabolism than was previously available, and we expect that this technique will be effective for deconstructing different aspects of enzyme activity.

In contrast to the majority of cases, which had reasonable agreement between the bulk experiments and the small oligomers, our efforts to engineer hyperactive chitinases were limited by the use of the 4MU-chitobioside substrate as a screening tool. Our best mutants from screening had significant increases in activity, but once the purified mutants were assayed by the chitO assay, the improvements were not present. In the case of the A239T/L364Q mutant, there was no quantifiable activity with bulk substrate. The classic maxim is that “in protein engineering you get what you screen for”, and in this case that was maximizing binding efficiency for the 4-methylumbelliferone fluorophore and the chitin dimer. The result underscores the need in the future for utilizing frequent counter-screening with bulk chitin when

performing selection experiments for chitin processing and matches well with previous results in engineering cellulases, which showed that screening with synthetic substrates had significant pitfalls compared to using insoluble substrates.²⁶ One challenge to accomplishing this is that, while the chitO assay is more sensitive and higher throughput than previous techniques, it is sensitive to free sugars and other components of the media that limits its utility for direct screening. With small scale purification, we may in the future be able to directly screen activity of mutants using the chitO method. In combination with recent advances in guiding small library directed evolution with machine learning²⁷, we may be able to effectively use this approach to find hyperactivating mutants without the requirement of using chitobioside substrates.

With the exception of the engineered mutants, the kinetic parameters measured with the 4MU, and bulk chitin assays were well aligned, with k_{cat} values that were remarkably similar, suggesting that the 4MU assay effectively captures the chemical step of hydrolysis, and K_M values that were on the order of 30 μM for the 4MU-chitobioside and 0.03% w/v for the bulk chitin assay. Under the approximation of infinite polymer length, there is one binding site per N-acetylglucosamine unit. Each chitin monomer unit has a molecular mass of 203.21 g/mol, so 0.03% w/v or 0.3 g/L would correspond to approximately 1.5 mM, 50 times greater than the K_M observed for the small oligomeric substrates. We hypothesize that the higher effective K_M reports on the relative crystallinity of the chitin, with a small proportion of theoretical substrate binding sites being accessible to the enzyme. In the future, it may be possible to alter this crystallinity, using partial deacetylation, coapplication of chitin-binding enzymes that might loosen the crystalline geometry, or physical milling to alter the surface area to volume ratio.

Using the new bulk activity measurements, we were able to discern a tradeoff between k_{cat} and K_M with the addition of the carbohydrate-binding domain of AMCase, as K_M improved

from $0.0333\% \pm 0.0056\%$ to $0.0172\% \pm 0.0049\%$ chitin w/v ($p = .02$), while k_{cat} decreased from 0.944 ± 0.111 1/s to 0.540 ± 0.083 1/s ($p = .007$). While the improved binding with the addition of additional binding sites for chitin is unsurprising, the difference in the k_{cat} is less clear.

Previous work has suggested that, for some chitinases, the rate limiting step in bulk catalysis is processivity.⁹ This result supports that hypothesis for AMCase as well, since the additional binding motif may inhibit the ability of the catalytic domain to effectively slide to new binding sites. If AMCase processivity proves to be rate limiting, given the closely matched k_{cat} for 4MU-chitobioside, with which processivity is not possible, and bulk chitin, it suggests that the rates of catalysis and processivity may be very similar in the mouse enzyme. This may be a result of selection optimizing the overall rate of the enzyme or the relative size of products generated by the enzyme. For example, larger oligomers could be produced if decrystallization and sliding were much faster than the rate of hydrolysis. These larger oligomers may be the relevant molecules sensed by the mammalian immune system, as seen in plants²⁸. The carbohydrate-binding domain may further impact other aspects of catalysis, such as selecting specific chitin local morphology, binding chitin in the correct orientation, modulating processivity, or releasing when strands of chitin become too short to further process. Additionally, the assayed constructs lack posttranslational modifications. Acidic mammalian chitinase is predicted to have multiple O-linked glycosylation sites in the linker between the catalytic domain and the carbohydrate-binding domain¹⁵, which may have significant effects on interactions with crystalline substrates. The methods developed here can give information about binding and catalysis with relevant substrates, but questions still remain about processivity, endo versus exo preference, and potential clustering and cooperative behavior between multiple enzymes. One avenue to more fully characterize these aspects of catalysis will be single-molecule measurements of kinetics.

Recently, significant progress has been made in measuring chitinase activities by single-molecule microscopy^{9,29,30}, and applying this approach to mammalian chitinases, ideally with native glycosylation, may help to break down the effects of different mutations on activity, give new insights into the function of the carbohydrate-binding domain, and help to differentiate the enzymatic role of chitotriosidase and acidic mammalian chitinase.

Tables

Table 1.1 | Measured rate constants for engineered mutants using 4MU-chitobioside assay.

Note: k_{cat} values are reported in units of 1/s. K_M values are reported in units of mM for 4MU assays. Fold changes are relative to wild-type enzyme.

	k_{cat}	Fold change (p -value)	K_M	Fold change (p -value)
WT	1.5 ± 0.3	N/A	33 ± 12	N/A
A239T/L364Q	4.5 ± 1.2	3.0 ($p = .014$)	19 ± 9	0.58 ($p = .18$)
V246A	3.6 ± 0.7	1.4 ($p = .009$)	32 ± 12	0.97 ($p = .92$)

Table 1.2 | Calculated rate constants for AMCase catalytic domain and full length enzyme.

Note: k_{cat} values are reported in units of 1/s. K_M values are reported in units of mM for 4MU assays and % w/v for colloidal clearance, ferricyanide, and chitO assays.

	Catalytic Domain		Full Length Enzyme	
	k_{cat}	K_M	k_{cat}	K_M
4MU-chitobioside	1.12 ± 0.09	28 ± 3	1.05 ± 0.07	25 ± 2
Colloidal Clearance	0.00140 ± 0.00008	0.09 ± 0.03	0.00106 ± 0.00002	0.07 ± 0.02
Ferricyanide	0.454 ± 0.042	0.046 ± 0.02	0.39 ± 0.02	0.029 ± 0.007
ChitO	0.9 ± 0.1	0.033 ± 0.006	0.54 ± 0.08	0.017 ± 0.005

Table 1.3 | Calculated rate constants for engineered mutants.

Note: k_{cat} values are reported in units of 1/s. K_M values are reported in units % w/v. Fold changes are relative to wild type enzyme.

	k_{cat}	Fold change (p value)	K_M	Fold change (p value)
WT	0.78 ± 0.05	N/A	0.030 ± 0.002	N/A
A239T/L364Q	N.D.	N/A	N.D.	N/A
V246A	0.80 ± 0.07	1.03 ($p = 0.7$)	0.035 ± 0.004	1.17 ($p = 0.1$)

Table 1.4 | Calculated rate constants for AMCase asthma-associated mutants.

Note: k_{cat} values are reported in units of 1/s. K_M values are reported in units of mM for 4MU assays and % w/v for chitO assays.

	4MU-chitobioside		Chitooligosaccharide Oxidase	
	k_{cat}	K_M	k_{cat}	K_M
WT	1.1 ± 0.10	30 ± 3	1.0 ± 0.2	0.032 ± 0.008
D45N	0.94 ± 0.07	35 ± 6	0.8 ± 0.1	0.032 ± 0.008
N47D	0.71 ± 0.05	35 ± 6	0.74 ± 0.03	0.032 ± 0.008
M61R	0.8 ± 0.2	60 ± 12	0.48 ± 0.05	0.041 ± 0.005
D45N/N47D	0.98 ± 0.09	37 ± 5	1.02 ± 0.04	0.027 ± 0.003
D45N/N47D/M61R	0.46 ± 0.05	50 ± 7	0.31 ± 0.04	0.047 ± 0.006

Table 1.5 | Calculated rate constants for AMCase and chitotriosidase.

Note: k_{cat} values are reported in units of 1/s. K_{M} values are reported in units of mM for 4MU assays and % w/v for chitO assays.

	AMCase		Chitotriosidase	
	k_{cat}	K_{M}	k_{cat}	K_{M}
4MU-chitobioside	1.02 ± 0.05	28 ± 4	1.1 ± 0.1	68 ± 15
4MU-chitotrioside	0.5 ± 0.1	24 ± 11	0.33 ± 0.03	25 ± 4
ChitO	1.06 ± 0.06	0.018 ± 0.002	0.82 ± 0.05	0.016 ± 0.003

Figures

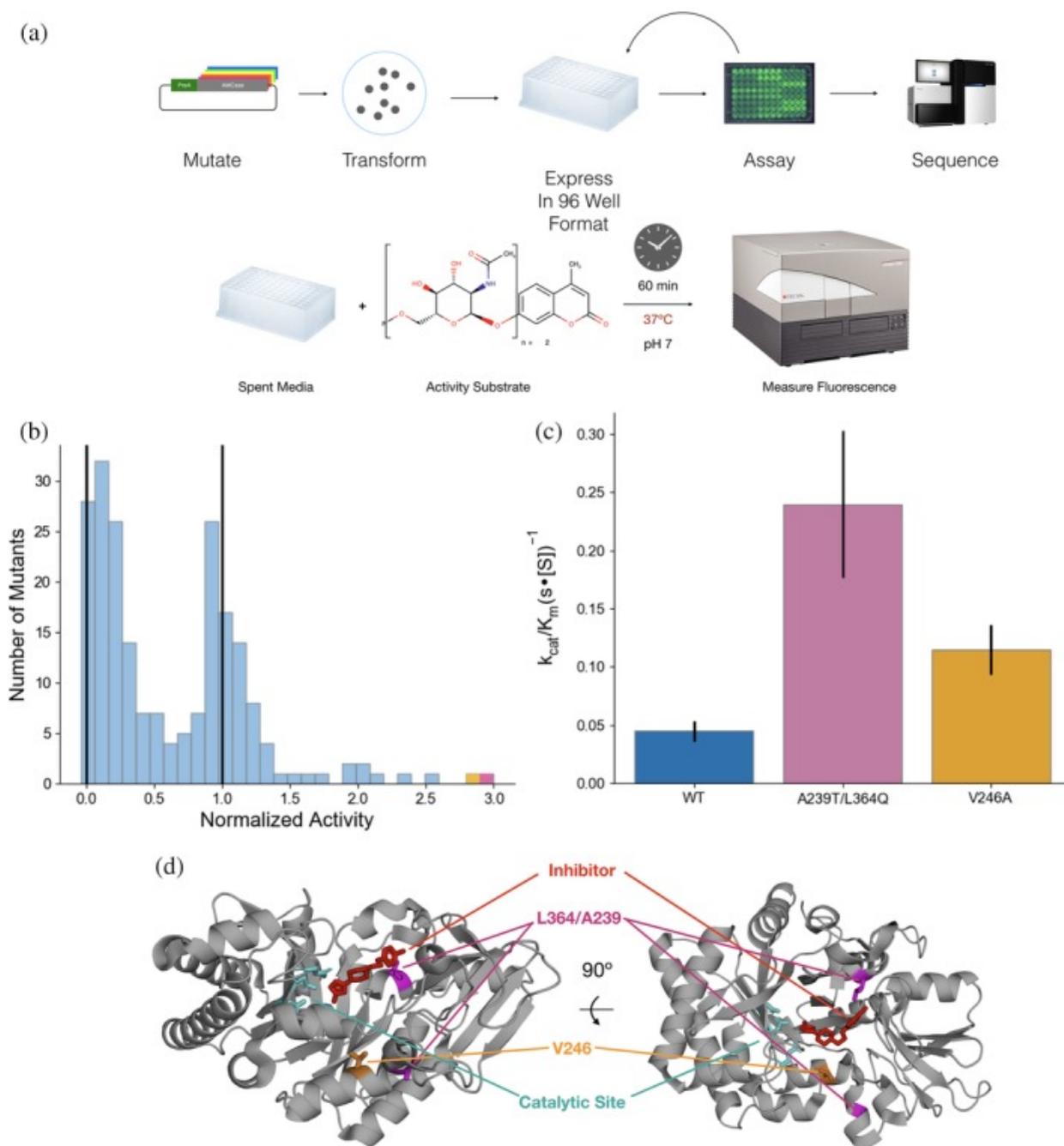


Figure 1.1 | Engineering of hyperactive AMCase mutants.

A) Workflow for directed evolution of AMCase. Mutants of AMCase were generated via error-prone PCR, then transformed and grown out from individual colonies in 96-well blocks. After expression, activity was measured using the 4MU-chitobioside substrate incubated with the expression media. **B)** Distribution of activity for mutants with 1–3 mutations per construct. Vertical lines at 0 and 1 represent a catalytically dead negative control and a wild type positive control, respectively. The best two results are highlighted in purple and orange. **C)** k_{cat}/K_M of purified hyperactive mutants using the 4MU-chitobioside assay. **D)** Structure of AMCase

catalytic domain (PDB: 3RM9) highlighting A239T/L364Q (pink) and V246A (orange). The active site catalytic network is highlighted in teal, and an inhibitor (5-(4-(2-[4-bromophenoxy]ethyl)piperazine-1-yl)-1H-1,2,4-triazol-3-amine)³¹ that binds to the active site cleft is shown in red.

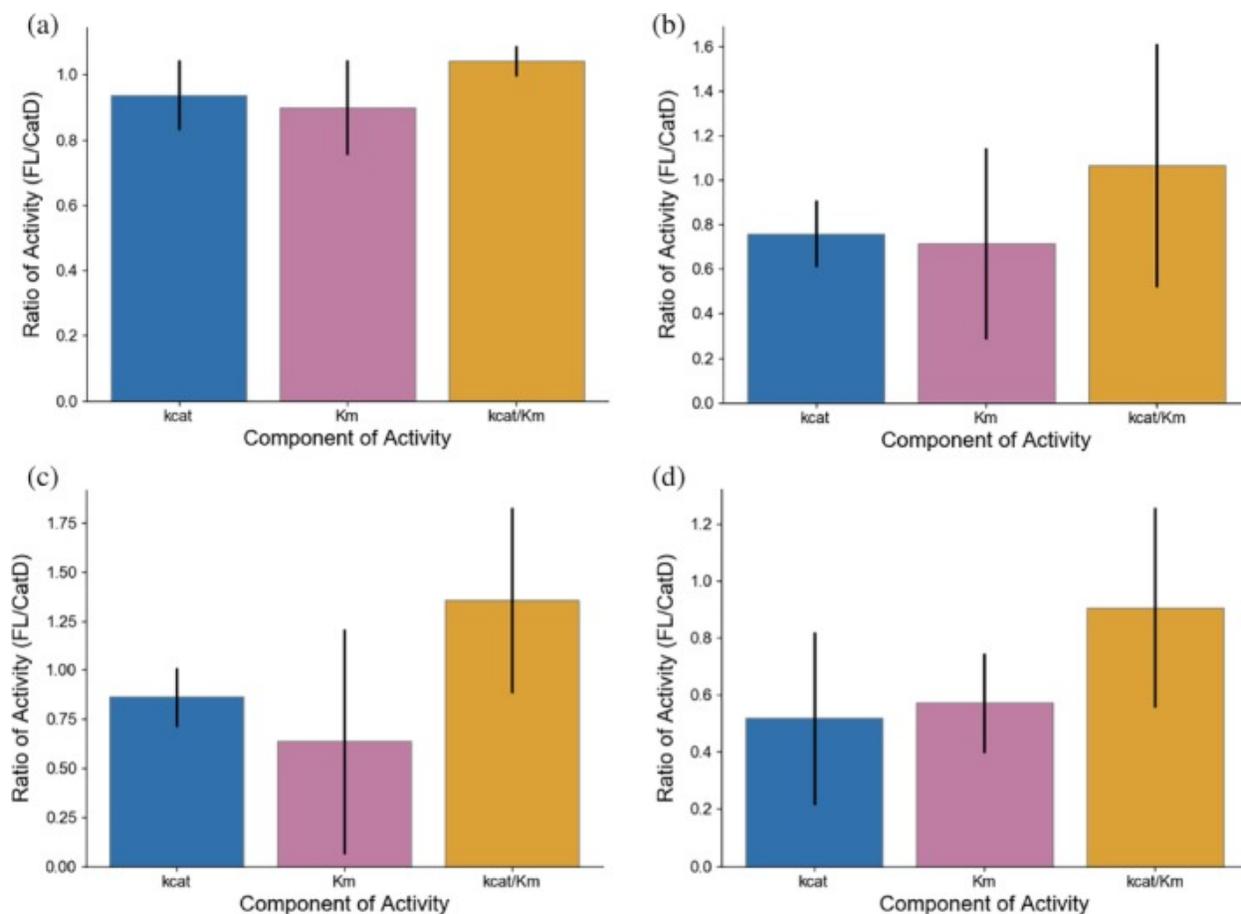


Figure 1.2 | Activity comparisons of AMCase catalytic domain and full length enzyme. Difference in k_{cat} , K_M , and k_{cat}/K_M of AMCase catalytic domain and full length enzyme generated via **A)** 4MU-chitobioside assay, **B)** colloidal chitin clearance assay, **C)** reducing sugar generation assay quantified with potassium ferricyanide, **D)** chitooligosaccharide oxidase coupled peroxidase assay. Error bars denote propagated SD of fit (accounting for covariance)

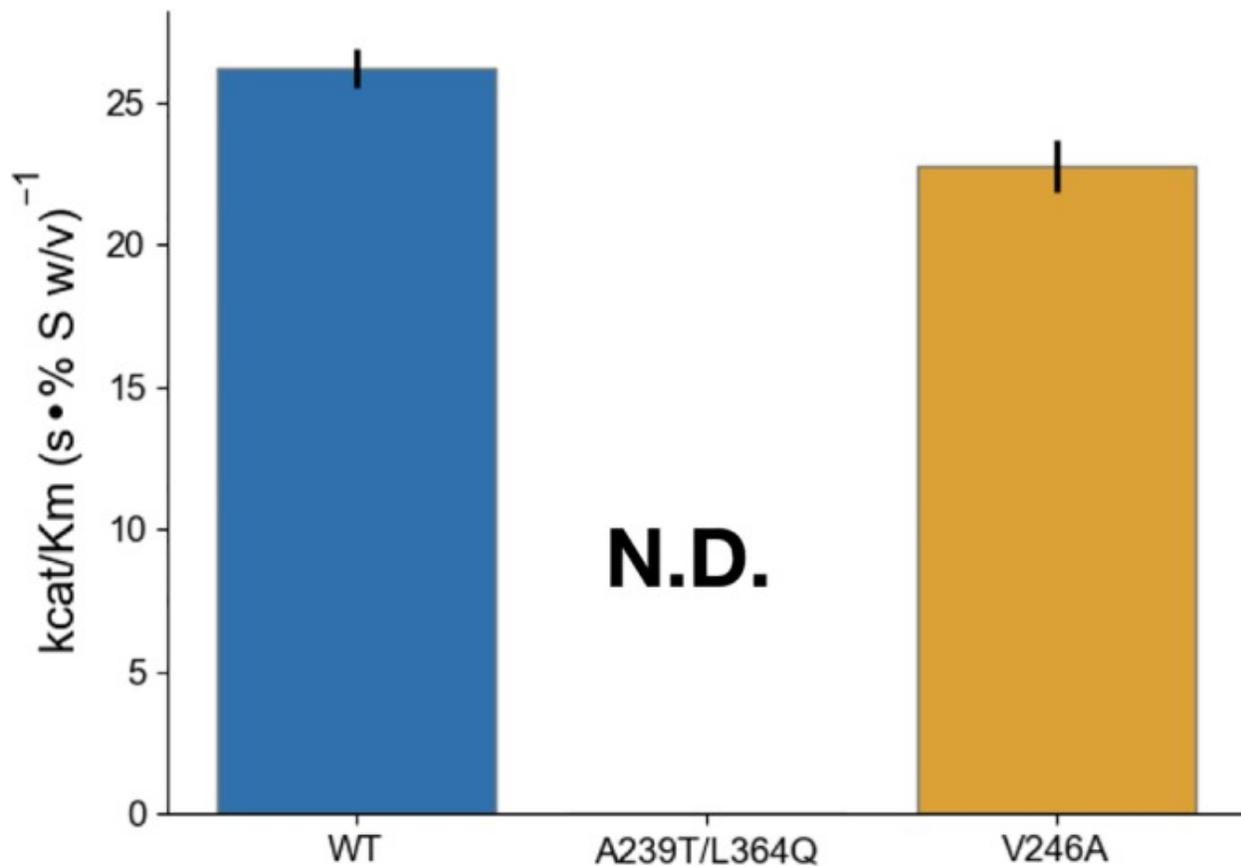


Figure 1.3 | Engineered mutant activity with the novel chito oligosaccharide oxidase assay. Difference in k_{cat}/K_M of purified hyperactive mutants using the 4MU-chitobioside assay. k_{cat} values are reported in units of 1/s. K_M values are reported in units % w/v. Error bars denote propagated *SD* of fit (accounting for covariance). The A239T/L364Q mutant had too little total activity to measure k_{cat} or K_M .

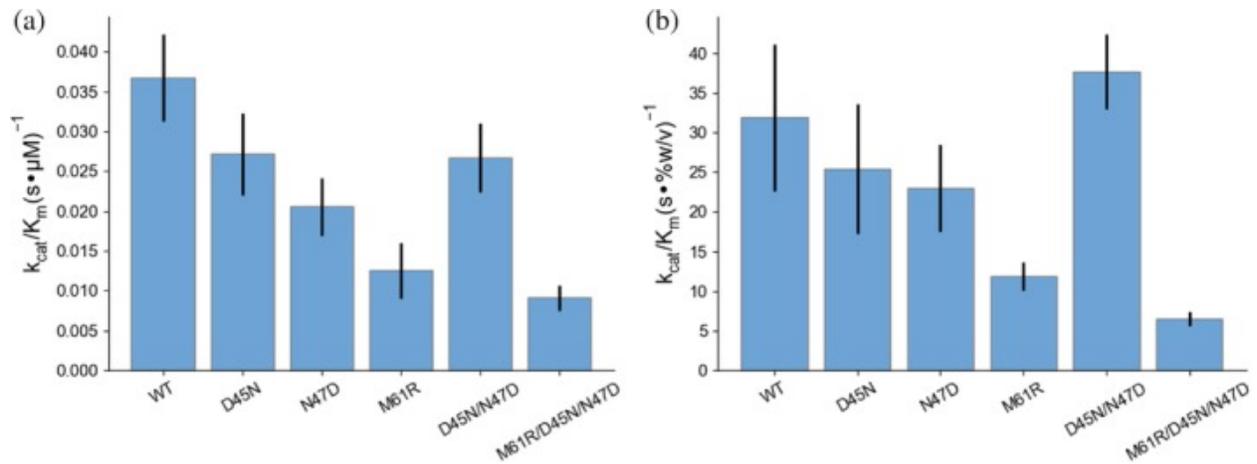


Figure 1.4 | Comparison of activity of AMCCase asthma-associated mutants.

Measurement of k_{cat}/K_M for reversed asthma-associated mutants in the mouse background using the **A)** 4MU-chitobioside and **B)** chitO assays. Error bars denote propagated *SD* of fit (accounting for covariance)

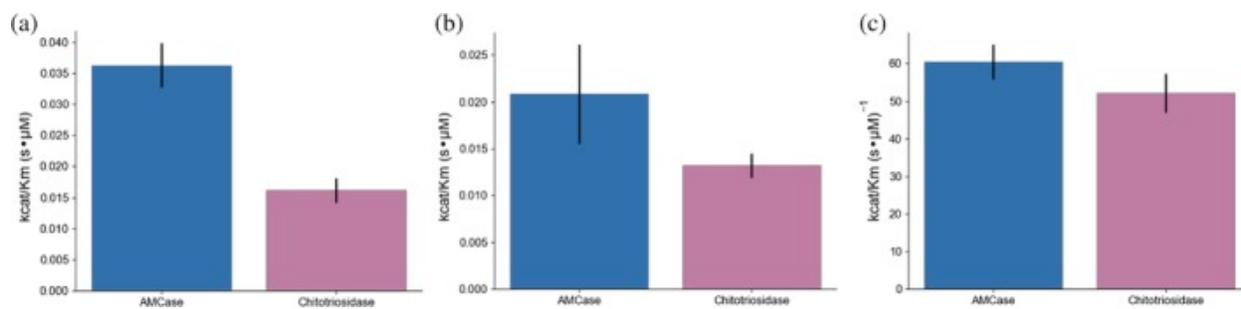


Figure 1.5 | Comparison of AMCCase and Chitotriosidase.

Differences in k_{cat}/K_M between AMCCase (blue) and chitotriosidase (magenta) using **A)** 4MU-chitobioside, **B)** 4MU-chitotrioside, **C)** chitooligosaccharide oxidase coupled peroxidase assay. Error bars denote propagated SD of fit (accounting for covariance)

Methods

Protein preparation

Constructs expressing a fusion of a protein A secretion sequence targeting periplasmic expression, AMCase or chitotriosidase, and a C-terminal V5-6xHIS as previously described²⁰ were ordered from Atum (Newark, CA). Mutants of AMCase were generated via PCR mutagenesis. Plasmids were transformed into BL21 cells and expressed overnight in ZY Autoinduction media at 37°C for 3 hr followed by 19°C overnight. We added protease inhibitor at the temperature change to minimize proteolysis of periplasmically expressed protein. Pelleted cells were lysed via osmotic shock in a two-step procedure. First, cells were resuspended in 20% Sucrose w/v, 20 mM Tris pH 6.5, 1 mg/mL lysozyme, 1 μ L universal nuclease, with a protease inhibitor tablet. The resuspended cells were incubated at 37°C for 1 hr, and then pelleted via centrifugation at 15000 \times g for 15 min. The supernatant was collected, and the pellet was resuspended in a wash buffer of 20 mM Tris pH 6.5 and 150 mM NaCl and incubated for 15 min at 4°C. The cells were centrifuged at 15000 \times g for 15 min, and the supernatant was combined with the supernatant from the first step to form the combined lysate. The combined lysate was bound to a HisTrap FF column, washed with 100 mM Tris pH 6.5, 150 mM NaCl, and then eluted with a gradient into 100 mM Tris pH 6.5, 150 mM NaCl, 500 mM imidazole. Fractions were selected for further purification based on activity assay with a commercial fluorogenic substrate (described below). Active fractions were pooled and subject to dialysis overnight into 100 mM Sodium Acetate pH 4.5, 150 mM NaCl, 5% glycerol w/v followed by filtration to remove insoluble aggregate and dialysis into 100 mM Tris pH 6.5, 150 mM NaCl, 5% glycerol w/v. The protein solution was concentrated and separated via size-exclusion chromatography on

a Superdex S75 16/600. Fractions were selected based on purity as assessed via SDS-page gel electrophoresis and based on activity as assayed with a commercial fluorogenic substrate.

Analysis of Kinetic Data

Kinetic measurements were made in a range of substrate concentrations outside of pseudo-first-order conditions. To robustly measure rates of catalysis, we fit our data using non-linear least-squares curve fitting to simple relaxation models for enzyme kinetics:

$$A(1 - e^{-k_1 t}) + B$$

$$A(1 - e^{-k_1 t}) + B(1 - e^{-k_2 t}) + C$$

where A shows the asymptotic signal from the clearance of substrate, k_1 is the rate constant of relaxation, and B is the background signal of the assay condition. To this end, we developed a small python library for relaxation modeling, which is available on GitHub:

<https://github.com/fraser-lab/relax>. Generally, a single-step relaxation model was required, but in cases where residuals showed significant structure, additional steps were added as either relaxation or linear fits (in cases where kinetics were pseudo-first-order). Specific data analysis scripts using relax.py are available at https://github.com/fraser-lab/chitin_analysis.

Continuous fluorescence measurements to quantify activity using commercial oligomeric substrates

Catalytic activity was assayed using 4-methylumbelliferyl chitobioside and 4-methylumbelliferyl chitotrioside as described previously³² with one critical modification. 10 nM chitinase enzyme was incubated with varying concentrations of 4MU-chitobioside or 4MU-chitotrioside up to

433 μM in McIlvaine Buffer³³ pH 7.0 at 37°C. The 4-methylumbelliferone (4MU) fluorophore is quenched by a β -glycosidic linkage to a short chitin oligomer, which is cleaved by a chitinase enzyme, which generates fluorescence with peak excitation at 360 nm and emission at 450 nm. Previously, the reaction was quenched, and the pH was raised to maximize the quantum yield of the 4MU substrate. To avoid noise introduced by quenching and substrate concentration, we measured fluorescence at regular intervals during the course of the reaction without a pH shift and determined the rate using a single step relaxation model. This allowed us to measure rates of catalysis under a large range of conditions without needing to account for the proper time to quench to maximize signal without the reaction reaching completion. The processing for data collected from this assay is illustrated in **Supplemental Figure 1.1**.

Bulk clearance activity assay

Borohydride-reduced colloidal chitin was purchased as a powder from Megazyme (Bray, Ireland) and resuspended to 4% w/v in pH 7.0 McIlvaine buffer. Higher concentrations did not stay in suspension effectively. To remove soluble oligomers, the suspension was pelleted by centrifugation at $3200 \times g$, the supernatant was discarded, and the pellet was resuspended in McIlvaine buffer. This wash step was repeated a total of five times. A concentration series was prepared by serial dilution of this washed 4% w/v stock in McIlvaine buffer, and 50 μL of each substrate concentration was incubated with 50 μL of 200 nM chitinase at 37°C in a clear-bottomed 96-well microplate with a lid that was sealed around the sides with parafilm to minimize evaporation. Clearance of substrate was monitored by reduction of scattering at OD_{680} for 72 hr with shaking between reads to maintain substrate suspension. The processing for data collected from this assay is illustrated in **Supplemental Figure 1.2**.

Potassium ferricyanide reduction assay

Four percentage w/v colloidal chitin was washed as above, and then diluted serially to generate a concentration range from was incubated with 1–100 nM chitinase for up to 18 hr at 37°C. At the endpoint of incubation, 50 µL of reaction mixture was quenched by the addition of 100 µL of 400 mM sodium carbonate. The insoluble chitin was pelleted by centrifugation at 4000 × g, and then 100 µL of supernatant was mixed with 100 µL of 0.6 g/mL potassium ferricyanide in a 96-well microplate with clear bottoms and a lid that was sealed around the sides with parafilm to minimize evaporation. The microplate was incubated for 4 hr at 42°C to maximize the rate of the nonenzymatic reduction of potassium ferricyanide by solubilized reducing sugars. During incubation absorbance at 420 nm was read out in 1 min intervals. We found that progress curve analysis gave poor results for this data, and instead ultimately found the difference between the maximum and minimum absorbance to be a more robust measure of total reducing sugar generation in the 18-hr incubation with chitinase. The processing of the data for this assay to generate rates is illustrated in **Supplemental Figure 1.3**.

Chitooligosaccharide oxidase coupled peroxidase assay

Processing of colloidal chitin and resultant generation of new reducing sugar moieties was monitored, as previously described¹², by oxidation by chitooligosaccharide oxidase (ChitO), producing as a byproduct peroxide, which in turn is converted into a fluorescent signal by horseradish peroxidase (HRP) and QuantaRed peroxidase substrate.³⁴ ChitO was purchased from Gecco Biotech (Groningen, the Netherlands), HRP and QuantaRed substrate were purchased from Sigma (St Louis, MO). Incorporating a fluorogenic HRP substrate improves the dynamic

range of the experiment and enables real-time observation of reducing sugar cleavage in a one-pot reaction incorporating insoluble chitin, chitinase, chitO, HRP, and QuantaRed substrate. Briefly, a 50 μ L solution containing 1–10 nM chitinase, 20 U/mL HRP, 100 nM ChitO, 0.5 μ L of QuantaRed substrate, and 10 μ L of QuantaRed enhancer solution in McIlvaine buffer pH 7.0 was mixed with 50 μ L of washed colloidal chitin substrate, as prepared above, in a black 96-well microplate with a lid to minimize evaporation. The plate was incubated with at 37°C and the fluorescence of the QuantaRed substrate was measured at 1-min intervals for 16 hr. The progression of fluorescence over time was modeled as a relaxation process as described above, after subtracting the signal from a chitinase-free control, which had signal that was modulated by the washing of the colloidal chitin. This enzyme-coupled reaction is sensitive to reaction conditions, with artifacts introduced by insufficient excess of chitO or HRP as well as by insufficiently washed colloidal chitin. With careful washing of the colloidal chitin and sufficient prewarming of both enzyme and substrate solutions, rates can be reliably measured for chitin concentrations ranging from 0.0005% to 2% colloidal chitin w/v, and for chitinase concentrations as low as 50 pM. The processing of data from this experiment is illustrated in **Supplemental Figure 1.4.**

Random mutagenesis and screening

Random mutations were generated using the commercial Genemorph II random mutagenesis kit (Agilent, Santa Clara, CA). The catalytic domain of acidic mammalian chitinase was amplified via error-prone PCR with varying amounts of parent plasmid present. We titrated the amount of parent plasmid until each clone carried 1–2 mutations. We then performed restriction digestion using StyI and Eco130I and ligation using Quick Ligase to generate plasmids containing our

mutations. We transformed these into electrocompetent BL21(DE3) *E. coli*. Individual colonies were picked and grown overnight in 96-well deep-well blocks, and then 20 μL of starter media was used to inoculate 300 μL of ZY media in deep well blocks, which was then used to express the protein at 30°C overnight. After expression, 50 μL of media from individual wells was mixed with 50 μL of 21.6 μM 4MU-chitobioside in McIlvaine buffer pH 7.0, which had been prewarmed to 37°C. The mixture was monitored by fluorescence as described above, and compared to positive and negative controls, which had been expressed in the same plate. Mutants with increased activity were grown out, mini-prepped, sequenced, retransformed, and expressed and rescreened in this manner in triplicate to confirm improved activity. Winners at this point were stored individually and pooled for further error-prone PCR and screening.

References

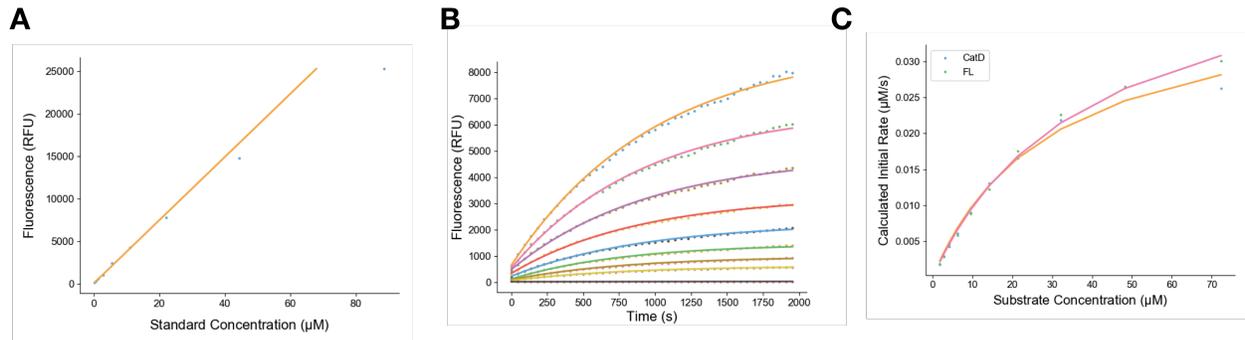
1. Cabib, E. & Bowers, B. Timing and function of chitin synthesis in yeast. *J. Bacteriol.* **124**, 1586–1593 (1975).
2. Zhu, K. Y., Merzendorfer, H., Zhang, W., Zhang, J. & Muthukrishnan, S. Biosynthesis, turnover, and functions of chitin in insects. *Annu. Rev. Entomol.* **61**, 177–196 (2016).
3. Tang, W. J., Fernandez, J., Sohn, J. J. & Amemiya, C. T. Chitin is endogenously produced in vertebrates. *Curr. Biol.* **25**, 897–900 (2015).
4. Bussink, A. P., Speijer, D., Aerts, J. & Boot, R. G. Evolution of mammalian chitinase(–like) members of family 18 glycosyl hydrolases. *Genetics* **177**, 959–970 (2007).
5. Boot, R. G. *et al.* Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *J. Biol. Chem.* **276**, 6770–6778 (2001).
6. Jang, M.-K., Kong, B.-G., Jeong, Y.-I., Lee, C. H. & Nah, J.-W. Physicochemical characterization of α -chitin, β -chitin, and γ -chitin separated from natural resources. *J Polym Sci A Polym Chem* **42**, 3423–3432 (2004).
7. Rinaudo, M. Chitin and chitosan: Properties and applications. *Prog. Polym. Sci.* **31**, 603–632 (2006).
8. Horn, S. J. *et al.* Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18089–18094 (2006).
9. Nakamura, A., Okazaki, K.-I., Furuta, T., Sakurai, M. & Iino, R. Processive chitinase is Brownian monorail operated by fast catalysis after peeling rail from crystalline chitin. *Nat. Commun.* **9**, 1–12 (2018).

10. Wakita, S. *et al.* Improved fluorescent labeling of chitin oligomers: Chitinolytic properties of acidic mammalian chitinase under somatic tissue pH conditions. *Carbohydr. Polym.* **164**, 145–153 (05/2017).
11. Okawa, K. *et al.* Loss and Gain of Human Acidic Mammalian Chitinase Activity by Nonsynonymous SNPs. *Mol. Biol. Evol.* **33**, 3183–3193 (2016).
12. Ferrari, A. R., Gaber, Y. & Fraaije, M. W. A fast, sensitive and easy colorimetric assay for chitinase and cellulase activity detection. *Biotechnol. Biofuels* **7**, 37 (2014).
13. Renkema, G. H. *et al.* Synthesis, sorting, and processing into distinct isoforms of human macrophage chitotriosidase. *Eur. J. Biochem.* **244**, 279–285 (1997).
14. Bigg, H. F., Wait, R., Rowan, A. D. & Cawston, T. E. The mammalian chitinase-like lectin, YKL-40, binds specifically to type I collagen and modulates the rate of type I collagen fibril formation. *J. Biol. Chem.* **281**, 21082–21095 (2006).
15. Steentoft, C., Vakhrushev, S. Y. & Joshi, H. J. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J* **32**, 1478–1488 (2013).
16. Reese, T. A. *et al.* Chitin induces accumulation in tissue of innate immune cells associated with allergy. *Nature* **447**, 92–96 (2007).
17. Bierbaum, S. *et al.* Polymorphisms and haplotypes of acid mammalian chitinase are associated with bronchial asthma. *Am. J. Respir. Crit. Care Med.* **172**, 1505–1509 (2005).
18. Seibold, M. A. *et al.* Differential enzymatic activity of common haplotypic versions of the human acidic Mammalian chitinase protein. *J. Biol. Chem.* **284**, 19650–19658 (2009).
19. Van Dyken, S. J. *et al.* Spontaneous Chitin Accumulation in Airways and Age-Related Fibrotic Lung Disease. *Cell* **169**, 497-509.e13 (2017).

20. Kashimura, A. *et al.* Protein A-Mouse Acidic Mammalian Chitinase-V5-His Expressed in Periplasmic Space of *Escherichia coli* Possesses Chitinase Functions Comparable to CHO-Expressed Protein. *PLoS One* **8**, e78669 (2013).
21. Tjoelker, L. W. *et al.* Structural and functional definition of the human chitinase chitin-binding domain. *J. Biol. Chem.* **275**, 514–520 (2000).
22. Limón, M. C., Margolles-Clark, E., Benítez, T. & Penttilä, M. Addition of substrate-binding domains increases substrate-binding capacity and specific activity of a chitinase from *Trichoderma harzianum*. *FEMS Microbiol. Lett.* **198**, 57–63 (2001).
23. Cole, S. W. The determination of reducing sugars by titration of ferricyanide. *Biochem. J.* **27**, 723–726 (1933).
24. Zhu, Z. *et al.* Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. *Science* **304**, 1678–1682 (2004).
25. Kimura, M. *et al.* Direct comparison of chitinolytic properties and determination of combinatory effects of mouse chitotriosidase and acidic mammalian chitinase. *Int. J. Biol. Macromol.* **134**, 882–890 (2019).
26. Zhang, P., Himmel, Y.-H. & Mielenz, M. E. Outlook for cellulase improvement: Screening and selection strategies. *Biotechnol Adv* **24**, 452–481 (2006).
27. Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 8852–8858 (2019).
28. Gubaeva, E., Gubaev, A. & Melcher, R. Slipped sandwich” model for chitin and chitosan perception in *Arabidopsis*. *Mol Plant Microbe Interact* **31**, 1145–1153 (2018).

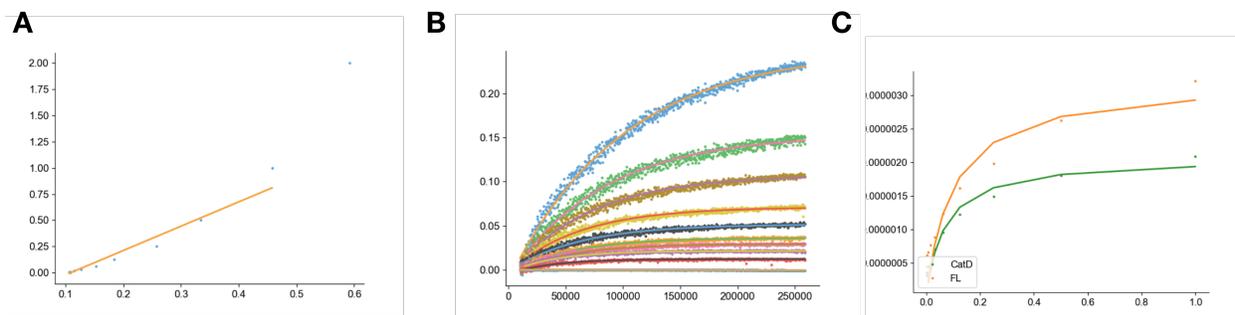
29. Nakamura, A. *et al.* Rate constants, processivity, and productive binding ratio of chitinase A revealed by single-molecule analysis. *Phys. Chem. Chem. Phys.* **20**, 3010–3018 (2018).
30. Fukuda, S. *et al.* High-speed atomic force microscope combined with single-molecule fluorescence microscope. *Rev. Sci. Instrum.* **84**, 073706 (2013).
31. Cole, D. C. *et al.* Identification and characterization of acidic mammalian chitinase inhibitors. *J. Med. Chem.* **53**, 6122–6128 (2010).
32. Renkema, G. H., Boot, R. G., Muijsers, A. O., Donker-Koopman, W. E. & Aerts, J. M. Purification and characterization of human chitotriosidase, a novel member of the chitinase family of proteins. *J. Biol. Chem.* **270**, 2198–2202 (1995).
33. McIlvaine, T. C. A buffer solution for colorimetric comparison. *J. Biol. Chem.* **49**, 183–186 (1921).
34. *QuantaRed Enhanced Chemifluorescent HRP Substrate Kit - ThermoFisher Scientific.*
Available from: <https://www.thermofisher.com/order/catalog/product/15159>. Accessed 3/2018 by BAB.

Supplemental Figures



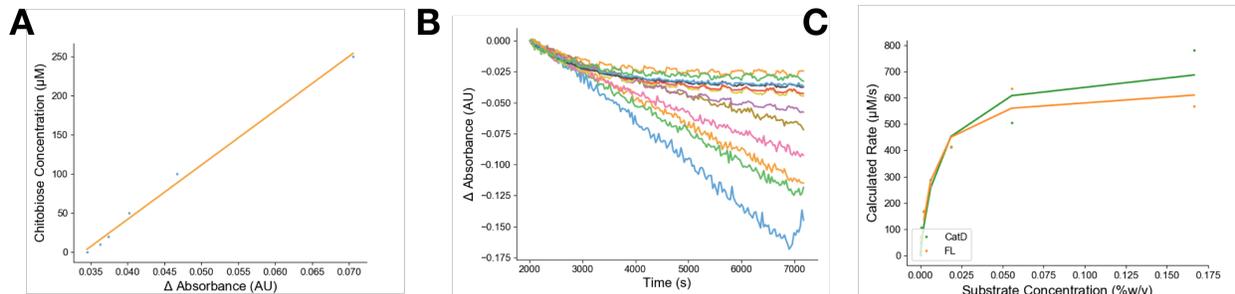
Supplemental Figure 1.1 | Data processing for 4MU assay.

A) Standards of 4MU were measured by fluorescence at 360 nm excitation and 420 nm emission and concentrations below 50 μM fit well to a linear regression. **B)** Progress curves of a concentration series of 4MU-chitobioside were fit by a non-linear relaxation analysis to extract initial rates. **C)** Initial rates were plotted against substrate concentration and were fitted via non-linear regression to a Michaelis-Menten curve to extract rate constants.



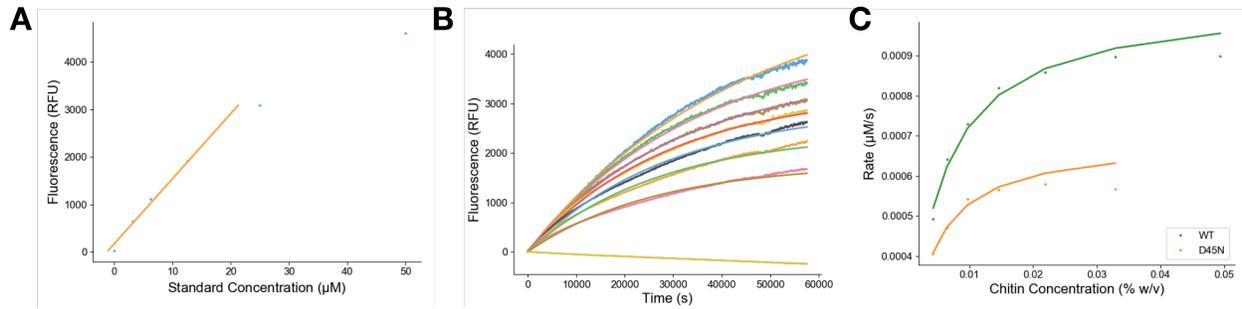
Supplemental Figure 1.2 | Data processing for colloidal chitin clearance assay.

A) Concentrations from enzyme-free controls were matched to absorbance, and for concentrations below 0.5% w/v a linear regression fit the data reasonably well. **B)** Progress curves of a concentration series of bulk chitin were subtracted from their initial state, then fit by a non-linear relaxation analysis to extract initial rates. **C)** Initial rates were plotted against substrate concentration and were fitted via non-linear regression to a Michaelis–Menten curve to extract rate constants.



Supplemental Figure 1.3 | Data processing for ferricyanide reduction assay.

A) Concentrations from chitobiose controls were matched to absorbance, and for concentrations below $250 \mu\text{M}$ a linear regression was fit the data. **B)** From progress curves for the non-enzymatic reaction with potassium ferricyanide, the maximum and minimum values were subtracted from each other and scaled by the incubation time to extract the rate of generation of soluble reducing sugars. **C)** Rates were plotted against substrate concentration and were fitted via non-linear regression to a Michaelis–Menten curve to extract rate constants.



Supplemental Figure 1.4 | Data processing for chitO assay.

A) Concentrations from chitobiose controls were matched to fluorescence after incubation with chitO, horseradish peroxidase, and QuantaRed, and for concentrations below 30 μM a linear regression was fit the data. **B)** From progress curves, a non-linear regression was used to fit relaxation parameters to extract initial rates for a concentration series of colloidal chitin. **C)** Rates were plotted against substrate concentration and were fitted via non-linear regression to a Michaelis–Menten curve to extract rate constants.

Acknowledgments

Funding

JSF was supported by a Pew Scholar Award from Pew Charitable Trusts, a pilot grant from the Sandler Asthma Basic Research Center (SABRE Center), and NIH GM123159. BAB was supported by an ARCS scholarship, a UCSF Discovery fellowship, and by NIH GM008284. RED was supported by the National Science Foundation Graduate Research Fellowship under Grant No. (1650113). SvD and RML were supported by NIH R01 HL128903 and HHMI.

Competing interests

BAB and JSF are inventors on a provisional patent application for the mutants described herein and their use in treating fibrotic lung disease. SJVD and RML are inventors on a pending patent application on the use of chitinases for treating fibrotic lung disease.

Chapter 2 Fragment binding to the Nsp3 macrodomain of SARS-CoV-2 identified through crystallographic screening and computational docking.

Marion Schuller^{1†}, Galen J. Correy^{2†}, Stefan Gahbauer^{3†}, Daren Fearon^{4†}, Taiasean Wu^{5,6},
Roberto Efraín Díaz^{2,7}, Iris D. Young^{2,8}, Luan Carvalho Martins⁹, Dominique H. Smith¹⁰, Ursula
Schulze-Gahmen⁸, Tristan W. Owens⁸, Ishan Deshpande⁸, Gregory E. Merz⁸, Aye C. Thwin⁸,
Justin T. Biel⁸, Jessica K. Peters⁸, Michelle Moritz⁸, Nadia Herrera⁸, Huong T. Kratochvil⁸,
QCRG Structural Biology Consortium^{8‡}, Anthony Aimon⁴, James M. Bennett¹¹, Jose Brandao
Neto⁴, Aina E. Cohen¹², Alexandre Dias⁴, Alice Douangamath⁴, Louise Dunnett⁴, Oleg
Fedorov¹¹, Matteo P. Ferla¹³, Martin Fuchs¹⁴, Tyler J. Gorrie-Stone⁴, James M. Holton^{15,16,12},
Michael G. Johnson¹⁷, Tobias Krojer^{11,18}, George Meigs^{15,16}, Ailsa J. Powell⁴, Johannes Gregor
Matthias Rack¹, Victor L Rangel^{11,18,19}, Silvia Russi¹², Rachael E. Skyner⁴, Clyde A. Smith¹²,
Alexei S. Soares²⁰, Jennifer L. Wierman¹², Kang Zhu¹, Peter O'Brien²¹, Natalia Jura²², Alan
Ashworth¹⁰, John Irwin³, Michael C. Thompson²³, Jason E. Gestwicki^{3,5}, Frank von
Delft^{4,11,18,24*}, Brian K. Shoichet^{3*}, James S. Fraser^{2*}, Ivan Ahel^{1*}

Affiliations

¹ Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1
3RE, UK.

² Department of Bioengineering and Therapeutic Sciences, University of California San
Francisco, CA 94158, USA.

³ Department of Pharmaceutical Chemistry, University of California San Francisco, CA 94158,
USA.

⁴ Diamond Light Source Ltd., Harwell Science and Innovation Campus, Didcot OX11 0DE, UK.

⁵ Institute for Neurodegenerative Disease, University of California San Francisco, CA 94158,
USA.

⁶ Chemistry and Chemical Biology Graduate Program, University of California San Francisco,
CA 94158, USA.

⁷ Tetrad Graduate Program, University of California San Francisco, CA 94158, USA.

⁸ Quantitative Biosciences Institute (QBI) Coronavirus Research Group Structural Biology
Consortium, University of California San Francisco, CA 94158, USA.

⁹ Biochemistry Department, Institute for Biological Sciences, Federal University of Minas
Gerais, Belo Horizonte, Brazil.

¹⁰ Helen Diller Family Comprehensive Cancer, University of California San Francisco, CA
94158, USA.

¹¹ Centre for Medicines Discovery, University of Oxford, South Parks Road, Headington OX3
7DQ, UK.

¹² Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Center, Menlo Park,
CA 94025, USA.

¹³ Wellcome Centre for Human Genetics, University of Oxford, Old Road Campus, Oxford OX3
7BN, UK.

¹⁴ National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973,
USA.

- ¹⁵ Department of Biochemistry and Biophysics, University of California San Francisco, CA
94158, USA.
- ¹⁶ Department of Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National
Laboratory, Berkeley, CA 94720, USA.
- ¹⁷ ChemPartner Corporation, South San Francisco, CA 94080, USA.
- ¹⁸ Structural Genomics Consortium, University of Oxford, Old Road Campus, Roosevelt Drive,
Headington OX3 7DQ, UK.
- ¹⁹ School of Pharmaceutical Sciences of Ribeirao Preto, University of Sao Paulo, São Paulo,
Brazil.
- ²⁰ Photon Sciences, Brookhaven National Laboratory, Upton, NY 11973, USA.
- ²¹ Department of Chemistry, University of York, Heslington, York YO10 5DD, UK.
- ²² Department of Cellular and Molecular Pharmacology, University of California San Francisco,
CA 94158, USA.
- ²³ Department of Chemistry and Chemical Biology, University of California Merced, CA 95343,
USA.

Contributions

The work in this chapter is published as Schuller, Correy, Gahbauer, Fearon et al. in *Science Advances* (2021). M. Schuller designed and cloned the construct that yielded the P4₃ crystals at Oxford/XChem; expressed, purified, and establishing crystallization conditions for P4₃ crystals at Oxford/XChem; assisted with data processing/analysis at XChem; set up the HTRF functional assay and performed data interpretation; prepared the manuscript. G. Correy cloned, expressed, and purified the P4₃ construct at UCSF; crystallized, performed fragment soaking, vitrified crystals, collected X-ray diffraction data and processed data for fragment screens at UCSF; modelled, refined and analyzed fragment structures at UCSF; purified, methylated and crystallised the C2 construct; performed ADP-ribose soaks; refined the apo and ADP-ribose bound structures determined using the P4₃ crystals; prepared the manuscript. S. Gahbauer performed docking screens against Mac1; performed the chemoinformatic analysis of fragment libraries; assisted with fragment-linking and -merging; prepared the manuscript. D. Fearon crystallized, prepared samples, collected X-ray diffraction data, refined, and analyzed fragment structures at XChem; prepared the manuscript. T. Wu performed and analyzed DSF experiments; prepared the manuscript. R.E. Díaz designed the construct that yielded the C2 crystals; expressed the C2 construct; performed and analyzed ITC experiments; prepared the manuscript. I. Young collected X-ray diffraction data and processed the diffraction data for crystals screened at UCSF. L. Martins assisted with docking screens against Mac1. D. Smith assisted with DSF experiments. U. Schulze-Gahmen crystallized the C2 construct at UCSF. T. Owens purified the C2 construct at UCSF. I. Deshpande collected X-ray diffraction data for C2 crystals at UCSF. G. Merz purified the C2 construct at UCSF. A. Thwin purified the C2 construct at UCSF. J. Biel supported the fragment soaking experiments performed at UCSF; supported the fragment

modeling and refinement at UCSF. J. Peters purified C2 construct at UCSF. M. Moritz purified C2 construct at UCSF. N. Herrera supported crystallization of the C2 construct at UCSF. H. Kratochvil supported crystallization of the C2 construct at UCSF. QCRG Structural Biology Consortium provided infrastructure and support for experiments performed at UCSF. A. Aimon prepared samples and collected X-ray diffraction data at XChem. J. Bennett set up the HTRF functional assay; analyzed and collected data for the HTRF assay. J. Neto collected X-ray diffraction data at XChem. A. Cohen supported X-ray diffraction experiments at the SSRL. A. Dias collected X-ray diffraction data at XChem. A. Douangamath crystallized, prepared samples, collected X-ray diffraction data, refined, and analyzed fragment structures at XChem. L. Dunnett collected X-ray diffraction data at XChem. O. Fedorov set up the HTRF functional assay; analyzed and collected data for the HTRF assay. M. Ferla performed fragment merging and linking. M. Fuchs supported X-ray diffraction experiments at the NSLS-II. T. Gorrie-Stone deposited fragments structures at XChem. J. Holton supported X-ray diffraction experiments at the ALS. M. Johnson analyzed and assigned structures to the N3- and N9-alkylated fragments. T. Krojer analyzed structural data at XChem. G. Meigs supported X-ray diffraction experiments at the ALS. A. Powell collected X-ray diffraction data at XChem. J. Rack provided feedback on the manuscript. V. Rangel refined fragment structures at XChem. S. Russi supported X-ray diffraction experiments at the SSRL. R. Skyner deposited fragments structures at XChem. C. Smith supported X-ray diffraction experiments at the SSRL. A. Soares supported X-ray diffraction experiments at the NSLS-II. J. Wierman supported X-ray diffraction experiments at the SSRL. K. Zhu expressed and crystallized the P4₃ construct at XChem. P. O'Brien designed and provided the York3D fragment library. N. Jura supervised work. A. Ashworth supervised work; prepared the manuscript. J. Irwin designed the UCSF_91 fragment library; prepared the

manuscript. M. Thompson vitrified C2 crystals at UCSF; collected X-ray diffraction data for C2 crystals at UCSF; refined the ultra-high resolution structure determined using C2 crystals; prepared the manuscript. J. Gestwicki supervised work. F. Delft provided the XChem facility; prepared the manuscript; supervised work. B. Shoichet guided and evaluated the docking work; prepared the manuscript; supervised work. J. Fraser supervised work; prepared the manuscript; arranged funding. I. Ahel supervised work; prepared the manuscript; arranged funding.

Abstract

The SARS-CoV-2 macrodomain (Mac1) within the non-structural protein 3 (Nsp3) counteracts host-mediated antiviral ADP-ribosylation signaling. This enzyme is a promising antiviral target because catalytic mutations render viruses non-pathogenic. Here, we report a massive crystallographic screening and computational docking effort, identifying new chemical matter primarily targeting the active site of the macrodomain. Crystallographic screening of diverse fragment libraries resulted in 214 unique macrodomain-binding fragments, out of 2,683 screened. An additional 60 molecules were selected from docking over 20 million fragments, of which 20 were crystallographically confirmed. X-ray data collection to ultra-high resolution and at physiological temperature enabled assessment of the conformational heterogeneity around the active site. Several crystallographic and docking fragment hits were also confirmed by solution binding using three biophysical techniques (DSF, HTRF, ITC). The 234 fragment structures presented explore a wide range of chemotypes and provide starting points for development of potent SARS-CoV-2 macrodomain inhibitors.

Introduction

Macrodomains are conserved protein domains found in all kingdoms of life and in several viruses¹. Viral macrodomains recognize and remove host-derived ADP-ribosylation, a post-translational modification of host and pathogen proteins^{2,3}. The innate immune response involves signaling by ADP-ribosylation, which contributes to the suppression of viral replication³⁻⁷. Upon viral infection, ADP-ribosylation is catalyzed by an interferon-induced subset of mammalian ADP-ribsyltransferases (ARTs), collectively termed ‘antiviral poly(ADP-ribsyl) polymerases’ (PARPs)^{3,8}. These enzymes transfer the ADP-ribose (‘ADPr’) moiety of NAD⁺ onto target proteins^{3,8}. For example, during coronavirus infection, PARP14 stimulates interleukin 4 (IL-4)-dependent transcription, which leads to the production of pro-inflammatory, antiviral cytokines⁹. Viral macrodomains, which are found primarily in corona-, alpha-, rubi- and herpes-viruses, can counteract this host defense mechanism via their (ADP-ribsyl)hydrolase activity, contributing to the host-viral arms race for control of cell signalling¹⁰.

Coronaviruses (CoVs) are important pathogens of livestock and humans. Three strains out of seven known to infect humans have caused major outbreaks within the last two decades: the severe acute respiratory sndrome (SARS) coronavirus, causing the SARS epidemic from 2002-2004, the Middle East respiratory sndrome (MERS) coronavirus, causing outbreaks in 2012, 2015 and 2018, and SARS-CoV-2, causing the current COVID-19 pandemic¹¹. The coronaviral conserved macrodomain (called ‘Mac1’ here; also known as ‘S2-MacroD’ or ‘X domain’) is encoded as part of the non-structural protein 3 (Nsp3), a 200 kDa multi-domain protein¹². While cell culture experiments suggest that SARS Mac1 is dispensable for viral replication in some cell lines^{5,13,14}, animal studies have shown that its hydrolytic activity promotes immune evasion and that it is essential for viral replication and pathogenicity in the

host^{6,7}. The critical role of macrodomains is further supported by experiments using catalytic null mutations of the murine hepatitis virus (MHV), which render that virus essentially non-pathogenic^{5,6,13}. Collectively, these findings support the idea that SARS-CoV-2 Mac1 is a promising drug target for disrupting the viral life cycle.

A barrier for macrodomain drug discovery has been the lack of well-behaved inhibitors for this domain. Making matters worse, there are few biochemical assays suitable for screening for such inhibitors. Thus far, PDD00017273, an inhibitor of the poly(ADP-ribose)glycohydrolase (PARG), a macrodomain-type (ADP-ribosyl)hydrolase, remains the only well-characterized inhibitor with convincing on-target pharmacology and selectivity¹⁵. The initial hit was discovered by a homogeneous time-resolved fluorescence (HTRF)-based assay that measures PARG activity, rendering the assay unsuitable for macrodomains that lack this activity¹⁶. A selective allosteric inhibitor targeting PARP14 was identified in an AlphaScreen-based high-throughput screen (HTS)¹⁷. While this inhibitor showed on-target activity in cells, its unique allosteric binding site is difficult to translate to other macrodomains. While potential Mac1 inhibitors have emerged with the advent of SARS-CoV-2¹⁸, their binding mechanisms and efficacy remain unclear, and the lack of a biochemical assay specific for Mac1 has hindered their development. Furthermore, structures of the new inhibitors bound to Mac1 have not yet been reported, making optimization of initial hits, however promising, difficult.

To address the lack of chemical matter against Mac1, we turned to fragment-based ligand discovery using crystallography as a primary readout (**Figure 2.1**). Fragment screens can efficiently address a large and relatively unbiased chemical space¹⁹. Despite typically weak overall affinity, fragments often have high ligand efficiency ($-\Delta G_b/HAC$), and can provide templates for further chemical elaboration into lead-like molecules²⁰. Crystallography can be

used as a primary screening method for fragment discovery²¹, and recent automation and processing software at synchrotron radiation sources has made this routinely possible at facilities like the XChem platform at Diamond Light Source²²⁻²⁵. As part of Diamond's contribution toward efforts to combat COVID-19, fragment screening expertise and infrastructure was made immediately available to any users working on SARS-CoV-2 targets²⁶. Similarly, synchrotron access for essential COVID-19-related research was also made available at the US Department of Energy light sources.

Because crystallographic fragment screens can generate hits that bind anywhere on the protein surface, we wanted to supplement those screens with molecular docking intentionally targeting the active site. Docking has the additional benefit of exploring a much larger chemical space than an empirical fragment library. While an empirical library of ~1000-to-2000 fragments can represent a chemical space as large as, or larger, than that of a classic HTS library of several million compounds, exploration of chemotypes, including those that are well-suited to a particular target subsite, will inevitably be limited²⁷. Conversely, docking a much larger virtual library allows finer grained sampling around many chemotypes. A potential drawback of molecular docking is doubt about its ability to predict weakly-binding fragment geometries with high fidelity. While docking has identified potent ligands from libraries of lead-like molecules (250 to 350 amu)²⁸⁻³⁰, such molecules offer more functional group handles for protein matching than do most fragments (150 to 250 amu), and docking is thought to struggle with the smaller, less complex, and geometrically more promiscuous fragments³¹. Thus, the pragmatism of this approach has been uncertain^{32,33}.

Here, we present a combination of experimental crystallographic-based and computational docking-based fragment screens performed against Nsp3 Mac1 of SARS-CoV-2

(Figure 2.1). Using X-ray crystallography, we screened fragment libraries of 2,683 compounds, yielding 214 unique fragment-bound Mac1 structures at atomic resolution. Docking of more than 20 million compounds prioritized 60 molecules for structure determination, yielding the structures of 20 additional compounds bound to Mac1. Additional X-ray data collection to ultra-high resolution and at physiological temperature illuminated the conformational heterogeneity in the Mac1 active site. We were able to confirm the binding of several fragments with differential scanning fluorimetry (DSF), isothermal titration calorimetry (ITC), and an ADPr-peptide displacement assay (HTRF), validating the activity of these molecules and providing a foundation for their optimization. The new fragments explore a wide range of chemotypes that interact with the catalytic site of Mac1. Together, these results create a roadmap for inhibitor development against Mac1, which may help to combat the pathogenicity of SARS-CoV-2.

Results

Two crystal forms of Nsp3 Mac1 reveal differences in active site accessibility

We sought a crystal system that enabled consistent ligand soaking for fragment screening and for testing docking predictions. Six Mac1 crystal forms have previously been reported (**Supplemental Data 2.1**). Initially, we designed a construct based on PDB entry 6VXS³⁴. This construct has been reported to crystallize in P1, C2 and P2₁ with either 1 or 2 molecules in the asymmetric unit (ASU) (**Supplemental Data 2.1**). This construct crystallized reproducibly in C2 with microseeding and diffracted to a maximum resolution of 0.77 Å (**Supplemental Data 2.1, Supplemental Figure 2.1, Supplemental Figure 2.2A**). This high resolution data yielded electron density maps at true atomic resolution with abundant alternative conformations (**Supplemental Figure 2.1**). The electron density maps also revealed features that are rarely observed in macromolecular crystallography, such as explicit hydrogen atoms, and covalent bond density (**Supplemental Figure 2.1**). Although the active site appears accessible (**Supplemental Figure 2.3B**), efforts to soak ADP-ribose into the crystals were unsuccessful. Additionally, soaking revealed that this crystal form suffers from inconsistent DMSO tolerance (**Supplemental Figure 2.2A**), which is problematic for fragment soaking. In attempts to overcome this problem, we experimented with lysine methylation³⁵, which increased DMSO tolerance (**Supplemental Figure 2.2A**), but unfortunately increased occlusion of the active site (**Supplemental Figure 2.3F,G**), and dehydration, which increased DMSO tolerance, at the cost of non-isomorphism (**Supplemental Figure 2.2A,C**).

In parallel, we designed a new Mac1 construct that crystallized in P4₃ with two molecules in the ASU (**Supplemental Data 2.1**). This construct crystallized reproducibly with microseeding and diffracted to a maximum resolution of 0.85 Å (**Supplemental Data 2.1**). The

sequence differences between the two constructs were slight (**Supplemental Data 2.1**), yet resulted in substantially different crystal packing (**Supplemental Figure 2.3B-E**). Although the active site of protomer B was obstructed, the active site of protomer A was accessible (**Supplemental Figure 2.3B**), and we were able to soak ADP-ribose into the crystals (**Supplemental Figure 2.4A**). This new structure also revealed a notable difference compared to previously reported Mac1-ADPr structures: the α -anomer of the terminal ribose was observed instead of the β -anomer (**Supplemental Figure 2.4A-D**). Despite this, alignment of ADP-ribose is excellent between all Mac1-ADPr structures (**Supplemental Figure 2.4D**), and the structures are similar overall (**Supplemental Figure 2.4E**). The DMSO tolerance of the P4₃ crystals was excellent (**Supplemental Figure 2.2A**). Accordingly, most of our fragment soaking work proceeded with this construct.

Identifying new ligands for Nsp3 Mac1 using crystallographic fragment screening and docking

Characterization of experimental and virtual screening libraries

Crystal soaking screens at the XChem facility were performed with the P4₃ crystals and a collection of fragment libraries (e.g. Diamond, SGC and iNEXT (DSI)-poised Library including 687 molecules³⁶ and the EU Open screen containing 968 molecules) totaling 2,122 molecules (see **Supplemental Data 2.1** for details). Crystals were screened at the Diamond Light Source. At UCSF, a fragment library composed of Enamine's Essential Fragment library with 320 compounds, augmented by an additional 91 molecules from an in-house library (UCSF_91), was screened against both the P4₃ and C2 crystal forms at the Advanced Light Source (ALS), the Stanford Synchrotron Radiation Lightsource (SSRL) and the National Synchrotron Light Source-

II (NSLS-II). On average, molecules across the X-Chem and UCSF collections had molecular weights of 192 ± 47 amu, cLogP values from -1.8 to 3.8, 13 ± 3 heavy atoms, and on average 2 rotatable bonds (**Supplemental Figure 2.5**).

Two fragment libraries were computationally docked against the structure of Mac1 (PDB: 6W02): a library of 722,963 fragments “in-stock” at commercial vendors, and the entire ZINC15 fragment library of 20,006,175 mainly make-on-demand fragments that have not been previously synthesized, but can readily be made, available predominantly from Enamine and Wuxi³⁴. Molecules from the ZINC15 fragment library had molecular weights ≤ 250 amu, cLogP ≤ 3.5 , with an average of 4 rotatable bonds, and typically 4 to 19 heavy atoms (**Supplemental Figure 2.5**). In addition, an “in-human” library of 20,726 drugs, investigational new drugs, and metabolites that have been tested in humans were included into the docking screen, with a view to potential repurposing opportunities. All three sets can be downloaded from ZINC15 (<https://zinc15.docking.org>)³⁷.

We investigated the fragment libraries for their diversity and their representation of chemotypes likely to bind at the adenine recognition site of Mac1 (**Supplemental Figure 2.5**). Bemis-Murcko (BM) scaffold³⁸ analysis revealed 179 unique scaffolds in the UCSF libraries, and 809 such scaffolds in the XChem fragment libraries. The in-stock fragment docking library contained 69,244 scaffolds, while 803,333 scaffolds were present in the entire ZINC15 20M fragment collection. Taken together, the experimentally screened libraries contained roughly two compounds per BM scaffold, while the docking libraries contained approximately ten fragments per scaffold, consistent with the expected higher granularity of the docking libraries afforded by their much larger size.

Since adenine-containing compounds are the only structurally characterized binders of Mac1, and fragment libraries are intended to cover a wide chemotype space, we assessed the prevalence of pyrimidines in the libraries. We found pyrimidines in 12 of the 411 fragments in the UCSF libraries, and in 72 of the 2,126 XChem fragments (3.39% of the physically-screened fragments (**Supplemental Figure 2.5**). Pyrimidines were found in 41,531 of the 722,963 (5.74%) in-stock fragments and in 890,199 molecules of the 20,006,175 compound fragment library (4.44%). While the percentages of molecules carrying the pyrimidine substructure were similar between the physical and docked fragments, the absolute numbers in the latter sets were far higher. Aside from bearing a pyrimidine substructure, these subsets were otherwise diverse: among the 890,199 pyrimidine-containing docking fragments, 60,919 distinct BM scaffolds were identified. Adenine itself was present in 5,457 fragments (582 different scaffolds). Furthermore, as ADP-ribose is negatively charged, anionic compounds were considered to exhibit favorable properties to bind to Mac1 by targeting the diphosphate region. Fortuitously, a substantial fraction (35%) of the UCSF fragment libraries is anionic (**Supplemental Figure 2.5**).

Hit rates and Mac1 interaction sites of fragments

Across both crystal forms and facilities, we collected diffraction data for Mac1 crystals soaked with 2,954 fragments (**Supplemental Data 2.1**). The diffraction characteristics of the P4₃ crystals were excellent: the average resolution was 1.1 Å, and 98% of crystals diffracted beyond 1.35 Å (**Figure 2.2C,E, Supplemental Figure 2.2B**). Although diffraction data was collected for 368 fragments soaked into the C2 crystals at UCSF, data pathologies meant that only 234 datasets could be analyzed. The datasets collected from C2 crystals had a mean resolution of 1.4 Å and ranged from 1.0 to 2.2 Å (**Figure 2.2A, Supplemental Figure 2.2B**). In total, we

identified 234 unique fragments binding to Mac1 using the PanDDA method (**Figure 2.2, Supplemental Data 2.1, Supplemental Data 2.2**)³⁹. Of these, 221 were identified using P4₃ crystals (hit rate of 8.8%) and 13 using C2 crystals (hit rate of 5.6%). 80% of the fragments were identified in the Mac1 active site, near to or overlapping with the regions occupied by the nucleoside (the adenosine site) or the phospho-ribose (the catalytic site) (**Figure 2.2G**). Additional fragments were scattered across the surface of the enzyme, with an enrichment at a distal macrodomain-conserved pocket near lysine 90 (the ‘K90 site’, 14 fragments) and with many others stabilized by crystal contacts (**Figure 2.2B,D,F, Supplemental Figure 2.6**). Coordinates, structure factors, and PanDDA electron density maps for all the fragments have been deposited in the Protein Data Bank (PDB) and are available through the Fragalysis webtool (<https://fragalysis.diamond.ac.uk>).

The unusually high hit rate for the adenosine site in the P4₃ form with the Enamine Essential library (21%) was in contrast to the relatively low hit rate with this library with the C2 form (1.3%). Of the five pairs of fragments identified in both crystal forms, two pairs were identified in the adenine subsite in both crystal forms, two in the adenine subsite in P4₃ crystals but in the K90 site in C2 crystals, and the remaining pair bound to a surface site in the P4₃ crystals and in the K90 site in the C2 crystals (**Supplemental Data 2.1**). Additional paired high-quality datasets were available for 54 fragments that were bound within the P4₃ crystals, but all showed no density for fragments in the C2 crystals (**Supplemental Data 2.1**). It is possible that competition for binding with the N-terminal residues may have contributed to the relatively low hit rate for the C2 form (**Supplemental Figure 2.3F**).

Docking hits mimic the adenine recognition pattern

Docking the entire (20 million) ZINC15 fragment library, after calibration of docking parameters using different control calculations (see **Methods**)^{37,40}, was completed in just under 5 hours of elapsed time on 500 cores. The 20,006,175 fragments were sampled in over 4.4 trillion complexes. Top-ranked molecules were inspected for their ability to form hydrogen bonds similar to adenine (e.g., with the side chain of Asp22 and with the backbones of Ile23 and Phe156), while molecules with internal molecular strain or unsatisfied hydrogen bond donors were deprioritized. Ultimately, we selected 54 fragments from the entire ZINC15 fragment library screen, 9 of which were immediately available for purchase from Enamine and 33 of 45 make-on-demand molecules were successfully synthesized *de novo*. Furthermore, 8 fragments were purchased from the ZINC15 in-stock fragment library screen, and an additional 10 compounds were sourced based on the ‘in-human’ library docking (**Supplemental Data 2.1**).

Of the 60 molecules tested for complex formation by crystal soaking, 20 were observed with unambiguous electron density in complex with Mac1 (**Supplemental Data 2.1**). Here too, the crystals diffracted to exceptionally high resolution, between 0.94 and 1.01 Å. The predicted docking poses typically superposed well on the observed crystallographic results (Hungarian method root mean square deviations⁴¹ ranging from 1-to-5 Å) and 19 out of the 20 docking hits bound to the adenine subsite of the Mac1, as targeted by docking (**Figure 2.3, Supplemental Figure 2.7**).

The most commonly observed scaffold among the docking hits was 7H-pyrrolo(2,3-d)pyrimidine occupying the adenine-binding subsite (**Figure 2.3A-C, Supplemental Figure 2.7A,B**). This ring system is typically hydrogen bonded with Asp22, Ile23 and Phe156. Fragments with this scaffold usually demonstrated high fidelity between the docking results and

the high resolution structures (RMSD 1.5 - 2.3 Å). For RMSD values >2 Å, indicating noticeable deviations between docking and crystallography⁴², visual inspection of docked and solved poses still revealed correct predictions of orientation and key interactions for most fragments in the targeted binding subsite (e.g. **Figure 2.3C,F,G**). Different substituents can be attached to this headgroup e.g. piperidine, adding a hydrophobic segment to the scaffold (e.g. ZINC336438345 (PDB: 5RSE)), occupying most of the adenosine binding site as shown in **Figure 2.3A,B** and **Supplemental Figure 2.7A,B**. In addition to hydrophobic variations, ZINC263392672 (PDB: 5RSG) attaches an anionic substituent to the pyrrolopyrimidine scaffold, offering additional hydrogen bonds within the binding pocket (**Figure 2.3C**). Interestingly, while docking predicted the carboxylic acid of compound ZINC263392672 to insert into the phosphate binding tunnel, forming a hydrogen bond to Val49, the crystal structure instead revealed hydrogen bonds to the backbone amines of Phe156 and Asp157 which we defined as the ‘oxyanion’ subsite within the adenosine site. Interactions with this backbone-defined oxyanion subsite were also observed for many other hits from both the docking and the crystallographically screened libraries (e.g. **Figure 2.3F**, **Supplemental Figure 2.7E**).

For a set of smaller, mainly adenine-like docking hits, modeled to only occupy the adenine subsite of the targeted adenosine binding site (**Figure 2.3D,E**, **Supplemental Figure 2.7C,D**), the comparison between docked and experimental poses revealed deviations between 1.3 and 4 Å. Making these somewhat larger deviations harder to interpret was that for several fragments the crystallographically observed pose, e.g. ZINC331945 (RMSD 3.97 Å, **Figure 2.3E**) and ZINC763250 (RMSD 3.78 Å, **Supplemental Figure 2.7D**), is partially stabilized by interactions with the symmetry mate (see below).

Another group of docking hits was selected for their close mimicry of the adenosine scaffold (**Figure 2.3F,G, Supplemental Figure 2.7I-L**). For these, the ultra-high resolution of the crystal structures was crucial, revealing that for four of these, the wrong purine isomer had been inadvertently synthesized, with alkyl derivatives from the N3 rather than the intended N9 nitrogen corresponding to the alkylation of adenine in ADP-ribose (**Supplemental Figure 2.7I-L**). Characterization of the original compound samples by HPLC/MS and NMR confirmed that the delivered compounds were >95% pure, mis-assigned positional isomers. For ZINC901381520 (**Figure 2.3F**), both N3 (PDB: 5RSK) and N9 (PDB: 5S6W) isomers were synthesized in different batches and confirmed to bind to the targeted adenosine binding site forming equal hydrogen-bond interactions with the protein (**Supplemental Figure 2.7I**). ZINC3888754 (PDB: 5RSC) (**Figure 2.3G**) contains an adenine-like heterocycle extended by methyl-groups at the C7 and C8 positions, revealing opportunities for expanding purine scaffolds beyond the adenine subsite to achieve Mac1 selectivity over other adenine-binding proteins.

In addition to hydrogen-bonding with residues involved in the adenine recognition of ADP-ribose, several docking hits hydrogen bond to the backbone carbonyl group of Ala154 (**Figure 2.3D,I, Supplemental Figure 2.7G**), revealing an intriguing accessory polar contact within this subsite. While most residues surrounding the adenosine-binding site adopted similar conformations in the fragment-bound crystal structures as in the ADPr-bound structure used for docking (PDB: 6W02)³⁴, Asp22 and Phe156 adopted multiple, alternative conformations. In most fragment-bound crystal structures, Phe156 rotated by approximately 90°, enabling improved face-to-face π - π stacking against the aromatic moieties in the bound fragments (**Figure 2.3C-G**). However, the docking template orientation of Phe156 was retained for other pyrimidine-containing fragment-bound crystal structures (**Figure 2.3B,H**).

Overall, two characteristics stand out from the docking screen: first, despite some important differences, there was high fidelity between the docking-predicted poses and those observed by crystallography. The docking hits explored the adenine subsite to which they were targeted. Second, these hits did so with relatively dense variations around several chemotypes, something afforded by the granularity of a >20 million fragment library. This density can be explored further, for example, 9,170 fragments (888 unique BM scaffolds) in the ZINC15 fragment library contained 7H-pyrrolo(2,3-d)pyrimidines, the functional group repeatedly observed in crystallographically confirmed docking hits (**Figure 2.3A-C**).

Analysis of key interactions between Mac1 and fragments from the crystallographic screens

Fragments binding to the adenine subsite

While docking was successful in targeting the adenine binding subsite, crystallographic fragment screening has the advantages of being binding site agnostic and has the potential to identify novel chemotypes at multiple sites. In total, crystallographic screening identified 99 adenine-subsite binding fragments that form subsets of the three hydrogen bonds found between Mac1 and ADP-ribose (**Figure 2.4A-C**). Fragments that formed at least two hydrogen bonds to the adenine subsite were separated into nine classes based on the number, nature and connectivity of atoms involved in such hydrogen bonding (**Figure 2.4D**). The most common class consisted of a 1,3-hydrogen bond donor/acceptor motif (**Figure 2.4D,E.I**). This resembles the kinase hinge binding motif, with the difference being the engagement of a side chain oxygen rather than a backbone carbonyl oxygen (**Supplemental Figure 2.8A-B**)⁴³. While 7 out of 18 fragments in this class were 4-amino-pyrimidine derivatives, other moieties were also found, including two 2-amino-thiazole-based fragments and several purine derivatives (**Supplemental**

Data 2.1). We also observed an unusual adenine-binding mode with a hydrogen bond formed between Ile23 and N7 instead of N1 (**Figure 2.4D,E.II**). The alternative binding mode can be explained by the N3 substitution of adenine on this fragment, which prevents formation of the canonical N1-Ile23 hydrogen bond. This pattern of hydrogen bonds to the protein has not been previously observed in adenines linked through N9.⁴⁴

We also observed diverse fragments without adenine-like motifs binding at this site, including succinimides, amides, thiazoles, diamino-pyridines, pyrazoles, pyrroles, and ureas (**Figure 2.4D,E.III-VIII**). These exploited, separately, and together, Asp22 and Ile23, Ala154, and occasionally all three adenine-defining hydrogen-bonding residues. Several fragments π - π stacked with Phe156, while those bearing a urea hydrogen-bonded with the carboxylate of Asp22 (**Figure 2.4D,E.VIII**). These interactions were reproduced by a series of benzimidazole-based fragments (**Figure 2.4D,E.IX**). These classes occupied what might be classified as an ‘upper’ subsite, above that defined by the adenine-ribose axis (**Figure 2.2G**), and may provide an opportunity to grow fragments away from the canonical site.

Fragments binding to the oxyanion subsite

In total, we identified 54 fragments that formed interactions with an unexpected “oxyanion” subsite, defined by the backbone nitrogens of Phe156 and Asp157 adjacent to the adenine subsite (**Figure 2.2G**, **Figure 2.5A**). As suggested by its name, most of these fragments (48/54) were anionic (**Supplemental Data 2.1**). Intriguingly, the defining backbone nitrogens adopted a similar orientation to those defining the classic oxyanion hole of serine hydrolases such as acetylcholinesterase (**Supplemental Figure 2.8D-F**). In the Mac1-ADPr structure, the C2 hydroxyl (2’OH) of the proximal ribose interacts with the oxyanion subsite via a bridging

water (**Figure 2.5A**). In total, 54 fragments formed at least one hydrogen bond to the oxyanion subsite (**Figure 2.5B**). Here too, the fragments were both geometrically (**Figure 2.5C**) and chemically diverse (**Figure 2.5D**): orienting groups either toward the phosphate tunnel, the lower site, or wrapped around toward the upper adenine subsite, providing multiple opportunities for further elaboration. Chemically, they interacted with the site using not only a carboxylate, but also sulfones, and isoxazole, α -keto acid, and a succinimide (**Figure 2.5E**). We suspect that the presence of the oxyanion subsite explains the higher hit rate for the Enamine Essential library versus the other crystallographic fragment libraries screened (27% versus 6%), as the former had a greater proportion of acids than the others (41% versus 4%) (**Supplemental Figure 2.5**).

Fragments binding to the catalytic and other potential allosteric sites

There were substantially fewer hits against the catalytic site (**Figure 2.2G**) versus the adenosine site (eight versus >100), though both appear to be accessible (**Supplemental Figure 2.3B**). The catalytic site consists of three subsites: the phosphate tunnel, which is occupied by the diphosphate of ADP-ribose, the ribose subsite, which is occupied by the terminal ribose of the molecule, and the outer subsite, which sits between Asn40 and Asn99 (**Figure 2.2G**, **Figure 2.7A**). Of the eight fragments binding in the catalytic site, seven bound in the outer subsite and one bound in the phosphate tunnel. Binding to the outer site was often defined by hydrophobic packing between the Tyr42 and Lys102 side chains, although POB0135 (PDB: 5S3W) and POB0128 (PDB: 5S3T) formed a salt bridge to Lys102 (e.g. **Figure 2.7A.I**). Interestingly, the latter fragment was also found to bind in the adenosine site. Other molecules, including Z2234920345 (PDB: 5S2L) and Z955123498 (PDB: 5S4A) stabilize an alternative conformation of Lys102 (**Figure 2.7A.II**). Three of the fragments, including Z85956652 (PDB: 5S2U),

positioned a halogen atom in the outer subsite (e.g. **Figure 2.7A.III**). The only fragment identified in the phosphate subsite was ZINC84843283 (PDB: 5RVI). This fragment was wedged between the Gly47/Ile131 loops, and increased the gap between the two loops by 1.6 Å (**Figure 2.7A.IV**). The absence of fragments binding to the ribose subsite, and the sparsity of fragments in the phosphate tunnel, means that designing a Mac1 inhibitor to occupy the catalytic site will rely more heavily on fragment growing than on fragment merging.

Both crystallographic screens also identified fragments binding to the ‘K90 site’, which is formed by a cleft between Lys31, Thr32 and Lys90 (**Figure 2.7B**). We identified seven fragments from the C2 crystal form and six from the P4₃ crystal form; intriguingly, none of the C2-derived fragments were found again when the UCSF libraries were rescreened in the P4₃ crystal condition. Although the K90 site is 15 Å from the adenosine site, it is connected to that subsite via a single alpha-helix (**Figure 2.7B**). Although there is no biochemical evidence for allosteric communication between these sites, the fragments provide starting points for designing chemical probes to test this possibility.

Fragment binding exploits protein conformational flexibility

To identify Mac1 flexibility associated with molecular recognition, we calculated the root-mean-square fluctuation (RMSF) of side-chain atoms across the P4₃ fragment-bound structures. Residues lining the adenosine site, especially Asp22 and Phe156, are the most flexible (**Figure 2.8A,B**). The flexibility of both residues is paralleled in previously reported crystal structures (**Figure 2.8C**) and also in the 0.77 Å apo structure, where multiple alternative conformations are clearly defined in electron density maps (**Figure 2.8D**, **Supplemental Figure 2.1A-C**). In the ultra-high resolution structure, residues 155-159 are modeled as a combination of

two distinct backbone conformations that diverge substantially at Phe156, which requires three distinct conformations of this residue to satisfy the observed density (**Figure 2.8D**, **Supplemental Figure 2.1C**). Despite this flexibility, hydrogen bonds to Asp22 are present in many fragments, including docking compounds that were chosen based on interactions with a static receptor (**Figure 2.8E**). Similarly, the flexibility of the aromatic side-chain of Phe156 enables adaptable stacking interactions with fragments (**Figure 2.8E,F**), with 46 fragments binding within 4 Å of Phe156. As with Asp22, the nature and geometry of these interactions are maintained for many soaked and docked fragments even as the residue moves relative to the rest of the protein.

In contrast to the adenosine site, little conformational heterogeneity is observed at the catalytic site, with only minimal changes in Lys102 and Tyr42 conformations (**Figure 2.8G**). Still, even in this site, there is more conformational heterogeneity observed in previously published structures (**Figure 2.8H**). In particular, a network of flexible side-chains encompassing Phe132, Asn99, and Lys102 is stabilized in a distinct conformation upon ADP-ribose binding (**Figure 2.8I**). To further probe the flexibility of the Phe132-Asn99-Lys102 network, we determined structures of Mac1 using the C2 crystal at human physiological temperature (37°C, 310 K) to 1.5 Å resolution (**Figure 2.8J**, **Supplemental Data 2.1**). As observed in other systems^{45,46}, we noticed that the cryogenic structure appeared more compact than the structure at higher temperatures. Specifically, we observed substantial loop displacements near the ribose-binding pocket of the active site, which are coupled to a global hinge-bending motion involving correlated motion of helices about the central β -sheet (**Supplemental Figure 2.4F,G**). The structure at physiological temperature more closely resembles the structure with ADPr-bound, with the backbone adopting a more open conformation (**Figure 2.8J**). However, the side chain

rotamers of Asn99 and Lys102 do not undergo the larger rearrangements. This temperature-dependent change in the width of the active site cleft can provide alternative, potentially more relevant, conformations for future ligand-discovery efforts targeting the catalytic site around the distal ribose.

Changes in water networks upon fragment binding

To assess the role of water networks in fragment binding, we first examined changes in water networks upon ADP-ribose binding. In the 0.85 Å P4₃ apo structure, the catalytic site contains 14 water molecules arranged in an ordered network that connects the Gly47 loop and the Ile131 loop, with an arc formed around the Phe132 side-chain (**Figure 2.9A**). In contrast, waters were more disordered in the adenosine site, with more diffuse electron density and higher B-factors (**Figure 2.9A,C**). Upon ADP-ribose binding, five waters were displaced from the catalytic site and the water network was disrupted (**Figure 2.9B**). This disruption is partly caused by altered conformation of the Phe132 and Asn99 side chains, which break the network between residues Asn40 and Asn99. Conversely, the network in the adenosine site was stabilized in the Mac1-ADPr complex (**Figure 2.9B**). The average B-factor decreased from 24 to 10 Å², and two networks connect the phosphate tunnel with the adenine/oxyanion subsites (**Figure 2.9C**). Although the adenine moiety only forms two direct hydrogen bonds to protein, it has four additional contacts via bridging water molecules (**Figure 2.9B**). Similar bridging waters were observed for fragments binding in the adenosine site including ZINC340465 (PDB: 5RSV), which forms only one direct hydrogen bond to the protein, but has an extensive hydrogen-bond network via water molecules (**Figure 2.9D**). Visualizing all water molecules within 3.5 Å of fragment atoms shows clusters near protein hydrogen bond acceptors and donors (**Figure 2.9E**).

Of particular interest is the cluster near the backbone carbonyl of Ala154. This site is occupied by a water molecule in the Mac1-ADPr structure and is bridged by adenine derivatives such as ZINC340465 (PDB: 5RSJ) (**Figure 2.9D**). In addition, five fragments occupy this site directly (**Figure 2.4A,D**), including the C2-amino-substituted adenine present in ZINC89254160_N3 (PDB: 5RSJ, **Figure 2.3D**). Extending fragments to displace the water molecules at other frequently populated sites could help to quantify the contribution of water networks to Mac1-binding, and to provide a test set for computational methods that seek to exploit solvent dynamics for ligand optimization^{47,48}.

Solution binding of fragment hits

To buttress the crystallographic studies, selected compounds were biophysically screened using differential scanning fluorimetry (DSF), isothermal titration calorimetry (ITC), and a homogeneous time-resolved fluorescence (HTRF) ADPr-peptide displacement assay (**Figure 2.10, Supplemental Data 2.1, Supplemental Data 2.2**). Because of their ready availability in useful amounts, most of these experiments focused on the docking hits. For DSF, in agreement with previous reports for this enzyme¹⁸, we observed substantial elevation of the apparent melting temperature (T_{m_a}) upon addition of ADP-ribose (**Figure 2.10C,D,G**). When tested in concentration-response from 0.188 to 3 mM, 10 of 54 docked fragments also induced small, but statistically significant and dose-responsive T_{m_a} elevation (**Figure 2.10C,D,G, Supplemental Data 2.1, Supplemental Data 2.2**). All 10 of these were also observed to bind in the crystallographic studies, providing relatively good agreement between these assays. However, the correlation was incomplete, as the remaining fragments observed by crystallography either decreased the T_{m_a} or had no significant effect (**Supplemental Data 2.1**).

To identify fragments with the most promising binding affinity for optimization, we tested the 19 crystallographically observed docking hits using ITC. Due to their small size, most of these fragments have low binding affinity and release little heat upon binding versus ADP-ribose. Thus, we only observed reliable thermodynamic measurements for four of the 19 fragments. These could be fit to a 1:1 binding stoichiometry with affinities in the low mM range (**Figure 2.10E, Supplemental Data 2.1, Supplemental Data 2.2**), consistent with the DSF results. Furthermore, the compounds measured by ITC that released the greatest amount of heat also induced the most significant T_{m_a} shifts in DSF.

Finally, we tested 57 docking-derived fragments and 18 crystallographic hits from the XChem library in an HTRF-based peptide displacement assay, which monitors displacement of a fluorescently labeled ADPr-conjugated peptide from the active site of Mac1 (**Figure 2.10F,G, Supplemental Data 2.1, Supplemental Data 2.2**). Eight of 57 docking hits (14%) and three of 18 crystallographic hits (17%) inhibited the enzyme with IC_{50} values between 180 μ M - 1 mM, with the most potent fragment being the docking-derived ZINC263392672 (PDB: 5RSG) with an IC_{50} of 183 μ M in this assay. Only five (ZINC3888754 (PDB: 5RSC), ZINC331945 (PDB: 5RSD), ZINC263392672 (PDB: 5RSG), ZINC336438345 (PDB: 5RSE) and ZINC6180281 (PDB: 5RSF), **Figure 2.3**) of the 10 docking hits that stabilized Mac1 as measured by DSF were inhibitory in the ADPr-peptide displacement assay. Interestingly, two docking hits that were not identified as binders by DSF or crystallography, ZINC1337772170 ($IC_{50} = 971$ μ M) and pterin ($IC_{50} = 784$ μ M), were found to be inhibitors in the peptide displacement assay (**Figure 2.10H**). This result might be explained by the use of a detergent in the peptide displacement assay that could increase compound solubility. With its ability to detect inhibition of Mac1, the ADPr-peptide displacement assay proved to be a sensitive and complementary

strategy for further characterization of the fragment hits obtained from the docking and crystallographic screens. Assuming that the HTRF-based peptide displacement assay produced the most reliable inhibition data, we estimated ligand efficiencies from IC_{50} values for hits for which we obtained reasonable dose-response curves. ADP-ribose, with an IC_{50} of 161 nM and 36 heavy atoms, has a ligand efficiency (LE) of 0.26 kcal/mol per non-hydrogen atom. The docking hits ZINC3888754 (PDB: 5RSC, LE=0.26), ZINC336438345 (PDB: 5RSE, LE=0.28), ZINC263392672 (PDB: 5RSG, LE=0.32) and ZINC331945 (PDB: 5RSD, LE=0.38) reveal similar or slightly improved ligand efficiencies, while the highest LE was calculated for the XChem library hit SF005 (PDB: 5S4G, **Figure 2.10H**), with 0.44 kcal/mol per heavy atom.

In summary, all crystallographically confirmed docking hits were tested using three complementary in-solution binding techniques - DSF, ITC, and an HTRF-based peptide displacement assay (**Supplemental Figure 2.9, Supplemental Data 2.1, Supplemental Data 2.2**). ZINC336438345 (PDB: 5RSE), ZINC331945 (PDB: 5RSD), ZINC263392672 (PDB: 5RSG) and ZINC26180281 (PDB: 5RSF) were the only four fragment hits for which binding data could be obtained by all three techniques (**Figure 2.10**). All of these fragments have key hydrogen bonds in the adenine subsite and π - π stack with Phe156. Furthermore, ZINC263392672 (PDB: 5RSG) interacts via its carboxyl group with the oxyanion subsite of Mac1. Finally, we note that crystallography, DSF, ITC all monitor binding, but do not measure function. The peptide displacement assay is thus of particular value for fragment characterization, since it measures specific displacement of an analog of the natural Mac1 substrate.

Opportunities for fragment linking and merging to optimize Mac1 inhibitors

Typically, one might be reluctant to speculate on optimization from fragment structures alone, but the unusually large number of structures perhaps supports some cautious inference here. Prior to modifying, linking, or merging fragments, it is important to consider the crystalline environment. In the P4₃ crystal form, the active site forms a bipartite enclosed pocket with a symmetry mate (**Figure 2.11A,B**). In particular, 24 fragments only hydrogen bond to Lys11 of the symmetry mate, and not with any residues in the adenosine site, indicating that these molecules should not be considered for fragment elaboration (**Figure 2.11C,D**). Based on the binding poses of remaining compounds, fragment pairs were linked into hypothetical scaffolds. These were used as templates to search the make-on-demand chemical space of the Enamine REAL database employing the Smallworld similarity (<https://sw.docking.org>) and Arthor substructure (<https://arthor.docking.org>) search engines (**Figure 2.11E,F**)⁴⁹. In a second approach, fragments with overlapping binding poses were merged into larger scaffolds, e.g. the purine of ZINC89254160_N3 (PDB: 5RSJ) interacting in the adenine binding subsite was replaced by ZINC26180281 (PDB: 5RSF) adding an additional hydrogen bond to Ala154 (**Figure 2.11F**). Whereas it remains speculative whether the suggested linked or merged molecules are indeed active against Mac1, the scaffolds observed here, and the key interactions they make with the enzyme, indicate a fruitful chemical space to further explore. Naturally, many of the fragments described here also merit investigation by alternative fragment growing or analoging strategies.

Discussion

Three key observations emerge from this study. Most noteworthy is the sheer number and the unusually high resolution of the 234 fragment-bound Mac1 structures, including 192 fragments identified in the active site. The fragments cover both stereotypical interactions (such as adenine-like hydrogen bonding to the Asp22 side chain/ Ile23 backbone and stacking interaction with Phe156) as well as diverse and unusual chemotypes that exploit active site flexibility (for instance by targeting the oxyanion-subsite). This abundance and diversity afford multiple starting points for future elaboration into lead-like molecules. Second, the high fidelity of docked poses to the subsequent crystallographic results supports the use of docking to explore the adenine recognition site, and importantly, demonstrates an ability of docking to prioritize fragments, at least for this target, something still debated in the field. Finally, with 234 diverse fragment structures determined, it should be possible to exploit the fortuitous juxtaposition of fragment pairs to design joined ligands that combine the affinities of both, leading to inhibitors with the low micromolar affinity needed for hit-to-lead optimization. One clear strategy involves extending molecules bound to the adenine subsite and with biophysically measurable binding affinities into the phosphate and ribose recognition regions.

In contrast to the large number of chemically diverse hits binding to the adenine subsite, the lack of fragments bound to the catalytic site is notable and may inform models of how ADP-ribosylated peptides bind to Mac1. The paucity of fragments is especially surprising given that three crystal environments (the A and B chains in the P4₃ crystal and the C2 crystal) were screened and that the site appears accessible in all lattices (**Supplemental Figure 2.2B**). The two major models for peptide-macrodomein interactions are either that the peptide binds along the widened cleft defined by Tyr42 and Lys102, or that it extends into solution through the flexible

Gly46-48 loop⁵⁰. Indeed, we observe fragments that bind in both locations (**Figure 2.7A**). Regardless of the binding mode, which could be distinct depending on the identity of the modified residue and target substrate, the lack of binding at this site suggests that the binding energy comes mostly from the ADP-ribose and not from the amino acids on the ADPr-conjugated protein. This hypothesis is also supported by the fact that Mac1 can hydrolyze a wide range of ADP-ribosylated substrates^{2,51}. Docking of larger ‘lead-like’ molecules, perhaps enabled by the expanded catalytic site revealed by the physiological temperature structure, and detailed description of solvent, may help to identify molecules exploiting this site.

The success of the fragment docking campaign contrasts, perhaps, with expectations of the field that fragments have too few functional-group handles to accurately dock or prioritize⁵². Not only were hit rates high (33%), so too was the fidelity of most docking poses to the crystallographic results. Even judged by potency, the most active fragment to emerge from this study, the 183 μ M inhibitor ZINC263392672 (PDB: 5RSG) (**Figure 2.3, Figure 2.10**), was a docking hit. Also, it was the docking hits that were most readily available for such functional testing, as they were sourced in 10 mg amounts, while the crystallographic screening compounds were often in short supply. This is a purely mechanical advantage of docking, and it is counterbalanced by the small numbers tested versus the crystallographic screens; still, having substantial material to work with is a pragmatic advantage. Admittedly, weaknesses also emerged from the docking. Intriguingly, the oxyanion site that featured so prominently among the crystallographic screening hits were not to be found among the docking predictions. This gap reflects both a failure of the docking scoring function to prioritize anions binding to this site (as they were at least sampled), and to some extent a failure of the docking group to pick the few molecules that did dock well to this site as likely candidates. More broadly, as we docked against

a single rigid structure of the protein, the subsequent conformational changes that the protein underwent, and the changes in the water network, were not captured in the docking predictions, and this was sometimes reflected in the larger RMSD differences between predicted and observed fragment poses (**Figure 2.3**). These caveats, important as they are, should not obscure a central observation from this study: the docking hit rate was not only high, but the hits were typically right for the right reasons; this may be something to build on for the field.

From the docked compounds, the most promising hits identified by in-solution binding experiments were also crystallographically confirmed. However, as expected, the majority of hits from crystallography did not show appreciable activity in the orthogonal biophysical assays within the tested concentration range (up to 10 mM in ITC, **Supplemental Data 2.1**). The macrodomain ADPr-peptide displacement assay also identified two docking hits not previously observed in soaking (ZINC1337772170 and pterin), which suggests that the crystal environment limited the ability of some fragments to bind. Yet, between solution experiments good consensus was observed for ZINC263392672 (PDB: 5RSG), ZINC336438345 (PDB: 5RSE) and ZINC331945 (PDB: 5RSD). While we are aware that obtaining high-quality binding data remains particularly challenging for weak binders such as fragments, the dose-response results obtained in the complementary assays for many of the identified hits provided convincing evidence for their true binding to Mac1. The inconsistency of fragment binding to different crystal systems of the same protein is apparent when comparing fragments that resulted in high quality data sets in both the P4₃ and C2 crystal systems. Surprisingly, only five of 59 possible fragments were observed in both systems, with three fragments binding with equivalent poses in the adenine subsite. This observation points to the value of having multiple measurements, and

even multiple crystal systems when they are available, in fragment-based drug discovery approaches.

Overall, this study has three main implications for the discovery of SARS-CoV-2 Nsp3 Mac1 inhibitors, and for antiviral efforts targeting macrodomains more broadly. First, we describe not only the first new chemical matter for this target but map its hotspots at high resolution. This provides a template for future inhibitor discovery and development against this enzyme. Such efforts will need to navigate selectivity over human macrodomains and other ATP-binding proteins including kinases (**Supplemental Figure 2.8**) and consider breadth across other viral macrodomains¹² (**Supplemental Figure 2.4**). Second, the specific fragments that we describe may lend themselves directly to optimization: several examples are discussed explicitly, amenable to make-on-demand chemistry (**Figure 2.11**); and the 234 structures should provide inspiration for countless other molecules. Finally, important technical advances emerged from this study: a crystal form that lends itself to ready structure determination, the creation of a reliable peptide-displacement assay for Mac1, and evidence supporting the ability of structure-based screening, such as molecular docking, to predict effective fragments. The ultra-high resolution X-ray diffraction data, which allowed hydrogen atoms to be refined explicitly, and electron density to be resolved on a subatomic scale, makes Mac1 an attractive candidate for in-depth computational dissection of its catalytic mechanism using approaches that integrate both classical and quantum calculations. Taken together, these advances will speed progress throughout the community to help validate this target and create effective antivirals.

Figures

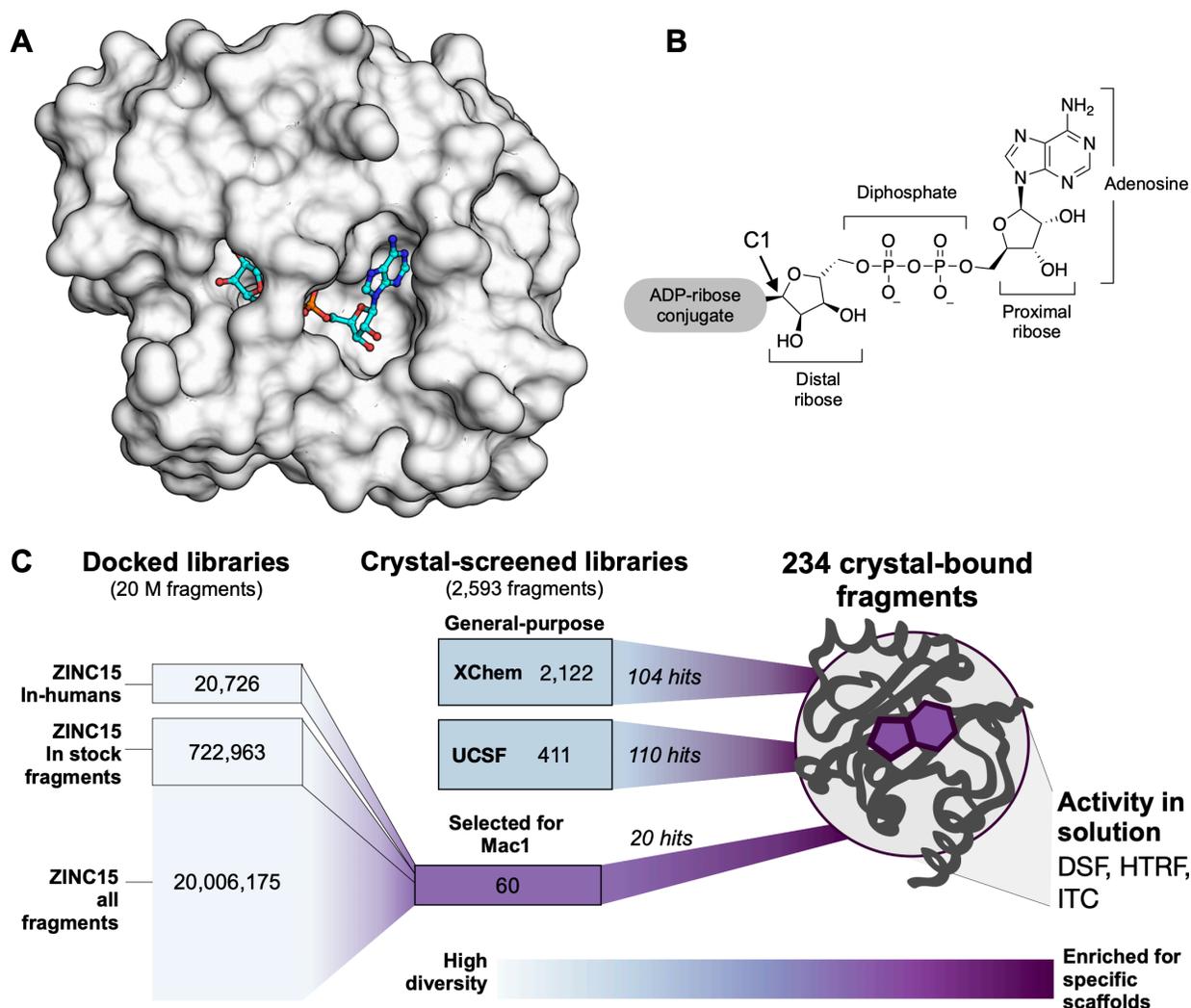


Figure 2.1 | Overview of the fragment discovery approach for SARS-CoV-2 Nsp3 Mac1 presented in this study.

A) Surface representation of Nsp3 Mac1 with ADP-ribose bound (cyan) in a deep and open binding cleft. **B)** Nsp3 Mac1 possesses ADP-ribosylhydrolase activity which removes ADP-ribosylation modifications attached to host and pathogen targets. ADP-ribose is conjugated through C1 of the distal ribose. **C)** Summary of the fragment discovery campaign presented in this work. Three fragment libraries were screened by crystallography: two general-purpose (XChem and UCSF), and a third bespoke library of 60 compounds, curated for Mac1 by molecular docking of over 20 M fragments. Crystallographic studies identified 214 unique fragments binding to Mac1, while the molecular docking effort yielded in 20 crystallographically confirmed hits. Several crystallographic and docking fragments were validated by ITC, DSF, and a HTRF-based ADPr-peptide displacement assay.

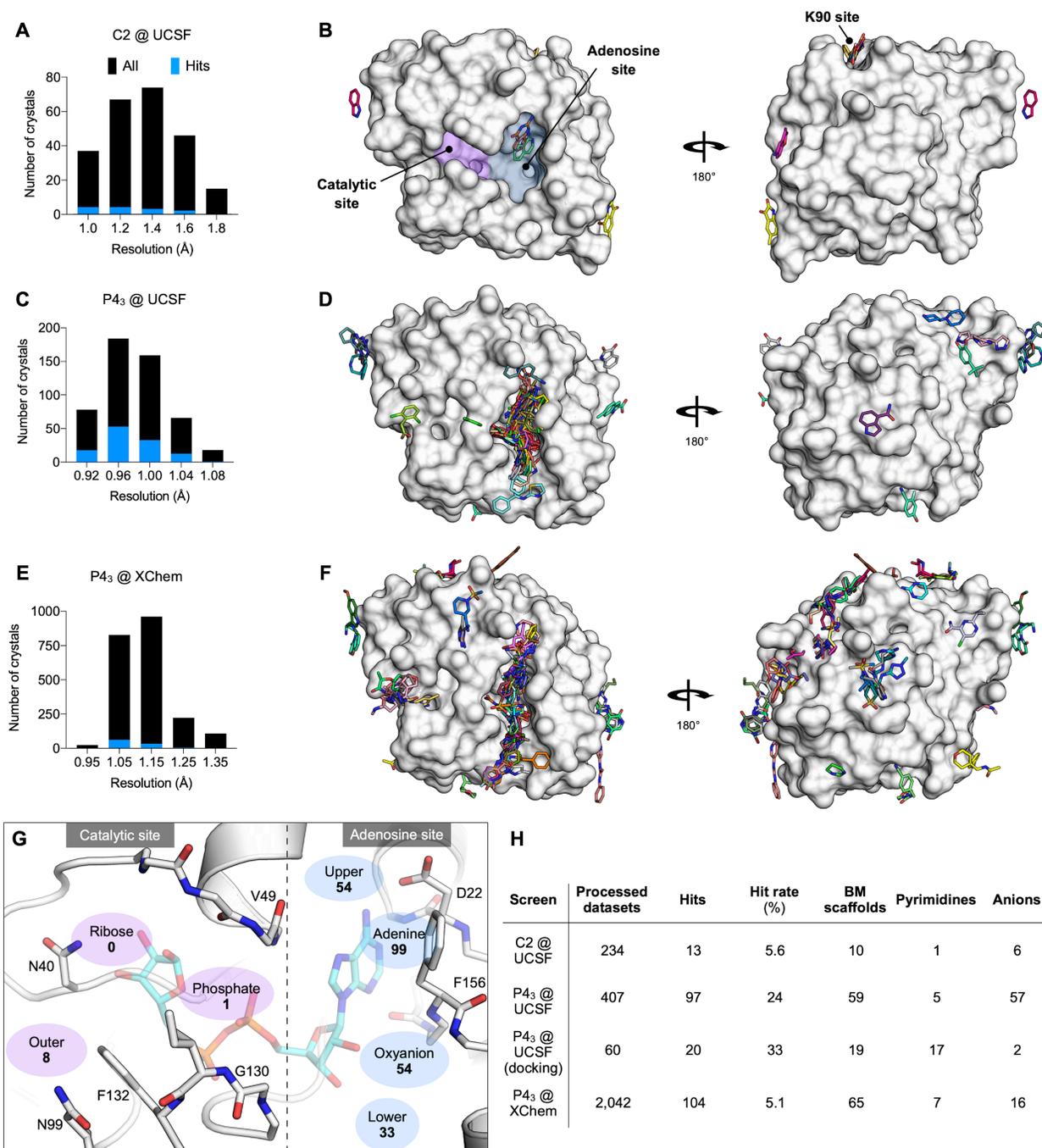


Figure 2.2 | Crystallographic screening identified 234 fragments bound to Mac1.

A,C,E) Histograms showing the resolution of the crystallographic fragment screening data. The resolution of datasets where fragments were identified are shown with blue bars. **B,D,F)** Surface representation of Mac1 with fragments shown as sticks. **G)** The Mac1 active site can be divided based on the interactions made with ADP-ribose. The ‘catalytic’ site recognizes the distal ribose and phosphate portion of the ADP-ribose, and harbors the catalytic residue Asn40¹⁰. The ‘adenosine’ site recognizes adenine and the proximal ribose. The number of fragments binding in each site is indicated. **H)** Summary of the fragments screened by X-ray crystallography, including the number of Bemis-Murcko (BM) scaffolds and anionic fragments identified as hits

in each screen. “Processed datasets” refers to the number of datasets that were analyzed for fragment binding with PanDDA. Out of the datasets collected for 2,954 fragments, 211 (7.1%) were not analyzed due to data pathologies.

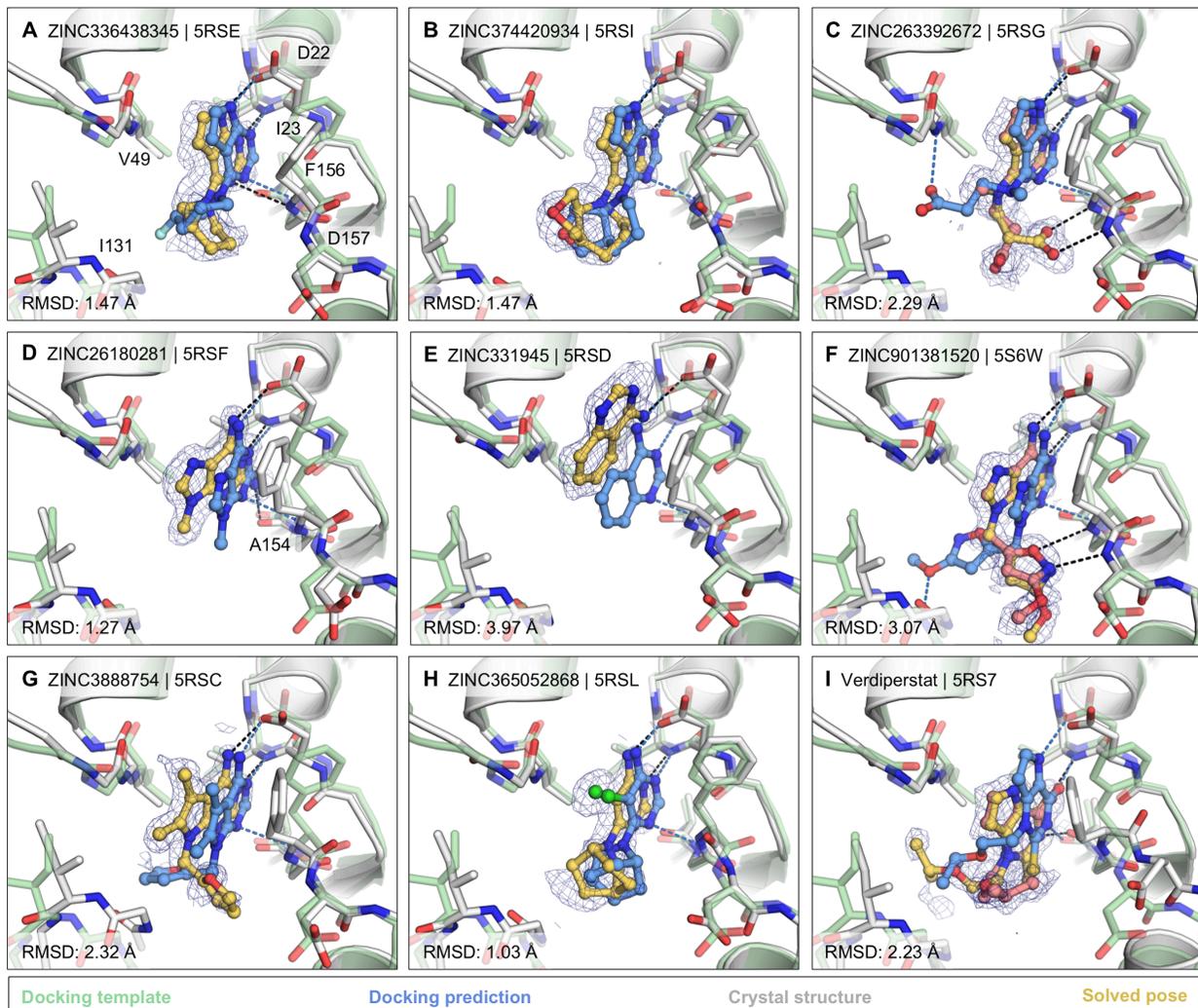


Figure 2.3 | Docking hits confirmed by high resolution crystal structures.

The protein structure (PDB: 6W02)³⁴, prepared for virtual screens is shown in green, predicted binding poses are shown in blue, the crystal protein structures are shown in grey, the solved fragment poses are shown in yellow, with alternative conformations shown in light pink. PanDDA event maps are shown as a blue mesh. Event maps were calculated prior to ligand modeling, and the maps are free from model bias towards any ligand³⁹. Protein-ligand hydrogen bonds predicted by docking or observed in crystal structures are colored light blue or black, respectively. Hungarian RMSD values are presented between docked and crystallographically determined ligand poses (binding poses for additional docking hits are shown in **Supplemental Figure 2.7**).

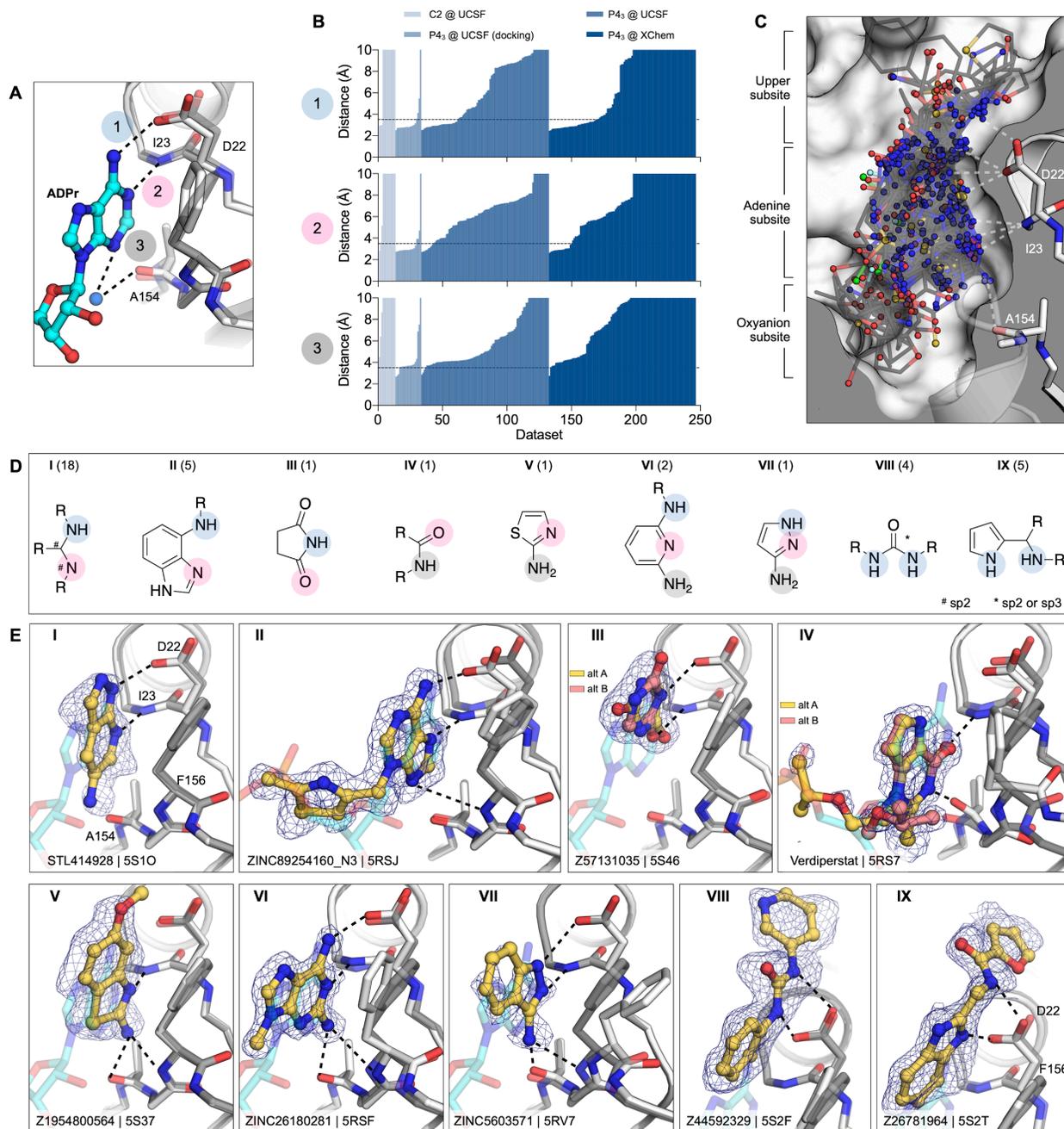


Figure 2.4 | Fragments binding to the adenine subsite.

A) Stick representation showing the interaction of the adenosine moiety of ADP-ribose with Mac1. The key hydrogen bonds are shown as dashed lines. **B)** Plot of the distances shown in **(A)** for all fragment hits. The distances, truncated to 10 Å, are for the closest non-carbon fragment atom. **C)** Stick representation showing all fragments interacting with Asp22-N, Ile23-N or Ala154-O. The surface is ‘sliced’ down a plane passing through Asp22. **D)** Structures of the nine unique motifs that make at least two hydrogen bonds to the adenine subsite. Colored circles match the interactions listed in **(A)** and **(B)**. The number of fragments identified for each motif are listed in parentheses. **E)** Examples of the nine structural motifs. The fragment is shown with yellow sticks and the PanDDA event map is shown as a blue mesh. ADP-ribose is shown as cyan transparent sticks. The apo structure is shown with dark gray transparent sticks.

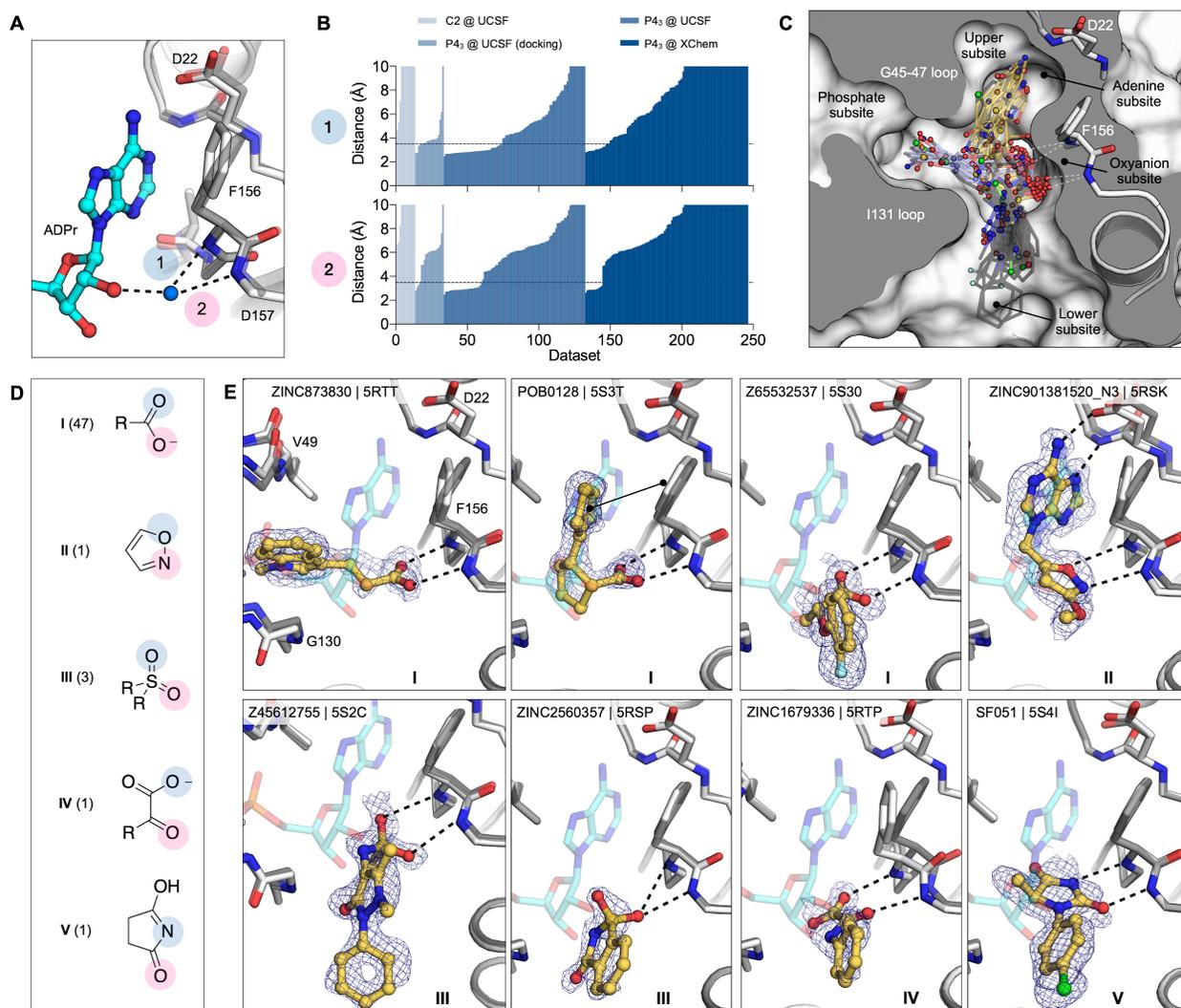


Figure 2.5 | Fragments binding to the oxyanion subsite.

A) Stick representation showing the interaction of ADP-ribose with the oxyanion subsite of Mac1. The water molecule bridging the ribose moiety and the oxyanion subsite is shown as a blue sphere. **B)** Plot of the distances highlighted in **(A)** for all fragment hits. Distances were calculated as described for **Figure 2.6B**. **C)** Stick representation showing all fragments interacting with Phe156-N and Asp157-N. Fragments are colored by secondary binding site with blue = phosphate, black = lower and yellow = adenine. The surface is “sliced” across a plane passing through Phe156 (white surface, grey interior). **D)** Structures of the five structural motifs that bind the oxyanion site. **E)** Examples of the five motifs. Three examples of motif I are shown, where the fragment also interacts with the phosphate, adenine or lower subsites. The fragment is shown with yellow sticks and the PanDDA event map is shown for reference as a blue mesh. ADP-ribose is shown with transparent cyan sticks. The apo structure is shown with transparent gray sticks.

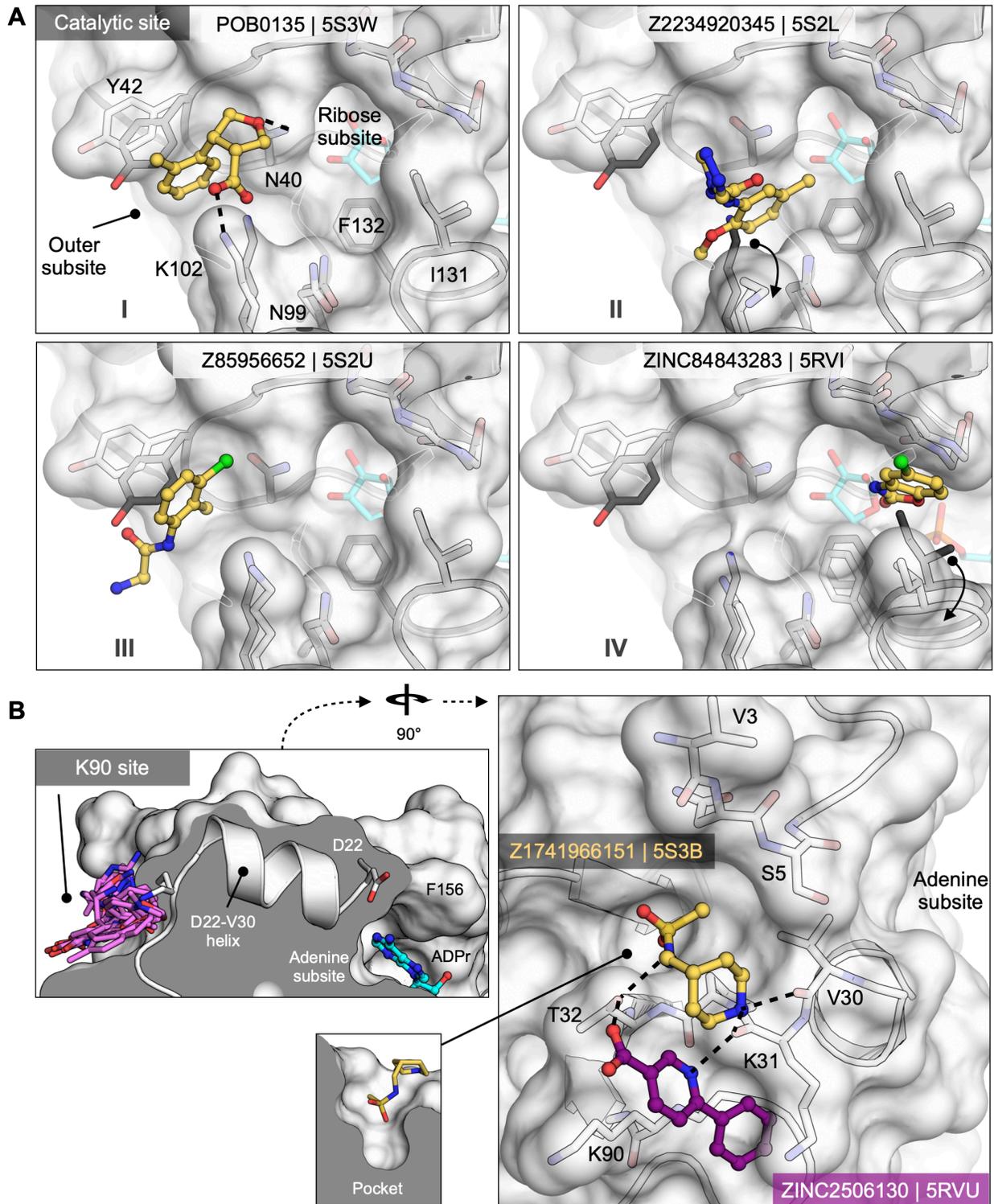


Figure 2.7 | Fragments targeting the catalytic and potential allosteric sites are sparsely populated compared to the adenosine site.

A) Surface representation showing fragments that bind near the catalytic site. The fragment POB0135 (PDB: 5S3W) bridges the gap between Asn40 and Lys102 via a hydrogen bond and a salt bridge, respectively. Although eight fragments bind in the outer subsite, the fragment

POB0135 makes the highest quality interactions. No fragments bind in the ribose subsite. The fragment in ZINC331715 (PDB: 5RVI) inserts into the phosphate subsite between Ile131 and Gly47. **B)** Left: the K90 site is connected to the adenosine site by the D22-V30 alpha-helix. Right: surface representation showing two fragments that bind to the K90 site. Hydrogen bonds are shown as dashed black lines. The fragment in Z1741966151 (PDB: 5S3B) is partially inserted in a nearby pocket (insert).

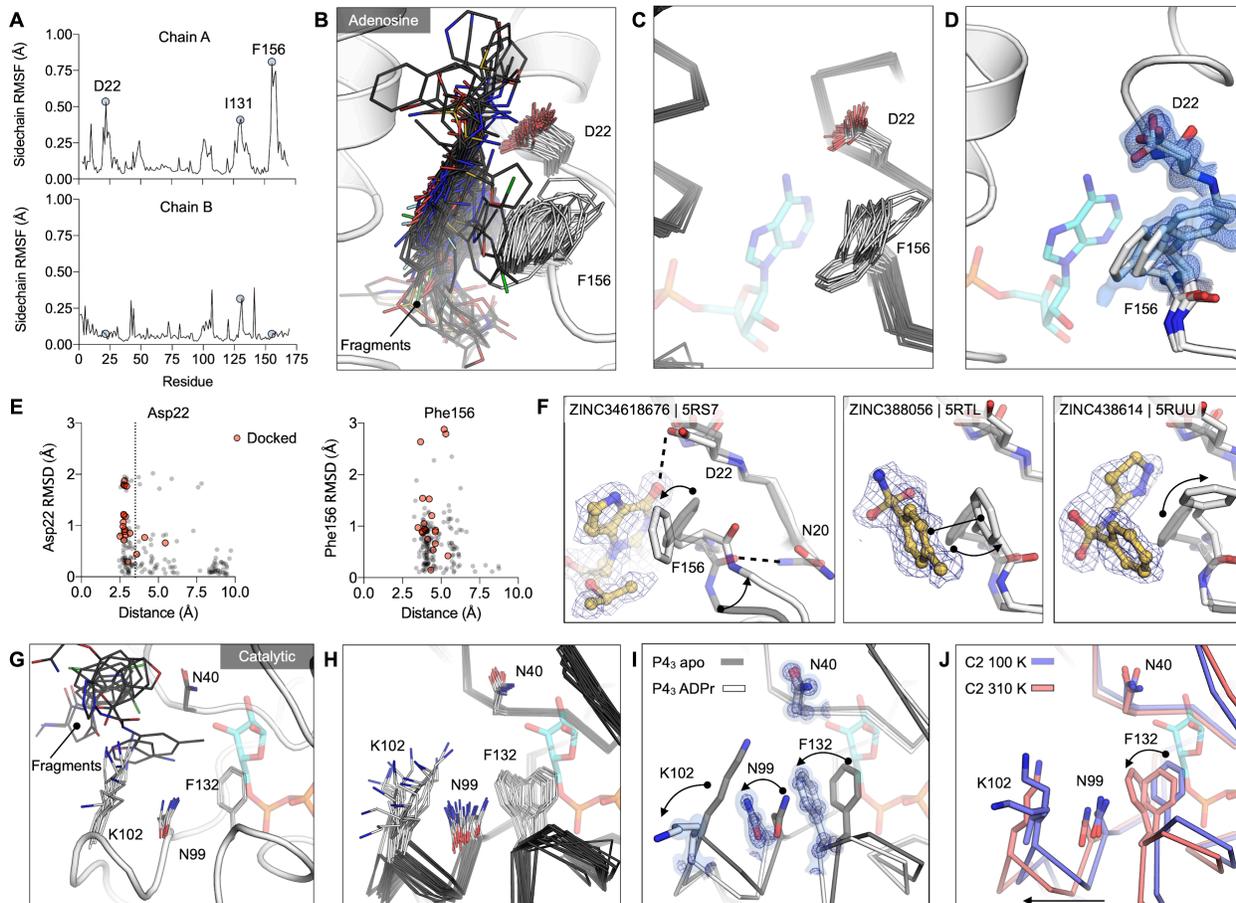


Figure 2.8 | Experimentally observed conformational heterogeneity is sampled by various fragments.

A) Plots of side chain RMSF for the 117 fragment structures from the UCSF screen using P4₃ crystals. **B)** Stick representation showing all fragments (black sticks) within 3.5 Å of the Asp22 carboxylate and 4 Å of the Phe156 ring (white sticks). **C)** Structural heterogeneity in the previously reported Mac1 structures. **D)** The Phe156 side chain is captured in three conformations in the C2 apo structure. Electron density maps ($2mF_o - DF_C$) are contoured at 0.5σ (blue surface) and 1σ (blue mesh). For reference, ADP-ribose is shown with blue sticks. **E)** Plots of side chain RMSD for Asp22 and Phe156 from the Mac1 apo structure as a function of ligand-protein distance. Structures were aligned by their C α atoms, before RMSDs were calculated for the Asp22 carboxylate and the Phe156 aromatic carbons. **F)** Fragment binding exploits preexisting conformational heterogeneity in the Phe156 side chain. The apo structure is shown with dark transparent gray sticks in each panel and the conformational changes are annotated with arrows. **G)** Stick representation showing all fragments (black sticks) in the outer subsite of the catalytic site. **H)** Conformational heterogeneity of residues in the catalytic site of the previously reported Mac1 crystal structures. **I)** ADP-ribose binding induces a coupled conformational change in the Phe132, Asn99 and Lys102 side chains, as well as a 2 Å shift in the Phe132 loop. Electron density maps ($2mF_o - DF_C$) are contoured at 1.5σ (blue surface) and 4σ (blue mesh). **J)** Mac1 structures determined at 100 K and 310 K using C2 crystals.

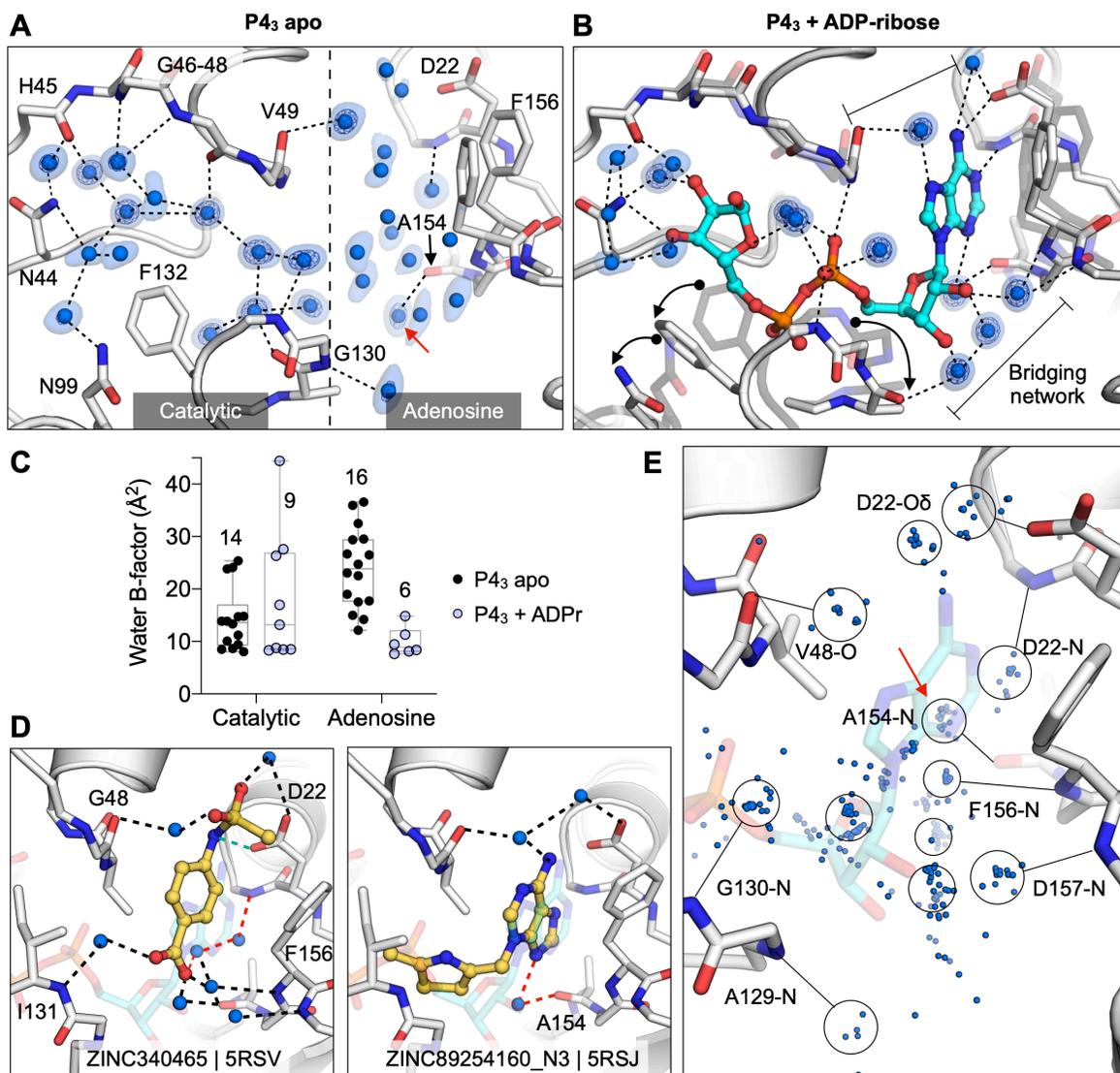


Figure 2.9 | Water networks in the active site are displaced as well as used by fragments for bridging interactions.

A) Water networks in the apo enzyme ($P4_3$ crystal form). Waters are shown as blue spheres, with electron density contoured at 5.0σ (blue mesh) and 1.5σ (blue surface). Hydrogen bonds are shown as dashed lines (distances are 2.6-3 \AA). **B)** Water networks in the Mac1-ADPr complex. ADP-ribose is shown as cyan sticks. Conformational changes upon ADP-ribose binding are highlighted with black arrows. **C)** Comparison of crystallographic B-factors of water molecules in the catalytic site and adenosine site. The range and 95% confidence interval are shown. **D)** Examples of the role of water networks in fragment binding. Left: ZINC340465 (PDB: 5RSV) forms a single hydrogen bond to the protein (green dashed line) but forms five hydrogen bonds via water molecules. Right: although few fragments hydrogen bond directly to the backbone oxygen of Ala154, several fragments interact with this residue via bridging water molecules (red dashed line) including ZINC89254160_N3 (PDB: 5RSJ). **E)** Plot showing all water molecules that lie within 3.5 \AA of a non-carbon fragment atom. Water molecules are shown as blue spheres, with the major clusters circled. The cluster in a red circle bridges fragments and the Ala154 backbone oxygen.

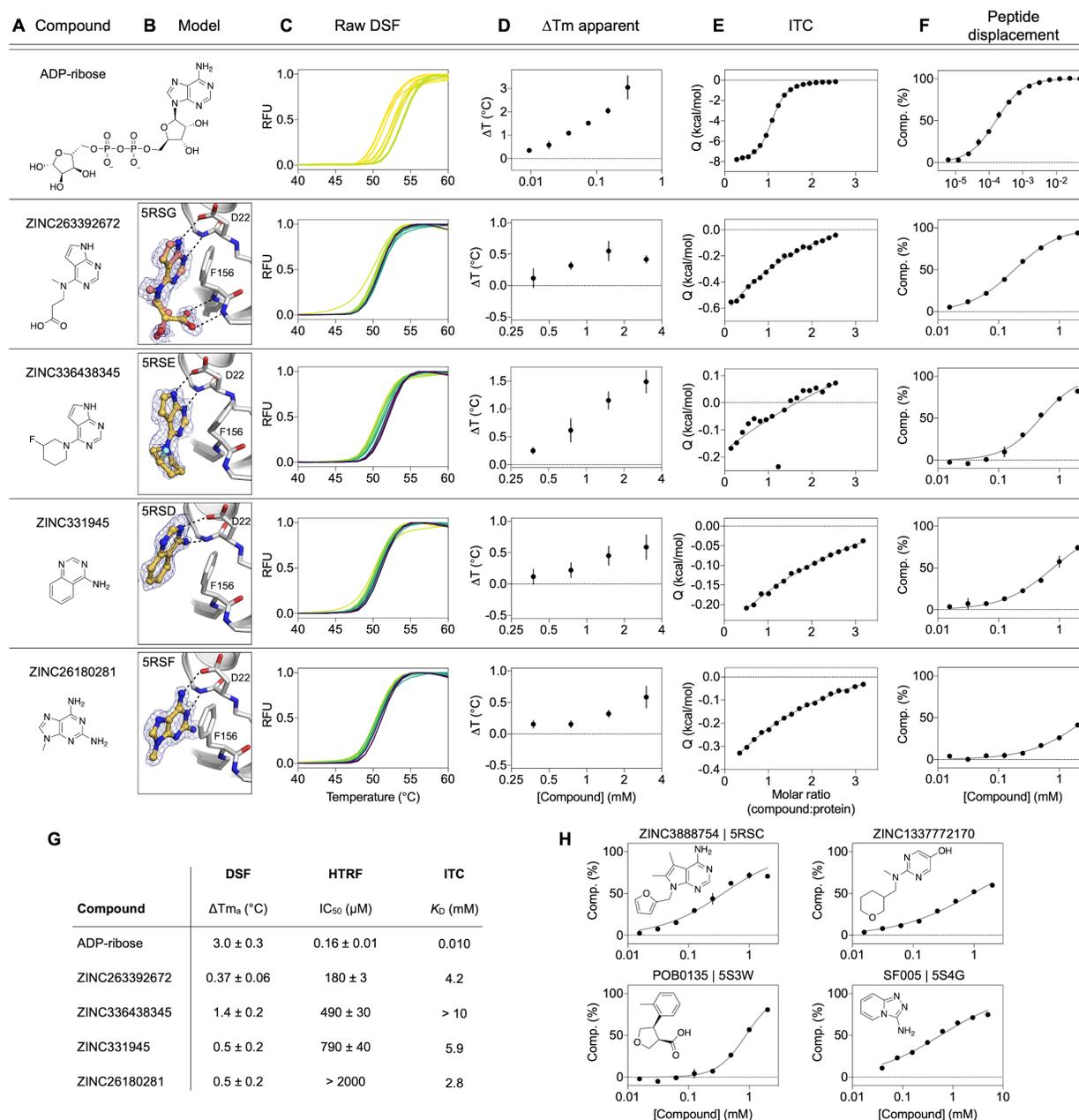


Figure 2.10 | Biophysical corroboration of solution binding of crystallographic fragment hits by DSF, ITC and ADPr-peptide displacement assay.

Top panel (**A-F**) shows performance of the most potent fragment hits in DSF, ITC, and ADPr-peptide displacement assay compared to ADP-ribose. **C,D**) Normalized raw DSF RFU data demonstrates canonical unfolding curves and minimal compound-associated curve shape aberrations. T_m elevation reveals Mac1 stabilization through fragment binding. Gradient color scale: 0 mM = yellow; 3 mM = purple. **E**) Integrated heat peaks measured by ITC as a function of binding site saturation. The black line represents a non-linear least squares (NLLS) fit using a single-site binding model. **F**) Peptide displacement assay measures ADPr-peptide displacement (i.e., % competition) from Mac1 by ligand. **G**) Summary of solution binding data for fragments from top panel. ΔT_m are given for the highest compound concentration in this assay. **H**) Additional fragment hits showing Mac1 peptide competition.

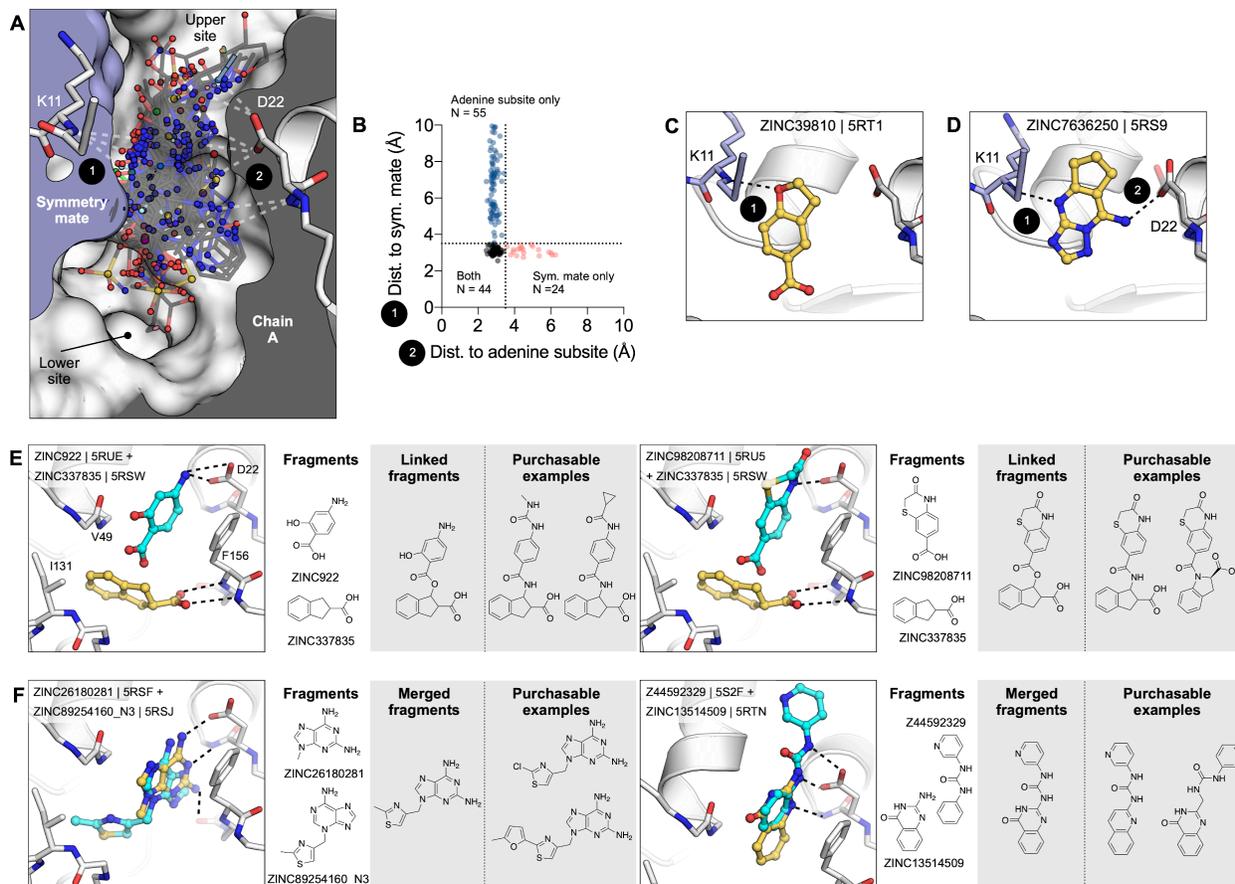


Figure 2.11 | Fragments bridging multiple adenosine sites provide direct merging opportunities.

A) Sliced view of the adenosine site (white surface, grey interior) and a symmetry mate (blue surface and interior) showing the deep pocket created by crystal packing in the P₄₃ crystals. The 66 fragments that hydrogen bond with the Lys11 backbone nitrogen are shown as sticks. **B)** Plot showing distances between the symmetry mate (Lys11-N) and the adenine subsite (Asp22-O δ , Ile23-N, Ala154-O) for all fragments identified in the adenosine site. Dashed lines show the 3.5 Å cut-off used to classify hydrogen-bonds. **C)** An example showing one of the 24 fragments that bound in the adenosine site, yet only formed a hydrogen bond with the symmetry mate. **D)** An example of one of the fragments that bridged the 9-11 Å gap between the adenine subsite and the symmetry mate. **E, F)** Opportunities for fragment linking and merging. Adjacent or overlapping fragments were initially merged into a single new compound. Examples of readily available make-on-demand compounds are shown.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and the **Supplemental Materials**. Crystallographic coordinates and structure factors for all structures have been deposited in the Protein Data Bank with the following accessing codes: 7KR0, 7KR1, 7KQW, 7KQO, 7KQP, 5RVJ, 5RVK, 5RVL, 5RVM, 5RVN, 5RVO, 5RVP, 5RVQ, 5RVR, 5RVS, 5RVT, 5RVU, 5RVV, 5RS7, 5RS8, 5RS9, 5RSB, 5RSC, 5RSD, 5RSE, 5RSF, 5RSG, 5RSH, 5RSI, 5RSJ, 5RSK, 5RSL, 5RSM, 5RSN, 5RSO, 5RSP, 5RSQ, 5RSR, 5RSS, 5RST, 5RSU, 5RSV, 5RSW, 5RSX, 5RSY, 5RSZ, 5RT0, 5RT1, 5RT2, 5RT3, 5RT4, 5RT5, 5RT6, 5RT7, 5RT8, 5RT9, 5RTA, 5RTB, 5RTC, 5RTD, 5RTE, 5RTF, 5RTG, 5RTH, 5RTI, 5RTJ, 5RTK, 5RTL, 5RTM, 5RTN, 5RTO, 5RTP, 5RTQ, 5RTR, 5RTS, 5RTT, 5RTU, 5RTV, 5RTW, 5RTX, 5RTY, 5RTZ, 5RU0, 5RU1, 5RU2, 5RU3, 5RU4, 5RU5, 5RU6, 5RU7, 5RU8, 5RU9, 5RUA, 5RUC, 5RUD, 5RUE, 5RUF, 5RUG, 5RUH, 5RUI, 5RUJ, 5RUK, 5RUL, 5RUM, 5RUN, 5RUO, 5RUP, 5RUQ, 5RUR, 5RUS, 5RUT, 5RUU, 5RUV, 5RUW, 5RUX, 5RUY, 5RUZ, 5RV0, 5RV1, 5RV2, 5RV3, 5RV4, 5RV5, 5RV6, 5RV7, 5RV8, 5RV9, 5RVA, 5RVB, 5RVC, 5RVD, 5RVE, 5RVF, 5RVG, 5RVH, 5RVI, 5S6W, 5S18, 5S1A, 5S1C, 5S1E, 5S1G, 5S1I, 5S1K, 5S1M, 5S1O, 5S1Q, 5S1S, 5S1U, 5S1W, 5S1Y, 5S20, 5S22, 5S24, 5S26, 5S27, 5S28, 5S29, 5S2A, 5S2B, 5S2C, 5S2D, 5S2E, 5S2F, 5S2G, 5S2H, 5S2I, 5S2J, 5S2K, 5S2L, 5S2M, 5S2N, 5S2O, 5S2P, 5S2Q, 5S2R, 5S2S, 5S2T, 5S2U, 5S2V, 5S2W, 5S2X, 5S2Y, 5S2Z, 5S30, 5S31, 5S32, 5S33, 5S34, 5S35, 5S36, 5S37, 5S38, 5S39, 5S3A, 5S3B, 5S3C, 5S3D, 5S3E, 5S3F, 5S3G, 5S3H, 5S3I, 5S3J, 5S3K, 5S3L, 5S3M, 5S3N, 5S3O, 5S3P, 5S3Q, 5S3R, 5S3S, 5S3T, 5S3U, 5S3V, 5S3W, 5S3X, 5S3Y, 5S3Z, 5S40, 5S41, 5S42, 5S43, 5S44, 5S45, 5S46, 5S47, 5S48, 5S49, 5S4A, 5S4B, 5S4C, 5S4D, 5S4E, 5S4F, 5S4G, 5S4H, 5S4I, 5S4J, 5S4K.

Methods

Fragment libraries

We screened 2,122 molecules from the XChem facility at Diamond Light Source against the Mac1 P4₃ crystal form, and 411 molecules from UCSF against the C2 and P4₃ crystal forms (**Supplemental Data 2.1**). The fragment library at XChem combined molecules from multiple fragment libraries: the Diamond, SGC and iNEXT (DSI)-poised Library (687 molecules³⁶), the Edelris fragment collection (132 molecules), the MiniFrag Probing Library (80 molecules⁵³), the FragLites collection (31 compounds⁵⁴), the PepLite library (22 molecules²⁶), the SpotFinder library (96 compounds), the York3D library (106 molecules⁵⁵) and the EU Open screen (968 molecules). The UCSF fragment library was composed of Enamine's Essential Fragment library (320 compounds) and 91 additional compounds from an in-house library (UCSF_91). To assemble the UCSF_91 library, we selected topologically diverse molecules having over 10,000 commercially available analogs in at least three points of substitution, allowing for rapid and extensive analog-by-catalog without having to resort to flask synthesis. We picked molecules that were also Bemis-Murcko scaffolds³⁸, stripped of acyclic terminal substituents. We thought simple, unsubstituted frameworks would be easier to optimize by adding chemical matter during analoging. From among these, we prioritized by eye scaffolds with various ring sizes and combinations including fused rings, spiro systems, with linkers of varying lengths between rings, in an attempt to sample a diverse range of compact shapes and properties. We added anions where the anionic moiety was a small acyclic substituent on the scaffold, again picking by eye for shape diversity. We chose molecules with 11-21 heavy atoms, with molecular weights between 200-300 amu and with a logP < 2.5 for solubility. Physical properties of all screened libraries are shown in **Supplemental Figure 2.5**.

Analyses of scaffolds and specific chemotypes in the employed chemical libraries are shown in **Supplemental Figure 2.5E**. Bemis-Murcko (BM) scaffold analysis was performed with the Molinspiration mib engine (<https://www.molinspiration.com>). Pyrimidines were identified using RDKit (<https://www.rdkit.org>) and molecular charges at pH 7.4 were approximated using ChemAxon Jchem version 2019.15 (<https://www.chemaxon.com/>) to identify anionic fragments.

C2 crystals at UCSF

Protein expression and purification

SARS-CoV-2 Nsp3 Mac1 (residues 2-170) was cloned into a pET22b(+) expression vector with an N-terminal His₆ tag and a TEV protease recognition site for removal of the tag (GenScript). In addition, a short linker (Asn-Ala-Gly) was included between the TEV recognition site and the Mac1 gene (**Supplemental Data 2.1**). To express Mac1, plasmid DNA was transformed into BL21(DE3) *E. coli*. After overnight growth on lysogeny broth (LB) agar supplemented with carbenicillin (100 ug/ml), starter cultures (10 mL LB) were grown at 37°C for 8 hours. Large scale cultures (1 l terrific broth (TB)) were grown at 37°C until an optical density of 0.8. Cultures were cooled at 4°C for 15 minutes, before protein expression was induced with 1 mM isopropyl 1-thio-D-galactopyranoside (IPTG), and the cultures were shaken at 20°C for 12 hours. Cells were collected by centrifugation and frozen at -80°C.

All purification steps were performed at 4°C using an AKTA FPLC system (Cytiva). Cells were resuspended in Ni-NTA binding buffer (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, 5% glycerol, 2 mM βME supplemented with 5 units/ml TurboNuclease (Sigma, T4330)) and lysed by sonication. Cell debris was collected by centrifugation and the lysate was

applied to a 5 mL HisTrap HP column (Cytiva, 17524802). The column was washed with 25 mL binding buffer followed by 25 mL 5% Ni-NTA elution buffer (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 500 mM imidazole, 5% glycerol, 2 mM β -ME), and then eluted with 100% elution buffer. Eluted protein was exchanged into TEV reaction buffer (50 mM Tris (pH 8.0), 100 mM NaCl, 1 mM DTT and 1% glycerol) using a HiPrep 26/10 desalting column (Cytiva, 17508701). To cleave the His₆ tag, Mac1 was diluted to 1.5 mg/ml using TEV reaction buffer and incubated with recombinant TEV protease⁵⁶ at a 1:20 ratio (Mac1:TEV) for 16 hours at 4°C. Cleaved Mac1 was separated from the uncleaved protein and TEV protease by re-running the sample over a HisTrap HP column (pre-equilibrated with TEV reaction buffer) and collecting the flow-through. The flow-through was supplemented with 10 mM DTT and concentrated to 2.5 mL using a 10 kDa molecular weight cut-off (MWCO) centrifugal concentrator (Amicon, UFC901024). The sample was further purified by size-exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 pg column (Cytiva, 28989333) equilibrated with SEC buffer (20 mM Tris (pH 8.0), 150 mM NaCl, 5% glycerol, 2 mM DTT). Eluted fractions were concentrated to 15 mg/ml and flash-frozen in liquid nitrogen and stored at -80°C. Protein used for ITC was purified in the same manner, but the SEC was run with 150 mM NaCl, 20 mM Tris (pH 8.0). Protein was concentrated to 10.8 mg/ml prior to flash freezing in liquid nitrogen and storage at -80°C.

Crystallization

Crystals were grown at 19°C using sitting-drop vapor diffusion with a reservoir solution containing 100 mM Tris (pH 8.5), 100 mM sodium acetate and 28% PEG 4000. Crystallization drops were set up with 200 nL protein and 200 nL reservoir. Initially, crystals were grown in MRC 2-well plates (SwissCI, MRC96TUVP) with a reservoir volume of 40 μ L. Crystals grew to

a maximum size after 1-2 days and were vitrified in liquid nitrogen without additional cryoprotection. For diffraction experiments at physiological temperatures, crystals were mounted using ALS-style goniometer bases (Mitegen, GB-B3S) and sealed with plastic capillary and vacuum grease (Mitegen, RT-T1). The capillary contained 4 μ l reservoir solution to prevent crystal dehydration.

Fragment soaking was performed using crystals grown with SwissCI 3-well plates (SwissCi, 3W96T-UVP). Microseeding was required to achieve consistent nucleation. Several large crystals grown in 100 mM Tris (pH 8.5), 100 mM sodium acetate and 28% PEG 4000 were transferred to a drop containing 5 μ l seed storage buffer (100 mM Tris (pH 8.5), 100 mM sodium acetate, 32% PEG 4000, 2 mM DTT) on a silicon coverslip (Hampton Research, HR3-233). Crystals were crushed using a flattened glass rod and transferred to 200 μ l of seed storage buffer, before being serially diluted 1:10 with seed storage buffer. Consistent nucleation was achieved with seeds at a 1:100 dilution, with crystallization drops containing 200 nL reservoir, 100 nl seed stock and 300 nl protein with 30 μ l in each reservoir.

Crystal dehydration and fragment soaking

Fragments were added to crystallization drops using acoustic dispensing with an Echo 650 liquid handler (Labcyte)²³. Two libraries were soaked at UCSF: the Enamine Essential fragment library (Enamine, 320 fragments), and the UCSF_91 library (91 fragments) (**Supplemental Data 2.1**). To limit DMSO-induced crystal damage, fragments were targeted to crystallization drops as far away from crystals as possible²³. Initial DMSO tolerance tests indicated that the C2 crystals were sensitive, rapidly disintegrating upon soaking with 10% DMSO (**Supplemental Figure 2.2B**). To enhance DMSO tolerance, 300 nl of a solution

containing 35% PEG 4000, 100 mM Tris (pH 8.5) and 100 mM sodium acetate was added to drops containing crystals using the Echo. Plates were resealed and incubated at 19°C for 6 hours. Fragment solutions (120 nl, 10% of the drop volume) were added using the Echo, and plates were re-sealed and incubated at 20°C for 3-8 hours. Crystals were vitrified directly from crystallization drops without additional cryoprotection.

Lysine methylation

Lysine methylation is a routine strategy for altering the crystallization properties of a protein³⁵. All reagents were added with the protein on ice and incubation steps were performed at 4°C with gentle shaking. First, 20 mg Mac1 was exchanged into lysine methylation buffer (50 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol) using a HiPrep 26/10 desalting column. The protein was diluted to 1 mg/ml with lysine methylation buffer, and 400 µl 1 M dimethylamine borane (DMAB, prepared in water) (Sigma, 180238) and 800 µl 1 M formaldehyde (prepared in water) (Sigma, F8775) were added to initiate the methylation reaction. The reaction was left to proceed for 2 hours, then 400 µl 1 M DMAB and 800 µl 1 M formaldehyde was added. After an additional 2 hours, 200 µl 1 M DMAB was added, and the reaction was left for a further 16 hours. To consume any remaining formaldehyde, and to cleave any intermolecular disulfide bonds, 2.5 mL of 1 M glycine (prepared in water) and 2.5 mL of 50 mM DTT (prepared in water) was added, and the reaction was incubated for an additional 2 hours. Next, the sample was concentrated to 2.5 mL using a 10 kDa MWCO concentrator, and purified by SEC. The methylated protein was concentrated to 15 mg/ml before flash freezing in liquid nitrogen and storage at -80°C.

To test the extent of lysine methylation, the purified sample was analyzed by liquid chromatography-mass spectrometry (LC-MS), using a Waters Acquity LC connected to a Waters TQ detector with electrospray ionization. The sample was separated on a C4 column held at 40°C using water with 0.1% formic acid as solvent A and acetonitrile with 0.1% formic acid as solvent B. After sample injection (5 µl at 10 µM diluted in 150 mM NaCl, 20 mM Tris (pH 8.0)), an isocratic elution was run with 95% solvent A and 5% solvent B for 1.5 min. Then, a linear gradient elution was run for 6.5 min to 95% solvent B. Finally, an isocratic elution was run with 95% solvent B for 2 min. The flow rate was 0.2 ml/min.

Crystallization of methylated Mac1

Crystals grew readily in the same conditions as the non-methylated protein (100 mM Tris (pH 8.5), 100 mM sodium acetate, 28% PEG 4000). Consistent nucleation was achieved using microseeding with the same protocol as the non-methylated protein. Crystallization drops were set up with 100 nl reservoir, 100 nl seed stocks and 200 nL protein using SwissCI 3-well plates. The methylated crystals displayed increased DMSO tolerance, so DMSO/fragment soaks were performed directly with 40 nl DMSO (10% of the drop volume).

Ultra high resolution data collection, refinement and modelling

To measure the diffraction at such high resolution, we employed a multi-pass, multi-crystal data collection strategy. We collected ultra high resolution X-ray diffraction data for Mac1 (C2 crystal form) by performing sequential high-energy (17000 eV) and low-energy (11111 eV) runs to accurately measure reflection intensities at high and low scattering angles respectively. The same data collection strategy (wedge, oscillation angle, exposure) was

implemented for multiple crystals, each held in different orientations relative to the X-ray beam and phi rotation axis.

The data sets were individually indexed and integrated with XDS⁵⁷. During data processing, we merged the high and low resolution datasets from multiple crystals in different orientations to maximize our coverage of reciprocal space given a square detector surface. A low-resolution cutoff of 2.5 Å was applied to the high resolution (high energy) data sets, because this cutoff simultaneously excludes potentially overlapping reflections at low scattering angles and allows for a significant number of shared observations between high and low resolution data sets, which facilitates robust scaling. Scaling and merging were performed using XSCALE, and the merged intensities were converted to structure factor magnitudes using XDSCONV⁵⁷.

We calculated phases by the method of molecular replacement, using the program Phaser⁵⁸ and a previous structure of Mac1 (PDB: 6WCF) as the search model. The model was manually adjusted in Coot⁵⁹ to fit the electron density map calculated from molecular replacement, followed by automated refinement of coordinates, atomic displacement parameters, and occupancies using *phenix.refine*⁶⁰ with optimization of restraint weights. Following two initial rounds of iterative model building and refinement using the aforementioned strategy, we began introducing additional parameters into the model, enabled by the extraordinarily high resolution of our diffraction data. First we implemented anisotropic atomic displacement parameters for heavy atoms (C,N,O,S), followed by refinement of explicit hydrogen atom positions. During early rounds of model building, we noticed $mF_o - DF_c$ difference density peaks appearing between heavy atom positions, suggesting that we are able to resolve covalent bonding densities (**Supplemental Figure 2.1E**). Indeed, atomic refinement that included a model for inter-atomic scatterers (IAS)⁶¹ was able to account for these densities and reduce the free-R value

by approximately 0.0043 (0.43%). Although the refined atomic coordinates do not differ significantly based on the inclusion or exclusion of IAS, the maximum-likelihood estimation of the phase error calculated by *phenix.refine* is 0.49° less when the IAS are included, suggesting an improvement in map quality (which may indirectly improve the model by aiding in subsequent manual interpretation of electron density features). Final refinement was performed without geometry or ADP weights (unrestrained).

Data collection at physiological temperature, refinement and modelling

We used a low-dose X-ray data collection strategy to acquire diffraction data from macromolecule crystals (C2 crystal form) at human physiological temperature (37°C, 310 K), which is the temperature most relevant to studies of SARS-CoV-2 infection. Using this strategy, we acquired data sets using an X-ray exposure of only 50 kGy - less than 1% of the total dose used at 100 K, which is essential to mitigate the rapid rate of radiation damage at 310 K compared to 100 K. The lower overall X-ray dose resulted in data with a lower overall resolution, extending to 1.5 Å.

Diffraction data from multiple crystals were merged using *xia2*⁶², implementing *DIALS*⁶³ for indexing and integration, and *Aimless*⁶⁴ for scaling and merging. We calculated phases by the method of molecular replacement, using the program *Phaser*⁵⁸ and our high resolution 100K structure as the search model. The model was manually adjusted in *Coot* to fit the electron density map calculated from molecular replacement, followed by automated refinement of coordinates, atomic displacement parameters, and occupancies using *phenix.refine*⁶⁰ with optimization of restraint weights.

Fragment data collection, refinement and modelling

Diffraction data was collected at ALS beamline 8.3.1 and SSRL beamlines 12-1 and 12-2. The data collection strategy is summarized in **Supplemental Data 2.1**. Fragment datasets were indexed, integrated and scaled using XDS⁵⁷ run through xia2⁶². Based on the space group and unit cell dimensions, six crystal forms were present (**Supplemental Figure 2.2C**). For each of the three C2 isoforms with one molecule in the ASU (isoform A, B and C), a single, high resolution dataset was selected to create a representative model for each isoform. Phases were obtained via molecular replacement with Phaser⁵⁸, using the ultra-high resolution C2 coordinates as the search model (PDB: 7KR0). Coordinates were refined with iterative rounds of manual model building in Coot and refinement with *phenix.refine*⁶⁰. Default refinement parameters were used, except five refinement macrocycles carried out per iteration and water molecules were automatically added to peaks in the $2mF_o-DF_c$ electron density map higher than 3.5σ . The minimum model-water distance was set to 1.8 Å and a maximum model-water distance to 6 Å. For later rounds of refinement, hydrogens were added to riding positions using *phenix.ready_set*, and B-factors were refined anisotropically for non-hydrogen and non-water atoms. Although these datasets were obtained from crystals soaked with fragments, there was no evidence for fragment binding in the mF_o-DF_c difference density maps, therefore the datasets were deemed acceptable as representative DMSO-only models for each isoform.

For the fragment datasets, molecular replacement was performed with Phaser⁵⁸ and initial refinement with Refmac⁶⁵, both run through the DIMPLE pipeline⁶⁶. The search model used for molecular replacement was selected to match the isoform of the dataset. Waters were included in the initial refinement by changing the HOH records in the PDB file to WWW. After refinement, waters were stripped from models and electron density maps were analyzed for fragment binding

using PanDDA³⁹. Electron density maps from 31 datasets were used to calculate the background electron density map for the A isoform, and 24 datasets were used for isoforms B and C (**Supplemental Data 2.1**). Datasets selected for background map calculation had the highest resolution and lowest R_{free} values. After PanDDA was run with default parameters, the threshold used to classify a hit was decreased by adjusting the Z-map analysis settings (contour_level = 2, min_blob_volume = 5, min_blob_z_peak = 2.5). Although there was a substantial increase in false positives, the decreased threshold allowed an additional seven fragments to be identified. Fragments were modelled into PanDDA event maps with COOT, using restraints generated by *phenix.elbow* from a SMILES string⁶⁷. Changes in protein conformation and solvation were also modeled. Because PanDDA can identify fragments binding with low occupancies, any changes in protein coordinates will have similar, low occupancies. If un-restrained refinement is performed on these low occupancy models, changes supported by PanDDA event maps are often reverted to the ground state model. In the past, this has been overcome by refining both ground-state (apo) and changed-state (fragment bound) structures simultaneously, with the changed state coordinates restrained. However, these multi-state models can be difficult to interpret. As an alternative, we modeled and refined the changed state only. To prevent reversion of the model into ground state density, coordinate refinement was switched off after fragments were modelled. Hydrogens were added with *phenix.ready_set*, waters were updated automatically and B-factors were refined anisotropically for non-hydrogen and non-water atoms. After one round of refinement, waters added into ground state electron density were removed. This was achieved by aligning the DMSO-only model to the refined model and removing any water molecules within 2.2 Å of the DMSO-only model. A final round of refinement was performed without updating water molecules.

P43 crystals at UCSF

Protein expression and purification

The C2 sequence in pET22b(+) was converted into the P4₃ sequence by removal of Glu170 and replacement of the N-terminal Asn-Ala-Gly-Glu motif with a methionine. Additionally, a Ser-Ser-Gly-Val-Asp-Leu-Gly-Thr linker was introduced between the His₆ tag and the TEV recognition sequence (**Supplemental Data 2.1**). All cloning steps were performed by PCR with overlapping primers and Gibson assembly⁶⁸. Protein was purified using the same protocol as the C2 protein, except that after SEC, the protein was concentrated to 40 mg/ml prior to flash freezing in liquid nitrogen.

Crystallization

Initially, crystals were grown by hanging-drop vapor diffusion with a reservoir solution containing 34% PEG 3000 and 100 mM CHES (pH 9.5). Screens were performed using pre-greased VDX plates (Hampton Research, HR3-142) with 0.5 mL reservoir solution in each well. Crystallization drops were set up on silicon coverslips (Hampton Research, HR3-233) with 2 µl Mac1 at 10 mg/ml and 2 µl reservoir. Crystals grew after 2-4 days at 19°C. As with the C2 crystals, microseeding was required to achieve consistent nucleation. Seed stocks were prepared as described previously, except the seed storage buffer used was 35% PEG 3000, 100 mM CHES (pH 9.5) and 2 mM DTT. Crystals for fragment soaking were grown using SwissCI 3-well sitting drop plates with reservoirs containing 30 µl 28% PEG 3000, 100 mM CHES (pH 9.5)). Crystallization drops were set up with 100 nl reservoir solution, 100 nl seed stocks (1:100,000

dilution) and 200 nL Mac1 at 40 mg/ml. Crystals were grown at 19°C and reached a maximum size after 24 hours.

Fragment and ADP-ribose soaking

Fragment soaks were performed using the same protocol as the C2 crystals, with soak times between 2-6 hours. ADP-ribose soaks were performed similarly, except that ADP-ribose was prepared in water to 100 mM, and crystals were soaked with 80 nl ADPr (20 mM final concentration). Crystals were vitrified directly after soaking using a Nanuq cryocooling device (Mitegen).

Fragment data collection, processing, modelling, and refinement

Diffraction data was collected at ALS beamline 8.3.1, SSRL beamline 12-1 and NSLS-II beamline 17-ID-2. The data collection strategy is summarized in **Supplemental Data 2.1**. Fragment datasets were indexed, integrated and scaled using XDS⁵⁷ and merged with Aimless⁶⁴. In addition to the fragment soaks, we collected diffraction data for 40 crystals soaked only with DMSO. To generate a DMSO-only model, a single high resolution dataset was selected, and phases were obtained by molecular replacement using the 0.77 Å C2 structure as a search model (PDB: 7KR0). Refinement and model building was performed as described previously for the C2 crystals. The fragment datasets were prepared for PanDDA analysis using the DIMPLE pipeline^{39,66}. Fragments were identified using PanDDA, with the background electron density map generated using 35 DMSO-only datasets (**Supplemental Data 2.1**). As with the analysis of C2 electron density maps, PanDDA was re-run with a decreased Z-map threshold (contour_level = 2.5, min_blob_volume = 5, min_blob_z_peak = 2.5). This strategy identified an additional 24

fragments. Fragment modeling and refinement was carried out using the same protocol as the experiment with C2 crystals.

P43 crystals at Oxford/XChem

Protein expression and purification

SARS-CoV-2 Nsp3 Mac1 (residues 3-169) was cloned into a pNIC28-Bsa4 expression vector which adds an N-terminal His₆-tag and a TEV protease recognition site for removal of the tag. For expression of protein used for crystallization, the constructs were transformed into the *E. coli* Rosetta strain BL21(DE3)-R3 and cells were grown at 37°C in LB medium (Miller) supplemented with 50 µg/ml of kanamycin and 35 µg/ml of chloramphenicol. After reaching an OD₆₀₀ of 0.5–0.6, the temperature was lowered to 18°C prior to induction of protein expression overnight by adding 0.5 mM IPTG. Harvested cells were resuspended in lysis buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 5% glycerol, 20 mM imidazole, 10 mM βME, cComplete EDTA-free protease inhibitors (Roche)) and stored at -20°C until purification. For protein purification, pellets were gently thawed in lukewarm water and lysed by high-pressure homogenization. DNA was digested using Benzonase. Proteins were purified by immobilized metal affinity chromatography (IMAC) using Ni-Sepharose resin (GE Healthcare) and eluted stepwise in binding buffer containing 40–500 mM imidazole. A high salt wash with 1 M NaCl was combined with the first elution step including 40 mM imidazole. Removal of the His₆ tag was carried out by addition of recombinant TEV protease during overnight dialysis into buffer without imidazole, followed by purification on a second IMAC column and finally by size-exclusion chromatography (SEC) (Superdex 75, GE Healthcare) in a buffer consisting of 20 mM HEPES (pH 8.0), 250 mM NaCl and 2 mM DTT. Macrodomain protein used for HTRF assay

was not subjected to TEV cleavage and purified after the IMAC step by SEC in a buffer consisting of 25 mM HEPES (pH 7.4), 300 mM NaCl, 5% glycerol and 0.5 mM TCEP. Proteins were characterized by SDS-PAGE, then flash frozen in liquid nitrogen and stored at -80°C until required.

Crystallographic fragment screening

SARS-CoV-2 Nsp3 Mac1 was concentrated to a final concentration of 47 mg/ml and *apo* crystals were grown in crystallization solution containing 100 mM CHES (pH 9.5) and 30% PEG 3000. Fragments were soaked into crystals as previously described²³ by adding dissolved compounds directly to the crystallization drops using an ECHO liquid handler (final concentration 10% DMSO); drops were incubated for approximately 1-3 hours prior to mounting and flash freezing in liquid nitrogen.

Data was collected at the beamline I04-1 at 100 K and automatically processed with Diamond Light Source's auto-processing pipelines using XDS⁵⁷ and either xia2⁶² or DIALS⁶³ with the default settings. Most Mac1 data processed to a resolution of approximately 1.1 Å. Further analysis was performed with XChemExplorer²⁴, electron density maps were generated with Dimple⁶⁶ and ligand-binding events were identified using PanDDA³⁹. Ligands were modelled into PanDDA-calculated event maps using Coot⁵⁹, restraints were calculated with AceDRG⁶⁹, and structures were refined with BUSTER⁷⁰. Coordinates, structure factors and PanDDA event maps for the structures discussed are deposited in the Protein Data Bank. Data collection and refinement statistics are summarized in **Supplemental Data 2.1**.

Molecular Docking Screens

Docking was performed against the crystal structure of SARS-CoV-2 Nsp3 Mac1 bound to ADP-ribose (PDB: 6W02³⁴). Chain B and all water molecules except for HOH324, HOH344, HOH384, and HOH406 were removed. These water molecules were included in the docking template structure since they were buried within the ADP-ribose binding site and formed bridging hydrogen bonds between ADP-ribose and the protein. The protein structure in complex with ADP-ribose and the four selected water molecules was capped at N- and C-termini and prepared for docking following the prepwizard protocol in Maestro (Schrödinger⁷¹). Accordingly, protons were added using Epik and protonation states were optimized with PropKa at pH 7. Finally, the structure was energetically minimized using the OPLS3e force field⁷¹. The maximum heavy-atom deviation from the initial structure was 0.3 Å.

Docking was performed with DOCK3.7 using pre-calculated scoring grids for rapid evaluation of docked molecules⁷². AMBER united atom charges⁷³ were assigned to the minimized protein structure and water molecules. Partial atomic charges of backbone amide hydrogen atoms for residues Ile23 and Phe156 were increased by 0.2 elementary charge units without changing the net charge of the residues, as described previously²⁹. The low dielectric constant of the protein environment was extended outwards from the protein surface by 1.9 Å using spheres generated by SPHGEN. Electrostatic potentials at the ligand-binding pocket were calculated by numerical solution of the Poisson-Boltzmann equation using QNIFFT⁷⁴, scoring grids for van der Waals potentials were generated with CHEMGRID. Ligand desolvation scoring grids were calculated by SOLVMAP⁷⁵, and the volume of the low protein dielectric was extended out 0.4 Å from the protein surface, as described previously⁴⁰. Since we specifically targeted the adenosine binding site of the full ADP-ribose binding pocket, atomic coordinates of

adenosine rather than the whole ADP-ribose molecule were used to generate 45 matching spheres, representing favorable positions for placing ligand atoms with docking⁷².

As ADP-ribose was the only known ligand for Mac1 when we started the docking campaign, the generated scoring grids and matching spheres were judged for their ability to place and score adenosine, adenine and ribose at the adenosine binding site of the ligand binding pocket compared to 250 property-matched decoys, generated following the DUDE-Z method⁷⁶. Decoys share similar physical properties as the control molecules but are topologically different, hence unlikely to ligate the binding pocket. Furthermore, an “extrema” set⁷⁶ of approximately 500,000 molecules including anionic, neutral and cationic compounds with molecular weights ranging from 250-350 Da was screened to ensure similar enrichments for monovalent anions and neutral molecules. We note that the lack of experimentally confirmed ligands for the macrodomain did not allow exhaustive control calculations.

Virtual compound libraries were downloaded from ZINC15 (www.zinc15.docking.org)³⁷. From the set of 722,963 in-stock fragments, 696,092 compounds were successfully docked, exploring on average 2,355 orientations and 63 conformations per compound in the binding pocket. Roughly 58 billion complexes were sampled in 88 core hours, or roughly 10 minutes on a 500 core cluster. Screening the entire 20 million ZINC15 fragment library resulted in the evaluation of ca. 4.4 trillion complexes within 2,342 core hours, or 4.7 hours on 500 cores. In that screen, 19,130,798 compounds were scored and sampled in ca. 2,145 orientations and 180 conformations each. From the relatively small “in-human” library, containing 20,726 molecules, 17,362 compounds were scored, and sampling was increased to roughly 16,615 orientations per compound. 84 billion complexes were evaluated in 27 core hours.

Compounds with DOCK scores < -20 (top 500,000 compounds from the entire fragment screen), were subsequently filtered for those with strained conformations, and inspected for their ability to form hydrogen bonds to residues Asp22, Ile23, Gly48, Val49, Gly130 or Phe156. Compounds with unsatisfied hydrogen bond donors or more than three unsatisfied hydrogen bond acceptors were deprioritized. From both fragment screens, 17 in-stock compounds (8 selected from the ZINC15 in-stock library docking screen) were purchased, and 45 make-on-demand fragments were ordered of which 33 were successfully synthesized, both from Enamine. The following compounds were selected from the “in-human” collection docking screen and purchased from different vendors: Pterin (Sigma-Aldrich, P1132), Verdiperstat (MedChem Express, HY-17646), Kinetin (Cayman Chemical, 20712), Irsogladine (Cayman Chemical, 30223), Diaveridine (Cayman Chemical, 29427), N6-Benzyladenine (Cayman Chemical, 21711), PP2 (Cayman Chemical, 13198), Temozolomide (Cayman Chemical, 14163), Chrysophanol (Cayman Chemical, 19870), Isoxanthopterin (Cayman Chemical, 17564).

Fragment linking and merging

Fragment mergers and linkers were generated using Fragmenstein (<https://github.com/matteoferla/Fragmenstein>), a python module that automatically joins fragments or places compounds based on fragments in way that is as faithful to the positions of the fragments as possible in a conformation that is energy acceptable. For merging, using RDKit⁷⁷, rings are temporarily collapsed into pseudo-atoms, one-to-one spatial overlapping atoms are identified, pseudo-atoms expanded with appropriate bonds to nearby atoms and various chemical corrections applied. For the constrained energy minimization, Pyrosetta is used⁷⁸. Interactive online summary of mergers was made at <https://michelangelo.sgc.ox.ac.uk/>⁷⁹.

Differential Scanning Fluorimetry (DSF)

Compounds were dissolved in DMSO to a final concentration of 100 mM and placed in a 384-well Echo source plate (Labcyte, PP0200). Using a LabCyte Echo. Each compound was dispensed into a 384-well storage plate (Greiner BioOne, 781280) in five stock concentrations in two-fold serial dilutions (compounds: 6.25-100 mM; ADP-ribose: 0.625-10 mM) and a final volume of 750 nl in triplicate. Two identical plates were created, with the second plate used to provide protein-free controls for all tested conditions. Echo dispensing instructions were created by an in-house app (https://gestwickilab.shinyapps.io/echo_layout_maker/).

DSF buffer was prepared by adding 10 μ L of SYPRO Orange (Thermo Scientific, S6650) to 10 mL buffer (50 mM Tris HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.01% Triton X-100), for a final dye concentration of 5X (10 μ M) SYPRO Orange. A compound plate (see above) was resuspended by the addition of 20 μ L of DSF buffer and set aside for 20 minutes in the dark. Purified Mac1 (P4₃ construct expressed at UCSF) was diluted to 10 μ M in DSF buffer, and 2 μ L of either protein solution or protein-free buffer was added to each well a 384-well white PCR plate (Axygen, PCR-384-LC480WNFBC) using an E1 ClipTip P125 electronic pipette. 8 μ L of resuspended compound was transferred to each well of the protein- and buffer-containing PCR plate using an Opentrons OT-2 liquid handling system, yielding the following final conditions: 2 μ M Mac1, 5X (10 μ M) SYPRO Orange, 3% DMSO, 0.1-3 mM fragments, and 0.1-1 mM ADP-ribose. The PCR plate was spun briefly in a salad spinner to remove bubbles, and sealed with optically clear film (Applied Biosystems, MicroAmp Optical Adhesive Film, 4311971). In an Analytik Jena qTower 384G qPCR instrument, plate was continuously heated from 25 - 94°C at a rate of 1°C/minute, and fluorescence was measured at each degree in

the TAMRA channel (535 nm / 580 nm). 53 of 54 fragments could be tested up to 3 mM without assay interference in these conditions (**Supplemental Data 2.1, Supplemental Data 2.2**). $T_{m_{as}}$ were calculated online at DSFworld⁸⁰, using fitting model 2.

Raw DSF data for the Mac1 construct used in this work was characterized by a major transition at $50.8 \pm 0.3^\circ\text{C}$, with a minor second transition at $67.0 \pm 3.6^\circ\text{C}$ (**Figure 2.10C,D, Supplemental Data 2.1, Supplemental Data 2.2**); results described refer to the major transition. Significance was defined as compounds with ANOVA p-values < 0.05 for T_{m_a} over the tested concentration regime.

Isothermal Titration Calorimetry (ITC)

All ITC titrations were performed on a MicroCal iTC 200 instrument (GE Healthcare). All reactions were performed in 20 mM Tris (pH 7.5), 150 mM NaCl using 300 - 600 μM of Mac1 (P4₃ construct expressed at UCSF) at 25°C . Titration of 4 mM ADP-ribose (Sigma-Aldrich, A0752) or 4-10 mM fragment contained in the stirring syringe included a single 0.2 μl injection, followed by 18 consecutive injections of 2 μl . Thermodynamic parameters were obtained from a non-linear least squares (NLLS) fit of a single-site binding model in the RITC package (<https://rdrr.io/cran/Ritc>).

Homogeneous Time-Resolved Fluorescence (HTRF)-based Peptide Displacement Assay

Fragment inhibitory activity on Mac1 was assessed by the displacement of an ADPr-conjugated biotin peptide from the His₆-tagged Mac1 using HTRF with a Eu³⁺-conjugated anti-His₆ antibody donor and streptavidin-conjugated acceptor. Compounds were dispensed into white ProxiPlate-384 Plus (PerkinElmer) assay plates using an Echo 525 liquid handler

(Labcyte). Binding assays were conducted in a final volume of 16 μ l with 12.5 nM Mac1, 400 nM peptide ARTK(Bio)QTARK(Aoa-RADP)S (synthesized by Cambridge Peptides (Birmingham, UK)), 1:125 Streptavidin-XL665 (Cisbio), 1:20000 Anti-His₆-Eu³⁺ cryptate (PerkinElmer) in assay buffer (25 mM HEPES (pH 7.0), 20 mM NaCl, 0.05% BSA, 0.05% Tween20). Assay reagents were dispensed into plates using a Multidrop combi (Thermo Scientific) and incubated at room temperature for 1 h. Fluorescence was measured using a PHERAstar microplate reader (BMG) using the HTRF module with dual emission protocol (A = excitation of 320 nm, emission of 665 nm, and B = excitation of 320 nm, emission of 620 nm). Raw data were processed to give an HTRF ratio (channel A/B \times 10,000), which was used to generate IC₅₀ curves by nonlinear regression using GraphPad Prism v8 (GraphPad Software, CA, USA).

References

1. Rack, J. G. M., Perina, D. & Ahel, I. Macrodomains: Structure, Function, Evolution, and Catalytic Activities. *Annu. Rev. Biochem.* **85**, 431–454 (2016).
2. Li, C. *et al.* Viral Macro Domains Reverse Protein ADP-Ribosylation. *J. Virol.* **90**, 8478–8486 (2016).
3. Fehr, A. R. *et al.* The impact of PARPs and ADP-ribosylation on inflammation and host-pathogen interactions. *Genes Dev.* **34**, 341–359 (2020).
4. Alhammad, Y. M. O. & Fehr, A. R. The Viral Macrodomain Counters Host Antiviral ADP-Ribosylation. *Viruses* **12**, (2020).
5. Fehr, A. R. *et al.* The Conserved Coronavirus Macrodomain Promotes Virulence and Suppresses the Innate Immune Response during Severe Acute Respiratory Syndrome Coronavirus Infection. *MBio* **7**, (2016).
6. Fehr, A. R. *et al.* The nsp3 macrodomain promotes virulence in mice with coronavirus-induced encephalitis. *J. Virol.* **89**, 1523–1536 (2015).
7. Grunewald, M. E. *et al.* The coronavirus macrodomain is required to prevent PARP-mediated inhibition of virus replication and enhancement of IFN expression. *PLoS Pathog.* **15**, e1007756 (2019).
8. Palazzo, L., Mikolčević, P., Mikoč, A. & Ahel, I. ADP-ribosylation signalling and human disease. *Open Biol.* **9**, 190041 (2019).
9. Caprara, G. *et al.* PARP14 Controls the Nuclear Accumulation of a Subset of Type I IFN-Inducible Proteins. *J. Immunol.* **200**, 2439–2454 (2018).
10. Fehr, A. R., Jankevicius, G., Ahel, I. & Perlman, S. Viral Macrodomains: Unique Mediators of Viral Replication and Pathogenesis. *Trends Microbiol.* **26**, 598–610 (2018).

11. Cui, J., Li, F. & Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
12. Rack, J. G. M. *et al.* Viral macrodomains: a structural and evolutionary assessment of the pharmacological potential. *Open Biol.* **10**, 200237 (2020).
13. Eriksson, K. K., Cervantes-Barragán, L., Ludewig, B. & Thiel, V. Mouse Hepatitis Virus Liver Pathology Is Dependent on ADP-Ribose-1"-Phosphatase, a Viral Function Conserved in the Alpha-Like Supergroup. *JVI* **82**, 12325–12334 (2008).
14. Putics, A., Filipowicz, W., Hall, J., Gorbalenya, A. E. & Ziebuhr, J. ADP-ribose-1"-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture. *J. Virol.* **79**, 12721–12731 (2005).
15. James, D. I. *et al.* First-in-Class Chemical Probes against Poly(ADP-ribose) Glycohydrolase (PARG) Inhibit DNA Repair with Differential Pharmacology to Olaparib. *ACS Chem. Biol.* **11**, 3179–3190 (2016).
16. Stowell, A. I. J. *et al.* A high-throughput screening-compatible homogeneous time-resolved fluorescence assay measuring the glycohydrolase activity of human poly(ADP-ribose) glycohydrolase. *Anal. Biochem.* **503**, 58–64 (2016).
17. Schuller, M. *et al.* Discovery of a Selective Allosteric Inhibitor Targeting Macrodomain 2 of Polyadenosine-Diphosphate-Ribose Polymerase 14. *ACS Chem. Biol.* **12**, 2866–2874 (2017).
18. Viridi, R. S., Bavisotto, R. V., Hopper, N. C. & Frick, D. N. Discovery of Drug-like Ligands for the Mac1 Domain of SARS-CoV-2 Nsp3. *bioRxiv* (2020)
doi:10.1101/2020.07.06.190413.

19. Hann, M. M., Leach, A. R. & Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comput. Sci.* **41**, 856–864 (2001).
20. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **1**, 187–192 (2009).
21. Krojer, T., Fraser, J. S. & von Delft, F. Discovery of allosteric binding sites by crystallographic fragment screening. *Curr. Opin. Struct. Biol.* **65**, 209–216 (2020).
22. Erlanson, D. A., Fesik, S. W., Hubbard, R. E., Jahnke, W. & Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **15**, 605–619 (2016).
23. Collins, P. M. *et al.* Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr D Struct Biol* **73**, 246–255 (2017).
24. Krojer, T. *et al.* The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr D Struct Biol* **73**, 267–278 (2017).
25. Wright, N. D. *et al.* The Low-Cost, Semi-Automated Shifter Microscope Stage Transforms Speed and Robustness of Manual Protein Crystal Harvesting. *Cold Spring Harbor Laboratory* 2019.12.20.875674 (2019) doi:10.1101/2019.12.20.875674.
26. Douangamath, A. *et al.* Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. 2020.05.27.118117 (2020) doi:10.1101/2020.05.27.118117.
27. Hall, R. J., Mortenson, P. N. & Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **116**, 82–91 (2014).
28. Manglik, A. *et al.* Structure-based discovery of opioid analgesics with reduced side effects. *Nature* **537**, 185–190 (2016).

29. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).
30. Stein, R. M. *et al.* Virtual discovery of melatonin receptor ligands to modulate circadian rhythms. *Nature* **579**, 609–614 (2020).
31. Bian, Y. & Xie, X.-Q. S. Computational Fragment-Based Drug Design: Current Trends, Strategies, and Applications. *AAPS J.* **20**, 59 (2018).
32. Chen, Y. & Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **5**, 358–364 (2009).
33. Teotico, D. G. *et al.* Docking for fragment inhibitors of AmpC beta-lactamase. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7455–7460 (2009).
34. Michalska, K. *et al.* Crystal structures of SARS-CoV-2 ADP-ribose phosphatase: from the apo form to ligand complexes. *IUCrJ* **7**, 814–824 (2020).
35. Walter, T. S. *et al.* Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* **14**, 1617–1622 (2006).
36. Cox, O. B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322–2330 (2016).
37. Sterling, T. & Irwin, J. J. ZINC 15--ligand discovery for everyone. *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
38. Bemis, G. W. & Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **39**, 2887–2893 (1996).
39. Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat. Commun.* **8**, 15123 (2017).

40. Mysinger, M. M. *et al.* Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5517–5522 (2012).
41. Allen, W. J. & Rizzo, R. C. Implementation of the Hungarian algorithm to account for ligand symmetry and similarity in structure-based design. *J. Chem. Inf. Model.* **54**, 518–529 (2014).
42. Ramírez, D. & Caballero, J. Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data? *Molecules* **23**, 1038 (2018).
43. Xing, L., Klug-Mcleod, J., Rai, B. & Lunney, E. A. Kinase hinge binding scaffolds and their hydrogen bond patterns. *Bioorg. Med. Chem.* **23**, 6520–6527 (2015).
44. Narunsky, A. *et al.* On the evolution of protein-adenine binding. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 4701–4709 (2020).
45. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16247–16252 (2011).
46. Keedy, D. A. *et al.* An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. *Elife* **7**, (2018).
47. Bayden, A. S., Moustakas, D. T., Joseph-McCarthy, D. & Lamb, M. L. Evaluating Free Energies of Binding and Conservation of Crystallographic Waters Using SZMAP. *J. Chem. Inf. Model.* **55**, 1552–1565 (2015).
48. Cappel, D., Sherman, W. & Beuming, T. Calculating Water Thermodynamics in the Binding Site of Proteins - Applications of WaterMap to Drug Discovery. *Curr. Top. Med. Chem.* **17**, 2586–2598 (2017).

49. Irwin, J. J. *et al.* ZINC20-A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* (2020) doi:10.1021/acs.jcim.0c00675.
50. Jankevicius, G. *et al.* A family of macrodomain proteins reverses cellular mono-ADP-ribosylation. *Nat. Struct. Mol. Biol.* **20**, 508–514 (2013).
51. Munnur, D. *et al.* Reversible ADP-ribosylation of RNA. *Nucleic Acids Res.* **47**, 5658–5669 (2019).
52. Lamoree, B. & Hubbard, R. E. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **61**, 453–464 (2017).
53. O'Reilly, M. *et al.* Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov. Today* **24**, 1081–1086 (2019).
54. Wood, D. J. *et al.* FragLites-Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *J. Med. Chem.* **62**, 3741–3752 (2019).
55. Downes, T. D. *et al.* Design and Synthesis of 56 Shape-Diverse 3D Fragments. *Chemistry* **26**, 8969–8975 (2020).
56. Tropea, J. E., Cherry, S. & Waugh, D. S. Expression and purification of soluble His(6)-tagged TEV protease. *Methods Mol. Biol.* **498**, 297–307 (2009).
57. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
58. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
59. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

60. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
61. Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., Lunin, V. Y. & Urzhumtsev, A. On macromolecular refinement at subatomic resolution with interatomic scatterers. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1194–1197 (2007).
62. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2009).
63. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85–97 (2018).
64. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
65. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
66. Wojdyr, M., Keegan, R., Winter, G. & Ashton, A. DIMPLE- a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands. *Acta Crystallographica Section A Foundations of Crystallography* vol. 69 s299–s299 Preprint at <https://doi.org/10.1107/s0108767313097419> (2013).
67. Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009).
68. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

69. Long, F. *et al.* AceDRG: a stereochemical description generator for ligands. *Acta Crystallogr D Struct Biol* **73**, 112–122 (2017).
70. Bricogne, G. Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives. *Acta Crystallogr. D Biol. Crystallogr.* **49**, 37–60 (1993).
71. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27**, 221–234 (2013).
72. Coleman, R. G., Carchia, M., Sterling, T., Irwin, J. J. & Shoichet, B. K. Ligand pose and orientational sampling in molecular docking. *PLoS One* **8**, e75992 (2013).
73. Weiner, S. J. *et al.* A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).
74. Gallagher, K. & Sharp, K. Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* **75**, 769–776 (1998).
75. Mysinger, M. M. & Shoichet, B. K. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.* **50**, 1561–1573 (2010).
76. Stein, R. M. *et al.* Property-unmatched decoys in docking benchmarks. *Understanding virtual solvent through large-scale ligand discovery* 114 (2020).
77. Tosco, P., Stiefl, N. & Landrum, G. Bringing the MMFF force field to the RDKit: implementation and validation. *J. Cheminform.* **6**, 37 (2014).
78. Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).

79. Ferla, M. P., Pagnamenta, A. T., Damerell, D., Taylor, J. C. & Marsden, B. D. MichelaNglo: sculpting protein views on web pages without coding. *Bioinformatics* **36**, 3268–3270 (2020).
80. Wu, T. *et al.* Three Essential Resources to Improve Differential Scanning Fluorimetry (DSF) Experiments. *Cold Spring Harbor Laboratory* 2020.03.22.002543 (2020) doi:10.1101/2020.03.22.002543.
81. Rad, M. N. S., Behrouz, S., Zarenezhad, E. & Kaviani, N. Highly efficient protocol for one-pot N-alkylation of nucleobases using alcohols in bmim[Br]: a rapid route to access acyclic nucleosides. *Journal of the Iranian Chemical Society* vol. 12 1603–1612 Preprint at <https://doi.org/10.1007/s13738-015-0633-9> (2015).
82. Morin, A. *et al.* Cutting edge: Collaboration gets the most out of software. *Elife* **2**, e01456 (2013).

Supplemental Information

Purity and structure determination of fragments ZINC901381520, ZINC82473428 and ZINC89254160 from Enamine

Samples of ZINC901391520, ZINC82473428 and ZINC89254160 obtained from Enamine were expected to be N₉-alkylated isomers but electron density of the fragments in X-ray structures indicated these samples were N₃-alkylated isomers instead (ZINC901391520_N3, ZINC82473428_N3 and ZINC89254160_N3, see **Supplemental Figure 2.7I-L**). The original samples of ZINC901391520, ZINC82473428 and ZINC89254160 used in fragment screening by X-ray crystallography were analyzed by HPLC-MS and ¹H NMR to confirm sample purity and corroborate structure. There is no reported characterization data to be used as reference for structure confirmation for N₉- or N₃-alkylated compounds ZINC901391520 and ZINC89254160. The N₉-alkylated structure ZINC82473428 is a previously prepared compound with tabulated NMR data reported by Rad *et al.*⁸¹

A re-supplied sample of ZINC901391520 from a new batch synthesized at Enamine was confirmed by ¹H NMR to be >95% purity and a different isomer than the original sample of ZINC901391520. The X-ray crystal structure of this fragment in complex with Mac1 revealed the fragment to be N₉-alkylated isomer (**Supplemental Figure 2.7I**).

The original samples of ZINC901391520, ZINC82473428 and ZINC89254160 from Enamine used in fragment screen were evaluated for purity by HPLC on an Agilent 1200 Binary SL system with diode array detection and mass spectrometric detection on an Agilent 6135B Quadrupole system in electrospray ionization mode (positive ion detection). One of two HPLC Methods A or B were used to determine sample purity using mobile phase linear gradients of

acetonitrile with 0.1% TFA in water with 0.1% TFA detailed below at 1.000 ml/min flow rate through a Phenomenex Gemini 3 mm C18 110 Å LC column (4.6 mm dia. x 150 mm length).

HPLC Method A mobile phase gradient: Gradient time points (minutes): 1.0-1.5-10.5-11.0-12.5-13.0-15.0; % acetonitrile at gradient time points: 5-5-20-95-95-5-5

HPLC Method B mobile phase gradient: Gradient time points (minutes): 1.0-7.0-8.0-10.0-10.5-12.0; % acetonitrile at gradient time points: 5-30-95-95-5-5

NMR experiments for samples ZINC901391520, ZINC82473428 and ZINC89254160

Original samples of ZINC901391520, ZINC82473428 and ZINC89254160 from Enamine used in the fragment screen were dissolved in d_6 -DMSO and analyzed by ^1H and ^{13}C NMR on a Bruker 400 MHz instrument with Avance III electronics. Data was obtained at ambient temperature (ca. 25°C) collecting 64 scans for proton experiments and 1024 scans for carbon experiments. Raw data was processed, and reports created using ACD Spectrus software.

Original sample ZINC901391520

A sample of 5.5 mg ZINC901391520 was dissolved in 0.75 mL d_6 -DMSO for NMR analysis and from this solution 50 μL was diluted in 0.45 mL acetonitrile to make up the analytical sample for HPLC-MS using HPLC Method A. The sample chromatogram from HPLC revealed a single peak with UV absorbance at both 214 and 254 nm at $t_R = 5.272$ minutes. Aside from a very strong UV_{214} peak at $t_R = 2.00$ minutes attributed to DMSO co-solvent in the sample, no other peaks were observed at these UV wavelengths and sample purity estimated >98% based on UV peak area. ^1H NMR (400 MHz, d_6 -DMSO, 25°C) δ ppm 8.49 (s, 1H), 7.91-8.26 (br d, 2H), 7.76 (s, 1H), 6.30 (s, 1H), 5.63 (s, 2H), 3.86 (s, 3H). ^{13}C NMR (101 MHz, d_6 -DMSO, 25°C)

δ ppm 172.05, 167.57, 155.01, 152.46, 149.39, 143.55, 120.18, 94.85, 57.13, 44.32. LRMS (ESI⁺) for peak at $t_R = 5.272$ minutes: observed $m/z = 247.3$ [MH]⁺ for C₁₀H₁₀N₆O₂ exact mass = 246.09.

Second batch sample ZINC901391520

A sample was dissolved in 0.75 mL *d*₆-DMSO for NMR analysis. ¹H NMR (400 MHz, *d*₆-DMSO, 25°C) δ ppm 8.24 (s, 1H), 8.16 (s, 1H), 7.31 (br s, 2H), 6.22 (s, 1H), 5.49 (s, 2H), 3.86 (s, 3H).

Sample ZINC82473428

A sample of 3.9 mg ZINC82473428 was dissolved in 0.75 mL *d*₆-DMSO for NMR analysis and from this solution 50 μ L was diluted in 0.45 mL acetonitrile to make up the analytical sample for HPLC-MS using HPLC Method B. The sample chromatogram from HPLC revealed a single peak with UV absorbance at both 214 and 254 nm at $t_R = 3.766$ minutes. Aside from a very strong UV₂₁₄ peak at $t_R = 2.00$ minutes attributed to DMSO cosolvent in the sample no other peaks were observed at these UV wavelengths and sample purity estimated >98% based on UV peak area. ¹H NMR (400 MHz, *d*₆-DMSO, 25°C) δ ppm 8.31 (s, 1H), 8.01 (br s, 2H), 7.86 (s, 1H), 4.44 (dd, $J=13.18, 3.39$ Hz, 1H), 4.31-4.40 (m, 1H), 4.20-4.30 (m, 1H), 3.75-3.87 (m, 1H), 3.58-3.70 (m, 1H), 1.93-2.07 (m, 1H), 1.75-1.92 (m, 2H), 1.58-1.73 (m, 1H). ¹³C NMR (101 MHz, *d*₆-DMSO, 25°C) δ ppm 154.78, 151.53, 149.56, 144.30, 75.35, 67.24, 52.54, 40.44, 28.23, 25.03. LRMS (ESI⁺) for peak at $t_R = 3.766$ minutes: observed $m/z = 220.3$ [MH]⁺ for C₁₀H₁₃N₅O exact mass = 219.11.

Reported NMR data for compound ZINC82473428_N9 from Rad et al., 2015⁸¹: ¹H NMR (400 MHz, *d*₆-DMSO, 25°C) δ ppm 7.91 (s, 1H), 7.83 (s, 1H), 7.01 (br s, 2H), 3.87-3.99 (m, 3H), 3.34-3.52 (m, 2H), 1.30-1.54 (complex m, 4H). ¹³C NMR (101 MHz, *d*₆-DMSO, 25°C) δ ppm 156.6, 152.9, 149.2, 144.7, 117.2, 80.6, 67.9, 57.8, 29.1, 25.1.

Sample ZINC89254160

A sample of 3.2 mg ZINC89254160 was dissolved in 0.75 mL *d*₆-DMSO for NMR analysis and from this solution 50 μL was diluted in 0.45 mL acetonitrile to make up the analytical sample for HPLC-MS using HPLC Method A. The sample chromatogram from HPLC revealed a major peak and a minor peak with UV absorbances at both 214 and 254 nm: major peak $t_R = 6.530$ minutes and minor peak $t_R = 6.751$ minutes. Relative peak area calculated as percentage of combined UV peak area at 254 nm was 93.3% major peak and 6.7% minor peak (corresponds to ca. 14:1 ratio). Aside from a very strong UV₂₁₄ peak at $t_R = 2.00$ minutes attributed to DMSO cosolvent in the sample no other peaks were observed at these UV wavelengths. Tabulated NMR data reported here for major peaks only. ¹H NMR (400 MHz, *d*₆-DMSO, 25°C) δ ppm 8.47 (s, 1H), 7.95 (br s, 2H), 7.73 (s, 1H), 7.47 (s, 1H), 5.55 (s, 2H), 2.60 (s, 3H). ¹³C NMR (101 MHz, *d*₆-DMSO, 25°C) δ ppm 166.21, 154.93, 152.47, 149.63, 149.51, 143.63, 120.43, 117.69, 48.08, 18.66. LRMS (ESI⁺) for major peak at $t_R = 6.530$ minutes: observed $m/z = 247.3$ [MH]⁺ for C₁₀H₁₀N₆S exact mass = 246.07. LRMS (ESI⁺) for minor peak at $t_R = 6.751$ minutes: observed $m/z = 247.3$ [MH]⁺ for C₁₀H₁₀N₆S exact mass = 246.07.

Major peak and minor peak have the same observed mass peak in LRMS and are presumed to be different N-alkylated isomers.

Conclusions based on HPLC-MS and NMR characterization of samples ZINC901391520, ZINC82473428 and ZINC89254160

HPLC-MS data confirmed that samples ZINC901391520 and ZINC82473428 are single compounds >98% purity with mass peak corresponding to either N₉- or N₃-alkylated isomers. Both ¹H and ¹³C NMR data corroborated initial samples ZINC901391520 and ZINC82473428 are >98% single compound. The very high purity determined for these two samples rules out the possibility that the structures determined by X-ray crystallography were the result of trace amounts of the alternative isomer in the samples. For ZINC89254160, HPLC-MS data confirmed that there was a 13:1 ratio of isomers in this sample and it is possible that the X-ray crystal structure of Mac1 obtained with ZINC89254160 was the result of protein complexed to trace/minor amounts of the alternative isomer (N₃-alkylated).

The NMR data obtained for sample ZINC82473428 used in crystallographic fragment screen does not match NMR data reported in the literature for the N₉-alkylated ZINC82473428 and thus this sample is presumed not to be N₉-alkylated isomer. NMR data is not sufficient to unambiguously assign N₃- or N₉-alkylated structures for ZINC901391520, ZINC82473428 or ZINC89254160 and the unambiguous structure assignment of ZINC901391520, ZINC82473428 and ZINC89254160 as N₃-alkylated isomers in this work was provided by the electron density observed for these fragments in the Mac1 X-ray crystal structures.

The crystal structure of Mac1 with ZINC400552187 additionally revealed the N₃-alkylated structure instead of the requested N₉-alkylated form. Using DSF and ITC, ZINC901391520, ZINC82473428, ZINC89254160, ZINC400552187 were initially screened as the N₃-alkylated isomer (ZINC901391520_N3 (PDB: 5RSK), ZINC82473428_N3 (PDB: 5RVF), ZINC89254160_N3 (PDB: 5RSJ), ZINC400552187_N3 (PDB: 5RVG)). In addition, the

N₉-alkylated ZINC901391520 (PDB: 5S6W) was tested in DSF and the peptide-competition assay (HTRF) (**Supplemental Data 2.1**).

QCRG Structural Biology Consortium authorship

This work was supported by the QCRG (Quantitative Biosciences Institute Coronavirus Research Group) Structural Biology Consortium. Listed below are the contributing members of the consortium listed by teams. Within each team the team leads are italicized (responsible for organization of each team, and for the experimental design utilized within each team), then the rest of team members are listed alphabetically.

Bacterial expression team: *Amy Diallo, Meghna Gupta, Erron W. Titus*, Jen Chen, Loan Doan, Sebastian Flores, Mingliang Jin, Huong T. Kratochvil, Victor L. Lam, Yang Li, Megan Lo, Gregory E. Merz, Joana Paulino, Aye C. Thwin, Zanlin Yu, Fengbo Zhou, Yang Zhang.

Protein purification team: *Daniel Asarnow, Michelle Moritz, Tristan W. Owens, Sergei Pourmal*, Caleigh M. Azumaya, Cynthia M. Chio, Bryan Faust, Meghna Gupta, Kate Kim, Joana Paulino, Jessica K. Peters, Kaitlin Schaefer, Tsz Kin Martin Tsui.

Crystallography team: *Nadia Herrera, Huong T. Kratochvil, Ursula Schulze-Gahmen, Iris D. Young*, Justin Biel, Ishan Deshpande, Xi Liu.

CryoEM grid freezing/collection team: *Caleigh M. Azumaya, Axel F. Brilot, Gregory E. Merz, Cristina Puchades, Alexandra N. Rizo, Ming Sun*, Julian R. Braxton, Meghna Gupta, Fei Li,

Kyle E. Lopez, Arthur Melo, Gregory E. Merz, Frank Moss, Joana Paulino, Thomas H. Pospiech, Jr., Sergei Pourmal, Amber M. Smith, Paul V. Thomas, Feng Wang, Zanlin Yu.

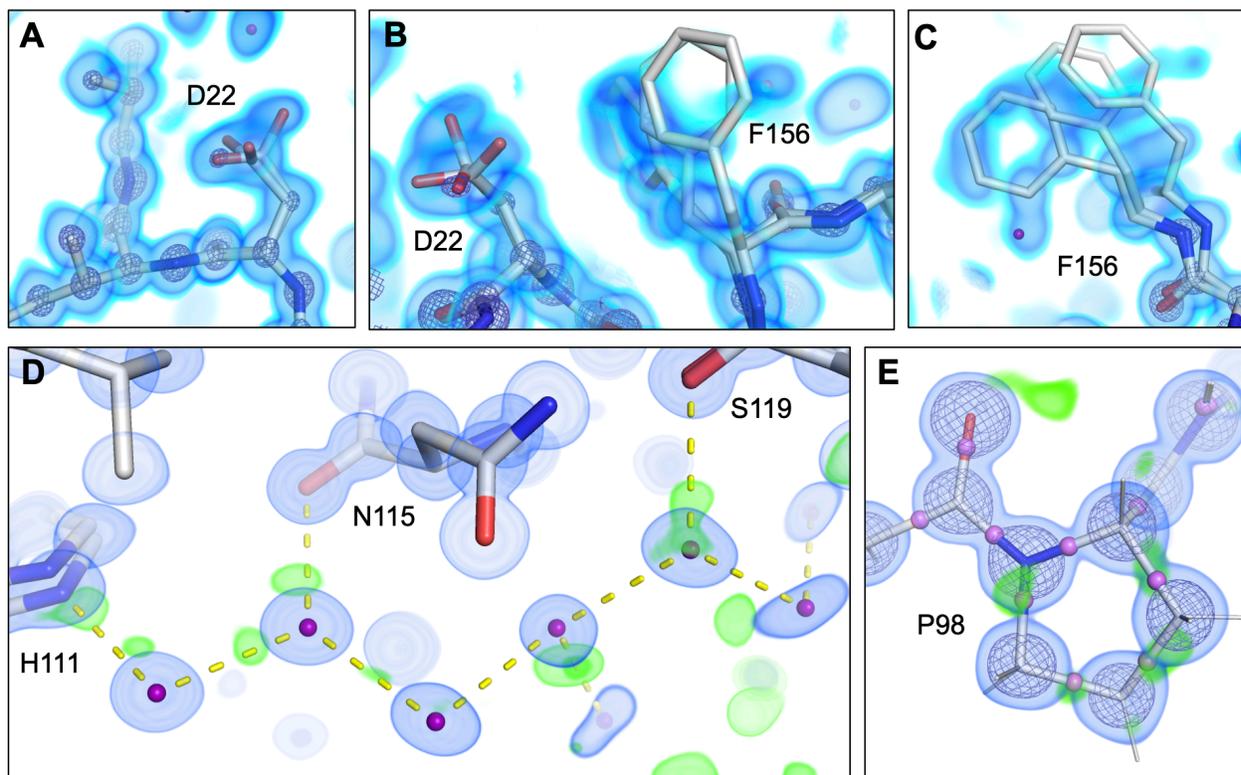
CryoEM data processing team: *Axel F. Brilot, Miles Sasha Dickinson, Gregory E. Merz, Henry C. Nguyen, Alexandra N. Rizo, Daniel Asarnow, Julian R. Braxton, Melody G. Campbell, Cynthia M. Chio, Un Seng Chio, Devan Diwanji, Bryan Faust, Meghna Gupta, Nick Hoppe, Mingliang Jin, Fei Li, Junrui Li, Yanxin Liu, Joana Paulino, Thomas H. Pospiech, Jr., Sergei Pourmal, Raphael Trenker, Donovan Trinidad, Eric Tse, Kaihua Zhang, Fengbo Zhou.*

Mammalian cell expression team: *Christian Billesboelle, Melody G. Campbell, Devan Diwanji, Carlos Nowotny, Amber M. Smith, Jianhua Zhao, Caleigh M. Azumaya, Alisa Bowen, Nick Hoppe, Yen-Li Li, Phuong Nguyen, Cristina Puchades, Mali Safari, Kaitlin Schaefer, Raphael Trenker, Tsz Kin Martin Tsui, Natalie Whitis.*

Infrastructure team: David Bulkley, Arceli Joves, Almarie Joves, Liam McKay, Mariano Tabios, Eric Tse.

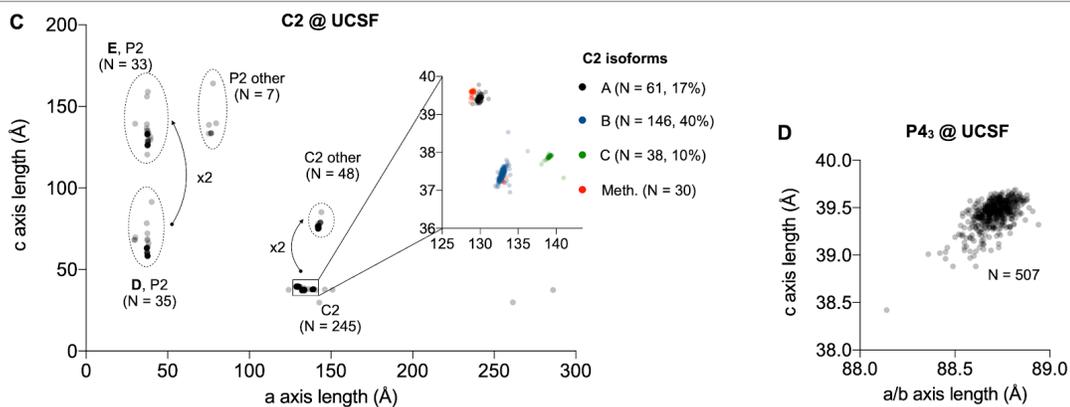
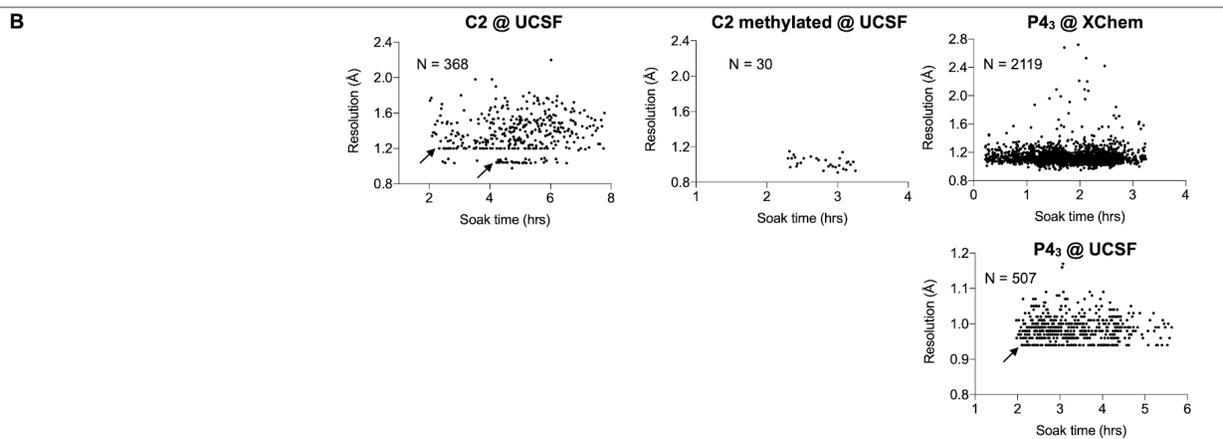
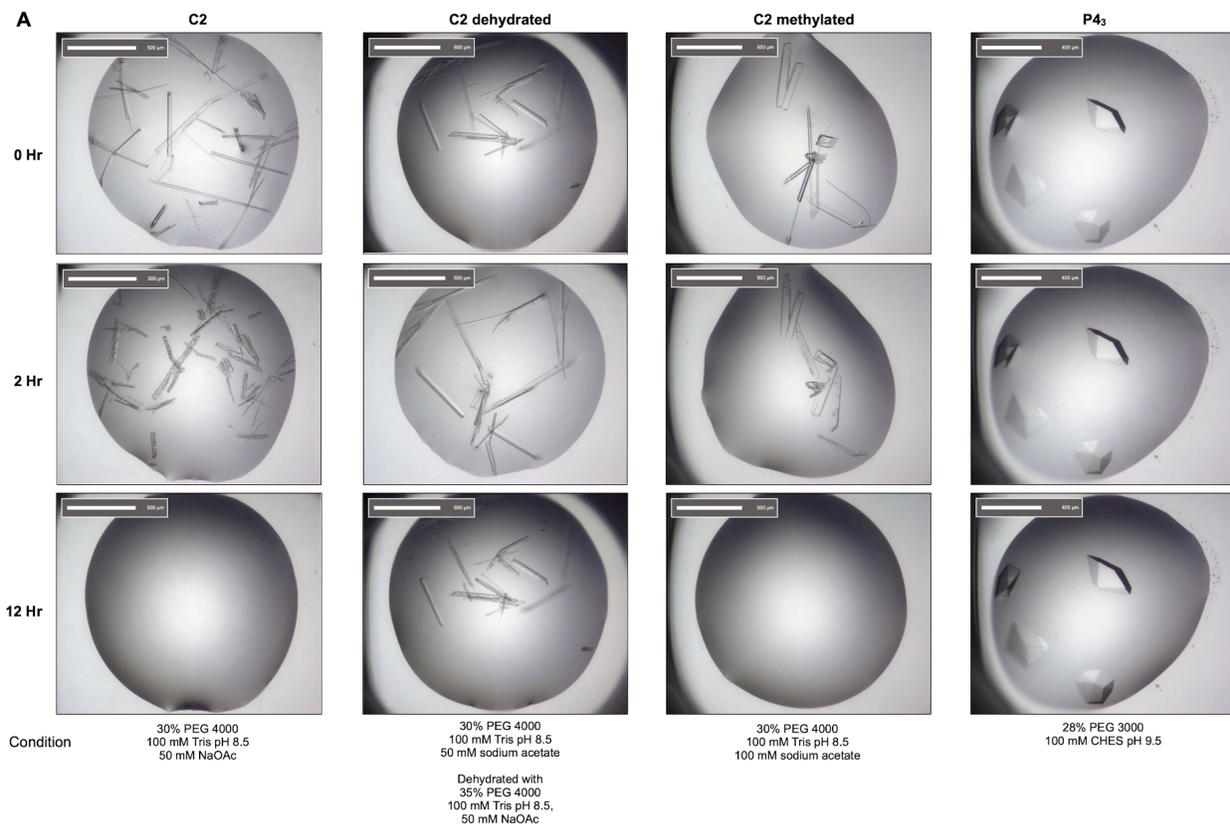
Leadership team: *Oren S Rosenberg, Kliment A Verba, David A Agard, Yifan Cheng, James S Fraser, Adam Frost, Natalia Jura, Tanja Kortemme, Nevan J Krogan, Aashish Manglik, Daniel R. Southworth, Robert M Stroud.*

Supplemental Figures



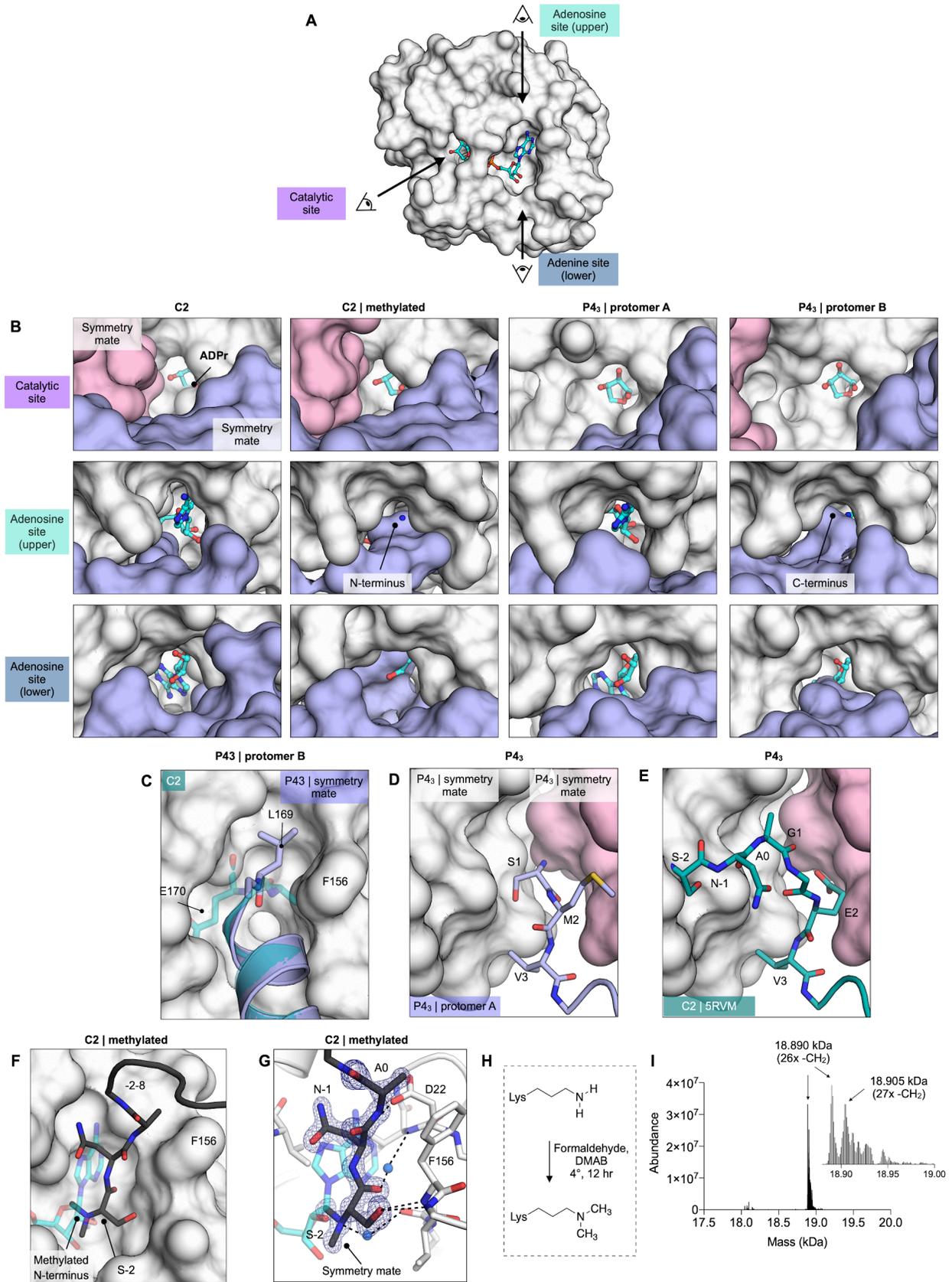
Supplemental Figure 2.1 | Ultra-high resolution features in Mac1 electron density maps.

A,B,C) Residues Asp22 and Phe156 both display conformational heterogeneity in high resolution electron density maps of apo Mac1. The ultra-high resolution nature of the electron density is evident in (A), where atoms belonging to residues Ile23 and Leu24, as well as the backbone, appear as separate spheres at high contour. In (A,B,C) three $2mF_o - Df_c$ maps are shown, contoured at 4.0σ (blue mesh), 1.0σ (blue volume), and 0.5σ (cyan volume). **D)** Mac1 electron density reveals ordered water networks, including difference density that suggests the positions of hydrogen atoms within the water network. In (D), a $2mF_o - Df_c$ map is shown, contoured at 1.5σ (blue volume), and a $mF_o - Df_c$ map is also shown, contoured at 2.5σ (green volume). Hydrogen bonding interactions are depicted as dashed yellow lines. **E)** Two electron density maps are shown for P98, a $2mF_o - Df_c$ map, contoured at 1.5σ (blue volume), and a $mF_o - Df_c$ map, contoured at 2.0σ (green volume). The green peaks of positive difference density between heavy atom positions suggest covalent bonding densities, and we refined a model of Mac1 that included interatomic scatterers (IAS), shown in violet in the figure.



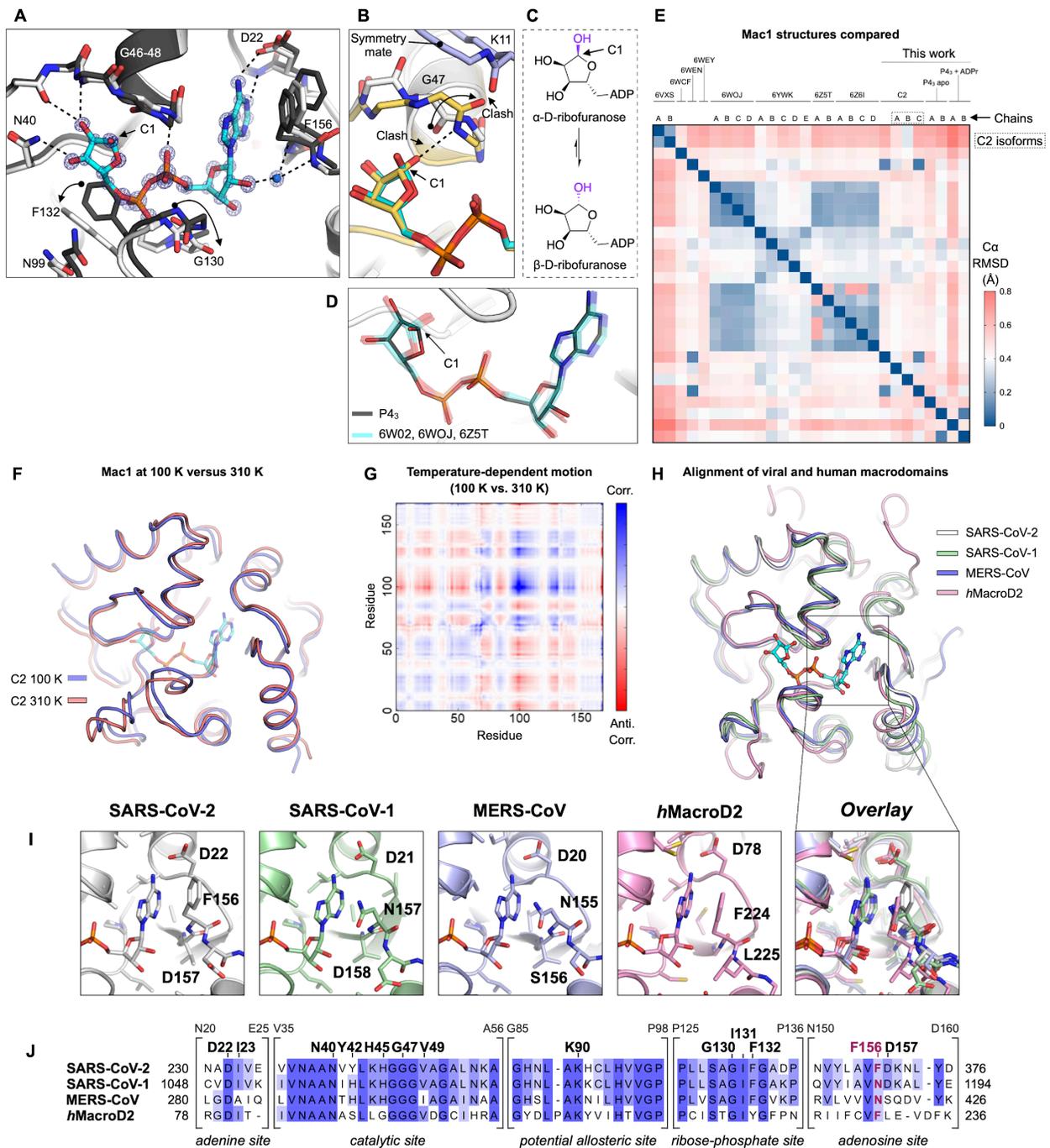
Supplemental Figure 2.2 | Comparison of isomorphism and DMSO tolerance of the C2 and P4₃ crystals.

A) Images of crystals after soaking with 10% DMSO for 0, 2 and 12 hours. **B)** Resolution of the three crystal forms as a function of soak time for the datasets collected at XChem and UCSF. The arrows indicate where the measurement of high-resolution reflections was limited by the experimental setup. **C)** Multiple isoforms were observed for the C2 crystals after dehydration. Isoforms were distinguished based on differences in the a and c unit cell lengths. Arrows indicate where doubling of the a or c axis occurred. Inset: the majority of the datasets that were indexed in C2 (245, 84%) could be clustered into three isoforms (A, B and C). Of the 30 datasets collected for crystals grown from methylated protein, the majority (28) were similar to the A isoform. **D)** The P4₃ crystals were isomorphous.



Supplemental Figure 2.3 | Crystal packing in Mac1 crystals determines active site accessibility.

A) Active site access in the C2 and P4₃ crystals. Mac1 is shown as a white surface with ADP-ribose bound in the active site shown as cyan sticks. The three access points are indicated with arrows. **B)** Crystal packing defines the three access points. The catalytic site is partially obstructed in the C2 crystals, but open in both protomers of the P4₃ crystals. In both the methylated C2 crystals and protomer B of the P4₃ crystals, the adenosine site is obstructed. **C)** The C-terminal leucine (Leu169) of the P4₃ construct occupies the adenosine site of a symmetry mate. The adenosine site is shown as a white surface and the C-terminal residues with blue sticks/cartoon. The C2 sequence (transparent teal cartoon/sticks) has an additional residue at the C-terminus (Glu170) and is therefore incompatible with the P4₃ crystal packing. **C)** The N-terminal residues of the P4₃ sequence (blue sticks) pack between two symmetry mates (white and pink surface). Compared to the P4₃ sequence, the C2 sequence contains a substitution (Met2Glu) and a three-residue insertion (Asn-Ala-Gly). These residues were typically disordered; however, they were resolved in one of the fragment structures (ZINC157088 | 5RVM) (shown aligned to the P4₃ protomer A in **(E)**). Like differences in the C-termini, differences in the N-termini may have contributed to the distinct crystal packing seen for the two Mac1 structures reported in this work. **F)** The adenosine site was obstructed by a symmetry mate in the structure determined from crystals grown using methylated C2 protein. **G)** In the structure of methylated Mac1, the side-chain hydroxyl of Ser-2 occupies the oxyanion subsite. Electron density ($2mF_o - DF_c$) is shown as a blue mesh, contoured at 1.5σ . **H)** Free amines were methylated using formaldehyde and dimethylamine borane (DMAB). The reaction is shown for lysine, however, based on the electron density shown in **(G)**, the N-terminal amine was methylated as well. The methylated amines would be protonated at the pH used to grow crystals (pH 8.5). **I)** LC/MS analysis of methylated Mac1 (C2 construct). The mass spectrum was deconvoluted using MaxEnt1. The major peak (18.89 kDa) is consistent with the methylation of 13 lysine residues ($26 \times -CH_2$). The minor peak (18.905 kDa + 15 Da) suggests that methylation was not 100% complete.

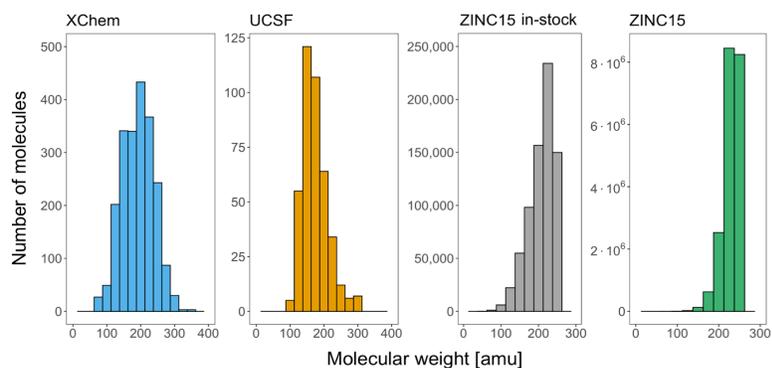


Supplemental Figure 2.4 | Structure and sequence comparison of Mac1 with related viral and human macrodomains.

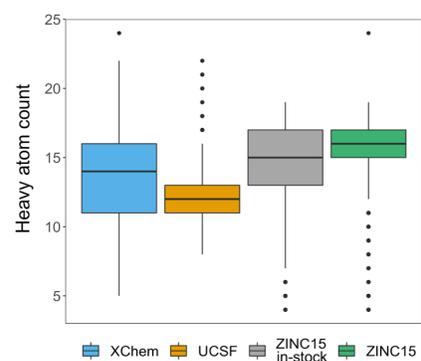
A) The structural changes previously reported to occur upon ADP-ribose binding are captured by the Mac1-ADPr structure determined using P4₃ crystals. The apo P4₃ structure is shown with dark gray sticks, with arrows indicating the changes in protein conformation upon ADP-ribose binding (white sticks). Electron density ($2mF_o - DF_c$) is contoured at 4σ (blue mesh). **B)** The α -anomer of the terminal ribose of ADP-ribose was observed in the P4₃ crystal form (cyan and white sticks). In previously reported structures (e.g., PDB 6W02, yellow sticks), a flip in Gly47 allows the β -anomer to bind by removing a steric block (red dashed line) and forming a new

hydrogen bond (black dashed line). However, the Gly47 flip is incompatible with the P4₃ crystal form because it would clash with the Lys11 carbonyl of a symmetry mate (blue sticks). In α -anomer, the anomeric hydroxyl is orientated away from Gly47, and binding can proceed without the peptide flip. **C)** Interconversion between ribose anomers in solution. **D)** Stick representation showing the previously reported Mac1-ADPr structures (cyan sticks) and the new structure determined using P4₃ crystals (grey sticks). The agreement between ADP-ribose is excellent, despite different anomers of the terminal ribose being present (α in the P4₃ structure, β in the previously reported structures). **E)** Heatmap showing the C α RMSD values after C α alignment for 10 previously reported SARS-CoV-2 Mac1 structures (6VXS, 6W02, 6W6Y, 6WCF, 6WEN, 6WOJ, 6WEY, 6YWK, 6Z5T, 6Z6I) and the new structures reported in this work. **F)** Comparison of SARS-CoV-2 Nsp3 Mac1 structures at 100 K (blue) and 310 K (red). The adenosine diphosphoribose ligand shown in the figure (cyan) is modeled according to its position in PDB 6W02. **G)** Correlation plot showing structural differences between the 100 K and 310 K structures. To generate the plot, the 100 K and 310 K structures were aligned, and difference vectors were calculated between identical C α atoms in the two structures. The plot shows all pairwise dot products between these difference vectors, revealing the extent to which temperature-dependent changes are correlated across the structure. Positive dot products (positive correlations) are colored blue, while dot products (negative correlations) are shown in red. The pattern of positive and negative correlations is characteristic of a hinge-bending motion. **H)** Alignment of three coronavirus macrodomain structures with a human macrodomain (hMacroD2). ADP-ribose from the SARS-CoV-2 structure is shown with cyan sticks. **I)** Comparison of the adenosine binding site highlighting key residues involved in adenine and fragment interaction. The adenine coordination by Phe156 is unique to SARS-CoV-2 amongst betacoronaviruses and replaced in SARS-CoV-1 (PDB: 2FAV) and MERS-CoV (PDB: 5HOL) with asparagine. Human macrodomains including MacroD2 (PDB: 4IQY) interact with adenine as SARS-CoV-2 Nsp3 Mac1 with a phenylalanine in this position which needs to be considered for achieving inhibitor selectivity for viral over human macrodomains. **J)** Sequence alignment showing conservation of residues in the ADP-ribose, catalytic and potential allosteric sites which are targeted by the fragments. Residue numbers on top refer to the construct residue numbering of SARS-CoV-2 Nsp3 Mac1. Numbers on either end of the alignment are residue numbers in the full-length proteins. The adenine coordinating Phe156 is highlighted in red.

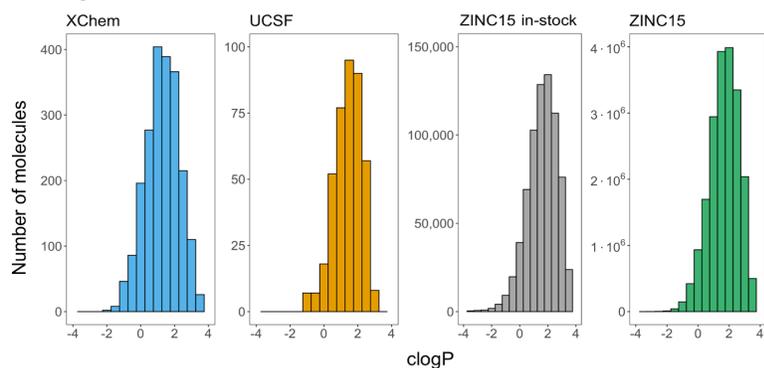
A Molecular weight



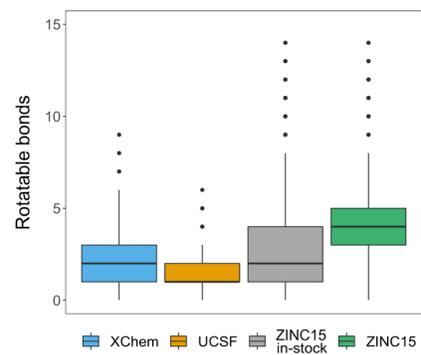
C Number of heavy atoms



B cLogP



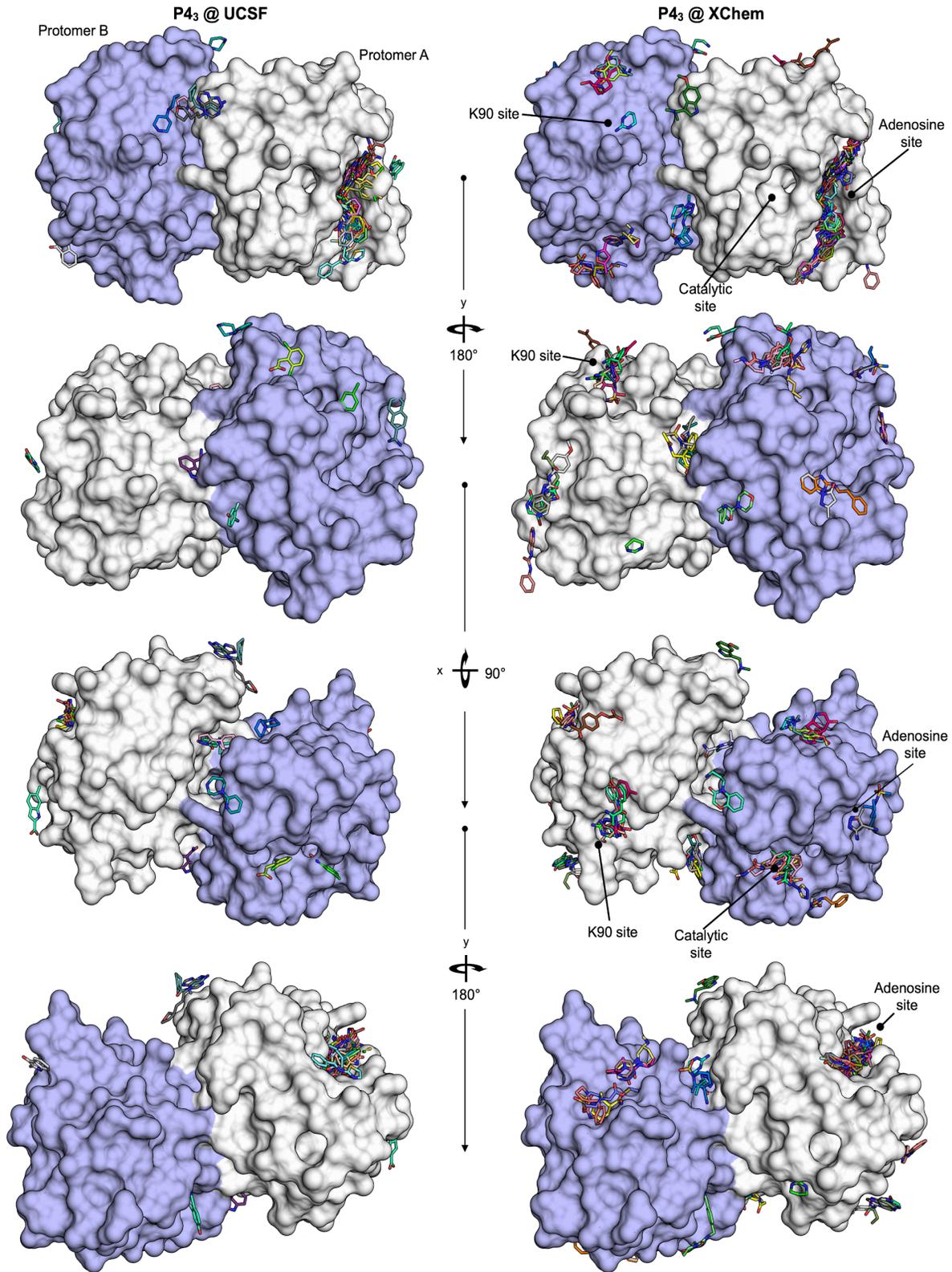
D Number of rotatable bonds



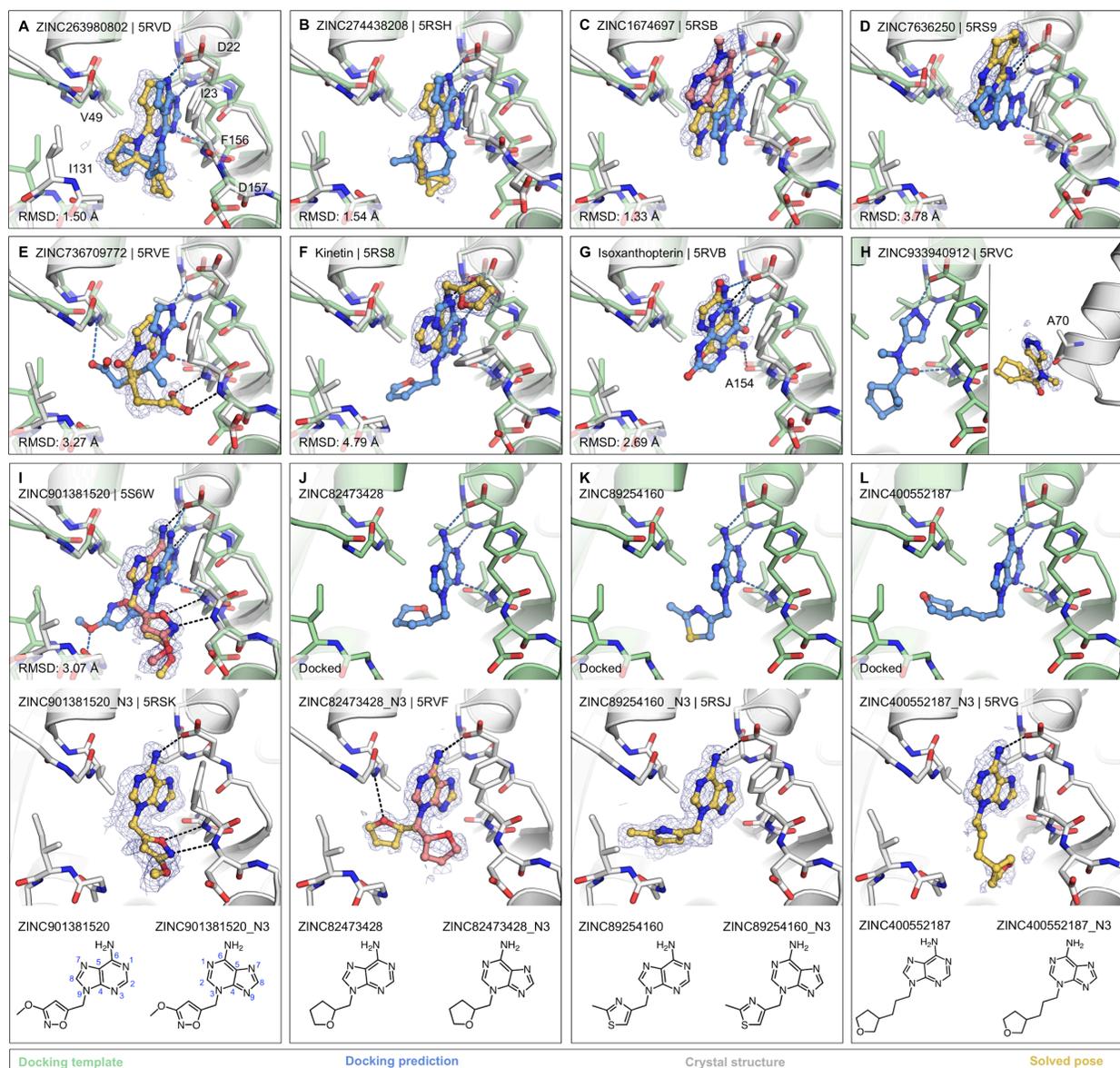
E Scaffold and chemotype analysis

Fragment Library	Number of molecules	Distinct Bemis-Murcko scaffolds	Number of pyrimidines	Number of anions
XChem	2,126	809	72 (3.39%)	115 (5.4%)
UCSF	411	179	12 (2.92%)	144 (35%)
ZINC15 (in-stock)	722,963	69,244	41,531 (5.74%)	85,398 (11.8%)
ZINC15	20,006,175	803,333	890,199 (4.44%)	739,184 (3.7%)

Supplemental Figure 2.5 | Physical properties, scaffold and chemotype analysis of screened fragment libraries.

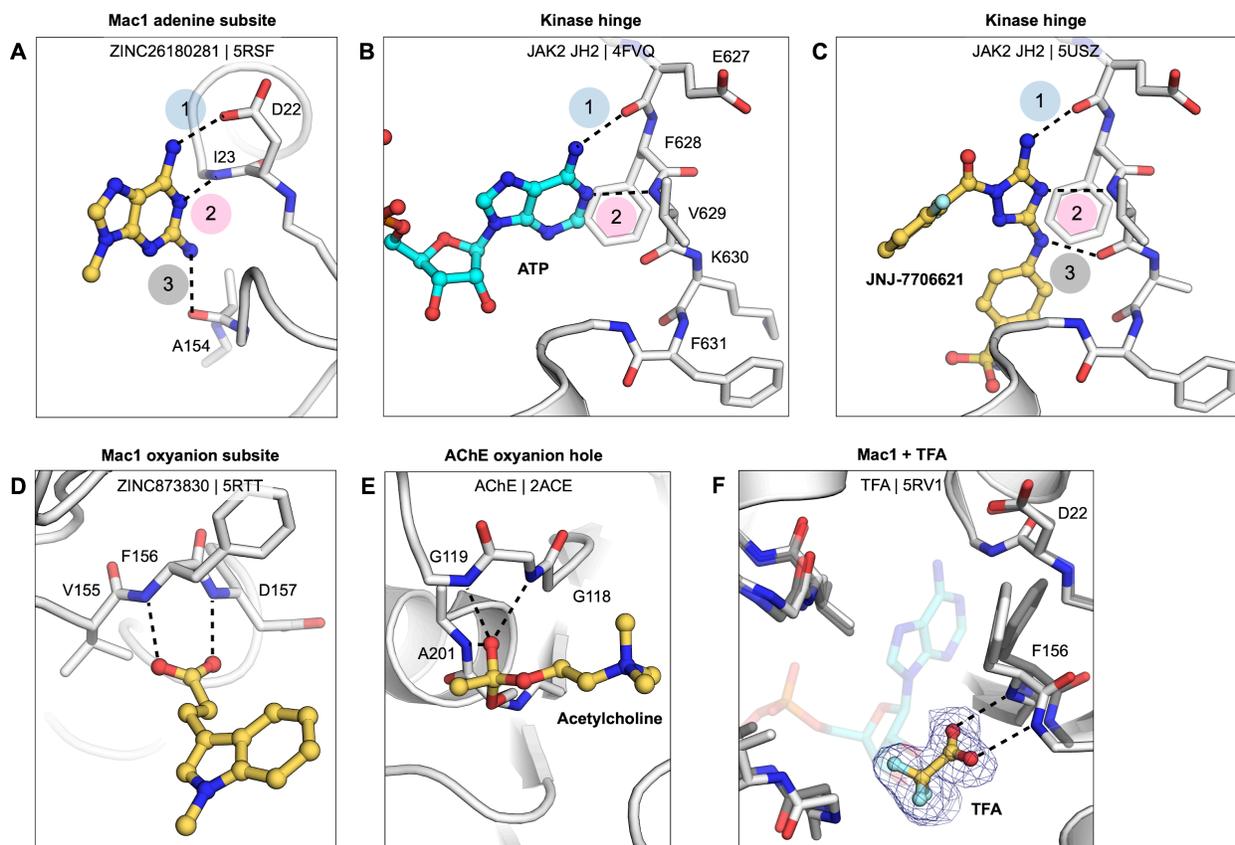


Supplemental Figure 2.6 | Overview of fragment binding to protomer A (white surface) and protomer B (blue surface) of the P43 crystals.



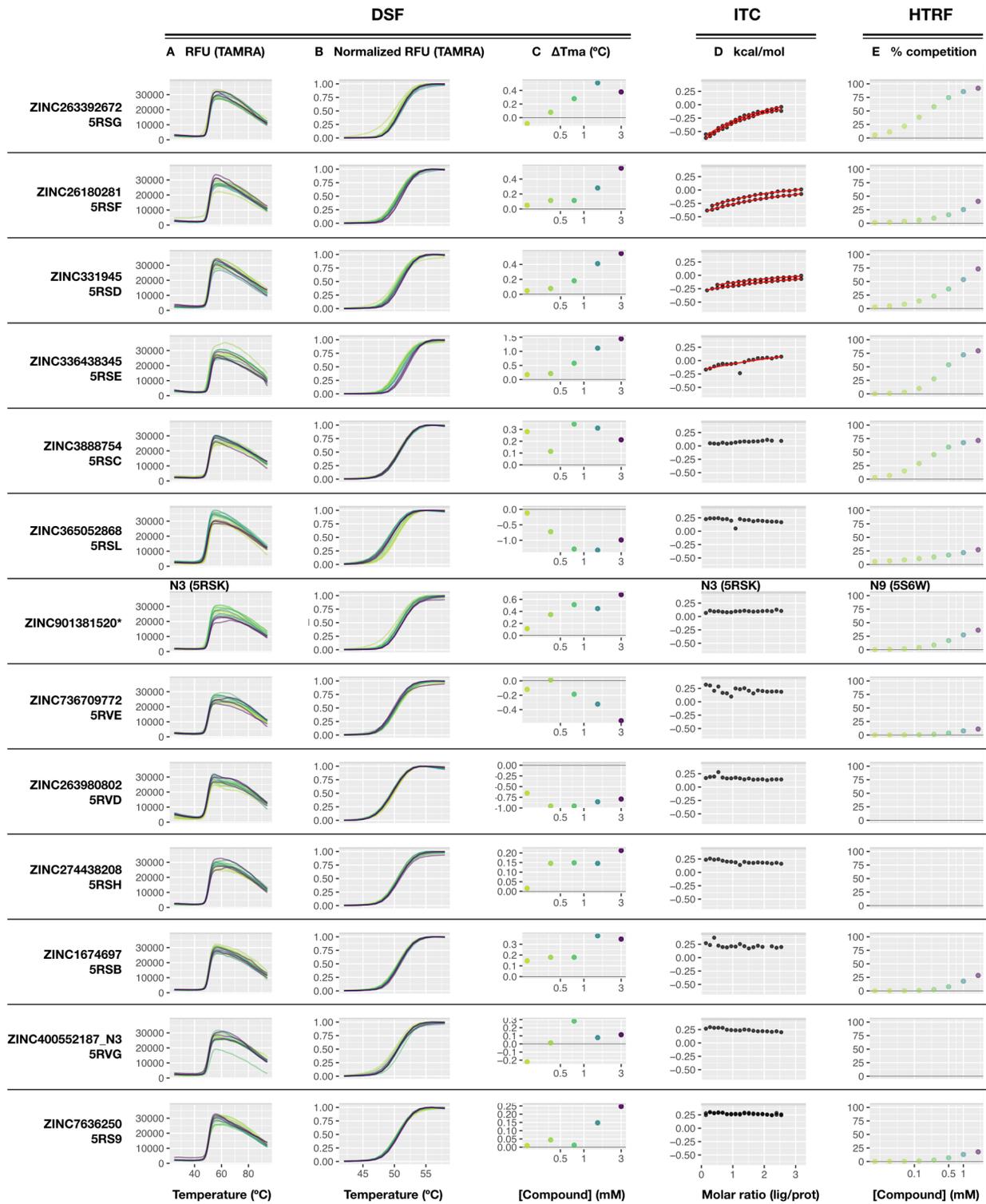
Supplemental Figure 2.7 | Additional soaking hits from docking and adenine-N3 vs -N9-alkylated isomers.

The protein structure (PDB: 6W02)³⁴, prepared for virtual screens is shown in green, predicted binding poses are shown in blue, the crystal protein structures are shown in grey, the solved fragment poses are shown in yellow, with alternative conformations shown in light pink. PanDDA event maps are shown as a blue mesh. Protein-ligand hydrogen bonds predicted by docking or observed in crystal structures are colored light blue or black, respectively. Hungarian RMSD values are presented between docked and crystallographically determined ligand poses.



Supplemental Figure 2.8 | Mac1 subsites compared to the adenine binding subsite in kinases and the oxyanion binding site in carboxylesterases.

A) Key features of the Mac1 adenine subsite are illustrated by the structure of ZINC26180281 (PDB: 5RSF). Hydrogen bonds are formed between the C6-amine of the adenine scaffold and the backbone nitrogen of Ile23, and between N1 of the adenine scaffold and the side-chain carboxylate of Asp22. The C2 amine of ZINC26180281 forms a non-canonical hydrogen bond to the backbone carbonyl oxygen of Ala154. **B)** Adenine recognition is similar in the pseudokinase domain of JAK2; however, the C6-amine forms a hydrogen bond to a backbone carbonyl oxygen rather than a side-chain carboxylate. Adenine binding occurs at the hinge residues that connect the N- and C-terminal lobes of the catalytic domain. Interactions that mimic adenine binding to the hinge residues are conserved in the majority of kinase inhibitors⁴³. Like ZINC2618028, kinase inhibitors exploit non-canonical hydrogen bonds. The 1,2,4-triazole derived inhibitor shown in **(C)** forms a hydrogen bond to the backbone carbonyl oxygen of Lys630. **D)** The fragment screens against Mac1 identified 47 oxyanions binding to the backbone nitrogens of Phe156 and Asp157. A comparable oxyanion recognition motif is present in acetylcholinesterase (AChE) **(E)**. In AChE, this motif stabilizes negative charge on the oxyanion transition state. **F)** Trifluoroacetic acid (TFA), present as a counter ion for ZINC3860798, was clearly defined in PanDDA event maps binding to the oxyanion subsite. TFA was also observed binding to the oxyanion subsite for fragments ZINC35185198 and ZINC51658946. The docking fragment ZINC263392672 also contained TFA, but no TFA was observed in the oxyanion subsite (PDB: 5RSG).



Supplemental Figure 2.9 | Comparison of DSF, HTRF, and ITC results for compounds tested in all assays.

A) Raw, un-normalized DSF data for the full measured temperature range (25 - 94 °C) demonstrates the absence of confounding changes in curve shape for all compounds. **B)** Normalized raw DSF data, enlarged to visualize compound-induced thermal shifts. **C)** Changes in T_{m_a} observed in the presence of fragments (0-3 mM fragment). **D)** Integrated heat peaks as a function of binding site saturation shown as black dots. The red line represents a non-linear least squares (NLLS) fit using a single-site binding model. **E)** Dose-response curves showing competition of the fragments with an ADPr-conjugated peptide for Mac1 binding. (*) ZINC901381520_N3 was tested in DSF and ITC, ZINC901381520_N9 was tested in HTRF.

Supplemental Data 2.1 | Summary of 1) summary of fragment screens, 2) amino acid sequences of constructs used for crystallography, 3) summary of the data collection strategy for the X-ray diffraction experiments, 4) data reduction and refinement statistics for all X-ray crystal structures reported in this work, 5) summary of all X-ray diffraction datasets collected, 6) classification of all fragment hits, and 7) solution binding data for selected fragments.

Supplemental Data 2.2 | Summary of 1) all fragments binding in the adenosine, catalytic and K90 sites of Mac1, 2) DSF data for all compounds tested, 3) ITC data for all compounds tested, and 4) HTRF peptide-displacement data for all compounds tested.

Acknowledgments

General

We thank all the staff of Diamond Light Source for providing support and encouragement which allowed us to carry out this work during the COVID-19 lockdown. We also thank Gemma Davison, Selma Dornen, James Sanderson, Matthew Martin, Mike Waring and Martin Noble (CRUK Newcastle Drug Discovery Unit, Newcastle University), Thomas Downes, Paul Jones, Hanna Klein, James Firth (York University), David Bajusz and Gyorgy Keseru (Hungarian Academy of Sciences) for providing fragment libraries. We also acknowledge EU-OPENSOURCE ERIC for providing its fragment library for the presented scientific work. EU-OPENSOURCE ERIC has received funding from European Union's Horizon 2020 research and innovation programme under grant agreement No 823893 (EU-OPENSOURCE-DRIVE). Structural biology applications used at UCSF were compiled and configured by SBGrid⁸². X-ray diffraction data processed at UCSF used the Wynton high-performance compute cluster.

Funding

This work was supported by NIH GM123159, NSF Rapid 2031205, and a TMC Award from the UCSF Program for Breakthrough Biomedical Research, funded in part by the Sandler Foundation (to J. Fraser); by the Wellcome Trust (grants 101794 and 210634), BBSRC (BB/R007195/1) and Cancer Research UK (C35050/A22284) (to I. Ahel); NIH R35GM122481 and DARPA HR0011-19-2-0020 (to B. Shoichet); GM071896 (to J. Irwin). The crystallographic screen at Oxford was supported by the XChem facility at Diamond Light Source (proposal ID MX27001). The Structural Genomics Consortium is a registered charity (number 1097737) that receives funds from AbbVie, Bayer Pharma AG, Boehringer Ingelheim, Canada Foundation for

Innovation, Eshelman Institute for Innovation, Genome Canada, Innovative Medicines Initiative (EU/EFPIA) [ULTRA-DD grant no. 115766], Janssen, Merck KGaA Darmstadt Germany, MSD, Novartis Pharma AG, Ontario Ministry of Economic Development and Innovation, Pfizer, São Paulo Research Foundation-FAPESP, Takeda, and Wellcome [106169/ZZ14/Z]. Beamline 8.3.1 at the Advanced Light Source is operated by the University of California Office of the President, Multicampus Research Programs and Initiatives grant MR-15-328599, NIH (R01 GM124149 and P30 GM124169), Plexxikon Inc., and the Integrated Diffraction Analysis Technologies program of the US Department of Energy Office of Biological and Environmental Research. The crystallographic fragment screens at UCSF were carried out using beamlines at the Advanced Light Source, the Stanford Synchrotron Radiation Lightsource and the National Synchrotron Light Source II. The Advanced Light Source (Berkeley, CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy under contract number DE-AC02-05CH11231, Office of Basic Energy Sciences. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). Extraordinary SSRL operations were supported in part by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act. This research used beamline 17-ID-2 of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National

Laboratory under Contract No. DE-SC0012704. The Center for BioMolecular Structure (CBMS) is primarily supported by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) through a Center Core P30 Grant (P30GM133893), and by the DOE Office of Biological and Environmental Research (KP1605010). The QCRG Structural Biology Consortium has received support from: Quantitative Biosciences Institute, Defense Advanced Research Projects Agency HR0011-19-2-0020 (to D. Agard. and K. Verba; B. Shoichet PI), FastGrants COVID19 grant (K. Verba PI), Laboratory for Genomics Research (O. Rosenberg PI) and Laboratory for Genomics Research LGR-ERA (R. Stroud PI). R.E. Díaz and T. Wu were supported by NSF GRFP. R.E. Díaz is a Howard Hughes Medical Institute Gilliam Fellow. I. Young was supported by NIH F32GM133129. M. Ferla was supported by the Wellcome Trust 203141/Z/16/Z and the NIHR Biomedical Research Centre Oxford. L. Martins was supported by CAPES-PriInt 88887.364628/2019-00.

Competing interests

N. Jura is a member of the SAB of Turning Point Therapeutics and SUDO Biosciences. A. Ashworth is a co-founder of Tango Therapeutics, Azkarra Therapeutics, Ovibio Corporation; a consultant for SPARC, Bluestar, ProLynx, Earli, Cura, GenVivo and GSK; a member of the SAB of Genentech, GLAdiator, Circle and Cambridge Science Corporation; receives grant/research support from SPARC and AstraZeneca; holds patents on the use of PARP inhibitors held jointly with AstraZeneca which he has benefited financially (and may do so in the future). B. Shoichet and J. Irwin are co-founders of a company, BlueDolphin LLC, that does fee-for-service docking. J. Fraser is a founder of Keyhole Therapeutics and a shareholder of Relay

Therapeutics and Keyhole Therapeutics. The Fraser laboratory has received sponsored research support from Relay Therapeutics.

Chapter 3 Structural characterization of ligand binding and pH-specific enzymatic activity of mouse Acidic Mammalian Chitinase.

Roberto Efraín Díaz^{1,2}, Andrew K. Ecker^{3,4,5}, Galen J. Correy¹, Pooja Asthana¹, Iris D. Young¹, Bryan Faust^{3, 6, 7}, Ian B. Seiple^{3,4}, Steven J. Van Dyken⁸, Richard M. Locksley^{9,10,11}, Michael C. Thompson¹², James S. Fraser^{1*}

Affiliations

¹ Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, USA

² Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

³ Department of Pharmaceutical Chemistry, University of California San Francisco, CA 94158, USA.

⁴ Cardiovascular Research Institute, University of California San Francisco, CA 94158, USA.

⁵ Chemistry and Chemical Biology Graduate Program, University of California San Francisco, CA 94158, USA.

⁶ Department of Biochemistry and Biophysics, University of California San Francisco, CA 94158, USA.

⁷ Biophysics Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

⁸ Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA

⁹ Department of Medicine, University of California, San Francisco, California 94143, USA

¹⁰ Department of Microbiology and Immunology, University of California, San Francisco,
California 94143, USA

¹¹ University of California, San Francisco, Howard Hughes Medical Institute, San Francisco,
California 94158, USA

¹² Department of Chemistry and Chemical Biology, University of California Merced, CA 95343,
USA.

Contributions

The work in this chapter is published as Díaz et al. in *bioRxiv* (2023). R.E. Díaz designed and cloned the construct that yielded the apo and holo mAMCase crystals; expressed, purified, and established crystallization conditions; collected X-ray diffraction data and processed data; modeled, refined, and analyzed structures; set up the 4MU-CB kinetic endpoint assay, collected, and analyzed data for the 4MU-CB assay; prepared the manuscript. A.K. Ecker designed the molecular dynamics simulation parameters, performed the simulations, and analyzed the data; developed the catalytic mechanism model for pH 2.0 and 6.5; prepared the manuscript. G.J. Correy vitrified crystals, collected X-ray diffraction data and processed data. P. Asthana collected X-ray diffraction data and processed data. I.D. Young processed the diffraction data. B. Faust assisted with mammalian cell culture for protein expression. I.B. Seiple provided access to the MOE software for MD simulations. S.J. Van Dyken guided the biochemistry work; prepared the manuscript. R. Locksley guided the biochemistry work; prepared the manuscript. M.C. Thompson guided the X-ray crystallography experiments, assisted with X-ray diffraction data processing and protein modeling. J.S. Fraser supervised work; prepared the manuscript; arranged funding.

CRedit Author Statement

Conceptualization, R.E.D., A.K.E., S.J.V.D., R.M.L., M.C.T., J.S.F. Methodology, R.E.D., A.K.E., B.F., M.C.T., J.S.F. Software, R.E.D. Investigation, R.E.D., A.K.E., G.J.C., P.A. Formal Analysis, R.E.D., A.K.E., G.J.C., I.D.Y., J.S.F. Visualization, R.E.D., A.K.E., J.S.F. Resources, B.F., I.B.S., J.S.F. Data Curation, R.E.D. Supervision, R.E.D., J.S.F. Project Administration,

J.S.F. Writing – Original Draft, R.E.D., A.K.E., J.S.F. Writing – Review & Editing, R.E.D.,
A.K.E., S.J.V.D., R.M.L., J.S.F. Funding Acquisition, R.E.D., J.S.F.

Abstract

Chitin is an abundant biopolymer and pathogen-associated molecular pattern that stimulates a host innate immune response. Mammals express chitin-binding and chitin-degrading proteins to remove chitin from the body. One of these enzymes, Acidic Mammalian Chitinase (AMCase), is known for its ability to function under acidic conditions in the stomach but is also active in tissues with more neutral pHs, such as the lung. Here, we used a combination of biochemical, structural, and computational modeling approaches to examine how the mouse homolog (mAMCase) can act in both acidic and neutral environments. We defined kinetic properties of mAMCase activity across a broad pH range, quantifying its unusual dual activity optima at pH 2 and 7. We also solved high resolution crystal structures of mAMCase in complex with chitin, where we identified extensive conformational ligand heterogeneity. Leveraging these data, we conducted molecular dynamics simulations that suggest how a key catalytic residue could be protonated via distinct mechanisms in each of the two environmental pH ranges. These results integrate structural, biochemical, and computational approaches to deliver a more complete understanding of the catalytic mechanism governing mAMCase activity at different pH. Engineering proteins with tunable pH optima may provide new opportunities to develop improved enzyme variants, including AMCase, for therapeutic purposes in chitin degradation.

Introduction

Chitin, a polymer of $\beta(1-4)$ -linked N-acetyl-D-glucosamine (GlcNAc), is the second most ubiquitous polysaccharide in nature. Chitin is abundant in numerous pathogens, such as nematode parasites, dust mites, and fungi¹⁻³, and is a pathogen-associated molecular pattern (PAMP) that activates mammalian innate immunity⁴. To mitigate constant exposure to environmental chitin, mammals have evolved unusual multi-gene loci that are highly conserved and encode chitin-response machinery, including chitin-binding (chi-lectins) and chitin-degrading (chitinases) proteins. Chitin clearance is particularly important for mammalian pulmonary health, where exposure to and accumulation of chitin can be deleterious. In the absence of AMCase, chitin accumulates in the airways, leading to epithelial stress, chronic activation of type 2 immunity, and age-related pulmonary fibrosis^{5,6}.

Humans express two active chitinases as well as five chitin-binding proteins that recognize chitin across many tissues⁷. Mitigating the negative effects of high chitin levels is particularly important for mammalian lung and stomach health. These tissues have distinct pH, with the lung environment normally \sim pH 7.0 and the stomach environment normally \sim pH 2.0, which raises the question of how chitin-response machinery has evolved to function optimally across such diverse chemical environments. Acidic Mammalian Chitinase (AMCase) was originally discovered in the stomach and named for its high enzymatic activity under acidic conditions. AMCase is also constitutively expressed in the lungs at low levels and overexpressed upon chitin exposure^{6,8,9}, suggesting this single enzyme has evolved to perform its function under vastly different chemical conditions.

AMCase is a member of the glycosyl hydrolase family 18 (GH18)¹⁰, and members of this family hydrolyze sugar linkages through a conserved two-step mechanism where the glycosidic

oxygen is protonated by an acidic residue and a nucleophile adds into the anomeric carbon leading to elimination of the hydrolyzed product (**Figure 3.1A**). This mechanism is corroborated by structures of different GH18 chitinases, most notably *S. marcescens* Chitinase A (PDB ID: 1FFQ)¹¹. In inhibitor bound structures for human AMCase (hAMCase; PDB ID: 3FY1), interactions mimicking the retentive post-cleavage transition state pre-hydrolysis of the oxazolinium intermediate are adopted by the nonhydrolyzable analogs^{12,13}.

Biochemical studies of mouse AMCase (mAMCase) measuring relative activity levels demonstrated a global maximum activity at acidic pH, but also a broad second local optimum near neutral pH¹⁴. This result suggested that mAMCase exhibits two distinct pH optima, which is unlike most enzymes that exhibit a shift or broadening of enzymatic activity across conditions¹⁵⁻¹⁷. For mAMCase the global maximum near pH 2.0 resembles the chemical environments of the stomach and the local maximum near pH 7.0 is similar to the environment of the lung. These two pH optima in the same enzyme suggest that mAMCase may employ different mechanisms to perform its function in different environments¹⁸. In contrast, the human homolog has maximal activity at pH 4.6 with sharply declining activity at more acidic and basic pH^{18,19}. This optimum corresponds with the pH of lung tissue in pulmonary fibrosis and other disease contexts, suggesting that hAMCase may have been selected for its ability to clear chitin from the lungs and restore healthy lung function.

The activity of mAMCase has been previously measured through endpoint experiments with limited insight into the rate of catalysis, substrate affinity, and potential substrate inhibition¹⁸. While the pH profile of mAMCase has been reported as a percentage of maximum activity at a given pH, it is unclear how the individual kinetic parameters (K_M or k_{cat}) vary¹⁴. These gaps have made it challenging to define the mechanism by which mAMCase shows

distinct enzymatic optima at different pHs. One possibility is that mAMCase undergoes structural rearrangements to support this adaptation. Alternatively, AMCase may have subtly different mechanisms for protonating the catalytic glutamic acid depending on the environmental pH .

In this work, we explore these hypotheses by employing biophysical, biochemical, and computational approaches to observe and quantify mAMCase function at different pHs. We measured the mAMCase hydrolysis of chitin, which revealed significant activity increase under more acidic conditions compared to neutral or basic conditions. To understand the relationship between catalytic residue protonation state and pH-dependent enzyme activity, we calculated the theoretical pKa of the active site residues and performed molecular dynamics (MD) simulations of mAMCase at various pHs. We also directly observed conformational and chemical features of mAMCase between pH 4.74 to 5.60 by solving X-ray crystal structures of mAMCase in complex with oligomeric GlcNAc_n across this range. Together these data support a model in which mAMCase employs two different mechanisms for obtaining a proton in a pH-dependent manner, providing a refined explanation as to how this enzyme recognizes its substrate in disparate environments.

Results

New assay confirms broad pH profile for mAMCase

Prior studies have focused on relative mAMCase activity at different pH^{18,20}, limiting the ability to define its enzymological properties precisely and quantitatively across conditions of interest. To expand upon these previous observations of dual optima in mAMCase activity at pH 2.0 and 7.0, we measured mAMCase activity *in vitro* across a broad pH range. We developed an approach that would enable direct measurement of k_{cat} and K_{M} for mAMCase across a broad pH range by modifying a prior assay that continuously measures mAMCase-dependent breakdown of a fluorogenic chitin analog, 4-methylumbelliferone (4MU) conjugated chitobioside. To overcome the pH-dependent fluorescent properties of 4MU-chitobioside, we reverted the assay into an endpoint assay, which allowed us to measure substrate breakdown across different pH²¹ (**Supplemental Figure 3.1A**).

We conducted our endpoint assay across a pH range of 2.0 to 7.4 to reflect the range of physiological conditions at its *in vivo* sites of action (**Figure 3.1B**; Data available at doi: 10.5281/zenodo.7968293). We then derived the Michaelis-Menten parameters at each pH unit measured (**Supplemental Figure 3.2A-C**). We found that mAMCase has maximum activity at pH 2.0 with a secondary local maximum at pH 6.5, pointing to a bimodal distribution of activity across pH. This is consistent with the relative activity measurements previously performed on mAMCase, but distinct from a single broad pH range, as has been observed for k_{cat} of hAMCase^{14,18}. The two maxima at pH 2.0 and 6.5 are an approximate match the pH at the primary *in vivo* sites of mAMCase expression, the stomach and lungs, respectively¹⁸. These observations raise the possibility that mAMCase, unlike other AMCase homologs, may have evolved an unusual mechanism to accommodate multiple physiological conditions.

We also found that pH primarily improves the rate of mAMCase catalysis 6.3-fold (k_{cat} ; **Figure 3.1C**), whereas K_M (**Figure 3.1D**) worsens 2.5-fold from pH 7.4 to pH 2.0. Similar to chitotriosidase the other active chitinase in mammals and also a GH18 chitinase, we observe an apparent reduction in the rate of mAMCase catalysis across all pH units measured at 4MU-chitobioside concentrations above 80 μM , which suggests that mAMCase may be subject to product inhibition²². The underlying mechanism for the observed product inhibition is that mAMCase can transglycosylate the catalysis products, as has been previously observed at pH 2.0 and 7.0²³. This potential product inhibition leads to a systematic underprediction of rates by the Michaelis-Menten model at high substrate concentrations. The catalytic efficiency (k_{cat}/K_M) of mAMCase may not capture the effects of product inhibition given that these constants reflect sub-saturating substrate concentrations. Independent of the potential for product inhibition, the trend that mAMCase has highest k_{cat} at very low pH and another local optimum at more neutral pH is clear. We hypothesize that these activity data resemble two overlapping activity distributions, suggesting that the rate at lower pH activity is dependent on the concentration of free protons in solution and that the higher pH optimum results from a distinct mechanism (**Figure 3.1E**).

Characterization of mAMCase ligand occupancy and conformational heterogeneity.

Our biochemical analyses led us to hypothesize that the pH-dependent activity profile of mAMCase is linked to the mechanism by which catalytic residues are protonated. Previous structural studies on AMCase have focused on interactions between inhibitors like methylallosamidin and the catalytic domain of the protein. We built on these efforts by solving the structure of mAMCase in complex with chitin oligomers of varying length (GlcNAc_n). We

used chitin oligomers because they are chemically identical to polymeric chitin found in nature but are soluble and therefore more amenable for co-crystallization than crystalline chitin is. We successfully determined high resolution X-ray crystal structures of the apo mAMCase catalytic domain at pH 5.0 or 8.0 (PDB ID: 8FG5, 8FG7) and holo mAMCase catalytic domain between pH 4.74 to 5.60 in complex with either GlcNAc₂ or GlcNAc₃ (PDB ID: 8GCA, 8FRC, 8FR9, 8FRB, 8FRD, 8FRG, 8FRA; **Supplemental Figure 3.3A,B; Table 3.1**).

Across these different datasets we observed complex ligand density in the active site of mAMCase. In all our datasets, we observed continuous ligand density that resembled higher order chitin oligomers (e.g., GlcNAc₄, GlcNAc₅, or GlcNAc₆). This observation was confusing given that these structures were co-crystallized with either GlcNAc₂ or GlcNAc₃ oligomers. For example, due to the continuous nature of ligand density observed in our mAMCase-GlcNAc₃ co-crystal structure at pH 4.74 (PDB ID: 8GCA, chain A), we initially modeled hexaacetylchitohexaose (H-(GlcNAc)₆-OH) into the -4 to +2 sugar-binding subsites, using the nomenclature for sugar-binding subsites from Davies et al.²⁴. This nomenclature defines the sugar-binding subsites as $-n$ to $+n$, with $-n$ corresponding to the non-reducing end and $+n$ the reducing end.

We next continued with a modeling approach that replaced higher order oligomer models with models that only used the chemically defined oligomers present in the crystallization drop. To accomplish this modeling of different binding poses, we placed multiple copies of these oligomers consistent with an interpretation of extensive conformational heterogeneity (**Supplemental Figure 3.5D**). In one sample co-crystallized with GlcNAc₃ at pH 4.74 (PDB ID: 8GCA, chains A-B), we identified ligand density that was consistent with GlcNAc₂, suggesting

that some hydrolysis occurs in the crystal. The resulting model includes compositional heterogeneity as there are both types of oligomers present.

Therefore, across all of our datasets, we modeled a combination of ligand binding events consisting of overlapping GlcNAc₂ or GlcNAc₃ molecules at each sugar-binding site, i.e., GlcNAc₂ ResID 401 Conf. A occupied subsites -3 to -2 while GlcNAc₂ ResID 401 Conf. C occupied subsites -2 to -1. By providing each ligand molecule with an alternative conformation ID, this allowed both occupancies and B-factors to be refined (**Figure 3.2A,B,C**; additional details in Methods). Across these different datasets, we observed ligand density for different combinations of occupancy over the -4 to +2 sugar-binding subsites (**Figure 3.2A**). While modeling chito-oligomers into strong electron density, we observed strong positive difference density between sugar-binding subsites near the C2 N-acetyl and the C6' alcohol moieties. Using the non-crystallographic symmetry (NCS) “ghost” feature in *Coot*, we were then able to observe that the positive difference density between ligand subsites in one chain could be explained by the dominant ligand pose observed in another associated crystallographic chain, suggesting the presence of a low-occupancy binding events. This observation led to the discovery that GlcNAc_n occupies intermediate subsites, which we label $n+0.5$, continuing to follow the nomenclature established by Davies et al., in addition to canonical sugar-binding subsites (**Figure 3.2B**)²⁴.

In addition to identifying novel $n+0.5$ sugar-binding subsites, we also observed strong positive difference density above the +1 subsite., which we label +1'. During ligand refinement, we observed density for both the α - and β -1,4-linked GlcNAc₂ anomers in the active site. This unexpected configurational heterogeneity, which is observable because of the high resolution of our datasets (1.30 - 1.95 Å), likely formed as a result of equilibration between the two anomers through an oxocarbenium close-ion-pair intermediate. The ability for the active site to

accommodate and stabilize these ligands is important given its role in degrading crystalline chitin, a complex and often recalcitrant substrate that likely requires multiple binding events by AMCase before degradation can occur. We did not identify consistent trends between the contents of the crystallization drop (pH, substrate identity, and substrate concentration), the diffraction parameters (space group, unit cell dimensions, resolution), and the resulting density in the active site; however, as outlined below, the protein conformations and substrate states are highly correlated. Collectively, modeling a combination of ligand binding modes, linkages, and anomers allowed us to interpret the resulting coordinates in a more complete model of how mAMCase coordinates and stabilizes polymeric chitin for catalysis (**Figure 3.2; Supplemental Figure 3.5D; Supplemental Figure 3.6A; Table 3.2**).

Structural characterization of mAMCase catalytic triad D₁xD₂xE.

We interpreted the protein-ligand interactions along the canonical binding sites (**Supplemental Figure 3.5C,D**). As with other chitinases, we observe a network of tryptophans consisting of Trp31, Trp360, Trp99, and Trp218 stabilizing the positioning of the ligand into the binding site through a series of H- π interactions with the -3, -1, +1, and +2 sugars, respectively²⁵⁻²⁷. These interactions are primarily with the axial hydrogens of the respective sugars but also include the N-H of the -3 and +1 sugar and the 6' O-H of the +2 sugar (**Supplemental Figure 3.6B**). Further, we observe Asp213 accepting a hydrogen bond with the 6' OH of the -1 sugar and Tyr141 acting as a hydrogen bond donor to the 6' OH of the +1 sugar. These two hydrogen bonds likely orient the ligand in the catalytically competent pose where the glycosidic oxygen bridging the -1 and +1 sugars is 2.8 Å away from the acidic Glu140 -OH (**Supplemental Figure 3.6C**). With this proximity, Glu140 can act as a hydrogen bond donor to the strained (122° bond

angle) bridging oxygen forming a hydrogen bond to promote the formation of an oxazolinium intermediate and subsequent cleavage of the glycosidic bond. We observed two interactions with the sugar in the -4 position supporting the ligand orientation far from the enzymatic active site. Residues involved in ligand binding and catalysis adopt similar side chain conformations in the absence of ligand (PDB ID: 8FG5, 8FG7), suggesting that the active site is organized prior to ligand binding and not subject to ligand-stabilized conformational changes.

We hypothesize that the +1' subsite is primarily occupied by the catalysis product GlcNAc₂ prior to its displacement from the active site by subsequent sliding of polymeric chitin (**Figure 3.2B**)²⁸. At this position, Trp99 and Trp218 engage in CH- π interactions with the +1 and +2 sugars, respectively while Asp213 forms a new H-bond with the carbonyl oxygen and Tyr141 retains an H-bond with the hydroxyl moiety on the +1 sugar. We are able to observe this post-catalysis binding mode due to the stabilizing interactions between GlcNAc₂ and Asp213, Trp99, Trp218, and Tyr141 (**Supplemental Figure 3.6C**). Together, these observations highlight the dynamic chitin binding modes within the mAMCase active site. Collectively, the observed non-canonical binding modes of these sugars is consistent with previous observations that once bound to polymeric chitin, GH18 chitinases engage in chain sliding from the reducing end of the substrate following catalysis²⁹.

In contrast to the largely static interactions outlined above, we observed conformational heterogeneity in the catalytically critical Asp138 residue, suggesting flipping between two equally stable states facing each of the other two residues in the catalytic triad (Asp136 or Glu140)³⁰. Using Ringer, we confirmed that there are two Asp138 conformations and only a single conformation for Asp136 and Glu140 (**Supplemental Figure 3.7**; Data available at doi: 10.5281/zenodo.7758815)³¹. Across 20 chains from the datasets derived from different pH and

co-crystallization conditions (**Table 3.1**), we quantified whether Asp138 is preferentially oriented towards Asp136 (*inactive* conformation) or preferentially oriented towards Glu140 (*active* conformation).

Prior work has suggested that Asp138 orients itself towards Glu140 to promote stabilization of the substrate's twisted boat conformation in the -1 subsite. Therefore, we explored if Asp138 conformation is correlated with ligand pose^{13,30,32,33}. As previously mentioned, we assign alternative conformation IDs to each ligand molecule based on its subsite positioning. We calculate subsite occupancy by taking the sum of all alternative ligand conformations at a given subsite, i.e., the occupancy of subsite -2 is equal to the occupancies of GlcNAc₂ ResID 401 Conf. A and GlcNAc₂ ResID 401 Conf. C (**Figure 3.3A**; see Methods for additional details; Data available at doi: 10.5281/zenodo.7905828). We observe a strong positive correlation between Asp138 conformation and ligand pose only in the -2 to +1 subsites (**Figure 3.3B**; **Table 3.2**). When the -1 subsite is at least 50% occupied, Asp138 prefers the *active* conformation (up towards Glu140). In this orientation, Asp138(HD2) forms a H-bond with Glu140(OE1) (2.6 Å) while Asp138(OD1) forms an H-bond with the amide nitrogen of GlcNAc in the -1 subsite (2.6 Å). Glu140(OE2) is 2.8 Å away from the glycosidic oxygen bridging the -1 and +1 sugars. We suspect that the inverse correlation between Asp138 *active* conformation and the -2.5 and -1.5 sugar-binding subsites represents ligand translocation towards the catalytic residues, prior to enzyme engagement with the ligand. When chitin occupies a canonical sugar-binding subsite, AMCcase forms stabilizing H-bonds with the ligand prior to catalysis. These observations are consistent with the proposed catalytic mechanism where upon protonation, the equilibrium between Asp138 conformations shifts to favor the *active* conformation (towards

Glu140) where Asp138 stabilizes Glu140 in proximity to the glycosidic oxygen prior to catalysis.

Theoretical pKa calculations of mAMCase catalytic triad D₁xD₂xE.

Based on the dual pH optimum observed in our kinetics assay and the conformational heterogeneity of Asp138, we calculated the theoretical pKa for catalytic D₁xD₂xE motif on mAMCase using PROPKA 3.0. PROPKA does not account for alternative conformations in its calculations, so we split our protein models to contain single conformations of the catalytic residues Asp136, Asp138, and Glu140. While PROPKA does account for ligands in its calculations, running the calculations with different alternative conformations of GlcNAc₂ or GlcNAc₃ had little effect on the calculated pKas for the active site residues (**Supplemental Figure 3.4**; Data available at doi: 10.5281/zenodo.7905863). Despite the observed ligand heterogeneity, we observe a relatively narrow range of pKa values for the catalytic triad. This suggests that the pKa of the catalytic residues is primarily influenced by the position of nearby residues and that the placement of solvent or ligand molecules has little effect. When Asp138 is oriented towards Asp136 (*the inactive* conformation), the pKa of the catalytic residues are 2.0, 13.0, 7.7 for Asp136, Asp138, and Glu140 respectively. Similarly, when Asp138 is oriented towards Glu140 (*the active* conformation), the pKa of the catalytic residues are 3.4, 12.4, 6.4 for Asp136, Asp138, and Glu140 respectively. Taking this information together, it is clear that the pKa of Asp136 and Glu140 are both affected by the orientation of Asp138 (**Figure 3.4A**; **Table 3.3**). The pKa of Asp136 suggests that at pH > 3.4, Asp136 is deprotonated, and its conjugate base is more stable. We observe a similar pKa distribution for the catalytic triad in human

AMCase and other GH18 chitinases with publicly available structures and optimum pH activity profiles (**Figure 3.4A-C**).

Given the pH range of our crystallization conditions, we expect that Asp136 is deprotonated while Asp138 and Glu140 are protonated. We hypothesize that this anionic aspartate can form a strong ionic hydrogen bond interaction with Asp138 orienting it in the *inactive* conformation. When Asp136 is protonated to its aspartic acid state, $\text{pH} < 3.2$, we expect that it is only capable of forming the relatively weaker neutral hydrogen bond with Asp138 lowering the favorability of the *inactive* conformation.

Additionally, when interpreting the pKa of Glu140, we hypothesize that under acidic conditions ($\text{pH} 2.0 - 6.5$), Glu140 can obtain its catalytic proton from solution. The accessibility of Asp138's proton to Glu140 progressively decreases as pH increases from pH 2.0 to 6.5. In contrast, under neutral and basic conditions ($\text{pH} 6.0 - 7.4$), Asp138 can shuttle a proton from Asp136 by rotating about its $\text{C}\alpha\text{-C}\beta$ bond to supply Glu140 with the proton. Glu140 subsequently uses the proton that it obtained from Asp138 to protonate the glycosidic bond in chitin, promoting hydrolysis as previously described in several chitinases^{30,34,35}. While this mechanism could explain how mAMCase has a local optimum at pH 2.0, it is insufficient to explain why we do not observe a similar optimum in hAMCase. The narrow range of pKa values across GH18 chitinases suggest that differences in optimal activity by pH may be influenced by other factors, such as protein stability, conformational dynamics, or coordination of distal GlcNAc residues by ionizable residues³⁶.

Molecular Dynamics

Based on our enzymology results suggesting the possibility of differential activity between acidic pH (pH 2.0) and near neutral pH (pH 6.5) and theoretical pKa calculations of the active site residues, we performed short atomistic molecular dynamics simulations to interrogate the movement of catalytic residues. While all the crystal structures we obtained were collected in a narrow acidic pH range between 4.74 - 5.60, we ran simulations at pH 2.0 and pH 6.5, ensuring that the protonation states of side chains populated by 3DProtonate were supported by our PROPKA calculations (Data available at doi: 10.5281/zenodo.7758821)^{37,38}. These simulations allowed us to investigate our hypothesis that at neutral pH mAMCase enzymatic activity is dependent on the protonation state of Asp136. We performed simulations using protein models that contain Asp138 in either the *inactive* (down towards Asp136; “*inactive* simulation”) or *active* conformation (up towards Glu140; “*active* simulation”) to avoid bias from the starting conformation.

In all our simulations, we observe that Glu140 orients its acidic proton towards the glycosidic bond between the -1 and +1 sugars and fluctuates between 1.5 to 2.3 Å for the duration of the simulation, with a median distance of 1.8 Å. The positioning of this proton is necessary to allow for the oxocarbenium cleavage of the glycosidic bond and recapitulates the positioning of Glu140 in our experimental structures. In simulations initiated from the *inactive* conformation at pH 2.0, we observe that Asp 138 is readily able to rotate about its C α -C β bond to adopt the *active* conformation forming the same hydrogen bond between Asp138 and Glu140. In contrast, from simulations at pH 6.5 started from the Asp138 *inactive* conformation, we observe that Asp138 remains hydrogen bonded to Asp136 throughout the duration of the simulation (*inactive* conformation; **Figure 3.5A-C**; Data available at doi:

10.5281/zenodo.7758821). This series of simulations allowed us to better visualize which catalytic side chains are dynamic and which catalytic side chains positioning are well maintained to help build our catalytic mechanism.

Discussion

mAMCase is an unusual enzyme that can bind and degrade polymeric chitin in very different pH environments. We hypothesized that mAMCase employs different mechanisms under acidic and neutral pH to protonate its catalytic glutamate. Through our analysis, we argue that the observed ligand and catalytic residue densities and occupancies in our crystal structures are consistent with the previously proposed GH18 catalytic mechanism³⁹. By modeling GlcNAc₂ as sequentially overlapping ligands in alternative conformations (**Figure 3.2**), we are able to visualize each step in the proposed catalytic cycle of mAMCase (**Figure 3.6**). This mechanism, which has been observed in other glycoside hydrolases, occurs when the glycosidic oxygen is protonated by an acidic residue and a nucleophile adds into the anomeric carbon leading to elimination of the hydrolyzed product.

Based on our crystal data and simulations, we envision that AMCase begins its catalytic cycle by searching for a ligand. At neutral pH, Asp136 is deprotonated (pKa = 2.1) forming an ionic hydrogen bond with Asp138 (pKa = 13.1). In contrast, at low pH Asp136 is protonated, yet continues to form a weaker hydrogen bond with Asp138 (**Figure 3.6** - Step 1). Glu140 (pKa = 7.7) is protonated across the enzyme's active pH range. Upon ligand binding (**Figure 3.6** - Step 2), Glu140 stabilizes the sugar at the -1 subsite. The ligand then translocates forward by one GlcNAc₂ to occupy the +1 and +2 subsites (**Figure 3.6** - Step 3). At neutral pH, Asp136 is predominantly deprotonated. When protonation of Asp136 occurs, this destabilizes the Asp136-Asp138 hydrogen bond and allows Asp138 to rotate about its C α -C β bond into the *active* conformation (towards Glu140). However, since Asp136 is always protonated at low pH, the Asp136-Asp138 hydrogen bond is less energetically favorable, therefore Asp138 can adopt the *active* conformation more readily (**Figure 3.6**- Step 4).

Once Asp138 is in the *active* conformation, Asp138 and Glu140 form stabilizing interactions with the N-acetyl of the ligand, priming it to become the nucleophile required for catalysis (**Figure 3.6** - Step 4). Glu140 provides its ionizable proton to the ligand's glycosidic oxygen, increasing the electrophilicity of the anomeric carbon (**Figure 3.6** - Step 5)⁴⁰. The carbonyl oxygen of the -1 sugar N-acetyl then nucleophilically adds into the anomeric carbon from the β face to cleave the glycosidic bond, forming the oxazolinium intermediate. At neutral pH, the resultant deprotonated Glu140 is then re-protonated through a proton shuttling where Asp136 donates its proton to Asp138 and Asp138 donates its ionizable proton to Glu140. At acidic pH we argue that Glu140 can be directly re-protonated by a proton in solution (**Figure 3.6** - Step 5). At a neutral pH this leads to Asp138 returning to an *inactive* conformation. However, at low pH Asp136 and Glu140 are both protonated due to the high concentration of protons in solution, allowing Asp138 to remain in the *active* conformation and form stabilizing interactions with the N-acetyl group on the ligand. The oxazolinium intermediate is then hydrolyzed by a water molecule, generating a GlcNAc₂ catalysis product in the +1 and +2 sugar subsites (**Figure 3.6** - Step 6). The GlcNAc₂ product dissociates from the +1 to +2 sugar subsites, then the ligand undergoes "decrystallization" and "chain sliding" before restarting its catalytic cycle, assuming AMCase is bound to its natural substrate²⁹. At neutral pH this catalytic mechanism is reset with Asp138 in its *inactive* conformation, however at low pH the catalytic mechanism is reset with Asp138 already in the *active* conformation. This could lead to faster rates of catalysis at lower pH compared to the neutral pH mechanism, providing a possible explanation for our observed pH profile.

While our model proposes a plausible explanation of why mAMCase is highly active at pH 2, it does not explain why hAMCase has a single activity optimum around pH 5. Prior work

by Kashimura et al. has demonstrated that *E. coli* expressed mAMCase is remarkably stable across a broad pH range⁴¹. Similar experiments have not yet been performed on hAMCase. Olland et al. previously identified Arg145, His208, and His269 as important for pH specificity¹³. Seibold et al. argued that hAMCase isoforms containing asthma protective mutations N45D, D47N, and M61R, which are wildtype in mAMCase, may influence the pKa of Asp138-Glu140 by undergoing structural rearrangement¹⁸. Tabata et al. identified mutations across the course of evolution in Carnivora that were inactivating or structurally destabilizing (loss of S-S bonds)⁴². To this end, we explored sequence differences between mouse and human AMCase homologs for insight into why mAMCase has such high enzymatic activity at pH 2.0 and 6.5 compared to hAMCase. We identified ionizable residues on mAMCase that likely contribute to its overall stability and are not present in hAMCase. Mutations Lys78Gln, Asp82Gly, and Lys160Gln result in the loss of surface stabilizing salt bridges in hAMCase and may contribute to its reduced activity at more acidic pH. Therefore, it is most likely that the dual pH optima of mAMCase is intrinsic to the catalytic mechanism, where Glu140 can be protonated directly from solution (at low pH) or through proton shuttling across the catalytic triad (at neutral pH; **Figure 3.1E**). However, hAMCase is likely too destabilized at low pH to observe an increase in k_{cat} . hAMCase may be under less pressure to activity at low pH due to humans' noninsect-based diet compared to other mammals with primarily insect-based diets⁴².

Together, these data demonstrate the importance of using structural and biochemical assays to develop our understanding of the catalytic mechanism governing mAMCase activity. Using biochemical and structural methods, we have begun to develop a detailed model of how AMCase fulfills its role in chitin recognition and degradation. Small chitin oligomers are ideal for measuring the ability of AMCase to cleave β -1,4-glycosidic linkages between GlcNAc units,

but these small oligomers do not represent the complex crystalline chitin encountered by AMCase in the lung. It is difficult to extrapolate the effects we observe using small chitin oligomers to binding (k_{on}), processivity (k_{proc}), catalysis (k_{cat}), or product release (k_{off}) on the native large and heterogeneous oligomeric substrates. In the future, we hope to be able to directly visualize the mAMCase-chitin interactions and characterize each step of the catalytic mechanism including decrystallization, degradation, product release, and chain sliding (also known as processivity).

To further understand the impact of pH on the structure of AMCase, it will be necessary to crystallize AMCase across a broader pH range that may expose conformational and structural changes that contribute to mAMCase's unique pH activity profile. Further, neutron diffraction crystallography could provide novel critical insight into the placement of protons across the active site and help to develop a more complete model of mAMCase's catalytic mechanism at different pH. Understanding the mechanistic basis behind an enzyme's dual pH optima will enable us to engineer proteins with tunable pH optima to develop improved enzyme variants for therapeutic purposes for diseases, such as asthma and lung fibrosis.

Tables

Table 3.1 | Data collection and refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses.

Dataset	Apo at 100 K	Apo at 277 K	Holo with GlcNAc3 at pH 4.74	Holo with GlcNAc2 at pH 4.91	Holo with GlcNAc2 at pH 5.08	Holo with GlcNAc2 at pH 5.25	Holo with GlcNAc2 at pH 5.25	Holo with GlcNAc2 at pH 5.43	Holo with GlcNAc2 at pH 5.60
PDB ID	8FG5	8FG7	8GCA	8FRC	8FR9	8FRB	8FRD	8FRG	8FRA
Diffraction Data DOI	10.18430/M38FG5	10.18430/M38FG7	10.18430/M38GCA	10.18430/M38FRC	10.18430/M38FR9	10.18430/M38FRB	10.18430/M38FRD	10.18430/M38FRG	10.18430/M38FRA
pH	5.00	8.00	4.74	4.91	5.08	5.25	5.25	5.43	5.60
Ligand	N/A	N/A	GlcNAc3	GlcNAc2	GlcNAc2	GlcNAc2	GlcNAc2	GlcNAc2	GlcNAc2
[Ligand] mM	N/A	N/A	12.67	29.00	19.33	19.33	29.00	29.00	19.33
Wavelength	1.117	1.116	1.116	1.116	1.116	1.116	1.116	1.116	1.116
Resolution range	46.8 - 1.3 (1.346 - 1.3)	50.88 - 1.64 (1.699 - 1.64)	61.83 - 1.7 (1.761 - 1.7)	69.52 - 1.92 (1.989 - 1.92)	69.59 - 1.5 (1.554 - 1.5)	57.29 - 1.7 (1.761 - 1.7)	58.67 - 1.68 (1.74 - 1.68)	69.59 - 1.741 (1.803 - 1.741)	86.27 - 1.95 (2.02 - 1.95)
Space group	P 1 21 1	P 21 21 21	P 21 21 2	P 2 21 21	P 2 21 21	P 21 21 21	P 2 21 21	P 21 21 2	P 21 21 21
Unit cell (length)	60.04 42.25 67.41	63.6466 71.8436 84.6724	76.0664 91.7195 106.132	70.9333 92.6896 105.123	71.1131 92.6412 105.423	91.9263 106.963 146.492	70.755 92.451 104.99	92.8934 105.041 70.8116	92.0659 106.705 146.57
Unit cell (angles)	90 95.18 90	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90	90 90 90
Total reflections	2099252 (194837)	620486 (61796)	516529 (48842)	339863 (33874)	702566 (63651)	1010525 (98078)	499250 (48902)	420425 (37138)	691049 (67775)
Unique reflections	83050 (8251)	47999 (4678)	82111 (8079)	53587 (5242)	109106 (10560)	158679 (15679)	78153 (7593)	71329 (6974)	105512 (10401)
Multiplicity	25.3 (23.6)	12.9 (13.2)	6.3 (6.0)	6.3 (6.5)	6.4 (6.0)	6.4 (6.3)	6.4 (6.4)	5.9 (5.3)	6.5 (6.6)
Completeness (%)	99.99 (99.98)	99.37 (98.65)	99.72 (99.42)	99.88 (99.79)	97.48 (95.47)	99.87 (99.88)	98.71 (97.03)	99.56 (99.03)	99.74 (99.62)
Mean I/sigma(I)	13.31 (1.88)	7.00 (1.19)	8.83 (3.12)	7.72 (3.21)	16.77 (5.46)	9.09 (3.10)	9.68 (3.09)	6.18 (2.56)	5.65 (1.26)
Wilson B-factor	15.81	16.38	12.17	13.44	9.16	12.47	11.55	15.76	12.64
R-merge	0.1342 (2.107)	0.2489 (2.119)	0.1811 (1.138)	0.1531 (0.5265)	0.06539 (0.2976)	0.1111 (0.5593)	0.1155 (0.569)	0.1321 (0.4674)	0.1619 (0.6276)
R-meas	0.137 (2.153)	0.2591 (2.203)	0.1972 (1.242)	0.1669 (0.5728)	0.07122 (0.3259)	0.121 (0.61)	0.126 (0.6197)	0.1448 (0.5188)	0.176 (0.6822)
R-pim	0.02718 (0.4382)	0.07097 (0.5968)	0.07709 (0.4917)	0.06573 (0.2233)	0.02784 (0.1311)	0.04745 (0.2411)	0.04965 (0.2425)	0.05834 (0.2207)	0.06836 (0.2647)
CC1/2	0.999 (0.858)	0.996 (0.502)	0.997 (0.805)	0.994 (0.884)	0.999 (0.943)	0.997 (0.888)	0.993 (0.68)	0.994 (0.845)	0.997 (0.845)
CC*	1 (0.961)	0.999 (0.818)	0.999 (0.944)	0.998 (0.969)	1 (0.985)	0.999 (0.97)	0.998 (0.9)	0.998 (0.957)	0.999 (0.957)
Reflections used in refinement	83046 (8251)	47968 (4677)	82030 (8059)	53543 (5242)	109065 (10557)	158531 (15678)	78103 (7592)	71295 (6967)	105380 (10401)

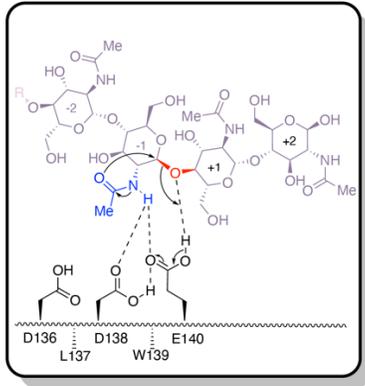
Dataset	Apo at 100 K	Apo at 277 K	Holo with GlcNAc3 at pH 4.74	Holo with GlcNAc2 at pH 4.91	Holo with GlcNAc2 at pH 5.08	Holo with GlcNAc2 at pH 5.25	Holo with GlcNAc2 at pH 5.25	Holo with GlcNAc2 at pH 5.43	Holo with GlcNAc2 at pH 5.60
Reflections used for R-free	4099 (422)	2328 (234)	4142 (427)	2738 (273)	5449 (559)	7978 (802)	3878 (334)	3561 (348)	5174 (542)
R-work	0.1317 (0.2361)	0.1469 (0.2707)	0.1598 (0.2428)	0.1472 (0.1616)	0.1376 (0.1615)	0.1423 (0.1850)	0.1396 (0.1724)	0.1657 (0.2194)	0.1695 (0.2074)
R-free	0.1519 (0.2613)	0.1717 (0.3244)	0.1978 (0.2952)	0.1898 (0.2065)	0.1644 (0.1932)	0.1778 (0.2315)	0.1689 (0.2113)	0.2083 (0.2737)	0.2056 (0.2463)
CC(work)	0.970 (0.583)	0.978 (0.789)	0.969 (0.819)	0.953 (0.846)	0.971 (0.922)	0.970 (0.878)	0.963 (0.903)	0.959 (0.749)	0.961 (0.869)
CC(free)	0.969 (0.558)	0.975 (0.729)	0.953 (0.775)	0.951 (0.793)	0.966 (0.910)	0.958 (0.791)	0.954 (0.882)	0.951 (0.757)	0.970 (0.846)
Number of non-hydrogen atoms	3583	3427	7330	6953	7507	13986	6951	7343	14428
macromolecules	3107	3097	6094	6016	6186	11938	6019	6286	11900
ligands	1	1	394	342	516	746	344	401	571
solvent	475	329	1034	763	1057	1666	756	852	2237
Protein residues	376	376	752	738	750	1478	738	738	1478
Nucleic acid bases									
RMS(bonds)	0.006	0.008	0.008	0.007	0.01	0.006	0.007	0.008	0.003
RMS(angles)	0.88	0.96	1.05	0.91	1.1	0.92	0.91	1.12	0.66
Ramachandran favored (%)	98.4	98.66	98.8	98.23	98.26	98.84	98.64	98.35	98.1
Ramachandran allowed (%)	1.6	1.34	1.2	1.77	1.74	1.16	1.36	1.65	1.9
Ramachandran outliers (%)	0	0	0	0	0	0	0	0	0
Rotamer outliers (%)	1.22	0.92	0.62	0.79	0.92	0.87	0.63	0.6	0.88
Clashscore	1.66	0.83	1.25	1.85	1.3	1.31	1.6	1.44	1.66
Average B-factor	21.71	19.1	16.09	14.55	12.73	15.72	14.2	17.9	15.9
macromolecules	19.83	17.9	13.9	13.24	10.3	13.76	12.5	16.36	13.88
ligands	98.88	46.35	23.57	18.87	15.73	17.53	15.9	23.5	19.18
solvent	33.82	30.3	27.53	23.9	26.25	29.3	27.32	27.98	26.24
Number of TLS groups									

Table 3.2 | Occupancy of each ligand subsite and Asp138 in the *active* conformation (separate file).

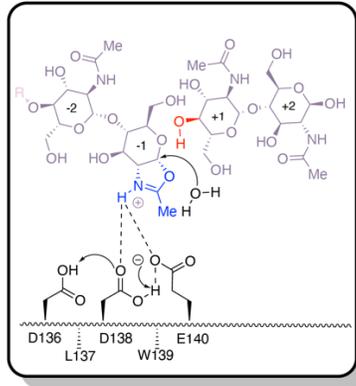
Table 3.3 | pKa across Asp136, Asp138, Glu140 of mAMCase structures in either Asp138 *inactive* or Asp138 *active* conformation (separate file).

Figures

A) Glycosidic oxygen is directed by acidic residue

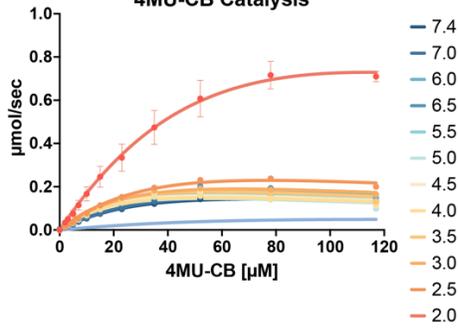


Glycosidic oxygen is eliminated through nucleophilic formation of oxazolinium

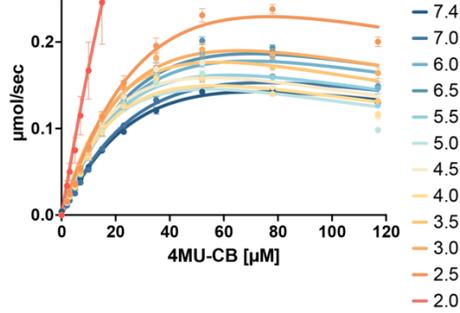


B)

Initial Rate of 4MU-CB Catalysis

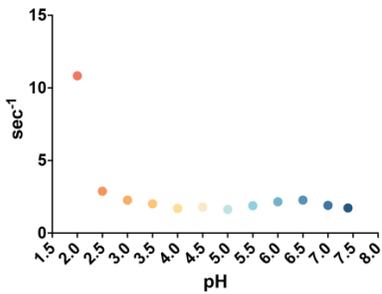


Initial Rate of 4MU-CB Catalysis



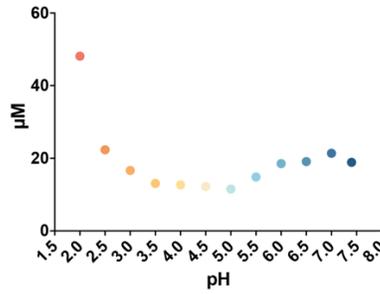
C)

k_{cat}



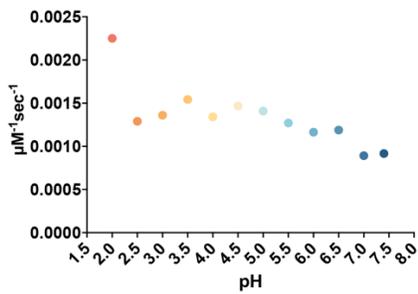
D)

K_M



E)

Catalytic Efficiency



F)

Catalytic Activity

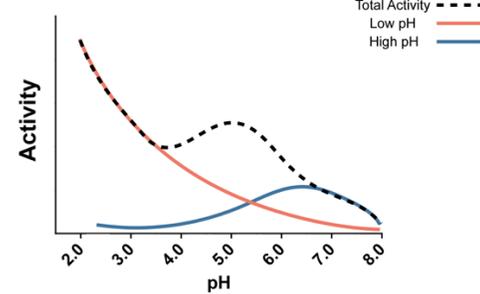


Figure 3.1 | Kinetic properties of mAMCase catalytic domain at various pH.

A) Chemical depiction of the conserved two-step mechanism where the glycosidic oxygen is protonated by an acidic residue and a nucleophile adds into the anomeric carbon leading to elimination of the hydrolyzed product. **B)** The rate of 4MU-chitobioside catalysis (1/sec) by mAMCase catalytic domain is plotted as a function of 4MU-chitobioside concentration (μM). Each data point represents $n = 4$ with error bars representing the standard deviation. Michaelis-Menten equation without substrate inhibition was used to estimate the k_{cat} and K_{M} from the initial rate of reaction at various substrate concentrations. **C)** The rate of substrate turnover (1/sec) by mAMCase catalytic domain is plotted as a function of pH. **D)** The Michaelis-Menten constant of mAMCase catalytic domain is plotted as a function of pH. **E)** The catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) of mAMCase catalytic domain is plotted as a function of pH. **F)** Hypothetical catalytic activity modeled explained by a low pH mechanism (red), and high pH mechanism (blue) and their corresponding total activity (dashed line).

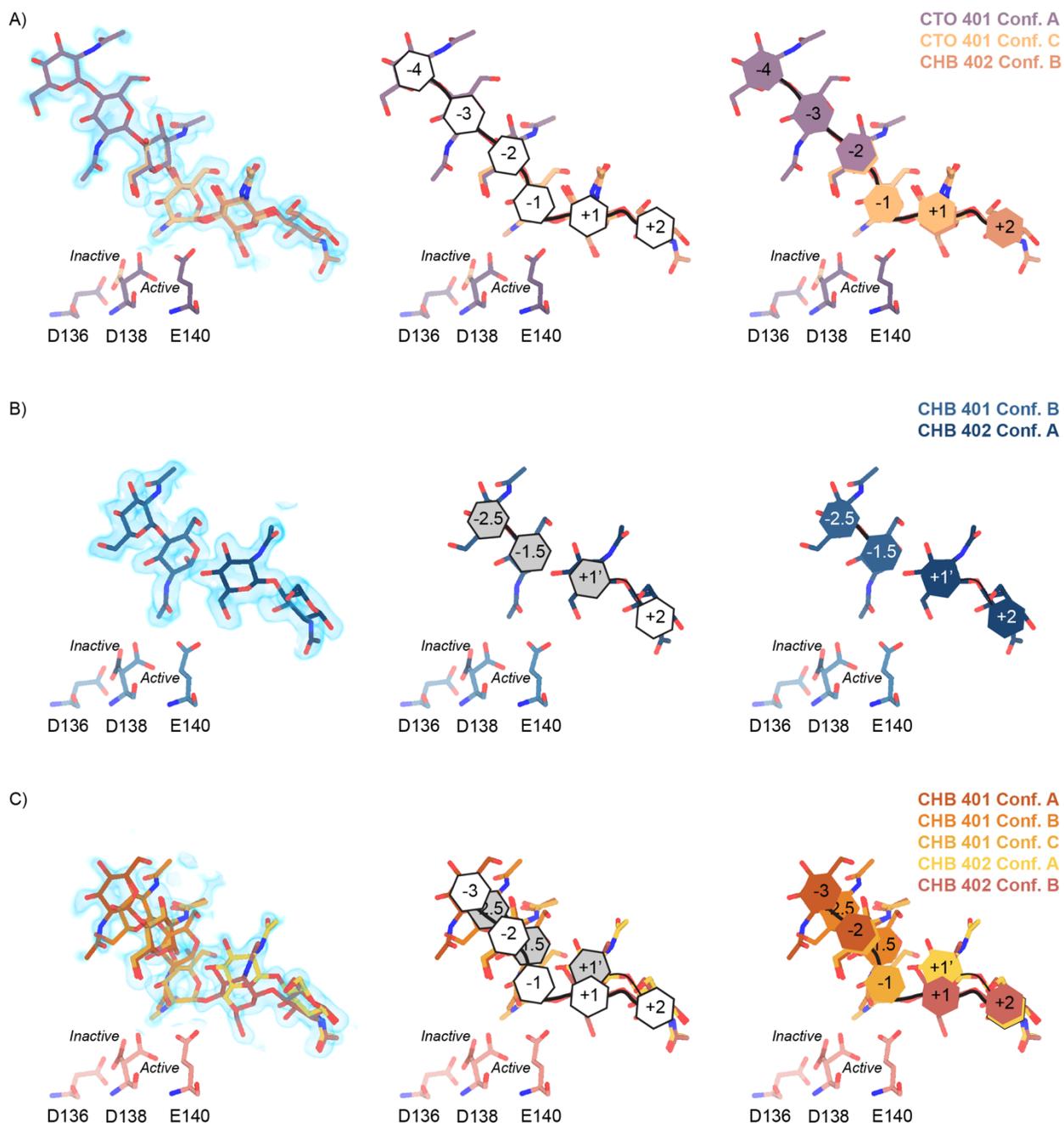


Figure 3.2 | Schematic representation of sugar-binding subsites in mAMCase.

A) PDB ID: 8GCA, chain A. Stick representation of all GlcNAc₂ sugar-binding events observed in n sugar-binding subsites with 2mFo-DFc map shown as a 1.2 Å contour (blue), the subsite nomenclature, and a schematic of alternative conformation ligand modeling. **B)** PDB ID: 8FRA, chain D. Stick representation of all GlcNAc _{n} binding events observed in $n+0.5$ sugar-binding subsites with 2mFo-DFc map shown as a 1.2 Å contour (blue), the subsite nomenclature, and a schematic of alternative conformation ligand modeling. **C)** PDB ID: 8FR9, chain B. Stick representation of all GlcNAc _{n} binding events observed in n and $n+0.5$ sugar-binding subsites with 2mFo-DFc map shown as a 1.2 Å contour (blue), the subsite nomenclature, and a schematic of alternative conformation ligand modeling.

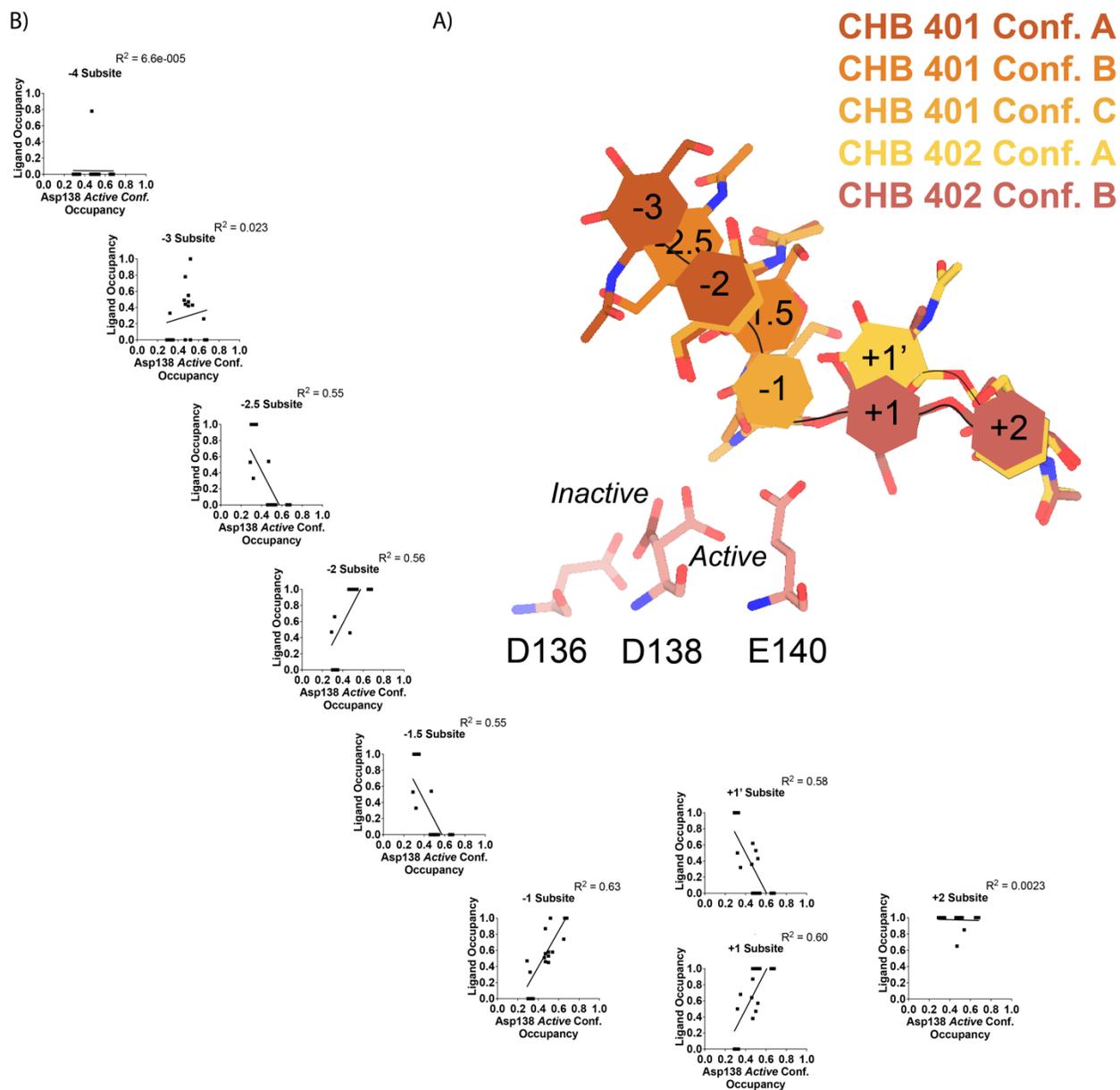


Figure 3.3 | Asp138 orientation correlates with ligand subsite occupancy.

A) PDB ID: 8FR9, chain B. Schematic of the alternative conformation ligand modeling. **B)** Linear correlation between sugar-binding subsite occupancy and Asp138 *active* conformation occupancy.

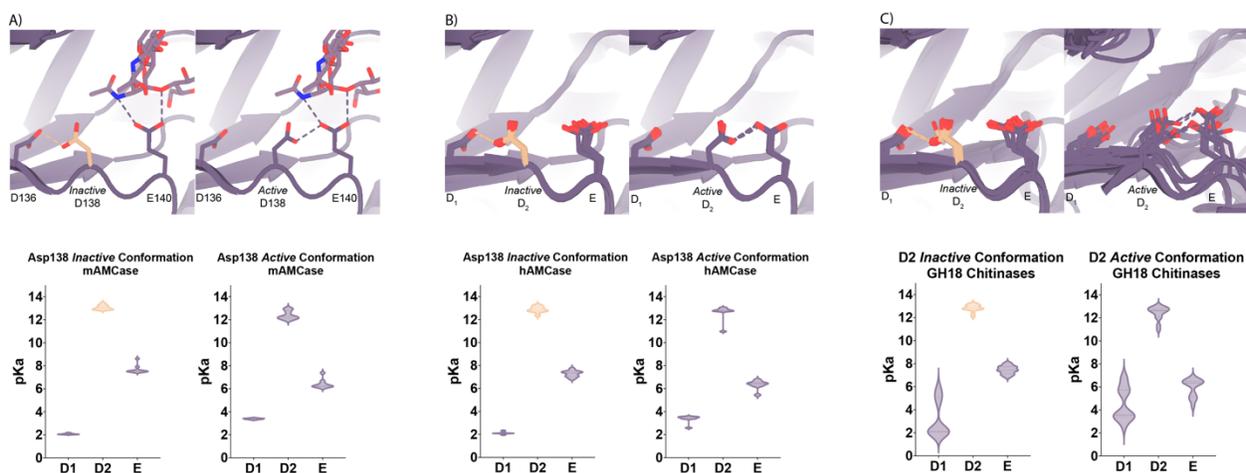


Figure 3.4 | pKa of GH18 chitinases in the D2 *inactive* and *active* conformation.

A) PDB ID: 8GCA, chain A. Distribution of pKa across Asp136, Asp138, Glu140 of mAMCase structures in either Asp138 *inactive* or Asp138 *active* conformation. **B)** PDB ID: 3FXE, 3RM4, 3RM8, 3RME (*inactive conformation*); 2YBU, 3FY1 (*active conformation*). Distribution of pKa across Asp136, Asp138, Glu140 of hAMCase structures in either Asp138 *inactive* or Asp138 *active* conformation. **C)** PDB ID: 3ALF, 3AQU, 3FXE, 3RM4, 3RM8, 3RME (*inactive conformation*); 2UY2, 2UY3, 2YBU, 4HME, 4MNJ, 4R5E, 4TXE (*active conformation*). Distribution of pKa across the catalytic triad D₁xD₂xE of GH18 chitinases in either D₂ *inactive* or *active* conformation.

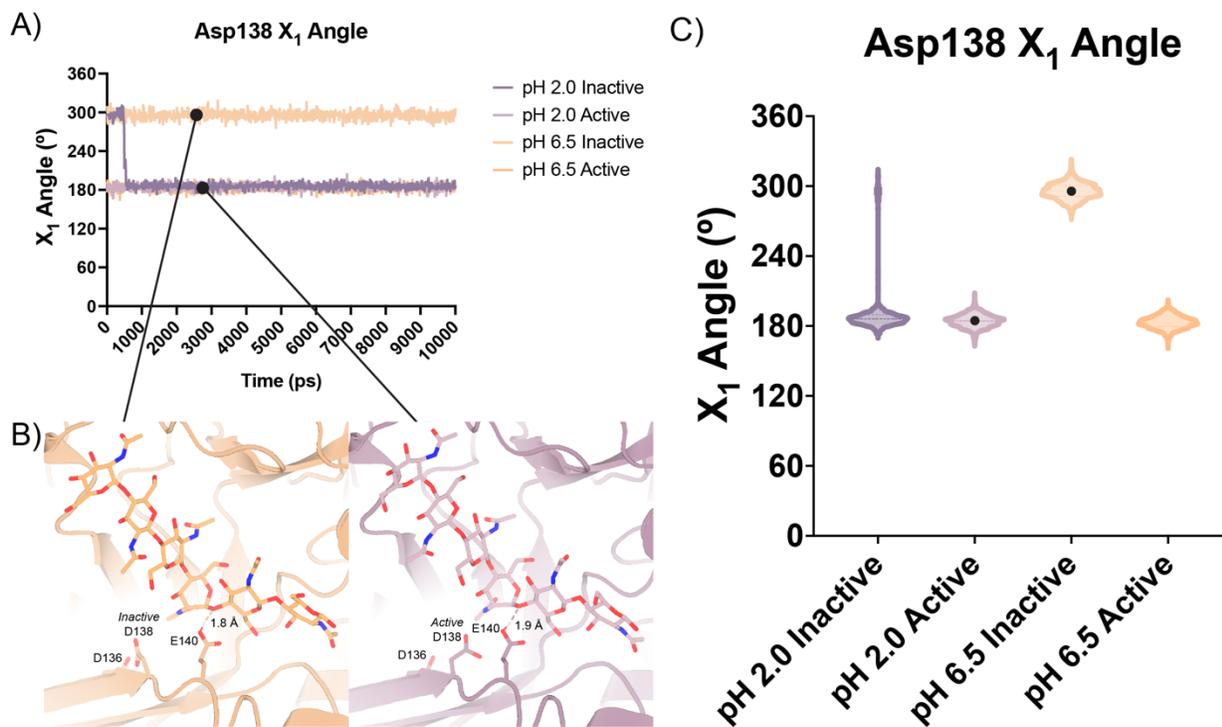
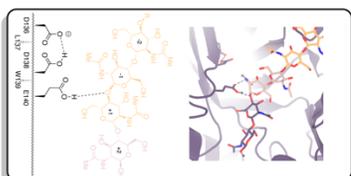
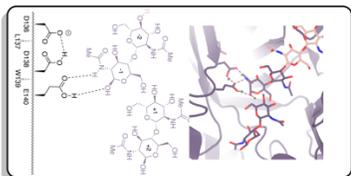
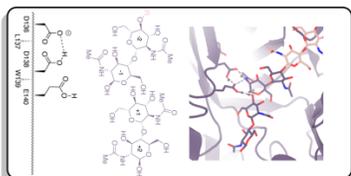
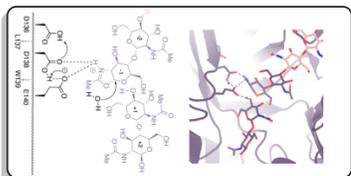
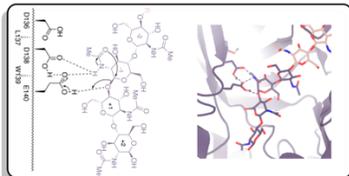
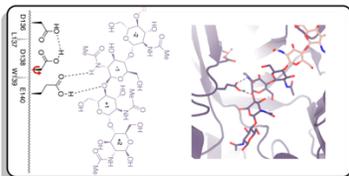
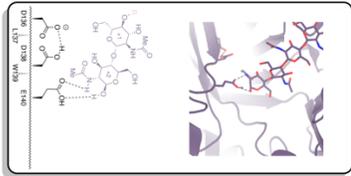
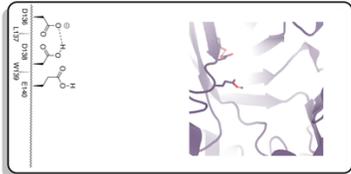


Figure 3.5 | Distribution of distances observed every 10 ps of each simulation and their respective time courses.

A) Asp138 χ_1 angles over a 10 ns simulation. **B)** Representative minimum distance snapshots of structure during pH 6.5 *inactive* simulation (left), and pH 2.0 *active* simulation (right). **C)** Distribution of Asp138 χ_1 angles over a 10 ns simulation.

pH 6.5



1
Prepriming

2
Ligand binding

3
Ligand translocation

4
D138 rotates into active conformation

5
Oxazolinium intermediate formed through intramolecular mechanism

6
Oxazolinium is hydrolyzed by water

7
GlcNAc₂ dissociates

8
Chitin chain slides forward

pH 2.0

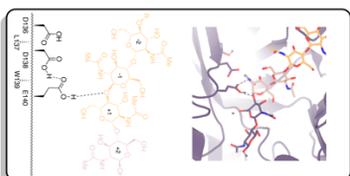
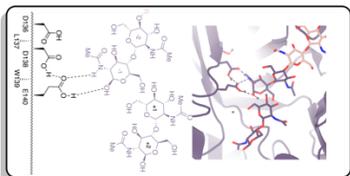
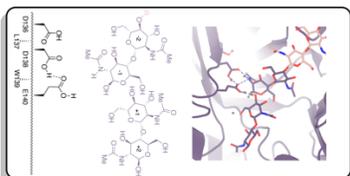
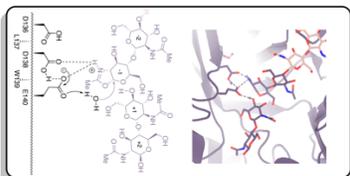
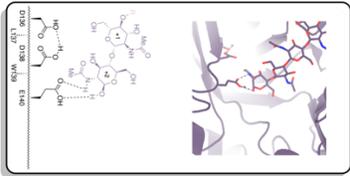
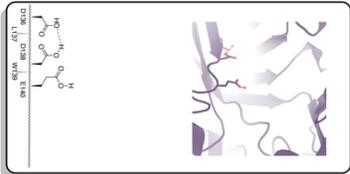


Figure 3.6 | Proposed model for ligand translocation towards the active site and ligand release post-catalysis.

A) PDB ID: 8GCA, chain A with no ligand (step 1); with GlcNAc₄ generated by *phenix.elbow* using PubChem ID: 10985690 (step 2); with GlcNAc₆ generated by *phenix.elbow* using PubChem ID: 6918014 (step 3-4, 8); with oxazolinium intermediate generated by *phenix.elbow* using PubChem ID: 25260046 (steps 5.1-5.2); with GlcNAc₂ and GlcNAc₄ generated by *phenix.elbow* using PubChem ID: 439544 and 10985690, respectively (steps 6-7). Chemical representation of GH18 catalytic cycle with corresponding molecular models of each step. Catalytic residues Asp136, Asp138, Glu140, and ligands are shown as sticks. Protons are shown as gray spheres. **B)** PDB ID: 8GCA, chain A. Animated movie of the mAMCase catalytic cycle at pH 2.0 (separate file) and **C)** at pH 6.5 (separate file). Catalytic residues Asp136, Asp138, Glu140, and ligands are shown as sticks. Protons are shown as gray spheres.

Data Availability

Raw experimental data, processing files, log files, GraphPad PRISM files, PyMOL scripts, and PyMOL sessions can be found on Zenodo or Protein Diffraction.

- **Figure 3.1** | Kinetic properties of mAMCase catalytic domain at various pH.
 - doi: 10.5281/zenodo.7968293
- **Figure 3.2** | Schematic representation of sugar-binding subsites in mAMCase.
 - doi: 10.5281/zenodo.7967930
- **Figure 3.3** | Asp138 orientation correlates with ligand subsite occupancy.
 - doi: 10.5281/zenodo.7905828
- **Figure 3.4** | pKa of GH18 chitinases in the D2 *inactive* and *active* conformation.
 - doi: 10.5281/zenodo.7905863
- **Figure 3.5** | Distribution of distances observed every 10 ps of each simulation and their respective time courses.
 - doi: 10.5281/zenodo.7758821
- **Figure 3.6** | Proposed model for ligand translocation towards the active site and ligand release post-catalysis.
 - doi: 10.5281/zenodo.7967958
- **Supplemental Figure 3.1** | pH of reaction solution before and after quenching with 0.1 M Gly-NaOH pH 10.7.
 - doi: 10.5281/zenodo.7968293
- **Supplemental Figure 3.2** | Kinetics of 4MU-chitobioside catalysis by mAMCase catalytic domain at various pH.
 - doi: 10.5281/zenodo.7968293

- **Supplemental Figure 3.3** | 96-well plate layout of crystallization conditions.
 - doi: 10.5281/zenodo.7905944
 - doi: 10.18430/M38FG5
 - doi: 10.18430/M38FG7
 - doi: 10.18430/M38GCA
 - doi: 10.18430/M38FRC
 - doi: 10.18430/M38FR9
 - doi: 10.18430/M38FRB
 - doi: 10.18430/M38FRD
 - doi: 10.18430/M38FRG
 - doi: 10.18430/M38FRA
- **Supplemental Figure 3.4** | pKa of apo and holo mAMCase in the D2 *inactive* and *active* conformation.
 - doi: 10.5281/zenodo.7905863
- **Supplemental Figure 3.5** | Overview of key residues for mAMCase activity.
 - doi: 10.5281/zenodo.7967978
- **Supplemental Figure 3.6** | Protein-ligand interactions between mAMCase and chitin.
 - doi: 10.5281/zenodo.7967954
- **Supplemental Figure 3.7** | Ringer analysis of catalytic triad confirms alternative Asp138 conformations.
 - doi: 10.5281/zenodo.7758815

Methods

Protein expression and purification

Protein expression and purification mAMCase catalytic domain (UniProt: Q91XA9; residues 22 to 391) was cloned into a pTwist CMV [pmRED006; Twist Biosciences; Addgene ID: 200228] or pcDNA3.1(+) [pmRED013; Genscript; Addgene ID: 200229] expression vector with a C-terminal 6xHis tag. To express mAMCase catalytic domain, 0.8-1 µg/mL plasmid DNA was transfected into ExpiCHO-S cells (ThermoFisher Scientific #A29127) using the Max Titer protocol (ThermoFisher Scientific MAN0014337). After cells were grown shaking at 37°C with 8% CO₂ for 18-22 hours, ExpiFectamine CHO Enhancer (ThermoFisher Scientific #A29129) and ExpiCHO feed (ThermoFisher Scientific #A29129) was added to the flask. Cells were then transferred to 32 °C with 5% CO₂ for an additional 9-13 days of growth, with a second volume of ExpiCHO feed added to the flask on day 5 post-transfection. Cells were removed by centrifugation at 4,000 RCF for 15 minutes at 4 °C, and the remaining supernatant was filtered using a 0.22 µm filter at 4 °C. Filtered supernatant was either dialyzed into Ni–nitrilotriacetic acid (NTA) loading buffer [100 mM Tris-HCl (pH 8.0), 150 mM NaCl] at 4 °C in a 10-kDa molecular weight cutoff (MWCO) Slide-A-Lyzer Dialysis Cassette, (ThermoFisher Scientific #66810) for 18-24 hours or concentrated in a 10-kDa MWCO centrifugal concentrator (Amicon #UFC901008) at 4,000 RCF in 5 min intervals until the final volume was equal to 10 mL, which was then diluted 1:10 with loading buffer for a total volume of 100 mL. The dialyzed supernatant volume was filtered using a 0.22 µm filter at 4 °C. All purification steps were performed at 4°C using an ÄKTA fast protein liquid chromatography system (Cytiva). The dialyzed supernatant was applied to a 5-ml HisTrap FF column (Cytiva, 17525501). The column was washed with 40 mL of loading buffer followed by 25 mL of 10% Ni-NTA elution buffer [100 mM Tris-HCl (pH

8.0), 150 mM NaCl, 500 mM imidazole] and then eluted over a 50 mL gradient from 10% to 100% elution buffer. Eluted protein was concentrated to 2.5 mL using a 10-kDa MWCO centrifugal concentrator (Amicon, UFC901024). The sample was further purified by size exclusion chromatography (SEC) using a HiLoad 16/600 Superdex 75 pg column (Cytiva, 28989333) equilibrated with SEC buffer [25 mM Tris-HCl (pH 8.0), 50 mM NaCl]. Eluted fractions were collected and stored at 4 °C for further use.

4MU-chitobioside Endpoint Assay

Chitinase catalytic activity has previously been assayed using 4-methylumbelliferyl chitobioside (4MU-CB; Sigma-Aldrich M9763) ^{43,44}. 100 nM chitinase enzyme was incubated with varying concentrations of 4MU-chitobioside up to 117 μM in McIlvaine Buffer at 37 °C ²¹. The 4-methylumbelliferone (4MU) fluorophore is quenched by a β-glycosidic linkage to a short chitin oligomer, which is cleaved by a chitinase enzyme, which generates fluorescence with peak excitation at 360 nm and emission at 450 nm. 4MU fluorescence is pH-dependent with peak excitation at 360 nm and emission at 450 nm at pH 7.0. It has been previously reported that 4MU peak excitation/emission increases and fluorescence intensity decreases as pH becomes more acidic ⁴⁵. Given the pH-dependent fluorescence properties of the 4MU fluorophore, we incubate the reaction at different pH, then quench with 0.1 M Gly-NaOH pH 10.7. Quenching the reaction with 0.1 M Gly-NaOH pH 10.7 stops the enzyme reaction and shifts the pH to maximize the quantum yield of the 4MU substrate.

A Tecan Spark multimode microplate reader is pre-heated to 37 °C. 4MU-chitobioside (Sigma-Aldrich M9763) and AMCase are separately pre-incubated at 37 °C for 15 minutes. 25 μL of 4MU-chitobioside or McIlvaine Buffer (Boston Bioproducts) is transferred into each well

in a Multiplate 96-Well PCR Plate, high profile, unskirted, clear (Bio-Rad MLP9601). Using a Multidrop Combi Reagent Dispenser (Thermo Scientific #5840300), 25 μ L of either 100 nM AMCase or McIlvaine Buffer (Boston Bioproducts) is dispensed into each well in the Multiplate 96-Well PCR Plate (Corning #3993). The Multiplate 96-Well PCR Plate is then incubated at 37 $^{\circ}$ C in a 96-well Non-Skirted PCR Plate Block (Thermo Scientific #88870120) in a digital dry bath (Thermo Scientific #88870006).

The reaction is quenched with 50 μ L 0.1 M Gly-NaOH pH 10.7 at timepoints 0", 15", 30", 45", 60", 90". 40 μ L of the quenched reaction is transferred to a 384-well Low Volume Black Flat Bottom Polystyrene NBS Microplate (Corning #3820), then immediately read using the following parameters:

- Excitation - 360 nm, 20 nm bandwidth
- Emission - 450 nm, 20 nm bandwidth
- Gain - 50
- Flashes - 20

This assay was performed in quadruplicate for each pH unit reported. This allowed us to reliably measure initial rates of catalysis across a large range of pH conditions. The workflow for this assay is illustrated in **Figure 3.1**.

Analysis of kinetic data

25 μ L of 200 μ M 4MU fluorophore (Sigma-Aldrich M1381) was serially diluted into 25 μ L McIlvaine Buffer (Boston Bioproducts) across the range of pHs to obtain 5 diluted ligand concentrations ranging from 100 μ M to 6.25 μ M as well as ligand free. This dilution series was performed in duplicate per 96-Well PCR plate for a total of 8 replicates per ligand concentration

at each given pH value. At the end of the experiment, the 4MU dilution series is quenched with 50 μ L 0.1 M Gly-NaOH pH 10.7 for a final dilution series ranging from 50 μ M to 3.125 μ M.

Relative fluorescence (RFU) was plotted against 4MU concentration, then a simple linear regression with the constraint $Y = 0$ when $X = 0$ was performed to obtain a standard curve. We then used the equation $Y = mX + b$, where m is the slope from the standard curve and Y is the RFU from a given experimental data point, to determine the concentration of 4MU [μ M] generated by AMCase at a given time point.

Average 4MU concentration [μ M] ($n = 4$) was plotted as a function of time with error bars representing the standard deviation. We then fit a simple linear regression with the constraint $Y = 0$ when $X = 0$ to obtain the initial rate of enzyme activity (4MU [μ M]/sec) at each concentration of 4MU-chitobioside [μ M]. Average initial rate ($n = 4$) was plotted as a function of 4MU-chitobioside concentration [μ M] with error bars representing the standard deviation. We fit our data to a Michaelis-Menten function without substrate inhibition to obtain V_{\max} and K_M parameters. We used the equation $k_{\text{cat}} = V_{\max}/[\text{Enzyme}]$ where $[\text{Enzyme}] = 0.1 \mu\text{M}$ to calculate k_{cat} . We calculated catalytic efficiency (CE) using the equation $\text{CE} = K_M/k_{\text{cat}}$. Kinetic parameters V_{\max} , K_M , k_{cat} , and catalytic efficiency were plotted as a function of pH.

Apo crystallization

Using hanging-drop vapor diffusion, crystallization screens were performed using a 96-well Clear Flat Bottom Polystyrene High Binding microplate (Corning CLS9018BC) with 0.5 mL of reservoir solution in each well. Crystallization drops were set up on 96-well plate seals (SPT Labtech 4150-05100) with 0.2 μ L of AMCase at 11 mg/ml and 0.2 μ L of reservoir using an SPT Labtech mosquito crystal. After 21 days at 20 $^{\circ}$ C, we observed crystals in a reservoir

solution containing 20% PEG-6000 , 0.1 M Sodium Acetate pH 5.0, and 0.2 M Magnesium Chloride (II) (MgCl₂) (NeXtal PACT Suite Well A10; #130718).

Apo data collection, processing, and refinement at cryogenic temperature

Diffraction data were collected at the beamline ALS 8.3.1 at 100 K. Diffraction data from multiple crystals were merged using xia2⁴⁶, implementing DIALS⁴⁷ for indexing and integration, and Aimless⁴⁸ for scaling and merging. We confirmed the space group assignment using DIMPLe⁴⁹. We calculated phases by the method of molecular replacement, using the program Phaser⁵⁰ and a previous structure of hAMCase (PDB: 3FXY) as the search model. The model was manually adjusted in Coot to fit the electron density map calculated from molecular replacement, followed by automated refinement of coordinates, atomic displacement parameters, and occupancies using phenix.refine⁵¹ with optimization of restraint weights. Default refinement parameters were used, except the fact that five refinement macrocycles were carried out per iteration and water molecules were automatically added to peaks in the 2mFo-DFc electron density map higher than 3.5 Å. The minimum model-water distance was set to 1.8 Å, and a maximum model-water distance was set to 6 Å. For later rounds of refinement, hydrogens were added to riding positions using *phenix.ready_set*, and B-factors were refined anisotropically for non-hydrogen and non-water atoms. Following two initial rounds of iterative model building and refinement using the aforementioned strategy, we began introducing additional parameters into the model, enabled by the extraordinarily high resolution of our diffraction data. First, we implemented anisotropic atomic displacement parameters for heavy atoms (C, N, O, and S), followed by refinement of explicit hydrogen atom positions. A final round of refinement was performed without updating water molecules.

Apo data collection, processing, and refinement at room temperature

Diffraction data were collected at the beamline ALS 8.3.1 at 277 K. Data collection, processing, refinement, and model building were performed as described previously for the apo crystals at cryogenic temperature.

Holo crystallization

Initially, crystals were grown by hanging-drop vapor diffusion with a reservoir solution containing 20% PEG-6000 (Hampton Research HR2533), 0.1 M Sodium Acetate (pH 3.6, Hampton Research HR293301; pH 4.1, Hampton Research HR293306; pH 5.0, Hampton Research HR293315; pH 5.6, Hampton Research HR293321), and 0.2 M Magnesium Chloride (II) (MgCl₂) (Hampton Research HR2559). Screens were performed using a 96-well Clear Flat Bottom Polystyrene High Binding microplate (Corning CLS9018BC) with 0.5 mL of reservoir solution in each well. Crystallization drops were set up on 96-well plate seals (SPT Labtech 4150-05100) with 0.2 μL of AMCcase at 11 mg/ml and 0.2 μL of reservoir using an SPT Labtech mosquito crystal. Crystals grew after 1-2 days at 20 °C.

Using hanging drop diffusion vapor, holo crystals grew after 12 hours at 20 °C. For the holo form with GlcNAc₂ (Megazyme O-CHI2), this construct crystallized in either P2₁2₁2 or P2₁2₁2₁ with either 2 or 4 molecules in the ASU and diffracted to a maximum resolution between 1.50 to 1.95 Å. For the holo form with GlcNAc₃ (Megazyme O-CHI3), this construct crystallized in P2₁2₁2 with 2 molecules in the ASU and diffracted to a maximum resolution of 1.70 Å.

Holo data collection, processing, and refinement at cryogenic temperature

Diffraction data were collected at the beamline ALS 8.3.1 and SSRL beamline 12-1 at 100 K. Data collection, processing, refinement, and model building were performed as described previously for the apo crystals.

Ligands were modeled into 2mFo-DFc maps with Coot, using restraints generated by *phenix.elbow* from an isomeric SMILES (simplified molecular input line-entry system) string⁵² using AM1 geometry optimization. Default refinement parameters were used, except the fact that five refinement macrocycles were carried out per iteration and water molecules were automatically added to peaks in the 2mFo-DFc electron density map higher than 3.5 Å. The minimum model-water distance was set to 1.8 Å, and a maximum model-water distance was set to 6 Å. Changes in protein conformation and solvation were also modeled. Hydrogens were added with *phenix.ready_set*, and waters were updated automatically. A final round of refinement was performed without updating water molecules⁵³.

Ligand modeling

For consistency, ligands were assigned an alternative conformation ID based on the sugar-binding subsites it occupied:

GlcNAc₂ ResID 401 Conf. A, -3 to -2

GlcNAc₂ ResID 401 Conf. B, -2.5 to -1.5

GlcNAc₂ ResID 401 Conf. C, -2 to -1

GlcNAc₂ ResID 402 Conf. D, -1 to +1

GlcNAc₂ ResID 402 Conf. B, +1 to +2

GlcNAc₂ ResID 402 Conf. A, +1' to +2

GlcNAc₃ ResID 401 Conf. A, -4 to -2

GlcNAc₃ ResID 401 Conf. B, -3 to -1

GlcNAc₃ ResID 401 Conf. C, -2 to +1

GlcNAc₃ ResID 402 Conf. B, -1 to +2

Ligand occupancies and B-factors using *phenix.refine*. Ligands with occupancies ≤ 0.10 were removed from the model.

Ringer analysis

Individual residues in each of the mAMCase structures were run through Ringer using *mmtbx.ringer*. Outputs from the csv file were then plotted using Matplotlib.

pKa Analysis

We used the APBS-PDB2PQR software suite (<https://server.poissonboltzmann.org/pdb2pqr>)⁵⁴. Each PDB model was separated into two separate models containing a single Asp138 conformation in either the *inactive* (down towards Asp136) or *active conformation* (up towards Glu140). Solvent and ligand molecules were not modified. The pH of the crystallization condition was provided for PROPKA to assign protonation states. The default forcefield PARSE was used. The following additional options were selected: Ensure that new atoms are not rebuilt too close to existing atoms; Optimize the hydrogen bonding network.

Molecular Dynamics

Simulations were performed using hexaacetyl-chitohexaose (PubChem Compound ID: 6918014) modeled into 8GCA with Asp138 in either the *inactive* (down towards Asp136) or *active conformation* (up towards Glu140). The model PDB file was opened in MOE and solvated in a sphere of water 10 Å away from the protein. This system then underwent structural preparation for simulations using the standard parameters with the AMBER14 forcefield. The system then was protonated to set pH {2.0, 6.5} based on side chain pKa predictions using the 3DProtonate menu followed by confirmation of appropriate protonation by PROPKA calculations. Protonated models underwent energy minimization by steepest descent before simulations were set up. Equilibration was performed for 10 ps followed by 100 ps of thermal gradient equilibration from 0K to 300K. A thermal bath equilibration was run for 100 ps before the production runs were started. Productions were run for 10 ns with a time step of 0.5 fs to not overshoot bond vibrations. The simulation was sampled every 10 ps for subsequent data analysis which was performed using the MOE database viewer and replotted using GraphPad Prism.

References

1. Cabib, E. & Bowers, B. Timing and function of chitin synthesis in yeast. *J. Bacteriol.* **124**, 1586–1593 (1975).
2. Zhu, K. Y., Merzendorfer, H., Zhang, W., Zhang, J. & Muthukrishnan, S. Biosynthesis, turnover, and functions of chitin in insects. *Annu. Rev. Entomol.* **61**, 177–196 (2016).
3. Tang, W. J., Fernandez, J., Sohn, J. J. & Amemiya, C. T. Chitin is endogenously produced in vertebrates. *Curr. Biol.* **25**, 897–900 (2015).
4. Elieh Ali Komi, D., Sharma, L. & Dela Cruz, C. S. Chitin and Its Effects on Inflammatory and Immune Responses. *Clin. Rev. Allergy Immunol.* **54**, 213–223 (2018).
5. Van Dyken, S. J. *et al.* Spontaneous Chitin Accumulation in Airways and Age-Related Fibrotic Lung Disease. *Cell* **169**, 497-509.e13 (2017).
6. Van Dyken, S. J. & Locksley, R. M. Chitins and chitinase activity in airway diseases. *J. Allergy Clin. Immunol.* **142**, 364–369 (2018).
7. Bussink, A. P., Speijer, D., Aerts, J. M. F. G. & Boot, R. G. Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases. *Genetics* **177**, 959–970 (2007).
8. Zhu, Z. *et al.* Acidic mammalian chitinase in asthmatic Th2 inflammation and IL-13 pathway activation. *Science* **304**, 1678–1682 (2004).
9. Reese, T. A. *et al.* Chitin induces accumulation in tissue of innate immune cells associated with allergy. *Nature* **447**, 92–96 (2007).
10. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859 (1995).

11. Papanikolaou, Y., Tavlas, G., Vorgias, C. E. & Petratos, K. De novo purification scheme and crystallization conditions yield high-resolution structures of chitinase A and its complex with the inhibitor allosamidin. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 400–403 (2003).
12. Cole, D. C. *et al.* Identification and characterization of acidic mammalian chitinase inhibitors. *J. Med. Chem.* **53**, 6122–6128 (2010).
13. Olland, A. M. *et al.* Triad of polar residues implicated in pH specificity of acidic mammalian chitinase. *Protein Sci.* **18**, 569–578 (2009).
14. Boot, R. G. *et al.* Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *J. Biol. Chem.* **276**, 6770–6778 (2001).
15. Yoong, P., Schuch, R., Nelson, D. & Fischetti, V. A. PlyPH, a bacteriolytic enzyme with a broad pH range of activity and lytic action against *Bacillus anthracis*. *J. Bacteriol.* **188**, 2711–2714 (2006).
16. Sajedi, R. H. *et al.* A Ca-independent α -amylase that is active and stable at low pH from the *Bacillus* sp. KR-8104. *Enzyme Microb. Technol.* **36**, 666–671 (2005).
17. Bhunia, B., Dutta, D. & Chaudhuri, S. Extracellular alkaline protease from *Bacillus licheniformis* NCIM-2042: Improving enzyme activity assay and characterization. *Eng. Life Sci.* **11**, 207–215 (2011).
18. Seibold, M. A. *et al.* Differential enzymatic activity of common haplotypic versions of the human acidic Mammalian chitinase protein. *J. Biol. Chem.* **284**, 19650–19658 (2009).
19. Chou, Y.-T. *et al.* Kinetic characterization of recombinant human acidic mammalian chitinase. *Biochemistry* **45**, 4444–4454 (2006).
20. Kashimura, A. *et al.* Functional properties of the catalytic domain of mouse acidic mammalian chitinase expressed in *Escherichia coli*. *Int. J. Mol. Sci.* **16**, 4028–4042 (2015).

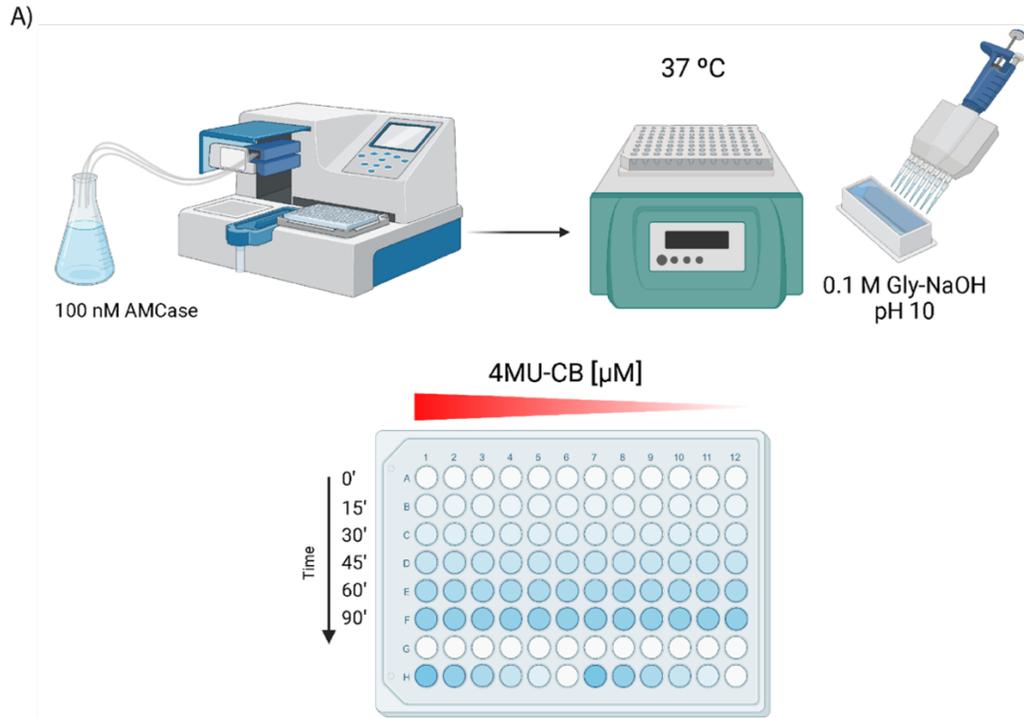
21. Barad, B. A. *et al.* Differences in the chitinolytic activity of mammalian chitinases on soluble and insoluble substrates. *Protein Sci.* **29**, 966–977 (2020).
22. Aguilera, B. *et al.* Transglycosidase activity of chitotriosidase: improved enzymatic assay for the human macrophage chitinase. *J. Biol. Chem.* **278**, 40911–40916 (2003).
23. Wakita, S. *et al.* Mouse acidic mammalian chitinase exhibits transglycosylation activity at somatic tissue pH. *FEBS Lett.* **591**, 3310–3318 (2017).
24. Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J* **321 (Pt 2)**, 557–559 (1997).
25. Watanabe, T. *et al.* Aromatic residues within the substrate-binding cleft of *Bacillus circulans* chitinase A1 are essential for hydrolysis of crystalline chitin. *Biochem. J* **376**, 237–244 (2003).
26. Horn, S. J. *et al.* Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18089–18094 (2006).
27. Zakariassen, H. *et al.* Aromatic residues in the catalytic center of chitinase A from *Serratia marcescens* affect processivity, enzyme activity, and biomass converting efficiency. *J. Biol. Chem.* **284**, 10610–10617 (2009).
28. Jiménez-Ortega, E., Kidibule, P. E., Fernández-Lobato, M. & Sanz-Aparicio, J. Structural inspection and protein motions modelling of a fungal glycoside hydrolase family 18 chitinase by crystallography depicts a dynamic enzymatic mechanism. *Comput. Struct. Biotechnol. J.* **19**, 5466–5478 (2021).
29. Nakamura, A., Okazaki, K.-I., Furuta, T., Sakurai, M. & Iino, R. Processive chitinase is Brownian monorail operated by fast catalysis after peeling rail from crystalline chitin. *Nat. Commun.* **9**, 1–12 (2018).

30. van Aalten, D. M. *et al.* Structural insights into the catalytic mechanism of a family 18 exochitinase. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8979–8984 (2001).
31. Lang, P. T. *et al.* Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* **19**, 1420–1431 (2010).
32. Fusetti, F. *et al.* Structure of human chitotriosidase. Implications for specific inhibitor design and function of mammalian chitinase-like lectins. *J. Biol. Chem.* **277**, 25537–25544 (2002).
33. Songsiriritthigul, C., Pantoom, S., Aguda, A. H., Robinson, R. C. & Suginta, W. Crystal structures of *Vibrio harveyi* chitinase A complexed with chitooligosaccharides: implications for the catalytic mechanism. *J. Struct. Biol.* **162**, 491–499 (2008).
34. Synstad, B. *et al.* Mutational and computational analysis of the role of conserved residues in the active site of a family 18 chitinase. *Eur. J. Biochem.* **271**, 253–262 (2004).
35. Bussink, A. P., Vreede, J., Aerts, J. M. F. G. & Boot, R. G. A single histidine residue modulates enzymatic activity in acidic mammalian chitinase. *FEBS Lett.* **582**, 931–935 (2008).
36. Mishra, P., Patni, D. & Jha, S. K. A pH-dependent protein stability switch coupled to the perturbed pKa of a single ionizable residue. *Biophys. Chem.* **274**, 106591 (2021).
37. Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **75**, 187–205 (2009).
38. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).

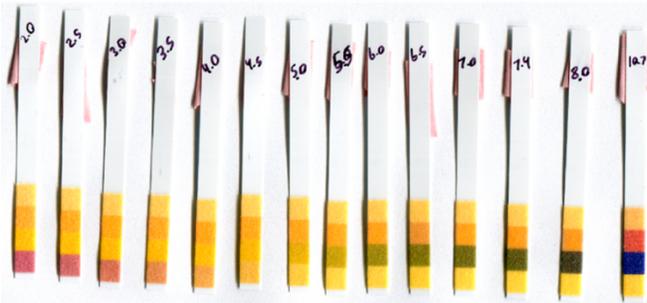
39. Meekrathok, P., Kukic, P., Nielsen, J. E. & Suginta, W. Investigation of Ionization Pattern of the Adjacent Acidic Residues in the DXDXE Motif of GH-18 Chitinases Using Theoretical pKa Calculations. *J. Chem. Inf. Model.* **57**, 572–583 (2017).
40. Iino, T., Sakurai, M. & Furuta, T. A novel ring-shaped reaction pathway with interconvertible intermediates in chitinase A as revealed by QM/MM simulation combined with a one-dimensional projection technique. *Phys. Chem. Chem. Phys.* **21**, 24956–24966 (2019).
41. Kashimura, A. *et al.* Protein A-mouse acidic mammalian chitinase-V5-His expressed in periplasmic space of Escherichia coli possesses chitinase functions comparable to CHO-expressed protein. *PLoS One* **8**, e78669 (2013).
42. Tabata, E. *et al.* Noninsect-Based Diet Leads to Structural and Functional Changes of Acidic Chitinase in Carnivora. *Mol. Biol. Evol.* **39**, (2022).
43. O'Brien, M. & Colwell, R. R. A rapid test for chitinase activity that uses 4-methylumbelliferyl-N-acetyl-beta-D-glucosaminide. *Appl. Environ. Microbiol.* **53**, 1718–1720 (1987).
44. Renkema, Boot & Muijsers. Purification and Characterization of Human Chitotriosidase, a Novel Member of the Chitinase Family of Proteins (*). *Boll. Soc. Ital. Biol. Sper.*
45. Zhi, H., Wang, J., Wang, S. & Wei, Y. Fluorescent Properties of Hymecromone and Fluorimetric Analysis of Hymecromone in Compound Dantong Capsule. *Journal of Spectroscopy* **2013**, (2013).
46. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2009).

47. Winter, G. *et al.* DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol* **74**, 85–97 (2018).
48. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 (2011).
49. Wojdyr, M., Keegan, R., Winter, G. & Ashton, A. DIMPLE-a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands. in *ACTA CRYSTALLOGRAPHICA A-FOUNDATION AND ADVANCES* vol. 69 S299–S299 (INTERNATIONAL UNION OF PURE AND APPLIED PHYSICS, INTERNATIONAL UNION OF CRYSTALLOGRAPHY 2 ABBEY SQ, CHESTER, CH1 2HU, ENGLAND, 2013).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
51. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
52. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
53. Wojdyr, M., Keegan, R., Winter, G. & Ashton, A. DIMPLE - a pipeline for the rapid generation of difference maps from protein crystals with putatively bound ligands. *Acta Crystallogr. A* **69**, 299–299 (2013).
54. Jurrus, E. *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
55. Morin, A. *et al.* Cutting edge: Collaboration gets the most out of software. *Elife* **2**, e01456 (2013).

Supplemental Figures



B) Mcllvaine Buffer



Mcllvaine Buffer + 0.1 M Gly-NaOH pH 10.7

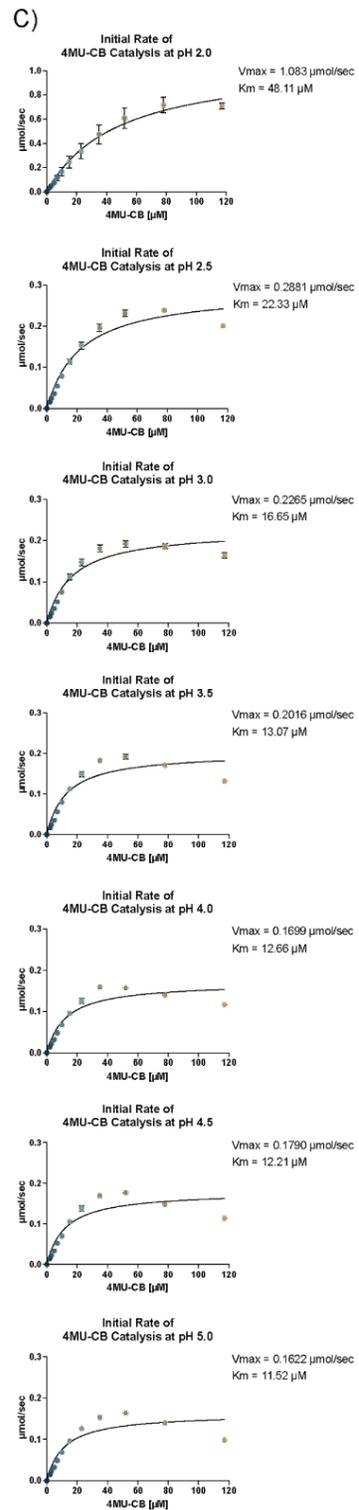
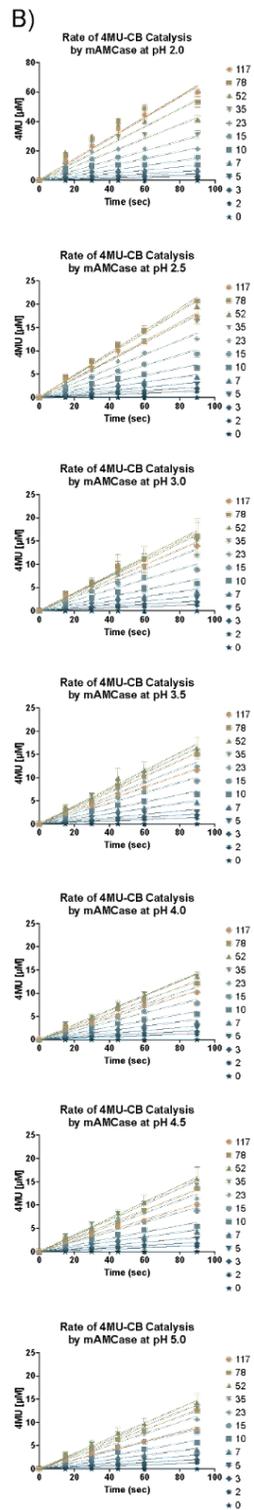
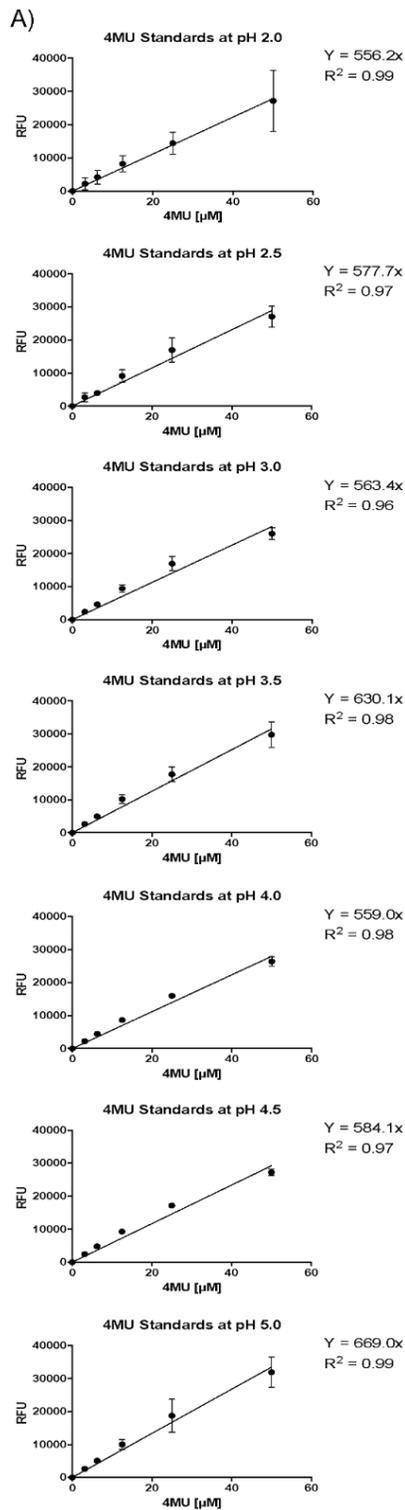


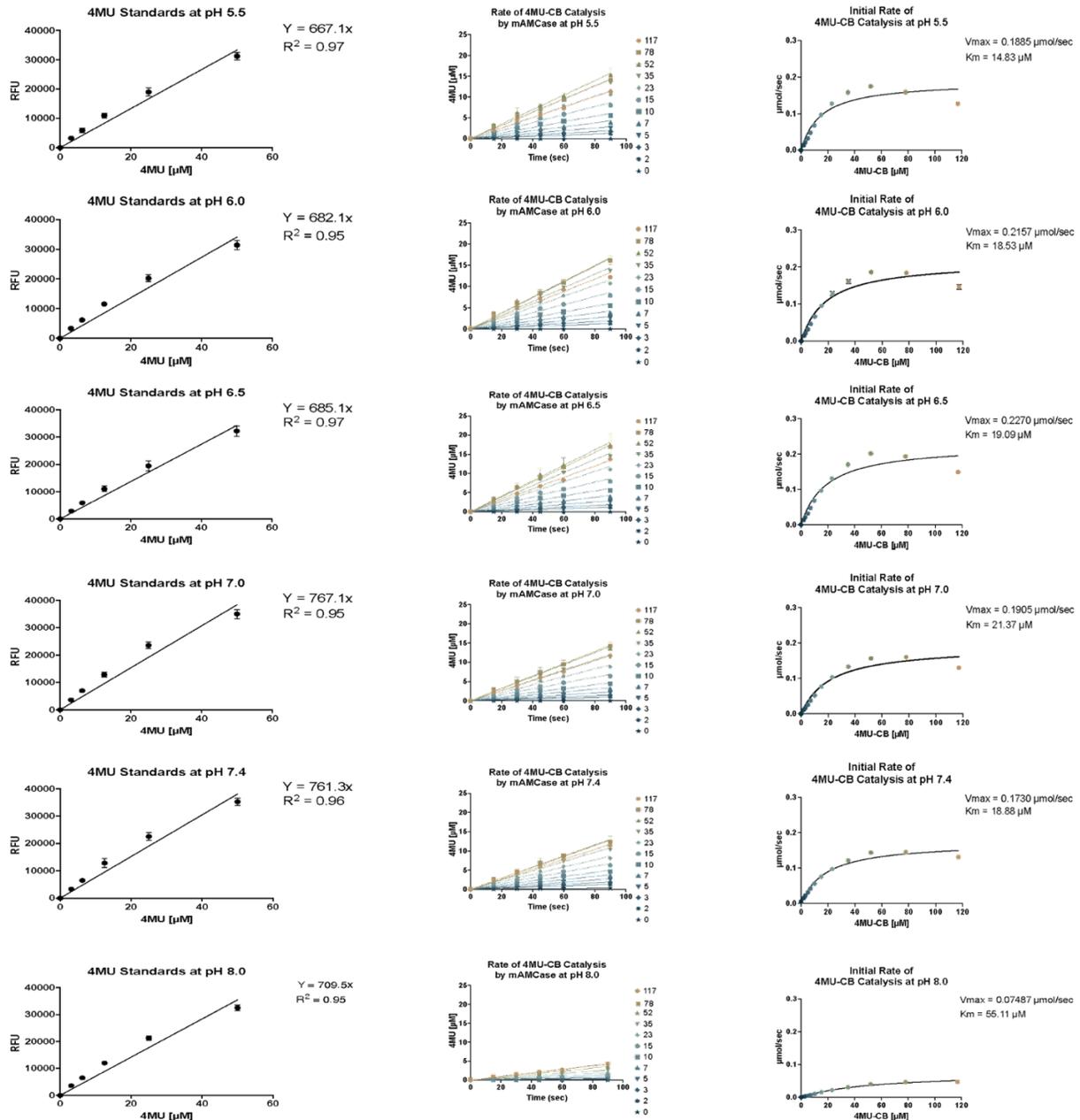
C) Reference



Supplemental Figure 3.1 | pH of reaction solution before and after quenching with 0.1 M Gly-NaOH pH 10.7.

A) Schematic of modified endpoint 4MU-chitobioside assay. **B)** Reaction pH before and after quenching with 0.1 M Gly-NaOH pH 10.7, and **C)** a pH strip reference sheet.





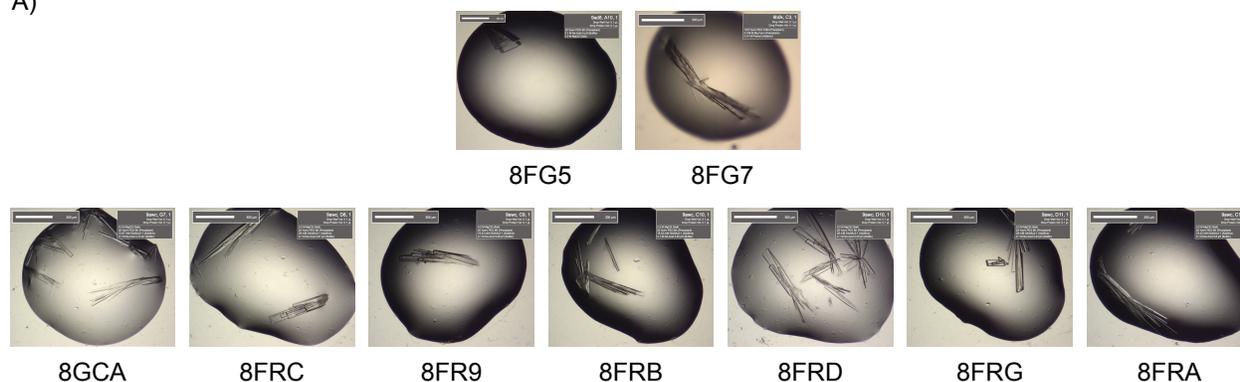
Supplemental Figure 3.2 | Kinetics of 4MU-chitobioside catalysis by mAMCase catalytic domain at various pH.

A) A linear fit forced through $Y = 0$ is used to generate the standard curve for converting RFU to 4MU [μM]. Each data point represents $n = 8$ with error bars representing the standard deviation.

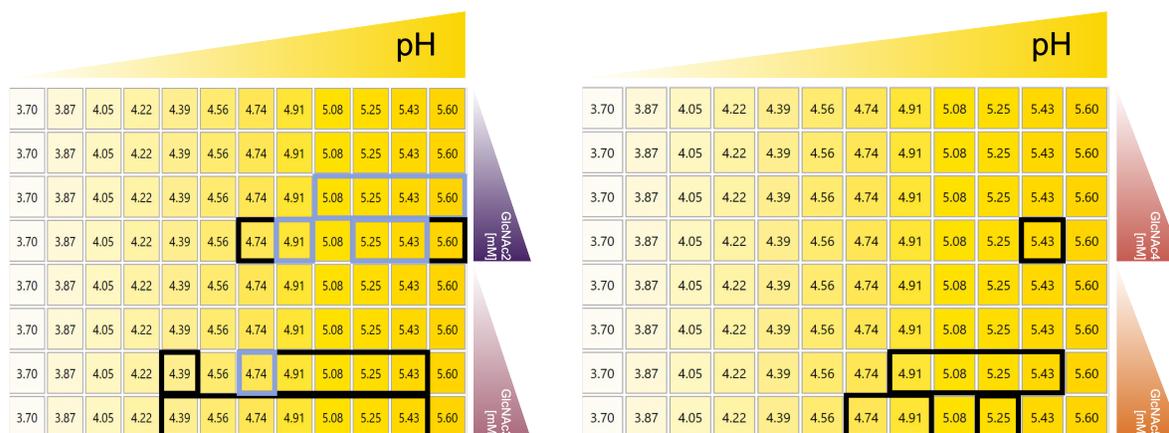
B) 4MU fluorescence (RFU) is plotted as a function of time (sec). Each data point represents $n = 4$ with error bars representing the standard deviation. A linear fit is applied to each concentration of 4MU-chitobioside to calculate an initial rate. RFU is converted to μM using a 4MU standard curve.

C) The rate of 4MU-chitobioside catalysis (1/sec) by mAMCase catalytic domain is plotted as a function of 4MU-chitobioside concentration (μM). Each data point represents $n = 4$ with error bars representing the standard deviation. Michaelis-Menten equation without substrate inhibition was used to estimate the k_{cat} and K_M from the initial rate of reaction at various substrate concentrations.

A)

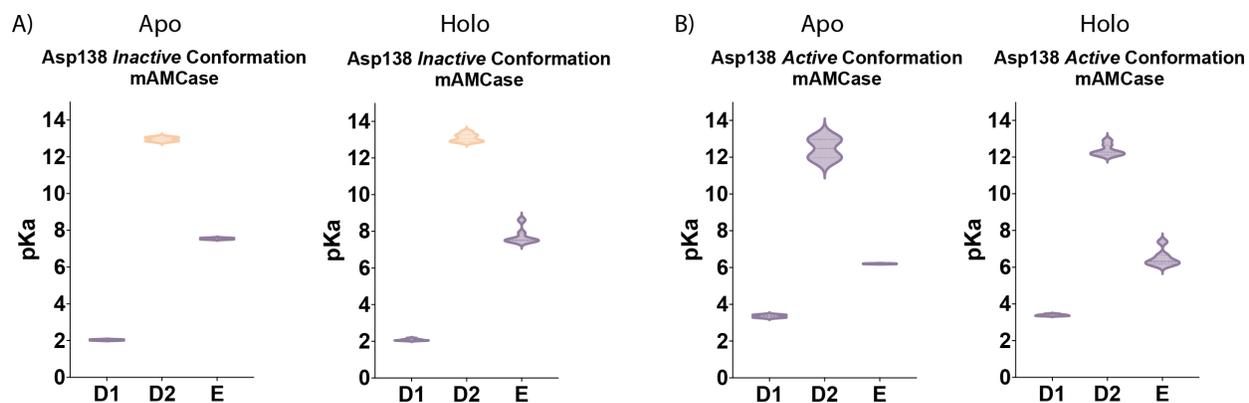


B)



Supplemental Figure 3.3 | 96-well plate layout of crystallization conditions.

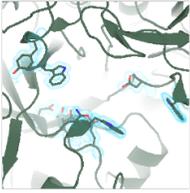
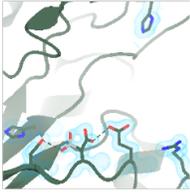
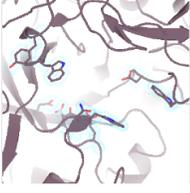
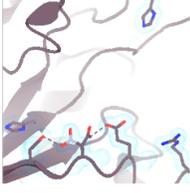
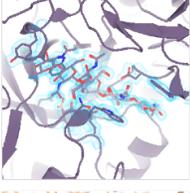
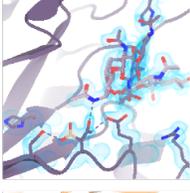
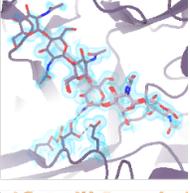
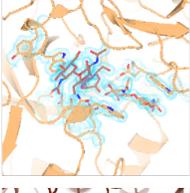
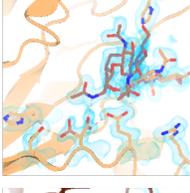
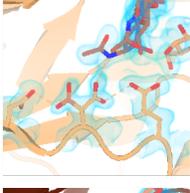
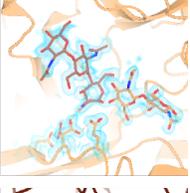
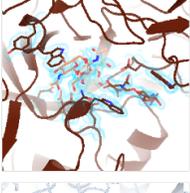
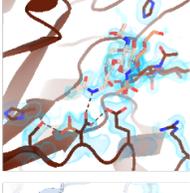
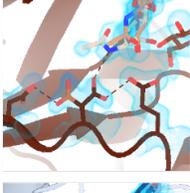
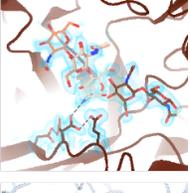
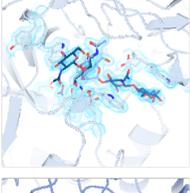
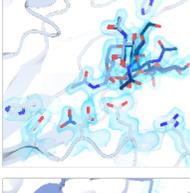
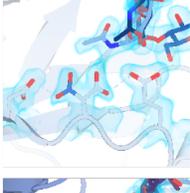
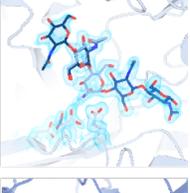
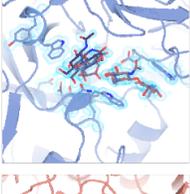
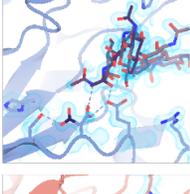
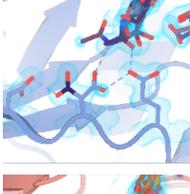
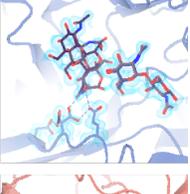
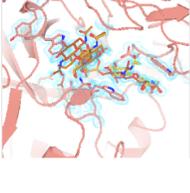
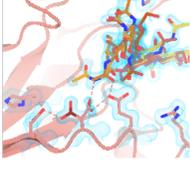
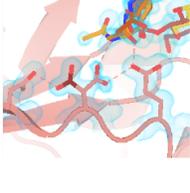
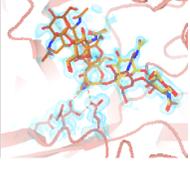
A) Brightfield view of crystals used to determine the structures reported in this paper. **B)** Hanging drop crystallization trays were set up as a 2-condition gradient to identify optimal crystallization conditions for AMCase + GlcNAc_n. pH increased along the X-axis from pH 3.70 to 5.60. Ligand concentration increased along the Y-axis from 0 mM to 29 mM [GlcNAc₂], 19 mM [GlcNAc₃], 10 mM [GlcNAc₄], or 8 mM [GlcNAc₅]. Black boxes indicate conditions where crystals grew. Lilac boxes indicate conditions for structures reported in this paper.

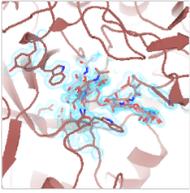
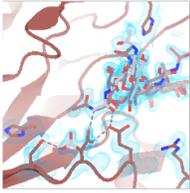
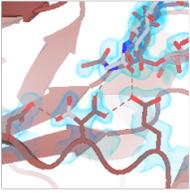
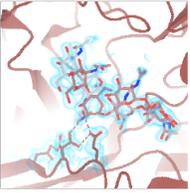
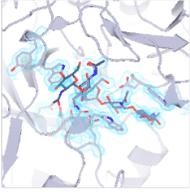
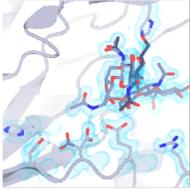
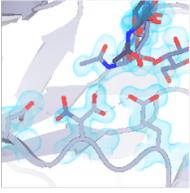
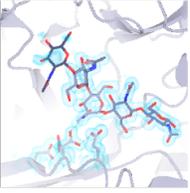
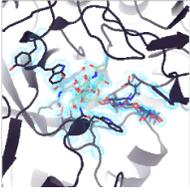
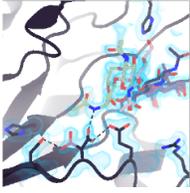
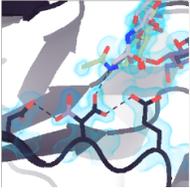
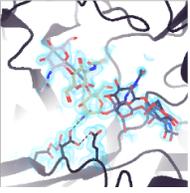
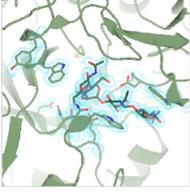
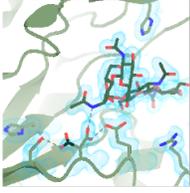
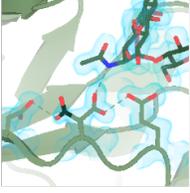
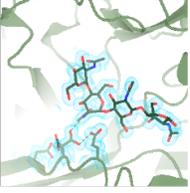
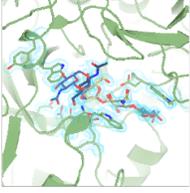
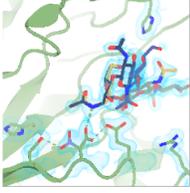
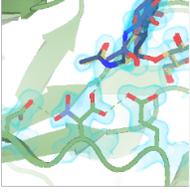
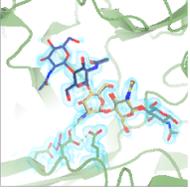
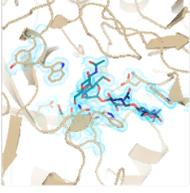
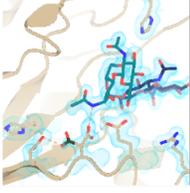
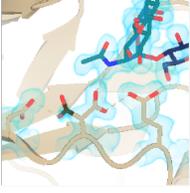
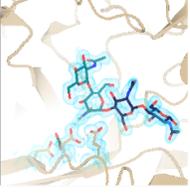
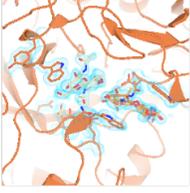
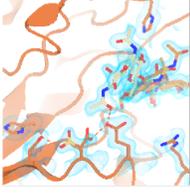
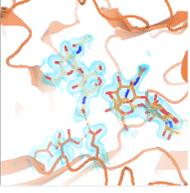
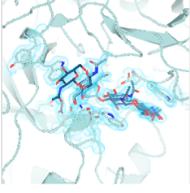
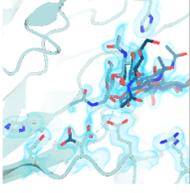
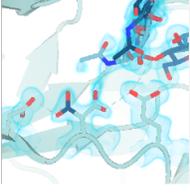
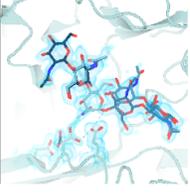


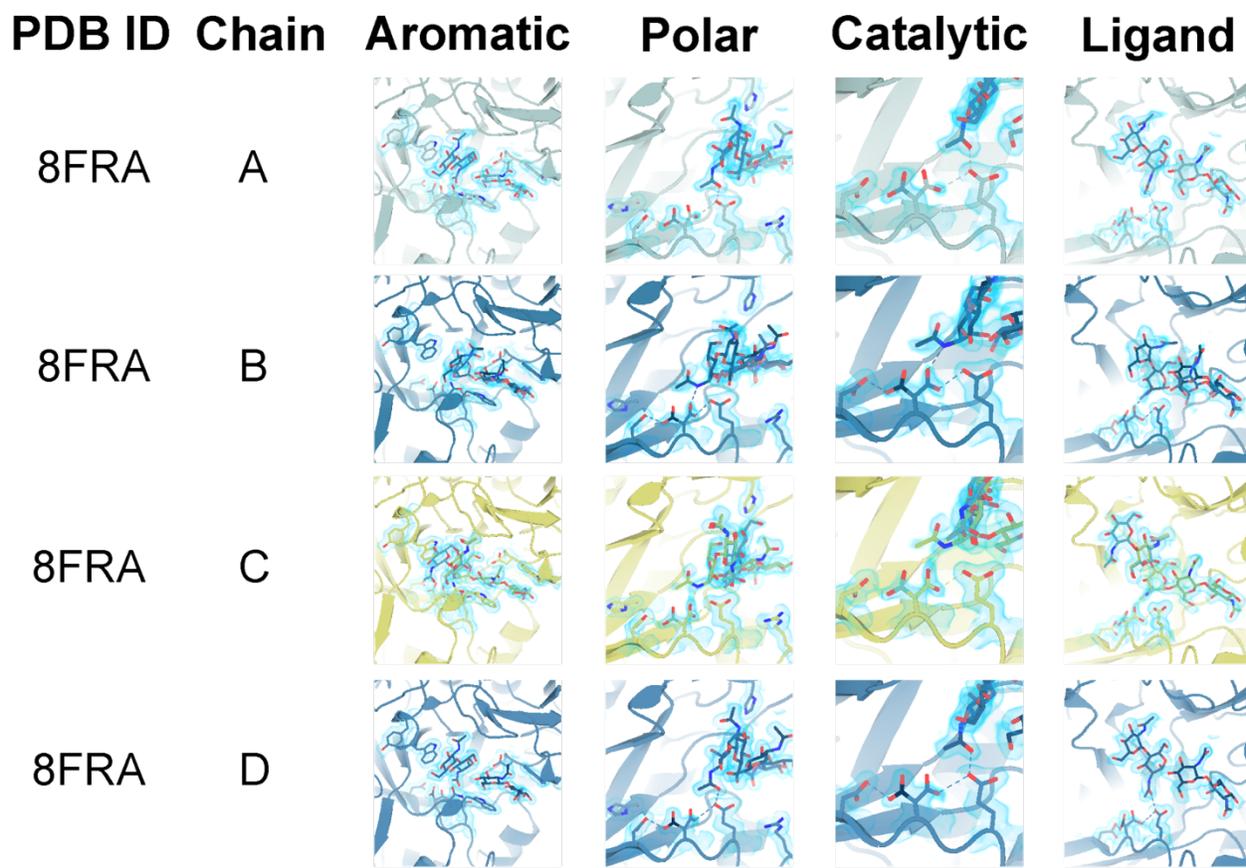
Supplemental Figure 3.4 | pKa of apo and holo mAMCase in the D2 *inactive* and *active* conformation.

PDB ID: 8FG5, 8FG7 (*apo*); 8GCA, 8FRC, 8FR9, 8FRB, 8FRD, 8FRG, 8FRA (*holo*).

Distribution of pKa across Asp136, Asp138, Glu140 between **A)** apo and **B)** holo mAMCase structures in the *inactive* or *active* conformation.

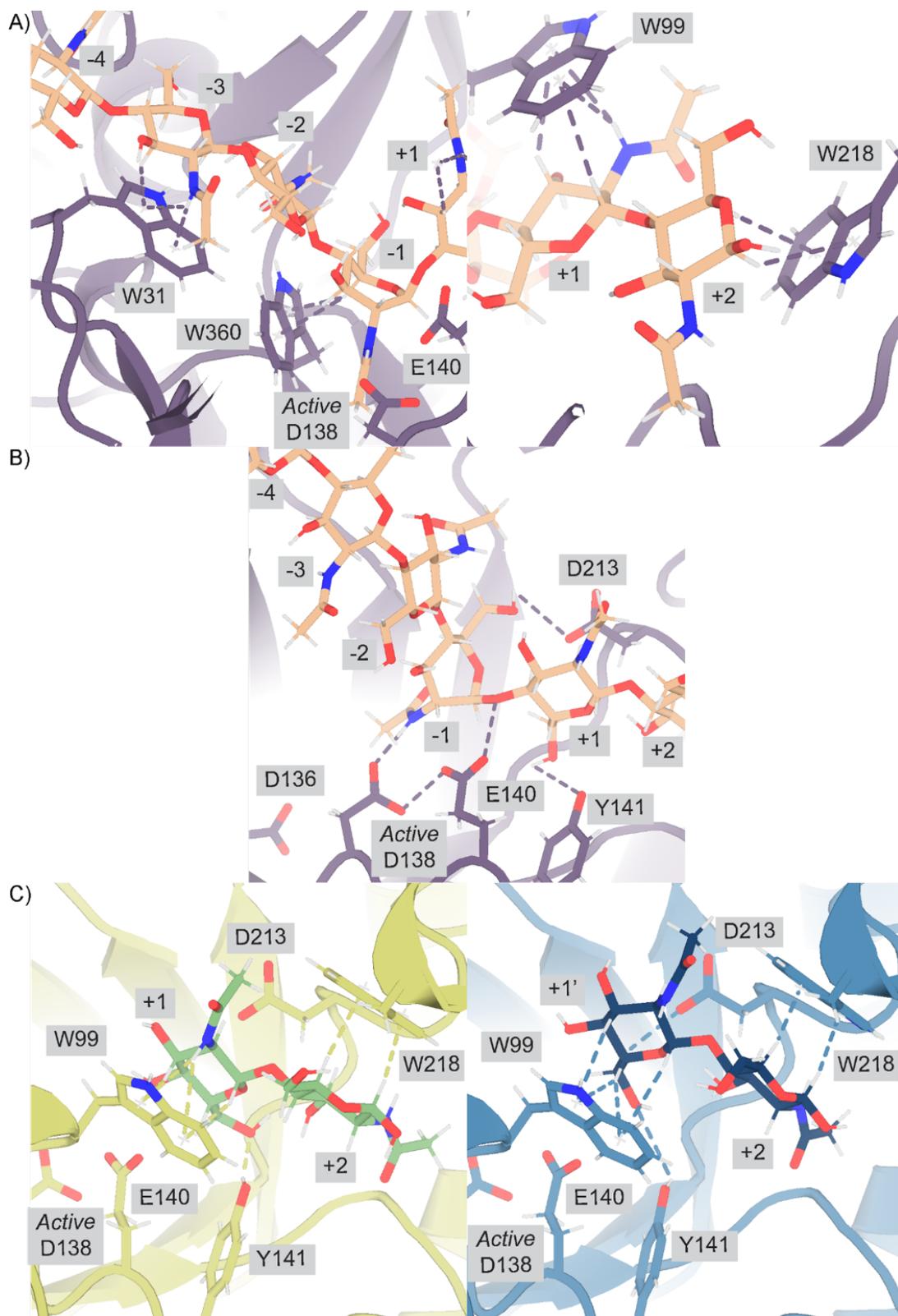
PDB ID	Chain	Aromatic	Polar	Catalytic	Ligand
8FG5	A				
8FG7	A				
8GCA	A				
8GCA	B				
8FRC	A				
8FRC	B				
8FR9	A				
8FR9	B				

PDB ID	Chain	Aromatic	Polar	Catalytic	Ligand
8FRB	A				
8FRB	B				
8FRB	C				
8FRB	D				
8FRD	A				
8FRD	B				
8FRG	A				
8FRG	B				



Supplemental Figure 3.5 | Overview of key residues for mAMCase activity.

A) Stick representation of ligand and aromatic residues Trp31, Tyr34, Trp99, and Trp218 in the active site with 2mFo-DFc map shown as a 1.2 Å contour (blue). **B)** Stick representation of ligand and polar residues Arg145, His208, Asp213, and His269 in the active site with 2mFo-DFc map shown as a 1.2 Å contour (blue). **C, D)** Stick representation of ligand and catalytic residues Asp136, Asp138, and Glu140 in the active site with 2mFo-DFc map shown as a 1.2 Å contour (blue).

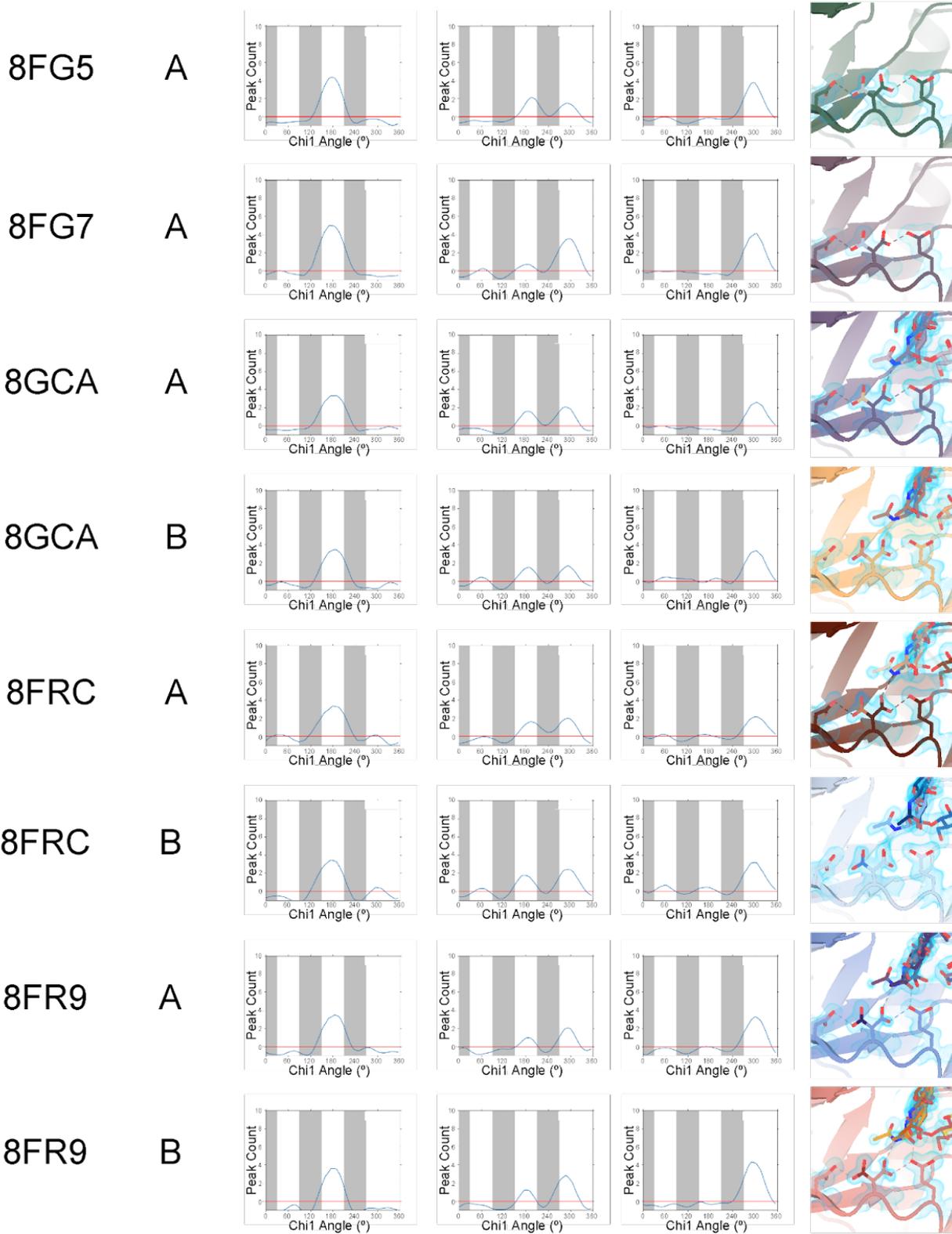


Supplemental Figure 3.6 | Protein-ligand interactions between mAMCase and chitin.

A) PDB ID: 8GCA, chain A with GlcNAc₆ modeled for viewing simplicity. Stick representation highlighting the stabilizing H- π interactions between Trp31, Trp360, and Trp218 and the -3, -1,

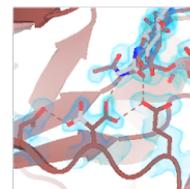
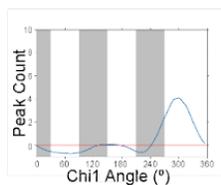
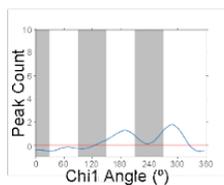
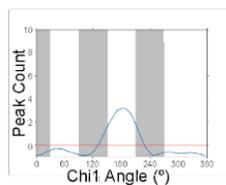
+1, and +2 sugars, respectively. **B)** PDB ID: 8GCA, chain A with GlcNAc₆ modeled for viewing simplicity. Stick representation highlighting the stabilizing hydrogen bond interactions between the -1 sugar and Asp138 (2.6 Å) and Asp213 (3.4 Å), and between the +1 sugar and Tyr141 (3.0 Å). Glu140 is 2.8 Å from the glycosidic oxygen bridging the -1 and +1 sugars. **C)** PDB ID: 8FRA, chains C (left) and D (right). Stick representation highlighting the stabilizing hydrogen bond interactions that we argue stabilize the +1 sugar (left; chain A) and the +1' sugar-binding subsite (right; chain B).

PDB ID Chain Asp136 Asp138 Glu140 Catalytic

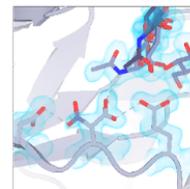
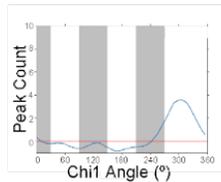
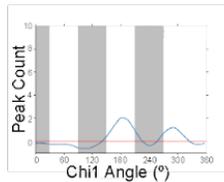
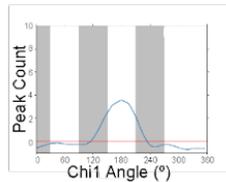


PDB ID Chain Asp136 Asp138 Glu140 Catalytic

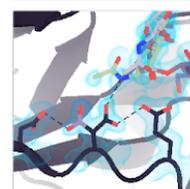
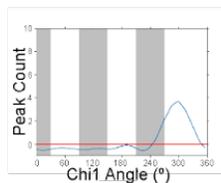
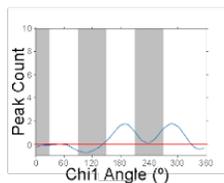
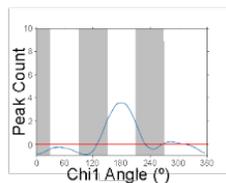
8FRB A



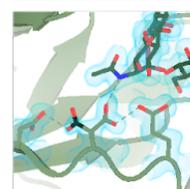
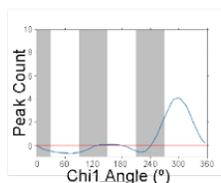
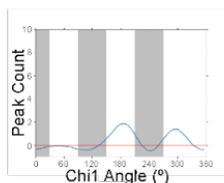
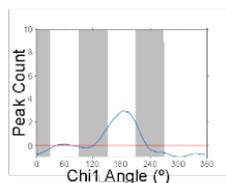
8FRB B



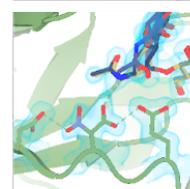
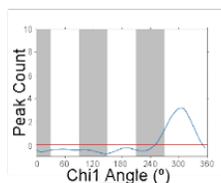
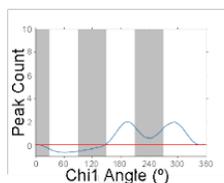
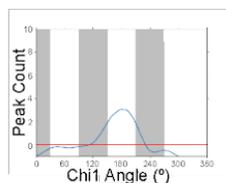
8FRB C



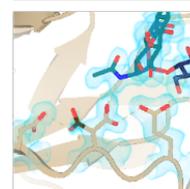
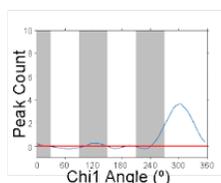
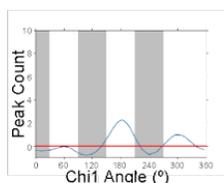
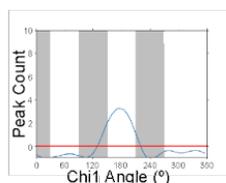
8FRB D



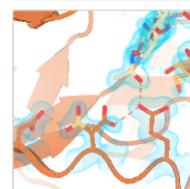
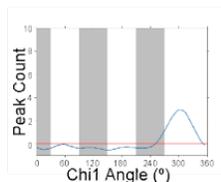
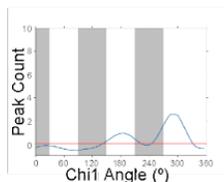
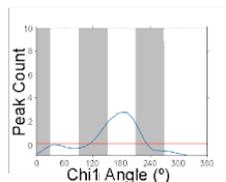
8FRD A



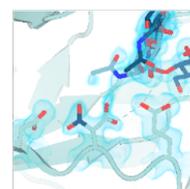
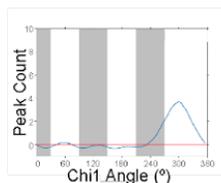
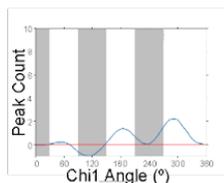
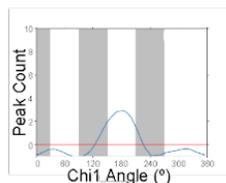
8FRD B

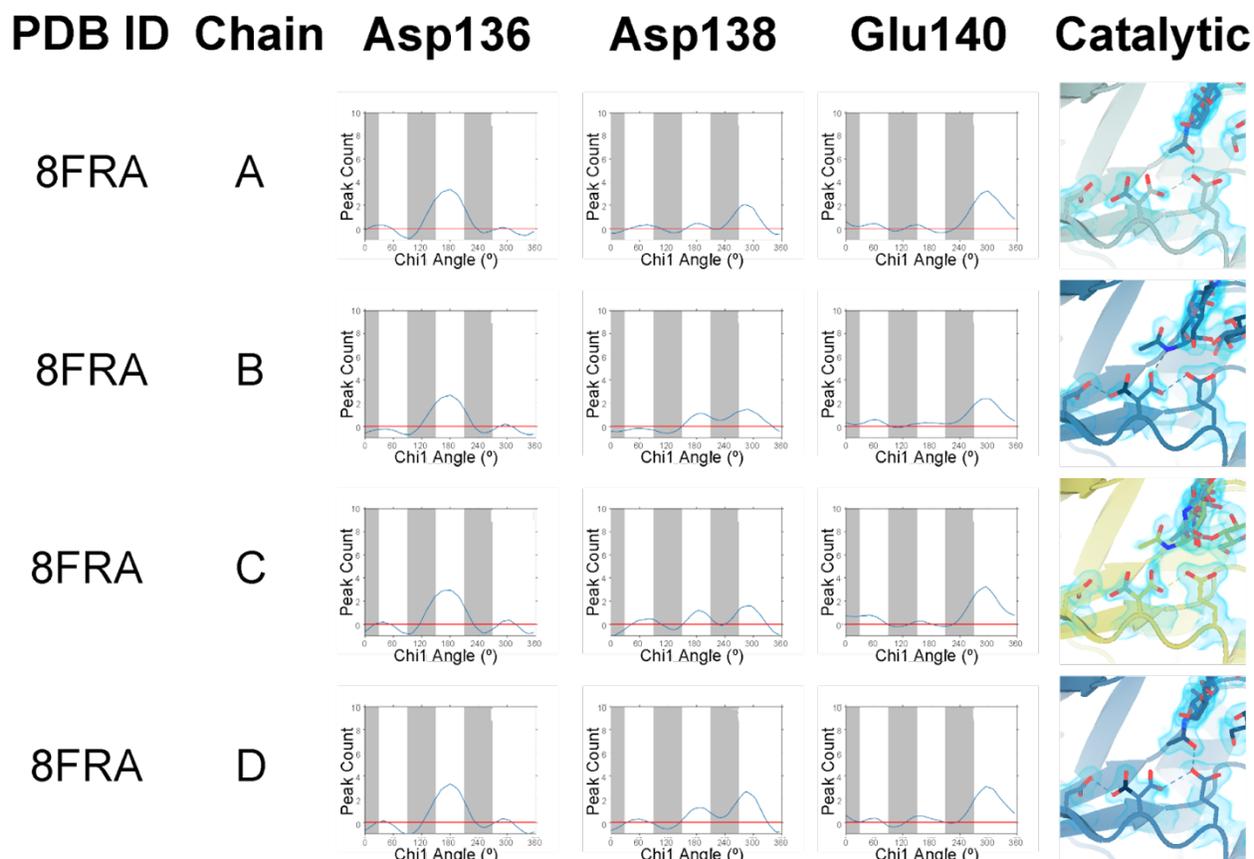


8FRG A



8FRG B





Supplemental Figure 3.7 | Ringer analysis of catalytic triad confirms alternative Asp138 conformations.

A) Ringer analysis to detect alternative conformations in electron density maps. Ringer detected one peak for Asp136 at $\chi_1 = 180^\circ$ and Glu140 at $\chi_1 = 300^\circ$, indicating only one conformation, whereas two peaks were detected for Asp138 at $\chi_1 = 180^\circ$ and $\chi_1 = 300^\circ$, indicating two alternative conformations. **B)** Stick representation of Asp136, Asp138, and Glu140 with 2mFo-DFc map volume shown as a 1.2 Å contour (blue).

Acknowledgments

General

We are grateful to Aashish Manglik and Mingliang Jin for providing ExpiCHO-S cells; to Liam McKay and Jose Luis Olmos, Jr. for support with the X-ray crystallography facility at UCSF; to George Meigs for assistance with X-ray data collection at ALS 8.3.1.; to Tzanko Doukov for assistance with X-ray data collection at SSRL 12-2; to Eric Greene, Duncan Muir, Stephanie Wankowicz, and Benjamin Barad for helpful discussions and critical feedback. Structural biology applications used at UCSF were compiled and configured by SBGrid.⁵⁵

Funding

This work was supported, in part, by California's Tobacco Related Disease Research Program (TRDRP) grant T29IP0554 (J.S.F). Research reported in this publication was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R01HL148033 (S.J.V.D., J.S.F.). Beamline 8.3.1 at the Advanced Light Source is operated by the University of California Office of the President, Multicampus Research Programs and Initiatives grant MR-15-328599, NIH (R01 GM124149 and P30 GM124169), Plexxikon Inc., and the Integrated Diffraction Analysis Technologies program of the US Department of Energy Office of Biological and Environmental Research. The crystallographic data was collected using beamlines at the Advanced Light Source, and the Stanford Synchrotron Radiation Lightsource. The Advanced Light Source (Berkeley, CA) is a national user facility operated by Lawrence Berkeley National Laboratory on behalf of the US Department of Energy under contract number DE-AC02-05CH11231, Office of Basic Energy Sciences. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is

supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. (1650113; R.E.D.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. R.E.D. is a Howard Hughes Medical Institute Gilliam Fellow. R.M.L. is supported by the Howard Hughes Medical Institute.

Competing interests

S.J.V.D. and R.M.L. are listed as inventors on a patent for the use of chitinases to treat fibrotic lung disease. S.J.V.D, R.M.L., and J.S.F. are listed as inventors on a patent for mutant chitinases with enhanced expression and activity.

Chapter 4 Contributions to Diversity, Equity, Inclusion, and Justice.

Roberto Efraín Díaz^{1,2}, D'Anne Duncan³

Affiliations

¹ Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94158, USA

² Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

³ Office of Diversity and Learner Success, Graduate Division, University of California, San Francisco, San Francisco, CA 94158, USA

Introduction

This chapter is a compilation of select projects focused on the advancement of diversity, equity, inclusion, and justice (DEIJ) in higher education and beyond. Academia, like all other social institutions, has been shaped by structural racism and oppressive politics. Its foundation relied upon the exclusion and exploitation of marginalized people for the benefit of the dominant class. In 1954, the Supreme Court of the United States ruled in *Brown v. Board of Education of Topeka* that “separate but equal” facilities were inherently unequal and violated the Equal Protection Clause of the Fourteenth Amendment¹. This decision ended the legalized racial segregation of children in public schools. While *Brown v. Board* exclusively focused on educational access and did not address the issue of hiring and other professional opportunities for racialized scholars, it became a legal precedent for subsequent lawsuits demanding the desegregation of higher education institutions. Despite legislative mandates forcing higher education institutions to allow Black men (*Brown v. Board*, 1954) then women of all races (Title IX, 1972)² equal access to education, their presence in academic spaces continues to be challenged by white men (and those who serve white supremacy).

Since the activist moment of the 1960s, there have been factions within academia that refuse to settle for tolerance and continue to fight for equity and justice. Student-led protests at San Francisco State University in 1968 resulted in the foundation of what is known today as the College of Ethnic Studies³. These protests challenged the dominant white perspective in academia by fighting for student-driven education that emphasized self-determination rather than simply integration.

Diversity, equity, and inclusion (DEI) in the 21st century echoes sentiments of inclusion and tolerance that were once championed during the 1960s civil rights movement. While DEI

efforts have marginally improved the representation of marginalized scholars in academia, these efforts have ultimately failed to undo centuries of anti-Black racism, sexism, and other oppressive ideologies integral to American society. This chapter will outline work done at UCSF that aims to go beyond tolerance and diversity in pursuit of equity and justice. I will discuss my efforts to promote LGBTQ+ safety in academic conferences, cultivate community amongst Queer scientists, and envision a more equitable and just graduate training environment.

Improving LGBTQ+ Inclusion in Academia

Graduate and Postdoc Queer Alliance

LGBTQ+ Coffee Hour

My life changed when I woke up in San Francisco on June 12th, 2016, and heard the news of the Pulse Shooting in Florida, my home state. Being 3,000 miles away, I felt isolated from my family, friends, and Queer community. But I went to the Castro District, San Francisco's historically LGBTQ+ neighborhood, and found myself surrounded by complete strangers who were experiencing the same feelings of loss and anger that I felt. It was in that moment that I understood the importance of community on one's mental health and sense of belonging, two elements essential to promoting social justice. At this point in my academic career, I began to recognize the interplay between academia and society and asked myself: What role does community play in our ability to make impactful change?

When I returned to UCSF for graduate school in 2017, I sought out opportunities to create community with fellow marginalized students. One of my first experiences building community at UCSF was through the revival of the Graduate and Postdoc Queer Alliance. The Graduate and Postdoc Queer Alliance (GPQA) is a trainee-led organization dedicated to providing community, professional development, and networking opportunities for LGBTQIA+ individuals at UCSF. I established a weekly coffee hour for LGBTQ+ people at UCSF to take respite from their work, enjoy a heavily discounted coffee, and meet new people. After a few months, I noticed that these coffee hours were almost exclusively attended by cis, white men. I questioned whether this was the result of who works at UCSF and/or if GPQA had failed to create a welcoming space for people of different genders and races. In collaboration with Adair

Borges and Iris D. Young, we began to focus our efforts on building a community space that welcomed and supported Queer and Trans people of all genders, races, abilities, and experiences.

This experience taught me to be more self-critical, intentional, and intersectional with how I approached community building, and diversity, equity, and inclusion work. Sometimes organizations will praise themselves for initiatives that make small improvements in representation but virtually no improvement in the lived experience of those marginalized people. It's disconcerting when someone claims to be improving LGBTQ+ representation in STEM, but then you examine their claims and see that all their work is focused on white, cisgender, gay men with very similar experiences. I feel like that's leaving out a lot of people who are oppressed by the same systems as white, cisgender men.

Petition to SACNAS

Scientific conferences need to be accessible to people from all backgrounds for scientific progress to continue unhindered. Conferences not only provide an important venue for sharing research with the greater community, but also offer early career advancement opportunities to early career researchers. Many scientists, especially from marginalized backgrounds, have shared stories of attending a conference where they have built personal and professional connections that encouraged them to persist in academia. These connections teach scientists how to communicate their ideas, shift their perspectives, and engage with people from different personal and professional backgrounds. The creation and nourishment of a scientific ecosystem within and across fields is critical to the advancement of science.

To varying degrees, many professional organizations routinely consider accessibility as they choose sites for their conferences. Some accessibility considerations include proximity of

the conference center to an airport, transportation options, and lodging expenses. For example, conference site options are further limited by the expected size of the meeting and the time of year. Thus, conference organizers are tasked with the difficult calculus of minimizing cost to attendees while maximizing the potential number of attendees and exhibitors. Unfortunately, this calculation often ignores, or rationalizes, the sociopolitical climate of a location and its effect on conference attendees.

Unfortunately, one's ability to attend a conference is often hampered by local and state politics. In September 2016, the state of California passed AB 1887 to "take action to avoid supporting or financing discrimination against lesbian, gay, bisexual, and transgender people"⁴. This law prohibits the use of California state funds for travel to states with discriminatory laws, such as Texas, Florida, and as of 2023, 21 additional states. Connecticut, Minnesota, New York, Vermont, and Washington have also enacted similar legislation. This legislation provides a financial barrier for LGBTQ+ researchers who want to attend a conference in any of these 23 states. For many early career scientists, not attending their discipline's conferences is highly undesirable, regardless of the financial or personal implications associated with attending. Queer* scientists must consider the potential benefits and risks associated with attending a conference in a state without LGBTQ+ protections. Do the benefits of attending outweigh the risk of legalized discrimination based on your sexuality or gender expression?

Professional organizations have increased their diversity, equity, and inclusion efforts focused on white women and, to a lesser extent, people of color, but largely fail to consider LGBTQ+ inclusion and equity. Organizations have begun to ask attendees their pronouns, offer

* For brevity, I use "Queer" to refer to both sexual orientation and gender minorities, including trans and non-binary people.

rainbow colored name tags, and even provide gender-neutral bathrooms at conference venues. But a Queer person's sense of safety and belonging should not be confined to the conference center. Queer attendees should not have to worry about being denied service at a local restaurant or being assaulted, knowing that the state's legislation will not rule in favor of the Queer person. In 24 states, there are no anti-discrimination protections for LGBTQ+ individuals seeking public accommodations, such as in restaurants or hotels. In 2023, a lack of protections is no longer a sign of neutrality as 46 states have proposed or enacted legislation that actively discriminates against people based on gender identity, gender expression, and/or sexual orientation⁵.

SACNAS 2018 was held in Texas, who enacted two laws (SB4, HB3859)^{6,7} that discriminate against undocumented and LGBTQ+ people, respectively. SACNAS strives to be an inclusive organization for Chicanos/Hispanics and Native Americans—yet the decision to host the conference in Texas disrupts the foundation for a truly inclusive environment. Moving forward, SACNAS needs to inhabit and promote environments that respect its members, their various identities, and the intersection that exists between these identities.

Many academic societies, including the Animal Behavior Society, International Association for Plant Taxonomy, Society for the Study of Evolution, American Society of Naturalists, and Society of Systematic Biologists, have amended their bylaws to prohibit hosting their annual conferences in locations with discriminatory laws. SACNAS should be held to the same, if not a higher, standard of diversity, equity, and inclusion given the marginalized groups this society serves.

In 2018, Melissa Spear, Ramiro Patiño, Sy Redding, and I developed a list of policies for SACNAS national to adopt that we believed demonstrate an explicit commitment to inclusivity of all identities, including LGBTQ+ identities and documentation status:

- Boycotting any state that has statewide discriminatory laws in effect, including the current states: Alabama, Kansas, Kentucky, Mississippi, North Carolina, Oklahoma, South Dakota, Tennessee, and Texas[†]
- Adding an Anti-Discrimination clause to conference-related contracts allowing SACNAS to relocate from that state with no financial loss if said state passes a discriminatory law at least (1) year ahead of the next conference date.

The lack of commitment from SACNAS to not hold their conferences in states with discriminatory laws has disadvantaged thousands of students from California, Connecticut, Minnesota, New York, Vermont, Washington, and other states who choose not to attend due to personal discomfort, safety concerns, and/or lack of financial resources.

Unfortunately, anti-LGBTQ+ legislation is not unique to the United States. In over 70 countries, same-sex relationships between consenting adults are criminalized and in 8 of those countries, gender expression “imitating the opposite sex” is criminalized. In 7 countries, same-sex relationships are punishable by life imprisonment or capital punishment. International organizations need to consider the accessibility of certain countries to its participants, whether due to travel restrictions or personal identity.

Attempts to broaden participation, however, are undercut by holding conferences in jurisdictions with discriminatory legislation, particularly targeting LGBTQ+ people. This legislation ranges from a lack of protection against discrimination to imprisonment and, in 8 countries, the death penalty⁸. In these cases, the resources and effort spent on outreach to recruit

[†] The list of states that California prohibits the use of state funds for travel to has expanded from 9 states in 2018 to 23 in 2023.

diverse attendees are undermined by putting professional meetings in places where it is not safe for all members of the professional society to participate.

There is no one ideal location to host a scientific meeting, and every location will likely exclude some section of the global scientific community through legislation as we describe above, visa restrictions, and/or cost, among other factors. However, when organizers actively make decisions that exclude marginalized members from academic spaces, these scholars are being told that their participation is not a priority, and that their intellectual contributions to their field are dispensable.

Improving Equity in Graduate Admissions

Diversity Network Initiative

Graduate school is an unequivocally difficult experience, which makes the decision of where to matriculate a daunting process especially for marginalized scholars. How does someone decide which prestigious (and racist) institution will provide them with the best training opportunities in the least harmful environment? What if applicants could talk to current students with shared identities about their experience to facilitate the decision-making process?

To foster a deeper sense of belonging and psychological safety at UCSF for marginalized prospective students, I created the Diversity Network Initiative (DNI) in collaboration with Dr. D'Anne Duncan, the Assistant Dean of Diversity and Learner Success in the Graduate Division. This initiative connects marginalized prospective students with current students from a similar background and promotes candid conversations about the culture at UCSF, especially as it pertains to diversity, equity, inclusion, and justice. Our primary goal is to increase the visibility of marginalized students in PhD programs, promote community building, and provide insight

into an institution's culture. We decided to implement the DNI at the interview stage since this is when applicants are not only being evaluated by UCSF faculty, but also evaluating UCSF faculty, students, staff, and culture. The applicants' assessment of UCSF as a potential future training environment is critical to their decision to matriculate.

Prospective and current students are asked a series of demographic questions, then we match the applicant with a student mentor based on shared identities and introduce them via email. We encourage the applicants and their mentors to engage in an open and honest dialogue (ideally in person) about the experiences of marginalized people at UCSF and the climate surrounding DEI. In the scenario where an applicant has an identity that their mentor match does not have, we assign them a second current student to address this unmet need.

In 2018, we piloted the DNI in Tetrad, which consists of Biochemistry and Molecular Biology, Cell Biology, and Genetics. Tetrad received over 400 applications and offered 79 interviews to applicants. Of those 79 applicants, 28 participated in the Diversity Network Initiative (35% participation rate). We solicited feedback from participants so we could improve the program for the following year. Some feedback we received was:

- "It gave me a clear sense of how content the students were with the program.... It allowed me to ask them questions naturally and interact with them as I might if I were a graduate student."
- "I greatly enjoyed the Diversity Network Program, and it helped make UCSF stand out among the schools I applied to."

Based on the positive feedback from participants of the pilot program, we presented the Initiative to the Basic Science graduate program administrators to expand to additional programs.

The Diversity Network Initiative was expanded to seven additional programs for the 2018-2019 admissions cycle. These programs included Neuroscience, Medical Scientist Training Program, Biophysics, Biomedical Informatics, Chemistry and Chemical Biology, and Pharmaceutical Sciences and Pharmacogenomics. Each of these programs received anywhere from 140 to over 500 applications, with a cumulative total of over 2,000 applications. Of those 2,079 applications, 340 applicants were offered an interview and 218 of those who interviewed participated in the Diversity Network Initiative (64% participation rate). Six of the programs had over 50% of their applicants participate, and two of those programs had over 75% participation. Again, we solicited feedback from participants:

- “I felt like this provided me a platform to ask questions that I might otherwise deem too personal or may not apply to my student host.”
- “My student was very helpful and honest about her experiences. It was very meaningful to speak to someone on the other end of such a long program, and especially someone who had shared many of my experiences/perspectives as an applicant.”
- “One of the reasons that helped me to make my final decision of accepting the offer from UCSF is its climate of diversity and inclusion. The proactive nature of UCSF when promoting diversity and inclusion impressed me.”

Based on the continued positive reception, we conducted the Diversity Network Initiative in the 2019-2020 admissions cycle but could not evaluate its impact due to the COVID pandemic.

Despite pausing the Diversity Network Initiative indefinitely, D’Anne and I have identified improvements we would like to implement if we restart the Initiative in the future. One

improvement is focused on automating the applicant-mentor pairing process. The current matching process requires someone to review each applicant's and current student's responses, then decide who should be paired together based on those responses. This is not only time-consuming but also subject to bias and inconsistency. Significant discussion is required to ensure that we are not integrating our own biases into the automation process. Another improvement would be integration of the DNI into the interview process as a core component. Currently, DNI mentors serve as secondary or tertiary contacts for prospective students. This can cause the applicant to feel overwhelmed and reduces accountability for the DNI mentor to contact their prospective student. If the DNI mentor also serves as the applicant's host during their interview visit, this would provide significantly more time for the applicant and mentor to interact and engage in various discussions that may influence the applicant's decision.

We anticipate that this initiative will continue to have positive effects on the recruitment and retention of marginalized students and are actively seeking opportunities to expand this program beyond UCSF Graduate Division Basic Science programs.

Conclusion

Initiatives should not claim to be promoting diversity, equity, and inclusion when they are neglecting Black people, Indigenous people, people of color, Disabled people, trans and non-binary people, and so many more marginalized identities. We—as an abstract collective of people committed to diversity, equity, inclusion, and justice—cannot just support one axis of identity. Support needs to apply to all the identities that community members bring with them to a shared space. If we're not familiar with the needs of certain identities, then we need to do the work and learn how to support them. Individuals with multiple identities deserve to feel welcomed and supported in all the communities that they belong to.

In closing, I want to share some words of wisdom based on my experiences advocating for DEIJ over the past 6 years.

1. Carry out your work slowly and thoughtfully. You may cause harm to those you intend to help if you don't proceed with intention.
2. You may not make as much progress as you'd like, and that's okay. Your primary role is as a PhD student. Any contributions to advancing DEIJ are important and should be recognized as such.
3. “Nothing about us without us.” Collaborate with the people who are impacted by your work. These people are experts on their own lived experiences and know what resources they need.

I hope this chapter has provided some context for how I approach DEIJ work in academia and beyond. To me, this work is not meant to sanitize an institution's reputation or to diversify the faces of white supremacy, but to redefine what is possible within the academy in service of

all marginalized people. Whether you have read my entire thesis or only this chapter, thank you for your attention.

“It is our duty to fight for our freedom.

It is our duty to win.

We must love each other and support each other.

We have nothing to lose but our chains.”

— **Assata Shakur**, *Assata: An Autobiography*

References

1. Brown v. Board of Education (1954). *National Archives*
<https://www.archives.gov/milestone-documents/brown-v-board-of-education> (2021).
2. Education Amendments Act of 1972, 20 U.S.C. §§1681 - 1688 (2018).
3. Sueyoshi, A. & Sujitparapitaya, S. Why Ethnic Studies. *Ethnic Studies Review* **43**, 86–102 (2020).
4. Mandated state-sponsored travel to states with discriminatory laws, Cal. Gov. Code § 11139.8, subd. (a)(5).
5. Mapping attacks on LGBTQ rights in U.s. state legislatures. *American Civil Liberties Union* <https://www.aclu.org/legislative-attacks-on-lgbtq-rights> (2023).
6. S.B. 4, 85th Legislature, 2017 Reg. Sess. (Tex. 2017).
<https://capitol.texas.gov/BillLookup/History.aspx?LegSess=85R&Bill=SB4>
7. H.B. 3859, 85th Legislature, 2017 Reg. Sess. (Tex. 2017).
<https://capitol.texas.gov/BillLookup/History.aspx?LegSess=85R&Bill=HB3859>
8. LGBT equality index. *Equaldex*. Available from: <https://www.equaldex.com/equality-index>. Accessed 05/2023 by RED.

Acknowledgments

General

The Graduate and Postdoc Queer Alliance would not exist in its current iteration without contributions from Martin Kampmann, Adair Borges, Iris D. Young, Steven A. Cincotta, Matthew J. Ryan, Yefim Zaltsman, Brian Bender, Branden Barger, Klint Jaramillo, Tracy Garcia, and Zara Weinberg. Thank you to Melissa L. Spear, Ramiro Patiño, Sy Redding, James Fraser, the UCSF Makers Lab, and SACNAS chapters across the country for standing up for LGBTQ+ rights; Alexander Bond, Itati SantaMaria, and Janet D. Stemwedel for collaborating on an article about LGBTQ+ Inclusion in Conferences; Toni Hurley for allowing D’Anne and I to pilot the Diversity Network Initiative in Tetrad, Geri Ehle, Nicole Flowers, Julia Molla, and Patricia Veitch for implementing the DNI in your graduate programs, and to Greyson Lewis for your work on automating the DNI. Thank you to Kelly Corso and Will Clark for critical feedback on this chapter.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:



DD22230EE09C499...

Author Signature

5/30/2023

Date