

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Maximizing Interpretability from Complex Experiments in Structural Biology and Biochemistry

Permalink

<https://escholarship.org/uc/item/8t7048v8>

Author

Barad, Benjamin Asher

Publication Date

2019

Peer reviewed|Thesis/dissertation

Maximizing Interpretability from Complex Experiments in Structural Biology and Biochemistry

by
Benjamin Barad

DISSERTATION

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

in

Biophysics

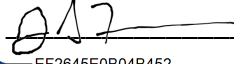
in the

GRADUATE DIVISION

of the

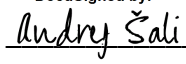
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

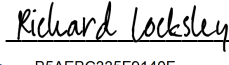
DocuSigned by:

EF2645E0B04B452...

James Fraser

Chair

DocuSigned by:


Andrej Šali

DocuSigned by:

B5AEBC335F9140F...

Richard Locksley

Committee Members

Copyright 2019

by

Benjamin Asher Barad

Acknowledgements

I first want to thank James Fraser, my thesis advisor, for his mentorship and support over the last six years. His scientific approach has been a standard for me to try to match, both in terms of his rigorous approach to quantifying structural data to maximize its interpretability, as well as his relentlessly positive and inclusive approach to doing science generally. James has constantly encouraged me to explore new methods and approaches without fear of failure, and I hope to carry forward that expansive approach to tackling biological problems with every tool available to me.

I further want to acknowledge the tremendous support and mentorship of faculty at UCSF and elsewhere who have given me advice, collaborated with me, sat on my committees, and provided opportunities to share my work. Elizabeth Sattely, my undergraduate research advisor at Stanford, gave me tremendous independence when I was just starting out in her lab, and her trust and belief in me was what made me want to go to graduate school. Yifan Cheng and Sophie Dumont took me into their labs for rotations, and the skills and intellectual approaches I learned from each have impacted this manuscript and will continue to have an effect on the work I produce for the rest of my scientific career. My committee members, Rich Locksley, Andrej Sali, and Michael Fischbach, gave me career advice and reminded me to always ask whether the juice is worth the squeeze.

Most of what I did in grad school I learned from the members of my lab. The spirit of collegiality and collaboration within the lab was tremendous when I joined and has only grown while I have been here. I especially want to thank Daniel Keedy, Lin Liu, David Mavor, Michael Thompson, and Rahel Woldeyes for teaching me along the way.

I would not have made it through grad school without friends and classmates who were excited for me when I succeeded and helped me keep my head up when I failed. Justin Biel, Evan Green, Lillian Kenner, Eugene Palovcak, Paul Thomas, Erin Thompson and Alex Wolff were classmates that reminded me how joyful biophysics research can be, and made UCSF a wonderful place to work. Ilan Chemmama was a classmate and a close friend who was always ready to pause work and go play squash. Nick Merrill was always ready to talk about new technologies and their impacts on the world, as well as PhD life outside of the life sciences. Adam Rothman, Laura Weaver, Gordon Dean, George Michopolous, and Jean Carlos Faust were my escape valve and my dungeons and dragons party.

My family's support was invaluable to me in the pursuit of this dissertation work. My parents, Kelsey and Mark, my step-parents, Pascale and Joel, my sisters, Maya and Lila, my stepbrothers, Seth and Sam, and my grandfather George all had faith in me and reminded me that there was more going on in the world than was happening in lab. They reminded me constantly why I want to learn more about the world.

This work is dedicated most of all to my partner in life, Rachel, who kept me sane through the 6 years of stress and hard work, and without whom I would not have nearly as much joy in the world.

Contributions

Chapter 1

Barad, B.A.; Echols, N.; Wang, R.Y.-R.; Cheng, Y.; DiMaio, F.; Adams, P.D.; Fraser, J.S. 2015. Side-chain-directed model and map validation for 3D Electron Cryomicroscopy. *bioRxiv*. 014738. doi:10.1101/014738. CC-BY-NC-ND 4.0 International license.

Chapter 2

Thompson, M.C.; Barad, B.A.; Wolff, A.M.; Cho, H.S.; Schotte, F.; Schwarz, D.M.C.; Anfinrud, P.; Fraser, J.S. 2018. Temperature-Jump Solution X-ray Scattering Reveals Distinct Motions in a Dynamic Enzyme. *bioRxiv*. 476432. doi:10.1101/476432. CC0 license.

Maximizing Interpretability from Complex Experiments in Structural Biology and Biochemistry

Benjamin Asher Barad

Abstract

Proteins are complex macromolecules whose structure informs their function and regulation in difficult to predict ways. Understanding their shape, dynamics, and regulation all pose major challenges in terms of collecting, analyzing, and interpreting data. In my dissertation I describe two contributions to data analysis for determining the structure and dynamics of proteins using novel approaches, as well as experimental work querying the function of an enzyme with a particularly recalcitrant substrate.

In the first chapter of this dissertation, I develop a tool, EMRinger, for the emerging field of high resolution electron microscopy that takes advantage of prior physical information about model geometry to more effectively determine if the model is built correctly into the map. This work adapted the tool Ringer, which had been previously developed in the Alber lab, in order to identify the dihedral angle for side chains with the greatest density, and confirm that the distribution of those peak positions does not violate the constraints of side chain dihedral angles to rotameric positions. This approach allows for orthogonal validation of backbone position in density (using the side chain density as a “lever”), which generally improves with refinement and is among the most sensitive model-in-map validation tools available for high resolution electron microscopy.

In the second chapter, I present progress on applying temperature jumps to folded proteins to quantify kinetics of the intrinsic motions in proteins that impact their function and

regulation. Using a pulsed infrared laser, we raise the temperature of a protein solution in nanoseconds, and follow the progression of the structure of the protein using solution x-ray scattering. As the protein changes temperature, the conformational equilibrium of the protein shifts as higher energy states become more accessible. By following this progress over the nanoseconds, microseconds and milliseconds following the heating pulse, we are able to reconstruct the relaxation of the protein into its new conformational equilibrium. With this information, we can gain kinetic information about the conformational landscapes of our sample, and using mutations we correlate the rates we observe with existing structural models of dynamics that have been characterized by x-ray crystallography.

In the third chapter, I investigate the mechanics of Acidic Mammalian Chitinase, which has the role of breaking down the recalcitrant polysaccharide chitin in the stomach and lungs of mammals. Mutations to acidic mammalian chitinase have previously been identified that lead to either protection against allergic asthma, and previous work has determined that these mutations lead to an increase in the activity of the enzyme. In order to better determine how these mutations affect activity, I developed new methods to assay chitinase activity. I use these methods to characterize the effects of the asthma-associated mutations, as well as investigating the role of the individual domains of acidic mammalian chitinase in degrading crystalline chitin and the differences in behavior of acidic mammalian chitinase and the other chitinase expressed in mammals, chitotriosidase. Additionally, I attempt to engineer hyperactive chitinases, and show that direct evolution based on screening with traditional fluorogenic oligomer substrates does not necessarily lead effectively to enzymes which are hyperactive against complex substrate, emphasizing the need for sensitive and high throughput methods to quantify degradation of crystalline chitin.

Table of contents

EMRinger: Side-chain directed density analysis for high resolution electron cryomicroscopy	1
Contributing authors	1
Preface.....	2
Abstract.....	3
Introduction.....	3
Results.....	6
Discussion.....	23
Methods.....	25
References.....	29
Temperature-Jump Solution X-ray Scattering Reveals Distinct Motions in a Dynamic Enzyme.....	36
Preface.....	36
Abstract.....	38
Introduction.....	39
Results.....	43
Discussion.....	67
Methods.....	72
References.....	80
Dissecting and Engineering Mammalian Chitinases	88
Introduction.....	88

Results.....	92
Discussion.....	102
Methods.....	106
References.....	114

List of figures

EMRinger: Side-chain directed density analysis for high resolution electron cryomicroscopy

Figure 1.1 EMRinger χ_1 map value sampling reports on backbone position and guides side-chain conformation.....	7
Figure 1.2 EMRinger reveals statistical enrichment at rotameric χ_1 angles in high resolution EM maps.	10
Figure S1.1 Atomic models in cryo-EM maps.	11
Figure S1.2 EMRinger Score is unaffected by model size.	12
Figure S1.3 Histogram of peak counts for EMRinger scan of T20S Proteasome (EMDB 5778, PDB 3J9I) at a map value threshold of $0.242 \text{ e}^-/\text{\AA}^3$	13
Figure S1.4 Adjusted EMRinger Score degrades rapidly with decreasing resolution.....	14
Figure 1.3 EMRinger Scores report on effective refinement of atomic models into EM maps.	18
Figure S1.5 Histograms of TrpV1 models at multiple map value thresholds.....	19
Figure 1.4 Acidic residues are differentially altered by radiation damage.	22

Temperature-Jump Solution X-ray Scattering Reveals Distinct Motions in a Dynamic Enzyme

Figure 2.1 Overview of T-jump SAXS/WAXS experiments.	45
Figure S2.1 Typical X-ray energy spectrum of the pink beam (3% energy bandwidth) used for the reported SAXS/WAXS measurements.	46

Figure 2.2 X-ray scattering from bulk water acts as a sensitive thermometer for T-jump experiments.....	49
Figure 2.3 T-jump data processing involves a combination of scaling and subtraction operations that produce time-resolved difference scattering curves.....	51
Figure 2.4 Time-resolved T-jump data allow kinetic modeling of conformational dynamics.	56
Figure S2.2 Kratky plots for CypA reveal a small thermal disorder effect without protein unfolding.....	57
Figure S2.3 Intermolecular interactions are not temperature dependent for CypA solutions.....	58
Figure S2.4 Comparison of static scattering differences between CypA solutions at 13°C and 23°C (black curve), and time resolved differences (100µs-laser off) for a T-jump spanning a temperature range of approximately 15°C and 26°C (red curve).....	58
Figure S2.5 Kinetic analysis of X-ray scattering by singular value decomposition (SVD).	61
Figure 2.5 Guinier analysis can be used to estimate changes in physical parameters of the average protein particle in the CypA ensemble from the time-resolved scattering data.	62
Figure 2.6 Linear Eyring plots for each of the two relaxation processes observed in our T-jump experiment with CypA.....	62
Figure 2.7 Kinetic analysis of two CypA mutants with distinct effects on the enzyme's function demonstrate the link between the observed T-jump signal and functional dynamics.	66
Figure S2.6 Residuals for kinetic fits of T-jump SAXS/WAXS data for CypA variants.....	67

Dissecting and Engineering Mammalian Chitinases

Figure 3.1 Activity comparisons of AMCase catalytic domain and full length enzyme.	94
Figure 3.2 Comparison of AMCase and Chitotriosidase.	97
Figure 3.3 Comparison of activity of AMCase asthma-associated mutants.	99
Figure 3.4 Engineering of hyperactive AMCase mutants.	101
Figure S3.1 Processing 4MU assay data.....	108
Figure S3.2 Data processing for colloidal chitin clearance assay.....	109
Figure S3.3 Data processing for ferricyanide reduction assay.....	110
Figure S3.4 Data processing for chitO assay.	112

List of tables

EMRinger: Side-chain directed density analysis for high resolution electron cryomicroscopy

Table S1.1 EMRinger analysis of selected maps above 5 Å resolution with atomic models.	15
Table 1.1 Statistics pre- and post-refinement.....	19
Table 2.1 Calculated rates for the fast (k_1) and slow (k_2) relaxation processes measured from all T-jump experiments reported here.	63
Table 2.2 Enthalpies (ΔH^\ddagger) and entropies (ΔS^\ddagger) of activation for the fast (k_1) and slow (k_2) processes observed for WT CypA, calculated from Eyring analysis.....	63

Dissecting and Engineering Mammalian Chitinases

Table 3.1 Calculated rate constants of AMCcase catalytic domain and full length enzyme.....	94
Table 3.2 Calculated rate constants for AMCcase and Chitotriosidase.....	97
Table 3.3 Calculated rate constants for AMCcase and chitotriosidase.....	99
Table 3.4 Comparison of calculated rates for engineered mutants.	102

Chapter I

EMRinger: Side-chain directed density analysis for high resolution electron cryomicroscopy

Contributing authors

Benjamin A Barad^{1,2}, Nathaniel Echols³, Ray Y-R Wang^{4,5}, Yifan Cheng⁶, Frank DiMaio^{5,7}, Paul D Adams^{3,8}, James S Fraser^{1,9}

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

² Graduate Group in Biophysics, University of California, San Francisco, CA USA

³ Lawrence Berkeley National Laboratory, Berkeley, CA USA

⁴ Graduate program in Biological Physics, Structure and Design, University of Washington, Seattle, WA USA

⁵ Department of Biochemistry, University of Washington, Seattle, WA USA

⁶ Keck Advanced Microscopy Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco, CA USA

⁷ Institute for Protein Design, Seattle, WA USA

⁸ Department of Bioengineering, University of California Berkeley, Berkeley, CA USA

⁹ james.fraser@ucsf.edu

Preface

The bulk of this chapter appears as Barad *et al.* preprinted in *bioRxiv* in 2015, and a version of which was ultimately published in *Nature Methods* later the same year.

The results presented in this chapter describe the construction of a novel approach for validating models built out of electron microscopy (cryoEM). In the early 2010's, developments in direct electron detectors and motion correction enabled the structural characterization of proteins by electron microscopy to reach resolutions better than 4Å, which enabled the building of *de novo* models. However, because of differences in the way the data is collected in real space from electron microscopy, in contrast with the reciprocal space data collection used in x-ray crystallography, the existing tools for validation, such as R. While models and maps are made iteratively in x-ray crystallography, with the electron density map being generated using the phases from the model, in cryoEM the data contains both phase and amplitude information and the map can be solved independently of a model. The building and refinement of the model happen independently of the refinement of the map. Because of this, three questions become relevant for validation: first, to what degree is the map a reasonable representation of the image data? Second, to what degree is the model physically reasonable? Third, is the model a good representation of the data encoded by the map? Our approach incorporates components of the second and third questions by asking whether the model is built into the map in such a way that it can be modeled in a physically realistic manner. To determine this, we extended the tool Ringer, which was used previously to identify and present statistical evidence for the resolvability of alternative states of side chains in high resolution x-ray crystallography. In the process we developed a tool which we have dubbed EMRinger.

Abstract

Advances in electron cryomicroscopy allow for the building of *de novo* atomic models into high resolution Coulomb potential maps. While established validation metrics independently assess map quality and model geometry, methods to assess the precise fitting of an atomic model into the map and to validate the interpretation of high resolution features are less well developed. Here, we present EMRinger, which tests model-to-map agreement using side-chain dihedral-directed map density measurements. These measurements reveal local map density peaks and show that peaks located at rotameric angles are a sensitive marker of whether the backbone is correctly positioned. The EMRinger Score can be improved by model refinement, suggesting its utility as an effective model-to-map validation metric. Additionally, EMRinger sampling identifies how radiation damage alters scattering from negatively charged amino acids during data collection. EMRinger will be useful in assessing how advances in cryo-EM increase the ability to resolve and model high-resolution features.

Introduction

Recent computational and experimental developments in single particle electron cryomicroscopy (cryo-EM) now make it possible, in some cases, to build atomic models without any reference structures¹. In particular, advances in direct electron detectors², algorithms to classify heterogeneous samples^{3,4}, and motion correction^{5,6} are positioning cryo-EM to become a dominant method for determining the structure of dynamic molecular machines^{7,8} and membrane proteins^{9,10}. Because these structures are otherwise inaccessible to X-ray crystallography or NMR¹¹, it is important to determine the reliability of the resulting atomic models, in particular

side chain placement, for their eventual use in directing detailed mechanistic studies or drug development¹².

All-atom *de novo* cryo-EM models present several unique challenges for validation¹³. First, the Coulomb potential map itself must be validated to ensure that the images are properly recombined and that the resolution estimate is accurate¹⁴. These validation challenges are primarily addressed by assessing the “gold standard” Fourier Shell Correlation (FSC) between two independently refined half maps¹⁵. Next the chemical reasonableness of the model is assessed using tools that are commonly applied in X-ray crystallography¹⁶. Similarly to crystallography, it is essential to balance the agreement to experimental data with the deviation from ideal geometry, while maintaining acceptable stereochemistry, Ramachandran statistics¹⁷, side chain rotamers¹⁸, and clash scores¹⁶.

The weighting between data and prior structural knowledge is key to the third step of model-to-map validation: determining whether the structure is accurately fitted, but not over-fitted, to the map¹⁹. Several cross validation schemes have been proposed recently¹⁹⁻²¹ and can help to ensure that the model is not only reasonable, but also well fitted to the map. However, real space correlation coefficient-based metrics are dominated by low-resolution, high-signal features and can be complicated by the map B-factor sharpening approaches used prior to model building and refinement²². Additionally, these considerations may complicate high resolution model-to-map validation and render it difficult to assess the reliability of the highest resolution features of EM maps, such as side chain or ligand conformations.

A potential solution for assessing the reliability of high resolution models is to examine statistical signatures of the weaker, high resolution, data. In particular, testing whether cryo-EM maps recapitulate the preferred rotameric distributions of protein side chains is particularly

appealing since side chains represent the highest resolution features modeled *de novo* by cryo-EM structures. For example, the position of $C\gamma$ is constrained to avoid “eclipsed” steric overlaps, predicting that a small map value peak, contributed by the scattering from $C\gamma$, should occur at rotameric χ_1 dihedral (N-C α -C β -C γ) angles near 60° , 180° , and 300° (-60°)²³. Previously, we have used Ringer^{24,25} to measure the electron density at all possible positions of the $C\gamma$ atom for each unbranched side chain under ideal stereochemistry and fixed backbone assumptions. The primary conformation, which is usually well modeled by the crystallographic structure, is defined by a local peak in the distribution of density vs. dihedral angle. In addition, secondary electron density peaks in this distribution can represent alternative side-chain conformations. Across >400 structures, we observed that these secondary peaks were strongly enriched at rotameric positions, which suggested that the secondary peaks represented unmodeled alternative conformations that are populated enough to rise above the noise levels in the electron density map²⁵.

Here, we examine whether significant side chain density can be observed in EM maps by measuring the distribution of map value peaks around the χ_1 dihedral angle and testing whether the primary peaks are enriched at rotameric positions. Our method, EMRinger, can be used as a global validation metric as structure refinement proceeds and highlights specific areas where manual intervention can be used to improve the local fit of the model. As an additional application, we use EMRinger to probe electron radiation damage to side chains, demonstrating how increased electron dose alters the scattering behavior of negatively-charged side chains. The EMRinger approach directly reveals the side chain information content of EM maps and is complementary to, but independent of, existing validation procedures that report on the

resolution of the map, the physical reasonableness of the model, and the detailed fit of the model to the map.

Results

Side-chain χ_1 map density sensitively reports on backbone positioning

EMRinger interpolates the normalized value of the cryo-EM map at each potential position of the $C\gamma$ position around the χ_1 dihedral angle, assuming the currently modeled N, $C\alpha$, and $C\beta$ atomic positions (**Fig. 1.1**). We next plot the distribution of map values by dihedral angle (**Fig. 1.1**), which reveals local information about both the map and correctness of the backbone of the atomic model. The peak in the distribution represents the most likely position of the $C\gamma$ atom of the side chain, even when it is not immediately obvious “by eye”. Based on steric constraints²⁶ and data mining from high resolution X-ray structures^{18,27}, we expected that high quality EM maps with well fit backbone models would be enriched in χ_1 peaks near the rotameric angles of 60° , 180° , and 300° .

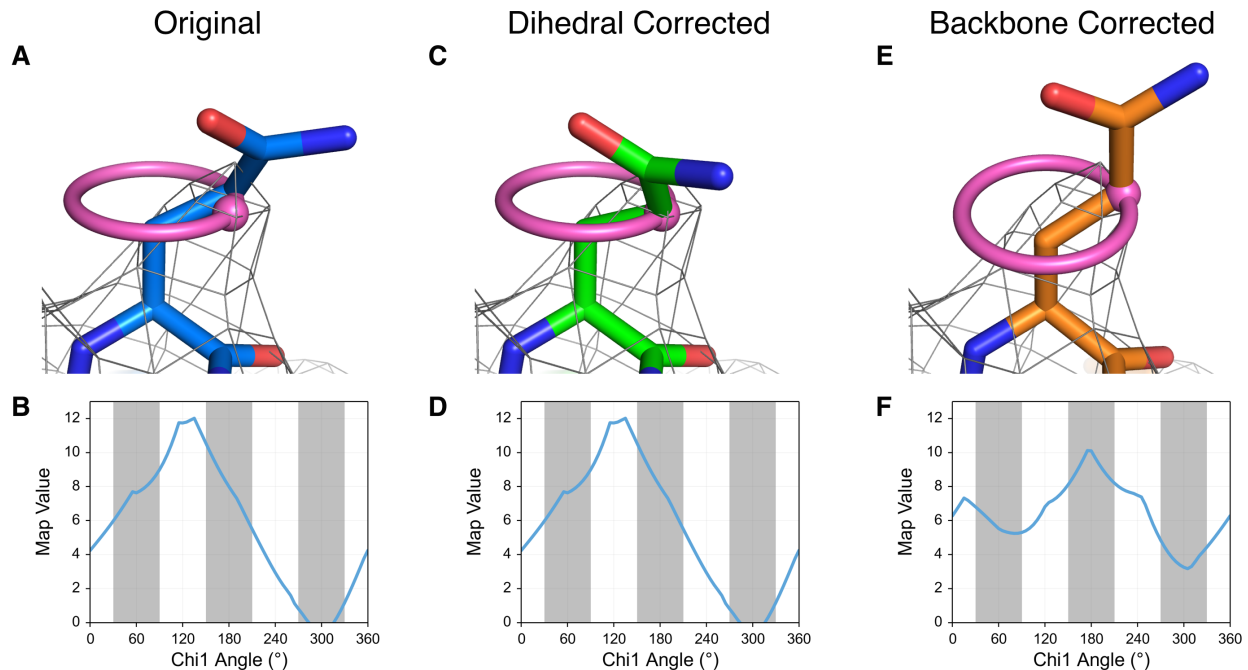


Figure 1.1 | EMRinger χ_1 map value sampling reports on backbone position and guides side-chain conformation.

(a) The side chain of TrpV1 Gln 519 (EMDB 5778, PDB 3J5P) is shown fitted, with a real space correlation coefficient (RSCC) of 0.590, to the potential map, shown at an isolevel of 10. (b) The EMRinger scan, reflected by the pink ring in a, for Gln 519 of Chain C reveals that the local map value peak (at 130°) occurs at a non-rotameric angle (white bars). This peak, shown as a pink dot in a, occurs 30° away from the modeled position. (c) The side chain can be rotated so that the χ_1 angle is at the map value peak (RSCC = 0.526). (d) The EMRinger results are unchanged as the sampling occurs relative to the backbone atoms, which have not moved. (e) Alternatively, the backbone position can be corrected with RosettaCM (DiMaio *et al*, Nature Methods, In Press) to place the model near a χ_1 map value peak a small reduction on the overall correlation of the residue to the map (RSCC = 0.442). (f) The peak at 175° is now in the rotameric region (grey bars).

However, there are several reasons, including noise in the map or an inaccurate model, why a side chain peak might occur at a non-rotameric angle. For example, residue Gln519 of TrpV1²⁸ (PDB: 3J5P) is modeled in a rotameric position, but has a peak at a non-rotameric angle in a 3.27 Å resolution map (EMDB: 5778) (Fig. 1.1). The distribution in map values by dihedral angle has a single dominant peak, suggesting that there is a local signal above the noise. The lack of a distinct peak can mean that the density threshold is too high, that the backbone is grossly mispositioned, or that the specific area has particularly local low resolution or high noise.

However, we observe singular peaks for most side chains in the TrpV1 map, which further suggests that noise is not the dominant reason why the peak occurs in a non-rotameric position. Alternatively, a peak in a non-rotameric position can indicate that the model is incorrect. If the N, C α , and C β atoms are improperly positioned in the strong potential surrounding the backbone, EMRinger will measure the map values in the wrong locations. It is important to note that the side chain is already modeled as rotameric and that changing the modeled side chain dihedral angle does not affect the result of EMRinger because the measurement relies only on the positions of the backbone and C β atoms (**Fig. 1.1**). In contrast, a small backbone adjustment places the C γ in the map value peak, while maintaining a rotameric side chain model, excellent stereochemistry, and a good map correlation (**Fig. 1.1**). Thus, EMRinger can identify well-fit backbone models because the local map value peaks will fall at rotameric angles. Our examination of EMRinger plots from several maps suggested that the enrichment of rotameric map value peaks could be used to assess the fit of the backbone model and the overall quality of the EM map.

EMRinger Score reports on the overall quality of the model and the map

To test the quality of model to map fit, we quantified the enrichment of EMRinger peaks within 30° of rotameric angles as a function of map value threshold. We recorded the position and map value of the peak for each side chain χ_1 angle in the 3.2Å resolution 20S proteasome map (EMDB 5623, PDB 3J9I) and observed that the distribution becomes more sharply peaked as the map value cutoff increases (**Fig. 1.2, S1.1**). At lower thresholds, noise flattens the results, with less enrichment for peaks in rotameric regions. Although rotameric regions are sampled more at higher thresholds, fewer residues have local map value peaks above these thresholds, and

noise from counting statistics dominates (**Fig. 1.2**). To quantify the relationship between sample size and rotameric enrichment, we used the normal approximation to the binomial distribution to generate a model-length independent validation statistic: the EMRinger score (**Fig. 1.2, S1.2**). For the 20S proteasome, the EMRinger score is maximized at the 0.242 normalized map value threshold and the signal is dominated by 1547 rotameric map value peaks, compared to 555 non-rotameric peaks (**Fig. S1.3**).

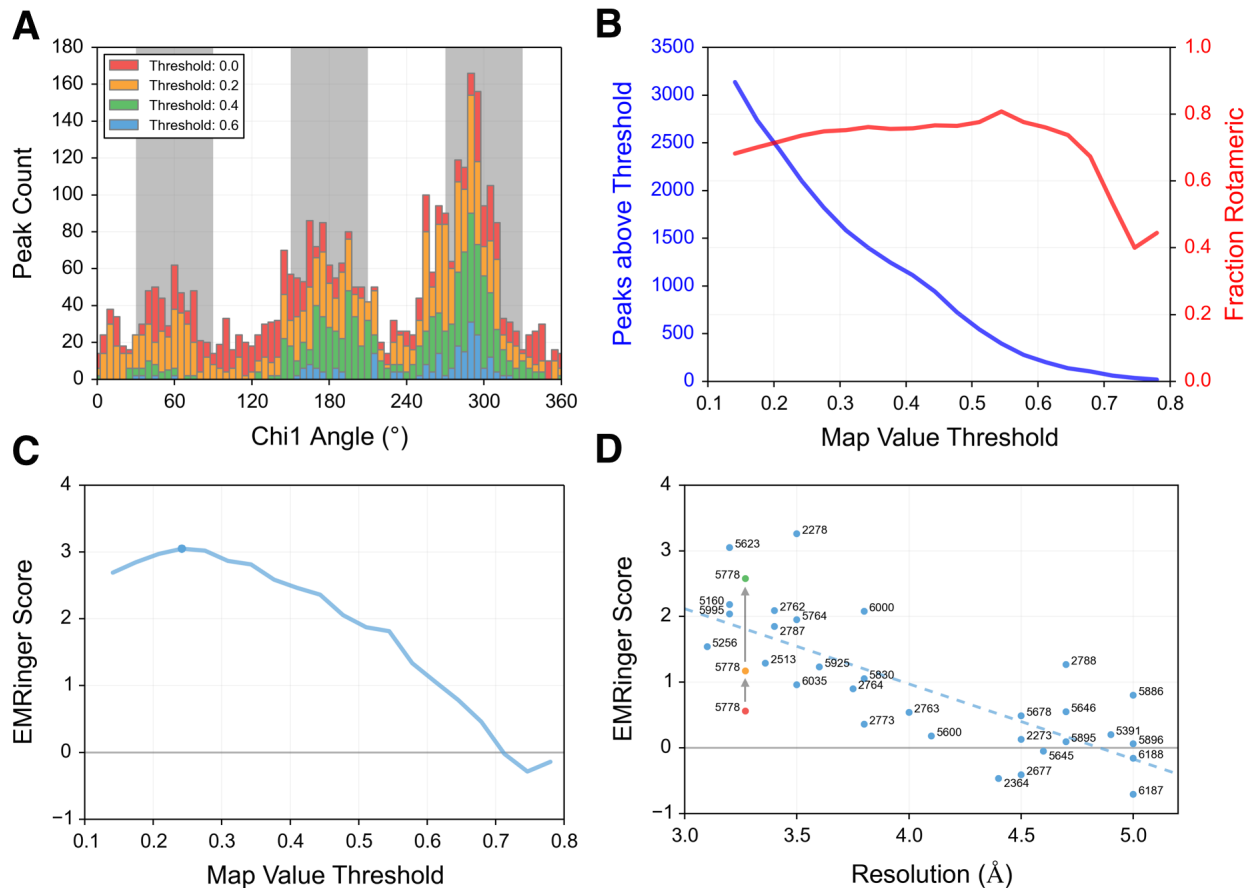


Figure 1.2 | EMRinger reveals statistical enrichment at rotameric χ_1 angles in high resolution EM maps.

(a) Histograms of EMRinger peaks observed above multiple map value thresholds. At high thresholds, more residues are located in the rotameric regions (grey bars). As the threshold lowers, relatively more peaks are added to the non-rotameric regions (white bars). (b) Scanning across map value thresholds demonstrates the tradeoff between sampled peaks (left) and fraction of rotameric peaks (right). (c) The EMRinger score balances the sample size and the rotameric enrichment and is maximized at a threshold of 0.242 for the proteasome structure (blue circle). (d) EMRinger scores for maps deposited in the EMDB with atomic models demonstrate the relationship between model quality and resolution. A linear fit ($R^2 = 0.549$) highlights how refinement of TrpV1 improves from the deposited model (red, PDB 3j5p), the transmembrane domain of the deposited model (orange), and a model refined by RosettaCM (green, PDB 3J9J) (DiMaio *et al*, Nature Methods, In Press).

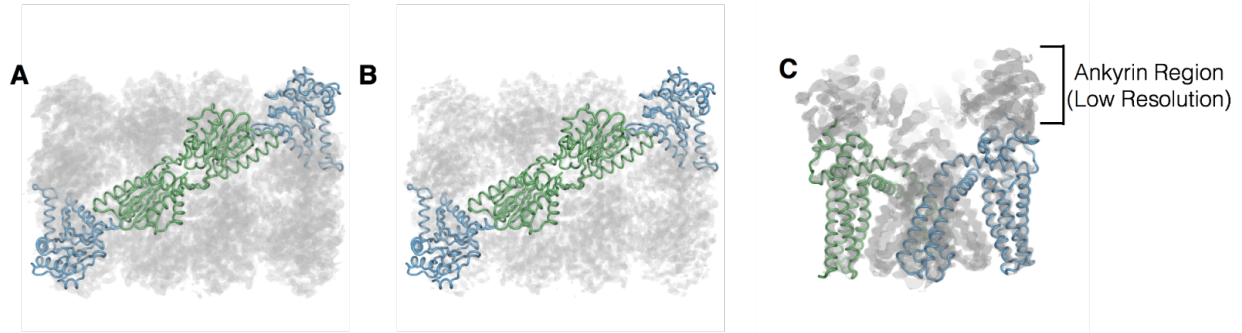


Figure S1.1 | Atomic models in cryo-EM maps.

(a) Two alpha (blue) and two beta (green) subunits of the T20S proteasome are shown as cartoon tubes fitted in a 3.2 Å potential map at isolevel 0.25 (EMDB 5623, PDB 3J9I). **(b)** The same subunits are depicted in density at a higher isolevel of 0.35, where sharper features of side chain density can be observed. **(c)** Two subunits of the TrpV1 tetramer are shown in green and blue in a 3.27 Å potential map at an isolevel of 10 (EMDB 5778, PDB 3J9J)

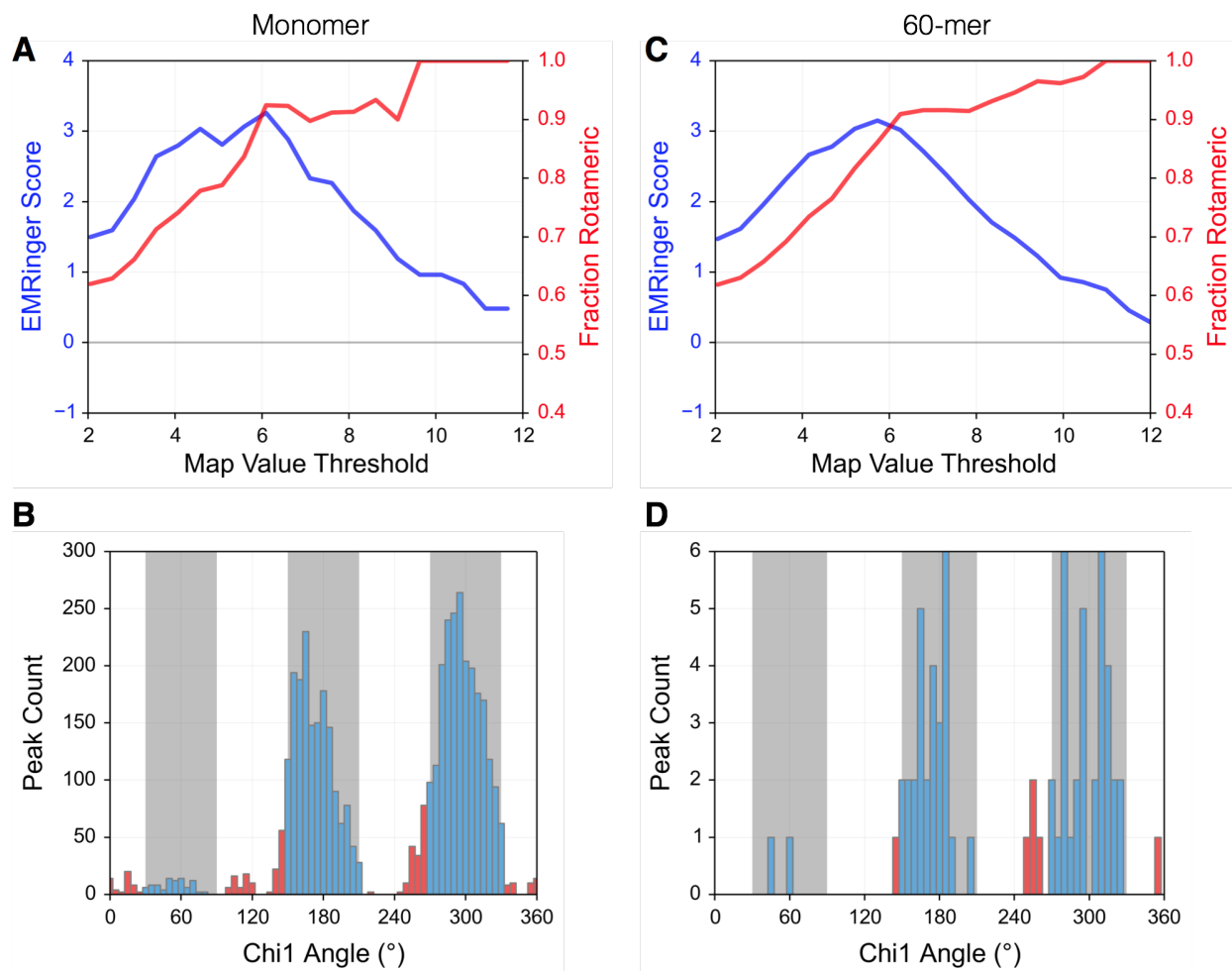


Figure S1.2 | EMRinger Score is unaffected by model size.

(a) EMRinger plot for a 366 amino acid monomer of the Hepatitis B virus capsid gives a peak EMRinger score of 3.25 (EMDB 2278, PDB 3J2V). **(b)** Histogram of EMRinger map value peaks above threshold 6.090 (the threshold of maximum EMRinger score) for the monomer in density. **(c)** EMRinger plot for the full biological 21960 amino acid 60-mer assembly of the Hepatitis B capsid gives a nearly identical set of scores to the monomer, with a peak score of 3.16. The smoother plot is likely due to the averaging out of artifacts due to grid sampling. **(d)** Histogram of EMRinger map value peaks above threshold 5.726 (the threshold of maximum EMRinger score) for the 60-mer in density.

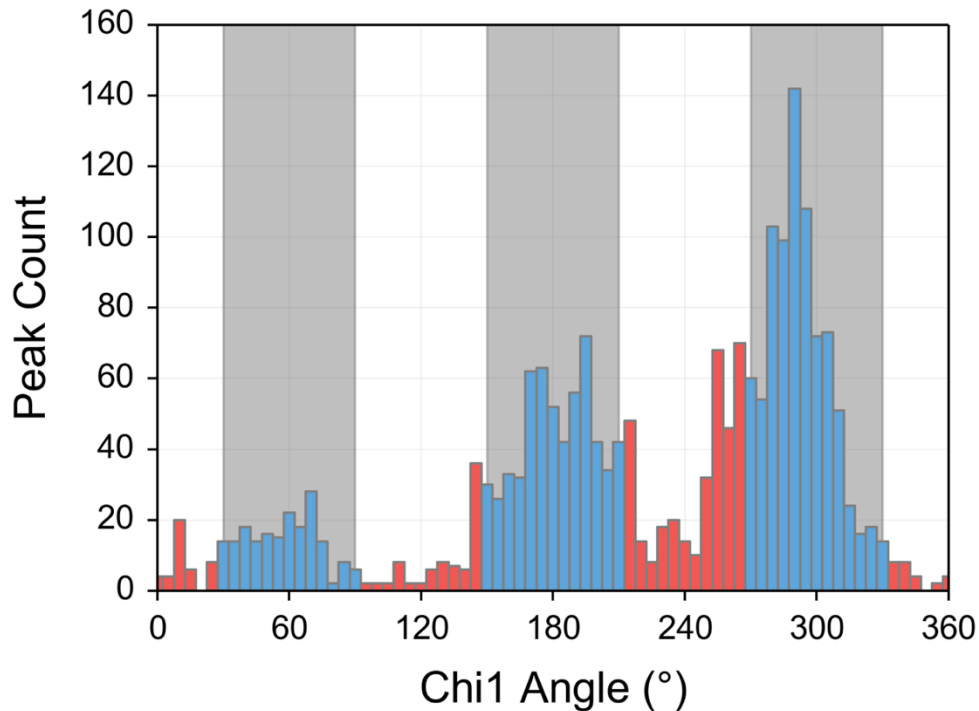


Figure S1.3 | Histogram of peak counts for EMRinger scan of T20S Proteasome (EMDB 5778, PDB 3J9I) at a map value threshold of $0.242 \text{ e}^-/\text{\AA}^3$. At this threshold, which maximizes the EMRinger score, 1547 rotameric peaks (blue) greatly outnumber 555 non-rotameric peaks (red).

Next, we sampled a series of cryo-EM maps deposited in the EMDB, spanning from 3-5 Å resolution, with atomic models built into the map density (**Fig. 1.2, Table S1.1**). The top scoring maps have scores above 3.0: the T20S proteasome, which used a crystallographic model with minimal refinement with MDFF⁶, and the hepatitis B viral capsid, which was built *de novo* and refined using real space refinement in Phenix²⁹. Both maps are consistently better than 3.5 Å local resolution³⁰, likely reflecting the underlying rigidity of the complexes. Recent mammalian ribosome structures^{7,31}, which are dynamic and have more variability in resolution, used masking to reconstruct the highest resolution regions. Refmac reciprocal-space refinement of *de novo* atomic models of these components results in EMRinger scores above 1.85²².

The EMRinger approach confirms the resolution dependence of side chain signals, with a strong correlation between decreasing resolutions and decreasing scores (**Fig. 1.2**). Since a random distribution should produce an EMRinger score of 0, the trend line suggests that the χ^1 angle of side chains can be resolved at 4.5 Å resolution or better. We observed similar trends in decreasing EMRinger score as maps of the T20S proteasome were progressively low-pass filtered (**Fig. S1.4**). These results demonstrate how the EMRinger score quantifies the standard visual check that side chains are resolved in high-resolution maps, providing insight into the quality of the high resolution features of the map and the model.

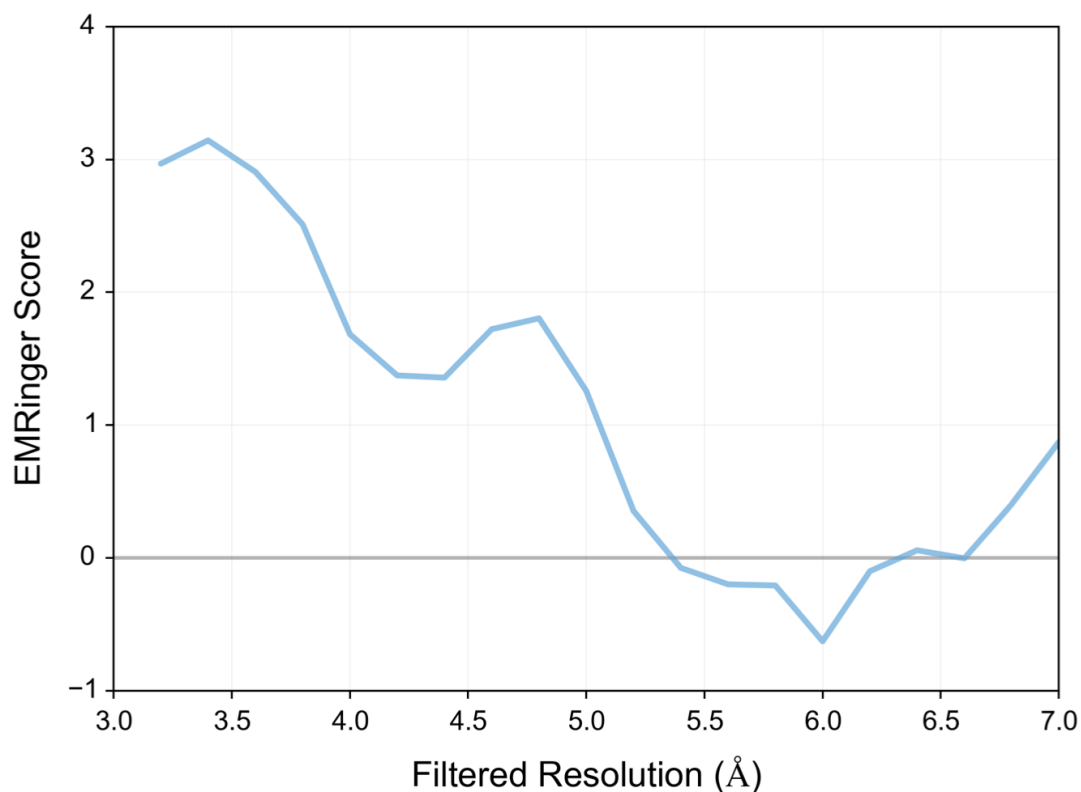


Figure S1.4| Adjusted EMRinger Score degrades rapidly with decreasing resolution.

The T20S proteasome map (EMDB 5623, PDB 1PMA) is low-pass filtered to resolutions ranging from 3.2 to 7 Å. EMRinger scores for each of these filtered maps show a resolution dependence and that by 5 Å resolution side chains are no longer distinguishable from noise and the EMRinger score is near 0.

Table S1.1 | EMRinger analysis of selected maps above 5 Å resolution with atomic models.
 For the transmembrane-only scan of the TrpV1 Channel (EMDB 5778), residues 381-695 of each chain of the deposited model (PDB 3J5P) were used.

EMDB ID	PDB ID	Resolution (Å)	Model Length	EMRinger Score	Description	Year
5256	3IZX	3.1	2427	1.54	Cytoplasmic Polyhedrosis Virus ⁴¹	2012
5995	3J7H	3.2	2616	2.04	Beta-Galactosidase ³³	2014
5160	3IYL	3.2	5708	2.18	Aquareovirus ⁴²	2010
5623	3J9I	3.2	3439	3.05	T20S Proteasome ⁶	2013
5778	3J5P	3.27	1484	0.56	TrpV1 Channel ¹	2014
5778 (TM only)	3J5P	3.27	792	1.17	TrpV1 Channel ¹	2014
5778 (Refined)	3J9J	3.27	876	2.58	TrpV1 Channel ¹	2015
2513	4CIO	3.36	521	1.29	F420 reducing hydrogenase ⁸	2013
2787	4V19, 4V1A	3.4	5326	1.85	Mammalian Mitochondrial Ribosome, Large Subunit ³¹	2014
2762	3J7Y	3.4	4806	2.09	Human Mitochondrial Ribosome Large Subunit ⁷	2014
6035	3J7W	3.5	1267	0.96	Bacteriophage T7 capsid ⁴³	2014
5764	3J4U	3.5	1757	1.95	Bordetella bacteriophage ⁴⁴	2014
2278	3J2V	3.5	366	3.26	Hepatitis B Virus Core ⁴⁵	2013
5925	3J6J	3.6	528	1.23	MAVS filament ⁴⁶	2014
2764	3J80	3.75	3060	0.9	40S-eIF1-eIF1A preinitiation complex ⁴⁷	2014
2773	4UY8	3.8	1976	0.36	TnaC stalled E.coli ribosome ⁴⁸	2014
5830	3J63	3.8	915	1.05	ASC Pyrin Domain ⁴⁹	2014
6000	3J7L	3.8	259	2.08	Brome Mosaic Virus ⁵⁰	2014

EMDB ID	PDB ID	Resolution (Å)	Model Length	EMRinger Score	Description	Year
2763	3J81	4	3225	0.54	Partial Yeast 48S preinitiation complex ⁴⁷	2014
5600	3J3I	4.1	604	0.18	Penicillium Chrysogenum Virus ⁵¹	2014
2364	4BTG	4.4	898	-0.47	Bacteriophage phi procapsid ¹⁰	2013
2677	4UPC	4.5	235	-0.41	Human Gamma-secretase ⁵²	2014
2273	3ZIF	4.5	7430	0.13	Bovine Adenovirus 3 ⁵³	2014
5678	3J40	4.5	1848	0.49	Bacteriophage epsilon15 ¹⁹	2013
5645	3J3X	4.6	4528	-0.05	Mm Chaperonin, Training ⁵⁴	2013
5895	3J6E	4.7	4705	0.09	GMPCPP Microtubule ⁵⁵	2014
5646	3J3X	4.7	4528	0.55	Mm Chaperonin, Testing ⁵⁴	2013
2788	4V1W	4.7	2976	1.27	Horse spleen apoferritin ⁵⁶	2014
5391	3J1B	4.9	4816	0.2	apo rATcpn-alpha ⁵⁷	2013
6187	3J8X	5	737	-0.71	Empty Microtubule/Kinesin ⁵⁸	2014
6188	3J8Y	5	744	-0.16	ADP-AIF3 Microtubule/Kinesin ⁵⁸	2014
5896	3J6F	5	4706	0.06	GDP microtubule ⁵⁵	2014
5886	3J69	5	579	0.8	nanobody/poliovirus ⁵⁹	2014

EMRinger score is highly sensitive to improvements during refinement

A notable exception to the trend of increasing score with higher resolution is TrpV1²⁸ (**Fig. 1.2**), which had a low EMRinger score (0.56) despite high resolution map (3.27 Å). This *de novo* model was built manually and not subjected to either real- or reciprocal-space refinement. Upon exclusion of the poorly resolved ankyrin domain, the Emringer score increases to 1.17, as only the atoms modeled into the highest resolution data remain (**Fig. S1.1, Table S1.1**). This suggests that atomic models may be more appropriate for the high resolution transmembrane region than for the 17nkyrin domain. Further rebuilding and refinement using RosettaCM (DiMaio *et al*, Nature Methods, In Press) gradually improved the EMRinger score in most trials (**Fig. 1.3**). Multiple refinement trajectories led to consistent improvements in EMRinger score from 1.17 to above 1.75. The best RosettaCM trajectory improves the EMRinger score to 2.58, while the validation metrics for an independent reconstruction improve by a small margin (**Fig. 1.3, S1.5, Table 1.1**). In contrast to existing measures, such as real-space correlation or FSC, the EMRinger score is sensitive to features at lower map values, amplifying improvements in the model that only show a minor impact in the agreement-to-density term used by RosettaCM. Consistent with the overlap between the geometrical and conformational components of the Molprobity score and the Rosetta energy function, refinement also improves MolProbity scores dramatically (**Table 1.1**).

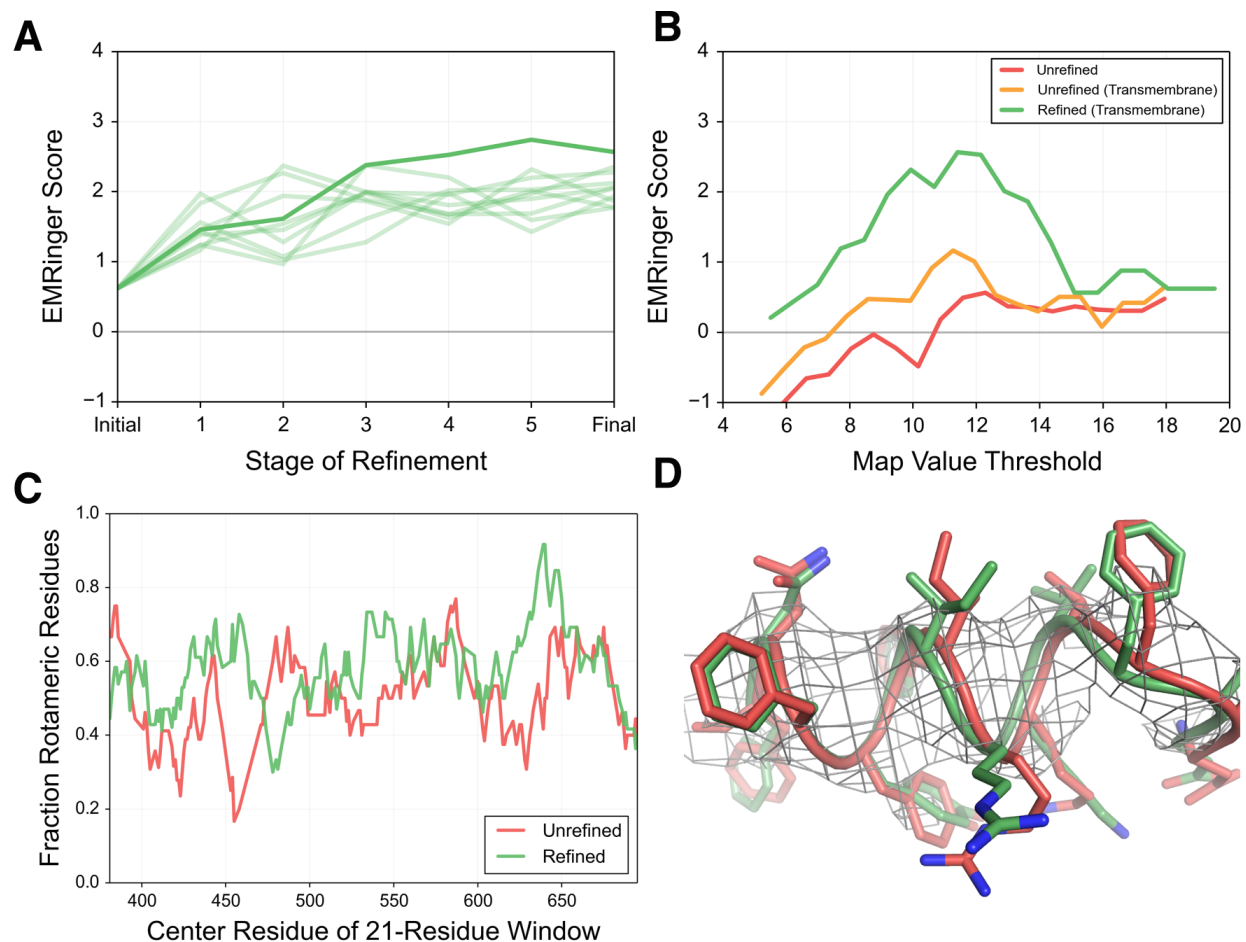


Figure 1.3 | EMRinger Scores report on effective refinement of atomic models into EM maps. **(a)** The EMRinger improves during refinement. RosettaCM (DiMaio *et al*, Nature Methods, In Press) trajectories for 9 trials are shown in light green with the final refinement shown in dark green. **(b)** Map value threshold scan for the unrefined model of TrpV1 (red, EMD5 5778, PDB 3J5P), the transmembrane region of the deposited TrpV1 model (orange), and for the model of TrpV1 refined by RosettaCM (green, PDB 3J9J) show the improvement during refinement. **(c)** Analyzing the unrefined (red) and refined (green) models in the transmembrane region highlights how portions of the model experience dramatic increases in rotameric peaks after refinement. **(d)** The unrefined (red) and refined (green) TrpV1 models are shown in density (isolevel of 10), revealing that small shifts in the placement of backbone of the alpha helix improves EMRinger statistics.

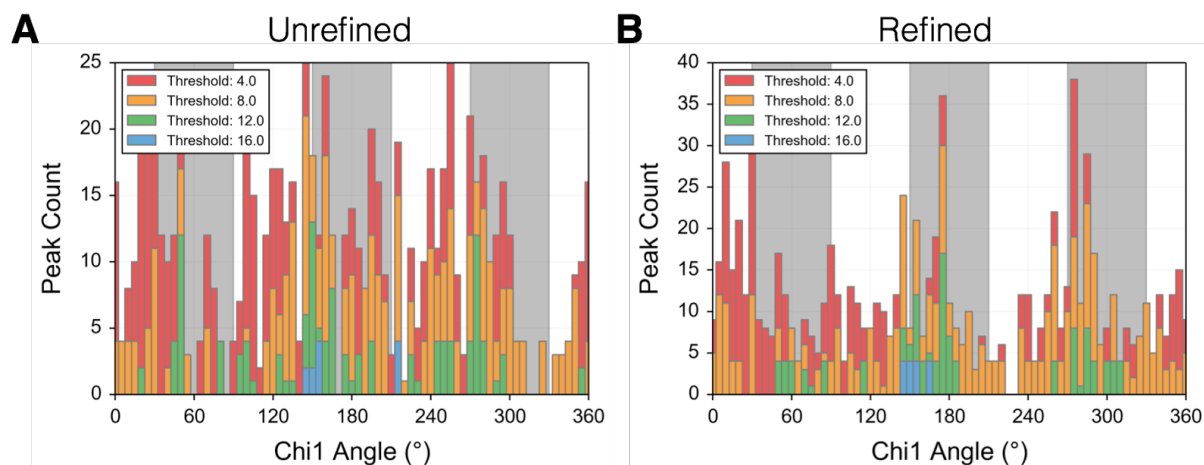


Figure S1.5 | Histograms of TrpV1 models at multiple map value thresholds. **(a)** Histograms at thresholds of 4, 8, 12, and 16 for EMRinger map value peaks of the transmembrane region of the deposited TrpV1 model (EMDB 5778, PDB 3J5P). **(b)** Histograms at thresholds of 4, 8, 12, and 16 for the EMRinger map value peaks of the transmembrane region of TrpV1 refined by RosettaCM show improved enrichment at rotameric positions at all thresholds.

Table 1.1 | Statistics pre- and post-refinement.

Cross correlation, FSC_{mask} , MolProbity scores and EMRinger score are calculated for the full unrefined TrpV1 model (EMDB 5778, PDB 3J5P), the transmembrane domain of the unrefined model, an intermediate model during refinement of the transmembrane region, and the final refined transmembrane region.

	Unrefined	Unrefined (Transmembrane Region)	Refinement Step 2 (Transmembrane Region)	Refinement Final (Transmembrane Region)
CC (3.27 Å Cutoff)	0.676	0.726	.715	0.728
CC (Training Map)	0.663	0.715	0.708	0.718
CC (Testing Map)	0.664	0.714	0.705	0.713
Integrated Model- Map FSC (15-3.4 Å)	0.473	0.553	0.513	0.526
All-atom Clashscore (MolProbity)	77.90	100.78	2.32	2.09
Modelled Rotamer Outliers (MolProbity)	26.6%	30.94%	0.35%	0%
Emringer Score	0.56	1.17	1.61	2.58

To identify the local changes responsible for these improvements, we analyzed 21-residue rolling windows along the length of the protein for the percent of peaks that occurred near rotameric angles (**Fig. 1.3**). The specific effects of the RosettaCM refinement can be seen in small backbone shifts, which move the C-beta atoms so that the peak value moves into a rotameric position (**Fig. 1.3**). These results demonstrate how small corrections of backbone position along secondary structures, introduced through independently-scored refinement procedures, can lead to improvements in EMRinger score and the accuracy of the resulting model.

EMRinger Score reveals the residue-specific effects of radiation damage

Radiation damage can severely limit the ability draw biological conclusions from EM data³². Because the electron beam also induces motion of the sample, the impact of radiation during data collection has been difficult to assess. Recent motion corrected analyses have indicated that high-resolution information degrades as a function of total electron dose, likely as a result of radiation damage⁸, and that the signal in the 5Å shell degrades rapidly in the second half of data collection⁶. In addition to these global metrics, previous work has hypothesized that differential radiation damage causes negatively charged glutamate and aspartate residues to have weaker density than neutral, but similarly shaped, glutamine and asparagine residues^{8,33,34}.

To quantify the effect of radiation damage on the high resolution features of the map and to address whether effects vary by residue type, we used EMRinger for dose-fractionated maps of the T20S proteasome. The overall EMRinger score degrades as a function of dose, with a sharp loss of signal beginning around the 15th frame, corresponding to a total dose of $\sim 18 \text{ e}^-/\text{\AA}^2$

(**Fig. 1.4**). Next, we performed EMRinger analysis on different subsets of amino acids. Amino acids with charged side chains generally lost signal more quickly as a function of dose than average, whereas aromatic residues were much more resistant to degradation (**Fig. 1.4**). Most notably, negatively charged side-chains appeared to lose signal much faster than positively charged side-chains, with EMRinger score dropping to zero by the map centered on the 8th frame.

The divergent results of EMRinger analysis of negatively charged side chains may be in part explained by the differential radiation damage effects that have been previously hypothesized. However, since a map comprised only of noise (in the extreme of radiation damage) should result in a score of zero, this effect is not sufficient to explain negative EMRinger scores observed in later frames. We examined the specific behavior of the negatively charged residues and observed that the initial map value peaks for some negatively charged residues inverted and became a local minimum in later frames (**Fig. 1.4**). This behavior is in contrast to the flattening effect, where a peak slowly degrades into noise, seen generally for other residue types (**Fig. 1.4**). The inversion of the peak may result from the electron scattering factors of negatively charged oxygen atoms, which are positive at high resolution but become negative at low resolution³⁵. This radiation damage effect would lead to a negative scattering contribution near the true (rotameric) position in subsequent maps. Because the rotameric peak of the original map can therefore be lowered below the baseline, EMRinger will then identify a new peak at a different local maximum in the damaged map. This new local maximum is more likely to occur at non-rotameric angles because the original rotameric angle is now suppressed by negative scattering contributions in the damaged map. The net effect of the negative scattering behavior

could therefore result in an enrichment of peaks at non-rotameric positions and, consequently, a negative EMRinger score after significant radiation damage has accumulated.

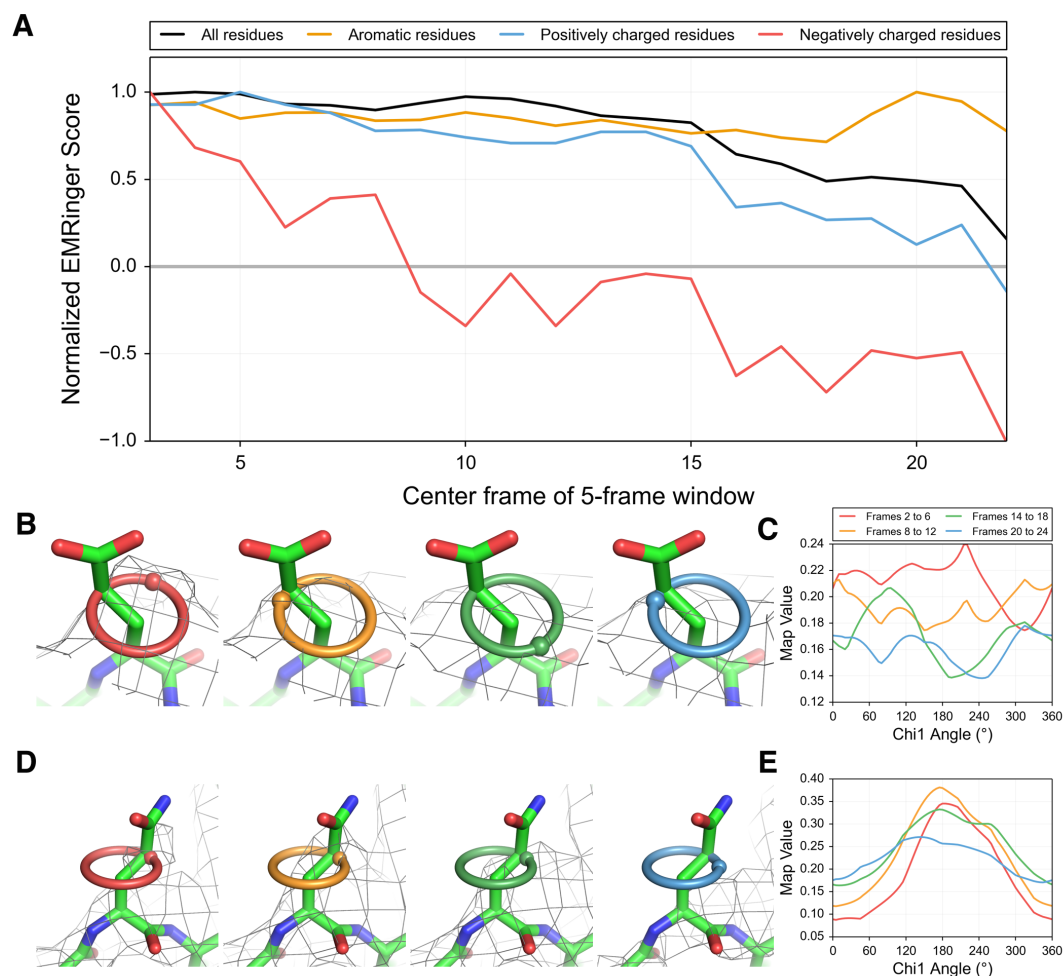


Figure 1.4 | Acidic residues are differentially altered by radiation damage.

(a) Normalized EMRinger scores are plotted for the T20S proteasome model (PDB: 3J9I) against maps calculated from 5 frames of data. Scores for the entire model (black), the aromatic residues (orange), and the positively charged residues (blue) slowly decrease as a function of dose. In contrast, negatively charged residues (red) experience a rapid drop and fall below a random score of 0. (b) Proteasome chain D residue Glu 99 shown in density (isolevel 0.18) for maps generated from frames 2-6 (red ring), 8-12 (orange ring), 14-18 (green ring), and 20-24 (blue ring), with spheres showing local map value peaks. (c) EMRinger plots for Glu 99 of Chain D corresponding to the maps in **b** show that peaks immediately flatten and eventually invert after high dose has accumulated. (d) Proteasome chain 1 residue Gln 36 shown in density (isolevel 0.3) as in **b**. (e) EMRinger plots corresponding to the maps in **d** show a gradual loss of signal as a function of dose.

Discussion

The dramatic advances in electron cryomicroscopy have created new challenges in building, refining, and validating atomic models. EMRinger extends and complements existing cryo-EM validation procedures at multiple levels. For example, the idea that high resolution features are detectable, confirming the resolution estimate, is quantified by the side chain enrichment. Moreover, the enrichment score tests the fine features of the side chain map density, which intersects with validating the physical correctness of the modeled backbone. While current methods test conformational features independently of agreement with the map, the EMRinger tests these features by querying the model and map together. This procedure is responsive to small backbone corrections that increase the accuracy of the model and the ability to draw mechanistic insights from it.

Our work confirms that side chain detail can be resolved in these maps by quantifying the statistical enrichment of map value peaks at rotameric positions of side chains. Although our analysis was restricted to χ_1 angles, similar statistical signatures may extend further out along many side chains. These statistical signatures, which are present in maps determined without model-biased phasing, are a strong indicator that the side chain density that has been identified is predominantly signal rather than noise. Our results confirm that recent advances in data collection, processing, and refinement are increasing the resolvability of atomic features and provide a new metric for assessing the reliability of atomic models generated *de novo* from high resolution cryo-EM maps.

Whereas model-to-map agreement metrics are normally dominated by low resolution features, the EMRinger score reports specifically on statistical signatures in high-resolution data. To validate the model-to-map correctness of atomic models from cryo-EM, refinement should

result in EM Ringer scores above 1.0 for well-refined structures with maps in the 3-4 Å range. EMRinger scores can be used in concert with cross validation procedures²¹ and with other measures, such as gold-standard FSC-based resolution¹³ and Molprobit statistics¹⁶. While it is unlikely that maps with highly variable resolution, generated by imaging more dynamic proteins, will display as much rotameric enrichment as more static molecules, successes in classification of images into different maps representing distinct biochemical states³⁶ should be accompanied by increases in EMRinger scores. Similarly, EMRinger scores should quantify improvements in resolvability of atomic features due to improved motion correction algorithms or improved balance between dose and radiation damage during data collection. The results of the EMRinger analysis on dose-fractionated data suggest that reconstructions based on different doses may be required to maximize the resolvability of different sets of side chains, just as different degrees of sharpening are commonly used now during model building.

Additionally, the high sensitivity of EMRinger suggests a natural direction for model-building and refinement. At the resolutions commonly used for model building in EM, there are many closely related backbone conformations that can fit the map density with nearly equal agreement. Given a nearly finalized backbone position, side chains with non-rotameric peaks can be adjusted to fix the C γ atom in the peak density. Subsequently, the backbone conformation and closure to adjacent residues can be optimized to maintain a rotameric side chain conformation, similar to the inverse rotamer approach used in some protein design applications³⁷. This procedure could, in principle, be iterated many times to converge on backbones that are consistent with the map and satisfy the rotameric peak constraints exploited by EMRinger. Similar approaches to quantifying statistical signatures in weakly resolved data may also prove

helpful for modeling of non-amino-acid structures at lower resolutions, including glycans and nucleic acids^{38,39}.

Methods

All scripts can be found at <https://github.com/fraser-lab/EMRinger> and can be run using Phenix/cctbx python (version numbers greater than 1894).

Map Values

We loaded CCP4 formatted maps using cctbx and used the map voxel values without further normalization. The wide range of normalization procedures used in constructing these maps explains the large differences in threshold values used for different model-map pairs in our study. However, because EMRinger calculations are based on the relative values of a single map, we can compare EMRinger scores between maps without further normalization.

EMRinger Map Analysis

EMRinger, as implemented in the Phenix software package⁴⁰, is an extension of the Ringer protocol developed previously^{24,25}. We adapted EMRinger to work with real-space maps and to rotate the C γ atom by increments of 5° around the χ_1 dihedral angle (starting at 0° relative to the amide nitrogen). EMRinger calculates and records the map value from a potential map at the position of the atom at each increment using the eight-point interpolation function supplied by Phenix. From this scan, EMRinger records the peak map value and the angle at which it is achieved. EMRinger is available as the *emringer.py* script. Real space correlation coefficients were measured by the *em_rsc.py* script.

EMRinger Score for Validation

We sampled all non- γ -branched, non-proline amino acids with a non-H γ atom, and measured the percent of map value peaks above a given noise-cutoff threshold that are near rotameric (60° , 180° , or 300°) positions. To determine the significance of this distribution, we calculated a Z-Score based on a normal approximation to the binomial distribution. EMRinger repeats this process across a range of map value thresholds, ranging from the minimum peak map value in any scan to the maximum, and returns the highest Z-score calculated in this range.

(Equation 1) In order to compare Z-scores between models of different structures, the Z-score is rescaled to the “EMRinger Score” to account for the total number of amino acids in the model (Equation 2).

$$(1) \quad Z\text{-score}_{threshold} = \frac{\text{Number Rotameric} - \frac{39}{72} \cdot \text{Number Above Threshold}}{\sqrt{\frac{39}{72} \left(1 - \frac{39}{72}\right) \cdot \text{Number Above Threshold}}}$$

$$(2) \quad \text{EMRinger Score}_{threshold} = \frac{10 \cdot Z\text{-score}_{threshold}}{\sqrt{\text{Model Length}}}$$

Adjusted EMRinger score does not change when the model and map are multiplied (e.g. in the case of a polymer with high symmetry), so that the score is definitive and no issues arise of how many monomers should be included in the analysis. An EMRinger score of 1.0 sets an initial quality goal for a model refined against a map in the 3.2-3.5Å range, while very high quality models at high resolution generate scores above 2.0. Maps that are highly variable in resolution may have lower EMRinger Scores unless poorly resolved regions of the map are masked out and excluded from the model. Calculation of the EMRinger score is accomplished by the *emringer_score.py* script. Rolling window EMRinger analysis is accomplished by the *emringer_rolling.py* script.

Refinement of TrpV1 with RosettaCM

Refinement of TRPV1 used an iterative local rebuilding procedure to improve local backbone geometry as well as fit to the experimental density data (DiMaio *et al*, Nature Methods, In Press). Refinement began with the deposited PDB structure of TRPV1 (PDB 3J5P). The model was trimmed to the transmembrane region (residues 381-695), and bond angles and bond lengths were given ideal geometry. During local rebuilding, 5 cycles of backbone rebuilding were run; in each cycle, regions with poor fit to density or poor local geometry were automatically identified, and rebuilding focused on these regions. Each rebuilding cycle was followed by side chain rotamer optimization and all-atom refinement with a physically realistic force field. Following this protocol, 1000 independent trajectories were run, and the final model was selected by filtering on two criteria: first, the 80 most nonphysical models were eliminated by assessing each model against the Rosetta all-atom force field; second, fit-to-density was used to rank models and select the best model from these 10.

Table Statistics

The cross-correlation was calculated using Chimera's "Fit in Map" tool across all contours and using a resolution cutoff for the calculated map. The integrated FSC was calculated between the model and an independent reconstruction over a masked region covering the protein only. The mask was truncated at 6 Å resolution, and we report the integrated FSC_{mask} over high-resolutions shells only (15 – ~3.4 Å). Molprobit statistics were calculated using the validate tool in Phenix nightly build 1894.

Radiation Damage Analysis

To identify the degradation of map signal with radiation damage, we used EMRinger with a single model across multiple dose-fractionated maps. For each dose-fractionated map, the EMRinger Score is calculated for the model. We calculated additional scores with the amino acids being sampled restricted to different classes (such as acidic or aromatic residues). This residue-specific sampling is accomplished by the *emringer_residue.py* script.

References

- 1 Liao, M., Cao, E., Julius, D. & Cheng, Y. Single particle electron cryo-microscopy of a mammalian ion channel. *Current opinion in structural biology* **27**, 1-7, doi:10.1016/j.sbi.2014.02.005 (2014).
- 2 Bai, X. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends in biochemical sciences* **40**, 49-57, doi:10.1016/j.tibs.2014.10.005 (2015).
- 3 Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *Journal of structural biology* **183**, 377-388, doi:10.1016/j.jsb.2013.07.005 (2013).
- 4 Scheres, S. H. Semi-automated selection of cryo-EM particles in RELION-1.3. *Journal of structural biology*, doi:10.1016/j.jsb.2014.11.010 (2014).
- 5 Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665, doi:10.7554/eLife.03665 (2014).
- 6 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature methods* **10**, 584-590, doi:10.1038/nmeth.2472 (2013).
- 7 Brown, A. *et al.* Structure of the large ribosomal subunit from human mitochondria. *Science* **346**, 718-722, doi:10.1126/science.1258026 (2014).
- 8 Allegretti, M., Mills, D. J., McMullan, G., Kuhlbrandt, W. & Vonck, J. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife* **3**, e01963, doi:10.7554/eLife.01963 (2014).

- 9 Yan, Z. *et al.* Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. *Nature* **517**, 50-55, doi:10.1038/nature14063 (2015).
- 10 Lu, P. *et al.* Three-dimensional structure of human gamma-secretase. *Nature* **512**, 166-170, doi:10.1038/nature13567 (2014).
- 11 Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159**, 995-1014, doi:10.1016/j.cell.2014.10.051 (2014).
- 12 Wong, W. *et al.* Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. *eLife* **3**, doi:10.7554/eLife.03080 (2014).
- 13 Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205-214, doi:10.1016/j.str.2011.12.014 (2012).
- 14 Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24-35, doi:10.1016/j.ultramic.2013.06.004 (2013).
- 15 Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature methods* **9**, 853-854, doi:10.1038/nmeth.2115 (2012).
- 16 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* **66**, 12-21, doi:10.1107/S09074444909042073 (2010).
- 17 Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *Journal of molecular biology* **7**, 95-99 (1963).
- 18 Lovell, S. C., Word, J. M., Richardson, J. S. & Richardson, D. C. The penultimate rotamer library. *Proteins* **40**, 389-408 (2000).

- 19 DiMaio, F., Zhang, J., Chiu, W. & Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein science : a publication of the Protein Society* **22**, 865-868, doi:10.1002/pro.2267 (2013).
- 20 Amunts, A. *et al.* Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485-1489, doi:10.1126/science.1249410 (2014).
- 21 Falkner, B. & Schroder, G. F. Cross-validation in cryo-EM-based structural modeling. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8930-8935, doi:10.1073/pnas.1119041110 (2013).
- 22 Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallographica Section D* **71**, 136-153, doi:doi:10.1107/S1399004714021683 (2015).
- 23 Dunbrack, R. L., Jr. Rotamer libraries in the 21st century. *Current opinion in structural biology* **12**, 431-440 (2002).
- 24 Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 237-242, doi:10.1073/pnas.1302823110 (2014).
- 25 Lang, P. T. *et al.* Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein science : a publication of the Protein Society* **19**, 1420-1431, doi:10.1002/pro.423 (2010).
- 26 Zhou, A. Q., O'Hern, C. S. & Regan, L. Predicting the side-chain dihedral angle distributions of nonpolar, aromatic, and polar amino acids using hard sphere models. *Proteins* **82**, 2574-2584, doi:10.1002/prot.24621 (2014).

- 27 Shapovalov, M. V. & Dunbrack, R. L., Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19**, 844-858, doi:10.1016/j.str.2011.03.019 (2011).
- 28 Liao, M., Cao, E., Julius, D. & Cheng, Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* **504**, 107-112, doi:10.1038/nature12822 (2013).
- 29 Yu, X., Jin, L., Jih, J., Shih, C. & Zhou, Z. H. 3.5Å cryoEM structure of hepatitis B virus core assembled from full-length core protein. *PloS one* **8**, e69729, doi:10.1371/journal.pone.0069729 (2013).
- 30 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature methods* **11**, 63-65, doi:10.1038/nmeth.2727 (2014).
- 31 Greber, B. J. *et al.* The complete structure of the large subunit of the mammalian mitochondrial ribosome. *Nature* **515**, 283-286, doi:10.1038/nature13895 (2014).
- 32 Glaeser, R. M. Limitations to significant information in biological electron microscopy as a result of radiation damage. *Journal of ultrastructure research* **36**, 466-482 (1971).
- 33 Bartesaghi, A., Matthies, D., Banerjee, S., Merk, A. & Subramaniam, S. Structure of beta-galactosidase at 3.2-Å resolution obtained by cryo-electron microscopy. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 11709-11714, doi:10.1073/pnas.1402809111 (2014).
- 34 Fioravanti, E., Vellieux, F. M., Amara, P., Madern, D. & Weik, M. Specific radiation damage to acidic residues and its relation to their chemical and structural environment. *Journal of synchrotron radiation* **14**, 84-91, doi:10.1107/S0909049506038623 (2007).

- 35 Mitsuoka, K. *et al.* The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution. *Journal of molecular biology* **286**, 861-882, doi:10.1006/jmbi.1998.2529 (1999).
- 36 Fernandez, I. S. *et al.* Molecular architecture of a eukaryotic translational initiation complex. *Science* **342**, 1240585, doi:10.1126/science.1240585 (2013).
- 37 Havranek, J. J. & Baker, D. Motif-directed flexible backbone design of functional interactions. *Protein science : a publication of the Protein Society* **18**, 1293-1305, doi:10.1002/pro.142 (2009).
- 38 Cowtan, K. Automated nucleic acid chain tracing in real time. *IUCrJ* **1**, 387-392, doi:10.1107/S2052252514019290 (2014).
- 39 Terwilliger, T. C. Rapid model building of alpha-helices in electron-density maps. *Acta crystallographica. Section D, Biological crystallography* **66**, 268-275, doi:10.1107/S0907444910000314 (2010).
- 40 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
- 41 Yu, X., Ge, P., Jiang, J., Atanasov, I. & Zhou, Z. H. Atomic model of CPV reveals the mechanism used by this single-shelled virus to economically carry out functions conserved in multishelled reoviruses. *Structure* **19**, 652-661, doi:10.1016/j.str.2011.03.003 (2011).
- 42 Zhang, X., Jin, L., Fang, Q., Hui, W. H. & Zhou, Z. H. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* **141**, 472-482, doi:10.1016/j.cell.2010.03.041 (2010).

- 43 Guo, F. *et al.* Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E4606-4614, doi:10.1073/pnas.1407020111 (2014).
- 44 Zhang, X. *et al.* A new topology of the HK97-like fold revealed in Bordetella bacteriophage by cryoEM at 3.5 Å resolution. *eLife* **2**, e01299, doi:10.7554/eLife.01299 (2013).
- 45 Yu, X., Jin, L., Jih, J., Shih, C. & Zhou, Z. H. 3.5Å cryoEM structure of hepatitis B virus core assembled from full-length core protein. *PloS one* **8**, e69729, doi:10.1371/journal.pone.0069729 (2013).
- 46 Wu, B. *et al.* Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Mol Cell* **55(4)**, 511-23, doi: 10.1016/j.molcel.2014.06.010 (2014)
- 47 Hussain, T. *et al.* Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell* **159**, 597-607, doi:10.1016/j.cell.2014.10.001 (2014).
- 48 Bischoff, L., Berninghausen, O. & Beckmann, R. Molecular basis for the ribosome functioning as an L-tryptophan sensor. *Cell reports* **9**, 469-475, doi:10.1016/j.celrep.2014.09.011 (2014).
- 49 Lu, A. *et al.* Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes. *Cell* **156**, 1193-1206, doi:10.1016/j.cell.2014.02.008 (2014).
- 50 Wang, Z. *et al.* An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nature communications* **5**, 4808, doi:10.1038/ncomms5808 (2014).

- 51 Luque, D. *et al.* Cryo-EM near-atomic structure of a dsRNA fungal virus shows ancient structural motifs preserved in the dsRNA viral lineage. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 7641-7646, doi:10.1073/pnas.1404330111 (2014).
- 52 Nemecek, D. *et al.* Subunit folds and maturation pathway of a dsRNA virus capsid. *Structure* **21**, 1374-1383, doi:10.1016/j.str.2013.06.007 (2013).
- 53 Cheng, L. *et al.* Cryo-EM structures of two bovine adenovirus type 3 intermediates. *Virology* **450-451**, 174-181, doi:10.1016/j.virol.2013.12.012 (2014).
- 54 Baker, M. L. *et al.* Validated near-atomic resolution structure of bacteriophage epsilon15 derived from cryo-EM and modeling. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 12301-12306, doi:10.1073/pnas.1309947110 (2013).
- 55 Alushin, G. M. *et al.* High-resolution microtubule structures reveal the structural transitions in alphabeta-tubulin upon GTP hydrolysis. *Cell* **157**, 1117-1129, doi:10.1016/j.cell.2014.03.053 (2014).
- 56 Russo, C. J. & Passmore, L. A. Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377-1380, doi:10.1126/science.1259530 (2014).
- 57 Zhang, K. *et al.* Flexible interwoven termini determine the thermal stability of thermosomes. *Protein & cell* **4**, 432-444, doi:10.1007/s13238-013-3026-9 (2013).
- 58 Shang, Z. *et al.* High-resolution structures of kinesin on microtubules provide a basis for nucleotide-gated force-generation. *eLife* **3**, doi:10.7554/eLife.04686 (2014).
- 59 Schotte, L. *et al.* Mechanism of action and capsid-stabilizing properties of VHHs with an in vitro antipoliioviral activity. *Journal of virology* **88**, 4403-4413, doi:10.1128/JVI.03402-13 (2014).

Chapter II

Temperature-Jump Solution X-ray Scattering Reveals Distinct Motions in a Dynamic Enzyme

Michael C. Thompson¹, Benjamin A. Barad^{1,2}, Alexander M. Wolff^{1,2}, Hyun Sun Cho³, Friedrich Schotte³, Daniel M.C. Schwarz^{1,4}, Philip Anfinrud^{3,#}, James S. Fraser^{1,*}

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

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

³Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 220892-0520, USA

⁴Chemistry and Chemical Biology Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

- anfinrud@nih.gov

* - jfraser@fraserlab.com

Preface

The bulk of this chapter appears as Thompson *et al.* preprinted in *bioRxiv* in 2018.

This project was a very challenging experiment and a successful collaboration both within the lab and with the Anfinrud lab at NIH, led by Michael Thompson. He worked with the Anfinrud lab to develop an approach for very rapid heating of protein solutions using an infrared laser (T-jumps), followed by capturing solution small and wide angle x-ray scattering curves (SAXS/WAXS) after short time delays. This process is similar to the time-resolved pump-probe x-ray scattering experiments that the Anfinrud lab and others have pursued previously. T-jumps have major advantages over traditional pump-probe techniques because all proteins respond to

temperature, while only a small proportion of proteins are responsive to visible light. T-jumps have been used previously in combination with spectroscopy as a method of studying protein folding. The novelty of this approach came from the application of the T-jump to folded proteins, in order to quickly change the protein's conformational equilibrium so that its relaxation to the new equilibrium can be observed and the different dynamic processes can be kinetically characterized. In this paper, we use this novel method to study the model protein cyclophilin A, which has multiple characterized dynamic modes that are associated with catalytic activity and regulation. Most importantly, we were able to study well-characterized mutants that disrupt individual dynamic modes, and in this way associate the kinetics we observed from the SAXS/WAXS curves with their respective structural dynamics.

My role in this project was primarily data analysis – the experiment generated large amounts of low resolution data that encoded the scattering of both protein sample and solvent. We collected interleaved pre- and post-T-jump SAXS/WAXS curves in order to isolate only the effects of temperature from a complex background of radiation damage, x-ray beam intensity, and drifting experiment geometry. We collected data for up to 28 time delays, and collected 50 repeats for each delay in order to improve our signal to noise. It was essential to determine at the beamline to what degree we were observing real dynamics, as well as to quantify the kinetics we observed as carefully as possible after the fact. The data encodes temperature response from both the protein and the solvent, and in order to isolate those signals I developed approaches to scale the SAXS/WAXS curves to account for fluctuating beam intensity, subtract on-off pairs, automatically report and remove outliers, and average data to maximize the signal-to-noise ratio of the time resolved difference data. I also developed the tools to integrate the SAXS/WAXS differences over the region that encoded the time resolved changes in the protein and developed

a library to automatically fit non-linear relaxation curves to this data to as precisely as possible quantify the kinetics of the two major processes we observed. The solvent information encodes temperature, since the liquid structure of water is responsive to temperature, and I created a precise thermometer using singular value decomposition of the solvent regime of the SAXS/WAXS curve that allowed us to determine the absolute temperature of the solution before and after temperature jumps, which allowed us to associate the relaxation kinetics observed at different starting temperatures with temperature to extract the activation entropy and enthalpy of the transition. Combining this work with other SAXS/WAXS analysis developed by Alex Wolff, we made this toolkit as straightforward to use as possible and made it open source, representing the first open source and freely available tool for processing of time resolved SAXS/WAXS data.

Abstract

Correlated motions of proteins and their bound solvent molecules are critical to function, but these features are difficult to resolve using traditional structure determination techniques. Time-resolved X-ray methods hold promise for addressing this challenge but have relied on the exploitation of exotic protein photoactivity, and are therefore not generalizable. Temperature-jumps (T-jumps), through thermal excitation of the solvent, have been utilized to study protein dynamics using spectroscopic techniques, but their implementation in X-ray scattering experiments has been limited. Here, we perform T-jump small- and wide-angle X-ray scattering (SAXS/WAXS) measurements on a dynamic enzyme, cyclophilin A (CypA), demonstrating that these experiments are able to capture functional intramolecular protein dynamics on the microsecond timescale. We show that CypA displays rich dynamics following a T-jump, and use the resulting time-resolved signal to assess the kinetics of conformational changes in the enzyme.

Two relaxation processes are resolved, which can be characterized by Arrhenius behavior. We also used mutations that have distinct functional effects to disentangle the relationship of the observed relaxation processes. A fast process is related to surface loop motions important for substrate specificity, whereas a slower process is related to motions in the core of the protein that are critical for catalytic turnover. These results demonstrate the power of time-resolved X-ray scattering experiments for characterizing protein and solvent dynamics on the μ s-ms timescale. We expect the T-jump methodology presented here will be useful for understanding kinetic correlations between local conformational changes of proteins and their bound solvent molecules, which are poorly explained by the results of traditional, static measurements of molecular structure.

Introduction

Protein motions are critical for functions such as enzyme catalysis and allosteric signal transduction¹, but it remains challenging to study excursions away from the most populated conformations². Traditional methods that utilize X-rays for structural characterization of biological macromolecules, such as crystallography and solution scattering, provide high-quality structural information, but this information is both spatially and temporally averaged because the measurements are performed on large ensembles of molecules and are typically slower than the timescales of molecular motion^{2,3}. To some extent, the spatial averaging inherent to X-ray experiments is advantageous, because it reveals the alternative local conformations of a molecule that are significantly populated at equilibrium; however, structural states that are not significantly populated at equilibrium, such as intermediates along a conformational transition pathway, are effectively invisible. The temporal averaging inherent to X-ray experiments also

results in a loss of information about how transitions between local alternative conformational states are coupled to one another. To gain kinetic information about molecular motion, researchers often turn to spectroscopic methods, but it can be difficult to correlate spectroscopic observables with high resolution structural models.

Time-resolved X-ray scattering and diffraction can overcome the limitations of traditional structure determination for studying the dynamics of biomolecules⁴⁻⁷. In these experiments, a fast perturbation is applied to the sample to remove it from conformational equilibrium and synchronize conformational changes in a significant fraction of the molecules. Ultrafast X-ray pulses, which are short relative to motions of interest, are then used to perform structural measurements in real time as the system relaxes to a new equilibrium, providing simultaneous structural and kinetic information at high spatial and temporal resolution. Time-resolved X-ray experiments can identify transiently-populated structural states along a conformational transition pathway, and reveal kinetic couplings between conformations⁸. Despite this potential to provide a wealth of information, especially when combined with molecular dynamics simulation⁹⁻¹², time-resolved experiments have not been broadly applied by structural biologists. To date, systems that have been most rigorously studied are those in which a protein conformational change is coupled to excitation of a photoactive ligand molecule, because the conformational change can be initiated with an ultrafast optical laser pulse (e.g. ¹³⁻¹⁸). Unfortunately, the number of proteins that undergo specific photochemistry as part of their functional cycle is small, and there is a fundamental need to develop generalized methods that can be used to synchronously excite conformational transitions in any protein molecule and expand the utility of time-resolved structural experiments¹⁹.

Protein structural dynamics are intimately coupled to the thermal fluctuation of the surrounding solvent (“solvent slaving”^{20,21}), and thermal excitation of the solvent by infrared (IR) laser temperature-jump has been used in numerous pump-probe experiments. These experiments work on the principle that absorption of IR photons excites the O-H stretching modes of water molecules, and the increased vibrational energy is dissipated through increased rotation and translation of the solvent molecules, effectively converting electromagnetic energy into kinetic (thermal) energy. Because this process of solvent heating and subsequent heat transfer to the protein is much faster⁹ than the large-scale molecular motions that define protein conformational changes, the sudden T-jump removes conformational ensembles of protein molecules from their thermal equilibrium so that their structural dynamics can be measured using relaxation methods (**Figure 2.1**). For example, T-jump perturbations have been coupled to ultrafast spectroscopic methods, including Fourier-transform infrared (FTIR) spectroscopy^{22,23}, nuclear magnetic resonance (NMR)^{24–26}, and various forms of fluorescence spectroscopy^{27,28}, for the study of protein folding and enzyme dynamics. While these methods provide detailed kinetic information, they yield only very limited structural information about the underlying atomic ensemble. In contrast, the application of T-jumps to time-resolved X-ray scattering and diffraction has been very limited. Nearly two decades ago, Hori, et al used temperature-jump Laue crystallography to study the initial unfolding step of 3-isopropylmalate dehydrogenase²⁹. That study explored only a single pump-probe time delay, which allowed them to observe laser-induced structural changes but precluded kinetic analysis. Within the last two years, the laser T-jump method has been paired with X-ray solution scattering to explore the oligomerization of insulin in non-physiological conditions^{30,31} and hemoglobin³². The results and analysis we present here expand

the role of the T-jump method in structural biology, by demonstrating that T-jump X-ray scattering experiments can be used as a general method to explore the functional, internal dynamics of proteins under solution conditions. Additionally, we provide a detailed outline of a data reduction and analysis procedure suitable for T-jump SAXS/WAXS experiments.

The T-jump SAXS/WAXS experiments we describe here used human cyclophilin A (CypA), a proline isomerase enzyme that functions as a protein folding chaperone and as a modulator of intracellular signaling pathways. CypA has been the subject of many NMR experiments that have identified two primary dynamic features of interest (**Figure 2.1**). First, the active site-adjacent loops (covering approximately residues 60-80 and hereafter referred to as the “loops” region)) are mobile on a ms-timescale³³. This region is especially interesting because evolutionarily selected mutations along these loops perturb the dynamics of the loop³⁴, alter the binding specificity of CypA for substrates such as HIV capsids³⁵, and restrict the host range of these viruses^{36,37}. Second, a group of residues that extends from the active site into the core of the protein has also been shown to be mobile on a ms-timescale³³. Subsequent work incorporating multi-temperature X-ray crystallography³⁸, mutagenesis³⁹, and further NMR experiments⁴⁰ have established a relationship between the conformational dynamics of a group of side chains in this region and catalysis. Motivated by the sensitivity of the conformational state of the active site-core network (hereafter referred to as the “core” region) to temperature³⁸, we performed infrared laser-driven T-jumps on buffered aqueous solutions of CypA and measured subsequent, time-dependent changes in small and wide angle X-ray scattering (SAXS/WAXS). While our measurements provide only low resolution structural information, we were able to measure the kinetics of protein conformational changes in CypA. We identified two

relaxation processes, and by performing T-jump experiments at a range of different temperatures, we were able to calculate thermodynamic properties of the transition states for the underlying conformational transitions. Specific mutants in the “loops” or the “core” regions of CypA show that the two processes are independent, each representing a distinct and uncoupled reaction coordinate on a complex conformational landscape. Collectively, our measurements and analysis demonstrate that a wealth of information about a protein’s conformational landscape can be obtained by pairing laser-induced T-jump with time-resolved X-ray scattering.

Results

A method for simultaneous measurement of structural and kinetic details of intrinsic protein dynamics

To measure protein structural dynamics, we utilized a pump-probe method that pairs an infrared laser-induced temperature-jump with global measurement of protein structure via X-ray solution scattering (**Figure 2.1**). We performed solvent heating in aqueous protein solutions by exciting the water O-H stretch with mid-IR laser pulses (1443nm, 7ns duration). At regularly defined time delays following the IR heating pulse (from 562ns to 1ms), we probed the sample with high-brilliance synchrotron X-ray pulses from a pink-beam undulator (3% bandwidth at 12keV, **Figure S2.1**) that were approximately 500ns in duration, and measured X-ray scattering using a large CCD detector that was capable of capturing small and wide scattering angles on a single panel. Because the duration of the IR pump pulse was sufficiently short compared to the duration of the X-ray probe, the heating was effectively instantaneous with respect to the relaxation processes we were able to observe. Data were collected as interleaved “laser on” and “laser off” X-ray scattering images, so that each pump-probe measurement could be paired to a

measurement made immediately before application of the pump laser (**Figure 2.1**). We measured 27 unique pump-probe time delays across four decades of time spanning from 562ns to 1ms, performing 50 repeat measurements for each time delay. For each detector image, the individual pixel values were azimuthally averaged as a function of the scattering vector magnitude, q , to give one-dimensional scattering intensity profiles ($I(q)$ curves). All scattering profiles were scaled to a single reference, and the data were analyzed as described below. These pump-probe measurements allowed us to monitor structural changes within the ensemble of heated molecules in real time as the system relaxed to a new thermal equilibrium following T-jump (**Figure 2.1A**).

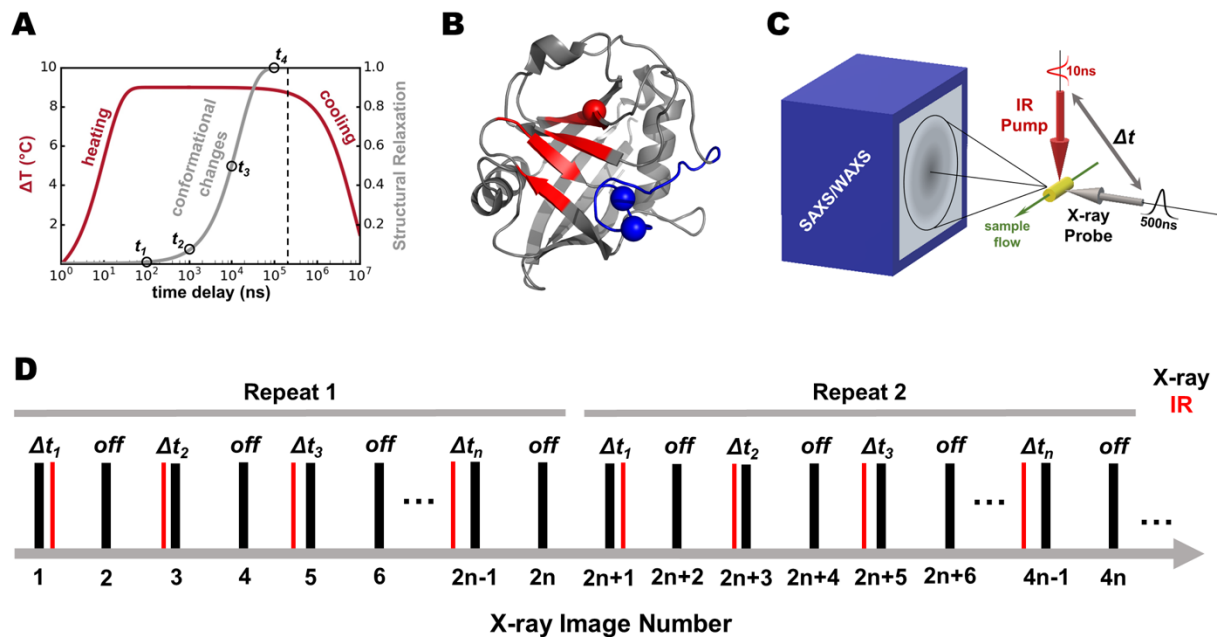


Figure 2.1 | Overview of T-jump SAXS/WAXS experiments.

(a) During a T-jump experiment, an infrared (IR) laser pulse, several nanoseconds in duration, vibrationally excites the water O-H stretch and rapidly heats an aqueous solution of protein molecules (red curve). Although heating is fast, the kinetic barriers to protein motions cause a lag in the structural relaxation to a new thermal equilibrium of conformational states (gray curve). (b) Ribbon diagram depicting a single cyclophilin A (CypA) molecule. The “core” dynamic residues that are linked to catalysis are colored red, and the site of a key mutation (S99T) is identified by a sphere at its $C\alpha$ position. Likewise, the “loop” region adjacent to the active site that helps determine substrate specificity is colored blue, and the site of key mutations (D66N/R69H) are also identified by spheres at their $C\alpha$ positions. (c) A schematic depicting the T-jump SAXS/WAXS instrumentation is shown with key features highlighted. A liquid sample flows horizontally through the interaction region, where it interacts with mutually perpendicular IR pump and X-ray probe beams. Both the pump and probe sources are pulsed, with a defined time delay between their arrival at the sample. Small- and wide-angle X-ray scattering (SAXS/WAXS) patterns are recorded on a single detector panel. (d) The diagram illustrates the data collection sequence used for the experiments described here. For each pump-probe time delay, a pair of images was collected such that the first image was a pump-probe measurement (“laser on”) and the subsequent image was collected with no application of the pump laser (“laser off”). On-off pairs with increasing pump-probe time delays were measured in succession until all of the desired delay times were acquired, and this sequence was repeated as many as 50 times to improve the signal-to-noise ratio of the data. Note that the first measurement within each repeat is a control measurement, wherein the probe pulse arrived at the sample before the pump pulse (negative time delay).

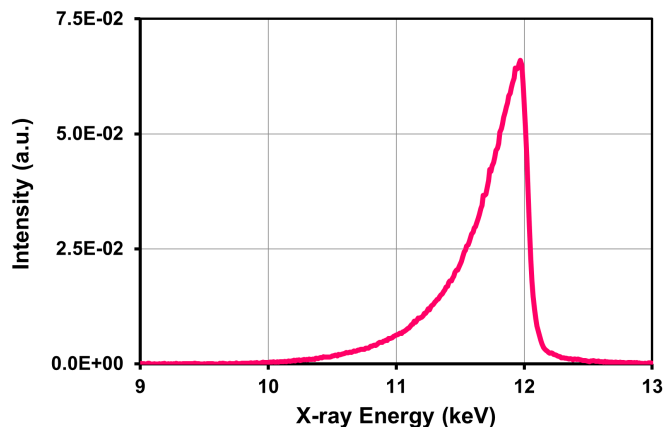


Figure S2.1 | Typical X-ray energy spectrum of the pink beam (3% energy bandwidth) used for the reported SAXS/WAXS measurements.

Calibrating the magnitude of the Temperature-Jump by Singular Value Decomposition

Because the isothermal compressibility of liquid water is highly temperature-dependent⁴¹, X-ray scattering from the bulk solvent acts as an exquisitely sensitive thermometer that can be used to calibrate the magnitude of the T-jump in our experiments^{9,30,32}. Our instrument configuration allowed us to measure low-angle protein scattering and high-angle solvent scattering simultaneously on the same detector image. To characterize the temperature-dependent behavior of the solvent scattering, we performed static SAXS/WAXS measurements of our CypA samples as a function of temperature (equilibrium, no IR laser), in addition to our time-resolved measurements. We pooled these static, temperature-dependent SAXS/WAXS curves (azimuthally integrated $I(q)$ v. q) with the time-resolved SAXS/WAXS curves from our T-jump measurements, and performed singular value decomposition (SVD) on a matrix constructed from the set of pooled curves (**Figure 2.2**). Specifically, each column of this matrix represents a scattering curve, with each row of the matrix corresponding to a q -bin and the entries in the matrix corresponding to measured scattering intensities. The SVD analysis, which was performed over the $q=0.07$ - 3.45 region of the scattering curves, identified a signal (a left

singular vector) whose prominent features were found in the q -region corresponding to the scattering of bulk water ($q > 1.0$) (**Figure 2.2**). By extracting the entries in the corresponding row of the V matrix (containing the right singular vectors as columns), we could determine how this singular vector contributed to each scattering curve and demonstrate that its contribution was strongly temperature-dependent. Specifically, in the static (no T-jump) scattering curves the contribution of this singular vector increased with temperature, and for T-jump measurements the contribution of this vector to the observed scattering curves is perfectly correlated to the application of the pump laser pulse over sequential laser on-off pairs of X-ray measurements (**Figure 2.2**), providing positive confirmation of a T-jump.

The identification of a temperature-dependent singular vector provided a simple way to measure the magnitude of the laser-induced T-jump. For each of the five static temperatures we explored, we calculated the average value of $v_{2,n}$, the entry in the matrix V that describes contribution of the temperature-dependent singular vector (U_2) to the n th scattering curve, across 32 individual X-ray scattering curves. We then plotted the average $v_{2,n}$ vs. temperature and fit the data using both linear and quadratic models (**Figure 2.2**). We examined the residuals for the two fits, determined the quadratic fit produced the most appropriate “standard curve” for estimation of the sample temperature from the SVD analysis, and used the resulting second-degree polynomial to estimate the temperature for each scattering curve in our series of time-resolved measurements. We compared the temperatures calculated for neighboring laser on and laser off scattering curves, and found the average T-jump produced by our IR heating pulse to be approximately 10.7°C on average. The SVD analysis also allows us to judge when cooling of the system becomes significant, so that we can identify the maximum pump-probe time delay that is valid for our relaxation analysis (**Figure 2.2**). We observe that $v_{2,n}$ is consistent as a function of

pump-probe time delay out to delay times of approximately 562 μ s, and decreases for longer time delays, implying that significant cooling of the sample takes place in less than 1 millisecond.

Consequently, we limited our subsequent analysis to time delays shorter than 562 μ s.

Additionally, we note that following laser T-jump, the solvent reaches a new thermal equilibrium faster than the measurement dead time of our experiment (562ns). This observation is consistent with other work, in which changes to the structure of bulk solvent following laser T-jump have been shown to equilibrate within roughly 200ns⁴².

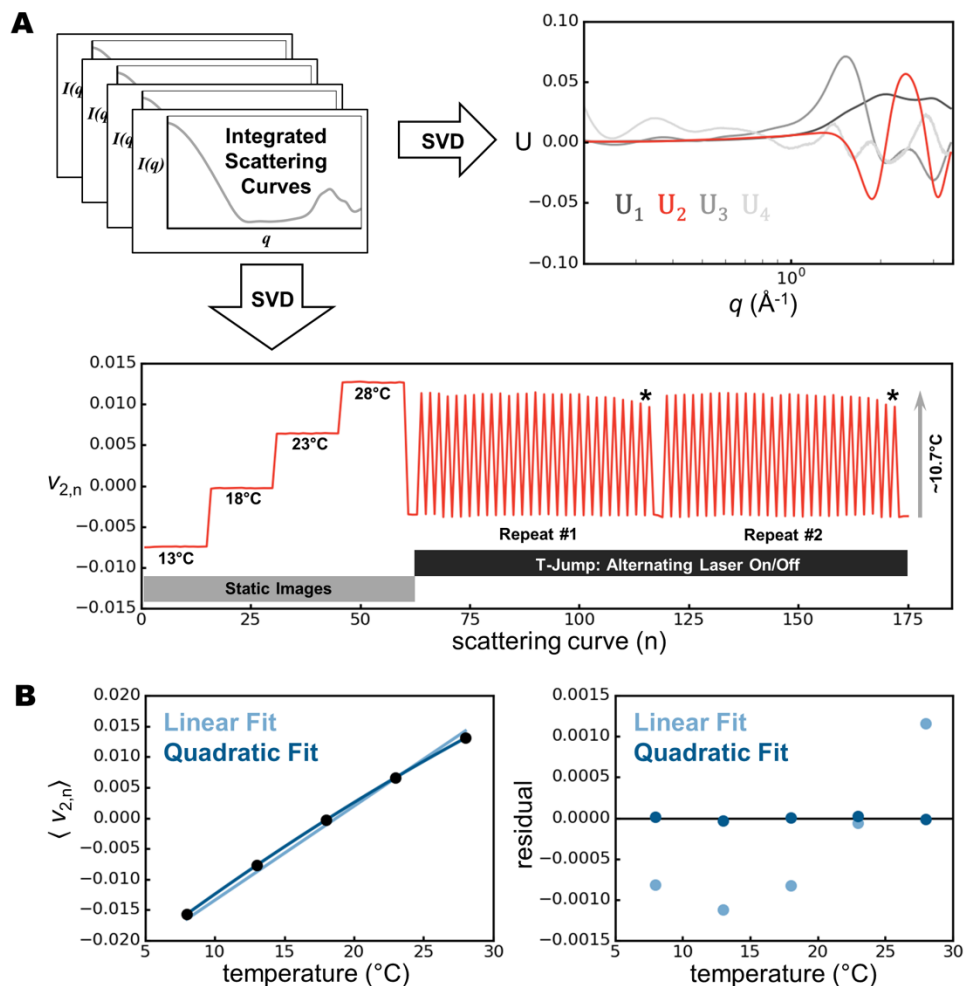


Figure 2.2 | X-ray scattering from bulk water acts as a sensitive thermometer for T-jump experiments.

(a) Using singular value decomposition (SVD), we can identify a signal whose contribution to each scattering curve is strongly dependent on the temperature. The left singular vectors with the four highest singular values are shown, with the vector corresponding to the temperature-dependent signal (U_2) colored red. The contribution of this vector ($v_{2,n}$) to each of 175 scattering curves is also shown. This set of 175 scattering curves includes static measurements (no pump laser) at four different temperatures, followed by two repeats of time-resolved T-jump measurements. The T-jump data were collected as “laser on-off” pairs, and within a single repeat each successive on-off pair was collected with an increasing pump-probe time delay. Cooling is evident at longer pump-probe time delays (denoted by *). (b) To calculate the magnitude of the laser-induced T-jump, we used the static data to determine the average value of $v_{2,n}$ as a function of temperature, and fit the data using both linear and quadratic models. Based on the residuals for the two fits, we chose to use the resulting quadratic equation to determine the magnitude of the laser-induced T-jump using the values of $v_{2,n}$ calculated for the time-resolved scattering curves by SVD.

T-Jump Produces Changes in the X-ray Scattering Profile of CypA

To determine the effect of the T-jump, we initially averaged all data for a given time delay, examined the scattering profiles for differences (**Figure 2.3**), and observed a small laser-induced change in the low-q region of the scattering profiles. Next, we sought to increase the sensitivity of the experiment by exploiting the structure of the interleaved data collection (**Figure 2.1**). We calculated the “on-off difference” between each set of paired “laser on” and “laser off” scattering profiles. Following subtraction, we binned the “on-off difference” scattering curves according to the associated pump-probe time-delay, performed an iterative chi-squared test to remove outliers ($\chi^2=1.5$), and averaged the calculated differences for all repeat measurements (**Figure 2.3**). This subtraction and averaging resulted in accurate measurements of “laser on-off difference” signals as a function of the pump-probe time delay. Scattering differences at high-q ($1.0\text{-}4.2\text{\AA}^{-1}$) were used to calibrate the final sample temperature after laser illumination as described in detail above. It is, however, worth noting here that the shape of the on-off difference signal at high-q is nearly identical to the left singular vector used to monitor the temperature by SVD. In addition to the high angle signals that were used to measure the temperature from the X-ray scattering, the averaging and subtraction also revealed time-resolved changes at low-q ($0.03\text{-}0.3\text{\AA}^{-1}$), which we analyze below in the context of the average physical dimensions and scattering density of the CypA “protein particle.” Here, we use the phrase “protein particle” to describe the protein molecule plus the ordered solvent bound to its surface, since both the protein and its hydration layer have an electron densities that differ from bulk solvent, and therefore contribute to the observed X-ray scattering by the CypA solution. The same T-jump

measurements were performed on protein samples and on samples consisting of buffer only without protein. After an additional scaling step, average on-off differences for the buffer alone were subtracted from average on-off differences for the buffer with protein, which isolated the signal changes at low- q due only to the protein.

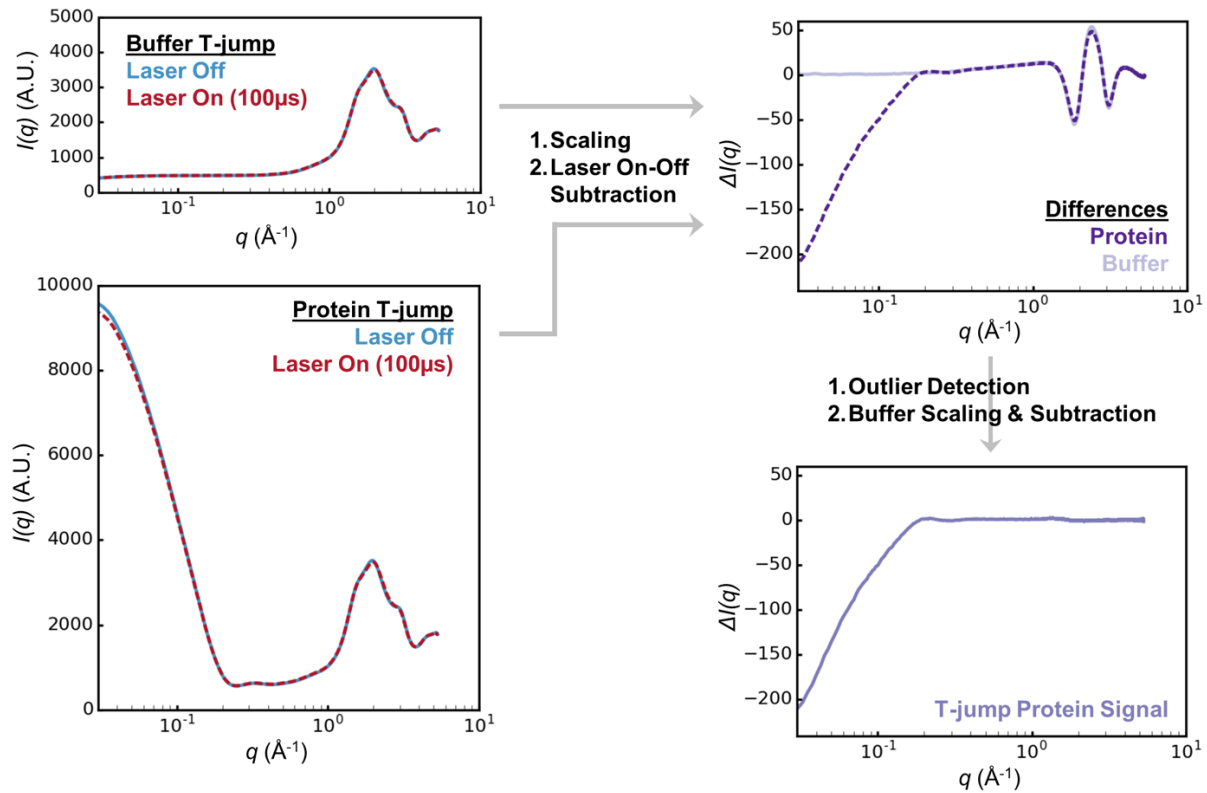


Figure 2.3 | T-jump data processing involves a combination of scaling and subtraction operations that produce time-resolved difference scattering curves. For each “laser on-off” pair, the recorded scattering curves are scaled to one another and the “laser off” curve is then subtracted from the “laser on” curve. This procedure is done independently for samples containing buffer only, and for protein samples. Next, the resulting difference curve for the buffer only sample is scaled to the difference curve obtained for the protein sample, and an additional buffer subtraction is performed to remove the thermal signal from the solvent. The result of this procedure is a difference scattering curve containing signal from the protein molecules only.

Time-Resolved Changes in Small Angle X-ray Scattering

Comparison of difference scattering curves calculated for 27 time delays revealed a time-dependent change in X-ray scattering by the protein, demonstrating that the modest T-jump we introduced was capable of exciting protein dynamics that could be observed in real time. The difference curves calculated from our data, showing the contribution of the protein to time-resolved changes in the SAXS/WAXS signal, have features in the low- q ($q=0.03-0.20 \text{ \AA}^{-1}$) region (**Figure 2.4**). Qualitatively, the time-resolved on-off differences show that the overall low-angle scattering and extrapolated value of $I(0)$ are reduced within the dead time of our experiment (562ns), and then begin to increase slightly over the next few microseconds before decreasing further at longer pump-probe time delays out to 562 μ s. Changes in low-angle scattering and $I(0)$ reflect changes in the overall size and shape of the particles in solution, with a reduction in both observables indicating of a loss of scattering mass and shrinkage of the scattering particles. The observed laser on-off difference in $I(0)$ is approximately 3% of the total observed signal, with one-third of that signal change occurring in the time regime that can be resolved by our measurements.

The on-off difference for our shortest pump-probe time delay (562ns) is significantly different from 0 at low scattering angles, which suggests the existence of structural changes in the system that are faster than the dead time of our measurements. The physical basis for the fast signal change is likely due to a combination of thermal expansion and change in the amount of ordered solvent surrounding the protein. First, thermal expansion of solvent results in the expulsion of some scattering mass from the X-ray beam path. Based on the volumetric thermal expansion coefficient of water (approximately $0.0003/\text{^\circ C}$)⁴³, this effect reduces the overall scattering mass by approximately 0.3% for our T-jumps, which were approximately 11 $^\circ$ C.

Second, protein thermal expansion coefficients are estimated to be larger than those of liquid water^{44,45}, so heating of the sample could result in a reduction in scattering contrast between the protein particles and the bulk solvent. Thermal expansion of solvent is well-known to occur within approximately 200ns, and it is reasonable to assume that protein thermal expansion may occur on a similar timescale. The thermal expansion coefficient of CypA is unknown, and rates of protein thermal expansion in general have not been studied explicitly, although studies of “protein-quake” motions in photoactive systems suggest these effects likely occur within hundreds of picoseconds^{9,46}. Kratky plots created from our static and T-jump data suggest a slight increase in protein flexibility without unfolding (**Figure S2.2**), which we interpret to be the result of protein thermal expansion and an overall increase in thermal disorder. This process appears to be faster than the dead time of our measurements, since the effect is temperature-dependent, but not time-dependent over the pump-probe time delays we explored. In addition to thermal expansion effects, the temperature change likely causes some of the ordered solvent around the protein to “melt” into the bulk, which could also lead to a fast decrease in the overall scattering mass and size of the protein particle. The kinetics of these fast processes, while potentially interesting, are invisible to our experiment. Therefore, our subsequent analysis is focused on structural dynamics that occur in the microsecond regime.

Because the main time-resolved signal change was confined to very low- q , we wanted to ensure our time-resolved signal was due a change in the protein’s form factor (infinite dilution), and not the structure factor of the protein solution. To test whether changes in the radial distribution function originated from structural changes within the individual protein particles and their associated solvent and not from changes in the relative arrangement of the CypA molecules in solution, we performed static SAXS/WAXS measurements of CypA as a function

of both temperature and CypA concentration (**Figure S2.3**). This control allowed us to characterize and correct for the effect of interparticle interactions. Using these data, we calculated the structure factor ($S(q)$) for a 50mg/mL CypA solution at multiple different temperatures, and determined there was no significant difference in the $q=0.03-0.2$ region of the scattering curves, consistent with other work on similarly-sized protein molecules in solution⁴⁷. Next, we plotted the second virial coefficient for CypA as a function of temperature, and noticed that this quantity shows only a very small temperature dependence that cannot account for the observed time-resolved differences. In contrast to our results for CypA, Bonneté, et al. performed similar calculations of second virial coefficients for lysozyme solutions at similar temperatures and buffer conditions, and calculated temperature-dependent changes that were 50-fold larger (or more) than what we determined for CypA⁴⁷. In addition to the direct measurements of interparticle interactions provided by concentration-dependent scattering measurements, we also used Guinier analysis to assess whether the radial distribution function (structure factor) of CypA particles in solution changes significantly upon temperature perturbation. We performed linear fits of $\ln[I(q)]$ vs. q^2 for averaged scattering curves derived from static temperature data and from time-resolved data, and observed that the residuals do not change substantially as a function of either temperature (in static experiments) or time (in time-resolved experiments). Because deviations from the linear Guinier fit are often the result of interparticle interactions, we concluded that the relative consistency of these residuals provides additional evidence that such interactions have a negligible effect on our observations. The Guinier analysis was also used for structural interpretation of the time-resolved signal, which is described in detail below.

Comparison of laser on and off scattering curves revealed scattering differences that were approximately the same in magnitude and direction as differences between static temperature measurements performed on samples equilibrated to temperatures that differ by 10°C (roughly corresponding to the magnitude of the laser-induced T-jump, **Figure S2.4**). The equilibrated signal change that we observe in our time-resolved measurements ($I(0)_{562\mu\text{s}} - I(0)_{\text{off}}$) is similar, but not identical, in magnitude (3.2% of total) and direction to differences calculated using static scattering curves collected at temperatures that approximate the laser on and laser off measurements in our time-resolved experiments (1.9% of total signal). The small discrepancy in the overall signal change induced by a ~10°C temperature change in static versus dynamic experiments could be due to additional relaxation processes which occur on timescales longer than we measure in our experiment (our measurements extended out to 562 μs), whereas some motions in CypA have been reported to have millisecond exchange rates³³, or due to systematic errors in comparing static and time-resolved measurements.

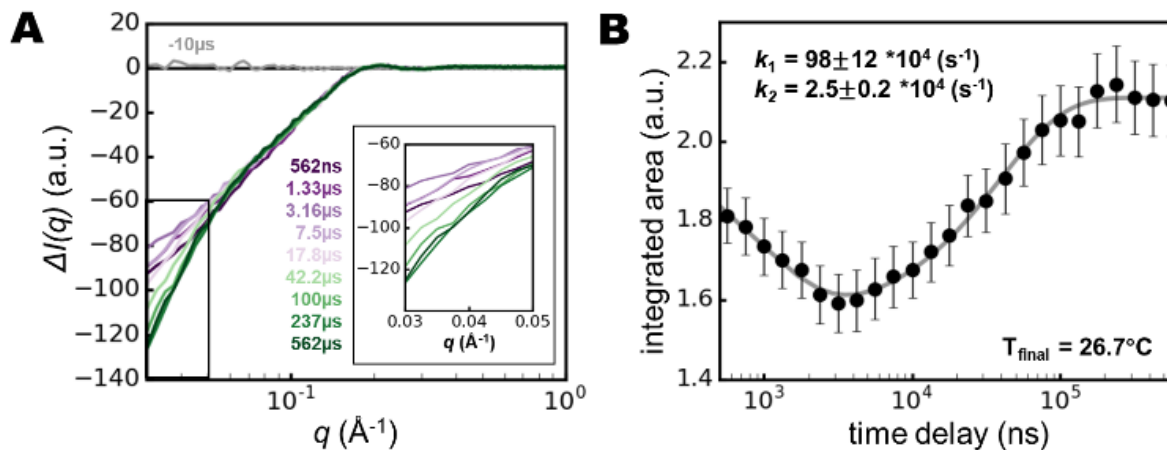


Figure 2.4 | Time-resolved T-jump data allow kinetic modeling of conformational dynamics. (a) A series of time-resolved difference X-ray scattering curves is shown for a subset of our T-jump data (10 out of 26 unique time delays). Data at low q are plotted on a linear q scale in the inset. (b) The area under the difference scattering curve in the $q=0.03\text{-}0.05\text{\AA}^{-1}$ region was integrated for all measured pump-probe time delays, and the resulting absolute values are plotted as a function of the pump-probe time delay. The plotted data suggest the existence of multiple relaxation processes, and we used a two-step model of relaxation kinetics to fit the observations (gray line). The rates calculated from the kinetic fit are provided.

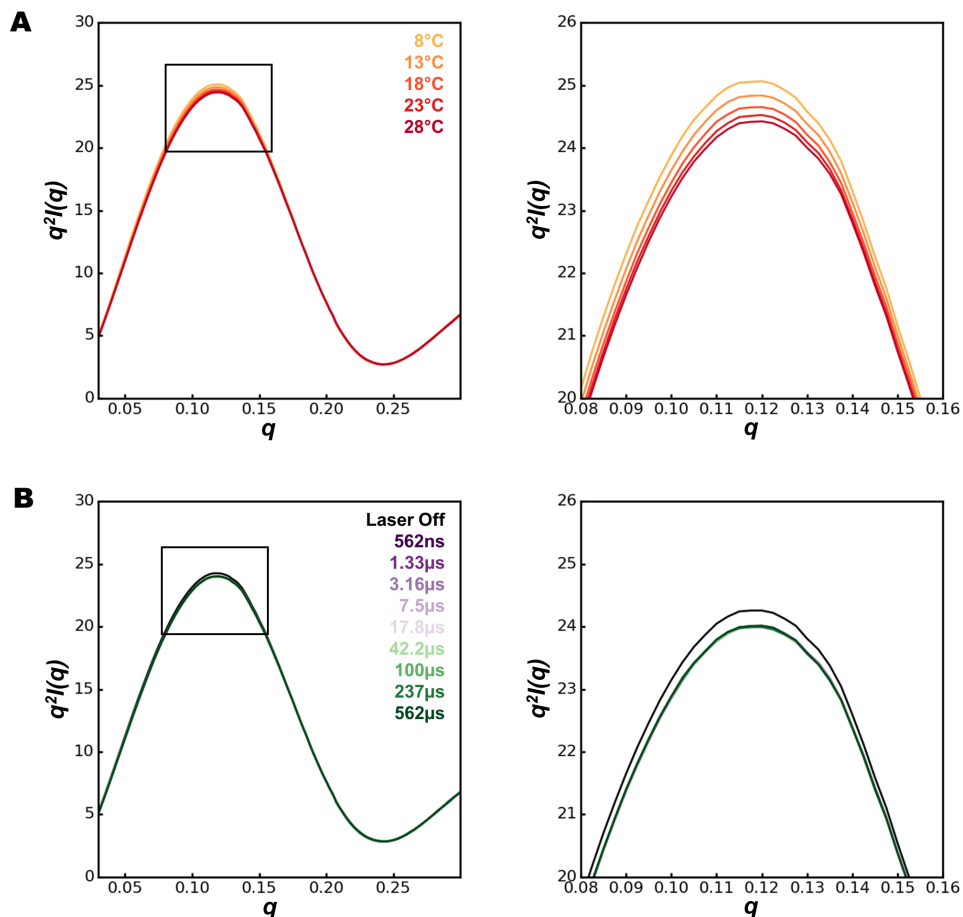


Figure S2.2 | Kratky plots for CypA reveal a small thermal disorder effect without protein unfolding.

(a) Kratky plots calculated as a function of static temperature, from 8°C to 28°C. The right panel shows an expanded view of the boxed region. (b) Kratky plots calculated as a function of time delay for time-resolved T-jump data (T-jump from approximately 15°C to 26°C). Again, the right panel shows an expanded view of the boxed region. All time delays show a similar difference relative to the “laser off” state, indicating that the underlying structural change is faster than the measurement dead time of our experiment.

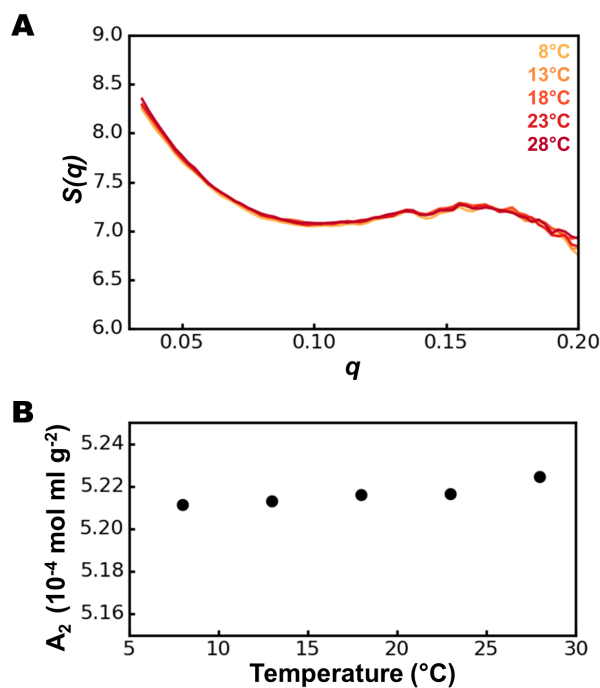


Figure S2.3 | Intermolecular interactions are not temperature dependent for CypA solutions. (a) Structure factors ($S(q)$) calculated for 50mg/mL CypA solutions (wild type) at temperatures ranging from 8-28°C. (b) Second virial coefficients (A_2) calculated for 50mg/mL CypA solutions (wild type) at temperatures ranging from 8-28°C.

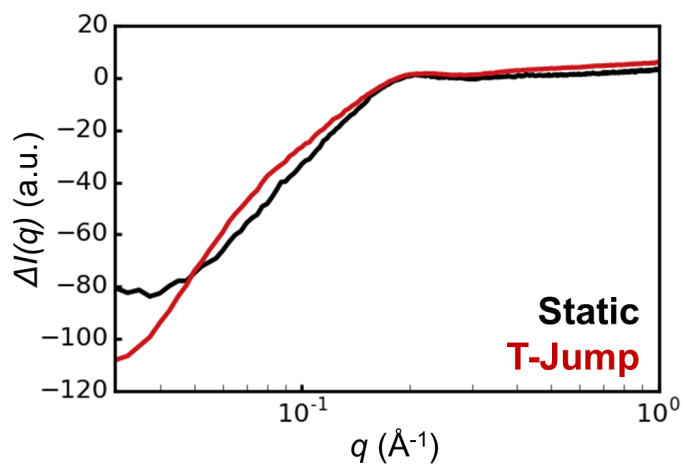


Figure S2.4 | Comparison of static scattering differences between CypA solutions at 13°C and 23°C (black curve), and time resolved differences (100 μ s-laser off) for a T-jump spanning a temperature range of approximately 15°C and 26°C (red curve).

Kinetic Modeling of Structural Dynamics from Time-Resolved Scattering Differences

Our time-resolved measurements of scattering differences allowed us to model the kinetics of global structural changes induced by the T-jump. For kinetic modeling, we integrated the area under each of our time-resolved difference curves in the $q=0.0275$ - 0.04 region and plotted the absolute value of the area as a function of the associated pump-probe time delay (**Figure 2.4**). Based on the apparent shape of the area vs. time delay plot, we reasoned that a two step kinetic model would be needed to fit the data, since the area first decreases, and then increases, as a function of time delay. We fit the observed data to a two-step model of relaxation kinetics (independent steps) using a non-linear least-squares curve fitting algorithm, and calculated rates of $9 \times 10^5 \text{ s}^{-1} \pm 2 \times 10^5 \text{ s}^{-1}$ for the fast process (k_1) and $2.7 \times 10^4 \text{ s}^{-1} \pm 0.2 \times 10^4 \text{ s}^{-1}$ for the slow process (k_2) at 26.7°C (299.7K). The errors calculated for these rates are the result of propagating measurement standard deviations through radial integration, scaling, on-off subtraction, averaging, buffer subtraction, difference curve integration, and kinetic fitting. It is worth noting that the errors calculated in our analyses are likely to overestimate the true error, as we considered all experimental errors to be random. In contrast, some experimental error is likely systematic, and would instead be removed, rather than propagated, by the subtractive operations employed during data processing.

In addition to our kinetic analysis of the on-off difference curves, we also used the time-resolved data to generate $I(q)_t$ scattering curves, which represent the time-dependent X-ray scattering from the CypA sample, but are generated in a manner that makes use of the many repeated paired “laser on” and “laser off” measurements to reduce the effects of systematic error (see Methods). We subsequently used these scattering curves for Guinier analysis to determine how the radius-of-gyration (R_g) of the average CypA particle in the conformational ensemble

changes as a function of time following the T-jump (**Figure 2.5**). This analysis demonstrated that after the T-Jump, the average CypA particle shrinks in the dead time of our experiment; however, within a few microseconds of the T-jump, a fast structural transition (described by k_1 in our kinetic analysis) causes the average CypA particle to expand subtly. While the increase in the calculated radius of gyration is small relative to the error on the Guinier fit for any single data point, our conclusion that the particle is expanding is supported by multiple time points and kinetic analysis of the integrated area under difference scattering curves. Following this fast increase in R_g , a second, slower process (described by k_2 in our kinetic analysis) reverses this trend, causing the average CypA particle to shrink again.

Next, to learn more about the conformational transitions in CypA that are excited by the T-jump, we repeated the experiment at multiple different jumped (final) temperatures ranging from 6.2°C to 29.9°C (279.2K to 302.9K). We modeled the kinetics of the SAXS/WAXS signal changes to observe how the relaxation rates changed as a function of temperature. The calculated rates (k_1 and k_2) for all temperatures are provided in **Table 2.1**. We analyzed the temperature-dependence of these rates using the Eyring equation, which provided insight into the thermodynamics of the transition states for the two processes. First, we plotted $\ln(k/T)$ versus $1/T$ (**Figure 2.6**), and noted that the relationships appeared to be linear. Therefore, we used the fitted slopes and y-intercepts to calculate the enthalpies and entropies of activation for each of the two processes according to the linearized Eyring equation:

$$\ln\left(\frac{k}{T}\right) = \left(\frac{\Delta H^\ddagger}{R} \cdot \frac{1}{T}\right) + \frac{\Delta S^\ddagger}{R} + \ln\left(\frac{k_B}{h}\right) \quad \text{Eq. 1}$$

where R is the gas constant, k_B is the Boltzmann constant, and h is Planck's constant. The enthalpies of activation (ΔH^\ddagger) and entropies of activation (ΔS^\ddagger) and their standard deviations are given in **Table 2.2**. The fast process (k_1) has a large, positive enthalpy of activation, but this is

partially offset by a slightly positive entropy of activation. Formation of the transition state during the slow process (k_2) has a smaller enthalpic cost, but is also entropically disfavored. We note that the lowest temperature measurement (279.2K) was not used in the Eyring analysis of the fast process (k_1) because the error on the measured rate was large due to the low magnitude of the overall time-resolved signal changes at this temperature.

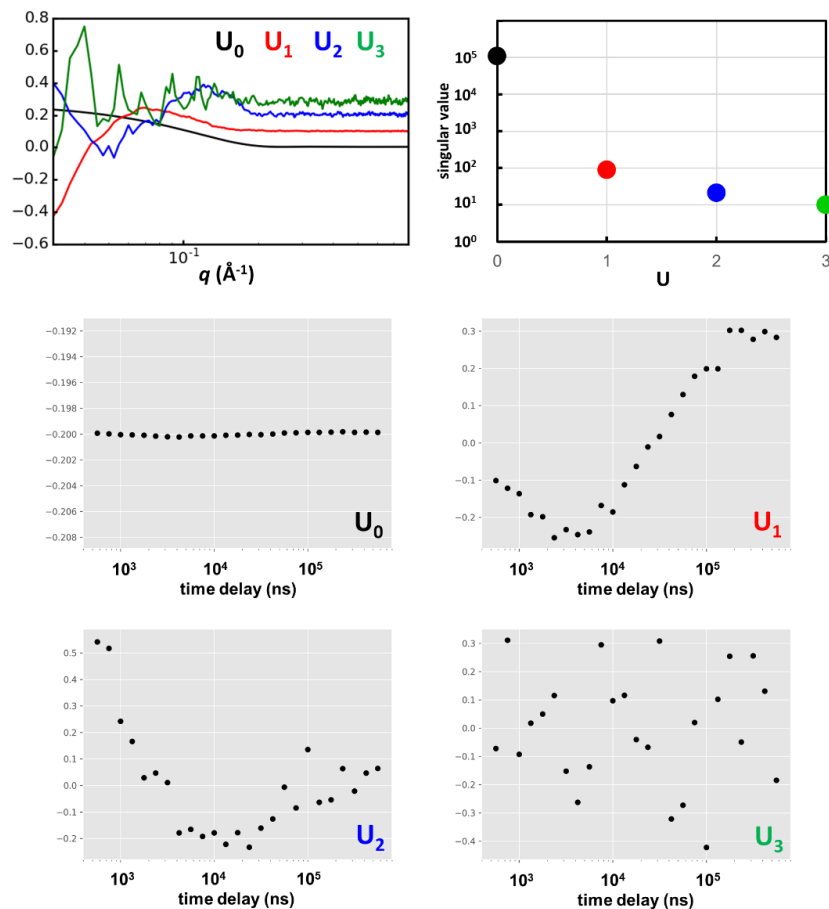


Figure S2.5 | Kinetic analysis of X-ray scattering by singular value decomposition (SVD). We used SAXS curves representing the time-resolved scattering at each time delay spanning 562ns and 562 μ s to construct a matrix, which was analyzed by SVD. The top left panel shows the top four left singular vectors, and the top right panel shows their corresponding singular values. The lower panels are constructed from the right singular vectors, and show the time-dependent contribution of each left singular vector to the total signal. We note that the primary time-resolved signal is dominated by a single singular vector, whose time-dependent behavior reflects that of the integration analysis described in the main text.

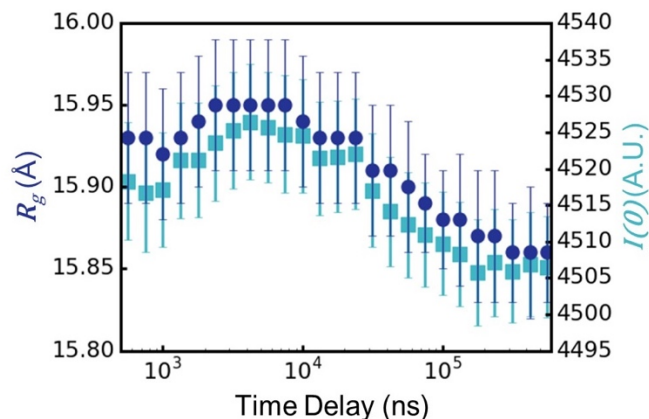


Figure 2.5 | Guinier analysis can be used to estimate changes in physical parameters of the average protein particle in the CypA ensemble from the time-resolved scattering data. Consistent with our kinetic analysis, the radius of gyration (R_g) of the average CypA particle in solution first increases and then decreases as a function of time following the T-jump. Additionally, the value of $I(0)$ extrapolated from the Guinier analysis shows an analogous increase and decrease, suggesting that the change in the particle size is coupled to a change in its average electron density, which is likely due to the acquisition and loss of water molecules from the solvation shell as the protein swells and then shrinks.

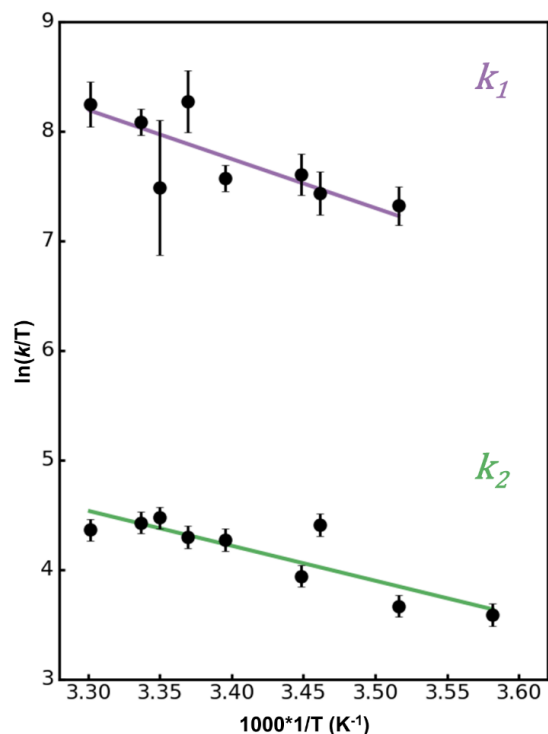


Figure 2.6 | Linear Eyring plots for each of the two relaxation processes observed in our T-jump experiment with CypA. The data are fit using Eq. 1. The linear fit for the fast process (k_1) is shown in purple, and the linear fit for the slow process (k_2) is shown in green

Table 2.1 | Calculated rates for the fast (k_1) and slow (k_2) relaxation processes measured from all T-jump experiments reported here.

Note that kinetic analyses for the S99T and NH variants were performed at 25.5°C and 27.1°C respectively.

CypA Variant	Temperature (°C)	Temperature (K)	k_1 (10^4 s $^{-1}$)	k_2 (10^4 s $^{-1}$)
WT	6.2 ± 0.2	279.2 ± 0.2	N/A	1.1 ± 0.3
WT	11.4 ± 0.2	284.4 ± 0.2	43 ± 7	1.2 ± 0.2
WT	15.9 ± 0.2	288.9 ± 0.2	49 ± 10	2.5 ± 0.3
WT	17.0 ± 0.1	290.0 ± 0.1	58 ± 11	1.6 ± 0.2
WT	21.5 ± 0.1	294.5 ± 0.1	57 ± 7	2.2 ± 0.2
WT	23.8 ± 0.2	296.8 ± 0.2	117 ± 33	2.2 ± 0.2
WT	25.5 ± 0.2	298.5 ± 0.2	53 ± 33	2.7 ± 0.9
WT	26.7 ± 0.1	299.7 ± 0.1	98 ± 12	2.5 ± 0.2
WT	29.9 ± 0.1	302.9 ± 0.1	116 ± 24	2.4 ± 0.2
S99T	25.5 ± 0.2	298.5 ± 0.2	34 ± 4	N/A
D66N/R69H (1)	27.1 ± 0.1	300.1 ± 0.1	N/A	4.1 ± 0.6

Table 2.2 | Enthalpies (ΔH^\ddagger) and entropies (ΔS^\ddagger) of activation for the fast (k_1) and slow (k_2) processes observed for WT CypA, calculated from Eyring analysis.

	ΔH^\ddagger (10^3 J mol $^{-1}$)	ΔS^\ddagger (J mol $^{-1}$ K $^{-1}$)
Fast Process (k_1)	37 ± 8	32 ± 26
Slow Process (k_2)	27 ± 3	-34 ± 11

CypA mutations with distinct effects on conformational dynamics alter time-resolved signal changes

The time-resolved signal changes that we attributed to WT CypA were observed only at low scattering angles, and therefore the resulting structural information had very limited resolution. To gain a better understanding of the structural transitions excited by the T-jump, we next studied two specific CypA mutants, S99T (in the “core” region, **Figure 1B**) and NH (D66N/R69H in the “loops”, **Figure 1B**). The conformational dynamics of these two variants of the enzyme each differ from the wild type in distinct ways: S99T is catalytically impaired due to a loss of rotameric exchange in a key network of residues, whereas NH alters the substrate specificity of CypA by enhancing the dynamics of the surface-exposed loops adjacent to the active site. Importantly, NMR relaxation measurements indicate that S99T perturbs the active site but not the loops³⁹, and that NH only perturb the loops³⁴.

We observed that both S99T and NH mutants showed time-resolved SAXS signal changes that differed from the wild type enzyme. Both mutants show a fast signal change that occurs within the measurement dead time of the experiment, which is similar to what we observed for the wild type enzyme and consistent with these changes being largely due to temperature-dependent changes to the solvation shell and thermal expansion. Beyond the initial fast loss of scattering intensity that was observed (and nearly identical) for all three CypA variants we studied (WT, S99T, and NH), the evolution of the time-resolved signals for each of the two mutants differ substantially from the wild type and from one another. In the S99T mutant (**Figure 2.7**), we observed only the fast decrease (k_1) of the integrated area under the difference curve ($q=0.03-0.05 \text{ \AA}^{-1}$), and a striking absence of the subsequent increase (k_2) in the integrated area at longer time delays that was observed for the wild type enzyme. In contrast, for NH, the

plot of integrated area under the difference curve as a function of time delay (**Figure 2.7**) appears to lack the initial fast decrease (k_1), but it does appear to retain the slower signal change (k_2) that results in an increase for this quantity at longer time delays. We initially fit the data from both the S99T and NH variants using a two-step relaxation model, as we did for the wild type. We found that for the mutants, the two-step kinetic model yielded at least one rate with a large error. For the S99T mutant, the first step (k_1) was well fit but the second step (k_2) was poorly fit, while the opposite was true for the NH variant. After visual inspection, we chose to use a single step kinetic model to fit the data for the S99T and NH mutants, and the calculated rates for the two mutants (k_1 for S99T and k_2 for NH) are also given in **Table 2.1**. Plots of the residuals for these fits revealed no structure, suggesting that a single-step kinetic model is sufficient to explain the data for the CypA mutants. In contrast, fitting kinetic data collected for the wild type enzyme using a single-step model results in residuals with exponential character, and a two-step kinetic model is needed to reduce the error in the fit (**Figure S2.5**).

Our measurements of the S99T and NH variants of CypA clearly demonstrated that mutations which are known to impinge on the activity and specificity of the enzyme also perturb the observed time-resolved signal relative to the wild type in our T-jump experiments. Most notably, the slow relaxation process (modeled by k_2) is shared only by the catalytically-competent wild type and NH variants, and its absence from the S99T variant suggests that the underlying conformational change is related to the catalytically-coupled motions that are arrested by the S99T mutation. These results indicate that T-jump experiments are capable of exciting and measuring functionally-relevant, intramolecular structural dynamics of proteins, even when the data are limited to relatively low scattering angles.

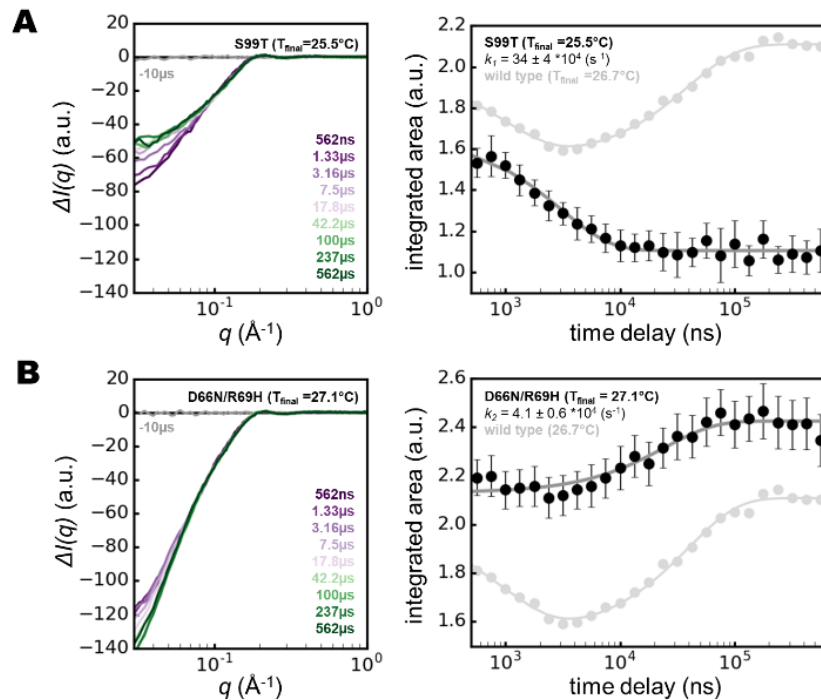


Figure 2.7 | Kinetic analysis of two CypA mutants with distinct effects on the enzyme's function demonstrate the link between the observed T-jump signal and functional dynamics.

The data are presented in the same manner as for the wild type enzyme, shown in Figure 4. In the plots of integrated area versus pump-probe time delay (right panels), the signal observed for the wild type enzyme is shown in light gray for comparison. **(a)** The S99T mutant, which displays defective catalytic function, shows only the fast relaxation process (k_1) and lacks the slower process (k_2). Note that in the right panel, the gray curve representing the wild type signal is offset by approximately -0.2 units, which accounts for the difference in integrated area due to a beamstop shift during the S99T measurements relative to measurements of other variants. **(b)** The D66N/R69H (NH) mutant, with altered substrate specificity, shows the slow relaxation process (k_2) and lacks the faster process (k_1).

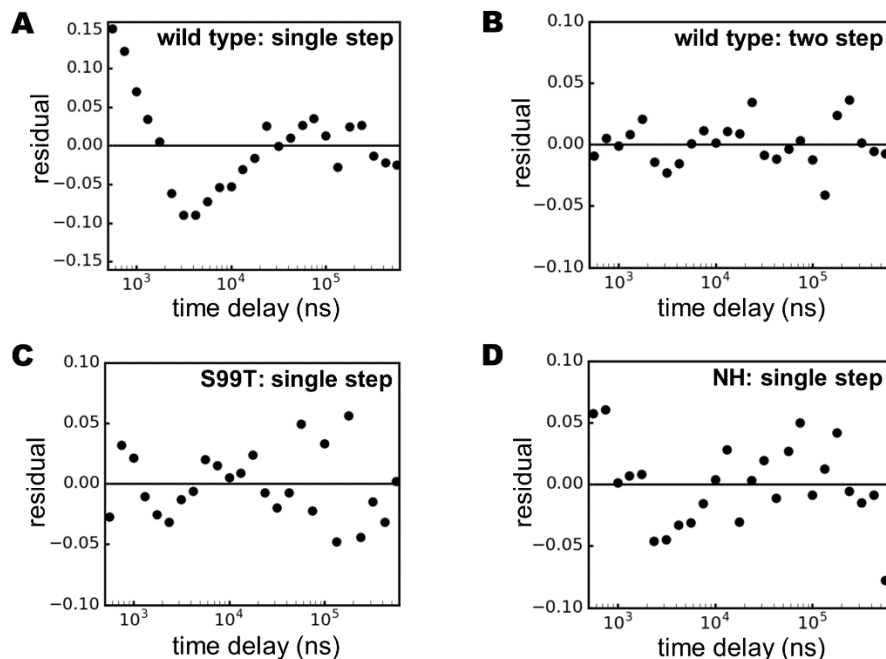


Figure S2.6 | Residuals for kinetic fits of T-jump SAXS/WAXS data for CypA variants. (a) Wild type, single-step relaxation. (b) Wild type, two-step relaxation. (c) S99T mutant, single-step relaxation. (d) NH double mutant, single-step relaxation.

Discussion

These results demonstrate the utility of T-jump X-ray scattering experiments for characterizing the intramolecular structural dynamics of proteins. Time-resolved T-jump X-ray scattering experiments have the potential to be a powerful tool for understanding the complex dynamics of protein molecules, such as the model enzyme CypA, which has no intrinsic photoactivity. In our experiments, T-jumps of approximately 10-11°C modified the CypA conformational ensemble, producing a clear, time-dependent change at very low scattering angles. Because the time-resolved scattering changes could be observed only at low- q , we attempted to control for the possibility that these changes were due to the temperature-dependence of interparticle spacings (quantified by the structure factor, $S(q)$). This process was challenging due to low signal-to-noise at high- q , as well as our inability to measure very low-angle scattering using the available instrumentation. However, static measurements allowed us to

identify the contribution of interparticle scattering, demonstrate that this contribution is invariant to temperature over the range of our measurement conditions, and correct for its effect.

Subsequent Guinier analysis of time-resolved scattering curves allowed us to track changes in the average radius-of-gyration (R_g) of the scattering particles, which include the CypA molecules plus their solvation shells of ordered water molecules. This signal change is comprised of an initial reduction in low-angle scattering that occurs within the measurement dead-time of our experiment, followed by a small increase in low angle scattering that equilibrates within a few microseconds ($k_1 = 98 \pm 12 \times 10^4 \text{ s}^{-1}$ at 26.7°C), and finally a further reduction in low angle scattering that equilibrates within tens of microseconds ($k_2 = 2.5 \pm 0.2 \times 10^4 \text{ s}^{-1}$ at 26.7°C). High-angle scattering differences required for atomistic structural interpretation were not observed due to signal-to-noise considerations. We suspect that high-angle features in time-resolved difference scattering curves may be especially weak for proteins such as CypA, in which conformational motions involve correlated shifts of atoms that can preserve many properties of short-range structure.

This analysis suggests a model in which the scattering density of the CypA particle (protein and ordered solvent) first increases and then decreases after excitation by the T-jump. By performing T-jump experiments over a range of temperatures, we discovered that both the fast and slow processes we observed could be described using Arrhenius kinetics. An Eyring analysis revealed relatively large, positive enthalpies of activation for both processes, consistent with the idea that conformational changes generally require breakage of existing interactions in both the protein and in the solvent. The activation enthalpy for the fast process (k_1) is larger, but the overall activation energy is lower because of a favorable activation entropy. The opposite is

true for the slower process (k_2), which has a smaller overall activation enthalpy, but has a disfavorable activation entropy.

We used the S99T and NH variants of CypA to disentangle the nature of these processes and their associated functions. The S99T mutant is capable of undergoing the fast (k_1) expansion process, but does not experience the subsequent slow (k_2) shrinkage. NMR and crystallography have shown that this mutation arrests the conformational exchange of the “core” catalytic network of residues in CypA by creating steric hindrance, strongly favouring a minor conformation of the wild type enzyme ⁴⁰. This interpretation suggests that the internal rearrangements are related to the k_2 process. However, there is a separation of timescales between the NMR results, which indicate ms dynamics in the “core” region, and the T-jump SAXS results here, which indicate μ s dynamics. This discrepancy may reflect coupled processes that are related by a population shuffling mechanism ⁴⁸ and agree with a broad timescale range of side chain dynamics in CypA uncovered by molecular dynamics experiments ⁴⁹. In contrast to the S99T mutant, the NH variant lacks the initial fast signal change (k_1) in our T-jump experiments, but clearly retains the slow (k_2) signal. NMR and crystallographic studies of the D66N/R69H (NH) double mutant demonstrated that it maintains wild type catalytic motions, but enhanced flexibility of a surface loop region adjacent to the active site ³⁴. This change is due to breaking of several hydrogen-bonding interactions and leads to changes in substrate specificity. Therefore, we hypothesize that the loop motions are responsible for the fast (k_1) signal in WT CypA and S99T, where these motions are known to be unperturbed by NMR ³⁹. However, NMR experiments with NH have shown that the loop motions still occur, but at an increased rate that renders them invisible to our experiments. Furthermore, the assignment of these motions by the mutational analysis is consistent with the R_g changes observed experimentally during each

process. X-ray crystal structures indicate that the minor conformational state of the catalytic network and associated solvent are smaller than the major state and its associated solvent. Using a room temperature X-ray crystal structure of wild type CypA (PDB:3K0N), we calculated the radius-of-gyration of the enzyme with the core catalytic network (Arg55, Met61, Ser99, and Phe113) in both the major and minor conformational states, and found that the predicted R_g of the minor state is 0.07 \AA smaller than the major state (14.09 \AA vs. 14.16 \AA). Additionally the increase in the average R_g during the faster process (k_1) is consistent with the loops sampling an expanded conformational ensemble, as indicated by recent exact-NOE NMR ensembles⁵⁰. Our kinetic modelling of the WT and mutant data suggest that two uncoupled dynamic modes are observed with different kinetics, each of which is individually perturbed by different mutations.

In addition to decreasing the R_g of the protein, conversion from the major to the minor state also results in a small reduction in solvent-exposed surface area, which would necessarily reduce the size of the protein solvation layer. While the reduction of the protein's R_g makes it more compact, and therefore should increase $I(0)$ because the protein has become more dense, the loss of material from the solvation layer opposes this effect, likely leading to the observed decrease in the scattering density of the CypA particle and reduction in $I(0)$. This coupling of R_g changes and changes in the solvent-exposed surface area of the protein can potentially explain the observed correlation between R_g and $I(0)$ changes calculated from our time-resolved scattering data. Additionally the increase in the average R_g and $I(0)$ during the faster process (k_1) is consistent with the loops sampling an expanded conformational ensemble (with increased surface area), as indicated by recent exact-NOE NMR ensembles. Our kinetic modelling of the WT and mutant data suggest that two uncoupled dynamic modes are observed with different kinetics, each of which is individually perturbed by different mutations. It is, however, unclear

from our measurements the extent to which the observed motions are coupled. The S99T variant clearly demonstrates that the fast motion can occur independently of the slower motion, because S99T is known to have arrested slow dynamics. On the other hand, the fast motion we observe is likely to be accelerated by the NH mutation, becoming too fast for us to observe rather than being impeded, as in the case for the slow motion in S99T. Therefore, we are unable to determine whether the fast motion is a required first step that precedes the slower motion.

Time-resolved X-ray structural measurements are critical for decoupling the experimental signatures of conformational changes that can become convoluted by the spatial and temporal averaging that is inherent to traditional X-ray experiments. If one were to assess traditional, static SAXS data for CypA, one would find that increasing the temperature of the sample results in a decrease in the average particle size at equilibrium. These static measurements as a function of equilibrium temperature fail to capture that the temperature change actually perturbs two distinct protein motions, which have the opposite effect on the enzyme's global structural characteristics. This information can only be obtained through a time-resolved experiment, which is able to separate the effects of these two motions because they have substantially different rates. The ability to dissect individual conformational motions and measure their rates using time-resolved X-ray measurements is important for understanding processes involving complex protein dynamics. Many of these dynamic processes, including allostery⁵¹⁻⁵⁴ and enzyme catalysis⁵⁵⁻⁵⁹, involve extensive reorganization of interactions between the protein and its ordered solvation shell, which are key contributors to the energetics that govern protein motions^{20,21,60-64}. Because X-ray solution scattering experiments report on the structure of a protein and the ordered solvent molecules that constitute its solvation shell⁶⁵⁻⁶⁸, the widespread application of time-resolved SAXS/WAXS experiments will enhance our understanding of how protein motions are driven by

solvent dynamics, especially when they can be combined with molecular dynamics simulations to provide atomic scale insight into the underlying structural changes⁹⁻¹². In order for these experiments to enter the mainstream of structural biology, it has become necessary to create perturbations that can be applied universally, to any protein of interest, and our results establish that T-jump can be used as a general perturbation method to excite functional intramolecular protein motions for time-resolved X-ray structural measurements. Looking forward, T-jumps can be paired with other perturbations, such as mutations and ligand binding, to answer important questions about how disease alleles or drug molecules impinge on protein dynamics.

Methods

Sample Preparation

CypA samples were prepared as described previously. Briefly, the recombinant protein was expressed in *E. coli* BL21(DE3) cells and purified by liquid chromatography. Cells were lysed by sonication at pH=6.5, the lysate was clarified by high-speed centrifugation, and CypA was captured from the clarified lysate using a HiTrap-SP cation-exchange column (GE Healthcare). The protein was eluted using a sodium chloride gradient, and fractions containing CypA were pooled, and the pH was shifted to 7.5. The resulting solution was applied to a HiTrap-Q anion exchange column (GE Healthcare), and CypA was collected in the column flow-through. Finally, a polishing step was performed using a Superdex-75 gel filtration column (GE Healthcare). The protein was concentrated to 50mg/mL in buffer containing 20mM HEPES (4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid) buffer at pH=7.5, 50mM sodium chloride, and 0.5mM TCEP (tris-hydroxyethylphosphine). CypA mutants (S99T and NH) were prepared following the same protocol used for the wild type protein. We note that while 50mg/mL is a relatively high protein concentration for in vitro experiments, this is much lower than the typical

intracellular protein concentration (approximately 300mg/mL). For all X-ray measurements performed on buffer only without protein, the buffer was taken from the concentrator filtrate.

T-Jump SAXS/WAXS Data Collection and Processing

Time-resolved SAXS/WAXS measurements of CypA were performed on the BioCARS beamline at the Advanced Photon Source, while the storage ring was operating in hybrid mode. Temperature-jump data were acquired using the pump-probe method, as described recently by Cho. et al.³². Fast temperature-jump was performed on a CypA solution (50mg/mL) in a silica capillary using an Opolette 355 II (HE) optical parametric oscillator (OPOTEK), which produced a 7ns laser pulse with a peak wavelength of 1443nm. The pump laser energy was approximately 1mJ per pulse, and the beam was focused to an elliptical spot with dimensions of 400 μ m by 60 μ m (FWHM, gaussian beam profile), yielding a photon fluence of \sim 50mJ/mm² at the sample, which heated a 50mg/mL CypA solution in a capillary. A suitably delayed X-ray pulse of 494ns duration (eight septuplets in APS hybrid mode) with a peak X-ray energy of 12keV and 3% energy bandwidth (pink beam, **Figure S2.1**), was used to probe the sample following the introduction of the T-jump, and the X-ray scattering was recorded using a Rayonix MX340-HS CCD detector. In our experiments, the temporal resolution is limited to approximately 500ns by the duration of the X-ray pulse, which is substantially longer than the duration of the IR pulse. To speed data acquisition, we utilized a sample holder and data collection scheme recently reported by Cho, et al, (Cho et al., 2018) which combined fast translation along the capillary axis with slow sample circulation via a peristaltic pump. The fast translation of the capillary allowed us to rapidly accumulate X-ray scattering from 41 pump-probe measurements on a single detector image by translating the capillary to a fresh position between each pump-probe pair. The slow circulation of the sample allowed us to replenish the protein solution and limit the extent of

X-ray radiation damage by spreading the X-ray dose over a relatively large volume during long data collection runs. Data were collected as pairs of alternating "laser on" and "laser off" X-ray images. The pump-probe time delay was systematically increased with each successive on/off pair of images. We measured pump-probe time delays spanning three logarithmic decades from 562ns to 1ms, at a time density of eight points per decade. A total of 50 replicate X-ray images were collected for each pump-probe time delay. It is important to note that time-resolved X-ray measurements referred to herein as "laser off," were followed (10 μ s) by application of an IR pulse to the sample, as described by Cho, et al. (Cho et al., 2018), which prevented the introduction of a temperature offset created by incomplete cooling in between "laser on" and "laser off" measurements. A temperature controller integrated into the sample holder allowed us to initiate the T-jump from different starting temperatures, and also allowed us to collect static temperature data. Static temperature images were collected in a manner similar to the time-resolved images, but without application of the pump laser pulse. Data collection protocols were identical for protein and buffer samples.

After acquiring the data we applied polarization, geometry, and detector non-uniformity corrections to the 2D X-ray images. The scattering intensities (photons/pixel) were binned and averaged as a function of the scattering vector magnitude (q), yielding isotropic scattering curves ($I(q)$ vs. q ; $q = 4\pi \cdot \sin(\theta)/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength) (Cho et al., 2018). Next, for each data collection run, we carried out outlier detection by performing singular value decomposition (SVD) on a matrix constructed from our integrated scattering curves. In this SVD, the left singular vector with the largest singular value represents the global average of all the scattering curves used to construct the input matrix. We analyzed the right singular vectors from the SVD to determine which images were irregular. Specifically, we

calculated the mean value of $v_{1,n}$, the entry in the matrix V that describes contribution of the right singular vector with the largest singular value (U_1) to the n^{th} scattering curve, across all the input X-ray scattering curves. Then, if the value of $v_{1,n}$ for any specific scattering was more than 2.5 standard deviations above or below the mean, that image was discarded. Our outlier detection procedure is implemented in a Python script called “SVD_Quarantine.py.” By inspecting the results of the SVD, we decided to remove the first 5 repeats from each data set, as well as some additional outliers that failed our statistical test. The same averaging and outlier detection method was used for both static and time-resolved measurements.

Scaling of X-ray Scattering Curves

All scaling of X-ray scattering curves was performed using an algebraic (least-squares) procedure. To determine the scale factor, A , which can be applied to a scattering curve $I(q)_a$ in order to scale it to a second scattering curve $I(q)_b$, we used the following equation:

$$A = \frac{\sum_q I(q)_a I(q)_b}{\sum_q I(q)_a^2} \quad \text{Eq. 2}$$

Although we used the equation above for scaling throughout our analysis, the q -range to which it was applied varied depending on the context, and details are provided below.

X-ray Thermometry

Following the initial data processing steps described above, we used singular value decomposition (SVD) to determine the magnitude of the T-jump introduced by the IR laser pulse. We pooled static, temperature-dependent SAXS/WAXS curves (azimuthally integrated $I(q)$ v. q) with the time-resolved SAXS/WAXS curves from time-resolved measurements, scaled them to a common reference over the $q=0.025-4.28 \text{ \AA}^{-1}$ region, and performed SVD on a matrix

built from these scaled curves. In this matrix, each column represents a single scattering curve, with the rows of the matrix corresponding to q -bins and the entries in the matrix consisting of azimuthally-averaged scattering intensities. The SVD analysis was performed using only the $q=0.07-3.45 \text{ \AA}^{-1}$ region of the scattering curves. As described in the Results section, the SVD identified a left singular vector whose contribution to the overall scattering signal was highly temperature dependent. This was the left singular vector with the second largest singular value (U_2). For each of the five temperatures used for static data collection, we calculated the average value of $v_{2,n}$, which is the entry in the matrix V describing the contribution of the temperature-dependent singular vector (U_2) to the n^{th} scattering curve. We then plotted the average $v_{2,n}$ vs. temperature and ultimately fit this data using a quadratic model. Finally, we used the resulting second-degree polynomial and the values $v_{2,n}$ for each time-resolved scattering curve to estimate the temperature for each T-jump measurement. By comparison of neighboring “laser on” and “laser off” scattering curves, we determined that the average T-jump was 10.7°C . Our thermometry procedure is implemented in a Python script called “thermometry_timepoints.py.”

Data reduction: On-Off Subtraction, Repeat Averaging, and Buffer Subtraction

We implemented a data reduction procedure that operated on the integrated scattering curves generated using our data collection protocol and produced several outputs that were subsequently used for our kinetic and structural analyses. This procedure, implemented in a Python script called “reduce_data.py,” took advantage of paired “laser on” and “laser off” measurements, redundant measurements of each pump-probe time-delay, and parallel T-jump experiments for samples containing protein and samples consisting of buffer only. The input for this script was essentially two data sets. The first, was the series of time-resolved scattering curves measured from a sample containing protein and consisting of paired “laser on/off”

measurements with multiple replicate measurements of each pump-probe time delay (see above). The second was a similar data set, only collected from a sample containing buffer only and no protein. All of the input scattering curves were scaled to a common reference over the $q=0.025$ - 4.28 \AA^{-1} range, and “laser off” curves were subtracted from their associated “laser on” curves to create a difference scattering curve ($\Delta I(q)$) for each “laser on/off” pair. Next, all replicate difference curves (i.e. same sample and time delay) were grouped together, an iterative chi-squared test was performed (using a cutoff of $\chi^2=1.5$), and the average difference curve was calculated for each pump-probe time delay in the series. For each time delay, the difference signal for the buffer only sample was scaled to the difference signal for the sample containing protein over the $q=1.5$ - 3.6 \AA^{-1} range, and then the buffer signal was subtracted from the protein signal to isolate the difference signal due only to the protein. Additionally, this script took all of the “laser off” scattering curves, performed an iterative chi-squared test (cutoff of $\chi^2=1.5$), and calculated their average. As was done for the difference curves, the average “laser off” scattering curve for buffer only was subtracted from the average “laser off” scattering curve for the protein sample after an additional scaling step ($q=1.5$ - 3.6 \AA^{-1} range). The output of this data reduction procedure was a single scattering curve ($I(q)$ vs. q) for the “laser off” state, and a difference scattering curve ($\Delta I(q)$ vs. q) for each pump-probe time delay. All output data were corrected for the contribution of the buffer, and errors were propagated from the initial measurement standard deviations.

Kinetic Analysis

The averaged difference curves produced by our data reduction procedure were used for kinetic analysis of the time-resolved signal changes, which was implemented in a Python script called “difference_dat_kinetics_bootstrap.py.” For each time delay, this script integrated the area

under the difference curve over the $q=0.03-0.05 \text{ \AA}^{-1}$ range, then fit the resulting data (integrated area vs. time) to calculate relaxation rates using non-linear least-squares curve fitting. We used the following equations, for single-step kinetic fits:

$$A(1 - e^{-k_1 t}) + B \quad \text{Eq.3}$$

And for two-step kinetic fits:

$$A(1 - e^{-k_1 t}) + B(1 - e^{-k_2 t}) + C \quad \text{Eq. 4}$$

The output of this analysis was a relaxation rate, or two rates, with standard errors calculated using a bootstrapping method (DiCiccio and Efron, 1996). In cases where we performed T-jumps at multiple temperatures, we used the observed rates and their standard deviations to perform an Eyring analysis by fitting Eq. 1 (above) using a least-squares method to determine the enthalpy and entropy of activation, and their standard errors (using the covariance matrix). We implemented the Eyring analysis in a Python script called “eyring_fit_k1_k2.py.”

Creation of High-Quality Time-Resolved Scattering Curves for Structural Analysis

To produce high-quality scattering curves that could be used for real space interpretation of the time-resolved X-ray scattering, we took the following steps. We used all of the “laser off” scattering curves from our on-off paired time-resolved measurements to create a single average curve. Then, for each of the time-delays reported, we added the average on-off difference (see above) to this average “laser off” scattering curve:

$$I(q)_t = \langle I(q)_{off} \rangle + \langle \Delta I(q)_t \rangle \quad \text{Eq. 5}$$

Next, we utilized static scattering measurements, as a function of both concentration and temperature, to characterize the effect of intermolecular interactions on the observed X-ray scattering and to calculate structure factors ($S(q)$) for our 50mg/mL CypA solutions at temperatures spanning a range relevant to our T-jump experiments. We calculated structure factors (and second virial coefficients) following the methods described by Bonnette, et al.⁴⁷. The scattering curves derived from summing the average “laser off” signal and the time-resolved differences were then divided by the calculated structure factors to correct for intermolecular interactions and extrapolate our measurements to infinite dilution. Because we discovered that the effect of intermolecular interactions were not temperature dependent, we did not need to model the time-dependence of structure factors for our protein solutions following the T-jump, and the structure factor calculated at 13°C was used for the infinite dilution extrapolation. The calculation of structure factors and the creation of the high-quality, corrected $I(q)$ curves were implemented in a pair of Python scripts called “packing_calc.py” and “packing_correction.py,” respectively.

Guinier Analysis and Calculation of R_g

In order to calculate radii-of-gyration (R_g) and to extrapolate the value of $I(0)$ from scattering curves, we used the linear Guinier approximation:

$$\ln[I(q)] = \ln[I(0)] - \frac{R_g^2}{3} \cdot q^2 \quad \text{Eq. 6}$$

Guinier analysis was performed over the q -region spanning 0.03-0.08Å⁻¹. We note that scattering curves were not placed on an absolute scale, however, this is not a requirement for Guinier analysis. The calculations were implemented in a Python script called “Rg_and_i0.py.”

References

1. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* 450, 964–972 (2007).
2. van den Bedem, H. & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution--it's about time. *Nat. Methods* 12, 307–318 (2015).
3. Bottaro, S. & Lindorff-Larsen, K. Biophysical experiments and biomolecular simulations: A perfect match? *Science* 361, 355–360 (2018).
4. Schmidt, M. Time-Resolved Macromolecular Crystallography at Modern X-Ray Sources. *Methods Mol. Biol.* 1607, 273–294 (2017).
5. Neutze, R. & Moffat, K. Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. *Curr. Opin. Struct. Biol.* 22, 651–659 (2012).
6. Schotte, F. *et al.* Watching a signaling protein function in real time via 100-ps time-resolved Laue crystallography. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19256–19261 (2012).
7. Cho, H. S. *et al.* Picosecond Photobiology: Watching a Signaling Protein Function in Real Time via Time-Resolved Small- and Wide-Angle X-ray Scattering. *J. Am. Chem. Soc.* 138, 8815–8823 (2016).
8. Schlichting, I. & Miao, J. Emerging opportunities in structural biology with X-ray free-electron lasers. *Curr. Opin. Struct. Biol.* 22, 613–626 (2012).
9. Arnlund, D. *et al.* Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser. *Nat. Methods* 11, 923–926 (2014).
10. Berntsson, O. *et al.* Sequential conformational transitions and α -helical supercoiling regulate a sensor histidine kinase. *Nat. Commun.* 8, 284 (2017).

11. Takala, H. *et al.* Signal amplification and transduction in phytochrome photosensors. *Nature* 509, 245–248 (2014).
12. Brinkmann, L. U. L. & Hub, J. S. Ultrafast anisotropic protein quake propagation after CO photodissociation in myoglobin. *Proc. Natl. Acad. Sci. U. S. A.* 113, 10565–10570 (2016).
13. Barends, T. R. M. *et al.* Direct observation of ultrafast collective motions in CO myoglobin upon ligand dissociation. *Science* 350, 445–450 (2015).
14. Coquelle, N. *et al.* Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* 10, 31–37 (2018).
15. Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* 352, 725–729 (2016).
16. Kern, J. *et al.* Structures of the intermediates of Kok’s photosynthetic water oxidation clock. *Nature* 563, 421–425 (2018).
17. Nogly, P. *et al.* Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science* 361, (2018).
18. Malmerberg, E. *et al.* Time-resolved WAXS reveals accelerated conformational changes in iodoretinal-substituted proteorhodopsin. *Biophys. J.* 101, 1345–1353 (2011).
19. Hekstra, D. R. *et al.* Electric-field-stimulated protein mechanics. *Nature* 540, 400–405 (2016).
20. Frauenfelder, H., Fenimore, P. W. & Young, R. D. Protein dynamics and function: insights from the energy landscape and solvent slaving. *IUBMB Life* 59, 506–512 (2007).

21. Fenimore, P. W., Frauenfelder, H., McMahon, B. H. & Parak, F. G. Slaving: solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16047–16051 (2002).
22. Wang, J. & El-Sayed, M. A. Temperature jump-induced secondary structural change of the membrane protein bacteriorhodopsin in the premelting temperature region: a nanosecond time-resolved Fourier transform infrared study. *Biophys. J.* 76, 2777–2783 (1999).
23. Wang, T., Lau, W. L., DeGrado, W. F. & Gai, F. T-jump infrared study of the folding mechanism of coiled-coil GCN4-p1. *Biophys. J.* 89, 4180–4187 (2005).
24. Akasaka, K., Naito, A. & Nakatani, H. Temperature-jump NMR study of protein folding: ribonuclease A at low pH. *J. Biomol. NMR* 1, 65–70 (1991).
25. Gillespie, B. *et al.* NMR and temperature-jump measurements of de novo designed proteins demonstrate rapid folding in the absence of explicit selection for kinetics. *J. Mol. Biol.* 330, 813–819 (2003).
26. Yamasaki, K. *et al.* Real-time NMR monitoring of protein-folding kinetics by a recycle flow system for temperature jump. *Anal. Chem.* 85, 9439–9443 (2013).
27. Meadows, C. W., Balakrishnan, G., Kier, B. L., Spiro, T. G. & Klinman, J. P. Temperature-Jump Fluorescence Provides Evidence for Fully Reversible Microsecond Dynamics in a Thermophilic Alcohol Dehydrogenase. *J. Am. Chem. Soc.* 137, 10060–10063 (2015).
28. Vaughn, M. B., Zhang, J., Spiro, T. G., Dyer, R. B. & Klinman, J. P. Activity-Related Microsecond Dynamics Revealed by Temperature-Jump Förster Resonance Energy

- Transfer Measurements on Thermophilic Alcohol Dehydrogenase. *J. Am. Chem. Soc.* 140, 900–903 (2018).
29. Hori, T. *et al.* The initial step of the thermal unfolding of 3-isopropylmalate dehydrogenase detected by the temperature-jump Laue method. *Protein Eng.* 13, 527–533 (2000).
 30. Rimmerman, D. *et al.* Direct Observation of Insulin Association Dynamics with Time-Resolved X-ray Scattering. *J. Phys. Chem. Lett.* 8, 4413–4418 (2017).
 31. Rimmerman, D. *et al.* Insulin hexamer dissociation dynamics revealed by photoinduced T-jumps and time-resolved X-ray solution scattering. *Photochem. Photobiol. Sci.* 17, 874–882 (2018).
 32. Cho, H. S. *et al.* Dynamics of Quaternary Structure Transitions in R-State Carbonmonoxyhemoglobin are Unveiled in Time-Resolved X-ray Scattering Patterns Following a Temperature Jump. *J. Phys. Chem. B* (2018). doi:10.1021/acs.jpcc.8b07414
 33. Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438, 117–121 (2005).
 34. Caines, M. E. C. *et al.* Diverse HIV viruses are targeted by a conformationally dynamic antiviral. *Nat. Struct. Mol. Biol.* 19, 411–416 (2012).
 35. Price, A. J. *et al.* Active site remodeling switches HIV specificity of antiretroviral TRIMCyp. *Nat. Struct. Mol. Biol.* 16, 1036–1042 (2009).
 36. Virgen, C. A., Kratovac, Z., Bieniasz, P. D. & Hatzioannou, T. Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3563–3568 (2008).

37. Wilson, S. J. *et al.* Independent evolution of an antiviral TRIMCyp in rhesus macaques. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3557–3562 (2008).
38. Keedy, D. A. *et al.* Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *Elife* 4, (2015).
39. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462, 669–673 (2009).
40. Otten, R. *et al.* Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* 9, 1314 (2018).
41. Clark, G. N. I., Hura, G. L., Teixeira, J., Soper, A. K. & Head-Gordon, T. Small-angle scattering and the structure of ambient liquid water. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14003–14007 (2010).
42. Gruebele, M., Sabelko, J., Ballew, R. & Ervin, J. Laser Temperature Jump Induced Protein Refolding †. *Acc. Chem. Res.* 31, 699–707 (1998).
43. Irvine, T. F. & Duignan, M. R. Isobaric thermal expansion coefficients for water over large temperature and pressure ranges. *Int. Commun. Heat Mass Transfer* 12, 465–478 (1985).
44. Frauenfelder, H. *et al.* Thermal expansion of a protein. *Biochemistry* 26, 254–261 (1987).
45. Hiebl, M. & Maksymiwi, R. Anomalous temperature dependence of the thermal expansion of proteins. *Biopolymers* 31, 161–167 (1991).
46. Levantino, M. *et al.* Ultrafast myoglobin structural dynamics observed with an X-ray free-electron laser. *Nat. Commun.* 6, 6772 (2015).
47. Bonneté, F., Finet, S. & Tardieu, A. Second virial coefficient: variations with lysozyme crystallization conditions. *J. Cryst. Growth* 196, 403–414 (1999).

48. Smith, C. A. *et al.* Population shuffling of protein conformations. *Angew. Chem. Int. Ed Engl.* 54, 207–210 (2015).
49. Wapeesittipan, P., Mey, A., Walkinshaw, M. & Michel, J. Allosteric effects in catalytic impaired variants of the enzyme cyclophilin A may be explained by changes in nano-microsecond time scale motions. *bioRxiv* 224329 (2018). doi:10.1101/224329
50. Chi, C. N. *et al.* A Structural Ensemble for the Enzyme Cyclophilin Reveals an Orchestrated Mode of Action at Atomic Resolution. *Angew. Chem. Int. Ed Engl.* 54, 11657–11661 (2015).
51. Kim, J. G. *et al.* Cooperative protein structural dynamics of homodimeric hemoglobin linked to water cluster at subunit interface revealed by time-resolved X-ray solution scattering. *Struct Dyn* 3, 023610 (2016).
52. Colombo, M. F., Rau, D. C. & Parsegian, V. A. Protein solvation in allosteric regulation: a water effect on hemoglobin. *Science* 256, 655–659 (1992).
53. Salvay, A. G., Grigera, J. R. & Colombo, M. F. The role of hydration on the mechanism of allosteric regulation: in situ measurements of the oxygen-linked kinetics of water binding to hemoglobin. *Biophys. J.* 84, 564–570 (2003).
54. Royer, W. E., Jr, Pardanani, A., Gibson, Q. H., Peterson, E. S. & Friedman, J. M. Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl. Acad. Sci. U. S. A.* 93, 14526–14531 (1996).
55. Fenwick, R. B., Oyen, D., Dyson, H. J. & Wright, P. E. Slow Dynamics of Tryptophan-Water Networks in Proteins. *J. Am. Chem. Soc.* 140, 675–682 (2018).
56. Grossman, M. *et al.* Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site. *Nat. Struct. Mol. Biol.* 18, 1102–1108 (2011).

57. Decaneto, E. *et al.* Solvent water interactions within the active site of the membrane type I matrix metalloproteinase. *Phys. Chem. Chem. Phys.* 19, 30316–30331 (2017).
58. Leidner, F., Kurt Yilmaz, N., Paulsen, J., Muller, Y. A. & Schiffer, C. A. Hydration Structure and Dynamics of Inhibitor-Bound HIV-1 Protease. *J. Chem. Theory Comput.* 14, 2784–2796 (2018).
59. Guha, S. *et al.* Slow solvation dynamics at the active site of an enzyme: implications for catalysis. *Biochemistry* 44, 8940–8947 (2005).
60. Dahanayake, J. N. & Mitchell-Koch, K. R. Entropy connects water structure and dynamics in protein hydration layer. *Phys. Chem. Chem. Phys.* 20, 14765–14777 (2018).
61. Wand, A. J. & Sharp, K. A. Measuring Entropy in Molecular Recognition by Proteins. *Annu. Rev. Biophys.* 47, 41–61 (2018).
62. Caro, J. A. *et al.* Entropy in molecular recognition by proteins. *Proc. Natl. Acad. Sci. U. S. A.* 114, 6563–6568 (2017).
63. Gavrilov, Y., Leuchter, J. D. & Levy, Y. On the coupling between the dynamics of protein and water. *Phys. Chem. Chem. Phys.* 19, 8243–8257 (2017).
64. Conti Nibali, V., D'Angelo, G., Paciaroni, A., Tobias, D. J. & Tarek, M. On the Coupling between the Collective Dynamics of Proteins and Their Hydration Water. *J. Phys. Chem. Lett.* 5, 1181–1186 (2014).
65. Hub, J. S. Interpreting solution X-ray scattering data using molecular simulations. *Curr. Opin. Struct. Biol.* 49, 18–26 (2018).
66. Svergun, D. I. *et al.* Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2267–2272 (1998).

67. Virtanen, J. J., Makowski, L., Sosnick, T. R. & Freed, K. F. Modeling the hydration layer around proteins: applications to small- and wide-angle x-ray scattering. *Biophys. J.* 101, 2061–2069 (2011).
68. Henriques, J., Arleth, L., Lindorff-Larsen, K. & Skepö, M. On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations. *J. Mol. Biol.* 430, 2521–2539 (2018).

Chapter III

Dissecting and Engineering Mammalian Chitinases

Contributing authors

Benjamin A. Barad^{1,2}, Lin Liu¹, Roberto E. Diaz^{1,3}, Ralph Basilio^{1,4}, Steven J. Van Dyken⁵,

Richard M. Locksley^{6,7,8}, James S. Fraser^{1*}

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

* - jfraser@fraserlab.com

Introduction

Polysaccharides are ubiquitous biopolymers which serve roles ranging from energy storage to signalling to structural rigidity.¹⁻³ Polysaccharide catabolism is achieved by enzymes in the glycosyl hydrolase, lytic polysaccharide monoxygenase, and glycosyltransferase families.⁴⁻⁶ Many polysaccharides assemble into higher order structures that complicate access by metabolic enzymes, and processive hydrolases are commonly employed by organisms to more effectively metabolize the substrates.^{7,8} Enzymatic hydrolysis is often quantified using

small oligomer analogues, which release a chromophore or fluorophore when cleaved.^{9,10} These simple substrates allow for high signal to noise and precise quantification, but activity measurement using these simple substrates is fundamentally limited in assessing activity with bulk substrates, due to their higher order crystalline structures, variable polymer length, branching and other modifications. An additional challenge is quantifying processivity - processive enzymes cut bulk substrate multiple times for each binding event. As simple substrates often only have one site at which hydrolysis will generate signal, the degree of processivity cannot be assessed at all with these substrates, and the total activity of the enzyme with bulk substrate can be very different in the presence vs absence of processivity.⁸

The difficulties in assessing catabolism of complex polysaccharide substrates is exemplified by Chitin, which is a ubiquitous polysaccharide, comprised of β -1,4-linked n-acetyl glucosamine, that is produced by fungi and arthropods for structural rigidity and water repulsion.^{3,11} Chitin polymers assemble into water-insoluble microcrystals, which have been observed in 3 different crystal forms, differentiated by the parallel or antiparallel orientation of neighboring particles.¹² The most common and lowest energy conformation, alpha-chitin, forms into antiparallel sheets which intercalate the N-acetyl groups of neighboring polymers and form tight hydrogen bonding networks.¹³ Strands of chitin must be extracted from this highly ordered structure in order 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.^{8,14} This observation makes it particularly challenging to effectively associate degradation with short oligomeric analogues with true catalytic efficacy. Beyond this, the insolubility and recalcitrance of bulk chitin makes it a particularly challenging substrate to quantify with high precision. Recently, several new methods have tackled this problem, both using labelled chitin substrates with

chromatography^{15,16} as well as enzyme coupled assays to generate colorimetric signal from reducing ends.¹⁷ 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.

Dissecting chitinolytic activity to understand its mechanisms has particularly interesting implications for understanding how mammalian chitinases function in innate immunity. Chitin is not expressed by mammals, but mammals have a conserved machinery to recognize and degrade environmental chitin that is inhaled or ingested.^{18,19} The molecular mechanism of recognition 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 mammals.¹⁸ Both are two domain family-18 glycosyl hydrolase 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 domain which is expected to be highly glycosylated, while Chitotriosidase has a shorter proline rich linker that has also been found to have glycosylation.²⁰⁻²² The role of the chitin binding domain, and of the linker, in processing chitin are not well understood. 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 substrate.²³ AMCase deficient mice build up chitin in their lungs and develop tissue fibrosis as an aging phenotype; external addition of recombinant chitinase to the airway reduces this phenotype.²⁴ This suggests that AMCase is predominantly responsible for clearance of chitin from airways, and further suggests that improvements to AMCase may reduce chitin load in airways.

Consistent with the reported role of AMCase in asthma, while the most common human haplotype of AMCase is significantly less active than the mouse AMCase (mAMCase), polymorphisms of AMCase which increase its activity have been associated previously with asthma protection.²⁵ A trio of mutations found in humans, N45D, D47N, and R61M, which change residues to the wild type identities of mAMCase, has been previously described to increase specific activity against model substrates.²⁶ Of these mutations, previous 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 mutation.¹⁶ The mechanism by which these mutations alter binding and catalysis, both with simple and complex substrates, remains unclear.

In this study, we develop and compare several techniques of measuring kinetics of mammalian chitinases, and provide improved methodology using commercial fluorogenic substrates as well as novel approaches to quantifying bulk chitin degradation. We develop a one-pot continuous-read fluorescent assay based on the previously developed enzyme-coupled assay.¹⁷ We use these methods to assay the impact of the carbohydrate binding domain on activity, and discover that it cause a minor K_m vs k_{cat} tradeoff but does not have a major effect on overall activity. We also compare the activity of murine acidic mammalian chitinase and chitotriosidase with different small oligomeric substrates and with bulk chitin. We compare the results of the asthma associated mutants in the mouse background, and find that the dominant effect is a k_{cat} decrease from the M61R mutation.

Finally, we endeavored to engineer hyperactive chitinase mutants, using a directed evolution approach based on simple fluorogenic substrates. We find mutations which dramatically increase the activity of the enzyme, both improving binding and catalysis. Using the novel enzyme coupled assay, we discovered that these mutations did not have the same effect

with bulk substrates, underscoring the importance of making kinetic measurements with real chitin substrates, both for screening and for characterization.

Results

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

In order to assess the role of the carbohydrate binding domain of AMCase in catalysis, we expressed and purified the isolated catalytic domain of AMCase, as well as the full length enzyme, using an *E. coli* periplasmic expression approach.²⁷ We first measured the ability of the enzyme to catalyze the breakdown of 4-methylumbelliferone (4MU) conjugated chitobiose, using a continuous read approach at pH 7. The activity of the two constructs was indistinguishable, either in binding or catalysis (**Figure 3.1, Table 3.1**). Due to the short length of the oligomeric substrate, this reaction is likely driven only by local interactions in the catalytic domain, and the presence of the carbohydrate binding domain does not affect it. Understanding the difference in activity between these constructs drove the development of new methodology for quantifying chitinase activity with complex substrates. We took advantage of the commercial availability of colloidal chitin substrates, which we found to be more consistent in size and shape and to have reduced settling times compared to traditional shrimp shell chitin. We first attempted to measure 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 enzymes 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 3.1, Table 3.1**). Each cutting event only minimally alters the scattering of chitin crystals, and many cuts are necessary to fully solubilize crystals.

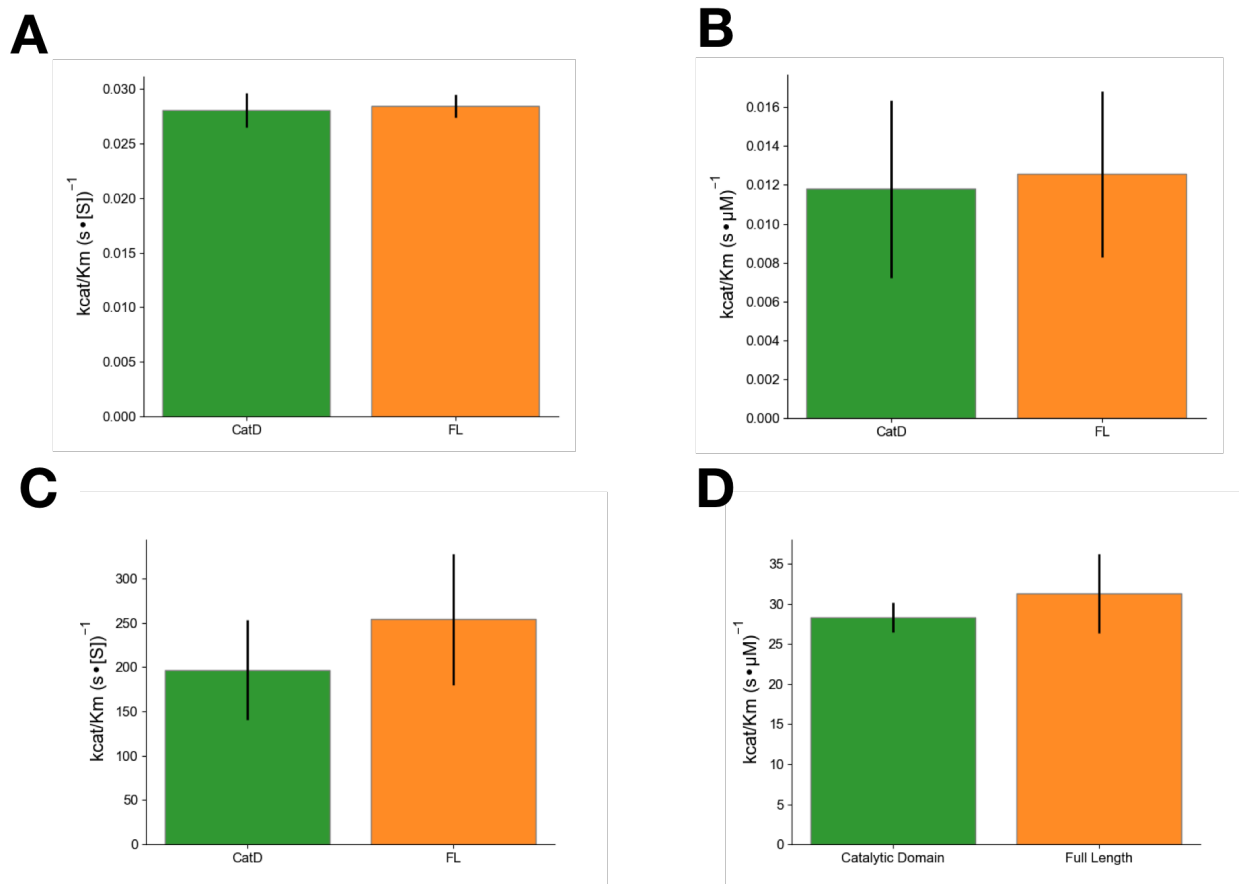


Figure 3.1 | Activity comparisons of AMCase catalytic domain and full length enzyme. Difference in k_{cat}/K_m of AMCase catalytic domain and full length enzyme generated via (a) 4MU-chitobiose assay (b) colloidal chitin clearance assay (c) reducing sugar generation assay quantified with potassium ferricyanide (d) chitooligosaccharide oxidase coupled peroxidase assay. Catalytic domain is in green, while full length AMCase including the linker and carbohydrate binding domain is in orange.

Table 3.1 | Calculated rate constants of AMCase catalytic domain and full length enzyme k_{cat} values are reported in units of 1/s. K_m values are reported in units of μM 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-Chitobiose	1.12 ± 0.09	28.0 ± 2.97	1.05 ± 0.07	25.2 ± 2.26
Colloidal Clearance	$.00140 \pm .00008$	$.0890 \pm .0274$	$.00106 \pm .00002$	$.0650 \pm 0.0178$
Ferricyanide	0.454 ± 0.042	0.0461 ± 0.0161	0.392 ± 0.023	0.0294 ± 0.0069
ChitO	0.944 ± 0.111	0.0333 ± 0.0056	0.540 ± 0.083	0.0172 ± 0.0049

We next attempted to quantify the production of soluble reducing ends, which we hypothesized to be more sensitive and to have an improved connection between individual catalytic events and improvement in signal. The first method we used to attempt this was a classical ferricyanide reduction assay:²⁸ after incubating colloidal chitin with AMCase at 37°C for up to 18 hours, we quenched the reaction with sodium carbonate, removed the insoluble chitin by centrifugation, and quantified the non-enzymatic reaction of soluble reducing sugars with potassium ferricyanide, read out by the disappearance of the yellow color by absorbance at 420nm. With this assay, we were not able to identify a significant difference in total activity, but were able to identify a small improvement in K_m that was offset by reduction in the k_{cat} of AMCase with the addition of the carbohydrate binding domain (**Figure 3.1, Table 3.1**). This tradeoff did not result in a large difference in activity, and due to the endpoint-based requirements of 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 ends.¹⁷ In order 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 and loss of maximal catalytic activity with the inclusion of the carbohydrate binding domain, resulting in no significant change in total activity (**Figure 3.1, Table 3.1**).

A caveat of this approach is that our enzymes, expressed in *E. coli*, lack post-translational modifications. Acidic Mammalian Chitinase is predicted to have multiple O-linked glycosylation sites in the linker between the catalytic domain and the chitin binding domain²⁸. Difference in binding, substrate specificity, and eventually processivity all may be affected by posttranslational modifications. In particular, the carbohydrate binding domain may interact directly with the glycosylations of AMCase, in *cis* or in *trans*, and alter the behavior of the enzyme in bulk. Further experiments to understand the role of the carbohydrate binding domain may require working with enzymes expressed in systems with native glycosylation.

Comparison of Acidic Mammalian Chitinase and Chitotriosidase

The different mechanisms employed by AMCase and Chitotriosidase to degrade chitin are not well understood, nor is their individual role in reacting to chitin in the lungs. Recent work has identified significant activity differences between acidic mammalian chitinase and chitotriosidase and has determined that in mice they do not have any epistatic effects in combination.²⁹ We sought to understand how differences in binding, substrate specificity, and hydrolytic activity differed between the two enzymes. We investigated substrate specificity by comparing the ability of the enzymes to cleave the terminal glycosidic linkage on 4MU-chitobiose and 4MU-chitotriose, representing hydrolysis in different substrate binding poses to generate chitobiose vs chitotriose as a substrate. When assayed the 4MU-chitobiose substrate, AMCase had more than double the activity of chitotriosidase, driven by a significant difference in K_m (**Figure 3.2, Table 3.2**). In contrast, the 4MU-chitotriose substrate led to tighter binding for both AMCase and chitotriosidase, but the difference was much larger with chitotriosidase, leading to a significantly smaller gap in activity between the two enzymes (**Figure 3.2, Table**

3.2). The reduction in observed k_{cat} for both enzymes was likely driven by the alternative, non-fluorogenic reaction trajectory in which the 4MU chitotriose is cleaved into chitobiose and 4MU-bound N-acetyl-glucosamine, leading to a systematic underestimate of k_{cat} . The differences in the K_m suggests that chitotriosidase benefits more from the extended binding interactions available with the larger 4MU-chitotriose 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_m differences (**Figure 3.2, Table 3.2**). The difference in the total activity measured between the 4MU assays and the bulk chitin experiment may be due to a variety of different mechanisms, including differential binding efficiency for extended chitin chains or differences in processivity.

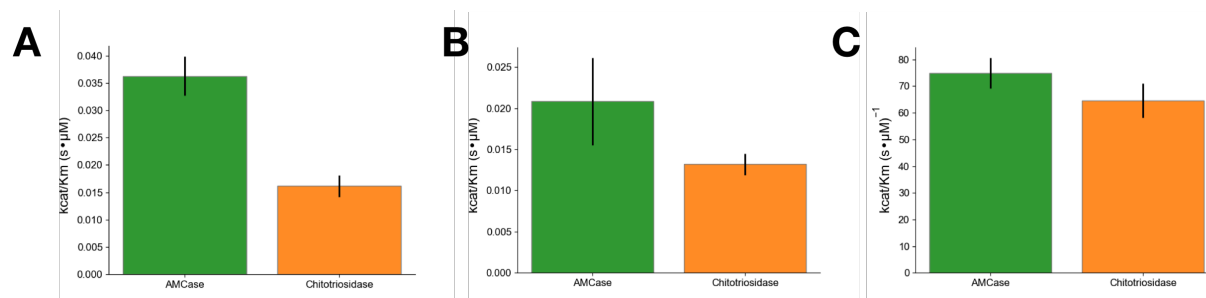


Figure 3.2 | Comparison of AMCCase and Chitotriosidase. Differences in k_{cat}/K_m between AMCCase (green) and Chitotriosidase (orange) using (a) 4MU-chitobiose (b) 4MU-chitotriose (c) chitooligosaccharide oxidase coupled peroxidase assay.

Table 3.2 | Calculated rate constants for AMCCase and Chitotriosidase

k_{cat} values are reported in units of 1/s. K_m values are reported in units of μM for 4MU assays and % w/v for chitO assays.

	AMCCase		Chitotriosidase	
	k_{cat}	K_m	k_{cat}	K_m
4MU-chitobiose	1.02 ± 0.05	28.12 ± 3.97	1.10 ± 0.12	68.02 ± 14.99
4MU-chitotriose	0.507 ± 0.117	24.32 ± 11.31	0.330 ± 0.029	25.00 ± 4.44
ChitO	0.908 ± 0.036	0.0176 ± 0.002	0.854 ± 0.028	0.0220 ± 0.0020

Effects of Human Asthma-Associated Mutants in the Mouse context

A trio of mutations in AMCase in humans, N45D, D47N, and R61M, confer significant increased activity to AMCase.^{16,26} In all three cases, the identity of the mutated residues becomes the same as the identity of the residues in the mouse context. In order to better understand the mutational landscape between the mouse and human enzymes, we sought to make the reverse mutations in the mouse background and quantify their effect on activity with 4MU-chitobiose as well as with bulk chitin. First we measured the activity of the mutations using 4MU-chitobiose, which gave largely similar results in the mouse background to what has previously been described in humans, with the majority of the activity difference being driven by the M61R mutation driving a significant decrease in k_{cat} and a small increase in K_m (**Figure 3.3, Table 3.3**). Smaller effects were observed for the individual D45N and N47D mutations, but the effects were reversed by the charge-swapped D45N/N47D construct, and the full triple mutant was the least active. These results suggest that the different residue identities have overall very similar effects in the mouse and human backgrounds, and that the low-activity wild-type identities in primates may have arisen due to reduced selective pressure for chitinase activity around pH 7. To understand how well these effects observed with the oligomeric substrate transferred to the case of bulk chitin, we assayed the activity of the mutations in the mouse context using the enzyme coupled chitO assay. For the most part, the results were similar, with the largest effect of any individual mutation and the majority of the effect of the triple mutation achieved by the M61R mutant (**Figure 3.3, Table 3.3**). The effects of the D45N and N47D mutations were less pronounced in the chitO background, while the M61R mutation had a similar effect on both K_m and k_{cat} .

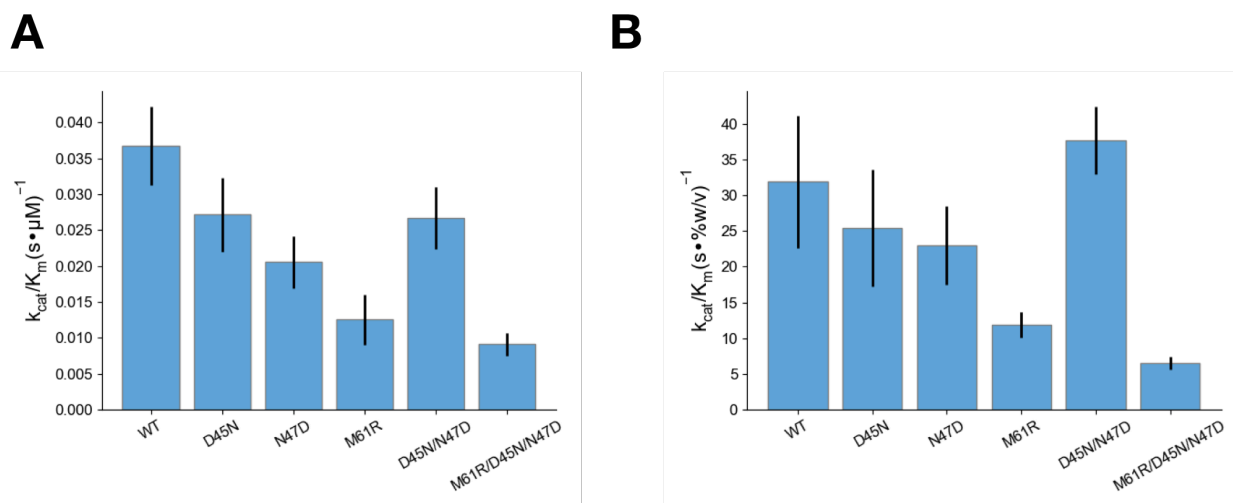


Figure 3.3 | Comparison of activity of AMCCase asthma-associated mutants. Measurement of k_{cat}/K_m for reversed asthma-associated mutants in the mouse background using the (a) 4MU-chitobiose and (b) chitooligosaccharide oxidase coupled peroxidase assays.

Table 3.3 | Calculated rate constants for AMCCase and chitotriosidase
 k_{cat} values are reported in units of 1/s. K_m values are reported in units of μM for 4MU-chitobiose and % w/v for chitO assays.

	4MU-Chitobiose		Chitooligosaccharide Oxidase	
	k_{cat}	K_m	k_{cat}	K_m
WT	1.09 ± 0.10	29.6 ± 3.4	1.02 ± 0.17	0.0321 ± 0.0077
D45N	0.94 ± 0.07	34.7 ± 6.0	0.80 ± 0.14	0.0316 ± 0.0084
N47D	0.71 ± 0.05	34.6 ± 5.6	0.74 ± 0.03	0.0320 ± 0.0076
M61R	0.76 ± 0.15	60.3 ± 11.5	0.48 ± 0.05	0.0406 ± 0.0045
D45N/N47D	0.98 ± 0.09	36.8 ± 4.9	1.02 ± 0.04	0.0272 ± 0.0032
D45N/N47D/M61R	0.46 ± 0.05	50.3 ± 7.0	0.31 ± 0.04	0.0469 ± 0.0056

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 knocked down²⁴. We sought to investigate whether we could improve the activity of mouse AMCCase in order to improve its efficacy in replacing or improving lung chitinase activity. We used error prone PCR to generate libraries of mAMCCase mutants, and expressed individual mutants in 96-well format (**Figure 3.4**). Our recombinant expression approach, utilizing periplasmic secretion as described previously, also results in significant enzyme secreted into the media.²⁷ Taking

advantage of this, we assayed the ability of the spent media of individual mutants after protein expression to cleave 4MU-chitobiose. 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 3.4**). Because these assays were done directly on spent media, they result from changes to the specific activity of the enzyme as well as to expression and secretion efficiency with different mutants. In order to determine whether our results represented real improvements in activity, we isolated and purified the two most active mutants:

A239T/L364Q (**Figure 3.4, orange**) was the most active mutant identified, with a 5-fold improvement in activity, and V246A (**Figure 3-4, blue**), which showed a 2-fold improvement in activity. After purification, we measured the specific activity of the assay with the 4MU-chitobiose substrate, and replicated the improvements we saw in the screening condition (**Figure 3.4**). The A239T/L364Q mutant improved significantly both in K_m and k_{cat} (**Table 3.4**). These results can be rationalized by the locations of the mutations in the structure. A239T and V246A are relatively distant from the active site, but L364Q is positioned immediately at the binding site for chitin, and may even coordinate the 4MU moiety in the fluorogenic substrate (**Figure 3.4**). Next, we investigated whether the activity increases observed with the 4MU-chitobiose 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 3.4**). The loss of activity of the double mutant suggests that the improvements were driven by the L364Q mutation interacting with the 4MU fluorophore. 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 at the stage of screening. Our attempts at directly screening activity of media with the chitO assay thus far have not been successful, due to a high background signal in the absence of active enzyme.

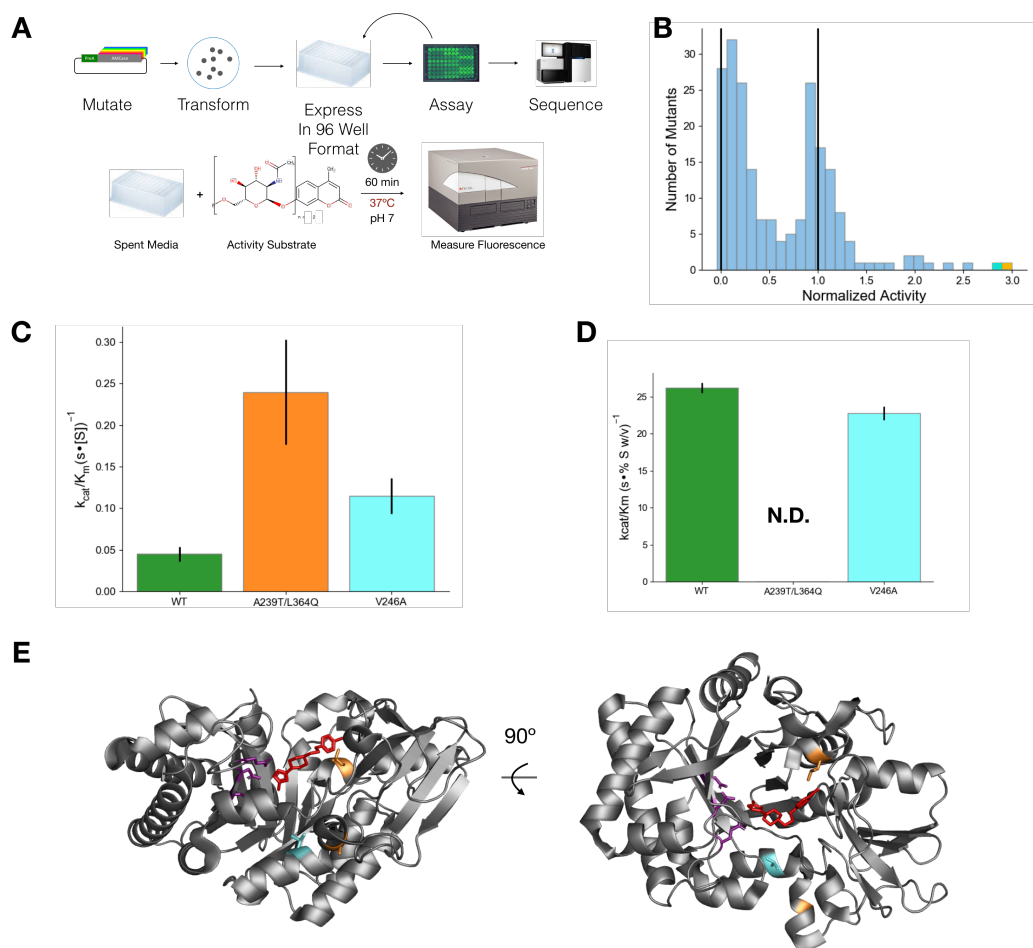


Figure 3.4 | Engineering of hyperactive AMCase mutants.

(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 chitobiose 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 blue and orange. (c) k_{cat}/K_m of purified hyperactive mutants using the 4MU-chitobiose assay and (d) the chitO assay. (e) Structure of AMCase catalytic domain highlighting A239T/L364Q (orange) and V246A (blue). The active site catalytic network is highlighted in purple, and an inhibitor that binds to the active site cleft is shown in red.

Table 3.4 | Comparison of calculated rates for engineered mutants.

k_{cat} values are reported in units of 1/s. K_m values are reported in units of μM for 4MU-chitobiose and % w/v for chitO assays.

	4MU-Chitobiose		Chitooligosaccharide Oxidase	
	k_{cat}	K_m	k_{cat}	K_m
WT	1.48 ± 0.26	32.97 ± 11.91	0.779 ± 0.048	0.0297 ± 0.0023
A239T/L364Q	4.46 ± 1.17	18.6 ± 9.43	N.D.	N.D.
V246A	3.62 ± 0.73	31.5 ± 11.8	0.800 ± 0.073	0.0351 ± 0.0042

Discussion

Broadly, these results demonstrate the value of pushing to quantify 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 to monitoring chitinase activity. While in some cases the results of the activity assays were more or less unchanged relative to the 4MU-chitobiose assays, in others the activities were tremendously different, underscoring the need for quantitative measures of bulk chitin catabolism. This proved to be particularly true both 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 a variety of experiments teasing apart the components driving enzyme activity.

Using the new bulk activity measurements, we were able to discern a k_{cat} to K_m tradeoff 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 ($p=0.02$), while k_{cat} decreased from 0.944

± 0.111 1/s to 0.540 ± 0.083 1/s ($p=0.007$). 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, ¹⁴ and this result appears to support 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 that proves to be true, given the closely matched k_{cat} for 4MU-chitobiose, in which processivity is not possible and can't be rate limiting, and bulk chitin, it appears that the rates of these two components of enzyme activity may be similar in mice. This may be a result of selection maximizing the overall rate of the enzyme, or may impact the relative size of products generated by the enzyme, since the likelihood of multiple sliding events between cutting events would be greater than if decrystallization and sliding were much slower than the rate of hydrolysis. Alternatively, 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. The secondary domain may also impact the stability of folding of the catalytic domain, leading to artificial reduction of measured k_{cat} by reducing the effective concentration of active sites in catalysis-competent conformations. Further studies may benefit by studying the effect of the secondary domain on activity in other chitinases, or using chimeras between different chitinase sequences to alter the sequence and function of the linker and carbohydrate binding domains. Beyond this, since both mammalian chitinases are expected to be natively glycosylated, the carbohydrate binding domain may create protein-protein interactions that could create assemblies that may be important for processing chitin by maximizing concentrations where crystallinity is breaking down.

Using the chitO assay, we are able to reliably tease out binding and catalysis effects for chitinases with real substrates, and compare them to the measurement we make with commercial oligomeric substrates. Qualitatively, the agreement between the 4MU and bulk chitin assays was very strong, 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 chitobiose and 0.03% w/v for the bulk chitin assay. Under the approximation of infinite polymer length, there is one potential binding site per monomer of n-acetylglucosamine. 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. Given the expectation that Acidic Mammalian Chitinase would be able to accommodate up to a hexamer of chitin in its binding cleft, we would expect that the actual binding should be tighter than with the shorter oligomers. 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, co-application of chitin binding enzymes that might loosen the crystalline geometry, or physical milling to alter the surface area to volume ratio, in order to test this hypothesis. Testing colloidal chitin generated from alternative sources, which may have different polymer crystallinity as well as average polymer length may also reveal new information about the mechanisms of chitin processing by chitinases.

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-chitobiose 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 and 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, as well as the smaller substrate overall. The result underscores the need in the future for utilizing frequent counter-screening with bulk chitin when performing selection experiments for chitin processing in the future. One challenge to accomplishing this is that, while the chitO assay is more sensitive and high throughput than previous techniques, it is sensitive to free sugars and other components of the media that limits its utility for direct screening. It is possible that with small scale purification using nickel resin, 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 learning,³⁰ we may be able to effectively identify use this approach to find activation mutants without the requirement of using chitobiose substrates.

Looking beyond this work developing methods to understand chitin hydrolysis in bulk, the methods developed here can give information about binding and catalysis with real substrates, but questions still remain about processivity, endo vs exo preference, and potential clustering and cooperative behavior between multiple enzymes. In order to fully characterize these aspects of catalysis, future work will require single-molecule measurements of kinetics. Recently, significant progress has been made in measuring chitinase activities by single molecule microscopy,^{14,31,32} 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.

Methods

Protein Preparation

Constructs expressing a fusion of a protein A secretion sequence targeting periplasmic expression, AMCase or chitotriosidase, and a c-terminal V5-his-tag as previously described²⁷ were ordered from Atum. Mutants of AMCase were generated via PCR mutagenesis. Chitinase-containing plasmids were transformed into BL21 cells and expressed overnight in ZY Autoinduction media at 37°C for 3 hours 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, 20 mM Tris pH 6.5, 1mg/mL lysozyme, 1 µL universal nuclease, with a protease inhibitor tablet. The resuspended cells were incubated at 37°C for 1 hour, then pelleted via centrifugation at 15000*g for 15 minutes. 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 minutes at 4°C. The cells were centrifuged at 15000*g for 15 minutes 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, 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 based on an activity assay with commercial fluorogenic substrate.

Analysis of Kinetic Data

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

$$A(1 - e^{-k_1t}) + B$$

$$A(1 - e^{-k_1t}) + B(1 - e^{-k_2t}) + C$$

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 free and open source 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). The product of A and k_1 yields a rate appropriate for k_{cat}/K_m determination. 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 chitobiose and 4-methylumbelliferyl chitotriose as described previously³³ with one critical modification. 10 nM chitinase enzyme was incubated with varying concentrations of 4MU-chitobiose or 4MU-chitotriose up to 433 μM in McIlvaine Buffer³⁴ pH 7 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. In order to avoid noise introduced by quenching and substrate concentration, we instead 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 **Figure S3.1**.

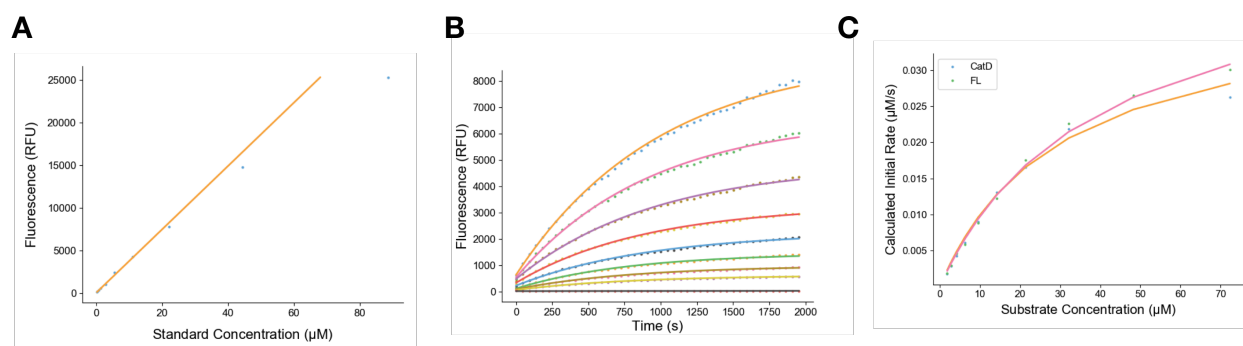


Figure S3.1 | Processing 4MU assay data

(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-chitobiose 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.

Bulk Clearance Activity Assay

Borohydride-reduced colloidal chitin was purchased as a powder from Megazyme and resuspended to 4% w/v in pH 7 McIlvaine buffer. Higher concentrations did not stay in suspension effectively. In order to remove soluble chitin oligomers, the suspension was pelleted by centrifugation at 3200*g, the supernatant was discarded, and the pellet was resuspended in McIlvaine buffer. This wash step was repeated a total of 5 times. A concentration series was prepared by serial dilution of this washed 4% stock in McIlvaine buffer, and 50 μ L of each substrate concentration was incubated with 50 μ L of 200 nM chitinase at 37C 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 OD680 for 72 hours with shaking between reads to maintain substrate suspension. The processing for data collected from this assay is illustrated in **Figure S3.2**.

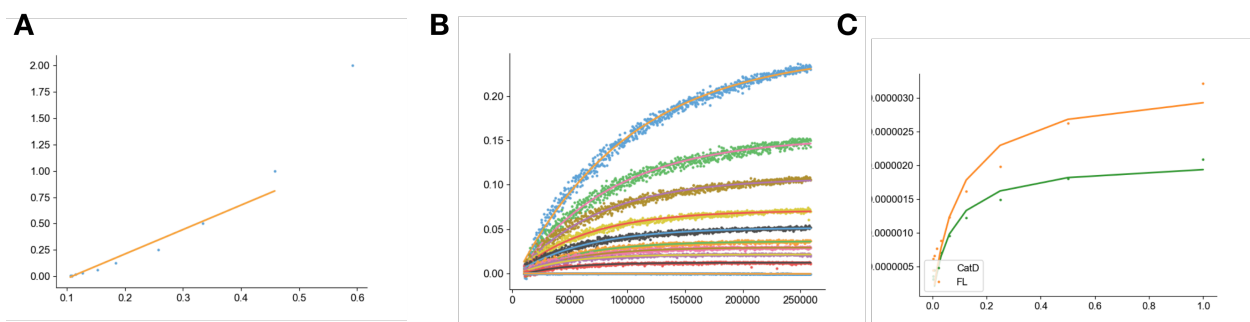


Figure S3.2 | Data processing for colloidal chitin clearance assay.

(a) Concentrations from enzyme-free controls were matched to absorbance, and for concentrations below 0.5% 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.

Potassium Ferricyanide Reduction Assay

4% w/v colloidal chitin was washed as above, then diluted serially to generate a concentration range from was incubated with 1-100nM chitinase for up to 18 hours at 37°C. At the endpoint of incubation, 50 μ L of reaction mixture was quenched by the addition of 100 μ of 400 mM sodium carbonate. The insoluble chitin was pelleted by centrifugation at 4000*g, then 100 μ L of supernatant was mixed with 100 μ L 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 hours at 42°C to maximize the rate of the non-enzymatic reduction of potassium ferricyanide by solubilized reducing sugars. During incubation absorbance at 420nm was read out in 1 minute intervals. We found that progress curve analysis gave 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 hour incubation with chitinase. The processing of the data for this assay to generate rates is illustrated in **Figure S3.3**.

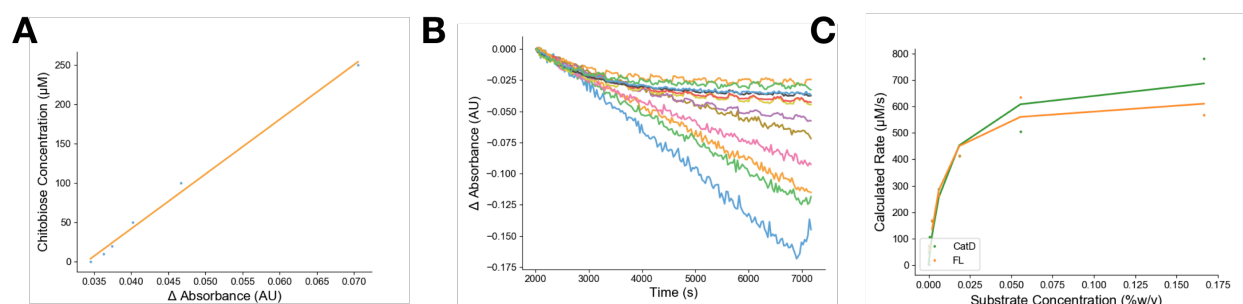


Figure S3.3 | Data processing for ferricyanide reduction assay.

(a) Concentrations from chitobiose controls were matched to absorbance, and for concentrations below 250 μ 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.

Chitooligosaccharide Oxidase Coupled Peroxidase Assay

Processing of colloidal chitin and resultant generation of new reducing sugar moieties was monitored, as previously described,¹⁷ 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.³⁵ ChitO was purchased from Gecco Biotech, HRP and quantared substrate were purchased from Sigma. 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, 100nM ChitO, 0.5 μ L of quantared substrate, and 10 μ L of quantared enhancer solution in McIlvaine buffer pH 7 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 minute intervals for 16 hours. 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 significant signal that was modulated by the washing of the colloidal chitin. This enzyme coupled reaction proved to be very 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 **Figure S3.4**.

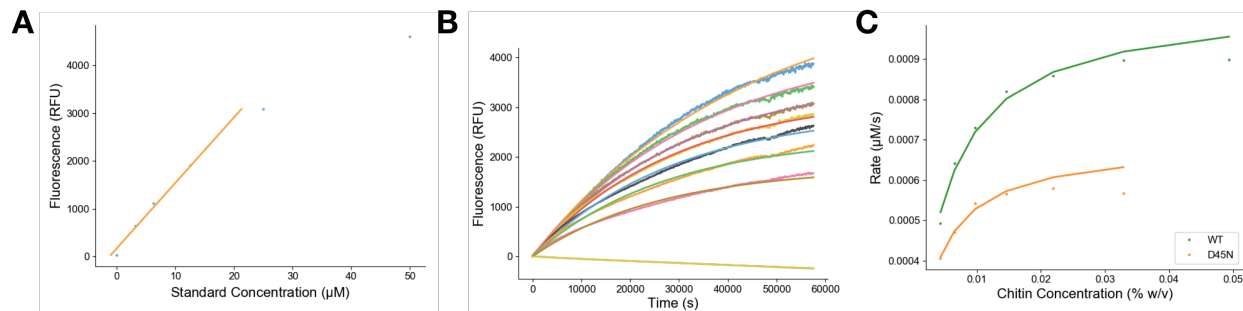


Figure S3.4 | Data processing for chitO assay.

(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.

Random Mutagenesis and Screening

Random mutations were generated using the commercial Genemorph II random mutagenesis kit.³⁶ Briefly, 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 quik 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, 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-chitobiose in McIlvaine buffer pH 7, 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 pooler for further error prone PCR and screening.

References

1. Ball, S. G. & Morell, M. K. From bacterial glycogen to starch: understanding the biogenesis of the plant starch granule. *Annu. Rev. Plant Biol.* 54, 207–233 (2003).
2. Liu, A.-J., Song, W., Yang, N., Liu, Y.-J. & Zhang, G.-R. Cartilage polysaccharide induces apoptosis in human leukemia K562 cells. *Cell Biol. Toxicol.* 23, 465–476 (2007).
3. Cabib, E. & Bowers, B. Timing and function of chitin synthesis in yeast. *J. Bacteriol.* 124, 1586–1593 (1975).
4. Davies, G. & Henrissat, B. Structures and mechanisms of glycosyl hydrolases. *Structure* 3, 853–859 (1995).
5. Hedegård, E. D. & Ryde, U. Molecular mechanism of lytic polysaccharide monooxygenases. *Chem. Sci.* 9, 3866–3880 (2018).
6. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555 (2008).
7. Diener, M. et al. Primary, Secondary, Tertiary and Quaternary Structure Levels in Linear Polysaccharides: From Random Coil, to Single Helix to Supramolecular Assembly. *Biomacromolecules* 20, 1731–1739 (2019).
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. Kim, Y. H., Park, S. K., Hur, J. Y. & Kim, Y. C. Purification and Characterization of a Major Extracellular Chitinase from a Biocontrol Bacterium, *Paenibacillus elgii* HOA73. *Plant Pathol. J.* 33, 318–328 (2017).
10. Kimura, M. et al. Functional Properties of Mouse Chitotriosidase Expressed in the Periplasmic Space of *Escherichia coli*. *PLoS One* 11, e0164367 (2016).

11. 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).
12. Jang, M.-K., Kong, B.-G., Jeong, Y.-I., Lee, C. H. & Nah, J.-W. Physicochemical characterization of β -chitin, γ -chitin, and δ -chitin separated from natural resources. *J. Polym. Sci. A Polym. Chem.* 42, 3423–3432 (2004).
13. Rinaudo, M. Chitin and chitosan: Properties and applications. *Prog. Polym. Sci.* 31, 603–632 (2006).
14. 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, 3814 (2018).
15. 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 (2017).
16. Okawa, K. et al. Loss and Gain of Human Acidic Mammalian Chitinase Activity by Nonsynonymous SNPs. *Mol. Biol. Evol.* 33, 3183–3193 (2016).
17. 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).
18. 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).
19. Boot, R. G. et al. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *J. Biol. Chem.* 276, 6770–6778 (2001).

20. Renkema, G. H. et al. Synthesis, sorting, and processing into distinct isoforms of human macrophage chitotriosidase. *Eur. J. Biochem.* 244, 279–285 (1997).
21. 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).
22. Steentoft, C. et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.* 32, 1478–1488 (2013).
23. Reese, T. A. et al. Chitin induces accumulation in tissue of innate immune cells associated with allergy. *Nature* 447, 92–96 (2007).
24. Van Dyken, S. J. et al. Spontaneous Chitin Accumulation in Airways and Age-Related Fibrotic Lung Disease. *Cell* 169, 497–509.e13 (2017).
25. 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).
26. 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).
27. 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).
28. Cole, S. W. The determination of reducing sugars by titration of ferricyanide. *Biochem. J.* 27, 723–726 (1933).
29. 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).

30. 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).
31. 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).
32. Fukuda, S. et al. High-speed atomic force microscope combined with single-molecule fluorescence microscope. *Rev. Sci. Instrum.* 84, 073706 (2013).
33. 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).
34. McIlvaine, T. C. A BUFFER SOLUTION FOR COLORIMETRIC COMPARISON. *J. Biol. Chem.* 49, 183–186 (1921).
35. QuantaRed Enhanced Chemifluorescent HRP Substrate Kit - Thermo Fisher Scientific. Available at: <https://www.thermofisher.com/order/catalog/product/15159>. (Accessed: 5th June 2019)
36. Random Mutagenesis Kit, Random Mutagenesis Kit | Agilent. Available at: <https://www.agilent.com/en/product/mutagenesis-cloning/mutagenesis-kits/random-mutagenesis-kits/genemorph-ii-233115>. (Accessed: 10th June 2019)

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

6/13/19

Date