Title

The role of protein, ligand and solvent conformational heterogeneity in protein function

Author

Woldeyes, Rahel A

Publication Date

2017

Woldeyes RA PhD Thesis

The role of protein, ligand and solvent conformational heterogeneity in protein function

by

Rahel A. Woldeyes

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry and Chemical Biology

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

[redacted]

To my family

Acknowledgements

First of all, I would like to express my deepest gratitude to my family. I cannot thank you enough for instilling in me values that served me so well in graduate school. Thank you for your unwavering support. I could not have completed my PhD without you; literally and figuratively :) To Anthony, thank you for your support and for just being amazing.

Thank you to the past and present membranes of the Fraser lab. I would like to thank Jaime for his support and guidance. I can’t say I won’t miss our presentation practice sessions. I would also like to thank professor Daniel Keedy for his insightful comments and delightful puns. To David Mavor and Andrew VanBenschoten, my class/labmates, thank you for joining the lab at the same time as me. It was a pleasure spending much of grad school with you. To Lin, Ben/Justin, Erin, Brandi, Tommy, Mike, Alex, and Kazu, thank you all for your help, your advice and most importantly for creating a great environment to work in. Also, I would like to thank my intern Saul Cortez. I had a lot of fun working with you. To Ken Hallenbeck, it was fun writing a paper with you. Thank you for being a great collaborator.

To my dearest friends, Si-Han Chen, Diego Garrido Ruiz and Sara Calhoun thank you for all the great times we had together and will continue to have. I would also like to thank my CCB, Biophysics and BMI classmates.

As a graduate student, I had the pleasure of working with and learning from many scientist. I would like to thank Aaron Brewster, Nick Sauter, Takanori Nakane, Mona Uervirojnangkoorn and Ben Barad for support with XFEL data processing. I am grateful

to Mark Kelly, Greg Lee and Chris Waddling for their support with NMR and X-ray crystallography experiments.

Last but not least, I would like to thank my mentors at UCSF and UMN-TC. Thank you to Bob Stroud and Charly Craik. I met both of you during interviews for grad school. Since then you have been my go-to people for advice. I truly appreciate all your help and support. I would also like to thank Danica Galonić Fujimori and Carol Gross for sharing their insights and for their guidance. I would like to thank my thesis committee members: Robert stroud, John Gross and Brian Shoichet. Thank you for providing valuable feedbacks . I am also grateful to my qualifying exam committee members, Andrej Sali and Natalia jura. Thank you for your helpful comments. I would also like to thank Alan Frankel and Michelle Arkin for great rotation and collaborations. I am grateful to my undergraduate advisors: Gianluigi Veglia, Margaret Titus, David Largaespada, Colin Campbell and Daniel Romero for providing me with the opportunity to fall in love with research. I would also like to thank UCSF and UMN-TC for creating a nurturing environment.

Contributions

Many of these chapters describe studies that have been published in the journals shown in the title pages. The works were done in a collaborations with co-authors of each of chapters. The first chapter ***E pluribus unum*, no more: from one crystal, many conformations.** Rahel wrote the manuscript, which appeared in Current Opinion in Structural Biology. **Probing the conformational dynamics of CypA using X-ray free electron lasers.** Rahel participated in the beamtime, processed the data, and contributed to the figures/manuscript of the paper, which appeared in eLife. **Divergent conformational dynamics controls allosteric ligand accessibility across evolutionary related αI-domain containing integrins.** Rahel conceived of the project, collected and analyzed data, and wrote the manuscript, which is not yet published. **Using XFELs to visualize solvent in the influenza A M2 Proton Channel.** Rahel coordinated the beamtime, processed the data, and contributed to the figures/ manuscript of the paper, which appeared in PNAS. She additionally is preparing a second manuscript where she conceived of the analyses and wrote the initial draft, which is not yet published.

The work of Rahel Woldeyes described herein was done under the supervision of James Fraser and meets the requirements for a standard PhD dissertation.

The role of protein, ligand and solvent conformational heterogeneity in protein function

Rahel A. Woldeyes

Abstract

The large scale motions that are necessary for the function of many proteins are controlled by atomic level fluctuations. Understanding what causes these fluctuations and how they are translated to the larger motions is essential for understanding the basic mechanism of many biological systems. The ability to control the fluctuations in a predictable way would significantly improve our rational drug and protein design efforts. But, often the important fluctuation that control conformational of proteins are not known. This is in part due to lack of experimental techniques that bridge these scales. Furthermore, computational efforts to meet this need are a work in progress. This thesis describes efforts to understand the mechanistic basis of large-scale motions in different biological systems by developing and applying new tools for the measurement of conformational dynamics. We probed the conformational dynamics of the enzyme CypA using X-ray free electron lasers (XFELs) to validate a known allosteric pathway of the enzyme with this radiation damage free technique. We studied differences in conformation dynamics and allosteric ligand accessibility in αI-domain containing integrins using NMR and room temperature crystallography. By comparing the conformational heterogeneity of two homologous integrins, we have found

evidence that the conformational landscape critically influences their ability to bind allosteric modulators. Finally, we visualized the solvent in the influenza M2 proton channel using XFELs in order to understand its mechanism of proton conduction. Taken together, our work highlights the essential role that conformational heterogeneity plays in the function of disparate biological systems, including the atomic-level motions that enable allosteric control of enzyme activity in CypA, the dependence of ligand binding on the conformational heterogeneity of LFA-1, and the solvent conformational hetrogenity that allows proton conduction through the M2 channel. This work is a significant advance towards a mechanistic understanding of the basis of conformational dynamics in the systems described, and will enable future work to manipulate these biological systems to fight disease and improve human health.

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 1

Chapter 2

Chapter 4

List of Supplementary Materials

Chapter 2

Chapter 3

Chapter 1

E pluribus unum, no more: from one crystal, many conformations

Citations:

Woldeyes RA, Sivak DA, Fraser JS. ***E pluribus unum, no mor*****e: From one crystal, many conformations**. *Curr Opin Struct Biol*. 2014 Oct;28:56-62

Summary

Several distinct computational approaches have recently been implemented to represent conformational heterogeneity from X-ray crystallography datasets that are averaged in time and space. As these modeling methods mature, newly discovered alternative conformations are being used to derive functional protein mechanisms. Room temperature X-ray data collection is emerging as a key variable for sampling functionally relevant conformations also observed in solution studies. Although concerns about radiation damage are warranted with higher temperature data collection, ‘diffract and destroy’ strategies on X-ray free electron lasers may permit radiation damage-free data collection. X-ray crystallography need not be confined to ‘static unique snapshots’; these experimental and computational advances are revealing how the many conformations populated within a single crystal are used in biological mechanisms.

Introduction

Macromolecular X-ray crystallography measures averaged intensities diffracted from ~10^{13} molecules in the crystal lattice. The resulting electron density map, which is used to locate the positions of atoms in the unit cell, is therefore an ensemble-averaged probability distribution. Traditionally, structural models are built into the highest peaks of the electron density distribution. The lower electron density values that surround these high signals are fit by the B-factor (the temperature factor, thermal factor, Debye-Waller factor, or atomic displacement parameter) [1] (**Figure 1.1a**), which models the isotropic fall-off of the density from the mean position as a Gaussian. Although the electron density distribution around each atom is often anisotropic, the additional parameters needed for anisotropic B-factors require a higher number of observations that is only afforded by very high resolution data [2].

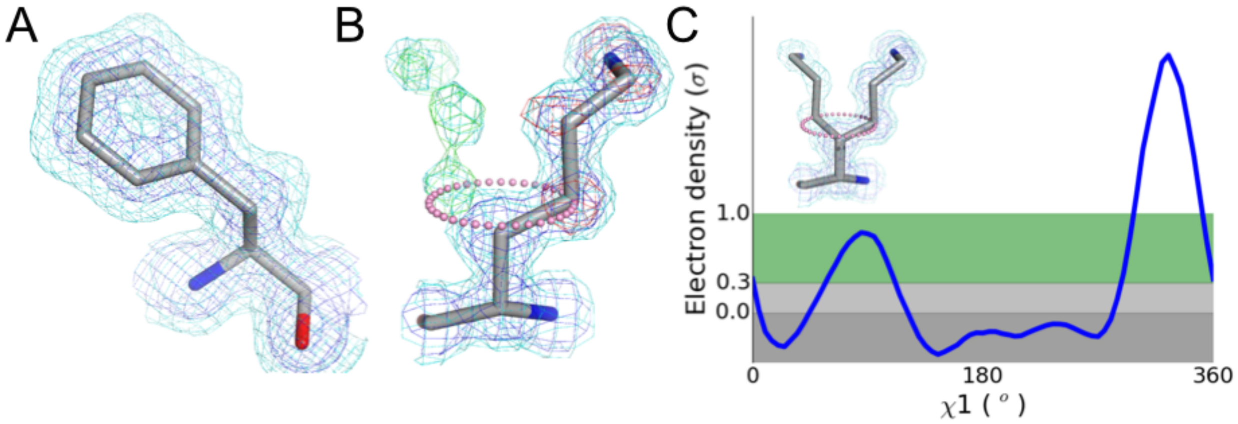


Figure 1.1 - Electron density maps contain ensemble-averaged information about multiple protein conformations.

(A) The spread in electron density (blue mesh, 1σ high contour; cyan mesh, 0.5 σ low contour) around each atom is approximated by a B-factor, which models the thermal motion as a Gaussian displacement about the mean position. (**B)** An electron density map with multiple maxima (blue mesh, high contour; cyan mesh, low contour) is inadequately modeled by B-factors, resulting in difference map peaks (red mesh, -1.5σ Fo-Fc peak underlying the model; green mesh, +1.5σ Fo-Fc peak indicating potential placement of alternative conformations). Because the alternative conformation partially overlaps with the primary conformation and is at lower occupancy, it is not visible at high contour. (**C)** Sampling the electron density around the χ_1 dihedral angle of the map shown in (b) (pink dots) reveals the presence of a distinct peak at the rotameric angle of -60°, providing an anchor point for manual model building of an alternative conformation. Automated model building is further complicated by the potential for backbone motions that can shift these peaks out of rotameric angles.

Additionally, even in the crystal lattice, proteins can adopt multiple discrete conformations [3]. Both anharmonic motion and static disorder can result in multiple relatively weak peaks in electron density maps [4]. Even individual anisotropic or grouped Translation Libration Screw (TLS) B-factors are an insufficient description of these multiple minima in the probability distribution [5-7] (**Figure 1.1b**). This review focuses on the emerging strategies for modeling conformational heterogeneity from X-ray data and the potential for modeled alternative conformations to generate new mechanistic insights into the function of macromolecules.

Electron density maps: More than meets the eye

Local maxima in electron density maps can be difficult to identify visually when electron density maps are rendered at a single threshold, as is common in the isosurface wire-frame representation used by Coot [8] and other graphics programs. Multiple contours or colour maps may be preferable for identifying conformations in weak, irregular electron density [9]. Maps modified by applying local feature enhancement [10], maximum entropy principles [11], and B-factor sharpening [12] may also aid in visually identifying important alternative conformations.

An alternative approach to visual inspection of electron density maps involves plotting electron density distributions as a function of dihedral angle (**Figure 1.1c**) [13]. Most side chain alternative conformations are confined to preferred rotameric torsion angles. The program Ringer identifies peaks originating from discrete alternative conformations sampled by protein side chains. Recent work demonstrates that these secondary peaks preferentially occur at low-energy rotameric positions [14]. Placing maps on an absolute electron-density scale and using new estimates of the noise throughout the map increases the power of Ringer to identify alternative conformations of protein and ligand atoms [15].

Transforming electron density into models reflecting heterogeneity

Performing multiple independent refinements from a slightly perturbed starting model can reveal conformational heterogeneity (**Figure 1.2a**). The starting diversity can be generated by multiple simulated annealing trajectories [16-18], alternative parameter sets [19], Monte Carlo sampling [20], or randomly-seeded automated rebuilding (using ARP/wARP [21], RAPPER [22], or Phenix Autobuild [23]). Blundell and colleagues interpreted the variability across multiple RAPPER rebuilding runs as yielding a “potentially more accurate representation of the true underlying structure than does a single model” [23,24]. However, Terwilliger later showed, using synthetic data and Phenix Autobuild, that repetitive rebuilding often does not accurately sample known conformations, but rather gives an estimate of the imprecision of the coordinates of the main conformation [25].

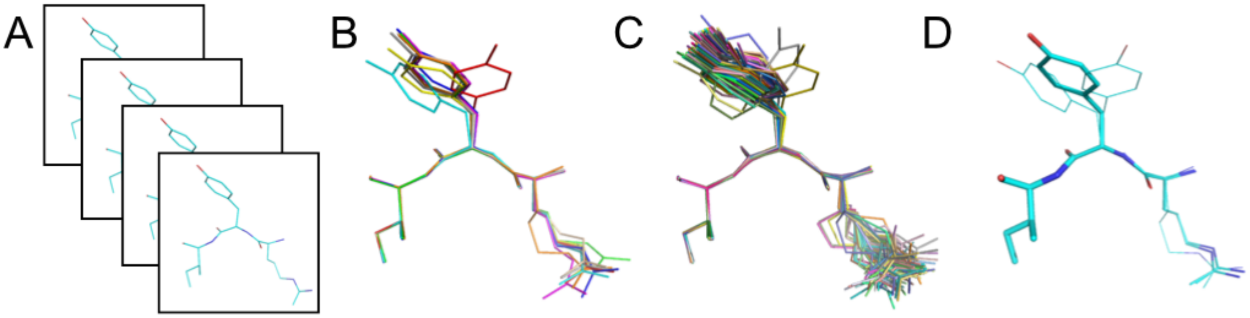


Figure 1.2 - Different model types are being used to interrogate conformational heterogeneity.

(A) In multiple independent refinements, each copy contributes to a distinct set of model structure factors. The distinct structures, separated here by boxes, yield an estimate of the relative precision of the refinement method. (**B)** In multi-copy ensemble refinement, a set number of copies of the protein, represented here by different colors, are refined together. (**C)** Similarly, in time-averaged ensemble refinement, multiple copies of the protein are selected from an MD simulation where the structure factors are averaged over a defined time window. (**D)** In multiconformer approaches, an optimal combination of between 1 and 4 conformations with associated occupancies (represented here by the thickness of the sticks) is constrained to sum to unit occupancy for each residue.

Averaging the structure factors from independently refined structures often results in a lower R_{free} than that from any individual refinement [25]. This result suggests a logical extension of the multiple independent copies approach: multiple conformations that are simultaneously refined together as an ensemble (**Figure 1.2b**). In an early application, Kuriyan showed that “twin” conformations, moving independently while contributing equally to the refinement, could improve R-factors relative to a single model [26]. But why stop at “twin” refinement? Many subsequent ensemble-modeling efforts scanned 2-20 copies of the protein [27-30]. In a recent, more comprehensive evaluation of ensemble refinement, Phillips and colleagues scanned ensembles of 2, 4, 8, and 16 copies across 50 structural genomics targets and selected a final ensemble based on the lowest R_{free} [31]. Their ensembles averaged 10.6 copies and improved R_{free} by 1.9%, suggesting that the conformations sampled across the ensemble agree well with the X-ray data. Although the most obvious improvements generally occurred in going from 1 to 2 or from 2 to 4 copies, the diminishing returns do not preclude moving beyond 16 copies. Could having even more structures further improve R_{free}?

Time-averaging can slow the increase in effective free parameters as the number of copies increases: the samples in the ensemble are not truly independent as they are generated by a single molecular dynamics simulation restrained by (time-averaged) agreement with X-ray structure factors (**Figure 1.2c**) [32]. Although any individual snapshot is generally a poor fit to the observed X-ray data, the agreement improves when averaged over many simulation snapshots. In practice, this creates attraction towards relatively undersampled regions (those with positive F_{obs}-F_{calc} difference density) and repulsion away from relatively oversampled local energy minima (those with negative F_{obs}-F_{calc} difference density). Although the original implementation, which did not include any B-factors, was susceptible to overfitting [33], including a reasonable B-factor model allowed for parallel reductions in R and R_{free} during the simulation [34-36].

Burnley and Gros have recently contributed a dramatically improved time-averaged ensemble refinement method [37], which is incorporated into the Phenix software suite [38]. To account for lattice disorder, an underlying TLS [39] B-factor model is first fit to the core of the molecule and then applied to the entire modeled structure. In addition to a bulk solvent model, explicitly-modeled solvent atoms are added and deleted throughout the simulation based on conventional real-space map criteria. In principle the restrained simulation could be extended to produce an arbitrarily large number of snapshots. Burnley and Gros keep only those simulation blocks of 250 consecutive snapshots with low R_{work} values. The still ungainly number of structures is further reduced to the final output ensemble, the smallest evenly-distributed set of structures that reproduces the time-averaged R_{free} to within 0.1%. This procedure resulted in ensembles containing between 50-800 structures across a wide variety of proteins [37]. In addition, the Gros group has recently applied this exciting method for an in-depth study of conformational heterogeneity of proteases [41].

A distinct model type, the multiconformer model, represents conformational diversity without creating multiple copies of the entire protein (**Figure 1.2d)**. In multiconformer models, if the electron density distribution for a continuous segment is well fit by a single conformation with an appropriate B-factor model, then only a single conformation is used. However, if the electron density distribution suggests discrete conformations, the heterogeneous atoms are copied and given an "ALTLOC" identifier in the PDB record. This second (or third or fourth) conformation is allowed to diverge and a new parameter "q" is refined reflecting the occupancies of the primary and secondary conformations. This tedious manual process can be subject to the whim and thoroughness of the model builder. Despite the small gains in R_{free}, modeling alternative conformations can reduce geometric distortions and rotamer-outlier side chain conformations. Fortunately, van den Bedem developed an automated approach to identify, build, and refine multiple conformations: qFit [42]. qFit enumerates many potential main and side chain conformations for each residue. Next, the program selects the optimal combination of conformations and associated initial occupancies based on combined fit to the density; often, only one conformation is selected. Fragments of neighboring residues are assembled using computational approaches borrowed from robotics to build the final model. Although correlated movement along directly adjacent residues sharing the same number of conformations is assumed, global correlated movements through tertiary contacts cannot be inferred directly from experimental data without reference to diffuse scattering features [43,44]. To address this problem, a companion approach, CONTACT, identifies networks of residues that can move between experimentally observed alternative conformations in a coupled manner [45]. In DHFR, residues with functionally relevant concerted motion (originally revealed using solution NMR experiments) were independently identified using the combined qFit - CONTACT approach. Both multiconformer and ensemble models present additional complications for validation. In particular, the use of R_{free} in the parameter optimization and validation steps highlights a potential need for a new generation of model selection criteria [40].

Warming up to different data collection strategies

As these modeling methods mature (**Table 1.1**), the major question is changing from "is there conformational heterogeneity?" to "what functions can conformational heterogeneity mediate?" Several recent studies, building on classic work by Petsko and others on the protein "glass transition" [46,47], highlight that the common practice of cryo-cooling can complicate the process of relating heterogeneity to function: cryo-cooling has obvious advantages in reducing radiation damage [48], but elegant theoretical studies by Halle suggest that the annealing that occurs during cryo-cooling may redistribute conformational heterogeneity [49]. The idea that conformational heterogeneity within the crystal can connect to biological function is dramatically exemplified by CypA [50]: data collection at room temperature allowed sampling of higher-energy conformations essential for catalysis also observed in solution by NMR experiments; in contrast, high-resolution cryogenic data revealed only a single conformational state. Similar changes to conformational ensembles were also observed across a larger sample of 30 proteins [51].

	Independent	**Multi-copy Ensemble**	**Time-averaged Ensemble**	**Multiconformer**
Overall goal	Many independent models that each independently explain the data	Fixed number of models that collectively explain the data	Ensemble of models related by dynamical simulation and fit to the data	Single model with locally-fit 1-4 conformations per segment
Diversity generation	Multi-start simulated annealing or automated rebuilding	Simulated annealing	Molecular dynamics simulation augmented with an X-ray energy term	Rotamer library extended from backbone atoms sampled along extent of anisotropic ellipsoids
Output	User-specified number of completely independent models	2-16 copies of the protein with equal occupancies, collectively contributing to F_{calc}	50-800 related models with equal occupancies, selected from the simulation	Single model with each residue having 0-3 alternative conformations, each with a partial occupancy
User-defined parameters	Rebuild fragment size, degree of cross-over between parallel models, and others depending on program	Ensemble size, B-factor model	TLS B-factor group selection, simulation relaxation time (default value determined by data resolution), X-ray:MD energy weight	Extent of sampling of backbone conformations from which to build rotamer library
Output structure selection	Retain all models within an R_{free} threshold	Vary ensemble size to minimize R_{free}	Minimize sampling frequency subject to maintaining R_{free} within 0.1% of final rolling average	Optimal combination of conformations for each residue that best explains local density features

	Independent	Multi-copy Ensemble	Time-averaged Ensemble	Multiconformer
Potential weaknesses	Biased to local energy minima; yields only an estimated precision of the refinement procedure	Low observation-to-parameter ratio for larger ensembles; occupancies fixed to number of models	Validation and interpretation of ensemble models requires new tools; coupling information may be limited	Limited backbone conformational sampling
Potential advantages	Extensive sampling of starting conformations	Sampling of anharmonic distributions	Time averaging may limit number of free parameters	Refinement of intermediate occupancies; limited number of free parameters
Implementation	RAPPER[22], phenix.autobuild[23], ExCoR[19]	CNS[31], custom scripts in other refinement protocols	phenix.ensemble-refinement[37]	qFit[42]

Table 1.1 - Comparison of approaches to address the challenge of modeling conformational heterogeneity.

However, concerns about radiation damage have historically reduced the widespread use of higher temperatures during data collection. Cryo-cooling makes complete datasets obtainable from a single crystal for many systems that otherwise could not have structures determined. Several strategies have been applied to outrun or reduce the damage at “room temperature,” where the temperature is generally maintained at 0-15°C. While it is always best to evenly expose the entire diffracting volume, the beam size is often limited by the physical setup of the beam line. In these cases, strategic translation can enable crystals to be exposed for longer and yield higher-quality data [52]. The ability to use higher dose rates and free-radical scavengers to outrun damage remains controversial, and there are many aspects of room temperature diffraction still available to optimize [53].

X-ray free electron lasers (XFELs), which can deliver short pulses of extremely large doses of X-rays [54], decouple the relationship between diffraction, radiation damage, and temperature. In “diffract and destroy” data collection strategies, diffraction occurs on a faster timescale than the radiation damage, affording an approximately radiation damage-free view of the molecule at any temperature. Indeed, differences are already being observed between the same molecules imaged at cryogenic temperatures at synchrotrons and room temperature at XFELs [55]. While XFELs provide obvious applications for viewing conformational dynamics of proteins within a crystal, there remain several roadblocks before “molecular movies” can be routinely recorded [56]. In particular, improved crystal delivery methods [57], synchronized triggering of conformational changes [58], and data processing schemes [59] are on the horizon.

Conclusions

What do the next 5-10 years hold? As the focus of our modeling efforts shift to representing the conformational ensemble, opportunities for integrative refinement and cross validation with solution experiments [60,61] will undoubtedly play a larger role. The major challenges will shift from describing conformational ensembles to understanding which of the populated conformations are important for biochemical functions. Time-resolved studies will likely be critical in this endeavor, but synchrotron-based Laue diffraction studies have previously been applied only to a limited set of systems [62-64]. The large changes in conformational ensembles often observed in the same crystal form [65,66] (**Figure 1.3**) coupled with the capabilities of XFELs for circumventing radiation damage at ambient temperatures suggest that the future will be dominated by teasing apart not only how the many conformations populated in a crystal relate to each other in space, but also how they relate to each other in time.

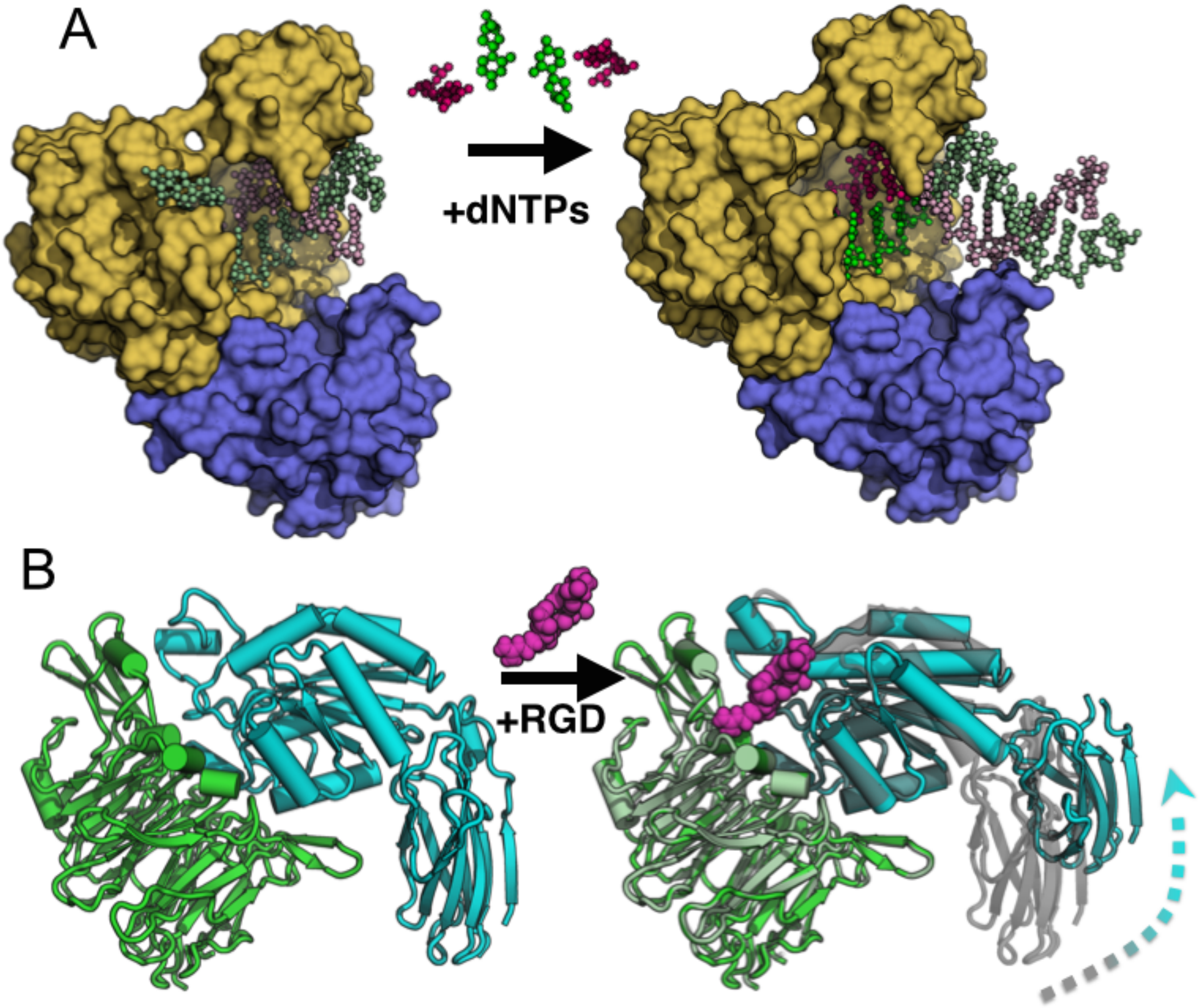


Figure 1.3 - Examples of functional conformational changes within a single crystal lattice.

(A) *Bacillus* DNA polymerase I can catalyze DNA synthesis in the crystal lattice. The initial complex (PDB: 1L3S) containing the polymerase domain (yellow) and exonuclease domain (blue) is soaked with dNTPs. During catalysis the nascent strands are extended (brighter colors) and the pre-existing strands are extruded towards the solvent channels (PDB: 1L3V). (**B**) The head domain of $\alpha_{IIb}\beta_3$ integrin (PDB: 3ZDX) undergoes a large allosteric conformational change when the ligand RGD peptide (magenta) is soaked into the crystal lattice (PDB: 3ZE2). Aligning the β-propeller domains (green) reveals how the β_3 inserted and hybrid domains extend and swing away (cyan) from the initial position (grey) upon peptide binding.

Acknowledgements

We thank Gira Bhabha, Nathaniel Echols, R. Bryn Fenwick, James Holton, Daniel Keedy, Timothy Springer, Andrew VanBenschoten, Henry van den Bedem, and Peter Wright for helpful comments. RAW is supported by an NSF Graduate Research Fellowship. DAS is supported by NIH GM081879. JSF is supported by NIH OD009180, GM110580 and NSF STC-1231306.

References

1. Jensen LH: **Protein model refinement based on x-ray data**. *Annu Rev Biophys Bioeng* 1974, **3**:81-93.
2. Merritt EA: **Expanding the model: anisotropic displacement parameters in protein structure refinement**. *Acta Crystallogr D Biol Crystallogr* 1999, **55**:1109-1117.
3. Phillips SE: **Structure and refinement of oxymyoglobin at 1.6 A resolution**. *J Mol Biol* 1980, **142**:531-554.
4. Smith JL, Hendrickson WA, Honzatko RB, Sheriff S: **Structural heterogeneity in protein crystals.** *Biochemistry* 1986, **25**:5018-5027.
5. Kuriyan J, Petsko GA, Levy RM, Karplus M: **Effect of anisotropy and anharmonicity on protein crystallographic refinement. An evaluation by molecular dynamics**. *J Mol Biol* 1986, **190**:227-254.
6. Vitkup D, Ringe D, Karplus M, Petsko GA: **Why protein R-factors are so large: a self-consistent analysis.** *Proteins* 2002, **46**:345-354.
7. Kuzmanic A, Pannu NS, Zagrovic B: **X-ray refinement significantly underestimates the level of microscopic heterogeneity in biomolecular crystals.** *Nat Commun* 2014, **5**:3220.
8. Emsley P, Lohkamp B, Scott WG, Cowtan K: **Features and development of Coot**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:486-501.
9. Maia FRNC, DeLano W, Szöke A, Van Der Spoel D: **Interactive visualization of electron density slices**. *Journal of Applied Crystallography* 2005, **38**:563-565.
10. Afonine PV: **FEM: Feature Enhanced Maps**. *Computational Crystallography Newsletter* 2013, **4(2)**:28-29.
11. Glykos NM, Kokkinidis M: **GraphEnt: a maximum-entropy program with graphics capabilities**. *Journal of Applied Crystallography* 2000, **33**:982-985.
12. Liu C, Xiong Y: **Electron density sharpening as a general technique in crystallographic studies**. *J Mol Biol* 2014, **426**:980-993.
13. Shapovalov MV, Dunbrack RL, Jr.: **Statistical and conformational analysis of the electron density of protein side chains**. *Proteins* 2007, **66**:279-303.
14. Lang PT, Ng HL, Fraser JS, Corn JE, Echols N, Sales M, Holton JM, Alber T: **Automated electron-density sampling reveals widespread conformational polymorphism in proteins**. *Protein Sci* 2010, **19**:1420-1431.
15. Lang PT, Holton JM, Fraser JS, Alber T: **Protein structural ensembles are**

revealed by redefining X-ray electron density noise. *Proc Natl Acad Sci U S A* 2014, **111**:237-242.

16. Rice LM, Shamoo Y, Brünger AT: **Phase Improvement by Multi-Start Simulated Annealing Refinement and Structure-Factor Averaging**. *Journal of Applied Crystallography* 1998, **31**:798-805.
17. Korostelev A, Laurberg M, Noller HF: **Multistart simulated annealing refinement of the crystal structure of the 70S ribosome**. *Proc Natl Acad Sci U S A* 2009, **106**:18195-18200.
18. Kohn JE, Afonine PV, Ruscio JZ, Adams PD, Head-Gordon T: **Evidence of functional protein dynamics from X-ray crystallographic ensembles**. *PLoS Comput Biol* 2010, **6**.
19. Nwachukwu JC, Southern MR, Kiefer JR, Afonine PV, Adams PD, Terwilliger TC, Nettles KW: **Improved crystallographic structures using extensive combinatorial refinement.** *Structure* 2013, **21**:1923-1930.
20. Knight JL, Zhou Z, Gallicchio E, Himmel DM, Friesner RA, Arnold E, Levy RM: **Exploring structural variability in X-ray crystallographic models using protein local optimization by torsion-angle sampling**. *Acta Crystallogr D Biol Crystallogr* 2008, **64**:383-396.
21. Perrakis A, Morris R, Lamzin VS: **Automated protein model building combined with iterative structure refinement**. *Nat Struct Biol* 1999, **6**:458-463.
22. Depristo MA, de Bakker PI, Johnson RJ, Blundell TL: **Crystallographic refinement by knowledge-based exploration of complex energy landscapes**. *Structure* 2005, **13**:1311-1319.
23. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD: **Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard**. *Acta Crystallogr D Biol Crystallogr* 2008, **64**:61-69.
24. Furnham N, Blundell TL, DePristo MA, Terwilliger TC: **Is one solution good enough?** *Nat Struct Mol Biol* 2006, **13**:184-185; discussion 185.
25. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Adams PD, Moriarty NW, Zwart P, Read RJ, Turk D, Hung LW: **Interpretation of ensembles created by multiple iterative rebuilding of macromolecular models**. *Acta Crystallogr D Biol Crystallogr* 2007, **63**:597-610.
26. Kuriyan J, Osapay K, Burley SK, Brunger AT, Hendrickson WA, Karplus M: **Exploration of disorder in protein structures by X-ray restrained molecular dynamics.** *Proteins* 1991, **10**:340-358.

27. Rader SD, Agard DA: **Conformational substates in enzyme mechanism: the 120 K structure of alpha-lytic protease at 1.5 A resolution**. *Protein Sci* 1997, **6**:1375-1386.

28. Burling FT, Weis WI, Flaherty KM, Brunger AT: **Direct observation of protein solvation and discrete disorder with experimental crystallographic phases**. *Science* 1996, **271**:72-77.

29. Wilson MA, Brunger AT: **The 1.0 A crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity**. *J Mol Biol* 2000, **301**:1237-1256.

30. Pellegrini M, Gronbech-Jensen N, Kelly JA, Pfluegl GM, Yeates TO: **Highly constrained multiple-copy refinement of protein crystal structures**. *Proteins* 1997, **29**:426-432.

31. Levin EJ, Kondrashov DA, Wesenberg GE, Phillips GN, Jr.: **Ensemble refinement of protein crystal structures: validation and application.** *Structure* 2007, **15**:1040-1052.

32. Gros P, van Gunsteren WF, Hol WG: **Inclusion of thermal motion in crystallographic structures by restrained molecular dynamics**. *Science* 1990, **249**:1149-1152.

33. Clarage JB, Phillips GN, Jr.: **Cross-validation tests of time-averaged molecular dynamics refinements for determination of protein structures by X-ray crystallography.** *Acta Crystallogr D Biol Crystallogr* 1994, **50**:24-36.

34. Schiffer CA, van Gunsteren WF: **Accessibility and order of water sites in and around proteins: A crystallographic time-averaging study**. *Proteins* 1999, **36**:501-511.

35. Schiffer CA, Gros P, van Gunsteren WF: **Time-averaging crystallographic refinement: possibilities and limitations using alpha-cyclodextrin as a test system.** *Acta Crystallogr D Biol Crystallogr* 1995, **51**:85-92.

36. Romo TD, Clarage JB, Sorensen DC, Phillips GN, Jr.: **Automatic identification of discrete substates in proteins: singular value decomposition analysis of time-averaged crystallographic refinements.** *Proteins* 1995, **22**:311-321.

37. Burnley BT, Afonine PV, Adams PD, Gros P: **Modelling dynamics in protein crystal structures by ensemble refinement.** *Elife* 2012, **1**:e00311.

38. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive Python-based system for macromolecular structure solution**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:213-221.

39. Painter J, Merritt EA: **Optimal description of a protein structure in terms of multiple groups undergoing TLS motion.** *Acta Crystallogr D Biol Crystallogr* 2006, **62**:439-450.

40. Kleywegt GJ: **Separating model optimization and model validation in statistical cross-validation as applied to crystallography**. *Acta Crystallogr D Biol Crystallogr* 2007, **63**:939-940.

41. Forneris F, Burnley BT, Gros P: **Ensemble refinement shows conformational flexibility in crystal structures of human complement factor D**. *Acta Crystallogr D Biol Crystallogr* 2014, **70**:733-743.

42. van den Bedem H, Dhanik A, Latombe JC, Deacon AM: **Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers**. *Acta Crystallogr D Biol Crystallogr* 2009, **65**:1107-1117.

43. Wall ME, Adams PD, Fraser JS, Sauter NK: **Diffuse X-ray scattering to model protein motions**. *Structure* 2014, **22**:182-184.

44. Kuzmanic A, Kruschel D, van Gunsteren WF, Pannu NS, Zagrovic B: **Dynamics may significantly influence the estimation of interatomic distances in biomolecular X-ray structures.** *J Mol Biol* 2011, **411**:286-297.

45. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS: **Automated identification of functional dynamic contact networks from X-ray crystallography.** *Nat Methods* 2013, **10**:896-902.

46. Tilton RF, Jr., Dewan JC, Petsko GA: **Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K**. *Biochemistry* 1992, **31**:2469-2481.

47. Frauenfelder H, Hartmann H, Karplus M, Kuntz ID, Jr., Kuriyan J, Parak F, Petsko GA, Ringe D, Tilton RF, Jr., Connolly ML, et al.: **Thermal expansion of a protein**. *Biochemistry* 1987, **26**:254-261.

48. Hope H: **Cryocrystallography of biological macromolecules: a generally applicable method.** *Acta Crystallographica Section B* 1988, **44**:22-26.

49. Halle B: **Biomolecular cryocrystallography: structural changes during flash-cooling.** *Proc Natl Acad Sci U S A* 2004, **101**:4793-4798.

50. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T: **Hidden alternative structures of proline isomerase essential for catalysis**. *Nature* 2009, **462**:669-673.

51. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature X-**

ray crystallography. *Proc Natl Acad Sci U S A* 2011, **108**:16247-16252.

52. Zeldin OB, Brockhauser S, Bremridge J, Holton JM, Garman EF: **Predicting the X-ray lifetime of protein crystals**. *Proc Natl Acad Sci U S A* 2013, **110**:20551-20556.

53. Warkentin M, Hopkins JB, Badeau R, Mulichak AM, Keefe LJ, Thorne RE: **Global radiation damage: temperature dependence, time dependence and how to outrun it**. *J Synchrotron Radiat* 2013, **20**:7-13.

54. Spence JC, Weierstall U, Chapman HN: **X-ray lasers for structural and dynamic biology**. *Rep Prog Phys* 2012, **75**:102601.

55. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A, et al.: **Serial femtosecond crystallography of G protein-coupled receptors**. *Science* 2013, **342**:1521-1524.

56. Neutze R, Moffat K: **Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges**. *Curr Opin Struct Biol* 2012, **22**:651-659.

57. Weierstall U, James D, Wang C, White TA, Wang D, Liu W, Spence JC, Bruce Doak R, Nelson G, Fromme P, et al.: **Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography**. *Nat Commun* 2014, **5**:3309.

58. Aquila A, Hunter MS, Doak RB, Kirian RA, Fromme P, White TA, Andreasson J, Arnlund D, Bajt S, Barends TR, et al.: **Time-resolved protein nanocrystallography using an X-ray free-electron laser**. *Opt Express* 2012, **20**:2706-2716.

59. Hattne J, Echols N, Tran R, Kern J, Gildea RJ, Brewster AS, Alonso-Mori R, Glockner C, Hellmich J, Laksmono H, et al.: **Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers**. *Nat Methods* 2014.

60. Fenwick RB, van den Bedem H, Fraser JS, Wright PE: **Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR**. *Proc Natl Acad Sci U S A* 2014, **111**:E445-454.

61. Rinaldelli M, Ravera E, Calderone V, Parigi G, Murshudov GN, Luchinat C: **Simultaneous use of solution NMR and X-ray data in REFMAC5 for joint refinement/detection of structural differences**. *Acta Crystallogr D Biol Crystallogr* 2014, **70**:958-967.

62. Aranda Rt, Levin EJ, Schotte F, Anfinrud PA, Phillips GN, Jr.: **Time-dependent atomic coordinates for the dissociation of carbon monoxide from**

myoglobin. *Acta Crystallogr D Biol Crystallogr* 2006, **62**:776-783.

63. Ihee H, Rajagopal S, Srajer V, Pahl R, Anderson S, Schmidt M, Schotte F, Anfinrud PA, Wulff M, Moffat K: **Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds**. *Proc Natl Acad Sci U S A* 2005, **102**:7145-7150.

64. Schlichting I, Almo SC, Rapp G, Wilson K, Petratos K, Lentfer A, Wittinghofer A, Kabsch W, Pai EF, Petsko GA, et al.: **Time-resolved X-ray crystallographic study of the conformational change in Ha-Ras p21 protein on GTP hydrolysis.** *Nature* 1990, **345**:309-315.

65. Johnson SJ, Taylor JS, Beese LS: **Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations.** *Proc Natl Acad Sci U S A* 2003, **100**:3895-3900.

66. Zhu J, Zhu J, Springer TA: **Complete integrin headpiece opening in eight steps.** *J Cell Biol* 2013, **201**:1053-1068.

Chapter 2

Probing the conformational dynamics of CypA using X-ray free electron lasers

Citations:

Keedy DA*, Kenner LR*, Warkentin M*, Woldeyes RA*, Hopkins JB, Thompson MC, Brewster AS, Van Benschoten AH, Baxter EL, Uervirojnangkoorn M, McPhillips SE, Song J, Alonso-Mori R, Holton JM, Weis WI, Brunger AT, Soltis SM, Lemke H, Gonzalez A, Sauter NK, Cohen AE, van den Bedem H, Thorne RE, Fraser JS. **Mapping the Conformational Landscape of a Dynamic Enzyme by XFEL and Multi-temperature Crystallography.** *eLife. 2015 Sep 30;4. pii: e07574*.

(*denotes equal contributions by the authors)

Summary

Determining the interconverting conformations of dynamic proteins in atomic detail is a major challenge for structural biology. Conformational heterogeneity in the active site of the dynamic enzyme cyclophilin A (CypA) has been previously linked to its catalytic function. Here we compare the conformational ensembles of CypA by fixed-target X-ray free electron laser (XFEL) crystallography and multitemperature synchrotron crystallography. The “diffraction-before-destruction” nature of XFEL experiments provides a radiation-damage-free view of the functionally important alternative conformations of CypA. We monitored the temperature dependences of these alternative conformations with eight synchrotron datasets spanning 100-310 K. Multiconformer models show that many alternative conformations in CypA are populated above, but not below, the glass transition temperature (~200 K) and reveal abrupt changes in protein flexibility that provide all-atom insight into conformational coupling. Together, our XFEL data and multitemperature analyses motivate a new generation of time-resolved experiments to structurally characterize the dynamic underpinnings of protein function.

Introduction

Current structural biology methods provide only incomplete pictures of how proteins can interconvert between distinct conformations [1,2]. X-ray crystallography reveals atomic coordinates with relatively high accuracy and precision. However, because X-ray diffraction is an ensemble experiment involving greater than millions of copies of the protein in one crystal, the resulting electron density maps may contain contributions from multiple alternative conformations [3–5]. At high resolution, it is often possible to detect and discretely model these alternative conformations [6–9]. Structural characterization of alternative conformations by X-ray crystallography can complement NMR [10,11] and computational simulations [12,13] in defining the role of protein dynamics in function [14].

Unfortunately, X-ray-induced radiation damage can complicate the interpretation of electron density maps [15–17]. As a result, more than 95% of crystal structures are determined at cryogenic conditions (~100 K) to minimize diffusion of reactive intermediates and any subsequent structural relaxations [15]. Although cryocooling minimizes radiation damage per unit X-ray dose, it can also modify mainchain and side chain conformational distributions, particularly near active sites and in distal regions important for allosteric function [18–20]. Recent studies have instead used a combination of room-temperature data collection and new computational tools to reveal a multitude of previously “hidden” alternative conformations, many of which can be directly linked to function [20–24]. For example, the interconversion between alternative conformations revealed by room-temperature crystallography in the active site of the proline isomerase cyclophilin A (CypA) is rate-limiting for catalysis [22]. However, radiation damage considerations limit the applicability of room-temperature crystallography only to systems that yield large and well-diffracting crystals [25,26]. To visualize the mechanisms by which perturbations such as ligand binding, mutation, and temperature shift the relative populations of alternative conformations, new technologies are needed to break the connection between radiation damage and diffraction resolution.

Fortunately, the advent of X-ray free electron lasers breaks this connection and expands our ability to observe conformational heterogeneity at room temperature while avoiding radiation damage [27,28]. XFELs deliver ultra-bright X-ray pulses, thereby allowing strong diffraction at a faster timescale than conventional radiation damage processes can occur [29]. This “diffraction-before-destruction” approach has been successful with two primary experimental configurations. In serial femtosecond crystallography (SFX), datasets are comprised of single exposures from thousands of individual nanocrystals injected from a slurry into the X-ray beam [30,31]. Room temperature SFX results have revealed important conformational differences compared to the cryogenic synchrotron structure of the human serotonin receptor GPCR [31]. The SFX technique has potential applications for monitoring irreversible reactions and overcoming additional technical limitations inherent to Laue crystallography at synchrotrons [32]. By contrast, in serial femtosecond rotation crystallography (SF-ROX), datasets are comprised of multiple exposures from each of dozens of individual macrocrystals mounted on fixed-target goniometers [33–35]. This technique more closely mirrors existing protocols at synchrotrons. However, because both the crystal is effectively stationary during data collection in both types of XFEL experiments only partial intensity measurements are recorded. In SFX this problem is overcome by massive redundancy and Monte Carlo integration [36]. Previous SF-ROX experiments carefully rotated the crystal between shots in steps that are finer than the mosaicity, which necessitates long crystals to accurately measure most reflections. However, recent advances in post-refinement can significantly reduce the number of images required for a complete dataset [37]. Additionally, previous SF-ROX experiments have used cryocooled macrocrystals, which counteracts many of the advantages of XFEL crystallography and forgoes opportunities for time-resolved experiments.

Although time-resolved experiments are a particularly promising direction for XFEL crystallography, they require a trigger to rapidly initiate the conformational change [28,32,37]. For example, time-resolved studies using laser excitation of the naturally occurring chromophore in PYP have yielded temporally resolved structural changes [38]. Because light-responsive chemical triggers that initiate the conformational changes are often system-specific, more general approaches are needed to observe how protein conformational ensembles respond to perturbations as a function of time [39]. Temperature is a promising and generalizable candidate perturbation because it modulates protein conformational distributions [18], and rapid temperature jumps can likely be triggered in crystals by laser-induced Raman excitation of the water in the solvent channels [40]. While time-dependent temperature perturbations have not been measured in protein crystals, classic work has examined the equilibrium temperature dependence of protein flexibility across individual structures determined at temperatures from ~80-320 K [41–44]. By using refined atomic B-factors as a proxy for flexibility, these studies identified a dramatic collective inflection point around 180-220 K [42,44]; however, the molecular origins of this “glass transition” remain incompletely understood and NMR studies indicate that distinct bands of motion have heterogeneous responses to temperature around the glass transition [45]. An all-atom, multi-conformer analysis of electron density maps at multiple temperatures could define whether there are collective shifts in the distributions of conformational substates and also set the stage for time-resolved structural experiments to define the rates and pathways of transitions between them.

Here, we combine room-temperature fixed-target XFEL and multi-temperature synchrotron experiments to examine the temperature-dependent conformational heterogeneity of the model enzyme CypA. First, we map the conformational distribution revealed at room temperature in the absence of radiation damage by XFEL SF-ROX diffraction using only 1,239 diffraction images and post-refinement using prime [37]. This XFEL data validates previous connections between alternative conformations imaged in the crystal and experiments performed in solution [22,46]. Next we measured the equilibrium conformational ensemble across a broad range of temperatures, revealing complex and abrupt changes as a function of temperature. Our work uncovers a heterogeneous temperature dependence of functional alternative protein conformations and suggests new ways to use XFEL crystallography to probe the dynamic underpinnings of protein function. These new XFEL capabilities will allow us to probe a wide range of radiation-damage-sensitive systems that are otherwise intractable for conventional room-temperature diffraction at synchrotrons.

Results

Fixed-Target Data Collection Using XFELs Reveals Radiation-Damage-Free Conformational Heterogeneity

To compare the distribution of alternative conformations between XFEL and synchrotron diffraction data, we collected two ambient temperature datasets: a 1.75 Å resolution radiation-damage-free dataset using serial femtosecond rotation crystallography (SF-ROX) [34,35,47] and a new 1.2 Å resolution synchrotron dataset(**Table 2.1,2.2**). For the XFEL experiment, we collected 1,239 individual frames, translating to unique unexposed regions of 71 crystals between each shot, and processed the data using *cctbx.xfel* [48]. Post-refinement in *prime* was essential for extracting a dataset with high completeness using the limited amount of images that can be recorded during short screening shifts at the LCLS [37]. Automated molecular replacement yielded interpretable electron density maps that allowed us to refine a single-conformer structural model with reasonable statistics (**Table 2.1**).

	XFEL
PDB ID	4YUP
Resolution range	43.98 - 1.75 (1.81 - 1.75)
Space group	$P2_12_12_1$
Unit cell (a, b, c)	42.42 51.82 87.96
Unique reflections	19942 (1894)
Completeness (%)	99 (96)
Wilson B-factor	21.12
Refinement resolution range	43.98 - 1.75 (1.93 - 1.75)
Reflections used in refinement	19936 (4811)
Reflections used for R-free	625 (151)
R-work (%)	20.0 (34.3)
R-free (%)	24.9 (36.1)
Number of non-hydrogen atoms	1762
Macromolecular atomcs	1559
Protein residues	164
RMS(bonds)	0.017
RMS(angles)	1.44
Ramachandran favored (%)	96
Ramachandran allowed (%)	3.6

	XFEL
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.8
Clashscore	1.92
Average B-factor	29.03
macromolecules	26.52
solvent	48.25
Number of TLS groups	3

Table 2.1: Crystallographic statistics for room-temperature XFEL dataset collected across 71 crystals. Statistics for the highest-resolution shell are shown in parentheses.

	1.2 Å synchrotron
PDB ID	4YUO
Wavelength	0.9795
Resolution range	44.60 - 1.20 (1.24 - 1.20)
Space group	P 21 21 21
Unit cell	42.9 52.43 89.11 90 90 90
Total reflections	307722 (18999)
Unique reflections	58118 (5122)
Multiplicity	5.3 (3.7)
Completeness (%)	91 (82)
Mean I/sigma(I)	10.99 (5.93)
Wilson B-factor	15.22
R-merge (%)	11.2 (20.4)
R-meas (%)	12.2 (23.4)
CC1/2	0.99 (0.96)
CC*	1.00 (0.99)
Refinement resolution range	45.19 - 1.20 (1.23 - 1.20)
Reflections used in refinement	58108 (3657)
Reflections used for R-free	2000 (126)
R-work (%)	12.7 (31.3)

	1.2 Å synchrotron
R-free (%)	14.6 (33.5)
Number of non-hydrogen atoms	2327
Macromolecular atoms	2143
Protein residues	163
RMS(bonds)	0.009
RMS(angles)	1.16
Ramachandran favored (%)	96
Ramachandran allowed (%)	4.1
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.84
Clashscore	0.98
Average B-factor	19.62
macromolecules	18.40
solvent	33.86

Table 2.2: Crystallographic statistics for room-temperature synchrotron dataset collected on a single crystal. Statistics for the highest-resolution shell are shown in parentheses.

In agreement with our previous room-temperature and solution studies [22], the XFEL and synchrotron mFo-DFc difference maps reveal evidence for catalytically-essential alternative conformations extending from the active site into the core of the protein (**Figure 2.1A,B)**. For example, the rotamer jump of Phe113, which is coupled to a backbone adjustment exemplifying the backrub motion [7,13], is apparent from a large positive mFo-DFc peak in both maps, which can be well modeled by a multi-conformer model (**Figure 2.1C,D**). Next, we performed automated electron-density scanning using Ringer [8,49], which identifies alternative conformations at low levels of electron density by evaluating the density value at each potential position for the γ atom about the χ1 dihedral angle. We focused on two residues, Ser99 and Leu98, which are key markers of the conformational exchange by NMR[46,50] and were implicated in our previous room-temperature X-ray and mutagenesis experiments [22]. Despite the lower resolution of the XFEL dataset, Ringer sampling revealed evidence for alternative conformations observed in the synchrotron dataset for core residue Ser99 (**Figure 2.1E**). We did not conclusively observe a peak for a discrete alternative conformation of Leu98 (**Figure 2.1F)**, but that is likely due to the lower resolution of the XFEL dataset. However, the asymmetric Ringer profile and the shift in the peak towards 300° indicate that the heterogeneity of Leu98 might be sufficiently modeled by the B-factors in the XFEL model.

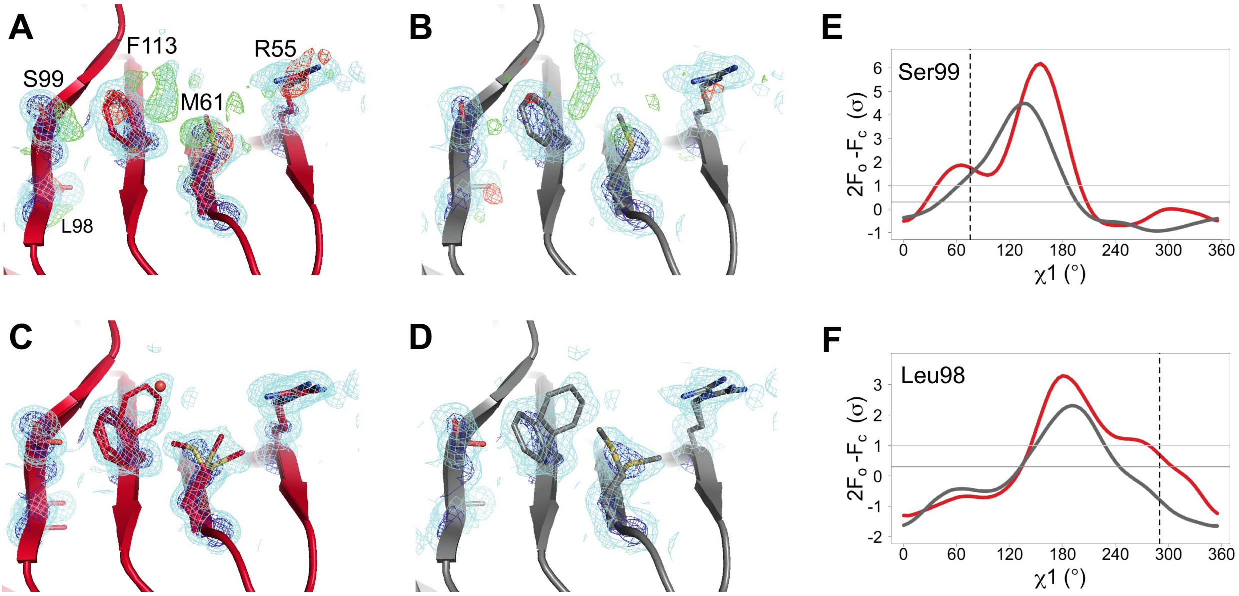


Figure 2.1: The active-site conformational ensemble of CypA imaged without radiation damage at room temperature. (A) Electron density maps for room-temperature synchrotron (red) and (**B**) XFEL (silver) single-conformer models reveal conformational heterogeneity extending from the protein core (Leu98 and Ser99) to the active site (Arg55) of CypA. The primary conformation is well supported by 2mFo-DFc electron density contoured at 0.6 σ (cyan mesh) and 3.0 σ (dark blue mesh). mFo-DFc difference electron density contoured at 3.0 σ (green mesh) and -3.0 σ (red mesh) suggest unmodeled alternative conformations. **(C,D)** Finalized multiconformer models explicitly model these alternative conformations, which are well-supported by 2mFo-DFc electron density. **(E,F)** Ringer electron-density sampling for the single-conformer models shows peaks representing alternative conformations for **(E)** Ser99 and **(F)** Leu98. The primary conformations of both residues are obvious as peaks for both models, but the minor conformations (dashed vertical line; as modeled in 3k0n) are also clearly evident, with 2mFo-DFc values well above the 0.3 σ (darker gray horizontal line) threshold. A backrub motion of -10° positions the backbone properly for Ringer to best detect the minor conformation for Ser99, but not for Leu98.

To interpret difference-map features and Ringer peaks, we built multi-conformer models for both datasets using *qFit* [9,24] and then finalized the models with *phenix.refine* and manual adjustments of alternative conformations [51] (**Figure 2.1C,D**, **Table 2.1,2.2**). The final models feature alternative conformations across the active-site network and are strongly supported by 2mFo-DFc electron density. Because of the lower resolution of the XFEL dataset, we were unable to definitively resolve the Leu98 alternative conformation and the partial-occupancy water overlaying the Phe113 alternative conformations; however, there is strong evidence for the other alternative conformations in the active-site network. These results indicate that XFEL crystallography can reveal native alternative conformations at high resolution and confirm that the alternative conformations observed in the synchrotron dataset are not an artifact of radiation damage. The observation of functionally important alternative conformations in CypA using fixed-target XFEL experiments further suggests that this approach can be expanded to studying the functional importance of conformational heterogeneity for other systems that are presently intractable for room-temperature diffraction at synchrotrons due to radiation damage.

Multitemperature X-ray Datasets Reveal Perturbed Conformational Ensembles of CypA

To probe the conformational landscape of CypA revealed by the room-temperature XFEL and synchrotron datasets, we collected eight high-resolution (1.34-1.58 Å) synchrotron crystallographic datasets across a wide range of temperatures from 100-310 K (**Table 2.3**). For each dataset, we initially refined single-conformer models.

Although the single-conformer models are very similar to each other, the accompanying electron density maps reveal differences throughout the protein. In the active-site network, the mFo-DFc difference electron density maps are relatively featureless below 200 K, suggesting that a single conformation is a valid fit below this temperature. In contrast, positive and negative mFo-DFc peaks become gradually more prevalent as temperature increases above 200 K, suggesting that multiple conformations are increasingly required to explain the data as temperature increases (**Figure S2.1**).

	100 K	150 K	180 K	240 K	260 K	280 K	300 K	310 K
PDB ID	4YUG	4YUH	4YUI	4YUJ	4YUK	4YUL	4YUM	4YUN
Wavelength	0.9767	0.9767	0.9767	0.9767	0.9767	0.9767	0.9767	0.9767
Resolution range	33.58 - 1.48 (1.53 - 1.48)	16.95 - 1.34 (1.39 - 1.34)	16.12 - 1.38 (1.43 - 1.38)	34.05 - 1.42 (1.47 - 1.42)	33.98 - 1.48 (1.53 - 1.48)	25.23 - 1.42 (1.47 - 1.42)	22.67 - 1.5 (1.55 - 1.50)	22.66 - 1.58 (1.64 - 1.58)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (a, b, c)	42.24 51.91 88.06	42.45 51.82 88.01	42.42 51.96 88.21	43.04 53.22 88.63	43.09 52.79 88.81	43.00 52.61 89.12	43.01 52.61 89.32	42.85 52.58 89.41
Total reflections	160129 (15842)	160780 (7437)	154202 (11295)	152578 (13600)	134699 (13381)	168932 (15019)	144734 (14433)	125225 (12326)
Unique reflections	32657 (3240)	42288 (3471)	39548 (3820)	38881 (3710)	34411 (3391)	38763 (3794)	32999 (3254)	28291 (2760)
Multiplicity	4.9 (4.9)	3.8 (2.1)	3.9 (3.0)	3.9 (3.7)	3.9 (3.9)	4.4 (4.0)	4.4 (4.4)	4.4 (4.5)
Completeness (%)	99 (100)	95 (80)	97 (95)	99 (96)	100 (100)	100 (100)	99 (100)	100 (100)
Mean I/ sigma(I)	14.07 (1.57)	25.95 (3.24)	16.47 (1.64)	12.86 (1.66)	10.09 (1.46)	15.51 (1.52)	16.90 (1.63)	13.26 (1.45)
Wilson B-factor	16.07	13.12	16.95	15.55	16.06	17.62	19.75	21.44
R-merge (%)	6.8 (99.4)	3.0 (29.4)	4.2 (71.8)	6.2 (99.2)	8.1 (104.3)	4.9 (100.0)	4.7 (101.7)	6.7 (127.3)
R-meas (%)	7.6 (111.0)	3.4 (36.9)	4.8 (85.6)	7.2 (116.8)	9.4 (120.8)	5.6 (115.3)	5.4 (115.9)	7.6 (144.5)
CC1/2	1.00 (0.62)	1.00 (0.90)	1.00 (0.60)	1.00 (0.50)	1.00 (0.52)	1.00 (0.52)	1.00 (0.59)	1.00 (0.56)
CC*	1.00 (0.88)	1.00 (0.97)	1.00 (0.87)	1.00 (0.82)	1.00 (0.83)	1.00 (0.83)	1.00 (0.86)	1.00 (0.85)
Refinement resolution range	33.09 - 1.48 (1.56 - 1.48)	19.12 - 1.34 (1.39 - 1.34)	17.00 - 1.38 (1.435 - 1.38)	34.06 - 1.42 (1.48 - 1.42)	33.98 - 1.48 (1.54 - 1.48)	25.23 - 1.42 (1.48 - 1.42)	22.67 - 1.5 (1.58 - 1.5)	25.22 - 1.58 (1.68 - 1.58)
Reflections used in refinement	32627 (4654)	42278 (3932)	39545 (4265)	38879 (4161)	34411 (4237)	38762 (4256)	32999 (4643)	28287 (4632)

	100 K	150 K	180 K	240 K	260 K	280 K	300 K	310 K
Reflections used for R-free	1028 (147)	1325 (125)	1238 (133)	1218 (130)	1080 (133)	1217 (133)	1036 (145)	889 (146)
R-work (%)	13.3 (20.4)	12.4 (16.4)	13.3 (25.4)	12.6 (26.3)	13.1 (26.0)	11.1 (22.6)	10.8 (20.0)	11.7 (21.8)
R-free (%)	18.3 (26.8)	15.6 (21.3)	17.5 (33.0)	15.6 (30.4)	16.8 (31.2)	14.3 (25.5)	14.4 (24.8)	15.0 (28.8)
Number of non-hydrogen atoms	2279	2433	1969	1993	2035	2120	2096	2172
Macromolecule atoms	1933	2132	1745	1750	1837	1924	1952	2061
Protein residues	165	164	164	163	163	163	163	163
RMS(bonds)	0.009	0.008	0.008	0.009	0.009	0.008	0.009	0.009
RMS(angles)	1.16	1.20	1.23	1.20	1.16	1.16	1.14	1.14
Ramachandran favored (%)	97	94	97	96	97	96	97	96
Ramachandran allowed (%)	3.3	5.7	2.7	4.1	3	4.2	3.3	3.9
Ramachandran outliers (%)	0	0	0	0	0	0	0	0
Rotamer outliers (%)	2.4	1.3	0.53	1.1	1.5	1.9	1.4	0.88
Clashscore	0.57	1.08	0.00	1.24	0.27	0.78	0.52	0.00
Average B-factor	21.74	17.25	21.85	20.14	20.00	21.48	24.09	25.77
Macromolecule Average B-factor	18.48	14.67	19.99	17.95	18.17	19.61	22.82	24.94
Solvent Average B-factor	39.99	35.54	36.34	35.89	37.01	39.89	41.23	41.30

Table 2.3: Crystallographic statistics for multitemperature synchrotron datasets collected on a single crystal per dataset. Statistics for the highest-resolution shell are shown in parentheses.

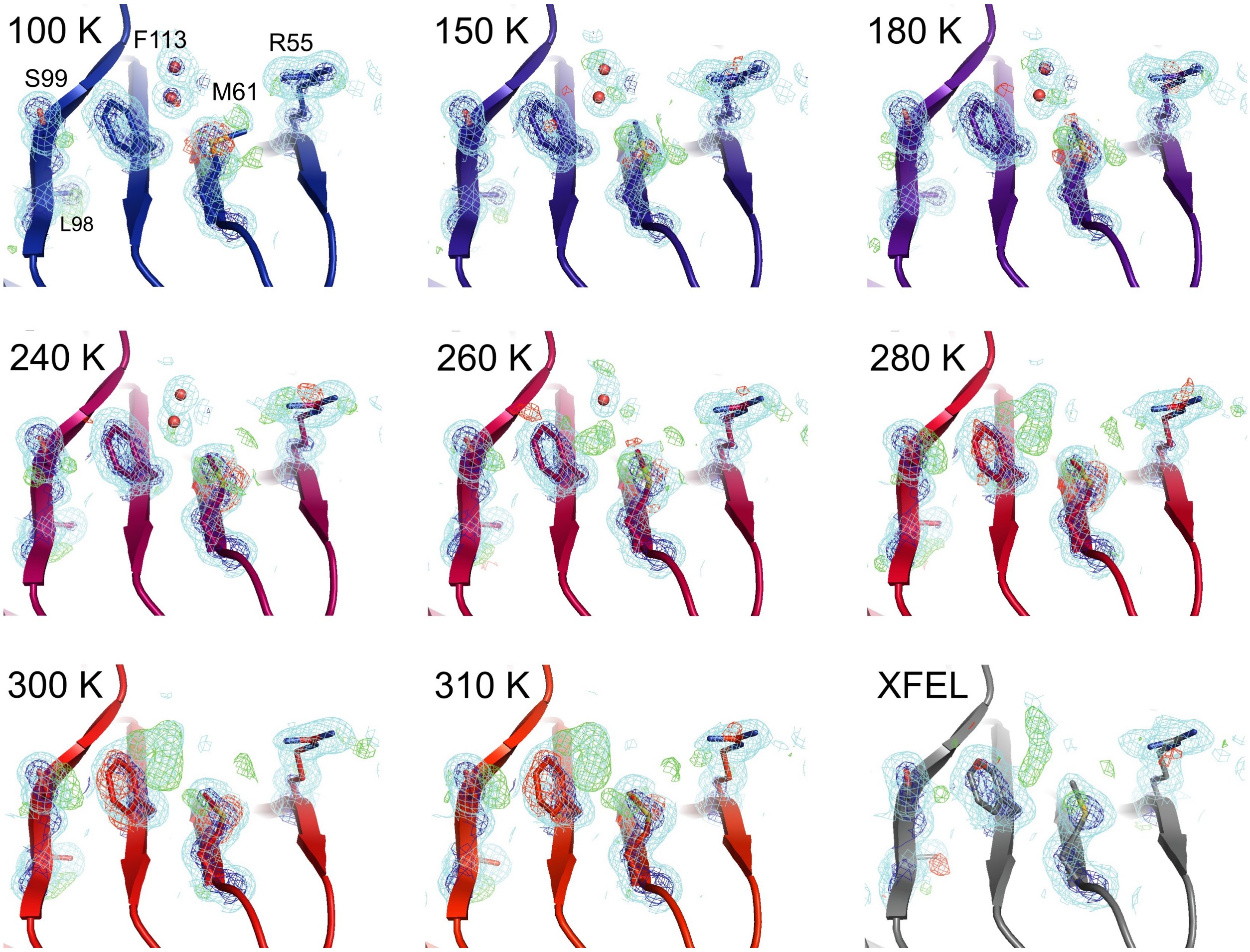


Figure S2.1: Single-conformer models cannot explain the crystallographic data at higher temperatures. The CypA dynamic network is shown after molecular replacement and refinement (including automated water placement) in PHENIX, before any manual rebuilding. The major state is well supported by 2mFo-DFc electron density contoured at 0.6 σ (cyan mesh) and 3.0 σ (dark blue mesh) for all datasets, but mFo-DFc difference electron density becomes more negative for the major state (-3.0 σ, red mesh) and more positive for the unmodeled minor state (+3.0 σ, green mesh) as temperature increases across the synchrotron datasets (blue to red); this is especially true above the glass transition (~200 K). Full-occupancy water molecules (red spheres) are automatically placed by PHENIX near the Phe113 minor state in lower-temperature, but not in higher-temperature synchrotron models because they are mutually exclusive with the secondary Phe113 conformation.

To ground this conformational redistribution in all-atom detail, we built a multiconformer model with *qFit* [9,24] for each multitemperature dataset. As with the room-temperature synchrotron and XFEL models, the finalized multiconformer models were improved over the single- conformer models (**Table 2.4**). Below the ~200 K glass transition temperature, the active- site network is best modeled as a single state, with ordered water molecules clearly evident adjacent to Phe113 (**Figure 2.2**, top row). Above 200 K, by contrast, multiple conformations provide a better explanation of the data. Interestingly, some partial- occupancy water molecules still co-occur with the major conformations (**Figure 2.2**, middle and bottom rows). Met61 appears to populate additional conformations above the glass transition, although it is difficult to precisely define changes in its conformational ensemble as temperature increases (see Methods). This residue bridges Phe113 and the catalytic Arg55 via steric contacts between alternative conformations in both directions, emphasizing the importance of modeling multiple conformations in all-atom detail for understanding inter-residue coupling. More generally, our multiconformer models separate harmonic from non-harmonic contributions to flexibility, which was a significant limitation of previous studies of the glass transition that relied on B-factors in single-conformer models [42,44].

		RT sync	**XFEL**	**100 K**	**150 K**	**180 K**	**240 K**	**260 K**	**280 K**	**300 K**	**310 K**
R_{free} (%)	*raw qFit*	16.7	25.2	19.0	16.9	18.5	17.5	17.9	15.7	16.3	16.1
	final	14.6	24.9	18.3	15.6	17.5	15.6	16.8	14.3	14.4	15.0
	Δ	-2.1	-0.3	-0.7	-1.3	-1.0	-1.9	-1.1	-1.4	-1.9	-1.1
Molprobity score	*raw qFit*	1.47	1.80	1.79	1.31	1.21	1.18	1.45	1.28	0.95	1.19
	final	1.08	1.39	1.19	1.29	0.63	1.14	0.91	1.25	0.99	0.76
	Δ	-0.39	-0.41	-0.80	-0.02	-0.58	-0.04	-0.54	-0.03	+0.04	-0.43

Table 2.4: Improvements in validation statistics from finalizing raw qFit models. Statistics calculated with *phenix.molprobity*.

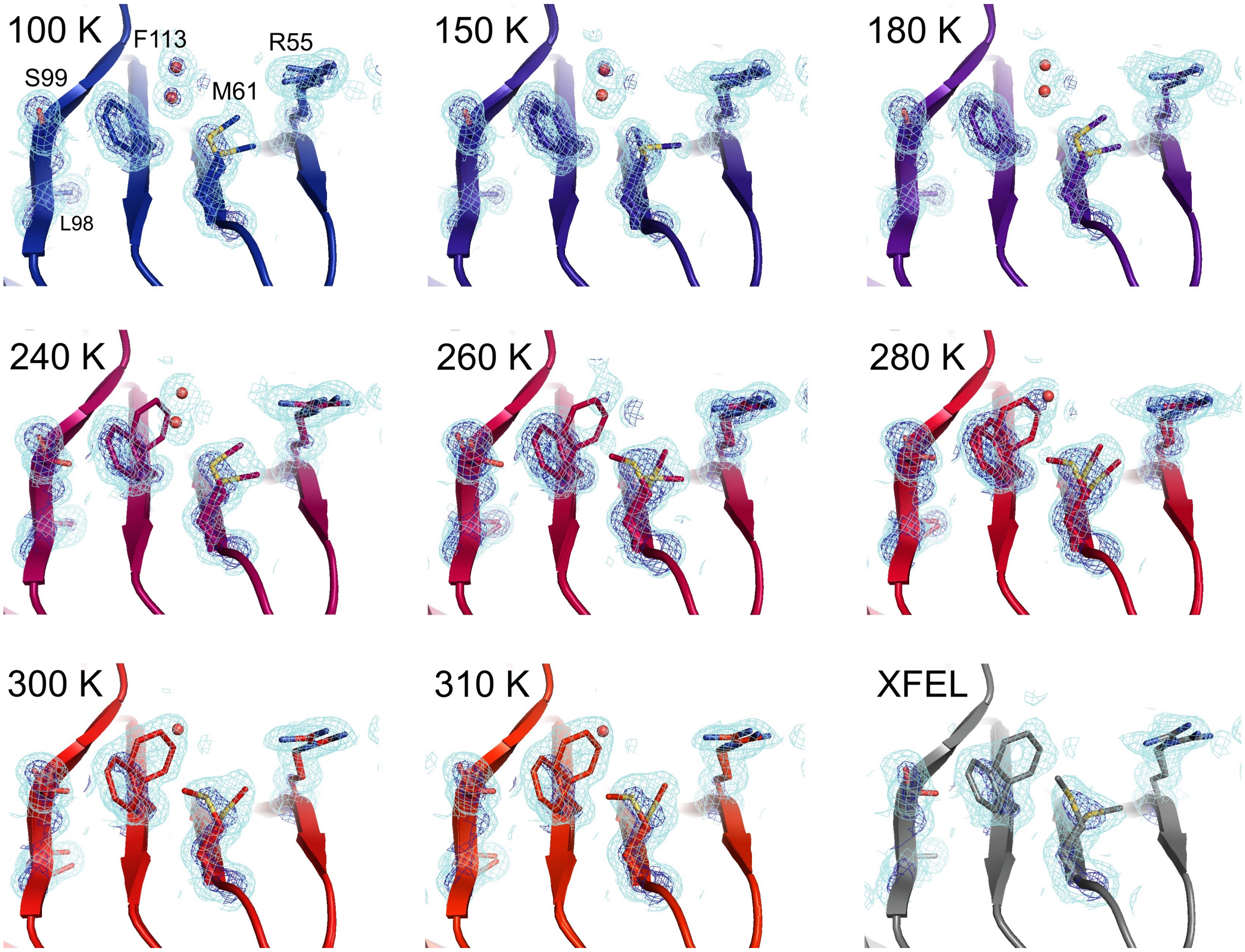


Figure 2.2: Multiconformer modeling across temperatures captures increasing conformational heterogeneity above the glass transition. Residues extending from the core to the active site of CypA adopt a single conformer at low temperatures, but gradually transition to increasing occupancy of secondary conformations as temperature increases above the glass transition. These conformations are well supported by 2mFo-DFc electron density contoured at 0.6 σ (cyan mesh) and 3.0 σ (dark blue mesh). This is corroborated by the room-temperature XFEL model (gray), which is free from conventional radiation damage and features the same secondary conformations. Water molecules (red spheres) are more fully ordered at low temperatures, but become only partially occupied at higher temperatures because they are mutually exclusive with the secondary Phe113 conformation.

Some Regions Feature Conformational Heterogeneity Only Below the Glass Transition Temperature

Although more conformational heterogeneity is evident at higher temperatures in the active site of CypA, theoretical studies suggest the counterintuitive idea that some additional conformations can be accessed only as temperature decreases [19]. We observed that one region of CypA exhibits backbone alternative conformations only below 200K: the loop containing residues 79-83 (**Figure 2.3**). This region is well fit by a single conformation at each temperature above the glass transition, but a secondary loop conformation is necessary to explain the electron density at 100, 150, and 180 K. Additionally, the loop is clearly single-state in the highest-resolution (1.2 Å) dataset, which is at room temperature (**Figure S2.1**) demonstrating that the slightly lower resolution of the elevated temperature datasets is not responsible for obscuring the secondary conformation.

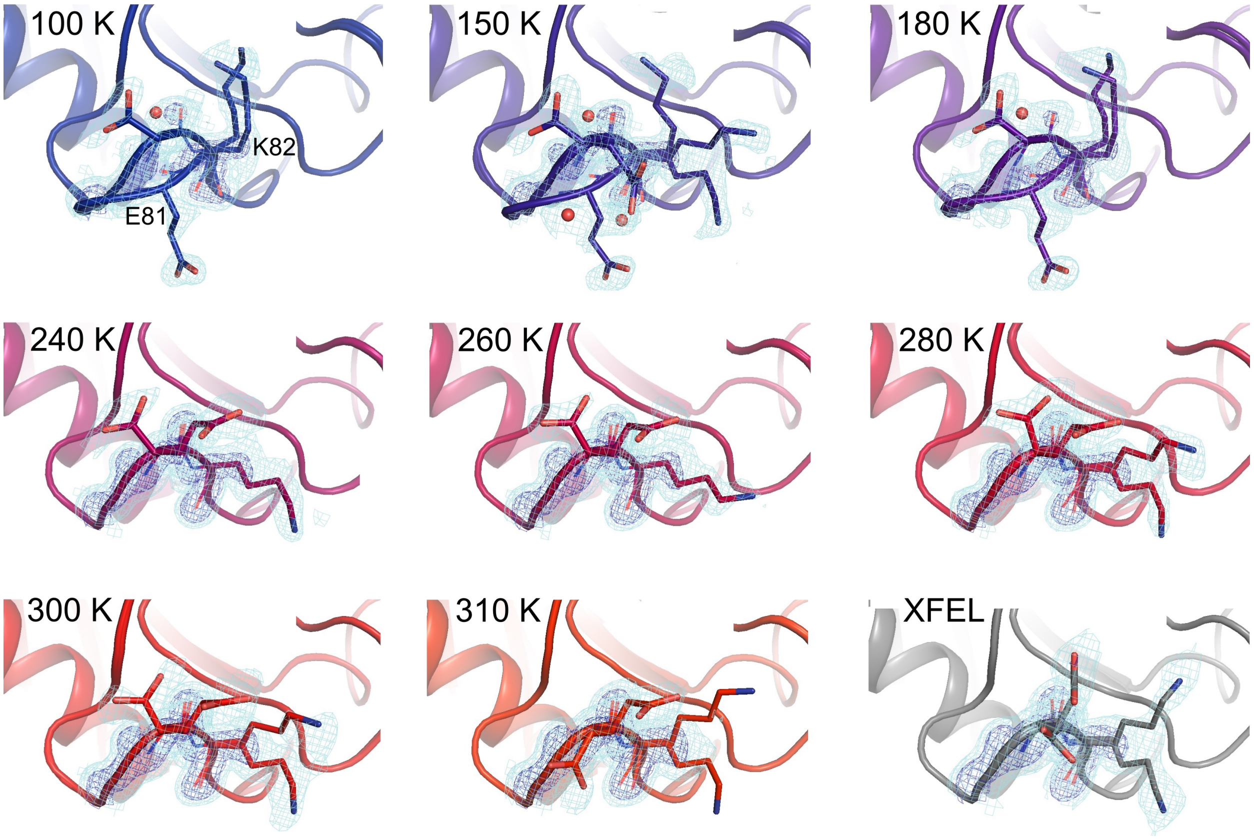


Figure 2.3: Low temperature "freezes in" an alternative loop conformation. The surface loop containing residues 79-83 adopts alternative conformations at temperatures below (top row) but not above (bottom two rows) the glass transition. The secondary loop conformation is separated from the body of the protein by an ordered water molecule (red sphere); the van der Waals interactions between the loop and the water may reflect an enthalpic stabilization that is more dominant at low temperatures. The electron density peak to the right of the water corresponds to the backbone carbonyl oxygen of Glu81. 2mFo-DFc electron density contoured at 0.6 σ (cyan mesh) and 2.0 σ (dark blue mesh).

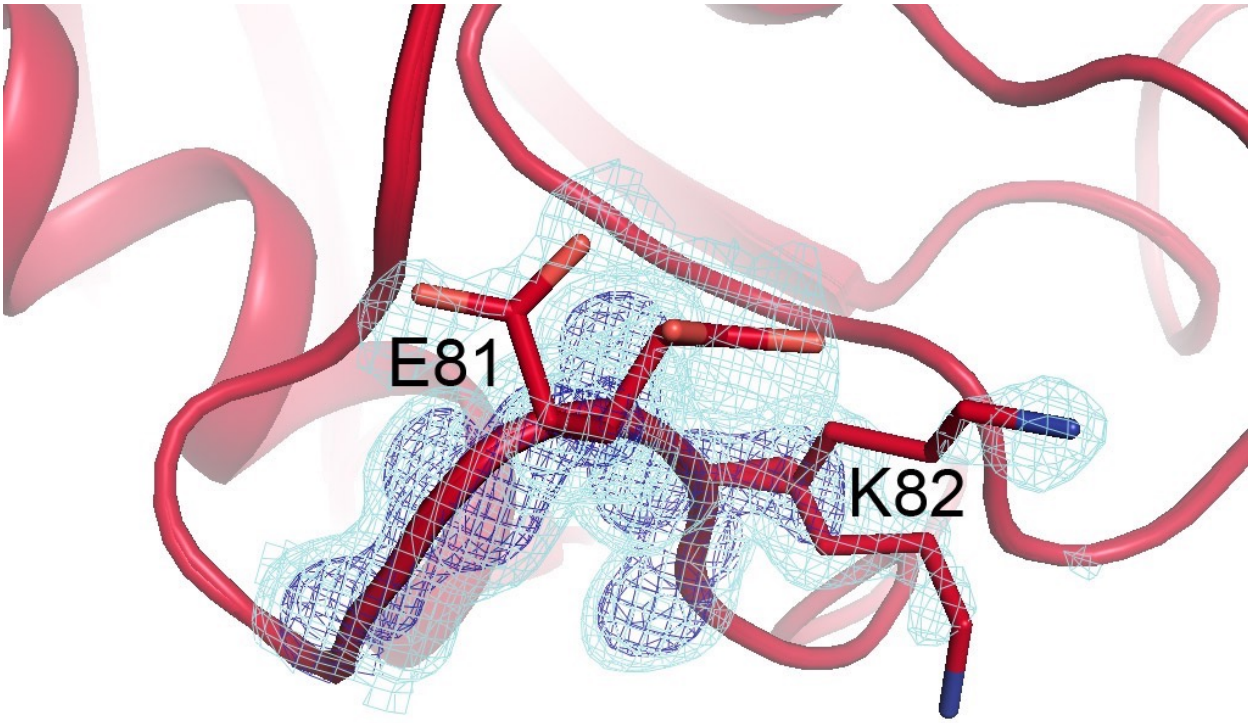


Figure S2.2: Temperature determines a loop conformational ensemble even in the highest resolution room temperature dataset. The surface loop containing residues 79-83 adopts a single conformation at room temperature in our highest-resolution (1.2 Å) synchrotron structure.

In the primary conformation, the 79-83 loop is not involved in any mainchain-mainchain hydrogen bonds to the rest of CypA, suggesting that the barrier to forming the secondary conformation does not involve breakage of cooperative secondary-structure-like interactions. Given that relative enthalpic contributions to free energy are increased at lower temperatures [19,20], these observations indicate that the appearance of the secondary state for residues 79-83 at 100-180 K, but not at 240-310 K, is enthalpically driven, in agreement with theoretical predictions [19,52]. Consistent with an enthalpic stabilization mechanism, the secondary conformation of the 79-83 loop is accompanied by an ordered, partial- occupancy water molecule (**Figure 2.3**, top row). This water molecule, which is clearly distinct from the carbonyl oxygen of the primary conformation of Glu81, wedges between the loop and the rest of the protein.

Discussion

Here we have mapped the conformational landscape of the dynamic enzyme CypA by a two-pronged strategy. First, we determined a high-resolution crystal structure of CypA using fixed-target SF-ROX data collection with the X-ray free electron laser. Because XFEL diffraction and data collection occur on a faster timescale than chemical damage to the crystal by incident X-rays, this structure is effectively free of radiation damage. As has been observed for the serotonin receptor GPCR, the conformations imaged by room temperature XFEL experiments can deviate from the cryogenic structrues determined at synchrotron sources [31]. In contrast to that SFX experiment, which used ~30,000 diffraction images, the advances made possible by post-refinement [37] and goniometer-based crystallography [33] at LCLS enabled data collection from only 1,239 images. Therefore our XFEL results bridge previous solution experiments with room- temperature synchrotron crystallography by demonstrating the structural origin of the active-site conformational heterogeneity of CypA free from any radiation damage artifacts.

Second, we characterized the conformational landscape of CypA by multitemperature crystallography. Previous approaches to studying the glass transition across the 180-220 K temperature regime were limited in different ways: Mössbauer spectroscopy [53] is restricted to metal sites, inelastic neutron scattering [54] gives only globally averaged information about motions, crystallographic B-factors [42,44] and NMR order parameters [45] are site-resolved but cannot distinguish between different types of motions, and molecular dynamics simulations [55] explicitly monitor all protein and solvent atoms but are subject to imperfections in the underlying energy function and limited sampling time. By contrast, our multiconformer crystallographic models include both B-factors and explicit occupancy-weighted alternative conformations, granting us experimentally-based, site-resolved insight into the conformational response to temperature perturbation. We observed many abrupt effects of temperature titration: for example, active-site conformational heterogeneity increases dramatically above 200 K, whereas one loop adopts alternative conformations only below 200 K. Most residues experience temperature-dependent increases in conformational heterogeneity only above 200 K; however, this increase does not originate from a single global transition temperature and even closely packed residues in the active site dynamic network exhibit heterogeneous responses.

These findings lend support to the idea that distinct classes of functionally important protein motions are activated as temperature increases across the glass transition [45]. In the dynamic active-site network of CypA, we observe evidence for imperfect coupling, with Ser99 and Phe113 more tightly coupled to each other than to the rest of the network. Our observations suggest that the catalytically limiting motion in the CypA active-site network, which was identified by NMR relaxation dispersion data [46,56], is likely comprised of many stochastic processes that occur on similar timescales. Indeed, analysis of the contacts between alternative conformations shows that many computed pathways of steric clash-relief events involve only subsets of the active-site network [24]. Such “frustration”, and subsequent “cracking” mechanisms, are likely at play for many systems that are thought to exhibit collective and functionally important dynamics [57,58]. These phenomena may arise because there is insufficient evolutionary pressure to “perfectly” sculpt the energy basins, energy barriers, and changes thereof along the entire trajectory of a conformational change. However, selection can act to exploit the distinct substates populated by the apo native enzyme for different parts of the catalytic cycle, such as substrate binding or product release [59].

However, to reveal the kinetic processes underlying these substates, time-resolved experiments, likely using the XFEL pump-probe setup [38] or next- generation synchrotron sources, will be necessary.These new technologies will test hypotheses about the relationship between macromolecular motion and function by offering increasing access to time-resolved experiments with all-atom detail. Such experiments can potentially reveal the structural basis of protein motions and the extent of conformational coupling during functional cycles. Temperature jumps are a potentially generalizable mechanism to expose the lower-population conformations used by ligand binding or shifted by mutation [5,18]. We have demonstrated here that both XFEL and multi-temperature crystallography are feasible for CypA, which are prerequisites for time-resolved temperature-jump XFEL experiments that could directly interrogate the degree and mechanism of coupling in the active-site network. The ability to use small crystals in XFEL experiments is a major advantage for syncing reactions and penetrating samples with “pump” laser pulses [38]. These advantages will likely extend to irreversible reactions triggered by the short diffusion times of substrates into nanocrystals [60]. As more complex reactions are probed, advances in identifying [8,61] and modeling [9,62] the component conformations will be required. These future studies will reveal protein conformational dynamics in atomic detail, advancing our ability to dissect native reaction mechanisms and effectively engineer artificial catalysts, biosensors, and other useful nanomachines.

Methods

Protein expression, purification, and crystallization

Wild-type CypA was produced and crystallized as previously reported [22]. Briefly, crystals were grown by mixing equal volumes of well solution (100 mM HEPES pH 7.5, 23% PEG 3350, 5 mM TCEP) and protein (60 mg/mL in 20 mM HEPES pH 7.5, 100 mM NaCl, 0.5 mM TCEP) in the hanging-drop format.

Crystallographic data collection

For the XFEL experiment, we collected multiple diffraction images per crystal using a 10 micron X-ray beam with each irradiation point separated by at least 25-40 microns to avoid diffusion of radiation damage. A total of 1,239 still diffraction images were collected from 71 CypA crystals over the course of two experiments using a goniometer setup and a Rayonix MX325HE detector at LCLS-XPP [33]. All data was collected at ambient temperature (approximately 273 K). To prevent dehydration, crystals were coated with paratone oil immediately after looping and mounted on the goniometer at the XPP end station of LCLS using the SAM sample exchange robot [63].

For the new 1.2 Å room-temperature synchrotron dataset, paratone oil was applied to cover a 2 µL hanging drop containing a single large crystal of CypA. The crystal was harvested through the paratone and excess mother liquor was removed using a fine paper wick. Attenuated data was collected at SSRL beamline 11-1 at 273 K controlled by the cryojet on the PILATUS 6M PAD detector.

For the multitemperature synchrotron datasets at 100, 150, 180, 240, 260, 280, 300, and 310 K, we collected data at the Cornell High Energy Synchrotron Source (CHESS) at beamline A1 with a 100 micron collimator using a wavelength of 0.9767 Å. Crystals were looped by hand, stripped of excess mother liquor (100 mM HEPES pH 7.5, 23% PEG 3350, 5 mM TCEP) using NVH oil [64], and cooled, which allowed collection of complete datasets at any temperature without ice formation while avoiding use of any penetrating cryoprotectants.

Crystallographic data processing

The XFEL data were processed using *cctbx.xfel* [48]. Of the 1,239 images collected, 772 were indexed and their intensities were integrated. Post-refinement, as implemented by *prime* (**p**ost-**r**ef**i**nement and **me**rging, version date: Nov 11 20:22:51 2014)[37], was used to correct the intensity measurements and merge the data. We optimized over the uc_tolerance, n- postref_cycle, sigma_min, partiality_min, and gamma_e values to obtain the final structure factor amplitudes. Data reduction statistics for the XFEL data are provided in **Table 2.1**.

The synchrotron datasets were indexed, integrated, and scaled using XDS and XSCALE, and intensities were subsequently converted to structure factor amplitudes using XDSCONV. All datasets were from single crystals. Data reduction statistics for the high-resolution, room-temperature dataset and multi-temperature datasets can be found in **Table 2.2** and **Table 2.3.**

To promote consistency between models derived from different data sets, Rfree flags were generated using *PHENIX* for the highest-resolution “reference” (1.2 Å 273 K) dataset first, and were subsequently copied to all other multitemperature and XFEL datasets for the automated molecular replacement and refinement pipeline.

Model building

For each dataset, we calculated initial phases by performing molecular replacement with *phenix.auto_mr* using PDB ID 2cpl as a search model. We next refined XYZs and ADPs of the initial model with *phenix.refine* for 4 macrocycles with XYZ and ADP weight optimization turned on; identified TLS groups with *phenix.find_tls_groups*; and refined optimized XYZs, ADPs, and TLS parameters for 6 more macrocycles. These single-conformer models and associated electron density maps were used as input for two subsequent steps.

First, the single-conformer models were analyzed with Ringer [8] via *mmtbx.ringer* using default settings. A coupled sidechain-backbone “backrub" motion [7] of -10° for Ser99 was necessary to match the Cα and Cβ positions of the minor conformation as modeled in PDB ID 3K0N; using this modified backbone indeed yielded maximal minor-conformation Ringer peaks for our multi-temperature datasets. No backrub motion was necessary for Leu98 due to the different type of backbone displacement [22] Correlation coefficients between pairs of Ringer curves were calculated using the *cor* function in R [65].

Second, the single-conformer models were used as input to *qFit [9,20]*. Subsequent to the automated model building, we manually deleted ill-fitting waters and altered alternative protein sidechain conformations based on fit to the electron density in *Coot* [66] and refinement with *phenix.refine*. For example, at 240 K, *qFit* automatically modeled Phe113 as single-state, but significant +mFo-DFc peaks remained, so we decided on a two-state model. Met61 was particularly difficult to model accurately across temperatures due to convolved issues of χ3 non-rotamericity for Met in general (Butterfoss et al., 2005), the relatively high electron count for sulfur, and likely temperature-modulated Met-specific radiation damage. For these reasons, visual inspection of the maps and manual building is currently essential for alternative backbone conformations with moderate displacements, as observed in residues 79-83. We are currently developing new methods to automatically detect and model such backbone excursions in multiscale multiconformer models. These efforts improved Rfree and MolProbity scores across datasets (**Table 2.4**). Because of the lower resolution, the XFEL model was refined with three TLS groups and with optimization of X-ray vs. geometry and ADP weights.

Acknowledgements

We thank Justin Biel, Bryn Fenwick, Robert Stroud, Ian Wilson, and Peter Wright for helpful conversations. DAK is supported by an A. P. Giannini Postdoctoral Research Fellowship. RAW is supported by a NSF Graduate Research Fellowship. MCT is supported by a BioXFEL Postdoctoral Fellowship. NKS and ASB are supported by NIH GM095887 and NIH GM102520. ATB and WIW acknowledge a Howard Hughes Medical Institute Collaborative Innovation Award (HCIA) that also provided funds for the purchase of the microdiffractometer for the goniometer setup. HvdB is sup- ported by the NIH Protein Structure Initiative U54GM094586 at the Joint Center for Structural Geno- mics and SLAC National Accelerator Laboratory grant SLAC-LDRD-0014-13-2. RET is supported by NSF MCB-1330685. JSF is a Searle Scholar, Pew Scholar, and Packard Fellow, and is supported by NIH OD009180, NIH GM110580, and NSF STC-1231306. Use of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The use of the Cornell High Energy Synchrotron Source (CHESS) is supported by NSF DMR-1332208; the Macromolecular Diffraction at CHESS (MacCHESS) facility is supported by NIH GM103485; and the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and NIH GM103393.

References

1. Motlagh HN, Wrabl JO, Li J, Hilser VJ: **The ensemble nature of allostery**. *Nature* 2014, **508**:331–339.

2. van den Bedem H, Fraser JS: **Integrative, dynamic structural biology at atomic resolution--it's about time**. *Nat Methods* 2015, **12**:307–318.

3. Rejto PA, Freer ST: **Protein conformational substates from X-ray crystallography**. *Prog Biophys Mol Biol* 1996, **66**:167–196.

4. Smith JL, Hendrickson WA, Honzatko RB, Sheriff S: **Structural heterogeneity in protein crystals**. *Biochemistry* 1986, **25**:5018–5027.

5. Woldeyes RA, Sivak DA, Fraser JS: **E pluribus unum, no more: from one crystal, many conformations**. *Curr Opin Struct Biol* 2014, **28**:56–62.

6. Burnley BT, Afonine PV, Adams PD, Gros P: **Modelling dynamics in protein crystal structures by ensemble refinement**. *Elife* 2012, **1**:e00311.

7. Davis IW, Arendall WB 3rd, Richardson DC, Richardson JS: **The backrub motion: how protein backbone shrugs when a sidechain dances**. *Structure* 2006, **14**:265–274.

8. Lang PT, Ng H-L, Fraser JS, Corn JE, Echols N, Sales M, Holton JM, Alber T: **Automated electron-density sampling reveals widespread conformational polymorphism in proteins**. *Protein Sci* 2010, **19**:1420–1431.

9. van den Bedem H, Dhanik A, Latombe JC, Deacon AM: **Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers**. *Acta Crystallogr D Biol Crystallogr* 2009, **65**:1107–1117.

10. Baldwin AJ, Kay LE: **NMR spectroscopy brings invisible protein states into focus**. *Nat Chem Biol* 2009, **5**:808–814.

11. Fenwick RB, van den Bedem H, Fraser JS, Wright PE: **Integrated description of protein dynamics from room-temperature X-ray crystallography and NMR**. *Proc Natl Acad Sci U S A* 2014, **111**:E445–54.

12. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE: **Biomolecular simulation: a computational microscope for molecular biology**. *Annu Rev Biophys* 2012, **41**:429–452.

13. Ollikainen N, Smith CA, Fraser JS, Kortemme T: **Flexible backbone sampling methods to model and design protein alternative conformations**. *Methods Enzymol* 2013, **523**:61–85.

14. Henzler-Wildman K, Kern D: **Dynamic personalities of proteins**. *Nature* 2007,

450:964–972.

15. Garman EF: **Radiation damage in macromolecular crystallography: what is it and why should we care?** *Acta Crystallogr D Biol Crystallogr* 2010, **66**:339–351.

16. Holton JM: **A beginner's guide to radiation damage**. *J Synchrotron Radiat* 2009, **16**:133–142.

17. Warkentin M, Hopkins JB, Badeau R, Mulichak AM, Keefe LJ, Thorne RE: **Global radiation damage: temperature dependence, time dependence and how to outrun it**. *J Synchrotron Radiat* 2013, **20**:7–13.

18. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature X-ray crystallography**. *Proc Natl Acad Sci U S A* 2011, **108**:16247–16252.

19. Halle B: **Biomolecular cryocrystallography: structural changes during flash-cooling**. *Proc Natl Acad Sci U S A* 2004, **101**:4793–4798.

20. Keedy DA, van den Bedem H, Sivak DA, Petsko GA, Ringe D, Wilson MA, Fraser JS: **Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR**. *Structure* 2014, **22**:899–910.

21. Deis LN, Pemble CW 4th, Qi Y, Hagarman A, Richardson DC, Richardson JS, Oas TG: **Multiscale conformational heterogeneity in staphylococcal protein a: possible determinant of functional plasticity**. *Structure* 2014, **22**:1467–1477.

22. Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T: **Hidden alternative structures of proline isomerase essential for catalysis**. *Nature* 2009, **462**:669–673.

23. Fukuda Y, Inoue T: **High-temperature and high-resolution crystallography of thermostable copper nitrite reductase.** *Chem Commun* 2015, **51**:6532–6535.

24. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS: **Automated identification of functional dynamic contact networks from X-ray crystallography.** *Nat Methods* 2013, **10**:896–902.

25. Holton JM, Frankel KA: **The minimum crystal size needed for a complete diffraction data set.** *Acta Crystallogr D Biol Crystallogr* 2010, **66**:393–408.

26. Zeldin OB, Brockhauser S, Bremridge J, Holton JM, Garman EF: **Predicting the X-ray lifetime of protein crystals.** *Proc Natl Acad Sci U S A* 2013, **110**:20551–20556.

27. Kern J, Tran R, Alonso-Mori R, Koroidov S, Echols N, Hattne J, Ibrahim M, Gul

S, Laksmono H, Sierra RG, et al.: **Taking snapshots of photosynthetic water oxidation using femtosecond X-ray diffraction and spectroscopy**. *Nat Commun* 2014, **5**:4371.

28. Spence JCH, Weierstall U, Chapman HN: **X-ray lasers for structural and dynamic biology**. *Rep Prog Phys* 2012, **75**:102601.

29. Neutze R, Wouts R, van der Spoel D, Weckert E, Hajdu J: **Potential for biomolecular imaging with femtosecond X-ray pulses**. *Nature* 2000, **406**:752–757.

30. Boutet S, Lomb L, Williams GJ, Barends TRM, Aquila A, Doak RB, Weierstall U, DePonte DP, Steinbrener J, Shoeman RL, et al.: **High-resolution protein structure determination by serial femtosecond crystallography**. *Science* 2012, **337**:362–364.

31. Liu W, Wacker D, Gati C, Han GW, James D, Wang D, Nelson G, Weierstall U, Katritch V, Barty A, et al.: **Serial femtosecond crystallography of G protein-coupled receptors**. *Science* 2013, **342**:1521–1524.

32. Neutze R, Moffat K: **Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges**. *Curr Opin Struct Biol* 2012, **22**:651–659.

33. Cohen AE, Soltis SM, González A, Aguila L, Alonso-Mori R, Barnes CO, Baxter EL, Brehmer W, Brewster AS, Brunger AT, et al.: **Goniometer-based femtosecond crystallography with X-ray free electron lasers**. *Proc Natl Acad Sci U S A* 2014, **111**:17122–17127.

34. Hirata K, Shinzawa-Itoh K, Yano N, Takemura S, Kato K, Hatanaka M, Muramoto K, Kawahara T, Tsukihara T, Yamashita E, et al.: **Determination of damage-free crystal structure of an X-ray-sensitive protein using an XFEL**. *Nat Methods* 2014, **11**:734–736.

35. Suga M, Akita F, Hirata K, Ueno G, Murakami H, Nakajima Y, Shimizu T, Yamashita K, Yamamoto M, Ago H, et al.: **Native structure of photosystem II at 1.95 Å resolution viewed by femtosecond X-ray pulses**. *Nature* 2015, **517**:99–103.

36. White TA, Barty A, Stellato F, Holton JM, Kirian RA, Zatsepin NA, Chapman HN: **Crystallographic data processing for free-electron laser sources**. *Acta Crystallogr D Biol Crystallogr* 2013, **69**:1231–1240.

37. Uervirojnangkoorn M, Zeldin OB, Lyubimov AY, Hattne J, Brewster AS, Sauter NK, Brunger AT, Weis WI: **Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals**. *Elife*

2015, **4**.

38. Tenboer J, Basu S, Zatsepin N, Pande K, Milathianaki D, Frank M, Hunter M, Boutet S, Williams GJ, Koglin JE, et al.: **Time-resolved serial crystallography captures high-resolution intermediates of photoactive yellow protein**. *Science* 2014, **346**:1242–1246.

39. Trincao J, Hamilton ML, Christensen J, Pearson AR: **Dynamic structural science: recent developments in time-resolved spectroscopy and X-ray crystallography**. *Biochem Soc Trans* 2013, **41**:1260–1264.

40. Moffat K: **Time-resolved macromolecular crystallography**. *Annu Rev Biophys Biophys Chem* 1989, **18**:309–332.

41. Doster W: **The protein-solvent glass transition**. *Biochim Biophys Acta* 2010, **1804**:3–14.

42. Frauenfelder H, Hartmann H, Karplus M, Kuntz ID Jr, Kuriyan J, Parak F, Petsko GA, Ringe D, Tilton RF Jr, Connolly ML: **Thermal expansion of a protein**. *Biochemistry* 1987, **26**:254–261.

43. Frauenfelder H, Petsko GA, Tsernoglou D: **Temperature-dependent X-ray diffraction as a probe of protein structural dynamics**. *Nature* 1979, **280**:558–563.

44. Tilton RF Jr, Dewan JC, Petsko GA: **Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320 K**. *Biochemistry* 1992, **31**:2469–2481.

45. Lee AL, Wand AJ: **Microscopic origins of entropy, heat capacity and the glass transition in proteins**. *Nature* 2001, **411**:501–504.

46. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D: **Intrinsic dynamics of an enzyme underlies catalysis**. *Nature* 2005, **438**:117–121.

47. Schlichting I: **Structural biology: Photosynthetic complex in close-up**. *Nature* 2015, **517**:26–27.

48. Hattne J, Echols N, Tran R, Kern J, Gildea RJ, Brewster AS, Alonso-Mori R, Glöckner C, Hellmich J, Laksmono H, et al.: **Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers**. *Nat Methods* 2014, **11**:545–548.

49. Lang PT, Holton JM, Fraser JS, Alber T: **Protein structural ensembles are revealed by redefining X-ray electron density noise**. *Proc Natl Acad Sci U S*

A 2014, **111**:237–242.

50. Eisenmesser EZ, Bosco DA, Akke M, Kern D: **Enzyme dynamics during catalysis**. *Science* 2002, **295**:1520–1523.

51. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive Python-based system for macromolecular structure solution**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:213–221.

52. Prell JS, Correra TC, Chang TM, Biles JA, Williams ER: **Entropy drives an attached water molecule from the C- to N-terminus on protonated proline**. *J Am Chem Soc* 2010, **132**:14733–14735.

53. Parak F, Frolov EN, Mössbauer RL, Goldanskii VI: **Dynamics of metmyoglobin crystals investigated by nuclear gamma resonance absorption**. *J Mol Biol* 1981, **145**:825–833.

54. Doster W, Cusack S, Petry W: **Dynamical transition of myoglobin revealed by inelastic neutron scattering**. *Nature* 1989, **337**:754–756.

55. Vitkup D, Ringe D, Petsko GA, Karplus M: **Solvent mobility and the protein "glass" transition**. *Nat Struct Biol* 2000, **7**:34–38.

56. Schlegel J, Armstrong GS, Redzic JS, Zhang F, Eisenmesser EZ: **Characterizing and controlling the inherent dynamics of cyclophilin-A**. *Protein Sci* 2009, **18**:811–824.

57. Bhabha G, Biel JT, Fraser JS: **Keep on moving: discovering and perturbing the conformational dynamics of enzymes**. *Acc Chem Res* 2015, **48**:423–430.

58. Ferreiro DU, Hegler JA, Komives EA, Wolynes PG: **On the role of frustration in the energy landscapes of allosteric proteins**. *Proc Natl Acad Sci U S A* 2011, **108**:3499–3503.

59. De Simone A, Aprile FA, Dhulesia A, Dobson CM, Vendruscolo M: **Structure of a low-population intermediate state in the release of an enzyme product**. *Elife* 2015, **4**.

60. Schmidt M: **Mix and Inject: Reaction Initiation by Diffusion for Time-Resolved Macromolecular Crystallography**. *Advances in Condensed Matter Physics* 2013, **2013**.

61. Rajagopal S, Schmidt M, Anderson S, Ihee H, Moffat K: **Analysis of experimental time-resolved crystallographic data by singular value decomposition**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**:860–871.

62. Shukla D, Hernández CX, Weber JK, Pande VS: **Markov state models provide**

insights into dynamic modulation of protein function. *Acc Chem Res* 2015, **48**:414–422.

63. Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP: **An automated system to mount cryo-cooled protein crystals on a synchrotron beam line, using compact sample cassettes and a small-scale robot**. *J Appl Crystallogr* 2002, **35**:720–726.

64. Warkentin M, Thorne RE: **Slow cooling of protein crystals**. *J Appl Crystallogr* 2009, **42**:944–952.

65. Team RC: **R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014**. 2014,

66. Emsley P, Lohkamp B, Scott WG, Cowtan K: **Features and development of Coot**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:486–501.

Chapter 3

Divergent conformational dynamics controls allosteric ligand accessibility across evolutionary related αI-domain containing integrins

Citations:

Woldeyes RA*, Hallenbeck KK*, Pfaff SJ, Cortez SV, Kelly MJ, Arkin MR, Fraser JS.

Divergent conformational dynamics controls allosteric ligand accessibility across evolutionary related αI-domain containing integrins. *In preparation*

(*denotes equal contributions by the authors)

Summary

Allosteric communications are essential in the regulation of many biological processes. From GPCRs to kinases, many protein families broadly share mechanisms of allosteric signaling. However, each member of the family might have their own idiosyncratic implementation. The β2 family of integrins, share a general allosteric mechanism that controls their biological response to ligand binding to their orthosteric site. However, the ability of these integrins to be modulated by allosteric modulators which bind to a pocket below the α7 helix differ. Here, we study the link between conformational heterogeneity and allosteric ligandability of two homologous integrins, Lfa-1 and Mac-1 αI-domains. Using small molecule covalent tethering, we determined that the allosteric pocket of Lfa-1 is more "ligandable" than Mac-1's. Biophysical studies showed that the Lfa-1 is more conformationally heterogeneous than Mac-1 and the differing conformational landscape of these homologous proteins critically influences their ability to bind allosteric inhibitors. Despite the restrictive conformational landscape of Mac-1, we were able to find allosteric inhibitors that engages the allosteric site. Taken together, our work suggest that these integrins sample a broadly conserved conformational minima with idiosyncratic energy landscapes connecting them. Furthermore, we show a link between the breadth of the conformational ensemble and the ability to find ligands that bind cryptic allosteric sites. For members of protein families, cryptic allosteric site binders can be more easily identified by importing homologous sequences that can pay the cost of exposing the hidden sites. This strategy could enable the targeting of the “undrugable” proteome.

Introduction

Allostery, the process by which binding at one site can influence the conformational ensemble at distant sites, is an important property in the regulation of proteins [1]. Although the evolutionary origins of allostery may lie in the adaptability of the conformational ensemble [2], many protein families are thought to share conserved allosteric mechanisms. For example, the allosteric activation of GPCRs is marked by an 11 Å displacement of helix 6 in the cytoplasmic region of the receptor in response to ligand binding at the orthosteric site about 40 Å away [3,4]. Similarly, many kinases use protein-protein interactions to stabilize the active state that is competent for catalysis [5]. Although the subtle difference in interactions at the allosteric site (the small molecule binding site for GPCRs or the protein protein interaction surface for kinases), the effect on the distant site is conserved. Whether other aspects of the intramolecular signal transduction mechanisms and the ability to impinge upon them with allosteric modulators are as well conserved remains less well studied.

The β2 family of integrins, which consist of the CD18 β chain and one of 4 α chains (CD11a/Lfa-1, CD11b/Mac-1, CD11c, or CD11d), share a general allosteric mechanism that governs their biological response to ligand binding at the metal ion dependent adhesion site (MIDAS) of the α chain (**Figure 3.1**). X-ray crystallography and Nuclear Magnetic Resonance (NMR) studies of isolated αI-domains show they constitutively adopt the un-bound, ‘low affinity’ conformation, but can be biased to sample ‘high affinity’ conformations with disulfide locks, point mutations, and truncations of the α7 helix [6]. While the allosteric shift of the α7 helix is conserved across αI-domains, they differ in their ability to be modulated by small molecules that bind to an allosteric site located underneath the α7 helix. Small molecules which bind to a pocket below the α7 helix are allosteric inhibitors of I-domain ligand binding, but have been reported only for Lfa-1 (αL, CD11a) [7][8]. These molecules prevent a ~10A shift in the C-terminal α7 helix shifts that normally accompanies ligand binding at the MIDAS site.

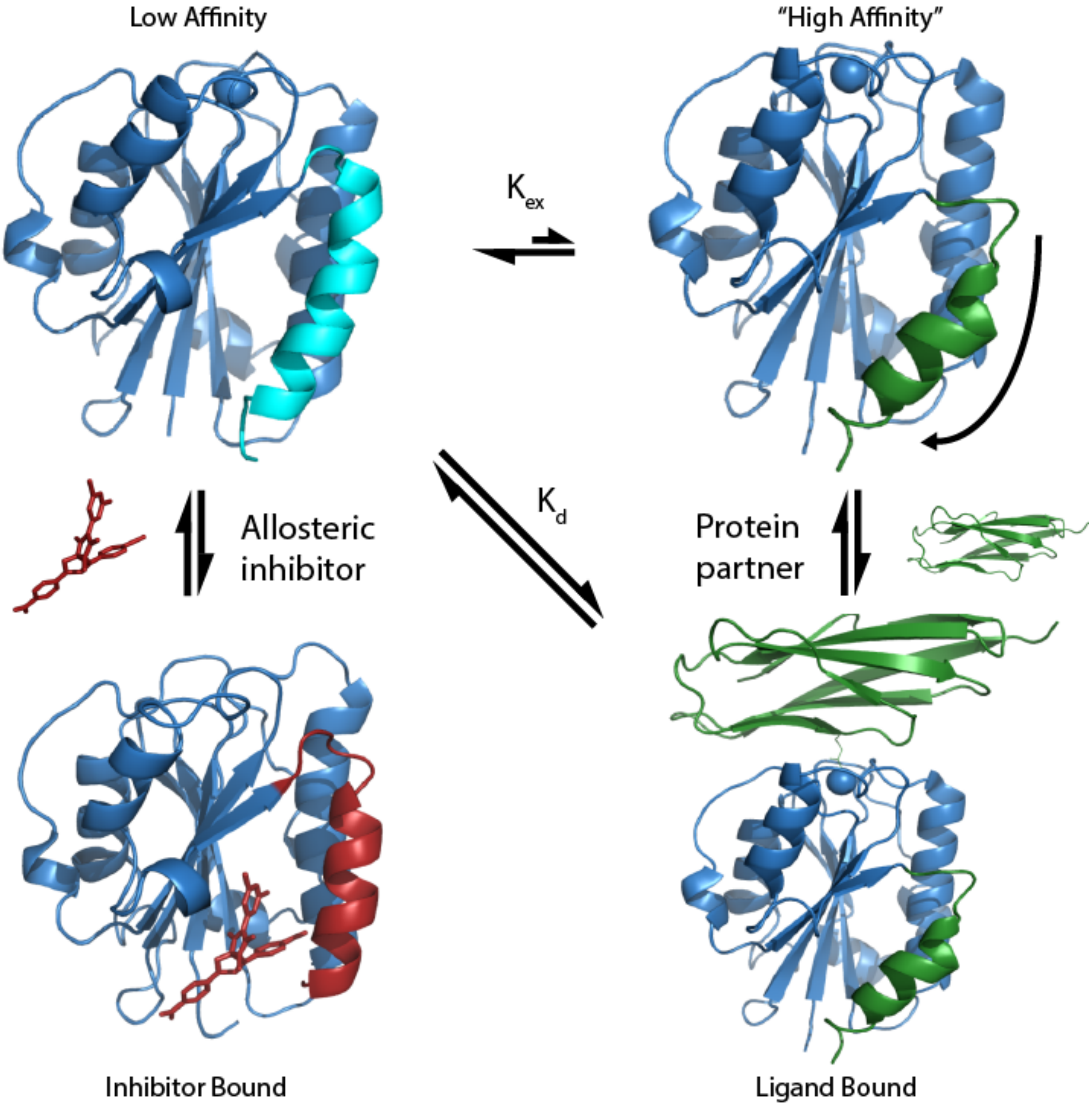


Figure 3.1: The classical thermodynamic cycle of Lfa-1 and Mac-1 I-domains. I-domains can adopt low or high affinity states as marked by the position of the α7 helix (colored cyan or green). Ligand binding to the MIDAS stabilizes the high affinity state. Allosteric ligand binding stabilizes the helix in a low affinity like state.

While the proposed conformational changes leading to integrin activation are thought to be conserved for this family of integrins, it is not clear that the α7 pocket is available in other integrins. For example, Mac-1 is a target for autoimmune diseases such as multiple sclerosis [9,10] but there are no structurally characterized molecules that bind to the Mac-1 α7 pocket. Indeed, although statin drugs interact with the allosteric pocket of Lfa1 [11,12], an X-ray structure of Mac-1 bound to simvastatin shows the the statin's carboxylate moiety coordinating the MIDAS metal ion [11]. This divergent mechanism is noteworthy, given that the Lfa-1 and Mac-1 I-domains adopt nearly identical overall folds.

Here, we have examined the link between conformational heterogeneity, evolutionary conservation, and "ligandability" of the α7 allosteric site for Lfa-1 and Mac-1αI-domains. We used site-directed, small-molecule covalent tethering to determine the "ligandibility" of the allosteric pocket and found that Lfa-1 captures more compounds than Mac-1 under the same experimental conditions. We used biophysical methods to show that the α7 helix of Lfa-1 is more conformationally heterogeneous than the equivalent region in Mac-1.We also showed how allosteric ligand binding can rearrange the MIDAS of Lfa-1 I-domain. Despite the more restrictive conformational landscape of Mac-1, a small number of molecules can be identified that bind to the allosteric site. Collectively, these results paint a picture of broadly conserved conformational minima, but idiosyncratic energy landscapes connecting them. LFA-1 samples a broad ensemble that enables promiscuous allosteric ligand binding whereas Mac-1 transitions between two sharply defined minima and has more stringent ligand binding criteria.

Results

Tethering identifies differential ligandablilty between Mac-1 and LFA-1 I-domains

To determine whether there is an intrinsic difference in the ability of ligands to bind to the allosteric pockets of Lfa-1 and Mac-1I-domains, we used a tethering screen. Tethering is a reversible, covalent labeling reaction using a custom library of disulfide-capped fragments. Under reducing conditions, fragments undergo disulfide-thiol exchange with solvent-exposed thiols (e.g. cysteine) on the target protein's surface, allowing thermodynamic equilibrium to determine the bound vs. unbound populations of protein. [13]. Covalent bond formation is then detected by the mass of the protein +adduct via Liquid Chromatography/Mass Spectrometry LC/MS (**Figure 3.2B-C**) [14].

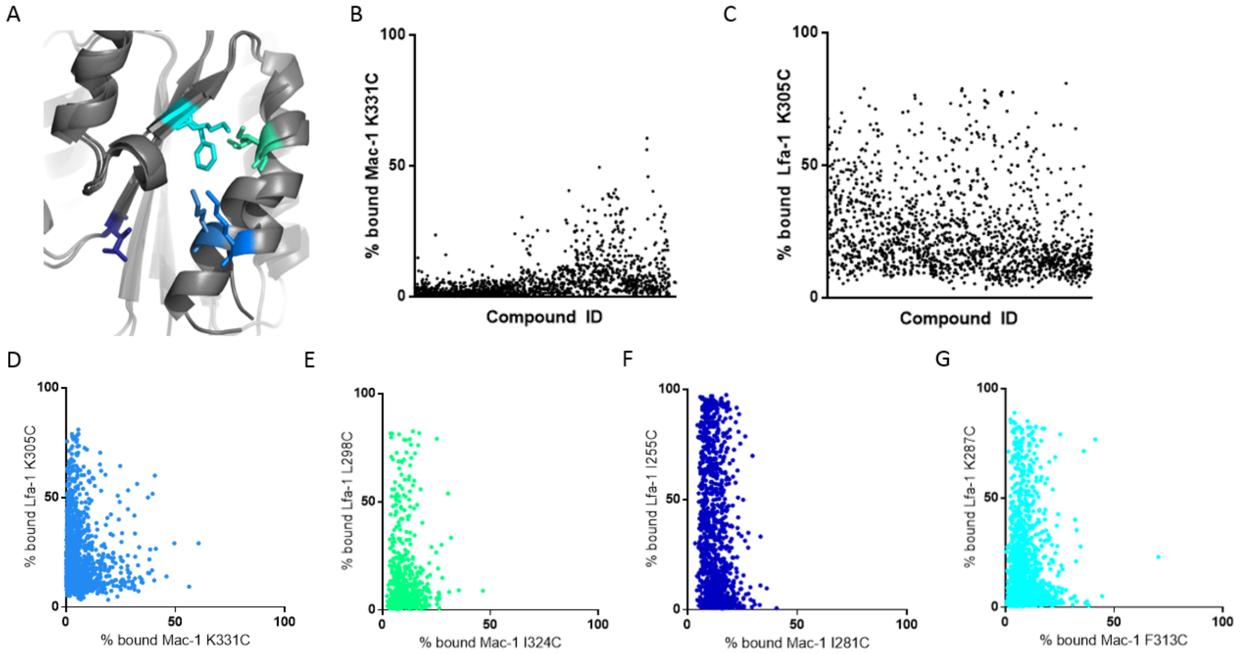


Figure 3.2: Tethering identifies differential ligandablilty. (A) Homologous sites of mutation to cysteine for tethering (**B)** Plots of tethering library molecules indicating the % bound to Mac1 K331C and **(C)** LFA-1 K305C, as determined by mass spectrometry **(D-G)** Correlation plots reflecting the % labeling by each molecule of Mac1 and LFA-1 shows the enrichment of labeling for LFA-1. Lfa-1 K305C and Mac-1 I281C (**Figure 3.2D y-axis, 2F x-axis**) have elevated background labeling, which could be caused by an increase in thiol nucleophilicity.

Because neither Lfa1 or Mac1 I-domains contain native cysteine residues, we introduced matching pairs of cysteines to four sites around the putative allosteric pocket via site-directed mutagenesis. We targeted residues on both the α7 helix itself and the opposing β-strand to remove biases based on individual thiol orientation(**Figure 3.2A**). Next, we screened each cysteine mutation against a custom library of 1600 tethering fragments. Using identical buffer, chromatography, and mass spectrometry conditions, we found the Lfa-1 residue was labeled by more disulfide fragments than its Mac-1 counterpart (**Figure 3.2D-G**).

To compare the screens, we used fragment rank-order to identify labeling efficiency of the 90th percentile fragment in the 1600-fragment library. Lfa-1 K305C, I255C and K287C had cutoffs of 42.5, 80.7, and 44.4 %-bound respectively, whereas Mac-1 K331C, I281C and F313C had much lower cutoffs: 11.3, 17.3, and 14.5 %-bound respectively. The most labeled Mac-1 mutation, K331C, is located on the α7 helix. In contrast, the highest labeling Lfa-1 mutation was I255C, located deep in the α7 pocket (**Figure 3.2F**). The greater accessibility in Lfa-1 of the buried cysteine residue to disulfide bond formation further indicates that the pocket is more open than in Mac-1. Thus, the high hit-rate, high %-binding of tethering hits, and the greater that the Lfa-1 pocket is more open. These comparative screens show that, surprisingly, the Mac1 allosteric site is much less accessible than is the analogous site on Lfa-1. suggest that the accessibility of the cysteine residue at the back of the Lfa-1 pocket may dominate the differences in ligandability.

X-ray crystallography suggests differences in flexibility surrounding α7 pocket

To explore the structural basis for the differential ligandability of the Lfa-1 and Mac-1 allosteric pockets, we analyzed the structures deposited in the PDB. Alignment of all the structures except for disulfide bond stabilized mutants of Lfa-1 in the PDB shows that heterogeneity is greatest in the α1- β2 loop and the α7 helix (Figure 3.3A-C). Alignment of Mac-1 I-domain structures show that the α1- β2 loop is also dynamic in Mac-1(**Figure 3.3B-D**). The α7 helix of Mac-1 can be segregated into two different classes, each of which has limited heterogeneity. In contrast, and in agreement with the multistate dynamics of Lfa-1 proposed by Vendruscolo, the α7 helix of Lfa-1 cannot be neatly clustered into two populations unless a ligand is bound to the MIDAS[15].

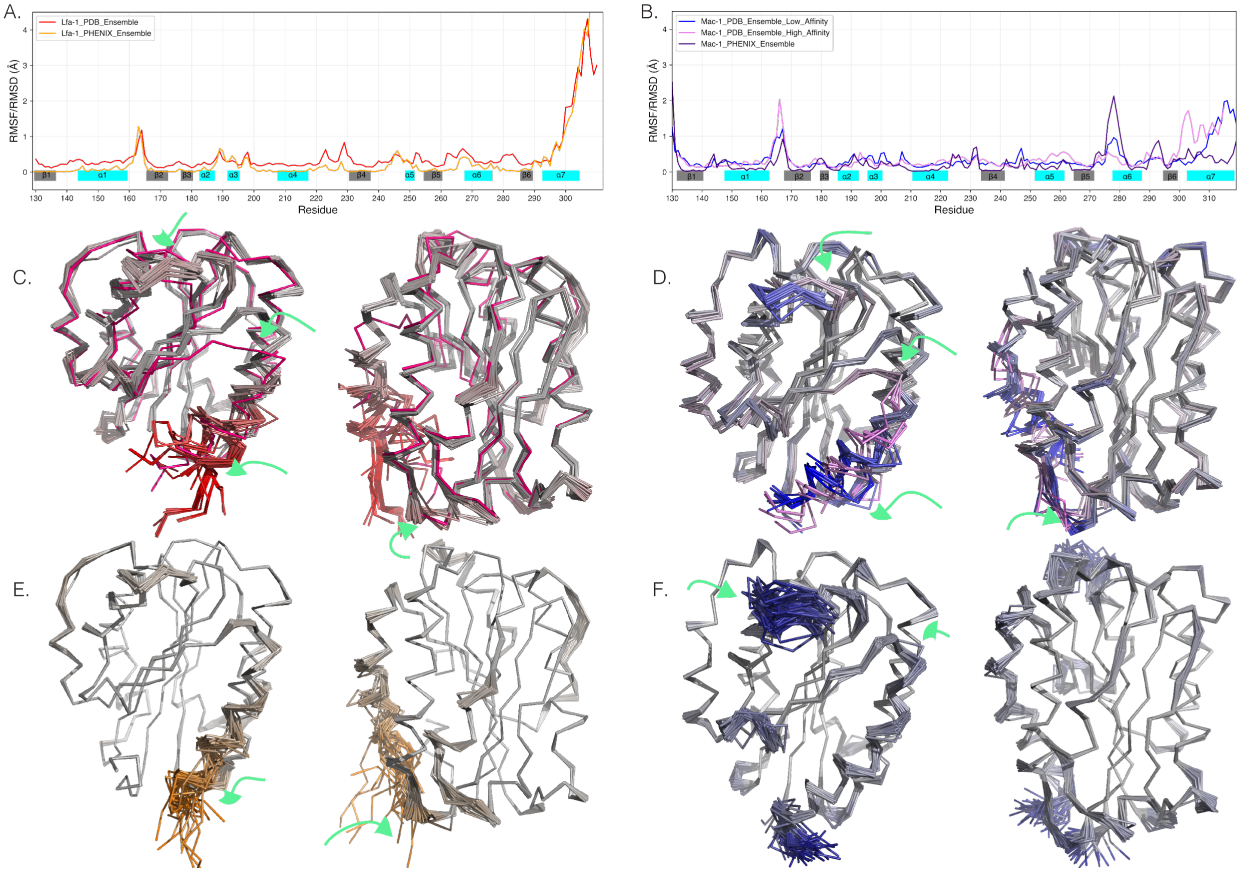


Figure 3.3. X-ray crystallography suggests differences in flexibility (A) Plot of c-α RMSD/RMSF by sequence for an alignment of all deposited Lfa-1 low affinity structures (red) and ensemble refinement of 1.8 Å room temperature data (orange) are consistent and identify the mobility of the α7 helix and the α1-β2 loop as regions of high mobility. (**B)** The time-averaged ensemble (purple) c-α RMSF results for Mac1 compared to all deposited WT Mac-1 PDB structures, which can be segregated into high (pale-pink) and low (blue) affinity states. Compared to LFA-1, this analysis highlights high variability in the β5-α6 loop and lower variability in the α7 helix. (**C)** Structural overlay of the PDB ensemble for Lfa-1 colored by RMSD values in a gradient (grey to red). Colored in pink is the ligand stabilized high affinity structure of Lfa-1 (**D)** Structural overlay of Mac-1 PDB ensemble colored by RMSD from (grey to blue or pale-pink) (**E)** The final time-averaged ensemble for Lfa-1 colored by RMSF (grey to orange) (**F)** The final phenix ensemble for Mac-1 colored by RMSF(grey to purple).The green arrows show areas with high conformational heterogeneity

To test the idea that the α7 helix of Lfa-1 is variable even within a single dataset, we collected high-resolution, room-temperature crystal structures of the Mac-1 and Lfa-1 I-domains (**Table S3.1,2**). These structures align well with previously reported structures that were collected on cryo-cooled crystals, but collection at room temperature increases our ability to visualize alternate conformations. For Lfa-1 the a7 helix has no density after residue 308, whereas backbone density extends to residue 322(extra five residues) in Mac-1, just four residues short of the construct C-terminus. The Lfa-1 structure adopted a native NMR and full headpiece like conformation of the α7 helix. Time averaged ensemble refinement, using phenix ensemble refinement, of Lfa-1 revealed a very similar heterogeneity profile as the PDB ensemble (**Figure 3.3C**). The most dynamic areas in both cases are the α1- β2 loop and the α7 helix. Similarly, the Mac-1 time-averaged ensemble also captures similar heterogeneity as the PDB ensemble for the low affinity conformation with reduced heterogeneity for the α7 helix. From these analyses, we conclude that Lfa-1 seems to sample a large range of conformations. This energy landscape allows Lfa-1 to readily create an allosteric pocket adjacent to the helix for ligands to bind. However in Mac-1, the two main conformations sampled create a high energetic cost to forming a homologous allosteric pocket.

	Lfa-1_RT	Lfa-1_RT_qFit	Lfa-1_RT_Ensemble
Wavelength	1.116		
Resolution range	44.90 - 1.80 (1.86 - 1.80)	44.90 - 1.80 (1.86 - 1.80)	44.90 - 1.80 (1.86 - 1.80)
Space group	P 32 2 1	P 32 2 1	P 32 2 1
Unit cell	104.90 104.90 51.64 90 90 120	104.90 104.90 51.64 90 90 120	104.90 104.90 51.64 90 90 120
Total reflections	182242 (13598)		
Unique reflections	30598 (2961)	30569 (2961)	30569 (2961)
Multiplicity	6.0 (4.5)		
Completeness (%)	91.90 (74.72)	91.90 (74.72)	91.89 (74.72)
Mean I/sigma(I)	10.55 (0.64)		
Wilson B-factor	22.67	22.67	22.67
R-merge	0.1205 (1.775)		
R-meas	0.1318 (2.007)		
R-pim	0.05274 (0.92)		
CC1/2	0.995 (0.296)		
CC*	0.999 (0.676)		
Reflections used in refinement	30569 (2240)	30569 (2240)	30569 (2240)
Reflections used for R-free	1827 (153)	1827 (153)	1826 (153)
R-work	0.1702 (0.3336)	0.1433 (0.2775)	0.1746 (0.3323)
R-free	0.2033 (0.3228)	0.2032 (0.3505)	0.2141 (0.3565)
CC(work)	0.960 (0.659)		
CC(free)	0.929 (0.712)		
Number of non-hydrogen atoms	1674	2745	57336
Macromolecules	1511	2582	54396
ligands	1	1	36
solvent	162	162	2904
Protein residues	188	188	188
RMS(bonds)	0.011	0.006	0.018
RMS(angles)	1.29	0.97	1.74
Ramachandran favored (%)	95.16	94.62	88.71

	Lfa-1_RT	Lfa-1_RT_qFit	Lfa-1_RT_Ensemble
Ramachandran allowed (%)	4.84	4.30	8.06
Ramachandran outliers (%)	0.00	1.08	3.23
Rotamer outliers (%)	1.20	2.29	11.24
Clashscore	2.64	3.76	0.00
Average B-factor	31.15	29.27	30.26
Macromolecules	30.20	28.61	30.22
ligands	20.72	20.75	26.19
solvent	40.09	39.96	31.15

Table S3.2. Data collection and refinement statistics for room temperature Lfa-1 I-domain structures.

	Mac-1_RT	Mac-1_RT_Ensemble
Wavelength	1.116	
Resolution range	46.04 - 1.71 (1.77 - 1.71)	46.04 - 1.71 (1.77 - 1.71)
Space group	P 21 21 21	P 21 21 21
Unit cell	37.57 51.35 103.92 90 90 90	37.57 51.35 103.92 90 90 90
Total reflections		
Unique reflections	22162 (2126)	22162 (2126)
Multiplicity		
Completeness (%)	98.74 (97.43)	98.74 (97.43)
Mean I/sigma(I)		
Wilson B-factor	23.34	23.34
R-merge		
R-meas		
R-pim		
CC1/2		
Reflections used in refinement	22162 (2124)	22162 (2124)
Reflections used for R-free	1113 (107)	1113 (107)
R-work	0.1729 (0.3213)	0.1372 (0.2820)
R-free	0.2104 (0.3749)	0.1956 (0.3675)
Number of non-hydrogen atoms	1664	90358
macromolecules	1560	87360
ligands	2	112
solvent	102	2886
Protein residues	193	193
RMS(bonds)	0.004	0.015
RMS(angles)	0.73	1.69
Ramachandran favored (%)	97.91	90.05
Ramachandran allowed (%)	2.09	6.81
Ramachandran outliers (%)	0.00	3.14
Rotamer outliers (%)	1.16	13.51
Clashscore	3.51	0.00

	Mac-1_RT	Mac-1_RT_Ensemble
Average B-factor	28.10	22.83
macromolecules	27.49	22.81
ligands	38.55	26.08
solvent	37.20	23.52

Table S3.2. Data collection and refinement statistics for room temperature Mac-1 I-domain structures.

NMR confirms differential α7 flexibility between Lfa-1 and Mac-1

To test whether the conformational heterogeneity observed by our room-temperature crystal structures were present in solution, we turned to NMR. We performed ^{15}N-^{1}H HSQC experiments on both I-domains and analyzed line broadening using peak volumes. This method, when combined with knowledge of secondary structure from crystallography, allows for the rapid assessment of which protein regions are exchanging between multiple conformations on the chemical shift time scale.

For Lfa-1, we observed significant line broadening around α1- β2 loop and the α7 helix as expected (**Figure 3.4A**). This observation confirms previous findings by Legge et al. who have shown the flexibility of the helix using heteronuclear NOE and deuterium exchange experiments and aligns well with the reduced restraints on the α7 helix reported in the Lfa-1 NMR structure[16]. The analysis also showed additional broadening around the α3-α4 loop which contains one of the five residues that make up the MIDAS.

To compare the Lfa-1 to Mac-1, we assigned the backbone resonances of Mac-1 I-domain, achieving 72% sequence coverage, including the entire α7 helix and allosteric pocket. Line broadening analysis shows that in contrast to the Lfa-1 α7 helix resonances, the Mac-1 α7 helix cross peaks remain sharp through to the C-terminus (**Figure 3.4B**). These data suggest the α7 helix of Mac-1 I-domain sits in a sharp local minima, and its conformational transitions occur on a timescale slower than $^{15}N/^{1}H$ signal relaxation. Collectively the NMR and X-ray results suggest that the differences in ligandability are paralleled by the differences in conformational heterogeneity between Mac-1 and Lfa-1.

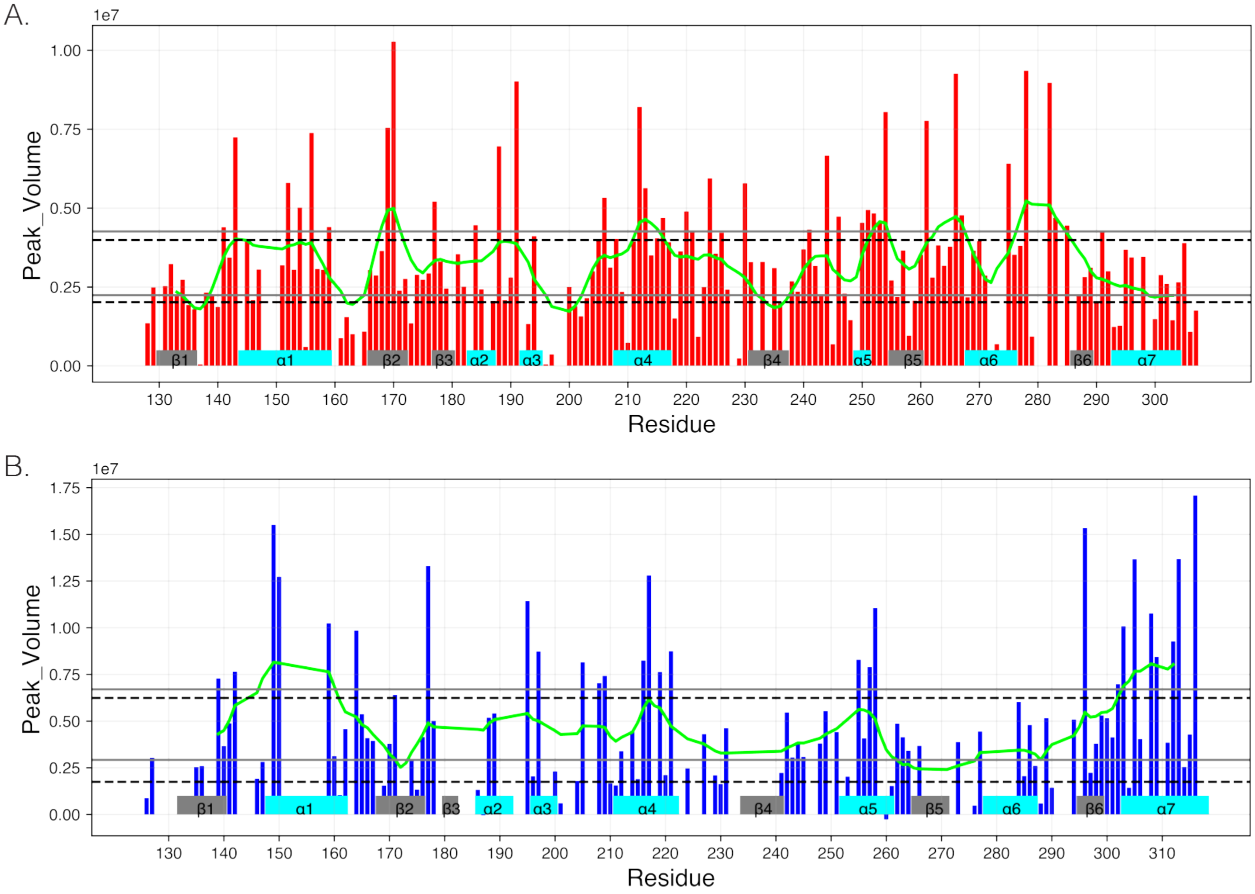


Figure 3.4 NMR confirms differential α7 flexibility (A) Peak volumes from a ^{15}N-^{1}H HSQC for assigned residues in Lfa-1 (red bars) with the rolling average (green). Residues along α7 helix are generally below +/- 1 standard deviation from the mean(grey solid lines) and +/- the iqr value from the median(black dotted lines). The rolling average is depressed to the bottom 25 percentile values for residues on the α7 helix. **(B)** A similar analysis for Mac1 peak volumes (blue bars) indicates that a7 helix is more rigid in Mac-1. The α7 helix residues have sharpened peaks with peak volume values that are mostly greater than +/- 1 standard deviation from the mean

Identifying mutation sites to destabilize Mac-1 α7 helix based on structures of LFA-1 bound to allosteric modulators.

We hypothesized the Mac-1 pocket ligandability could be modulated with a mutation designed to decrease the stability of Mac-1, particularly at the α7 helix. To identify candidate sites in Mac-1 which could influence pocket ligandability, we determined co-crystal structure of Lfa-1 bound to BMS-68852 and Lovastatin at room temp (**Table S3.3**). In addition to the expected local changes around α7 helix, we observed rearrangements around the MIDAS site in the BMS-68852 structure (**Figure 3.5**). This rearrangement would prevent productive MIDAS formation resulting in inhibition of ligand binding at the site. We also see a similarly perturbed MIDAS site in 1 of the 2 mol in the crystallographic asymmetric unit when bound to Lovastatin (**Figure 3.5C**).

	Lfa-1 BMS-688521	Lfa-1 BMS-688521_qFit	Lfa-1 Lovastatin	Lfa-1 Lovastatin_qFit
Wavelength	1.116		1.116	
Resolution range	65.64 - 1.86 (1.93 - 1.86)	65.64 - 1.86 (1.93 - 1.86)	61.56 - 1.85 (1.92 - 1.85)	61.56 - 1.85 (1.92 - 1.85)
Space group	P 2 21 21	P 2 21 21	I 1 2 1	I 1 2 1
Unit cell	36.51 75.41 133.29 90 90 90	36.51 75.41 133.29 90 90 90	66.89 46.08 127.23 90 104.61 90	66.89 46.08 127.23 90 104.61 90
Total reflections	61375 (5972)		64475 (6363)	
Unique reflections	30777 (2994)	30771 (2994)	32276 (3188)	32272 (3188)
Multiplicity	2.0 (2.0)		2.0 (2.0)	
Completeness (%)	96.54 (94.93)	96.55 (94.93)	99.82 (99.72)	99.81 (99.72)
Mean I/sigma(I)	6.28 (2.79)		5.19 (0.75)	
Wilson B-factor	16.99	16.99	20.00	20.00
R-merge	0.05167 (0.2254)		0.1019 (1.055)	
R-meas	0.07307 (0.3188)		0.1441 (1.491)	
R-pim	0.05167 (0.2254)		0.1019 (1.055)	
CC1/2	0.993 (0.791)		0.993 (0.37)	
CC*	0.998 (0.94)		0.998 (0.735)	
Reflections used in refinement	30771 (2994)	30771 (2994)	32272 (3186)	32272 (3186)
Reflections used for R-free	2000 (195)	1999 (195)	2000 (198)	2000 (198)
R-work	0.1672 (0.2002)	0.1372 (0.1583)	0.1830 (0.3099)	0.1450 (0.2221)
R-free	0.1999 (0.2542)	0.2096 (0.2605)	0.2130 (0.3372)	0.2274 (0.3182)
CC(work)	0.952 (0.872)		0.965 (0.624)	
CC(free)	0.947 (0.871)		0.949 (0.703)	
Number of non-hydrogen atoms	3697	5437	3543	5121
macromolecules	3379	5119	3244	4822
ligands	75	75	60	60

	Lfa-1 BMS-688521	Lfa-1 BMS-688521_qFit	Lfa-1 Lovastatin	Lfa-1 Lovastatin_qFit
solvent	243	243	239	239
Protein residues	364	364	367	367
RMS(bonds)	0.012	0.012	0.006	0.024
RMS(angles)	1.12	1.68	0.91	1.10
Ramachandran favored (%)	95.28	93.61	94.77	95.32
Ramachandran allowed (%)	4.17	5.28	4.96	4.68
Ramachandran outliers (%)	0.56	1.11	0.28	0.00
Rotamer outliers (%)	2.92	2.02	1.39	2.67
Clashscore	7.23	4.82	2.26	10.10
Average B-factor	22.23	19.32	26.11	24.47
macromolecules	21.59	18.75	25.20	23.59
ligands	16.00	16.06	35.20	45.95
solvent	33.06	32.47	36.09	36.66

Table S3.3. Data collection and refinement statistics for room temperature co-crystal structures of Lfa-1 I-domain

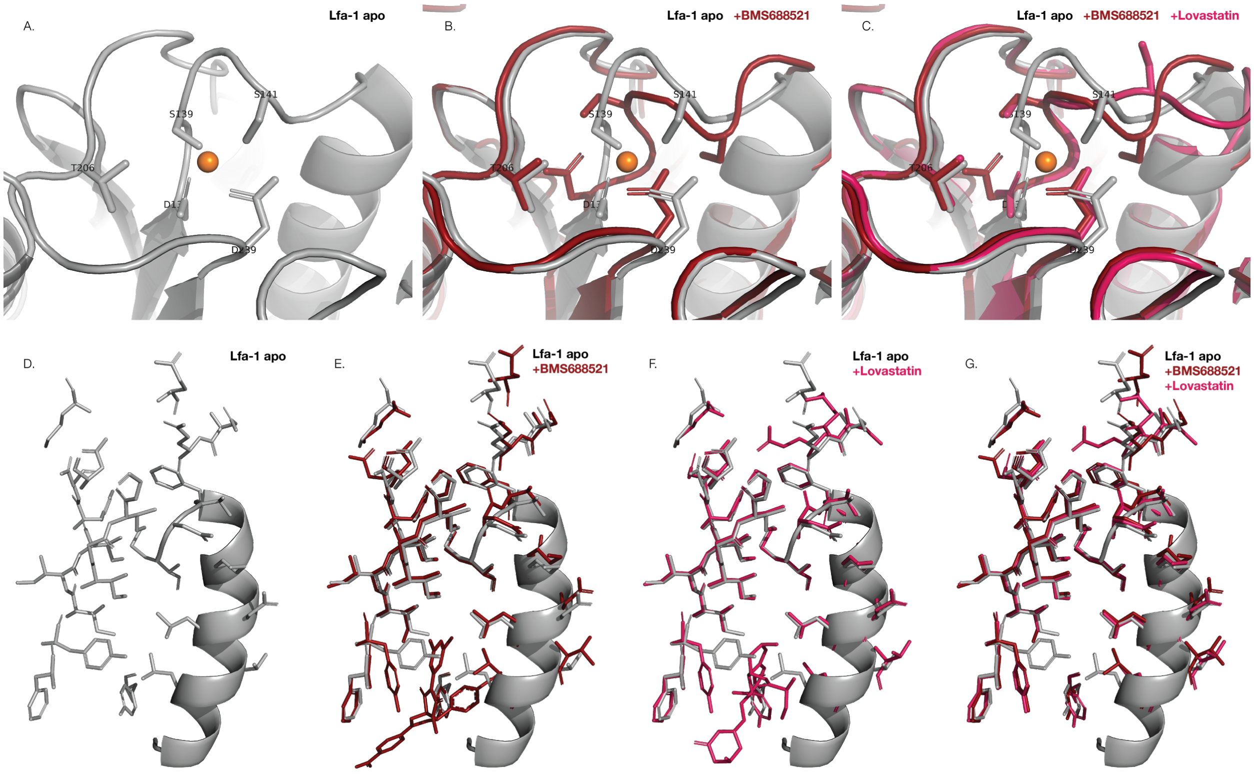


Figure 3.5. MIDAS and allosteric site rearrangements seen in the X-ray structures of inhibitor bound Lfa-1 I-domain (A) Lfa-1 apo MIDAS(grey) with the coordinating magnesium(orange)**. (B)** BMS-68852 binding rearranges the MIDAS residues(crimson)**. (C)** Lovastatin binding can also rearrange the MIDAS(ruby). Bottom panels show perturbation of the unliganded Lfa-1 I-domain **(D).** when bound to BMS-68852 **(E)** and bound to Lovastatin **(F).** Overlay of all the structures is shown on the last panel**(G) .**

Consistent with this long range structural perturbation, we observed that titrating BMS-68852 or Lovastatin into Lfa-1 shifts backbone NMR peaks throughout the protein, including the MIDAS site (**Figure 3.6A,B & Figure S3.1**). To connect these shifts to molecular mechanism we used CONTACT to identify networks of sterically incompatible alternative conformations. The analysis showed that there are network of residues that can connect the allosteric site of Lfa-1 to the MIDAS site (**Figure 3.6C,D & Figure S3.2**). We see similar sets of residues contributing to the coupling between the allosteric site and the MIDAS in our CSP and CONTACT analysis, extending from the core of the protein to MIDAS residue Asp 137 and Ser 139 on the β1-α1 loop (**Figure 3.6C-D**). This suggests that the effect of these compounds is two-fold: they both block the movement of α7 helix and promote a distorted MIDAS site.

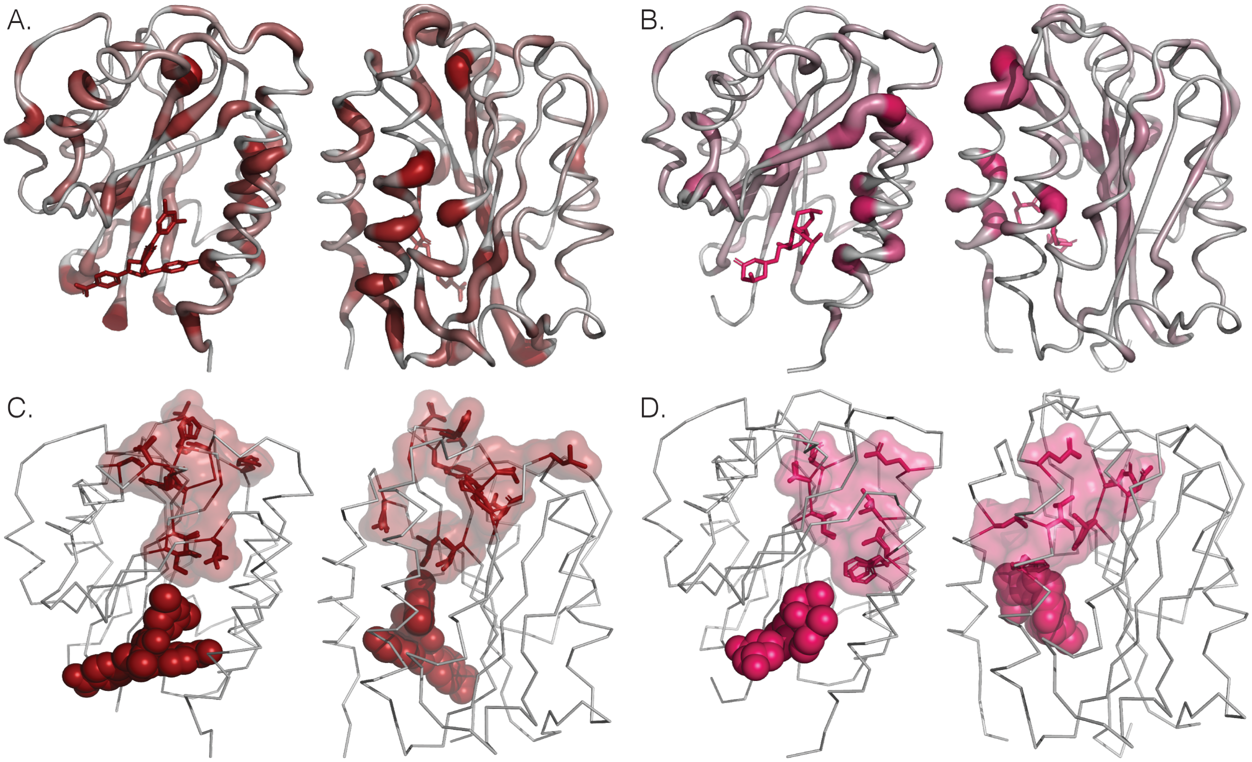
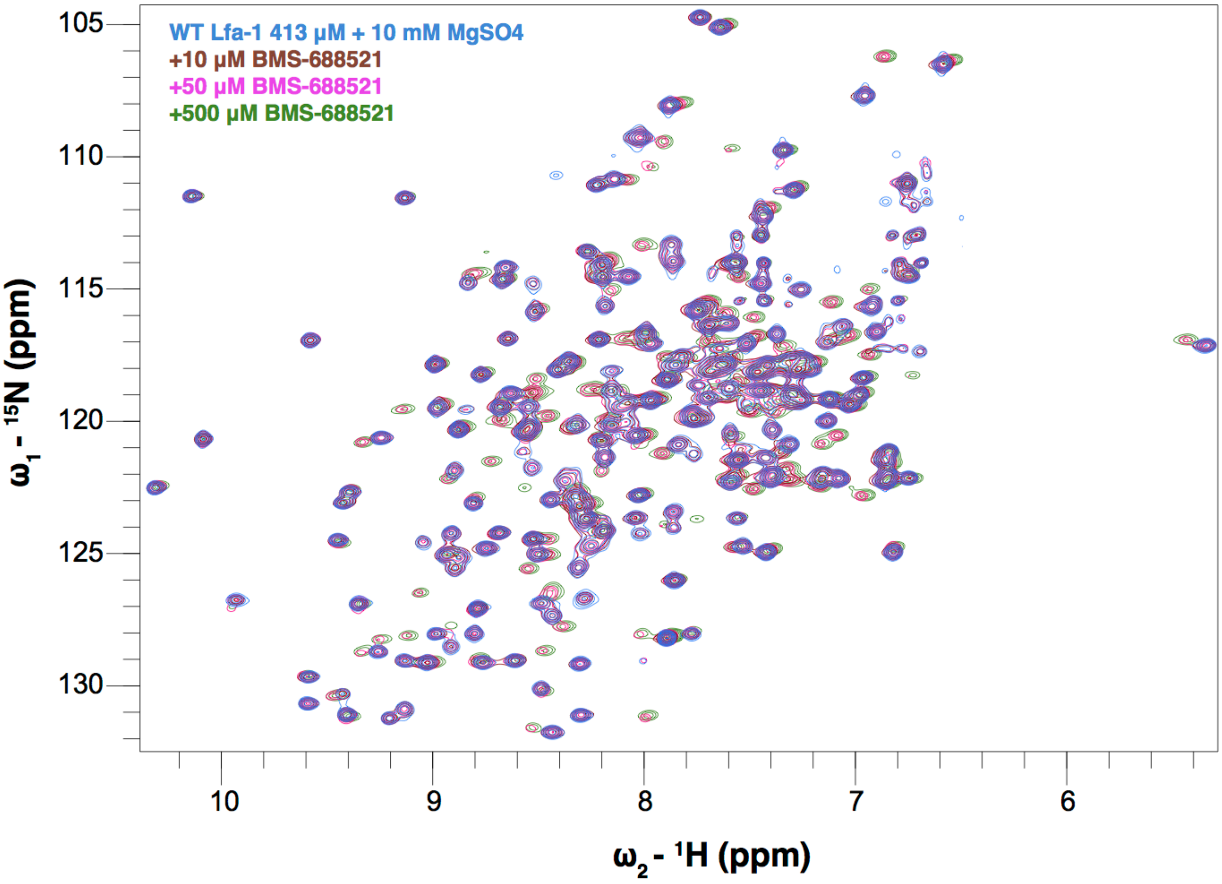


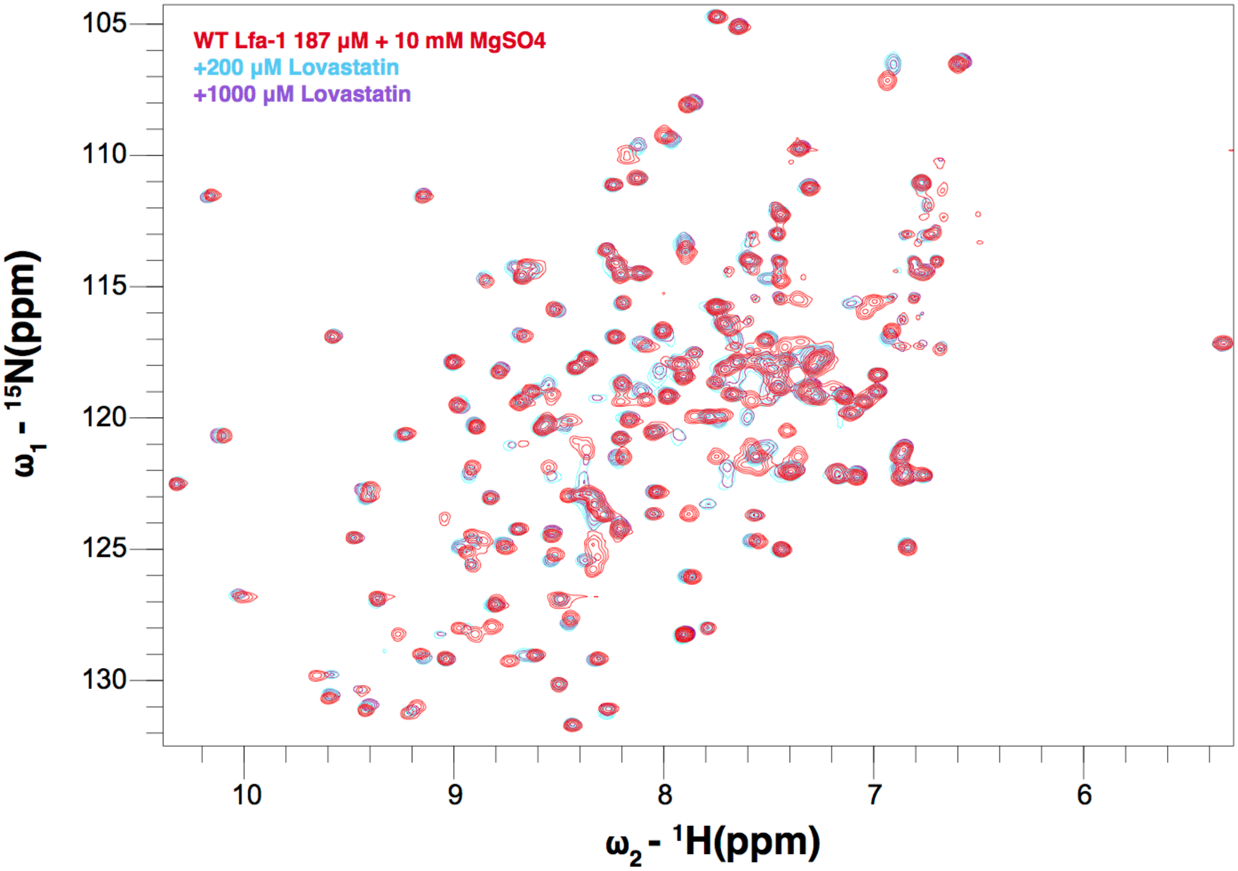
Figure 3.6. Network of perturbed residues connecting Lfa-1 I-domain allosteric site to the MIDAS (A) Chemical shift perturbation values from BMS-68852 titration experiments mapped onto the inhibitor bound structure. CSP values are displayed in a putty representation and a color gradient from (grey to crimson)(from 0 to 0.25 ppm). **(B)** CSP values from titration experiment using Lovastatin shown on Lovastatin bound structure. The values are shown in a putty representation and are colored by CSP (grey to Ruby) **(C)** Network of residues that connect the allosteric site to the MIDAS when BMS-68852 is bound are shown as sticks with a surface rendering of the volume of the residues(crimson). BMS-68852 is shown in a sphere representation **(D)** Contacting networks when Lovastatin is bound(ruby). Lovastatin is shown as spheres.

Figure S3.1. Spectra from titration experiments with BMS-68852(A) or Lovastatin(B). Cross peaks for the MIDAS**(C)**, “Ratchet pocket”**(D)** and α7 helix residues**(E)**

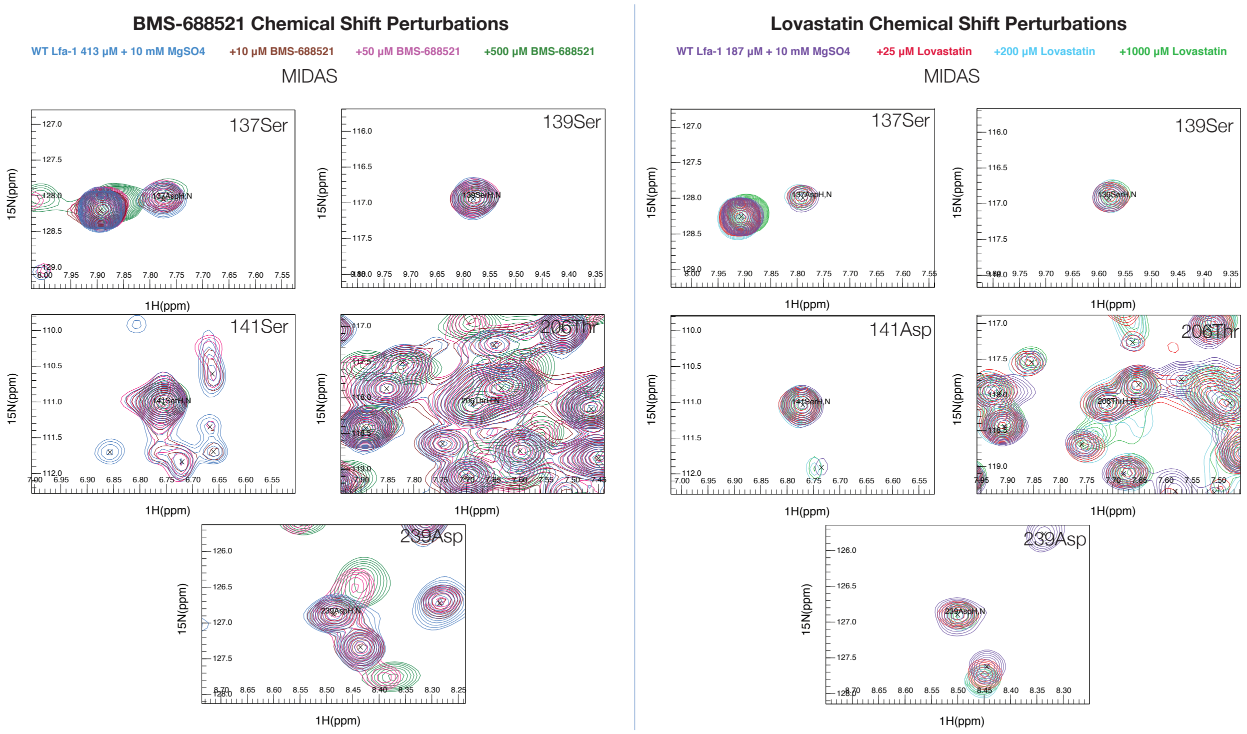
A.



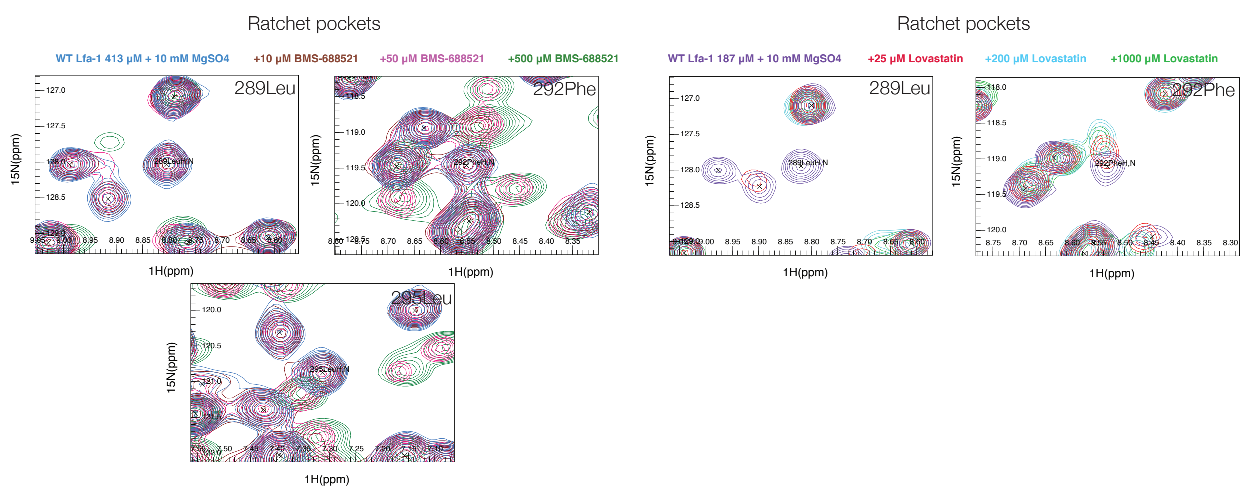
B.


WT Lfa-1 187 µM + 10 mM MgSO4
+200 µM Lovastatin
+1000 µM Lovastatin
105
110
115
120
125
130
10
9
8
7
6
ω1 - 15N(ppm)
ω2 - 1H(ppm)

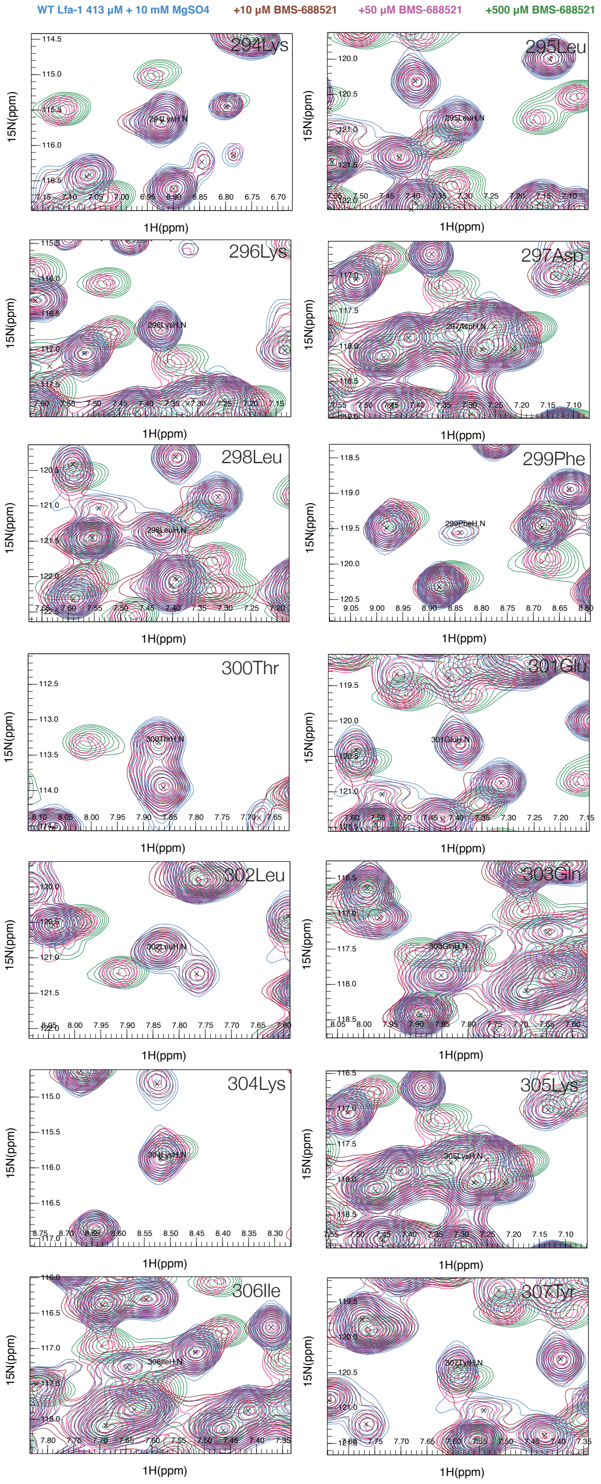
C.


BMS-688521 Chemical Shift Perturbations
WT Lfa-1 413 µM + 10 mM MgSO4
+10 µM BMS-688521
+50 µM BMS-688521
+500 µM BMS-688521
MIDAS
137Ser
139Ser
141Ser
206Thr
239Asp
Lovastatin Chemical Shift Perturbations
WT Lfa-1 187 µM + 10 mM MgSO4
+25 µM Lovastatin
+200 µM Lovastatin
+1000 µM Lovastatin
MIDAS
137Ser
139Ser
141Asp
206Thr
239Asp
15N(ppm)
1H(ppm)

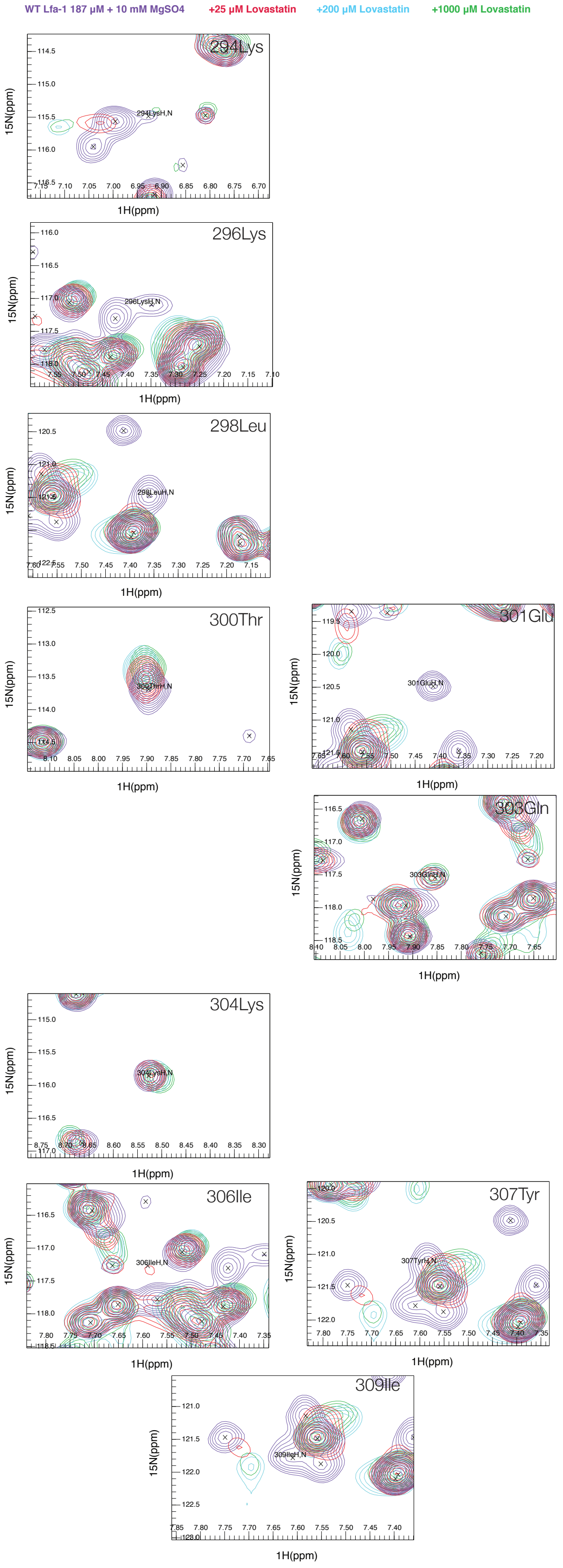
D.


Ratchet pockets
WT Lfa-1 413 µM + 10 mM MgSO4
+10 µM BMS-688521
+50 µM BMS-688521
+500 µM BMS-688521
289Leu
292Phe
295Leu
Ratchet pockets
WT Lfa-1 187 µM + 10 mM MgSO4
+25 µM Lovastatin
+200 µM Lovastatin
+1000 µM Lovastatin
289Leu
292Phe
15N(ppm)
1H(ppm)

Alpha 7 helix (294-309)



Alpha 7 helix (294-309)



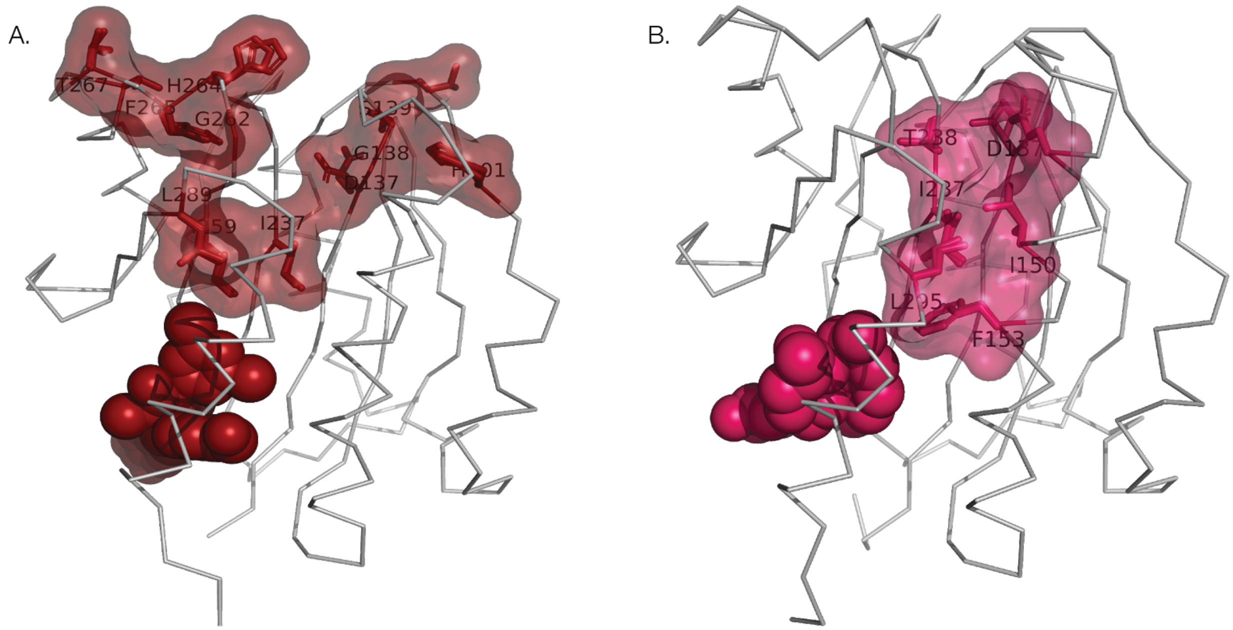


Figure S3.2. Network of contacting residues with BMS-68852(A) or Lovastatin(B) bound. Residues that are part of the network are shown as sticks and the allosteric ligands are shown as spheres.

We then inspected the backbone and sidechain conformational changes between apo Lfa-1 and BMS-68852-bound Lfa-1 and cross-referenced candidate residues where the amino acid identity between Mac1 and Lfa1 diverge. These analyses identified Mac-1 Q325/Lfa-1 F299, which sits on the solvent-exposed side of the α7 helix and packs against the nearby α1 helix, as a good candidate.

Destabilizing mutation in Mac-1 does not alter α7 pocket ligandability

Based on the analysis of the LFA-1 co-crystal structures and sequence differences, we mutated Q325 to Phe in Mac-1. Mac-1 Q325F expressed and purified similarly to wild type. To test whether the mutation had a global effect on stability, we performed thermal denaturation melts and monitored the effect by circular dichroism spectroscopy. These assays revealed that Q325F is thermally destabilized by 3.4°C relative to wild type (**Figure 3.7B**). We confirmed the mutant protein had similar affinity for canonical Mac-1 protein partner C3dg as WT I-domain using surface plasmon resonance (SPR). To test whether Q325F increased conformational heterogeneity in the Mac-1 α7 helix, we expressed U-15N K331C/Q325F, collected a 1H/15N HSQC spectrum, and performed line-broadening analysis (**Figure 3.7D**). In WT Mac1, the α7 helix was the most intense region of the spectrum, but in Q325F the α3 helix (**Figure 3.4B**), suggesting the sequence swap did destabilize the Mac-1 a7 helix, but were not as broadened as those observed in the homologous region of Lfa-1 (**Figure 3.4A**). This result suggested that the stability of this region was subtly perturbed, but that it is only a partial contributor to the 3.4 °C change in melting temperature.

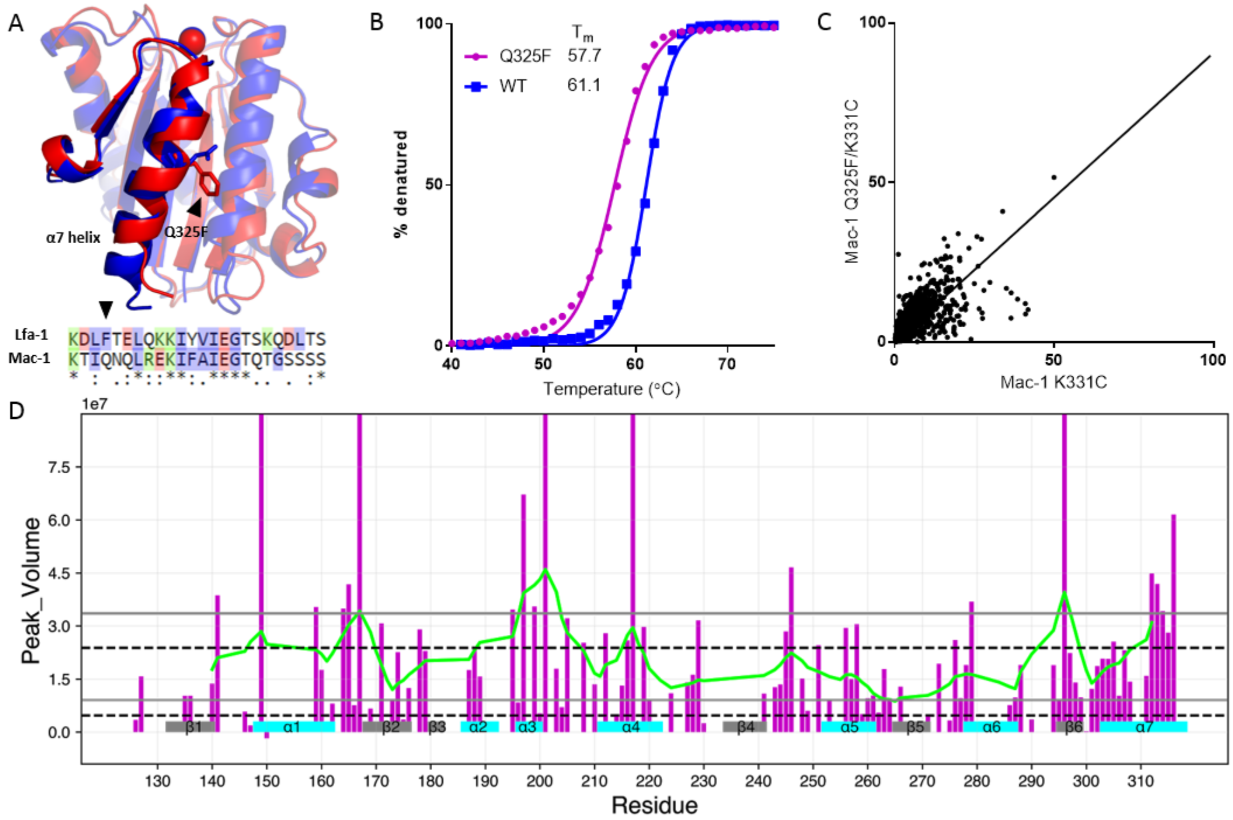


Figure 3.7. Destabilizing mutation in Mac-1 does not alter α7 pocket ligandability **(A)** Design for a Glu to Phe point mutation to destabilize Mac-1 based on a comparison of the Mac-1 (blue) and Lfa-1 (red) α7 helix sequences. **(B)** The Mac-1 Q325F (pink) mutation results in decreased thermal stability detected by CD. **(C)** Disulfide tethering screen comparing Mac-1 K331C/Q325F (y-axis) to the single mutant K331C (x-axis), does not show a difference in ligandability. **(D)** 15N-HSQC peak volumes of Mac-1 K331C/Q325F mutant. Compared to WT, the α7-helix of K331C/Q309F has fewer residues above the moving average (green line).

Next, we tested whether this variant (with an additional K331C mutation to enable disulfide tethering) increased the ligandability of the α7 pocket. Surprisingly, we found no significant impact on the ligandability of the Mac-1 pocket compared to Mac-1 K331C (**Figure 3.7C**). This result suggests that primary result of the mutation was to destabilize conformations relative to the unfolded state and that conformations that promote the allosteric pocket remain energetically disfavored.

Mac-1 α7 pocket can bind fragments

Although analysis of WT and Q325F Mac-1 variants identified a much smaller number of molecules capable of covalently labeling the protein, a small number of compounds (42) labelled K331C. K331C also had highest-labeling fragment at 60.6 %-bound (**Figure 3.2D**). We therefore focused on dose-response experiments to identify hits capable of fully labeling K331C. From these experiments, we identified SMDC917607 (**Figure 3.8B**), a methyl-hydroxy quinoline capable of fully labeling K331C with a EC50 of 87 µM in 500uM βME (**Figure 3.8A**). Notably, several analogs of SMDC917607 contained in our library showed no appreciable binding to K331C (**Figure 3.8B**). Extending the aliphatic linker from 2 to 3 carbons, removing the substitutions to the quinoline ring, or consolidating to a single ring system all ablate compound binding. Increasing the reducing potential of the buffer to 1 mM βME also ablated binding (data not shown), implying a weak interaction driven by non-covalent contributions of the fragment rather than disulfide bond formation .

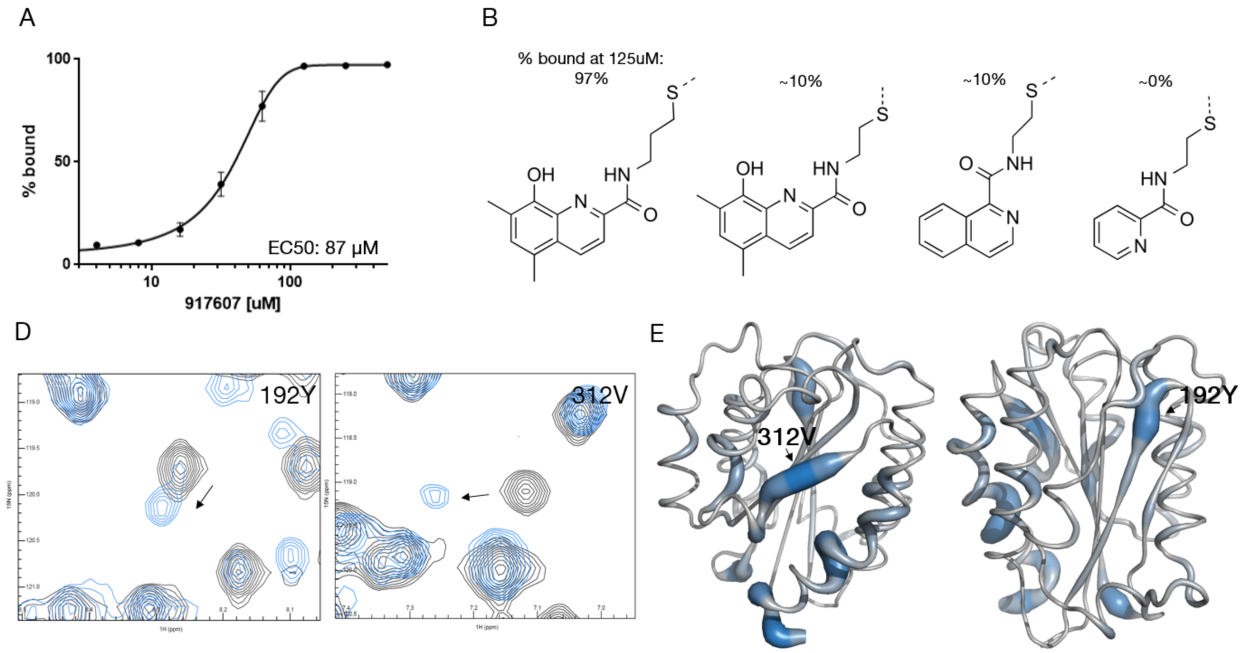


Figure 3.8. Hit Validation and Binding Mechanism (A) Dose response binding of SMDC917607 and analogs to Mac-1 K331C detected by intact protein LC/MS. **(B)** Chemical structure of the fragment SMDC917607 and analogs present in screening library. **(C)** 15N-HSQC titration of SMDC917607 into U-15N WT Mac-1 I-domain, showing chemical shift perturbations. **(D)** Mapping of CSP from panel C to the Mac-1 I-domain X-ray structure.

To identify the binding mechanism of SMDC917607, we collected 15N-HSQC NMR chemical shift perturbation (CSP) experiments in K331C Mac-1 I-domain (**Figure 3.8C & Figure S3.3**). The residues demonstrating CSP >1 STD from the mean at 1.1 mM 917607 were mapped to the Mac-1 X-ray structure (**Figure 3.8E).** Because the lone pair donors of SMDC917607's hydroxy quinoline could potentially chelate metal ions, such as the Mg2+ present in the Mac-1 MIDAS, we included a 10 mM excess of $MgCl_2$ during data collection. V312, near K331C, undergoes the most intense CSP of 9.5-fold the mean CSP, suggesting V312 makes direct contact with SMDC917607 (**Figure 3.8E**). Nearby aromatic residues F250, F313 display CSP, as well I335 and G337, residues in the unstructured C-terminus of the α7 helix. Y192 is the only residue which shifts in regions distant from the tethering site on the α7 helix. Notably, Y192 sits directly below the MIDAS, providing evidence for allosteric coupling of the α7 and MIDAS, and suggesting molecules which bind Mac-1 can induce conformational change at distant sites.

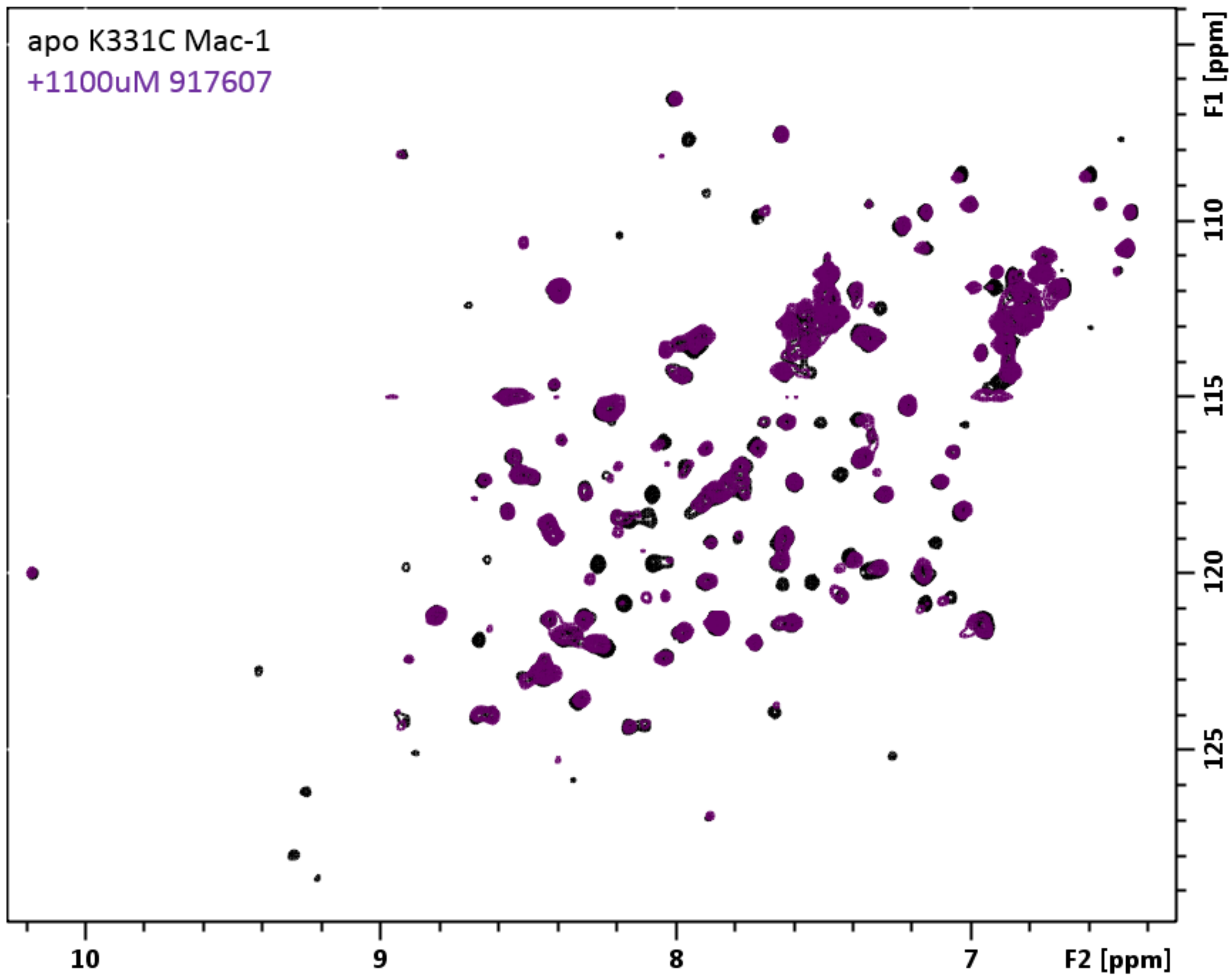


Figure S3.3: Mac-1 titration experiment spectra with the unbound Mac-1 K331C(black) and bound to SMDC 917607(purple)

Discussion

Protein ligandability is a function of both the chemistry at the binding site and the accessibility of the site[17]. While Mac-1 and Lfa-1 αI-domains both share a hydrophobic surface underneath the dynamic α7 helix, ligands have only previously been discovered for Lfa-1. Our studies demonstrate that this difference in ligandability is no accident: the two family members differ in their ability to be covalently labeled by disulfide containing fragments when homologous residues are mutated to cysteine. Since the chemistry of the two sites are quite similar, we hypothesized that the difference in accessibility was responsible for the differences in ligandability. Despite the broadly similar allosteric mechanism, involving helix 7 conformational change, our results point to a difference in the breadth of the conformational ensemble as the primary determinant of ligandability. Our data are consistent with previous studies that suggest that the Lfa-1 I-domain samples a continuum of α7 conformations in its unliganded state. Our model (**Figure 3.9**) also indicates that Mac-1 samples more defined two state landscape, that is probably more iso-energetic between the extremes, with a higher energetic penalty for the excursions that form the allosteric pocket.

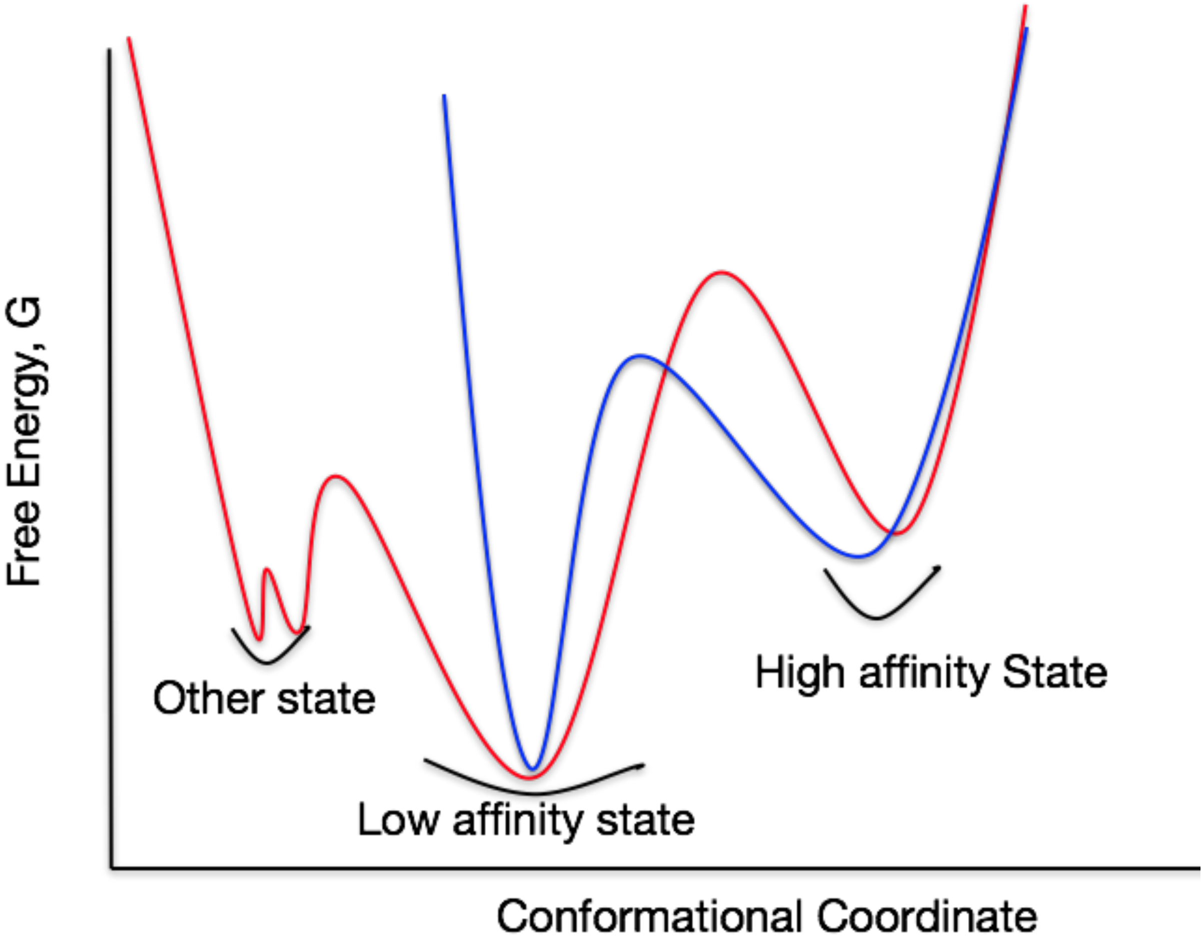


Figure 3.9. Proposed energy landscapes of Lfa-1(Red) and Mac-1(blue) I-domains. Lfa-1 samples multiple conformations in the ligand unbound state while Mac-1 mostly samples around the low affinity state and adopts the high affinity conformation when the energy penalty is paid by introducing mutations or truncations.

Our efforts to create a more ligandable Mac-1 α7 helix by importing an equivalent Lfa-1 sequence did not succeed. We were able to destabilize the protein and shift the breadth of the conformational ensemble of helix 7. This result as well as biophysical data from ligand bound Lfa-1 indicated that ligand binding is not just a product of flexible α7, but also the ability to distort the MIDAS. The accessibility of specific conformations that are ligandable is needed to generate high hit rates in assays that directly monitor binding, such as the covalent tethering experiment performed here.

In pharmaceutical settings, we care more about generating a single high quality hit than a high hit rate overall. Finding these needles in the haystack is more difficult when the conformational ensemble is narrow (as in Mac-1). For Mac-1, the ligand with the best results across multiple screens does engage the allosteric site. This result demonstrates that the energetic penalty for opening the pocket can be overcome by specific ligands with the added energetic benefit of increased local concentration through covalent tethering. Two important caveats merit mentioning here: 1) a covalent hit tethering hit can be quite difficult to advance to a free standing molecule with sufficient potency, 2) the energy landscape of the α7 helix that controls accessibility of this site may be quite different in the context of the full protein.

Despite these caveats, our study demonstrates a link between the breadth of the conformational ensemble and the ability to identify ligands that bind at cryptic allosteric sites. In the future, some of the costs for exposing these sites can be paid by mutations that would increase the probability of identifying hits. Although this idea, combined with covalent tethering, was unsuccessful here, it may be possible to apply this strategy to enable targeting of difficult sites on otherwise inaccessible targets.

Methods

Protein expression and purification

Wild-type Lfa-1 I-domain (E124-D316) was expressed in BL21(DE3) cells from a pET15b vector with an N-terminal hexahistidine (His6)- tag followed by a TEV protease cleavage site. Cells were grown in LB media(for crystallography) or M9 media(for NMR) and expression was induced with 1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 4hrs at 37°C.

For co-crystallization experiments, we expressed a smaller construct of wild-type Lfa-1 I-domain(S125-309I) in Rosetta 2(DE3) cells using a pET15b vector containing an N-terminal hexahistidine (His6)- tag followed by a TEV protease cleavage site. The cells were grown in LB or in M9 minimal media containing 1 g/mL of 15N NH4Cl(Sigma), until an OD600 of 0.5 -0.7 at 37°C . Protein expression was induced using 200 uM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 18°C overnight. The cells were then pelleted and flash frozen in liquid N2 .

For crystallization and NMR experiments, Lfa-1 I-domain was purified from inclusion bodies as previously described with some modifications[16]. The pellet of cells were dissolved in unfolding buffer containing 6 M Gdn-HCl, 20 mM Phosphate buffer, 10 mM Imidazole, 500 mM NaCl, and 1 mM TCEP. The cells were lysed using a sonicator and the inclusion bodies were solubilized for 2hrs at room-temp. The cell debris was pelleted and HisTrap HP column (GE Healthcare) was used to affinity purify the I-domain from the soluble fraction. The protein was refolded using the rapid dilution method previously described in a buffer with 50 mM Tris pH 8.5, 1 mM MgSO4, 1mM

TCEP and 5% (v/v)glycerol [16]. The refolded I-domain was then concentrated using a HisTrap HP column. We then added TEV protease to cleave the His6 tag and dialyzed the sample overnight at 4°C into the crystallization buffer(50 mM Tris pH 8.5, 1 mM MgSO4 and 1mM TCEP). We removed the cleaved tag and uncleaved products by running the sample through an additional HisTrap step. The untagged flow through was loaded onto Superdex 75 gel filtration column equilibrated with crystallization buffer for a final purification step. For NMR experiments, we buffer exchanged into an NMR buffer (10 mM Sodium phosphate pH 7.2, 10 mM $MgSO_4$, 150 mM NaCl and 0.05 % NaN_3) and added 10% D_2O.

For co-crystallization experiments, the pelleted Rosetta-2 cells were dissolved in a buffer containing 50 mM Tris-HCl pH 7.4, 3 mM Imidazole, 500 mM NaCl, 10 mM MgSO4, 1 mM TCEP, and 5% Glycerol in the presence of cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail(Roche). The cells were lysed using sonication and the soluble fraction was loaded onto a HisTrap HP column (GE Healthcare) for affinity purification. The eluted fractions were dialyzed overnight into a buffer containing 50 mM Tris pH 8.5, 10 mM MgSO4, 1 mM TCEP and 5% (v/v)glycerol while removing the His-tag from the protein using TEV protease. The untagged protein was ran through a Superdex 75 gel filtration column using the co-crystallization buffer(50 mM Tris pH 8.5, 10 mM MgSO4 and 1 mM TCEP).

The Mac-1 I-domain sequence coding for residues 146-342 was codon-optimized for expression in Escherichia coli by DNA 2.0 and subcloned into a pET15b plasmid containing a 6xHis affinity tag followed by a TEV protease cleavage site at the N-terminus. Point mutations were made via Megawhop PCR20 or QuikChange Site Directed Mutagenesis Kit (Agilent). All constructs were verified by DNA sequencing.

Recombinant Mac-1 I-domain was essentially as described [18] without an ion exchange step. Rosetta 2(DE3) Escherichia coli were grown in 2XYT (for crystallography and tethering) or M9 (for NMR) at 37 °C until OD600 reached 0.3. The temperature was reduced to 25 °C, and at OD600 = 0.5-0.7, expression was induced with 0.25 mM isopropylβ-D-1-thiogalactopyranoside (IPTG) followed by culture for 16-20 hours. Cells were harvested by centrifugation; resuspended in 50 mL buffer per L of culture of 50 mM HEPES (pH 7.5), 500 mM NaCl, 10 mM MgCl, 0.25 mM tris(2-carboxyethyl)phosphine (TCEP), 10 mM imidazole, and 5% w/v glycerol. Cells were lysed by sonication while on ice, and the insoluble lysate fraction was removed by centrifugation. The soluble lysate fraction was collected and incubated with HisPur Cobalt resin (Thermo) for 1-2 hours at 4°C, washed, and eluted by gravity flow in lysis buffer containing 150 mM imidazole. To remove the 6xHis affinity tag, purified protein was incubated and dialyzed overnight at 4 °C with 0.5 mg recombinant Tobacco etch virus (TEV) protease, engineered to contain its own 6xHis affinity tag. Cleavage/dialysis buffer was 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl, 0.25 mM TCEP, and 5% w/v glycerol. TEV protease, contaminants, and uncleaved protein were removed by repass over a HisPur Cobalt resin column equilibrated in lysis buffer. Cleaved and repassed I-domain was further purified by size exclusion chromatography (SEC) on a Superdex 75 16/600 column (GE Healthcare). For disulfide tethering mutations, SEC was done with 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl, and 5% w/v glycerol. For crystallography, SEC was done with 20 mM HEPES (pH 7.5), 250 mM NaCl, 10 mM MgCl. For NMR, SEC was done with 10 mM HEPES, 150 mM NaCl, 10 mM MgCl, 0.05% (w/v) NaN3 and 5% D2O to avoid further buffer exchange. Protein purity of SEC fractions was examined via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and fractions representing >95% purity were combined and concentrated to >5 mg/mL. Preparations containing 5% glycerol were flash frozen in LN2 and stored at -80 °C. Preparations without 5% glycerol were stored at 4°C. WT protein identity and cysteine mutation presence were confirmed by intact protein (LC/MS) on a Xevo G2-S (Waters). Yield of pure protein varied across constructs, but was 4.2 mg/L for codon-optimized WT Mac-1 I-domain.

Crystallization

Lfa-1 crystals were obtained via hanging drop vapor diffusion method by mixing 1µL of 20-22 mg/ml of protein and 1µL reservoir solution containing 30% (w/v) PEG 3350 and 0.1 M potassium phosphate dibasic. Crystals were formed in 2-5 days at 295K.

Lfa-1 co-crystals trays were set up using 22.5 mg/mL(1.0 mM) of protein in 50 mM Tris pH 8.5, 10 mM MgSO4, 1 mM TCEP and 1 mM BMS-68852 (KeyOrganics) or 8 mM Lovastatin (Cayman Chemical). A hanging drop vapor diffusion method was used to obtain the crystals. The Lfa-1_BMS-688521 co-crystal formed by mixing 0.1 µL of the reservoir (1 M LiCl, 0.1 M Na_3Cit pH 4 and 30%(w/v) PEG 6000) and 0.1 µL the protein-inhibitor complex solutions. BMS-688521 binds Lfa-1 with about 2.5 nM affinity while Lovastatin binds to the I-domain with about 3.8 µM affinity [8,19]. The Lfa-1_lovastatin co-crystal grew in a solution containing 0.1 µL of the well solution (0.2 M NH_4NO_3 and

20%(w/v) PEG 3350) and 0.1 µL of the protein-inhibitor complex solution. In both inhibitor bound conditions, the crystals grew within 14-21 days at 293 K.

Mac-1 crystallization was induced by hanging drop vapor diffusion at room temperature. 2 µL of 10 mg/mL Mac1 I-domain was mixed with an equal volume of well solution: 10 mM $MgCl_2$, 1.3 M $(NH_4)_2SO_4$ and 1.6 M NaCl. Crystals formed overnight and grew to full size in 4-5 days.

X-ray data collection, processing and modeling

Data sets were collected at the Advanced Light Source (ALS) beamline 8.3.1 on crystals maintained at 273K (RT) or 100K (cryo). The Lfa-1 I-domain diffraction data was processed and scaled with HKL2000 [20]. The co-crystal structures were processed using DIALS [21]. Space group determination and scaling were done using POINTLESS, AIMLESS and CTRUNCATE in CCP4 [22–25].

The structure was solved by molecular replacement using the program Phaser in PHENIX [26]. The structure of Wild-type LFA-1 I-domain (PDB 3F74; residues G128-Y307) was used as a search model. Six N-terminal residues and the α7 –helix were manually rebuilt using Coot [27] and the model was refined using PHENIX [28]. For the co-crystal structures we used the new cryogenic Lfa-1 structure as a search model and used the same pipeline to solve the structures.

X-ray structural heterogeneity modeling

Multi-conformer models were built using the electron density sampling algorithm qFit 2.0 [29,30]. qFit first builds in up to 4 different conformations as well as occupancies per residue into the electron density and then connects the conformations to give a multi-conformers model. We did a grid search over backbone amplitude ranges 0.05 to 0.35 and sigma value ranges 0.01-0.05 and selected the multi-conformer model with the lowest Rfree value.

The CONTACT algorithm was used to discover network of residues undergoing collective conformational exchange [31]. The algorithms identifies network of residues that can switch between alternate conformations to relieve clashes caused by overlapping van der Waals radii(vdW). A threshold value for clashes (Tstress) was set to the worst 25% of vdW overlaps and the maximum vdW overlap allowed after relief was 10%. We used contactApp which uses Cytoscape and Chimera to visualize the networks (generously supplied by Scooter Morris) [32,33]. We made the CONTACT network figures using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

Ensembles of conformations were built using phenix.ensemble_refinement [34]. We did a grid search over pTLS(0.1-1), wXRAY(2.5-10) and tX(0.5-1.5) refinement parameters to obtain optimal values. We selected the ensemble with the lowest Rfree value. Root mean square fluctuation (RMSF) were calculated using the CA atoms of each residue in the ensemble. The ensemble structures are colored based on RMSF values using the Spectrumany function in PyMOL.

PDB ensemble generation

The Lfa-1 PDB ensemble was created using PDB ID and CHAIN: 5E6S AND CHAIN A, 5E6S AND CHAIN C, 5E6S AND CHAIN E, 5E6U AND CHAIN A, 5E6R AND CHAIN A, 1RD4 AND CHAIN A, 1RD4 AND CHAIN B, 1RD4 AND CHAIN C, 1RD4 AND CHAIN D,

1ZOP AND CHAIN A, 1ZOP AND CHAIN B, 1ZOO AND CHAIN A, 1ZOO AND CHAIN B, 1ZON AND CHAIN A, 1LFA AND CHAIN A, 1LFA AND CHAIN B, 4IXD AND CHAIN A, 1XUO AND CHAIN A, 1XUO AND CHAIN B, 1XDG AND CHAIN A, 1XDG AND CHAIN B, 1XDD AND CHAIN A, 1XDD AND CHAIN B, 3E2M AND CHAIN A, 3E2M AND CHAIN B, 3EOB AND CHAIN I, 3EOB AND CHAIN J, 3EOA AND CHAIN I, 3EOA AND CHAIN J, 3BQN AND CHAIN B, 3BQN AND CHAIN C, 3BQM AND CHAIN B, 3BQM AND CHAIN C, 1CQP AND CHAIN A, 1CQP AND CHAIN B, 3M6F AND CHAIN A, 2O7N AND CHAIN A, 2ICA AND CHAIN A, 3F78 AND CHAIN A, 3F78 AND CHAIN B, 3F78 AND CHAIN C, 3F74 AND CHAIN A, 3F74 AND CHAIN B, 3F74 AND CHAIN C, 1ZON CHAIN A, and 1DGQ AND CHAIN A(the first chain from the NMR ensemble). We aligned these structures to the cryogenic Lfa-1 structure we have solved using residue ranges 120-310 and calculated root mean square deviation(RMSD) values using the CA atoms of each residue. The ligand stabilized high affinity structure with a native α7–helix(3TCX) was not used in the RMSD calculations but is shown on the image of the ensemble(pink). The PDB ensembles are colored by RMSD values.The ensemble included all the structures currently (Oct. 2017) found in the PDB except for the disulfide bond stabilized structures and a structure that had the α7–helix pointing the opposite direction compared to other structures in the PDB. We didn't include the disulfide bond stabilized structures because we wanted to measure the conformational heterogeneity of Lfa-1 close to a native condition.

The PDB ensemble for Mac-1 I-domain was generated using PDB ID and CHAIN: 1BHO AND CHAIN 1, 1BHO AND CHAIN 2, 1BHQ AND CHAIN 1, 1BHQ AND CHAIN 2,

1IDN AND CHAIN 1, 1IDN AND CHAIN 2, 1IDO AND CHAIN A, 1JLM AND CHAIN A, 1M1U AND CHAIN A, 1MF7 AND CHAIN A, 1N9Z AND CHAIN A, 1NA5 AND CHAIN A, 3Q3G AND CHAIN E, 3Q3G AND CHAIN G, 3Q3G AND CHAIN I, 3Q3G AND CHAIN L, 3QA3 AND CHAIN E, 3QA3 AND CHAIN G, 3QA3 AND CHAIN I, 3QA3 AND CHAIN L, 4M76 AND CHAIN B, and 4XW2 AND CHAIN A. We used residue ranges 130-319. We split the ensembles to open and closed state ensembles. The open state is defined as structures with an overall CA RMSD values greater than 1 Å and included PDB ID and CHAIN: 1N9Z_A 4M76_B 1M1U_A 1IDO_A and 4XW2_A. This high RMSD deviation is a result of the presence of activating point mutations or truncations. We aligned and calculated RMSD values of these structures as described for the Lfa-1 PDB ensembles.

NMR data collection

All Lfa-1 NMR experiments were conducted on a Bruker 800 MHz spectrometer at 300 K. We conducted a 2D ^{1}H-^{15}N HSQC TROSY experiments as implemented in trosyetf3gpsi pulse sequence; avance-version (12/01/11) [35–40].For Lfa-1 BMS-688521 titration experiments, 413 μM of uniformly ^{15}N-labeled protein in the NMR buffer(10 mM Sodium phosphate pH 7.2, 10 mM MgSO4, 150 mM NaCl and 0.05 % NaN3) was used to collect spectra at 0 μM, 10 μM, 50 μM, 100 μM, 150 μM and 500 μM concentrations of BMS-68852. For the Lovastatin experiments, we titrated in 0 μM, 25 μM, 125 μM, 200 μM and 1000 μM of lovastatin into 187 μM of the ^{15}N-labeled protein in the NMR buffer.

Mac1 NMR samples were purified in or buffer exchanged into 5% ^2H_2O, 10mM phosphate buffer pH 7.2 with 150mM NaCl, 10mM $MgSO_4$ and 0.05% w/v NaN_3. Several samples were used to acquire backbone amide resonance assignments, a uniformly $^{13}C/^{15}N/^2H$-labeled sample at 1.3 mM/0.5 mL and a uniformly $^{13}C/^{15}N$-labeled and ~70% ^2H-labeled sample at 0.7 mM/0.5 mL. A suite of three-dimensional triple-resonance experiments were collected on each: HNCA, HNcoCA and ^{15}N-NOESY-HSQC on the 70% ^2H sample, and HNCO, HNcoCACB and HNCACB on the U-$^{13}C/^{15}N/^2H$-labeled sample. Additionally, $^{15}N/^1H$ HSQC spectra were collected on five additional samples containing either ^{15}N-glycine, leucine, alanine, isoleucine or valine to achieve single-residue $^{15}N/^1H$-labeling. A Transverse relaxation optimized spectroscopy (TROSY) experiment was used to deconvolute overlapped sidechain resonances.

All data were collected at 298K on Bruker Avance 500 MHz and 800 MHz instruments equipped with cryogenic probes. Data were processed and analyzed in NMRPipe [41], CCPNMR [42]and SPARKY (Goddard TD, Kneller DG. Sparky. Vol. 3. University of California; San Francisco: 1999).

NMR spectral details and processing parameters

The spectral offsets (widths) for the HNCA, HNcoCA, and: ^1H, 4.701 ppm (20 ppm); ^{13}C, 54 ppm (30 ppm), ^{15}N, 116.3 ppm (32 ppm), using 64 dummy scans with 1024 ^1H, 56 ^{13}C and 156 ^{15}N complex data points collected in Echo-AntiEcho mode for the ^{13}C-dimension, States-TPPI mode for the ^{15}N dimension.

The spectral offsets (widths) for the HNCO, HNcoCACB, and HNCACB were: 4.750 ppm (14 ppm); ^{13}C, 39 ppm (36 ppm), ^{15}N, 119 ppm (75 ppm), using 32 dummy scans and 1024 ^{1}H, 98 ^{13}C and 98 ^{15}N complex data points collected in Echo-AntiEcho mode for the ^{13}C-dimension, States-TPPI mode for the ^{15}N dimension with ^{2}H coupling applied. The total acquisition time for the HNCO was 21.5 hrs, HNcoCA was 38.5 hrs, HNcoCACB was 39 hrs.

NMR spectral analysis for titration experiments

The Lfa-1 NMR titration spectra were processed using NMRPipe [41]. We automatically processed the data series using autoProc.tcl script and used nmrDraw to find peaks from the series. The backbone chemical shift assignments for Lfa-1 I-domain were obtained using both BMRB entry 18941 and 4553 [43,44]. We transferred the assignments to the titration series using ipap.tcl script in NMRpipe. To export the chemical shift values for our series, we used view2D.tcl. The chemical shift perturbation (CSP) values were calculated using modelTitr.tcl script with coefficient values of 1.0 and 0.2 for HN and 15N respectively (CSP = Sqrt[1.0*(X shift)^2 + 0.2*(Y shift)^2]). CSP values are displayed in the figures using a putty representation and a color gradient from grey to crimson/ruby. A wider tube and a higher intensity color indicate high CSP values. Full and per peak spectral images were made using CcpNmr Analysis using a macro generously provided by Wayne Boucher [42] .

NMR volume calculations

Peak volumes were calculated using nmrDraw. Volumes are calculated by averaging the intensities from multiple points around the peak volume [41]. In the figures, the calculated volumes are shown as bar graphs with a rolling average value shown as green line graphs.

Tethering experiments

I-domain constructs containing target cysteines were diluted to 500 nM in 20 mM Tris (pH 8.0) containing 500 µM β-mecaptoethanol (βME). 25 µL of the reaction mixture was dispensed into 384-well low-volume V-shape plate (Grenier Bio). Using a Biomek FX (Beckman), 50 nL of tethering fragment was pinned into each well from a 384-well source plate containing 50 mM fragments in DMSO. The reaction mixture was incubated for 3 hours at room temperature, and covalent adduct formation was detected via intact protein LC/MS as previously described [14].

Acknowledgements

R.A.W. and K.K.H were supported by NSF Graduate Research Fellowships. J.S.F. is supported by NSF STC-1231306, NIH GM110580, a Searle Scholar Award from the Kinship Foundation, a Pew Scholar Award from the Pew Charitable Trusts, a Packard Fellowship from the David and Lucile Packard Foundation. We would like to thank Daniel Keedy, Lin Liu, Ben Barad, Justin Biel, Greg Lee, Chris Waddling, Frank Delaglio and all Fraser lab members for helpful comments.

References

1. Motlagh HN, Wrabl JO, Li J, Hilser VJ: **The ensemble nature of allostery**. *Nature* 2014, **508**:331–339.

2. Raman AS, White KI, Ranganathan R: **Origins of Allostery and Evolvability in Proteins: A Case Study**. *Cell* 2016, **166**:468–480.

3. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, et al.: **Structure of a nanobody-stabilized active state of the β2 adrenoceptor**. *Nature* 2011, **469**:175–180.

4. Changeux J-P, Christopoulos A: **Allosteric Modulation as a Unifying Mechanism for Receptor Function and Regulation**. *Cell* 2016, **166**:1084–1102.

5. Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J: **Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms**. *Mol Cell* 2011, **42**:9–22.

6. Shimaoka M, Xiao T, Liu J-H, Yang Y, Dong Y, Jun C-D, McCormack A, Zhang R, Joachimiak A, Takagi J, et al.: **Structures of the αL I Domain and Its Complex with ICAM-1 Reveal a Shape-Shifting Pathway for Integrin Regulation**. *Cell* 2003, **112**:99–111.

7. Shimaoka M, Springer TA: **Therapeutic antagonists and conformational regulation of integrin function**. *Nat Rev Drug Discov* 2003, **2**:703–716.

8. Watterson SH, Xiao Z, Dodd DS, Tortolani DR, Vaccaro W, Potin D, Launay M, Stetsko DK, Skala S, Davis PM, et al.: **Small molecule antagonist of leukocyte function associated antigen-1 (LFA-1): structure-activity relationships leading to the identification of 6-((5S,9R)-9-(4-cyanophenyl)-3-(3,5-dichlorophenyl)-1-methyl-2,4-dioxo-1,3,7-triazaspiro[4.4]nonan-7-yl)nicotinic acid (BMS-688521)**. *J Med Chem* 2010, **53**:3814–3830.

9. Ley K, Rivera-Nieves J, Sandborn WJ, Shattil S: **Integrin-based therapeutics: biological basis, clinical use and new drugs**. *Nat Rev Drug Discov* 2016, **15**:173–183.

10. Ryu JK, Petersen MA, Murray SG, Baeten KM, Meyer-Franke A, Chan JP, Vagena E, Bedard C, Machado MR, Rios Coronado PE, et al.: **Blood coagulation protein fibrinogen promotes autoimmunity and demyelination via chemokine release and antigen presentation**. *Nat Commun* 2015, **6**:8164.

11. Jensen MR, Bajic G, Zhang X, Laustsen AK, Koldsø H, Skeby KK, Schiøtt B, Andersen GR, Vorup-Jensen T: **Structural Basis for Simvastatin Competitive Antagonism of Complement Receptor 3**. *J Biol Chem* 2016, **291**:16963–

16976.

12. Kallen J, Welzenbach K, Ramage P, Geyl D, Kriwacki R, Legge G, Cottens S, Weitz-Schmidt G, Hommel U: **Structural basis for LFA-1 inhibition upon lovastatin binding to the CD11a I-domain**. *J Mol Biol* 1999, **292**:1–9.

13. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, Wells JA: **Site-directed ligand discovery**. *Proc Natl Acad Sci U S A* 2000, **97**:9367–9372.

14. Hallenbeck KK, Davies JL, Merron C, Ogden P, Sijbesma E, Ottmann C, Renslo AR, Wilson C, Arkin MR: **A Liquid Chromatography/Mass Spectrometry Method for Screening Disulfide Tethering Fragments**. *SLAS Discov* 2017,

15. Kukic P, Alvin Leung HT, Bemporad F, Aprile FA, Kumita JR, De Simone A, Camilloni C, Vendruscolo M: **Structure and dynamics of the integrin LFA-1 I-domain in the inactive state underlie its inside-out/outside-in signaling and allosteric mechanisms**. *Structure* 2015, **23**:745–753.

16. Legge GB, Kriwacki RW, Chung J, Hommel U, Ramage P, Case DA, Dyson HJ, Wright PE: **NMR solution structure of the inserted domain of human leukocyte function associated antigen-1**. *J Mol Biol* 2000, **295**:1251–1264.

17. Surade S, Blundell TL: **Structural biology and drug discovery of difficult targets: the limits of ligandability**. *Chem Biol* 2012, **19**:42–50.

18. Bajic G, Yatime L, Sim RB, Vorup-Jensen T, Andersen GR: **Structural insight on the recognition of surface-bound opsonins by the integrin I domain of complement receptor 3**. *Proc Natl Acad Sci U S A* 2013, **110**:16426–16431.

19. Ullrich T, Baumann K, Welzenbach K, Schmutz S, Camenisch G, Meingassner JG, Weitz-Schmidt G: **Statin-derived 1,3-oxazinan-2-ones as submicromolar inhibitors of LFA-1/ICAM-1 interaction: stabilization of the metabolically labile vanillyl side chain**. *Bioorg Med Chem Lett* 2004, **14**:2483–2487.

20. Otwinowski Z, Minor W: **Processing of X-ray diffraction data collected in oscillation mode**. *Methods Enzymol* 1997, **276**:307–326.

21. **[No title]**. [date unknown],

22. Evans P: **Scaling and assessment of data quality**. *Acta Crystallogr D Biol Crystallogr* 2005, **62**:72–82.

23. Evans PR: **An introduction to data reduction: space-group determination, scaling and intensity statistics**. *Acta Crystallogr D Biol Crystallogr* 2011, **67**:282–292.

24. Evans PR, Murshudov GN: **How good are my data and what is the**

resolution? *Acta Crystallogr D Biol Crystallogr* 2013, **69**:1204–1214.

25. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, et al.: **Overview of theCCP4 suite and current developments**. *Acta Crystallogr D Biol Crystallogr* 2011, **67**:235–242.

26. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive Python-based system for macromolecular structure solution**. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:213–221.

27. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics**. *Acta Crystallogr D Biol Crystallogr* 2004, **60**:2126–2132.

28. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive Python-based system for macromolecular structure solution**. In *International Tables for Crystallography*. . 2012:539–547.

29. van den Bedem H, Dhanik A, Latombe JC, Deacon AM: **Modeling discrete heterogeneity in X-ray diffraction data by fitting multi-conformers**. *Acta Crystallogr D Biol Crystallogr* 2009, **65**:1107–1117.

30. Keedy DA, Fraser JS, van den Bedem H: **Exposing Hidden Alternative Backbone Conformations in X-ray Crystallography Using qFit**. *PLoS Comput Biol* 2015, **11**:e1004507.

31. van den Bedem H, Bhabha G, Yang K, Wright PE, Fraser JS: **Automated identification of functional dynamic contact networks from X-ray crystallography**. *Nat Methods* 2013, **10**:896–902.

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res* 2003, **13**:2498–2504.

33. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera--a visualization system for exploratory research and analysis**. *J Comput Chem* 2004, **25**:1605–1612.

34. Burnley BT, Afonine PV, Adams PD, Gros P: **Modelling dynamics in protein crystal structures by ensemble refinement**. *Elife* 2012, **1**:e00311.

35. Czisch M, Boelens R: **Sensitivity enhancement in the TROSY experiment**. *J Magn Reson* 1998, **134**:158–160.

36. Pervushin KV, Wider G, Wüthrich K: **Single Transition-to-single Transition Polarization Transfer (ST2-PT) in [15N,1H]-TROSY**. *J Biomol NMR* 1998,

12:345–348.

37. Meissner A, Schulte-Herbrüggen T, Briand J, Sørensen OW: **Double spin-state-selective coherence transfer. Application for two-dimensional selection of multiplet components with long transverse relaxation times.** *Mol Phys* 1998, **95**:1137–1142.

38. Weigelt J: **Single Scan, Sensitivity- and Gradient-Enhanced TROSY for Multidimensional NMR ExperimentsJ. Am. Chem. Soc.1998,120, 10778–10779**. *J Am Chem Soc* 1998, **120**:12706–12706.

39. Rance M, Loria JP, Palmer AG3rd: **Sensitivity improvement of transverse relaxation-optimized spectroscopy**. *J Magn Reson* 1999, **136**:92–101.

40. Zhu G, Kong XM, Sze KH: **Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments.** *J Biomol NMR* 1999, **13**:77–81.

41. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A: **NMRPipe: a multidimensional spectral processing system based on UNIX pipes**. *J Biomol NMR* 1995, **6**:277–293.

42. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED: **The CCPN data model for NMR spectroscopy: development of a software pipeline**. *Proteins* 2005, **59**:687–696.

43. Zimmerman T, Oyarzabal J, Sebastián ES, Majumdar S, Tejo BA, Siahaan TJ, Blanco FJ: **ICAM-1 peptide inhibitors of T-cell adhesion bind to the allosteric site of LFA-1. An NMR characterization.** *Chem Biol Drug Des* 2007, **70**:347–353.

44. Kriwacki RW, Legge GB, Hommel U, Ramage P, Chung J, Tennant LL, Wright PE, Dyson HJ: **Assignment of 1H, 13C and 15N resonances of the I-domain of human leukocyte function associated antigen-1**. *J Biomol NMR* 2000, **16**:271–272.

Chapter 4

Using XFELs to visualize solvent in the influenza A M2 Proton Channel

Citations:

Woldeyes RA, Thomaston JL, Nakane T, Brewster AS, Uervirojnangkoorn M, Barad BA, Tanaka T, Yamashita A, Koiwai K, Chen Y, Arima T, Kobayashi J, Masuda T, Suzuki M, Nureki O, Tono K, Joti Y, Sugahara M, Yumoto F, Sauter NK, Tanaka R, Nango E, Iwata S, DeGrado WF, Fraser JS. **Comparing XFEL data processing methods highlights sensitivity/robustness of water electron density interpretation within the flu M2 channel.** *In preparation*

Summary

The influenza A virus must acidify its interior to release its RNA. This important step in the viral life cycle relies on matrix protein 2 (M2), a pH-regulated proton channel embedded in the lipid envelope of the virus. M2 has been a subject of intense study both because it the target of adamantane compounds that can be used to treat flu and because it is an ideal model system to study proton conduction across membranes. The structural location of ordered water within the M2 channel plays a key role in both of these areas. Ordered water molecules are proposed to make key bridging interactions between adamantane compounds and the protein. Additionally, there is significant debate about the existence "water wires" of ordered solvent molecules that allow proton conduction by connecting the viral exterior and interior. It is therefore critical to determine the distribution of water molecules within the channel. Previous high resolution X-ray data of the M2 channel crystallized in the lipid sponge phase revealed interesting differences in the distributions of water as a function of pH. However, preliminary room temperature data called into question whether data collected from cryo-cooled crystals was informative for biologically relevant mechanisms of proton transport. Unfortunately, damage to the small M2 crystals from the radiation dose at room temperature precluded mechanistic interpretations of the solvent density. To circumvent these issues and to observe the functionally important structural dynamics of M2, we used X-ray Free Electron Lasers to collect radiation-damage free, 1.4 Å dataset of M2 at low, intermediate and high pHs. By avoiding the artifacts associated with cryo-cooling and eliminating the effects of radiation damage, the resulting room temperature data have allowed us to draw new conclusions about

the solvent conformational heterogeneity. We will build on this work and use XFEL to monitor changes in the solvation of the pore of M2 as a function of both pH and time. Furthermore, these structures will help in the design of next-generation of antivirals address continuing problems with drug resistance.

Introduction

Water plays essential structural and dynamic roles in biology [1-3]. However, we lack atomic resolution details of how the conformational dynamics of water molecules modulate biological processes [4]. This is impart due to limitations of conventional experimental techniques and computational techniques that attempt to meet this need are a work in progress. However, X-ray Free Electron Laser (XFEL) sources offer new and unique opportunity to look at these weak feature in atomic detail effectively without any radiation damage.

To explore the role water molecules play in proton conduction, we used XFELs to study the solvent in the influenza A M2 proton channel. The influenza A virus must acidify its interior to release its RNA (**Figure 4.1**) [5]. This acidification step relies on matrix protein 2 (M2) channel. M2 is a pH regulated proton channel embedded in the lipid envelope of the influenza A virus. Although it is known that acidification of the endosome induces this channel to conduct protons the mechanism by which protons are shuttled through the pore is not clear.

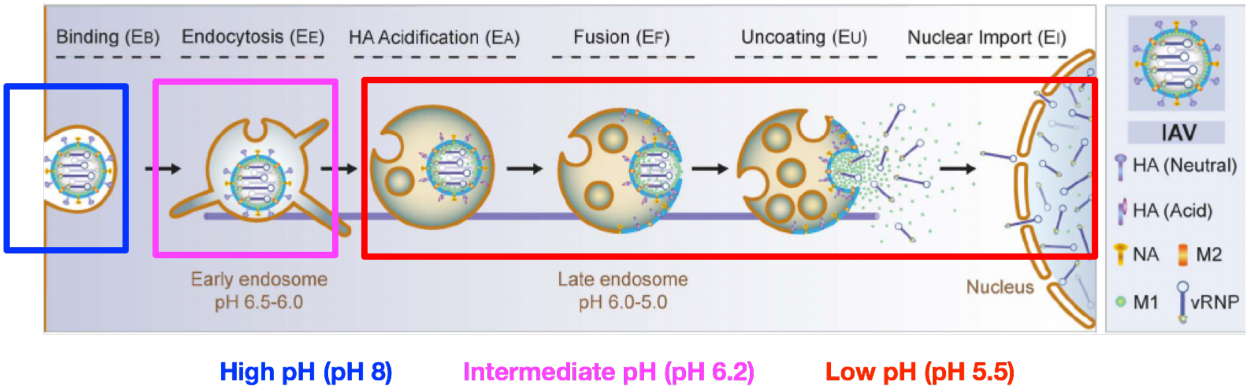


Figure 4.1: The influenza A virus must acidify its interior to release its RNA (Figure adapted from [5])

A hypothesis from simulations suggest that proton conduction across the membrane is mediated by ordered waters within the channel [6]. However, there is limited experimental evidence for the position and distribution of water molecules within the channel. Furthermore, currently there are two major competing hypothesis for the mechanism of proton conduction through the channel; the “water wire” and “proton shuttle” hypothesis. In the water wire hypothesis, protons hop through the channel only using water molecules lining the channel. The opening and closing of the channel is controlled by electrostatic repulsion caused by protonation of the selective filter, His37[6]. Conversely in the second hypothesis, protons need to be shuttled into the channel through protonation and deprotonation of His37 followed by a ring flip [6]. Therefore, the waters that are within a hydrogen bonding distance from His37 are important for the proton conduction. In an effort to differentiate between these hypothesis, we asked: What is the distribution of water molecules within the channel and how does the water distribution change as a function of pH?

Previous work has shown that highly diffracting (1.1 Å) crystals of M2 can be obtained in physiologically relevant lipid environment [7]. Using these crystals, the DeGrado lab solved cryogenic structures of the M2 channel at low (pH 5.5) and high pH (pH 8.0) conditions [7]. However, these structures were not able to explain why only the low pH condition supported proton conduction. It is possible that cryo-cooling of the crystals at liquid nitrogen temperature (100 K) trapped the waters in non-native states during the freezing process. This concept was previously visited by Fraser et al and others who have shown that this cryo-cooling process is too slow to preserve the room temperature conformational distribution [8,9]. Furthermore the buffers used during crystallization to maintain the pH are sensitive to temperature changes and therefore it is not clear what is the pH of the frozen protein crystal.

To make sure the waters in the channel exist at room temperature, we collected room temperature datasets. However during data collection, we saw radiation dose dependent fading of high resolution Bragg spots. The size of the crystals were less than 40 microns and therefore we were not able to translate over the crystal to minimize radiation damage. Using the first set of frames that were “undamaged”, we solved the structures albeit to the limited resolutions; 1.4 Å and 1.7 Å resolutions for low and high pH conditions respectively [7]. The room temperature datasets called into question the results from cryogenic datasets. Unfortunately, damage to the small M2 crystals precluded mechanistic interpretations of the solvent density.

To avoid artifacts associated with cryo-cooling and eliminate radiation damage, we used XFELs to visualize the solvent in the M2 channel. The femtosecond pulses of XFELs allow for “diffraction before destruction” where the X-rays move through the sample and give diffraction before radiation damage occurs and the crystals blow up [10,11]. We collected 1.4 Å datasets at low, intermediate and high pH conditions. To obtain complete datasets with high correlation coefficients, we performed metrology refinement of SPring-8 Angstrom Compact Free Electron Laser (SACLA) multi-port readout charge coupled device (MPCCD) data. The refinement allowed us to minimize noise in intensity measurement and increased indexing rates. Using the new XFEL structures as ground truth, we found that previous diffraction data collected at synchrotron sources either suffer from artifact associate with cryo-cooling or suffer from radiation damage. Furthermore, the new structures allowed us to draw new conclusions about the solvent in the M2 channel. At low pH the pore contains more water molecules that can facilitate proton conduction through the channel. Waters found within a hydrogen bonding distance of the selective filter, His37, stabilize the conducting state only at low pH conditions. The next step is to study the reorganization of the solvent as a function of pH and time using time resolved XFEL experiments.

Results

Joint crystal model and geometry refinement minimizes noise in intensity measurements and allow for increased indexing rates

To obtain effectively artifact and radiation damage free structures of the M2 channel, we collected three 1.4 Å resolution room temperature datasets using the XFEL source at SACLA [12]. The crystals were maintained in three pH conditions that correspond to pHs that M2 experiences at different stages of the viral life cycle, pH 8.0, 6.2 and 5.5. Images with diffraction spots were identified using CHEETAH [13]. After finding the hits, dials.stills_process was used to index measured intensities [14,15]

Overall, we had the highest indexing rate for the low and high pH conditions while the intermediate condition gave a much lower indexing rate. For the low pH condition (pH 5.5), we obtained 108,680 images crystal diffraction hits (**Table 4.1**). Of these hits, we were able to index 60,054 images with an overall indexing rate of 55.2% (**Table 4.1).** The indexing rate was limited by salt crystal formation during some of the runs (Figure 4.2A: red arrow) and images with multiple lattices (**Figure 4.2A**: green arrow). For the intermediate condition our indexing rate was down to 24.1% (**Table 4.1**). This low indexing rate is mainly due to the quality of our crystals and salt crystal formation during injection (**Figure 4.2B**: red arrow). In addition, we again saw images with multiple lattices. We were able to avoid salt crystal formation for the high pH condition (**Figure 4.2C**). We obtained an indexing rate of 46.9% (**Table 4.1**). The indexing rate was mainly affected by images with multiple lattices.

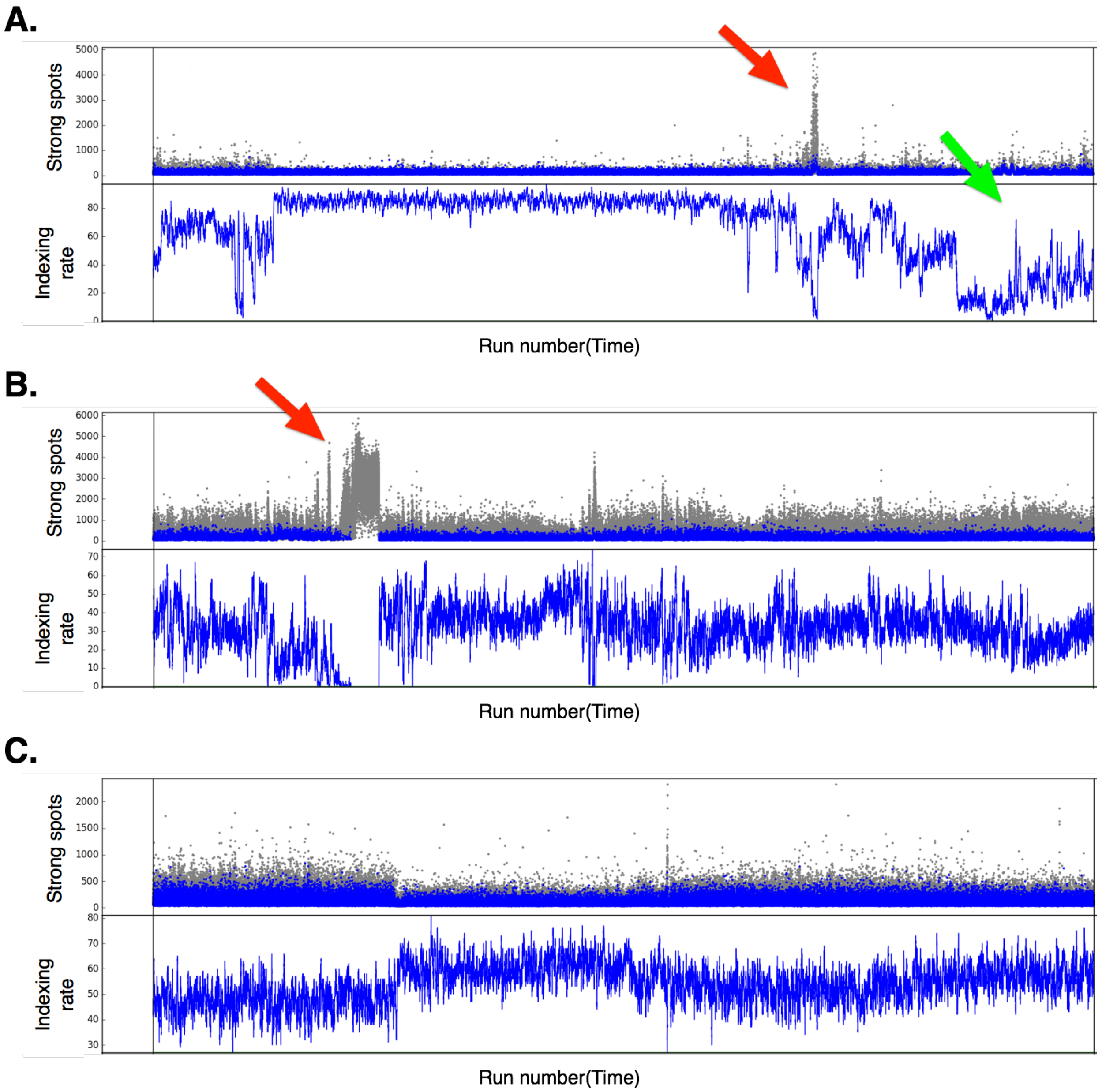


Figure 4.2: Summary of the data collection statistics for the low, intermediate and high pH datasets. The top rows display the number of strong spots per run number with the spots that were indexed shown in **blue**. High number of strong spots that didn't index (**grey**) indicate formation of salt crystals (**red arrows**). In the low pH **(A)** and intermediate pH **(B)** conditions, salt crystals formed during sample injection. We didn't see any salt crystal formation for the high pH condition **(C).** The second rows show changes in indexing rates as a function of run numbers. We saw high indexing rates for the low and high pH conditions. The indexing rate for the intermediate condition was significantly lower. As a result of multiple lattice formation(**green arrow**), we saw a significant drop in indexing rates for the low pH **(A)** condition towards the

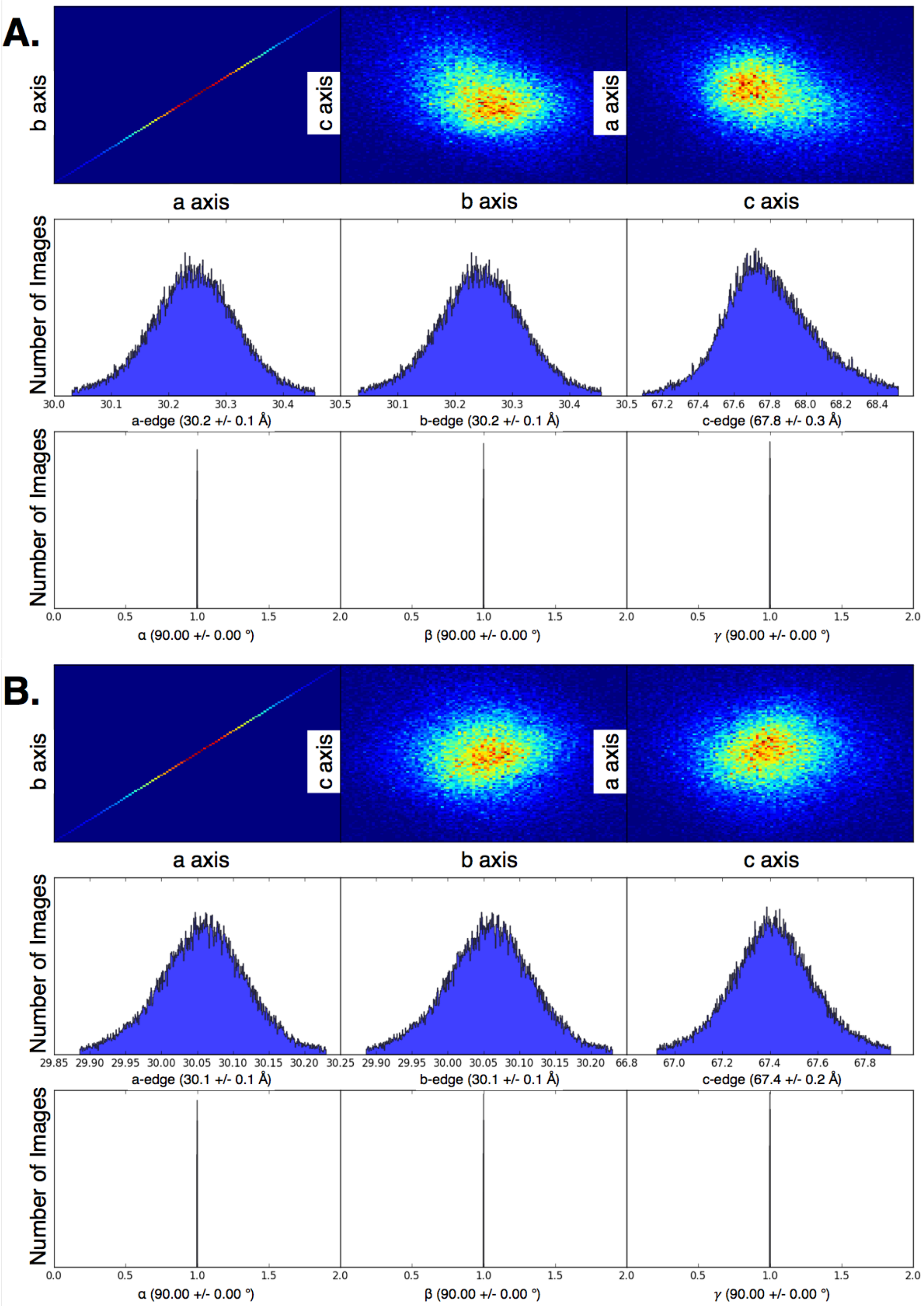
end of the run.

Condition	# Hits	# Integrated IMG initial	Indexing Rates Initial
Low pH	108,680	60,054	55.2%
Inter pH	285,657	68,765	24.1%
High pH	159,167	74,707	46.9%

Table 4.1: Summary of the data collection statistics. High indexing rates were obtained for the low and high pH conditions while the intermediate condition gave a much lower indexing rate.

To minimize differences between batches of crystals and variation in the geometry of the detector from run to run, we jointly refined the crystal models (unit cell and crystal orientation) with the geometry of the detector using DIALS [14,15]. We modified the geometry refinement algorithm designed for the LCLS's CSPAD detector to work for SACLA's MPCCD [14].The target function is the difference between observed and predicted spot locations. We refined detector rotation and tilt as a whole and as quadrants along with detector distance and filtered the images using RMSD.

The metrology refinement minimized noise in our intensity measurements and allowed for higher indexing rates for all of our datasets. For the low pH condition, the joint refinement fixed the skewed unit cell size distribution we initially saw (**Figure 4.3**). We also saw a slight change in the unit cell size and an improvement in deviation from the mean unit cell length after the refinement (**Figure 4.3B,C**). Using the refined geometry, we indexed additional 4000 images for the low pH condition (**Table 4.2**). In general, we see increased indexing rates with metrology refinement. For the intermediate and high pH conditions, we obtained at least 10,000 additional indexed images (**Table 4.2**).


A.
b axis
c axis
a axis
a axis
b axis
c axis
Number of Images
30.0
30.1
30.2
30.3
30.4
30.5
30.1
30.2
30.3
30.4
30.5
67.2
67.4
67.6
67.8
68.0
68.2
68.4
a-edge (30.2 +/- 0.1 Å)
b-edge (30.2 +/- 0.1 Å)
c-edge (67.8 +/- 0.3 Å)
Number of Images
0.0
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
α (90.00 +/- 0.00 °)
β (90.00 +/- 0.00 °)
γ (90.00 +/- 0.00 °)
B.
b axis
c axis
a axis
a axis
b axis
c axis
Number of Images
29.85
29.90
29.95
30.00
30.05
30.10
30.15
30.20
30.25
29.90
29.95
30.00
30.05
30.10
30.15
30.20
66.8
67.0
67.2
67.4
67.6
67.8
a-edge (30.1 +/- 0.1 Å)
b-edge (30.1 +/- 0.1 Å)
c-edge (67.4 +/- 0.2 Å)
Number of Images
0.0
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
0.5
1.0
1.5
2.0
α (90.00 +/- 0.00 °)
β (90.00 +/- 0.00 °)
γ (90.00 +/- 0.00 °)

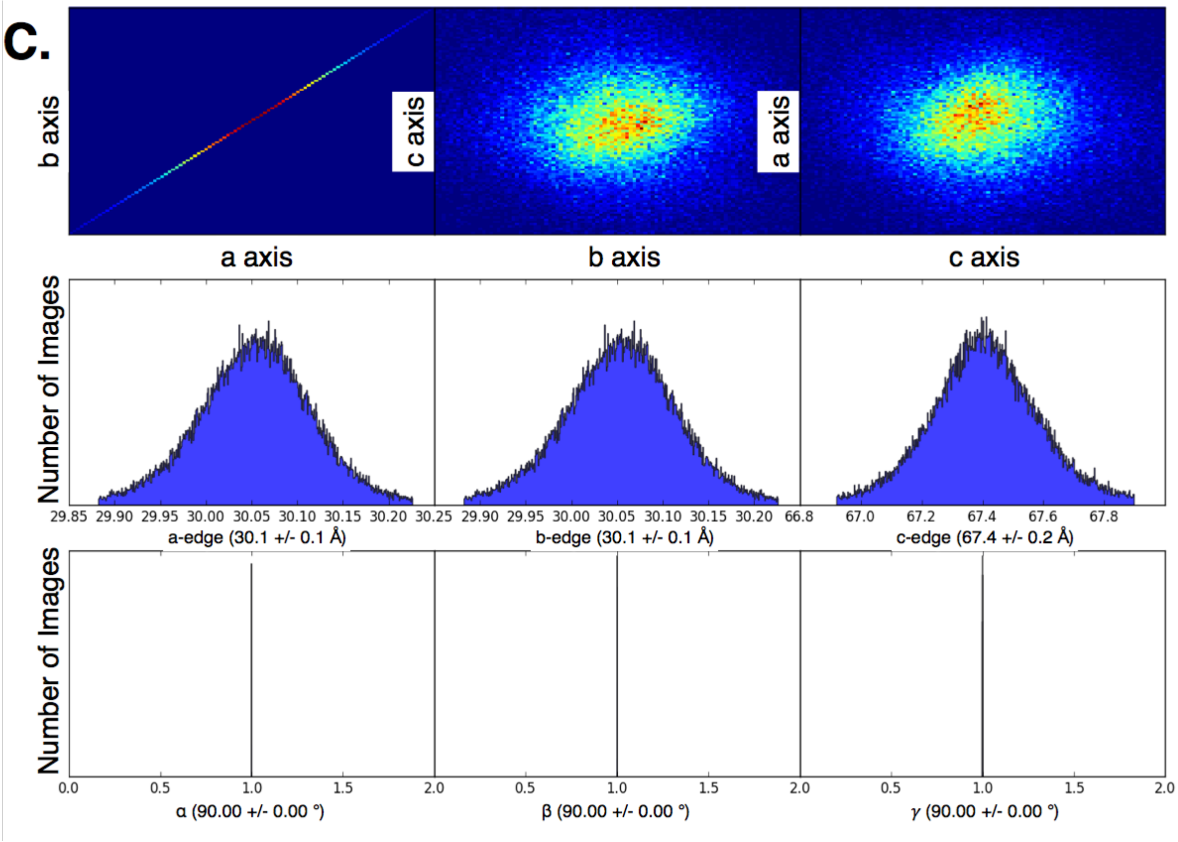


Figure 4.3: Joint crystal model and geometry refinement minimizes noise in intensity measurements. Before metrology refinement **(A)**, the unit cell size distribution of the crystals is skewed on the c-edge. After the first **(B)** and the second (**C)** rounds of metrology refinement, the skewness in the distribution is fixed and the unit cell parameters slightly change.

Condition	# Integrated IMG initial	# Integrated IMG Met_1	# Integrated IMG Met_2
Low pH	60,054	64,333	64,388
Inter pH	68,765	81,424	81,951
High pH	74,707	83,925	84,361

Table 4.2: Increased indexing rates with metrology refinement in all of the pH conditions.

Comparing the unit cell size distribution for our datasets, we saw pH dependent shift in the the unit cell size distributions. The low pH condition has the longest a and b edge mean length and the smallest c edge mean length (**Figure 4.4)**. Conversely, the high pH condition had the opposite trend **(Figure 4.4**). While the intermediate condition length distribution is in between (**Figure 4.4**). These size distributions indicate that the solvent in the pore of the channel is biggest for the low pH condition and requires the expansion of the unit cell size in a and b edges relative to the other conditions. The average unit cell size for the high pH condition is slimmer, but longer.

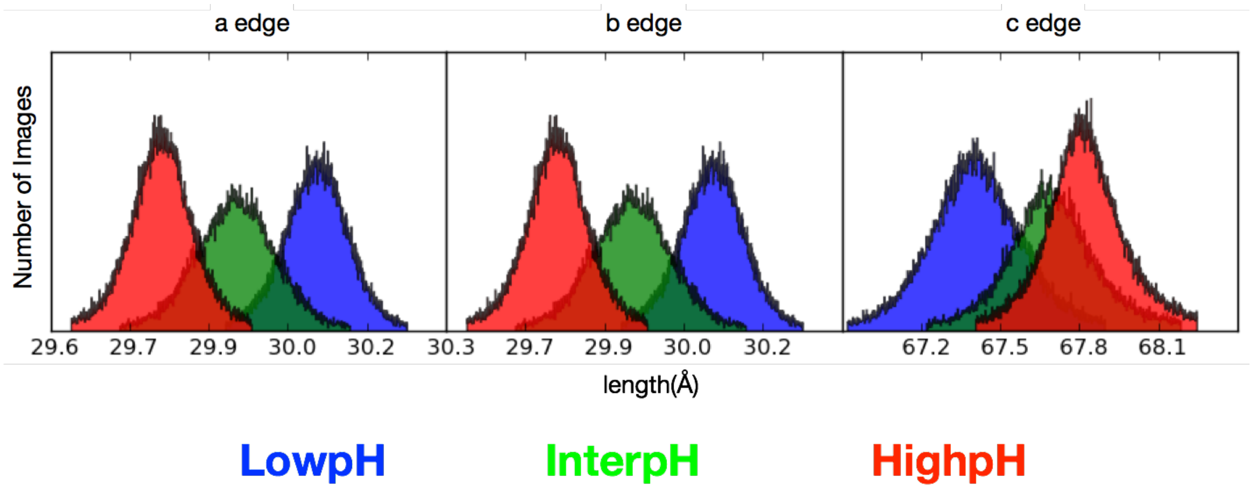


Figure 4.4: Changes in unit cell size distribution as a function of pH. The low pH **(blue)** condition has largest a and b edge and the smallest c edge lengths. The opposite is true for the high pH condition indicating that there are less number of molecules within the pore.

With the metrology refined images, we proceeded to solving the indexing ambiguity for our datasets. The space group for our crystals, I4, has two possible indexing modes. In our case, solving the indexing ambiguity was especially difficult because the crystals have small unit cell size so fewer diffraction spots per image. It is hard to find common reflections between the images. However, we were able to using the algorithm developed by Brehm and Diederichs and implemented in cxi.Brehm_Diederichs [16] to solve it through iterative clustering of the two indexing modes.

Using the solved re-indexing operators, we constructed full intensities out of multiple partial measurements using the post-refinement algorithm PRIME [16,17]. To obtain optimal values for post-refinement parameters, we did a grid search of different unit cell sizes, I/sigI values and varying spread of the energy spectrum. The grid search over the energy spectrum is especially important since there is a significant amount of shot to shot variation of the X-ray energies that delivered to the crystals. We used improvement in the correlation coefficient between two halves of our datasets to score our grid searches. The merged data sets had 1.4 Å resolution with high completeness and CC 1/2 values (**Table 4.3**).

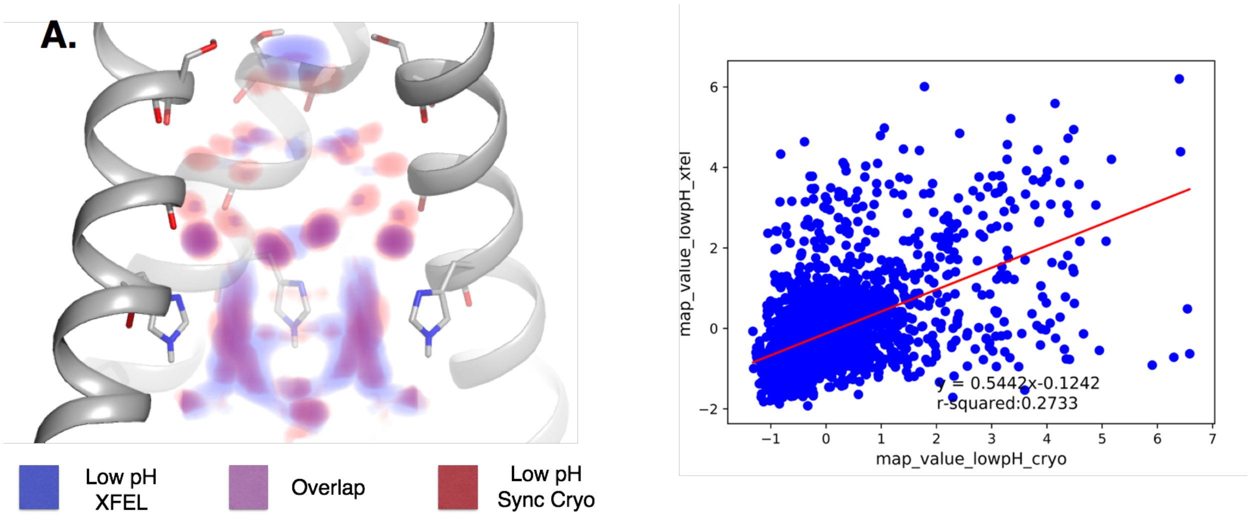
Condition	Resolution(Å)	Completness (%)	CC 1/2 (%)
Low pH	1.40 - 21.20	99.95	99.86(77.82)
Inter pH	1.40 - 21.12	99.25	99.88(58.99)
High pH	1.40 - 21.04	99.93	99.93(75.35)

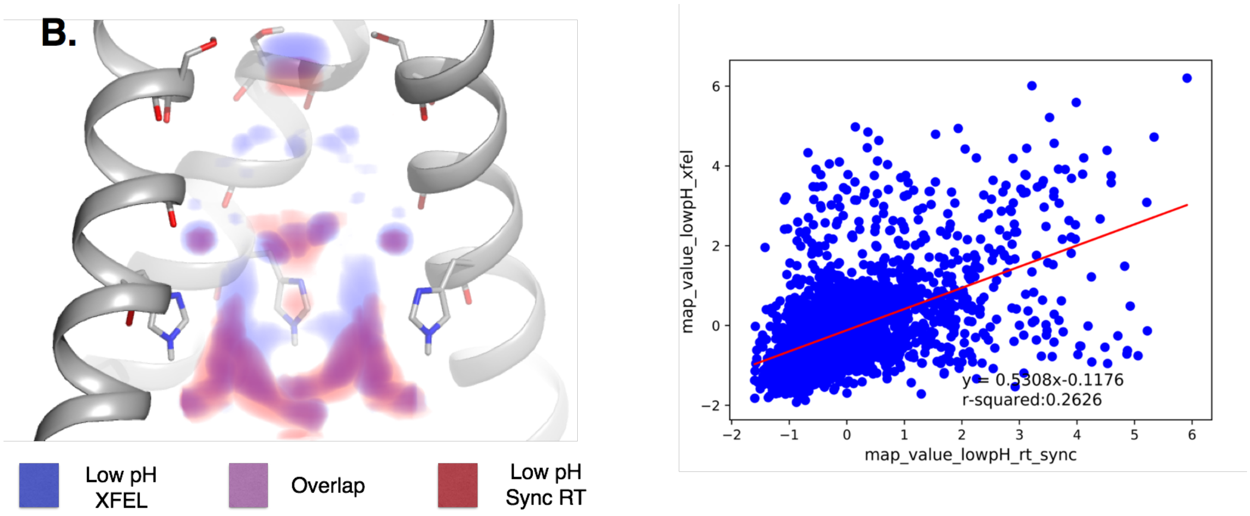
Table 4.3: Summary of refinement statistics for the three 1.4 Å resolution datasets. The datasets had high resolution, completeness and correlation.

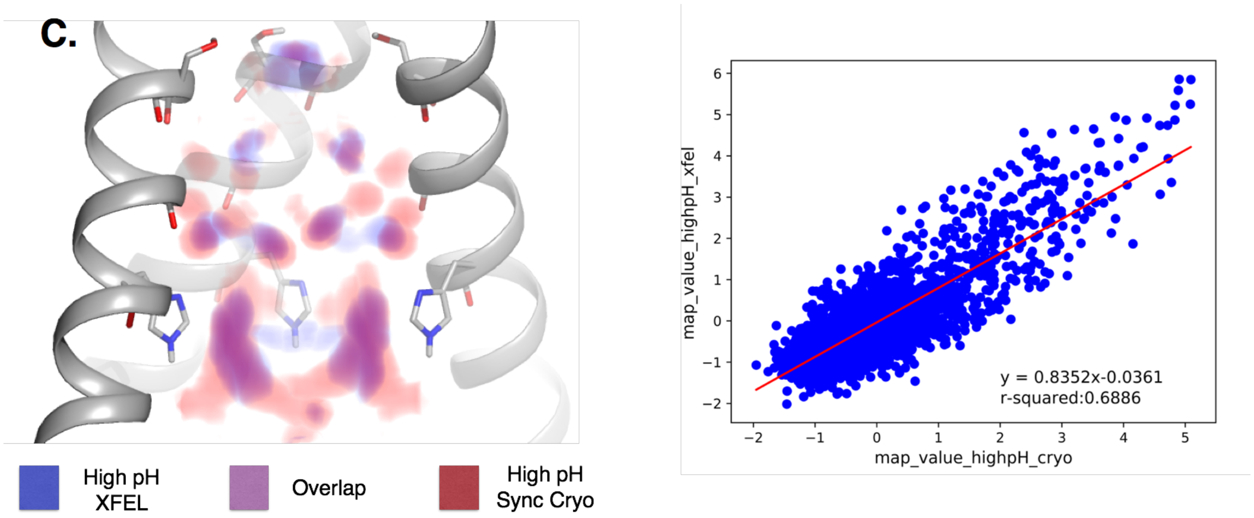
Diffraction data collected at synchrotron sources either contain artifacts associate with cryo-cooling or suffer from radiation damage

Treating the XFEL derived electron density as the ground truth, we can detect both of the challenges associated with synchrotron X-ray diffraction data collection; artifacts of cryo-cooling and radiation damage associated with room temperature data collection. Comparing the cryogenic electron density for the pore of the channel to the XFEL electron density, we identified densities in the pore that are unique to the cryogenic datasets in both low pH and high pH conditions (**Figure 4.5A,C**). We concluded that these unique densities are due to non-native trapping of waters during cryo-cooling of the crystals used to obtain the diffraction patterns. Conversely, for the datasets that were collected at room temperature using a synchrotron source showed loss and re-distribution of some densities as a result of radiation damage (**Figure 4.5B,D**).

Quantification of the election density in the pore further shows limitations of our synchrotron structures. The correlation plots for the electron densities show that the low pH conditions (**Figure 4.5A,B: right**) are more susceptible to temperature and radiation dose dependent water redistribution compared to the high pH conditions (**Figure 4.5C,D: right**). This observation is consistent with the idea that as the temperature drop during freezing the apparent pH of some buffer (MES and Tris) should increase [18]. And therefore, the high pH condition would not be affected by the change in pH of the buffer as much as the low pH condition.


A.
Low pH XFEL
Overlap
Low pH Sync Cryo
map_value_lowpH_xfel
map_value_lowpH_cryo
y = 0.5442x-0.1242
r-squared:0.2733


B.
Low pH XFEL
Overlap
Low pH Sync RT
map_value_lowpH_xfel
map_value_lowpH_rt_sync
y = 0.5308x-0.1176
r-squared:0.2626


C.
High pH XFEL
Overlap
High pH Sync Cryo
map_value_highpH_xfel
map_value_highpH_cryo
y = 0.8352x-0.0361
r-squared:0.6886

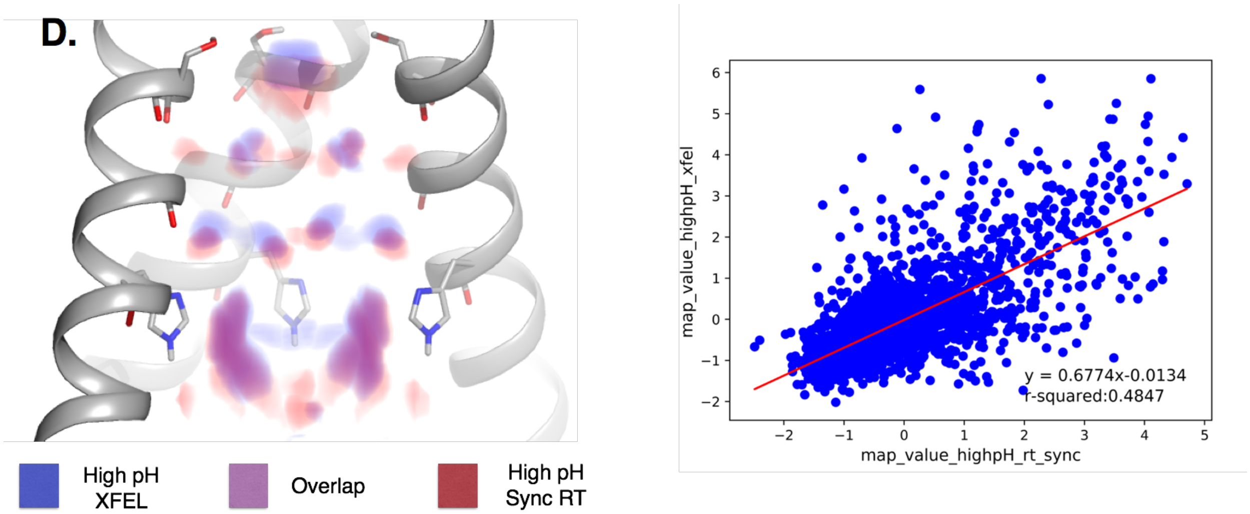


Figure 4.5: Diffraction data collected at synchrotron sources either contain artifacts associate with cryo-cooling or suffer from radiation damage. The electron density for the solvent in the channel obtained at cryogenic conditions **(A, C:** left**)** show additional water densities that correspond to water molecules trapped in non-native states. Conversely, datasets collected at room temperature using a synchrotron source show loss of water density at the top of the channel due to radiation damage and re-distribution of some densities (**B, D:** left). Quantification of the electron density of pore show that the low pH conditions **(A, B:** right) are more susceptible to temperature and radiation dose dependent water redistribution compared to the high pH conditions **(C, D:** right).

At low pH the pore of the M2 channel contains more water molecules

Our new radiation damage free, XFEL structures allow us to draw new conclusions about the solvent in the M2 channel. At the low pH condition, the pore contains more water compared to the high pH condition (**Figure 4.6**: left). The extra waters form layers of water molecules capable of conducting a proton or a hydronium ion through the channel. For most of the pore, there is a good correspondence between the the two pH conditions (**Figure 4.6**: right). This indicates that the energetic cost of interconverting between the two states are minimal.

Furthermore, the electron density in the intermediate pH state shows features that are more similar to the high pH state (**Figure 4.7**). Moreover, the solvent density is diffused at the top of the pore and covers most of the positions of the waters adopted by the other two pH conditions. Interestingly, the diffused density covers most of the top part of the channel at a about equal occupancy. Possibly, at this intermediate state the waters are sampling multiple energetically comparable states.

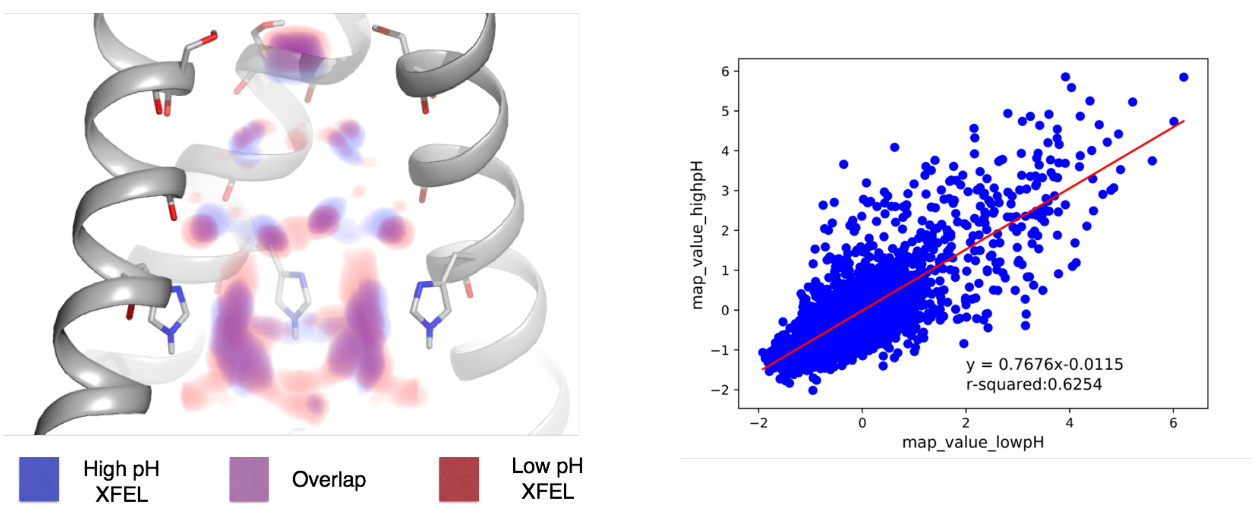


Figure 4.6: At low pH the pore of the M2 channel contains more water molecules. The extra low pH waters can interact with the backbone and side-chains of the protein to stabilize the proton conducting state **(left).** There is a good correspondence for the rest of the solvent density **(left and right**) which indicates that the energy penalty for redistribution of the pore solvent to create a conducting state is not very high.

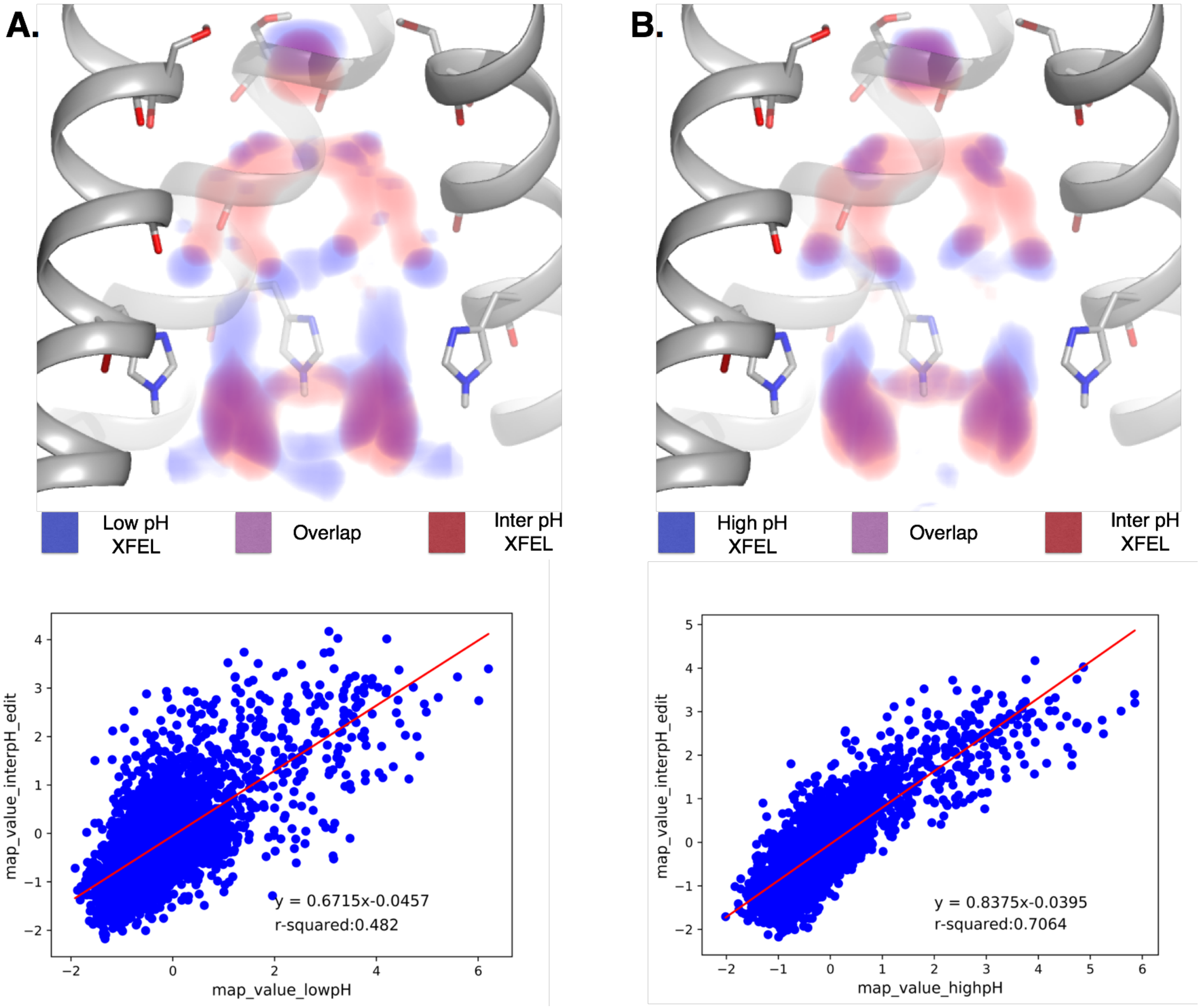
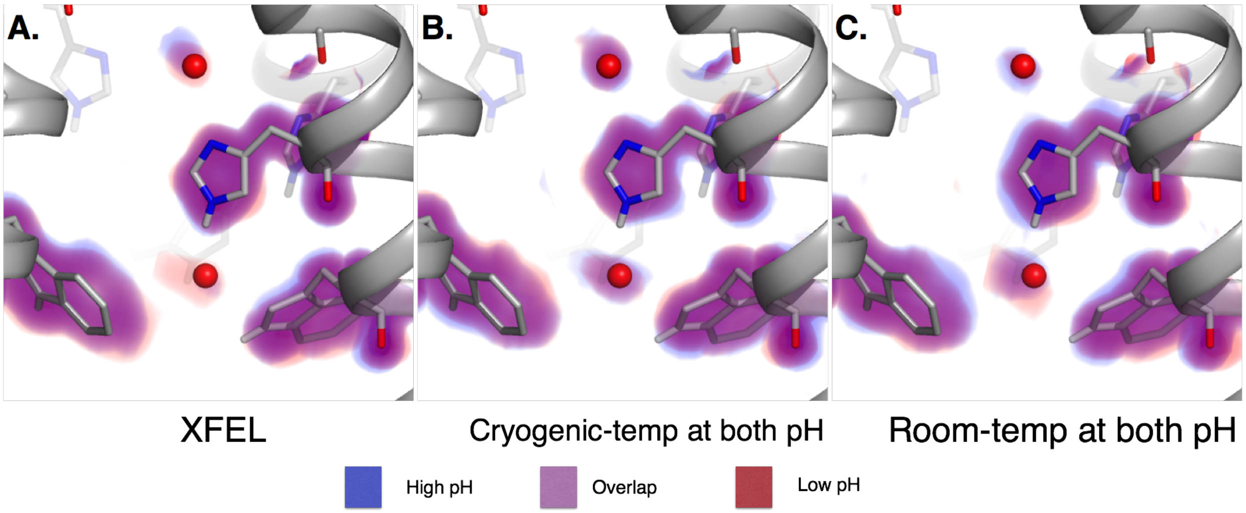


Figure 4.7: The electron density in the intermediate pH state is more similar to the high pH state. Comparing the intermediate condition **(A,B:** red density) to low and high pH conditions **(A,B:** blue densities)**,** we show that the intermediate density is more similar to that of the high pH condition **(A,B:** bottom graphs**).** However, there is a diffused solvent density at the top of the channel that seems to cover most of the possible positions of the waters at about the same occupancy.

The XFEL structures reveal key waters near the selective filter of the channel, His 37, that can stabilize the conducting state only at low pH conditions

In addition, the solvent in the channel at low pH can interact with the backbone and side-chains of the protein to stabilize the conducting state. At high pH, we see more protein conformational heterogeneity. His 37 which is a key residue in the proton conduction mechanism is stabilized through interaction with two waters found within a hydrogen bonding distance from the residue at low pH (**Figure 4.8**). This interaction allows proton to be shuttled through the histidine into the virus. Surprisingly, the water found below the histidine is only absent in the high pH XFEL structure. The absence allows the residue to adopt multiple conformations (**Figure 4.8A**). Both cryogenic and room temperature datasets collected at a synchrotron contain densities that correspond to the same water (**Figure 4.8B,C**) indicating the importance of using XFELs to visualize radiation and temperature sensitive features of proteins.



XFEL

Cryogenic-temp at both pH

Room-temp at both pH

Figure 4.8: The XFEL structures reveal that key waters near the selective filter of the channel, His 37, stabilize the conducting state only at low pH conditions. Electron density for the water below His 37 is only absent in the high pH, XFEL structure **(A)**.

Discussion

Water plays essential roles in many biological processes [1]. However, atomic resolution insights to how the conformational dynamics of water molecules contribute had remained unclear mostly due to limitation of conventional techniques. Here, we use XFELs to study the conformational heterogeneity of water molecules within the M2 proton channel as a function of pH. We solved three 1.4 Å resolution structures at low (pH 5.5), intermediate (pH 6.2) and high (pH 8.0) pH conditions using serial femtosecond crystallography [19,20]. We were able to obtain complete datasets with high correlation coefficients with our data processing pipeline that included metrology refinement. The refinement raised our indexing rates and minimize noise in intensity measurement. The resolution was only limited by the geometry of our experimental setup.

The XFEL structures deviate from cryogenic and room temperature structures determined using a synchrotron source. We found that the cryogenic structures contain water molecules that have been trapped in non native conformations during the freezing process. This observation is consistent with with previous studies [8,21]. Conversely, the room temperature synchrotron structures suffered from radiation damage. We saw loss of water density throughout the pore.

The XFEL structures give insights to why the proton conduction occurs at low pH conditions. There are more water molecules that can facilitate proton conduction at this pH conditions in the channel. This observation is also supported by the trend we see in unit cell distributions for the different conditions. The low pH unit cell distribution is wider and shorter indicating while for the high pH condition the unit cell distribution is slimmer and longer. Furthermore, we have found that waters found within hydrogen bonding distance from the selective filter of the protein, His37, are essential in the proton conduction mechanism. At low pH, these two waters stabilize His 37 through hydrogen bonding interactions. At high pH, the bottom water is absent allowing the residue to adopt multiple conformations.

Overall, our work show that temperature is a key factor for visualizing physiologically important water distribution. Furthermore, we saw that radiation damage can significantly perturb conformational distribution of sensitive features such as waters in a channel. The next steps include determining how robust the water distribution are to different XFEL data processing techniques. Furthermore, visualizing the coordinate reorganization of water and protein using pH jump experiments would give unprecedented insight into the role structure and dynamics of water molecules in proton conduction. Once established pH jump and other types of time resolved experiments can be used to study other roles water molecules play in biology at atomic resolution.

Methods

Protein synthesis and crystallization

M2 transmembrane residues 22 to 46 (Ac-SSDPLVVAASIIGILHLILWILDRLNH3) were synthesized and crystalized as previously described [7]. The crystals were maintained in three pH conditions that correspond to pHs that M2 experiences at different stages of the viral life cycle, pH 8.0, 6.2 and 5.5.

XFEL data collection

We collected serial femtosecond crystallography datasets using the LCP injector placed inside of the Diverse application platform for hard X-ray diffraction in SACLA (DAPHNIS) in a Helium chamber [12]. To achieve the viscosity needed for the LCP injector, we added equal volume of monoolein to our crystal slurry. The datasets were collected at room temperature (20°C). The beam had 1.5 μm by1.5 μm spot size with 10.5 KeV energy for the low pH condition and 10.7 KeV for intermediate and high pH conditions. The diffraction data was collected on MPCCD detector using 10 fs pulse duration at 30 Hz and 510 uJ/pulse.

XFEL data processing

Images with diffraction spots were identified using CHEETAH [13]. The hits were then indexed using optimal spot shape model, I/sigI values and outlier recognition algorithms using dials.stills_process [14]. To obtain the optimal parameter values, we did a grid search and selected the values that gave us the highest number of indexed images. To minimize differences between batches of crystals and varying geometry of the detector from run to run, we jointly refined the crystal model and the geometry of the detector using DIALS ([14,15]. We repurposed the algorithm developed for the LCLS's CSPAD detector to do the refinement on SACLA MPCCD data. To minimize the difference between observed and predicted spot locations, we performed a metrology refinement that included unit cell and crystal orientation, detector rotation, tilt and distance as a whole and as quadrants. The results were filtered by RMSD. This step helped to minimize noise during the integration of the spots and allowed us to integrate additional images (**Table 4.2**).

Next, we used the algorithm developed by Brehm and Diederichs to solve the indexing ambiguity in our datasets [16]. Our crystals had an I4 space group symmetry and therefore has two possible indexing modes. We iteratively clustered the two indexing modes using the cxi.Brehm_Diederich implementation of the algorithm [22].

To overcome the challenge of constructing full intensities out of multiple partial measurement, we used post-refinement [17]. The post-refinement procedure uses information about our crystal lattice as well as the distribution of the peaks to give the full spot shape . To obtain optimal values for the post-refinement, we did a grid search of different unit cell sizes, I/sigI values and different spread of the energy spectrum. We scored our grid by looking at improvements in the correlation coefficient between two halves of our datasets (CC1/2). After merging our datasets, we were able to obtain 1.4 Å resolution structures with high completeness and CC 1/2 values (**Table 4.3**).

Model building and refinement

The structure was solved by molecular replacement using the program Phaser in PHENIX [23] Previously structures processed using CCTBX.XFEL (low pH: 5JOO, intermediate pH: 5UMI; and high pH: 5TTC) were used as search models [24,25] . Positions and occupancies of the waters within the channel and of His37 were adjusted using Coot [26] and refined using PHENIX [23,26].

Electron density visualisation

For easy visualization of solvent in the M2 channel, we chose to show the electron density around the solvent in a volume representation as implemented in PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). This representation of the electron density allowed us to capture weak electron density since there is no need to contour the election density to a specific cut-off. The figures show electron density values ranging from 0.01 σ to 4.99 σ values.

Electron density quantification

To quantify the electron density within the channel, we aligned the structures and measured the map values at the same points using phenix.superpose_pdbs and phenix.map_value_at_point. The coordinates for the points were obtained from pseudo atoms created with 1 Å map grid spacing. We plotted the correlation for these coordinates at varying pH and data collection conditions.

Acknowledgements

R.A.W. is supported by the National Science Foundation (NSF) Graduate Research Fellowship Program; J.S.F. is a Searle Scholar, Pew Scholar, and Packard Fellow, and is supported by NIH Grants OD009180 and GM110580 and NSF Grant STC-1231306. S.I., O.N., and F.Y. were supported by the X-Ray Free-Electron Laser Priority Strategy Program (Ministry of Education, Culture, Sports, Science and Technology). We acknowledge computational support from SACLA high-performance computing system and Mini-K supercomputer system. J.L.T.,W.F.D., and experimental work were supported by NIH Grants GM122603 and GM117593. N.K.S. acknowledges NIH Grant GM117126 for computational methods. Preliminary XFEL diffraction experiments were carried out at Linac Coherent Light Source (LCLS) X-ray pump–probe (Protein Crystal Screening Proposal LG53). Use of the LCLS, SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-76SF00515. The XFEL datasets described in this paper were collected at BL3 of SACLA with the approval of the Japan Synchrotron Radiation Research Institute (Proposals 2015A8048, 2015B8028, and 2016A8030).

References

1. Ball P: **Water as an active constituent in cell biology**. *Chem Rev* 2008, **108**:74–108.

2. Grossman M, Born B, Heyden M, Tworowski D, Fields GB, Sagi I, Havenith M: **Correlated structural kinetics and retarded solvent dynamics at the metalloprotease active site.** *Nat Struct Mol Biol* 2011, **18**:1102–1108.

3. Dielmann-Gessner J, Grossman M, Nibali VC, Born B, Solomonov I, Fields GB, Havenith M, Sagi I: **Enzymatic turnover of macromolecules generates long-lasting protein–water-coupled motions beyond reaction steady state**. *Proceedings of the National Academy of Sciences* 2014, **111**:17857–17862.

4. Ball P: **Water is an active matrix of life for cell and molecular biology.** *Proc Natl Acad Sci U S A* 2017, **114**:13327–13335.

5. Banerjee I, Yamauchi Y, Helenius A, Horvath P: **High-Content Analysis of Sequential Events during the Early Phase of Influenza A Virus Infection**. *PLoS One* 2013, **8**:e68450.

6. Gu R-X, Liu LA, Wei D-Q: **Structural and energetic analysis of drug inhibition of the influenza A M2 proton channel**. *Trends Pharmacol Sci* 2013, **34**:571–580.

7. Thomaston JL, Alfonso-Prieto M, Woldeyes RA, Fraser JS, Klein ML, Fiorin G, DeGrado WF: **High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction**. *Proc Natl Acad Sci U S A* 2015, **112**:14260–14265.

8. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature X-ray crystallography**. *Proc Natl Acad Sci U S A* 2011, **108**:16247–16252.

9. Halle B: **Biomolecular cryocrystallography: structural changes during flash-cooling.** *Proc Natl Acad Sci U S A* 2004, **101**:4793–4798.

10. Chapman HN, Fromme P, Barty A, White TA, Kirian RA, Aquila A, Hunter MS, Schulz J, DePonte DP, Weierstall U, et al.: **Femtosecond X-ray protein nanocrystallography.** *Nature* 2011, **470**:73–77.

11. Neutze R, Wouts R, van der Spoel D, Weckert E, Hajdu J: **Potential for biomolecular imaging with femtosecond X-ray pulses**. *Nature* 2000, **406**:752–757.

12. Tono K, Nango E, Sugahara M, Song C, Park J, Tanaka T, Tanaka R, Joti Y,

Kameshima T, Ono S, et al.: **Diverse application platform for hard X-ray diffraction in SACLA(DAPHNIS): application to serial protein crystallography using an X-ray free-electron laser**. *J Synchrotron Radiat* 2015, **22**:532–537.

13. Barty A, Kirian RA, Maia FRNC, Hantke M, Yoon CH, White TA, Chapman H: **Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data**. *J Appl Crystallogr* 2014, **47**:1118–1131.

14. Brewster AS, Waterman DG, Parkhurst JM, Gildea RJ, Michels-Clark TM, Young ID, Bernstein HJ, Winter G, Evans G, Sauter NK: **Processing XFEL data with cctbx.xfel and DIALS**. *Computational Crystallography Newsletter* 2016,

15. Waterman DG, Winter G, Gildea RJ, Parkhurst JM, Brewster AS, Sauter NK, Evans G: **Diffraction-geometry refinement in the DIALS framework**. *Acta Crystallogr D Struct Biol* 2016, **72**:558–575.

16. Brehm W, Diederichs K: **Breaking the indexing ambiguity in serial crystallography**. *Acta Crystallogr D Biol Crystallogr* 2014, **70**:101–109.

17. Uervirojnangkoorn M, Zeldin OB, Lyubimov AY, Hattne J, Brewster AS, Sauter NK, Brunger AT, Weis WI: **Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals**. *Elife* 2015, **4**.

18. Williams-Smith DL, Bray RC, Barber MJ, Tsopanakis AD, Vincent SP: **Changes in apparent pH on freezing aqueous buffer solutions and their relevance to biochemical electron-paramagnetic-resonance spectroscopy**. *Biochem J* 1977, **167**:593–600.

19. Boutet S, Lomb L, Williams GJ, Barends TRM, Aquila A, Doak RB, Weierstall U, DePonte DP, Steinbrener J, Shoeman RL, et al.: **High-resolution protein structure determination by serial femtosecond crystallography**. *Science* 2012, **337**:362–364.

20. Liu W, Ishchenko A, Cherezov V: **Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography**. *Nat Protoc* 2014, **9**:2123–2134.

21. Keedy DA, van den Bedem H, Sivak DA, Petsko GA, Ringe D, Wilson MA, Fraser JS: **Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR**. *Structure* 2014, **22**:899–910.

22. Sauter NK, Hattne J, Brewster AS, Echols N, Zwart PH, Adams PD: **Improved crystal orientation and physical properties from single-shot XFEL stills**. *Acta Crystallogr D Biol Crystallogr* 2014, **70**:3299–3309.

23. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive**

Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 2010, **66**:213–221.

24. Hattne J, Echols N, Tran R, Kern J, Gildea RJ, Brewster AS, Alonso-Mori R, Glöckner C, Hellmich J, Laksmono H, et al.: **Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers**. *Nat Methods* 2014, **11**:545–548.

25. Thomaston JL, Woldeyes RA, Nakane T, Yamashita A, Tanaka T, Koiwai K, Brewster AS, Barad BA, Chen Y, Lemmin T, et al.: **XFEL structures of the influenza M2 proton channel: Room temperature water networks and insights into proton conduction**. *Proc Natl Acad Sci U S A* 2017, **114**:13357–13362.

26. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics.** *Acta Crystallogr D Biol Crystallogr* 2004, **60**:2126–2132.

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.

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 Rahel Woldeyes Date 12\28\2017