

Atomistic Insights into the Regulation of a Thermophilic CRISPR-Cas9 by Anti-CRISPR and Intrinsic Protein Dynamics

Alexa Loretta Knight

B.S., University of Washington Tacoma, 2022

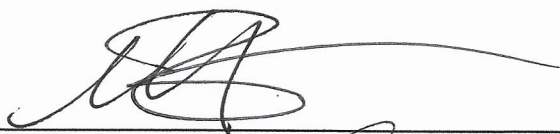
B.A., University of Massachusetts Boston, 2017

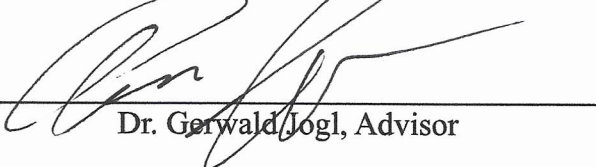
*A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in the Molecular Biology, Cell Biology, and Biochemistry Graduate Program at
Brown University*

April 2026

© Copyright 2026 by Alexa Loretta Knight

This dissertation by Alexa Knight is accepted in its present form by the Molecular Biology, Cell Biology, and Biochemistry Graduate Program as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

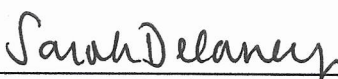
Date 4/23/26 
Dr. George Lisi, Advisor

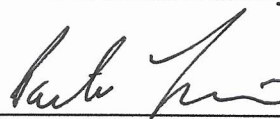
Date 4/23/26 
Dr. Gerwald Jogl, Advisor

Recommended to the Graduate Council

Date 4/23/26 
Dr. Mark Johnson, Reader

Date 4/23/26 
Dr. Erica Larschan, Reader

Date 4/23/26 
Dr. Sarah Delaney, Reader

Date 4/23/26 
Dr. Patrick Loria, Outside Reader

Approved by the Graduate Council

Date _____
Dr. David P. Lindstrom, Dean of the Graduate School

Alexa Knight

Providence, RI 02903 • alexa_knight@brown.edu

ORCID ID: 0000-0002-5407-0157

LinkedIn: <https://www.linkedin.com/in/alexa-knight-567384215/>

EDUCATION

Brown University

Ph.D. Candidate

Providence, RI

2022 - Present

University of Washington

B.Sc., Biomedical Sciences

Summa Cum Laude

Tacoma, WA

2018 - 2022

University of Massachusetts

B.A., Anthropology

Summa Cum Laude

Boston, MA

2013 - 2017

RESEARCH EXPERIENCE

Brown University

Ph.D. Candidate

Department of Molecular Biology and Biochemistry

Advisor: George Lisi, Ph.D. & Gerwald Jogl Ph.D.

Providence, RI

Sept 2022 - Present

Project: *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR* - Structural and biochemical studies of anti-CRISPR proteins in complex with Cas9, using solution NMR, X-ray crystallography, cryo-EM, and biochemical techniques

University of Washington

Undergraduate Researcher

Department of Sciences and Mathematics

Advisor: Sarah Alaei, Ph.D.

Tacoma, WA

2019- 2022

Project: *Investigating the Roles of LPS Modifications in Outer Membrane Vesicle Biogenesis* - Interrogated molecular mechanisms that drive production and function of *Porphyromonas gingivalis* outer membrane vesicles using transmission electron microscopy and confocal microscopy

Project: *Isolation and Identification of Arsenic Resistant Bacteria from Washington Lakes* - Investigated presence and abundance of arsenic resistant bacteria through 16s rRNA sequencing

PROFESSIONAL EXPERIENCE

Quantitative Consultant

University of Washington | 2020-2022

- Provided one-on-one tutoring to undergraduate students
- Developed and ran weekly calculus workshops for undergraduate students
- Specialized in tutoring subjects including general chemistry, general biology, organic chemistry, microbiology, calculus I, II, & III, and matrix algebra

Family and Community Partnerships Coordinator

Denise Louie Education Center | 2017-2018

- Assisted homeless and low-income families in accessing resources related to housing, education, and health
- Designed organizational tools to track program outcome data for grant writing purposes
- Responsible for conducting audits of early learning sites and preschools
- Managed a team of 12 staff members

AWARDS AND SCHOLARSHIPS

- Blavatnik Family Graduate Fellowship in Biology and Medicine (2025)
- Elected **Chair** of the Computational Aspects of Biomolecular NMR Gordon Research Seminar (to be held 2027)
- NSF Graduate Research Fellowship (Honorable Mention, 2024)
- New England Structural Biology Symposium Poster Award (2024)
- NIH T-32 Pre-Doctoral Fellowship (2023, 2024)
- Mary Gates Research Endowment Scholarship (2021)
- Departmental Senior Thesis Honors Award – Anthropology (2017)

PUBLICATIONS

9) Martinez-Guzman, M., Belato, H.B., Nagle I., Chaubey, M., Rose, T.G., **Knight, A.L.**, Fick, O., Lisi, G.P., Kizer, M.E.* Structural characterization of the novel Sso7d-based glycan binding protein's molecular determinants of ligand recognition: probing the role of CH- π and hydrogen bonding in carbohydrate binding. *Submitted*.

8) **Knight, A.L.**; Belato, H.B.; Dresser, C.; Luo, J.; Pindi, C.; D'Ordine, A.M.; Arantes, P.; Mercado, B.; Palermo G.; Lasekan; Arantes, P.; P.; Jogl, G.; Lisi, G.P.* (2025). Orthosteric and allosteric effects of anti-CRISPR II-C1 inhibition of GeoCas9 from integrated structural biology. *Submitted*.

7) Ahsan, M.; Saha, A.; Ramos, D.; Strohkendl, I.; **Knight, A.L.**; Skeens, E.; Lisi, G.P.; Taylor, D.W.; Palermo, G.* A Cryptic Binding Pocket Regulates the Metal-dependent Activity of Cas9 *bioRxiv*. 2025. DOI: 10.1101/2025.08.25.672025 *Under Review*.

6) Molina Roca, C.E.†, **Knight, A.L.**, & Lisi, G.P.* (2025). Thermodynamic, kinetic, and ensemble aspects distinguishing long-range allosteric activation of DNA cleavage in Cas proteins. *Physical Biology*. 2026. 23. 021001-021012

5) **Knight, A.L.**, Luo, J.; and Lisi, G.P.* Assessing Temperature-dependent DNA Cleavage by a Thermophilic CRISPR-Cas9. *Bio-protocol*. (2025). 15. e5463-e5474

4) Alaci, S.R., King, A.J., Banani, K., Reddy, A., Ortiz, J., **Knight, A.L.**, Haldeman, J., Su, T.H., Park, H., Coats, S.R., & Jain, S. (2025). Lipid A remodeling modulates outer membrane vesicle biogenesis by *Porphyromonas gingivalis*. *Journal of Bacteriology*, 207, e00336-24. <https://doi.org/10.1128/jb.00336-24>

3) **Knight, A.L.**, & Lisi, G.P.* (2025). Spy-ing on nucleic acids: Atomic resolution of the *S. pyogenes* CRISPR-Cas9 surveillance state structure. *Structure*, 33, 636–638. (Structure Spotlight)

2) Belato, H.B.†, **Knight, A.L.**†, D'Ordine, A.M., Pindi, C., Fan, Z., Luo, J., Palermo, G., Jogl, G., & Lisi, G.P.* (2024). Structural and dynamic impacts of single-atom disruptions to guide RNA interactions within the recognition lobe of *Geobacillus stearothermophilus* Cas9. *eLife*. <https://doi.org/10.7554/eLife.99275.1> †These authors contributed equally.

1) **Knight, A.L.**†, Widjaja, V.†, & Lisi, G.P.* (2023). Temperature as a modulator of allosteric motions and crosstalk in mesophilic and thermophilic enzymes. *Frontiers in Molecular Biosciences*. <https://doi.org/10.3389/fmolb.2023.1281062> (Thematic issue: "Allosteric Functions and Inhibitions: Structural Insights") †These authors contributed equally.

PRESENTATIONS

16) **Poster** – New England Science Symposium (NESS) – Harvard Medical School – Boston, MA **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

15) **Poster** – Biophysical Society Annual Meeting – San Francisco, CA (2026) **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

14) **Oral Presentation** - Invited Speaker for Chemistry Seminar Series - Oxford University (2025) **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

13) **Poster** - Molecular Biophysics in the Northeast - Durham, NH (2025) **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

12) **Oral Presentation** - Invited Speaker at The Bavarian NMR Center (BNMRZ) at The Technical University of Munich (TUM) (2025) **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

11) **Poster** – Gordon Research Conference – Computational Aspects of Biomolecular NMR - Les Diablerets, Switzerland (2025) **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

10) **Oral Presentation** – MCB Data Club (Brown University) - Providence, RI (2025 & 2024) **Knight, A.L.**....Lisi, G.P. *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

9) **Poster** – The Protein Society Annual Meeting (TPS) - Vancouver, British Columbia (2024) **Knight, A.L....Lisi, G.P.** *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

8) **Poster** – New England Structural Biology Symposium (NESS) - Farmington, CT (2024) **Knight, A.L....Lisi, G.P.** *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

7) **Poster** – MCB Retreat (Brown University) - Warren, RI (2023) **Knight, A.L....Lisi, G.P.** *Investigation of Anti-CRISPR Inhibition of a Thermostable GeoCas9 by Solution NMR*

6) **Poster** – The Biophysical Society Annual Meeting - San Diego, CA (2023) **Knight, A.L....Lisi, G.P.** *Characterizing the Structure and Dynamics of the Recognition Lobe in a Thermophilic Cas9.*

5) **Poster** – American Society of Microbiology Annual Meeting (ASM) - Washington, DC (2022) **Knight, A.L....Alaei, S.** *Investigating the Roles of LPS Modification in Outer Membrane Vesicle Biogenesis.*

4) **Poster**– Undergraduate Research Symposium - Seattle, WA (2022) **Knight, A.L....Alaei, S.** *Investigating the Roles of LPS Modification in Outer Membrane Vesicle Biogenesis.*

3) **Oral Presentation** – Undergraduate Research Symposium - Virtual (2021) **Knight, A.L....Alaei, S.** *Identifying and Characterizing Arsenic Resistant Bacteria Species in Periphyton Communities*

2) **Oral Presentation** – Undergraduate Thesis Defense - Boston, MA (2017) **Knight, A.L....Clarkin, P.** *Correlates of Low Birth Weight Infants and Public Health*

1) **Oral Presentation** – Harvard Peabody Museum Docent Presentation – Boston, MA (2017) **Knight, A.L.** *The Secret Sex Life of Anthropological Artifacts: Gender and Race in the Museum*

TEACHING, MENTORING, AND OUTREACH EXPERIENCE

- Teaching Assistant for Intro Microbiology (BIOL 0510)
Brown University, Providence RI (2024)
- Invited Lecturer for Microbial Physiology (TBIOL 401)
University of Washington, Tacoma WA (2023, 2024)
- GRC Computational Aspects of Biomolecular NMR Discussion Leader
Les Diablerets, Switzerland (2025)
- MCBGP Admissions Committee
Graduate student committee member for admissions to MCBGP (2024, 2025)
- Brown Student Parent Community President
Founded an organization for graduate student parents to access community (2024)

- Brown University Undergraduate Poster Day
Poster Judge (2023 – present)
- Diversity and Inclusion Committee (MCB Graduate Program)
Committee Member (2022-2024)

Rotating Graduate Students:

- Briana Mercado ~6 months
- Praise Lasekan ~6 months

Undergraduate Students:

- Charlotte Dresser ~1 year

SKILLS

Site-directed mutagenesis, Spectrophotometry, PCR, protein purification, Pymol, GraphPad Prism, X-Ray Crystallography, Vitrobot grid preparation, Cryo-EM grid screening, Microscale Thermophoresis (MST), CD, FPLC, Oral Science Communication, solution NMR, Sparky, nmrPipe, nmrDraw, ITC, DSF

Acknowledgements

I would like to express my sincere gratitude to my co-advisors Dr. George Lisi and Dr. Gerwald Jogl. George, thank you for giving me the opportunity to explore and fall in love with the field of structural biology when I knew nothing going in. Your unending support has been invaluable throughout this experience. Gerwald, thank you for always encouraging my enthusiastic (and at times unrealistic) scientific goals with laughter and a “let’s find out” mindset. I am incredibly grateful to you both for all you have taught me about structural biology, biophysics, and how to navigate the academic world.

Thank you to my committee members, Dr. Mark Johnson, Dr. Sarah Delaney, and Dr. Erica Larschan. I greatly appreciate your guidance and encouragement throughout my PhD.

I would like to thank my past advisors Dr. Patrick Clarkin and Dr. Sarah Alaei. Your guidance and encouragement were instrumental in me becoming the scientist I am today. I truly would not be here without you, and I am forever grateful. You both changed my life.

To my lab mates, I am eternally grateful for your mentorship, encouragement, and friendship. I have learned so much from all of you and I am constantly in awe of your hard work and dedication to your science. I am so excited to see where you all go, I am certain the future is bright. A special thank you to Dr. Helen Belato for being my first rotation mentor and continued teacher, I owe so much to you.

To my Tripod, thank you for grounding me and giving me the gift of laughter and joy throughout grad school. Thank you for reminding me that there is life outside the lab, and that a lot of my problems can be solved by companionship, CrossFit, and a Caesar salad with a side of fries. Thank you for being such sweet aunties to my girls.

To my family, thank you for being such incredible supports, especially over the past few years. I couldn’t have done this without your countless dinners, school pick-ups, and late-night phone chats from the lab. Thank you from the bottom of my heart for being there for my girls and Sam, you are the backbone of this whole operation. Thank you for reminding me that the world is a beautiful place and magical things can happen if you don’t give up.

To Graihagh and Aoife, thank you for being my motivation and inspiration. You have given this whole process meaning. Whenever I wanted to give up and whenever I felt afraid, I would think of you, and you would make me strong. This Ph.D. is for you, because I want you both to know that you can do anything if you don’t give up. You are smart, strong, and capable girls. I cannot wait to see what the future holds for you. I love you and I am so incredibly proud of you both.

To my Sam, there are not enough pages in the world to describe my gratitude for all you've done.

*Long was the way that fate them bore,
O'er stony mountains cold and grey,
Through halls of iron and darkling door,
And woods of nightshade morrowless.
The Sundering Seas between them lay,
And yet at last they met once more,
And long ago they passed away
In the forest singing sorrowless.*

Table of Contents

Title Page...	i
Copyright Page...	ii
Signature Page...	iii
Curriculum Vitae...	iv
Acknowledgements...	x
Table of Contents.	xii
List of Figures...	xvi

CHAPTER ONE

Introduction...	1
The Allosteric Regulation of Proteins Links Dynamics and Function...	2
NMR Provides High Resolution, Atomistic, Insights into the Structure and Dynamics of Allosteric Enzymes...	4
CRISPR-Cas9 is an RNA-guided Endonuclease and Revolutionary Biological Editing Tool.....	7
Cas9 Mismatch Tolerance can lead to Deleterious Off-target Effects.....	12
CRISPR-Cas9 Activity is Dependent on Allosteric Signaling between Spatially Distant Domains.....	13
Thermophilic Cas9 Provides a Promising Alternative to <i>Spy</i> Cas9.....	14
Anti-CRISPR Inhibitors Offer Spatial and Temporal Control of Cas9 Editing.....	17
References.....	20

CHAPTER TWO

Molecular Tuning of the Thermophilic *Geo*Cas9 Recognition Lobe Modulates Allosteric

Motions and Guide RNA Interactions...	29
Abstract...	30
Introduction...	31
Results...	34
Discussion...	67
Materials and Methods	70
References...	79

CHAPTER THREE

Orthosteric and Allosteric Effects of Anti-CRISPR II-C1 Inhibition of *Geo*Cas9 from

Integrated Structural Biophysics...	88
Abstract...	89
Introduction...	89
Results...	92
Discussion...	119
Materials and Methods	123
References...	134

CHAPTER FOUR

Future Considerations for the Study of Allostery in Thermophilic Enzymes.....	142
Abstract.....	143
Introduction.....	143
Current Solution Nuclear Magnetic Resonance Spectroscopy Tools for Interrogating Allosteric Systems via Temperature.....	147
The Effect of Temperature on Allosteric Communication	150
Case Study: Temperature-dependent Allosteric Activation of a Thermophilic Enzyme.....	151
Case Study: Temperature Adaptation of Dynamic Allostery	156
Conclusions and Considerations for Further Biophysical Interrogations of <i>GeoCas9</i>	157
References.....	160

CHAPTER FIVE

Discussion.....	165
Studies of the Allosteric Regulation of <i>GeoCas9</i> and the Inhibitory Mechanism of AcrIIC1..	166
Summary of Findings and Future Directions.....	167
Contributions of this Dissertation.....	172
References... ..	173

CHAPTER SIX

Contributed Works...	174
Abstract...	175
Introduction...	175
Results and Discussion...	178
Methods...	188
References...	191

List of Figures

CHAPTER ONE

Figure 1. The ensemble model of allostery acknowledges the continuum of conformational states and exchange rates that span both “fast” and “slow” timescales...	3
Figure 2. Architectures and critical structural elements of Cas9 guide RNA...	10
Figure 3. AlphaFold model of <i>GeoCas9</i> ...	16

CHAPTER TWO

Figure 1. Arrangement of <i>GeoCas9</i> domains across the primary sequence, cryo-EM structure, and <i>GeoRec</i> NMR spectra...	35
Figure 2. <i>SpyCas9</i> and <i>GeoCas9</i> Rec domain alignments, assigned HSQC spectra of <i>GeoRec</i> and subdomains...	37
Figure 3. <i>GeoRec</i> subdomains overlay nicely with ^1H - ^{15}N HSQC and ^1H - $^{13}\text{CH}_3$ ILV-methyl spectra...	38
Figure 4. Temperature-dependent CD unfolding profiles of <i>GeoRec1</i> , <i>GeoRec2</i> , and <i>GeoRec</i> reveal that the unfolding profile of the individual subdomains are conserved within that of <i>GeoRec</i> ...	38
Figure 5. Sites of K267E and R332A mutations modeled from <i>NmeCas9</i> and chemical shift perturbations induced by these mutations on <i>GeoRec</i> spectra...	41
Figure 6. HSQC spectra of WT <i>GeoRec2</i> overlaid with K267E and R332A in <i>GeoRec2</i> ...	41
Figure 7. Raw CD spectra and CD melt of WT <i>GeoRec2</i> , K267E <i>GeoRec2</i> , and R332A <i>GeoRec2</i> ...	43

Figure 8. R_1 , R_2 , and ^1H - ^{15}N NOE measurements for WT GeoRec2 compared to K267E and R332A GeoRec2 at 600 and 850 MHz.....	46
Figure 9. CPMG and order parameter measurements for WT <i>GeoRec2</i> , K267E <i>GeoRec2</i> , and R332A <i>GeoRec2</i>	48
Figure 10. NMR HSQC RNA titrations of GeoRec... ..	50
Figure 11. Structures of full-length and truncated GeoCas9 gRNAs used in this study... ..	50
Figure 12. NMR chemical shift perturbations caused by RNA binding to WT, K267E, and R332A <i>GeoRec</i>	52
Figure 13. MST-derived profiles of WT <i>GeoRec2</i> , K267E <i>GeoRec2</i> , and R332A <i>GeoRec2</i> binding 39nt portion of gRNA.....	52
Figure 14. $^1\text{H}^{15}\text{N}$ TROSY HSQC NMR spectral overlays of apo and RNA-bound WT, K267E, and R332A <i>GeoRec2</i>	54
Figure 15. RNA-induced NMR chemical shift perturbations to WT <i>GeoRec</i> overlaid with those of K267E and R332A <i>GeoRec</i>	55
Figure 16. Effects of mutations in full-length <i>GeoCas9</i> revealed by MD simulations	59
Figure 17. Analysis of MD simulations.....	60
Figure 18. <i>In vitro</i> cleavage assays show no significant difference between WT, K267E, and R332A <i>GeoCas9</i> activity.....	64
Figure 19. Off-target <i>In vitro</i> cleavage assays... ..	64
Figure 20. MST-derived binding affinities of WT, K267E, and R332A <i>GeoCas9</i> binding to a full-length Tnnt2 gRNA.....	66

CHAPTER THREE

Figure 1. Overlay of X-ray structure of AcrIIC1 bound to <i>Nme</i> HNH and to <i>Geo</i> HNH.....	93
Figure 2. X-ray crystal structure of <i>Geo</i> HNH bound to AcrIIC1.....	94
Figure 3. Snapshots of ^1H - ^{15}N HSQC titrations of <i>Geo</i> HNH with AcrIIC1 inhibitor... ..	96
Figure 4. ^1H - ^{15}N HSQC NMR spectra of <i>Geo</i> HNH titrated with WT, S78A, C79P, or S78A/C79P AcrIIC1... ..	97
Figure 5. ^1H - ^{15}N HSQC spectral titration of ^{15}N -AcrIIC1 with <i>Geo</i> HNH.....	98
Figure 6. ^1H - ^{15}N HSQC and CD spectra of WT AcrIIC1, S78A AcrIIC1, and C79P AcrIIC1.....	100
Figure 7. Summary of R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE relaxation parameters.....	102
Figure 8. NMR CPMG and order parameter measurements of WT and S78A AcrIIC1 in complex with <i>Geo</i> HNH.....	104
Figure 9. <i>In vitro</i> DNA cleavage inhibition assays.....	107
Figure 10. <i>In vitro</i> DNA cleavage inhibition assays when RNP is formed prior to addition of Acr.....	108
Figure 11. Effects of AcrIIC1 interfacial mutations on structural stability, binding energetics, and interface integrity of the <i>Geo</i> HNH-AcrIIC1 complex.....	110
Figure 12. MD derived mutation-induced changes <i>Geo</i> HNH-AcrIIC1 complex.....	111
Figure 13. Structures of sgRNA-bound and sgRNA:DNA-bound <i>Geo</i> Cas9 revealing potential steric clash with RuvC domain.....	115
Figure 14. NMR spectral overlay of ^{15}N - <i>Geo</i> HNH revealing a small number of CSPs at saturating concentrations of AcrIIC1.....	117

CHAPTER FOUR

Figure 1. Kinetic and thermodynamic aspects of protein conformational equilibria are temperature-sensitive.....146

Figure 2. Overlay of ^1H - $^{13}\text{CH}_3$ methyl-ILV NMR spectra collected at 303 K, 323 K, and 343 K depicting temperature-dependent variations in exchange regime.....153

Figure 3. Differences in Eigenvector Centrality within IGPS, intrinsic to the binding of PRFAR (holo enzyme) at 303 K or 323 K and without effector binding (apo enzyme) at the elevated temperature of 323 K.....155

CHAPTER SIX

Figure 1. Cas9 activation under varying concentrations of Mg^{2+} ions. 176

Figure 2. NMR HSQC titrations of HNH-L2 with Mg^{2+} at 298 K and 310 K. 179

Figure 3. Dynamic Impacts of Mg^{2+} Binding to HNH-L2 by NMR at 298 K and 310 K..... 180

Figure 4. Mg^{2+} binding to HNH-L2 shows negligible effect on hydrogen bonding network. ...182

Figure 5. NMR HSQC titrations of HNH-L2 with Li^+ and Al^{3+} 185

Figure 6. *In vitro* DNA cleavage assays show incongruent effect of varying concentrations of Mg^{2+} , Li^+ , and Al^{3+} on *Spy*Cas9 nuclease activity.....186

Figure 7. DSF-derived melting temperatures of Cas9 variants. 187

CHAPTER ONE

Introduction

The Allosteric Regulation of Proteins Links Dynamics and Function

Protein allostery is a fundamental and ubiquitous biomolecular regulatory mechanism by which spatially distinct sites communicate through amino acid “networks”. The transmission of these signals is coordinated by intrinsic protein dynamics through specific residues, often forming contiguous pathways^(1, 2) that are activated by endogenous or exogenous stimuli⁽¹⁾. Historically, allostery has been understood through an emphasis on the “active” and “inactive” states, which often represented extremes of a continuum visualized by static crystallographic structures⁽³⁾. However, classical models failed to describe how the structures themselves facilitated allosteric communication between distant sites. Furthermore, the models lacked biophysical or mathematical descriptors of allosteric ensembles of states, or how these states could be tuned through exogenous and endogenous factors⁽³⁾. As such, an ensemble view of allostery offers a high degree of temporal and spatial control of the conformational ensemble and activity of protein complexes through the coupling of nucleic acids or small molecules, as well as changes in pH, temperature, and chemical modification state⁽⁴⁾.

Within allosteric proteins, activity and subsequent transformation of the primary ligand (substrate) at the orthosteric (active) site is regulated (activated or suppressed) by the binding of a secondary ligand at a spatially distinct site⁽¹⁾. Allosteric signaling perturbs a protein’s equilibrium ensembles, altering the free-energy landscape of its conformational space⁽⁵⁾ (**Figure 1**). Allosteric effectors generally either affect the substrate binding affinity (K-type) or catalytic conversion rate (V-type) by altering structure and dynamics within the protein matrix ⁽⁵⁾. Accordingly, we can therefore view allostery as manifest through the alteration of the breadth of the protein conformational distribution, highlighting the role of entropic contributions to allostery.

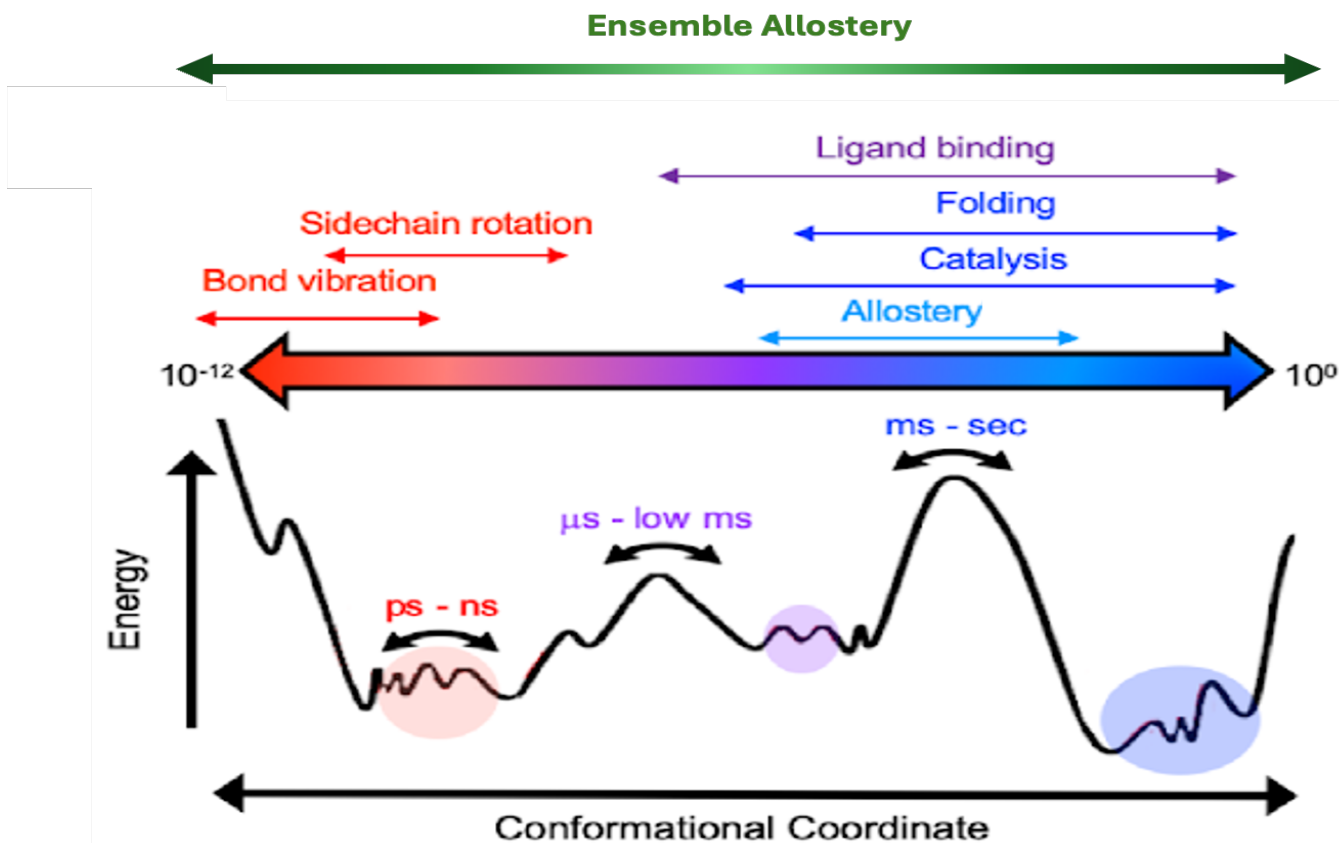


Figure 1. The ensemble model of allostery acknowledges the continuum of conformational states and exchange rates that span both “fast” and “slow” timescales, not simple the end states of “active” and “inactive”.

As allosteric enzymes regulate a wide variety of catalytic reactions, they are implicated in many crucial metabolic and physiological processes. As such, the characterization and understanding of allosteric pathways and the intrinsic protein dynamics that govern their activity have become of great interest, due to the potential to harness these mechanisms for high efficiency biological tools and alternative drug targeting sites^(2, 6, 7). Some of the most significant experimental advances in the analysis of allostery have risen from the field of nuclear magnetic resonance (NMR)⁽⁸⁾. This technique allows for the simultaneous interrogation of structure, dynamics, and thermodynamics which are crucial for the elucidation of allosteric networks. Many allosteric systems are large, multimeric, protein and nucleic acid complexes whose activation is dependent on crosstalk between disparate domains⁽⁹⁻¹¹⁾. This thesis will examine fundamental biophysical aspects of allostery in such a system, the CRISPR-Cas9 enzyme machine. This work hinges on experimental and computational tools that have been critical for resolving allosteric pathways, coupling NMR, X-ray crystallography, molecular dynamics (MD) simulation, and a variety of other biophysical techniques.

NMR Provides High Resolution, Atomistic, Insights into the Structure and Dynamics of Allosteric Enzymes

NMR delivers unparalleled atomic detail about the dynamic interconversion of unique chemical environments, referred to as chemical or conformational exchange⁽¹²⁻¹⁴⁾. The introduction of isotopic spin labels at the protein backbone amides (¹⁵N nuclei) or backbone and side chain carbons (¹³C nuclei) provide complete atomic level access to the protein fold. Any given NMR-active nucleus will have a Larmor or resonant frequency (ω_0) dependent on contributions from its local magnetic field (B), the external magnetic field (B_0), and local electrochemical environment

(B_i), where $B=B_0+B_i$. These resonant frequencies are referenced to a standard frequency in order to quantify the chemical shift (δ , in parts-per-million or Hertz). These chemical shifts are highly sensitive to the local electrochemical environment, thus small changes in protein structure can be measured with precision^(13, 14). Conformational exchange of particular interest in allosteric enzymes, as environments representing the active (R) and inactive (T) states can be elucidated⁽¹⁴⁾. Chemical exchange can be categorized into three general regimes: slow, intermediate, and fast, dependent on whether the rate of exchange (k_{ex}) is slower than, approximately equal to, or faster than the difference in chemical shift ($\Delta\omega$) between resonances in disparate chemical environments^(13, 15). Thus, changes in chemical shifts in a given allosteric system may approximate a shift in the inactive to active equilibrium.

The most fundamental protein NMR experiment is the ^1H - ^{15}N HSQC (heteronuclear single quantum coherence), a two-dimensional experiment which couples the amide N-H pairs within the protein backbone. As such, each peak resolved within the spectrum corresponds to a single amino acid. The HSQC spectrum is often referred to as the “fingerprint” spectrum, as peak position or resonance frequency can provide information about distinct chemical environments “felt” by individual residues. Three-dimensional NMR experiments are used to assign each amino acid in the HSQC fingerprint accurately by linking each amide nitrogen, alpha carbon, beta carbon, and carbonyl carbon within the protein sequence to the ones preceding and following it. As such, exogenous changes to the protein chemical environment can result in intensity (line broadening) or positional (chemical shift perturbation, CSP) changes within the HSQC spectrum. Additionally, transverse relaxation optimized spectroscopy-heteronuclear single quantum coherence (TROSY-HSQC) can be utilized for larger protein systems susceptible to line broadening through fast

transverse relaxation. This optimized HSQC selectively detects elements of the signal with the slowest relaxation, resulting in a higher resolution spectrum.

NMR probes structural and dynamic fluctuations across a wide range of biologically relevant timescales, and a suite of relaxation experiments have been designed to probe each regime^(5, 14, 15). Of particular interest to the elucidation of allosteric pathways are motions on the μs - ms time regime, although this trend is not unyielding, as contributions of fast (ps - ns) timescale motions to allostery, driven by entropic modulation, have been observed in the literature^(16, 17), and within the course of this thesis research.

Larger conformational rearrangements typically associated with allosteric transitions occur on the μs - ms timescale and are often commensurate with rates of enzymatic catalysis⁽¹⁸⁾. In this case, transitions between inactive and active states must cross significant energy barriers resulting in the broadening of NMR resonances and increasing the transverse relaxation rate. These phenomena are best characterized by Carr-Purcell-Meiboom-Gill (CPMG) or $R_{1\rho}$ NMR relaxation dispersion experiments that quantify chemical exchange between the ground (major) state and sparsely populated “invisible” states^(19, 20). In practice, these experiments have been widely employed to investigate allosteric mechanisms in kinases, DNA binding proteins, and the family of RNA-guided DNA endonucleases known as CRISPR-Cas9⁽²¹⁻²⁶⁾, among many other systems.

Historically, slower timescales of motion are coupled with allosteric mechanisms due to the direct link between allosteric alteration of enzyme conformations, ligand binding, and the rate of catalysis. However, advances in experimental structural and dynamic approaches have provided greater evidence for “dynamic allostery.”. This ensemble view treats allosteric activation as the population-weighted sum of many conformational states, rather than the behavior of a single “active” structure⁽³⁾. Effector binding stabilizes only some states but redistributes the entire

ensemble, so each populated conformation, including functionally silent ones, contributes to how strongly allosteric signals modulate activity. As such, contributions from *ps-ns* (low energy barrier) motions play crucial roles in modulating the thermodynamic equilibrium of allosteric enzymes, thus affecting catalytic activity.

Fluctuations in bond vectors of individual amino acids and changes in chemical shift anisotropy or dipolar interactions between nuclei can be quantified by measuring motions on the *ps-ns* timescale^(13, 27). These are associated with entropically driven allostery and have been described for numerous systems^(28, 29). Generally, these motions are an important component of the allosterically altered free energy landscape that regulates active and inactive enzymatic populations⁽¹³⁾. Protein backbone and sidechain flexibility can therefore be probed through the lens of a changing energy landscape with longitudinal (T_1) and transverse (T_2) spin relaxation experiments, coupled with the heteronuclear ^1H - ^{15}N Nuclear Overhauser Effect (NOE). The implementation and data analysis derived from these experiments are discussed at length in chapters 2 & 3.

CRISPR-Cas9 is an RNA-guided Endonuclease and Revolutionary Biological Editing Tool

Clustered Regularly Interspaced Short Palindromic Repeat-associated (CRISPR-Cas) proteins have evolved as the bacterial adaptive immune system, targeting and degrading invasive phage DNA. The most extensively studied CRISPR enzyme, CRISPR-Cas9 (Cas9) from *S. pyogenes* (*SpyCas9*) has since been co-opted as one of the most revolutionary gene editing tools in history, with novel applications, mechanisms, and Cas effectors still being discovered today^(30, 31). The Cas9 enzyme complexes with a single-guide RNA (sgRNA) which supplies a highly specific sequence for the targeting and cleavage of complementary DNA. Cas9 first recognizes a

protospacer adjacent motif (PAM) in the DNA, a short sequence of nucleotides that precede the cleavage site. Upon sgRNA and target DNA strand (TS) complementarity, an RNA:DNA hybrid is formed while displacing the non-target strand (NTS)^(30, 32-34). The Cas9 enzyme is comprised of a lobe containing its nucleic acid recognition (REC) and PAM-interacting (PI) domains, which are connected to the two catalytic nuclease domains HNH and RuvC by a small “Wedge” domain (WED). The cooperation of each domain within the greater Cas9 architecture generates double stranded cuts within DNA complementary to its sgRNA. These cuts are catalyzed by the HNH and RuvC nuclease domains that cleave the target and non-target strands, respectively. The REC domain is responsible for orienting the RNA:DNA hybrid and allosteric activation of the nuclease domains. The WED domain functions in sgRNA positioning, while the PI domain recognizes PAM sequences on target DNA, often functioning as the first filter of potential DNA targets^(30, 32-34).

Class-II Cas enzymes are generally monomeric species composed of multiple functional domains. While the Class-II, Type-II *Spy*Cas9 is the most established Cas effector, years-long efforts to expand DNA target recognition and biological applications have introduced *Geobacillus stearothermophilus* (*Geo*Cas9), *Staphylococcus aureus* (*Sa*Cas9), *Neisseria meningitidis* (*Nme*Cas9) Cas9 orthologs^(32, 35, 36). The canonical *Spy*Cas9 utilizes a ~100 nucleotide fusion of trans-activating CRISPR RNA (tracrRNA) and CRISPR RNA (crRNA) as an engineered sgRNA⁽³⁷⁾.

*Spy*Cas9 sgRNA is a fusion of a 5'-DNA targeting sequence (endogenously known as the spacer region or in engineered systems, crRNA) and a 3'-scaffold sequence (tracrRNA) that promotes *Spy*Cas9 uptake. Fusion between crRNA and tracrRNAs requires complementary binding between the crRNA repeat and tracrRNA anti-repeat regions, producing the core stem loop binding scaffold for *Spy*Cas9⁽³⁹⁾ (**Figure 2A**). Isothermal titration calorimetry (ITC) experiments,

which quantitatively describe the thermodynamics of protein-ligand interactions⁽⁴⁰⁾, show sgRNA binding by *Spy*Cas9 to be enthalpically driven by $\Delta H \sim -400$ kcal/mol, resulting in a $\Delta G \sim -8$ kcal/mol⁽⁴¹⁾ (ΔH and ΔG are the change in enthalpy and free energy, respectively)⁽⁴⁰⁾. Formation of the *Spy*Cas9-ribonucleoprotein complex (RNP) causes the rearrangement of the REC domain that opens the lobe to promote a catalytically competent structure⁽⁴²⁾. This catalytically active state initiates target DNA interactions, where DNA binding occurs between the NGG PAM site on the non-target strand (NTS) and the *Spy*Cas9 PI domain⁽⁴³⁾. PAM binding triggers the initial unwinding of the DNA, while the WED domain flanking the DNA TS and targeting region of the sgRNA further unwinds the strands^(44, 45). R-loop formation is initiated by the hybridization of the DNA TS with the first few PAM-proximal bases of the sgRNA, the seed region, resulting in a 10-12 base pair (bp) R-loop adjacent to the PAM site. As such, the R-loop is structure of hybridized RNA:DNA complex forming a helix and single-stranded DNA bubble^(46, 47). R-loop formation is the rate-limiting step of *Spy*Cas9-mediated DNA cleavage and its formation allosterically stabilizes the Cas9:sgRNA:DNA complex in a catalytically active state^(46, 47).

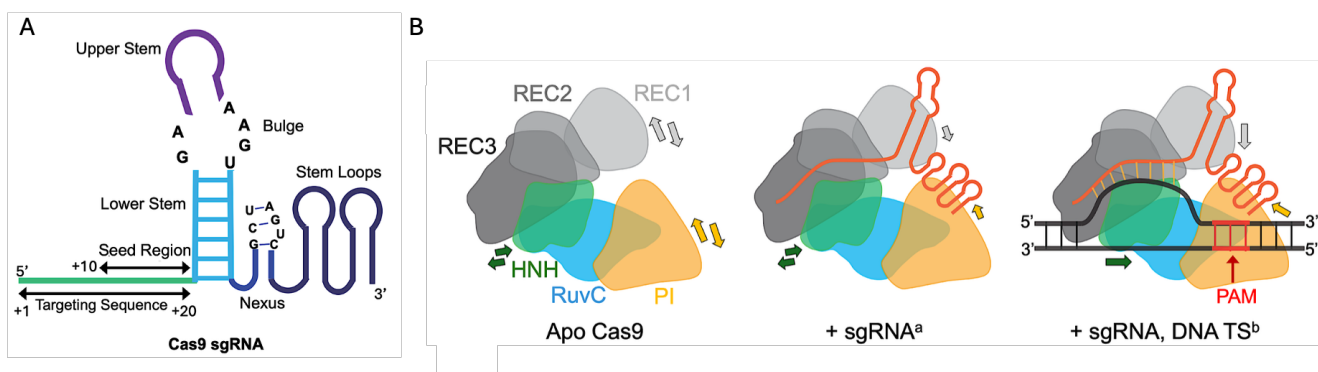


Figure 2. (A) Architectures and critical structural elements of Cas9 guide RNAs. (B) A dynamic apo structure is partially closed by sgRNA ($\Delta G = -7.7$ kcal/mol at 25 °C), where HNH can sample an active conformation. DNA binding further stabilizes the structure and twisting of HNH primes both nucleases for cleavage.

SpyCas9 undergoes conformational fluctuations between inactive, pseudo-active (sometimes called “surveillance”)⁽⁴⁸⁾, and active states that involve multiple domains of the Cas protein traversing large atomic distances. Each state has a unique energetic landscape established by a multitude of studies with the goal of leveraging thermodynamic drivers of *SpyCas9* domain dynamics to “fine-tune” activity and specificity^(41, 49-51). In its unliganded state, *SpyCas9* adopts an auto-inhibited conformation that shifts dramatically, primarily via REC, upon sgRNA binding. The *SpyCas9*-RNP will transiently bind DNA substrates through random collision, as combined biochemical and single-molecule fluorescence assays have observed *SpyCas9* to utilize both one-dimensional (1D, lateral diffusion on DNA substrate) and three-dimensional (3D, randomized collision) diffusion to do so⁽⁵²⁾. These transient interactions are stabilized by the PAM binding site residues, and mutating the charged lysines to alanines has been shown to specifically disrupt 1D lateral diffusion, slowing the target search rate. Further, *SpyCas9*-RNP search time in live *E. coli* is not rapid (up to 6 hours) due to its interactions with several PAM and off-target sites in the search for the correct sequence, as well as a greater endogenous availability of partially complementary target sequences⁽⁵²⁾.

Upon identification of a complementary target DNA duplex, *SpyCas9* again undergoes a large conformational change from the pseudo-active to active state. Cryo-EM along with enhanced sampling molecular dynamics (a method to accelerate conformational sampling such that rare events with high energy barriers can be captured) have captured this transition involving the translocation of the HNH domain, facilitated by its allosteric communication with REC^(25, 53-55). The step-by-step dynamics of *SpyCas9* activation have been described structurally, often with DNA-bound structures of *SpyCas9* displaying two states of the HNH nuclease domain and MD simulations capturing HNH rotating 180° toward the DNA TS with respect to the RNA:DNA

hybrid⁽⁴²⁾. The L1/L2 loops (*i.e.* hinges) connecting HNH to the adjacent RuvC nuclease domain also rearrange to establish critical hydrogen bonds with the RNA:DNA hybrid⁽⁴²⁾. Following this step, the catalytic H840 within HNH is oriented toward the scissile phosphate of the DNA TS in the active state. The intrinsic flexibility of the HNH domain keeps the energy barrier of the structural transition between pre-catalytic and catalytic states quite low (~1-2 kcal/mol), while the overall energies of all conformers, depending on the point-of-reference, differ by as little as ~8 kcal/mol. The ability to maintain a balance of energetically similar active and inactive conformers within a 1,400 residue enzyme highlights the exquisite coordination of domain motions within Cas effectors and dependence on allosteric signaling pathways.

Cas9 Mismatch Tolerance can lead to Deleterious Off-target Effects

The mismatch tolerance of Cas effectors has been widely studied, as has its relationship to rates of DNA cleavage. Indeed, the initial report of *SpyCas9* activity showed that mismatches between the sgRNA and DNA proximal to the PAM drastically diminished cleavage, while PAM-distal mismatches were at least partially tolerated⁽⁴⁹⁾. This has also been seen in Cas9s from other organisms⁽⁵⁶⁾. More recently, this effect was quantified using PAM-distal mismatches of varying lengths while measuring *SpyCas9* cleavage rates⁽⁵⁷⁾. As little as 3 base pair mismatches occurring 18-20 nucleotides distal from the PAM reduce the *SpyCas9* cleavage rate 40-fold, while the same 3 base pair mismatch 6-8 nucleotides from the PAM causes a 2,000-fold rate reduction⁽⁵⁷⁾. The reduced cleavage rates have been attributed to a disruption of allosteric control over the nuclease conformation by the PAM-distal end of the RNA:DNA heteroduplex⁽⁵⁷⁾. Recently published cryo-EM structures confirm that off-target substrates can retain the integrity of the RNA:DNA heteroduplex past the conformational checkpoint of nuclease activation⁽⁴⁶⁾, explaining why the

mismatches are tolerated during cleavage, albeit at a reduced rate. The basis for the observed reduction has been attributed to a significant population of *SpyCas9* being locked into the catalytically incompetent structure from early R-loop formation in the presence of mismatched substrates⁽⁴⁶⁾. The hypothesis that domain immobilization leads to reductions in catalytic activity suggests that intrinsic dynamics serve as a specificity checkpoint for *SpyCas9* and should therefore play an integral role in efforts to engineer specificity enhancement⁽⁴⁸⁾. The mismatch tolerance of *SpyCas9* remains a substantial barrier to the development of highly accurate, high specificity editors. Although many enhanced specificity variants exist today, they are still plagued by reduced overall activity due to factors described above. As such, studies of the allosteric mechanism of *SpyCas9* and other orthologs are vital to overcoming these barriers, as tuning intra- and interdomain communication within Cas9 may alter mismatch tolerance while maintaining catalytic activity.

CRISPR-Cas9 Activity is Dependent on Allosteric Signaling between Spatially Distant Domains

As Cas9 domains are functionally coupled yet spatially distant, allosteric signaling through the protein and dynamic rearrangements are crucial to efficient and accurate enzymatic function. Allostery within Cas9 was originally probed through the insertion of exogenous PDZ domains, small structural domains that act as scaffolding proteins⁽⁵⁸⁾. The insertion of these domains highlighted the existence of allosteric “hotspots” within Cas9. Computational and biochemical studies have likewise indicated that allosteric pathways within Cas9 synchronize DNA recognition, binding, and catalytic activity^(9, 42, 59, 60). Previous biophysical work has shown that the HNH domain serves as the core of allosteric relay from the REC domains, and that intrinsic dynamics within HNH are responsible for signal transduction, and by extension, conformational

regulation of double-stranded DNA cleavage^(25, 54). Additionally, studies have shown how REC binding to off-target DNA sequences alter HNH dynamics and Cas9 specificity, highlighting both the REC and HNH as targets for mutagenesis in the development of enhanced specificity Cas9 variants^(26, 53, 60, 61). Although our understanding of the allosteric communication within *SpyCas9* has led to the development of enhanced specificity variants, it is necessary to continue to study divergent Cas9 orthologs if we hope to expand what CRISPR can do biologically, technically, and therapeutically.

Thermophilic Cas9 Provides a Promising Alternative to SpyCas9

Although a multitude of Cas9 effectors exist today, the majority of innovative structural and biochemical studies have centered around *SpyCas9*. However, as a human therapeutic, *SpyCas9* maintains narrow potential due to its instability in human plasma⁽³⁶⁾. Recently, a Cas9 from *Geobacillus stearothermophilus* (*GeoCas9*) (**Figure 3**) was found to be stable and functional in human plasma while remaining enzymatically active at temperatures exceeding 70 °C⁽³⁶⁾. The heightened stability and increased cleavage temperature range (15 – 75 °C for *GeoCas9* vs 35 – 45 °C for *SpyCas9*) of the thermophilic *GeoCas9* offers an increased stability and longevity in cells⁽³⁶⁾. *GeoCas9* exhibits a smaller size (<1100 amino acids) compared to *SpyCas9* (1368 a.a.) which makes it a better candidate for *in vivo* delivery by viral vectors or other vesicle delivery systems⁽³⁶⁾. *GeoCas9* recognizes a relatively long and specific PAM sequence (5'-NNNNCRAA-3') in comparison to *SpyCas9* (5'-NGG-3') with a preference for AT genomes, which confers a natural enhancement of specificity and the expansion of target genomes⁽³⁶⁾. These findings indicate *GeoCas9* to be a promising alternative to *SpyCas9*, with the added potential as a therapeutic editor in human disease. Furthermore, there is little known about mechanisms of thermophilic Cas9

homologs that can offer unique advantages to genome editing. As such, it is crucial to understand the allosteric mechanism governing *GeoCas9* in order to establish hotspots for functional control and develop enhanced variants with high genome editing proficiency.

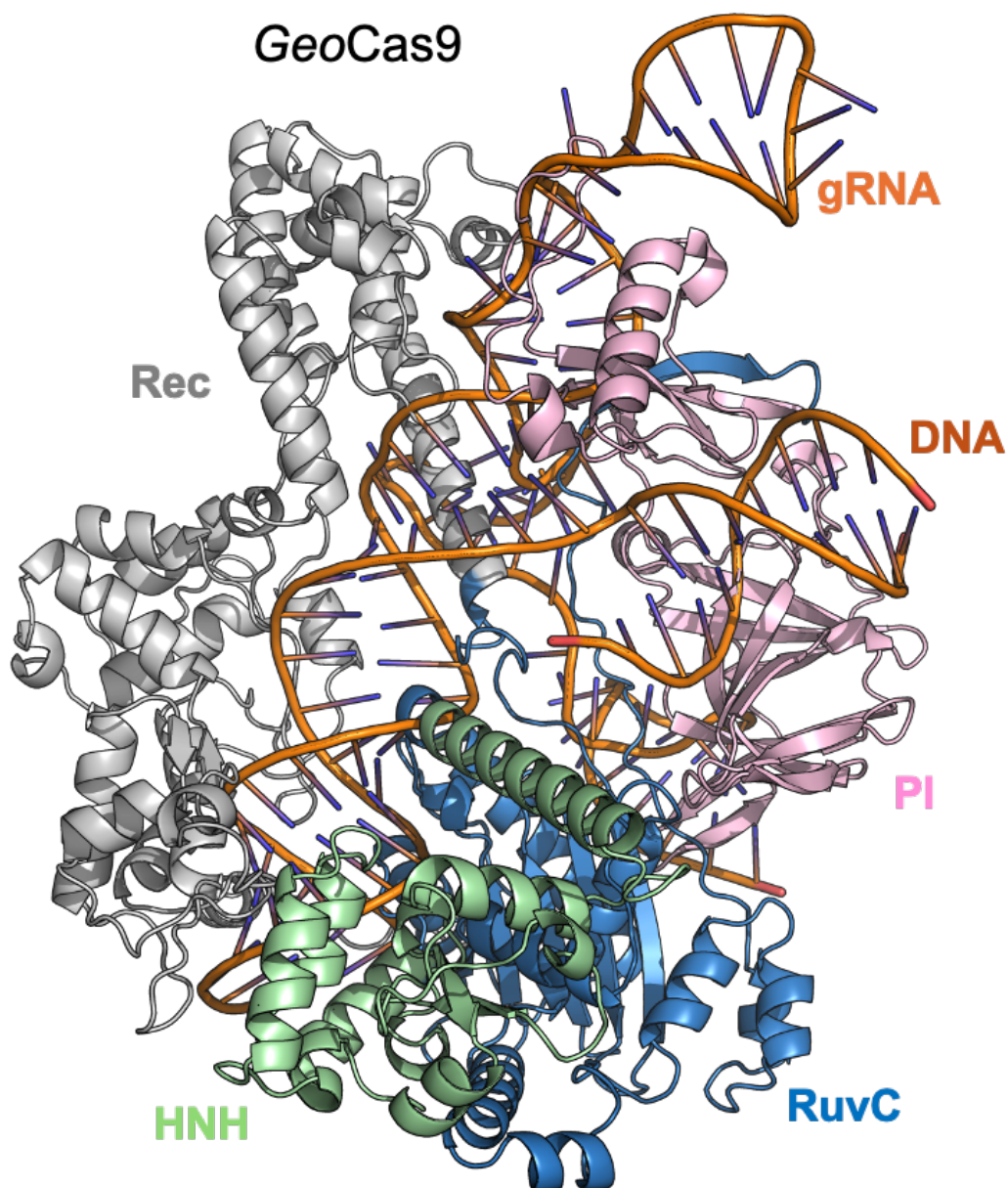


Figure 3. AlphaFold* model of *GeoCas9* showing domain arrangements complexed with both guide RNA and DNA. *A Cryo-EM structure of *GeoCas9* has been solved and deposited to the PDB, (PDB ID: 8UZA) however the *GeoHNH* domain remains completely unresolved in the experimental structure. As this thesis is largely centered around the *GeoHNH* domain, the AlphaFold model is utilized to show relative positioning of the domains. Overlays between the experimental structure and AlphaFold model show remarkable agreement.

Initial biophysical experiments pioneered by our lab showed that although structurally homologous, the HNH domains of *Spy*- And *Geo*Cas9 were divergent in their intrinsic dynamics^(62, 63). The μ s-ms motions that govern allosteric communication within *Spy*Cas9 are absent from *Geo*HNH, while faster ps-ns motions are dominant, highlighting conformational and flexibility differences between Cas9 species that govern enzymatic activity^(25, 63). This thesis seeks to further probe the allosteric differences between *Geo*Cas9 and *Spy*Cas9, and how methods of controlling editing and off-target effects can be leveraged with *Geo*Cas9 to advance the potential of this thermophilic Cas9 as an effective genome editor.

Anti-CRISPR Inhibitors Offer Spatial and Temporal Control of Cas9 Editing

Control of Cas9 editing and mediation of off-target effects has primarily centered on mutagenesis to alter allosteric communication and protein-protein or protein-nucleic acid contacts within the enzyme. The most widely used high-fidelity *Spy*Cas9 variants implemented mutagenesis to destabilize non-specific DNA/Rec interactions or to tighten the conformational proofreading of HNH by altering its allosteric communication with Rec^(26, 50, 51, 53, 61). As such, one goal of this thesis will indeed be to explore analogous mutations within *Geo*Cas9's REC domain to assess whether a thermophilic Cas9 with altered intrinsic flexibility can benefit by similar mutagenesis strategies.

Additionally, there has been growing interest in leveraging exogenous factors to control Cas9 editing, namely small protein inhibitors called Anti-CRISPRs (Acrs)⁽⁶⁴⁾ that evolved in viruses and phage to escape CRISPR-mediated destruction⁽⁶⁵⁾. Acrs are small polypeptides that are classed based on the type of Cas system they target (Acr1 for Class-I, etc..)⁽⁶⁵⁾. Thus far, over 40 different Acrs have been discovered with novel inhibitors still being discovered and categorized⁽⁶⁴⁾.

Mechanisms of Acr inhibition are numerous and diverse, with some Acr proteins broadly inhibiting multiple Cas orthologs. However, the inhibition mechanism of Acrs can be categorized into four main subgroups: (1) direct binding and steric occlusion of functional sites, (2) dynamic remodeling of Cas effectors through nucleic acid mimicry, (3) allosteric inhibition of Cas functional sites, and (4) enzymatic degradation of CRISPR-Cas protein components⁽⁶⁴⁻⁶⁶⁾.

Just as CRISPR-Cas technology has been adapted for use in bioengineering, anti-CRISPR inhibitors provide an opportunity to leverage an evolved mechanism for spatiotemporal control of Cas systems⁽⁶⁴⁾. These small proteins provide regulatory promise without the need of specificity enhancing mutations or fusion proteins within the Cas itself, which demand irreversible modification of the enzyme. Acrs can effectively act as an on/off switch, limiting off-target effects or fine tuning the duration of Cas activity⁽⁶⁴⁾. Anti-CRISPRs characterized with *Spy*Cas9 and other orthologs have already shown promise. For example, AcrIIA4 has been used experimentally for CRISPR-based gene regulation circuits, where the transcription of genes can be halted by the addition of this Acr in mammalian cells⁽⁶⁷⁾. Other broad applications of Acrs include antibacterials⁽⁶⁸⁾, modulation of gene expression⁽⁶⁹⁾, enhancing microbial gene-editing strategies by enhancing transformation efficiency⁽⁷⁰⁾, reducing off-target effects while maintaining catalytic activity⁽⁷¹⁾, and as biosensors via dCas9s⁽⁷²⁾. Furthermore, the relatively small size of Acrs (~50-200 amino acids) makes them amenable to adeno-associated viral vectors, a common delivery strategy to target cells in specific tissues⁽⁶⁴⁾. Beyond the general categories of Acrs from a functional perspective, there remains a lack of structural and biophysical mechanistic understanding. Furthermore, there is a lack of interrogation into whether “broad-spectrum” inhibitors act similarly, from a molecular standpoint, in disparate Cas effectors.

A recent study profiled the inhibitory activity of a wide range of Acrs toward phylogenetically distant Cas enzymes⁽⁷³⁾. A type II inhibitor called AcrIIC1 was found to suppress double stranded DNA cleavage by *GeoCas9*, as well as close structural orthologs. This “broad spectrum” inhibitor binds the HNH domain at conserved active site residues, effectively “blocking” catalytic activity. However, the specific molecular details were established with the *NmeCas9*, and no atomistic information exists for AcrIIC1 with *GeoCas9*. Furthermore, it is unclear if even a broad spectrum inhibitor acts identically upon distinct Cas systems. Therefore, this thesis investigates the inhibition mechanism of AcrIIC1 on *GeoCas9* using structural and biophysical strategies to understand AcrIIC1-*GeoCas9* interactions. This work informs strategies for refining the inhibitory targeting of *GeoCas9* and broadens the molecular understanding of anti-CRISPR mechanisms.

References

1. Liu J, Nussinov R. Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLOS Computational Biology*. 2016;12(6):e1004966.
2. Changeux J-P. The concept of allosteric modulation: an overview. *Drug Discovery Today: Technologies*. 2013;10(2):e223–e8.
3. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature*. 2014;508(7496):331–9.
4. Laskowski RA, Gerick F, Thornton JM. The structural basis of allosteric regulation in proteins. *FEBS Letters*. 2009;583(11):1692–8.
5. East KW, Skeens E, Cui JY, Belato HB, Mitchell B, Hsu R, et al. NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys Rev*. 2020;12(1):155–74.
6. Changeux J-P. 50 years of allosteric interactions: the twists and turns of the models. *Nature Reviews Molecular Cell Biology*. 2013;14(12):819–29.
7. Nussinov R, Tsai C-J. Allostery in Disease and in Drug Discovery. *Cell*. 2013;153(2):293–305.
8. Sekhar A, Kay LE. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proceedings of the National Academy of Sciences*. 2013;110(32):12867–74.
9. Palermo G, Ricci CG, Fernando A, Basak R, Jinek M, Rivalta I, et al. Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *Journal of the American Chemical Society*. 2017;139(45):16028–31.

10. Adhireksan Z, Palermo G, Riedel T, Ma Z, Muhammad R, Rothlisberger U, et al. Allosteric cross-talk in chromatin can mediate drug-drug synergy. *Nature Communications*. 2017;8(1):14860.
11. Ricci CG, Silveira RL, Rivalta I, Batista VS, Skaf MS. Allosteric Pathways in the PPAR γ -RXR α nuclear receptor complex. *Scientific Reports*. 2016;6(1):19940.
12. Hu Y, Cheng K, He L, Zhang X, Jiang B, Jiang L, et al. NMR-Based Methods for Protein Analysis. *Analytical Chemistry*. 2021;93(4):1866–79.
13. East KW, Skeens E, Cui JY, Belato HB, Mitchell B, Hsu R, et al. NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophysical Reviews*. 2020;12(1):155–74.
14. Lisi GP, Loria JP. Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chemical Reviews*. 2016;116(11):6323–69.
15. Chao F-A, Byrd RA. Protein dynamics revealed by NMR relaxation methods. *Emerging Topics in Life Sciences*. 2018;2(1):93–105.
16. McDonald LR, Whitley MJ, Boyer JA, Lee AL. Colocalization of Fast and Slow Timescale Dynamics in the Allosteric Signaling Protein CheY. *Journal of Molecular Biology*. 2013;425(13):2372–81.
17. Tzeng S-R, Kalodimos CG. The role of slow and fast protein motions in allosteric interactions. *Biophysical Reviews*. 2015;7(2):251–5.
18. Whittier SK, Hengge AC, Loria JP. Conformational Motions Regulate Phosphoryl Transfer in Related Protein Tyrosine Phosphatases. *Science*. 2013;341(6148):899–903.

19. Loria JP, Rance M, Palmer AG. A Relaxation-Compensated Carr–Purcell–Meiboom–Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *Journal of the American Chemical Society*. 1999;121(10):2331–2.
20. Massi F, Johnson E, Wang C, Rance M, Palmer AG. NMR ρ Rotating-Frame Relaxation with Weak Radio Frequency Fields. *Journal of the American Chemical Society*. 2004;126(7):2247–56.
21. Akke M. NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis. *Current Opinion in Structural Biology*. 2002;12(5):642–7.
22. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS, et al. Dynamics connect substrate recognition to catalysis in protein kinase A. *Nature Chemical Biology*. 2010;6(11):821–8.
23. Capdevila DA, Braymer JJ, Edmonds KA, Wu H, Giedroc DP. Entropy redistribution controls allostery in a metalloregulatory protein. *Proceedings of the National Academy of Sciences*. 2017;114(17):4424–9.
24. Vise PD. NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. *Nucleic Acids Research*. 2005;33(7):2061–77.
25. East KW, Newton JC, Morzan UN, Narkhede YB, Acharya A, Skeens E, et al. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J Am Chem Soc*. 2020;142(3):1348–58.
26. Nierzwicki L, East KW, Morzan UN, Arantes PR, Batista VS, Lisi GP, et al. Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *eLife*. 2021;10.

27. Sapienza PJ, Lee AL. Using NMR to study fast dynamics in proteins: methods and applications. *Current Opinion in Pharmacology*. 2010;10(6):723–30.
28. Wand AJ. Bringing disorder and dynamics in protein allostery into focus. *Proceedings of the National Academy of Sciences*. 2017;114(17):4278–80.
29. Rivalta I, Lisi GP, Snoeberger N-S, Manley G, Loria JP, Batista VS. Allosteric Communication Disrupted by a Small Molecule Binding to the Imidazole Glycerol Phosphate Synthase Protein–Protein Interface. *Biochemistry*. 2016;55(47):6484–94.
30. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012;337(6096):816–21.
31. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346(6213):1258096.
32. Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*. 2017;46(1):505–29.
33. Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, et al. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*. 2014;343(6176):1247997–.
34. Wang J, Arantes PR, Ahsan M, Sinha S, Kyro GW, Maschietto F, et al. Twisting and swiveling domain motions in Cas9 to recognize target DNA duplexes, make double-strand breaks, and release cleaved duplexes. *Frontiers in Molecular Biosciences*. 2023;9.
35. Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. *Annu Rev Biophys*. 2017;46:505–29.

36. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*. 2017;8(1).
37. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*. 2014;346(6213):1258096.
38. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015;163(3):759–71.
39. Jiang M, Ye Y, Li J. Core Hairpin Structure of SpCas9 sgRNA Functions in a Sequence- and Spatial Conformation-Dependent Manner. *SLAS Technol*. 2021;26(1):92–102.
40. Robertson AD, Murphy KP. Protein Structure and the Energetics of Protein Stability. *Chem Rev*. 1997;97(5):1251–68.
41. Baranova SV, Zhdanova PV, Golyshev VM, Lomzov AA, Pestryakov PE, Chernonosov AA, et al. Thermodynamic parameters obtained for the formation of the Cas12a-RNA/DNA complex. *Biochem Biophys Res Commun*. 2025;743:151176.
42. Palermo G, Miao Y, Walker RC, Jinek M, McCammon JA. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc Natl Acad Sci U S A*. 2017;114(28):7260–5.
43. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014;513(7519):569–73.
44. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*. 2014;507(7490):62–7.

45. Eggers AR, Chen K, Soczek KM, Tuck OT, Doherty EE, Xu B, et al. Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *Cell*. 2024;187(13):3249–61 e14.
46. Pacesa M, Lin CH, Clery A, Saha A, Arantes PR, Bargsten K, et al. Structural basis for Cas9 off-target activity. *Cell*. 2022;185(22):4067–81 e21.
47. Liu MS, Gong S, Yu HH, Jung K, Johnson KA, Taylor DW. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat Commun*. 2020;11(1):3576.
48. De Paula VS, Dubey A, Arthanari H, Sgourakis NG. Dynamic sampling of a surveillance state enables DNA proofreading by Cas9. *Cell Chem Biol*. 2025;32(2):267–79 e5.
49. Baranova SV, Zhdanova PV, Pestryakov PE, Chernonosov AA, Koval VV. Key thermodynamic characteristics of Cas9 and Cas12a endonucleases' cleavage of a DNA substrate containing a nucleotide mismatch in the region complementary to RNA. *Biochem Biophys Res Commun*. 2025;768:151892.
50. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016;529(7587):490–5.
51. Schmid-Burgk JL, Gao L, Li D, Gardner Z, Strecker J, Lash B, et al. Highly Parallel Profiling of Cas9 Variant Specificity. *Mol Cell*. 2020;78(4):794–800 e8.
52. Yang M, Sun R, Deng P, Yang Y, Wang W, Liu JG, et al. Nonspecific interactions between SpCas9 and dsDNA sites located downstream of the PAM mediate facilitated diffusion to accelerate target search. *Chem Sci*. 2021;12(38):12776–84.

53. Skeens E, Sinha S, Ahsan M, D'Ordine AM, Jogl G, Palermo G, et al. High-fidelity, hyper-accurate, and evolved mutants rewire atomic-level communication in CRISPR-Cas9. *Sci Adv.* 2024;10(10):eadl1045.
54. Wang J, Maschietto F, Qiu T, Arantes PR, Skeens E, Palermo G, et al. Substrate-independent activation pathways of the CRISPR-Cas9 HNH nuclease. *Biophys J.* 2023;122(24):4635–44.
55. Wang J, Arantes PR, Ahsan M, Sinha S, Kyro GW, Maschietto F, et al. Twisting and swiveling domain motions in Cas9 to recognize target DNA duplexes, make double-strand breaks, and release cleaved duplexes. *Front Mol Biosci.* 2022;9:1072733.
56. Belato HB, Knight AL, D'Ordine AM, Pindi C, Fan Z, Luo J, et al. Structural and dynamic impacts of single-atom disruptions to guide RNA interactions within the recognition lobe of *Geobacillus stearothermophilus* Cas9. *Elife.* 2025;13.
57. Bravo JPK, Liu MS, Hibshman GN, Dangerfield TL, Jung K, McCool RS, et al. Publisher Correction: Structural basis for mismatch surveillance by CRISPR-Cas9. *Nature.* 2022;604(7904):E10.
58. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL. Hidden dynamic allostery in a PDZ domain. *Proceedings of the National Academy of Sciences.* 2009;106(43):18249–54.
59. Palermo G. Structure and Dynamics of the CRISPR–Cas9 Catalytic Complex. *Journal of Chemical Information and Modeling.* 2019;59(5):2394–406.
60. Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, et al. Key role of the REC lobe during CRISPR–Cas9 activation by ‘sensing’, ‘regulating’, and ‘locking’ the catalytic HNH domain. *Quarterly Reviews of Biophysics.* 2018;51.

61. Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016;351(6268):84–8.
62. Belato HB, Norbrun C, Luo J, Pindi C, Sinha S, D'Ordine AM, et al. Disruption of electrostatic contacts in the HNH nuclease from a thermophilic Cas9 rewires allosteric motions and enhances high-temperature DNA cleavage. *J Chem Phys*. 2022;157(22):225103.
63. Belato HB, D'Ordine AM, Nierzwicki L, Arantes PR, Jogl G, Palermo G, et al. Structural and dynamic insights into the HNH nuclease of divergent Cas9 species. *Journal of Structural Biology*. 2022;214(1):107814.
64. Marino ND, Pinilla-Redondo R, Csörgő B, Bondy-Denomy J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature Methods*. 2020;17(5):471–9.
65. Belato HB, Lisi GP. The Many (Inter)faces of Anti-CRISPRs: Modulation of CRISPR-Cas Structure and Dynamics by Mechanistically Diverse Inhibitors. *Biomolecules*. 2023;13(2):264.
66. Choudhary N, Tandi D, Verma RK, Yadav VK, Dhingra N, Ghosh T, et al. A comprehensive appraisal of mechanism of anti-CRISPR proteins: an advanced genome editor to amend the CRISPR gene editing. *Frontiers in Plant Science*. 2023;14.
67. Song G, Zhang F, Tian C, Gao X, Zhu X, Fan D, et al. Discovery of potent and versatile CRISPR–Cas9 inhibitors engineered for chemically controllable genome editing. *Nucleic Acids Research*. 2022;50(5):2836–53.
68. Louwen R, Staals RHJ, Endtz HP, Van Baarlen P, Van Der Oost J. The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. *Microbiology and Molecular Biology Reviews*. 2014;78(1):74–88.
69. Adli M. The CRISPR tool kit for genome editing and beyond. *Nature Communications*. 2018;9(1).

70. Mayo-Muñoz D, He F, Jørgensen JB, Madsen PK, Bhoobalan-Chitty Y, Peng X. Anti-CRISPR-Based and CRISPR-Based Genome Editing of *Sulfolobus islandicus* Rod-Shaped Virus 2. *Viruses*. 2018;10(12):695.
71. Shin J, Jiang F, Liu J-J, Bray NL, Rauch BJ, Baik SH, et al. Disabling Cas9 by an anti-CRISPR DNA mimic. *Science Advances*. 2017;3(7):e1701620.
72. Li J, Xu Z, Chupalov A, Marchisio MA. Anti-CRISPR-based biosensors in the yeast *S. cerevisiae*. *Journal of Biological Engineering*. 2018;12(1).
73. Harrington LB, Doxzen KW, Ma E, Liu J-J, Knott GJ, Edraki A, et al. A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell*. 2017;170(6):1224–33.e15.

CHAPTER TWO

Molecular Tuning of the Thermophilic *GeoCas9* Recognition Lobe Modulates Allosteric Motions and Guide RNA Interactions

This chapter is adapted from a published manuscript in eLife (2024) titled:

Structural and dynamic impacts of single-atom disruptions to guide RNA interactions within the recognition lobe of *Geobacillus stearothermophilus* Cas9

*Helen B. Belato^{1#}, *Alexa L. Knight¹, Jinping Luo², Alexandra M. D’Ordine¹, Gerwald Jogl¹, George P. Lisi¹

* Co-first authors

¹Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI USA

²Brown University Transgenic Mouse & Gene Targeting Facility, Providence, RI USA

[#]Present address: Department of Chemistry, Brown University, Providence, RI USA

Author Contributions

ALK: writing – revisions, data curation, formal analysis, produced *GeoRec3*, *GeoRec1*, *GeoRecFL*, and *GeoCas9* proteins and sgRNA, conducted the NMR. **HBB:** writing – original draft, data curation, formal analysis, produced *GeoRec3*, *GeoRec1*, *GeoRecFL*, and *GeoCas9* proteins and sgRNA, conducted the NMR and biophysical experiments. **AMD:** solved X-ray crystal structure of *GeoRec3*. **GJ:** supervision – X-ray crystallography. **JL:** *in vitro* cleavage assay. **GPL:** conceptualization, supervision – NMR spectroscopy, writing – review and editing.

Abstract

The intuitive manipulation of specific amino acids to alter the activity or specificity of CRISPR-Cas9 has been a topic of great interest. As a large multi-domain RNA-guided endonuclease, the intricate molecular crosstalk within the Cas9 protein hinges on its conformational dynamics, but a comprehensive understanding of the extent and timescale of the motions that drive its allosteric function and association with nucleic acids remains elusive. Here, we investigated the structure and multi-timescale molecular motions of the recognition (Rec) lobe of *Geo*Cas9, a thermophilic Cas9 from *Geobacillus stearothermophilus*. Our results provide new atomic details about the *Geo*Rec subdomains (*Geo*Rec1, *Geo*Rec2) and the full-length domain in solution. Two rationally designed mutants, K267E and R332A, enhanced and redistributed micro-millisecond flexibility throughout *Geo*Rec, and NMR studies of the interaction between *Geo*Rec and its guide RNA showed that mutations reduced this affinity and the stability of the ribonucleoprotein complex. Despite measured biophysical differences due to the mutations, DNA cleavage assays reveal no functional differences in on-target activity, and similar specificity. These data suggest that guide RNA interactions can be tuned at the biophysical level in the absence of major functional losses but also raise questions about the underlying mechanism of *Geo*Cas9, since analogous single-point mutations have significantly impacted on- and off-target DNA editing in mesophilic *S. pyogenes* Cas9. A K267E/R332A double mutant also did not enhance *Geo*Cas9 specificity, highlighting the robust tolerance of mutations to the Rec lobe of *Geo*Cas9 and species-dependent complexity of Rec across Cas9 paralogs. Ultimately, this work provides an avenue by which to modulate the structure, motion, and guide RNA interactions at the level of the Rec lobe of *Geo*Cas9, setting the stage for future studies of *Geo*Cas9 variants and their effect on its allosteric mechanism.

Introduction

The vast majority of Cas systems explored as genome editors originate from mesophilic hosts. The emergence of the thermophilic *GeoCas9*, with DNA cleavage function up to 85°C, can expand CRISPR technology to higher temperature regimes and stabilities,^(1,2) but its regulatory mechanism relative to canonical Cas9s must be established. The *SpyCas9*, which originates from the mesophilic *Streptococcus pyogenes*, as well as *GeoCas9*, are both effectors of Type-II CRISPR systems. Interestingly, the Type II-A *SpyCas9* has been by far the most used Cas enzyme, including in ongoing clinical trials.^(3,4) But Cas9 homologs of the Type II-C class, such as *Neisseria meningitis* (*NmeCas9*) and *Campylobacter jejuni* (*CjeCas9*), to which *GeoCas9* belongs, have been validated for mammalian genome editing,^(2,5,6) reinforcing the need to better understand this CRISPR class.

The similar domain arrangements of *GeoCas9* and *SpyCas9* led us to initially speculate that these could share atomic level mechanistic similarities.⁽⁷⁾ *GeoCas9* utilizes a guide RNA (gRNA) to localize and unwind a double-stranded DNA (dsDNA) target after recognition of its 5'NNNNCRAA-3' protospacer adjacent motif (PAM).^(2,8) Upon recognition of the PAM sequence by the PAM-Interacting domain (PI), Cas9-bound guide (gRNA) forms an RNA:DNA hybrid with the target DNA strand. Initially thought to be part of the PI domain⁽²⁾, the wedge (WED) domain recognizes the repeat:anti-repeat region of the gRNA and the dsDNA upstream of the target region.⁽⁹⁾ The Rec lobe of Cas9 is responsible for orienting the RNA:DNA hybrid, as well as the adjacent nuclease domains, into their active conformations.⁽¹⁰⁻¹³⁾ Coordinated cleavage of the target and non-target DNA strand then occurs via the HNH and RuvC nucleases, respectively. The *GeoCas9* nuclease active sites within HNH and RuvC are spatially distinct from the PAM recognition site in the PI domain, necessitating structural and dynamic changes that allosterically

couple dsDNA binding to cleavage. Biochemical^(11,14,15) and structural^(16,17) experiments using the extensively studied *SpyCas9* have revealed that its function is governed by a sophisticated allosteric mechanism that transfers gRNA and dsDNA binding information from the Rec lobe to the distal catalytic sites. A dynamically driven allosteric signal spans the HNH domain of *SpyCas9*, enabled by the plasticity of the Rec lobe, which orchestrates the conformational activation required for DNA cleavage.^(14,18) Our prior work revealed a divergence in the timescales of allosteric motions in the *SpyCas9* and *GeoCas9* HNH domains^(7,17) suggesting an unusually flexible HNH and unique allosteric mode of regulation for *GeoCas9*. It is therefore also possible that docking of the gRNA with *GeoCas9*, and thus its interaction with the RNA:DNA hybrid, may differ from the *SpyCas9* system, as *GeoCas9* contains a truncated Rec lobe with only two of the three canonical subdomains.

The high thermal stability and more compact size of *GeoCas9* (it is 281 residues shorter than *SpyCas9*) can be especially important for *in vivo* delivery applications, since promising viral vectors (*i.e.* adeno-associated virus, AAV) have cargo capacities of ~4.7kb,⁽¹⁹⁾ which prevents *SpyCas9*-gRNA packaging into a single AAV vector but permits “all-in-one” delivery of *GeoCas9*-gRNA.⁽¹²⁾ Until the very recent cryo-EM structures of *GeoCas9*,^(20,21) little was known about specific residues that influence its structure, gRNA binding, or function. Our recent NMR work with *SpyCas9* uncovered pathways of micro-millisecond timescale motions that propagate chemical information related to allostery and specificity through *SpyRec* and its RNA:DNA hybrid,^(16,17) prompting us to investigate this phenomenon in *GeoRec*.

An atomic-level structural understanding of specificity in large multi-domain protein-nucleic acid complexes like Cas9 is often difficult to address by NMR spectroscopy. Although dynamic ensembles in DNA repair enzymes have provided some insight,²² many efforts to improve

Cas9 specificity and reduce off-target activity have relied on large mutational screens⁽²³⁾ or error-prone PCR⁽²⁴⁾, which are less intuitive. Inter-subunit allosteric communication between the catalytic HNH domain and the Rec lobe is critical to Cas9 specificity, as the binding of off-target DNA sequences at Rec alter HNH dynamics to affect DNA cleavage.^(8,25,26) To further probe the fundamental role of protein motions in the function and specificity of *GeoCas9*, as well as the effect of protein-nucleic acid interactions on its structural signatures, we engineered two mutations in *GeoRec* (K267E and R332A, housed within *GeoRec2*). We hypothesized that these variants could enhance *GeoCas9* specificity (*i.e.* limit its off-target cleavage) for two reasons. First, the chosen mutation sites are homologous to those of specificity-enhancing variants of *SpyCas9*.^(24,27) Second, altered Cas9-gRNA interactions have been shown to be a consequence of specificity-enhancement and these charged residues appear to directly interact with the gRNA.^(10,11,15,28) Balancing these two points is the fact that Type-II Cas systems generally have conserved nuclease domains, but are delineated by highly varied Rec lobes.⁽¹²⁾ This implies that the structural and dynamic properties of Rec may play an outsized role in differentiating the functions of *SpyCas9* and *GeoCas9*, which may not be identical. Nevertheless, our work provides new insight into the biophysical, biochemical, and functional role of the *GeoRec* lobe and how mutations modulate the domain itself and its interaction with gRNA in full-length *GeoCas9*.

Results

The structural similarity of GeoRec1, GeoRec2, and GeoRec facilitates NMR analysis of protein dynamics and RNA affinity

GeoCas9 is a 1087 amino acid polypeptide, thus we employed a “divide and conquer” approach for NMR studies, which we previously showed to be useful for quantifying allosteric structure and motion in *SpyCas9*.^(16,17,29,30) The *GeoRec* lobe is comprised of subdomains *GeoRec1* and *GeoRec2*, which likely work together to recognize nucleic acids. We engineered constructs of the *GeoRec1* (136 residues, 16 kDa) and *GeoRec2* (212 residues, 25 kDa) subdomains and solved the X-ray crystal structure of *GeoRec2* at 1.49 Å, which aligns remarkably well with the structure of the *GeoRec2* domain within the AlphaFold model (RMSD 1.03 Å) and new cryo-EM structure of *GeoCas9* (RMSD 1.10 Å, **Figure 1A**). We were neither able to crystallize *GeoRec1* nor full-length *GeoCas9* in the apo state, but our *GeoRec2* crystal structure represents the structure of the subdomain within the full-length *GeoCas9* protein quite well. Our previous studies of *GeoHNH* also show identical superpositions of X-ray crystal structures with full-length Cas complexes.⁷ In addition to the individual subdomains, we also generated an NMR construct of the intact *GeoRec* (370 residues, 43 kDa).

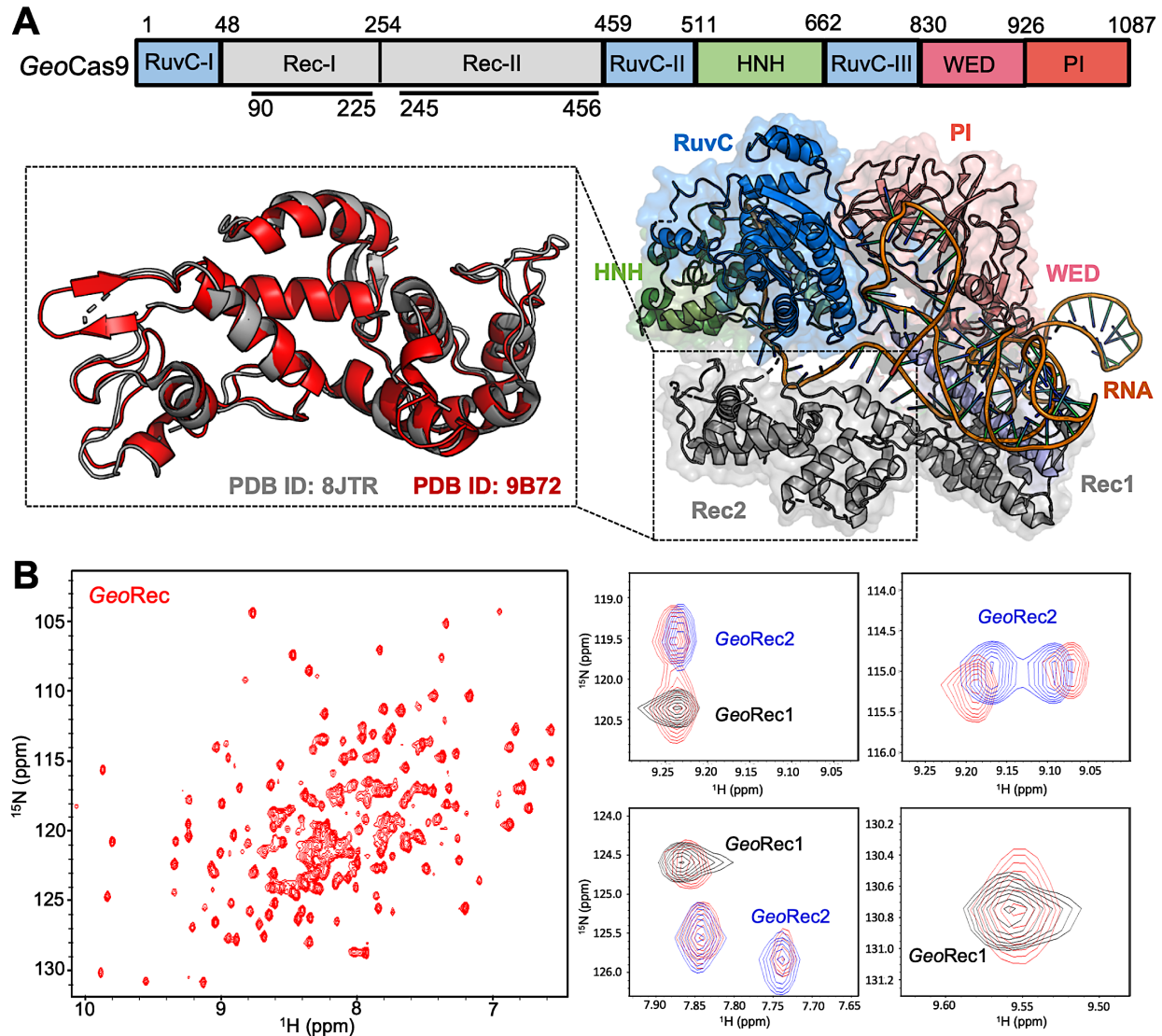


Figure 1. (A) Arrangement of *GeoCas9* domains across the primary sequence. The cryo-EM structure of *GeoCas9* in complex with gRNA (PDB: 8JTR) shows poor resolution of HNH. The *GeoRec2* domain from PDB: 8JTR (gray) is overlaid with our X-ray structure of *GeoRec2* (red, PDB: 9B72). **(B)** $^1\text{H}/^{15}\text{N}$ TROSY HSQC NMR spectrum of *GeoRec* collected at 850 MHz. Overlays of this spectrum with resonances from spectra of *GeoRec1* (black) and *GeoRec2* (blue) demonstrate a structural similarity between the isolated subdomains and intact *GeoRec*.

Despite only 22% sequence identity, the structure of *SpyRec3* and *GeoRec2* are highly similar (RMSD 2.00 Å, **Figure 2**). The structure of *GeoRec1*, in contrast, does not align perfectly with *SpyRec1*, instead, it partially aligns with both *SpyRec1* and *SpyRec2* (**Figure 2**). Thus, the nearly identical *SpyRec3* and *GeoRec2* architectures and their intrinsic dynamics may be a common thread among Type II Cas9s of different size and PAM preference. To capture atomic-level signatures of *GeoRec*, we obtained well-resolved ^1H - ^{15}N NMR fingerprint spectra for all three protein constructs and assigned the amide backbones (**Figure 2**). ^1H - ^{15}N amide and ^1H - $^{13}\text{CH}_3$ Ile, Leu, and Val (ILV)-methyl NMR spectra (**Figure 3**) of *GeoRec* overlay very well with those of its individual subdomains, suggesting that the linkage of subdomains within the full-length *GeoRec* polypeptide does not alter their individual folds. Consistent with this observation, circular dichroism (CD) thermal unfolding profiles of *GeoRec1* ($T_m \sim 34^\circ\text{C}$) and *GeoRec2* ($T_m = 61.50^\circ\text{C}$) are distinct and occur as separate events in the unfolding profile of *GeoRec* (**Figure 4**). The dumbbell shape of *GeoRec*, with its two globular subdomains connected by a short flexible linker, is a likely contributor to these biophysical properties.

Rationally designed GeoRec2 mutants do not substantially impact the GeoRec structure

To understand how the structure and gRNA interactions of *GeoCas9* can be modulated at the level of *GeoRec*, we engineered two charge-altering point mutants in the *GeoRec2* subdomain, K267E and R332A. Based on the AlphaFold2 model of *GeoCas9*, both of these residues are $< 5\text{\AA}$ from the bound RNA:DNA hybrid and were predicted to interface with the nucleic acids directly (**Figure 5A/B**). A new experimental cryo-EM structure of *GeoCas9* confirmed the interaction between K267 and the gRNA but does not report a $< 5\text{\AA}$ interaction of R332 with the gRNA.⁽³¹⁾ The rationale for our designed mutations was also that removal of positive charge would weaken the interactions between *GeoCas9* and the gRNA, affecting K_d via the electrostatics or dynamics

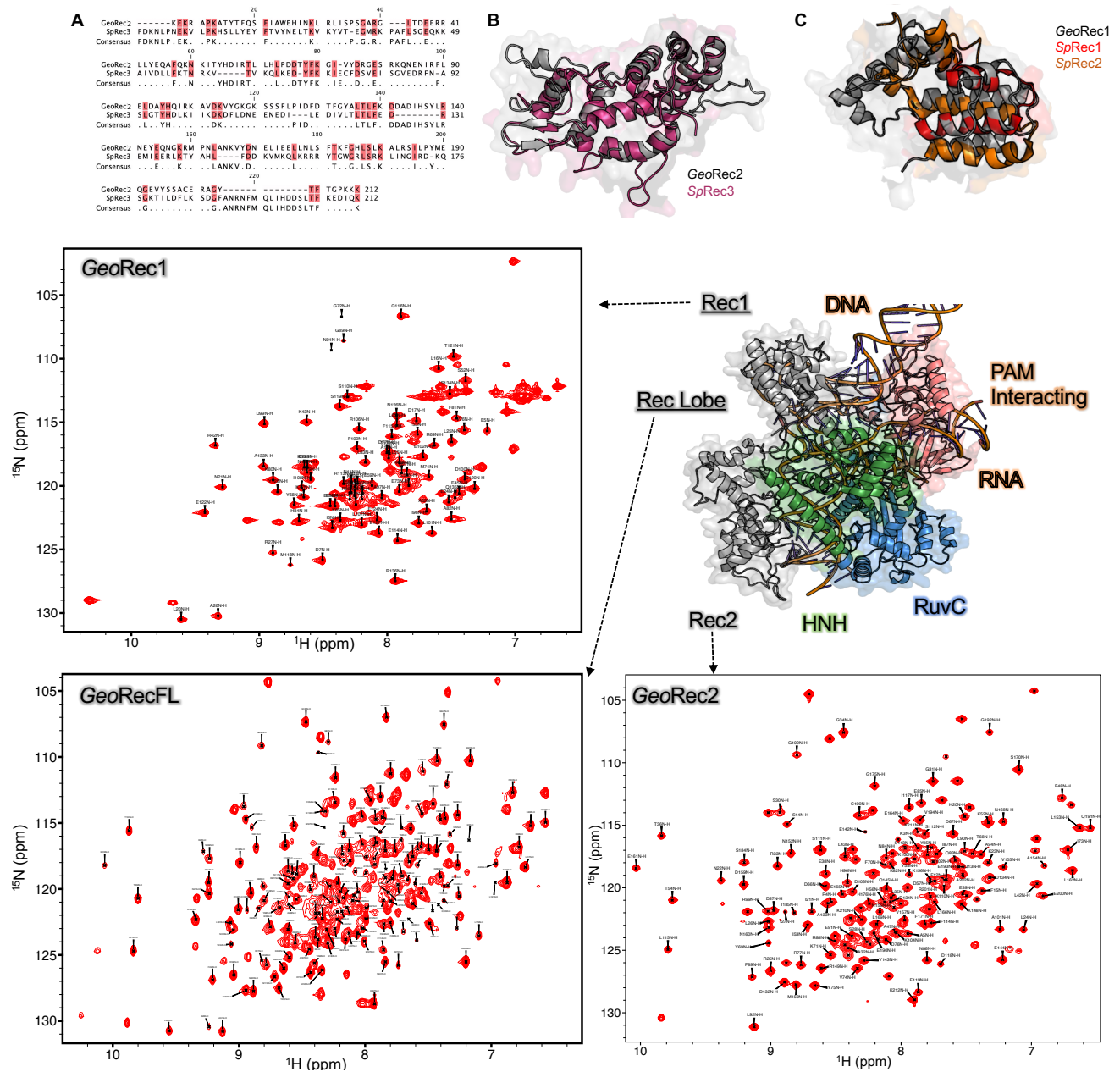


Figure 2. (A) Sequence alignment of *GeoRec2* and *SpyRec3*. Conserved residues are highlighted pink and are listed as consensus. (B) Overlay of the *GeoRec2* X-ray crystal structure (gray) with *SpyRec3* from full-length *SpyCas9* (pink, PDB: 4UN3). (C) Overlay of the *GeoRec1* structure derived from full-length *GeoCas9* modeled with AlphaFold2 (gray) with the homologous portion of *SpyRec1* and *SpyRec2* from *SpyCas9* (PDB: 4UN3) in red and orange, respectively. Assigned ^1H - ^{15}N TROSY HSQC spectrum of *GeoRec1* (top left), *GeoRec* (bottom left), and *GeoRec2* (bottom right).

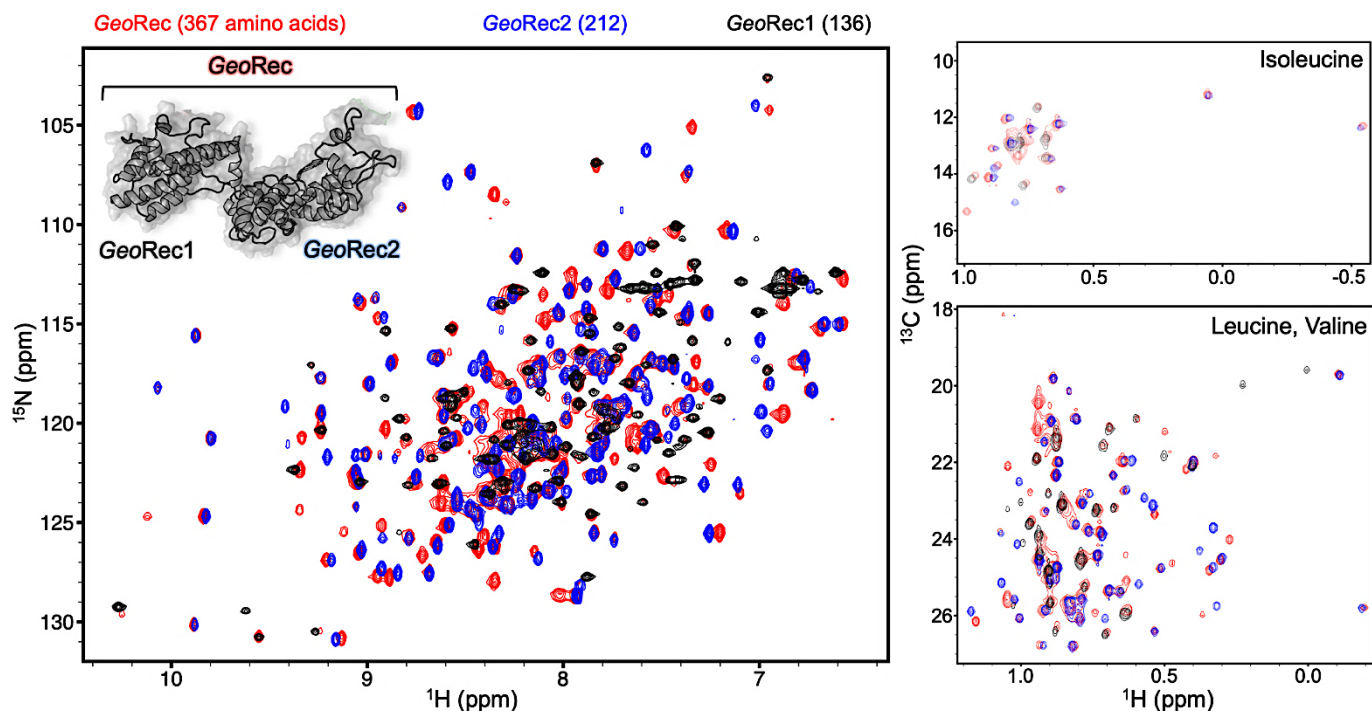


Figure 3. The dumbbell shape of *GeoRec* is composed of the *GeoRec1* and *GeoRec2* subdomains. Overlay of ^1H - ^{15}N TROSY HSQC spectra of *GeoRec* (red), *GeoRec1* (black), and *GeoRec2* (blue) collected at 850 MHz shows that spectra of isolated *GeoRec* subdomains overlay nicely with the spectrum of *GeoRec*. The number of amino acids that make up each construct are indicated above the spectrum. Overlaid ^1H - ^{13}C ILV-methyl spectra of *GeoRec* (red), *GeoRec1* (black), and *GeoRec2* (blue), show a similar property of the subdomains.

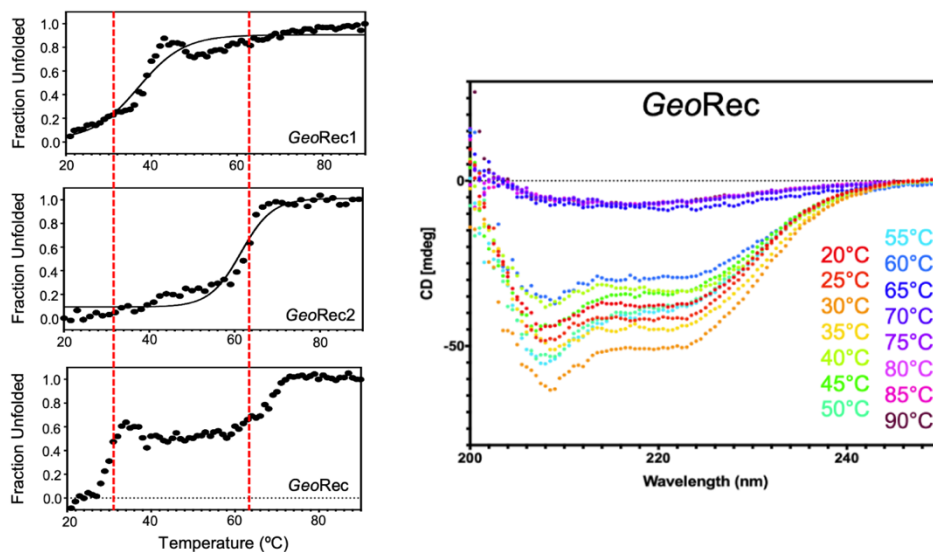


Figure 4. Temperature-dependent CD unfolding profiles of *GeoRec1*, *GeoRec2*, and *GeoRec* reveal that the unfolding profile of the individual subdomains are conserved within that of *GeoRec*. CD spectra spanning 200 – 250 nm at increasing temperatures (20-90 °C, right) show a gradual loss of *GeoRec* secondary structure to ~60 °C, followed by an abrupt and complete unfolding at 65 °C, beyond the T_m of *GeoRec2*. The T_m of *GeoRec1* and 2 were determined by fitting the CD data as described in the Materials and Methods. The red dashed lines indicate the T_m values of *GeoRec1* and *GeoRec2*, to guide the interpretation of the full-length *GeoRec* unfolding profile.

of the *GeoRec* lobe. Studies of *SpyCas9* revealed that interaction of *SpyRec3* (analogous to *GeoRec2*) with its RNA:DNA hybrid triggers conformational rearrangements that allow the catalytic HNH domain to sample its active conformation.¹¹ Thus, *SpyRec3* acts as an allosteric effector that recognizes the RNA:DNA hybrid to activate HNH. Mismatches (*i.e.* off-target DNA sequences) in the target DNA generally prevent *SpyRec3* from undergoing the full extent of its required conformational rearrangements, leaving HNH in a “proofreading” state with its catalytic residues too far from the DNA cleavage site. Off-target DNA cleavage by Cas9 remains an area of intense study and substantial effort from various groups has gone into mitigating such effects.^(15,23,28,32,33) Indeed, many high-specificity *SpyCas9* variants contain mutations within *SpyRec3* that increase the threshold for its conformational activation, reducing the propensity for HNH to sample its active state in the presence of off-target DNA sequences.^(11,15,23) Studies of flexibility within Rec itself, as well as its gRNA interactions in the presence of mutations, are therefore essential to connecting biophysical properties to function and specificity in related Cas9s.

The K267E *GeoRec2* variant is sequentially and structurally similar to a specificity enhancing site in *SpyCas9* (K526E), within the *evoCas9* system.⁽²⁷⁾ The *SpyCas9* K526E mutation substantially reduced off-target activity alone, but was even more effective in conjunction with three other single-point mutations in *SpyRec3*.⁽²⁷⁾ The R332A *GeoRec2* variant also resembles one mutation within a high-specificity *SpyCas9* variant, an early iteration of HiFi *SpyCas9* called HiFi Cas9-R691A.⁽²⁴⁾ We assessed mutation-induced changes to local structure in *GeoRec* via NMR chemical shift perturbations in ¹H-¹⁵N HSQC backbone amide spectra. Consistent with experiments using *GeoRec2* alone, chemical shift perturbations and line broadening are highly

localized to the mutation sites. (**Figure 5C-F**). Perturbation profiles of the *GeoRec2* subdomain and intact *GeoRec* also implicate the same residues as sensitive to the mutations (**Figure 6**).

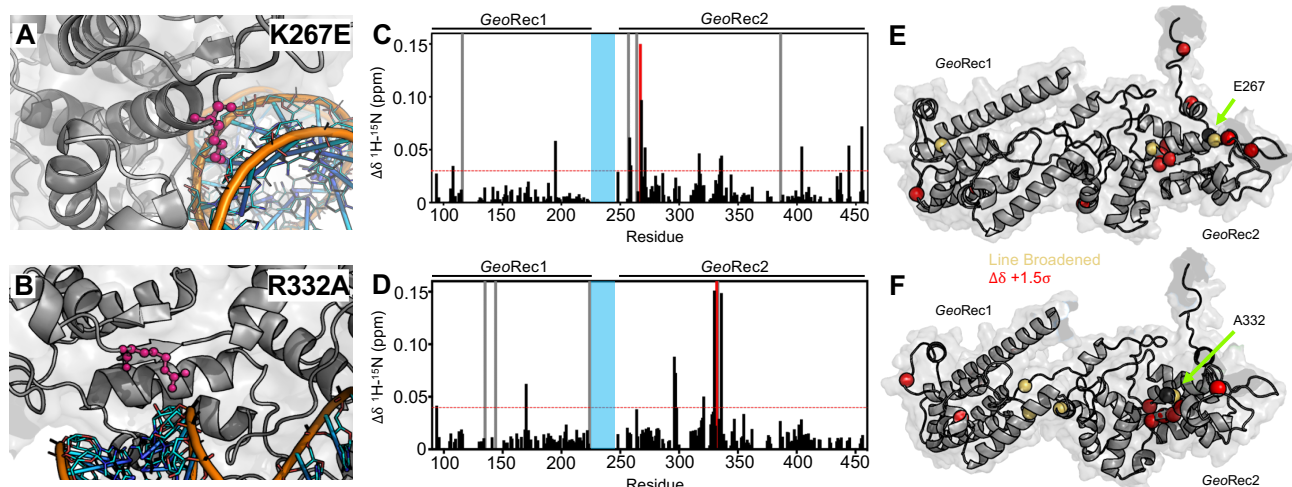


Figure 5. (A, B) Sites of selected mutations within *GeoRec2*, K267 and R332, are highlighted as purple sticks directly facing the RNA and DNA modeled from *NmeCas9* (PDB ID: 6JDV), allowing for prediction of the binding orientation within *GeoCas9*. NMR chemical shift perturbations caused by the K267E (C) or R332A (D) mutations are plotted for each residue of *GeoRec*. Gray bars denote sites of line broadening, the blue bar denotes an unassigned region of *GeoRec* corresponding to the native Rec1-Rec2 linker, and the red bar indicates the mutation site. The red dashed line indicates 1.5σ above the 10% trimmed mean of the data. Chemical shift perturbations 1.5σ above the 10% trimmed mean are mapped onto K267E (E) and R332A (F) *GeoRec* (red spheres). Resonances that have broadened beyond detection are mapped as yellow spheres and the mutation sites are indicated by a black sphere and green arrow.

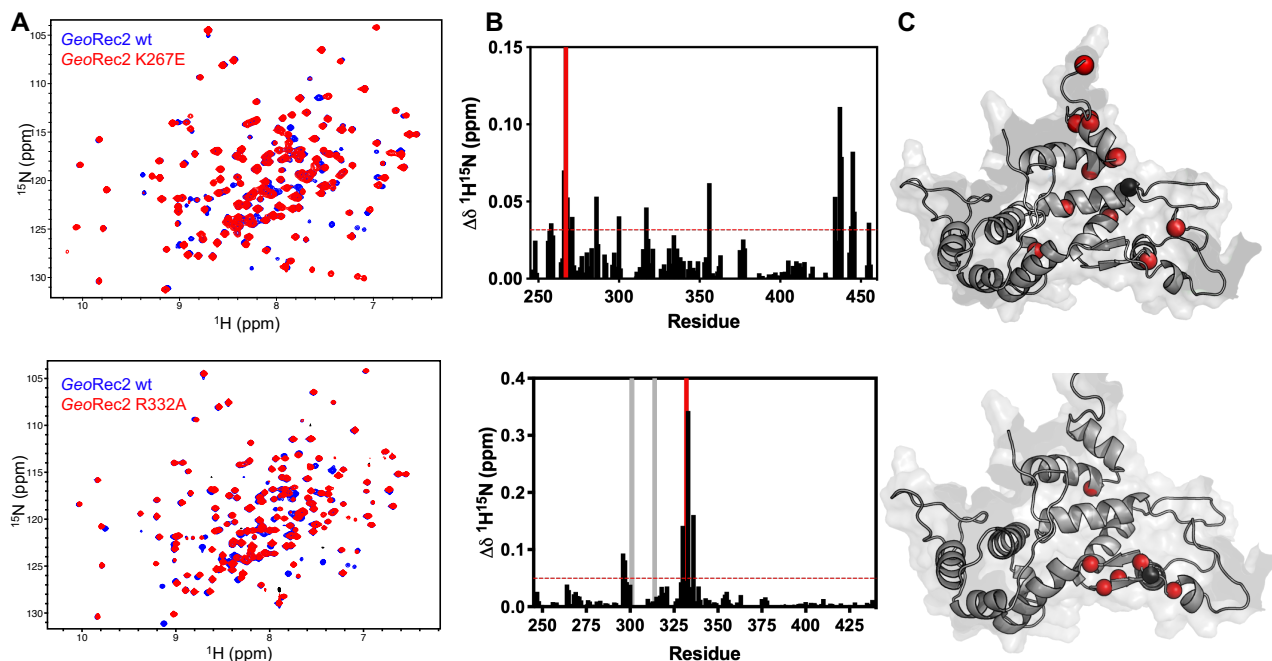


Figure 6. (A) 600 MHz ^1H - ^{15}N TROSY HSQC spectra of WT *GeoRec2* (blue) overlaid with K267E *GeoRec2* (red, top) and R332A *GeoRec2* (red, bottom). (B) Chemical shift perturbations caused by K267E (top) or R332A (bottom) mutations are plotted for each residue. Resonances broadened beyond detection are marked with gray bars. The mutation site is marked with a red bar. (C) Chemical shift perturbations $>1.5\sigma$ of the 10% trimmed mean of all shifts are shown as red spheres on the *GeoRec2*. The mutation site is shown as black sphere. 1.5σ above the 10% trimmed mean are mapped onto K267E (E) and R332A (F) *GeoRec* (red spheres). Resonances that have broadened beyond detection are mapped as yellow spheres and the mutation sites are indicated by a black sphere and green arrow.

CD spectroscopy revealed that wild-type (WT), K267E, and R332A *GeoRec2* maintained similar alpha-helical secondary structure, though the thermostability of both variants was slightly reduced from that of WT *GeoRec2* (**Figure 7**). The T_m of WT *GeoRec2* is ~62 °C, consistent with the T_m of the full-length *GeoCas9*, while that of K267E *GeoRec2* was decreased to ~55 °C. Though the R332A *GeoRec2* T_m remains ~62 °C, this variant underwent a smaller unfolding event near 40 °C before completely unfolding. These data suggest that despite small structural perturbations, both mutations are destabilizing to *GeoRec2*, which led us to expect a change in NMR-detectable protein dynamics.

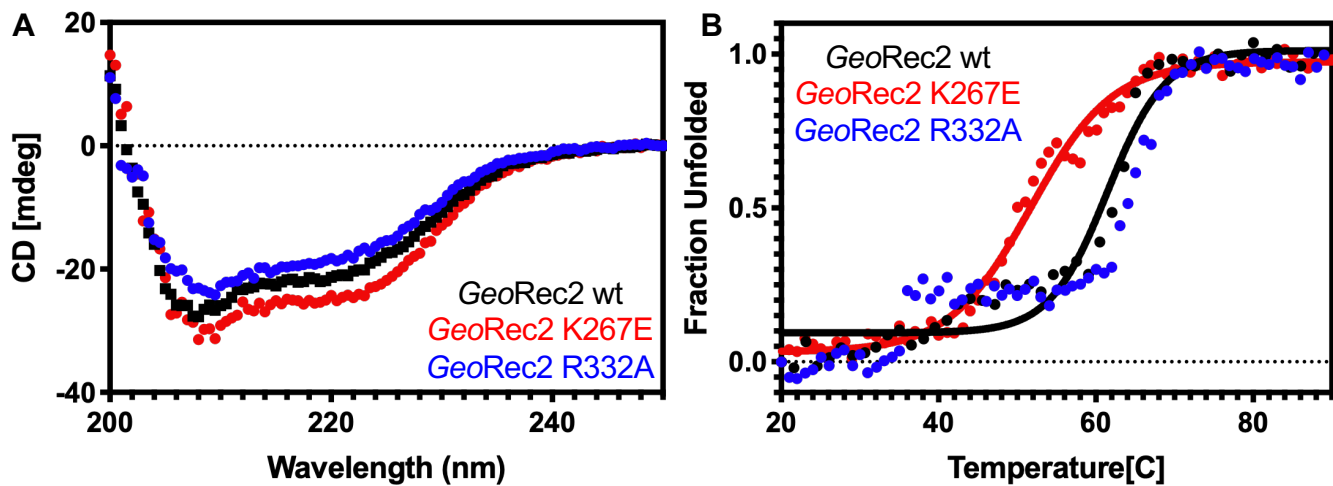


Figure 7. (A) Raw CD spectra of WT *GeoRec2* (black), K267E *GeoRec2* (red), and R332A *GeoRec2* (blue). **(B)** Temperature-dependent CD spectra reveal that K267E *GeoRec2* unfolds at a lower temperature (~55 °C, red) than WT *GeoRec2* (~62 °C, black). R332A *GeoRec2* undergoes a smaller unfolding event around 40 °C before completely unfolding ~62 °C (blue). The data were fit as described in the Materials and Methods.

Mutations enhance and redistribute molecular motions within GeoRec2

Due to the high molecular weight of the intact *GeoRec* lobe, decays in NMR signal associated with spin relaxation experiments were significant and hampered data quality. Thus, we focused on quantifying the molecular motions of the *GeoRec2* subdomain, where the K267E and R332A mutations reside, and the chemical shift perturbations are most apparent. To obtain high-quality per-residue information representative of *GeoRec*, we measured longitudinal (R_1) and transverse (R_2) relaxation rates and heteronuclear ^1H - ^{15}N NOEs (**Figure 8**), then used these data in a Model-free analysis of per-residue order parameters (S^2). Previous measurements of S^2 across the adjacent *GeoHNH* nuclease revealed substantial *ps-ns* timescale flexibility,⁽⁷⁾ leading us to wonder whether a similar observation would be made for *GeoRec2*, which abuts *GeoHNH*. Such a finding could suggest that HNH-Rec2 crosstalk in *GeoCas9* is driven primarily by rapid bond vector fluctuations. However, unlike *GeoHNH*, S^2 values for *GeoRec2* are globally elevated, suggesting that the *ps-ns* motions of this subdomain arise primarily from global tumbling of the protein in solution. We therefore carried out Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR experiments to assess the flexibility of *GeoRec2* on slower timescales, which has been linked to chemical information transfer in the well-studied *SpyCas9*.^(10,16,17,29) Evidence of $\mu\text{s-ms}$ motions (*i.e.* curved relaxation dispersion profiles) is observed in 17 residues within the *GeoRec2* core, spanning its interfaces to Rec1 and HNH (**Figure 9A/B**). Such motions are completely absent from *GeoHNH*, thus two neighboring domains, *GeoRec2* and *GeoHNH*, diverge in their intrinsic flexibility (at least in isolation), raising questions about the functional implications of these motions in *GeoRec2*. We previously showed that heightened flexibility of *SpyRec3* via specificity-enhancing mutations concomitantly narrowed the conformational space sampled by *SpyHNH*, highlighting a “motional trade-off” between the domains. Manipulation of the flexibility

of *Spy*Cas9 and *Geo*Cas9 domains by mutagenesis also impacts aspects of nucleic acid binding and cleavage,^(1,10,14,15,29) which led us to investigate similar perturbations in *Geo*Rec2.

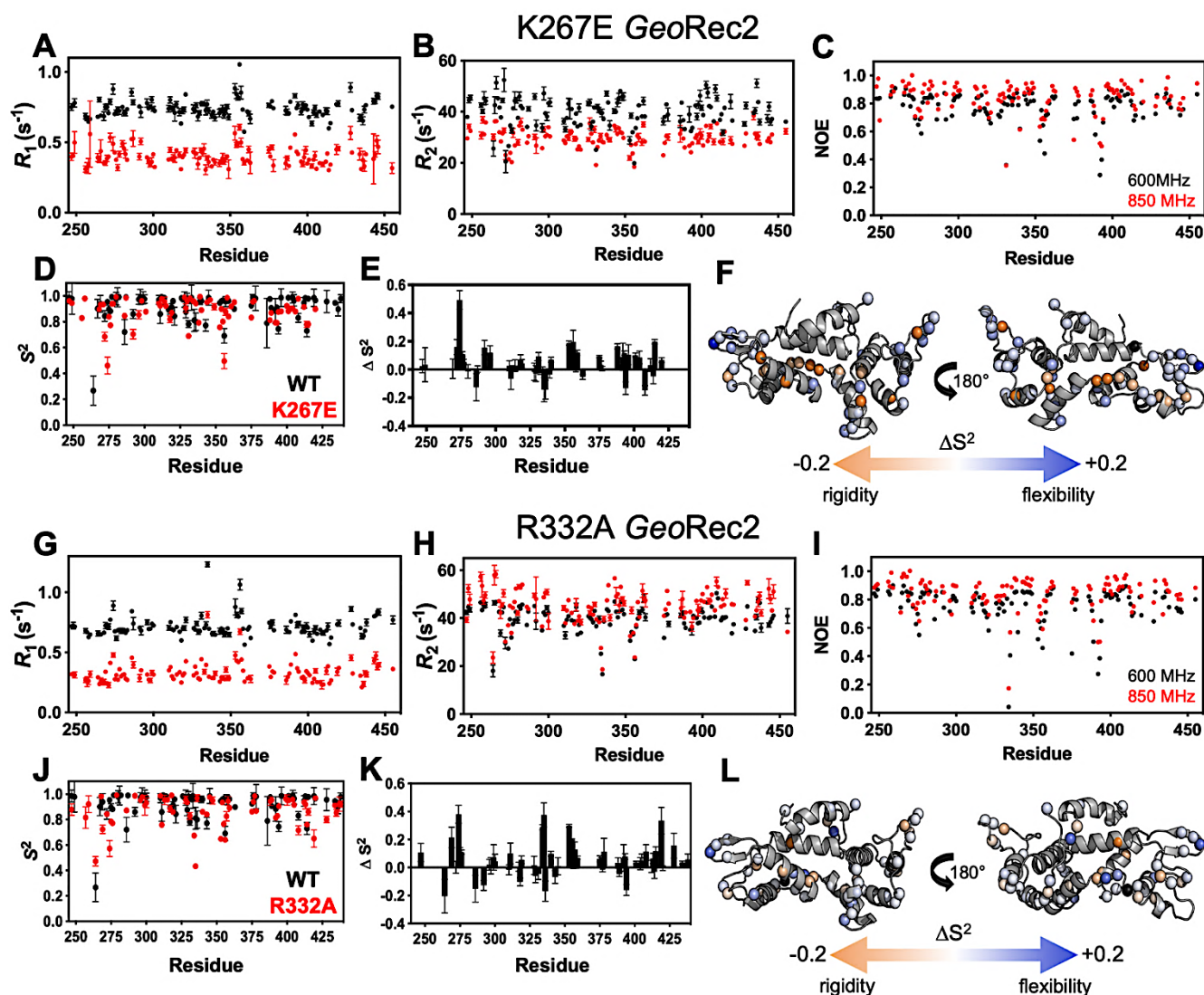


Figure 8. R_1 (A), R_2 (B), and ^1H - ^{15}N NOE (C) values for K267E *GeoRec2* measured by NMR at 600 MHz (black) and 850 MHz (red). (D) Order parameters (S^2) for K267E *GeoRec2* determined from Model-free analysis of R_1 , R_2 , and ^1H - ^{15}N NOE measurements (red) overlaid with S^2 values of WT *GeoRec2* (black). ΔS^2 (plotted as WT – mutant) for K267E *GeoRec2* (E), where a negative or positive values correspond to suppressed or heightened ps-ns flexibility of that site, respectively. K267E ΔS^2 values are mapped onto the *GeoRec2* X-ray crystal structure in (F). R_1 (G), R_2 (H), and ^1H - ^{15}N NOE (I) values for R332A *GeoRec2* measured by NMR at 600 MHz (black) and 850 MHz (red). (J) Order parameters (S^2) for R332A *GeoRec2* (red) overlaid with S^2 values of WT *GeoRec2* (black). ΔS^2 (plotted as WT – mutant) for R332A *GeoRec2* (K) and R332A ΔS^2 values are mapped onto the *GeoRec2* X-ray crystal structure in (L).

Since *SpyRec3* and *GeoRec2* have similar structures and μs - ms flexibility, we speculated that charge-altering mutations would modulate the biophysical properties of *GeoRec* and the function of *GeoCas9*, as observed for *SpyCas9*. We investigated K267E and R332A *GeoRec2* with NMR spin relaxation, as described for WT *GeoRec2* (*vide supra*). An analysis of chemical exchange rates, k_{ex} , derived from dual-field CPMG relaxation dispersion show a global k_{ex} for WT *GeoRec2* of $147 \pm 41 \text{ s}^{-1}$. The K267E mutation, which directly contacts the nucleic acids, shifts the globally fitted k_{ex} to $376 \pm 89 \text{ s}^{-1}$, while the R332A variant maintains a global k_{ex} similar to that of WT *GeoRec2* ($142 \pm 28 \text{ s}^{-1}$) and consistent with its similar thermal stability. The global fit of the K267E variant is based on CPMG profiles of 33 residues, while that of R332A is derived from 18 residues (**Figure 9**). Interestingly, the residues participating in the global motions of both variants are distinct from those of WT *GeoRec2*, demonstrating that residue-specific flexibility is redistributed throughout *GeoRec2*, which suggests an altered intradomain molecular crosstalk within the larger *GeoRec*. Indeed, perturbation to NMR-detectable motions in *SpyCas9* rewired its allosteric signaling and enzymatic function.^(16,29) A similar dynamic modulation of *GeoCas9* may fine-tune its DNA cleavage, which has been demonstrated within the *GeoHNH* nuclease¹ and wedge (WED) domains.⁽³²⁾ We also assessed the ps-ns fluctuations of *GeoRec2* variants (a negligible contribution to the WT *GeoRec2* dynamic profile) and calculated order parameters from R_1 , R_2 and ^1H - ^{15}N NOE relaxation measurements (**Figure 9C**). Bond vector fluctuations on the ps-ns timescale are only locally altered, thus the mutation-induced reshuffling of these motions is negligible ($\langle \Delta S^2 \rangle \leq 0.1$) and suggests that, like WT *GeoRec2*, *ps*-*ns* motion arises primarily from global tumbling in solution.

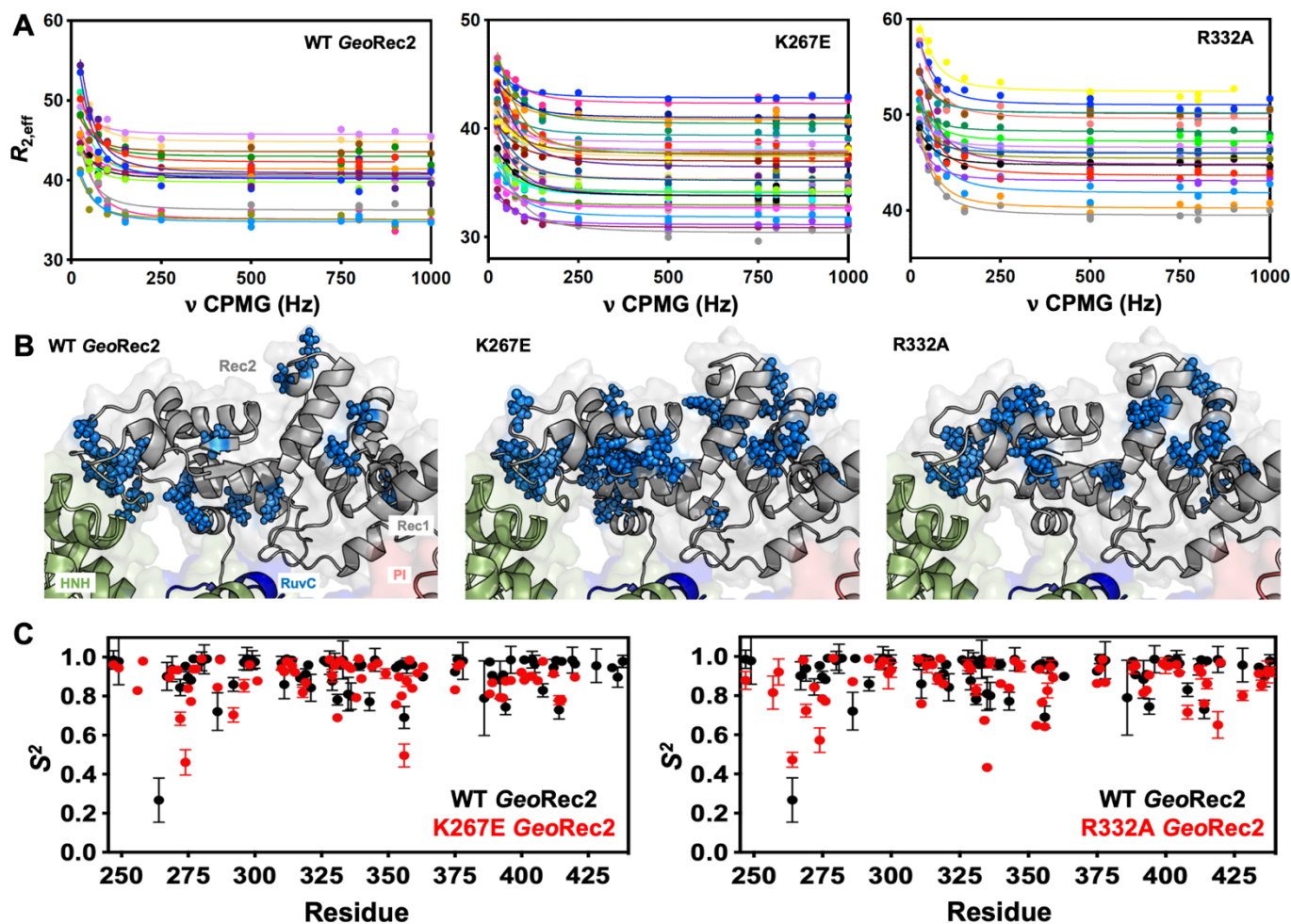


Figure 9. (A) CPMG relaxation dispersion profiles of all residues with evidence of μ s-ms motion, fit to a global k_{ex} of 147 ± 41 s $^{-1}$ (WT *GeoRec2*, left), 376 ± 89 s $^{-1}$ (K267E *GeoRec2*, center), and 142 ± 28 s $^{-1}$ (R332A *GeoRec2*, right) collected at 25 °C and 600 MHz. Relaxation dispersion profiles for individual resonances are shown in Figures S8-S10. (B) Sites exhibiting CPMG relaxation dispersion in (A) are mapped to *GeoRec2* as blue spheres. Adjacent domains within the cryo-EM structure of *GeoCas9* are also shown. (C) Per-residue NMR order parameters of WT (black), K267E, and R332A (red, separate plots) *GeoRec2*.

Mutations within GeoRec alter its affinity for RNA

The role of the Rec lobe in orienting the RNA:DNA hybrid within Cas9 is crucial to its function.⁽¹⁰⁻¹³⁾ Thus, the structure, motions, and nucleic acid interactions of Rec represent a critical piece of the Cas9 signaling machinery. Previous studies of *Spy*Cas9 revealed that gRNA binding to the Rec lobe induces a global structural rearrangement of the protein that positions the adjacent HNH into its “proofreading” state,⁽¹⁵⁾ after which target DNA binding positions the nucleases into active conformations for cleavage.^(15,34) We wondered if the atomistic details of the apo *Geo*Cas9-to-RNP transition could be captured by NMR using the *Geo*Rec construct. In our previous studies, we used an *in vitro* DNA cleavage assay with *Geo*Cas9 and a 141nt gRNA containing a 21nt spacer targeting the mouse *Tnnt2* gene locus.⁽¹⁾ Since this assay was already established, we utilized the same gRNA sequence. However, truncating this gRNA was necessary to optimize binding studies for NMR analysis. We focused on the 5' end of the gRNA, which includes the spacer sequence, based on the *Geo*Cas9 AlphaFold2 model and structural data from *Nme*Cas9 and *Spy*Cas9 showing interactions between the Rec lobe and this region of the gRNA. The subsequent cryo-EM structure of *Geo*Cas9 corroborated this interaction.⁽²⁰⁾ Initial attempts using a truncated 101nt portion of the full gRNA resulted in poor NMR spectra. An overlay of the ¹H-¹⁵N HSQC NMR spectra of apo *Geo*Rec and *Geo*Rec-RNP at a 1:1 molar ratio showed extensive line broadening (**Figure 10**), likely due to the large size of the complex (75.5 kDa). To mitigate this issue, a 39nt RNA containing the 21bp spacer sequence was selected for its ability to maintain the NMR signal while being long enough to interact fully with the Rec lobe, as suggested by prior structures (**Figure 11**). When bound to *Geo*Rec, this complex is 55.6 kDa and a ¹H-¹⁵N NMR spectral overlay of apo *Geo*Rec and the domain bound to 39nt RNA shows clear, resolved resonances with significant chemical shift perturbations and line broadening (**Figure 12A/B**). The strongest chemical shift

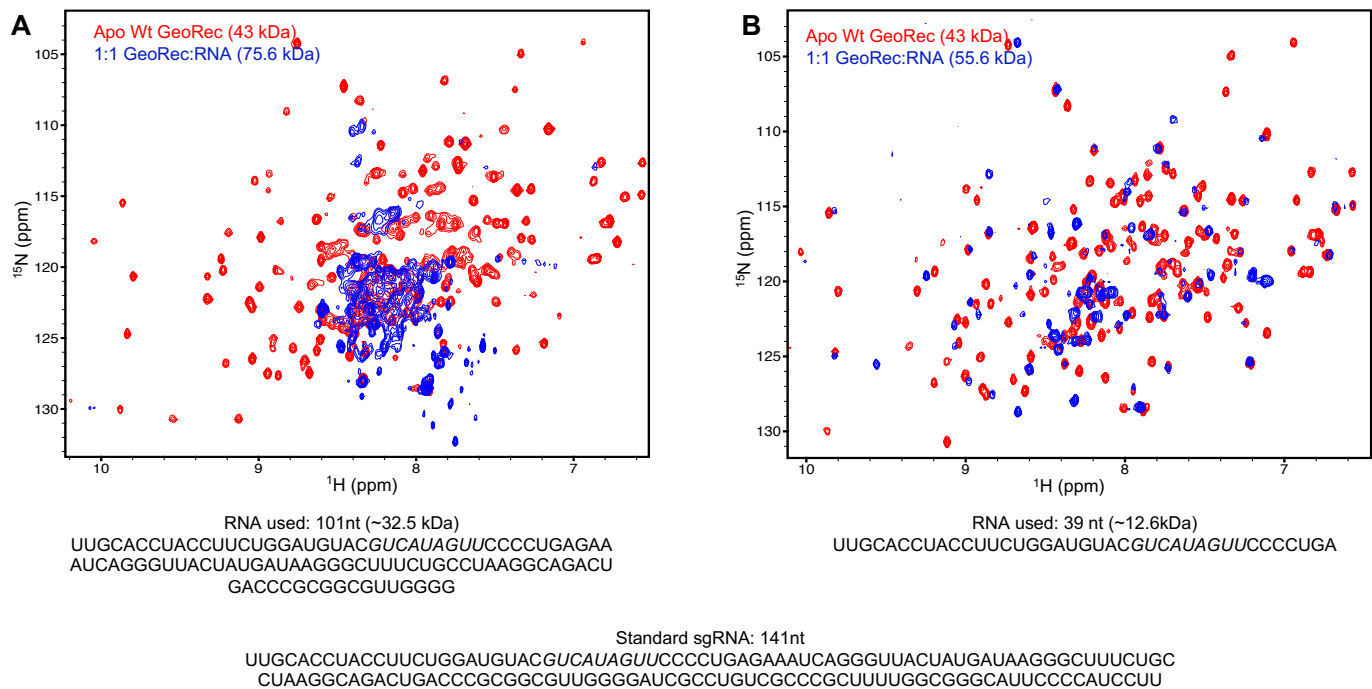


Figure 10. 850 MHz ^1H - ^{15}N TROSY HSQC NMR spectra of apo-GeoRec (red) and GeoRec in complex with either a 101-nt (**A**) or 39-nt portion of the full gRNA (**B**) at a 1:1 molar ratio (blue). The RNA sequence used in each NMR binding experiment is shown below each spectrum, along with the full-length gRNA sequence.

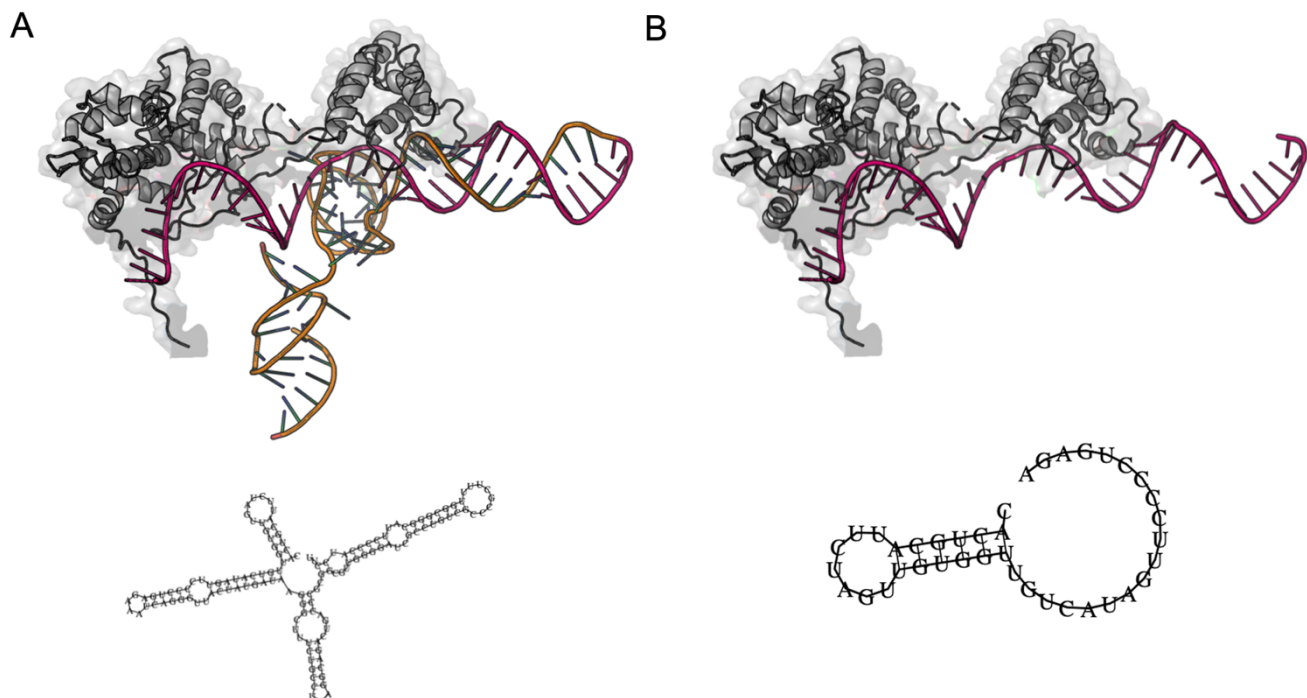


Figure 11. Structures of full-length (**A**) and truncated (**B**) GeoCas9 gRNAs used in this work, based on the cryo-EM structure PDB: 8UZA. The 2D structure cartoons below each representation were predicted from sequence by the RNAfold web server (<http://rna.tbi.univie.ac.at/cgi-bin/NAWebSuite/RNAfold.cgi>) maintained by the Institute for Theoretical Chemistry, University of Vienna.

perturbations are localized to the *GeoRec2* subdomain that interfaces with the RNA:DNA hybrid at the PAM distal end, where previous studies of specificity-enhancing variants of *SpyCas9* have identified alterations in nucleic acid binding to *SpyRec3*.⁽¹⁶⁾ It is not known whether specific residues at the PAM distal binding interface of *GeoRec2* play a similar role. Line broadening is evident in both *GeoRec1* and *GeoRec2*, primarily localized to the RNA:DNA hybrid interface revealed in recent *GeoCas9* structures. Microscale thermophoresis (MST) experiments quantified the affinity of *GeoRec* for this RNA, producing a $K_d = 3.3 \pm 1.5 \mu\text{M}$ that is consistent with the concentration-dependent NMR chemical shift perturbations (**Figure 5A**).

To understand how the K267E and R332A mutants impact RNA binding to *GeoRec*, we conducted RNA titration experiments via NMR and observed that chemical shift perturbations were attenuated in both variants, relative to WT *GeoRec*. Despite this muted structural effect, the impact from RNA-induced line broadening remains substantial in the *GeoRec1* subdomain. Our NMR data revealed that a three-fold greater concentration of RNA was required to induce the maximal structural and dynamic effects in the variants than is required for WT *GeoRec* (**Figure 12A/C**), suggesting that the variants have a reduced RNA affinity. MST experiments showed statistically significant reductions in RNA affinity for the K267E and R332A constructs, relative to WT *GeoRec*, where K267E *GeoRec* produced a $K_d = 7.2 \pm 1.0 \mu\text{M}$ and R332A *GeoRec* produced a $K_d = 7.2 \pm 1.5 \mu\text{M}$ (**Figure 13B**). The ~2-fold increase in K_d may also be due, in part, to a change in the binding mode of the RNA, such as a faster k_{off} . Collectively, these data reveal that mutations within *GeoRec* primarily alter its structure around the mutation site with weaker distal effects, but more significantly impact protein dynamics and in turn, the RNA interaction. NMR experiments also demonstrate that the presence of RNA impacts both subdomains of *GeoRec*, providing a significant structural interface for additional molecular tuning of nucleic acid binding.

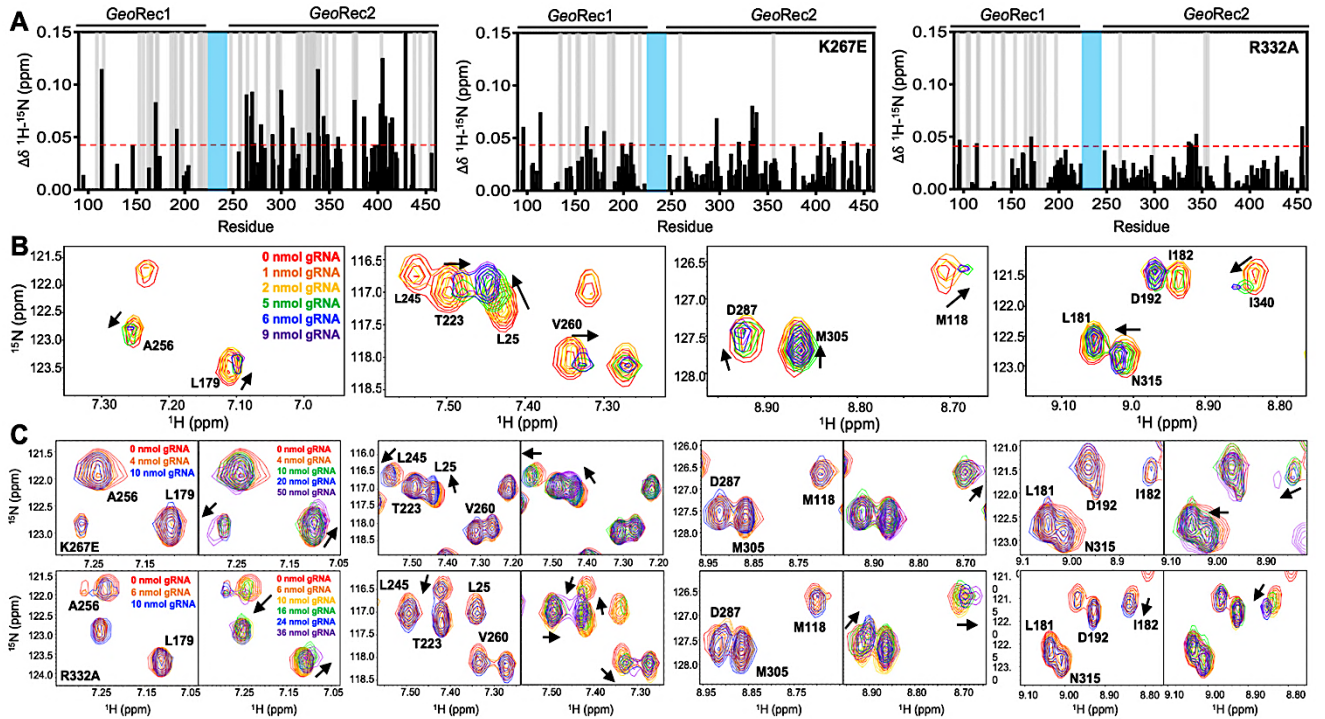


Figure 12. (A) NMR chemical shift perturbations caused by RNA binding to WT, K267E, and R332A *GeoRec*. Gray bars denote sites of line broadening, and the blue bar denotes an unassigned region of *GeoRec* corresponding to the flexible Rec1-Rec2 linker. The red dashed line indicates 1.5σ above the 10% trimmed mean of the data. (B) Representative NMR resonance shifts caused by titration of a 39nt portion of the full gRNA into WT *GeoRec*. (C) NMR titration of 39nt RNA into K267E (top) and R332A (bottom) *GeoRec*. The left panel of each pair demonstrates that minimal change in NMR chemical shift or resonance intensity is apparent at RNA concentrations mimicking the WT titration. The right panel of each pair depicts the titration over a three-fold wider concentration range of RNA, where shifts and line broadening are visible. Representative resonances are colored by increasing RNA concentration in the legend.

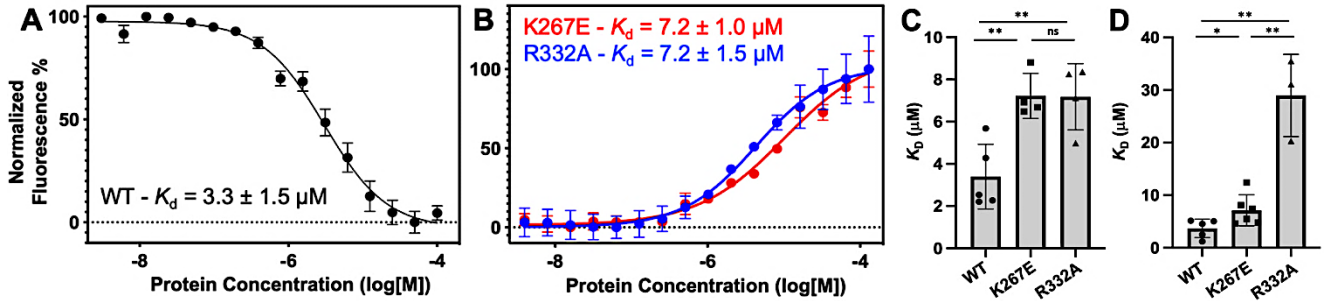


Figure 13. Representative MST-derived profiles of WT (A), K267E, and R332A (B) *GeoRec* binding to a Cy5-labeled 39nt portion of a gRNA, yielding $K_d = 3.3 \pm 1.5 \mu\text{M}$, $K_d = 7.2 \pm 1.0 \mu\text{M}$ and $K_d = 7.2 \pm 1.5 \mu\text{M}$, respectively. Bar graphs comparing K_d values across $n \geq 3$ replicate samples are shown for Tnnt2 RNA (C) and 8UZA RNA from a recent cryo-EM structure (D). * $p < 0.05$, ** $p < 0.004$

To investigate the impact of RNA binding on *GeoRec2* in greater detail, we conducted NMR titration experiments using the isolated domain, which yielded even clearer NMR spectra. **Figure 14** shows NMR spectra of WT, K267E, and R332A *GeoRec2* overlaid with their corresponding RNA-bound spectra (39nt Tnnt2 RNA). At protein:RNA molar ratios used for full-length *GeoRec* studies, the WT *GeoRec2* spectrum exhibited significant line broadening across the *GeoRec2* sequence. Plots of NMR peak intensities ($I_{\text{bound}}/I_{\text{free}}$) show substantial resonance intensity losses (**Figure 14**) with many residues likely in the intermediate exchange regime, in addition to the assumed changes in rotational correlation of the domain. In comparison, spectra of the RNA-bound K267E and R332A *GeoRec2* variants showed less pronounced signal decay at the same levels of titrant, retaining nearly double the $I_{\text{bound}}/I_{\text{free}}$ ratio across these spectra (**Figure 14**). These data are consistent with the results of NMR experiments with the 43 kDa *GeoRec*, supporting the premise that *GeoRec2* mutations weaken its interaction with RNA. The less crowded NMR spectrum of isolated *GeoRec2* facilitated the resolution of distinct structural features that explain the impact of the mutations on RNA binding (**Figure 9**). For example, residue I53 adopts a similar conformation in RNA-bound WT and K267E *GeoRec2* but assumes a different structural state in R332A. Conversely, residue R25 populates a WT-like structure in RNA-bound R332A *GeoRec2*, unlike K267E. Additionally, two resonances are observed for residue K71 in the RNA-bound R332A NMR spectrum, indicating real-time equilibration between two structural states. This effect is unique to the R332A variant and underscores subtle structural and dynamic changes to *GeoRec* during RNA binding.

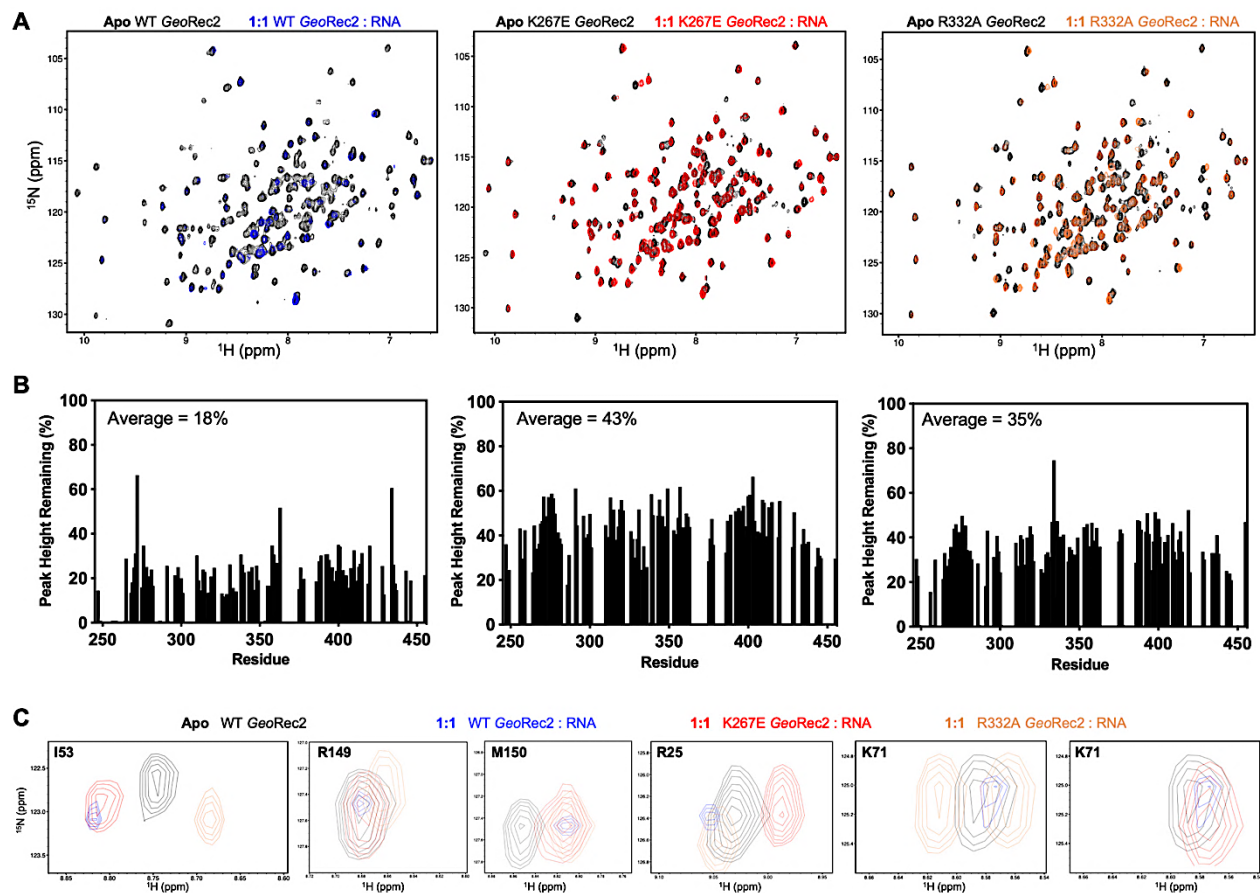


Figure 14. (A) $^1\text{H}/^{15}\text{N}$ TROSY HSQC NMR spectral overlays of apo (black) and RNA-bound WT (blue), K267E (red), and R332A (orange) *GeoRec2*. (B) Per-residue residual peak heights after addition of equimolar RNA to WT (left), K267E (center), and R332A (right) *GeoRec2*. (C) NMR snapshots of apo WT *GeoRec2* (black) and gRNA-bound WT (blue), K267E (red), and R332A (orange) *GeoRec2*, highlighting unique RNA-induced structural perturbations of each variant.

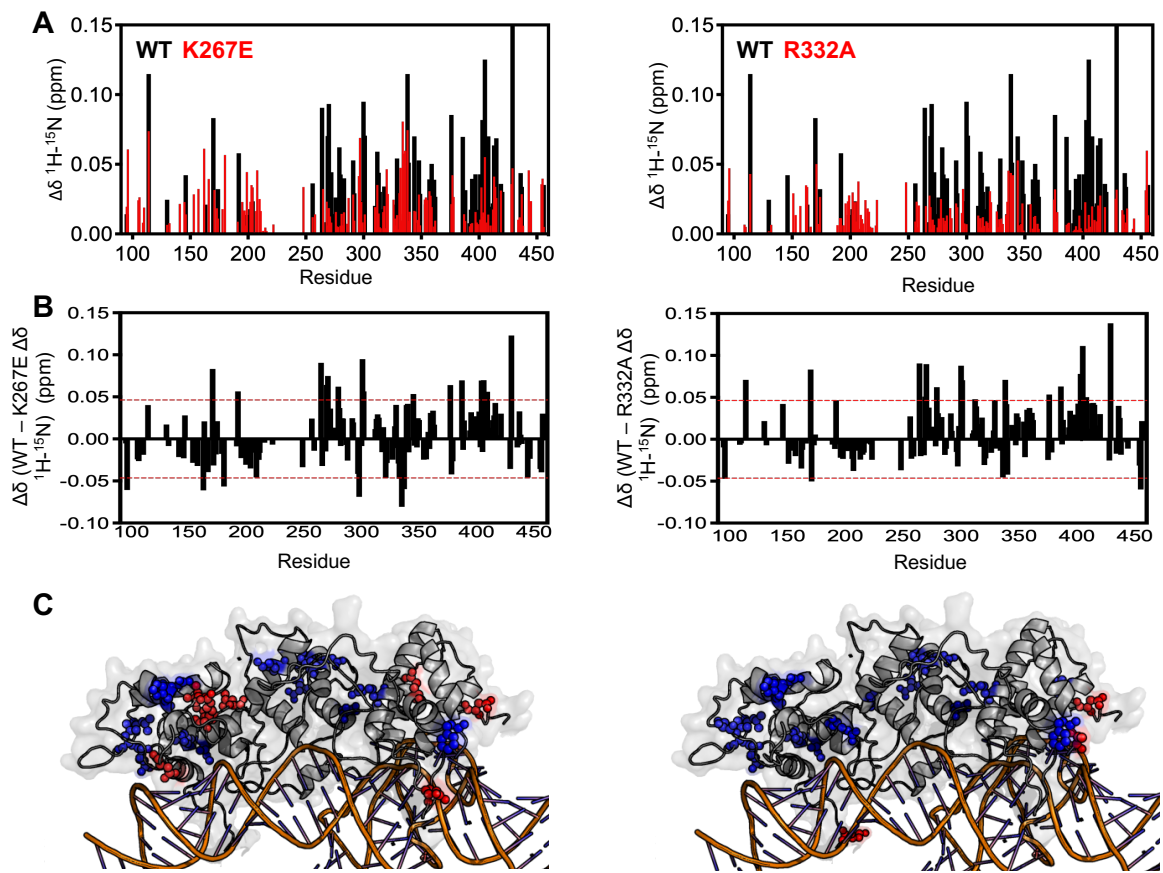


Figure 15. (A) RNA-induced NMR chemical shift perturbations to WT *GeoRec* (black) overlaid with those of K267E (left) and R332A (right) *GeoRec*, each in red. Similar plots are also found in Figure 1D and 4B of the main text. **(B)** The difference in RNA-induced chemical shift perturbation ($\Delta\delta$ WT – mutant) comparing WT and K267E *GeoRec* (left), and WT and R332A *GeoRec* (right). The red dashed lines indicate 1σ above and below the 10% trimmed mean of all shifts. **(C)** Chemical shift perturbations $+1.0\sigma$ above the trimmed mean of all data are mapped onto *GeoRec* as blue spheres, while perturbations -1.0σ are mapped onto *GeoRec* as red spheres, representing key residues influencing RNA interactions.

We further examined the NMR data to attempt to identify residues most critical for RNA binding to *GeoRec*. In an overlay of the WT and mutant RNA-induced chemical shift perturbations ($\Delta\delta$, **Figure 15**), it became clear that the effect of RNA binding to *GeoRec* variants was muted, where even at saturating concentrations, the chemical shift perturbations across the K267E and R332A *GeoRec2* sequences were weaker than those of same residues in WT *GeoRec*. The residual $\Delta\delta$ (WT - mutant) was plotted (**Figure 15**), where positive values indicate that residues in a *GeoRec* variant are weakly affected by RNA, relative to WT. Negative residual $\Delta\delta$ denote sites where *GeoRec* variants experience a greater structural impact from RNA than corresponding sites in WT. Of particular interest are the positive residuals that hint at the sites in *GeoRec* most critical for tight RNA binding. These residues were mapped onto the *GeoRec* structure (**Figure 15**) and termed allosteric hotspots, as many are not at the RNA interface. Mutations of these hotspots in future studies offers a potential means of precisely tuning the affinity of *GeoRec* to its gRNA. Notably, residues with positive residual $\Delta\delta$ (suggested as critical for tight RNA binding) largely overlap in the analysis of both variants. Specifically, residues F170, R192, H264, R269, L270, L279, H300, D301, E368, D376, D403, E405, E408, and I429 appear as allosteric hotspots (with CPMG relaxation dispersion) critical to WT-like RNA interaction.

Having observed a reduced affinity of *GeoRec* variants for RNA by NMR and MST, we next quantified the impact of the K267E and R332A mutations on RNP formation and stability in full-length *GeoCas9*. The thermal unfolding midpoint of full-length WT *GeoCas9* determined by CD is $\sim 60^\circ\text{C}$ and the K267E and R332A mutations do not change the T_m of the apo protein (**Figure 7**). Upon formation of an RNP (using a full-length gRNA), the T_m of WT *GeoCas9* increases to 73°C . K267E *GeoCas9* retains a similar T_m increase to 70°C , while R332A *GeoCas9* forms a less stable RNP with T_m of 61°C . The trend of these data is consistent with NMR and MST, which

highlight that although K267E and R332A mutations within *GeoRec* have somewhat muted structural effects, these changes alter protein dynamics and the interaction with gRNA.

Mutations in full-length GeoCas9 alter its structural dynamics and interaction with gRNA

To further investigate the effects of mutations on protein dynamics, we performed molecular dynamics (MD) simulations based on the cryo-EM structure of full-length *GeoCas9* (PDB: 8UZA) in complex with gRNA and target DNA. We simulated the full-length WT *GeoCas9* and its K267E and R332A mutants as well as a double mutant combining K267E and R332A (**Figure 16A**), in three replicates of approximately 2 μ s each. Multi-microsecond simulations revealed substantial changes in the dynamics of the *GeoCas9* mutants compared to the WT (**Figure 16**). Specifically, we observed that mutations in the REC domain significantly altered the dynamics of both the Rec and the adjacent HNH (**Figure 17**). Differential root-mean-square fluctuations (Δ RMSF) analysis of protein residues between the WT and variants further highlighted these alterations, showing increased dynamics in the HNH and Rec domains induced by the mutations (**Figure 16B**). To quantify the impact of these mutations, we analyzed protein-RNA interactions by calculating the number of contacts between *GeoCas9* and gRNA in the WT and mutant systems. A contact was defined as a distance between two atoms of ≤ 4.5 Å. The number of contacts was significantly reduced in all variants compared to the WT, with the most pronounced reduction observed in the K267E variant (**Figure 16C**). We next quantified gRNA-Rec domain binding by calculating the binding free energy using the MM-GBSA method over ~ 200 ns of stable simulation trajectories (details in Materials and Methods). Consistent with a reduction in gRNA contacts, variants with the K267E mutation exhibited a substantial reduction in binding free energy (>60 kcal mol⁻¹) relative to the WT, whereas the R332A variant displayed a smaller reduction (<20 kcal mol⁻¹, **Figure 16D**). Notably, protein-DNA interactions remained largely unaffected, suggesting that these mutations do not impair *GeoCas9* DNA cleavage ability.

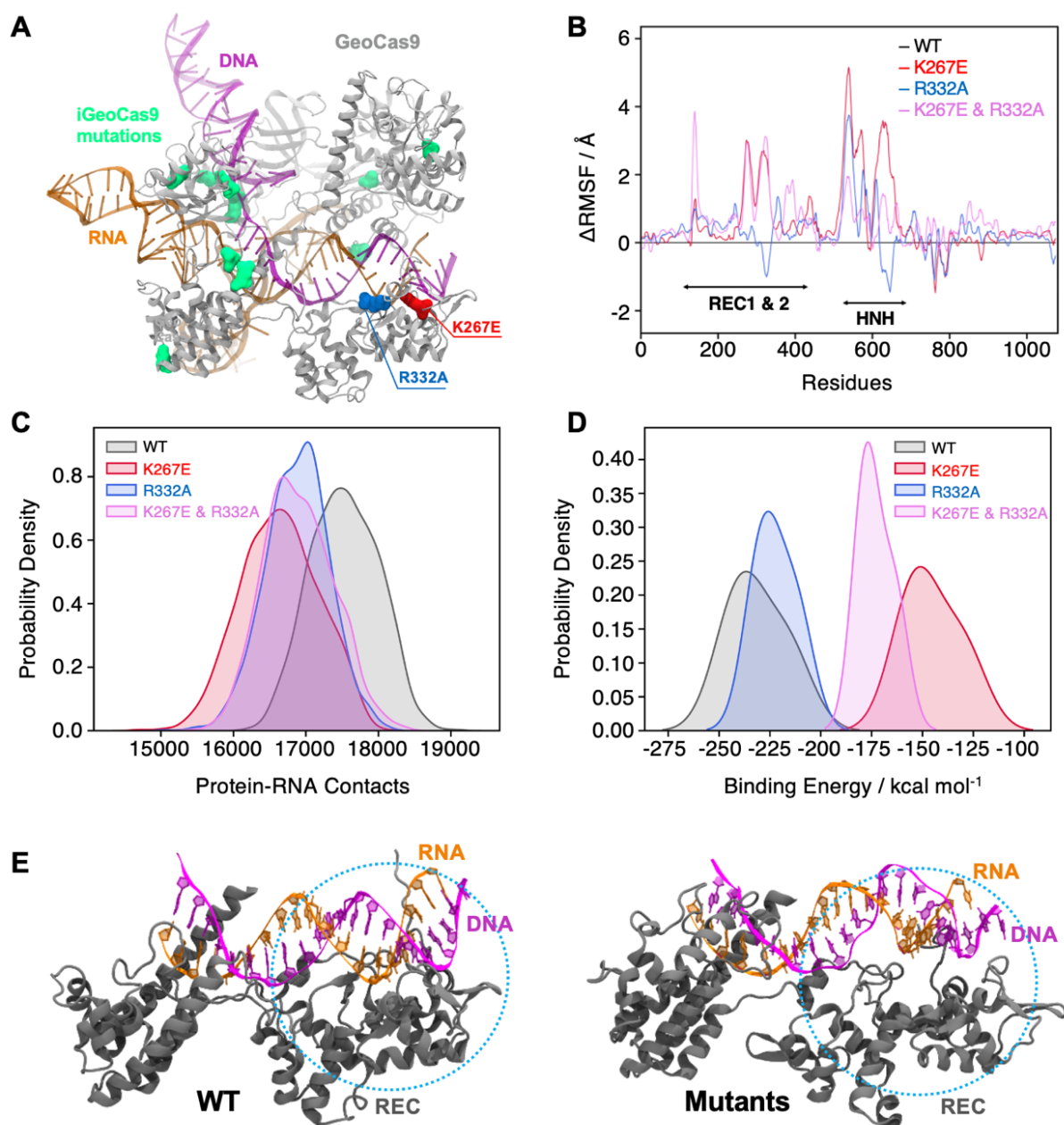


Figure 16. Effects of mutations in full-length *GeoCas9* revealed by MD simulations. **(A)** The structure of *GeoCas9* (PDB: 8UZA, protein in gray) bound to gRNA (orange) and DNA (magenta) is shown. Mutations studied include K267E (red), R332A (blue), and the 10 mutations of i*GeoCas9* (lime green), all highlighted in surface representation. **(B)** Differential root-mean-square fluctuations (ΔRMSF) of protein residues computed between WT *GeoCas9* and the K267E (red), R332A (blue), and double mutant (pink). **(C)** Distribution of protein-RNA contacts for WT and *GeoCas9* variants computed over the 6 μs simulation ensemble. **(D)** Comparison of gRNA binding free energy to the Rec domain in WT *GeoCas9* and variants. **(E)** Representative snapshots from MD simulations illustrating structural changes in Rec-gRNA association in WT *GeoCas9* (left) and variants (right).

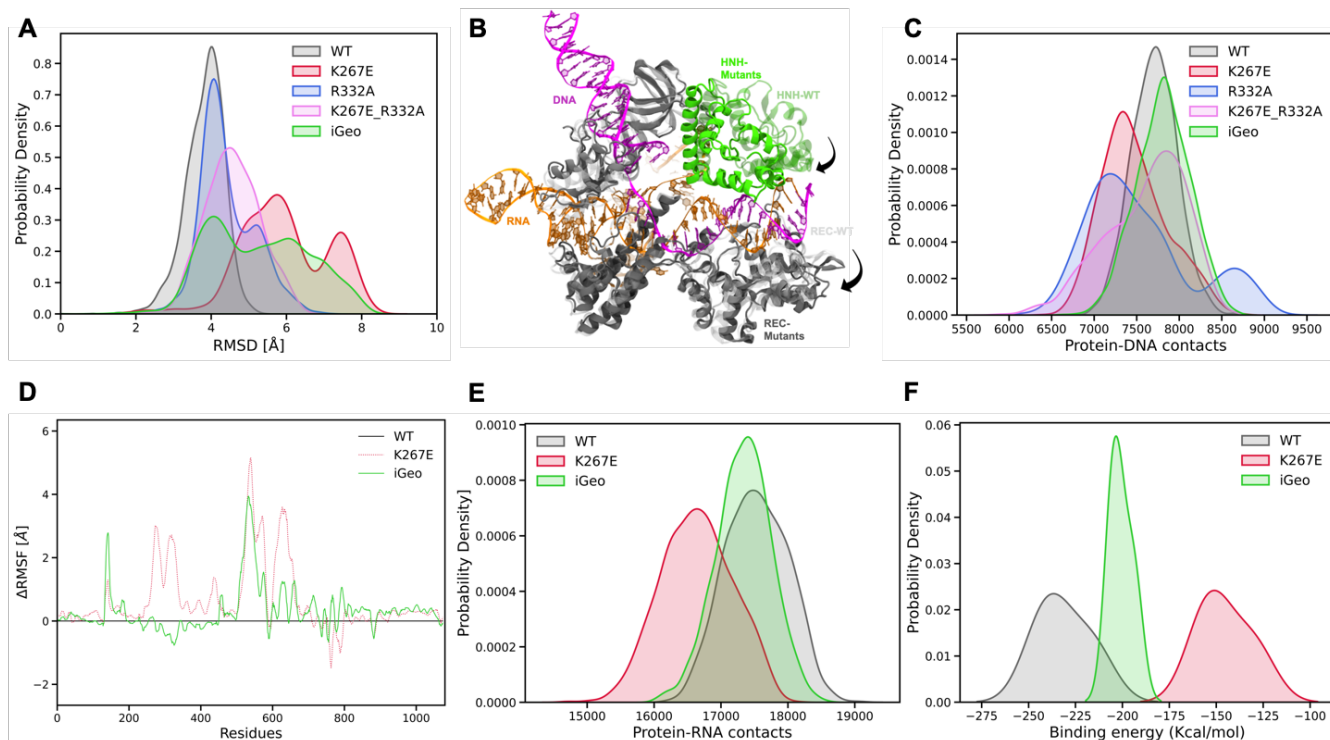


Figure 17. Analysis of MD simulations. **(A)** Root-mean-square deviation (RMSD) distribution of WT (gray), K267E (red), R332A (blue), K267E/R332A double mutant (pink) and iGeoCas9 (green). **(B)** Conformational changes in the HNH and Rec domains of the mutant systems (black arrows), compared to WT *GeoCas9*. **(C)** Distribution of protein-DNA contacts for WT and mutants computed over the 6 μ s ensemble. **(D)** Differential RMSD (Δ RMSF) of the protein residues computed between the WT *GeoCas9* and K267E (red) and iGeoCas9 (green). **(E)** Distribution of protein-RNA contacts for WT, K267E and iGeoCas9 computed over the 6 μ s ensemble. **(F)** Comparison of binding free energy of RNA with Rec domain of *GeoCas9* between WT, K267E and iGeoCas9.

Additionally, we simulated a novel variant, i*Geo*Cas9 (PDB: 8UZB), containing mutations in the Rec1 and WED domains (**Figure 16A**, mutations highlighted in lime green). This variant was recently demonstrated to have enhanced specificity in genome-editing.⁽²¹⁾ Intriguingly, i*Geo*Cas9 exhibited increased dynamics in the HNH and Rec domains, along with a reduction in gRNA binding free energy similar to the K267E and R332A mutants. These results suggest a distinct allosteric pathway involving additional residues that enables i*Geo*Cas9 to maintain improved DNA cleavage activity despite reduced gRNA binding affinity. In fact, i*Geo*Cas9 samples the greatest conformational space of any variant tested (**Figure 17**), suggesting a high level of flexibility is critical to enhanced specificity in *Geo*Cas9. Collectively, our simulations reveal that mutations K267E and R332A destabilize the *Geo*Cas9 interaction with gRNA, consistent with NMR observations. Furthermore, the enhanced dynamics and altered binding affinities observed in i*Geo*Cas9 indicate potential allosteric mechanisms that optimize its genome-editing functionality, where K267E and R332A evoke a similar, but lesser degree of biophysical change.

DNA cleavage assays suggest the highly stable GeoCas9 is resistant to functional changes by K267E or R332A mutations

The dynamic impact of the *Geo*Rec mutations and their altered gRNA interactions at the biophysical level led us to speculate that either mutation incorporated into full-length *Geo*Cas9 would also alter its DNA cleavage function, especially at elevated temperatures where WT *Geo*Cas9 is most active. Temperature-dependent functional alterations were previously observed for single-point mutations within *Geo*HNH.⁽¹⁾ Although the K267E and R332A mutations slightly diminished on-target DNA cleavage by *Geo*Cas9, the effect was very subtle and these overall

cleavage activities followed the temperature dependence of WT *GeoCas9* quite closely (**Figure 18**).

To assess the impact of the K267E and R332A mutations on *GeoCas9* specificity, we assayed the propensity for off-target cleavage using DNA substrates with mismatches 5-6 or 19-20 base pairs from the PAM site (**Figure 19**). As a control for on- and off-target activity, we assayed WT *SpyCas9* alongside the widely used high-specificity HiFi-*SpyCas9* variant⁽²⁴⁾ (**Figure 19**) and found a lower percent of digested off-target (mismatched) DNA sequences when compared to WT *SpyCas9*. As expected, WT *GeoCas9* was increasingly sensitive to mismatched target sequences closer to the seed site, which has been demonstrated with *SpyCas9* and other Cas systems.^(2,6,8,35) No significant differences in activity were observed with digestion durations ranging from 1-60 minutes,⁽²⁾ implying that a 1-minute digestion is sufficient for *in vitro* activity of *GeoCas9* with the target DNA template. While these findings generally align with prior investigations of off-target DNA cleavage,^(2,6,8) there are nuanced differences. Specifically, a previous study reported ~10% cleavage of off-target DNA with a mismatch 5-6 base pairs from the PAM by WT *GeoCas9*.⁽²⁾ Our results showed nearly 50% cleavage for the same off-target mismatch, but still a significant decrease in cleavage from on-target or 19-20 base pair distal mismatches. This could be due to the relatively high RNP concentrations (600-900 nM) in our assay (for clear visibility on the gel), compared to prior studies with RNP concentrations ≤ 500 nM.⁽²⁾ Our results corresponded closely to those of prior studies with a 19-20 base pair mismatch, where off-target cleavage is tolerated by WT *GeoCas9*.⁽²⁾ Single-point mutants K267E and R332A *GeoCas9* have negligible impact on *GeoCas9* specificity (both variants follow the trend of WT *GeoCas9*, **Figure 19**), which contrasts prior work with *SpyCas9* that demonstrated robust specificity enhancement with single-point mutations in Rec.^(24,27) Additionally, a *GeoCas9* double

mutant K267E/R332A exhibits decreased on-target cleavage efficiency, which has been noted in high-specificity Cas systems.^(15,23,24,27,36-38) However, the additive effect of the K267E/R332A double mutant still does not enhance *GeoCas9* specificity in our assay.

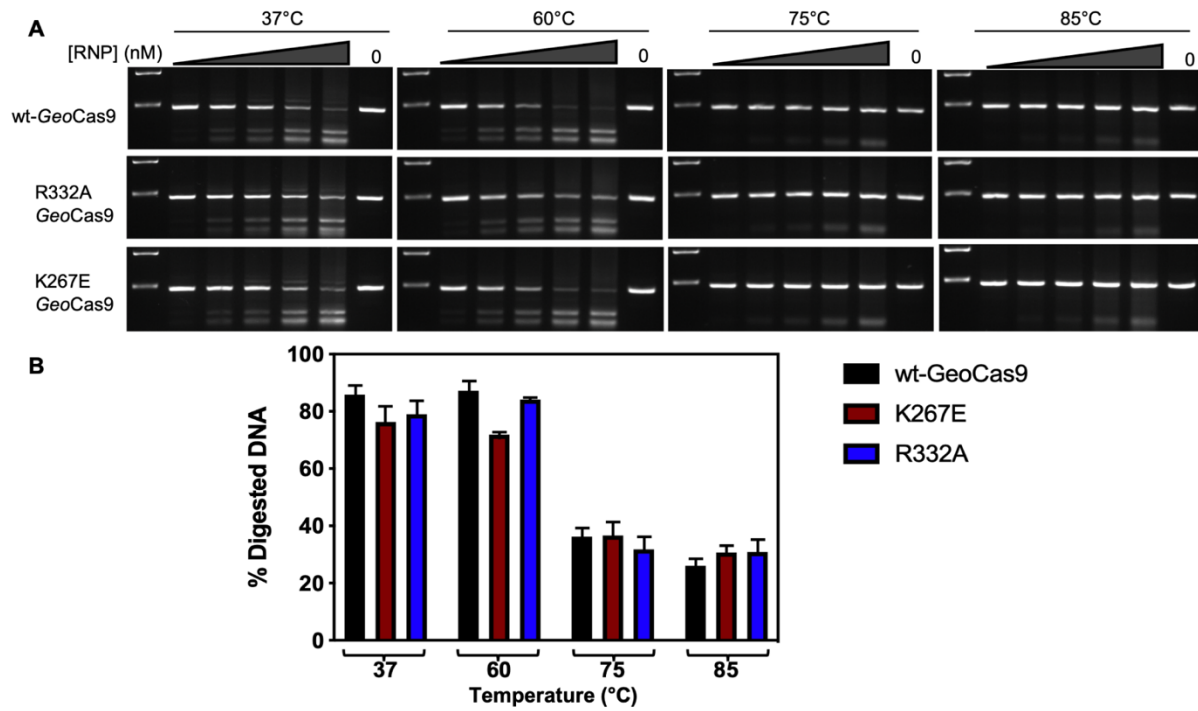


Figure 18: RNPs of varying concentrations of WT, K267E, or R332A *GeoCas9* and gRNA were incubated at 37, 60, 75, or 85 °C for 30 min, after which the RNPs were used for individual cleavage reactions. *In vitro* cleavage assays indicate no significant difference in temperature-dependent activity between WT, K267E, and R332A *GeoCas9*. **(A)** RNP concentration in each lane (left-to-right) is 100, 200, 300, 600, or 900 nM and 0 nM (control) at each temperature tested. Molecular weight markers on agarose gels (top-to-bottom) are 600 and 400 basepairs. **(B)** Graph quantifying the DNA band intensity measurements on the agarose gel (A) using ImageJ. Data plotted as mean $\pm$ SD.

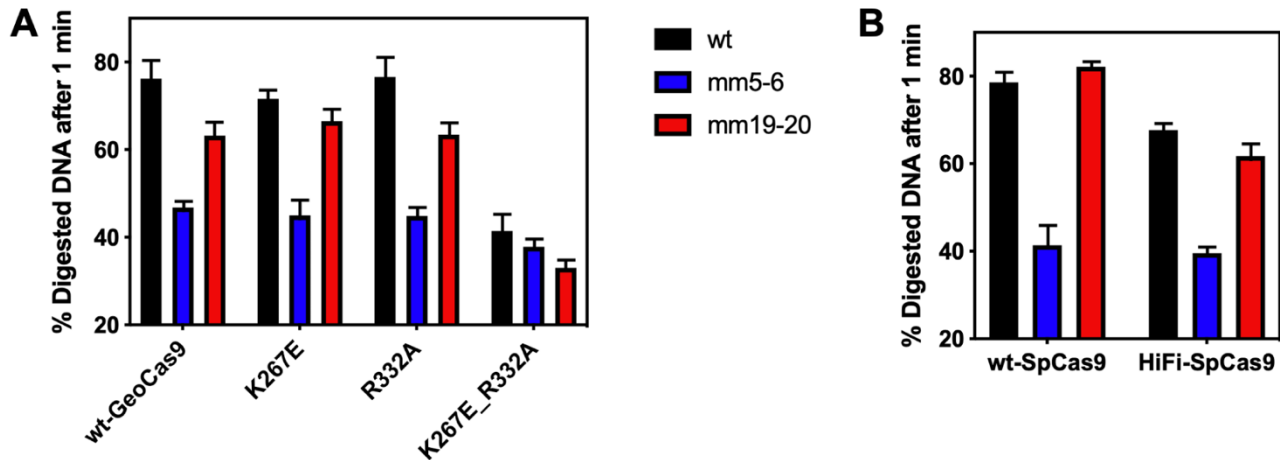


Figure 19: **(A)** Off-target *in vitro* cleavage assay with WT, K267E, R332A, and K267E/R332A *GeoCas9*. **(B)** Off-target *in vitro* cleavage assay with WT and HiFi-*SpCas9*. Data plotted as mean $\pm$ SD. Legend: WT = on-target DNA substrate at the mouse *Tnnt2* gene locus. mm5-6 = off-target DNA with the 5th and 6th nucleotide mismatches from the PAM seed site. mm19-20 = off-target DNA with the 19th and 20th nucleotide mismatches from the PAM seed site.

MST-derived binding affinities using full-length Tnnt2 gRNA and full-length WT, K267E, or R332A *GeoCas9* indicate that all three proteins have similar affinities for the gRNA used in the functional assays (**Figure 20**). Thus, mutations do not substantially alter full-length *GeoCas9* binding to Tnnt2 gRNA, supporting similar cleavage activities for these proteins. We repeated the MST experiments using a different gRNA sequence derived from the new cryo-EM structure of *GeoCas9* (PDB:8UZA), which has a different spacer sequence. We observed a similar trend, as all *GeoCas9* variants exhibited comparable affinities for this gRNA. Our functional studies illustrate an apparent resilience of *GeoCas9* to major functional changes at the level of Rec, despite comparable mutations having profound functional impacts in mesophilic Cas9s.

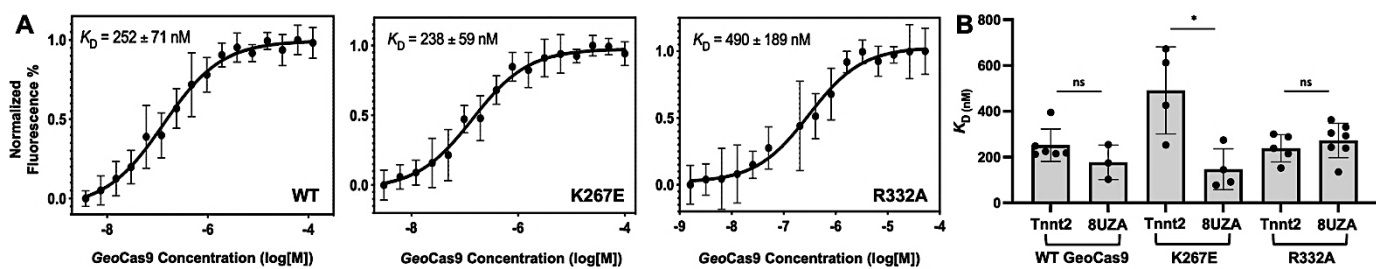


Figure 20. (A) Representative MST-derived profiles of WT, K267E, and R332A *GeoCas9* binding to a Cy5-labeled full-length Tnnt2 gRNA. **(B)** Bar graph comparing K_d values across $n \geq 3$ replicate samples are shown for Tnnt2 and 8UZA gRNA from a recent cryo-EM structure of *GeoCas9*. * $p < 0.01$

Discussion

CRISPR-Cas9 is a powerful tool for targeted genome editing with high efficiency and modular specificity.^(2,15,24,27) Allosteric signals propagate DNA binding information to the HNH and RuvC nuclease domains, facilitating their concerted cleavage of double-stranded DNA.^(8,10,14,15) The intrinsic flexibility of the nucleic acid recognition lobe plays a critical role in this information transfer, exerting a measure of conformational control over catalysis.⁽¹⁰⁾ This study provides new insights into the structural, dynamic, and functional role of the thermophilic *GeoCas9* recognition lobe. Novel constructs of subdomains *GeoRec1* and *GeoRec2*, as well as intact *GeoRec* show a high structural similarity to the domains in full-length *GeoCas9*, facilitating solution NMR experiments that captured the intrinsic allosteric motions across *GeoRec2*. These studies revealed the existence of μ s-ms timescale motions that are classically associated with allosteric signaling and enzyme function, which span the entire *GeoRec2* domain to its interfaces with *GeoRec1* and the adjacent *GeoHNH* domain.

Based on homology to specificity-enhancing variants of the better studied *SpyCas9*, the biophysical and biochemical consequences of two mutations were tested in *GeoRec2*, the larger *GeoRec* lobe, and full-length *GeoCas9*. We speculated that removing positively charged residues with potential to interact with negatively charged nucleic acids could disrupt *GeoCas9*-gRNA complex formation, stability, and subsequent function by altering the protein or nucleic acid motions. Indeed, CPMG relaxation dispersion experiments revealed that mutations enhanced and reorganized the μ s-ms flexibility of *GeoRec2*. Further, NMR titrations showed the affinity of K267E and R332A *GeoRec* for 39nt portions of two gRNAs to be weaker than that of WT *GeoRec*, consistent with MST-derived K_d values using the isolated domain. The mutations also diminished

the stability of the full-length *GeoCas9* RNP complex, though this could result from an allosteric effect that destabilizes the *GeoCas9* structure without appreciably altering gRNA binding.

The collective changes to protein dynamics, gRNA binding, and RNP thermostability suggested that mutations could modulate *GeoCas9* function, as observed in similar studies of *SpyCas9* reporting that gRNA dynamics, affecting the potential for the RNA:DNA hybrid to dissociate, have affected function.^(11,15,39) Yet, the functional impact of single-point and double mutations in this work were negligible, despite homologous K-to-E and R-to-A single point mutations enhancing specificity of the mesophilic *SpyCas9*. The biophysical impact of mutations within *GeoRec2* and *GeoRec* may be tempered by its evolutionary resilience and the highly stable neighboring domains in the context of full-length *GeoCas9*, reflected in an unchanged affinity for target DNA once the RNP was formed.^(40,41) Thus, a greater number of additive (or synergistic) mutations within *GeoRec* would be required to fine-tune activity or specificity to a large degree.

It should be noted that the effects of these and other *GeoRec* mutations may vary *in vivo* or with alternative target cleavage sites and cell types. Such studies will be the subject of future work, as will biochemical assays of homologous mutations across diverse Cas9s, which have contributed to the wide use of CRISPR technology.⁽⁴²⁾ We also note that despite the homology between *GeoRec2* and *SpyRec3* and the latter's role in evo- and HiFi-*SpyCas9* variants that inspired the K267E and R332A mutations, the maximally enhanced *SpyCas9* variants contain four mutations each. Presumably each individual substitution plays a small role modulating specificity. However, there is no consistent pattern that discerns whether multiple mutations will have additive or synergistic impacts on Cas9 function. NMR and MD studies of high-specificity *SpyCas9* variants (HF-1, Hypa, and Evo, each with distinct mutations in the *SpyRec3* domain) reveal universal structural and dynamic variations in regions of *SpyRec3* that interface with the

RNA;DNA hybrid.⁽¹⁶⁾ Notably, a recently published variant, i*GeoCas9*,⁽³²⁾ demonstrated enhanced genome-editing capabilities in HEK293T cells with eight mutations, though none in the Rec2 subdomain. This study highlighted the functional adaptability of i*GeoCas9* under low magnesium conditions, a trait beneficial in mammalian cells, distinguishing it from WT *GeoCas9*. These very recently published data, as well as the findings reported here still advance our molecular understanding of the functional handles in *GeoCas9*, relevant to the design of new enhanced variants.

This study marks the first phase of mapping allosteric motions and pathways of information flow in the *GeoRec* lobe with solution NMR experiments. Such information transfer is critical to the crosstalk between Rec and HNH in several Cas9s. Despite NMR advancements in perdeuteration,⁽⁴³⁾ transverse relaxation-optimized spectroscopy (TROSY),⁽⁴⁴⁾ and sparse isotopic labeling,⁽⁴⁵⁾ per-residue dynamics underlying allosteric signaling in large multi-domain proteins such as *GeoCas9* (~126 kDa) have remained challenging to characterize. Novel cryo-EM structures of *GeoCas9*⁽³²⁾ will facilitate the merging of future NMR and MD simulation studies to report on RNP dynamics and atomic level networks of communication. The identification of additional (or synergistic) allosteric hotspots within *GeoRec* using an integrated workflow will help to further resolve the balance between structural flexibility and the unusually high stability of *GeoCas9*, leading to new insight into targeted manipulation of RNA affinity and enhanced variants.

Here, we set out to biophysically characterize the Rec lobe of *GeoCas9* to obtain new understanding of its function (in the context of well-studied mesophilic Cas9s). Using an AlphaFold2 model, and later a cryo-EM structure of *GeoCas9*, we introduced mutations based on proximal gRNA interactions and homology to specificity-enhancing sites in *SpyCas9*. However,

the mutations did not affect *GeoCas9* function as expected, highlighting the complicated interplay between the biophysics of mesophilic and thermophilic Cas enzymes and the difficulty of applying universal functional predictions to Cas9. The very recent report of the *iGeoCas9* variant further reinforces this point.⁽²¹⁾ While high-specificity *SpyCas9* variants are heavily mutated in Rec3 (analogous to *GeoRec2*), *iGeoCas9* lacks mutations in Rec2 entirely, raising new questions about the functional role of *GeoRec*. MD simulations of WT *GeoCas9*, *iGeoCas9*, and the Rec variants revealed that while K267E and R332A induce dynamic effects on a similar trajectory to *iGeoCas9*, a true high-specificity variant samples a very wide conformational space with displacements of both Rec and HNH (**Figures 16 and 17**). Through further study of the fundamental mechanism of *GeoCas9*, it remains possible that engineering of *GeoRec* may produce high-specificity variants.

Materials and Methods

Expression and purification of GeoRec1, GeoRec2, GeoRec, and GeoCas9

The *GeoRec1* (residues 90-225) and *GeoRec2* (residues 245-456) subdomains, as well as the entire *GeoRec* lobe (residues 90-456) of *G. stearothermophilus* Cas9 were engineered into a pET28a vector with a N-terminal His₆-tag and a TEV protease cleavage site. The K267E and R332A mutations were separately introduced into the *GeoRec2* plasmid. Plasmids were transformed into BL21 (DE3) cells (New England Biolabs). Protein samples for CD spectroscopy, MST, and functional assays were grown in Lysogeny Broth (LB, Fisher), while isotopically labeled samples for NMR were grown in M9 minimal media (deuterated for *GeoRec2* and *GeoRec*) containing CaCl₂, MgSO₄, MEM vitamins, and 1.0 g/L ¹⁵N ammonium chloride and 2.0 g/L ¹³C glucose (Cambridge Isotope Laboratories), as the sole nitrogen and carbon sources, respectively. Cells were induced with 1 mM IPTG after reaching an OD₆₀₀ of 0.8–1.0 and grown

for 4 hours at 37 °C post induction. The cells were harvested by centrifugation, resuspended in a buffer of 50 mM Tris-HCl, 250 mM NaCl, 5 mM imidazole, and 1 mM PMSF at pH 7.4, lysed by ultrasonication, and purified by Ni-NTA affinity chromatography. Following TEV proteolysis of the terminal His-tag, the samples were further purified on a Superdex75 size exclusion column. NMR samples were dialyzed into a buffer containing 20 mM NaPi, 80mM KCl, 1mM DTT, and 1mM EDTA at pH 7.4.

The full-length *GeoCas9* plasmid was acquired from Addgene (#87700), expressed in TB media and was expressed and purified as previously described.⁽²⁾ The K267E, R332A, and K267E/R332A variants were introduced into full-length *GeoCas9* by modifying the original plasmid acquired from Addgene.

NMR spectroscopy

Backbone resonance assignments of *GeoRec1* and *GeoRec2* were carried out on a Bruker Avance NEO 600 MHz spectrometer at 25 °C. The following triple resonance experiments were collected for each sample: ¹H-¹⁵N TROSY-HSQC, HNCA, HN(CO)CA, HN(CA)CB, HN(COCA)CB, HN(CA)CO and HNCO. All spectra were processed in NMRPipe⁽⁴⁶⁾ and analyzed in Sparky⁽⁴⁷⁾. Three-dimensional correlations and assignments were made in CARA⁽⁴⁸⁾ and *GeoRec1* and *GeoRec2* backbone assignments were deposited in the BMRB under accession numbers 52363 and 51197, respectively. Backbone resonance assignments of *GeoRec* were completed by transferring assignments from the individually assigned spectra of *GeoRec1* and *GeoRec2*, as done previously for other large Cas9 fragments.^(49,50)

NMR spin relaxation experiments were carried out in a temperature-compensated manner at 600 and 850 MHz on Bruker Avance NEO and Avance III HD spectrometers, respectively. CPMG experiments were adapted from the report of Palmer and coworkers⁽⁵¹⁾ with a constant

relaxation period of 20 ms and ν_{CPMG} values of 0, 25, 50, 75, 100, 150, 250, 500, 750, 800, 900, and 1000 Hz. Exchange parameters were obtained from global fits of the data carried out with RELAX⁽⁵²⁾ using the R2eff, NoRex, and CR72 models, as well as in-house fitting in GraphPad Prism with the following models:

Model 1: No exchange

$$R2_{\text{eff}} = R20 \quad (1)$$

Model 2: Two-state, fast exchange (Meiboom equation ⁵³)

$$R2_{\text{eff}} = R20 + k_{\text{ex}}[1 - 4\text{CPMG}k_{\text{ex}}k_{\text{ex}}4\text{CPMG}] \quad (2)$$

Global fitting of CPMG profiles was determined to be superior to individual fits based on the Akaike Information Criterion.⁽⁵⁴⁾ Uncertainties in these rates were determined from replicate spectra with duplicate relaxation delays of 0, 25, 50 (×2), 75, 100, 150, 250, 500 (×2), 750, 800 (×2), 900, and 1000 Hz.

Longitudinal and transverse relaxation rates were measured with randomized T_1 delays of 0, 20, 60, 100, 200, 600, 800, and 1200 ms and T_2 delays of 0, 16.9, 33.9, 50.9, 67.8, 84.8, and 101.8 ms. Peak intensities were quantified in Sparky and the resulting decay profiles were analyzed in Sparky with errors determined from the fitted parameters. Uncertainties in these rates were determined from replicate spectra with duplicate relaxation delays of 20 (×2), 60 (×2), 100, 200, 600 (×2), 800, and 1200 ms for T_1 and 16.9, 33.9 (×2), 50.9 (×2), 67.8 (×2), 84.8, 101.8 (×2) ms for T_2 . Steady-state ^1H - ^{15}N NOE were measured with a 6 second relaxation delay followed by a 3 second saturation (delay) for the saturated (unsaturated) experiments and calculated by $I_{\text{sat}}/I_{\text{ref}}$. All relaxation experiments were carried out in a temperature-compensated interleaved manner.

Model-free analysis was carried out by fitting relaxation rates to five different forms of the spectral density function with local τ_m , spherical, prolate spheroid, oblate spheroid, or ellipsoid diffusion tensors.⁽⁵⁵⁻⁶⁰⁾ The criteria for inclusion of resonances in the diffusion tensor estimate was based on the method of Bax and coworkers.⁶¹ N-H bond lengths were assumed to be 1.02 Å and the ^{15}N chemical shift anisotropy tensor was -160 ppm. Diffusion tensor parameters were optimized simultaneously in RELAX under the full automated protocol.⁽⁵²⁾ Model selection was iterated until tensor and order parameters did not deviate from the prior iteration.

NMR titrations were performed on a Bruker Avance NEO 600 MHz spectrometer at 25 °C by collecting a series of ^1H - ^{15}N TROSY HSQC spectra with increasing ligand (*i.e.* RNA) concentration. The ^1H and ^{15}N carrier frequencies were set to the water resonance and 120 ppm, respectively. Samples of WT, K267E, and R332A *GeoRec* were titrated with a 39nt portion of a gRNA until no further spectral perturbations were detected. NMR chemical shift perturbations were calculated as:

$$\Delta\delta=(\Delta\delta_{\text{HN2}}+\Delta\delta_{\text{NH225}})/2$$

Microscale thermophoresis (MST)

MST experiments were performed on a Monolith X instrument (NanoTemper Technologies), quantifying WT, K267E, R332A, and K267E/R332A *GeoRec* binding to a 39-nt Cy5-labeled RNA at a concentration of 20 nM in a buffer containing 20 mM sodium phosphate, 150 mM KCl, 5 mM MgCl_2 , and 0.1% Triton X-100 at pH 7.6. The *GeoRec* proteins were serially diluted from a 200 μM stock into 16 microcentrifuge tubes and combined in a 1:1 molar ratio with serially diluted gRNA from a 40 nM stock. After incubation for 5 minutes at 37 °C in the dark, each sample was loaded into a capillary for measurement. K_d values for the various complexes

were calculated using the MO Control software (NanoTemper Technologies). Statistical significance was calculated using a two-tailed T-test.

Circular dichroism (CD) spectroscopy

All *GeoCas9* and *GeoRec* proteins were buffer exchanged into a 20 mM sodium phosphate buffer at pH 7.5, diluted to 1 μ M, and loaded into a 2 mm quartz cuvette (JASCO instruments). A CD spectrum was first measured between 200 - 250 nm, after which the sample was progressively heated from 20 – 90 °C in 1.0 °C increments while ellipticity was monitored at 222 and 208 nm. Phosphate buffer baseline spectra were subtracted from the sample measurements. Prior to CD measurements, *GeoCas9*-RNP was formed by incubating 3 μ M *GeoCas9* with its gRNA at a 1:1.5 molar ratio at 37 °C for 10 minutes. The unfolding CD data was fit in GraphPad Prism to:

$$\text{Ellipticity}(T)=[(mfT+bf)+(muT+bu)]\exp[(-\Delta H_D,vHR)(1T-1T_m)]/1+\exp[(-\Delta H_D,vHR)(1T-1T_m)]$$

X-ray crystallography

GeoRec protein purified as described above was crystallized by sitting drop vapor diffusion at room temperature by mixing 1.0 μ L of 15 mg/mL *GeoRec* in a buffer of 20 mM HEPES and 100 mM KCl at pH 7.5 with 2.0 μ L of crystallizing condition: 0.15 M calcium chloride, 15 % polyethylene glycol 6000, 0.1 M HEPES at pH 7.0. Crystals were cryoprotected in crystallizing condition supplemented with 30% ethylene glycol. Diffraction images were collected at the NSLS-II AMX beamline at Brookhaven National Laboratory under cryogenic conditions. Images were processed using XDS⁶² and Aimless in CCP4.⁶³ Chain A of the *N. meningitidis* Cas9 X-ray structure (residues 249-445 only, PDB ID: 6JDQ) was used for molecular replacement with Phaser followed by AutoBuild in Phenix.⁶⁴ Electron density was only observed for the *GeoRec2* subdomain. The *GeoRec2* structure was finalized through manual building in Coot⁶⁵ and refinement in Phenix.

Molecular dynamics (MD) simulations

Molecular Dynamics (MD) simulations were based on the cryo-EM structure of full-length *GeoCas9* (PDB: 8UZA, resolution 3.17 Å) in complex with gRNA and target DNA with two mutations in *GeoCas9* (at residues 8 and 582). Four systems were considered for the MD studies: WT, K267E, R332A, K267E/R332A and *iGeoCas9*. We generated the WT *GeoCas9* by back-mutating A8D and A582H from the cryo-EM structure (PDB: 8UZA), followed by introducing the mutations K267E, R332A, or a double mutation (with both K267E and R332A) for the variant systems. Subsequently, we performed MD simulation of *iGeoCas9* (PDB: 8UZB, resolution 2.63 Å) consisting of 10 mutations (D8A, E149G, T182I, N206D, P466Q, H582A, Q817R, E843K, E854G, K908R). All systems were solvated with explicit water in a periodic box of ~ 134 Å $\times$ ~ 154 Å $\times$ ~ 151 Å resulting in $\sim 276,000$ atoms. Counter ions were added to neutralize the systems. MD simulations were performed using a protocol tailored for protein-nucleic acid complexes,⁶⁶ previously applied in studies of CRISPR-Cas systems.⁶⁷⁻⁶⁹ All the simulations were performed by using Amber ff19SB force field for protein,⁷⁰ ff99bsc1 corrections and χ OL3 corrections for DNA and RNA, respectively.^{71,72} Water molecules were described by TIP3P model.⁷³ All bonds involving hydrogens were constrained using the LINCS algorithm. A particle mesh Ewald method (PME) with a 10 Å cutoff was used to calculate electrostatics. Energy minimization was performed to relax the water molecules and counterions, keeping the protein-nucleic acid complex fixed with harmonic potential restraints of 100 kcal/mol Å². Equilibration was performed by gradually increasing the temperature from 0 to 100 K and then to 200 K in canonical NVT ensemble and isothermal-isobaric NPT ensemble. A final temperature of 300 K was maintained via Langevin dynamics with a collision frequency $\gamma = 1/\text{ps}$ and a reference pressure of 1 atm was achieved through Berendsen barostat. Production runs were carried out in NVT ensemble for 2 μs for each

system in three replicates, resulting in 6 μ s per system (totaling 30 μ s for all systems). The equations of motion were integrated with the leapfrog Verlet algorithm with a time step of 2 fs. All simulations were conducted using the GPU-empowered version of AMBER 22.⁷⁴ Analysis was performed on the aggregate ensemble (i.e., $\sim$ 6 μ s per system).

To characterize the protein-nucleic acid interactions in all the systems under investigation, we performed contact analysis. A contact was considered to form between two atoms within a cutoff distance of ≤ 4.5 Å. The binding free energy of gRNA and DNA with *GeoCas9* was calculated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method.⁷⁵ ⁷⁷ This approach was used to compare the Rec-gRNA binding affinity of WT *GeoCas9* with its mutants. For each system, the binding energies were calculated over the $\sim$ 200 ns ensemble of the stable trajectories at an interval of $\sim$ 20 ns.

DNA cleavage assays

GeoCas9 gRNA templates containing 21-nt spacers targeting the mouse *Tnnt2* gene locus were introduced into EcoRI and BamHI sites in pUC57 (Genscript). The plasmid was transformed into BL21(DE3) cells (New England BioLabs) and subsequent restriction digest of the plasmid DNA was carried out using the BamHI restriction enzyme (New England BioLabs) according to the manufacturer's instructions. Linearized plasmid DNA was immediately purified using the DNA Clean and Concentrator-5 kit (Zymo Research) according to the manufacturer's instructions. RNA transcription was performed *in vitro* with the HiScribe T7 High Yield RNA Synthesis Kit (New England BioLabs). DNA substrates containing the target cleavage site (479 base pairs, **Figure S13**) were produced by polymerase chain reaction (PCR) using mouse genomic DNA as a template and primer pairs 5'CAAAGAGCTCCTCGTCCAGT3' and 5'ATGGACTCCAGGACCCAAGA3' followed by a column purification using the NucleoSpin®

Gel and PCR Clean-up Kit (Macherey-Nagel). For the *in vitro* activity assay, RNP formation was achieved by incubating 3 μ M *GeoCas9* (WT, K267E, R332A, or K267E/R332A mutant) and 3 μ M gRNA at either 37 °C, 60 °C, 75 °C, or 85 °C for 30 minutes in a reaction buffer of 20 mM Tris, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 5% glycerol at pH 7.5. The 10 μ L cleavage reactions were set up by mixing RNP at varying concentrations with 149 nanograms of PCR products on ice followed by incubation at 37 °C for 30 minutes. The reaction was quenched with 1 μ L of proteinase K (20 mg/mL) and subsequent incubation at 56 °C for 10 minutes. 6x DNA loading buffer was added to each reaction and 10 μ L of reaction mixture per lane was loaded onto an agarose gel. DNA band intensity measurements were carried out with ImageJ.

For *in vitro* off-target activity assays, RNP formation was achieved by incubating 10 μ M *GeoCas9* (WT, K267E, R332A, or K267E/R332A mutant) and 10 μ M gRNA at 37 °C for 30 minutes in the reaction buffer described above. The 10 μ L cleavage reactions were set up by mixing 1 μ M RNP with 150 nanograms of PCR products on ice followed by incubation at 37 °C for varying time points. The reaction was quenched with 1 μ L of proteinase K (20 mg/mL) and subsequent incubation at 56 °C for 10 minutes. 6x DNA loading buffer was added to each reaction and 10 μ L of reaction mixture per lane was loaded onto an agarose gel. DNA band intensity measurements were carried out with ImageJ. WT and HiFi *SpyCas9* control proteins were purchased from Integrated DNA Technologies (IDT, cat. No. 108158 and No. 108160, respectively), as was the associated *SpyCas9* gRNA, Alt-R™ CRISPR-Cas9 gRNA, with an RNA spacer sequence complementing 5'- TGGACAGAGCCTTCTTCTTC-3'. The on-target and off-target DNA sequences used for the *SpyCas9 in vitro* cleavage assay can be found in **Table S3**.

Acknowledgments

This work was supported by NIH grant R01 GM 136815 (to GP and GPL) and NSF grant MCB 2143760 (to GPL). GP acknowledges support from the NIH (Grant No. R01GM141329) and the NSF (CHE- 2144823), as well as from the Sloan Foundation (FG-2023-20431) and the Camille and Henry Dreyfus Foundation (TC-24-063). This research used the AMX beamline of the National Synchrotron Light Source II, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Brookhaven National Laboratory under Contract No. DE-SC0012704. The Center for BioMolecular Structure (CBMS) is primarily supported by NIGMS through a Center Core P30 Grant (P30 GM133893), and by the DOE Office of Biological and Environmental Research (KP1607011). Computational studies were carried out using Expanse at the San Diego Supercomputing Center through allocation MCB160059 and Bridges2 at the Pittsburgh Supercomputer Center through allocation BIO230007 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by NSF grants #2138259, #2138286, #2138307, #2137603, and #2138296.

References

- 1 Belato, H. B. *et al.* Disruption of electrostatic contacts in the HNH nuclease from a thermophilic Cas9 rewires allosteric motions and enhances high-temperature DNA cleavage. *The Journal of Chemical Physics* **157**, 225103, doi:10.1063/5.0128815 (2022).
- 2 Harrington, L. B. *et al.* A thermostable Cas9 with increased lifetime in human plasma. *Nat Commun* **8**, 1424, doi:10.1038/s41467-017-01408-4 (2017).
- 3 Zhang, S. *et al.* Current Trends of Clinical Trials Involving CRISPR/Cas Systems. *Front. Med.* **10**, DOI: 10.3389/fmed.2023.1292452 (2023).
- 4 Li, T. *et al.* CRISPR/Cas Therapeutics: Progress and Prospects. *Signal Trans. Targ. Ther.* **8**, DOI: 10.1038/s41392-41023-01309-41397 (2023).
- 5 Kim, E. *et al.* In vivo genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat Commun* **8**, 14500, doi:10.1038/ncomms14500 (2017).
- 6 Lee, C. M., Cradick, T. J. & Bao, G. The *Neisseria meningitidis* CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. *Mol Ther* **24**, 645-654, doi:10.1038/mt.2016.8 (2016).
- 7 Belato, H. B. *et al.* Structural and dynamic insights into the HNH nuclease of divergent Cas9 species. *J Struct Biol* **214**, 107814, doi:10.1016/j.jsb.2021.107814 (2021).
- 8 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
- 9 Eggers, A. R. *et al.* Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *Cell* **187**, 3249-3261.e3214, doi:10.1016/j.cell.2024.04.031 (2024).

- 10 Palermo, G. *et al.* Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q Rev Biophys* **51**, doi:10.1017/s0033583518000070 (2018).
- 11 Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027, doi:10.1126/sciadv.aao0027 (2017).
- 12 Mir, A., Edraki, A., Lee, J. & Sontheimer, E. J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **13**, 357-365, doi:10.1021/acscchembio.7b00855 (2018).
- 13 Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481, doi:10.1126/science.aab1452 (2015).
- 14 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113, doi:10.1038/nature15544 (2015).
- 15 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410, doi:10.1038/nature24268 (2017).
- 16 Skeens, E. *et al.* High-fidelity, hyper-accurate, and evolved mutants rewire atomic-level communication in CRISPR-Cas9. *Sci Adv* **10**, eadl1045, doi:10.1126/sciadv.adl1045 (2024).
- 17 East, K. W. *et al.* Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *Journal of the American Chemical Society*, doi:10.1021/jacs.9b10521 (2020).
- 18 Palermo, G., Miao, Y., Walker, R. C., Jinek, M. & McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent Sci* **2**, 756-763, doi:10.1021/acscentsci.6b00218 (2016).

- 19 Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol Ther* **18**, 80-86, doi:10.1038/mt.2009.255 (2010).
- 20 Shen, P. *et al.* Structure of *Geobacillus stearothermophilus* Cas9: Insights into the Catalytic Process and Thermostability of CRISPR-Cas9. *ACS Catalysis* **14**, 13227-13235 (2024).
- 21 Eggers, A. R. *et al.* Rapid DNA Unwinding Accelerates Genome Editing by Engineered CRISPR-Cas9. *Cell* **187**, 3249-3261 (2024).
- 22 Lisi, G. P., East, K. W., Batista, V. S. & Loria, J. P. Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. *Proc Natl Acad Sci U S A* **114**, E3414-e3423, doi:10.1073/pnas.1700448114 (2017).
- 23 Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016).
- 24 Vakulskas, C. A. *et al.* A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat Med* **24**, 1216-1224, doi:10.1038/s41591-018-0137-0 (2018).
- 25 Guzman, I., Ghaemi, Z., Baranger, A., Luthey-Schulten, Z. & Gruebele, M. Native Conformational Dynamics of the Spliceosomal U1A Protein. *The Journal of Physical Chemistry B* **119**, 3651-3661, doi:10.1021/jp511760m (2015).
- 26 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).
- 27 Casini, A. *et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* **36**, 265-271, doi:10.1038/nbt.4066 (2018).
- 28 Ricci, C. G. *et al.* Deciphering Off-Target Effects in CRISPR-Cas9 through Accelerated Molecular Dynamics. *ACS Cent Sci* **5**, 651-662, doi:10.1021/acscentsci.9b00020 (2019).

- 29 Nierzwicki, L. *et al.* Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *Elife* **10**, doi:10.7554/eLife.73601 (2021).
- 30 Nierzwicki, Ł. *et al.* Principles of target DNA cleavage and the role of Mg²⁺ in the catalysis of CRISPR–Cas9. *Nature Catalysis* **5**, 912-922, doi:10.1038/s41929-022-00848-6 (2022).
- 31 Shen, P. *et al.* Structure of *Geobacillus stearothermophilus* Cas9: Insights into the Catalytic Process and Thermostability of CRISPR-Cas9. *ACS Catalysis* **14**, 13227-13235, doi:10.1021/acscatal.4c03278 (2024).
- 32 Eggers, A. R. *et al.* Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *bioRxiv*, 2023.2012.2014.571777, doi:10.1101/2023.12.14.571777 (2023).
- 33 Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229-236, doi:10.1038/s41586-020-1978-5 (2020).
- 34 Palermo, G. *et al.* Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J Am Chem Soc* **139**, 16028-16031, doi:10.1021/jacs.7b05313 (2017).
- 35 Hou, Z. *et al.* Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15644-15649, doi:10.1073/pnas.1313587110 (2013).
- 36 Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).
- 37 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63, doi:10.1038/nature26155 (2018).
- 38 Lee, J. K. *et al.* Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat Commun* **9**, 3048, doi:10.1038/s41467-018-05477-x (2018).

- 39 Ma, E., Harrington, L. B., O'Connell, M. R., Zhou, K. & Doudna, J. A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**, 398-407, doi:10.1016/j.molcel.2015.10.030 (2015).
- 40 Nguyen, V. *et al.* Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* **355**, 289-294, doi:10.1126/science.aah3717 (2017).
- 41 Saavedra, H. G., Wrabl, J. O., Anderson, J. A., Li, J. & Hilser, V. J. Dynamic allostery can drive cold adaptation in enzymes. *Nature* **558**, 324-328, doi:10.1038/s41586-018-0183-2 (2018).
- 42 Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017).
- 43 Venters, R. A., Farmer, B. T., 2nd, Fierke, C. A. & Spicer, L. D. Characterizing the use of perdeuteration in NMR studies of large proteins: ¹³C, ¹⁵N and ¹H assignments of human carbonic anhydrase II. *Journal of molecular biology* **264**, 1101-1116, doi:10.1006/jmbi.1996.0699 (1996).
- 44 Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12366-12371, doi:10.1073/pnas.94.23.12366 (1997).
- 45 Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* **1**, 749-754, doi:10.1038/nprot.2006.101 (2006).
- 46 Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293, doi:10.1007/bf00197809 (1995).

- 47 Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325-1327, doi:10.1093/bioinformatics/btu830 (2015).
- 48 Keller, R. L. J. *Optimizing the process of nuclear magnetic resonance spectrum analysis and computer aided resonance assignment* Doctoral Thesis thesis, ETH Zürich, (2005).
- 49 Nerli, S., De Paula, V. S., McShan, A. C. & Sgourakis, N. G. Backbone-independent NMR Resonance Assignments of Methyl Probes in Large Proteins. *Nat. Commun.* **12**, 691 (2021).
- 50 De Paula, V. S., Dubey, A., Arthanari, H. & Sgourakis, N. G. Dynamic Sampling of a Surveillance State Enables DNA Proofreading by Cas9. *Cell Chem. Biol.* doi.org/10.1016/j.chembiol.2024.10.001 (2024).
- 51 Loria, J. P., Rance, M. & Palmer, A. G. A Relaxation-Compensated Carr–Purcell–Meiboom–Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *Journal of the American Chemical Society* **121**, 2331-2332, doi:10.1021/ja983961a (1999).
- 52 Bieri, M., d'Auvergne, E. J. & Gooley, P. R. relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μ s motion of proteins. *J Biomol NMR* **50**, 147-155, doi:10.1007/s10858-011-9509-1 (2011).
- 53 Luz, Z. & Meiboom, S. Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *The Journal of Chemical Physics* **39**, 366-370, doi:10.1063/1.1734254 (1963).
- 54 Cavanaugh, J. E. & Neath, A. A. The Akaike Information Criterion: Background, Derivation, Properties, Application, Interpretation, and Refinements. *WIREs Comp. Mol. Sci.* **11**, e1460 (2019).

- 55 Brüschweiler, R., Liao, X. & Wright, P. E. Long-range Motional Restrictions in a Multidomain Zinc-finger Protein from Anisotropic Tumbling. *Science* **268**, 886-889 (1995).
- 56 Mandel, A. M., Akke, M. & Palmer, A. G. Backbone Dynamics of Escherichia coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **246**, 144-163 (1995).
- 57 Fushman, D., Cahill, S. & Cowburn, D. The Main-chain Dynamics of the Dynamin Pleckstrin Homology (PH) Domain in Solution: Analysis of ¹⁵N Relaxation with Monomer/Dimer Equilibration. *J. Mol. Biol.* **266**, 173-194 (1997).
- 58 Orekhov, V. Y., Korzhnev, D. M., Diercks, T., Kessler, H. & Arseniev, A. S. ¹H-¹⁵N NMR Dynamic Study of an Isolated α -helical Peptide (1-36)bacteriorhodopsin Reveals the Equilibrium Helix-coil Transitions. *J. Biomol. NMR* **14**, 345-356 (1999).
- 59 Korzhnev, D. M., Billeter, M., Arseniev, A. S. & Orekhov, V. Y. NMR Studies of Brownian Tumbling and Internal Motions in Proteins. *Prog. NMR Spectros.* **38**, 197-266 (2001).
- 60 Zhuravleva, A. V. *et al.* Gated Electron Transfers and Electron Pathways in Azurin: A NMR Dynamic Study at Multiple Fields and Temperatures. *J. Mol. Biol.* **342**, 1599-1611 (2004).
- 61 Tjandra, N., Feller, S. E., Pastor, R. W. & Bax, A. Rotational Diffusion Anisotropy of Human Ubiquitin from N-¹⁵ NMR Relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566 (1995).
- 62 Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/s0907444909047337 (2010).
- 63 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/s0907444910045749 (2011).

- 64 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877, doi:10.1107/s2059798319011471 (2019).
- 65 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/s0907444910007493 (2010).
- 66 Sinha, S., Pindi, C., Ahsan, M., Arantes, P. R. & Palermo, G. Machines on Genes through the Computational Microscope. *J. Chem. Theor. Comput.* **19**, 1945-1964 (2023).
- 67 Saha, A. *et al.* An Alpha-helical Lid Guides the Target DNA Toward Catalysis in CRISPR-Cas12a. *Nat. Commun.* **15**, DOI: 10.1038/s41467-41024-45762-41466 (2024).
- 68 Arantes, P. R. *et al.* Dimerization of the Deaminase Domain and Locking Interactions with Cas9 Boost Base Editing Efficiency in ABE8e. *Nucl. Acids Res.* **52**, 13931-13944 (2023).
- 69 Sinha, S. *et al.* Unveiling the RNA-mediated Allosteric Activation Discloses Functional Hotspots in CRISPR-Cas13a. *Nucl. Acids Res.* **52**, 906-920 (2024).
- 70 Tian, C. *et al.* fl9SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theor. Comput.* **16**, 528-552 (2019).
- 71 Galindo-Murillo, R. *et al.* Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theor. Comput.* **12** (2016).
- 72 Zgarbova, M. *et al.* Refinement of the Cornell *et al.* Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theor. Comput.* **7**, 2886-2902 (2011).
- 73 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926-935 (1983).

- 74 Case, D. A. *et al.* The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **26**, 1668-1688, doi:10.1002/jcc.20290 (2005).
- 75 Mongan, J., Simmerling, C., McCammon, J. A., Case, D. A. & Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theor. Comput.* **3**, 156-169 (2007).
- 76 Nguyen, H., Pérez, A., Bermeo, S. & Simmerling, C. Refinement of Generalized Born Implicit Solvation Parameters for Nucleic Acids and their Complexes with Proteins. *J. Chem. Theor. Comput.* **11**, 3714-3728 (2015).
- 77 Nguyen, H., Roe, D. R. & Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theor. Comput.* **9**, 2020-2034 (2013).

Chapter Three

Orthosteric and Allosteric Effects of Anti-CRISPR II-C1 Inhibition of *GeoCas9* from Integrated Structural Biophysics

This chapter will be published as an original research article

Alexa L. Knight,¹ Helen B. Belato,^{1,\$} Charlotte S. Dresser,¹ Chinmai Pindi,² Briana J. Mercado,¹ Praise Lasekan,¹ Jinping Luo,³ Pablo R. Arantes,² Gerwald Jogl,^{1,4,*} Giulia Palermo,^{2,*} and George P. Lisi^{1,4,*}

¹Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI USA

²Departments of Bioengineering & Chemistry, University of California Riverside, Riverside, CA USA

³Transgenic & Genome Editing Facility, Brown University, Providence, RI USA

⁴Giuliani RNA Center, Brown University, Providence, RI USA

^{\$}Present address: Department of Chemistry, Brown University, Providence, RI USA

Author Contributions

ALK: writing – original draft, data curation, formal analysis, produced *GeoHNNH*, *AcrIIC1*, *AcrIIC1* variants, and *GeoCas9* proteins and sgRNA, conducted the NMR, X-ray crystallography, and biophysical experiments. **HBB:** produced *GeoHNNH*, *AcrIIC1* proteins, conducted NMR. **CSD:** produced *GeoHNNH*, *AcrIIC1*, *AcrIIC1* variant proteins. **CP:** performed molecular dynamic simulations, MD figure development. **PRA:** performed molecular dynamic simulations. **BJM:** produced *GeoHNNH* and *AcrIIC1* proteins, conducted NMR experiments. **PL:** produced *GeoHNNH*, *AcrIIC1*, and *AcrIIC1* variant proteins, X-ray crystallography of *GeoHNNH:AcrIIC1* S78A variant, MST. **AMD:** crystal preparation and shooting of *GeoHNNH:AcrIIC1*. **GJ:** supervision – X-ray crystallography. **JL:** *in vitro* cleavage assay. **GPL:** conceptualization, supervision – NMR spectroscopy, writing – review and editing.

Abstract

Anti-CRISPRs (Acrs) are small protein inhibitors of CRISPR-Cas effectors that originate from the translated genetic material of bacteriophage. Harnessing the natural ability of Acrs to bind and disrupt CRISPR-Cas editing can provide enhanced spatiotemporal control of gene editing. Recent studies have revealed diverse structures and functions of Acrs, however, atomistic studies of the specific molecular mechanisms behind Acr inhibition are lacking. Here, we reveal how structure, function, and dynamics govern AcrIIC1 inhibition of Cas9 from *G. stearothermophilus* (*GeoCas9*) via its HNH nuclease domain. An X-ray crystal structure of the *Geo*HNH-AcrIIC1 complex reveals a conserved binding interface at the catalytic site and disruption of crucial electrostatic contacts known to modulate the thermostability of *GeoCas9*. AcrIIC1 binding also rewires the intrinsic dynamics of the *Geo*HNH domain, stimulates millisecond motions that are absent from the unliganded nuclease, and attenuates the guide RNA affinity of *GeoCas9*. Subsequent AcrIIC1 mutations in residues at its crystallographic binding interface uncouple Acr binding from inhibition, providing new insight into mechanism by which AcrIIC1 acts on *GeoCas9*.

Introduction

Applications of CRISPR-associated protein 9 (CRISPR-Cas9) have expanded from a biological tool for exploring gene editing to a promising therapeutic for human disease, including recent efforts to mitigate antibiotic resistance in bacterial infections⁽¹⁻⁶⁾. However, spatiotemporal control over CRISPR-Cas function remains a primary limitation, leading to off-target effects that could lead to large-scale genomic instability^(7, 8). Specificity-enhancing mutations and directed evolution pipelines have mitigated this issue, but many of these insights have yet to translate to Cas effectors that function outside of mesophilic regimes⁽⁹⁻¹¹⁾. An emergent thermophilic Cas9 from *Geobacillus stearothermophilus* (*GeoCas9*) has an expanded temperature range for double stranded DNA cleavage of ≥ 75 °C as well as increased thermostability and longevity in human

serum^(12, 13). Recent work has characterized some aspects of the intricate mechanism of *GeoCas9*, including that structural fluctuations on multiple timescales govern the allosteric communication between its spatially separated domains^(12, 14, 15). Most recently, cryo-EM structures have revealed a critical role for the “wedge” (WED), between the RuvC nuclease and PAM-interacting domains, in DNA unwinding and mutations in WED have produced the first high-specificity *GeoCas9*, termed i*GeoCas9*^(15, 16).

In addition to protein engineering campaigns to improve the specificity and efficiency of Cas effectors, exogenous means of spatiotemporal control have leveraged bacteriophage-encoded anti-CRISPR proteins (Acrs). Several groups have investigated the structure and biochemistry of Acrs, which emerged naturally from an “evolutionary arms race” to bind and inhibit Cas effectors by diverse mechanisms⁽¹⁷⁻²²⁾. Here we provide atomic details of an Acr originating from *Neisseria meningitidis*, AcrIIC1, and its complex with *GeoCas9*. Prior work reported AcrIIC1 to be a broad-spectrum inhibitor of related Type II-C Cas effectors from *N. meningitidis* (*NmeCas9*), *Campylobacter jejuni* (*CjeCas9*), and *GeoCas9*. While AcrIIC1 does exploit some of the conserved features between *GeoCas9* and *NmeCas9* for binding, we also report that AcrIIC1 inhibition of *GeoCas9* imparts a distinct modulation of its HNH nuclease dynamics and function⁽¹⁷⁾. As a consequence, AcrIIC1 not only inhibits cleavage activity through steric occlusion and rewiring of *GeoHNH* flexibility, but also by tampering with the guide RNA affinity of the full-length *GeoCas9*.

To visualize the structural and biochemical signatures of AcrIIC1 inhibition of *GeoCas9*, we combined nuclear magnetic resonance (NMR), X-ray crystallography, and all-atom molecular dynamics (MD) simulations with biophysical and functional assays. We resolve the interface

between *Geo*HNH and AcrIIC1 by X-ray crystallography, identifying residues S78 and C79 of AcrIIC1 that directly interact with D581 and H582 of *Geo*HNH, respectively. We describe how AcrIIC1 binding to the HNH domain of *Geo*Cas9 modulates its structure and rewires its intrinsic motions. We demonstrate that beyond a direct interaction that sequesters key residues from the *Geo*HNH active site (orthosteric), AcrIIC1 propagates long-range structural perturbations through *Geo*HNH, modulates guide RNA affinity of *Geo*Cas9, and potentially binds a secondary (allosteric) location within the *Geo*HNH domain. To assess the contribution of individual amino acids to the binding interface, we engineered S78A, C79P, and S78A/C79P mutations within AcrIIC1. The Cys-to-Pro substitution at residue 79 was designed to simultaneously disrupt multiple interactions of AcrIIC1 with the *Geo*HNH backbone observed crystallographically. NMR spectra of ^{15}N -*Geo*HNH with S78A AcrIIC1 recapitulate the perturbations of wild-type (WT) AcrIIC1, while C79P has no effect. NMR relaxation experiments point to a dynamic rewiring of *Geo*HNH, the extent of which is tuned by the AcrIIC1 mutations. Similarly, functional studies show none of our mutant variants inhibit *Geo*Cas9 cleavage activity. All-atom MD simulations of WT and mutant *Geo*HNH-AcrIIC1 complexes show that WT AcrIIC1 stabilizes the HNH catalytic loop, while AcrIIC1 mutants increase backbone flexibility and conformational heterogeneity. Consistent with functional studies, MD confirms that AcrIIC1 inhibition of *Geo*Cas9 requires both active-site occupancy and disruption of a coordinated dynamic network within *Geo*HNH, with variants suggesting that subtle perturbations to the protein-protein contacts can partially uncouple binding from inhibition.

Results

A crystal structure of the GeoHNH-AcrIIC1 complex reveals a conserved binding interface

To elucidate the binding mode of AcrIIC1 to *GeoCas9*, we determined a 1.7 Å resolution X-ray crystal structure of AcrIIC1 bound to the HNH domain of *GeoCas9*. Aligning almost identically with a previously published crystal structure of *NmeHNH* bound to AcrIIC1 (**Figure 1**), the inhibitor directly interacts with the active site of *GeoHNH*. Crucially, the active site residues D581 and H582 form hydrogen bonds with AcrIIC1 residue S78, as well as the sidechain and backbone amide of C79 (**Figure 2A**). Second sphere charged residues of *GeoHNH* further stabilize the inhibitor. Structural studies of closely related HNH domains, including *NmeHNH* that shares 44% sequence identity with *GeoHNH*, showed similar charge stabilization of anti-CRISPRs via the same residues (**Figure 2B, C**)^(17, 21).

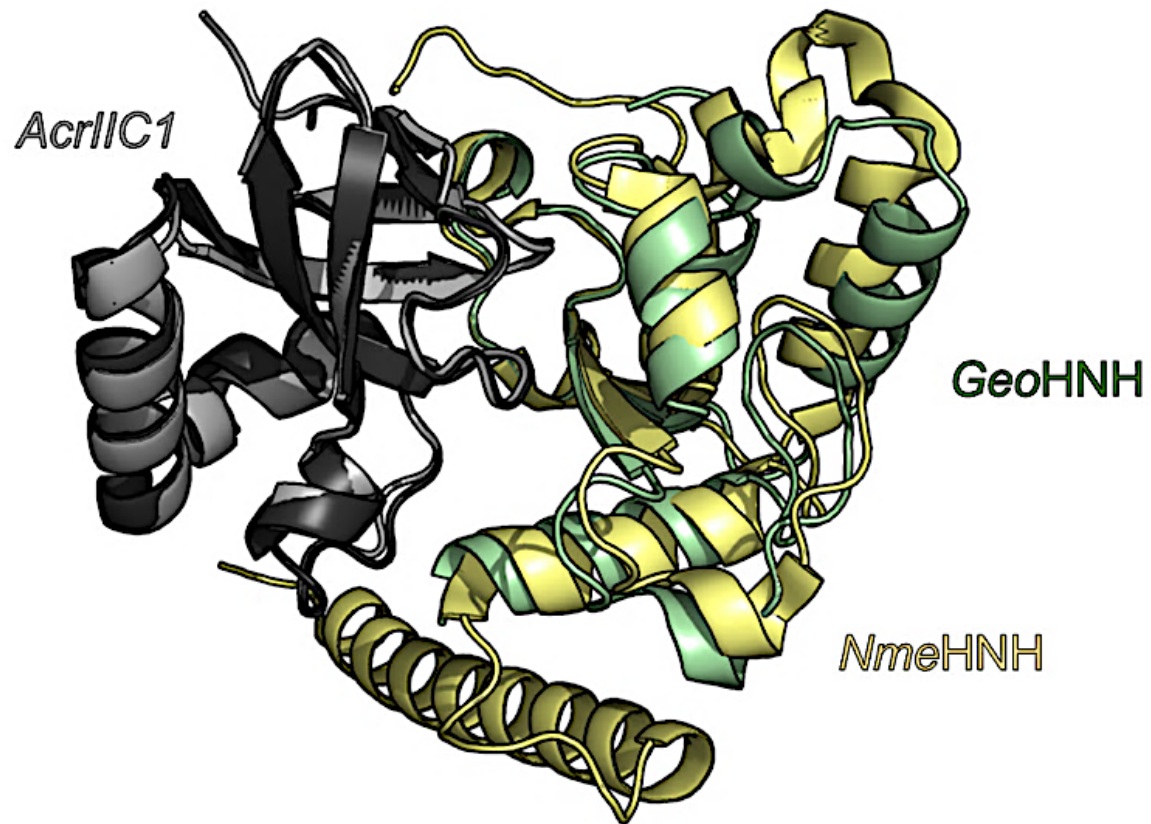


Figure 1. Overlay of X-ray crystal structures of AcrIIC1 bound to *NmeHNH* (PDB ID: 5VGB) and to *GeoHNH* (PDB ID: 10VC). A very high degree of similarity is noted.

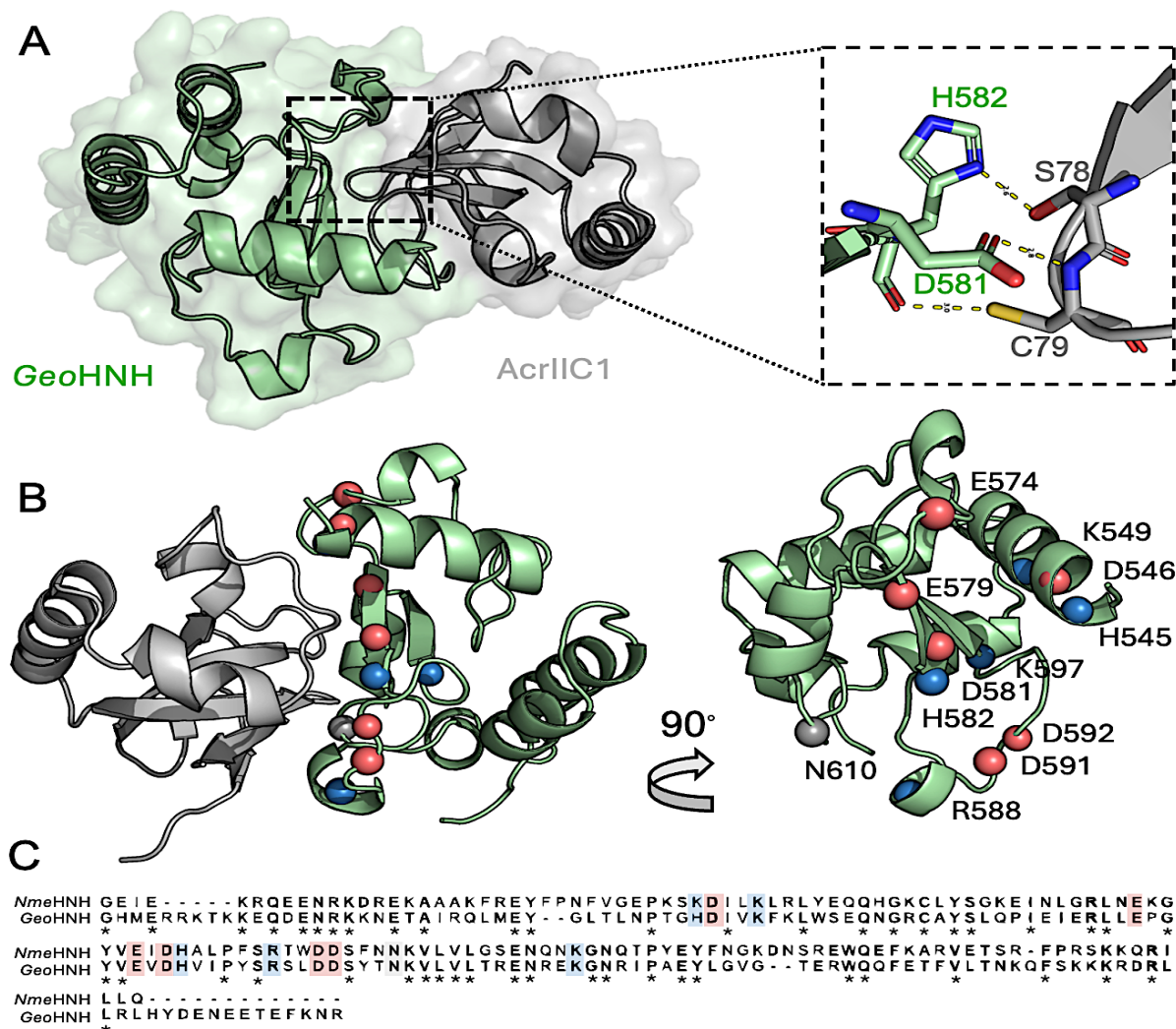


Figure 2. (A) X-ray crystal structure of the *GeoHNNH*-AcrIIC1 complex. The inset highlights a direct interaction of AcrIIC1 with the *GeoHNNH* catalytic residues. **(B)** The X-ray structure positions conserved charged residues at the HNNH-Acr interface to stabilize the complex, as in the AcrIIC1 complex with *NmeHNNH*, a *GeoCas9* ortholog. **(C)** Sequence alignments of *NmeHNNH* and *GeoHNNH* show a high degree of conservation at residues forming the protein-protein interface with AcrIIC1. Charged residues highlighted in (B) are colored and other conserved residues are denoted by asterisks.

Solution NMR studies of the GeoHNNH-AcrIIC1 complex reveals direct and long-range structural perturbations

To assess the effect of AcrIIC1 on the atomic structure of *GeoHNNH*, we first used NMR to visualize the *GeoHNNH* amide backbone under the influence of increasing concentrations of AcrIIC1. ^1H - ^{15}N transverse relaxation optimized spectroscopy-heteronuclear single quantum coherence (TROSY-HSQC) spectra of *GeoHNNH* revealed that AcrIIC1 binding induced chemical shift perturbations (CSPs) throughout the nuclease domain, despite interacting with only two residues in the X-ray crystal structure. We observed the titration of *GeoHNNH* with wild-type (WT) AcrIIC1 to be under slow exchange in the large majority of *GeoHNNH* resonances, suggesting a tight binding interaction (**Figure 3A, Figure 4**) The greatest magnitude and concentration of these CSPs are observed near the *GeoHNNH* catalytic site, consistent with the crystal structure, though the CSPs caused by AcrIIC1 propagate well into the core of *GeoHNNH* (**Figure 3B, C**). Importantly, a reciprocal NMR titration using ^{15}N -labeled AcrIIC1 produced a similar level of slow-exchange CSPs occurring predominantly at the expected crystallographic interface (**Figure 5**).

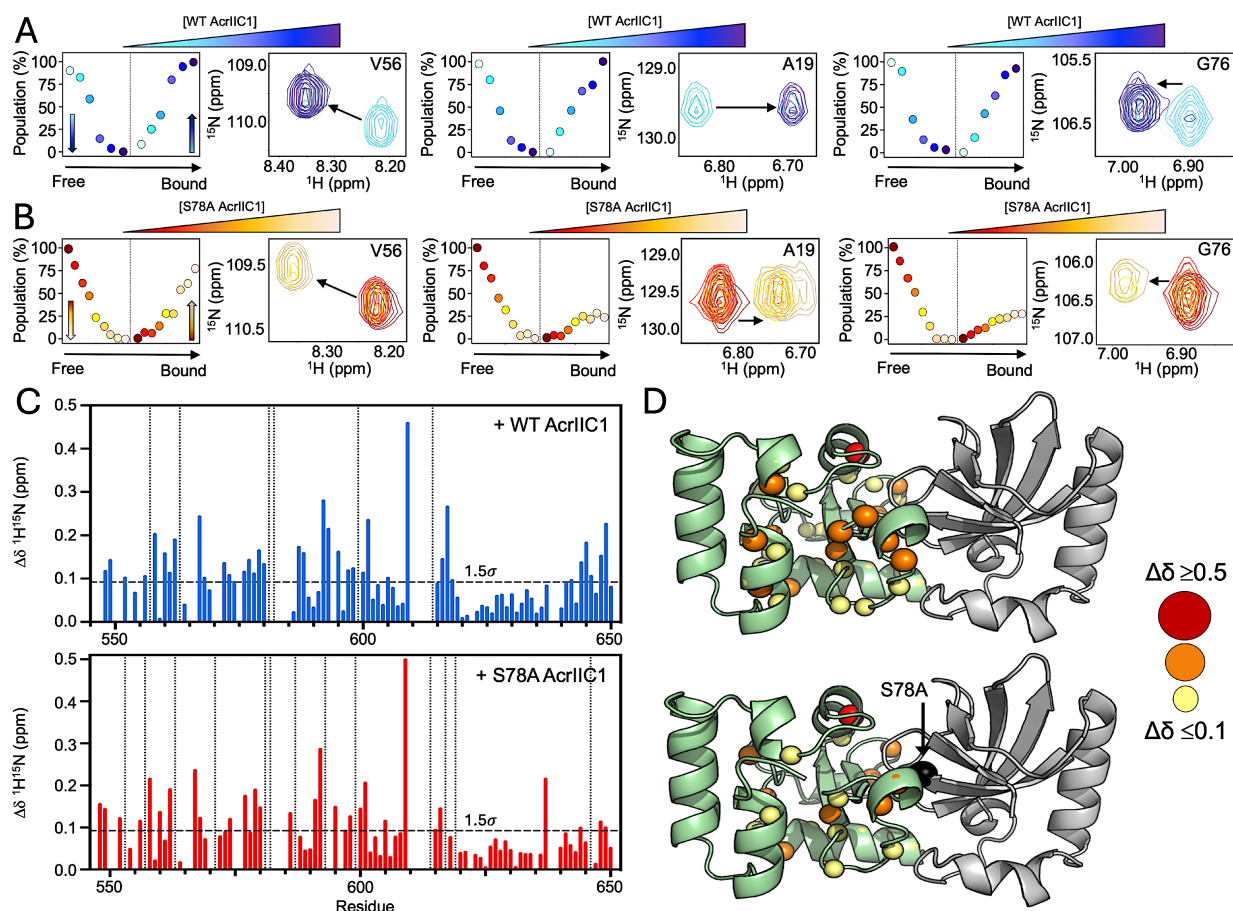


Figure 3. Snapshots of ^1H - ^{15}N HSQC titrations of *GeoHNNH* with *AcrIIC1* inhibitor. *GeoHNNH* is titrated with (A) WT *AcrIIC1* (light-to-dark blue) and (B) S78A *AcrIIC1* (dark-to-light red). Slow exchange CSPs suggest tight binding and populations of free and *AcrIIC1*-bound *GeoHNNH* are derived from resonance volumes shown as dot plots. The inflection between ligation states is indicated by vertical dashed lines. (C) Per-residue NMR CSPs of *GeoHNNH* caused by WT *AcrIIC1* (upper, blue bars). The same concentration of S78A *AcrIIC1* induces very weak CSPs in *GeoHNNH* (lower, black bars), while further addition of S78A *AcrIIC1* rescues the CSPs (red bars). Dashed vertical lines denote sites of NMR line broadening in each plot. (D) CSPs $>1.5\sigma$ of the 10% trimmed mean of all shifts (horizontal dashed line in (B)) are mapped onto the *GeoHNNH* structure as spheres, where the size and color correlate with intensity of the CSP. The S78A mutation is noted with a black sphere.

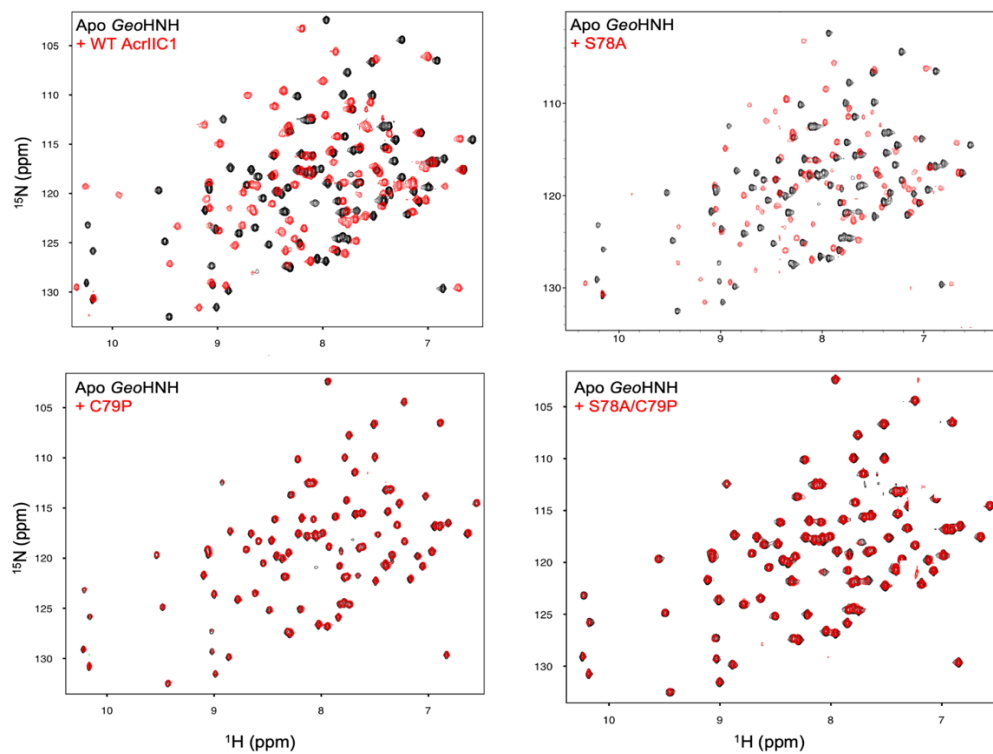


Figure 4. ^1H - ^{15}N HSQC NMR spectra of *GeoHNNH* titrated with WT, S78A, C79P, or S78A/C79P AcrIIC1. C79P and S78A/C79P AcrIIC1 inhibitors do not induce any CSPs in *GeoHNNH*.

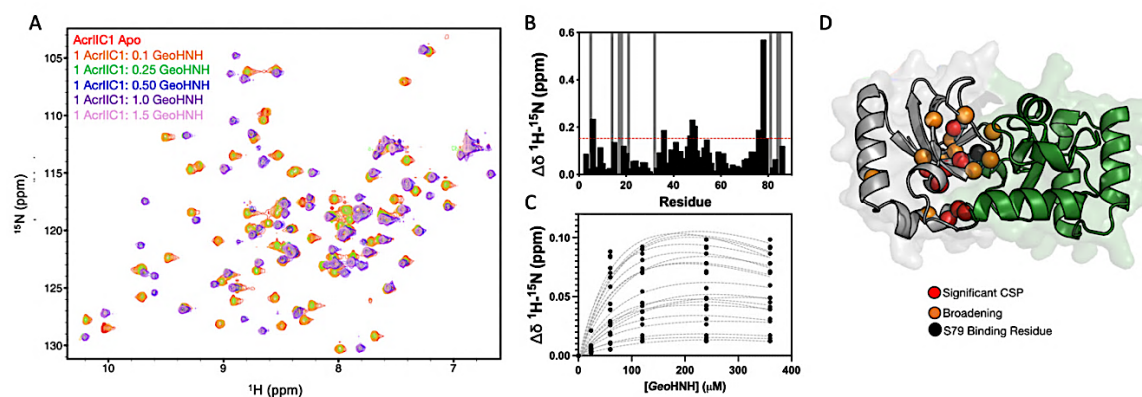


Figure 5. (A) ^1H - ^{15}N HSQC spectral titration of ^{15}N -AcrIIC1 with GeoHNNH. (B) Per-residue CSPs within AcrIIC1 caused by GeoHNNH. Grey bars denote sites of NMR line broadening. Red dashed vertical line are CSPs $>1.5\sigma$ of the 10% trimmed mean of all shifts. (C) Apparent binding affinity of AcrIIC1 for GeoHNNH, based on a global fit of fast-exchange NMR chemical shift trajectories, is ~ 15 mM. (D) CSPs $>1.5\sigma$ of the 10% trimmed mean of all shifts (horizontal dashed line in (B)) are mapped onto the AcrIIC1 structure as spheres, where red denotes CSPs $>1.5\sigma$, orange denotes line broadening and black denotes the binding site residue S78.

To validate the contribution of individual AcrIIC1 residues depicted in the X-ray structure to the *Geo*HNH binding interaction, we introduced single-point mutations to S78 (S78A) and C79 (C79P, to disrupt both the Cys side chain and amide backbone interactions) of AcrIIC1. Circular dichroism and ^1H - ^{15}N NMR spectra confirm that S78A and C79P variants remain folded with intact secondary structure (**Figure 6**). The S78A AcrIIC1 variant induced a similar effect on *Geo*HNH in ^1H - ^{15}N TROSY-HSQC titrations as WT AcrIIC1 (**Figure 3B-D, Figure 4**), though line broadening is enhanced in titrations with S78A AcrIIC1. The C79P AcrIIC1 variant induced no CSPs in *Geo*HNH, highlighting the hydrogen bonding between the sidechain and backbone of C79 with *Geo*HNH as most crucial for maintaining the complex. Additionally, we engineered a S78A/C79P negative control and confirmed this variant induced no CSPs in the ^1H - ^{15}N NMR spectrum of *Geo*HNH.

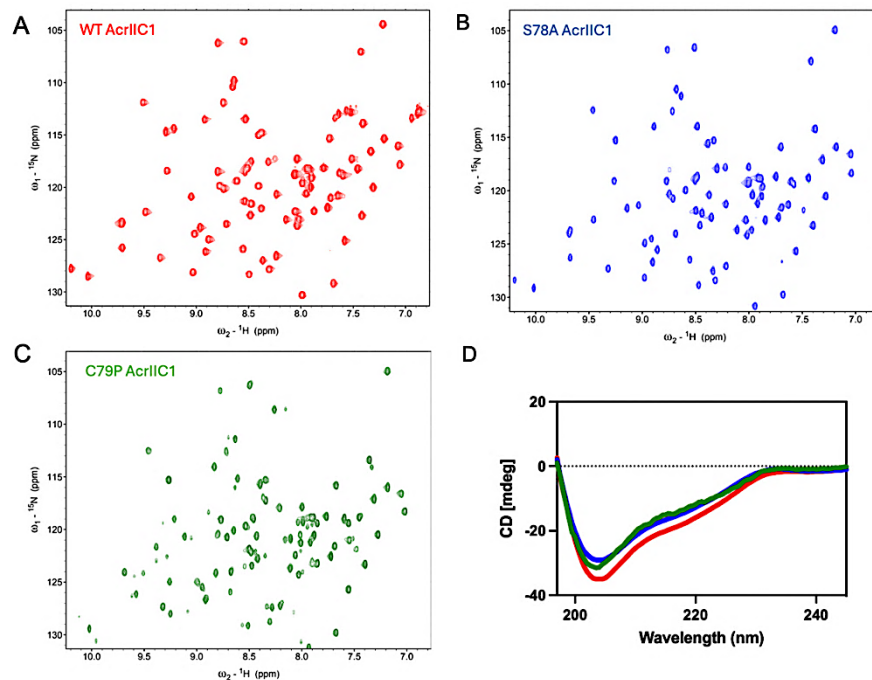


Figure 6. ^1H - ^{15}N HSQC spectra of (A) WT AcrIIC1 (B) S78A AcrIIC1 and (C) C79P AcrIIC1. (D) CD spectra of WT (red), S78A (blue) C79P AcrIIC1 (green) showing maintenance of secondary structure in the presence of mutations.

To further assess the interaction between *Geo*HNH and WT AcrIIC1 or variants, we measured R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE relaxation parameters that are sensitive to fast timescale conformational motions and report on global tumbling (**Figure 7**). When compared to *Geo*HNH alone, *Geo*HNH in complex with both WT AcrIIC1 and S78A showed globally depressed R_1 and elevated R_2 values consistent with an increase in molecular weight. As expected from initial NMR titrations, R_1 and R_2 values of *Geo*HNH in the presence of C79P AcrIIC1 showed no change from the apo protein. ^1H - ^{15}N NOEs of both WT and S78A AcrIIC1 revealed some local fluctuation in backbone motion due to AcrIIC1, while the C79P NOE profile resembled that of apo *Geo*HNH. We then used these data to calculate per-residue order parameters (S^2) for the free and Acr-bound states of *Geo*HNH (**Figure 8A**, **Figure7**).

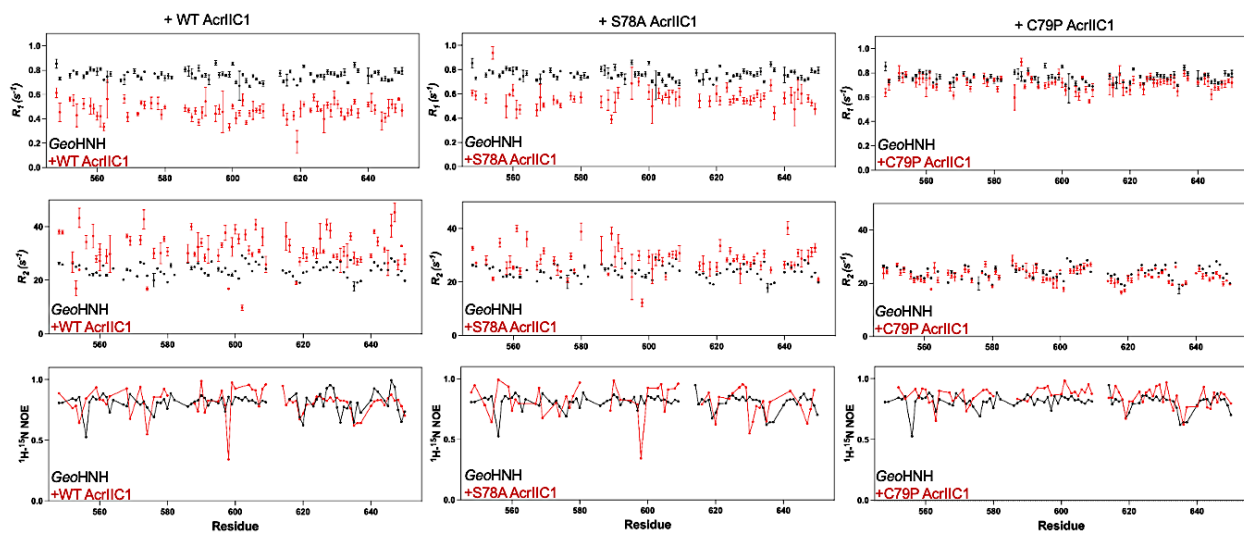


Figure 7. Summary of R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE relaxation parameters. The per-residue R_1 , R_2 , and NOE values are plotted for *GeoHNNH* in the presence of WT (left), S78A (center), and C79P (right) AcrIIC1.

Most notably, regions in close sequence and spatial proximity to the *Geo*HNH active site experienced sharp increases in S^2 in the presence of AcrIIC1, suggesting the inhibitor suppresses *ps-ns* timescale flexibility in these regions. Collectively, NMR titrations and relaxation experiments point to the interactions of AcrIIC1 residue C79 with *Geo*HNH as critical to the inhibited complex, while S78 plays a lesser role. Additionally, we observed a suppression of flexibility of the *Geo*HNH domain on the *ps-ns* timescale, which has previously been shown to control nuclease function and allostery⁽¹²⁾.

AcrIIC1 disrupts salt bridge networks of GeoHNH, stimulating millisecond timescale motions

In a prior study of *Geo*HNH, we analyzed an active site salt bridge network and showed that mutation of a central residue (K597) disrupted many charged contacts, leading to significant structural and functional perturbations⁽¹²⁾. Namely, a K597A mutation induced widespread and globally synchronous millisecond timescale motions that were absent from WT *Geo*HNH. Further, these motions rewired *Geo*HNH such that a single Lys-to-Ala mutation enhanced the thermostability of full-length *Geo*Cas9 and increased its functional range for DNA cleavage. Our X-ray crystal structure of the *Geo*HNH-AcrIIC1 complex reveals the same extent of salt bridge disruption. Stabilizing contacts between K597, H582, and D592 are now broken, allowing the loop containing the catalytic histidine to flip outward, while K597 rotates from the *Geo*HNH surface toward the core of the protein (**Figure 8B**).

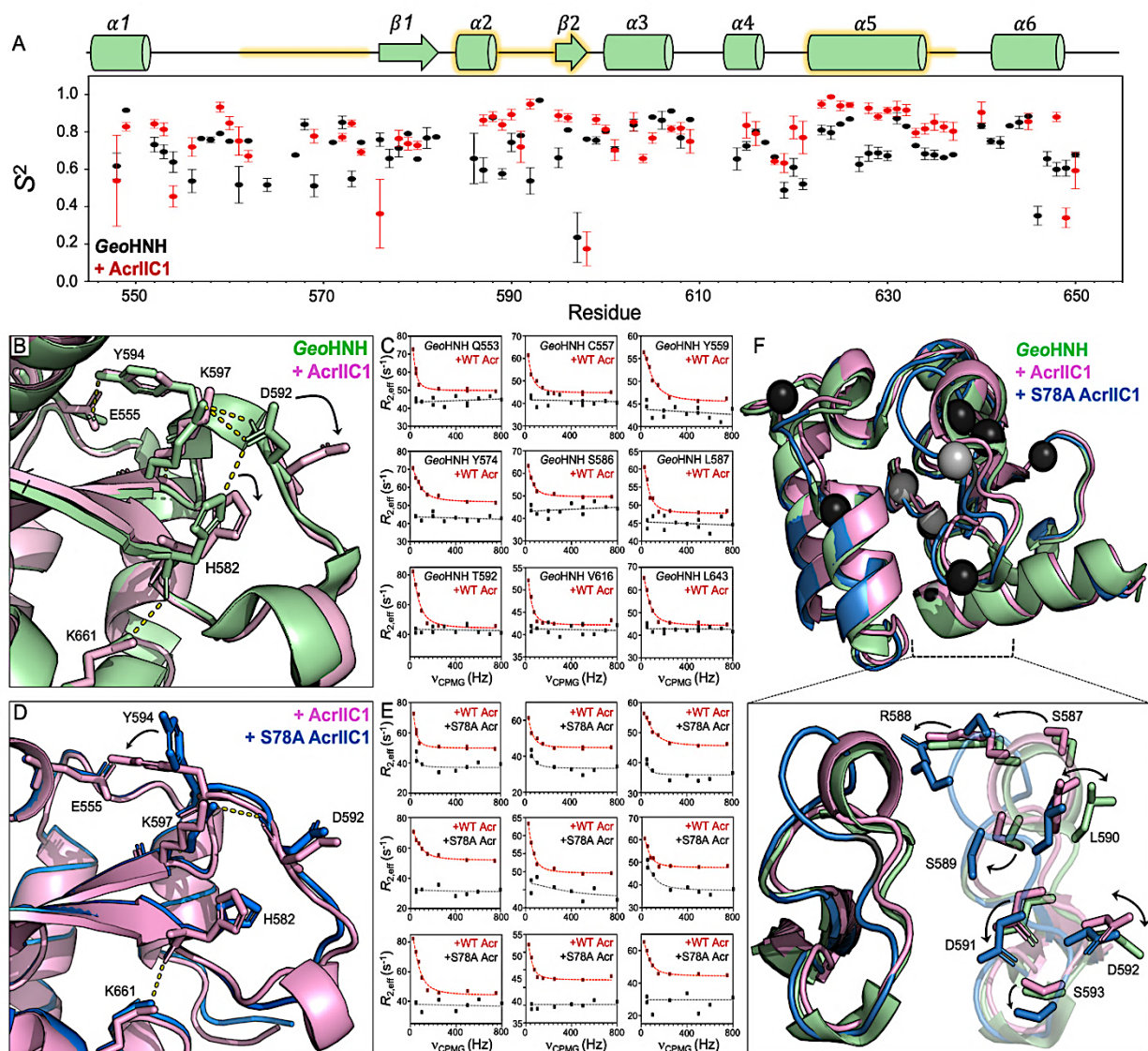


Figure 8. (A) NMR order parameters (S^2) of *GeoHNNH* in the absence (black) and presence (red) of AcrIIC1. A cartoon of *GeoHNNH* structure (top) highlights AcrIIC1-induced suppression of fast timescale flexibility (yellow glow). (B) X-ray crystal structure of *GeoHNNH* (green, PDB: 7MPZ) overlaid with that of AcrIIC1-bound *GeoHNNH* (pink, PDB: 10VC). AcrIIC1 disrupts electrostatic contacts within a salt bridge network (dashes) via displacement of key catalytic side chains (arrows). (C) CPMG relaxation dispersion profiles of *GeoHNNH* (black, flat profiles) and AcrIIC1-bound *GeoHNNH* (red curves), highlighting the stimulation of *ms* motions upon disruption of the salt bridge network. (D) X-ray crystal structures of WT (pink, PDB: 10VC) and S78A AcrIIC1-bound *GeoHNNH* (blue, PDB: 10VB) are nearly identical at the *GeoHNNH* active site. (E) CPMG relaxation dispersion experiments indicate that the S78A mutation limits the ability of AcrIIC1 to stimulate *ms* motions in *GeoHNNH*. (F) Sites of *ms* timescale flexibility are mapped onto a structural overlay of *GeoHNNH* (green, PDB: 7MPZ), AcrIIC1-bound *GeoHNNH* (pink, PDB: 10VC), and S78A AcrIIC1-bound *GeoHNNH* (blue, PDB: 10VB). Black spheres indicate flexible residues in the WT Acr-bound complex, while gray spheres are flexible residues in the S78A Acr-bound complex. CPMG relaxation dispersion localizes to a region of crystallographic structural heterogeneity among the three *GeoHNNH* domains, highlighted for the backbone and side chains.

We then interrogated the impact of salt bridge disruption by AcrIIC1 on *Geo*HNH flexibility with Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR experiments, noting that millisecond motions, as in our prior work, were now apparent in *Geo*HNH (**Figure 8C**). Ten residues in the *Geo*HNH-AcrIIC1 complex now display curved relaxation dispersion profiles, while no such profiles exist in *Geo*HNH alone. Residues undergoing chemical exchange are not isolated to the binding interface with AcrIIC1, but rather span the length of the *Geo*HNH, including sites of interaction with nucleic acids and the recognition (REC) domain within *Geo*Cas9. This suggests the potential of AcrIIC1 to alter the communication between REC and HNH, which has been shown to be crucial for coordinating double stranded DNA cleavage^(14, 23-25). An assessment of the S78A AcrIIC1 variant showed a nearly identical active site arrangement of the inhibitor bound state, though the extent of *ms* motion was not quite as pronounced (**Figure 8D, E**). Interestingly, the region of *Geo*HNH harboring the bulk of its *ms* motion and greatest decreases in *ps-ns* motion, also displays conformational heterogeneity between our *Geo*HNH, WT AcrIIC1-bound, and S78A AcrIIC1-bound X-ray crystal structures (**Figure 8F**) which may explain the subtle differences in NMR-detected flexibility.

Mutations at the crystallographic binding interface limit the ability of AcrIIC1 to inhibit the sgRNA interactions and DNA cleavage of GeoCas9

We next tested the functional consequences of mutations at the crystallographic *Geo*HNH-AcrIIC1 binding interface with *in vitro* DNA cleavage assays of *Geo*Cas9 in the presence of WT, S78A, C79P, or S78A/C79P AcrIIC1. WT AcrIIC1 inhibits target DNA cleavage by *Geo*Cas9 in a concentration dependent manner, consistent with prior reports that also demonstrated a significant molar excess of AcrIIC1 is required for inhibition⁽¹⁷⁾. At similar concentrations, we

observed essentially no inhibition by the S78A, C79P, or S78A/C79P (negative control) AcrIIC1 variants (**Figure 9A, B**).

We also noted that the order in which components were added to the reaction mixture was critical to the efficiency of inhibition. When the *Geo*Cas9 ribonucleoprotein complex (RNP) was formed prior to the addition of AcrIIC1, little-to-no inhibition was observed (**Figure 10**). This suggests that sgRNA-induced conformational changes, which typically suppress the flexibility of apo Cas enzymes, may occlude the AcrIIC1 binding site.

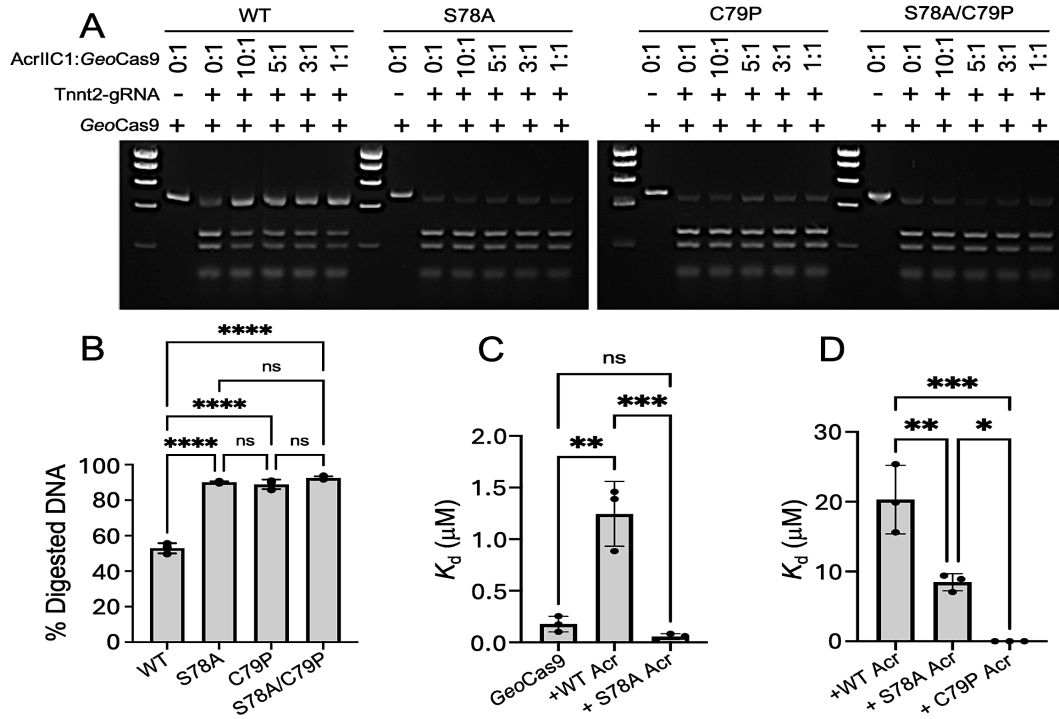


Figure 9. (A) *In vitro* DNA cleavage assays indicate that AcrIIC1 mutations at the crystallographic binding interface ablate its inhibition. *GeoCas9* was incubated with varying concentrations of WT, S78A, C79P, and S78A/C79P AcrIIC1, after which RNPs were formed by incubating the AcrIIC1-bound *GeoCas9* with sgRNA and assayed for target DNA cleavage at 37 °C. **(B)** Bar graph quantifying the DNA cleavage via band intensities from the agarose gel. **** $p < 0.0001$. $n = 3$ technical replicates for each sample. **(C)** MST-derived K_d values of *GeoCas9* for its sgRNA in the absence and presence of WT and S78A AcrIIC1. ** $p < 0.0011$; *** $p < 0.0006$. $n = 3$ technical replicates for each sample. **(D)** MST-derived K_d values of *GeoHNH* for WT, S78A, and C79P AcrIIC1. * $p < 0.0282$; ** $p < 0.0061$; *** $p < 0.0004$. $n = 3$ technical replicates for each sample.

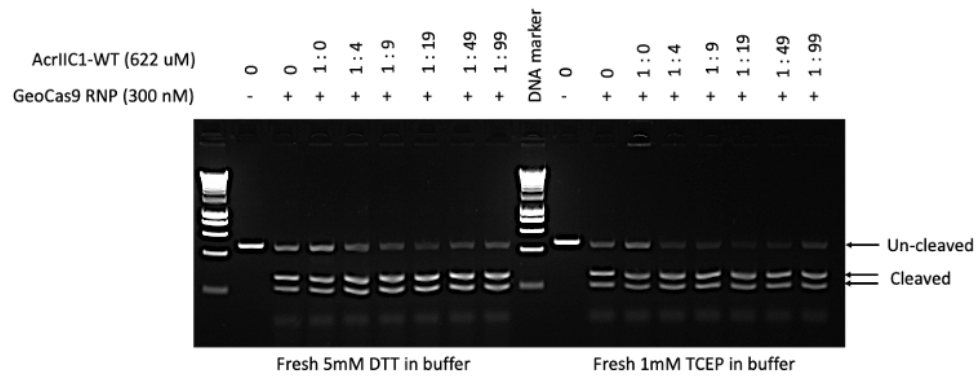


Figure 10. Representative gel monitoring *in vitro* DNA cleavage. If the *GeoCas9* RNP is formed prior to the addition of *AcrIIC1*, a significant reduction in the inhibition efficacy of *AcrIIC1* is observed. Thus, while *AcrIIC1* can inhibit *GeoCas9* at higher concentrations, it is most effective if it binds *GeoCas9* prior to sgRNA.

These data led us to wonder if the affinity of *GeoCas9* for its sgRNA was modulated by the inhibitor. We used microscale thermophoresis (MST) to measure binding affinities between *GeoCas9* and a Cy5-labeled sgRNA from a recently published cryo-EM structure (PDB:8UZA), using *GeoCas9* alone or in the presence of WT or S78A AcrIIC1. Consistent with functional assays, MST showed that WT AcrIIC1 significantly reduced the affinity of *GeoCas9* for its sgRNA ($K_d = 124 \pm 91$ nM for *GeoCas9* compared to $K_d = 1.2 \pm 0.4$ μ M for *GeoCas9*-AcrIIC1) (**Figure 9C**). Surprisingly, we found that despite WT-like X-ray structures and CSPs in NMR titrations of *GeoHNH* with S78A AcrIIC1, this variant did not attenuate the affinity of *GeoCas9* for sgRNA ($K_d = 52.2 \pm 23.7$ nM). These data are also consistent with functional assays, which show little-to-no inhibition from the S78A AcrIIC1 variant. Interestingly, we found that MST assessing protein-protein interaction between *GeoHNH* and WT or S78A AcrIIC1 showed little difference in binding affinity (**Figure 9D**), with S78A returning a slightly tighter K_d than WT AcrIIC1 to *GeoHNH* ($K_d = 19.9 \pm 12.5$ μ M and 7.93 ± 4.5 μ M, respectively). Despite its structural impact in NMR titrations, differences between an isolated *GeoHNH* domain and full-length *GeoCas9* suggest that direct side chain-specific interactions between S78 of AcrIIC1 and the catalytic H582 is in fact required for inhibition.

Mutations in AcrIIC1 cause dynamic destabilization of GeoHNH-AcrIIC1 interface

To complement our experimental structural studies, we performed extensive all-atom molecular dynamics (MD) simulations of the *GeoHNH*-AcrIIC1 complex to investigate how Acr binding and the interfacial mutations reshaped the conformational landscape of *GeoHNH*. Simulations were carried out for the WT AcrIIC1 complex as well as the S78A, C79P, and S78A/C79P variants in triplicate, enabling robust quantitative comparisons of structural stability, flexibility, and allosteric coupling.

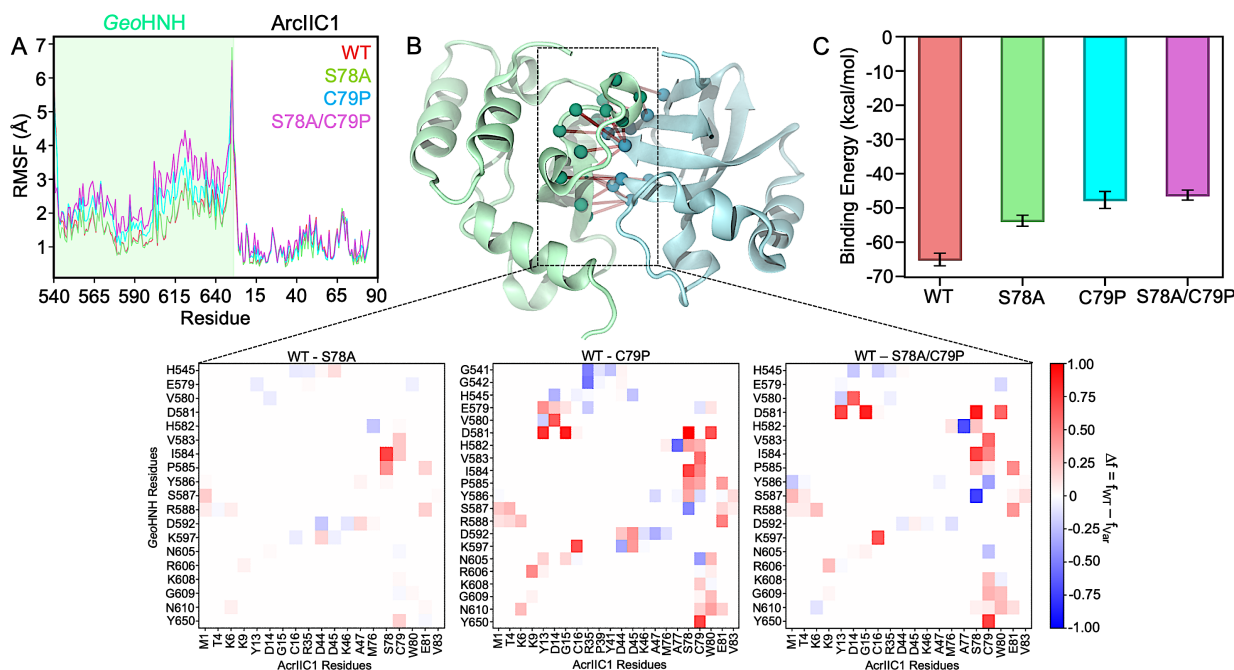


Figure 11. Effects of AcrIIC1 mutations on structural stability, binding energetics, and interface integrity of the *GeoHNNH*-AcrIIC1 complex. **(A)** Per-residue root-mean-square fluctuations (RMSF) of the *GeoHNNH*-AcrIIC1 complex across WT and mutant systems (S78A, C79P, and S78A/C79P). The shaded green region highlights the *GeoHNNH* domain within the complex. **(B)** Representative structural snapshot of the *GeoHNNH*-AcrIIC1 complex highlighting the primary binding interface. Spheres denote key residues participating in stable interfacial interactions (red lines) in the WT complex, corresponding to contacts enriched in WT (red squares) identified from the differential contact maps. Lower panels depict residue-residue contact difference maps (Δf) for S78A, C79P, and S78A/C79P variants relative to WT ($\Delta f = \text{WT} - \text{variant}$). Positive values (red) indicate contacts enriched or stable in the WT complex, whereas negative values (blue) indicate contacts newly formed in the mutant complexes. **(C)** Binding energies (kcal/mol) between *GeoHNNH* and WT or mutant AcrIIC1 systems, showing reduced binding energy upon mutation ($\pm$ standard error of mean).

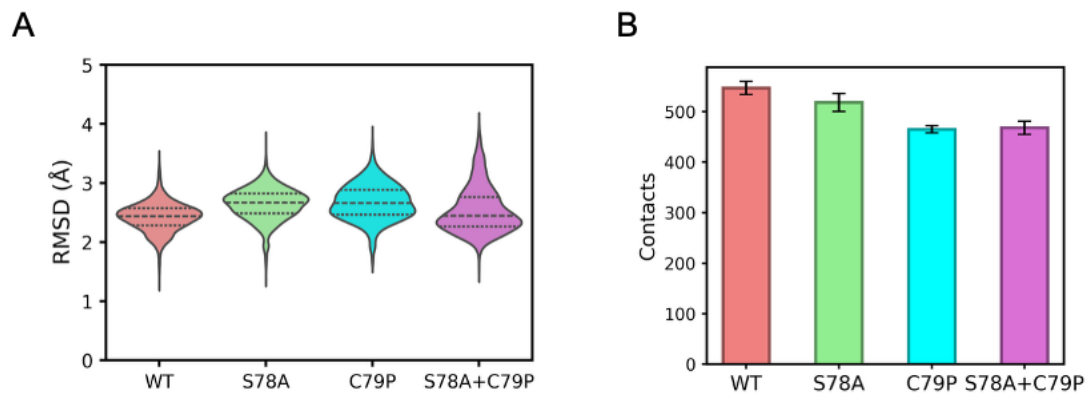


Figure 12. Mutation-dependent changes GeoHNNH-AcrIIC1 complex. A) Backbone RMSD distributions from triplicate MD simulations for WT and mutant complexes (S78A, C79P, and S78A/C79P). B) Average number of intermolecular contacts between GeoHNNH and AcrIIC1 in WT and mutant systems, defined using a 4.5 Å heavy-atom distance cutoff. Error bars represent the standard error of the mean (SEM) calculated across three independent 2 μ s trajectories for each system.

Across all replicas, the *Geo*HNH-WT AcrIIC1 complex remained structurally stable, as reflected by low backbone root-mean-square deviations (RMSDs) relative to the crystallographic structure (**Figure 12**). The variants exhibited RMSDs values comparable to WT, with S78A and C79P exhibiting slightly increased variability. Per-residue root-mean-square fluctuation (RMSF) analysis revealed higher intrinsic flexibility of *Geo*HNH than AcrIIC1 (**Figure 11A**). While the S78A variant largely preserved WT-like backbone dynamics, both the C79P and double mutant complexes exhibited elevated fluctuations (**Figure 11A**). Notably, increased fluctuations extended beyond the immediate binding interface, consistent with partial propagation of local perturbations into broader regions of the *Geo*HNH domain, consistent with NMR CSP analysis.

To further assess the stability of the *Geo*HNH-AcrIIC1 complex, we calculated the total number of contacts formed between them. Here, a contact is defined using a distance cut-off of 4.5 Å between heavy atoms of HNH residues and AcrIIC1 inhibitor residues. The WT complex maintained the highest number of intermolecular contacts, whereas S78A showed a modest reduction, and both C79P and the double mutant exhibited a substantial loss of contacts (**Figure 12B**). These observations are consistent with subtle disruptions in interfacial packing rather than complete dissociation. To further characterize changes in intermolecular interactions, we examined differential residue-residue contact maps relative to the WT complex (**Figure 10B, 11B**). The WT system is characterized by a dense network of persistent contacts at the *Geo*HNH–AcrIIC1 interface, several of which are substantially weakened or lost in the mutant systems. These disrupted interactions are reflected as regions of positive Δf , highlighting key interaction hotspots that are sensitive to mutation. While some compensatory contacts emerge in the mutants (negative Δf), these are sparse and do not recapitulate the cohesive interaction network observed in the WT complex. Structural mapping of these WT-enriched contacts onto the complex reveals that residues

surrounding positions 78 and 79 form a central interaction hub, underscoring their importance in maintaining interface integrity (**Figure 10B, 11B**).

Consistent with these structural and dynamical changes, interaction energy calculations revealed a progressive reduction in binding affinity upon mutation. The WT complex exhibited the most favorable interaction energy, followed by S78A, whereas C79P and the double mutant showed significantly weaker binding (**Figure 10C**). Together, these results demonstrate that while S78A induces only modest perturbations, the C79P and double mutations substantially disrupt interfacial packing, increase conformational flexibility, and weaken binding, thereby compromising the inhibitory efficiency of AcrIIC1 against *Geo*HNH.

Collectively, the MD simulations indicate that interfacial mutations increase conformational heterogeneity and destabilize the catalytic-loop geometry observed in the crystallographic complex. In agreement with X-ray and NMR analyses, effective inhibition by AcrIIC1 depends not only on active-site engagement but also on stabilization of a coordinated dynamic interaction with *Geo*HNH. The behavior of the variants further suggests that subtle perturbations at the interface can partially decouple Acr binding from inhibition, highlighting the importance of dynamic allosteric regulation in this system.

These observations are consistent with functional assays, which reveal that the S78A AcrIIC1 variant exhibits little to no inhibitory activity despite retaining binding to *Geo*HNH. Notably, MST measurements of the *Geo*HNH-AcrIIC1 interaction show comparable binding affinities for WT and S78A AcrIIC1, with S78A displaying a slightly tighter apparent dissociation constant ($K_d = 7.93 \mu\text{M}$) relative to WT ($K_d = 19.9 \mu\text{M}$) (**Figure 9D**). This apparent discrepancy between binding affinity and inhibitory function suggests that the S78A mutation does not

significantly impair complex formation but instead perturbs the functional coupling between binding and inhibition. In the context of our MD simulations, these results support a model in which S78A preserves overall binding but subtly alters interfacial dynamics and contact organization, leading to a loss of inhibitory efficiency without substantial destabilization of the complex. Together, the experimental and computational data highlight that effective inhibition by AcrIIC1 depends not only on binding affinity but also on the precise structural and dynamic integrity of the interface required to stabilize the catalytically inactive conformation of *Geo*HNH.

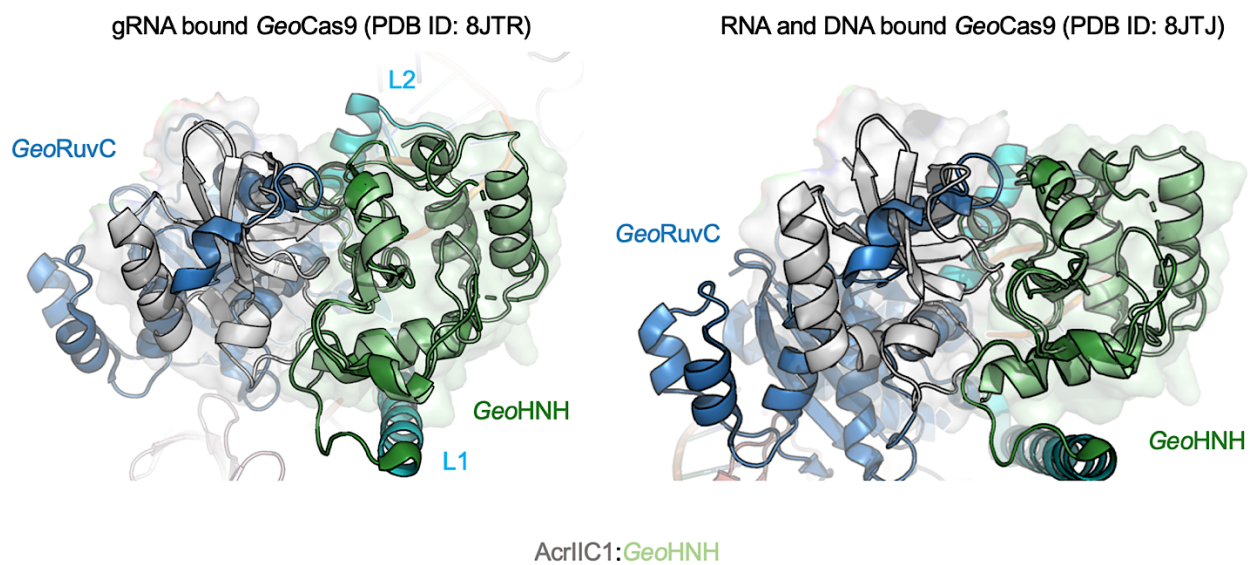


Figure 13. Structures of sgRNA-bound (left) and sgRNA:DNA-bound *GeoCas9* revealing potential steric clash of AcrIIC1 with RuvC when these structures are overlaid with the *GeoHNH*-AcrIIC1 crystal structure.

NMR and AlphaFold3 modeling suggest a second binding location for AcrIIC1 on GeoHNH

We have noted in several prior studies that a shift in the intrinsic dynamics of *GeoHNH* to slower timescales is detrimental to its ability to propagate long-range allosteric information^(14, 26). Balancing this point is our prior speculation, supported by functional assays, that despite this motional shift, mutation-induced disruption of salt bridges could modify or even strengthen contacts between HNH and the adjacent RuvC or REC domains of full-length *GeoCas9*⁽¹²⁾. The influence of the AcrIIC1 binding partner, as opposed to a mutation, could be more complicated. As has been observed in related Cas9-Acr complexes, when our X-ray structure of *GeoHNH*-AcrIIC1 is overlaid with structures of single-guide RNA (sgRNA)-bound or sgRNA:DNA-bound *GeoCas9*, the steric clash is very high between AcrIIC1 and the RuvC domains of these static snapshots (**Figure 13**). However, slower timescale and larger amplitude motions of the *GeoHNH* domain could serve to accommodate the inhibitor at its primary binding site. Thus, AcrIIC1 may prefer to bind apo *GeoCas9* and prevent additional conformational dynamics that activate the nucleases.

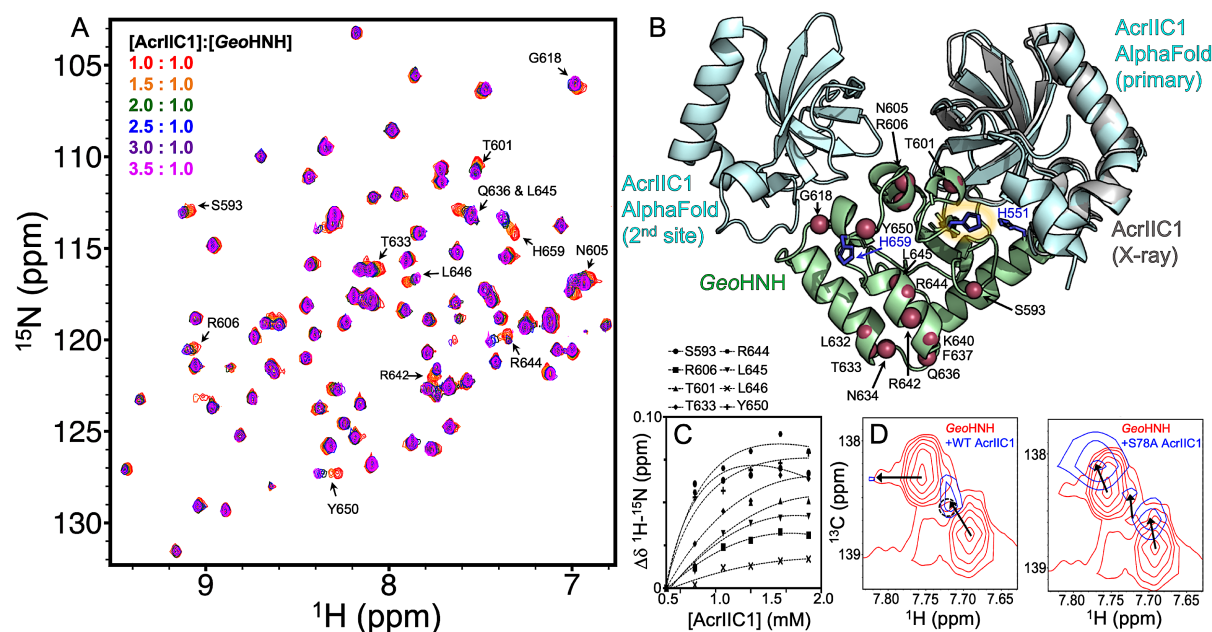


Figure 14. (A) NMR spectral overlay of ^{15}N -GeoHNNH revealing a small number of CSPs at super-saturating concentrations of AcrIIC1. The spectra are colored according to the Legend (inset). (B) AlphaFold3 model of GeoHNNH (green) bound to two molecules of AcrIIC1 (cyan). The primary site matches our crystallographic interface (gray cartoon), while the second site localizes to regions of NMR CSPs (red spheres). Histidine side chains are shown as blue sticks and the catalytic H582 is circled. (C) Apparent binding affinity for the second site, based on a global fit of NMR chemical shift trajectories, is ~ 710 mM. (D) ^1H - ^{13}C NMR spectra of the His e1 side chain show AcrIIC1-induced CSPs (arrows) or line broadening (circles) in all three native His residues of GeoHNNH, despite two of these being outside the primary binding site.

AcrIIC1 induced large and widespread CSPs in NMR spectra of ^{15}N -*Geo*HNH (**Figure 3A**, **Figure 4**), which ceased at equimolar equivalents of AcrIIC1: *Geo*HNH. However, further titration to significant molar excesses of AcrIIC1 induced linear shifts in fast exchange at a small number of resonances (**Figure 14A**). Distinct from the initial addition of AcrIIC1, these shifts map to two discrete regions of the *Geo*HNH domain (**Figure 14B**) that would be solvent exposed in the context of full-length *Geo*Cas9. We therefore speculated that *Geo*HNH may accommodate a second AcrIIC1 inhibitor at another interface. An AlphaFold3 model predicted a secondary binding location that aligned with one of the regions of NMR CSPs, with specific interactions between AcrIIC1 and the N605 and R606 sidechains, as well as the G618 backbone⁽²⁷⁾. Further, our AlphaFold3 model strongly resembles an existing structure (PDB: 6J9N, 8IF0) of *Nme*HNH bound to both AcrIIC1 (identical to our crystallographic interface) and AcrIIC3, though the secondary Acrs have subtle differences in orientation^(22, 28). NMR demonstrates the secondary interaction to be weak and/or transient, with an apparent $K_d = 700\ \mu\text{M}$ from a global fit of CSPs in fast exchange (**Figure 14C**). ^1H - ^{13}C HMQC NMR experiments visualizing the histidine side chains of *Geo*HNH showed that in addition to expected line broadening of the catalytic H582 due to its direct interaction with WT AcrIIC1, CSPs and/or line broadening of H551 and H659, which are distal to primary binding site, are also apparent (**Figure 14D**). H659 is especially close to the AlphaFold3-predicted secondary interface. It should be noted, however, many CSPs from our NMR experiments also localized to the base of two α -helices on the face opposite the *Geo*HNH catalytic site (**Figure 14B**), making it an equally plausible surface for interaction.

Discussion

Anti-CRISPRs have emerged as natural protein regulators of CRISPR-Cas effectors, but most mechanistic work has focused on mesophilic Cas systems and/or static views of Acr inhibition. Furthermore, previous biophysical studies of Acrs have generally focused on the structure of the Acrs themselves, rather than atomistic interactions with Cas proteins. Structural insights into several Cas9s in complex with AcrIIC1 have framed it as a broad-spectrum inhibitor sterically occluding a conserved HNH active site⁽¹⁷⁾. Building on these prior insights, new atomic level information from NMR and MD simulations have refined a view of AcrIIC1 binding that reshapes the internal communication pathways within the thermophilic *Geo*Cas9 enzyme. This study presents evidence for a dynamic aspect of anti-CRISPR inhibition in which Acrs modulate the same networks within *Geo*HNH that govern RuvC and REC coordination and sgRNA affinity⁽²⁶⁾. Our prior work revealed that the core dynamics of *Geo*HNH are consistent up to temperatures of 40 °C, highlighting an evolutionary pressure to preserve fast motions under high entropic burden. We found no specific motional pathway comprised by *Geo*HNH residues, but rather a robust preservation of the internal dynamics themselves⁽²⁶⁾. We previously speculated that rewiring these core motions toward the slower *ms* timescale would increase the activity of *Geo*Cas9, as similar flexibility was found to be widespread and principal in the canonical *Sp*Cas9. However, such a rewiring of the *Geo*HNH domain by AcrIIC1 in fact has an inhibitory effect, highlighting a key distinction between the intrinsic dynamics that govern mesophilic and thermophilic Cas9, and thus the need to establish detailed, system-specific mechanistic rules for Cas effectors.

This study suggests that AcrIIC1 inhibits *Geo*Cas9 through a combination of orthosteric and allosteric effects that are more nuanced than pure active-site steric hinderance. An X-ray

crystal structure of the *Geo*HNH-AcrIIC1 complex highlights an interface that engages the HNH catalytic loop via AcrIIC1 residues S78 and C79, disrupting an electrostatic network within *Geo*HNH that rewires its local structure and dynamics⁽¹²⁾. A region of conformational heterogeneity across several HNH crystal structures is observed, and NMR relaxation experiments reveal that AcrIIC1 suppresses *ps-ns* motions throughout this region (and others) while stimulating new, broadly distributed *ms* motion. Functionally, the interaction between *Geo*HNH residue H582 and S78 of AcrIIC1 is crucial to inhibition, while disruption of the D581-C79 interaction abolishes the binding of AcrIIC1 to *Geo*HNH and its inhibition of *Geo*Cas9. At high inhibitor concentrations, NMR and AlphaFold3 modeling indicate the existence of a weak secondary binding site on *Geo*HNH. We speculate that this secondary site could most plausibly affect closed conformations of sgRNA-bound *Geo*Cas9, altering communication between *Geo*HNH and adjacent RuvC and REC domains.

Anti-CRISPRs have been co-opted as a biological research tool for spatiotemporal control of Cas9 editing through strategic employment of inhibition⁽²⁹⁾. This work highlights the necessity for atomistic structural and biophysical studies of Acr mechanisms, and that even broad-spectrum inhibitors like AcrIIC1 maintain atomistic subtleties that may differentiate their effect on related Cas systems. We showed that AcrIIC1 attenuates the affinity of *Geo*Cas9 for sgRNA, also demonstrating that timing and order of assembly of the Cas9 RNP are critical parameters for controlling Cas9 editing windows. However, occupancy of Acr in the *Geo*HNH catalytic site alone is not sufficient to diminish sgRNA affinity, but rather specific contacts between S78 and H582 are required. These findings argue that spatiotemporal regulation of *Geo*HNH by an Acr cannot be explained solely by steric occlusion, rather, specific side-chain chemistry must perturb long-range communication between *Geo*HNH, REC, and RuvC. The stimulation of *ms* motions via salt-bridge

disruption parallel earlier observations of altered *Geo*HNH dynamics, which in prior work extended the thermostability and editing range of *Geo*Cas9, suggesting that natural Acrs and engineered *Geo*Cas9 variants access overlapping allosteric switches, with outcomes dependent on the specific motional landscape.

The potential for a weak secondary AcrIIC1 binding site occupied at high Acr concentrations in NMR titrations and supported by AlphaFold3 modeling is plausible when considering the inhibitor stoichiometries required for *Geo*Cas9 inhibition. Our results suggest that AcrIIC1 inhibition is far less efficient toward preformed *Geo*Cas9 RNPs than toward apo *Geo*Cas9, thus it is possible that evolutionary pressure would select for Acrs capable of binding at multiple sites on a Cas effector. Indeed, *acr* gene expression increases up to 100-fold within the first 20 minutes of infection, effectively flooding the bacterial cell with high concentrations of Acr protein⁽³⁰⁾. Although exact intracellular Acr concentrations are unknown, overexpression of the *acr* gene and high Acr:Cas9 ratios would favor promiscuous binding of Acrs to many regions of the Cas9 enzyme, including HNH. Such speculation draws some support from the wide diversity of reported Acr mechanisms, including between structurally similar orthologs^(22, 29). Further studies are required to unambiguously map additional binding sites for AcrIIC1 on *Geo*Cas9, though it is possible additional weak contacts between the enzyme and inhibitor influence flexibility as a means of accessing the catalytic site.

Together, the integrated structural biophysics presented here argue that even broad-spectrum Acrs, like AcrIIC1, can tune Cas9 activity in system-specific ways, providing a framework for engineering next-generation inhibitors that exploit both orthosteric active-site engagement and targeted manipulation of allosteric networks to achieve precise, context-dependent control of genome editing. These data also highlight the critical importance of

atomistic studies for continued elucidation of conformational dynamics and allosteric pathways within Cas systems, thereby informing the rational design of controllable Cas9-Acr complexes.

Materials and Methods

Protein expression and purification

Separately, the HNH domain of *Geobacillus stearothermophilus* Cas9 (residues 511-662), AcrIIC1 from *N. meningitidis* and S78A and C79P AcrIIC1 variants were cloned into pET28a vectors with N-terminal His₆-tags and TEV protease cleavage sites. The plasmids were transformed into BL21 (DE3) cells (New England BioLabs), expressed, and purified as previously described in LB media ^(12, 17). Isotopically labeled samples of *Geo*HNH and AcrIIC1 were grown in M9 minimal media containing CaCl₂, MgSO₄, and MEM vitamins and supplemented with ¹⁵N-labeled ammonium chloride and ¹³C-labeled glucose (1.0 g/L and 2.0 g/L, respectively; Cambridge Isotope Laboratories). Cells were grown to an OD₆₀₀ of 0.8–1.0, induced with 1 mM IPTG, and then incubated for a further 4 hr at 37 °C. Cells containing *Geo*HNH were harvested by centrifugation, resuspended in a buffer of 20 mM HEPES, 150 mM KCl, 5 mM imidazole, and 1 mM PMSF at pH 8.0 (lysis buffer), lysed by ultrasonication, and loaded onto a Ni-NTA column equilibrated with lysis buffer. The column was washed with 1 column volume of lysis buffer and *Geo*HNH was eluted with a linear gradient of imidazole from 0.2-250 mM. The His₆-tag was cleaved by dialyzing the resulting fractions containing *Geo*HNH against a buffer of 20 mM HEPES, 150 mM KCl, and 5 mM imidazole at pH 8.0 overnight at 4 °C with TEV. The sample was further purified on a Hi-Load Superdex 75 size-exclusion column and dialyzed into a buffer of 20 mM HEPES, 80 mM KCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, and 10% (v/v) D₂O at pH 7.4.

Cells containing AcrIIC1 and variants were also harvested by centrifugation, resuspended in a buffer of 50 mM Tris-HCl, 500 mM NaCl, 1 mM TCEP, 20 mM imidazole, and 1 mM PMSF at pH 7.5 (Acr lysis buffer), lysed by ultrasonication, and loaded onto a Ni-NTA column equilibrated

with Acr lysis buffer. The column was washed with 1 column volume of Acr lysis buffer and eluted with a linear gradient of imidazole from 0.2-250 mM. The His₆-tag was cleaved by dialyzing the resulting fractions containing AcrIIC1 against a buffer of 50 mM HEPES, 150 mM KCl, 5%(v/v) glycerol and 1 mM DTT at pH 7.5 overnight at 4 °C with TEV. The sample was further purified on a Hi-Load Superdex 75 size-exclusion column. AcrIIC1 protein for NMR spectroscopy was exchanged into a buffer of 20 mM HEPES, 80 mM KCl, 1 mM EDTA, 1 mM DTT, and 5% glycerol at pH 7.5 by Amicon centrifugal cell.

The full-length *GeoCas9* plasmid was acquired from Addgene (#87700). *GeoCas9* was expressed in TB media and was expressed and purified as previously described⁽¹³⁾.

NMR spectroscopy

Two-dimensional ¹H-¹⁵N HSQC NMR spectra were collected at 600 MHz and NMR spin relaxation experiments were carried out at 600 and 850 MHz on Bruker Avance NEO and Avance III HD spectrometers, respectively. All experiments were performed at 25 °C. NMR spectra were processed with NMRPipe and analyzed in NMRFAM-SPARKY^(31, 32). NMR assignments of *GeoHNNH* were obtained from BMRB entry 50887⁽¹⁴⁾, while those of AcrIIC1 were obtained from BMRB entry 36471 ⁽³³⁾.

NMR titrations were performed on a Bruker Avance NEO 850 MHz spectrometer via a series of ¹H-¹⁵N TROSY HSQC spectra with increasing ligand (*i.e.* AcrIIC1) concentration. The ¹H and ¹⁵N carrier frequencies were set to the water resonance and 120 ppm, respectively. Samples of WT *GeoHNNH* were separately titrated with a WT, S78A, or C79P AcrIIC1 until no further spectral perturbations were detected. Combined ¹H-¹⁵N NMR chemical shift perturbations were calculated using the method of Bax and coworkers ⁽³⁴⁾.

$$\Delta\delta = \sqrt{\frac{(\Delta\delta_{HN}^2 + \Delta\delta_{NH}^2/25)}{2}}$$

Longitudinal ⁽³⁵⁾ and transverse ⁽³⁶⁾ relaxation rates were measured with randomized T_1 delays of 0, 20, 60, 100, 200, 600, 800, and 1200 ms and T_2 delays of 0, 16.9, 33.9, 50.9, 67.8, 84.8, and 101.8 ms. Peak intensities were quantified in Sparky and the resulting decay profiles were analyzed in Sparky with errors determined from the fitted parameters ⁽³²⁾. Uncertainties in these rates were determined from replicate spectra with duplicate relaxation delays of 20 (x2), 60 (x2), 100, 200, 600 (x2), 800, and 1200 ms for T_1 and 16.9, 33.9 (x2), 50.9 (x2), 67.8 (x2), 84.8, 101.8 (x2) ms for T_2 . Steady-state ^1H - ^{15}N NOEs ⁽³⁶⁾ were measured with a 6 s relaxation delay followed by a 3 s saturation (delay) for the saturated (unsaturated) experiments and calculated by $I_{\text{sat}}/I_{\text{ref}}$.

Model-free analysis was conducted via fits to five different forms of the spectral density function using local t_m , spherical, prolate spheroid, oblate spheroid, or ellipsoid diffusion tensors⁽³⁷⁻⁴²⁾. The criteria for inclusion of resonances in the diffusion tensor estimate were based on the method of Bax and coworkers⁽⁴³⁾. N-H bond lengths were assumed to be 1.02 Å and the ^{15}N chemical shift anisotropy tensor was -160 ppm. Diffusion tensor parameters were optimized simultaneously in RELAX under the full automated protocol ⁽⁴⁴⁾. Model selection was iterated until tensor and order parameters did not deviate from the prior iteration.

CPMG relaxation dispersion experiments were carried out with a constant relaxation period of 20 ms and ν_{CPMG} values of 0, 25, 50 (x2), 75, 100, 150, 250, 500 (x2), 750, 800 (x2), 900, and 1000 Hz⁽⁴⁵⁾.

(2)

Uncertainties in these rates were determined from replicate spectra with duplicate relaxation delays of 0, 25, 50 ($\times 2$), 75, 100, 150, 250, 500 ($\times 2$), 750, 800 ($\times 2$), 900, and 1000 Hz. All relaxation experiments were carried out in a temperature-compensated interleaved manner.

Histidine $\epsilon 1$ sidechain carbon chemical shifts were measured via HMQC (*hmqcphpr* pulse program) using NMR samples of *Geo*HNH bound to WT or S78A AcrIIC1. The $\epsilon 1$ carbons of the *Geo*HNH histidine sidechains were assigned by mutagenesis to alanine.

Microscale thermophoresis (MST)

MST experiments were performed on a NanoTemper Monolith X instrument (NanoTemper Technologies) at 25 °C. Experiments quantifying protein-guide RNA interactions used full-length *Geo*Cas9 in complex with WT or S78A AcrIIC1. Signal was detected using a 39 nucleotide Cy5-labeled guide RNA at a concentration of 20 nM in a buffer of 20 mM sodium phosphate, 150 mM KCl, 5 mM MgCl₂, and 0.1% Triton X-100 at pH 7.6. The *Geo*Cas9 protein was serially diluted from a 115 μ M stock into 16 microcentrifuge tubes and combined in a 1:1 molar ratio with serially diluted guide RNA from a 40 nM stock. After incubation for 5 min at 37 °C in the dark, each sample was loaded into an MST capillary (NanoTemper Technologies) for measurement. K_d values for the various complexes were calculated using the MO Control software (NanoTemper Technologies). Statistical significance was calculated using a two-tailed T-test.

MST experiments quantifying protein-protein interactions used *Geo*Cas9 or *Geo*HNH and WT, S78A, or C79P AcrIIC1. WT AcrIIC1 and its mutants contained uncleaved 6x-His affinity tags that were used for attachment of Red-tris as a label. The binding affinity of the Red-tris dye for the His-tag was 2 nM. Full-length *Geo*Cas9 or *Geo*HNH were serially diluted from stock concentrations of 95 μ M and 200 μ M, respectively into microcentrifuge tubes and 100 nM of Red-

tris labeled AcrIIC1 was added to each tube. After incubation for 5 min at 37 °C in the dark, each sample was loaded into an MST capillary for measurement. K_d values for the various complexes were calculated using the MO Control software (NanoTemper Technologies). Statistical significance was calculated using a two-tailed T-test.

Circular dichroism (CD) spectroscopy

AcrIIC1 proteins were exchanged into a buffer of 20 mM sodium phosphate at pH 7.5 via Amicon centrifugal cell. Samples were then diluted to 1 μ M and loaded into a 1 mm quartz cuvette (JASCO Instruments). A CD spectrum was first measured between 200 - 250 nm, after which the sample was progressively heated from 25 - 90 °C in 1.5 °C increments while ellipticity was monitored at 222 and 208 nm. Measurements made on a buffer-only baseline were subtracted from the sample measurements.

Generation of the GeoHNH:AcrIIC1:AcrIIC1 homology model

Protein complex models were generated with the multimer mode of AlphaFold3. Protein sequences for each chain were taken from GeoHNH and WT AcrIIC1, with signal peptides and affinity tags removed. All chains were submitted together to specify the intended 2 AcrIIC1:1 GeoHNH stoichiometry of the complex.

AlphaFold3 was run with its standard pipeline to build multiple sequence alignments and search for structural templates. Both template-based and no-template runs were performed using default settings, unless stated otherwise. For each complex, 5–10 models were generated using different random seeds.

Models were ranked using AlphaFold's confidence scores, including pLDDT for local per-residue confidence and combined ipTM/pTM for overall complex and interface quality. The highest-ranked model was chosen for analysis. Predicted aligned error (PAE) plots and visual inspection in PyMOL were used to assess the reliability of domain and subunit positions; regions with low confidence were interpreted cautiously. Figures of the subsequent AlphaFold structures were generated in PyMOL.

X-ray Crystallography

*Geo*HNH-AcrIIC1 protein complexes used for crystallization were purified as described above and stored in a crystallization buffer of 20 mM HEPES and 80 mM KCl at pH 8.0. The protein complex was formed by incubation *Geo*HNH and AcrIIC1 proteins at a 1:1 molar ratio for 30 minutes on ice. The complex mixture was then loaded onto an EnrichSEC column and eluted using crystallization buffer. Fractions containing both proteins were used to grow crystals using sitting drop vapor diffusion at room temperature by mixing 0.2 μ L of 23 mg/mL *Geo*HNH with 0.4 μ L of crystallizing condition from the PEG/Ion (Hampton Research) well E6: containing 0.2 M sodium malonate at pH 6.0 and 20% (w/v) polyethylene glycol 3350. Crystals were looped and frozen using 20% polyethylene glycol as a cryoprotectant. Diffraction images were collected using a Rigaku MicroMax-003i sealed tube X-ray generator with a Saturn 944 HG CCD detector. Images were processed using XDS and Aimless in CCP4^(47, 48). An X-ray structure of AcrIIC1 bound to the *N. meningitidis* Cas9 HNH domain (PDB ID: 8IF0) was used for molecular replacement with Phaser followed by AutoBuild in Phenix⁽⁴⁹⁾. The *Geo*HNH-AcrIIC1 structure was finalized through manual building in Coot and refinement in Phenix^(49, 50).

DNA cleavage assays

The inhibition of AcrIIC1 and its variants against *GeoCas9* cleavage of DNA substrates was assayed as described previously⁽¹⁴⁾. Briefly, a 1 μ L volume of 0, 6, 18, 30, or 60 μ M AcrIIC1 was mixed with 6 μ M *GeoCas9* in a 1x reaction buffer of 20 mM Tris, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, and 5% glycerol at pH 7.5 and incubated at 37 °C for 10 min. The mixture was then placed on ice and 1 μ L of 6 μ M Tnnt2 sgRNA was added along with 1 μ L of 140.3 ng/ μ L PCR-amplified Tnnt2 DNA substrates and 5 μ L of 2x reaction buffer. After gentle mixing and a transient spin, the tubes were incubated at 37 °C for 30 min, followed by an incubation at 85 °C for 10 min to terminate the cleavage reaction. Finally, 2 μ L of 6x DNA loading buffer was added and 5.5 μ L of that reaction mixture was subsequently run on a 1.5% agarose gel. DNA band intensity quantitation was carried out in ImageJ. A negative control reaction without sgRNA and any form of the AcrIIC1 inhibitor was also included.

The sgRNA sequence, with 23-nt spacer underlined, was:

UUGCACCUACCUUCUGGAUGUACGUCAUAGUUCCCCUGAGAAAUCAGGGUUACUA
UGAUAAGGGCUUUCUGCCUAAGGCAGACUGACCCGCGGCGUUGGGGAUCGCCUGU
CGCCCGCUUUUGGCGGGCAUUCCCCAUCCUU

The Tnnt2 DNA sequence, with **PAM** in bold and the 23-nt spacer underlined, was:

CAAAGAGCTCCTCGTCCAGTGGGAAGAGAGCTGATCTCATTTGTAAGGAATACCCTC
TTCATCCCCCACCCTTGCCATTGATCTATTCATTCCATCTCCATGACAACAGGAAGA
GAGGGCCCGGCGTGAGGAGGAGGAGAACAGGAGGAAGGCTGAGGATGAGGCCCCGG
AAGAAGAAGGCTCTGTCCAACATGATGCACT**TTGGAGGGTACATCCAGAAGGTAG**
GTGCAAAGCAGCATCGGGCACCAGGACACCCCAGTGTATCCTCAAGGCCGCCTTTG
CTTGGATCCATGAAGAAATTCCCAACTGCTGTGGCTGAAGTCTAAGGTCTGCTCATG

TCTAGCCCCTGAGCTGTCTATCAGCCTGACCATGGCTTCAGTAGGAGGGCTCTGCTG
TGTGTGACAGTTAGAACACTAATATGTCTCCAAATTCTGGCTCCCCAAAGGGACAAC
TGGGAGAATCTTGGGTCCTGGAGTCCAT)

MD simulations

All-atom molecular dynamics simulations were performed on the *Geo*HNH–AcrIIC1 complex in its WT form and its three variants; S78A, C79P, and S78A/C79P. Initial coordinates were derived from the corresponding X-ray crystal structures of the WT and mutant complexes. Each system consisted of the *Geo*HNH domain bound to AcrIIC1 and was simulated independently in three replicas.

System preparation and simulation protocols were adapted from workflows previously developed and validated for CRISPR–Cas nucleoprotein complexes^(23, 51-53). Each complex was solvated in explicit water using rectangular periodic boxes extending at least 12 Å from any solute atom, yielding systems of approximately 32,000 atoms. Appropriate numbers of counterions were added to ensure electrical neutrality and to reproduce physiological salt conditions. All simulations were carried out using the AMBER ff19SB19 force field for proteins, together with OL15 and OL3 corrections for DNA and RNA, respectively. Solvent molecules were represented using the TIP3P water model⁽⁵⁴⁻⁵⁷⁾. Hydrogen-containing bonds were constrained using the LINCS algorithm, enabling a 2 fs integration time step. Long-range electrostatic interactions were evaluated using the particle mesh Ewald method with a real-space cutoff of 10 Å and van der Waals interactions were truncated at the same distance. Energy minimization was first performed to eliminate unfavorable contacts while restraining the protein backbone with harmonic potentials of 100 kcal·mol⁻¹·Å⁻². Systems were then gradually heated from 0 K to 300 K under constant-volume

conditions, followed by equilibration in the isothermal–isobaric ensemble. Temperature regulation was achieved using Langevin dynamics with a collision frequency of 1 ps⁻¹, and pressure was maintained at 1 atm using the Berendsen barostat^(58, 59). After equilibration, restraints were removed and production simulations were carried out in the NVT ensemble. Each system was simulated for 2 μs per replica, with three independent replicas, yielding 6 μs of sampling per variant and a total simulation time of 24 μs. Equations of motion were propagated using the leapfrog Verlet algorithm. All simulations were performed using the GPU-accelerated version of AMBER 22⁽⁶⁰⁾. Trajectory analyses were conducted on the combined ensembles using standard AMBER tools and custom scripts.

To characterize differences in the stability of HNH-AcrIIC1 intermolecular contacts across systems, residue-residue contact maps were first computed for each trajectory. A contact between two residues was defined as formed when any pair of heavy atoms were within a distance cutoff of ≤ 4.5 Å. For each residue pair, a contact persistence frequency (f) was calculated as the fraction of trajectory frames in which the contact was present relative to the total number of analyzed frames. To enable a direct comparison between systems, only **common contacts** (those observed at least once in both the WT and variants) were considered. The difference in contact stability was then quantified as:

$$\Delta f = f_{WT} - f_{var}$$

where f_{WT} and f_{var} represent the contact persistence frequencies in the GeoHNH-WT AcrIIC1 and GeoHNH–mutant AcrIIC1 complexes, respectively. Positive values of Δf indicate contacts that are more stable in the WT system, whereas negative values indicate contacts that are more stable in the variant.

Binding free energy calculations

The binding free energy (ΔG_{bind}) of HNH and AcrIIC1 was calculated using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method. For each system, the average binding energy (ΔG_{bind}) values were calculated from the independent windows chosen at a regular interval of 2 ns during the 20 ns of the stable trajectory. The binding free energy was obtained from the difference between the free energies of the protein-protein complex (G_{complex}), unbound protein (G_{HNH}) and the protein (G_{AcrII}):

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{HNH}} - G_{\text{AcrII}}$$

In Amber, these binding energy values are computed from the changes in the molecular mechanical energy (ΔE_{MM}), entropic contribution and the change in solvation free energy due to the substrate binding to the protein for complex formation:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S$$

$$\Delta H = \Delta E_{\text{MM}} + \Delta G_{\text{polar}} + \Delta G_{\text{nonpolar}}$$

The term (ΔE_{MM}) includes electrostatic and van der Waals interactions between the substrate and protein. The ΔG_{polar} is estimated by using the Generalized born (GB) model and the $\Delta G_{\text{nonpolar}}$ is obtained from the hydrophobic contributions calculated from the solvent accessible surface area. To identify sequence-specific contributions, MM/GBSA energy decomposition was performed at the per-base level.

Acknowledgments

This work was supported by NIH grant R01 GM136815 (to GP and GPL) and NSF grant (MCB 2143760 to GPL). ALK was supported by a Blavatnik Family Graduate Fellowship in Biology & Medicine.

Data Availability

X-ray crystal structures of *Geo*HNH bound to WT AcrIIC1 and S78A AcrIIC1 were deposited under PDB IDS 10VC and 10VB, respectively.

References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*. 2012;337(6096):816–21.
2. Chen K, Han H, Zhao S, Xu B, Yin B, Lawanprasert A, et al. Lung and liver editing by lipid nanoparticle delivery of a stable CRISPR–Cas9 ribonucleoprotein. *Nature Biotechnology*. 2025;43(9):1445–57.
3. Li T, Yang Y, Qi H, Cui W, Zhang L, Fu X, et al. CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduction and Targeted Therapy*. 2023;8(1).
4. Zhang S, Wang Y, Mao D, Wang Y, Zhang H, Pan Y, et al. Current trends of clinical trials involving CRISPR/Cas systems. *Frontiers in Medicine*. 2023;10.
5. Sousa AA, Terrey M, Sakai HA, Simmons CQ, Arystarkhova E, Morsci NS, et al. In vivo prime editing rescues alternating hemiplegia of childhood in mice. *Cell*. 2025;188(16):4275–94.e23.
6. Anastassopoulou C, Tsakri D, Panagiotopoulos A-P, Saldari C, Sagona AP, Tsakris A. Armed Phages: A New Weapon in the Battle Against Antimicrobial Resistance. *Viruses*. 2025;17(7):911.
7. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology*. 2013;31(9):822–6.
8. Banani SF, Lee HO, Hyman AA, Rosen MK. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol*. 2017;18(5):285–98.

9. Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature*. 2017;550(7676):407–10.
10. Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nature Biotechnology*. 2018;36(3):265–71.
11. Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016;529(7587):490–5.
12. Belato HB, Norbrun C, Luo J, Pindi C, Sinha S, D’Ordine AM, et al. Disruption of electrostatic contacts in the HNH nuclease from a thermophilic Cas9 rewires allosteric motions and enhances high-temperature DNA cleavage. *The Journal of Chemical Physics*. 2022;157(22):225103.
13. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*. 2017;8(1).
14. Belato HB, Knight AL, D’Ordine AM, Pindi C, Fan Z, Luo J, et al. Structural and dynamic impacts of single-atom disruptions to guide RNA interactions within the recognition lobe of *Geobacillus stearothermophilus* Cas9. *eLife*. 2025;13.
15. Eggers AR, Chen K, Soczek KM, Tuck OT, Doherty EE, Xu B, et al. Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *Cell*. 2024;187(13):3249–61.e14.

16. Shen P, Zhang L, Liu B, Li X, Wang C, Min J, et al. Structure of *Geobacillus stearothermophilus* Cas9: Insights into the Catalytic Process and Thermostability of CRISPR-Cas9. *ACS Catalysis*. 2024;14(17):13227–35.
17. Harrington LB, Doxzen KW, Ma E, Liu J-J, Knott GJ, Edraki A, et al. A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell*. 2017;170(6):1224–33.e15.
18. Zheng I, Learn B, Bailey S. Structural basis for inhibition of SpyCas9 by the anti-CRISPR protein AcrIIA26. *Biochemical Journal*. 2026;483(3):289–300.
19. Kang YW, Park HH. The anti- CRISPR protein AcrIE8 .1 inhibits the type I-E CRISPR-Cas system by directly binding to the. *FEBS Letters*. 2025.
20. Yu L, Yin M, Zhu Y, Lu Z, Xiao B, Zhou F, et al. An anti-CRISPR targets the sgRNA to block Cas9 and guides the design of enhanced genome editors. *Nature Structural & Molecular Biology*. 2026.
21. Kim Y, Lee SJ, Yoon HJ, Kim NK, Lee BJ, Suh JY. Anti-CRISPR AcrIIC3 discriminates between Cas9 orthologs via targeting the variable surface of the HNH nuclease domain. *The FEBS Journal*. 2019;286(23):4661–74.
22. Zhu Y, Gao A, Zhan Q, Wang Y, Feng H, Liu S, et al. Diverse Mechanisms of CRISPR-Cas9 Inhibition by Type IIC Anti-CRISPR Proteins. *Molecular Cell*. 2019;74(2):296–309.e7.
23. Skeens E, Sinha S, Ahsan M, D’Ordine AM, Jogl G, Palermo G, et al. High-fidelity, hyper-accurate, and evolved mutants rewire atomic-level communication in CRISPR-Cas9. *Science Advances*. 2024;10(10).

24. Wang J, Maschietto F, Qiu T, Arantes PR, Skeens E, Palermo G, et al. Substrate-independent activation pathways of the CRISPR-Cas9 HNH nuclease. *Biophysical Journal*. 2023;122(24):4635–44.
25. Wang J, Arantes PR, Ahsan M, Sinha S, Kyro GW, Maschietto F, et al. Twisting and swiveling domain motions in Cas9 to recognize target DNA duplexes, make double-strand breaks, and release cleaved duplexes. *Frontiers in Molecular Biosciences*. 2023;9.
26. Belato HB, D'Ordine AM, Nierzwicki L, Arantes PR, Jogl G, Palermo G, et al. Structural and dynamic insights into the HNH nuclease of divergent Cas9 species. *Journal of Structural Biology*. 2022;214(1):107814.
27. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*. 2024;630(8016):493–500.
28. Zhu YL. NmeHNH-AcrIIC1 Complex Protein Data Bank, PDB. 2024.
29. Marino ND, Pinilla-Redondo R, Csörgő B, Bondy-Denomy J. Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature Methods*. 2020;17(5):471–9.
30. Li Y, Bondy-Denomy J. Anti-CRISPRs go viral: The infection biology of CRISPR-Cas inhibitors. *Cell Host & Microbe*. 2021;29(5):704–14.
31. Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*. 1995;6(3).
32. Lee W, Tonelli M, Markley JL. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*. 2015;31(8):1325–7.
33. Zhao Y, Hu J, Yang S-S, Zhong J, Liu J, Wang S, et al. A redox switch regulates the assembly and anti-CRISPR activity of AcrIIC1. *Nature Communications*. 2022;13(1).

34. Grzesiek S, Stahl SJ, Wingfield PT, Bax A. The CD4 Determinant for Downregulation by HIV-1 Nef Directly Binds to Nef. Mapping of the Nef Binding Surface by NMR. *Biochemistry*. 1996;35(32):10256–61.
35. Lakomek N-A, Ying J, Bax A. Measurement of ¹⁵N relaxation rates in perdeuterated proteins by TROSY-based methods. *Journal of Biomolecular NMR*. 2012;53(3):209–21.
36. Zhu G, Xia Y, Nicholson LK, Sze KH. Protein Dynamics Measurements by TROSY-Based NMR Experiments. *Journal of Magnetic Resonance*. 2000;143(2):423–6.
37. Brüschweiler R, Liao X, Wright PE. Long-range Motional Restrictions in a Multidomain Zinc-finger Protein from Anisotropic Tumbling. *Science*. 1995;268:886–9.
38. Mandel AM, Akke M, Palmer AG. Backbone Dynamics of Escherichia coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J Mol Biol*. 1995;246:144–63.
39. Fushman D, Cahill S, Cowburn D. The Main-chain Dynamics of the Dynamin Pleckstrin Homology (PH) Domain in Solution: Analysis of ¹⁵N Relaxation with Monomer/Dimer Equilibration. *J Mol Biol*. 1997;266:173–94.
40. Orekhov VY, Korzhnev DM, Diercks T, Kessler H, Arseniev AS. ¹H-¹⁵N NMR Dynamic Study of an Isolated α -helical Peptide (1-36)bacteriorhodopsin Reveals the Equilibrium Helix-coil Transitions. *J Biomol NMR*. 1999;14:345–56.
41. Korzhnev DM, Billeter M, Arseniev AS, Orekhov VY. NMR Studies of Brownian Tumbling and Internal Motions in Proteins. *Prog NMR Spectros*. 2001;38:197–266.
42. Zhuravleva AV, Korzhnev DM, Kupce E, Arseniev AS, Billeter M, Orekhov VY. Gated Electron Transfers and Electron Pathways in Azurin: A NMR Dynamic Study at Multiple Fields and Temperatures. *J Mol Biol*. 2004;342:1599–611.

43. Tjandra N, Feller SE, Pastor RW, Bax A. Rotational Diffusion Anisotropy of Human Ubiquitin from N-15 NMR Relaxation. *J Am Chem Soc.* 1995;117:12562–6.
44. Bieri M, d'Auvergne EJ, Gooley PR. relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μ s motion of proteins. *J Biomol NMR.* 2011;50(2):147–55.
45. Loria JP, Rance M, Palmer AG. A Relaxation-Compensated Carr–Purcell–Meiboom–Gill Sequence for Characterizing Chemical Exchange by NMR Spectroscopy. *Journal of the American Chemical Society.* 1999;121(10):2331–2.
46. Luz Z, Meiboom S. Nuclear Magnetic Resonance Study of the Protolysis of Trimethylammonium Ion in Aqueous Solution—Order of the Reaction with Respect to Solvent. *The Journal of Chemical Physics.* 1963;39(2):366–70.
47. Kabsch W. *XDS*. *Acta Crystallographica Section D Biological Crystallography.* 2010;66(2):125–32.
48. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the *CCP4* suite and current developments. *Acta Crystallographica Section D Biological Crystallography.* 2011;67(4):235–42.
49. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallographica Section D Structural Biology.* 2019;75(10):861–77.
50. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of *Coot*. *Acta Crystallographica Section D Biological Crystallography.* 2010;66(4):486–501.

51. Patel AC, Sinha S, Arantes PR, Palermo G. Unveiling Cas8 dynamics and regulation within a transposon-encoded Cascade–TniQ complex. *Proceedings of the National Academy of Sciences*. 2025;122(14).
52. Sinha S, Adrian, Pablo, Patel A, Mitchell, Palermo G. Unveiling the RNA-mediated allosteric activation discloses functional hotspots in CRISPR-Cas13a. *Nucleic Acids Research*. 2024;52(2):906–20.
53. Pablo, Chen X, Sinha S, Saha A, Amun, Sample M, et al. Dimerization of the deaminase domain and locking interactions with Cas9 boost base editing efficiency in ABE8e. *Nucleic Acids Research*. 2024;52(22):13931–44.
54. Tian C, Kasavajhala K, Belfon KAA, Raguet L, Huang H, Migués AN, et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation*. 2020;16(1):528–52.
55. Galindo-Murillo R, Robertson JC, Zgarbová M, Šponer J, Otyepka M, Jurečka P, et al. Assessing the Current State of Amber Force Field Modifications for DNA. *Journal of Chemical Theory and Computation*. 2016;12(8):4114–27.
56. Zgarbová M, Otyepka M, Šponer J, Mládek A, Banáš P, Cheatham TE, et al. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *Journal of Chemical Theory and Computation*. 2011;7(9):2886–902.
57. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983;79(2):926–35.

58. Turq P, Lantelme F, Friedman HL. Brownian dynamics: Its application to ionic solutions. *The Journal of Chemical Physics*. 1977;66(7):3039–44.
59. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*. 1984;81(8):3684–90.
60. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. *Journal of Chemical Information and Modeling*. 2023;63(20):6183–91.

CHAPTER FOUR

Future Considerations for the Study of Allostery in Thermophilic Enzymes

This chapter is adapted from a published manuscript in Frontiers in Molecular Biosciences (2023) titled:

Temperature as a Modulator of Allosteric Motions and Crosstalk in Mesophilic and Thermophilic Enzymes

Alexa L. Knight^{1,*}, Vinnie Widjaja^{1,*}, George P. Lisi^{1,2}

¹Department of Molecular Biology, Cell Biology, & Biochemistry, Brown University, Providence, RI 02903 USA

²Giuliani RNA Center, Brown University, Providence, RI USA

*Co-first authors

Author Contributions

AK: Writing—original draft, Writing—review and editing, Figure generation, Data curation, Funding acquisition, Investigation. **VW:** Data curation, Funding acquisition, Investigation, Writing—original draft, Writing—review and editing. **GPL:** Writing—review and editing, Conceptualization, Supervision.

Abstract

Mesophilic and thermophilic enzyme counterparts are often studied to understand how proteins function under harsh conditions. To function well outside of standard temperature ranges, thermophiles often tightly regulate their structural ensemble through intra-protein communication (via allostery) and altered interactions with ligands. It has also become apparent in recent years that the enhancement or diminution of allosteric crosstalk can be temperature-dependent and distinguish thermophilies from their mesophilic paralogs. Since most studies of allostery utilize chemical modifications from pH, mutations, or ligands, the impact of temperature on allosteric function is comparatively understudied. As this thesis outlines the study of a thermophilic enzyme compared to its highly studied mesophilic counterpart, it is important to discuss the wealth of biophysical methods, as well as critical case studies, that dissect temperature-dependent function of mesophilic-thermophilic enzyme pairs and their allosteric regulation across a range of temperatures.

Introduction

Allostery is a regulatory mechanism that spatially and functionally couples distinct regions of a protein, allowing multiple sites to communicate. The use of allosteric modulators, such as small molecules, peptides, or mutations, can alter enzymatic activity or substrate binding affinity via changes in structure or dynamics of the protein. The idea that subtle conformational motions can affect the transmission of chemical information over small and large molecular distances has been refined over the past decade, as an increasing number of allosteric systems only partially conform to classical paradigms necessitating specific sequences of biomolecular events. Experiments paired with theoretical underpinnings have confirmed that allostery can be driven

primarily through protein motions and that allosteric effectors do not necessarily activate a protein through conformational change⁽¹⁻³⁾. Central to this observation is the existence of amino acid networks that relay chemical information to modulate the degree of coupling between active and allosteric sites. “Ensemble” allostery has been observed in both mesophilic and thermophilic organisms, where the effect of temperature on conformational ensembles, protein motions, and molecular crosstalk has been investigated to understand enzyme function at physiological extremes.

Temperature plays an important role in allosteric regulation, particularly in thermophilic enzymes that remain active over very wide environmental ranges⁽⁴⁻⁷⁾. For example, K-type enzymes (allosteric effector modulates K_D) functioning at elevated temperatures have been found in some cases to exhibit an enhancement of allosteric function, while in others the allosteric effect is inverted^(7, 8). Temperature-induced allosteric modulation in most K-type enzymes, as expected, is driven by changes in substrate affinity upon allosteric ligand binding⁽⁸⁻¹⁰⁾. Beyond this effect, correlations between temperature elevation and enzymatic activity do not reveal consistent patterns regarding allosteric activation^(10, 11), thus, the molecular mechanisms underlying temperature-modulated allostery remain unclear.

As a consequence, allosteric drug design against target proteins from extremophilic organisms (for example, opportunistic pathogens that survive processes like pasteurization) or for enzyme regulation under harsh conditions is understudied⁽⁹⁾. Of particular relevance to this thesis, is the consideration that although multiple thermophilic Cas enzymes have been discovered there is little structural, biophysical, or mechanistic information known about these systems. Furthermore, atomic-level interrogations at elevated temperatures are completely lacking. Additionally, the temperature dependence of a classical Boltzmann distribution describing kinetic

energies for a set of molecules (**Figure 1**) can be extended to a description of protein conformational dynamics⁽¹²⁾. Energy levels within such thermodynamic equilibria become increasingly degenerate as temperature increases, thereby allowing an enzyme to escape kinetics barriers that drastically suppress active populations near room temperature. Allosteric communication and processes resulting from energetic connections between distal molecular sites correlate to the Boltzmann distribution of state transitions as temperature increases. In this chapter, the importance of temperature in distinguishing characteristics of allosteric enzymes and highlight case studies of enzymes that have temperature-dependent structures, dynamics, or functions as well as meso-thermophilic paralogs that maintain structure and stability but differ in key biophysical ways will be discussed. This chapter lays the foundation for a major arm of future study for *GeoCas9*, as detailed temperature dependent mechanistic studies have yet to be undertaken. Also highlighted, are current solution NMR methods for interrogating the influence of temperature on protein conformation and dynamics.

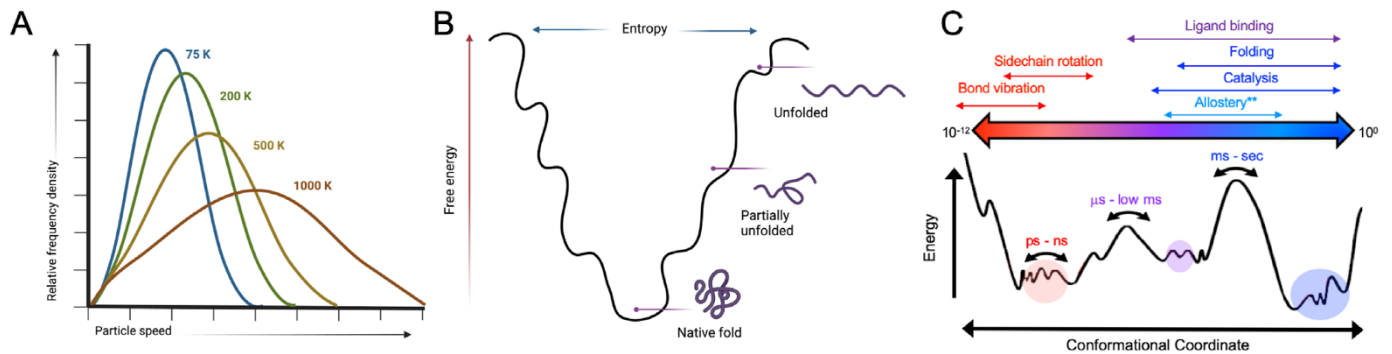


Figure 1. Kinetic and thermodynamic aspects of protein conformational equilibria are temperature-sensitive. (A) Temperature dependence of a Boltzmann distribution showing increased degeneracy at higher temperatures. Created with BioRender. (B) Protein folding landscapes rely on a delicate balance of enthalpy and entropy that are sensitive to temperature. Created with BioRender. (C) Many biological processes have distinct temporal and energetic requirements that define protein conformational states. Variable temperatures can skew energetic barriers that rewire protein motions that directly influence biochemical mechanisms.

Current Solution Nuclear Magnetic Resonance (NMR) Spectroscopy Tools for Interrogating Allosteric Systems via Temperature

As discussed in the Introduction, solution NMR spectroscopy is a premier molecular tool to elucidate structural and dynamic changes in allosteric systems. NMR is highly sensitive to local environments (*i.e.* subtle changes to protein structures) and can probe molecular motion on a wide range of biological timescales (picosecond-second). Of critical importance are the biologically slow motions that traditionally define conformational equilibria associated with protein folding, enzyme catalysis, and allosteric communication (**Figure 1**)^(13, 14). NMR experiments designed to quantify protein motion related to these processes are called relaxation dispersion. These experiments provide kinetic, thermodynamic, and structural information as micro-millisecond protein motions broaden NMR resonances by increasing the transverse relaxation rate (R_2) by an amount R_{ex} . In the fast exchange limit, R_{ex} can be described as

$$R_{ex} = \frac{P_A P_B \Delta\omega^2}{k_{ex}} \quad (1)$$

where k_{ex} is the chemical exchange rate between the assumed two conformational states (kinetics), P_A and P_B are the equilibrium populations of the conformers (thermodynamics), and $\Delta\omega$ is the chemical shift difference between conformers (structure). The use of Eqn. 1 in the mathematical expressions for single- and multiple-quantum Carr-Purcell-Meiboom-Gill (CPMG) and $R_{1\rho}$ relaxation experiments has been derived elsewhere^(13, 14).

Solution NMR can be performed under a variety of physiological pH or solvent conditions⁽¹⁴⁾, but of particular interest here is the effect of temperature on the parameters described above. Indeed, some earlier NMR studies have outlined structural and dynamic alternations in proteins across wide temperature ranges⁽¹⁵⁻¹⁷⁾, while others have focused on utilizing temperature variations to probe the underlying causes of evolutionary divergence in pairs of thermophilic and

mesophilic enzymes⁽¹⁸⁻²⁰⁾. In the latter case, it is suggested that temperature-dependent allosteric regulation (i.e. where allosteric effects “accelerate” or “dampen” with temperature) may be an evolutionary feature to maintain maximal catalytic efficiency of an enzyme over a wide range^(21, 22). Enzymes classified as either mesophilic or thermophilic often display similar dynamic signatures under “traditional” regimes of 273-290 K. Conformational fluctuations that distinguish these enzyme systems from one another become more apparent as the temperature approaches substantially higher (or lower) regimes. This can manifest as a change in k_{ex} , structure ($\Delta\omega$), or population of active conformer at a given temperature. The relationship between k_{ex} and $\Delta\omega$ influences how the NMR resonances appear and may be manipulated by temperature to provide a quantitative link between protein motion and activation (via $\Delta G^\ddagger$, which is related to E_a)^(23, 24).

Scenarios where k_{ex} is much greater than, equal to, or much less than $\Delta\omega$ dictate the fast, intermediate, or slow exchange regimes, respectively. At the extremes of this continuum, an NMR resonance appears as a population-weighted average (fast exchange) or as two resonances with populations defined by the peak volumes (slow exchange).

The reshuffling of structural populations is critical to the activation of enzymes as allosteric modulators bias the ensemble toward an active state. Enzyme activation pathways can therefore be observed in real-time by NMR, and changes in sample temperature can alter the way enzymes achieve their functional conformation. In the fast exchange limit, where $k_{ex} \gg \Delta\omega$ and NMR resonances appear as a population-weighted average, the single N-H chemical shift contains information about the major and minor states.

$$\delta_{NH} = p_b \omega_b + p_a \omega_a \quad (2)$$

where $\Delta\omega$ is defined as $|\omega_b - \omega_a|$, the characteristic chemical shifts of each conformer. The relationship between populations of exchanging states can define K ^(25, 26), the equilibrium constant related to ΔG by

$$\Delta G = -RT \ln \left(\frac{p_b}{p_a} \right) \quad (3)$$

The transition state energy for the conversion between states (*i.e.* change in equilibrium populations) can be estimated from the temperature dependence of the NMR linewidth ($\Delta\nu_{1/2}$), fitting to Eqn. 4.

$$\Delta\nu_{1/2} = \frac{1}{\pi T_2} + \frac{K(\omega_b - \omega_a)^2}{4\pi(1+K)} \left(\frac{h}{k_B T} \right) e^{\Delta G^\ddagger / RT} \quad (4)$$

Here, K is defined by the ratio of P_B and P_A . Calculation of $\Delta G^\ddagger$ provides access to the individual components of k_{ex} ($k_{ex} = k_f + k_r$) through the Eyring Equation⁽²⁴⁾.

$$k_f = \kappa \frac{k_B T}{h} e^{-\Delta G^\ddagger / RT} \quad (5)$$

where κ is assumed to be equal to 1. Analysis of k_f and k_r can dissect aspects of k_{ex} that are commensurate with catalytic rates, open-to-close motions of activation loops, or enzyme breathing. Indeed, thermodynamic contributions to the transition state between inactive and active enzyme can be estimated from the linear form of the Eyring equation

$$\ln \frac{k_f}{T} = \frac{\Delta H^\ddagger}{R} \frac{1}{T} + \ln \frac{\kappa_B}{h} + \frac{\Delta S^\ddagger}{R} \quad (6)$$

where transition state enthalpy and entropy are the slope and y-intercept of the plot, respectively. The Arrhenius equation ($k = Ae^{-E_a/RT}$) describing the effect of temperature on reaction rates via

transition state energy is similar and can be considered if k_{ex} is commensurate with reaction rate⁽²⁷⁾. Here, the pre-exponential factor, A , is slightly temperature-dependent, rather than constant and reactions with large E_a increase rapidly with temperature. Some similarity between the Eyring and Arrhenius relationships requires an assumption of constant enthalpies and entropies of activation, nonetheless, reaction rates altered by the presence of allosteric ligands may be highly temperature-dependent, suggesting that NMR studies at multiple temperatures may be required to define a mechanism, particularly for enzymes derived from extremophilic organisms⁽²⁸⁻³⁰⁾.

As allosteric effectors shift the structural ensemble of an enzyme, the weak interactions surrounding the enzyme active site may also be altered substantially. Subtle bond vibrations and side chain reorientations that are essential to the active states of most enzymes require the breaking and formation of van der Waal's contacts and hydrogen bonds. NMR has been successful in estimating hydrogen bond lengths and propensities via temperature-dependent ^1H - ^{15}N chemical shifts. Straightforward HSQC measurements over a wide temperature range provide $\Delta\delta_{\text{HN}}/\Delta T$, where values below a certain threshold (often defined as -4.6 ppb/K in the literature) indicate a significant hydrogen bond propensity for a given residue⁽³¹⁾. The magnitude of $\Delta\delta_{\text{HN}}/\Delta T$ has been correlated to the hydrogen bond length, and evaluation of temperature-dependent ^1H - ^{15}N chemical shifts across a series of point-mutants, or between meso- and thermophilic enzymes of the same class, can provide fairly detailed structural information to complement relaxation dispersion studies^(21, 22, 32).

The Effect of Temperature on Allosteric Communication

Allosteric mechanisms, essential for the transmission of perturbations from an effector binding site to other distant functional sites, are widespread and often depend on protein

conformational fluctuations⁽³³⁾. In fact, biomolecular motions involved in conformational selection provided some of the early insight into dynamically driven allostery that did not conform to classical paradigms⁽³³⁾. The extent and timescale of protein motion is intimately linked to ligand affinity or the degree of enzyme activation via the effective sampling of productive conformational states. The temperature-dependent conformational or kinetic equilibria that define allosteric mechanisms can be tied to entropy⁽³⁴⁻³⁶⁾, change in heat capacity⁽³⁷⁾, or enthalpy⁽³⁸⁾, but many elegant studies have demonstrated that the nature of this long-range coupling is not universal, and the highly varied thermal properties between and within enzyme families can strongly affect allostery. The latter observation prompts questions as to whether allostery is even the primary mechanism at the physiological temperature of many enzymes, particularly thermophiles that are often studied at lower, non-native temperatures. Thus, the temperature-dependence of allostery requires a deeper consideration, particularly for thermophilic enzymes with wide activity ranges. Studies of the adaptability to temperature in several enzymes has shown that enthalpic, and more critically, entropic tuning is a viable method to modulate the energetic favorability of allosteric pathways, which will be discussed in the case studies below.

Case Study: Temperature-dependent Allosteric Activation of a Thermophilic Enzyme

Biophysical studies of thermophilic enzymes offer unique insights into the molecular factors that distinguish them from mesophilic homologs, which is a decades old question of protein biochemistry. Thermophiles are often easier to work with than corresponding mesophiles, and in this case study, provide an opportunity to scrutinize temperature-sensitive aspects of allostery. Loria and coworkers have shown in several studies that the glutamine amidotransferase imidazole glycerol phosphate synthase (IGPS) is a model for V-type allostery that can be structurally and

enzymatically characterized over a very wide temperature range^(4, 39). *Thermatoga maritima* IGPS is a heterodimeric complex of the HisH (glutaminase) and HisF (cyclase) enzymes, where the two catalytic sites are separated by >25 Å yet are functionally synchronized. HisF binds the allosteric activator N'-[(5'-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR), which is entropically driven at room temperature, and enhanced glutamine hydrolysis in HisH nearly 5000-fold over basal levels.

Allosteric activation of Gln hydrolysis by PRFAR is driven by milisecond (ms) timescale motions throughout the HisF domain that propagate to the dimer interface and reposition the glutaminase active site architecture for catalysis. Mutations along the “dynamic pathway” spanning HisF and HisH diminished ms motion and catalytic rates to a similar extent⁽⁴⁶⁾. The link between PRFAR-induced allosteric activation and protein dynamics was established at 303 K, well below the physiological temperature of *T. maritima*⁽⁴⁶⁾. Temperature-dependent NMR studies of ¹H¹³CH₃ ILV-methyl labeled apo IGPS over a range of 303-343 K revealed linear, temperature-dependent shifts in fast exchange ($k_{\text{ex}} \gg \Delta\omega$)⁽⁴⁷⁾. At 343 K, surprisingly, a cluster of methyl resonances appeared to experience slow exchange (**Figure 2**).

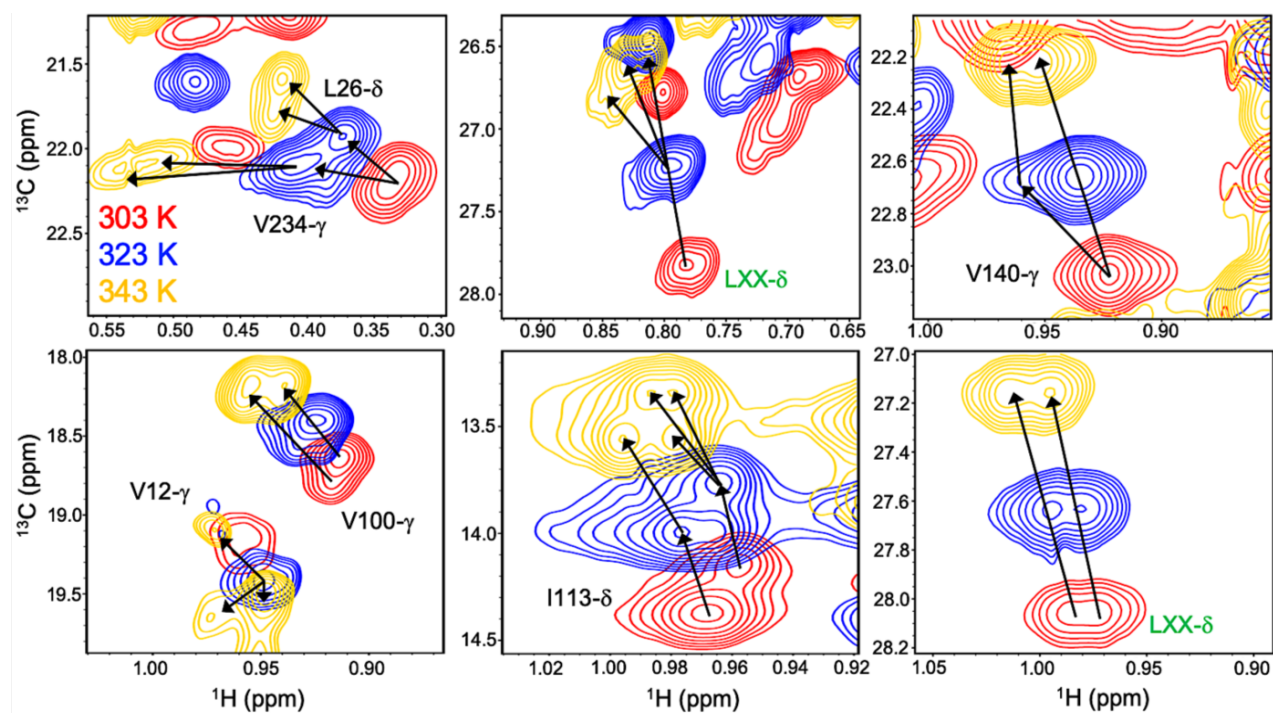


Figure 2. Overlay of ^1H - $^{13}\text{CH}_3$ methyl-ILV NMR spectra collected at 303 K (red), 323 K (blue) and 343 K (gold) depicting temperature-dependent variations in exchange regime. Arrows following chemical shift trajectories bifurcate when minor populations of a given resonance become visible. Here, elevated temperatures more clearly resolve multiple conformations within the HisF subunit of IGPS. Resonances labeled in green were unassigned in the original data presented by Loria and coworkers in *Front. Mol. Biosci.* **2017**.

PRFAR-bound IGPS displayed similar behavior based on ^1H - ^{13}C temperature coefficients, though sites of slow exchange at 343K were redistributed to different residues. Sites found to undergo ms motion in response to PRFAR binding were notable outliers when comparing apo- and PRFAR-bound IGPS temperature coefficients, suggesting that IGPS retains its flexibility at high temperatures.

Interestingly, elevated temperatures enhanced the flexibility of apo IGPS substantially, so much so that the apo enzyme appeared dynamically similar to PRFAR-bound IGPS at 343 K, based on methyl-TROSY relaxation dispersion. Concomitant with this biophysical change, the degree of catalytic rate enhancement afforded by PRFAR at 343 K is only 65-fold (compared with ~5000-fold at 303 K), despite a largely temperature-independent PRFAR binding affinity. Thus, at temperatures more closely approximating the native environment of *T. maritima*, IGPS appears to rely much less heavily on dynamic allosteric activation. To understand the molecular basis for this effect, MD simulations at elevated temperatures showed that temperature-driven increases in Eigenvector Centrality (EC), an established metric to assess dynamic connectivity along an allosteric pathway⁽⁴⁸⁾, mimicked those induced by PRFAR at room temperature. This suggested that temperature alone can compensate for the effect of PRFAR in establishing the allosterically activated IGPS conformation. When PRFAR binding was examined with EC at elevated temperatures, the crosstalk within the IGPS heterodimer encompassed the entire structure, consistent with biochemical data indicating a shift from well-defined allosteric rate enhancement to a “pre-activated” system with a highly flexible apo state (**Figure 3**).

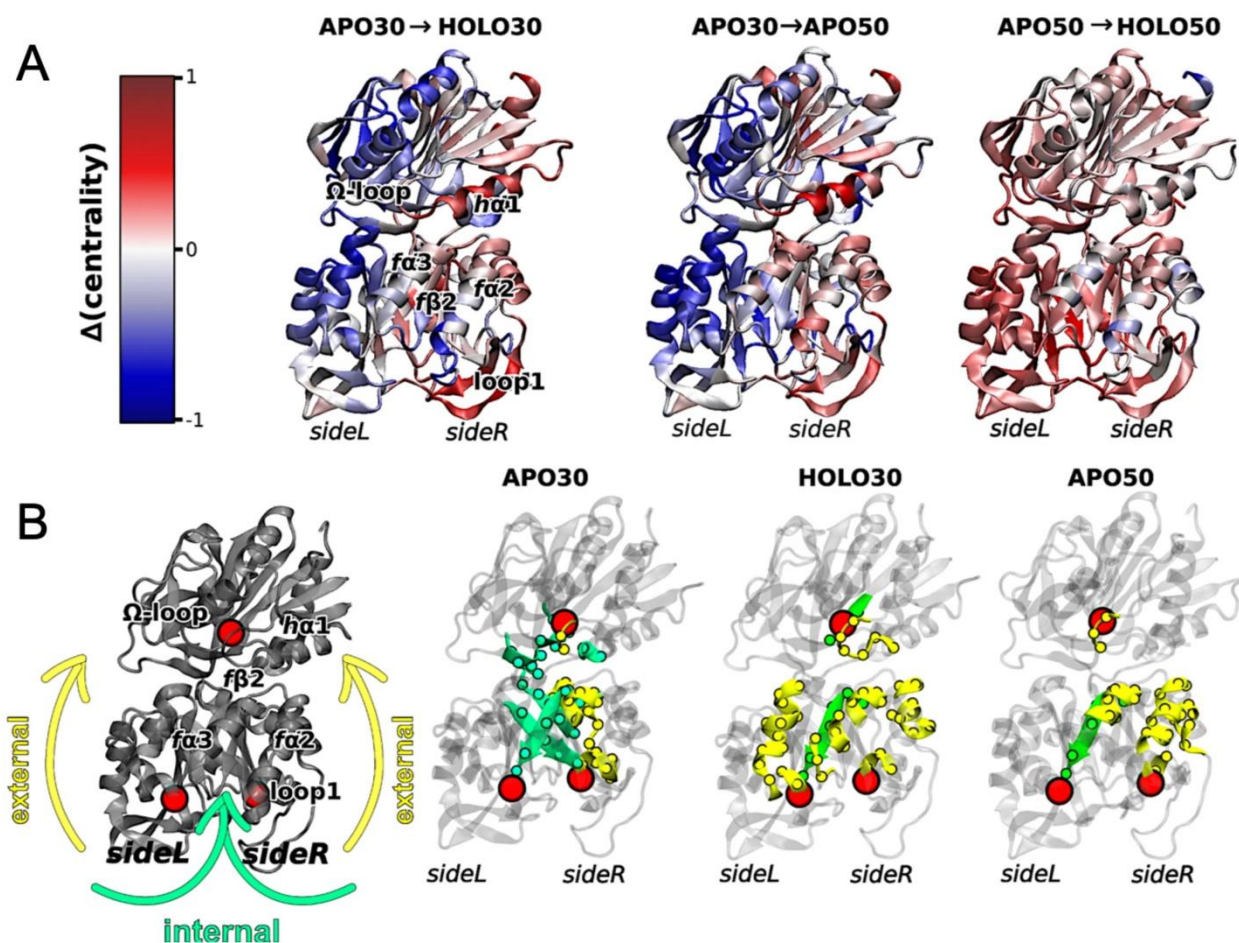


Figure 3. (A) Differences in Eigenvector Centrality within IGPS, intrinsic to the binding of PRFAR (holo enzyme) at 303 K (left) or 323 K (right) and without effector binding (apo enzyme) at the elevated temperature of 323 K (center). Communication pathways induced by PRFAR at 303 K are recapitulated solely due to the effect of temperature in the apo enzyme at 323 K. (B) Comparative transmission of distinct mutual information pathways linking the PRFAR binding site to the active site ($hG50$ in HisH, red sphere). Areas of secondary structure involved in communication pathways are separately highlighted as solvent exposed amino acids (yellow, "external") and buried (green, "internal"). Similar trends are observed in for the holo enzyme at 303 K and apo enzyme at 323 K when compared to apo IGPS (303 K), specifically a transition from internal pathways to external communication due to effector binding and/or temperature increase. Figure was reproduced from Maschietto et al, *Nature Communications*, under a Creative Commons Attribution 4.0 International License.

Case Study: Temperature Adaptation of Dynamic Allostery

Thermal adaptation of organisms to their environmental niches is a hallmark of evolution. When two descendants evolve under different temperature extremes, physicochemical differences (*i.e.* adaptive mutations) arise that change how those enzymes catalyze essential reactions. Primary sequence variations are often found in residues that are surface exposed, while active sites in the protein interior are rarely structurally perturbed, preserving overall catalytic rates. An open question exists regarding how related enzymes with distinct thermally adapted sequences use allosteric changes to propagate information to/from the protein periphery to the active site. One explanation for this phenomenon may be “entropy-tuning,” a dynamics-based allosteric mechanism that allows for the spatial separation of the control of key enzymatic parameters. Hilser and coworkers have demonstrated this effect by engineering the *E. coli* adenylate kinase (Adk) enzyme, where the conformational entropy of several distal regions serves as a metric for dynamically tuned ligand affinity and catalysis⁽⁴⁰⁾. The three domains of Adk, the CORE, AMP-binding, and LID, work in concert to catalyze effective phosphoryl transfer. The highly stable CORE positions catalytic residues to form the substrate binding site, while the flexible LID and AMP-binding domains are spatially locked during binding and until product release. It was initially assumed that conformational changes in the LID and AMP-binding domains were rigid-body motions, but more recent work revealed a ‘local unfolding’ of the LID domain via uncorrelated motions throughout the entire subunit.

Differential scanning calorimetry (DSC) studies of single-point mutations within the LID domain, created to increase the probability of local unfolding without disrupting the ground state structure, showed a clear reduction in the major unfolding transition of wild-type (wt) Adk. A smaller low-temperature transition in LID mutants corresponding to an increased population

(~40%) of a locally unfolded intermediate state was also observed, which comprises only 5% of wt-Adk at 305 K. The thermodynamic modulation of the minor LID conformer was recapitulated in temperature-dependent NMR relaxation experiments of apo- and holo-Adk LID variants. The stabilization of a locally unfolded minor state was attributed to the entropic favorability of removing bulky side chains that would normally restrict partial unfolding. Lower NMR sample temperatures return the LID conformational equilibrium to a wt-Adk level, illustrating quite elegantly that strategies for temperature adaptation can leverage the energetic (*i.e.* entropic) balance of local unfolding, as only single-point mutations on the protein surface are sufficient to energetically bias the stability of the folded state of a domain within an allosteric complex. Olsson and Wolf-Watz provided a similar example using point mutations to partially unfold the ATP binding subdomain that would otherwise remain folded, revealing a novel open-to-close (or unfold-to-refold) molecular mechanism in *E. coli* Adk⁽⁴¹⁾. Additionally, as chaperone machinery has evolved to recognize unfolded states of proteins, local unfolding is an attractive mechanism for the regulation of enzymes via K_m or k_{cat} that are sensitive to a folded-unfolded equilibrium. Additional biophysical characterizations of local disorder in canonical enzymes are required to tie protein-level regulation to that of proteome-level regulation via chaperones.

Conclusions and Considerations for Further Biophysical Interrogations of GeoCas9

In this chapter, an overview of relevant methods, theoretical considerations, tools, and case studies for investigating the temperature-dependent function of mesophilic-thermophilic enzyme pairs and the allosteric regulation of enzymes across a wide temperature range was discussed. The modulation of temperature in solution NMR studies can alter the protein conformational equilibrium and biologically relevant ligand binding configurations that remain elusive in regimes

distant from those of optimal activity. Exploring thermodynamic and structural principles via temperature is crucial for expanding our understanding of the importance of adaptation in protein function, particularly in thermophilic organisms that are often studied at room temperature in the laboratory. The multi-faceted nature of protein dynamics and their interplay with observable structural change makes studies of temperature-dependent allostery highly complex. Indeed, temperature can enhance, suppress, or have no effect on allosteric coupling, and such outcomes cannot be universally ascribed to or predicted by a certain sequence, structure, intrinsic motional profile, or mechanism. This is further underscored by a lack of broadly defined biophysical principles that distinguish mesophilic from thermophilic proteins. Continued research in this field is essential to better understand how laboratory and native environmental temperatures alter chemical crosstalk within proteins.

Furthermore, this work highlights the necessity of temperature-dependent interrogation of *GeoCas9* allostery and DNA cleavage. Although practical use of CRISPR-Cas9 editors often involve incubatory periods at room temperature or on ice, the optimal active temperature for *GeoCas9* is $\sim 65^{\circ}\text{C}$ ⁽⁴²⁾. This chapter makes clear that temperature-dependent motions and energetic landscapes of thermophiles are unique and may only be uncovered at elevated temperatures. Similar studies on *GeoCas9* should probe the biophysical and structural qualities discussed within this thesis at elevated temperatures to fully characterize its mechanism. The human body contains spatially organized temperature gradients, with temperatures differing as much as $\sim 10^{\circ}\text{C}$ between regions and when the body is under stress⁽⁴³⁾. As *GeoCas9* has been highlighted for its potential use as a human therapeutic, studies into any temperature dependent mechanistic fluctuations will provide important context for what body systems and disease contexts would be best suited for *GeoCas9* deployment. Furthermore, as the anti-CRISPR protein AcrIIC1 studied within this thesis

was isolated from the mesophilic *NmeCas9*, future work in exploring mutagenesis-derived thermostable variants of the inhibitor should complement high temperature studies of *GeoCas9*.

References

1. Amaro RE. Toward Understanding "the Ways" of Allosteric Drugs. *ACS Cent Sci.* 2017;3(9):925–6
2. Fenton AW. Allostery: an illustrated definition for the 'second secret of life'. *Trends Biochem Sci.* 2008;33(9):420–5.
3. Kremer DM, Lyssiotis CA. Targeting allosteric regulation of cancer metabolism. *Nat Chem Biol.* 2022;18(5):441–50.
4. Lisi GP, Currier AA, Loria JP. Glutamine Hydrolysis by Imidazole Glycerol Phosphate Synthase Displays Temperature Dependent Allosteric Activation. *Front Mol Biosci.* 2018;5:4.
5. Braxton BL, Tlapak-Simmons VL, Reinhart GD. Temperature-induced inversion of allosteric phenomena. *J Biol Chem.* 1994;269(1):47–50.
6. Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ. Dynamic allostery can drive cold adaptation in enzymes. *Nature.* 2018;558(7709):324–8.
7. McGresham MS, Lovingshimer M, Reinhart GD. Allosteric regulation in phosphofructokinase from the extreme thermophile *Thermus thermophilus*. *Biochemistry.* 2014;53(1):270–8.
8. McGresham MS, Reinhart GD. Enhancing allosteric inhibition in *Thermus thermophilus* Phosphofructokinase. *Biochemistry.* 2015;54(3):952–8.
9. Maschietto F, Morzan UN, Tofoleanu F, Gheeraert A, Chaudhuri A, Kyro GW, et al. Turning up the heat mimics allosteric signaling in imidazole-glycerol phosphate synthase. *Nat Commun.* 2023;14(1):2239.

10. Katava M, Marchi M, Madern D, Sztucki M, Maccarini M, Sterpone F. Temperature Unmasks Allosteric Propensity in a Thermophilic Malate Dehydrogenase via Dewetting and Collapse. *J Phys Chem B*. 2020;124(6):1001–8.
11. Katava M, Maccarini M, Villain G, Paciaroni A, Sztucki M, Ivanova O, et al. Thermal activation of 'allosteric-like' large-scale motions in a eukaryotic Lactate Dehydrogenase. *Sci Rep*. 2017;7:41092.
12. Skrdla PJ. Crystallizations, solid-state phase transformations and dissolution behavior explained by dispersive kinetic models based on a Maxwell-Boltzmann distribution of activation energies: theory, applications, and practical limitations. *J Phys Chem A*. 2009;113(33):9329–36.
13. Cavanagh J. *Protein NMR Spectroscopy: Principles and Practice*. 2nd Ed (9780121644918): Academic Press; 2007.
14. Lisi GP, Loria JP. Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem Rev*. 2016;116(11):6323–69.
15. Morino Y, Yamasaki M, Tanase S, Nagashima F, Akasaka K, Imoto T, et al. ¹H NMR studies of aspartate aminotransferase. Histidyl residues of cytosolic and mitochondrial isoenzymes. *J Biol Chem*. 1984;259(6):3877–82.
16. Weininger U, Modig K, Akke M. Ring Flips Revisited: ¹³C Relaxation Dispersion Measurements of Aromatic Side Chain Dynamics and Activation Barriers in Basic Pancreatic Trypsin Inhibitor. *Biochemistry*. 2014;53(28):4519–25.
17. Jung A, Bamann C, Kremer W, Kalbitzer HR, Brunner E. High-temperature solution NMR structure of TmCsp. *Protein Sci*. 2004;13(2):342–50.

18. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, Kern D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. *Nat Struct Mol Biol.* 2004;11(10):945–9.
19. Butterwick JA, Loria JP, Astrof NS, Kroenke CD, Cole R, Rance M, et al. Multiple time scale backbone dynamics of homologous thermophilic and mesophilic ribonuclease HI enzymes. *J Mol Biol.* 2004;339(4):855–71.
20. Toth K, Amyes TL, Wood BM, Chan KK, Gerlt JA, Richard JP. An examination of the relationship between active site loop size and thermodynamic activation parameters for orotidine 5'-monophosphate decarboxylase from mesophilic and thermophilic organisms. *Biochemistry.* 2009;48(33):8006–13.
21. Palmer AG, 3rd. A dynamic look backward and forward. *J Magn Reson.* 2016;266:73–80.
22. Neudecker P, Lundström P, Kay LE. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J.* 2009;96(6):2045–54.
23. Westwood MN, Johnson CC, Oyler NA, Meints GA. Kinetics and thermodynamics of BI-BII interconversion altered by T:G mismatches in DNA. *Biophys J.* 2022;121(9):1691–703.
24. Ruminski RR, Underwood S, Vallety K, Smith SJ. Synthesis and Characterization of [Ru(dpop')(2)](PF(6))(2). A Ruthenium(II) Complex with the Novel Mixed-Denticity Bridging Ligand Dipyrrodo(2,3-a:3',2'-j)phenazine (dpop'). *Inorg Chem.* 1998;37(25):6528–31.
25. Bain AD, Duns GJ, Rathgeb F, Vanderkloet J. A Study of Chemical Exchange in Unequally Populated Systems by Novel NMR Methodologies. Application to the Cis-Trans Isomerization in Furfural. *The Journal of Physical Chemistry.* 1995;99(48):17338–43.
26. Krishnan VV. Molecular Thermodynamics Using Nuclear Magnetic Resonance (NMR) Spectroscopy. *Inventions (Basel).* 2019;4(1).

27. Carvalho-Silva VH, Coutinho ND, Aquilanti V. Temperature Dependence of Rate Processes Beyond Arrhenius and Eyring: Activation and Transitivity. *Front Chem*. 2019;7:380.
28. Busi B, Yarava JR, Bertarello A, Freymond F, Adamski W, Maurin D, et al. Similarities and Differences among Protein Dynamics Studied by Variable Temperature Nuclear Magnetic Resonance Relaxation. *J Phys Chem B*. 2021;125(9):2212–21.
29. Siemer AB, Huang KY, McDermott AE. Protein linewidth and solvent dynamics in frozen solution NMR. *PLoS One*. 2012;7(10):e47242.
30. Holliday MJ, Camilloni C, Armstrong GS, Vendruscolo M, Eisenmesser EZ. Networks of Dynamic Allostery Regulate Enzyme Function. *Structure*. 2017;25(2):276–86.
31. Cierpicki T, Otlewski J. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J Biomol NMR*. 2001;21(3):249–61.
32. Belato HB, Norbrun C, Luo J, Pindi C, Sinha S, D'Ordine AM, et al. Disruption of electrostatic contacts in the HNH nuclease from a thermophilic Cas9 rewires allosteric motions and enhances high-temperature DNA cleavage. *J Chem Phys*. 2022;157(22):225103.
33. Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KF, Becker S, et al. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science*. 2008;320(5882):1471–5.
34. Diehl C, Engström O, Delaine T, Håkansson M, Genheden S, Modig K, et al. Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc*. 2010;132(41):14577–89.
35. Caro JA, Harpole KW, Kasinath V, Lim J, Granja J, Valentine KG, et al. Entropy in molecular recognition by proteins. *Proc Natl Acad Sci U S A*. 2017;114(25):6563–8.

36. Wand AJ, Sharp KA. Measuring Entropy in Molecular Recognition by Proteins. *Annu Rev Biophys.* 2018;47:41–61.
37. Hobbs JK, Jiao W, Easter AD, Parker EJ, Schipper LA, Arcus VL. Change in heat capacity for enzyme catalysis determines temperature dependence of enzyme catalyzed rates. *ACS Chem Biol.* 2013;8(11):2388–93.
38. Monod J, Wyman J, Changeux JP. ON THE NATURE OF ALLOSTERIC TRANSITIONS: A PLAUSIBLE MODEL. *J Mol Biol.* 1965;12:88–118.
39. Lisi GP, East KW, Batista VS, Loria JP. Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. *Proc Natl Acad Sci U S A.* 2017;114(17):E3414–e23.
40. Schrank TP, Wrabl JO, Hilser VJ. Conformational heterogeneity within the LID domain mediates substrate binding to Escherichia coli adenylate kinase: function follows fluctuations. *Top Curr Chem.* 2013;337:95–121.
41. Olsson U, Wolf-Watz M. Overlap between folding and functional energy landscapes for adenylate kinase conformational change. *Nat Commun.* 2010;1:111.
42. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications.* 2017;8(1).
43. Rodicheva, M & Abramov, A & Gneusheva, E. (2021). Development of a Passive System Model of Human Heat Exchange with the Environment. *IOP Conference Series: Materials Science and Engineering.* 1079. 042040. 10.1088/1757-899X/1079/4/042040.

CHAPTER FIVE

Discussion

Studies of Allosteric Regulation of GeoCas9 and the Inhibitory Mechanism of AcrIIC1

The co-opting of CRISPR-Cas9 for genome editing has been one of the most revolutionary scientific advancements in recent history. It has propelled the functional study of genomes in a multitude of model organisms, leading to routine targeting of genetic modifications and large-scale genetic screens. Furthermore, it has led many groups to leverage Cas9 editing for human therapeutic strategies, advances that have already led to the first FDA approved CRISPR-based treatment of sickle cell disease⁽¹⁾.

Extensive biophysical and structural studies of *SpyCas9* have revealed sophisticated allosteric mechanisms that trigger activation upon target DNA recognition that subsequently signals the formation and careful orientation of the RNA:DNA hybrid by the Rec lobe⁽²⁻⁴⁾. Chemical signals from these spatially distinct domains propagate toward the nuclease lobe, where HNH and RuvC catalyze the cleavage of the target and non-target strand, respectively. Successful manipulation of this allosteric crosstalk has given rise to multiple high-specificity variants of *SpyCas9*, but the need to target diverse genomes and its lack of stability in human plasma makes *SpyCas9* less than ideal for future therapeutic developments. Here, the recently discovered *GeoCas9* offers a promising alternative, as it has proven stable in human plasma and functional at elevated temperatures⁽⁵⁾. We are only now beginning to understand the mechanistic determinants of long-range communication within *GeoCas9*, and how this signaling is distinct from that of the extensively characterized *SpyCas9*. The work presented in this thesis highlights the need for species specific interrogations of CRISPR-Cas9 mechanisms, as rules for specificity enhancement and interactions with broad-spectrum anti-CRISPR inhibitors exhibit unique conformational pressure on *GeoCas9*.

Summary of Findings and Future Directions

Chapter Two is a comprehensive study of the atomic-level structural and biophysical impacts of two known specificity enhancing mutations (K267E and R332A) in *SpyCas9*, which are introduced into the Rec domain of *GeoCas9*. Furthermore, this chapter provided the first dynamic and structural characterization of the native Rec domain in *GeoCas9*. Novel constructs were developed of the full-length Rec domain, along with two separate constructs of the two subdomains, *GeoRec1* and *GeoRec2*. A potential dynamic pathway within *GeoRec2* was identified spanning the length of the domain that includes 24 residues exhibiting μ s-ms timescale flexibility, similar to *SpyCas9*. This led to the question of whether allosteric tuning of this pathway in *GeoCas9* could be altered by mutations of the same amino acids that have enhanced specificity of *SpyCas9*. It has been shown that removing positively charged amino acids that interact with negatively charged nucleic acids reduces off-target effects due to the weakening of Cas9-sgRNA interaction with its DNA substrate and alteration of Cas9 protein dynamics.

HSQC spectra of the two subdomains *GeoRec1* and *GeoRec2* overlay identically with the full-length domain. In addition, circular dichroism showed a preservation of the unfolding profiles of *GeoRec1* and *GeoRec2* within the full-length *GeoRec* thermal denaturation profile, providing support for the two subdomains behaving as independent domains. Furthermore, chemical shift perturbations induced by the K267E and R332A in the individual *GeoRec2* domain aligned with shifts within the full-length domain, suggesting dynamic and biophysical properties observed with the individual domains of the mutants could be extrapolated to the full-length domain.

Longitudinal (R_1) and transverse (R_2) relaxation rates and heteronuclear ^1H - ^{15}N NOEs were measured for a Model-free analysis of per-residue order parameters (S^2) in *GeoRec2*, K267E, and R332A. Although we observed negligible bond vector functions (S^2) between *GeoRec2* and

the mutants, there was significant enhancement of μ s-ms motions as measured by CPMG experiments. This indicated an enhancement of chemical information transfer throughout the Rec domain; a similar phenomenon was observed in *SpyCas9* enhanced variants. In RNA titrations with the Rec domain we found the majority of significant chemical shifts perturbations to be concentrated in the isolated *GeoRec2* domain. Further studies on the isolated domain showed that both K267E and R332A mutants showed a reduction in intensity loss when RNA was added, indicating a weakened interaction as compared to WT *GeoRec2*. Furthermore, MST confirmed both K267E and R332A had a two-fold reduction in guide RNA affinity compared to WT. However, when incorporated into the full-length *GeoCas9* protein, neither mutation resulted in an enhancement of cleavage efficiency, even when probed at elevated temperatures. Furthermore, neither single-point mutations nor double mutants within the full-length *GeoCas9* altered the specificity, while inducing dramatic specificity enhancement within *SpyCas9*. This study highlighted the resilience of the thermophilic Cas9 to major functional changes at the level of the Rec domain. Importantly, this study provides justification for variant specific mutagenesis strategies and interrogation of intrinsic dynamics.

In **Chapter Three** the atomistic details of the interaction between the anti-CRISPR protein (Acr) AcrIIC1 and *GeoCas9* are described. This chapter was a first of its kind examination of the structural and dynamic effects of an Acr on a Cas effector. Prior to this study, AcrIIC1 was thought to simply sterically occlude the catalytic site thereby inhibiting target DNA cleavage and stalling Cas9. However, mutagenesis studies of binding residues in AcrIIC1 show that binding alone is not sufficient for inhibition, but rather an intricate rewiring of HNH domain dynamics is a crucial aspect of the Acr's inhibitory affect.

A novel 1.7 Å resolution crystal structure of the *Geo*HNH domain in complex with AcrIIC1 was solved, aligning almost identically with a previous published study of *Nme*HNH bound to AcrIIC1. This structure identified conserved binding residues S78 and C79 in AcrIIC1 that interact with the active site residues D581 and H582 of *Geo*HNH. TROSY-HSQC spectra of *Geo*HNH showed AcrIIC1 binding induced significant chemical shift perturbations throughout the entirety of the domain, with the largest shifts concentrating around the binding interface between the two proteins. Single-point mutations within AcrIIC1 (S78A, and C79P) were engineered to probe the individual contribution of each residue to the inhibition mechanism. HSQC spectra of *Geo*HNH with S78A AcrIIC1 showed remarkably similar chemical shift patterning to WT AcrIIC1. A crystal structure of the S78A mutant bound to *Geo*HNH also showed no large structural deviations from the WT bound structure. The C79P bound *Geo*HNH HSQC spectra showed no chemical shift perturbations indicating C79 is crucial for complex formation.

To further assess the interaction between *Geo*HNH and WT AcrIIC1 or variants, we measured R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE relaxation parameters that are sensitive to fast timescale conformational motions and report on global tumbling. Globally depressed R_1 and elevated R_2 of both WT and S78A bound *Geo*HNH confirmed the formation of the complex in solution, while the C79P bound profile resembled Apo *Geo*HNH. We used this data to calculate order parameters and found that regions in close sequence and spatial proximity to the *Geo*HNH active site had sharp increases in S^2 in the presence of AcrIIC1, suggesting the suppression of ps-ns timescale flexibility in these regions. To complement these findings, we ran CPMG experiments to probe motions on the us-ms timescale, as we knew these to be absent in the Apo *Geo*HNH domain. Interestingly, we saw the rewiring towards these “slow” timescale motions in the presence of WT AcrIIC1 in the same region we observed the suppression of “fast” motion. Despite inducing

almost identical chemical shift perturbations as WT AcrIIC1 on the *Geo*HNH domain, S78A did not promote dynamic rewiring on the scale of WT AcrIIC1.

After observing disparate structural and dynamic influences of the Acr variants on *Geo*HNH, we next assessed their inhibition of full-length *Geo*Cas9. Interestingly, only WT AcrIIC1 inhibited DNA cleavage activity of *Geo*Cas9, indicating that binding alone is not sufficient for inhibition, but that a substantial dynamic rewiring of the domain is required. Surprisingly, we also observed that WT AcrIIC1 binding greatly reduces the affinity of *Geo*Cas9 for its guide RNA, a trend not observed for the mesophilic *Nme*Cas9.

Previous studies have shown that for many Acr variants, translation of *acr* gene is dramatically increased upon secondary infection⁽⁶⁾. Although exact intercellular concentrations of Acr proteins within the cell are unknown at any time, it has been suggested that part of the phage inhibition mechanism toward Cas9 is to flood the cell with anti-CRISPRs. Within the literature and our own studies it is shown that efficient inhibition is dependent on high concentrations of Acr, with some papers reporting up to tenfold increase in Acr:Cas9 ratios to achieve complete inhibition in functional assays⁽⁷⁾. Interestingly, NMR studies have shown unique chemical shift behavior at high concentrations of Acr with the *Geo*HNH domain, including the broadening of all three histidine side chain resonances of *Geo*HNH, despite only two of those histidines being near the crystallographic binding interface. Taken together, this evidence suggests the possibility for a secondary binding location within the domain. It is possible that multiple Acrs binding to a single Cas9 represents an important and understudied aspect of Acr inhibition. Future studies should focus binding studies of novel domain constructs of *Geo*Cas9 to see if Acr binding is exclusive to the HNH domain.

Chapter Four provides a comprehensive review of future considerations for the study of thermophilic Cas enzymes. This chapter emphasizes recent advances in the comparative studies of mesophilic and thermophilic enzymes where it has been found that enhancement or diminution of allosteric crosstalk can be temperature dependent and influence any enzyme differently. However, the impact of temperature on allostery is largely understudied. Therefore, this chapter outlines the importance of temperature in differentiating characteristics of allosteric enzymes by highlighting case studies with temperature dependent structure and dynamics, as well as NMR methods suited for these analyses.

NMR relaxation dispersion and temperature dependent chemical shift measurements are discussed as effective strategies for quantifying exchange kinetics, conformational populations, and entropic/enthalpic contributions to differing regimes of k_{ex} vs $\Delta\omega$. This allows for the linkage of temperature modulated motions and hydrogen bonding changes to enzymatic activity, particularly in thermophiles.

The case studies within this chapter illustrate how modulating temperature can either induce or modulate allostery. In the thermophilic IGPS, elevating the temperature was sufficient to recapitulate the PRFAR-induced allosteric network and push the apo enzyme into a “pre-active” state. In the case study of adenylate kinase, the authors showed how thermal adaptations can operate through allosteric pathways, rather than changes at an enzymes active site. Single-point mutations in the LID domain of adenylate kinase increase the population of an unfolded intermediate without disrupting the ground-state structure. This stabilization of the minor state establishes how small mutations can drive subtle shifts in the conformational equilibria, thereby entropically tuning allosteric parameters.

Contributions of this dissertation

This doctoral dissertation has provided several new contributions to the field of structural biology. Broadly, it has highlighted the importance of establishing allosteric models on a per-species basis within Cas editors for precise and effective alteration of residue pathways with the goal of specificity enhancement. This work is an innovative exploration of specificity-enhancing mutations in a mesophilic Cas system and the ability to translate those to a thermophilic system, while also uncovering novel dynamic pathways within the *GeoCas9*. Additionally, this thesis described the first atomic-level interrogation of the interaction and subsequent dynamic and structural effects of Anti-CRISPR binding to *GeoCas9*. I showed that the Anti-CRISPR AcrIIC1 inhibited *GeoCas9* cleavage activity not just by simply blocking the catalytic site, but by altering the intrinsic dynamics of the *GeoHNH* domain through the diminishment of fast timescale motions. Future studies of this system should delve into the thermophilic qualities of *GeoCas9* by interrogating not only the intrinsic dynamics at elevated temperatures but also how specificity enhancing mutations and interactions with anti-CRISPR inhibitors are modulated at elevated temperatures. Furthermore, an exploration into structure and dynamics of additional thermophilic homologs could give insights into whether allosteric signaling within thermophiles are governed by categorical rules.

References

1. Singh A, Irfan H, Fatima E, Nazir Z, Verma A, Akilimali A. Revolutionary breakthrough: FDA approves CASGEVY, the first CRISPR/Cas9 gene therapy for sickle cell disease. *Annals of Medicine & Surgery*. 2024;86(8):4555–9.
2. Wang J, Arantes PR, Ahsan M, Sinha S, Kyro GW, Maschietto F, et al. Twisting and swiveling domain motions in Cas9 to recognize target DNA duplexes, make double-strand breaks, and release cleaved duplexes. *Frontiers in Molecular Biosciences*. 2023;9.
3. Palermo G, Chen JS, Ricci CG, Rivalta I, Jinek M, Batista VS, et al. Key role of the REC lobe during CRISPR–Cas9 activation by ‘sensing’, ‘regulating’, and ‘locking’ the catalytic HNH domain. *Quarterly Reviews of Biophysics*. 2018;51.
4. Nierzwicki L, East KW, Morzan UN, Arantes PR, Batista VS, Lisi GP, et al. Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *eLife*. 2021;10.
5. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, et al. A thermostable Cas9 with increased lifetime in human plasma. *Nature Communications*. 2017;8(1).
6. Stanley SY, Borges AL, Chen K-H, Swaney DL, Krogan NJ, Bondy-Denomy J, et al. Anti-CRISPR-Associated Proteins Are Crucial Repressors of Anti-CRISPR Transcription. *Cell*. 2019;178(6):1452–64.e13.
7. Harrington LB, Doxzen KW, Ma E, Liu J-J, Knott GJ, Edraki A, et al. A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell*. 2017;170(6):1224–33.e15.

CHAPTER SIX

Contributed work:

NMR Studies of the Effect of Metal Ions on *SpyCas9* HNH-L2 Domain

This work is adapted from a manuscript submitted to Nat Comm., 2025 titled:

A Cryptic Binding Pocket Regulates the Metal-Dependent Activity of Cas9

Mohd Ahsan, Aakash Saha, Delisa Ramos, Isabel Strohkendl, Alexa L. Knight, Erin Skeens, George P. Lisi, David W. Taylor, Giulia Palermo

¹*Department of Bioengineering, Riverside, CA 92521, USA*

²*Department of Molecular Biosciences, Austin, TX 78712, USA*³

³*Department of Molecular Biology, Cell Biology & Biochemistry, Brown University, Providence, RI 02903, USA*

Author Contributions:

MA performed molecular simulations, analyzed the data, and wrote the manuscript. **AS** analyzed the data. **DS** performed and analyzed biochemical experiments, aided by **IS**. **AK** carried out NMR sample preparation and experiments, MST experiments, DSF experiments, analyzed the data, and contributed to manuscript writing. **ES** carried out NMR sample preparation and experiments, analyzed the data, and contributed to manuscript writing. **GPL** supervised the NMR experiments. **DWT** supervised the biochemical experiments. **GP** supervised the computational studies, conceived the research, and contributed to manuscript writing. All authors critically reviewed and edited the manuscript.

Abstract

Cas9 is a metal-dependent nuclease that has revolutionized gene editing across diverse cells and organisms exhibiting varying ion uptake, metabolism, and concentrations. However, how divalent metals impact its catalytic function, and consequently its editing efficiency in different cells, remains unclear. Here, extensive molecular simulations, Markov State Models, biochemical and NMR experiments, demonstrate that divalent metals – Mg^{2+} , Ca^{2+} , and Co^{2+} – promote activation of the catalytic HNH domain by binding within a dynamically forming divalent metal binding pocket (DBP) at the HNH-RuvC interface. Mutations in DBP residues disrupt HNH activation and impair the coupled catalytic activity of both nucleases, identifying this cryptic DBP as a key regulator of Cas9's metal-dependent activity. The ionic strength thereby promotes Cas9's conformational activation, while its catalytic activity is metal-specific. These findings are critical to improving the metal-dependent function of Cas9 and its use for genome editing in different cells and organisms.

Introduction

CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats and associated protein Cas9) has revolutionized life sciences by introducing a facile genome editing technology^(1,2). At its core, Cas9 is a metal-dependent nuclease enabling precise genetic modifications across a wide range of cells and organisms with diverse ion uptake, metabolism, and intracellular metal concentrations⁽³⁾. A key distinction between bacterial and mammalian cells is the availability of free divalent ions, which can influence protein-nucleic acid interactions and the specificity of Cas effectors⁽⁴⁻⁷⁾. However, the extent to which divalent metals regulate the catalytic function of Cas9, and consequently its genome editing capability in different cellular

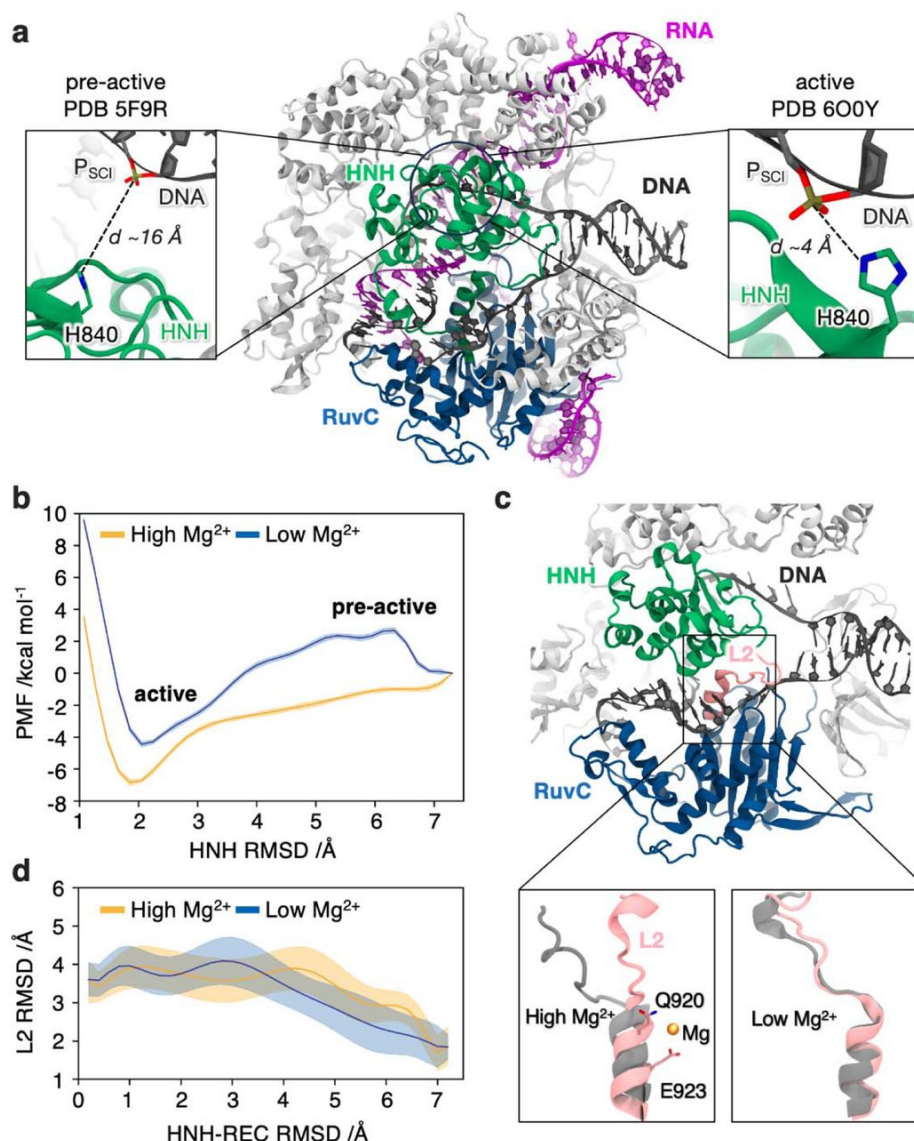


Figure 1. Cas9 activation under varying concentrations of Mg²⁺ ions. (A) Structure of *S. pyogenes* Cas9 bound to RNA (magenta) and DNA (black), highlighting the HNH (green) and RuvC (blue) catalytic domains⁽¹⁴⁾. Close-up views show the HNH nuclease in its pre-active state (PDB 5F9R⁽¹³⁾, left), where the catalytic H840 is $\sim 16 \text{ \AA}$ away from the scissile phosphate (P_{SCI}) on the DNA target strand, and in its active conformation (PDB 6O0Y⁽¹⁴⁾, right), where HNH is docked at the cleavage site. (B) Free energy profiles depicting the conformational change of HNH from its pre-active to active state at high (10 mM, yellow) and low (10 μ M, blue) concentration of Mg²⁺. The Potential of Mean Force (PMF, in kcal/mol) was computed along the root-mean-square deviation (RMSD) of C α atoms in the HNH domain and portions of the recognition lobe (REC) relative to both states. Error estimation was performed using Monte Carlo bootstrap analysis of free energy simulations. (C) Rearrangements of the L2 region (residues 907–925) in response to Mg²⁺ concentration. At high Mg²⁺ levels (left), residues Q920 and E923 form interactions with the ion, which are lost under low Mg²⁺ conditions. (D) RMSD of the L2 region along the free energy profiles reveals that as HNH moves toward the active state, conformational changes in L2 are more pronounced at high Mg²⁺ concentration. Errors are reported as the standard deviation for each free energy simulation window.

environments, remains unclear. Understanding the metal-dependent function of Cas9 is therefore important to fully exploit its function across different biological applications, aiding more precise and versatile genome editing technologies.

The *S. pyogenes* Cas9 (*Spy*Cas9) is an RNA-guided nuclease that utilizes the sequence specificity of its guide RNA to target and cleave complementary DNA sequences in the presence of metal ions⁽⁸⁾. Double-stranded DNA cleavage is carried out by two catalytic domains, HNH and RuvC (**Figure 1. A**), which cleave the DNA target strand (TS, complementary to the guide RNA) and the non-target strand (NTS), respectively^(3,9). The HNH domain exhibits considerable conformational flexibility and structural rearrangements, in contrast to the RuvC domain, which maintains a stable structure⁽¹⁰⁾. Upon DNA binding, the HNH nuclease transitions from an inactive state, in which its catalytic triad (comprising H840, N863, D839) is oriented away from the cleavage site on the DNA TS^(11,12), to a pre-active state, where H840 shifts closer to the scissile phosphate but remains ~16 Å away (**Figure 1. A, left**)⁽¹³⁾. While this transition can occur in the absence of metal ions, the final shift from the pre-active to the active state, where H840 docks at the cleavage site (**Figure 1. A, right**)^(14–16), is dependent on Mg²⁺ ions^(17–19). Single-molecule studies have shown that Mg²⁺ concentrations between 5 and 10 mM (i.e., high concentrations) are required for HNH to transition into its active state⁽¹⁸⁾. At low Mg²⁺ concentrations (10 µM), this conformational shift is hindered, preventing HNH activation for DNA cleavage. These studies also revealed that other divalent cations, such as Ca²⁺ and Co²⁺, can promote this conformational activation when present at high concentrations. These findings suggest that high concentrations of divalent cations may promote the docking of HNH at the DNA target strand for cleavage. However, the impact of ionic strength on this critical conformational change, which is essential for Cas9 function, remains unclear. Additionally, how

different divalent metal ions – such as Mg^{2+} , Ca^{2+} , and Co^{2+} , as well as mono- and trivalent ions such as Li^+ and Al^{3+} alter the chemical mechanism of DNA cleavage, either facilitating or inhibiting catalysis⁽³⁾, remains an open question. Addressing these questions is key to understanding how Cas9 function is altered in cells with varying ion uptake, metabolism, and intracellular concentrations, as well as for optimizing Cas9-mediated genome editing across diverse biological contexts, including plants and mammalian cells.

In this study we investigated the effect of mono-, di-, and trivalent metal ions on the structural and dynamic profile of the HNH-L2 domain using MD simulation (not included here) and nuclear magnetic resonance (NMR) spectroscopy. MD simulations reveal the existence of a cryptic DBP, formed through interactions with divalent metal ions, is responsible for promoting HNH activation within *SpyCas9*. NMR titrations with various mono-, di-, and trivalent metal ions were performed to validate binding, hydrogen bond propensity, and dynamics of the HNH-L2 domain. These results show the disparate effects of various metal ions on the structure, dynamics, and function of the HNH-L2 domain.

Results and Discussion

In order to probe the effect of Mg^{2+} binding on an isolated HNH-L2 construct, solution NMR experiments were performed. In the presence of high Mg^{2+} , a distinct structural state was observed (**Figure 2.**) consistent with computational findings (presented in full version of the manuscript). All-atom simulations of HNH found that in the presence of high Mg^{2+} the HNH domain exhibited a more pronounced movement towards the scissile phosphate on the DNA target strand. In the presence of high Mg^{2+} , a unique structural state was persistent at both 298 K and 310 K by NMR. To assess the dynamic profile of the Mg^{2+} bound state we measured R_1 , R_2 ,

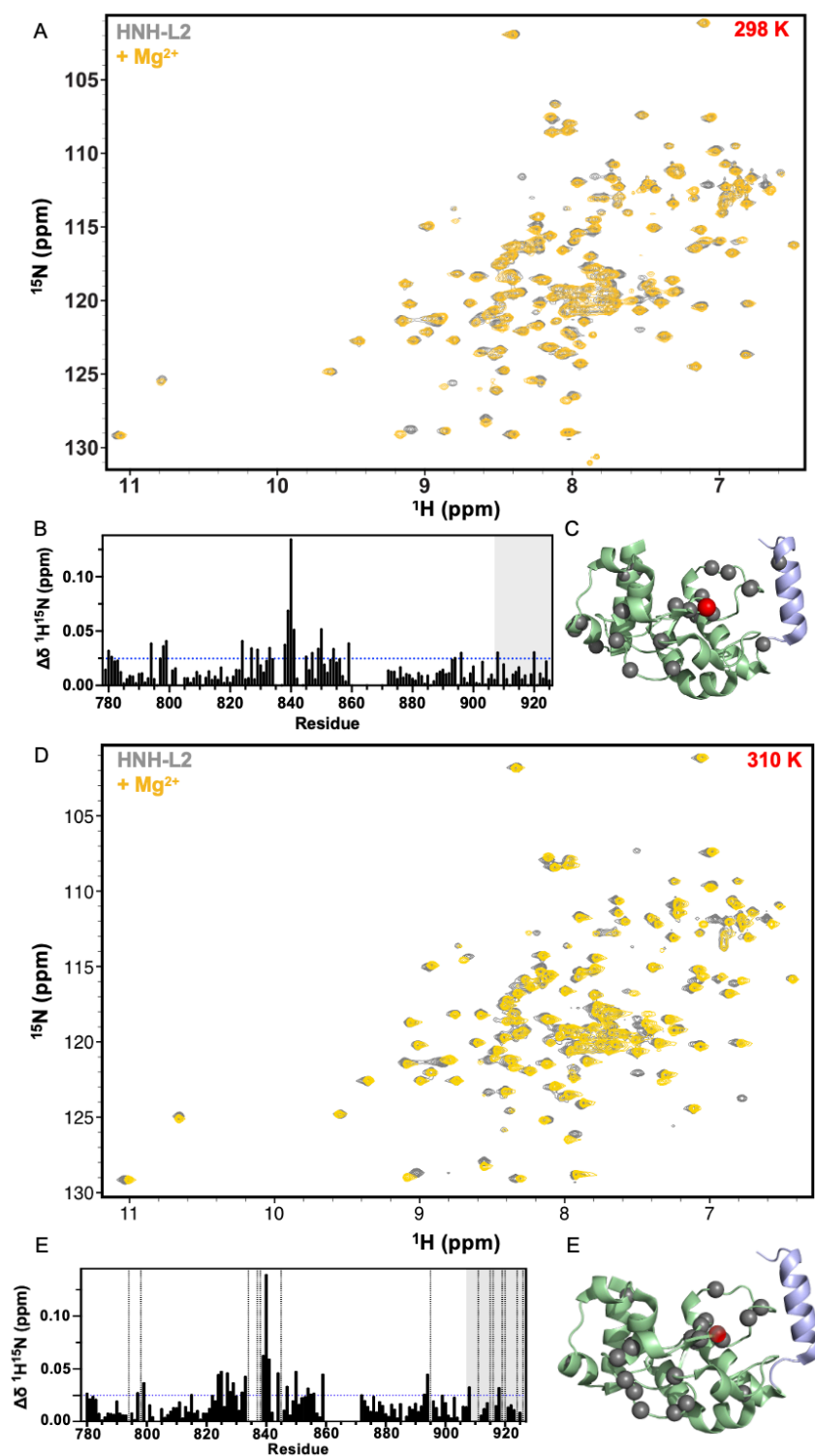


Figure 2. NMR HSQC titrations of HNH-L2 with Mg^{2+} at 298 K and 310 K. HSQC spectra of apo HNH-L2 (grey) overlaid with 140x Mg^{2+} (yellow) at 298 K (A) and 310 K (B). Chemical shift perturbation (CSP) plots at 298 K (B) and 310 K (E) show little difference. CSPs $>1.5\sigma$ of the 10% trimmed mean of all shifts (horizontal dashed line) are plotted as spheres on the HNH-L2 structure at 298 K (C) and 310 K (E).

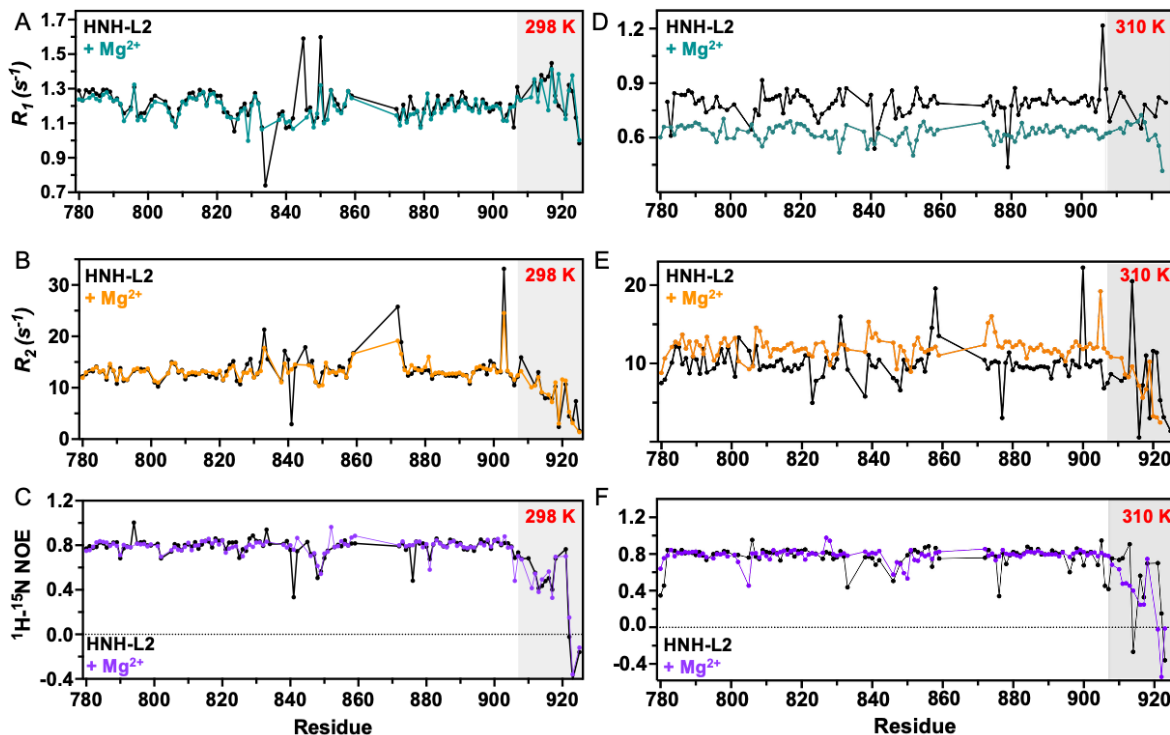


Figure 3. Dynamic Impacts of Mg^{2+} Binding to HNH-L2 by NMR at 298 K (A-C) and 310 K (D-F). Fast timescale dynamics (ps – ns) are plotted for longitudinal (R_1) and transverse (R_2) relaxation rates and ^1H - ^{15}N NOE values for apo HNH-L2 (black) and Mg^{2+} bound HNH-L2 (teal, orange, and purple for R_1 , R_2 , and NOE, respectively) at 298 K and 310 K.

and ^1H - ^{15}N heteronuclear NOE relaxation parameters that are sensitive to fast timescale conformational motions and report on global tumbling at 298 K and 310 K. At 298 K R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE values were largely consistent between apo and Mg^{2+} bound HNH-L2 (**Figure 3. A-C**). At 310 K we observed globally suppressed R_1 values compared to R_1 values at 298 K, a phenomenon consistent with a sharp increase in temperature. Upon Mg^{2+} binding, R_1 values were further globally suppressed indicative of faster overall tumbling due to possible reorganization of solvation layer or reduced repulsion of charged residues. This observation was consistent across R_1 , R_2 , and ^1H - ^{15}N heteronuclear NOE values of HNH at 298 K and 310 K, apart from the L2 region which shows evidence of increased flexibility at 310 K (**Figure 3. D-F**).

Next, we wanted to assess whether and how Mg^{2+} was reshaping the hydrogen-bonding network and stability of the HNH-L2 domain. To do so, we performed NMR temperature coefficient experiments on apo and Mg^{2+} bound HNH-L2 (**Figure 4**). ^1H - ^{15}N HSQC experiments at 288 K, 293 K, 298 K, 303 K, and 308 K for both apo and Mg^{2+} bound HNH-L2 revealed a linear shifting pattern consistent with a gradual increase in temperature (**Figure 4. A, B**). We plotted $\Delta\delta^1\text{H}/\Delta T$ and $\Delta\delta^{15}\text{N}/\Delta T$ per residue and correlation plots (**Figure 4. C-F**) and found Mg^{2+} binding did induce profound rearrangements of the hydrogen bonding network. A few residues within the helical core of the protein (res. ~840-860) showed slightly altered hydrogen-bond propensity indicating a minor reshuffling of the network (**Figure 4. G**).

Although extensive biochemical and theoretical studies of metal-dependent phosphodiester hydrolysis demonstrates that catalysis is reliant on divalent metal ion, we wondered whether the metal-dependent conformational activation could be uncoupled from catalysis. As such, we tested the effect of the mono- and trivalent metal ions Li^+ and Al^{3+} on the isolated HNH-L2 structure in conjecture with MD simulations (not included here). ^1H - ^{15}N HSQC

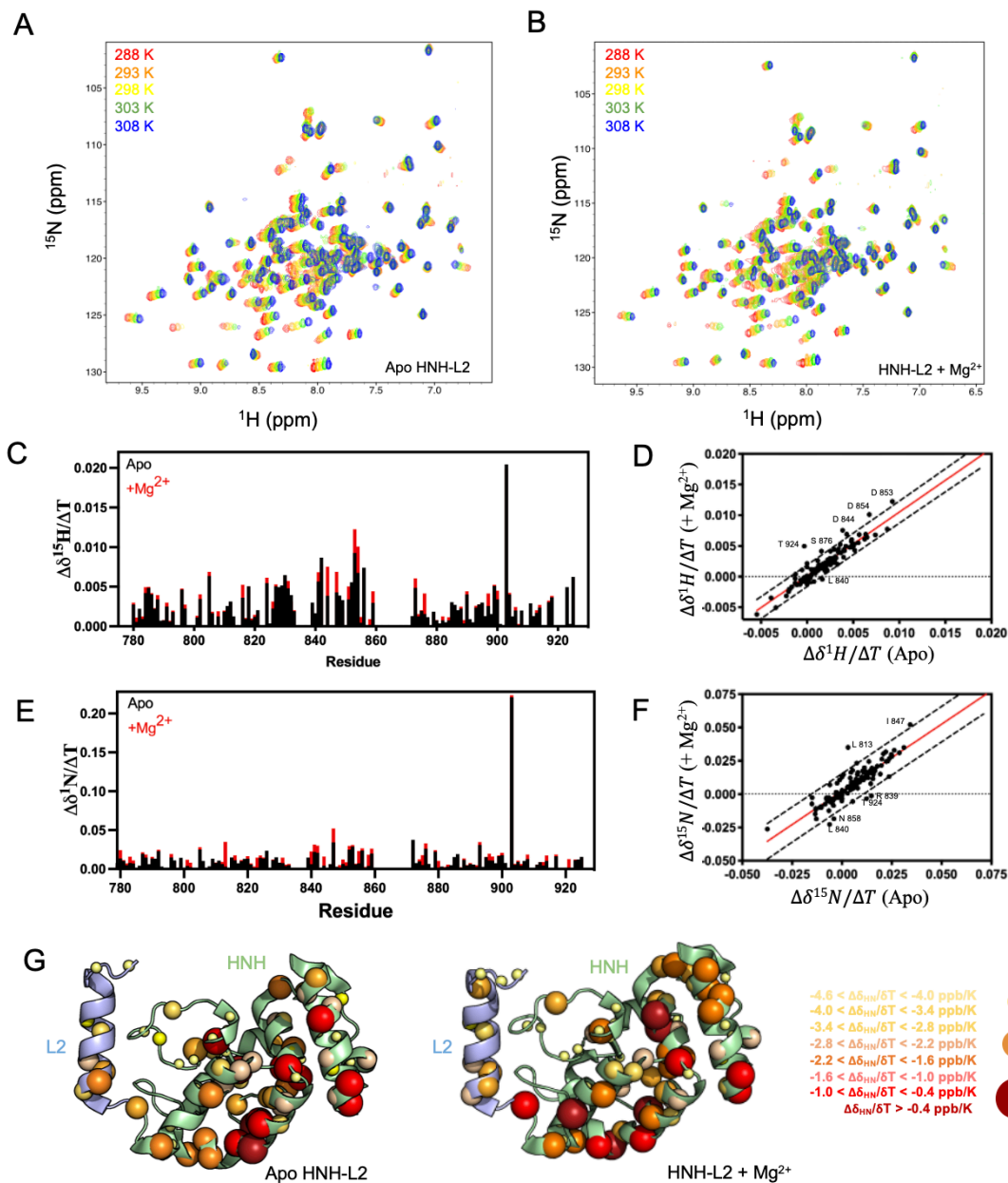


Figure 4. Mg^{2+} binding to HNH-L2 shows negligible effect on hydrogen bonding network. ^1H - ^{15}N HSQC spectra show apo (A) and Mg^{2+} bound HNH-L2 at 288 K, 293 K, 298 K, 303 K, and 308 K. $\Delta\delta^1\text{H}/\Delta T$ (C) and $\Delta\delta^1\text{N}/\Delta T$ (E) values are shown per residue in the apo (black) and Mg^{2+} bound (red). (G) Temperature coefficient ($\Delta\delta/\Delta T$) values are plotted as spheres on the HNH-L2 structure.

titrations of Li^+ into HNH-L2 reveal no significant chemical shift perturbations (CSPs), even at high Li^+ concentrations (**Figure 5. A-C**). Due to the monovalency and weaker charge density, it is likely Li^+ cannot coordinate a multi-residue binding interaction within HNH-L2. ^1H - ^{15}N HSQC titrations of Al^{3+} into HNH-L2 show significant globally distributed CSPs at lower concentrations than the divalent ions tested, although the highest CSPs cluster to the domain core (**Figure 5. D-F**). As Al^{3+} is a weakly paramagnetic metal ion, we were unable to titrate to as high of concentrations of ions as done so with the previous divalent metals in NMR. Although Al^{3+} induces many significant CSPs it is difficult to determine by NMR whether this is a result of true conformational coordination of metal ions or an effect of promiscuous binding to the HNH-L2 domain.

In order to assess the functional effect of these mono-, di-, and trivalent metal ions on the full-length *SpyCas9*, we performed *in vitro* DNA cleavage assays substituting various metals in the reaction buffer at increasing concentrations (**Figure 6**). We tested Li^+ , Mg^{2+} , and Al^{3+} ions at 5 mM, 10 mM, and 15 mM consistent with, and slightly above, published concentrations used in standard *SpyCas9 in vitro* DNA cleavage assays. Cleavage buffer containing Mg^{2+} induced successful and efficient DNA cleavage across all three concentrations (**Figure 6. B**). Surprisingly, cleavage buffer containing Li^+ was able to induce cleavage of DNA substrates, albeit at a reduced effectiveness in a concentration independent manner. Despite the significant impact Al^{3+} induced on the NMR spectra of HNH-L2, no concentration of Al^{3+} was sufficient to induce cleavage of the DNA substrate. In fact, the Al^{3+} seems to completely digest the DNA substrate entirely, as none was observable in the agarose gel (**Figure 6. A**), a phenomenon previously observed in the literature⁽²⁹⁾.

Finally, to assess functional relevance of the DBP, three *SpyCas9* variants were engineered with mutations (M1, M2, and M3) in residues lining the DBP. MD simulations and *in vitro* biochemical experiments were performed in order to assess the catalytic influence of these mutations (not included here). To confirm any effect of these mutations was not due to a destabilization of the protein, we performed differential scanning fluorimetry (DSF) melts to assess the thermostability of the WT *SpyCas9* and mutant protein (**Figure 7**). We assessed the thermostability of these proteins under apo, high Mg^{2+} , and high Mg^{2+} and gRNA conditions. These experiments concluded that the intrinsic stability of each variant was within ± 2 °C of each other. Furthermore, the presence of high Mg^{2+} made no difference to the thermostability, while gRNA binding increased the thermostability of all *SpyCas9* variants. As such, any variability in catalytic activity of the variants was due to the effects of the mutations within the DBP, and not due to instability of the protein itself.

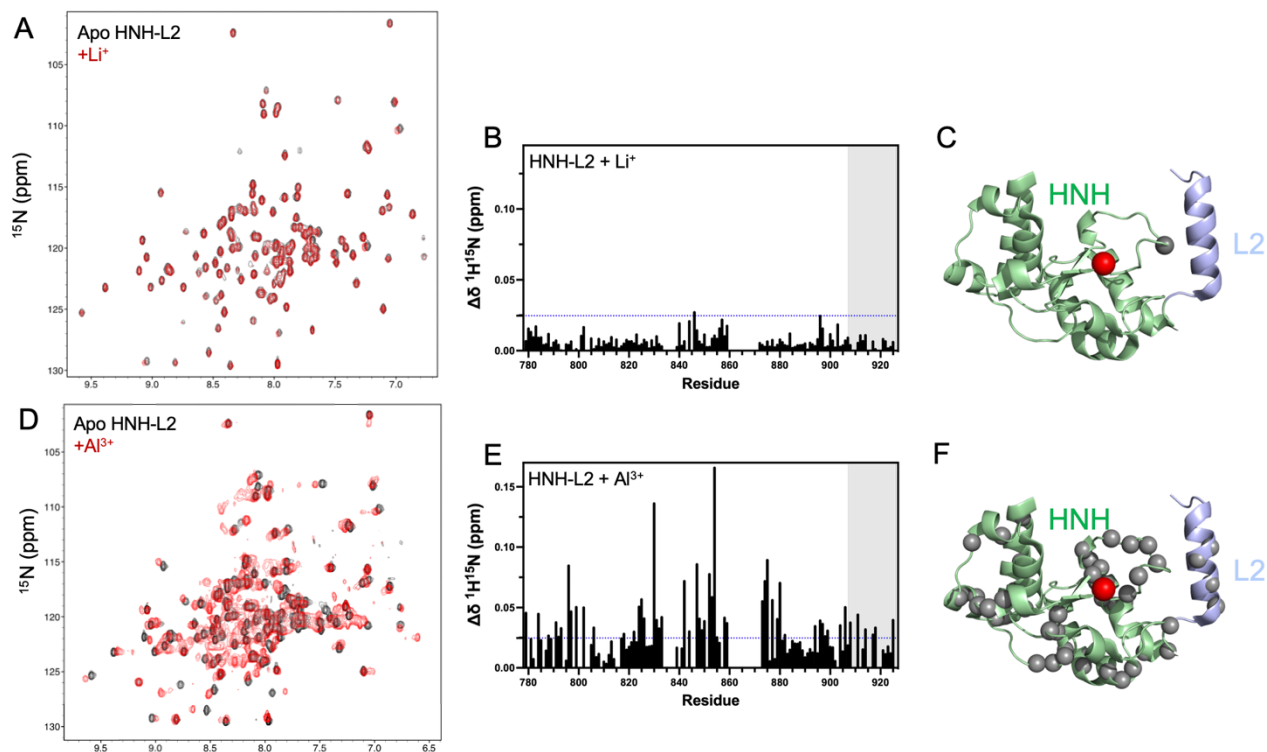


Figure 5. Monovalent and trivalent metal ions induce disparate NMR chemical shift behavior when bound to HNH-L2. (A-C) ^1H - ^{15}N HSQC titrations with Li^+ very few chemical shift perturbations. (D-F) ^1H - ^{15}N HSQC titrations with Al^{3+} induce significant chemical shift perturbations throughout the entire domain architecture.

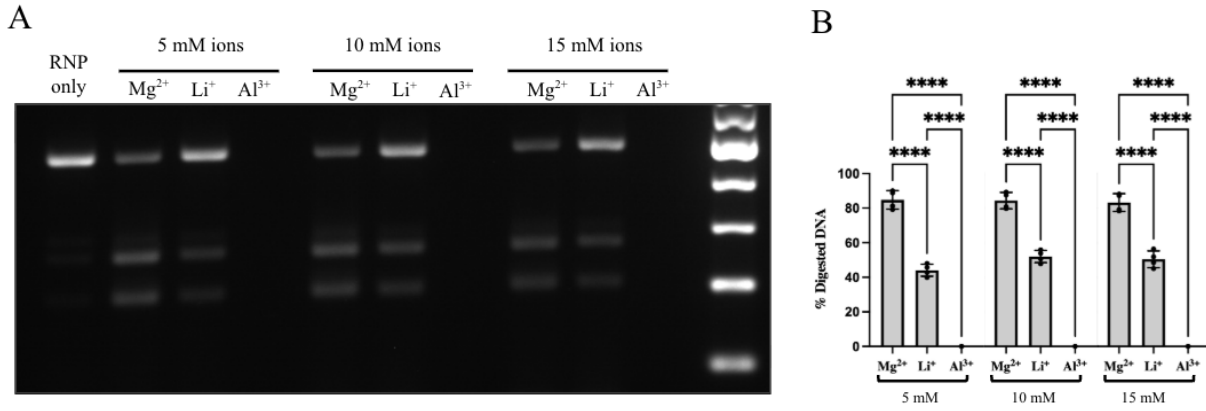


Figure 6. *in vitro* DNA cleavage assays show incongruent effect of varying concentrations of Mg²⁺, Li⁺, and Al³⁺ on *Spy*Cas9 nuclease activity. (A) Representative gel image showing the cleavage of DNA substrate in presence of varying concentration of metal ions on an agarose gel. (B) Quantification of total digested DNA. *****p*<0.0001. *n*=3 technical replicates

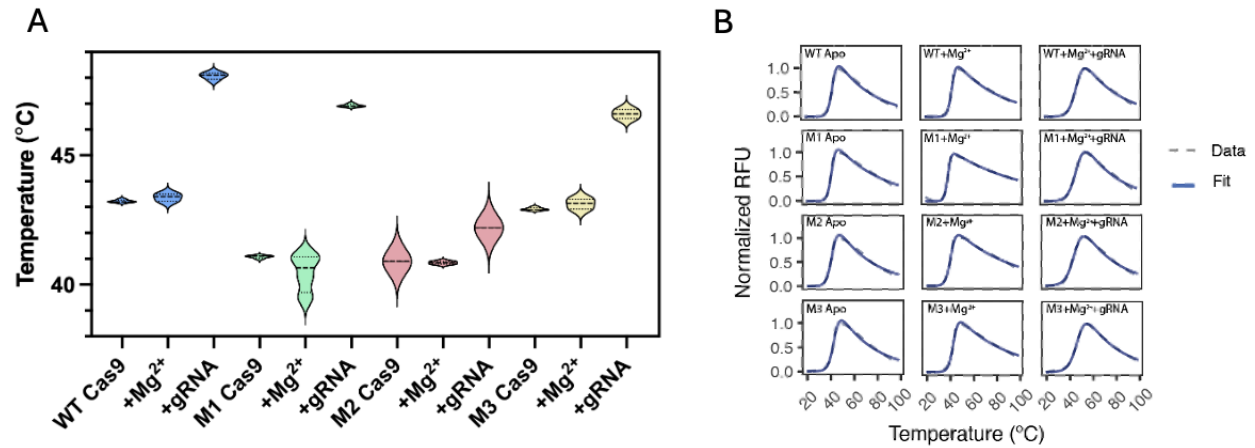


Figure 7. DSF-derived melting temperatures of Cas9 variants. (A) Violin plots show the distribution of melting temperatures (T_m) measured by differential scanning fluorimetry for WT Cas9 and variants M1, M2, and M3 in the absence or presence of Mg^{2+} and guide RNA (gRNA). Dotted horizontal lines indicate the median and quartiles for each condition. The data show condition dependent shifts in thermostability across the Cas9 variants. (B) Representative raw DSF traces (dashed grey lines) with sigmoidal best fit overlaid (blue line).

Methods

NMR Spectroscopy

The HNH-L2 domain (residues 772–925) of *S. pyogenes* Cas9 was cloned into pET-28a(+) with an N-terminal His₆-tag and TEV cleavage site. Protein was expressed in M9 minimal medium supplemented with ¹⁵NH₄Cl and ¹³C-glucose, induced with 1 mM IPTG, and purified using Ni-NTA affinity and size-exclusion chromatography. Post-purification, protein was dialyzed into buffers containing KCl, MgCl₂, or CaCl₂ depending on the sample condition. NMR experiments were conducted at 600 MHz and 25°C using a Bruker Avance NEO spectrometer unless indicated otherwise. Backbone assignments were obtained using standard triple-resonance experiments⁽²¹⁾ and validated against BMRB entry 27949⁽²²⁾. Chemical shift perturbations were computed from ¹H–¹⁵N HSQC spectra⁽²³⁾ and parameters from T_1 , T_2 , ¹H–[¹⁵N]-NOE experiments were measured using established protocols⁽²⁴⁾ with uncertainties were derived from replicate spectra^(25–27). Two-dimensional ¹H–¹⁵N HSQC spectra were acquired on a 850 MHz Bruker spectrometer equipped with a cryogenic probe over a temperature range of 288–308 K, in 5 K increments. Samples were equilibrated at each temperature for at least 10–15 minutes prior to data acquisition. Amide proton chemical shifts were tracked across temperatures for assigned residues. Temperature coefficients ($\Delta\delta/\Delta T$) were obtained by linear regression of the amide proton chemical shift as a function of temperature.

Differential scanning fluorimetry (DSF)

Differential scanning fluorimetry (DSF) was used to assess the thermal stability of four Cas9 variants under three conditions: apo, with Mg²⁺, and with guide RNA (gRNA). Each condition was measured in four independent replicates ($n = 4$). Proteins were assayed in 20 mM HEPES,

200 mM KCl, and 1 mM DTT at pH 7.5 with SYPRO Orange at 10× final concentration. Samples were heated from 25 to 95 °C, and fluorescence was monitored throughout the melt. DSF data were analyzed using publicly available software on the DSFworld website. Melting temperatures were obtained by fitting fluorescence traces to sigmoidal curves. Mean T_m values and standard deviations from the four replicates were used for plotting.

In Vitro Cleavage Assays

RNP formation

S. pyogenes Cas9 ribonucleoprotein (RNP) complexes were formed by mixing Cas9 protein (12 μM stock) with annealed crRNA (55 μM, IDT 26002-g1_crRNA) and tracrRNA (89.4 μM, IDT Alt-R CRISPR-Cas9 tracrRNA) at 1:1.2:1 molar ratio (Cas9:crRNA:tracrRNA) in annealing buffer (10 mM Tris-HCl (pH 7.5), 100 mM KCl). The crRNA and tracrRNA were first annealed by mixing equimolar amounts (10 μM final), heating to 95°C for 5 min, and cooling to room temperature over 30 min. The RNP complex (1 μM final) was incubated at room temperature for 10 min prior to use.

Cleavage reactions

Double-stranded DNA substrates (483 bp linear amplicon containing PAM and guide RNA spacer sequence, 50 ng/μL = 16.8 nM) were generated by PCR using DreamTaq Green PCR Master Mix with mouse genomic DNA template (60°C annealing, 45 s extension, 35 cycles). DNA substrates were incubated with RNP (500 nM final) in cleavage buffer (20 mM Tris-HCl pH (7.5), 100 mM KCl, 5% glycerol v/v, 1 mM DTT) supplemented with divalent cations (Mg²⁺, Li⁺, Co²⁺, or Al³⁺ at 5, 10, or 15 mM final). Reactions (10 μL total volume) were assembled on

ice, incubated at 37°C for 30 min, and terminated by heating to 65°C for 10 min. Products (6 μ L + 2 μ L 6X DNA loading dye) were resolved on 1.5% agarose gels stained with GelRed nucleic acid dye and visualized by UV transillumination. No-RNP controls were included for each condition. All reactions were performed in triplicate. Cleavage buffer components were prepared as 4X base buffer (80 mM Tris-HCl pH 7.5, 400 mM KCl, 20% glycerol), 8X divalent cation stocks, and 8X DTT (freshly prepared from 1 M stock).

References

1. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. *Cell* **187**, 1076–1100 (2024).
2. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).
3. Jinek, M. et al. Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1247997–1247997 (2014).
4. Eggers, A. R. et al. Rapid DNA unwinding accelerates genome editing by engineered CRISPR-Cas9. *Cell* **187**, 3249–3261 (2024).
5. Nguyen, G. T. et al. CRISPR-Cas12a exhibits metal-dependent specificity switching. *Nucleic Acids Res* **52**, 9343–9359 (2024).
6. Sundaresan, R., Parameshwaran, H. P., Yogesha, S. D., Keilbarth, M. W. & Rajan, R. RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. *Cell Rep* **21**, 3728–3739 (2017).
7. Newsom, S. N. et al. Differential Divalent Metal Binding by SpyCas9's RuvC Active Site Contributes to Nonspecific DNA Cleavage. *CRISPR J* **6**, 527–542 (2023).
8. Jinek, M. et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).
9. Ahsan, M., Pindi, C. & Palermo, G. Emerging Mechanisms of Metal-Catalyzed RNA and DNA Modifications. *Annu Rev Phys Chem* **76**, 497–518 (2025).

10. Palermo, G., Miao, Y., Walker, R. C., Jinek, M. & McCammon, J. A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proceedings of the National Academy of Sciences* **114**, 7260–7265 (2017).
11. Nishimasu, H. et al. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **156**, 935–949 (2014).
12. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
13. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
14. Zhu, X. et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat Struct Mol Biol* **26**, 679–685 (2019).
15. Bravo, J. P. K. et al. Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **603**, 343–347 (2022).
16. Pacesa, M. et al. R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191–196 (2022).
17. Raper, A. T., Stephenson, A. A. & Suo, Z. Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. *J Am Chem Soc* **140**, 2971–2984 (2018).
18. Dagdas, Y. S., Chen, J. S., Sternberg, S. H. & Doudna, J. A. A Conformational Checkpoint between DNA Binding and Cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao002 (2017).
19. Wang, Y. et al. Real-time observation of cas9 post-catalytic domain motions. *Proc Natl Acad Sci U S A* **118**, e2010650118 (2021).

20. Nierzwicki, Ł. et al. Principles of target DNA cleavage and the role of Mg^{2+} in the catalysis of CRISPR–Cas9. *Nat Catal* **5**, 912–922 (2022).
21. Grzesiek, S. & Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J Magn Res* **99**, 201–207 (1992).
22. Belato, H. B., East, K. W. & Lisi, G. P. 1H , ^{13}C , ^{15}N backbone and side chain resonance assignment of the HNH nuclease from *Streptococcus pyogenes* CRISPR-Cas9. *Biomol NMR Assign* **13**, 367–370 (2019).
23. Grzesiek, S., Stahl, S. J., Wingfield, P. T. & Bax, A. The CD4 Determinant for Downregulation by HIV-1 Nef Directly Binds to Nef. Mapping of the Nef Binding Surface by NMR. *Biochemistry*
24. Kay, L. E., Torchia, D. A. & Bax, A. Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*
25. East, K. W. et al. Allosteric Motions of the CRISPR-Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J Am Chem Soc* **142**, 1348–1358 (2020).
26. Nierzwicki, Ł. et al. Enhanced specificity mutations perturb allosteric signaling in CRISPR-Cas9. *Elife* **10**, e73601 (2021).
27. Skeens, E. et al. High-Fidelity, Hyper-Accurate, and Evolved Mutants Rewire Atomic Level Communication in CRISPR-Cas9. *Sci. Adv.* **8**, ead11045 (2024).
28. Markley, J. L. Observation of histidine residues in proteins by nuclear magnetic resonance spectroscopy. *Acc Chem Res* **8**, 70–80 (1975).

29. Poot-Poot W, Rodas-Junco BA, Muñoz-Sánchez JA, Hernández-Sotomayor SMT. Protoplasts: a friendly tool to study aluminum toxicity and coffee cell viability. SpringerPlus. 2016;5(1).