

UNDERSTANDING SPECIFICITY OF SMALL MOLECULE INHIBITORS OF REGULATORS OF
G-PROTEIN SIGNALING (RGS) PROTEINS

By

Vincent Shaw

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Pharmacology and Toxicology—Doctor of Philosophy

2019

ABSTRACT

UNDERSTANDING SPECIFICITY OF SMALL MOLECULE INHIBITORS OF REGULATORS OF G-PROTEIN SIGNALING (RGS) PROTEINS

By

Vincent Shaw

Regulators of G-protein Signaling (RGS) proteins terminate G-Protein Coupled Receptor (GPCR) signaling by binding to active G α subunits and accelerating hydrolysis of GTP. Targeting RGS proteins with inhibitors is a strategy to increase receptor-mediated signaling. There are several existing RGS inhibitors, which including the thiadiazolidinones (TDZDs). All RGS inhibitors discovered to date are covalent modifiers of cysteine residues and these act preferentially on RGS4 over other RGS isoforms. To widen the scope of therapeutic potential of RGS inhibitors, it would be useful to have inhibitors with specificities for other isoforms. To aid in the development of new inhibitors, it will be important to understand what factors are responsible for RGS isoform selectivity. While RGS isoforms vary in their number and location of cysteines, cysteines that are shared among most RGS proteins are buried beneath the protein surface. We hypothesize that there is a dual role for cysteine complement and protein dynamics that drives specificity of TDZD inhibitors.

Interestingly, representative RGS proteins RGS4, RGS8, and RGS19 have dramatic differences in potency of inhibition when mutated to contain a single cysteine. Hydrogen-deuterium exchange (HDX) was used to evaluate differences in flexibility among RGS proteins, and deuterium incorporation was found to be correlated with TDZD potency. Molecular dynamics studies supported these differences in flexibility, and illustrated that flexibility differences may underlie solvent accessibility of shared cysteines. To understand what structural elements control

RGS domain flexibility, we focused on interhelical salt bridge-forming residues that differ among the RGS isoforms. Mutations that induced salt bridge formation in RGS19 decreased its flexibility and decreased potency of TDZD inhibition, while salt bridge removal in RGS8 and RGS4 increased flexibility and increased potency of inhibition. This suggests a causative relationship between protein dynamics and inhibitor potency. The movements observed in these proteins suggest that cysteines may be exposed to solvent by formation of a transient pocket, which may be taken advantage of in the design of non-covalent inhibitors. Finally, the role of individual conserved cysteines was evaluated. NMR studies of single-cysteine RGS8 mutants demonstrated that inhibitors can interact with either cysteine. Mass spectrometry studies showed that a TDZD inhibitor may mediate an interaction between the α_4 and α_7 cysteines in WT RGS8 by formation of a disulfide bond. As a whole, this work demonstrates a role for both cysteine interaction and protein dynamics in the control of RGS inhibitor selectivity.

ACKNOWLEDGEMENTS

I have a lot of people to thank that have supported me in the past five years. All of the lab members, current and former, were there to offer advice on countless occasions. These include, but are by no means limited to: Erika, Benita, Jeff, Kate, Behirda, Tom, Sean, Jade, Cassie, Hoa, Yajing, Maria, Nils, Maja, Indi, Zhangzhe, Clarissa, Charuta, and Melissa. Their presence has kept lab life fun and spirits high. Special thanks also to Josiah, my undergraduate mentee who has helped with countless protein preps and experiments. Thanks to the guidance of Dr. Benita Sjögren, Dr. Harish Vashisth, Dr. Karen Liby, and Dr. Jon Kaguni. In addition, having such a supportive department has been invaluable to getting by as a graduate student. I'm grateful that all of the other grad students, our GSO, the faculty, our administrative staff, and Dr. Anne Dorrance, our Graduate Program Director, have always been in my corner.

Much of this work was made possible through collaboration. This thesis features contributions that include MD simulations and analysis done by Drs. Mohammadjavad Mohammadi and Hossein Mohammadiarani in the lab of Dr. Harish Vashisth, UNH; NMR work done in conjunction with Ryan Puterbaugh and Dr. Krisztina Varga, also at UNH; and ongoing screening efforts from Dr. Arzu Uyar and Dr. Alex Dickson at MSU. I was helped greatly by Dr. Schillmiller and Dr. Jones in the RTSF Mass Spectrometry and Metabolomics Core and Dr. Sundari Chodavarapu in the lab of Dr. Jon Kaguni, who helped me develop a workflow for mass spectrometry detection of hydrogen/deuterium exchange.

I'm lucky to have had friends close at hand, including Alex, Hannah, Charlotte, Erin, Kelsey, Lynne, Dan, Shane, Kim, Sarah, Tim, Megan, and Steven. It's a wonder I got anything done with Steven roping me into side projects. I'm also very fortunate to have family that dou-

ble as my closest friends: Mom, Dad, Andy, and Margaret; Rich, Sandy, Patrick, and Laurel; and especially, my wife Kate.

Finally, I owe a lot of thanks to my mentor, Dr. Richard Neubig. Rick's Socratic method always helped me arrive at a clearer understanding and led me to ask better scientific questions on my own. I always came away from our meetings with solidified plans for further experiments and feeling motivated.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
KEY TO ABBREVIATIONS	xxi
CHAPTER 1: Introduction	1
Challenging drug targets	2
Protein-protein interactions	3
Importance of protein dynamics	5
Covalent modifiers	7
RGS proteins as therapeutic targets	8
The G-protein signaling pathway	8
RGS protein diversity	12
Physiology of RGS proteins in disease	14
RGS inhibitors	17
Recent discovery efforts	17
Thiadiazolidinone characterization	19
Contribution of this work	22
CHAPTER 2: Differential Protein Dynamics of Regulators of G-Protein Signaling: Role in Specificity of Small-Molecule Inhibitors	23
Introduction	24
Materials and Methods	25
Protein expression and purification	25
Flow cytometry protein interaction assay	26
Hydrogen/deuterium exchange	26
System setup and simulation details	29
RMSD, RMSF, and SASA Measurements	29
Results	30
Discussion	38
Conclusions	41
CHAPTER 3: An Interhelical Salt Bridge Controls Flexibility and Inhibitor Po- tency For Regulators of G-protein Signaling (RGS) Proteins 4, 8, and 19	42
Introduction	43
Materials and Methods	44
Materials	44
Protein Expression and Purification	44
Differential Scanning Fluorimetry	45
Flow Cytometry Protein Interaction Assay (FCPIA)	45

Hydrogen-Deuterium Exchange	46
Molecular Dynamics (MD) Simulation	46
Dynamic cross-correlation analysis	47
Analysis of salt-bridge interactions	48
Statistical Analysis	48
Results	49
Discussion	61
CHAPTER 4: Distinct Roles of Individual Cysteines in Covalent Inhibition of RGS	
Proteins	64
Introduction	65
Materials and Methods	66
Protein purification and expression	66
NMR Spectroscopy	67
Iodoacetamide alkylation and trypsin digestion	68
Protection of RGS8 from iodoacetamide alkylation by CCG-203769	69
Protein mass spectrometry	69
Flow cytometry protein interaction assay	69
Non-reducing SDS-PAGE	70
Results and Discussion	70
Cys148 in RGS4 is more accessible to a covalent modifier than Cys95	70
CCG-203769 can directly act upon either cysteine in RGS8	71
Cys ¹⁶⁰ RGS8 is more sensitive to compound-induced denaturation than WT.	75
Functional inhibition by CCG-203769 is altered in cysteine mutants	75
CCG-203769 induces an intra-protein disulfide in WT RGS8.	76
Among single-cysteine RGS8 mutants (Cys ¹⁰⁷ and Cys ¹⁶⁰), CCG-203769 induces dimerization via an inter-protein disulfide.	78
CCG-203769 induces inter-protein disulfide in RGS4	79
Conclusions	80
CHAPTER 5: Identification of Transient Pockets in RGS4 and RGS19	83
Introduction	84
Approach and Results	86
Pocket Identification	86
Pocket Clustering	88
Frames for screening	90
Conclusions	91
CHAPTER 6: Conclusions and Future Directions	93
Role of individual cysteines in action of inhibitors	94
Role of protein dynamics in RGS inhibitor selectivity	94
Future research in understanding action of TDZD inhibitors	96
Continuing discovery of non-covalent inhibitors	98
APPENDIX	99

LIST OF TABLES

Table 2-1:	Summary of MD simulations.	34
Table 3-2:	Details of MD simulations.	47
Table 3-3:	The salt-bridge interaction within the α_4 - α_7 bundle of helices in single-cysteine structure of RGS4, RGS8, and RGS19 from MD simulations and potency of CCG-50014 inhibition of single-cysteine RGS proteins in our previous work. ¹⁴⁴	50
Table 3-4:	Interaction affinities between $G\alpha_o$ and RGS proteins and mutants	58
Table A-1:	Model definitions and corresponding metrics. Among models reported in the literature are models M1 through M6 (empirical models) and the model M7 (a fractional population model). For models reported in this work, M8 is an empirical model and M9 is a fractional population model. Additional details on models M8 and M9 are provided in supporting information.	104
Table A-2:	Summary of MD simulations.	124
Table A-3:	Models proposed in this work	127
Table A-4:	Details on all protection factor correlation models with the default and re-optimized values of their parameters. Optimized values based upon simulations conducted using CHARMM and AMBER force-fields are listed with superscripts <i>ch</i> and <i>am</i> , respectively. In addition, details on two new models M8 and M9 proposed in this work are listed.	128

LIST OF FIGURES

Figure 1-1:	Activation of G-protein signaling upon agonist binding at GPCR.	10
Figure 1-2:	Activation of different signaling pathways is mediated by different G-protein subtypes.	11
Figure 1-3:	RGS are GTPase-Activating Proteins (GAPs). They terminate G-protein signaling by catalyzing hydrolysis of GTP on $G\alpha$	12
Figure 1-4:	The circuitry of the motor pathway. RGS4 is expressed in the striatum. In the Parkinson's disease state, dopaminergic input from the substantia nigra to the striatum is lost.	16
Figure 1-5:	The role of RGS4 in response to dopamine signaling in the indirect and direct pathway spiny projection neurons of the striatum.	18
Figure 1-6:	Thiadiazolidinones CCG-50014, the lead compound, and CCG-203769, an analog with improved specificity for RGS4.	20
Figure 1-7:	Locations of cysteines in RGS protein. $G\alpha$ is shown in gray spheres, RGS in shown in light blue. Cysteines 71 and 132 in RGS4 (red) are not conserved among RGS proteins. Cysteine 148 in RGS4 (blue) is shared by RGS8 and RGS4. Cysteine 95 in RGS4 (green) is the best conserved cysteine among RGS proteins, found in all isoforms except RGS6 and RGS7.	21
Figure 2-1:	Alignment of fragments observed by mass spectrometry following cleavage of RGS proteins by pepsin. Horizontal bars indicate length and position of observed fragments. The two N-terminal residues of each fragment were excluded from analysis due to rapid back-exchange. Vertical gray boxes indicate approximate positions of helices within the RGS domain.	28
Figure 2-2:	(A) Locations of cysteines in RGS4, RGS8, and RGS19. (B) Potency of CCG-50014 against RGS19, which has only one cysteine, and mutant RGS4 and RGS8 containing only the shared α_4 helix cysteine. n=3.	31
Figure 2-3:	(A-E) Kinetics of deuterium exchange in selected protein fragments from (A) α_4 , (B) α_5 , (C) α_5 - α_6 interhelical region, (D) α_6 and (E) α_7 . Sequences of observed fragments are aligned with residue numbers of each fragment indicated. Cysteine locations are marked in red. n=3.	32

Figure 2-4:	(A) Global kinetics of deuterium exchange. Deuterium incorporation (DI) is expressed as a percent of exchangeable amide hydrogen positions. Where fragments overlap, data is displayed as average DI of observed fragments. (B) Degree of DI at 300 minutes in 90% D ₂ O is mapped onto protein structure of RGS4, RGS8, and RGS19. n=3.	33
Figure 2-5:	Root mean squared fluctuations (RMSF) per residue during 2 μ s MD simulations of (A) RGS4 (PDB: 1AGR), (B) RGS8 (PDB: 2ODE), and (C) RGS19 (PDB: 1CMZ). The RMSF trends for each protein for the simulation set 2 are shown in Fig. 2-6. Gray bars indicate helical regions.	35
Figure 2-6:	Root mean squared fluctuations across protein sequence during 3 μ s MD simulations of (A) RGS4 (PDB: 1AGR), (B) RGS8 (PDB: 2ODE), and (C) RGS19 (PDB: 1CMZ). Gray bars indicate helical regions.	35
Figure 2-7:	Solvent-accessible surface areas (SASA) are shown for sulfur atoms in shared cysteines on α_4 helix for simulation set 1 (A) and set 2 (B) in RGS4, RGS8, and RGS19, and for shared cysteines on α_6 - α_7 interhelical loop in simulation set 1 (C) and set 2 (D) in RGS4 and RGS8.	36
Figure 2-8:	Conformational changes during molecular dynamics simulations. Root mean square deviations of α_6 helix and α_6 - α_7 loop, starting conformation, and a snapshot conformation during MD simulation are shown for (A, D, G) RGS4, (B, E, H) RGS8, and (C, F, I) RGS19. Protein regions plotted in MD trajectories are depicted in color in protein structures. Arrows indicate locations of notable solvent exposure during simulation.	37
Figure 2-9:	Snapshot of RGS19 from simulation set 2 at 240 ns. Cleft opening observed in simulation set 1 (Fig 6I) was recapitulated in this simulation.	38
Figure 3-1:	(A) Alignment of RGS19, RGS4, and RGS8 sequences in α_4 - α_7 helix bundle. Charged residues that make interhelical contacts are indicated in red and blue. RGS19 has 1, RGS4 has 3, and RGS8 has 4 salt bridges. Structural alignments of α_4 - α_5 (B), α_5 - α_6 (C), and α_6 - α_7 (D) helix pairs are shown, with highlighted residues in panel a rendered as sticks. RGS19 (PDB 1CMZ) is in green, RGS4 (PDB 1AGR) is in yellow, and RGS8 (PDB 5DO9) is in cyan. Black brackets in panel A indicate residues depicted in panels B, C, and D	50
Figure 3-2:	L118D mutation increases thermal stability of RGS19, but Q183K mutation has no significant effect (n = 3, 1-way ANOVA with Sidak's multiple comparison test. ****p < 0.001). L118D mutation in RGS19 has reduced potency of inhibition of CCG-50014, but Q183K mutation does not. K _i , calculated using a Cheng-Prusoff correction, ²³² is reported to account for effect of mutations in RGS on G α_o affinity.	51

Figure 3-3:	Thermal stability was determined by differential scanning fluorimetry. (A) The L118D mutation in RGS19 increased melting temperature by 7 °C compared to WT. (B) The E84L mutation in RGS8 decreased melting temperature by 8 °C. (C) The RGS4 D90L mutation introduced a biphasic melt curve and increased melting temperature by 5 °C. For each pair, the three replicate derivative melt curves are shown on the left and average melt temperatures are shown on the right. Error bars represent SD. n = 3. Analyzed by 1-way ANOVA with Sidak's Multiple Comparisons test. ****p < 0.0001	53
Figure 3-4:	The traces of root-mean-squared-deviation (RMSD) vs. simulation time (μs) for (a) RGS4 D90L, (b) RGS8 E84L, and (c) RGS19 L118D. Two independent simulation runs for each structure are presented, and the wild-type runs are presented from our previous work. ¹⁴⁴	54
Figure 3-5:	ΔRMSF in between WT and mutant simulation trajectories	55
Figure 3-6:	Dynamic cross correlation matrix calculated for the Cα atoms of (A) RGS19/RGS19 L118D, (B) RGS8/RGS8 E84L, (C) RGS4/RGS4 D90L. Horizontal dotted lines indicate the regions of the α ₄ helix, while vertical solid lines indicate the regions of the α ₅ helix for each protein. The color scheme ranges from anticorrelation (-1.0, blue), no correlation (0, green), and positive correlation (+1.0, red). Values are the average for the two independent simulation runs.	57
Figure 3-7:	Difference in %deuterium incorporation (Δ%DI) between mutated and unmutated proteins in RGS19 L118D (A), RGS8 E84L (B), and RGS4 D90L (C) fragments, as measured by HDX. Red arrows indicate fragments containing mutated residue, and black arrows indicate fragments containing conserved α ₄ cysteine. Kinetics of deuterium incorporation in these fragments for individual constructs are shown below. n = 3. Error bars represent SD. Analyzed by 2-way ANOVA with Sidak's multiple comparisons test. *p < 0.05, **p < 0.01, ****p < 0.0001.	59
Figure 3-8:	Potency of inhibition of CCG-50014 against α ₄ is altered in salt bridge mutants of RGS proteins. (A) RGS4 IC ₅₀ : 8.8 μM, RGS4 D90L IC ₅₀ : 2.2 μM. (B) RGS8 IC ₅₀ : 29 μM, RGS8 E84L IC ₅₀ : 4.6 μM. (C) RGS19 IC ₅₀ : 7.0 μM, RGS19 L118D IC ₅₀ : 1.1 μM. n=3.	60
Figure 4-1:	(A) Locations of cysteines in RGS protein based on structure of RGS4 (PDB: 1AGR). α ₄ and α ₇ cysteines, conserved across multiple RGS proteins, are marked in blue. The α ₃ and α ₆ helix cysteines, unique to RGS4, are marked in red. (B) Degree of IAA alkylation at Cys71 (α ₃), Cys95 (α ₄), Cys132 (α ₆), and Cys148 (α ₇) in RGS4.	71

Figure 4-2:	WT RGS8 protein NMR spectra. (A) ^1H - ^{15}N HSQC NMR spectrum of WT RGS8. (B) The structure of ligand CCG-203769. (C) Overlay of ^1H - ^{15}N HSQC NMR spectra of WT RGS8 before (red spectrum) and after the addition of its ligand CCG-203769 at 1:1, 1:2, and 1:4 RGS8:ligand ratio (grey spectra). Shifted residues are highlighted in the zoomed spectrum. Spectra were acquired at 25 °C on a Bruker AVANCE III HD 800 MHz NMR spectrometer equipped with a TCI Cryoprobe at the CUNY Advanced Science Research Center NMR facility.	73
Figure 4-3:	Chemical shift perturbation of WT and single-cysteine RGS8 protein NMR spectra upon the addition of ligand CCG-203769 ^1H - ^{15}N HSQC NMR spectra of RGS8 were overlaid before (red spectrum) and after the addition of its ligand CCG-203769 at 1:1 RGS8:ligand ratio (black spectra) for (A) WT RGS8 (B) Cys ¹⁰⁷ RGS8, and (C) Cys ¹⁶⁰ RGS8. (D) The magnitude of chemical shift perturbation. Spectra were acquired at 25 °C on a Bruker AVANCE III HD 800 MHz (WT and Cys ¹⁰⁷ RGS8) or a Bruker AVANCE III HD 700 MHz (Cys ¹⁶⁰ RGS8) NMR spectrometers equipped with Cryoprobes at the CUNY Advanced Science Research Center NMR facility.	74
Figure 4-4:	Inhibition of RGS-Gα binding for WT, Cys ¹⁶⁰ , and Cys ¹⁰⁷ RGS8 in response to increasing concentrations of CCG-203769 was measured by FCPIA. WT IC ₅₀ = 25 μM), Cys ¹⁶⁰ IC ₅₀ = 2.2 μM, and Cys ¹⁰⁷ was not inhibited.	76
Figure 4-5:	CCG-203769 masks cysteine alkylation by IAA by inducing disulfide bond. (A) Deconvoluted mass spectra of WT RGS8 (first column), Cys ¹⁶⁰ RGS8 (second column), and Cys ¹⁰⁷ RGS8 (third column). Spectra were taken before treatment (first row), after excess of IAA (second row), and pretreated with CCG-203769 before addition of IAA (third row). (B) WT, Cys ¹⁶⁰ , and Cys ¹⁰⁷ RGS8 mass changes analyzed by SDS-PAGE after treatment with vehicle, 250 μM CCG-203769, or CCG-203769 followed by 1 mM DTT. Monomer mass indicated with black arrow and dimer mass indicated with red arrow.	77
Figure 4-6:	RGS4, RGS8, and RGS19 mass changes analyzed by SDS-PAGE after treatment with vehicle, 250 μM CCG-203769, or CCG-203769 followed by 1 mM DTT.	80
Figure 4-7:	Proposed mechanism of disulfide bond induction by CCG-203769 in RGS8	81
Figure 5-1:	Locations of pocket-forming residues in RGS4 (top) and RGS19 (bottom). Color indicates frequency with which each atom touches a pocket alpha sphere. Blue is less frequent and red is more frequent.	87
Figure 5-2:	Pocket volume and mean local hydrophobic density (MLHD) plotted over the simulation trajectory for RGS19 (A) and RGS4 (B). Pockets in RGS19 were larger and more frequent than those in RGS4.	88

Figure 5-3:	Clustering of pocket states for RGS19 (A) and RGS4 (B). Volume is plotted against MLHD, and color indicates distinct clusters. An ensemble of pockets representing clusters with high MLHD and a variety of pocket volumes were selected for structure based screening.	90
Figure 5-4:	Pocket states that are representative of cluster 4 and 7 in RGS19 (A and B) and cluster 2 and 6 in RGS4 (C and D). Pocket-forming atoms illustrated with white surface.	92
Figure A-1:	Kinetic scheme for HDX is highlighted. A conformational fluctuation in the protein exposes buried amide groups (blue) (closed state) to solvent (open state) where amide hydrogens (white) are exchanged by deuterium (yellow) with an intrinsic rate constant k_{int}	102
Figure A-2:	Sequence and structural views of RGS proteins. (A) Sequence alignment of RGS4, RGS8, and RGS19 is shown with conserved residues highlighted in red; blue boxes indicate residues that are conserved between at least two among three RGS proteins. (B) Shown are front and back views of the overlay of RGS4 (PDB code 1AGR), RGS8 (PDB code 2ODE), and RGS19 (PDB code 1CMZ) structures with each of the nine helices uniquely colored. Regions rendered as white cartoons are interhelical loops.	105
Figure A-3:	Comparisons of model predictions of HDX-MS data across all three RGS proteins. Performance metrics (relative error, E , and correlation coefficient, CC) for different models are shown based upon data averaged from all trajectories of RGS4, RGS8, and RGS19 conducted with the CHARMM-FF (data in panels A and B) and the AMBER-FF (data in panels C and D). (A, C) The relative error between the predicted and observed %DI [$E(x, y) = \sum_{i=0}^n x_i - y_i / \sum_{i=0}^n y_i$]. (B, D) Correlation coefficient between the predicted and observed %DI [$CC(x, y) = \sum (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$]. Gray bars are for models with the default parameters reported in the literature, blue bars are their re-optimized versions based upon our experimental data, and red bars are for new models proposed in this work. No performance data for the original model M5 are reported because the parameter values were not available from the original work, ⁴² but the performance data are reported for the optimized version of this model (M5*) based upon our experimental data.	109
Figure A-4:	Comparisons of model predictions of HDX-MS data for each RGS protein. The definitions of E and CC , and other details are the same as in Figure 3. Colored bars distinguish data for each RGS protein: black bars, RGS4; blue bars, RGS8; and magenta bars, RGS19	110

Figure A-5:	The exposure of amide hydrogens in the NMR structures of RGS proteins. Shown are the maximum (open circles) and the average (solid circles) values of the solvent accessible surface area for all amide hydrogens in the NMR structures of RGS4 (panel A) and RGS19 (panel B). In both panels, the absence of filled circles for certain amides as well as the absence of open circles in panel B, is due to the approximately nil SASA values for those amides. The absence of open circles for RGS4 in panel A is due to the lack of availability of more than 1 conformer in the NMR structure of RGS4 as opposed to 20 conformers in the NMR structure of RGS19.	116
Figure A-6:	Mean residence times for the open and closed states of amide hydrogens. Data are shown from all simulations of RGS4, RGS8, and RGS19 conducted with the CHARMM-FF (panel A) and the AMBER-FF (panel B). The MRT calculations were carried out using our proposed fractional population model M9 that showed consistent predictions with the HDX-MS data.	118
Figure A-7:	Experimentally measured percentage deuterium incorporation (%DI) of fragments in RGS proteins at $t = 0, 3, 10, 30, 100, 300,$ and 1000 minutes (RGS4: top row; RGS8: middle row; RGS19: bottom row).	129
Figure A-8:	Definitions of fragments for each RGS protein. Each fragment comprises residues whose color determines their location in nine α helices of each RGS protein. Residue names in connecting loops are highlighted in black, but shown as white cartoons in the protein structure. All helices are colored and labeled in the protein rendering.	129
Figure A-9:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1AGR and AMBER force-field.	130
Figure A-10:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1EZT and AMBER force-field	130
Figure A-11:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2IHD and AMBER force-field	131
Figure A-12:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2ODE and AMBER force-field	131

Figure A-13:	Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and AMBER force-field.	132
Figure A-14:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and AMBER Force-field	132
Figure A-15:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and AMBER Force-field	133
Figure A-16:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and AMBER Force-field.	134
Figure A-17:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and AMBER Force-field.	135
Figure A-18:	Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and AMBER Force-field.	136
Figure A-19:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and AMBER Force-field.	136
Figure A-20:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and AMBER Force-field.	136
Figure A-21:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and AMBER Force-field.	137

Figure A-22:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and AMBER Force-field.	137
Figure A-23:	Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and AMBER Force-field.	137
Figure A-24:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1AGR and CHARMM Force-field.	138
Figure A-25:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1EZT and CHARMM Force-field.	138
Figure A-26:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2IHD and CHARMM Force-field.	139
Figure A-27:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2ODE and CHARMM Force-field.	139
Figure A-28:	Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and CHARMM Force-field.	140
Figure A-29:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and CHARMM Force-field.	140
Figure A-30:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and CHARMM Force-field.	141

Figure A-31:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and CHARMM Force-field.	142
Figure A-32:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and CHARMM Force-field.	143
Figure A-33:	Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and CHARMM Force-field.	144
Figure A-34:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and CHARMM Force-field.	144
Figure A-35:	Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and CHARMM Force-field.	144
Figure A-36:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and CHARMM Force-field.	145
Figure A-37:	Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and CHARMM Force-field.	145
Figure A-38:	Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and CHARMM Force-field.	145
Figure A-39:	Deuterium incorporation is mapped on RGS proteins at $t = 1000$ min as observed in experiments and as predicted by the models M7, M8, and M9. Data are presented for the CHARMM-FF simulations of RGS4, RGS8, and RGS19.	146

Figure A-40:	Root mean squared fluctuations (RMSF) per residue across protein sequences are shown from 2- μ s long MD simulations of (A) RGS4 (PDB: 1AGR, 1EZT), (B) RGS8 (PDB: 2IHD, 2ODE), and (C) RGS19 (PDB: 1CMZ). Color bars indicate helical regions.	147
Figure A-41:	Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS4, CHARMM-FF).	148
Figure A-42:	Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS8, CHARMM-FF).	149
Figure A-43:	Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS4, AMBER-FF).	150
Figure A-44:	Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS8, AMBER-FF).	151
Figure A-45:	Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS19, CHARMM-FF).	151
Figure A-46:	Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS19, AMBER-FF).	152
Figure A-47:	The residues protected by hydrogen-bonds or salt-bridging interactions are highlighted (panels A and B). The traces for distances between the centers-of-masses of residue pairs are shown in panel C (S120-Q122) and panel D (E84-R119 and E111-R119).	152
Figure A-48:	SASA data similar to Fig. A-6 are shown from MD simulations of all RGS proteins for both force-fields (CHARMM-FF, panel A; AMBER-FF, panel B). Color and labeling details are similar to Fig. A-6	153
Figure A-49:	Corrected mean residence times for open-states of amide hydrogens are shown. Other details are similar to Fig. A-6.	154
Figure A-50:	Residue-residue correlations among open states of all amide-hydrogens (CHARMM- FF, RGS4 (PDB code 1AGR), model M7). The correlation matrix is calculated based on the probability that two amide hydrogens simultaneously explore open states; $C(i, j) = (P(i, j) - P(i)P(j))/(P(i)P(j)(1 - P(i))(1 - P(j)))^{0.5}$	155
Figure A-51:	Data similar to A-50 are shown for RGS8 (CHARMM-FF, RGS8 (PDB code 2ODE), model M7).	156

Figure A-52: Probability of a closed to open transition in a given amide vs. simulation length (μs) is presented based upon Poisson statistics. Data are shown for PFs = 102, 104, 106, and 1011 with $\tau_O = 20$ ps and 100 ps. 157

KEY TO ABBREVIATIONS

2-AG—2-arachadonylglycerol

5HT—5-hydroxytryptamine (serotonin)

6-OHDA—6-hydroxydopamine

AC—adenylyl cyclase

BME—2-mercaptoethanol

BPTI—bovine pancreatic trypsin inhibitor

cAMP—cyclic adenosine monophosphate

DAG—diacylglycerol

DCC—dynamic cross-correlation

DEP—Dishevelled, Egl-10, Pleckstrin)

DI—deuterium incorporation

DMSO—dimethyl sulfoxide

dSPN—direct pathway spiny projection neuron

DSS—4,4-dimethyl-4-silapentane

DTT—dithiothreitol

eCB—endocannabinoid

FCPIA—flow cytometry protein interaction assay

GAIP—G alpha interacting protein

GAP—GTPase-activating protein

GDI—guanine nucleotide dissociation inhibitors

GDP—guanosine diphosphate

GEF—guanine nucleotide exchange factor

GGL—G gamma-like

GPCR—G-protein coupled receptor

GPe—globus pallidus externus

GPi—globus pallidus internus

GTP—guanosine triphosphate

HDX—hydrogen/deuterium exchange

HEPES—4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HSQC—heteronuclear single quantum coherence

IAA—iodoacetamide

IP₃—inositol trisphosphate

iSPN—indirect pathway spiny projection neuron

LTD—long-term depression

LTP—long-term potentiation

MC—Monte Carlo

MD—molecular dynamics

MLHD—mean local hydrophobic density

MRT—mean residence time

MS—mass spectrometry

NMR—nuclear magnetic resonance

NPT—number of particles, pressure, temperature

NVT—number of particles, volume, temperature

PDB—protein data bank

PEPCK—phosphoenolpyruvate carboxykinase

PF—protection factor

PIP₂—phosphatidylinositol 4,5-bisphosphate

PLC—phospholipase C

PKA—protein kinase A

RGS—regulator of G-protein signaling

RH—RGS Homology

RMSD—root-mean-square deviation

RMSF—root-Mean-Square Fluctuation

SASA—solvent-accessible surface area

SDS-PAGE—sodium dodecyl sulfate polyacrylamide gel electrophoresis

SPN—spiny projection neuron

TDZD—thiadiazolidinone

WT—wild type

CHAPTER 1:

Introduction

Protein dynamics play a major role in protein-ligand interactions.¹⁻³ While dynamics have a role to play in virtually all molecular interactions,^{1,2,4-6} this facet of binding may be especially worth considering in the context of inhibition of challenging targets such as intracellular protein-protein interactions, where there are generally no binding pockets evolutionarily built for receiving small-molecule signals.^{7,8} In this thesis, I discuss the role of protein dynamics in specificity of an interaction between a series of covalently-acting small molecules and their targets, Regulators of G-protein Signaling (RGS) proteins. These compounds inhibit by disrupting the interaction between RGS proteins and their binding partner, the alpha subunit of the heterotrimeric G-protein ($G\alpha$). While this system may be unique in many respects, giving similar consideration to protein dynamics when probing the mechanism of other biologically active chemicals will provide valuable insight into the drivers of drug specificity.

Challenging drug targets

Intermolecular interactions are behind the function all biological molecules. Historically, the field of pharmacology has been built on studying receptors and molecules that can bind to them.⁹⁻¹¹ Receptors are proteins that have evolved to receive chemical signals from the outside of the cell and relay them to the interior. As such, this obviates the need for the stimulus molecule to enter the cell itself. In addition, receptors generally have a ready-made binding pocket, making them convenient to target with exogenous chemicals. These pockets take very specific shapes, which will only allow certain molecules to bind. Ligand specificity can be compared to key in a lock: the compound must be just the right shape, or it will not fit in the binding pocket.

Receptors have historically been thought of as readily druggable targets: proteins for which small molecules that bind will be relatively easy to find. However, only a small portion of

all proteins are considered druggable, and only a subset of these may have any medical utility.¹² Most existing drugs target enzymes or receptors, which have pockets for binding substrates or external chemicals respectively. As this low-hanging fruit gets exhausted, however, fewer and fewer drugs are developed for new targets.^{12,13} If we can find ways to inhibit unique candidate proteins, it will increase the potential for continued discovery of small-molecule therapeutics.

One alternative to traditional targeting of cell-surface receptors, while still using small molecules, is to develop compounds that target intracellular signaling proteins. This is often more difficult because these proteins have not evolved to respond directly to chemical signals, so they may lack dedicated binding pockets. Some intracellular proteins, however, do have binding pockets and are considered receptors. These include nuclear receptors, a receptor family noted for mediating endocrine signals such as androgens, estrogen, thyroid hormones, and more, including many yet-undiscovered ligands.¹⁴ Still other intracellular proteins, while not binding external chemical signals as a part of their canonical function, bind to one another to mediate signaling cascades. These may still have cavities that can be exploited in developing inhibitors. One example is kinases, a protein family for which there has been a sudden rise in the number of available inhibitors, most of which act by competitive inhibition of ATP at the nucleotide binding site.^{15,16} The ability to pursue intracellular proteins as targets will widen the scope of druggable candidates, and greatly expand the possibilities for pharmacological modification of disease.

Protein-protein interactions

One key way in which proteins transmit signals is by binding to one another. By modulating protein-protein interactions (PPIs) among signal transducing proteins, these signals can be tweaked. However, protein-protein interactions can be difficult to target. This is evidenced

by the observation that high throughput screening libraries have had lower success rates in identifying PPI inhibitors than in identifying inhibitors of traditional receptors.¹⁷ One reason is that the interface between proteins is quite large and often relatively flat. The average interface size between a small molecule of 500 Da and the binding pocket on its protein target is about 300 Å²,¹⁸ an order of magnitude smaller than protein-protein interfaces, which range in area from 1500-3000 Å².^{19,20} Despite these interfaces being flat and lacking deep binding pockets, there are now several examples of molecules that bind to these interfaces.²¹⁻²³ However, the binding surface between proteins and inhibitors acting at protein-protein interfaces are more spread out, necessitating larger molecules to maintain the same number of contacts.^{23,24} One difficulty in the identification of new inhibitors of PPIs is that most discovery efforts, including high-throughput screens, use compound libraries that are biased toward smaller molecules similar to those that bind existing receptors.^{25,26}

Protein function may also be tweaked by binding of a molecule at a location distant from its protein-protein binding interface, its substrate binding pocket, or orthosteric small-molecule binding site.^{2,27,28} This mechanism is dependent on protein allostery, where a conformational change at one part of the protein may induce conformational changes at an active site on another part of the protein. An argument in favor of the use of allosteric regulators is that they do not necessarily preclude binding of an endogenous ligand, substrate, or protein binding partner, so they may modulate the intensity of existing signaling when and where it already occurs rather than induce or block signaling globally.²⁹ There are many examples of allosteric modulators, both for receptors³⁰ and for other protein targets.^{31,32} This may be important when targeting protein-protein interactions because any cavity on a protein surface may be sufficient to alter the protein's function, even if a binding pocket is not available directly at the protein-protein

interaction surface.

Importance of protein dynamics

Our understanding of the shape that a ligand and receptor take on upon binding, and the fit of the former into the latter, is driven by the field of structural biology. Advances in technology using such techniques as X-ray crystallography, cryogenic electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR) have allowed high resolution determination of three-dimensional structures of receptor-ligand complexes. Using these protein structural models, a pharmacologist may understand how shape and molecular interactions drive binding, and may hypothesize ways the ligand might be altered to improve affinity or specificity.^{33,34} However, attempts to predict ligand binding using *in silico* docking techniques may be limited when using a single static structure, because they do not account for protein flexibility.^{35–37} The role of dynamics may explain why expected binding results obtained by *in silico* docking and virtual screening techniques using a static structure are often far removed from affinities and structures that are experimentally observed.³⁸ Solution NMR offers an ensemble of structural states, but these still may not be representative of the wide variety of transient movements a protein may make in solution. Although they are more computationally intensive, there are now methods for *in silico* evaluation of binding that account for protein flexibility.^{37,39,40} Protein dynamics are well worth considering, as these might have a strong influence on ligand binding kinetics and affinity.³

There are two models for the role dynamics might play in binding of a molecule in a pocket, called conformational selection and induced fit.⁴¹ In the conformational selection model, a protein's conformation in solution, including the shape and properties of its binding pocket, are undergoing continuous fluctuations. Occasionally, this pocket will be amenable to compound

binding. Therefore, a ligand's on-rate may be influenced by how frequently the protein takes a certain conformation. Many apo-proteins exhibit conformations in solution similar to conformations found when ligand is bound. One such protein is adenylate kinase (AdK). By locking the protein AdK in a conformation similar to its ligand-bound state, ligand affinity was drastically increased, providing evidence for this model.⁴²

In the induced fit model, a compound's interaction with the protein surface or pocket may induce a conformational change that allows the compound to bind with higher affinity.¹ Proteins in which an induced fit-like mode of ligand binding apply include lid-gated enzymes, in which the active site is covered by a "lid" that closes around it. It would be difficult for such a complex to be compatible with a pure conformational selection model, since even if the apo-protein sampled a conformation similar to the substrate-bound one in solution, the lid would sterically occlude compound entry.^{43,44} Induced fit would require an initial binding event in which the compound first makes contact with the protein before inducing a conformational change. Structural evidence for such an "encounter complex" has been seen in the phosphoenolpyruvate carboxykinase (PEPCK) enzyme.⁴⁵

The actual behavior of a protein-compound interaction may not purely fit with one or the other of these models. Some would consider the two models less different than they appear, or in fact just different perspectives of the same mechanism.⁴⁶ For example, a combination could exist in which a certain conformation of the apo-protein is required for a compound to bind, and as the compound enters, it pushes on the residues forming the pocket to cause a secondary conformational adjustment.⁴⁷ Importantly, regardless of which model is more relevant to a particular protein-compound interaction, failure to account for protein dynamics may prevent a full structural understanding of ligand binding.

Some pockets that are potentially druggable may in fact not be present in the crystal structures, but will open occasionally as the protein moves in solution. These are referred to as transient pockets. A related concept is that of cryptic pockets, which are pockets that only become apparent once a compound is bound. As with conformational selection and induced fit, there may be much overlap between transient and cryptic pockets. Many cryptic pockets may in fact be more flexible than the surrounding residues and sample open-like states, even in the absence of ligand.⁴⁸ It may be helpful to identify transient pockets to develop PPI inhibitors that act directly at the protein interface.^{7,8} However, it is also possible to find transient or cryptic pockets that affect protein activity allosterically. A great example is in the case of K-Ras, which is frequently mutated in cancers to become constitutively active and makes a very desirable drug target. Many attempts at finding inhibitors have failed and K-Ras was long considered undruggable.⁴⁹ However, a covalent inhibitor was discovered that binds to the cysteine in a G12C K-Ras mutant.⁵⁰ Analysis of the protein-adduct structure revealed that the compound resided in a cryptic pocket: one that was not previously apparent in apo structures of K-Ras. Additionally, this inhibitor acts allosterically, binding adjacent to, as opposed to obscuring, the GTP binding site.⁵⁰⁻⁵² As such, the role of dynamics in the formation of transient or cryptic pockets has gained recognition for its importance in discovery of molecules that bind to difficult targets.

Covalent modifiers

Another mechanism of inhibition is covalent modification. Covalent binding of an inhibitor may improve potency by reducing the off-rate to a negligible level, and there are several examples of well-known drugs that act by covalent mechanisms. In fact, some of the most successful and widely used drugs are covalent inhibitors.⁵³ Some classic examples include aspirin,

which inhibits cyclooxygenase (COX) by acetylation of a serine in the active site,⁵⁴ and clopidogrel, which is converted to an active metabolite by liver enzymes and inhibits P2Y₁₂ adenosine receptors by thiol-based cysteine modification.⁵⁵ In more recent years, success in development of covalent inhibitors has been met in the field of kinase inhibitors, with several covalent drugs finding FDA approval.^{56–59}

However, concerns about off-target effects and toxicity can be a barrier to development of covalent modifiers, and these compounds are generally avoided by the pharmaceutical industry.^{60,61} One concern is development of an immune response to a covalent drug-protein adduct.^{62,63} For example, allergy to penicillin is mediated by IgE and T-cell responses to penicillin-modified peptides.⁶⁴ This does not occur in all patients, but there is a fear that such idiosyncratic reactions may only be discovered once the drug is brought to a larger patient pool. Although there are some cases where covalent modification is tolerable or even useful in an inhibitor, finding non-covalent inhibitors is highly desirable to reduce risk and ease drug development.

RGS proteins as therapeutic targets

The G-protein signaling pathway

Perhaps the most pharmacologically important class of cell surface receptors is G-Protein Coupled Receptors, or GPCRs. GPCRs are not unique to humans; they have been playing a role in responding to stimuli since the dawn of multicellular organisms. They are found not only among animals, but in other kingdoms including fungi⁶⁵ and possibly in plants.⁶⁶ The human genome contains over 800 GPCRs,⁶⁷ which play a myriad of physiological roles, from the nervous system (including neurotransmission, sensation of taste, pain, and vision) to cell to cell signals in other

physiological systems such as cytokines signaling in the immune system and hormones in the endocrine system. As such, GPCRs and their associated signaling partners make attractive drug targets for multiple applications.

The GPCR protein has seven transmembrane domains. The extracellular loops between these helices form a binding pocket for a ligand. Generally ligand binding is a reversible, non-covalent interaction, and the receptor may become active when a ligand is bound. Much cell-cell signaling throughout the body, and particularly in the CNS, is mediated by such reversible ligand-receptor interactions with GPCRs. These include glutamate with mGluRs,^{68,69} GABA at GABA_B receptors,^{70,71} monoamines at their respective receptors,⁷²⁻⁷⁴ and more. However, diverse variations also exist. For example, retinal is a ligand for the opsin family of GPCRs. It can remain bound to the receptor, and change conformation to activate the receptor in the presence of light. Another example is protease-activated receptors, a family with an N-terminal tail that may be cleaved to form a tethered ligand: a ligand that is a part of the receptor rather than an external signal.⁷⁵

The intracellular loops form a binding site for an intracellular effector: the heterotrimeric G-protein. The G-protein is comprised of the alpha subunit (denoted $G\alpha$), which binds a guanine nucleotide, and the beta and gamma subunits, which form an obligate dimer ($G\beta\gamma$). The C-terminal tail of the of a GDP-bound alpha subunit may dock to the intracellular side of the receptor.⁷⁶ Once an agonist binds to the receptor, a conformational change in the receptor allows the GDP to dissociate from the $G\alpha$ subunit and GTP to bind in its place. The GTP-bound G-protein is now said to be in its active conformation. It dissociates from the receptor and the alpha subunit dissociates from the beta-gamma dimer, allowing each to initiate signaling by binding to downstream receptors.

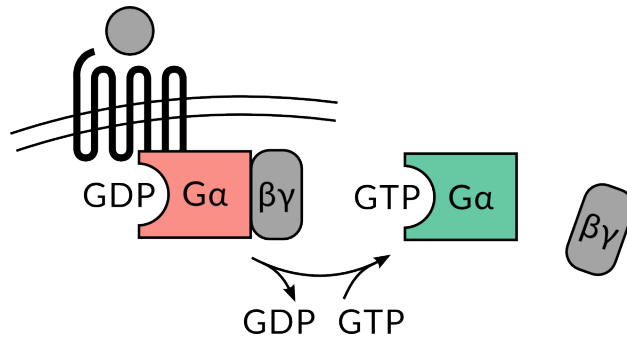


Figure 1-1: Activation of G-protein signaling upon agonist binding at GPCR.

G α proteins are divided into four main categories: the G $_s$, the G $_{i/o}$, the G $_q$, and the G $_{12}$ families. G $_s$ proteins have adenylyl cyclase (AC) as their effector, binding to AC and stimulating production of cyclic AMP, or cAMP. cAMP is a necessary signaling molecule for a signaling cascade that starts with activation of Protein Kinase A (PKA), which is also known as cAMP-dependent protein kinase. Thus, in the extracellular presence of an agonist for G $_s$ coupled GPCRs, G $_s$ can elicit an increase in intracellular cAMP. G $_{i/o}$ proteins have the opposite effect: when activated, they reduce intracellular cAMP. This effect may be mediated by the G $\beta\gamma$ subunit dimer rather than the G $_{i/o}$ alpha subunit itself. In fact, G $\beta\gamma$ is responsible for much of the G-protein signaling. In the case of G $_{i/o}$, the “active” G α could exert its effects merely by releasing G $\beta\gamma$.^{77,78} It should be noted that G $\beta\gamma$ is released during G $_s$ signaling as well, but any inhibitory effects of the G $\beta\gamma$ dimer, if present, are overshadowed by the activation induced by G $_s$. G $_q$ and G $_{12}$ families act through cAMP-independent mechanisms. G $_q$, when active, is known for induction of calcium release from the endoplasmic reticulum (ER). G $_q$ binds to Phospholipase C (PLC), which cleaves the phospholipid phosphatidylinositol 4,5-bisphosphate (PIP $_2$) into diacylglycerol (DAG) and inositol trisphosphate (IP $_3$). IP $_3$ binds to the IP $_3$ receptor on the ER, causing calcium release into the cytosol.⁷⁹ One well-known example of G $_q$ mediated signaling is endocrine regulation of smooth muscle, where calcium release is necessary for contraction of actin and myosin.⁸⁰ It should also

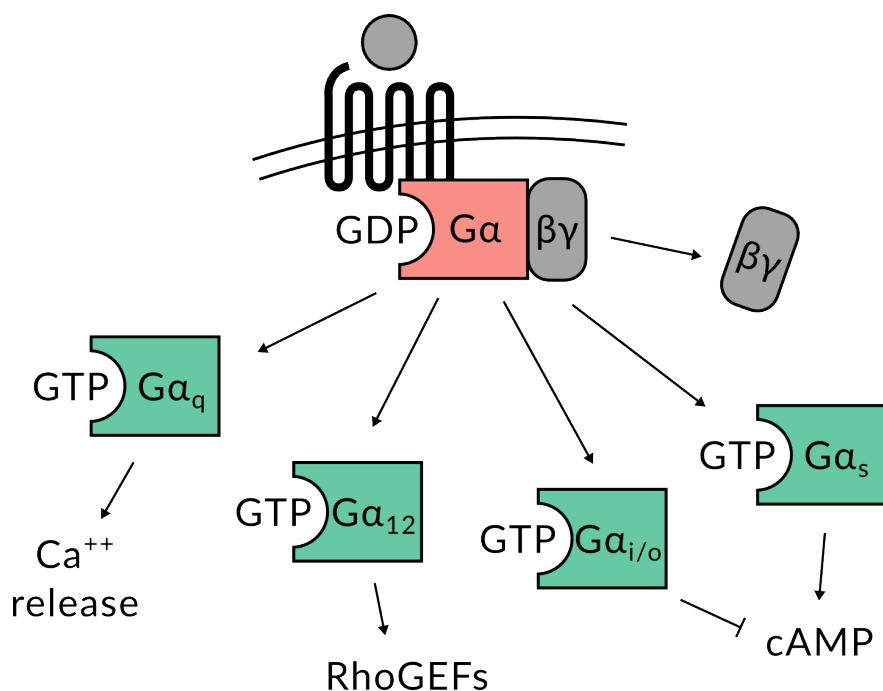


Figure 1-2: Activation of different signaling pathways is mediated by different G-protein subtypes.

be noted that DAG, the other product of phospholipid cleavage by PLC, can also go on to initiate signaling of its own. Finally, active G₁₂ family proteins activate RhoGEFs, nucleotide exchange factors for small Rho family GTPase proteins (as opposed to heterotrimeric G-proteins). These small G-proteins go on to induce phosphorylation cascades and cause changes in cytoskeleton regulation and gene transcription.⁸¹ Although other receptor types exist, heterotrimeric G-proteins and their receptors are a signaling powerhouse capable of initiating a vast array of cellular functions.

The Regulators of G-protein Signaling (RGS) proteins are negative modulators of the G-protein pathway. They bind to active Gα subunit, increasing the rate of GTP hydrolysis. This is known as GTPase-Activating Protein, or GAP, activity. Although the Gα subunit has some intrinsic ability to mediate hydrolysis, it is greatly accelerated when bound to an RGS protein. By returning the alpha subunit to its GDP-bound state, it allows reassociation of the complex

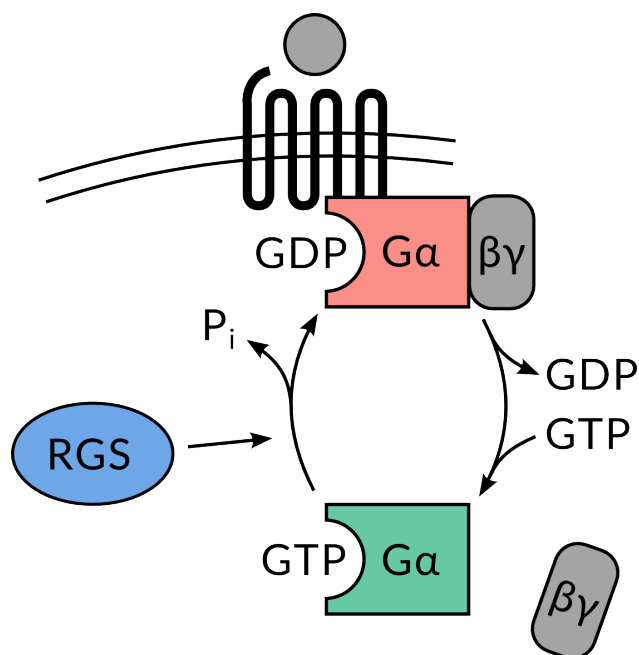


Figure 1-3: RGS are GTPase-Activating Proteins (GAPs). They terminate G-protein signaling by catalyzing hydrolysis of GTP on $G\alpha$.

between the alpha subunit, the beta-gamma subunits, and the receptor. G-protein signaling may be thought of as a cycle, in which the receptor acts as a Guanine nucleotide Exchange Factor (GEF), activating the G protein, and the RGS protein acts as a GAP, deactivating the G-protein (Fig. 1-3). Many drugs exist that manipulate the G-protein cycle from the GPCR side, by positively or negatively altering G-protein activation, but compounds that manipulate G-protein signaling by altering GAP activity remain largely unexplored.

RGS protein diversity

The GAP activity of RGS proteins is carried out by the RGS homology (RH) domain, also referred to as the RGS domain or RGS box. The RGS domain is comprised of about 130 conserved amino acids, which form nine helices.^{82,83} While more proteins with RGS domains may exist, there are twenty that are considered canonical RGS proteins. These RGS isoforms are numbered 1-21, with no RGS15. These in turn are divided into four families: R4, R7, R12, and RZ. R4 is the largest

family, encompassing RGS isoforms 1-5, 8, 13, 16, 18, and 21. These proteins are relatively small, with no additional domains other beyond the RGS domain. However, the N-terminal tail, while not large, may play a role in targeting the RGS protein to certain receptors.⁸⁴ The R4 family act as GAPs for most $G_{i/o}$ and G_q proteins. One notable exception is RGS2, which binds only to G_q and not $G_{i/o}$.⁸⁵

The R7 family consists of RGS6, 7, 9, and 11. These proteins are unique in that they contain two more domains, DEP and GGL, that confer additional functions beyond GAP activity. The DEP domain (Dishevelled, Egl-10, and Pleckstrin; named for the proteins in which it was first identified) is a small domain that may help target the whole protein to specific subcellular locations by binding to membrane anchor proteins.⁸⁶⁻⁸⁸ GGL (G-gamma-like) is a domain that, as its name implies, shares sequence identity and structural features with $G\gamma$. This domain binds to $G\beta\gamma$,⁸⁹ and may help increase GAP activity by colocalizing the RGS protein with $G\alpha_o$.^{90,91}

The R12 family consists of RGS10, RGS12, and RGS14. These proteins are grouped based on their sequence identity, but vary considerably in their number of domains. RGS10 is relatively short, without additional well-characterized domains. RGS12 and RGS14 both have tandem repeats of Ras-binding domains, allowing them to scaffold with small GTPase and MAP kinase proteins, and a G-protein Regulatory (GPR) motif, also known as a GoLoco motif, which bind to $G\alpha_{i/o}$ subunits and act as guanine nucleotide dissociation inhibitors (GDIs).⁹²⁻⁹⁴ RGS12 is the longest full-length RGS protein, at 1387 amino acids, containing in addition a PTB (phosphotyrosine-binding) and PDZ domain, which assists in protein localization.^{93,95}

The RZ family contains RGS17, RGS19, and RGS20. RGS19 is also known as GAIP (G alpha interacting protein). Like most other RGS proteins, they are capable of GAP activity toward G_i and G_o , and are, in many respects, similar to the R4 family. They are relatively small, having no

additional domains other than a cysteine-rich region on the C-terminal tail. However, they are unique in their ability to act as GAPs for $G\alpha_z$, a G protein that is a part of the $G_{i/o}$ family, but not affected by other RGS proteins. This aspect gives the family its name.

In addition to differences in molecular function, RGS isoforms also differ in their tissue distribution,^{96,97} allowing each isoform to play a unique physiological role. A specific RGS inhibitor will act only where its target isoform is expressed, improving tissue specificity beyond that which could be achieved by a GPCR agonist distributed throughout the body. In addition, rather than inducing signaling at the GPCR, use of an RGS inhibitor may prolong endogenous signaling where it is already occurring, further reducing off-target effects.

Physiology of RGS proteins in disease

With RGS proteins playing such diverse roles, it is not surprising that they are involved in the pathogenesis or modulation of many disease states. Since many RGS proteins are highly expressed in the brain,⁹⁶ there is high potential for modulation of RGS proteins in treatment of CNS disorders. For example, RGS4 has been implicated in seizures. Endogenous adenosine has a protective effect on kainate induced seizures, which may be reduced by negative regulation by scaffolding between the A1 receptor, neurabin, and RGS4. Both neurabin knockout and RGS4 inhibition are able to reduce kainate-induced seizures.⁹⁸ RGS4 has also been implicated in reward and addiction. Male RGS4 knockout mice have reduced cocaine induced reward effects compared to WT.⁹⁹ RGS proteins may also have use in treatment of depression. Mice expressing RGS insensitive $G\alpha_i$ show an antidepressant-like phenotype. This is likely mediated by the 5HT-1A receptor, as the effect is blocked by a 5HT1A antagonist.¹⁰⁰ RGS19 is capable of attenuating 5HT-1A receptor signaling, indicating it may be a potential target for treatment of depression.¹⁰¹ RGS19

attenuates mu-opioid signaling, indicating it may play a role in pain modulation as well.¹⁰²

RGS17 is a potential target in treatment of certain cancer types.¹⁰³ One example is in lung cancer: lung cancer susceptibility has been associated with mutations in the first intron of RGS17. RGS17 is heavily upregulated in as many as 80% of lung cancers, while RGS17 knockdown reduced the rate of proliferation in a human lung tumor cell line.¹⁰⁴ Similarly, RGS17 is also upregulated in prostate cancers,¹⁰⁵ which could be another indication for RGS17 inhibition.

One of the most promising applications for targeting an RGS protein is RGS4 in Parkinson's disease. RGS4 is very highly expressed in the striatum,⁹⁷ which is a critical part of the motor control pathway. Motor signals originate in the cortex and are modulated by parts of the basal ganglia, specifically the substantia nigra, the striatum, and the globus pallidus. The striatum receives glutamatergic input from the cortex and dopaminergic input from the substantia nigra. There are two types of spiny projection neurons (SPNs) in the striatum, those that form the direct pathway (dSPN) and those forming the indirect pathway (iSPN). The dSPNs express excitatory G_{olf} coupled (D1-type) dopamine receptors, and project to the internal globus pallidus (GPi). Meanwhile, the iSPNs express inhibitory G_i coupled (D2-type) dopamine receptors, and project to the external globus pallidus (GPe), which in turn projects to the GPi (Fig. 1-4). In the Parkinson's disease state, dopamine-producing neurons in the substantia nigra die, and total dopamine input is reduced. This causes disinhibition of the indirect pathway and reduced excitation of direct pathway. This imbalance between direct and indirect pathways is thought to underlie the motor deficits observed in Parkinson's disease.¹⁰⁶

Synaptic plasticity may play a significant role in regulation of motor control. Long-term depression (LTD) is a form of synaptic plasticity mediated by the release of endocannabinoids (eCBs) such as 2-arachadonylglycerol (2-AG) and anandamide. CB1 cannabinoid receptors are

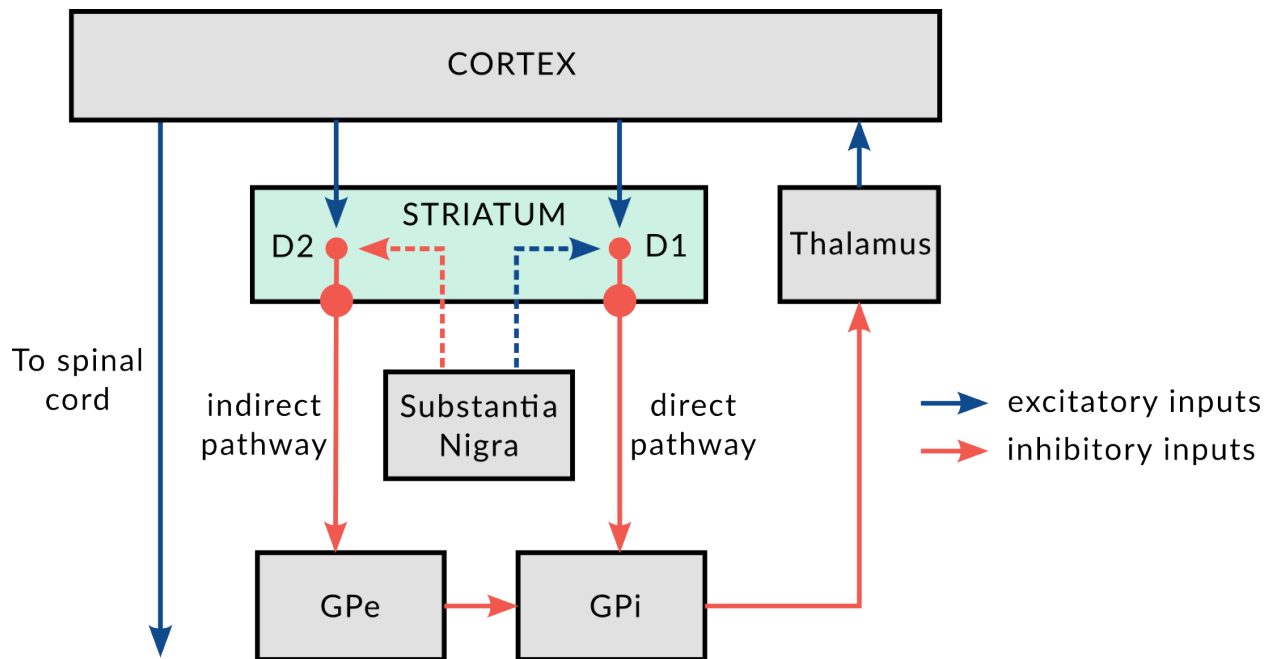


Figure 1-4: The circuitry of the motor pathway. RGS4 is expressed in the striatum. In the Parkinson's disease state, dopaminergic input from the substantia nigra to the striatum is lost.

expressed in glutamatergic nerve terminals projecting to the striatum,¹⁰⁷ and anandamide is released in the striatum upon activation of D2-like but not D1-like receptors.¹⁰⁸ In the Parkinson's disease state, lack of dopamine may reduce endocannabinoid release from iSPNs, thus disinhibiting the indirect pathway and contributing to the imbalance between direct and indirect pathways.

Lerner and Kreitzer (2012) proposed a model in which RGS4 acts as a link between D2 receptor activity and synaptic plasticity in iSPNs.¹⁰⁹ In this model, endocannabinoid release is stimulated in response to glutamatergic signaling through $G\alpha_q$ -coupled mGluR receptors. RGS4 acts as a GAP for $G\alpha_q$, negatively modulating endocannabinoid release. RGS4 activity is enhanced upon phosphorylation by PKA.¹¹⁰ Dopamine D2 receptor activity inhibits cAMP production, thus reducing PKA activity and RGS4 phosphorylation (Fig. 1-5A). In a Parkinson's disease state, lack of dopamine may allow unchecked cAMP and PKA activity, leading to excessive activity of RGS4 and reduced endocannabinoid-mediated long-term depression. Indeed, $RGS4^{-/-}$ mice

have increased LTD over wild-type mice, even in the presence of a D2 antagonist. In addition, in a 6-Hydroxydopamine (6-OHDA) lesion model of Parkinson's disease, RGS4^{-/-} mice were less susceptible than wild-type mice to Parkinson's-like motor deficits.¹⁰⁹

A major problem in treatment of Parkinson's disease is dyskinesia. L-DOPA, a precursor to dopamine, is a standard way of treating motor symptoms in Parkinson's patients. However, a major problem with L-DOPA is that with continued use, its efficacy wanes and dosage is increased, causing dyskinesia. There is a different form of synaptic plasticity that may also play a role in this L-DOPA-induced dyskinesia. While dopamine (and L-DOPA) activity causes LTD in iSPNS, it can induce long-term potentiation (LTP) of dSPN activity. This effect is dependent on D1 receptors; application of D1 antagonists blocks LTP.¹¹¹ Shen et al. (2015) showed that application of an RGS4 inhibitor can induce LTD in dSPNs. Therefore, RGS4 inhibition will not contribute to, and may functionally oppose, the dyskinetic effect of D1 receptor-dependent LTP (Fig. 1-5B),¹¹² which would otherwise be induced upon L-DOPA administration.

RGS4 inhibitors may shine as a combination therapy with L- DOPA. In the indirect pathway, they will complement the effect of L-DOPA in dampening excessive activity. In the direct pathway, however, they may counteract the long-term potentiation responsible for L-DOPA induced dyskinesia.

RGS inhibitors

Recent discovery efforts

In light of this developing rationale for targeting RGS proteins, there have been efforts to develop inhibitors. Roman et al., 2007 performed a high throughput screening campaign run to

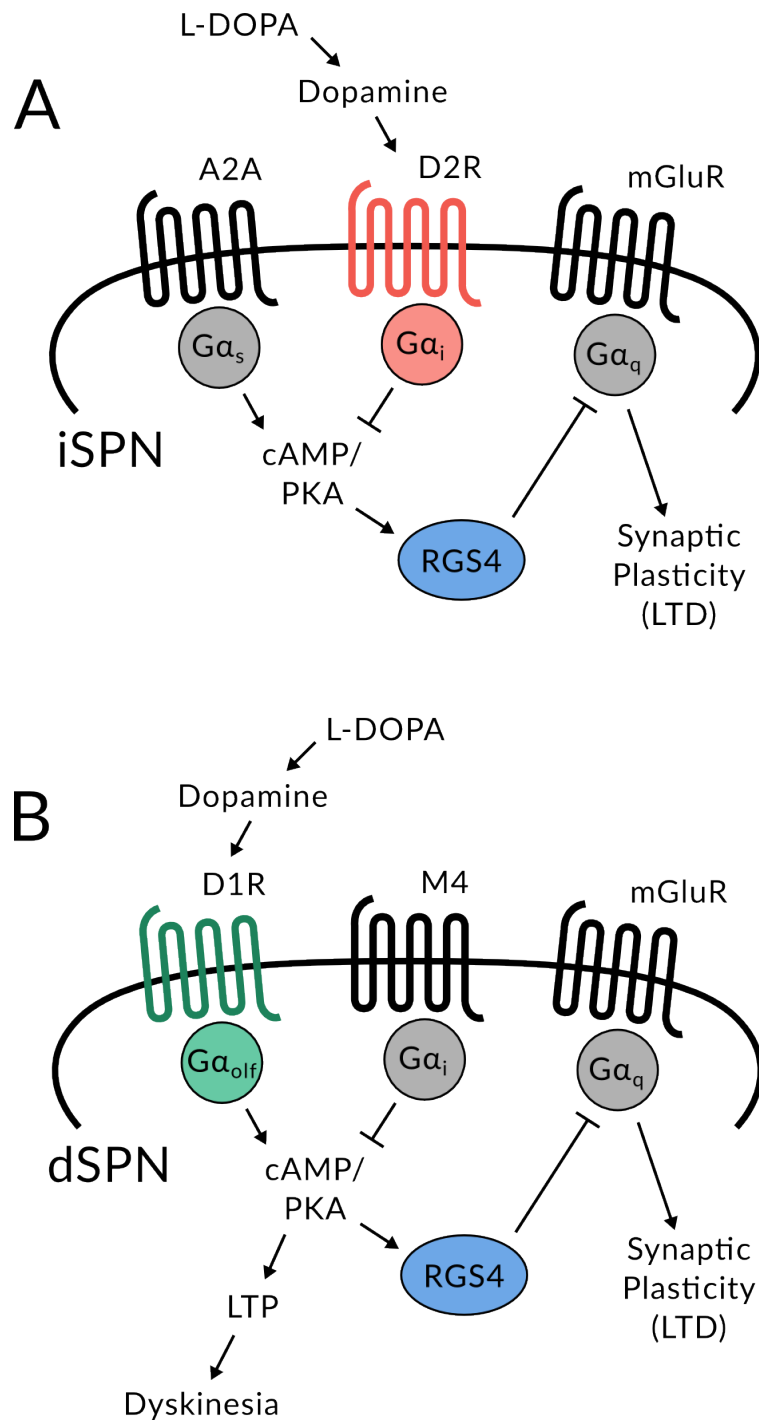


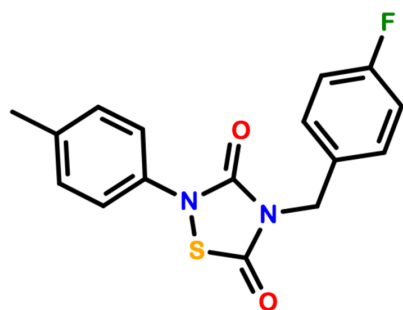
Figure 1-5: The role of RGS4 in response to dopamine signaling in the indirect and direct pathway spiny projection neurons of the striatum.

discover inhibitors of the RGS/G α interaction. This led to the discovery of CCG-4986, the first reported RGS inhibitor, which has specificity for RGS4 over RGS8.¹¹³ Further investigation revealed that this compound acted by covalent modification of cysteine residues. Interestingly, some cysteines are not located near the RGS/G α interface, but binding at these cysteines is still sufficient for inhibition of binding between RGS and G α . This indicates that these compounds inhibit the protein-protein interaction allosterically.³¹ A later screen, multiplexed to determine effects on different RGS proteins, led to the discovery of CCG-50014, the lead compound of the thiadiazolidinones (TDZDs).¹¹⁴ This compound also acts by covalent modification of cysteine residues, and like CCG-4986, is specific for RGS4. To date, no noncovalent RGS inhibitors have been discovered.

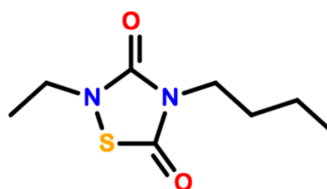
Thiadiazolidinone characterization

RGS inhibitor CCG-50014 has been fairly well characterized in its action against RGS8. Blazer et al., 2011, showed that at least one cysteine is necessary for the compound to inhibit the RGS-G α interaction, and an adduct between the compound and the protein can be detected by mass spectrometry. Thus, it is well established that the thiadiazolidinone inhibition is mediated by covalent modification.¹¹⁵ Interestingly, it was also shown that general cysteine modifiers such as iodoacetamide and n-ethyl maleimide act far less potently on RGS proteins than the thiadiazolidinones. Further, on a cysteine-dependent protease, general cysteine alkylators acted far more potently than CCG-50014.¹¹⁵ These results suggest that the interaction between CCG-50014 and RGS proteins is unique, and lends strength to the concept that a cysteine modifier is capable of acting specifically, without indiscriminate adducts at other proteins.

Because of the reduced potential for adverse effects in with a highly isoform-specific inhibitor, a further effort has been made to develop thiadiazolidinones with improved specificity



CCG-50014



CCG-203769

Figure 1-6: Thiadiazolidinones CCG-50014, the lead compound, and CCG-203769, an analog with improved specificity for RGS4.

for RGS4 over other isoforms. This resulted in the discovery of another thadiazolidinone, CCG-203769, which has aliphatic chains in place of the aromatic rings found in CCG-50014. It is more RGS4 selective than CCG-50014 because it is less potent against RGS8.¹¹⁶

Because CCG-203769 is more selective for RGS4 than other TDZDs, it may have use as a treatment for disease states in which reduction of RGS4 activity is desirable, most notably, Parkinson's disease. Blazer et al., 2015 demonstrated that this compound has *in vivo* effects on motor coordination. In this study, Parkinson's-like bradykinesia (slowness of movement) was induced using the D2-type receptor antagonist raclopride. Two tests were used to analyze reversal of this impairment by CCG-203769: the drag test, which counts steps taken by a mouse as drawn backward by the tail, and the bar test, in which mice were evaluated for the latency in removing their forepaws from an elevated block. In each of these, a dose of as low as 0.1 mg/kg CCG-203769 was sufficient to induce reversal of the bradykinetic effect of raclopride.¹¹⁷

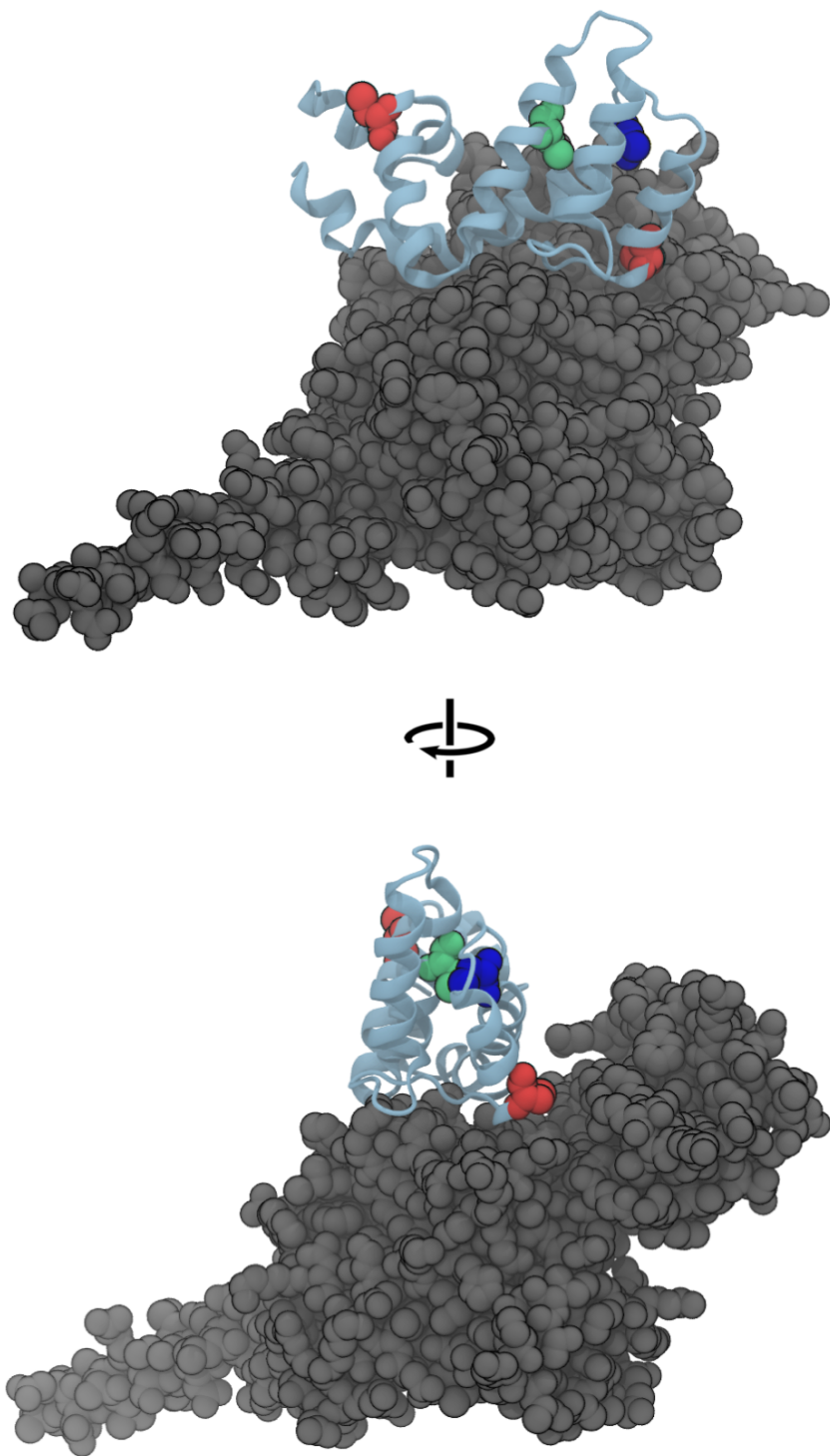


Figure 1-7: Locations of cysteines in RGS protein. $G\alpha$ is shown in gray spheres, RGS is shown in light blue. Cysteines 71 and 132 in RGS4 (red) are not conserved among RGS proteins. Cysteine 148 in RGS4 (blue) is shared by RGS8 and RGS4. Cysteine 95 in RGS4 (green) is the best conserved cysteine among RGS proteins, found in all isoforms except RGS6 and RGS7.

Contribution of this work

While several RGS inhibitors have now been discovered, all are most potent for either RGS4 or RGS1.¹¹⁸ This is not too surprising, given that these isoforms contain relatively high numbers of cysteines in the RGS domain compared to other isoforms, with four and three cysteines respectively. In light of this, it is likely that the RGS4 or RGS1 selectivity is driven primarily by cysteine complement. Interestingly, however, TDZDs can also act on many RGS proteins, even those with only one cysteine. For example, RGS19 contains one cysteine, but CCG-50014 can still inhibit its interaction with G α with an IC₅₀ of 120 nM.¹¹⁵ This suggests that the factors influencing TDZD selectivity are more complex than the simple quantity of cysteines in the RGS domain. The cysteine found in RGS19 is well conserved, found in 18 out of the 20 canonical RGS proteins. From its position on the α_4 helix, it angles toward the center of the α_4 - α_7 helical bundle, so it is buried rather than at the protein surface (green spheres in Fig. 1-7). In order for a covalent inhibitor to access this cysteine, it may be necessary for the protein to undergo a motion exposing this cysteine to the solvent. This implies that protein dynamics is an important yet unexplored factor in RGS inhibitor specificity.

This work aims to develop a better understanding of other factors that drive TDZD selectivity, especially at a structural and dynamic level, which will enable the development of noncovalent inhibitors.

CHAPTER 2:

**Differential Protein Dynamics of Regulators of G-Protein Signaling: Role in Specificity
of Small-Molecule Inhibitors**

Reprinted with permission from *J. Am. Chem. Soc.* 2018, 140, 3454-3460
Copyright 2018 American Chemical Society

Vincent S Shaw*, Hossein Mohammadiarani*, Harish Vashisth, Richard R Neubig

*Co-first authors

V.S. expressed protein and performed HDX-MS and FCPIA. H.M. performed MD simulations and calculated RMSF, RMSD, and SASA.

Introduction

Protein-protein interactions (PPIs) remain a poorly tapped pool of potential targets for small-molecule inhibitors. Targeting PPIs has been challenging because many protein-protein interfaces are flat and lack a dedicated small-molecule binding pocket.^{23,119,120} However, it may be possible to interrupt PPIs by binding to transiently exposed pockets,^{121,122} either at the protein-protein interface⁷ or at allosteric sites.^{32,123} Targeting of allosteric sites, as they are less evolutionarily conserved, may confer better specificity than directly targeting interfaces.¹²⁴ In addition, there may be variation in dynamic exposure of allosteric pockets among members of a protein family. Such differences in protein dynamics could drive inhibitor specificity.¹²⁵

G-protein signaling is critical in pharmacology. Approximately thirty percent of marketed drugs target GPCRs and related pathways.¹² Regulators of G-protein Signaling (RGS) proteins control GPCR signaling by binding to active, GTP-bound $G\alpha$ subunits, thereby accelerating GTP hydrolysis. This terminates G-protein signaling. Inhibition of an RGS protein can amplify signaling through a GPCR. We previously identified thiadiazolidinone (TDZD) inhibitors of the RGS- $G\alpha$ interaction in a high-throughput screen.¹¹⁶ They allosterically inhibit RGS proteins by covalent modification of cysteine residues at sites distant from the RGS- $G\alpha$ interface. The TDZD inhibitor CCG-50014 is most potent against RGS4, followed by RGS19 and distantly by RGS8.¹¹⁵ RGS4 inhibitors may be valuable as therapeutics for Parkinson's disease. RGS4 is highly expressed in the striatum,^{96,97} where it regulates synaptic plasticity in response to dopamine signaling.^{109,112} A TDZD inhibitor with enhanced specificity for RGS4, CCG-203769, reduces bradykinesia in a raclopride model of certain Parkinson's-like motor deficits in mice.¹¹⁷

The RGS domain, which is responsible for the GTP-ase accelerating activity of RGS proteins, is present in 20 human RGS proteins as well as some proteins with a similar fold that lack

Gα binding properties.¹²⁶ The RGS domain is a 120-amino acid domain consisting of nine α-helices (Fig. 2-2A).^{126,127} Differences in TDZD potency may be due to different locations or numbers of cysteines among RGS isoforms or to differential transient cysteine exposure. RGS4, RGS8, and RGS19 all share an α₄ helix cysteine, while RGS4 and RGS8 share one on the α₆-α₇ interhelical loop (Fig. 2-2A). Notably, these cysteines are buried beneath the protein surface in crystal structures.^{83,128} Therefore, it may be necessary for dynamic pockets to open to expose these cysteines for TDZD interaction. Understanding dynamic pockets will be beneficial, as such a pocket may be exploited in rational design of novel non-covalent inhibitors using a docking-based virtual screen. We previously showed that the α₅-α₆ helical pair is flexible using enhanced sampling MD simulations.¹²⁹ Covalent modification by TDZD inhibitors could lock the α₅-α₆ interhelical loop in a position that prevents the RGS interaction with Gα proteins. We hypothesize that differential transient exposure of buried cysteine residues drives TDZD selectivity. Here, we used hydrogen/deuterium exchange with mass spectroscopy (HDX-MS) and long time-scale classical unbiased molecular dynamics (MD) studies to examine differences in dynamics between RGS4, RGS8, and RGS19. These RGS protein isoforms represent a range of potencies of TDZD inhibitors (RGS4 > RGS19 > RGS8). HDX-MS and MD studies make a powerful combined experimental and computational approach for evaluating protein dynamics.^{130,131} These revealed a dual role of protein dynamics and cysteine complement in the selectivity of TDZDs against RGS proteins.

Materials and Methods

Protein expression and purification

N-terminally truncated (Δ51) rat RGS4 with 6xHis tag in pET23d vector, RGS homology

domain of human RGS8 (42-173) with 6xHis tag in pQE80 vector, RGS homology domain of human RGS19 (89-206) with 6xHis tag in pET15b vector, and $G\alpha_o$ with 6xHis tag in pQE-6 vector¹³² were individually transformed into BL21(DE3) E. coli. Single-cysteine mutant RGS protein constructs were generated by mutating cysteines to alanines using QuikChange II mutagenesis kit (Agilent). Protein expression was induced by addition of 200 μ M IPTG. Expression was carried out overnight at 25°C and cells were harvested by centrifugation. Pellets were resuspended in 50 mM HEPES 100 mM NaCl pH7.4 and lysed by sonication. Lysate was centrifuged, supernatants were applied to nickel affinity column, and protein was eluted with 300 mM imidazole. RGS4 was further purified by SP sepharose column. Column was equilibrated with 50 mM Na Acetate, 40 mM NaCl, 1 mM DTT, 1 mM EDTA, and 1mM EGTA (pH 5.5) and protein eluted by linear gradient to buffer including 1M NaCl. RGS8 and RGS19 were purified by Q sepharose column. Column was equilibrated with 20 mM NaCl, 50 mM Tris, and 1 mM DTT (pH 8.0), and protein eluted with linear gradient to buffer including 1M NaCl.

Flow cytometry protein interaction assay

FCPIA was performed as previously described.¹³³ Briefly, RGS proteins were biotinylated and bound to xMAP LumAvidin microspheres (Luminex). $G\alpha_o$ protein labeled with AlexaFluor-532 was exposed to beads in presence of GDP and aluminum fluoride to stabilize the transition state. Bead fluorescence was read using Luminex 200 flow cytometer.

Hydrogen/deuterium exchange

HDX-MS was performed as described in Chodavarapu et al., 2015.¹³⁴ In principle, after exposure to D₂O for different times, the exchange of amide hydrogens for deuterium was quenched by acidification then samples were digested with pepsin and separated by LC-MS for analysis

of deuterium content. Specifically, proteins were incubated on ice for desired time in 90% D₂O containing 100 mM NaCl and 5 mM HEPES, pH 7.4. All columns and valves were kept on ice to reduce back exchange. Exchange was quenched by 1:1 addition of ice cold 1% (v/v) formic acid in H₂O, bringing the pH to 2.5. 100 µl samples were immediately loaded at 0.1 ml/min, using external pump (LC-20AD; Shimadzu), to an Enzymate pepsin column (2.1 x 30 mm, Waters) equilibrated with cold 0.1% formic acid in H₂O. After 1 min, the pump was stopped and proteins were digested on-column for 1 min (See pepsin cleavage pattern in Fig. 2-6). Following digestion, the resulting peptides were eluted at 0.5 ml/min onto an Xbridge BEH C18 VanGuard trap column (2.1 x 5 mm, Waters). The peptides were then eluted from the trap column by valve switching of liquid flow, using a Waters 2777c autosampler, onto an Ascentis Express Peptide ES-C18 column (2.1 x 50 mm, Supelco). Flow through the 2777c autosampler valve and the Peptide ES-C18 column was controlled by a Waters Acquity Binary Solvent Manager. The peptides were initially washed for 1 min at 0.3 ml/min with 99% solvent A (0.1% formic acid in H₂O) and 1% solvent B (acetonitrile). Peptides were then separated by elution with a gradient from 1% B to 30% B at 3 min, then to 99 % B at 6 min and held at 99% B for 1 min. Eluted peptides were analyzed using a Xevo G2-XS QToF mass spectrometer (Waters) by electrospray ionization operating in positive-ion mode. Fragments observed following cleavage of RGS proteins by pepsin are shown in Fig. 2-1. Mass spectra were acquired in continuum mode over m/z 100-2000. Data were analyzed using Microsoft Excel, HX Express,¹³⁵ and GraphPad Prism software. Deuterium incorporation was determined by the increase in centroid mass of each peptide's isotope distribution compared to undeuterated control.

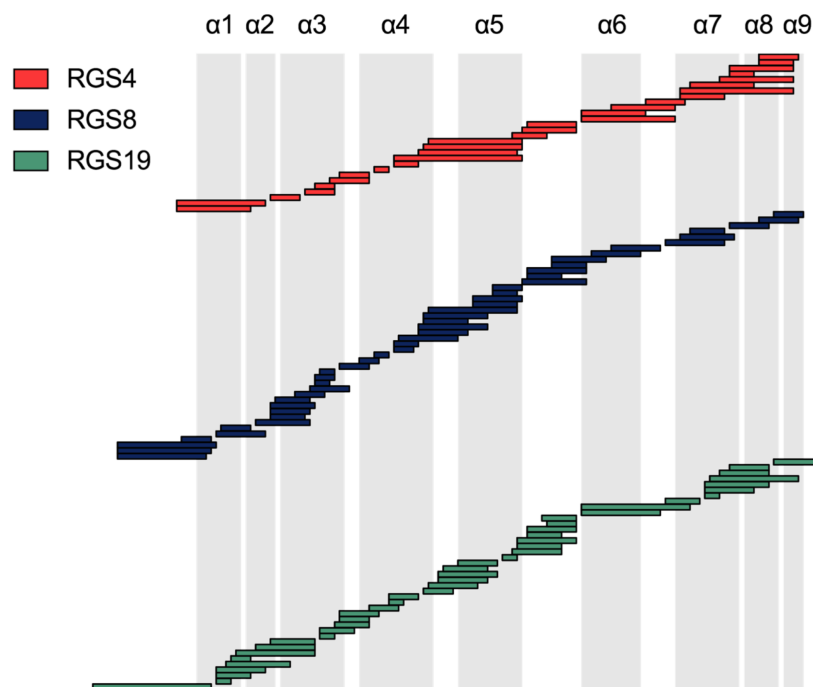


Figure 2-1: Alignment of fragments observed by mass spectrometry following cleavage of RGS proteins by pepsin. Horizontal bars indicate length and position of observed fragments. The two N-terminal residues of each fragment were excluded from analysis due to rapid back-exchange. Vertical gray boxes indicate approximate positions of helices within the RGS domain.

System setup and simulation details

Molecular dynamics (MD) simulations were carried out by the Vashisth Lab at UNH. Trajectory calculations and their analysis were done using NAMD/VMD software suite^{136,137} with the CHARMM force-field and CMAP correction.^{138,139} The initial coordinates for RGS4, RGS8, and RGS19, respectively, were taken from the protein data bank entries 1AGR, 2ODE, and 1CMZ. Each protein was initially modeled using the psfgen tool in VMD and further solvated in a simulation box ($\sim 65\text{\AA} \times \sim 70\text{\AA} \times \sim 65\text{\AA}$) of TIP3P water molecules and charge-neutralized with NaCl. The final solvated and ionized simulation domains contained 28160 atoms (RGS4), 30731 atoms (RGS8), and 29560 atoms (RGS19), respectively. The box volume was then optimized in the NPT ensemble by initially applying 500 cycles of a conjugate-gradient minimization scheme followed by a short 40 ps MD run with a 2 fs time step in which the temperature was controlled using the Langevin thermostat and the pressure was controlled by the Nose-Hoover barostat. We carried out all simulations using periodic boundary conditions. These briefly equilibrated systems of all RGS proteins were further subjected to two independent MD simulation sets in the NVT ensemble. The first set (Set 1 in Table 2-1) of simulations were 2 μs -long for each RGS protein, and the second set of simulations (Set 2 in Table 2-1) were 3 μs -long for each RGS protein. Results from Set 1 are discussed in Fig. 2-5 and 2-7, and from Set 2 are shown in Fig. 2-7, 2-6, and 2-9. Computations were performed on Trillian, a Cray XE6m-200 Supercomputer and using Premise, a UNH in-house GPU based cluster. In addition, this work used the Extreme Science and Engineering Discovery Environment (XSEDE).¹⁴⁰

RMSD, RMSF, and SASA Measurements

We carried out the analyses on per-residue root-mean-squared-fluctuation (RMSF) and

root-mean-squared-deviation (RMSD), as reported in Fig. 2-4 and 2-5, by aligning each frame of MD trajectories to the initial frame based upon all C α atoms of each protein. The solvent-accessible surface area (SASA) of sulfur atoms in cysteines were calculated using a probe radius of 1.4 Å.

Results

Previous work has demonstrated a role for the number and position of cysteine residues in the potency of RGS inhibitors.³¹ To eliminate this confounding variable and allow better assessment of the role of protein dynamics, the potency of CCG-50014 was compared among RGS19 and mutants of RGS4 and RGS8 containing only the shared α_4 cysteine. These mutants are termed RGS4 95C and RGS8 107C respectively. While removal of additional cysteines reduced potency in both RGS4 and RGS8, dramatic differences in TDZD potency still exist among single-cysteine proteins. RGS19 was most potently inhibited with an IC₅₀ of 1.1 μ M, while RGS4 95C was inhibited with an IC₅₀ of 8.5 μ M, and RGS8 107C had an IC₅₀ of >100 μ M (Fig. 2-2B).

To compare solvent exposure kinetics on the α_4 helix, we performed HDX-MS on RGS4, RGS8, and RGS19 apo-proteins. A map of pepsin cleavage fragments observed in each protein is shown in Fig. 2-1. Consistent with the higher potency of inhibition by the TDZD, the cysteine-containing fragment from α_4 (residues 92-97) in RGS4 shows significantly higher exchange than that from RGS8 (residues 86-91). After a 1000-minute incubation in D₂O, the 92-97 fragment of RGS4 had 35% deuterium incorporation (DI), while the analogous fragment in RGS8 had only 8% DI. Further strengthening the correlation of dynamics with selectivity, RGS19 had much faster exchange than RGS4 or RGS8 in the α_4 helix. It reached 48% DI by only 100 minutes, while RGS4 and RGS8 had 9% and 1% respectively (Fig. 2-3A). A similar trend was observed in the α_5 helix.

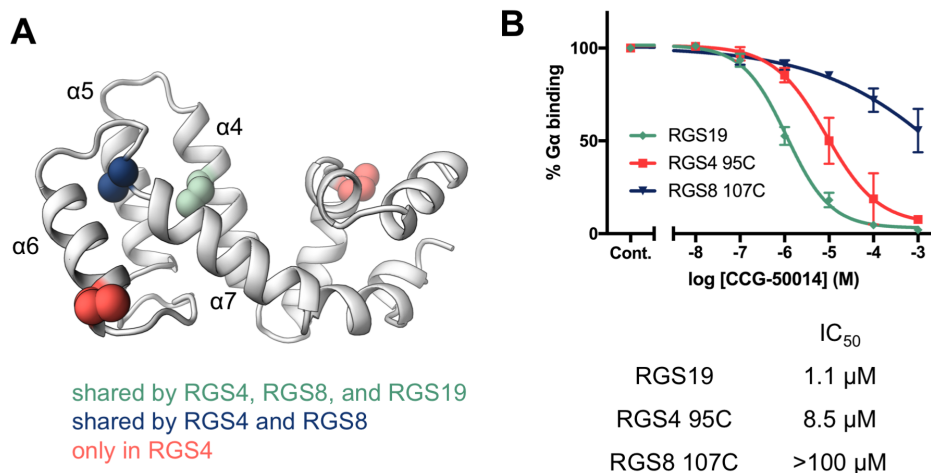


Figure 2-2: (A) Locations of cysteines in RGS4, RGS8, and RGS19. (B) Potency of CCG-50014 against RGS19, which has only one cysteine, and mutant RGS4 and RGS8 containing only the shared α_4 helix cysteine. $n=3$.

RGS8 had the least exchange after 1000 minutes (24% DI), followed by RGS4 and RGS19 (38% and 49% DI, respectively, Fig. 2-3B). One pattern consistent among all three isoforms is high exchange in the α_5 - α_6 interhelical loop, indicating that RGS proteins are flexible in this region. Those fragments in all three proteins exceeded 50% DI by 100 minutes (Fig. 2-3C). This was not surprising, as the α_5 - α_6 loop is the longest unstructured region within the RGS domain. In the α_6 helix, RGS19 again had higher exchange than RGS8 and RGS4. RGS8 was particularly protected in the residue 126-136 fragment, reaching only 7% DI after 1000 minutes. However, higher exchange was observed in the residue 130-140 fragment of RGS8, likely because this fragment also contains residues that are a part of the α_6 - α_7 loop (Fig. 2-3D). A similar effect was seen in RGS4 near the α_7 helix, in which a fragment wholly within α_7 (residues 150-159) had much slower exchange than a fragment partially overlapping the α_6 - α_7 loop (residues 143-151) (Fig. 2-3E).

According to these results, RGS8 had low deuterium exchange relative to other RGS proteins throughout the helices surrounding its cysteines. This is indicative of rigidity of these he-

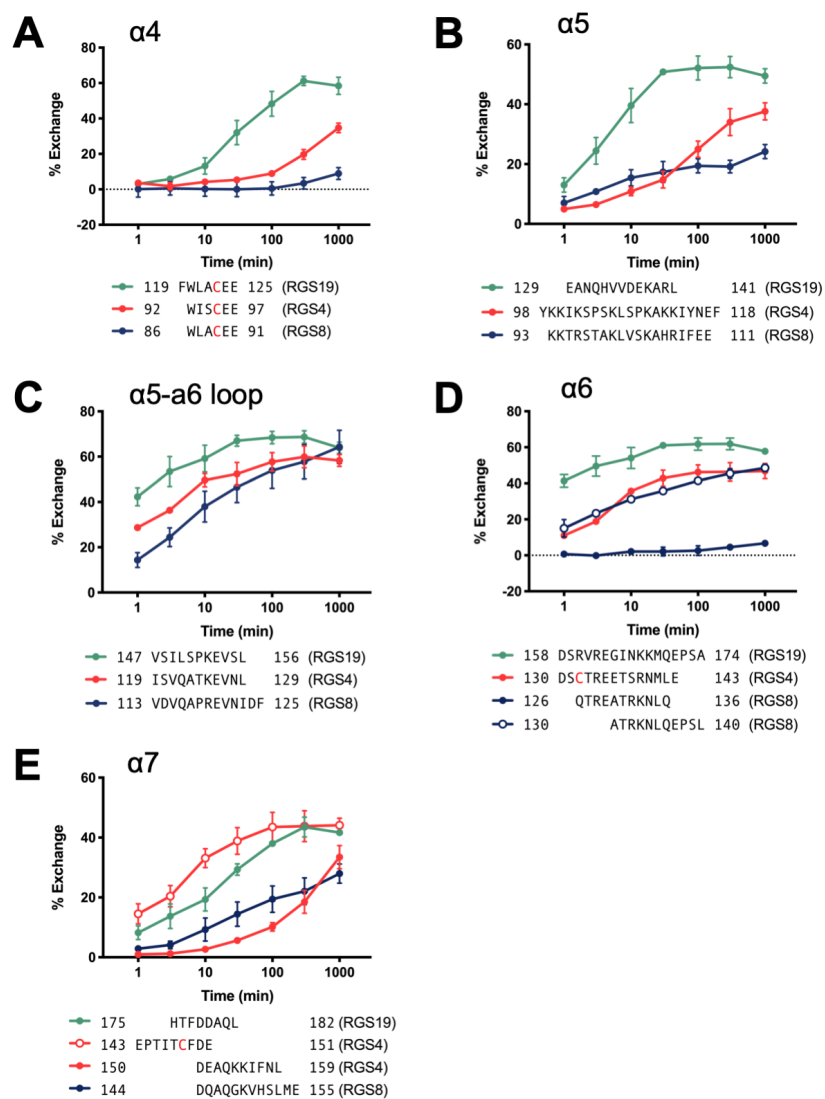


Figure 2-3: (A-E) Kinetics of deuterium exchange in selected protein fragments from (A) α_4 , (B) α_5 , (C) α_5 - α_6 interhelical region, (D) α_6 and (E) α_7 . Sequences of observed fragments are aligned with residue numbers of each fragment indicated. Cysteine locations are marked in red. n=3.

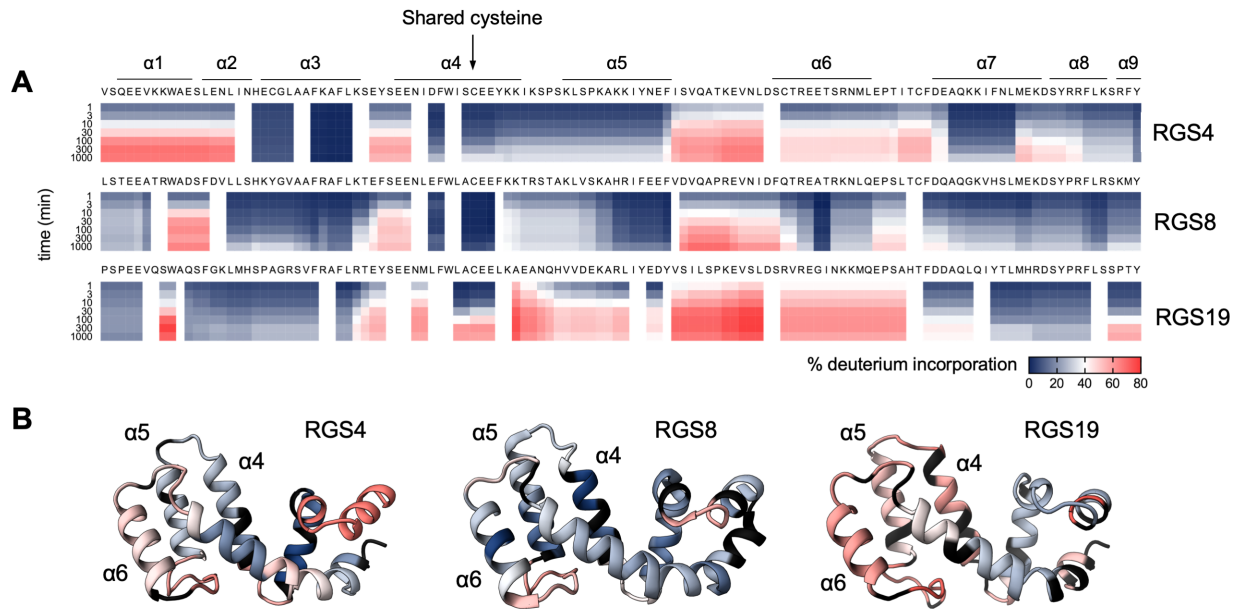


Figure 2-4: (A) Global kinetics of deuterium exchange. Deuterium incorporation (DI) is expressed as a percent of exchangeable amide hydrogen positions. Where fragments overlap, data is displayed as average DI of observed fragments. (B) Degree of DI at 300 minutes in 90% D_2O is mapped onto protein structure of RGS4, RGS8, and RGS19. $n=3$.

lices in RGS8, which likely prevents exposure of cysteines to solvent. This observation also could explain the low potency of TDZDs against RGS8 relative to other RGS isoforms. The α_6 helix of RGS4 has more deuterium exchange than the α_4 , α_5 , and α_7 helices (Fig. 2-4A and B). Rapid exchange in the α_6 helix may be due to movement away from neighboring helices or unfolding of the helix itself. Such a movement could increase solvent exposure of the otherwise buried cysteine 148 on the α_6 - α_7 loop. This would allow access by TDZD inhibitors. Because the higher exchange on α_6 compared to other nearby helices is unique to RGS4, this potentially contributes to the increased potency of TDZDs against wild type RGS4 versus RGS8. In the α_4 , α_5 , and α_6 helices, RGS19 shows higher deuterium exchange than RGS4 or RGS8, indicating that RGS19 is highly dynamic. For example, in a fragment of the α_5 helix, RGS19 had 51% DI after 30 minutes, while similar fragments in RGS4 and RGS8 had 15% and 17% incorporation, respectively (Fig. 2-

Protein	initial coordinates	# of atoms	trajectory length	
			set1	set 2
RGS4	PDB: 1AGR	28160	2 μ s	3 μ s
RGS8	PDB: 2ODE	30731	2 μ s	3 μ s
RGS19	PDB: 1CMZ	29560	2 μ s	3 μ s

Table 2-1: Summary of MD simulations.

3B). This fits with functional data showing that RGS19 is more potently inhibited by CCG-50014 than single-cysteine RGS4 and RGS8 (2-2B). Although RGS19 lacks cysteines on the α_6 helix and α_6 - α_7 loop which may contribute to potency of inhibition of RGS4 by TDZDs (Cys 132 and Cys 148 in RGS4), it has the highest potency of inhibition among single-cysteine RGS proteins. This may be due to a pronounced movement of the α_4 , α_5 , and α_7 helices, allowing TDZDs to access RGS19's cysteine on the α_4 helix.

To probe the molecular details of dynamic motions in RGS4, RGS8, and RGS19 that underlie the flexibility differences observed in HDX-MS as well as to evaluate possible routes of access to cysteines by TDZDs, we performed long time-scale classical MD simulations in explicit-solvent. Our previous short time-scale classical MD simulations did not show any major conformational changes; but enhanced sampling simulations did show changes.¹²⁹ Here, we conducted microsecond time-scale classical MD simulations through which the flexibility in key helices became apparent.

The first set of simulations that were 2 μ s long (set 1 in Table 2-1) showed regions of pronounced movement in all three proteins. RGS4 showed unique motions within the α_6 helix (Fig. 2-5A), while in RGS8 and RGS19, movement was primarily within the α_6 - α_7 interhelical loop (Fig. 2-5B and C). A second independent set of simulations that were 3 μ s long (set 2 in Table 2-1) showed the largest movement in RGS19, again particularly prominent in the α_6 - α_7 interhelical loop, with the α_6 helix and α_5 - α_6 interhelical loop also relatively flexible (Fig 2-6). However, RGS4

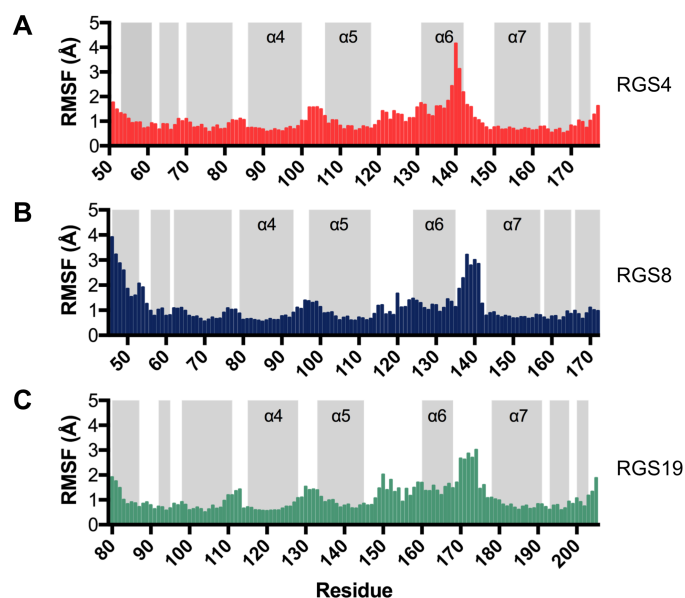


Figure 2-5: Root mean squared fluctuations (RMSF) per residue during 2 μ s MD simulations of (A) RGS4 (PDB: 1AGR), (B) RGS8 (PDB: 2ODE), and (C) RGS19 (PDB: 1CMZ). The RMSF trends for each protein for the simulation set 2 are shown in Fig. 2-6. Gray bars indicate helical regions.

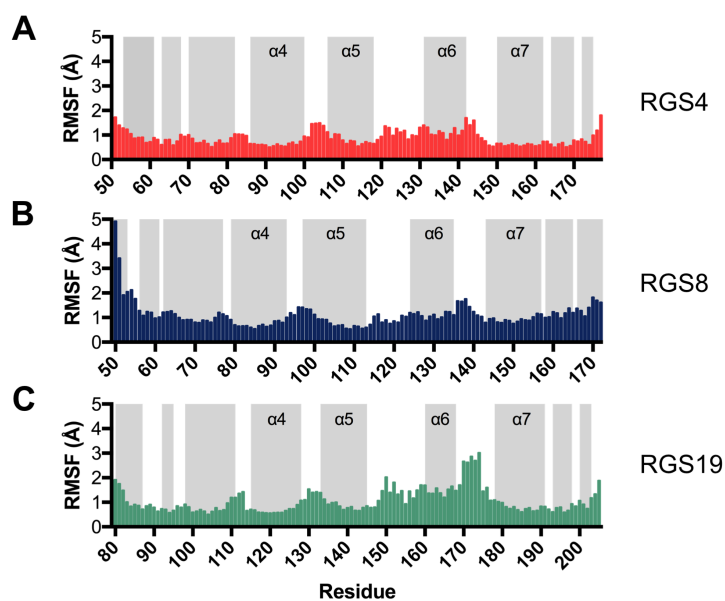


Figure 2-6: Root mean squared fluctuations across protein sequence during 3 μ s MD simulations of (A) RGS4 (PDB: 1AGR), (B) RGS8 (PDB: 2ODE), and (C) RGS19 (PDB: 1CMZ). Gray bars indicate helical regions.

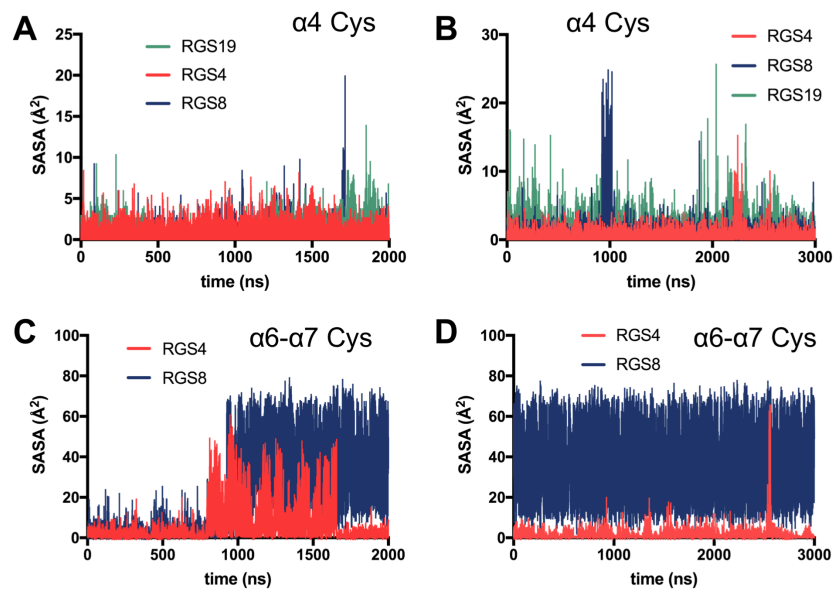


Figure 2-7: Solvent-accessible surface areas (SASA) are shown for sulfur atoms in shared cysteines on α_4 helix for simulation set 1 (A) and set 2 (B) in RGS4, RGS8, and RGS19, and for shared cysteines on α_6 - α_7 interhelical loop in simulation set 1 (C) and set 2 (D) in RGS4 and RGS8.

and RGS8 were relatively stable; simulation set 2 did not recapture the α_6 helix movement in RGS4 observed in simulation set 1. This illustrates a limitation of MD simulations in observation of movements that occur infrequently or on long time scales. Taken together, these simulation sets indicate highest flexibility in RGS19, with potential for flexibility in distinct regions in RGS8 and RGS4. In all simulations for each protein, pronounced movements also occurred in the residues located in terminal helices. This is likely an effect of free terminal ends; residues outside of the RGS homology domains were not included in the simulations.

Analysis of solvent exposure of sulfur atoms reveals exposure of initially buried cysteines. (Fig. 2-7). Cys123 in RGS19 is more exposed than analogous cysteines in RGS4 and RGS8 in the 2 μ s simulation set (Fig. 2-7A) and again in the 3 μ s simulation set (Fig. 2-7B). This may explain the potency of RGS19 relative to the analogous single-cysteine RGS4 and RGS8. Pronounced exposure of the α_6 - α_7 interhelical Cys160 in RGS8 was observed in both sets of simulations (Fig.

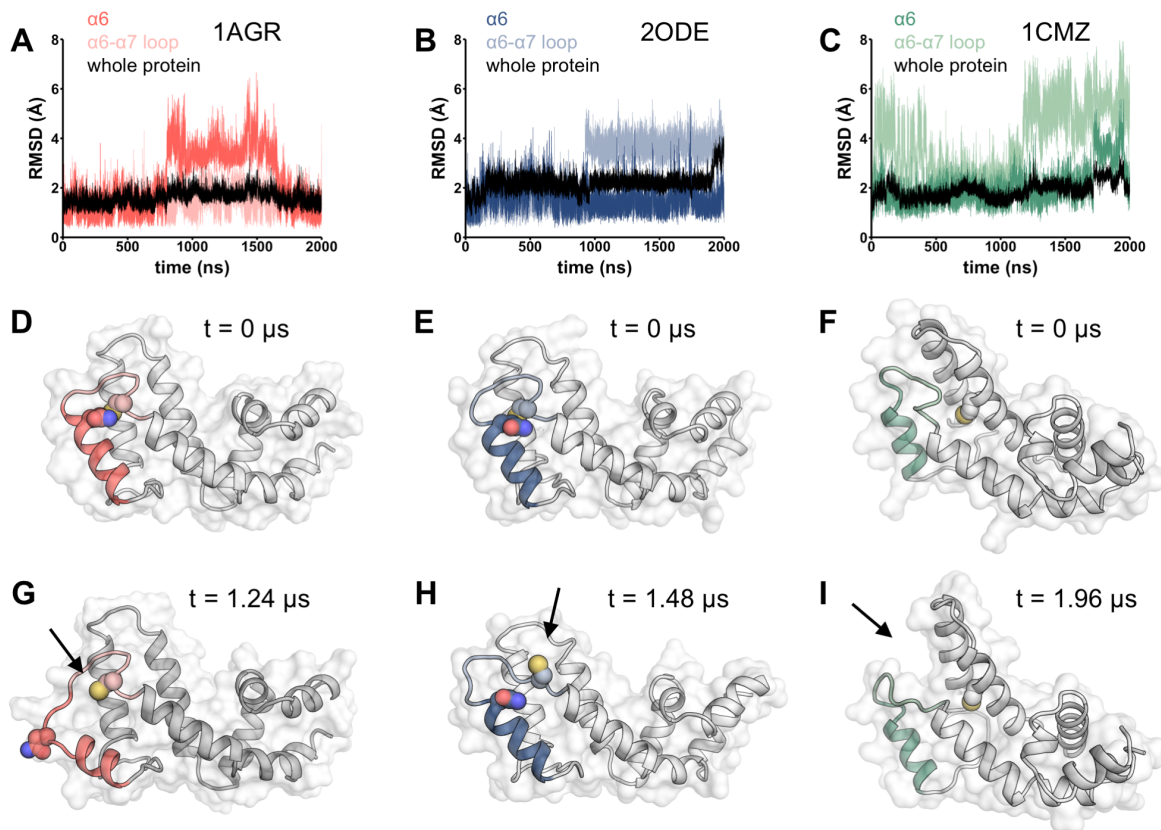


Figure 2-8: Conformational changes during molecular dynamics simulations. Root mean square deviations of α_6 helix and α_6 - α_7 loop, starting conformation, and a snapshot conformation during MD simulation are shown for (A, D, G) RGS4, (B, E, H) RGS8, and (C, F, I) RGS19. Protein regions plotted in MD trajectories are depicted in color in protein structures. Arrows indicate locations of notable solvent exposure during simulation.

2-7C and D).

In addition, the conformations observed during movements of the α_6 helix and α_6 - α_7 loop show distinct routes of cysteine exposure among the three RGS proteins. In the RGS4 crystal structure (PDB: 1AGR),⁸³ Asn140 occludes Cys148 from exposure to the protein surface (Fig. 2-8D). In the MD simulation set 1 using 1AGR as initial coordinates, a transient movement of the α_6 helix was observed, reaching 15.1 Å between α -carbons at 1.24 μs (Fig. 2-8G), versus 5.9 Å at baseline. This movement coincided with a high solvent exposure of Cys148 (Fig. 2-7C). In MD simulation set 1 of RGS8 (using PDB code 2ODE¹⁴¹ as initial coordinates), helices α_4 , α_5 , α_6 , and α_7

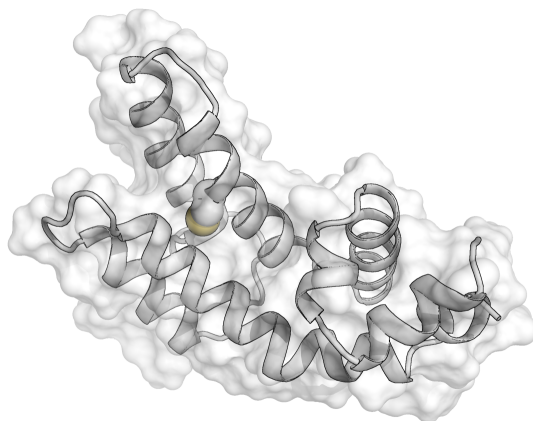


Figure 2-9: Snapshot of RGS19 from simulation set 2 at 240 ns. Cleft opening observed in simulation set 1 (Fig 6I) was recapitulated in this simulation.

were stable relative to the same helices in other proteins tested. However, the α_6 - α_7 interhelical region, which includes cysteine 160, underwent a pronounced movement (Fig. 2-8B). Cys160 rotated toward the protein surface at 1 μ s, and remained exposed to solvent for the remainder of the trajectory (Fig. 2-7B and 2-8H). This cysteine exposure was observed again for the duration of simulation set 2 (Fig. 2-7D). RGS19 lacks the cysteine in the α_6 - α_7 interhelical loop, having only Cys123 on α_4 . Both MD simulations of RGS19 (starting with the PDB code 1CMZ¹⁴²) revealed a movement of the α_6 - α_7 interhelical loop away from the α_4 and α_5 helices, resulting in an open groove in the protein surface (arrow in Fig. 2-8I and 2-9). This observation likely explains the higher observed DI of α_4 and α_5 helices in RGS19 compared to RGS4 and RGS8, but additional changes, perhaps induced by compound binding, may be required for full exposure of Cys123.

Discussion

RGS protein flexibility, as measured both by DI and solvent exposure of the α_4 cysteine in MD simulations, is correlated with the potency of TDZDs to inhibit RGS proteins containing

only a single shared cysteine. RGS19 had the most pronounced DI throughout the α_4 - α_7 helix bundle, and it was more potently inhibited by CCG-50014 than single-cysteine RGS4 or RGS8. Such flexibility could result in increased likelihood of binding of TDZDs at the α_4 cysteine. This would be expected to lead to perturbation of residues involved in G-protein binding, as suggested by previous NMR experiments.¹²⁹

There was also good concordance between regional protein flexibility in the HDX-MS studies and in MD simulations. In RGS8, helices α_4 , α_5 , α_6 , and α_7 were protected from deuterium exchange and were also stable during MD simulations. The dramatic movement of the RGS4 α_6 helix in simulation set 1 mirrors its high solvent exposure in HDX studies. This suggests that movement of the α_6 helix is likely responsible for solvent exposure of Cys 148 in RGS4, providing a plausible route of access by TDZD inhibitors. Indeed, cysteine 148 was the most important single cysteine for inhibition of RGS4 by our other cys-linking inhibitor, CCG-4986.³¹

Deuterium exchange is measured on a much longer timescale than is accessible by MD simulations. In order for exchange to occur, amide hydrogens must be in a conformation amenable to exchange, requiring both interruption of H-bonds and proximity of solvent waters. These exchange-competent states are short lived, often existing on a 10-100 picosecond timescale.¹³¹ They are frequent enough to be readily observed in microsecond timescale simulations; however, the rate of intrinsic hydrogen exchange is much slower than the rate of hydrogen solvent exposure. This is termed EX2 kinetics, in which an amide hydrogen may make multiple visits to a solvent-exposed state before an exchange event occurs.¹⁴³ While exchange is still representative of the time spent in an open state, this allows observation of exchange on much longer timescales than those of dynamic motions.

Interestingly, dynamic cysteine exposure varied among protein isoforms. In RGS4, move-

ment of helix 6 exposed the α_6 - α_7 cysteine, while in RGS8, helix 6 was stable and that cysteine rotated toward solvent in during a movement of the α_6 - α_7 loop. RGS19 lacks a cysteine on the α_6 - α_7 loop, but opens a cleft toward a deeply buried α_4 helix cysteine. These results suggest that the route of modification by covalent inhibitors varies among RGS isoforms, even at shared cysteine locations.

These differences in dynamic motions among RGS isoforms may contribute to differences in potency of TDZD inhibition by two ways. First, differences in the rate of covalent modification or the magnitude of effect on $G\alpha$ binding may be driven by differences between RGS isoforms in the direction of cysteine solvent exposure. Second, distinct transient conformations occurring more frequently in certain RGS isoforms may permit unique non-covalent docking to drive covalent modification. In such a scenario, the open state could be taken advantage of in a docking-based virtual screen, permitting the discovery of non-covalent RGS inhibitors. Although additional future work is required to fully understand the inhibitor access routes and mechanisms (e.g. conformational selection versus induced fit), we have previously shown¹²⁹ using nuclear magnetic resonance (NMR) and MD simulation analyses that an open conformation of RGS4 facilitates covalent docking of CCG-50014 and leads to significant perturbations in residues near the binding pocket and at the protein-protein interface. This is because inhibitor binding only allows a partial recovery of the open conformation to an apo-like conformation as opposed to a nearly complete recovery in the absence of the inhibitor. Because conformational changes induced by compound binding may be a factor in inhibition, we aim to undertake studies involving docking of other TDZD and non-TDZD analogs¹¹⁶ using conformations of RGS proteins reported in this work. These possibilities remain an object of future investigations.

Conclusions

The application of HDX-MS and MD methods reveal that RGS isoforms differ in their mechanism of transient cysteine exposure, suggesting distinct routes of access by covalent inhibitors. These differences are potentially responsible for the selective potency of TDZD inhibitors among RGS isoforms. Importantly, the conformations of RGS proteins in which cysteine residues are transiently exposed could be potentially useful for designing the next generation of inhibitory small-molecules.

CHAPTER 3:

An Interhelical Salt Bridge Controls Flexibility and Inhibitor Potency For Regulators of G-protein Signaling (RGS) Proteins 4, 8, and 19

Vincent Shaw performed protein expression, DSF, HDX, and FCPIA. Mohammadjavad Mohammadi contributed MD simulations and RMSF, RMSD, and DCC analyses. Josiah Quinn assisted with mutagenesis and protein expression and performed DSF.

Introduction

Drug specificity is often considered to be like a key fitting into a complementary shaped lock. It has become clear recently that protein dynamics can play an important role in drug discovery.³ Regulators of G-protein Signaling (RGS) proteins bind to activated G α subunits of G-proteins, thereby accelerating GTP hydrolysis and attenuating G-protein signaling. In regulating GPCR signaling, RGS proteins play a role in the physiology of numerous systems. By inhibiting RGS proteins, signaling via a GPCR may be enhanced. There are twenty RGS isoforms, each with a different tissue distribution. Combination of GPCR agonists with inhibitors specific for a single RGS isoform should limit effects on GPCR signaling to the subset of target tissues with intersecting distributions of the RGS isoform and the GPCR. This has the potential to reduce agonist off-target effects and makes RGS proteins an attractive target for modulation of GPCR signaling.

The potent RGS inhibitors discovered to date are all covalent modifiers of cysteine residues and are selective for RGS4 and RGS1.^{31,116,118} These proteins have four and three cysteines, respectively, in the RGS homology domain, which is more than most other RGS proteins. RGS4 has been linked to nervous system related disease states in which RGS4 inhibition may be desirable, including seizures⁹⁸ and Parkinson's disease.^{109,112,117} Continued efforts to seek non-covalent inhibitors are worth pursuing, because the lower risk associated with non-covalent inhibitors is considered safer and may facilitate further development.⁶⁰ In addition, it would be valuable to discover RGS inhibitors with other specificities since other RGS proteins which are not potently inhibited by covalent modifiers have been implicated as potential targets, including RGS17 in cancer^{103,105} and RGS19 in depression.¹⁰¹ To identify noncovalent inhibitors with novel specificities, it will be useful to understand what factors apart from the number of cysteines in the RGS

domain drive selectivity of RGS inhibitors.

The RGS homology domain contains nine alpha helices. A cysteine residue on the α_4 helix, which faces the interior of the α_4 - α_7 helical bundle, is conserved among 18 of the 20 RGS isoforms, excepting only RGS6 and RGS7.¹²⁶ Interestingly, when RGS proteins are mutated to contain only this single, shared cysteine, there are still dramatic differences in the potency by which different isoforms are inhibited.¹⁴⁴ RGS19, which contains only the shared α_4 cysteine, is more potently inhibited than single-cysteine versions of RGS4 and RGS8.^{144,145}

Previously, we found using molecular dynamics (MD) simulations that RGS19 is more flexible than RGS4 and RGS8.¹⁴⁴ In these modeling studies, we also found that salt bridge interactions were perturbed in response to inhibitor binding¹⁴⁶ In this work, we hypothesized that mutations that alter salt bridge interactions will both enhance RGS protein flexibility and increase the potency of RGS inhibitors such as CCG-50014.

Materials and Methods

Materials

Chemicals were purchased from Sigma-Aldrich (St. Louis, MO). QuikChange II Mutagenesis kit was purchased from Agilent (Santa Clara, CA). BL21(DE2) competent cells and Protein Thermal Shift Dye Kit was purchased from Thermo Fisher Scientific (Waltham, MA). Lumavidin Microspheres were purchased from Luminex (Austin, TX). CCG-50014 {4-[(4- fluorophenyl)methyl]-2-(4-methylphenyl)-1,2,4-thiadiazolidine-3,5-dione} was synthesized as previously described.¹¹⁵

Protein Expression and Purification

RGS proteins were produced as previously described.¹⁴⁴ Briefly, a his-tagged RGS domain

of RGS8 in a pQE80 vector, a his-tagged RGS domain of RGS19 in a pET15b vector, and a his-tagged $\Delta 51$ N-terminally truncated RGS4 in a pET23d vector were transformed into BL21(DE3) competent *E. coli* cells (Sigma-Aldrich, St. Louis, MO). At an OD₆₀₀ of 2.0, protein production was induced by addition of 200 μ M IPTG, and incubation was continued at 25 °C for 16 hours. Cells were lysed and the protein was purified by nickel affinity chromatography. Mutations were induced with a QuikChange mutagenesis kit (Agilent) and verified by Sanger sequencing. All RGS proteins, including those with mutations in salt bridge-forming residues, were produced on a single-cysteine background (WT RGS19, C160A RGS8, and C74A C132A C148A RGS4). $G\alpha_o$ protein was expressed and purified as described.¹³²

Differential Scanning Fluorimetry

Differential scanning fluorimetry was performed using the Protein Thermal Shift Dye Kit (ThermoFisher Scientific, Waltham, MA). Dye was added at 1X to 10 μ M protein samples in 50 mM HEPES and 100 mM NaCl buffer, pH 7.4 in a volume of 20 μ L. Fluorescence was read using a QuantStudio 7 Flex Real-Time PCR System while the temperature was ramped from 20 °C to 80 °C at a rate of 0.05 °C/s. Peak melting temperatures were defined as the point of fastest increase in fluorescence with respect to temperature. Data was analyzed using Protein Thermal Shift software v1.3 (Thermo Fisher Scientific, Waltham, MA) and Prism 7 (GraphPad Inc, LaJolla, CA).

Flow Cytometry Protein Interaction Assay (FCPIA)

FCPIA was performed as described¹³³ with minor modifications. RGS proteins were biotinylated by incubation at 1:1 molar ratio with EZ-link NHS-LC-biotin (Thermo Fisher Scientific) for two hours on ice, then excess biotin was removed using Amicon Ultra centrifugal filters (catalog no. UFC501096, Millipore, Burlington, MA). RGS proteins at 50 nM were incubated with

xMAP LumAvidin beads (Luminex, Austin, TX) while shaking at room temperature for 1 hour. Beads were washed and incubated with varying concentrations of CCG-50014, followed by addition of 50 nM G α o labeled with AF-532.¹³³ Samples were read in a Luminex 200 flow cytometer as described¹³³ and analysis performed in GraphPad Prism 7.

Hydrogen-Deuterium Exchange

Hydrogen-deuterium exchange was performed as previously described.^{134,144} Briefly, proteins were incubated on ice at 1.2 μ M in 90% D₂O solvent with 5 mM HEPES and 100 mM NaCl, pH 7.4 for the desired time (1, 3, 10, 30, or 100 minutes). Exchange was quenched by 1:1 addition of ice cold 1% formic acid. A Shimadzu pump was used to load 100 μ L of each sample onto a pepsin column (Waters, Milford, MA) followed by incubation for 1 minute for digestion. Samples were then loaded to an Xbridge BEH C18 VanGuard trap column (Waters) and eluted and separated using an Ascentis Express Peptide ES-C18 column (Sigma-Aldrich) with a gradient of 0.1% formic acid to acetonitrile. All columns and solvents were maintained on ice. Peaks were detected with a Xevo G2-XS QToF mass spectrometer (Waters). Data were analyzed using MassLynx (Waters), HX-Express2,¹³⁵ and GraphPad Prism 7.

Molecular Dynamics (MD) Simulation

The Vashisth Lab at UNH performed two sets of classical all-atom and explicit-solvent MD simulations for single-cysteine RGS4 and RGS4 D90L, single-cysteine RGS8 and RGS8 E84L, and WT RGS19 and RGS19 L118D (Table 3-2) using the NAMD software¹³⁶ on a high-performance computing cluster (Towns et al., 2014) using the CHARMM force-field with the CMAP correction.^{138,139} We used Visual Molecular Dynamics (VMD) for system creation and post-simulation analysis.¹³⁷ The initial coordinates were obtained from the protein data bank

Run No.	System	Initial structure	Run length (μ s)	System size (atoms)	No. of runs
1	RGS4 D90L	1AGR	1	30031	2
2	RGS4	1AGR	1	30031	2
3	RGS8 E84L	2ODE	1	32257	2
4	RGS8	2ODE	1	32257	2
5	RGS19 L118D	1CMZ	1	25077	2
6	RGS19	1CMZ	1	25077	2

Table 3-2: Details of MD simulations.

files with codes 1AGR (RGS4), 2DOE (RGS8), and 1CMZ (RGS19). Except for Cys95 in RGS4 and Cys89 in RGS8, all cysteines were changed to alanines. Each protein was then solvated in a simulation box of TIP3P water molecules¹⁴⁷ and charge-neutralized with NaCl. The final solvated and ionized simulation domains contained 30031 atoms (RGS4), 32257 atoms (RGS8), and 25077 atoms (RGS19). Each solvated and ionized system was energy minimized for ~500-1000 cycles via conjugate-gradient optimization, then equilibrated via 1 μ s MD simulations conducted with a time-step (Δt) of 2 fs. The NPT (constant number, pressure, temperature) ensemble with a Langevin thermostat and a damping coefficient of 5 ps⁻¹ was used for temperature control and the Nosé-Hoover barostat was used for pressure control. Periodic boundary conditions were used throughout; non-bonded interactions were accounted for with a cut-off of 10 Å where smooth switching was initiated at 8 Å. Long-range electrostatic interactions were handled using the Particle Mesh Ewald (PME) method.

Dynamic cross-correlation analysis

The dynamic cross-correlation (DCC) maps of each system were calculated based on the C α atoms of residues using the MD-TASK package.¹⁴⁸ Each cell value (C_{ij}) in the matrix of the DCC map was calculated using the following formula:

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\left(\sqrt{\langle \Delta r_i^2 \rangle} \cdot \sqrt{\langle \Delta r_j^2 \rangle} \right)}$$

Where Δr_i represents the displacement from the mean position of atom i, and $\langle \rangle$ denotes the time average over the whole trajectory. Positive values of C_{ij} show correlated motion between residues i and j, moving in the same direction, whereas negative values of C_{ij} show anti-correlated motion between residues i and j, moving in the opposite direction.

Analysis of salt-bridge interactions

Salt-bridge interaction analysis was carried out using VMD based on a distance criterion uniformly applied to determine the existence of salt-bridges for each frame in all trajectories. Specifically, salt-bridge interactions were considered to be formed if the distance between any of the oxygen atoms of acidic residues and the nitrogen atoms of basic residues were within a cut-off distance of 4 Å.

Statistical Analysis

All deuterium exchange and functional inhibition data were done with an n of 3 independent experiments. Sample sizes were predetermined before experiments were done. Changes in thermal stability were analyzed by 1-way ANOVA with Sidak's multiple comparisons post-test. H_0 : There is no difference in thermal stability between WT and mutant RGS proteins. Differences in deuterium incorporation were analyzed using 2-way ANOVA with Sidak's multiple comparisons post-test. H_0 : There is no difference in deuterium incorporation between WT and mutant RGS proteins. Error bars represent means $\pm$ SD. In saturation binding experiments, RGS-G α inhibition was determined by fitting total and nonspecific binding. In functional inhibition experiments, IC_{50} was determined by fitting a four-parameter logistic curve. All curve fitting and statistical analysis was done using GraphPad Prism 7 (GraphPad Inc, LaJolla, CA).

Results

Comparison of the structures for RGS19 (PDB 1CMZ),¹⁴² RGS4 (PDB 1AGR),⁸³ and RGS8 (PDB 5DO9)¹²⁸ shows that there are differing numbers of interhelical salt bridges on the exteriors of their α_4 - α_7 helix bundles. Some of these may contribute to differences in stability and dynamics among the RGS isoforms.

RGS19 has only one interhelical salt bridge in this bundle, between E125 (α_4) and K138 (α_5) (Fig. 3-1A and B). However, this salt bridge is well conserved among all three proteins (Fig. 3-1A-D), so it is unlikely to contribute to observed differences in flexibility.¹⁴⁴ A salt bridge network that connects α_4 , the α_5 - α_6 interhelical loop, and α_5 is present in RGS8 (E84-R119-E111) and RGS4 (D90-K125-E117) but absent in RGS19 (Fig. 3-1A and B). The residues that form this network are present in 7 of the 20 RGS protein family members, all in the R4 subfamily. Between the α_5 and α_6 helices, a salt bridge is present in RGS8 (D114-R132), but absent in both RGS4 and RGS19 (Fig. 3-1A and C). Finally, a charged pair between the α_6 and α_7 helices is present in RGS8 (E91-K104) and RGS4 (D130-K155), but is absent in RGS19 (Fig. 3-1A and D).

To estimate the relevance of each of these salt bridges in maintenance of helix bundle rigidity, the time each amino acid in a charged pair spent within a 4Å of one another over the course of a long timescale (2 μ s) MD simulation¹⁴⁴ was measured. The α_6 - α_7 salt bridge, which is present in RGS4 and RGS8 but absent in RGS19, occupied a salt bridge-forming distance for 31.5% of the simulation in RGS4 and 36.1% in RGS8. The salt bridge interaction between residues of α_4 and α_5 - α_6 interhelical loop, also not present in RGS19, was maintained for 58.7% of time in RGS4 and 44.2% in RGS8 (Table 3-3). The charged pair that is unique to RGS8 between α_5 and α_6 helices remained in contact for 47.5% of the simulation.

We elected to make mutations that altered interhelical salt bridges to test their functional

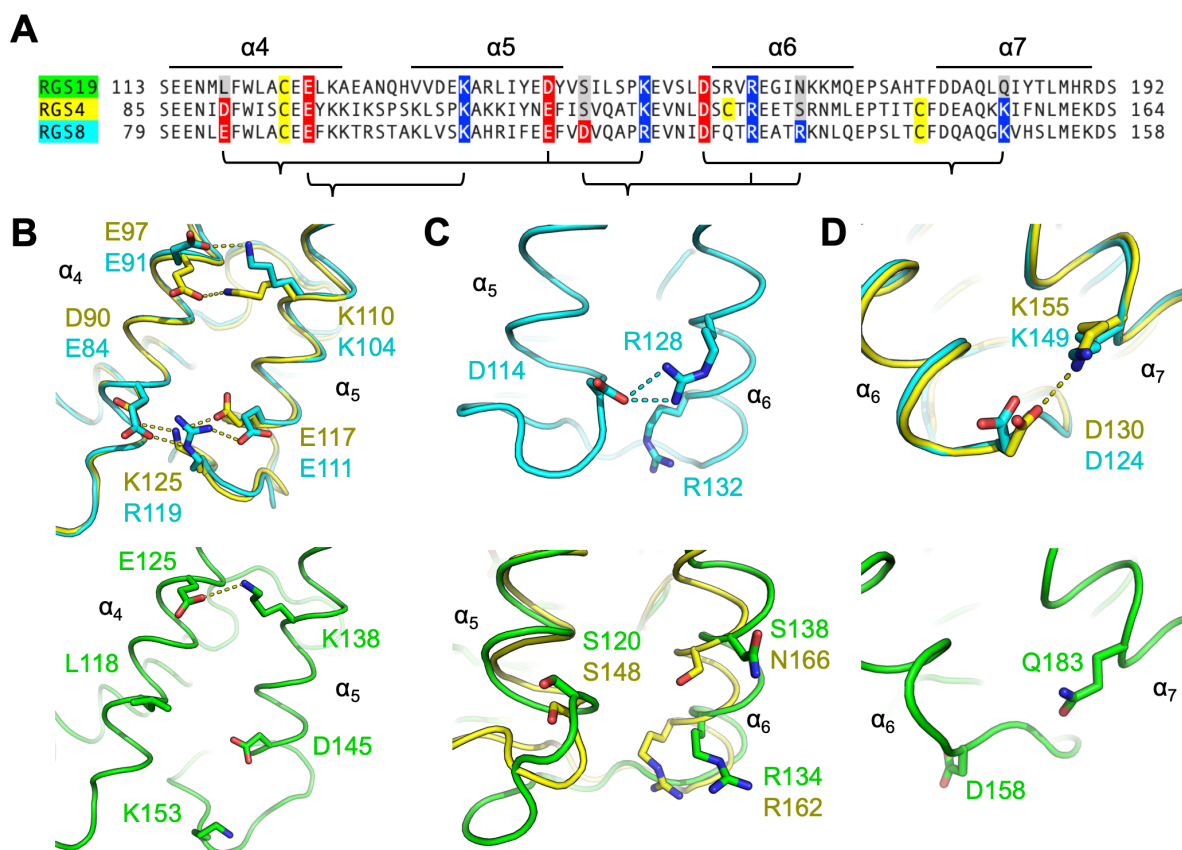


Figure 3-1: (A) Alignment of RGS19, RGS4, and RGS8 sequences in α_4 - α_7 helix bundle. Charged residues that make interhelical contacts are indicated in red and blue. RGS19 has 1, RGS4 has 3, and RGS8 has 4 salt bridges. Structural alignments of α_4 - α_5 (B), α_5 - α_6 (C), and α_6 - α_7 (D) helix pairs are shown, with highlighted residues in panel a rendered as sticks. RGS19 (PDB 1CMZ) is in green, RGS4 (PDB 1AGR) is in yellow, and RGS8 (PDB 5DO9) is in cyan. Black brackets in panel A indicate residues depicted in panels B, C, and D

	α_4 - α_5		% of sim within 4Å	α_5 - α_6		% of sim within 4Å	α_6 - α_7		% of sim within 4Å	CCG-50014 IC ₅₀ (μM)
RGS4	D90	K125	58.7	S120	S138	-	D130	K155	31.5	8.5
RGS8	E84	R119	44.2	D114	R128	47.5	D124	K149	36.1	>1000
RGS19	L118	K153	-	S148	N166	-	D158	Q183	-	1.1

Table 3-3: The salt-bridge interaction within the α_4 - α_7 bundle of helices in single-cysteine structure of RGS4, RGS8, and RGS19 from MD simulations and potency of CCG-50014 inhibition of single-cysteine RGS proteins in our previous work.¹⁴⁴

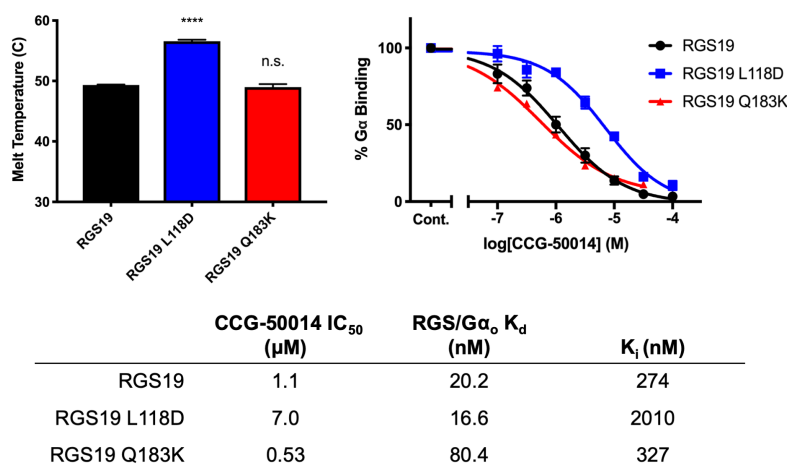


Figure 3-2: L118D mutation increases thermal stability of RGS19, but Q183K mutation has no significant effect ($n = 3$, 1-way ANOVA with Sidak's multiple comparison test. **** $p < 0.001$). L118D mutation in RGS19 has reduced potency of inhibition of CCG-50014, but Q183K mutation does not. K_i , calculated using a Cheng-Prusoff correction,²³² is reported to account for effect of mutations in RGS on $G\alpha_o$ affinity.

roles. There are two positions at which interhelical salt bridges are shared by RGS4 and RGS8 but are absent in RGS19: α_4 - α_5 (Fig. 3-1B) and α_6 - α_7 (Fig. 3-1D). In the α_4 helix of RGS19, L118 was mutated to an aspartate to introduce the α_4 - α_5 salt bridge found in RGS4 and RGS8 (Fig 3-1B). In helix α_7 of RGS19, Q183 was mutated to a lysine to introduce the α_6 - α_7 salt bridge found in RGS4 and RGS8 (Fig 3-1D). In order to eliminate confounding effects due to multiple cysteines in inhibitor potency experiments, all proteins, with and without salt-bridge mutations, used a single-cysteine protein background. Each construct has only the conserved cysteine in helix α_4 of the RGS domain.

To determine how disruption or addition of a salt bridge may alter protein structure or dynamics, thermal stability was measured by differential scanning fluorimetry. Addition of a salt bridge in RGS19 by the L118D mutation caused a 7 °C increase in thermal stability compared

to WT (Fig 3-3A). In contrast, the Q183K mutation in RGS19 did not alter thermal stability or inhibitor potency (Fig. 3-2). Removal of a salt bridge in RGS8 by the E84L mutation caused an 8 °C decrease in thermal stability (Fig 3-3B). Unexpectedly, RGS4 showed a more complex pattern in which the D90L mutation resulted in a biphasic melt curve and a 5 °C increase in melting temperature rather than a decrease (Fig 3-3C).

To probe the molecular details of changes in structural flexibility in the mutant proteins, we conducted microsecond timescale classical MD simulations in explicit-solvent for RGS19 L118D, RGS8 E84L, and RGS4 D90L. Root-mean-square deviations (RMSDs) of these simulations are shown in Fig. 3-4. To understand the effect of the mutations on the protein structures, particularly in helices in the vicinity of the mutated site, we computed the root-mean-square fluctuation (RMSF) per residue from two independent MD simulations of mutated and WT RGS4, RGS8, and RGS19. The calculated change in RMSF per residue of the mutant RGS19 L118D from wild-type RGS19 reveals a strong stabilization and decrease in fluctuations of residues located in helices α_4 - α_7 and in the interhelical loops between these helices. There is a particularly pronounced decrease in motion in the α_5 - α_6 interhelical loop (Fig. 3-5A). We find a modest increase in fluctuation of residues in mutant RGS8 E84L vs. the wild-type structure (Fig. 3-5B). These changes are in the loop region connecting helices α_5 and α_6 , the α_6 helix, and the loop connecting helices α_6 and α_7 . Similar changes but of lesser extent were found in the mutant RGS4 D90L (Fig. 3-5C). Additionally, small decreases were observed in the RMSF values of residues in helices α_3 and α_8 of the mutated RGS19 (Fig. 3-5A), but not in the mutated RGS8 and RGS4 (Fig. 3-5B and C).

To further investigate whether salt bridge-modifying mutations in RGS4, RGS8, and RGS19 affect residue-residue interactions, we calculated dynamic cross-correlation matrices

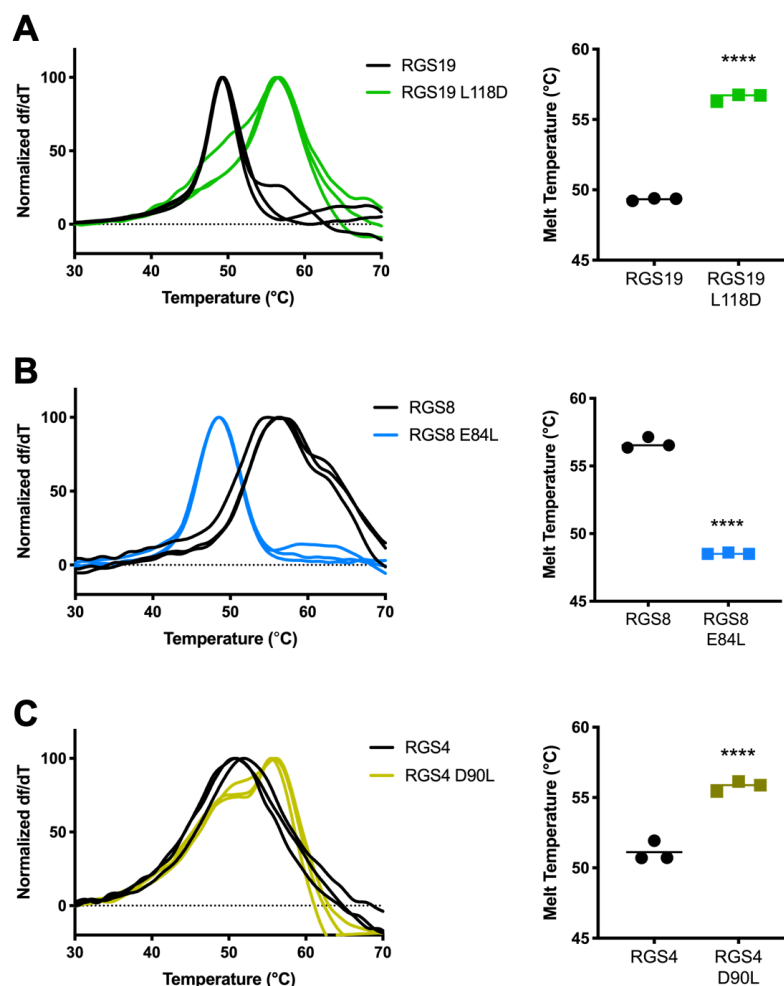


Figure 3-3: Thermal stability was determined by differential scanning fluorimetry. (A) The L118D mutation in RGS19 increased melting temperature by 7 $^{\circ}C$ compared to WT. (B) The E84L mutation in RGS8 decreased melting temperature by 8 $^{\circ}C$. (C) The RGS4 D90L mutation introduced a biphasic melt curve and increased melting temperature by 5 $^{\circ}C$. For each pair, the three replicate derivative melt curves are shown on the left and average melt temperatures are shown on the right. Error bars represent SD. $n = 3$. Analyzed by 1-way ANOVA with Sidak's Multiple Comparisons test. **** $p < 0.0001$

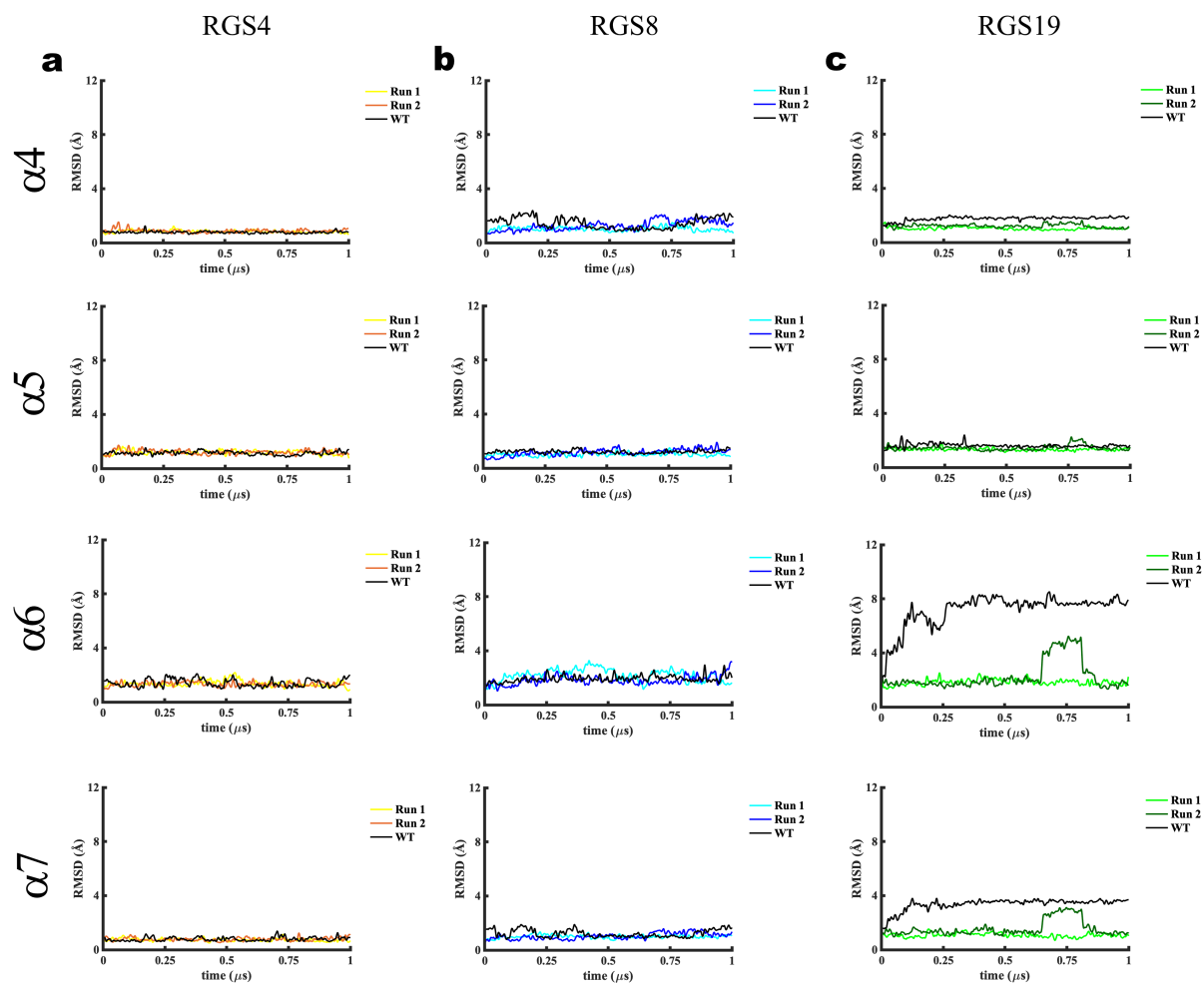


Figure 3-4: The traces of root-mean-squared-deviation (RMSD) vs. simulation time (μs) for (a) RGS4 D90L, (b) RGS8 E84L, and (c) RGS19 L118D. Two independent simulation runs for each structure are presented, and the wild-type runs are presented from our previous work.¹⁴⁴

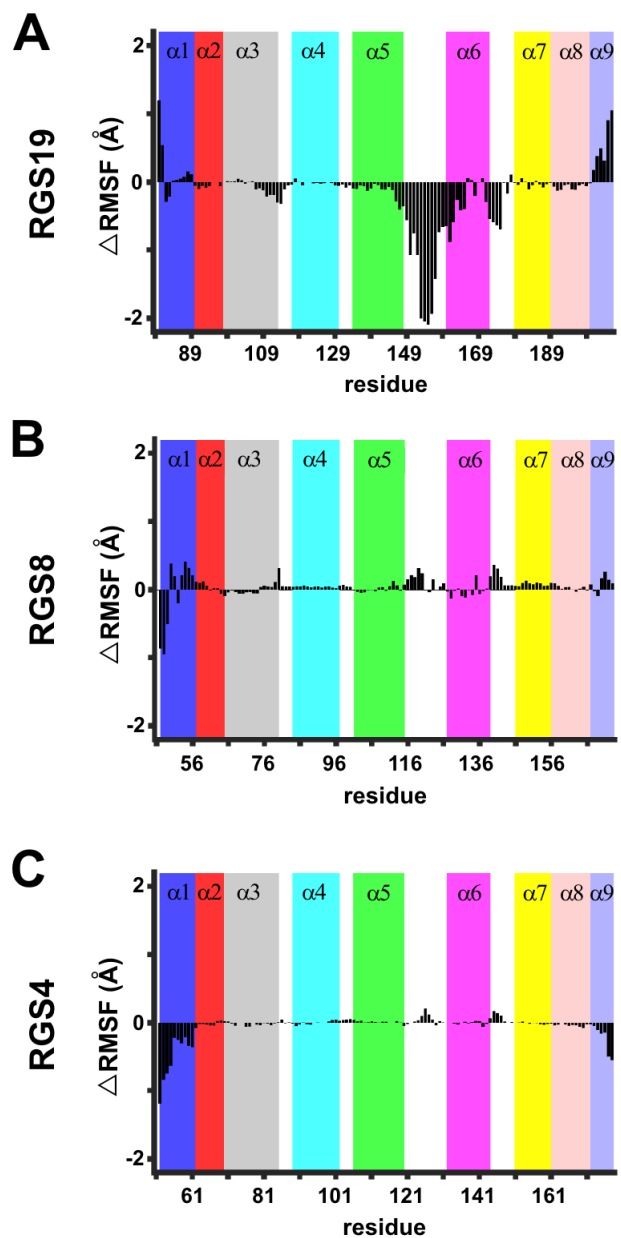


Figure 3-5: Change in RMSF per residue (ΔRMSF) between wild-type RGS proteins and RGS proteins with mutation in the α_4 - α_5 salt bridge forming residue. (A) L118D in RGS19 (B) E84L in RGS8 and (C) D90L in RGS4. Data represent differences in RMSF from two independent MD simulations of the mutated forms of RGS proteins.

for the C α atoms in all MD trajectories. For WT RGS19, RGS8 and RGS4, there is a modest positive correlation between the motions of residues of the α_4 helix and the residues of the α_5 helix (Fig. 3-6A-C). For the RGS19 L118D mutant, we find higher residue-residue correlations between helices α_4 and α_5 in comparison to unmutated RGS19 (see arrows, Fig 3-6A). There was no appreciable change between WT and mutant RGS4 (Fig 3-6B). For wild-type RGS8, we find that the motions of residues in the α_4 helix (aa 79-93) and the α_5 helix (aa 97-113) are marginally positively correlated (see arrows, Fig. 3-6C). This positive correlation between the α_4 and α_5 helices remains in the RGS8 E84L mutant, but shows a modest shift in areas of correlation away from the loop connecting α_4 - α_5 to mid-regions of the α_4 and α_5 helices (see arrows, Fig. 3-6C).

In order to experimentally determine which regions in WT and mutant proteins were affected by the salt bridge mutations, hydrogen-deuterium exchange studies were performed. After exposure to solvent containing 90% D₂O, proteins were digested with pepsin and deuterium incorporation (DI) was measured by mass spectrometry as previously reported.¹⁴⁴ In RGS4, the fragment surrounding the salt-bridge mutation site (aa 88-91) took up deuterium very slowly in both the WT and D90L mutant constructs, reaching 8.1% and 6.7% DI, respectively. However, the D90L mutation led to a substantial increase in deuterium exchange in the 92-97 fragment surrounding Cys95, from 17.5% to 37.0% DI. The RGS4 D90L mutant also trended toward increased DI across all protein fragments compared to WT RGS4, especially at higher timepoints (Fig. 3-7A). In RGS8, removal of the salt-bridge forming residue by the E84L mutation did not cause a significant change in DI in either of the fragments of the α_4 helix but trended toward a global increase in DI throughout the protein (Fig. 3-7B). In RGS19, mutation of L118 to a salt bridge-forming residue, aspartic acid, caused significant decreases in DI in both α_4 helical fragments, aa 116-119 and aa 120-125. In the 116-119 fragment, WT RGS19 had reached 43.1% DI by 10 minutes,

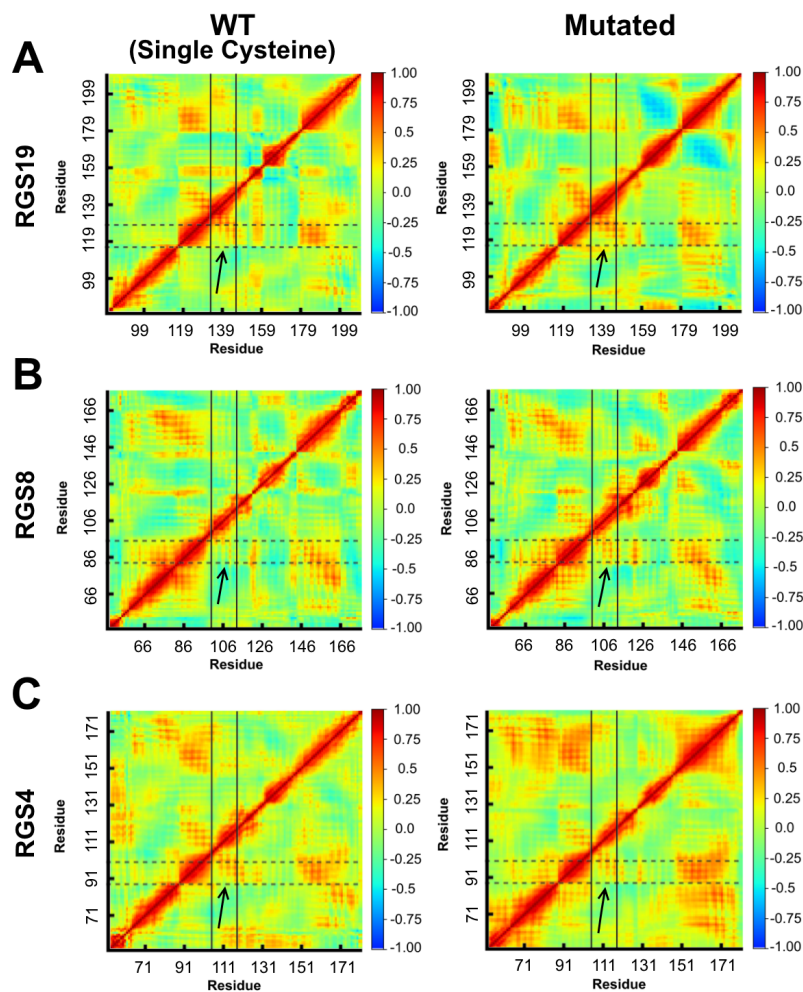


Figure 3-6: Dynamic cross correlation matrix calculated for the C α atoms of (A) RGS19/RGS19 L118D, (B) RGS8/RGS8 E84L, (C) RGS4/RGS4 D90L. Horizontal dotted lines indicate the regions of the α_4 helix, while vertical solid lines indicate the regions of the α_5 helix for each protein. The color scheme ranges from anticorrelation (-1.0, blue), no correlation (0, green), and positive correlation (+1.0, red). Values are the average for the two independent simulation runs.

	$G\alpha_o$ K_D (nM)	CCG-50014 IC_{50} (μ M)	CCG-50014 K_i (μ M)
RGS19	16.6	1.1	0.27
RGS19 L118D	20.2	7.0	2.01
RGS8	5.9	29.0	3.06
RGS8 E84L	4.8	4.6	0.40
RGS4	5.2	8.8	0.83
RGS4 D90L	3.9	2.2	0.16

Table 3-4: Interaction affinities between $G\alpha_o$ and RGS proteins and mutants, and IC_{50} and K_i of inhibition of RGS- $G\alpha_o$ binding by CCG-50014. K_i values were calculated by Cheng-Prusoff correction of the IC_{50} values.

while the RGS19 L118D mutant showed less than half as much DI (18.7%). In fragment 120-125, WT RGS19 reached 18.5% DI at 10 minutes, while the RGS19 L118D mutant reached only 6.2%. Unlike RGS4 and RGS8, the RGS19 L118D mutant's changes in DI were more restricted to fragments from helices neighboring the mutation site, and were most pronounced in the early (1 to 10 minute) timescale (Fig. 3-3C).

Finally, to assess the functional relevance of the α_4 salt-bridge forming residues, we used a flow-cytometry based protein-protein interaction assay (FCPIA)^{113,133} to measure the binding of RGS proteins to $G\alpha_o$ and the potency of inhibition by CCG-50014. The L118D mutation in RGS19 induced an increase in IC_{50} from 1.1 μ M (WT) to 7.0 μ M (L118D) (Fig. 3-8A). Conversely, removal of this charged α_4 residue in RGS4 and RGS8 induced a decrease in IC_{50} (Fig. 3-8B and C). CCG-50014 inhibited the RGS- $G\alpha$ interaction with an IC_{50} of 8.8 μ M for WT RGS4 and 2.2 μ M for the RGS4 D90L mutant. It showed a potency of 29 μ M for WT RGS8 and 4.6 μ M for the RGS8 E84L mutant. None of the mutations to salt bridge-forming residues on the α_4 helix caused notable changes in affinity between $G\alpha_o$ and RGS proteins. The L118D mutation in RGS19 shifted the K_d of the $G\alpha_o$ interaction from 17 nM to 20 nM, the E84L mutation in RGS8 shifted the K_d from 5.9 nM to 4.8 nM, and the D90L mutation in RGS4 shifted the K_d from 5.2 nM to 3.9 nM (Table 3-4).

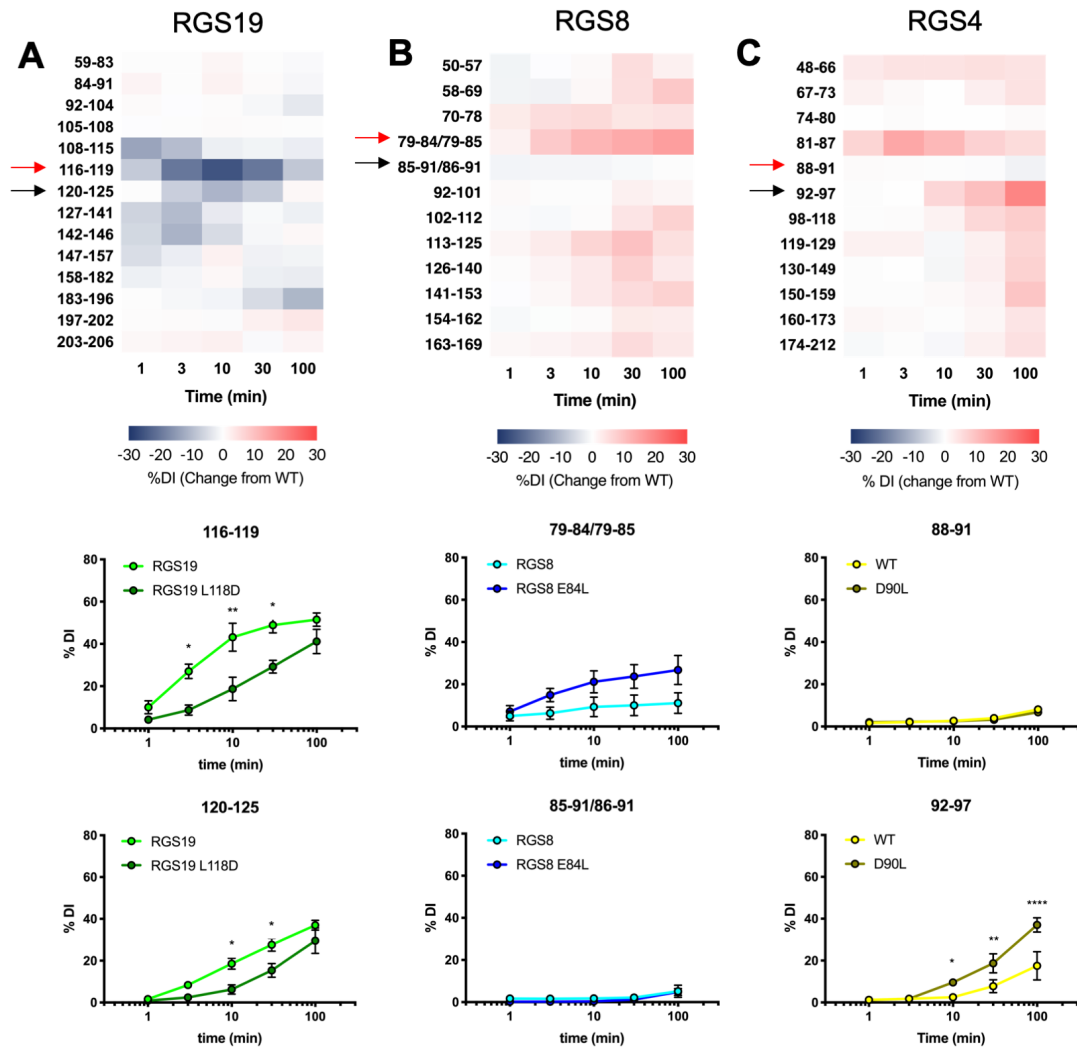


Figure 3-7: Difference in %deuterium incorporation ($\Delta\%DI$) between mutated and unmutated proteins in RGS19 L118D (A), RGS8 E84L (B), and RGS4 D90L (C) fragments, as measured by HDX. Red arrows indicate fragments containing mutated residue, and black arrows indicate fragments containing conserved $\alpha 4$ cysteine. Kinetics of deuterium incorporation in these fragments for individual constructs are shown below. $n = 3$. Error bars represent SD. Analyzed by 2-way ANOVA with Sidak's multiple comparisons test. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$.

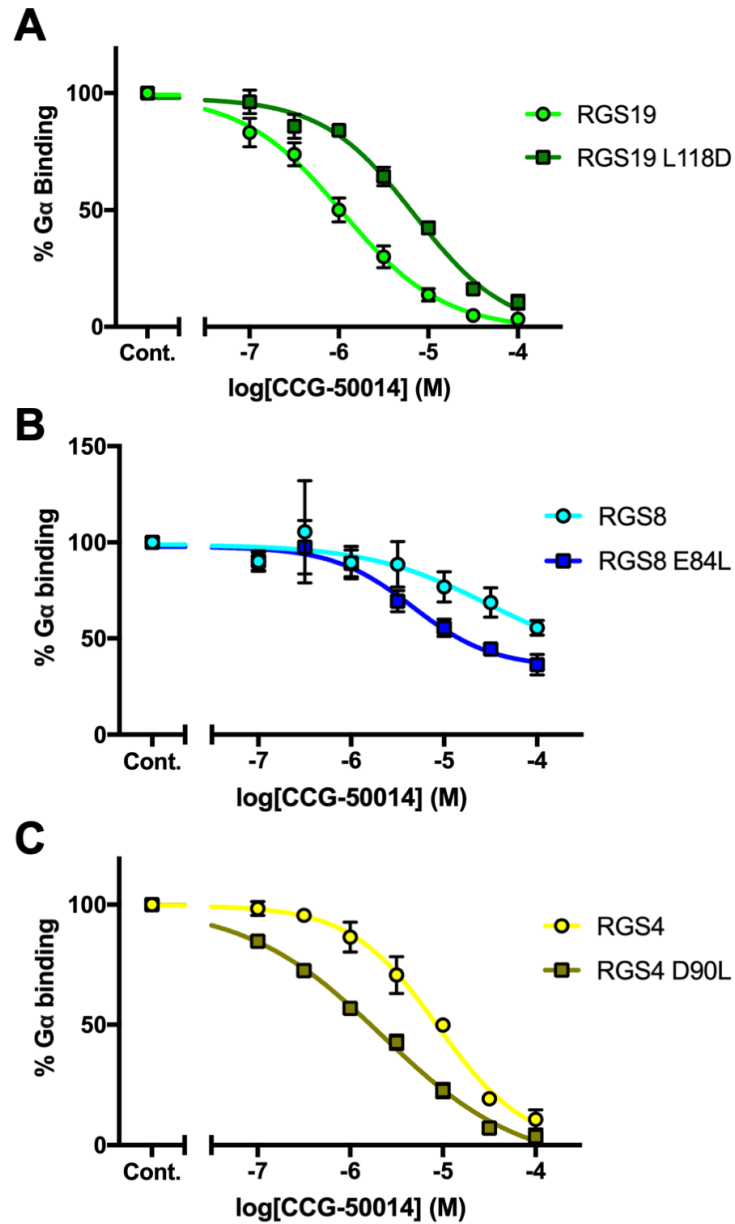


Figure 3-8: Potency of inhibition of CCG-50014 against $\alpha 4$ is altered in salt bridge mutants of RGS proteins. (A) RGS4 IC₅₀: 8.8 μ M, RGS4 D90L IC₅₀: 2.2 μ M. (B) RGS8 IC₅₀: 29 μ M, RGS8 E84L IC₅₀: 4.6 μ M. (C) RGS19 IC₅₀: 7.0 μ M, RGS19 L118D IC₅₀: 1.1 μ M. n=3.

Discussion

A comparison of the crystal structures of the three RGS proteins studied here revealed several differences in charged residue contacts among the proteins. We first observed that RGS19 has fewer interhelical salt bridges in its α_4 - α_7 helical bundle than RGS4 or RGS8. This may be responsible for the high flexibility previously observed in WT RGS19.¹⁴⁴ RGS8 has four distinct interhelical salt bridges within the helical bundle, while RGS4 has three and RGS19 has one (Fig 3-1A), correlating with previously observed flexibility differences. RGS19 is most flexible, followed by RGS4 and RGS8.¹⁴⁴ This further supports a role of salt bridges in RGS protein flexibility.

The changes in thermal stability in response to mutations in the α_4 helix salt bridge-forming residues suggest that this location may be responsible for differences in stability and dynamics among the isoforms. This is supported by the increase in thermal stability in response to the L118D mutation in RGS19, and destabilization in RGS8 response to the E84L mutation. While the D90L mutation altered thermal stability in RGS4, it stabilized rather than destabilized the protein. The biphasic melt curves in D90L RGS4 make the thermal stability data difficult to interpret. HDX clarifies the effect of the D90L mutation in RGS4 by showing localized increases flexibility of the protein. The lack of effect on thermal stability with the Q183K mutation in RGS19 correlates with the observation that the α_6 - α_7 salt bridges in RGS4 and RGS8 were less stably maintained in simulations than were the α_4 - α_5 salt bridges. In light of these results, we found it unlikely that the difference between Q183 in α_6 of RGS19 and the lysines found in RGS4 and RGS8 (K155 and K149 respectively) play a major role in the flexibility differences between these proteins. Rather, the salt bridge-forming residue on α_4 is a stronger driver of differences in protein flexibility.

To determine the effects of mutations in salt bridge-forming residues on protein dynamics,

both an *in silico* approach (all-atom MD simulations) and an experimental approach (hydrogen-deuterium exchange) were employed. In simulations, the increase in positive correlation between residues in the α_4 and α_5 helices in the RGS19 L118D mutant likely results from the introduced interhelical salt-bridge. The decrease in DI in the α_4 helix of RGS19 in the HDX studies is consistent with reduced solvent exposure. This is of particular interest given that the Cys123 target of the TDZD compounds is located in that helix. Conversely, mutations that eliminated salt bridges in RGS4 and RGS8 increased DI in some fragments from their α_4 helices (Fig. 3-7A and B), suggesting that this results in increased solvent exposure and greater compound accessibility at the buried cysteine. Surprisingly, the RGS4 D90L mutant did not have increased DI in the fragment spanning the mutation site (Fig. 3-7C). In addition, the μ s timescale MD simulations captured positive residue-residue (C α -C α) correlations between the α_4 and α_5 helices of that were similar in WT and mutated RGS4 D90L. This fits with the thermal stability data and suggests that the effect of the D90L mutation in RGS4 is more complex than simple disruption of an ionic contact.

In MD simulations, the RGS4 D90L and RGS8 E84L mutations did not have as large an effect on the magnitude of residue fluctuations as did the L118D mutation in RGS19 (Fig. 3-5A and B). This may be because differences become apparent on shorter timescales in RGS19 than in RGS4 and RGS8, so simulations on μ s timescales may not have captured all of the differences in dynamics caused by mutations in RGS4 D90L and RGS8 E84L. Indeed, in HDX studies, stronger differences in DI were observed between RGS19 and RGS19 L118D at shorter timepoints (1 and 3 minutes) than in RGS4 D90L and RGS8 E84L (Fig 3-7A-C).

Finally, to determine how changes in protein flexibility affected the potency of inhibition by an RGS inhibitor, we used FCPIA to evaluate the inhibition of G α binding by CCG-50014. Importantly, manipulation of RGS protein flexibility induced the expected changes in the potency

of inhibition by TDZD covalent modifiers. Thus, enhancing flexibility by removal of salt bridge-forming residues increased the potency of inhibition by CCG-50014 while reducing protein flexibility reduced potency of inhibition by CCG-50014. These results support a causal relationship between RGS protein flexibility and potency of inhibition.

In conclusion, differences in flexibility among RGS isoforms appear to drive differences in the potency of a covalent inhibitor, CCG-50014. The differences in isoform flexibility in turn are strongly influenced by the presence or absence of an α_4 - α_5 salt bridge and manipulation of this salt bridge is sufficient to induce changes in inhibitor potency among single-cysteine RGS proteins. Developing a deeper understanding of these differences in flexibility may enable the development of a new generation of RGS inhibitors with novel specificities.

CHAPTER 4:

Distinct Roles of Individual Cysteines in Covalent Inhibition of RGS Proteins

Vincent Shaw expressed proteins, performed FCPIA, MS, and SDS-PAGE, and analyzed data. Ryan Puterbaugh and Dr. Kriszina Varga performed NMR and prepared spectra.

Introduction

Signaling via heterotrimeric G-proteins is a pathway critical to pharmacology.¹² G-proteins are activated upon agonist binding to GPCR, allowing GDP release from the $G\alpha$ subunit and GTP association. This puts the G-protein in its active conformation, initiating downstream signaling via the $G\alpha$ subunit and the $G\beta\gamma$ dimer. Regulators of G-protein signaling (RGS) proteins end signaling through the G-protein by binding to the active $G\alpha$ subunit and accelerating hydrolysis of GTP. This GTPase-Activating Protein (GAP) activity is mediated by the RGS domain, a 130 aa domain with nine α helices.^{82,83}

There has been interest in targeting RGS proteins as a strategy for modulating G-protein signaling. By inhibiting the GAP activity of RGS proteins, GPCR-mediated signaling may be increased. Because there are many RGS isoforms with unique physiological roles, isoform selectivity will be particularly important to limit off-target effects in the therapeutic use of RGS inhibitors.¹⁴⁹ Covalently acting inhibitors have been developed that prevent binding between the RGS domain and the $G\alpha$ subunit by modification of cysteines in the RGS domain.^{113,114,116,150–152} These include CCG-203769, a thiadiazolidinone (TDZD) inhibitor that is selective for RGS4 and may hold promise for treatment of Parkinson's disease.^{109,117} A better understanding of the role of individual cysteines in RGS inhibition by CCG-203769 will help define the molecular mechanism of isoform specificity.

Previous work demonstrates a relationship between RGS isoform dynamics and potency of inhibition among three RGS proteins (RGS4, RGS8, and RGS19) when mutated to contain a single cysteine.^{129,144,146} However, many RGS isoforms contain additional cysteines that may influence the potency of covalent modifiers. A cysteine on the α_4 helix is very well conserved, shared by all of the RGS proteins with the exception of RGS6 and RGS7. This cysteine is found

in all members of the R4 and RZ families. Only one other cysteine, on the α_7 helix, is conserved among the some of the RGS domains of RGS proteins. This is present in eight of the ten R4 family members, including RGS4 and RGS8. It is not found in RZ family members, such as RGS19. Although the α_4 and α_7 cysteines are near one another on adjacent helices in the 3D structure, existing crystal structural information does not indicate the presence of a disulfide bond in the apo structure.^{83,128,141,153}

Our previous work demonstrates that protein dynamics plays a role in the isoform specificity of TDZDs when compound action is restricted to a shared, single cysteine on the α_4 helix.¹⁴⁴ However, many RGS proteins, including RGS4 and RGS8, have additional cysteines in the RGS domain that may contribute to potency of covalent modifiers. The RGS proteins most potently inhibited by the TDZD CCG-50014 are RGS4 and RGS1, both of which have additional cysteines in the RGS domain beyond the well-conserved α_4 and α_7 cysteines.¹¹⁸. While a correlation between number of cysteines and potency of inhibition has been noted among TDZDs and several other inhibitors,^{31,118} the way that individual cysteines contribute to compound-induced changes in protein conformation have not been fully elucidated, particularly at cysteines beyond the conserved α_4 cysteine. In this work, we provide evidence that the TDZD CCG-203769 can act at multiple cysteines, and may mediate a unique interaction leading to the induction of a disulfide bond between cysteines common to many RGS isoforms on the α_4 helix and the α_7 helix.

Materials and Methods

Protein purification and expression

Single-cysteine constructs of RGS8 containing either Cys107 or Cys160 were generated

by individual mutation of each cysteine to serine using QuikChange mutagenesis (Agilent, Santa Clara, CA). An RGS8 C160S mutant is termed Cys¹⁰⁷ RGS8 and an RGS8 C107S mutant is termed Cys¹⁶⁰ RGS8. His-tagged expression constructs of the RGS domain of RGS8 and RGS19, a Δ 51 N-terminally truncated RGS4, and G α_o were used to prepare the tagged proteins as previously described.^{132,144} Isotopically labeled proteins were expressed by plasmid transfection into BL21(DE3) competent cells (Sigma-Aldrich, St. Louis, MO). These were grown at 37 °C in LB to OD₆₀₀ 0.7, followed by centrifugation at 8000×g for 20 minutes and resuspension at 25% of the original volume in phosphate-buffered minimal media (recipe described by Storaska and Neubig, 2013),¹⁵⁴ supplemented with 4 g/L D-Glucose and 1 g/L (¹⁵NH₄)₂SO₄ (Sigma-Aldrich, St. Louis, MO). Cells were incubated in minimal media for 30 minutes at 37°C, 200 μ M IPTG was added, and the temperature was lowered to 25 °C and protein was induced for 12 hrs. Cells were lysed by sonication and centrifuged at 120,000×g for 1 hr. The cell lysates were batch purified on a nickel affinity column and eluted with 300 mM imidazole in 50 mM HEPES and 500 mM NaCl, pH 7.4. Protein was further purified by cation exchange chromatography. An SP sepharose column (GE, Chicago, IL) was equilibrated with 50 mM sodium phosphate, 40 mM NaCl, and 1 mM DTT (pH 6.9) and protein was eluted using linear gradient to buffer including 1M NaCl.

NMR Spectroscopy

Protein was dialyzed to buffer containing 50 mM sodium phosphate, 40 mM NaCl, pH 6.0 and concentrated to 50 μ M using Amicon 10,000 Da MWCO centrifugal filter columns (Millipore, Burlington, MA). D₂O, NaN₃, and 4,4-dimethyl-4-silapentane (DSS) were added to achieve 5% v/v, 4 mM, and 0.2 mM final concentration, respectively. Titrations of the protein with the CCG-203769 ligand were prepared, with the ligand protein-ligand concentration of 1:1, 1:2 and 1:4

(molar ratio). Samples were then packed into Shigemi NMR tubes, and ^1H - ^{15}N HSQC (Heteronuclear Single Quantum Correlation) NMR spectra were collected at 25 °C at the CUNY Advanced Science Research Center (ASRC). WT RGS8 and Cys¹⁰⁷ RGS8 data were collected using a Bruker AVANCE III HD 800 MHz NMR spectrometer equipped with a Bruker Ascend UltraShield Plus 18.8 Tesla standard bore magnet and a TCI Cryoprobe. Cys¹⁶⁰ RGS8 data were collected using a Bruker AVANCE III HD 700 MHz NMR spectrometer equipped with a Bruker UltraShield 16.4 Tesla standard bore magnet and a QCI-F CryoProbe. All NMR data were processed using the programs Bruker Topspin and NMRpipe.¹⁵⁵ All processed NMR data were analyzed using the program NMRFAM Sparky.¹⁵⁶

Iodoacetamide alkylation and trypsin digestion

RGS4 was treated with varying concentrations of iodoacetamide (IAA) in 50 mM HEPES and 100 mM NaCl buffer, pH 7.4 and incubated at room temperature while shaking, protected from light. Free IAA was removed by buffer exchange using Amicon 10,000 Da MWCO centrifugal filter columns (Millipore) into digestion buffer (400 mM ammonium bicarbonate, 5 mM DTT, pH 7.5) with 8 M urea. Samples were diluted to 0.5 M urea in digestion buffer. Trypsin from porcine pancreas (Sigma-Aldrich catalog no. T0134) was added at a ratio of 1:1 protein:trypsin. The mixture was incubated at 37 °C overnight before analysis by mass spectrometry (see methods below). Peak height may be used as measurement of peptide quantity.¹⁵⁷ The peak intensities of iodoacetamide-alkylated peptides were determined for the most abundant RGS4 fragments resulting from trypsin digestion that contained one cysteine. The four cysteines in the RGS domain of RGS4 and the associated fragments are: Cys71, aa 59-76; Cys95, aa 78-99; Cys132, aa 126-134; and Cys148, aa 140-155. The intensities of alkylated fragments are expressed as a percent of the

sum of the intensities of alkylated and unalkylated fragments.

Protection of RGS8 from iodoacetamide alkylation by CCG-203769

RGS8 WT, Cys¹⁶⁰, or Cys¹⁰⁷ proteins at 50 μ M in 100 mM HEPES and 100 mM NaCl (pH 7.4) were treated with 100 μ M CCG-203769 or DMSO vehicle (final concentration 1%) at room temperature for 1 hr. An excess of iodoacetamide (500 μ M) was added to quench the action of CCG-203769 by alkylation of any free cysteine thiols. The mixture was incubated in the dark at room temperature for 1 hr, then diluted 10-fold in urea (final concentration 9 M) to ensure access of iodoacetamide to free cysteines.

Protein mass spectrometry

Samples were injected using a Waters 2777c autosampler and desalted by trapping on a Hypersil Gold CN guard column (1 x 10 mm, Thermo Fisher Scientific, Waltham, MA) for full proteins or separated using an Ascentis Express Peptide ES-C18 column (2.1 x 50 mm, Supelco, Bellefonte, PA) for protein fragments using a gradient of 0.1% formic acid in H₂O and acetonitrile. Proteins were ionized by electrospray ionization using a Xevo G2-XS QToF mass spectrometer (Waters, Milford, MA) in positive ion mode, collecting data in continuum mode over m/z 100-2000. Full length protein spectra were deconvoluted and analyzed using the MaxEnt1 algorithm in MassLynx (Waters).

Flow cytometry protein interaction assay

The flow cytometry protein interaction assay (FCPIA) was performed as previously described¹³³ to measure RGS-G α binding. Briefly, biotinylated WT, Cys¹⁰⁷, or Cys¹⁶⁰ RGS8 was linked to Lumavidin microspheres (Luminex, Austin, TX). These were incubated with varying

concentrations of CCG-203769 or vehicle (DMSO) for 30 minutes. AF-532-labeled $G\alpha_o$ (50 nM final concentration) was added and bead fluorescence was read using a Luminex 200 flow cytometer.

Non-reducing SDS-PAGE

Protein samples at 5 μ M were pretreated with vehicle or 250 μ M CCG-203769. Where indicated, disulfides were reduced by addition of 1 mM dithiothreitol (DTT). Samples were mixed with SDS sample buffer (Bio-Rad) devoid of BME or other reducing agent, and separated by SDS-PAGE using a 15% polyacrylamide gel. Proteins were visualized using Coomassie Blue stain.

Results and Discussion

Cys148 in RGS4 is more accessible to a covalent modifier than Cys95

Differences in accessibility of the cysteines to the solvent may contribute to variation in TDZD action at different cysteines in RGS domain proteins. There are two cysteines that are conserved among R4 family members, one on each of the the α_4 and α_7 helices of the RGS domain (Fig. 4-1A). RGS4 is a representative RGS protein that has both of these cysteines, but also has two others. Previous studies have suggested mechanisms by which otherwise buried cysteines on RGS proteins may access solvent.^{129,144} In MD simulations, the α_7 helix cysteine was found to have more solvent accessible surface area than the α_4 cysteine in both RGS4 and RGS8.¹⁴⁴

To verify this experimentally, cysteine accessibility in RGS4 was measured by assessing the degree of modification by a general cysteine alkylator, iodoacetamide. By fragmenting the protein with trypsin, the degree of modification at each fragment can provide an indication of the relative exposure of individual cysteines; however, the rate of alkylation may also be affected

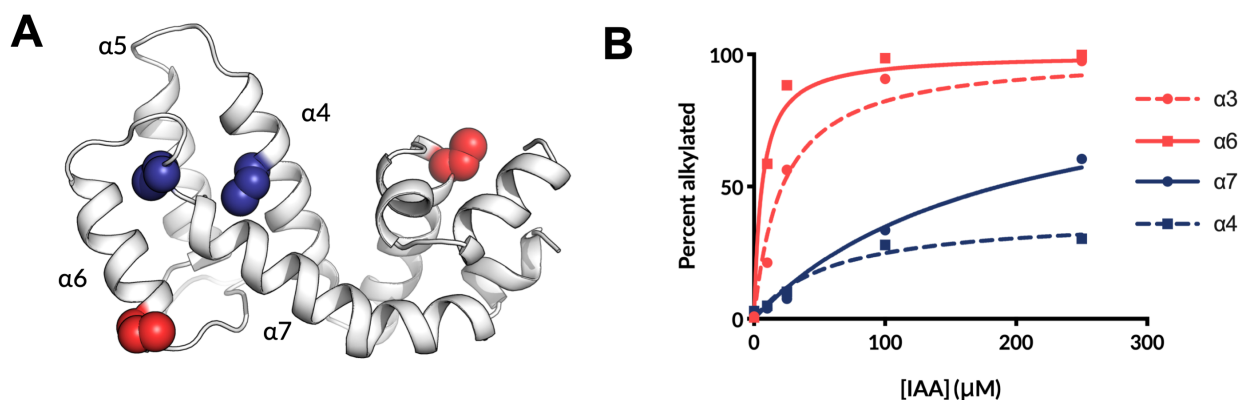


Figure 4-1: (A) Locations of cysteines in RGS protein based on structure of RGS4 (PDB: 1AGR). α_4 and α_7 cysteines, conserved across multiple RGS proteins, are marked in blue. The α_3 and α_6 helix cysteines, unique to RGS4, are marked in red. (B) Degree of IAA alkylation at Cys71 (α_3), Cys95 (α_4), Cys132 (α_6), and Cys148 (α_7) in RGS4.

by factors such as how readily cysteine thiols convert to the thiolate anion.¹⁵⁸ Cysteines on the α_3 and α_6 helices appear to be more exposed, as they became 97.4% and 99.8% alkylated respectively when exposed to 250 μ M IAA (Fig. 4-1B). This is consistent with their high degree of solvent exposure in crystal structures (Fig. 4-1A).^{83,153} Cysteines on α_4 and α_7 , which are conserved in RGS8 and other R4 family proteins, were less readily alkylated. The α_7 cysteine reached 60.4% alkylation with 250 μ M IAA, while the α_4 cysteine reached only 30.3% alkylation (Fig. 4-1B). This indicates that the α_7 cysteine may be more solvent-exposed than the α_4 cysteine, which is consistent with our previous modeling data showing that the α_4 cysteine has less solvent-exposed surface area than the α_7 cysteine in both RGS4 and RGS8.¹⁴⁴

CCG-203769 can directly act upon either cysteine in RGS8

To determine the individual roles of the two conserved cysteines (on α_4 and α_7 helices), we used the RGS8 protein and CCG-203769 ligand in ^1H - ^{15}N HSQC NMR spectroscopy. This protein was chosen because of the evidence supporting the stability of RGS8.¹⁴⁴ Some RGS proteins may

be sensitive to denaturation upon interaction with small molecules.¹⁵⁴ RGS8 has been observed to have high thermal stability relative to other RGS proteins (Fig. 3-2), which may make it a better candidate for studies requiring a long durations in solution or at elevated temperatures. CCG-203769 was chosen as ligand. Despite inhibiting RGS8 with a lower potency than CCG-50014, CCG-203769 is more soluble in aqueous solution.¹¹⁶

High quality 2D ^1H - ^{15}N HSQC NMR spectra of the RGS8 proteins (WT, Cys¹⁰⁷, and Cys¹⁶⁰) were obtained. In the WT spectrum (Fig. 4-2A), peak count corresponds well with the expected ^1H - ^{15}N correlations. The wide chemical shift dispersion, well-defined peaks with roughly uniform intensities and line shapes reflect a folded, homogeneous protein tertiary structure.

To probe the effect of ligand binding with RGS8, WT RGS8 (^{15}N -enriched) was mixed with CCG-203769 (unlabeled) at 1:1, 1:2, and 1:4 RGS8:ligand ratio. The ligand induced changes in the RGS8 ^1H - ^{15}N HSQC spectra (Fig. 4-2C and 4-3A) and chemical shift perturbations increased in magnitude with increasing concentrations of compound, indicating binding of CCG-203769 to the protein. Once the assignments of the RGS8 spectra are completed, we will identify which residues in RGS8 are perturbed in response to the small molecule inhibitor.

The titration of the Cys¹⁰⁷ RGS8 with the ligand also yielded high quality spectra (Fig. 4-3B). Similarly, several peaks in Cys¹⁰⁷ RGS8 were perturbed in response to CCG-203769, indicating that CCG-203769 can also act at Cys107. Interestingly, many of the same peaks that were perturbed in WT RGS8 were also perturbed in Cys¹⁰⁷ RGS8 (Fig. 4-3B and D), indicating that CCG-203769 affects protein conformation similarly between WT and Cys¹⁰⁷ RGS8. This suggests that Cys107 likely is involved in inhibition of WT protein function.

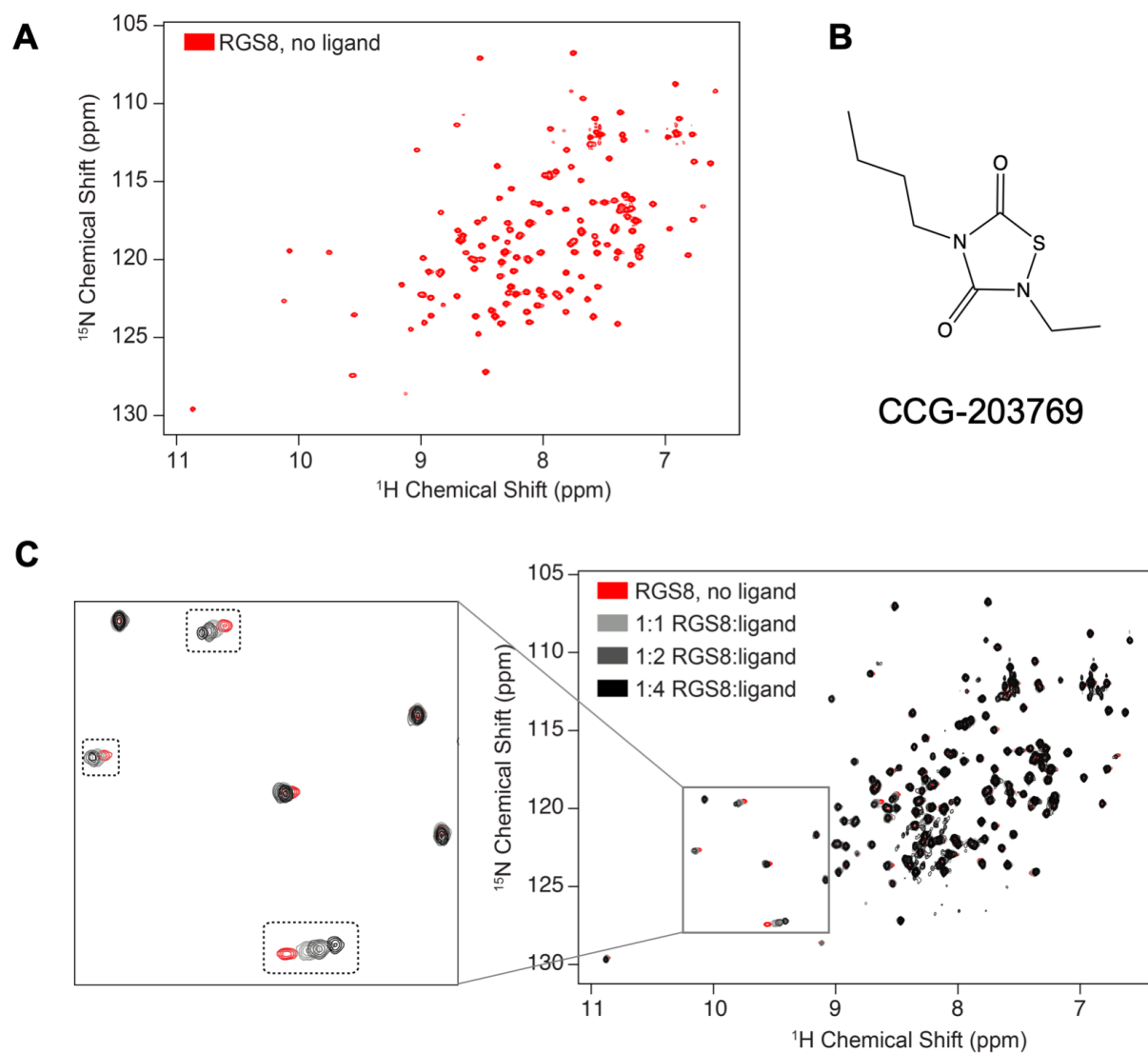


Figure 4-2: WT RGS8 protein NMR spectra. (A) ^1H - ^{15}N HSQC NMR spectrum of WT RGS8. (B) The structure of ligand CCG-203769. (C) Overlay of ^1H - ^{15}N HSQC NMR spectra of WT RGS8 before (red spectrum) and after the addition of its ligand CCG-203769 at 1:1, 1:2, and 1:4 RGS8:ligand ratio (grey spectra). Shifted residues are highlighted in the zoomed spectrum. Spectra were acquired at 25 °C on a Bruker AVANCE III HD 800 MHz NMR spectrometer equipped with a TCI Cryoprobe at the CUNY Advanced Science Research Center NMR facility.

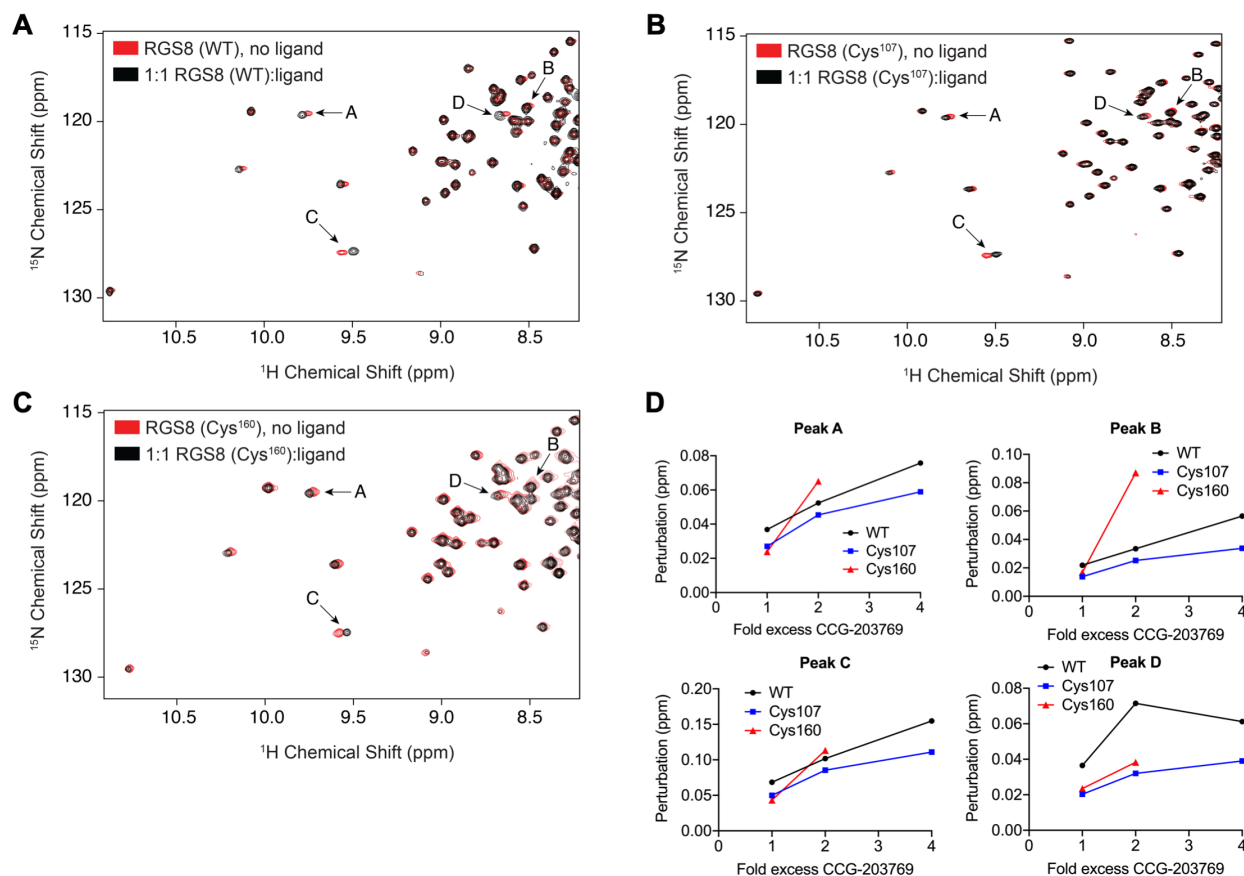


Figure 4-3: Chemical shift perturbation of WT and single-cysteine RGS8 protein NMR spectra upon the addition of ligand CCG-203769 ^1H - ^{15}N HSQC NMR spectra of RGS8 were overlaid before (red spectrum) and after the addition of its ligand CCG-203769 at 1:1 RGS8:ligand ratio (black spectra) for (A) WT RGS8 (B) Cys¹⁰⁷ RGS8, and (C) Cys¹⁶⁰ RGS8. (D) The magnitude of chemical shift perturbation. Spectra were acquired at 25 °C on a Bruker AVANCE III HD 800 MHz (WT and Cys¹⁰⁷ RGS8) or a Bruker AVANCE III HD 700 MHz (Cys¹⁶⁰ RGS8) NMR spectrometers equipped with Cryoprobes at the CUNY Advanced Science Research Center NMR facility.

Cys¹⁶⁰ RGS8 is more sensitive to compound-induced denaturation than WT.

To determine effects of CCG-203769 on protein conformation mediated by the α_7 cysteine, Cys¹⁶⁰ RGS8 was also exposed to compound in ¹H-¹⁵N HSQC NMR studies. Some chemical shift perturbations were observed at 1:1 and 1:2 molar ratios of Cys¹⁶⁰ RGS8 to CCG-203769. However, higher concentrations of compound resulted in signal loss at protein:ligand ratios above 1:2, most likely due to protein denaturation. The loss of signal with increasing concentrations of CCG-203769 precluded measurement of peak perturbations. (Fig. 4-3D) The decrease in Cys¹⁶⁰ RGS8 stability in the presence of CCG-203769 compared to WT RGS8 is surprising, given that there is only one cysteine at which the compound may act. However, this fits with data demonstrating that Cys¹⁶⁰ RGS8 is more potently inhibited than WT in functional inhibition studies with CCG-203769 (Fig. 4-4) and with CCG-50014.¹¹⁵

Functional inhibition by CCG-203769 is altered in cysteine mutants

Previous studies have illustrated that manipulation of cysteines alters potency of inhibition by covalent modifiers. In RGS4, removal of individual cysteines leads to a decrease in potency of inhibition by CCG-4986³¹ and by CCG-50014.¹⁴⁴ Interestingly, however, mutation of RGS8 to the single cysteine Cys¹⁶⁰ has been shown to cause an increase in the potency of CCG-50014 compared to WT, while mutation to the single cysteine Cys¹⁰⁷ RGS8 caused a decrease in potency.¹³³ This is consistent with the observation that Cys¹⁶⁰ RGS8 is more prone to denaturation in response to compound exposure in NMR studies.

To test whether CCG-203769 acting at individual cysteines inhibits RGS-G α binding similarly to CCG-50014, a flow cytometry-based protein-protein interaction assay was used. As expected, the Cys¹⁰⁷ RGS8 mutant was minimally inhibited by CCG-203769, retaining 87% of G α

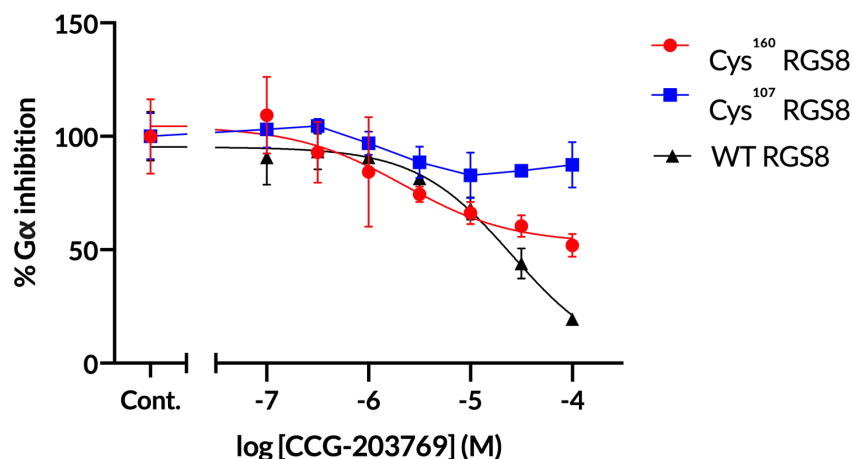


Figure 4-4: Inhibition of RGS-G α binding for WT, Cys¹⁶⁰, and Cys¹⁰⁷ RGS8 in response to increasing concentrations of CCG-203769 was measured by FCPIA. WT IC₅₀ = 25 μ M), Cys¹⁶⁰ IC₅₀ = 2.2 μ M, and Cys¹⁰⁷ was not inhibited.

binding at 100 μ M CCG-203769, while WT RGS8 only had 19% binding remaining at that concentration. Cys¹⁶⁰ RGS8 showed only partial inhibition of G α binding (retaining 52% G α binding) even at the highest CCG-203769 concentration used (Fig. 4-4). CCG-203769 inhibited Cys¹⁶⁰ RGS8 with an increased potency (IC₅₀ = 2.2 μ M) compared to WT (IC₅₀ = 25 μ M), which is consistent with that previously seen with CCG-50014.¹³³ In addition, while Cys¹⁶⁰ RGS8 was inhibited, Cys¹⁰⁷ RGS8 was minimally affected even at the highest concentration of CCG-203769 (Fig. 4-4). This suggests that Cys160 is more readily acted upon than Cys107.

CCG-203769 induces an intra-protein disulfide in WT RGS8.

No mass adduct was directly observed upon incubation of CCG-203769 with RGS8, so an excess of IAA was used to label free cysteine thiols. In protein not treated with CCG-203769, IAA caused a mass increase of 114.5 Da (2 times the mass of the acetamide adduct), indicating that iodoacetamide accesses and forms an adduct at both cysteines. When protein was pretreated with CCG-203769, this 114.5 Da mass increase was largely absent, indicating that CCG-203769 protects

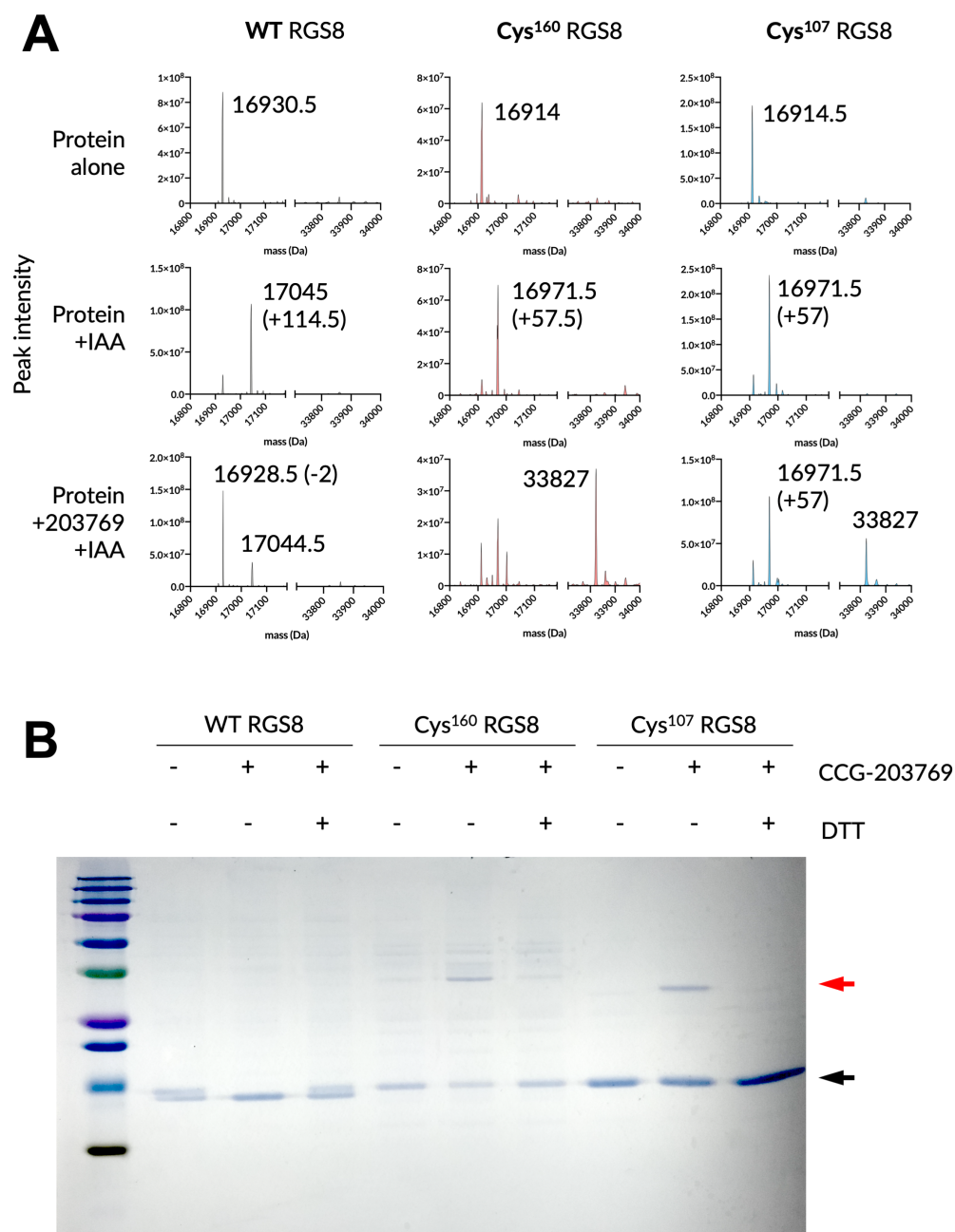


Figure 4-5: CCG-203769 masks cysteine alkylation by IAA by inducing disulfide bond. (A) Deconvoluted mass spectra of WT RGS8 (first column), Cys¹⁶⁰ RGS8 (second column), and Cys¹⁰⁷ RGS8 (third column). Spectra were taken before treatment (first row), after excess of IAA (second row), and pretreated with CCG-203769 before addition of IAA (third row). (B) WT, Cys¹⁶⁰, and Cys¹⁰⁷ RGS8 mass changes analyzed by SDS-PAGE after treatment with vehicle, 250 μ M CCG-203769, or CCG-203769 followed by 1 mM DTT. Monomer mass indicated with black arrow and dimer mass indicated with red arrow.

RGS8 from IAA alkylation. Surprisingly, however, a peak of the expected mass of a protein-compound adduct with CCG-203769 was not detected. In fact, there was a 2 Da decrease from 16930.5 to 16928.5 (Fig. 4-5A). This suggests that two hydrogens were lost, which is consistent with the formation of a disulfide bond between the two cysteines.

Mass accuracy in protein MS is only about 0.01%,^{159,160} so it is difficult to draw definitive conclusions from a 2 Da loss in mass in a 17 kDa protein. Another method for detecting disulfides is labeling of free cysteines with IAA; cysteines participating in a disulfide bond are unavailable for alkylation. Pretreatment of RGS8 with CCG-203769 only partially protects the protein from IAA alkylation (Fig. 4-5A). However, in the smaller population of protein with cysteines accessible to IAA, both cysteines were alkylated, with no population of protein having a mass corresponding to a single alkylation (Fig. 4-5A). This “all-or-nothing” response to alkylation of the two cysteines by IAA after CCG-203769 pretreatment is also consistent with induction of a disulfide bond in the population of protein that was not modified by IAA.

Among single-cysteine RGS8 mutants (Cys¹⁰⁷ and Cys¹⁶⁰), CCG-203769 induces dimerization via an inter-protein disulfide.

To determine how CCG-203769 may act differently on distinct cysteines in RGS8, the single-cysteine mutant proteins Cys¹⁰⁷ RGS8 and Cys¹⁶⁰ RGS8 were also tested to determine whether CCG-203769 protected individual cysteines from IAA adduct formation. As expected, treatment of proteins with an excess of IAA caused an increase in mass of 57 Da in both Cys¹⁰⁷ and Cys¹⁶⁰ RGS8, consistent with a single alkylation at each mutant’s only cysteine. When each single-cysteine protein was pretreated with CCG-203769, there was a decrease in the amount of protein alkylated by IAA, but no corresponding increase in the mass of unmodified protein. Instead, a mass appeared that corresponds to two times the mass of the protein (Fig. 4-5A). This

suggests that compound may be inducing a covalently linked dimer.

To test whether the dimer-inducing effect of CCG-203769 is mediated by a disulfide bond, protein treated with compound was analyzed by SDS-PAGE with reducing agent absent from the sample buffer. As observed by mass spectrometry, addition of CCG-203769 caused a dimer-sized mass in single-cysteine but not WT RGS8. This was reversed by addition of dithiothreitol to the CCG-203769-treated Cys¹⁰⁷ and Cys¹⁶⁰ RGS8 (Fig. 4-5B), consistent with a dimer mediated by a disulfide bond. A slight difference was observed between band positions of Cys¹⁰⁷ and Cys¹⁶⁰ RGS8 dimer masses (Fig. 4-5B). This is likely due to differently positioned disulfides altering the shape of the denatured protein, resulting in a gel shift.

Both single-cysteine mutants were sensitive to compound-induced dimerization. However, the Cys¹⁰⁷ RGS8 was only partially dimerized in response to CCG-203769 addition, while a large population was still alkylated by IAA. Cys¹⁶⁰ RGS8 pretreated with CCG-203769 had a larger proportion of the dimer mass (Fig. 4-5A). This, combined with data indicating that Cys¹⁶⁰ RGS8 is more susceptible to compound-induced denaturation, suggests that Cys160 is more readily dimerized than Cys107. This also fits with data indicating that the α_7 cysteine is more readily alkylated than the α_4 cysteine in RGS4 (Fig. 4-1).

CCG-203769 induces inter-protein disulfide in RGS4

To determine whether CCG-203769 exhibits disulfide-inducing behavior against other RGS proteins, RGS4 and RGS19 were also tested for an increase in size mediated by CCG-203769. RGS19 has only one cysteine, Cys123, the α_4 cysteine analogous to Cys107 in RGS8. As anticipated, it behaves much like Cys¹⁰⁷ RGS8; CCG-203769 induces a mass double the size of monomer protein, and this is reversible by addition of 1 mM DTT (Fig. 4-6). Interestingly, RGS4 also formed

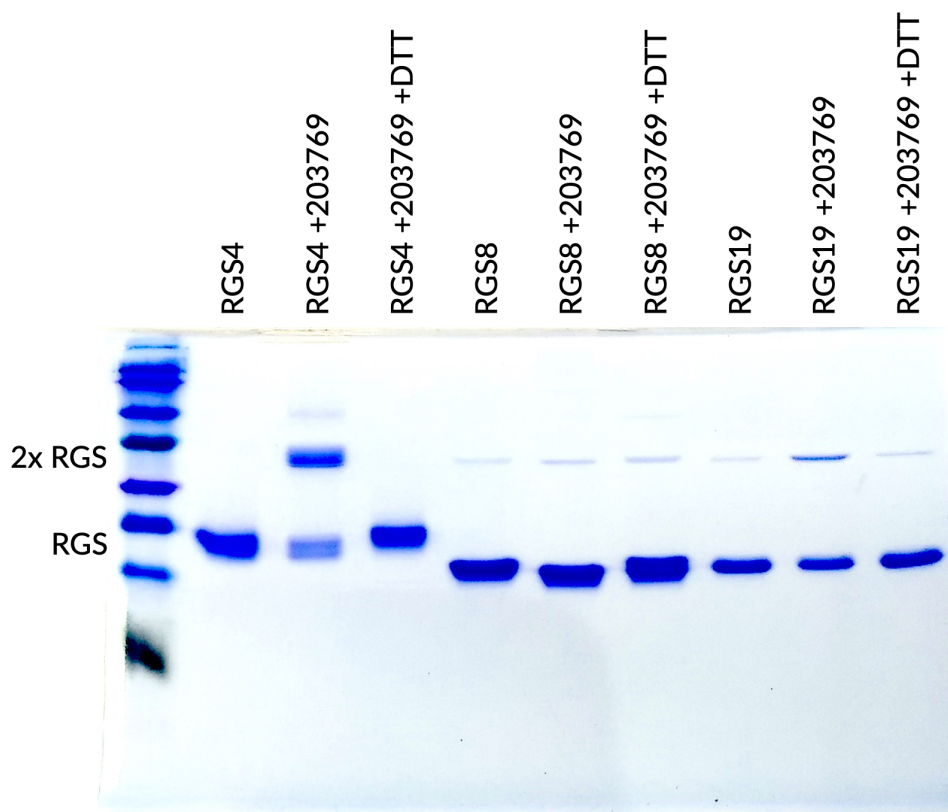


Figure 4-6: RGS4, RGS8, and RGS19 mass changes analyzed by SDS-PAGE after treatment with vehicle, 250 μ M CCG-203769, or CCG-203769 followed by 1 mM DTT.

a dimer after addition of CCG-203769, despite having multiple additional cysteines. These also were reversible by DTT, indicating they are disulfide-mediated. Of the proteins tested, RGS4 was the only protein with multiple cysteines that was susceptible to covalent dimer formation. It is possible that this is mediated by one or more of the cysteines unique to RGS4, namely Cys71 on α_3 and Cys132 on α_6 .

Conclusions

This work demonstrates that both of the conserved cysteines in RGS proteins can play a role in inhibition by CCG-203769. Both Cys¹⁰⁷ and Cys¹⁶⁰ RGS8 exhibited chemical shift perturbations in response to CCG-203769 in ¹H-¹⁵N HSQC studies. They were also prone to compound-

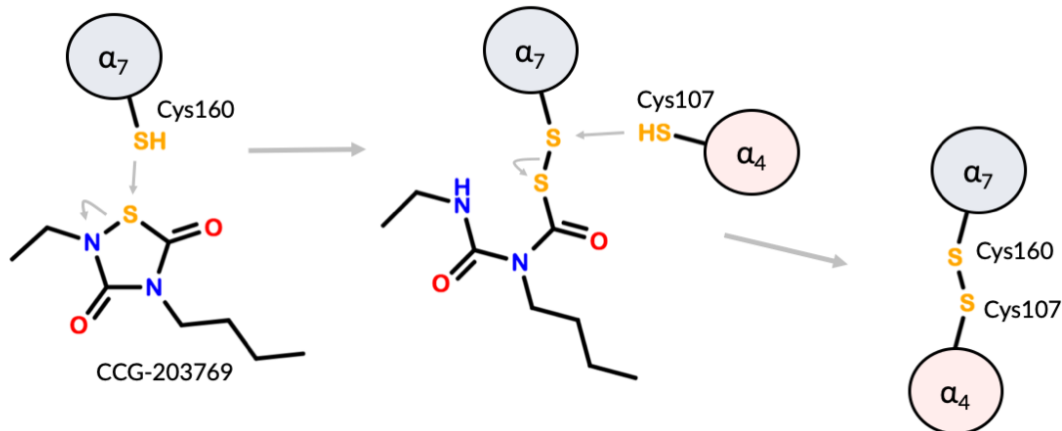


Figure 4-7: Proposed mechanism of disulfide bond induction by CCG-203769 in RGS8

induced formation of a disulfide-linked dimer. The α_7 cysteine is more readily alkylated than the α_4 cysteine in RGS4 (Fig. 4-1) and mutant RGS8 containing only this cysteine (Cys160) was more susceptible to both denaturation and inhibition of G α binding than was Cys¹⁰⁷ RGS8. This, combined with earlier data showing that the α_7 cysteine has more solvent exposed surface area in MD simulations of RGS4 and RGS8,¹⁴⁴ suggests that the α_7 cysteine is more likely to be the site of primary compound adduct formation than the α_4 cysteine. Importantly, CCG-203769 was found to be capable of inducing a disulfide bond between free cysteines, a mechanism not previously known to be effected by these compounds. In single-cysteine proteins, this effect resulted in formation of a disulfide-linked dimer which likely causes the instability observed in Cys¹⁶⁰ RGS8.

We propose a model in which CCG-203769 first forms an adduct at the more accessible α_7 cysteine. The protein-compound adduct contains a disulfide bond which is then displaced by the free thiol of the α_4 cysteine (Fig. 4-7), leaving an internal disulfide bond between the α_4 and α_7 helices. In single-cysteine RGS proteins, a similar mechanism may mediate disulfide formation with initial CCG-203769 adduct taking place on the only available cysteine, and the disulfide may

be formed by a thiol from a separate molecule.

Interestingly, these compounds have activity in cells and even in *in vivo*.^{112,117} This occurs despite a reductive intracellular environment due to the presence of glutathione and other reducing agents.^{161–163} It is possible that in the cell, the disulfide-forming behavior of these compounds is reversible. Compound-induced dimers between single-cysteine RGS proteins likely are an artifact of the *in vitro* environment, where the only free thiols available for disulfide formation are on other RGS proteins. However, it is possible that an intra-protein disulfide between Cys107 and Cys160 in WT RGS8 is responsible for a conformational change in RGS8 that prevents the RGS-G α interaction, both *in vitro* and in living systems. Further work will be necessary to determine whether a disulfide bond mediates inhibition by CCG-203769 in the cellular environment, how such a disulfide affects activity of other RGS proteins, and whether other cysteine-modifying inhibitors may act by the same mechanism.

CHAPTER 5:

Identification of Transient Pockets in RGS4 and RGS19

Vincent Shaw performed pocket identification, analysis, and clustering. Hossein Mohammadi-arani performed MD simulations and Mohammadjavad Mohammadi prepared trajectory files.

Introduction

Protein dynamics play a critical role in molecular recognition.^{3,164} Whether a ligand selects a specific protein conformation that alters its function or induces a functional change by pushing a protein into a unique conformation, proteins do not remain stationary in solution. The dynamic motions of a target protein play a key role in the specificity of its ligand.^{7,165,166}

In some cases, pockets that might be capable of binding small molecules are not visible in available structures, but exist in conformations taken by the protein in solution.^{8,122} It may be possible to design drugs that take advantage of transient pockets, which are not present in static structures but are sampled by a protein in solution; or cryptic pockets, which become apparent once a ligand is bound.⁵⁰ There are several proteins with dynamically fluctuating pockets that have been targeted using virtual screens informed by molecular dynamics.^{32,167}

Heterotrimeric G-protein signaling is pathway of enormous pharmacological significance, with a high proportion of known drugs targeting G-proteins or related pathways.¹² Agonist binding to a GPCR results in dissociation of the GDP nucleotide of the G-protein alpha subunit. This allows GTP to bind, putting the G-protein in its active conformation. The $G\alpha$ and $G\beta\gamma$ subunits dissociate, each mediating downstream signaling. A key part of this pathway is termination of signaling, which is catalyzed by Regulators of G-Protein Signaling (RGS) proteins. These bind to the active $G\alpha$ subunits and accelerate hydrolysis of the GTP to GDP, allowing a return to the inactive form and re-recruitment of $G\beta\gamma$.

Because of their status as critical component of the G-protein cycle, RGS proteins make attractive drug targets. There are 20 canonical RGS isoforms, with yet more proteins having RGS homology domains. Each has different tissue distributions and diverse physiological roles. Identification of inhibitors with high isoform specificity will permit targeting of certain pathways

and disease states with reduced off-target effects. While existing inhibitors that are selective for RGS4 may be useful in treatment of Parkinson's disease,^{109,117} there may be uses for targeting of other RGS proteins as well. One potential target is RGS19, which has been implicated in pain regulation¹⁰² and depression.¹⁰¹

Several series of RGS inhibitors have been already identified, and all act by covalent modification.^{113,114} Interestingly, these inhibitors do show isoform specificity, but all are selective for RGS4 and/or RGS1. Most likely, this specificity is largely due to differences in the number of cysteines among RGS isoforms; both RGS4 and RGS1 have additional cysteines that are not well-conserved among RGS proteins.¹¹⁸ While this inhibition of RGS4 or RGS1 may be desirable for treatment of some disease states, it may make targeting other RGS proteins with a smaller cysteine complement impossible as long as we are limited to use of covalent modifiers. Discovery of non-covalent RGS inhibitors may open the doors to drugs with novel specificities in addition to reducing any toxicity risk associated with the use of covalent drugs.

Previously, we identified a role for protein dynamics in the specificity of inhibitors acting at a cysteine on the α_4 helix.^{129,144,146} MD studies from this work suggested that the structure of the RGS domain may open sufficiently to allow covalent inhibitors to access this otherwise buried cysteine. Also, differences in flexibility among isoforms play a role in driving inhibitor selectivity.¹⁴⁴ If the apo-protein forms a binding pocket with sufficient frequency and druggability, it may be exploited in the design of non-covalent inhibitors. In this work, we seek to identify transient binding pockets in RGS proteins. This will permit discovery of new compounds by virtual screening.

Approach and Results

Pocket Identification

Previous studies have showed flexibility in RGS4 and RGS19.¹⁴⁴ In MD simulations, RGS4 showed a pronounced movement in the α_6 helix, in which it partially lost its helical structure and moved away from the helical bundle. RGS19 also showed a dramatic movement, in which the α_6 - α_7 loop separates from the α_4 and α_5 helices, creating a groove.¹⁴⁴ In either case, compound access may be permitted by the development of a transient pocket. If a pocket conformation can be identified that is both amenable to small-molecule binding and frequently occurring, it will be useful in the rational design of non-covalently binding inhibitors.

Pockets were found using MDpocket,¹⁶⁸ a part of the Fpocket suite of pocket-finder software.¹⁶⁹ Fpocket defines pockets by filling cavities with alpha spheres: spheres with an external boundary touching four atoms, with no atoms inside them.¹⁷⁰ Sphere radii are restricted to between 3 and 6 Å to prevent large spaces (like external surfaces) or small spaces from being included as part of pockets.¹⁶⁹

In RGS4, frequent pockets were found to occur between the α_5 and α_6 helices (Fig. 5-1). This was expected, considering the flexibility and movement previously observed in α_6 of RGS4.¹⁴⁴ In RGS19, MDpocket most frequently identified pockets formed by atoms in the α_6 - α_7 loop (Fig. 5-1). This makes sense, considering the groove formed where the α_6 - α_7 loop separates from the rest of the α_4 - α_7 helix bundle.¹⁴⁴ Because pockets were identified with the highest frequency on these parts of the structures, these locations were used for extracting descriptors of pocket shape and characteristics, clustering of pocket conformations, and choosing a state for virtual screening.

Descriptors of the pocket for each frame of the trajectory were generated by MDpocket.

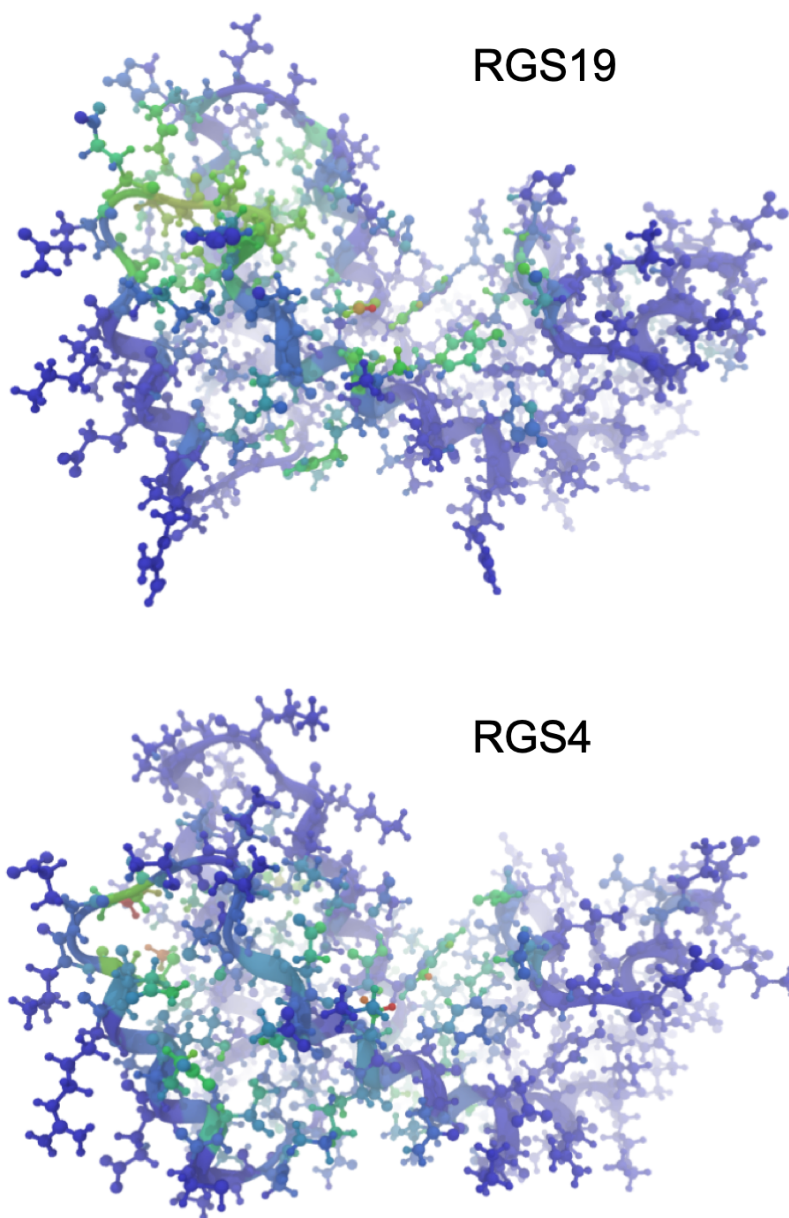


Figure 5-1: Locations of pocket-forming residues in RGS4 (top) and RGS19 (bottom). Color indicates frequency with which each atom touches a pocket alpha sphere. Blue is less frequent and red is more frequent.

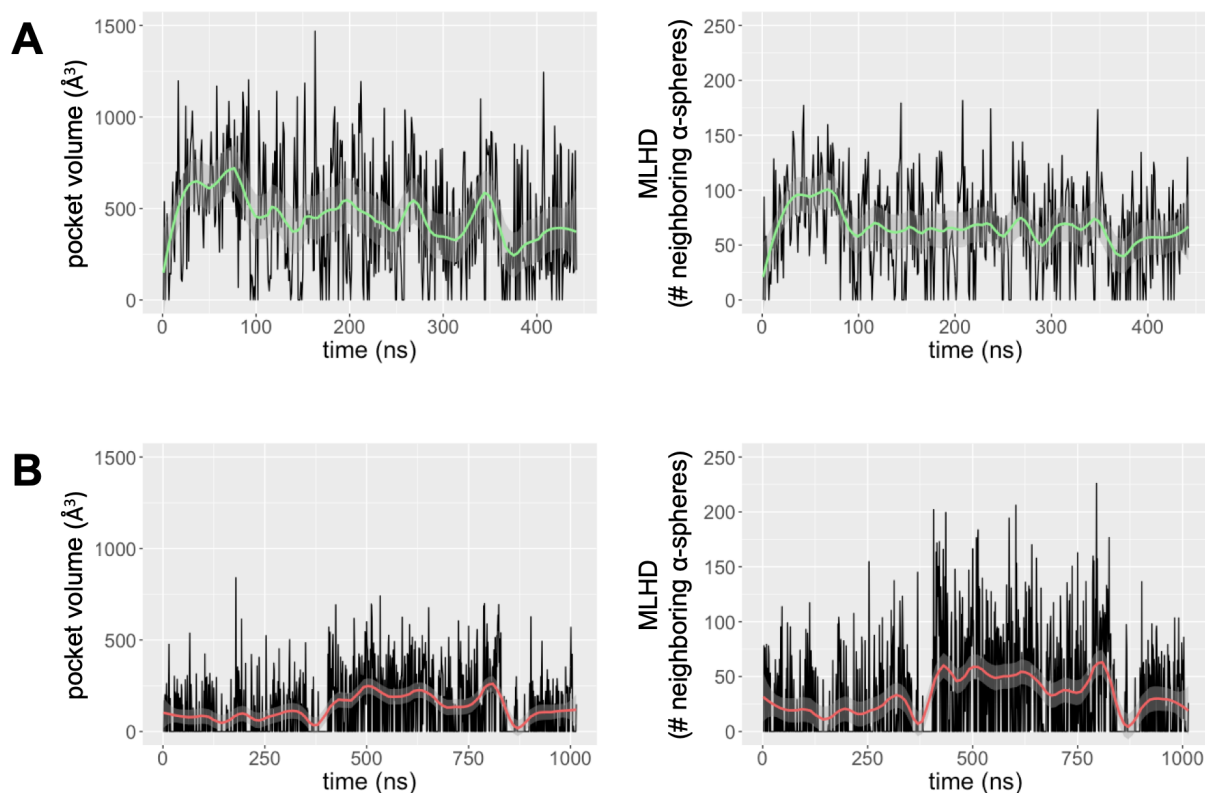


Figure 5-2: Pocket volume and mean local hydrophobic density (MLHD) plotted over the simulation trajectory for RGS19 (A) and RGS4 (B). Pockets in RGS19 were larger and more frequent than those in RGS4.

By plotting these parameters over time, trends in the pocket's size, shape, and druggable potential can be identified.

Pocket Clustering

To use a conformation from an MD simulation in a virtual screen, we wanted to choose a conformation that was both druggable (i.e. amenable to small molecule binding) and representative of a frequently occurring state. One option is to choose a state that has the highest mean local hydrophobic density, an index closely correlated with a pocket's druggability.^{168,171} Although there is a precedent for a successful virtual screen being performed with such a strategy,¹⁶⁷ it may not be ideal for choosing a transient pocket from a long time scale simulation. Even if a compound

were identified that had a favorable energy of binding in the static structure of the most druggable conformation sampled in a long time scale simulation, if that state was very rarely sampled by the protein in solution it might have a low on-rate and therefore low affinity. Therefore, we clustered pocket states to identify populations of distinct conformations taken by the protein in solution. By choosing states that are most similar to cluster centers, we identify states that are most representative of a conformational group, ensuring use of a state that is druggable but not anomalous.

Clustering of pocket states was performed using R and Rstudio. First, pocket descriptors were normalized to equal scales to avoid uneven weighting of pocket parameters. The pocket states were clustered based on seven pocket descriptors: pocket volume, nonpolar surface area, polar surface area, number of alpha spheres, average alpha sphere radius, maximum distance between any two alpha spheres, and mean local hydrophobic density. These descriptors were chosen to separate states based on their size, shape, and complexity. Mean local hydrophobic density is a measurement of how densely packed hydrophobic areas are, based on the degree with which nonpolar alpha spheres overlap one another. This descriptor has been found to be closely related to druggability.^{168,171}

The NbClust package¹⁷² was used to determine the optimal number of clusters. In order to limit the number of groups of pocket states, the cluster numbers were limited to between 5 and 10. In the RGS4 trajectory, the optimal number of clusters was 6 according to a plurality of indices, while in RGS19, the optimal number of clusters was 7. States were clustered using kmeans clustering.

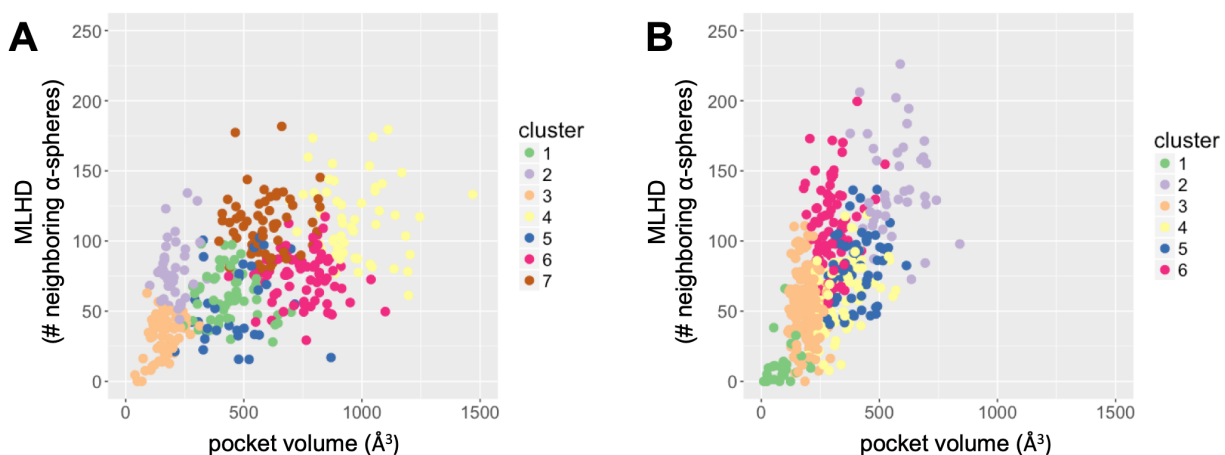


Figure 5-3: Clustering of pocket states for RGS19 (A) and RGS4 (B). Volume is plotted against MLHD, and color indicates distinct clusters. An ensemble of pockets representing clusters with high MLHD and a variety of pocket volumes were selected for structure based screening.

Frames for screening

For structure-based screening, it will be beneficial to use states that are amenable to small molecule binding. Because mean local hydrophobic density has been found to be a strong indicator of small-molecule binding potential,^{168,171} clusters that are high in this index were chosen. In addition, pocket volume may play a role in which compounds may bind. The volumes of pockets in a large set of ligand binding proteins in the protein data bank has been found to cover a wide variety of pocket volumes.¹⁷³ The median volume was 536 Å³, with first and third quartiles at 375 and 715 Å³ respectively,¹⁷³ suggesting that pockets that are excessively large or small may not make ideal drug targets. As such, the pocket populations that form clusters 4 and 7 in RGS19 and clusters 2 and 6 in RGS4, which have high MLHDs and differing but moderate pocket volumes, may make the most promising populations of pockets for virtual screening.

States most similar to the cluster center will best represent the actual pocket conformation in solution. The transient pocket conformations shown in Figure 5-4 are the trajectory frames most similar to cluster centers 2 and 6 in RGS4 and 4 and 7 in RGS19. These states will go on to

be used for virtual screening to identify non-covalent compounds that bind to RGS4 and RGS19.

One limitation of virtual screening methodology is usage of static structures, where the conformation used may not be representative of those occurring in solution. The approach used in this work partially addresses this problem by clustering populations of similar pocket-like states, and choosing only the most representative conformation of each cluster for use in virtual screening. However, compound docking methods still often use a static protein structure. Even if the conformation used realistically occurs in solution, effects of the compound binding on the protein structure are not accounted for. This may bias screening results toward compounds that bind in an conformational selection mode rather than an induced fit mode, resulting in a smaller proportion of hits that validate experimentally. It may be possible to resolve this issue by using flexible docking, in which protein flexibility is taken into account. This may improve the quality of hits and likelihood of successful identification of new chemical matter by virtual screen.

Conclusions

This work identifies transient pockets between α_5 and α_6 helices of RGS4 and RGS19. This will enable identification of compounds that bind non-covalently by virtual screen. Rather than consideration of static structures only, which may not be representative of conformations taken by a protein in solution, this work chooses druggable, transient pockets that are representative of frequently occurring conformations, and may be useful in the discovery of non-covalent compounds.

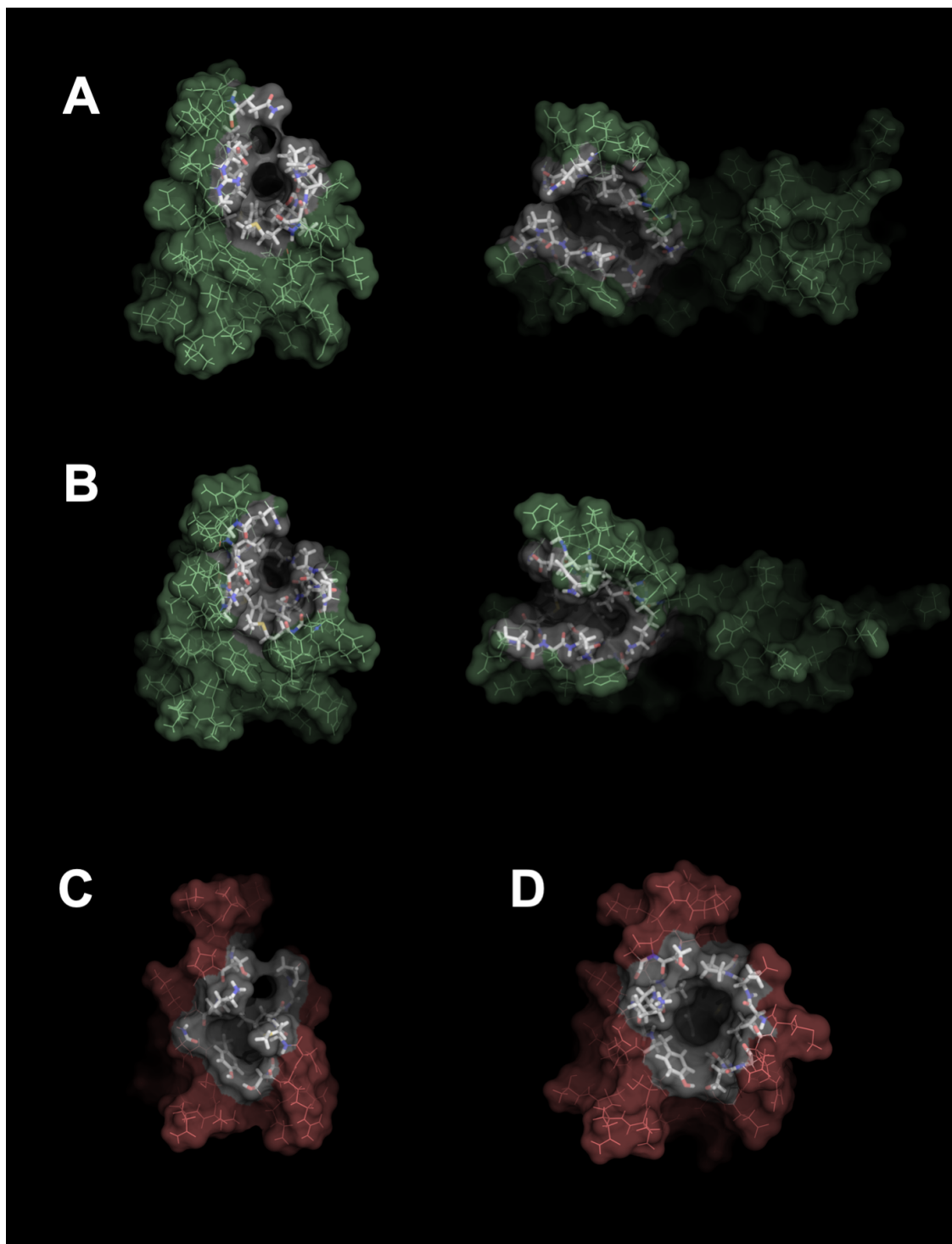


Figure 5-4: Pocket states that are representative of cluster 4 and 7 in RGS19 (A and B) and cluster 2 and 6 in RGS4 (C and D). Pocket-forming atoms illustrated with white surface.

CHAPTER 6:
Conclusions and Future Directions

In this thesis, I aimed to understand the drivers of selectivity of covalent modifiers of RGS proteins. This led to the pursuit of two main hypotheses: that RGS inhibitor isoform specificity is determined by cysteine complement, and that RGS inhibitor isoform specificity is determined by protein flexibility. While these hypotheses at first appear to be in conflict, this body of work demonstrates that both are essential pieces of the full picture of isoform selectivity.

Role of individual cysteines in action of inhibitors

A primary determinant of selectivity among covalent modifiers of RGS proteins is the number and location of cysteines in the RGS domain. Previous work has shown that conserved cysteines on the α_4 and α_7 helices each are capable of mediating inhibition by CCG-50014.¹¹⁵ In Chapter 4, I propose that the TDZD CCG-203769 induces a disulfide bond between these two cysteines. NMR studies have demonstrated that CCG-203769 perturbs protein structure in both single-cysteine mutants of RGS8, indicating it can act at either cysteine. Mass spectrometry studies indicated that CCG-203769 could induce a dimer sized mass among single-cysteine RGS8. This was reversible by DTT, indicating it is disulfide mediated. Interestingly, without forming any apparent adduct in WT RGS8, CCG-203769 prevented iodoacetamide alkylation and induced a 2 Da reduction in mass, suggesting it may be inducing a disulfide between cysteines. This is a unique covalent interaction between the two well-conserved cysteines present in RGS8 that is induced by the inhibitor.

Role of protein dynamics in RGS inhibitor selectivity

In the course of this work, it also became apparent that there were other drivers of selectivity beyond the cysteine complement, namely protein flexibility. While the effects of differences

in flexibility on potency of inhibition may be largely masked by differences in cysteine complement among covalent inhibitors, it will be particularly important to understand protein dynamics in order to identify inhibitors that act non-covalently. By eliminating the reliance on covalent interactions, it may be possible to target RGS proteins that have fewer cysteines.

In Chapter 2, I demonstrate a correlation between protein flexibility and potency of the TDZD CCG-50014 when it acts at a single, shared cysteine. Among RGS4, RGS8, and RGS19 mutants that contain only the shared α_4 cysteine, RGS19 is most potently inhibited (IC_{50} of 1.1 μM), followed by RGS4 and RGS8 (IC_{50} s of 8.5 μM and >100 μM respectively). When solvent exposure was measured by deuterium exchange, RGS19 was found to have faster deuterium exchange in the α_4 helix, followed by RGS4 and RGS8. In addition, MD simulations shed light on movements that may lead to these differences in solvent exposure, and how cysteines may become accessible covalent inhibitors. In RGS19, MD simulations showed movement throughout the α_4 - α_7 helix bundle, opening a cleft between α_4 - α_5 and α_6 - α_7 helices. This was supported by HDX data showing high deuterium incorporation throughout this helix bundle in RGS19. Likewise, an α_7 helical cysteine in RGS4 was exposed by an outward movement of the α_6 helix in simulations, which matched the high deuterium exchange observed in this helix. Finally, RGS8 was least flexible in MD simulations and had least deuterium exchange, which may explain the limited potency with which it is inhibited.

While this work identified a correlative relationship between protein flexibility and potency of inhibition, we wanted to go beyond correlation and demonstrate that direct manipulation of flexibility could induce changes in potency of inhibition. To this end, we sought to identify interacting residues within RGS proteins that are responsible for differences in flexibility among isoforms. Of particular interest is a salt bridge network linking the α_4 helix, the α_5 - α_6 interhe-

lical loop, and the α_5 helix. This network is shared by RGS4 and RGS8, but absent in RGS19, which lacks a charged residue on α_4 , having instead a leucine. This lack of a salt bridge may be responsible for the observed flexibility of RGS19. Mutation in this residue in RGS19 to add a salt bridge-forming residue (aspartate) increased thermal stability, reduced deuterium incorporation, and, importantly, decreased potency of inhibition by CCG-50014. Conversely, mutation to remove the salt bridge in RGS4 and RGS8 increased deuterium incorporation and increased the potency of CCG-50014. This strongly supports a causative relationship between RGS protein flexibility and potency of inhibition.

Access of inhibitors to buried cysteines hinted at the existence of transient pockets. Pockets were identified from MD simulations of RGS4 and RGS19. In RGS4 a pocket opens between the α_5 and α_6 helices, and in RGS19, a cleft opened between the α_4 - α_5 and α_6 - α_7 sets of helices. In order to move forward with virtual screening targeting these pockets, conformations were clustered to ensure selection of a structure that was both druggable and representative of frequently occurring states. Identification of transient pockets will enable rational design of non-covalent inhibitors by structure-based screening.

Future research in understanding action of TDZD inhibitors

While some questions on the mode of compound action and the basis for specificity have been answered, new avenues for future research have opened. One area is in better understanding the role of the disulfide bonding induced by CCG-203769. While this may occur in RGS proteins in these assay conditions, it remains to be understood how readily such a change can affect G-protein binding and GAP activity, both in biochemical assays as well as in the cellular environment.

Experimental efforts to understand the molecular mechanism of the interaction between

RGS proteins and TDZD inhibitors is being undertaken in collaboration with Dr. Krisztina Varga at the University of New Hampshire. While we already have demonstrated that HSQC spectra in RGS proteins are perturbed by CCG-203769, it will be useful to understand which parts of the protein are altered in response to inhibitor action. To this end, I have produced uniformly labeled ^{13}C ^{15}N -RGS8, and efforts are already underway using these samples to assign HSQC spectra peaks to individual amides on the protein backbone.

One open question that remains is whether the induction of a disulfide bond by CCG-203769 is sufficient to prevent binding between RGS proteins and $\text{G}\alpha$. We would hypothesize that peaks corresponding to amides of residues near cysteines 107 and 160 in RGS will be perturbed, but it will be interesting to see if there are also peaks perturbed corresponding to residues involved in $\text{G}\alpha$ binding. MD simulations also provide a useful avenue for answering these questions, work which is currently being carried forward by the Vashisth lab at the University of New Hampshire. In particular, simulations that illustrate the effect of compound adduct at each cysteine in RGS8 will be valuable for comparison with the HSQC NMR studies of single cysteine mutants. Simulation work may also be able to shed light on how protein conformation may be altered upon induction of a disulfide bond.

It would be interesting to see whether the compound-induced disulfide in RGS8 can occur in cells. It may be possible to test whether RGS8 protein expressed in mammalian cells treated with CCG-203769 are also protected from IAA alkylation, and whether this is reversible by DTT. Finally, another open question is whether the disulfide bond-inducing behavior of CCG-203769 is relevant to other TDZDs, other covalent inhibitors, and among different RGS proteins. For example, RGS19 is inhibited by CCG-203769 and other TDZDs, but because it lacks a cysteine on the α_7 helix, this cannot be mediated by an intraprotein disulfide. It remains to be seen whether

an adduct between a TDZD and a single-cysteine RGS proteins such as RZ family member is maintained or displaced by a free thiol, in both in vitro and in cell environments.

Continuing discovery of non-covalent inhibitors

Research should continue in discovery of non-covalent inhibitors. In collaboration with the Dickson Lab, virtual screening efforts are under way using the transient pockets described in Chapter 5. Using a pharmacophore-based screening campaign, a library of compounds will be extracted from the Zinc library that block interactions between residues that make contact in the closed state but are separated in the open-like state. These compounds will then be docked against the open states in a structure-based screen using Schrödinger Glide. Because docking against a static structure does not account for protein movement induced by compound docking, this may bias discovery against compounds that bind in an induced-fit-like mode. To combat this, hits may be further refined with flexible docking, in which movement of both protein and compound are simulated during binding. Finally, hit compounds will be ordered or synthesized and tested for activity in inhibition of G α binding or inhibition of GAP activity. This work may yield new non-covalent inhibitors that take advantage of transient pockets in RGS proteins that we have defined here.

In conclusion, this work shows a dual role for number of cysteines and protein dynamics in specificity of RGS protein activity. Understanding drivers of RGS protein selectivity will allow future researchers to better predict the action of current inhibitors as well as develop chemical matter with new specificities. By laying out a path forward for targeting a transient pocket in RGS proteins, we may be able to break the cysteine dependence of RGS inhibitors, allowing novel selectivities and opening the doors to new applications as chemical probes or therapeutics.

APPENDIX

**Interpreting Hydrogen-Deuterium Exchange Events in Proteins Using Atomistic
Simulations: Case Studies on Regulators of G-protein Signaling Proteins**

Reprinted with permission from *J. Phys. Chem. B* 2018, 122, 40, 9314-9323
Copyright 2018 American Chemical Society

Hossein Mohammadiarani*, Vincent Shaw*, Richard R Neubig, Harish Vashisth

*Co-first authors

H.M. performed MD simulations and developed computational models. V.S. expressed protein and performed HDX-MS.

Introduction

Hydrogen-deuterium exchange (HDX) is a widely used protein labeling reaction in which an amide hydrogen in the backbone of amino acids in proteins is exchanged with a deuterium atom. To probe the locations of exchanged hydrogens in the protein backbone, HDX is often accompanied by other techniques including nuclear magnetic resonance (NMR) spectroscopy and mass-spectrometry (MS).¹⁷⁴ HDX methodologies have been successfully applied to understand protein-protein interactions,^{175–177} conformational changes in proteins,^{178–182} protein folding,¹⁸⁰ and ligand binding.^{183,184} Early applications of HDX on the A-chain of hormone insulin showed that intramolecular hydrogen bonds were a hindrance for hydrogen exchange because of their role in stabilization of the helical structure.¹⁸⁵ Since then many investigations have been conducted to characterize the mechanism of exchange events. These include studies of: deuterium exchange of poly-DL-alanine in aqueous solution at different temperatures and pH,^{186,187} the influence of residue side chains on the HDX rate of peptide groups,¹⁸⁸ modeling amides and peptides in a chemical exchange step,^{189–191} development of empirical rules for acid and base catalytic rate constants,^{188,192} development of general models for recognizing hydrogen exchange process between the folded states and the unfolded states using temperature variation,^{193–197} the negative effect of static solvent accessibility on exchanging protons,¹⁹⁸ and the correlation between apparent adiabatic compressibility and hydrogen exchange rates.¹⁹⁹ Bai et al.²⁰⁰ carried out experiments to formulate inductive and steric blocking effects of neighboring amino acids on the amide group hydrogen exchange. Their comprehensive dipeptide models included all 20 amino acids and have informed values of intrinsic kinetic rates used in many previous studies.^{130,131,201} The qualitative and quantitative interpretation of HDX events is becoming an increasingly important tool for studying dynamics in proteins which are challenging to study using other experimen-

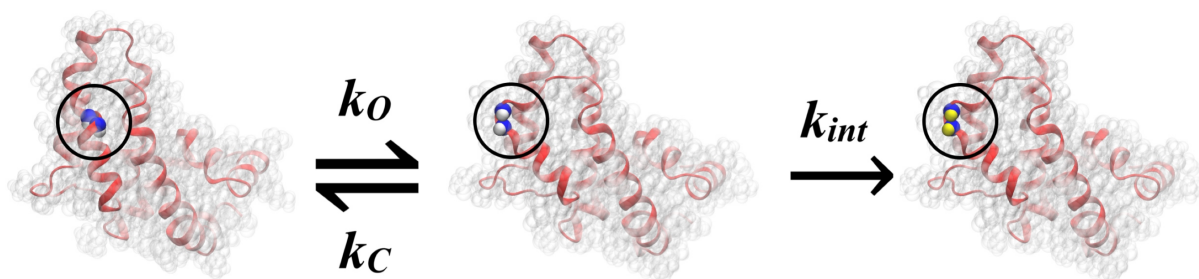


Figure A-1: Kinetic scheme for HDX is highlighted. A conformational fluctuation in the protein exposes buried amide groups (blue) (closed state) to solvent (open state) where amide hydrogens (white) are exchanged by deuterium (yellow) with an intrinsic rate constant k_{int} .

tal methods.^{130,202–204} These investigations, over the past half-century, have resulted in various interpretations of the HDX mechanism^{204–207} primarily via different models used to rationalize exchange events.^{130,131,200,208–212} The general mechanism of HDX is described by a dynamic equilibrium between closed and open states (Figure A-1) of amide hydrogens with rate constants k_c and k_o , respectively, and a first order reaction in the exchange competent or open state¹³⁰ (denoted as an intrinsic rate constant, k_{int} , in Figure A-1). The normal exchange mode for proteins that do not undergo global unfolding events is the EX2 exchange limit, in which $k_c \gg k_{int}$.¹⁹¹ This mechanism suggests that steric hindrance protects amide hydrogens from exchanging with deuterium. In addition to the physical protection, amide hydrogens that are involved in hydrogen-bonded (H-bonded) structures are protected and show decreased exchange rates.^{205,207,213,214} Therefore, HDX rates implicitly involve structural changes and dynamics in proteins.¹³⁰ A variety of models have been used to determine protein conformational states using Monte Carlo (MC)^{208,215} or molecular dynamics (MD)^{130,131,201,209,210,212,216–224} approaches.

In these models, the protection factor (PF) (ranging between 0 and 10^{10}) is a key parameter that correlates conformational dynamics in proteins with the overall HDX rate (k_{hdX}).²²⁵ In Table

A-1, we summarize various PF correlations for seven different models (M1 through M7) that have been proposed previously; detailed descriptions of these models are provided in the supplemental introduction. The parameters and criteria in PF correlations can be tuned either using MD simulations²¹⁰ or using structures refined from experiments (e.g. the NMR method). There are two general approaches to obtain the PFs for amide hydrogens by sampling conformations using simulation methods. In the first approach, PFs empirically correlate to metrics of the protein structure (e.g. models M1 to M6 in Table A-1). In the second approach,¹³¹ the PF is defined as a fractional population of the closed state to the open state for each amide hydrogen (e.g. model M7 Table A-1). As a complement to HDX experiments, MD simulations not only provide details on exchanging amide hydrogens, but also capture frequencies of open states which may occur on a much shorter time scale than the hydrogen exchange itself.^{131,210} As it remains challenging to conduct long time-scale atomistic MD simulations, the modeling of hydrogen exchange using MD simulations has generally been limited to coarse-grained and/or empirical models with implicit solvent.^{208,217,226} Several studies have employed short time scale MD simulations to predict HDX rates.^{130,131,201,227} To date, only Persson et al.¹³¹ used a millisecond long MD simulation²²⁸ for HDX analysis of a 58-residue protein, bovine pancreatic trypsin inhibitor (BPTI). They suggest that the mean residence times for the open states of all amides in BPTI are on the sub 100 ps time-scale.

However, the ability of existing models of PF correlations (Table A-1) to predict HDX trends, when applied to identical experimental dataset(s), is yet to be systematically analyzed. Furthermore, it would be useful to determine whether any of the existing models (based upon their default or reoptimized parameters) can faithfully distinguish differences in HDX patterns of homologous proteins. Finally, comparing the predictive performance of various models for widely used interatomic potentials (force-fields) for proteins (e.g. CHARMM and AMBER) will

Model	criteria								Protection Factor Definitions
	1	2	3	4	5	6	7	8	
M1 ^{ref. 225}	✓	✓							$\log(PF_i) = u \cdot (SA_i) + v / (HB_i)$
M2 ^{ref. 208}	✓		✓						$\ln(PF_i) = (\beta_c N_i^C + \beta_h N_i^h)$
M3 ^{ref. 209}	✓			✓					$\ln(PF_i) = (\beta_c N_i^C + \beta_h N_i^h)$
M4 ^{ref. 210}			✓		✓				$\ln(PF_i) = (\beta_c N_i^C + \beta_r (N_i^r)^{-1})$
M4 ^{ref. 212}	✓								$PF_i = (C_o NH_i^{sol} + C_c NH_i^\beta) / C NH_i^{sol}$
M6 ^{ref. 130}	✓								$PF_i = base / (1 + (\sqrt{base})^{1-NHstat_i})$
M7 ^{ref. 131}						✓	✓		$PF_i = \tau_C / \tau_O$
M8 [†]							✓	✓	$\ln(PF_i) = (\beta_s SASA_i^{-\gamma_s} + \beta_p D_i^{-\gamma_p})$
M9 [†]							✓	✓	$PF_i = \tau_C / \tau_O$

¹Hydrogen bond; ²Distance from the surface; ³# of residues in the vicinity; ⁴# of heavy atoms in the vicinity; ⁵RMSF; ⁶# of waters in the vicinity; ⁷polar atoms in the vicinity; ⁸SASA; [†]new model proposed in this work.

Table A-1: Model definitions and corresponding metrics. Among models reported in the literature are models M1 through M6 (empirical models) and the model M7 (a fractional population model). For models reported in this work, M8 is an empirical model and M9 is a fractional population model. Additional details on models M8 and M9 are provided in supporting information.

likely provide further guidance for future studies combining MD simulations and HDX experiments. In this work, we have investigated these issues by conducting a series of atomistic MD simulations of three homologous regulators of G-protein signaling (RGS) proteins (RGS4, RGS8, and RGS19) (Fig. A-2) using CHARMM and AMBER force-fields (CHARMM-FF and AMBER-FF). We compared the predictive performance of seven existing models (Table A-1 with our recently reported HDX-MS data for all three proteins,¹⁴⁴ and reoptimized parameters of these existing models for improved predictions. We also found solvent accessible surface area (SASA) as a useful metric to better predict protection factors in combination with the open-state definition of Persson et al.¹³¹ This was surprising because some existing models have reported SASA as a poor predictor. Based upon this latter observation, we derived two new models (M8 and M9; see supplemental methods and Table A-3, A-4) for better reproducing our experimentally observed HDX trends in three RGS proteins.

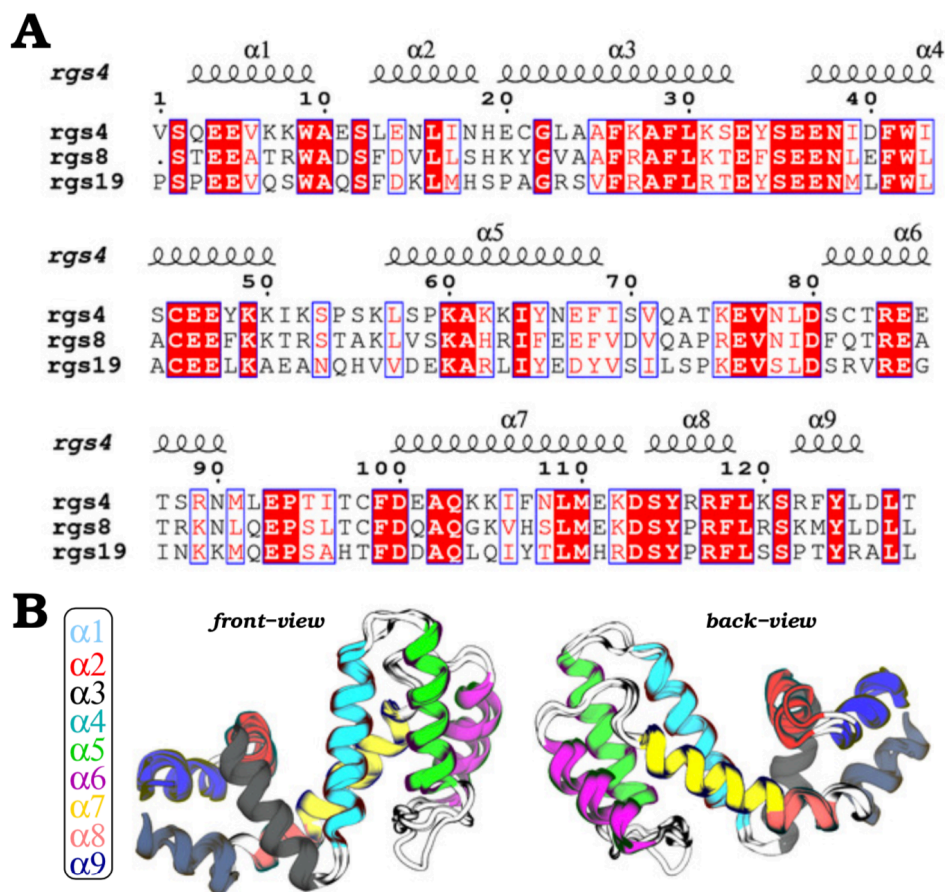


Figure A-2: Sequence and structural views of RGS proteins. (A) Sequence alignment of RGS4, RGS8, and RGS19 is shown with conserved residues highlighted in red; blue boxes indicate residues that are conserved between at least two among three RGS proteins. (B) Shown are front and back views of the overlay of RGS4 (PDB code 1AGR), RGS8 (PDB code 2ODE), and RGS19 (PDB code 1CMZ) structures with each of the nine helices uniquely colored. Regions rendered as white cartoons are interhelical loops.

Materials and Methods

We carried out all MD simulation trajectories and their analyses using NAMD and VMD software suite^{136,137} as well as python,²²⁹ and used both the CHARMM36 force-field with the CMAP correction^{138,139} and the AMBER force-field (ff14SB).²³⁰ For all MD trajectories, 50000 frames were generated for each μ s of dynamics. For RGS4 and RGS8, simulations were conducted with two different initial coordinates, while for RGS19 only one experimental structure is currently known, the coordinates of which were used in simulations. In particular, the initial coordinates for RGS4, RGS8, and RGS19, respectively, were taken from the following protein data bank entries (RGS4: 1AGR and 1EZT; RGS8: 2IHD, 2ODE; RGS19: 1CMZ). Each protein was initially modeled using the psfgen tool in VMD, and then further solvated in a simulation box (~ 65 Å x ~ 70 Å x ~ 65 Å) of TIP3P water molecules and charge-neutralized with NaCl. All system sizes are provided in Table A-2. The volume of simulation domains was then optimized in the NPT ensemble by initially applying 500 cycles of a conjugate-gradient minimization scheme followed by a short 40 ps MD run with a 2 fs time step in which the temperature was controlled at 310K using the Langevin thermostat and the pressure was controlled by the Nose-Hoover barostat. We carried out all simulations using periodic boundary conditions. These briefly equilibrated systems of all RGS proteins were further subjected to long time scale (2 μ s for each protein) MD simulations in the NVT ensemble. For all proteins and both force-fields, we generated 10 total MD simulations with 20 μ s of MD simulation data (Table A-2). All details on protein expression, purification, and data collection using HDX-MS are provided in our previous work.¹⁴⁴ Briefly, deuterium incorporation (DI) for RGS4, RGS8 and RGS19 was measured at a fragment resolution using HDX-MS experiments at $t = 1, 3, 10, 30, 100, 300,$ and 1000 minutes (Fig. A-7 and A-8).¹⁴⁴ We note that incubations were carried out in a 90% D₂O solution containing 5 mM HEPES and

100 mM NaCl. We provide further description of protocols for HDX modeling in supplemental methods.

Results and Discussion

Comparison of predicted and experimentally-observed deuterium incorporation trends for RGS4, RGS8, and RGS19: To evaluate the predictive performance of various existing models for PF correlations (see Table A-2 and Model Details), we conducted 10 independent all-atom, explicit-solvent, and μ s-timescale MD simulations for all RGS proteins (Table A-2 and supplemental methods). For each 2 μ s timescale simulation, we analyzed 100,000 conformations of each protein by applying criteria reported previously for each model (Table A-4) and combined calculations on those metrics to obtain protection factors (PFs) for each residue. These PFs, when combined with the intrinsic exchange rates,²⁰⁰ were then used to predict and compare the percentage of deuterium incorporation (%DI) at $t = 0, 3, 10, 30, 100, 300,$ and 1000 minutes for each experimentally observed fragment of each protein (Fig. A-7 and A-8).¹⁴⁴ Then, we reoptimized parameters of models M1 through M7 (the reoptimized models hereafter are referred to as M1* through M7*) by minimizing a fragment-based objective function that compares the predicted and measured values of DI (see supplemental methods). The reoptimization procedure was carried out for simulations conducted with both force-fields (CHARMM-FF and AMBER-FF). The default as well as re-optimized parameters of all 9 models are listed in Table A-4.

We quantified the comparisons between the predicted and experimentally measured deuterium incorporation (%DI) using the relative error (E) and correlation-coefficient (CC) analyses. E measures the discrepancy between the exact values of DI that were measured via HDX-MS experiments and the values that were calculated from MD simulations. However, CC measures

the linear relationship between the measured DI and the modeled DI. It is a measurement of the interdependence or association of two variables and ranges between -1 (negative correlation) and 1 (positive correlation). Therefore, both E and CC are taken into account for the evaluation of each model. In Fig. A-3 and Fig. A-4, we present the statistics of performance of each model via calculations on E and CC for the CHARMM-FF and the AMBER-FF. Specifically, Fig. A-3 shows the performance metrics computed by averaging over data from all MD simulations of all RGS proteins (RGS4, RGS8, and RGS19), while Figure 4 shows the same metrics computed by averaging over all MD simulations of each RGS protein. For additional details, we show the traces of the predicted vs. measured %DI for all fragments of each RGS protein for both force-fields (Figure S3 to Figure S32).

For discussion in the following, we refer to models M1 through M6 as empirical models, and the model M7 as a fractional population model (see supplemental introduction). Overall, we observe that the models M1 through M6 show larger errors and lower correlations in comparison to the model M7 for simulations with both force-fields (gray bars in Fig. A-3). Among empirical models, the model M6 has the smallest error for simulations with the CHARMM-FF (Fig. A-3A), while the model M4 has the smallest error for simulations with the AMBER-FF (Fig. A-3B). The CC values are comparable for the model M6 in the CHARMM-FF and for the model M4 in the AMBER-FF. After re-optimizing the parameters for these models (see supplemental methods and Table A-4), the models M1* and M2* showed significant improvement (lower E and higher CC) for both force-fields in comparison to other models (M3* to M6*), that only moderately improved (blue bars in Fig. A-3). After the reoptimization, even though the E values for the model M7* marginally decreased in comparison to the model M7 (with default parameters), the CC values are similar in both force-fields. The E and CC values for our proposed models (M8 and M9), both

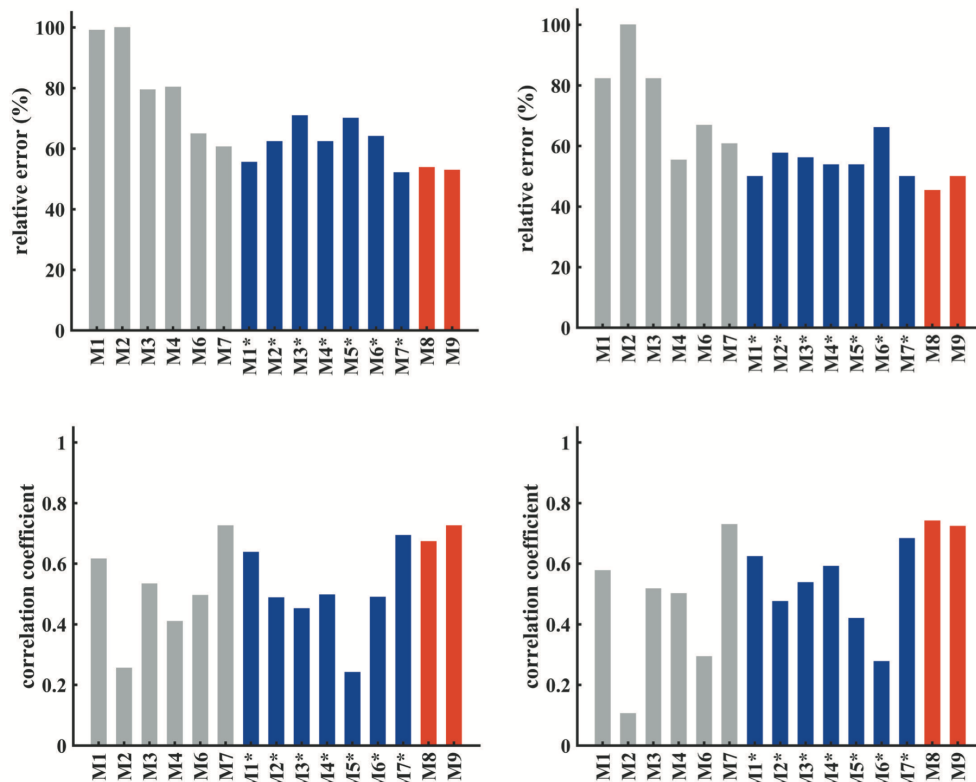


Figure A-3: Comparisons of model predictions of HDX-MS data across all three RGS proteins. Performance metrics (relative error, E , and correlation coefficient, CC) for different models are shown based upon data averaged from all trajectories of RGS4, RGS8, and RGS19 conducted with the CHARMM-FF (data in panels A and B) and the AMBER-FF (data in panels C and D). (A, C) The relative error between the predicted and observed %DI [$E(x, y) = \sum_{i=0}^n |x_i - y_i| / \sum_{i=0}^n y_i$]. (B, D) Correlation coefficient between the predicted and observed %DI [$CC(x, y) = \sum (x_i - \bar{x})(y_i - \bar{y}) / \sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}$]. Gray bars are for models with the default parameters reported in the literature, blue bars are their re-optimized versions based upon our experimental data, and red bars are for new models proposed in this work. No performance data for the original model M5 are reported because the parameter values were not available from the original work,⁴² but the performance data are reported for the optimized version of this model (M5*) based upon our experimental data.

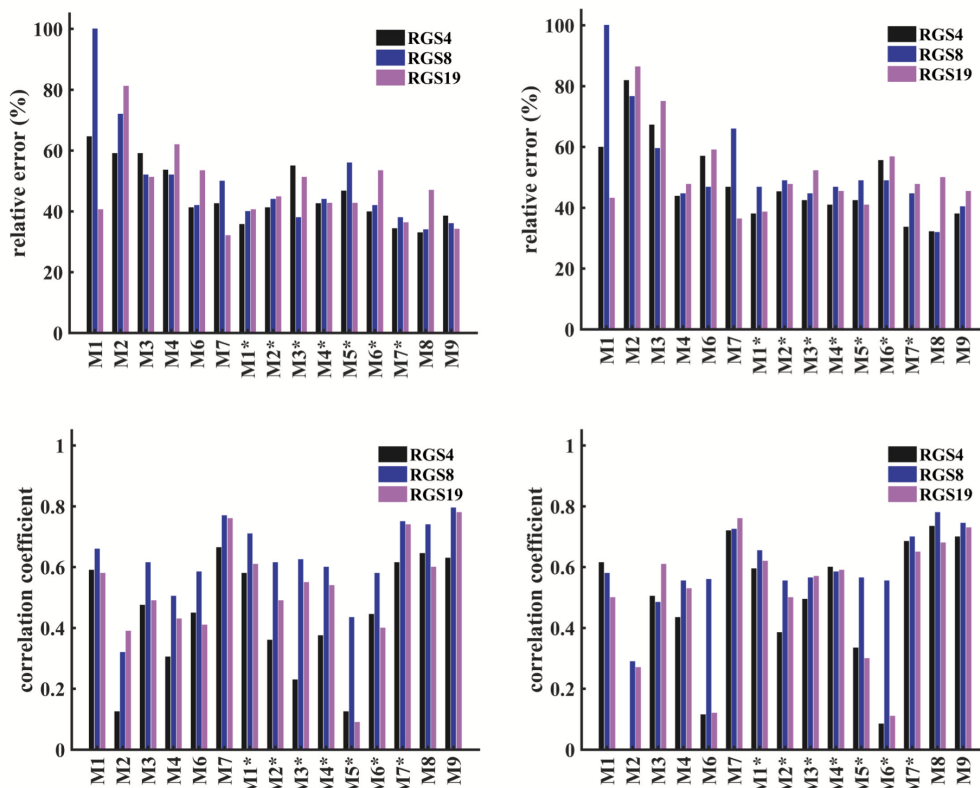


Figure A-4: Comparisons of model predictions of HDX-MS data for each RGS protein. The definitions of E and CC , and other details are the same as in Figure 3. Colored bars distinguish data for each RGS protein: black bars, RGS4; blue bars, RGS8; and magenta bars, RGS19

of which are based on the SASA of each amide hydrogen and its distance from the first polar atom (see supplemental methods), show results comparable to the fractional population model M7 and its reoptimized version M7*. Both of our proposed models consistently predict DI trends with lower E values and higher CC values for both force-fields. Taken together, these data suggest that the proposed models M8 and M9 as well as the models M7 and M7* predict experimentally observed HDX trends better than the other models (M1/M1* through M6/M6*).

On comparing the performance of all empirical models for each RGS protein (Fig. A-4), we observe that the DI trends in RGS4 and RGS8 for the CHARMM-FF are best described (lower E and higher CC values) by the model M6, and for the AMBER-FF are best described by the model M4 (for RGS4) and equally well described by the models 4 and 6 (for RGS8). For RGS19, the

model M1 captures DI trends better than other empirical models (M2 through M6) for both force-fields, but this model is a poor predictor for RGS4 and RGS8. We also observe that the model M2 poorly predicts DI trends (higher E and lower CC values) for all three proteins, and the model M7, a fractional population model, consistently shows better predictions (lower E and higher CC values) for both force-fields. On re-optimizing, all empirical models (M1* through M6*) show improvement (lower E and higher CC values) over their default parameter versions (M1 through M6), but both versions of the fractional population model (M7 and M7*) provide consistently better predictions than the empirical models. The performance of our proposed models M8 and M9 is comparable to the model M7*, but for all three models (M7*, M8, and M9) the performance is marginally poorer (i.e. E values are marginally higher and CC values marginally lower) for RGS19 in comparison to RGS4 and RGS8.

The time-dependence of model predictions contributes significantly to differences in the ability of each model to predict HDX-DI results for each experimentally observed fragment (24 fragments for RGS4, 38 fragments for RGS8, and 26 fragments for RGS19; Fig. A-8).¹⁴⁴ The models show significant variation between shorter time points ($t= 0, 3, 10, 30$, and 100 minutes) and longer time points ($t= 300$ and 1000 minutes) when comparing predicted DI trends at the level of individual fragments for both force-fields (Fig. A-9 to Fig. A-38). For example, models M3, M4, and M6 under-predicted experimentally observed DI trends at shorter time points, but the trends at longer time points are predicted reasonably well (Fig. A-24 and A-25). Similarly, the re-optimized models including M2* through M6* under-predicted DI trends at shorter time points for RGS4 simulations (Fig. A-30 and A-35). Unlike these models, our proposed models M8 and M9 overall show better agreement with the HDX data across all time points and fragments for RGS4 and RGS8 with both force-fields (Fig. A-19 to A-22 and A-34 to A-37). However, for RGS19,

except fragments 18 to 26, each model under-predicts DI trends for both force-fields (Fig. A-23 and A-38).

Our HDX-MS data showed that the amide hydrogens exchanged rapidly in RGS19 in comparison to RGS4 and RGS8 (Fig. A-7), especially in helices α_4 , α_5 , and α_6 (fragments 10 to 23; Fig. A-8).¹⁴⁴ At $t = 1000$ minutes and for models M7, M8, and M9, the mapping of the predicted vs. measured DI on protein structures (Fig. A-39) shows that these models under-predicted DI trends in the α_4 helix of RGS19, but predicted well in the α_6 helix as well as in the α_5 - α_6 / α_6 - α_7 interhelical loops. Importantly, the structural motifs in RGS proteins that showed poor agreement between the predicted and measured DI trends also showed significantly lower residue fluctuations in MD simulations (Fig. A-40) in comparison to those motifs that showed higher fluctuations and as a result better agreement with the experiments.

In summary, each model has unique metrics for estimating the PFs and some of these metrics are shared among different models. For example, the number of polar atoms or residues in the vicinity of an amide hydrogen indirectly assess the likelihood of existence of hydrogen bonds between amide hydrogens and other atoms in proteins. Therefore, different models are directly or indirectly correlated to hydrogen bonds. Our analyses show that the fractional population modeling (e.g. models M7/M7* and M9) is more robust than empirical approaches. In particular, the fractional population models are broadly applicable to newer systems without reoptimization of parameters (e.g. the model M7 makes reasonably accurate predictions both before or after optimization). In our new models (M8 and M9), combining two metrics, SASA and the number of polar protein atoms in the vicinity of a given amide hydrogen, shows better predictions both for the empirical model (M8) and the fractional population model (M9). We also suggest that our new models are potentially applicable to other protein systems for efficient interpretation of HDX

data because these models only require coordinates of the protein atoms. These can be readily extracted from the solvated simulation trajectories for rapid analysis.

Comparison of predicted and measured HDX data at a single-residue resolution

Our HDX-MS data was collected at a fragment resolution for each protein (Fig. A-7 and A-8),¹⁴⁴ but atomistic MD simulations complement these data by providing additional details on the protections of amide hydrogens at a single-residue resolution. At $t = 1000$ minutes for models M7, M7*, M8, and M9, we show in Fig. A-41 to A-46 a color-coded mapping of DI trends for each residue of RGS4, RGS8, and RGS19 for both force-fields. These data show that the amide hydrogens in the N-terminus of the α_3 helix (containing 12 residues; see Fig. A-2) are fully exchanged and some residues are partially exchanged. MD simulations show that the unexchanged or partially exchanged amide hydrogens are participating in hydrogen bonds and are therefore largely protected. Consistent with HDX experiments, these protection effects are observed in fragments 2 and 3 in RGS4 (Fig. A-29), fragments 8, 10, 11 in RGS8 (Fig. A-31), and the fragment 6 in RGS19 (Fig. A-33). In HDX-MS experiments, we observed that the residues in the N-terminus of the α_4 -helix show high exchange propensity in all RGS systems which is accurately predicted by models M7, M7*, and M8. However, all models underpredicted amide hydrogen exchanges in other parts of the α_4 helix (e.g. fragment 6 in RGS4, fragments 14, 15, 16 in RGS8, and fragments 11 and 12 in RGS19) (Fig. A-30, A-31, A-33, A-34, A-36, and A-38). Analyses of our MD simulations showed that the amide hydrogens in these fragments are strongly protected via hydrogen bonds, and therefore local unfolding of the helical structure, even if very transiently, is perhaps required to facilitate any exchange event. Through MD simulations, similar protection effects were identified in the α_5 helix of RGS8 (fragments 24 and 25) (Fig. A-31, A-32, A-36, and A-37)

and RGS19 (fragment 18) (Fig. A-33 and A-38).

The models accurately predicted experimentally-observed exchanges in amide hydrogens in the connecting loops between helices, particularly for the α_5 - α_6 loop (e.g. fragments 12 and 13 for RGS4 in Fig. A-30, A-31, A-34, and A-35; fragment 27 for RGS8 in Fig. A-31, A-32, A-36, and A-37; and fragments 20, 21, and 22 for RGS19 in Fig. A-33 and A-38) which is the longest unstructured region in RGS proteins (Fig. A-2). However, our models showed partial protection for the amide hydrogen of Q122, a residue located in the α_5 - α_6 interhelical loop of RGS4, even though the side chain of this residue is solvent exposed. The amide hydrogen in Q122 forms a long lasting hydrogen bond with S120 leading to a significant protection of this amide hydrogen (Fig. A-47A and C). We also observed complete protection of the amide hydrogen in the residue R119 of RGS8, which is located in the α_5 - α_6 interhelical loop (Fig. A-42). We attribute this to strong salt bridging interactions between the residue R119 and residues E84/E111 (Fig. A-47B and D). For residues located near the protein surface as well as in flexible loops, the ability to remain protected is consistent with the earlier observations on Staphylococcal nuclease²¹¹ showing that the proximity to the surface of the protein does not usually produce fast exchange and therefore a detailed hydrogen by hydrogen analysis is needed, as we have carried out here via MD simulations. These results also provide testable predictions for future HDX-NMR studies aimed at resolving residue-level exchanges since HDX-MS results only provide fragment-level resolution.

Solvent accessible surface area as a metric

In our proposed models M8 and M9, SASA is a key metric in determination of the exposure of amide hydrogens to solvent that consequently contributes to the calculation of protection factors. Since the hydrogen atoms are resolved in the NMR structures of RGS4 (PDB code 1EZT

containing only 1 conformer) and RGS19 (PDB code 1CMZ containing 20 conformers), we computed the maximum and average SASA of all amide hydrogens from the NMR structures (Fig. A-5). Given that all missing hydrogens are included in our MD simulations, we also calculated similar SASA measures of all amide hydrogens of RGS4, RGS8, and RGS19 from all MD trajectories conducted using both force-fields (Fig. A-48). The NMR structures show that only a few amide hydrogens are exposed to solvent and those are located in the connection loops between helices. The maximum values of SASA among all amide hydrogens are $\sim 8 \text{ \AA}^2$ and $\sim 14 \text{ \AA}^2$ for RGS4 (PDB code 1EZT) and RGS19 (PDB code 1CMZ), respectively.

Our model M9 showed that the SASA threshold beyond which the experimental HDX trends are well predicted are $8.02 \text{ \AA}^2$ and $9.15 \text{ \AA}^2$ for CHARMM and AMBER force-fields, respectively. Given these values, none of the residues in the NMR structure of RGS4, and only 4 residues in the NMR structure of RGS19 have enough exposure for competent exchange. However, amide hydrogens show larger exposure to solvent in MD simulations (Fig. A-48) with maximum values up to $\sim 20 \text{ \AA}^2$. For interhelical loops, the average SASA of amide hydrogens in simulations is about two times that of helical motifs in RGS proteins. The residues within well-folded and stable helices never adopt SASA values beyond the threshold SASA values (*vide supra*), thereby suggesting strong protection effects for these amide hydrogens. Given that the SASA values of amide hydrogens in the initial structures of RGS proteins (Fig. A-5) and in MD simulations (Fig. A-48) are different as well as given the consistent performance of our SASA-based proposed models (M8 and M9; Fig. A-3 and A-4), we find SASA computed from MD simulations as a useful metric in modeling of HDX-MS data.

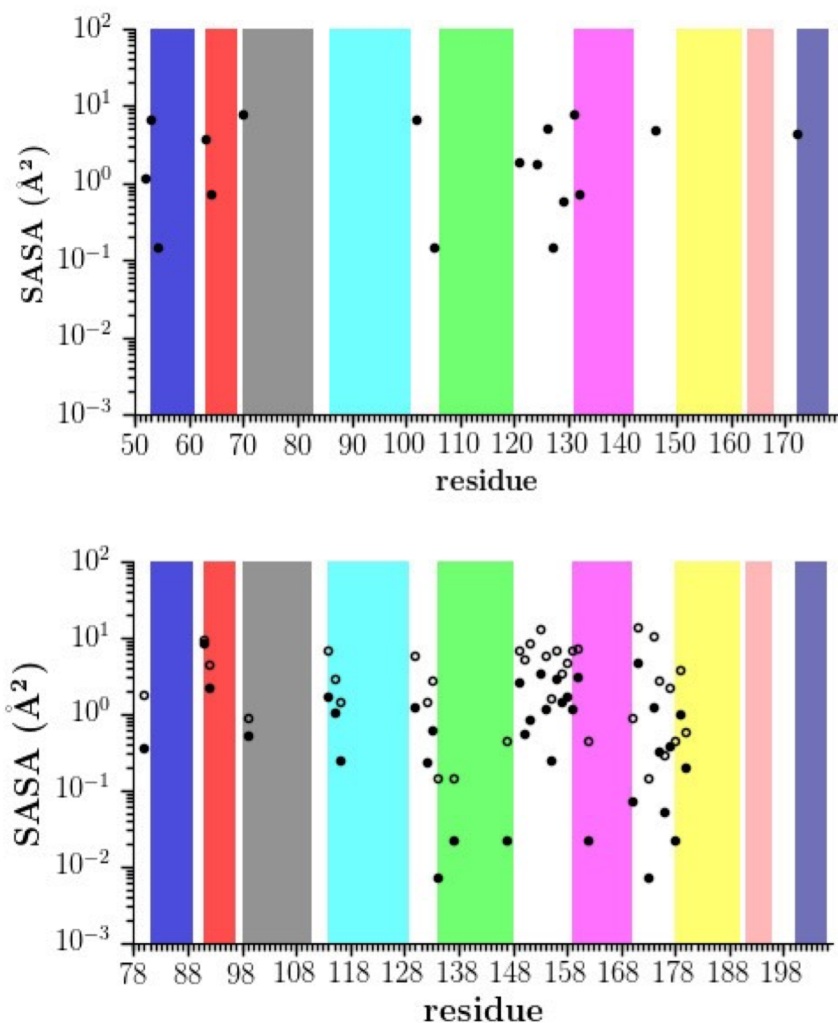


Figure A-5: The exposure of amide hydrogens in the NMR structures of RGS proteins. Shown are the maximum (open circles) and the average (solid circles) values of the solvent accessible surface area for all amide hydrogens in the NMR structures of RGS4 (panel A) and RGS19 (panel B). In both panels, the absence of filled circles for certain amides as well as the absence of open circles in panel B, is due to the approximately nil SASA values for those amides. The absence of open circles for RGS4 in panel A is due to the lack of availability of more than 1 conformer in the NMR structure of RGS4 as opposed to 20 conformers in the NMR structure of RGS19.

Mean residence times and cooperativity of amide hydrogens in the open and closed states

In the fractional-population models (M7/M7* and M9), the kinetics of fluctuations between the open and closed states are characterized by the mean residence time (MRT) which is defined, in an MD simulation, as the average number of consecutive frames in each state multiplied by the time-step.¹³¹ Therefore, computing the MRT at residue-resolution provides information on the tendency of each amide hydrogen to be in the open and the closed state. Two specific criteria (Table A-4) were evaluated to classify amides as being in the open or closed states for each frame in MD trajectories. Then, the MRT values of the closed state and the open state are used to calculate the protection factors ($PF = \tau_C/\tau_O$). To calculate the PF for model M9, we divided the number of frames in which an amide hydrogen is in a closed state (N_{FC}) by the number of frames in which an amide hydrogen is in an open state (N_{FO}). If N_O and N_C are the number of visits to the open state and the closed state during the MD trajectory, respectively, and T_O and T_C are the total time that each amide is in the open or the closed state, respectively, it can be written that $T_O = N_{FO}\Delta\tau = N_O\tau_O$ and $T_C = N_{FC}\Delta\tau = N_C\tau_C$, where $\Delta\tau$ is the time-step (which is 2 fs in our MD simulations). This results in the protection factor, $PF = T_C/T_O$ by assuming that $N_O = N_C - 1.30$. In Fig. A-6, we show the MRT values of the open and the closed states of all residues from MD trajectories of all proteins conducted using the CHARMM-FF and the AMBER-FF. These values were calculated using equations: $\tau_O = N_{FO}\Delta\tau/N_O$ and $\tau_C = N_{FC}\Delta\tau/N_C$.

Since the open states of amide hydrogens may occur at time scales shorter than the time-step ($\Delta\tau$) used in MD simulations, it was previously shown that the MRT values can be quantitatively corrected to account for the sampling-resolution systematic binning error. The corrected values are given by $\tau^c_O = -\Delta\tau/\ln(1-N_O/N_{FO})$ and $\tau^c_C = N_{FC}\Delta\tau/(1-N_{FO}\ln(1-N_O/N_{FO}))$.¹³¹ We show the corrected MRT values in Fig. A-49. These data show that τ_O ranges between 20 to 50 ps while τ^c_O

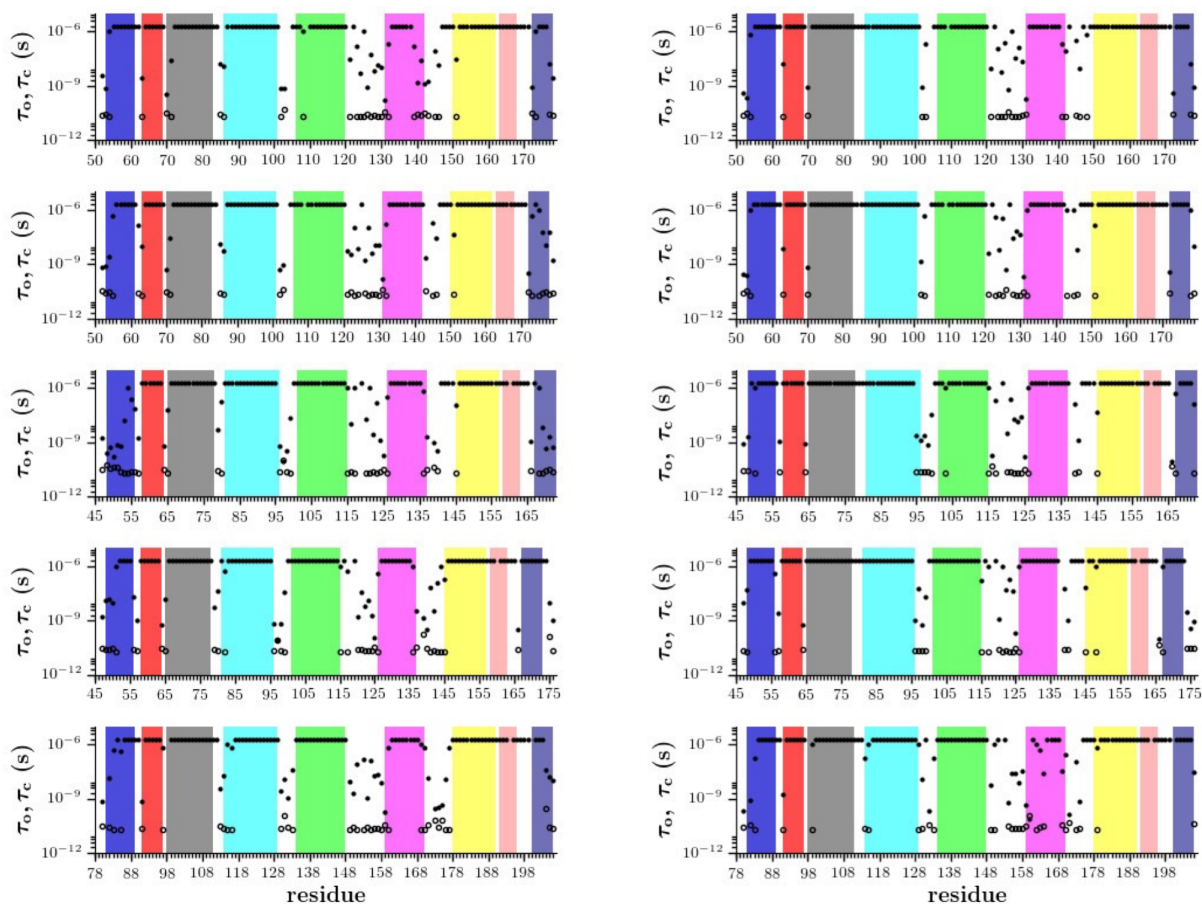


Figure A-6: Mean residence times for the open and closed states of amide hydrogens. Data are shown from all simulations of RGS4, RGS8, and RGS19 conducted with the CHARMM-FF (panel A) and the AMBER-FF (panel B). The MRT calculations were carried out using our proposed fractional population model M9 that showed consistent predictions with the HDX-MS data.

ranges between 5 to 50 ps and τ_C ranges between 170 ps to 2 μ s while τ^c_C ranges between 110 ps to 2 μ s. The observation that the open states of amides occur on a sub-100 ps time scale is consistent with similar earlier observations on the protein BPTI.¹³¹ As suggested previously,¹³¹ these time scales are orders of magnitude shorter than the MRT values of globally unfolded proteins and therefore highlight the concept that amides can exchange by highly localized and short-lived fluctuations without the need for global unfolding. We further examined whether the open states of amide hydrogens are truly localized or if they are allosterically coupled and cooperative. Specifically, we computed the open state residue-residue correlation matrix for two simulations that have shown significant per-residue fluctuations in RGS4 (PDB:1AGR) and RGS8 (PDB:2ODE) using the CHARMM-FF. We observed that the correlation matrix varies in a short-range for both systems (Fig. A-50 and A-51) indicating that the open states for amides are largely uncorrelated between residue pairs, as also has been previously observed for BPTI.¹³¹ These observations are consistent with the amide hydrogen exchanges occurring in the EX2 exchange limit.¹⁹¹ Furthermore, the probability of observing open states of amides for a trajectory of given length can be analyzed using Poisson statistics.¹³¹ We present this analysis in Fig. A-52 for the PF-values of 10^2 , 10^4 , 10^6 , and 10^{11} with $\tau_O = 20$ ps and 100 ps. The analysis shows that the open states of amides with the PFs ranging between 10^2 and 10^6 can be observed in MD trajectories of simulation lengths ranging between 10^{-3} μ s and 10 μ s. This is consistent with the results on the DI observed in experiments and predicted by simulations for RGS proteins. However, the amides that are highly protected and are not observed to exchange in experiments likely have protection factors of 10^{11} or higher (as predicted by our simulations) and would require trajectories on time scales of millisecond or higher for observing open states. We suggest that the probability of observing sufficient opening events for amides can be further enhanced by conducting simulations

with multiple force-fields and different initial structures of proteins, as we have carried out in this work for RGS proteins.

Conclusion

We used MD simulations to study hydrogen-deuterium exchange events in three isoforms of RGS proteins. Specifically, we analyzed various existing models from the literature to assess their ability in accurately predicting experimentally observed exchange patterns in these homologous RGS proteins. These analyses revealed significant variation among models in accuracy of predictions and showed that empirical models (termed models M1 through M6 in Table A-1) with their previously reported criteria made inconsistent predictions, while a fractional population model (Model M7) predicted experimentally-observed trends with good accuracy. Even though we found that reoptimizing previous empirical models using our data on RGS proteins improves their prediction accuracy, the performance of the fractional population model is less sensitive to parameters. We further assessed the usefulness of a previously ignored metric, SASA of amide hydrogens determined from MD simulations, and combined it with the distance of a given amide hydrogen from the first polar atoms in proteins to propose two new models (models M8 and M9) that show good predictions for observed HDX patterns. Importantly, the proposed models only require the coordinates of protein atoms from solvated trajectories providing improved computational efficiency. We also find that the amide hydrogens often transiently visit open states on sub-100 ps time scales, which is significantly shorter than time scales for global unfolding. This therefore suggests that there is localized exposure of the amide-hydrogens, especially given that open states among amide hydrogens of a given protein are uncorrelated.

Model Details

In the following, we provide details on seven existing models for protection factor (PF) correlations, as shown in Table 1.

Model M1: Resing et al.²²⁵ conducted early studies to predict exchange rates in a kinase protein (ERK2) by fitting protection factors to an equation of the form $\log(PF_i) = \log(k_{int}/k_{hdx}) = u \cdot (SA_i) + v/(HB_i)$, where k_{hdx} is the experimentally measured exchange rate of an amide hydrogen, k_{int} is the intrinsic exchange rate calculated according to Bai et al.,²⁰⁰ SA_i is the distance of each amide hydrogen from the surface of protein in Å, and HB_i is the hydrogen bond length of backbone amide nitrogens to an acceptor. They also used deuterium exchange rates measured by Milne et al.²³¹ for horse heart cytochrome c.

Model M2: Vendruscolo et al.²⁰⁸ proposed a model for predictions of HDX rates based on the exploration of conformations using Monte Carlo (MC) sampling biased by experimental data. They speculated that the protection of amide hydrogens comes from buried part of the amide group and also from the hydrogen bonding in the secondary structure which resulted in a phenomenological expression including the number of contacts of residue i with other residues (N_i^c) and the number of hydrogen bonds formed by the amide hydrogens of residues (N_i^h), respectively. According to their definition, hydrogen bonds are present if the angle between the NH vector and the OH vector is below 0.7 rad and the OH distance is below 2.4 Å. Also, two residues are in contact if any pair of their atoms are closer than 8.5 Å.

Model M3: Best et al.²⁰⁹ used the same phenomenological expression that Vendruscolo et al.²⁰⁸ had proposed but with minor changes in definition of N_i^c and N_i^h . The contribution of burial in the model is the number of heavy atoms within a distance of 6.5 Å from the amide nitrogen. A cutoff of 2.4 Å between the donor hydrogen and the acceptor was used for identifying a hydrogen

bond without an angle criterion. They optimized the parameters of their model using experimental protection factors and the corresponding protection factors from a 1 ns conventional MD simulation of seven different proteins. They acknowledge that major protein fluctuations were elusive from short MD simulations which motivated them to conduct a biased simulation of the protein bovine pancreatic trypsin inhibitor (BPTI) by using hydrogen exchange restraints with varying values of the parameters.

Model M4: Kieseritzky et al.²¹⁰ used MD simulations as a complement for hydrogen exchange experiments. They simulated oxidized c-type cytochrome under native conditions (PDB code 1K3H) with the CHARMM22 force-field using explicit water molecules modeled using the TIP3P water model. The simulation was 3 ns long. They proposed a protection factor definition based on a linear combination of protection factors $\log(PF_i) = \log(k_{int}/k_{hdx}) = \beta_1 PFE_1 + \beta_2 PFE_2$. They optimized parameters β_1 and β_2 to arrive at an agreement between computed (based on MD simulation data) and measured hydrogen exchange protection factors. The nine different protection factor correlations in their paper show varieties of error and Pearson's correlation coefficient out of which $PFE_1 = [\text{the number of residues which are in contact with corresponding residue}]$ and $PFE_2 = [\text{the inverse of the backbone atom RMSF}]$ show the least error and the best correlation.

Model M5: A model was suggested by Ma et al.²¹² where NH_i^β is the average number of hydrogen bonds between the NH atom of residue i and C=O backbone oxygen within 2.6 Å distance, and NH_i^{sol} is the average number of hydrogen bonds between NH and water oxygen within 3.0 Å distance of residue i . In the original model, NH^β is measured in β -sheets and the correlation is marginal $PF_i = (NH_i^{sol} + NH_i^\beta)/C.NH_i^{sol}$. They used CHARMM27 force-field to do MD simulations of different β -sheet conformations, each of which was for 60 ns.

Model M6: Park et al.¹³⁰ recently developed a novel model based on a comprehensive HDX-MS experimental data using Amber 11 ff99SB force-field and a 100 ns long simulation. Their logistics growth function HDX model consist of one fitting parameter called “base”. $NHstat_i$ is defined as ([the number of snapshots showing H-bonding of amide hydrogen to protein]-[the number of snapshots showing H-bonding amide hydrogen to water])/[the total number of snapshots]. They provided three amide hydrogen bond models out of which model HB2 has been compared with other models in their work. In the HB2 model, H-bonding of a given amide hydrogen to the side chain as well as C=O group in the backbone are counted as H-bonding of amide hydrogens to protein. The fraction of deuterium incorporation (DI) for each amide hydrogen was computed by the first order reaction kinetics $DI_i^{res} = 1 - \exp(-k_{int,i} t / PF_i)$.

Model M7: Persson et al.¹³¹ used a significantly long MD simulation of protein BPTI (0.262 ms long) generated by Shaw et al.²²⁸ using Amber ff99SB-I/TIP4P-Ew force-field. They start with a description of the standard model in which each amide can be exposed to solvent in an open state or buried within the protein by a closed state: $(N-H)_c \xrightleftharpoons[k_c]{k_o} (N-H)_o \xrightarrow{k_{int}} (N-D)_o$ in which HDX rate is given as $k_{hdx} = k_o k_{int} / (k_o + k_c + k_{int})$. The assumption of $k_{int} \ll k_c + k_o$, which is an applicable assumption for HDX experiment, results in a simple and practical phenomenological model $k_{hdx} = k_{int} / (PF + 1)$. The protection factor here is the key for the calculation of hydrogen-deuterium exchange rate and it is defined as the ratio of residence time in the closed state to residence time in the open state which is applicable to MD simulations. The criteria for the open state and the closed state play an important role in computing protection factors. They speculate that a direct access to external solvent and disruption of any intramolecular H-bond with the N-H group are key factors in defining the open state. A residue is in an open state when the amide hydrogen has at least two water oxygens within 2.6 Å and that the amide hydrogen has no other

	PDB	system size (atoms)	force-field (trajectory length)
RGS4	1AGR	28160	CHARMM36 (2 μ s), AMBER (2 μ s)
	1EZT	29275	CHARMM36 (2 μ s), AMBER (2 μ s)
RGS8	2IHD	27490	CHARMM36 (2 μ s), AMBER (2 μ s)
	2ODE	30731	CHARMM36 (2 μ s), AMBER (2 μ s)
RGS19	1CMZ	29560	CHARMM36 (2 μ s), AMBER (2 μ s)

Table A-2: Summary of MD simulations.

polar protein atoms (except in neighboring residues) within 2.6 Å.

Other studies: In addition to models highlighted above, Craig et al.²¹⁷ modeled deuterium incorporation of three different proteins using coarse-grained MD simulations. The open state criteria were evaluated by the number of contacts per residue and the distance changes between the H- bonded residues compared to their native conformations. Petruk et al.²¹⁸ studied a kinase protein (ERK2MAP) using all-atom explicit-water MD simulations and showed that both the whole dynamically averaged solvent accessible surface area (SASA) and the number of waters in the first solvation shell of each amide nitrogen can be used as metrics for predicting deuterium incorporation. Recently, Adhikary et al.²⁰¹ have modeled deuterium incorporation using multiple MD simulations (each 450 ns long) of neurotransmitter sodium symporters.

Supplemental Methods

System Setup: MD simulations

A summary of all MD simulations for RGS4, RGS8, and RGS19 is provided in Table A-2. Specifically, 10 independent MD simulations, each 2 μ s long, were conducted using both CHARMM and AMBER force-fields for all apo RGS proteins.

The HDX-MS experiments provided fragment-based DI whereas in MD simulations, it is feasible to calculate DI at a residue resolution. In Fig. A-8, we show details on all fragments and their residues for RGS4, RGS8, and RGS19. To compare DI between experiments and simulations, DI of residues (except prolines that do not have amide hydrogens) were averaged over the corresponding fragment using Eq. (1):

$$DI_i^{frag} = \frac{\sum_{j=1, \neq PRO}^m DI_j^{res}}{m} \quad (1)$$

where m is the number of residues in the fragment.

For all models, we calculated the intrinsic HDX kinetic rates per Bai et al.²⁰⁰ at 273 K, the temperature at which our HDX-MS experiments were conducted. Initially, we analyzed 100,000 frames for each 2 μ s MD trajectory by applying the default criteria reported in the literature for models M1 through M7 to compute PFs of amides for all RGS proteins. We then re-optimized the parameters of all models by minimizing an objective function (Eq. (2)) which incorporates HDX-MS data and MD simulations of all RGS proteins. It should be noted here that the optimization of parameters were carried out separately for each force-field due to the fact that CHARMM and AMBER force-fields are parameterized differently for studies of protein dynamics.

$$OF = \sum_{SYS=1}^5 \left(\sum_{frag=1}^n \left| DI_{exp}^{frag} - DI_{sim}^{frag} \right| \right)_{SYS} \quad (2)$$

where OF is the objective function, DI is deuterium incorporation, SYS is the number of simulations for an RGS protein using the same force-field, and $frag$ is the fragment number. All default and re-optimized parameters of models M1 through M7 are listed in Table A-4.

In addition to existing models (M1-M7), we revisited and evaluated SASA of amide hydrogens as a metric in prediction of amide PFs because contradictory observations regarding the use of SASA as a metric have been proposed in the literature. Published studies indicate that SASA of amide hydrogens reasonably predicts the number of exchanged hydrogens²¹⁸ or is an even better indicator for protected hydrogens than using H-bonds.²²¹ Contrary to this view, a lack of agreement between HDX experiments and MD simulations based on SASA has been reported.¹³¹ Besides, although anticorrelations between the SASA of amide hydrogens and the residue-resolution protection factors from experiments existed, Park et al.¹³⁰ chose H-bonds as a metric for HDX modeling to overcome the limitation of using SASA and they concluded that H-bonds are a generic and suitable metric for the estimation of PFs.

We therefore developed two new models (listed as M8 and M9 in Table A-3 and A-4) using the distances of amide hydrogens from the first polar atom as an alternative metric along with SASA of each amide hydrogen to comply with the theory of HDX in which a residue may be protected by polar atoms despite having large enough SASA.^{131,205} This assertion comes from the fact that surface exposed hydrogens (with higher values of SASA) can be significantly protected from hydrogen exchange.²²⁵ Surprisingly, these two metrics in combination have resulted in trends and values consistent with experiments.

Specifically, model M8 is an empirical model (similar to models M1 through M6) based upon SASA of amide hydrogens and distances of amide hydrogens to the first polar atom (except in the neighboring residues) (D_i) and $\ln(PF_i)$ is a power function of $SASA_i$ and D_i . However, model M9 is a fractional population model¹³¹ where the same metrics ($SASA_i$ and D_i) were used for distinguishing between the open and closed states of amides. We define the open state in model M9 for each amide hydrogen when its SASA crosses a threshold value (d_{sasa}) and that the

Model	Protection factor criteria
M8	$\ln(PF_i) = (\beta_s SASA_i^{-\gamma_s} + \beta_p D_i^{-\gamma_p})$
M9	$PF_i = \tau_C / \tau_O$

Table A-3: Models proposed in this work.

amide hydrogen has no other polar protein atom (except in neighboring residues) within a threshold distance (d_p). The values of thresholds/cut-offs in model M9 and four correlation coefficients in model M8 are obtained by minimizing the objective function in Eq. (2). The intrinsic exchange rates in new models were also calculated according to Bai et al.²⁰⁰

Model	Criteria
M1 (1999-Resing)	$\log(PF_i) = u \cdot (SA_i) + v/(HB_i)$ $u = 0.76, v = 8.2$ $u^{ch} = 6.15, v^{ch} = 5.32$ $u^{am} = 5.18, v^{am} = 4.92$
M2 (2003-Vendruscolo)	$\ln(PF_i) = (\beta_c N_i^C + \beta_h N_i^h)$ $\beta_c = 1, \beta_h = 5$ $\beta_c^{ch} = 0.49, \beta_h^{ch} = 0.85$ $\beta_c^{am} = 0.5, \beta_h^{am} = 0.9$
M3 (2006-Best)	$\ln(PF_i) = (\beta_c N_i^C + \beta_h N_i^h)$ $\beta_c = 0.35, \beta_h = 2$ $\beta_c^{ch} = 0.23, \beta_h^{ch} = 5.40$ $\beta_c^{am} = 0.23, \beta_h^{am} = 4.00$
M4 (2006-Kieseritzky)	$\ln(PF_i) = (\beta_c N_i^C + \beta_r (N_i^r)^{-1})$ $\beta_c = 0.5, \beta_r = 0.9$ $\beta_c^{ch} = 0.45, \beta_r^{ch} = 1.31$ $\beta_c^{am} = 0.19, \beta_r^{am} = 6.45$
M5 (2011-Ma)	$PF_i = (C_o NH_i^{sol} + C_c NH_i^\beta) / CNH_i^{sol}$ $C_o^{ch} = 8.48e^{-6}, C_c^{ch} = 2.50$ $C_o^{am} = 0.15, C_c^{am} = 1.47e^4$
M6 (2011-Park)	$PF_i = base / (1 + (\sqrt{base})^{1-NHstat_i})$ $base = 10^8$ $base^{ch} = 1.3e^8$ $base^{am} = 0.4e^8$
M7 (2015-Persson)	$PF_i = \tau_C / \tau_O$ $d_w = 2.60, d_p = 2.60$ $d_w^{ch} = 2.43, d_p^{ch} = 2.73$ $d_w^{am} = 2.40, d_p^{am} = 2.73$
M8	$\ln(PF_i) = (\beta_s SASA_i^{-\gamma_s} + \beta_p D_i^{-\gamma_p})$ $\beta_s^{ch} = 0.72, \beta_p^{ch} = 2.60e^1$ $\gamma_s^{ch} = 0.53, \gamma_p^{ch} = 0.99$ $\beta_s^{am} = 1.30e^{-3}, \beta_p^{am} = 3.65e^1$ $\gamma_s^{am} = 2.64, \gamma_p^{am} = 1.27$
M9	$PF_i = \tau_C / \tau_O$ $d_{sasa}^{ch} = 9.15^2, d_p^{ch} = 3.00$ $d_{sasa}^{am} = 8.02^2, d_p^{am} = 2.99$

Table A-4: Details on all protection factor correlation models with the default and reoptimized values of their parameters. Optimized values based upon simulations conducted using CHARMM and AMBER force-fields are listed with superscripts *ch* and *am*, respectively. In addition, details on two new models M8 and M9 proposed in this work are listed.

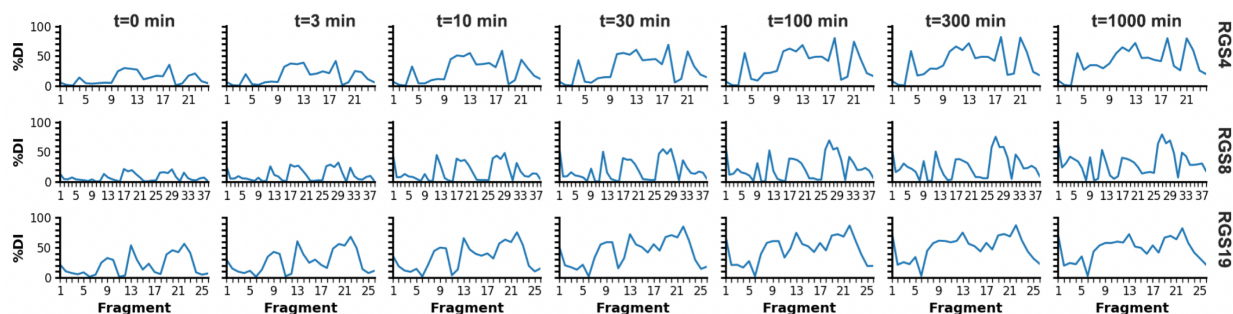


Figure A-7: Experimentally measured percentage deuterium incorporation (%DI) of fragments in RGS proteins at $t = 0, 3, 10, 30, 100, 300$, and 1000 minutes (RGS4: top row; RGS8: middle row; RGS19: bottom row).

RGS4			RGS19			RGS8		
Fragment Number	Residue ID PDB : 1AGR	Sequences	Fragment Number	Residue ID PDB : 1CMZ	Sequences	Fragment Number	Residue ID PDB : 2ODE	Sequences
1	67 to 73	INHECGL	1	84 to 87	VQSW	1	51 to 57	TRWADSF
2	74 to 80	AAFKAFL	2	87 to 91	WAQSF	2	58 to 68	DVLLSHKYGVAA
3	76 to 80	FKAFL	3	88 to 104	AQSFDKLMHSPAGRSVF	3	58 to 69	DVLLSHKYGVAA
4	81 to 87	KSEYSEE	4	92 to 104	DKLMHSPAGRSVF	4	61 to 68	LSHKYGVAA
5	88 to 91	NIDF	5	95 to 104	MHSPAGRSVF	5	61 to 69	LSHKYGVAA
6	92 to 97	WISCEE	6	105 to 108	RAFL	6	61 to 70	LSHKYGVAA
7	92 to 118	WISCEEYKKIKSPKLSPKAKKIYNEF	7	105 to 112	RAFLRTEY	7	62 to 69	SHKYGVAA
8	97 to 117	EYKKIKSPKLSPKAKKIYNE	8	108 to 115	LRTYESEE	8	66 to 72	GVAAFRA
9	98 to 118	YKKIKSPKLSPKAKKIYNEF	9	109 to 115	RTEYSEE	9	69 to 77	AFRAFLKTE
10	116 to 123	NEFISVQA	10	109 to 117	RTEYSEENM	10	70 to 74	FRFL
11	118 to 129	FISVQATKEVNL	11	119 to 122	FWLA	11	71 to 74	RAFL
12	119 to 129	ISVQATKEVNL	12	119 to 125	FWLACEE	12	75 to 81	KTEFSEE
13	123 to 129	ATKEVNL	13	126 to 132	LKAEANQ	13	79 to 83	SEENL
14	130 to 143	DSCRTRETSRNMLE	14	127 to 137	KAQANQHVDE	14	82 to 85	NLEF
15	130 to 149	DSCRTRETSRNMLEPTITCF	15	129 to 141	EANQHVDEKARL	15	86 to 90	WLACE
16	136 to 149	ETSRNMLEPTITCF	16	130 to 139	ANQHVDEKA	16	86 to 91	WLACEE
17	143 to 151	EPTITCFDE	17	133 to 141	HVVDEKARL	17	91 to 101	EFKKTRSTAKL
18	144 to 149	PTITCF	18	142 to 145	IYED	18	91 to 105	EFKKTRSTAKLVSKA
19	150 to 159	DEAQKKIFNL	19	145 to 154	DYVSILSPKE	19	92 to 101	FKKTRSTAKL
20	150 to 173	DEAQKKIFNLMEKDSYRRFLKSRF	20	147 to 154	VSILSPKE	20	92 to 105	FKKTRSTAKLVSKA
21	158 to 173	NLMKDSYRRFLKSRF	21	147 to 157	VSILSPKEVSL	21	93 to 111	KKTRSTAKLVSKAHRIFEE
22	160 to 165	MEKDSY	22	151 to 157	SPKEVSL	22	102 to 111	VSKAHRIFEE
23	160 to 173	MEKDSYRRFLKSRF	23	158 to 174	DSRVREGINKMQEPSA	23	102 to 112	VSKAHRIFEEF
24	166 to 173	RRFLKSRF	24	175 to 182	HTFDQAQL	24	106 to 111	HRIFEE
			25	183 to 196	QIYTLMHRDYSYPRF	25	106 to 112	HRIFEEF
			26	188 to 196	MHRDYSYPRF	26	112 to 125	FVDVQAPREVNIDF
						27	113 to 120	VDVQAPRE
						28	113 to 125	VDVQAPREVNIDF
						29	118 to 125	PREVNIDF
						30	118 to 129	PREVNIDFQTRF
						31	126 to 136	QTREATRKNLQ
						32	130 to 140	ATRKNLQEPSL
						33	141 to 153	TCFDQAQGVHSL
						34	144 to 155	DQAQGVHSLME
						35	146 to 153	AQGVHSL
						36	154 to 162	MEKDSYPRF
						37	154 to 169	MEKDSYPRFLRSKMYL
						38	160 to 168	PRFLRSKMY

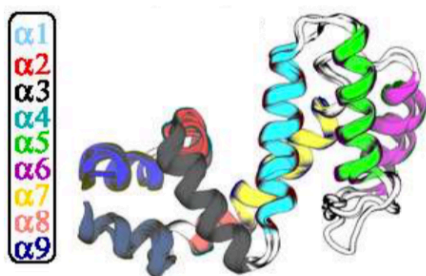


Figure A-8: Definitions of fragments for each RGS protein. Each fragment comprises residues whose color determines their location in nine α helices of each RGS protein. Residue names in connecting loops are highlighted in black, but shown as white cartoons in the protein structure. All helices are colored and labeled in the protein rendering.

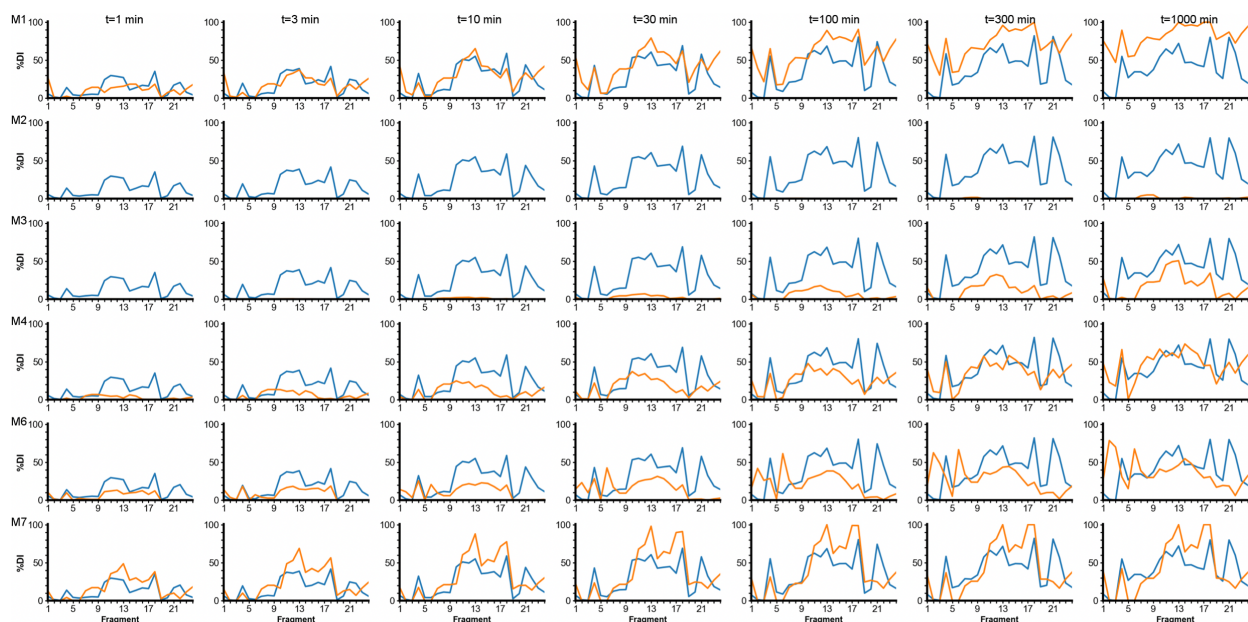


Figure A-9: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1AGR and AMBER force-field.

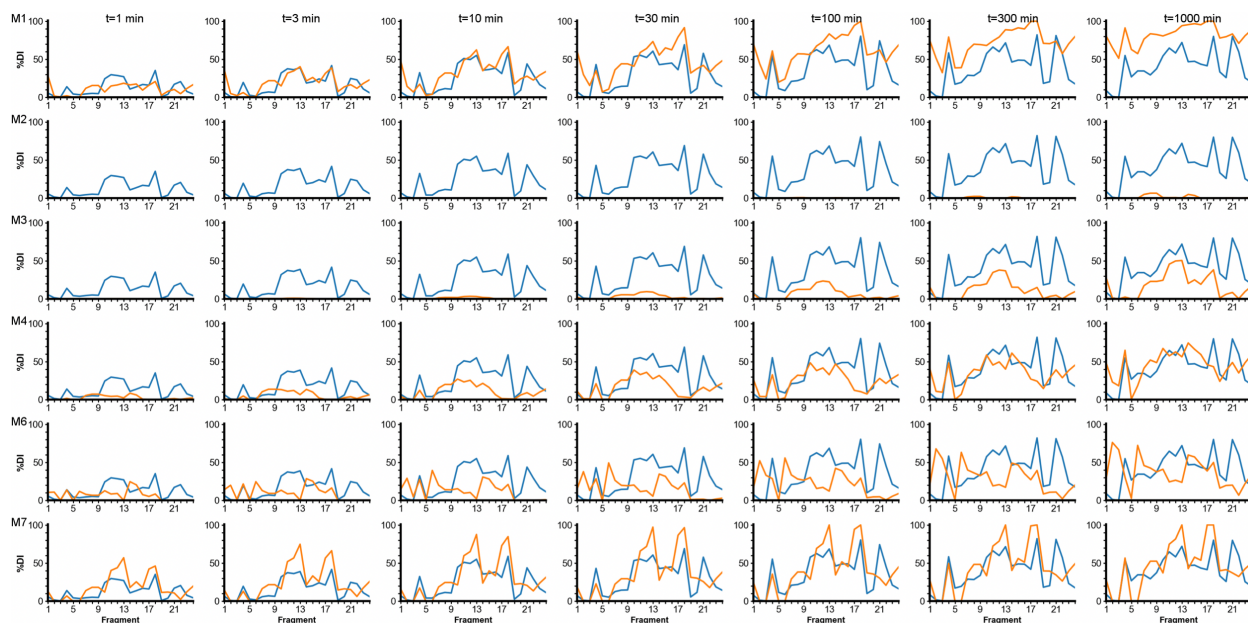


Figure A-10: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1EZT and AMBER force-field.

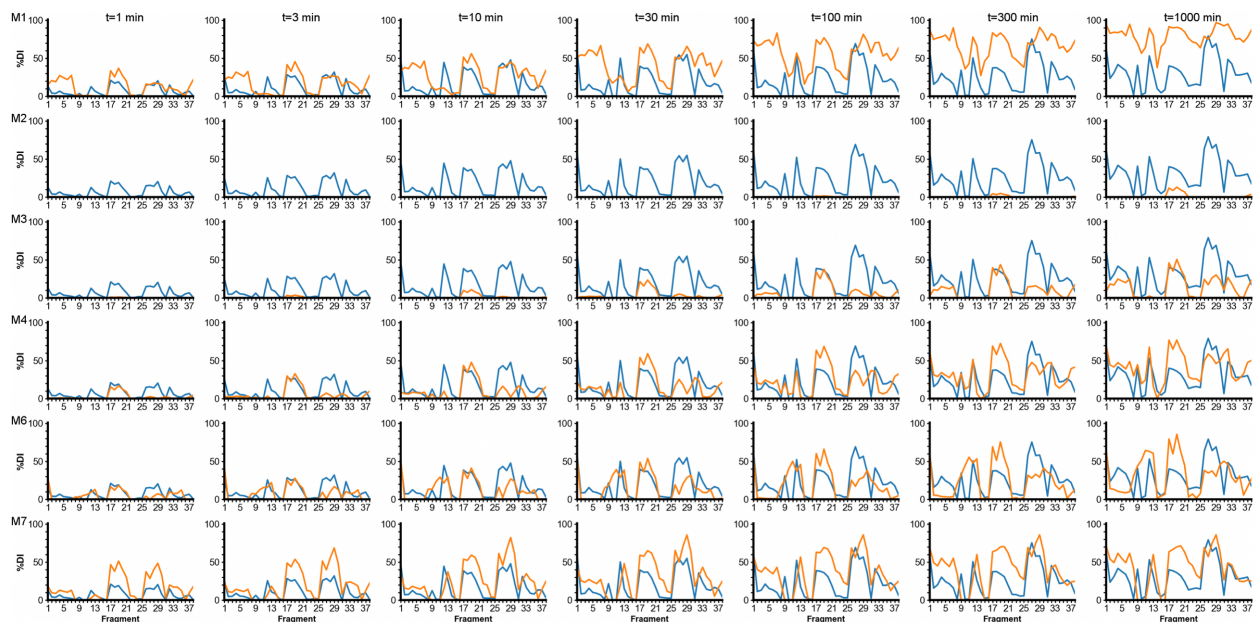


Figure A-11: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2IHD and AMBER force-field

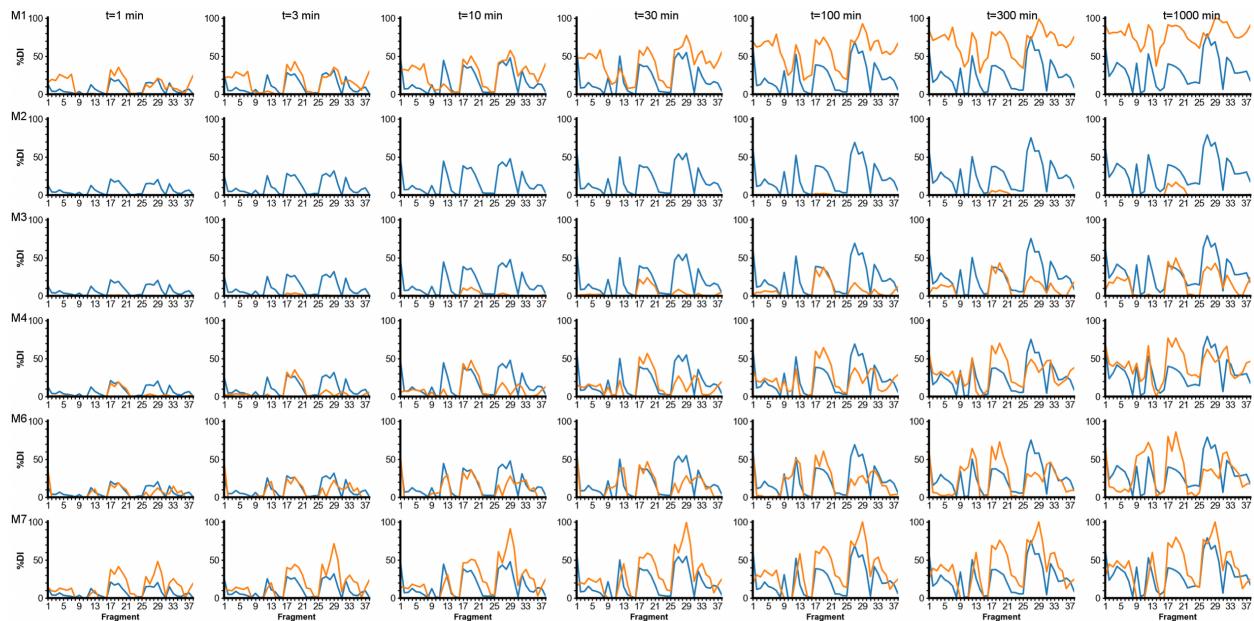


Figure A-12: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2ODE and AMBER force-field

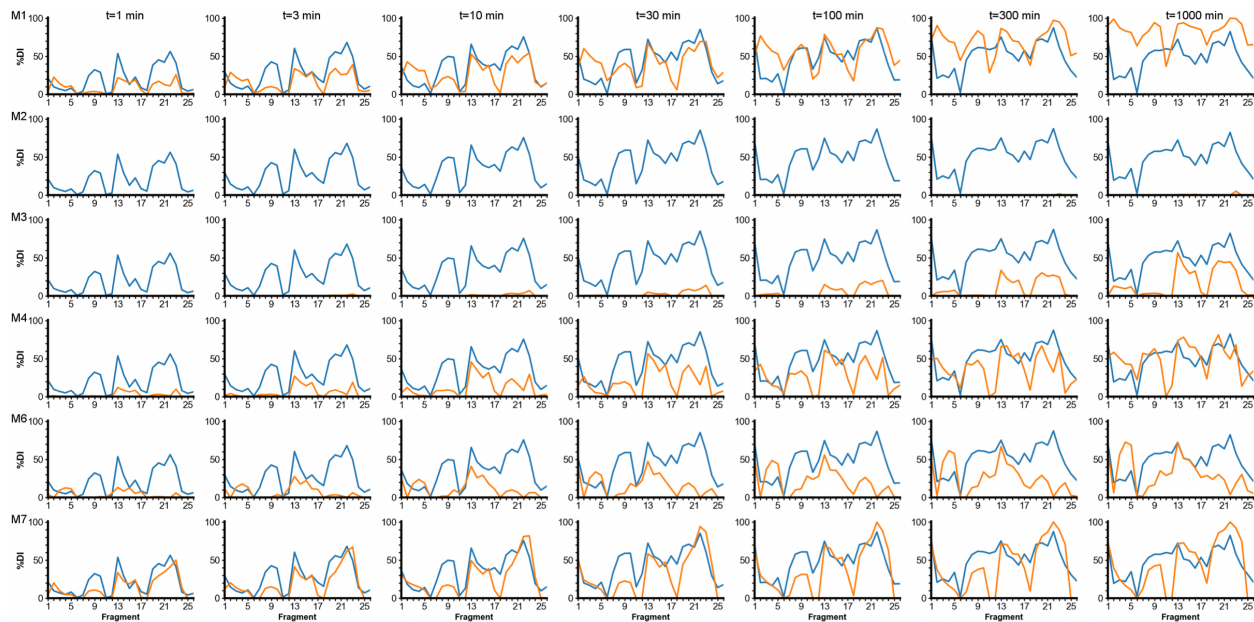


Figure A-13: Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and AMBER force-field.

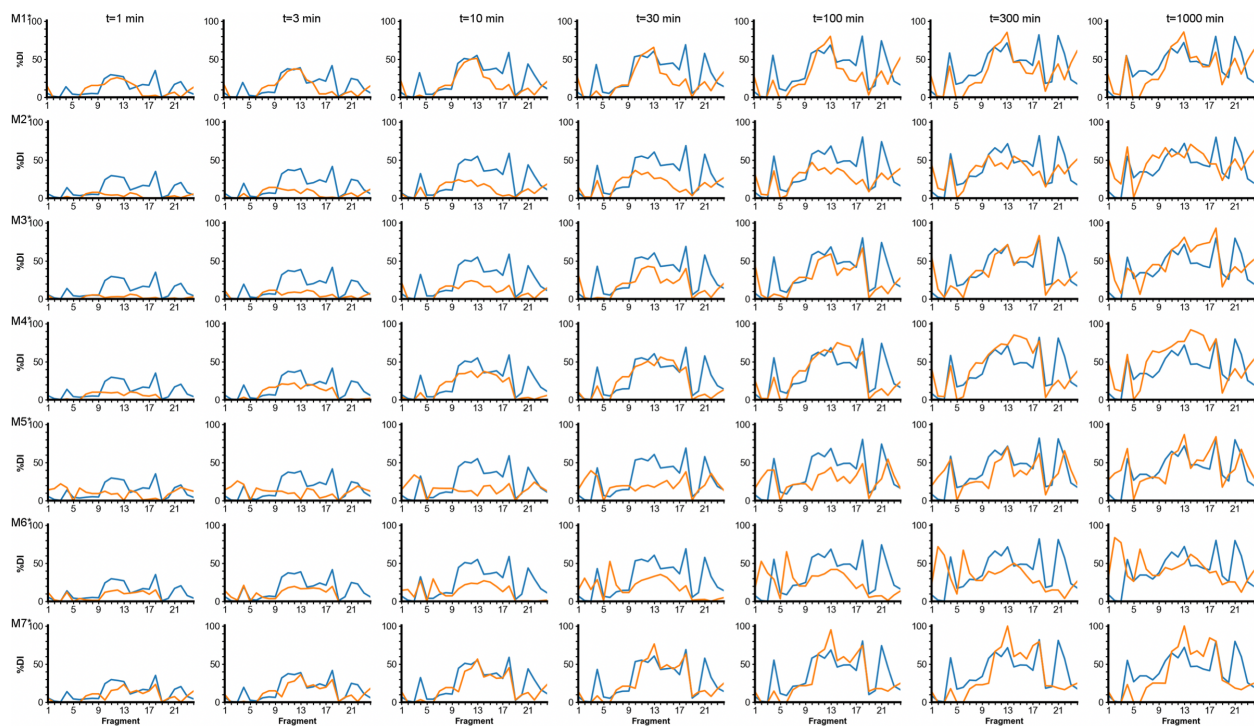


Figure A-14: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and AMBER Force-field

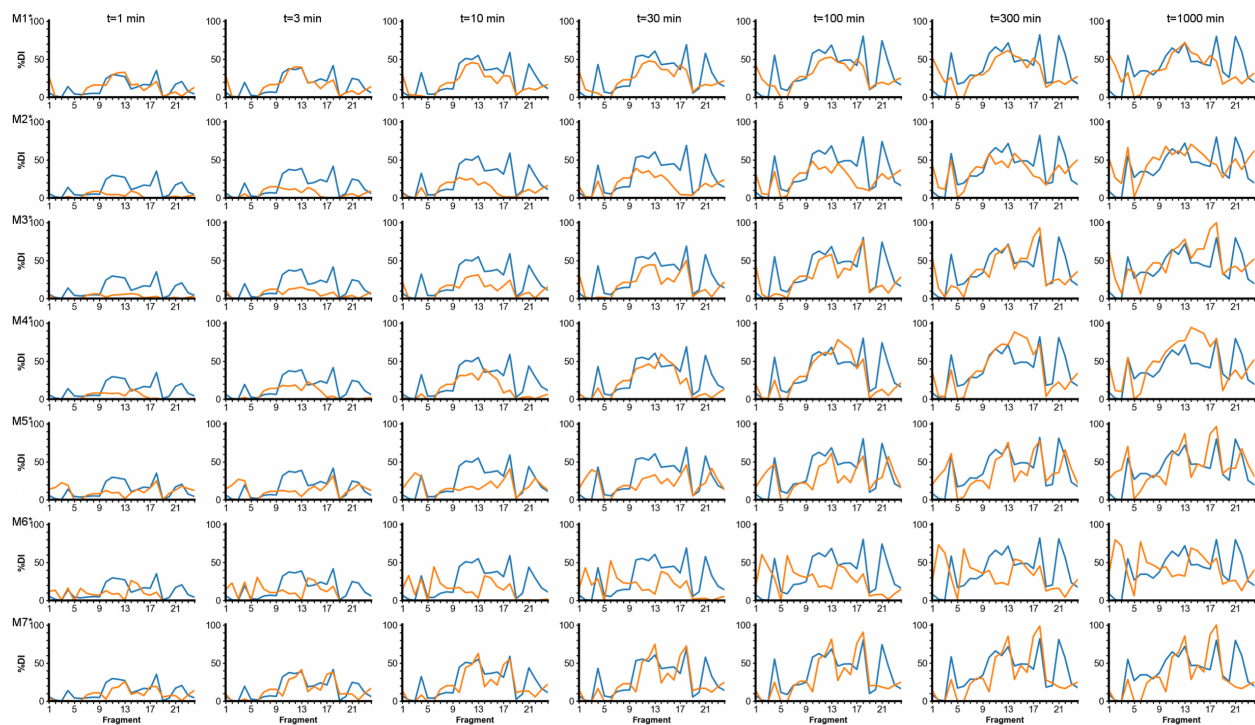


Figure A-15: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and AMBER Force-field

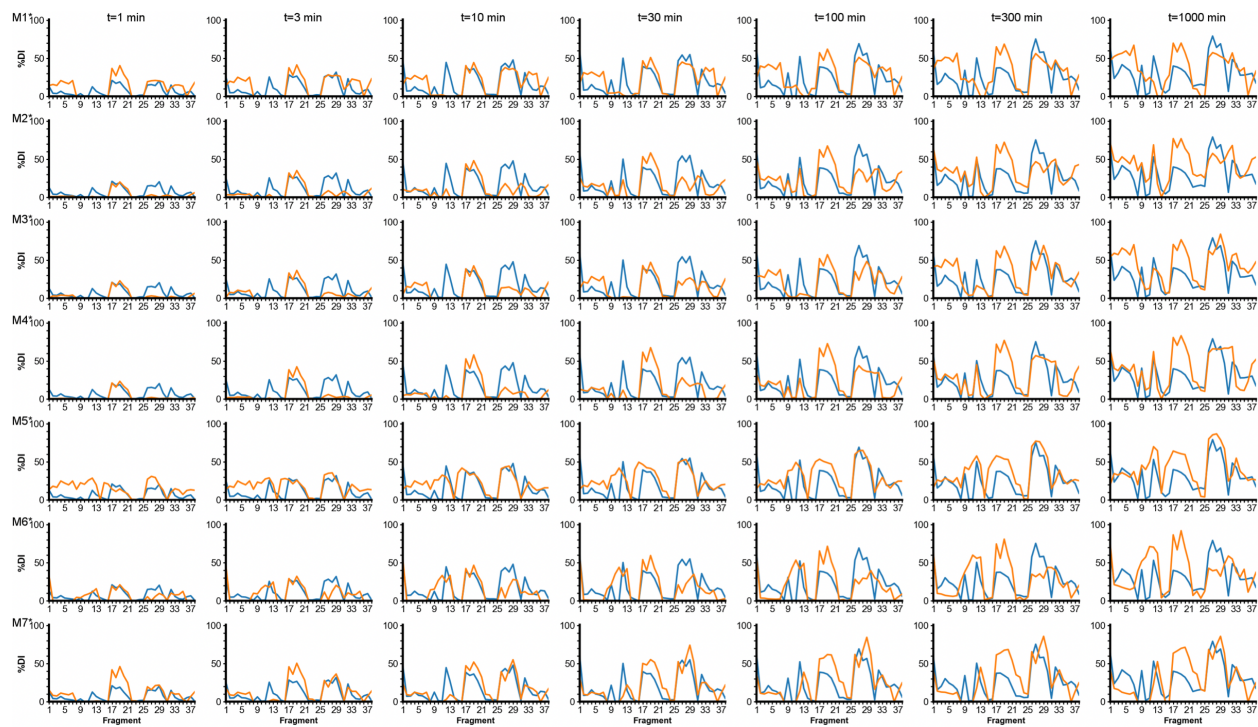


Figure A-16: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and AMBER Force-field.

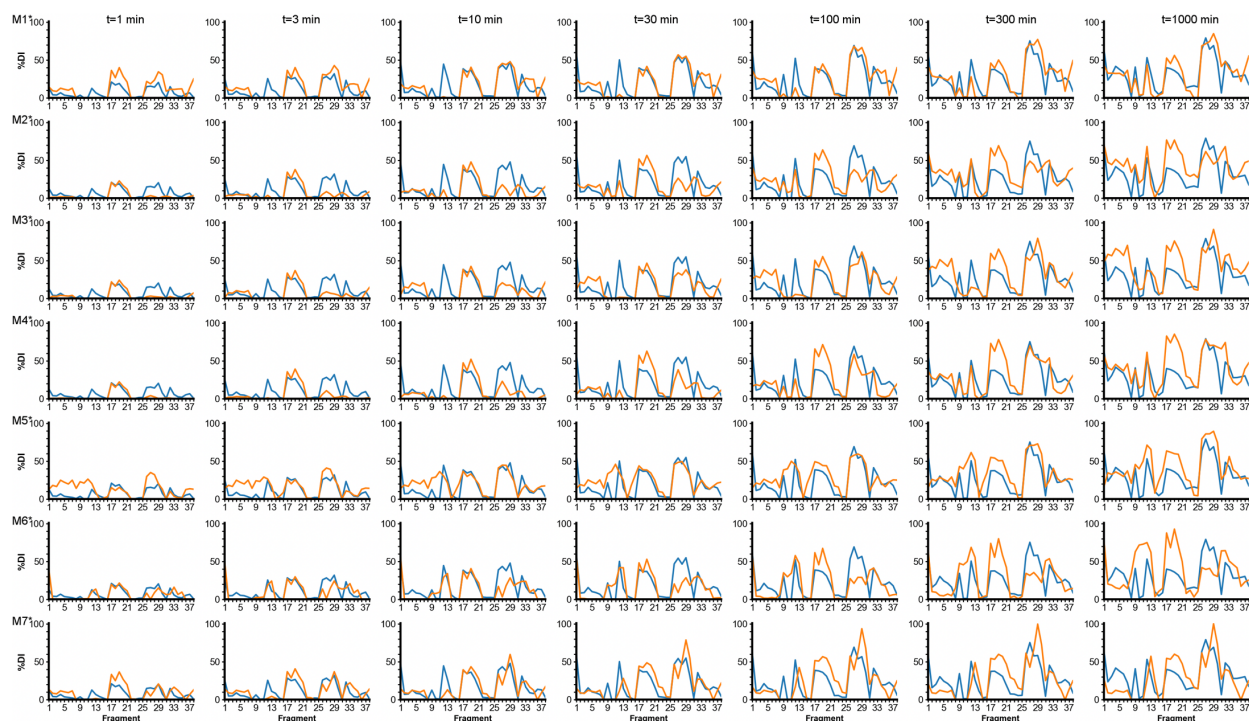


Figure A-17: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and AMBER Force-field.

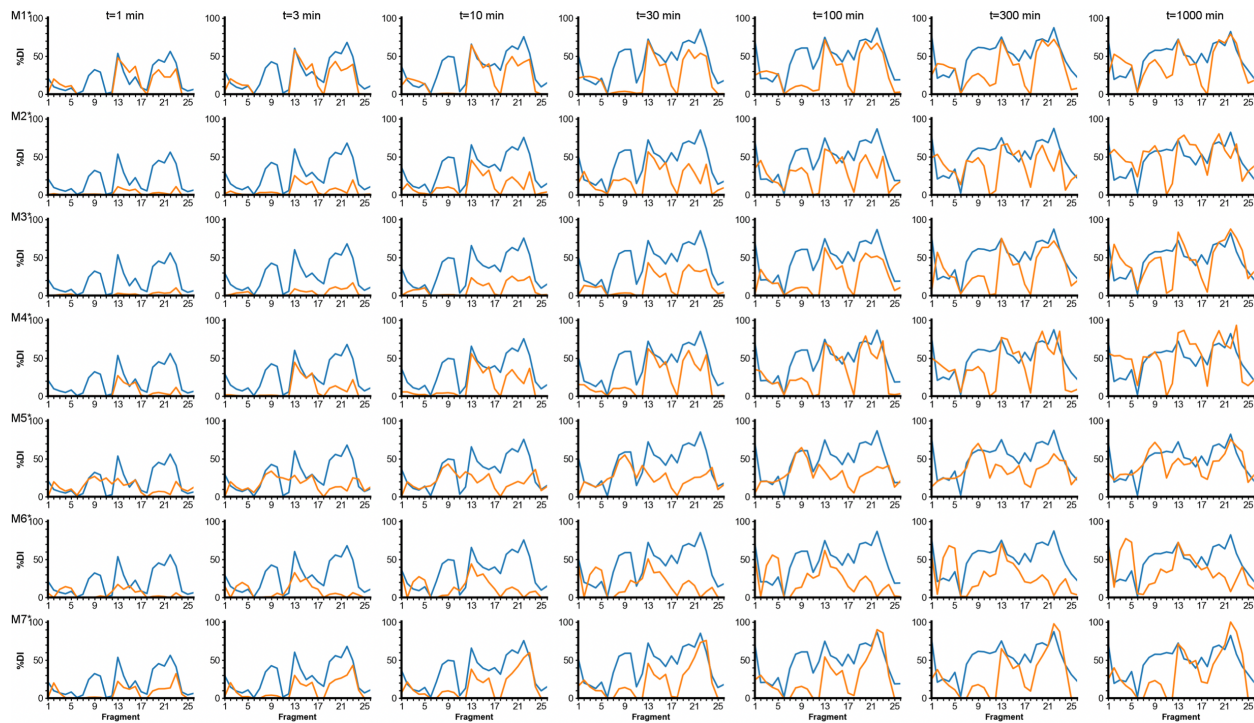


Figure A-18: Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and AMBER Force-field.

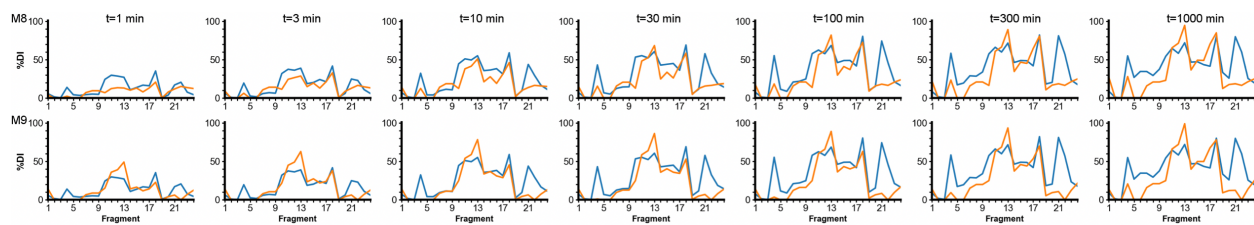


Figure A-19: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and AMBER Force-field.

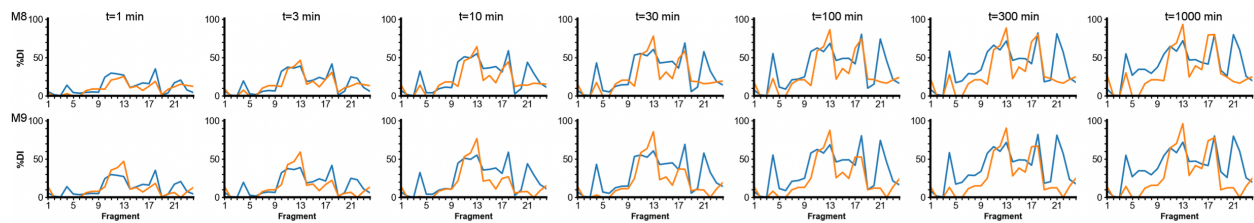


Figure A-20: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and AMBER Force-field.

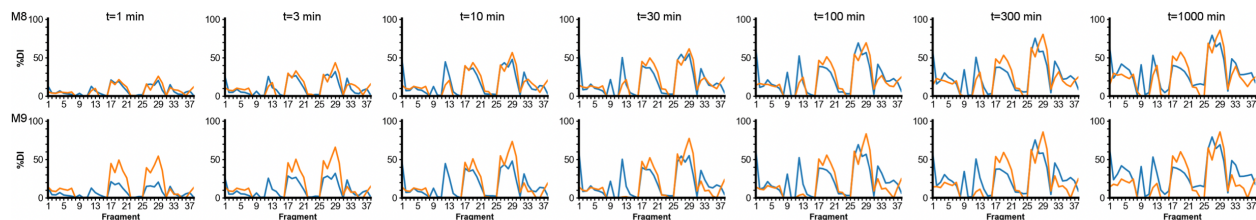


Figure A-21: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and AMBER Force-field.

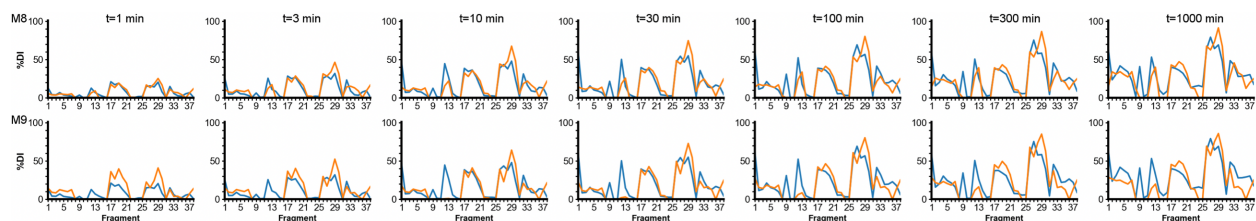


Figure A-22: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and AMBER Force-field.

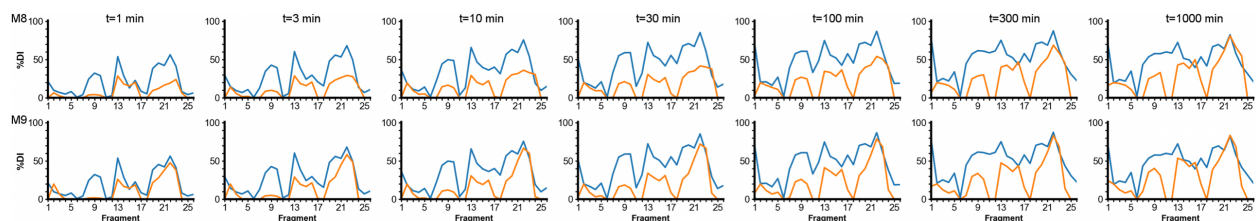


Figure A-23: Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and AMBER Force-field.

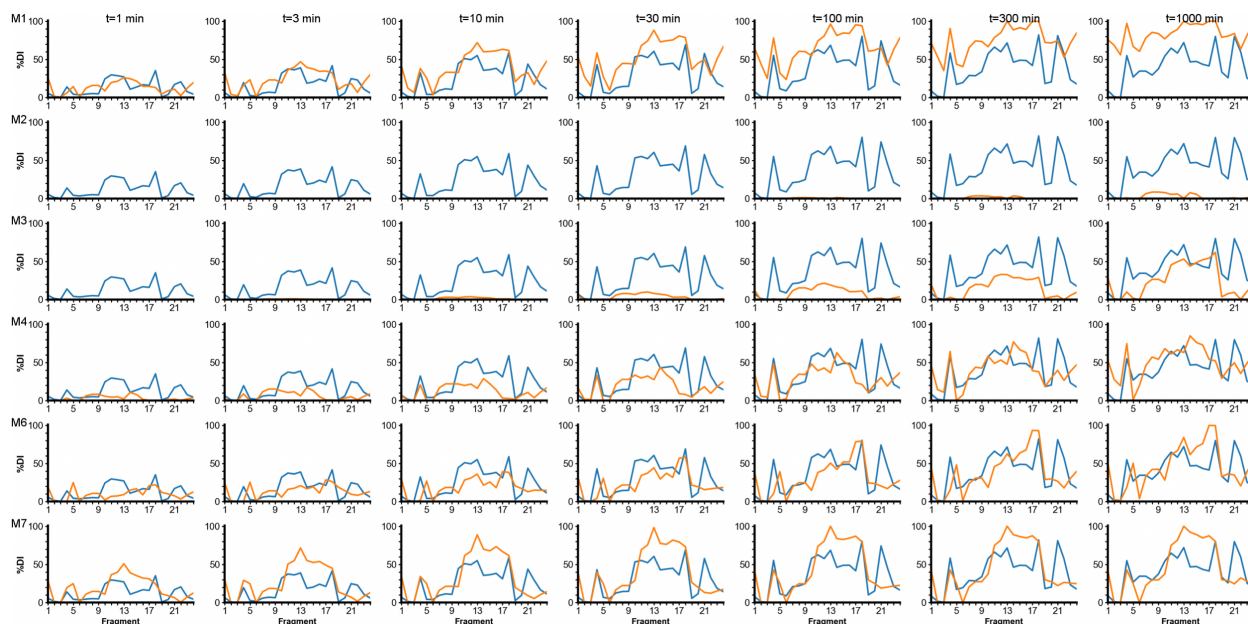


Figure A-24: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1AGR and CHARMM Force-field.

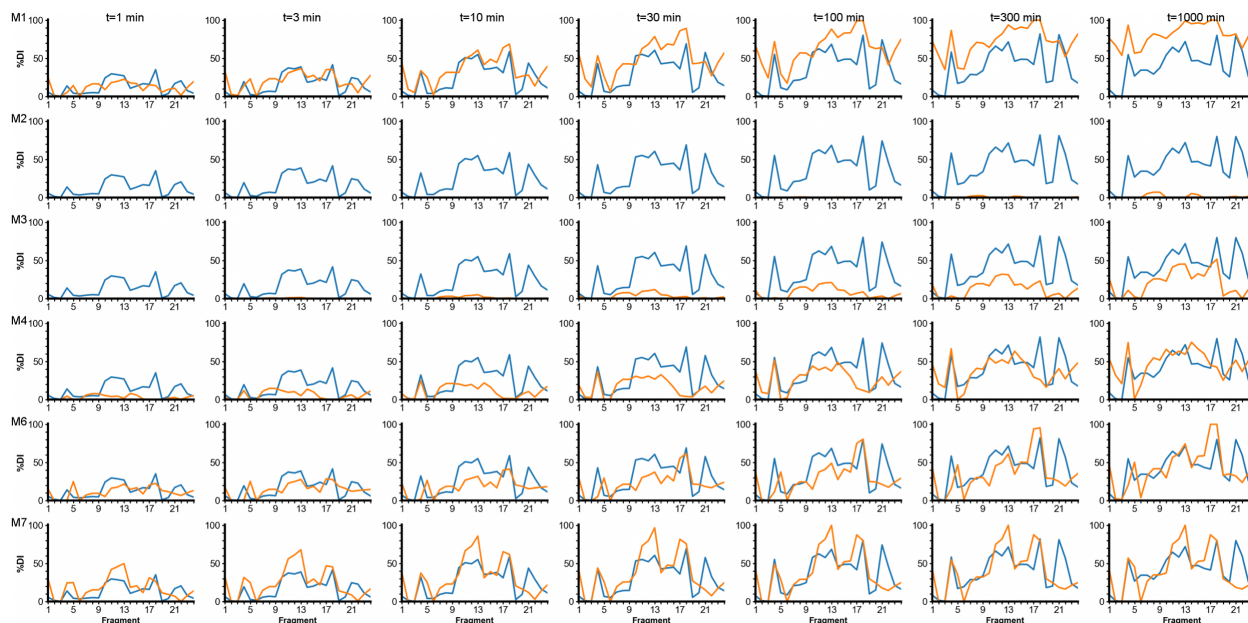


Figure A-25: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1EZT and CHARMM Force-field.

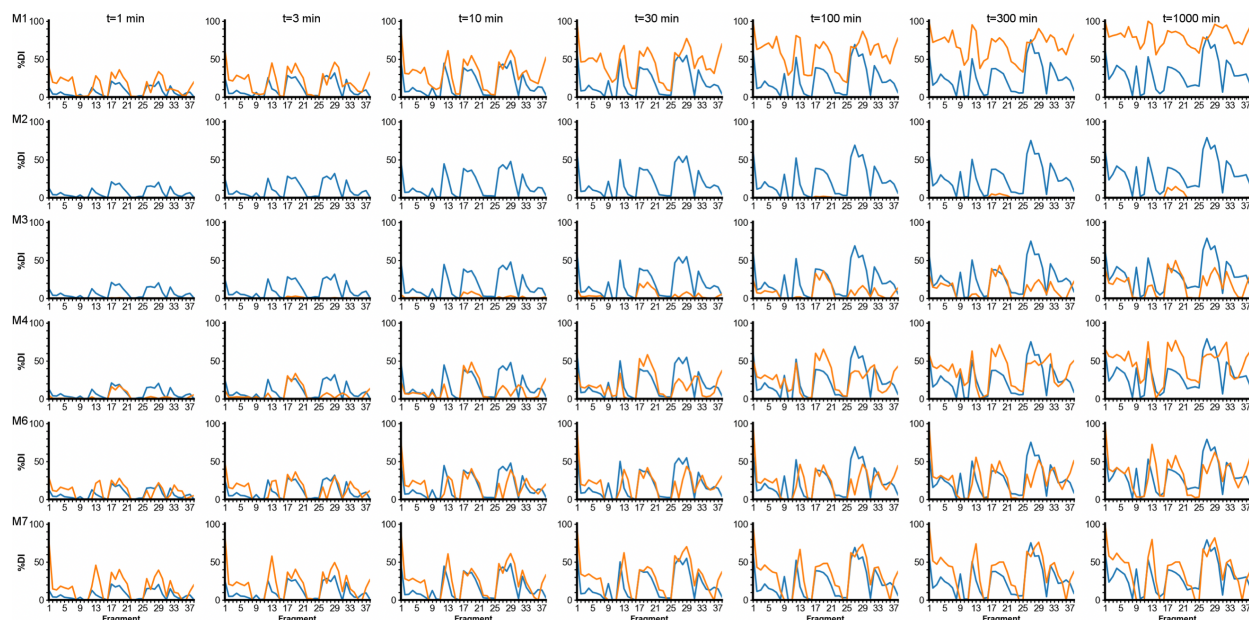


Figure A-26: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2IHD and CHARMM Force-field.

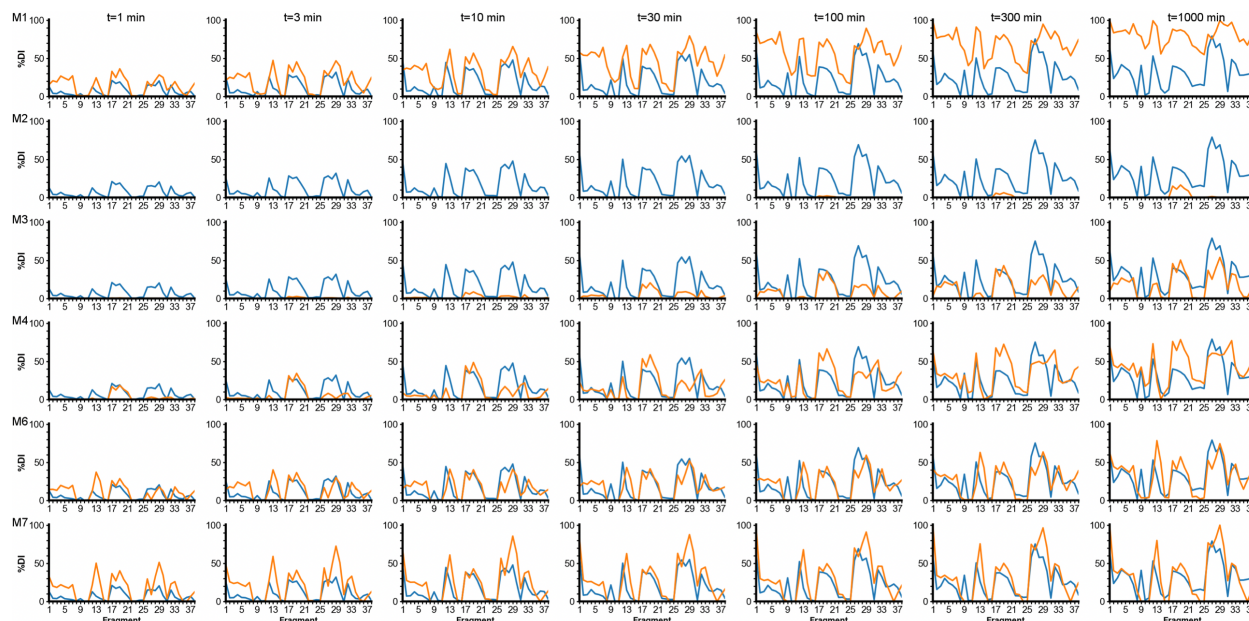


Figure A-27: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:2ODE and CHARMM Force-field.

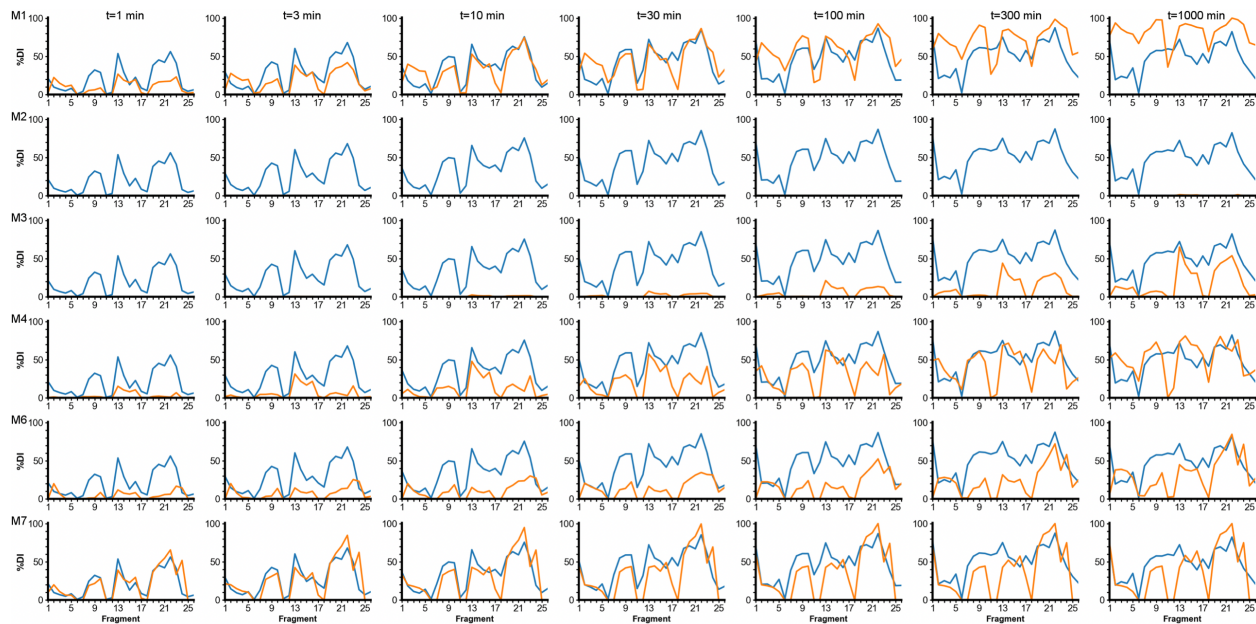


Figure A-28: Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with default parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and CHARMM Force-field.

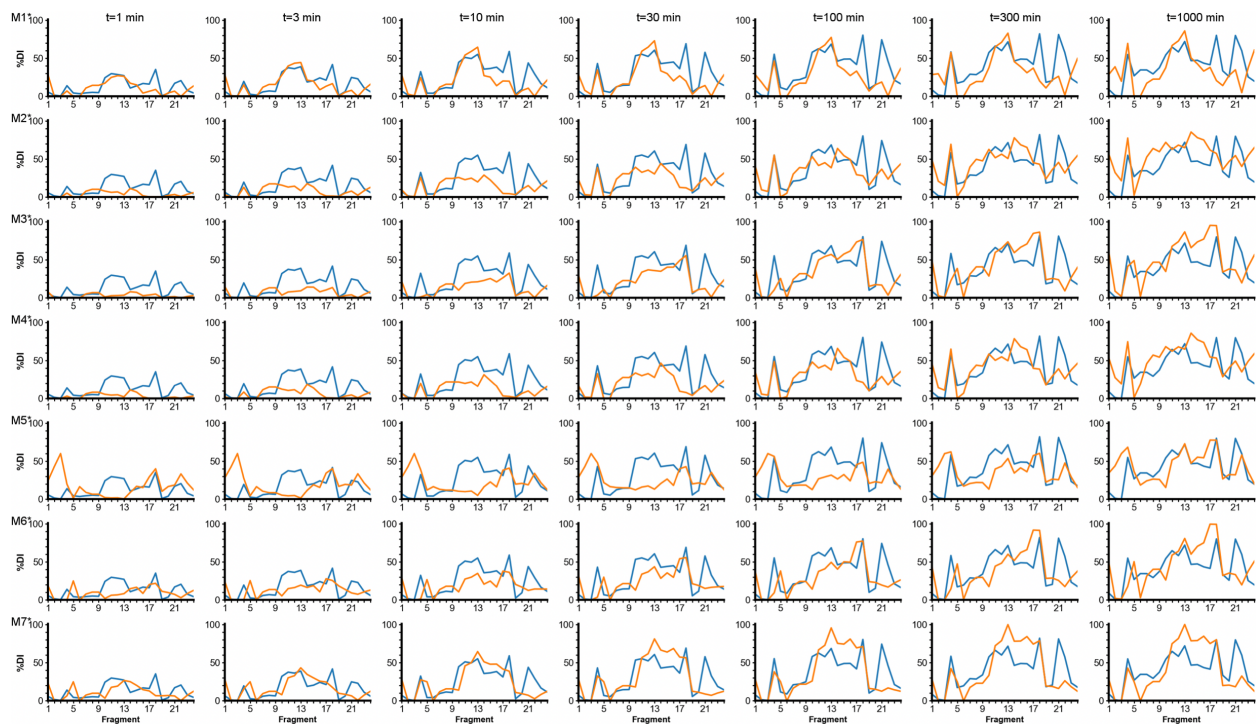


Figure A-29: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and CHARMM Force-field.

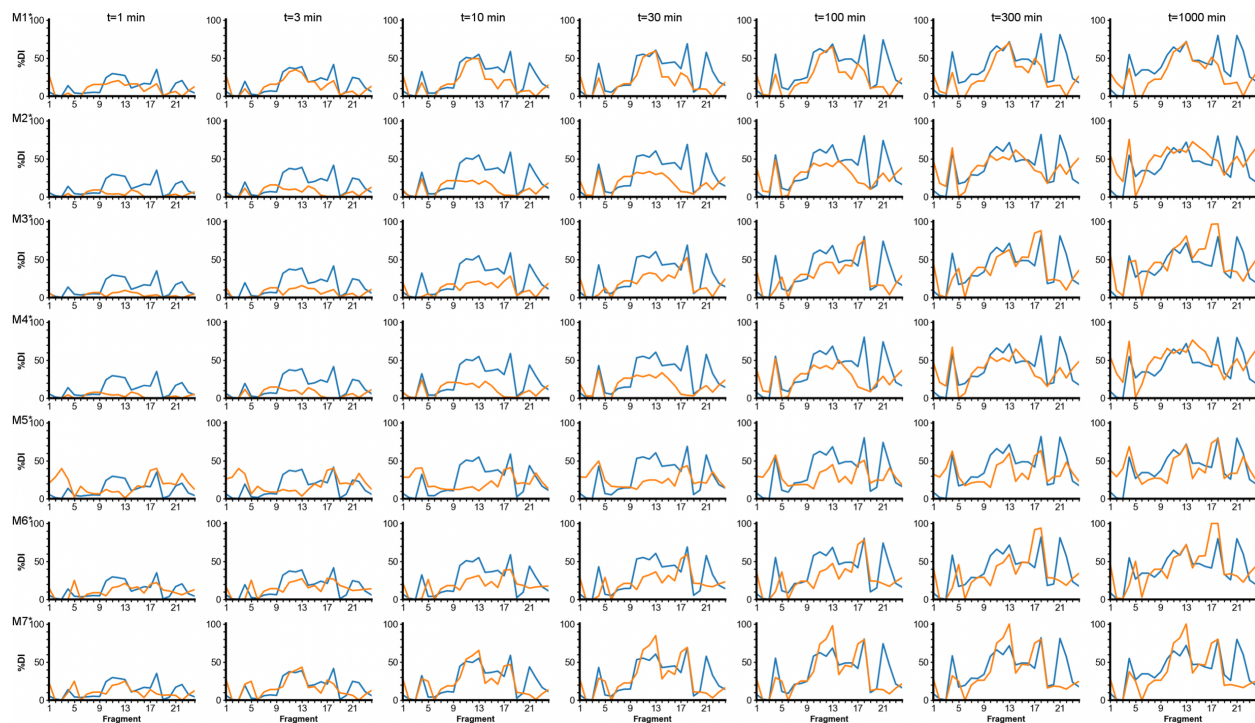


Figure A-30: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and CHARMM Force-field.

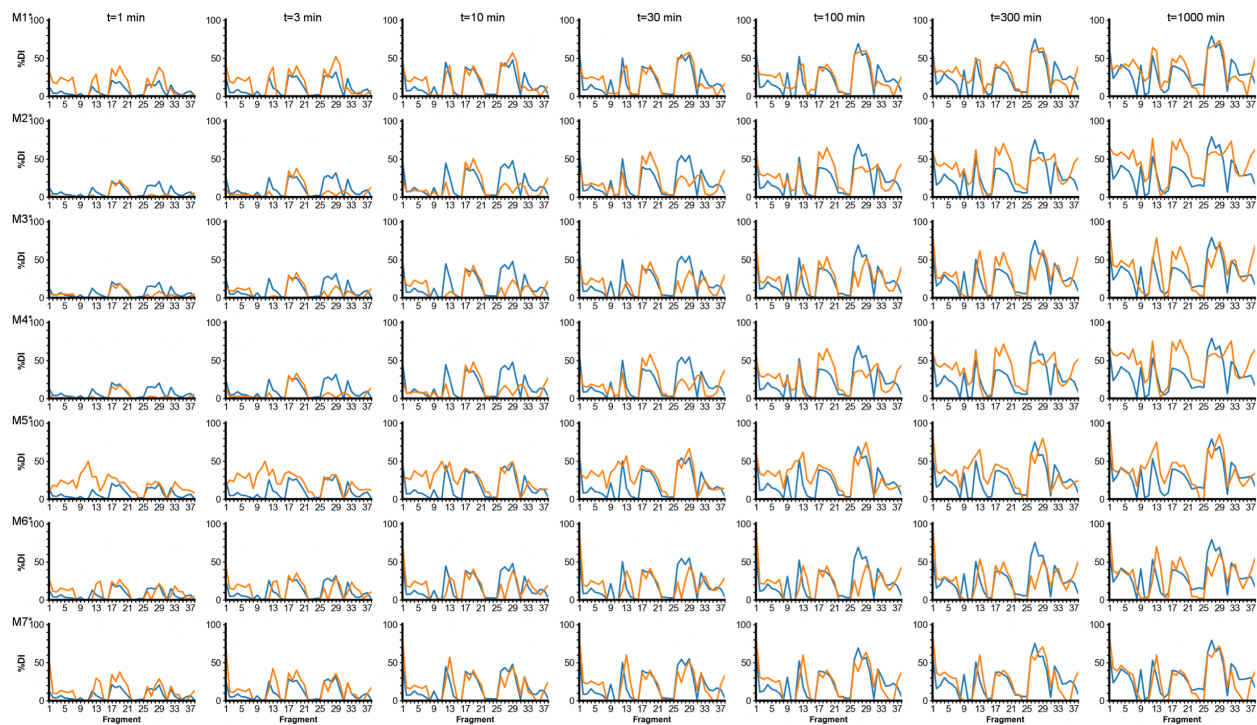


Figure A-31: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and CHARMM Force-field.

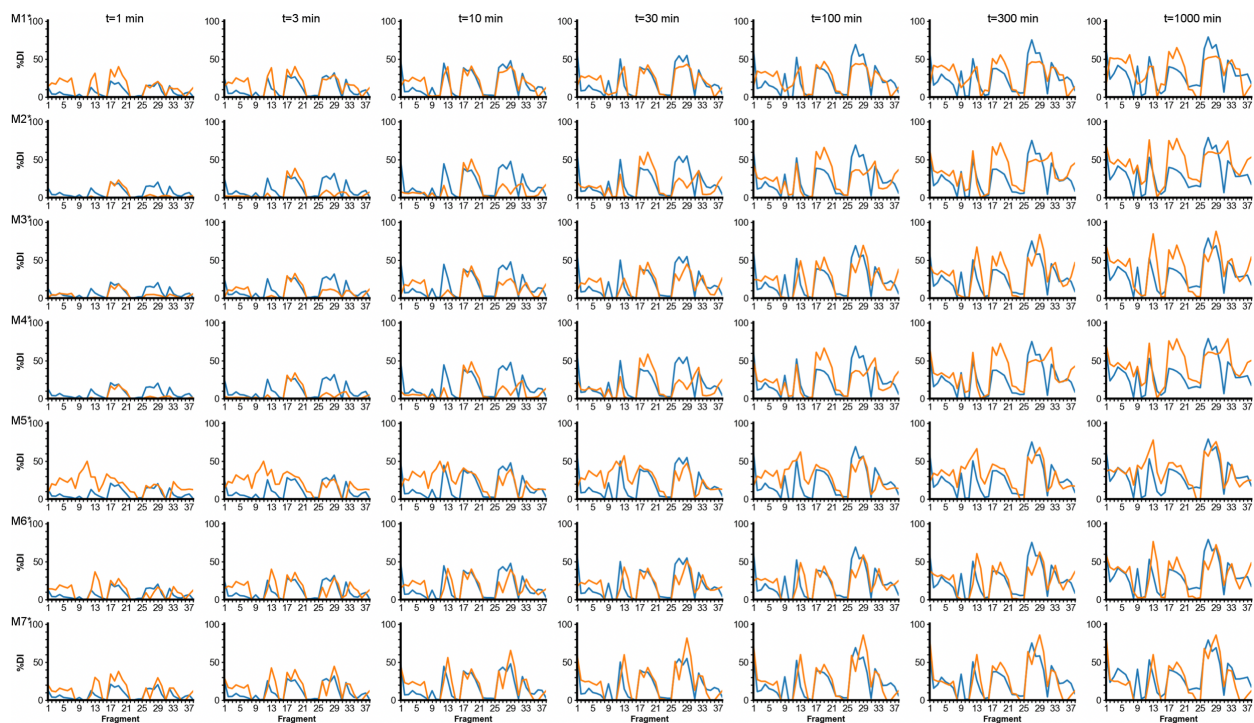


Figure A-32: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and CHARMM Force-field.

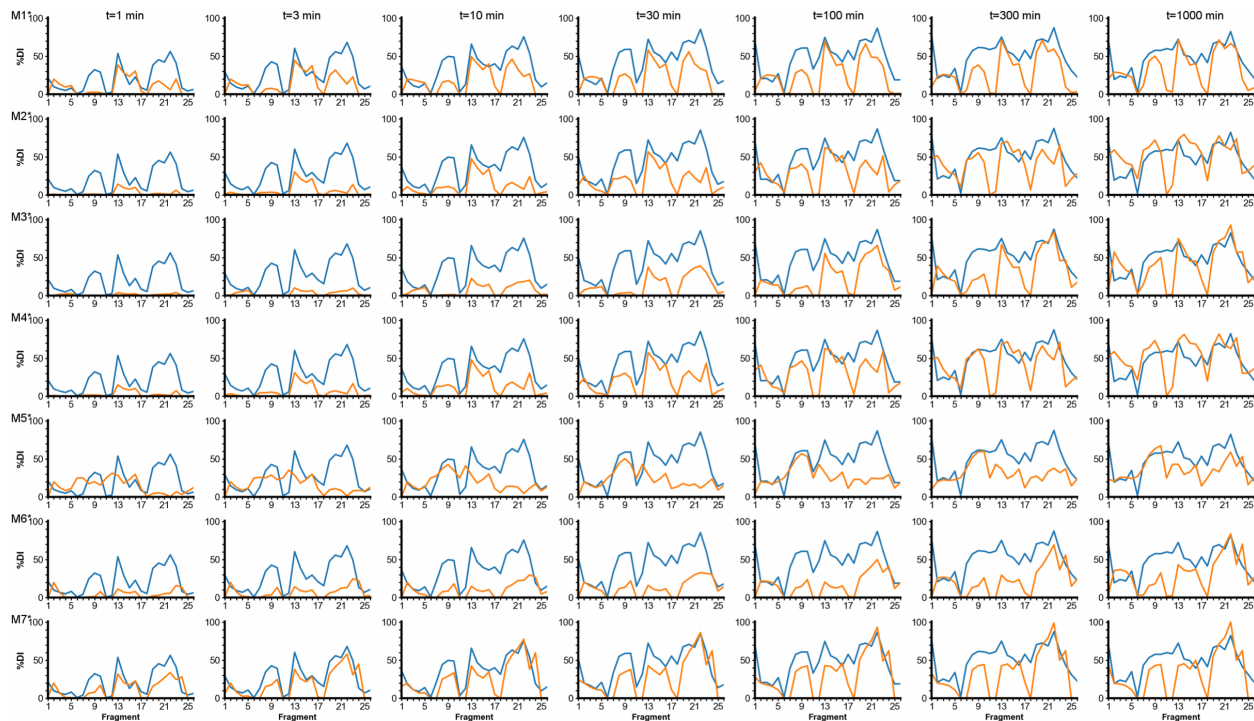


Figure A-33: Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown seven discrete times, alongside each different model with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and CHARMM Force-field.

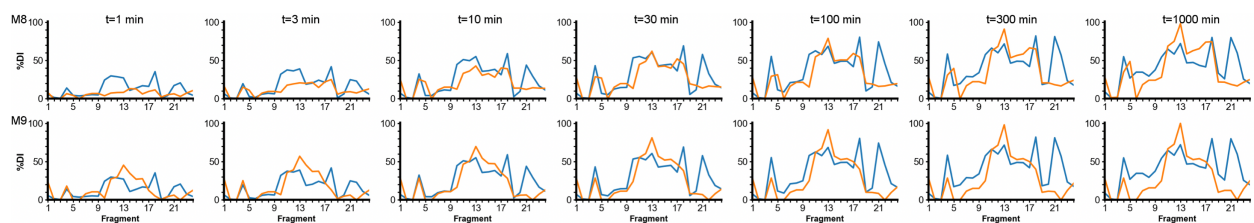


Figure A-34: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1AGR and CHARMM Force-field.

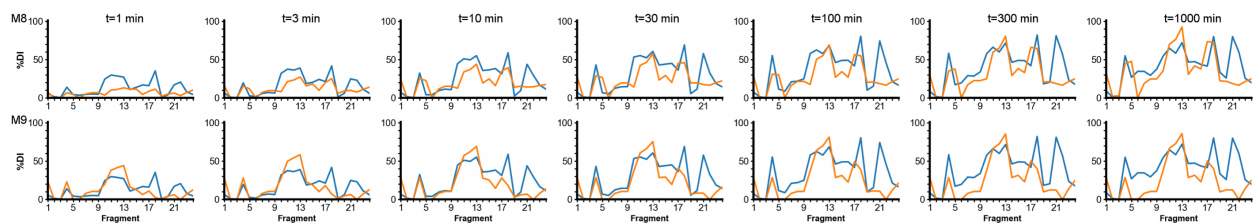


Figure A-35: Modeled deuterium incorporation of fragments in RGS4. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1EZT and CHARMM Force-field.

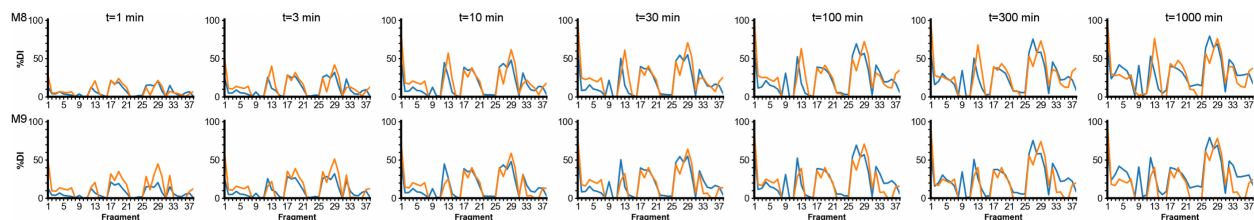


Figure A-36: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2IHD and CHARMM Force-field.

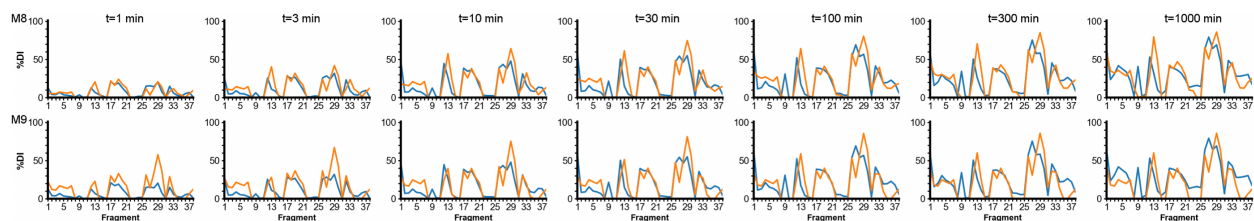


Figure A-37: Modeled deuterium incorporation of fragments in RGS8. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:2ODE and CHARMM Force-field.

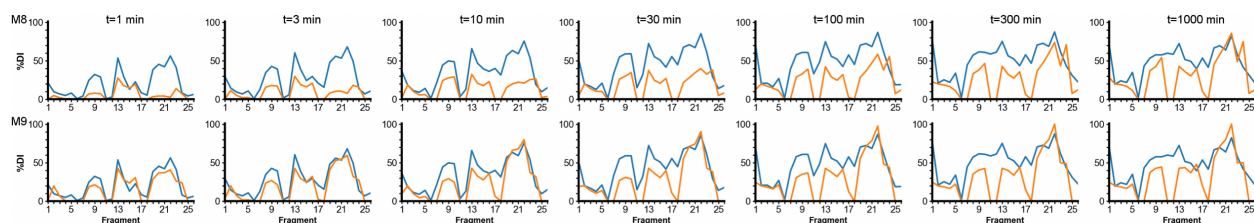


Figure A-38: Modeled deuterium incorporation of fragments in RGS19. The HDX experiment (blue) is shown twice, alongside new models (M8, M9) with optimized parameters (orange). This figure shows the MD simulation results for PDB:1CMZ and CHARMM Force-field.

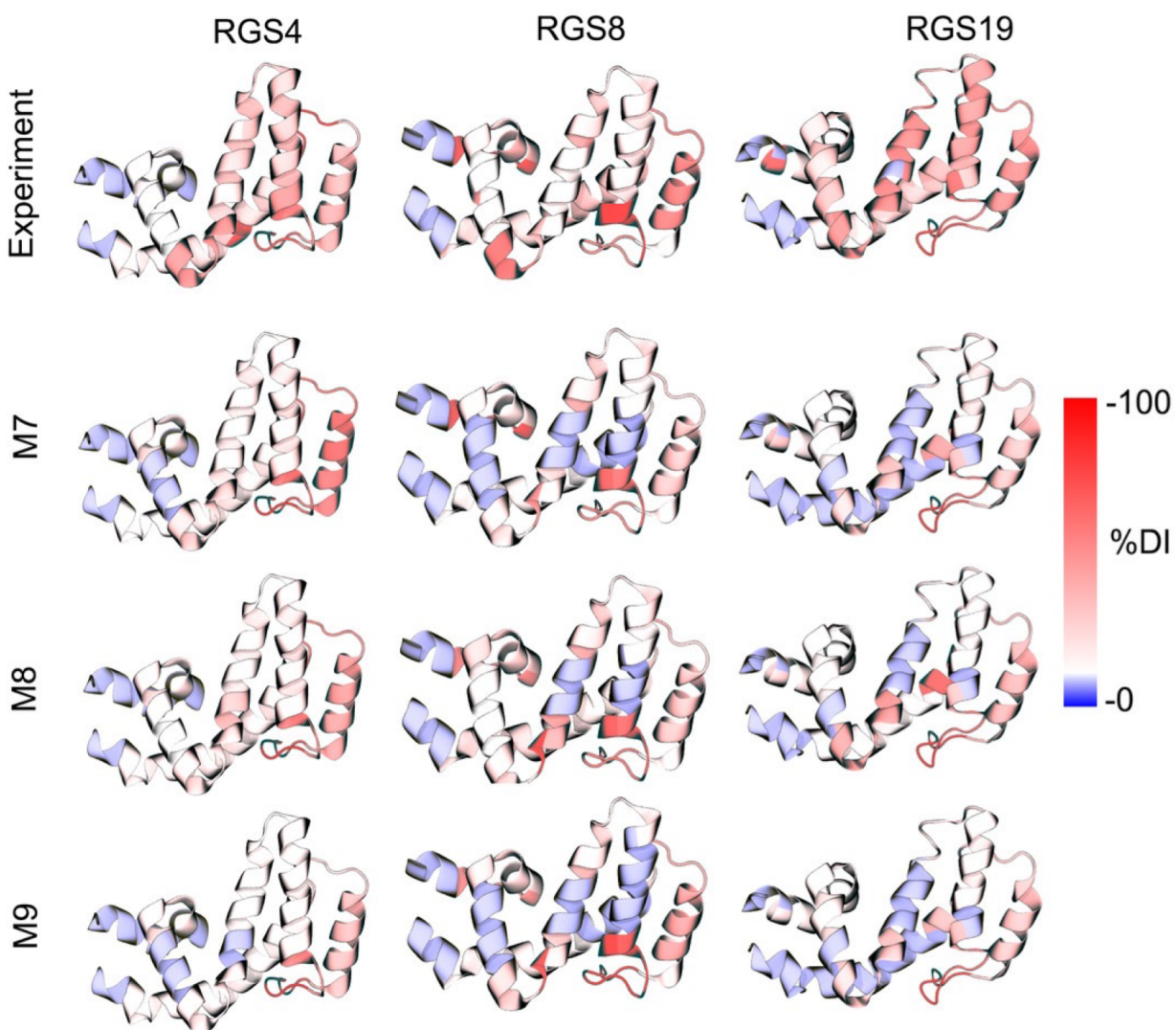


Figure A-39: Deuterium incorporation is mapped on RGS proteins at $t = 1000$ min as observed in experiments and as predicted by the models M7, M8, and M9. Data are presented for the CHARMM-FF simulations of RGS4, RGS8, and RGS19.

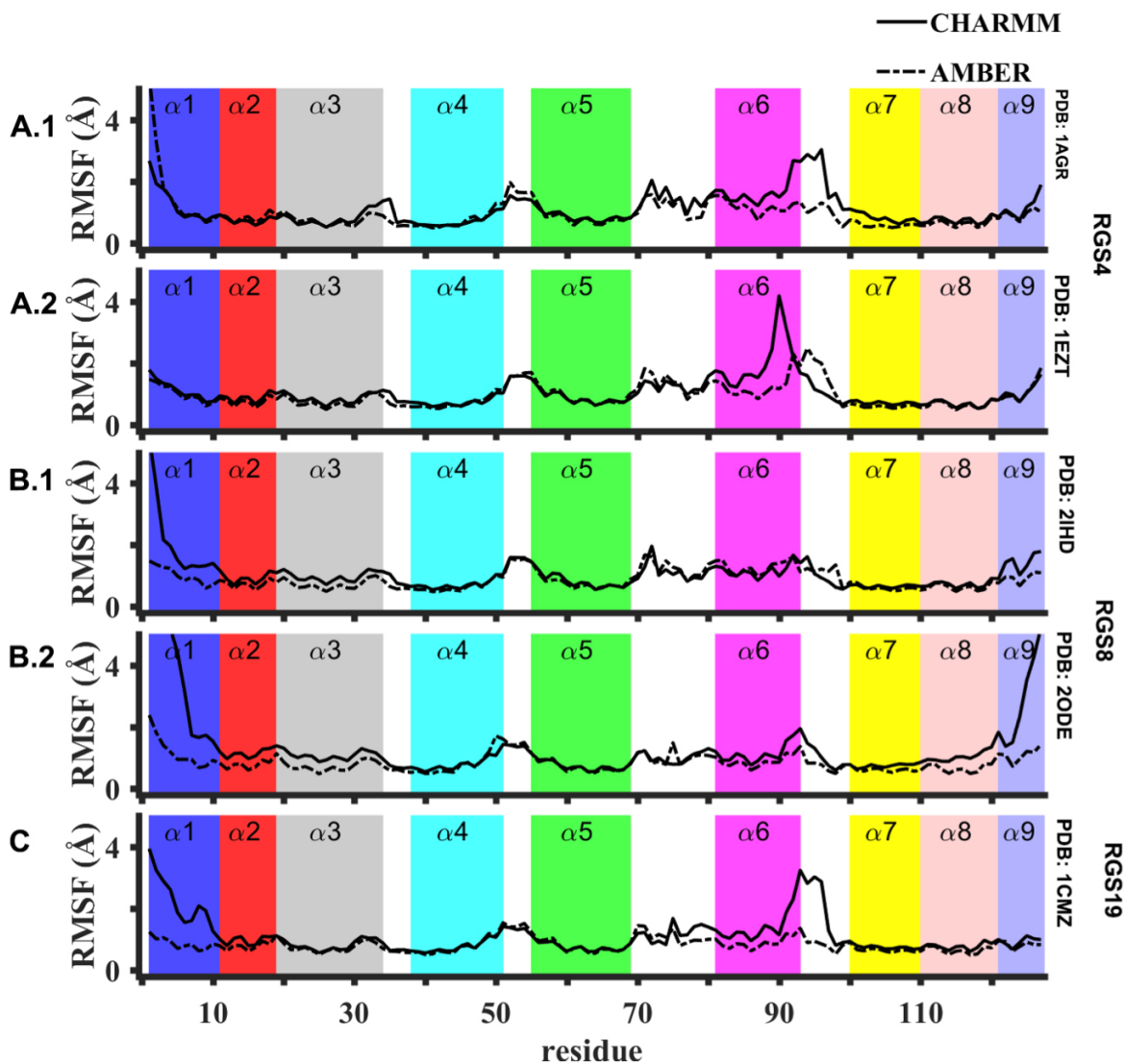


Figure A-40: Root mean squared fluctuations (RMSF) per residue across protein sequences are shown from 2- μ s long MD simulations of (A) RGS4 (PDB: 1AGR, 1EZT), (B) RGS8 (PDB: 2IHD, 2ODE), and (C) RGS19 (PDB: 1CMZ). Color bars indicate helical regions.

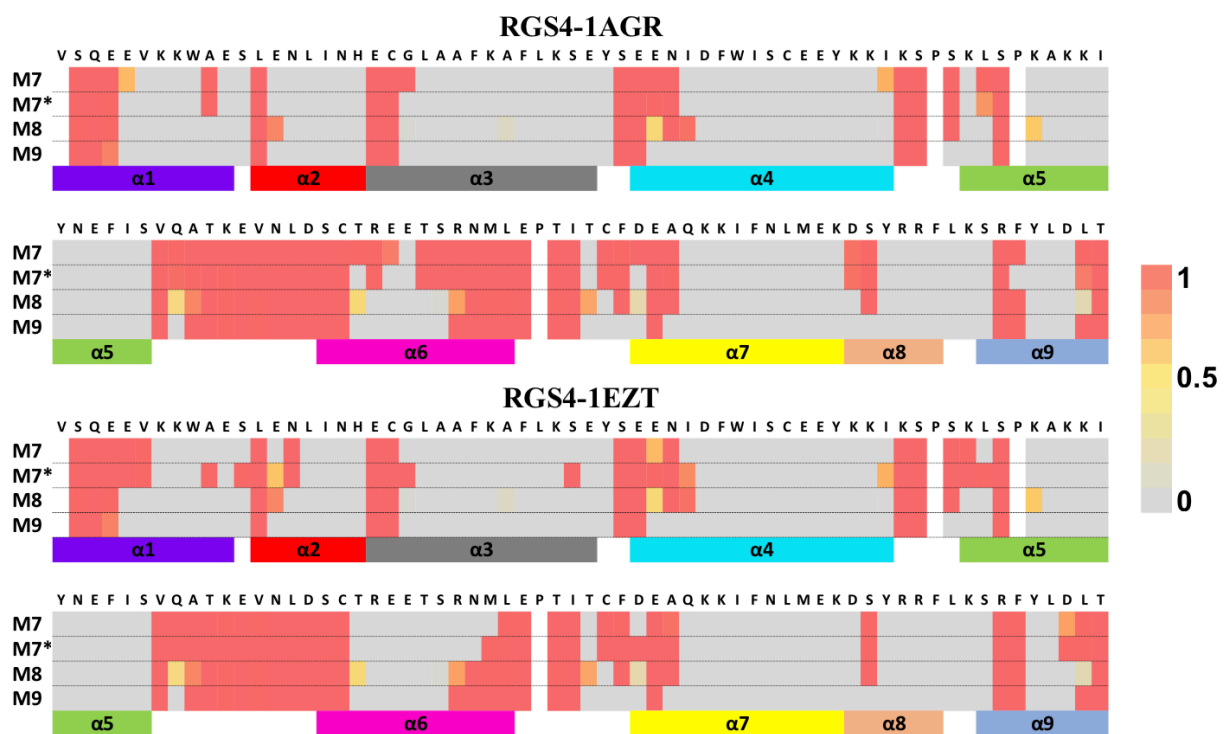


Figure A-41: Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS4, CHARMM-FF).

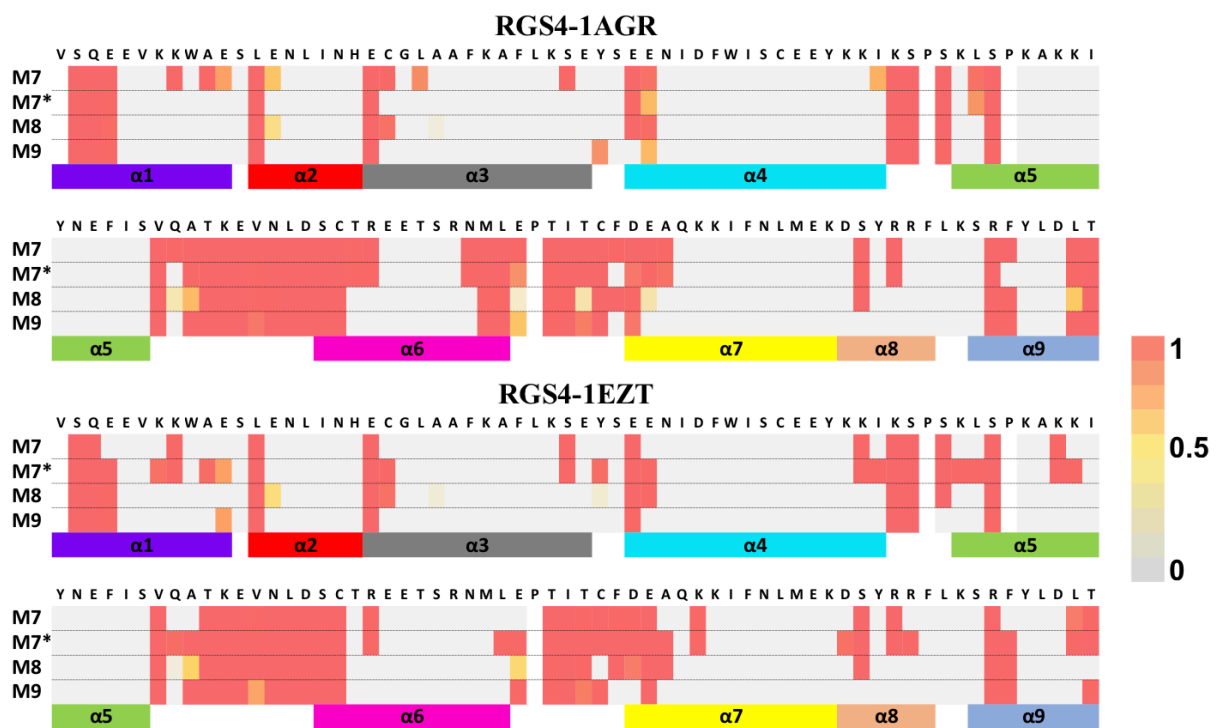


Figure A-43: Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS4, AMBER-FF).

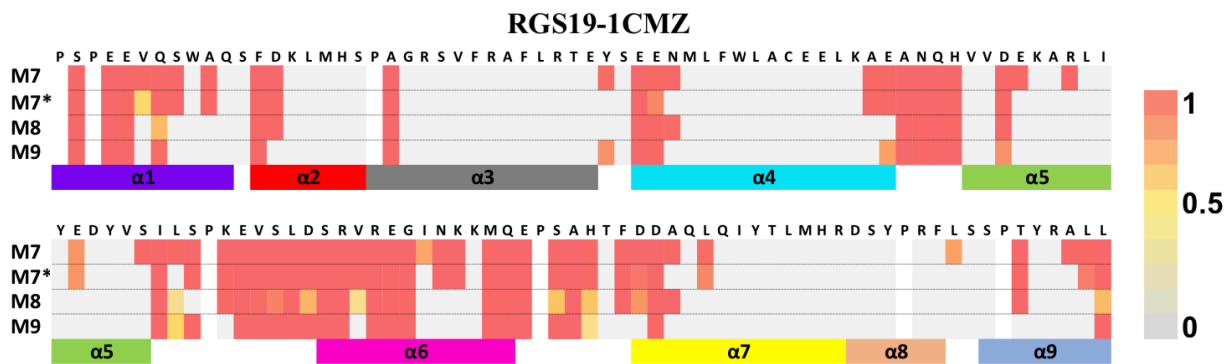


Figure A-46: Modeled deuterium incorporation at $t = 1000$ min at a single-residue resolution (RGS19, AMBER-FF).

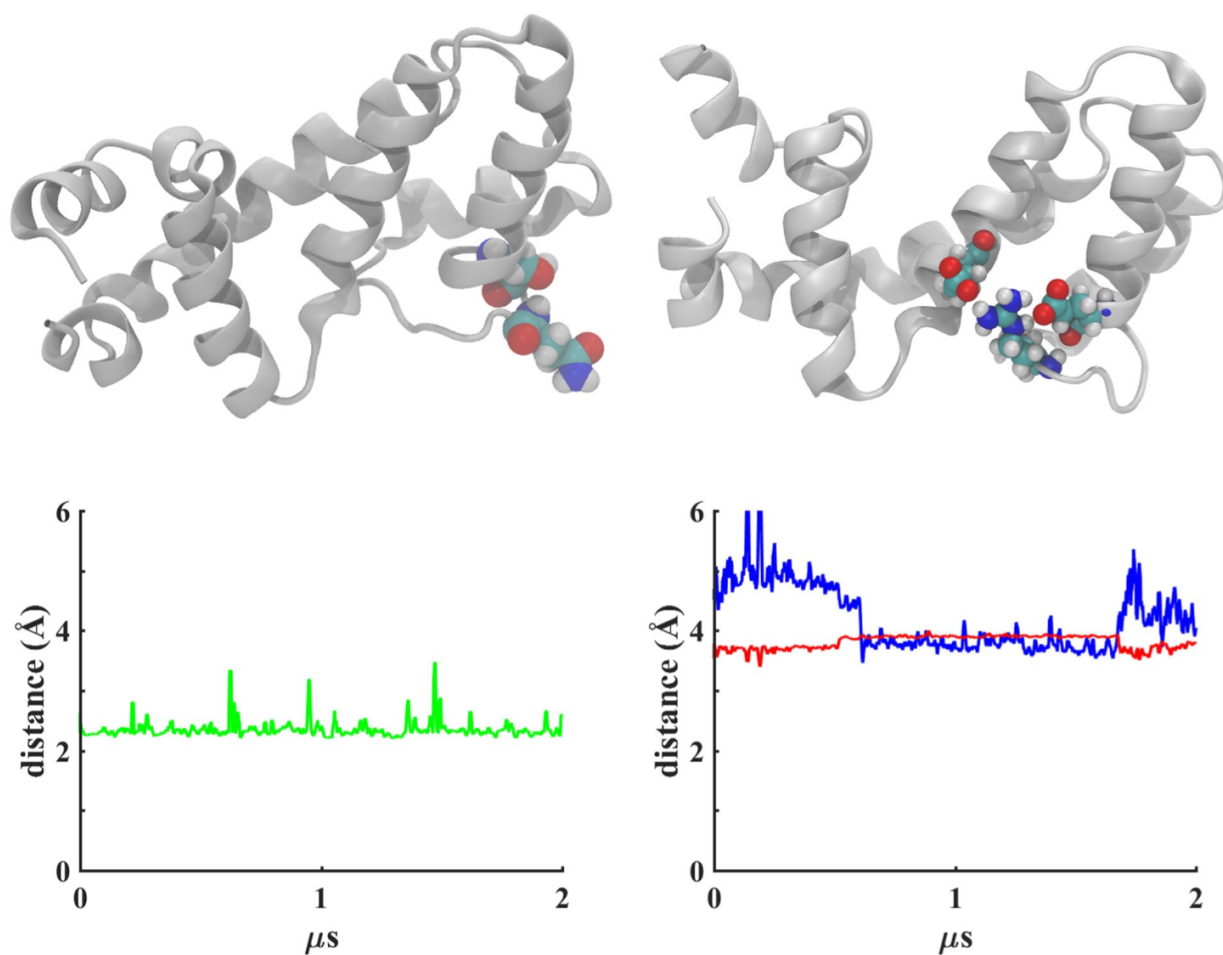


Figure A-47: The residues protected by hydrogen-bonds or salt-bridging interactions are highlighted (panels A and B). The traces for distances between the centers-of-masses of residue pairs are shown in panel C (S120-Q122) and panel D (E84-R119 and E111-R119).

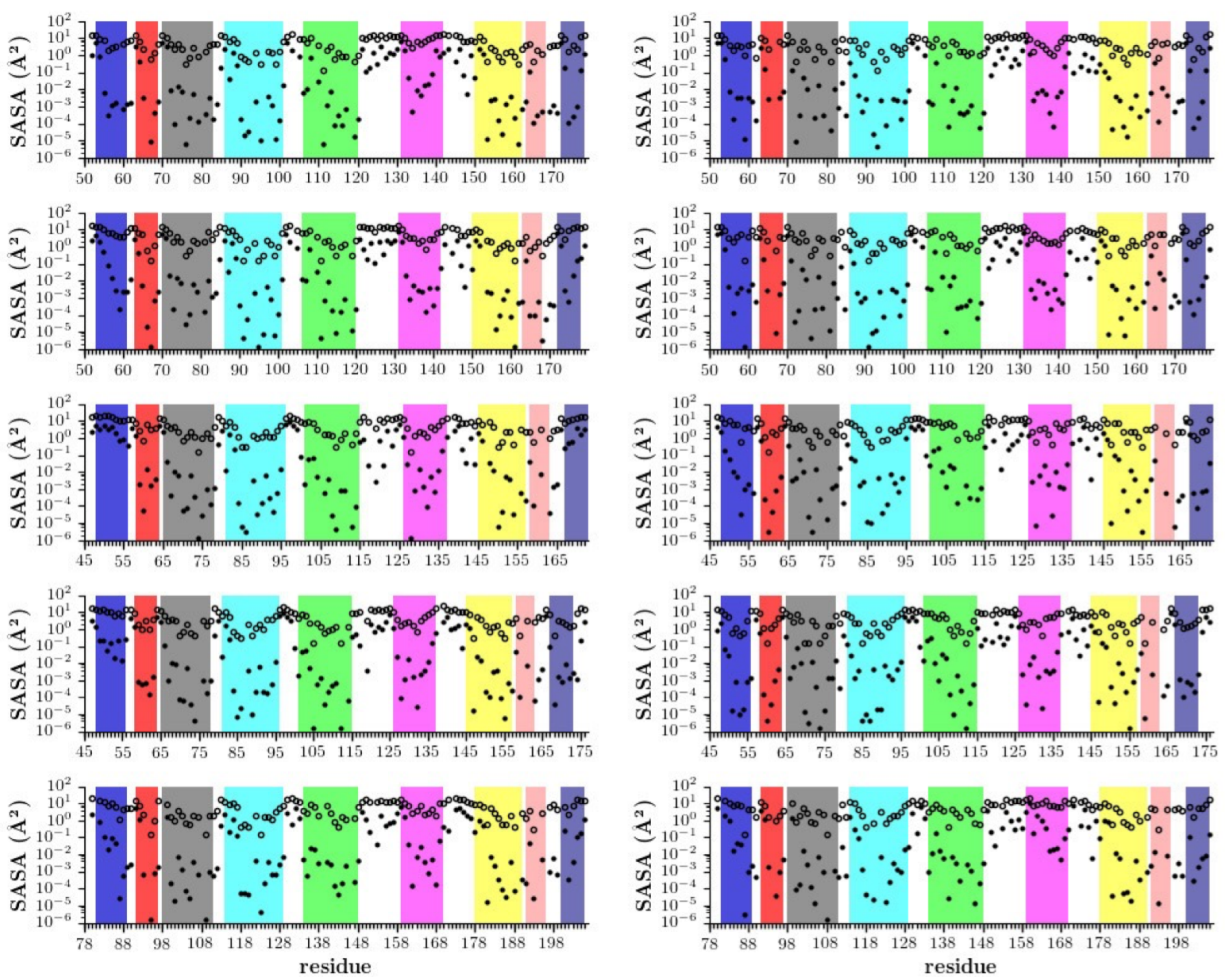


Figure A-48: SASA data similar to Fig. A-6 are shown from MD simulations of all RGS proteins for both force-fields (CHARMM-FF, panel A; AMBER-FF, panel B). Color and labeling details are similar to Fig. A-6

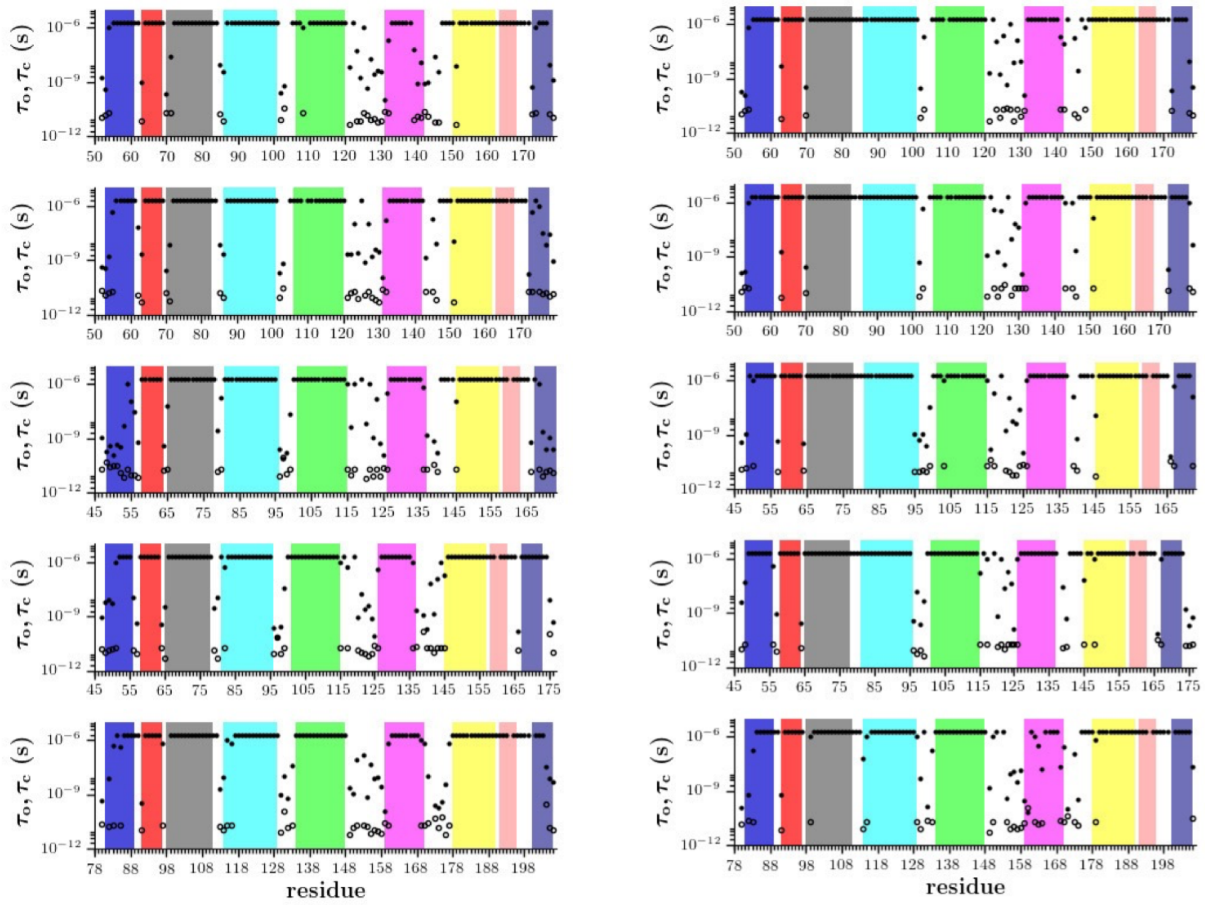


Figure A-49: Corrected mean residence times for open-states of amide hydrogens are shown. Other details are similar to Fig. A-6.

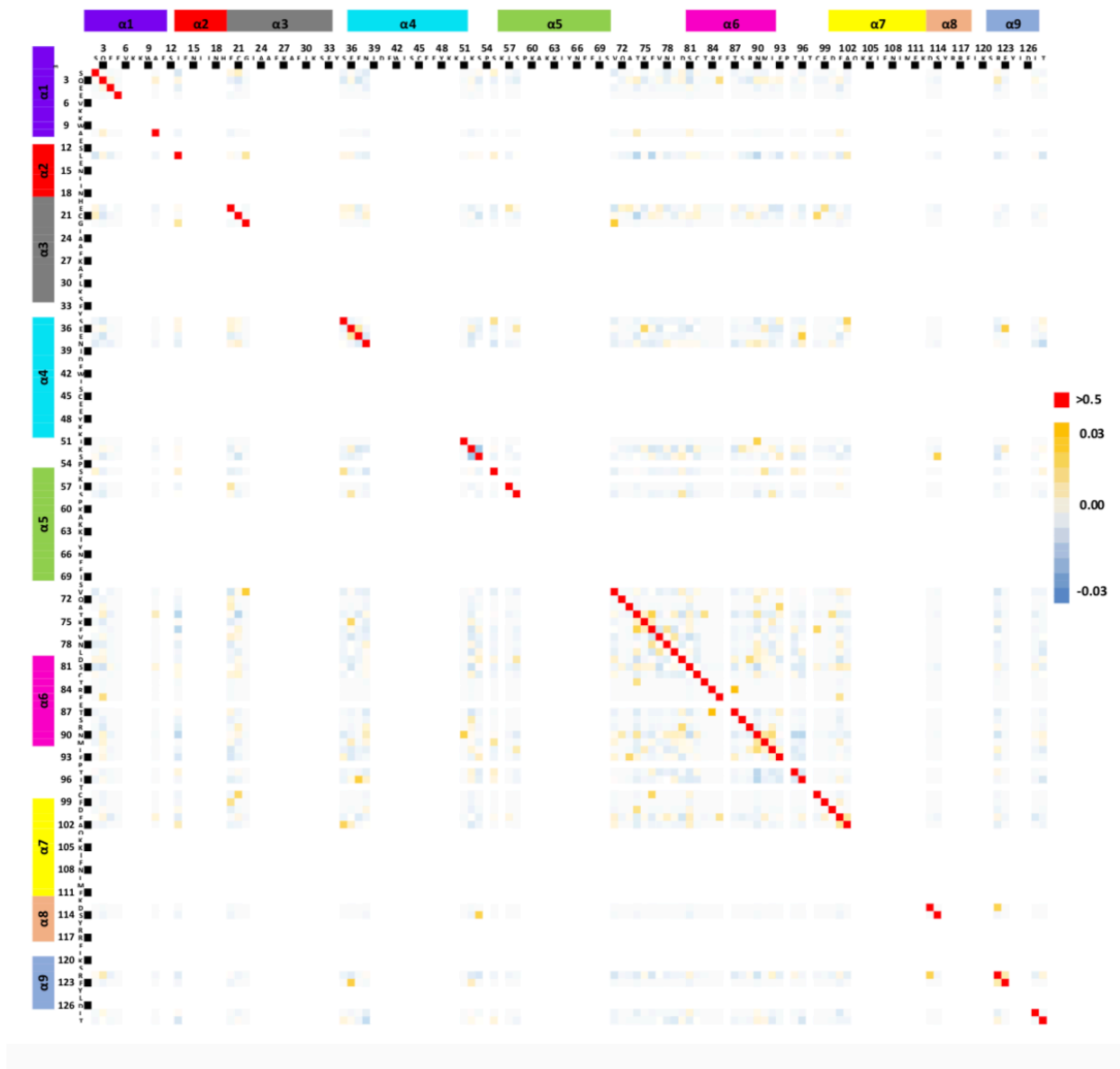


Figure A-50: Residue-residue correlations among open states of all amide-hydrogens (CHARMM-FF, RGS4 (PDB code 1AGR), model M7). The correlation matrix is calculated based on the probability that two amide hydrogens simultaneously explore open states; $C(i, j) = (P(i, j) - P(i)P(j)) / (P(i)P(j)(1 - P(i))(1 - P(j)))^{0.5}$



Figure A-51: Data similar to A-50 are shown for RGS8 (CHARMM-FF, RGS8 (PDB code 2ODE), model M7).

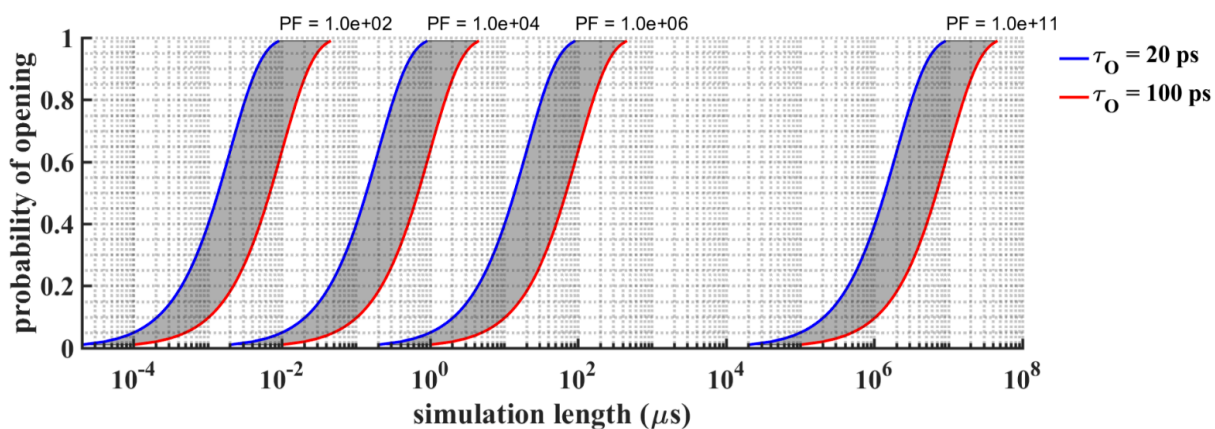


Figure A-52: Probability of a closed to open transition in a given amide vs. simulation length (μs) is presented based upon Poisson statistics. Data are shown for PFs = 102, 104, 106, and 1011 with $\tau_{\text{O}} = 20 \text{ ps}$ and 100 ps .

REFERENCES

REFERENCES

1. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 98–104 (1958).
2. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* **12**, 88–118 (1965).
3. Feixas, F., Lindert, S., Sinko, W. & McCammon, J. A. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophysical Chemistry* **186**, 31–45 (2014).
4. Tobin, D. & Bahar, I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. *Proceedings of the National Academy of Sciences* **102**, 18908–18913 (2005).
5. Mittag, T., Kay, L. E. & Forman-Kaya, J. D. Protein dynamics and conformational disorder in molecular recognition. *Journal of Molecular Recognition* **23**, 105–116 (2010).
6. Latorraca, N. R., Venkatakrisnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chemical Reviews* **117**, 139–155 (2017).
7. Eyrisch, S. & Helms, V. Transient pockets on protein surfaces involved in protein-protein interaction. *Journal of Medicinal Chemistry* **50**, 3457–3464 (2007).
8. Metz, A. *et al.* Hot Spots and Transient Pockets: Predicting the Determinants of Small-Molecule Binding to a Protein-Protein Interface. *Journal of Chemical Information and Modeling* **52**, 120–133 (2012).
9. Langley, J. N. On the reaction of cells and of nerve-endings to certain poisons, chiefly as regards the reaction of striated muscle to nicotine and to curari. *The Journal of Physiology* **33**, 374–413 (1905).
10. Parascandola, J. Origins of the receptor theory. *Trends in Pharmacological Sciences* **1**, 189–192 (1979).
11. Maehle, A.-H. A binding question: the evolution of the receptor concept. *Endeavour* **33**, 135–40 (2009).
12. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nature Reviews Drug Discovery* **1**, 727–730 (2002).
13. Taylor, D. Fewer new drugs from the pharmaceutical industry. *British Medical Journal*

(*Clinical Research Edition*) **326**, 408–9 (2003).

14. Evans, R. M. & Mangelsdorf, D. J. Nuclear Receptors, RXR, and the Big Bang. *Cell* **157**, 255–66 (2014).

15. Cohen, P. & Alessi, D. R. Kinase drug discovery - What's next in the field? **8**, 96–104 (2013).

16. Botta, M. New Frontiers in Kinases: Special Issue. *ACS Medicinal Chemistry Letters* **5**, 270–270 (2014).

17. Sperandio, O., Reynès, C. H., Camproux, A.-C. & Villoutreix, B. O. Rationalizing the chemical space of protein–protein interaction inhibitors. *Drug Discovery Today* **15**, 220–229 (2010).

18. Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotechnology* **25**, 71–75 (2007).

19. Wells, J. A. & McClendon, C. L. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **450**, 1001–1009 (2007).

20. Conte, L. L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *Journal of Molecular Biology* **285**, 2177–98 (1999).

21. Stanton, B. Z. *et al.* A small molecule that binds Hedgehog and blocks its signaling in human cells. *Nature Chemical Biology* **5**, 154–156 (2009).

22. Ran, X. & Gestwicki, J. E. Inhibitors of protein–protein interactions (PPIs): an analysis of scaffold choices and buried surface area. *Current Opinion in Chemical Biology* **44**, 75–86 (2018).

23. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chemistry and Biology* **21**, 1102–1114 (2014).

24. Gowthaman, R., Deeds, E. J. & Karanicolas, J. Structural Properties of Non-Traditional Drug Targets Present New Challenges for Virtual Screening. *Journal of Chemical Information and Modeling* **53**, 2073–2081 (2013).

25. Dandapani, S. & Marcaurelle, L. A. Grand Challenge Commentary: Accessing new chemical space for 'undruggable' targets. *Nature Chemical Biology* **6**, 861–863 (2010).

26. Bauer, R. A., Wurst, J. M. & Tan, D. S. Expanding the range of 'druggable' targets with natural product-based libraries: an academic perspective. *Current Opinion in Chemical Biology* **14**, 308–14 (2010).

27. Changeux, J.-P. & Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **308**, 1424–8 (2005).
28. Tsai, C.-J., Sol, A. del & Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Molecular Biosystems* **5**, 207 (2009).
29. Conn, P. J., Christopoulos, A. & Lindsley, C. W. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nature Reviews Drug Discovery* **8**, 41–54 (2009).
30. May, L. T., Leach, K., Sexton, P. M. & Christopoulos, A. Allosteric Modulation of G Protein–Coupled Receptors. *Annual Review of Pharmacology and Toxicology* **47**, 1–51 (2007).
31. Roman, D. L., Blazer, L. L., Monroy, C. A. & Neubig, R. R. Allosteric inhibition of the regulator of G protein signaling-G α protein-protein interaction by CCG-4986. *Molecular Pharmacology* **78**, 360–365 (2010).
32. Kunze, J. *et al.* Targeting dynamic pockets of HIV-1 protease by structure-based computational screening for allosteric inhibitors. *Journal of Chemical Information and Modeling* **54**, 987–991 (2014).
33. Appelt, K. *et al.* Design of enzyme inhibitors using iterative protein crystallographic analysis. *Journal of Medicinal Chemistry* **34**, 1925–34 (1991).
34. Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **257**, 1078–1082 (1992).
35. Warren, G. L. *et al.* A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry* **49**, 5912–5931 (2006).
36. Englebienne, P. & Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 4. Are Popular Scoring Functions Accurate for this Class of Proteins? *Journal of Chemical Information and Modeling* **49**, 1568–1580 (2009).
37. Lexa, K. W. & Carlson, H. A. Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics* **45**, 301–43 (2012).
38. Pantsar, T. & Poso, A. Binding Affinity via Docking: Fact and Fiction. *Molecules* **23**, (2018).
39. Jain, A. N. Effects of protein conformation in docking: improved pose prediction through protein pocket adaptation. *Journal of Computer-Aided Molecular Design* **23**, 355–74 (2009).

40. Subramanian, J., Sharma, S. & B-Rao, C. A Novel Computational Analysis of Ligand-Induced Conformational Changes in the ATP Binding Sites of Cyclin Dependent Kinases. *Journal of Medicinal Chemistry* **49**, 5434–5441 (2006).
41. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: A flux description of reaction mechanism. *Proceedings of the National Academy of Sciences* **106**, 13737–13741 (2009).
42. Kovermann, M., Grundström, C., Sauer-Eriksson, A. E., Sauer, U. H. & Wolf-Watz, M. Structural basis for ligand binding to an enzyme by a conformational selection pathway. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 6298 (2017).
43. Fersht, A. R., Knill-Jones, J. W., Bedouelle, H. & Winter, G. Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: a mobile loop envelopes the transition state in an induced-fit mechanism. *Biochemistry* **27**, 1581–1587 (1988).
44. Joseph, D., Petsko, G. & Karplus, M. Anatomy of a conformational change: hinged “lid” motion of the triosephosphate isomerase loop. *Science* **249**, 1425–1428 (1990).
45. Sullivan, S. M. & Holyoak, T. Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13829–34 (2008).
46. Weikl, T. R. & Paul, F. Conformational selection in protein binding and function. *Protein Science: A Publication of the Protein Society* **23**, 1508–18 (2014).
47. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences* **35**, 539–46 (2010).
48. Beglov, D. *et al.* Exploring the structural origins of cryptic sites on proteins. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E3416–E3425 (2018).
49. O’Bryan, J. P. Pharmacological targeting of RAS: Recent success with direct inhibitors. *Pharmacological research* **139**, 503–511 (2019).
50. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548–51 (2013).
51. Milburn, M. V. *et al.* Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939–45 (1990).

52. Comitani, F. & Gervasio, F. L. Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH. *Journal of Chemical Theory and Computation* **14**, 3321–3331 (2018).
53. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nature Reviews Drug Discovery* **10**, 307–317 (2011).
54. Roth, G. J., Stanford, N. & Majerus, P. W. Acetylation of prostaglandin synthase by aspirin. *Proceedings of the National Academy of Sciences of the United States of America* **72**, 3073 (1975).
55. Savi, P. *et al.* Identification and biological activity of the active metabolite of clopidogrel. *Journal of Thrombosis and Haemostasis* **84**, 891–6 (2000).
56. Pan, Z. *et al.* Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. *ChemMedChem* **2**, 58–61 (2007).
57. Miller, V. A. *et al.* Afatinib versus placebo for patients with advanced, metastatic non-small-cell lung cancer after failure of erlotinib, gefitinib, or both, and one or two lines of chemotherapy (LUX-Lung 1): a phase 2b/3 randomised trial. *The Lancet Oncology* **13**, 528–538 (2012).
58. Yver, A. Osimertinib (AZD9291)—a science-driven, collaborative approach to rapid drug design and development. *Annals of Oncology* **27**, 1165–1170 (2016).
59. Zhao, Z. & Bourne, P. E. Progress with covalent small-molecule kinase inhibitors. *Drug Discovery Today* **23**, 727–735 (2018).
60. Potashman, M. H. & Duggan, M. E. Covalent modifiers: an orthogonal approach to drug design. *Journal of Medicinal Chemistry* **52**, 1231–1246 (2009).
61. Evans, D. C., Watt, A. P., Nicoll-Griffith, D. A. & Baillie, T. A. Drug-Protein Adducts: An Industry Perspective on Minimizing the Potential for Drug Bioactivation in Drug Discovery and Development. **17**, 3–16 (2004).
62. Naisbitt, D. J., Gordon, S., Pirmohamed, M. & Park, B. Immunological Principles of Adverse Drug Reactions. *Drug Safety* **23**, 483–507 (2000).
63. Johnson, D. S., Weerapana, E. & Cravatt, B. F. Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Medicinal Chemistry* **2**, 949–964 (2010).
64. Padovan, E. T-cell response in penicillin allergy. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **28 Suppl 4**, 33–6 (1998).

65. Krishnan, A., Almén, M. S., Fredriksson, R. & Schiöth, H. B. The origin of GPCRs: identification of mammalian like Rhodopsin, Adhesion, Glutamate and Frizzled GPCRs in fungi. *PLOS ONE* **7**, e29817 (2012).
66. Taddese, B. *et al.* Do plants contain g protein-coupled receptors? *Plant Physiology* **164**, 287–307 (2014).
67. Lagerström, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Reviews Drug Discovery* **7**, 339–357 (2008).
68. Sugiyama, H., Ito, I. & Hirono, C. A new type of glutamate receptor linked to inositol phospholipid metabolism. *Nature* **325**, 531–533 (1987).
69. Conn, P. J. & Pin, J.-P. PHARMACOLOGY AND FUNCTIONS OF METABOTROPIC GLUTAMATE RECEPTORS. *Annual Review of Pharmacology and Toxicology* **37**, 205–237 (1997).
70. Dunlap, K. Functional and pharmacological differences between two types of GABA receptor on embryonic chick sensory neurons. *Neuroscience Letters* **47**, 265–270 (1984).
71. Andrade, R., Malenka, R. C. & Nicoll, R. A. A G protein couples serotonin and GABAB receptors to the same channels in hippocampus. *Science* **234**, 1261–5 (1986).
72. Goldstein, D. S., Eisenhofer, G. & McCarty, R. *Catecholamines : bridging basic science with clinical medicine*. 1084 (Academic Press, 1998).
73. Missale, C., Nash, S. R., Robinson, S. W., Jaber, M. & Caron, M. G. Dopamine Receptors: From Structure to Function. *Physiological Reviews* **78**, 189–225 (1998).
74. Barnes, N. M. & Sharp, T. A review of central 5-HT receptors and their function. *Neuropharmacology* **38**, 1083–1152 (1999).
75. Chen, J., Ishii, M., Wang, L., Ishii, K. & Coughlin, S. R. Thrombin receptor activation. Confirmation of the intramolecular tethered liganding hypothesis and discovery of an alternative intermolecular liganding mode. *The Journal of Biological Chemistry* **269**, 16041–5 (1994).
76. Rasmussen, S. G. F. *et al.* Crystal structure of the β 2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–55 (2011).
77. Gilman, A. G. *G Proteins and Dual Control of Adenylate Cyclase*. **36**, 577–579 (1984).
78. Tang, W. J. & Gilman, A. G. Type-specific regulation of adenylyl cyclase by G protein beta gamma subunits. *Science* **254**, 1500–3 (1991).

79. Berridge, M. J. Inositol trisphosphate and calcium signalling. *Nature* **361**, 315–25 (1993).
80. Wynne, B. M., Chiao, C.-W. & Webb, R. C. Vascular Smooth Muscle Cell Signaling Mechanisms for Contraction to Angiotensin II and Endothelin-1. *Journal of the American Society of Hypertension* **3**, 84–95 (2009).
81. Siehler, S. Regulation of RhoGEF proteins by G12/13-coupled receptors. *British Journal of Pharmacology* **158**, 41–9 (2009).
82. Druey, K. M., Blumer, K. J., Kang, V. H. & Kehrl, J. H. Inhibition of G-protein-mediated MAP kinase activation by a new mammalian gene family. *Nature* **379**, 742–746 (1996).
83. Tesmer, J. J. G., Berman, D. M., Gilman, A. G. & Sprang, S. R. Structure of RGS4 bound to ALF4-activated G(i alpha1): stabilization of the transition state for GTP hydrolysis. *Cell* **89**, 251–261 (1997).
84. Zeng, W. *et al.* The N-terminal domain of RGS4 confers receptor-selective inhibition of G protein signaling. *The Journal of Biological Chemistry* **273**, 34687–90 (1998).
85. Heximer, S. P., Watson, N., Linder, M. E., Blumer, K. J. & Hepler, J. R. RGS2/G0S8 is a selective inhibitor of Gqalpha function. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 14389–93 (1997).
86. Patil, D. N. *et al.* Structural organization of a major neuronal G protein regulator, the RGS7-Gβ5-R7BP complex. *eLife* **7**, (2018).
87. Hu, G. & Wensel, T. G. R9AP, a membrane anchor for the photoreceptor GTPase accelerating protein, RGS9-1. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 9755–60 (2002).
88. Drenan, R. M. *et al.* Palmitoylation regulates plasma membrane-nuclear shuttling of R7BP, a novel membrane anchor for the RGS7 family. *Journal of Cell Biology* **169**, 623–633 (2005).
89. Snow, B. E. *et al.* A G protein gamma subunit-like domain shared between RGS11 and other RGS proteins specifies binding to Gbeta5 subunits. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 13307–12 (1998).
90. Kovoov, A. *et al.* Co-expression of Gβ5 Enhances the Function of Two Gγ Subunit-like Domain-containing Regulators of G Protein Signaling Proteins. *Journal of Biological Chemistry* **275**, 3397–3402 (2000).
91. Sondek, J. & Siderovski, D. P. Gγ-like (ggl) domains: new frontiers in g-protein signaling and β-propeller scaffolding. *Biochemical Pharmacology* **61**, 1329–1337 (2001).

92. Siderovski, D. P., Diversé-Pierluissi, M. A. & De Vries, L. The GoLoco motif: a Gαi/o binding motif and potential guanine-nucleotide exchange factor. *Trends in Biochemical Sciences* **24**, 340–341 (1999).
93. Willard, M. D. *et al.* Selective role for RGS12 as a Ras/Raf/MEK scaffold in nerve growth factor-mediated differentiation. *The EMBO Journal* **26**, 2029–2040 (2007).
94. Kimple, R. J. *et al.* RGS12 and RGS14 GoLoco Motifs Are Gα_i Interaction Sites with Guanine Nucleotide Dissociation Inhibitor Activity. *Journal of Biological Chemistry* **276**, 29275–29281 (2001).
95. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *Journal of Cell Science* **114**, 3219–31 (2001).
96. Gold, S. J., Ni, Y. G., Dohlman, H. G. & Nestler, E. J. Regulators of G-protein signaling (RGS) proteins: region-specific expression of nine subtypes in rat brain. *The Journal of Neuroscience* **17**, 8024–8037 (1997).
97. Larminie, C. *et al.* Selective expression of regulators of G-protein signaling (RGS) in the human central nervous system. *Molecular Brain Research* **122**, 24–34 (2004).
98. Chen, Y. *et al.* Neurabin Scaffolding of Adenosine Receptor and RGS4 Regulates Anti-Seizure Effect of Endogenous Adenosine. *Journal of Neuroscience* **32**, 2683–2695 (2012).
99. Rorabaugh, B. R. *et al.* Regulators of G-protein signaling 2 and 4 differentially regulate cocaine-induced rewarding effects. *Physiology & Behavior* **195**, 9–19 (2018).
100. Talbot, J. N. *et al.* RGS inhibition at G(α)_{i2} selectively potentiates 5-HT_{1A}-mediated antidepressant effects. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 11086–91 (2010).
101. Wang, Q., Terauchi, A., Yee, C. H., Umemori, H. & Traynor, J. R. 5-HT_{1A} receptor-mediated phosphorylation of extracellular signal-regulated kinases (ERK1/2) is modulated by regulator of G protein signaling protein 19. *Cellular Signalling* **26**, 1846–1852 (2014).
102. Wang, Q. & Traynor, J. R. Modulation of μ-opioid receptor signaling by RGS19 in SH-SY5Y cells. *Molecular Pharmacology* **83**, 512–20 (2013).
103. Bodle, C. R., Mackie, D. I. & Roman, D. L. RGS17: an emerging therapeutic target for lung and prostate cancers. *Future Medicinal Chemistry* **5**, 995–1007 (2013).
104. You, M. *et al.* Fine Mapping of Chromosome 6q23-25 Region in Familial Lung Cancer Families Reveals RGS17 as a Likely Candidate Gene. *Clinical Cancer Research* **15**, 2666–2674 (2009).

105. James, M. A., Lu, Y., Liu, Y., Vikis, H. G. & You, M. RGS17, an Overexpressed Gene in Human Lung and Prostate Cancer, Induces Tumor Cell Proliferation Through the Cyclic AMP-PKA-CREB Pathway. *Cancer Research* **69**, 2108–2116 (2009).
106. DeLong, M. R. Primate models of movement disorders of basal ganglia origin. *Trends in Neurosciences* **13**, 281–285 (1990).
107. Morera-Herreras, T., Miguelez, C., Aristieta, A., Ruiz-Ortega, J. Á. & Ugedo, L. Endocannabinoid modulation of dopaminergic motor circuits. *Frontiers in Pharmacology* **3**, 110 (2012).
108. Giuffrida, A. *et al.* Dopamine activation of endogenous cannabinoid signaling in dorsal striatum. *Nature Neuroscience* **2**, 358–363 (1999).
109. Lerner, T. N. & Kreitzer, A. C. RGS4 is required for dopaminergic control of striatal LTD and susceptibility to parkinsonian motor deficits. *Neuron* **73**, 347–359 (2012).
110. Huang, J., Zhou, H., Mahavadi, S., Sriwai, W. & Murthy, K. S. Inhibition of $G\alpha_q$ -dependent PLC- $\beta 1$ activity by PKG and PKA is mediated by phosphorylation of RGS4 and GRK2. *American Journal of Physiology-Cell Physiology* **292**, C200–C208 (2007).
111. Picconi, B. *et al.* Loss of bidirectional striatal synaptic plasticity in L-DOPA-induced dyskinesia. *Nature Neuroscience* **6**, 501–506 (2003).
112. Shen, W. *et al.* M4 muscarinic receptor signaling ameliorates striatal plasticity deficits in models of L-DOPA-induced dyskinesia. *Neuron* **88**, 762–773 (2015).
113. Roman, D. L. *et al.* Identification of small-molecule inhibitors of RGS4 using a high-throughput flow cytometry protein interaction assay. *Molecular Pharmacology* **71**, 169–175 (2007).
114. Roman, D. L., Ota, S. & Neubig, R. R. Polyplexed Flow Cytometry Protein Interaction Assay: A Novel High-Throughput Screening Paradigm for RGS Protein Inhibitors. *Journal of Biomolecular Screening* **14**, 610–619 (2009).
115. Blazer, L. L., Zhang, H., Casey, E. M., Husbands, S. M. & Neubig, R. R. A nanomolar-potency small molecule inhibitor of regulator of G-protein signaling proteins. *Biochemistry* **50**, 3181–3192 (2011).
116. Turner, E. M., Blazer, L. L., Neubig, R. R. & Husbands, S. M. Small Molecule Inhibitors of Regulators of G Protein Signaling (RGS) Proteins. *ACS Medicinal Chemistry Letters* **3**, 146–150 (2011).
117. Blazer, L. L. *et al.* Selectivity and Anti-Parkinson's Potential of Thiadiazolidinone RGS4 Inhibitors. *ACS Chemical Neuroscience* **6**, 911–919 (2015).

118. Hayes, M. P., Bodle, C. R. & Roman, D. L. Evaluation of the Selectivity and Cysteine Dependence of Inhibitors across the Regulator of G Protein-Signaling Family. *Molecular Pharmacology* **93**, 25–35 (2018).
119. Whitty, A. & Kumaravel, G. Between a rock and a hard place? *Nature Chemical Biology* **2**, 112–118 (2006).
120. Jin, L., Wang, W. & Fang, G. Targeting Protein-Protein Interaction by Small Molecules. *Annual Review of Pharmacology and Toxicology* **54**, 435–456 (2014).
121. Oleinikovas, V., Saladino, G., Cossins, B. P. & Gervasio, F. L. Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *Journal of the American Chemical Society* **138**, 14257–14263 (2016).
122. Stank, A., Kokh, D. B., Fuller, J. C. & Wade, R. C. Protein Binding Pocket Dynamics. *Accounts of Chemical Research* **49**, 809–815 (2016).
123. Gunasekaran, K., Ma, B. & Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics* **57**, 433–443 (2004).
124. Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nature Reviews Drug Discovery* **1**, 198–210 (2002).
125. Mittal, A. & Johnson, M. E. Conformational diversity of bacterial FabH: Implications for molecular recognition specificity. *Journal of Molecular Graphics and Modelling* **55**, 115–122 (2015).
126. Tesmer, J. J. G. Structure and function of regulator of G protein signaling homology domains. *Progress in Molecular Biology and Translational Science* **86**, 75–113 (2009).
127. Neubig, R. R. & Siderovski, D. P. Regulators of G-Protein Signalling As New Central Nervous System Drug Targets. *Nature Reviews Drug Discovery* **1**, 187–197 (2002).
128. Taylor, V. G., Bommarito, P. A. & Tesmer, J. J. G. Structure of the Regulator of G Protein Signaling 8 (RGS8)-G α q Complex: MOLECULAR BASIS FOR G α SELECTIVITY. *Journal of Biological Chemistry* **291**, 5138–5145 (2016).
129. Vashisth, H., Storaska, A. J., Neubig, R. R. & Brooks 3rd, C. L. Conformational dynamics of a regulator of G-protein signaling protein reveals a mechanism of allosteric inhibition by a small molecule. *ACS Chemical Biology* **8**, 2778–2784 (2013).
130. Park, I.-H. *et al.* Estimation of hydrogen-exchange protection factors from MD simulation based on amide hydrogen bonding analysis. *Journal of Chemical Information and Modeling* **55**, 1914–1925 (2015).

131. Persson, F. & Halle, B. How amide hydrogens exchange in native proteins. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 10383–10388 (2015).
132. Lee, E., Linder, M. E. & Gilman, A. G. Expression of G-protein alpha subunits in *Escherichia coli*. *Methods in Enzymology* **237**, 146–164 (1994).
133. Blazer, L. L., Roman, D. L., Muxlow, M. R. & Neubig, R. R. Use of flow cytometric methods to quantify protein-protein interactions. *Current Protocols in Cytometry* 11–13 (2010). doi:[10.1002/0471142956.cy1311s51](https://doi.org/10.1002/0471142956.cy1311s51)
134. Chodavarapu, S., Jones, A. D., Feig, M. & Kaguni, J. M. DnaC traps DnaB as an open ring and remodels the domain that binds primase. *Nucleic Acids Research* **44**, 210–220 (2015).
135. Guttman, M., Weis, D. D., Engen, J. R. & Lee, K. K. Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. *Journal of the American Society for Mass Spectrometry* **24**, 1906–1912 (2013).
136. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).
137. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *Journal of Molecular Graphics* **14**, 33–38 (1996).
138. MacKerell Jr, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B* **102**, 3586–3616 (1998).
139. MacKerell Jr, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly Reviews of Biophysics* **40**, 191–285 (2004).
140. Towns, J. *et al.* XSEDE: accelerating scientific discovery. *Computing in Science and Engineering* **16**, 62–74 (2014).
141. Soundararajan, M. *et al.* Structural diversity in the RGS domain and its interaction with heterotrimeric G protein α -subunits. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6457–6462 (2008).
142. Alba, E. de, De Vries, L., Farquhar, M. G. & Tjandra, N. Solution structure of human GAIP (G α interacting protein): a regulator of G protein signaling. *Journal of Molecular Biology* **291**, 927–939 (1999).
143. Weis, D. D., Wales, T. E., Engen, J. R., Hotchko, M. & Ten Eyck, L. F. Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis.

Journal of the American Society for Mass Spectrometry **17**, 1498–1509 (2006).

144. Shaw, V. S., Mohammadiarani, H., Vashisth, H. & Neubig, R. R. Differential Protein Dynamics of Regulators of G-Protein Signaling: Role in Specificity of Small-Molecule Inhibitors. *Journal of the American Chemical Society* **140**, 3454–3460 (2018).

145. Mohammadiarani, H., Shaw, V. S., Neubig, R. R. & Vashisth, H. Interpreting Hydrogen–Deuterium Exchange Events in Proteins Using Atomistic Simulations: Case Studies on Regulators of G-Protein Signaling Proteins. *The Journal of Physical Chemistry B* **122**, 9314–9323 (2018).

146. Mohammadi, M., Mohammadiarani, H., Shaw, V. S., Neubig, R. R. & Vashisth, H. Interplay of cysteine exposure and global protein dynamics in small-molecule recognition by a regulator of G-protein signaling protein. *Proteins: Structure, Function, and Bioinformatics* (2018). doi:[10.1002/prot.25642](https://doi.org/10.1002/prot.25642)

147. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926–935 (1983).

148. Brown, D. K. *et al.* MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **33**, 2768–2771 (2017).

149. Roman, D. L. & Traynor, J. R. Regulators of G protein signaling (RGS) proteins as drug targets: modulating G-protein-coupled receptor (GPCR) signal transduction. *Journal of Medicinal Chemistry* **54**, 7433–40 (2011).

150. Blazer, L. L. *et al.* Reversible, allosteric small-molecule inhibitors of regulator of G protein signaling proteins. *Molecular Pharmacology* **78**, 524–33 (2010).

151. Storaska, A. J. *et al.* Reversible inhibitors of regulators of G-protein signaling identified in a high-throughput cell-based calcium signaling assay. *Cellular Signalling* **25**, 2848–55 (2013).

152. Monroy, C. A., Mackie, D. I. & Roman, D. L. A High Throughput Screen for RGS Proteins Using Steady State Monitoring of Free Phosphate Formation. *PLOS ONE* **8**, (2013).

153. Moy, F. J. *et al.* NMR structure of free RGS4 reveals an induced conformational change upon binding Galpha. *Biochemistry* **39**, 7063–73 (2000).

154. Storaska, A. J. & Neubig, R. R. NMR Methods for Detection of Small Molecule Binding to RGS4. *Methods in Enzymology* **522**, 133–152 (2013).

155. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based

on UNIX pipes. *Journal of Biomolecular NMR* **6**, 277–293 (1995).

156. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

157. Zhang, G. *et al.* Protein Quantitation Using Mass Spectrometry. *Methods in Molecular Biology* **673**, 211 (2010).

158. Britto, P. J., Knipling, L. & Wolff, J. The local electrostatic environment determines cysteine reactivity of tubulin. *The Journal of Biological Chemistry* **277**, 29018–27 (2002).

159. Jardine, I. Molecular weight analysis of proteins. *Methods in Enzymology* **193**, 441–55 (1990).

160. Ho, C. S. *et al.* Electrospray ionisation mass spectrometry: principles and clinical applications. *The Clinical Biochemist. Reviews* **24**, 3–12 (2003).

161. Meister, A. & Anderson, M. E. Glutathione. *Annual Review of Biochemistry* **52**, 711–760 (1983).

162. Holmgren, A. Thioredoxin and glutaredoxin systems. *The Journal of Biological Chemistry* **264**, 13963–6 (1989).

163. Yang, J., Chen, H., Vlahov, I. R., Cheng, J.-X. & Low, P. S. Evaluation of disulfide reduction during receptor-mediated endocytosis by using FRET imaging. *Proceedings of the National Academy of Sciences* **103**, 13872–13877 (2006).

164. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology* **5**, 789–96 (2009).

165. McCammon, J. A. Target flexibility in molecular recognition. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1754**, 221–224 (2005).

166. Cozzini, P. *et al.* Target flexibility: an emerging consideration in drug discovery and design. *Journal of Medicinal Chemistry* **51**, 6237–55 (2008).

167. Rahimova, R. *et al.* Identification of allosteric inhibitors of the ecto-5'-nucleotidase (CD73) targeting the dimer interface. *PLOS Computational Biology* **14**, e1005943 (2018).

168. Schmidtke, P., Bidon-Chanal, A., Luque, F. J. & Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **27**, 3276–3285 (2011).

169. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC bioinformatics* **10**, 168 (2009).
170. Liang, J., Woodward, C. & Edelsbrunner, H. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Science* **7**, 1884–1897 (1998).
171. Schmidtke, P. & Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *Journal of Medicinal Chemistry* **53**, 5858–5867 (2010).
172. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* **61**, 1–36 (2014).
173. Sheridan, R. P., Maiorov, V. N., Holloway, M. K., Cornell, W. D. & Gao, Y.-D. Drug-like Density: A Method of Quantifying the ‘Bindability’ of a Protein Target Based on a Very Large Set of Pockets and Drug-like Ligands from the Protein Data Bank. *Journal of Chemical Information and Modeling* **50**, 2029–2040 (2010).
174. Konermann, L., Vahidi, S. & Sowole, M. A. Mass Spectrometry Methods for Studying Structure and Dynamics of Biological Macromolecules. *Analytical Chemistry* **86**, 213–232 (2014).
175. Mandell, J. G., Baerga-Ortiz, A., Falick, A. M. & Komives, E. A. in *Protein-ligand interactions* **305**, 065–080 (Humana Press, 2005).
176. Mandell, J. G., Falick, A. M. & Komives, E. A. Identification of protein-protein interfaces by decreased amide proton solvent accessibility. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14705–10 (1998).
177. Lee, T. *et al.* Docking motif interactions in MAP kinases revealed by hydrogen exchange mass spectrometry. *Molecular Cell* **14**, 43–55 (2004).
178. Bonnington, L. *et al.* Application of Hydrogen/Deuterium Exchange-Mass Spectrometry to Biopharmaceutical Development Requirements: Improved Sensitivity to Detection of Conformational Changes. *Analytical Chemistry* **89**, 8233–8237 (2017).
179. Houde, D., Berkowitz, S. A. & Engen, J. R. The Utility of Hydrogen/Deuterium Exchange Mass Spectrometry in Biopharmaceutical Comparability Studies. *Journal of Pharmaceutical Sciences* **100**, 2071–2086 (2011).
180. Pirrone, G. F., Iacob, R. E. & Engen, J. R. Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Analytical Chemistry* **87**, 99–118 (2015).

181. Bobst, C. E. *et al.* Detection and Characterization of Altered Conformations of Protein Pharmaceuticals Using Complementary Mass Spectrometry-Based Approaches. *Analytical Chemistry* **80**, 7473–7481 (2008).
182. Huang, R. Y.-C. & Chen, G. Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry. *Analytical and Bioanalytical Chemistry* **406**, 6541–6558 (2014).
183. Barksdale, A. D. & Rosenberg, A. Acquisition and interpretation of hydrogen exchange data from peptides, polymers, and proteins. *Methods of Biochemical Analysis* **28**, 1–113 (1982).
184. Abaturov, L., Jinoria, K., Varshavsky, Y. & Yakobashvily, N. Effect of ligand and heme on conformational stability (intramolecular conformational motility) of hemoglobin as revealed by hydrogen exchange. *FEBS Letters* **77**, 103–106 (1977).
185. Hvidt, A. & Linderstrøm-Lang, K. The kinetics of the deuterium exchange of insulin with D₂O. An amendment. *Biochimica et Biophysica Acta* **16**, 168–169 (1955).
186. *Deuterium Exchange of Poly-DL-alanine in Aqueous Solution.* **69**, 106–118 (1957).
187. Bryan, W. P. & Nielsen, S. O. Hydrogen-deuterium exchange of poly-dl-alanine in aqueous solution. *Biochimica et Biophysica Acta* **42**, 552–553 (1960).
188. Molday, R. S., Englander, S. W. & Kallen, R. G. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150–158 (1972).
189. Woodward, C. K. & Hilton, B. D. Hydrogen Exchange Kinetics and Internal Motions in Proteins and Nucleic Acids. *Annual Review of Biophysics and Bioengineering* **8**, 99–127 (1979).
190. Englander, S. W., Downer, N. W. & Teitelbaum, H. Hydrogen Exchange. *Annual Review of Biochemistry* **41**, 903–924 (1972).
191. Hvidt, A. & Nielsen, S. O. Hydrogen Exchange in Proteins. *Advances in Protein Chemistry* **21**, 287–386 (1966).
192. Takahashi, T., Nakanishi, M. & Tsuboi, M. Hydrogen–deuterium exchange study of amino acids and proteins by 200- to 230-nm spectroscopy. *Analytical Biochemistry* **110**, 242–9 (1981).
193. Hilton, B. D. & Woodward, C. K. On the mechanism of isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biochemistry* **18**, 5834–41 (1979).

194. Ellis, L. M., Bloomfield, V. A. & Woodward, C. K. Hydrogen-tritium exchange kinetics of soybean trypsin inhibitor (Kunitz). Solvent accessibility in the folded conformation. *Biochemistry* **14**, 3413–9 (1975).
195. Woodward, C. K., Ellis, L. M. & Rosenberg, A. Solvent accessibility in folded proteins. Studies of hydrogen exchange in trypsin. *The Journal of Biological Chemistry* **250**, 432–9 (1975).
196. Woodward, C. K. & Rosenberg, A. Studies of hydrogen exchange in proteins. VI. Urea effects on ribonuclease exchange kinetics leading to a general model for hydrogen exchange from folded proteins. *The Journal of Biological Chemistry* **246**, 4114–21 (1971).
197. Rosenberg, A. & Chakravarti, K. Studies of hydrogen exchange in proteins. I. The exchange kinetics of bovine carbonic anhydrase. *The Journal of Biological Chemistry* **243**, 5193–201 (1968).
198. Levitt, M. HYDROGEN BOND AND INTERNAL SOLVENT DYNAMICS OF BPTI PROTEIN. *Annals of the New York Academy of Sciences* **367**, 162–181 (1981).
199. Eden, D., Matthew, J. B., Rosa, J. J. & Richards, F. M. Increase in apparent compressibility of cytochrome c upon oxidation. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 815–9 (1982).
200. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins: Structure, Function, and Genetics* **17**, 75–86 (1993).
201. Adhikary, S. *et al.* Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer. *Proceedings of the National Academy of Sciences* **114**, E1786–E1795 (2017).
202. Brock, A. Fragmentation hydrogen exchange mass spectrometry: A review of methodology and applications. *Protein Expression and Purification* **84**, 19–37 (2012).
203. Zhang, Q. *et al.* Epitope Mapping of a 95 kDa Antigen in Complex with Antibody by Solution-Phase Amide Backbone Hydrogen/Deuterium Exchange Monitored by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Analytical Chemistry* **83**, 7129–7136 (2011).
204. Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. Mechanisms and uses of hydrogen exchange. *Current Opinion in Structural Biology* **6**, 18–23 (1996).
205. Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. Protein hydrogen exchange mechanism: Local fluctuations. *Protein Science: A Publication of the Protein Society* **12**, 153 (2003).
206. Wytenbach, T. & Bowers, M. T. Gas phase conformations of biological molecules: the hydrogen/deuterium exchange mechanism. *Journal of the American Society for Mass Spectrometry* **10**, 9–14 (1999).

207. Miller, D. W. & Dill, K. A. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein Science: a Publication of the Protein Society* **4**, 1860–73 (1995).
208. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Rare Fluctuations of Native Proteins Sampled by Equilibrium Hydrogen Exchange. *Journal of the American Chemical Society* **125**, 15686–15687 (2003).
209. Best, R. B. & Vendruscolo, M. Structural Interpretation of Hydrogen Exchange Protection Factors in Proteins: Characterization of the Native State Fluctuations of CI2. *Structure* **14**, 97–106 (2006).
210. Kieseritzky, G., Morra, G. & Knapp, E.-W. Stability and fluctuations of amide hydrogen bonds in a bacterial cytochrome c: a molecular dynamics study. *Journal of Biological Inorganic Chemistry* **11**, 26–40 (2006).
211. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: Testing current models. *Protein Science* **21**, 987–995 (2012).
212. Ma, B. & Nussinov, R. Polymorphic triple beta-sheet structures contribute to amide hydrogen/deuterium (H/D) exchange protection in the Alzheimer amyloid beta42 peptide. *The Journal of Biological Chemistry* **286**, 34244–53 (2011).
213. Bentley, G. A. *et al.* Exchange of individual hydrogens for a protein in a crystal and in solution. *Journal of Molecular Biology* **170**, 243–7 (1983).
214. Clarke, J., Itzhaki, L. S. & Fersht, A. R. Hydrogen exchange at equilibrium: a short cut for analysing protein-folding pathways? *Trends in Biochemical Sciences* **22**, 284–7 (1997).
215. Liu, T. *et al.* Quantitative Assessment of Protein Structural Models by Comparison of H/D Exchange MS Data with Exchange Behavior Accurately Predicted by DXCOREX. *Journal of The American Society for Mass Spectrometry* **23**, 43–56 (2012).
216. Kuwata, K. *et al.* NMR-detected hydrogen exchange and molecular dynamics simulations provide structural insight into fibril formation of prion protein fragment 106-126. *Proceedings of the National Academy of Sciences* **100**, 14790–14795 (2003).
217. Craig, P. O. *et al.* Prediction of Native-State Hydrogen Exchange from Perfectly Funneled Energy Landscapes. *Journal of the American Chemical Society* **133**, 17463–17472 (2011).
218. Petruk, A. A. *et al.* Molecular Dynamics Simulations Provide Atomistic Insight into Hydrogen Exchange Mass Spectrometry Experiments. *Journal of Chemical Theory and Computation* **9**, 658–669 (2013).
219. Hsu, Y.-H. *et al.* Fluoroketone Inhibition of Ca²⁺ -Independent Phospholipase A2

through Binding Pocket Association Defined by Hydrogen/Deuterium Exchange and Molecular Dynamics. *Journal of the American Chemical Society* **135**, 1330–1337 (2013).

220. Radou, G., Dreyer, F. N., Tuma, R. & Paci, E. Functional Dynamics of Hexameric Helicase Probed by Hydrogen Exchange and Simulation. *Biophysical Journal* **107**, 983–990 (2014).

221. McAllister, R. G. & Konermann, L. Challenges in the Interpretation of Protein H/D Exchange Data: A Molecular Dynamics Simulation Perspective. *Biochemistry* **54**, 2683–2692 (2015).

222. McAllister, R. G. & Lars Konermann, S. From Solution Into the Gas Phase: Studying Protein Hydrogen Exchange and Electrospray Ionization Using Molecular Dynamics Simulation. (The University of Western Ontario, 2015).

223. Mao, Y. *et al.* Hydrogen/Deuterium Exchange and Molecular Dynamics Analysis of Amyloid Fibrils Formed by a D69K Charge-Pair Mutant of Human Apolipoprotein C-II. *Biochemistry* **54**, 4805–4814 (2015).

224. Khakinejad, M., Kondalaji, S. G., Donohoe, G. C. & Valentine, S. J. Ion Mobility Spectrometry-Hydrogen Deuterium Exchange Mass Spectrometry of Anions: Part 2. Assessing Charge Site Location and Isotope Scrambling. *Journal of the American Society for Mass Spectrometry* **27**, 451–61 (2016).

225. Resing, K. A., Hoofnagle, A. N. & Ahn, N. G. Modeling deuterium exchange behavior of ERK2 using pepsin mapping to probe secondary structure. *Journal of the American Society for Mass Spectrometry* **10**, 685–702 (1999).

226. Hilser, V. J. & Freire, E. Structure-based Calculation of the Equilibrium Folding Pathway of Proteins. Correlation with Hydrogen Exchange Protection Factors. *Journal of Molecular Biology* **262**, 756–772 (1996).

227. García, A. E. & Hummer, G. Conformational dynamics of cytochrome c: correlation to hydrogen exchange. *Proteins* **36**, 175–91 (1999).

228. Shaw, D. E. *et al.* Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **330**, 341–346 (2010).

229. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: a structure for efficient numerical computation. *Computing in Science and Engineering* **13**, 22–30 (2011).

230. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation* **11**, 3696–3713 (2015).

231. Milne, J. S., Mayne, L., Roder, H., Wand, A. J. & Englander, S. W. Determinants of protein hydrogen exchange studied in equine cytochrome c. *Protein Science: A Publication of the*

Protein Society **7**, 739–45 (1998).

232. Yung-Chi, C. & Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochemical Pharmacology* **22**, 3099–3108 (1973).