Conformational Dynamics of a Regulator of G-protein Signaling Protein Reveals a Mechanism of Allosteric Inhibition by a Small Molecule

Harish Vashisth,[†] Andrew J. Storaska,[‡,‖] Richard R. Neubig,[¶] and Charles L. Brooks III[*,§]

[†]*Department of Chemical Engineering, University of New Hampshire, Durham, NH, USA*
[‡]*Department of Pharmacology, University of Michigan, Ann Arbor, MI, USA*
[¶]*Department of Pharmacology and Toxicology, Michigan State University, East Lansing, MI, USA*
[§]*Department of Chemistry and Biophysics Program, University of Michigan, Ann Arbor, MI, USA*

[‖] co-first author
[*] corresponding author; Email: brookscl@umich.edu

SUPPLEMENTAL METHODS

Molecular Dynamics (MD) Simulations: System Setup

All MD trajectories were generated using NAMDv2.8[1] and the CHARMM force-field with CMAP correction.[2,3] VMDv1.9[4] was used for system creation and protein rendering. Details of all simulations appear in Table S1. Initial coordinates from two different structures of RGS4 with Protein Data Bank (PDB) codes 1AGR (crystal structure) and 1EZT (NMR structure) were used. All systems were solvated (Fig. S1a, S1b, and S1c) using explicit (TIP3P) water and all hydrogen atoms were included. Charge neutrality was maintained by adding counterions. Similar protocols were followed for all equilibration simulations: 500-1000 cycles of energy-minimization via conjugate-gradient optimization and a constant temperature (310K) via the Langevin thermostat with damping coefficient of 5 ps^{-1}. The equilibration phases were carried out initially in the NPT ensemble to adjust the box volume, and thereafter continued in the NVT ensemble using a time-step of 2-fs with rigid bonds in all simulations. Periodic boundary conditions were used throughout. Non-bonded interactions were cut-off beyond 10 Å with smooth switching taking effect at 8.0 Å. Long-range electrostatic interactions were handled using the particle mesh Ewald (PME) method. The equilibration runs for wild-type RGS4 were used to sample initial conditions for all temperature accelerated molecular dynamics (TAMD) trajectories. Details of TAMD simulations are described in the following. The inhibitor molecule (CCG-50014; see Fig. S1e) was parameterized using MATCH.[5]

Temperature Accelerated Molecular Dynamics

Temperature-accelerated molecular dynamics (TAMD) is an enhanced sampling approach to explore the physical free-energy landscape in a pre-defined set of collective variables (CVs).[6–8] In this work, we have used a conformational sampling algorithm for proteins based upon TAMD,[9] which has been successfully applied to several systems recently.[10–13] The coupled system of equations describing TAMD are as follows:

$$\begin{aligned} m_i\ddot{x}_i &= -\frac{\partial V(x)}{\partial x_i} - \kappa\sum_{j=1}^{m}[\theta_j^*(x) - \theta_j]\frac{\partial\theta_j^*(x)}{\partial x_i} - \gamma m_i\dot{x}_i + \eta_i(t;\beta) \\ \bar{\gamma}\bar{m}_j\dot{\theta}_j &= \kappa[\theta_j^*(x) - \theta_j] + \xi_j(t;\bar{\beta}) \end{aligned} \tag{S1}$$

where $\theta^*(x)$ = ($\theta_1^*(x)$,$\theta_2^*(x)$,.....,$\theta_m^*(x)$) are collective variables that are functions of atom Cartesian coordinates, m_i and $\bar{m}_j$ are the masses of x_i and θ_j, *V(x)* is the interatomic MD potential, κ is the "coupling spring-constant", γ is the Langevin friction coefficient, η is the white noise satisfying fluctuation-dissipation theorem at physical temperature β^{-1}, $\bar{\gamma}$ and ξ respectively are fictitious friction and thermal noise at artificial temperature $\bar{\beta}^{-1}$.

The aforementioned set of equations describe the motion of *x(t)* and *θ(t)* over the extended potential

$$U_\kappa(x,\theta) = V(x) + \frac{\kappa}{2}\sum_{j=1}^{m}[\theta_j^*(x) - \theta_j]^2 \tag{S2}$$

As shown earlier by Maragliano and Vanden-Eijnden,[6] by choosing κ such that $\theta^*(x(t)) \approx \theta(t)$ and fictitious friction coefficient $\bar{\gamma}$ such that θ moves slower than *x*, one can generate a trajectory at artificial temperature $\bar{\beta}^{-1}$ subject to the free energy computed at the physical temperature β^{-1}. To ensure these conditions, we have chosen a TAMD friction $\bar{\gamma}$ of 500 ps^{-1} and a spring constant κ of 100 kcal/mol·$Å^2$. In this work, we choose the Cartesian coordinates of centers-of-mass of spatially contiguous groups of residues as CVs. The entire RGS4 structure (residues 51-178) was divided into 6 subdomains (18 CVs). Residue memberships for mutually exclusive and collectively exhaustive subdomains of RGS4 are listed in Table S2 and the location of each subdomain is shown in Fig. S1d. A total of four independent TAMD trajectories were carried out (Table S1).

Theoretical Chemical Shift Perturbation (CSP) Analysis

We used software packages SHIFTX2[14] and SPARTA+[15] for theoretical prediction of chemical shifts for ^{15}N and ^{1}H atoms based upon our MD simulations. Specifically, starting with two different initial structures (PDB codes 1AGR and 1EZT), we carried out at least three independent simulations of mutant RGS4 with and without inhibitor CCG-50014 (runs#2, 7, 11 and 13 in Table S1). The predicted values of chemical shifts were further time-averaged over each trajectory and ensemble-averaged over independent trajectories. We used following equation to predict chemical shift perturbation (CSP) for each residue

$$\mathrm{CSP} = \sqrt{\left(\frac{\Delta\delta^{15}\mathrm{N}}{5}\right)^2 + (\Delta\delta^{1}\mathrm{H})^2} \tag{S3}$$

The predicted CSP for each residue (in ppm) is shown in Fig. S13 and Fig. S14, while key residues are summarized in Table S3. These results are also discussed in the main article.

Protein Expression and Purification

RGS4 was expressed and purified for NMR studies as previously described.[16] Briefly, N-terminally truncated (Δ51) RGS4 was expressed with an N-terminal maltose binding protein (MBP) fusion containing a 10x histidine tag and tobacco etch virus protease (TEV) site using a pMalC2H10 vector. Two RGS4 mutants were prepared with either all native Cysteines mutated to alanine (cysless RGS4), or all native Cysteines except Cysteine 95 mutated to alanine (Cys95 RGS4), using methods employed previously.[17] Purified ^{15}N-labeled Δ51 RGS4 (cysless and Cys95) was prepared by removing the MBP tag via addition of TEV protease and purified by ion exchange chromatography. The samples were buffer exchanged into 50 mM sodium phosphate buffer with 50 mM NaCl at pH 6.0 by dialyzing overnight at 4°C. The concentration was determined using a NanoDrop spectrophotometer.

NMR Spectroscopy

NMR samples were prepared with 0.1 mM ^{15}N-labeled RGS4 in the same buffer with the addition of 7% (v/v) D_2O and either DMSO (2% v/v) or 0.1 mM CCG-50014. The heteronuclear single-quantum coherence (HSQC) NMR data were collected at 25°C with a 600 MHz Bruker Avance III magnet with cryogenically cooled sample probe. Data were processed with Bruker TopSpin 1.3 software and analyzed with computer aided resonance assignment (CARA) software. The ^{15}N chemical shifts were referenced from the previous solution structure data of RGS4.[18] ^{1}H-^{15}N HSQC spectra of cysless RGS4 and Cys95 RGS4 are shown in Fig. S11, while normalized intensity ratios of the HSQC peaks after and before addition of inhibitor CCG-50014 are shown in Fig. 6. We note that chemical shift changes for each residue in the HSQC spectra have not been identified as the irreversible nature of the compound (CCG-50014) prevents titrating peak shift trajectories as a function of molar ratio of compound to protein. Instead, peak perturbations are analyzed via a loss in intensity (Fig. 6) resulting from the peaks becoming split between two (or possibly more) chemical states. We also note that the HSQC spectra of cysless RGS4 and Cys95 RGS4 show similar signal dispersion compared to wild-type RGS4 indicating that the overall fold is preserved. Furthermore, peak coordinates in the HSQC spectra are similar to wild-type RGS4 for more than 80% of the residues in both mutant proteins. Peak intensities in each spectrum were normalized to residue 201 in order to compare the intensity change before and after addition of CCG-50014. Peak intensities were compared either as a ratio, or as % peak attenuation (100-((I_{50014}/I_0)*100)). A number of peaks were removed from the peak intensity analysis due to ambiguity resulting from overlap in the HSQC spectrum: 58, 73, 78, 79, 95, 96, 97, 107, 114, 117, 119, 132, 133, 135, 136, 152, 156, 161, 163, 183, 186, 188, 189, 192, 203, 204. Nearly all of these peaks are found to be overlapped in the WT RGS4 spectrum, published previously. The results from NMR-HSQC experiments are discussed in the main article.

Principal Component Analysis of the NMR ensemble (PDB code 1EZY)

We carried out principal component analysis (PCA) of the NMR ensemble (PDB code 1EZY) to better understand correspondence of principal modes with elastic network model based normal modes. Such analysis has been quite useful in other proteins as well.[19,20] These data are presented in Table S4. By matching a single principal component (PC) with a single low-frequency mode via computation of overlaps, we find that PC 1 is significantly correlated to low-frequency Mode 2. This suggests that principal motions present in RGS4 can be explained by a single low-frequency mode of RGS4. This also provides a structure-based explanation of underlying dynamics.

SUPPLEMENTAL REFERENCES

1. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
2. MacKerell, Jr., A. D. et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
3. MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
4. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
5. Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III, *J. Comput. Chem.* **2012**, *33* , 189–202.
6. Maragliano, L.; Vanden-Eijnden, E. *Chem. Phys. Lett.* **2006**, *426*, 168–175.
7. Maragliano, L.; Vanden-Eijnden, E. *J. Chem. Phys.* **2008**, *128*, 184110.
8. Vanden-Eijnden, E. *J. Comput. Chem.* **2009**, *30*, 1737–1747.
9. Abrams, C. F.; Vanden-Eijnden, E. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4961–4966.
10. Vashisth, H.; Maragliano, L.; Abrams, C. F. *Biophys. J.* **2012**, *102*, 1979–1987.
11. Vashisth, H.; Skiniotis, G.; Brooks, C. L., III, *Structure* **2012**, *20*, 1453–1462.
12. Vashisth, H.; Brooks, C. L., III, *J. Phys. Chem. Lett.* **2012**, *3* , 3379–3384.
13. Vashisth, H.; Abrams, C. F. *Proteins* **2013**, *81*, 1017–1030.
14. Han, B.; Liu, Y.; Ginzinger, S.; Wishart, D. *J. Biomol. NMR* **2011**, *50*, 43–57.
15. Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13–22.
16. Storaska, A. J.; Neubig, R. R. *Methods Enzymol.* **2012**, *522*, 133–152.
17. Roman, D. L.; Blazer, L. L.; Monroy, C. A.; Neubig, R. R. *Mol. Pharmacol.* **2010**, *78*, 360–365.
18. Moy, F. J.; Chanda, P. K.; Cockett, M. I.; Edris, W.; Jones, P. G.; Powers, R. *J. Biomol. NMR* **1999**, *15*, 339–340.
19. Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. *Structure* **2008** , *16* , 321–330.
20. Yang, L.; Eyal, E.; Bahar, I.; Kitao, A. *Bioinformatics* **2009** , *25* , 606–614.

TABLE S1. Details of all MD and TAMD simulations.

Run No.	System	Initial coordinates	Run type	Run length (ns)	Ligand (CCG-50014)	Construct type	System size (atoms)
1	RGS4	1AGR	MD	20	×	wild-type	28076
2	RGS4	1AGR	MD[c]	20	×	mutant[a]	28073
3	RGS4	1AGR	TAMD[b]	35	×	wild-type	28076
4	RGS4	1AGR	TAMD	35	×	wild-type	28076
5	RGS4	1AGR	TAMD	25	×	wild-type	28076
6	RGS4	TAMD (run#4)	MD	40	×	mutant	31585
7	RGS4	TAMD (run#4)	MD[c]	40	√	mutant	31529
8	RGS4-G_α	1AGR	MD	20	×	wild-type	67368
9	RGS4-G_α	1AGR[d]	MD	45	√	wild-type	69259
10	RGS4	1EZT	MD	20	×	wild-type	29269
11	RGS4	1EZT	MD[c]	20	×	mutant	29269
12	RGS4	1EZT	TAMD	50	×	wild-type	29269
13	RGS4	TAMD (run#11)	MD[c]	40	√	mutant	39436

[a] all cysteines but Cys95 were mutated to Ala in all mutant simulations
[b] $\bar{\beta}^{-1}$ = 2 kcal/mol; all other TAMD runs were carried out at $\bar{\beta}^{-1}$ = 3 kcal/mol
[c] three independent simulations were carried out in each case except for run#13 where four independent simulations were carried out
[d] initial coordinates of RGS4 were taken from TAMD (run#4), and of G_α were taken from the PDB coordinate file 1AGR; restraints on all C_α-atoms were enforced for the first 20 ns to maintain RGS4-G_α interaction which was followed by a 25-ns long unrestrained MD equilibration

TABLE S2. Subdomain Memberships in RGS4. Subdomain partitions are shown in Fig. S1d

Subdomain	Mass (kDa)	Residues
1	1.56	103-114, 143-145
2	1.63	115-128
3	1.63	129-142
4	3.32	51-74, 174-178
5	3.43	88-102, 146-158
6	3.49	75-87, 159-173

TABLE S3. Summary of perturbed residues on CCG-50014 binding to RGS4. Each residue with chemical shift perturbation (CSP) larger than mean in Figures 6, S13, and S14 is listed. All residues with CSPs larger than 1 standard deviation (SD) are marked with asterisk. Listed in Columns 1 and 2 are residues predicted by software packages SHIFTX2 and SPARTA+, respectively, based upon MD simulations, while Column 3 lists residues based upon ^{1}H-^{15}N HSQC experiments. In Column 3, each underlined residue has been predicted by at least one software to have a CSP larger than mean in MD simulations.

SHIFTX2				SPARTA+				Measured			
1AGR		1EZT		1AGR		1EZT					
59	145*	52	148*	59	128	67*	134	52	92	139	176
66	146*	60	149*	60	130*	87	137	53	93	141	177
67	147	67*	150	62	139	91	141	61*	94	142*	178*
71	148*	68	152	66	141	92	142	62	100	143	
72	149*	87	153*	92	142	95	145*	63*	101*	145	
87	150	91	156	93	143*	96	146	64*	103	146	
91	151	95	162	95	145*	98	147*	65	108	147	
95	152	98	173	97	146*	99	148*	66*	110*	148	
98*	153*	99*		98*	147*	101	149*	68	111	149	
99*	157	118*		99*	148	117	150	69	112	150	
105	169	121		102	149*	118*	151*	70	113	151	
108	170	122*		103	150*	119	159	72	115	153*	
111	172*	123		105	151*	120*	160	74	118	157	
115	173*	124		107	152	121*		76	120	158	
118*	174*	126*		117	153	122*		77*	123	159	
121*	177	127		118*	169	123*		81	124*	160	
122	178	130		119*	172	124		82	125	165	
123		134		120	173	125		83	126*	167*	
124*		140		121*	174	126		86	127	168	
126*		141		122	175	127		87	128	169*	
128		143		123*	176	129		88	131	172	
130		145*		124*	177*	130		89	134	174	
141		147*		126		131		91*	137	175	

TABLE S4. Overlaps between the first three principal components (PCs) and the first three low-frequency normal modes.

	Mode 1	Mode 2	Mode 3
PC 1	0.01	**0.41**	0.23
PC 2	0.12	0.20	0.03
PC 3	0.10	0.17	0.15

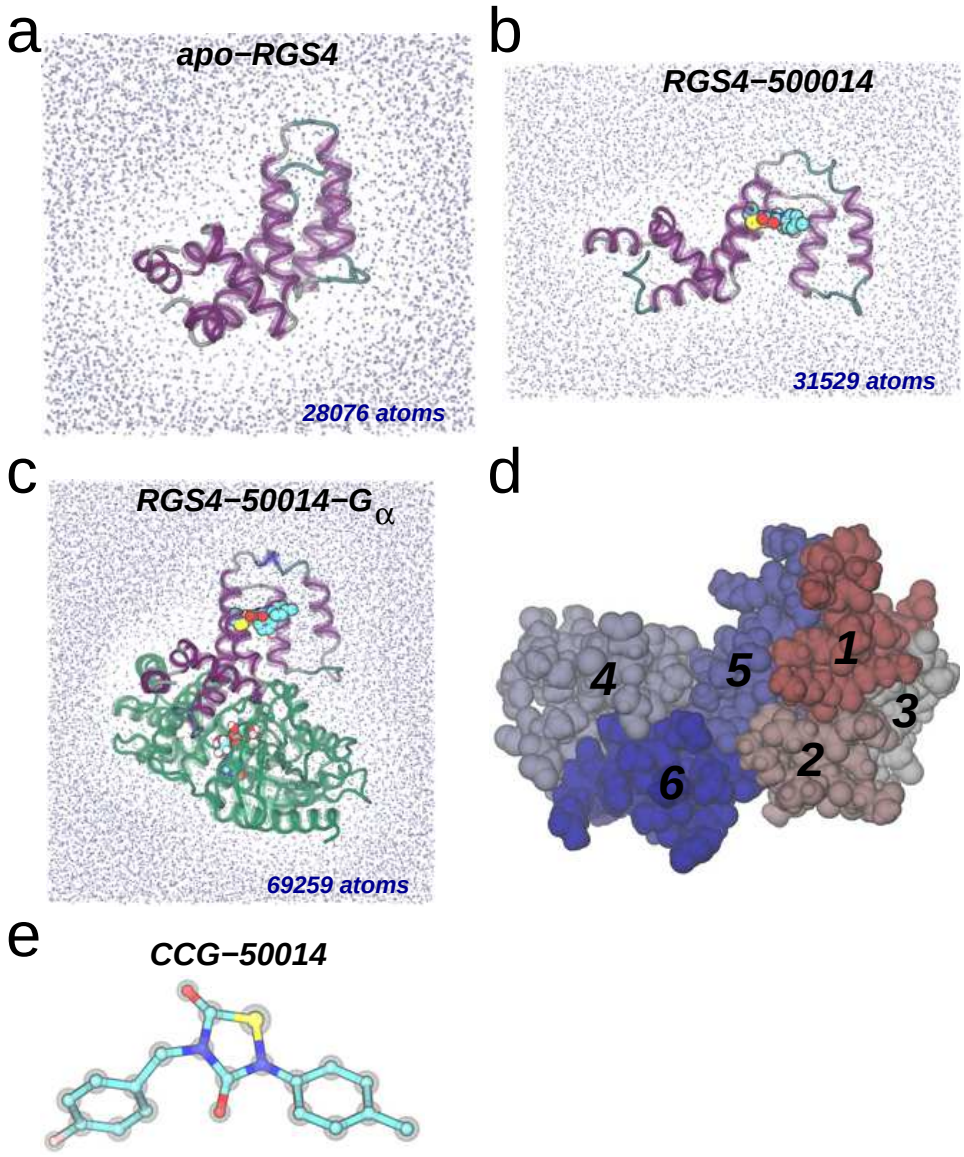

FIGURE S1. (a, b, and c) Schematic representations of typical simulation-domains (runs#1, 7, and 9 in Table S1). Protein is shown in cartoon representation with solvent and ions as spheres. A single molecule of CCG-50014 bound to Cys95 is highlighted in space filling representations in panels b and c. (d) Subdomain partitions of RGS4 are shown for TAMD simulations. Each subdomain is colored and labeled with a total of 6 subdomains (18 CVs). Residue memberships for each subdomain are listed in Table S2. (e) Snapshot of an unconjugated CCG-50014 molecule.

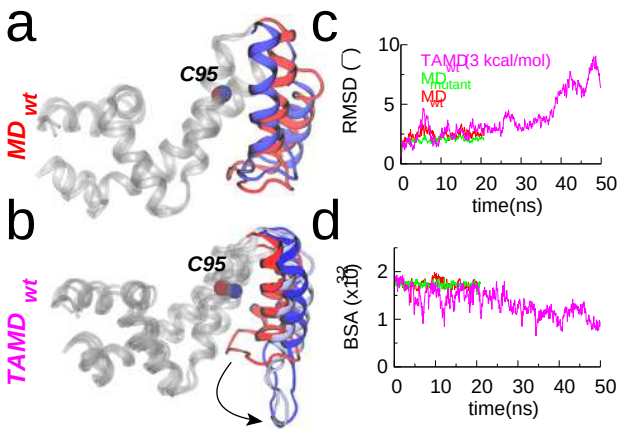

FIGURE S2. Data for MD and TAMD simulations starting with initial coordinates of RGS4 taken from PDB code 1EZT (runs#10-12 in Table S1). See Fig. 2 in the main article for coloring scheme and caption details.

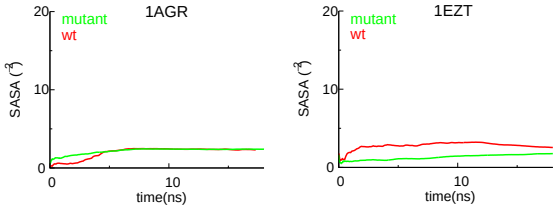

FIGURE S3. Solvent accessible surface area (SASA) for residue Cys95 from MD equilibration trajectories of wild-type and mutant-RGS4 (runs#1, 2, 10 and 11 in Table S1). Data for MD simulations starting with two different initial coordinates (PDB codes 1AGR and 1EZT) are shown. The SASA traces for mutant-RGS4 runs are averages over three independent MD runs.

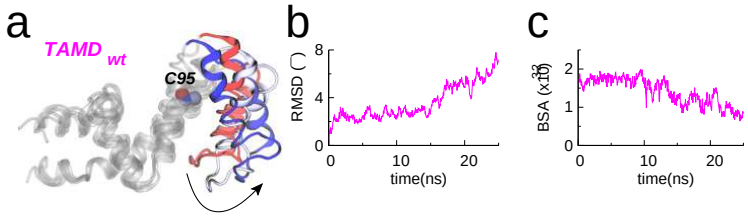

FIGURE S4. Data for an additional TAMD simulation starting with initial coordinates taken from PDB code 1AGR (run#5 in Table S1). See Fig. 2 in the main article for coloring scheme as well for data from TAMD runs#3 and 4 in Table S1.

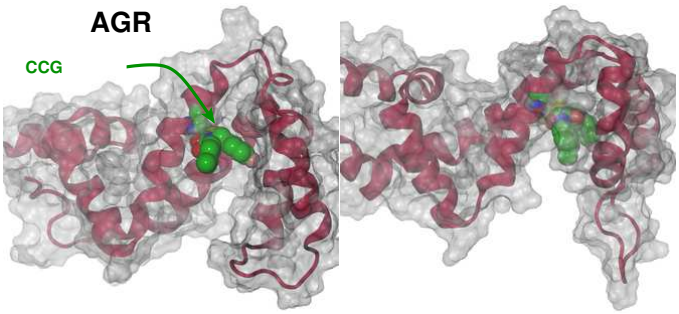

FIGURE S5. CCG-50014-docked and energy-minimized initial conformations of RGS4.

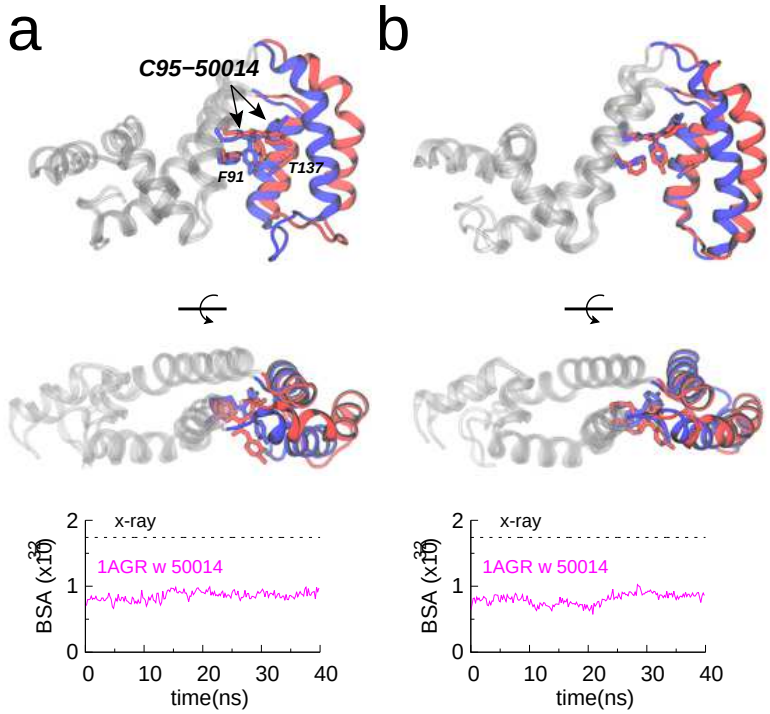

FIGURE S6. (top panels) Snapshots from the beginning (red) and end (blue) of two additional independent 40-ns long CCG-50014 bound MD simulations (run#7 in Table S1). (bottom panels) The trace of same buried surface area (BSA) as shown in Fig. 2d and Fig. 4d in the main article.

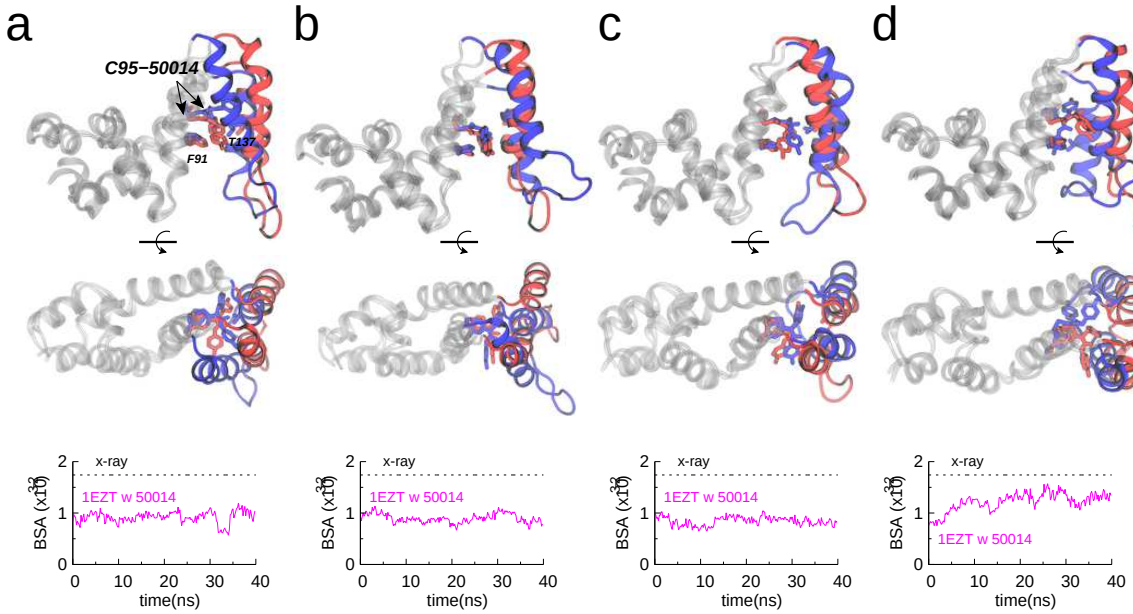

FIGURE S7. (top panels) Snapshots from the beginning (red) and end (blue) of four independent 40-ns long CCG-50014 bound MD simulations (run#13 in Table S1). (bottom panels) The trace of same buried surface area (BSA) as shown in Fig. 2d and Fig. 4d in the main article.

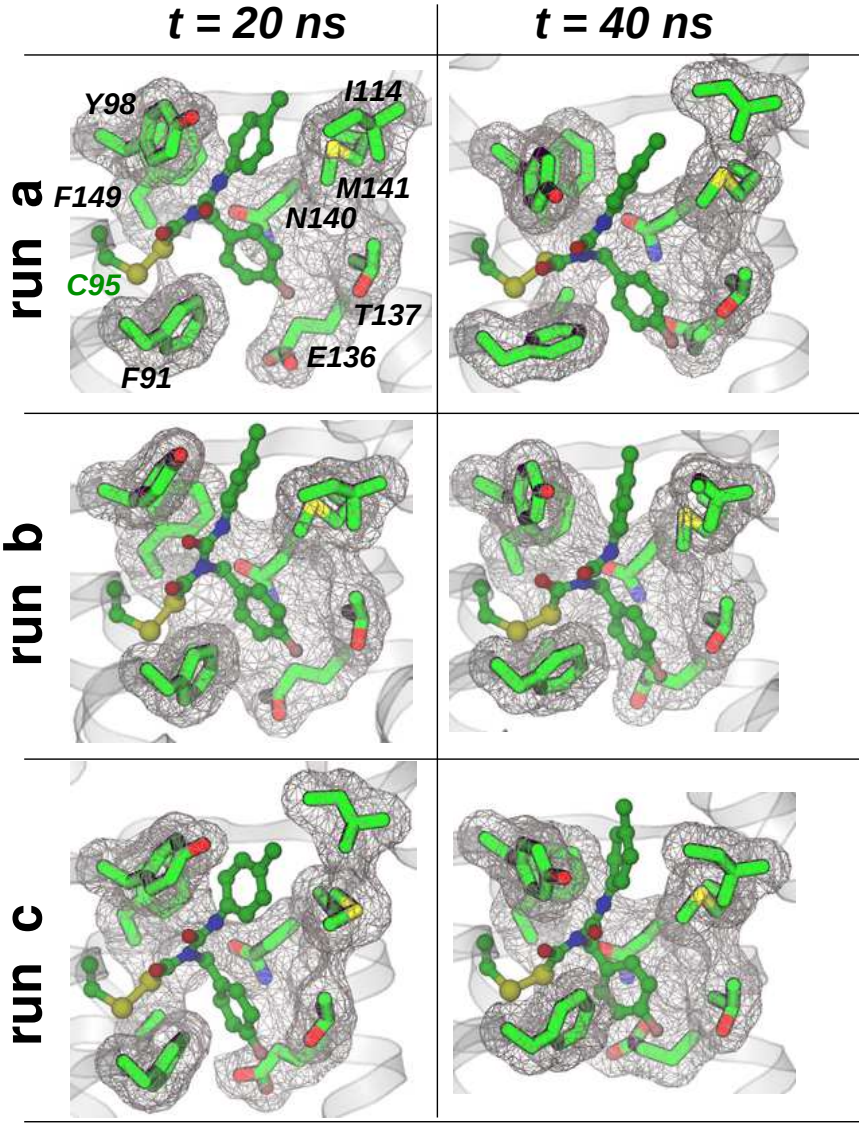

FIGURE S8. Residues interacting with the inhibitor molecule (CCG-50014). At two different time-points during MD equilibration, highlighted are the positions of the side-chains of residues directly in contact with CCG-50014. Covalently linked (to Cys95) CCG-50014 is shown in small spheres, and the labeled side-chains of RGS4 residues are rendered as transparent green sticks. The surface corresponding to each RGS4 residue is also rendered as a gray mesh. The snapshots are based upon three independent 40-ns long MD simulations (run#7 in Table S1).

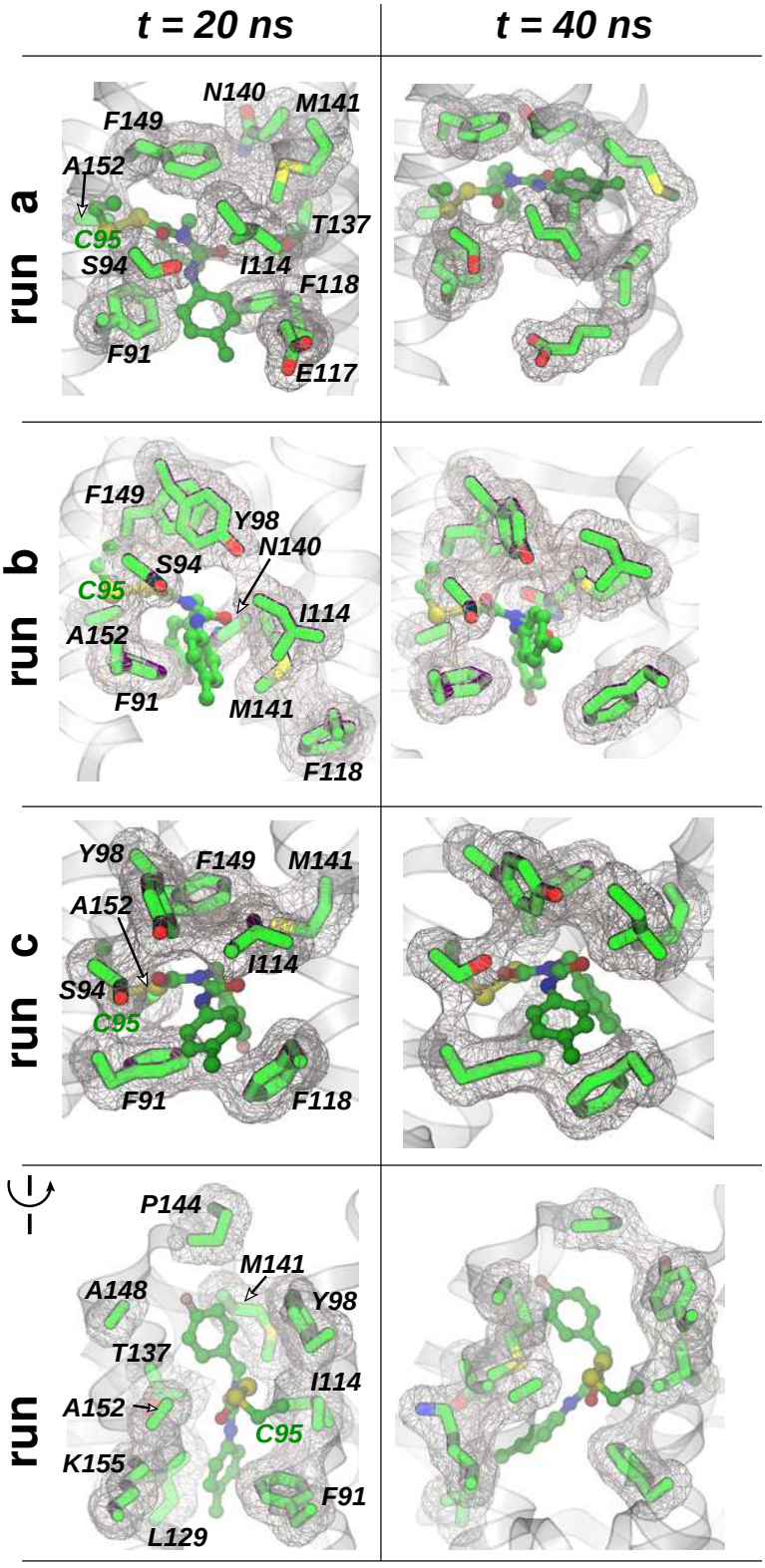

FIGURE S9. Same data as in Fig. S8 are shown for four independent 40-ns long MD simulations (run#13 in Table S1). See caption of Fig. S8 for details. The bottom-most panels are ∼90° rotated in comparison to all top panels for a better view of CCG-50014 and surrounding residues.

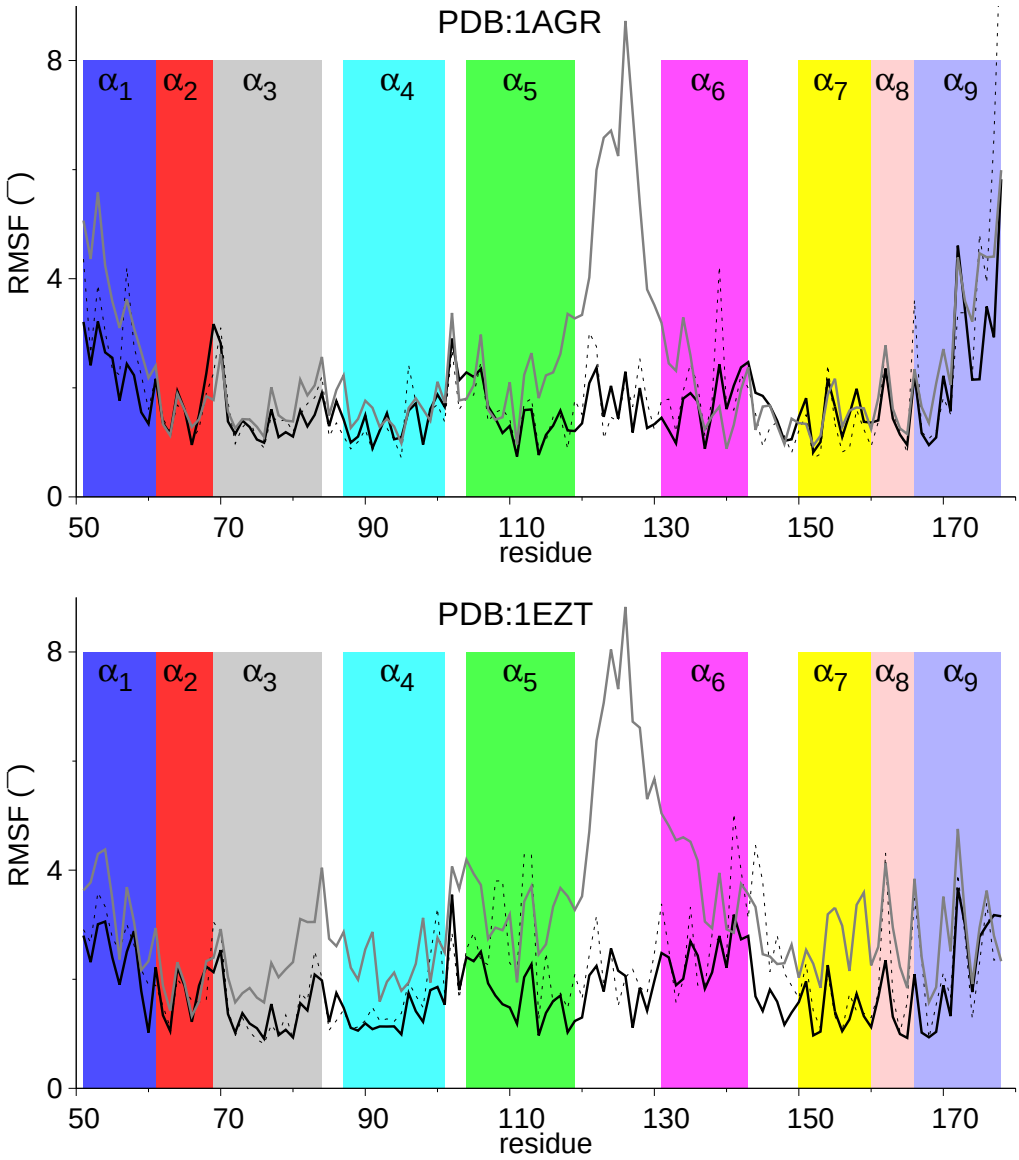

FIGURE S10. Per-residue root mean square fluctuation (RMSF) computed from MD equilibration trajectories (runs#1, 2, 7, 10, 11 and 13 in Table S1) of wild-type RGS4 (dotted line), mutant RGS4 with CCG-50014 (solid gray line) and mutant RGS4 without CCG-50014 (solid black line). RMSF for mutant RGS4 runs with and without 50014 are averages over at least three independent MD trajectories. See Fig. 5 in the main article for change in RMSF per residue (ΔRMSF) on binding of CCG-50014.

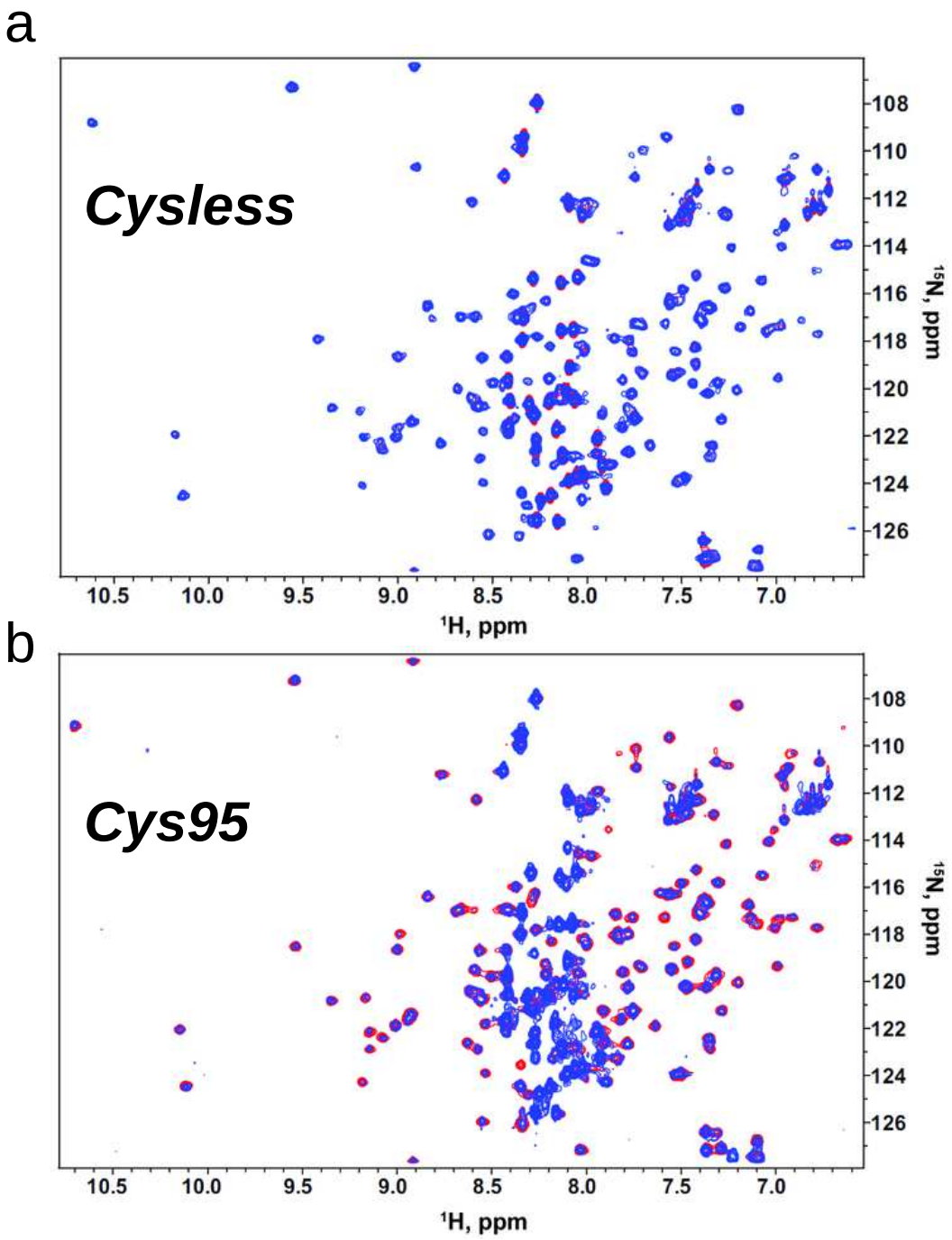

FIGURE S11. ^{1}H-^{15}N HSQC spectra of Cysteine less (Cysless) and single Cysteine (Cys95) RGS4 are shown in the top and bottom panels. Spectra were recorded before (red) and after (blue) CCG-50014 exposure.

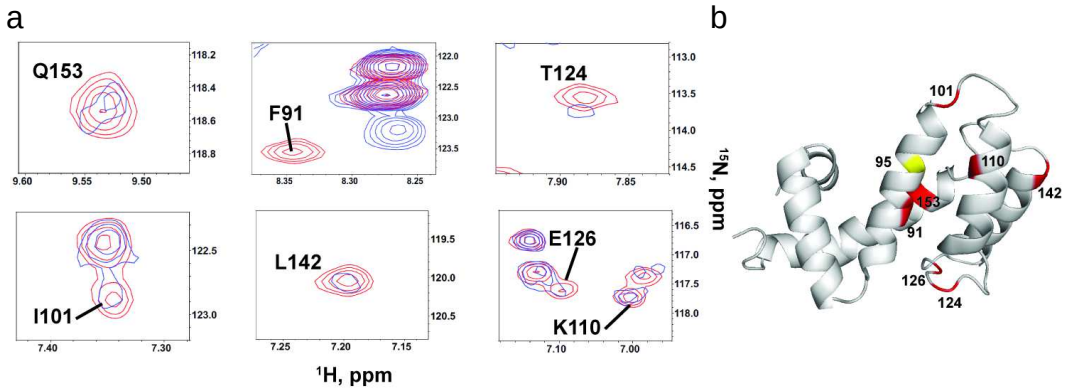

FIGURE S12. (a) Regions of the ^{1}H-^{15}N HSQC spectra for Cys95-RGS4 that show residues in the four helix bundle (α_4-α_7) as perturbed greater than 1 SD in the presence of CCG-50014 (blue) overlaid onto the reference spectra (red). (b) Key residues (red; 91, 101, 110, 142, and 153) of RGS4 that are potentially in contact with CCG-50014 bound at Cys95 (yellow) or may be perturbed as a part of the allosteric conformational changes that result in disruption of the G-protein binding surface of RGS4 (residues 124 and 126).

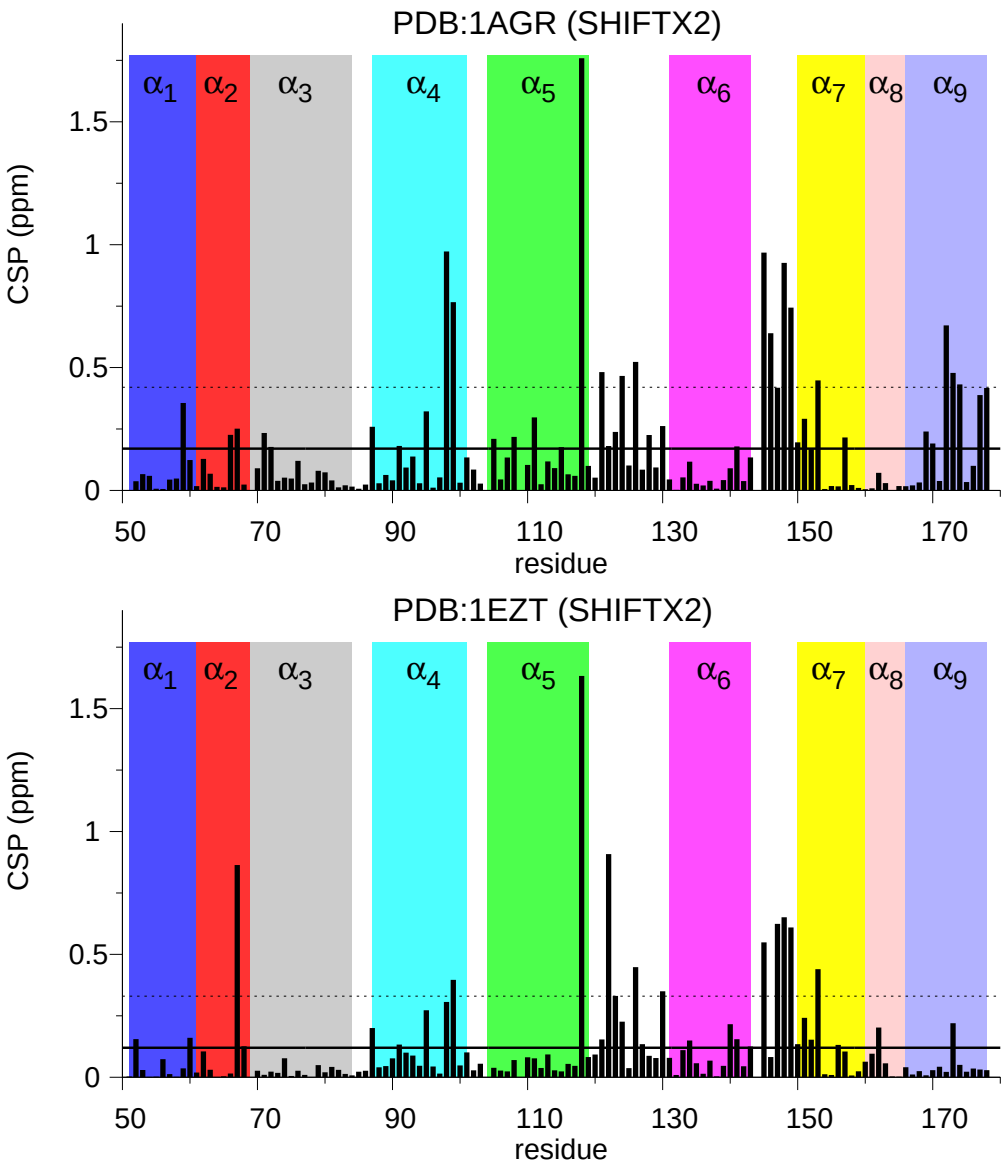

FIGURE S13. . Chemical shift perturbation (CSP; ppm) per residue computed using software package SHIFTX2. Computed CSPs are based upon MD simulations of mutant (Cys95) RGS4 with and without CCG-50014 (see Supplemental Methods for CSP prediction details). Top and bottom panels, respectively, correspond to CSP predictions based upon MD trajectories starting with PDB structures 1AGR and 1EZT. Solid and dotted horizontal lines indicate values at mean and 1 standard deviation (SD). Residues that are predicted to be perturbed larger than mean values are listed in Table S3.

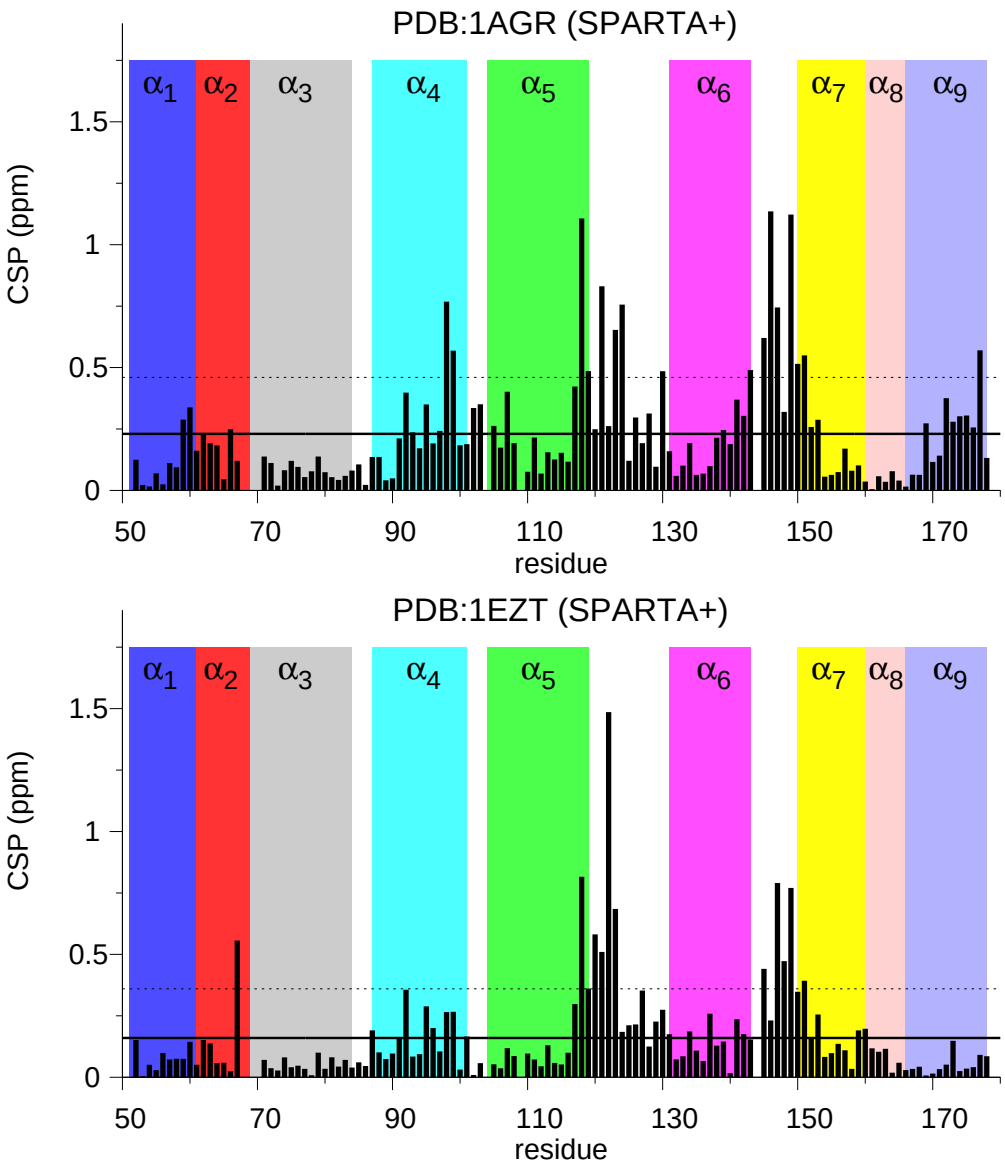

FIGURE S14. Same data as in Fig. S13 are shown for predictions using software package SPARTA+. See caption of Fig. S13 for additional details.

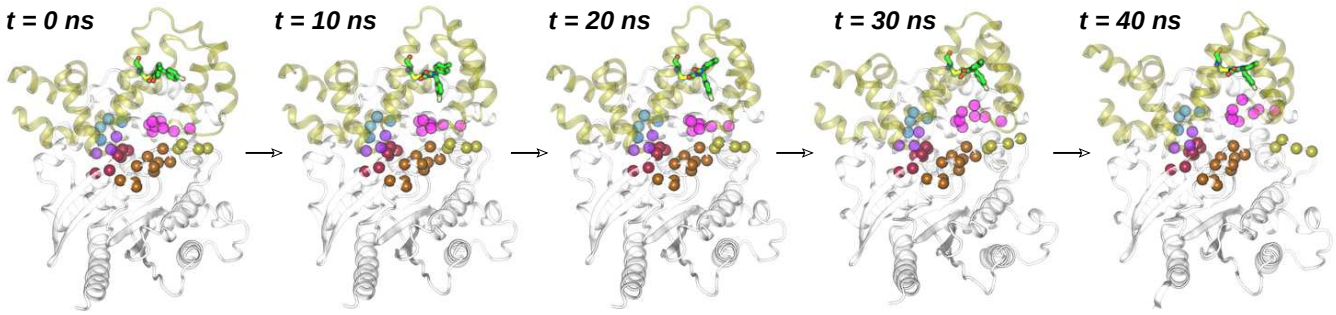

FIGURE S15. Snapshots at different time-points from a 45-ns long MD simulation of CCG-500014 bound RGS4-G_α complex (run#9 in Table S1). See Fig. 8 in the main article for coloring and labeling details.