Received: 20 March 2021 | Revised: 19 June 2021 | Accepted: 25 July 2021

DOI: 10.1002/jcc.26747

FULL PAPER

Journal of COMPUTATIONAL CHEMISTRY WILEY

Mixed-solvent molecular dynamics simulation-based discovery of a putative allosteric site on regulator of G protein signaling 4

Wallace K. B. Chan[1] | Debarati DasGupta[2] | Heather A. Carlson[2] | John R. Traynor[1,2]

[1]Department of Pharmacology, Edward F Domino Research Center, University of Michigan, Ann Arbor, Michigan, USA

[2]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
John R. Traynor, Department of Pharmacology, Edward F Domino Research Center, University of Michigan, Ann Arbor, MI 48109-5632, USA.
Email: jtraynor@umich.edu

Funding information
National Institute of General Medical Sciences, Grant/Award Number: R01 GM065372; National Institute on Drug Abuse, Grant/Award Numbers: R01 DA035316, R37 DA039997

Abstract

Regulator of G protein signaling 4 (RGS4) is an intracellular protein that binds to the G_α subunit ofheterotrimeric G proteins and aids in terminating G protein coupled receptor signaling. RGS4 has been implicated in pain, schizophrenia, and the control of cardiac contractility. Inhibitors of RGS4 have been developed but bind covalently to cysteine residues on the protein. Therefore, we sought to identify alternative druggable sites on RGS4 using mixed-solvent molecular dynamics simulations, which employ low concentrations of organic probes to identify druggable hotspots on the protein. Pseudo-ligands were placed in consensus hotspots, and perturbation with normal mode analysis led to the identification and characterization of a putative allosteric site, which would be invaluable for structure-based drug design of non-covalent, small molecule inhibitors. Future studies on the mechanism of this allostery will aid in the development of novel therapeutics targeting RGS4.

KEYWORDS
allosteric site prediction, computational solvent mapping, regulator of G protein signaling

1 | INTRODUCTION

The regulator of G protein signaling (RGS) family of proteins is responsible for modulating G protein-coupled receptor (GPCR)-mediated signaling. Members of the family are characterized by a 125 amino-acid RGS homology (RH) domain. Upon binding an agonist, a $G_{\alpha/io}$-coupled GPCR undergoes a conformational change leading to recruitment of heterotrimeric G protein complex ($G_{\alpha\beta\gamma}$), exchange of GDP for GTP on the G_α subunit, and dissociation from the $G_{\beta\gamma}$ heterodimer. Both G_α-GTP and $G_{\beta\gamma}$ each interact with various downstream signaling partners. RGS proteins interact with the active G_α-GTP subunit and operate as GTPase-accelerating proteins (GAPs); the ensuing hydrolysis of GTP to GDP leads to the reformation of $G_{\alpha\beta\gamma}$ complex and termination of signaling. Consisting of over 20 members, the RGS family of proteins can be further divided into several subfamilies based on sequence alignment of the conserved RH domains.[1] RGS4 is a small member of the family that consists of the RH domain plus a short N-terminal tail. RGS4 has been implicated in several disease states, including schizophrenia,[2] Parkinson's disease,[3] and drug addiction.[4] Because of its wide ranging physiological roles, numerous research efforts have explored the utilization of inhibitors as pharmacological tools and potential novel drugs (such as CCG-50014 which operates through covalent attachment to certain cysteine residues present in RGS4).[5–8] Such inhibitors have been used to confirm roles for RGS4 in Parkinson's disease[9] and acute pain signaling.[10] However, a noncovalent, reversible inhibitor against RGS4 has yet to be discovered.

While targeting the interface between RGS4 and G_α-GTP (the so-called A-site) seems obvious, there have been no small molecules developed against this region, though there has been growing interest in targeting an allosteric site on RGS4.[11] Phosphatidylinositol-3,4,5-trisphosphate (PIP_3) and calmodulin are two known endogenous allosteric modulators of RGS4; PIP_3 inhibits GAP activity of RGS4, while calmodulin reverses PIP_3-mediated inhibition.[12,13] Mutation of a pair of lysine residues (Lys-99 and Lys-100) to alanine residues was shown to abrogate binding of both PIP_3 and calmodulin.[14] The

scaffolding protein axin contains an RH domain. The structure of this protein was solved in complex with a peptide fragment from the adenomatous polyposis coli (APC) protein, in which a serine residue from the peptide makes a backbone contact with the side group of Lys-163.[15] This binding site for the APC protein on the RH domain was termed the B-site and is found in proteins with the RH domain fold.[11] Lys-163 in axin corresponds with Lys-100 in RGS4, so since Lys-99 and Lys-100 are well conserved across members of the RGS family,[14] the functional importance of these residues are very likely to be significant.[11] However, the putative B-site has yet to be fully characterized on the RGS family of proteins and remains an area of investigation. Targeting the B-site and other potential allosteric sites on RGS4 would be of potential therapeutic value.

In the present study, we employ mixed-solved molecular dynamics (MixMD) simulations[16] to identify putative allosteric sites on RGS4. This technique uses a low concentration of each of three organic probes (pyrimidine, acetonitrile, and isopropyl alcohol) to find ligand-binding hotspots on a protein based on the binding preference of the probes. This methodology has the advantage of utilizing protein dynamics to find pockets that might otherwise be inaccessible with a single snapshot crystal structure. After identifying hotspots on the wild type and a mutant in which both Lys-99 and Lys-100 were replaced with alanine residues (K99A/K100A), we probed the predicted hotspots with normal mode analysis (NMA). Pseudo-ligands were placed in high-confidence hotspots and perturbation of the NMA was used to identify which were potential regulatory sites. We identified five predicted sites, including the A-site and a site found to contain one of the residues responsible for the binding of PIP_3 and calmodulin.

2 | METHODS

2.1 | Structure preparation and system setup

The ensemble-averaged, solution NMR structure of rat RGS4 (PDB: 1EZT) was employed for the current study,[17] and its structure was prepared with Molecular Operating Environment (MOE, version 2019.0101) for molecular dynamics simulations. The N- and C-termini were capped with acetate and *N*-methyl amide, respectively. Hydrogens were added, and the protonation states of all residues were set to pH 7.0. In silico mutations (K99A/K100A) for RGS4 were performed using the Protein Builder module from MOE.

From AmberTools18, tleap was utilized to prepare the relevant input files using the prepared structure with the AMBER ff14SB force field.[18] The protein was surrounded with a layer of chemical probes (pyrimidine, acetonitrile, or isopropyl alcohol) before layering the rest of the system with TIP3P water. The concentration of probe solvent was 5% vol/vol.[16] The resulting systems were neutralized with counterions and had a dimension of ~78 Å × 75 Å × 71 Å. An example is shown for RGS4 layered with pyrimidine in Figure S1.

2.2 | MixMD simulations

Molecular dynamics simulations were run using the GPU implementation of PMEMD from AMBER18.[19] A first round of minimization was performed on the systems with a 10 kcal/mol·Å^2 harmonic restraint on the protein, which consisted of 2500 steps of steepest descent followed by 2500 steps of conjugate gradient. The second round of minimization was the same as the first apart from removing the harmonic restraint. Subsequently, the systems were gradually heated in the NVT ensemble from 0 to 100 K for 12.5 ps, then from 100 to 310.15 K in the NPT ensemble for 125 ps at 1 bar; a 1 fs timestep was used, along with a 10 kcal/mol·Å^2 harmonic restraint on the protein. After reaching the target temperature, the systems were equilibrated in the NPT ensemble at 310.15 K and 1 bar with a 2 fs time step. Beginning with a 5 kcal/mol·Å^2 harmonic restraint on the protein, the restraint was reduced by 1 kcal/mol·Å^2 every 500 ps for a total of 2.5 ns, then by 0.1 kcal/mol·Å^2 every 500 ps for a total of 5 ns. Production simulations had no constraints and were run in the NPT ensemble at 310.15 K and 1 bar for 50 ns with a 2 fs time step. All bond lengths concerning hydrogen were constrained with the SHAKE algorithm. Non-bonded interactions were set to cut off at 9.0 Å, while Particle mesh Ewald summation was used for long-range electrostatics. The Langevin thermostat and Monte Carlo barostat were used where applicable. Periodic boundary conditions were applied to all heating, equilibration, and production runs. For both wild-type and mutant RGS4, 10 independent production runs were carried out for each of the three probes (pyrimidine, acetonitrile, and isopropyl alcohol), resulting in a total of 3 μs simulation time.

2.3 | Probe occupancy calculations

Using the ptraj program from AmberTools18, all of the resulting trajectories for each probe were centered, imaged, and aligned, then the locations of the probe atoms from the last 25 ns of each independent simulation were combined and binned onto a grid with 0.5 Å spacing. To compare across conditions, Z-scores were then calculated using the following equation to normalize the occupancies:

$$z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

where x_i is the occupancy at grid point i, μ is the mean occupancy of all grid points, and σ is the SD of occupancy at all grid points. The normalized occupancy maps for each probe could then be visualized like electron densities with PyMOL (version 2.3.5).[20]

2.4 | Assessment of predicted allosteric sites

MixMD Probeview[21] was used to identify potential allosteric sites from the occupancy data using the averaged protein structure from each probe in PyMOL. The default parameters for DBSCAN

clustering were set at occupancy cutoff = 0.1, $\varepsilon = 3$, and minimum number of points = 10.

To simulate the effect of having a bound ligand in each site, an octahedron composed of six carbon dummy atoms was manually placed in regions where all three probes had overlapping occupancies, adapted from that performed in Panjkovich and Daura.[22] The edges of the octahedron (i.e., carbon–carbon distance) were 1.5 Å in length. With these pseudo-ligands positioned on the protein, further analyses involving its effects on protein dynamics could then be performed.

Custom R scripts were written using the Bio3D package (version 2.4-1) to assess the impact of the pseudo-ligand on overall protein motion.[23,24] NMA calculations were performed using an all-atom elastic network model (ENM) with the *aaenm* force field. The lowest-frequency modes, which correspond with large-amplitude conformational changes, were used in the study. Theoretical temperature factors were calculated as follows for each non-hydrogen atom:

$$B = \frac{8\pi^2}{3}\langle \mu^2 \rangle \quad (2)$$

where $\langle \mu^2 \rangle$ is the mean squared displacement. Wilcoxon signed-rank tests were used for an atom-wise comparison of predicted temperature factors between the bound (holo) and unbound states (apo), generating two-tailed p values. p Values <0.001 were considered significant. Additionally, coarse-grained ENM using the *calpha* force field was used to generate the dynamic cross-correlation matrix.

2.5 | Probe kinetics

Fragments have a low binding affinity and so it is challenging to assess the binding affinities of the probes from simulation studies. However, direct evaluation of on–off rates and binding kinetics could be done if there is constant exchange of probes with bulk solvent. This resembles an equilibrium scenario of enough binding and unbinding phenomena. Caflisch and coworkers have reported an elegant route for assessment of probe binding kinetics based on on–off rates from MD simulations.[25] In a recent publication, Pan et al extracted on–off rates for drug like fragments with millimolar affinity using long scale simulation data.[26] They performed microsecond long simulations and observed spontaneous binding and unbinding repeatedly occurring throughout the production runs. In this work, the large number of binding and unbinding events to the allosteric site detected, allows for extraction on on–off rates from the MixMD simulations.

The K_D values (mM) were estimated using the equation:

$$K_D = \frac{k_{off}}{k_{0n}} = \frac{t_{on}}{t_{off}} \times [\text{Probe}] \quad (3)$$

2.6 | Conformational clustering of Site 3

POVME[27] was employed to explore the possible binding site conformations that Site 3 could adopt. One hundred evenly spaced frames were extracted from the last 25 ns of each trajectory and aligned to the starting structure. An inclusion sphere with a radius of 12 Å was centered on Site 3 with a grid spacing of 1 Å. Convex hull exclusion was applied with the first frame as reference to remove extraneous points outside the potential allosteric site (i.e., out in solution). Tanimoto coefficients were calculated on overlapping grid points between each frame, and hierarchical clustering with a Kelly penalty was performed.

2.7 | Molecular docking

An SDF file consisting of 10,240 compounds from the "Discovery Diversity Set" was obtained from Enamine (New Jersey), which was then imported into a MOE database. All compounds were subjected to the "Wash" module; the dominant protonation state at pH 7.4 was set for each compound, followed by the generation and minimization of the 3D structure.

GOLD[28] was used for molecular docking experiments. Representative structures from the top 5 most populated conformational clusters of Site 3 from RGS4 were used for ensemble docking. A point was specified in the center of Site 3, and all residues within 12 Å were considered part of the binding site. Ring conformations were explored using the "flip ring corners" and "match template conformations" options, while pyramidal nitrogens and amide bonds were allowed to flip. For each compound, 10 independent genetic algorithm runs were used with the "ensemble" search efficiency, and conformations were assessed with the ChemPLP scoring function. The top-scoring compound was ultimately saved for analysis, in addition to the RGS4 conformer with which it was associated.

3 | RESULTS

3.1 | Prediction of allosteric sites on RGS4

MixMD simulations were performed using the solution NMR structure of rat RGS4 (PDB: 1EZT).[17] Utilizing the three small organic probes, pyrimidine, acetonitrile, and isopropyl alcohol, binding hotspots were identified through visual analysis of probe occupancies from the simulations (Figure 1A). In order to identify and rank the predicted binding sites, we applied MixMD Probeview[21] to the probe occupancies, and subsequently the top-five predicted binding sites were chosen for further analysis (Figure 1B). As compared to using a more stringent level of significance (35σ), a more complete manifestation of the sites was observed at a lower level (20σ). In particular, Sites 1–3 possessed very strong probe density, while Sites 4 and 5 did not map as well (Figure 1). This suggests that the former three sites have a greater propensity for being physiologically relevant. A small pocket was mapped to 35σ by all three probe hotspots at Site 2 (Figure S3). This site is known from a previous study of the structure of $G_{i\alpha 1}$ in complex with RGS4 (PDB: 1AGR) where $G_{i\alpha 1}$ interacted with RGS4 primarily through three switch regions.[29] The Thr-182 from switch I

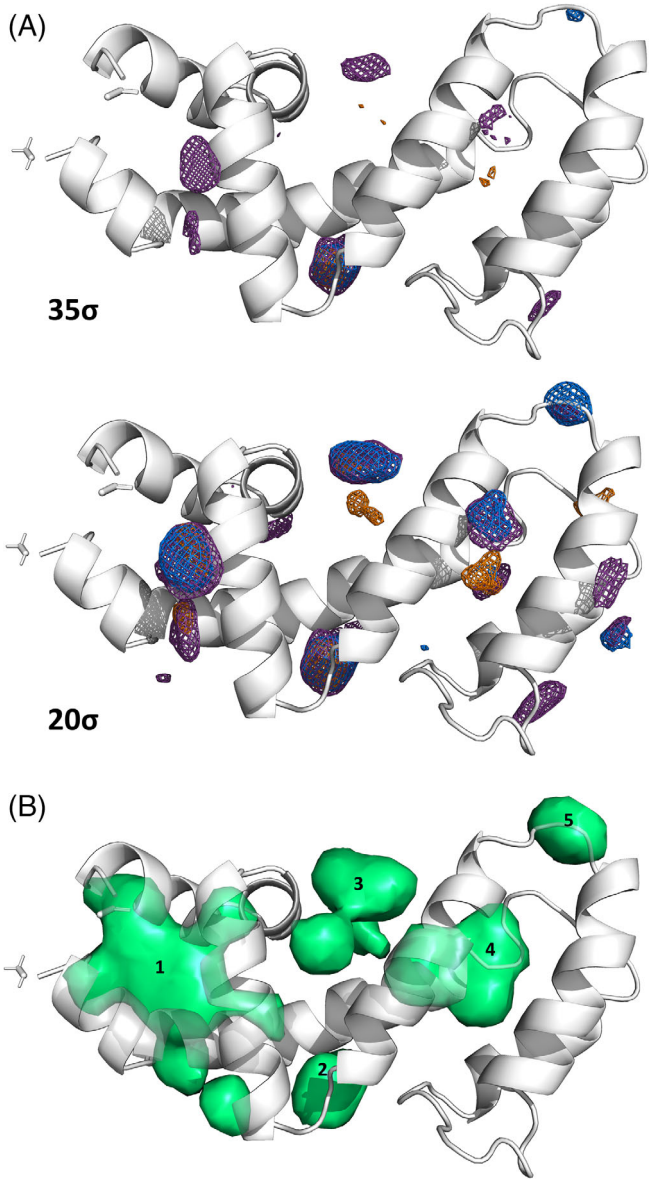

FIGURE 1 Mixed-solved molecular dynamics (MixMD) analysis of regulator of G protein signaling 4. (A) Hotspots for pyrimidine (purple), acetonitrile (orange), and isopropyl alcohol (blue) are shown contoured at 35σ and 20σ. (B) Predicted allosteric sites generated with MixMD ProbeView are shown (green)

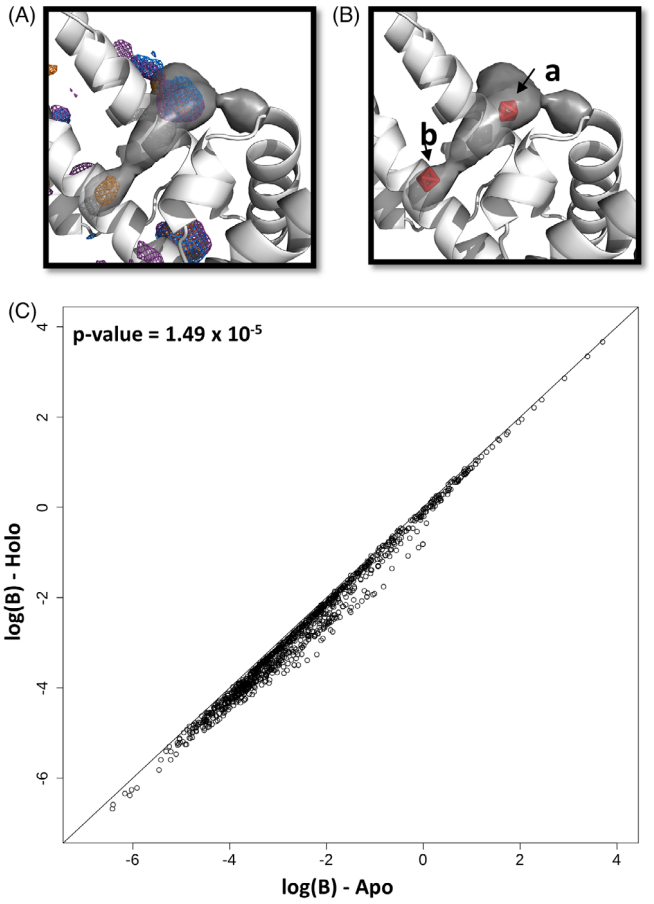

FIGURE 2 Perturbation of Site 3 using normal mode analysis. (A) Hotspots for pyrimidine (purple), acetonitrile (orange), and isopropyl alcohol (blue) are shown contoured at 20σ. The average structure from the pyrimidine simulations was used. (B) Carbon octahedron pseudo-ligands were placed in the hotspots (a + b) within the predicted allosteric site to simulate a bound ligand. (C) All-atom normal mode analysis using the first 10 non-trivial modes resulted in an overall reduction of predicted temperature factors in the holo state as compared to the apo state

(residues 176–184), which is completely buried by a pocket formed by conserved residues from RGS4,[29] overlaps with the predicted pocket from the 35σ contour map (Figure S2). By lowering the contour to 20σ, we revealed additional sites where hotspots from multiple probes overlapped, albeit with increased noise (Figure 1A, bottom). The binding site designated as Site 2 was consistent with the pocket that accommodated Thr-182 from switch I of $G_{\alpha i1}$.

In addition to Site 2, we also examined the remaining sites in detail. Site 1 was star shaped and contained two distinct hotspots; one was strongly mapped by all three probes, while the other was mapped by pyrimidine and acetonitrile (Figure S3). Similarly, Site 3 was located on a cleft opposite that of Site 2 and also contained two distinct hotspots; one was strongly mapped by all three probes, while the other was mapped only by acetonitrile (Figure 2A). Though hotspots within Sites 4 and 5 were present at 20σ, their probe occupancies were much more spurious at 35σ, unlike the first three sites (Figures 1 and S3).

To supplement our MixMD-based predictions, we submitted the RGS4 structure (PDB: 1EZT) to both the FTSite and PARS servers, both of which operate on a static protein structure.[30,31] The PARS server uses the method from Panjkovich and Daura, which predicts allosteric sites based on NMA.[22,31] Conversely, FTSite docks chemical probes onto a static protein structure, and sites are identified by consensus clusters of probes.[32] Three sites were found using the FTSite algorithm (Figure S4A); the top two sites could be considered one continuous site and overlapped with Site 3 (Figure S4A, pink and green), while the last site overlapped with Site 2 (Figure S4A, blue). Independently, the PARS server identified binding sites that also

corresponded with the same sites (Figure S4B, orange). Combination of the three approaches resulted in a consensus of Sites 2 and 3.

3.2 | Effect of octahedron pseudo-ligands on protein flexibility with NMA

To examine the impact of a simulated allosteric modulator on protein flexibility, a single carbon octahedron pseudo-ligand was manually placed into locations in which the probe occupancies had the greatest agreement (Figures 2Ba and S3, top). This method was previously shown by Panjkovich and Daura to aid in the positive identification of known allosteric sites in a retrospective study without a priori knowledge.[22] Subsequently, NMA was performed to compare differences in flexibility between the apo and holo structures. Using the first 10, 20, and 30 non-redundant normal modes, no difference was observed for Sites 1, 4, and 5 (Table S1). However, the placement of an octahedron at the high-confidence hotspot at Site 3 produced a statistically significant change in flexibility for RGS4 within all relevant ranges of normal modes (Table S1, Site 3—a). Known to interact with $G_{i\alpha1}$, Site 2 produced an appreciable difference in flexibility using the first 10 non-redundant normal modes (p value = 0.076), though it was not statistically significant. The mechanism by which $G_{i\alpha1}$ interacts with RGS4 to enhance its GTPase activity requires interaction with all three of its switch regions,[29] so it is expected that the sole perturbation of Site 2 on switch region 1 will not change the dynamics of RGS4 substantially, though it did appear to have a marginal impact on protein flexibility. Both Sites 2 and 3 were consistent with predictions from the PARS server and FTSite (Figure S3). Taken together, we hypothesized that Site 3 potentially harbors allosteric activity.

3.3 | Identification of Site 3 of RGS4 as potential allosteric site

From visual inspection of the hotspots at Site 3, we observed additional acetonitrile occupancy at 20σ (Figure 2Bb), which could be connected to the high-confidence hotspot based on the occupancies using MixMD Probeview to form a contiguous site (Figure 2A). Interestingly, this site encompasses one (Lys-99) of the two residues (Lys-99/Lys-100) that are known to interact with the endogenous modulators, PIP_3 and calmodulin.[14] Hence, we performed MixMD on a mutant RGS4 with both residues mutated to alanine. This resulted in Site 3 largely remaining the same, though a slight increase in probe occupancy was observed where Lys-99 was replaced with alanine (Figure S5); this suggests that the lysine residues are likely not solely responsible for forming an allosteric site. Since the high-confidence hotspot at Site 3 was shown earlier to affect protein flexibility with the presence of the octahedron (Figure 2Ba), we decided to place an additional octahedron on the acetonitrile hotspot in Site 3 to simulate a ligand interacting at both subsites (Figure 2Bb). This produced a more significant effect on protein flexibility than placing an octahedron at the high-confidence hotspot on Site 3 alone and decreased the overall predicted temperature factors in the holo state as compared with the apo state (Figure 2C; Table S1, Site 3—a + b). Additionally, it is evident that RGS4 does not exhibit drastic global movements in the apo state based on NMA, but the octahedrons appear to reduce RMSF in most residues in the holo state (Figure 3). Consequently, we sought to elucidate the effect of the pseudo-ligands on perturbation among each of the individual modes.

As the low-frequency modes represent the collective dynamics of a protein,[33] we examined the impact of the octahedrons on each of the first 10 non-redundant modes. Normal modes 8 and 12 resulted in the most statistically significant changes in protein flexibility, though modes 7, 10, 13, and 14 had a noticeable impact (Table 1). NMA trajectories for modes 8 and 12 are shown in Figure 4. The motion of RGS4 could be summarized by designating dynamic units comprised by two halves of the protein: H1 contains both N- and C-termini (Figure 4, left half), while H2 consists of the remainder of the protein (Figure 4, right half). Mode 8 describes a twisting motion between H1 and H2 of the protein demarcated by the cleft formed by Site 3. Conversely, mode 12 demonstrates a similar twisting motion as mode 8 with H1, while H2 simultaneously moves inward toward the cleft. From the NMA trajectories, we observed what appeared to be correlated motions between H1 and H2 (Figure 4). The movement between the two halves from both modes 8 and 12 are such that the presence of the octahedrons could potentially impinge overall movement. Therefore, we next examined the effect on regions of RGS4 with correlated motion.

We generated a dynamic cross-correlation map from NMA to assess whether the presence of the octahedron pseudo-ligands in Site 3 would affect coupled atomic fluctuation from non-contiguous residues (Figure 5A). Overall, correlated motion was decreased in the holo state, which was consistent with our observation of decreased

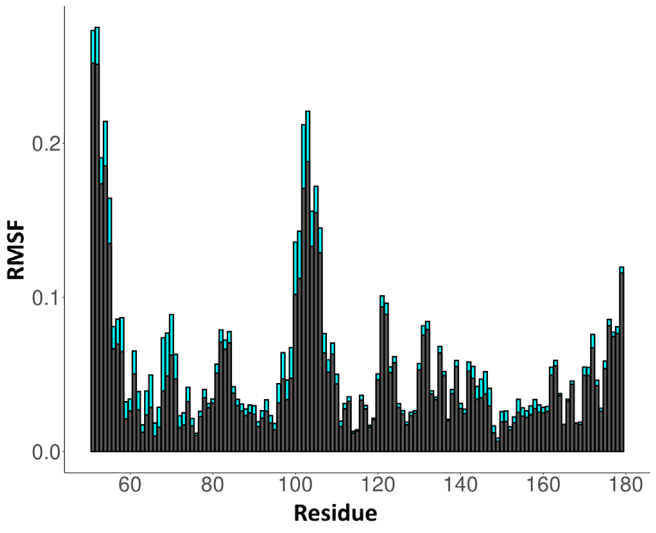

FIGURE 3 Root mean square fluctuation analysis of RGS4 with pseudo-ligands bound in Site 3. All-atom normal mode analysis using the first 10 non-trivial modes revealed an overall decrease in atomic fluctuations in the holo state (black) as compared to the apo state (cyan). C-alpha positions are shown for clarity

TABLE 1 Effect of pseudo-ligands on Site 3 with respect to the first ten non-redundant modes

Normal mode	*p* Value[a]	Normal mode	*p* Value[a]
7	0.005	12	1.290×10^{-5}
8	8.078×10^{-6}	13	0.044
9	0.874	14	0.011
10	0.007	15	0.499
11	0.095	16	0.066

[a]Two-tailed *p* values from Wilcoxon signed-ranked test.

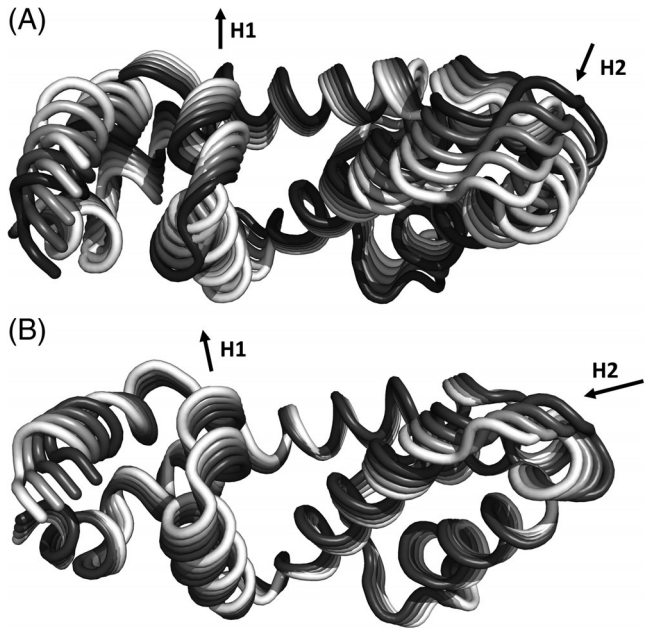

FIGURE 4 Normal mode analysis of RGS4. The dynamics of (A) mode 8 and (B) mode 12 are shown. H1 and H2 represent the movement of the two halves of RGS4, corresponding with the arrows (black to white)

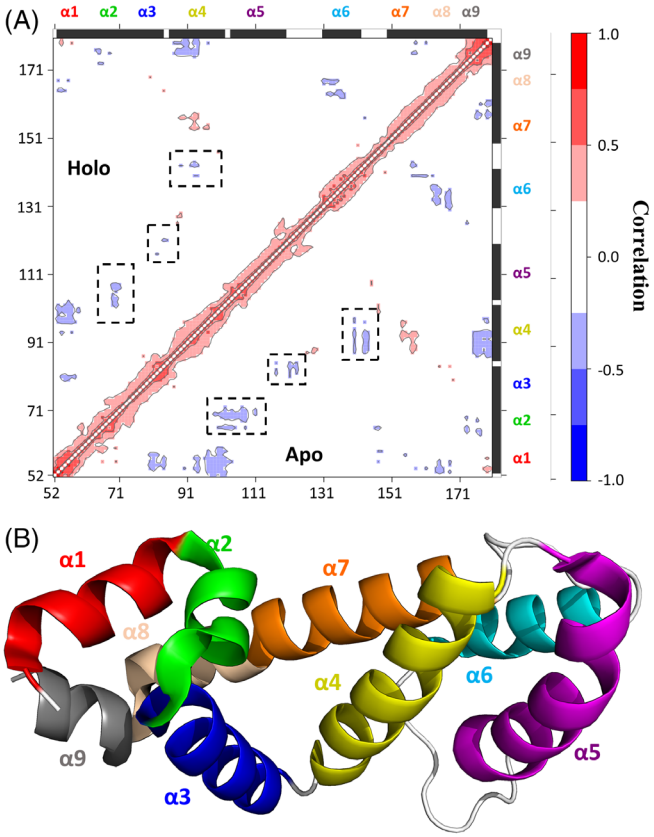

FIGURE 5 Effect of octahedron pseudo-ligands on correlated motion of RGS4. (A) A dynamic cross-correlation matrix was generated for RGS4 based on normal mode analysis (NMA). Analysis was conducted with (holo) and without (apo) pseudo-ligands bound in Site 3. The first 10 non-redundant modes were used in the calculations. (B) RGS4 contains 9 α helices, colored individually for clarity

predicted temperature factors. More specifically, coupling was markedly reduced in three regions: (1) between α2 and α4-α5, (2) between α3 and α5, and (3) between α4 and α6 (Figure 5A,B). Intriguingly, these regions correspond with the motion described with normal modes 8 and 12, which were negatively impacted by the presence of the octahedron pseudo-ligands (Figure 4). Moreover, α2, α4, and α7 form Site 3, of which the former two helices overlapped with the regions of reduced coupling. Overall, this indicates a potential role of Site 3 in an allosteric mechanism.

3.4 | Dissociation constants and analyses on on–off rates kinetics of Site 3

To assess the probe binding kinetics to the allosteric pocket, we used the method Caflisch proposed for analyzing DMSO binding-unbinding rates to the FKBP protein.[25] Our water miscible probes are all less than six heavy atoms, and they diffuse well in the 20–25 ns simulation time. The production runs for each probe type were analyzed to track the closest distance between the center of the hotspot and the probe in question (Figure 6). The analyses were done using two different distance thresholds; probes were considered bound when they were within 3 Å of the hotspot and unbound when they were over 6 Å. For verifying robustness of the calculations, a 7 Å distance cut-off was also employed for defining unbound states. Apart from our main goal of estimating K_D values from the simulations, usage of two different distance cut-offs showed that the affinities computed were not sensitive to the distance metric used to define the bound and unbound states. Using these thresholds, the trajectories were partitioned into bound and unbound states. The time lengths of each event were tallied, and a distribution of dwell times was estimated. The cumulative distributions of the binding and unbinding events were plotted and fitted to exponential decay plots and on–off times extracted. The fitting equation and plots are in shown in Figures S6–S9, while the values are reported in Table 2. We also used 7 Å as the distance threshold to probe unbinding and the values obtained were similar (Table 2). This reinforces Caflsich's hypothesis[25] that the binding

affinities are not extremely sensitive to the cut-off distances employed in our calculations. The values for on–off rates for using 3 and 7 Å cutoffs are tabulated in Table 2. We do find that the concentration of the cosolvent in the simulation directly impacts the calculated affinities.

3.5 | Characterization of Site 3 of RGS4

The full predicted allosteric site (Site 3) is shown in Figure 7 and is comprised of the following residues: Glu-64, Ile-67, Asn-68, Trp-92, Glu-96, Lys-99, Asp-150, Lys-154, Asn-158, and Glu-161. Three of the α-helices from RGS4 (α2, α4, and α7) contributed residues to form this cleft. Site 3 consists largely of polar residues, while also harboring five charged residues (Glu-64, Glu-96, Lys-99, Lys-154, and Glu-161). Also, the bottom of the cleft is lined with aromatic and nonpolar residues. The cleft itself is saddle shaped and appears as though it would best accommodate a "linear" small molecule (Figure 1B); this is examined below.

To assess the conformational diversity of Site 3, we performed clustering on the cleft from the last 25 ns of all trajectories for each probe with POVME.[27] Overall, the cleft appeared to maintain an average volume of ~1400 Å^3 (Figure S10A). Moreover, much conformational diversity was observed among the 25 clusters, though the top 5 most populated clusters harbored the majority of structures (Figure S10B). Cluster representatives from each of the top five clusters are displayed in surface representation in Figure S10C.

Apart from the conformational diversity of Site 3, we were also interested in identifying potential compounds which seek out this site. As a result, we performed molecular docking with GOLD[28] against the "Discovery Diversity Set" from Enamine using a representative structure from each of the top five conformational clusters. The docking scores followed a normal distribution centered around a docking score of ~55 (Figure S11). Moreover, manual inspection of the top 100 ranked compounds revealed that 65 of them interacted with RGS4 where the 2 pseudoligand-bound hotspots were located in Site 3 (Figures 2B and 8A). Perhaps unsurprisingly, most of them could be described as "linear" small molecules. The top-scoring "hit," Z346633068, had a docking score of 82.48, significantly higher than the next highest compound, Z2495891309, which had a docking score of 75.34 (Figures 8B and S11). While the docking score for Z346633068 was very high, we observed a cis-peptide bond with its pose (Figure 8C); though this is generally an unfavorable conformation, the remaining interactions made with RGS4 likely offset the energetic penalty. Much of the predicted interaction between Z346633068 and RGS4 appears to be governed by hydrophobic interactions with the nonpolar/aromatic residues in the cleft. Furthermore, the phenyl-2-pyrrolidinone portion of the compound overlaid with all three probe densities of Site 3a, while the tetrahydroisoquinolone moiety interacted with Site 3b (Figures 8C and 2B). However, an important insight from this pose is the potential for the formation of a salt bridge between the tertiary amine of the tetrahydroisoquinolone moiety of Z346633068, which is charged at pH 7.4, and Glu-64 of RGS4 (Figure 8C). In fact, this same salt bridge is seen for 17 other compounds out of the 65 potential hits, suggesting that Glu-64 could play a key role in ligand binding to Site 3.

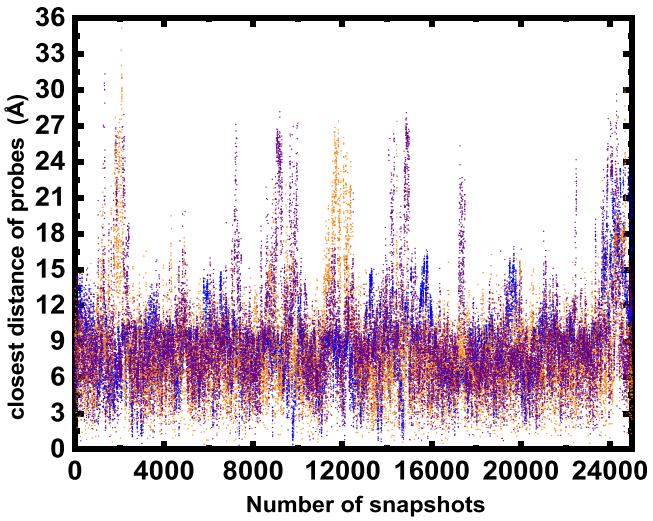

FIGURE 6 Closest distance of the probes (acetonitrile—orange, isopropyl alcohol—blue, pyrimidine—purple) to the geometric center of Site 3 on RGS4

4 | DISCUSSION

Altogether, we predicted five binding sites on RGS4 then independently assessed the effects of octahedron pseudo-ligands on protein flexibility. Sites 3 and 4 form a putative allosteric site commonly

TABLE 2 On–off rates and corresponding dissociation constants (K_D) estimated for probe binding kinetics

	Probe	Conc. (mM)	# Binding events	# Unbinding events	T_{on} (ns)	T_{off} (ns)	K_D (mM)
<3 Å bound states >6 Å unbound states	C3N	712	721	636	0.003	0.032	610.2
	IPA	603	385	358	0.008	0.063	76.6
	PYR	583	2042	1008	0.004	0.500	4.5
<3 Å bound states >7 Å unbound states	C3N	712	721	590	0.003	0.034	610.2
	IPA	603	385	337	0.008	0.120	41.59
	PYR	583	2042	877	0.004	0.629	3.71

Abbreviations: C3N, acetonitrile; IPA, isopropyl alcohol; PYR, pyrimidine.

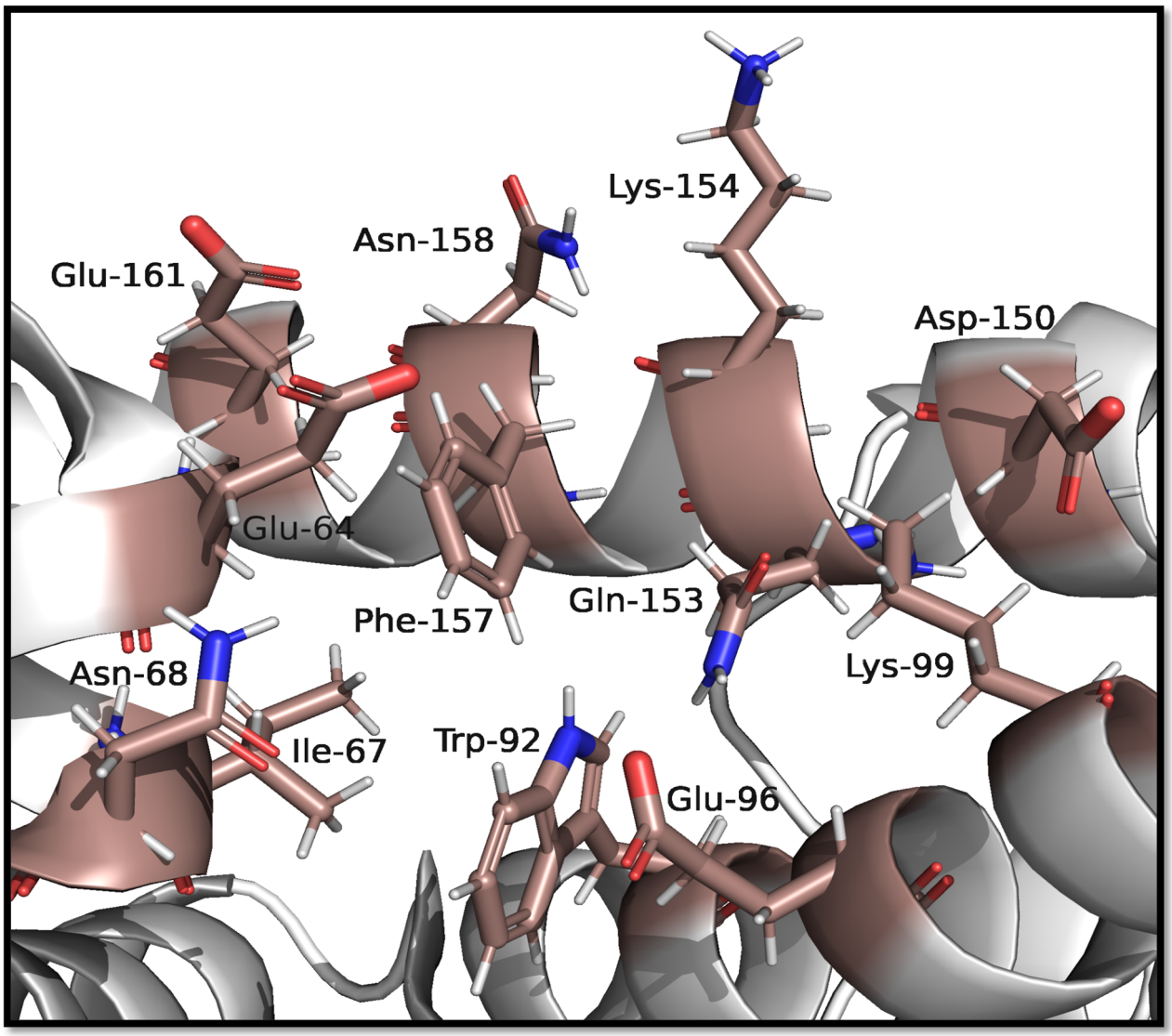

FIGURE 7 Proposed allosteric site of RGS4. Relevant residues from Site 3 are shown in dark salmon

referred to as the B-site.[11] Site 3 contains the residues, Lys-99 and Lys-100, that are required for interacting with PIP_3 and calmodulin,[14] but the full extent of this B-site has not yet been thoroughly characterized. Here, we show that occupancy of Site 3 with an octahedron pseudo-ligand altered protein flexibility and so is predicted to be an allosteric region. No such effect was seen at Site 4. It is possible that Site 4, which is analogous to the APC peptide-binding site on the RH domain of axin, contributes more to binding specificity over function. Site 2 is in agreement with a known pocket that interacts with switch I from $G_{i\alpha1}$ (Figure S2).[29] While novel, Sites 1 and 5 were not predicted by our method to affect protein flexibility, and though they may still bind small molecules, it is possible that no effects on function would result.

From our theoretical results, placement of the octahedron pseudo-ligands in Site 3 reduced RGS4 dynamics (Table S1). Using the low-frequency modes, it has previously been shown that protein flexibility is affected by the presence of an allosteric modulator in an allosteric site.[34] Accordingly, we believe Site 3 to be critical in an allosteric mechanism with RGS4 because it contains Lys-99, which is known to be crucial for PIP_3 and calmodulin binding.[14] Unfortunately, evidence from the literature has been scant regarding how both endogenous allosteric modulators bind to RGS4, but given its therapeutic potential, it would be advantageous to design a small molecule or peptide that could interact in a manner that mimics that of PIP_3. Complicating this is the fact that several PIP_3 analogues, including PIP_2, IP_4, and PIP_3 with truncated acyl chains, have been shown to be unable to inhibit GAP activity.[35] From this brief structure–activity relationship, it is evident that both the number of phosphate groups and length of the acyl chains are essential for inhibition of GAP activity. The importance of retaining the full or physiologically relevant acyl chains suggests that localization of RGS4 to the plasma membrane might play a part in the function of PIP_3. Moreover, the additional phosphate present in PIP_3 but not PIP_2 seems to be necessary for inhibition of GAP activity on RGS4. On the other hand, calmodulin

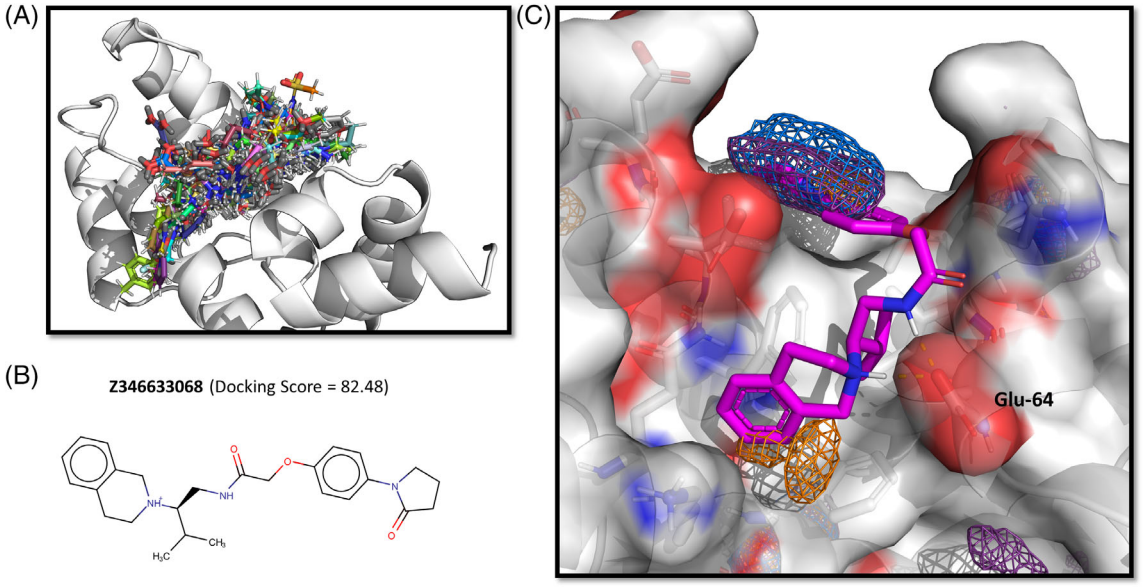

FIGURE 8 Molecular docking against the "Discovery Diversity Set" from Enamine. (A) Out of the top 100 ranked compounds, 65 of them (various colors) were observed to occupy both pseudoligand-bound hotspots from Site 3. (B) Z346633068 was the top-scoring, potential hit, with a docking score significantly greater than the next highest compound (Z2495891309). (C) In addition to interacting with RGS4 at both important hotspots on Site 3, the tertiary amine in Z346633068 can potentially form a salt bridge with Glu-64. Hotspots for pyrimidine (purple), acetonitrile (orange), and isopropyl alcohol (blue) are shown contoured at 20σ

may interact with RGS4 by binding the cleft formed by Site 3, which contains two basic residues, Lys-99 and Lys-154. Given that the copious amount of acidic residues on calmodulin aid in its charged interaction with targets,[36] it is possible that both of these basic residues may be involved in calmodulin binding, in addition to other basic residues within the vicinity. However, it is also possible that calmodulin interacts with these basic residues without binding within the cleft, acting to outcompete PIP_3. Taken together, it would be of therapeutic interest to further our understanding of how PIP_3 and calmodulin interact with Site 3 of RGS4 by solving their structures in complex with each of the allosteric modulators.

To the best of our knowledge, the identification and characterization of Site 3 represents the first prediction and description of an allosteric site on RGS4 that could potentially bind small molecules in a noncovalent fashion (Figure 7). Since calmodulin is known to require lysine residues for interaction,[14] it is possible that the binding of a small molecule inhibitor at Site 3 could partially abrogate calmodulin binding, though the same could be seen for PIP_3. This could either inhibit (i.e., such as PIP_3) or maintain (i.e., such as calmodulin) GAP activity on RGS4. Nonetheless, how small molecules interact with Site 3 and what type of functional effect a noncovalent inhibitor would possess remains unclear, and future experiments will be required to address these questions. Nonetheless, our data suggests that perturbation of Site 3 affects the conformational dynamics of RGS4 and that this site could be employed in future structure-based drug design efforts to develop novel, small molecule inhibitors against RGS4.

In addition to RGS4, the RGS family of proteins consists of over 20 members, many of which can be targeted by covalent inhibitors (e.g., RGS8 and RGS19).[37] It has been previously shown that RGS4 and RGS8 are more rigid than RGS19 due to a comparatively larger number of interhelical salt bridges and are consequently less susceptible to covalent inhibitors, such as CCG-50014; the greater flexibility of RGS19 makes it more likely for its single conserved cysteine to be exposed for attack by covalent inhibitors.[37] The relative rigidity of RGS4 was supported by our RMSF analysis of Site 3 in the apo state, while the inclusion of the octahedron pseudo-ligands in Site 3 was predicted to make the protein even more rigid (Figure 3). Moreover, the presence of a large-scale conformational change of RGS4 involving the opening of α5 and α6 using temperature-accelerated molecular dynamics simulations, which exposed Cys-95 for covalent attachment with CCG-50014, has been observed.[38] This would involve the breaking of two salt bridges each between α4-α5 and α6-α7, which would require the traversal of a high energy barrier. Though increased flexibility of RGS4 appears to be advantageous for certain covalent inhibitors, the effect of decreasing the flexibility of RGS4 by binding non-covalent inhibitors to Site 3 is currently unknown.

In a recent study, allosteric pathways originating from key conserved cysteine residues were elucidated across a panel of RGS proteins employing MD simulations.[39] Among the allosteric pathways stemming from Cys-95 on RGS4 to other residues that were known to make contact with the switch regions on $G_{i\alpha1}$, two residues (Trp-92 and Phe-157) agreed with our predicted allosteric site on multiple paths (Figure 7), the former residue of which the study suggests is important for allosteric regulation.[39] Additionally, mutation of Trp-92 to alanine was previously found to partially impair GAP activity,[40] inferring the residue's importance in function. While covalent attachment of CCG-50014 to Cys-95 perturbs allosteric communication on RGS4, it is certainly possible that a small molecule interacting

noncovalently with our predicted site would confer a similar effect through disruption of the allosteric pathway, as this concept has been similarly demonstrated with another protein, imidazole glycerol phosphate synthase.[41] As such, it would be of interest to perform in vitro mutations on Trp-92 or Phe-157 to see if there is an effect on allostery mediated by PIP_3 or calmodulin binding.

One of the most desirable characteristics of an inhibitor is target specificity. The residues that comprise Site 3 in RGS4 are well conserved among the RGS family of proteins, as shown from sequence alignment (Figure S12). Therefore, it would be a challenge to design a small molecule or peptide inhibitor that would preferentially target RGS4. On the other hand, this problem could potentially be circumvented by devising an allosteric modulator that would also interact with Site 4 similar to the way the APC protein associates with axin.[15] However, Site 4 is also well conserved among the RGS family (Figure S12), so the non-conserved residues from both sites involved in establishing interactions with the hypothetical inhibitor would have to be carefully considered. It is interesting to note that PIP_3 was previously demonstrated to inhibit GAP activity with RGS4, RGS10, and RGS19 (i.e., GAIP) but not RGS16.[12] While RGS10 and GAIP are in different families, RGS4 and RGS16 both belong to the R4 family of RGS proteins and share 44% sequence identity.[42] Moreover, the corresponding residue properties shared between the two proteins at Site 3 are very similar (Figure S12), so whatever is responsible for the specificity of PIP_3 to Site 3 on the RGS family of proteins is subtle and not immediately discernible by sequence alignment. One avenue that can be pursued is the systematic mutation of basic residues on RGS4, especially those near Lys-99. Additionally, Lys-99 and Lys-100 have only ever been mutated together as a pair,[14] so the importance of each residue individually on PIP_3 binding has yet to be established; to the best of our knowledge, no other basic residues on RGS4 have been mutated. Given that PIP_3 but not PIP_2 can inhibit GAP activity, the number of interacting basic residues should ostensibly correspond with all headgroup phosphates present on PIP_3.

The present study adapted the method of Panjkovich and Daura for assessing protein flexibility with pseudo-ligands.[22] However, several modifications were made to their protocol to better accommodate binding site predictions from MixMD simulations. First, the edges of the octahedron (i.e., carbon–carbon distance) in the original study were ~5.6 Å, while we chose 1.5 Å. In general, the smaller octahedrons fit better onto probe hotspots, as opposed to the bulkier octahedrons, and better represented where a fragment might bind. Second, Panjkovich and Daura employed a coarse-grained NMA approach and only used backbone Cα atoms, while we used all-atom NMA, which has been previously shown to have better agreement with fluctuations from MD simulations.[43] Third, the original study used known sites, though it was later adapted into the PARS server[31] and used binding site predictions from the algorithm, LIGSITE^{CSC},[44] on static protein structures. Conversely, our method used MixMD simulations to generate allosteric site predictions, which considers protein flexibility and has the ability to tease out cryptic sites not observable in crystal structures.[45] Due to the focus of the present study, we are unable to provide benchmark results against the PARS server and FTSite on a larger dataset of proteins-allosteric modulator complexes, though a future study would be of broader interest to systematically validate our method.

5 | CONCLUSIONS

In conclusion, we have demonstrated through MixMD simulations and perturbation of NMA using octahedron pseudo-ligands that RGS4 harbors a potential allosteric site. Furthermore, we fully characterized this site and found it to be consistent with multiple lines of evidence from the literature; this represents the first prediction of an allosteric site on RGS4 that could potentially bind small molecules noncovalently. However, further experimental investigation is required to elucidate its allosteric mechanism, especially in relation to PIP_3-mediated inhibition of GAP activity. Moreover, utilization of Site 3 in structure-based drug design could lead to the discovery of small molecule inhibitors of GAP activity against RGS4 and hence potentially a novel class of drug for medical conditions, such as pain management or schizophrenia.

ACKNOWLEDGMENTS

This research was supported by the National Institute of Drug Abuse (R01 DA035316 and R37 DA039997 to J. R. T.) and the National Institute of General Medical Science (R01 GM065372 to H. A. C.).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Wallace K. B. Chan https://orcid.org/0000-0001-5104-4159
Debarati DasGupta https://orcid.org/0000-0002-6289-5749
Heather A. Carlson https://orcid.org/0000-0002-7495-1699
John R. Traynor https://orcid.org/0000-0002-1849-8316

REFERENCES

[1] S. Hollinger, J. R. Hepler, *Pharmacol. Rev* **2002**, *54*(3), 527.
[2] P. Levitt, P. Ebert, K. Mirnics, V. L. Nimgaonkar, D. A. Lewis, *Biol. Psychiatry* **2006**, *60*(6), 534.
[3] T. N. Lerner, A. C. Kreitzer, *Neuron* **2012**, *73*(2), 347.
[4] D. Terzi, E. Stergiou, S. L. King, V. Zachariou, *Prog. Mol. Biol. Transl. Sci.* **2009**, *86*, 299.
[5] D. L. Roman, J. N. Talbot, R. A. Roof, R. K. Sunahara, J. R. Traynor, R. R. Neubig, *Mol. Pharmacol.* **2007**, *71*(1), 169.
[6] E. M. Turner, L. L. Blazer, R. R. Neubig, S. M. Husbands, *ACS Med. Chem. Lett.* **2012**, *3*(2), 146.
[7] C. A. Monroy, D. I. Mackie, D. L. Roman, *PLoS One* **2013**, *8*(4), e62247.
[8] A. J. Storaska, J. P. Mei, M. Wu, M. Li, S. M. Wade, L. L. Blazer, B. Sjögren, C. R. Hopkins, C. W. Lindsley, Z. Lin, *Cell. Signal.* **2013**, *25*(12), 2848.
[9] L. L. Blazer, A. J. Storaska, E. M. Jutkiewicz, E. M. Turner, M. Calcagno, S. M. Wade, Q. Wang, X.-P. Huang, J. R. Traynor, S. M. Husbands, *ACS Chem. Neurosci.* **2015**, *6*(6), 911.
[10] S.-Y. Yoon, J. Woo, J.-O. Park, E.-J. Choi, H.-S. Shin, D.-H. Roh, K.-S. Kim, *Anesth. Anal.* **2015**, *120*(3), 671.

[11] H. Zhong, R. R. Neubig, *J. Pharmacol. Exp. Ther.* **2001**, *297*(3), 837.
[12] S. G. Popov, U. M. Krishna, J. Falck, T. M. Wilkie, *J. Biol. Chem.* **2000**, *275*(25), 18962.
[13] M. Ishii, A. Inanobe, Y. Kurachi, *Proc. Natl. Acad. Sci. USA* **2002**, *99*(7), 4325.
[14] M. Ishii, S. Fujita, M. Yamada, Y. Hosaka, Y. Kurachi, *Biochem. J.* **2005**, *385*(1), 65.
[15] K. E. Spink, P. Polakis, W. I. Weis, *EMBO J.* **2000**, *19*(10), 2270.
[16] P. Ghanakota, H. A. Carlson, *J. Phys. Chem. B* **2016**, *120*(33), 8685.
[17] F. J. Moy, P. K. Chanda, M. I. Cockett, W. Edris, P. G. Jones, K. Mason, S. Semus, R. Powers, *Biochemistry* **2000**, *39*(24), 7063.
[18] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*(8), 3696.
[19] D. A. Case, H. M. Aktulga, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, III, G. A. Cisneros, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, C. Jin, K. Kasavajhala, M. C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K. M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K. A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, H. Wei, R. M. Wolf, X. Wu, Y. Xue, D. M. York, S. Zhao, P. A. Kollman, Amber 2021, University of California, San Francisco **2021**.
[20] Schrodinger, LLC. 2021. The PyMOL Molecular Graphics System, Version 2.3.5.
[21] S. E. Graham, N. Leja, H. A. Carlson, *J. Chem. Inf. Model.* **2018**, *58*(7), 1426.
[22] A. Panjkovich, X. Daura, *BMC Bioinform.* **2012**, *13*(1), 273.
[23] B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon, L. S. Caves, *Bioinformatics* **2006**, *22*(21), 2695.
[24] L. Skjærven, X.-Q. Yao, G. Scarabelli, B. J. Grant, *BMC Bioinform.* **2014**, *15*(1), 399.
[25] D. Huang, A. Caflisch, *ChemMedChem* **2011**, *6*(9), 1578.
[26] A. C. Pan, H. Xu, T. Palpant, D. E. Shaw, *J. Chem. Theory Comput.* **2017**, *13*(7), 3372.
[27] J. R. Wagner, J. Sørensen, N. Hensley, C. Wong, C. Zhu, T. Perison, R. Amaro, *J. Chem. Theory Comput.* **2017**, *13*(9), 4584.
[28] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Protein Struct. Funct. Bioinform.* **2003**, *52*(4), 609.
[29] J. J. Tesmer, D. M. Berman, A. G. Gilman, S. R. Sprang, *Cell* **1997**, *89*(2), 251.
[30] D. Kozakov, L. E. Grove, D. R. Hall, T. Bohnuud, S. E. Mottarella, L. Luo, B. Xia, D. Beglov, S. Vajda, *Nat. Protoc* **2015**, *10*(5), 733.
[31] A. Panjkovich, X. Daura, *Bioinformatics* **2014**, *30*(9), 1314.
[32] C.-H. Ngan, D. R. Hall, B. Zerbe, L. E. Grove, D. Kozakov, S. Vajda, *Bioinformatics* **2011**, *28*(2), 286.
[33] I. Bahar, A. Rader, *Curr. Opin. Struct. Biol.* **2005**, *15*(5), 586.
[34] S. Mitternacht, I. N. Berezovsky, *PLoS Comput. Biol.* **2011**, *7*(9), e1002148.
[35] Y. Tu, T. M. Wilkie, *Methods in Enzymology*, Elsevier, Amsterdam, Netherlands **2004**, p. 89.
[36] I. André, T. Kesvatera, B. Jönsson, K. S. Åkerfeldt, S. Linse, *Biophys. J.* **2004**, *87*(3), 1929.
[37] V. S. Shaw, M. Mohammadi, J. A. Quinn, H. Vashisth, R. R. Neubig, *Mol. Pharmacol.* **2019**, *96*(6), 683.
[38] H. Vashisth, A. J. Storaska, R. R. Neubig, C. L. Brooks III., *ACS Chem. Biol.* **2013**, *8*(12), 2778.
[39] Y. Liu, H. Vashisth, *Biophys. J.* **2021**, *120*(3), 517.
[40] S. P. Srinivasa, N. Watson, M. C. Overton, K. J. Blumer, *J. Biol. Chem.* **1998**, *273*(3), 1529.
[41] I. Rivalta, G. P. Lisi, N.-S. Snoeberger, G. Manley, J. P. Loria, V. S. Batista, *Biochemistry* **2016**, *55*(47), 6484.
[42] N. B. Senese, R. Kandasamy, K. E. Kochan, J. R. Traynor, *Front. Mol. Neurosci.* **2020**, *13*, 5.
[43] X.-Q. Yao, L. Skjærven, B. J. Grant, *J. Phys. Chem. B* **2016**, *120*(33), 8276.
[44] B. Huang, M. Schroeder, *BMC Struct. Biol.* **2006**, *6*(1), 19.
[45] S. R. Kimura, H. P. Hu, A. M. Ruvinsky, W. Sherman, A. D. Favia, *J. Chem. Inf. Model.* **2017**, *57*(6), 1388.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: W. K. B. Chan, D. DasGupta, H. A. Carlson, J. R. Traynor, *J. Comput. Chem.* **2021**, *42*(30), 2170. https://doi.org/10.1002/jcc.26747