Electronic Supporting Information

Materials and Methods

Preparation of NMR Samples. The G12C, G12S, G12D, G12V and G13D mutants were generated by site-directed mutagenesis of the wild-type human H-Ras (residues 1-166). The uniformly ^{15}N- and $^{15}N,^{13}C$-labeled wild-type and mutant Ras and unlabeled Sos (residues 564-1049) were overexpressed and purified as described previously.[1,2] The wild-type Ras and mutants were loaded with native GTP or GDP as required in individual experiments, with trace amount of Sos and excess free GTP included in the GTP-bound samples sustaining the active form.[3] The buffer for NMR samples contained 20 mM Tris (pH 7.5), 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 5 μM GDP, 0.01% NaN_3 and 10% D_2O. Residual dipolar couplings of GDP-bound G12C, G12S, and G12D were measured in two different alignment media, 17 mg/mL Pf1 phage[4] and 4.2% PEG ($C_{12}E_5$/hexanol at a molar ratio of 0.83) bicelle doped with 1/30 sodium octyl sulfate.[5] RDCs of Ras•GTP were measured in 17 mg/mL Pf1 phage and 5% PEG (C_8E_5/octanol at a molar ratio of 1.2) bicelle.

NMR Spectroscopy and Data Analysis. All NMR experiments were performed on a Bruker spectrometer (14.1 T) equipped with a cryoprobe. The high-resolution $^1H^N$ and ^{15}N chemical shifts of WT and individual mutants in both GDP- and GTP-bound forms, used for direct chemical shift comparisons (Fig. S1), were extracted from the ^1H-^{15}N HSQC spectra recorded at 15 °C with assignments confirmed using HNCA spectra. For

NMR-based ensemble modeling (eqn S3), the $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ chemical shifts and ^{1}H-^{15}N RDCs were measured using a set of standard multi-dimensional experiments including HSQC, HNCA, HN(CO)CACB, HNCO, and IPAP-HSQC recorded at 15 °C (for WT•GTP) or 22 °C (for G12C•GDP, G12S•GDP, and G12D•GDP). The one-bond ^{1}H-^{15}N splittings were measured in both weakly oriented and isotropic samples, with RDCs determined based on their differences.[6] All spectra were processed using the NMRPipe software package[7] and visualized with Sparky.[8] Individual peaks in the 2D and 3D spectra were fit to 2D and 3D Gaussian functions respectively for accurate quantification of RDCs and chemical shifts.

The amide hydrogen exchange rates of GDP-bound G12C, G12S, G12D and WT were measured using the CLEANEX-PM experiment[9] at 22 °C with an RF field of 7.2 kHz for the spin-lock sequence and mixing times (τ_m) from 2.8 to 98.5 ms. The apparent decay rate of water magnetization (R_w) during the mixing time and the fraction of water magnetization (*fw*) at the beginning of the mixing time were measured as described previously.[9,10] Reference signal intensities (I_0) were measured using an HSQC with a 12-sec recycle delay and consistent settings with the CLEANEX-PM experiment. The $k_{H\text{-}EX}$ rates were extracted from the τ_m-dependent intensities,[9–11]

$$\frac{I(\tau_m)}{I_0} = \frac{f_w k_{H-EX}\left[\exp(-R_w\tau_m) - \exp\left(-R_p\tau_m - k_{H-EX}\tau_m\right)\right]}{R_p + k_{H-EX} - R_w} \qquad \text{(S1)}$$

where R_p represents the apparent relaxation rates of protein magnetizations during τ_m.

Ensemble Construction. $E_{S2\text{•}GTP}$, $E_{G12C\text{•}GDP}$, $E_{G12S\text{•}GDP}$, and $E_{G12D\text{•}GDP}$ were constructed via NMR-guided optimization using the structural pools sampled by accelerated

molecular dynamics (aMD) simulations. The crystal structure of GTP-bound H-Ras[12] was used as the starting conformation for the aMD simulation of active Ras. The crystal structure of GDP-bound wild-type H-Ras[13] was used as the starting conformation for simulations of inactive G12C, G12S and G12D after mutating residue 12 from glycine to cysteine, serine and aspartate, respectively. The AMBER 16 software package was used for all simulations with the AMBER ff14SB force field[14] and TIP3P water model.[15] The bound GDP/GTP and Mg^{2+} were included in the simulations with the previously published force field for GDP/GTP[16] employed. The temperature was kept at 300 K using the Langevin thermostat with a collision frequency of 5.0 ps^{-1} and the pressure coupled to 1 bar using isotropic position scaling with a relaxation time of 2.0 ps. Non-bonded interactions were cut off at 10 Å, with the long-range electrostatic interactions treated by the particle mesh Ewald method.[17] Each system was solvated in a cubic box with a minimum distance of 10 Å from the protein to the edge of the box. Other details about the system setup and relaxation are fully analogous to those published previously.[18] The total potential energy ($V(\mathbf{r})$) of a system, when lower than a threshold energy (E), was added with a boost potential ($\Delta V(\mathbf{r})$) in the aMD simulations,[19]

$$\Delta V(\mathbf{r}) = \frac{\left(E - V(\mathbf{r})\right)^2}{\alpha + E - V(\mathbf{r})} \tag{S2}$$

where $\alpha = 0.2N_{\text{atoms}}$ kcal/mol and $E = \alpha + \langle V_{total} \rangle$. N_{atoms} is the number of total atoms in the system, and $\langle V_{total} \rangle$ is the average total potential energy in a short (100 ns) conventional MD simulation performed for each mutant. For each G12 mutant, before the aMD production run at 300 K under the NPT condition, a 50 ns position restrained aMD simulation was carried out, in which residues that do not show chemical shift

changes with respect to the wild-type protein were kept frozen. The 1.1 μs unrestrained aMD simulation was stored every 10 ps, with the first 0.1 μs treated as additional equilibration and the rest as the structural pool (100,000 snapshots) for each mutant. It is worth noting that, although free energy landscapes, in principle, are available from these trajectories by reweighting, the size of the current systems (166 residues) prevents their constructions owing to the high statistical noise[20–22]. Conformational ensembles (each comprising 128 conformers) that were optimized against the experimental RDC and chemical shift data were subsequently selected from the pools using the genetic algorithm (GA) with the following fitness function,[23,24]

$$\chi^2 = \sum_i \left(RDC_i^{\text{calc}} - RDC_i^{\text{expt}} \right)^2 + \sum_j \left(\delta_{j,13\text{C}}^{\text{calc}} - \delta_{j,13\text{C}}^{\text{expt}} \right)^2 + \frac{1}{4} \sum_k \left(\delta_{k,15\text{N}}^{\text{calc}} - \delta_{k,15\text{N}}^{\text{expt}} \right)^2 \qquad \text{(S3)}$$

where i represents the RDCs for individual N-H bond vectors, and j (k) represents experimental $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and $^{13}C'$ (^{15}N) chemical shifts except those within 5 Å of the bound GDP and Mg^{2+}. SPARTA+[25] was used for predicting ^{13}C and ^{15}N chemical shifts from the snapshots. The back-calculation of RDCs and the numerical scheme for optimizing χ^2 are as described previously.[24,26]

***Ten-Fold Cross Validation*.** The above described ensemble modeling was validated by randomly dividing the chemical shifts and RDCs into 10 data sets, of which one was used as the testing set and the remainder as the training set. In each of the 10 iterations, an ensemble was generated using the training set and subsequently used to predict the data in the testing set. The levels of agreement between experimental and back-calculated data for the testing sets were found to be close to those of the training sets

(Figs. S10-S13), lending support to the robustness of the modeling.

Cryptic Binding Site Analysis. MDpocket[27] was used to detect SI/II-P and SII-P for each conformer in the ensembles and calculate their volumes and proportions of apolar α spheres. The relative solvent accessibilities (*RSA*) of residue 12 in individual ensembles were calculated to evaluate its levels of exposure,

$$\langle RSA \rangle = \langle ASA \rangle / ASA_{\max} \tag{S4}$$

where $ASA_{\max}$ is the theoretical maximum accessible surface area with values of 104 Å^2, 193 Å^2, 167 Å^2, and 155 Å^2 for glycine, aspartate, cysteine, and serine, respectively.[28] The angular brackets indicate ensemble averaging.

The spatial connectivity between residue 12 and SII-P was measured using the surface area of the methylene (for G12), thiol (for C12), carboxylate (for D12), or hydroxyl (for S12) group accessible to the SII-P pocket (SII-P-ASA). Conformers, in which the corresponding groups are in direct contact with (disconnected from) SII-P, have values of SII-P-ASA larger than (equal to) zero. The boundary of the SII-P pocket in each conformer was calculated using the Fpocket program.[29]

Dihedral Angle Principal Component Analysis (PCA). The PCA was performed using the backbone dihedral angles (φ,ψ) of the three ensembles ($E_{S2\cdot GTP}$, $E_{S1\cdot GTP}$, and $E_{Ras\cdot GDP}$), represented by cos(φ), sin(φ), cos(ψ) and sin(ψ) for each residue.[30] Each conformer of the ensembles was then projected onto the 3D space spanned by the first three principal components (Fig. 1B, main text). The relative motional amplitude (*A*)

of a given residue along a principal component is obtained from its 4 corresponding elements (e_1,e_2,e_3,e_4) in the eigenvector, $A = \sqrt{\sum_{i=1}^{4} e_i^{\,2}}$, as shown in Fig. S5.

Analysis of Cross-Regional and Inter-Residue Contacts. The level of packing between the N-termini of switch II (residues 60-70) and α3 (residues 87-96) could affect the exposure of residue 12 and was denoted by the number of cross-regional atom pairs with the cutoff distance of 5 Å (Fig. 3B in the main text).

Inter-residue contact maps, which provide information on both local and long-range interactions, were calculated for individual ensembles (Figs. S7 and S8). Two residues, of which the minimum distance is less (or larger than) 5 Å, are considered to be (or not to be) in contact, with the assigned value of 1 (or 0). The value for each pair of residues was averaged over the ensemble.

Back-Calculation of Hydrogen Exchange Protection Factor. The hydrogen exchange protection factors (*PF*) of a given amide were computed from each ensemble using the approximate relationship,[31]

$$\langle \ln PF \rangle = \langle \beta_C N_C + \beta_H N_H \rangle \qquad \text{(S5)}$$

where N_C is the number of heavy atoms within 6.5 Å from the amide nitrogen, and N_H is the number of hydrogen bonds. The weights (β_C = 0.35, β_H = 2) suggested by Best and Vendruscolo[31] were used. The angular brackets indicate ensemble averaging.

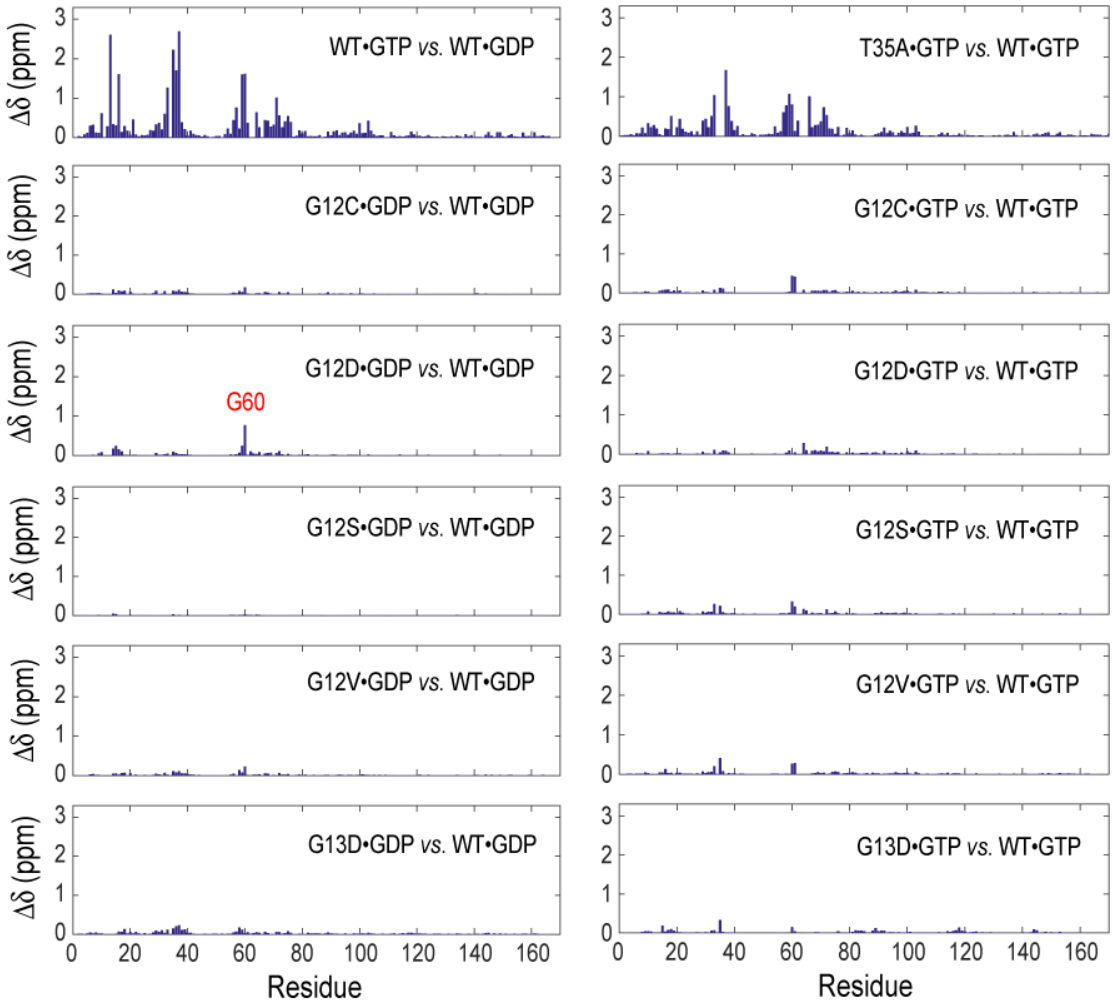

Fig. S1 Chemical shift perturbations caused by the mutations and the changes of nucleotide-bound or conformational states. Data used for comparing state 2 and state 1 (T35A•GTP) were taken from the previous work [3]. The combined chemical shift differences were calculated as $\Delta\delta=\sqrt{\Delta\delta_H{}^2+(0.14\Delta\delta_N)^2}$. The $\Delta\delta$ values for the mutated and directly adjacent residues are not shown.

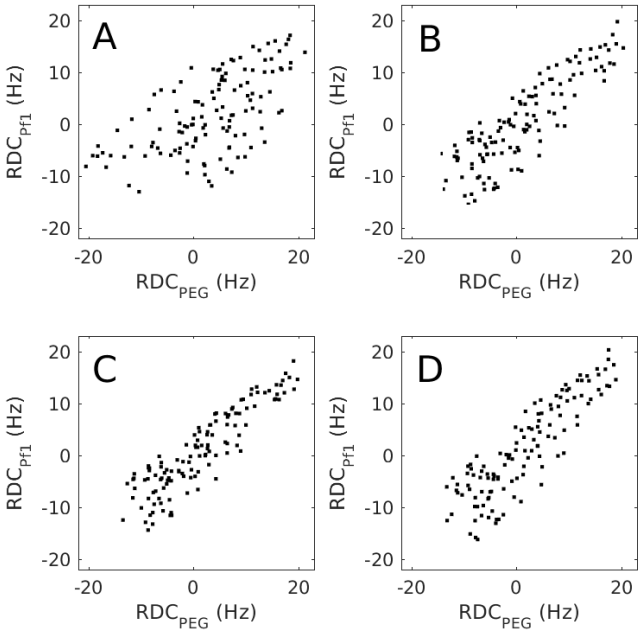

Fig. S2 Scatter plots of experimental RDCs for WT•GTP (A), G12C•GDP (B), G12D•GDP (C), and G12S•GDP (D), measured in Pf1 and PEG that produced different alignment.

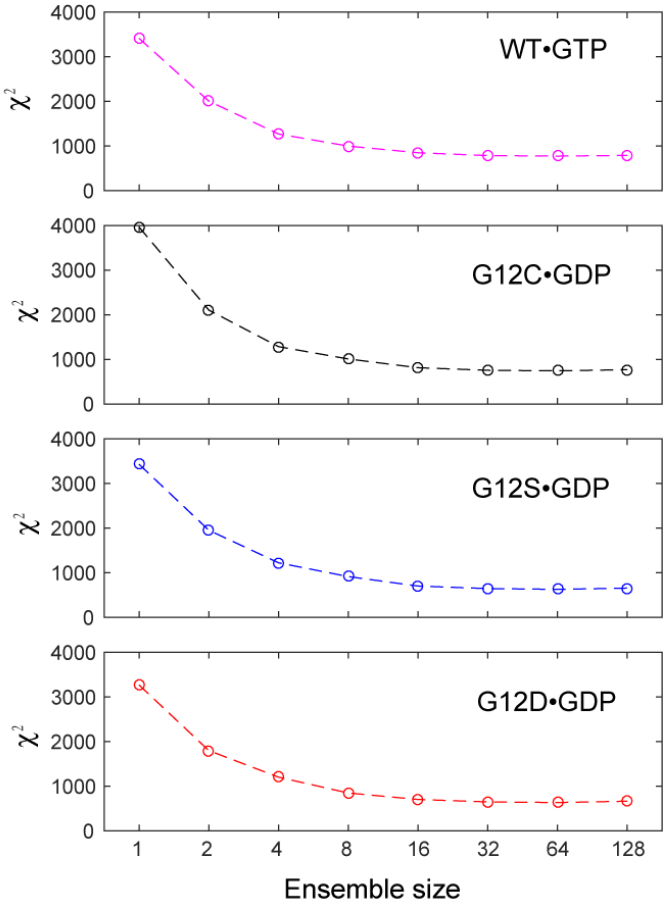

Fig. S3 The level of agreement with the experimental data (χ^2) in the GA-based refinement as a function of the ensemble size for WT•GTP (magenta), G12C•GDP (black), G12S•GDP (blue), and G12D•GDP (red).

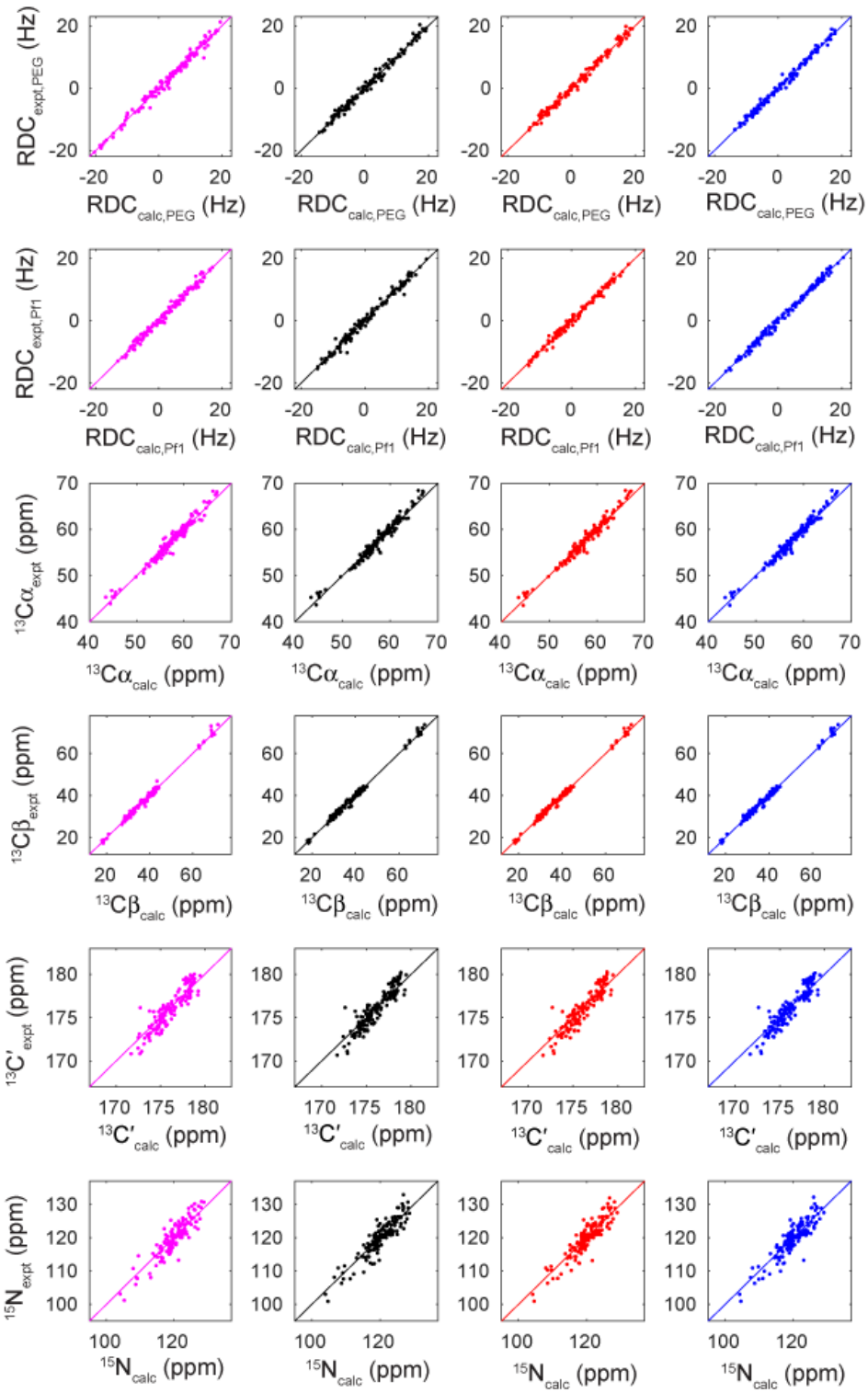

Fig. S4 Comparisons of experimental and back-calculated RDCs and chemical shifts for E_{S2}•GTP (magenta '.'), E_{G12C}•GDP (black '.'), E_{G12D}•GDP (red '.'), and E_{G12S}•GDP (blue '.'). The corresponding RDC *Q* values and chemical shift root-mean-square deviations (RMSDs) are listed in Table S1.

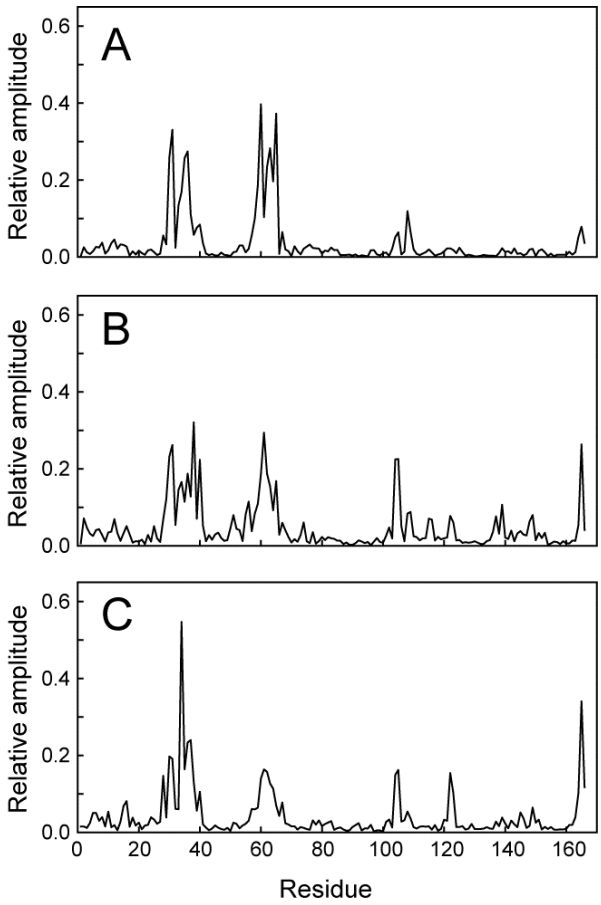

Fig. S5 The relative motional amplitudes of individual residues along the first 3 dihedral angle principal components – PC 1 (A), PC 2 (B) and PC 3 (C).

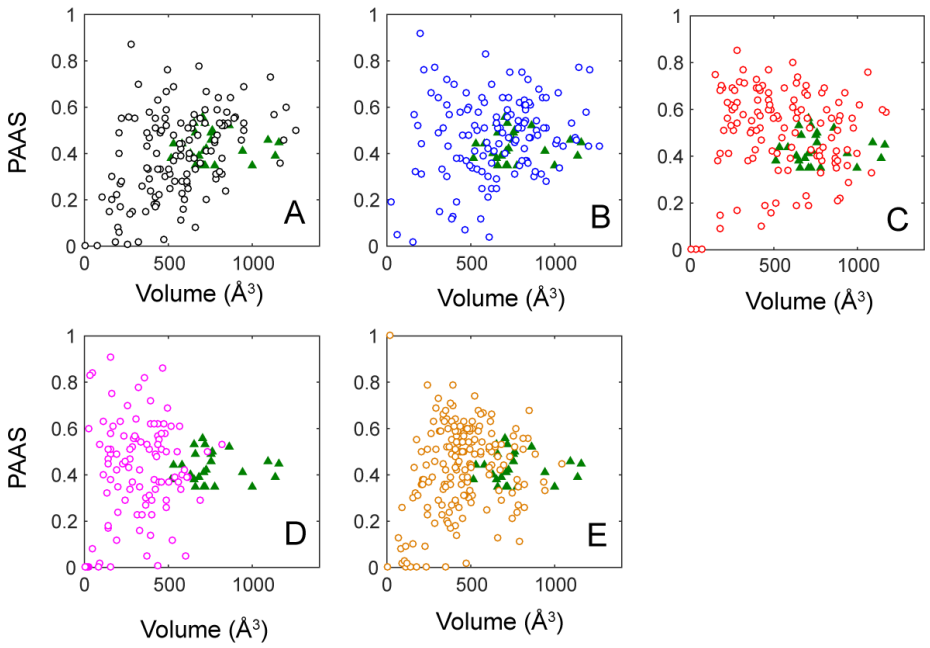

Fig. S6 Distributions of the pocket volumes and proportions of apolar α spheres (PAAS) of SII-P in $E_{G12C \cdot GDP}$ (A), $E_{G12S \cdot GDP}$ (B), $E_{G12D \cdot GDP}$ (C), $E_{S2 \cdot GTP}$ (D), and $E_{S1 \cdot GTP}$ (E), in comparison with those in the co-crystal structures of Ras in complex with SII-P inhibitors (green triangles; PDB codes: 4LUC, 4M1S, 4M1Y, 4LV6, 4LYF, 4LYJ, 4M1O, 4M1T, 4M1W, 4M22, 5F2E, 5V9O, 5V9U, 6PGP, 6TAM, 6USX, 6UT0, 5V9L, 6N2J, 6N2K, 6P8W, 6P8X, 6P8Z, 6TAN, 6USZ).

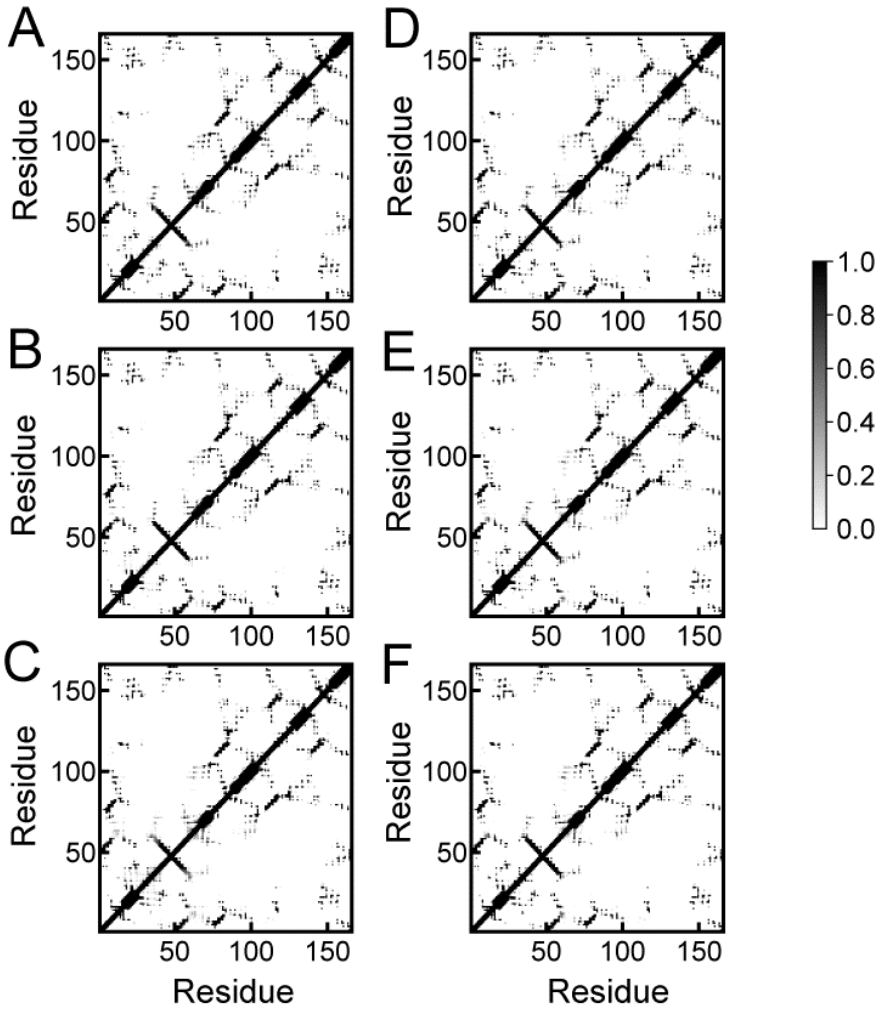

Fig. S7 Inter-residue contact maps, $C_{S2•GTP}$ (A), $C_{S1•GTP}$ (B), $C_{Ras•GDP}$ (C), $C_{G12D•GDP}$ (D), $C_{G12S•GDP}$ (E) and $C_{G12C•GDP}$ (F), calculated for the ensembles $E_{S2•GTP}$, $E_{S1•GTP}$, $E_{Ras•GDP}$, $E_{G12D•GDP}$, $E_{G12S•GDP}$ and $E_{G12C•GDP}$, respectively. The overall patterns of inter-residue contacts are highly conserved across the states and mutants.

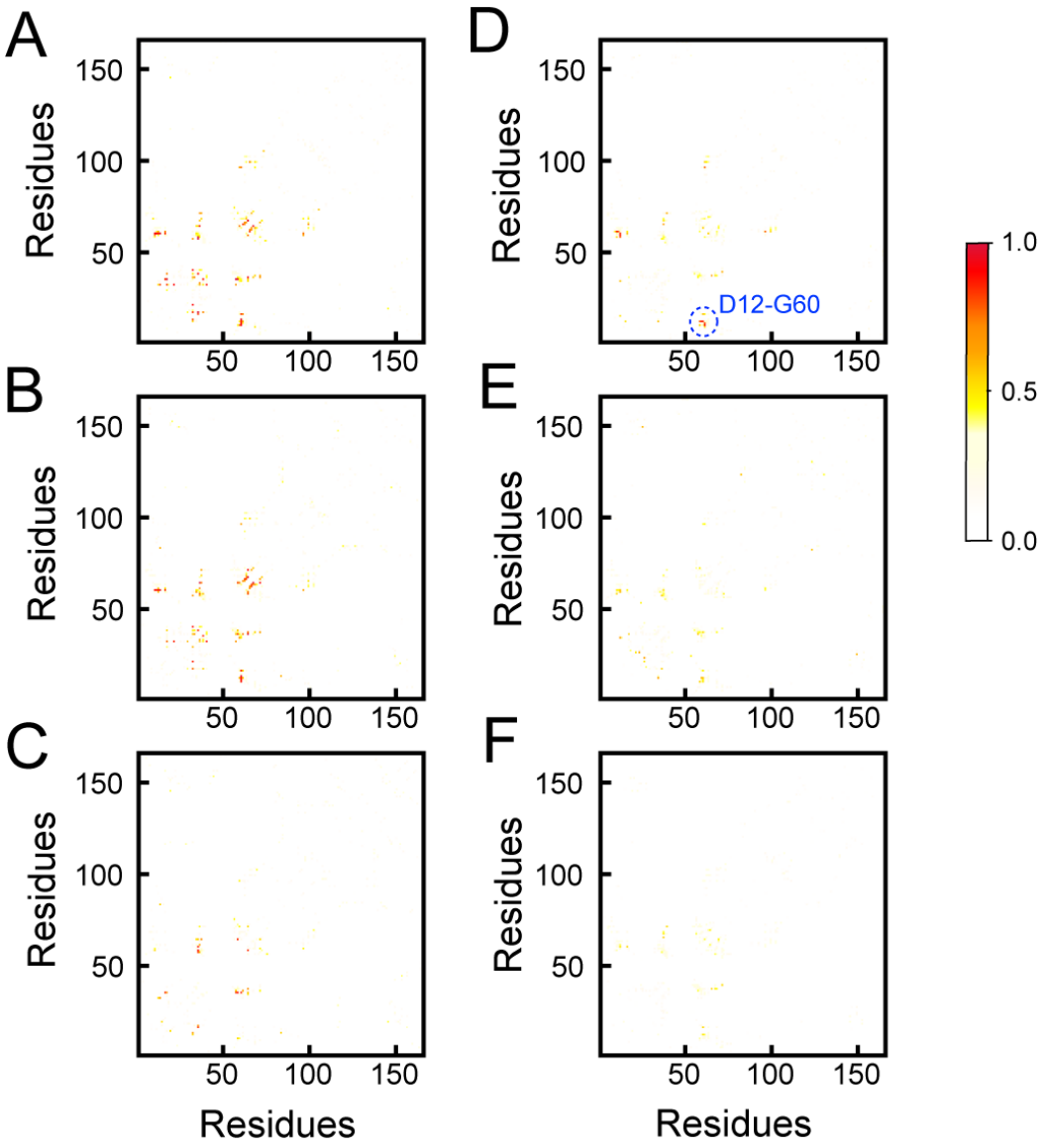

Fig. S8 The contact map differences, $|C_{S2\bullet GTP}-C_{Ras\bullet GDP}|$ (A), $|C_{S1\bullet GTP}-C_{Ras\bullet GDP}|$ (B), $|C_{S2\bullet GTP}-C_{S1\bullet GTP}|$ (C), $|C_{G12D\bullet GDP}-C_{Ras\bullet GDP}|$ (D), $|C_{G12C\bullet GDP}-C_{Ras\bullet GDP}|$ (E), and $|C_{G12S\bullet GDP}-C_{Ras\bullet GDP}|$ (F). The inter-residue contact differences caused by G12 mutations are much smaller than those by change of the states. The most obvious mutation-induced change is observed at the D12-G60 residue pair in G12D•GDP (D).

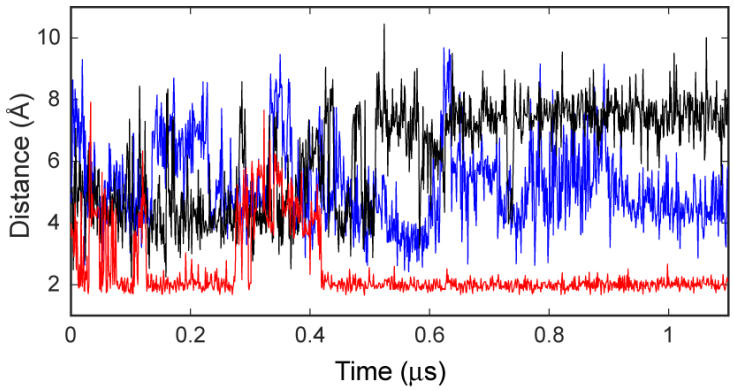

Fig. S9 The minimum distances between the NH group of G60 and the sidechain of residue 12 in the aMD simulations of G12C•GDP (black), G12S•GDP (blue), and G12D•GDP (red).

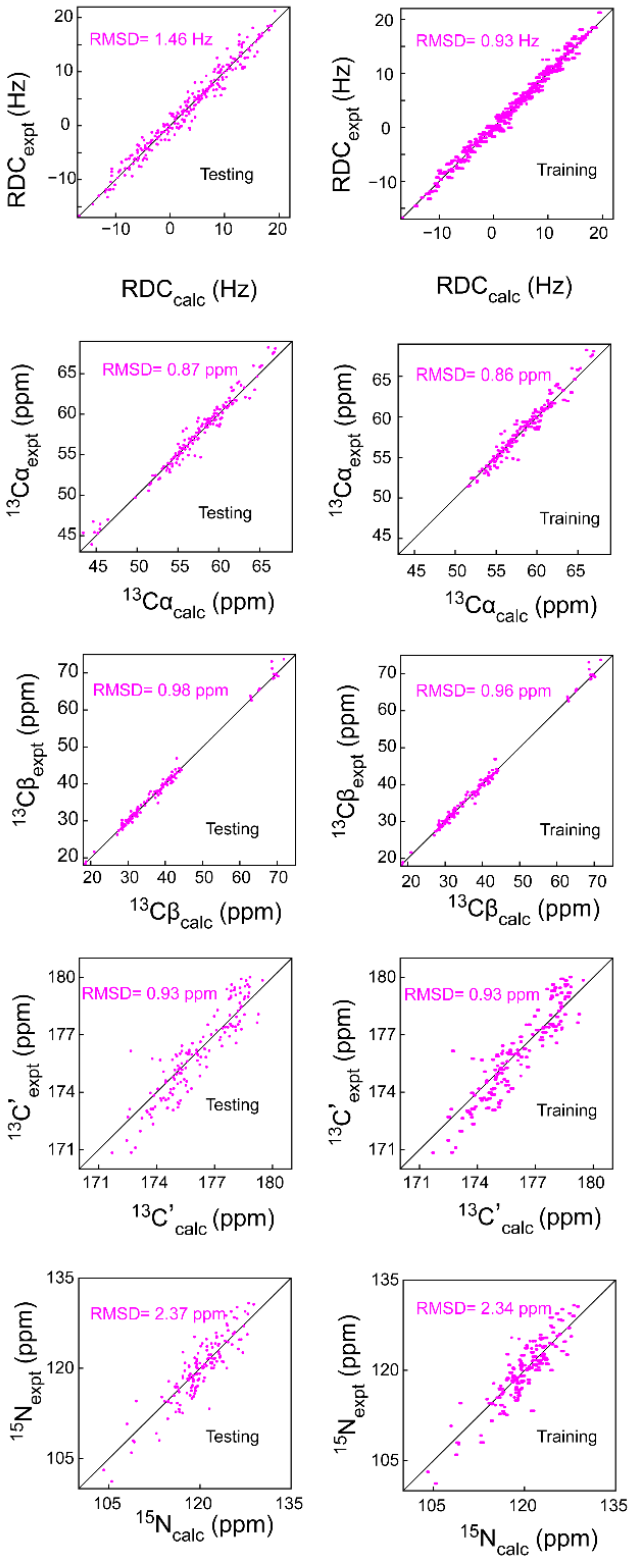

Fig. S10 Comparisons of experimental and calculated data of Ras•GTP for the testing and training sets in the 10-fold cross validation.

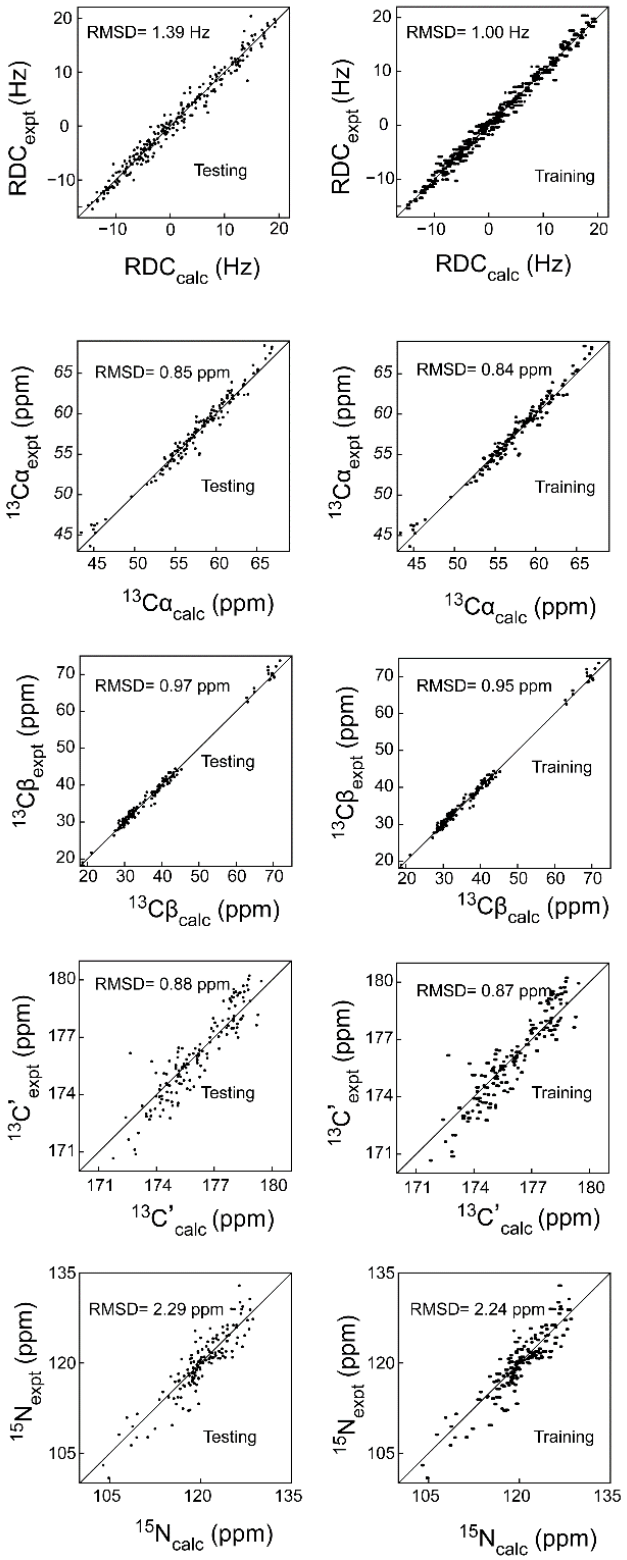

Fig. S11 Comparisons of experimental and calculated data of G12C•GDP for the testing and training sets in the 10-fold cross validation.

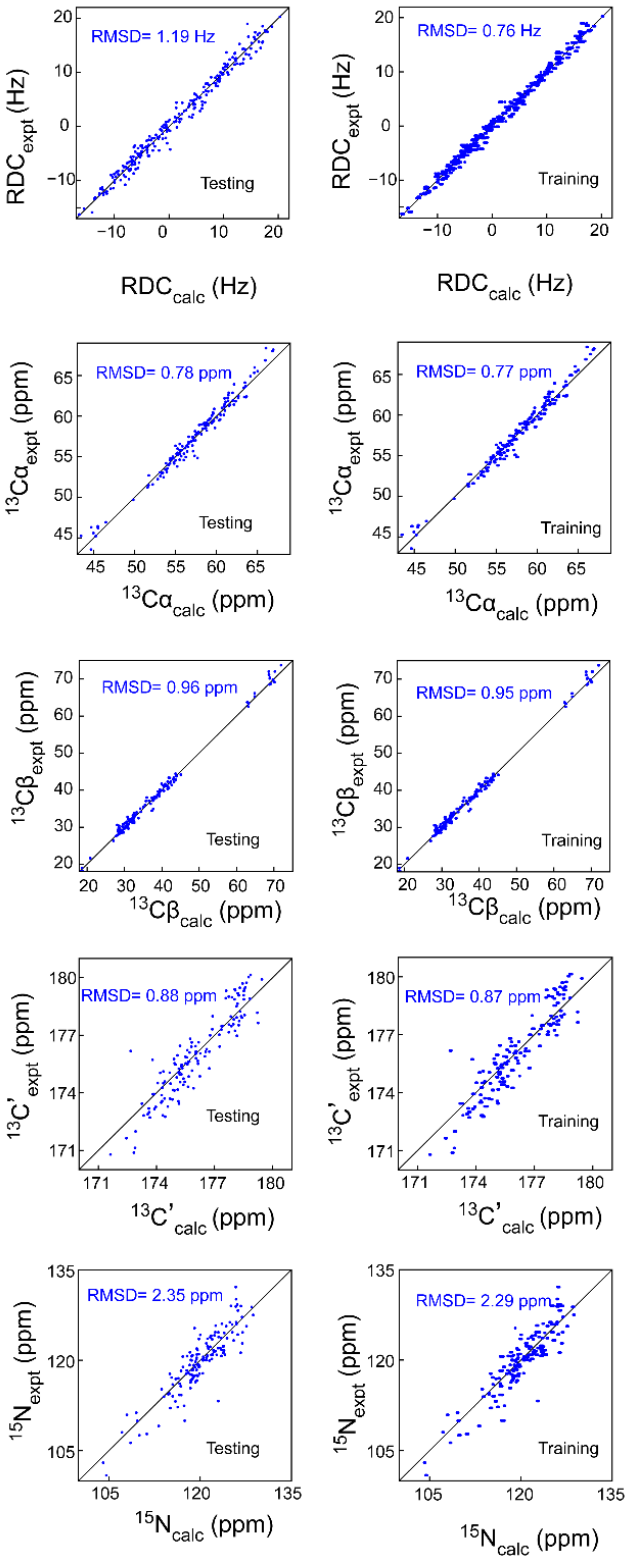

Fig. S12 Comparisons of experimental and calculated data of G12S•GDP for the testing and training sets in the 10-fold cross validation.

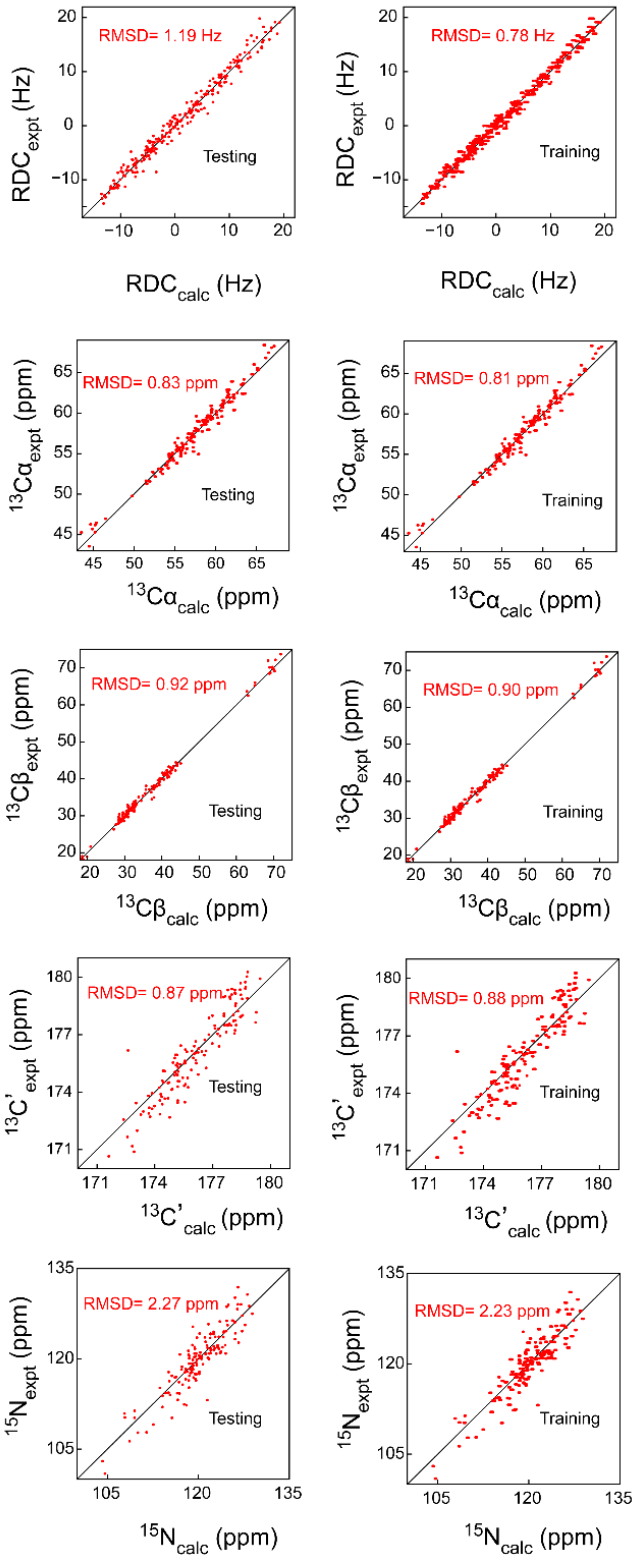

Fig. S13 Comparisons of experimental and calculated data of G12D•GDP for the testing and training sets in the 10-fold cross validation.

Examining the hydrogen exchange protection in K-Ras(G12D)•GDP

As a proof of concept, the catalytic domains (G-domain) of the H-Ras isoform was used as a representative for examining the state and allele specificity in this work. As the G-domains, harboring the cryptic pockets for direct inhibitors, are nearly identical among Ras isoforms, the results obtained here may provide general insights for other isoforms as well. To check whether the largest mutational effect on the SII-P accessibility we observed in H-Ras also exist in K-Ras, we measured the $k_{H\text{-}EX}$ of K-Ras(WT)•GDP and K-Ras(G12D)•GDP using consistent experimental conditions with those for the H-Ras samples, followed by calculation of the ρ values. The almost identical hydrogen exchange protection pattern with that of H-Ras shows that the unique (G60)N-H···O(D12) hydrogen bond is well conserved (Fig. S14).

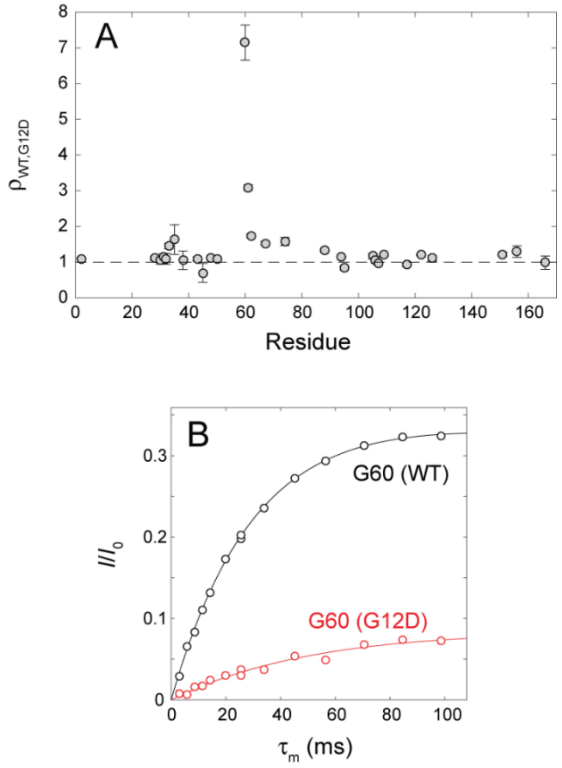

Fig. S14 Comparison of the $k_{H\text{-}EX}$ rates of K-Ras•GDP with those of K-Ras(G12D)•GDP. (A) The ratios (ρ) of $k_{H\text{-}EX}$ of K-Ras•GDP to those of

K-Ras(G12D)•GDP, plotted against the residue number. (B) Intensities of the G60 resonance as a function of τ_m for K-Ras•GDP (black) and K-Ras(G12D)•GDP (red).

Table S1. Degree of agreement between the experimental data and those back-calculated from the optimized ensembles/original pools of G12C•GDP, G12D•GDP, and G12S•GDP and WT•GTP (state2).

	G12C•GDP	G12D•GDP	G12S•GDP	WT•GTP
$Q_{RDC,PEG}$	0.11 (0.34)	0.11 (0.33)	0.10 (0.32)	0.11 (0.28)
$Q_{RDC,Pf1}$	0.13 (0.34)	0.10 (0.30)	0.08 (0.28)	0.12 (0.25)
$^{13}C^{\alpha}$ RMSD (ppm)	0.84 (0.87)	0.82 (0.87)	0.78 (0.86)	0.86 (0.91)
$^{13}C^{\beta}$ RMSD (ppm)	0.95 (0.99)	0.91 (0.93)	0.98 (1.01)	0.96 (1.00)
$^{13}C'$ RMSD (ppm)	0.88 (0.90)	0.87 (0.89)	0.86 (0.91)	0.93 (0.96)
^{15}N RMSD (ppm)	2.25 (2.31)	2.24 (2.41)	2.32 (2.49)	2.34 (2.48)

Numbers in and out of the parentheses correspond to the values for the original aMD pools and optimized ensembles, respectively.

References

1 M. J. Smith, B. G. Neel and M. Ikura, NMR-based functional profiling of RASopathies and oncogenic RAS mutations, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 4574–4579.

2 H. Sondermann, S. M. Soisson, S. Boykevisch, S. S. Yang, D. Bar-Sagi and J. Kuriyan, Structural analysis of autoinhibition in the Ras activator son of sevenless, *Cell*, 2004, **119**, 393–405.

3 X. Chen, H. Yao, H. Wang, Y. Mao, D. Liu and D. Long, Extending the lifetime of native GTP-Bound Ras for site-resolved NMR measurements: quantifying the allosteric dynamics, *Angew. Chemie - Int. Ed.*, 2019, **58**, 2730–2733.

4 M. R. Hansen, L. Mueller and A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol.*, 1998, **5**, 1065–1074.

5 M. Rückert and G. Otting, Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, *J. Am. Chem. Soc.*, 2000, **122**, 7793–7797.

6 M. Ottiger, F. Delaglio and A. Bax, Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, *J. Magn. Reson.*, 1998, **131**, 373–378.

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

8 T. D. Goddard and D. G. Kneller, *SPARKY 3*, University of California, San Francisco.

9 T. L. Hwang, P. C. van Zijl and S. Mori, Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme., *J. Biomol. NMR*, 1998, **11**, 221–226.

10 H. Wang, S. Wang, C. Li, H. Li, Y. Mao, W. Liu, C. Xu and D. Long, Probing Transient Release of Membrane-Sequestered Tyrosine-Based Signaling Motif by Solution NMR Spectroscopy, *J. Phys. Chem. Lett.*, 2017, **8**, 3765–3769.

11 J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, Investigation of exchange processes by two‐dimensional NMR spectroscopy, *J. Chem. Phys.*, 1979, **71**, 4546–4553.

12 A. J. Scheidig, C. Burmester and R. S. Goody, The pre-hydrolysis state of p21(ras) in complex with GTP: New insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins, *Structure*, 1999, **7**, 1311–1324.

13 M. V Milburn, L. Tong, A. M. DeVos, A. Brunger, Z. Yamaizumi, S. Nishimura and S. H. Kim, Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins, *Science*, 1990, **247**, 939–945.

14 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theory Comput*, 2015, **11**, 3696–3713.

15 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.*, 1983, **79**, 926–935.

16 K. L. Meagher, L. T. Redman and H. A. Carlson, Development of polyphosphate parameters for use with the AMBER force field, *J. Comput. Chem.*, 2003, **24**, 1016–1025.

17 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

18 L. C. T. Pierce, R. Salomon-Ferrer, C. Augusto F. De Oliveira, J. A. McCammon and R. C. Walker, Routine access to millisecond time scale events with accelerated molecular dynamics,

J. Chem. Theory Comput., 2012, **8**, 2997–3002.

19 D. Hamelberg, J. Mongan and J. A. McCammon, Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules, *J. Chem. Phys.*, 2004, **120**, 11919–11929.

20 P. R. L. Markwick and J. A. Mccammon, Studying functional dynamics in bio-molecules using accelerated molecular dynamics, *Phys. Chem. Chem. Phys*, 2011, **13**, 20053–20065.

21 Y. Miao and J. A. Mccammon, Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications, *Annu. Rep. Comput. Chem.*, 2017, **13**, 231–278.

22 Y. Miao, W. Sinko, L. Pierce, D. Bucher, R. C. Walker and J. A. McCammon, Improved reweighting of accelerated molecular dynamics simulations for free energy calculation, *J. Chem. Theory Comput.*, 2014, **10**, 2677–2689.

23 G. Nodet, L. Salmon, V. Ozenne, S. Meier, M. R. Jensen and M. Blackledge, Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings, *J. Am. Chem. Soc.*, 2009, **131**, 17908–17918.

24 D. Liu, Y. Mao, X. Gu, Y. Zhou and D. Long, Unveiling the 'invisible' druggable conformations of GDP-bound inactive Ras, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2024725118.

25 Y. Shen and A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, *J. Biomol. NMR*, 2010, **48**, 13–22.

26 D. Liu, X. Chen and D. Long, NMR-Derived Conformational Ensemble of State 1 of Activated Ras Reveals Insights into a Druggable Pocket, *J. Phys. Chem. Lett.*, 2020, **11**, 3642–3646.

27 P. Schmidtke, A. Bidon-chanal, F. J. Luque and X. Barril, MDpocket: Open-source cavity

detection and characterization on molecular dynamics trajectories, *Bioinformatics*, 2011, **27**, 3276–3285.

28 M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman and C. O. Wilke, Maximum allowed solvent accessibilites of residues in proteins, *PLoS One*, 2013, **8**, e80635.

29 V. Le Guilloux, P. Schmidtke and P. Tuffery, Fpocket: An open source platform for ligand pocket detection, *BMC Bioinformatics*, 2009, **10**, 168.

30 A. Altis, P. H. Nguyen, R. Hegger and G. Stock, Dihedral angle principal component analysis of molecular dynamics simulations, *J. Chem. Phys.*, 2007, **126**, 244111.

31 R. B. Best and M. Vendruscolo, Structural interpretation of hydrogen exchange protection factors in proteins: characterization of the native state fluctuations of CI2, *Structure*, 2006, **14**, 97–106.