An expanded classification of active, inactive, and druggable RAS conformations

Mitchell I. Parker[1,2], Joshua E. Meyer[1,3], Erica A. Golemis[1,4], Roland L. Dunbrack, Jr.[1]

[1]*Program in Molecular Therapeutics, Fox Chase Cancer Center, Philadelphia, PA 19111, USA*
[2]*Molecular and Cell Biology and Genetics (MCBG) Program, Drexel University College of Medicine, Philadelphia, PA 19102, USA*
[3]*Department of Radiation Oncology, Fox Chase Cancer Center, Philadelphia, PA 19111, USA*
[4]*Department of Cancer and Cell Biology, Lewis Katz School of Medicine, Philadelphia, PA 19140*, USA
e-mail: Roland.dunbrack@fccc.edu

RAS (KRAS, NRAS, and HRAS) proteins have widespread command of cellular circuitry and are high-priority drug targets in cancers and other diseases. Effectively targeting RAS proteins requires an exact understanding of their active, inactive, and druggable conformations, and the structural impact of mutations. Here we define an expanded classification of RAS conformations by clustering all 699 available human KRAS, NRAS, and HRAS structures in the Protein Data Bank (PDB) by the arrangement of their catalytic switch 1 (SW1) and switch 2 (SW2) loops. This enabled us to clearly define the geometry of closely related RAS conformations, many of which were not previously described. We determined the catalytic impact of the most common RAS mutations and identified several novel druggable RAS conformations. Our study expands the topography of characterized RAS conformations and will help inform future structure-guided RAS drug design.

Mutations in the RAS isoforms, KRAS, NRAS, and HRAS, drive oncogenesis in ~20% of human cancers, and cause a variety of tumor predisposition syndromes, making these proteins high-priority therapeutic targets[1]. Over the past 30 years, our molecular understanding of RAS mutations and our ability to drug these proteins has considerably improved, owing, in part, to hundreds of structural studies examining wild-type (WT) and mutated RAS in complex with various signaling effector and regulatory proteins, or with small molecule and designed protein inhibitors (reviewed in ref. [2]). However, our structural understanding of RAS mutations is incomplete, and, except for KRAS G12C, G12D, and G13C, all mutated RAS forms have not yet been selectively targeted by therapeutics[3,4].

KRAS, NRAS, and HRAS are conformational switches that modulate growth (and other) signaling pathways by transitioning between active, GTP-bound and inactive, GDP-bound states[5]. In normal tissues, RAS conformational cycling is tightly regulated by the catalytic (CDC25) domain of guanine exchange factors (GEFs; e.g., SOS1), which remove GDP allowing subsequent GTP rebinding[6], and GTPase activating proteins (GAPs; e.g., NF1), which catalyze the otherwise slow intrinsic rate of GTP hydrolysis to GDP[7]. GEFs and GAPs bind to the conformationally dynamic RAS switch 1 (SW1) and switch 2 (SW2) loops, which also provide binding sites for signal effector proteins[4] (e.g., RAF1) and direct RAS inhibitors[3,8]. RAS targeted therapies mainly bind to an SW1/SW2 pocket (SP12) to block RAS protein interactions or an SW2 pocket (SP2) to lock RAS in an inactive, GDP-bound conformation. Overall, the conformations of SW1 and SW2 are essential to RAS function and the druggability of these proteins.

Most tumor-associated RAS mutations modify SW1 and SW2 conformational preferences in ways that reduce the rate of intrinsic and GAP-mediated hydrolysis (residues 12, 13, and 61) and/or enhance the rate of GEF-mediated exchange (residues 13, 61, and 146)[9-13]. The net effect of these mutations is to increase the steady state cellular concentration of active, GTP-bound RAS that is capable of stimulating signaling pathways via the following mechanisms: (1) by binding to and activating RAS effectors[14]; (2) by binding to the allosteric (REM) domain of the GEF SOS1, which functions to accelerate GDP release at the CDC25 domain of SOS1[15]; and (3) by promoting homodimerization of RAS monomers at their helices α4 and α5, which is required for activation of certain dimeric effectors, such as RAF1[16-19]. However, the exact SW1 and SW2 conformations that can form each

RAS complex and their potential druggability are unknown, making it difficult to design therapeutics that block the activities of WT and mutated RAS proteins.

While all published RAS structures have been made publicly available through the Protein Data Bank (PDB)[20,21], this structural dataset has not been leveraged in a comprehensive way to improve our understanding of RAS conformations and inform RAS drug discovery. Therefore, we analyzed all 699 available human KRAS, NRAS, and HRAS structures in the PDB to define a more comprehensive classification of active, inactive, and druggable RAS conformations. We first annotated the molecular contents of each RAS structure, including their mutation status, nucleotide state and bound protein (e.g., effector, GAP, GEF) or inhibitor site (e.g., SP12, SP2). Second, we conformationally clustered these structures based on the arrangement of their SW1 and SW2 loops to create a biologically and therapeutically informed map of the RAS conformational landscape. Overall, our study defines an expanded RAS conformational classification and provides a valuable resource for analyzing RAS structures in ways that will further improve understanding of RAS mutations and the ability to drug these proteins. We have created a web database presenting our analysis of RAS structures in the PDB, which includes a page for conformationally classifying user inputted structures (http://dunbrack.fccc.edu/rascore/).

Results

Preparing RAS structures from the PDB. We identified 699 human KRAS (N=421), HRAS (N=268), and NRAS (N=10) structures from 392 PDB entries (some entries contain multiple copies of the structure, sometimes in different conformations). Subsequently, we created an automated system for annotating RAS structures by their molecular contents, including their mutation status, nucleotide state ("3P" for GTP or any triphosphate analog, "2P" for GDP, and "0P" for nucleotide-free), bound proteins (effector, GAP, GEF CDC25 and REM domains, designed protein "binders" such as an Affimer[22] or DARPin[23,24], nanodiscs, and others), small molecule inhibitor sites (SP12, SP2, and others), and whether the α4α5 homodimer is present in the protein crystal (only X-ray structures). In following subsections, we identify the conformations of SW1 and SW2 in the prepared RAS structures and associate these conformations with the described molecular contents to define an expanded RAS conformational classification (Fig. 1a).

Clustering RAS Conformations. Several RAS conformations have been previously described based on the arrangement of SW1 (residues 25-40) and SW2 (residues 56-76). For SW1, the conformations are named by their nucleotide state and include: the "canonical" GDP-bound[25,26], nucleotide-free[6,15], and GTP-bound "state 1" and "state 2" (Fig. 1b). State 2 is considered the active conformation of RAS, since it has a higher binding affinity for signaling effectors than state 1[27-33]. For SW2, two GTP-bound conformations have been characterized, including an active, "R" state" and an inactive, "T" state"[32-34] (reidentified in ref. [35] and called state 2*) (Fig. 1c). Furthermore, other "unnamed" GDP-bound SW2 conformations have been differentiated based on their ability to bind certain small molecule inhibitors[36-39]. However, only a few RAS structures in the PDB have been classified into the previously named SW1 and SW2 conformations, and there is no systematic method for differentiating these conformations from each other and from potentially unidentified RAS conformations.

Prior studies clustered RAS structures in the PDB (121 entries at most) and only identified the four SW1 conformations, likely because they used principal component analysis (PCA) with a distance metric that integrates atomic distances across the entire protein[40,41]. Hence, we took a more sensitive approach by separately clustering the arrangements of SW1 and SW2 based on their backbone dihedral angle values: φ (phi), ψ (psi), and ω (omega). Across RAS structures in the PDB, SW1 and SW2 displayed the most dihedral variability on the Ramachandran map[42] (φ versus ψ plot) (Supplementary Fig. 1), and were the most conformationally diverse loops when structures were visualized altogether and by nucleotide state (Fig. 1d-g).

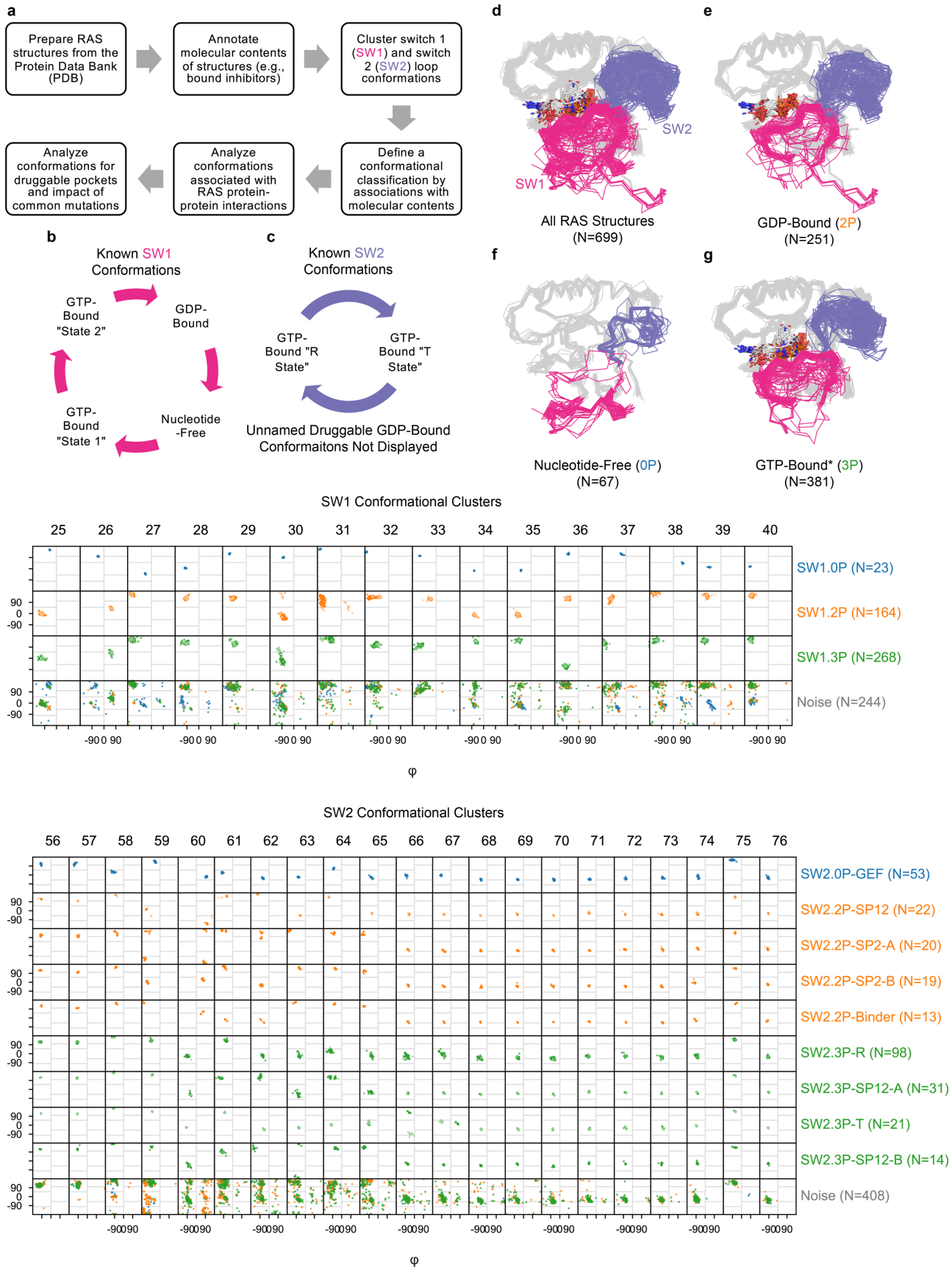

a
Prepare RAS structures from the Protein Data Bank (PDB)
Annotate molecular contents of structures (e.g., bound inhibitors)
Cluster switch 1 (SW1) and switch 2 (SW2) loop conformations
Define a conformational classification by associations with molecular contents
Analyze conformations associated with RAS protein-protein interactions
Analyze conformations for druggable pockets and impact of common mutations
b
Known SW1 Conformations
GTP-Bound "State 2"
GDP-Bound
GTP-Bound "State 1"
Nucleotide-Free
c
Known SW2 Conformations
GTP-Bound "R State"
GTP-Bound "T State"
Unnamed Druggable GDP-Bound Conformaitons Not Displayed
d
SW2
SW1
All RAS Structures (N=699)
e
GDP-Bound (2P) (N=251)
f
Nucleotide-Free (0P) (N=67)
g
GTP-Bound* (3P) (N=381)
h
SW1 Conformational Clusters
25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
SW1.0P (N=23)
SW1.2P (N=164)
SW1.3P (N=268)
Noise (N=244)
ψ
φ
i
SW2 Conformational Clusters
56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76
SW2.0P-GEF (N=53)
SW2.2P-SP12 (N=22)
SW2.2P-SP2-A (N=20)
SW2.2P-SP2-B (N=19)
SW2.2P-Binder (N=13)
SW2.3P-R (N=98)
SW2.3P-SP12-A (N=31)
SW2.3P-T (N=21)
SW2.3P-SP12-B (N=14)
Noise (N=408)
ψ
φ

Fig. 1 | Conformational clustering of RAS switch 1 (SW1) and switch 2 (SW2) loop structures across the Protein Data Bank (PDB). **a**, Study outline. **b**, Previously named switch 1 (SW1) and, **c**, switch 2 (SW2) conformations. **d**, Visualization of all human RAS (HRAS, KRAS, and NRAS) structures in the Protein Data Bank (PDB) separated by bound nucleotide: **e**, GDP-bound ("2P"), **f**, nucleotide-free ("0P"), and, **g,** GTP-bound ("3P", * for GTP and GTP analogs). Ramachandran maps (φ versus ψ backbone dihedrals) for, **h**, SW1 and, **i**, SW2 conformational clusters in the PDB. Points on Ramachandran maps corresponding to loop structures with slight backbone dihedral variation from the remainder of the cluster are displayed with a lighter hue.

After removing SW1 and SW2 loops with incomplete modeling or poor electron density, we arrived at 487 SW1 (70.0% of 699 structures) and 412 SW2 (58.9% of 699 structures) loops for conformational clustering (Supplementary Fig. 2, column 2). In our analysis, we used the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm[43] with a distance metric that locates the maximum backbone dihedral difference upon pairwise comparison of loop residues (previously implemented in refs. [44-47] for other proteins). DBSCAN finds major clusters and removes outliers (labeling them as "noise"). We first separated RAS structures by their nucleotide state (0P, 2P, and 3P) and subsequently clustered the conformations of SW1 and SW2 for each nucleotide state using DBSCAN. We then assigned a small number of poorly or incompletely modeled loops to the clusters obtained from DBSCAN by using a nearest neighbors (NN) approach (Supplementary Fig. 2).

The results for the SW1 and SW2 conformational clustering (with NN assignments added in) are displayed as Ramachandran maps per residue of each cluster compared to noise in Fig. 1h, i. We identified three SW1 and nine SW2 conformations, each of which was found across multiple RAS isoforms, PDB entries, and crystal forms (CFs; entries with the same space group and very similar unit cell dimensions and angles) (Supplementary Table 1, including the mean dihedral distance and loop α carbon atom root-mean-square deviation for each conformation). Overall, we were able to conformationally cluster 81% (N=395 out of 487) of SW1 and 58.8% (N=242 of 412) of SW2 loops that passed the completeness and electron density checks (Supplementary Fig. 2, columns 4 versus 3).

SW1 and SW2 Conformational Clusters. For clarity and brevity in our classification, we named each SW1 and SW2 conformational cluster by its loop name and nucleotide state and then added further designations as needed (Supplementary Fig. 3). The correlation of the SW1 and SW2 conformations to each other are provided in Table 1 and reported throughout the text.

Table 1 | Correlation of SW1 and SW2 conformational clusters.

Cluster Name		SW1.0P	Noise	SW1.2P	Noise	SW1.3P				All
						WaterHB	DirectHB	NoHB	Noise	
SW2.0P	**GEF**	23	30							53
	Noise		14							14
SW2.2P	**SP12**			22						22
	SP2-A			10	10					20
	SP2-B			19						19
	Binder			12	1					13
	Noise			101	76					177
SW2.3P	**R**					62	22	10	4	98
	SP12-A					29	1		1	31
	SP12-B					11	1	1	1	14
	T						20	1		21
	Noise					53	35	22	107	217
All		23	44	164	87	155	79	34	113	699

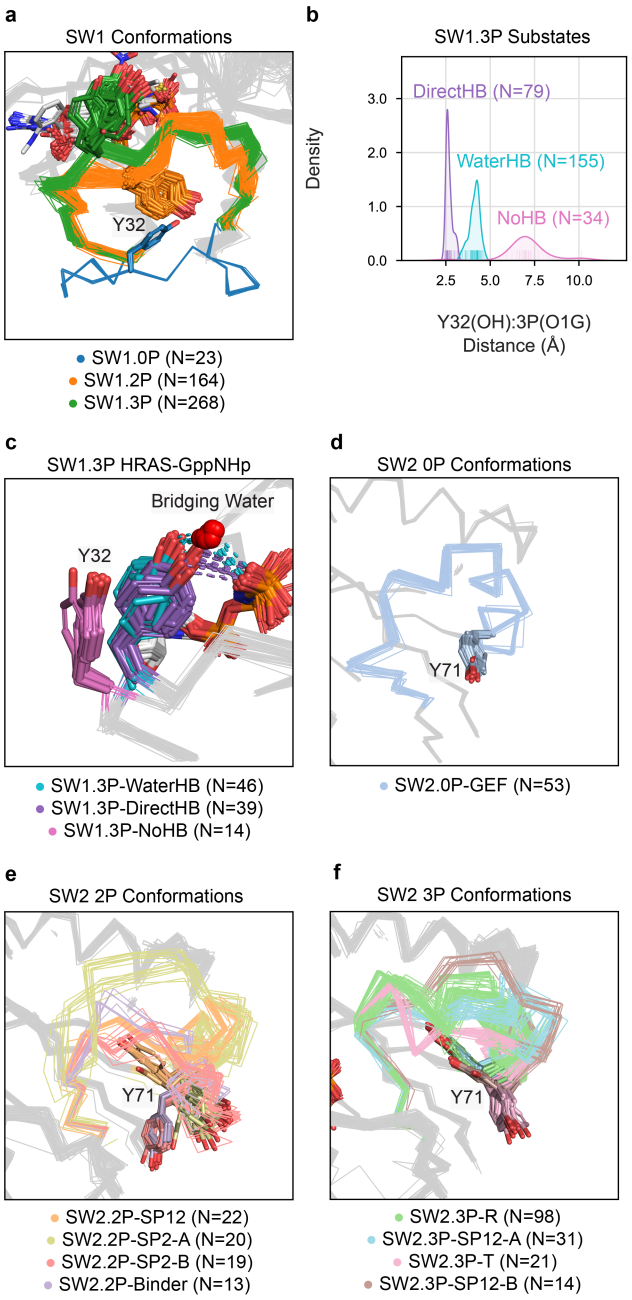


Fig. 2 | SW1 and SW2 conformational clusters. a, SW1 conformations. In the SW1.0P conformation, the central Y32 residue in SW1 is ~12-13 Å from the active site. In the SW1.2P conformation, SW1 is "closed" and interacts with the nucleotide through the backbone atoms of residues 28-32. In the GTP-bound conformation, SW1.3P, further interactions are made with the nucleotide involving the side chains of residues Y32 and T35. **b**, Y32(OH):3P(O1G) distance distribution for SW1.3P substates. **c**, SW1.3P substates within HRAS bound to GppNHp. SW2 conformations within, **d**, 0P, **e**, 2P, and, **f**, 3P states. In the SW2.0P-GEF conformation, residues 58-60 of SW2 are pulled towards the nucleotide site, and the side chains of residues Q61 and Y71 form an intra-SW2 hydrogen bond (not displayed), which is not seen in other SW2 conformations. In all SW2.2P conformations, except for SW.2P-SP12, Y71 is exposed to the solvent; the opposite trend is observed in all SW2.3P conformations where Y71 is buried in the hydrophobic core of the protein, except in SW2.3P-T where it is exposed.

For the SW1 conformations, there was a one-to-one correspondence with the nucleotide state, and we, therefore, labeled these conformations SW1.0P, SW1.2P, and SW1.3P (Fig. 1h, Supplementary Fig. 3a, and Supplementary Table 1). These SW1 conformations are visualized in Fig. 2a and can be differentiated by position of residue Y32 in SW1, which was the original method for classifying these conformations[2]. Two SW1 conformations were removed from our clustering based on their infrequency of occurrence in the PDB, but are displayed in Supplementary Fig. 4; these include structures labeled by some authors as the GTP-bound "state 1" [27,29,30] (N=6) and a "non-canonical" GDP-bound "β' or Mg-free"[12,48] (N=4). Further, we do not specify whether SW1.3P is GTP-bound state 1 or state 2, since multiple structures in this cluster have been called by both of these conformations in the literature (namely 5P21, 1CTQ, 1JAH, and 1WQ1; see refs. [25,33,49]).

Previously, hydrolytically-relevant substates of GTP-bound RAS have been described based on differences in hydrogen-bonding (HB) of the hydroxyl (OH) atom of Y32 with one of the γ-phosphate oxygen (O1G) atoms of GTP: direct (hydrolytically incompetent), water-mediated (prefers intrinsic hydrolysis), and absent (prefers GAP-mediated hydrolysis)[33,50,51]. Examination of the distribution of distances between the Y32(OH) atom and the closest 3P(O1G) atom across RAS structures in the GTP-bound cluster (SW1.3P) revealed three peaks at distances of 3, 4.5, and 7 Å (Fig. 2b); we associated these peaks with the GTP-bound substates, naming them SW1.3P-Direct, SW1.3P-WaterHB, and SW1.3P-NoHB, respectively (Fig. 2c).

For the SW2 conformations, we extended the nomenclature based on nucleotide state because of the greater complexity of patterns observed (Fig. 1i, Supplementary Fig. 3b, and Supplementary Table 1). There were nine SW2 conformations in total, including the previously described R state (SW2.3P-R) and T state (SW2.3P-T), and the SW2 conformation found in nucleotide-free structures (which we named SW2.0P-GEF for its binding to GEFs); and six previously unclassified druggable conformations, which we named by their associated bound protein (only SW2.2P-Binder) or inhibitor site (SP12 or SP2) and, in some cases, an indicator of cluster size order (A or B). The SW2 conformations are visualized by nucleotide state in Fig. 2d-f with residue Y71 displayed, because we later demonstrate that the position of this residue relates to RAS druggability.

SW1 and SW2 conformations associated with biological RAS interactions. Based on biochemical studies in conjunction with analyses of RAS-protein co-crystals, researchers have resolved the SW1, and sometimes, SW2 conformations associated with certain RAS protein-protein interactions: (1) SW1 nucleotide-free (SW1.0P) binds to GEF CDC25 domains[6]; (2) SW1 GTP-bound (SW1.3P, preferably state 2) and SW2 R state (SW2.3P-R) binds to signaling effectors[32-35,52]; (3) SW1 GTP-bound (SW1.3P) binds to GEF REM domains[15]; and (4) the SW1 GTP-bound substate with no HB between Y32(OH) and 3P(O1G) (SW1.3P-NoHB) binds to GAPs[10,50]. However, we do not know the combination of SW1 and SW2 conformations in RAS proteins bound to each RAS interacting partner. Therefore, we analyzed which RAS conformational clusters are found in RAS-protein complexes currently available in the PDB (Table 2).

As expected, the nucleotide-free conformations, SW1.0P and SW2.0P-GEF, only exist in structures bound to the GEF CDC25 domain of SOS1, which has SW1 held open by a SOS1 region called the "helical hairpin"[6] (Fig. 3a, b and Table 2). All three SW1.3P substates bind to signaling effectors (Fig. 3c and Table 2), and we confirmed that only the SW2.3P-R conformation is found as well in these complexes (Fig. 3d and Table 2). Surprisingly, we discovered that the GEF REM domain of SOS1 preferably associates with SW1.3P-WaterHB (Fig. 3e and Table 2) and binds to the SW2.3P-R conformation (Fig. 3f and Table 2). Furthermore, we found that the GAP NF1 interacts with both SW1.3P-WaterHB (Fig. 3g, left and Table 2) and SW1.3P-NoHB (Fig. 3g, right and Table 2), with SW1.3P-WaterHB precluding the catalytic GAP "arginine finger"[7] from the active site and SW1.3P-NoHB enabling its direct interaction with GTP. This observation was similarly made in two previous studies, but not connected to the previously described GTP-bound (our SW1.3P) substates: one which identified these substates in RAS-NF1 complexes and called them the ground and transition states, respectively[10]; and another which found them across monomeric RAS structures and called them *Tyr32in* and *Tyr32out*, respectively[50].

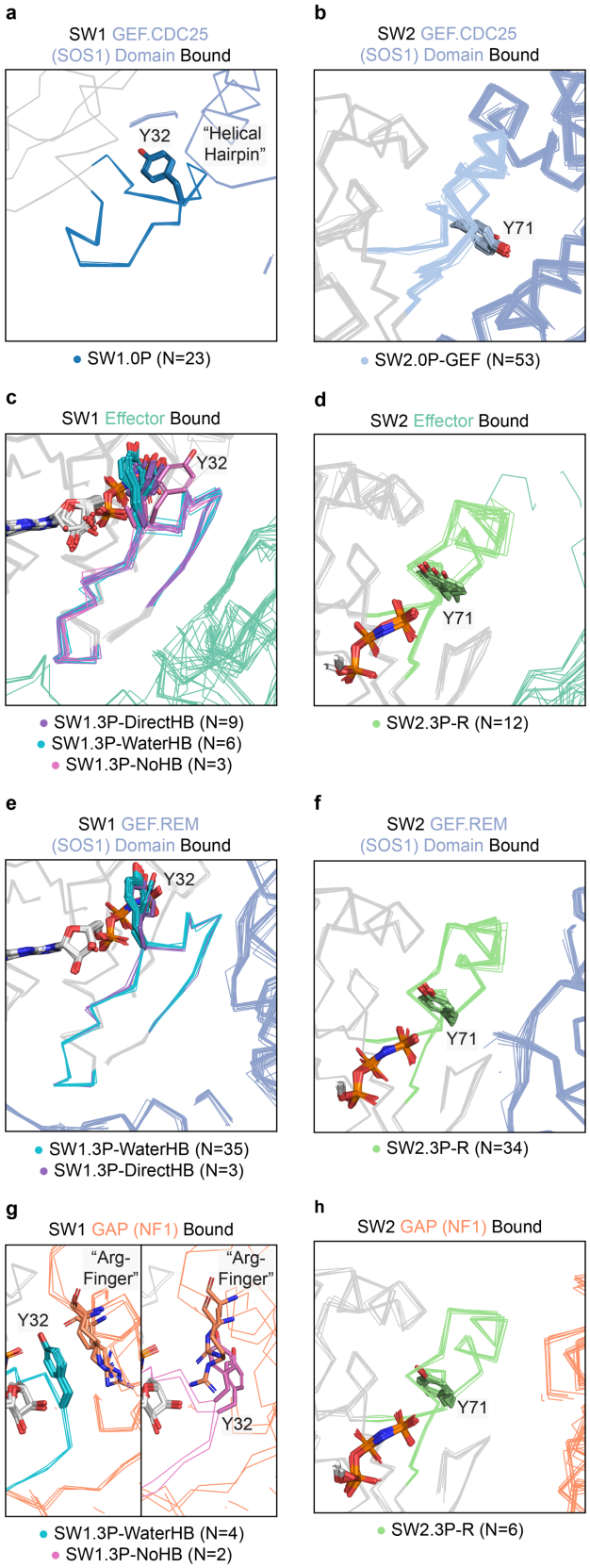

a
SW1 GEF.CDC25
(SOS1) Domain Bound
Y32
"Helical
Hairpin"
SW1.0P (N=23)
b
SW2 GEF.CDC25
(SOS1) Domain Bound
Y71
SW2.0P-GEF (N=53)
c
SW1 Effector Bound
Y32
SW1.3P-DirectHB (N=9)
SW1.3P-WaterHB (N=6)
SW1.3P-NoHB (N=3)
d
SW2 Effector Bound
Y71
SW2.3P-R (N=12)
e
SW1 GEF.REM
(SOS1) Domain Bound
Y32
SW1.3P-WaterHB (N=35)
SW1.3P-DirectHB (N=3)
f
SW2 GEF.REM
(SOS1) Domain Bound
Y71
SW2.3P-R (N=34)
g
SW1 GAP (NF1) Bound
"Arg-
Finger"
Y32
"Arg-
Finger"
Y32
SW1.3P-WaterHB (N=4)
SW1.3P-NoHB (N=2)
h
SW2 GAP (NF1) Bound
Y71
SW2.3P-R (N=6)

Fig. 3 | SW1 and SW2 conformations associated with RAS effector and regulatory protein interactions. SW1 and SW2 conformations, **a** and **b**, bound to the GEF CDC25 (catalytic) domains of SOS, **c** and **d**, effectors, **e** and **f**, the GEF REM (allosteric) domain of SOS1, and, **g** and **h**, the GAP NF1. **a,** "Helical hairpin" of SOS1 opening SW1 of RAS. **g**, Comparison of catalytic "arginine (Arg) finger" position for the GAP NF1 when bound to SW1.3P-WaterHB (left) and SW1.3P-NoHB (right, PDB: 6OB3 chains A and C, which are identical to the transition state stabilized in 1WQ1).

Table 2 | Distribution of SW1 and SW2 conformations by bound proteins.

Cluster Name		Effector	GAP	GEF		Binder	Nanodisc	Other	None	All
				CDC25	REM					
SW1.0P				23						23
Noise				37		1			6	44
SW1.2P						13		1	150	164
Noise		1	1	1		11	2	3	68	87
SW1.3P	**WaterHB**	6	4		35			3	107	155
	DirectHB	9			3	10			57	79
	NoHB	3	2			2	1		26	34
	Noise	3				19	10	2	79	113
SW2.0P	**GEF**			53						53
	Noise			7		1			6	14
SW2.2P	**SP12**								22	22
	SP2-A					8			12	20
	SP2-B							1	18	19
	Binder					11			2	13
	Noise	1	1	1		5	2	3	164	177
SW2.3P	**R**	12	6		34	4		2	40	98
	SP12-A								31	31
	SP12-B						2		12	14
	T								21	21
	Noise	9			4	27	9	3	165	217
All		22	7	61	38	56	13	9	493	699

SW1 and SW2 conformations found in α4α5 homodimers. Homodimerization of GTP-bound RAS monomers at their α4 and α5 helices is required in some cases for signal effector activation (e.g., RAF1)[16-19], and has been identified across RAS crystal structures in the PDB[53,54], but the conformations that can homodimerize are entirely unknown. Therefore, we identified the RAS α4α5 homodimer using the method employed in our Protein Common Interface Database (ProtCID) server[53]. We found the α4α5 homodimer across 140 HRAS, KRAS, and NRAS structures (31% of X-ray experiments; 115 PDB entries and 19 CFs) (Supplementary Data 1). The functional relevance of the α4α5 homodimer is further supported by the observance of RAS α4α5 homodimers in co-crystal complexes with the signaling effectors RAF1 (N=5), PLCε1 (N=2), and RASSF1 (N=1) as well as the GEF, GRP4 (N=6) (Supplementary Data 1).

All SW1.3P substates form the homodimer complex with more than half of them being SW1.3P-DirectHB (55%; N=40 of 73 SW1.3P dimers) (Supplementary Fig. 5a and Supplementary Table 2). Altogether, the SW1.3P structures most commonly form the α4α5 homodimer (52%; N=73 of 140 dimers), with SW1.2P forming this complex but less commonly (14%; N=20 of 140 dimers), and the remainder found in structures assigned to noise (34%; N=47 of 140 dimers) (Supplementary Fig. 5a and Supplementary Table 2). Both GTP-bound and GDP-bound α4α5 homodimers were also observed in NMR experiments (PDB: 6W4E and 6W4F, respectively)[19]. Surprisingly, we found that both active, SW2.3P-R and inactive, SW2.3P-T are the most common SW2

conformations (at approximately equal rates) that form the α4α5 homodimer (Supplementary Fig. 5b and Supplementary Table 2), contrary to the expectation that only active, GTP-bound RAS would form this complex. Of note, the SW2.3P-R conformation co-occurs with all SW1.3P substates, while 95.5% of the SW2.3P-T conformations (N=21 of 22 structures in the cluster) are found in conjunction with the hydrolytically incompetent SW1.3P-DirectHB substate (Table 1). This SW1-SW2 pairing suggests that RAS proteins may first occupy SW2.3P-T to prevent inactivation, then α4α5 homodimerize, and subsequently transition to SW2.3P-R to bind to signaling effectors[32-35].

SW1 and SW2 conformations possessing druggable pockets. RAS proteins are notoriously difficult to drug, because of their conformational variability and lack of deep surface pockets[3,4]. However, through NMR experiments and other techniques, druggable pockets have been identified in certain RAS conformations[55-58]. Therefore, we analyzed the available RAS structures in the PDB for druggable pockets with the Fpocket software[59], to associate the presence of inhibitor-bound and -unbound pockets with the identified SW1 and SW2 conformations.

We first obtained pocket descriptors for observed inhibitor-bound sites on RAS structures, including their pocket volumes and druggability scores. Out of the 699 available structures, 177 were bound to inhibitors: 48% at the SW1/SW2 pocket (SP12) site (N=85), 46% at the SW2 pocket (SP2) site (N=81), and the remaining 6% at other or multiple sites, which included the SP12 and SP2 sites as well as the base or center of the nucleotide site, a site near residue P110, or an allosteric site at the C-terminal end of the protein. We subsequently focused our analysis on the most targeted pockets, SP12 and SP2. With Fpocket, we were able to detect and calculate pocket descriptors for 93% of SP12 and 90% of SP2 inhibitor-bound sites (Fig. 4a, b). We then used Fpocket to predict potentially druggable pockets in inhibitor unbound structures and classified which of these predictions were found at the SP12 or SP2 sites based on similarity of their residue contacts. In all, we identified 203 SP12 and 215 SP2 inhibitor-unbound sites, which translated to more than 70% of these sites existing without inhibitors present in complex.

Examining SP12, SP2, and other sites, we found that inhibitor-bound pockets correlated with higher druggability scores (>0.5) than inhibitor-unbound pockets (Fig. 4c). Although SP12 inhibitor-bound and -unbound sites had no difference in pocket volumes (~250-500 $Å^3$), SP2 inhibitor bound sites displayed greater pocket volumes than their unbound counterparts (>500 $Å^3$). Importantly, the presence of SP12 and SP2 sites correlated with different side chain χ_1 (chi1) dihedrals (i.e., rotamers) for residue Y71 in SW2. The SP12 inhibitor-bound structures prefer a ~60°, or gauche-plus (g+) rotamer of Y71, which buries Y71 within the RAS domain, exposing the SP12 site and occluding the SP2 site; this trend for Y71 was previously described for a select few SP12 inhibitor-bound structures[8,38]. The SP2 inhibitor -bound structures, by contrast, contain a ~300°, or gauche-minus (g-) rotamer of Y71, which exposes Y71 to solvent thereby opening the SP2 site and occluding the SP12 site (Fig. 4d and Supplementary Fig. 6a, b).

Looking at the SW1 and SW2 conformations in RAS proteins possessing inhibitor-bound and -unbound pockets, we found that SP12 inhibitors preferentially bind to the SW1.3P-WaterHB substate with SW2 conformations SW2.3P-R, SW2.3P-SP12-A, SW2.3P-SP12-B, and SW2.2P-SP12 (Fig. 4 e, f and Table 3). SP2 drugs alternatively prefer binding to SW1.2P structures with SW2 conformations SW2.2P-SP2-A and SW2.2P-SP2-B (Fig. 4 g, h and Table 3). Across RAS structures in the PDB, we found that 3P structures prefer a g+ rotamer for Y71, while 2P structures prefer a g- rotamer for Y71 (Supplementary Fig 6c), which may explain why SP12 inhibitors mostly bind to 3P conformations, while SP2 inhibitors mainly bind to 2P conformations. For both SP12 and SP2 sites, inhibitor-bound and -unbound structures had a very similar distribution of SW1 and SW2 conformations (Table 3). For the most part, any SW1 or SW2 conformation that was found in both inhibitor-bound and -unbound structures had higher druggability scores in the inhibitor unbound setting than those conformations exclusive to inhibitor unbound structures (Supplementary Figure 7), indicating that these inhibitor-binding conformations may be better targets for small molecule inhibition.

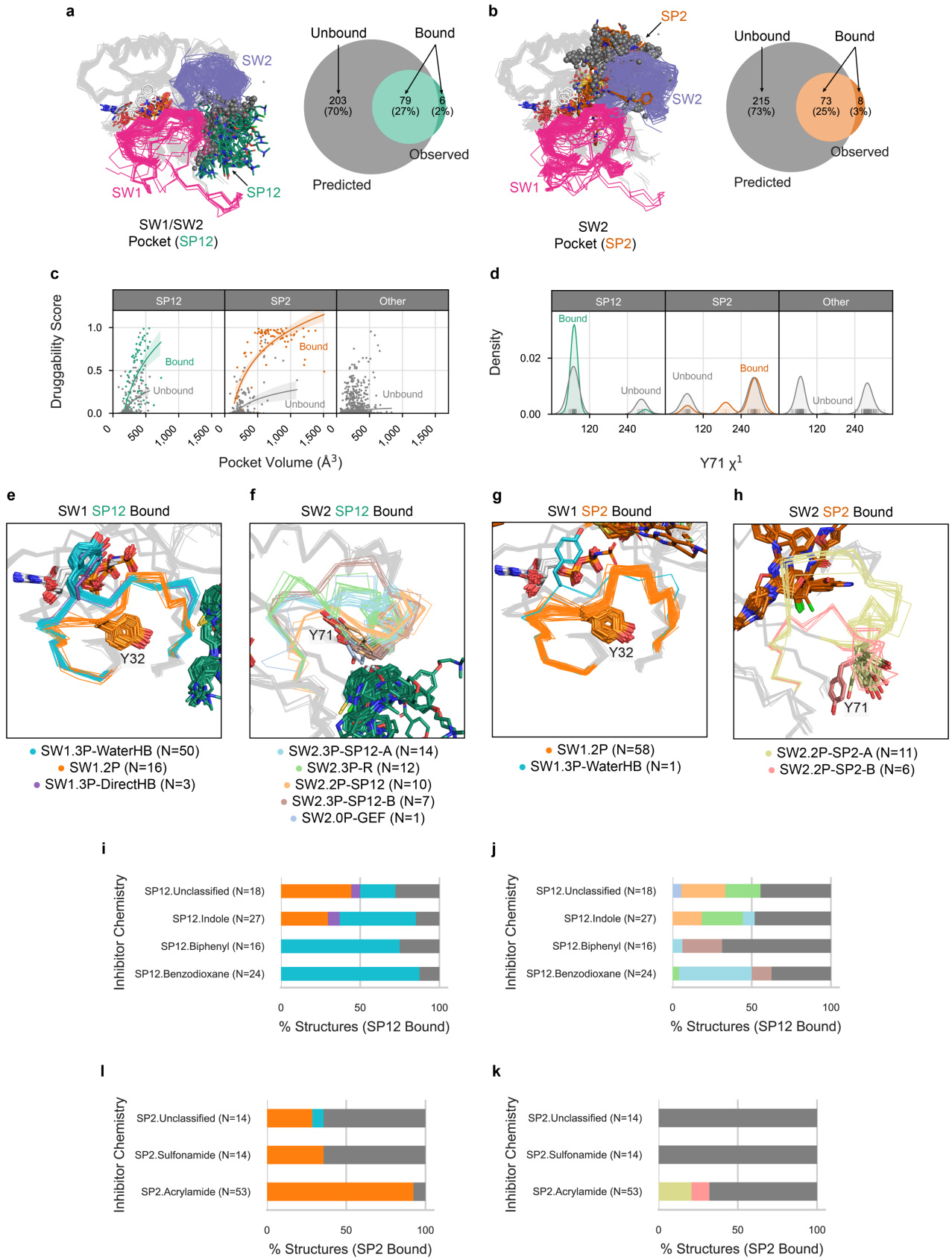

a
SW2
SW1
SP12
SW1/SW2
Pocket (SP12)
Unbound
Bound
203
(70%)
79
(27%)
6
(2%)
Observed
Predicted
b
SP2
SW2
SW1
SW2
Pocket (SP2)
Unbound
Bound
215
(73%)
73
(25%)
8
(3%)
Observed
Predicted
c
Druggability Score
SP12
SP2
Other
Bound
Unbound
Pocket Volume (Å3)
d
Density
Y71 χ1
e
SW1 SP12 Bound
Y32
SW1.3P-WaterHB (N=50)
SW1.2P (N=16)
SW1.3P-DirectHB (N=3)
f
SW2 SP12 Bound
Y71
SW2.3P-SP12-A (N=14)
SW2.3P-R (N=12)
SW2.2P-SP12 (N=10)
SW2.3P-SP12-B (N=7)
SW2.0P-GEF (N=1)
g
SW1 SP2 Bound
Y32
SW1.2P (N=58)
SW1.3P-WaterHB (N=1)
h
SW2 SP2 Bound
Y71
SW2.2P-SP2-A (N=11)
SW2.2P-SP2-B (N=6)
i
Inhibitor Chemistry
SP12.Unclassified (N=18)
SP12.Indole (N=27)
SP12.Biphenyl (N=16)
SP12.Benzodioxane (N=24)
% Structures (SP12 Bound)
j
Inhibitor Chemistry
SP12.Unclassified (N=18)
SP12.Indole (N=27)
SP12.Biphenyl (N=16)
SP12.Benzodioxane (N=24)
% Structures (SP12 Bound)
l
Inhibitor Chemistry
SP2.Unclassified (N=14)
SP2.Sulfonamide (N=14)
SP2.Acrylamide (N=53)
% Structures (SP2 Bound)
k
Inhibitor Chemistry
SP2.Unclassified (N=14)
SP2.Sulfonamide (N=14)
SP2.Acrylamide (N=53)
% Structures (SP2 Bound)

Fig. 4 | SW1 and SW2 conformations possessing observed and predicted druggable pockets. Observed (i.e., inhibitor bound) and predicted (not inhibitor bound), **a**, SW1/SW2 pockets (SP12) and, **b**, SW2 pockets (SP2) across RAS structures in the PDB. **c**, Pocket volumes and druggability scores and, **d**, residue Y71 χ_1 (chi1) rotamer distributions across inhibitor bound and unbound SP2, SP12, or other sites. SW1 and SW2 conformations in RAS structures with an inhibitor bound SP12 site, **e** and **f**, respectively, and inhibitor bound SP2 site, **g** and **h**, respectively. Percent of each SW1 and SW2 conformation bound to inhibitors with different chemistries at, **i** and **j**, respectively, the SP12 site and, **k** and **l**, respectively, the SP2 site. **i-l,** colored by the same scheme as **e**-**h** with gray indicating structures labeled noise.

Table 3 | Distribution of SW1 and SW2 conformations by inhibitor site.

Cluster Name		SP12		SP2		Multiple		Other		All
		Bound	Unbound	Bound	Unbound	Bound	Unbound	Bound	Unbound	
SW1.0P					23					23
Noise		2	1		31		2	7	1	44
SW1.2P		16	6	58	37	1	21		25	164
Noise		3	8	16	10		13		37	87
SW1.3P	**WaterHB**	50	52	1	8	3	10		31	155
	DirectHB	3	33		5		12		26	79
	NoHB		7		16		2		9	34
	Noise	11	26	6	15		10		45	113
SW2.0P	**GEF**	1			49			3		53
	Noise	1	1		5		2	4	1	14
SW2.2P	**SP12**	10					12			22
	SP2-A			11					9	20
	SP2-B			6	10		1		2	19
	Binder		1		8				4	13
	Noise	9	13	57	29	1	21		47	177
SW2.3P	**R**	12	36		8		8		34	98
	SP12-A	14	11		2		4			31
	SP12-B	7	3		1		1		2	14
	T		11		1		2		7	21
	Noise	31	57	7	32	3	19		68	217
All		85	133	81	145	4	70	7	174	699

Recently, SP2 inhibitors with divergent chemistries were found to bind different conformations[36,37], but the conformational preferences of other SP2 and SP12 inhibitor chemistries are unknown. Therefore, we subdivided SP2 and SP12 inhibitors by chemistries (focusing on inhibitor classes discussed repeatedly in the literature) and examined SW1 and SW2 conformational preferences by inhibitor chemistries and binding sites. The analyzed inhibitor chemistries included acrylamide and sulfonamide for SP2 inhibitors[39,60] and indole, benzodioxane, and biphenyl for SP12 inhibitors[61-66]. Indole SP12 compounds preferred binding to structures with the SW1.3P-WaterHB, SW1.3P-DirectHB, or SW1.2P SW1 conformations and the SW2.3P-R, SW2.2P-SP12, or SW2.3P-SP12-A SW2 conformations (Figure 4i, j and Supplementary Table 3). The indole SP12 compounds included RAS-SOS1 inhibitors (e.g., DCAI)[67,68] and those that block multiple key RAS interactions (e.g., BI-2852, Cmpd2)[69]. Of note, most indole SP12 inhibitors targeted KRAS G12D mutated structures in the PDB (Supplementary Data 1). Benzodioxane and biphenyl SP12 inhibitors instead preferentially bound to structures with the SW1.3P-WaterHB SW1 conformation and the SW2.3P-SP12-A or SW2.3P-SP12-B SW2 conformations (Figure 4i, j and Supplementary Table 3); these SP12 inhibitors function to block RAS-effector interactions (derived from intracellular antibody fragments, e.g., PPIN-1, PPIN-2)[62], and mostly target KRAS Q61H structures

in the PDB (Supplementary Data 1). Lastly, acrylamide SP2 compounds, which include the well-known KRAS G12C (covalent) inhibitors (e.g., sotorasib/AMG 510, adagrasib/MRTX849)[39,60], preferentially bind to structures with the SW1.2P SW1 conformation and the SW2.2P-SP2-A or SW2.2P-SP2-B SW2 conformations (Figure 4k, l and Supplementary Table 3). Sulfonamide SP2 compounds, which are KRAS G12C (covalent) inhibitors as well, solely bound to structures that the DBSCAN algorithm assigned to noise.

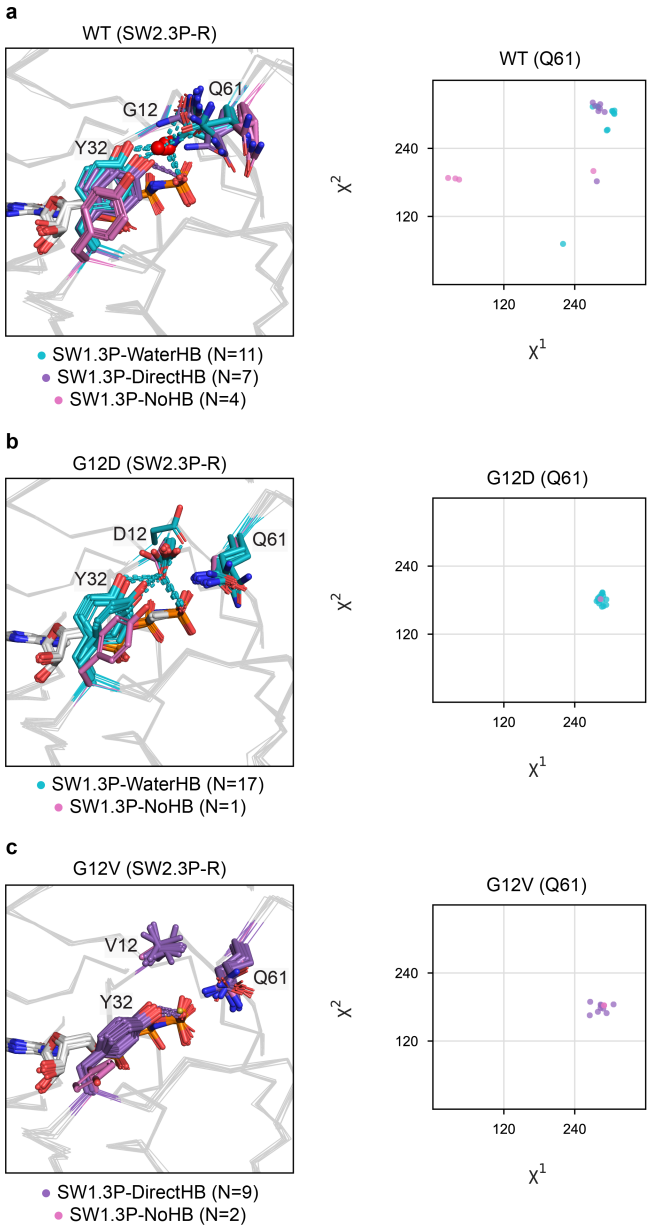


Fig. 5 | Structural impact of G12D and G12V mutations on intrinsic hydrolysis. SW1.3P substates within SW2.3P-R structures and their residue Q61 χ_1 (chi1) and χ_2 (chi2) rotamers by, **a**, WT, **b**, G12D, and, **c**, G12V forms.

Structural impact of G12D and G12V mutations on intrinsic hydrolysis. Mutations have been shown to impact RAS conformational preferences. Therefore, we sought to use our classification of RAS structures in the PDB to test a hypothesis proposed by Marcus and Mattos regarding the structural impact G12D and G12V mutations[32]. Through observing the placement of residue Y32 in a few WT, G12D, and G12V HRAS structures, Marcus and Mattos proposed that G12V mutations sterically push Y32 into a hydrolytically incompetent position

(our SW1.3P-DirectHB), severely impairing intrinsic hydrolysis, while G12D mutations stabilize Y32 (with residue D12) in a position preferential for intrinsic hydrolysis (our SW1.3P-WaterHB). Knowing that the active, SW2.3P-R conformation (1) can occur in conjunction with all three SW1.3P substates (Table 1), (2) is the suggested conformation to biologically undergo hydrolysis[32-35], and (3) is the only conformation in the PDB to include several WT, G12D, and G12V mutated structures (Supplementary Table 4), we chose to examine the SW1.3P substate preferences for WT, G12D, and G12V structures within the SW2.3P-R cluster (included HRAS and KRAS for all forms). While WT structures had a near even distribution of SW1.3P substates (Fig. 5a, left), 94% of G12D were 3P-WaterHB and 81% of G12V were 3P-DirectHB (Fig. 5b, c, left). In the analyzed G12D structures, residue D12 stabilizes Y32 in a position preferential for intrinsic hydrolysis (our SW1.3P-WaterHB), just as Marcus and Mattos proposed[32] (Fig. 5b, left).

To identify further insights regarding the selected WT, G12D, and G12V structures, we examined their residue Q61 side chain arrangements, since Q61 is considered the direct mediator of intrinsic hydrolysis for RAS proteins[51]. We found that for both SW1.3P-DirectHB and SW1.3P-WaterHB WT structures the Q61 side chain rotates towards the Y32(OH) and 3P(O1G) atoms ({χ_1, χ_2} {~300°, ~300°}, {g-, g-} rotamer), while the Q61 side chain rotates away from these atoms in most SW1.3P-NoHB WT structures ({χ_1, χ_2} {~60°, ~180°}, {g+, trans(t)} rotamer) (Fig. 5a, right). Since SW1.3P-DirectHB and SW1.3P-WaterHB have different propensities for hydrolysis, and both these substates have a different Y32 position but the same Q61 side chain placement, we suggest that Y32, instead of Q61, is the direct mediator of intrinsic hydrolysis. The hydrolytic importance of Y32 over Q61 is further supported by the fact that all analyzed G12D and G12V mutated structures both have their Q61 side chain partially rotated away ({χ_1, χ_2} {~300°, ~180°}, {g-, t} rotamer) from Y32(OH) and 3P(O1G) (akin to SW1.3P-NoHB WT structures) despite preferring SW1.3P substates with different hydrolytic propensities (Figure 5b, c, right). Overall, our extended analysis of G12D and G12V mutations clarifies the structural arrangement required for intrinsic hydrolysis and solidifies the central role of the SW1.3P substates for RAS activity.

Discussion

Since the first HRAS structure was experimentally solved in 1990[26,70], researchers have focused on characterizing the RAS conformational landscape through examining the structural arrangements of their SW1 and SW2 loops. In this study, we used an extended dataset (699 KRAS, NRAS, and HRAS structures), and an approach that differs from previous studies (which analyzed 121 entries at most)[40,41], to create a data-driven classification of three SW1 and nine SW2 RAS conformations. This approach can be used to automatically conformationally classify and annotate the molecular contents of additional RAS structures as they are experimentally solved and provides a clear and consistent method for comparing WT and mutated structures across various biological and inhibitory contexts. To facilitate future analyses of RAS structures, we have created a web database presenting our analysis of RAS structures in the PDB, which includes a page for classifying user inputted structures (http://dunbrack.fccc.edu/rascore/).

One uncertainty faced in defining a RAS conformational classification was identifying which GTP-bound SW1 conformations are state 1 and state 2. These SW1 conformations were discovered in the early 2000s with the observation of two peaks in the ^{31}P NMR spectra for the GTP α and γ phosphates[52]. Later studies found that mutations in residues Y32 and T35, as well as the common G12V mutation, cause a shift to the state 1-associated peaks, while other mutations, such as G12D, and the presence of the signaling effector, RAF1, cause a shift to state 2[71-73]. Ten years after the state 1 and state 2 conformations were described, researchers experimentally solved a potential state 1 structure using a T35S mutant construct[29], and later for WT, G12V, Q61L, and other mutations[27,30]. However, the previously identified state 1 structures were too infrequently occurring in our analysis to unambiguously name them as the actual state 1 conformation. Moreover, as alluded to by Mattos and colleagues[33,49], the identified GTP-bound substates (our SW1.3P-WaterHB, SW1.3P-DirectHB, and SW1.3P-NoHB), could also explain the split state 1 and state 2 peaks in NMR spectra. Considering the NMR studies described above, and that we found G12D and G12V mutated structures prefer the SW1.3P-WaterHB and SW1.3P-DirectHB substates, respectively, we propose that the SW1.3P-WaterHB substate is state 2 and that state 1 is either the SW1.3P-DirectHB or SW1.3P-NoHB substates, the other non-clustered state 1 structures, SW1-disordered structures, or some mixture of these structural configurations.

In contrast to other studies[40,41], we associated each SW1 and SW2 conformation with RAS interactions involving proteins and small molecule inhibitors. This analysis helped confirm previously held hypotheses about RAS conformations in a large dataset and uncovered some new hidden trends. For example, it has been hypothesized that RAS preferentially binds to signaling effector proteins and the GEF REM domain of SOS1 when its SW1 conformation is "GTP-bound" (our SW1.3P) and to signaling effectors when the SW2 conformation is in the "R state" (our SW2.3P-R)[15,34,35]. We found these hypotheses to be true, but further discovered that all SW1.3P substates (based on their hydrogen bonding of Y32 to GTP) and the SW2.3P-R conformation binds to signaling effectors while the GEF REM domain of SOS1 preferentially binds to the SW1.3P-WaterHB substate and SW2.3P-R conformations. Under the same SW1.3P substate classification, we clarified that the previously described GAP-binding conformations, namely the ground and transition states[10] and *Tyr32in* and *Tyr32out* states[50], are actually the SW1.3P-WaterHB and SW1.3P-NoHB substates, respectively. Similarly, we defined that the inactive SW2 conformations "T state"[32-34] and "state 2*[35] are in actuality identical structural arrangements and that both belong to our SW2.3P-T conformation. Importantly, by comparing all inactive SW2.3P-T structures with their active counterparts in cluster SW2.3P-R, we confirmed the hypothesis that, unlike SW2.3P-R, SW2.3P-T does not bind to signaling effector proteins. We also found that both these conformations can form the RAS α4α5 homodimer complex required for activation of dimeric signaling effectors, such as RAF1. Furthermore, we confirmed that both GTP-bound and GDP-bound SW1 structures (here SW1.3P and SW1.2P) can form the α4α5 homodimer, as previously shown through NMR experiments of KRAS[19], but we demonstrated as well that all three SW1.3P substates and all RAS isoforms can complex as an α4α5 homodimer at least in crystals.

A major value of this study is the definition of a comprehensive set of RAS conformations that are known targets for small molecule or designed protein inhibitors. Six out of seven of these druggable SW2 conformations are newly characterized (all except for SW2.3P-R); these include: GTP-bound SW2.3P-SP12-A and SW2.3P-SP12-B, and GDP-bound SW2.2P-SP12, SW2.2P-SP2-A, SW2.2P-SP2-B, and SW2.2P-Binder. We associated each of these conformations with their preference for binding inhibitors with certain chemistries, which is information researchers can use to select appropriate structural templates for structure-guided drug design. One overall finding from our analysis of these druggable conformations was that all of them exist in the absence of inhibitors, indicating that these structural arrangements may occur naturally within a biological context and are not solely the product of drug binding, which was an uncertainty prior to this study. In addition, we found that the SP2 inhibitor site is only present in structures with Y71 exposed to solvent, while the SP12 inhibitors site appears in structures with Y71 buried into the protein core. Although this trend for Y71 was previously described for a select few SP12 inhibitor-bound structures[8,38], the consistency of this finding among many inhibitor-bound structures suggests it is an essential determinant of SP2 and SP12 druggability.

While this study has expanded our understanding of RAS conformations, it only marks the beginning of mapping the RAS conformational landscape. We hope that the RAS conformational classification system described here will be paired with further structure-activity relationship data to create machine learning models for RAS drug discovery. The pharmaceutical industry has over six times as many RAS inhibitor-bound structures as there are available in the PDB[74], and analysis of these structures using the conformational clustering approach developed here would likely help identify further druggable RAS conformations. Most importantly, having all RAS structures in the PDB consistently annotated for their molecular contents and conformation will enable simple utilization of this growing structural dataset for informing future RAS drug discovery and studies examining RAS mutations.

Methods

In this study, we conformationally clustered the available human KRAS, NRAS, and HRAS structures in the PDB and determined structural features associated with each conformation. Here we describe the methods for preparing the available RAS structures, annotating their molecular contents, clustering SW1 and SW2 conformations, and performing further interrogative structural analyses.

Software Utilized. All analyses were performed using various packages in Python with versions provided in our code in GitHub (https://github.com/mitch-parker/rascore). BioPython and PyMOL were used for structure and

sequence calculations and visualizations. Pandas and Numpy were used for dataset preparation and querying. SciKit Learn was used for clustering. RDKit was used for chemical searching and visualizations. Matplotlib and Seaborn were used for plotting.

Preparing Available RAS Structures. PDB entries containing human KRAS (all are 4B isoform), NRAS, and HRAS were identified by SWISS-PROT[75] identifier in the *pdbaa* file (December 1 2021) in the PISCES webserver[76]. For each PDB entry, the asymmetric unit and all biological assemblies were downloaded and renumbered according to UniProt[77] scheme using PDBrenum[78]. In addition, electron density of individual atom (EDIA) scores (a per atom measure of model quality)[79] for each PDB entry were downloaded from the ProteinPlus webserver[80]. Since some PDB entries contain multiple RAS polypeptide chains, we separated each RAS chain of the asymmetric unit (only first model for NMR) with its corresponding bound ligands and/or proteins. Ligands were labeled as biological, pharmacological, or chemical compounds, metal ions, residue modifications, or membrane components using a custom dictionary prepared in considering annotations from BioLiP[81] and FireDB[82], which is included in our code in GitHub. Subsequently, ligands were assigned to a RAS chain if they (1) had the same chain label (only biological and chemical compounds, metal ions, or residue modifications); (2) had more than 5 residue contacts within 4 Å of the chain (only pharmacological compounds); or (3) linked the chain to a nanodisc (only membrane components). Proteins were assigned to a RAS chain if it had more than 5 Cβ contacts within 12 Å and 1 atom contact within 5 Å or if it had more than 5 atom contacts within 5 Å of the chain, except for the protein component of nanodiscs which were included irrespective of the number of contacts. Bound protein assignments were checked against the biological assemblies and discrepancies were corrected. We treated each RAS chain as a unique RAS structure in subsequent analyses.

Annotating RAS Structures. We annotated RAS structures by various molecular contents, many of which are not reported in PDB entries. Mutation status was identified by comparison of the sequence in the PDB entry and human UniProt[77] sequences for KRAS (P01116-1 and -2), NRAS (P01111), and HRAS (P01112). Pharmacological compounds were further classified by binding site based on the presence of one or more predefined residue contacts within 4 Å of the structure: 12, 96, or 99 for SP2; 5, 39, or 54 for SP12; 85, 118, and 119 for the base of the nucleotide site; 29, 30, and 32 for the center the nucleotide site; 106, 108, and 110 for the P110 site; and 4, 49, and 164 for the allosteric site. The SMILES strings used in searching for inhibitor chemistries (performed with RDKit) are included as a Supplementary Note. Bound proteins were labeled by Pfam[83] based on SWISS-PROT[75] identifier and further classified as an effector, (Pfams: RBD, RA, PI3K_rbd), GEF (Pfam: RasGEF), or GAP (Pfam: RasGAP), or other. Any bound protein without a SWISS-PROT identifier was classified as a designed protein “binders” or RAS linked to a nanodisc. To identify the α4α5 homodimer, we used the protocol employed in the ProtCID web server[53], requiring an average Q-score greater than 0.3 to the α4α5 homodimer found in PDB: 3K8Y. In addition, each PDB entry was assigned to a crystal form using are method previously described and implemented in ProtCID[53,84].

Conformationally Clustering SW1 and SW2. RAS structures were first separated by nucleotide state: 0P, 2P, or 3P. Within each nucleotide state, we clustered the completely modeled SW1 and SW2 loops possessing carbonyl (O) atom EDIA scores greater than 0.4 (i.e., well modeled) using the DBSCAN algorithm with a dihedral-based distance metric. DBSCAN finds major clusters and removes outliers[43], which is ideal for conformationally clustering structural datasets since they usually contain several outliers that were poorly modeled or solved under rare experimental conditions. Variations of our conformational clustering algorithm have been described in our previous works[44-47] . In this study, we used a distance metric that locates the maximum angular difference (*d*) upon pairwise comparison of the backbone dihedral angle values phi (φ), psi (ψ), and omega (ω) for residues 1 through *n* of compared loops *i* versus *j,* where $d(\theta_i, \theta_j) = 2(1 - \cos(\theta_j - \theta_i))$:

$$D_{max}(i,j) = \max(d(\varphi_i^1, \varphi_j^1), d(\psi_i^1, \psi_j^1), d(\omega_i^1, \omega_j^1) \ldots d(\varphi_i^n, \varphi_j^n), d(\psi_i^n, \psi_j^n), d(\omega_i^n, \omega_j^n))$$

For SW1, we calculated D_{max} for residues 25-40. For SW2, we calculated D_{max} for residues 56-76 and included chi1 (χ^1) of residue 71 in the calculation, since its position indicates SW2 flexibility[85,86] and because we found that it affects the SP2 and SP12 sites. These residue ranges were selected because they both quantitatively and qualitatively encompass the extent of SW1 and SW2 conformational variability across RAS structures in the PDB. Like a previous study[44], we ran DBSCAN across a grid of parameters and applied a set of quality control filters to generate a robust consensus clustering. This procedure was necessary since a single setting of DBSCAN does not identify all possible clusters due to their varying shapes and sizes. We elaborate on our conformational clustering pipeline as a Supplementary Note.

Detecting Hydrogen Bonds. Conformational substates were defined by the hydrogen bond (HB) interaction between the Y32 OH atom and γ phosphate of GTP or its analogs: direct (DirectHB) or water-mediate (Water HB). DirectHB and WaterHB cutoffs were based on a previous analysis of protein structures in the PDB[87]. DirectHB was defined as a 2.0-3.2 Å donor-acceptor distance with 90-180° carbon-donor-acceptor and carbon-acceptor-donor angles. WaterHB was defined as 2.0-3.0 Å donor-water and acceptor-water distances with 80-140° carbon-water-acceptor and carbon-water-donor angles. In the absence of a bridging-water, WaterHB was defined as a 2.8-5.6 Å donor-acceptor distance, which was arrived at using the Law of Cosines with the previously specified cutoffs and examining the distance distributions of WaterHB across RAS structures in the PDB.

Analyzing Druggable Pockets. Pocket descriptors (i.e., pocket volumes and druggability scores) for inhibitor bound SP12 and SP2 sites were calculated with the Fpocket software[59]. Fpocket was then used to predict pockets in inhibitor unbound structures. Predicted pockets were assigned to the SP12 or SP2 sites if the average Simpson similarity of their residue contacts to bound inhibitors at those respective sites was greater than 0.6.

Data Availability

All data are available as a static table in Supplementary Data 1. In addition, we have created a web database called “rascore” that presents a continually updated dataset of annotated and conformationally classified RAS structures from the PDB. Further, the rascore database includes a page for conformationally classifying user inputted structures. The link to our rascore database and accompanying code for conformationally classifying RAS structures via the command line can be found in GitHub (https://github.com/mitch-parker/rascore).

Code Availability

All open-source code can be obtained from GitHub **(**https://github.com/mitch-parker/rascore) under the MIT license. Software used in this study are included in Methods with more extensive details provided in GitHub, such as package versions and computational environment setup.

References

1. Prior, I.A., Hood, F.E. & Hartley, J.L. The Frequency of Ras Mutations in Cancer. *Cancer Research* **80**, 2969 (2020).
2. Lu, S. et al. Ras Conformational Ensembles, Allostery, and Signaling. *Chem Rev* **116**, 6607-65 (2016).
3. Moore, A.R., Rosenberg, S.C., McCormick, F. & Malek, S. RAS-targeted therapies: is the undruggable drugged? *Nature Reviews Drug Discovery* **19**, 533-552 (2020).
4. Hofmann, M.H., Gerlach, D., Misale, S., Petronczki, M. & Kraut, N. Expanding the Reach of Precision Oncology by Drugging All KRAS Mutants. *Cancer Discov* (2022).
5. Simanshu, D.K., Nissley, D.V. & McCormick, F. RAS Proteins and Their Regulators in Human Disease. *Cell* **170**, 17-33 (2017).
6. Boriack-Sjodin, P.A., Margarit, S.M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337-43 (1998).
7. Scheffzek, K. et al. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333-8 (1997).
8. Kessler, D. et al. Drugging all RAS isoforms with one pocket. *Future Medicinal Chemistry* **12**, 1911-1923 (2020).

9. Hunter, J.C. et al. Biochemical and Structural Analysis of Common Cancer-Associated KRAS Mutations. *Molecular Cancer Research* **13**, 1325 (2015).
10. Rabara, D. et al. KRAS G13D sensitivity to neurofibromin-mediated GTP hydrolysis. *Proceedings of the National Academy of Sciences* **116**, 22122 (2019).
11. Johnson, C.W. et al. Isoform-Specific Destabilization of the Active Site Reveals a Molecular Mechanism of Intrinsic Activation of KRas G13D. *Cell Reports* **28**, 1538-1550.e7 (2019).
12. Poulin, E.J. et al. Tissue-Specific Oncogenic Activity of KRAS(A146T). *Cancer Discov* **9**, 738-755 (2019).
13. Buhrman, G., Wink, G. & Mattos, C. Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. *Structure (London, England : 1993)* **15**, 1618-1629 (2007).
14. Nakhaeizadeh, H., Amin, E., Nakhaei-Rad, S., Dvorsky, R. & Ahmadian, M.R. The RAS-Effector Interface: Isoform-Specific Differences in the Effector Binding Regions. *PloS one* **11**, e0167145-e0167145 (2016).
15. Margarit, S.M. et al. Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* **112**, 685-95 (2003).
16. Ambrogio, C. et al. KRAS Dimerization Impacts MEK Inhibitor Sensitivity and Oncogenic Activity of Mutant KRAS. *Cell* **172**, 857-868.e15 (2018).
17. Rudack, T. et al. The Ras dimer structure. *Chemical Science* **12**, 8178-8189 (2021).
18. Nan, X. et al. Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway. *Proceedings of the National Academy of Sciences* **112**, 7996 (2015).
19. Lee, K.Y. et al. Two Distinct Structures of Membrane-Associated Homodimers of GTP- and GDP-Bound KRAS4B Revealed by Paramagnetic Relaxation Enhancement. *Angew Chem Int Ed Engl* **59**, 11037-11045 (2020).
20. Pantsar, T. The current understanding of KRAS protein structure and dynamics. *Computational and Structural Biotechnology Journal* **18**, 189-198 (2020).
21. ww, P.D.B.c. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Research* **47**, D520-D528 (2019).
22. Haza, K.Z. et al. RAS-inhibiting biologics identify and probe druggable pockets including an SII-α3 allosteric site. *Nature Communications* **12**, 4045 (2021).
23. Bery, N. et al. KRAS-specific inhibition using a DARPin binding to a site in the allosteric lobe. *Nature Communications* **10**, 2607 (2019).
24. Guillard, S. et al. Structural and functional characterization of a DARPin which inhibits Ras nucleotide exchange. *Nature Communications* **8**, 16111 (2017).
25. Dharmaiah, S. et al. Structures of N-terminally processed KRAS provide insight into the role of N-acetylation. *Sci Rep* **9**, 10512 (2019).
26. Milburn, M.V. et al. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939 (1990).
27. Muraoka, S. et al. Crystal structures of the state 1 conformations of the GTP-bound H-Ras protein and its oncogenic G12V and Q61L mutants. *FEBS Lett* **586**, 1715-8 (2012).
28. Araki, M. et al. Solution structure of the state 1 conformer of GTP-bound H-Ras protein and distinct dynamic properties between the state 1 and state 2 conformers. *J Biol Chem* **286**, 39644-53 (2011).
29. Shima, F. et al. Structural basis for conformational dynamics of GTP-bound Ras protein. *J Biol Chem* **285**, 22696-705 (2010).
30. Lu, J., Bera, A.K., Gondi, S. & Westover, K.D. KRAS Switch Mutants D33E and A59G Crystallize in the State 1 Conformation. *Biochemistry* **57**, 324-333 (2018).
31. Kalbitzer, H.R. & Spoerner, M. State 1(T) inhibitors of activated Ras. *Enzymes* **33 Pt A**, 69-94 (2013).
32. Marcus, K. & Mattos, C. Direct Attack on RAS: Intramolecular Communication and Mutation-Specific Effects. *Clin Cancer Res* **21**, 1810-8 (2015).
33. Johnson, C.W. & Mattos, C. Chapter Three - The Allosteric Switch and Conformational States in Ras GTPase Affected by Small Molecules. in *The Enzymes*, Vol. 33 (ed. Tamanoi, F.) 41-67 (Academic Press, 2013).
34. Holzapfel, G., Buhrman, G. & Mattos, C. Shift in the equilibrium between on and off states of the allosteric switch in Ras-GppNHp affected by small molecules and bulk solvent composition. *Biochemistry* **51**, 6114-26 (2012).
35. Matsumoto, S. et al. Molecular Mechanism for Conformational Dynamics of Ras·GTP Elucidated from In-Situ Structural Transition in Crystal. *Scientific Reports* **6**, 25931 (2016).

36. Gentile, D.R. et al. Ras Binder Induces a Modified Switch-II Pocket in GTP and GDP States. *Cell Chem Biol* **24**, 1455-1466.e14 (2017).
37. Zeng, M. et al. Potent and Selective Covalent Quinazoline Inhibitors of KRAS G12C. *Cell Chemical Biology* **24**, 1005-1016.e3 (2017).
38. Nyíri, K., Koppány, G. & Vértessy, B.G. Structure-based inhibitor design of mutant RAS proteins—a paradigm shift. *Cancer and Metastasis Reviews* **39**, 1091-1105 (2020).
39. Ostrem, J.M., Peters, U., Sos, M.L., Wells, J.A. & Shokat, K.M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* **503**, 548-551 (2013).
40. Li, H., Yao, X.-Q. & Grant, B.J. Comparative structural dynamic analysis of GTPases. *PLOS Computational Biology* **14**, e1006364 (2018).
41. Gorfe, A.A., Grant, B.J. & McCammon, J.A. Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins. *Structure* **16**, 885-96 (2008).
42. Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95-9 (1963).
43. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* 226–231 (AAAI Press, Portland, Oregon, 1996).
44. Kelow, S.P., Adolf-Bryfogle, J. & Dunbrack, R.L. Hiding in plain sight: structure and sequence analysis reveals the importance of the antibody DE loop for antibody-antigen binding. *MAbs* **12**, 1840005 (2020).
45. Modi, V. & Dunbrack, R.L. Defining a new nomenclature for the structures of active and inactive kinases. *Proceedings of the National Academy of Sciences* **116**, 6818 (2019).
46. Shapovalov, M., Vucetic, S. & Dunbrack, R.L., Jr. A new clustering and nomenclature for beta turns derived from high-resolution protein structures. *PLoS Comput Biol* **15**, e1006844 (2019).
47. North, B., Lehmann, A. & Dunbrack, R.L., Jr. A new clustering of antibody CDR loop conformations. *Journal of molecular biology* **406**, 228-256 (2011).
48. Bera, A.K. et al. Structural basis of the atypical activation mechanism of KRAS(V14I). *J Biol Chem* **294**, 13964-13972 (2019).
49. Buhrman, G. et al. Analysis of binding site hot spots on the surface of Ras GTPase. *J Mol Biol* **413**, 773-89 (2011).
50. Li, Y. et al. Specific Substates of Ras To Interact with GAPs and Effectors: Revealed by Theoretical Simulations and FTIR Experiments. *J Phys Chem Lett* **9**, 1312-1317 (2018).
51. Buhrman, G., Holzapfel, G., Fetics, S. & Mattos, C. Allosteric modulation of Ras positions Q61 for a direct role in catalysis. *Proceedings of the National Academy of Sciences* **107**, 4931 (2010).
52. Spoerner, M., Herrmann, C., Vetter, I.R., Kalbitzer, H.R. & Wittinghofer, A. Dynamic properties of the Ras switch I region and its importance for binding to effectors. *Proceedings of the National Academy of Sciences* **98**, 4944 (2001).
53. Xu, Q. & Dunbrack, R.L. ProtCID: a data resource for structural information on protein interactions. *Nature Communications* **11**, 711 (2020).
54. Spencer-Smith, R. et al. Inhibition of RAS function through targeting an allosteric regulatory site. *Nat Chem Biol* **13**, 62-68 (2017).
55. Prakash, P., Hancock, J.F. & Gorfe, A.A. Binding hotspots on K-ras: consensus ligand binding sites and other reactive regions from probe-based molecular dynamics analysis. *Proteins* **83**, 898-909 (2015).
56. Grant, B.J. et al. Novel Allosteric Sites on Ras for Lead Generation. *PLOS ONE* **6**, e25711 (2011).
57. Marshall, C.B. et al. NMR in integrated biophysical drug discovery for RAS: past, present, and future. *J Biomol NMR* **74**, 531-554 (2020).
58. Donohue, E. et al. Second harmonic generation detection of Ras conformational changes and discovery of a small molecule binder. *Proceedings of the National Academy of Sciences* **116**, 17290 (2019).
59. Le Guilloux, V., Schmidtke, P. & Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).
60. Ostrem, J.M. & Shokat, K.M. Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. *Nat Rev Drug Discov* **15**, 771-785 (2016).
61. Cruz-Migoni, A. et al. Structure-based development of new RAS-effector inhibitors from a combination of active and inactive RAS-binding compounds. *Proceedings of the National Academy of Sciences* **116**, 2545 (2019).
62. Quevedo, C.E. et al. Small molecule inhibitors of RAS-effector protein interactions derived using an intracellular antibody fragment. *Nature Communications* **9**, 3169 (2018).

63. Bery, N. et al. BRET-based RAS biosensors that show a novel small molecule is an inhibitor of RAS-effector protein-protein interactions. *eLife* **7**, e37122 (2018).
64. Chen, F.-Y., Li, X., Zhu, H.-P. & Huang, W. Regulation of the Ras-Related Signaling Pathway by Small Molecules Containing an Indole Core Scaffold: A Potential Antitumor Therapy. *Frontiers in pharmacology* **11**, 280-280 (2020).
65. Burns, M.C. et al. Approach for targeting Ras with small molecules that activate SOS-mediated nucleotide exchange. *Proceedings of the National Academy of Sciences* **111**, 3401 (2014).
66. Abbott, J.R. et al. Discovery of Aminopiperidine Indoles That Activate the Guanine Nucleotide Exchange Factor SOS1 and Modulate RAS Signaling. *Journal of Medicinal Chemistry* **61**, 6002-6017 (2018).
67. Sun, Q. et al. Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angew Chem Int Ed Engl* **51**, 6140-3 (2012).
68. Maurer, T. et al. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proceedings of the National Academy of Sciences* **109**, 5299 (2012).
69. Kessler, D. et al. Drugging an undruggable pocket on KRAS. *Proceedings of the National Academy of Sciences* **116**, 15823 (2019).
70. Pai, E.F. et al. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *Embo j* **9**, 2351-9 (1990).
71. Spoerner, M., Wittinghofer, A. & Kalbitzer, H.R. Perturbation of the conformational equilibria in Ras by selective mutations as studied by 31P NMR spectroscopy. *FEBS Lett* **578**, 305-10 (2004).
72. Smith, M.J., Neel, B.G. & Ikura, M. NMR-based functional profiling of RASopathies and oncogenic RAS mutations. *Proceedings of the National Academy of Sciences* **110**, 4574 (2013).
73. Parker, J.A., Volmar, A.Y., Pavlopoulos, S. & Mattos, C. K-Ras Populates Conformational States Differently from Its Isoform H-Ras and Oncogenic Mutant K-RasG12D. *Structure* **26**, 810-820.e4 (2018).
74. Mullard, A. What does AlphaFold mean for drug discovery? *Nat Rev Drug Discov* **20**, 725-727 (2021).
75. Duvaud, S. et al. Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. *Nucleic Acids Research* **49**, W216-W227 (2021).
76. Wang, G. & Dunbrack, R.L., Jr. PISCES: a protein sequence culling server. *Bioinformatics* **19**, 1589-91 (2003).
77. Consortium, T.U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506-D515 (2018).
78. Faezov, B. & Dunbrack, R.L. PDBrenum: a webserver and program providing Protein Data Bank files renumbered according to their UniProt sequences. *bioRxiv*, 2021.02.14.431128 (2021).
79. Meyder, A., Nittinger, E., Lange, G., Klein, R. & Rarey, M. Estimating Electron Density Support for Individual Atoms and Molecular Fragments in X-ray Structures. *Journal of Chemical Information and Modeling* **57**, 2437-2447 (2017).
80. Fährrolfes, R. et al. ProteinsPlus: a web portal for structure analysis of macromolecules. *Nucleic Acids Research* **45**, W337-W343 (2017).
81. Yang, J., Roy, A. & Zhang, Y. BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic acids research* **41**, D1096-D1103 (2013).
82. Maietta, P. et al. FireDB: a compendium of biological and pharmacologically relevant ligands. *Nucleic Acids Research* **42**, D267-D272 (2014).
83. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Research* **49**, D412-D419 (2021).
84. Xu, Q. et al. Statistical analysis of interface similarity in crystals of homologous proteins. *J Mol Biol* **381**, 487-507 (2008).
85. Bandaru, P. et al. Deconstruction of the Ras switching cycle through saturation mutagenesis. *eLife* **6**, e27810 (2017).
86. Guterres, H. & Mattos, C. Ras Residue Y71 Promotes Flexibility of Switch I and Switch II. *The FASEB Journal* **31**, 913.15-913.15 (2017).
87. Jiang, L., Kuhlman, B., Kortemme, T. & Baker, D. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins* **58**, 893-904 (2005).

Acknowledgements

We thank Bulat Faezov for providing the program PDBrenum in advance of publication, Vivek Modi for sharing scripts for calculating dihedral angles, and Simon Kelow for guidance on writing the conformational clustering algorithm. This work was funded by NIH F30 GM142263 (to M.I.P.), NIH R35 GM122517 (to R.L.D.), Colon Cancer Alliance funding (to J.E.M.), and the NCI Core Grant P30 CA006927 (to Fox Chase Cancer Center).

Author contributions

Conceptualization, Investigation, Methodology, Project Administration, Resources, Validation, and Writing – Original Draft, M.I.P and R.L.D.; Data Curation, Formal Analysis, Software, and Visualization, M.I.P.; Supervision, J.E.M., E.A.G., R.L.D.; Funding Acquisition and Writing - Review & Editing, M.I.P, J.E.M, E.A.G, and R.L.D.

Competing interests

The authors declare no competing interests.

Supplementary Notes

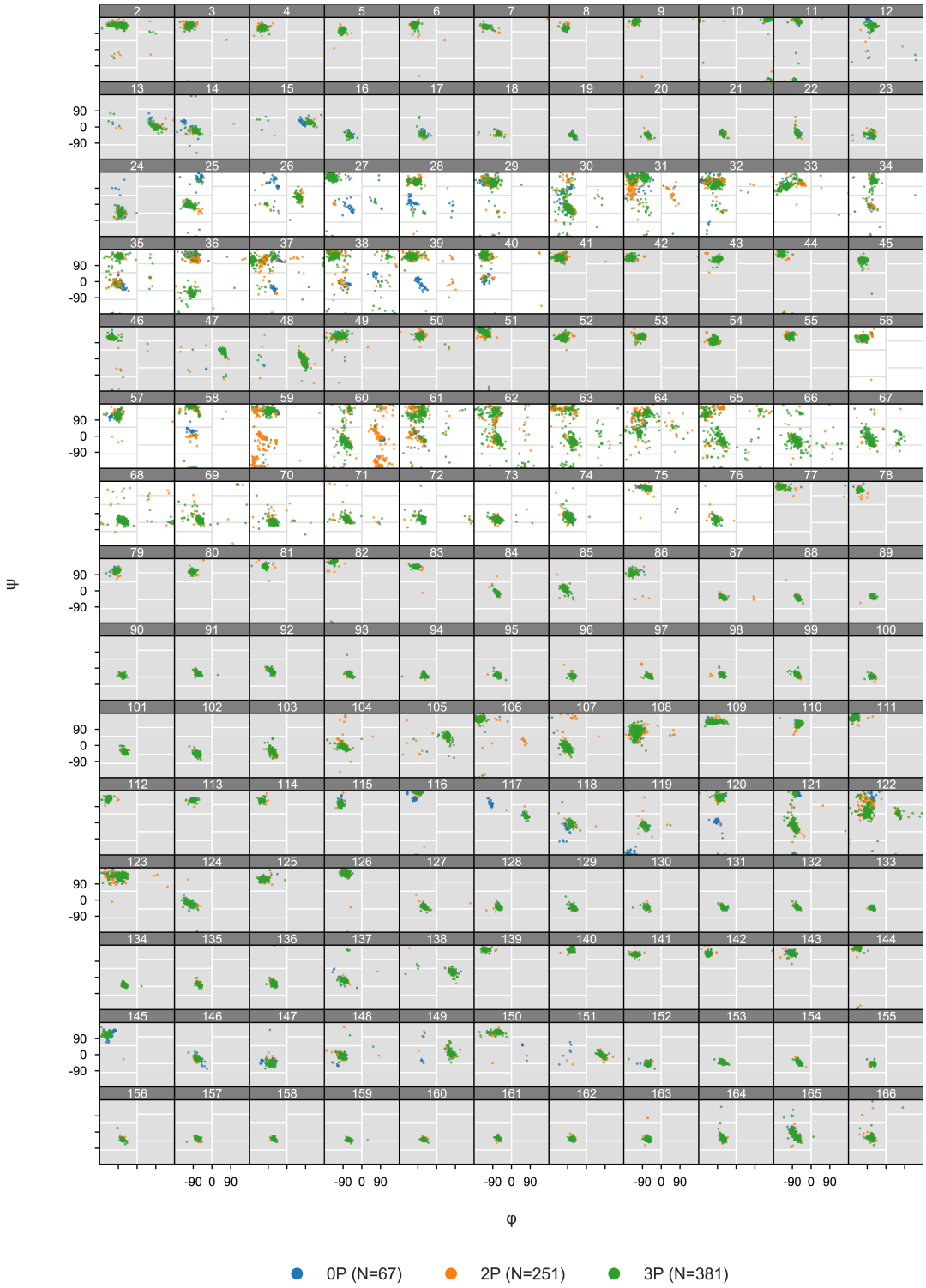
Conformationally Clustering SW1 and SW2. We clustered well-modeled SW1 and SW2 loops using the DBSCAN algorithm with a dihedral-based distance metric. DBSCAN finds major clusters and removes outliers[1]. Similar to a previous study[2], we ran DBSCAN across a grid of parameters and applied a set of quality control filters to generate a robust consensus clustering. This procedure was necessary since a single setting of DBSCAN cannot identify all possible clusters due to their varying shapes, densities, and sizes. Below, we list and explain the steps involved in our conformational clustering pipeline:

1. **Run DBSCAN on Well Modeled Loops.** Different parameters of DBSCAN can produce slightly divergent clustering results with merging, splitting, or disappearance of clusters. Therefore, we ran DBSCAN across a grid of parameters D=0.1-1.6 for ε (~20-80°): with steps of 0.1 and minimum samples 3-15 with steps of 1 and, following, took the consensus of these clustering results. The ε range covers the smallest regional subdivision of the Ramachandran map that residues with similar dihedrals can belong to[3]. We found this range to be ideal for conformational clustering in a past study[2].
2. **Find Passing Clusters Across Runs.** Not all DBSCAN clustering runs produce ideal separation of clusters. Therefore, we used two quality control filters to remove non-optimal clusters across runs before performing the consensus procedure: (a) mean silhouette score and (b) maximum dihedral distance:
    a. Silhouette score is a measure incorporating the similarity of an object to members of its own cluster (cohesion) and difference from other clusters (separation)[4]. The score can range from -1 (poor match to cluster) to 1 (good match to cluster). We removed clusters with mean silhouette score less than 0.6.
    b. Some larger DBSCAN parameters can merge similar conformational clusters that are separate conformations. Therefore, removed clusters with a maximum dihedral distance greater than D=3.75 (~150°). Clusters with points this far apart tend to be a mix of two Ramachandran regions and therefore unsuitable for our purposes.
3. **Get Union of Similar Clusters Across Runs.** Upon removal of poor clusters through quality filters, we found the union of clusters across runs with a Simpson similarity score greater than 0.9. The Simpson similarity score is the number of points two clusters have in common divided by the size of the smaller cluster. In most cases, in our DBSCAN runs, a cluster at one value of ε is often a subset of another cluster (Simpson score of 1.0) at larger ε, as outlying points of the cluster are incorporated.
4. **Merge Clusters with Close Loop Cα-RMSD.** The maximum dihedral metric is highly sensitive to peptide flips that an author may accidentally structurally model but does not signify a different conformation. This often happens at low resolution when the electron density can be modeled in two different ways. Therefore, we merged clusters with a loop Cα-RMSD less than 1.2 Å, which we found to be an appropriate cutoff through trial and error, by testing a range of 0.5-2.0 Å and visualizing the outputted results.
5. **Prune Cluster Members.** At certain DBSCAN settings, some bordering structures can find their way into clusters that visually appear to be outliers. In consequence, we created a step for pruning cluster members with a nearest neighbor dihedral distance greater than D=0.45 (~40°), which covers half of the smallest regional subdivision of the Ramachandran map, and loop Cα-RMSD greater than 1.2 Å.
6. **Remove Small Clusters.** Since a wide range of DBSCAN parameters are traversed during clustering, some small conformational clusters can be identified that a) either have no functional or binding corollaries or b) are duplicates from a single study or set of experimental conditions. To filter out these unmeaningful conformations, we removed structures possessing less than seven chains or found in less than five PDB entries.
7. **Classify Poorly Modeled Loops.** Since the poorly modeled loops were not included in clustering, we used a reversal of the pruning approach to assign them to clusters. We only classified poorly modeled loops if their NN dihedral distance was less than 0.45 (~40°) or loop Cα-RMSD was less than 1.2 Å in reference to a single conformational cluster. This approach can be used to conformationally classify the

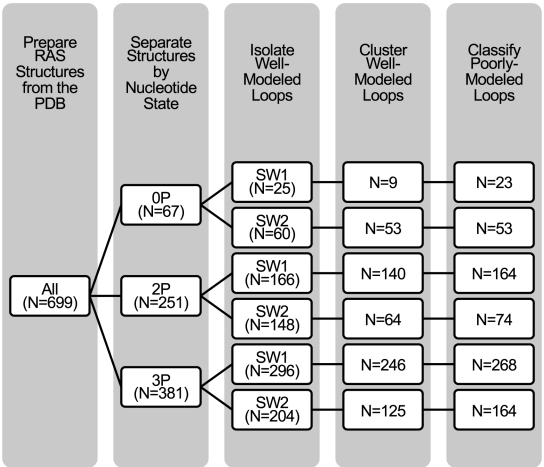
additional RAS structures that will be experimentally solved and deposited to the PDB, or ones produced through computational simulations.

Inhibitor Chemistry Search. We used the following canonical SMILES strings retrieved from PubChem[5] in searching for inhibitor chemistry with RDKit:

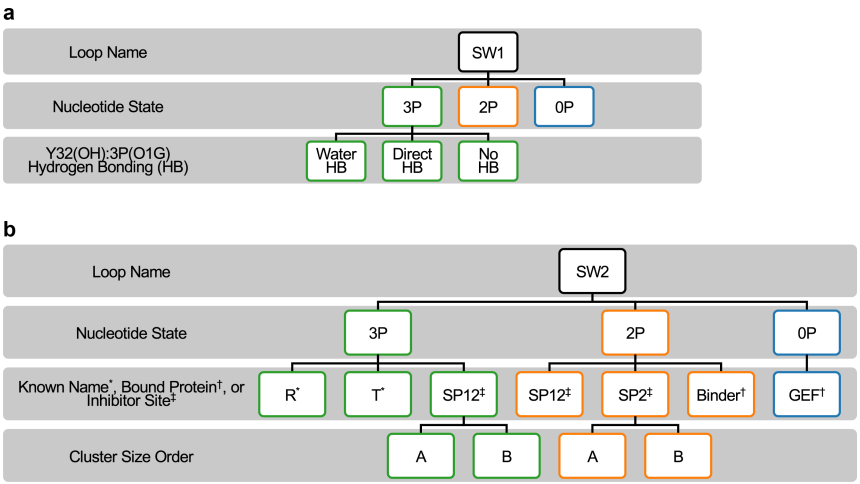
- **SW2 pocket (SP2)**
  - *Acrylamide*
    - CC(=O)N1CCNCC1 (1-Acetylpiperazine)
    - CC(=O)N1CCC1 (1-Azetidin-1-YL-ethanone)
    - CC(=O)N1CCCC1 (N-Acetylpyrrolidine)
  - *Sulfonamide*
    - CCS(=O)(=O)N (Ethanesulfonamide)
    - CC(=O)N1CCCCC1 (1-Acetylpiperidine)
- **SW1/SW2 Pocket (SP12)**
  - *Indole*
    - C1=CC=C2C(=C1)C=CN2 (Indole)
  - Benzodioxane
    - C1COC2=CC=CC=C2O1 (1,4-Benzodioxane)
  - *Biphenyl*
    - C1=CC=C(C=C1)C2=CC=CC=C2 (Biphenyl)



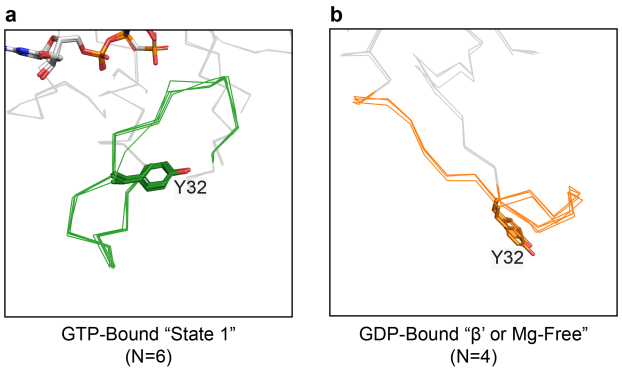
Supplementary Fig. 1 | Ramachandran map (φ versus ψ backbone dihedrals) for available RAS structures in the PDB. Includes the RAS GTP-binding (G) domain (residues 1-166). SW1 (residues 25-40) and SW2 (residues 56-76) loops highlighted.



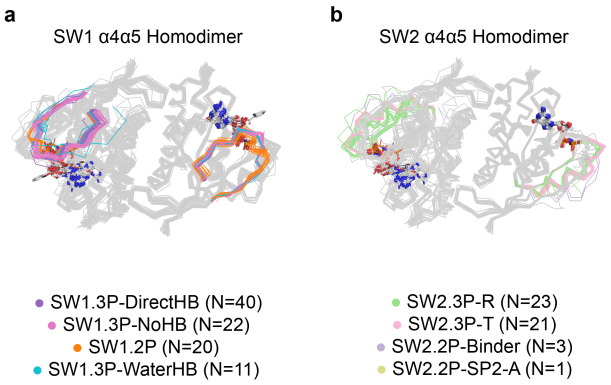
Supplementary Fig. 2 | RAS conformational clustering pipeline. Total RAS structures at each step of the conformational clustering pipeline.



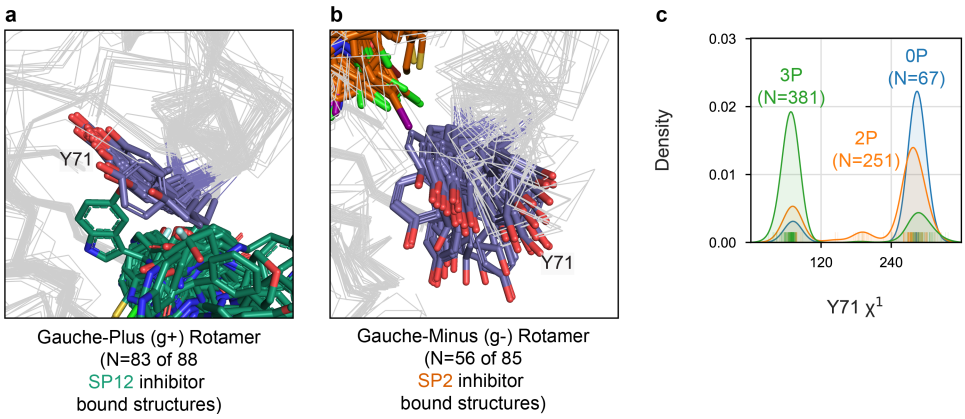
Supplementary Fig. 3 | RAS conformational classification and nomenclature. a, SW1 and, **b**, SW2 conformational clusters and naming system.



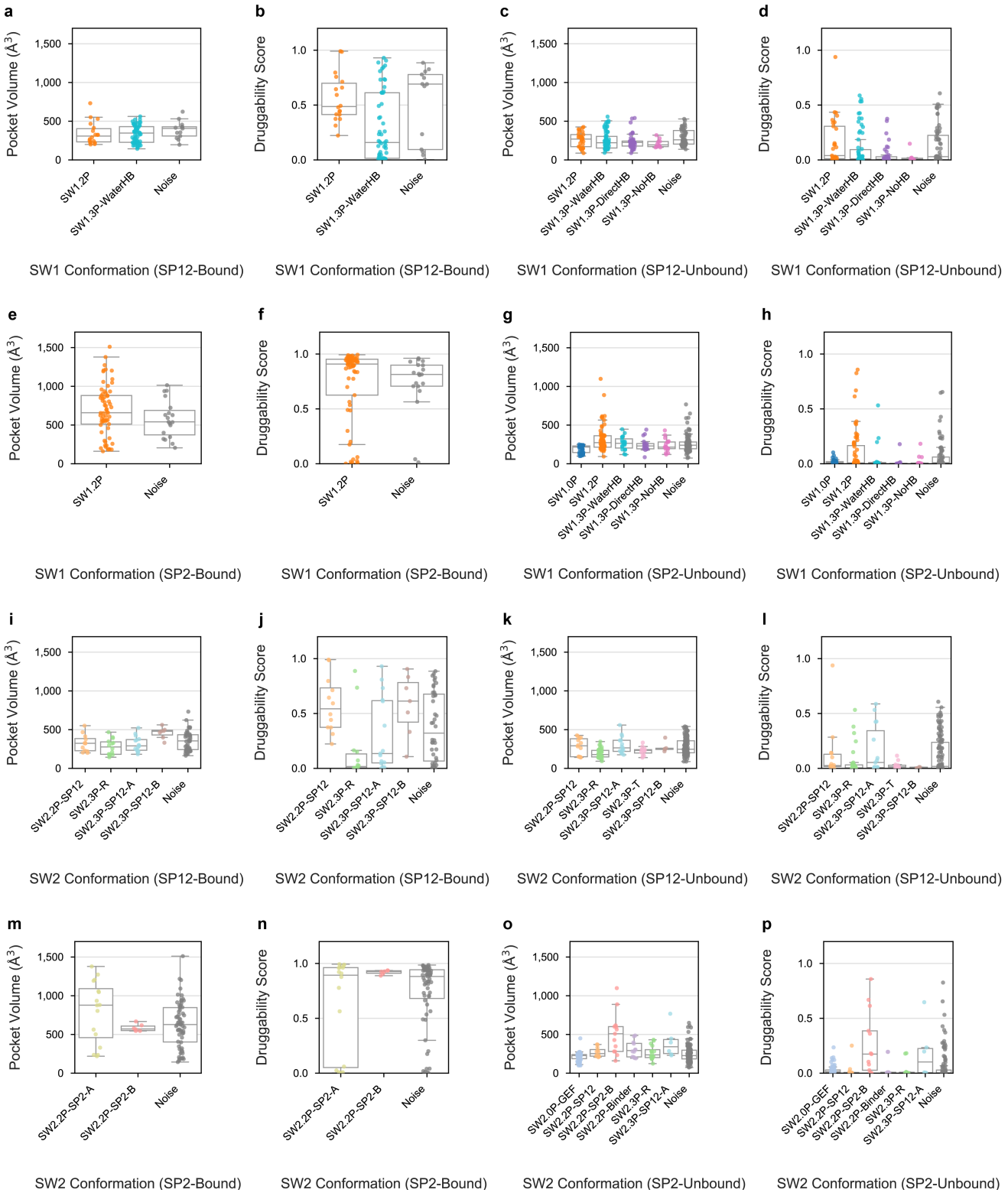
Supplementary Fig. 4 | SW1 conformations described previously that were not clustered. a, GTP-bound "state 1" (PDB: 1XCM, 3KKN, 4EFL, 4EFM, 4EFN, 6BP1) and, **b**, GDP-bound "β' or Mg-free" (PDB: 6BOF, 6M9W, and 6MQG).



Supplementary Fig. 5 | SW1 and SW2 conformations involved in RAS α4α5 homodimerization. a, SW1 and, **b**, SW2 conformations in RAS structures from the PDB that form the α4α5 homodimer.



Supplementary Fig. 6 | Importance of Y71 Position for Inhibitor Binding. a, Gauche-plus (g+) rotamer, and, **b**, gauche-minus (g-) rotamer for residue Y71 in SP12 and SP2 inhibitor bound structures, respectively. **c**, Y71 rotamer for nucleotide-free (0P), GDP-bound (2P), and GTP-bound (3P) nucleotide states.



Supplementary Fig. 7 | Pocket volumes and druggability scores for SP12 and SP2 bound and unbound structures. Pocket volumes and druggablity scores for SW1 conformations with SP12 inhibitor sites that are bound, **a** and **b**, respectively, and unbound, **c** and **d**, respectively; for SW1 conformations with SP2 inhibitor sites that are bound, **e** and **f**, respectively, and unbound, **g** and **h**, respectively; for SW2 conformations with SP12 inhibitor sites that are bound, **i** and **j**, respectively, and unbound, **k** and **l**, respectively; for SW2 conformations with SP2 inhibitor sites that are bound, **m** and **n**, respectively, and unbound, **o** and **p**, respectively. Only conformations with more than three structures with pockets are included

Supplementary Table 1 | Summary of SW1 and SW2 conformational clusters.

Cluster Name		N	PDB	CF	HRAS	KRAS	NRAS	Mean NN Dihedral Distance	Mean NN Loop Cα-RMSD
SW1.0P		23	23	1	23			8.1°	0.1 Å
Noise		44	37	8	33	11		89.7	3.6
SW1.2P		164	96	45	20	138	6	15.2	0.6
Noise		87	49	31	17	70		96.0	6.7
SW1.3P		268	173	40	130	136	2	12.8	0.5
Noise		113	69	30	45	66	2	79.9	3.6
SW2.0P	**GEF**	53	53	4	47	6		9.9	0.3
	Noise	14	7	6	9	5		87.1	2.3
SW2.2P	**SP12**	22	17	3		22		11.5	0.4
	SP2-A	20	8	6	1	19		20.8	0.8
	SP2-B	19	9	6		17	2	15.2	0.4
	Binder	13	5	4		13		9.9	0.4
	Noise	177	110	59	36	137	4	80.8	2.9
SW2.3P	**R**	98	84	24	65	30	3	15.2	0.6
	SP12-A	31	16	6	1	30		23.1	0.3
	T	21	21	1	21			16.3	0.3
	SP12-B	14	14	2		14		22.3	0.6
	Noise	217	124	48	88	128	1	81.1	1.9

Note. N = structures; PDB = entries;
CF = crystal forms; NN = nearest neighbor.

Supplementary Table 2 | Distribution of SW1 and SW2 conformations by α4α5 homodimerization status.

Cluster Name		α4α5 Homodimer	None	All
SW1.0P			23	23
Noise		6	38	44
SW1.2P		20	144	164
Noise		10	77	87
SW1.3P	**WaterHB**	11	144	155
	DirectHB	40	39	79
	NoHB	22	12	34
	Noise	31	82	113
SW2.0P	**GEF**		53	53
	Noise	6	8	14
SW2.2P	**SP12**		22	22
	SP2-A	1	19	20
	SP2-B		19	19
	Binder	3	10	13
	Noise	26	151	177
SW2.3P	**R**	23	75	98
	SP12-A		31	31
	SP12-B		14	14
	T	21		21
	Noise	60	157	217
All		140	559	699

Supplementary Table 3 | Distribution of SW1 and SW2 conformations by inhibitor binding site and chemistry.

Cluster Name		SP12				SP2			Multiple	Other	None	All
		BEN	BIP	IND	UNCL	ACR	SUL	UNCL				
SW1.0P											23	23
Noise					2					7	35	44
SW1.2P				8	8	49	5	4	1		89	164
Noise				2	1	4	9	3			68	87
SW1.3P	**WaterHB**	21	12	13	4			1	3		101	155
	DirectHB			2	1						76	79
	NoHB										34	34
	Noise	3	4	2	2			6			96	113
SW2.0P	**GEF**				1					3	49	53
	Noise				1					4	9	14
SW2.2P	**SP12**			5	5						12	22
	SP2-A					11					9	20
	SP2-B					6					13	19
	Binder										13	13
	Noise			5	4	36	14	7	1		110	177
SW2.3P	**R**	1		7	4						86	98
	SP12-A	11	1	2							17	31
	SP12-B	3	4								7	14
	T										21	21
	Noise	9	11	8	3			7	3		176	217
All		24	16	27	18	53	14	14	4	7	522	699

Note. BEN=benzodioxane; BIP=biphenyl; IND=indole; ACR=acrylamide; SUL=sulfonamide; UNCL=unclassified

Supplementary Table 4 | Distribution of SW1 and SW2 conformations by mutation status.

Cluster Name		WT	G12D	G12V	G12C	G12A	G13D	Q61H	Other	All
SW1.0P		23								23
Noise		29		1	9		1		4	44
SW1.2P		26	9	26	65	4	2		32	164
Noise		14	8	14	30		4		17	87
SW1.3P	**WaterHB**	35	32	1			1	50	36	155
	DirectHB	27	3	13	5	2		7	22	79
	NoHB	14	3	4			3	1	9	34
	Noise	32	13	16		4	7	1	40	113
SW2.0P	**GEF**	45			5		1		2	53
	Noise	7		1	4				2	14
SW2.2P	**SP12**	8		7	2		2		3	22
	SP2-A			8	11				1	20
	SP2-B	2	2	5	6				4	19
	Binder	2	1	8		2				13
	Noise	28	14	12	76	2	4		41	177
SW2.3P	**R**	23	18	12	1		3	1	40	98
	SP12-A	1	7					22	1	31
	SP12-B	1		1				12		14
	T	10							11	21
	Noise	73	26	21	4	6	8	24	55	217
All		200	68	75	109	10	18	59	160	699

Supplementary Data

Supplementary Data 1 | Available human KRAS, NRAS, and HRAS structures in the PDB annotated by SW1 and SW2 conformations and molecular contents.

Supplementary References

1. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* 226–231 (AAAI Press, Portland, Oregon, 1996).
2. Kelow, S.P., Adolf-Bryfogle, J. & Dunbrack, R.L. Hiding in plain sight: structure and sequence analysis reveals the importance of the antibody DE loop for antibody-antigen binding. *MAbs* **12**, 1840005 (2020).
3. Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95-9 (1963).
4. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **20**, 53-65 (1987).
5. Kim, S. et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research* **49**, D1388-D1395 (2021).