Online Methods for

Universal allosteric mechanism for Gα activation by GPCRs

Tilman Flock[1,*], Charles N. J. Ravarani[1,#], Dawei Sun[2,3#], A. J. Venkatakrishnan[1,$], Melis Kayikci[1], Christopher G. Tate[1], Dmitry B. Veprintsev[2,3] and M. Madan Babu[1,*]

[1]*MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK*
[2]*Laboratory of Biomolecular Research, Paul Scherrer Institut, Villigen, Switzerland*
[3]*Department of Biology, ETH Zurich, Zurich, Switzerland*

*E-mail: tflock@mrc-lmb.cam.ac.uk OR madanm@mrc-lmb.cam.ac.uk
[#]Equal contribution
[$]*Current address: Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA*

Online Methods

Generation of sequence and structure datasets

Identification of relevant Gα protein structures. All structures related to the Gα protein family (Pfam family: PF00503) were collected from Pfam (release 27.0) and Ensembl[39] using the R biomaRt interface[40]. In addition, the identified 973 Gα homolog sequences (see below) were aligned against the entire PDB database using the BLAST algorithm[41] to ensure all Gα-containing structures were identified. 91 PDB entries (146 Gα chains) were identified, of which two were obsolete (2PZ3 retracted, 2EBC superseded by 3UMR). Crystallographic data and coordinate files were retrieved from the RCSB Protein Database (PDB) API (Tuesday Feb 04, 2014 at 4 PM PST). Gα structures from the parasite *Entamoeba histolytica* (4FID) and *Arabidopsis thaliana* (2XTZ), as well as non-full length Gα (1AQG and 1LVZ are solution NMR studies of the C-terminal helix of Gα, 3RBQ contains the 11aa N-terminal part of transducin bound to UNC119) were excluded from the analyses. Four structures of the last 10 C-terminal amino acid residues of Gαt bound to rhodopsin (2X72, 3DQB, and 3PQR) or meta-rhodopsin (4A4M) were used for the GPCR-Gα interface analysis. Five PDB entries had no publication associated and were manually traced back to their original articles: 3UMR was published in Johnston et al[42] and 4G5O, 4G5Q, 4G5R, 4G5S were discussed in a study by Jia et al[43]. The final set of structures in our analyses span orthologs from human, mouse, rat, and cow and encompass twelve different Gα genes from eight different Gα subfamilies (GNAI1, GNAI3, GNAO, GNAS2, GNAT1, GNAQ, GNA12, GNA13), thereby representing all Gα families (Gαs, Gαi, Gαq, and Gα12). A full list of all retrieved PDBs is provided in **Supplementary Table 1**.

Identification of canonical human Gα protein sequences and paralog alignment. All relevant human Gα protein isoforms and variants were obtained from Ensembl[39] using R (full list in **Supplementary Table 1**). The 'canonical' protein sequences for each of the 16 human Gα genes, as defined by Uniprot[44], were used as representative sequences for each human Gα gene throughout this work. The sequences were aligned using Muscle[45] and were manually refined using the consensus secondary structure as a guide (see below). Phylogenetic relationships of Gα were obtained from Treefam[46] (family TF300673). The cladogram of the 16 canonical human Gα protein alignment was built with the Phylogeny.fr web service[47] choosing the PhyML v3.0 algorithm[48] with the SH-like Approximate Likelihood-Ratio Test using the Jones-Taylor-Thornton substitution matrix and TreeDyn[49] for visualization.

Ortholog alignments of one-to-one Gα orthologs of 16 human Gα genes. Phylogenetic relationships of Gα sequences were collected from TreeFam[46], the Orthologous MAtrix (OMA) database[50] and EnsemblComparaGeneTrees (Compara)[51] using R scripts. Compara had the highest fraction of complete Gα sequences for each human Gα gene, except for Gαs, for which OMA had a better sequence coverage. In total, 973 genes from 66 organisms were used, of which 773 were one-to-one orthologs. To build an accurate, low-gap alignment of such a number of sequences, 16 independent orthologous alignments for each human Gα gene were first created by aligning one-to-one ortholog groups using the PCMA algorithm[52] followed by manual refinement. Subsequently, each ortholog alignment was cross-referenced to the CGN (see below) by referencing its respective human sequence to the human paralog alignment. Conservation scores of each CGN position were calculated using both sequence identity and sequence similarity, based on the BLOSUM62 substitution matrix, (**Supplementary Note**) using all complete sequences of the cross-referenced alignments (561 sequences). Sequence conservation was mapped onto PDB structures (**Supplementary Data**) and visualized by generating PDB files with b-factors substituted by conservation scores.

Phylogenetic distances. The evolutionary distance of the retrieved sequences relative to human was evaluated with TimeTree[53]. Gα one-to-one orthologs extend back to Chordate (sea squirts; *Ciona savignyi, Ciona intestinalis* for Gα15), separated around 722.5 million years from *Homo sapiens*, and the most ancestral one-to-many ortholog extends back to Opisthokonta (yeast; *S. cerevisiae*), separated

by 1,215 million years from human. In this work, we only investigated G proteins from organisms that have a GPCR/G-protein system. Since plants do not encode GPCRs and the heterotrimeric G proteins are known to be auto-activated, we did not consider the plant G proteins in our analysis.

Development of a Common Gα Numbering (CGN) system

Common Gα numbering system. Comparative analyses of different protein structures and sequences to infer general principles of a protein family require a way of relating structural, genomic, or experimental data from different studies to each topologically equivalent position on homologous proteins. For GPCRs, the Ballerstors-Weinstein (BW) numbering scheme[54] enables the referencing of positions in the transmembrane helices of different GPCRs, not considering loop regions. We sought to develop a common G protein numbering (CGN) system that includes loop regions and describes Gα residues in three levels of detail (D.S.P), similar to a postal address. D refers to the structural domain and is optional (catalytic GTPase domain: G; helical domain: H), S stands for one of the 37 consensus secondary-structure elements (including loops) of the conserved Gα topology, and P relates to the corresponding residue position within the consensus secondary structure element mapped to an alignment of all 'canonical' human Gα sequences (**Extended Figure 1**). For a detailed guide of how to use the CGN and map any Gα protein, please refer to the CGN webserver (http://www.mrc-lmb.cam.ac.uk/CGN) and **Supplementary Note**.

Mapping structures to Uniprot sequences. Since several Gα protein structures represent chimeric G proteins, have peptide tags, or contain point mutations, each residue/position in the PDB structures was mapped to its Uniprot sequence(s) using the Structure Integration with Function, Taxonomy and Sequence (SIFTS)[55] webserver followed by a manual validation for missing positions. This allowed assigning residue positions of each Gα structure to their equivalent positions in the human paralog alignment and the ortholog alignments.

Determination of domain D and consensus secondary structure S. Secondary structure assignments were calculated for each Gα structure with the STRIDE algorithm[56]. The consensus secondary structure elements (SSE) were determined by considering the most prominent secondary structure type at each topologically equivalent Gα position when comparing the secondary structure assignment of all 80 Gα structures (mean and standard deviation of secondary structure type at each CGN position were calculated). Topologically equivalent positions had a high agreement in their SSE assignment and showed well-defined flanking regions (**Supplementary Note**). In addition, the assigned consensus SSEs were manually confirm through a 3D-structure alignment using MUSTANG[57], from which the domains (G-domain and H-domain) were defined. The Gα-SSE nomenclature follows a standardized expansion of the previously defined nomenclature[58]: Capital H and S represent helices or sheets, respectively. SSEs of the G domain follow a numerical identifier (H1, H2, …, H5 and S1, S2, …, S6 with the exception of HG), SSEs of the H domain have an alphabetical identifier (HA, HB, … HF), starting from the N- to the C-terminus of Gα. The N-terminal region that forms a membrane-anchored helix is defined as HN. Systematic identifiers for historical names of some loop regions (switch regions, P-loop, etc.) were derived by concatenating the flanking SSE names using lower case; for instance s6h5 refers to the loop between S6 and H5. A reference table including the historical loop names is provided in **Extended Figure 1** and **Supplementary Table 2**.

Determination of position P. P describes the amino acid position within an SSE, as determined by mapping the consensus secondary structure to the human paralog alignment (**Extended Figure 1** and **Supplementary Table 2**). Insertions in orthologs are annotated P-i, where i stands for the number of inserted residues after position P, for instance Arg334$^{H4.27-2}$ for the second amino acid of an insertion after helix H4 found in Pufferfish (*Tetraodon nigroviridis*) Gαs (**Supplementary Note**).

Consensus non-covalent contact networks between conserved residues

Non-covalent residue contact networks (RCNs). Non-covalent contacts between residues of a protein define its topology, conformation, and stability. For each of the 80 Gα protein structures, a local version of the RINerator 0.5 package from 2014[59] was used to calculate H-bonds and van der Waals interactions between residues. Matrices of the all-against-all atomic distances of all residue contacts within each structure were computed using R and the bio3d package[60]. Non-canonical interaction such as pi-pi stacking were identified with NCI[61]. All other calculations, analysis, and processing were performed in R.

Assignment of Gα structures to signaling states. Structural differences between Gα seem to arise from a convolution of the conformational state, binding partner, and Gα protein type and species (**Supplementary Note**). To identify which non-covalent contacts of each PDB are crucial for each signaling state independently of the Gα protein type and species, all Gα structures were assigned to one of the four different Gα signaling states depending on (a) the bound ligand, and (b) the interaction partner (**Supplementary Table 1**). The four states are (1) heterotrimeric GDP-bound state (inactive state), (2) nucleotide-free heterotrimeric receptor bound complex (GEF-bound state), (3) GTPγS and potentially 'effector'-bound state (active state), and (4) RGS-bound GDP+ALF hydrolysis transition state (GAP-bound state). Eleven structures represent the inactive state, one full-length structure (and four structures of the C-terminal Gα peptide in complex with a GPCR) the GPCR-bound state, 25 have GTPγS bound or/and are co-crystallized with their downstream effectors and 40 structures have Gα in the GTP-hydrolysis transition state with GDP and aluminum fluoride (ALF) bound (GDP+ALF) and/or are co-crystalized with their RGS or a GTP-hydrolysis promoting peptide mimicking the RGS binding interface (e.g. Go-Loco motif). Two structures (1CIP, 1SVS) had non-standard Gα ligands bound, and 2ZJZ[62] did not have a detailed description of its biochemical relevance, and thus were not assigned to any signaling state. Eleven structures were identified as chimeras and 21 included mutations (**Supplementary Note**). The publication of each PDB was checked to confirm the relevance of the assigned signaling state.

Consensus contacts between conserved residues. To compare residue contact networks (RCNs) from different structures, topologically equivalent positions were cross-referenced with the CGN system. All RCN analyses, consensus RCN calculation, and conservation analysis were conducted using customized R scripts: Matrices representing the absence or presence of non-covalent contacts between each possible pair of CGN residues in each PDB structure were computed (**Supplementary Figure S1**). The consensus contacts of each signaling state were computed as the probability to find a contact in all structures of the state. Since structure models can differ in their number of equivalent residues due to missing electron density, not fully fitted models, or truncations for crystallographic purposes, each consensus contact probability was normalized by the number of structures of the state that have the respective residue pair, in order distinguish the absence of a contact from the absence of an equivalent position in a single PDB. To expand the structural analysis to other Gα proteins with only sequence data available, sequence conservation was mapped to each CGN residue (see above).

Visualization of consensus contacts and identification of universal structural motifs. The consensus contacts between conserved residues in the different signaling states were visualized to investigate the contact re-organization in detail. For 2D visualization, the respective consensus RINs were exported to Cytoscape[63] using the RCytoscape interface[64]. For 3D visualization, R was used to create consensus RCNs in PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.) by creating pseudo PDBs that show residues as spheres from their C-alpha atoms and lines/edges between them via the CONECT entries. Information on sequence conservation was mapped via the B-factor field of the pseudo PDB structures. For simplification, only contacts present in more than 90% of all structures with a sequence identity > 90% were shown as 'consensus contacts' between conserved residues – this threshold was chosen based on the bimodal distribution of contact occurrence (**Supplementary Figure S2**). In addition, only long-range interactions (>i+4) were shown for the consensus RINs. It is important to note that these cut-offs were just applied for visualization, while for

the analysis, no cut-off was needed and all relevant consensus contacts were additionally visually inspected in all 80 PDB structures by creating automated PyMol sessions from R that superimpose all 80 structures. To generate RINs between SSE, the sums over all contact of the respective SSE as defined by the consensus SSE of the CGN were computed. Chimera[65] was used to manually re-evaluate atomic contacts, and PyMol was used to create publication-quality images.

Interface analysis

Buried surface area and inter-Gα-GPCR residue contact networks. Inter-chain RCNs between Gα and the receptor (Gαs and β_2-Adrenergic Receptor (β_2AR) chains A and R in 3SN6, Gαt C-terminal peptide and rhodopsin from chains B and A in 2X72, 3DQB, 3PQR, 4A4M) were calculated as described above. The buried surface area (BSA) was obtained from the PDBe PISA (Proteins, Interfaces, Structures, and Assemblies)[66] XML repository and normalized by the accessible surface area for each residue position. BSA and Gα-GPCR RCNs were mapped to the CGN and the Ballosteros-Weinstein numbering, respectively. Sequence conservation from 561 complete Gα homolog sequences and 249 human non-olfactory class A GPCRs was mapped onto the interface to determine the conserved 'hotspot' residues in the interface and visualized in PyMol (**Supplementary Data**). The BSA histogram, the visualization of the residue interaction network per secondary structure elements, and the correlation of BSA per residue vs. conservation were produced in R and ggplot2.

Force field-based energy estimations. The per-residue energy contributions to Gα monomer and Gα–GPCR complex stability were calculated using FoldX 3.0, which uses energy terms weighted by empirical data from protein engineering experiments to provide a quantitative estimation of each residue contribution to protein stability and protein complex stability (http://foldx.crg.es/). For the interface analysis, the 3SN6 structure was energy minimized with the FoldX 'repair pdb' function and subsequently, the per-residue energy contributions for both the Gαs-β_2-AR complex and the monomers in isolation were calculated using the FoldX 'sequence detail', 'analyse complex', and 'stability' functions at 298K, pH 7.0, and 0.05M ion strength. The per-residue energy contributions to complex stability were calculated as the difference between the energy contributions of each residue in the monomer and complex ($\Delta\Delta G_{interface}$) and visualized with R (**Extended Figure 2a**). For energy contributions of each residue within Gα monomers (**Extended Figure 4b)**, the average energy contribution and standard deviation for each Gα position was computed after running the FoldX 'stability' and 'sequence detail' functions at 298K, pH 7.0, and 0.05M ion strength for each of the 79 non-complex structures.

Disorder propensity calculations for all Gα homolog sequences and structures. The disorder propensity of each of the 561 complete Gα homolog sequences was calculated with IUPred[67] (prediction-type-setting: 'short disorder'). The missing structure positions were identified with bio3d package[60] (**Extended Figure 2b**).

New and published mutational studies of different Gα classes

Identification of mutations, mutant structures and chimeras. Additional literature on Gα mutations was retrieved with the text mining tool MutationMapper[68] and manually validated and filtered for correct hits/search results. Disease mutations were retrieved from the Database of Single Nucleotide Polymorphisms (dbSNP) and the Catalogue of Somatic Mutations in Cancer (COSMIC)[69] with biomaRt[40], and from the Human Gene Mutation Database (HGDM)[70]. Mutations, chimeras and peptide tags in the analyzed structures were identified by comparing their Uniprot sequence to their PDB sequence using SIFTS[55] and mining PDBe annotations. All mutation data were mapped back to the CGN and visualized on their respective human Gα structure.

Alanine scanning and stability of Gai. The alanine scanning expression library of Gαi1 was prepared as reported before[71]. The recombinant Gαi1 alanine mutants were expressed in 24 well plates, purified by standard Ni-NTA affinity chromatography followed by buffer exchange using 96-well filter plates.

The bovine rhodopsin and βγ subunits were prepared from bovine retinas[72]. The melting temperature of each alanine mutant upon addition of GDP or GTPγS was measured by differential scanning fluorimetry assay. The effect of each alanine Gαi1 mutant on R*-Gi complex formation and complex stability were measured by the HTP assay based on native gel electrophoresis method. Detailed methods and protocols are provided in the accompanying paper by Sun et al[25].

Gα versus Ras comparison

The Ras conformational cycle was featured in the RSCB PDB[73] April 2012 PDB-101 Molecule of the Month by David Goodsell (doi: 10.2210/rcsb_pdb/mom_2012_4), with high-resolution structures showing human HRas in its active GTPγS bound state (PDB-ID: 5P21[74]) and the GDP-bound inactive state (PDB-ID: 4Q21[75]). 1BDK[76] was used as representative of the HRas GEF-bound state. These Ras representative structures were combined with an alignment of all human HRas paralogs identified from the OMA database[50]. A structural alignment between the identified active and inactive Ras structures and the corresponding active and inactive Gα (1GOT and 3UMS) was used to accurately map Ras positions to the CGN (**Supplementary Data**) despite the low sequence identity (<6%) between Ras and Gα. The number of atomic non-covalent contacts between helix H5 and helix H1 in the Ras and Gα structures was manually compared in Chimera for the structures 1GOT and 4Q21.

References

39 Flicek, P. *et al*. Ensembl 2014. *Nucleic Acids Res* **42**, D749-755, doi:10.1093/nar/gkt1196 (2014).

40 Kasprzyk, A. BioMart: driving a paradigm change in biological data management. *Database (Oxford)* **2011**, bar049, doi:10.1093/database/bar049 (2011).

41 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

42 Johnston, C. A. *et al*. Structural determinants underlying the temperature-sensitive nature of a Galpha mutant in asymmetric cell division of Caenorhabditis elegans. *J Biol Chem* **283**, 21550-21558, doi:10.1074/jbc.M803023200 (2008).

43 Jia, M. *et al*. Crystal structures of the scaffolding protein LGN reveal the general mechanism by which GoLoco binding motifs inhibit the release of GDP from Galphai. *J Biol Chem* **287**, 36766-36776, doi:10.1074/jbc.M112.391607 (2012).

44 Magrane, M. & Consortium, U. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* **2011**, bar009, doi:10.1093/database/bar009 (2011).

45 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

46 Ruan, J. *et al*. TreeFam: 2008 Update. *Nucleic Acids Res* **36**, D735-740, doi:10.1093/nar/gkm1005 (2008).

47 Dereeper, A. *et al*. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**, W465-469, doi:10.1093/nar/gkn180 (2008).

48 Guindon, S. *et al*. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

49 Chevenet, F., Brun, C., Banuls, A. L., Jacq, B. & Christen, R. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* **7**, 439, doi:10.1186/1471-2105-7-439 (2006).

50 Altenhoff, A. M., Schneider, A., Gonnet, G. H. & Dessimoz, C. OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res* **39**, D289-294, doi:10.1093/nar/gkq1238 (2011).

51 Vilella, A. J. *et al*. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19**, 327-335, doi:10.1101/gr.073585.107 (2009).

52 Pei, J., Sadreyev, R. & Grishin, N. V. PCMA: fast and accurate multiple sequence alignment based on profile consistency. *Bioinformatics* **19**, 427-428 (2003).

53 Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971-2972, doi:10.1093/bioinformatics/btl505 (2006).

54 Ballesteros, J. A. & Weinstein, H. *Receptor Molecular Biology*. Vol. 25 (Elsevier, 1995).

55 Velankar, S. *et al*. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Res* **41**, D483-489, doi:10.1093/nar/gks1258 (2013).

56 Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* **32**, W500-502, doi:10.1093/nar/gkh429 (2004).

57 Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J. & Lesk, A. M. MUSTANG: a multiple structural alignment algorithm. *Proteins* **64**, 559-574, doi:10.1002/prot.20921 (2006).

58 Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 A crystal structure of transducin-alpha complexed with GTP gamma S. *Nature* **366**, 654-663, doi:10.1038/366654a0 (1993).

59 Doncheva, N. T., Klein, K., Domingues, F. S. & Albrecht, M. Analyzing and visualizing residue networks of protein structures. *Trends Biochem Sci* **36**, 179-182, doi:10.1016/j.tibs.2011.01.002 (2011).

60 Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695-2696, doi:10.1093/bioinformatics/btl461 (2006).

61 Babu, M. M. NCI: A server to identify non-canonical interactions in protein structures. *Nucleic Acids Res* **31**, 3345-3348 (2003).

62 Morikawa, T. *et al*. Crystallization and preliminary X-ray crystallographic analysis of the receptor-uncoupled mutant of Galphai1. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **63**, 139-141, doi:10.1107/S1744309107003363 (2007).

63 Lopes, C. T. *et al*. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26**, 2347-2348, doi:10.1093/bioinformatics/btq430 (2010).

64 Shannon, P. T., Grimes, M., Kutlu, B., Bot, J. J. & Galas, D. J. RCytoscape: tools for exploratory network analysis. *BMC Bioinformatics* **14**, 217, doi:10.1186/1471-2105-14-217 (2013).

65 Pettersen, E. F. *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

66 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

67 Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434, doi:10.1093/bioinformatics/bti541 (2005).

68 Vohra, S. & Biggin, P. C. Mutationmapper: a tool to aid the mapping of protein mutation data. *PLoS One* **8**, e71711, doi:10.1371/journal.pone.0071711 (2013).

69 Forbes, S. A. *et al*. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805-811, doi:10.1093/nar/gku1075 (2015).

70 Stenson, P. D. *et al*. The Human Gene Mutation Database: 2008 update. *Genome Med* **1**, 13, doi:10.1186/gm13 (2009).

71 Sun, D. *et al*. AAscan, PCRdesign and MutantChecker: a suite of programs for primer design and sequence analysis for high-throughput scanning mutagenesis. *PLoS One* **8**, e78878, doi:10.1371/journal.pone.0078878 (2013).

72 Maeda, S. *et al*. Crystallization scale preparation of a stable GPCR signaling complex between constitutively active rhodopsin and G-protein. *PLoS One* **9**, e98714, doi:10.1371/journal.pone.0098714 (2014).

73 Berman, H. M. *et al*. The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000).

74 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**, 2351-2359 (1990).

75 Milburn, M. V. *et al*. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* **247**, 939-945 (1990).

76 Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337-343, doi:10.1038/28548 (1998).