Table of Contents

Supplementary Note 1

Accessing Protein Contacts Atlas

Protein Contacts Atlas can be accessed online at http://www.mrc-lmb.cam.ac.uk/pca. The main page provides two ways to access data: (i) entering a PDB ID or (ii) uploading co-ordinates in PDB format. If the user is interested in a structure but does not know the corresponding PDB ID, they can perform a keyword search of the PDBe database[1]; if the user is interested in a domain rather than a specific protein, they can specify a PFAM domain and select from a list of structures containing that domain. Both of these options can be accessed through the "Advanced Features" link at the top of the main page. Users can also download a text file containing all the atomic contacts (per PDB file) for all the structures containing a PFAM domain of interest. The user also has a number of different options to parameterize, filter and/or choose specific contact types for visualization and analysis in Protein Contacts Atlas (see **Methods** for details).

Interactive visualization panels for sequence, 3D structure and contacts

Clicking on any node (to view contacts within one chain) or link (to view contacts between two chains) in the biomolecular complex network redirects to the “Visualization and Analysis” page. This has three interlinked panels, displaying representations of the sequence, 3D structure, and non-covalent contacts (**Supplementary Fig. 2**).

Sequence panel: When a user clicks on a chain, its secondary structure elements (SSE) and sequence information is displayed in the sequence panel at the top of the page. The first line shows each of the secondary structures (helices and sheets but without loops) in distinct colors and provides a quick snapshot of the SSEs in the chain. Clicking on the individual SSEs (or residues) updates the relevant information in the other two panels. The second line shows the same information with loops and is aligned with the protein sequence, which is displayed in the third line. If the user clicks on an interface from the biomolecular complex network, then the first line shows the SSEs for both chains, only highlighting those that form residue contacts at the interface. The subsequent lines present similar information about SSEs with loops and their amino acid sequence for both chains. Besides the types of data explained above, naturally occurring human single nucleotide polymorphisms and disease mutations data from Uniprot are also displayed in the sequence panel. Detailed information on the SNP/mutation can be viewed for the residues in the selected chain and structure by moving the mouse over the residue.

3D structure panel: The 3D structure panel is by default on the right side of the page, and shows a cartoon representation of the 3D structure of the selected chain(s) rendered using WebGLmol (http://webglmol.osdn.jp/index-en.html). Clicking on an SSE or an individual residue in the sequence panel highlights the relevant information in the cartoon representation (**Supplementary Fig. 2**). The residues of interest (e.g. a position mutated in human disease) can be selected and highlighted on the 3D structure using different representations (e.g. spheres, sticks or network). These options can be selected by clicking on the buttons below the 3D structure panel. In the 3D residue contact networks representation, the backbone of the structure is shown as a Cα trace with the residues as small spheres. Upon selecting a residue, their interacting neighbors are displayed as a network. The contacts are represented as edges and

their thickness denotes the number of atomic contacts between residues, thereby providing an estimate of the extent of packing between contacting residues. While we only display contacting residues that are at least 4 residues apart in the protein sequence, all the non-covalent contacts are calculated and are available as a downloadable file. As an example we show Leu388 on the last helix of Gαs, which interacts with the Adenosine A2a receptor[2], a G-protein coupled receptor (GPCR) (PDB: 5G53). The contacts between L388 and I200 and A204 are intuitively visualized in the network representation as shown in **Fig. 3a.** Clicking on the "Deselect and reset to default" icon on the contacts panel displays the entire structure as a 3D residue contact network. The user can also directly click on any residue of interest in the 3D structure panel to view and obtain information about the residue of interest (e.g. chain, residue name and position). Any view and representation that is shown in the 3D structure panel can be downloaded by clicking on the "Save as PyMOL File" link and opening this file in PyMOL will reproduce the highlighting and orientation of the structure. A publication-ready image of the 3D structure can also be created and downloaded by taking a snapshot (camera icon; **Supplementary Fig. 2** within the 3D structure panel).

Contacts panel: The contacts panel has three tabs. The first two directly represent the structure of the contact network, using a chord plot, residue contact matrix, and asteroid plot; the third represents network metrics calculated for each residue (see main text for details). In this panel, there are options to download detailed atom-level contact information for the entire structure as a text file, a file that is compatible with the Cytoscape[3] network viewer (http://www.cytoscape.org/) (*Downloadable Data* section), and publication-quality images of the representation on the screen. If the user updates any of the contact panels by clicking on either the SSE in the chord plot, individual residues in the asteroid plot or groups of residues in the scatter plot, then the sequence panel and the 3D structure panel are automatically updated. The three different panels are thus linked, allowing the interactive visualization and analysis of distinct types of information and entities in an integrated way.

Supplementary Note 2

Here, we highlight the different features of Protein Contacts Atlas using various biomolecular structures as case studies.

Multi-level visualization of non-covalent contacts

We highlight the multi-level visualization of non-covalent contacts in the context of a protein-protein interaction (between the Adenosine A2a receptor and an engineered mini Gαs protein[2]) and a protein-nucleic acid interaction (between p53 and DNA[4]). The Adenosine A2a receptor is an important protein that is expressed in the brain and in the immune system, and couples to intracellular signaling pathways through Gαs. This interaction is critical to ensure fidelity in downstream response when the receptor is activated by endogenous ligands. Protein Contacts Atlas allows easy visualization and identification of the key secondary structures and the important residue and atomic contacts that mediate this interaction (**Fig. 3**). Analyzing the interface of chain A (GPCR) and C (Gαs protein) of the complex shows that Helix 39 of Gαs contains residues that interact with Helix 11 of the A2a receptor (**Fig. 3a;** PDB: 5G53). The residue contact matrix view highlights residues that engage in contacts with the A2a receptor (**Fig. 3a**). The multi-level visualization of non-covalent contacts in Protein Contacts Atlas provides a non-covalent contact based context for generating testable hypotheses for understanding the molecular mechanisms of disease-associated mutations causing pseudo-hyperthyroidism and Albright hereditary osteodystrophy. For example, G protein position $L388^{G.H5.20}$ (superscript denotes common G protein numbering system[5]) of Helix 39 contacts positions $A204^{5.65}$ and $I200^{5.61}$ (superscripts denote GPCRdb numbering[6,7]) on the Helix 11 (TM5) of the receptor. A mutation of 388 (L->R) causes the disease pseudo-hypoparathyroidism[8]. Similarly, G protein position $R385^{G.H5.17}$ of Helix 39 contacts Q207 on the Helix 11 of the receptor. A

variation of position R385 to H decouples the receptor interaction from its effector, causing Albright hereditary osteodystrophy[9]. Correspondingly, Protein Contacts Atlas also enables analysis of contacts with nucleic acids, including both DNA and RNA. p53 is a DNA binding transcription factor and an important tumour suppressor gene. In p53, mutation of a critical Arginine residue (R273) to a Histidine has been implicated in loss of its tumour suppressive function and this position is also recurrently mutated in many somatic cancers[10-12]. The interaction of p53 with its cognate DNA sequence is critical for its function. Protein Contacts Atlas provides a context for the visualization of non-covalent contacts between this residue and the phosphate group of the DNA backbone, providing a molecular interpretation of why a mutation in this position results in disease; i.e. by directly disrupting protein-DNA interactions (**Fig. 3b;** PDB: 4MZR).

Asteroid plot enables analysis of the ligand-receptor interaction in β2 adrenergic receptor

Beta-blockers are important, intensive care drugs that are given to protect the heart from a second heart attack after an initial attack. They act through the beta-adrenergic receptors. Here, we illustrate how the beta-blocker carazolol acts at its target (**Fig. 4a**; PDB 2RH1) through the asteroid plot representation of the crystal structure of human β2 adrenergic receptor[13]. Asteroid plot (**Fig. 4b**) reveals the contacting residues, and the second-shell residues that may stabilize/assist these in ligand binding. Coloring them by secondary structure elements reveals the importance of the residues in the different helices (and loops) in binding the ligand. It highlights that residues D113$^{3.32}$ (GPCRdb numbering) and N312$^{7.38}$ from TM helices 3 and 7 make the largest number of direct contacts with the ligand in the first shell. Such representation also provides the basis for generating hypotheses regarding the molecular basis of disease. For instance, mutations have been observed in T164$^{4.56}$, which has been linked with asthmatic patients; while this residue does not directly contact the ligand, it does contact T118$^{3.37}$, which makes a direct contact with the ligand. In other words, T164$^{4.56}$ is a second-shell residue whose mutation might affect ligand binding indirectly. We can also note from the ligand-residue interaction matrix that T118$^{3.37}$ makes a contact with the C11 atom of the ligand (**Fig. 4c**). Therefore, in patients carrying the mutation at T164, the ligand-mediated effects of the mutation may manifest allosterically through T118.

Scatter plot matrix allows quantitative analysis of per-residue properties

A number of mutations in rhodopsin are associated with human diseases. 68 of the known human disease mutations in rhodopsin[14] were mapped on to the bovine rhodopsin structure (1GZM) and visualized as scatter plot matrix, shown in **Supplementary Fig. 3**. Clicking on the name of the property (*x* and *y* axes labels) colors the structure in the 3D structure panel according to their values as a colour spectrum (cyan, low to magenta, high), revealing residues with extreme values for a property of interest (i.e. outliers).

Interpreting stability measurement of point mutations in G protein

G proteins are the key intracellular adaptors by which each of the ~800 human GPCRs communicate with diverse intracellular signaling pathways to bring a physiological response. G proteins undergo nucleotide exchange upon binding to the activated receptor and regulate effector molecules. Thus studying the residues that contribute to G protein stability is critical for understanding how G proteins function. We highlight this feature by mapping experimental data from a recent study that mutated each of the 354 positions in Gαi to alanine and quantified the stability of each mutant in the presence of GDP[15]. This study therefore provides the contribution of each residue position to stabilizing the G protein in the GDP-bound conformation. **Supplementary Fig. 4** shows a scatter plot matrix after mapping this external information via a colour spectrum. A key residue Phe336$^{G.H5.8}$ (text in superscript refers to Common G protein Numbering[5]) was identified to be important for G protein activation. Its mutation to alanine increases the rate of spontaneous GDP release and results in substantial loss of Gαi stability and complete impairment of its ability to bind nucleotides. This can be rationalized using Protein Contacts Atlas because Phe336$^{G.H5.8}$

makes a large number of contacts and may thus have a critical role in stabilizing the Gαi subunit in the nucleotide-bound conformation (**Supplementary Fig. 4b**). Mutating this residue to alanine appears to destabilize nucleotide binding by affecting the stability of the G protein, possibly by mimicking GDP release triggered by interaction of GPCRs with G proteins[5,16].

References

1. Velankar, S. et al. PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res* (2015).
2. Carpenter, B., Nehme, R., Warne, T., Leslie, A.G. & Tate, C.G. Structure of the adenosine A(2A) receptor bound to an engineered G protein. *Nature* **536**, 104-7 (2016).
3. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-504 (2003).
4. Emamzadah, S., Tropia, L., Vincenti, I., Falquet, B. & Halazonetis, T.D. Reversal of the DNA-binding-induced loop L1 conformational switch in an engineered human p53 protein. *J Mol Biol* **426**, 936-44 (2014).
5. Flock, T. et al. Universal allosteric mechanism for Galpha activation by GPCRs. *Nature* **524**, 173-9 (2015).
6. Isberg, V. et al. Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends Pharmacol Sci* **36**, 22-31 (2015).
7. Isberg, V. et al. GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res* **44**, D356-64 (2016).
8. Thiele, S. et al. Functional characterization of GNAS mutations found in patients with pseudohypoparathyroidism type Ic defines a new subgroup of pseudohypoparathyroidism affecting selectively Gsalpha-receptor interaction. *Hum Mutat* **32**, 653-60 (2011).
9. Schwindinger, W.F., Miric, A., Zimmerman, D. & Levine, M.A. A novel Gs alpha mutant in a patient with Albright hereditary osteodystrophy uncouples cell surface receptors from adenylyl cyclase. *J Biol Chem* **269**, 25387-91 (1994).
10. Somers, K.D. et al. Frequent p53 mutations in head and neck cancer. *Cancer Res* **52**, 5997-6000 (1992).
11. Chehab, N.H., Malikzay, A., Stavridi, E.S. & Halazonetis, T.D. Phosphorylation of Ser-20 mediates stabilization of human p53 in response to DNA damage. *Proc Natl Acad Sci U S A* **96**, 13777-82 (1999).
12. Wright, J.D. & Lim, C. Mechanism of DNA-binding loss upon single-point mutation in p53. *J Biosci* **32**, 827-39 (2007).
13. Cherezov, V. et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-65 (2007).
14. Rakoczy, E.P., Kiel, C., McKeone, R., Stricher, F. & Serrano, L. Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations. *J Mol Biol* **405**, 584-606 (2011).
15. Sun, D. et al. Probing Galphai1 protein activation at single-amino acid resolution. *Nat Struct Mol Biol* **22**, 686-94 (2015).
16. Alexander, N.S. et al. Energetic analysis of the rhodopsin-G-protein complex links the alpha5 helix to GDP release. *Nat Struct Mol Biol* **21**, 56-63 (2014).