SUPLYMENTARY MATERIAL

Data Content

The GPCRRD contains sequences, mutations, electron microscopy, neutron diffraction, site-directed mutagenesis, FTIR, Disulfide Bridge and X-ray data that are regularly collected from the literature and from the primary sources [GPCRDB (Horn, et al., 2003), UniProt (Yip, et al., 2008), EMBL (Kanz, et al., 2005), TinyGRAP (Beukers, et al., 1999)]. Table S1 summarizes the statistics about the data content of the GPCRRD.

Table S1. Statistics of resources in GPCRRD.

Experiment Source	Information	Derived restraints	Number of restraints
Electron microscopy	2-D electron density projection map	Positions of helix centroids	10
Neutron diffraction	2-D projection map	Positions of helix centroids	2
Functional site-directed mutagenesis	Detection of residues that are functionally important	Orientation of the helix to position the mutated functional residue toward inside of the seven-helix bundle.	5588
FTIR difference spectroscopy	Detection of functionally active residues	Orientation of the helix to position functionally active residues inside the seven-helix bundle	16
Disulfide bridge detection	Detection of Cys residues participating	Position of the identified Cys residues next to each other	38
X-ray	3-D coordinates	Positions of heavy atoms	15

Data Collections

Experimental data are extracted from online literature and other resources using an automated procedure outlined in Figure S1. Documents are retrieved from the Medline database using the PubMed query system (Schuler, et al., 1996). Medline abstracts are used when full texts are not available. Pattern matching with regular expressions was used to identify point mutation data. The pattern must start with one amino acid in the one- or three-letter code followed by a number. The regular expression we use is:

$$([A-Z][1-9][0-9]+\$)\,|\,([A-Z][1-9][0-9]*[A-Z]\$)|$$
$$([A-Z][a-z][a-z][1-9][0-9]*\$)\,| \quad \text{(S1)}$$
$$([A-Z][a-z][a-z][1-9][0-9]*[A-Z][a-z][a-z]\$)$$

where [A–Z] and [A–Z][a–z][a–z] must belong to set of 20 amino acids in one- or three-letter code, respectively (Horn, et al., 2004). We then combined these data with those from the primary sources [GPCRDB (Horn, et al., 2003), UniProt (Yip, et al.,

2008), EMBL (Kanz, et al., 2005), TinyGRAP (Beukers, et al., 1999)] and the redundant entries were removed.

Three different filters were applied to validate the experimental information. Firstly, we applied sequence filter to check whether the wild-type amino acids in the extracted point mutations are found at the indicated positions in the corresponding sequences. Secondly, we used function filter to find the function related mutation. When the residue number of the mutation and the function related words, such as ability, mediate, select, agonist antagonist or binding etc., occur in the same sentence and there is no "not", we consider this mutation is functional mutation. Otherwise, we do not consider it as a function related residue. Finally, we manually validated those data which do not belong to the two categories. Electron microscopy, neutron diffraction, FTIR, Disulfide Bridge and X-ray data are collected and validated from the literature and from the original sources.

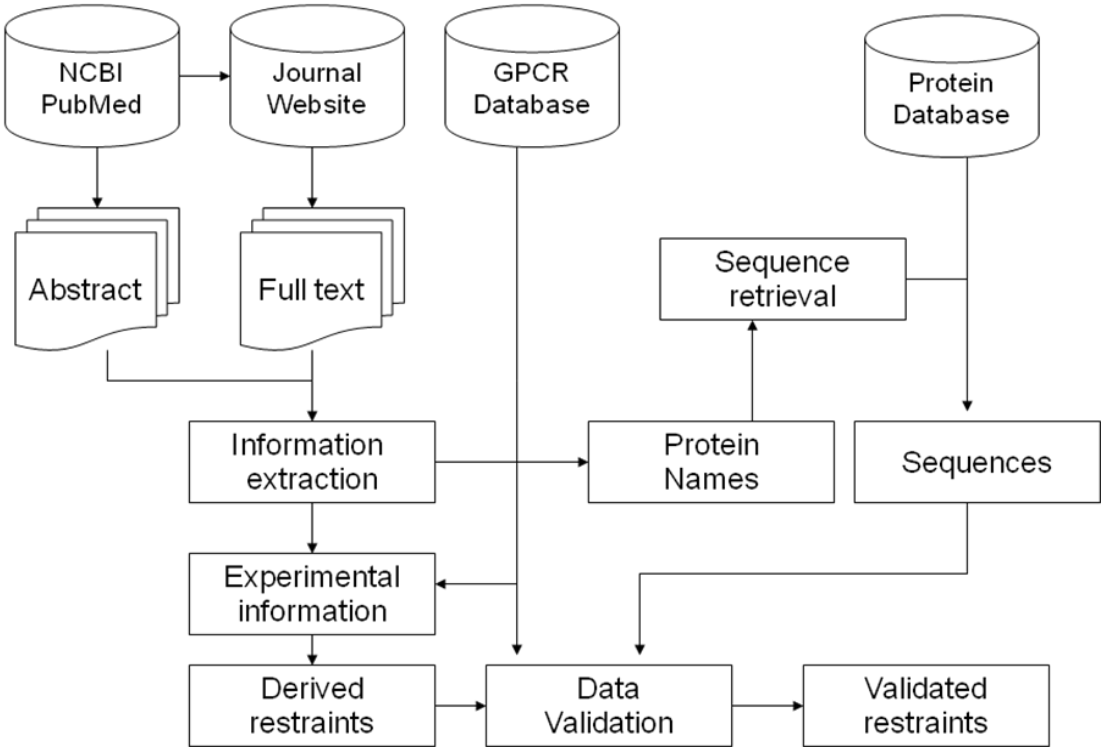

Figure S1. The flowchart for extracting experimental data and spatial restraints from literature and other resources.

Computationally Derived Data

Residue orientation. In order to derive the spatial restraints from the experimental data, we first predict the transmembrane helices using TMHMM 2.0 software package (Krogh, et al., 2001). If we assume that the orientation of mutated functional residues are towards inside of the seven-helix bundle (Schushan, et al., 2010), the orientation restraint of mutation data can be obtained as showed in Figure S2. It is intuitively understandable since hydrophobic residues of the TM helix are in close contact with the membrane wall and the interior part of the bundle is the only room left for ligand binding.

If we define a plane $O_aO_bO_c$, which is perpendicular to the transmembrane helix TM_b and passing through the alpha carbon atom of a query residue, we can have three points O_a, O_b and O_c, which are the intersections of plane $O_aO_bO_c$ and three axis of the neighboring transmembrane helices. These three points provide graphical representation of the lower and upper limits of the orientation restraint. If the alpha carbon atom of a query residue is positioned inside the angle Φ, low energy can be assigned to this restraint. Otherwise, high energy will be applied to penalize the outliers. The illustration of orientation restraint of muscarinic acetylcholine M1 receptor was shown in Figure S3. We can obtain the orientation restraint from mutation of D71N (Lameh, et al., 1992).

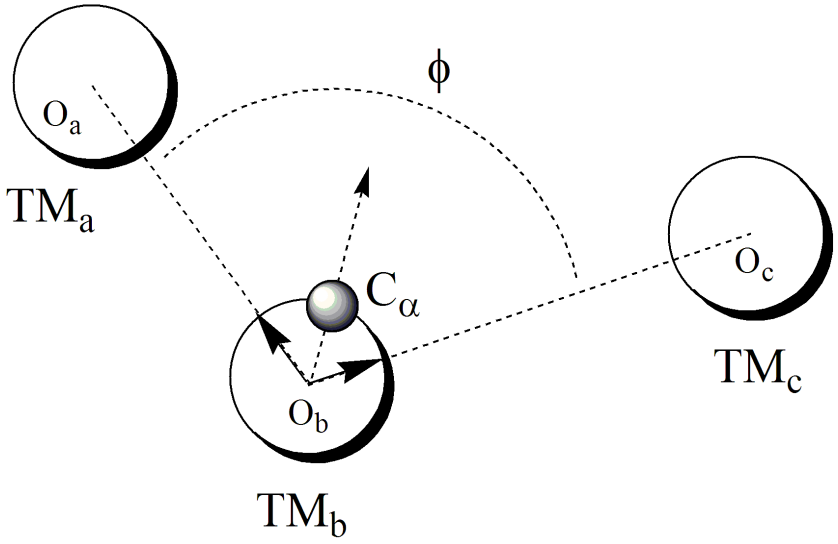

Figure S2. The illustration of orientation restraint.

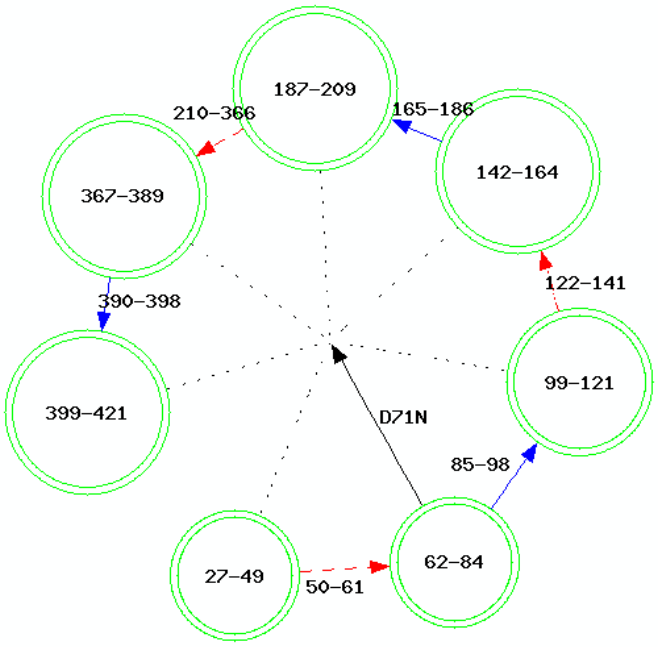

Figure S3. The illustration plot of orientation restraint.

Residue-residue contact restraints. We also collected and verified the ligand binding information based on the original published literature. This information together with mutagenesis data can be converted into the pair-wise side-chain contact restraints:

$$d_{ij} < D_{ij} \qquad \text{(S2)}$$

Here d_{ij} is the distance of side-chain centers of mass between the involved residue pairs. D_{ij} is the maximum distance of residue pairs estimated based on the mutagenesis experiments. For a known binding ligand, D_{ij} is equal to the maximum size of the ligand plus δ, where δ is a parameter to count for the van der Waals distances based on the ligand size. For unknown ligands, a uniform cutoff (~10Å) is used as a loose constraint. It is suggested that these parameters should be fine-tuned in the assembly simulations to tolerate the maximum uncertainty from the experimental data. For example, the contact restraint of Y179-W400 in muscarinic acetylcholine M1 receptor was obtained from primary literature (Ma, et al., 2009) since residues Y179 in TM5 and W400 in TM7 are involved in binding the same ligand (benzyl quinolone carboxylic acid).

The restraints of FTIR data can be generated in the same way as mutagenesis data. For example, orientation restraint FTIR E134 in human rhodopsin (opsd_human) can be converted from functional residue 134 (Vogel, et al., 2006).

Distance restraints. For the disulfide bridge, we can easily generate distance restraints according to geometry of disulfide bond. For example, disulfide bridge Cys124-Cys202 was detected in chicken P2Y purinoceptor 1 (p2ry1_chick) (Hoffmann, et al., 1999). The 2-D electron density maps of electron microscopy and neutron diffraction can be converted into 2-D position restraints in the membrane surface plane. For example, the 2-D electron density map of bovine rhodopsin (opsd_bovin), as shown in Figure S4, can be converted to 2-D coordinates (32.36 31.87 36.29 44.02 45.84 42.47 48.30 32.68 54.93 26.76 46.38 21.52 39.76 31.31) and position restraints (Unger and Schertler, 1995).

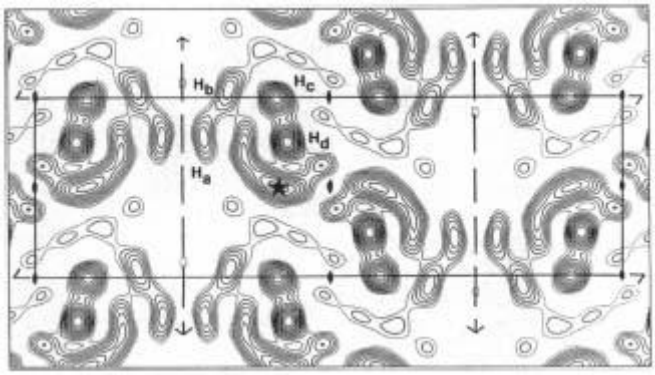

Figure S4. 2-D electron density map of bovine rhodopsin (opsd_bovin)

References

Beukers, M.W., Kristiansen, K., Ijzerman, A.P. and Edvardsen, O. (1999) TinyGRAP database: a bioinformatics tool to mine G-protein-coupled receptor mutant data, *Trends Pharmacol Sci*, **20**, 475-477.

Hoffmann, C., Moro, S., Nicholas, R.A., Harden, T.K. and Jacobson, K.A. (1999) The role of amino acids in extracellular loops of the human P2Y1 receptor in surface expression and activation processes, *J Biol Chem*, **274**, 14639-14647.

Horn, F., Bettler, E., Oliveira, L., Campagne, F., Cohen, F.E. and Vriend, G. (2003) GPCRDB information system for G protein-coupled receptors, *Nucleic Acids Res*, **31**, 294-297.

Horn, F., Lau, A.L. and Cohen, F.E. (2004) Automated extraction of mutation data from the literature: application of MuteXt to G protein-coupled receptors and nuclear hormone receptors, *Bioinformatics*, **20**, 557-568.

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Diez, F.G., Harte, N., Kulikova, T., Lin, Q., Lombard, V., Lopez, R., Mancuso, R., McHale, M., Nardone, F., Silventoinen, V., Sobhany, S., Stoehr, P., Tuli, M.A., Tzouvara, K., Vaughan, R., Wu, D., Zhu, W.M. and Apweiler, R. (2005) The EMBL nucleotide sequence database, *Nucleic Acids Res*, **33**, D29-D33.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes, *Journal of Molecular Biology*, **305**, 567-580.

Lameh, J., Philip, M., Sharma, Y.K., Moro, O., Ramachandran, J. and Sadee, W. (1992) Hm1 muscarinic cholinergic receptor internalization requires a domain in the third cytoplasmic loop, *J Biol Chem*, **267**, 13406-13412.

Ma, L., Seager, M.A., Wittmann, M., Jacobson, M., Bickel, D., Burno, M., Jones, K., Graufelds, V.K., Xu, G., Pearson, M., McCampbell, A., Gaspar, R., Shughrue, P., Danziger, A., Regan, C., Flick, R., Pascarella, D., Garson, S., Doran, S., Kreatsoulas, C., Veng, L., Lindsley, C.W., Shipe, W., Kuduk, S., Sur, C., Kinney, G., Seabrook, G.R. and Ray, W.J. (2009) Selective activation of the M1 muscarinic acetylcholine receptor achieved by allosteric potentiation, *Proc Natl Acad Sci U S A*, **106**, 15950-15955.

Schuler, G.D., Epstein, J.A., Ohkawa, H. and Kans, J.A. (1996) Entrez: Molecular biology database and retrieval system, *Computer Methods for Macromolecular Sequence Analysis*, **266**, 141-162.

Schushan, M., Barkan, Y., Haliloglu, T. and Ben-Tal, N. (2010) Cα-trace model of the transmembrane domain of human copper transporter 1, motion and functional implications, *Proceedings of the National Academy of Sciences*, **107**, 10908-10913.

Unger, V.M. and Schertler, G.F. (1995) Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy, *Biophys J*, **68**, 1776-1786.

Vogel, R., Siebert, F., Yan, E.C., Sakmar, T.P., Hirshfeld, A. and Sheves, M. (2006) Modulating rhodopsin receptor activation by altering the pKa of the retinal Schiff base, *J Am Chem Soc*, **128**, 10503-10512.

Yip, Y.L., Famiglietti, M., Gos, A., Duek, P.D., David, F.P.A., Gateau, A. and Bairoch, A. (2008) Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase, *Hum Mutat*, **29**, 361-366.