Structure preparation. After downloading each structure from the Protein Data Bank (PDB), missing atoms were added with Charmm22 all-hydrogen parameters (*1*). In cases where only a single chain was present in the PDB file but the oligomeric state was higher than 1, symmetry operations were used to build the other chains in the oligomer (while imposing a 5-Å cutoff). The oligomeric structure was used to define the transmembrane (TM) domain and to calculate the solvent accessibility of each residue. However, for each oligomeric protein, only a single chain (the first chain containing the TM domain) was used to count the number of glycines and other residues as well as for the backbone statistics.

For serving as references for calculating whole-residue and backbone accessibility, ideal poly-Gly, poly-Ala, and poly-Asp helices were generated using backbone dihedral angles of ϕ = -60° and ψ = -45°.

Solvent accessible surface areas (SASAs). Solvent accessible surface areas of individual atoms were calculated using the NACCESS program (*2*). Relative accessibility (i.e., percentage of the nominal maximum SASA, irrespective of secondary structure) was calculated for either a whole residue or for the backbone polar atoms (C, O, and N).

Sequence conservation scores. The ConSurf server (*3*) was used to generate conservation scores for individual residues. Default parameters were used for all the proteins except for 3EAM. For the latter the minimal sequence identity for sequence alignment was lowered from the default 35% to 25%. Residues with the top 20% conservation scores in each protein were referred to as conserved.

Locating membrane central plane. The membrane central plane was located such that the lipid-facing charged residues (Asp, Glu, Arg, and Lys) of the TM domain on the two sides of the central plane had maximal separations from it. 2Q7M, 2QTS, and 3HGC apparently had serious distortions (see the main text), which prevented a precise determination of the membrane central plane. For these structures, a membrane central plane is displayed in Fig. 1 based on visual inspection. For 2KYV, no charged residues were present on one side of the TM domain, thereby preventing the implementation of our algorithm, so the central plane was defined as the plane passing though the center of the TM domain helices and perpendicular to the oligomeric symmetry axis. For the other 27 structures (1FX8, 1L7V, 1P7B, 2A65, 2AHY, 2BL2, 2EI4, 2GIF, 2J7A, 2L0J, 2NS1, 2OAR, 2ONK, 2R9R, 2YVX, 2ZW3, 3DQB, 3DWW, 3EAM, 3HD6, 3M71, 3MKT, 3NCY, 3ND0, 3O7Q, 2KSF, and 2GFP), the membrane central plane was determined according to the following procedure.

(1) The principal axes were calculated, and the one generally parallels to the TM helices was aligned with the z axis (i.e., the membrane normal).

(2) For each structure, the list of charged residues was trimmed to keep only the surface exposed ones of the TM domain. First, all buried residues were eliminated (with the

following SASA criteria: whole residue, side chain, backbone, and polar portion all had ≤ 20% accessibility). Second, those in extramembranous domains, exposed to internal pores, or in loops far into the extracellular or intracellular space were removed. Removed residues were:

1L7V: residues 56, 59, 87, 144, 232, 291
1P7B: residues 106, 115, 130, 148
2A65: residues 30, 62, 67, 91, 112, 192, 274, 287, 288, 290, 369, 370, 385, 398, 401, 402, 404, 419, 435, 487
2AHY: residues 66, 77, 97
2BL2: residues 32, 39, 52, 112, 139
2EI4: residues 13, 87, 90, 101, 120, 139, 199, 209, 217, 221
2GIF: residue 567
2J7A: residues 57, 100, 117, 126, 131, 133, 134, 138, 141, 147, 149, 154, 155
2ONK: residues 84, 88, 92, 148, 149, 159, 183, 195, 199, 248
2NS1: residues 9, 121, 122, 160, 185, 194, 249, 309, 310, 313, 390, 395, 401
2R9R: chains B, D, F, and H TM rsidues 183, 220, 226, 236, 240, 259, 293, 296, 299, 302, 323, 346
2YVX: residues 45, 80, 91, 112, 224, 234, 246, 285, 340, 358, 359, 432
2ZW3: residues 2, 32, 41, 42, 47, 66, 75, 143, 147, 184, 187, 188, 209
3DH6: residues 65, 129, 145, 166, 201, 210, 218, 248, 298, 336, 432, 59, 123, 176, 177
3DQB: residues 83, 113, 122, 134, 135, 181, 190, 296
3DWW: residues 26, 60, 66, 70, 73, 75, 77, 110, 126
3EAM: residues 222 and 272
3M71: residues 62, 95, 123, 181, 189, 198, 228
3MKT: residues 11, 36, 91, 151, 155, 255, 301, 305, 371, 382, 385, 443
3ND0: residues 50, 60, 109, 127, 144, 197, 198, 286, 332, 395, 406, 409, 225, 228, 306, 309
3NCY: residues 67, 145, 149, 174, 208, 348, 349, 384, 408, 409
3O7Q: residues 46, 86, 135, 154, 283, 312, 379, 381, 415
2KSF: residue 467
2GFP: residues 33, 42, 118, 227, 268, 269.

The remaining charged residues formed two clusters, one on the extracullar side and one on the intracellualr side.

(3) The membrane central plane was represented by the following equation: $ax + by - z + d = 0$, and optimal values of the cofficients a, b, and d were determined by scanning. Specifically, a and b were scanned from -0.8 to 0.8, and d was scanned from $d_0 - 4$ to $d_0 + 4$, where d_0 was an initial guess located midway between the extracellular and intracellular clusters of charged residues. The increment for scanning each coefficient was 0.02. Among the 81 × 81 × 401 sets of coefficients, the optimization was selected according to three creteria (distances were calculated using C_α atoms):

(*i*) All the charged residues in the two clusters were at least 10 Å away from the central plane. For 2GIF, the threshold was reduced to 7 Å, and for 2NS1 and 2YVX, to 8 Å.
(*ii*) The difference between the average distance of the extracullar cluster of charged residues from the central plane and that of the intracelullar cluster was less than 1 Å.
(*iii*) After satisfying the above two criteria, the optimum was selected as the one that maximized the total of the distances of the two clusters of charges from the central plane.

The signed distance (positive toward the extracellualr side and negative toward the intracellular side) from the central plane of a residue is denoted as z. In collecting the statistics displayed in Fig. 3a, b, glycine and aspartate residues with $|z| < 10$ Å and exposed to internal pores were excluded. These include 5 glycines (residues 199 in 1FX8, 34 in 2L0J, 343 in 3ND0, and 395 and 402 in 3O7Q) and 2 aspartates (residues 404 in 2A65 and 46 in 3O7Q). In Fig. 3d, some additional residues with $|z| < 10$ Å and exposed to internal pores were excluded. These include Ala201 in 1FX8, Gly65 in 2AHY, Leu891 in 2GIF, Pro94 in 2ONK, Pro321 in 2YVX, Ala428 in 2YVX, Ala117 in 3DQB, Phe82 in 3DWW, Pro60 in 3MKT, Asn22, Ile23, Met24, Gly25, and Thr155 in 3NCY, and Ser279 and Gly394 in 3O7Q.

Helix-helix contacts. Two helices were defined as forming contacts if there was one or more inter-helical distances between heavy atoms that were < 5 Å and these contact atoms were located in the hydrophobic region (i.e., $|z| < 10$ Å). A glycine was considered as being found in forming helix-helix contacts when any heavy atom of the glycine was separated from a heavy atom in the partner helix by no more than 110% of the minimum distance between the helices.

1. A. D. Mackerell, M. Feig and C. L. Brooks, J Comput Chem, 2004, 25, 1400-1415.

2. S. J. Hubbard, J. M. Thornton, 1993, ‘NACCESS’, Computer Program, Department of Biochemistry and Molecular Biology, University College London.

3. N. Ben-Tal, H. Ashkenazy, E. Erez, E. Martz and T. Pupko, Nucleic Acids Res, 2010, 38, W529-W533.