Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Allosteric Na^+-binding site modulates CXCR4 activation

Xiaojing Cong*[a] and Jérôme Golebiowski*[a,b]

[a] Institute of Chemistry - Nice, UMR 7272 CNRS - University Côte d'Azur, Nice cedex 2, France.
[b] Department of Brain & Cognitive Sciences, DGIST, Dalseong-gun, Daegu, Republic of Korea.
* To whom correspondence may be addressed: xiaojing.cong@unice.fr or jerome.golebiowski@unice.fr

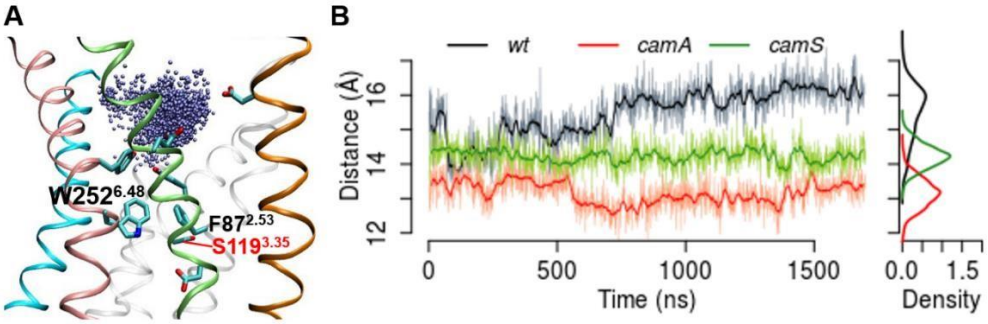

Figure S1. (A) Density of Na^+ ion within *camS* during 1.7 µs of brute-force MD simulations. **(B)** Time-series and distribution plots of the $F87^{2.53}$-$W252^{6.48}$ Cα distance during the brute-force MD, showing shrinkage of the Na^+ passage in the CAM receptors.

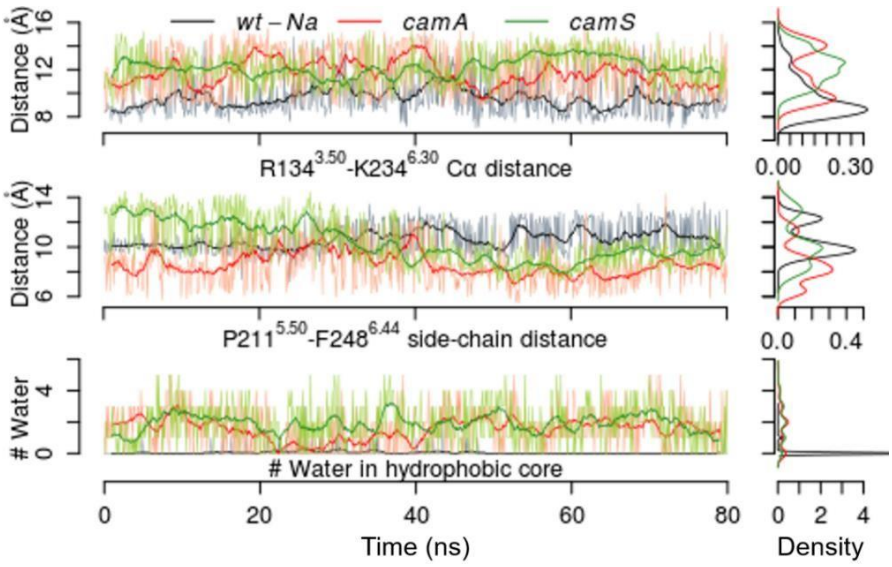

Figure S2. CAM receptors displayed active-like features during the REST2 MD, contrasting *wt-Na*. Top to bottom: intracellular cleft opening (as measured by $R134^{3.50}$-$K234^{6.30}$ Cα distance), transmission switch activation (as measured by $F248^{6.44}$-$P211^{5.50}$ side-chain distance), and hydration in receptor hydrophobic core (number of water molecules within the TM bundle below $S123^{3.39}$ and above $I230^{6.26}$).

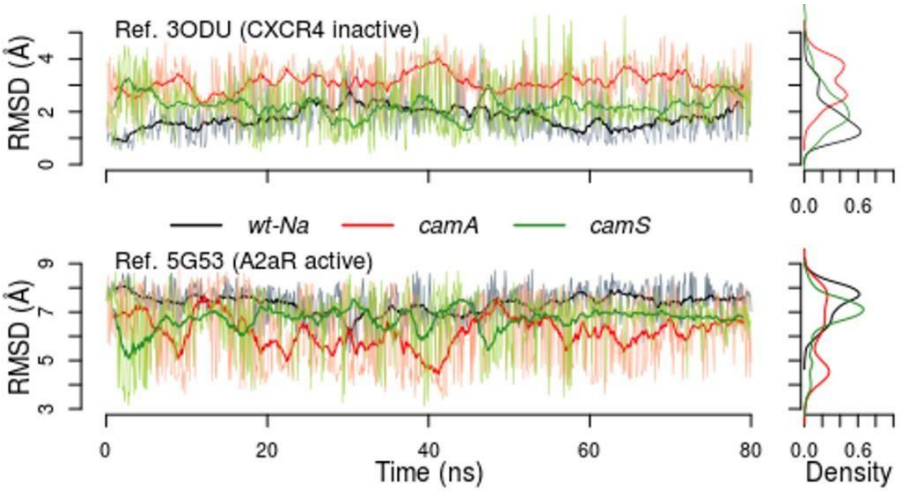

Figure S3. Cα RMSD of the TM6 intracellular end (residues 6.23–6.30) of the three receptors during REST2 MD with respect to that in the X-ray crystal structure of CXCR4 in inactive state (PDB ID: 3ODU) and A2aR in active state (PDB ID: 5G53). While *wt-Na* preserved the initial inactive state (top, black), the CAMs (red and green) showed regular outward movement of TM6 with the RMSD approaching the active state (bottom).

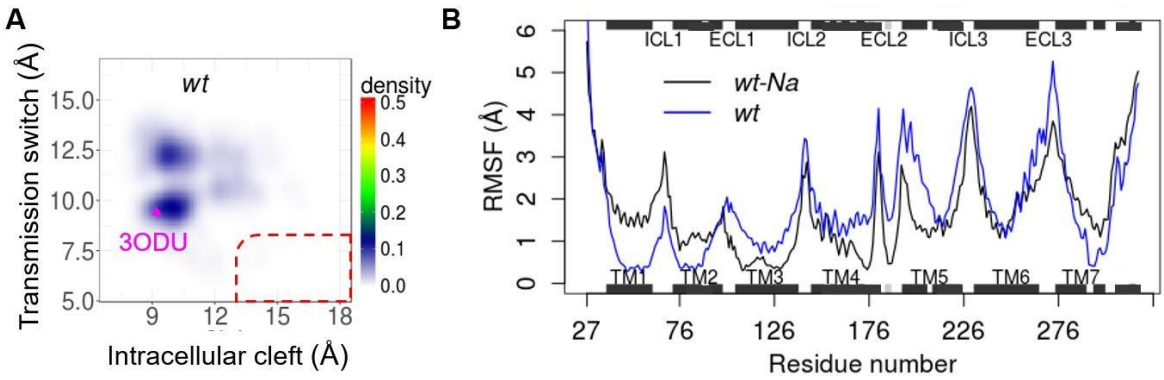

Figure S4. The *wt* receptor without allosteric Na^+ showed almost no active-like state (0.5% of trajectory) during the REST2 MD but more flexibility than *wt-Na* from TM3 to TM7. **(A)** Contour map of the transmission switch conformation (as measured by $F248^{6.44}$-$P211^{5.50}$ side-chain distance) and the intracellular cleft (as measured by $R134^{3.50}$-$K234^{6.30}$ Cα distance). **(B)** Cα's RMSF of *wt* (blue) versus *w-Na* (black).

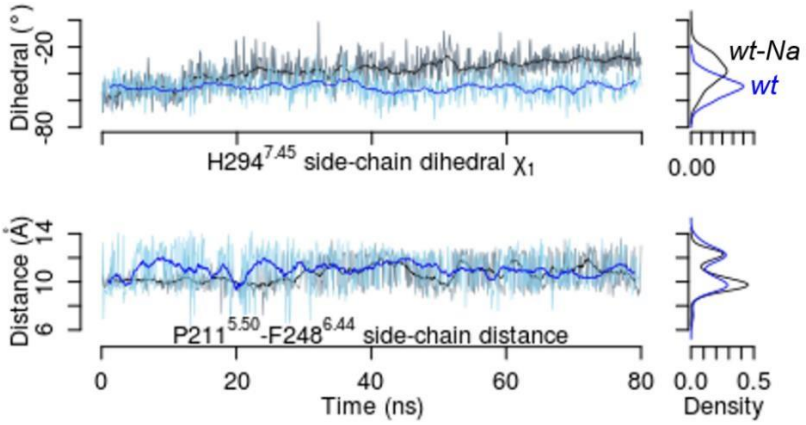

Figure S5. The *wt* receptor without allosteric Na^+ showed slight differences from *wt-Na* in the allosteric site and the transmission switch during the REST2 MD, especially in the orientation of $H294^{7.45}$ (as measured by the side- chain dihedral χ_1) and $F248^{6.44}$ (as measured by the side-chain distance to $P211^{5.50}$).

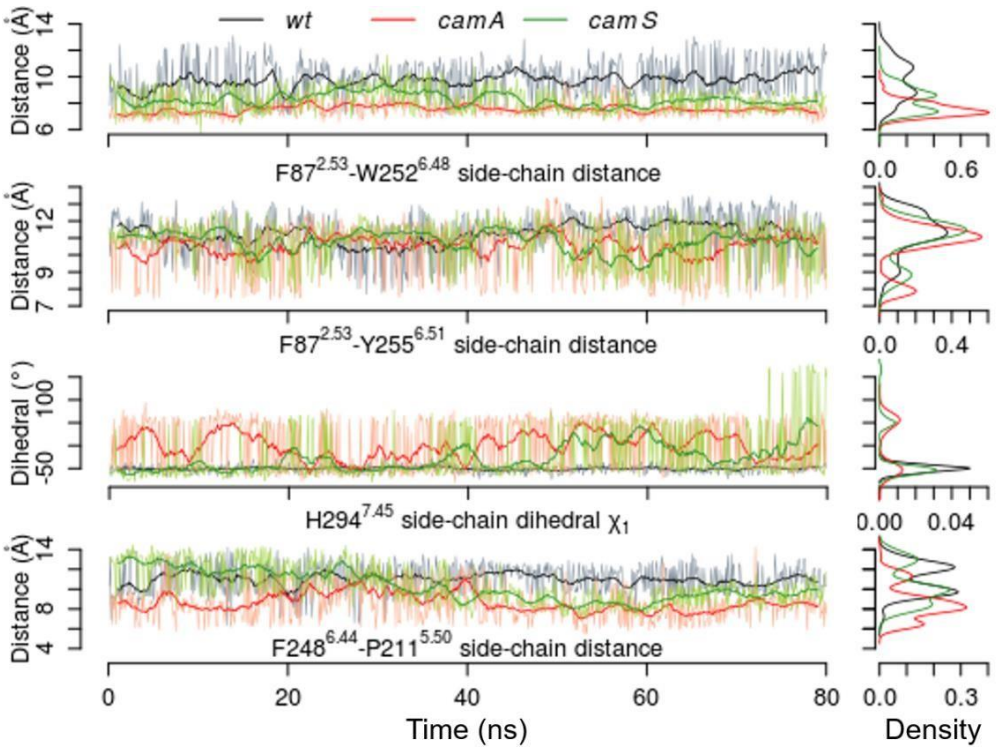

Figure S6. Time-series and distribution plots of the rearrangements of microswitches at the receptor core during the REST2 MD of the three receptors without allosteric Na^+.

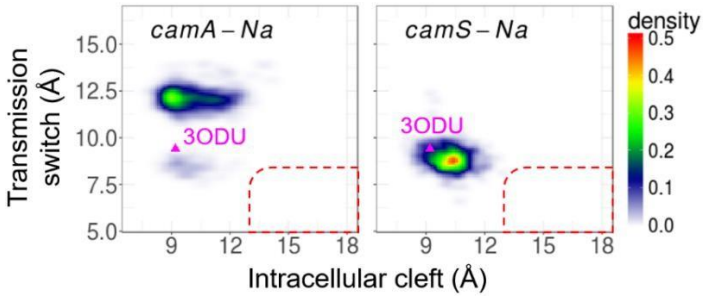

Figure S7. A Na^+ ion artificially placed at the allosteric site of *camA* and *camS* locked the receptors in inactive state during the REST2 MD, as shown here by contour maps of the transmission switch conformation (as measured by $F248^{6.44}$-$P211^{5.50}$ side-chain distance) and the intracellular cleft (as measured by $R134^{3.50}$-$K234^{6.30}$ Cα distance).

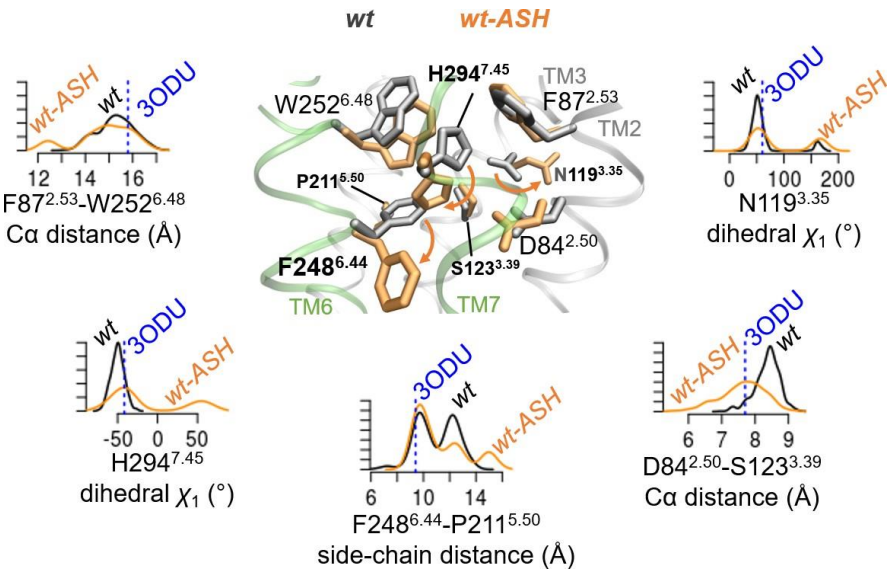

Figure S8. Protonated D84$^{2.50}$ induced rearrangements of the allosteric site in *wt-ASH* (orange) compared to *wt* (black, D84$^{2.50}$ deprotonated). In *wt-ASH*, the F87$^{2.53}$-W252$^{6.48}$ and D84$^{2.50}$-S123$^{3.39}$ distances decreased while H294$^{7.45}$ and F248$^{6.44}$ rotated. In particular, N119$^{3.35}$ rotated toward the outer face of the TM bundle, same as in the active-state structures of the μ-opioid receptor (PDB IDs: 6DDE [1] and 5C1M [2]). The rearrangements were correlated and loosely coupled to the intracellular cleft opening (Table S3).

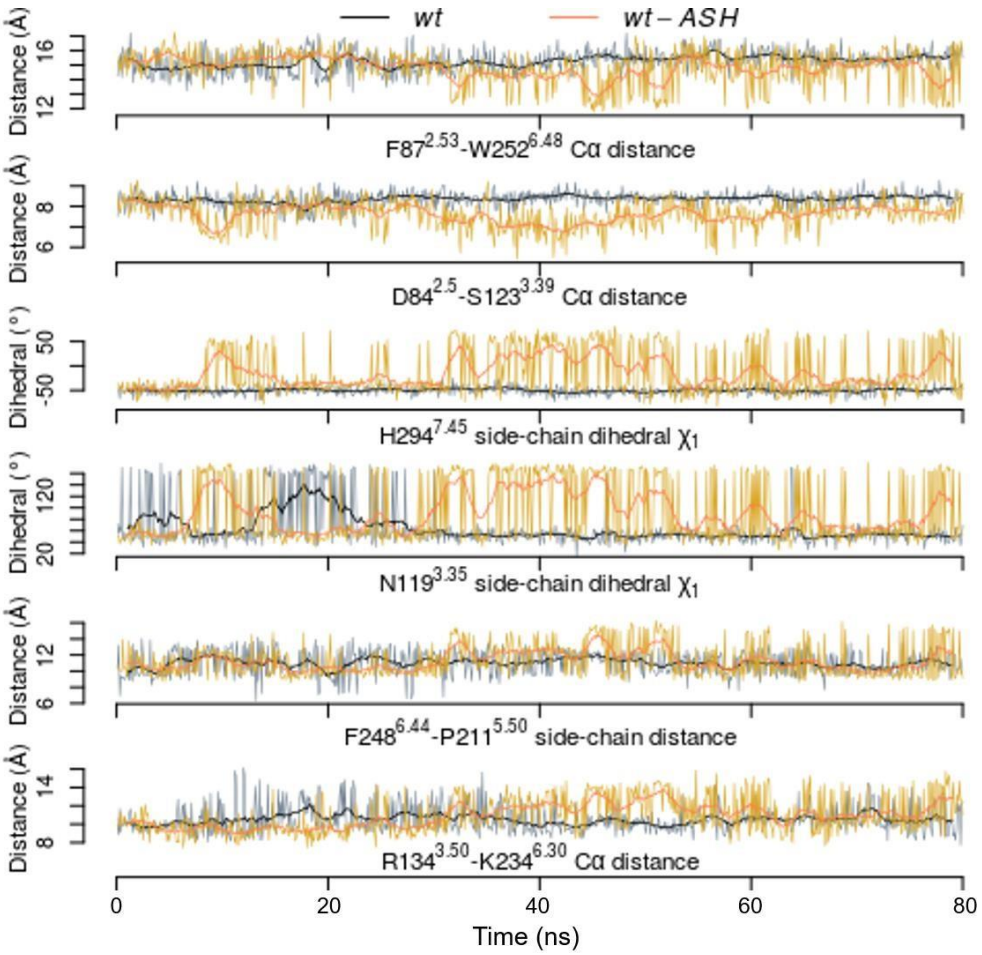

Figure S9. Time series plots of the microswitches and the intracellular cleft sampled during the REST2 MD of the *wt* receptor with D84$^{2.50}$ deprotonated (*wt*, black) or protonated (*wt-ASH*, orange).

Table S1. Allosteric Na^+-binding site collapses in experimental structures of class A GPCRs in active state. Listed here are Cα distances of the most conserved Na^+-coordinating residues, $D^{2.50}$ and $S^{3.39}$.

Receptor	State	β2-adrenergic	μ-opioid	M2 muscarinic	Adenosine 2a	Cannabinoid 1
PDB ID	Active	3SN6 [3]	6DDE[1]/5C1M[2]	4MQS [4]	5G53 [5]	5XR8 [6]
Distance (Å)		6.4	6.5 / 6.1	6.8	5.9	6.7
	Inactive	2RH1 [7]	4DKL [8]	3UON [9]	4EIY [10]	5TGZ [11]
		7.8	8.0	8.3	8.2	7.8

Table S2. Spearman's correlation test[a] between the rotations of $H294^{7.45}$ (allosteric site) and $F248^{6.44}$ (transmission switch), and the intracellular cleft opening during the REST2 MD. The $H294^{7.45}$ and $F248^{6.44}$ rotations appeared to be correlated in the CAMs but not in the *wt*. They were less correlations with the cleft opening, indicating loose coupling.

Pairs of features tested	*camA*		*camS*		*wt*	
	r	*P-value*	*r*	*P-value*	*r*	*P-value*
Rotations of $H294^{7.45}$ ~ $F248^{6.44}$	-0.72	< 2.2e-16	-0.55	< 2.2e-16	-0.11	< 0.002
$H294^{7.45}$ rotation ~ cleft opening	-0.46	< 2.2e-16	0.20	2.5e-08	0.05	0.16
$F248^{6.44}$ rotation ~ cleft opening	0.17	7.0e-07	-0.19	7.5e-08	0.02	0.67

[a] Tested on the $H294^{7.45}$ side-chain dihedral χ_1, the $F248^{6.44}$-$P211^{5.50}$ side-chain distance and the $R134^{3.50}$-$K234^{6.30}$ Cα distance, with a sample size of 800 simulation frames.

Table S3. Spearman's correlation test[a] of the movements of the microswitches[b] and the intracellular cleft opening in *wt-ASH*. Listed here are the *r* values, while all the *P*-values were < 2.2e-16 unless otherwise noted.

r	$F87^{2.53}$-$W252^{6.48}$	$D84^{2.50}$-$S123^{3.39}$	$N119^{3.35}$ rotation	$H294^{7.45}$ rotation	$F248^{6.44}$ rotation
$D84^{2.50}$-$S123^{3.39}$	0.59				
$N119^{3.35}$ rotation	-0.47	-0.50			
$H294^{7.45}$ rotation	-0.61	-0.49	0.62		
$F248^{6.44}$ rotation	-0.60	-0.40	0.60	0.49	
Cleft opening	-0.39	-0.26[c]	0.48	0.45	0.51

[a] Tested with a sample size of 800 simulation frames.

[b] See Fig. S8 caption for the measurements.

[c] *P*-value = 8.58e-14.

Supplementary Methods

Simulation systems. The initial receptor models were built with Modeller v9.15 [12] using the crystal structure of human CXCR4 (PDB ID 3ODU [13]) as template. The intracellular loop 3 (residues 271-296), missing in the crystal structures, was trimmed by 18 residues (residues 273-290) in our model as it is presumably flexible and not mandatory for G protein signaling [14]. The Desmond-Maestro (v2016.1, non-commercial distribution) tools [15] were used to predict the protonation states and to embed the receptor in a bilayer of POPC: the most abundant phospholipid in animal cell membranes [16]. The system was solvated in a periodic 75 × 75 × 105 $Å^3$ box of explicit water and neutralized with 0.15 M of Na^+ and Cl^- ions. Finally, each simulation system consisted of ~50,000 atoms, including ~10,000 water molecules, 31 Na^+ and 39 Cl^- ions. The Amber lipid 14 [17], Amber99SB-ildn [18], TIP3P [19] and the Joung-Cheatham [20] force field parameters were used for the lipids, the protein, the water molecules and the ions, respectively.

Brute-force MD details. After energy minimization, each system was gradually heated to 310 K with 200 $kcal \cdot mol^{-1} \cdot Å^{-1}$ restraints on the protein atoms. Bonds involving hydrogen atoms were constrained using the LINCS algorithm [9], allowing for a 2-fs time step. Van der Waals and short-range electrostatic interactions were cut off at 8 Å. Long- range electrostatic interactions were computed using Particle Mesh Ewald summation (PME). Two phases of 10- ns equilibration with 15 and 5 $kcal \cdot mol^{-1} \cdot Å^{-1}$ restraints on the protein, respectively, were performed in the *NPT*- ensemble (P = 1 bar, T= 310 K) using the Andersen-Parrinello-Rahman barostat [6,7] and the Nose-Hoover thermostat

[8]. Semi-isotropic pressure coupling was applied throughout the MD simulations, which allows the simulation box in the *z*-axis (perpendicular to the lipid bilayer) to vary independently of the *x-y* plane. Non-restrained MD were then carried out and production trajectory was collected after 50 ns of equilibration.

Supporting references

1. A. Koehl, H. Hu, S. Maeda, Y. Zhang, Q. Qu, J. M. Paggi, N. R. Latorraca, D. Hilger, R. Dawson, H. Matile, G. F. X. Schertler, S. Granier, W. I. Weis, R. O. Dror, A. Manglik, G. Skiniotis and B. K. Kobilka, *Nature*, 2018, DOI: 10.1038/s41586-018-0219-7.
2. W. Huang, A. Manglik, A. J. Venkatakrishnan, T. Laeremans, E. N. Feinberg, A. L. Sanborn, H. E. Kato, K. E. Livingston, T. S. Thorsen, R. C. Kling, S. Granier, P. Gmeiner, S. M. Husbands, J. R. Traynor, W. I. Weis, J. Steyaert, R. O. Dror and B. K. Kobilka, *Nature*, 2015, **524**, 315-321.
3. S. G. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I.

Weis, R. K. Sunahara and B. K. Kobilka, *Nature*, 2011, **477**, 549-555.

4. A. C. Kruse, A. M. Ring, A. Manglik, J. Hu, K. Hu, K. Eitel, H. Hubner, E. Pardon, C. Valant, P. M. Sexton, A. Christopoulos, C. C. Felder, P. Gmeiner, J. Steyaert, W. I. Weis, K. C. Garcia, J. Wess and B. K. Kobilka, *Nature*, 2013, **504**, 101-106.
5. B. Carpenter, R. Nehmé, T. Warne, A. G. W. Leslie and C. G. Tate, *Nature*, 2016, **536**, 104-107.
6. T. Hua, K. Vemuri, S. P. Nikas, R. B. Laprairie, Y. Wu, L. Qu, M. Pu, A. Korde, S. Jiang, J. H. Ho, G. W. Han, K. Ding, X. Li, H. Liu, M. A. Hanson, S. Zhao, L. M. Bohn, A. Makriyannis, R. C. Stevens and Z. J. Liu, *Nature*, 2017, **547**, 468-471.
7. V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka and R. C. Stevens, *Science*, 2007, **318**, 1258-1265.
8. A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka and S. Granier, *Nature*, 2012, **485**, 321-326.
9. K. Haga, A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga and T. Kobayashi, *Nature*, 2012, **482**, 547-551.
10. W. Liu, E. Chun, A. A. Thompson, P. Chubukov, F. Xu, V. Katritch, G. W. Han, C. B. Roth, L. H. Heitman, I. J. AP, V. Cherezov and R. C. Stevens, *Science*, 2012, **337**, 232-236.
11. T. Hua, K. Vemuri, M. Pu, L. Qu, G. W. Han, Y. Wu, S. Zhao, W. Shui, S. Li, A. Korde, R. B. Laprairie, E. L. Stahl, J. H. Ho, N. Zvonok, H. Zhou, I. Kufareva, B. Wu, Q. Zhao, M. A. Hanson, L. M. Bohn, A. Makriyannis, R. C. Stevens and Z. J. Liu, *Cell*, 2016, **167**, 750-762 e714.
12. N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper and A. Sali, *Current protocols in bioinformatics*, 2006, **Chapter 5**, Unit-5 6.
13. B. Wu, E. Y. Chien, C. D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F. C. Bi, D. J. Hamel, P. Kuhn, T. M. Handel, V. Cherezov and R. C. Stevens, *Science*, 2010, **330**, 1066-1071.
14. P. Egloff, M. Hillenbrand, C. Klenk, A. Batyuk, P. Heine, S. Balada, K. M. Schlinkmann, D. J. Scott, M. Schutz and A. Pluckthun, *Proceedings of the National Academy of Sciences of the United States of America*, 2014, **111**, E655-662.
15. *Journal*, 2016.
16. N. H. Tattrie, J. R. Bennett and R. Cyr, *Can J Biochem Cell B*, 1968, **46**, 819-&.
17. C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J Chem Theory Comput*, 2014, **10**, 865-879.
18. K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950-1958.

19. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys*, 1983, **79**, 926-935.
20. I. S. Joung and T. E. Cheatham, *The Journal of Physical Chemistry B*, 2008, **112**, 9020-9041.