Supporting Information for: Understanding Ligand Binding Selectivity in a Prototypical GPCR Family

Giulio Mattedi[1], Francesca Deflorian[2], Jonathan S. Mason[2], Chris de Graaf[2], and Francesco L. Gervasio[1,3]

[1]Department of Chemistry, University College London, London WC1E 6BT, UK
[2]Sosei Heptares, Granta Park, Great Abington, Cambridge, CB21 6DG, UK
[3]Institute of Structural and Molecular Biology, University College London, London WC1E 6BT, UK

1 Alignment of Hat and Human A_1R and $A_{2a}R$

Alignment of the rat and human proteins, as calculated with Clustal Omega.[1] Pocket-lining residues for the four sequences are marked with asterisks. It is to be noted that murine and human A_1R and $A_{2a}R$ have high sequence similarity (82.9% and 94.8% for $A_{2a}R$ and A_1R, respectively), particularly in the orthosteric site region, which allows us to extrapolate useful information for the studying ligand selectivity on the human receptors. In the case of $A_{2a}R$, for which binding affinities of ZM241385 are available both for human and rat receptor, the selectivity pattern is closely preserved, as can be seen in Table 1 of the main text.

A_1R

```
sp|P25099|AA1R_RAT      MPPYISAFQAAYIGIEVLIALVSVPGNVLVIWAVKVNQALRDATFCFIVSLAVADVAVGA 60
sp|P30542|AA1R_HUMAN    MPPSISAFQAAYIGIEVLIALVSVPGNVLVIWAVKVNQALRDATFCFIVSLAVADVAVGA 60
sp|P25099|AA1R_RAT      LVIPLAILINIGPQTYFHTCLMVACPVLILTQSSILALLAIAVDRYLRVKIPLRYKTVVT 120
sp|P30542|AA1R_HUMAN    LVIPLAILINIGPQTYFHTCLMVACPVLILTQSSILALLAIAVDRYLRVKIPLRYKMVVT 120
                                                          **          *
sp|P25099|AA1R_RAT      QRRAAVAIAGCWILSLVVGLTPMFGWNNLSVVEQDWRANGSVGEPVIKCEFEKVISMEYM 180
sp|P30542|AA1R_HUMAN    PRRAAVAIAGCWILSFVVGLTPMFGWNNLSAVERAWAANGSMGEPVIKCEFEKVISMEYM 180
sp|P25099|AA1R_RAT      VYFNFFVWVLPPLLLMVLIYLEVFYLIRKQLNKKVSASSGDPQKYYGKELKIAKSLALIL 240
sp|P30542|AA1R_HUMAN    VYFNFFVWVLPPLLLMVLIYLEVFYLIRKQLNKKVSASSGDPQKYYGKELKIAKSLALIL 240
                          *  ** ** *      *    *  *
sp|P25099|AA1R_RAT      FLFALSWLPLHILNCITLFCPTCQKPSILIYIAIFLTHGNSAMNPIVYAFRIHKFRVTFL 300
sp|P30542|AA1R_HUMAN    FLFALSWLPLHILNCITLFCPSCHKPSILTYIAIFLTHGNSAMNPIVYAFRIQKFRVTFL 300
sp|P25099|AA1R_RAT      KIWNDHFRCQPKPPIDEDLPEEKAED 326
sp|P30542|AA1R_HUMAN    KIWNDHFRCQPAPPIDEDLPEERPDD 326
```

$A_{2a}R$

```
sp|P29274|AA2AR_HUMAN     MPIMGSSVYITVELAIAVLAILGNVLVCWAVWLNSNLQNVTNYFVVSLAAADIAVGVLAI 60
sp|P30543|AA2AR_RAT       ---MGSSVYITVELAIAVLAILGNVLVCWAVWINSNLQNVTNFFVVSLAAADIAVGVLAI 57
sp|P29274|AA2AR_HUMAN     PFAITISTGFCAACHGCLFIACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGTR 120
sp|P30543|AA2AR_RAT       PFAITISTGFCAACHGCLFFACFVLVLTQSSIFSLLAIAIDRYIAIRIPLRYNGLVTGVR 117
```

```
                                                                  **    *
sp|P29274|AA2AR_HUMAN      AKGIIAICWVLSFAIGLTPMLGWNNCGQPKEGKNHSQGCGEGQVACLFEDVVPMNYMVYF 180
sp|P30543|AA2AR_RAT        AKGIIAICWVLSFAIGLTPMLGWNNCSQKD--GNSTKTCGEGRVTCLFEDVVPMNYMVYY 175

sp|P29274|AA2AR_HUMAN      NFFACVLVPLLLMLGVYLRIFLAARRQLKQMESQPLPGERARSTLQKEVHAAKSLAIIVG 240
sp|P30543|AA2AR_RAT        NFFAFVLLPLLLMLAIYLRIFLAARRQLKQMESQPLPGERTRSTLQKEVHAAKSLAIIVG 235
                              *  ** **  *       *     *   *
sp|P29274|AA2AR_HUMAN      LFALCWLPLHIINCFTFFCPDCSHAPLWLMYLAIVLSHTNSVVNPFIYAYRIREFRQTFR 300
sp|P30543|AA2AR_RAT        LFALCWLPLHIINCFTFFCSTCRHAPPWLMYLAIILSHSNSVVNPFIYAYRIREFRQTFR 295

sp|P29274|AA2AR_HUMAN      KIIRSHVLRQQEPFKAAGTSARVLAAHGSDGEQVSLRLNGHPPGVWANGSAPHPERRPNG 360
sp|P30543|AA2AR_RAT        KIIRTHVLRRQEPFQAGGSSAWALAAHSTEGEQVSLRLNGHPLGVWANGSATHSGRRPNG 355

sp|P29274|AA2AR_HUMAN      YALGLVSGGSAQESQGNTGLPDVELLSHELKGVCPEPPGLDDPLAQDGAGVS-------- 412
sp|P30543|AA2AR_RAT        YTLGLGGGGSAQGSPRDVELPTQER------QEGQEHPGLRGHLVQARVGASSWSSEFAP 409

sp|P29274|AA2AR_HUMAN      - 412
sp|P30543|AA2AR_RAT        S 410
```

2 Supplementary Methods

2.1 System Setup

The crystal structures of the human A_1 and A_{2a} receptors were downloaded from the Protein Data Bank (A_1R: 5N2S,[2] A_{2a}R: 5IU4[3]). After correcting mutations and removing the apocytochrome b562 (bRIL) inserted in the ICL3 loop or at the N-terminus, missing loops were modelled with MODELLER 9.19[4] and the proteins were embedded in a pre-equilibrated POPC membrane. The membrane was then aligned to the XY plane, resulting in the main axis of the GPCRs and the Z axis being close to parallel.

With the exception of the A_{2a}R-ZM241385 system, the ligands were placed in the binding pockets by superposition of the proteins onto the crystal structure of A_{2a}R. Histidines exposed to the solvent were protonated. The complexes were solvated with TIP3P water and chloride ions were added to balance the net charge. Ligands were parametrised with the generalised AMBER force field (GAFF)[5] and charges were calculated using Gaussian09[6] with a 6-31G* basis set at the Hartree-Fock level. Protein, water and ion parameters were generated with AMBER14SB force field,[7,8] and the phospholipid topology was downloaded from LipidBook.[9] All simulations were run using GROMACS 5.1.4[10] with the PLUMED 2.3.1 plugin[11] in the NPT ensemble. Temperature and pressure were enforced with the V-rescale thermostat,[12] and the Parrinello-Rahman barostat.[13] Electrostatics were treated with the PME-Switch algorithm and the cutoff for van der Waals and electrostatic interactions was set to 1.0 nm.

2.2 Equilibration

Clashes were fixed by energy minimization using a steepest descent integrator until the maximum force exerted on the atoms dropped below 1000.0 kJ/mol/nm. Over the span of 60 ns, the system were then progressively heated from 100 K to 300 K in the NPT ensemble, applying positional restraints to the α carbons of the proteins. A 25 ns-long unrestrained molecular dynamics simulation then followed.

2.3 Unbiased Molecular Dynamics

For the unbiased molecular dynamics runs, a single 1 μs-long simulation was run for each system, starting from the equilibrated conformations, at 300 K.

2.4 Metadynamics

The metadynamics[14] simulations were run using a parallel tempering scheme.[15] Six replicas for each system were equilibrated for 250 ps to temperatures from 300 K to 310 K, and in the production run exchanges were attempted every 1000 steps. Similarly to the setup used in our previous experience of the reconstruction of binding free energy landscapes in GPCRs,[16] the pocket-ligand distance vector was calculated using the α carbon of the conserved $W^{6.48}$ at the bottom of the binding site and the center of mass of the carbon and nitrogen atoms at the interface between the triazole and triazine rings of the ligands. *Z-projection* was defined as the projection of the vector onto the Z-axis, whereas *XY-projection* as the projection onto the XY-plane. In all four systems, the POPC bilayer is parallel to the XY-plane, positioning the opening of the orthosteric binding site along the Z-axis.

During the production metadynamics runs Gaussian hills were deposited every 2 ps in the well-tempered scheme[17] with a bias factor of 15. The Gaussian width was set to 0.1 nm for the *Z-projection* and *XY-projection* and to 0.033 nm for the salt bridge distance. In the simulations where two CVs were biased the initial height was 1 kJ/mol, whereas in that biasing three CVs it was set to 1.5 kJ/mol.

When reconstructing the binding free energy surface of the ligands with A_1R biasing *Z-projection* and *XY-projection* is was clear that the $E172^{ECL2}$-$K265^{ECL3}$ salt bridge opening is a slowly equilibrating degree of freedom that greatly hinders the diffusion of the ligand, and biasing it was necessary to reach convergence. For ZM241385 the *Z-projection* - salt bridge distance combination was adequate, while for LUF5452 adding *XY-projection* was needed. The $E172^{ECL2}$-$K265^{ECL3}$ salt bridge distance was calculated using the position of the carbonyl carbon of the glutamate side chain and ϵN of the histidine imidazole ring.

The simulations were considered converged when the binding free energy estimate over time adopted an asymptotic behaviour. The final binding affinity was computed by calculating average and standard deviation of the binding free energy over the last 100 ns of metadynamics simulation. A funnel-like restraint was applied in order to limit the exploration of degenerate fully solvated states of the ligand.[16,18] The functional form of the restrain is a sigmoid in the space of the *XY-projection* CV:

$$r = h \cdot \frac{1}{1 + e^{s(z - z_0)}} + b \tag{1}$$

Where r is the *XY-projection*, $h = 1.6$ nm the funnel width (modulo b),

$s = 2$ nm^{-1} controls the steepness of the function, z is the *Z-projection* CV, $z_0 = 4$ nm the inflection point and $b = 0.15$ nm the minimum width. At the funnel boundary, a quadratic repulsive potential was applied. The binding free energy was then corrected for the presence of the restraint using Eq. 3 in the paper by Limongelli et al.[18]

3 Supplementary Figures

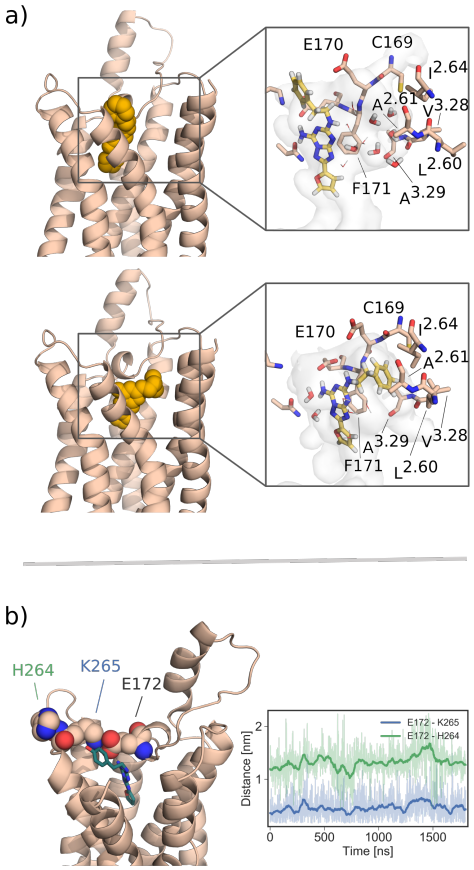

Figure S1: **a)** LUF5452 bound to the main binding site (top) and interacting with the hydrophobic ECV site (bottom) during the parallel tempering metadynamics simulation. The residues that are in close proximity to the phenyl group of the ligand are labelled. **b)** Comparison of the distance between the cabonyl carbon of E172 and ϵN of H264 or K265 of A_1R, recorded in a parallel tempering metadynamics simulation of the A_1R-ZM241385 system where only *Z-projection* and *XY-projection* were biased. Throughout the dynamics, K265 is likely to interact with the glutamate side chain and influence the binding and unbinding processes.

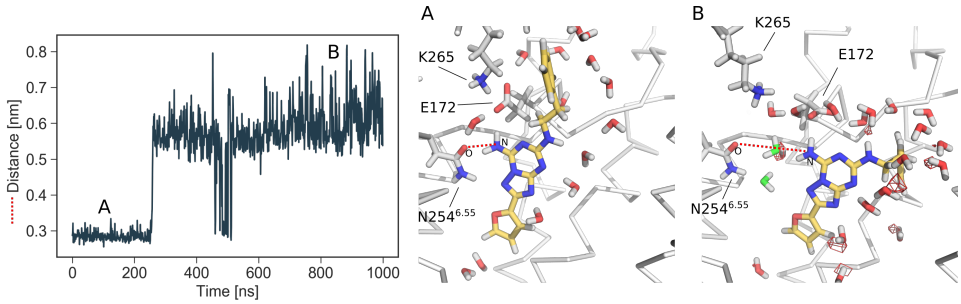

Figure S2: Distance between the primary exocyclic amine nitrogen of LUF5452 and the carbonyl oxygen of N245$^{6.55}$ of A_1R throughout the unbiased molecular dynamics simulation. Upon the rearrangement of the ligand the distance between the two atoms increases from 3 Å to 6 Å, allowing for one or more water molecules to mediate the interaction between the compound and the residue side chain. The water hotspots calculated with VolMap[19] (red mesh) indicate a solvent hotspot between the two partners, suggesting that the interaction is favourable.

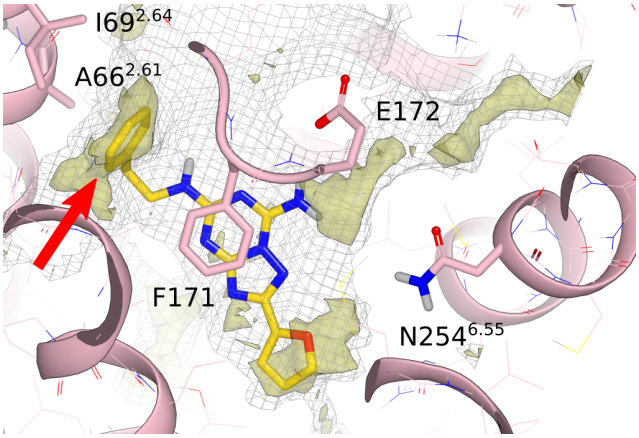

Figure S3: GRID[20] analysis of the protein cavity of A_1R, from a representative snapshot of the unbiased molecular dynamics with LUF5452, with the ligand interacting with the hydrophobic ECV pocket (red arrow). The CH3 methyl probe (gray mesh, contoured at 1 kcal/mol) was used to define how close a ligand carbon atom can get to the surface of the protein. The lipophilic C1= probe (yellow solid surface, contoured at -2.8 kcal/mol) highlights the lipophilic hotspots in the pocket. Note the overlap with the phenyl ring of LUF5452. VIDA (OpenEye) was used for producing the figure.

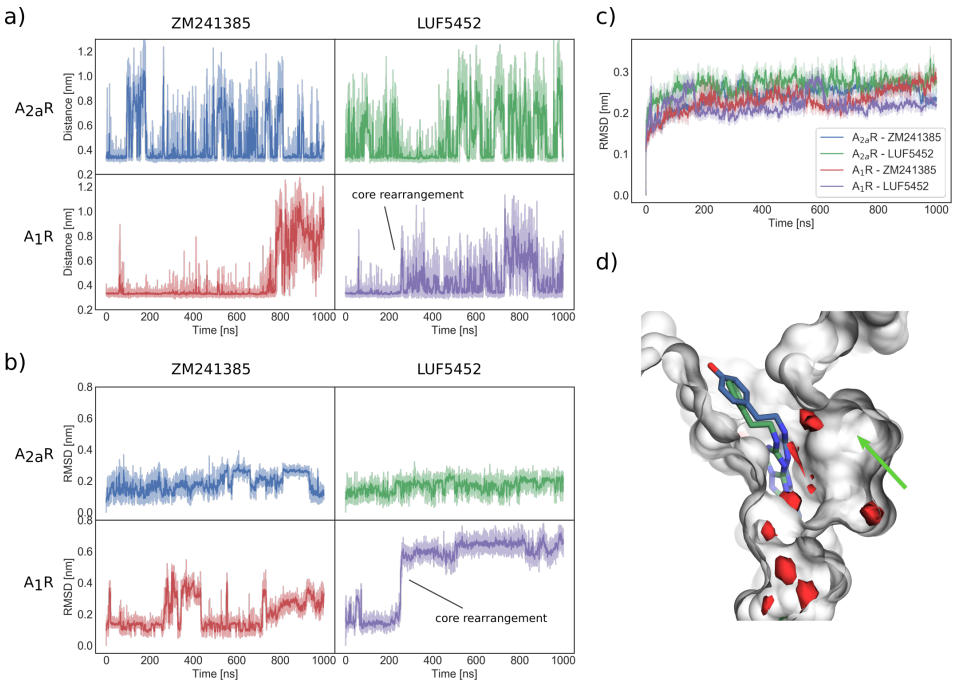

Figure S4: Analysis of the unbiased MD simulations: **a)** Salt bridge distance in the four systems over the simulations. **b)** Ligand RMSD. The difference in stability between A_1R and $A_{2a}R$ is clear from the plots. In the A_1R-LUF5452 simulation, at t=250 ns, the ligand spontaneously rearranged in order to interact with the hydrophobic ECV pocket. **c)** Protein Cα RMSD over the simulation. **d)** Superposition of A_1R-LUF5452 (green) and A_1R-ZM241385 (blue). The water hotspots (at 1 σ) are shown as red surfaces. The lack of sites of strong interaction for the solvent in the hydrophobic ECV pocket (green arrow) is suggested by the lack of a hotspot.

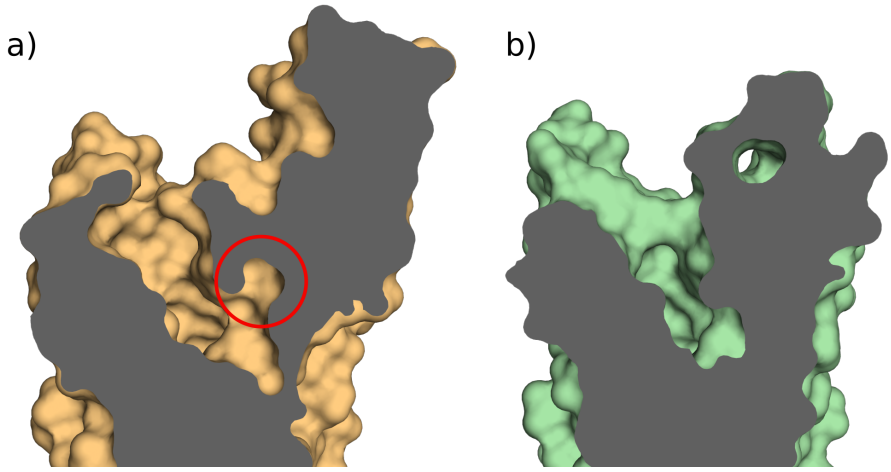

Figure S5: Cross sectional view of A_1R (**a**, PDB 5N2S[2]) and $A_{2a}R$ (**b**, PDB 5IU4[3]). Ligands not shown for clarity. The red circle indicates the accessory hydrophobic ECV pocket, absent in $A_{2a}R$.

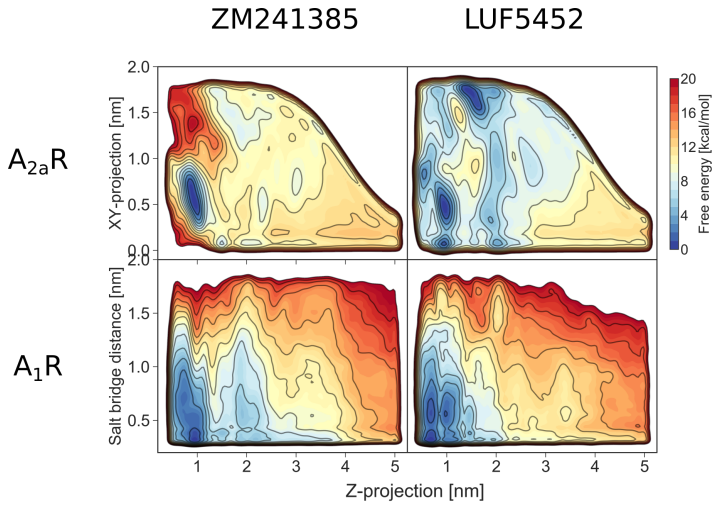

Figure S6: Binding free energy landscapes of the systems onto the biased variables. A_1R-LUF5452 is the projection of the free energy onto *Z-projection* and the $E172^{ECL2}$-$K265^{ECL3}$ salt bridge distance.

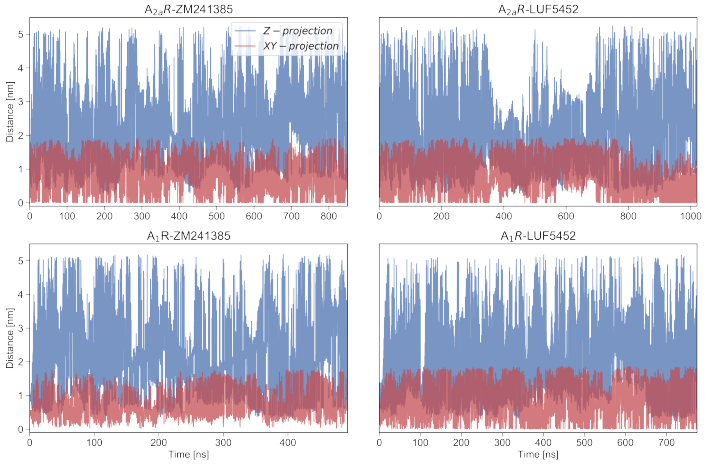

Figure S7: Diffusion of the parallel tempering metadynamics simulations in the *Z-projection* - *XY-projection* CV space.

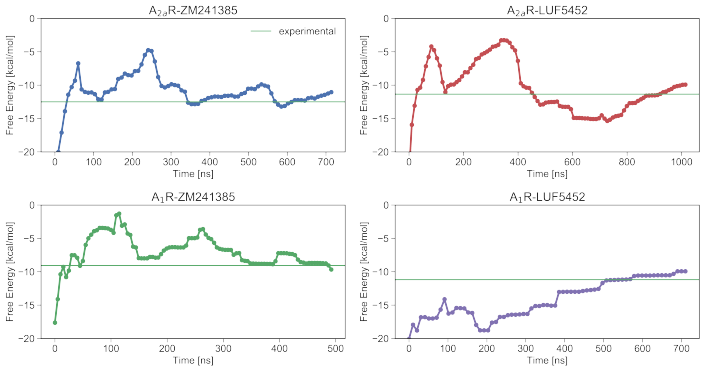

Figure S8: Evolution of the computed binding free energy over the metadynamics simulation time in comparison with experimental data.

References

(1) Larkin, M.; Blackshields, G.; Brown, N.; Chenna, R.; McGettigan, P.; McWilliam, H.; Valentin, F.; Wallace, I.; Wilm, A.; Lopez, R.; Thompson, J.; Gibson, T.; Higgins, D. Clustal W and Clustal X Version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.

(2) Cheng, R. K. Y.; Segala, E.; Robertson, N.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Fiez-Vandal, C.; Marshall, F. H.; Cooke, R. M. Structures of Human A_1 and A_{2A} Adenosine Receptors with Xanthines Reveal Determinants of Selectivity. *Structure* **2017**, *25*, 1275–1285.e4.

(3) Segala, E.; Guo, D.; Cheng, R. K. Y.; Bortolato, A.; Deflorian, F.; Doré, A. S.; Errey, J. C.; Heitman, L. H.; Ijzerman, A. P.; Marshall, F. H.; Cooke, R. M. Controlling the Dissociation of Ligands from the Adenosine A_{2A} Receptor Through Modulation of Salt Bridge Strength. *J. Med. Chem.* **2016**, *59*, 6470–6479.

(4) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M.; Eramian, D.; Shen, M.-y.; Pieper, U.; Sali, A. Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinf.* **2006**, *15*, 5.6.1–5.6.30.

(5) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a.; Case, D. a. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(6) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas,; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision B.01. 2009.

(7) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.;

Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker,; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. Amber 14. University of California, San Francisco, 2014.

(8) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713, PMID: 26574453.

(9) Domański, J.; Stansfeld, P. J.; Sansom, M. S. P.; Beckstein, O. Lipidbook: A Public Repository for Force-Field Parameters Used in Membrane Simulations. *J. Membr. Biol.* **2010**, *236*, 255–258.

(10) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(11) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(12) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J Chem Phys* **2007**, *126*, 14101.

(13) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196–1199.

(14) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(15) Sugita, Y.; Okamoto, Y. Replica-exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(16) Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T. An Efficient Metadynamics-based Protocol to Model the Binding Affinity and the Transition State Ensemble of G-protein-coupled Receptor Ligands. *J. Chem. Inf. Model.* **2017**, *57*, 1210–1217.

(17) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered Metadynamics: A Smoothly Converging and Tunable Free-energy Method. *Phys. Rev. Lett.* **2008**, *100*, 1–4.

(18) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6358–6363.

(19) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

(20) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.