nature methods

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41592-020-0884-y

In the format provided by the authors and unedited.

GPCRmd uncovers the dynamics of the 3D-GPCRome

Ismael Rodríguez-Espigares[1,24], Mariona Torrens-Fontanals[1,24], Johanna K. S. Tiemann[2,3], David Aranda-García[1], Juan Manuel Ramírez-Anguita[1], Tomasz Maciej Stepniewski[1], Nathalie Worp[1], Alejandro Varela-Rial[4,5], Adrián Morales-Pastor[1], Brian Medel-Lacruz[1], Gáspár Pándy-Szekeres[6], Eduardo Mayol[7], Toni Giorgino[8,9], Jens Carlsson[10], Xavier Deupi[11,12], Slawomir Filipek[13], Marta Filizola[14], José Carlos Gómez-Tamayo[7], Angel Gonzalez[7], Hugo Gutiérrez-de-Terán[15], Mireia Jiménez-Rosés[7], Willem Jespers[15], Jon Kapla[10], George Khelashvili[16,17], Peter Kolb[18], Dorota Latek[13], Maria Marti-Solano[18,19], Pierre Matricon[10], Minos-Timotheos Matsoukas[7,20], Przemyslaw Miszta[13], Mireia Olivella[7], Laura Perez-Benito[7], Davide Provasi[14], Santiago Ríos[7], Iván R. Torrecillas[7], Jessica Sallander[15], Agnieszka Sztyler[13], Silvana Vasile[15], Harel Weinstein[16,17], Ulrich Zachariae[21,22], Peter W. Hildebrand[2,3,23], Gianni De Fabritiis[4,5], Ferran Sanz[1], David E. Gloriam[6], Arnau Cordomi[7], Ramon Guixà-González[11,12] ✉ and Jana Selent[1] ✉

[1]Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute—Department of Experimental and Health Sciences, Pompeu Fabra University, Barcelona, Spain. [2]Institute of Medical Physics and Biophysics, Charite University Medicine Berlin, Berlin, Germany. [3]Institute of Medical Physics and Biophysics, Medical University Leipzig, Leipzig, Sachsen, Germany. [4]Computational Science Laboratory, Universitat Pompeu Fabra, Barcelona Biomedical Research Park, Barcelona, Spain. [5]Acellera, Barcelona, Spain. [6]Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark. [7]Laboratori de Medicina Computacional, Unitat de Bioestadistica, Facultat de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain. [8]Biophysics Institute, National Research Council of Italy, Milan, Italy. [9]Department of Biosciences, University of Milan, Milan, Italy. [10]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden. [11]Laboratory of Biomolecular Research, Paul Scherrer Institute (PSI), Villigen PSI, Switzerland. [12]Condensed Matter Theory Group, PSI, Villigen PSI, Switzerland. [13]Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland. [14]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [15]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden. [16]Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, NY, USA. [17]Institute for Computational Biomedicine, Weill Cornell Medical College of Cornell University, New York, NY, USA. [18]Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marburg, Germany. [19]MRC Laboratory of Molecular Biology, Cambridge, UK. [20]Department of Pharmacy, University of Patras, Patras, Greece. [21]Computational Biology, School of Life Sciences, University of Dundee, Dundee, UK. [22]Physics, School of Science and Engineering, University of Dundee, Dundee, UK. [23]Berlin Institute of Health, Berlin, Germany. [24]These authors contributed equally: Ismael Rodríguez-Espigares, Mariona Torrens-Fontanals. ✉e-mail: ramon.guixa@psi.ch; jana.selent@upf.edu

GPCRmd uncovers the dynamics of the 3D-GPCRome

Ismael Rodríguez-Espigares[1*], Mariona Torrens-Fontanals[1*], Johanna K.S. Tiemann[2,3], David Aranda-García[1], Juan Manuel Ramírez-Anguita[1], Tomasz Maciej Stepniewski[1], Nathalie Worp[1], Alejandro Varela-Rial[4,5], Adrián Morales-Pastor[1], Brian Medel Lacruz[1], Gáspár Pándy-Szekeres[6], Eduardo Mayol[7], Toni Giorgino[8,9], Jens Carlsson[10], Xavier Deupi[11,12], Slawomir Filipek[13], Marta Filizola[14], José Carlos Gómez-Tamayo[7], Angel Gonzalez[7], Hugo Gutierrez-de-Teran[15], Mireia Jimenez[7], Willem Jespers[15], Jon Kapla[10], George Khelashvili[16,17], Peter Kolb[18], Dorota Latek[13], Maria Marti-Solano[18,21], Pierre Matricon[10], Minos-Timotheos Matsoukas[7,22], Przemyslaw Miszta[13], Mireia Olivella[7], Laura Perez-Benito[7], Davide Provasi[14], Santiago Ríos[7], Iván Rodríguez-Torrecillas[7], Jessica Sallander[15], Agnieszka Sztyler[13], Silvana Vasile[15], Harel Weinstein[16,17], Ulrich Zachariae[23,24], Peter W. Hildebrand[2,3,25], Gianni De Fabritiis[4,5], Ferran Sanz[1], David E. Gloriam[6], Arnau Cordomi[7], Ramon Guixà-González[11,12,✉], Jana Selent[1,✉]

---

[1]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM) – Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), Barcelona, Spain, [2]Institute of Medical Physics and Biophysics, Charite University Medicine Berlin, Berlin 10117, Germany, [3]Institute of Medical Physics and Biophysics, Medical University Leipzig, Leipzig, Sachsen 04107, Germany, [4]Computational Science Laboratory, Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), Carrer del Dr. Aiguader 88, 08003 Barcelona, Spain, [5]Acellera, C / Carrer Dr. Trueta, 183, 08005 Barcelona, Spain, [6]Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100, Copenhagen, Denmark, [7]Laboratori de Medicina Computacional, Unitat de Bioestadistica, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain, [8]Biophysics Institute (IBF-CNR), National Research Council of Italy, Milan, Italy, [9]Department of Biosciences, University of Milan, Milan, Italy, [10]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden, [11]Laboratory of Biomolecular Research, Paul Scherrer Institute (PSI), 5232 Villigen PSI, Switzerland, [12]Condensed Matter Theory Group, Paul Scherrer Institute (PSI), 5232 Villigen PSI, Switzerland, [13]Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland, [14]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [15]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24, Uppsala, Sweden, [16]Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, NY 10065, USA, [17]Institute for Computational Biomedicine, Weill Cornell Medical College of Cornell University, New York, NY 10065, USA, [18]Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6, 35032 Marburg, Germany, [21]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK, [22]Department of Pharmacy, University of Patras, 26504, Patras, Greece, [23]Computational Biology, School of Life Sciences, University of Dundee, Dundee, UK, [24]Physics, School of Science and Engineering, University of Dundee, Dundee, UK, [25]Berlin Institute of Health, 10178 Berlin, Germany. * These authors contributed equally. ✉ Correspondence to ramon.guixa@psi.ch or jana.selent@upf.edu.

Supplementary Note 1. Sustainability

Projects such as the GPCRmd face huge challenges to overcome in order to be accepted, used and sustained in the future. As previously shown, structural biology or (proteo-)genomics projects like the PDB (rcsb.org[1]) or the Galaxy project[2], respectively, have demonstrated the need for strong community support rather than individual laboratories to preserve adequate sustainability. Therefore, we created the GPCRmd consortium under the umbrella of the newly granted ERNEST network (GLISTEN COST action, https://ernest-gpcr.eu), with the support of the GPCR community. Additionally, the focus on a specific research area such as GPCRs rather than a general database for all MD simulations reduces hurdles like deposition space problems, too general and likely unused analysis tools and problems in the findability due to too broad keywords/labels. An important asset promoting sustainability is the automated update of GPCRmd with new MD simulations for newly published GPCR structures. These updates will follow a standardized protocol developed by the GPCRmd community. On the other hand, GPCRmd is designed to serve as a revision platform in the near future allowing the editor and reviewers to evaluate dynamics data on the fly, making them more transparent and trustful and allowing upon acceptance for an easy deposition into the database.

Supplementary Note 2. System set-up and simulation protocol

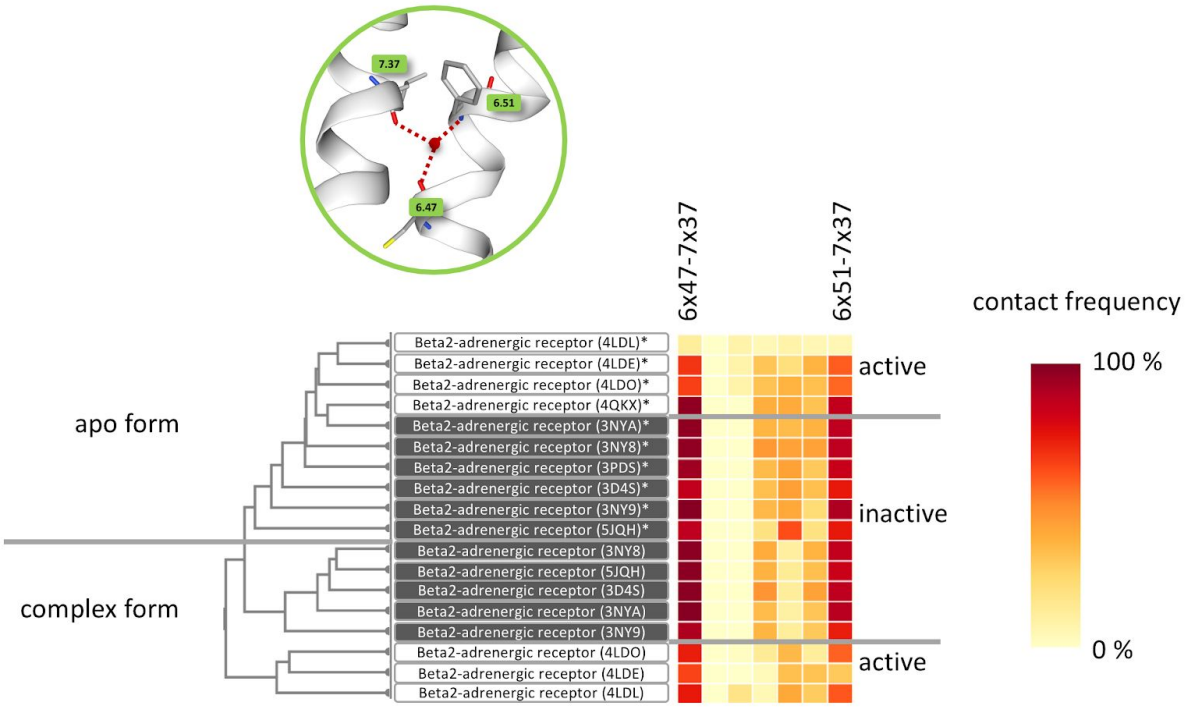
Protein structure preparation. The structures of simulated GPCRs were obtained from the Protein Data Bank (PDB) (rcsb.org[1]) (see Table S1 for PDB IDs). Auxiliary proteins and long unresolved or truncated N- and C-terminal regions were removed and stabilizing mutations were reverted back to the native sequence. In receptors where ICL3 was longer than 10 residues, a chain break was introduced in the middle and five residues at each end were modeled using MODELLER[3]. Modelling and refinement of the structures was performed using the methodology described by Pándy-Szekeres et al[4]. Steric clashes were energy minimized using the MOE software[5]. Crystallographic waters and lipid modifications present in the X-ray structure were preserved. The rest of the co-crystallized molecules were discarded. Receptor residue protonation and tautomeric states were assigned using PROPKA[6,7] as implemented in PDB2PQR[8] at pH 7 and subsequently curated by members of the GPCRmd consortium. Residue D2x50 was either kept deprotonated or protonated for antagonist- or agonist-bound receptor complexes, respectively.

Ligand parameterization. Tripos Mol2 File Format files were taken from the PDB using the MOE software[5]. Protonation and tautomeric states at pH 7 were predicted and assigned using the Marvin Calculator Plugin[9]. The obtained Mol2 files were used to generate parameters by analogy using the ParamChem server 1.0.0 (https://cgenff.umaryland.edu) and CGenFF 3.0.1[10–13]. Parameters for inorganic phosphate (PO_4^{2-}) were obtained from CGenFF and SwissParam[14] web server. Finally, adenosine parameters where taken from the RNA CHARMM 36 force-field[15]

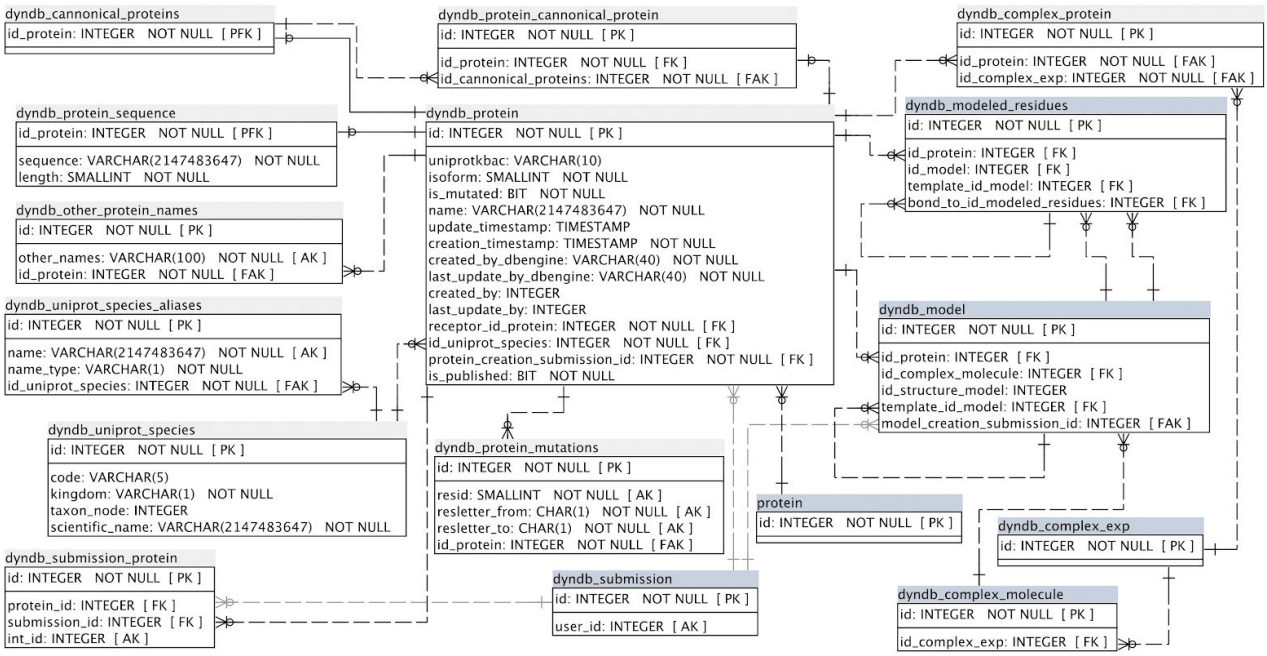
Placement of additional internal waters: For each structure we used HomolWat[16] (http://lmc.uab.cat/homolwat/) to incorporate internal water molecules not determined in other structures of the same or parent receptors. To this end, we used water molecules with a circular variance[19] > 0.6 measured using vectors from the oxygen atom of a water molecule to the surrounding atoms up to 10 Å. The algorithm uses *blastp* (ncbi-blast v2.6.0+)[18] to generate an ordered list of receptors with determined structures that contain resolved water molecules and subsequently tries to incorporate into the model all water molecules that do not clash (> 2.4 Å) with receptor atoms or previously introduced water molecules, starting with 1) receptors with the largest sequence identity, 2) structures with the best resolution and 3) water molecules with the smallest B-factors. Water hydrogens are added using the PDB2PQR software[8]. Non-coincident water molecules (distance > 2 Å) predicted by Dowser+ software[20] were also incorporated into the structure.

System set-up. Structures generated by the GPCRmd consortium were checked for protonation consistency. Acetylated and charged N terminus were used for incomplete and complete N terminus capping, respectively. Amidated and charged C terminus were used for incomplete and complete C terminus capping, respectively. Each GPCR model was aligned to its respective orientation taken from the Orientations of Proteins in Membranes database[21] using STAMP 4.4[22,23] . Receptors were then embedded into a POPC bilayer and solvated ensuring a 20 Å distance between protein periodic distances, considering also receptor diffusional rotation. Furthermore, the system was checked for lipids inserted into aromatic rings. Finally, systems were solvated with TIP3 water molecules and the ionic strength of the solution was adjusted to 0.15 M using NaCl ions. Parameters for the simulation were obtained from the CHARMM36m force field[24,25].

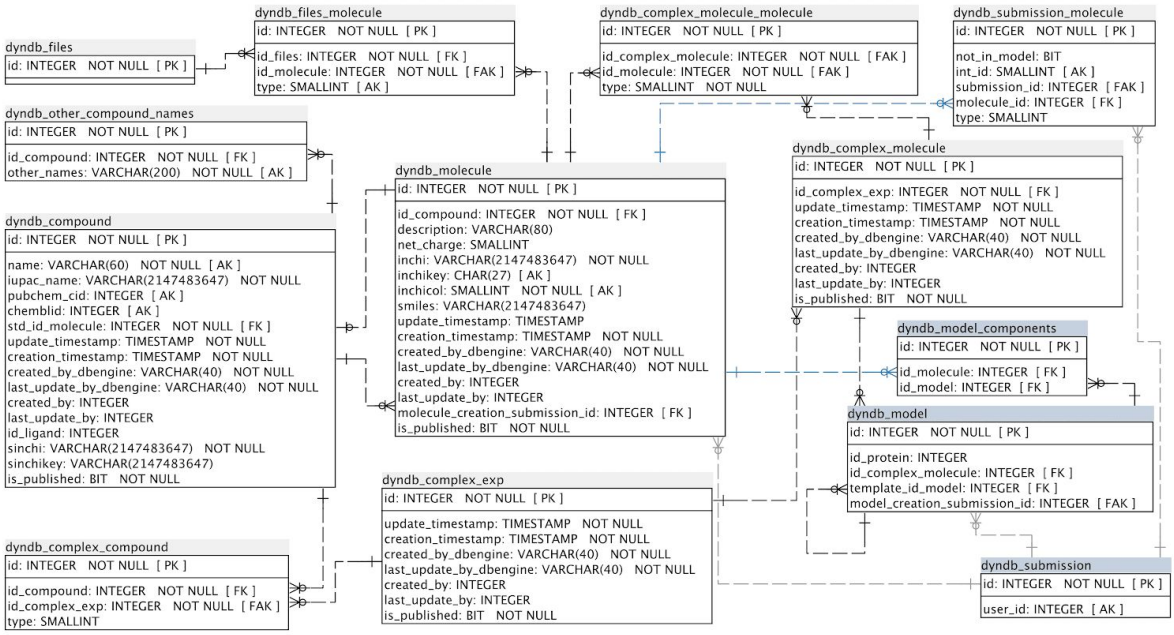
Molecular dynamics (MD) simulations. Systems were first energy minimized during 5000 step and then equilibrated at constant pressure (NPT, 1.01325 bar) using the Berendsen barostat[26] with a pressure relaxation time of 800 fs and a compressibility factor of 4.57×10^{-5} bar^{-1} during 30 ns. In a first step, harmonic restraints of 1.0 kcal/(mol·Å^2) were set on protein backbone and water oxygen atoms during 10 ns. Then, restrains were progressively released in a ramp of -0.095 kcal/(mol·Å^2·ns) during 10 ns followed by a restraints-free equilibration step of 10 ns. Production simulations were performed at constant volume (NVT) in 3 replicates of 500 ns per system using ACEMD[27] and GPUGRID[28]. Time-step of 2 and 4 fs were used during the equilibration and production runs, respectively. Non-bonded interactions were cut-off at 9 Å. A smooth switching function for the cut-off was applied, starting at 7.5 Å. Long-distance electrostatic forces were calculated using the Particle Mesh Ewald algorithm[29] with a grid spacing of 1 Å. Bond lengths of hydrogen atoms were kept constrained using the RATTLE algorithm[30]. All simulations were carried out at a temperature of 310K using the Langevin thermostat[31] with damping constants γ of 1 ps^{-1} and 0.1 ps^{-1} for NPT and NVT simulations, respectively.



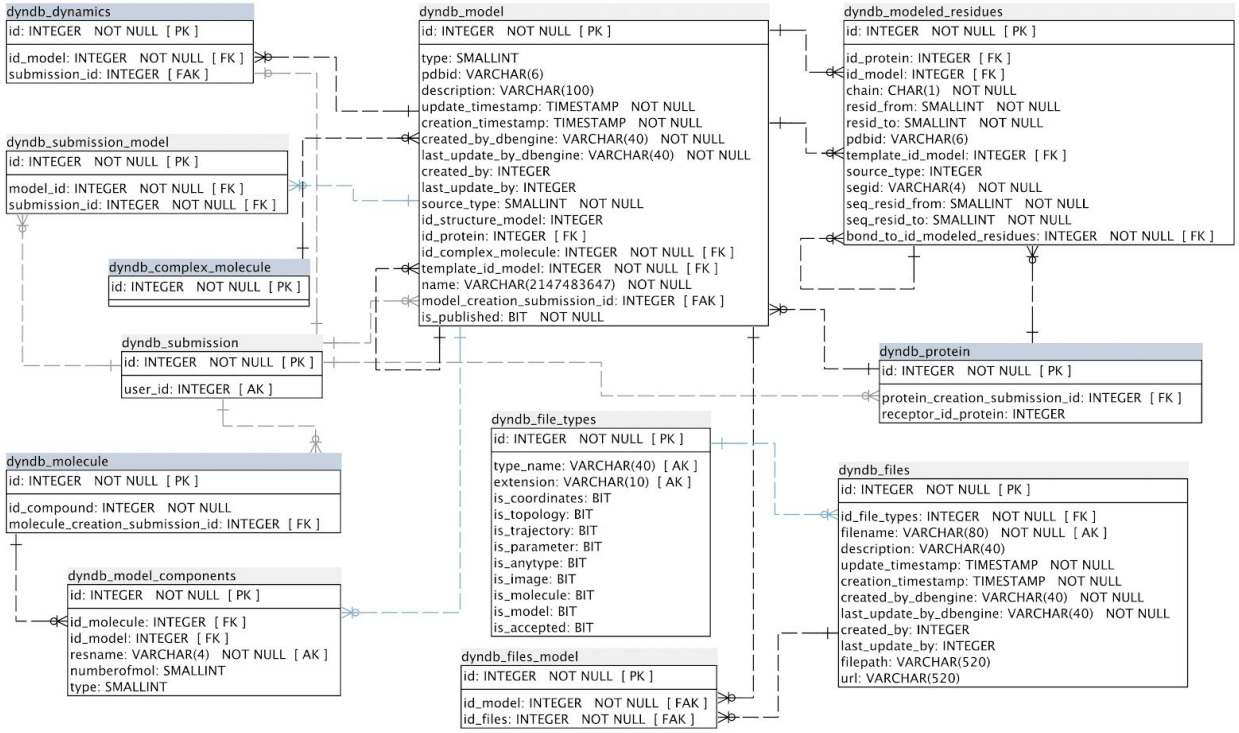
Supplementary Figure 1. Water-mediated contact map of the β2AR highlights the bifurcated network between TM6 and TM7. Systems marked with an asterisk indicate apo-form simulations which are nicely separated from the ligand-receptor complex simulations by the clustering analysis. Within the clustered groups for apo- and complex forms, we see also a separation of active and inactive structures. The bifurcated network which links TM6 to TM7 can be seen in inactive as well as active structures. Interestingly, contact frequencies are reduced in active structures indicating a network loosening. The water-mediated contact map is computed over the accumulated simulation time of 1.5 μs (3 x 0.5 μs) per system.



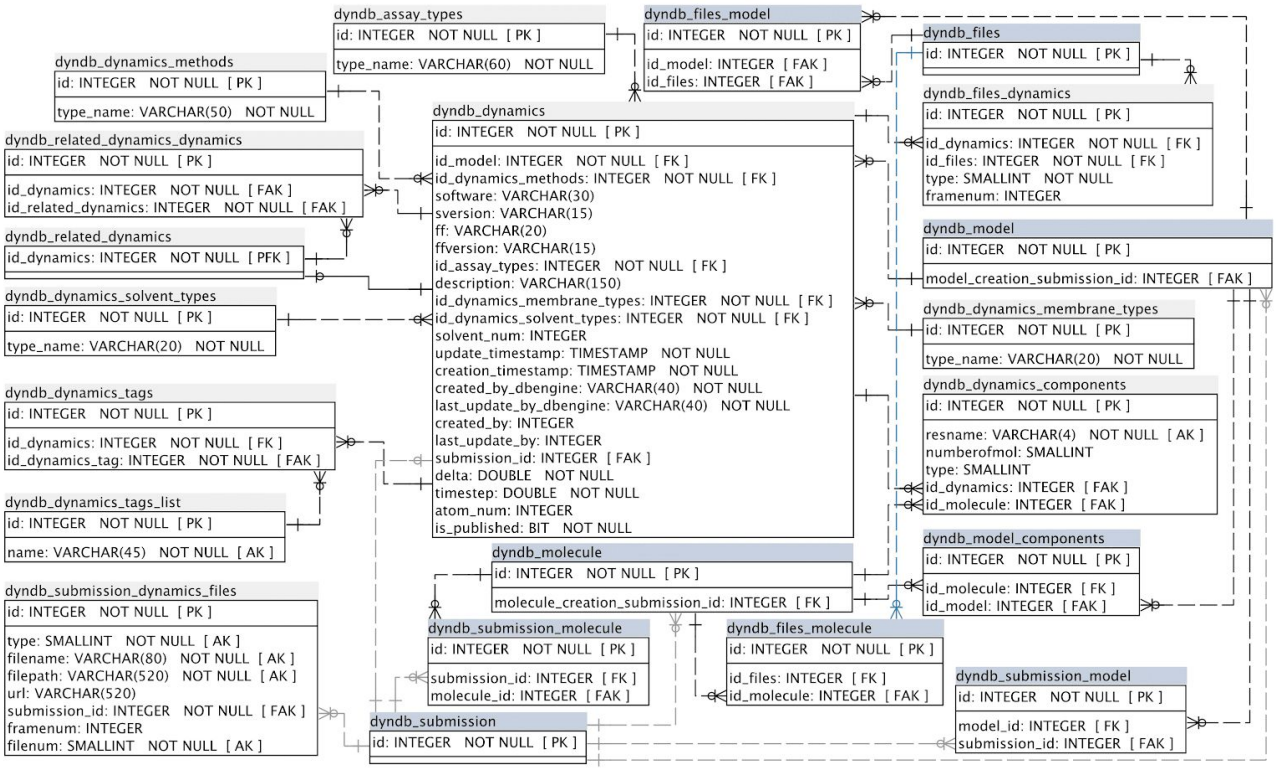
Supplementary Figure 2. GPCRmd entity-relationship (ER) diagram of entities related to protein objects. Tables in blue only display fields taking part in the relationships.



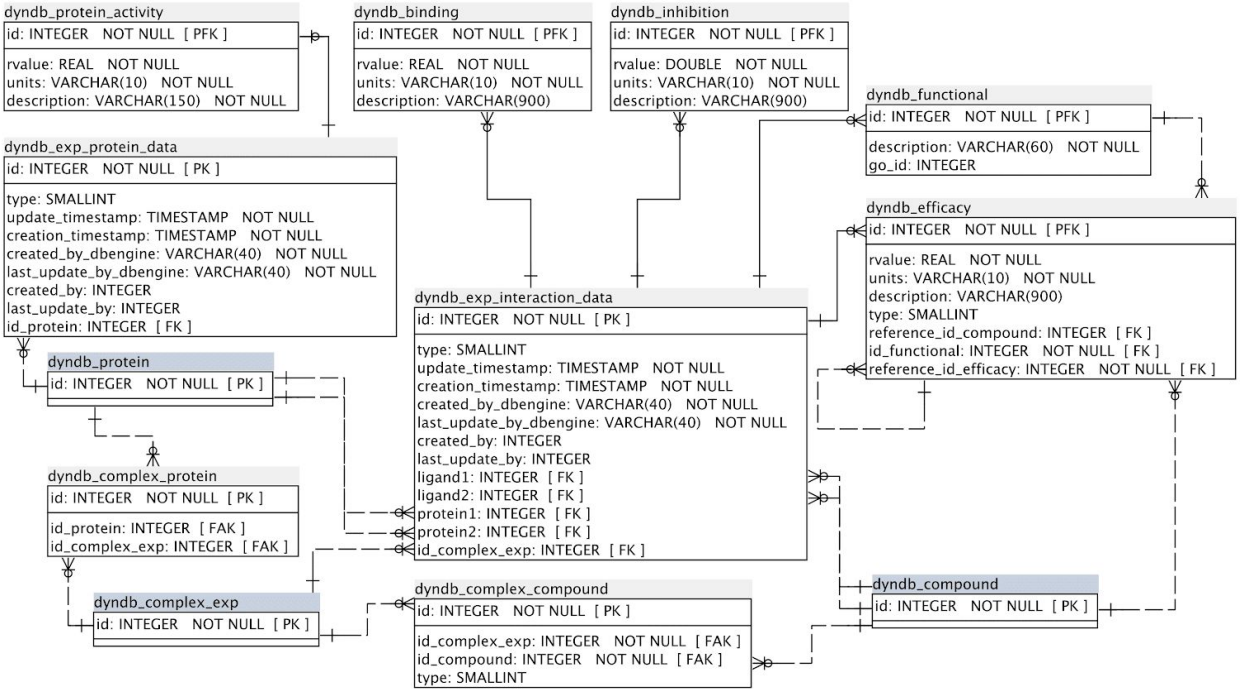
Supplementary Figure 3. GPCRmd ER diagram of entities related to molecule objects. Tables in blue only display fields taking part in the relationships.



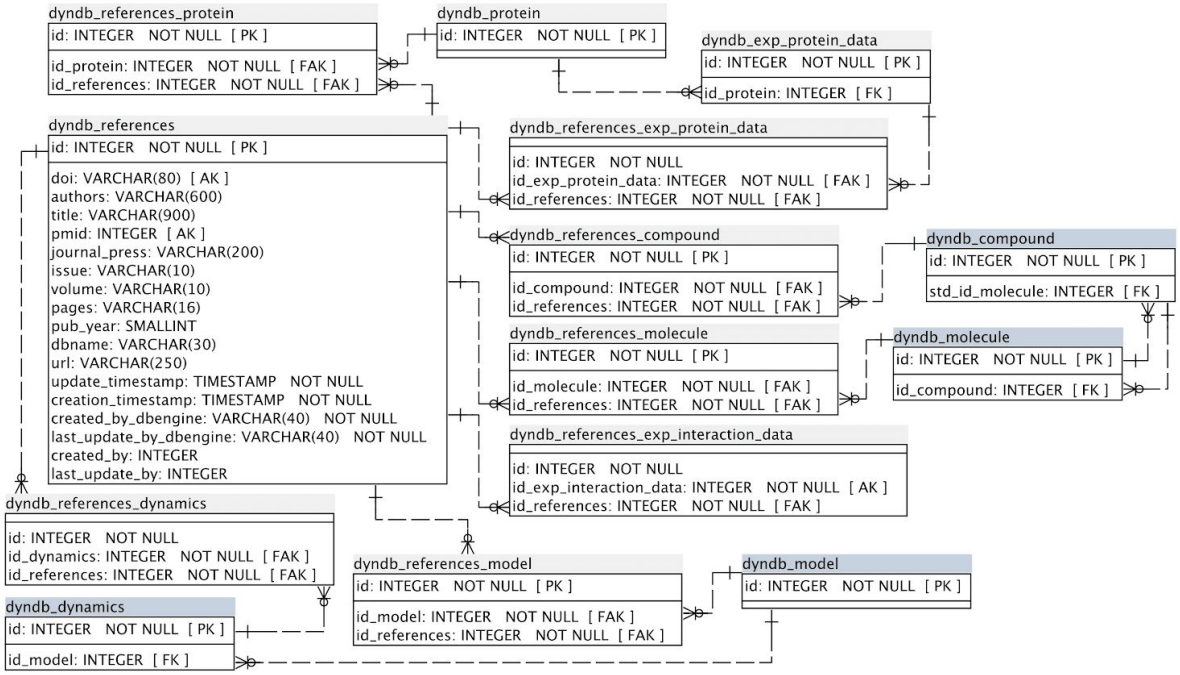
Supplementary Figure 4. GPCRmd ER diagram of entities related to model objects. Tables in blue only display fields taking part in the relationships.



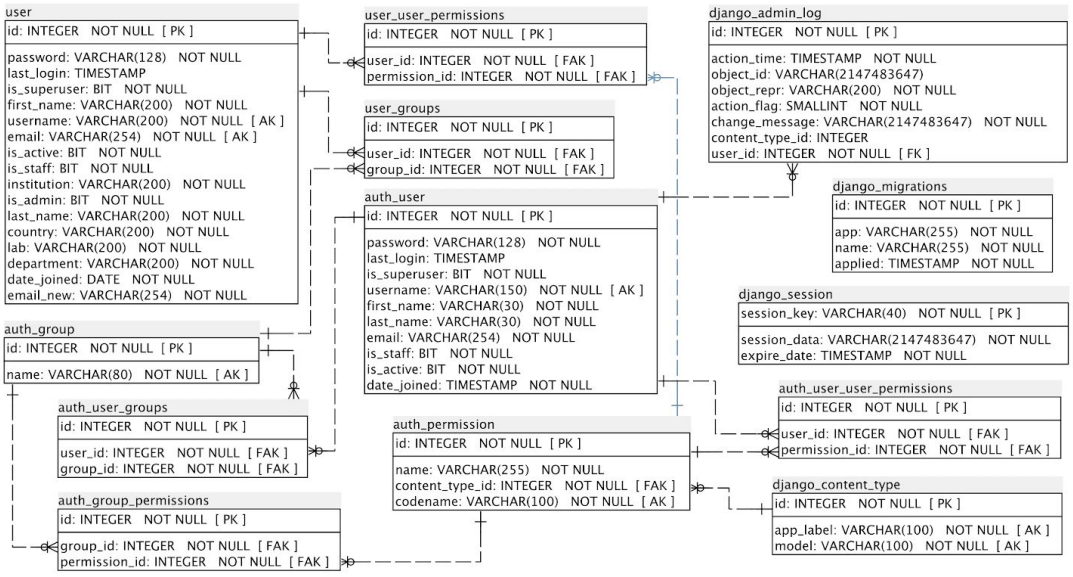
Supplementary Figure 5. GPCRmd ER diagram of entities related to dynamics objects. Tables in blue only display fields taking part in the relationships.



Supplementary Figure 6. GPCRmd ER diagram of entities related to experimental data. Tables in blue only display fields taking part in the relationships.



Supplementary Figure 7. GPCRmd ER diagram of entities related to reference objects. Tables in blue only display fields taking part in the relationships.



Supplementary Figure 8. GPCRmd ER diagram of user and Django tables.

References

1. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
2. Boekel, J. *et al.* Multi-omic data analysis using Galaxy. *Nat. Biotechnol.* **33**, 137–139 (2015).
3. Webb, B. & Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **54**, 5.6.1-5.6.37 (2016).
4. Munk, C. *et al.* GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, 440–446 (2017).
5. Chemical Computing Group ULC. *Molecular Operating Environment (MOE)*. (2014).
6. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **7**, 2284–2295 (2011).
7. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
8. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. a. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–7 (2004).
9. ChemAxon. *Marvin 14.11.24.0*. (2014).
10. Vanommeslaeghe, K. *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
11. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.* **33**, 2451–2468 (2012).
12. Vanommeslaeghe, K. & MacKerell, a D. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–54 (2012).
13. Vanommeslaeghe, K., Raman, E. P. & MacKerell, a D. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52**, 3155–68 (2012).
14. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).
15. Denning, E. J., Priyakumar, U. D., Nilsson, L. & Mackerell, A. D. Impact of 2′-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929–1943 (2011).
16. Mayol, E. et al. HomolWat: a web server tool to incorporate “homologous” water molecules into GPCR structures. *Nucleic Acids Res.* (in press), doi: 10.1093/nar/gkaa440.
17. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 2.0.5*.
18. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
19. Mezei, M. A new method for mapping macromolecular topography. *J. Mol. Graph. Model.* **21**, 463–472 (2003).

20. Morozenko, A., Leontyev, I. V. & Stuchebrukhov, A. A. Dipole Moment and Binding Energy of Water in Proteins from Crystallographic Analysis. *J. Chem. Theory Comput.* **10**, 4618–4623 (2014).
21. Lomize, M. A., Lomize, A. L., Pogozheva, I. D. & Mosberg, H. I. OPM: Orientations of Proteins in Membranes database. *Bioinformatics* **22**, 623–625 (2006).
22. Russell, R. B., Walsh, T. & Barton, G. *STAMP Structural Alignment of Multiple Proteins Version 4.4*.
23. Russell, R. B. & Barton, G. J. Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins Struct. Funct. Bioinforma.* **14**, 309–323 (1992).
24. Huang, J. *et al.* Charmm36M: an Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* (2016) doi:10.1038/nMeth.4067.
25. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
26. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, a. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684 (1984).
27. Harvey, M. J., Giupponi, G. & Fabritiis, G. D. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **5**, 1–9 (2009).
28. Buch, I., Harvey, M. J., Giorgino, T., Anderson, D. P. & De Fabritiis, G. High-Throughput All-Atom Molecular Dynamics Simulations Using Distributed Computing. *J. Chem. Inf. Model.* **50**, 397–403 (2010).
29. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
30. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
31. Loncharich, R. J., Brooks, B. R. & Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylalanyl-N′-methylamide. *Biopolymers* **32**, 523–535 (1992).