Supporting Information to: Identification of Allosteric Mutation Sites and Ligand Effects in the Dopamine Receptor with a Large-Scale Alchemical Mutation Scan

Lisa Schmidt and Bert L. de Groot*

Max Planck Institute for Multidisciplinary Sciences, Department of Theoretical and Computational Biophysics, Group of Computational Biomolecular Dynamics

E-mail: bgroot@gwdg.de

1 FDA analysis and mutational scan of ECC highlight importance of ECL2 for the activation process.

Because of the changed communication in the ECC the cavity openings are slightly different in the active and the inactive state. While the ECC is less deep in the active state and covered by the ECL2, the inactive cavity is protrudes deeper into the receptor.[1–6] This change in ECC opening and communication between the active and the inactive state might also explain the different interactions of agonists and antagonists as reported in the experimental pdb structures of ligand bound D2 receptors.[1–6] While most conventional antagonists (stabilizing inactive state) bind deeper in the ECC[1,4] and form contacts to TM2,TM3,TM5,TM6 and TM7, agonists mainly form contacts to TM3, TM5, TM6 and TM7.[3,5,6]

Since the FDA analysis showed larger changes in the ECC region, we first chose 8 residues

from ECL2 (N175, N176, A177, Q179, N180, C45.50,182, I45.51,183, I45.52,184) as well as W23.50,100 from ECL1 and N402 from ECL3 for our initial mutational scan. N175, A177, I45.52,184 and W23.50,100 all turned out to be strongly mutation inactivated, while I45.51,183 was strongly mutation activated and C45.50,182 was ambivalent. In our wt simulations we had observed long lasting contacts between W23.50,100 in ECL1 and C45.50,182 and I45.51,183 in ECL2. The prediction that mutation of W23.50,100 leads to a stabilization of the inactive over the active state matches with the hypothesis that antagonist interaction with W23.50,100[1,4] leads to a stabilization of the inactive state. For C45.50,182 we observed a stabilization of the active state for the aromatic mutants because of increased π interactions between the aromatic rings of C45.50,182F or C45.50,182Y mutants and W23.50,100. Interesting in ECL2 is that residues from the CII motif (C45.50,182,I45.51,183,I45.52,184), which plays a role in agonists and antagonists interaction[1–5] all displayed multiple high $\Delta\Delta G$ values in different directions.

The attraction between ECL2 and ECL3 that is visible in the FDA in Fig. 2B comes from a higher flexibility of ECL2 in the active state. The main contributing residues here are N175, in ECL2 and N402 in ECL3. This attraction is much weaker in the inactive state because of the formation of the small helix in ECL2 and the following decrease in flexibility of ECL2. N175 forms more interactions with residues in the extracellular TM4 region in the inactive state.

The free energy scan showed that N175 is strongly mutation inactivated, possibly by weakening the ECL2-ECL3 attraction. We do not detect larger $\Delta\Delta G$ values for N402.

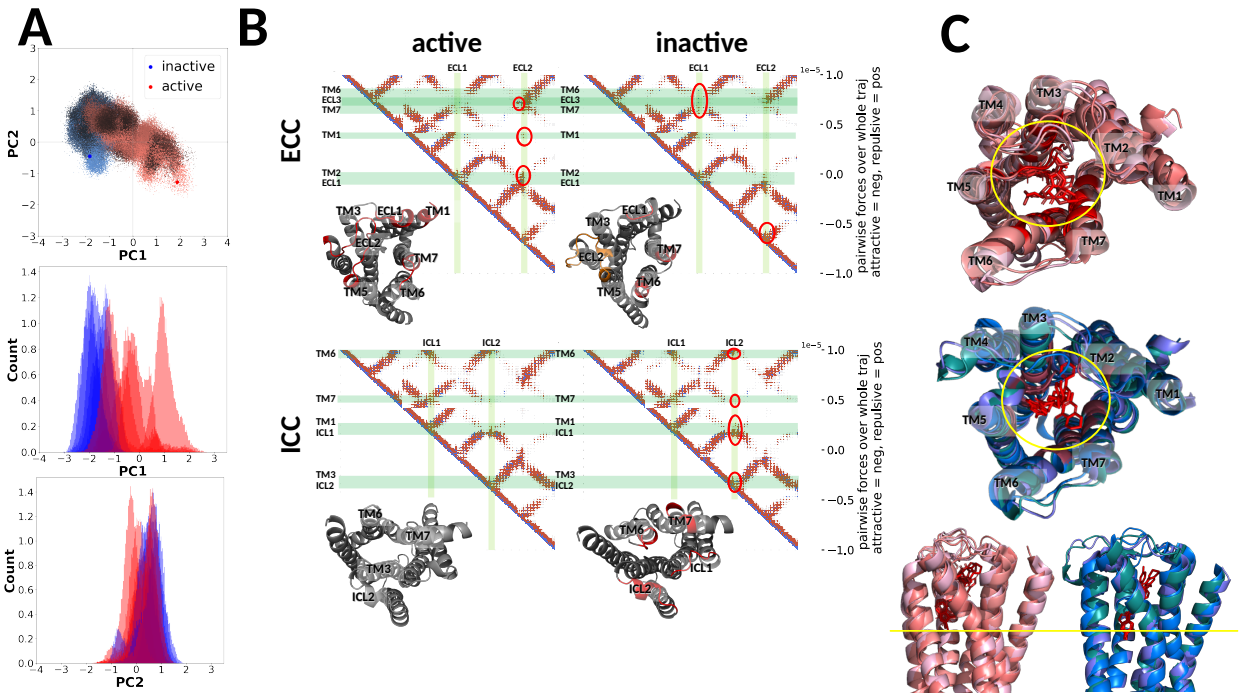

Figure 1: **FDA analysis of the DRD2 receptor.** The FDA analysis shows different behaviour in inactive and active state. Attractive forces are shown in red, repulsive forces in blue.

Another feature of ECL2 is the formation of a small helix in the inactive state and a unfolded an expanded lid-like conformation in the active state.[1,3,7] We analysed the ECL2 dynamics using PCA of only this loop. The first principal component vector corresponds to helix unfolding, the second PC more to a slight twisting of the loop. While there is a more clear separation between active and inactive state simulations in PC1, both states behave similar in PC2.

Interestingly we do observe a tendency towards the inactive more "condensed" state of ECL2 in two (out of 5) active state simulations. In the inactive state we observe partial unfolding of ECL2 only once (out of 5). This can be seen as another indication of the stability of the inactive compared to the active state. However, the receptor can undergo inactivation defined by the inward movement of TM6 without fully "folding" of ECL2.

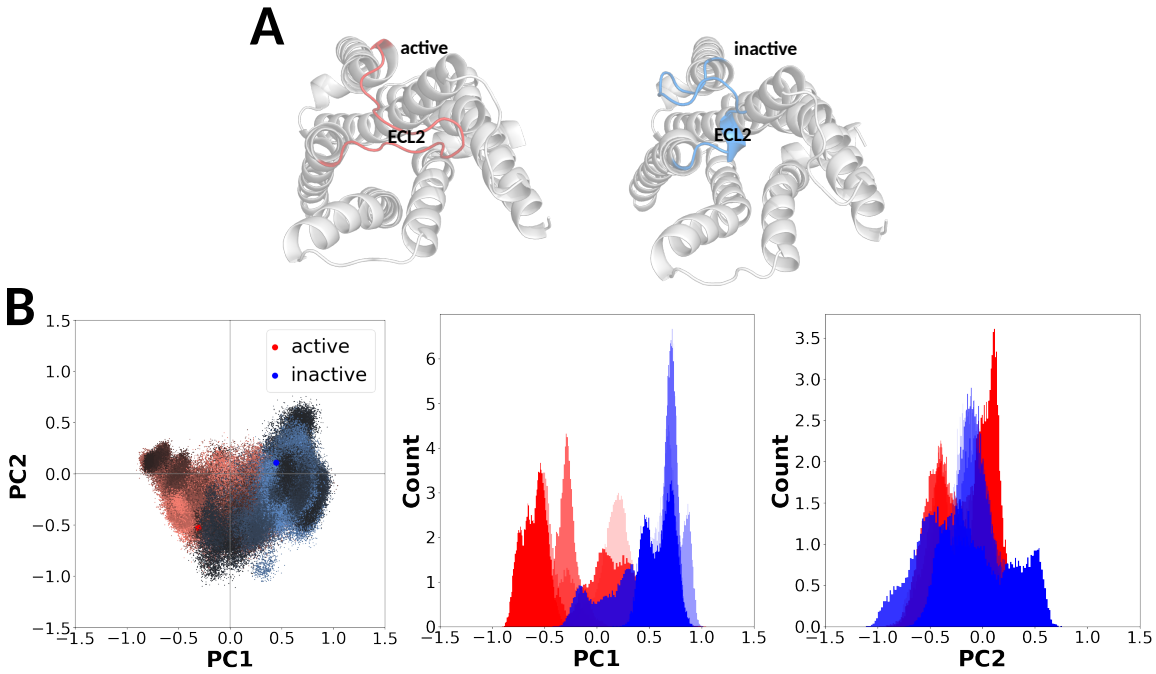

Figure 2: **PCA analysis of ECL2 of the DRD2 receptor.** .

2 Important residues locate differently along the membrane normal.

If our hypothesis were valid that strongly mutation activated residues are found more often closer to the ICC and G-protein binding site, we would expect that this trend holds up for most of our sampled residues. We mapped the conserved residues to the structure and found that there is an increase in conserved residues along the membrane normal towards the ICC, matching with the region of G-protein/arrestin interaction (FIG 3, SI FIG 3).[8,9] Interestingly, the distribution of high $\Delta\Delta G$ residues, along with their mutation activity (active/inactive stabilizing), varies along the membrane normal. Specifically, an accumulation of high $\Delta\Delta G$ residues is observed around the central region of the transmembrane helices. Further, mutation active stabilizing residues are mainly located at both the extracellular and intracellular ends of the receptor, with the majority found in the ICC (5). In contrast, mutation inactivated and ambivalent residues are predominantly located at the center of the transmembrane region (FIG 3). For many mutants we observe a change in residue interaction patterns. Assuming that mutation of high $\Delta\Delta G$ residues not only alters interaction patterns but also disrupts inter-residue contacts, we can explain this localization along the membrane normal with the different opening of ECC and ICC in the active/inactive state. While the ECC remains similarly open for both states, the ICC widens in the active state, resulting in fewer interactions of residues within the ICC.[8,10] This correlates with our findings that mutation of residues in this region more frequently leads to active state stabilization, which is likely due to interaction disruption. Furthermore, although the ECC opening at the extracellular sides appears similar in both states, the extracellular cleft extends deeper into the receptor in the inactive state.[1,4,8,11] This aligns with the accumulation of mutation inactive stabilizing residues in this region. Here, we observe that contact patterns are slightly changed between the states and mutation favors the inactive state by disrupting some interactions.[8,9,12]

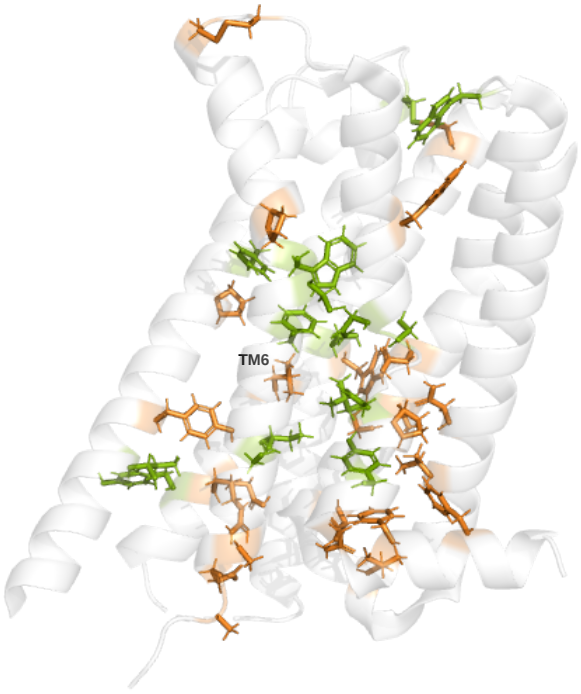

Figure 3: **Conserved residues mapped to the DRD2 structure** Residues that are conserved and have multiple high $\Delta\Delta G$ values are shown in green, conserved residues that have not been sampled or did not show multiple high $\Delta\Delta G$ are shown in orange.

3 Class A sequence alignment

For the sequence alignment we used the sequences from the UniProt database[13] for human GPCR class A aminergic receptors including the Dopamine receptors 1 to 5 (DRD1-5), the Alpha-2A, B and C adrenergic receptors (ADA2A/B/C), the Rhodopsin receptor (OPSD), the Cannabinoid receptor 1 and 2 (CNR1-2), the 5-hydroxytryptamine receptor A, B, D, F and E (5HT1A/B/D/F/E), the Muscarinic acetylcholine receptor M1 to 5 (ACM1-5), the adenosine receptor A2a, A2b, A1 and A3 (AA2AR, AA2BR, AA1R, AA3R). We used the sequence alignement tool of the UniProt database for sequence alignement.[13–17]

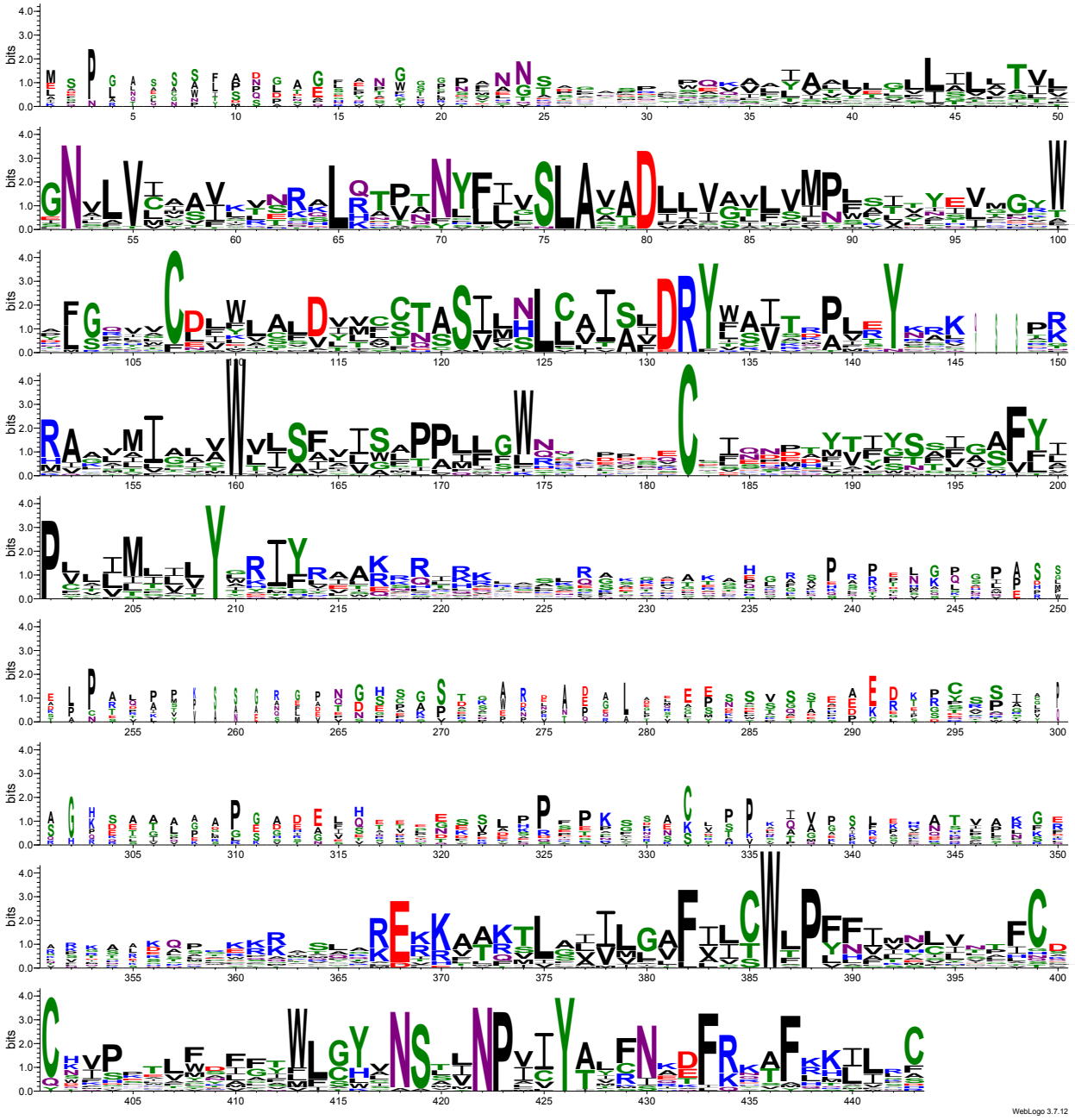

Figure 4: **Sequence logo of the aligned class A GPCRs.** The sequence logo was created with the online tool WebLogo3.[18]

4 Alternative representations of the $\Delta\Delta G$ mutation scan and chemical properties of the DRD2 receptor

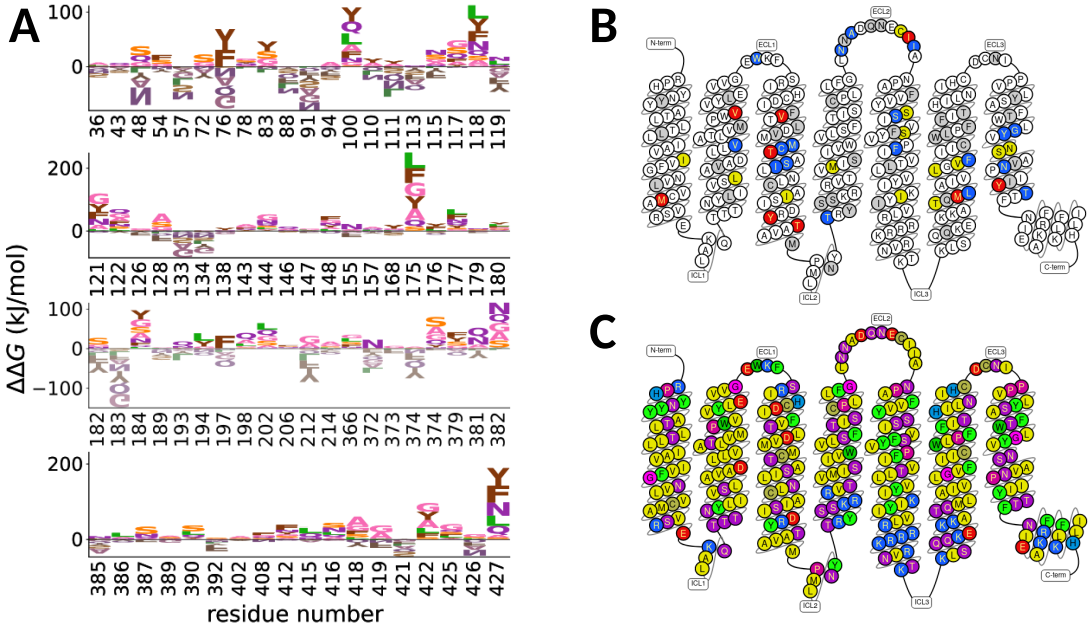

Figure 5: **Sequence Logo of the mutated residues (A), snake plots with residues colored after the relative free energy trends (B) and colored according to their chemical properties(C).** A shows an alternative representation of the mutation scan results. Negative values indicate a stabilization of the active state and positive values a stabilization of the inactive state. The amino acids are colored according to their chemical properties (F, Y brown; N, Q purple; S yellow; A, G pink, L green). The logo was created using the python package logomaker.[19] The residues that are strongly mutation activated (red), strongly mutation inactivated (blue) and ambivalent (yellow) are shown in the snake plot in B. Residues that were mutated but did not show multiple high relative free energies in our scan are colored in grey. C shows the DRD2 residues colored according to their chemical properties. Apolar residues are colored in yellow, glycine in pink, polar residues in purple, negatively charged residues in red, positively charged residues in blue and aromatic residues in green.

5 An overview of the DRD2 residue properties

The following figure shows the results of our analysis. The first part shows the free energy mutation scan and the resulting trends.

The following 4 columns show different aspects of residue positioning in the receptor, including: the position in the Z-axis (membrane normal), the secondary structural element (trans membrane helix TM or loop ECL/ICL), the $C\alpha$ backbone atom distance between the aligned structures of active and inactive receptor state in Å (see main text) as well as the orientation of the residue relative to the receptor center (inward, outward or tangential, see main text).

The next 3 columns show the comparison to literature data and conservance which includes: comparison to known microswitches (green if residue shows multiple high $\Delta\Delta G$ and belongs to the microswitches, red if residue belongs to the microswitches but did not show multiple high $\Delta\Delta G$), comparison to conserved residues according to the class A GPCR alignment (see above; green if conserved and multiple high $\Delta\Delta G$, red if conserved but no multiple high $\Delta\Delta G$) and finally comparison to experimental data from the GPCRdb[20–22] (green if the trends are matching, red if the trends are not matching).

The following column shows the outside pockets that the scanned residues belong to. Residues that are within outside pockets and have multiple high $\Delta\Delta G$ are colored in the respective multiple high $\Delta\Delta G$ trend color (red active, blue inactive, yellow ambivalent).

Results of the structure alignment comparison (DBASS analysis, see main text) are found in the next column. Residues that show differences between the active and inactive state environments and have multiple high $\Delta\Delta G$ are colored in light purple, those that show differences between active and inactive state environments but without multiple high $\Delta\Delta G$ are shown in dark purple.

The last column shows the 12 new residues that do not have experimental comparison but show a clear influence on the active-inactive state-equilibrium when mutated to different amino acids (show multiple high $\Delta\Delta G$).

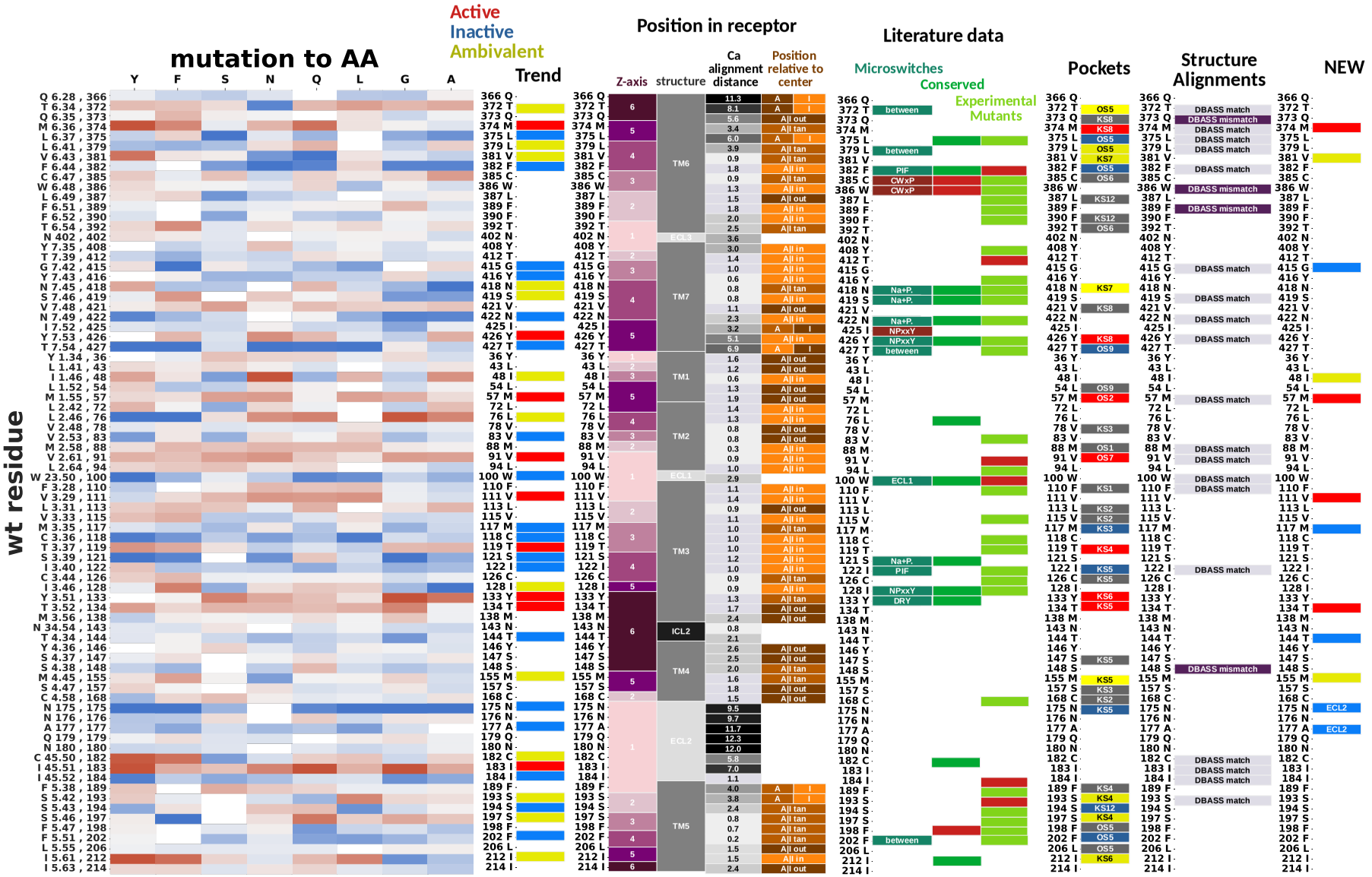

Figure 6: **Relative free energy scan results and residue properties**

5.1 Simulations of mutants show similar trends as free energy scan.

We simulated L2.46,76A (ambivalent, Ala mutant activating), L1.52,54A (no effect) and N7.45,418A (ambivalent, Ala mutant inactivating) for 1000ns (2 replicates per state) and analyzed the trajectories with PCA. Matching the predictions from our free energy scan we find a stabilization of the active state and a slight destabilization of the inactive state in L2.46,76A compared to the wt. For N7.45,418A the trend is less clear, however we find that the active state is less stable than the inactive state which matches the free energy prediction. For L1.52,54A we did not detect high $\Delta\Delta G$ and expected to find no difference in active/inactive state stabilization. Interestingly, we do not find a big difference between the stability of the active/inactive state of this mutant, but compared to the wt we find that both states are stabilized. This indicates that this mutant does not effect the state equilibrium, but stabilizes both states equally.

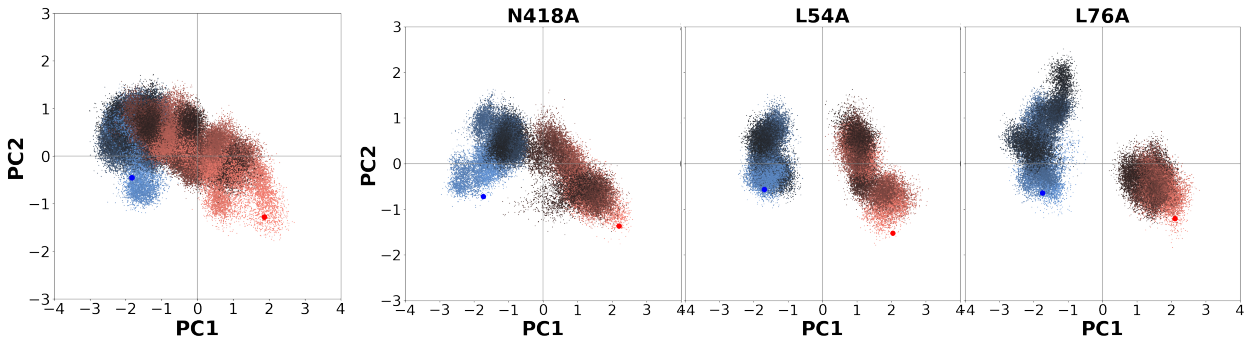

Figure 7: **PCA of wt system and mutant systems.** L2.46,76A (ambivalent, Ala mutant activating), L1.52,54A (no effect) and N7.45,418A (ambivalent, Ala mutant inactivating) were simulated for 1000ns (2 replicates per state). To compare the wt and mutant simulations we only plotted the wt PCA for t=1000ns in this figure.

6 Risperidone Data

6.1 The antagonist risperidone affects different residues.

To examine if ligands effect the mutational relative free energy differences we conducted alanine scans with the ligand risperidone bound to the inactive state binding pocket.

For the ligand scan we used the risperidone position as reported by Wang and coworkers[1] in the inactive receptor and compared the mutational relative free energy of the apo inactive receptor against the risperidone bound inactive receptor. The resulting $\Delta\Delta G$ can be compared to the experimental $\Delta\Delta G$ of ligand binding between wt and mutants. However, since we worked with restraints on the protein backbone and sodium ions, it is expected that the pmx calculated $\Delta\Delta G$ would underestimate the experimental $\Delta\Delta G$. Even though we calculated a Pearson correlation coefficient of 0.97 and a slope of 1.29 for a liner fit of our pmx data, it is clearly visible that we underestimate the binding effect (FIG 8). Since deleting residue-ligand interactions by mutation to alanine should result in destabilization of the ligand bound to the apo state, important residues for ligand binding are expected to show a more positive $\Delta\Delta G$ (stabilization of the unbound state) upon mutation. The addition of risperidone in the inactive state led to a stabilization of the unbound state for most mutants. However, the addition of risperidone had the biggest effect on C6.47,385A and W6.48,386A from the CWxP motif where we detected a difference in $\Delta\Delta G$ between the apo and the risperidone scan of more than 30 kJ/mol. Interestingly both C6.47,385 and W6.48,386 did not show multiple high $\Delta\Delta G$ values in the other scan, even though they are highly conserved and have shown to be important within the activation process.[8,11,23,24] Moreover, all available structures for the D2 receptor on the pdb with different ligands (agonists and antagonists) show ligand-D2 interaction at the CWxP motif. Since the CWxP motif is the only motif where we did not detect any multiple high $\Delta\Delta G$ values in our initial scan but observed a very high difference between the inactive state with and without the ligand we propose that this motif plays a more important role in ligand binding than in stabilization of inactive/

active state without ligand and sodium ions.

A stabilization of the apo inactive state was also observed for Y3.51,133A (DRY motif), F5.47,198A (RISP binding side), C45.50,182A(ECL2, CII ligand binding motif) and I7.52,425A (between NPxxY and DRY motif). We also observe an effect of the ligand for residues that are further apart from the binding pocket, which indicates an allosteric coupling between those residues (Y3.51,133A, I7.52,425A) and ligand binding. This matches very well with the finding that both Y3.51,133 (DRY motif) and I7.52,425 (NPxxY motif) are part of the known allosteric microswitch system.[8,25]

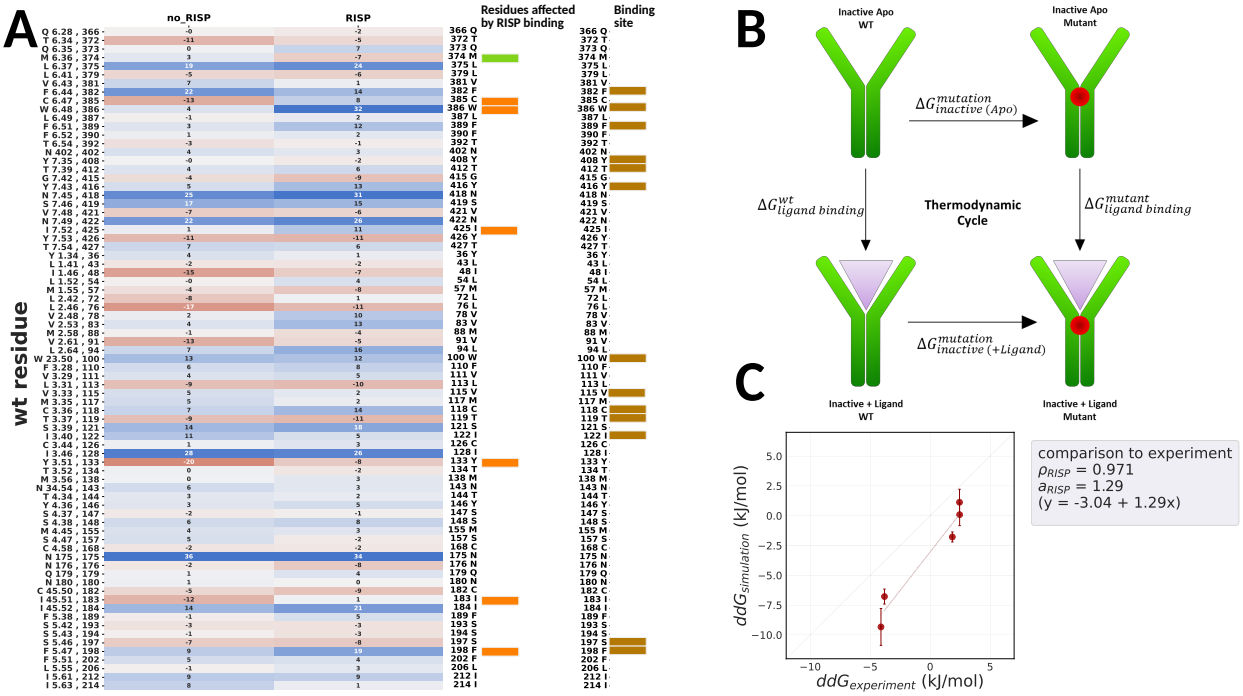

Figure 8: **Risperidone Scan (A), the thermodynamic cycle of ligand binding (B) and comparison to the experiment (C).** The alanine mutation scan shows the relative free energy change of mutation in presence of risperidone in the binding pocket (RISP) and without risperidone (no RISP). Residues with a difference of more than 10 kJ/mol in $\Delta\Delta G$ between the two scans were classified as 'affected by risperidone binding' and are marked in orange (risperidone binding leads to more positive $\Delta\Delta G$, stabilization inactive state) and green (risperidone binding leads to more negative $\Delta\Delta G$, stabilization active state). For comparison the binding site residues are shown in the last map in panel A. The comparison with experimental data is shown in panel C using the thermodynamic cycle from panel B.

6.2 Simulation of the ligand bound systems

The active receptor was solved with the ligand bromocriptine bound (pdb 7jvr)[3] and the inactive structure was solved with risperidone bound (pdb 6cm4).[1] We parameterized risperidone using CGenFF.[26,27] Wang et al. had already docked risperidone into the inactive receptor structure,[1] so we used this experimentally defined position as a starting point for our mutations. We applied position restraints on all C_α protein backbone atoms and all Na^+ as described above. There were no position restraints applied to risperidone. There are only five experimental data points available for the Kd of risperidone binding to different single mutants of DRD2. We used these and the wild type data to calculate the experimental $\Delta\Delta G^{ligandbinding}_{mutation}$ of ligand binding in mutated systems with the following equations:

$$\Delta G = -\ln(K_d)\mathrm{R}T$$

$$\Delta\Delta G^{ligandbinding}_{mutation} = \Delta G^{ligandbinding}_{mutant} - \Delta G^{ligandbinding}_{wt}$$

$$= \Delta G^{ligand}_{mutation} - \Delta G^{noligand}_{mutation}$$

where T is the temperature where the experiments were performed 298 K, R is the gas constant 8.314472 J/mol $\cdot$ K

The simulations were compared to the experimental data with the second equation.

References

(1) Wang, S.; Che, T.; Levit, A.; Shoichet, B. K.; Wacker, D.; Roth, B. L. Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* **2018**, *555*.

(2) Im, D. et al. Structure of the dopamine D2 receptor in complex with the antipsychotic

drug spiperone. *Nature Communications* **2020**, *11*, 6442.

(3) Zhuang, Y. et al. Structural insights into the human D1 and D2 dopamine receptor signaling complexes. *Cell* **2021**, *184*, 931–942.e18.

(4) Fan, L.; Tan, L.; Chen, Z.; Qi, J.; Nie, F.; Luo, Z.; Cheng, J.; Wang, S. Haloperidol bound D2 dopamine receptor structure inspired the discovery of subtype selective ligands. *Nature Communications* **2020**, *11*, 1074.

(5) Yin, J.; Chen, K.-Y. M.; Clark, M. J.; Hijazi, M.; Kumari, P.; Bai, X.-c.; Sunahara, R. K.; Barth, P.; Rosenbaum, D. M. Structure of a D2 dopamine receptor–G-protein complex in a lipid membrane. *Nature* **2020**, *584*, 125–129.

(6) Xu, P. et al. Structural genomics of the human dopamine receptor system. *Cell Research* **2023**, *33*, 604–616.

(7) Nicoli, A.; Dunkel, A.; Giorgino, T.; de Graaf, C.; Di Pizio, A. Classification Model for the Second Extracellular Loop of Class A GPCRs. *Journal of Chemical Information and Modeling* **2022**, *62*, 511–522, PMID: 35113559.

(8) Zhou, Q. et al. Common activation mechanism of class A GPCRs. *eLife* **2019**, *8*, e50279.

(9) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Heydenreich, F. M.; Flock, T.; Miljus, T.; Balaji, S.; Bouvier, M.; Veprintsev, D. B.; Tate, C. G.; Schertler, G. F. X.; Babu, M. M. Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. *Nature* **2016**, *536*, 484–487.

(10) Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chemical Reviews* **2017**, *117*, 139–155.

(11) Selent, J.; Sanz, F.; Pastor, M.; De Fabritiis, G. Induced Effects of Sodium Ions on

Dopaminergic G-Protein Coupled Receptors. *PLOS Computational Biology* **2010**, *6*, 1–6.

(12) Nygaard, R. et al. The Dynamic Process of 2-Adrenergic Receptor Activation. *Cell* **2013**, *152*, 532–542.

(13) Pundir, S.; Martin, M. J.; O'Donovan, C.; Consortium, T. U. UniProt Tools. *Current Protocols in Bioinformatics* **2016**, *53*, 1.29.1–1.29.15.

(14) Consortium, T. U. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* **2022**, *51*, D523–D531.

(15) Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **2018**, *47*, D506–D515.

(16) Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research* **2020**, *49*, D480–D489.

(17) Pundir, S.; Martin, M. J.; O'Donovan, C.; Consortium, T. U. UniProt Tools. *Current Protocols in Bioinformatics* **2016**, *53*, 1.29.1–1.29.15.

(18) Gavin E. Crooks, J.-M. C., Gary Hon; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Research* **2004**, *14*, 1188–1190.

(19) Tareen, A.; Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **2019**, *36*, 2272–2274.

(20) Isberg, V.; Vroling, B.; van der Kant, R.; Li, K.; Vriend, G.; Gloriam, D. E. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Research* **2013**, *42*, D422–D425.

(21) Munk, C.; Harpsøe, K.; Hauser, A. S.; Isberg, V.; Gloriam, D. E. Integrating structural and mutagenesis data to elucidate GPCR ligand binding. *Current Opinion in Pharmacology* **2016**, *30*, 51–58.

(22) Pándy-Szekeres, G.; Munk, C.; Tsonkov, T. M.; Mordalski, S.; Harpsøe, K.; Hauser, A. S.; Bojarski, A. J.; Gloriam, D. E. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Research* **2017**, *46*, D440–D446.

(23) Michino, M.; Free, R. B.; Doyle, T. B.; Sibley, D. R.; Shi, L. Structural basis for Na+-sensitivity in dopamine D2 and D3 receptors. *Chemical Communications* **2015**, *51*, 8618–8621.

(24) Draper-Joyce, C. J.; Verma, R. K.; Michino, M.; Shonberg, J.; Kopinathan, A.; Klein Herenbrink, C.; Scammells, P. J.; Capuano, B.; Abramyan, A. M.; Thal, D. M.; Javitch, J. A.; Christopoulos, A.; Shi, L.; Lane, J. R. The action of a negative allosteric modulator at the dopamine D2 receptor is dependent upon sodium ions. *Scientific Reports* **2018**, *1208*.

(25) Hauser, A. S.; Kooistra, A. J.; Munk, C.; Heydenreich, F. M.; Veprintsev, D. B.; Bouvier, M.; Babu, M. M.; Gloriam, D. E. GPCR activation mechanisms across classes and macro/microscales. *Nature Structural Molecular Biology* **2021**, *28*, 879–888.

(26) Vanommeslaeghe, K.; MacKerell, A. D. J. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144–3154.

(27) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. J. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *Journal of Chemical Information and Modeling* **2012**, *52*, 3155–3168.