



Fine-tuning activation specificity of G-protein-coupled receptors via automated path searching

Rujuan Ti^{a,1} , Bin Pang^{b,1} , Leiye Yu^b, Bing Gan^c , Wenzhuo Ma^a, Arie Warshel^{d,2}, Ruobing Ren^{b,e,2} , and Lizhe Zhu^{a,2}

Contributed by Arie Warshel; received October 15, 2023; accepted January 12, 2024; reviewed by Ioan Andricioaei and Xiangyu Liu

Physics-based simulation methods can grant atomistic insights into the molecular origin of the function of biomolecules. However, the potential of such approaches has been hindered by their low efficiency, including in the design of selective agonists where simulations of myriad protein–ligand combinations are necessary. Here, we describe an automated input-free path searching protocol that offers (within 14 d using Graphics Processing Unit servers) a minimum free energy path (MFEP) defined in high-dimension configurational space for activating sphingosine-1-phosphate receptors (S1PRs) by arbitrary ligands. The free energy distributions along the MFEP for four distinct ligands and three S1PRs reached a remarkable agreement with Bioluminescence Resonance Energy Transfer (BRET) measurements of G-protein dissociation. In particular, the revealed transition state structures pointed out toward two S1PR3 residues F263/I284, that dictate the preference of existing agonists CBP307 and BAF312 on S1PR1/5. Swapping these residues between S1PR1 and S1PR3 reversed their response to the two agonists in BRET assays. These results inspired us to design improved agonists with both strong polar head and bulky hydrophobic tail for higher selectivity on S1PR1. Through merely three *in silico* iterations, our tool predicted a unique compound scaffold. BRET assays confirmed that both chiral forms activate S1PR1 at nanomolar concentration, 1 to 2 orders of magnitude less than those for S1PR3/5. Collectively, these results signify the promise of our approach in fine agonist design for G-protein-coupled receptors.

GPCR | activation path | drug specificity | agonist design | sphingosine-1-phosphate

G-protein-coupled receptors (GPCRs) represent the largest superfamily of membrane proteins in human genome (1). Distributed in most cell types, GPCRs play essential roles in numerous diverse physiological processes (2) and are currently the targets for ~35% of FDA (Food and Drug Administration)-approved drugs (3). The binding of an agonist to this receptor triggers a change of the packing of its 7 transmembrane helices (TM), which is further transmitted intracellularly, causing the downstream G-protein to associate with the receptor, disassemble due to exchange of Guanosine diphosphate (GDP) by Guanosine triphosphate (GTP), and then initiates downstream signaling.

As the 800 members of the GPCR family can be activated by a plethora of ligands, specificity lies in the core of agonist design for GPCRs, especially those in the same subfamily. Note that the class A GPCR alone already contains 30 subfamilies with more than three members (Fig. 1*A*). The high sequence similarity among the members in the same subfamily poses grave challenges for selective agonist design. Agonists with moderate structural differences may prefer to activate distinct subsets of the members. For instance, the five members of the Sphingosine 1-phosphate receptor (S1PR) family (4) share ~70% sequence similarity in their agonist-binding pockets (Fig. 1*B*). As a result, most existing agonists, such as Fingolimod-Phosphate, CBP-307, and Siponimod (BAF-312), do not activate S1PR1 exclusively, but also some of the other four members (5–7) (Fig. 1*C*). Such promiscuity undermines the efficacy of these ligands in extending indications and reducing side effects when the endogenous S1P signals are hijacked indistinguishably for disease treatments, calling for the design of highly selective agonist for S1PR1 (8–10). However, elaborate design requires a thorough understanding of the activation mechanism, which is grueling for existing technology. Structural biology techniques typically provide precious atomistic structures for the inactive receptor and active receptor–ligand complex but are unable to capture the crucial short-lived transition state (TS) for activation. Besides, structural biology, and likewise mutagenesis of the receptor, remains expensive and far from high-throughput to empower systematic dissections of multiplex receptor–ligand combinations.

Physics-driven molecular dynamics simulations (MD) are able in principle to describe the fine details of GPCR activation but have been suffering from low efficiency mainly due to the expensive force calculations. Consequently, brute force MD has been mostly

Significance

G-protein-coupled receptors (GPCRs) are the targets for ~35% of Food and Drug Administration-approved drugs. Yet designing selective GPCR agonists remains challenging, because it requires a thorough understanding of the activation mechanism of several GPCRs by various ligands. Molecular dynamics simulations can offer fine details of activation but suffer from low efficiency. To tackle this efficiency bottleneck, we present an automated input-free protocol, built upon the travelling-salesman automated path searching (TAPS). Our approach not only dissected the activation of Sphingosine-1-phosphate receptors (S1PRs) by existing agonists for treating autoimmune diseases but also enabled the design of a unique compound that activates S1PR1 exclusively. This mechanism-based design minimized wet-lab costs through three computational iterations and a single synthesis of the final compound.

Author contributions: R.T., B.P., A.W., R.R., and L.Z. designed research; R.T., B.P., L.Y., and B.G. performed research; R.T. contributed new reagents/analytic tools; R.T., B.P., W.M., A.W., R.R., and L.Z. analyzed data; and R.T., B.P., A.W., R.R., and L.Z. wrote the paper.

Reviewers: I.A., University of California Irvine; and X.L., Tsinghua University.

The authors declare no competing interest.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹R.T. and B.P. contributed equally to this work.

²To whom correspondence may be addressed. Email: warshel@usc.edu, renruobing@fudan.edu.cn, or zhulizhe@cuhk.edu.cn.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2317893121/-DCSupplemental>.

Published February 12, 2024.

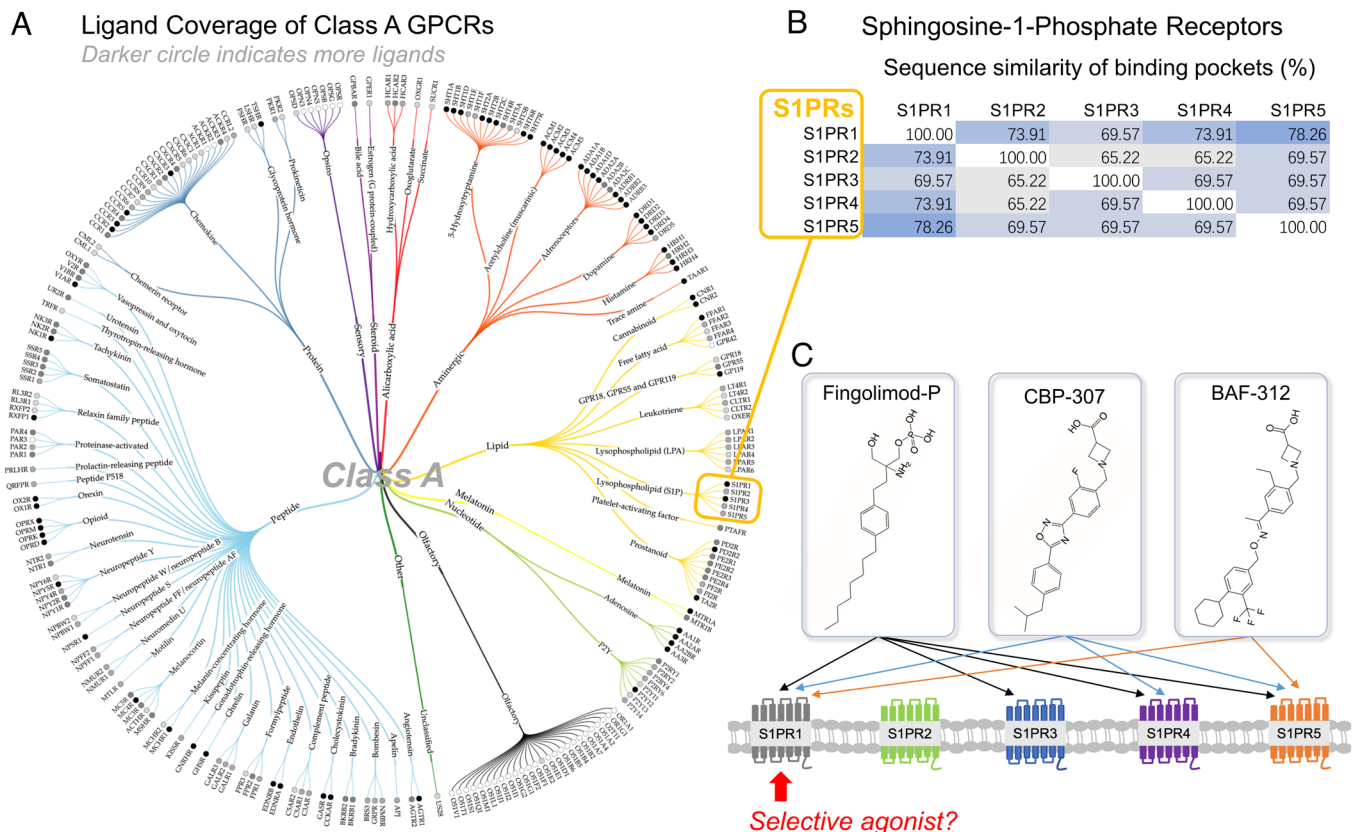


Fig. 1. Problem of specificity in the agonist design for class A GPCRs. (A) Subfamilies of Class A GPCRs and the corresponding coverage of known ligands. The figure is drawn by GPCRdb (<https://gpcrdb.org/>). (B) Sequence similarity of the agonist-binding pocket among the five members of the sphingosine-1-phosphate receptor family (S1PRs). (C) Pattern of the activation of S1PRs by three representative agonists Fingolimod-Phosphate, CBP-307, and Siponimod (BAF-312), arrows indicate activation.

used for local structure optimization or single case study after wet-lab discovery (11–17) rather than as a systematic prediction tool that could reduce the number of necessary wet-lab attempts. Nevertheless, the use of specialized computers has provided important insight (11, 15). Alternative coarse-grained approaches (18, 19) offer ways to gain molecular insight on the landscape of GPCR activation process (20, 21). However, obviously it is desirable to exploit the increase in computer power and the advance in simulation strategies to obtain a more microscopic description.

Results

Automated Search of GPCR Activation Path. MD-based path methods are algorithms that locate the minimum free energy path (MFEP) closest to a given initial path for a conformational change of interest (22–27). Focusing the sampling process on the transition instead of the stable states, path methods can in principle provide rigorous characterization of the conformational change in a cost-effective manner (28). Given the abundance of resolved inactive and active forms of various GPCRs (29) and the tremendous progress in structure prediction by AlphaFold2 (30) or RosettaFold (31), path methods are particularly suitable for dissecting the activation process between the two states.

Nonetheless, most path methods require guessed mechanism (collective variables with physical meanings, CVs) as input (23–26) and therefore cyclic sampling for input validation, which prevented their large-scale applications (*SI Appendix, Fig. S1*) (32). Moreover, as the MFEP therein is typically defined in a low dimensional CV space, the interpretation of the activation mechanism

could be incomplete and therefore unable to inspire ligand modifications for enhanced selectivity.

To tackle these issues, some of us have recently developed a Travelling-salesman based Automated Path Searching (TAPS) approach (33) that has successfully located the MFEP for biomolecular systems with hundreds of residues at cost of sub-microsecond simulations (34–36). TAPS only requires a distance metric as input, i.e., the set of atoms used to compute the root-mean-square-distance (RMSD) between any pair of conformations (Fig. 2A). Such a feature is particularly useful for a general prediction of an arbitrary ligand-receptor complex, where the vital atoms underlying the activation process are unknown a priori. Full automation can be achieved by using all heavy atoms of the complex for RMSD calculations, yielding an MFEP composed of complete complex structures.

Simulation-Guided GPCR Agonist Design. Centered around the TAPS approach, we devised the following protocol for designing selective GPCR agonists. All steps for the path-searching process are designed input-free to predict the activation potency of arbitrary ligands and reveal the corresponding activation mechanism (Fig. 2B ①–④). Without assuming prior knowledge of the ligand-binding pocket, blind docking is first performed to obtain initial structural models of the receptor-ligand complex for both the inactive and active states, followed by short simulations for local structural optimization (Fig. 2B ①). Targeted MD (37) is then used to drag the complex from the inactive to the active model, with the dragging force applied to all heavy atoms of the complex (Fig. 2B ②). The activation MFEP is finally obtained by

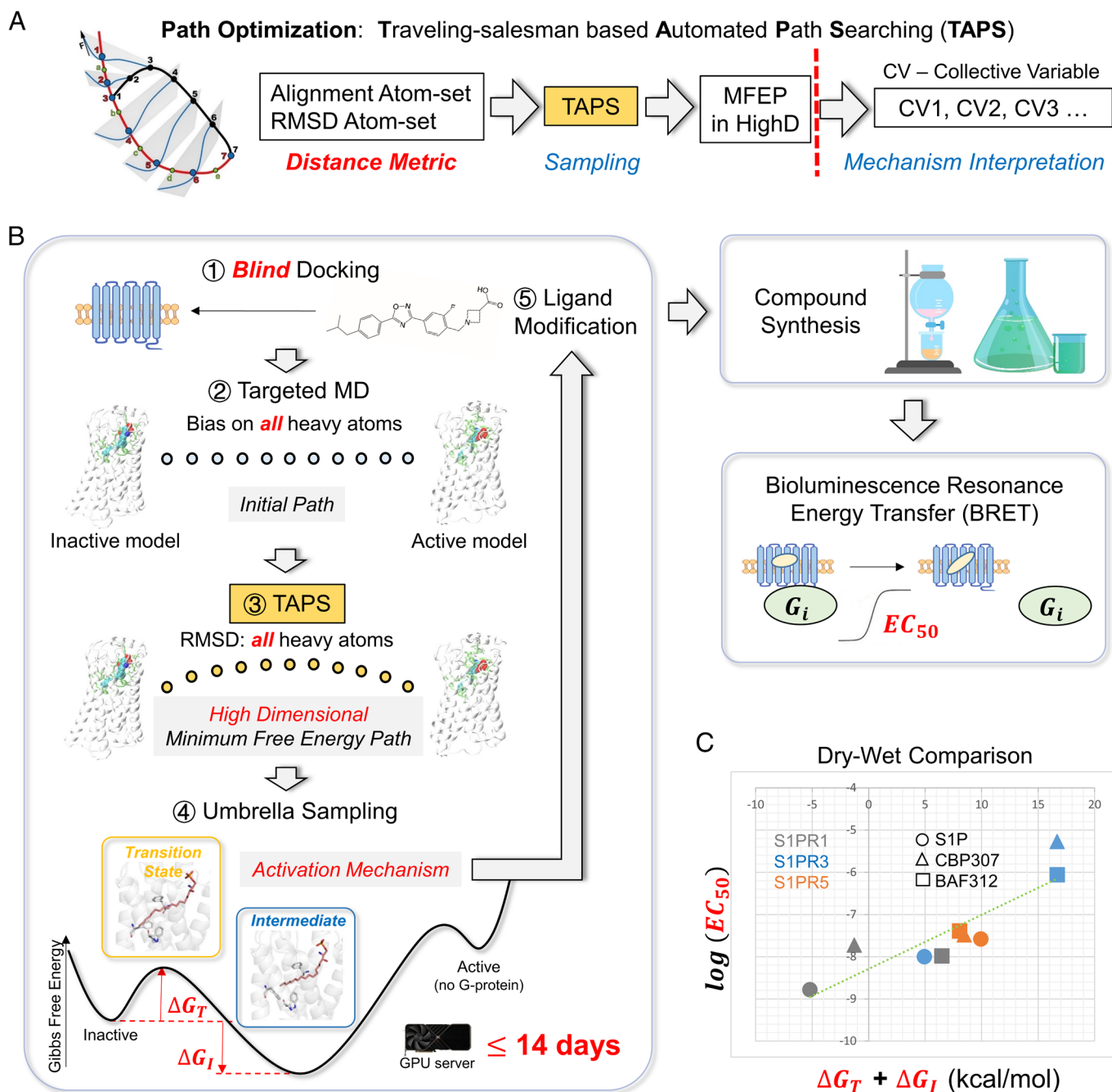


Fig. 2. Automated search of activation path for rational GPCR agonist design. (A) Travelling-salesman-based automated path searching needs only a distance metric as input and yields a path composed by complete biomolecular structures with full information (in high-dimension configurational space) for mechanism interpretation. (B) Protocol for rational design of selective GPCR agonist: guess the inactive and active structure of the receptor-ligand complex via blind docking (no G-protein); use targeted MD that bias on all heavy atoms of the complex to generate an initial activation path; use all heavy atoms to define RMSD and TAPS to find the MFEP; perform umbrella sampling to obtain the free energy distribution along the MFEP; modify the ligand structure for enhanced selectivity based on the TS and intermediate structures; pick the most selective compound for synthesis and BRET assay for validation. (C) Comparison between computational prediction (x-axis, summation of activation barrier height and stability of IS) and wet-lab measurements (y-axis, logarithm of EC_{50} in BRET assay) for nine receptor-ligand combinations (colors indicate receptors; shapes indicate ligands).

TAPS with RMSD defined by all heavy atoms (Fig. 2B ③). We then calculated the path collective variable (PCV) (38) of the MFEP, with RMSD defined again by all heavy atoms. The two variables s and z of the PCV approximate the progress along the MFEP and the average distance from a high dimensional point to the MFEP, respectively (38). Umbrella sampling (39) is then performed on PCV-s, yielding the Gibbs free energy distribution along the MFEP (Fig. 2B ④).

Note that here for minimizing computational cost, our simulations do not involve the downstream G-protein, which assumes that the coupling of G-protein occurs after the ligand drives the receptor toward an intermediate state (IS) which almost resembles the structural features of the fully active state, e.g., outward movement of transmembrane helix 6 (TM6). This assumption generally holds at least for GPCRs without obvious basal activity (12). For GPCRs with basal activity, the G-protein can be included in the

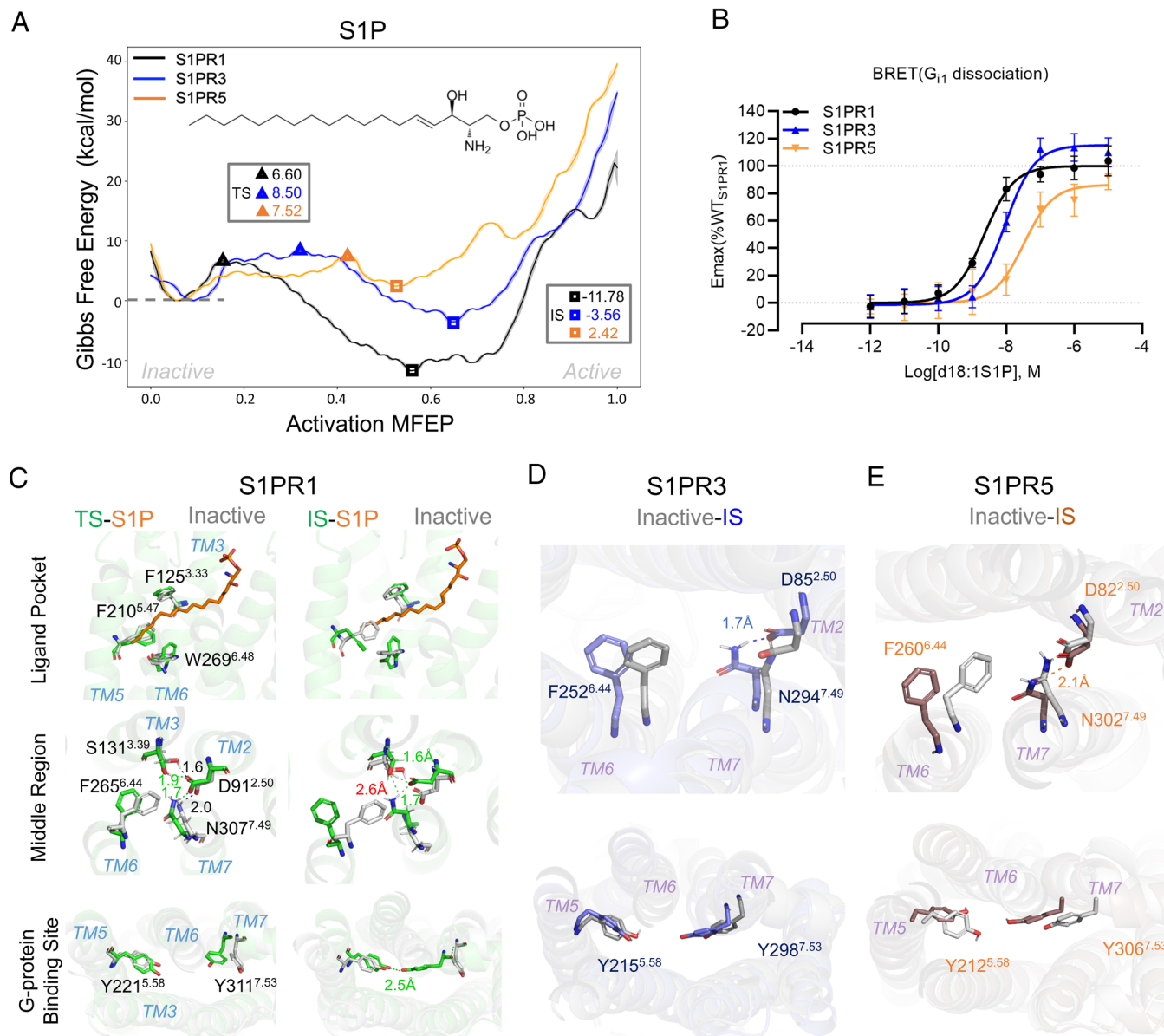


Fig. 3. Activation mechanism of S1PR1/3/5 by endogenous agonist S1P. (A) Predicted Gibbs free energy curves along the activation paths with the transparent bands denoting the errors, the height of activation barriers, and free energy difference between the IS and the inactive state are labeled by triangles/TS and squares/IS respectively; (B) BRET measurements of G_{11} dissociation; (C) structural comparison of the inactive state (gray), TS (green), and IS (green) for S1P-S1PR1; (D) structural comparison of the inactive (gray) and IS (purple) for S1P-S1PR3; (E) structural comparison of the inactive (gray) and IS (orange) for S1P-S1PR5. Errors in the free energy are plotted in the same color as the free energy curves but transparent.

simulation with initial path generated by dragging the complex from the G-protein-coupled active state to the inactive one.

For our current setup without the G-protein, the free energy distribution along the MFEP typically reveals an activation TS, an intermediate that is structurally similar to the fully active state (Fig. 2B). The active state will exhibit “apparent instability” due to the absence of G-protein stabilization. The activation potency of the tested ligand can be measured by the sum of the barrier height ΔG_T and the stability of the intermediate with respect to the inactive state ΔG_I .

The full information of the activation mechanism, encoded in the complete complex structures of the TS and the intermediate, can be used to guide the modification of the ligand for tuning its activation potency for a series of receptors and, therefore, specificity enhancement. Notably, on a state-of-the-art 4/8-card Graphics Processing Unit (GPU) server, one iteration of such in silico prediction takes less than 14 d on the same timescale of

subsequent compound synthesis and functional assay. Here, we used Bioluminescence Resonance Energy Transfer (BRET) (40) to measure the dissociation of the G-protein for functional assay and validation of the designed agonists.

We first tested this protocol on three S1PRs (S1PR1/3/5) and three existing agonists, including CBP-307, Siponimod (BAF-312), and the endogenous agonist S1P. As shown in Fig. 2C, our computational prediction ($\Delta G_T + \Delta G_I$) reached remarkable agreement with the EC₅₀ of BRET assay for all 9 receptor-ligand combinations. Such accuracy signifies the potential of this approach in reducing wet-lab costs and offers strong support for subsequent design of new agonists.

Activation Mechanism of S1PR by Endogenous Agonist S1P. For the natural agonist S1P, all three S1PRs (S1PR1/3/5) share a similar activation barrier of height 6 to 8 kcal/mol in our prediction (Fig. 3A). The major differences among the receptors were found in the

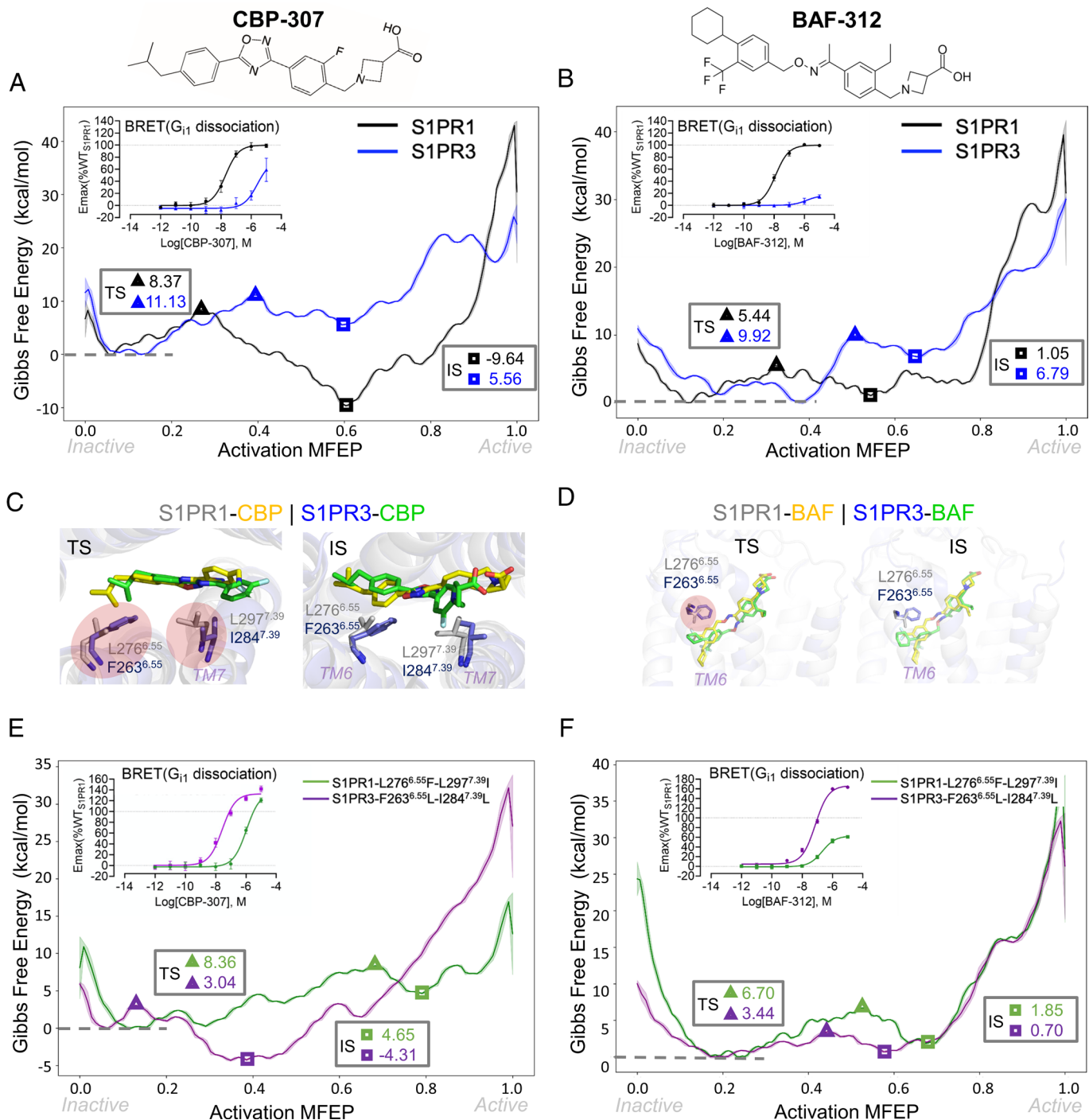


Fig. 4. Activation mechanism of S1PR1/3 by CBP-307 and BAF-312. Free energy distributions along the activation paths are plotted with the transparent bands denoting the errors: (A) CBP-307, S1PR1/3; (B) BAF-312, S1PR1/3. BRET measurements of G_{i1} dissociation are given as the *Insets*; structures of the transition and IS are presented for (C) CBP-307, S1PR1/3; and (D) BAF-312, S1PR1/3. Free energy distributions of the receptor mutants are plotted: (E) CBP-307, S1PR1/3 mutants; (F) BAF-312, S1PR1/3 mutants.

stability of the IS: -11.78, -3.56, and 2.42 kcal/mol for S1PR1/3/5, respectively. Consistent with BRET measurements (Fig. 3B, EC50 reported in *SI Appendix, Table S1*), these results pointed to decreasing activation potency of S1P for S1PR1, S1PR3, and S1PR5.

Throughout the activation process, the strong polar head of S1P is stably captured by all three receptors (Fig. 3C). In S1PR1, the hydrophobic tail of S1P interacts with the bulky side chains of F210^{5.47}, F125^{3.33}, and W269^{6.48}. To cross the activation barrier, S1P has to induce rotation of W269^{6.48} and F210^{5.47}, providing additional space for F265^{6.44} to reorient to the other side of TM6.

These changes destabilize the interactions between TM3 and TM6, resulting in the outward movement of TM6. Accordingly, the middle region of S1PR1 experienced a series of changes. In the inactive state, D91^{2.50} on TM2 has a stronger interaction with S131^{3.39} on TM3 than N307^{7.49} on TM7, yet the movement of F265^{6.44} creates room for N307^{7.49} to move toward the receptor center and form tighter contact with D91^{2.50}, promoting the inward movement of TM7.

After crossing the barrier, the hydrophobic tail of S1P swings toward TM3 in the IS, causing F210^{5.47} to rotate substantially

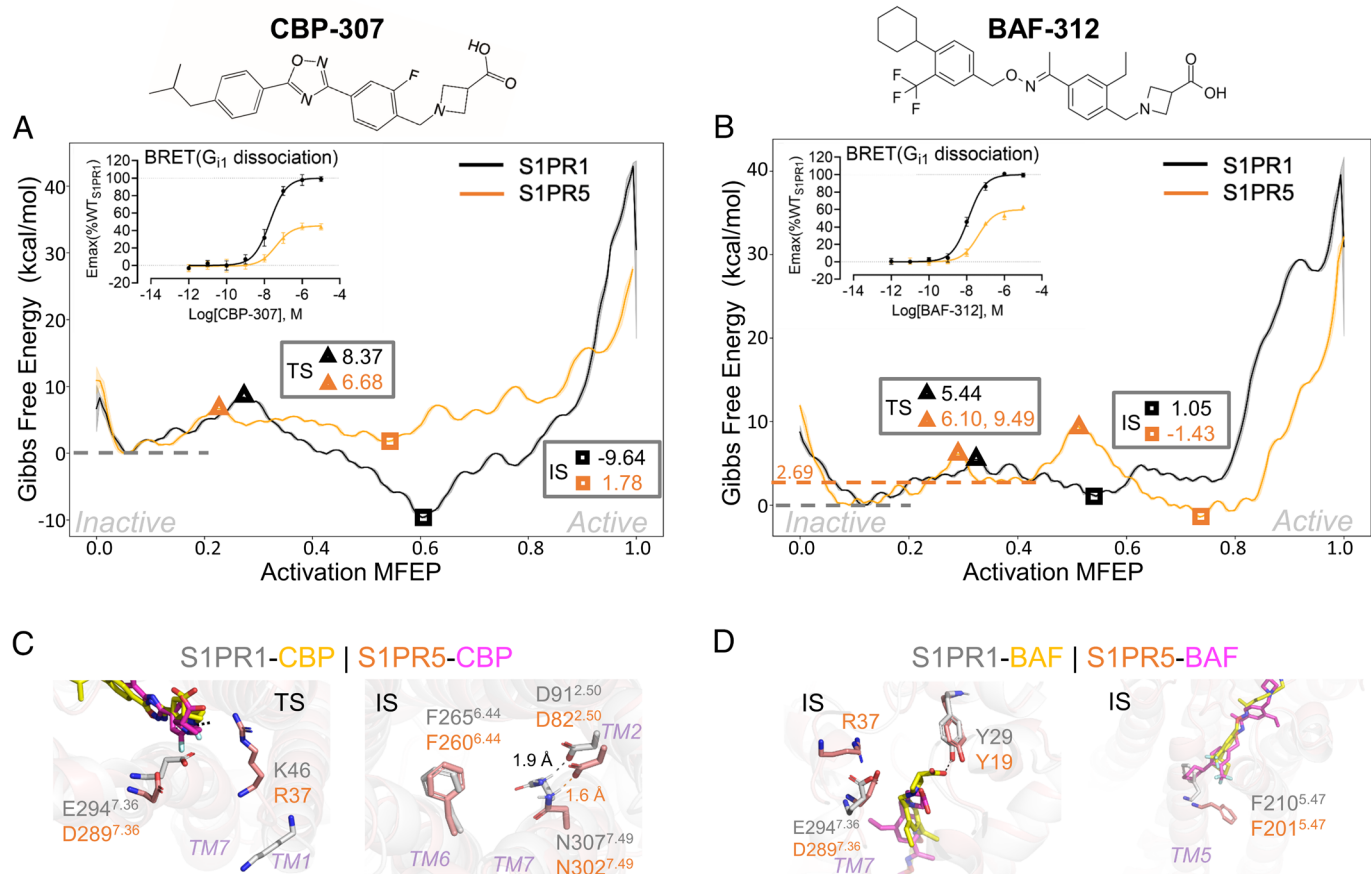


Fig. 5. Activation mechanism of S1PR1/5 by CBP-307 and BAF-312. Free energy distributions along the activation paths are plotted with the transparent bands denoting the errors: (A) CBP-307, S1PR1/5; (B) BAF-312, S1PR1/5; BRET measurements of G₁₁ dissociation are given as the insets; structures of the transition and IS are presented for (C) CBP-307, S1PR1/5; and (D) BAF-312, S1PR1/5.

toward TM6. Meanwhile, F265^{6.44} rotates further toward TM5, stabilizing the outward movement of TM6 with a new polar contact formed between N307^{7.49} and S131^{3.39}. Together with D91^{2.50}, this polar interaction stabilizes the inward movement of TM7 and finally caused Y311^{7.53} in the TM7-NPxxY region near the G protein binding site to coordinate with Y221^{5.58} on TM5, further disconnecting the interaction between TM6 and TM3.

Since all residues mentioned above are conserved in the S1PR family, they represent the common activation mechanism of all subtypes. The major difference between S1PR1 and S1PR3/5 is that, in the IS of S1PR3/5, no stable polar contacts are formed between N294^{7.49}/N302^{7.49}, the serine on TM3 (S131^{3.39}), Y298^{7.53}/Y306^{7.53}, or Y215^{5.58}/Y212^{5.58} (Fig. 3 D and E), although these residues all arrive at similar positions as in S1PR1. This explains why the S1PR3/5 intermediates are less stable than that of S1PR1. Therefore, we conclude that S1P activates S1PR1/3/5 by the same molecular mechanism, but with lower efficiency in S1PR3/5 due to the lack of a series of polar contacts in their intermediates.

Origin of CBP/BAF's Preference on S1PR1 over S1PR3. CBP-307 is a unique, orally available S1PR agonist currently in phase II clinical trials targeting ulcerative colitis and Crohn's disease (5). BAF-312 (Siponimod) is a specific S1PR agonist approved in 2019 by the US Food and Drug Administration (FDA) to replace Fingolimod for treating multiple sclerosis (41) via reducing the efflux of lymphocytes from lymph nodes and therefore suppressing immune response. Being also amphiphilic, CBP-307 and BAF-312 both hold a polar hydroxyl group as head and a ring-like

hydrophobic portion, except that CBP-307 has more rings in its middle while BAF-312 is ring-rich at its tail (Fig. 4 A and B).

CBP-307 has been reported to activate S1PR1/4/5 only (Fig. 1 C) with significantly lower potency on S1PR5 than S1PR1 (5). BAF312 was approved because of its selective activation on S1PR1/5 and its 1,000-fold higher potency for S1PR1 than S1PR2/3/4 (7). Such preferences of both compounds of S1PR1 over S1PR3 are well reproduced by our BRET assays (Inset of Fig. 4 A and B) and the free energy distributions along the activation paths (Fig. 4 A and B). CBP-307 not only has a lower barrier for S1PR1 (8.37 kcal/mol) than S1PR3 (11.13 kcal/mol) but also a significantly more stable intermediate (-9.64 kcal/mol) for S1PR1 than S1PR3 (5.56 kcal/mol). BAF-312 activates S1PR1 via a small barrier of 5.44 kcal/mol and an intermediate of stability 1.05 kcal/mol. The corresponding values for BAF-S1PR3 are both significantly higher (9.92 and 6.79 kcal/mol).

For both agonists, the two bulky S1PR3 residues F263^{6.55} on TM6 (L276^{6.55} of S1PR1) and I284^{7.39} on TM7 (L297^{7.39} of S1PR1) seem crucial in discouraging its activation. In S1PR1, both agonists are more flexible, with their hydrophobic tails able to insert into the pocket deeper than in S1PR3 (Fig. 4 C and D). This difference is mainly caused by the shrinking of space in S1PR3 at the presence of the large ring of F263^{6.55} and bulky I284^{7.39}. F263^{6.55} and I284^{7.39} forced the hydrophobic part of CBP-307 to detach from TM6 and its polar head to rearrange. Such difference is further transmitted to the central part and G-protein binding region of S1PR1 and S1PR3. In the central part, the side chain of Y98^{2.57} is able to rotate to form a stable polar contact with the main chain of S1PR1-TM7, stabilizing its inward shift (SI Appendix, Fig. S7A). No such contact is

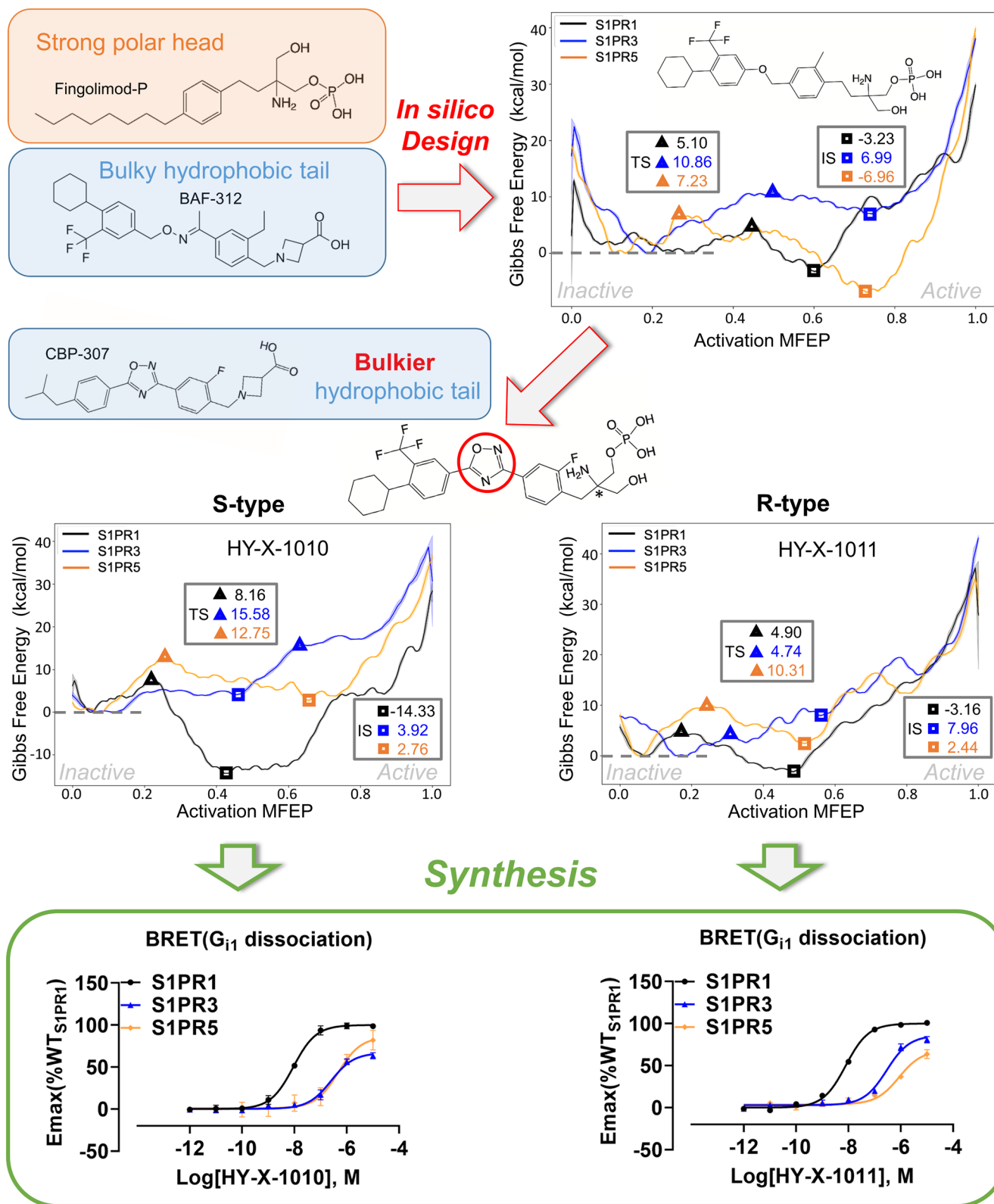


Fig. 6. Rational design of S1PR1-specific agonists. Free energy distributions along the activation paths are plotted with the transparent bands denoting the errors. The free energy distributions and BRET measurements of G_{i1} dissociation for S1PR1/3/5 are presented in black, blue, and orange, respectively. The chiral carbon of the second designed compound (Middle of figure) is labeled with an asterisk.

however observed in S1PR3. In the deeper central region, F265^{6.44} of S1PR1 experienced a notable rotation to allow an obvious inward shift of TM7. No such rotation can be observed for F252^{6.44} in S1PR3. Therefore, the presence of F263^{6.55} and I284^{7.39} appears to be the dominant factor that weakens the hydrophobic interaction

between CBP/BAF and TM6, the inward shift of TM7, and prevents the activation of S1PR3.

To verify this hypothesis, we designed two receptor mutants that swapped L276^{6.55}/L297^{7.39} and F263^{6.55}/I284^{7.39} between S1PR1 and S1PR3. For CBP-307, our calculations revealed a high barrier

ΔG_T of 8.36 kcal/mol and a positive ΔG_I of 4.65 kcal/mol for S1PR1-L276^{6,55}F-L297^{7,39}I, indicating low activation potency (Fig. 4E). By contrast, CBP-307 activates S1PR3-F263^{6,55}L-I284^{7,39}L via a small barrier ($\Delta G_T = 3.04$ kcal/mol) and a stable intermediate ($\Delta G_I = -4.31$ kcal/mol). Similarly, as shown in Fig. 4F, BAF-312 is able to activate S1PR3-F263^{6,55}L-I284^{7,39}L ($\Delta G_T = 3.44$ kcal/mol, $\Delta G_I = 0.7$ kcal/mol) more efficiently than S1PR1-L276^{6,55}F-L297^{7,39}I ($\Delta G_T = 6.70$ kcal/mol, $\Delta G_I = 1.85$ kcal/mol). As expected, BRET assays confirmed these predictions (*Inset* of Fig. 4 E and F and *SI Appendix, Table S1*).

Strong Polar Head May Discourage S1PR5 Activation. For S1PR5, both CBP-307 and BAF-312 exhibited lower activation potency than S1PR1. CBP-307 activates S1PR5 via similar barrier of 6.68 kcal/mol to that of S1PR1, yet with a less stable intermediate (1.78 kcal/mol) than S1PR1 (Fig. 5A). BAF-312 activates S1PR5 via two TS where the first has similar height (6.10 kcal/mol) with S1PR1 and the second being the rate-limiting step (9.49 kcal/mol). Interestingly, the intermediate for BAF-S1PR5 is slightly more stable (-1.43 kcal/mol) than S1PR1 (Fig. 5B). Overall, due to the more stable intermediate, BAF-312 seems to activate S1PR5 more efficiently than CBP-307, as confirmed by BRET assays (Fig. 5 A and B, *Inset* and *SI Appendix, Table S1*). Meanwhile, the activation potencies of BAF-312 for S1PR1/5 are comparable.

The major differences in the sequence of S1PR1/5 can be found near the polar head of the ligand, i.e., R37 in S1PR5 with stronger polarity and longer side chain than K46 in S1PR1 and D289^{7,36} in S1PR5 with a shorter side chain than E294^{7,36} in S1PR1. In the transition structures of CBP-S1PR1, K46 points away from the polar head of CBP, while E294^{7,36} seems coordinated with the N-H group on the polar head of CBP. In the opposite, in S1PR5, R37 points to the polar head with D289^{7,36} loses its contact with the N-H group. Accordingly, in the intermediate of CBP-S1PR1, an obvious inward movement of TM7 can be observed, resembling the structural features of the S1P-S1PR1 activation partly (Fig. 5C and *SI Appendix, Fig. S7B*). In CBP-S1PR5, the inward motion of TM7 seems prevented by the tighter polar interactions between N307^{7,49} and D91^{2,50} in the central region of the receptor than S1PR1 (Fig. 5C) and the loss of coordination between Y89^{2,57} and the TM7 backbone (*SI Appendix, Fig. S7B*). These results indicate that ligands with a stronger polar head may resemble more structural features of the S1P-S1PR1 activation while further discouraging the inward motion of TM7 and therefore activation of S1PR5.

For BAF-312, while the first transition state in S1PR5 looks similar to the one in S1PR1, in the second transition state, the ligand inserts deeper into the receptor than in S1PR5 (*SI Appendix, Fig. S8B*). Such insertion persists in the intermediate of BAF-S1PR5 possibly due to the loss of the contact between the polar head and Y19 on ECL1 observed in S1PR1 and induces the rotation of F201^{5,47} (F210^{5,47} in S1PR1) to form a stable packing with the surrounding bulky side chains of hydrophobic residues, stabilizing the intermediate structure (Fig. 5D). This suggests that introducing a stronger polar head into the ligand may encourage its coordination with R37/Y19, preventing its insertion into the receptor and the activation of S1PR5.

S1PR1-Specific Agonists. Based on the knowledge learned for S1P, CBP-307, and BAF-312, we first designed a unique compound scaffold with strongly polar phosphate-group similar to that of S1P and Fingolimod-P and a bulky ring-rich hydrophobic tail that resembles the one of BAF-312 (Fig. 6). The bulky tails as shown by CBP/BAF shall prevent the activation of

S1PR3, while the phosphate group may discourage the insertion of the ligand into S1PR5 and therefore its activation. Indeed, this compound is unable to activate S1PR3 ($\Delta G_T = 10.86$ kcal/mol, $\Delta G_I = 6.99$ kcal/mol). However, it appears to activate S1PR5 more efficiently ($\Delta G_T = 7.23$ kcal/mol, $\Delta G_I = -6.96$ kcal/mol) than S1PR1 ($\Delta G_T = 5.10$ kcal/mol, $\Delta G_I = -3.23$ kcal/mol). After visual inspection of the structures along its activation paths, we hypothesized that the apparent higher activation potency for S1PR5 might be due to the lack of an extra ring in the middle of the ligand, similar to that of CBP-307. This results in a second scaffold design with even bulkier ring-rich hydrophobic tails (*Middle* of Fig. 6).

Since this new scaffold has two chiral forms, S (HY-X-1010) and R (HY-X-1011), we calculated the free energy curves along the activation paths for both. Both forms activate S1PR1 significantly more efficiently than S1PR3/5. The S-type is featured by an 8.16 kcal/mol barrier and a very stable intermediate (-14.33 kcal/mol) for S1PR1, but higher barriers (15.58, 12.75 kcal/mol) and barely stable intermediates (3.92, 2.7 kcal/mol) for S1PR3/5. For the R-type, the activation potency of S1PR1 appears comparable to the S-type: though with a less stable intermediate ($\Delta G_I = -3.16$ kcal/mol), the activation barrier $\Delta G_T = 4.90$ kcal/mol is considerably lower than the S-type. For S1PR3/5, the R-type exhibits either a high barrier (S1PR5 10.31 kcal/mol) and an unstable intermediate (S1PR3 7.96 kcal/mol), indicating the ability for both forms to activate only S1PR1.

The predicted selectivity encouraged us to synthesize the two forms of the designed compound. BRET assays confirmed that, for both forms, the activation of S1PR1 occurs at tens of nanomolar concentration, around two orders of magnitude smaller than that for S1PR5.

Discussion

The present work demonstrated the long-expected promise of physics-driven computation in rational drug design. The ability of our approach to reveal the detailed activation mechanism of GPCRs by arbitrary agonists at affordable time-cost enables fine-tuning of the ligand scaffold for higher selectivity on S1PR1. Such mechanism-based modification also simplifies the design procedure to only a few iterations and helps minimizing wet-lab costs. As a general approach, our path-searching protocol can be readily applied in the agonist design for other GPCRs, e.g., S1PR3/5, or other target families where abundant inactive and active structures are available. In the present work, ligand modification has been made manually. Automatic generation of ligand scaffold through generative learning could be an important future direction for further automation of the design procedure.

Materials and Methods

In the present work, automated predictions of the activation potency of S1PRs by arbitrary ligands were achieved by integrating homology modeling, blind docking, all-atom MD simulations, TAPS, and umbrella sampling for calculating the free energy distribution along the activation path. BRET assays were applied for in vitro validation of the predictions. Details of the experiments, e.g., molecular cloning and cell culture and computational setup are described in *SI Appendix, Materials and Methods*.

Data, Materials, and Software Availability. TAPS method is wrapped within a python script publicly available at <https://github.com/liusong299/TAPS> (42). All other data are included in the manuscript and/or *SI Appendix*.

ACKNOWLEDGMENTS. We are grateful to Xiaogan 3D Scientific Computing Center for the computational support and Connect Biopharm for donating the agonist CBP-307. This work was supported by grant from the State Key Development Program of China (2022YFA0806700, R.R.); National Natural Science Foundation of China Projects 31971218 (R.R.) and 31971179 (L.Z.); Science, Technology, and Innovation Commission of Shenzhen Municipality Projects JCYJ-2020010915-0003938 (L.Z.), RCYX20200714114645019 (L.Z.), GXWD2020123115722002-20200831175432002 (L.Z.). L.Z. was also supported in part by a Presidential Fellowship, and R.T. and B.G. were supported by the

Ganghong Young Scholar Development Fund at the Chinese University of Hong Kong, Shenzhen.

Author affiliations: ^aWarshel Institute for Computational Biology, School of Medicine, The Chinese University of Hong Kong, Shenzhen 518172, China; ^bShanghai Key Laboratory of Metabolic Remodeling and Health, Institute of Metabolism and Integrative Biology, Fudan University, Shanghai 200437, China; ^cKobilka Institute of Innovative Drug Discovery, School of Medicine, The Chinese University of Hong Kong, Shenzhen 518172, China; ^dDepartment of Chemistry, University of Southern California, Los Angeles, CA 90089; and ^eShanghai Qi Zhi Institute, Shanghai 200030, China

1. T. K. Bjarnadottir *et al.*, Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* **88**, 263–273 (2006).
2. D. Hilger, M. Masureel, B. K. Kobilka, Structure and dynamics of GPCR signaling complexes. *Nat. Struct. Mol. Biol.* **25**, 4–12 (2018).
3. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schioth, D. E. Gloriam, Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug. Discov.* **16**, 829–842 (2017).
4. H. Rosen, R. C. Stevens, M. Hanson, E. Roberts, M. B. A. Oldstone, Sphingosine-1-phosphate and its receptors: Structure, signaling, and influence. *Annu. Rev. Biochem.* **82**, 637–662 (2013).
5. Anonymous, Potency and selectivity of CBP-307 on S1P receptors (DiscoveRx Corporation, Connect Biopharma CBP-307 investigator Brochure 5.0 ed, 2020).
6. V. Brinkmann *et al.*, The immune modulator FTY720 targets sphingosine 1-phosphate receptors. *J. Biol. Chem.* **277**, 21453–21457 (2002).
7. P. Gergely *et al.*, The selective sphingosine 1-phosphate receptor modulator BAF312 redirects lymphocyte distribution and has species-specific effects on heart rate. *Br. J. Pharmacol.* **167**, 1035–1047 (2012).
8. A. M. Subei, J. A. Cohen, Sphingosine 1-phosphate receptor modulators in multiple sclerosis. *CNS Drugs* **29**, 565–575 (2015).
9. S. J. Park, D. S. Im, Sphingosine 1-phosphate receptor modulators and drug discovery. *Biomol. Ther.* **25**, 80–90 (2017).
10. R. Ren, B. Pang, Y. Han, Y. Li, A glimpse of the structural biology of the metabolism of sphingosine-1-phosphate. *Contact* **4**, 2515256421995601 (2021).
11. R. O. Dror *et al.*, Activation mechanism of the beta(2)-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18684–18689 (2011).
12. N. R. Latorraca, A. J. Venkatakrishnan, R. O. Dror, GPCR dynamics: Structures in motion. *Chem. Rev.* **117**, 139–155 (2017).
13. S. Lu *et al.*, Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nat. Commun.* **12**, 4721 (2021).
14. Y. L. Miao, J. A. McCammon, G-protein coupled receptors: Advances in simulation and drug discovery. *Curr. Opin. Struct. Biol.* **41**, 83–89 (2016).
15. A. S. Powers *et al.*, Structural basis of efficacy-driven ligand selectivity at GPCRs. *Nat. Chem. Biol.* **19**, 805–814 (2023).
16. P. Xiao *et al.*, Ligand recognition and allosteric regulation of DRD1-Gs signaling complexes. *Cell* **184**, 943–956.e18 (2021).
17. L. Y. Yu *et al.*, Structural insights into sphingosine-1-phosphate receptor activation. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117716119 (2022).
18. R. Alhadeff, I. Vorobyov, H. W. Yoon, A. Warshel, Exploring the free-energy landscape of GPCR activation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 10327–10332 (2018).
19. A. Warshel, Multiscale modeling of biological functions: From enzymes to molecular machines (Nobel lecture). *Angew. Chem. Int. Ed. Engl.* **53**, 10020–10031 (2014).
20. C. Bai *et al.*, Exploring the activation process of the beta 2AR-G(s) complex. *J. Am. Chem. Soc.* **143**, 11044–11051 (2021).
21. D. Mondal, V. Kolev, A. Warshel, Exploring the activation pathway and G(i)-coupling specificity of the mu-opioid receptor. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 26218–26225 (2020).
22. P. G. Bolhuis, D. Chandler, C. Dellago, P. L. Geissler, Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
23. C. J. Chen, Y. Z. Huang, X. W. Jiang, Y. Xiao, A fast tomographic method for searching the minimum free energy path. *J. Chem. Phys.* **141**, 154109 (2014).
24. G. D. Leines, B. Ensing, Path finding on high-dimensional free energy landscapes. *Phys. Rev. Lett.* **109**, 020601 (2012).
25. L. Maragliano, A. Fischer, E. Vanden-Eijnden, G. Ciccotti, String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **125**, 24106 (2006).
26. A. C. Pan, D. Sezer, B. Roux, Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B* **112**, 3432–3440 (2008).
27. E. Vanden-Eijnden, Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **61**, 391–420 (2010).
28. L. T. Chong, A. S. Saglam, D. M. Zuckerman, Path-sampling strategies for simulating rare events in biomolecular systems. *Curr. Opin. Struct. Biol.* **43**, 88–94 (2017).
29. Q. T. Zhou *et al.*, Common activation mechanism of class A GPCRs. *Elife* **8**, e50279 (2019).
30. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
31. M. Baek *et al.*, Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
32. M. Moradi, G. Enkavi, E. Tajkhorshid, Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate: Phosphate antiporter. *Nat. Commun.* **6**, 8393 (2015).
33. L. Z. Zhu *et al.*, TAPS: A traveling-salesman based automated path searching method for functional conformational changes of biological macromolecules. *J. Chem. Phys.* **150**, 124105 (2019).
34. K. Xi *et al.*, Assessing the performance of traveling-salesman based automated path searching (TAPS) on complex biomolecular systems. *J. Chem. Theory Comput.* **17**, 5301–5311 (2021).
35. L. Y. Wang, K. Xi, L. Z. Zhu, L. T. Da, DNA deformation exerted by regulatory DNA-binding motifs in human alkyladenine DNA glycosylase promotes base flipping. *J. Chem. Inf. Model.* **62**, 3213–3226 (2022).
36. K. Xi, L. Z. Zhu, Automated path searching reveals the mechanism of hydrolysis enhancement by T4 lysozyme mutants. *Int. J. Mol. Sci.* **23**, 14628 (2022).
37. J. Schlitter, M. Engels, P. Kruger, Targeted molecular-dynamics—A new approach for searching pathways of conformational transitions. *J. Mol. Graphics* **12**, 84–89 (1994).
38. D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).
39. G. M. Torrie, J. P. Valleau, Non-physical sampling distributions in monte-carlo free-energy estimation—Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
40. R. H. Olsen *et al.*, TRUPATH, an open-source biosensor platform for interrogating the GPCR transducerome. *Nat. Chem. Biol.* **16**, 841–849 (2020).
41. L. J. Scott, Siponimod: A review in secondary progressive multiple sclerosis. *CNS Drugs* **34**, 1191–1200 (2020).
42. L. Zhu, S. Liu, K. Xi, Traveling-salesman based Automatic Path-Searching (TAPS). Github. <https://github.com/liusong299/TAPS>. Accessed 22 January 2024.