A $^{19}$F-qNMR-Guided Mathematical Model for G Protein-Coupled Receptor Signaling

Jesús Giraldo$^{1,2,3,\#,*}$, Jesper J. Madsen$^{4,5,6,\#}$, Xudong Wang$^7$, Lei Wang$^8$, Cheng Zhang$^8$, Libin Ye$^{7,9,\ast}$

$^1$Laboratory of Molecular Neuropharmacology and Bioinformatics, Unitat de Bioestadística and Institut de Neurociències, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain
$^2$Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, Spain
$^3$Unitat de Neurociència Trasllacional, Parc Taulí Hospital Universitari, Institut d'Investigació i Innovació Parc Taulí (I3PT), Institut de Neurociències, Universitat Autònoma de Barcelona, Spain
$^4$Global and Planetary Health, College of Public Health, University of South Florida, Tampa, FL 33612, USA
$^5$Center for Global Health and Infectious Diseases Research, College of Public Health, University of South Florida, Tampa, FL 33612, USA
$^6$Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
$^7$Department of Molecular Biosciences, University of South Florida, Tampa, FL 33620, USA
$^8$Department of Pharmacology and Chemical Biology, University of Pittsburgh School of Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA
$^9$H. Lee Moffitt Cancer Center & Research Institute, 12902 USF Magnolia Drive, Tampa, FL 33612, USA

$\#$Equal contribution
Abbreviations:

$^{19}$F-qNMR  $^{19}$Florine quantitative nuclear magnetic resonance
A$_{2A}$R  adenosine A$_{2A}$ receptor
β$_{2A}$R  β2 adrenergic receptor
BTFMA  2-bromo-N-(4-((trifluoromethyl)phenyl)acetamide
CHS  cholesteryl hemisuccinate
cAMP  cyclic adenosine monophosphate
DDM  n-dodecyl β-D-maltoside
DEER  double electron-electron resonance
DTT  dithiothreitol
EDTA  ethylenediaminetetraacetic acid
GDP  guanosine diphosphate
GPCR  G protein-coupled receptor
GRKs  G protein-coupled receptor kinases
GTP  guanosine triphosphate
HEPES  4-(2-hydroxyethyl)-1-piperazine ethanesulfonic acid
HEK293  human embryonic kidney 293 cell line
LUF  A$_{2A}$R partial agonist LUF5834
MNG-3  lauryl maltose neopentyl glycol
NECA  A$_{2A}$R full agonist 5'-N-Ethylcarboxamidoadenosine
S/N  signal/noise ratio
smFRET  single-molecule fluorescence resonance energy transfer
sf9  Spodoptera frugiperta
TCEP  tris(2-carboxyethyl)phosphine
Abstract
GPCRs exhibit a wide range of pharmacological efficacies, yet the molecular mechanisms responsible for the differential efficacies in response to various ligands remain poorly understood. This lack of understanding has hindered the development of a solid foundation for establishing a mathematical model for signaling efficacy. However, recent progress has been made in delineating and quantifying receptor conformational states and associating function with these conformations. This progress has allowed us to construct a mathematical model for GPCR signaling efficacy that goes beyond the traditional ON/OFF binary switch model. In this study, we present a quantitative conformation-based mathematical model for GPCR signaling efficacy using the adenosine A\textsubscript{2A} receptor (A\textsubscript{2A}R) as a model system, under the guide of \textsuperscript{19}F quantitative NMR (\textsuperscript{19}F-qNMR) experiments. This model encompasses two signaling states, a fully activated state, and a partially activated state, defined as being able to regulate the cognate G\alpha\textsubscript{S} nucleotide exchange with respective G protein recognition capacity. By quantifying the population distribution of each state, we can now in turn examine GPCR signaling efficacy. This advance provides a foundation for assessing GPCR signaling efficacy using a conformation-based mathematical model in response to ligand binding.
Significance statement

Mathematical models to describe signaling efficacy of G protein-coupled receptors (GPCRs) mostly suffer from considering only two states (ON/OFF). However, research indicates that a GPCR possesses multiple active-(like) states that can interact with $\mathrm{G} \alpha \beta \gamma$ independently, regulating varied nucleotide exchanges. With the guide of $^{19}$F-qNMR, the transitions among these states are quantified as a function of ligand and $\mathrm{G} \alpha \beta \gamma$, serving as a foundation for a novel conformation-based mathematical signaling model.
Introduction

Cell signaling is commonly initiated by a ubiquitous class of seven-transmembrane receptors, referred to as G protein-coupled receptors (GPCRs). Typically, when an agonist binds to a GPCR, it triggers the receptor to engage its cognate G-proteins, leading to activation of the α-subunit by replacing GDP with GTP, which then in turn leads to the release of the activated α-subunit (Weis and Kobilka, 2018). Despite advances in methods of structural biology and the release of over 250 structures of more than 60 GPCRs (Liang et al., 2018a; Liang et al., 2018b; Liang et al., 2017; Madsen et al., 2022; Rasmussen et al., 2011; Zhang et al., 2017; Zhao et al., 2019) that have been obtained using X-ray and cryo-EM methods (Carpenter and Tate, 2017; Lebon et al., 2011; Liu et al., 2012; Nehme et al., 2017; Rosenbaum et al., 2007; Wan et al., 2018), these methods have limitations in capturing transient or intermediate states and protein dynamics and interactions. Consequently, these structural snapshots may not fully reveal the conformational transition process and associated signaling mechanisms. As a result, a mathematical model has been developed here to determine the pharmacological efficacy of GPCRs based on their conformational responses to ligands. The model uses the adenosine A$_{2A}$ receptor (A$_{2A}$R) as a prototypical receptor system.

The submaximal signal output resulting from the partial agonist LUF5834 (denoted as LUF in this paper), in reference to the full agonism signaling induced by full agonist NECA, was in the range of 50-60% in some study (Guo et al., 2012). This was consistent with our cAMP-based measurement shown in Fig. 1G. Although multistate models for class A GPCRs based on NMR observations have been proposed (Katritch et al., 2014; Susac et al., 2018; Ye et al., 2016), these studies do not clarify how activation intermediates, such as partially activated state, contribute to the observed signaling efficacy. Using site-directed mutagenesis and $^{19}$F-labeling, a three-state quantitative model has been established for A$_{2A}$R. This model suggests that the receptor adopts at least two distinct activation states, which we refer to as the partially active intermediate state as (P) and the fully active state as (F), based on their
respective roles in the signaling process. We remark that these states have previously been defined as S3 and S3’ (Ye et al., 2018; Ye et al., 2016), and renamed in this manuscript in order to correspond to their functions in the receptor activation. The subpopulations of these states are preferentially sampled by partial (LUF) or full agonists (NECA) (Ye et al., 2018; Ye et al., 2016). Based on the quantitative response of these two signaling states to ligands, we have established, for the first time, a three-state mathematical model for the A2A signaling output that is based on the quantitative conformational equilibrium with the guide of 19F-qNMR. This model provides a foundation for developing a more sophisticated mathematical model in the future as the conformational resolution of receptor states continuously get improved and the function of each conformational state is being elucidated.
Materials and Methods

Plasmid Construction and Transformants Screening
The C-terminal peptides of the $\alpha$ were purchased from ChinaPeptide. The construct A$_{2A}R_{V229C}$ for $^{19}$F NMR experiments was performed as previously described (Ye et al., 2018; Ye et al., 2016). The mutant gene was sequenced and verified by the DNA sequencing facility (The Center for Applied Genomics, Sick Kids Hospital, Toronto, Canada) with the AOX1 primer pair of PF$_{AOX1}$ and PR$_{AOX1}$ prior to expression transformant screening. Freshly prepared competent cells of a strain of Pichia pastoris SMD 1163 (Δhis4 Δpep4 Δprb1, Invitrogen) were electro-transformed with individually Pmel-HF (New England Biolabs) linearized plasmid using a Gene Pulse II (Bio-Rad). Clone selection was performed as previously described by an in-house two-step transformant screening strategy (Draper-Joyce et al., 2018) and high-yield constructs were utilized for subsequent expression according to the specific requirements of NMR spectroscopy. (Scorer et al., 1994)

Receptor Expression, Purification and $^{19}$F Labelling
A single yeast colony on YPD plates was used to inoculate 4 mL of YPD medium and was then grown for 12 h. Subsequently, the culture was used to inoculate 200 mL of BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) YNB (yeast nitrogen base) without amino acids, 0.00004% (w/v) biotin, 1% (w/v) glycerol, 0.1 M PB (phosphate buffer) at pH 6.5) and grown for an additional 24 h. Cells were spun down and resuspended in 1 L of BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) YNB without amino acids, 0.00004% (w/v) biotin, 0.5% (w/v) methanol, 0.1 M phosphate buffer at pH 6.5, 0.04% (w/v) histidine and 3% (v/v) DMSO, 10 mM theophylline) at 20 °C. Methanol (0.5% (v/v)) was added every 18 h. Sixty hours after induction, yeast cells were subsequently harvested.

Cell pellets were collected by centrifugation and washed with 50 mM HEPES, 10% glycerol, pH 7.4 before addition of breaking buffer (50 mM HEPES, pH 7.4, 100 mM NaCl, 2 mM EDTA, 10% glycerol, 100 μM theophylline). The homogenized cell pellets
were applied to LM20 Microfluidizer for disrupting the yeast cell walls and releasing the membrane fraction parts. It is worthy to note that the temperatures were consistently kept at 4 °C by siting the sample in an ice-water bath and the disruption pressure was set to 15,000 psi. The passages of disruption were conducted in order to maximize the membrane release and as such increase the receptor production. It was usually executed for three times. Undisrupted yeast cells and cell debris were separated from the membrane suspension by centrifugation (8,000 g) for 30 min. The supernatant was collected and centrifuged at 100,000g for 1 h, and the precipitated cell membrane was then immediately dissolved in 50 mM HEPES, pH 7.4, 100 mM NaCl, 1% MNG-3 (lauryl maltose neopentyl glycol) and 0.02% CHS (cholesteryl hemisuccinate), 100 μM theophylline, and 20 mM imidazole) under continuous agitation for 1–2 h at 4°C until the solution was transparent. Subsequently, Talon resin (Clontech) was added to the solubilized membranes and incubated for 2 h. The A2AR-bound Talon resin was washed with 50 mM HEPES buffer, pH 7.4, containing 100 mM NaCl, 0.1% MNG-3, and 0.02% CHS and resuspended in the same buffer, followed by addition of 100 μM TCEP reducing agent and incubation for 20 min. TCEP was washed out immediately with a buffer consisting of 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.1% MNG-3, and 0.02% CHS. The A2AR-bound Talon resin was then resuspended in 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.1% MNG-3, and 0.02% CHS, and combined with the BTFMA(Ye et al., 2015). While A2AR was bound to a Talon (metal affinity) resin via the C-terminal poly HIS tag, the N-terminal FLAG tag was removed via Tobacco etch virus (TEV) protease. After its removal, the A2AR-bound Talon resin was extensively washed, using a disposable column, with 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.1% MNG-3, and 0.02% CHS. Apo A2AR was then eluted from the Talon resin with 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.1% MNG-3, 0.02% CHS, and 250 mM imidazole. Sodium chloride and imidazole in the sample were removed by dialysis against 50 mM HEPES, pH 7.4, 0.1% MNG-3, 0.02% CHS for 3 h. The XAC-agarose gel (antagonist xanthine amine congener (XAC) conjugated to Affi-Gel 10 resin) and A2AR were subsequently incubated together for 2 h with gentle nutation. Functional A2AR was eluted with 50 mM HEPES, pH 7.4, 0.1% MNG-3, 0.02% CHS, 100 mM NaCl, 20 mM theophylline. Talon resin was added to the eluted sample and incubated for another 2 h to bind functional A2AR.
Conjugated A$_{2A}$R was washed extensively with 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.1% MNG-3, 0.02% CHS, and 20 mM imidazole, to remove all theophylline. The functional apo A$_{2A}$R was eluted with 50 mM HEPES, pH 7.4, 100 mM NaCl, 0.1% MNG-3, 0.02% CHS, and 250 mM imidazole, and the sample was dialyzed to remove imidazole for NMR experiments.

**Expression and Purification of G Proteins**

Bovine G$_{α_s}$ short with an N-terminal 6xHistidine tag followed by a thrombin protease site was cloned into the pQE60 vector. Promoter and construct were sequenced and verified by Eton Bioscience (San Diego, CA, USA). *E. coli* WK6 cells were transformed with the vector, from a plate of LB and 50 μg/mL carbenicillin and a single colony was used to inoculate 8 mL of LB with 50 μg/mL ampicillin and grown shaking at 220 rpm in a 37 °C incubator for 8 hours. 1 mL of the 8 mL seed culture was used to inoculate a 100 mL LB with 50 μg/mL ampicillin culture that was allowed to grow shaking at 220 rpm in a 37 °C incubator 15 hours. 10 mL of this 100 mL culture was used to inoculate 1 L of TB with 50 μg/mL ampicillin. 1 L cultures were grown shaking at 200 rpm in a 37 °C incubator and sampled for OD$_{600}$ every 30 mins after 3 hrs. Once OD$_{600}$ 0.6 was reached temperature was lowered to 20°C and IPTG was added to a final concentration of 30 μM, additionally chloramphenicol was added to 1 μg/mL. Growth continued for 16 hrs and was arrested by centrifugation at 4500 g for 15 mins.

Cell pellets were resuspended in 15 mL/lysis buffer (50 mM Tris, pH 8.0, 100 mM NaCl, 5 mM MgCl$_2$, 5 mM imidazole, 10 μM GDP, 20 μM β-mercaptoethanol, 35 μg/mL phenylmethanesulphonyl fluoride, 32μg/ml tosyl phenylalanyl chloromethyl ketone, 32 μg/ml tosyl lysyl chloromethyl ketone 3.2 μg/ml leupeptin and 3.2 μg/ml soybean trypsin inhibitor). Cells were broken by sonication and lysate was clarified by centrifugation at 4 °C for 40 mins at 125,000 g. Supernatant was applied to a nickel column by gravity and washed with 10 column volumes wash 1 buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 5 mM MgCl$_2$, 5 mM imidazole) then 10 column volumes of wash 2 buffer (50 mM Tris, pH 8.0, 100 mM NaCl, 5 mM MgCl$_2$, 10 mM imidazole) and finally 10 column volumes of wash 3 buffer (50 mM Tris, pH 8.0, 1 mM MgCl$_2$, 10 mM imidazole). Protein was eluted
with 15 column volumes of elution buffer (50 mM Tris, pH 8.0, 250 mM imidazole, 1 mM MgCl2) upon elution GDP was added to 10 μM and dithiothreitol was added to 2 mM. Elution was diluted 3x with Q buffer A (50 mM Tris, pH 8.0, 1 mM MgCl2, 2 mM dithiothreitol) and applied to a Q sepharose column. Nearly pure Gαs eluted from the Q sepharose column in ~20% Q buffer B (50 mM Tris, pH 8.0, 1M NaCl, 1 mM MgCl2, 2 mM dithiothreitol) after a 20 column volume wash with 95% buffer A and 5% buffer B and 20 column volume gradient from 5-30% buffer B, GDP to 10 μM final concentration was added to fractions containing Gαs. The peak containing Gαs was pooled and concentrated with a 10-kDa molecular weight cut off concentrator (Amicon) to ~2 mL. Concentrated Gαs was polished 200 μL at a time over a superdex 75 column in running buffer containing 50 mM Hepes, pH 7.4, 2 mM DTT, 100 mM NaCl, 1 mM MgCl2 and 10 μM GDP. The fractions containing pure, as visualized by Coomassie stained gel, Gαs were again pooled and concentrated in a fresh 10-kDa molecular weight cut off concentrator (Amicon), aliquoted and flash cooled in liquid N2 for storage at -80 °C.

Gβ and Gγ were produced as a heterodimer in Spodoptera frugiperda (sf9) cells grown in EX-cell 420 serum-free medium (Sigma). Cultures were grown to a density of 2.0 million cells per ml then infected with two separate Autographa californica nuclear polyhedrosis viruses, one containing the gene for Gβ1 with an N-terminal 6xHistidine tag and one containing a gene for Gγ2 at a multiplicity of infection of 1:1. After 48 hours of shaking at 27 °C at 135 rpm the cells were harvested by centrifugation at 1000 g for 10 min. Cells were resuspended in lysis buffer (50mM HEPES, 8.0, 65 mM NaCl, 1 mM EDTA, 5mM β-mercaptoethanol, 35 μg/mL phenylmethanesulphonyl fluoride, 32 μg/mL tosyl phenylalanyl chloromethyl ketone, 32 μg/mL tosyl lysyl chloromethyl ketone 3.2 μg/mL leupeptin and 3.2 μg/mL soybean trypsin inhibitor) placed in a precooled nitrogen cavitation bomb (Parr Instrument Company). The nitrogen cavitation bomb was pressurized with 650 p.s.i. N2 for 30 min at 4 °C before being rapidly depressurized and all liquid collected. The lysate was subjected to a 1000 g centrifugation at 4 °C for 10 min to remove unlysed cells and organelles. The supernatant from the 1000 g centrifugation was then centrifuged at 125,000 g for 35 min at 4 °C and the supernatant was discarded. The pellet was resuspended in wash buffer (50 mM HEPES, pH 8.0, 50
mM NaCl, 5 mM MgCl₂, 5 mM β-mercaptoethanol, 35 μg/mL phenylmethanesulphonyl fluoride, 32 µg/mL tosyl phenylalanyl chloromethyl ketone, 32 µg/mL tosyllysyl chloromethyl ketone 3.2 µg/mL leupeptin and 3.2 μg/mL soybean trypsin inhibitor) and centrifuged again at 125,000 g for 35 min at 4°C. The pellet was again resuspended in wash buffer, homogenized with 30 strokes of a tight-fitting glass on glass Dounce homogenizer and diluted to ~5 mg/mL total protein prior to flash cooling in liquid N₂. The frozen membranes were thawed and diluted to final concentrations 20 mM HEPES, pH 8.0, 200 mM NaCl, 5 mM MgCl₂, 10 mM imidazole, 3% sodium cholate detergent, 5 mM β-mercaptoethanol, 35 µg/mL phenylmethanesulphonyl fluoride, 32 µg/mL tosyl phenylalanyl chloromethyl ketone, 32 µg/mL tosyllysyl chloromethyl ketone 3.2 µg/mL leupeptin and 3.2 μg/mL soybean trypsin inhibitor. The membranes were allowed to solubilize while gently stirring at 4°C for 1 hour prior to a 125,000 g 40 min at 4°C centrifugation. Supernatant was kept and detergent was exchanged from cholate to DDM by slow addition until five-fold diluted with n-dodecyl β-D-maltoside (DDM) buffer (20 mM HEPES, pH 8.0, 200 mM NaCl, 5 mM MgCl₂, 10 mM imidazole, 0.05% w/v DDM, 5 mM β-mercaptoethanol). This DDM solubilized solution was then loaded onto a nickel column by gravity that was subsequently washed with 20 column volumes of wash 1 buffer (20 mM HEPES, pH 8.0, 300 mM NaCl, 5 mM MgCl₂, 10 mM imidazole, 30 μM AlCl₃, 5 mM NaF, 10 μM GDP, 5 mM β-mercaptoethanol, 0.05% DDM) then 20 column volumes of were 2 (same as wash 1 but with 50 mM NaCl). Protein was eluted from the nickel column with 12 column volumes of elution buffer (20 mM HEPES, pH 8.0, 40 mM NaCl, 1 mM MgCl₂, 250 mM imidazole, 30 μM AlCl₃, 5 mM NaF, 10 μM GDP, 5 mM β-mercaptoethanol, 0.05% DDM) and loaded onto a MonoQ column. After loading the column was washed with 20 column volumes of Q buffer A (20 mM HEPES, pH 8.0, 40 mM NaCl, 100 μM MgCl₂, 30 μM AlCl₃, 5 mM NaF, 10 μM GDP, 5 mM β-mercaptoethanol, 0.05% DDM) then eluted into fractions with a linear gradient from 0-100% buffer B (same as buffer A except 1 M NaCl) over 50 column volumes. Gβγ heterodimer elutes around 12% buffer B as a pure protein according to a Coomassie stained gel. Fractions containing Gβγ were pooled and concentrated with a 10-kDa molecular weight cut off concentrator (Amicon) to 11 mg/mL.
**19F NMR Experiments**

NMR samples consisted of 250 µL volumes of 20-100 µM A2AR_V229C in 50 mM HEPES buffer and 100 mM NaCl, doped with 10% D2O as well as 10 µL bendroflumethazide as an internal reference. All 19F NMR experiments were performed on a 600 MHz Varian Inova spectrometer equipped with a cryogenic triple resonance probe with the high frequency channel tunable to 19F, at temperature of 20 °C. The typical 1D 19F NMR experimental setup included a 12-15 µs 90° excitation pulse, an acquisition time of 200 ms, a spectral width of 15 kHz, and a repetition time of 1 s. Most spectra were acquired with 5000--15,000 scans depending on sample concentrations, which provided a signal to noise ratio of roughly 50-100. The deconvolution of 19F NMR spectra were performed using the software MestReNova 12.0.3, in which the chemical shifts for states I1-2, P and F of mutant V229C were set -61.05±0.05, 61.64±0.05, -61.85 ±0.05 ppm with reference to -59.05 ppm while there was no restrains for linewidth. The spectrum processing typically involved zero filling, and exponential apodization equivalent to 15 Hz line broadening.

**Mathematical Modelling**

A rectangular hyperbola function \( \frac{E}{E_m} = \frac{S}{(S+K_E)} \) was proposed for the relationship between fractional effect \( \frac{E}{E_m} \) and stimulus \( S \), with \( E_m \) being the maximum possible effect of the system and \( K_E \) the transducer factor of stimulus into response. The stimulus results from the sum of the stimuli provided by each of the active receptor species as the product of its concentration times the intrinsic efficacy. Three receptor states were considered: inactive (I), partially active (P) and fully active (F), to which the ligand may bind, and whose populations for free and ligand-bound receptors are regulated by well-defined equilibrium constants. Basal response is included in the model because partially and fully active states of the free receptor have intrinsic efficacy. It is considered that ligand concentration largely exceeds receptor concentration and, then, the concentration of free ligand is approximately equal to the concentration of ligand added to the system.
Structural Modeling and Conformational State Prediction

AlphaFold2 is used to predict and explore structural models of A2AR conformational states. We use the publicly available Jupyter notebooks to access ColabFold/AlphaFold2 (Jumper et al., 2021; Mirdita et al., 2022; Steinegger and Soding, 2017). For proteins that can adopt multiple conformations, modeling with AlphaFold2 can present challenges (Heo and Feig, 2022). To handle these, we sample models of apo-A2AR broadly from the inactive state to the active state by seeding active (PDB ID: 5G53, 6GDG, 5G53+6GDG), intermediate (PDB ID: 7ARO), or inactive (PDB ID: 5UIG) templates to AlphaFold2 for its prediction. We found by manual inspection (of especially the TM6 orientation) that seeding 6GDG (for the active conformation) and 7ARO (for the intermediate conformation) yielded good results. Relaxation of the structures by gradient descent in the Amber99sb force field (Hornak et al., 2006) was done with OpenMM v.7.3.1 (Eastman et al., 2017). Structural visualization was performed in PyMOL (The PyMOL Molecular Graphics System, Version 2.5.0 Schrödinger, LLC). Protein-protein interaction between A2AR and the insertion helix from miniGs were characterized by PLIP (Adasme et al., 2021). The definitions of key interactions employed by PLIP are as follows (Salentin et al., 2015). Hydrogen bonds: Maximum distance between acceptor and donor (HBOND_DIST_MAX) = 4.1 Å, and minimum angle at the hydrogen bond donor (D-H…A) (HBOND_DON_ANGLE_MIN) = 100° (Hubbard and Haider, 2010). Hydrophobic contacts: Maximum distance of carbon atoms (HYDROPH_DIST_MAX) = 4.0 Å; Salt bridges: Maximum distance between two centers of charges (SALTBRIDGE_DIST_MAX) = 5.5 Å (Barlow and Thornton, 1983).

Radioligand Binding Assay

Competition binding studies for partial agonist LUF and full agonist NECA against hot ligand [3H]CGS21680 in the A2AR constructs were performed by incubating purified A2AR receptors with a certain amount of [3H]CGS21680 determined from saturation experiments that reached to the saturation level. A range of test agonists (LUF or NECA) from 10^{-2} to 10^{-8} M were then used to compete with [3H]CSG21680 for the orthosteric binding pocket. The mixture samples were conducted at room temperature for 2 hrs before applied to PEI pretreated Whatman GF/C filters onto the manifold. The sample
loaded Whatman GF/C filters were then washed three times to remove unbound hot and cold ligands. The GF/C filters were then carefully transferred into scintillation tubes and count per minute (CPM) numbers were recorded for each sample using LS 6500 Multi-Purpose Scintillation Counter. The competition curves were fit using Prism GraphPad 9.0 with log (agonist) vs. response in a nonlinear regress with variable slope and least squares fit.

**cAMP Measurements**

The cAMP measurements were performed in HEK293 cell lines. We use the cAMP-Glo™ assay kit to measure the cAMP levels. The full-length WT* construct was engineered from PRESTO-Tango-A2AR from Addgene. The plasmid was then transfected the HEK293 cells from ATCC, and the cAMP products as a function of different ligands (NECA and LUF) were measured.
Results

Quantification of Inactive, Partially Active, and Fully Active States Using $^{19}$F-qNMR

We used $^{19}$F-qNMR to investigate how the LUF-selected receptor conformation interacts with G proteins. Specifically, we examined the effects of the G$\alpha_s$B$\gamma$ heterotrimer with the motion of the transmembrane domain VI (TM6) on the site of V229C$^{6,31}$. The results, as shown in Figs.1C,1D and 1E, reveal that the addition of full agonist into the G$\alpha_s$B$\gamma$ saturated receptor only slightly increase the F fraction while reducing the I$_{1-2}$ and P states. On the other hand, the addition of partial agonist resulted in a significant shift in the conformational population towards the P conformer with decreased I$_{1-2}$ and F states, leading to a higher subpopulation of F fraction in the basal signaling (apo+G$\alpha_s$B$\gamma$ sample) than in the partial agonism signaling (LUF+A$_{2A}$R+G$\alpha_s$B$\gamma$ sample). Previous studies have shown that A$_{2A}$R signaling is almost four times as pronounced with partial agonist stimulation compared to basal levels (Lane et al., 2012), which was also indicated in the cAMP measurement (Fig.1G). However, the spectra in Fig. 1D and 1E suggest that the F population of LUF-saturated A$_{2A}$R is even much lower (18% vs 68%) than that of apo-A$_{2A}$R, indicating that the P state must be competent for G$\alpha_s$B$\gamma$ to achieve a higher signaling level for partial agonism signaling than for basal signaling (Fig.1G) since this is the only distinguishing feature of the partial agonist-saturated spectrum in this three state model system.

The results of the $^{19}$F NMR experiments suggest that the interactions between partial agonist- and full agonist-bound A$_{2A}$R with G$\alpha_s$B$\gamma$ are clearly distinguishable. However, the resolved structure is only available for the NECA-A$_{2A}$R-miniG$\alpha_s$B$\gamma$ complex. To investigate the differences between the interactions of full agonist-bound (F state) and partial agonist-bound (P state) A$_{2A}$R with the G proteins, we used AlphaFold2 to model the complex between the receptor and the C$\alpha$S peptide of a mini-Gs construct by seeding the LUF-bound structure (PDB ID: 7ARO) or the NECA-bound structure (PDB
ID:6GDG). The resulting intermediate conformation complex is clearly distinct from the fully active complex (Fig. 2A and 2B). Comparison of the specific receptor-peptide interactions reveals a contrast between the fully active and intermediate partially active cases (Figs. 2D and 2E). The number of hydrophobic interactions, hydrogen bonds, and salt bridges is less in the intermediate case (8 hydrophobic interactions, 6 hydrogen bonds, and 1 salt bridge in the partially active vs. 10 hydrophobic interactions, 9 hydrogen bonds, and 2 salt bridges in the fully active). A decreased number of interactions is consistent with a lower-affinity binding pattern upon partial agonist binding that likely reduces downstream signaling efficacy compared to that of the full agonist-bound complex. The amino acids at the interface that interact with the G protein binding cavity also differ significantly in these two scenarios (Figs. 2D and 2E). Additionally, the electrostatic surface of A2AR show differences when Cα5 engages with partial and full agonist bound receptors, as seen in Fig. 2C.

Considering the evidence presented in the above sections, our next step is to create a mathematical model that can quantify the signaling efficacy for the A2AR based on current three conformational states in response to partial or the full agonists.

Mathematical Model Formulation
A three-state receptor model designed to account for partial agonism is constructed as follows (Fig. 3). The model includes two cycles, I and II, depicting the induction/selection of partial (RPA, ARPA) and full (RF, ARF) active states of the receptor, respectively, from a common inactive state (RI, ARI). Thus, the model extends prior models in the literature by integrating two concepts: constitutive receptor activity (Slack and Hall, 2012) and multiple receptor states (Leff et al., 1997). Distinct receptor states can be associated to multiple signaling pathways or considered to embody differential intrinsic efficacies for a single signaling pathway (Kenakin, 2002). The latter assumption is herein considered because of the experimental NMR data in which the present model is based on. See the Supplemental Appendix for a detailed description of the equilibrium constants, pharmacological parameters and stimulus-effect transduction function.
Modeling Experimental Results

Revisiting the concentrations of the receptor species from Fig. 1

From the inactive, partially active and fully active relative populations of apo and LUF- and NECA-bound receptors depicted in Fig. 1, the equilibrium constants and binding parameters of the system can be obtained as follows (see Supplemental Appendix).

- Apo

\[ [R_I] = 10; [R_{PA}] = 22; [R_{FA}] = 68 \]

\[ K_1 = \frac{[R_{PA}]}{[R_I]} = \frac{22}{10} = 2.2 \]

\[ K_2 = \frac{[R_{FA}]}{[R_I]} = \frac{68}{10} = 6.8 \]

- LUF

\[ [R_I] + [AR_I] = 1; [R_{PA}] + [AR_{PA}] = 81; [R_{FA}] + [AR_{FA}] = 18 \]

We assume that the concentrations of free receptor species are negligible with respect to the ligand-bound receptor because the experiments were performed upon ligand saturation (Fig. 1).

\[ [AR_I] \approx 1; [AR_{PA}] \approx 81; [AR_{FA}] \approx 18 \]

\[ [AR_{PA}] = \alpha K_1 [AR_I]; 81 = \alpha 2.2 \times 1; \alpha = 81/2.2 = 36.8 \]

\[ [AR_{FA}] = \beta K_2 [AR_I]; 18 = \beta 68 \times 1; \beta = 18/68 = 2.6 \]

\[ \gamma = \beta/\alpha = 2.6/36.8 = 0.07 \]

\( \alpha = 36.8 \) measures the induction/selection of partially activated receptor conformation by LUF with respect to inactive one.

\( \beta = 2.6 \) measures the induction/selection of fully activated receptor conformation by LUF with respect to inactive one.

\( \gamma = 0.07 \) measures the induction/selection of fully activated receptor conformation by LUF with respect to partially activated one.

The \( \alpha, \beta, \gamma \) values reflect the order of concentrations of the LUF-bound receptor species:
Partially activated > Fully activated > Inactive

- **NECA**

\[ [R_i] + [AR_i] = 1; [R_{PA}] + [AR_{PA}] = 15; [R_{FA}] + [AR_{FA}] = 84 \]

In the same way as with LUF, we assume that the concentrations of free receptor species are negligible with respect to the ligand-bound receptor because the experiments were performed upon ligand saturation.

\[ [AR_i] \approx 1; [AR_{PA}] \approx 15; [AR_{FA}] \approx 84 \]

\[ AR_{PA} = \alpha K_{1}[AR_i]; 15 = \alpha 2.2 \times 1; \alpha = 15/2.2 = 6.8 \]
\[ AR_{FA} = \beta K_{2}[AR_i]; 84 = \beta 6.8 \times 1; \beta = 84/6.8 = 12.4 \]
\[ \gamma = \beta/\alpha = 12.4/6.8 = 1.8 \]

\( \alpha = 6.8 \) measures the induction/selection of partially activated receptor conformation by LUF with respect to inactive one.

\( \beta = 12.4 \) measures the induction/selection of fully activated receptor conformation by NECA with respect to inactive one.

\( \gamma = 1.8 \) measures the induction/selection of fully activated receptor conformation by NECA with respect to partially activated one.

The \( \alpha, \beta, \gamma \) values reflect the order of concentrations of the NECA-bound receptor species:

Fully activated > Partially activated > Inactive.

**Intrinsic Efficacies Based on the Model and Asymptotic Pharmacological Responses**

As previously mentioned, it has been shown that A2AR signaling is almost four times stronger when the receptor is stimulated by partial agonists compared to basal levels (Lane et al., 2012). However, based on the spectra in Fig. 1, the F state population of LUF-saturated A2AR is even lower than that of apo-A2AR. This suggests that the P activation intermediate must be capable of interacting with \( \text{G}_{\alpha}\beta\gamma \) in order to achieve a higher level of signaling for partial agonism than for basal agonism. This is the only
distinguishing feature of the partial agonist-saturated spectrum. In terms of the model, this assessment can be expressed as the assertion that the intrinsic efficacy of the LUF-bound partially activated receptor is greater than the intrinsic efficacy of the apo receptor: \( \varepsilon_{\text{ARPA}} > \varepsilon_{\text{RFA}} \). Because of the relationship between active conformations and G protein-bound receptor, the above expression for LUF can be rewritten as \( \varepsilon_{\text{ARPA}} \text{-GDP}_{\text{nc}} > \varepsilon_{\text{RFA}} \text{-GDP}_{\text{o}} \).

To illustrate the utility of the model, we can simulate basal and maximum responses by giving values to the parameters in the model. The simulation of functional responses through a mathematical model allows the biological system to be interrogated under different conditions and, although it does not correspond to a real experiment, it provides new knowledge about the system and the possible formulation of hypotheses.

Assuming \( \chi = 0.04 \), \( \varepsilon_{\text{RPA}} = 1 \) and \( \varepsilon_{\text{RFA}} = 5 \) (and recovering \( K_1 = 2.2 \) and \( K_2 = 6.8 \) from the analysis of Fig. 1), then (see Supplemental Appendix) fractional basal response is equal to

\[
\frac{E}{E_m} = \frac{\chi(\varepsilon_{\text{RPA}}K_1+\varepsilon_{\text{RFA}}K_2)}{1+K_1+K_2+\chi(\varepsilon_{\text{RPA}}K_1+\varepsilon_{\text{RFA}}K_2)} = \frac{0.04(2.2+5\times6.8)}{1+2.2+6.8+0.04(2.2+5\times6.8)} = 0.13
\]

If, in addition, we assume that for LUF \( \varepsilon_{\text{ARPA}} = 30 \) and \( \varepsilon_{\text{ARFA}} = 50 \) (and recovering that \( \alpha = 36.8 \) and \( \beta = 2.6 \) from the analysis of Fig. 1), then the maximum fractional response for LUF is equal to

\[
\lim_{[A] \to \infty} \frac{E}{E_m} = \frac{\chi(\varepsilon_{\text{ARPA}}\alpha K_1+\varepsilon_{\text{ARFA}}\beta K_2)}{1+\alpha K_1+\beta K_2+\chi(\varepsilon_{\text{ARPA}}\alpha K_1+\varepsilon_{\text{ARFA}}\beta K_2)} = \\
\frac{0.04(30\times36.8\times2.2+50\times2.6\times6.8)}{1+36.8\times2.2+2.6\times6.8+0.04(30\times36.8\times2.2+50\times2.6\times6.8)} = 0.57
\]

Moreover, if we assume that for NECA \( \varepsilon_{\text{ARPA}} = 40 \) and \( \varepsilon_{\text{ARFA}} = 1000 \) (and recovering that \( \alpha = 6.8 \) and \( \beta = 12.4 \) from the analysis of Fig. 1), then (see Supplemental Appendix) the maximum fractional response for NECA is equal to
\[
\lim_{[A]\to\infty} E_m = \frac{\chi(e_{ARPA}\alpha K_1 + e_{ARFA}\beta K_2)}{1 + \alpha K_1 + \beta K_2 + \chi(e_{ARPA}\alpha K_1 + e_{ARFA}\beta K_2)} = 0.04\frac{(40 \times 6.8 \times 2.2 + 1000 \times 12.4 \times 6.8)}{1 + 6.8 \times 2.2 + 12.4 \times 6.8 + 0.04(40 \times 6.8 \times 2.2 + 1000 \times 12.4 \times 6.8)} = 0.97
\]

Note the criteria used for the proposal of intrinsic efficacies: (i) the intrinsic efficacy of the fully active is greater than that of the partially active state for each of the receptor species (apo, LUF-bound and NECA-bound); (ii) the intrinsic efficacy of the NECA-bound fully active state is greater than that of the LUF-bound fully active state and the latter greater than that of the APO fully active state: \(\varepsilon_{NECARFA} > \varepsilon_{LUFRFA} > \varepsilon_{RFA}\); (iii) the intrinsic efficacy of the NECA-bound partially active state is greater than that of the LUF-bound partially active state and the latter greater than that of the APO partially active state: \(\varepsilon_{NECARPA} > \varepsilon_{LUFRPA} > \varepsilon_{RPA}\); (iv) the intrinsic efficacy of the LUF-bound partially active state is greater than that of the APO fully active state: \(\varepsilon_{LUFRPA} > \varepsilon_{RFA}\).

**Modelled Concentration-Effect Curves for LUF and NECA**

The graphs in Fig. 4 display the concentration-effect curves for LUF and NECA either from experimental settings (Fig. 4A) or simulated by utilizing the parameter values mentioned earlier (Fig. 4C), which are consistent with published experimental equilibrium dissociation constants (Navarro et al., 2020) and signaling efficacies (Lane et al., 2012). These curves illustrate the pharmacological profiles of the ligands, with LUF showing partial agonism in comparison to NECA, and a leftward shift due to its higher receptor affinity. It is worth noting that while the values for the equilibrium constants between receptor species were derived from the NMR data, the values for the equilibrium dissociation constant between the ligand and the inactive state of the receptor as well as the values for the parameters related to function were simulated. Because of the many parameters included in Equation 3 there can appear a problem of overfitting in the case we want to estimate the parameter values from experimental functional responses. This problem is inherent to operational models of agonism (Black and Leff, 1983) and can be addressed by including additional concentration-effect
curves with decreasing maximum responses for each of the ligands considered (Leff et al., 1990a). Concentration-effect curves with decreasing maximum responses can be obtained by including an irreversible antagonist at different concentrations. Inclusion of irreversible antagonists would have the effect of decreasing the value of χ (by decreasing available total receptor concentration) without affecting the rest of parameters thus helping parameter estimation (Leff et al., 1990b).
Discussion

Our $^{19}$F NMR spectroscopy findings are in agreement with other studies showing that GPCRs can adopt a diverse range of conformational states (Calebiro and Grimes, 2019; Gregorio et al., 2017; Manglik et al., 2015; Susac et al., 2018; Wingler et al., 2019; Ye et al., 2018; Ye et al., 2016). These states enable the recognition of ligands as well as facilitate binding to G proteins (or other transducer partners), ultimately leading to G protein (or other transducer partner) activation (Manglik et al., 2015; Nygaard et al., 2013; Ye et al., 2016). Ligands with different efficacies stabilize distinct receptor activation intermediates which, in turn, can interact with various partners such as G proteins, GRKs, and β-arrestins, resulting in a signal response with varying intensity and bias (Rosenbaum et al., 2009). These distinct conformations of the receptor have allosteric signaling capacity that spans from the orthosteric pocket to the nucleotide exchange sites. Studies involving a purified, fluorescently labelled β2 adrenergic receptor (β2AR) have demonstrated that partial and full agonists stabilize distinct conformational states by engaging specific micro-switches in the receptor (Ghanouni et al., 2001; Swaminath et al., 2005; Swaminath et al., 2004). Furthermore, and very recently, by using double electron-electron resonance (DEER) and single-molecule fluorescence resonance energy transfer (smFRET), the conformational ensemble of the μ-opioid receptor including Gi activation and β-arrestin-1 recruitment has been characterized and the intrinsic efficacies associated to ligand-specific receptor conformations dissected (Zhao et al., 2023). Quantification of these ensembles within a mathematical modeling framework (Kenakin, 2019; Roche et al., 2013) would allow simulation and testing of hypotheses under various conditions thus leading to a better understanding of biophysical and biochemical experiments. Indeed, research on the β2AR receptor has indicated that the use of negative and positive nanobodies to modulate the conformational equilibrium can lead to a 3-state activation model encompassing two inactive states along with a fully active state that can form ternary complex with Gαβγ (Staus et al., 2016).

It is well known that the binding of various ligands to the identical orthosteric pocket can result in a broad range of downstream signaling responses, including varied efficacy.
and pathway biases (Kahsai et al., 2011). A conformational equilibrium rebalancing and dynamics could be key driving forces that result in distinct signaling, a notion that remains largely unexplored. Therefore, a mathematical model that can rigorously connect the conformational profile upon ligand binding to signaling outputs is a valuable tool for exploration of the signaling mechanism. The mathematical model established in this study enables quantification of the signaling efficacy of the A$_{2A}$R receptor based on subpopulations of $^{19}$F NMR-delineated conformational states upon ligand binding. This model may also be applicable to other GPCRs as a general approach of $^{19}$F-qNMR in dissecting and quantifying conformational states of various receptors. We are actively working towards achieving this objective and anticipate that the model’s progress will align closely with advancements in technology (e.g., improved $^{19}$F-tags and GPCR expression systems). Ultimately, we anticipate ligand-based receptor conformational equilibrium changes can be established in an appropriate cell-based model. It is important to note that the current model is solely based on those conformational states that interact with G$\alpha$$_{s}$βγ, without considering additional G protein subtypes that co-occur in the signaling process. These signaling convergence and biases must be considered in future models. There is no doubt that continuous methodological innovation is required to establish more advanced models. Furthermore, it is worth noting as well that the dynamics of individual conformational states play a significant role in signaling and should also be included in a future model when it becomes possible to investigate the dynamics of individual sub-states. We expect that the initial model established here will set a conceptual cornerstone for future development of a high-resolution conformation-based mathematical model for signaling prediction that will take into account both conformational transitions and dynamics of the receptor upon ligand binding.

Acknowledgements
We thank Kun Yan and Griffin Parker in assisting with cell culture.

Data, Materials, and Software Availability
All study data are included in the main text.
Author contributions:
Participated in research design: Ye, Giraldo
Conducted experiments: Ye, Giraldo, Madsen, Wang(X)
Contributed new reagents: Wang (L), Zhang
Performed data analysis: Ye, Giraldo, Madsen, Wang(X)
Wrote or contributed to the writing of the manuscript: Ye, Giraldo, Madsen with input from all authors.

References


Footnotes:
L. Y. acknowledged the NIH support by the grant 1R01GM149659 and startup funds and internal proposal enhancement grant from University of South Florida. J. G. has received funding by the grant PID2020-119136RB-I00 funded by MCIN/AEI/10.13039/501100011033. No author has an actual or perceived conflict of interest with the contents of this article.

**Legends for Figures**

**Fig. 1.** $^{19}$F NMR spectra of V229C-A$_{2A}$R as a function of G$\alpha$$_{s}$$\beta$$\gamma$, partial agonist (LUF) and full agonist (NECA) and related signaling levels. (A) Topological location of $^{19}$F probe. (B) $^{19}$F NMR spectra of V229C-A$_{2A}$R as a function of partial agonist LUF and full agonist NECA; these three spectra were adapted from Libin Ye, et al., Nature, 2016. (C) $^{19}$F NMR spectra of V229C-A$_{2A}$R 5×G$\alpha$$_{s}$$\beta$$\gamma$, 50×LUF and 100×NECA. (D) Highlighted subpopulations of different conformational states upon saturations with both ligands and G$\alpha$$_{s}$$\beta$$\gamma$. (E) Histograms of population distributions for different states. Of note, the P and F states were distinguished by measuring T2 values of peaks as well as a function of ligands as shown in the figure and the reference (Libin Ye et al, Nature, 2016). Considering the effects of spectral coalescence and broadening from G protein binding, we set the chemical shift range for each conformer within ±0.05 ppm variation using the references published before stated in method section. Data with error bars are presented as state population±SD. The SD values were determined based on spectral S/Ns and fitting errors of the deconvolutions. (F) Competition radioligand binding assays of partial and full agonists against the hot ligand $[^3]$HCGS21680 and (G) cAMP measurements as a function of ligands concentrations. Of note, the maximal efficacy of NECA was normalized to 100. NECA curve was then used as a reference for LUF efficacy curve creation.

**Fig. 2.** Structural view of the interaction between A$_{2A}$R and the C$\alpha$5 of mini-G. (A and B) Interaction profiles in the fully active 6gdg-like complex (FA) and in the intermediate partially active complex (PA) differ markedly. (C) Comparison of the interaction between C$\alpha$5 and the intracellular receptor cavity with the receptor surface colored according to the vacuum electrostatic potential. (D and E) A closer view on the C$\alpha$5 with residues in direct contact with the receptor in sticks. Interaction types are depicted as: solid blue line (hydrogen bond), dashed gray line (hydrophobic), yellow dashed line (salt bridge).

**Fig. 3.** Cycles I and II depict the interconversion between inactive and partially active (Cycle I) and inactive and fully active (Cycle II) receptor species. Cycle III depicts the interconversion between partially and fully active receptor species. In each of the cycles, horizontal arrows correspond to the induction concept (a receptor state is induced by another one) whereas vertical arrows correspond to the selection concept (receptor states are selected by ligand binding). The constants included in Cycle III can be obtained from those included in cycles I and II. For the mathematical model derived in this study only the constants depicted in cycles I and II are included. Note that because we are using cooperativity parameters, the equilibrium constant for a particular equilibrium can be written in different ways depending on which equilibrium is referred to. For instance, the right side of Panel II and the right side of Panel III are governed by the same equilibrium constant, but they are expressed in a different way because they are referred to different equilibria (K$_{AI}$ and K$_{AP}$, respectively).
Fig. 4. Conformation-based mathematical modeling for GPCR signaling and its validation. (A) Experimental/empirical data of signaling responses upon partial and full agonist bound to the GPCR, along with basal signaling of the receptor. (B) Molecular basis of GPCR signaling efficacy used for modeling buildup. The population distributions of inactive states (I$_{1-2}$) partially activated state (P), and fully activated state (F) were re-equilibrated as partial agonist, full agonist, and $G_{o}$ bound to the receptor. (C) Modelled concentration-effect curves for LUF and NECA agonists by using the mathematical model depicted in Fig. 4 and Equation 3 (the equation for the asymptotic maximum fractional effect (efficacy) is shown in the figure, with the following parameters for the $A_{2A}$R system: (i) the receptor system: $\chi$=0.04, $K_{1}$=2.2, $K_{2}$=6.8, $\epsilon_{RPA}$=1, $\epsilon_{RFA}$=5; (ii) LUF: $K_{AI}$=$10^{-9}$, $\epsilon_{ARPA}$=30, $\epsilon_{ARFA}$=50, $\alpha$=36.8, $\beta$=2.6; (iii) NECA: $K_{AI}$=$10^{-7}$, $\epsilon_{ARPA}$=40, $\epsilon_{ARFA}$=1000, $\alpha$=6.8, $\beta$=12.4.
Fig. 1
Fig. 2
Fig. 3
A. Pharmacological responses

- Full agonism signaling
- Partial agonism signaling
- Basal signaling

B. Quantified substates profile

- $I_{1-2}$
- $P$
- $F$

  - +Full agonist + $G_{\alpha \beta \gamma}$
  - +Partial agonist + $G_{\alpha \beta \gamma}$
  - apo + $G_{\alpha \beta \gamma}$

C. Output of conformation-based mathematical model

$$\lim_{[A] \to \infty} \frac{E}{E_m} = \frac{\chi (\varepsilon_{ARPA} \alpha K_1 + \varepsilon_{ARFA} \beta K_2)}{1 + \alpha K_1 + \beta K_2 + \chi (\varepsilon_{ARPA} \alpha K_1 + \varepsilon_{ARFA} \beta K_2)}$$

Fig. 4