Supporting information for

Identification of CB1 Receptor Allosteric Sites Using Force-Biased MMC Simulated Annealing and Validation by Structure–Activity Relationship Studies

Dow P. Hurst, [Φ] Sumanta Garai, [†] Pushkar M Kulkarni, [†] Peter C. Schaffer, [†] Patricia H. Reggio* [Φ] and Ganesh A. Thakur*[†]

[Φ] Department of Chemistry and Biochemistry, University of North Carolina at Greensboro, Greensboro NC USA

[†] Department of Pharmaceutical Sciences, Bouvé College of Health Sciences, Northeastern University, Boston, Massachusetts 02115, USA

g.thakur@northeastern.edu

phreggio@uncg.edu

Supporting Figure:

Figure S1A shows the large simulation box in which the CB1 model was embedded. Here the box is filled with the indole ring (Fragment B; light green). Figure S1B shows the system later in the annealing schedule. Here, one copy of Fragment B (light green, tube) has established a position against the lipid face of TMH2 in a pocket at the intracellular (IC) end of TMH1-2-4 on CB1. In addition to showing the location of Fragment B, Figure S1C shows Fragments A_R, and C of GAT228 (tube display; mauve). These each are the result of separate MMC runs that identified a binding site for each fragment in the IC TMH1-2-4 exosite in the position shown. The order of the fragments, in this case, reflects the structure of GAT228.

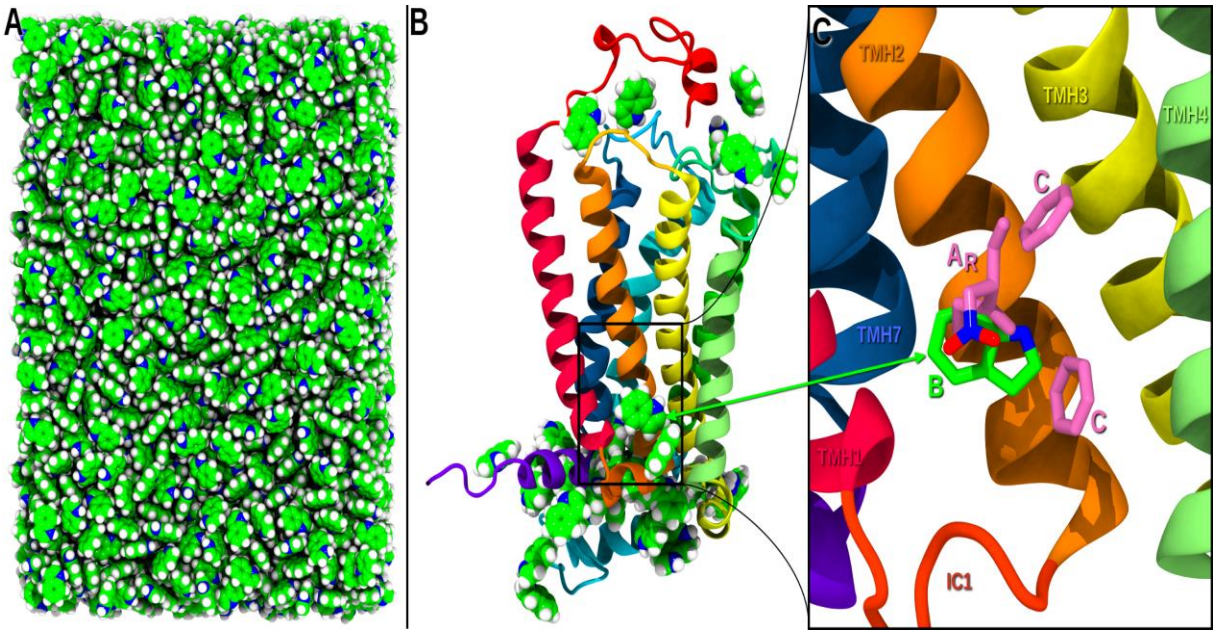

Figure S1. (A) The large simulation box in which the CB1 model was embedded and filled with the GAT228 indole ring fragment (Fragment B, light green) is shown. **(B)** This figure shows the system later in the annealing schedule. Here, one copy of Fragment B (light green, tube) has established a position against the lipid face of TMH2 in a pocket at the intracellular (IC) end of TMH1-2-4 on CB1. **(C)** In addition to showing the location of Fragment B, Figure 3C shows Fragments A_R, and C of GAT228 (tube display; mauve). These each are the result of separate MMC runs that identified a binding site for each fragment in the IC TMH1-2-4 exosite in the position shown. The order of the fragments, in this case, reflects the structure of GAT228.

METHODS

MMC. To identify potential binding site(s) for GAT228 (*R*) and GAT229 (*S*) binding at CB1, we used the Force-Biased Metropolis Monte Carlo simulated annealing program (MMC) as a first step. [1, 2] The MMC method has been used successfully to identify water binding sites in proteins and on DNA [1]; to identify potent and novel p38 kinase inhibitors [3]; to identify thermolysin and T4 lysozyme binding sites [2] ;and, to identify the binding site of the CB1 negative allosteric modulator, pregnenolone. [4] In MMC, the molecule of interest is first divided into smaller fragments. A series of grand canonical ensembles of a molecular fragment interacting with the protein in a large simulation box are created. The chemical potential of the system is then annealed at descending chemical potential levels, with each new level starting from the last ensemble generated from the previous one. At each step, fragment poses are sampled throughout the box and over the entire protein. Fragments are treated as rigid solvents that are inserted and deleted millions of times until the lowest energy configuration is found at the explored annealing level. As the chemical potential is annealed, the number of fragments in the box decreases because the method eliminates any fragment that has a poorer free energy of interaction than the annealing level. The output of the calculation is an ensemble of fragment poses at each chemical potential level in the annealing schedule. The method is repeated for each molecular fragment. Data analysis is performed using the GENS tool in the MMC program available from the Mihaly Mezei Laboratory (see http://inka.mssm.edu/~mezei/mmc/).

Of particular interest in MMC calculations are fragments that persist at particular sites on the protein throughout the annealing schedule, since these fragments clearly have high affinity for those sites. Receptor regions at which all molecular fragments of a studied ligand collect are identified from MMC output. This set of sites is then refined to include only those sites at which the order of the fragments reflects the structure of the entire molecule.

Both GAT228 and GAT229 were built in Maestro (Suite 2018-4, Schrödinger, Inc.). Conformational analyses of these compounds were performed using the OPLS3e force field as implemented in the Schrödinger Suite 2018-4[5]. Parameterization of three missing dihedrals (present in the two structures) was performed using the Schrödinger Force Field Builder tool for OPLS3e. Conformational searches were performed (using 3 to 8-fold rotations) for each rotatable bond and all unique conformers identified. To calculate the difference in energy between the global minimum energy conformer of each compound and its final docked conformation, rotatable bonds in the global minimum energy conformer were driven to their corresponding value in the final docked conformation and the single-point energy of the resultant structure was calculated using the OPLS3e force field.

CB1 HOMOLOGY MODEL USED FOR THE *IN SILICO* STUDIES

MMC runs required the identification of a solvent fragment and the presence of a CB1 receptor model. The CB1 receptor model used in this study was based on the CB1 activated state crystal structure (PDB-ID: 5XRA)[6] and the cryo-EM CB1/Gi bound structure (PDB-ID: 6N4B)[7].

Structures were prepared with the Protein Preparation protocol (Suite 2019-1, Schrödinger, Inc.), mutations in the 5XRA structure were returned to WT, and both models were inspected for close contacts and crystal packing effects. Modification of the 5XRA structure to create the CB1 model involved calculating low free energy conformations for TMH2/7 to accommodate mutation data for S7.39 [8], F2.61[9], F2.64[9]. In addition, because of loop compression in the 5XRA structure, the IC1 loop from the 6N4B Gi bound structure was used in the model. TMH4 in the 5XRA CB1 R* structure[6] has tight crystal contacts on its IC end with TMH1 of another bundle because of antiparallel packing. For this reason, the conformation of TMH4 in the 6N4B Gi bound structure [7] was used in the model. Modeller[10] was used to remodel the EC3 loop (D6.58 to T7.33) and allow K(373) to interact with D2.63, consistent with mutation results[11]. Modeller was also used to extend and model the N-terminus to S(88) with inclusion of the C(98) to C(107) disulfide bridge. [12, 13] The receptor/ligand complex was energy minimized in Prime (Suite 2019-1, Schrödinger, Inc.). The Prime implicit membrane functionality was employed. Hydrophilic residues facing the binding crevice and within the low dielectric region of the implicit membrane were excluded from the low dielectric via exclusion spheres placed on each residue. The Generalized Born/Surface Area (GB/SA) continuum solvation model for water was used with the dielectric set to 80 outside of the implicit membrane region and 1 within. A truncated Newton conjugate gradient minimization was performed using the OPLS3e force field for one iteration up to a maximum of 1000 steps and with a 0.1 kcal/mol gradient endpoint. Constraints of 1 kcal/mol were placed on the C-alpha atoms of residues R3.50, Y5.58, L6.33, and Y7.53 to prevent the intracellular opening present in the R* structures from closing during the minimization.

Docking. Once the two sites were identified, the structure of GAT228 was docked in the intracellular TMH1-2-4 exosite identified by MMC, with R(149), H2.41, F4.46, and W4.50 used as direct interaction sites. GAT229 was docked in the TMH2-3-EC1 site identified by MMC, with Y2.59 and D2.63 as direct interaction sites. The receptor/ligand complexes were energy minimized in Prime (Suite 2019-1, Schrödinger, Inc.). The Prime implicit membrane functionality was employed. Hydrophilic residues facing the binding crevice and within the low dielectric region of the implicit membrane were excluded from the low dielectric via exclusion spheres placed on each residue. The Generalized Born/Surface Area (GB/SA) continuum solvation model for water was used with the dielectric set to 80 outside of the implicit membrane region and 1 within. A truncated Newton conjugate gradient minimization was performed using the OPLS3e force field for one iteration, up to a maximum of 1000 steps and with a 0.1 kcal/mol gradient endpoint. In order to prevent the intracellular opening in the R* model from closing during the minimization, constraints of 1 kcal/mol were placed on the C-α atoms of R3.50, Y5.58, L6.33, and Y7.53. The resulting docks were refined with the Induced Fit protocol (Suite 2019-1, Schrödinger, Inc.). The Glide box size was set to 12 $Å^3$ centered on the ligand and the SP docking algorithm employed. Residues within 5 Å of the docked ligand were included in the Prime refinement stage, except in the case of the TMH2-3-EC1 site where S2.60 and K3.28 were excluded based on mutation data. [8, 14] The implicit membrane previously used during the initial Prime minimization was employed here as well.

Assessment of Pair-wise Interaction Energies. After defining the atoms of GAT228 (or GAT229) as one group (Group 1) and the atoms corresponding to a residue that lines the respective binding site in the final ligand/CB1 R* complex as another group (Group 2), Macromodel (version 8.6, Schrödinger, LLC, New York, NY) was used to output the pair-wise interaction energy (coulombic and Van der Waals) for a given pair of atoms. The pairs corresponding to Group 1 (ligand) and Group 2 (residue of interest) were then summed to yield the interaction energy between the ligand and that residue. The dielectric constant used for the electrostatic term followed the five-slab membrane model developed by Sengupta et al.[15] , with ε=2 or ε=10 used depending on the environment of the residue and ligand.

Pairwise interaction energies for GAT228 at the CB_1 R* Ago-Allosteric Site

Table S1. GAT228/CB_1-R* Complex

Residue	VdW (kcal/mol)	Electrostatic (kcal/mol)	Total Energy (kcal/mol)
R(148)	-6.06	-3.18	-9.24
H2.41	-3.33	-3.30	-6.63
F4.46	-0.13	-4.84	-4.97
L1.58	0.01	-3.48	-3.47
V2.48	0.03	-3.25	-3.21
W4.50	-0.13	-2.53	-2.66
G2.44	-0.08	-2.13	-2.22
S2.45	0.08	-1.59	-1.51
V4.43	0.13	-1.49	-1.37
I1.57	0.12	-1.40	-1.28
L1.54	0.04	-1.16	-1.12
Y2.40	-0.03	-0.40	-0.43
A2.47	0.00	-0.22	-0.22
C4.47	0.04	-0.25	-0.21
L2.52	0.01	-0.12	-0.11
A2.49	0.02	-0.13	-0.11
A4.42	0.00	-0.10	-0.10
		Subtotal	-38.86
	GAT228 Conformational Cost		1.68
		Total	**-37.18**

Pairwise interaction energies for GAT229 at the CB_1 R* PAM Site

Table S2. GAT229/CP55,940/CB_1-R* Complex

Residue	VdW (kcal/mol)	Electrostatic (kcal/mol)	Total Energy (kcal/mol)
Y2.59	-2.29	-4.45	-6.74
R(182)	-0.03	-5.82	-5.85
V3.24	0.02	-4.99	-4.97
D2.63	-0.96	-3.38	-4.35
D(184)	-0.40	-1.79	-2.19
V2.66	-0.00	-2.01	-2.02
I2.62	0.12	-1.95	-1.83
N3.23	-0.03	-1.55	-1.57
K(183)	0.00	-1.39	-1.39
F(180)	-0.03	-0.89	-0.92
S(185)	0.01	-0.40	-0.39
F3.27	-0.13	-0.23	-0.36
H(181)	0.03	-0.28	-0.26
S2.60	-0.01	-0.21	-0.22
K(373)	0.45	-0.57	-0.12
		Subtotal	-33.18
	GAT229 Conformational Cost		0.45
		Total	**-32.73**

CHEMISTRY

Materials and Methods: All commercial chemicals and solvents were purchased from Sigma Aldrich, Inc. (St. Louis, MO), Alfa Aesar, and Combi-blocks as reagent grade and unless otherwise specified were used without further purification. Biotage® Initiator microwave system was used for the synthesis of a few of the intermediates of the final covalent probes. Reaction progress was monitored by thin-layer chromatography (TLC) using commercially prepared silica gel 60 F254 glass plates. Compounds were visualized under ultraviolet (UV) light or by staining with iodine, phosphomolybdic acid, or *p*-anisaldehyde reagent. Flash column chromatography was carried out on a Biotage® SP1, Biotage® Isolera, or Interchim purification unit using prepacked columns from Reveleris, Biotage, and Luknova. Solvents used include hexanes, ethyl acetate, acetone, methanol, and dichloromethane. Characterization of compounds and their purity were established by a combination of HPLC, TLC, mass spectrometry, and NMR analyses. NMR spectra were recorded in chloroform-d on a Varian NMR spectrometer (^{1}H NMR at 500 MHz and 400 MHz). Chemical shifts were recorded in parts per million (δ) relative to tetramethylsilane (TMS; 0.00 ppm) or solvent peaks as the internal reference. Multiplicities are indicated as bs (broad singlet), s (singlet), d (doublet), t (triplet), q (quartet), quin (quintet), sept (septet), or m (multiplet). Coupling constants (*J*) are reported in hertz (Hz). All test compounds were greater than 95% pure as determined by LC/MS analysis performed using an Agilent Technologies 1260 Infinity reverse phase HPLC, with a dual-wavelength UV−visible detector and an Agilent Technologies 6120 Quadrupole mass spectrometer (electrospray ionization). HRMS was done on SCIEX TOF/TOFTM 5800 MALDI system ion mode with a delay time of 100 ns. Each sample well was surveyed to find a "sweet spot", and then 400 laser pulses were averaged to generate a spectrum.

2-phenyl-2-(2-phenyl-1H-indol-3-yl)ethan-1-amine (2): The nitro compound **1** (100 mg, 0.292 mmol) was dissolved in THF (15 ml) and methanol (2 ml), $NiCl_2.6H_2O$ (130 mg, 0.547 mmol) was added to the solution under argon atmosphere. The reaction mixture was stirred at room temperature for 45 min then cooled -5 ^{0}C. Slowly and portion wise $NaBH_4$ (50 mg, 1.32 mmol) was added to the reaction mixture and stirred it for 2 hours at 0 ^{0}C then 1 hour at room temperature. After the full consumption of starting material, the reaction mixture was quenched with sat. aq. solution of NH_4Cl, the organic solvent was removed under reduced pressure. The aq. layer was extracted with ethyl acetate (3*10 ml) and combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under vacuum. The crude product obtained was purified by

silica gel flash column chromatography to obtain pure amine **2**[16] (74 mg, 81%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.22 (brs, 1H), 7.63 (d, *J* = 7.8 Hz, 1H), 7.48 (d, *J* = 6.9 Hz, 2H), 7.43-7.35 (m, 6H), 7.30-7.26 (m, 2H), 7.20 (t, *J* = 7.3 Hz, 2H), 7.07 (t, *J* = 7.8 Hz, 1H), 4.44 (dd *J* = 9.4, 6.2 Hz, 1H), 3.57 (br, 1H), 3.46 (bs, 1H), 2.05 (bs, 2H); MS-ESI (m/z): 313 $[M+H]^+$.

3-(2-Nitroethyl)-1H-indole (5): 1H-Indole-3-carbaldehyde, **3** (500 mg, 2.04 mmol) was taken in anhydrous nitromethane (15 ml), then 3.0 gm ammonium acetate was added and stirred at 115 ^{0}C for 4 hours under argon atmosphere. The reaction mixture was cooled, and excess nitromethane was removed under vacuum. The residue was partitioned in ethyl acetate and water. The combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under vacuum. The crude product obtained was purified by silica gel flash column chromatography to obtain unsaturated nitro compound **4** (446 mg, 82%).

The unsaturated nitro compound, **4** (150 mg, 0.567 mmol) was dissolved in methanol (20 ml) and water (2 ml) and cooled to 0 ^{0}C. $NaBH_4$ (200 mg, 5.28 mmol) was added to the reaction mixture and stirred at that temperature for 2 hours and 1 hour at room temperature. The reaction mixture was quenched with sat. aq. solution of NH_4Cl then organic solvent was evaporated under vacuum. The reaction mixture was extracted with ethyl acetate (3*10 ml) and combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under vacuum. The crude product was purified by recrystallization from methanol yielded saturated nitro compound **5** (109 mg, 72%). ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.15 (s, 1H), 7.62 (d, *J*=7.5 Hz, 1H), 7.54-7.51 (m, 4H), 7.45-7.43 (m, 1H), 7.41 (d, *J*=8.0 Hz, 1H), 7.28-7.24 (m, 1H), 7.22-7.18 (m, 1H), 4.63-4.60 (m, 2H), 3.67-3.64 (m, 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 136.0, 135.7, 132.2, 129.2, 128.4, 128.2, 128.0, 122.8, 120.3, 118.3, 111.1, 106.2, 75.0, 23.3; HRMS (TOF ESI): calcd for $C_{16}H_{14}N_2O_2$: 267.1128 $[M + H]^+$; found: 267.1150.

3-(2-nitro-1-phenylethyl)-1H-indole (8): In a 100 ml round bottom flask, 1H-Indole, **6** (1.1 gm, 9.39 mmol), β-nitrostyrene (1.54 gm, 10.33 mmol), Et_4NBr (2.17 gm, 10.33 mmol) were dissolved in dioxane:water (2:1, 10 ml) and refluxed for 10 hours. The organic solvent was evaporated under reduced pressure and crude mixture was extracted with ethyl acetate (3*10 ml) and combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under vacuum. The crude product was purified by recrystallization from methanol yielded Michael adduct **8**[17, 18] (1.7 gm, 68%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.05 (bs, 1H), 7.44 (d, *J*=8.4 Hz, 1H), 7.35-7.29 (m,

5H), 7.26-7.23 (m, 1H), 7.21-7.17 (m, 1H), 7.07 (t, *J*=7.5 Hz, 1H), 7.01 (d, *J*=1.9 Hz, 1H), 5.19 (t, *J*=7.9 Hz, 1H), 5.05 (dd, *J*=13.1, 8.2 Hz, 1H), 4.93 (dd, *J*=12.7, 8.9 Hz, 1H); MS-ESI (m/z): 267 $[M+H]^+$.

1-methyl-3-(2-nitro-1-phenylethyl)-2-phenyl-1H-indole (9): In a 100 ml round bottom flask, 1-methyl-2-phenyl-1H-indole, **7** (500 mg, 2.41 mmol), β-nitrostyrene (350 mg, 2.34 mmol), Et_4NBr (450 mg) were dissolved in dioxane:water (2:1, 10 ml) and refluxed for 10 hours. The organic solvent was evaporated under reduced pressure and crude mixture was extracted with ethyl acetate (3*10 ml) and combined organic layers were dried over anhydrous Na_2SO_4, filtered and concentrated under vacuum. The crude product was purified by recrystallization from methanol yielded Michael adduct **9**[17] (526 mg, 61%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.56 (d, *J*=8.7 Hz, 1H), 7.49-7.48 (m, 3H), 7.36 (d, *J*=8.7 Hz, 1H), 7.29-7.25 (m, 7H), 7.20-7.19 (m, 1H), 7.13 (t, *J*=7.9 Hz, 1H), 5.10-5.09 (m, 2H), 5.05-5.02 (m, 1H), 3.53 (s, 3H); MS-ESI (m/z): 281 $[M+H]^+$.

3-(2-nitro-1-phenylethyl)-1-phenyl-1H-indole (10): In a microwave vial 3-(2-nitro-1 -phenylethyl)-1H-indole, **8** (200 mg, 0.751 mmol), iodobenzene (306 mg, 1.50 mmol), K_3PO_4 (478 mg, 2.25 mmol) in DMSO were taken, the vial was sealed and heated to 150 ^{0}C under microwave irradiation for 90 min. Reaction mixture was cooled to room temperature, diluted with water and extracted with dichloromethane. The organic layer was separated and washed with brine and dried (Na_2SO_4). Evaporation of volatiles under reduced pressure followed by purification by silica gel flash column chromatography to give **10** (65 mg, 25.3%); ^{1}H NMR (500 MHz, $CDCl_3$): δ 7.54-7.45 (m, 6H), 7.42-7.32 (m, 5H), 7.27 (tt, *J*=7.5, 1.5 Hz, 1H), 7.22 (td, *J*=8.0, 1.0 Hz, lH),7.17(s, 1H), 7.13 (td, *J*= 7.5, 1.5 Hz, 1H), 5.25 (t, *J*=7.5 Hz, 1H), 5.11 (dd, *J*= 12.5, 7.5 Hz, 1H), 4.98 (dd, J= 12.5 8.5 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 139.2, 138.9, 136.4, 129.6, 128.9, 127.8, 127.6, 127.4, 126.6, 125.3, 124.3, 123.0, 120.5, 119.2, 115.3, 110.8, 79.4, 41.4; HRMS (TOF ESI): calcd for $C_{16}H_{14}N_2O_2$: 343.1441 $[M + H]^+$; found: 343.1444.

2-cyclohexyl-3-(2-nitro-1-phenylethyl)-1H-indole (12): In a 100 ml round bottom flask, 2-cyclohexyl-1H-indole, **11**[19] (200 mg, 1.08 mmol), β-nitrostyrene (270 mg, 1.62 mmol) and ammonium trifluoro acetate (70 mg, 0.54 mmol) were dissolve with 10% aq. ethanol (5 ml).[20] The reaction mixture was refluxed at 110 ^{0}C for 12 hours. After completion of reaction, ethanol was evaporated under reduced pressure and quenched with water. The reaction mixture was diluted with ethyl acetate and extracted. The combined organic layers were evaporated under vacuum and

the reaction mixture was purified by silica gel column chromatography to Michael adduct, **12** (220 mg, 63%); ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.94 (bs, 1H), 7.35-7.20 (m,7H), 7.11 (t, *J*=7.73 Hz, 1H), 7.01 (t, *J*=7.47 Hz, 1H), 5.30-5.22 (m, 2H), 5.14-5.05 (m, 1H), 2.92-2.87 (m, 1H), 1.95-1.75 (m, 5H), 1.49-1.35 (m, 4H), 1.30-1.24 (m, 1H); ^{13}C NMR (100 MHz, DMSO d_6): δ 142.3, 140.5, 135.6, 128.4, 127.1, 126.5, 125.8, 120.2, 118.7, 118.5, 110.9, 106.4, 78.3, 39.8, 35.3, 32.8, 32.4, 26.2, 26.1, 25.6; HRMS (TOF ESI): calcd for $C_{22}H_{24}N_2O_2$: 349.1911 $[M + H]^+$; found: 349.2104.

3-(2-nitro-1-phenylethyl)-2-phenyl-1H-pyrrole (15):

To a 250 ml round bottom flask were added hydroxylamine hydrochloride (4.623 g, 66.56 mmol), sodium acetate (6.829 g, 83.2 mmol,) and 80% aqueous ethanol (60 ml). The reaction mixture was stirred at room temp for 40 minutes. Ensuing 40 minutes of room temp stirring, acetophenone, **13** (41.6 mmol, 4.85 ml) was added to the reaction mixture; the reaction was then refluxed at 105-110 °C. Following 1 hour of reflux, the reaction mixture was filtered via vacuum filtration; filtrate collected was purified by flash chromatography using a silica gel column chromatography to yield acetophenone oxime (5 gm, 90%).

To a 2-neck, 100 mL round bottom flask were added acetophenone oxime (500 mg, 3.7 mmol), potassium hydroxide (207 mg, 3.7 mmol,), and DMSO (20 ml). The reaction mixture was stirred and heated up at 150°C. Calcium carbide (approx. 20 g) was added to an Erlenmeyer flask; the flask was then capped with a septum. Acetylene gas was generated in a separate, capped Erlenmeyer flask containing calcium carbide; water was slowly added to the calcium carbide containing flask via a syringe to generate acetylene gas which was transferred and directly bubbled into the stirred and heated reaction mixture via a canula. Water was added in 45-minute intervals over 4 hours to the calcium carbide flask to transfer an ample amount of acetylene gas to the reaction mixture. Following 4 hours of heating under acetylene gas, the reaction mixture was cooled to room temp, diluted with water, then extracted with diethyl ether (5* 20 ml). Combined organic extracts were washed with brine, dried over anhydrous Na_2SO_4, then concentrated under reduced pressure. The organic resin was purified via flash chromatography using a silica gel to obtain desired product, 2-phenylpyrrole, **14**[21] (235 mg, 44%).

To a 100 ml round bottom flask were added 2-phenylpyrrole, **14** (225 mg, 1.57 mmol), β-nitrostyrene (345 mg, 2.31 mmol), ammonium trifluoroacetate (101 mg, 0.77 mmol,), and 10%

aqueous ethanol (15 ml) The reaction mixture was refluxed at 115 °C for 19 hours. Following reflux, the reaction mixture was concentrated under reduced pressure, diluted with EtOAc, then washed with water; aq. phase was extracted 3X with ethyl acetate. Combined organic extracts were washed with brine, dried over anhydrous Na_2SO_4 then concentrated under reduced pressure. Organic resin was purified by flash chromatography using a silica gel to obtain the desired product, 3-(2-nitro-1-phenylethyl)-2-phenylpyrrole, **15** (418 mg, 91%).^{1}H NMR (500 MHz, $CDCl_3$): δ 8.05 (s, 1H), 7.39 – 7.34 (m, 4H), 7.33 – 7.28 (m, 3H), 7.28 – 7.25 (m, 2H), 7.17 (ddt, *J* = 8.1, 6.9, 1.4 Hz, 1H), 6.43 (t, *J* = 3.5 Hz, 1H), 6.14 – 6.11 (m, 1H), 4.99 (dd, *J* = 12.0, 7.5 Hz, 1H), 4.93 (t, *J* = 7.5 Hz, 1H), 4.80 (dd, *J* = 12.0, 7.5 Hz, 1H) ^{13}C NMR (100 MHz, Chloroform-*d*) δ 137.8, 132.6, 132.2, 130.0, 129.3, 128.8, 128.2, 127.9, 126.4, 123.7, 107.7, 106.2, 79.1, 43.0; HRMS (TOF ESI): calcd for $C_{18}H_{16}N_2O_2$: 293.1285 $[M + H]^+$; found: 293.1341.

Reference:

1. Guarnieri, F.; Mezei, M., Simulated annealing of chemical potential: A general procedure for locating bound waters. Application to the study of the differential hydration propensities of the major and minor grooves of DNA. *J. Am. Chem. Soc.* **1996,** *118*, 8493-8494.
2. Clark, M.; Guarnieri, F.; Shkurko, I.; Wiseman, J., Grand canonical Monte Carlo simulation of ligand-protein binding. *J Chem Inf Model* **2006,** *46*, 231-42.
3. Moore, W. R., Maximizing discovery efficiency with a computationally driven fragment approach. *Curr. Opin. Drug Disc. Devel.* **2005,** *8*, 355-364.
4. Vallee, M.; Vitiello, S.; Bellocchio, L.; Hebert-Chatelain, E.; Monlezun, S.; Martin-Garcia, E.; Kasanetz, F.; Baillie, G. L.; Panin, F.; Cathala, A.; Roullot-Lacarriere, V.; Fabre, S.; Hurst, D. P.; Lynch, D. L.; Shore, D. M.; Deroche-Gamonet, V.; Spampinato, U.; Revest, J. M.; Maldonado, R.; Reggio, P. H.; Ross, R. A.; Marsicano, G.; Piazza, P. V., Pregnenolone can protect the brain from cannabis intoxication. *Science* **2014,** *343*, 94-8.
5. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of chemical theory and computation* **2016,** *12*, 281-96.
6. Hua, T.; Vemuri, K.; Nikas, S. P.; Laprairie, R. B.; Wu, Y.; Qu, L.; Pu, M.; Korde, A.; Jiang, S.; Ho, J. H.; Han, G. W.; Ding, K.; Li, X.; Liu, H.; Hanson, M. A.; Zhao, S.; Bohn, L. M.; Makriyannis, A.; Stevens, R. C.; Liu, Z. J., Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* **2017**.
7. Krishna Kumar, K.; Shalev-Benami, M.; Robertson, M. J.; Hu, H.; Banister, S. D.; Hollingsworth, S. A.; Latorraca, N. R.; Kato, H. E.; Hilger, D.; Maeda, S.; Weis, W. I.; Farrens, D. L.; Dror, R. O.; Malhotra, S. V.; Kobilka, B. K.; Skiniotis, G., Structure of a Signaling Cannabinoid Receptor 1-G Protein Complex. *Cell* **2019,** *176*, 448-458 e12.
8. Kapur, A.; Hurst, D. P.; Fleischer, D.; Whitnell, R.; Thakur, G. A.; Makriyannis, A.; Reggio, P. H.; Abood, M. E., Mutation Studies of Ser7.39 and Ser2.60 in the Human CB1 Cannabinoid Receptor: Evidence for a Serine-Induced Bend in CB1 Transmembrane Helix 7. *Mol Pharmacol* **2007,** *71*, 1512-24.
9. Shim, J. Y.; Bertalovitz, A. C.; Kendall, D. A., Identification of essential cannabinoid-binding domains: structural insights into early dynamic events in receptor activation. *J Biol Chem* **2011,** *286*, 33422-35.

10. Fiser, A.; Do, R. K.; Sali, A., Modeling of loops in protein structures. *Protein Sci* **2000,** *9*, 1753-73.
11. Marcu, J.; Shore, D. M.; Kapur, A.; Trznadel, M.; Makriyannis, A.; Reggio, P. H.; Abood, M. E., Novel insights into CB1 cannabinoid receptor signaling: a key interaction identified between the extracellular-3 loop and transmembrane helix 2. *J Pharmacol Exp Ther* **2013,** *345*, 189-97.
12. Fay, J. F.; Farrens, D. L., The membrane proximal region of the cannabinoid receptor CB1 N-terminus can allosterically modulate ligand affinity. *Biochemistry* **2013,** *52*, 8286-94.
13. Fay, J. F.; Farrens, D. L., Structural dynamics and energetics underlying allosteric inactivation of the cannabinoid receptor CB1. *Proc Natl Acad Sci U S A* **2015,** *112*, 8469-74.
14. Song, Z. H.; Bonner, T. I., A lysine residue of the cannabinoid receptor is critical for receptor recognition by several agonists but not WIN55212-2. *Mol Pharmacol* **1996,** *49*, 891-6.
15. Sengupta, D.; Smith, J. C.; Ullmann, G. M., Partitioning of amino-acid analogues in a five-slab membrane model. *Biochim Biophys Acta* **2008,** *1778*, 2234-43.
16. Noland, W. E.; Lange, R. F., The Nitroethylation of Indoles. III.1-3 A Synthetic Route to Substituted Tryptamines. *Journal of the American Chemical Society* **1959,** *81*, 1203-1209.
17. Gu, Y.; Barrault, J.; Jérôme, F., Glycerol as An Efficient Promoting Medium for Organic Reactions. *Advanced Synthesis & Catalysis* **2008,** *350*, 2007-2012.
18. Fan, Y.; Kass, S. R., Enantioselective Friedel–Crafts Alkylation between Nitroalkenes and Indoles Catalyzed by Charge Activated Thiourea Organocatalysts. *The Journal of Organic Chemistry* **2017,** *82*, 13288-13296.
19. La Regina, G.; Bai, R.; Rensen, W. M.; Di Cesare, E.; Coluccia, A.; Piscitelli, F.; Famiglini, V.; Reggio, A.; Nalli, M.; Pelliccia, S.; Da Pozzo, E.; Costa, B.; Granata, I.; Porta, A.; Maresca, B.; Soriani, A.; Iannitto, M. L.; Santoni, A.; Li, J.; Miranda Cona, M.; Chen, F.; Ni, Y.; Brancale, A.; Dondio, G.; Vultaggio, S.; Varasi, M.; Mercurio, C.; Martini, C.; Hamel, E.; Lavia, P.; Novellino, E.; Silvestri, R., Toward Highly Potent Cancer Agents by Modulating the C-2 Group of the Arylthioindole Class of Tubulin Polymerization Inhibitors. *Journal of Medicinal Chemistry* **2013,** *56*, 123-149.
20. Kulkarni, P. M.; Ranade, A.; Garai, S.; Thakur, G. A., Microwave-accelerated Conjugate Addition of 2-Arylindoles to Substituted β-Nitrostyrenes in the Presence of Ammonium Trifluoroacetate: An Efficient Approach for the Synthesis of a Novel Class of CB1 Cannabinoid Receptor Allosteric Modulators. *Journal of Heterocyclic Chemistry* **2017,** *54*, 2079-2084.
21. Mikhaleva, A. I.; Petrova, O. V.; Sobenina, L. N., 2-Phenylpyrrole: one-pot selective synthesis from acetophenone oxime and acetylene by a Trofimov reaction*. *Chemistry of Heterocyclic Compounds* **2012,** *47*, 1367-1371.