Biochemistry

Article

Membrane Phospholipid Analogues as Molecular Rulers to Probe the Position of Hydrophobic Contact Point of Lysophospholipid Ligands on the Surface of G-Protein-coupled Receptor during Membrane Approach

Misa Sayama, Akiharu Uwamizu, Yuko Otani, Asuka Inoue, Junken Aoki, Masakazu Sekijima, and Tomohiko Ohwada

Biochemistry, **Just Accepted Manuscript** • DOI: 10.1021/acs.biochem.0c00061 • Publication Date (Web): 03 Mar 2020

Membrane Phospholipid Analogues as Molecular Rulers to Probe the Position of Hydrophobic Contact Point of Lysophospholipid Ligands on the Surface of G-Protein-coupled Receptor during Membrane Approach

Misa Sayama [1], Akiharu Uwamizu [2], Yuko Otani [1], Asuka Inoue [2,3,4], Junken Aoki [2,4,5,*], Masakazu Sekijima [6,7], Tomohiko Ohwada [1,*]

[1]*Laboratory of Organic and Medicinal Chemistry, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
[2]*Laboratory of Molecular and Cellular Biochemistry, Graduate School of Pharmaceutical Sciences, Tohoku University, 6-3, Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980-8578, Japan*
[3]*AMED-PRIME, Japan Science and Technology Corporation, Kawaguchi, 332-0012, Japan*
[4]*AMED-LEAP, Japan Science and Technology Corporation, Kawaguchi, 332-0012, Japan*
[5]*AMED-CREST, Japan Science and Technology Corporation, Kawaguchi, 332-0012, Japan*
[6]*Department of Computer Science, Tokyo Institute of Technology, 4259-J3-23, Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan*
[7]*Advanced Computational Drug Discovery Unit, Tokyo Institute of Technology, 4259-J3-23, Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa 226-8501, Japan*

*To whom correspondence should be addressed.
Chemical study: For T.O.: phone, +81-3-5841-4730; fax, +81-3-5841-4735;
E-mail, ohwada@mol.f.u-tokyo.ac.jp
Biological study: For J.A.: phone, +81-22-795-6860; fax, +81-22-795-6859;
E-mail, jaoki@m.tohoku.ac.jp

Abstract

When lipid mediators bind to G-protein-coupled receptors (GPCRs), the ligand first enters the lipid bilayer, then diffuses laterally in the cell membrane to make hydrophobic contact with the receptor protein, and finally enters the receptor's binding pocket. In this process, the location of the hydrophobic contact point on the surface of the receptor has been little discussed even in cases where the crystal structure has been solved, because the ligand-binding pocket is buried inside the transmembrane (TM) domains. Here, we coupled an activator ligand to a series of membrane phospholipid surrogates, which constrain the depth of ligand entry into the lipid bilayer. Consequently, by measuring the receptor-activating activity as a function of depth of entry into the membrane, these surrogates can be used as molecular rulers to estimate the location of the hydrophobic contact point on the surface of GPCR. We focused on lysophosphatidylserine (LysoPS) receptor, GPR34 and prepared a series of simplified membrane-lipid-surrogate-conjugated lysophospholipid analogues by attaching alkoxy amine chains of varying lengths to the hydrophobic tail of a potent GPR34 agonist. As expected, the activity of these lipid-conjugate LysoPS analogues was chain-length-dependent. The predicted contact position matches the position of the terminal benzene ring of a non-lipidic ligand that protrudes between TMs 4 and 5 of the receptor. We further found that the nature of the terminal hydrophilic functional group of the conjugated membrane lipid surrogate strongly influences the activity, suggesting that lateral hydrophilic contact of LysoPS analogues with the receptor's surface is also crucial for ligand-GPCR binding.

Introduction

G-Protein-coupled receptors (GPCRs) are membrane receptors containing seven transmembrane (TM) helices. Various types of compounds activate GPCRs, including biogenic amines, peptides, photons, inorganic ions, and lipids. While GPCRs for biogenic amines such as adrenaline, acetylcholine, histamine, dopamine and serotonin have been well characterized as therapeutic targets, GPCRs that bind lipid ligands are also emerging as candidate targets for drug discovery. A representative example is sphingosine-1-phosphate receptor subtype 1 ($S1P_1$), and a synthetic functional antagonist (superagonist) of this receptor, fingolimod, has been developed to treat multiple sclerosis.[1] From both structural and physicochemical points of views, endogenous lipidic ligands are different from hydrophilic aminergic ligands, in that the former have a long hydrophobic fatty acid moiety. Therefore, the ligand-binding pathways of lipidic ligands are expected to be different from those of small hydrophilic aminergic ligands: a small hydrophilic ligand such as adrenaline can directly approach the adrenaline receptor (AR) from the extracellular region,[2] whereas lipid-type ligands such as sphingosine-1-phosphate are thought to take different pathway, i.e., membrane approach.[3] In membrane approach, the ligand first enters the lipid bilayer (Figure 1, state A), then diffuses laterally in the cell membrane to make hydrophobic contact with the receptor protein (Figure 1, state B), and finally enters the receptor's binding pocket (Figure 1, state C).

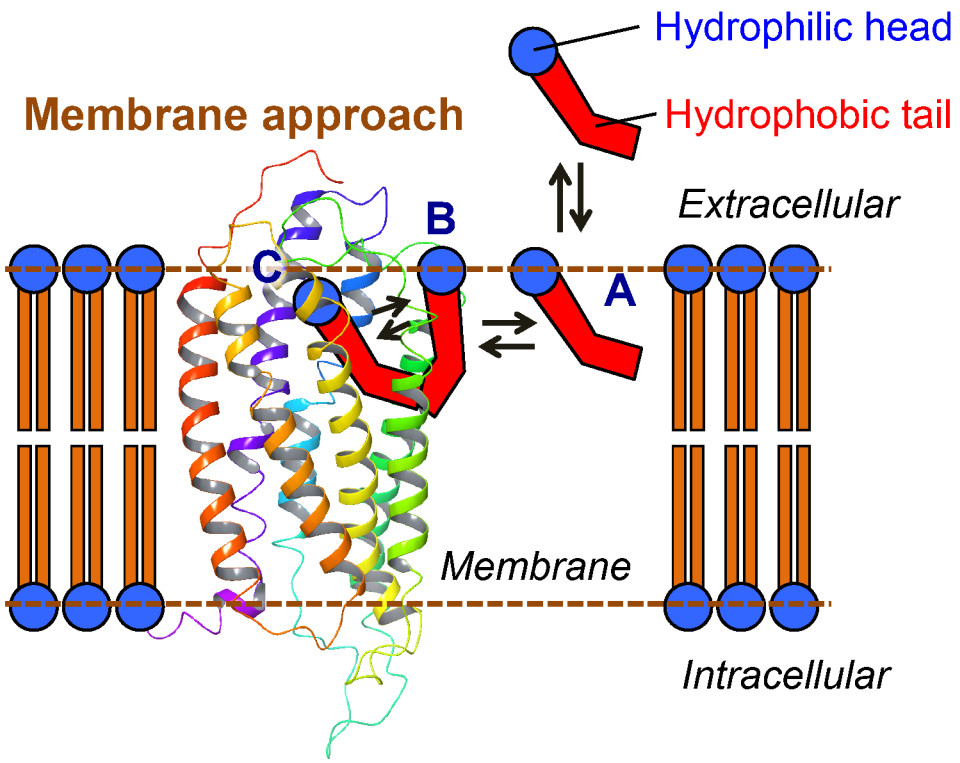


Figure 1. Possible entry/exit pathways of lipid ligands to GPCRs

In the membrane approach pathway, the ligand first enters the lipid bilayer (state A), then diffuses laterally in the cell membrane to make hydrophobic contact with the receptor (state B), and finally enters the receptor's binding pocket (state C).

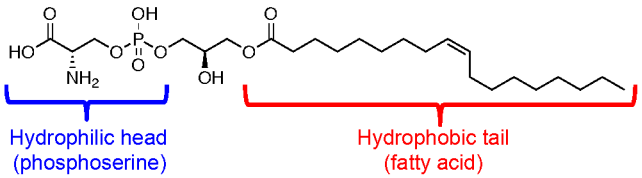
Furthermore, in the crystal structure of free fatty acid receptor GPR40, the hydrophobic part of the bound allosteric agonist protrudes between TMs 3 and 4.[4,5] The crystal structure of zebrafish LPA_6 ($zLPA_6$) also has a hydrophobic crevice between TMs 4 and 5, opening toward the membrane side, which may provide a gateway for entry of the ligand's hydrophobic moiety.[6] Such openings on the receptor surface might be available as an entry point for ligands that enter the membrane and diffuse laterally in it, or accidentally reach the vicinity of the receptor. On the other hand, such openings are relatively extended along the TMs, and therefore, the putative contact position of the hydrophobic moiety of the ligand is unclear. The membrane approach of synthetic small lipidic ligands has been computationally reproduced by

molecular dynamics (MD) simulations in the cases of homology-modeled structures of CB_2[7] and CB_1[8] and the crystal structures of $S1P_1$[3] and EP_4.[9] Such simulation studies support the idea that the first contact position of the hydrophobic moiety is different from the binding position of the hydrophobic moiety in the final binding pose.

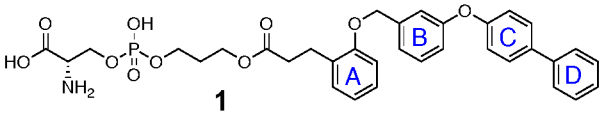
GPR34/LPS_1 is one of the lipid-liganded GPCRs, though its crystal structure has not yet been reported. GPR34 is specifically activated by endogenous lysophosphatidylserines (LysoPSs).[10,11] LysoPS is biosynthesized through hydrolysis of one of the ester linkages of a membrane glycerophospholipid, phosphatidylserine (PS). LysoPS is amphiphilic, being composed of hydrophilic serine, a phosphodiester linkage, glycerol, and a hydrophobic fatty acid moiety (Figure 2a). In the course of ligand design with the aim of creating chemical tools to investigate the function of GPR34, as well as for drug discovery, we found that the ligand activity depends on the structure of the hydrophobic fatty acid moiety, and in particular, we observed potent activation of GPR34 by the LysoPS analogue **1** (Figure 2a), which contains four benzene rings as a fatty acid surrogate.[12] One possible pathway for LysoPS to bind to GPR34 is membrane approach (Figure 1, from state A to C). Before binding to the pocket, the lipid ligand is assumed to make hydrophobic contact on the surface of the receptor protein (Figure 1, state B), because modification of the size and length of the hydrophobic moiety of lipid ligands often dramatically alters the ligand-receptor binding activity. The location of the first hydrophobic contact on the surface of the receptor is difficult to access experimentally, because the binding pocket is buried inside the transmembrane (TM) domains. Therefore, in the present work, we aimed to employ membrane phospholipid surrogates of various chain lengths to constrain the

depth of ligand entry into the lipid bilayer. Then, by measuring the receptor-activating activity as a function of depth of entry into the membrane, these surrogates can be used as molecular rulers to estimate the position (depth) of the hydrophobic contact point on the surface of a lysophospholipid-specific GPCR (Figure 2b). The analogues were synthesized by attaching alkoxy amine chains of varying lengths, as simplified surrogates of membrane phospholipid, to the hydrophobic tail of a potent GPCR agonist **1**. We then carried out an experimental study of the receptor-activating activity of the membrane-lipid-surrogate-conjugated LysoPS analogues, combined with molecular dynamic simulations of the membrane protein system. We found that the activity of these lipid-conjugate LysoPS analogues is indeed length-dependent: conjugation of a short membrane lipid surrogate abolished the receptor-activation activity, while conjugation of a long membrane lipid surrogate retained activity comparable to that of the prototype lysophospholipid analogue. The results suggest that the contact position is deep, that is, near the middle of the membrane bilayer (width ~20 Å), which matches the position of the terminal benzene ring of a receptor-bound non-lipidic ligand that protrudes between TMs 4 and 5. Furthermore we found that the nature of the terminal functional group of the conjugated membrane lipid surrogate strongly influenced the activity of LysoPS analogues, suggesting that lateral association of the LysoPS analogue on the receptor's surface in the lipid-water boundary, probably through hydrophilic contact, is also crucial for ligand-GPCR binding.

a



Lysophosphatidylserine (LysoPS(18:1))



1

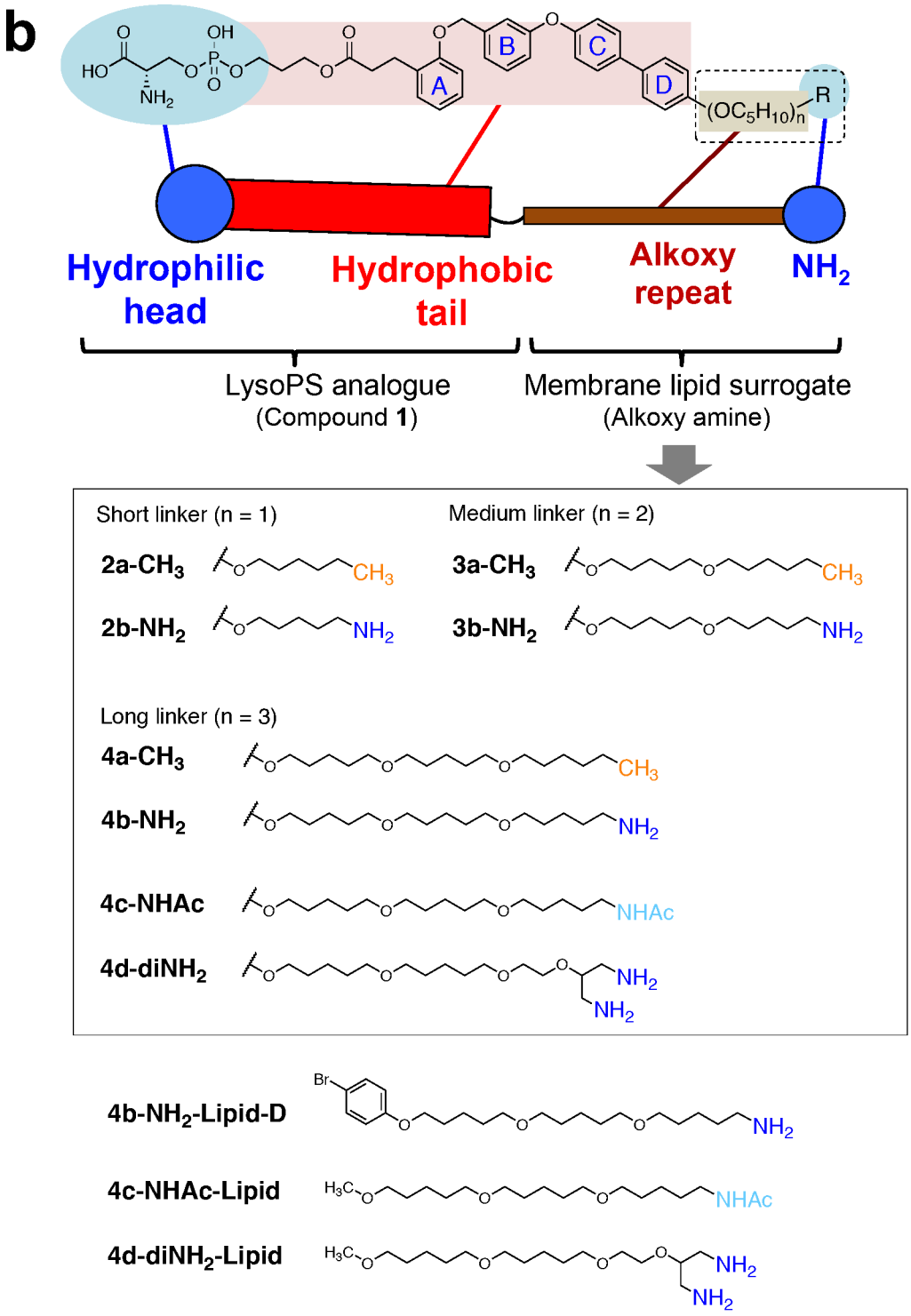


Figure 2. The ligand design concept

(a) Structures of LysoPS (18:1) and LysoPS analogue **1**. (b) General structure of lipid-surrogate-conjugated LysoPS derivatives and chemical structures of the synthesized lipid-surrogate-conjugated LysoPS derivatives

Materials/Experimental Details

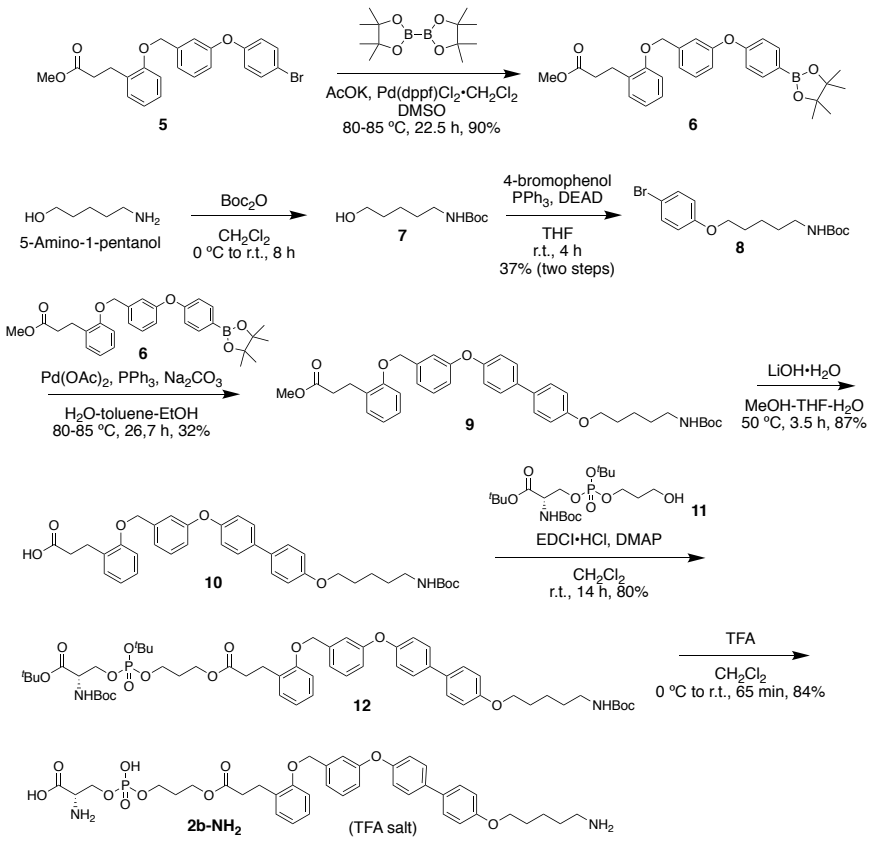
Chemical synthesis

Details of synthesis of compounds **2b-NH_2**, **3a-CH_3**, **3b-NH_2**, **4a-CH_3**, **4b-NH_2**, **4c-NHAc**, **4d-diNH_2** and **4b-NH_2-Lipid-D** (The synthesis of **2a-CH_3** was described previously)[12] are described.

General Synthetic Procedures:

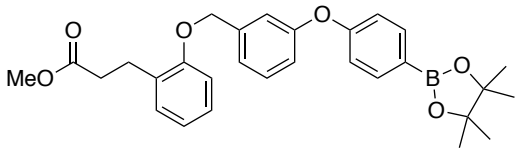
Melting points were determined with a Yanaco micro melting point apparatus without correction. ^{1}H (400 MHz), ^{13}C (100 MHz) and ^{31}P (161 MHz) NMR spectra were recorded on a Bruker Avance400. Chemical shifts were calibrated with tetramethylsilane as an internal standard or with the solvent peak, and are shown in ppm (δ) values; coupling constants are shown in hertz (Hz). The following abbreviations are used: s = singlet, d = doublet, t = triplet, q = quartet, quin = quintet, dd = double doublet, m = multiplet, and brs = broad singlet. An electron spray ionization time-of-flight (ESI-TOF) mass spectrometer (Bruker micrOTOF-05) was used to obtain high-resolution mass spectra (HRMS). All commercially available compounds and solvents were used as received except PPh_3 which was used after recrystallization. Combustion analyses were carried out in the Microanalytical Laboratory of the Graduate School of Pharmaceutical Sciences, the University of Tokyo. All the tested compounds showed ≥95% purity on the basis of combustion analysis. The purity of some of the compounds that contain water in combustion analysis due to their hydroscopic nature were also examined by HPLC analysis, showing at least >88% purity on the basis of reverse phase HPLC analysis. HPLC conditions: Imtakt Unison UK-C18, 3 μm, 150×4.6 mm, acetonitrile: H_2O containing 0.1% HCO_2H as an eluent, detection: 275 nm UV, flow rate 1.0 mL/min.

Synthesis of 2b-NH_2 (Scheme 1)



Scheme 1. Synthesis of 2b-NH_2

Compound **6**

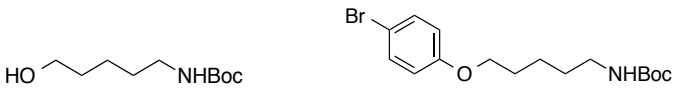


Compound **5** (892.0 mg, 2.021 mmol), bis(pinacolato)diborane (561.7 mg, 2.212 mmol), AcOK (990.0 mg, 10.09 mmol) and Pd(dppf)Cl_2•CH_2Cl_2 (167.0 mg, 0.2045 mmol) were dissolved in DMSO (15 mL) which had been deoxygenated by N_2 bubbling for 90 minutes. The mixture was stirred at 80-85˚C for 22.5 hours under an Ar atmosphere. The mixture was filtered on Celite®. AcOEt (120 mL) was added and the organic layer was washed with water twice (75 mL×2) and brine (50 mL), dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a black oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=4:1)

to yield compound **6** as pale green oil (883.5 mg, 1.809 mmol, 90%), which became a pale green solid after cooling.

^{1}H NMR (400 MHz, $CDCl_3$): 7.785 (2H, d, J = 8.6 Hz), 7.351 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.200-7.145 (3H, m), 7.093 (1H, s), 7.011-6.953 (3H, m), 6.913-6.847 (2H, m), 5.064 (2H, s), 3.630 (3H, s), 2.976 (2H, t, J = 7.7 Hz), 2.616 (2H, t, J = 7.7 Hz), 1.342 (12H, s). ^{13}C NMR (100 MHz, $CDCl_3$): 173.70, 159.84, 156.93, 156.32, 139.36, 136.64, 130.09, 129.95, 129.14, 127.55, 121.98, 120.84, 118.47, 117.90, 117.72, 111.54, 83.73, 69.26, 51.46, 33.99, 26.15, 24.83. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{29}H_{33}BNaO_6^+$: 511.2262. Found: 511.2239. Mp: 65.4-66.2 ˚C (colorless plates, recrystallized from *n*-hexane/CH_2Cl_2).

Compounds **7** and **8**

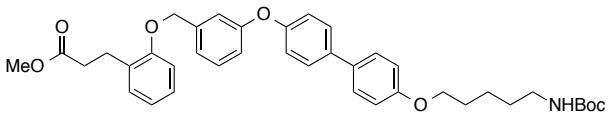


5-Amino-1-pentanol (1024.4 mg, 9.9297 mmol) was dissolved in CH_2Cl_2 (15 mL). To the solution, Boc_2O (2300 μL) was added dropwise at 0 ˚C. The whole was stirred at room temperature for 8 hours. H_2O (10 mL) was added. The organic layer was separated, washed with brine (5 mL), dried over Na_2SO_4, evaporated and vacuum dried to yield compound **7** as a yellow oil (2517.7 mg), which was used without further purification.

To a solution of **7** (2375.3 mg), 4-bromophenol (1970.6 mg, 11.390 mmol) and PPh_3 (2964.0 mg, 11.300 mmol) in anhydrous THF (40 mL) was added DEAD (2M in toluene) (6.0 mL) dropwise at 0 ˚C. Then the mixture was stirred at room temperature under an Ar atmosphere for 4 hours. H_2O (30 mL) was added and the aqueous layer was extracted three times with CH_2Cl_2 (30 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a sticky orange oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=6:1). To remove residual 4-bromophenol, the crude product (2092.3 mg) was dissolved in CH_2Cl_2 (10 mL) and the solution was washed with 2M aqueous NaOH three times (10 mL×3). The organic layer was dried over Na_2SO_4, evaporated and vacuum dried to yield compound **8** as a yellow oil, which later became a colorless solid (1332.4 mg, 3.7189 mmol, 37%, two steps).

^{1}H NMR (400 MHz, $CDCl_3$): 7.347 (2H, d, J = 9.0 Hz), 6.752 (2H, d, J = 9.0 Hz), 4.554 (1H, brs), 3.904 (2H, t, J = 6.4 Hz), 3.141-3.127 (2H, m), 1.778 (2H, quin, J = 6.9 Hz), 1.560-1.435 (13H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.09, 155.96, 132.16, 116.23, 112.62, 79.08, 67.91, 40.41, 29.80, 28.77, 28.38, 23.26. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{16}H_{24}BrNNaO_3^+$: 380.0832. Found: 380.0859. Anal. Calcd. For: $C_{16}H_{24}BrNO_3$: C, 53.64; H, 6.75; N, 3.91. Found: C, 53.59; H, 6.90; N, 3.74. Mp: 40.0-40.7 ˚C (colorless plates, recrystallized from *n*-hexane).

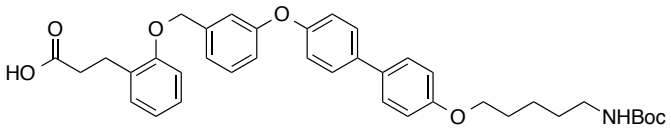
Compound **9**



Compound **6** (297.9 mg, 0.6100 mmol), compound **8** (228.5 mg, 0.6378 mmol), Na_2CO_3 (267.0 mg, 2.519 mmol), PPh_3 (10.4 mg, 0.0397 mmol) and $Pd(OAc)_2$ (5.5 mg, 0.024 mmol) were dissolved in EtOH (1.2 mL), H_2O (1.2 mL) and toluene (2.8 mL). The mixture was degassed by freeze-pump-thaw and stirred at 80-85 ˚C under an Ar atmosphere for 26.7 hours. H_2O (10 mL) was added and the aqueous layer was extracted three times with AcOEt (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a black oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=4:1) to yield compound **9** as a yellow sticky oil (125.3 mg, 0.1958 mmol, 32%), which later became a pale yellow solid.

^{1}H NMR (400 MHz, $CDCl_3$): 7.529-7.466 (4H, m), 7.355 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.192-7.149 (3H, m), 7.110 (1H, s), 7.066 (2H, d, J = 8.7 Hz), 6.992 (1H, dd, J = 8.1 Hz, 1.0 Hz), 6.953 (2H, d, J = 8.8 Hz), 6.914-6.859 (2H, m), 5.073 (2H, s), 4.562 (1H, brs), 3.999 (2H, t, J = 6.4 Hz), 3.618 (3H, s), 3.168-3.138 (2H, m), 2.978 (2H, t, J = 7.7 Hz), 2.614 (2H, t, J = 7.7 Hz), 1.827 (2H, quin, J = 6.8 Hz), 1.612-1.494 (4H, m), 1.450 (9H, s). ^{13}C NMR (100 MHz, $CDCl_3$): 173.70, 158.43, 157.70, 156.34, 155.98, 155.84, 139.29, 136.28, 132.95, 130.11, 129.90, 129.14, 127.98, 127.86, 127.56, 121.55, 120.83, 119.37, 117.88, 117.05, 114.76, 111.55, 79.09, 69.30, 67.77, 51.44, 40.45, 33.99, 29.83, 28.92, 28.40, 26.15, 23.35. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{39}H_{45}NNaO_7^+$: 662.3088. Found: 662.3102. Mp: 68.2-69.2 ˚C (colorless powder, recrystallized from *n*-hexane/CH_2Cl_2).

Compound **10**

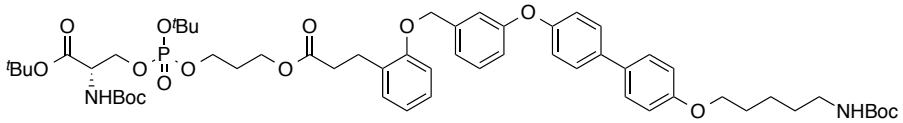


Compound **9** (109.4 mg, 0.1710 mmol) was dissolved in MeOH (0.5 mL), anhydrous THF (0.5 mL) and H_2O (0.5 mL). To the solution was added LiOH•H_2O (22.7 mg, 0.541 mmol). The solution was stirred at 50˚C for 3.5 hours. 2M aqueous HCl was added to acidify the mixture to pH 2. The aqueous layer was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated and vacuum-dried to yield compound **10** as a white powder (93.2 mg, 0.149 mmol, 87%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.517-7.462 (4H, m), 7.343 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.185-7.149 (3H, m), 7.085 (1H, s), 7.056 (2H, d, *J* = 8.7 Hz), 6.986 (1H, dd, *J* = 8.2 Hz, 0.9 Hz), 6.939 (2H, d, *J* = 8.7 Hz), 6.904-6.852 (2H, m), 5.062 (2H, s), 4.556 (1H, s), 3.986 (2H, t, *J* = 6.3 Hz), 3.157-3.143 (2H, m), 2.966 (2H, t, *J* = 7.7 Hz), 2.635 (2H, t, *J* = 7.6 Hz), 1.812 (2H, quin, *J* = 6.8 Hz), 1.597-1.478 (4H, m), 1.445 (9H, s).

^{13}C NMR (100 MHz, $CDCl_3$): 178.31, 158.41, 157.74, 156.33, 156.01, 155.81, 139.22, 136.28, 132.93, 130.11, 129.94, 128.89, 127.97, 127.86, 127.63, 121.53, 120.84, 119.45, 117.89, 116.98, 114.77, 111.54, 79.15, 69.30, 67.76, 40.47, 33.80, 29.80, 28.90, 28.40, 25.87, 23.33. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{38}H_{42}NO_7^-$: 624.2967. Found: 624.2997. Anal. Calcd. For: $C_{38}H_{43}NO_7$•0.05CH_2Cl_2: C, 72.54; H, 6.90; N, 2.22. Found: C, 72.38; H, 7.01; N, 2.35. Mp: 121.0-122.6 ˚C (colorless needles, recrystallized from *n*-hexane/CH_2Cl_2).

Compound **12**

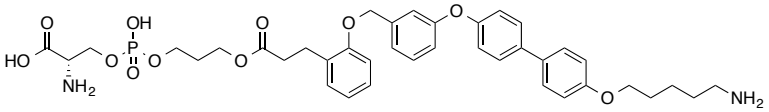


Compound **10** (83.4 mg, 0.133 mmol), compound **11** (52.8 mg, 0.116 mmol) and DMAP (4.5 mg, 0.037 mmol) were dissolved in anhydrous CH_2Cl_2 (1.0 mL). EDCI•HCl (39.8 mg, 0.208 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 14 hours. Anhydrous CH_2Cl_2 (1.0 mL),

EDCI•HCl (17.8 mg) and MeOH (0.4 mL) were added and the reaction mixture was stirred for 1.5 hours. H_2O (10 mL) was added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a colorless sticky oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1 to 0:1) to yield compound **12** as a colorless sticky oil (98.7 mg, 0.0928 mmol, 80%).

^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.529-7.466 (4H, m), 7.354 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.185-7.149 (3H, m), 7.102 (1H, s), 7.062 (2H, d, J = 8.7 Hz), 6.995-6.934 (3H, m), 6.910-6.858 (2H, m), 5.507 (1H, d, J = 7.7 Hz), 5.080 (2H, s), 4.595 (1H, brs), 4.369-4.304 (2H, m), 4.226-4.181 (1H, m), 4.155-4.111 (2H, m), 4.033-3.982 (4H, m), 3.162-3.151 (2H, m), 2.982 (2H, t, J = 7.6 Hz), 2.629 (2H, t, J = 7.6 Hz), 1.964-1.889 (2H, m), 1.827 (2H, quin, J = 6.8 Hz), 1.633-1.388 (40H, m). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.55, -5.70. ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.04, 168.30, 158.41, 157.64, 156.31, 155.95, 155.80, 155.19, 139.28, 136.25, 132.87, 130.08, 129.89, 128.99, 127.94, 127.82, 127.55, 121.49, 120.78, 119.33, 117.79, 117.04, 114.73, 111.56, 83.64, 83.57, 82.62, 82.59, 79.87, 79.02, 69.27, 67.73, 67.36, 64.01, 63.96, 60.37, 60.32, 54.42, 54.33, 40.44, 33.99, 29.81, 29.74, 29.70, 29.50, 29.43, 28.88, 28.37, 28.25, 27.89, 26.07, 23.32. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{57}H_{79}N_2NaO_{15}P^+$: 1085.5110. Found: 1085.5085.

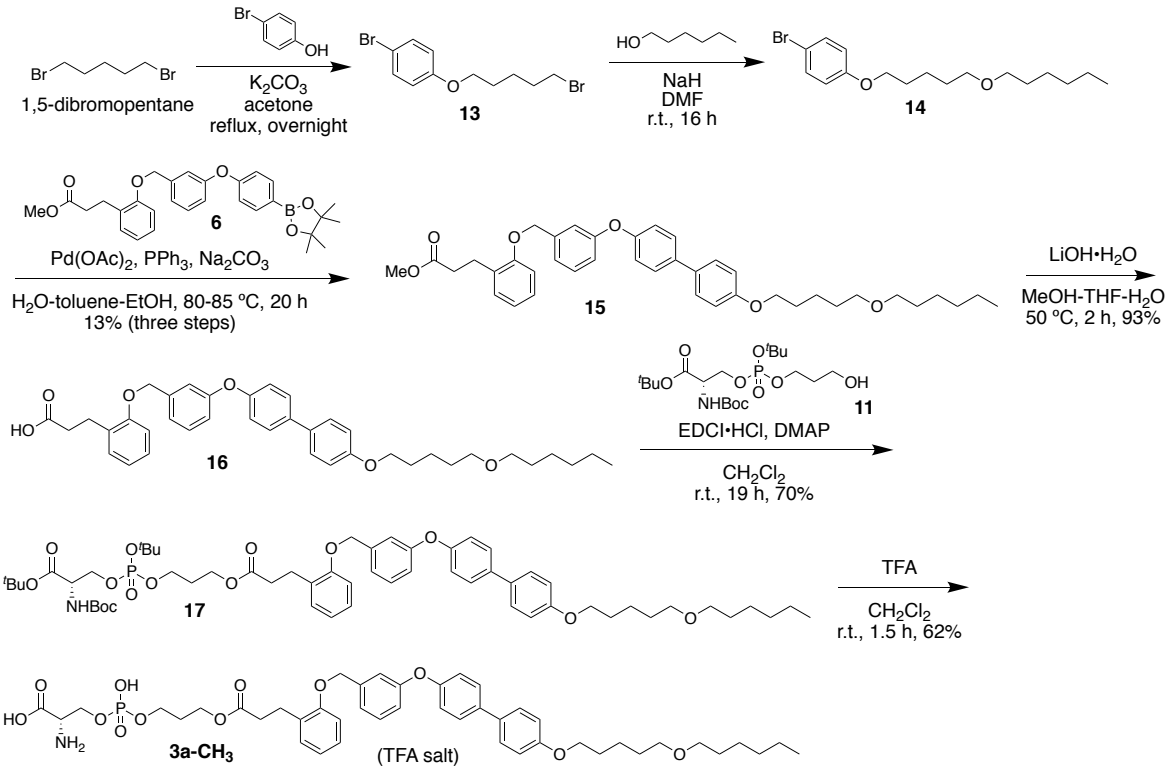
Compound **2b-NH$_2$**



Compound **12** (82.4 mg, 0.0775 mmol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL). TFA (1.0 mL) was added at 0 ˚C. The reaction mixture was stirred at 0 ˚C for 10 minutes and at room temperature for 65 minutes. The mixture was then diluted with CH_2Cl_2 (5 mL) and evaporated. The residue was flash-chromatographed on a column of silica-gel ($CHCl_3$:MeOH:AcOH=8:1:1, 7:1:2, 6:1:3 to 5:1:4). The resulting colorless oil was dissolved in CH_2Cl_2 and TFA, evaporated and vacuum dried at 40 ˚C overnight to yield **2b-NH$_2$** as the TFA-salt (pale brown powder, 48.7 mg, 0.0649 mmol, 84%).

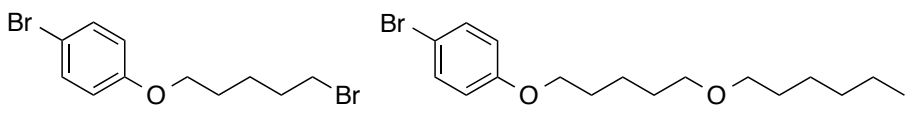
^{1}H NMR (400 MHz, $CDCl_3$/TFA-*d*): 7.680 (2H, brs), 7.532 (4H, brs), 7.384 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.230-7.188 (2H, m), 7.131-7.030 (7H, m), 6.950-6.908 (2H, m), 6.763 (2H, brs), 5.122 (2H, s), 4.678 (2H, brs), 4.575 (1H, brs), 4.190 (4H, brs), 4.069 (2H, brs), 3.226 (2H, brs), 2.983 (2H, t, *J* = 7.0 Hz), 2.733 (2H, t, *J* = 7.0 Hz), 1.957-1.849 (6H, m), 1.668-1.632 (2H, m). HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{39}H_{46}N_2O_{11}P^-$: 749.2845. Found: 749.2869. Anal. Calcd. For: $C_{39}H_{47}N_2O_{11}P\bullet1.9CF_3CO_2H\bullet0.4H_2O$: C, 52.75; H, 5.14; N, 2.87. Found: C, 52.62; H, 5.54; N, 2.84. Hygroscopic. HPLC purity: 89%.

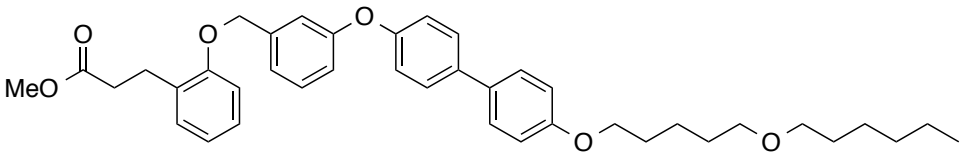
Synthesis of 3a-CH_3 (Scheme 2)



Scheme 2. Synthesis of 3a-CH_3

Compounds **13**, **14** and **15**





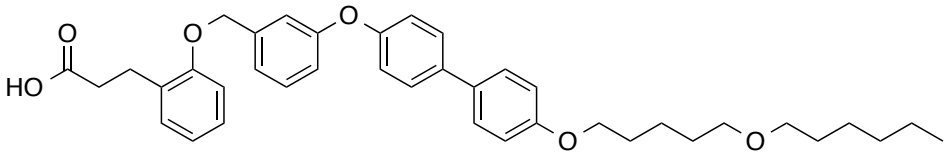
1,5-Dibromopentane (999.5 mg, 4.347 mmol), 4-bromophenol (357.1 mg, 2.064 mmol), K_2CO_3 (288.0 mg, 2.084 mmol) and 18-crown-6 (31.2 mg) were dissolved in acetone (10 mL). The mixture was heated to reflux and stirred for 15.5 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a yellow oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=20:1) to yield crude compound **13** (770.0 mg).

1-Hexanol (412.1 mg, 4.033 mmol) was dissolved in anhydrous DMF (5.0 mL). NaH (60% dispersion in mineral oil, 158.8 mg, 3.970 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 40 minutes. Crude **13** (549.7 mg) in anhydrous DMF (5.0 mL) was added to the solution at 0 ºC and the whole was stirred at room temperature for 16 hours. H_2O (10 mL) was added and the whole was extracted three times with Et_2O (20 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a colorless oil. The residue was chromatographed on an open column of silica-gel twice (first, *n*-hexane:AcOEt=30:1 to 20:1, second, *n*-hexane:AcOEt=30:1) to yield compound **14** as a mixture (220.4 mg).

Crude **14** (177.4 mg), compound **6** (408.4 mg, 0.8362 mmol), PPh_3 (8.6 mg, 0.033 mmol) and $Pd(OAc)_2$ (7.3 mg, 0.033 mmol) were dissolved in EtOH (1.2 mL), toluene (2.4 mL) and 2M aqueous Na_2CO_3 (2.4 mL). The mixture was degassed by freeze-pump-thaw cycles and stirred at 80-85 ˚C under an Ar atmosphere for 20 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a black oil. The residue was chromatographed on an open column of silica-gel three times (first and second, *n*-hexane:AcOEt=7:1, third, *n*-hexane:AcOEt=10:1 to 3:1) to yield compound **15** as a yellow oil (162.4 mg, 0.2599 mmol, 13%, three steps).

^{1}H NMR (400 MHz, $CDCl_3$): 7.527-7.463 (4H, m), 7.353 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.191-7.147 (3H, m), 7.107 (1H, s), 7.063 (2H, d, *J* = 8.7 Hz), 6.990 (1H, dd, *J* = 8.1 Hz, 1.0 Hz), 6.954 (2H, d, *J* = 8.8 Hz), 6.912-6.857 (2H, m), 5.071 (2H, s), 4.003 (2H, t, *J* = 6.5 Hz), 3.616 (3H, s), 3.457-3.391 (4H, m), 2.977 (2H, t, *J* = 7.7 Hz), 2.612 (2H, t, *J* = 7.7 Hz), 1.871-1.801 (2H, m), 1.698-1.508 (6H, m), 1.377-1.255 (6H, m), 0.887 (3H, t, *J* = 6.9 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): 173.71, 158.52, 157.73, 156.36, 155.84, 139.30, 136.33, 132.88, 130.14, 129.93, 129.16, 128.00, 127.87, 127.58, 121.55, 120.84, 119.39, 117.89, 117.06, 114.78, 111.55, 71.07, 70.67, 69.30, 67.93, 51.47, 34.00, 31.72, 29.74, 29.54, 29.14, 26.17, 25.87, 22.81, 22.63, 14.05. HRMS (ESI-TOF, $[M+Na]^+$):Calcd for $C_{40}H_{48}NaO_6^+$, 647.3343. Found: 647.3362. Mp: < 30-32.0 ˚C.

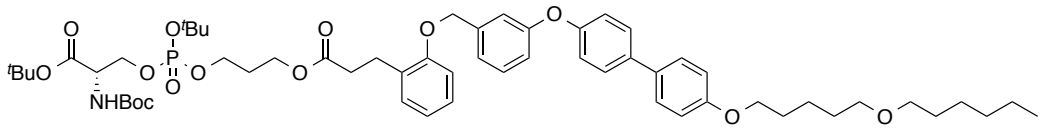
Compound **16**



Compound **15** (152.6 mg, 0.2442 mmol) was dissolved in MeOH (0.8 mL), THF (0.8 mL) and H_2O (0.8 mL). LiOH•H_2O (31.7 mg, 0.755 mmol) was added and the solution was stirred at 50 ºC for 2 hours. 2M aqueous HCl was added to acidify the mixture to pH 2. The whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave compound **16** as a pale yellow solid (139.1 mg, 0.2277 mmol, 93%), which was used without further purification.

^{1}H NMR (400 MHz, $CDCl_3$): 7.518-7.454 (4H, m), 7.343 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.191-7.151 (3H, m), 7.093 (1H, s), 7.056 (2H, d, *J* = 8.7 Hz), 6.985 (1H, dd, *J* = 7.8 Hz, 0.9 Hz), 6.943 (2H, d, *J* = 8.8 Hz), 6.907-6.854 (2H, m), 5.062 (2H, s), 3.990 (2H, t, *J* = 6.5 Hz), 3.454-3.390 (4H, m), 2.969 (2H, t, *J* = 7.7 Hz), 2.644 (2H, t, *J* = 7.7 Hz), 1.860-1.790 (2H, m), 1.693-1.604 (2H, m), 1.587-1.499 (4H, m), 1.373-1.258 (6H, m), 0.884 (3H, t, *J* = 6.7 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): 178.46, 158.49, 157.75, 156.34, 155.81, 139.21, 136.31, 132.86, 130.13, 129.95, 128.82, 127.98, 127.85, 127.67, 121.54, 120.85, 119.44, 117.90, 117.01, 114.77, 111.54, 71.06, 70.66, 69.31, 67.91, 33.75, 31.70, 29.70, 29.50, 29.12, 25.85, 14.05. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{39}H_{45}O_6^-$: 609.3222. Found: 609.3194. Mp: 54.5-55.8 ˚C.

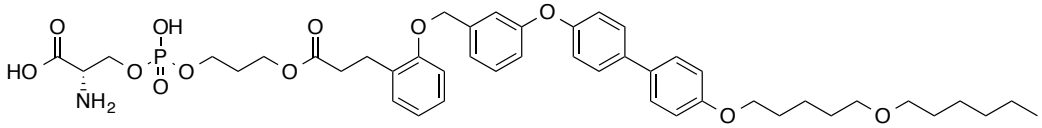
Compound **17**



Compound **16** (72.6 mg, 0.119 mmol), compound **11** (49.8 mg, 0.109 mmol) and DMAP (6.7 mg, 0.055 mmol) were dissolved in anhydrous CH_2Cl_2 (1.0 mL). EDCI•HCl (32.6 mg, 0.170 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 19 hours. EDCI•HCl (14.1 mg), CH_2Cl_2 (1.0 mL) and MeOH (0.5 mL) were added, and the reaction mixture was stirred for 1.5 hours. H_2O (10 mL) was added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a pale yellow oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1 to 0:1) to yield compound **17** as a colorless wax (79.7 mg, 0.0760 mmol, 70%).

^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.527-7.461 (4H, m), 7.351 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.182-7.143 (3H, m), 7.098 (1H, dd, J = 1.8 Hz, 1.8 Hz), 7.058 (2H, d, J = 8.8 Hz), 6.992-6.934 (3H, m), 6.907-6.852 (2H, m), 5.499 (1H, d, J = 8.4 Hz), 5.078 (2H, s), 4.366-4.300 (2H, m), 4.222-4.182 (1H, m), 4.152-4.110 (2H, m), 4.033-3.985 (4H, m), 3.456-3.390 (4H, m), 2.980 (2H, t, J = 7.6 Hz), 2.647-2.605 (2H, m), 1.961-1.887 (2H, m), 1.870-1.800 (2H, m), 1.699-1.619 (2H, m), 1.605-1.525 (4H, m), 1.473-1.431 (27H, m), 1.360-1.256 (6H, m), 0.886 (3H, t, J = 6.9 Hz). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.55, -5.69. ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.07, 168.33, 158.50, 157.67, 156.34, 155.81, 155.22, 139.31, 136.31, 132.83, 130.11, 129.92, 129.02, 127.97, 127.84, 127.58, 121.51, 120.81, 119.36, 117.81, 117.07, 114.76, 111.58, 83.67, 83.60, 82.64, 82.62, 79.90, 71.04, 70.64, 69.30, 67.91, 67.39, 64.04, 63.98, 60.40, 54.44, 54.37, 34.02, 31.69, 29.76, 29.75, 29.71, 29.51, 29.46, 29.11, 28.28, 27.92, 26.11, 25.85, 22.79, 22.60, 14.03 HRMS (ESI-TOF, $[M+Na]^+$):Calcd for $C_{58}H_{82}NNaO_{14}P^+$, 1070.5365. Found: 1070.5356.

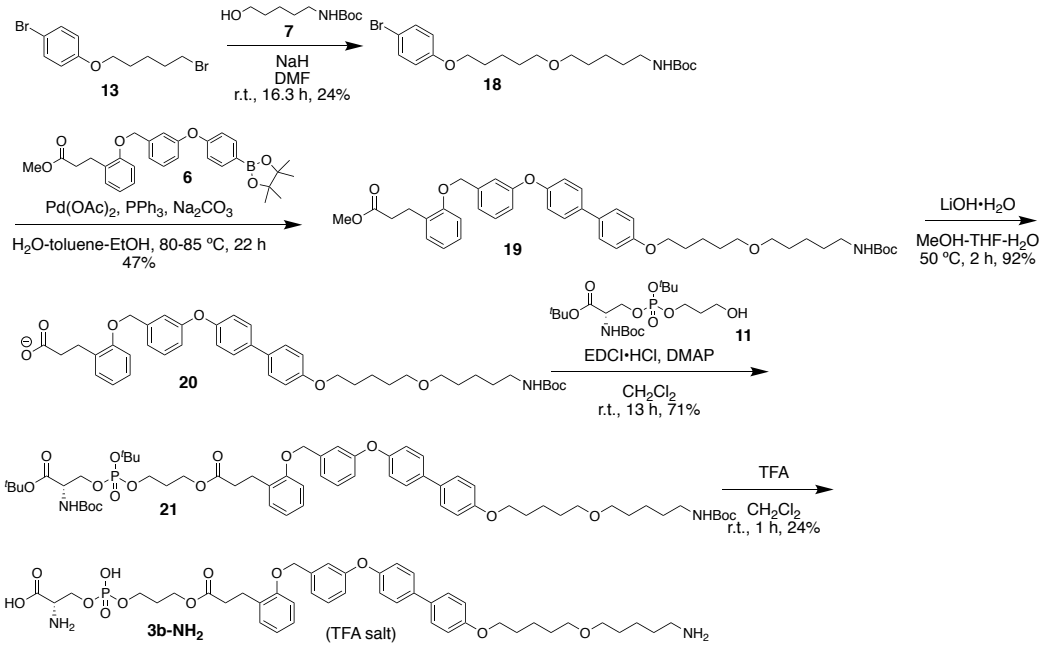
Compound **3a-CH$_3$**



Compound **17** (72.5 mg, 0.0691 mmol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL) and TFA (1.0 mL) at 0 ˚C and the reaction mixture was stirred at 0 ˚C for 5 minutes and at room temperature for 90 minutes. The mixture was then diluted with CH_2Cl_2 (5 mL) and evaporated. The residue was flash-chromatographed on a column of silica-gel ($CHCl_3$:MeOH:AcOH=8:1:1 to 7:1:2). The resulting colorless oil was dissolved in CH_2Cl_2 and TFA and vacuum dried at 40 ˚C overnight to yield **3a-CH$_3$** as the TFA-salt (colorless powder, 35.6 mg, 0.0426 mmol, 62%).

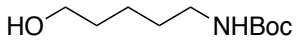
^{1}H NMR (400 MHz, $CDCl_3$/TFA): 7.519-7.488 (4H, m), 7.348 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.204-7.146 (2H, m), 7.100-7.040 (4H, m), 7.012-6.991 (3H, m), 6.902-6.882 (2H, m), 5.084 (2H, s), 4.556 (2H, brs), 4.440 (1H, brs), 4.169 (2H, brs), 4.094 (2H, t, *J* = 6.3 Hz), 3.981 (2H, brs), 3.697-3.632 (4H, m), 2.960 (2H, t, *J* = 7.2 Hz), 2.710 (2H, t, *J* = 7.1 Hz), 1.910 (2H, brs), 1.882-1.811 (2H, m), 1.778-1.693 (2H, m), 1.669-1.601 (2H, m), 1.572-1.487 (2H, m), 1.345-1.270 (6H, m), 0.887 (3H, t, *J* = 6.7 Hz). ^{31}P NMR (161 MHz, $CDCl_3$/TFA): -1.94. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{45}H_{57}NO_{12}P^-$: 834.3624. Found: 834.3650. Anal. Calcd. For: $C_{45}H_{58}NO_{12}P\bullet 0.9CF_3CO_2H$: C, 59.89; H, 6.33; N, 1.49. Found: C, 60.04; H, 6.51; N, 1.62.

Synthesis of 3b-NH$_2$ (Scheme 3)



Scheme 3. Synthesis of 3b-NH_2

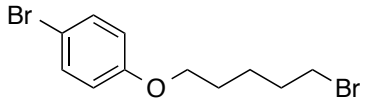
Compound **7**



5-Amino-1-pentanol (2085.8 mg, 20.218 mmol) was dissolved in CH_2Cl_2 (50 mL). Boc_2O (4.8 mL) was added at 0 ºC and the mixture was stirred at room temperature for 1.8 hours. H_2O (70 mL) was added. The aqueous layer was separated and extracted twice with CH_2Cl_2 (50 mL×2). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=4:1 to CH_2Cl_2:MeOH=10:1) to afford compound **7** as a colorless oil (3885.1 mg, 19.112 mmol, 95%).

^{1}H NMR (400 MHz, $CDCl_3$): 3.642 (2H, t, J = 6.5 Hz), 3.121 (2H, t, J = 7.0 Hz), 1.624-1.529 (2H, m), 1.514-1.476 (2H, m), 1.442 (9H, s), 1.419-1.372 (2H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 156.09, 79.16, 62.57, 40.50, 32.19, 29.80, 28.37, 22.86. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{10}H_{21}NNaO_3^+$: 226.1414. Found: 226.1419.

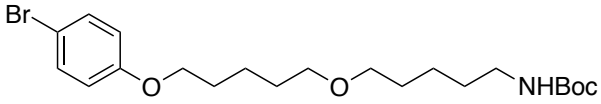
Compound **13**



1,5-Dibromopentane (4601.7 mg, 20.012 mmol), 4-bromophenol (872.8 mg, 5.045 mmol), K_2CO_3 (830.5 mg, 6.009 mmol) and 18-crown-6 (106.6 mg) were dissolved in acetone (35 mL). The mixture was heated to reflux and stirred for 2 hours. H_2O (30 mL) was added and the whole was extracted three times with AcOEt (20 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a pale yellow oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=1:0 to 40:1) to yield compound **13** as a colorless solid which easily melted to give a colorless oil at room temperature (melting point < 34 ºC) (1456.7 mg, 4.5234 mmol, 90%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.362 (2H, d, J = 8.9 Hz), 6.767 (2H, d, J = 9.0 Hz), 3.932 (2H, t, J = 6.3 Hz), 3.438 (2H, t, J = 6.7 Hz), 1.935 (2H, quin, J = 7.2 Hz), 1.840-1.770 (2H, m), 1.656-1.580 (2H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.05, 132.22, 116.25, 112.73, 67.80, 33.52, 32.42, 28.33, 24.77.

Compound **18**

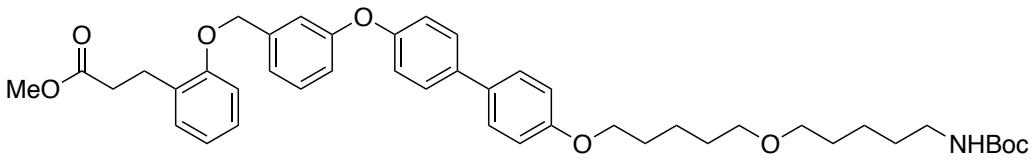


Compound **7** (124.9 mg, 0.6144 mmol) was dissolved in anhydrous DMF (1.5 mL). NaH (60% dispersion in mineral oil, 28.0 mg, 0.700 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 20 minutes. Compound **13** (160.1 mg, 0.4971 mmol) in anhydrous DMF (1.5 mL) was added to the solution at 0 ºC and the whole was stirred at room temperature for 16.3 hours. H_2O (10 mL) was added and the whole was extracted three times with Et_2O (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=5:1) to yield compound **18** as a colorless oil (52.8 mg, 0.119 mmol, 24%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.354 (2H, d, J = 9.0 Hz), 6.765 (2H, d, J = 9.0 Hz), 4.517 (1H, brs), 3.921 (2H, t, J = 6.5 Hz), 3.433-3.381 (4H, m), 3.136-3.088 (2H, m), 1.794 (2H, quin, J = 7.0 Hz), 1.668-1.475 (8H, m), 1.437 (9H, s), 1.405-1.327 (2H,

m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.15, 155.93, 132.14, 116.24, 112.54, 78.97, 70.69, 70.64, 68.04, 40.98, 40.50, 29.86, 29.43, 29.35, 28.96, 28.39, 23.47, 22.72. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{21}H_{34}BrNNaO_4^+$, 466.1563. Found: 466.1569.

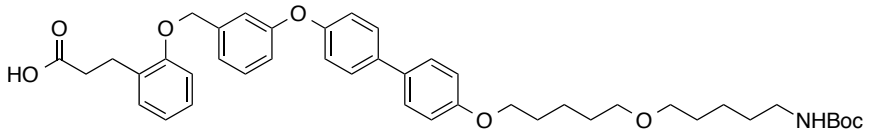
Compound **19**



Compound **18** (49.7 mg, 0.112 mmol), compound **6** (107.6 mg, 0.2203 mmol), PPh_3 (3.1 mg, 0.012 mmol) and $Pd(OAc)_2$ (1.4 mg, 0.0062 mmol) were dissolved in EtOH (0.4 mL), toluene (0.6 mL) and 2M aqueous Na_2CO_3 (0.6 mL). The mixture was degassed by freeze-pump-thaw cycles and stirred at 85 ˚C under an Ar atmosphere for 22 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to leave a grey oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=4:1 to 3:1) to yield compound **19** as a colorless oil (37.8 mg, 0.0521 mmol, 47%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.511 (2H, d, J = 8.7 Hz), 7.484 (2H, d, J = 8.8 Hz), 7.354 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.192-7.148 (3H, m), 7.109 (1H, brs), 7.065 (2H, d, J = 8.7 Hz), 7.004-6.938 (3H, m), 6.913-6.858 (2H, m), 5.073 (2H, s), 4.005 (2H, t, J = 6.5 Hz), 3.617 (3H, s), 3.451-3.394 (4H, m), 3.139-3.094 (2H, m), 2.978 (2H, t, J = 7.7 Hz), 2.614 (2H, t, J = 7.7 Hz), 1.835 (2H, quin, J = 7.0 Hz), 1.695-1.337 (19H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 173.69, 158.49, 157.70, 156.34, 155.93, 155.82, 139.28, 136.30, 132.87, 130.11, 129.90, 129.13, 127.97, 127.85, 127.56, 121.54, 120.82, 119.37, 117.87, 117.04, 114.76, 111.53, 79.01, 70.72, 69.28, 67.89, 51.45, 40.52, 33.98, 29.87, 29.49, 29.38, 29.11, 28.40, 26.15, 23.49, 22.79. HRMS (ESI-TOF, $[M+Na]^+$):Calcd for $C_{44}H_{55}NNaO_8^+$, 748.3820. Found: 748.3849.

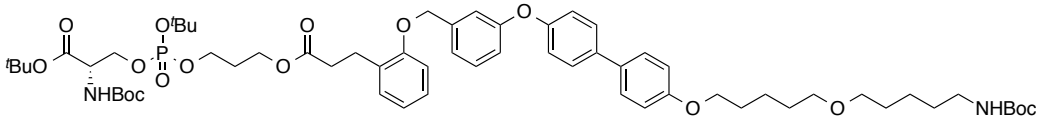
Compound **20**



Compound **19** (68.7 mg, 0.0946 mmol) was dissolved in MeOH (1.0 mL), THF (1.0 mL) and H_2O (1.0 mL). LiOH•H_2O (12.5 mg, 0.298 mmol) was added and the solution was stirred at 50 ºC for 3.8 hours. 2M aqueous solution of HCl was added to acidify the mixture to pH 2. The whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel ($CHCl_3$:MeOH=20:1) to afford compound **20** as a colorless oil (56.2 mg, 0.0789 mmol, 83%).

^{1}H NMR (400 MHz, DMSO-*d*6): 12.087 (1H, brs), 7.622 (2H, d, *J* = 8.8 Hz), 7.563 (2H, d, *J* = 8.8 Hz), 7.420 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.233 (1H, d, *J* = 7.7 Hz), 7.186-7.121 (3H, m), 7.076 (2H, d, *J* = 8.8 Hz), 7.009-6.987 (4H, m), 6.883-6.844 (1H, m), 6.756 (1H, t, *J* = 5.2 Hz), 5.133 (2H, s), 3.999 (2H, t, *J* = 6.5 Hz), 3.377-3.323 (4H, m), 2.887 (2H, q, *J* = 6.5 Hz), 2.795 (2H, t, *J* = 7.7 Hz), 2.454 (2H, t, *J* = 7.7 Hz), 1.738 (2H, quin, *J* = 7.0 Hz), 1.593-1.524 (2H, m), 1.505-1.435 (4H, m), 1.401-1.311 (11H, m), 1.291-1.201 (2H, m).^{13}C NMR (100 MHz, $CDCl_3$): 178.00, 158.46, 157.73, 156.33, 155.79, 139.21, 136.29, 132.87, 130.10, 129.92, 128.92, 127.96, 127.84, 127.60, 121.52, 120.82, 119.43, 117.89, 116.97, 114.77, 111.53, 79.06, 70.69, 69.29, 67.87, 40.50, 33.80, 29.83, 29.66, 29.43, 29.34, 29.07, 28.39, 25.89, 23.46, 22.76. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{43}H_{52}NO_8^-$: 710.3698. Found: 710.3716.

Compound **21**

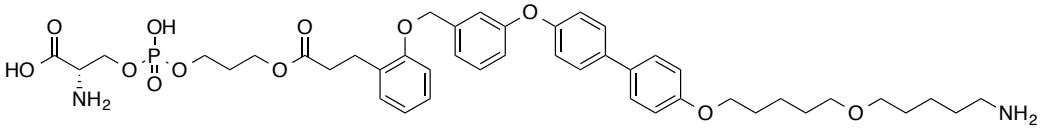


Compound **20** (30.5 mg, 0.0428 mmol), compound **11** (17.3 mg, 0.0380 mmol) and DMAP (1.9 mg, 0.016 mmol) were dissolved in anhydrous CH_2Cl_2 (0.5 mL). EDCI•HCl (16.2 mg, 0.0845 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 13 hours. EDCI•HCl (7.3 mg) and MeOH (0.2 mL) were added the reaction mixture was stirred for 1.5 hours. H_2O (10 mL) was

added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1 to 1:2) to yield compound **21** as a colorless oil (30.9 mg, 0.0269 mmol, 71%).

^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.508 (2H, d, *J* = 8.7 Hz), 7.480 (2H, d, *J* = 8.7 Hz), 7.350 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.182-7.146 (3H, m), 7.096 (1H, brs), 7.057 (2H, d, *J* = 8.7 Hz), 6.989-6.935 (3H, m), 6.906-6.854 (2H, m), 5.495 (1H, d, *J* = 7.9 Hz), 5.077 (2H, s), 4.537 (1H, brs), 4.364-4.299 (2H, m), 4.229-4.174 (1H, m), 4.151-4.109 (2H, m), 4.028-3.968 (4H, m), 3.448-3.391 (4H, m), 3.137-3.090 (2H, m), 2.979 (2H, t, *J* = 7.6 Hz), 2.646-2.605 (2H, m), 1.962-1.884 (2H, m), 1.867-1.797 (2H, m), 1.692-1.360 (46H, m). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.55, -5.70. ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.07, 168.33, 158.49, 157.66, 156.33, 155.81, 155.21, 139.31, 136.30, 132.83, 130.10, 129.92, 129.01, 127.97, 127.84, 127.58, 121.50, 120.80, 119.36, 117.80, 117.07, 114.75, 111.57, 83.67, 82.65, 82.62, 79.90, 78.98, 70.71, 69.30, 67.88, 67.39, 64.03, 63.97, 60.40, 54.43, 54.35, 40.51, 34.01, 29.87, 29.76, 29.72, 29.49, 29.37, 29.11, 28.39, 28.28, 27.91, 26.10, 23.49, 22.78. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{62}H_{89}N_2NaO_{16}P^+$, 1171.5842. Found: 1171.5852.

Compound **3b-NH$_2$**

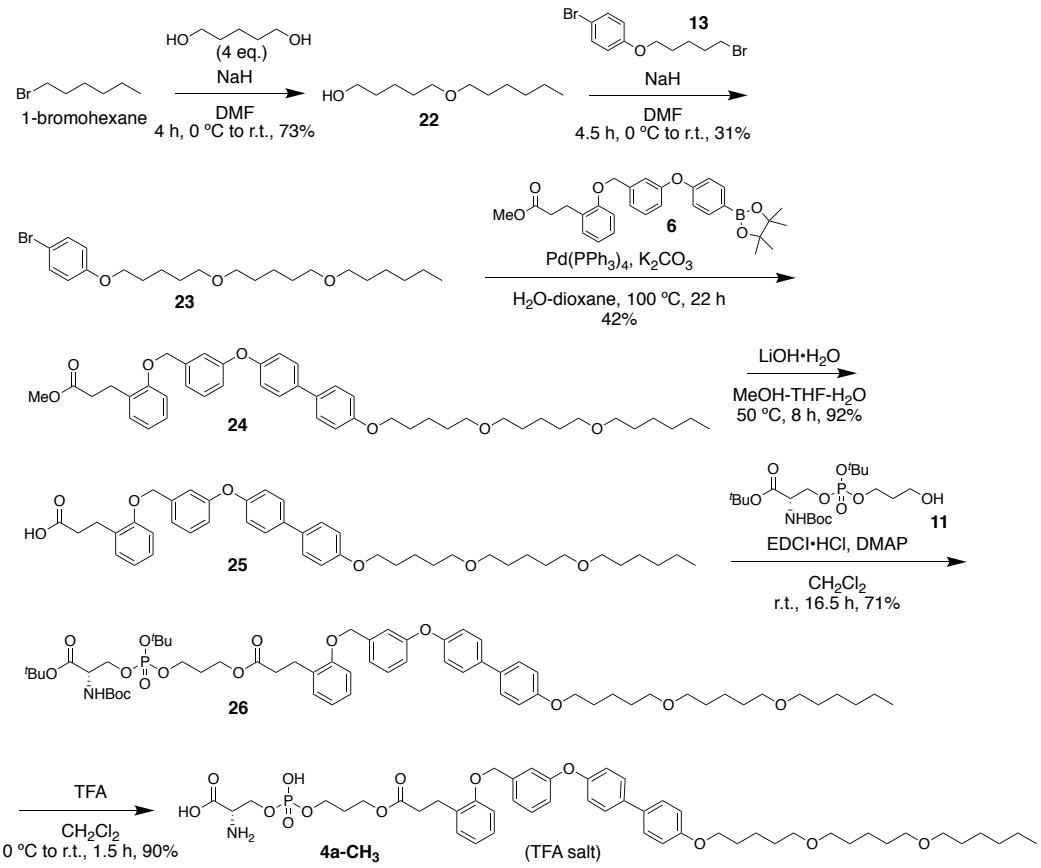


Compound **21** (30.2 mg, 0.0263 mmol) was dissolved in anhydrous CH_2Cl_2 (0.2 mL) and TFA (0.5 mL) at 0 ˚C and the reaction mixture was stirred at 0 ˚C for 10 minutes and at room temperature for 1 hour. The mixture was then diluted with CH_2Cl_2 (5 mL) and evaporated. The residue was flash-chromatographed on a column of silica-gel ($CHCl_3$:MeOH:AcOH=7:2:1, 6:3:1 to 5:4:1). The resulting colorless oil was dissolved in $CDCl_3$ and TFA and vacuum dried at 40 ˚C overnight to yield **3b-NH$_2$** as the TFA-salt (pale yellow powder, 5.3 mg, 0.0063 mmol, 24%).

^{1}H NMR (400 MHz, $CDCl_3$/TFA): 7.677 (2H, brs), 7.512-7.493 (4H, m), 7.365 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.208-7.147 (2H, m), 7.110-6.981 (7H, m), 6.923-6.888 (2H,

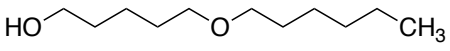
m), 6.711 (3H, brs), 5.097 (2H, s), 4.626-4.523 (3H, m), 4.148-4.137 (4H, m), 4.024 (2H, brs), 3.692-3.658 (4H, m), 3.163 (2H, brs), 2.953 (2H, brs), 2.704 (2H, brs), 1.917 (2H, brs), 1.846 (2H, quin, J = 7.0 Hz), 1.760-1.703 (6H, m), 1.546-1.450 (4H, m). HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{44}H_{56}N_2O_{12}P^-$: 835.3576. Found: 835.3589. Anal. Calcd. For: $C_{44}H_{57}N_2O_{12}P•2CF_3CO_2H•H_2O$: C, 53.24; H, 5.68; N, 2.59. Found: C, 53.64; H, 6.00; N, 2.49. Hygroscopic. HPLC purity: 92%.

Synthesis of 4a-CH_3 (Scheme 4)



Scheme 4. Synthesis of 4a-CH_3

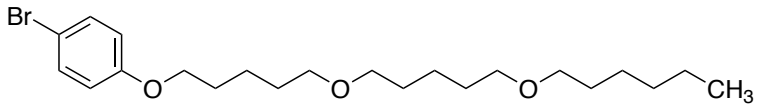
Compound **22**



Pentane-1,5-diol (2082.6 mg, 19.996 mmol) was dissolved in anhydrous DMF (10 mL). NaH (60% dispersion in mineral oil, 302.8 mg, 7.571 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 5 minutes. 1-Bromohexane (826.1 mg, 5.004 mmol) in anhydrous DMF (5.0 mL) was added to the solution at 0 ºC and the whole was stirred at room temperature for 4 hours. H_2O (15 mL) was added and the whole was extracted three times with Et_2O (20 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=1:0 to 1:1) to afford compound **22** as a colorless oil (690.0 mg, 3.664 mmol, 73%).

^{1}H NMR (400 MHz, $CDCl_3$): 3.650 (2H, t, J = 6.5 Hz), 3.432-3.381 (4H, m), 1.639-1.528 (7H, m), 1.469-1.408 (2H, m), 1.351-1.298 (6H, m), 0.887 (3H, t, J = 6.9 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): 71.05, 70.73, 62.81, 32.48, 31.69, 29.68, 29.41, 25.83, 22.60, 22.43, 14.03. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{11}H_{24}NaO_2^+$: 211.1669. Found: 211.1669.

Compound **23**

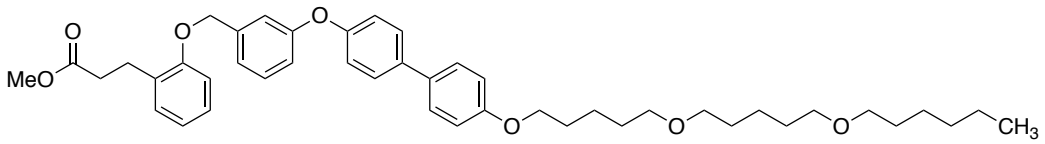


Compound **22** (288.6 mg, 1.533 mmol) was dissolved in anhydrous DMF (3.0 mL). NaH (60% dispersion in mineral oil, 68.7 mg, 1.72 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 15 minutes. Compound **13** (579.2 mg, 1.799 mmol) in anhydrous DMF (3.0 mL) was added to the solution at 0 ºC and the whole was stirred at room temperature for 4.5 hours. H_2O (10 mL) was added and the whole was extracted three times with Et_2O (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=40:1 to 10:1) to afford compound **23** as a colorless oil (203.2 mg, 0.4732 mmol, 31%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.352 (2H, d, J = 9.0 Hz), 6.762 (2H, d, J = 9.0 Hz), 3.917 (2H, t, J = 6.5 Hz), 3.435-3.366 (8H, m), 1.792 (2H, quin, J = 7.1 Hz), 1.668-1.486 (10H, m), 1.432-1.370 (2H, m), 1.361-1.258 (6H, m), 0.882 (3H, t, J = 6.9 Hz).

^{13}C NMR (100 MHz, $CDCl_3$): 158.17, 132.17, 116.26, 112.57, 71.01, 70.88, 70.75, 70.65, 68.07, 31.71, 29.72, 29.59, 29.58, 29.47, 28.99, 25.86, 22.83, 22.74, 22.62, 14.05. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{22}H_{37}BrNaO_3^+$: 451.1818. Found: 451.1824.

Compound **24**



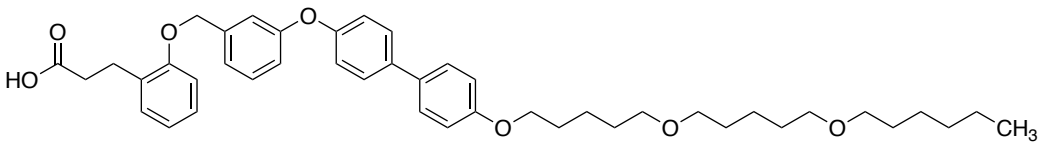
Compound **23** (170.4 mg, 0.3968 mmol), compound **6** (203.3 mg, 0.4163 mmol), K_2CO_3 (88.4 mg, 0.640 mmol) and $Pd(PPh_3)_4$ (23.2 mg, 0.0201 mmol) were dissolved in dioxane (5.0 mL) and H_2O (1.0 mL). The mixture was degassed by freeze-pump-thaw and stirred at 100 ºC under an Ar atmosphere for 22 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a black oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=10:1, 4:1 to 2:1) to afford compound **24** (70.4 mg, 0.0990 mmol) and crude **25** (89.7 mg).

The crude **25** and DMAP (7.7 mg, 0.063 mmol) were dissolved in CH_2Cl_2 (1.0 mL). MeOH (1.0 mL) and EDCI•HCl (67.5 mg, 0.352 mmol) were added and the whole was stirred at room temperature under an Ar atmosphere for 6.3 hours. H_2O (10 mL) was added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a pale brown oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=10:1 to 5:1) to afford compound **24** as a colorless oil (46.9 mg, 0.0660 mmol). The total yield of compound **24** is 42% (117.3 mg, 0.1650 mmol).

^{1}H NMR (400 MHz, $CDCl_3$): 7.508 (2H, d, J = 8.7 Hz), 7.480 (2H, d, J = 8.8 Hz), 7.353 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.191-7.147 (3H, m), 7.107 (1H, brs), 7.062 (2H, d, J = 8.7 Hz), 6.988 (1H, dd, J = 8.0 Hz, 1.3 Hz), 6.953 (2H, d, J = 8.8 Hz), 6.912-6.857 (2H, m), 5.072 (2H, s), 4.001 (2H, t, J = 6.5 Hz), 3.616 (3H, s), 3.453-3.369 (8H, m), 2.976 (2H, t, J = 7.7 Hz), 2.612 (2H, t, J = 7.7 Hz), 1.833 (2H, quin, J = 7.0 Hz), 1.695-1.520 (10H, m), 1.439-1.268 (8H, m), 0.880 (3H, t, J = 6.9 Hz). ^{13}C NMR

(100 MHz, $CDCl_3$): 173.71, 158.50, 157.71, 156.34, 155.83, 139.29, 136.31, 132.87, 130.12, 129.91, 129.14, 127.98, 127.85, 127.57, 121.54, 120.83, 119.38, 117.88, 117.05, 114.76, 111.53, 71.00, 70.87, 70.76, 70.70, 69.29, 67.91, 51.46, 33.99, 31.70, 29.72, 29.58, 29.52, 29.12, 26.16, 25.85, 22.82, 22.79, 22.61, 14.04. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{45}H_{58}NaO_7^+$: 733.4075. Found: 733.4057.

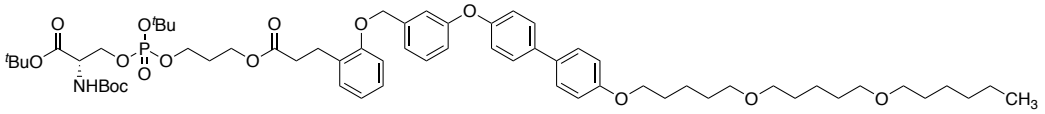
Compound **25**



Compound **24** (103.4 mg, 0.1454 mmol) was dissolved in MeOH (0.5 mL), THF (0.5 mL) and H_2O (0.5 mL). LiOH•H_2O (19.8 mg, 0.472 mmol) was added and the solution was stirred at 50 ºC for 8 hours. H_2O (1 mL) and 2M aqueous solution of HCl (1 mL) werxe added to acidify the mixture to pH 2. The whole was extracted three times with AcOEt (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a white solid. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1 to CH_2Cl_2:MeOH=10:1) to afford compound **25** as a colorless oil (93.4 mg, 0.134 mmol, 92%), which later solidified.

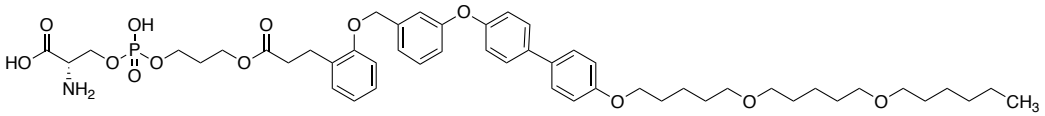
^{1}H NMR (400 MHz, $CDCl_3$): 7.502 (2H, d, *J* = 8.7 Hz), 7.474 (2H, d, *J* = 8.7 Hz), 7.345 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.187-7.152 (3H, m), 7.089 (1H, brs), 7.058 (2H, d, *J* = 8.7 Hz), 6.986 (1H, dd, *J* = 8.1 Hz, 0.9 Hz), 6.944 (2H, d, *J* = 8.8 Hz), 6.908-6.854 (2H, m), 5.063 (2H, s), 3.990 (2H, t, *J* = 6.5 Hz), 3.451-3.372 (8H, m), 2.969 (2H, t, *J* = 7.7 Hz), 2.642 (2H, t, *J* = 7.7 Hz), 1.824 (2H, quin, *J* = 7.0 Hz), 1.689-1.499 (10H, m), 1.438-1.257 (8H, m), 0.879 (3H, t, *J* = 6.9 Hz). ^{13}C NMR (100 MHz, $CDCl_3$): 178.12, 158.48, 157.74, 156.34, 155.81, 139.21, 136.31, 132.86, 130.13, 129.94, 128.84, 127.97, 127.85, 127.65, 121.53, 120.84, 119.44, 117.90, 117.00, 114.77, 111.54, 70.99, 70.86, 70.75, 70.69, 69.30, 67.89, 33.71, 31.69, 29.69, 29.55, 29.48, 29.11, 25.87, 25.84, 22.80, 22.77, 22.61, 14.04. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{44}H_{55}O_7^-$: 695.3953. Found: 695.3956. Mp: 38.3-42.0 ˚C.

Compound **26**



Compound **25** (78.2 mg, 0.112 mmol), compound **11** (48.9 mg, 0.107 mmol) and DMAP (6.6 mg, 0.054 mmol) were dissolved in anhydrous CH_2Cl_2 (1.0 mL). EDCI•HCl (37.4 mg, 0.195 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 16.5 hours. Further CH_2Cl_2 (0.5 mL), EDCI•HCl (10.2 mg) and MeOH (0.4 mL) were added and the reaction mixture was stirred for 1.5 hours. H_2O (10 mL) was added and the whole was extracted three times with CH_2Cl_2 (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=3:1 to 1:1) to afford compound **26** as a colorless sticky oil (86.9 mg, 0.0766 mmol, 71%). ^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.511 (2H, d, J = 8.7 Hz), 7.482 (2H, d, J = 8.8 Hz), 7.353 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.185-7.149 (3H, m), 7.100 (1H, brs), 7.060 (2H, d, J = 8.7 Hz), 6.992-6.944 (3H, m), 6.909-6.854 (2H, m), 5.501 (1H, d, J = 7.9 Hz), 5.081 (2H, s), 4.367-4.302 (2H, m), 4.222-4.177 (1H, m), 4.155-4.112 (2H, m), 4.032-3.986 (4H, m), 3.455-3.371 (8H, m), 2.982 (2H, t, J = 7.6 Hz), 2.650-2.612 (2H, m), 1.964-1.889 (2H, m), 1.870-1.799 (2H, m), 1.695-1.382 (39H, m), 1.364-1.270 (6H, m), 0.882 (3H, t, J = 6.9 Hz). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.55, -5.71. ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.09, 168.33, 158.50, 157.67, 156.34, 155.81, 155.22, 139.31, 136.31, 132.83, 130.11, 129.92, 129.02, 127.97, 127.84, 127.58, 121.51, 120.81, 119.36, 117.81, 117.08, 114.76, 111.58, 83.68, 82.66, 82.63, 79.91, 70.98, 70.87, 70.74, 70.69, 69.30, 67.90, 67.41, 64.04, 63.99, 60.41, 54.44, 34.02, 31.69, 29.77, 29.73, 29.71, 29.57, 29.51, 29.46, 29.12, 28.28, 27.92, 26.11, 25.84, 22.81, 22.78, 22.60, 14.03. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{63}H_{92}NNaO_{15}P^+$: 1156.6097. Found: 1156.6096.

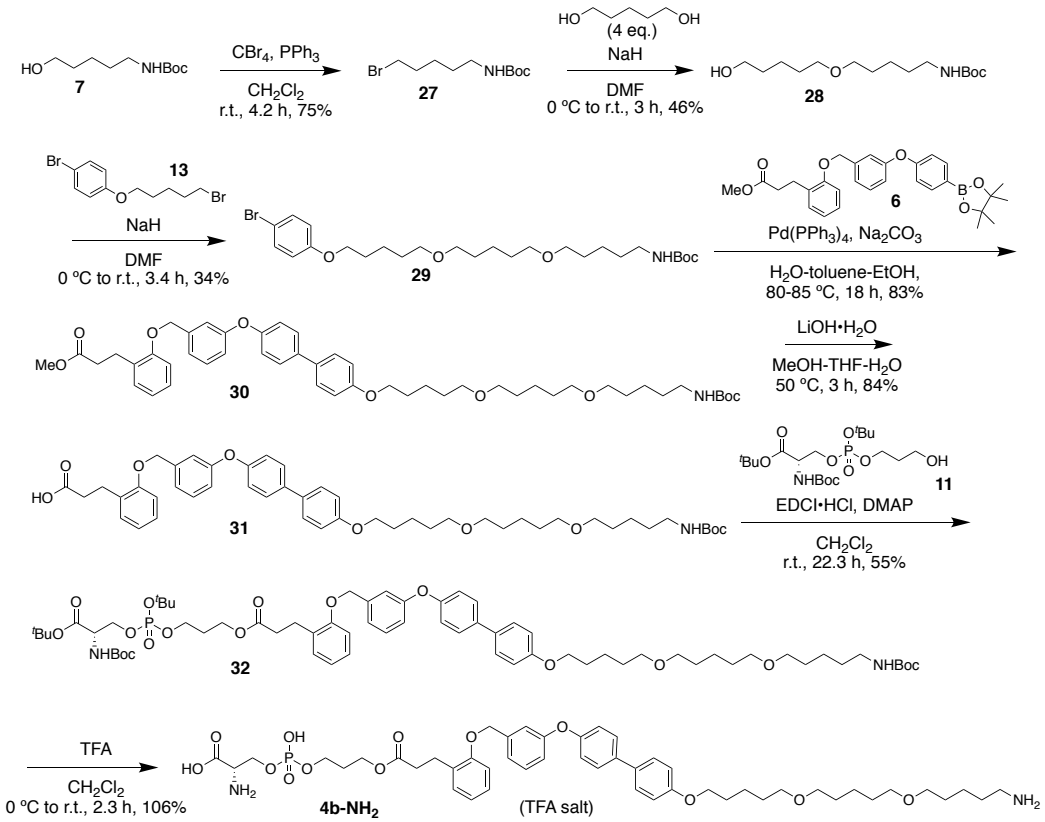
Compound **4a-CH$_3$**



Compound **26** (19.1 mg, 0.0168 mmol) was dissolved in anhydrous CH_2Cl_2 (0.2 mL) and TFA (0.5 mL) at 0 ºC and the reaction mixture was stirred at 0 ºC for 5 minutes and at room temperature for 1.5 hours. The mixture was then diluted with CH_2Cl_2 (5 mL) and evaporated. The residue was subjected to flash column chromatography ($CHCl_3$:MeOH:AcOH=9:1:0, 8:1:1 to 7:1:2). The resulting colorless oil was dissolved in CH_2Cl_2 and TFA and vacuum dried at 40 ˚C to afford **4a-CH$_3$** as the TFA-salt (colorless powder, 14.0 mg, 0.0152 mmol, 90%).

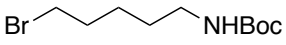
^{1}H NMR (400 MHz, $CDCl_3$/TFA-*d*): 7.503 (4H, brs), 7.360 (1H, dd, *J* = 7.8 Hz, 7.8 Hz), 7.216-7.152 (2H, m), 7.110-6.995 (7H, m), 6.919-6.899 (2H, m), 5.095 (2H, s), 4.622 (2H, brs), 4.490 (1H, brs), 4.175 (2H, brs), 4.104 (2H, t, *J* = 6.0 Hz), 4.027 (2H, brs), 3.682-3.609 (8H, m), 2.961 (2H, brs), 2.710 (2H, brs), 1.934 (2H, brs), 1.880-1.810 (2H, m), 1.757-1.503 (10H, m), 1.398-1.288 (8H, m), 0.880 (3H, t, *J* = 6.7 Hz). ^{31}P NMR (161 MHz, $CDCl_3$/TFA): -1.44. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{50}H_{67}NO_{13}P^-$: 920.4356. Found: 920.4330. Anal. Calcd. For: $C_{50}H_{68}NO_{13}P\bullet 0.7CF_3CO_2H$: C, 61.62; H, 6.91; N, 1.40. Found: C, 61.26; H, 7.27; N, 1.45.

Synthesis of 4b-NH$_2$ (Scheme 5)



Scheme 5. Synthesis of 4b-NH_2

Compound **27**

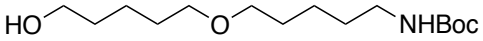


Compound **7** (2027.8 mg, 9.9753mmol) and CBr_4 (3650.1 mg, 11.007 mmol) were dissolved in CH_2Cl_2 (50 mL). PPh_3 (2885. 5 mg, 11.001 mmol) was added at 0 ºC and the mixture was stirred for 4.2 hours at room temperature. Silica-gel (15 mL) was added and the solvent was evaporated. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=10:1 to 5:1) to afford compound **27** as a colorless solid (1992.9 mg, 7.4871 mmol, 75%).

^{1}H NMR (400 MHz, $CDCl_3$): 4.541 (1H, brs), 3.408 (2H, t, J = 6.7 Hz), 3.133-3.120 (2H, m), 1.877 (2H, quin, J = 7.0 Hz), 1.526-1.423 (13H, m). ^{13}C NMR (100 MHz,

$CDCl_3$): 155.95, 79.16, 40.30, 33.61, 32.30, 29.25, 28.39, 25.32. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{10}H_{20}BrNNaO_2^+$: 288.0570. Found: 288.0570.

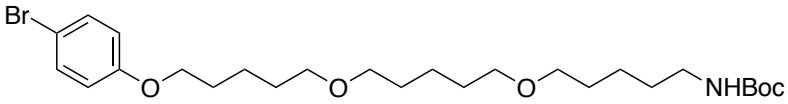
Compound **28**



Pentane-1,5-diol (3110.8 mg, 29.869 mmol) was dissolved in anhydrous DMF (15 mL). NaH (60% dispersion in mineral oil, 443.1 mg, 11.08 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 20 minutes. Compound **27** (1979.4 mg, 7.4363 mmol) was added to the solution at 0 ºC and the whole was stirred at room temperature for 3 hours. H_2O (60 mL) was added and the whole was extracted three times with Et_2O (50 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1 to 1:1) to afford compound **28** as a colorless oil (996.4 mg, 3.443 mmol, 46%).

^{1}H NMR (400 MHz, $CDCl_3$): 4.579 (1H, brs), 3.650 (2H, t, J = 5.9 Hz), 3.427-3.385 (4H, m), 3.109 (2H, t, J = 7.0 Hz), 2.020-1.700 (1H, brs), 1.639-1.546 (6H, m), 1.517-1.353 (15H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 156.04, 79.14, 70.72, 70.62, 62.76, 40.68, 32.43, 29.82, 29.36, 29.30, 28.40, 23.46, 22.46. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{15}H_{31}NNaO_4^+$: 312.2145. Found: 312.2144.

Compound **29**

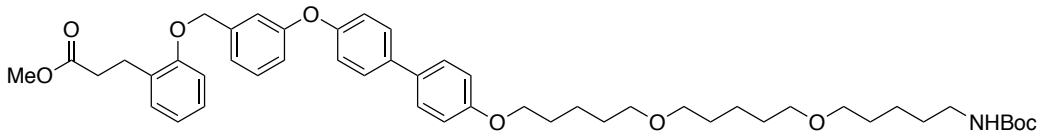


Compound **28** (739.4 mg, 2.555 mmol) was dissolved in anhydrous DMF (2.0 mL). NaH (60% dispersion in mineral oil, 118.7 mg, 2.968 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 25 minutes. Compound **13** (909.2 mg, 2.823 mmol) in anhydrous DMF (3.0 mL) was added to the solution at 0 ºC and the whole was stirred at room temperature for 3.4 hours. H_2O (20 mL) was added and the whole was extracted three times with Et_2O (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=5:1 to 4:1) to afford compound **29** as a

colorless oil, which later changed to a colorless powder (465.0 mg, 0.8765 mmol, 34%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.352 (2H, d, J = 9.0 Hz), 6.763 (2H, d, J = 9.0 Hz), 4.538 (1H, brs), 3.919 (2H, t, J = 6.5 Hz), 3.437-3.369 (8H, m), 3.102 (2H, brs), 1.793 (2H, quin, J = 7.0 Hz), 1.668-1.321 (25H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.16, 155.98, 132.16, 116.25, 112.56, 79.08, 70.85, 70.81, 70.66, 70.65, 68.06, 40.60, 29.88, 29.56, 29.46, 29.37, 28.98, 28.41, 23.49, 22.81, 22.73. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{26}H_{44}BrNNaO_5^+$: 552.2295. Found: 552.2324. Anal. Calcd For $C_{26}H_{44}BrNO_5$: C, 58.86; H, 8.36; N, 2.64. Found: C, 58.86; H, 8.44; N, 2.58. Mp: 32.2-33.1 ˚C.

Compound **30**

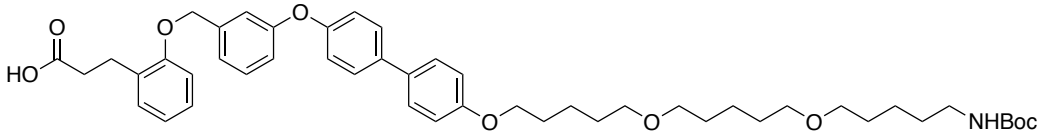


Compound **29** (277.3 mg, 0.5227 mmol), compound **6** (380.3 mg, 0.7787 mmol), Na_2CO_3 (506.9 mg, 4.783 mmol) and $Pd(PPh_3)_4$ (18.8 mg, 0.0163 mmol) were dissolved in H_2O (2.4 mL), toluene (2.4 mL) and EtOH (1.2 mL). The mixture was degassed by freeze-pump-thaw cycles and stirred at 80-85 ºC under an Ar atmosphere for 18 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (15 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated. The residue was chromatographed on an open column of silica-gel four times (first, *n*-hexane:AcOEt=4:1 to 2:1, second to forth, *n*-hexane:AcOEt=5:2) to afford compound **30** (351.1 mg, 0.4324 mmol, 83%) as a colorless oil.

^{1}H NMR (400 MHz, $CDCl_3$): 7.507 (2H, d, J = 8.7 Hz), 7.479 (2H, d, J = 8.8 Hz), 7.350 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.187-7.143 (3H, m), 7.105 (1H, brs), 7.061 (2H, d, J = 8.7 Hz), 6.987 (1H, dd, J = 8.2 Hz, 1.9 Hz), 6.951 (2H, d, J = 8.8 Hz), 6.908-6.853 (2H, m), 5.067 (2H, s), 4.538 (1H, brs), 3.998 (2H, t, J = 6.5 Hz), 3.612 (3H, s), 3.453-3.369 (8H, m), 3.100 (2H, brs), 2.974 (2H, t, J = 7.7 Hz), 2.610 (2H, t, J = 7.7 Hz), 1.831 (2H, quin, J = 7.0 Hz), 1.708-1.337 (25H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 173.68, 158.48, 157.69, 156.32, 155.95, 155.81, 139.27, 136.29, 132.85, 130.10, 129.89, 129.11, 127.96, 127.84, 127.55, 121.53, 120.81, 119.37, 117.86, 117.03, 114.75, 111.52, 78.98, 70.84, 70.80, 70.69, 70.64, 69.26, 67.89, 53.39, 51.44,

40.50, 33.97, 29.86, 29.55, 29.50, 29.36, 29.11, 28.39, 26.14, 23.47, 22.80, 22.77. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{49}H_{65}NNaO_4^+$: 834.4552. Found: 834.4539.

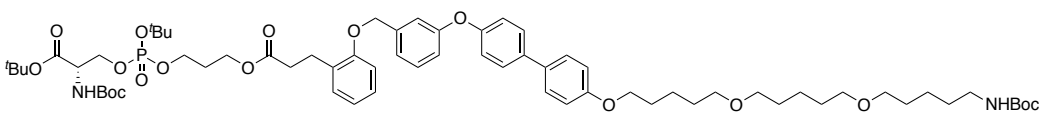
Compound **31**



Compound **30** (141.3 mg, 0.1693 mmol) was dissolved in MeOH (0.6 mL), THF (0.6 mL) and H_2O (0.6 mL). LiOH•H_2O (25.3 mg, 0.603 mmol) was added and the solution was stirred at 50 ºC for 3 hours. 2M aqueous solution of HCl (4 mL) was added to acidify the mixture to pH 2. The whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1 to CH_2Cl_2:MeOH=10:1) to afford compound **31** as a colorless sticky oil (113.4 mg, 0.1421 mmol, 84%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.499 (2H, d, *J* = 8.8 Hz), 7.472 (2H, d, *J* = 8.8 Hz), 7.341 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.180-7.146 (3H, m), 7.085 (1H, brs), 7.055 (2H, d, *J* = 8.7 Hz), 6.984 (1H, dd, *J* = 8.1 Hz, 1.8 Hz), 6.941 (2H, d, *J* = 8.8 Hz), 6.903-6.849 (2H, m), 5.057 (2H, s), 4.566 (1H, brs), 3.988 (2H, t, *J* = 6.5 Hz), 3.449-3.365 (8H, m), 3.089 (2H, brs), 2.964 (2H, t, *J* = 7.7 Hz), 2.630 (2H, t, *J* = 7.7 Hz), 1.821(2H, quin, *J* = 7.0 Hz), 1.684-1.308 (25H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 177.86, 158.47, 157.74, 156.34, 155.99, 155.80, 139.22, 136.30, 132.87, 130.11, 129.93, 128.93, 127.97, 127.85, 127.61, 121.54, 120.83, 119.43, 117.91, 116.99, 114.77, 111.53, 79.11, 70.82, 70.67, 70.66, 69.30, 67.89, 40.57, 33.76, 29.83, 29.54, 29.52, 29.46, 29.34, 29.09, 28.40, 25.89, 23.45, 22.80, 22.77. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{48}H_{62}NO_9^-$: 796.4430. Found: 796.4404.

Compound **32**

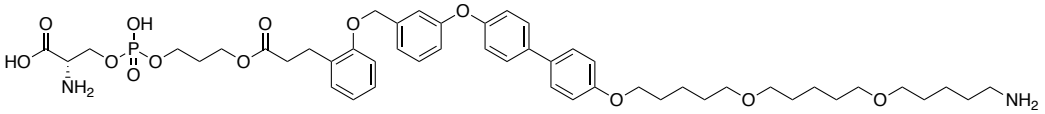


Compound **31** (84.8 mg, 0.106 mmol), compound **11** (46.2 mg, 0.101 mmol) and DMAP (4.2 mg, 0.034 mmol) were dissolved in anhydrous CH_2Cl_2 (1.6 mL).

EDCI•HCl (37.6 mg, 0.196 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 22.3 hours. EDCI•HCl (11.0 mg) and MeOH (0.2 mL) were added and the reaction mixture was stirred for 1 hr. H_2O (10 mL) was added and the whole was extracted five times with CH_2Cl_2 (10 mL×5). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1, 1:1 to 1:2) to afford compound **32** as a colorless oil (69.2 mg, 0.0560 mmol, 55%).

^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.510 (2H, d, *J* = 8.7 Hz), 7.482 (2H, d, *J* = 8.8 Hz), 7.352 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.183-7.148 (3H, m), 7.100 (1H, brs), 7.060 (2H, d, *J* = 8.7 Hz), 6.991-6.943 (3H, m), 6.908-6.856 (2H, m), 5.510 (1H, d, *J* = 7.8 Hz), 5.079 (2H, s), 4.576 (1H, brs), 4.367-4.302 (2H, m), 4.224-4.179 (1H, m), 4.154-4.112 (2H, m), 4.031-3.985 (4H, m), 3.456-3.372 (8H, m), 3.113-3.098 (2H, m), 2.981 (2H, t, *J* = 7.6 Hz), 2.649-2.608 (2H, m), 1.963-1.888 (2H, m), 1.834 (2H, quin, *J* = 7.0 Hz), 1.695-1.322 (52H, m). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.57, -5.72 (mixture of diastereomers). ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.04, 168.29, 158.46, 157.63, 156.30, 155.93, 155.78, 155.19, 139.28, 136.26, 132.79, 130.07, 129.89, 128.98, 127.93, 127.80, 127.55, 121.48, 120.78, 119.32, 117.77, 117.04, 114.73, 111.55, 83.66, 83.59, 82.59, 79.87, 78.92, 70.81, 70.76, 70.65, 70.61, 69.27, 67.86, 67.36, 67.30, 64.01, 63.96, 60.36, 54.41, 54.33, 40.49, 33.98, 29.83, 29.73, 29.69, 29.52, 29.47, 29.33, 29.08, 28.37, 28.24, 27.88, 26.07, 23.44, 22.78, 22.74. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{67}H_{99}N_2NaO_{17}P^+$: 1257.6574. Found: 1257.6580.

Compound **4b-NH$_2$**

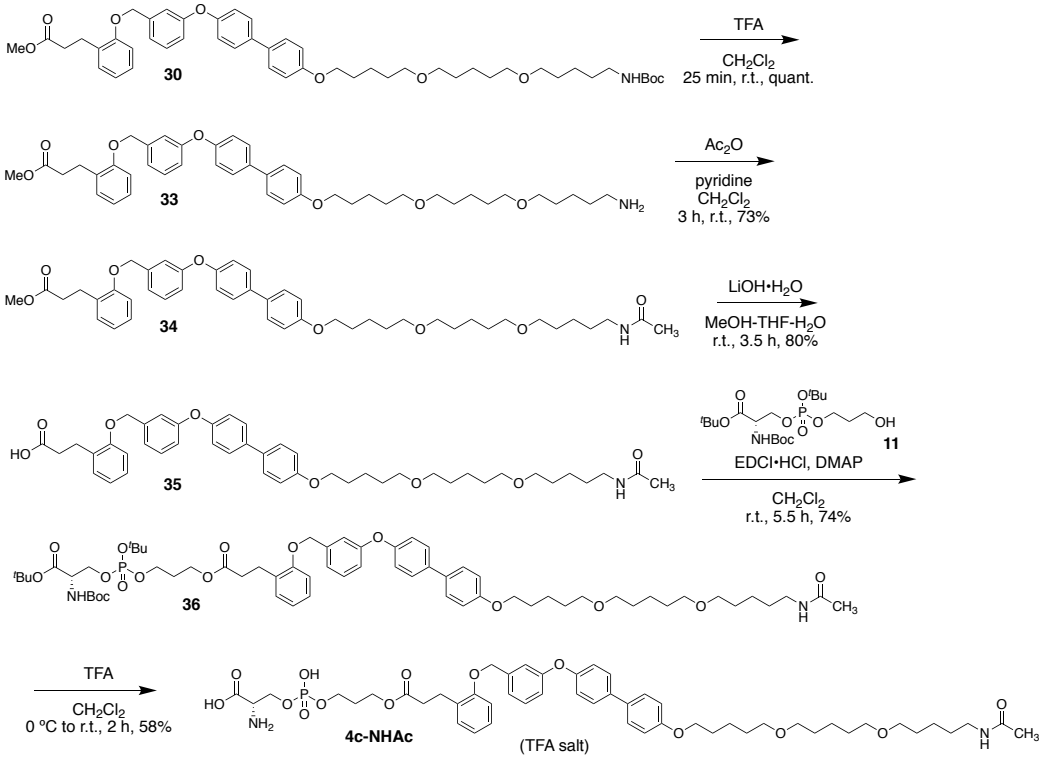


Compound **32** (62.3 mg, 0.0504 mmol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL) and TFA (1.0 mL) at 0 ºC and the reaction mixture was stirred at 0 ºC for 10 minutes and at room temperature for 2.3 hours. The mixture was then diluted with CH_2Cl_2 (5 mL) and evaporated. The residue was subjected to flash column chromatography ($CHCl_3$:MeOH:AcOH=7:1:2, 6:1:3 to 5:1:4). The resulting colorless oil was dissolved

in CH_2Cl_2 and TFA and vacuum dried at 40 ˚C to afford **4b-NH_2** as the TFA-salt (colorless sticky oil, 49.5 mg, 0.0536 mmol, 106%).

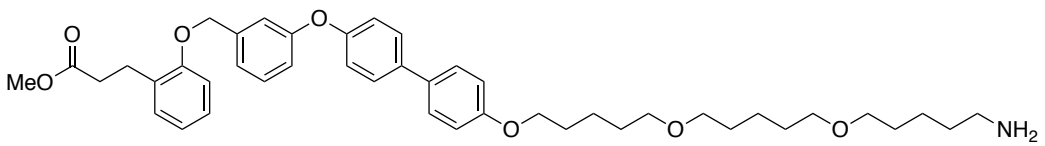
^{1}H NMR (400 MHz, $CDCl_3$/TFA-*d*): 7.540-7.483 (4H, m), 7.356 (1H, dd, *J* = 7.8 Hz, 7.8 Hz), 7.206-7.129 (2H, m), 7.096-6.974 (7H, m), 6.901-6.881 (2H, m), 5.080 (2H, s), 4.553 (2H, brs), 4.437 (1H, brs), 4.120-4.082 (4H, m), 4.033-3.960 (2H, m), 3.664-3.574 (8H, m), 3.107 (2H, brs), 2.932 (2H, brs), 2.667 (2H, brs), 1.883-1.795 (4H, m), 1.724-1.629 (10H, m), 1.540-1.299 (8H, m). ^{31}P NMR (161 MHz, $CDCl_3$/TFA-*d*): -1.25. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{49}H_{66}N_2O_{13}P^-$: 921.4308. Found: 921.4291. Anal. Calcd. For: $C_{49}H_{67}N_2O_{13}P\bullet1.7CF_3CO_2H$: C, 56.35; H, 6.20; N, 2.51. Found: C, 56.04; H, 6.56; N, 2.34.

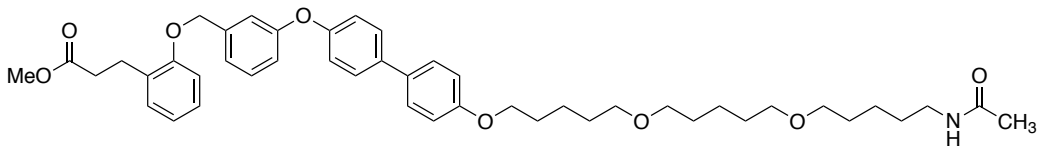
Synthesis of 4c-NHAc (Scheme 6)



Scheme 6. Synthesis of 4c-NHAc

Compounds **33** and **34**

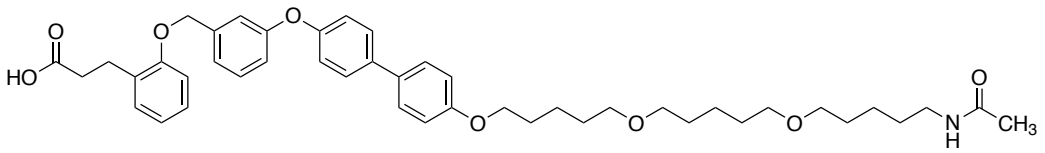




Compound **30** (142.1 mg, 0.1750 mmol) was dissolved in anhydrous CH_2Cl_2 (5.0 mL) and TFA (1.0 mL) at 0 ºC. Then the reaction mixture was stirred at room temperature for 25 minutes. The mixture was evaporated. AcOEt (15 mL) and 10% aqueous solution of Na_2CO_3 (15 mL) were added. The water layer was separated and extracted twice with AcOEt (10 mL×2). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered, and the solvent was evaporated to afford compound **33** as a colorless oil (124.6 mg, 0.1750 mmol, 100%), which was used without further purification.

Compound **33** (118.9 mg, 0.1670 mmol) was dissolved in anhydrous CH_2Cl_2 (1.0 mL) Ac_2O (26 μL) and pyridine (42 μL) was added and the mixture was stirred at room temperature under an Ar atmosphere for 3 hours. Saturated solution of $NaHCO_3$ was added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered, and the solvent was evaporated. The residue was flash-chromatographed on a column of silica-gel (*n*-hexane:AcOEt=1:3) to afford compound **34** (91.7 mg, 0.122 mmol, 73%). ^{1}H NMR (400 MHz, $CDCl_3$): 7.509 (2H, d, *J* = 8.8 Hz), 7.482 (2H, d, *J* = 8.8 Hz), 7.354 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.191-7.147 (3H, m), 7.107 (1H, brs), 7.064 (2H, d, *J* = 8.7 Hz), 7.002-6.976 (1H, m), 6.954 (2H, d, *J* = 8.8 Hz), 6.912-6.855 (2H, m), 5.530 (1H, brs), 5.072 (2H, s), 4.002 (2H, t, *J* = 6.5 Hz), 3.616 (3H, s), 3.456-3.377 (8H, m), 3.234 (2H, q, *J* = 6.6 Hz), 2.976 (2H, t, *J* = 7.7 Hz), 2.612 (2H, t, *J* = 7.7 Hz), 1.958 (3H, s), 1.834 (2H, quin, *J* = 7.1 Hz), 1.694-1.478 (12H, m), 1.440-1.356 (4H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 173.72, 169.98, 158.49, 157.70, 156.34, 155.84, 139.29, 136.30, 132.88, 130.12, 129.92, 129.13, 127.98, 127.86, 127.57, 121.56, 120.83, 119.38, 117.88, 117.06, 114.77, 111.54, 70.84, 70.72, 70.60, 69.29, 67.91, 51.46, 39.58, 33.99, 29.57, 29.55, 29.50, 29.34, 29.33, 29.12, 26.16, 23.66, 23.33, 22.84, 22.79. HRMS (ESI-TOF, $[M+H]^+$): Calcd for $C_{46}H_{60}NO_8^+$: 754.4313. Found: 754.4292. Anal. Calcd For $C_{46}H_{59}NO_8$•0.2H_2O: C, 72.93; H, 7.90; N, 1.85. Found: C, 72.74; H, 7.68; N, 1.78.

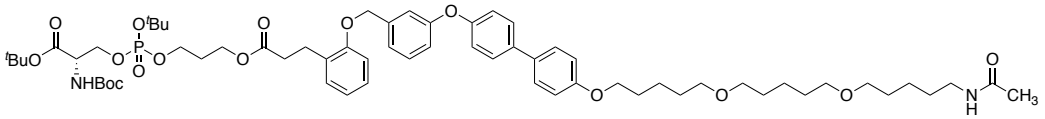
Compound **35**



Compound **34** (85.5 mg, 0.113 mmol) was dissolved in MeOH (0.6 mL), THF (0.6 mL) and H_2O (0.6 mL). LiOH•H_2O (14.7 mg, 0.350 mmol) was added and the solution was stirred at room temperature for 3.5 hours. 2M aqueous solution of HCl was added to acidify the mixture to pH 2. The whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:acetone=3:2, 1:1 to 0:1) to afford compound **35** as a colorless oil (67.4 mg, 0.0911 mmol, 80%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.501 (2H, d, *J* = 8.7 Hz), 7.475 (2H, d, *J* = 8.8 Hz), 7.342 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.182-7.145 (3H, m), 7.090 (1H, brs), 7.056 (2H, d, *J* = 8.7 Hz), 6.985 (1H, dd, *J* = 8.1 Hz, 1.8 Hz), 6.942 (2H, d, *J* = 8.8 Hz), 6.900-6.849 (2H, m), 5.628 (1H, brs), 5.057 (2H, s), 3.991 (2H, t, *J* = 6.5 Hz), 3.452-3.372 (8H, m), 3.221 (2H, q, *J* = 6.6 Hz), 2.966 (2H, t, *J* = 7.7 Hz), 2.627 (2H, t, *J* = 7.7 Hz), 1.996 (0.15H, s), 1.952 (2.85H, s), 1.822 (2H, quin, *J* = 7.0 Hz), 1.686-1.464 (12H, m), 1.433-1.325 (4H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 177.44, 170.36, 158.45, 157.71, 156.33, 155.81, 139.22, 136.26, 132.86, 130.10, 129.92, 129.00, 127.96, 127.84, 127.57, 121.56, 120.81, 119.42, 117.91, 116.99, 114.76, 111.51, 70.81, 70.80, 70.67, 70.57, 69.28, 67.88, 39.61, 33.79, 29.51, 29.48, 29.44, 29.26, 29.23, 29.07, 25.92, 23.62, 23.21, 22.82, 22.77. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{45}H_{56}NO_8^-$: 738.4011. Found: 738.3997. Anal. Calcd For $C_{45}H_{57}NO_8$•0.4H_2O: C, 72.34; H, 7.80; N, 1.87. Found: C, 72.18; H, 7.87; N, 1.73.

Compound **36**

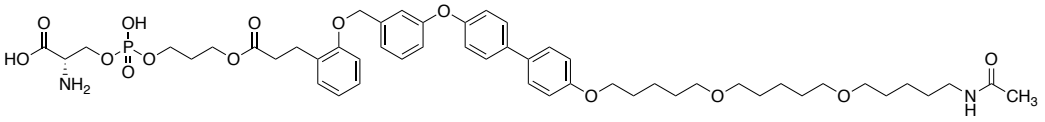


Compound **35** (61.5 mg, 0.0831 mmol), compound **11** (38.3 mg, 0.0841 mmol) and DMAP (4.2 mg, 0.034 mmol) were dissolved in anhydrous CH_2Cl_2 (1.0 mL). EDCI•HCl (33.3 mg, 0.174 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 5.5 hours. H_2O (10 mL) was added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer

was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:acetone=3:2) to afford compound **36** as a colorless sticky oil (72.8 mg, 0.0618 mmol, 74%).

^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.510 (2H, d, *J* = 8.7 Hz), 7.482 (2H, d, *J* = 8.8 Hz), 7.353 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.184-7.149 (3H, m), 7.099 (1H, brs), 7.060 (2H, d, *J* = 8.7 Hz), 6.991-6.943 (3H, m), 6.908-6.856 (2H, m), 5.619 (1H, brs), 5.509 (1H, d, *J* = 7.8 Hz), 5.080 (2H, s), 4.364-4.299 (2H, m), 4.222-4.176 (1H, m), 4.153-4.110 (2H, m), 4.030-3.970 (4H, m), 3.456-3.376 (8H, m), 3.231 (2H, q, *J* = 6.6 Hz), 2.980 (2H, t, *J* = 7.6 Hz), 2.649-2.607 (2H, m), 1.956-1.886 (5H, m), 1.868-1.798 (2H, m), 1.694-1.495 (12H, m), 1.472-1.431 (27H, m), 1.418-1.354 (4H, m). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.58, -5.73. ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.08, 169.99, 168.32, 158.47, 157.64, 156.32, 155.81, 155.21, 139.30, 136.27, 132.83, 130.09, 129.91, 128.99, 127.95, 127.83, 127.58, 121.51, 120.80, 119.35, 117.80, 117.07, 114.75, 111.57, 83.71, 83.63, 82.66, 82.63, 79.92, 70.82, 70.69, 70.58, 69.29, 67.89, 67.38, 64.04, 63.98, 60.39, 54.42, 54.35, 39.55, 34.00, 29.75, 29.74, 29.71, 29.69, 29.54, 29.52, 29.48, 29.44, 29.31, 29.23, 29.10, 28.26, 27.90, 26.09, 23.63, 23.29, 22.82, 22.77. HRMS (ESI-TOF, $[M+H]^+$): Calcd for $C_{64}H_{94}N_2O_{16}P^+$: 1177.6335. Found: 1177.6346.

Compound **4c-NHAc**

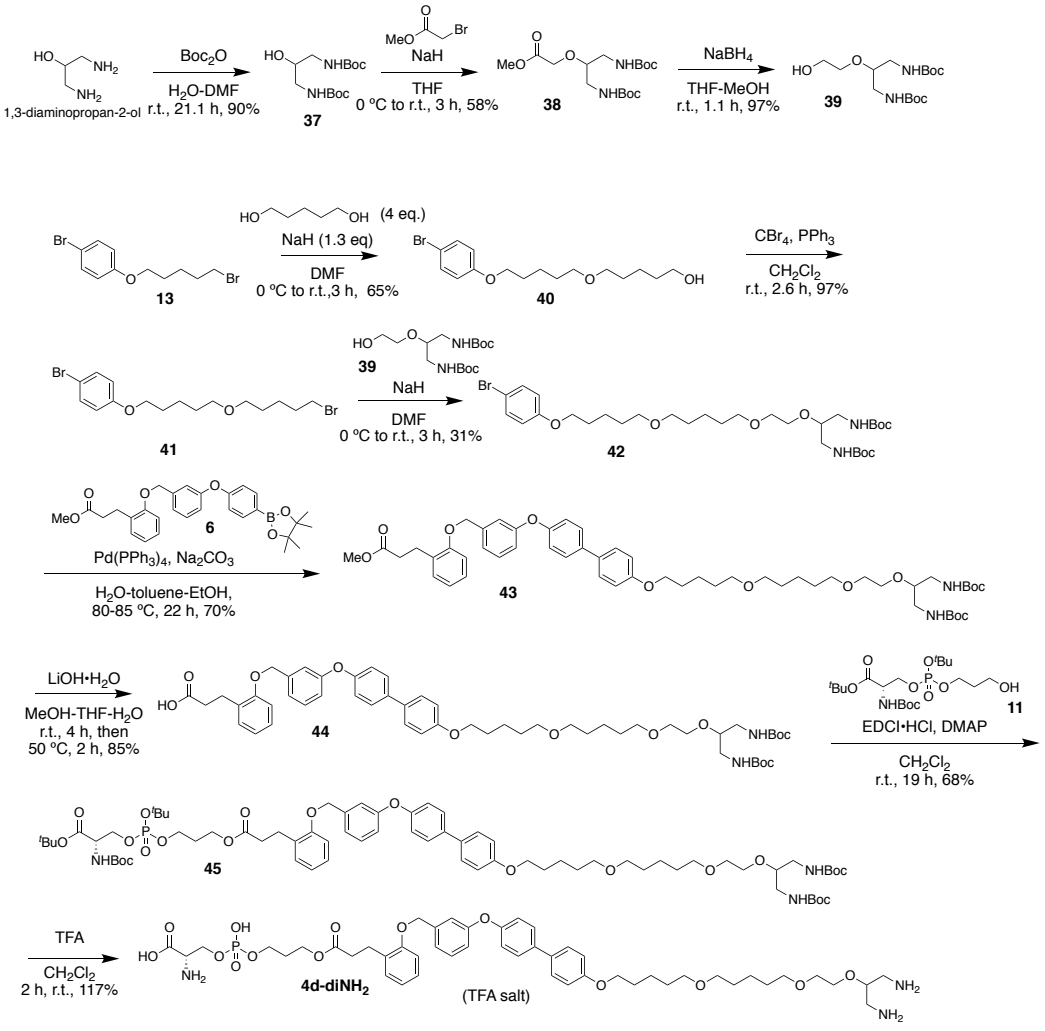


Compound **36** (69.7 mg, 0.0592 mmol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL) and TFA (1.0 mL) at 0 ºC and the reaction mixture was stirred at room temperature for 2 hours. The mixture was then diluted with CH_2Cl_2 (5 mL) and evaporated. The residue was flash-chromatographed on a column of silica-gel ($CHCl_3$:MeOH:AcOH=8:1:1 to 7:1:2). The resulting colorless oil was dissolved in CH_2Cl_2 and TFA, evaporated and vacuum dried at 40 ˚C to afford **4c-NHAc** as TFA-salt (colorless powder, 33.1 mg, 0.0343 mmol, 58%).

^{1}H NMR (400 MHz, $CDCl_3$/TFA-*d*): 7.512 (4H, brs), 7.369 (1H, t, *J* = 7.6 Hz, 7.6 Hz), 7.220-7.007 (9H, m), 6.930-6.910 (2H, m), 5.103 (2H, s), 4.652 (2H, brs), 4.535

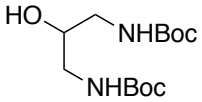
(1H, brs), 4.174-4.052 (6H, m), 3.690-3.613 (8H, m), 3.399 (2H, brs), 2.966 (2H, t, *J* = 7.0 Hz), 2.714 (2H, t, *J* = 7.0 Hz), 2.315 (3H, s), 1.942 (2H, brs), 1.851 (2H, quin, *J* = 6.8 Hz), 1.761-1.624 (10H, m), 1.566-1.491 (2H, m), 1.431-1.365 (4H, m). ^{31}P NMR (161 MHz, $CDCl_3$/TFA-*d*): -1.34. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{51}H_{68}N_2O_{14}P^-$: 963.4414. Found: 963.4436. Anal. Calcd For $C_{51}H_{69}N_2O_{14}P\bullet 0.8CF_3CO_2H$: C, 59.81; H, 6.66; N, 2.65. Found: C, 59.81; H, 6.89; N, 2.49.

Synthesis of 4d-diNH_2 (Scheme 7)



Scheme 7. Synthesis of 4d-diNH_2

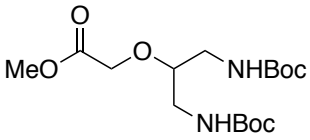
Compound **37**



1,3-Diaminopropan-2-ol (1020.7 mg, 11.325 mmol) was dissolved in H_2O (20 mL) and DMF (10 mL). Et_3N (9.2 mL) and Boc_2O (5814.8 mg, 26.643 mmol) in DMF (10 mL) were added and the mixture was stirred at room temperature for 21.1 hours. H_2O (30 mL) was added and the whole was extracted twice with AcOEt (35 mL×2). The combined organic layer was washed with 0.1 M aqueous solution of HCl (50 mL), saturated aqueous solution of $NaHCO_3$ (50 mL), H_2O (50 mL) and brine, then dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was flash-chromatographed on a column of silica-gel (*n*-hexane:AcOEt=2:1, 3:2 to 1:1) to afford compound **37** as a colorless solid (2962.6 mg, 10.203 mmol, 90%).

^{1}H NMR (400 MHz, $CDCl_3$): 5.151 (2H, brs), 3.768-3.694 (2H, m), 3.292-3.133 (4H, m), 1.443 (18H, s). ^{13}C NMR (100 MHz, $CDCl_3$): 157.27, 79.82, 71.16, 43.55, 28.34. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{13}H_{26}N_2NaO_5^+$: 313.1734. Found: 313.1742. Mp: 101.0-102.0 ˚C (colorless powder, recrystallized from *n*-hexane/CH_2Cl_2).

Compound **38**

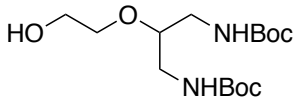


Compound **37** (2288.1 mg, 7.8802 mmol) was dissolved in anhydrous THF (30 mL). NaH (60% dispersion in mineral oil, 477.2 mg, 11.93 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 30 minutes. Methyl bromoacetate (1.0 mL) was added at 0 ºC and the whole was stirred at room temperature under an Ar atmosphere for 3 hours. Saturated solution of $NaHCO_3$ (30 mL) and H_2O (20 mL) was added. The water layer was separated and extracted twice with AcOEt (30 mL×2). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a yellow oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=5:2 to 2:1) to afford compound **38** as a colorless powder (1647.7 mg, 4.5463 mmol, 58%).

^{1}H NMR (400 MHz, $CDCl_3$): 5.329 (2H, brs), 4.198 (2H, s), 3.763 (3H, s), 3.486 (1H, quin, J = 5.1 Hz), 3.388-3.357 (2H, m), 3.136-3.101 (2H, m), 1.444 (18H, s). ^{13}C

NMR (100 MHz, $CDCl_3$): 171.55, 156.41, 79.40, 78.87, 66.99, 51.99, 40.49, 28.35. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{16}H_{30}N_2NaO_7^+$: 385.1945. Found: 385.1940. Mp: 98.2-100.0 ˚C (colorless plates, recrystallized from *n*-hexane/CH_2Cl_2).

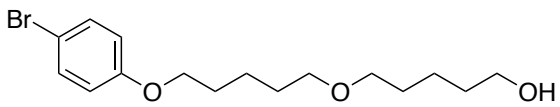
Compound **39**



Compound **38** (1628.2 mg, 4.4925 mmol) was dissolved in anhydrous THF (20 mL). $NaBH_4$ (429.5 mg, 11.353 mmol) was added. MeOH (4.0 mL) was added dropwise for 3 minutes and the mixture was stirred at room temperature for 1.1 hours. H_2O (30 mL) was added. The organic and water layers were separated and the water layer was extracted with AcOEt (60 mL). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=2:1) to afford compound **39** as a colorless sticky oil (1457.7 mg, 4.3590 mmol, 97%).

^1H NMR (400 MHz, $CDCl_3$): 5.254 (2H, brs), 3.726-3.705 (2H, m), 3.683-3.663 (2H, m), 3.443 (1H, quin, J = 5.0 Hz), 3.315-3.175 (4H, m), 3.029 (1H, t, J = 5.1 Hz), 1.442 (18H, s). ^{13}C NMR (100 MHz, $CDCl_3$): 156.58, 79.49, 77.66, 71.11, 61.85, 40.33, 28.31. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{15}H_{30}N_2NaO_6^+$: 357.1996. Found: 357.2017.

Compound **40**

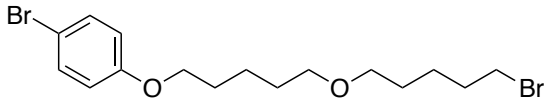


1,5-Pentanediol (711.4 mg, 6.831 mmol) was dissolved in anhydrous DMF (2.0 mL). NaH (60% dispersion in mineral oil, 88.4 mg, 2.21 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 20 minutes. Compound **13** (562.5 mg, 1.747 mmol) was added with anhydrous DMF (3.0 mL) and the whole was stirred at room temperature under an Ar atmosphere for 3 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (10 mL×3). The combined

organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=3:1, 2:1 to 1:1) to afford compound **40** as a colorless oil (389.0 mg, 1.127 mmol, 65%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.354 (2H, d, J = 9.0 Hz), 6.764 (2H, d, J = 9.0 Hz), 3.920 (2H, t, J = 6.5 Hz), 3.646 (2H, t, J = 6.5 Hz), 3.444-3.402 (4H, m), 1.794 (2H, quin, J = 7.1 Hz), 1.673-1.387 (12H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.15, 132.16, 116.26, 112.58, 70.82, 70.69, 68.06, 62.82, 32.47, 29.43, 29.39, 28.97, 22.73, 22.43. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{16}H_{26}BrO_3^+$: 345.1060. Found: 345.1072.

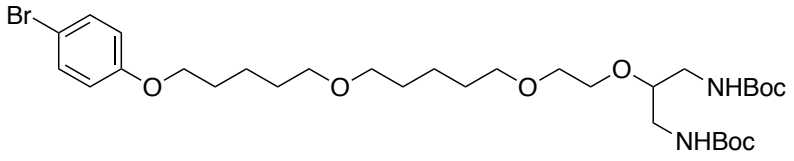
Compound **41**



Compound **40** (371.4 mg, 1.076 mmol) and CBr_4 (498.5 mg, 1.503 mmol) were dissolved in anhydrous CH_2Cl_2 (5.0 mL). PPh_3 (396.1 mg, 1.510 mmol) was added at 0 ºC and the whole was stirred at room temperature under an Ar atmosphere for 2.6 hours. Silica-gel (2.3 g) was added and the whole was evaporated. The residue was chromatographed on an open column of silica-gel (*n*-hexane:AcOEt=10:1) to afford compound **41** as a colorless oil (425.6 mg, 1.043 mmol, 97%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.354 (2H, d, J = 9.0 Hz), 6.764 (2H, d, J = 9.0 Hz), 3.922 (2H, t, J = 6.5 Hz), 3.441-3.393 (6H, m), 1.883 (2H, quin, J = 7.1 Hz), 1.796 (2H, quin, J = 7.0 Hz), 1.672-1.458 (8H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.17, 132.17, 116.26, 112.58, 70.71, 70.56, 68.06, 33.76, 32.59, 29.45, 28.98, 28.88, 24.93, 22.75. HRMS (ESI-TOF, $[M+H]^+$): Calcd for $C_{16}H_{25}Br_2O_2^+$: 407.0216. Found: 407.0209. Anal. Calcd For $C_{16}H_{24}Br_2O_2$: C, 47.08; H, 5.93. Found: C, 47.02; H, 5.82.

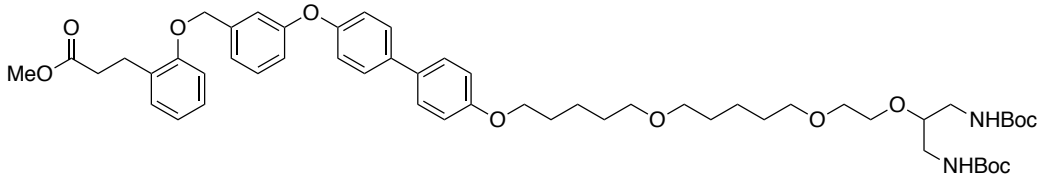
Compound **42**



Compound **39** (242.3 mg, 0.7246 mmol) was dissolved in anhydrous DMF (1.0 mL). NaH (60% dispersion in mineral oil, 37.5 mg, 0.938 mmol) was added at 0 ºC and the mixture was stirred at 0 ºC under an Ar atmosphere for 20 minutes. Compound **41** (325.5 mg, 0.7975 mmol) was added with anhydrous DMF (1.5 mL) and the whole was stirred at room temperature under an Ar atmosphere for 3 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was flash-chromatographed on a column of silica-gel (*n*-hexane:AcOEt=5:2) to afford compound **42** as a colorless oil (149.3 mg, 0.2256 mmol, 31%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.352 (2H, d, J = 9.1 Hz), 6.762 (2H, d, J = 9.0 Hz), 5.227-4.895 (2H, m), 3.918 (2H, t, J = 6.5 Hz), 3.685 (2H, t, J = 4.6 Hz), 3.523 (2H, t, J = 4.6 Hz), 3.476-3.384 (7H, m), 3.284-3.242 (2H, m), 3.138-3.103 (2H, m), 1.791 (2H, quin, J = 7.0 Hz), 1.665-1.371 (10H, m), 1.434 (18H, s). ^{13}C NMR (100 MHz, $CDCl_3$): 158.15, 156.32, 132.15, 116.24, 112.56, 79.26, 71.32, 70.81, 70.67, 70.16, 68.79, 68.04, 40.70, 29.51, 29.44, 29.35, 28.97, 28.38, 22.72. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{31}H_{53}BrN_2NaO_8^+$: 683.2878. Found: 683.2895. Anal. Calcd For $C_{31}H_{53}BrN_2O_8$: C, 56.27; H, 8.07; N, 4.23. Found: C, 56.05; H, 7.78; N, 4.06.

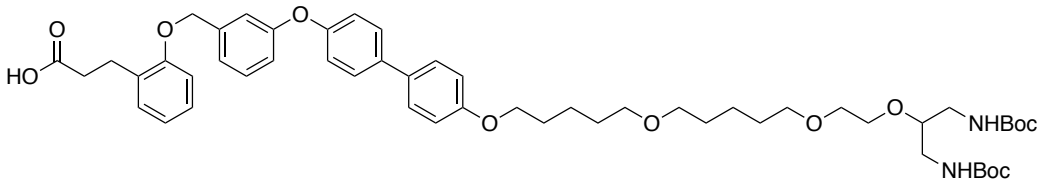
Compound **43**



Compound **42** (25.3 mg, 0.0382 mmol), compound **6** (28.5 mg, 0.0584 mmol), Na_2CO_3 (84.5 mg, 0.797 mmol) and $Pd(PPh_3)_4$ (2.7 mg, 0.0023 mmol) were dissolved in H_2O (0.4 mL), toluene (0.4 mL) and EtOH (0.2 mL). The mixture was degassed by freeze-pump-thaw cycles and stirred at 80-85 ºC under an Ar atmosphere for 22 hours. H_2O (10 mL) was added and the whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated. The residue was flash-chromatographed on a column of silica-gel twice (*n*-hexane:acetone=3:1) to afford compound **43** as a colorless oil (25.3 mg, 0.0268 mmol, 70%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.507 (2H, d, J = 8.8 Hz), 7.480 (2H, d, J = 8.8 Hz), 7.351 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.188-7.144 (3H, m), 7.106 (1H, brs), 7.062 (2H, d, J = 8.7 Hz), 7.000-6.941 (3H, m), 6.909-6.853 (2H, m), 5.239 (2H, brs), 5.069 (2H, s), 3.999 (2H, t, J = 6.5 Hz), 3.687 (2H, t, J = 4.5 Hz), 3.613 (3H, s), 3.526 (2H, t, J = 4.6 Hz), 3.482-3.397 (7H, m), 3.320-3.286 (2H, m), 3.153-3.105 (2H, m), 2.975 (2H, t, J = 7.7 Hz), 2.611 (2H, t, J = 7.7 Hz), 1.832 (2H, quin, J = 7.0 Hz), 1.692-1.518 (8H, m), 1.433-1.365 (20H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 173.68, 158.46, 157.68, 156.31, 155.81, 139.27, 136.26, 132.85, 130.09, 129.89, 129.10, 127.95, 127.83, 127.54, 121.52, 120.80, 119.36, 117.85, 117.03, 114.73, 111.51, 79.21, 71.30, 70.80, 70.71, 70.14, 69.25, 68.72, 67.87, 51.43, 40.65, 33.96, 29.51, 29.48, 29.35, 29.10, 28.36, 26.13, 22.77, 22.70. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{54}H_{74}N_2NaO_{12}^+$: 965.5134. Found: 965.5130.

Compound **44**

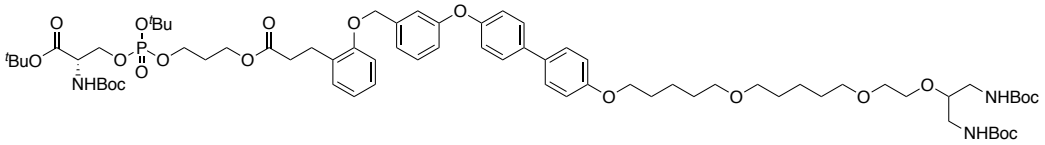


Compound **43** (126.9 mg, 0.1345 mmol) was dissolved in MeOH (0.7 mL), THF (0.7 mL) and H_2O (0.7 mL). LiOH•H_2O (16.8 mg, 0.400 mmol) was added and the solution was stirred at room temperature for 4 hours and then at 50 ºC for 2 hours. 2M aqueous solution of HCl (1 mL) was added to acidify the mixture to pH 2. The whole was extracted three times with AcOEt (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated to give a colorless oil. The residue was chromatographed on an open column of silica-gel three times (first, *n*-hexane:acetone=3:2, second, $CHCl_3$:MeOH=20:1, third, *n*-hexane:acetone=2:1) to afford compound **44** as a colorless oil (106.0 mg, 0.1141 mmol, 85%).

^{1}H NMR (400 MHz, $CDCl_3$): 7.500 (2H, d, J = 8.8 Hz), 7.474 (2H, d, J = 8.8 Hz), 7.345 (1H, dd, J = 7.9 Hz, 7.9 Hz), 7.183-7.148 (3H, m), 7.089 (1H, brs), 7.056 (2H, d, J = 8.7 Hz), 6.989 (1H, dd, J = 8.0 Hz, 2.0 Hz), 6.943 (2H, d, J = 8.8 Hz), 6.906-6.852 (2H, m), 5.250 (2H, brs), 5.062 (2H, s), 3.993 (2H, t, J = 6.5 Hz), 3.680 (2H, brs), 3.520 (2H, t, J = 4.5 Hz), 3.476-3.393 (7H, m), 3.312-3.278 (2H, m), 3.136-3.102 (2H, m), 2.965 (2H, t, J = 7.7 Hz), 2.633 (2H, t, J = 7.7 Hz), 1.823 (2H, quin, J

= 7.0 Hz), 1.685-1.514 (8H, m), 1.429-1.361 (20H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 177.50, 158.47, 157.73, 156.35, 156.33, 155.81, 139.21, 136.29, 132.87, 130.12, 129.94, 128.91, 127.97, 127.85, 127.62, 121.54, 120.83, 119.43, 117.93, 116.98, 114.76, 111.53, 79.36, 71.33, 70.79, 70.70, 70.15, 69.29, 68.83, 67.88, 40.81, 33.68, 29.51, 29.46, 29.35, 29.09, 28.38, 25.89, 22.78, 22.72. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{53}H_{71}N_2O_{12}^-$: 927.5012. Found: 927.5028. Anal. Calcd For $C_{53}H_{72}N_2O_{12}\bullet 0.1H_2O$: C, 68.38; H, 7.82; N, 3.01. Found: C, 68.05; H, 7.83; N, 2.97.

Compound **45**

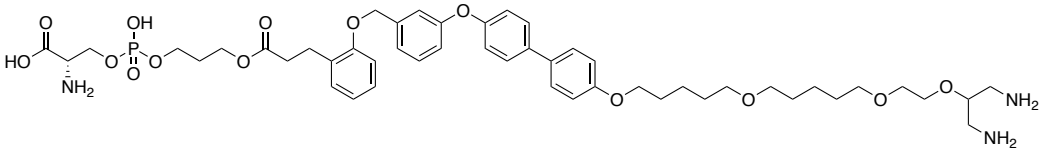


Compound **44** (85.7 mg, 0.0922 mmol), compound **11** (42.6 mg, 0.0935 mmol) and DMAP (3.5 mg, 0.029 mmol) were dissolved in anhydrous CH_2Cl_2 (1.0 mL). EDCI•HCl (35.9 mg, 0.187 mmol) was added and the whole was stirred at room temperature under an Ar atmosphere for 19 hours. H_2O (10 mL) was added and the whole was extracted three times with CH_2Cl_2 (10 mL×3). The combined organic layer was washed with brine, dried over Na_2SO_4 and filtered. The solvent was evaporated. The residue was flash-chromatographed on a column of silica-gel ($CHCl_3$:AcOEt=5:2) to afford compound **45** as a colorless sticky oil (85.1 mg, 0.0623 mmol, 68%).

^{1}H NMR (400 MHz, $CDCl_3$, a mixture of diastereomers): 7.510 (2H, d, *J* = 8.7 Hz), 7.483 (2H, d, *J* = 8.8 Hz), 7.353 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.184-7.149 (3H, m), 7.101 (1H, brs), 7.060 (2H, d, *J* = 8.7 Hz), 6.992-6.944 (3H, m), 6.906-6.857 (2H, m), 5.513 (1H, d, *J* = 7.7 Hz), 5.243 (2H, brs), 5.079 (2H, s), 4.368-4.303 (2H, m), 4.225-4.180 (1H, m), 4.155-4.112 (2H, m), 4.032-3.972 (4H, m), 3.688 (2H, t, *J* = 4.5 Hz), 3.528 (2H, t, *J* = 4.5 Hz), 3.485-3.400 (7H, m), 3.308-3.275 (2H, m), 3.152-3.118 (2H, m), 2.982 (2H, t, *J* = 7.6 Hz), 2.650-2.612 (2H, m), 1.964-1.888 (2H, m), 1.834 (2H, quin, *J* = 7.0 Hz), 1.695-1.514 (8H, m), 1.475-1.387 (47H, m). ^{31}P NMR (161 MHz, $CDCl_3$, a mixture of diastereomers): -5.56, -5.71. ^{13}C NMR (100 MHz, $CDCl_3$, a mixture of diastereomers): 173.02, 168.27, 158.43, 157.61, 156.28, 155.77, 155.17, 139.26, 136.23, 132.78, 130.05, 129.87, 128.95, 127.91, 127.78, 127.53, 121.46, 120.76, 119.31, 117.75, 117.02, 114.70, 111.53, 83.64, 83.57, 82.60, 82.57, 79.85,

79.18, 71.26, 70.75, 70.66, 70.11, 69.24, 68.75,, 67.83, 67.34, 63.99, 63.93, 60.34, 54.39, 54.32, 40.71, 33.96, 29.71, 29.69, 29.67, 29.47, 29.44, 29.40, 29.31, 29.06, 28.33, 28.23, 27.87, 26.06, 22.73, 22.67. HRMS (ESI-TOF, $[M+Na]^+$): Calcd for $C_{72}H_{108}N_3NaO_{20}P^+$: 1388.7156. Found: 1388.7133.

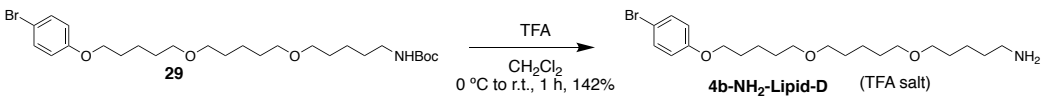
Compound **4d-diNH$_2$**



Compound **45** (70.3 mg, 0.0514 mmol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL) and TFA (1.0 mL) at 0 ºC. The reaction mixture was stirred at room temperature for 2 hours. The mixture was then diluted with CH_2Cl_2 (ca. 4 mL) and evaporated. The residue was subjected to flash column chromatography ($CHCl_3$:MeOH:AcOH=5:1:4 to 1:2:2). The resulting colorless oil was dissolved in CH_2Cl_2 and TFA and vacuum dried at 40 ˚C to afford **4d-diNH$_2$** as the TFA-salt (colorless wax, 57.4 mg, 0.0602 mmol, 117%).

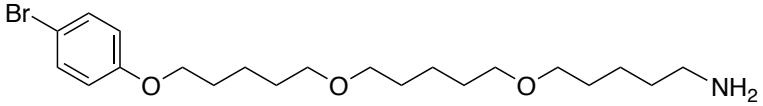
^{1}H NMR (400 MHz, MeOD): 7.551 (2H, d, *J* = 8.7 Hz), 7.519 (2H, d, *J* = 8.8 Hz), 7.387 (1H, dd, *J* = 7.9 Hz, 7.9 Hz), 7.203-7.114 (3H, m), 7.098 (1H, brs), 7.033 (2H, d, *J* = 8.7 Hz), 6.999-6.942 (4H, m), 6.879-6.840 (1H, m), 5.117 (2H, s), 4.258-4.175 (3H, m), 4.091 (2H, t, *J* = 6.5 Hz), 4.030 (2H, t, *J* = 6.3 Hz), 3.991-3.935 (1H, m), 3.891-3.809 (4H, m), 3.665-3.643 (2H, m), 3.542 (2H, t, *J* = 6.8 Hz), 3.480 (2H, t, *J* = 6.2 Hz), 3.456 (2H, t, *J* = 6.3 Hz), 3.260 (2H, dd, *J* = 13.5 Hz, 3.8 Hz), 3.083 (2H, dd, *J* = 13.5 Hz, 7.4 Hz), 2.892 (2H, t, *J* = 7.5 Hz), 2.535 (2H, t, *J* = 7.6 Hz), 1.910-1.790 (4H, m), 1.692-1.531 (8H, m), 1.472-1.394 (2H, m). ^{31}P NMR (161 MHz, MeOD): -0.16. HRMS (ESI-TOF, $[M-H]^-$): Calcd for $C_{49}H_{67}N_3O_{14}P^-$: 952.4366. Found: 952.4382. Anal. Calcd For $C_{49}H_{68}N_3O_{14}P\bullet 4CF_3CO_2H$: C, 48.55; H, 5.15; N, 2.98. Found: C, 48.41; H, 5.22; N, 3.08.

Synthesis of 4b-NH$_2$-Lipid-D (Scheme 8)



Scheme 8. Synthesis of 4b-NH_2-Lipid-D

Compound **4b-NH_2-Lipid-D**



Compound **29** (8.3 mg, 0.016 mmol) was dissolved in anhydrous CH_2Cl_2 (0.5 mL). TFA (0.05 mL) was added at 0 ºC and the reaction mixture was stirred at room temperature for 1 hr. The mixture was diluted with CH_2Cl_2 (ca. 2 mL) and evaporated to afford **4b-NH_2-Lipid-D** as the TFA-salt (colorless oil, 9.5 mg, 0.017 mmol, 142%). ^{1}H NMR (400 MHz, $CDCl_3$): 7.862 (3H, brs), 7.349 (2H, d, J = 9.0 Hz), 6.759 (2H, d, J = 9.0 Hz), 4.048 (1.5H, brs), 3.914 (2H, t, J = 6.5 Hz), 3.441-3.375 (8H, m), 2.936 (2H, brs), 1.787 (2H, quin, J = 7.0 Hz), 1.718-1.334 (16H, m). ^{13}C NMR (100 MHz, $CDCl_3$): 158.17, 132.17, 116.27, 112.58, 70.90, 70.79, 70.67, 70.22, 68.05, 39.77, 29.45, 29.40, 29.39, 28.97, 28.75, 27.15, 22.98, 22.73, 22.70. HRMS (ESI-TOF, $[M+H]^+$): Calcd for $C_{21}H_{37}BrNO_3^+$: 430.1951. Found: 430.1954. Anal. Calcd For $C_{21}H_{36}BrNO_3 \bullet 1.1CF_3CO_2H$: C, 50.13; H, 6.73; N, 2.52. Found: C, 50.30; H, 6.78; N, 2.56.

TGFα shedding assay

TGFα shedding assay was performed as described previously with minor modifications.[12] In this assay, GPCR activation is evaluated in terms of the amount of alkaline phosphatase-tagged TGFα (AP-TGFα) released from the cellular membrane by ecto-domain shedding of its pro-form into the supernatant. Shedding of AP-TGFα occurs efficiently downstream of $G\alpha_q$ signaling, so that a kind of chimeric Gα subunit, $G\alpha_{q/i1}$, was co-expressed with GPR34, which exclusively couples with $G\alpha_{i1}$.

In short, HEK293A cells were suspended in 10% (v/v) fetal bovine serum-containing Dulbecco's modified Eagle's medium at a density of 2×10^5 cells/mL, seeded (10 mL per dish) in 100 mm dishes, and cultured in a CO_2 incubator. After 24 hours, the cells were transfected with a mixture of plasmids (AP-TGFα (2.5 μg), GPR34 (2.5 μg), and $G\alpha_{q/i1}$ subunit (0.5 μg)) in 1 mL of 20 μg/mL Polyethyleneimine Max Solution (Polysciences). For mock-transfected cells, empty plasmid was transfected instead of GPR34-coding plasmid. After incubation for 24 hours, the cells were detached with trypsin, resuspended in 5 mM HEPES (pH 7.4)-containing Hank's balanced salt

solution (HBSS; 60 mL per dish) and seeded in 96-well half-area plates in which compounds and Ki16425 (an antagonist for LPA_1 and LPA_3) were predispensed. After 1.5 hours incubation, the plates were centrifuged (190×g, 2 minutes), and the supernatant was transferred to other 96-well half-area plates. Then, *p*-nitrophenyl phosphate was added to both the supernatant and cell plates, and the absorbance at 405 nm (OD_{405}) was measured with a microplate reader (SpectraMax 340 PC384, Molecular Devices) before and after 30 minutes incubation. In the present study we employed 0.1 μM (0.001%) BSA, which is much lower than the physiological concentration of albumin in plasma (~ 3%).

GPCR activation was calculated as follows.

$$\text{AP-TGF}\alpha\ \text{release}_{\text{Sup}}(\%) = \Delta\text{OD405}_{\text{Sup}}/(\Delta\text{OD405}_{\text{Sup}} + \Delta\text{OD405}_{\text{Cell}})\times 125$$

$$\text{GPCR activation}(\%) = \text{AP-TGF}\alpha\ \text{release}_{\text{Sup}}\ \text{under stimulated conditions}(\%) - \text{AP-TGF}\alpha\ \text{release}_{\text{Sup}}\ \text{under non-stimulated conditions}(\%)$$

$LogEC_{50}$, EC_{50}, and E_{max} values for active compounds were calculated by fitting data to four-parameter sigmoid using Graphpad Prism 6 (Graphpad, USA). RIA (relative intrinsic activity) values[36] were represented as E_{max}/EC_{50} for active compounds relative to that for the control compound.

Generation of GPR34 binding models complexed with compound 1 derivatives

We first docked compound **2a-CH_3** to the GPR34 homology model we previously constructed and then manually modified the hydrophobic terminus of receptor-bound **2a-CH_3** as appropriate in order to create the binding models of other compounds. 3D structure of compound **2a-CH_3** was prepared by using Ligprep (version 35015, Schrödinger 2015-3, Schrödinger, LLC, NY). The output structure having total charge –1 was docked to the GPR34 homology model we previously created (model 1-1)[12] using Glide SP mode (version 68015, Schrodinger 2015-3, Schrödinger, LLC, NY).[13,14] Box center was defined as the centroid of the bound agonist **46** and docking position was constrained as the core shown in Figure S1a. The terminal alkoxy group of the obtained model for **2a-CH_3** was manually modified so that the alkoxy repeat

went toward extracellular part of the membrane (Figure S1b). These modified structures were used for the setup of following molecular dynamics simulation.

MD simulation of GPR34-unbound ligands and receptor-ligand complex

To simulate the insertion of GPR34-unbound LysoPS analogues in the lipid bilayer, the ligand was manually extracted from the ligand-receptor complex model and placed in the POPC membrane solvated with TIP3P water molecules and 0.15 M NaCl. We performed MD simulations in the presence or in the absence of the receptor (Figures S1c and S1d). Simulation was performed by Desmond[24] using OPLS2005 force field in the NPγT ensemble at 310 K and 1.01325 bar and 4000 bar•Å as surface tension using a Langevin thermostat and barostat. The simulation time is either 30 ns (without receptor) or 100 ns (with receptor). The position of *p*-$\mathbf{O}^{D}$ was plotted. The receptor-ligand complexes themselves were also simulated for 100 ns to check the movement of the terminus of the lipid-conjugate part of LysoPS analogues.

The figures were created on the graphical user interface of Maestro (Schrödinger 2018-4, Schrödinger, LLC, NY).

Simulating dissociation of 2a-CH_3 from GPR34 model by SMD

The GPR34-**2a-CH_3** complex (the N and C terminus cap structures were removed for the SMD simulation) was embedded in the POPC membrane (110 Å × 110 Å, 283 molecules) solvated with 23688 molecules of TIP3P water molecules and 0.15 M NaCl by using VMD (version 1.9.4a8).[15] The CHARMM36 force field[16–18] was used and parameters for **2a-CH_3** were obtained from the MATCH webserver.[19] The system was first equilibrated in four stages by NAMD[20] using the constant pressure and temperature algorithm.[21] The SHAKE algorithm[22] and the Smooth Particle Mesh

Ewald (PME) algorithm[23] with nonbonded cutoffs of 12 Å were used. A time integration step was set to 2 fs except the first stage of four equilibration stages (1 fs). To equilibrate the system, first, after 1000 steps of minimization, 0.5 ns MD simulation at 300 K was conducted, in which the atoms except the POPC aliphatic tails were fixed to 'melt' the lipid tails. In the second stage of 0.5 ns of the MD run following 1000 steps of minimization, harmonic constraint was applied to protein and ligand atoms. Then the constraint was released and the system was heated at 310 K in the following 0.5 ns MD simulation (third stage). After these runs, the area in the xy-plane was set constant and 0.5 ns MD simulation was performed (fourth stage).

After equilibration of the system, SMD simulation was performed by using NAMD in NPT ensemble at 310 K with 1.01325 bar using Langevin dynamics. A virtual spring with a spring constant 250 pN/Å was attached to the terminal carbon atom of **2a-CH$_3$** and pulled laterally to the direction between TMs 4 and 5 at a constant velocity of 2.5 Å/ns. The positions of Cα atoms at the internal terminal residues of receptor transmembrane helices (Leu80, Ser88, Ile154, Gln170, Leu246, Arg266 and Met326) were fixed. Simulation for 20 ns was long enough to remove **2a-CH$_3$** from GPR34 model. SMD simulation was conducted five times changing the initial velocity and performed on the supercomputer TSUBAME3.0 at the Tokyo Institute of Technology.

Results/Discussion

Molecular Design of Membrane-lipid-surrogate-conjugated Lysophospholipid Derivatives

We aimed to experimentally evaluate the bioactivity of a series of membrane-lipid-surrogate-conjugated LysoPS analogues by means of TGFα shedding assay[25] with human GPR34 (Figure 2b). In order to simplify the experimental system and synthetic

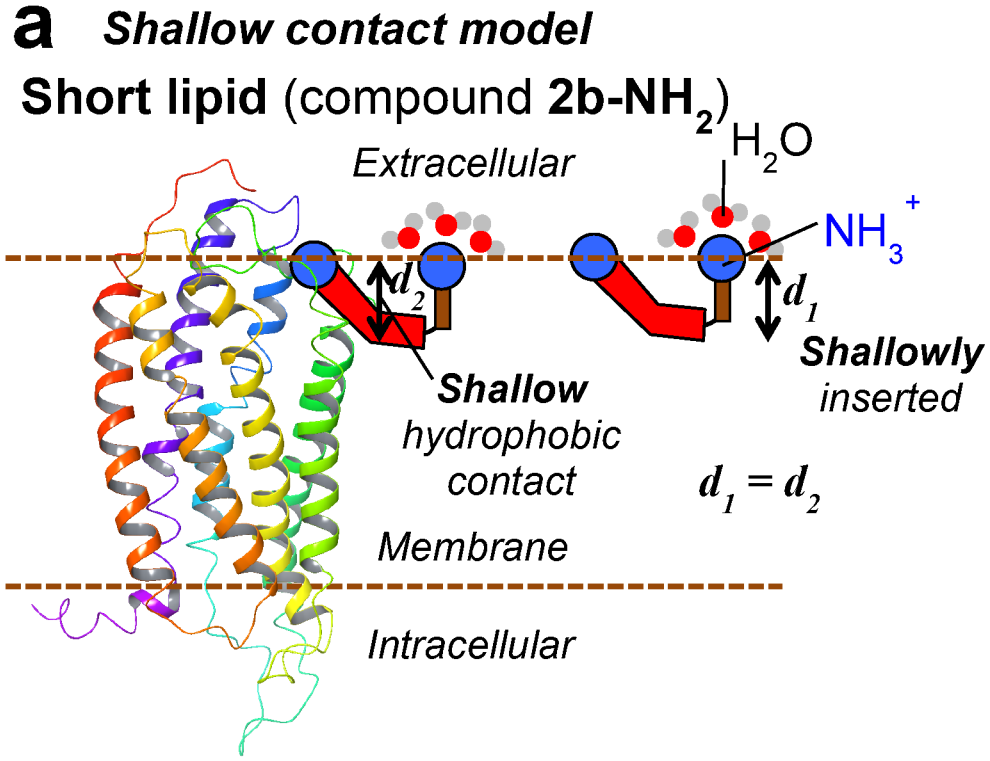
procedures, we used an alkoxy chain repeat, $(OC_5H_{10})_n$, as a mimic of the hydrophobic tail and a terminal amino group (forming an ammonium group upon protonation) as a mimic of the hydrophilic head of phospholipids. This membrane lipid-mimetic moiety was attached to the hydrophobic terminus (ring D) of LysoPS analogue **1** (**2b-NH_2, 3b-NH_2** and **4b-NH_2**, Figure 2b). The use of alkoxy amine chains of different lengths should provide suitable molecular rulers to examine the distance (depth) of the hydrophobic contact point of the LysoPS analogues on the receptor from the surface of the lipid bilayer during membrane approach. The synthetic routes were similar to that previously reported for compound **2a-CH_3**[12]; for details, see the Experimental Section.

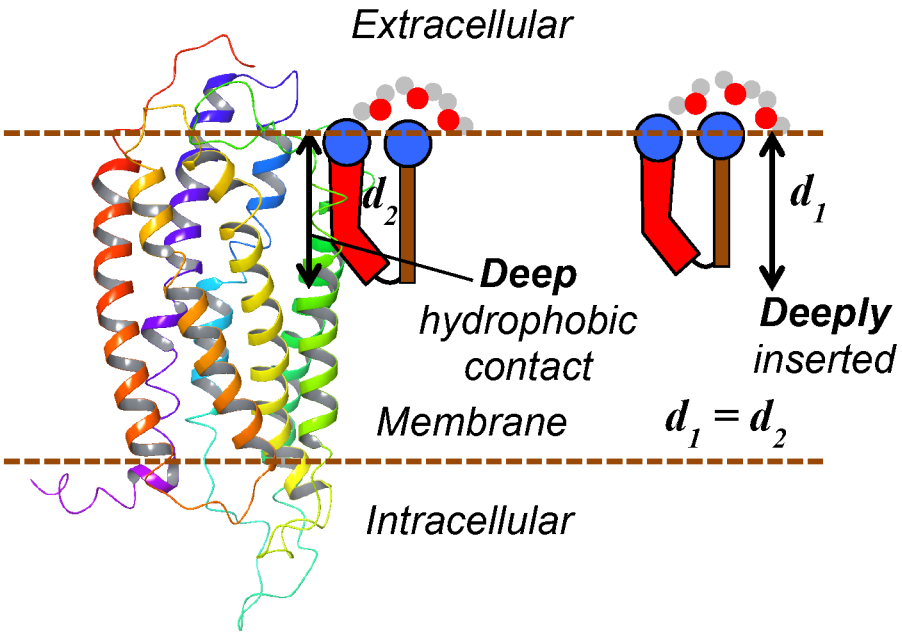
The terminal amino group should remain in the aqueous environment because it would be protonated and solvated by water, and the resultant ammonium ion would not readily enter the lipid bilayer. We examined the behavior of the alkoxy amine chains themselves when attached at the terminus of the LysoPS analogue in a 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine (POPC) membrane bilayer system by means of unbiased MD, and found that these lipid surrogates take extended structures, not folded structures. This result supports the validity of using the alkoxy chain as a ruler for the present purpose. The activity of the analogue should depend on the position of the ligand–receptor hydrophobic contact (d_2), that is, shallow contact (Figure 3a and 3c) and deep contact (Figure 3b). In our system, the length of the conjugate membrane lipid surrogate (d_1), that is, the length of the alkoxy chain also affects the activity (Figure 3). When the chain is short (as in compound **2b-NH_2**), the LysoPS analogue can be only shallowly embedded in the lipid bilayer, assuming that the hydrophilic ammonium moiety is excluded (Figure 3a). Thus, if the compound is active, this indicates that the position of the ligand–receptor hydrophobic contact is

shallow (near the bilayer surface, i.e., $d_1 \sim d_2$ (Figure 3a). If the position of the ligand–receptor hydrophobic contact is deep, the short chain analogue should be inactive.

On the other hand, in the case of a longer chain (as in compound **4b-NH$_2$**), the LysoPS moiety can be inserted deeper into the lipid bilayer (Figure 3b). Thus, as the chain length is increased in this series of compounds, the chain length at which ligand activity appears should correspond to the depth of the hydrophobic contact point on the receptor, that is, in the case of deep contact, $d_1 \sim d_2$ (Figure 3b); in the case of shallow contact, $d_1 > d_2$ (Figure 3c).

The distance (d_1) between the oxygen atom attached to ring D (*p*-**O**^{D}) and the terminal nitrogen atom (**N**) of the amino group is estimated to be 7.36 Å for **2b-NH$_2$**, 14.64 Å for **3b-NH$_2$** and 21.91 Å for **4b-NH$_2$** in the extended (all-*anti*) conformations (Figure S2).


a Shallow contact model
Short lipid (compound 2b-NH2)
Extracellular
H2O
NH3+
d2
d1
Shallow
hydrophobic
contact
Shallowly
inserted
d1 = d2
Membrane
Intracellular


b Deep contact model
Long lipid (compound 4b-NH2)
Extracellular
d2
d1
Deep
hydrophobic
contact
Deeply
inserted
Membrane
d1 = d2
Intracellular

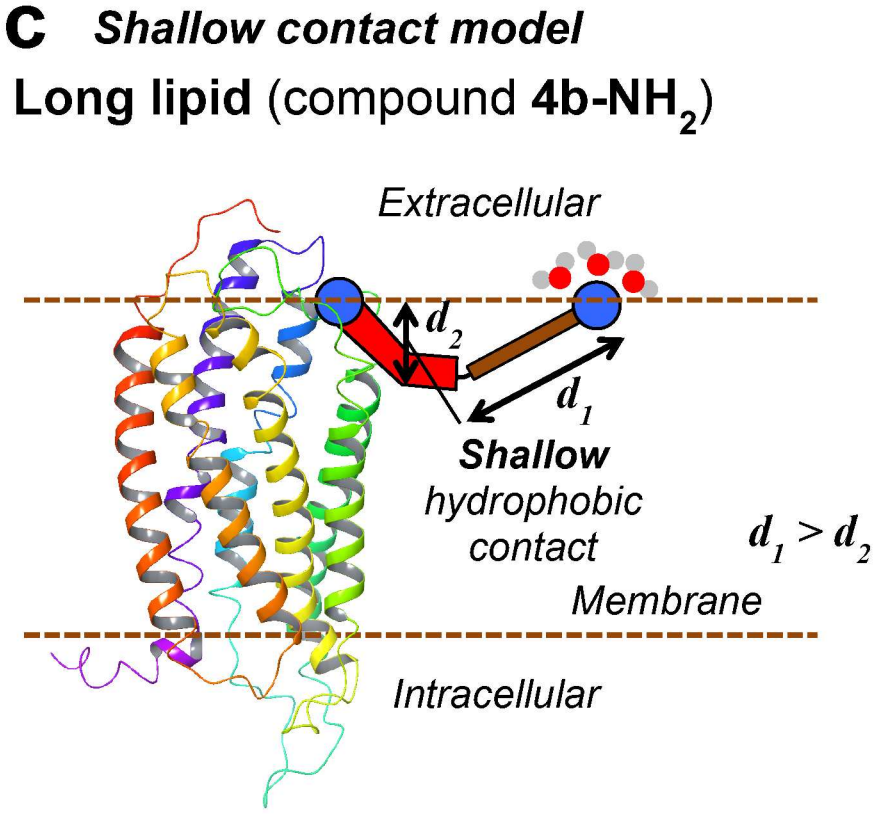


Figure 3. Association of the lipid-surrogate-conjugated LysoPS derivatives with the receptor hydrophobic surface: Shallow contact model and deep contact model. (a) Binding of *short-chain* LysoPS derivative, for which only shallow hydrophobic contact is allowed. A deep contact model is not acceptable. (b) Binding of *long-chain* LysoPS derivative, for which deep hydrophobic contact is allowed. (c) Binding of *long-chain* LysoPS derivative, for which shallow hydrophobic contact is allowed.

Chain-length Dependence of Receptor Activation by the Lysophospholipid Analogues

To validate our experimental approach, we replaced the terminal amino group of the alkoxy amine chain in the conjugate lipids with a methyl group (Figure 2b) and

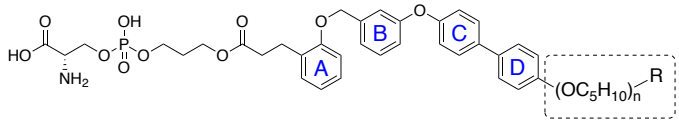
examined the bioactivity of these compounds towards GPR34. We prepared a series of methyl replacement compounds (**2a-CH_3, 3a-CH_3** and **4a-CH_3**) in which the length of the alkoxy repeat ranges from short (OC_5H_{10}) (**2a-CH_3**), to medium ((OC_5H_{10})$_2$) (**3a-CH_3**), to long ((OC_5H_{10})$_3$) (**4a-CH_3**) (Figure 2b), and evaluated their activity (Table 1 and Figure 4). **2a-CH_3** was a potent agonist for GPR34 (EC_{50} = 66 nM), as reported previously.[12] In contrast, the corresponding LysoPS analogue **2b-NH_2**, which contains an amino group in place of the terminal methyl group of **2a-CH_3**, showed no activity toward GPR34 (EC_{50} >> 10 μM) (Table 1 and Figure 4). This indicates that the position of hydrophobic contact should be deeper than the length of the membrane lipid surrogate moiety of **2b-NH_2**.

Compound **3a-CH_3**, having the medium-length alkoxy chain, (OC_5H_{10})$_2$ (EC_{50} = 60 nM), was as active toward GPR34 as **2a-CH_3** (EC_{50} = 66 nM) (Table 1); thus, the additional alkoxy repeat compared to **2a-CH_3** (= **3a-CH_3**) has little effect on the activity of **3a-CH_3** (EC_{50} = 60 nM). On the other hand, compound **3b-NH_2**, which bears an amino group at the terminus, showed marked GPR34 agonistic activity (EC_{50} = 340 nM) compared to **2b-NH_2** (EC_{50} >> 10 μM), though it was still much less potent than **3a-CH_3** (EC_{50} = 60 nM) (Table 1 and Figure 4).

When the alkoxy chain was further elongated from (OC_5H_{10})$_2$ (medium) to (OC_5H_{10})$_3$ (long), we found that **4b-NH_2** showed higher activity (EC_{50} = 31 nM) than **4a-CH_3** (EC_{50} = 200 nM), suggesting that **4b-NH_2** may make more frequent hydrophobic contacts with the surface of the receptor. We confirmed that compound **4b-NH_2-Lipid-D** (Figure 2b), which lacks the LysoPS analogue moiety except for ring D, was inactive toward GPR34. Thus, the membrane-lipid-surrogate part alone did not show agonistic activity toward GPR34.

The experimental structure–activity relationship of LysoPS derivatives bearing conjugated surrogate membrane lipids of different lengths therefore indicates that when the lipidic alkoxy repeat is as long as that of **4b-NH$_2$**, the hydrophobic part of LysoPS can interact with the surface of the receptor sufficiently, enabling the ligand to enter the binding pocket.

Table 1. Activity of lipid-surrogate-conjugated LysoPS analogues towards human GPR34.

	GPR34						
	EC50						
	(LogEC50)						
	[Emax]						
	<RIA>						
	n =						
Positive Control	240 nM	**2a-CH$_3$**	66 nM	**2b-NH$_2$**	>> 10 µM	**4c-NHAc**	2.5 nM
	(-6.61 ± 0.04)	(n = 1, short)	(-7.18 ± 0.20)	(n = 1, short)	(NA)	(n = 3, long)	(-8.60 ± 0.05)
	[21.4 ± 1.2%]	(R = CH$_3$)	[20.3 ± 1.8%]	(R = NH$_2$)	[NA]	(R = NHAc)	[19.9 ± 0.8%]
	<1 ± 0>		<4.2 ± 1.9>		<NA>		<95 ± 19>
	6		3		3		3
18:1-LysoPS	> 3 µM	**3a-CH$_3$**	60 nM	**3b-NH$_2$**	340 nM	**4d-diNH$_2$**	150 nM
	(> -5.5)	(n = 2, medium)	(-7.22 ± 0.04)	(n = 2, medium)	(-6.47 ± 0.06)	(long)	(-6.83 ± 0.08)
	[NA]	(R = CH$_3$)	[18.4 ± 0.8%]	(R = NH$_2$)	[17.7 ± 1.5%]	(see **Figure 2b**)	[20.7 ± 0.6%]
	<NA>		<3.6 ± 0.5>		<0.62 ± 0.11>		<1.7 ± 0.5>
	6		3		3		3
1	24 nM	**4a-CH$_3$**	200 nM	**4b-NH$_2$**	31 nM		
(n = 0)	(-7.61 ± 0.09)	(n = 3, long)	(-6.70 ± 0.02)	(n = 3, long)	(-7.51 ± 0.11)		
(R = H)	[23.7 ± 0.9%]	(R = CH$_3$)	[16.3 ± 1%]	(R = NH$_2$)	[20.6 ± 0.5%]		
	<12 ± 3>		<0.95 ± 0.13>		<8.2 ± 2.4>		
	3		3		3		

EC_{50}, $LogEC_{50}$ (M) and E_{max} (% AP-TGFα release) values are calculated from sigmoidal concentration–response curves and shown as mean ± SEM of the indicated numbers of independent experiments (n). RIA (relative intrinsic activity, a dimensionless parameter)[26] is an E_{max}/EC_{50} value relative to that of a positive control compound (whose RIA value is 1)[27] (Figure S12). "NA" means "not available" because of very low activity

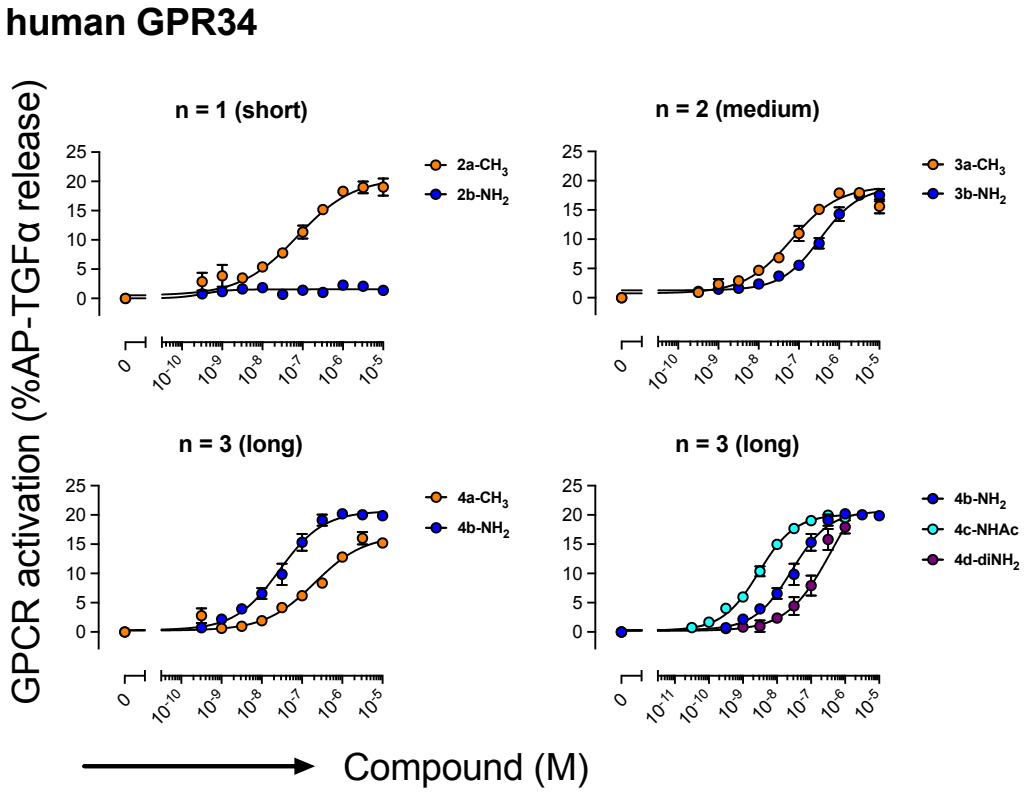


Figure 4. Activity of membrane lipid-surrogate-conjugated LysoPS analogues towards human GPR34. HEK293A cells were transfected with human-GPR34-coding plasmid or an empty plasmid and treated with compounds. Receptor-specific GPCR activation (% AP-TGFα release) was determined by subtracting background responses in the mock-transfected cells (also see Figure S12) from the responses in the GPR34-transfected cells. Data are mean and SEM (standard error of the mean) for 3 independent experiments.

Location of Hydrophobic Terminal Groups of LysoPS Analogues in the Lipid Bilayer

To examine the relationship between the chain length of the membrane lipid surrogate and the depth of the conjugated LysoPS analogues in the lipid bilayer, we performed MD simulations of compounds **2a-CH$_3$**, **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$** in a POPC

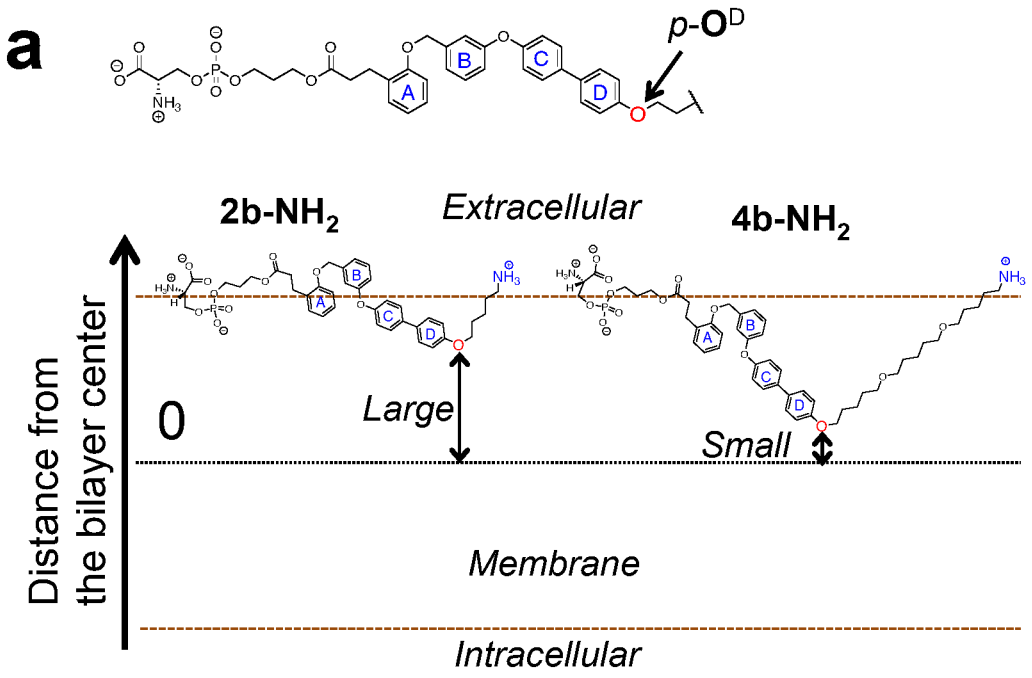
lipid bilayer environment (Figure 5). The terminal amino group was protonated. We monitored the position of the oxygen atom of the aromatic ring D (*p*-$\mathbf{O}^{D}$) of the LysoPS moiety in each analogue, and plotted the distance of the oxygen atom (*p*-$\mathbf{O}^{D}$) from the bilayer center during the MD simulation (Figure 5a); a large distance means that ring D of the LysoPS analogue is inserted shallowly in the lipid bilayer, while when the distance approaches zero, ring D lies close to the center of lipid bilayer (Figure 5b).

MD simulations of these LysoPS analogues in the lipid bilayer indicated that the terminal methyl group of **2a-CH$_3$** moves freely in the lipid bilayer (Figure S3), and the *p*-$\mathbf{O}^{D}$ atom can reach the center of the lipid bilayer (Figure 5b(1)). In contrast, the ionized terminal amino group of receptor-unbound **2b-NH$_2$** remains at the membrane surface (Figure S3) and the distance of the *p*-$\mathbf{O}^{D}$ atom of **2b-NH$_2$** from the bilayer center is greater than that in the case of **2a-CH$_3$** (Figure 5b(2)), indicating that ring D of **2b-NH$_2$** cannot penetrate into the lipid bilayer as deeply as that of **2a-CH$_3$**. This is as expected, based on the molecular design (Figure 2). Indeed, the distance of the *p*-$\mathbf{O}^{D}$ atom in **3b-NH$_2$** (Figure 5b(3)) and **4b-NH$_2$** (Figure 5b(4)) from the bilayer center can occasionally take a value close to zero, indicating that the ring D of **3b-NH$_2$** and **4b-NH$_2$** can be deeply inserted into the lipid bilayer. Thus, these simulations indicate that the hydrophobic moiety of the LysoPS part of **3b-NH$_2$** and **4b-NH$_2$** can approach the deep hydrophobic surface of the receptor without the need for desolvation of water from the terminal hydrophilic group (Figure 3b).

We also conducted MD simulations with monitoring of the position of the *p*-$\mathbf{O}^{D}$ atom of receptor-unbound **2b-NH$_2$, 3b-NH$_2$** and **4b-NH$_2$** in the presence of the receptor, considering the effect of ligand position on remote association with the receptor (Figure S4). There are positive charges on the surface of the receptor around TMs 4, 5

and 6 (Figure S5), and attractive interactions are assumed to exist between these charged surfaces and the hydrophilic ionized part of the ligand. Indeed, in these simulations we found association of the ligand hydrophilic head and the surface of the receptor in all cases. This observation is consistent with the after-mentioned SMD simulation (see Figure 9). About the distance of the *p*-$\mathbf{O}^{D}$ atom from the bilayer center, we obtained similar results to those in the absence of the receptor (Figures 5b (absence of the receptor) and S4 (presence of the receptor)). Therefore, the surface charge of the receptor does not appear to have much influence on the insertion of the LysoPS moiety into the lipid bilayer. In the absence and in the presence of the receptor, the alkoxy amine (for **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$**) takes extended structures. We measured the distance (d_1) between the *p*-$\mathbf{O}^{D}$ atom of the ligand and the terminal nitrogen atom of **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$** in an all-*anti* conformation (Figure S2) and the maximum vertical distance (depth) (d_2) in MD simulation between the *p*-$\mathbf{O}^{D}$ atom of the ligand **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$** and the C_α atom of Leu194 which is located at the tip of TM4, i. e., in the water-lipid boundary, respectively. The length d_1 represents a maximum length of the alkoxy amine chain. The receptor-activation activity of each ligand in terms of RIA (E_{max}/EC_{50} relative to that of the positive reference agonist, Table 1) was plotted against two kinds of distance (d_1 and d_2) (Figure 5c). The two graphs resemble each other, and at short length of the lipid surrogate, the activity is low, and at long length of the lipid surrogate, the activity is high. This is consistent with a deep contact model (Figure 3(b)). The sharp kink in the case of long lipid derivative **4b-NH$_2$** suggested that the depth of the hydrophobic contact region is at around 18-22 Å, which is comparable to the thickness of phospholipids (~21 Å) as a result, that is, a half of the thickness of a lipid bilayer (~40 Å) composed of phospholipid

(dipalmitoylphosphatidylcholine) (hydrophobic thickness: ~30 Å, Luzzati thickness: ~40 Å).[28]



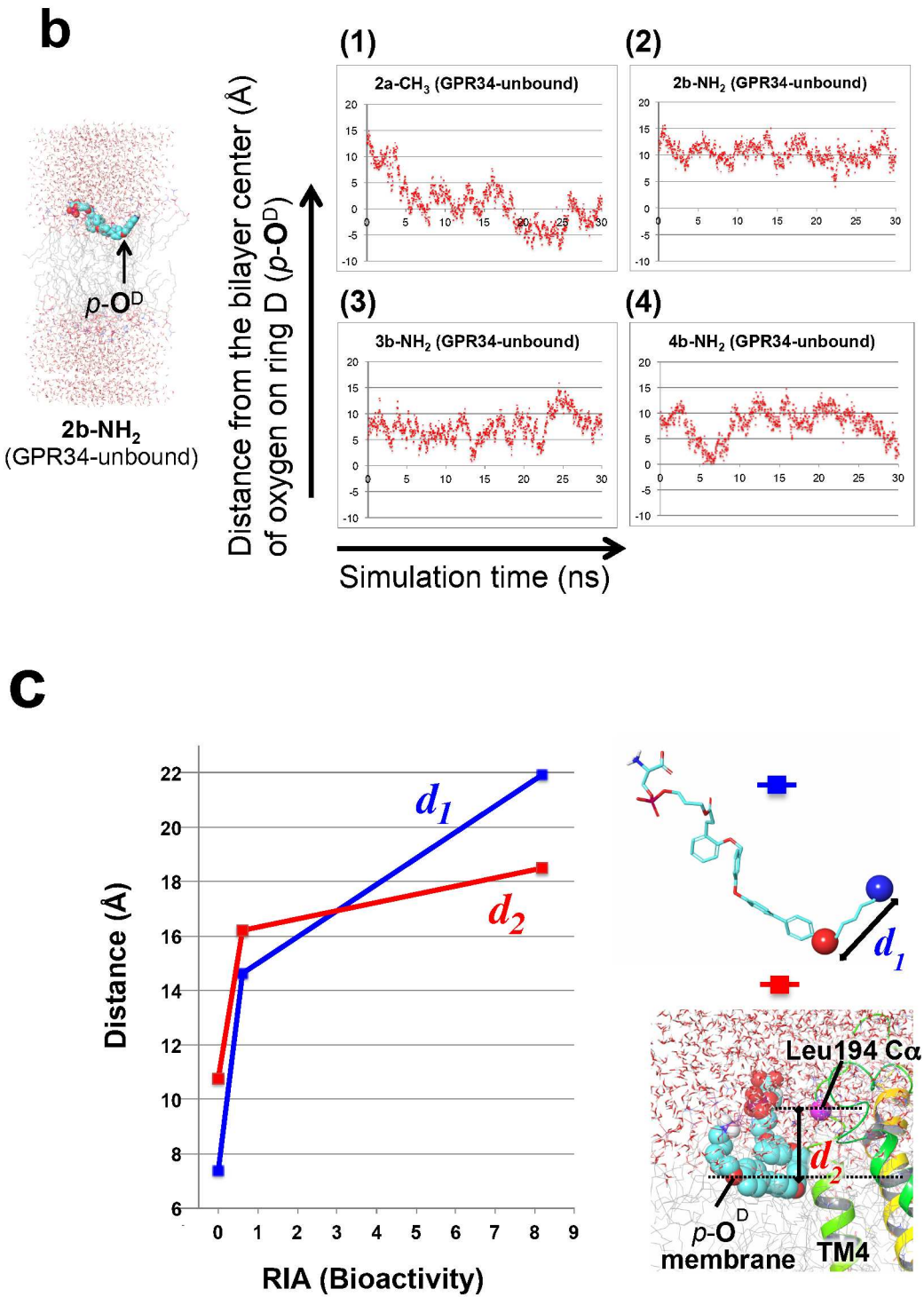


Figure 5. Insertion of membrane lipid-surrogate-conjugated LysoPS analogues into membrane lipids. (a) Schematic view of the insertion of lipid conjugate LysoPS analogues into the lipid bilayer. Distance from the bilayer center (black dotted line) of

oxygen on ring D (*p*-$\mathbf{O}^{D}$) is large (left) when the analogue is shallowly inserted in the lipid bilayer, but small (right) when the analogue is inserted deeply.

(b) The simulated system of receptor-*unbound* **2b-NH_2** (as a representative) is shown (left). The distance from the bilayer center of *p*-$\mathbf{O}^{D}$ during 30 ns MD simulations of this receptor-*unbound* state is plotted. (right, (1) **2a-CH_3**, (2) **2b-NH_2**, (3) **3b-NH_2** and (4) **4b-NH_2**).

(c) The activity of each analogue represented as RIAs (E_{max}/EC_{50} values relative to that of a positive control compound, Table 1) was plotted against two selected distances (d_1 and d_2) shown below the graph—the distance between *p*-$\mathbf{O}^{D}$ atom (red sphere) and the terminal nitrogen atom (blue sphere) in all-*anti* extended conformation (d_1, blue, Figure S2) or the maximum vertical distance between *p*-$\mathbf{O}^{D}$ atom and Cα atom of Leu194, which is located at the tip of TM4 (d_2). The RIA value of **2b-NH_2**, which is inactive toward GPR34, is set to 0.

Depth of the Binding Hydrophobic Moieties in the Binding Poses of LysoPS Analogues with GPR34

While the crystal structure of GPR34 has not been reported yet, a binding model structure of GPR34 and LysoPS analogues has been proposed on the basis of the experimental structure–activity relationship of LysoPS analogues.[12] This model suggests that GPR34 has an opening between TMs 4 and 5 facing the membrane lipids, and the hydrophobic tail of the receptor-bound ligand protrudes from this opening (Figure 6).

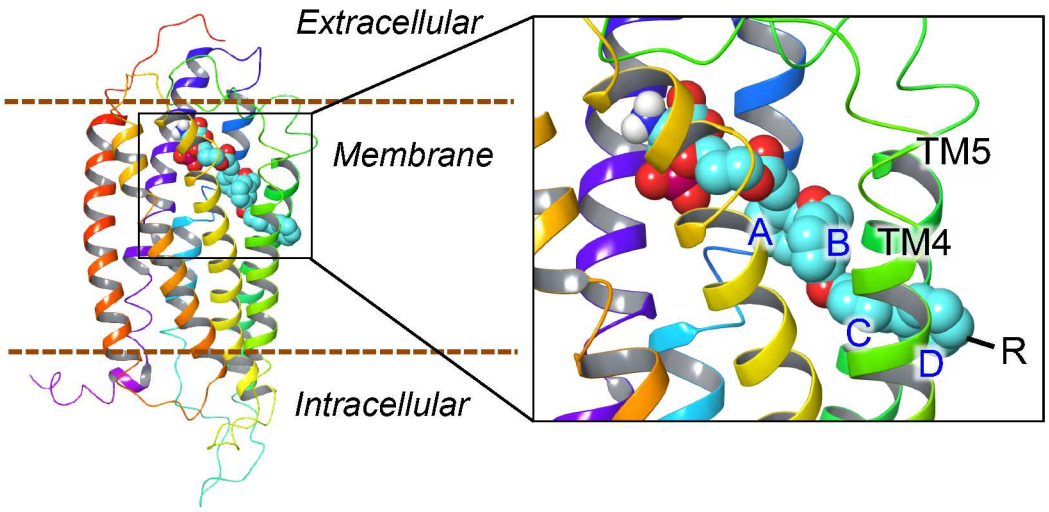


Figure 6. Docking model of **1** (cyan) with GPR34 model (ribbon). The R group attached to ring D will be placed in the lipid bilayer.

To further evaluate the role of the hydrophobic contact region, we created binding models of compounds **2a-CH$_3$**, **4a-CH$_3$**, **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$** to the homology-modeled GPR34 receptor. MD simulation for GPR34-*bound* **2a-CH$_3$** in POPC shows that the *p*-**O**^{D} atom of **2a-CH$_3$** consistently resides near the membrane center (Figure 7); this location should correspond to that of the receptor's hydrophobic pocket, which accommodates ring D of LysoPS analogues. Next, we created binding models of compounds **4a-CH$_3$**, **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$** with GPR34. The conjugated membrane lipid surrogate moiety (alkoxy chain (for **4a-CH$_3$**) or alkoxy amine (for **2b-NH$_2$**, **3b-NH$_2$** and **4b-NH$_2$**)) was manually added in an all-*anti* conformation so that the added surrogates take extended structures and protrude through TM4 and TM5 toward the extracellular region (Figure S1, see Experimental Procedure). Then, the membrane-lipid-surrogate-conjugated LysoPS analogues were embedded into the POPC membrane box and the whole was subjected to 100 ns MD simulation.

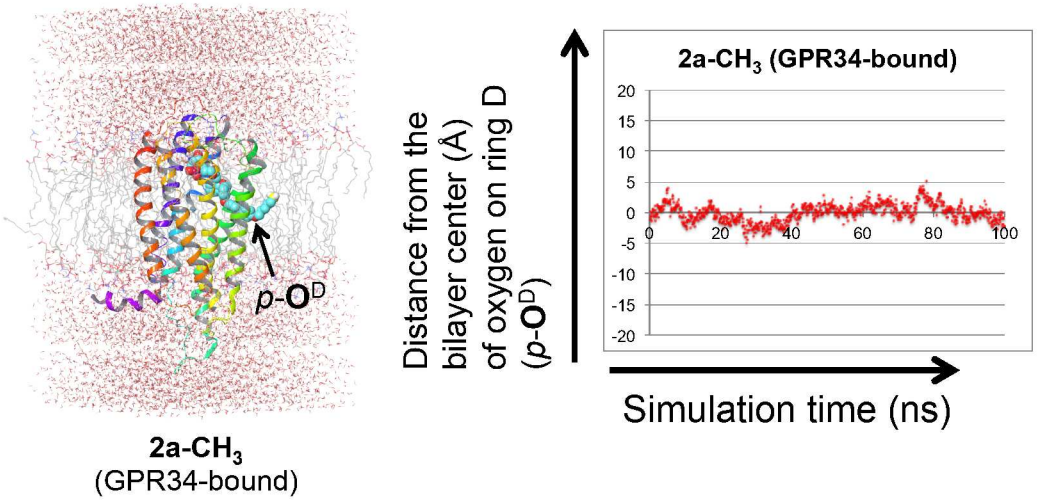


Figure 7. Position of the ligand–receptor hydrophobic contact in the binding model. The simulated system of receptor-*bound* **2a-CH$_3$** is shown (left). The distance from the bilayer center of *p*-**O**^{D} during 100 ns MD simulations of this receptor-*bound* state is plotted (right). The *p*-**O**^{D} atom stayed near the center of the bilayer during simulation.

In the cases of the short and medium alkoxy chains (**2b-NH$_2$** or **3b-NH$_2$**), the protonated amino group was placed in the lipid bilayer before MD simulation (Figure 8). Even after MD simulation, the protonated amino group remained in the lipid bilayer, and as a result, the membrane surface plane was disturbed, and water molecules entered the membrane region to solvate the ammonium group on the time scale of the MD simulations (Figure 8). Also, the *p*-**O**^{D} atom of **2b-NH$_2$** or **3b-NH$_2$** moved up from the membrane center (Figure S6). On the other hand, in the case of a long alkoxy chain (**4b-NH$_2$**), the ammonium group was anchored at the surface of the membrane and the external water molecules solvated the terminal ammonium head group without disturbing the membrane surface plane (Figure 8). The terminal ammonium head group of **4b-NH$_2$** was located stably at the membrane surface and its

p-**O**^{D} atom resides near the membrane center like the position in **2a-CH$_3$** (Figure S6). In contrast, the terminal methyl group of **4a-CH$_3$**, in which the terminal NH_2 is replaced with CH_3, fluctuates within the lipid bilayer (Figure S7).

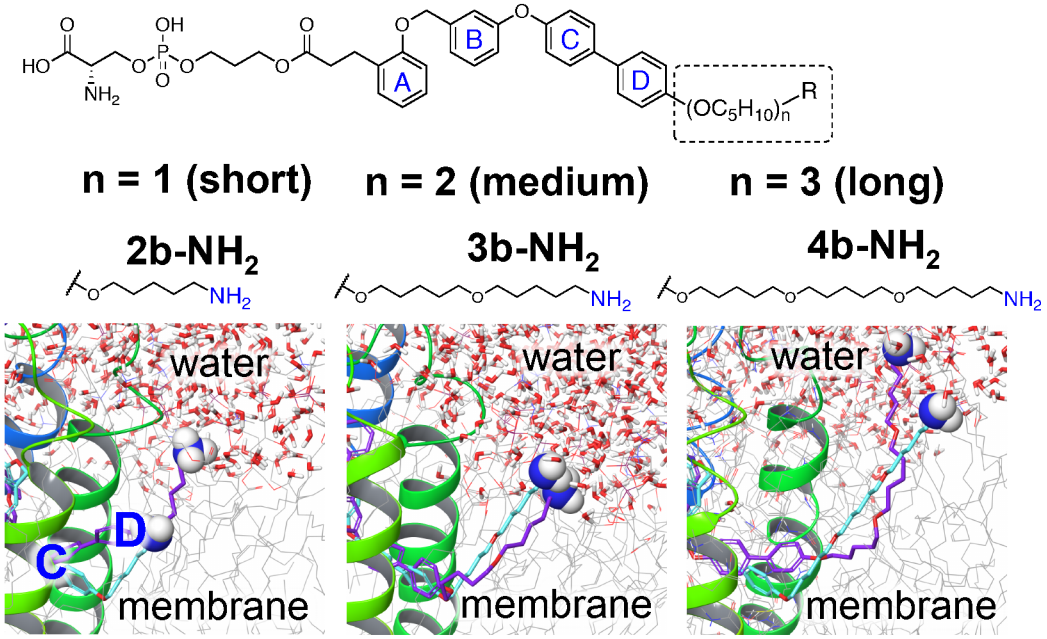


Figure 8. MD simulations of binding models of the membrane lipid-surrogate-conjugated LysoPS analogues.

The images illustrate the receptor TMs 4 and 5 (green ribbons) and membrane-lipid-surrogate conjugate part of the ligand (purple sticks) after 100 ns MD simulation. Lipid bilayer (lines) and water molecules (sticks) in the system are shown. The ligand at the initial position (before MD simulation) is superimposed (cyan sticks). Water molecules entered the membrane region in the simulations of **2b-NH$_2$** and **3b-NH$_2$**.

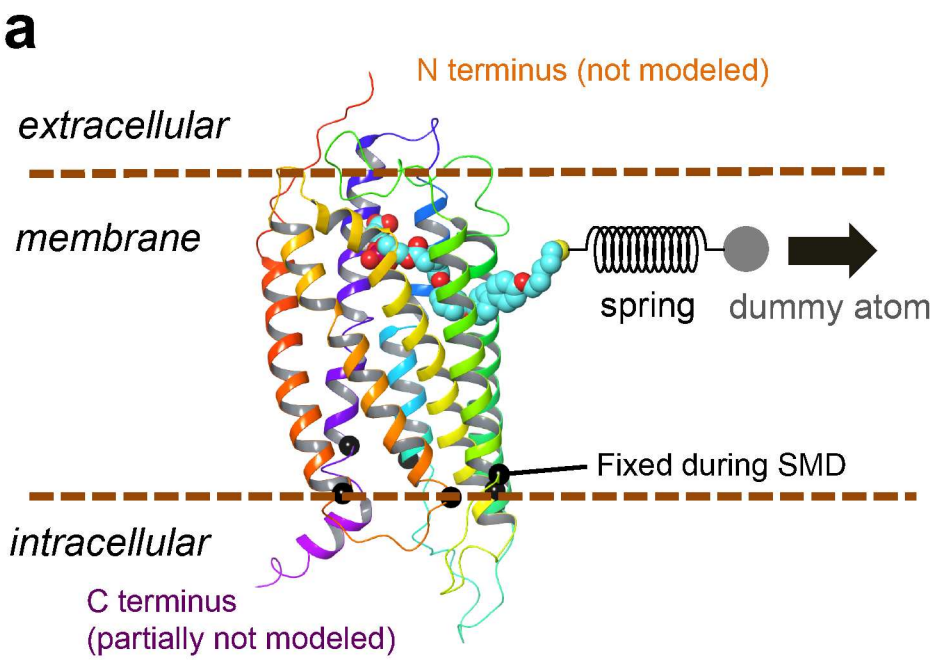
Thus the depth of the protruding aromatic moiety of the LysoPS ligands, i.e., the *p*-**O**^{D} atom, from the membrane surface is estimated to 21 Å (Figure 7). This is consistent with the depth of the hydrophobic contact (18~21 Å). Hydrophobic

residues like Tyr139, Leu181, Phe219, Ile222 and Leu223 forms the hydrophobic contact region on TMs 4 and 5 (Figure S8).

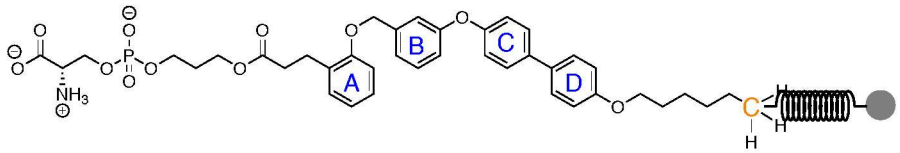
Entry/Exit Pathway of Lysophospholipid Analogues to GPCR

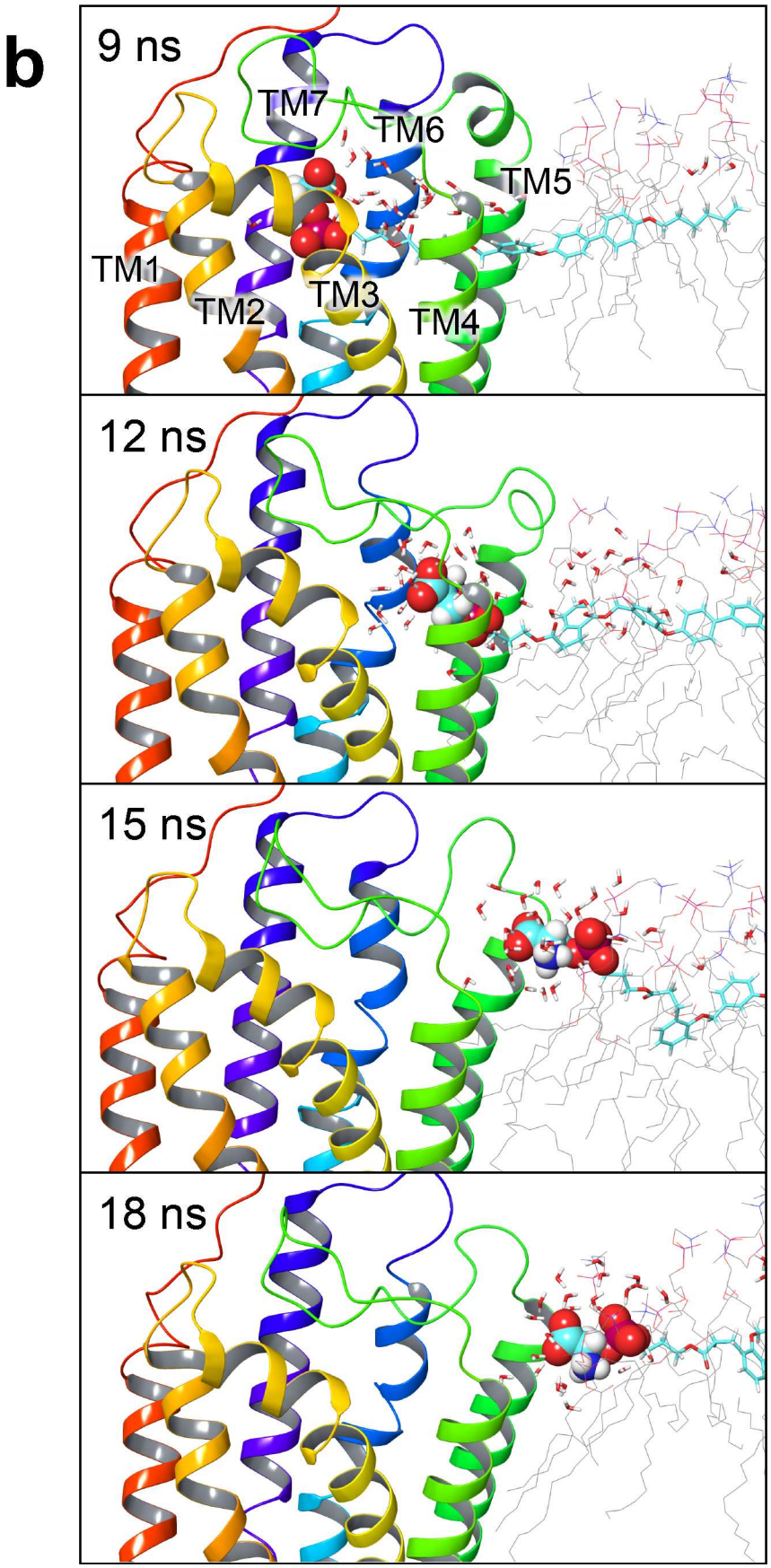
In order to examine the detail of membrane approach at the atomic level, we used steered molecular dynamics (SMD) simulation inspired from single-molecule pull experiments[29–31] to simulate the dissociation process of the ligands through TMs 4 and 5. In these simulations, a dummy atom is attached to a specified atom via a virtual spring (Figure 9a). The dummy atom is pulled in a certain direction at a constant velocity during MD simulations. Starting from the GPR34 receptor-**2a-CH$_3$** complex model, a dummy atom was attached to the terminal carbon atom of **2a-CH$_3$** and used to pull the complex in the direction of TMs 4 and 5 (Figure 9a). Compound **2a-CH$_3$** was successfully dissociated from the receptor (Figure 9b). In the SMD simulations, we can see first of all the detachment of the hydrophobic moiety of the ligand from the receptor embedded in the lipid membrane, followed by detachment of the hydrophilic moieties of the ligand from the surface of the receptor at the water-lipid boundary region (see Supplementary movie). The hydrophilic moiety, which initially interacts with the receptor hydrophilic pocket around TMs 2, 3 and 7, passes between the extracellular tip of TMs 4 and 5 to reach the surface of the membrane (water-lipid boundary) outside the receptor, while retaining its hydration by water molecules. The hydrophobic tail takes an extended structure, because we pulled the terminus of the hydrophobic tail. During five repeated simulations, we detected two energy barriers (at around 9 ns and 18 ns in Figure 9c, also see Figure S9) in terms of the force on the virtual spring (Figure 9a). The first one at 9 ns was the highest, and probably reflects the breaking of polar interactions between the hydrophilic head of **2a-CH$_3$** and the

receptor (Figure 9b). The appearance of the second barrier at 18 ns was unexpected; its magnitude is lower than or similar to that of the first one, and may reflect the breaking of interactions between the hydrophilic head of **2a-CH$_3$** and the polar amino acid residues of TMs 3, 4, 5 and 6 on the dissociation pathway (Figure 9b). Amino acid residues that might directly interact with the hydrophilic head of **2a-CH$_3$** in several SMD simulation runs were identified as Arg208, Lys214, Glu216, Asn220 and Arg286 (Figure S10). Thus, these amino acid residues may constitute a secondary binding site or a hydrophilic contact, located inside or on the surface of the receptor. A similar secondary binding site, called a "vestibule", has been proposed not only in aminergic GPCRs, β_2AR[2,32] and M3 muscarinic acetylcholine receptor,[33] but also in another lipid-liganded GPCR, S1P$_1$,[3] based on binding or dissociation simulations. Notably, the hydrophilic head of the ligand is solvated by water molecules throughout the dissociation process, as well as outside the receptor after dissociation (Figure 9b). Thus, after the ligand dissociates from the receptor, the hydrophilic group of the ligand is favorably moved into an aqueous environment, while the hydrophobic moiety of the ligand remains in the lipid bilayer. While we could not directly simulate the ligand-binding process to GPR34, the present ligand dissociation simulation is consistent with the idea that the opening between TMs 4 and 5 serves as a point of entry for the ligand to reach its binding site, and there is a hydrophilic contact region between the ligand and the receptor.


a
N terminus (not modeled)
extracellular
membrane
spring
dummy atom
Fixed during SMD
intracellular
C terminus
(partially not modeled)

2a-CH_3


A
B
C
D


b
9 ns
TM7
TM6
TM5
TM1
TM2
TM3
TM4
12 ns
15 ns
18 ns

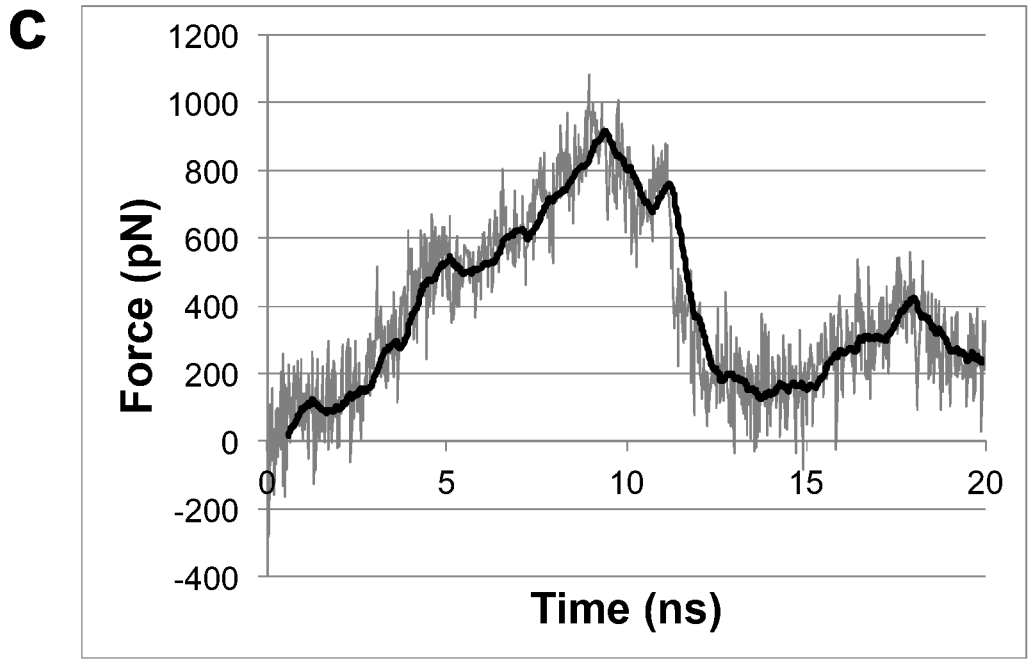


Figure 9. Simulation of the unbinding process of **2a-CH$_3$** by SMD. (a) Setup of SMD simulations (side view). The binding model of GPR34 (ribbon) and **2a-CH$_3$** (spheres) after equilibration MD is shown. A yellow-colored atom in **2a-CH$_3$** is the carbon atom to which a virtual spring was attached. The black spheres on the GPR34 model represent the Cα atoms whose positions were fixed during SMD simulation. (b) Snapshots from SMD simulation showing the extracellular part of the receptor (ribbon) and dissociating **2a-CH$_3$** (cyan sticks and spheres). For the POPC membrane (lines) and water (sticks), only the molecules within 5 Å from **2a-CH$_3$** are shown. The atoms forming the hydrophilic head of **2a-CH$_3$** are emphasized as spheres. Water molecules hydrated the hydrophilic head of **2a-CH$_3$** during the ligand dissociation process. (c) Plot of force on the virtual spring during the SMD simulation in which **2a-CH$_3$** was pulled toward the membrane. Moving average (period 0.6 ns) is shown as a black line.

Effect of Terminal Polar Head Group of the Conjugated Membrane Lipid Surrogate

As we expected, a conjugated membrane lipid surrogate with a long alkoxy repeat, $(OC_5H_{10})_3$, is crucial for anchoring the terminal amine functionality to the membrane surface. Furthermore, such a terminal hydrophilic head group of the conjugated surrogate membrane lipid might increase the GPR34-activating activity, since **4b-NH$_2$** (EC_{50} = 31 nM) showed higher activity than **4a-CH$_3$** (EC_{50} = 200 nM) (Table 1 and Figure 4). Therefore, we next introduced other terminal groups in place of the amino group to examine in more detail the effect of the terminal hydrophilic head group of the conjugated surrogate membrane lipid (Figure 2b, **4c-NHAc** and **4d-diNH$_2$**). Because an amino group can be protonated and has a positive charge, we examined the effect of the magnitude of the positive charge at the lipid terminus. We synthesized **4c-NHAc**, in which the amino group of **4b-NH$_2$** is acetylated, and **4d-diNH$_2$**, which has two amino groups at the terminus (Figure 2b). We expected that anchoring of the terminal hydrophilic head group of the conjugate membrane lipid surrogate to the membrane surface during the ligand binding might serve to increase the agonistic activity toward GPR34. Thus, we expected that the diamino derivative might show higher activity because the diamino groups would anchor the LysoPS analogue more strongly to the membrane surface, whereas the N-acetyl derivative might show reduced activity compared to **4b-NH$_2$** due to its lack of a positive charge. But, unexpectedly, diamine **4d-diNH$_2$** (EC_{50} = 150 nM) showed lower activity than monoamine **4b-NH$_2$** (EC_{50} = 31 nM), and N-acetamide **4c-NHAc** (EC_{50} = 2.5 nM) showed higher activity than monoamine **4b-NH$_2$** (EC_{50} = 31 nM) (Table 1 and Figure 4).

To understand these observations at the molecular level, we carried out MD simulations of model complexes of the whole membrane-lipid-surrogate-conjugated LysoPS derivatives (**4c-NHAc** and **4d-diNH$_2$**) with GPR34 embedded in the lipid

bilayer, focusing on the behavior of the terminal hydrophilic groups (Figure 10). We found that the terminal N-acetamide in **4c-NHAc** interacts favorably with the positively charged receptor surface (Figure 10a). On the other hand, the diamino group of **4d-diNH$_2$** tends to move away from the receptor surface, probably because of Coulombic repulsion between the positive charges of **4d-diNH$_2$** and the positively charged receptor surface (Figure 10b). We also carried out MD simulations of the isolated membrane lipid surrogate moieties, N-acetyl alkoxy chain (**4c-NHAc-Lipid**) and diamine alkoxy chain (**4d-diNH$_2$-Lipid**) (Figure 2b), in the presence of GPR34 embedded in the lipid bilayer (Figure S11). Similarly to the simulations of the whole membrane-lipid-surrogate-conjugated LysoPS derivatives, only the conjugated membrane lipid surrogate part of **4c-NHAc-Lipid** stayed on the receptor surface, while the diamino lipid part of **4d-diNH$_2$-Lipid** tended to move away from the receptor surface. Thus, these MD simulations reproduced the affinity of the N-acetamide-bearing lipid mimic **4c-NHAc** for the receptor surface, as well as the dissociation of the diamino derivative **4d-diNH$_2$** from the receptor surface. These results suggest that not only the hydrophobic group of the membrane-lipid-surrogate-conjugated lipid, but also the hydrophilic part plays a role in guiding LysoPS analogue binding to the receptor, not just in anchoring it to the membrane surface. Dependence of ligand activity on the phospholipid head group has been proposed for β_2AR[34] and membrane lipids might control receptor activation by interacting with the intracellular region, such as IL3, of the receptor, as judged from long-time-scale MD simulation.[35,36] Our experimental findings that membrane-lipid-surrogate-conjugated LysoPS analogues show different activities depending on the head group of the conjugated lipid suggest that binding of lipid ligands to GPCRs could be controlled by

the interaction between the membrane lipid and the charge on the extracellular part of the receptor.

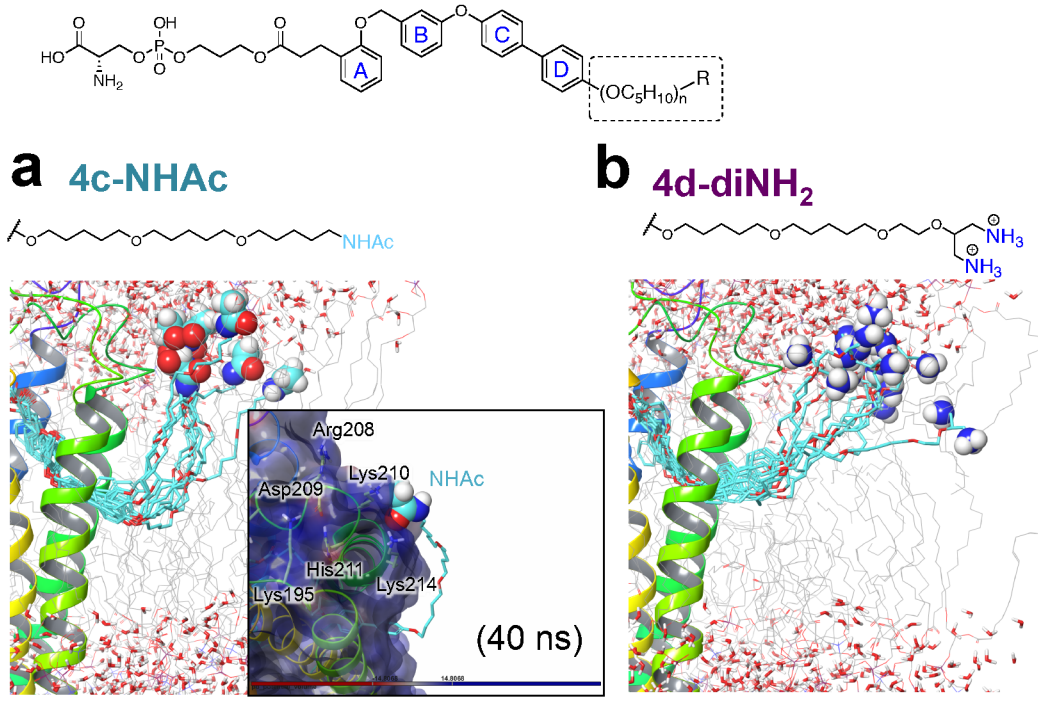


Figure 10. Movement of the terminal R-group attached at the conjugate lipid part having a long alkoxy repeat.

Side views of the binding models of GPR34-bound (a) **4c-NHAc** (b) **4d-diNH$_2$** (receptor: ribbon, ligand: sticks and spheres). Lipid bilayer (lines) and water molecules (sticks) in the system are also shown. Ten ligand structures were extracted every 10 ns from the 100 ns MD trajectory and superimposed on the initial system. The terminal group of the ligand is shown by spheres (red: oxygen, blue: nitrogen, white: hydrogen). The interaction of the terminal N-acetamide group with the receptor positively charged surface (colored blue, Poisson-Boltzmann electrostatic potential surface calculated by using Maestro function (Schrödinger 2018-4, Schrödinger, LLC, NY)) taken from the snapshot after 40 ns (a, inset) is shown. There is a cluster of positively charged residues near the extracellular terminus of TM5 and extracellular loop 2.

Conclusion

Experimentally, we found that the presence of a short chain membrane lipid surrogate in the hydrophobic tail of LysoPS analogue **1** (compound **2b-NH_2**) reduced the ability to activate GPR34. On the other hand, introduction of a long chain membrane lipid surrogate (compound **4b-NH_2**) resulted in similar activity toward GPR34, as compared with the prototype **1**. These experimental observations can be rationalized in terms of the postulate that the conjugated membrane lipid surrogate determines the ability of the hydrophobic moiety of LysoPS to make contact with the deep receptor surface, which is required for entry into the binding pocket. The lack of activity of **2b-NH_2** can be attributed to insufficient hydrophobic contact because the hydrophobic moiety of **2b-NH_2** is only shallowly inserted into the lipid bilayer if the amino group remains on the membrane surface (Figure 3a). On the other hand, the hydrophobic moiety of **4b-NH_2** can enter more deeply than **2b-NH_2** into the lipid bilayer, even when the amino group remains on the surface (Figure 3b). Therefore we can estimate the position of the hydrophobic contact of the LysoPS ligand as being near the middle of the lipid bilayer (at 18~21 Å from the surface of the membrane lipid). This is consistent with the depth of the terminal hydrophobic moiety of the LysoPS analogue observed in the binding pose (~21 Å). The SMD simulation result indicated that the unbinding pathway of **2a-CH_3** from the receptor to the membrane region, passing between TMs 4 and 5, is feasible. There is also a secondary binding pocket, or a hydrophilic contact located between the original binding pocket and the surface of the receptor. Positively charged residues at this secondary binding site may contribute to guiding the hydrophilic site of LysoPS to the hydrophilic site of the receptor. This would account for the high activity of **4c-NHAc**, which has an acetamide moiety as the hydrophilic head of the conjugated membrane lipid surrogate.

Furthermore, the ligand binding pose closely matches the shape of the contact surface, possibly because of the long size of LysoPS molecule and the narrowness of the passage inside the GPR34 receptor.[12]

Therefore, the scenario of membrane approach may involve hydrophilic contact of the polar sites of the ligand on the positively-charged receptor surface at the lipid-water boundary, followed by the hydrophobic contact of the hydrophobic moiety of the ligand on the surface of the receptor at the middle of membrane bilayer along the opening between TMs 4 and 5. Consequently the whole ligand leans into the binding pocket of the receptor through the opening between TMs 4 and 5. The contact conformation and binding conformation are very similar.

This strategy that a conjugated membrane lipid surrogate can measure the depth (position) of hydrophobic contact of LysoPS to the opening of the receptor surface might also be applicable to other lysophospholipid GPCRs, such as S_1P, LPA and lysophosphatidylinositol (LPI), and other hydrophobic endogenous ligands such as prostaglandins and leukotrienes, and fatty acids. These findings should be helpful for the rational design of length and shape of the hydrophobic moiety of functional lipid ligands.

References and Notes

(1) Brinkmann, V., Billich, A., Baumruker, T., Heining, P., Schmouder, R., Francis, G., Aradhye, S., and Burtin, P. (2010) Fingolimod (FTY720): Discovery and development of an oral drug to treat multiple sclerosis. *Nat. Rev. Drug Discov. 9*, 883–897.
(2) Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., Xu, H., and Shaw, D. E. (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A. 108*, 13118–13123.
(3) Stanley, N., Pardo, L., and Fabritiis, G. De. (2016) The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep. 6*, 22639.
(4) Srivastava, A., Yano, J., Hirozane, Y., Kefala, G., Gruswitz, F., Snell, G., Lane, W., Ivetac, A., Aertgeerts, K., Nguyen, J., Jennings, A., and Okada, K. (2014)

receptor bound to allosteric agonist TAK-875. *Nature 513*, 124–127.
(5) Lu, J., Byrne, N., Wang, J., Bricogne, G., Brown, F. K., Chobanian, H. R., Colletti, S. L., Di Salvo, J., Thomas-Fowlkes, B., Guo, Y., Hall, D. L., Hadix, J., Hastings, N. B., Hermes, J. D., Ho, T., Howard, A. D., Josien, H., Kornienko, M., Lumb, K. J., Miller, M. W., Patel, S. B., Pio, B., Plummer, C. W., Sherborne, B. S., Sheth, P., Souza, S., Tummala, S., Vonrhein, C., Webb, M., Allen, S. J., Johnston, J. M., Weinglass, A. B., Sharma, S., and Soisson, S. M. (2017) Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat. Struct. Mol. Biol. 24*, 570–577.
(6) Taniguchi, R., Inoue, A., Sayama, M., Uwamizu, A., Yamashita, K., Hirata, K., Yoshida, M., Tanaka, Y., Kato, H. E., Nakada-Nakura, Y., Otani, Y., Nishizawa, T., Doi, T., Ohwada, T., Ishitani, R., Aoki, J., and Nureki, O. (2017) Structural insights into ligand recognition by the lysophosphatidic acid receptor LPA_6. *Nature 548*, 356–360.
(7) Hurst, D. P., Grossfield, A., Lynch, D. L., Feller, S., Romo, T. D., Gawrisch, K., Pitman, M. C., and Reggio, P. H. (2010) A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem. 285*, 17954–17964.
(8) Jakowiecki, J., and Filipek, S. (2016) Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. *J. Chem. Inf. Model. 56*, 2457–2466.
(9) Toyoda, Y., Morimoto, K., Suno, R., Horita, S., Yamashita, K., Hirata, K., Sekiguchi, Y., Yasuda, S., Shiroishi, M., Shimizu, T., Urushibata, Y., Kajiwara, Y., Inazumi, T., Hotta, Y., Asada, H., Nakane, T., Shiimura, Y., Nakagita, T., Tsuge, K., Yoshida, S., Kuribara, T., Hosoya, T., Sugimoto, Y., Nomura, N., Sato, M., Hirokawa, T., Kinoshita, M., Murata, T., Takayama, K., Yamamoto, M., Narumiya, S., Iwata, S., and Kobayashi, T. (2019) Ligand binding to human prostaglandin E receptor EP_4 at the lipid-bilayer interface. *Nat. Chem. Biol. 15*, 18–26.
(10) Sugo, T., Tachimoto, H., Chikatsu, T., Murakami, Y., Kikukawa, Y., Sato, S., Kikuchi, K., Nagi, T., Harada, M., Ogi, K., Ebisawa, M., and Mori, M. (2006) Identification of a lysophosphatidylserine receptor on mast cells. *Biochem. Biophys. Res. Commun. 341*, 1078–1087.
(11) Kitamura, H., Makide, K., Shuto, A., Ikubo, M., Inoue, A., Suzuki, K., Sato, Y., Nakamura, S., Otani, Y., Ohwada, T., and Aoki, J. (2012) GPR34 is a receptor for lysophosphatidylserine with a fatty acid at the *sn*-2 position. *J. Biochem. 151*, 511–518.
(12) Sayama, M., Inoue, A., Jung, S., Ikubo, M., Otani, Y., Uwamizu, A., Kishi, T., Makide, K., Aoki, J., Hirokawa, T., and Ohwada, T. (2017) Probing the hydrophobic binding pocket of G protein-coupled lysophosphatidylserine receptor GPR34/LPS1 by docking-aided structure−activity analysis. *J. Med. Chem. 60*, 6384–6399.
(13) Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem. 47*, 1739–1749.
(14) Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem. 47*, 1750–1759.
(15) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. 14*, 33–38.
(16) Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L.,

Grubmüller, H., and MacKerell Jr, A. D. (2016) CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods 14*, 71–73.
(17) Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell Jr, A. D., and Pastor, R. W. (2010) Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B 114*, 7830–7843.
(18) Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 Dihedral Angles. *J. Chem. Theory Comput. 8*, 3257–3273.
(19) Yesselman, J. D., Price, D. J., Knight, J. L., and Brooks, C. L. (2012) MATCH: An atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem. 33*, 189–202.
(20) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802. The equilibration protocol was optimized referring to NAMD Membrane Protein Tutorial (http://www.ks.uiuc.edu/Training/Tutorials/). Minimization and following 1000 steps of the MD simulation were conducted using NAMD version 2.12 and the other simulations were conducted using NAMD version 2.13 installed on the supercomputer TSUBAME3.0 at the Tokyo Institute of Technology.
(21) Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684–3690.
(22) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys. 23*, 327–341.
(23) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089–10092.
(24) Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. (2007) Scalable algorithms for molecular dynamics simulations on commodity clusters. *ACM/IEEE SC 2006 Conf.*
(25) Inoue, A., Ishiguro, J., Kitamura, H., Arima, N., Okutani, M., Shuto, A., Higashiyama, S., Ohwada, T., Arai, H., Makide, K., and Aoki, J. (2012) TGFα shedding assay: An accurate and versatile method for detecting GPCR activation. *Nat. Methods 9*, 1021–1029.
(26) Ehlert, F. J., Griffin, M. T., Sawyer, G. W., and Bailon, R. (1999) A simple method for estimation of agonist activity at receptor subtypes: Comparison of native and cloned M_3 muscarinic receptors in guinea pig ileum and transfected cells. *J. Pharmacol. Exp. Ther. 289*, 981–992.
(27) Jung, S., Inoue, A., Nakamura, S., Kishi, T., Uwamizu, A., Sayama, M., Ikubo, M., Otani, Y., Kano, K., Makide, K., Aoki, J., and Ohwada, T. (2016) Conformational constraint of the glycerol moiety of lysophosphatidylserine affords compounds with receptor subtype selectivity. *J. Med. Chem. 59*, 3750–3776.
(28) Nagle, J. F., and Tristram-Nagle, S. (2000) Lipid bilayer structure. *Curr. Opin. Struct. Biol. 10*, 474–480.
(29) Grubmüller, H., Heymann, B., and Tavan, P. (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science 271*, 997–999.
(30) Park, S., and Schulten, K. (2004) Calculating potentials of mean force from

steered molecular dynamics simulations. *J. Chem. Phys. 120*, 5946–5961.
(31) Li, M. S., and Mai, B. K. (2012) Steered molecular dynamics-a promising tool for drug design. *Curr. Bioinform. 7*, 342–351.
(32) González, A., Perez-Acle, T., Pardo, L., and Deupi, X. (2011) Molecular basis of ligand dissociation in β-adrenergic receptors. *PLoS One 6*, e23815.
(33) Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., Green, H. F., Liu, T., Chae, P. S., Dror, R. O., Shaw, D. E., Weis, W. I., Wess, J., and Kobilka, B. K. (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature 482*, 552–556.
(34) Dawaliby, R., Trubbia, C., Delporte, C., Masureel, M., Van Antwerpen, P., Kobilka, B. K., and Govaerts, C. (2016) Allosteric regulation of G protein-coupled receptor activity by phospholipids. *Nat. Chem. Biol. 12*, 35–39.
(35) Bruzzese, A., Gil, C., Dalton, J. A. R., and Giraldo, J. (2018) Structural insights into positive and negative allosteric regulation of a G protein-coupled receptor through protein-lipid interactions. *Sci. Rep. 8*, 4456.
(36) Díaz, Ó., Dalton, J. A. R., and Giraldo, J. (2019) Revealing the mechanism of agonist-mediated cannabinoid receptor 1 (CB1) activation and phospholipid-mediated allosteric modulation. *J. Med. Chem. 62*, 5638–5654.

ASSOCIATED CONTENT

Supporting Information Available:

Supporting Figures S1-12

Supporting Movie

These materials are available free of charge via the internet at http://pubs.asc.org.

AUTHOR INFORMATION

M. Sayama., Y. O. and T. O. (The University of Tokyo) performed chemical studies including design and synthesis of compounds. A. I., A. U. and J. A. (Tohoku University) performed biological studies including TGFα shedding assay. M. Sayama (The University of Tokyo) and M. Sekijima (Tokyo Institute of Technology) performed computational studies.

Corresponding Authors

Chemical study: For T.O.: phone, +81-3-5841-4730; fax, +81-3-5841-4735;
E-mail, ohwada@mol.f.u-tokyo.ac.jp. Laboratory of Organic and Medicinal Chemistry, Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.

Biological study: For J.A.: phone, +81-22-795-6860; fax, +81-22-795-6859;

E-mail, jaoki@m.tohoku.ac.jp. Laboratory of Molecular and Cellular Biochemistry, Graduate School of Pharmaceutical Sciences, Tohoku University, 6-3, Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980-8578, Japan

NOTES

The authors declare no competing financial interests.

ACKNOWLEDGMENTS

The SMD simulation was performed on the supercomputer TSUBAME3.0, supported by TSUBAME Encouragement Program for Young/Female Users of Global Scientific Information and Computing Center at Tokyo Institute of Technology (M. Sayama).

This work was supported by JSPS KAKENHI Grant Numbers JP 26104508 and 16K08157 (Y. Otani), 17K08264 (A. Ionue), and 26293002 (T. Ohwada). Also M. Sayama thanks a Research Fellowship for Young Scientists from JSPS (Grant number 16J08059).

This research was partially supported by Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) Grant Number JP18am0101095, 19ck0106258 (J. Aoki), and 19am0101112 (support number 0365) (M. Sekijima), the PRIME Grant number JP18gm5910013 (A. Inoue), and the LEAP Grant number JP19gm0010004 (A. Inoue and J. Aoki) from Japan Agency for Medical Research and Development (AMED).

For Table of Contents use only