Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2701283 C 2014/07/15

(11)(21) 2 701 283

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2008/10/22
(87) Date publication PCT/PCT Publication Date: 2009/04/30
(45) Date de délivrance/Issue Date: 2014/07/15
(85) Entrée phase nationale/National Entry: 2010/03/30
(86) N° demande PCT/PCT Application No.: US 2008/080847
(87) N° publication PCT/PCT Publication No.: 2009/055512
(30) Priorités/Priorities: 2007/10/22 (US60/999,951); 2007/10/24 (US61/000,325); 2008/06/09 (US61/060,107); 2008/10/01 (US61/194,961)

(51) Cl.Int./Int.Cl. *C07K 1/02* (2006.01), *C07K 1/00* (2006.01), *C07K 14/705* (2006.01), *G01N 33/68* (2006.01)

(72) Inventeurs/Inventors:
STEVENS, RAYMOND C., US;
HANSON, MICHAEL A., US;
CHEREZOV, VADIM, US;
KUHN, PETER, US

(73) Propriétaire/Owner:
THE SCRIPPS RESEARCH INSTITUTE, US

(74) Agent: FASKEN MARTINEAU DUMOULIN LLP

(54) Titre : PROCEDES ET COMPOSITIONS POUR OBTENIR DES CRISTAUX HAUTE RESOLUTION DE PROTEINES MEMBRANAIRES

(54) Title: METHODS AND COMPOSITIONS FOR OBTAINING HIGH-RESOLUTION CRYSTALS OF MEMBRANE PROTEINS

(57) Abrégé/Abstract:
The invention describes compositions and method useful for the crystallization of membrane proteins.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 April 2009 (30.04.2009)

PCT

(10) International Publication Number
WO 2009/055512 A3

(51) International Patent Classification:
G01N 1/00 (2006.01)

(21) International Application Number:
PCT/US2008/080847

(22) International Filing Date:
22 October 2008 (22.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/999,951	22 October 2007 (22.10.2007)	US
61/000,325	24 October 2007 (24.10.2007)	US
61/060,107	9 June 2008 (09.06.2008)	US
61/194,961	1 October 2008 (01.10.2008)	US

(71) Applicant *(for all designated States except US)*: **THE SCRIPPS RESEARCH INSTITUTE** [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **STEVENS, Raymond, C.** [US/US]; 7096 Caminito Valverde, La Jolla, CA 92037 (US). **HANSON, Michael, A.** [US/US]; 1218 Avenida Amistad, San Marcos, CA 92069 (US). **CHEREZOV, Vadim** [RU/US]; 12604 Torrey Bluff Dr., Apt. 399, San Diego, CA 92130 (US). **KUHN, Peter** [DE/US]; 708 Barbara Ave., Solana Beach, CA 92075 (US).

(74) Agents: **ULLSPERGER, Chris** et al.; Fenwick & West LLP, Silicon Valley Center, 801 California Street, Mountain View, CA 94041 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(88) Date of publication of the international search report:
30 December 2009

WO 2009/055512 A3

(54) Title: METHODS AND COMPOSTIONS FOR OBTAINING HIGH-RESOLUTION CRYSTALS OF MEMBRANE PROTEINS

(57) Abstract: The invention describes compositions and method useful for the crystallization of membrane proteins.

PATENT APPLICATION

METHODS AND COMPOSITIONS FOR OBTAINING HIGH-RESOLUTION CRYSTALS OF MEMBRANE PROTEINS

Inventors: Raymond C. Stevens
Citizen of the United States
7096 Caminito Valverde
La Jolla, California 92037

Michael A. Hanson
Citizen of the United States
1218 Avenida Amistad
San Marcos, California 92069

Vadim Cherezov
Citizen of Russia
12604 Torrey Bluff Dr., Apt. 399
San Diego, California 92130

Peter Kuhn
Citizen of Germany
708 Barbara Ave
Solana Beach, California 92075

TSRI 1326.1

METHODS AND COMPOSITIONS FOR OBTAINING HIGH-RESOLUTION CRYSTALS OF MEMBRANE PROTEINS

RELATED APPLICATIONS

[0001] This invention claims the benefit of U.S. provisional applications 60/999,951, filed Oct. 22, 2007; US provisional application 61/000,325, filed Oct. 24, 2007; U.S. provisional appliction 61/060,107, filed June 9, 2008; and U.S. provisional application 61/194,961, filed Oct. 1, 2008, each of which is incorporated herein by reference, in its entirety, for all purposes.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with the support of the United States government under the NIH Roadmap Initiative grant P50 GM073197, Protein Structure Initiative P50 GM62411 (to R.C.S.), NIH Roadmap Initiative grant R21 GM075811 and NINDS Grant NS028471, NIH grant F32 GM082028; and NIH grant R01 GM056169. The GM/CA-CAT beamline (23-ID) is supported by the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). The Government has certain rights in the invention.

[0003] Coordinates and structure factors have been deposited in the Protein Data Bank with identification code 2RH1.

BACKGROUND OF THE INVENTION

[0004] Naturally occurring polypeptides or proteins often fold into complex, three-dimensional shapes that determine both chemical and physiological functionality. Thus a thorough understanding of proteins necessarily involves a detailed representation of their spatial topography. The field of protein crystallography has flourished over the last 20 years resulting in a rapid increase in the knowledge bas of protein structure enabling great strides in other disciplines including biochemistry, pharmaceutical development and cell biology. However, the structural biology field has largely been restrained to working with protein that is naturally soluble in aqueous media, or made soluble by incorporation into surfactant micelles. The present invention provides methods and compositions that allow for the study of membrane-embedded proteins (i.e., integral membrane proteins) in a more natural membrane bilayer environment. The present invention enable a more detailed analysis of important classes of membrane –embedded polypeptides that play key roles in a variety of cellular processes including energy and signal transduction.

SUMMARY OF THE INVENTION

[0005] Other objects, features and advantages of the methods and compositions described herein will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples, while indicating specific embodiments, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

[0006] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference for all purposes and to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

[0007] In one aspect the invention provides compositions, e.g., useful for the crystallization of membrane protein. In some embodiments, the composition is suitable for lipidic cubic phase crystallization. In some embodiments the compositions contain 10 – 60 %v/v of a polyethylene glycol, 0.01- 0.5 M of a salt, 1 -20 %v/v of an organic compound, and 1 – 50 %w/w of a lipid additive in a host lipid. In some embodiments of the compositions, the protein to be crystallized is present at a concentration of 1 to 100 mg/mL. In some embodiments, the protein to be crystallized is present at a concentration of 50 mg/mL.

[0008] In some embodiments of the compositions of the inventions, the polyethylene glycol is PEG or modified PEG at a molecular size of 10 -8000. In some embodiments, the PEG or modified PEG has an average molecular weight of 400 – 8000. In some embodiments, the PEG or modified PEG has an average molecular weight of 400 -2000. In some embodiments, the PEG or modified PEG has an average molecular weight of 400 -1000. In some embodiments, the PEG or modified PEG has an average molecular weight of 400. Examples of modified PEG include but are not limited to PEG laurate, PEG dilaurate, PEG oleate, PEG dioleate, PEG stearate, PEG distearate, PEG glyceryl trioleate, PEG glyceryl laurate, PEG glyceryl stearate, PEG glyceryl oleate, PEG palm kernel oil, PEG hydrogenated castor oil, PEG castor oil, PEGcorn oil, PEG caprate/caprylate glycerides, PEG caprate/caprylate glycerides, PEG cholesterol, PEG phyto sterol, PEG soya sterol, PEG trioleate, PEG sorbitan oleate, PEG sorbitan laurate, PEG succinate, PEG nonyl phenol series, PEG octyl phenol series, Methyl-PEG, PEG-Maleimide, PEG4-NHS Ester and methoxypoly(ethylene glycol) (mPEG).

[0009] In some embodiments of the compositions, the salt is a sulfate salt. In some embodiments, the sulfate salt is sodium sulfate. In some embodiments, the salt is present at a concentration of 0.1- 0.5 M. In some embodiments, the salt is present at a concentration of 0.1- 0.2 M. In some embodiments, the compositions

of the invention contain a buffer. In some embodiments, the buffer is present at a concentration of 0.01 – 0.5 M. In some embodiments, the buffer is present at a concentration of 0.1 – 0.2 M. In some embodiments, the buffer is present at a concentration of 0.1 M. In some embodiments, the buffer is Bis-tris propane. In some embodiments, the buffer has a pH 6.5 -7.0. In some embodiments of the compositions, an organic compound is present at a concentration of 1 – 10 %v/v. In some embodiments, the organic compound is present at a concentration of 5-7 %v/v. In some embodiments, the organic compound is 1,4-butanediol.

[0010] In some embodiments of the compositions, the lipid additive is present at a concentration of 1 – 20 %w/w in a host lipid. In some embodiments, the lipid additive is present at a concentration of 8-10 %w/w in a host lipid. Examples of lipid additives include but are not limited to cholesterol, DOPE, DOPE-Me, DOPC, and Asolectin. In some embodiments, the lipid additive is cholesterol. Examples of host lipids include, but are not limited to monopalmitolein, monovaccenin and monoolein. In some embodiments, the host lipid is monoolein.

[0011] In some embodiments of this aspect, the composition of the invention comprises 30-35 %v/v PEG400, 0.1-0.2 M Na sulfate, 0.1 M Bis-tris propane pH 6.5 -7.0, 5-7 %v/v 1,4-butanediol using 8-10 %w/w cholesterol in monoolein as the host lipid.

[0012] In another aspect the invention includes compositions suitable for lipidic cubic phase crystallization. In some embodiments, the compositions of the inventions contain a lipid additive. In some embodiments, the lipid additive is present at a concentration of 1 – 50 %w/w in a host lipid. In some embodiments, the lipid additive is present at a concentration of 1 – 20 %w/w in a host lipid. In some embodiments, the lipid additive is present at a concentration of 8-10 %w/w in a host lipid. Examples of lipid additives include, but are not limited to, cholesterol, DOPE, DOPE-Me, DOPC, and Asolectin. In some embodiments, the lipid additive is cholesterol. Examples of host lipids include, but are not limited to monopalmitolein, monovaccenin and monoolein. In some embodiments the host lipid is monoolein.

[0013] In another aspect the invention includes methods for crystallization of membrane proteins. In some embodiments, the method for crystallization of membrane proteins comprises adding a lipid additive to a lipidic cubic phase. Examples of lipid additives include, but are not limited to, cholesterol, DOPE, DOPE-Me, DOPC, and Asolectin. In some embodiments, the lipid additive is cholesterol. In some embodiments, the lipid additive is present at a concentration of 1 – 50 %w/w in a host lipid. In some embodiments, the lipid additive is present at a concentration of 1 – 20 %w/w in a host lipid. In some embodiments, the lipid additive is present at a

concentration of 8-10 %w/w in a host lipid. Examples of host lipids include, but are not limited to monopalmitolein, monovaccenin and monoolein. In some embodiments the host lipid is monoolein.

[0014] In another aspect, the invention provides for methods of crystallization of a protein. In some embodiments of this aspect, the method comprises, providing said protein in a lipidic cubic phase composition, filling a plate comprising a material that does not interfere with imaging (such as, e.g., a transparent glass or plastic) with said composition, placing said plate containing said composition under conditions suitable for crystallization of said protein and detecting the presence of a crystal of said protein in said plate. In some embodiments, the method further comprises covering said plate with a second plate comprising a material that does not interfere with imaging (such as, e.g., a transparent glass or plastic).

[0015] In some embodiments of the methods, the protein is a non-colored protein. In some embodiments, the protein is a G protein-coupled receptor (GPCR). In some embodiments, the protein comprises a β_2AR, a CXCR4, or a human adenosine A_{2A} receptor. In some embodiments the protein comprises a stabilizing point mutation or a T4 lysozyme fusion or both.

[0016] In some embodiments of the methods, the lipidic cubic phase composition comprises a lipid additive. Examples of lipid additives include, but are not limited to, cholesterol, DOPE, DOPE-Me, DOPC, and Asolectin. In some embodiments, the lipid additive is cholesterol. In some embodiments, the lipid additive is present at a concentration of 1 – 50 %w/w in a host lipid. In some embodiments, the lipid additive is present at a concentration of 1 – 20 %w/w in a host lipid. In some embodiments, the lipid additive is present at a concentration of 8-10 %w/w in a host lipid. Examples of host lipids include, but are not limited to monopalmitolein, monovaccenin and monoolein. In some embodiments the host lipid is monoolein.

[0017] In some embodiments of the methods, the first plate and second plate are made of glass. In some embodiments, the plate allows for the control of crystallization conditions, such as the humidity and temperature of said lipidic cubic phase composition.

[0018] In some embodiments of the methods, the crystals are harvested directly from the plate. In some embodiments of the methods, the crystals are harvested between the cubic and the sponge phase of the lipidic cubic phase composition. In some embodiments of the methods, the crystals are harvested directly from said lipidic cubic phase composition and placing said crystals in liquid nitrogen.

[0019] In another aspect the invention provides methods for screening a crystal present in a liquid cubic phase composition. In some embodiments, the method comprises exposing the composition to a first beam and

determining a change of the first beam, exposing the composition to a second beam and determining a change of the second beam, and identifying an area where the crystal is present in said composition. Examples of changes in the beams, include but are not limited to, change in direction and/or intensity of the beams. In some embodiments the crystals are non-colored.

[0020] In some embodiments of the methods, the first beam and second beam are attenuated. In some embodiments the beams are attenuated 10 times. In some embodiments, the first beam is a slitted 100 x 25 μm beam. In some embodiments, the methods comprise exposing said lipidic cubic phase composition to a third beam. In some embodiments, the methods comprise exposing said lipidic cubic phase composition to up to ten extra beams. In some embodiments, the exposure of composition to the beams is 2 seconds. In some embodiments, the beams are beams of visible light.

[0021] In another aspect the invention includes a crystal of a membrane protein. In some embodiments, the invention includes a crystal of a G protein-coupled receptor (i.e., a "GPCR") non-covalently bound to a ligand. In some embodiments, the extracellular domain of said crystalline GPCR is resolvable by X-Ray crystallography. In some embodiments, the ligand is a diffusible ligand.

[0022] In some embodiments of this aspect, the volume of the crystal is greater than 15 x 5 x 1 μm. In some embodiments, the volume of said crystal is greater than 30 x 5 x 5 μm. In some embodiments of this aspect, the volume of crystal is greater than 40 x 20 x 5 μm. In some embodiments, volume of the crystal is estimated assuming that each of the stated dimensions are orthogonal so that the volume estimate is the product of the dimensions. In some embodiments, the crystal is suitable for X-ray crystallography. In some embodiments, X-ray crystallographic analysis can be carried out to determine the structure of a protein comprising said crystal.

[0023] In some embodiments of this aspect, the crystal is crystallized using liquid cubic phase crystallization. In some embodiments, the crystal is obtainable by harvesting the crystal from a glass sandwich plate. In some embodiments, the crystal diffract to a resolution of 1.0 to 10.0 Å. In some embodiments, the crystal diffract to a resolution of 2.0 to 5.0 Å. In some embodiments, the crystal diffract to a resolution of 2.2 Å. In some embodiments, the structure of said crystal is solved and refined at resolution of less than about 3.2, 2.8, 2.6 or 2.4 Å. In some embodiments, the structure of said crystal is solved and refined at resolution of less than about 2.8, 2.6 or 2.4 Å. In some embodiments, the structure of said crystal is solved and refined at resolution of less than about 2.4 Å.

[0024] In some embodiments, the G protein-coupled receptor is a β_2AR protein, a CXCR4 protein, or a human adenosine A_{2A} receptor protein.

[0025] In another aspect the invention provides for a crystal of β_2AR. In some embodiments, the structure of an extracellular domain of said crystal is capable of being resolved by X-ray crystallography. In some embodiments, the crystal comprises 442 amino acids, a palmitic acid covalently bound to Cys341 and an acetamide molecule bound to $Cys265^{6.27}$, a diffusible ligand, up to 10 molecules a lipid additive, up to five salt ions and up to 10 butanediol molecules. In some embodiments, the lipid additive is cholesterol. In some embodiments, the crystal comprises three cholesterol molecules. In some embodiments, the salt ion is a sulfate ion. In some embodiments, the crystal comprises two sulfate ions. In some embodiments, the diffusible ligand is carazol. In some embodiments the crystal comprises two butanediol molecules. In some embodiments of this aspect, the volume of the crystal is greater than 15 x 5 x 1 μm. In some embodiments, the volume of said crystal is greater than 30 x 5 x 5 μm. In some embodiments of this aspect, the volume of the crystal is greater than 40 x 20 x 5 μm. In some embodiments the volume of the crystal is estimated by assuming that each dimension is orthogonal to the other dimensions so that the volume is the product of the three lengths. In some embodiments, the crystal is suitable for X-ray crystallography. In some embodiments, the structure of a β_2AR protein can be determined from said crystal using X-ray crystallographic analysis. In some embodiments, the crystal is crystallized using liquid cubic phase crystallization. In some embodiments, the crystal is obtainable by harvesting the crystal from a glass sandwich plate. In some embodiments of this aspect, the crystal is crystallized using liquid cubic phase crystallization. In some embodiments, the crystal is obtainable by harvesting the crystal from a glass sandwich plate. In some embodiments, the crystal diffract to a resolution of 1.0 to 10.0 Å. In some embodiments, the crystals diffract to a resolution of 2.0 to 5.0 Å. In some embodiments, the crystal diffract to a resolution of 2.2 Å. In some embodiments, the structure of said crystal is solved and refined at resolution of less than about 3.2, 2.8, 2.6 or 2.4 Å. In some embodiments, the structure of said crystal is solved and refined at resolution of less than about 2.8, 2.6 or 2.4 Å. In some embodiments, the structure of said crystal is solved and refined at resolution of less than about 2.4 Å.

[0026] In another embodiment, the invention provides a composition for lipidic cubic phase crystallization of a membrane protein, comprising a polyethylene glycol or modified polyethylene glycol; 0.01- 1M of a salt; a host lipid; a lipid additive, wherein said lipid additive is present at 10 – 60 %v/v ratio relative to the host lipid; a buffer; and 1 to 100 mg/ml of a membrane protein. In a related embodiment, the polyethylene glycol is PEG

or modified PEG, wherein said PEG or modified PEG has an average molecular weight of 200 -20,000, 400 – 8000, or 400 -2000. In yet another related embodiment, the PEG or modified PEG in the composition has an average molecular weight of 400. In another related embodiment, the salt is selected from the group consisting of a sulfate salt, a citrate salt. a malonate salt. a tartrate salt, an acetate salt, and a formate salt. In certain embodiments of the composition, the salt is present at a concentration of 0.1 - 0.2 M. In another related embodiment, the buffer is present at a concentration of 0.05 - 0.5 M in the composition. In certain embodiments, the buffer is Bis-tris propane.or sodium citrate. In other related embodiments of the composition, the buffer has a pH between 4.5 - 8.0.

[0027] In still other related embodiments of the composition for lipidic cubic phase crystallization of a membrane protein, the composition further comprising an alcohol present at a concentration of 1 - 10 %v/v or 5-7 %v/v. In certain embodiments, the alcohol is a diol or triol. In other embodiments, the alcohol is 1,4-butanediol or 2,6-hexanediol.

[0028] In still other related embodiments of the composition for lipidic cubic phase crystallization of a membrane protein, the lipid additive is present at a concentration of 1 – 20 %w/w in a host lipid or 8-10 %w/w in a host lipid. In yet another related embodiment of the composition, the lipid additive is selected from the group consisting of 2-monoolein, phosphotidylcholine, cardiolipin, lyso-PC, a polyethylene glyocol-lipid, dioleoylphosphatidylethanolamine ("DOPE"), DOPE-Me, dioleoyl phosphatidylcholine ("DOPC"), Asolectin, and a sterol. In still other embodiments, lipid composition is a sterol. In related embodiments, the lipid additive is cholesterol.

[0029] In still another related embodiment of the composition for lipidic cubic phase crystallization of a membrane protein, the host lipid is selected from the group consisting of monopalmitolein, monovaccenin and monoolein. In a related embodiment, the host lipid is monoolein. In still another related embodiment, the membrane protein to be crystallized in said composition is present at a concentration of 1 to 100 mg/mL. In yet another embodiment, the memrane protein to be crystallized in said composition is present at a concentration of 40-60 mg/mL.

[0030] In still another related embodiment of the composition for lipidic cubic phase crystallization of a membrane protein membrane protein is a G-protein coupled receptor, such as a β_2AR protein, a human adenosine A2A receptor protein, a CXCR4-T4L protein, or a β_2AR-T4L protein. In related embodiments, the G-protein coupled receptor is a protein comprising or consisting of a $\beta_2AR^{(E122W)}$, a $\beta_2AR^{(E122W)}$-T4L, a human

adenosine A2A receptor-T4L, a CXCR4-T4L or β_2AR-T4L. In still another related embodiment of the composition for lipidic cubic phase crystallization of a membrane protein, the composition comprises a ligand selected from the group consisting of carazolol, timolol, alprenolol, and clenbutorol.

[0031] In another embodiment, the invention provides a method of generating crystals of a membrane protein comprising: mixing a lipid additive with a host lipid to form a lipid mixture, wherein said lipid additive is selected from the group consisting of a sterol, DOPE, DOPE-Me, DOPC, and Asolectin, and wherein said lipid additive is 5 to 20 %w/w in said host lipid; and combining said lipid mixture with a membrane protein solution under conditions suitable for the formation of a lipidic cubic phase composition. In a related embodiment of the method, said protein is a non-colored protein. In a related embodiment, the amount of said lipid additive is 8 to 10 % w/w in said lipid. In another related embodiment, the invention further comprises: filling a plate with said lipidic cubic phase composition, wherein said plate is compatible with imaging; placing said plate containing said lipic cubic phase composition under conditions suitable for crystallization of said protein; and detecting the presence of a crystal of said protein in said plate. In another embodiment, the method further comprises covering said plate with a second plate.

[0032] In a related embodiment of the method of generating crystals of a membrane protein, the protein is a GPCR. In yet another related embodiments, the protein comprises a β_2AR. In yet another related embodiment, the . β_2AR protein is selected from the group consisting of $\beta_2AR^{(E122W)}$, $\beta_2AR^{(E122W)}$-T4L, and β_2AR –T4L. In yet another related embodiment, the GPCR is a human adenosine A2A receptor or a CXCR4 receptor where the proteins may comprise, in still other related embodiments, a T4 lysozome..

[0033] In yet another related embodiment of method of generating crystals of a membrane protein, the lipid additive is present at a concentration of 1 – 20 %w/w or 8-10 %w/w in a host lipid. In yet another related embodiment of the method, the second plate comprises a glass. In yet another related embodiment, the method further comprises harvesting crystals directly from said plate. Another related embodiment of the method comprises harvesting crystals from between the cubic and the sponge phase of said lipidic cubic phase composition. In anothre related embodiment, the method comprises harvesting crystals directly from said lipidic cubic phase composition and placing said crystals in liquid nitrogen. In yet another related embodiment, the method comprising a step of soaking into said crystal a diffusable ligand or candidate ligand.

[0034] The invention also provides a method of screening a crystal of a GPCR present in a liquid cubic phase composition comprising: preparing a liquid cubic phase composition comprising a GPCR protein, a host

lipid, and a lipid additive; exposing said composition to a first X-ray beam and determining a change in direction or intensity of said first X-ray beam; exposing said composition to a second beam and determining a change in direction or intensity of said second X-ray beam; identifying an area where said GPCR crystal is present in said composition; and exposing said identified area to at least a third X-ray beam. In a related embodiment, the first beam is a slitted 100 x 25 μm beam. In another related embodiment, the crystal is colorless. In yet another embodiment, the GPCR crystal is $\beta_2AR^{(E122W)}$-T4L, $\beta_2AR^{(E122W)}$, β_2AR, or β_2AR–T4L protein. In related embodiments, the crystal is human adenosine A2A receptor or a CXCR4 receptor where the proteins may comprise, in still other related embodiments, a T4 lysozome.

[0035] In another embodiment, the invention provides a crystal of a human β_2AR protein wherein the extracellular loop ECL2 of said β_2AR is sufficiently ordered to produce interpretable electron density in a crystallographically-derived electron density map. In yet another related embodiment, each β_2AR molecule in said crystal comprises three non-covalently bound cholesterol molecules and at least one salt ion. In yet another related embodiment, the at least one salt ion is a sulfate ion. In yet anothre related embodiment, each β_2AR molecule in said crystal said crystal comprises two sulfate ions. In yet anothre related embodiment, each β_2AR molecule in said crystal said crystal further comprises carazol. In yet another related embodiment of the crystal, each β_2AR molecule in said crystal comprises between 1 and 10 butanediol molecules. In yet another related embodiment of the crystal, the volume of said crystal exceeds 15 x 5 x 1 μm, 30 x 5 x 5 μm or 40 x 20 x 5 μm. In yet another related embodiment, the specific surface area of the crystal is 0.0001 - 5 m^2/g. In yet another related embodiment, the crystal is crystallized using liquid cubic phase crystallization. In yet another related embodiment, the crystal is obtainable by harvesting the crystal from a glass sandwich plate. In yet another related embodiment, the crystal diffracts to a resolution of 2.0 to 10.0 Å, 2.0 to 5.0 Å, or 2.2 to 2.4 Å. In yet another related embodiment, the structure of said crystal is solved and refined at a resolution higher than about 3.2, higher than about 2.8, or higher than about 2.4 Å.

[0036] In yet another embodiment, the invention provides a crystalline form of a human β_2AR protein having an atomic arrangement of coordinates comprising the β_2AR coordinates set forth in Appendix I.

[0037] In another embodiment, the invention provides a crystalline form of a human β_2AR protein, where said form has unit cell dimensions of a=106.3 Angstroms, b=169.2 Angstroms, and c= 40.2 Angstroms. In arelated embodiment, said β_2AR protein is β_2AR-T4L. In another related embodiment, the β_2AR-T4L crystal further comprises a carazolol ligand.

[0038] In another embodiment, the invention provides a crystalline form of a human β_2AR protein, wherein said space group of said crystalline form is C2. In a related embodiment, said β_2AR protein is β_2AR-T4L. In another related embodiment, the β_2AR-T4L crystal further comprises a carazolol ligand.

[0039] In another embodiment, the invention provides a crystalline form of a human β_2AR protein, wherein said crystalline form diffracts X-rays to a resolution of 2.4 Angstroms. In a related embodiment, said β_2AR protein comprises a point mutation that stabilizes said β_2AR protein. In another related embodiment, said β_2AR protein is β_2AR-T4L. In a related embodiment, the β_2AR-T4L crystal further comprises a carazolol ligand.

[0040] In another embodiment, the invention provides a crystalline form of a human β_2AR protein wherein each β_2AR molecule in said crystal comprises 442 amino acids, a palmitic acid covalently bound to Cys341, an acetamide molecule bound to $Cys265^{6.27}$, a diffusible ligand, one to ten molecules of a lipid additive, one to five salt ions and one to ten butanediol molecules.

[0041] In another embodiment, the invention provides a method of identifying a compound that binds to a ligand binding site of a human β_2AR protein, comprising comparing a set of three-dimensional structures representing a set of candidate compounds with a three-dimensional molecular model of said ligand binding site, comprising: receiving a three-dimensional model of a ligand binding site on said human β_2AR protein, wherein said three-dimensional model of said ligand binding site comprises atomic co-ordinates for a plurality of ligand-binding residues, wherein said atomic co-ordinates are taken from **Appendix I**; determining, for each of the set of compound three-dimensional models, a plurality of distance values indicating distances between said atomic co-ordinates of said candidate compound of the set of candidate compounds and said atomic coordinates of said ligand-binding residues comprising said ligand binding site; determining, for each of the set of compound three-dimensional models, a binding strength value based on the plurality of distance values determined for the compound three-dimensional model, wherein the binding strength value indicates the stability of a complex formed by said human β_2AR protein and a compound represented by the compound three-dimensional model; and storing a set of results indicating whether each candidate compound binds to the three-dimensional model based on the binding strength values. In a related embodiment, said ligand-binding residues comprise a plurality of residues selected from the group consisting of Y199, A200, S204, T118, V117, W286, Y316, F290, F289, N293, W109, F193,and Y308. In another related embodiment, said ligand-binding residues comprise a plurality of residues selected from the group consisitng of W109, V117, T118, F193, Y199, A200, W286, F289, F290,

Y316. In another related embodiment of the method of identifying a compound that binds to a ligand binding site of a human β_2AR protein, said binding strength value is based on one or more of a hydrogen bonding strength, a hydrophobic interaction strength, or a Coulombic interaction binding strength. In another related embodiment, one or more of said receiving, determining, or storing steps is carried out using a commercially-available software program. In yet another related embodiment, the commercially-available software program is selected from the group consisting of DOCK, QUANTA, Sybyl, CHARMM, AMBER, GRID, MCSS, AUTODOCK, CERIUS II, Flexx, CAVEAT, MACCS-3D, HOOK, LUDI, LEGEND, LeapFrog, Gaussian 92, QUANTA/CHARMM, Insight II/Discover, and ICM. In yet another related embodiment, the method further comprising the step of contacting a human β_2AR protein with a molecule comprising an identified candidate compound. In yet another related embodiment, the molecule further comprises a moiety capable of competitively displacing a ligand from said human β_2AR protein, wherein said ligand binds to said ligand binding site of said human β_2AR protein. In yet another related embodiment, the method further comprising characterizing a binding interaction between said human β_2AR protein and said molecule comprising said identified candidate compound, and storing a result of said characterizing. In yet another related embodiment, said characterization comprises determining an activation of a function of said human β_2AR protein, an inhibition of a function of said human β_2AR protein, an increase in expression of said human β_2AR protein, a decrease in expression of said human β_2AR protein, a displacement of a ligand bound to said ligand binding site, or a stability measure for said human β_2AR protein.

[0042] In another embodiment, the invention provides a method for selecting a library of potential modulators of β_2AR to be screened, comprising calculating a structure of a first potential modulator using at least a portion of the structure co-ordinates of Appendix I, correlating said structure of said first potential modulator with a library of modulators identified as comprising said structure said first potential modulator, and storing or transmitting information about the identified library.

[0043] In yet another embodiment, the invention provides a method of solving the structure of a crystalline form of a protein, comprising: using at least a portion of the structure co-ordinates of Appendix I to solve the structure of the crystalline form of a test protein, wherein said test protein has significant amino acid sequence homology to any functional domain of β_2AR; and transmitting or storing data descriptive of the structure of said test protein.

[0044] In another embodiment, the invention provides a method of identifying from a set of candidate compound three-dimensional models a compound that binds to a ligand binding site of a GPCR or β_2AR protein comprising: receiving a three-dimensional model of a ligand binding site on said GPCR or β_2AR protein, wherein said three-dimensional model of said ligand binding site comprises atomic co-ordinates for a plurality of ligand binding residues; determining, for each candidate compound of the set of candidate compound three-dimensional models, a plurality of distance and angle values indicating distances and angles between atomic co-ordinates of said candidate compound of the set of candidate compound three-dimensional models and said ligand binding site comprising atomic coordinates of said ligand-binding residues; determining, for each of the set of candidate compound three-dimensional models, a binding strength value based on the plurality of distance and angle values determined for the candidate compound three-dimensional model, wherein the binding strength value indicates the stability of a complex formed by said human GPCR or β_2AR protein and a compound represented by the compound three-dimensional model; storing a set of results indicating whether each candidate compound binds to the three-dimensional model based on the binding strength values; searching a database of small organic molecules for compounds exhibiting shape, chemistry, or electrostatic similarity with the candidate compounds indicated to bind to the three-dimensional model based on the binding strength values; and identifying the set of small organic molecules exhibiting shape, chemistry, or electrostatic similarity with the candidate compounds indicated to bind to the three-dimensional model based on binding strength values as likely to also bind to the GPCR or β_2AR.the database of small organic molecules is the available chemicals database. In a related embodiment, the shape, chemistry or electrostatic similarity is determined using a program selected from the group consisting of BROOD (openeye), EON (openeye), ROCS (openeye), ISIS Base, and SciFinder.

[0045] In another embodiment, the invention provides a method of identifying a ligand that binds to a membrane protein comprising: preparing a lipid meso phase, wherein said lipid meso phase composition comprises (1) a host lipid; (2) said membrane protein; (3) a lipid additive selected from the group consisting of consisting of a sterol, cholesterol, DOPE, DOPE-Me, DOPC, and Asolectin, wherein said lipid additive is 1 to 50% w/w in a lipid host; subjecting said lipid meso phase to humidity and temperature conditions to grow crystals of said membrane protein; contacting said membrane protein crystals with diffusible ligands or a mixture of diffusible ligands; determining the three-dimensional structure of said diffusible ligand contacted membrane protein crystals by X-ray crystallography to obtain an electron density map; and identifying bound

mixture of diffusible ligands; determining the three-dimensional structure of said diffusible ligand contacted membrane protein crystals by X-ray crystallography to obtain an electron density map; and identifying bound ligands by inspection of the electron density map. In a related embodiment, the ligands are substantially insoluble in water.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046] A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0047] **Figure 1.** Crystals of β_2AR-T4L obtained in 30-35 %v/v PEG400, 0.1-0.2 M Na sulfate, 0.1 M Bis-tris propane pH 6.5 -7.0, 5-7 %v/v 1,4-butanediol using 8-10 %w/w cholesterol in monoolein as the host lipid. **Figure 1a** shows β_2AR-T4L crystals in the crystallization mixture drop (upper left) and in the loop. **Figure 1b** shows crystals in the "sponge" phase.

[0048] **Figure 2.** Before (top) and after (bottom) images of a lipidic cubic phase crystal harvested directly from a well in a previously sealed glass sandwich plate, according to the method described herein (see, e.g., Example 1).

[0049] **Figure 3.** Diffraction pattern (2.8 Å resolution) from β_2AR-T4L crystals grown in lipid cubic phase. The crystal size was approximately 25 x 5 x 5 Å; space group C2 (a=106.8 Å, b=169.5 Å, c=40.5 Å; β= 105.3°, $\alpha=\gamma=90°$). Beam diameter was 10 μm, exposure 10 s, oscillation: 1°.

[0050] **Figure 4.** Gallery of crystals of various GPCRs obtained using LCP/cholesterol mixtures and in combination with a variety of ligands. The top panel corresponds to non-optimized initial hits, whereas the bottom panel shows diffraction quality for optimized crystals. From left to right, β_2AR-T4L (bound to carazolol), diffracted to 2.4 Å resolution; β_2AR(E122W)T4L (bound to carazolol), diffracted to 3.5 Å resolution; β_2AR(E122W)T4L (bound to alprenolol), diffracted to 3.5 Å resolution; β_2AR(E122W)T4L (bound to timolol), diffracted to 2.8 Å resolution; β_2AR(E122W) (bound to carazolol); β_2AR(E122W)T4L (bound to clenbuterol), diffracted to 6 Å, anisotropic; human A_{2A} adenosine receptor-T4L (bound to ZM241385), diffracted to 2.6 Å resolution.

[0051] **Figure 5**. **A.** β_2AR-T4L Crystals grown from bicelle conditions. **B.** Diffraction image from bicelle grown microcrystals of β_2AR-T4L recorded using 10 μm minibeam on 23ID-B beamline at APS. Black circle is drawn at resolution 3.5 Å.

[0052] **Figure 6**. **A.** Microcrystals of β_2AR-T4L grown in lipidic mesophase. **B.** Diffraction image from lipidic cubic phase grown microcrystals of β_2AR-T4L recorded using a 10 μm minibeam on 23ID-B beamline at APS. The white circle is drawn at resolution 2.2 Å.

[0053] **Figure 7**. Detailed representation of the carazolol binding site in β_2AR-T4L. **Figure 7A, B, and C** are representations of the electron density of the ligand binding site at three different orientations. Residues are labeled by their Ballesteros-Weinstein numbers as superscripts. Electon density is contoured at 1σ from a 2 F_o-F_c difference map. Both **B** and **C** are generated by rotating the field of view 90° about the y-axis clockwise and counterclockwise respectively.

[0054] **Figure 8**. Electron density of: **A.** Cholesterol molecules shown with a F_o-F_c electron density contoured at 2σ omitting the lipid from phase calculation, palmitic acid is also shown. **B.** Helix-kinked region with $2F_o$-F_c electron density contoured at 1.5σ.

[0055] **Figure 9.** Overall fold of the β_2AR-T4L fusion with its predicted orientation in the plasma membrane and key intramolecular interactions. **A.** Stereoview of the overall fold of β_2AR-T4L. The receptor and T4L are colored gray and green, respectively. Carazolol is colored blue and the lipid molecules bound to the receptor are colored yellow. **B.** The receptor is aligned to a rhodopsin model that was positioned in a lipid membrane (boundaries indicated by horizontal black lines) as found in the orientations of proteins in membranes (OPM) database (M. A. Lomize et al., *Bioinformatics* **22**, 623 (2006)). T4L is fused internally into the third intracellular loop of β_2AR and maintains minimal intramolecular packing interactions by tilting away from the receptor. **C.** Specific intramolecular interactions between β_2AR and T4L are represented.

[0056] **Figure 10.** Crystal packing interactions in the lipidic mesophase crystallized β_2AR-T4L. **A.** There are four main contact areas, two of which are mediated by T4L in the plane of the membrane with itself through a two-fold symmetry axis and translation. The third interaction is normal to the membrane plane between T4L and lumen exposed loops of β_2AR. The fourth interaction is generated by the two-fold symmetry axis, packing one receptor to receptor in the plane of the membrane. **B.** The receptor crystal packing interface is composed mainly of lipids with two cholesterol molecules and two palmitic acid molecules forming the majority of the interactions. A network of ionic charge interactions exists on the cytoplasmic end of the

interface forming the only inter-receptor protein contacts. **C.** Comparison between β_2AR-T4L and rhodopsin (PDB ID Code 2135) parallel receptor association interface. Helices I (blue) and VIII (magenta) are highlighted in both structures. Only one monomer is shown for each receptor representation along with helices I' and VIII' only from the opposing symmetry related molecule. The rhodopsin interface is twisted significantly relative to β_2AR-T4L resulting in a significant offset from the parallel orientation required for a physiological dimer interface. β_2AR-T4L associated monomers are in a highly parallel orientation.

[0057] **Figure 11.** Surface representation of β_2AR colored by calculated charge from red (-10 k_bT/e_c) to blue (+10 k_bT/e_c) using a dielectric constant of 70. **A.** Three main areas of interest are indicated. The binding site cleft is negatively charged as is a groove between helices III, IV and V. The third region is an overall positive charge in the region of the ionic lock and DRY motif on the cytoplasmic face. The overall result is a highly polarized molecule that may utilize its negative charge to facilitate binding of catecholamine ligands. The presence of a negative charge in the groove between helices III, IV and V is unexpected as it is in the middle of the lipid membrane. This charge may be partially derived from the presence of an unpaired glutamate at position $122^{3.41}$. The effective charge in this region is likely greater than shown here due to its location in the low dielectric environment of the lipid membrane. **B.** View rotated 90° from A. Showing both the negatively charged binding site cleft (top) and positively charged cytoplasmic face (bottom). Poisson-Boltzmann electrostatics were calculated using the program APBS (Baker et al., *Proc Natl Acad Sci U S A* , **98**, 10037 (2001)) as implemented in Pymol (The PyMOL Molecular Graphics System (2002) on World Wide Web http://www.pymol.org). Pymol was used exclusively in the preparation of all figures.

[0058] **Figure 12.** Comparison of the extracellular sides of β_2AR-T4L and rhodopsin. **A.** The N-terminus is missing from the experimental density in the β_2AR-T4L structure and is not shown. ECL2 is shown in green and contains a short α-helix and two disulfide bonds (yellow). The intraloop disulfide bond constrains the tip of ECL2 which interacts with ECL1. The second disulfide bond links ECL2 with helix III. There is one interaction between ECL2 and carazolol (blue) through $Phe193^{5.32}$. The entire loop is held out of the ligand binding site by a combination of the rigid helical segment and the two disulfide bonds. **B.** In contrast, ECL2 (green) in rhodopsin assumes a lower position in the structure that occludes direct access to the retinal-binding site and forms a small β-sheet in combination with the N-terminal region (magenta) directly above the bound retinal (pink).

[0059] **Figure 13.** Ligand binding characterization and comparison to rhodopsin. **A.** A view looking down on the plane of the membrane from the extracellular surface showing a detailed representation of the

carazolol binding site in β_2AR-T4L. Carazolol is shown as sticks with carbon atoms colored yellow. β_2AR-T4L residues contributing to carazolol binding are shown in green and labeled. Electron density is contoured at 5σ from an F_o-F_c omit map calculated without the contribution of carazolol. **B.** Binding orientation comparison between 11-*cis*-retinal in rhodopsin and carazolol in β_2AR-T4L. Van der Waals' surfaces for carazolol and retinal are represented as dots to accentuate the close packing interactions. Retinal in the all-*cis* conformation (pink), binds deep in the active site of rhodopsin as compared to carazolol (blue), packing its β-ionone ring between Tyr268$^{6.51}$ and Phe212$^{5.47}$ (cyan), blocking movement of Trp265$^{6.48}$ (magenta) into the space. The β-ionone ring of *trans*-retinal in activated rhodopsin would not block Trp265$^{6.48}$ from rotating into the space allowing a rotameric shift into its proposed active form. **C.** There are four residues involved in the toggle switch mechanism of β_2AR-T4L as shown. Phe290$^{6.52}$ (magenta) is sandwiched between Phe208$^{5.47}$ (tan) and Phe289$^{6.51}$ (tan) forming a ring-face aromatic interaction. Like rhodopsin, an activation step is thought to occur by a rotameric change of Trp286$^{6.48}$ (magenta) which would displace Phe290$^{6.52}$. Carazolol is shown to interact extensively with the sandwich motif as shown: however, few interactions are seen with Trp286$^{6.48}$. The 6.52 position in β_2AR-T4L is occupied by Phe290$^{6.52}$ as opposed to Ala269$^{6.52}$ in rhodopsin where the β-ionone ring replaces an aromatic protein side chain in forming the sandwich interactions. The aromatic character of the sandwich is otherwise maintained by Phe289$^{6.51}$ and Phe208$^{5.47}$ in β_2AR-T4L.

[0060] **Figure 14.** Comparison of β_2AR-T4L helical orientations with rhodopsin (PDB ID Code 1U19). **A.** β_2AR-T4L is rendered as a ribbon trace colored with a blue to red spectrum corresponding to observed distances between Cα positions in the two structures (RMSD 2.7 Å between all residues in the transmembrane region). Helix II shows very little movement, whereas the entire lengths of helices III, IV, V shift significantly. Helix VIII and loops were not included in the comparison and are colored in tan. **B.** Movements of helices I and V of rhodopsin (grey) are shown relative to β_2AR-T4L. **C.** Movements of helices III, IV and VI. **D.** Ligand binding site representation. Carazolol is shown with yellow carbons. Entire helices are assigned a single designation based on their divergence from the rhodopsin position in the area of the ligand binding site as shown. Helix I is highly divergent, Helices II and VI are similar to rhodopsin. Helices IV and VII are moderately constant. Helices III and V are moderately divergent.

[0061] **Figure 15.** Affinity curves for adrenergic ligands binding to β_2AR-T4L and wildtype β_2AR. Saturation curves for the antagonist [^{3}H]DHA is shown at left, next to competition binding curves for the natural ligand (-)-Epinephrine and the high-affinity synthetic agonist Formoterol. Binding experiments on membranes isolated from Sf9 insect cells expressing the receptors were performed as described above.

[0062] **Figure 16**. Comparison of the proteolytic stability between the wild-type β_2AR and β_2AR-T4L in a limited trypsin proteolysis assay. TPCK-trypsin was added to carazolol-bound, purified, dodecylmaltoside-solubilized receptor at a 1:1000 ratio (wt:wt), and samples were analyzed by SDS-PAGE. Intact β_2AR-T4L (56.7 kD) and FLAG-tagged wild-type β_2AR (47.4 kD) migrate similarly as ~55 kD bands. Markers are Biorad low-range SDS-PAGE protein standards.

[0063] **Figure 17**. Stability comparison of unliganded β_2AR365 and β_2AR-T4L. For dodecylmaltoside-solubilized receptor preparations, maintenance of the ability to specifically bind [^{3}H]DHA after incubation at 37 °C is taken as a measure of stability.

[0064] **Figure 18**. Superimposed Cα traces of the receptor component of β_2AR-T4L (in blue) and β_2AR365 (in yellow). Common modeled transmembrane helix regions 41-58, 67-87, 108-137, 147-164, 204-230, 267-291, 312-326, 332-339 were used in the superposition by the program Lsqkab (The CCP4 Suite, *Acta Crystallogr D Biol Crystallogr* **50**, 760 (1994)) (RMSD = 0.8 Å).

[0065] **Figure 19**. Carazolol dissociation from β_2AR365. Dodecylmaltoside-solubilized carazolol-bound receptor (at 50 μM) was dialyzed in a large volume of buffer containing 300 micromolar alprenonol as a competing ligand, and aliquots were removed from the dialysis cassette at different time points. Remaining bound carazolol was measured (in a relative sense) by collecting fluorescence emission with excitation at 330 nm and emission from 335 to 400 nm. For each carazolol fluorescence measurement, data was normalized for the protein concentration in the dialysis cassette (measured with the Bio-Rad Protein DC kit). The Y-axis represents carazolol fluorescence emission Intensity (in cps) at 341 nm. The exponential decay of carazolol concentration in the receptor dialysis cassette was fit using Graphpad Prism software, giving a half-life of 30.4 hrs.

[0066] **Figure 20**. Comparison of β_1 and β_2AR sequences. After aligning the β_1 and β_2AR sequences, positions that have different amino acids between the two receptors were mapped onto the high-resolution structure of β_2ART4L (shown as red sticks). The carazolol ligand is shown as green sticks (with nitrogens in blue and oxygens in red). Highlighted residues Ala85$^{2.56}$, Ala92$^{2.63}$ and Tyr308$^{7.35}$ are homologous to amino acids Leu110$^{2.56}$, Thr117$^{2.63}$ and Phe359$^{7.35}$ of the β_1AR, which were shown to be primarily responsible for its selectivity over β_2AR for the compound RO363 (Sugimoto et al., *J Pharmacol Exp Ther* 301, 51 (2002)). In the β_2AR-T4L structure, only Tyr308$^{7.35}$ faces the ligand, while Ala85$^{2.56}$ lies at the interface between helices II and III. Of all the divergent amino acids, only Tyr308$^{7.35}$ is found within 4 Å of any atom of carazolol.

[0067] **Figure 21.** Design and optimization of the β_2AR-T4L fusion protein (SEQ ID NO: 2). **A.** The sequence of the region of the β_2AR targeted for insertion of a crystallizable domain is shown, and the positions of the junctions between the receptor and T4L (in red) for various constructs are indicated. The sequences that were initially replaced or removed are faded. Red lines are shown after every tenth residue. Peptide 'LNKYADWT' disclosed as SEQ ID NO: 3. **B.** Immunofluorescence images of HEK293 cells expressing selected fusion constructs. Panels on the left shows M1 anti-FLAG signal corresponding to antibody bound to the N-terminus of the receptor. Panels on the right show the same signal merged with blue emission from DAPI (nuclear staining for all cells). Plasma membrane staining is observed in the positive control, D3 and D1, while C3 and D5 are retained in the endoplasmic reticulum.

[0068] **Figure 22.** Functional characterization of β_2AR-T4L. **A.** Affinity competition curves for adrenergic ligands binding to β_2AR-T4L and wild-type β_2AR. Binding experiments on membranes isolated from Sf9 insect cells expressing the receptors were performed as described in the methods section of Example 4. **B.** β_2AR-T4L is still able to undergo ligand-induced conformational changes. Bimane fluorescence spectra (excitation at 350 nm) of detergent-solubilized β_2AR-T4L and wild-type β_2AR truncated at 365, labeled under conditions that selectively modify Cys265$^{6.27}$ (see methods section of Example 4), were measured after incubating unliganded receptor with compounds for 15 min at room temperature. The cartoon illustrates that the observed changes in fluorescence can be interpreted as a movement of the bimane probe from a more buried, hydrophobic environment to a more polar, solvent-exposed position.

[0069] **Figure 23**. **A.** Side-by-side comparison of the crystal structures of the β_2AR-T4L fusion protein and the complex between β_2AR365 and a Fab fragment. The receptor component of the fusion protein is shown as a blue cartoon (with modeled carazolol as red spheres), while the receptor bound to Fab5 is in yellow. **B.** Differences in the environment surrounding Phe264$^{6.26}$ (shown as spheres) for the two proteins. **C.** The analogous interactions to the "ionic lock" between the E(D)RY motif and Glu247$^{6.30}$ seen in rhodopsin (right panel, darkened) are broken in both structures of the β_2AR (left panel). Pymol (W. L. DeLano, *The PyMOL Molecular Graphics System (2002) on the World Wide Web http://www.pymol.org*) was used for the preparation of all figures.

[0070] **Figure 24.** Schematic representation of the interactions between β_2AR-T4L and carazolol at the ligand binding pocket. Residues shown have at least one atom within 4 Å of the ligand in the 2.4 Å resolution crystal structure.

[0071] **Figure 25.** The ligand binding pocket of β_2AR-T4L with carazolol bound. **A.** Residues within 4 Å of the ligand are shown as sticks, with the exception of A200, N293, F289, and Y308. Residues that form polar contacts with the ligand (distance cutoff 3.5 Å) are in green, other residues are gray (in all panels, oxygens are colored red and nitrogens are blue). **B.** Same as panel A, except that the ligand is oriented with its amine facing out of the page. W109 is not shown. **C.** Packing interactions between carazolol and all residues within 5 Å of the ligand. View is from the extracellular side of the membrane. Carazolol is shown as yellow spheres, receptor residues are shown as sticks within van der Waals dot surfaces. **D.** Model of (-)-isoproterenol (magenta sticks) in the ligand binding pocket observed in the crystal structure. A model of the agonist with optimal bond lengths and angles was obtained from the PRODRG server (Schuettelkopf, et al., *Acta Crystallogr D Biol Crystallogr* D60, 1355 (2004)), and the dihedral angles were adjusted to the values observed in the homologous atoms of bound carazolol (16-22 in **Figure 24**). The one remaining unaccounted dihedral in (-)-isoproterenol was adjusted in order to place the catechol ring in the same plane as the C_{16}-C_{15}-O_{14} plane in carazolol. Residues known to specifically interact with agonists are shown as green sticks.

[0072] **Figure 26.** Packing interactions in the β_2AR that are likely to be modulated during the activation process. **A.** On the left, residues previously demonstrated to be CAMs (Rasmussen et al., *Mol Pharmacol* 56, 175 (1999); Tao, *et. al., Mol Endocrinol* 14, 1272 (2000); Jensen et al., *J Biol Chem* 276, 9279 (2001); Shi et al., *J Biol Chem* 277, 40989 (2002); Zuscik, et. al., (1998)) or UCMs (Strader et al., *Proc Natl Acad Sci U S A* 84, 4384 (1987); Chung, et al., *J Biol Chem* 263, 4052 (1988); Moro, et. al., *J Biol Chem* 269, 6651 (1994); Green, et. al., *J Biol Chem* 268, 23116 (1993); Gabilondo et al., *Proc Natl Acad Sci U S A* 94, 12285 (1997)) are shown as van der Waals spheres mapped onto a backbone cartoon of the β_2AR-T4L structure. On the right, residues that are found within 4Å of the CAMs Leu124$^{3.43}$ and Leu272$^{6.34}$ are shown as yellow spheres or dot surfaces. A vertical cross-section through the structure illustrates that these surrounding residues connect the CAMs on helices III and VI with the UCMs on helix VII through packing interactions. **B.** In both β_2AR-T4L (blue) and rhodopsin (purple), a network of ordered water molecules is found at the interface between the transmembrane helices at their cytoplasmic ends. **C.** Network of hydrogen bonding interactions between water molecules and β_2AR-T4L residues (sidechains as blue sticks), notably the UCMs on helix VII (orange cartoon).

DETAILED DESCRIPTION OF THE INVENTION

[0073] The invention described herein provides methods and compositions for generating crystal structures of membrane proteins that diffract to resolutions as low as 2 to 3 Ångstroms. In one embodiment, the methods involve the crystallization of proteins in a lipidic cubic phase, wherein the host lipid comprises an

additive, e.g., a sterol, such as cholesterol. The invention also provides the crystallized membrane proteins themselves, wherein the crystallized membrane proteins include GPCRs or modified GPCRs. The crystallized proteins can also include bound ligands, natural agonists, antagonists, and/or allosteric effectors. The invention additionally provides methods of using the 3-dimensional structures of the proteins (obtained from the crystals) to screen for novel ligands, drugs, and other useful molecules that affect the conformation and/or activity of the proteins *in vitro* or *in vivo*.

[0074] More specifically, the invention provides particular crystal forms of GPCRs diffracting to high resolutions. GPCRs have been grouped into five classes (Fredriksson, et al., *Mol Pharmacol* 63, 1256 (2003)) based on sequence conservation, with class A GPCRs, including β_2AR, being the largest and most studied. β_2AR agonists are used in the treatment of asthma and preterm labor (DeLano, *The PyMOL Molecular Graphics System* (2002) on World Wide Web at pymol.org). The crystal forms provided by the invention include several diffraction-quality class A GPCR crystals, including crystals comprising β2AR and crystals comprising the human adenosine A_{2A} receptor.

[0075] The invention provides a three-dimensional structure of a human β_2AR protein comprising a T4-lysozyme (T4L) in place of the third intracellular loop ("β_2AR-T4L") that has been solved in the presence of carazolol (2-propanol, 1-(*9H*-carbazol-4-yloxy)-3-[(1-methylethyl)amino] at 2.4 Å resolution. Additional class A GPCR structures make it possible to correlate sequence differences between GPCRs, e.g., between rhodopsin and β2AR, with emperically determined structural differences and extrapolate to other class A GPCRs. Highlighting interactions that constrain class A receptors into each of the two observed states allows a more comprehensive analysis of structural divergence and, therefore, more accurate models. Furthermore, GPCR structures provide an alternative signaling state on which to base homology models that will be more relevant for virtual ligand screening and structure-based drug design (Bissantz, *et. al, Proteins* 50, 5 (2003); Gouldson et al., *Proteins* 56, 67 (2004)).

Definitions

[0076] Terms used in the claims and specification are defined as set forth below unless otherwise specified. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs..

[0077] As used herein, the term "binding site" or "binding pocket" refers to a region of a protein that binds or interacts with a particular compound.

[0078] As used herein, the terms "binding" or "interaction" refers to a condition of proximity between a chemical entity, compound, or portions thereof, with another chemical entity, compound or portion thereof. The association or interaction can be non-covalent--wherein the juxtaposition is energetically favored by hydrogen bonding or van der Waals or electrostatic interactions--or it can be covalent.

[0079] As used herein, the term "residue" refers to an amino acid that is joined to another by a peptide bond. Residue is referred to herein to describe both an amino acid and its position in a polypeptide sequence.

[0080] As used herein, the term "surface residue" refers to a residue located on a surface of a polypeptide. In contrast, a buried residue is a residue that is not located on the surface of a polypeptide. A surface residue usually includes a hydrophilic side chain. Operationally, a surface residue can be identified computationally from a structural model of a polypeptide as a residue that contacts a sphere of hydration rolled over the surface of the molecular structure. A surface residue also can be identified experimentally through the use of deuterium exchange studies, or accessibility to various labeling reagents such as, e.g., hydrophilic alkylating agents.

[0081] As used herein, the term "polypeptide" refers to a single linear chain of 2 or more amino acids. A protein is an example of a polypeptide.

[0082] As used herein, the term "homolog" refers to a gene related to a second gene by descent from a common ancestral DNA sequence. The term, homolog, can apply to the relationship between genes separated by the event of speciation or to the relationship between genes separated by the event of genetic duplication.

[0083] As used herein, the term "conservation" refers to a high degree of similarity in the primary or secondary structure of molecules between homologs. This similarity is thought to confer functional importance to a conserved region of the molecule. In reference to an individual residue or amino acid, conservation is used to refer to a computed likelihood of substitution or deletion based on comparison with homologous molecules.

[0084] As used herein, the term "distance matrix" refers to the method used to present the results of the calculation of an optimal pairwise alignment score. The matrix field (i,j) is the score assigned to the optimal alignment between two residues (up to a total of i by j residues) from the input sequences. Each entry is calculated from the top-left neighboring entries by way of a recursive equation.

[0085] As used herein, the term "substitution matrix" refers to a matrix that defines scores for amino acid substitutions, reflecting the similarity of physicochemical properties, and observed substitution frequencies. These matrices are the foundation of statistical techniques for finding alignments.

[0086] As used herein, the term "pharmacophore" refers to an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger or block a biological response. A pharmacophore can be used to design one or more candidate compounds that comprise all or most of the ensemble of steric and electronic features present in the pharmacophore and that are expected to bind to a site and trigger or block a biological response.

[0087] As used herein, the term "G-protein coupled receptor" (or "GPCR") refers to a member of a family of heterotrimeric guanine-nucleotide binding protein ("G-protein") coupled receptors (Pierce, et al., *Nat. Rev. Mol. Cell. Biol.* 3:630 (2002)). GPCRs share a common structural signature of seven membrane-spanning helices with an extra-cellular N terminus and an intracellular C terminus. The family has been grouped into at least five classes (designated A, B, C, D, E, etc.; see, e.g., Fredriksson, et al., *Mol Pharmacol* 63, 1256 (2003)) based on sequence conservation. When used without a descriptive limitation, the term "a G-protein couple receptor" includes GPCRs with native amino acid sequences as well as genetically engineered or otherwise mutated GPCR proteins. Mutated GPCR proteins include those comprising point mutations, truncations, inserted sequences or other chemical modifications, while retaining ligand binding activity. One example of a GPCR referred to herein that comprises a point mutation is $\beta 2AR^{E122W}$. An example of a GPCR referred to herein that comprises an inserted T4 lysozyme sequence is the human A_{2a} adenosine receptor -T4L.

[0088] Adrenergic receptors in the class A or amine group are some of the most thoroughly investigated GPCRs (Kobilka, *Annu Rev Neurosci* 15, 87 (1992); Caron, et al., *Recent Prog Horm Res* 48, 277 (1993); Strosberg, *Protein Sci* 2, 1198 (1993); Hein, et al., *Trends Cardiovasc Med* 7, 137 (1997); Rohrer, *J Mol Med* 76, 764 (1998); Xiang, et al., *Adrenergic Receptors*, 267 (2006)), and are composed of two main subfamilies, α and β, which differ in tissue localization and ligand specificity, as well as in G protein coupling and downstream effector mechanisms (Milligan, et al., *Biochem Pharmacol* 48, 1059 (1994)). Some representative class A receptors include the human A_{2A} adenosine receptor and the beta-2 adrenergic receptor. The term "beta-2 adrenergic receptor" (or "β_2AR" or "β2AR") refers to a class A GPCR that responds to diffusable hormones and neurotransmitters and resides predominantly in smooth muscles throughout the body. When used without a descriptive limitation, the term "β2AR" includes β2ARs with native amino acid sequences as well as genetically engineered or otherwise mutated β2AR proteins. Mutated β2AR proteins include those comprising point mutations, truncations, inserted sequences or other chemical modifications, while retaining ligand binding activity. One example of a β2AR referred to herein that comprises a point mutation is $\beta 2AR^{E122W}$. An example

of a β2AR referred to herein that comprises an inserted T4 lysozyme sequence is the human adenosine receptor $\beta2AR^{E122W}$-T4L.

[0089] The term "diffracts to a resolution of xx-yy Angstroms" means that diffraction data exceeding a predetermined signal to noise ratio can be obtained within the stated resolution range. In some embodiments, that diffraction data can be obtained using synchrotron radiation. Also, in some embodiments, that diffraction data can be obtained following freezing of the crystal in liquid nitrogen.

[0090] As used herein, the term "atomic co-ordinates" refers to a set of three-dimensional co-ordinates for atoms within a molecular structure. In one embodiment, atomic-co-ordinates are obtained using X-ray crystallography according to methods well-known to those of ordinarily skill in the art of biophysics. Briefly described, X-ray diffraction patterns can be obtained by diffracting X-rays off a crystal. The diffraction data are used to calculate an electron density map of the unit cell comprising the crystal; said maps are used to establish the positions of the atoms (i.e., the atomic co-ordinates) within the unit cell. Those of skill in the art understand that a set of structure co-ordinates determined by X-ray crystallography contains standard errors. In other embodiments, atomic co-ordinates can be obtained using other experimental biophysical structure determination methods that can include electron diffraction (also known as electron crystallography) and nuclear magnetic resonance (NMR) methods. In yet other embodiments, atomic co-ordinates can be obtained using molecular modeling tools which can be based on one or more of ab initio protein folding algorithms, energy minimization, and homology-based modeling. These techniques are well known to persons of ordinary skill in the biophysical and bioinformatic arts, and are described in greater detail below.

[0091] Atomic co-ordinates for binding pockets, such as, e.g., the ligand binding pocket of β2AR, and/or other agonist/antagonist binding sites of the present invention are intended to encompass those co-ordinates set out in the .pdb file (**Appendix I**) (SEQ ID NOS 4-5, 1 and 6-9, respectively in order of appearance) incorporated into this specification, as well as co-ordinates that are substantially equivalent. Substantially equivalent co-ordinates are those that can be related to a reference set of co-ordinates by transformation reflecting differences in the choice of origin or inter-axis angels for one or more axes used to define the coordinate system. Operationally, co-ordinates are "substantially equivalent" when the structures represented by those co-ordinates can be superimposed in a manner such that root mean square deviations (RMSD) of atomic positions for the structures differs by less than a predetermined threshold. In some embodiments that threshold is less than about 5 Angstroms, or less than about 4 Angstroms, or less than about 3 Angstroms, or less than about 2 Angstroms, or

less than about 1 Angstrom, or less than about 0.9 Angstrom, or less than about 0.8 Angstrom, or less than about 0.7 Angstrom, or less than about 0.6 Angstrom, or less than about 0.5 Angstrom, or less than about 0.4 Angstrom, or less than about 0.3 Angstrom. Preferably, co-ordinates are considered "substantially equivalent" when the RMSD is less than about 1 Angstrom. Methods for structure superpositioning and RMSD calculations are well known to those of ordinary skill in the art, and can be carried out using programs such as, e.g., the programs listed in **Table 5** below.

[0092] Structural similarity can be inferred from, e.g., sequence similarity, which can be determined by one of ordinary skill through visual inspection and comparison of the sequences, or through the use of well-known alignment software programs such as CLUSTAL (Wilbur et al., *J. Proc. Natl. Acad. Sci. USA*, 80, 726 730 (1983)) or CLUSTALW (Thompson et al., *Nucleic Acids Research*, 22:4673 4680 (1994)) or BLAST.RTM. (Altschul et al., *J Mol. Biol.*, October 5;215(3):403 10 (1990)), a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. CLUSTAL W is available at the EMBL-EBI website (http://www.ebi.ac.uk/clustalw/); BLAST is available from the National Center for Biotechnology website (http ://www.ncbi.nlm.nih.gov/BLAST/). A residue within a first protein or nucleic acid sequence corresponds to a residue within a second protein or nucleic acid sequence if the two residues occupy the same position when the first and second sequences are aligned.

[0093] The term percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent "identity" can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[0094] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence co-ordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0095] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., infra).

[0096] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (NCBI web-site).

[0097] The term "sterol" refers to a subgroup of steroids with a hydroxyl group at the 3-position of the A-ring. *See* Fahy E. Subramaniam S et al., "A comprehensive classification system for lipids," *J. Lipid Res.* 46 (5):839-861 (2005)). Sterols are amphipathic lipids synthesized from acetyl-coenzyme A via the HMG-CoA reductase pathway. The overall molecule is quite flat. Sterols can include, e.g., cholesterol or cholesteryl hemisuccinate ("CHS").

[0098] The term "atomic co-ordinates for residues" refers to co-ordinates for all atoms associated with a residue, or for some of the atoms such as, e.g., side chain atoms.

[0099] The term "atomic co-ordinates of a candidate compound" refers to co-ordinates for all atoms comprising the compound or a subset of atoms comprising the compound.

[00100] The term "characterizing a binding interaction" refers to characterizing any observable property of a first molecule and determining an whether there is a change in that observable property after contacting the first molecule with a second molecule under conditions in which said first and second molecules can potentially bind.

[00101] The term "antagonist" refers to molecules that bind to and block the active site of a protein, but do not affect the equilibrium between inactive and active states. In contrast, an "agonist" is a ligand that shifts the equilibrium to an active receptor state. An "inverse agonist" is a ligand that acts to reduce the basal activity of a receptor through interactions that shift the equilibrium to more of an inactive state.

[00102] Ballesteros-Weinstein numbering appears in the text and Figures as superscripts to the protein numbering. Within each helix is a single most conserved residue among the class A GPCRs. This residue is designated X.50, where x is the number of the transmembrane helix. All other residues on that helix are numbered relative to this conserved position.

[00103] It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

Method for Producing Diffraction Quality Crystals of Membrane Proteins

[00104] In one aspect, the present invention discloses a modified lipidic cubic mesophase method for crystallizing proteins (see, e.g., Cherezov et al., *Biophysical J.*, v.83, 3393-3407 (2002)). The novel method described herein yields diffraction quality crystals of membrane proteins and is particularly useful for generating crystals of G-protein coupled receptor proteins ("GPCRs"). The method has now been successfully applied to diverse members of this important family of proteins, yielding crystals that diffract to resolutions in the 2.5 Å range. Among other advantages, this method allows diffraction-quality crystals of membrane proteins to be generated in the absence of any stabilizing antibodies bound to the protein in the crystal.

[00105] The LCP/sterol crystallization method described herein includes a step of mixing a solution containing the protein of interest with a host lipid or a host lipid mixture that includes a lipid additive. Given the teaching provided herein, one skilled in the art will recognize that a variety of host lipids may suffice for the generation of a cubic mesophase, e.g., hydrated monounsaturated monoacylglycerols such as monoolein, monopalmitolein, and/or monovacennin. The host lipid 1-monoolein is a preferred host lipid for certain applications of the method. In embodiments utilizing a lipid mixture, a lipid additive that is distinct from the host lipid is included, e.g., monounsaturated monoacylglycerols or other hydrophobic molecules known to interact with membranes or membrane-associated proteins such as, 2-monoolein, phosphotidylcholine, cardiolipin, lyso-PC, a polyethylene glyocol-lipid, dioleoylphosphatidylethanolamine ("DOPE"), DOPE-Me, dioleoyl phosphatidylcholine ("DOPC"), Asolectin, or a sterol (e.g., cholesterol, ergosterol, etc.). An example of a lipid mixture for GPCR crystallization is one comprising cholesterol as lipid additive in a ratio between 1 and 50%w/w relative to the host lipid, more preferably between 5 and 20%, and even more preferably between 8 and 12%. The protein mixture may include ligands of physiological interest and/or ligands that stabilize the protein. In the case of GPCRs, the ligands may include various agonists and antagonists known to the artisan, including well-known agonists such as carazolol (an inverse agonist), timolol, and other molecules including, without limitation,

Examples of ligands, include but are not limited to carazolol, light and olfactory stimulatory molecules; adenosine, bombesin, bradykinin, endothelin, y-aminobutyric acid (GABA), hepatocyte growth factor, melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), glucagon, acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., calcitonin, C5a anaphylatoxin, follicle stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin releasing hormone (TRH), and oxytocin).

[00106] A typical concentration of protein in the protein mixture is 25-75 mgs/ml but this concentration may vary according to protein identity and purification methods. As will be recognized by the skilled artisan, the concentration must be high enough to result in a degree of insolubility sufficient for nucleation to occur after a precipitation solution is combined with the protein-laden lipid solution; on the other hand, concentrations of protein that are too high may prevent the orderly growth of high-quality crystals.

[00107] The lipid mixture is preferably combined with the protein mixture and homogenized, e.g., using a syringe mixer, spontaneously yielding a homogenous cubic phase. Typically, the lipid mixture is added to the protein solution at a ratio of 1:1, 3:2, 4:2 w/w lipid:protein, but this ratio may be varied by the skilled artisan as desired, depending on various parameters, e.g., the concentration of protein in the protein mixture. The protein-laden lipidic cubic phase preparation thus obtained is then combined with precipitation solution (also referred to as crystallization solution) on or in an appropriate surface or container, e.g., a glass sandwich plate with wells where the mixed solutions can incubate while crystallization occurs. A typical volume of the protein-laden lipidic cubic phase used in the method is between 10 and 100 nL, with 40 to 60 nL preferred in certain embodiments. A typical volume of precipitation solution is 20 to 100 times greater, e.g., for a 20 nL volume of protein-laden lipidic cubic phase, approximately 1 μL of precipitation solution would be added to initiate crystallization.

[00108] The precipitation solution used in the crystallization method is an appropriately buffered solution (i.e., buffered to approximate the physiological conditions of the native protein) comprising polyethylene glycol, a salt, and optionally a small soluble molecule such as an alcohol.

[00109] With respect to the polyethylene glycol in the precipitation solution, useful PEG molecules include PEG 300, PEG 400, PEG 550, PEG 550mme, PEG 1000, and PEG 1500, as well as other PEG molecules with

average molecular weights less than 2000. In certain embodiments, larger average molecular weight PEG molecules (up to 20,000) or modified PEG molecules may be preferred. In some embodiments, the PEG or modified PEG has an average molecular weight of 400. Examples of modified PEG include but are not limited to PEG laurate, PEG dilaurate, PEG oleate, PEG dioleate, PEG stearate, PEG distearate, PEG glyceryl trioleate, PEG glyceryl laurate, PEG glyceryl stearate, PEG glyceryl oleate, PEG palm kernel oil, PEG hydrogenated castor oil, PEG castor oil, PEGcorn oil, PEG caprate/caprylate glycerides, PEG caprate/caprylate glycerides, PEG cholesterol, PEG phyto sterol, PEG soya sterol, PEG trioleate, PEG sorbitan oleate, PEG sorbitan laurate, PEG succinate, PEG nonyl phenol series, PEG octyl phenol series, Methyl-PEG, PEG-Maleimide, PEG4-NHS Ester and methoxypoly(ethylene glycol) (mPEG).

[00110] PEG may be present in the crystallization solution in concentrations between 10-60 %v/v, and most typically between 20-40 %v/v. The preferred concentration will vary depending on the average molecular weight of PEG utilized, i.e., 10-60 %v/v of PEG will be preferred for PEG $\leq$ 1000 whereas 10-30 %w/v will be preferred for PEG > 1000 (larger average molecular weight PEG formulations are described in %w/v rather than %v/v).

[00111] With respect to the salt used in the method, an optimum cation can usually be found for a given crystal. Both sodium and lithium sulfate have proven useful for obtaining high resolution proteins of GPCRs. Again, the concentrations may be varied up to 1M, with lower concentrations of approximately 50-200 mM typically preferred. Other organic salts, e.g., citrate, malonate, tartrate, formate and acetate, may also be screened for their effects on crystal formation. In certain embodiments, the precipitation solution additionally comprises a small organic molecule such as an alcohol, a diol or a triol, e.g., a hexanediol, a butanediol, or derivative thereof. These molecules may be present in the precipitation solution in various concentrations, as appropriate, but typically in the range of 1-20 %v/v, more typically in the 5-10 %v/v range. In certain embodiments, preferred combinations of lipid additives (in the protein-laden lipidic cubic phase mixture) and small molecules (in the precipitation solution) yield optimal results. Examples of such combinations include 1,4-butanediol in combination with DOPE or cholesterol, and 2,6-hexanediol in combination with cholesterol.

[00112] In optimizing the conditions from micro-crystals to larger crystals for a given system (e.g., a protein/ligand system), the choice and concentration of a specific sterol(s) and specific lipid(s), as well the pH, buffer system, salt, and salt concentration may be varied, as in other types of crystallization formats. As noted above, small organic additives, especially alcohols and diols such as 1,4 butanediol, 1,6 hexanediol, etc., can be particularly useful in generating large diffraction quality crystals. Also, due to the membrane fluidity-altering

properties of cholesterol and other sterols, sterol and precipitant concentration should be treated as dependent variables. For example, increasing concentrations of cholesterol in monoolein serve to rigidify the membranes, potentially slowing diffusion of the membrane protein within the lipid matrix. Conversely, increasing concentrations of PEG 400 swell the cubic phase, thereby increasing the lattice parameter of the matrix and speeding diffusion within the lipid. The former scenario would slow the rate of crystallization while the latter would increase the rate. The two effects should therefore be balanced for optimal nucleation and also for optimal growth of large, well-ordered crystals that diffract to a high resolution.

[00113] The mixing of the protein-laden lipidic cubic phase solution and the precipitation solution typically occurs at room temperatures. After set-up, the plates containing the mixed crystallization solutions can be monitored as often as desired for the appearance of crystal growth. One skilled in the art will recognize that further optimization of these conditions may be desirable, for example, to maximize the size and number of diffraction quality crystals that are obtained. In making determinations as to the preferred molecules and conditions for crystallization, the skilled artisan may rely on well-known phase diagrams and other previously determined physical constants, in addition to the novel methodology and Examples described herein. For certain lipid mixtures, pre-screening their phase behavior by microscope visualization and/or by X-ray prior to being mixed with the protein solution may facilitate the process of optimization. An *in meso* crystallization robot and automatic imager combined with multiple 96-well optimization screens can be used to run thousands of trials in a relatively facile manner.

[00114] It also possible to achieve additional stabilization of proteins and improve the yield of diffraction-quality crystals using the LCP/sterol method described herein by modification of the protein. For example, an unstable region of the protein may be replaced or stabilized by incorporation of a portion of a stable protein, e.g., a T4 lysozyme, whose structure is previously known but which does not (when fused) significantly affect the biochemical activity of the protein of interest. For example, the ECL2 and ECL3 regions of a β2AR can be stabilized by such modifications, as described herein (Examples 3 and 4). Other modifications include one or more point mutations that do not significantly alter the properties of the protein of interest except to increase its stability and/or tendency to crystallize well. For example, $\beta 2AR^{(E122W)}$ comprises an E122W point mutation and yields crystals with the LPC/sterol method. Analogous residues in other GPCRs could be modified in the same way. One advantage of the LCP method applied to both modified and unmodified proteins, as noted above, is that it allows (but does not preclude) the crystallization of proteins in the absence of heterologous proteins, such as antibodies, that may not be of interest to the crystallographer.

Method of ligand screening by lipidic cubic phase crystallization

[00115] Aspects of the lipidic cubic phase crystallization methodology described above can be modified for the purpose of determining low affinity lipid binding sites within integral membrane proteins through co-crystallization trials within a lipidic cubic phase matrix. In this method various lipids of different composition are incorporated at a variety of concentrations into monoolein, wherein the monoolein is solubilized in chloroform or heated to its fluid isotropic phase. Crystal growth is then assessed by visual inspections and diffraction data collected on any crystalline material within the experiment. Because the lipid is low affinity the method requires an environment conducive for free exchange of lipid from annular to non-annular protein binding sites. The presence of interpretable electron density not associated with crystal packing interfaces allows the inference of specific binding sites for a particular lipid within the context of the membrane protein in a membrane environment. Because the binding occurs within a membrane the complicating factor of detergent partitioning is eliminated and thermodynamics of association are more realistic. The method thus allows one to characterize in detail previously inaccessible regions of membrane proteins, as well as describe and exploit binding interactions that might otherwise remain undetected. Furthermore, the technique can be applied to ligand binding studies where the ligand occupies a site on the membrane protein that is juxtaposed to the lipid plane relying on partitioning into the aqueous phase to enable saturation of the site. This limits the exposure of the protein to harsh organic co-solvents and may also find utility for soluble proteins that crystallize within the lipidic cubic phase solvent channels.

[00116] By wan of example, existing crystallization conditions for a protein can be utilized as a starting point for screening novel ligands to the β2-adrenergic receptor. In the first instance, cholesterol solubilized in chloroform can be incorporated into chloroform-solubilized monoolein at weight ratio of 10%. After drying and desiccating the mixture, protein at 30-80 mg/mL can incorporated at 2/3 volume ratio and used for crystallization trials. A similar protocol was used for other lipid like molecules, including cholesteryl hemisuccinate and a variety of other cholesterol analogs. In each case protein was incorporated into the resulting mixture and screened for crystallization. Binding of the novel ligand to the receptor is indicated by diffraction quality crystals and ultimately by three-dimensional structural data. By incorporating analogues of cholesterol we are able to map out the binding specificities based on the unique structural features of their respective sterol rings and polar moieties and, if their incorporation led to diffraction quality crystals, the interactions between the protein and cholesterol analogue are determined.

[00117] This method of ligand screening is not limited to lipid-like molecules, as we can use the lipidic cubic phase as a host for other highly hydrophobic molecules that act at orthosteric binding sites. One problem with structure based or fragment based design of novel ligands is the hydrophobicity often associated with potential drug leads. This is a problem in aqueous based crystallization schemes because the solubility of the ligand is often less than 1 mM and unless there is a slow off rate from the protein of interest the binding site will be in a ligand depleted state at crystallization conditions which often involve protein concentrations between 0.5 and 1mM. One can attempt to co-solubilize the hydrophobic ligand in aqueous miscible organic solvent such as dimethyl sulfoxide (DMSO) or dimethyl formamide (DMF). However, these often interfere with the stability or crystallization properties of the protein and their usefulness is not general. Therefore, this method allows one to incorporate the hydrophobic ligand directly into the lipidic cubic phase where its accessibility to the protein will be limited by the partitioning between the lipid and aqueous phase and/or the accessibility of the binding.

[00118] Additional guidance relating to these methods is provided by the working and prophetic examples of protein crystallization presented herein.

The Crystal Structure of Human β2AR Bound to Carazolol and Uses Thereof

[00119] G-protein coupled receptors are cell surface receptors that indirectly transduce extracellular signals to downstream effectors, e.g., intracellular signaling proteins, enzymes, or channels. G-protein coupled receptor membrane proteins are grouped into one of at least 6 classes (i.e., A, B, C, D, E, and F). An example of a mammalian G-protein coupled receptor is the β2A receptor, a receptor in the Class A subfamily of GPCRs.

[00120] Class A GPCRs function in a variety of physiological processes such as vasodilation, bronchodilation, neurotransmitter signaling, stimulation of endocrine secretions, gut peristalsis, development, mitogenesis, cell proliferation, cell migration, immune system function, and oncogenesis. Accordingly, class A GPCRs can be used as screening targets to identify modulators of these processes which can then function to ameliorate diseases associated with these processes, e.g., cancer and autoimmunity.

[00121] The 2.4 Angstrom structure of β2AR bound to carazolol, described herein (PDB coordinates appear in **Appendix I)** (SEQ ID NOS 4-5, 1 and 6-9, respectively in order of appearance) can be used as a model for rationally designing pharmacophore and/or candidate compounds, either de novo or by modification of known compounds. As noted below, the multiple ligand binding sites in this structure include amino acids that are highly conserved across a large number of class A G protein coupled receptors (GPCRs) indicating that the 2.4 Angstrom structure of β2AR can be used for the rational designing of ligands (*e.g.*, therapeutic compounds)

that bind to this receptor and others. Pharmacophore and candidate compounds identified through the use of the crystal structure co-ordinates will have utility as pharmaceuticals due to their ability to alter the structure and/or binding properties of β2AR. Pharmacophores and candidate compounds can be determined according to any method known in the art, including the methods described in U.S. Pat. No. 5,888,738 to Hendry, and the methods described in U.S. Pat. No. 5,856,116 to Wilson et al. the disclosures of which both are incorporated by reference in their entirety for all purposes.

[00122] The structure data provided herein can be used in conjunction with computer-modeling techniques to develop models of sites on the human β2AR or related GPCRs selected by analysis of the crystal structure data. The site models characterize the three-dimensional topography of site surface, as well as factors including van der Waals contacts, electrostatic interactions, and hydrogen-bonding opportunities. Computer simulation techniques can be used to map interaction positions for functional groups including protons, hydroxyl groups, amine groups, divalent cations, aromatic and aliphatic functional groups, amide groups, alcohol groups, etc. that are designed to interact with the model site. These groups can be designed into a pharmacophore or candidate compound with the expectation that the candidate compound will specifically bind to the site. Pharmacophore design thus involves a consideration of the ability of the candidate compounds falling within the pharmacophore to interact with a site through any or all of the available types of chemical interactions, including hydrogen bonding, van der Waals, electrostatic, and covalent interactions, although, in general, and preferably, pharmacophores interact with a site through non-covalent mechanisms.

[00123] The ability of a pharmacophore or candidate compound to bind to the human β2AR can be analyzed prior to actual synthesis using computer modeling techniques. Only those candidates that are indicated by computer modeling to bind the target with sufficient binding energy (*i.e.*, binding energy corresponding to a dissociation constant with the target on the order of 10^{-2} M or tighter) can be synthesized and tested for their ability to bind to the human β2AR using binding assays or functional assays known to those of skill in the art. The computational evaluation step thus avoids the unnecessary synthesis of compounds that are unlikely to bind β2AR or one or more of its constitutive binding sites, or the related binding sites of another GPCR with adequate affinity.

[00124] A human β2AR or candidate compound(s) can be computationally evaluated and designed by means of a series of steps in which chemical entities or fragments are screened and selected for their ability to associate with individual binding target sites on β2AR or binding site thereof, including, but not limited to a binding

pocket of the human β2AR. One skilled in the art can use one of several methods to screen chemical entities or fragments for their ability to associate with one or more of these human β2AR binding sites. For example, increased affinity and specificity may be designed into caffeine and other xanthine molecules by combining interactions with both xanthine and non-xanthine binding sites.

[00125] The process can begin by visual inspection of, for example a target site on a computer screen, based on the human β2AR co-ordinates, or a subset of those co-ordinates (*e.g.*, binding pocket residues V117, T118, F193, Y199, A200, W286, F289, F290, and Y316), as set forth in **Appendix I** (SEQ ID NOS 4-5, 1 and 6-9, respectively in order of appearance). Selected fragments or chemical entities can then be positioned in a variety of orientations or "docked" within a target site of the human β2AR as defined from analysis of the crystal structure data. Docking can be accomplished using software such as Quanta (Molecular Simulations, Inc., San Diego, Calif.) and Sybyl (Tripos, Inc. St. Louis, Mo.) followed by energy minimization and molecular dynamics with standard molecular mechanics forcefields such as CHARMM (Molecular Simulations, Inc., San Diego, Calif.), ICM (Molsoft, San Diego, Calif.), and AMBER (University of California, San Francisco).

[00126] Specialized computer programs can also assist in the process of selecting fragments or chemical entities. These include but are not limited to: GRID (Goodford, P. J., "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules," *J. Med. Chem.*, 28, pp. 849 857 (1985)); GRID is available from Oxford University, Oxford, UK; MCSS (Miranker, A. and M. Karplus, "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method," *Proteins: Structure, Function and Genetics*, 11, pp. 29 34 (1991)); MCSS is available from Molecular Simulations, Inc., San Diego, Calif.; AUTODOCK (Goodsell, D. S. and A. J. Olsen, "Automated Docking of Substrates to Proteins by Simulated Annealing," Proteins: Structure, Function, and Genetics, 8, pp. 195 202 (1990)); AUTODOCK is available from Scripps Research Institute, La Jolla, Calif.; DOCK (Kuntz, I. D., et al. "A Geometric Approach to Macromolecule-Ligand Interactions," *J. Mol. Biol.*, 161, pp. 269 288 (1982)); DOCK is available from University of California, San Francisco, Calif.; CERIUS II (available from Molecular Simulations, Inc., San Diego, Calif.); and Flexx (Raret, et al. *J. Mol. Biol.* 261, pp. 470 489 (1996)).

[00127] After selecting suitable chemical entities or fragments, they can be assembled into a single compound. Assembly can proceed by visual inspection of the relationship of the fragments to each other on a three-dimensional image of the fragments in relation to the human β2AR or its binding sites or those of a related

GPCR receptor structure or portion thereof displayed on a computer screen. Visual inspection can be followed by manual model building using software such as the Quanta or Sybyl programs described above.

[00128] Software programs also can be used to aid one skilled in the art in connecting the individual chemical entities or fragments. These include, but are not limited to CAVEAT (Bartlett, P. A., et al. "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules" In "Molecular Recognition in Chemical and Biological Problems," Special Publ, *Royal Chem. Soc.*, 78, pp. 182 196 (1989)); CAVEAT is available from the University of California, Berkeley, Calif.; 3D Database systems such as MACCS-3D (MDL Information Systems, San Leandro, Calif.); this area is reviewed in Martin, Y. C., "3D Database Searching in Drug Design," *J. Med. Chem.*, 35:2145 2154 (1992)); and HOOK (available from Molecular Simulations Inc., San Diego, Calif.).

[00129] As an alternative to building candidate pharmacophores or candidate compounds up from individual fragments or chemical entities, they can be designed de novo using the structure of the β2AR, its constituent ligand binding pocket, or the homologous cavities in a related GPCR, optionally, including information from co-factor(s) or known activators or inhibitor(s) that bind to the target site. De novo design can be implemented by programs including, but not limited to LUDI (Bohm, H. J., "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors," *J. Comp. Aid. Molec. Design*, 6, pp. 61 78 (1992)); LUDI is available from Molecular Simulations, Inc., San Diego, Calif.; LEGEND (Nishibata, Y., and Itai, A., *Tetrahedron* 47, p. 8985 (1991); LEGEND is available from Molecular Simulations, San Diego, Calif.; and LeapFrog (available from Tripos Associates, St. Louis, Mo.).

[00130] The functional effects of known β2AR also can be altered through the use of the molecular modeling and design techniques described herein. This can be carried out by docking the structure of the known ligand on a human A_{2A} adenosine receptor or a model structure of one or more binding sites of the human β2AR (*e.g.*, the binding pocket described herein) and modifying the shape and charge distribution of the ligand or protein model structure to optimize the binding interactions between the ligand and protein. The modified structure can be synthesized or obtained from a library of compounds and tested for its binding affinity and/or effect on ribosome function. Of course, where the crystal structure of a complex between a human β2AR (or subunit thereof) and a ligand is known, comparisons between said complex and the structures of the present invention can be made to gain additional information about alterations in human β2AR conformation that occur upon ligand binding. This information can be used in design of optimized ligands. Compounds that interfere or

activate human β2AR function (e.g., by interacting with a binding pocket) are especially well suited for the docking, co-crystallization, and optimization applications of the present invention.

[00131] Additional molecular modeling techniques also can be employed in accordance with the invention. See, e.g., Cohen, N. C., et al. "Molecular Modeling Software and Methods for Medicinal Chemistry," *J. Med. Chem.*, 33, pp. 883 894 (1990); Hubbard, Roderick E., "Can drugs be designed?" *Curr. Opin. Biotechnol.* 8, pp. 696 700 (1997); and Afshar, et al. "Structure-Based and Combinatorial Search for New RNA-Binding Drugs," *Curr. Opin. Biotechnol.* 10, pp. 59 63 (1999).

[00132] Following pharmacophore or candidate compound design or selection according to any of the above methods or other methods known to one skilled in the art, the efficiency with which a candidate compound falling within the pharmacophore definition binds to the human β2AR or its ligand binding site, or alternatively binds to a related GPCR or homologous portions thereof, can be tested and optimized using computational evaluation. A candidate compound can be optimized, e.g., so that in its bound state it would preferably lack repulsive electrostatic interaction with the target site. These repulsive electrostatic interactions include repulsive charge-charge, dipole-dipole, and charge-dipole interactions. It is preferred that the sum of all electrostatic interactions between the candidate compound and the human β2AR, including its ligand binding site when the candidate compound is bound to the target make a neutral or favorable contribution to the binding enthalpy or free energy.

[00133] Specific computer software is available in the art to evaluate compound deformation energy and electrostatic interactions. Examples of programs designed for such uses include, but are not limited to Gaussian 92, revision C (Frisch, M. J., Gaussian, Inc., Pittsburgh, Pa. (1992)); AMBER, version 4.0 (Kollman, P. A., University of California at San Francisco, (1994)); QUANTA/CHARMM (Molecular Simulations, Inc., San Diego, Calif. (1994)); and Insight II/Discover (Biosym Technologies Inc., San Diego, Calif. (1994)). These programs can be run, using, e.g., a Silicon Graphics workstation, Indigo, 02-R10000 or IBM RISC/6000 workstation model 550. Other hardware and software combinations can be used to carry out the above described functions, and are known to those of skill in the art. In general, the methods described herein, particularly computer-implemented methods, comprise a step of recording or storing data onto a medium, wherein the medium can include a computer-readable medium. Additionally, or alternatively, the methods comprise a step of reporting or communicating the data to a user of interest, e.g., an operator of the device and/or computer that is employed in the method; or the computer can perform an additional useful task, e.g.,

alert the operator of the computer that a function has been completed, upon completing one or more determining steps of the method.

[00134] Once a pharmacophore or candidate compound has been optimally selected or designed, as described above, substitutions can then be made in some of its atoms or side groups to improve or modify its binding properties. Generally, initial substitutions are conservative in that the replacement group will have approximately the same size, shape, hydrophobicity and charge as the original group. Components known in the art to alter conformation should be avoided in making substitutions. Substituted candidates can be analyzed for efficiency of fit to the human β2AR (or one or more binding sites of the human β2AR) using the same methods described above.

[00135] **Assays**

[00136] Any one of a number of assays of function known to those of skill in the art can be used to determine the biological activity of candidate compounds.

[00137] Candidate compound interaction with the human β2AR (or one or more binding sites of human β2AR) or to a related GPCR or portion thereof can be evaluated using direct binding assays including filter binding assays, such as are known to those skilled in the art. Binding assays can be modified to evaluate candidate compounds that competitively inhibit the binding of, e.g., known human β2AR binding compounds including xanthine and xanthine-based compounds such as theophylline, theobromine and caffeine. These and other assays are described in International Publication WO 00/69391, the entire disclosure of which is incorporated by reference in its entirety for all purposes. Methods of assaying for modulators of ligand binding and signal transduction include in vitro ligand binding assays using GPCRs, such as human β2AR (or one or more binding sites selected from the binding pockets I, II and III of the human β2AR), portions thereof such as the extracellular domain, or chimeric proteins comprising one or more domains of a GPCR, oocyte GPCR expression or tissue culture cell GPCR expression, either naturally occurring or recombinant; membrane expression of a GPCR, either naturally occurring or recombinant; tissue expression of a GPCR; expression of a GPCR in a transgenic animal, etc.

[00138] As noted above, GPCRs and their alleles and polymorphic variants are G-protein coupled receptors that participate in signal transduction and are associated with cellular function in a variety of cells, e.g., neurons, immune system cells, kidney, liver, colon, adipose, and other cells. The activity of GPCR polypeptides can be assessed using a variety of in vitro and in vivo assays to determine functional, chemical, and physical effects,

e.g., measuring ligand binding, (e.g., radioactive ligand binding), second messengers (e.g., cAMP, cGMP, IP_3, DAG, or Ca^{2+}), ion flux, phosphorylation levels, transcription levels, neurotransmitter levels, and the like. Such assays can be used to test for inhibitors and activators of a GPCR. In particular, the assays can be used to test for compounds that modulate natural ligand-induced GPCR activity, for example, by modulating the binding of the natural ligand to the receptor and/or by modulating the ability of the natural ligand to activate the receptor. Typically in such assays, the test compound is contacted with the GPCR in the presence of the natural ligand. The natural ligand can be added to the assay before, after, or concurrently with the test compound. The results of the assay, for example, the level of binding, calcium mobilization, etc. is then compared to the level in a control assay that comprises the GPCR and natural ligand in the absence of the test compound.

[00139] Screening assays of the invention are used to identify modulators that can be used as therapeutic agents, e.g., antagonists of GPCR activity. For example, carazolol is a known high-affinity inverse agonist of human β2AR.

[00140] The effects of test compounds upon the function of the GPCR polypeptides can be measured by examining any of the parameters described above. Any suitable physiological change that affects GPCR activity can be used to assess the influence of a test compound on the GPCRs and natural ligand-mediated GPCR activity. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects such as transmitter release, hormone release, transcriptional changes to both known and uncharacterized genetic markers (e.g., northern blots), changes in cell metabolism such as cell growth or pH changes, and changes in intracellular second messengers such as Ca^{2+}, IP_3 or cAMP.

[00141] For a general review of GPCR signal transduction and methods of assaying signal transduction, *see*, e.g., Methods in Enzymology, vols. 237 and 238 (1994) and volume 96 (1983); Bourne et al., Nature 10:349:117-27 (1991); Bourne et al., Nature 348:125-32 (1990); Pitcher et al., Annu. Rev. Biochem. 67:653-92 (1998).

[00142] Modulators of GPCR activity are tested using GPCR polypeptides, either recombinant or naturally occurring. The protein can be isolated, expressed in a cell, expressed in a membrane derived from a cell, expressed in tissue or in an animal, either recombinant or naturally occurring. For example, neurons, cells of the immune system, adipocytes, kidney cells, transformed cells, or membranes can be used. Modulation is tested using one of the in vitro or in vivo assays described herein or others as generally known in the art. Signal transduction can also be examined in vitro with soluble or solid state reactions, using a chimeric molecule such

as an extracellular domain of a receptor covalently linked to a heterologous signal transduction domain, or a heterologous extracellular domain covalently linked to the transmembrane and or cytoplasmic domain of a receptor. Furthermore, ligand-binding domains of the protein of interest can be used in vitro in soluble or solid state reactions to assay for ligand binding.

[00143] Ligand binding to a human β2AR (or one or more binding sites thereof) or a chimeric protein derivative can be tested in a number of formats. For example, binding can be performed in solution, in a bilayer membrane, attached to a solid phase, in a lipid monolayer, or in vesicles. Typically, in an assay of the invention, the binding of the natural ligand to its receptor is measured in the presence of a candidate modulator. Alternatively, the binding of the candidate modulator can be measured in the presence of the natural ligand. Often, competitive assay that measure the ability of a compound to compete with binding of the natural ligand to the receptor are used. Binding can be measured by assessing GPCR activity or by other assays: binding can be tested by measuring e.g., changes in spectroscopic characteristics (e.g., fluorescence, absorbance, refractive index), hydrodynamic (e.g., shape) changes, or changes in chromatographic or solubility properties.

[00144] Receptor-G-protein interactions can also be used to assay for modulators. For example, in the absence of GTP, binding of an activator such as the natural ligand will lead to the formation of a tight complex of a G protein (all three subunits) with the receptor. This complex can be detected in a variety of ways, as noted above. Such an assay can be modified to search for inhibitors. For example, a ligand can be added to the human β2AR and G protein in the absence of GTP to form a tight complex. Inhibitors can be identified by looking at dissociation of the receptor-G protein complex. In the presence of GTP, release of the alpha subunit of the G protein from the other two G protein subunits serves as a criterion of activation.

[00145] An activated or inhibited G-protein will in turn alter the properties of downstream effectors such as proteins, enzymes, and channels. The classic examples are the activation of cGMP phosphodiesterase by transducin in the visual system, adenylate cyclase by the stimulatory G-protein, phospholipase C by G_q and other cognate G proteins, and modulation of diverse channels by Gi and other G proteins. Downstream consequences such as generation of diacyl glycerol and IP_3 by phospholipase C, and in turn, for calcium mobilization e.g., by IP_3 can also be examined. Thus, modulators can be evaluated for the ability to stimulate or inhibit ligand-mediated downstream effects. In other examples, the ability of a modulator to activate a GPCR expressed in adipocytes in comparison to the ability of a natural ligand, can be determined using assays such as lipolysis (*see*, e.g., WO01/61359).

[00146] Activated GPCRs become substrates for kinases that phosphorylate the C-terminal tail of the receptor (and possibly other sites as well). Thus, activators will promote the transfer of ^{32}P from gamma-labeled GTP to the receptor, which can be assayed with a scintillation counter. The phosphorylation of the C-terminal tail will promote the binding of arrestin-like proteins and will interfere with the binding of G-proteins. The kinase/arrestin pathway plays a key role in the desensitization of many GPCR receptors. Modulators can therefore also be identified using assays involving beta-arrestin recruitment. Beta-arrestin serves as a regulatory protein that is distributed throughout the cytoplasm in unactivated cells. Ligand binding to an appropriate GPCR is associated with redistribution of beta-arrestin from the cytoplasm to the cell surface, where it associates with the GPCR. Thus, receptor activation and the effect of candidate modulators on ligand-induced receptor activation, can be assessed by monitoring beta-arrestin recruitment to the cell surface. This is frequently performed by transfecting a labeled beta-arrestin fusion protein (e.g., beta-arrestin-green fluorescent protein (GFP)) into cells and monitoring its distribution using confocal microscopy (*see*, e.g., Groarke et al., J. Biol. Chem. 274(33):23263-69 (1999)).

[00147] Receptor internalization assays can also be used to assess receptor function. Upon ligand binding, the G-protein coupled receptor--ligand complex is internalized from the plasma membrane by a clathrin-coated vesicular endocytic process; internalization motifs on the receptors bind to adaptor protein complexes and mediate the recruitment of the activated receptors into clathrin-coated pits and vesicles. Because only activated receptors are internalized, it is possible to detect ligand-receptor binding by determining the amount of internalized receptor. In one assay format, cells are transiently transfected with radiolabeled receptor and incubated for an appropriate period of time to allow for ligand binding and receptor internalization. Thereafter, surface-bound radioactivity is removed by washing with an acid solution, the cells are solubilized, and the amount of internalized radioactivity is calculated as a percentage of ligand binding. *See*, e.g., Vrecl et al., Mol. Endocrinol. 12:1818-29 (1988) and Conway et al., J. Cell Physiol. 189(3):341-55 (2001). In addition, receptor internalization approaches have allowed real-time optical measurements of GPCR interactions with other cellular components in living cells (*see*, e.g., Barak et al., Mol. Pharmacol. 51(2)177-84 (1997)). Modulators can be identified by comparing receptor internalization levels in control cells and cells contacted with candidate compounds. For example, candidate modulators the human β2AR are assayed by examining their effects on receptor internalization upon binding of the natural ligand.

[00148] Another technology that can be used to evaluate GPCR-protein interactions in living cells involves bioluminescence resonance energy transfer (BRET). A detailed discussion regarding BRET can be found in Kroeger et al., J. Biol. Chem., 276(16):12736-43 (2001).

[00149] Receptor-stimulated guanosine 5'-O-(γ-Thio)-Triphosphate ([^{35}S]GTPγS) binding to G-proteins can also be used as an assay for evaluating modulators of GPCRs. [^{35}S]GTPγS is a radiolabeled GTP analog that has a high affinity for all types of G-proteins, is available with a high specific activity and, although unstable in the unbound form, is not hydrolyzed when bound to the G-protein. Thus, it is possible to quantitatively assess ligand-bound receptor by comparing stimulated versus unstimulated [^{35}S]GTPγS binding utilizing, for example, a liquid scintillation counter. Inhibitors of the receptor-ligand interactions would result in decreased [^{35}S]GTPγS binding. Descriptions of [^{35}S]GTPγS binding assays are provided in Traynor and Nahorski, Mol. Pharmacol. 47(4):848-54 (1995) and Bohn et al., Nature 408:720-23 (2000).

[00150] The ability of modulators to affect ligand-induced ion flux can also be determined. Ion flux can be assessed by determining changes in polarization (i.e., electrical potential) of the cell or membrane expressing a GPCR. One means to determine changes in cellular polarization is by measuring changes in current (thereby measuring changes in polarization) with voltage-clamp and patch-clamp techniques, e.g., the "cell-attached" mode, the "inside-out" mode, and the "whole cell" mode (*see*, e.g., Ackerman et al., New Engl. J. Med. 336:1575-1595 (1997)). Whole cell currents are conveniently determined using the standard methodology (*see*, e.g., Hamil et al., Pflügers. Archiv. 391:85 (1981). Other known assays include: radiolabeled ion flux assays and fluorescence assays using voltage-sensitive dyes (*see*, e.g., Vestergarrd-Bogind et al., J. Membrane Biol. 88:67-75 (1988); Gonzales & Tsien, Chem. Biol. 4:269-277 (1997); Daniel et al., J. Pharmacol. Meth. 25:185-193 (1991); Holevinsky et al., J. Membrane Biology 137:59-70 (1994)). Generally, the compounds to be tested are present in the range from 1 pM to 100 mM.

[00151] Preferred assays for G-protein coupled receptors include cells that are loaded with ion or voltage sensitive dyes to report receptor activity. Assays for determining activity of such receptors can also use known agonists and antagonists for other G-protein coupled receptors and the natural ligands disclosed herein as negative or positive controls to assess activity of tested compounds. In assays for identifying modulatory compounds (e.g., agonists, antagonists), changes in the level of ions in the cytoplasm or membrane voltage are monitored using an ion sensitive or membrane voltage fluorescent indicator, respectively. Among the ion-sensitive indicators and voltage probes that can be employed are those disclosed in the Molecular Probes 1997 Catalog. For G-protein coupled receptors, promiscuous G-proteins such as Gα15 and Gα16 can be used in the

assay of choice (Wilkie et al., Proc. Nat'l Acad. Sci. USA 88:10049-10053 (1991)). Such promiscuous G-proteins allow coupling of a wide range of receptors to signal transduction pathways in heterologous cells.

[00152] Receptor activation by ligand binding typically initiates subsequent intracellular events, e.g., increases in second messengers such as IP_3, which releases intracellular stores of calcium ions. Activation of some G-protein coupled receptors stimulates the formation of inositol triphosphate (IP_3) through phospholipase C-mediated hydrolysis of phosphatidylinositol (Berridge & Irvine, Nature 312:315-21 (1984)). IP_3 in turn stimulates the release of intracellular calcium ion stores. Thus, a change in cytoplasmic calcium ion levels, or a change in second messenger levels such as IP_3 can be used to assess G-protein coupled receptor function. Cells expressing such G-protein coupled receptors can exhibit increased cytoplasmic calcium levels as a result of contribution from both intracellular stores and via activation of ion channels, in which case it can be desirable although not necessary to conduct such assays in calcium-free buffer, optionally supplemented with a chelating agent such as EGTA, to distinguish fluorescence response resulting from calcium release from internal stores.

[00153] Other assays can involve determining the activity of receptors which, when activated by ligand binding, result in a change in the level of intracellular cyclic nucleotides, e.g., cAMP or cGMP, by activating or inhibiting downstream effectors such as adenylate cyclase. There are cyclic nucleotide-gated ion channels, e.g., rod photoreceptor cell channels and olfactory neuron channels that are permeable to cations upon activation by binding of cAMP or cGMP (*see*, e.g., Altenhofen et al., Proc. Natl. Acad. Sci. U.S.A. 88:9868-9872 (1991) and Dhallan et al., Nature 347:184-187 (1990)). In cases where activation of the receptor results in a decrease in cyclic nucleotide levels, it can be preferable to expose the cells to agents that increase intracellular cyclic nucleotide levels, e.g., forskolin, prior to adding a receptor-activating compound to the cells in the assay. Cells for this type of assay can be made by co-transfection of a host cell with DNA encoding a cyclic nucleotide-gated ion channel, GPCR phosphatase and DNA encoding a receptor (e.g., certain glutamate receptors, muscarinic acetylcholine receptors, dopamine receptors, serotonin receptors, and the like), which, when activated, causes a change in cyclic nucleotide levels in the cytoplasm.

[00154] In one embodiment, changes in intracellular cAMP or cGMP can be measured using immunoassays. The method described in Offermanns & Simon, *J. Biol. Chem.* 270:15175-15180 (1995) can be used to determine the level of cAMP. Also, the method described in Felley-Bosco et al., *Am. J. Resp. Cell and Mol. Biol.*, 11:159-164 (1994) can be used to determine the level of cGMP. Further, an assay kit for measuring cAMP and/or cGMP is described in U.S. Pat. No. 4,115,538, herein incorporated by reference.

[00155] In another embodiment, phosphatidyl inositol (PI) hydrolysis can be analyzed according to U.S. Pat. No. 5,436,128, herein incorporated by reference. Briefly, the assay involves labeling of cells with ^{3}H-myoinositol for 48 or more hrs. The labeled cells are treated with a test compound for one hour. The treated cells are lysed and extracted in chloroform-methanol-water after which the inositol phosphates are separated by ion exchange chromatography and quantified by scintillation counting. Fold stimulation is determined by calculating the ratio of cpm in the presence of agonist to cpm in the presence of buffer control. Likewise, fold inhibition is determined by calculating the ratio of cpm in the presence of antagonist to cpm in the presence of buffer control (which can or can not contain an agonist).

[00156] In another embodiment, transcription levels can be measured to assess the effects of a test compound on ligand-induced signal transduction. A host cell containing the protein of interest is contacted with a test compound in the presence of the natural ligand for a sufficient time to effect any interactions, and then the level of gene expression is measured. The amount of time to effect such interactions can be empirically determined, such as by running a time course and measuring the level of transcription as a function of time. The amount of transcription can be measured by using any method known to those of skill in the art to be suitable. For example, mRNA expression of the protein of interest can be detected using northern blots or their polypeptide products can be identified using immunoassays. Alternatively, transcription based assays using reporter genes can be used as described in U.S. Pat. No. 5,436,128, herein incorporated by reference. The reporter genes can be, e.g., chloramphenicol acetyltransferase, firefly luciferase, bacterial luciferase, beta-galactosidase and alkaline phosphatase. Furthermore, the protein of interest can be used as an indirect reporter via attachment to a second reporter such as green fluorescent protein (*see*, e.g., Mistili & Spector, Nature Biotechnology 15:961-964 (1997)).

[00157] The amount of transcription is then compared to the amount of transcription in either the same cell in the absence of the test compound, or it can be compared with the amount of transcription in a substantially identical cell that lacks the protein of interest. A substantially identical cell can be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA. Any difference in the amount of transcription indicates that the test compound has in some manner altered the activity of the protein of interest.

[00158] Samples that are treated-with a potential GPCR inhibitor or activator are compared to control samples comprising the natural ligand without the test compound to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative GPCR activity value of 100. Inhibition of a

GPCR is achieved when the GPCR activity value relative to the control is about 90%, optionally 50%, optionally 25-0%. Activation of a GPCR is achieved when the GPCR activity value relative to the control is 110%, optionally 150%, 200-500%, or 1000-2000%.

[00159] In one embodiment the invention provides soluble assays using molecules such as a domain, e.g., a ligand binding domain, an extracellular domain, a transmembrane domain (e.g., one comprising seven transmembrane regions and cytosolic loops), the transmembrane domain and a cytoplasmic domain, an active site, a subunit association region, etc.; a domain that is covalently linked to a heterologous protein to create a chimeric molecule; a GPCR; or a cell or tissue expressing a GPCR, either naturally occurring or recombinant. In another embodiment, the invention provides solid phase based in vitro assays in a high throughput format, where the domain, chimeric molecule, GPCR, or cell or tissue expressing a GPCR is attached to a solid phase substrate.

[00160] Certain screening methods involve screening for a compound that modulates the expression of the GPCRs described herein, or the levels of natural ligands, e.g., ASP and stanniocalcins. Such methods generally involve conducting cell-based assays in which test compounds are contacted with one or more cells expressing the GPCR or ligand and then detecting an increase or decrease in expression (either transcript or translation product). Such assays are typically performed with cells that express the endogenous GPCR or ligand.

[00161] Expression can be detected in a number of different ways. As described herein, the expression levels of the protein in a cell can be determined by probing the mRNA expressed in a cell with a probe that specifically hybridizes with a transcript (or complementary nucleic acid derived therefrom) of the GPCR or protein ligand. Probing can be conducted by lysing the cells and conducting Northern blots or without lysing the cells using in situ-hybridization techniques (see above). Alternatively, protein can be detected using immunological methods in which a cell lysate is probed with antibodies that specifically bind to the protein.

[00162] Other cell-based assays are reporter assays conducted with cells that do not express the protein. Certain of these assays are conducted with a heterologous nucleic acid construct that includes a promoter that is operably linked to a reporter gene that encodes a detectable product. A number of different reporter genes can be utilized. Some reporters are inherently detectable. An example of such a reporter is green fluorescent protein that emits fluorescence that can be detected with a fluorescence detector. Other reporters generate a detectable product. Often such reporters are enzymes. Exemplary enzyme reporters include, but are not limited to, beta-

glucuronidase, CAT (chloramphenicol acetyl transferase), luciferase, beta-galactosidase and alkaline phosphatase.

[00163] In these assays, cells harboring the reporter construct are contacted with a test compound. A test compound that either modulates the activity of the promoter by binding to it or triggers a cascade that produces a molecule that modulates the promoter causes expression of the detectable reporter. Certain other reporter assays are conducted with cells that harbor a heterologous construct that includes a transcriptional control element that activates expression of the GPCR or ligand and a reporter operably linked thereto. Here, too, an agent that binds to the transcriptional control element to activate expression of the reporter or that triggers the formation of an agent that binds to the transcriptional control element to activate reporter expression, can be identified by the generation of signal associated with reporter expression.

[00164] In one embodiment the invention provides soluble assays using molecules such as a domain, e.g., a ligand binding domain, an extracellular domain, a transmembrane domain (e.g., one comprising seven transmembrane regions and cytosolic loops), the transmembrane domain and a cytoplasmic domain, an active site, a subunit association region, etc.; a domain that is covalently linked to a heterologous protein to create a chimeric molecule; a GPCR; or a cell or tissue expressing a GPCR, either naturally occurring or recombinant. In another embodiment, the invention provides solid phase based in vitro assays in a high throughput format, where the domain, chimeric molecule, GPCR, or cell or tissue expressing a GPCR is attached to a solid phase substrate.

[00165] In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators or ligands in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator, or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (e.g., 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100-1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 different compounds are possible using the integrated systems of the invention.

[00166] The molecule of interest can be bound to the solid state component, directly or indirectly, via covalent or non covalent linkage e.g., via a tag. The tag can be any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest (e.g., the signal transduction molecule of interest) is attached to the solid support by interaction of the tag and the tag binder.

[00167] A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, etc.). Antibodies to molecules with natural binders such as biotin are also widely available and are appropriate tag binders; *see,* SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis Mo.).

[00168] Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody. In addition to antibody-antigen interactions, receptor-ligand interactions are also appropriate as tag and tag-binder pairs. For example, agonists and antagonists of cell membrane receptors (e.g., cell receptor-ligand interactions such as transferrin, c-kit, viral receptor ligands, cytokine receptors, chemokine receptors, interleukin receptors, immunoglobulin receptors and antibodies, the cadherin family, the integrin family, the selectin family, and the like; *see*, e.g., Pigott & Power, The Adhesion Molecule Facts Book I (1993). Similarly, toxins and venoms, viral epitopes, hormones (e.g., opiates, steroids, etc.), intracellular receptors (e.g. which mediate the effects of various small ligands, including steroids, thyroid hormone, retinoids and vitamin D; peptides), drugs, lectins, sugars, nucleic acids (both linear and cyclic polymer configurations), oligosaccharides, proteins, phospholipids and antibodies can all interact with various cell receptors.

[00169] Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

[00170] Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly-gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to persons of skill in the art. For example, poly(ethylene glycol) linkers are available from Shearwater Polymers, Inc. Huntsville, Ala. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

[00171] Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical

reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example, groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature. *See*, e.g., Merrifield, J. Am. Chem. Soc. 85:2149-2154 (1963) (describing solid phase synthesis of, e.g., peptides); Geysen et al., J. Immun. Meth. 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank & Doring, Tetrahedron 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor et al., Science, 251:767-777 (1991); Sheldon et al., Clinical Chemistry 39(4):718-719 (1993); and Kozal et al., Nature Medicine 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Non-chemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

Modulators

[00172] Inhibitors and/or activators identified according to the methods of the invention can be provided from libraries of compounds available from a number of sources or can be derived by combinatorial chemistry approaches known in the art. Such libraries include but are not limited to the available Chemical Director, Maybridge, and natural product collections. In one embodiment of the invention libraries of compounds with known or predicted structures can be docked to the human β2AR structures of the invention. In another embodiment, the libraries for ligands binding to the ligand binding site can include carazolol and related compounds. In another embodiment, the libraries can include a linker component or moiety. In some embodiments, the linker can include from about 10-22 atoms and can include one or more of C, O, N, S, and/or H atoms. In another embodiment, the libraries can include a ligand binding site (also known as the ligand, agonist, or antagonist binding pocket) component or moiety. In some embodiments, the libraries can include drug-like molecules, i.e., molecules having structural attributes of one or more compounds known to bind to and/or affect a physiologic function of a GPCR.

[00173] In some embodiments, the invention includes compounds that can be tested as modulators of GPCR activity. Compounds tested as modulators of GPCRs can be any small chemical compound or biological entity. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds can be dissolved in aqueous or organic (especially DMSO-based) solutions. The assays are designed to screen

large chemical libraries by automating the assay steps. The assays are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, Mo.), Aldrich (St. Louis, Mo.), Sigma-Aldrich (St. Louis, Mo.), Fluka Chemika-Biochemica Analytika (Buchs Switzerland) and the like.

[00174] In one preferred embodiment, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator or ligand compounds). Such "combinatorial chemical libraries" or ligand libraries are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

[00175] A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[00176] Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Pat. No. 5,010,175, Furka, Int. J. Pept. Prot. Res. 37:487-493 (1991) and Houghton et al., Nature 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (e.g., PCT Publication No. WO 91/19735), encoded peptides (e.g., PCT Publication WO 93/20242), random bio-oligomers (e.g., PCT Publication No. WO 92/00091), benzodiazepines (e.g., U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs et al., Proc. Nat. Acad. Sci. USA 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara et al., J. Amer. Chem. Soc. 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann et al., J. Amer. Chem. Soc. 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen et al., J. Amer. Chem. Soc. 116:2661 (1994)), oligocarbamates (Cho et al., Science 261:1303 (1993)), and/or peptidyl phosphonates (Campbell et al., J. Org. Chem. 59:658 (1994)), nucleic acid libraries (see Ausubel, Berger and Russell & Sambrook, all supra), peptide nucleic acid libraries (see, e.g., U.S. Pat. No. 5,539,083), antibody libraries (see, e.g., Vaughn et al., Nature Biotechnology, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries

(see, e.g., Liang et al., Science, 274:1520-1522 (1996) and U.S. Pat. No. 5,593,853), small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN, January 18, page 33 (1993); isoprenoids, U.S. Pat. No. 5,569,588; thiazolidinones and metathiazanones, U.S. Pat. No. 5,549,974; pyrrolidines, U.S. Pat. Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Pat. Nos. 5,506,337; benzodiazepines, 5,288,514, and the like).

[00177] Devices for the preparation of combinatorial libraries are commercially available (*see*, e.g., 357 MPS, 390 MPS, Advanced Chem Tech, Louisville Ky., Symphony, Rainin, Woburn, Mass., 433A Applied Biosystems, Foster City, Calif., 9050 Plus, Millipore, Bedford, Mass.). In addition, numerous combinatorial libraries are themselves commercially available (see, e.g., ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, Mo., 3D Pharmaceuticals, Exton, Pa., Martek Biosciences, Columbia, Md., etc.).

[00178] It is noted that modulators that compete with the binding and/or activity of the known ligands for to human β2AR can be used to treat various diseases including, but not limited to, coronary artery disease, atherosclerosis, thrombosis, obesity, diabetes, stroke, and other diseases.

[00179] In one embodiment, a modulator binds to a site on a GPCR, e.g., a human β2AR. In one aspect, the site is a carazolol binding site. In a related aspect, the site is a ligand binding site. In another aspect, the modulator has a first moiety that binds to a binding site. In another aspect, the first moiety is connected to a linker. In another aspect, the first moiety and the linker are connected to at least one additional moiety that binds to a site other than that bound by the first moiety. In another aspect, the two or more moieties are not connected by a linker and are both present in a composition.

Computer-based modeling of β2AR

[00180] Protein-ligand docking aims to employ principles by which protein receptors, e.g., human β2AR, recognize, interact, and associate with molecular substrates and compounds to predict the structure arising from the association between a given compound and a target protein of known three-dimensional structure.

[00181] In protein-ligand docking, the search algorithm can allow the degrees of freedom of the protein-ligand system to be sampled sufficiently as to include the true binding modes. Three general categories of algorithms have been developed to address this problem of ligand flexibility: systematic methods; random or stochastic methods; and simulation methods.

[00182] Systematic search algorithms attempt to explore all degrees of freedom in a molecule. These algorithms can be further divided into three types: conformational search methods, fragmentation methods, and database methods.

[00183] In conformational search methods, all rotatable bonds in the ligand are systematically rotated through 360° using a fixed increment, until all possible combinations have been generated and evaluated. As the number of structures generated increases immensely with the number of rotatable bonds (combinatorial explosion), the application of this type of method, in its purest form, is very limited.

[00184] Fragmentation methods use two different approaches to incrementally grow the ligands into the active site. One approach is by docking the several fragments into a site and linking them covalently to recreate the initial ligand ("the place-and-join approach"). Another approach is by dividing the ligand into a rigid core-fragment that is docked in first place and flexible regions that are subsequently and successively added ("the incremental approach"). DOCK (*see* above) is an example of s docking programs that use a fragmentation search method.

[00185] Database methods using libraries of pre-generated conformations or conformational ensembles to address the combinatorial explosion problem. A example of a docking program using database methods is FLOG which generates a small set of 25 database conformations per molecule based on distance geometry, that are subsequently subject to a rigid docking protocol.

[00186] Random search algorithms sample the conformational space by performing random changes to a single ligand or a population of ligands. At each step, the alteration performed is accepted or rejected based on a predefined probability function. There are three basic types of methods based on random algorithms: Monte Carlo methods (MC), Genetic Algorithm methods (GA), and Tabu Search methods.

[00187] Simulation methods employ a rather different approach to the docking problem, based on the calculation of the solutions to Newton's equations of motion. Two major types exist: molecular dynamics (MD) and pure energy minimization methods.

[00188] Scoring functions normally employed in protein-ligand docking are generally able to predict binding free energies within 7-10 kJ/mol and can be divided into three major classes: force field-based, empirical, and knowledge-based scoring functions.

[00189] In force-field based scoring, standard force fields quantify the sum of two energies: the interaction energy between the receptor and the ligand, and the internal energy of the ligand. The energies are normally

accounted through a combination of a van der Waals with an electrostatic energy terms. A Lennard-Jones potential is used to describe the van der Waals energy term, whereas the electrostatic term is given by a Coulombic formulation with a distance-dependent dielectric function that lessens the contribution from charge-charge interactions.

[00190] Empirical scoring functions are based on the idea that binding energies can be approximated by a sum of several individual uncorrelated terms. Experimentally determined binding energies and sometimes a training set of experimentally resolved receptor-ligand complexes are used to determine the coefficients for the various terms by means of a regression analysis.

[00191] Knowledge-based scoring functions focus on following the rules and general principles statistically derived that aim to reproduce experimentally determined structures, instead of binding energies, trying to implicitly capture binding effects that are difficult to model explicitly. Typically, these methods use very simple atomic interactions-pair potentials, allowing large compound databases to be efficiently screened. These potentials are based on the frequency of occurrence of different atom-atom pair contacts and other typical interactions in large datasets of protein-ligand complexes of known structure. Therefore, their derivation is dependent on the information available in limited sets of structures.

[00192] Consensus Scoring combines the information obtained from different scores to compensate for errors from individual scoring functions, therefore improving the probability of finding the correct solution. Several studies have demonstrated the success of consensus scoring methods in relation to the use of individual functions schemes.

[00193] Using the Protein-ligand docking methods described above, a predicted association can be made between a selected chemical library compound (*see* above for examples) and the binding sites in the human β2AR structure described in **Appendix I** (SEQ ID NOS 4-5, 1 and 6-9, respectively in order of appearance). These methods will therefore allow the generation of a binding profile for any known compound in any of the binding sites or cavities of the human β2AR based on the simulated docking of the compound.

[00194] In another embodiment, a form of computer-assisted drug design is employed in which a computer system is used to generate a three-dimensional structure of the candidate class A GPCR based on the structural information encoded by the amino acid sequence. This will allow use of the methods described above to identify candidate compounds based on their ability to dock in one or more of the predicted GPCR structure binding sites. In one aspect, the input amino acid sequence of the GPCR interacts directly and actively with a

pre-established algorithm in a computer program to yield secondary, tertiary, and quaternary structural models of the class A GPCR. The models of the class A GPCR structure are then examined to identify the position and structure of the binding sites, e.g., a binding pocket. The position and structure of the predicted binding site(s) is then used to identify various compounds that modulate ligand-receptor binding using the methods described above.

[00195] The three-dimensional structural model of the GPCR is generated by entering protein amino acid sequences of at least 10 amino acid residues or corresponding nucleic acid sequences encoding a GPCR polypeptide into the computer system. The amino acid sequence represents the primary sequence or subsequence of the protein, which encodes the structural information of the protein. At least 10 residues of the amino acid sequence (or a nucleotide sequence encoding 10 amino acids) are entered into the computer system from computer keyboards, computer readable substrates that include, but are not limited to, electronic storage media (e.g., magnetic diskettes, tapes, cartridges, and chips), optical media (e.g., CD ROM), information distributed by internet sites, and by RAM. The three-dimensional structural model of the GPCR is then generated by the interaction of the amino acid sequence and the computer system, using software known to those of skill in the art. Any method of protein structure modeling such as *ab-initio* modeling, threading or sequence-sequence based methods of fold recognition. In one embodiment, the AS2TS system of protein structure modeling is used. In other embodiments, a sequence alignment in combination with a threshold protein sequence similarity to determine a set of protein sequences for which to model protein structure is used. In one aspect, sequence alignments are generated for the set of sequences to be modeled with sequences of proteins with solved empirical structure in a protein structure databank known to one of skill in the art. If the sequences to be modeled have a sufficient similarity to one or more sequences with known protein structure, then the three dimensional structure of the sequence can be modeled.

[00196] The amino acid sequence represents a primary structure that encodes the information necessary to form the secondary, tertiary and quaternary structure of the GPCR of interest. In one embodiment, software can look at certain parameters encoded by the primary sequence to generate the structural model. These parameters are referred to as "energy terms," and primarily include electrostatic potentials, hydrophobic potentials, solvent accessible surfaces, and hydrogen bonding. Secondary energy terms include van der Waals potentials. Biological molecules form the structures that minimize the energy terms in a cumulative fashion. The computer program is therefore using these terms encoded by the primary structure or amino acid sequence to create the secondary structural model.

[00197] The tertiary structure of the protein encoded by the secondary structure is then formed on the basis of the energy terms of the secondary structure. The user at this point can enter additional variables such as whether the protein is membrane bound or soluble, its location in the body, and its cellular location, e.g., cytoplasmic, surface, or nuclear. These variables along with the energy terms of the secondary structure are used to form the model of the tertiary structure. In modeling the tertiary structure, the computer program matches hydrophobic faces of secondary structure with like, and hydrophilic faces of secondary structure with like.

[00198] In another embodiment, protein structure alignments can be used to determine the structure of GPCRs using the known structure of the β2AR (**Appendix I**) (SEQ ID NOS 4-5, 1 and 6-9, respectively in order of appearance). Protein structure alignments preferably are sets of correspondences between spatial co-ordinates of sets of carbon alpha atoms which form the 'backbone' of the three-dimensional structure of polypeptides, although alignments of other backbone or side chain atoms also can be envisioned. These correspondences are generated by computationally aligning or superimposing two sets of atoms order to minimize distance between the two sets of carbon alpha atoms. The root mean square deviation (RMSD) of all the corresponding carbon alpha atoms in the backbone is commonly used as a quantitative measure of the quality of alignment. Another quantitative measure of alignment is the number of equivalent or structurally aligned residues.

[00199] In another embodiment, a GPCR structure is calculated based on the solved structure of the human β2AR by computationally aligning or superimposing two sets of atoms to minimize distance between the two sets of carbon alpha atoms (i.e., the alpha carbon atoms of the human β2AR and an unknown GPCR structure), followed by one or more of simulated annealing and energy minimization. The result of this calculation is a computed structure for a GPCR that provides atomic co-ordinates for the alpha carbon backbone as well as side chain atoms.

[00200] A variety of methods for generating an optimal set of correspondences can be used in the present invention. Some methods use the calculation of distance matrices to generate an optimal alignment. Other methods maximize the number of equivalent residues while RMSD is kept close to a constant value.

[00201] In the calculation of correspondences, various cutoff values can be specified to increase or decrease the stringency of the alignment. These cutoffs can be specified using distance in Angstroms. Depending on the level of stringency employed in the present invention, the distance cutoff used is less than 10 Angstroms or less than 5 Angstroms, or less than 4 Angstroms, or less than 3 Angstroms. One of ordinary skill will recognize that the utility of stringency criterion depends on the resolution of the structure determination.

[00202] In another embodiment of the present invention, the set of residue-residue correspondences are created using a local-global alignment (LGA), as described in US Patent Publication Number 2004/0185486. In this method, a set of local superpositions are created in order to detect regions which are most similar. The LGA scoring function has two components, LCS (longest continuous segments) and GDT (global distance test), established for the detection of regions of local and global structure similarities between proteins. In comparing two protein structures, the LCS procedure is able to localize and superimpose the longest segments of residues that can fit under a selected RMSD cutoff. The GDT algorithm is designed to complement evaluations made with LCS searching for the largest (not necessary continuous) set of 'equivalent' residues that deviate by no more than a specified distance cutoff.

[00203] Using the protein structure alignments described above, the structure of human β2AR in **Appendix I** (SEQ ID NOS 4-5, 1 and 6-9, respectively in order of appearance) can be used as a model on which to discern the structure of other GPCRs and/or their predicted ligand-binding sites.

[00204] Once the GPCR structure has been generated, a binding pocket can be identified by the computer system. Computational models seek to identify the regions by characterization of the three dimensional structure of the GPCR. Some methods of identifying a binding pocket use triangulation such as weighted Delaunay triangulation to determine pocket volumes (castP). Other methods use spheres to determining protein pocket volumes (Q-site-finder, UniquePocket). Conserved binding-site identification seeks to identify conserved regions such as a binding pocket through associating the residues which form the aforementioned regions with conserved residues in homologous protein sequences or structures, e.g., through the use of sequence alignments.

[00205] One method of identifying a binding pocket in a GPCR entails filling the three dimensional protein structures with spheres, creating a "negative image" of the structure. A cutoff distance, such as 8 Angstroms, is used to determine spheres which interact with residues. Spheres are labeled as conserved or not-conserved based on their interaction with residues which form a conserved binding site. The conserved spheres are clustered based on their three dimensional co-ordinates to identify a set of spheres with interact with conserved residues and are proximal in three dimensional space forming a cluster. Three-dimensional structures for potential compounds are generated by entering chemical formulas of compounds. The three-dimensional structure of the potential compound is then compared to that of the GPCR protein ligand-binding site(s) (e.g., a binding pocket) to identify compounds that bind to the GPCR binding site(s). Binding affinity between the

GPCR binding site(s) and the compound is determined using energy terms to determine which ligands have an enhanced probability of binding to the protein.

[00206] It should be noted that the language used in the specification has been principally selected for readability and instructional purposes, and can not have been selected to delineate or circumscribe the inventive subject matter. Accordingly, the disclosure of the present invention is intended to be illustrative, but not limiting, of the scope of the invention, which is set forth in the claims.

[00207] Reference will now be made in detail to particularly preferred embodiments of the invention. Examples of the preferred embodiments are illustrated in the following Examples section.

EXAMPLES

[00208] Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

[00209] The practice of the present invention will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols. A and B (1992).

Example 1: Crystallization of a β2AR protein using a lipidic cubic phase/sterol method

[00210] This Example describes the generation of diffraction-quality crystals of a β2AR protein, specifically crystals of β2AR-T4L, a fusion protein of human β2AR with T4 lysozyme, bound to carazolol. A detailed description of the protein and its synthesis is provided in Example 4. Briefly, T4 lysozyme was generated by three distinct modifications to β2AR: (1) a fusion protein was created by replacement of the third intracellular loop with T4L, (2) the carboxyl terminal 48 amino acids were deleted, and (3) a glycosylation site

at Asn187 was eliminated through a glutamate substitution. β2AR-T4L was expressed in Sf9 insect cells, solubilized in 1 % dodecylmaltoside, and purified by sequential antibody and ligand affinity chromatography.

LCP crystallization setup

[00211] Protein solution (30 mg/ml) was mixed with a host lipid or lipid mixture typically in 2:3 ratio by volume ratio using a syringe mixer (Cheng, et al., *Chem Phys Lipids* 95, 11 (1998). Upon mixing (~100 passages, 2-3 min) the sample spontaneously formed homogeneous transparent cubic phase, which was directly used in crystallization trials. Robotic trials were performed using an *in meso* crystallization robot (Cherezov, et al., *Acta Cryst D* 60, 1795 (2004)). Six-well glass sandwich plates (Cherezov, et al., *Acta Cryst D* 60, 1795 (2004); Cherezov, *J Appl Cryst* 36, 1372 (2003)), were filled with 25 or 50 nL protein-laden lipidic cubic phase drops overlaid by 800 μL of precipitant solution in each well and sealed with a glass coverslip. Manual setups were performed in Impact microbatch plates (Hampton Research cat# HR3-293), Innovaplate SD-2 sitting drop plates (Hampton Research cat# HR3-083) or VDX48 hanging drop plates (Hampton Research cat# HR3-275). Modified repetitive syringe dispenser (Cherezov, et al., *J Appl Cryst* 38, 398 (2005)), coupled with a gas-tight 10 μL syringe was used to deliver 70 nL of cubic phase drops per well to which 1-2 μL of precipitant solution was added with a microvolume pipette. Reservoirs of the Innovaplate and VDX48 plates were filed with 50 and 100 μL of precipitant respectively. All operations starting from mixing lipid and protein were performed at room temperature (~21-23° C). After setup, plates were transferred into an automatic incubator/imager (RockImager 1000, Formulatrix Inc.) maintained at 20° C. Plates were imaged every 12 hours for the first 3 days, then every day until 7 days and after that on the 10th and on the 14th day.

Initial hits

[00212] Initial trials were performed using protein solution at a concentration of 30 mg/mL mixed with monoolein as a host lipid against a set of 6 commercial screens (Index HT, SaltRx HT and MemFac HT from Hampton Research, JCSG+ and MbClass from Nextal, and MemSys&MemStart from Molecular Dimensions) set up in duplicates. Initial hits detected in three different wells contained extremely small, <5 μm, needle-like birefringent crystal showers. The detection of such small colorless protein crystals in LCP was made feasible by specially developed glass sandwich plates with optimized optical properties (Cherezov, et al., *Acta Cryst D* 60, 1795 (2004); Cherezov et al., *J Appl Cryst* 36, 1372 (2003)). Hit conditions were similar by chemical composition containing 30 %v/v PEG 400 as a precipitant, low concentration of Li sulfate and a buffer at pH 6.5 or 7.5.

Optimization

[00213] Crystal condition optimization is well known to practioners in the art of x-ray crystallography. What follows is a specific example of a generally applicable optimization approach in which one or more of the crystallization mixture components is systematically varied in concentration or substituted by a chemical analog. Initial rounds of optimization were focused on varying concentration of the main precipitant, PEG 400, buffer pH and identity, and salt identity and concentration. As a result, Li sulfate was replaced with Na sulfate and useful concentration and pH ranges were established. Crystals were still rather small reaching ~15x5x1 μm in size.

[00214] Further, lipid and soluble additives were searched for and optimized simultaneously. Three different host lipids (monopalmitolein, monovaccenin and monoolein), five lipid additives to monoolein host (DOPE, DOPE-Me, DOPC, Asolectin and cholesterol) and 96-well soluble additives screen were tried in different combinations along with the previously found basic crystallization conditions. One of the soluble additives, 1,4-butanediol, stood out, but only when it was used in combination with lipid additives, DOPE or cholesterol. When DOPE was used as an additive the crystals grew as thin plates (40x7x2 μm), while when cholesterol was used the crystals grew as small rods (30x5x5 μm). DOPE was dropped out in favor of cholesterol in the subsequent optimization steps.

[00215] Final optimization required fine tuning concentrations of all components (protein, PEG 400, Na sulfate, 1,4-butanediol, cholesterol, buffer pH). At the final stages of optimization, higher concentrations of protein, e.g., 50 mg/mL protein solution, were preferred. Decreasing the volume of lipidic cubic phase per trial from 50 to 20 nL consistently produced larger crystals. The best crystals (40 x 20 x 5 μm; **Figure 1**) were obtained in 30-35 %v/v PEG400, 0.1-0.2 M Na sulfate, 0.1 M Bis-tris propane pH 6.5 -7.0, 5-7 %v/v 1,4-butanediol using 8-10 %w/w cholesterol in monoolein as the host lipid. Thus, in another aspect, the invention provides a method of directly adding a lipid additive (e.g., cholesterol, DOPE) to the host lipid prior to combining with the protein mixture, thereby significantly improving the size and quality of LCP grown crystals. Previously, exogenous lipid had been added only to the protein solution prior to combining the protein solution with the host lipid (Luecke, et al., *Science* 293, 1499 (2001)).

[00216] To limit the range of precipitant and additives concentrations used for screening, previously published phased diagrams were used, e.g., for monoolein (Qiu, et al., *Biomaterials* 21, 223 (2000)), and monovaccenin (Qiu, et al., *J. Phys. Chem. B* 102, 4819 (1998)), and the effects on monoolein phase behavior of soluble (Cherezov, et al., *Biophys J* 81, 225 (2001)), and lipid additives (Cherezov, et al., Biophys J 83, 3393

(2002)). Certain lipid mixtures required pre-screening their phase behavior by microscope visualization and by X-ray, prior to being mixed with the protein for crystallization trials. Thirty-three 96-well optimization screens were used in combinations with different lipid mixtures, resulting in over 15,000 trials. This throughput was made feasible through the use of an *in meso* crystallization robot and automatic imager.

[00217] Typically, the best crystals grown under lipidic cubic phase conditions appear at the boundary between the cubic and the sponge phase (Cherezov, et al., *J Mol Biol* 357, 1605 (2006); Wadsten, et al., *J Mol Biol* 364, 44 (2006)). When crystals are close to the phase boundary, equally good crystals are obtained in either phase.

Crystal harvesting

[00218] Crystals were harvested directly from the glass sandwich plates (**Figure 2**) because this method provided results superior to those obtained with microbatch or vapor diffusion plates. These plates have been specifically designed to perform optimally only at the screening and optimization stages (Cherezov, et al., *Acta Cryst D* 60, 1795 (2004); Cherezov and M. Caffrey, *J Appl Cryst* 36, 1372 (2003)). No harvesting has been previously attempted from them, due to the difficulties in separating glass slides strongly bound by a high performance double sticky tape. Thus, in another aspect, the invention disclosed herein provides a special technique for opening individual wells and harvesting crystals from them. A corner of a capillary cutting stone (in this instance, from Hampton Research) was used to scratch the top glass around the perimeter of the well. Gently pressing the glass slide just outside the scratch allows propagation of the scratch through the depth of the glass. The glass slide was then broken in one point just outside of the scratched perimeter using a sharp needle. This hole was used to lift up the glass slide and expose the cubic phase for harvesting. An extra drop of ~5 μL of precipitant was added to the well to reduce dehydration. Using this technique, it was possible to open up and harvest crystals successfully from more than 80% of attempted wells.

[00219] Crystals were scooped directly from the lipidic cubic phase using 30 μm aperture MiTeGen MicroMounts and plunged into liquid nitrogen. Care was taken to drag as little as possible lipid around the crystal to decrease the unwanted background. Attempts to dissolve the lipids, either by increasing concentration of PEG400 or using a mineral oil, typically resulted in decreasing the diffraction power of the crystals.

Data Collection

[00220] During screening for diffraction of these crystals at APS beamline GM/CA CAT (**Figure 3**), the crystals themselves could not be observed in the mounted loops (**Figure 1a**). Therefore, a systematic screening of the loop material with varying beams was conducted to identify the crystal in the loop. Optimization of the

diffraction with a low x-ray dose was then used to center the crystals and eventually allow for complete data collection using the 10 μm x 10 μm minibeam setup at GM/CA CAT. The complete data set is then compared to data filtered by a sigma cutoff (see **Table 1**). All of the data was used in structure solution and refinement.

Table 1

Signal/noise -3

Resolution	# Observed refl	# Unique refl	Redundancy	Completeness	R_{sym}	I/SIGMA	R-meas	Rmrgd-F
10	3352	333	10.1	87.40%	6.60%	23.25	6.90%	2.70%
8	3591	354	10.1	99.40%	7.60%	22.6	8.00%	2.80%
6	10480	1003	10.4	99.60%	9.90%	20.01	10.50%	4.00%
3	126008	1'968	10.5	99.80%	13.30%	14.01	14.00%	6.60%
2.8	33'58	3130	10.6	100.10%	38.00%	6.42	39.90%	18.70%
2.7	19702	1893	10.4	99.70%	49.50%	4.96	52.10%	24.90%
2.6	23772	2275	10.4	99.90%	60.20%	4.07	63.40%	30.10%
2.5	14'08	2558	5.5	99.30%	58.90%	2.69	65.10%	51.80%
2.4	14672	3060	4.8	99.10%	67.30%	2.18	75.70%	62.80%
total	248843	26574	9.4	99.50%	12.70%	9.62	13.40%	11.40%

Signal/noise 0

Resolution	# Observed refl	# Unique refl	Redundancy	Completeness	R_{sym}	I/SIGMA	R-meas	Rmrgd-F
10	3352	333	10.1	87.40%	6.60%	23.25	6.90%	2.70%
8	3591	354	10.1	99.40%	7.60%	22.6	8.00%	2.80%
6	10480	1003	10.4	99.60%	9.90%	20.01	10.50%	4.00%
3	125628	1'923	10.5	99.40%	13.30%	14.06	14.00%	6.50%
2.8	32679	3077	10.6	98.40%	37.30%	6.54	39.50%	17.70%
2.7	19346	1849	10.5	97.40%	48.30%	5.08	51.30%	23.40%
2.6	23201	2210	10.5	97.00%	58.00%	4.2	62.00%	28.00%
2.5	13461	2406	5.6	93.40%	56.40%	2.88	62.30%	45.90%
2.4	13833	2827	4.9	91.50%	64.10%	2.4	71.50%	54.30%
total	245571	25982	9.5	97.30%	12.30%	9.85	13.30%	10.80%

Example 2. Using the the LCP/sterol method to generate additional membrane protein crysals

[00221] In addition to the β_2AR-T4L/carazolol structure (Examples 1, 3, and 4), the LCP/sterol matrix has successfully been used to crystallize a diversity of receptor-ligand systems.

[00222] **1. β_2AR-T4L $^{(E122W)}$** A thermally-stabilized construct of β_2AR-T4L comprising an E122W mutation has been crystallized in the presence of both agonist and antagonist ligands including: alprenolol, timolol, clenbuterol and carazolol. For lipidic cubic phase (LCP) crystallization of β_2AR-T4L $^{(E122W)}$, robotic trials were performed using an *in meso* crystallization robot (Cherezov et al., 2004). Glass sandwich plates in 96-well format (Cherezov and Caffrey, 2003; Cherezov et al., 2004) were filled with 25 or 50 nL protein-laden LCP drops overlaid by 0.8 μL of precipitant solution in each well and sealed with a glass coverslip. All operations starting from mixing lipid and protein were performed at room temperature (~21-23 °C). Crystals were obtained in 28% (v/v) PEG 400, 0.3 M potassium formate, 0.1 M Bis-tris propane pH 7.0 and saturating concentrations of ligand (e.g., 2 mM in the case of timolol) using 10% (w/w) cholesterol in monoolein as the

host lipid. Diffraction data were collected on all four ligand complexes (see **Figure 4**), and structures were determined for alprenolol (3.2 Å), timolol (2.8 Å), and carazolol (2.8 Å).

[00223] **2. A_{2A}R-T4L** The applicability of the monoolein cholesterol system in the crystallization of non-biogenic amine receptors has also been demonstrated with the structural determination of the human A_{2A} adenosine receptor (A_{2A}R-T4L) bound to a high affinity selective antagonist, ZM241385, to 2.6 Å resolution. See **Figure 4**. For lipidic cubic phase (LCP) crystallization of the human A2A adenosine receptor *in meso*, glass sandwich plates (Cherezov, *et al., Acta Crystallogr D Biol Crystallogr,* 60, 1795 (2004)) were filled with 50 nl receptor-cholesterol-monoolein LCP drops overlaid by 0.8 µl of precipitant solution in each well and sealed with a glass coverslip. Lipid:receptor LCP mixture typically contained monoolein:cholesterol (54%:6% (w/w)) and receptor (40% (w/w)). Crystallization set-ups were performed at ambient temperature (22±2 °C). Plates were incubated and imaged at 20 °C using an automated incubator/imager (RockImager 1000, Formulatrix). Data-collection quality crystals (~ 100 µm x 10 µm x 5 µm) were obtained in 30% (v/v) PEG 400 (range of 28-32%), 186 mM Lithium sulfate (range of 180 to 220 mM), 100 mM Sodium citrate (pH 6.5) (Range of 5.5 to 6.5) and 200 µM ZM241385. The protein crystallized in the primitive monoclinic space group $P2_1$ with one molecule per asymmetric unit and an estimated solvent content of 52%.

[00224] **3. $\beta_2AR^{(E122W)}$** Initial crystals of $\beta_2AR^{(E122W)}$, lacking inserted T4 lysozyme, have also been obtained. The protein was extracted from insect cell membranes using a mixture of 0.5% w/v dodecyl maltoside (DDM), 0.1% w/v cholesteryl hemisuccinate (CHS) and 1mM timolol. Timolol was maintained at 1mM throughout the first steps of the purification. The extracted protein was purified by binding overnight to Talon™ immobilized metal affinity resin followed by a standard washing and elution with 200mM imidazole. Adenosine triphosphate at 5 mM in combination with 10mM $MgCl_2$ was used to eliminate chaperone protein contamination. Eluted protein was concentrated to 2.5mL and desalted into a 0mM imidazole buffer using a PD10 desalting column (GE-Biosciences). Protein was then bound to 100µL of Ni-sepharose immobilized metal affinity resin in the presence of PNGase (New England Biolabs) to remove glycosylation, and incubated overnight. After incubation the protein was washed on the column and timolol was replaced by carazolol for structure solution. The protein bound to carazolol was eluted from the Ni-Sepharose column, treated with 100mM Nacitrate pH 7.5, and concentrated to 50mg/mL.

[00225] The protein solution was then mixed with monoolein containing 10% cholesterol at a ration of 40:60 % w/w protein to lipid to generate the lipidic cubic phase used in crystallization trials. The LCP lipid containing protein was dispensed onto glass sandwich crystallization plates at a volume of 20 nL to which 1 µL

of crystallization solution was added. The entire experiment in 96 well format was then covered by an additional glass plate which was fastened to the first by virtue of double back sticky tape. Initial crystals have been obtained after 24 hours by addition of a solution containing 35% v/v PEG 400, 100mM NaSO4 100mM Bis tris propane pH 7 and 8% 2,6 hexanediol.

[00226] By way of a prophetic example, optimized crystals of $\beta_2AR^{(E122W)}$ obtained by this method are screened for their ability to diffract at high resolution, e.g., less than 3.5 Å or, more preferably, less than 3 Å. Guidance for optimization is provided by the optimization protocols set forth herein and in the examples. In combination with the teaching provided herein, one skilled in the art will readily identify appropriate beam settings to obtain diffraction patterns from which a detailed molecular structure of the optimally crystallized protein can be solved.

[00227] **4. CXCR4-T4L** CXCR4, also called fusin, is a GPCR protein specific for stromal-derived-factor-1 (SDF-1 also called CXCL12), a molecule endowed with potent chemotactic activity for lymphocytes. This Example teaches prophetically how the methods of the invention may be used to generate diffraction quality crystals of a fusion protein comprising CXCR4 (CXCR4-T4L).

[00228] The cDNA encoding CXCR4 is synthesized by outsourcing to DNA2.0 where the DNA was optimized for human codon usage, elimination of transcribed RNA secondary structure, elimination of ribosome binding sites and avoidance of common restrictions sites used in subsequent cloning procedures. Two initial variants are contracted to be synthesized, the first encoding a wild-type full-length receptor and the second a full length receptor with a fusion protein located between transmembrane helix V and helix VI, effectively eliminating the third intracellular loop (3IL) region of the receptor. In the case of CXCR4, T4-lysozyme (T4L) is the fusion protein fused in the place of the 3IL. A set of guidelines is followed for the incorporation of T4L into the fusion protein which minimizes the possibility of structural disruption and concomitant effects on protein expression and stability. Briefly, the 5' insertion point for the fusion protein takes precedence over the 3' insertion point and is located 66 nucleotides (22 residues) downstream of the codon for a conserved proline on helix V of the receptor. If the 3IL section of the receptor is large, the 3' fusion point is set 87 nucleotides (29 residues) upstream from the codon for the family conserved proline on helix VI. However, as is the case for CXCR4, where the 3IL loop is small, cDNA for the fusion protein is inserted directly into the 3IL loop position dictated by spacing from helix V with no resulting excision of intervening nucleotides. Specifically, T4L is inserted into CXCR4 based on spacing between a conserved proline on helix V and a C-terminal truncation was generated based on literature precedence. Each synthesized cDNA is flanked by an out of frame AscI

(GGCGCGCCG) restriction site on the 5' end and an in frame FseI (GGCCGGCC) on the 3' end for sub – cloning into a set of four expression vectors. Viral DNA is then generated, amplified from these vectors according to standard protocols and titered using flow cytometry to measure the population of cells expressing the virally encoded GP64 protein.

Protein expression of CXCR4-T4L

[00229] With titered virus in hand, small scale expression trials are carried out in a volume of 5 mL / experiment. Expression levels are assessed using flow cytometry to measure the mean fluorescence intensity (MFI) and percentage of cells expressing the FLAG epitope encoded by the expression screening vectors. Expressing cells are tested with and without permeabilization to generate a ratio between protein inserted in the plasma membrane and protein inside the cellular trafficking machinery. Acorrelation between cell surface expression and overall protein expression is demonstrated, as well as a correlation between stability and the ratio of cell surface expression/total expression. In addition to these assays, small scale purification after solubilization with dodecyl maltoside (DDM) is carried out to determine the quantity of recoverable protein as well as the quality as measured by size exclusion chromatography. Based on these data it was apparent that the T4L fused receptor is expressing and that it is dimerizing in a ligand independent manner, an indicator of C-terminal non-specific interactions in other receptors. Thus, a C-terminal truncation mutant of CXCR4 is generated.

Protein purification

[00230] The C-terminal truncation of T4L fused CXCR4 was scaled up to production sized expression (5-10L of cell culture) and further processed by large scale purification efforts intended for crystallization trials. Briefly 5-10 L of cells culture are centrifuged and washed with PBS followed by freezing at -80°C. The frozen cellular material is then resuspended in 820 mL of lysis buffer (10mM Hepes pH 7.5, 10mM $MgCl_2$, 20mM KCl) supplemented with protease inhibitor (Roche). The cell suspension is lysed by 20 strokes of a dounce homogenizer and centrifuged at 45,000 rpm in Ti45 ultracentrifuge for 30 minutes. The resulted pellet was separated from the supernatant, resuspended and the process repeated six times to ensure complete removal of soluble protein material. On the final resuspension step the membranes were resuspended in 100mL of lysis buffer containing 40% v/v glycerol, homogenized with 20 strokes of a dounce homogenizer and frozen in 10mL aliquots at -80 °C for indefinite storage.

[00231] For solubilization and purification, each 10mL aliquot of frozen membranes is resuspended to 25 mL using lysis buffer to which 100uM AMD070 ligand is added in addition to protease inhibitor at 2x working concentration and 2 mg/mL iodoacetamide. The membranes are allowed to thaw and incubate with ligand at an appropriate temperature for at least 30 minutes. After the incubation the mixture is diluted two-fold with a 2x solubilization buffer containing 100mM Hepes pH 7.5, 1M NaCl, 2% w/v DDM 0.2% w/v CHS. The solubilization is allowed to proceed with agitation for at least 2 hours at 4°C after which insoluble material is separated by centrifugation and discarded. The supernatant is isolated and allowed to bind to 0.5 mL of Talon (Clontech) IMAC resin charged with Co^{2+} in the presence of 20mM imidizole buffered to 7.5 and 800mM NaCl. Binding to the Talon IMAC resin is allowed to proceed with agitation at least 4 hours but most commonly overnight. After binding, the slurry is poured into a gravity column and the resin is separated from the supernatant. The resin is then washed with 80 column volumes (CV) of wash buffer (50mM Hepes pH 7.5, 800mM NaCl, 20mM Imidizole, 0.1% w/v DDM, 0.01% w/v CHS and 100 uM AMD070 (or receptor appropriate ligand). After the initial wash the resin is further treated to adjust the NaCl concentration to 500mM and to increase the ligand concentration to 300uM. The protein is then eluted from the resin using 200mM Imidazole and concentrated to 2.5 mL for removal of the excess imidizole with a PD10 desalting column (GE Biosciences). The ligand concentration is increased to 500uM and the protein is bound to 100uL of Ni-Sepharose IMAC resin in the presence of 20,000 units of PNGase (NEB) an endoglycosidase capable of removing N-linked glycosylation. The protein is allowed to bind to the resin and deglycosylate for 6 hours after which the resin is washed with imidizole free elution buffer (50mM Hepes pH 7.5, 500mM NaCl, 0.05%w/v DDM, 0.01% w/v CHS and 1mM AMD 070). After the washing step the protein was eluted from the resin using the same buffer but including 200mM imidizole. After elution the protein is normally concentrated to approximately 50mg/mL and tested for integrity by SEC. Crystallizable protein should be >90% free of heterogeneity as judged by SDS-PAGE and contain no detectable aggregated species at high protein concentrations as judged by SEC. If the protein remains of high quality it is reconstituted into lipidic cubic phase containing cholesterol. The reconstituted protein is then dispensed onto glass sandwich crystallization plates and tested for crystallization using the screening methodology described in this Example and Example 1. After mixing, the protein-laden lipidid cubic phase mixture will comprise 3.6 - 7.2% w/w sterol, 56.5 - 52.8 w/w % Monoolein and 40% w/w protein solution (a 3:2 ratio of lipid mixture to protein). Initial crystallization conditions use PEG 400 between 25-35%, a salt between 50-500 mM, and a pH between 5.0 - 7.5.

Example 3. High Resolution Crystal Structure of a Human β_2-Adrenergic G protein-Coupled Receptor T4 Lysozyme Fusion Protein

[00232] The engineering, functional properties, expression and purification of crystallization grade β2AR-T4L protein are described in more detail in Example 4. Briefly, β2AR-T4L was generated by three distinct modifications to β2AR: (1) a fusion protein was created by replacement of the third intracellular loop with T4L, (2) the carboxyl terminal 48 amino acids were deleted, and (3) a glycosylation site at Asn187 was eliminated through a glutamate substitution. β2AR-T4L was expressed in Sf9 insect cells, solubilized in 1 % dodecylmaltoside, and purified by sequential antibody and ligand affinity chromatography. Using the modified lipidic cubic phase (LCP) crystallization procedure described herein, wherein crystals are grown from a cholesterol-doped monoolein cubic phase, β2AR-T4L crystals were obtained that diffract to a resolution of 2.2 Å. The structure was solved at 2.4 Å resolution. Compared to crystallization in detergents, LCP provides a more native, lipid environment for crystallization, as well as a confinement of protein molecules to two-dimensional membrane sheets that may facilitate the crystallization process through the formation of Type I packing interactions (Caffrey, *Curr Opin Struct Biol* 10, 486 (2000); Deisenhofer, *EMBO J* 8, 2149 (1989); Landau et al., *Proc Natl Acad Sci U S A* 93, 14532 (1996)).

Methods

[00233] **Lipidic cubic phase crystallization.** Crystals of engineered human β2AR (β2AR-T4L) grown from bicelles could not be optimized beyond 3.5 Å resolution (**Figure 5**). Lipidic cubic phase (LCP) crystallization trials were therefore performed using an *in meso* crystallization robot (Cherezov, et al., *Acta Crystallogr D Biol Crystallogr* 60, 1795 (2004)). 96-well glass sandwich plates (Cherezov, et al., *Acta Crystallogr D Biol Crystallogr* 60, 1795 (2004); Cherezov, *et al, J Membr Biol* 195, 165 (2003)) were filled with 25 or 50 nL protein-laden LCP drops overlaid by 0.8 µL of precipitant solution in each well and sealed with a glass coverslip. All operations starting from mixing lipid and protein were performed at room temperature (~21-23 °C). Crystals were obtained in 30-35% (v/v) PEG 400, 0.1-0.2 M sodium sulfate, 0.1 M Bis-tris propane pH 6.5-7.0 and 5-7% (v/v) 1,4-butanediol using 8-10% (w/w) cholesterol in monoolein as the host lipid (**Figure 6A**). Addition of cholesterol and 1,4-butanediol dramatically improved crystals size and shape, thereby enabling high-resolution diffraction. In this instance, additions of phospholipids (dioleoylphosphatidylcholine, dioleoylphosphatidylethanolamine, asolectin) to the main host LCP lipid monoolein (either alone or in combination with cholesterol) failed to improve crystal quality.

Crystal harvesting

[00234] The average size of the harvested crystals was 30 x 15 x 5 μm (largest crystal was 40 x 20 x 7 μm). Crystals were harvested directly from the glass sandwich plates, even though these plates have been specifically designed for screening and optimization (Cherezov, et al., *Acta Crystallogr D Biol Crystallogr* 60, 1795 (2004); Cherezov, *et al, J Membr Biol* 195, 165 (2003)). Crystals were scooped directly from the LCP using 30 or 50 μm aperture MiTeGen MicroMounts and plunged into liquid nitrogen. Care was taken to drag as little as possible lipid around the crystal to decrease unwanted background scattering. Attempts to dissolve the lipids, either by increasing concentration of PEG 400 or using a mineral oil, typically resulted in a decrease in diffraction power of the crystals.

Data collection

[00235] X-ray data were collected on the 23ID-B beamline (GM/CA CAT) at the Advanced Photon Source, Argonne, IL using a 10 μm minibeam (wavelength 1.0332 Å) and a MarMosaic 300 CCD detector (**Figure 6B**). Several complete datasets were collected from single crystals at resolution between 2.8 and 3.5 Å using 5x attenuated beam, 5 s exposure and 1° oscillation per frame. However, some crystals diffracted to a maximum of 2.2 Å resolution upon 5 s exposure with 1x attenuated beam. Therefore, 10-20° wedges of high-resolution data were collected from more than 40 crystals (some of the crystals were large enough to allow 2-3 translations). 31 of the best datasets from 27 independent crystals were then combined and scaled against the lower resolution full dataset to obtain complete 2.4 Å data.

[00236] One of the challenges during data collection was visualization of colorless microcrystals within an opaque frozen lipid phase and aligning them with the 10 μm minibeam. Because the crystals could not be adequately visualized through the inline optics at the beamline, alignment-by-diffraction techniques were employed. The present invention provides, in one aspect, an optimized crystal search algorithm to locate the crystals without the minibeam. First, the area of the loop containing lipid was scanned in the vertical direction with a highly attenuated and slitted 100 x 25 μm beam. When diffraction was found, the crystal location was further confined by two additional exposures to an area of ~50 x 25 μm. This area was further coarse-scanned with the collimated and 10x attenuated minibeam using 15 μm steps, following by fine-tuning the position using 5 and 2 μm steps. After locating the crystal in one orientation the loop was rotated 90° and the procedure was repeated. Typically during alignment the crystal was exposed ~10 times using 10x attenuated beam and 2 s exposures.

Data processing

[00237] A 90% complete, 2-fold redundant monoclinic dataset was processed from one crystal diffracting to 2.8 Å resolution. Initial indexing of lattice parameters in spacegroup C2 and crystal orientation were performed using HKL2000 (Otwinowski, *et al, in Methods in Enzymology* C. W. J. Carter, R. M. Sweet, Eds. (Academic Press, New York, 1997), vol. 276, pp. 307-326). The refined lattice parameters and space group were implemented in the data processing program XDS for spot integration which models error explicitly for radiation decay, absorption, and rotation (Kabsch, *J Appl Crystallogr* 26, 795 (1993)). Because data was collected using a 10 μm beam from microcrystals, maintaining the crystal orientation at the beam center during data collection was especially problematic. It appeared that XDS modeled the crystal orientation error upon rotation about the phi axis better than other data processing programs that were tried, resulting in better merging statistics. In addition to rotational error, the radiation decay was also an issue that was partially corrected by the XDS processing program, enabling a more reliable scaling of datasets from different crystals and translations of crystals. The 2.8 Å data was used as a scaling reference for incorporation of additional wedges of data collected at a much higher exposure. Each new dataset was indexed in XDS using the original unit cell parameters as constants which were then refined along with the crystal orientation, beam geometry, and mosaicity parameters. The refinement was generally stable, resulting in very similar unit cell constants which enabled subsequent scaling. All of the integrated wedges of data were then tested individually against the scaling reference set and included in the final scaled dataset if the merging statistics remained acceptable upon incorporation of the data. In total, 31 wedges of data from 27 crystals were combined with the scaling reference dataset, 22 of which diffracted to a resolution of 2.4 Å or better. Each of the higher resolution datasets were exposed to a much larger dose of radiation resulting in a rapid decay in intensity. Typically 10° – 20° wedges were collected from each crystal or translation, 5° – 7° of which had diffraction data to 2.4 Å. The final merging statistics for the dataset are shown in **Table 2**. Based on the mean $F/\sigma(F)$ of reflections near the three crystallographic axes, the effective resolution is estimated to be 2.4 Å along b* and c* and 2.7 Å along a*. The anisotropy results in the high merging R factors in the last few resolution shells despite the significant $I/\sigma(I)$ values. The anisotropy is either an inherent property of the crystals or the result of a preferential orientation of the crystals within the mounting loop. Thus, the higher resolution shells were filled in anisotropically by incorporation of the additional data at high exposure levels, while the lower resolution shells have a very high redundancy and low anisotropy.

Table 2. Data processing statistics from XDS. A comparison is made between data filtered by a sigma cutoff and the complete set. All of the data was used in structure solution and refinement.

Signal / noise ≥ -3

Resolution	# Observed refl	# Unique refl	Redundancy	Completeness	R_{sym}	I/ SIGMA	R-meas	Rmrgd-F
10	3352	333	10.1	87.40%	6.60%	23.25	6.90%	2.70%
8	3591	354	10.1	99.40%	7.60%	22.6	8.00%	2.80%
6	10480	1003	10.4	99.60%	9.90%	20.01	10.50%	4.00%
3	126008	11968	10.5	99.80%	13.30%	14.01	14.00%	6.60%
2.8	33158	3130	10.6	100.10%	38.00%	6.42	39.90%	18.70%
2.7	19702	1893	10.4	99.70%	49.50%	4.96	52.10%	24.90%
2.6	23772	2275	10.4	99.90%	60.20%	4.07	63.40%	30.10%
2.5	14108	2558	5.5	99.30%	58.90%	2.69	65.10%	51.80%
2.4	14672	3060	4.8	99.10%	67.80%	2.18	75.70%	62.80%
Total	248843	26574	9.4	99.50%	12.70%	9.62	13.40%	11.40%

Signal / noise ≥ 0

Resolution	# Observed refl	# Unique refl	Redundancy	Completeness	R_{sym}	I/SIGM A	R-meas	Rmrgd-F
10	3352	333	10.1	87.40%	6.60%	23.25	6.90%	2.70%
8	3591	354	10.1	99.40%	7.60%	22.6	8.00%	2.80%
6	10480	1003	10.4	99.60%	9.90%	20.01	10.50%	4.00%
3	125628	11923	10.5	99.40%	13.30%	14.06	14.00%	6.50%
2.8	32679	3077	10.6	98.40%	37.60%	6.54	39.50%	17.70%
2.7	19346	1849	10.5	97.40%	48.80%	5.08	51.30%	23.40%
2.6	23201	2210	10.5	97.00%	58.90%	4.2	62.00%	28.00%
2.5	13461	2406	5.6	93.40%	56.40%	2.88	62.30%	45.90%
2.4	13833	2827	4.9	91.50%	64.10%	2.4	71.50%	54.30%
Total	245571	25982	9.5	97.30%	12.60%	9.85	13.30%	10.80%

$R_{sym} = \Sigma_{hkl} |I(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}(hkl)$, where <I(hkl)> is the mean of the symmetry equivalent reflections of I(hkl); R-meas = redundancy independent Rsym; Rmrgd-F= indication of amplitude quality. See, e.g., Murshudov, et al, *Acta Crystallogr D* D53, 240 (1997).

Structure solution and refinement

[00238] Initial phases for β2AR-T4L were obtained by molecular replacement using both T4-lysozyme (PDB ID Code 2LZM) and a polyalanine model of the rhodopsin seven-transmembrane bundle (PDB ID Code 1U19) as search models. It was necessary to trim the lysozyme search model to remove residues 12-71 as that domain had shifted conformations relative to the larger section. This domain was later reintroduced to the model by fitting into observed density. Molecular replacement was carried out using the program Phaser by first placing the truncated lysozyme (RFZ= 3.74; TFZ= 3.65) followed by the rhodopsin model (RFZ= 5.2; TFZ= 7) (McCoy, *Acta Crystallogr D Biol Crystallogr* **63**, 32 (2007)). In order to optimize placement of the receptor each of the seven helices was refined independently by rigid body maximum likelihood refinement as implemented in Refmac (Initial Rwork/Rfree = 0.50 / 0.51) (Murshudov, *et al, Acta Crystallogr D* **D53**, 240 (1997)).

[00239] Initial rounds of refinement were carried out using restrained parameters in Refmac. Model rebuilding was performed in Coot utilizing 2 Fo-Fc sigma-A weighted maps, as well as density modified maps calculated using Resolve prime-and-switch phasing which reduces model bias introduced by model based phasing methods (Terwilliger, *Acta Crystallogr D* **D56**, 965 (2000)). The Resolve calculated maps were superior to the sigma-A weighted ones in that more of the main chain density could be traced. Density for the bound ligand was visible early in the refinement but was not modeled immediately to allow an unbiased assessment of the phase quality through the improvement of the signal/noise of the observed ligand density. The structure quality is excellent (**Table 3**), with strong electron density in particular observed in the ligand binding site (**Figure 7**), cholesterol binding sites (**Figure 8A**), and the proline helix kinks (**Figure 8B**).

Table 3 Data collection and refinement statistics

	β_2AR-T4L
Data collection (APS GM/CA CAT 23ID-B, 10 μm beam)*	
Space group	C2
Cell dimensions	
a, *b*, *c* (Å)	106.32, 169.24, 40.15
β (°)	105.62
No. of reflections processed	245,571
No. unique reflections	26,574
Resolution (Å)	50 – 2.4 (2.5 – 2.4)
R_{sym}	12.7 (67.8)
Mean I/σ(I)	9.6 (2.2)
Completeness (%)	99.5(99.1)
Redundancy	9.4 (4.8)
Refinement*	
Resolution (Å)	20 – 2.4 (2.46 – 2.4)
No. reflections (test set)	25,247 (1,310)
R_{work} / R_{free}	19.8(27.0) / 23.2(30.1)
No. atoms	3,805
Protein	3,544
Ions, lipids, ligand and other	213
Water	48
Overall *B*-values (Å^2)	82
β_2AR	77
T4-Lysozyme	75
Carazolol	55
Lipid	100
R.m.s deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.5
Ramachandran plot statistics (%) (excl. Gly, Pro):	
Most favored regions	94.8
Additionally allowed regions	5.0
Generously allowed regions	0.2
Disallowed regions	0

*Highest resolution shell is shown in parenthesis.$R_{sym} = \Sigma_{hkl} |I(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}(hkl)$, where <I(hkl)> is the mean of the symmetry equivalent reflections of I(hkl).

Analysis of β2A receptor topology

[00240] The final model of β2AR-T4L includes 442 amino acids. β2AR-T4L was treated with iodoacetamide during purification to eliminate free thiols. The model includes a palmitic acid covalently bound to Cys341 (GPCRs are frequently post-translationally modified with palmitoylate on cysteine residues at the C-terminal tail) and an acetamide molecule bound to $Cys265^{6..27}$. Throughout the description, residues are designated by their position within the β2AR sequence and their Ballesteros-Weinstein designation as a superscript where applicable. Ballesteros-Weinstein numbering is used throughout the text as superscripts to the protein numbering. Within each helix is a single most conserved residue among the class A GPCRs. This residue is designated x.50 where x is the number of the transmembrane helix. All other residues on that helix

are numbered relative to this conserved position. The model also includes one carazolol molecule, three cholesterol molecules, two sulfate ions and two butanediol molecules that interact with β2AR. There are also four sulfate ions, a putative disaccharide (modeled as maltose) and a molecule of PEG 400 bound to T4L. For β2AR, excellent electron density is observed for residues 29-342, including the ligand carazolol and the two disulfide bonds Cys106$^{3.25}$-Cys191$^{5.30}$ and Cys184$^{4.76}$-Cys190$^{5.29}$. The palmitic acid at Cys341 is clearly visible in Fo-Fc omit maps; however, the quality of the electron density is lower than for the rest of the receptor. The N-terminus (residues 1 to 28) and the majority of the C-terminus (residues 343 to 365) are disordered and not visible in the structure.

[00241] The β2AR has a fold composed of seven transmembrane helices forming a helical bundle (**Figure 9A**). The residues that make up the helices (I to VII) in β2AR are as follows: helix I 29$^{1.28}$ to 60$^{1.59}$, helix II 67$^{2.38}$ to 96$^{2.67}$, helix III 103$^{3.22}$ to 136$^{3.55}$, helix IV 147$^{4.39}$ to 171$^{4.63}$, helix V 197$^{5.36}$ to 229$^{5.68}$, helix VI 267$^{6.29}$ to 298$^{6.60}$, and helix VII 305$^{7.32}$ to 328$^{7.55}$. The residues forming the intracellular loops (ICL) and extracellular loops (ECL) of β2AR are: ICL1 61$^{1.60}$ to 66$^{2.37}$, ECL1 97$^{2.68}$ to 102$^{3.21}$, ICL2 137$^{3.56}$ to 146$^{4.38}$, ECL2 172$^{4.64}$ to 196$^{5.35}$, ICL3 230$^{5.69}$ to 266$^{6.28}$ (residues 231 to 262 are replaced by T4-lysozyme residues 2 to 161), and ECL3 299$^{6.61}$ to 304$^{7.31}$. Helices II, V, VI and VII each have a proline-induced kink at conserved positions along the span of the transmembrane segments. These kinks are thought to enable the structural rearrangements required for activation of G protein effectors (Yohannan et al., *Proc Natl Acad Sci U S A* 101, 959 (2004)). In addition to the seven membrane spanning helices, β2AR possesses two other helical segments: helix VIII, which is believed to be common to all rhodopsin-like GPCRs (Katragadda et al., *Biochim Biophys Acta* 1663, 74 (2004)), and an unexpected, short helical segment in the middle of ECL2, which is not present in rhodopsin, and was not predicted by computational secondary structure analysis (**Figure 9A**).

[00242] In the β2AR-T4L construct, T4L is fused to the truncated cytoplasmic ends of helices V and VI. In the crystal structure, the T4L moiety is tilted slightly away from the center axis of β2AR drawn normal to the membrane (**Figure 9B**). As a result, interactions between T4L and β2AR are minimal, with only 400Å2 of surface area buried between them. The intramolecular contacts between T4L and β2AR include salt bridges between the side chains of T4L-Asp159 and the side-chain amine of β2AR-Lys227$^{5.66}$ (distance 3.4 Å) and between the guanidinium group of T4L-Arg8 with the side-chain carboxyl of β2AR-Glu268$^{6.30}$ on helix VI (distance 3.2 Å) (**Figure 9C**, **Table 4**). The latter interaction is noteworthy because it differs from rhodopsin where Glu6.30 forms an ionic bond with Arg3.50 of the conserved D(E)RY motif. This interaction is postulated to be important for maintaining rhodopsin in the inactive state, but the charged groups of the two residues

[Arg131$^{3.50}$ (NH1) and Glu268$^{6.30}$ (OE1)] are 10 Å apart in the β2AR-T4L structure. The remainder of the lysozyme molecule provides important crystal packing interactions, but does not appear to influence significantly the receptor structure.

Table 4: Direct contacts between β_2AR and T4L

van der Waals Contacts		
β_2AR atom	T4 Lysozyme	
Leu230$^{5.69}$	Trp 158	
Lys263$^{6.25}$	Asp159	
Cys265$^{6.27}$	Ile9	
Leu266$^{6.28}$	Ile9	
Leu266$^{6.28}$	Glu5	
Hydrogen Bond and Salt Bridge Contacts		
β_2AR atom	T4 Lysozyme	Distance (Å)
Lys227$^{5.66}$(NZ)	Asp159(OD1)	3.4
Gln229$^{5.68}$(O)	Asn2 (N)	3.1
Gln229$^{5.68}$(O)	Asn2 (ND2)	3.2
Gln268$^{6.30}$(OE2)	Arg8 (NH2)	3.2
Covalent bonds		
β_2AR atom	T4 Lysozyme	
Leu230$^{5.69}$	Asn2	
Lys263$^{6.25}$	Tyr161	

Crystal packing interactions

[00243] The β2AR-T4L protein is packed in a C-centered monoclinic lattice with one molecule per asymmetric unit (**Figure 10A**). Membrane protein generally can form two types of crystal packing: Type I represents stacks of two dimensional crystals ordered in the third dimension via interactions of hydrophilic parts of membrane proteins. Type II crystals are composed of membrane proteins whose hydrophobic part is shielded by a detergent micelle and all crystal contacts are formed through hydrophilic, solvent exposed parts of protein molecules. As observed in all previous lipidic mesophase grown crystals (Schwartz, et al., *Annu Rev Pharmacol Toxicol* 46, 481 (2006)), the β2AR-T4L crystals adopt Type I packing, featuring a multilayered arrangement in accordance with proposed crystallization mechanism (Caffrey, *Curr Opin Struct Biol* 10, 486 (2000); P. Nollert, et al., *FEBS Lett* 504, 179 (2001)). Within each layer, protein molecules form arrays of parallel, symmetry-related dimers. There are four distinct crystal-packing interactions within each layer, three of which are mediated by T4L. The fourth interaction in the array is between two receptor molecules related by a crystallographic two-fold rotation axis. This is the sole interaction between symmetry-related receptors, and is mediated primarily by ordered lipids consisting of six cholesterol and two palmitic acid molecules, the latter being covalently attached to Cys341 in the C-terminal portion of the receptor (O'Dowd et al., *J Biol Chem* 264, 7564 (1989)) (**Figure 10B**). These eight lipid molecules form a two-fold symmetric sheet between receptors.

The only direct receptor-receptor contact involves a 2.7 Å pair of ionic interactions between the charged amine group of $Lys60^{1.59}$ in helix I and the carboxylate of Glu^{338} in helix VIII from the symmetry-related receptor. Remarkably, of the 515 Å2 buried at the receptor symmetry interface, 73% of the crystal contact surface area is mediated by ordered lipid, while only 27% is contributed by protein-protein contacts. The stacking interactions between layers are formed between T4L and extracellular loops ECL2 and ECL3 of the receptor (**Figure 10A**). It is unlikely that these contacts affect the orientation of these loops due to the small size of ECL3 and the rigid architecture of ECL2.

Lipid mediated receptor association

[00244] Many GPCRs including β2AR are thought to exist as dimers in the plasma membrane, although the location of the dimer interface and the functional significance of dimerization is not clear (Milligan, *Mol Pharmacol* 66, 1 (2004)). The observation of ordered lipids in the helix I and VIII interface between two symmetry related molecules suggests the association is physiologically relevant (Angers, et al., *Proc Natl Acad Sci U S A* 97, 3684 (2000); Javitch, *Mol Pharmacol* 66, 1077 (2004); Mercier, et al., *J Biol Chem* 277, 44925 (2002)). Associations between the equivalent regions of rhodopsin have been found in crystal structures (Salom et *al., Proc Natl Acad Sci U S A* 103, 16123 (2006); Schertler, *Curr Opin Struct Biol* 15, 408 (2005)) (**Figure 10C**).

[00245] The role of cholesterol in the physiologic function of β2AR is well documented. Depletion of cholesterol from the membranes of neonatal cardiac myocytes alters the signaling behavior of endogenous β2AR (Xiang, et al., *J Biol Chem* 277, 34280 (2002)). In untreated cells, activation of β2AR results in sequential coupling to the G proteins Gs and Gi, producing a biphasic effect on myocyte contraction rate. Upon depletion of cholesterol, the β2AR couples more strongly to Gs. This effect may be due to a role of cholesterol in regulating interactions between the β2AR and G proteins, or an effect of cholesterol on β2AR dimerization. The β2AR couples efficiently to Gs as a monomer (Mialet-Perez, et al., *J Biol Chem* 279, 38603 (2004)), so cholesterol mediated association (dimerization) may reduce the efficiency of β2AR coupling to Gs. The effects of cholesterol depletion on β2AR signaling may also be a secondary effect of altering subcellular signaling compartments. There is evidence that cells may concentrate signaling molecules, such as GPCRs and their cognate G proteins, by way of membrane microdomains or compartments, such as caveolae (Ostrom, et al., *Br J Pharmacol* 143, 235 (Sep, 2004)). This compartmentalization may be a major regulator of receptor-effector coupling. Thus, the importance of cholesterol in forming the observed crystallographic association is consistent with its role in β2AR signaling.

Electrostatic charge distribution

[00246] Electrostatic charge distribution was calculated using APBS (Gether, *Endocr Rev* 21, 90 (2000)) and mapped onto a molecular surface representation of β2AR. The analysis reveals three polarized areas within the molecule (**Figure 11A**). First, the cytoplasmic face of the receptor is involved in G protein interaction and carries a net positive charge even in the absence of ICL3, which also has a predicted overall positive charge (**Figure 11B**). The second site is an electrostatically negative region located within the membrane between helices III, IV and V potentially exposed to the lipid alkyl chains, which is unexpected as the burial of charge within the plasma membrane is thermodynamically unfavorable. A glutamate residue at position $122^{3.41}$ may partially account for the observed charge distribution. Finally, the binding site cleft is negatively charged and exposed to solvent by an unusual ECL2 architecture and lack of N-terminal interactions. This negative charge may facilitate ligand binding through electrostatic funneling of positively charged catecholamines (**Figure 11B**).

Extracellular region

[00247] The ECLs and amino termini of GPCRs, together with the extracellular halves of the transmembrane helices, are believed to define the ligand-binding site of each receptor (Angers et al., *Proc Natl Acad Sci U S A* 97, 3684 (2000)). Therefore, the ECLs play an important role in the overall pharmacology of any particular receptor. In general, small molecule ligands are thought to bind deeper within the space created by the transmembrane domain helices, whereas larger ligands such as peptides bind closer to the membrane surface near the ECLs (Ji, et al., *J Biol Chem* 273, 17299 (1998); Gether, *Endocr Rev* 21, 90 (2000)). Mutagenesis studies suggest that the β2AR binds its ligand deep within the transmembrane helix bundle, which may be related to the observation that the extracellular regions have a rather simple structure with short loops connecting transmembrane helices II and III, and VI and VII (**Figure 12A**). ECL2, which links helices IV and V, has a somewhat more extensive architecture that is unanticipated. In contrast to the buried, β-sheet structure of this loop in rhodopsin (**Figure 12B**), ECL2 in β2AR is more exposed to the solvent and contains an extra helical segment. Additionally, there is an intra-loop disulfide bond between $Cys184^{4.76}$ and $Cys190^{5.29}$ that may help stabilize the more exposed ECL2. A second disulfide bond between $Cys191^{5.30}$ and $Cys106^{3.25}$ in helix III effectively ties ECL2 to the transmembrane core (Noda, et al., *J Biol Chem* 269, 6743 (1994)). The distal portion of ECL2 makes close contacts with ECL1 and contains a glycosylation site at $Asnl87^{5.26}$ (Mialet-Perez, et al., *J Biol Chem* 279, 38603 (2004)), which may serve to mask a grouping of aromatic residues on ECL1; in this construct, Asn1875.26 has been mutated to glutamate to aid in crystallization.

[00248] Electron density corresponding to the N-terminus was not apparent in the maps and, therefore, residues 1-28 are not included in the model. This disorder contrasts with rhodopsin, in which the N-terminus interacts extensively with the ECLs, forming a small four-strand β-sheet in conjunction with ECL2. This sheet structure forms a cap that effectively isolates the retinal binding site in a hydrophobic pocket (**Figure 13B**). The lack of interactions between the N-terminus of β2AR and ECL2 further enables diffusible ligand access to the binding site. However a completely disordered N-terminus may be an artifact induced by the presence of the N-terminal Flag tag which carries an overall positive charge and may disrupt N-terminal interactions.

[00249] The short helical region on ECL2 adds a rigid structural element that, along with the two disulfide bonds, constrains the loop to a small range of conformations and helps stabilize the receptor by linking three transmembrane helices (**Figure 13A**). This rigid conformation may help to stabilize the core of the receptor and lock ECL2 in a conformation that does not hinder access to the binding pocket.

Ligand binding site and comparison to rhodopsin

[00250] Carazolol is a partial inverse agonist that binds with picomolar affinity to β_2AR-T4L producing a reduction of the basal activity of the receptor. The crystal structure reveals extensive interactions between the receptor and carazolol that position the carbazole moiety adjacent to Phe289$^{6.51}$, Phe290$^{6.52}$, and Trp286$^{6.48}$ (**Figure 13A**, **Figure 7**, and **Table 5**). In contrast, *cis*-retinal is a full inverse agonist covalently bound to rhodopsin, which suppresses all activity towards transducin (Palczewski, *Annu Rev Biochem* 75, 743 (2006)). Carazolol and retinal occupy similar spaces in their respective receptors, with significant overlap of the non-aromatic regions of carazolol. However, the β-ionone ring of retinal extends deep into the binding pocket of rhodopsin and contacts residues on helix V and VI, where it is sandwiched between Phe212$^{5.47}$ and Tyr268$^{6.51}$, and interacts with the highly conserved Trp265$^{6.48}$ (**Figure 13B**). It has been proposed that changes in the rotamer of Trp265$^{6.48}$ occur upon activation of rhodopsin and related family members, and constitutes the "toggle switch" for receptor activation (Schwartz, et al., *Annu Rev Pharmacol Toxicol* 46, 481 (2006)). Accordingly, the interactions between c/s-retinal and Trp265$^{6.48}$ are likely to contribute to the absence of basal activity in rhodopsin. Carazolol does not interact directly with the toggle switch on helix VI, however it lowers the basal activity of the receptor, and may do so by interacting with Phe289$^{6.51}$ and Phe290$^{6.52}$, which form an extended aromatic network surrounding the highly conserved Trp286$^{6.48}$. As a result, Trp286$^{6.48}$ adopts the rotamer associated with the inactive state. Thus, the steric constraints imposed by Phe290$^{6.52}$ appear to structurally mimic the interaction of the β-ionone ring of retinal with the conserved Trp265$^{6.48}$ and Phe212$^{5.47}$ on rhodopsin (Shi et al., *J Biol Chem* 277, 40989 (2002)) (**Figure 13C**).

Table 5: Direct contacts between β_2AR and carazolol.

Hydrogen Bond and Salt Bridge Contacts		
β_2AR atom	Carazolol	Distance (Å)
Asp113$^{3.32}$ (OD2)	N19	2.9
Asp113$^{3.32}$ (OD1)	O17	2.6
Ser203$^{5.42}$ (OG)	N7	3.2
Asn312$^{7.39}$ (ND2)	O17	2.9
Asn312$^{7.39}$ (OD1)	N19	2.9
Tyr316$^{7.43}$ (OH)	N19	3.4
Hydrophobic and Aromatic Interactions (closest distance for each residue, <4 Å)		
β_2AR atom	Carazolol	Distance (Å)
Trp109$^{3.28}$ (CH2)	C21	3.8
Val114$^{3.33}$ (CG1)	C11	3.9
Val117$^{3.36}$ (CG1)	C12	4.0
Thr118$^{3.37}$ (OG1)	C11	3.9
Phe193$^{5.32}$ (CE2)	C6	3.5
Tyr199$^{5.38}$ (CE2)	C2	3.9
Ser207$^{5.46}$ (CB)	C10	3.6
Trp286$^{6.48}$ (CH2)	O17	3.4
Phe289$^{6.51}$ (CE2)	O14	3.7
Phe290$^{6.52}$ (CZ)	C12	3.5
Asn293$^{6.55}$ (ND2)	C5	3.6
Tyr308$^{7.35}$ (OH)	C6	3.6

Structural alignment and helix bundle reorganization

[00251] It has long been thought that class A GPCRs share a similar architecture due to their predicted seven transmembrane helical bundles and sequence conservation within the membrane spanning regions (Lefkowitz, *Nat Cell Biol* 2, E133 (2000)). We aligned the structure of β2AR-T4L to highest resolution structure of rhodopsin (PDB ID Code 1U19) to evaluate the similarities and differences in ligand binding modes. We used difference distance matrices to select non-divergent areas between the two structures that align to reveal the differences in helix orientation between β2AR-T4L and rhodopsin. For the alignment, residues on β2AR were aligned to equivalent residues on Rhodopsin, respectively: 43-59 to 47-63; 67-95 to 71-99; 122-135 to 126-139; 285-296 to 264-275.

[00252] Relative to rhodopsin, the following helical shifts are seen in β2AR-T4L: the extracellular portions of helices I and III angle away from the center of the receptor, helix IV is translated away from the center of the receptor, helix V is translated closer to the center of the receptor and helix VI angles away from the receptor on the cytoplasmic end (**Figure 14**). The largest difference is in helix I, which lacks a proline-induced kink found in rhodopsin and is comparatively straight. The angle between the rhodopsin and β2AR positions of helix I is approximately 18° with a shift of 7 Å at the apex on the extracellular face. This structural difference may arise from the need for an accessible binding site in β2AR, which is provided in part by a lack of interactions between

the N-terminus and extracellular loop segments. In contrast the N-terminal region in rhodopsin occludes the retinal-binding site through extensive interactions with the extracellular loops (**Figure 12B**). Helix V of β2AR is closer to the binding pocket by approximately 3.5 Å on average and its lumenal end is angled more towards helix VI. Helix IV of β2AR is further from the binding site, possibly to remove steric clashes resulting from the modified position of helix V (**Figure 14B, 14C**). Helix III pivots further from the binding site about a fulcrum located close to the cytoplasmic end (**Figure 14C**). The angle formed between rhodopsin helix III and the β2AR helix III is approximately 7°, yielding a 4 Å displacement out of the binding pocket at the cytoplasmic end of the helix. Helix VI is positioned further from the center of the receptor at the cytoplasmic end as compared to rhodopsin, which is caused by a slight difference in the angle about the proline-induced kink in the helix (**Figure 14C**).

[00253] The ligand-binding pocket is formed by both structurally conserved and divergent helices as compared to rhodopsin (**Figure 14D**). Helices III and V are two of the most conformationally shifted helices and contain the canonical catecholamine binding residues associated with activation of adrenergic family of receptors (Strader et al., *J Biol Chem* 263, 10267 (1988); Strader, et al., *J Biol Chem* 264, 13572 (1989); Liapakis et al., *J Biol Chem* 275, 37779 (2000)). The comparison with rhodopsin shows that the structurally conserved helices provide a common core present throughout the class A GPCRs, whereas the variable helices confer binding site plasticity with a resulting architecture capable of binding a large spectrum of ligands.

Comparison to rhodopsin-based GPCR models

[00254] Since the determination of the inactive dark-state rhodopsin structure (Palczewski et al., *Science* 289, 739 (2000)), a number of homology models of other class A GPCRs have been reported (Bissantz, et al., *Proteins* 50, 5 (2003); Fano, et al., *J Chem Inf Model* 46, 1223 (2006); Hobrath, et al., *J Med Chem* 49, 4470 (2006); Nowak, et al., *J Med Chem* 49, 205 (2006); Zhang, et al., *PLoS Comput Biol* 2, e13 (2006)). Typically, homology models start by alignment of so-called fingerprint motifs that are common among the family. These fingerprint motifs are extrapolated to assign coordinates for the entire helical bundle. Loop regions are either ignored or modeled based on databases of loop conformations depending on the application (Bissantz, *et. al, Proteins* 50, 5 (2003)). A number of models exist for β2AR, some of which have been improved upon with supporting biochemical data (Bissantz, *et. al, Proteins* 50, 5 (2003); Zhang, et al., *PLoS Comput Biol* 2, e13 (2006); Freddolino et al., *Proc Natl Acad Sci U S A* 101, 2736 (2004); Furse, et al., *J Med Chem* 46, 4450 (2003); Gouldson et al., *Proteins* 56, 67 (2004)). When compared to the β2AR structure reported here (according to the methods described above in this Example), however, all of these models were more similar to

rhodopsin, as were models for other receptors (e.g. dopamine, muscarinic, and chemokine). This highlights a general shortcoming in homology models generated from a single structural template. The structural divergence between β2AR and rhodopsin would be quite difficult to predict accurately using only rhodopsin as a template.

Example 4: Structural Insights into β_2 Adrenergic Receptor Function

Methods

Molecular biology for generation of mammalian and Sf9 expression constructs.

[00255] The insect cell expression plasmid that was used as a template for modification of the human β2AR gene has been described previously (Yao et al., *Nat Chem Biol* 2, 417 (2006)): the wild-type coding sequence of the human β2AR (starting at Gly2) was cloned into the pFastbac1 Sf-9 expression vector (Invitrogen) with the HA signal sequence followed by the Flag epitope tag at the amino terminus and the third glycosylation site mutated as N187E. Using this template, a TAA stop codon was placed between Gly365 and Tyr366, terminating translation without the 48 C-terminal residues of the wild-type β2AR ("β2AR365"). A synthetic DNA cassette encoding the T4 Lysozyme (WT*- C54T, C97A) protein was made by overlapping extension PCR of 50-base oligonucleotides. This cassette was amplified and inserted into the β2AR365 construct between $Ile233^{5.72}$ and $Arg260^{6.22}$ (**Figure 21A**), using the Quickchange Multi protocol (Stratagene). The corresponding mammalian cell expression plasmid was made by amplifying the entire fusion gene and cloning it into pCDNA3 (Invitrogen). Further deletions in the Sf9 and mammalian cell constructs were made using appropriate synthetic oligonucleotides in the Quickchange Multi protocol (Stratagene). All constructs were confirmed by sequencing.

HEK293 cell staining and immunofluorescence staining.

[00256] HEK293 cells were cultured on plastic dishes at 37 °C with 5% CO2 in Dulbecco's modified Eagle's medium (Cellgro) with 5% fetal bovine serum. For an individual expression experiment, cells at confluency were split, and approximately 100,000 cells were used to seed glass cover slips in the same medium. After 2 d, cells were transfected with the addition of 1 μg of a given pCDNA3-receptor plasmid and 3 μl of Fugene 6 reagent (Roche). 48 h after transfection, cells were washed with PBS, fixed with 4% paraformaldehyde, blocked with PBS + 2% goat serum, permeabilized with PBS + 2% goat serum + 0.5% Nonidet P-40 (Sigma), stained with Alexa488- conjugated M1 anti-FLAG antibody (for receptor) plus DAPI (nuclear) in blocking buffer, and washed with blocking buffer. Cover slips were mounted on microscope slides with Vectashield (Vector Labs) and dried overnight. Staining was visualized with an Axioplan 2 fluorescence imaging system, using a 63x objective and either green (Alexa488/FITC) or blue (DAPI/Hoechst) filter sets. A

plasmid pCDNA3-β1AR, expressing an N-terminal FLAG-tagged β1 adrenergic receptor, was used as a positive control for cell-surface staining. Empty pCDNA3 was used as a negative control to assess background staining.

Expression and purification of β_2AR-T4L from baculovirus-infected Sf9 cells.

[00257] Recombinant baculovirus was made from pFastbac1-β2AR-T4L using the Bac-to-Bac system (Invitrogen), as described previously (Yao et al., *Nat Chem Biol* 2, 417 (2006)). The β2AR-T4L protein was expressed in Sf9 insect cells infected with this baculovirus, and solubilized according to previously described methods (Kobilka, *Anal Biochem* 231, 269 (1995)). Dodecylmaltoside-solubilized receptor with the N-terminal FLAG epitope (DYKDDDA) (SEQ ID NO: 1) was purified by M1 antibody affinity chromatography (Sigma), treated with TCEP/iodoacetamide, and further purified by alprenolol-Sepharose chromatography (Kobilka, *Anal Biochem* 231, 269 (1995)) to isolate only functional GPCR. Eluted alprenolol-bound receptor was re-bound to M1 FLAG resin, and ligand exchange with 30 μM carazolol was performed on the column. β2AR-T4L was eluted from this final column with 0.2 mg/ml FLAG peptide in HLS buffer (0.1% dodecylmaltoside, 20 mM Hepes, 100 mM NaCl, pH 7.5) plus 30 μM carazolol and 5 mM EDTA. N-linked glycolsylations were removed by treatment with PNGaseF (NEB). Protein was concentrated from ~5 mg/ml to 50 mg/ml with a 100 kDa molecular weight cut-off Vivaspin concentrator (Vivascience), and dialyzed against HLS buffer plus 10 μM carazolol.

Binding measurements on wild-type β_2AR and β_2AR-T4L from membranes.

[00258] Membrane preparation from baculovirus-infected Sf9 cells was performed as described previously (Swaminath, et al., *Mol Pharmacol* 61, 65 (2002)). For each binding reaction, membranes containing 0.7 μg total membrane protein were used. Saturation binding of [^{3}H]-dihydroalprenolol (DHA) was measured by incubating membranes resuspended in 500 μl binding buffer (75 mM Tris, 12.5 mM $MgCl_2$, 1 mM EDTA, pH 7.4, supplemented with 0.4 mg/ml BSA) with 12 different concentrations of [^{3}H]DHA (Perkin Elmer) between 20 pM and 10 nM. After 1 h incubation with shaking at 230 rpm, membranes were filtered from the binding reactions with a Brandel harvester, washed with binding buffer, and measured for bound [^{3}H]DHA with a Beckman LS6000 scintillation counter. Non-specific binding was assessed by performing identical reactions in the presence of 1 μM alprenolol. For competition binding, membranes resuspended in 500 μl binding buffer were incubated with 0.5 nM [^{3}H]DHA plus increasing concentrations of the competing ligand (all compounds were purchased from Sigma). For (-)-isoproterenol and (-)-epinephrine, concentrations were 100 pM – 1mM, each increasing by a factor of 10. For salbutamol, concentrations were 1 nM – 10 mM. For ICI-118,551 and formoterol, concentrations were 1 pM – 10 μM. Non-specific binding was measured by using 1 μM unlabeled

alprenolol as competing ligand. Each data point in the curves in Figures 2A and S1 represents the mean of three separate experiments, each done in triplicate. Binding data were analyzed by nonlinear regression analysis using Graphpad Prism. The values for K_d of [^{3}H]DHA and K_i of other ligands are shown in **Table 6**.

Table 6.

Saturation Binding

[^{3}H]DHA	$K_d \pm SE$ (nM)	Bmax (pmol/mg)
β_2AR	0.161 ± 0.012	30.0 ± 0.5
β_2AR-T4L	0.180 ± 0.016	21.6 ± 0.5

Competition Binding

Ligand	K_i [S.E. interval] for β_2AR (nM)	K_i [S.E. interval] for β_2AR-T4L (nM)
(-)-isoproteronol	50.6 [48.9-52.3]	15.7 [15.2-16.2]
(-)-epinephrine	175 [163-188]	56.0 [52.8-59.4]
salbutamol	728 [708-750]	307 [291-323]
ICI-118,551	0.617 [0.570-0.668]	0.626 [0.591-0.662]
formoterol	3.60 [3.39-3.83]	1.68 [1.55-1.81]

Binding affinities of different ligands for the wild-type β_2AR and the fusion protein β_2AR-T4L. The saturation and competition binding curves shown in **Figure 22** were fit to theoretical saturation and one-site competition binding models, using the program Graphpad Prism. K_i values were calculated using the Cheng-Prusoff equation: $K_i = IC_{50}/(1+[ligand]/K_d)$.

Bimane fluorescence experiments on purified, detergent-solubilized receptors

[00259] β_2AR-T4L and β_2AR365 were purified as described above, with two important differences. First, prior to iodoacetamide treatment, FLAG-pure receptor at 2.5 μM (measured by soluble [^{3}H]DHA binding) was incubated with 5 μM monobromobimane for 1 h at 4 °C. Second, after binding the bimane-labeled alprenolol-Sepharose-purified receptor to M1 antibody resin, the column was washed extensively with ligand-free buffer before elution. Based on previous precedent (Ghanouni, et al., *Proc Natl Acad Sci U S A* 98, 5997 (2001)), this protocol is expected to target primarily Cys265$^{6.27}$ for fluorophore derivitization. Fluorescence spectroscopy was performed on a Spex FluoroMax-3 spectrofluorometer (Jobin Yvon Inc.) with photon-counting mode, using an excitation and emission bandpass of 5 nm. All experiments were done at 25 °C. For emission scans, we set excitation at 350 nm and measured emission from 417 to 530 nm with an integration time of 1.0 s nm^{-1}. To determine the effect of ligands, spectra were measured after 15 min incubation with different compounds (at saturating concentrations: [(-)-isoproterenol] = 100 μM; [ICI-118,551] = 10 μM; [salbutamol] = 500 μM). Fluorescence intensity was corrected for background fluorescence from buffer and ligands in all experiments. The curves shown in **Figure 22B** are each the average of triplicate experiments performed in parallel. λ_{max} values and intensity changes for β_2AR-T4L and β_2AR365, each incubated with different ligands, are tabulated in **Table 7**.

Table 7.

Ligand	λmax ± SD for (β_2AR365 (nm)	λmax ± SD for β_2AR-T4L (nm)
none	448 ± 2	447 ±2
(-)-isoproteronol	453 ± 2	455 ±2
ICI-118,551	447 ± 1	446 ± 1
salbutamol	449 ± 1	449 ±1

Intensity at λmax_{Ligand} / Intensity at λmax_{none}

Ligand	β_2AR365	β_2AR-T4L
(-)-isoproteronol	0.758 ± 0.007	0.824 ± 0.006
ICI-118,551	1.013 ± 0.008	1.028 ± 0.008
salbutamol	0.950 ± 0.013	0.928 ± 0.009

Bimane fluorescence responses for unliganded β_2AR365 and β_2AR-T4L, incubated for 15 min with different ligands. Top panel shows the λ_{max} for fluorescence emission spectra (excitation at 350 nm and emission from 417 to 530 nm) collected after 15 min incubation with ligand. Each value is mean ± standard deviation for triplicate experiments performed in parallel. Bottom panel shows the change in fluorescence intensity after incubation with ligand, represented as the ratio of Intensity at λmax of the ligand to Intensity at λ_{max} of the control no ligand ("none") response.

Comparing the proteolytic stability of unliganded β_2AR and β_2AR-T4L.

[00260] The limited trypsin proteolysis protocol was adapted from Jiang *et al.*, *Biochemistry* 44, 1163 (2005). Carazolol-bound β_2AR-T4L or wild-type β_2AR (each at 30 mg/ml) were diluted 10-fold into HLS buffer (see above) and TPCK-trypsin was added at a 1:1000 ratio (wt:wt). The digests were incubated at room temperature. At various time points, aliquots were removed and flash frozen on dry ice/ethanol. After the last aliquot was removed, all samples were thawed, and an equal volume of 10% SDS/PAGE loading buffer was added to each. Samples were then analyzed by electrophoresis on 12% polyacrylamide gels, followed by staining with Coomassie blue (**Figure 16**).

Comparing the stability of unliganded β_2AR and β_2AR-T4L

[00261] Unliganded β2AR365 and β2AR-T4L were each purified as described above for the bimane experiments. 200 µl 0.02 mg/ml receptor in HLS buffer was incubated at 37 °C on a heating block. At the time points indicated in **Figure 17**, samples were briefly spun and gently vortexed and 16.5 µl was removed and diluted 18.2-fold in HLS (300 µl total). Then 4 x 5 µl was removed for determination of total binding and 2 x 5 µl was removed for nonspecific binding. To measure soluble binding, 5 µl diluted receptor was added to 105 µl HLS (400-fold final dilution of receptor) containing 10 nM [3H]DHA ± 10 µM cold alprenolol. Reactions were incubated 30 min at RT, then on ice until processing. 100 µl of each reaction was applied to a 1 ml G50 column to separate protein from residual unbound [3H]DHA, and receptor was eluted using 1.1 ml ice-cold HLS. Bound [3H]DHA was quantified on a Beckman LS6000 scintillation counter.

Carazolol dissociation from the "wild-type" receptor β_2AR365

[00262] β2AR365 was purified with carazolol bound, according to the protocol described above for β2ART4L. Carazolol-bound receptor (at approximately 50 µM concentration) was dialyzed in the dark against 1L dialysis buffer (20mM HEPES pH7.5, 100mM NaCl, 0.1% dodecylmaltoside, 300 micromolar alprenolol) at room temperature with stirring. At indicated time points, two samples were removed from the parafilm-sealed open-ended dialysis chamber, diluted into fresh dialysis buffer, and carazolol emission spectra were obtained on a Spex FluoroMax spectrofluorometer (using excitation at 330 nm and emission from 335 to 400 nm). As internal standards for every time point, samples were removed for determination of protein concentration using the Bio-Rad Protein DC kit (**Figure 19**).

CAM and UCM mutants

[00263] The CAMs (constitutively active mutants) described in the literature that are the basis for **Figure 26A** and the associated discussion are: L124A (Tao, et al., Mol Endocrinol 14, 1272 (2000)), C116F (Zuscik, et

al., J Biol Chem 273, 3401 (1998)), D130A (Rasmussen et al., *Mol Pharmacol* 56, 175 (1999)), L272C (Jensen et al., *J Biol Chem* 276, 9279 (2001)), and C285T (Shi et al., *J Biol Chem* 277, 40989 (2002)). The UCMs (uncoupling mutations) from the literature that were used for **Figure 26C** are: D79N (Chung, et al., *J Biol Chem* 263, 4052 (1988)), F139A (Moro, et al., *J Biol Chem* 269, 6651 (1994)), T164I (Green, et al., *J Biol Chem* 268, 23116 (1993)), N318K (Strader et al., *Proc Natl Acad Sci U S A* 84, 4384 (1987)), N322A (Barak, et al., *Biochemistry* 34, 15407 (1995)), P323A (Barak, et al., *Biochemistry* 34, 15407 (1995)), Y326A (Barak, et al., *Biochemistry* 34, 15407 (1995)), L339A (Gabilondo et al., *Proc Natl Acad Sci U S A* 94, 12285 (1997)), and L340A (Gabilondo et al., *Proc Natl Acad Sci U S A* 94, 12285 (1997)).

Biochemical and structural analysis of β_2AR-T4L

[00264] The β_2AR fusion protein in which T4 Lysozyme replaces most of the third intracellular loop of the GPCR ("β_2AR-T4L") retains near-native pharmacologic properties. The β_2AR-T4L protein was crystallized in lipidic cubic phase, as described in the Examples above, and the resulting 2.4 Å resolution crystal structure reveals the interface between the receptor and the ligand carazolol, a partial inverse agonist. The efficacy of a ligand describes the effect of the ligand on the functional properties of a GPCR. For purposes of the Examples only, agonists are defined as ligands that fully activate the receptor; partial agonists induce submaximal activation even at saturating concentrations; inverse agonists inhibit basal receptor activity, and antagonists have no effect on basal activity, but competitively block access of other ligands. Carazolol, is defined as a partial inverse agonist because it suppresses only 50% of the basal activity of the β_2AR. Analysis of mutagenesis data in light of the structure clarifies the roles of different amino acids in inverse agonist binding, and implies that rearrangement of the binding pocket accompanies agonist binding. In addition, the structure reveals how mutations known to cause constitutive activity or uncoupling of agonist binding and G-protein activation are distributed between the ligand-binding pocket and the cytoplasmic surface of the protein, such that changes in side chains due to interaction with the ligand can be transmitted through the structure to the site of G protein interaction.

Cloning of β_2AR-T4L

[00265] DNA encoding the T4L protein (C54T, C97A) (Matsumura, et al., *Proc Natl Acad Sci U S A* 86, 6562 (1989)) was initially cloned into the human β2AR gene, guided by comparison of ICL3 length and sequence among class A GPCRs (Horn et al., *Nucleic Acids Res* 31, 294 (2003)): residues $234^{5.73}$- $259^{6.21}$ of the β2AR were replaced by residues 2-164 of T4L (construct "E3" in **Figure 21A**). In addition, the receptor was truncated at position 365, which aligns approximately with the position of the rhodopsin carboxyl terminus.

Although these modifications resulted in a receptor that was expressed efficiently in Sf9 cells, further optimization was carried out to reduce the length of the junction between the receptor and the T4L termini, as described in the methods above. Several candidate constructs are illustrated in **Figure 21A**, and selected immunofluorescence images of transfected, permeabilized HEK293 cells are shown in **Figure 21B**. Relative to the initial construct, we could remove three residues from the cytoplasmic end of helix V, three residues from the C-terminal end of T4L, and three residues from the N terminus of helix VI, all without losing significant cell-surface expression. The final construct used for crystallization trials ("β2AR -T4L") has residues $231^{5.70}$-$262^{6.24}$ of the β2AR replaced by amino acids 2-161 of T4L ("1D" in **Figure 21A**). Similar reduction of flexibility through minimization of linker length has been important in previous crystallization studies on soluble fusion proteins (Smyth, et al., *Protein Sci* 12, 1313 (2003)).

Functional properties of β_2AR-T4L

[00266] We measured saturation binding of [3H]DHA to the β2AR-T4L, as well as competition binding of the inverse agonist ICI-118,551 and several agonists (**Figure 22A**, **Figure 15** and **Table 6**). The results show that β2AR-T4L has wild-type affinity for the antagonist [3H]DHA and the inverse agonist ICI-118,551, whereas the affinity for both agonists (isoproterenol, epinephrine, formoterol) and a partial agonist (salbutamol) is two to three-fold higher relative to wild-type β2AR. Higher agonist binding affinity is a property associated with constitutively active mutants (CAMs) of GPCRs. CAMs of the β2AR also exhibit elevated basal, agonist-independent activation of Gs, and typically have lower expression levels and reduced stability (Gether et al., *J Biol Chem* 272, 2587 (1997); Rasmussen et al., *Mol Pharmacol* 56, 175 (1999)). β2ART4L exhibits binding properties of a CAM, but it expresses at levels exceeding 1 mg per liter of Sf9 cell culture, is more resistant to trypsin proteolysis than the wild-type β2AR (**Figure 16**), and retains binding activity in detergent at 37 °C as well as the wild-type receptor (**Figure 17**).

[00267] β2AR-T4L did not couple to Gs, as expected due to the replacement of ICL3 by T4L. To assess whether the fused protein alters receptor function at the level of its ability to undergo conformational changes, we used a covalently attached fluorescent probe as a reporter for ligand-induced structural changes. Fluorophores attached at $Cys265^{6.27}$, at the cytoplasmic end of helix VI, detect agonist-induced conformational changes that correlate with the efficacy of the agonist towards G protein activation (Ghanouni et al., *J Biol Chem* 276, 24433 (2001); Ghanouni, et al., *Proc Natl Acad Sci U S A* 98, 5997 (2001); Swaminath et al., *J Biol Chem* 279, 686 (2004); Swaminath et al., *J Biol Chem* 280, 22165 (2005)). Detergent-solubilized β2AR365 (wild-type receptor truncated at 365) and β2AR-T4L were each labeled with monobromobimane. Addition of

the agonist isoproterenol to purified β2AR365 induces a decrease in fluorescence intensity and a shift in λmax for the attached bimane probe (**Figure 22B** and **Table 7**). These changes in intensity and λmax are consistent with an agonist-induced increase in polarity around bimane. A smaller change is observed with the partial agonist salbutamol, while the inverse agonist ICI-118,551 had little effect. For the β2AR-T4L, there are subtle differences in the baseline spectrum of the bimane-labeled fusion protein, as might be expected if the environment around Cys265$^{6.27}$ is altered by T4L. However, the full agonist isoproterenol induces a qualitatively similar decrease in intensity and rightward shift in λmax. Thus the presence of the fused T4L does not prevent agonist-induced conformational changes. The partial agonist salbutamol induced larger responses in β2AR-T4L than were observed in wild-type β2AR, and there was a small increase in fluorescence in response to the inverse agonist ICI-118,551. These are properties observed in CAMs (Gether et al., *J Biol Chem* 272, 2587 (1997); Samama, et al., *J Biol Chem* 268, 4625 (1993)) and are consistent with the higher affinities for agonists and partial agonists exhibited by β2AR-T4L. Therefore, we conclude that the T4L fusion induces a partial constitutively active phenotype in the β2AR, likely caused by changes at the cytoplasmic ends of helices V and VI.

Comparison between β_2AR-T4L and β_2AR-Fab Structures

[00268] The β_2AR-T4L fusion strategy is validated by comparison of its structure to the structure of wild-type β_2AR complexed with a Fab that recognizes a three dimensional epitope consisting of the amino and carboxyl-terminal ends of ICL3, determined at an anisotropic resolution of 3.4 Å / 3.7 Å (Rasmussen *et al.*, *Nature,* 7168:355-6 (2007)). **Figure 23A** illustrates the similarity between the fusion and antibody complex approaches to β_2AR crystallization, in that both strategies rely on attachment (covalent or non-covalent, respectively) of a soluble protein partner between helices V and VI. A major difference between the two structures is that the extracellular loops and the carazolol ligand could not be modeled in the β_2AR-Fab complex, whereas these regions are resolved in the structure of β_2AR-T4L. Nonetheless, it is clear that the T4L insertion does not significantly alter the receptor. Superposition of the two structures (**Figure 18**) illustrates that the transmembrane helices of the receptor components are very similar (RMSD = 0.8 Å for 154 common modeled transmembrane Cα positions, versus 2.3 Å between β_2AR-T4L and the 154 equivalent residues in rhodopsin), especially when the modest resolution of the Fab complex is taken into account.

[00269] There is one significant difference between the Fab-complex and chimeric receptor structures that can be attributed to the presence of T4L. The cytoplasmic end of helix VI is pulled outward as a result of the fusion to the carboxyl terminus of T4L, which alters the packing of Phe264$^{6.26}$ at the end of helix VI

(**Figure 23B**). In the Fab-complex β_2AR, interactions between Phe264$^{6.26}$ and residues in helix V, helix VI, and ICL2 may be important in maintaining the β_2AR in the basal state. The loss of these packing interactions in β_2AR-T4L could contribute to the higher agonist binding affinity characteristic of a CAM.

[00270] An unexpected difference between the structure of rhodopsin and the β_2AR-T4L involves the sequence E/DRY found at the cytoplasmic end of helix III in 71% of class A GPCRs. In rhodopsin, Glu134$^{3.49}$ and Arg135$^{3.50}$ form a network of hydrogen bond and ionic interactions with Glu247$^{6.30}$ at the cytoplasmic end of helix VI. These interactions have been referred to as an "ionic lock" that stabilizes the inactive state of rhodopsin and other class A members (Ballesteros et al., *J Biol Chem* 276, 29171 (2001)). However, the arrangement of the homologous residues is significantly different in β_2AR-T4L: Arg131$^{3.50}$ interacts primarily with Asp130$^{3.49}$ and a sulfate ion rather than with Glu268$^{6.30}$, and the distance between helix III and helix VI is greater than in rhodopsin (**Figure 23C**). The fact that similar ionic lock structures were obtained using two different approaches suggests that a broken ionic lock is a genuine feature of the carazolol-bound state of the receptor.

Ligand binding to the β_2AR

[00271] The β_2AR-T4L fusion protein was purified and crystallized in complex with the inverse agonist carazolol. Carazolol stabilizes the β_2AR against extremes of pH and temperature, perhaps related to its unusually high binding affinity (K_d < 0.1 nM) and slow dissociation kinetics ($t_{1/2}$ ~30 h) (**Figure 19**). The interactions between carazolol and β_2AR-T4L are depicted schematically in **Figure 24**. The carbazole ring system is oriented roughly perpendicular to the plane of the membrane, and the alkylamine chain (atoms 15-22 in the model) is nearly parallel to the heterocycle (**Figure 25A-B**). As described in Example 3, above, carazolol was modeled into the electron density as the (S)-(-) isomer due to the higher affinity of this enantiomer, despite the fact that a racemic mixture of the ligand was used in crystallization. Asp113$^{3.32}$, Tyr316$^{7.43}$, and Asn312$^{7.39}$ present a constellation of polar functional groups to the alkylamine and alcohol moieties of the ligand, with Asp113$^{3.32}$ and Asn312$^{7.39}$ sidechains forming close contacts (<3Å) with O_{17} and N_{19} atoms of carazolol (**Figures 24 and 25A-B**). Asp113$^{3.32}$ was one of the first β_2AR residues shown to be important for ligand binding; notably the D113N mutation causes complete loss of detectable affinity for antagonists (Strader et al., *Proc Natl Acad Sci U S A* 84, 4384 (1987)) and a decrease in the potency of agonists towards cell-based G protein activation by over 4 orders of magnitude (Strader et al., *J Biol Chem* 263, 10267 (1988)). Likewise, mutations of Asn312$^{7.39}$ perturb β_2AR binding to agonists and antagonists: changes to nonpolar amino acids (Ala or Phe) reduce affinities to undetectable levels, while retention of a polar functionality (Thr or Gln) gives partial

affinity (Suryanarayana, et al., *Mol Pharmacol* 44, 111 (1993)). On the opposite end of the ligand near helix V, N_7 of the carbazole heterocycle forms a hydrogen bond with the side chain hydroxyl of Ser203$^{5.42}$. Interestingly, mutations of Ser203$^{5.42}$ specifically decrease β_2AR affinity towards catecholamine agonists and aryloxyalkylamine ligands with nitrogen-containing heterocycles such as pindolol (Liapakis et al., *J Biol Chem* 275, 37779 (2000)), and by implication carazolol. Thus, the polar interactions between carazolol and the receptor observed in the crystal structure agree with the known biochemical data. The contribution of Tyr316$^{7.43}$ to antagonist and agonist affinity remains to be tested; this residue is conserved as tyrosine in all sequenced adrenergic receptor genes (Horn et al., *Nucleic Acids Res* 31, 294 (2003)).

[00272] **Figure 25C** shows the tight packing between carazolol and surrounding amino acids that buries 790 Å^2 of surface area from solvent; specific contacts are depicted schematically in **Figure 24**. Notable among the hydrophobic residues contacting carazolol are Val114$^{3.33}$, Phe290$^{6.52}$, and Phe193$^{5.32}$. The side chain of Val114$^{3.33}$ from helix III makes multiple contacts with the C_8-C_{13} ring of the carbazole heterocycle, and Phe290$^{6.52}$ from helix VI forms an edge-to-face aromatic interaction with the same ring. As a result, these two amino acids form a hydrophobic "sandwich" with the portion of the aryl moiety that is common to many adrenergic antagonists. Mutation of Val114$^{3.33}$ to alanine was shown to decrease β_2AR affinity towards the antagonist alprenolol by an order of magnitude, as well as lowering affinity for the agonist epinephrine 300-fold (P. Chelikani et al., *Proc Natl Acad Sci U S A* 104, 7027 (2007)). Phe193$^{5.32}$ is different from other carazolol contact residues in that it is located on the ECL2, in the path of hormone accessibility to the binding pocket. This amino acid contributes more buried surface area than any other residue to the interface between β_2AR-T4L and carazolol (**see Table 8**). Therefore, Phe193$^{5.32}$ is likely to contribute significantly to the energy of β_2AR-carazolol complex formation, and the position of this residue on the extracellular side of the binding site may allow it to act as a gate that contributes to the unusually slow dissociation of the ligand (**Figure 19**).

Table 8.

β_2AR residue	Surface area buried (Å^2)
Trp109$^{3.28}$	21.4
Thr110$^{3.29}$	5.7
Asp113$^{3.32}$	19.3
Val114$^{3.33}$	25.5
Val117$^{3.36}$	8.5
Thr118$^{3.37}$	1.9
Phe193$^{5.32}$	51.2
Thr195$^{5.34}$	7.4
Tyr199$^{5.38}$	7.6
Ala200$^{5.39}$	10.0
Ser203$^{5.42}$	9.0
Ser204$^{5.43}$	4.6
Ser207$^{5.46}$	6.3
Trp286$^{6.48}$	3.1
Phe289$^{6.51}$	20.0
Phe290$^{6.52}$	19.0
Phe293$^{6.55}$	18.7
Tyr308$^{7.35}$	14.4
Asn312$^{7.39}$	22.5
Tyr316$^{7.43}$	6.5

Buried surface area contributions at the β_2AR-T4L/carazolol interface. Solvent accessible surface area calculations were done with the CNS software package (Brunger et al., *Acta Crystallogr D Biol Crystallogr* 54, 905 (1998)), using a probe radius of 1.4 Å. Buried surface area contributions of individual residues were determined by calculating solvent-accessible surface area per residue for the full β_2ART4L/carazolol model, and subtracting these numbers from the calculated values for the receptor model without carazolol.

[00273] Analysis of the binding pocket provides insights into the structural basis for pharmacologic selectivity between the β_2AR and closely related adrenergic receptors such as the β_1AR. The affinities of these two receptors for certain ligands, such as ICI-118,551, betaxolol and RO363 (Sugimoto et al., *J Pharmacol Exp Ther* 301, 51 (2002)), differ by up to 100-fold. Curiously, all of the amino acids in the carazolol binding pocket are conserved between the β_1AR and β_2AR (see **Figure 20**). The majority of the 94 amino acid differences between the β_1AR and β_2AR are found in the cytoplasmic and extracellular loops. While residues that differ in the transmembrane segments generally face the lipid bilayer, eight residues lie at the interface between helices and may influence helix packing. The structural basis for pharmacologic differences between β_1AR and β_2AR must, therefore, arise from amino acid differences in the entrance to the binding pocket or subtle differences in the packing of helices. Evidence for the latter comes from chimeric receptor studies (Frielle, et al., *Proc Natl Acad Sci U S A* 85, 9494 (1988)) in which successive exchange of helices between β_1AR and β_2ARs led to a gradual change in affinity for the β_2AR selective ICI-118,551 and the β_1AR selective betaxolol.

[00274] As discussed above, β_2AR-T4L shows CAM-like properties with respect to agonist binding affinities, suggesting that the unliganded β_2AR-T4L may exist in a more active conformation than the wild type-β_2AR. Nevertheless, as shown in **Figure 22B**, β_2AR-T4L can be stabilized in an inactive conformation by an inverse agonist. Since β_2AR-T4L was crystallized with bound carazolol, a partial inverse agonist, the structure most likely represents an inactive state. This is consistent with the similarity of the β_2AR-T4L and β_2AR-Fab5 carazolol-bound structures. To assess whether conformational changes are required to accommodate catecholamines, a model of isoproterenol was placed in the binding site such that common atoms (16-22 in **Figure 24**) were superimposed onto the analogous carazolol coordinates in the crystal structure (**Figure 25D**). Residues Ser204$^{5.43}$ and Ser207$^{5.46}$ are critical for catecholamine binding and activation of the β_2AR, with Ser204$^{5.43}$ hydrogen bonding to the meta-hydroxyl and Ser207$^{5.46}$ to the para-hydroxyl of the catechol ring, respectively (Strader, et al., *J Biol Chem* 264, 13572 (1989)). In our model, the catechol hydroxyls of isoproterenol face the appropriate serines on helix V, but the distances are too long for hydrogen bonding (6.8 Å from meta-hydroxyl oxygen to the sidechain oxygen of Ser204$^{5.43}$, 4.8 Å from the para-hydroxyl oxygen to the sidechain oxygen of Ser207$^{5.46}$). In addition, Asn293$^{6.55}$ and Tyr308$^{7.35}$, two residues expected to form selective interactions with agonists based on the literature (Wieland, et al., *Proc Natl Acad Sci U S A* 93, 9276 (1996); Kikkawa, et al., *Mol Pharmacol* 53, 128 (1998)), are too distant to form productive polar or hydrophobic contacts with the modeled isoproteronol molecule. These observations suggest that agonist binding requires

changes in the binding site relative to the carazolol-bound structure, unless common structural components of agonists and inverse agonists bind in a significantly different manner.

Structural insights into β_2AR activation

[00275] Analysis of mutations that affect β_2AR function provides insights into structural rearrangements that are likely to occur during receptor activation. **Figure 26A** illustrates the location of amino acids for which mutations lead to elevated basal, agonist-independent activity (constitutively active mutations, CAMs), as well as amino acids for which mutations impair agonist activation (uncoupling mutations, UCMs). Residues for which CAMs have been described are likely to be involved in interactions that maintain the receptor in the inactive conformation. These amino acids are centrally located on helices III and VI. In contrast, positions in which UCMs have been observed are likely to form intramolecular interactions that stabilize the active state. A cluster of UCMs are found at the cytoplasmic end of helix VII. Neither CAMs nor UCMs are directly involved in agonist binding. Although the CAMs and UCMs are not directly connected in sequence, it is evident from the structure that they are linked through packing interactions, such that movements in one will likely affect the packing of others. For example, **Figure 26A** (right panel) shows all amino acids with atoms within 4 Å of the two centrally located CAMs, Leu124$^{3.43}$ (Tao, et al., *Mol Endocrinol* 14, 1272 (2000)) and Leu272$^{6.34}$ (Jensen et al., *J Biol Chem* 276, 9279 (2001)). Several amino acids that pack against these CAMs also interact with one or more UCMs. Trp286$^{6.48}$ lies at the base of the binding pocket. It has been proposed that agonist binding leads to a change in the rotameric state of Trp286$^{6.48}$ with subsequent changes in the angle of the helical kink formed by Pro288$^{6.50}$ (Shi et al., *J Biol Chem* 277, 40989 (2002)). It is likely that an agonist-induced change in the rotameric state of Trp286$^{6.48}$ will be linked to changes in sidechains of CAMs and UCMs through packing interactions and propagated to the cytoplasmic ends of the helices and the associated intracellular loops that interact with G proteins and other signaling molecules.

[00276] In the structures of both rhodopsin and the β_2AR, a cluster of water molecules lies near the most highly conserved class A GPCR residues (**Figure 26B**). It has been proposed that these water molecules may play a role in the structural changes involved in receptor activation (Pardo, et al., *Chembiochem* 8, 19 (2007)). **Figure 26C** shows the network of potential hydrogen bonding interactions that link Trp286$^{6.48}$ with conserved amino acids extending to the cytoplasmic ends of helices. UCMs have been identified for three amino acids linked by this network - N322$^{7.49}$, P323$^{7.50}$, and Y326$^{7.53}$ (Barak et al., *Biochemistry* 34, 15407 (1995)). This relatively loose-packed, water filled region is likely to be important in allowing conformational transitions, as there will be fewer steric restraints to sidechain repacking.

[00277] All references, issued patents and patent applications cited within the body of the instant specification are hereby incorporated by reference in their entirety, for all purposes.

CLAIMS

1. A composition for lipidic cubic phase crystallization of a membrane protein, comprising

a polyethylene glycol or modified polyethylene glycol, wherein said polyethylene glycol has an average molecular weight of less than 2,000 g/mole;

0.1 to 0.5 M of a salt;

a host lipid, wherein said host lipid is a monoacylglycerol;

a lipid additive, wherein said lipid additive is a sterol, and wherein said sterol is present at a concentration of 5-20% w/w in said host lipid;

a buffer, wherein said buffer has a pH between 4.5-8.0; and

1 to 100 mg/ml of the membrane protein.

2. The composition of claim 1, wherein said salt is selected from the group consisting of a sulfate salt, a citrate salt, a malonate salt, a tartrate salt, an acetate salt, and a formate salt.

3. The composition of claim 2, wherein said salt is present at a concentration of 0.1 - 0.2 M.

4. The composition of claim 1, wherein said buffer is present at a concentration of 0.05 - 0.5 M.

5. The composition of claim 1, further comprising an alcohol present at a concentration of 1 - 10 %v/v.

6. The composition of claim 5, wherein said alcohol is present at a concentration of 5-7 %v/v.

7. The composition of claim 5, wherein said alcohol is a diol or triol.

8. The composition of claim 6, wherein said alcohol is 1,4-butanediol.

9. The composition of claim 6, wherein said alcohol is 2,6-hexanediol.

10. The composition of claim 1 wherein said host lipid is selected from the group consisting of monopalmitolein, monovaccenin and monoolein.

11. The composition of claim 10 wherein said host lipid is monoolein.

12. The composition of claim 1 wherein said protein to be crystallized in said composition is present at a concentration of 40-60 mg/mL.

13. The composition of claim 1 wherein said membrane protein is a G-protein coupled receptor.

14. The composition of claim 1, wherein said membrane protein is a β_2AR protein.

15. The composition of claim 13, wherein said G-protein coupled receptor is selected from the group consisting of $\beta_2AR^{(E122W)}$ and $\beta_2AR^{(E122W)}$-T4L.

16. The composition of claim 14 or 15, further comprising a ligand selected from the group consisting of carazolol, timolol, alprenolol, and clenbutorol.

17. A method of generating crystals of a membrane protein comprising:

mixing a lipid additive with a host lipid to form a lipid mixture, wherein said host lipid is a monoacylglycerol and wherein said lipid additive is a sterol, and wherein said lipid additive is 5 to 20 % w/w in said host lipid; and

combining said lipid mixture with a membrane protein solution to form a protein-lipid mixture, wherein said membrane protein is present in said protein-lipid mixture at a concentration of between 1 and 100 mg/ml; and

overlaying said protein-lipid mixture with a precipitant solution, wherein said precipitant solution comprises a polyethylene glycol or modified polyethylene glycol, wherein said polyethylene glycol has an

average molecular weight of less than 2,000 g/mole; 0.1 to 0.5 M of a salt; and a buffer, wherein said buffer has a pH between 4.5-8.0.

18. The method of claim 17, wherein the amount of said lipid additive is 8 to 10 % w/w in said host lipid.

19. A method of screening a crystal of a G-protein coupled receptor (GPCR) present in a lipidic cubic phase composition comprising:

preparing a lipidic cubic phase composition, wherein said composition comprises a GPCR protein; a polyethylene glycol or modified polyethylene glycol, wherein said polyethylene glycol has an average molecular weight of 300 to 2,000 g/mole; 0.1 to 0.5 M of a salt; a host lipid, wherein said host lipid is a monoacylglycerol; a lipid additive, wherein said lipid additive is a sterol and wherein said sterol is present at a concentration of 5-20% w/w in said host lipid; and a buffer, wherein said buffer has a pH between 4.5-8.0;

exposing said composition to a first X-ray beam and determining a change in direction or intensity of said first X-ray beam;

exposing said composition to a second beam and determining a change in direction or intensity of said second X-ray beam;

identifying an area where said GPCR crystal is present in said composition; and

exposing said identified area to at least a third X-ray beam.

20. The method of claim 19, wherein said first beam is a slitted 100 x 25 μm beam.

21. The method of claim 19, wherein said crystal is colorless.

22. The method of claim 19, wherein said GPCR crystal is $\beta_2AR^{(E122W)}$ or $\beta_2AR^{(E122W)}$-T4L.

Application number/numéro de demande: US2008/080847

Figures: 1A, 1B, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 18, 20, 21, 23, 25, 26

Pages:

Unscannable Item(s) received with this application.

To inquire is you can order a copy of the unscannable item(s),

Please visit the CIPO Website at HTTP://CIPO.GC.CA

Item(s) ne pouvant être balayés.

Documents reçus avec cette demande ne pouvant être balayés.

Pour vous renseigner si vous pouvez commander

une copie des items ne pouvant être balayes,

veuillez visiter le site web de l' OPIC au HTTP://CIPO.GC.CA

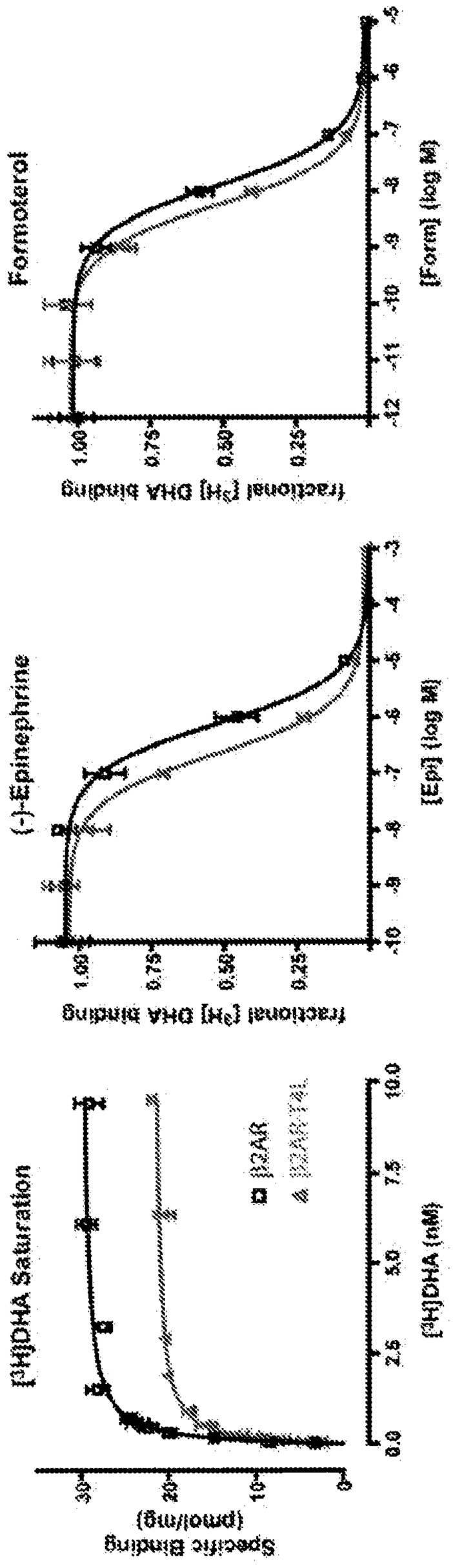
Formoterol
[Form] (log M)
fractional [3H] DHA binding
(-)-Epinephrine
[Epi] (log M)
fractional [3H] DHA binding
[3H]DHA Saturation
β2AR
BAR-/-
[3H]DHA (nM)
Specific Binding
(pmol/mg)

FIG. 15

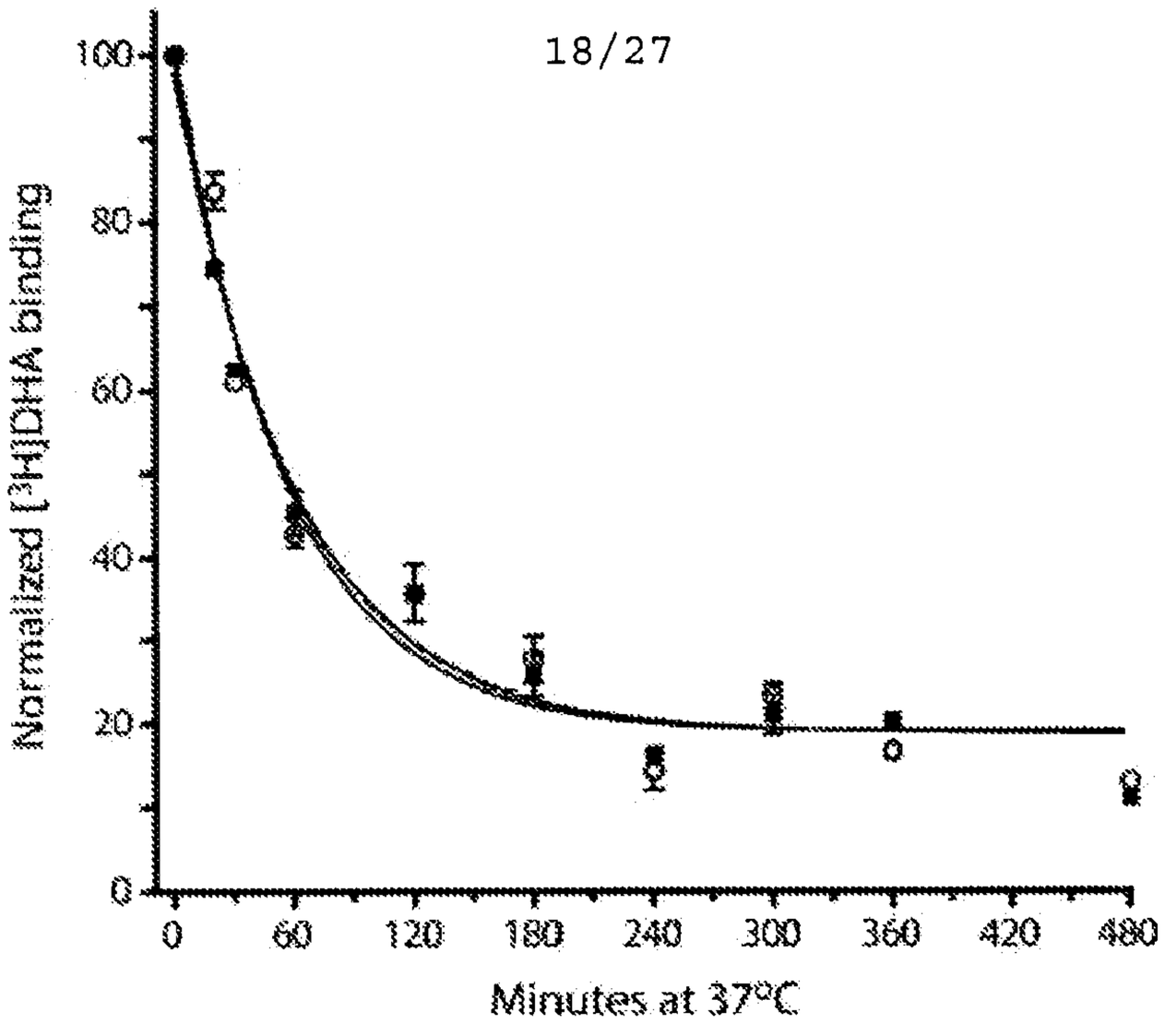
Normalized [3H]DHA binding
100
80
60
40
20
0
0
60
120
180
240
300
360
420
480
Minutes at 37°C

FIG. 17

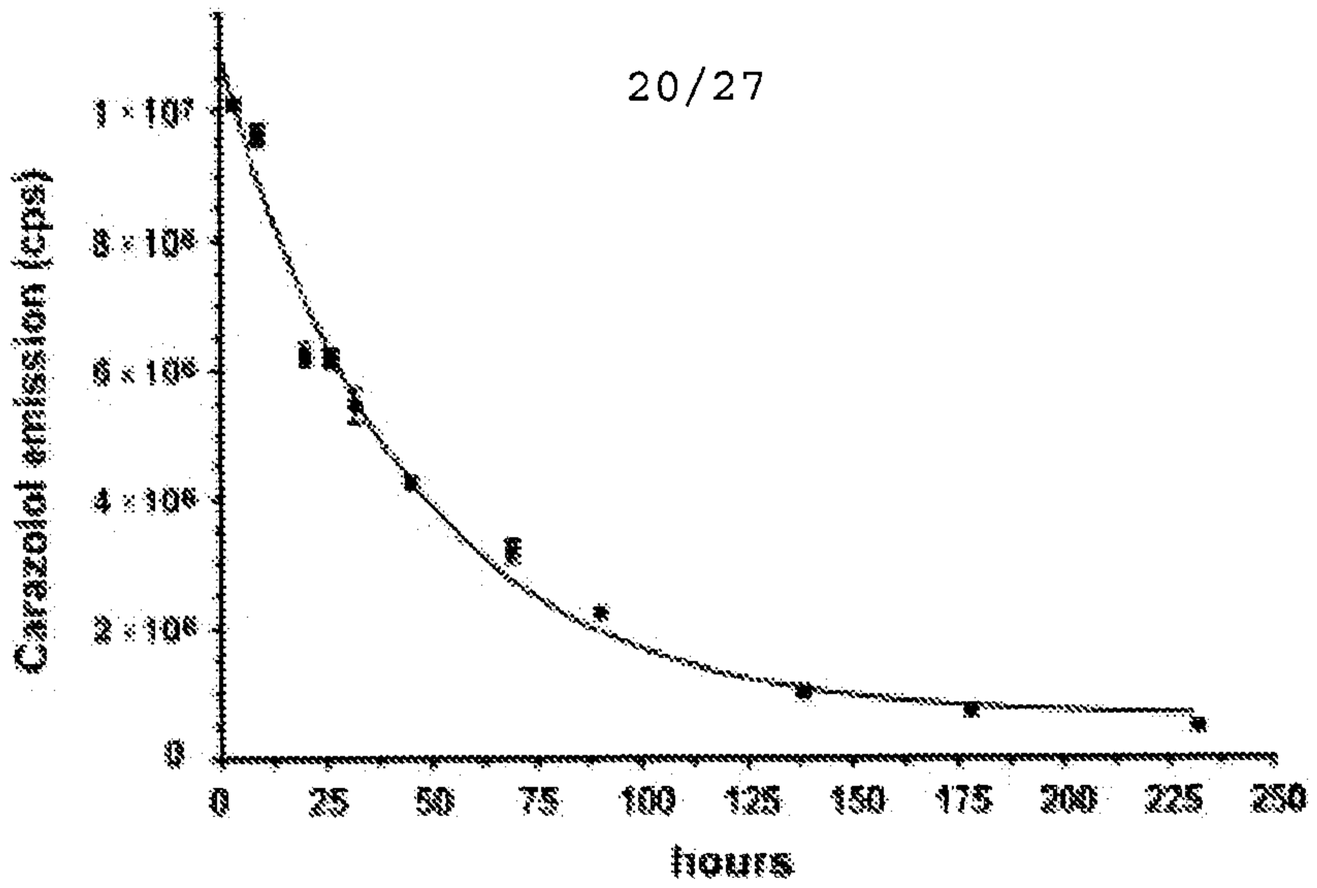
Carazolol emission (cps)
1 x 10⁸
8 x 10⁷
6 x 10⁷
4 x 10⁷
2 x 10⁷
0
0
25
50
75
100
125
150
175
200
225
250
hours

FIG. 19

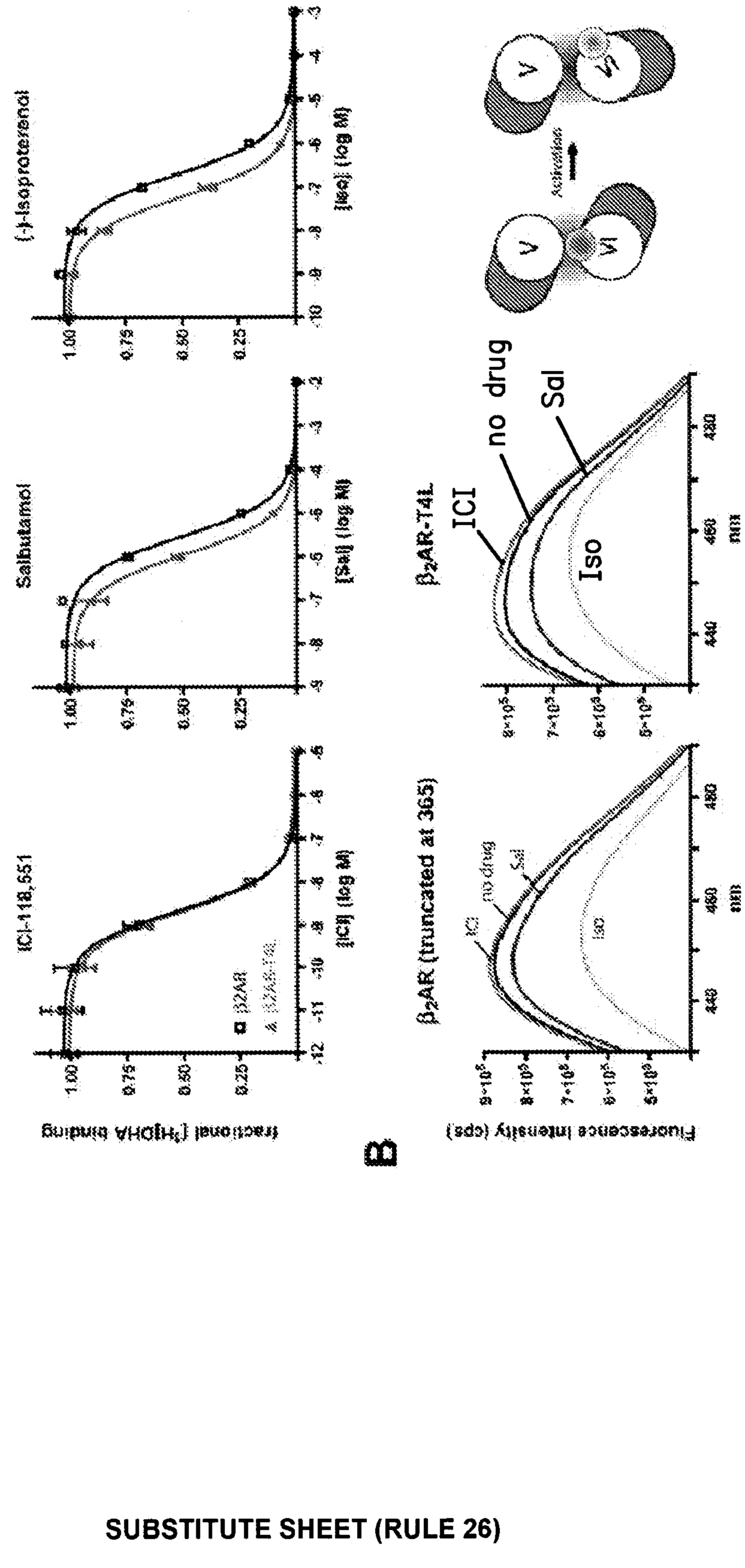
A
(-)-Isoproterenol
Salbutamol
ICI-118,551
fractional [3H]DHA binding
[ICI] (log M)
[Sal] (log M)
[Iso] (log M)
β2AR
β2AR-T4L
B
β2AR (truncated at 365)
ICI
no drug
Sal
Iso
Fluorescence intensity (cps)
nm

FIG. 22

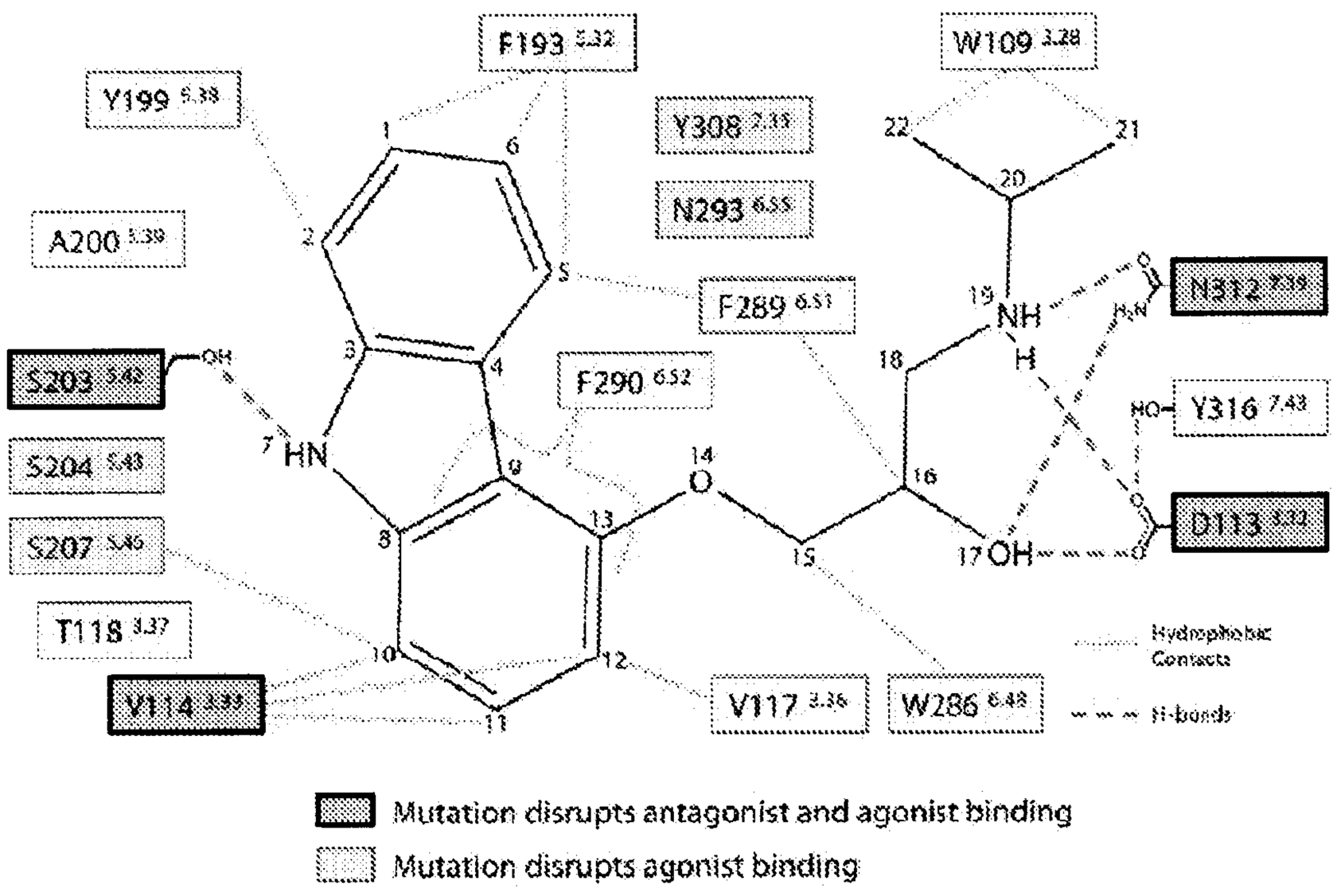
F193 5.32
W109 3.28
Y199 5.38
Y308 7.35
N293 6.55
A200 5.39
F289 6.51
N312 7.39
S203 5.42
F290 6.52
Y316 7.43
S204 5.43
D113 3.32
S207 5.46
T118 3.37
V114 3.33
V117 3.36
W286 6.48
Hydrophobic Contacts
H-bonds
Mutation disrupts antagonist and agonist binding
Mutation disrupts agonist binding

FIG. 24