

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 August 2007 (09.08.2007)

PCT

(10) International Publication Number
WO 2007/089899 A2

(51) International Patent Classification:
C07K 14/705 (2006.01)

(21) International Application Number:
PCT/US2007/002766

(22) International Filing Date: 1 February 2007 (01.02.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/764,676 1 February 2006 (01.02.2006) US

(71) Applicant (for all designated States except US): **OKLAHOMA MEDICAL RESEARCH FOUNDATION** [US/US]; 755 Research Parkway, Suite 410, Oklahoma City, OK 73104 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **ZHANG, Xuejun, C.** [US/US]; 1116 Charlton Road, Edmond, OK 73003 (US).

(74) Agents: **JENKINS, Kenneth, E.** et al.; Paul, Hastings Janofsky & Walker LLP, 3579 Valley Centre Drive, San Diego, CA 92130 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, 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, LV, LY, MA, MD, 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, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

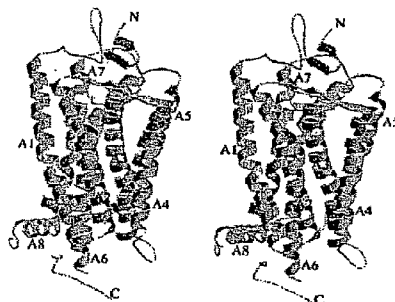
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, 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:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: WATER-SOLUBLE (G PROTEIN)-COUPLED RECEPTOR PROTEIN



(57) Abstract: The present invention provides novel water-soluble GPCR proteins and methods of making the same.



WO 2007/089899 A2

WATER-SOLUBLE (G PROTEIN)-COUPLED RECEPTOR PROTEIN

BACKGROUND OF THE INVENTION

[0001] The GPCR family is one of the largest and most diverse groups of proteins. For example, the human genome alone encodes ~950 GPCR proteins. There is a wealth of information about this protein family in the literature as well as online databases. GPCRs respond to a variety of different extracellular stimuli and activate G proteins on the cytosol side of the plasma membrane. The stimuli can be Ca^{2+} , small chemicals, hormones, peptides, proteases, and even photons. The activated G proteins, in turn, evoke down-stream intracellular responses. GPCRs are involved in many physiological processes and thus are attractive targets for pharmacological intervention for modifying these processes in normal and pathological states. It is estimated that GPCRs account for 30–50% of the current therapeutic targets. Mammalian GPCRs are commonly divided into a few distinct classes. Each represents heptahelical receptors of related gene and amino acid sequence; however, there is no clear sequence relationship across classes. The major classes of GPCRs include the rhodopsin family (also called class A), the glucagons receptor family (class B), and the metabotropic glutamate receptor family (class C). The rhodopsin family is by far the largest and most studied GPCR subfamily. It constitutes ~90% of all GPCRs.

[0002] Currently, the only available experimental 3D structure of the GPCR family is that of bovine rhodopsin (Fig. 1). The structure reveals that its seven-helix ($\alpha 1$ – $\alpha 7$) bundle TM region, which spans ~40 Å in length along the axis of the helical bundle, is stabilized by a number of inter-helical hydrogen bonds and hydrophobic interactions, with most of them being mediated by highly conserved residues in the rhodopsin subfamily. Primary sequence analysis suggests that these structural features are likely to be shared by other members of the class A GPCRs and even members of other classes. The implication is that structural information obtained from a typical GPCR protein regarding the activation mechanism is likely applicable to the whole GPCR family. Many GPCRs function as either homodimers or hetero-dimers on the plasma membrane. In some cases, a receptor dimer is shown to be the functional unit that interacts with one heterotrimeric G-protein to activate it. For example, dopamine D2 receptor monomers can be oxidatively or chemically crosslinked via Cys168 (equivalent to Ala169 in bovine rhodopsin) and other engineered Cys residues at a symmetrical interface in helix $\alpha 4$. Whether oligomerization is a general requirement for GPCR activation remains to be investigated.

[0003] While it is widely recognized that understanding the specificities between many GPCRs and their cognate, physiological ligands, agonist, and/or antagonist are of extreme importance for development of new therapeutic methods and drugs, a generally applicable approach to study GPCR structures at high resolution remains to be developed. Furthermore, many fundamental questions remain open regarding the GPCR structure-function relationship, including the mechanism of GPCR activation, and the structural basis of the GPCR-G protein interaction.

[0004] Human PAR1 (425 amino acid residues) was identified as a thrombin receptor in early 1990s by Coughlin and colleagues. Subsequently, three other protease-activated receptors, PARs 2-4, were characterized. All of them belong to the class A GPCR family thus are almost certain to share similar folding with rhodopsin. PAR1 comprises an N-terminal peptide (which functions in the proteolytic activation), seven-transmembrane helix domain, and cytosolic tail containing the sorting signal targeting to lysosome during signal desensitization. Fig. 2 shows sequence alignment between human PAR1-4 and bovine rhodopsin whose crystal structure has been reported. The alignment shows that while the TM helices and most their connecting loops are conserved in their length and folding, the N-terminus of PAR1 is 70-residues longer than that in rhodopsin. This is consistent with the fact that the two GPCR proteins are activated by completely different mechanisms (photon absorption vs. proteolytic cleavage).

[0005] Among GPCRs, the protease-activation mechanism is unique to PARs. After proteolytic cleavage by a specific serine protease (e.g. thrombin for PAR1), the new N-terminus serves as a tethered ligand for receptor activation. Conceptually, the activation mechanism is similar to that of many serine proteases in which formation of the active site is triggered by the insertion of a proteolytically produced nascent peptide N-terminus. In human PAR1, the activation cleavage site is between Arg41 and Ser42. Peptides that mimic the nascent N-terminus are able to activate the receptor in the absence of a protease. The ligand binding site is contributed partly by the N-terminal peptide (between the cleavage site and $\alpha 1$) and the second extracellular loop (i.e. EC2 between helices $\alpha 4$ and $\alpha 5$). PAR1-mediated thrombin functions have been directly linked to many physiological and pathological processes. The PAR1 activator, thrombin, is a key serine protease in the coagulation cascade. Depending on its interactions with other proteins, thrombin can exhibit either pro- or anti-coagulation activity. Many of its effects are consistent with a primary role in vessel wound

healing and revascularization. Such a role not only includes clot formation, but also has effects upon a multitude of cell types involved in the systemic response to vascular damage.

[0006] The most thoroughly studied thrombin–PAR1 interaction is that of platelet aggregation, in which thrombin causes platelets to change shape, adhere to each other, and secrete the contents of their storage granules. As PAR1 plays important roles in tissue remodeling, its failure has been linked with many diseases. PAR1 is highly expressed in tumor cells, invasive cell lines, and in breast carcinoma specimens. Anti-sense cDNA directed against PAR1 is shown to be able to inhibit breast carcinoma invasion in a model system. PAR1 expression is up-regulated in prostate carcinoma compared with normal prostate tissue and is hypothesized to play a central role in prostate tumorigenesis. Thrombin is also shown to activate astrocytes through PAR1, and particularly microglia in propagating local inflammation and producing potential neuro-toxic side-effects. In Alzheimer's disease, the levels of an endogenous inhibitor of thrombin, protease nexin-1, is shown to be reduced.

[0007] Other PAR proteins, e.g. PAR2, have also been implicated in neurological disorders and inflammatory diseases. Given the fact that thrombin also regulates coagulation, specific therapeutic regulation of PAR1 seems to represent an adjunct or alternative approach to thrombin inhibition in modulating downstream cellular functions. For example, a PAR1 antagonist has an advantage over a direct thrombin inhibitor since it does not inhibit enzymatic action of thrombin in the coagulation cascade. Thus, the side effect of excessive bleeding can be eliminated.

[0008] Like other GPCRs, activated PARs convey information to intracellular heterotrimeric G-proteins. The G-protein is a heterotrimer comprised of a single α (40–50 kDa), β (~35 kDa), and γ (~10 kDa) subunit. The α -subunit ($G\alpha$) is a GTPase which is structurally related to Ras-like small GTPases. The $G\alpha$ subunit is composed of two domains: a nucleotide binding domain with high structural homology to Ras-like small GTPases, and an all- α -helical domain as an insertion between the helix α_1 and strand β_2 of the core Ras-like domain. There are three flexible regions in a $G\alpha$ nucleotide binding domain, designated as switch-I, -II, and -III. They change conformation in response to GTP binding and hydrolysis. In addition, the N-terminal region is disordered in the $G\alpha$ crystal structure but becomes ordered when interacting with a $G\beta\gamma$ complex. In their native forms, $G\alpha$ proteins are usually N-terminally modified by the covalent attachment of the fatty acids myristate and/or palmitate. These posttranslational modifications in a $G\alpha$ subunit affect its targeting to specific

cellular membrane domains (e.g. raft domains) and thus regulate its interactions with other proteins such as adenylyl cyclase, $G\beta\gamma$ complex, and GPCRs.

[0009] G-proteins that interact with GPCRs are commonly grouped into four subfamilies, namely Gs, Gi, Gq, and G12, on the basis of their $G\alpha$ domain amino acid sequences and

5 functions. The Gs and Gi proteins stimulate and inhibit cAMP formation, respectively; members of the Gq family stimulate β isoforms of phospholipase C (PLC); and members of the G12 family regulate the platelet actin cytoskeleton. PAR1 has been shown to couple to multiple heterotrimeric G-proteins, including the Gi, Gq, and G12 subfamilies. It serves as the guanine-nucleotide exchange factor (GEF) for these G-proteins and facilitates GDP
10 dissociation and GTP reloading in $G\alpha$. The GTP-bound $G\alpha$ dissociates from its partner, the tightly-bound $\beta\gamma$ -subunits ($G\beta\gamma$), allowing both $G\alpha$ and $G\beta\gamma$ to interact with down stream effectors. Fatty acylation of $G\alpha$ and prenylation of $G\beta\gamma$ cause them to remain associated with the plasma membrane regardless the nucleotide-binding states. A direct interaction between PAR1 and Gq (and Gi2) has been demonstrated by immuno-precipitation. Mediated by
15 PARs, thrombin has been shown to have GEF activity for members of the G12 subfamily. Moreover, thrombin has at least two cellular effects: (1) it inhibits cAMP signaling; and (2) it stimulates PLCcatalyzed hydrolysis of poly-phosphoinositides, resulting in the formation of InsP3, mobilization of intracellular Ca^{2+} , and generation of diacylglycerol (the endogenous activator of protein kinase C).

20 [0010] Distinct cytosolic domains of PAR1 couple to different G-proteins and induce different intracellular signals. For example, the third intracellular domain (i.e. the IC3 loop connecting helices $\alpha 5$ and $\alpha 6$) appears to couple to a Gi-protein and activate mitogen-activated protein (MAP) kinases in CHO (Chinese hamster ovary) cells. Similarly, the C-terminal tail of PAR1 is a critical site for PLC activation via a Gq-protein. These two regions
25 are the least conserved in class A GPCR subfamily. Disrupting either of these interactions in CHO cells interferes with the thrombin-induced cell proliferation. Associated with the unique irreversible proteolytic activation, a PAR protein requires a special desensitization process, including steps of acute shutoff and internalization, to terminate the effects of its non-diffusible ligand. The acute shutoff of PAR1 signal is usually performed via the
30 phosphorylation of the cytoplasmic C-terminus, which contains consensus GRK (GPCR kinase) phosphorylation sites.

[0011] Phosphorylation within such cytosolic regions may cause dissociation of the tethered ligand from the receptor activation site on the extracellular side or simply disrupt the G-protein binding. Furthermore, extracellular proteolytic cleavage may also terminate the PAR1 signal. For example, the key serine protease in fibrinolysis, plasmin, has been shown to desensitize thrombin-dependent Ca^{2+} signaling through cleavage at sites distal to PAR1 Arg41. Desensitized PAR1 proteins are further internalized into lysosome for degradation.

[0012] Because PAR-mediated signal transduction pathways play important roles in many physiological and pathological processes, development of PAR specific inhibitors are of great interest. It requires detailed structural information on PAR activation and desensitization.

Although molecular modeling could provide such information to some extent, detailed structural studies are necessary for more insightful analyses such as structure-based drug design, for example, of small molecules of agonists and antagonists.

[0013] PAR1 (protease activated receptor 1) belongs to the guanine nucleotide-binding protein (G protein)-coupled receptor (GPCR) family of membrane proteins. Thrombin-mediated proteolysis activates its extracellular domain thus inducing G-protein activation on the intracellular side of the plasma membrane and in turn activating down-stream effectors. Detailed biochemistry and cell-biology studies on PAR1 are hindered by lack of reliable three-dimensional information about this membrane protein. Currently, the only available crystal structure of the GPCR family is that of rhodopsin in its inactive form, which shares less than 20% sequence identity with PAR1.

[0014] The present invention provides a solution to these and other needs in the art.

BRIEF SUMMARY OF THE INVENTION

[0015] Interaction of an activated GPCR with a $\text{G}\alpha$ protein is an essential step for signal transduction across the membrane. Therefore, elucidation of such a complex is of great interest to the field. By providing novel water-soluble GPCR proteins, the present invention provides solutions to these and other needs in the art. In contrast to detergent solubilized membrane proteins, each of the solubilized GPCR variants provides significantly more surface area that may form specific interactions during crystal packing.

[0016] In one aspect, the present invention provides a method of making a water-soluble (G Protein)-Coupled Receptor (GPCR) Protein. The method includes (a) performing a sequence alignment between a subject GPCR protein and a control GPCR protein thereby identifying a

set of helical transmembrane amino acid residues forming five transmembrane helices of the subject GPCR protein. In step (b), the solvent accessibility of amino acid residues within the set of helical transmembrane amino acid residues is assessed. Step (c) involves selecting a hydrophobic helical transmembrane amino acid residue from at least two transmembrane helices of the subject GPCR protein. Finally, in step (d), the two hydrophobic helical transmembrane amino acid residues are independently replaced with two hydrophilic amino acid residues by performing site directed mutagenesis of the subject GPCR protein, thereby making the water-soluble GPCR protein.

[0017] In another aspect, the present invention provides a method of making a water-soluble (G Protein)-Coupled Receptor (GPCR) Protein. The method includes step (a) in which a sequence alignment is performed between a subject GPCR protein and a control GPCR protein thereby identifying a set of solvent-exposed hydrophobic helical transmembrane amino acids. In step (b), five solvent-exposed hydrophobic helical transmembrane amino acid residues within the set of solvent-exposed helical transmembrane amino acid residues are replaced with five independently selected hydrophobic amino acid residues, thereby making the water-soluble GPCR. Each of the five solvent-exposed hydrophobic helical transmembrane amino acid residues form part of a different transmembrane helix within the subject GPCR.

[0018] In another aspect, the present invention provides a water-soluble GPCR protein produced by the methods of the present invention described above.

[0019] In another aspect, the present invention provides a water-soluble PAR-1 protein comprising at least 11 amino acid substitutions. The amino acid substitutions include replacing a hydrophobic amino acid with a hydrophilic amino acid residue. The hydrophobic amino acids may be selected from Phe104, Gly111, Val114, Val 115, Leu117, Leu119, Ile121, Ile128, Val149, Leu150, Phe157, Phe177, Ile198, Phe221, Leu224, Ala225, Ala228, Leu229, Ile231, Val235, Ala276, Phe280, Val281, Ile284, Val288, Val291, Leu355, Val359, Ile362, Ile366, and Leu369.

BRIEF DESCRIPTION OF THE DRAWINGS

[0020] Figure 1. Stereo ribbon diagram of bovine rhodopsin crystal structure [Protein Data Bank (PDB) file 1L9H]. The extracellular region is on top, and cytosolic region is at the

bottom. The helices are labeled as A1–A8. Among them, A1–A7 are TM helices. The amino and carboxyl termini are labeled as N and C, respectively.

[0021] Figure 2. Amino acid sequence alignment between bovine rhodopsin (GenBank accession #: P02699) and human PAR1–4 (AAA36742, P55085, O00254, and Q96RI0). The helical secondary structures (A1–A8) of bovine rhodopsin based on its crystal structure (PDB file 1L9H) are shown on the top. Intracellular and extracellular loops are labeled as IC1–3 and EC1–3, respectively. Selected residue numbers of rhodopsin and PAR1 are shown above and below the sequences, respectively. Residues identical to that of rhodopsin are highlighted. The percentage “solvent” accessibility for each residue calculated from the rhodopsin model is shown at the bottom, with black as completely buried and white as maximally accessible. The most conserved residue across the GPCR superfamily in each TM helix is boxed. Collectively, they are used as the registration positions in the BW numbering system Surratt, C.K. and W.R. Adams, *G protein-coupled receptor structural motifs: relevance to the opioid receptors*. Curr Top Med Chem, 2005. 5(3): p. 315-24. This figure was drawn with the programs AlScript Barton, G.J., *ALSCRIPT a tool to format multiple sequence alignments*. Protein Engineering, 1993. 6(1): p. 37-40. & EdPDB Zhang, X. and B.W. Matthews, *EDPDB: A multifunctional tool for protein structure analysis*. Journal of Applied Crystallography, 1995. 28: p. 624-630.

[0022] Figure 3. Schematic diagram of restriction site distribution in a WT silent-mutation construct (residues 23–425). This figure was output from the program VectorNTI.

[0023] Figure 4. Expression of PAR1 variants in a cell-free *E. coli* based *in vitro* translation system. PAR1 variants M1 (23–425) and M27 (81–425) were expressed as both MBP and GST-fusion proteins concomitantly with His6-tag in the presence of 0.2% Brij35. The samples were analyzed using 12% SDS-PAGE followed by western blot against anti-His6. Lanes are labeled as total reaction mixture (T), soluble fraction (S), and pellet (P). Samples of negative controls (empty vectors) are shown as total reaction.

[0024] Figure 5. High affinity protein-fragment complementation assay (PCA) based on α -complementation of β -galactosidase (β -Gal). MBP- α fragment fusion (labeled as +) results in blue colonies in the IPTG/X-Gal plate, and two M1-PAR1- α fragment fusion clones (#6 and #7) result in white colonies. The negative control vector contains MBP but not the α -fragment of β -Gal.

[0025] Figure 6. Schematic diagram of a high affinity PCA experiment.

[0026] Figure 7. Schematic diagram of a low affinity PCA experiment.

DETAILED DESCRIPTION OF THE INVENTION

A. Definitions

5 [0027] "Peptide" refers to a polymer in which the monomers are amino acids and are joined together through amide bonds, alternatively referred to as a "polypeptide." The terms "peptide" and "polypeptide" encompass proteins. Unnatural amino acids, for example, β -alanine, phenylglycine and homoarginine are also included under this definition. Amino acids that are not gene-encoded may also be used in the present invention. Furthermore, amino
10 acids that have been modified to include reactive groups may also be used in the invention. All of the amino acids used in the present invention may be either the D - or L -isomer. The L -isomers are generally preferred. In addition, other peptidomimetics are also useful in the present invention. For a general review, *see*, Spatola, A. F., in CHEMISTRY AND BIOCHEMISTRY OF AMINO ACIDS, PEPTIDES AND PROTEINS, B. Weinstein, eds., Marcel
15 Dekker, New York, p. 267 (1983).

[0028] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ -
20 carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical
25 structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

[0029] The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by
30 the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example,

recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

5 [0030] An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

B. Description of the Embodiments

10 [0031] In one aspect, the present invention provides a method of making a water-soluble (G Protein)-Coupled Receptor (GPCR) Protein. The method includes (a) performing a sequence alignment between a subject GPCR protein and a control GPCR protein thereby identifying a set of helical transmembrane amino acid residues forming five transmembrane helices of the subject GPCR protein. In some embodiments, a set of helical transmembrane amino acid
15 residues forming six or seven transmembrane helices is identified. The three-dimensional structure of the control GPCR protein is known thereby facilitating the identification of the set of helical transmembrane amino acid residues. In step (b), the solvent accessibility of each amino acid in the set of helical transmembrane amino acid residues is assessed.

[0032] Step (c) involves selecting a hydrophobic helical transmembrane amino acid residue
20 from at least two transmembrane helices of the subject GPCR protein. Thus, at least two hydrophobic helical transmembrane amino acid residues are selected. The selecting is based at least in part on the solvent accessibility assessment of step (b). The hydrophobic helical transmembrane amino acid residue forms part of the set of helical transmembrane amino acid residues. In some embodiments, a hydrophobic helical transmembrane amino acid residue
25 from at least three transmembrane helices are selected. Thus, in this embodiment, at least three hydrophobic helical transmembrane amino acid residues are selected. In some embodiments, a hydrophobic helical transmembrane amino acid residue from at least four transmembrane helices are selected. In some embodiments, a hydrophobic helical transmembrane amino acid residue from at least five transmembrane helices are selected. In
30 some embodiments, a hydrophobic helical transmembrane amino acid residue from at least six transmembrane helices are selected. In other embodiments, where two, three, four, or five

hydrophobic helical transmembrane amino acid residues are selected, the two, three, four, or five hydrophobic helical transmembrane amino acid residues form part of two, three, four, or five, respectively, different transmembrane helices of the subject GPCR protein, and the two, three, four, or five hydrophobic helical transmembrane amino acid residues form part of the set of helical transmembrane amino acid residues.

[0033] In some embodiments, step (c) includes selecting ten hydrophobic helical transmembrane amino acid residues forming part of five different transmembrane helices. Step (c) may include selecting six hydrophobic helical transmembrane amino acid residues forming part of six different transmembrane helices of the subject GPCR protein. The selecting is based at least in part on the assessment in step (b), and the six hydrophobic helical transmembrane amino acid residues form part of the set of helical transmembrane amino acid residues. In other embodiments, step (c) includes selecting 20 to 40 hydrophobic helical transmembrane amino acid residues. The 20 to 40 hydrophobic helical transmembrane amino acid residues form part of the set of helical transmembrane amino acid residues. In still other embodiments, step (c) includes selecting 30 to 35 hydrophobic helical transmembrane amino acid. The 30 to 35 hydrophobic helical transmembrane amino acid residues form part of the set of helical transmembrane amino acid residues.

[0034] Finally, in step (d), the two, three, four, five, or six hydrophobic helical transmembrane amino acid residues are independently replaced with two, three, four, five, or six hydrophilic amino acid residues, respectively, by performing site directed mutagenesis of the subject GPCR protein, thereby making the water-soluble GPCR protein. Because the hydrophobic helical transmembrane amino acid residues are independently replaced with hydrophilic amino acid residues, each of the hydrophilic amino acid residues are optionally the same or different.

[0035] In another aspect, the present invention provides a method of making a water-soluble (G Protein)-Coupled Receptor (GPCR) Protein. The method includes step (a) in which a sequence alignment is performed between a subject GPCR protein and a control GPCR protein thereby identifying a set of solvent-exposed hydrophobic helical transmembrane amino acids. In step (b), five solvent-exposed hydrophobic helical transmembrane amino acid residues within the set of solvent-exposed helical transmembrane amino acid residues are replaced with five independently selected hydrophobic amino acid residues, thereby making the water-soluble GPCR. Each of the five solvent-exposed

hydrophobic helical transmembrane amino acid residues form part of a different transmembrane helix within the subject GPCR. Thus, where five solvent-exposed hydrophobic helical transmembrane amino acid residues are replaced, five of the seven transmembrane helices of the subject GPCR contain hydrophobic helical transmembrane amino acid residue replacements.

[0036] In some embodiments, step (b) include selecting ten solvent-exposed hydrophobic helical transmembrane amino acid residues within the set of solvent-exposed helical transmembrane amino acid residues with ten independently selected hydrophobic amino acid residues. At least five of the ten solvent-exposed hydrophobic helical transmembrane amino acid residues form part of a different transmembrane helix within the subject GPCR. Step (b) may include selecting six solvent-exposed hydrophobic helical transmembrane amino acid residues within the set of solvent-exposed helical transmembrane amino acid residues with six independently selected hydrophobic amino acid residues. The six solvent-exposed hydrophobic helical transmembrane amino acid residues each form part of a different transmembrane helix within the subject GPCR. In other embodiments, step (b) includes selecting from 20 to 40 solvent-exposed hydrophobic helical transmembrane amino acid residues within the set of solvent-exposed helical transmembrane amino acid residues with 20 to 40 independently selected hydrophobic amino acid residues. At least five of the 20 to 40 solvent-exposed hydrophobic helical transmembrane amino acid residues form part of a different transmembrane helix within the subject GPCR. In still other embodiments, step (b) includes selecting from 30 to 35 solvent-exposed hydrophobic helical transmembrane amino acid residues within the set of solvent-exposed helical transmembrane amino acid residues with 30 to 35 independently selected hydrophobic amino acid residues. At least five of the 20 to 40 solvent-exposed hydrophobic helical transmembrane amino acid residues form part of a different transmembrane helix within the subject GPCR.

[0037] A "water-soluble GPCR," as used herein, refers to a GPCR protein variant that is more soluble in its folded form in an aqueous solution than the corresponding folded native, or wild-type GPCR protein. In some embodiments, the water-soluble GPCR is at least partially soluble in an aqueous solution without the use of detergents. In some embodiments, the water-soluble GPCR is completely soluble in an aqueous solution, where detergents are absent from the aqueous solution.

[0038] A "control GPCR protein" is a GPCR protein whose primary sequence is known and whose three-dimensional structure has been identified using generally known and accepted methods (i.e. NMR analysis and/or X-ray crystallography). A "subject GPCR," as used herein, refers to a GPCR protein that is subjected to the methods of the present invention described above.

[0039] A "helical transmembrane amino acid residue," as used herein, refers to an amino acid residue of a GPCR protein that forms part of one of the seven transmembrane helices of the GPCR protein. A "hydrophobic helical transmembrane amino acid residue," as used herein, refers to a helical transmembrane amino acid residue having a non-polar side chain that dissolves poorly in water. Examples of hydrophobic helical transmembrane amino acid residues may include G, A, V, L, I, M, P, F, and W. A "hydrophilic amino acid residue," as used herein, is an amino acid residue containing a side chain that is not hydrophobic (e.g. an amino acid residue with a charged or polar side chain). Examples of hydrophilic amino acid residues include S, T, N, Q, Y, C, K, H, D and E. A "solvent-exposed helical transmembrane amino acid residue," as used herein refers to a hydrophobic helical transmembrane amino acid residue that has been identified as a being surface amino acid within the helical transmembrane regions of the subject GPCR using the methods disclosed herein.

[0040] In some embodiments, at least one hydrophobic helical transmembrane amino acid residues are selected from at least two, three, four, five, six, or seven transmembrane helices of the subject GPCR protein. In other embodiments, at least two hydrophobic helical transmembrane amino acid residues are selected from at least two, three, four, five, six, or seven transmembrane helices of the subject GPCR protein. The total number of the hydrophobic helical transmembrane amino acid residues or the solvent-exposed helical transmembrane amino acid residues may be from 20 to 40, or from 30 to 35. The hydrophobic helical transmembrane amino acid residues or the solvent-exposed hydrophobic helical transmembrane amino acid residues may be selected from the middle three or four helical turns of the transmembrane helix.

[0041] In some embodiments, the subject GPCR protein is a class A GPCR protein, such as PAR1. The control GPCR protein may be rhodopsin.

[0042] In some embodiments, the methods further include stabilizing the water-soluble GPCR by engineering one or more inter-helix bonds (e.g. ionic bonds such as salt bridges,

disulfide bonds, and/or hydrogen bonds) between two or more of the seven transmembrane helices of the water-soluble GPCR protein. In other embodiments, the methods further includes assessing the functionality of the water-soluble GPCR using a protein-fragment complementation assay. In some related embodiments, the methods further include
5 improving the functionality of the water-soluble GPCR by restoring at least one of the hydrophilic amino acid residues with the wild type hydrophobic helical transmembrane amino acid residue present in the subject GPCR protein. Thus, a functionally optimized water-soluble GPCR is produced.

10 [0043] In another aspect, the present invention provides a water-soluble GPCR protein produced by the methods of the present invention described above.

[0044] In another aspect, the present invention provides a water-soluble PAR-1 protein having at least 11 amino acid substitutions. Each substitution replaces a hydrophobic amino acid with a hydrophilic amino acid residue. The hydrophobic amino acids to be replaced may be selected from Phe104, Gly111, Val114, Val 115, Leu117, Leu119, Ile121, Ile128, Val149,
15 Leu150, Phe157, Phe177, Ile198, Phe221, Leu224, Ala225, Ala228, Leu229, Ile231, Val235, Ala276, Phe280, Val281, Ile284, Val288, Val291, Leu355, Val359, Ile362, Ile366, and Leu369. The numbering system for the above referenced amino acids is consistent with the sequence of PAR1 as set forth in Figure 2.

[0045] In some embodiments, the water-soluble PAR-1 protein includes at least or
20 approximately 20 of the amino acid substitutions. In other embodiments, the water-soluble PAR-1 protein includes at least or approximately 30 of the amino acid substitutions. The water-soluble PAR-1 protein may further include one or more engineered inter-helix bonds.

[0046] In general, the seven-TM helix bundle of GPCRs expose a large hydrophobic surface area that is suitable for membrane insertion but makes the protein incompatible with
25 water. By inverting the surface property of a GPCR from hydrophobic to hydrophilic without drastically disrupting the overall folding, systematic hydrophobic-to-polar/charged (H-P) amino acid residue substitution is employed. The result is a model system for studying GPCR in solution that in many ways represents the intact protein in the membrane. This model system is based in part on the observation that soluble globular proteins usually have
30 significant amount of polar residues covering their surface and mostly hydrophobic residues forming the protein core.

[0047] The fact that the seven-TM helix bundle structure of the GPCR family is highly conserved provides a means to assess the solvent accessibility of TM surface regions (e.g. of amino acids in the helical transmembrane region). More specifically, the crystal structure of a control GPCR protein (e.g. inactive bovine rhodopsin) may be used to identify surface amino acids within the helical transmembrane regions of the subject GPCR (e.g. PAR-1 protein). Although the amino acid sequence identity between the subject GPCR and the control GPCR may be relatively low (e.g. 20%), recognizable patterns are found between their amino acid sequences especially in the regions of TM helices.

[0048] In the methods of the present invention, sequence alignments are performed as comparisons between the amino acid (or nucleic acid) sequences of the subject GPCR protein and the control GPCR proteins. An example is provided in Figure 2 showing a sequence alignment between a control GPCR, rhodopsin, and a subject GPCR, PAR-1. This particular alignment is consistent with a multi-sequence alignment of 270 class A GPCRs reported by Mirzadegan et al. (Mirzadegan, T., et al., *Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin*. Biochemistry, 2003. 42(10): p. 2759-67). A number of signature motifs are well conserved in such a multi-sequence alignment. For example, Asn55 (1.50) (100% conserve) in $\alpha 1$, Leu79 (2.46) (98%) and Asp83 (2.50) (93%) in $\alpha 2$, Trp161 (4.50) (98%) in $\alpha 4$, Pro215 (5.53) (91%) in $\alpha 5$, and the N/DPxxY motif in $\alpha 7$. The majority of these amino acids are located in the cytoplasmic half of the TM region of the GPCR. Because of the diversity of ligands and G-proteins associated with GPCRs, these signature structural motifs are more likely involved in common properties (e.g. overall folding, membrane-orientation, trafficking, and/or signal-transduction mechanism) of this protein family rather than essential for ligand and/or G-protein specificity. For example, a strong inter-helix hydrogen bond between Asn55(1.50)–Asp83(2.50) in rhodopsin is likely also conserved in PAR1. Moreover, a critical disulfide bond in the extracellular region of rhodopsin between residues 110 and 187 appears conserved in PAR1 between residues 175 and 254, suggesting that not only the seven-TM helix bundle but also the very organization of exo-membrane domains is conserved to some extent between rhodopsin and PAR1.

[0049] In addition to the sequence alignment, previous saturation mutagenesis of each of the seven TM helices of another class A GPCR, C5aR, also suggests that the observed helix-helix interacting surfaces in the rhodopsin structure are conserved among members of the class A family. Furthermore, helix packing moment analysis introduced by Liu et al. (Liu, W.,

et al., *Helix packing moments reveal diversity and conservation in membrane protein structure*. J Mol Biol, 2004. 337(3): p. 713-29), further supports this alignment. Helix packing moment analysis is based on observations that in membrane proteins small and/or weakly polar residues such as Ala, Gly, Ser, Thr, and Cys are more likely to be involved in helix-helix packing. All helices, except $\alpha 6$, in rhodopsin and in a sequence alignment based PAR1 model show a clear distribution of these small residues in the helix-helix interfaces, supporting the validity of the latter. The TM helix $\alpha 6$ is unusual in that its helix-packing moment vector does not point to helix-helix interface in the rhodopsin crystal structure. The amino acid sequences do not show a clear pattern in this region in the family-wide alignment. Partly because of these anomalies, it is suspected that $\alpha 6$ is involved in a conformational change associated with GPCR activation. The orientation prediction of the transmembrane helices, however, need not be performed for all transmembrane regions. In some embodiments, 2, 3, 4, 5, or 6 of the 7 transmembrane helices are analyzed.

[0050] To estimate the number of point mutations needed to convert a GPCR protein into a soluble one, the surface hydrophobicity of rhodopsin crystal structure with that of some typical soluble proteins was compared. Of the total 342 residues in one rhodopsin molecule, 145 are found on the surface of its isolated crystal structure (PDB file 1L9H); and 84 of them (58%) are hydrophobic. This surface hydrophobicity ratio is 29%, 28%, 34%, and 39% for Rab5, T4 lysozyme, GST, and the serine protease domain of plasminogen, respectively, which is significantly lower than the 58% value of rhodopsin. If the number of surface hydrophobic residues of rhodopsin is reduced from the current 58% to ~35%, then ~33 (i.e. $84 - 145 \times 35\%$) point mutations of H-P substitution are needed. This equates to ~four to five substitutions in each of the seven TM helices. Similar estimation is likely to be valid for other GPCR proteins including PAR1, because they share a similar overall structure in the transmembrane region. In some embodiments, the number of amino acid substitutions is at least 2 per transmembrane helix. In some embodiments, the number of transmembrane helices modified is at least 4, 5, or 6.

[0051] In some embodiments, additional surface mutations are added to provide spared solubility for future functional studies. Here, an earlier point mutation for solubilization is restored (i.e. reversed) to the wild type hydrophobic helical transmembrane amino acid residue present in the subject GPCR protein to optimize functionality.

[0052] Because transmembrane surface residues are among the most variable ones in both the GPCR superfamily and individual subfamilies (e.g. PARs), mutations in this region are unlikely to interrupt the overall structure of PAR1. For example, one of the most well known soluble counterparts is T4 lysozyme in which most surface point mutations have essentially
 5 no effects on the protein stability and overall structure.

[0053] Typically, the soluble GPCR variant will include minimal structural disturbance due to mutations. Thus, in some embodiments, the method is initiated using a small number of point mutations in the middle of surface helices. In other embodiments, the mutations are first made in the middle of the helix to maximize the solubilization effect, where the surface of the
 10 native protein is more hydrophobic in general than the flanking regions. A recent study on membrane insertion of a potassium-channel voltage sensor protein demonstrates that introducing polar residues, e.g. arginine, in the middle of a TM helix has the largest effect in increasing the free energy requirement for membrane insertion (i.e. thermodynamically most unfavorable) (Hessa, T., S.H. White, and G. von Heijne, *Membrane insertion of a potassium-*
 15 *channel voltage sensor*. Science, 2005. 307(5714): p. 1427).

[0054] Typically, a TM helix ranges in length between 25 and 35 amino acid residues, depending on the angle the helix makes with the membrane. In some embodiments, for each helix, ~2 positions are selected from each of the middle three or four helical turns.

[0055] In some embodiments, the selection of the hydrophobic helical transmembrane amino acid residue for replacement is based on visual inspection of the GPCR crystal
 20 structure (e.g. rhodopsin PDB file 1L9H) and calculation of its solvent-accessible surface (see e.g. Fig. 2). Based on this methodology, over 30 positions were identified in the homology model of PAR1 for mutagenesis (Table 1). In addition, a mutated sequence of PAR1 to a web based program (TMpred (<http://www.ch.embnet.org>)) which predicted that all TM helices in
 25 this variant would lose their transmembrane tendency. More web-based programs for related purposes have been reviewed by other researchers (Ahram, M. and D.L. Springer, *Large-scale proteomic analysis of membrane proteins*. Expert Rev Proteomics, 2004. 1(3): p. 293-302).

Table 1. Certain mutation sites on surface of TM helices of PAR1.

<u>Helix</u>	<u>Potential mutation sites</u>
$\alpha 1$	Phe104(1.34), Gly111(1.41), Val114(1.44), Val115(1.45), Leu117(1.47), Leu119(1.49), Ile121(1.51), and Ile128(1.58)

	$\alpha 2$	Val149(2.51), Leu150(2.52), and Phe157(2.59)
	$\alpha 3$	Phe177(3.27) and Ile198(3.48)
	$\alpha 4$	Phe221(4.44), Leu224(4.47), Ala225(4.48), Ala228(4.51), Leu229(4.52), Ile231(4.54), and Val235(4.58)
5	$\alpha 5$	Ala276(5.44), Phe280(5.48), Val281(5.49), Ile284(5.52), Val288(5.56), and Val291(5.59)
	$\alpha 7$	Leu355(7.37), Val359(7.41), Ile362(7.44), Ile366(7.48), and Leu369(7.51)

[0056] The number following the residue type is that in the amino acid sequence of native
10 PAR1. The number in parenthesis is that of the BW numbering system, and is consistent with
the PAR1 sequence in Figure 2.

[0057] In some embodiments, certain structural elements are introduced simultaneously to
enhance protein stability in the aqueous environment. For example, engineering surface
hydrogen bonds, particularly inter-helix salt-bridges are provided to stabilize soluble GPCR
15 variants. In some embodiments, maintaining and/or enhancing helix propensity is used to
stabilize the soluble variant. In some embodiments, the point mutations are not be located in
the N- or C-terminal cap range to minimize capping effects of the mutation on helix stability.
In other embodiments, all cysteine residues that are not in positions forming disulfide bridges
in the transmembrane region are mutated to serine residues to reduce complexity during
20 protein expression.

[0058] In some embodiments, conserved proline residues (and adjacent residues) playing
important structural roles by maintaining a kink in a long TM helix and providing certain
flexibility between the separated segments are conserved. For example, the side chain of Ser,
Thr or Cys residue at the (i -1) position relative to the Pro residue (i) may form a hydrogen
25 bond with the backbone carbonyl group of the (i -4) position.

[0059] In some embodiments, the crystal structures of existing soluble α -helical bundle
proteins are used as a template for designing multiple mutations on the helix surface. For
example, the Rabaptin5 four-helix bundle structure employs numerous inter-helix hydrogen
bonds. Each of the helices in the antiparallel four helix bundle consists of more than 70
30 residues and thus provides choices for templates.

[0060] The overall structure of a typical PAR1, like other GPCR proteins, contains a well packed TM domain and short loops connecting the helices outside of both sides of the membrane. For example, the N-terminal peptide of PAR1 contains the thrombin cleavage site and is significantly different from that of rhodopsin in both length and the amino acid sequence. Although a proper thrombin cleavage site may be required for studying the proteolytic activation, similar activation effects may be mimicked in the absence of the cleavage site by using agonist peptides. It has been shown that plasmin cleavage at Lys82 of PAR1 does not desensitize Ca²⁺ response of platelets or COS7 cells to the PAR1-specific agonist peptide of sequence SFLLRN (Kuliopulos, A., et al., *Plasmin desensitization of the PAR1 thrombin receptor: kinetics, sites of truncation, and implications for thrombolytic therapy*. *Biochemistry*, 1999. 38(14): p. 4572-85), suggesting that the ligand binding site is located toward the C-terminal side of Lys82. Thus, in some embodiments, this region (i.e. residues 1–80) is removed for both structural and functional studies. PAR1 variants having variable lengths of N-terminal peptides may be constructed to select for more soluble, stable variants. For example, the first N-terminal 20 residues of native PAR1 are extremely hydrophobic (Fig. 2), presumably functioning as a signal peptide to interact with signal-recognition particle for targeting translocon during biogenesis. In some embodiments, the region is deleted to increase PAR1 solubilization. Furthermore, the C-terminal tail (residues Val382–Thr425) of PAR1 is shown to be dispensable for thrombin-induced MAP kinase activation. Thus, in some embodiments, this region is truncated without disrupting the overall structure of PAR1.

[0061] In some embodiments, fusion proteins of G α with solubilized PAR1 variants are constructed. The benefits of using such fusion proteins include the defined 1:1 stoichiometry of PAR1 and G α (which is believed to be biologically relevant by some researchers) and proper physical proximity of the C-terminus of GPCR to the N-terminus of G α which has been indicated to be required for GPCR mediated G-protein activation.

[0062] In some embodiments, co-crystal structures of PAR1 and G α proteins are prepared. Recombinant G α proteins may be expressed that have been shown to bind with PAR1, including Gq/11, Gi2, G12, and G13. Insect cell and bacteria-cell based expression systems have been used for G α over-expression in other investigations. For example, G α 12 and G α 13 can be expressed and purified from Sf9 insect cells 154. (Kozasa, T. and A.G. Gilman, *Purification of recombinant G proteins from Sf9 cells by hexahistidine tagging of associated*

subunits. *Characterization of alpha 12 and inhibition of adenylyl cyclase by alpha z*. J Biol Chem, 1995. 270(4): p. 1734-41; Singer, W.D., R.T. Miller, and P.C. Sternweis, *Purification and characterization of the alpha subunit of G13*. J Biol Chem, 1994. 269(31): p. 19796-802). E. coli expression of other G α proteins have been documented by Lee et al. (Lee, E., M.E. Linder, and A.G. Gilman, *Expression of G-protein alpha subunits in Escherichia coli*. Methods Enzymol, 1994. 237: p. 146-64). Recently, Ishihara et al. reported several examples using a cell-free translational system to express GPCR-G α fusion proteins (Ishihara, G., et al., *Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors*. Protein Expr Purif, 2005. 41(1): p. 27-37). Crystal structures of a number of G α subunits have been reported in PDB (about half dozen entries are found as of late 2005). The co-crystal structure will provide, among others, detailed information on the G α -PAR1 interface and structural bases of GEF activity of PAR1 towards G-proteins.

[0063] In some embodiments, functional assays are used to evaluate PAR1 variants and/or to guide their optimization. In some embodiments, functionality is assessed using a protein-fragment complementation assay (PCA). In the so-called low affinity complementation assay, two separately synthesized fragments of the reporter protein can not spontaneously reconstitute the functional reporter protein. Instead, two fusion interacting proteins are needed to bring them together for reconstitution of the reporter function. In some embodiments, the survival dihydrofolate reductase (DHFR) based PCA is employed (Fig. 6). Prokaryotic and eukaryotic DHFRs are central to cellular one-carbon metabolism and are absolutely required for cell survival. Specifically, they catalyze the reduction of dihydrofolate to tetrahydrofolate for use in transfer of one-carbon units required for biosynthesis of serine, methionine, purines, and thymidylate. Reconstitution of enzyme activity of murine DHFR (mDHFR) can be monitored in vivo by survival of E. coli cells under a condition that bacterial DHFR activity is selectively suppressed (i.e. trimethoprim at 1 μ g/mL).

[0064] The two complementation fragments of mDHFR are called F[1,2] (residues 1-107) and F[3] (residues 108-159) (Gegg, C.V., K.E. Bowers, and C.R. Matthews, *Probing minimal independent folding units in dihydrofolate reductase by molecular dissection*. Protein Sci, 1997. 6(9): p. 1885-92). To bring the two fragments together, the G α N-terminus may be fused to the C-terminus of F[1,2] construct, and the C-terminus of the GPCR variant may be fused to the N-terminus of the F[3] construct. This design allows a functional reconstitution

of mDHRF when GPCR and $G\alpha$ interact with each other. Alternative connections may also be constructed.

[0065] Correctly folded, soluble GPCR variants may be selected from a library by interacting with $G\alpha$ from *E. coli* cell culture and/or agar-plates and identified further by full-length DNA sequencing from the (trimethoprim resistant) colonies. The GPCR library containing saturated random mutations may be constructed using methods well known in the art. As alternative approaches, either a β -lactamase based low affinity PCA or the commercial HIS3-aadA based Bacterio Match II Two-Hybrid Vector Kit (Stratagene) may be used to screen for functional GPCR variants.

[0066] Before studying the interaction between GPCR variants and their partners, a number of techniques may be employed to verify the quality of the solubilized GPCR variants, including for example, (i) circular dichroism (CD) to determine the secondary structure contents and thermal stability of the recombinant protein, (ii) native gel, sizing chromatography and/or dynamic laser scattering (DLS) to verify its aggregation state, and (iii) analytical ultra centrifugation (AUC) to determine the oligomerization state of PAR1 molecule in solution.

[0067] An important property of the GPCR–G-protein interaction is the ability of the GPCR protein to cause release of GDP from the $G\alpha$ subunit of the heterotrimeric G-protein and initiate binding of GTP to $G\alpha$. For example, the $G\alpha$ subunits of a number of G-proteins (e.g. Gi2, Gq/11,, Gq/16, G12, and G13) have been shown to bind with activated GPCR proteins, such as PAR1, and this interaction stimulates the GTP loading to the G-protein. Although the $G\beta\gamma$ complex may influence the GPCR– $G\alpha$ binding, $G\alpha$ alone is sufficient to respond to agonist binding.

[0068] In some embodiments, a functional test for solubilized GPCR variants is used in which $G\alpha$ proteins are expressed in HEK293 cells (see C2.3) or Sf9 cells and an affinity pull-down assay and nucleotide-loading assay is performed. A more quantitative measurement of the affinity may be carried out with surface plasmon resonance (SPR) using the BIACORE instrument and/or isothermal titration calorimetry. Since many GPCR proteins, such as PAR, play a GEF role for $G\alpha$, the activated recombinant GPCR protein binds with the nucleotide free-form or the GDP-bound form of $G\alpha$ better than with the GTP-bound form in these affinity assays.

[0069] Measurement of the GTP-loading may be carried out using a modified version of [35S]GTP γ S-based assay described by McIntire et al. (McIntire, W.E., et al., *Reconstitution of G protein-coupled receptors with recombinant G protein alpha and beta gamma subunits*. Methods Enzymol, 2002. 343: p. 372-93), which was used for studying interactions between
5 membrane-bound GPCR and G-proteins before. Here, a kinetic nucleotide exchange assay is used to analyze the GEF activity of solubilized GPCR variants in the presence and absence of an agonist.

[0070] In some embodiments, an N-terminal truncated G α variant is used to reduce structural flexibility and to test its interaction with solubilized GPCR variant in solution.

10 cDNAs of many GPCR -interacting G α proteins are in the public domain.

[0071] In some embodiments, a second generation of soluble GPCR protein is produced to achieve positive results in the functional assays. In some embodiments, second generation design of soluble PAR1 variants preserve any dimer interface by avoiding mutations that have drastic effects.

15 [0072] In some embodiments, GPCR dimerization is tested using a pull-down assay between recombinant, soluble proteins of two distinguishable GPCR constructs. For example, a GST-fusion PAR1 may be used to pull-down an excessive amount of the same PAR1 variant without a tag. The results are analyzed using SDS-PAGE followed by western blot against PAR1. Once a potential homo-dimerization interface is identified, alanine-scanning
20 mutagenesis is used to confirm the finding.

[0073] The functional consequence of the dimerization may be further analyzed by comparing a pro-dimerization variant with dimer-breaking mutants in assays such as G α -binding. Since GPCRs have a uniform orientation relative to the membrane, the parallel orientation of GPCR in a dimer may be verified using the PCA technique outlined herein.

25 [0074] GPCR homodimerization may be studied by fluorescence resonance energy transfer (FRET) in intact COS7 cells, using live-cell microscopy techniques. As the efficiency of FRET is dependent on the inverse sixth power of the intermolecular separation, FRET is an valuable technique for investigating the changes in molecular proximity of biological macromolecules and has been widely used to study GPCR oligomerization. Two different
30 donor and acceptor dyes or a donor and a quencher may be used, allowing detection of FRET by the appearance of sensitized fluorescence of the acceptor or by quenching of donor

fluorescence. Two approaches to identify GPCR homodimerization by FRET include, for example, (1) using Fab fragments of the monoclonal anti human GPCR antibody labeled with dye (e.g. Alexa488 and Cy3) for measurements of FRET-induced sensitized emission, and (2) making GPCR GFP variants fusion proteins by tagging GPCR with variants of GFP that will form a suitable pair for FRET experiments (e.g. cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP)).

II. Examples

[0075] The following point mutations were constructed in a parental background of N-terminal truncation variant of PAR1 (i.e. residues 23–425) using the QuikChange II Kit (Stratagene).

Table 2 Human PAR1 variants containing multiple point mutations

<u># ID</u>	<u>Mutations</u>
1. M1	F221D
2. M3	V114D, V115R, F221D
3. M4	G111K, V114D, V115R, F221D
4. M6	G111K, V114D, V115R, F221D, F280D, V281R
5. M7	G111K, V114D, V115R, F221D, F280D, V281R, I324N
6. M8	G111K, V114D, V115R, F221D, F280D, V281R, I324N, V331N
7. M9	G111K, V114D, V115R, F221D, L224N, F280D, V281R, I324N, V331N
8. M10	G111K, V114D, V115R, L150R, F221D, L224N, F280D, V281R, I324N, V331N

[0076] cDNA of human PAR1 was obtained from Dr. S. Coughlin. Fragments of residues 23–425 were cloned into a number of vectors for mutagenesis and expression trials.

[0077] To prepare for cassette mutagenesis, a WT PAR1 cDNA was constructed containing over 20 unique restriction sites by either adding new sites or converting double cleavage sites (i.e. one restriction endonuclease cleaves at two places) into single cleavage sites using silent mutations (Fig. 3). Two mutant PAR1 containing 27 hydrophobic-to-charge/polar substitutions (M27) and 32 hydrophobic-to-charge/polar substitutions (M32) were

construction based on this silent PAR1 variant. M27 consists of residues 81-425 and contains the following point mutations: L104E, F105R, G112D, V115K, L118K, V126D, L151R, S154E, F158R, F178R, I199E, F222E, L225K, A226D, A229E, A277D, F281H, V282N, I285E, V289K, V292N, V314N, C321K, I325E, V332D, L356E, V360D, plus an N-terminal
5 modification of L81M. M32 consists of residues 21-425 and contains the following point mutations: L104E, F105R, G112D, V115K, V116T, L108K, L120R, I122T, I129T, V150T, L151R, F158R, I199E, F222E, L225K, A226E, A229E, L230N, I232T, V236T, A277D, F281H, V282N, I285E, V289K, V292N, L356E, V360D, I363T, I367T, L370N plus an N-terminal modification of L21M.

10 A. PAR1 expression in a cell-free coupled transcription-translation system

[0078] A cell-free, coupled transcription-translation expression system was tested for protein expression of a number of PAR1 variants. We have tested the effects of primer optimization, linear vs. circular form of templates, and *E. coli*. vs. wheat-germ based cell
15 extraction on PAR1 expression as suggested by Roche (the manufacturer of the commercial cell-free expression kits). While the effects of primer and DNA forms are minimal, *E. coli*-based expression system appears to work best for PAR1. A selection of detergents was tested, and Treen-20 and Brij35 were shown to improve solubility of some PAR1 variants. PAR1 variant M27, was tested in comparison with M1 (Fig. 3). The result shows that with a MBP-
20 or GST-fusion construct, M27 mostly stays in the soluble form while M1 is mostly in the aggregated form and pellet. Furthermore, GST-M27 fusion protein was purified using glutathione (GSH) affinity resin (data not shown). Following cleavage of MBP-tagged M32 from solid support, the solutions was centrifuged and the supernatant was subjected to gel electrophoresis, which showed a band corresponding to M32. This indicates that M32 is soluble in aqueous solution.

25 B. Expression of Gα12 recombinant protein

[0079] To study the interaction between PAR1 and Gα proteins, the cDNA of human Gα12 was acquired (GenBank Access #L01694) from the cDNA Resource Center of University of Missouri at Rolla (UMR). The recombinant protein was expressed in HEK293 cell as described before (Zhu, G., et al., *Crystal structure of the human GGA1 GAT domain*.
30 Biochemistry, 2003. 42(21): p.6392-9).

C. High affinity protein-fragment complementation assay

[0080] To screen a saturated random library for soluble PAR1 variants, a high affinity protein-fragment complementation assay (PCA) based on α -complementation of β -galactosidase (β -Gal) was performed. Tests showed that having a soluble protein, MBP, in the plasmid vector (pMAL-C2X) results in blue colonies, while insoluble WT or M1 mutant (see Table 2) of PAR1 results in white colonies (Fig. 4), suggesting that the screening works properly. A vector containing both β -lactamase and β -Gal α -fragment has also been constructed for double screening and maintaining the random library.

D. Cassette Mutagenesis

[0081] The following is a prophetic example of using cassette mutagenesis in the generation of a water soluble PAR1 protein. Cassette mutagenesis has been successfully used in structure-function studies of GPCRs with ancestral gene reconstruction. The use of this technique in the current project is not only convenient, but in many cases essential, for example, where a large number of point mutations or saturated random mutations are introduced into a relative small region (e.g. in one TM helix).

[0082] Unique restriction endonuclease sites that flank each of the TM helices of interest are constructed and utilized. A large piece of synthetic DNA containing multiple point mutations is inserted between two unique restriction sites in a given TM helix. A TM helix usually ranges in length between 25 and 35 residues, corresponding to 75–105 bases. If both pre-constructed restriction sites are located within the helix, the inserted DNA piece is constructed with two pieces of ~60-base, staggered oligomers using PCR amplification.

[0083] The above discussed Cys-to-Ser mutations can also be introduced at this step. In case that this technique is used for a saturated random mutagenesis, special attention will be given to avoid degenerated codons that potentially introduce unexpected restriction sites thus interfere with existing ones. The experimental protocols of synthetic gene design and construction, oligo-nucleotide synthesis, stepwise PCR, cloning of PCR product, and expression of synthetic genes have been documented in detail (Chang, B.S., M.A. Kazmi, and T.P. Sakmar, *Synthetic gene technology: applications to ancestral gene reconstruction and structure-function studies of receptors*. Methods Enzymol, 2002. 343: p.274-94). Numerous silent mutation sites for commercial restriction endonucleases can be identified in the PAR1 gene accessing the online program NEBcutter at the New England BioLabs' web site

(www.neb.com). Simply by introducing silent restriction sites and reducing double-cleavage sites to single-cleavage sites, we have constructed a cDNA encoding wild-type PAR1 yet containing over 20 unique restriction sites (Fig. 3). The restriction sites are chosen such that all potential cleavage sites produce overhang ends promoting directional ligation.

5 **E. Directed evolution**

[0084] The following is a prophetic example of using the directed evolution method in the as a complementary approach to creating soluble PAR1 variants. The directed evolution usually include two steps: diversity generation and screening. Common techniques of diversity generation include saturated mutagenesis and DNA shuffling.

10 [0085] A PAR1 cDNA library is constructed containing saturated random mutations at selected positions, for example, from the list in Table 1. To screen for soluble PAR1 variants from such a library, the protein-fragment complementation assay (PCA) technique is employed. A functional reporter protein is rationally split into two fragments. Association of the two fragments provides information on the status of peptide fused with the fragments. The high affinity complementation technique uses two separately synthesized fragments that spontaneously reconstitute to the functional reporter protein. This technique is used to detect the existence of functional GPCR in soluble form. The β -galactosidase (β -Gal) α -complementation is employed with the PAR1 protein fused to the N-terminus of the α -fragment of *E. coli* β -Gal (residues 7–58). The screening is performed on X-gal plates by identifying blue colonies of *E. coli* DH5 α (lacZ Δ M15, Invitrogen) containing ω -fragment of β -Gal and co-transfected with the fusion of β -Gal α -fragment. The positive control is a soluble protein, such as MBP, in the place of PAR1, which produces all colonies in blue color (see Fig. 4).

[0086] By comparing the color density of colonies in plates of varied X-Gal concentrations, the relative level of expression quantity and solubility of different PAR1 variants is determined. The protocol described by Wigley and coworkers is instructive (Wigley, W.C., et al., *Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein*. Nat Biotechnol, 2001. 19(2): p. 131-6).

F. Methods and Procedures

30 [0087] The methods and procedures discussed in this section are generally well established techniques.

[0088] *Construct generation.* Full length cDNA clones for the proteins described here have been obtained from individual laboratories or UMR cDNA Resource Center. All subsequent constructs are generated by PCR using high-fidelity polymerase and sub-cloned into variety of vectors for expression. Primers are designed with standard considerations in mind, such as minimizing hairpins, primer duplexes, mispriming, and optimizing the melting temperature. Site directed point mutations will be incorporated into the PAR1 cDNA using the modified QuikChange (Stratagene) protocol recently described by Zheng et al. (Zheng, L., U. Baumann, and J.L. Reymond, *An efficient one-step site-directed and site-saturation mutagenesis protocol*. Nucleic Acids Res, 2004. 32(14): p. e115). In all cases, constructs are verified by sequencing the entire coding region using the OMRF DNA Sequencing Core Facility.

[0089] *Expression and purification.* Constructs are tested for optimal soluble expression in either BL21(DE3) CodonPlus cells (Stratagene) or BL21(DE3) Rosetta cells (Novagen) in 2xLB media under the following conditions: i) cells are grown at room temperature or 28–37°C to an OD_{600nm} of 0.6–1.0 and induced with 1 mM IPTG for 3 h; and ii) cells are grown at room temperature to an OD_{600nm} of 0.2–0.4 and induced with 0.1 mM IPTG for 14 h. Where necessary, temperature and length of induction are subsequently varied to achieve optimal production of soluble/functional proteins. For purification, cells are suspended in lysis buffer and disrupted by frozen and thaw. Triton X100 is added to a final concentration of 0.5%, and the crude cell lysates are centrifuged at ~35,000g for 40 min. All tagged proteins are first purified using affinity columns following manufacturer recommended protocols and further purified with either Resource Q or Resource S (Amersham) followed by gel filtration chromatography, and the purity is judged by SDS-PAGE. To generate an untagged form, the fusion protein at a concentration of 2–4 mg/mL is incubated with the minimal amount of proper protease overnight at 4° to achieve 90% cleavage. Incubation with a proper affinity resin eliminates residual uncleaved fusion protein, and the subsequent chromatographic steps eliminate the protease from the preparation.

[0090] *Cell-free coupled transcription-translation protein expression.* For a small scale (50 µL) reaction, PCR generated linear template or midi-prep purified (Qiagen) plasmid DNA (0.5 µg), is added to the reaction mixture containing E. coli S30 extract, an energy generating system, tRNAs, amino acids, nucleotides, and T7 RNA polymerase. Messenger RNA coding for the PAR1 protein is synthesized by in vitro transcription of the DNA template under the

control of T7 RNA polymerase promoter. The reaction is carried out with constant temperature (30°C) and mixing. Each 50 µL reaction generates ~10–20 µg fusion protein. After separating any aggregated, insoluble protein from soluble one by centrifugation of the total reaction mixture, expression of the target protein is analyzed using SDS-PAGE followed by western blotting; detection is by monoclonal anti-PAR1 antibody against the PAR1 42–48 region (Santa Cruz Biotechnology, Inc.) or by anti-bodies against other tags if the PAR1 epitope is not available.

[0091] Protein samples can be quantified according to its specific (mutation dependent) molar extinction coefficient by measuring the UV absorbance at 280 nm. Guanidine hydrochloride (6 M at ~pH 7) can be used to solubilize those samples of less soluble PAR1 variants during a UV measurement. Because PAR1 contains many cysteine residues, most of which do not form disulfide bonds in the native protein, a reducing agent (e.g. dithiothreitol (DTT)) may be employed for the solubilization process and proper folding of the protein unless they have been systematically mutated.

[0092] *Circular dichroism (CD) analysis.* Melting temperature (T_m) is measured as a means of quality control. α -Helices, which are the dominant secondary structure elements in PAR1, have the strongest CD signal around 222 nm. Thus, stability is determined by recording the CD222 nm as a function of temperature in a buffer that supports solubility of PAR1 variants in a wide temperature range.

[0093] A well folded globular protein is usually characterized by a two-state temperature curve in a CD scan indicating cooperative folding, and T_m is defined by the middle point of the transition region between the two states. The folding reversibility is checked by overlap of spectra before and after heating.

[0094] *Analytical ultra centrifugation.* Beckman XLA and XLI instruments are employed to perform both sedimentation velocity (SV) and sedimentation equilibrium (SE) measurements. The sedimentation equilibrium measurements allows the study of oligomerization of a PAR1 variant. A sample is dialyzed against a reference buffer without reducing agents and centrifuged to equilibrium in the Beckman Optima XLI. The absorbance (A) at 230 or 280 nm is measured as a function of the radial distance (r) from the axis of rotation. Data are analyzed by fitting with the function $A(r) = (A_0 - A_\infty) \exp[-n s_m (r^2 - r_0^2) / 2] + A_\infty$, where A_0 and A_∞ are constants, n represents the oligomeric state, r_0 is a reference

data point where $A(r)=A_0$, and sm is calculated using the monomer molecular mass. By carefully selecting a matching solution, one can compensate the effect of detergent on the apparent molecular weight of the PAR1 complex.

[0095] *Affinity pull-down assay.* To test the functions of solubilized PAR1, recombinant fusion proteins of, for example, GST-PAR1 variants are used to pull-down full-length G α 12. All three forms of G α 12 are tested. First, GST-PAR1 variants are incubated with glutathione (GSH) beads in the presence or absence of synthesized PAR1 agonist peptide (e.g. the peptide SFLLRNP of KD $\sim 1 \mu M$) to obtain active and inactive forms of immobilized PAR1, respectively. Then, the G α 12 recombinant protein in a certain nucleotide-binding form is loaded and incubated. The sample is washed to get rid of non-specific binding and analyzed with SDS-PAGE and western blots against G α 12 (Biogenesis, UK).

[0096] *Nucleotide loading.* Guanine nucleotides (GDP, GTP, GppNHp, or [35S]GTP γ S) are loaded by incubating G α for 30 min in buffer containing 20 mM HEPES (pH 7.5), 150 mM NaCl, 1 mM EDTA, and a 20 fold excess of nucleotide. Unbound nucleotide is removed by gel filtration using either a D-Salt column (Pierce) or a Superdex-75 column. Due to variations in pI, the buffer conditions are adjusted as needed to avoid precipitation during the exchange reaction.

[0097] *Surface plasmon resonance (SPR).* Binding affinity, particularly the on- and off rates, between PAR1 variants and recombinant G α proteins is quantitatively determined using the BIACORE 3000 biosensor (BIACORE Inc). The standard method of binding the ligand to a carboxymethylated dextran matrix (CM5, Cat. #:Br-100-12) is used. Two other BIACORE methodologies for alternate, comparative purposes and have been used in binding assays to study interactions between Rab5 variants and their potential binding partners.

[0098] The first is based on binding of GST-fusion protein to the sensor chip through a GST-mediated binding mode, and the second is based on poly-His tags. Analyte (G α) at varied concentration is injected in the flow cell (90 s) followed by the buffer-only dissociation time (180 s), and the sensorgram is recorded as the chipbound protein molecules associate and dissociate with the analyte. For a simple unimolecular dissociation process, the SPR signal follows an exponential decay, $R(t) = R_0 \exp(-k_{off} t)$, where R_0 is the initial SPR signal (in resonance units, RU) and k_{off} is the dissociation rate constant. For a simple bimolecular association process, the rate of change in the SPR signal is given by the equation $dR/dt = k_{on}$

$C R_{eq} - (k_{on} C + k_{off}) R$, where R_{eq} is the SPR signal at equilibrium, C is the concentration of the immobilized protein (the 'ligand'), and k_{on} is the association rate constant. First the dissociation rate constant (k_{off}) and then the association rate constant (k_{on}) will be calculated using these homogeneous kinetics models. For a simple two-component binding reaction, the dissociation constant (K_D) is obtained from the ratio of the rate constants, $K_D = k_{off}/k_{on}$. In cases where the association and dissociation rates are both fast, R_{eq} can be extracted directly from the data plotted as a function of analyte concentration $[C]$, and K_D obtained from a non-linear least-squares fit to $R_{eq} = R_{max}[C]/(K_D + [C])$.

[0099] Computational analysis of the binding curves is performed using the BIAevaluation software as previously described. Mean values and standard errors will be determined from at least three independent experiments. To determine the extent to which mass transport contributes to the observed data, sensorgrams will be measured over a wide range of flow rates and analyte concentrations. Flow rates and analyte concentrations will be adjusted where necessary to minimize contributions from rebinding or mass transport. Only sensorgrams clearly in the kinetically controlled regime (i.e. not limited by mass transport or rebinding) are included in the analysis.

[0100] *Nucleotide exchange kinetics.* GTP/GDP exchange kinetics is measured by monitoring the binding of $[35S]GTP\gamma S$ to the GDP-loaded $G\alpha$. In this exchange reaction, $G\alpha$ is diluted to 10 nM (containing 0.5 μM final GDP), and the PAR1: $G\alpha$ ratio ranges between 0 and 1. After 10 min incubation at 25°C, about 7×10^6 cpm $[35S]GTP\gamma S$ is added, which brings the final reaction volume to 500 μL . At this point (time=0), the mixture is split into two aliquots: one of 210 μL (for the control) and the other 290 μL (for the PAR1/agonist). A baseline rate of $[35S]GTP\gamma S$ binding is established by taking samples over a 15–20 min incubation period from the control tube. At 8 min after the zero time point, varied concentration (0–100 nM) of PAR1 agonist are added. The receptor-activated time course is established by removing 30 μL aliquots every 60 s. All samples are filtered through nitrocellulose filters (Millipore).

[0101] The filters are washed three times with 4 mL of an ice-cold buffer containing 5 mM $MgCl_2$ and counted by liquid scintillation counting. The binding rate of $[35S]GTP\gamma S$ (k_{obs}) is determined from the linear region of the binding curve for each PAR1 or agonist

concentration. The kinetics efficiency, k_{cat}/K_m , can be derived from $k_{obs} = (k_{cat}/K_m)[PAR1] + k_{intr}$, where k_{intr} is the basal binding-rate and $[PAR1]$ is kept as $\ll K_m$.

[0102] *FRET experiments.* Nonfluorescent acceptors are used such as the QSY dyes from Molecular Probes (Eugene, OR). The FRET experiments are carried out on COS cells grown on coverslips until subconfluent levels, and transferred to the FCS2 thermostated chamber. Two different FRET approaches are employed to test the dimerization both on living and on fixed cells. First, an acceptor-sensitized emission FRET is used to follow PAR-1 dimerization in real-time on living cells. Donor fluorescence (CFP) is excited by the emission of acceptor. In parallel, control cell lines expressing only the acceptor (YFP-PAR1) are subjected to the same experiment to take into account the bleeding of the excitation energy from the donor (CFP) into the acceptor (YFP) channel. This image is considered as the background for the experiment. Second, intensity – based FRET detection is employed using fixed cells, based on donor de-quenching after specific photobleaching of the acceptor, as described in a paper from Lupu lab on coendocytosis of β -secretase and the amyloid precursor protein (Huang, X.P., et al., *J Biol Chem*, 2004. 279(36): p. 37886-94).

[0103] *Crystallization and data collection.* Prior to a crystallization experiments, the protein sample is exchanged into a low ionic strength buffer and concentrated to 10–20 mg/mL. Sparse matrix screens (from Hampton Research, Emerald BioStructures, and/or Jena Biosciences) for initial crystallization conditions are performed by vapor diffusion in hanging or sitting drops. For optimization, precipitants are systematically varied as a function of pH and temperature in the presence and absence of various additives (e.g. monovalent, divalent or trivalent salts). Other variables (such as protein concentration, detergents, and chaotropic agents) are also considered. Micro-seeding is employed as needed to control nucleation and improve crystal morphology. Stabilizer solutions for heavy atom soaks and/or cryo-solutions for freezing are empirically determined. For data collection, crystals are flash frozen in liquid nitrogen and transferred to a nitrogen cryostream (Oxford Cryosystems) maintained at 100°K.

[0104] Diffraction data are collected using an in-house Rigaku generator equipped with Osmic mirrors and a MAR345 image plate detector or at a synchrotron source. Crystals are maintained at a constant temperature of 100°K in a nitrogen cryostream to minimize radiation damage and allow complete data sets to be collected on a single crystal. Raw intensity data are indexed/processed with HKL2000.

[0105] *Heavy atom substitution.* Heavy atom screens begin with mercurial reagents (e.g. CH₃HgCl, Hg(OAc)₂, and PCMB), which have large isomorphous differences and moderate anomalous signals at the CuK α edge. Other reagents to be screened include lanthanides (e.g. GdCl₃ and TbCl₃). GdCl₃ and TbCl₃ derivatives have moderate isomorphous differences and large anomalous signals at the CuK α edge. Soaks are conducted for 2 d at initial concentrations of 1 mM (mercurials) or 10 mM (lanthanides).

[0106] The concentration of heavy atom reagents and/or length of the soak are adjusted empirically to achieve optimal substitution. If necessary, the screen will extend to other common heavy atom reagents. For each heavy atom soak, a small wedge of data (typically five 1° oscillation images) is collected and scaled directly to the native data set. For candidate derivatives showing resolution dependent intensity differences consistent with heavy atom binding, complete data sets are collected.

[0107] Alternatively, seleno-methionine based multi-wavelength anomalous diffraction (MAD) experiment will be used to determine the phases. As few as one methionine per hundred residues is sufficient for this technique to provide an interpretable map, provided the crystal exhibits strong diffraction and the data are accurately measured. A typical experiment involves data collection at three or four wavelengths around the selenium absorption edge at a tunable synchrotron X-ray source. The location of the edge is determined for each crystal by scanning the X-ray emission spectrum. If the crystals cannot be oriented to collect Friedel pairs on the same image, a reverse beam strategy is used. Experimental maps are calculated with phases derived from MAD alone, in combination with other derivatives, or combined with MR phases.

[0108] *Crystallographic software and procedures.* Heavy atom sites are located by Patterson methods (using SOLVE (Terwilliger, T.C., *SOLVE and RESOLVE: automated structure solution and density modification*. Methods Enzymol, 2003. 374: p. 22-37)), direct methods (SHELXD) (Schneider, T.R. and G.M. Sheldrick, *Substructure solution with SHELXD*. Acta Crystallogr D Biol Crystallogr, 2002. 58(Pt 10 Pt 2): p. 1772-9) and Shake-and-Bake (Miller, R., et al., *SnB: Crystal structure determination via Shake-and-Bake*. Journal of Applied Crystallography, 1994. 27: p. 613-621), or from isomorphous or anomalous difference Fourier. Heavy atom parameters are refined with SHARP. Initial phases

are improved by solvent flipping and, where appropriate, non-crystallographic symmetry averaging.

[0109] Manual chain-tracing, model building and manipulation of masks are facilitated by the molecular graphics program O (Jones, T.A., et al., *Improved methods for binding protein models in electron density maps and the location of errors in these models*. Acta Crystallogr A, 1991. 47(Pt 2): p. 110-9). For high resolution data set, automated model building is conducted with ARP/wARP. Molecular replacement solutions are identified using PHASER or AMORE. Simulated annealing and positional refinement are carried out with CNS. Publication quality figures are rendered with MolScript. Unless otherwise referenced, programs are used as implemented in CCP4 (Bailey, S., *The CCP4 suite: programs for protein crystallography*. Acta Crystallographica, 1994. D50: p. 760-763).

[0110] *Cytosolic free-calcium measurement*. Effects of PAR1 mutants on (presumably Gq-mediated) PLC activation are analyzed by measuring the intracellular free Ca^{2+} . First, cells overexpressing a PAR1 mutant are loaded with Fluo-3, a fluorescent indicator of intracellular free calcium, following the protocols from the manufacturer (Molecular Probe). After adding agonist peptide (e.g. SFLLRN-NH₂, ~1 μM) or thrombin (~10 nM) to the Fluo-3 loaded cells (~ 2×10^5 per mL), fluorescence measurements are monitored using a Perkin-Elmer fluorescence spectrometer with an excitation wavelength of 480 nm and an emission recorded at 530 nm. The data are normalized with regard to receptor density on the cell surface which are determined using an ELISA assay. The larger the normalized fluorescent signal, the stronger is the PAR1 response to the agonist.

[0111] *Inositol phosphate hydrolysis assay*. The hydrolysis of inositol phosphates by activated PLC are measured after thrombin-induced activation of PAR1. The stably transfected CHO (or COS7) cells were sub-cultured in multi-well culture dishes and labeled with 3 $\mu\text{Ci/mL}$ myo-[3H]-inositol (Amersham-Biosciences) for 2 d at 37°C. Then, the cells are stimulated with ~10 nM thrombin (or PAR1 agonist peptide) for 1 min at 37°C in the presence of 10 mM LiCl. Cell extracts are loaded onto a 1-mL column of AG 1-X8 anion-exchange gel resin (BioRad). The column is washed with 3 mL of 40 mM NH₄OH (pH 9.0) and eluted with 4 mL of 2 M ammonium formate and 0.1 M formic acid. The collected inositol mono-, bis-, and triphosphates are quantified by scintillation counting.

[0112] *MAP kinase activity assay.* PAR1 activates MAP kinase through both a pertussis toxin (PTX)-sensitive Gi-dependent pathway and a Gq- and PKC-dependent pathway. The MAP kinase activity is measured by the amount of phosphate group transferred from ATP to peptides. The PAR1 variant-expressing cells are plated (into a 10-cm dish at 10⁶ cells) and
5 cultured overnight. They are then incubated with serum-starved medium for 2 d, and the cells are exposed to PAR1-agonists (1 min) and lysed at 4°C. After centrifugation (at ~20,000g for 15 min at 4°C), the supernatant is used for MAP kinase activity assay using the p42/p44 MAP kinase enzyme assay system (Amersham-Biosciences). The reaction is initiated by adding of [γ³²P]-ATP. Incubation proceeds for 30 min, and the phosphorylated peptide is separated
10 from the unincorporated radioactivity on binding paper. After washing the paper, the extent of phosphorylation is measured by scintillation counting. Meanwhile, the total protein concentration in the cell lysate is measured to verify that the same amounts of proteins are used for the MAP kinase activity assay.

[0113] *DNA synthesis assay.* To measure DNA synthesis as a marker of thrombin-induced
15 cell proliferation, the PAR1-overexpressing CHO cells are cultured overnight and then starved for 1–2 d; the cells are then incubated with or without thrombin for 24 h and in the presence bromodeoxyuridine (BrdU, 10 mM) for the last 18 h at 37°C. After removing the culture medium, the cells are fixed and incubated with the peroxidase-labeled anti-BrdU antibody for 1.5 h at room temperature using the Labeling and Detection Kit (Roche). The
20 immune complexes of newly synthesized DNA are detected with a fluorescently labeled secondary antibody and fluorescent spectrophotometer.

WHAT IS CLAIMED IS:

1 1. A method of making a water-soluble (G Protein)-Coupled Receptor
2 (GPCR) Protein, said method comprising the steps of:

3 (a) performing a sequence alignment between a subject GPCR protein and a
4 control GPCR protein thereby identifying a set of helical transmembrane amino acid residues
5 forming five transmembrane helices of said subject GPCR protein;

6 (b) assessing the solvent accessibility of each amino acid in said set of helical
7 transmembrane amino acid residues;

8 (c) selecting five hydrophobic helical transmembrane amino acid residues
9 forming part of five different transmembrane helices of said subject GPCR protein, wherein
10 said selecting is based at least in part on said assessing, and wherein said five hydrophobic
11 helical transmembrane amino acid residues form part of said set of helical transmembrane
12 amino acid residues;

13 (d) replacing said five hydrophobic helical transmembrane amino acid residues
14 with five independently selected hydrophobic amino acid residues by performing site directed
15 mutagenesis of said subject GPCR protein, thereby making said water-soluble GPCR protein.

1 2. The method of claim 1, wherein step (c) comprises selecting ten
2 hydrophobic helical transmembrane amino acid residues forming part of five different
3 transmembrane helices.

1 3. The method of claim 1, wherein step (a) comprises performing a
2 sequence alignment between said subject GPCR protein and said control GPCR protein
3 thereby identifying said set of helical transmembrane amino acid residues forming six
4 transmembrane helices of said subject GPCR protein, and

5 step (c) comprises selecting six hydrophobic helical transmembrane amino acid
6 residues forming part of six different transmembrane helices of said subject GPCR protein,
7 wherein said selecting is based at least in part on said assessing, and wherein said six
8 hydrophobic helical transmembrane amino acid residues form part of said set of helical
9 transmembrane amino acid residues.

1 4. The method of one of claims 1-3, wherein step (c) comprises selecting
2 20 to 40 hydrophobic helical transmembrane amino acid residues, wherein said 20 to 40

3 hydrophobic helical transmembrane amino acid residues form part of said set of helical
4 transmembrane amino acid residues.

1 5. The method of one of claims 1-3, wherein step (c) comprises selecting
2 30 to 35 hydrophobic helical transmembrane amino acid residues, wherein said 30 to 35
3 hydrophobic helical transmembrane amino acid residues form part of said set of helical
4 transmembrane amino acid residues.

1 6. The method of claim 1, wherein said five hydrophobic helical
2 transmembrane amino acid residues are selected from the middle three or middle four helical
3 turns of the transmembrane helix.

1 7. The method of claim 1, wherein said subject GPCR protein is a class A
2 GPCR protein.

1 8. The method of claim 1, wherein said subject GPCR protein is PAR1.

1 9. The method of claim 1, further comprising stabilizing said water-
2 soluble GPCR by engineering one or more inter-helix bonds.

1 10. The method of claim 1, wherein said control GPCR protein is
2 rhodopsin.

1 11. The method of claim 1, further comprising (e) assessing the
2 functionality of said water-soluble GPCR using a protein-fragment complementation assay.

1 12. The method of claim 1, further comprising (f) improving the
2 functionality of said water-soluble GPCR by restoring at least one of said hydrophilic amino
3 acid residues with the wild type hydrophobic helical transmembrane amino acid residue
4 present in said subject GPCR protein, thereby producing a functionally optimized water-
5 soluble GPCR.

1 13. A water-soluble GPCR protein produced by the method of claim 1.

1 14. A water-soluble PAR-1 protein comprising 11 amino acid substitutions,
2 wherein each substitution replaces a hydrophobic amino acid with a hydrophilic amino acid
3 residue, said hydrophobic amino acids selected from the group consisting essentially of

4 Phe104, Gly111, Val114, Val115, Leu117, Leu119, Ile121, Ile128, Val149, Leu150, Phe157,
5 Phe177, Ile198, Phe221, Leu224, Ala225, Ala228, Leu229, Ile231, Val235, Ala276, Phe280,
6 Val281, Ile284, Val288, Val291, Leu355, Val359, Ile362, Ile366, and Leu369.

1 15. The water-soluble PAR-1 protein of claim 14, comprising at least 20 of
2 said amino acid substitutions.

1 16. The water-soluble PAR-1 protein of claim 14, comprising at least 30 of
2 said amino acid substitutions.

1 17. The water-soluble PAR-1 protein of claim 14, further comprising one
2 or more engineered inter-helix bonds.

1 18. A method of making a water-soluble (G Protein)-Coupled Receptor
2 (GPCR) Protein, said method comprising the steps of:

3 (a) performing a sequence alignment between a subject GPCR protein and a
4 control GPCR protein thereby identifying a set of solvent-exposed hydrophobic helical
5 transmembrane amino acids; and

6 (b) replacing five solvent-exposed hydrophobic helical transmembrane amino
7 acid residues within said set of solvent-exposed helical transmembrane amino acid residues
8 with five independently selected hydrophobic amino acid residues, thereby making said
9 water-soluble GPCR,

10 wherein each of said five solvent-exposed hydrophobic helical transmembrane
11 amino acid residues form part of a different transmembrane helix within said subject GPCR.

1 19. The method of claim 18, wherein step (b) comprises selecting ten
2 solvent-exposed hydrophobic helical transmembrane amino acid residues within said set of
3 solvent-exposed helical transmembrane amino acid residues with ten independently selected
4 hydrophobic amino acid residues, wherein at least five of said ten solvent-exposed
5 hydrophobic helical transmembrane amino acid residues form part of a different
6 transmembrane helix within said subject GPCR.

1 20. The method of claim 18, wherein step (b) comprises selecting six
2 solvent-exposed hydrophobic helical transmembrane amino acid residues within said set of
3 solvent-exposed helical transmembrane amino acid residues with six independently selected
4 hydrophobic amino acid residues, wherein said six solvent-exposed hydrophobic helical

5 transmembrane amino acid residues form part of a different transmembrane helix within said
6 subject GPCR.

1 21. The method of claim 18, wherein step (b) comprises selecting from 20
2 to 40 solvent-exposed hydrophobic helical transmembrane amino acid residues within said set
3 of solvent-exposed helical transmembrane amino acid residues with 20 to 40 independently
4 selected hydrophobic amino acid residues, wherein at least five of said 20 to 40 solvent-
5 exposed hydrophobic helical transmembrane amino acid residues form part of a different
6 transmembrane helix within said subject GPCR.

1 22. The method of claim 18, wherein step (b) comprises selecting from 30
2 to 35 solvent-exposed hydrophobic helical transmembrane amino acid residues within said set
3 of solvent-exposed helical transmembrane amino acid residues with 30 to 35 independently
4 selected hydrophobic amino acid residues, wherein at least five of said 20 to 40 solvent-
5 exposed hydrophobic helical transmembrane amino acid residues form part of a different
6 transmembrane helix within said subject GPCR.

1 23. The method of claim 18, wherein said five solvent-exposed
2 hydrophobic helical transmembrane amino acid residues are selected from the middle three or
3 middle four helical turns of the transmembrane helix.

1 24. The method of claim 18, wherein said subject GPCR protein is a class
2 A GPCR protein.

1 25. The method of claim 18, wherein said subject GPCR protein is PAR1.

1 26. The method of claim 18, further comprising stabilizing said water-
2 soluble GPCR by engineering one or more inter-helix bonds.

1 27. The method of claim 1, wherein said control GPCR protein is
2 rhodopsin.

1 28. The method of claim 18, further comprising (e) assessing the
2 functionality of said water-soluble GPCR using a protein-fragment complementation assay.

1 29. The method of claim 18, further comprising improving the
2 functionality of said water-soluble GPCR by restoring at least one of said hydrophilic amino

3 acid residues with the wild-type hydrophobic helical transmembrane amino acid residue
4 present in said subject GPCR protein, thereby producing a functionally optimized water-
5 soluble GPCR.

1 30. A water-soluble GPCR protein produced by the method of claim 18.

1/7

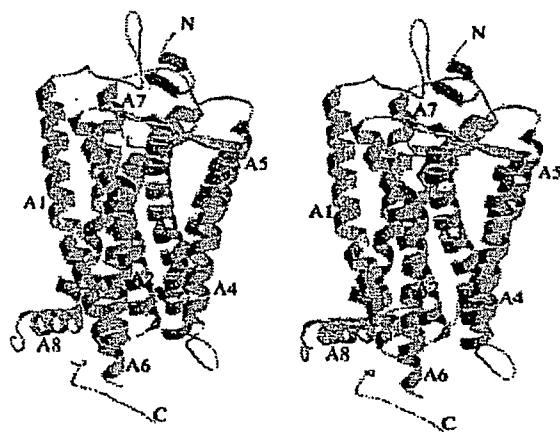


FIG. 1

1000
 1001
 1002
 1003
 1004
 1005
 1006
 1007
 1008
 1009
 1010
 1011
 1012
 1013
 1014
 1015
 1016
 1017
 1018
 1019
 1020
 1021
 1022
 1023
 1024
 1025
 1026
 1027
 1028
 1029
 1030
 1031
 1032
 1033
 1034
 1035
 1036
 1037
 1038
 1039
 1040
 1041
 1042
 1043
 1044
 1045
 1046
 1047
 1048
 1049
 1050
 1051
 1052
 1053
 1054
 1055
 1056
 1057
 1058
 1059
 1060
 1061
 1062
 1063
 1064
 1065
 1066
 1067
 1068
 1069
 1070
 1071
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107
 1108
 1109
 1110
 1111
 1112
 1113
 1114
 1115
 1116
 1117
 1118
 1119
 1120
 1121
 1122
 1123
 1124
 1125
 1126
 1127
 1128
 1129
 1130
 1131
 1132
 1133
 1134
 1135
 1136
 1137
 1138
 1139
 1140
 1141
 1142
 1143
 1144
 1145
 1146
 1147
 1148
 1149
 1150
 1151
 1152
 1153
 1154
 1155
 1156
 1157
 1158
 1159
 1160
 1161
 1162
 1163
 1164
 1165
 1166
 1167
 1168
 1169
 1170
 1171
 1172
 1173
 1174
 1175
 1176
 1177
 1178
 1179
 1180
 1181
 1182
 1183
 1184
 1185
 1186
 1187
 1188
 1189
 1190
 1191
 1192
 1193
 1194
 1195
 1196
 1197
 1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218
 1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244
 1245
 1246
 1247
 1248
 1249
 1250
 1251
 1252
 1253
 1254
 1255
 1256
 1257
 1258
 1259
 1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276
 1277
 1278
 1279
 1280
 1281
 1282
 1283
 1284
 1285
 1286
 1287
 1288
 1289
 1290
 1291
 1292
 1293
 1294
 1295
 1296
 1297
 1298
 1299
 1300
 1301
 1302
 1303
 1304
 1305
 1306
 1307
 1308
 1309
 1310
 1311
 1312
 1313
 1314
 1315
 1316
 1317
 1318
 1319
 1320
 1321
 1322
 1323
 1324
 1325
 1326
 1327
 1328
 1329
 1330
 1331
 1332
 1333
 1334
 1335
 1336
 1337
 1338
 1339
 1340
 1341
 1342
 1343
 1344
 1345
 1346
 1347
 1348
 1349
 1350
 1351
 1352
 1353
 1354
 1355
 1356
 1357
 1358
 1359
 1360
 1361
 1362
 1363
 1364
 1365
 1366
 1367
 1368
 1369
 1370
 1371
 1372
 1373
 1374
 1375
 1376
 1377
 1378
 1379
 1380
 1381
 1382
 1383
 1384
 1385
 1386
 1387
 1388
 1389
 1390
 1391
 1392
 1393
 1394
 1395
 1396
 1397
 1398
 1399
 1400
 1401
 1402
 1403
 1404
 1405
 1406
 1407
 1408
 1409
 1410
 1411
 1412
 1413
 1414
 1415
 1416
 1417
 1418
 1419
 1420
 1421
 1422
 1423
 1424
 1425
 1426
 1427
 1428
 1429
 1430
 1431
 1432
 1433
 1434
 1435
 1436
 1437
 1438
 1439
 1440
 1441
 1442
 1443
 1444
 1445
 1446
 1447
 1448
 1449
 1450
 1451
 1452
 1453
 1454

FIG. 2

3/7

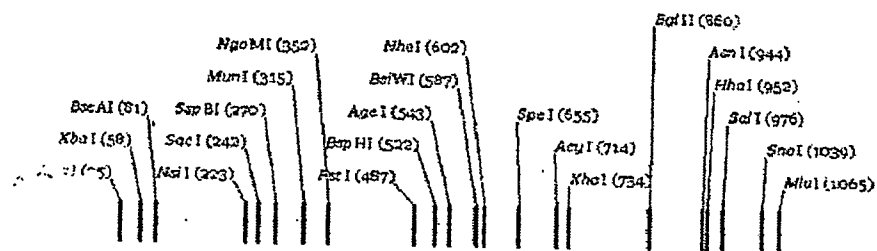


FIG. 3

4/7

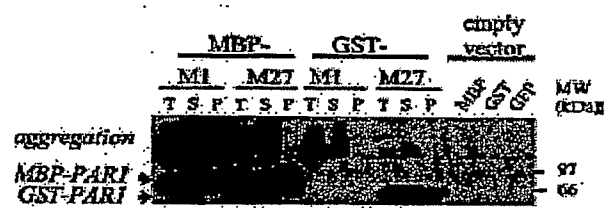


FIG. 4

5/7

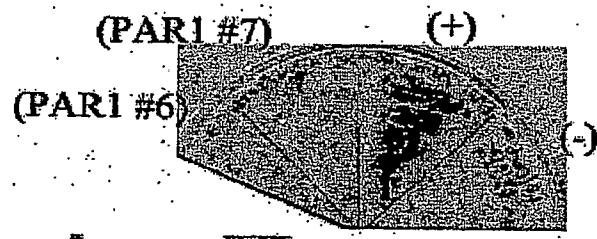


FIG. 5

6/7

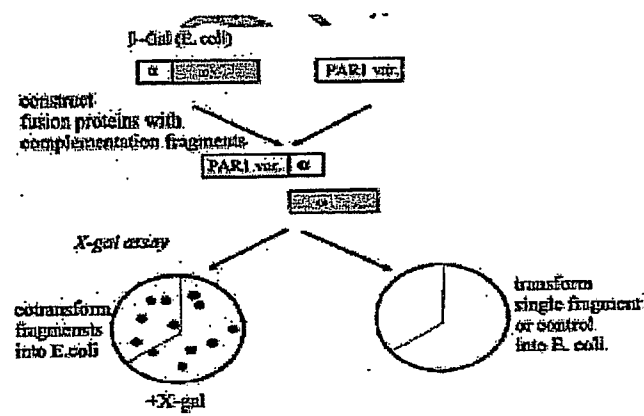


FIG. 6

7/7

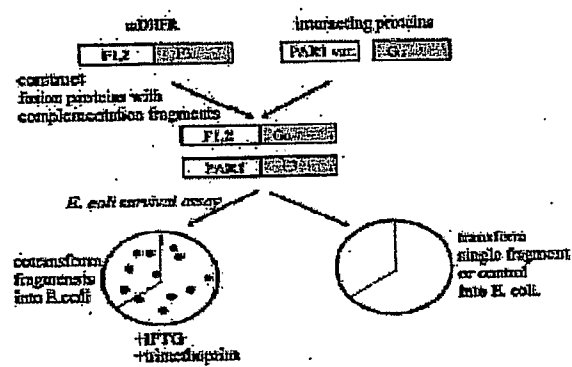


FIG. 7