

CLAIMS

1. A computer implemented method for executing a procedure to select a water-soluble variant of a G Protein-Coupled Receptor (GPCR), the method comprising:
 - (1) entering a sequence of the GPCR for analysis;
 - (2) obtaining a variant of the GPCR, wherein a plurality of hydrophobic amino acids in the transmembrane (TM) domain alpha-helical segments ("TM regions") of the GPCR are substituted, wherein:
 - (a) said hydrophobic amino acids are selected from the group consisting of Leucine (L), Isoleucine (I), Valine (V), and Phenylalanine (F);
 - (b) each said Leucine (L) is independently substituted by Glutamine (Q), Asparagine (N), or Serine (S);
 - (c) each said Isoleucine (I) and said Valine (V) are independently substituted by Threonine (T), Asparagine (N), or Serine (S); and,
 - (d) each said Phenylalanine is substituted by Tyrosine (Y); and, subsequently,
 - (3) obtaining an α -helical secondary structure result for the variant to verify maintenance of α -helical secondary structures in the variant;
 - (4) obtaining a trans-membrane region result for the variant to verify water solubility of the variant,thereby selecting the water-soluble variant of the GPCR.
2. The method of claim 1, wherein step (3) is performed prior to, concurrently with, or after step (4).
3. The method of claim 1 or 2, wherein in step (2), one subset of said plurality of hydrophobic amino acids in one and the same TM region of the GPCR are substituted to generate one member of a library of potential variants, and one or more different subsets of said plurality of hydrophobic amino acids are substituted to generate additional members of the library.
4. The method of claim 3, further comprising ranking all members of said library based on a combined score, wherein the combined score is a weighed combination of the α -helical secondary structure prediction result and the trans-membrane region prediction result.

5. The method of claim 1 further comprising ranking the variant using a ranking function.
6. The method of claim 1 further comprising performing the method with a data processor.
7. The method of claim 6 further comprising a memory connected to the data processor.
8. The method of claim 5 wherein the ranking function includes a secondary structure component and a water solubility component.
9. The method of claim 8 wherein the ranking function includes a weighting value for the secondary structure component and/or the water solubility component.
10. The method of claim 4, further comprising: selecting N members with the highest combined scores to form a first library of potential variants for said TM region, wherein N is a pre-determined integer (e.g., 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more).
11. The method of claim 10, further comprising generating one library of potential variants for 1, 2, 3, 4, 5, or all 6 other TM regions of the GPCR.
12. The method of claim 11 further comprising: replacing two or more TM regions of the GPCR with corresponding TM regions from the libraries of potential variants, to create a library of combinatorial variants.
13. The method of any one of claims 1-12, wherein substantially all (e.g., all) said leucines are substituted by glutamines.
14. The method of any one of claims 1-13, wherein substantially all (e.g., all) said isoleucines are substituted by threonines.
15. The method of any one of claims 1-14, wherein substantially all (e.g., all) said valines are substituted by threonines.
16. The method of any one of claims 1-15, wherein substantially all (e.g., all) said phenylalanines are substituted by tyrosines.

17. The method of any one of claims 1-16, wherein one or more (e.g., 1, 2, or 3) said leucines are not substituted.
18. The method of any one of claims 1-17, wherein one or more (e.g., 1, 2, or 3) said isoleucines are not substituted.
19. The method of any one of claims 1-18, wherein one or more (e.g., 1, 2, or 3) said valines are not substituted.
20. The method of any one of claims 1-19, wherein one or more (e.g., 1, 2, or 3) said phenylalanines are not substituted.
21. The method of claim 1 or 2, wherein in step (2), all (100% of) said leucines in the TM regions are substituted by glutamines, all (100% of) said isoleucines in the TM regions are substituted by threonines, all (100% of) said valines in the TM regions are substituted by threonines, and all (100% of) said phenylalanines in the TM regions are substituted by tyrosines.
22. The method of claim 1, wherein all said leucines are substituted by glutamines, all said isoleucines are substituted by threonines, all said valines are substituted by threonines, or all said phenylalanines are substituted by tyrosines.
23. The method of any one of claims 1-22, further comprising producing / expressing said combinatorial variants.
24. The method of any one of claims 1-23, further comprising testing said combinatorial variants for ligand binding (e.g., in yeast two-hybrid system), wherein those having substantially the same ligand binding compared to that of the GPCR are selected.
25. The method of any one of claims 1-24, further comprising testing said combinatorial variants for a biological function of the GPCR, wherein those having substantially the same biological function compared to that of the GPCR are selected.
26. The method of any one of claims 1-25, wherein the library of combinatorial variants comprises less than about 2 million members.
27. The method of any one of claims 1-26, wherein said sequence of the GPCR comprises information about the TM regions of the GPCR.

28. The method of any one of claims 1-27, wherein said sequence of the GPCR is obtained from a protein structure database (e.g., PDB, UniProt).
29. The method of any one of claims 1-28, wherein the TM regions of the GPCR are predicted based on the sequence of the GPCR.
30. The method of claim 29, wherein the TM regions of the GPCR are predicted using TMHMM 2.0 (TransMembrane prediction using Hidden Markov Models) software module.
31. The method of claim 30, wherein said TMHMM 2.0 software module utilizes a dynamic baseline for peak searching.
32. The method of any one of claims 1-31, further comprising providing a polynucleotide sequence for each variants of the GPCR.
33. The method of claim 32, wherein the polynucleotide sequence is codon optimized for expression in a host (e.g., a bacterium such as *E. coli*, a yeast such as *S. cerevisiae* or *S. pombe*, an insect cell such as Sf9 cell, a non-human mammalian cell, or a human cell).
34. The method of any one of claims 1-33, wherein the scripted procedure comprises VBA scripts.
35. The method of any one of claims 1-34, wherein the scripted procedure is operable with a Linux system (e.g., Ubuntu 12.04 LTS), a Unix system, a Microsoft Windows operative system, an Android operative system, or an Apple iOS operative system.
36. A non-transitory computer readable medium having stored thereon a sequence of instructions to perform the method of any of claims 1-35.
37. A data processing system operative to select a water soluble variant of a membrane protein, comprising a data processor operative to perform substitution of amino acids,
wherein the system ranks a protein variant with a ranking function, the ranking function including a secondary structure component and a water solubility component; and,

wherein the system is operative to process steps (a), (b), (c) and (d) of any one of claims 1-35.

38. A computer implemented method for executing a procedure to select a water-soluble variant of a method comprising:

- (1) operating a data processing to identify a sequence of a membrane protein for analysis;
- (2) obtaining a variant of the membrane protein, wherein a plurality of hydrophobic amino acids in the transmembrane (TM) domain alpha-helical segments ("TM regions") of the membrane protein are substituted, wherein the data processor:
- (3) determines an α -helical secondary structure result for the variant to verify maintenance of α -helical secondary structures in the variant;
- (4) determines trans-membrane region result for the variant to verify water solubility of the variant; and
- (5) selects a water-soluble variant of the membrane protein,

wherein the substitution include:

- (a) hydrophobic amino acids are selected from the group consisting of Leucine (L), Isoleucine (I), Valine (V), and Phenylalanine (F);
- (b) each Leucine (L) is independently substituted by Glutamine (Q), Asparagine (N), or Serine (S);
- (c) each Isoleucine (I) and said Valine (V) are independently substituted by Threonine (T), Asparagine (N), or Serine (S); and,
- (d) each said Phenylalanine is substituted by Tyrosine (Y).

39. The method of claim 38, wherein one subset of said plurality of hydrophobic amino acids in one and the same TM region of a GPCR are substituted to generate one member of a library of potential variants, and one or more different subsets of said plurality of hydrophobic amino acids are substituted to generate additional members of the library.