TITLE: Locating the ligand binding sites for the G-protein coupled estrogen receptor (GPER) using combined information from docking and sequence conservation

Authors: Ashley R. Vidad[1], Stephen Macaspac[1] & Ho Leung Ng[1,2]*

[1]University of Hawaii at Manoa, Department of Chemistry, Honolulu, HI. USA
[2]University of Hawaii Cancer Center, Honolulu, HI. USA

* Corresponding author. hng@hawaii.edu

Abstract

High concentrations of estrogenic compounds can overstimulate estrogen receptors and potentially lead to breast, ovarian, and cervical cancers. Recently, a G-protein coupled estrogen receptor (GPER/GPR30) was discovered that has no structural similarity to the well-characterized, classical estrogen receptor ERα. The crystal structure of GPER has not yet been determined, and the ligand binding sites have not yet been experimentally identified. The recent explosion of GPCR crystal structures now allow homology modeling with unprecedented reliability. We create, validate, and describe a homology model for GPER. We describe and apply ConDock, the first hybrid scoring function to use information from protein surface conservation and ligand docking, to predict binding sites on GPER for four ligands, estradiol, G1, G15, and tamoxifen. ConDock is a simple product function of sequence conservation and binding energy scores. ConDock predicts that all four ligands bind to the same location on GPER, centered on L119, H307, and N310; this site is deeper in the receptor cleft than are ligand binding sites predicted by previous studies. We compare the sites predicted by ConDock and traditional methods analyzing surface geometry, surface conservation, and ligand chemical interactions. Incorporating sequence conservation information in ConDock avoids errors resulting from physics-based scoring functions and modeling.

Introduction

Estrogens are hormones with multiple physiological and pathological functions in both men and women. The primary reproductive hormone in females, estrogen is responsible for ovarian development, oocyte and endometrial maturation, and uterine contraction[1]. Overexpression of estrogens is associated with estrogen-sensitive ovarian, endometrial, and breast cancers[2]. In addition to its role in the female reproductive system, estrogen is also an important regulator in the skeletal, nervous, cardiovascular, endocrine, and immune systems[3]. Estrogen functions in production of bone, insulin secretion, T cell differentiation, neuroprotection, and modulation of pain sensation, and cardiomyocyte contractility[3–5]. Imbalance of estrogen levels can lead to various pathological disorders such as osteoporosis, diabetes mellitus, autoimmune diseases, and dilated myopathy[3–6].

The most important estrogen is 17β-estradiol (E2). The primary estrogen receptors, estrogen receptor-α (ERα) and estrogen receptor-β (ERβ) are dimeric nuclear receptor transcription factors that activate gene expression when bound to E2[7,8]. These two receptors are structurally very similar. Depending on tissue and ligand, ERα and ERβ can act redundantly, synergistically, and/or antagonistically. Studies suggest that ERα and ERβ have minimal participation in rapid and membrane-associated estrogen signaling activities as shown by the continued activation by estrogen in the presence of ER antagonists[9]. This suggested the existence of unidentified membrane-associated estrogen receptors. G-protein coupled estrogen receptor (GPER, formerly known as GPR30) is a membrane-bound ER that was discovered recently; GPER is proposed to be a primary mediator of rapid estrogen-associated effects[10–12]. It has also been shown to mediate other estrogen-activated responses such as cAMP regeneration and nerve growth factor expression[13,14]. Because of the technical challenges involved in experiments with

G-protein coupled receptors, biochemical experiments are difficult and no crystal structure of GPER is available, and thus details of ligand binding are unknown.

Estrogen receptors can recognize a range of structurally diverse ligands (Fig. 1). E2 is the primary ligand of well-characterized estrogen receptors. E2 binds to all known estrogen receptors and is thus non-specific for GPER. Estrogen receptor binding ligands, such as tamoxifen, are also used as drugs to treat estrogen-responsive breast cancer[15,16]. Tamoxifen acts as an ER antagonist in some tissues and as an ER agonist in others, which is why it is considered a selective estrogen receptor modulator (SERM)[17]. Other drugs such as fulvestrant act as ER antagonists in all tissues; these drugs are considered to be selective estrogen receptor downregulators (SERD)[18]. Some SERMs and SERDs are agonists of GPER[9], and recently GPER-specific ligands G1 and G15 were discovered[19,20]. G1 and G15 are structurally similar; they differ by an acetyl group (Fig. 1). G1 is an agonist, whereas G15 is an antagonist.
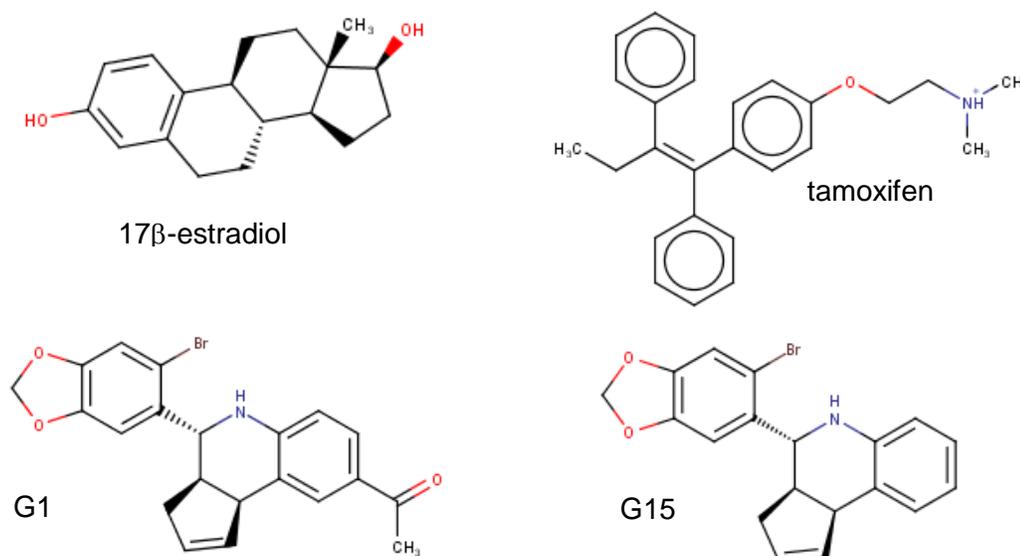

Figure 1. Four experimentally verified GPER ligands

The ligand binding sites in GPER have not yet been experimentally identified. Various computational approaches have been used to predict ligand binding sites in other G-protein coupled receptors. Traditional docking methods compute the lowest energy orientation of a ligand fit to a receptor surface. Such methods are highly dependent on the form of the energy scoring function and accuracy of the receptor model structure[21–23]. These methods have been used to identify ligand binding sites and build pharmacophores for G-protein coupled receptors (GPCRs)[24–26], but the lack of diverse GPCR crystal structures presents serious challenges to using docking methods for identification of ligand binding sites.

Another feature that can be used to predict ligand binding sites is surface or sequence conservation. Binding sites for particular ligands are often conserved, although systematic sequence variation can encode ligand specificity[27–29]. The massive abundance of genomic data for GPCRs strongly constrains possibilities for ligand binding sites even without chemical or structural information[30–32].

There has been less research on methods that combine information from chemical interactions, geometric surface analysis, and bioinformatics. Hybrid strategies, such as Concavity[33], have demonstrated superior performance in predicting ligand binding sites compared to single-mode approaches. Concavity scores binding sites by evolutionary sequence conservation as quantified by the Jensen-Shannon divergence[29] and employs geometric criteria of size and shape. Here, we describe and apply a new hybrid scoring function, ConDock, which combines information from surface conservation with intermolecular interactions from docking calculations, to predict ligand binding sites for GPER. We compare our results from those previously published using purely docking based methods[34,35].

Results

Because there are no crystal structures yet available for GPER, we created a homology model using GPCR-I-TASSER[36], among the most reliable GPCR homology modeling software packages. GPCR-I-TASSER identified the closest matching crystal structure to GPER as that of the CCR5 chemokine receptor (PDB 4mbs) with 23% sequence identity. GPCR-I-TASSER used this crystal structure along with 9 other GPCR crystal structures as templates for homology modeling. The GPER homology model differs from chain A of the crystal structure of CCR5 chemokine receptor with RMSD of 0.96 Å across $C\alpha$ atoms (Fig. S1). The primary differences are in the extracellular loop between helices 4 and 5 and the intracellular loops between helices 5 and 6, and after helix 7. These two intracellular loops are those predicted by ERRAT[37] to be most likely in error based on the likelihood of atom pair type interactions from high resolution crystal structures. It should be noted that ERRAT was not developed for analyzing membrane proteins so the reliability of its error analysis for the GPER homology model is uncertain. Furthermore, as the sites with the greatest predicted errors are on the intracellular face of GPER, far from the ligand binding site, they are less likely to affect our ligand prediction study.

Using the SwissDock server[38], we docked structures of the four ligands E2, G1, G15, and tamoxifen (Fig. 1) to a homology model of GPER, as no crystal structure is available. Most of the docked sites from SwissDock were not located on the extracellular face of GPER and thus were considered nonviable (Fig. S3). The shortcomings of a purely physics based scoring function such as that used by SwissDock in predicting ligand binding is not surprising given the lack of an experimental crystal structure and well known limitations of current computational methodology[39–42].

We next ranked all ligand binding sites generated by SwissDock that were located on the extracellular face of GPER using the combined ConDock score. The ConDock score is the simple product of the ConSurf[43,44] binding surface sequence conservation score and the SwissDock FullFitness energy score[45]. A highly negative ConDock score is associated with a more probable ligand binding site. For all four ligands, the ConDock score identified one or two ligand binding sites and poses that clearly outscored other candidates (Table 1). ConDock identified the same approximate binding site for all four ligands, although this was not an explicit criterion in the calculations (Fig. 2). The average ConSurf conservation score across the four ligand binding sites is 0.82, indicating that the site is highly but not completely conserved. The binding site is located deep in the receptor cleft, although this too was not a criterion in the prediction calculation. Given the lack of additional experimental evidence for the location of the ligand binding site, the proposed ConDock sites are physically reasonable.

Ligand	ConSurf conservation score	SwissDock energy score	Combined ConDock score
E2, pose 1	0.84	-964.9	-813.2
E2, pose 2	0.80	-967.9	-774.3
G1	0.85	-970.7	-825.1
G15	0.8	-974.6	-779.7
Tamoxifen, pose 1	0.81	-947.9	-763.6
Tamoxifen, pose 2	0.81	-947.5	-763.6

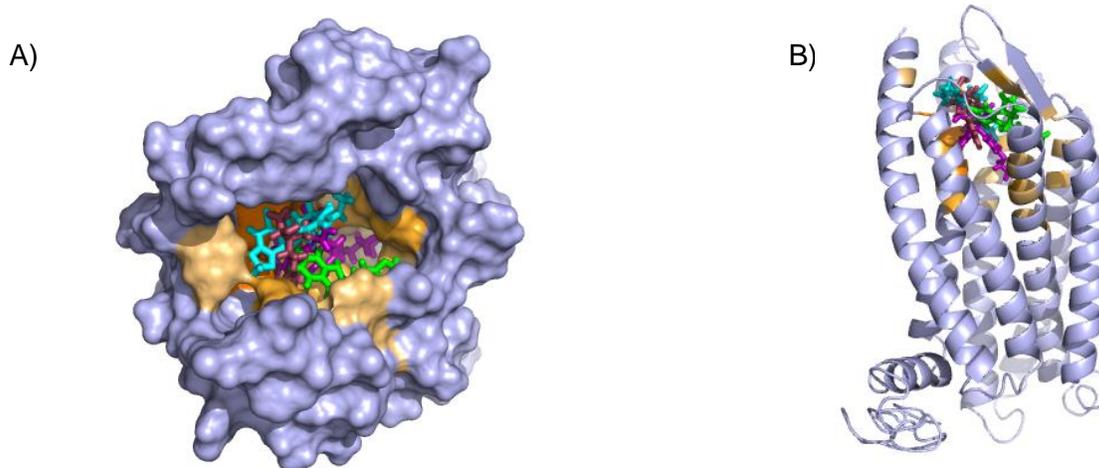**Table 1. Scores for predicted binding sites and poses for GPER ligands.**

Figure 2. Predicted ligand binding pockets in GPER. A) Extracellular perspective of GPER showing amino acids predicted to contact ligands E2 (maroon), G1 (cyan), G15 (green), and tamoxifen (violet). Residues colored orange are predicted to contact one or more ligands with darker hue indicating interaction with multiple ligands. B) Predicted binding pocket viewed from a 90° rotation.

We found two promising binding sites for E2 in GPER. The two sites are 4.4 Å apart, located deep in the receptor cleft (Fig. 3). E2 is oriented perpendicular to the lipid membrane and rotated about 180° between the two poses. The conservation scores for these two poses are 0.84 and 0.80. The energy scores of the two poses are similar. The amino acids contacting E2 in pose 1 are conserved in GPERs from six species, and only one residue contacting pose 2, H282, varies across species. In the top ranked pose, there is a hydrogen bond between the inward pointing D-ring hydroxyl group of E2 and the carboxyl terminal on E115. Hydrophobic interactions are present between E2 and non-polar residues L119, Y123, P303, and F314. In the second ranked pose, the inward pointing A-ring hydroxyl group of E2 makes a hydrogen bond with N310. This pose is in a less hydrophobic environment, contacting primarily H282 and P303.
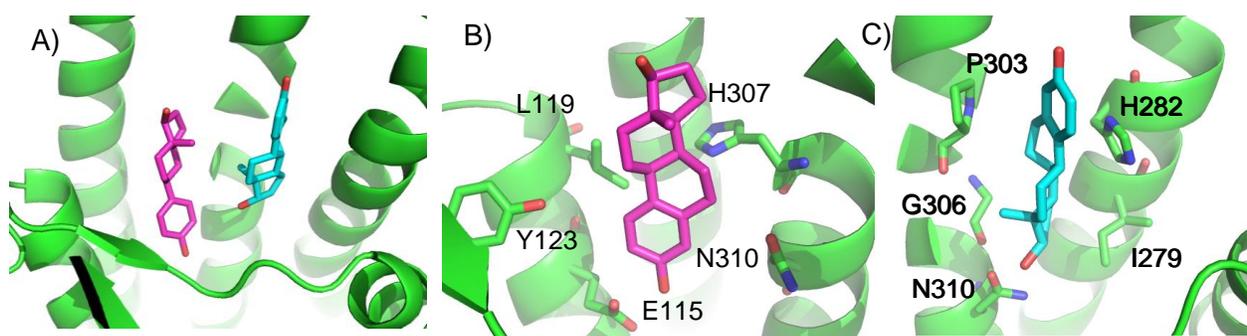

Figure 3. Predicted E2 binding sites in GPER. A) The two highest scoring docking poses for E2. B) Receptor-ligand interactions for E2 pose 1. C) Receptor-ligand interactions for E2 pose 2.

ConDock predicts that G1 and G15 bind in adjacent but distinct binding sites separated by 2.3 Å despite the chemical similarity of the two ligands. The top predicted binding site for G1 is found within the pocket bound by Y55, L119, F206, Q215, I279, P303, H307, and N310 (Fig. 4). This orientation had the highest conservation score of all predicted binding sites at 0.85. In this pose, N310 makes a long hydrogen bond with the acetyl oxygen of G1. The predicted

binding site for G15 is found within the pocket bound by L119, Y123, M133, S134, L137, Q138, P192, V196, F206, C207, F208, A209, V214, E218, H307, and N310. This pose had a conservation score of 0.8. Hydrogen bonding is not observed between GPER and G15. Hydrophobic interactions are observed with L119, Y123, F206, and V214. An interesting aspect of this binding site is that the amino acid at position 214 is valine in humans but is isoleucine in Atlantic croaker and zebrafish. These two hydrophobic amino acids differ by only a methyl group.
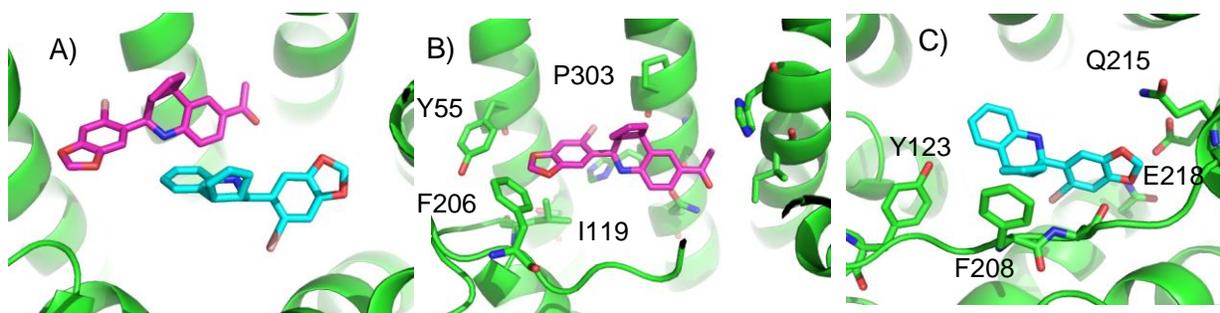

Figure 4. Predicted G1 and G15 binding sites in GPER. A) The highest scoring docking poses for G1 (maroon) and G15 (cyan). B) Receptor-ligand interactions for G1. C) Receptor-ligand interactions for G15.

ConDock predicted two equally high scoring, overlapping poses for tamoxifen, near E115, L119, Y123, L137, Q138, M141, Y142, Q215, E218, W272, E275, I279, P303, G306, H307, and N310 (Fig. 5). The conservation score of this orientation is 0.81. Hydrophobic interactions are observed between tamoxifen and non-polar residues L119, Y123, Y142, P303, and F314. Notably, the amine group of tamoxifen is neutralized by E218 and E275.
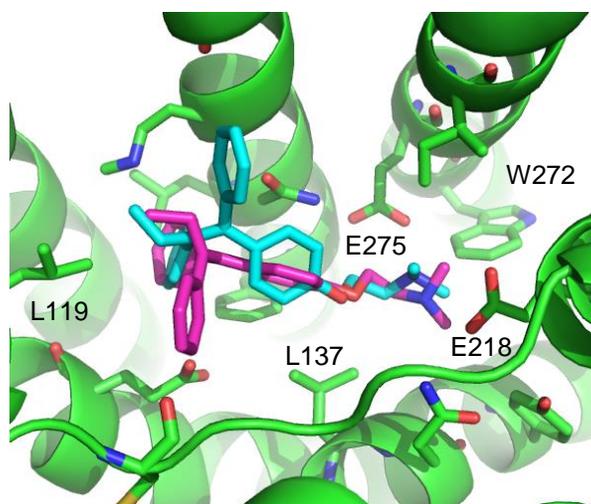
Figure 5. Predicted tamoxifen binding sites in GPER. A) The highest scoring docking poses for tamoxifen, pose 1 (maroon) and pose 2 (cyan).

In comparison with the ligand binding sites predicted by traditional methods based on surface geometry and conservation, the site predicted by ConDock is more detailed and of higher resolution due to the information from chemical interactions from ligand docking. Moreover, prediction methods based on surface geometry and conservation cannot differentiate between binding sites for different ligands. We compared the ligand binding site predicted by ConDock to those predicted by three other software packages representing different approaches: CASTp[46], which analyzes surface geometry, SiteHound[47], which maps surfaces with a chemical probe, and Concavity[33], which analyzes surface geometry and conservation (Fig. 6). All three methods were able to identify a ligand binding site roughly matching that from ConDock. The pocket predicted by ConDock is deeper than the other pockets, which while intuitively attractive is not necessarily correct. SiteHound performed particularly poorly, with the top scoring site located on the GPER intracellular face. The site identified by SiteHound closest to the ConDock site was scored third and is a shallow binding pocket near H52-G58, E275-H282, and R299-H307 (Fig. 6C). In

contrast, the Concavity site was smaller and more shallow than the ConDock site (Fig. 6D).

Surprisingly, the site predicted by the simpler CASTp method best matched the ConDock site

but is also smaller and more shallow (Fig. 6B). For proteins such as GPCRs with large, concave

binding pockets, geometry based prediction methods such as Concavity and CASTp can easily

identify the general location of the binding site. However, such methods may have more

difficulty recovering the specific, ligand-specific binding site. It is also surprising that ConDock

more closely matched the results of the geometry based methods given that ConDock does not

take surface geometry into account. Without experimental data, it is not possible to conclude

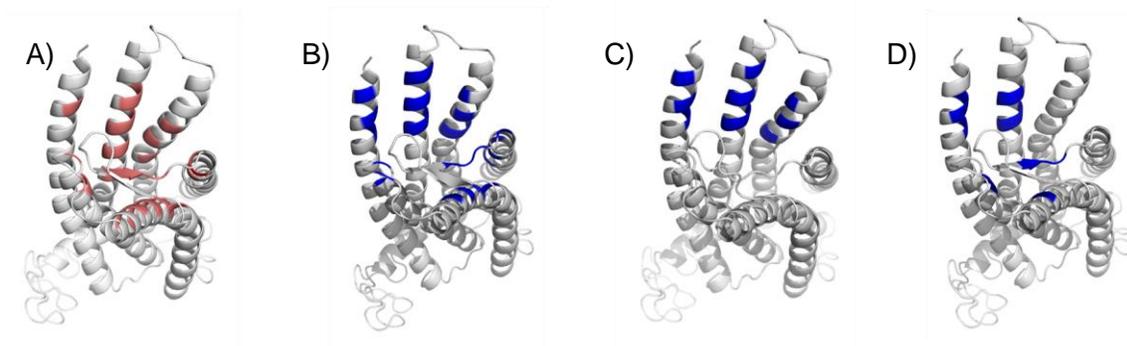which of the predicted binding sites is correct at this time.

Figure 6. Predicted ligand binding sites by ConDock, CASTp, SiteHound, Concavity. Ligand binding sites are colored, predicted by A) ConDock, B) CASTp, C) SiteHound, D) Concavity.

Discussion

The ConDock scoring method, incorporating information from both surface conservation

and docking binding energy, was used to predict viable ligand binding sites for four different

GPER ligands. In contrast to more typical geometry-based ligand binding site prediction

methods, ConDock scoring takes advantage of chemistry-specific information about the ligand-

receptor interface. The poor performance of SiteHound in predicting ligand binding sites on GPER suggests that a method based only on chemical interactions or docking is highly susceptible to error. Surface conservation data not only provides orthogonal knowledge but also dampens the influence from the shortcomings of current computational methods in homology modeling, docking, and predicting binding affinity. How best to mathematically combine these multiple data sources has been debated[29,33], but we demonstrate here that a simple product function is effective. Although the ligands analyzed differ greatly in chemical structure, the ConDock scoring method predicted that all four bind to the same approximate region, deep in the extracellular cleft of the receptor. Undoubtedly, further refinement of a hybrid scoring function will lead to improved predictions.

Recent GPER modeling studies using molecular dynamics simulations and docking identified different potential binding sites for E2, G1, and G15 near F206 and F208; the interaction with this region was described as driven primarily by π-π stacking interactions[34,35]. Figure 7 compares the ConDock binding site against that predicted in the molecular dynamics simulation and docking study. The ConDock binding site is located deeper in the extracellular cleft; the other proposed site mostly involved surface exposed loops. Mendez-Luna *et al*. proposed that Q53, Q54, G58, C205, and H282 all interact with G1 and G15; however, none of these residues are conserved across the six species we analyzed. Experimental data is not currently available to support one model over the other.
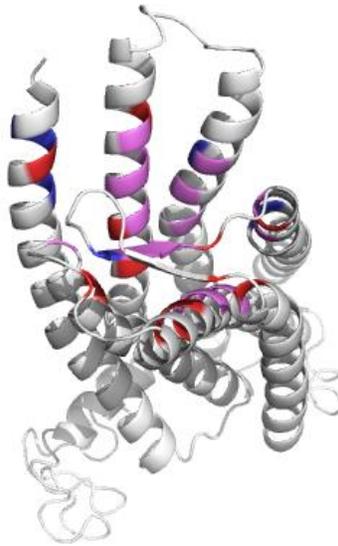
Figure 7. Comparison of proposed ligand binding sites. Comparison of the ConDock ligand binding site (red) with that proposed by Mendez-Luna *et al.* (blue). Residues found in both sites are colored violet.

In summary, the simple ConDock hybrid scoring model predicts physically plausible ligand binding sites by combining information from ligand docking and surface conservation. Using multiple orthogonal sources of information avoids errors introduced by modeling, especially in a case where a crystal structure of the receptor is unavailable. Using this hybrid method, we identified a site in the extracellular cleft of GPER that has the potential to bind four known GPER ligands. Further optimization of hybrid scoring functions should yield significantly improved predictions.

Methods

Protein conservation

GPER amino acid sequences were acquired from the UniProt protein sequence database[48]. Only Swiss-Prot[49] curated sequences were compared using the Align function of UniProt. The amino acid sequences aligned were from human (*Homo sapiens*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), Rhesus macaque (*Macaca mulatta*), zebrafish (*Danio rerio*), and Atlantic croaker (*Micropogonias undulatus*). The multiple sequence alignment file was submitted to ConSurf[43,44]. ConSurf assesses conservation using Bayesian reconstruction of a phylogenetic tree. Each sequence position is scored from 0-9, where 9 indicates that the amino acid was retained in all of the organisms (Fig. S4). Values from ConSurf were mapped onto the receptor surface with Chimera[50].

Homology modeling and docking

The crystal structure of GPER has not yet been determined. We created a model using GPCR I-TASSER (Iterative Threading Assembly Refinement), among the most accurate homology modeling softwares customized for GPCRs[36]. GPCR I-TASSER modeled the GPER structure using templates from the ten closest related GPCR crystal structures (PDB 4mbs, 2ks9, 1kpn, 1u19, 2ziy, 1kp1, 3odu, 4ea3, 4iaq, 2y00). The homology model was validated with ERRAT[37]. Coordinates for E2, G1, G15, and tamoxifen were downloaded from the ZINC ligand database[51] and submitted to SwissDock[38] for docking. SwissDock is a web interface to the EADock DSS[45] engine, which performs blind, global (does not require targeting of a particular surface) docking using the physics based CHARMM22 force field[52]. The "FullFitness Score" calculated by SwissDock using clustering and the FACTS implicit solvent model[53] was used as the "Energy Score" for our calculations.

Combined analysis

SwissDock poses were manually screened for those binding sites located on or near the extracellular side of the protein. Ligand binding surfaces included residues within 3.5 Å from the docked ligand. The average conservation score of the amino acids that were highlighted served as the "Conservation Score" of that specific orientation (Fig. 8). The combined ConDock score is defined as the product of the Conservation and Energy Scores. As the Energy Score is a modified free energy function, a highly negative ConDock score is associated with a more probable ligand binding site. Binding sites predicted by ConDock results were compared with those predicted by CASTp[46], SiteHound[47], and Concavity[33]. For CASTp, SiteHound, and Concavity, ligand binding pockets were defined as residues within 4 Å of the selected probe/cluster.

$$\textbf{Combined ConDock Score} = (\text{Conservation Score}) * (\text{Energy Score})$$

$$\textbf{Conservation Score} = \frac{1}{10} \frac{\sum_{k=1}^{n}(\text{Amino Acid ConSurf Score})_k}{n}$$

$$\textbf{Energy Score} = \text{SwissDock FullFitness Score}$$

Figure 8. Calculation of combined ConDock scores for ligand binding sites. The Conservation Score is calculated over the n residues in a binding site, indexed by k.

Acknowledgements

Author contributions

A.R.V., S.M., and H.L.N. performed the analysis and calculations. A.R.V., S.M., and

H.L.N. wrote the manuscript. H.L.N. supervised the project.

Competing financial interests

The authors have no competing financial interests.

References

1. Melmed, S., Polonsky, K. S., Larsen, P. R. & Kronenberg, H. M. *Williams Textbook of Endocrinology*. (Elsevier Health Sciences, 2011).

2. Yager, J. D. & Leihr, J. G. Molecular Mechanisms of Estrogen Carcinogenesis. *Annu. Rev. Pharmacol. Toxicol.* **36,** 203–232 (1996).

3. Simpson, E. R. Sources of estrogen and their importance. *J. Steroid Biochem. Mol. Biol.* **86,** 225–230 (2003).

4. Cenci, S. *et al.* Estrogen deficiency induces bone loss by enhancing T-cell production of TNF-α. *J. Clin. Invest.* **106,** 1229–1237 (2000).

5. Costrini, N. V. & Kalkhoff, R. K. Relative effects of pregnancy, estradiol, and progesterone on plasma insulin and pancreatic islet insulin secretion. *J. Clin. Invest.* **50,** 992–999 (1971).

6. Donaldson, C. *et al.* Estrogen Attenuates Left Ventricular and Cardiomyocyte Hypertrophy by an Estrogen Receptor–Dependent Pathway That Increases Calcineurin Degradation. *Circ. Res.* **104,** 265–275 (2009).

7. Heldring, N. *et al.* Estrogen Receptors: How Do They Signal and What Are Their Targets. *Physiol. Rev.* **87,** 905–931 (2007).

8. Harrington, W. R. *et al.* Activities of estrogen receptor alpha- and beta-selective ligands at diverse estrogen responsive gene sites mediating transactivation or transrepression. *Mol. Cell. Endocrinol.* **206,** 13–22 (2003).

9. Prossnitz, E. R. & Barton, M. The G-protein-coupled estrogen receptor GPER in health and disease. *Nat. Rev. Endocrinol.* **7,** 715–726 (2011).

10. Carmeci, C., Thompson, D. A., Ring, H. Z., Francke, U. & Weigel, R. J. Identification of a Gene (GPR30) with Homology to the G-Protein-Coupled Receptor Superfamily Associated with Estrogen Receptor Expression in Breast Cancer. *Genomics* **45,** 607–617 (1997).

11. Kvingedal, A. M. & Smeland, E. B. A novel putative G-protein-coupled receptor expressed in lung, heart and lymphoid tissue. *FEBS Lett.* **407,** 59–62 (1997).

12. O'Dowd, B. F. *et al.* Discovery of Three Novel G-Protein-Coupled Receptor Genes. *Genomics* **47,** 310–313 (1998).

13. Filardo, E. J., Quinn, J. A., Frackelton, A. R. & Bland, K. I. Estrogen Action Via the G Protein-Coupled Receptor, GPR30: Stimulation of Adenylyl Cyclase and cAMP-Mediated Attenuation of the Epidermal Growth Factor Receptor-to-MAPK Signaling Axis. *Mol. Endocrinol.* **16,** 70–84 (2002).

14. Kanda, N. & Watanabe, S. 17β-Estradiol Enhances the Production of Nerve Growth Factor in THP-1-Derived Macrophages or Peripheral Blood Monocyte-Derived Macrophages. *J. Invest. Dermatol.* **121,** 771–780 (2003).

15. Howell, A. *et al.* Fulvestrant, Formerly ICI 182,780, Is as Effective as Anastrozole in Postmenopausal Women With Advanced Breast Cancer Progressing After Prior Endocrine Treatment. *J. Clin. Oncol.* **20,** 3396–3403 (2002).

16. Osborne, C. K. Tamoxifen in the Treatment of Breast Cancer. *N. Engl. J. Med.* **339,** 1609–1618 (1998).

17. Dhingra, K. Antiestrogens – Tamoxifen, SERMs and Beyond. *Invest. New Drugs* **17,** 285–311 (1999).

18. Wardell, S. E., Marks, J. R. & McDonnell, D. P. The turnover of estrogen receptor α by the selective estrogen receptor degrader (SERD) fulvestrant is a saturable process that is not required for antagonist efficacy. *Biochem. Pharmacol.* **82,** 122–130 (2011).

19. Bologa, C. G. *et al.* Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2,** 207–212 (2006).

20. Dennis, M. K. *et al.* In vivo Effects of a GPR30 Antagonist. *Nat. Chem. Biol.* **5,** 421–427 (2009).

21. Shoichet, B. K. & Kobilka, B. K. Structure-based drug screening for G-protein-coupled receptors. *Trends Pharmacol. Sci.* **33,** 268–272 (2012).

22. Katritch, V. & Abagyan, R. GPCR agonist binding revealed by modeling and crystallography. *Trends Pharmacol. Sci.* **32,** 637–643 (2011).

23. Katritch, V., Rueda, M., Lam, P. C.-H., Yeager, M. & Abagyan, R. GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins* **78,** 197–211 (2010).

24. Sanders, M. P. A. *et al.* Snooker: A Structure-Based Pharmacophore Generation Tool Applied to Class A GPCRs. *J. Chem. Inf. Model.* **51,** 2277–2292 (2011).

25. Kratochwil, N. A. *et al.* G protein-coupled receptor transmembrane binding pockets and their applications in GPCR research and drug discovery: a survey. *Curr. Top. Med. Chem.* **11,** 1902–1924 (2011).

26. Tang, H., Wang, X. S., Hsieh, J.-H. & Tropsha, A. Do crystal structures obviate the need for theoretical models of GPCRs for structure-based virtual screening? *Proteins* **80,** 1503–1521 (2012).

27. Wass, M. N. & Sternberg, M. J. E. Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins Struct. Funct. Bioinforma.* **77,** 147–151 (2009).

28. Kalinina, O. V., Gelfand, M. S. & Russell, R. B. Combining specificity determining and conserved residues improves functional site prediction. *BMC Bioinformatics* **10,** 174 (2009).

29. Capra, J. A. & Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics* **23,** 1875–1882 (2007).

30. Madabushi, S. *et al.* Evolutionary Trace of G Protein-coupled Receptors Reveals Clusters of Residues That Determine Global and Class-specific Functions. *J. Biol. Chem.* **279,** 8126–8132 (2004).

31. Levit, A., Barak, D., Behrens, M., Meyerhof, W. & Niv, M. Y. Homology model-assisted elucidation of binding sites in GPCRs. *Methods Mol. Biol. Clifton NJ* **914,** 179–205 (2012).

32. Sanders. ss-TEA: Entropy based identification of receptor specific ligand binding residues from a multiple sequence alignment of class A GPCRs. *BMC Bioinformatics* **12,** 332–343 (2011).

33. Capra, J. A., Laskowski, R. A., Thornton, J. M., Singh, M. & Funkhouser, T. A. Predicting Protein Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D Structure. *PLoS Comput Biol* **5,** e1000585 (2009).

34. Méndez-Luna, D. *et al.* Deciphering the GPER/GPR30-agonist and antagonists interactions using molecular modeling studies, molecular dynamics, and docking simulations. *J. Biomol. Struct. Dyn.* **33,** 2161–2172 (2015).

35. Arnatt, C. K. & Zhang, Y. G Protein-Coupled Estrogen Receptor (GPER) Agonist Dual Binding Mode Analyses Toward Understanding of Its Activation Mechanism: A Comparative Homology Modeling Approach. *Mol. Inform.* **32,** 647–658 (2013).

36. Zhang, J., Yang, J., Jang, R. & Zhang, Y. GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. *Structure* **23,** 1538–1549 (2015).

37. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2,** 1511–1519 (1993).

38. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **39,** W270–W277 (2011).

39. Merz, K. M. Limits of Free Energy Computation for Protein−Ligand Interactions. *J. Chem. Theory Comput.* **6,** 1769–1776 (2010).

40. Wan, S., Knapp, B., Wright, D. W., Deane, C. M. & Coveney, P. V. Rapid, Precise, and Reproducible Prediction of Peptide–MHC Binding Affinities from Molecular Dynamics That Correlate Well with Experiment. *J. Chem. Theory Comput.* **11,** 3346–3356 (2015).

41. Li, Y., Hou, T. & Goddard III, W. Computational Modeling of Structure-Function of G Protein-Coupled Receptors with Applications for Drug Design. *Curr. Med. Chem.* **17,** 1167–1180 (2010).

42. Smith, R. D. *et al.* CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. *J. Chem. Inf. Model.* (2015). doi:10.1021/acs.jcim.5b00387

43. Armon, A., Graur, D. & Ben-Tal, N. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information1. *J. Mol. Biol.* **307,** 447–463 (2001).

44. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38,** W529-533 (2010).

45. Grosdidier, A., Zoete, V. & Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *J. Comput. Chem.* **32,** 2149–2159 (2011).

46. Dundas, J. *et al.* CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34,** W116-118 (2006).

47. Hernandez, M., Ghersi, D. & Sanchez, R. SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Res.* **37,** W413–W416 (2009).

48. Consortium, T. U. UniProt: a hub for protein information. *Nucleic Acids Res.* **43,** D204–D212 (2015).

49. Boeckmann, B. *et al.* Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *Comptes Rendus Biol.* **328,** 882–899 (2005).

50. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

51. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **52,** 1757–1768 (2012).

52. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30,** 1545–1614 (2009).

53. Haberthür, U. & Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *J. Comput. Chem.* **29,** 701–715 (2008).