

# Locating ligand binding sites in G-protein coupled receptors using combined information from docking and sequence conservation

Ashley Ryan Vidad<sup>1,\*</sup>, Stephen Macaspac<sup>1,\*</sup> and Ho Leung Ng<sup>2</sup>

<sup>1</sup> Department of Chemistry, University of Hawaii at Manoa, Honolulu, Hawaii, United States of America

<sup>2</sup> Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, Kansas, United States of America

\* These authors contributed equally to this work.

## ABSTRACT

GPCRs (G-protein coupled receptors) are the largest family of drug targets and share a conserved structure. Binding sites are unknown for many important GPCR ligands due to the difficulties of GPCR recombinant expression, biochemistry, and crystallography. We describe our approach, ConDockSite, for predicting ligand binding sites in class A GPCRs using combined information from surface conservation and docking, starting from crystal structures or homology models. We demonstrate the effectiveness of ConDockSite on crystallized class A GPCRs such as the beta2 adrenergic and A2A adenosine receptors. We also demonstrate that ConDockSite successfully predicts ligand binding sites from high-quality homology models. Finally, we apply ConDockSite to predict the ligand binding sites on a structurally uncharacterized GPCR, GPER, the G-protein coupled estrogen receptor. Most of the sites predicted by ConDockSite match those found in other independent modeling studies. ConDockSite predicts that four ligands bind to a common location on GPER at a site deep in the receptor cleft. Incorporating sequence conservation information in ConDockSite overcomes errors introduced from physics-based scoring functions and homology modeling.

**Subjects** Biochemistry, Bioinformatics, Biophysics, Computational Biology, Pharmacology

**Keywords** Ligand binding sites, GPCR, GPER, Drug binding sites, Bioinformatics, Modeling, Docking, Structure conservation, Molecular evolution

## INTRODUCTION

GPCRs (G-protein coupled receptors) are the largest family of drug targets and the targets of >30% of all drugs. Because they are membrane proteins with flexible and dynamic structures, biochemical and crystallography experiments are difficult. Only ~87 GPCRs out of ~800 in the human genome have been crystallized despite their great pharmacological importance. GPCR homology modeling remains challenging due to conformational flexibility and the abundance of flexible loops (*Lai et al., 2017*). Crystal structures have shown that the large majority of ligands bind in the large central, extracellular cavity of GPCRs, but the specific binding sites in the cavity can vary widely between different

Submitted 16 September 2020

Accepted 6 September 2021

Published 24 September 2021

Corresponding author

Ho Leung Ng, [hng@ksu.edu](mailto:hng@ksu.edu)

Academic editor

Pedro Silva

Additional Information and  
Declarations can be found on  
page 15

DOI [10.7717/peerj.12219](https://doi.org/10.7717/peerj.12219)

© Copyright

2021 Vidad et al.

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

ligands even for the same or closely related receptors (*Wacker, Stevens & Roth, 2017; Chan et al., 2019*).

Various computational approaches have been used to predict ligand binding sites in G-protein coupled receptors. Traditional docking methods compute the lowest energy pose 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 (*Katritch et al., 2010; Katritch & Abagyan, 2011; Shoichet & Kobilka, 2012; Weiss et al., 2016; Lim et al., 2018*). These methods have been used to identify ligand binding sites and build pharmacophores for GPCRs (*Kratochwil et al., 2011; Sanders et al., 2011; Tang et al., 2012*), but the lack of diverse GPCR crystal structures presents serious challenges to using docking methods for identification of ligand binding sites. In particular, few crystal structures of non-class A GPCRs have been determined. Moreover, homology models usually cannot be used to identify ligand binding sites or for docking without extensive optimization, such as with advanced molecular dynamics sampling methods (*Katritch et al., 2010; Lai et al., 2017; Zou, Ewalt & Ng, 2019*). An underappreciated feature that can be used to predict ligand binding sites is surface or sequence conservation. Binding sites for particular ligands are often conserved, and systematic sequence variation can encode ligand specificity (*Capra & Singh, 2007; Kalinina, Gelfand & Russell, 2009; Wass & Sternberg, 2009*). While highly conserved receptors often share similar ligand binding sites, such direct relationships often do not apply between less conserved receptors. Yet, the massive abundance of genomic data for GPCRs can provide strong constraints for possible ligand binding sites even without chemical or structural information (*Madabushi et al., 2004; Sanders, 2011; Levit et al., 2012*). The binding sites for synthetic, non-physiological ligands can also be identified as they often share some or even most of their binding sites with physiological ligands (*Wacker, Stevens & Roth, 2017*). However, binding site conservation information alone usually cannot predict ligand binding sites, as often, very large parts of the receptor are highly conserved, including key structural elements.

There has been less research on methods that combine information from chemical interactions, geometric surface analysis, and bioinformatics. Hybrid strategies, such as Concavity (*Capra et al., 2009*), 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 (*Capra & Singh, 2007*), and employs geometric criteria of size and shape. Here, we describe a new hybrid strategy we have developed, called ConDockSite, to predict ligand binding sites from combined information from surface conservation and docking calculations. We compare our results with those previously published using purely docking-based and other hybrid methods (*Arnatt & Zhang, 2013; Méndez-Luna et al., 2015*). ConDockSite is not intended to be used for docking, *i.e.*, predicting ligand binding poses, which are highly sensitive to small structural details in crystal structures. We demonstrate the effectiveness of ConDockSite for identifying ligand binding sites for the two best characterized class A GPCRs with known crystal structures, the  $\beta_2$  adrenergic

and A2A adenosine receptors. We then demonstrate the effectiveness of ConDockSite with high quality GPCR homology models.

Finally, we apply ConDockSite to predict the hypothetical binding sites of four ligands to the less characterized class A G-protein coupled estrogen receptor (GPER, formerly known as GPR30), a membrane-bound estrogen receptor. GPER is proposed to mediate rapid estrogen-associated effects, cAMP regeneration, and nerve growth factor expression (Kvingedal & Smeland, 1997; Carmeci et al., 1997; O'Dowd et al., 1998; Filardo et al., 2002; Kanda & Watanabe, 2003). GPER is known to bind estradiol and the estrogen receptor inhibitors, tamoxifen and fulvestrant, that are used to treat breast cancer (Fig. S1). Recently, GPER-specific ligands G1 and G15 were discovered (Bologa et al., 2006; Dennis et al., 2009). G1 and G15 are structurally similar, differing by only an acetyl group. G1 is an agonist, whereas G15 is an antagonist. No crystal structure of GPER is available, and details of ligand binding are unknown. ConDockSite predictions can be tested experimentally by measuring the effects of mutagenesis of predicted ligand binding sites on ligand binding. Such efforts should be straightforward given our previous publication describing methods for recombinant expression and ligand binding assays for GPER (Souza et al., 2019).

## METHODS

### Protein surface conservation

GPCR protein sequences were acquired from the SwissProt database (Boeckmann et al., 2005). For this study, we chose the protein sequences manually for consistency and reproducibility. For the A2A adenosine receptor, the protein sequences aligned were from *Homo sapiens*, *Canis familiaris*, *Xenopus tropicalis*, *Myotis davidii*, *Loxodonta africana*, *Gallus gallus*, *Anolis carolinensis*, *Oncorhynchus mykiss*, *Ailuropoda melanoleuca*, and *Alligator mississippiensis*. For the  $\beta_2$  adrenergic receptor, the protein sequences aligned were from *Homo sapiens*, *Oncorhynchus mykiss*, *Myotis brandtii*, *Callorhynchus milii*, *Ophiophagus hannah*, *Canis familiaris*, *Loxodonta africana*, *Ailuropoda melanoleuca*, *Ficedula albicollis*, and *Xenopus laevis*. GPER protein sequences aligned were from diverse species: *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Macaca mulatta*, *Danio rerio*, and *Micropogonias undulatus*. Sequences were chosen to represent a diverse range of animal species. Sequences for the other receptors studied were chosen automatically by ConSurf from the NCBI "NR" non-redundant database. We obtained qualitatively similar results when using ConSurf to automatically select protein sequences for analysis. Multiple sequence alignment files were submitted to ConSurf (<https://consurf.tau.ac.il>) (Armon, Graur & Ben-Tal, 2001; Ashkenazy et al., 2010). 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 the organisms (Fig. S2). Values from ConSurf were mapped onto the receptor surface with Chimera (Pettersen et al., 2004).

### Homology modeling and docking

The crystal structures for the A2A adenosine receptor and the  $\beta_2$  adrenergic receptor were acquired from the RCSB protein data bank. The crystal structures used were of the  $\beta_2$

adrenergic receptor bound to the agonist, epinephrine (PDB 4ldo), the agonist, BI-167107 (PDB 3p0g), the inverse agonist, carazolol (PDB 2rh1), and the inverse agonist, ICI 118,551 (PDB 3ny8). The crystal structures used were of the A2A adenosine receptor bound to the agonist, adenosine (PDB 2ydo), the agonist, CGS21680 (PDB 4ug2), the inverse agonist, ZM241385 (PDB 5k2a), and the inverse agonist, compound 12X (PDB 5iub). The crystal structures of the mu opioid, serotonin 5HT2B, dopamine D2 with haloperidol, dopamine D2 with risperidone, ghrelin, histamine H1, and muscarinic M1 receptors were taken from PDB entries 5c1m, 6drz, 6luq, 6cm4, 6ko5, 3rze, and 5cxv. Structures were prepared for docking with Chimera by removing extraneous chains, solvent atoms, and bound ligands with the DockPrep protocol. Ligands were docked into receptors with SwissDock (<http://www.swissdock.ch>) (Grosdidier, Zoete & Michielin, 2011a).

The homology models for all the receptors were generated by I-TASSER (<https://zhanglab.dcmf.med.umich.edu/I-TASSER>) (Yang *et al.*, 2015). I-TASSER generates composite homology models using multiple crystal structures. Only the predicted 7-transmembrane regions were input to I-TASSER for modeling. Models with >60% sequence identity were excluded as modeling templates so that I-TASSER would not simply retrieve the corresponding crystal structures for modeling. The PDB crystal structures used for modeling the A2A receptor were 6oij, 6e59, and 6me6. The PDB crystal structures used for modeling the  $\beta$ 2 adrenergic receptor were 6me8, 7bts, 5zbh, 6oij, 6kuw, 6ibl, and 2ks9. The PDB crystal structures used for modeling the mu opioid receptor were 4n6h and 4djh. The PDB crystal structures used for modeling the serotonin 5HT2B receptor were 6me6, 6kuw, 4iaq, 5wiv, 6a93, 5zbh, 6kp6, 7lcy, 6a94, and 6vdp. The PDB crystal structures used for modeling the dopamine D2 receptor were 6kux, 6kp6, and 6kuw. The PDB crystal structures used for modeling the ghrelin receptor were 4n6h and 6e59. The PDB crystal structures used for modeling the histamine H1 receptor were 6oij, 6kp6, 6me6, and 6kux. The PDB crystal structures used for modeling the muscarinic M1 receptor were 6me6, 6kp6, and 6me2.

The crystal structure of GPER has not yet been determined. We created a homology model using GPCR I-TASSER (Iterative Threading Assembly Refinement), the most accurate homology modeling software customized for GPCRs (<https://zhanglab.dcmf.med.umich.edu/GPCR-I-TASSER>) (Zhang *et al.*, 2015). GPCR I-TASSER modeled the GPER structure using template fragments automatically selected from the closest related GPCR crystal structures (CCR5: PDB 4mbs, sphingosine 1-phosphate: PDB 3v2y, CXCR4: PDB 3odu, delta opioid: PDB 4n6h). The homology model was validated with ERRAT (Colovos & Yeates, 1993). Coordinates for E2, G1, G15, and tamoxifen were downloaded from the ZINC ligand database (Irwin *et al.*, 2012) and submitted to SwissDock (Grosdidier, Zoete & Michielin, 2011a) for docking. SwissDock is a web interface to the EADock DSS (Grosdidier, Zoete & Michielin, 2011b) engine, which performs blind, global (does not require targeting of a particular surface) docking using the physics-based CHARMM22 force field (Brooks *et al.*, 2009). The “FullFitness Score” calculated by SwissDock using clustering and the FACTS implicit solvent model (Habberthür & Caflisch, 2008) was used as the “Energy Score” for our calculations.

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

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

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

**Scheme 1** Calculation of combined ConDockSite scores for ligand binding sites. The Conservation Score is calculated over the  $n$  residues in a binding site, indexed by  $k$ .

Full-size  DOI: 10.7717/peerj.12219/fig-8

SwissDock was chosen both for its high effectiveness as well as ease of use by students. For consistency, we performed all docking studies in this paper with SwissDock although we obtained qualitatively similar docking results with AutoDock Vina, the most popular docking software, in our preliminary studies.

### Combined analysis

SwissDock poses were analyzed for ligand binding sites near highly conserved surfaces. Ligand binding surfaces included residues with atoms 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 (Scheme 1). The combined ConDockSite 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 ConDockSite score is associated with a more probable ligand binding site. Binding sites predicted by ConDockSite results were compared with those predicted by CASTp (<http://sts.bioe.uic.edu/castp/index.html>) (Dundas *et al.*, 2006), SiteHound (Hernandez, Ghersi & Sanchez, 2009), and Concavity (Capra *et al.*, 2009). For CASTp, SiteHound, and Concavity, ligand binding pockets were defined as residues within 4 Å of the selected probe/cluster.

### Binding site prediction benchmarks

The distances between the centers of mass of the predicted and experimental ligand coordinates served as a benchmark of comparison for the sites predicted by the ConDockSite scoring function. This is a more appropriate measure than the ligands RMSD used for pose comparison as ConDockSite is intended to predict binding sites rather than precise binding poses.

## RESULTS

We developed ConDockSite to predict ligand binding pockets using information from surface conservation and docking calculations. ConDockSite uses a simple scoring function that is the product of surface conservation scores from ConSurf (Armon, Graur & Ben-Tal, 2001) and docking scores from SwissDock (Grosdidier, Zoete & Michielin, 2011a). A highly negative ConDockSite score is associated with a more probable ligand binding site.

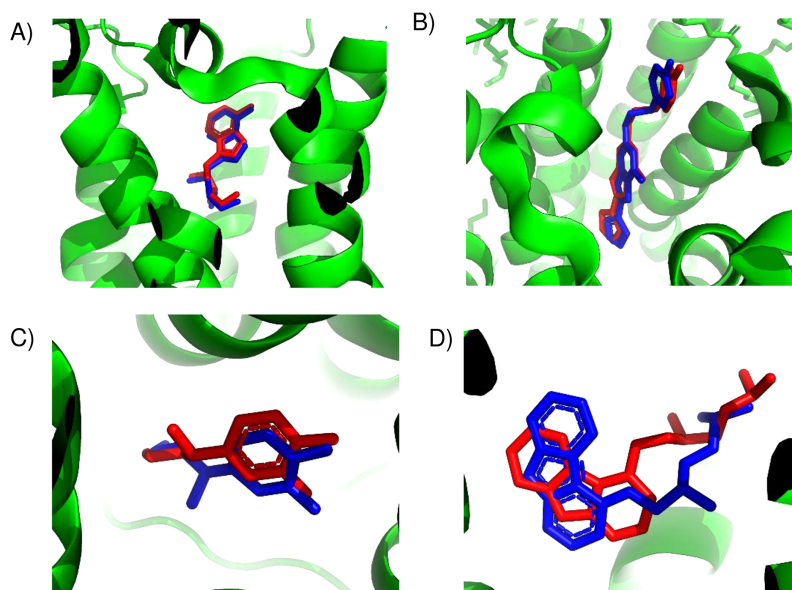
The A2A adenosine and β2 adrenergic receptors are by far the most heavily studied GPCRs by crystallography. First, we show that the ConSurf surface conservation

information alone is inadequate for identifying binding sites. Conservation varies greatly associated throughout the two receptors and is not directly associated with the ligand binding sites (Fig. S3). Many of the conserved regions are instead associated with the internal packing of the seven transmembrane helices. Next, we show that geometric binding pocket analyses, such as CASTp (Dundas *et al.*, 2006), often work poorly for GPCRs. For the two receptors, CASTp merely predicts the entire GPCR central cavity as the ligand binding site (Fig. S4). It fails to localize the ligands to specific parts of the central cavity.

Finally, we used the crystal structures of the two receptors as positive control experiments to validate the effectiveness of ConDockSite for predicting ligand binding sites. For both receptors, we performed cross-docking of an agonist and inverse agonist against a crystal structure of the receptor bound to a different agonist or inverse agonist: ligands were cross-docked rather than self-docked into its own crystal structure. Self-docking is often trivial for modern docking methods and thus, was not studied. Docking was performed with SwissDock, which has demonstrated high accuracy in docking ligands into receptors without prior knowledge of the binding site (also known as global or blind docking), and also includes a user-friendly web interface suitable for students (Grosdidier, Zoete & Michielin, 2011a). SwissDock docking results were then ranked by the ConDockSite scoring function (Table S1). Residues within 3.5 Å of the highest scoring predicted ligand sites were compared with the binding surfaces associated with the ligand poses in the crystal structures.

As a convenient metric for the distances between predicted and experimental ligand binding sites, we use the distances between the ConDockSite-scored ligand poses and those observed in the crystal structures. For the A2A adenosine receptor, the agonist, adenosine, was cross-docked into the crystal structure of the receptor with the agonist CGS21680 (PDB 4ug2). The highest ConDockSite-ranked pose for adenosine within the A2A adenosine receptor was within 0.4 Å of the ligand position (distance between centers of mass) in the crystal structure (PDB 2ydo). (Fig. 1A). The ConDockSite-predicted binding site had a ConSurf conservation score of 0.86 and is essentially the same as the experimental binding site. The inverse agonist, ZM241385, was cross-docked into the crystal structure of the receptor with the inverse agonist, compound 12X (PDB 5iub) (Segala *et al.*, 2016). The highest ranked site for ZM241385 within the A2A adenosine receptor was within 1.0 Å of the ligand's position in the crystal structure (PDB 5k2a). In this top pose, ZM241385 is found within the same binding site as that observed in the crystal structure (Fig. 1B), with a ConSurf conservation score of 0.86. For both adenosine and ZM241385, the ConDockSite-predicted site corresponded to the top site predicted by SwissDock. In this case, docking alone was adequate to identify the binding site. The correct binding site also had the highest surface conservation. The A2A adenosine receptor structures are easy tests for ligand prediction, passed by both SwissDock and ConDockSite.

For the  $\beta$ 2 adrenergic receptor, the agonist, epinephrine, was cross-docked into the crystal structure of the receptor with the agonist, BI-167107 (PDB 3p0g) (Rasmussen *et al.*, 2011). The highest ranked pose for epinephrine within the  $\beta$ 2 adrenergic receptor was



**Figure 1** Predicted and experimental ligand binding sites in A2A adenosine and  $\beta$ 2 adrenergic receptors. Superposition of crystal structure with ligand bound (red) with ConDockSite predicted pose (blue). (A) Adenosine with A2A receptor. (B) ZM241385 with A2A receptor. (C) Epinephrine with  $\beta$ 2 adrenergic receptor. (D) Carazolol with  $\beta$ 2 adrenergic receptor.

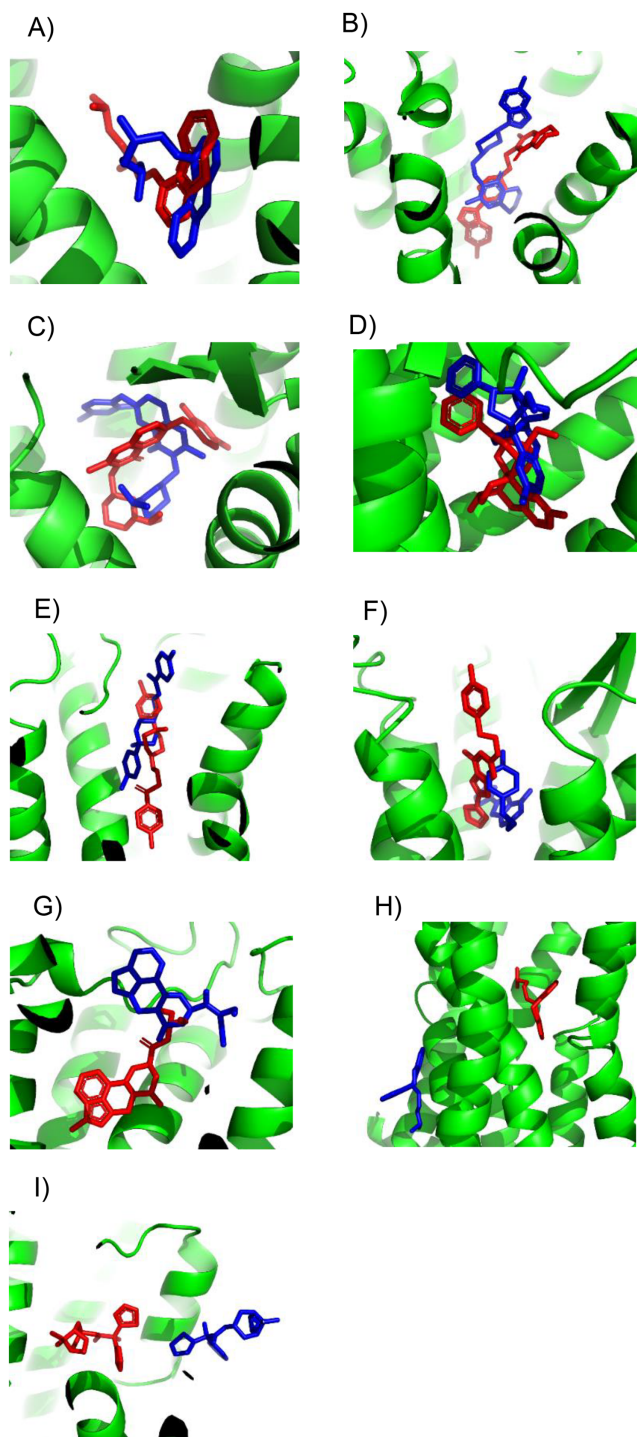
Full-size  DOI: [10.7717/peerj.12219/fig-1](https://doi.org/10.7717/peerj.12219/fig-1)

within 0.4 Å of the ligand position within the crystal structure (PDB 4ldo). This binding site for epinephrine was again essentially the same as the observed binding pocket (Fig. 1C) and had a ConSurf conservation score of 0.85. This site was also scored the highest by SwissDock. For the  $\beta$ 2 adrenergic receptor, the inverse agonist, carazolol, was cross-docked into the crystal structure of the receptor with the inverse agonist ICI 118,551 (PDB 3ny8) (Wacker *et al.*, 2010). The highest ranked pose for carazolol within the  $\beta$ 2 adrenergic receptor was within 1.0 Å of the ligand's position within the crystal structure (PDB 2rh1). This binding site for carazolol was essentially the same as that in the crystal structure (Fig. 1D). This binding pocket has a ConSurf conservation score of 0.78. The correct carazolol binding site was scored the third highest by SwissDock. The use of surface conservation information allowed selection of the proper binding site. The extremely accurate placement of both agonists and inverse agonists for the A2A adenosine and the  $\beta$ 2 adrenergic receptors demonstrates ConDockSite's effectiveness when crystal structures are available.

Unfortunately, crystal structures are not available for most GPCRs. The most valuable use of ConDockSite is to predict drug binding sites in homology models. By using surface conservation information, ConDockSite is less sensitive to homology model inaccuracies than other ligand binding site prediction methods that are based purely on geometric methods. To demonstrate the ability of ConDockSite to work with homology models, we created models of eight GPCRs, the  $\beta$ 2 adrenergic, A2A adenosine, 5HT2B serotonin, mu opioid receptors, D2 dopamine, ghrelin, H1 histamine, and M1 muscarinic receptors, while excluding their known X-ray structures from the templates used for

modeling. We used I-TASSER ([Yang et al., 2015](#)) for homology modeling which does not use GPCR-specific structural constraints but allows for custom selection of templates. I-TASSER created fairly accurate models of all eight receptors, with RMSDs across C $\alpha$  atoms between the models and crystal structures ranging from a best of 1.5 Å for the histamine H1 receptor (PDB 3rze) to a respectable 2.1 Å for the muscarinic M1 (PDB 5cxv) receptors. We used ConDockSite to predict the binding sites of the  $\beta$ 2 adrenergic receptor with carazolol, A2A adenosine receptor with ZM241385, 5HT2B serotonin receptor with methysergide, mu opioid receptor with BU72, D2 dopamine receptor with risperidone and haloperidol, ghrelin receptor with the antagonist Compound 21, histamine H1 receptor with doxepin, and muscarinic M1 receptor with tiotropium. ConDockSite performed best with the  $\beta$ 2 adrenergic receptor homology model, with only 2.3 Å between the centers of mass of the predicted and crystal structure ligand poses ([Fig. 2](#), [Table S2](#)), supporting the prediction of very similar binding pockets. The binding pocket was predicted correctly although the ligand pose was inaccurate. There was a weak correspondence between ConDockSite performance with the accuracy of the homology models. The H1 histamine receptor model was the best homology model, with RMSD of 1.5 Å between the model and the crystal structure C $\alpha$  atoms. However, ConDockSite was unable to correctly predict the binding site for the antagonist doxepin, which was predicted >10 Å from the crystal structure site. The poor performance was due to the inaccurate modeling of a loop containing residues 162–165 over the top of the ligand binding site as well as multiple errors in the conformation of side chains in the transmembrane helices. The D2 dopamine receptor model was the second-best homology model, with RMSD of 1.6–1.8 Å between the model and the crystal structure C $\alpha$  atoms (PDB 6luq bound to haloperidol and 6cm4 bound to risperidone). ConDockSite predicted the binding site of risperidone well, only 2.4 Å from the crystal structure site, but less well for haloperidol, which was predicted 4.3 Å from the crystal structure site. With this medium level of accuracy, the ligand poses are generally incorrect, but most of the binding residues in the predicted pockets are the same as those in the crystal structures, supporting successful ConDockSite predictions. The ghrelin receptor homology model was also predicted well with RMSD of 1.7 Å between the model and the crystal structure C $\alpha$  atoms (PDB 6ko5). The ghrelin antagonist, Compound 21, was predicted 3.1 Å away from the crystal structure site. The predicted ligand binding sites for the A2A adenosine and mu opioid ligands (PDB 5c1m) are 3–4 Å away from the crystal structures. ConDockSite performs less well with the serotonin 5HT2B receptor (PDB 6drz) where the distance between the predicted and actual ligand binding sites was 7.0 Å. In the serotonin 5HT2B receptor structure, the ligand, methysergide, binds very deep in the receptor. Serious errors in homology modeling of the 5HT2B receptor side chains make it difficult or impossible to dock the ligand into the deep, restricted binding site. ConDockSite also fails with the muscarinic M1 receptor (PDB 5cxv) where the distance between the predicted and actual tiotropium binding sites was >10 Å. Overall, our results with ConDockSite are consistent with benchmark modeling results that show that GPCR homology models of modest accuracy from templates with low sequence identity are still sometimes useful for docking and virtual screening ([Lim et al., 2018](#); [Costanzi et al.,](#)





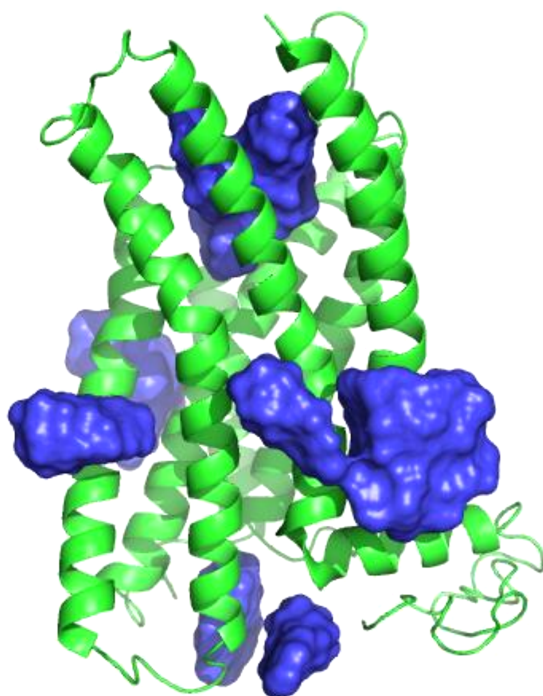
**Figure 2** Predicted and experimental ligand binding sites for homology models of eight GPCRs. Models are shown in order of best to worst predictions. Superposition of crystal structure with ligand bound (red) with ConDockSite predicted pose (blue). (A) Carazolol with  $\beta_2$  adrenergic receptor. (B) Risperidone with dopamine D2. (C) Compound 21 with ghrelin receptor. (D) BU72 with mu opioid. (E) Haloperidol with dopamine D2. (F) ZM241385 with A2A adenosine. (G) Methysergide with serotonin 5HT2B. (H) Doxepin with histamine H1. (I) Tiotropium with muscarinic M1.

Full-size  DOI: 10.7717/peerj.12219/fig-2

2019). In some of the failed cases (histamine H1, muscarinic M1, and serotonin 5HT2B), while the homology models were sometime accurate at the level of backbone atoms in the 7<sup>th</sup> region, the loops were modeled poorly and disrupted the modeled ligand binding pocket. In other cases, the homology model backbones were modeled well but differences in the side chain conformations disrupted the integrity of the ligand binding sites. In these cases, the homology models are not accurate enough for docking or ConDockSite. In retrospect, ConDockSite sometimes performed better with alternative homology models generated by I-TASSER with lower RMSD measures but with alternative loop placements.

After demonstrating the applicability of ConDockSite for homology models, we applied ConDockSite to predict the binding sites in a GPCR, GPER (G-protein coupled estrogen receptor), which has not yet been crystallized. GPER is of great interest to us and other researchers due to its unusual physiological and pharmacological roles in estrogen-related biology. As there is no experimental data on the ligand binding sites in GPER, we cannot validate our predictions. They should be considered speculative at this point, but we hope they can guide future experiments. To predict the potential ligand binding sites in GPER, we first created a homology model using GPCR-I-TASSER ([Zhang et al., 2015](#)). GPCR-I-TASSER has been shown to be among the most accurate GPCR homology modeling software package. We used the generic I-TASSER in our validation studies because of its fine-grained options for template selection that are lacking in GPCR-I-TASSER. Because GPCR-I-TASSER uses GPCR-specific structural constraints, it is expected to outperform the generic I-TASSER ([Zhang et al., 2015](#)). GPCR-I-TASSER identified the closest matching crystal structure to GPER to be the CCR5 chemokine receptor (PDB 4mbs) with 23% sequence identity. GPCR-I-TASSER used this crystal structure along with 9 other GPCR 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. S5](#)) and has a Ramachandran plot with 95.6% of residues, excluding glycine, in preferred regions ([Fig. S6](#)). The primary differences are in the extracellular loop between helices 4 and 5 (ECL2) and the intracellular loops between helices 5 and 6 (ICL3), and after helix 7. These two intracellular loops are predicted by ERRAT ([Colovos & Yeates, 1993](#)) to be the least reliable based on the likelihood of atom pair type interactions from high-resolution crystal structures ([Fig. S7](#)).

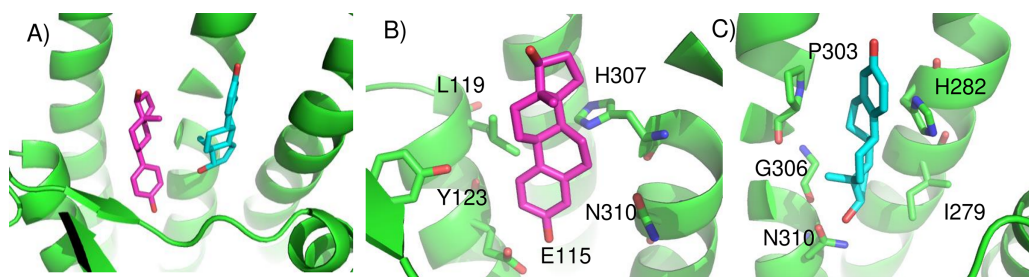
Using the SwissDock server ([Grosdidier, Zoete & Michielin, 2011a](#)), we docked structures of the four ligands E2, G1, G15, and tamoxifen ([Fig. S1](#)) to the homology model of GPER. The docked sites from SwissDock, including those that were scored the highest, were located throughout the receptor surface and thus were considered mostly nonviable ([Fig. 3](#)). The shortcomings of a purely physics-based scoring function such as that used by SwissDock in predicting ligand binding are not surprising given the lack of an experimental crystal structure and well-known limitations of current homology modeling and docking methodology ([Li, Hou & Goddard, 2010](#); [Merz, 2010](#); [Wan et al., 2015](#); [Smith et al., 2016](#)).



**Figure 3** E2 binding sites calculated by SwissDock. E2 poses are in blue. The top of the figure corresponds to the extracellular face of GP62. [Full-size !\[\]\(ba1b80118482ccef74a5d718ca4d7242\_img.jpg\) DOI: 10.7717/peerj.12219/fig-3](https://doi.org/10.7717/peerj.12219/fig-3)

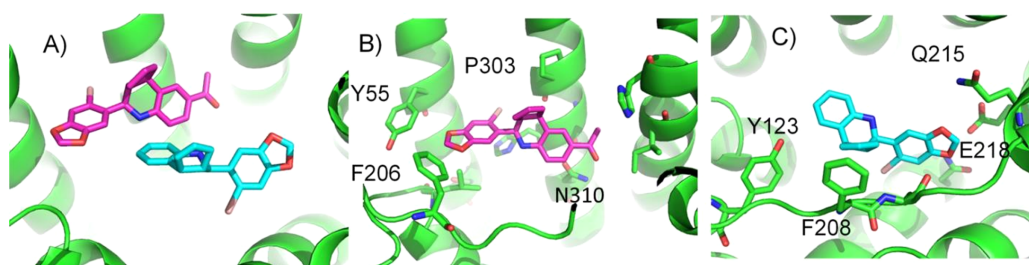
We then ranked all ligand binding sites generated by SwissDock using the combined ConDockSite score. For all four ligands, the ConDockSite score identified one or two ligand binding sites that clearly outscored (more negative) other candidates (Table S1). ConDockSite identified the same approximate binding site for all four ligands, although this was not an explicit criterion in the calculations (Fig. S8). The average ConSurf conservation score across the four ligand binding sites is 0.82 (1.0 represents complete conservation), indicating that the site is highly but not completely conserved. The binding site is located deep in the receptor cleft, although depth 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 ConDockSite sites are physically reasonable.

We found two promising potential binding sites for E2 in GP62. The two sites are 4.4 Å apart, located deep in the receptor cleft (Fig. 4). 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 GP62s 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. This binding site approximately corresponds to that predicted by Lappano et al using docking (Lappano et al., 2010). In the second ranked pose, the inward pointing A-ring hydroxyl group of E2 makes a hydrogen



**Figure 4** 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.

Full-size [DOI: 10.7717/peerj.12219/fig-4](https://doi.org/10.7717/peerj.12219/fig-4)



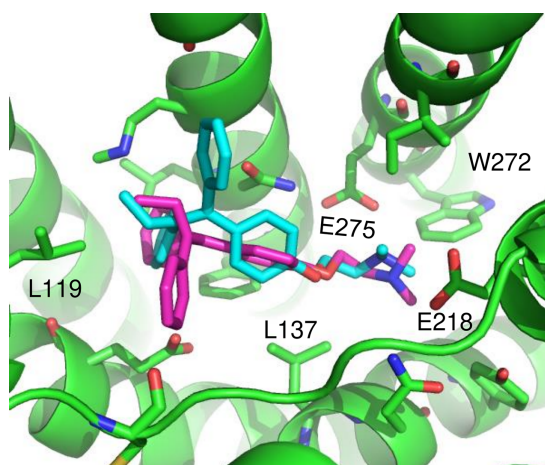
**Figure 5** Predicted G1 and G15 binding sites in GPER. (A) The highest scoring docking poses for G1 (maroon) and G15 (cyan). (B) Receptor-ligand interactions for G1. (C) Receptor-ligand interactions for G15.

Full-size [DOI: 10.7717/peerj.12219/fig-5](https://doi.org/10.7717/peerj.12219/fig-5)

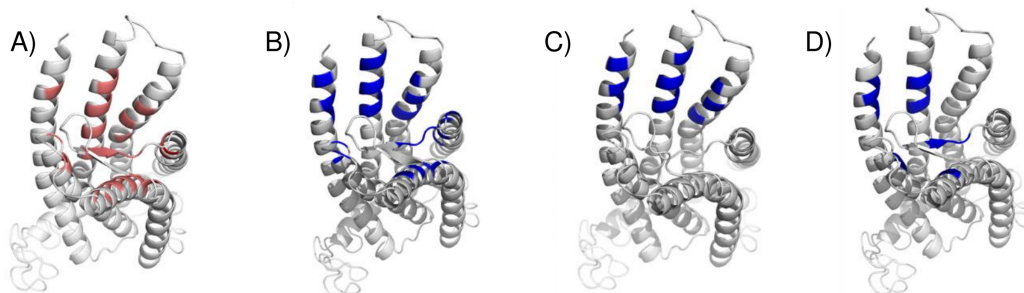
bond with N310. This pose is in a less hydrophobic environment, contacting primarily H282 and P303.

ConDockSite predicts that G1 and G15 bind in adjacent but distinct binding sites separated by 2.3 Å. The top predicted binding site for G1 is found within the pocket bound by Y55, L119, F206, Q215, I279, P303, H307, and N310 (Fig. 5). This orientation had the highest conservation score of all predicted binding sites at 0.85. In this pose, N310 makes a long hydrogen bond (3.6 Å N-O distance) with the acetyl oxygen of G1. The predicted binding site for G15 is found within the pocket is surrounded 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. The ConDockSite G1 result correspond to the binding sites predicted by recent studies using docking and molecular dynamics simulations and validated by design and activity testing of new G1 derivatives (Méndez-Luna, Bello & Correa-Basurto, 2016; Martínez-Muñoz et al., 2018).

ConDockSite 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. 6). The conservation score of this orientation is 0.81. Hydrophobic interactions are observed between tamoxifen and non-polar residues L119,



**Figure 6** Predicted tamoxifen binding sites in GPER. The highest scoring docking poses for tamoxifen, pose 1 (maroon) and pose 2 (cyan). [Full-size !\[\]\(b345a1c4255362eec3746050dd71ccac\_img.jpg\) DOI: 10.7717/peerj.12219/fig-6](https://doi.org/10.7717/peerj.12219/fig-6)



**Figure 7** Predicted E2 binding sites by ConDockSite, CASTp, SiteHound, concavity. Ligand binding sites are colored, predicted by (A) ConDockSite. (B) CASTp. (C) SiteHound. (D) Concavity. [Full-size !\[\]\(0c0f8cc6eca4f663c17a652926046967\_img.jpg\) DOI: 10.7717/peerj.12219/fig-7](https://doi.org/10.7717/peerj.12219/fig-7)

Y123, Y142, P303, and F314. Notably, the amine group of tamoxifen makes ionic interactions with E218 and E275.

We compared the GPER ligand binding sites predicted by ConDockSite to those predicted by three other software packages representing different approaches: CASTp (*Dundas et al., 2006*), which analyzes surface geometry, SiteHound (*Hernandez, Ghersi & Sanchez, 2009*), which maps surfaces with a chemical probe, and Concavity (*Capra et al., 2009*), which analyzes surface geometry and conservation (*Fig. 7*). All three methods could identify a ligand binding site very roughly matching that from ConDockSite. In comparison with the ligand binding sites predicted by traditional methods based on surface geometry and conservation (*Fig. 7*), the sites predicted by ConDockSite are more detailed in shape due to the information from chemical interactions from ligand docking. The pocket predicted by ConDockSite 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 ConDockSite site was scored third and is a shallow binding pocket near H52-G58, E275-H282, and R299-H307 (*Fig. 7C*). In contrast, the

Concavity site was smaller and shallower than the ConDockSite site (Fig. 7D). Surprisingly, the site predicted by the simpler CASTp method best matched the ConDockSite site but is also smaller and shallower (Fig. 7B). For proteins such as GPCRs with large, concave binding pockets, geometry-based prediction methods such as Concavity and CASTp can easily identify the general, approximate location of the ligand binding site. However, such methods may have more difficulty recovering the specific, ligand-specific binding site. It is also surprising that ConDockSite more closely matched the results of the geometry-based methods given that ConDockSite does not take surface geometry into account. As described previously, the G1 and G15 binding sites predicted by ConDockSite more closely match those made using docking against very computationally expensive molecular dynamics simulations (Méndez-Luna, Bello & Correa-Basurto, 2016).

## DISCUSSION

The ConDockSite scoring method, incorporating information from both surface conservation and docking binding energy, demonstrated high accuracy in predicting ligand binding sites from the crystal structures of two class A GPCRs, the A2A adenosine and  $\beta$ 2 adrenergic receptors. ConDockSite also successfully predicted the ligand binding sites for many high-quality homology models but failed for models for which loops were modeled poorly and disrupted the ligand binding site. Better homology models with improved loop and side chain placement should allow better ConDockSite performance. ConDockSite was also used to predict viable ligand binding sites for four different GPER ligands. In contrast to more typical geometry-based ligand binding site prediction methods, ConDockSite 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, most likely due to the inadequate accuracy of homology models. 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 (Capra & Singh, 2007; Capra et al., 2009), but we demonstrate here that a simple product scoring function is already effective. The four GPER ligands studied here differ greatly in chemical structure, but the ConDockSite 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.

Earlier 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  $\pi$ - $\pi$  stacking interactions (Arnatt & Zhang, 2013; Méndez-Luna et al., 2015). Figure S9 compares the ConDockSite binding site against that predicted in the molecular dynamics simulation and docking study. The ConDockSite binding site is located deeper in the extracellular cleft; the other proposed site involved more surface-exposed loops. It was 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. More recent studies using better homology models and computationally expensive long time-scale molecular dynamics simulations predict E2, G1, and G15 binding sites that approximately match those predicted by ConDockSite (Lappano *et al.*, 2010; Méndez-Luna, Bello & Correa-Basurto, 2016). The ConDockSite binding site predictions can be tested experimentally by performing site-directed mutagenesis and ligand binding assays.

In summary, the simple ConDockSite hybrid scoring model predicts physically plausible ligand binding sites by combining information from ligand docking and surface conservation. Using multiple orthogonal sources of information partially avoids errors introduced by modeling (Capra *et al.*, 2009). Given a homology model of modest quality, ConDockSite can sometimes accurately predict ligand binding sites. Using this hybrid method, we identified a site in the extracellular cavity of GPER that has the potential to bind four known GPER ligands. Further optimization of hybrid scoring functions and homology modeling methods should yield significantly improved predictions. Extension of this approach may allow analysis of non-class A GPCRs.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was funded by the Victoria S. and Bradley L. Geist Foundation (Ho Leung Ng), NSF CAREER Award 1833181 (Ho Leung Ng), and the Undergraduate Research Opportunities Program at the University of Hawaii at Manoa (Ashley R Vidad, Stephen Macaspac). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Geist Foundation: 1833181.

University of Hawaii at Manoa.

### Competing Interests

Ho Leung Ng. is an Academic Editor for PeerJ.

### Author Contributions

- Ashley Ryan Vidad conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Stephen Macaspac conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ho Leung Ng. conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

Our analysis was performed using structures and software tools that are publicly available in databases (PDB) or web servers (I-TASSER, ConSurf, SwissDock).

Web servers:

ConSurf: <https://consurf.tau.ac.il>

I-TASSER: <https://zhanglab.dcmf.med.umich.edu/I-TASSER>

GPCR-I-TASSER: <https://zhanglab.dcmf.med.umich.edu/GPCR-I-TASSER/>

SwissDock: <http://www.swissdock.ch>

CASTp: <http://sts.bioe.uic.edu/castp/index.html>

PDB files:

beta2 adrenergic receptor: 4ldo, 3p0g, 2rh1, 3ny8

A2A adenosine receptor: 2ydo, 4ug2, 5k2a, 5iub

mu opioid receptor: 5c1m

serotonin 5HT2B receptor: 6drz

dopamine D2 receptor: 6luq, 6cm4

ghrelin receptor: 6ko5

histamine H1 receptor: 3rze

muscarinic M1 receptor: 5cxv.

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.12219#supplemental-information>.

## REFERENCES

- Armon A, Graur D, Ben-Tal N. 2001.** ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *Journal of Molecular Biology* 307(1):447–463 DOI 10.1006/jmbi.2000.4474.
- Arnatt CK, Zhang Y. 2013.** G protein-coupled estrogen receptor (GPER) agonist dual binding mode analyses toward understanding of its activation mechanism: a comparative homology modeling approach. *Molecular Informatics* 32(7):647–658 DOI 10.1002/minf.201200136.
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010.** ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research* 38(Web Server):W529–W533 DOI 10.1093/nar/gkq399.
- Boeckmann B, Blatter M-C, Famiglietti L, Hinz U, Lane L, Roechert B, Bairoch A. 2005.** Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *Comptes Rendus Biologies* 328(10–11):882–899 DOI 10.1016/j.crv.2005.06.001.
- Bologa CG, Revankar CM, Young SM, Edwards BS, Arterburn JB, Kiselyov AS, Parker MA, Tkachenko SE, Savchuck NP, Sklar LA, Oprea TI, Prossnitz ER. 2006.** Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nature Chemical Biology* 2(4):207–212 DOI 10.1038/nchembio775.
- Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W,**



- York DM, Karplus M. 2009. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry* 30(10):1545–1614 DOI 10.1002/jcc.21287.
- Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. 2009. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLOS Computational Biology* 5(12):e1000585 DOI 10.1371/journal.pcbi.1000585.
- Capra JA, Singh M. 2007. Predicting functionally important residues from sequence conservation. *Bioinformatics* 23(15):1875–1882 DOI 10.1093/bioinformatics/btm270.
- Carmeci C, Thompson DA, Ring HZ, Francke U, Weigel RJ. 1997. Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. *Genomics* 45(3):607–617 DOI 10.1006/geno.1997.4972.
- Chan HCS, Li Y, Dahoun T, Vogel H, Yuan S. 2019. New binding sites, new opportunities for GPCR drug discovery. *Trends in Biochemical Sciences* 44(4):312–330 DOI 10.1016/j.tibs.2018.11.011.
- Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science: A Publication of the Protein Society* 2:1511–1519 DOI 10.1002/pro.5560020916.
- Costanzi S, Cohen A, Danfora A, Dolatmoradi M. 2019. Influence of the structural accuracy of homology models on their applicability to docking-based virtual screening: the  $\beta_2$  adrenergic receptor as a case study. *Journal of Chemical Information and Modeling* 59(7):3177–3190 DOI 10.1021/acs.jcim.9b00380.
- Dennis MK, Burai R, Ramesh C, Petrie WK, Alcon SN, Nayak TK, Bologna CG, Leitao A, Brailoiu E, Deliu E, Dun NJ, Sklar LA, Hathaway HJ, Arterburn JB, Oprea TI, Prossnitz ER. 2009. In vivo effects of a GPR30 antagonist. *Nature Chemical Biology* 5(6):421–427 DOI 10.1038/nchembio.168.
- Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. 2006. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research* 34(Web Server):W116–W118 DOI 10.1093/nar/gkl282.
- Filardo EJ, Quinn JA, Frackelton AR, Bland KI. 2002. 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. *Molecular Endocrinology* 16(1):70–84 DOI 10.1210/mend.16.1.0758.
- Grosdidier A, Zoete V, Michielin O. 2011a. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research* 39(suppl):W270–W277 DOI 10.1093/nar/gkr366.
- Grosdidier A, Zoete V, Michielin O. 2011b. Fast docking using the CHARMM force field with EADock DSS. *Journal of Computational Chemistry* 32(10):2149–2159 DOI 10.1002/jcc.21797.
- Haberthür U, Cafilisch A. 2008. FACTS: fast analytical continuum treatment of solvation. *Journal of Computational Chemistry* 29:701–715 DOI 10.1002/(ISSN)1096-987X.
- Hernandez M, Ghersi D, Sanchez R. 2009. SITEHOUND-web: a server for ligand binding site identification in protein structures. *Nucleic Acids Research* 37(Web Server):W413–W416 DOI 10.1093/nar/gkp281.
- Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. 2012. ZINC: a free tool to discover chemistry for biology. *Journal of Chemical Information and Modeling* 52(7):1757–1768 DOI 10.1021/ci3001277.

- Kalinina OV, Gelfand MS, Russell RB. 2009.** Combining specificity determining and conserved residues improves functional site prediction. *BMC Bioinformatics* **10**(1):174  
DOI [10.1186/1471-2105-10-174](https://doi.org/10.1186/1471-2105-10-174).
- Kanda N, Watanabe S. 2003.**  $17\beta$ -Estradiol enhances the production of nerve growth factor in THP-1-derived macrophages or peripheral blood monocyte-derived macrophages. *Journal of Investigative Dermatology* **121**(4):771–780 DOI [10.1046/j.1523-1747.2003.12487.x](https://doi.org/10.1046/j.1523-1747.2003.12487.x).
- Katritch V, Abagyan R. 2011.** GPCR agonist binding revealed by modeling and crystallography. *Trends in Pharmacological Sciences* **32**(11):637–643 DOI [10.1016/j.tips.2011.08.001](https://doi.org/10.1016/j.tips.2011.08.001).
- Katritch V, Rueda M, Lam PC-H, Yeager M, Abagyan R. 2010.** GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins-Structure Function and Bioinformatics* **78**(1):197–211 DOI [10.1002/prot.22507](https://doi.org/10.1002/prot.22507).
- Kratochwil NA, Gatti-McArthur S, Hoener MC, Lindemann L, Christ AD, Green LG, Guba W, Martin RE, Malherbe P, Porter RHP, Slack JP, Winnig M, Dehmlow H, Grether U, Hertel C, Narquizian R, Panousis CG, Kolczewski S, Steward L. 2011.** G protein-coupled receptor transmembrane binding pockets and their applications in GPCR research and drug discovery: a survey. *Current Topics in Medicinal Chemistry* **11**(15):1902–1924  
DOI [10.2174/156802611796391267](https://doi.org/10.2174/156802611796391267).
- Kvingedal AM, Smeland EB. 1997.** A novel putative G-protein-coupled receptor expressed in lung, heart and lymphoid tissue. *FEBS Letters* **407**(1):59–62  
DOI [10.1016/S0014-5793\(97\)00278-0](https://doi.org/10.1016/S0014-5793(97)00278-0).
- Lai JK, Ambia J, Wang Y, Barth P. 2017.** Enhancing structure prediction and design of soluble and membrane proteins with explicit solvent-protein interactions. *Structure* **25**(11):1758–1770 e8  
DOI [10.1016/j.str.2017.09.002](https://doi.org/10.1016/j.str.2017.09.002).
- Lappano R, Rosano C, De Marco P, De Francesco EM, Pezzi V, Maggiolini M. 2010.** Estriol acts as a GPR30 antagonist in estrogen receptor-negative breast cancer cells. *Molecular and Cellular Endocrinology* **320**(1–2):162–170 DOI [10.1016/j.mce.2010.02.006](https://doi.org/10.1016/j.mce.2010.02.006).
- Levit A, Barak D, Behrens M, Meyerhof W, Niv MY. 2012.** Homology model-assisted elucidation of binding sites in GPCRs. *Methods in Molecular Biology (Clifton, N.J.)* **914**:179–205  
DOI [10.1007/978-1-62703-023-6](https://doi.org/10.1007/978-1-62703-023-6).
- Li Y, Hou T, Goddard III W. 2010.** Computational modeling of structure-function of G protein-coupled receptors with applications for drug design. *Current Medicinal Chemistry* **17**(12):1167–1180 DOI [10.2174/092986710790827807](https://doi.org/10.2174/092986710790827807).
- Lim VJY, Du W, Chen YZ, Fan H. 2018.** A benchmarking study on virtual ligand screening against homology models of human GPCRs. *Proteins: Structure, Function, and Bioinformatics* **86**(9):978–989 DOI [10.1002/prot.25533](https://doi.org/10.1002/prot.25533).
- Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG, Lichtarge O. 2004.** Evolutionary trace of G protein-coupled receptors reveals clusters of residues that determine global and class-specific functions. *Journal of Biological Chemistry* **279**(9):8126–8132  
DOI [10.1074/jbc.M312671200](https://doi.org/10.1074/jbc.M312671200).
- Martínez-Muñoz A, Prestegui-Martel B, Méndez-Luna D, Fragoso-Vázquez MJ, García-Sánchez JR, Bello M, Martínez-Archundia M, Chávez-Blanco A, Dueñas-González A, Mendoza-Lujambio I, Trujillo-Ferrara J, Correa-Basurto J. 2018.** Selection of a GPER1 ligand via ligand-based virtual screening coupled to molecular dynamics simulations and its anti-proliferative effects on breast cancer cells. *Anti-Cancer Agents in Medicinal Chemistry* **18**(11):1629–1638 DOI [10.2174/1871520618666180510121431](https://doi.org/10.2174/1871520618666180510121431).
- Méndez-Luna D, Bello M, Correa-Basurto J. 2016.** Understanding the molecular basis of agonist/antagonist mechanism of GPER1/GPR30 through structural and energetic analyses. *The Journal*

- of Steroid Biochemistry and Molecular Biology* **158(12)**:104–116  
DOI 10.1016/j.jsbmb.2016.01.001.
- Méndez-Luna D, Martínez-Archundia M, Maroun RC, Ceballos-Reyes G, Fragoso-Vázquez MJ, González-Juárez DE, Correa-Basurto J. 2015.** Deciphering the GPER/GPR30-agonist and antagonists interactions using molecular modeling studies, molecular dynamics, and docking simulations. *Journal of Biomolecular Structure and Dynamics* **33(10)**:2161–2172  
DOI 10.1080/07391102.2014.994102.
- Merz KM. 2010.** Limits of free energy computation for protein–ligand interactions. *Journal of Chemical Theory and Computation* **6(5)**:1769–1776 DOI 10.1021/ct100102q.
- O’Dowd BF, Nguyen T, Marchese A, Cheng R, Lynch KR, Heng HHQ, Kolakowski Jr. LF, George SR. 1998.** Discovery of three novel G-protein-coupled receptor genes. *Genomics* **47(2)**:310–313 DOI 10.1006/geno.1998.5095.
- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004.** UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**:1605–1612 DOI 10.1002/(ISSN)1096-987X.
- Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch A, Steyaert J, Weis WI, Kobilka BK. 2011.** Structure of a nanobody-stabilized active state of the  $\beta$  2 adrenoceptor. *Nature* **469(7329)**:175–180 DOI 10.1038/nature09648.
- Sanders. 2011.** ss-TEA: entropy based identification of receptor specific ligand binding residues from a multiple sequence alignment of class A GPCRs. *BMC Bioinformatics* **12(1)**:332–343  
DOI 10.1186/1471-2105-12-332.
- Sanders MPA, Verhoeven S, de Graaf C, Roumen L, Vroling B, Nabuurs SB, de Vlieg J, Klomp JPG. 2011.** Snooker: a structure-based pharmacophore generation tool applied to class A GPCRs. *Journal of Chemical Information and Modeling* **51(9)**:2277–2292  
DOI 10.1021/ci200088d.
- Segala E, Guo D, Cheng RKY, Bortolato A, Deflorian F, Doré AS, Errey JC, Heitman LH, IJzerman AP, Marshall FH, Cooke RM. 2016.** Controlling the dissociation of ligands from the adenosine A2A receptor through modulation of salt bridge strength. *Journal of Medicinal Chemistry* **59(13)**:6470–6479 DOI 10.1021/acs.jmedchem.6b00653.
- Shoichet BK, Kobilka BK. 2012.** Structure-based drug screening for G-protein-coupled receptors. *Trends in Pharmacological Sciences* **33(5)**:268–272 DOI 10.1016/j.tips.2012.03.007.
- Smith RD, Damm-Ganamet KL, Dunbar JB, Ahmed A, Chinnaswamy K, Delproposito JE, Kubish GM, Tinberg CE, Khare SD, Dou J, Doyle L, Stuckey JA, Baker D, Carlson HA. 2016.** CSAR benchmark exercise 2013: evaluation of results from a combined computational protein design, docking, and scoring/ranking challenge. *Journal of Chemical Information and Modeling* **56(6)**:1022–1031 DOI 10.1021/acs.jcim.5b00387.
- Souza SA, Kurohara DT, Dabalos CL, Ng HL. 2019.** G protein-coupled estrogen receptor production using an escherichia coli cell-free expression system. *Current Protocols in Protein Science* **97(1)**:e88 DOI 10.1002/cpps.88.
- Tang H, Wang XS, Hsieh J-H, Tropsha A. 2012.** Do crystal structures obviate the need for theoretical models of GPCRs for structure-based virtual screening? *Proteins: structure, Function, and Bioinformatics* **80(6)**:1503–1521 DOI 10.1002/prot.24035.
- Wacker D, Fenalti G, Brown MA, Katritch V, Abagyan R, Cherezov V, Stevens RC. 2010.** conserved binding mode of human  $\beta$ 2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *Journal of the American Chemical Society* **132(33)**:11443–11445 DOI 10.1021/ja105108q.

- Wacker D, Stevens RC, Roth BL. 2017.** How ligands illuminate GPCR molecular pharmacology. *Cell* **170**(3):414–427 DOI [10.1016/j.cell.2017.07.009](https://doi.org/10.1016/j.cell.2017.07.009).
- Wan S, Knapp B, Wright DW, Deane CM, Coveney PV. 2015.** Rapid, precise, and reproducible prediction of peptide-MHC binding affinities from molecular dynamics that correlate well with experiment. *Journal of Chemical Theory and Computation* **11**(7):3346–3356 DOI [10.1021/acs.jctc.5b00179](https://doi.org/10.1021/acs.jctc.5b00179).
- Wass MN, Sternberg MJE. 2009.** Prediction of ligand binding sites using homologous structures and conservation at CASP8. *Proteins: Structure, Function, and Bioinformatics* **77**(S9):147–151 DOI [10.1002/prot.22513](https://doi.org/10.1002/prot.22513).
- Weiss DR, Bortolato A, Tehan B, Mason JS. 2016.** GPCR-bench: a benchmarking set and practitioners' guide for G protein-coupled receptor docking. *Journal of Chemical Information and Modeling* **56**(4):642–651 DOI [10.1021/acs.jcim.5b00660](https://doi.org/10.1021/acs.jcim.5b00660).
- Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015.** The I-TASSER suite: protein structure and function prediction. *Nature Methods* **12**(1):7–8 DOI [10.1038/nmeth.3213](https://doi.org/10.1038/nmeth.3213).
- Zhang J, Yang J, Jang R, Zhang Y. 2015.** GPCR-I-TASSER: a hybrid approach to G protein-coupled receptor structure modeling and the application to the human genome. *Structure* **23**(8):1538–1549 DOI [10.1016/j.str.2015.06.007](https://doi.org/10.1016/j.str.2015.06.007).
- Zou Y, Ewalt J, Ng HL. 2019.** Recent insights from molecular dynamics simulations for G protein-coupled receptor drug discovery. *International Journal of Molecular Sciences* **20**(17):4237 DOI [10.3390/ijms20174237](https://doi.org/10.3390/ijms20174237).