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

Ashley R Vidad [Equal first author, 1] , **Stephen Macaspac** [Equal first author, 1] , **Ho Leung Ng** [Corresp. 2]

1 Department of Chemistry, University of Hawaii at Manoa, Honolulu, Hawaii, United States of America

2 Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, Kansas, United States of America

Corresponding Author: Ho Leung Ng
Email address: hng@ksu.edu

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.

1 **TITLE: Locating ligand binding sites in G-protein coupled receptors using combined**
2 **information from docking and sequence conservation**
3

5 Authors: Ashley R. Vidad*[1], Stephen Macaspac*[1] & Ho Leung Ng[2#]
6
7
8 [1]University of Hawaii at Manoa, Department of Chemistry, Honolulu, HI. USA
9 [2]Kansas State University, Department of Biochemistry & Molecular Biophysics, Manhattan, KS.
10 USA
11
12 * Authors are considered equal contributors to this paper.
13 [#] Corresponding author. hng@ksu.edu

15
16
17
18 Keywords: G protein-coupled receptor (GPCR), binding sites, homology modeling

20 **Abstract**
21
22   GPCRs (G-protein coupled receptors) are the largest family of drug targets and share a

23 conserved structure. Binding sites are unknown for many important GPCR ligands due to the

24 difficulties of GPCR recombinant expression, biochemistry, and crystallography. We describe

25 our approach, ConDockSite, for predicting ligand binding sites in class A GPCRs using

26 combined information from surface conservation and docking, starting from crystal structures or

27 homology models. We demonstrate the effectiveness of ConDockSite on crystallized class A

28 GPCRs such as the beta2 adrenergic and A2A adenosine receptors. We also demonstrate that

29 ConDockSite successfully predicts ligand binding sites from high-quality homology models.

30 Finally, we apply ConDockSite to predict the ligand binding sites on a structurally

31 uncharacterized GPCR, GPER, the G-protein coupled estrogen receptor. Most of the sites

32 predicted by ConDockSite match those found in other independent modeling studies.

33 ConDockSite predicts that four ligands bind to a common location on GPER at a site deep in the

34 receptor cleft. Incorporating sequence conservation information in ConDockSite overcomes

35 errors introduced from physics-based scoring functions and homology modeling.

37
38
39
40
41
42
43
44

Introduction

48   GPCRs (G-protein coupled receptors) are the largest family of drug targets and the

49   targets of >30% of all drugs. Because they are membrane proteins with flexible and dynamic

50   structures, biochemical and crystallography experiments are difficult. Only ~87 GPCRs out of

51   ~800 in the human genome have been crystallized despite their great pharmacological

52   importance. GPCR homology modeling remains challenging due to conformational flexibility

53   and the abundance of flexible loops (Lai et al., 2017). Crystal structures have shown that the

54   large majority of ligands bind in the large central, extracellular cavity of GPCRs, but the specific

55   binding sites in the cavity can vary widely between different ligands even for the same or closely

56   related receptors (Wacker, Stevens & Roth, 2017; Chan et al., 2019).

57   Various computational approaches have been used to predict ligand binding sites in G-

58   protein coupled receptors. Traditional docking methods compute the lowest energy pose of a

59   ligand fit to a receptor surface. Such methods are highly dependent on the form of the energy

60   scoring function and accuracy of the receptor model structure (Katritch et al., 2010; Katritch &

61   Abagyan, 2011; Shoichet & Kobilka, 2012; Weiss et al., 2016; Lim et al., 2018). These methods

62   have been used to identify ligand binding sites and build pharmacophores for GPCRs

63   (Kratochwil et al., 2011; Sanders et al., 2011; Tang et al., 2012), but the lack of diverse GPCR

64   crystal structures presents serious challenges to using docking methods for identification of

65   ligand binding sites. In particular, few crystal structures of non-class A GPCRs have been

66   determined. Moreover, homology models usually cannot be used to identify ligand binding sites

67   or for docking without extensive optimization, such as with advanced molecular dynamics

68   sampling methods (Katritch et al., 2010; Lai et al., 2017; Zou, Ewalt & Ng, 2019). An

69   underappreciated feature that can be used to predict ligand binding sites is surface or sequence

70   conservation. Binding sites for particular ligands are often conserved, and systematic sequence

71   variation can encode ligand specificity (Capra & Singh, 2007; Kalinina, Gelfand & Russell,

72   2009; Wass & Sternberg, 2009). While highly conserved receptors often share similar ligand

73   binding sites, such direct relationships often do not apply between less conserved receptors. Yet,

74   the massive abundance of genomic data for GPCRs can provide strong constraints for possible

75   ligand binding sites even without chemical or structural information (Madabushi et al., 2004;

76   Sanders, 2011; Levit et al., 2012). The binding sites for synthetic, non-physiological ligands can

77   also be identified as they often share some or even most of their binding sites with physiological

78   ligands (Wacker, Stevens & Roth, 2017). However, binding site conservation information alone

79   usually cannot predict ligand binding sites, as often, very large parts of the receptor are highly

80   conserved, including key structural elements.

81   There has been less research on methods that combine information from chemical

82   interactions, geometric surface analysis, and bioinformatics. Hybrid strategies, such as Concavity

83   (Capra et al., 2009), have demonstrated superior performance in predicting ligand binding sites

84   compared to single-mode approaches. Concavity scores binding sites by evolutionary sequence

85   conservation, as quantified by the Jensen-Shannon divergence (Capra & Singh, 2007), and

86   employs geometric criteria of size and shape. Here, we describe a new hybrid strategy we have

87   developed, called ConDockSite, to predict ligand binding sites from combined information from

88   surface conservation and docking calculations. We compare our results with those previously

89   published using purely docking-based and other hybrid methods (Arnatt & Zhang, 2013;

90   Méndez-Luna et al., 2015). ConDockSite is not intended to be used for docking, ie., predicting

91   ligand binding poses, which are highly sensitive to small structural details in crystal structures.

92   We demonstrate the effectiveness of ConDockSite for identifying ligand binding sites for the two

93   best characterized class A GPCRs with known crystal structures, the β2 adrenergic and A2A

94   adenosine receptors. We then demonstrate the effectiveness of ConDockSite with high quality

95   GPCR homology models.

96   Finally, we apply ConDockSite to predict the hypothetical binding sites of four ligands to

97   the less characterized class A G-protein coupled estrogen receptor (GPER, formerly known as

98   GPR30), a membrane-bound estrogen receptor. GPER is proposed to mediate rapid estrogen-

99   associated effects, cAMP regeneration, and nerve growth factor expression (Kvingedal &

100 Smeland, 1997; Carmeci et al., 1997; O'Dowd et al., 1998; Filardo et al., 2002; Kanda &

101 Watanabe, 2003). GPER is known to bind estradiol and the estrogen receptor inhibitors,

102 tamoxifen and fulvestrant, that are used to treat breast cancer (Fig. S1). Recently, GPER-specific

103 ligands G1 and G15 were discovered (Bologa et al., 2006; Dennis et al., 2009). G1 and G15 are

104 structurally similar, differing by only an acetyl group. G1 is an agonist, whereas G15 is an

105 antagonist. No crystal structure of GPER is available, and details of ligand binding are unknown.

106 ConDockSite predictions can be tested experimentally by measuring the effects of mutagenesis

107 of predicted ligand binding sites on ligand binding. Such efforts should be straightforward given

108 our previous publication describing methods for recombinant expression and ligand binding

109 assays for GPER (Souza et al., 2019).

110

111 **Methods**

112 **Protein surface conservation**

113   GPCR protein sequences were acquired from the SwissProt database (Boeckmann et al.,

114 2005). For this study, we chose the protein sequences manually for consistency and

115 reproducibility. For the A2A adenosine receptor, the protein sequences aligned were from *Homo*

116 *sapiens, Canis familiaris, Xenopus tropicalis, Myotis davidii, Loxodonta africana, Gallus gallus,*

117 *Anolis caronlinesis, Oncorhynchus mykiss, Ailuropoda melanoleuca*, and *Alligator*

118 *mississippiensis*. For the β2 adrenergic receptor, the protein sequences aligned were from *Homo*

119 *sapiens, Oncorhynchus mykiss, Myotis brandtii, Callorhinchus milii, Ophiophagus hannah,*

120 *Canis familiaris, Loxodonta africana, Ailuropoda melanoleuca, Ficedula albicollis*, and *Xenopus*

121 *laevis*. GPER protein sequences aligned were from diverse species: *Homo sapiens, Rattus*

122 *norvegicus, Mus musculus, Macaca mulatta, Danio rerio*, and *Micropogonias undulatus*.

123 Sequences were chosen to represent a diverse range of animal species. Sequences for the other

124 receptors studied were chosen automatically by ConSurf from the NCBI "NR" non-redundant

125 database. We obtained qualitatively similar results when using ConSurf to automatically select

126 protein sequences for analysis. Multiple sequence alignment files were submitted to ConSurf

127 (https://consurf.tau.ac.il) (Armon, Graur & Ben-Tal, 2001; Ashkenazy et al., 2010). ConSurf

128 assesses conservation using Bayesian reconstruction of a phylogenetic tree. Each sequence

129 position is scored from 0-9, where 9 indicates that the amino acid was retained in all the

130 organisms (Fig. S2). Values from ConSurf were mapped onto the receptor surface with Chimera

131 (Pettersen et al., 2004).

132

133 **Homology modeling and docking**

134   The crystal structures for the A2A adenosine receptor and the β2 adrenergic receptor

135 were acquired from the RCSB protein data bank. The crystal structures used were of the β2

136 adrenergic receptor bound to the agonist, epinephrine (PDB 4ldo), the agonist, BI-167107 (PDB

137 3p0g), the inverse agonist, carazolol (PDB 2rh1), and the inverse agonist, ICI 118,551 (PDB

138 3ny8). The crystal structures used were of the A2A adenosine receptor bound to the agonist,

139   adenosine (PDB 2ydo), the agonist, CGS21680 (PDB 4ug2), the inverse agonist, ZM241385

140   (PDB 5k2a), and the inverse agonist, compound 12X (PDB 5iub). The crystal structures of the

141   mu opioid, serotonin 5HT2B, dopamine D2 with haloperidol, dopamine D2 with risperidone,

142   ghrelin, histamine H1, and muscarinic M1 receptors were taken from PDB entries 5c1m, 6drz,

143   6luq, 6cm4, 6ko5, 3rze, and 5cxv. Structures were prepared for docking with Chimera by

144   removing extraneous chains, solvent atoms, and bound ligands with the DockPrep protocol.

145   Ligands were docked into receptors with SwissDock (http://www.swissdock.ch) (Grosdidier,

146   Zoete & Michielin, 2011a).

147   The homology models for all the receptors were generated by I-TASSER

148   (https://zhanglab.dcmb.med.umich.edu/I-TASSER) (Yang et al., 2015). I-TASSER generates

149   composite homology models using multiple crystal structures. Only the predicted 7-

150   transmembrane regions were input to I-TASSER for modeling. Models with > 60% sequence

151   identity were excluded as modeling templates so that I-TASSER would not simply retrieve the

152   corresponding crystal structures for modeling. The PDB crystal structures used for modeling the

153   A2A receptor were 6oij, 6e59, and 6me6. The PDB crystal structures used for modeling the β2

154   adrenergic receptor were 6me8, 7bts, 5zbh, 6oij, 6kuw, 6ibl, and 2ks9. The PDB crystal

155   structures used for modeling the mu opioid receptor were 4n6h and 4djh. The PDB crystal

156   structures used for modeling the serotonin 5HT2B receptor were 6me6, 6kuw, 4iaq, 5wiv, 6a93,

157   5zbh, 6kp6, 7lcy, 6a94, and 6vdp. The PDB crystal structures used for modeling the dopamine

158   D2 receptor were 6kux, 6kp6, and 6kuw. The PDB crystal structures used for modeling the

159   ghrelin receptor were 4n6h and 6e59. The PDB crystal structures used for modeling the

160   histamine H1 receptor were 6oij, 6kp6, 6me6, and 6kux. The PDB crystal structures used for

161   modeling the muscarinic M1 receptor were 6me6, 6kp6, and 6me2.

162   The crystal structure of GPER has not yet been determined. We created a homology

163   model using GPCR I-TASSER (Iterative Threading Assembly Refinement), the most accurate

164   homology modeling software customized for GPCRs

165   (https://zhanglab.dcmb.med.umich.edu/GPCR-I-TASSER) (Zhang et al., 2015). GPCR I-

166   TASSER modeled the GPER structure using template fragments automatically selected from the

167   closest related GPCR crystal structures (CCR5: PDB 4mbs, sphingosine 1-phosphate: PDB 3v2y,

168   CXCR4: PDB 3odu, delta opioid: PDB 4n6h). The homology model was validated with ERRAT

169   (Colovos & Yeates, 1993). Coordinates for E2, G1, G15, and tamoxifen were downloaded from

170   the ZINC ligand database (Irwin et al., 2012) and submitted to SwissDock (Grosdidier, Zoete &

171   Michielin, 2011a) for docking. SwissDock is a web interface to the EADock DSS (Grosdidier,

172   Zoete & Michielin, 2011b) engine, which performs blind, global (does not require targeting of a

173   particular surface) docking using the physics-based CHARMM22 force field (Brooks et al.,

174   2009). The "FullFitness Score" calculated by SwissDock using clustering and the FACTS

175   implicit solvent model (Haberthür & Caflisch, 2008) was used as the "Energy Score" for our

176   calculations. SwissDock was chosen both for its high effectiveness as well as ease of use by

177   students. For consistency, we performed all docking studies in this paper with SwissDock

178   although we obtained qualitatively similar docking results with AutoDock Vina, the most

179   popular docking software, in our preliminary studies.

180

181   **Combined analysis**

182   SwissDock poses were analyzed for ligand binding sites near highly conserved surfaces.

183   Ligand binding surfaces included residues with atoms within 3.5 Å from the docked ligand. The

184   average conservation score of the amino acids that were highlighted served as the "Conservation

185 Score" of that specific orientation (Scheme 1). The combined ConDockSite score is defined as

186 the product of the Conservation and Energy Scores. As the Energy Score is a modified free

187 energy function, a highly negative ConDockSite score is associated with a more probable ligand

188 binding site. Binding sites predicted by ConDockSite results were compared with those predicted

189 by CASTp (http://sts.bioe.uic.edu/castp/index.html) (Dundas et al., 2006), SiteHound

190 (Hernandez, Ghersi & Sanchez, 2009), and Concavity (Capra et al., 2009). For CASTp,

191 SiteHound, and Concavity, ligand binding pockets were defined as residues within 4 Å of the

192 selected probe/cluster.

193
194 **Scheme 1. Calculation of combined ConDockSite scores for ligand binding sites.** The
195 Conservation Score is calculated over the n residues in a binding site, indexed by k.
196

197

198 **Binding site prediction benchmarks**

199 The distances between the centers of mass of the predicted and experimental ligand

200 coordinates served as a benchmark of comparison for the sites predicted by the ConDockSite

201 scoring function. This is a more appropriate measure than the ligands RMSD used for pose

202 comparison as ConDockSite is intended to predict binding sites rather than precise binding

203 poses.

204

205 **Results**

206 We developed ConDockSite to predict ligand binding pockets using information from

207 surface conservation and docking calculations. ConDockSite uses a simple scoring function that

208 is the product of surface conservation scores from ConSurf (Armon, Graur & Ben-Tal, 2001) and

209 docking scores from SwissDock (Grosdidier, Zoete & Michielin, 2011a). A highly negative

210 ConDockSite score is associated with a more probable ligand binding site.

211 The A2A adenosine and β2 adrenergic receptors are by far the most heavily studied

212 GPCRs by crystallography. First, we show that the Consurf surface conservation information

213 alone is inadequate for identifying binding sites. Conservation varies greatly associated

214 throughout the two receptors and is not directly associated with the ligand binding sites (Fig. S3).

215 Many of the conserved regions are instead associated with the internal packing of the seven

216 transmembrane helices. Next, we show that geometric binding pocket analyses, such as CASTp

217 (Dundas et al., 2006), often work poorly for GPCRs. For the two receptors, CASTp merely

218 predicts the entire GPCR central cavity as the ligand binding site (Fig. S4). It fails to localize the

219 ligands to specific parts of the central cavity.

220 Finally, we used the crystal structures of the two receptors as positive control

221 experiments to validate the effectiveness of ConDockSite for predicting ligand binding sites. For

222 both receptors, we performed cross-docking of an agonist and inverse agonist against a crystal

223 structure of the receptor bound to a different agonist or inverse agonist: ligands were cross-

224 docked rather than self-docked into its own crystal structure. Self-docking is often trivial for

225 modern docking methods and thus, was not studied. Docking was performed with SwissDock,

226 which has demonstrated high accuracy in docking ligands into receptors without prior knowledge

227 of the binding site (also known as global or blind docking), and also includes a user-friendly web

228 interface suitable for students (Grosdidier, Zoete & Michielin, 2011a). SwissDock docking

229 results were then ranked by the ConDockSite scoring function (Table S1). Residues within 3.5 Å

230 of the highest scoring predicted ligand sites were compared with the binding surfaces associated

231 with the ligand poses in the crystal structures.

232   As a convenient metric for the distances between predicted and experimental ligand

233   binding sites, we use the distances between the ConDockSite-scored ligand poses and those

234   observed in the crystal structures. For the A2A adenosine receptor, the agonist, adenosine, was

235   cross-docked into the crystal structure of the receptor with the agonist CGS21680 (PDB 4ug2).

236   The highest ConDockSite-ranked pose for adenosine within the A2A adenosine receptor was

237   within 0.4 Å of the ligand position (distance between centers of mass) in the crystal structure

238   (PDB 2ydo). (Fig. 1A). The ConDockSite-predicted binding site had a ConSurf conservation

239   score of 0.86 and is essentially the same as the experimental binding site. The inverse agonist,

240   ZM241385, was cross-docked into the crystal structure of the receptor with the inverse agonist,

241   compound 12X (PDB 5iub) (Segala et al., 2016). The highest ranked site for ZM241385 within

242   the A2A adenosine receptor was within 1.0 Å of the ligand's position in the crystal structure

243   (PDB 5k2a). In this top pose, ZM241385 is found within the same binding site as that observed

244   in the crystal structure (Fig. 1B), with a ConSurf conservation score of 0.86. For both adenosine

245   and ZM241385, the ConDockSite-predicted site corresponded to the top site predicted by

246   SwissDock. In this case, docking alone was adequate to identify the binding site. The correct

247   binding site also had the highest surface conservation. The A2A adenosine receptor structures are

248   easy tests for ligand prediction, passed by both SwissDock and ConDockSite.

249   For the β2 adrenergic receptor, the agonist, epinephrine, was cross-docked into the

250   crystal structure of the receptor with the agonist, BI-167107 (PDB 3p0g) (Rasmussen et al.,

251   2011). The highest ranked pose for epinephrine within the β2 adrenergic receptor was within 0.4

252   Å of the ligand position within the crystal structure (PDB 4ldo). This binding site for epinephrine

253   was again essentially the same as the observed binding pocket (Fig. 1C) and had a ConSurf

254   conservation score of 0.85. This site was also scored the highest by SwissDock. For the β2

255 adrenergic receptor, the inverse agonist, carazolol, was cross-docked into the crystal structure of

256 the receptor with the inverse agonist ICI 118,551 (PDB 3ny8) (Wacker et al., 2010). The highest

257 ranked pose for carazolol within the β2 adrenergic receptor was within 1.0 Å of the ligand's

258 position within the crystal structure (PDB 2rh1). This binding site for carazolol was essentially

259 the same as that in the crystal structure (Fig. 1D). This binding pocket has a ConSurf

260 conservation score of 0.78. The correct carazalol binding site was scored the third highest by

261 SwissDock. The use of surface conservation information allowed selection of the proper binding

262 site. The extremely accurate placement of both agonists and inverse agonists for the A2A

263 adenosine and the β2 adrenergic receptors demonstrates ConDockSite's effectiveness when

264 crystal structures are available.

265

266 **Figure 1. Predicted and experimental ligand binding sites in A2A adenosine and β2 adrenergic receptors from cross-docked crystal structures.** Superposition of crystal structure with ligand bound (red) with ConDockSite predicted pose (blue). A) A2A receptor with the

267 agonist, adenosine. B) A2A receptor with the inverse agonist, ZM241385. C) β2 adrenergic receptor with the agonist, epinephrine. D) β2 adrenergic receptor with the inverse agonist,

268 carazolol.

269 Unfortunately, crystal structures are not available for most GPCRs. The most valuable

270 use of ConDockSite is to predict drug binding sites in homology models. By using surface

271 conservation information, ConDockSite is less sensitive to homology model inaccuracies than

272 other ligand binding site prediction methods that are based purely on geometric methods. To

273 demonstrate the ability of ConDockSite to work with homology models, we created models of

274 eight GPCRs, the β2 adrenergic, A2A adenosine, 5HT2B serotonin, mu opioid receptors, D2

275 dopamine, ghrelin, H1 histamine, and M1 muscarinic receptors, while excluding their known X-

276 ray structures from the templates used for modeling. We used I-TASSER (Yang et al., 2015) for

277 homology modeling which does not use GPCR-specific structural constraints but allows for

278   custom selection of templates. I-TASSER created fairly accurate models of all eight receptors,

279   with RMSDs across Cα atoms between the models and crystal structures ranging from a best of

280   1.5 Å for the histamine H1 receptor (PDB 3rze) to a respectable 2.1 Å for the muscarinic M1

281   (PDB 5cxv) receptors. We used ConDockSite to predict the binding sites of the β2 adrenergic

282   receptor with carazalol, A2A adenosine receptor with ZM241385, 5HT2B serotonin receptor

283   with methysergide, mu opioid receptor with BU72, D2 dopamine receptor with risperidone and

284   haloperidol, ghrelin receptor with the antagonist Compound 21, histamine H1 receptor with

285   doxepin, and muscarinic M1 receptor with tiotropium. ConDockSite performed best with the β2

286   adrenergic receptor homology model, with only 2.3 Å between the centers of mass of the

287   predicted and crystal structure ligand poses (Fig. 2, Table S2), supporting the prediction of very

288   similar binding pockets. The binding pocket was predicted correctly although the ligand pose

289   was inaccurate. There was a weak correspondence between ConDockSite performance with the

290   accuracy of the homology models. The H1 histamine receptor model was the best homology

291   model, with RMSD of 1.5 Å between the model and the crystal structure Cα atoms. However,

292   ConDockSite was unable to correctly predict the binding site for the antagonist doxepin, which

293   was predicted > 10 Å from the crystal structure site. The poor performance was due to the

294   inaccurate modeling of a loop containing residues 162-165 over the top of the ligand binding site

295   as well as multiple errors in the conformation of side chains in the transmembrane helixes. The

296   D2 dopamine receptor model was the second-best homology model, with RMSD of 1.6 - 1.8 Å

297   between the model and the crystal structure Cα atoms (PDB 6luq bound to haloperidol and 6cm4

298   bound to risperidone). ConDockSite predicted the binding site of risperidone well, only 2.4 Å

299   from the crystal structure site, but less well for haloperidol, which was predicted 4.3 Å from the

300   crystal structure site. With this medium level of accuracy, the ligand poses are generally

301   incorrect, but most of the binding residues in the predicted pockets are the same as those in the

302   crystal structures, supporting successful ConDockSite predictions. The ghrelin receptor

303   homology model was also predicted well with RMSD of 1.7 Å between the model and the crystal

304   structure Cα atoms (PDB 6ko5). The ghrelin antagonist, Compound 21, was predicted 3.1 Å

305   away from the crystal structure site. The predicted ligand binding sites for the A2A adenosine

306   and mu opioid ligands (PDB 5c1m) are 3-4 Å away from the crystal structures. ConDockSite

307   performs less well with the serotonin 5HT2B receptor (PDB 6drz) where the distance between

308   the predicted and actual ligand binding sites was 7.0 Å. In the serotonin 5HT2B receptor

309   structure, the ligand, methysergide, binds very deep in the receptor. Serious errors in homology

310   modeling of the 5HT2B receptor side chains make it difficult or impossible to dock the ligand

311   into the deep, restricted binding site. ConDockSite also fails with the muscarinic M1 receptor

312   (PDB 5cxv) where the distance between the predicted and actual triotropium binding sites was >

313   10 Å. Overall, our results with ConDockSite are consistent with benchmark modeling results that

314   show that GPCR homology models of modest accuracy from templates with low sequence

315   identity are still sometimes useful for docking and virtual screening (Lim et al., 2018; Costanzi et

316   al., 2019). In some of the failed cases (histamine H1, muscarinic M1, and serotonin 5HT2B),

317   while the homology models were sometime accurate at the level of backbone atoms in the 7tm

318   region, the loops were modeled poorly and disrupted the modeled ligand binding pocket. In other

319   cases, the homology model backbones were modeled well but differences in the side chain

320   conformations disrupted the integrity of the ligand binding sites. In these cases, the homology

321   models are not accurate enough for docking or ConDockSite. In retrospect, ConDockSite

322   sometimes performed better with alternative homology models generated by I-TASSER with

323   lower RMSD measures but with alternative loop placements.

324   **Figure 2. Predicted and experimental ligand binding sites for homology models of**
325   **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)
326   Carazolol with β2 adrenergic receptor. B) Risperidone with dopamine D2. C) Compound
      21 with ghrelin receptor. D) BU72 with mu opioid. E) Haloperidol with dopamine D2. F)
327   ZM241385 with A2A adenosine. G) Methysergide with serotonin 5HT2B. H) Doxepin
      with histamine H1. I) Tiotropium with muscarinic M1.

328

329   After demonstrating the applicability of ConDockSite for homology models, we applied

330   ConDockSite to predict the binding sites in a GPCR, GPER (G-protein coupled estrogen

331   receptor), which has not yet been crystallized. GPER is of great interest to us and other

332   researchers due to its unusual physiological and pharmacological roles in estrogen-related

333   biology. As there is no experimental data on the ligand binding sites in GPER, we cannot

334   validate our predictions. They should be considered speculative at this point, but we hope they

335   can guide future experiments. To predict the potential ligand binding sites in GPER, we first

336   created a homology model using GPCR-I-TASSER (Zhang et al., 2015). GPCR-I-TASSER has

337   been shown to be among the most accurate GPCR homology modeling software package. We

338   used the generic I-TASSER in our validation studies because of its fine-grained options for

339   template selection that are lacking in GPCR-I-TASSER. Because GPCR-I-TASSER uses GPCR-

340   specific structural constraints, it is expected to outperform the generic I-TASSER (Zhang et al.,

341   2015). GPCR-I-TASSER identified the closest matching crystal structure to GPER to be the

342   CCR5 chemokine receptor (PDB 4mbs) with 23% sequence identity. GPCR-I-TASSER used this

343   crystal structure along with 9 other GPCR structures as templates for homology modeling. The

344   GPER homology model differs from chain A of the crystal structure of CCR5 chemokine

345   receptor with RMSD of 0.96 Å across Cα atoms (Fig. S5) and has a Ramachandran plot with

346   95.6% of residues, excluding glycine, in preferred regions (Fig. S6). The primary differences are

347   in the extracellular loop between helices 4 and 5 (ECL2) and the intracellular loops between

348   helices 5 and 6 (ICL3), and after helix 7. These two intracellular loops are predicted by ERRAT

349   (Colovos & Yeates, 1993) to be the least reliable based on the likelihood of atom pair type

350   interactions from high-resolution crystal structures (Fig. S7).

351   Using the SwissDock server (Grosdidier, Zoete & Michielin, 2011a), we docked

352   structures of the four ligands E2, G1, G15, and tamoxifen (Fig. S1) to the homology model of

353   GPER. The docked sites from SwissDock, including those that were scored the highest, were

354   located throughout the receptor surface and thus were considered mostly nonviable (Fig. 3). The

355   shortcomings of a purely physics-based scoring function such as that used by SwissDock in

356   predicting ligand binding are not surprising given the lack of an experimental crystal structure

357   and well-known limitations of current homology modeling and docking methodology (Li, Hou &

358   Goddard III, 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 GPER.

359

360   We then ranked all ligand binding sites generated by SwissDock using the combined

361   ConDockSite score. For all four ligands, the ConDockSite score identified one or two ligand

362   binding sites that clearly outscored (more negative) other candidates (Table S1). ConDockSite

363   identified the same approximate binding site for all four ligands, although this was not an explicit

364   criterion in the calculations (Fig. S8). The average ConSurf conservation score across the four

365   ligand binding sites is 0.82 (1.0 represents complete conservation), indicating that the site is

366   highly but not completely conserved. The binding site is located deep in the receptor cleft,

367   although depth was not a criterion in the prediction calculation. Given the lack of additional

368 experimental evidence for the location of the ligand binding site, the proposed ConDockSite sites

369 are physically reasonable.

370   We found two promising potential binding sites for E2 in GPER. The two sites are 4.4 Å

371 apart, located deep in the receptor cleft (Fig. 4). E2 is oriented perpendicular to the lipid

372 membrane and rotated about 180° between the two poses. The conservation scores for these two

373 poses are 0.84 and 0.80. The energy scores of the two poses are similar. The amino acids

374 contacting E2 in pose 1 are conserved in GPERs from six species, and only one residue

375 contacting pose 2, H282, varies across species. In the top ranked pose, there is a hydrogen bond

376 between the inward pointing D-ring hydroxyl group of E2 and the carboxyl terminal on E115.

377 Hydrophobic interactions are present between E2 and non-polar residues L119, Y123, P303, and

378 F314. This binding site approximately corresponds to that predicted by Lappano et al using

379 docking (Lappano et al., 2010). In the second ranked pose, the inward pointing A-ring hydroxyl

380 group of E2 makes a hydrogen bond with N310. This pose is in a less hydrophobic environment,

381 contacting primarily H282 and P303.

382

383 **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
384 interactions for E2 pose 2.

385

386   ConDockSite predicts that G1 and G15 bind in adjacent but distinct binding sites

387 separated by 2.3 Å. The top predicted binding site for G1 is found within the pocket bound by

388 Y55, L119, F206, Q215, I279, P303, H307, and N310 (Fig. 5). This orientation had the highest

389 conservation score of all predicted binding sites at 0.85. In this pose, N310 makes a long

390 hydrogen bond (3.6 Å N-O distance) with the acetyl oxygen of G1. The predicted binding site for

391   G15 is found within the pocket is surrounded by L119, Y123, M133, S134, L137, Q138, P192,

392   V196, F206, C207, F208, A209, V214, E218, H307, and N310. This pose had a conservation

393   score of 0.8. Hydrogen bonding is not observed between GPER and G15. Hydrophobic

394   interactions are observed with L119, Y123, F206, and V214. The ConDockSite G1 result

395   correspond to the binding sites predicted by recent studies using docking and molecular

396   dynamics simulations and validated by design and activity testing of new G1 derivatives

397   (Méndez-Luna, Bello & Correa-Basurto, 2016; Martínez-Muñoz et al., 2018).

398

399   **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.
400   C) Receptor-ligand interactions for G15.

401

402   ConDockSite predicted two equally high scoring, overlapping poses for tamoxifen, near

       Figure 6. Predicted tamoxifen binding sites in GPER. A) The highest scoring
       docking poses for tamoxifen, pose 1 (maroon) and pose 2 (cyan).

403   E115, L119, Y123, L137, Q138, M141, Y142, Q215, E218, W272, E275, I279, P303, G306,

404   H307, and N310 (Fig. 6). The conservation score of this orientation is 0.81. Hydrophobic

405   interactions are observed between tamoxifen and non-polar residues L119, Y123, Y142, P303,

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

407

408   We compared the GPER ligand binding sites predicted by ConDockSite to those

409   predicted by three other software packages representing different approaches: CASTp (Dundas et

410   al., 2006), which analyzes surface geometry, SiteHound (Hernandez, Ghersi & Sanchez, 2009),

411   which maps surfaces with a chemical probe, and Concavity (Capra et al., 2009), which analyzes

412   surface geometry and conservation (Fig. 7). All three methods could identify a ligand binding

413   site very roughly matching that from ConDockSite. In comparison with the ligand binding sites

414   predicted by traditional methods based on surface geometry and conservation (Fig. 7), the sites

415   predicted by ConDockSite are more detailed in shape due to the information from chemical

416   interactions from ligand docking. The pocket predicted by ConDockSite is deeper than the other

417   pockets, which while intuitively attractive, is not necessarily correct. SiteHound performed

418   particularly poorly, with the top scoring site located on the GPER intracellular face. The site

419   identified by SiteHound closest to the ConDockSite site was scored third and is a shallow

420   binding pocket near H52-G58, E275-H282, and R299-H307 (Fig. 7C). In contrast, the Concavity

421   site was smaller and shallower than the ConDockSite site (Fig. 7D). Surprisingly, the site

422   predicted by the simpler CASTp method best matched the ConDockSite site but is also smaller

423   and shallower (Fig. 7B). For proteins such as GPCRs with large, concave binding pockets,

424   geometry-based prediction methods such as Concavity and CASTp can easily identify the

425   general, approximate location of the ligand binding site. However, such methods may have more

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

427   ConDockSite more closely matched the results of the geometry-based methods given that

428   ConDockSite does not take surface geometry into account. As described previously, the G1 and

429   G15 binding sites predicted by ConDockSite more closely match those made using docking

430   against very computationally expensive molecular dynamics simulations (Méndez-Luna, Bello &

431   Correa-Basurto, 2016).

432
433 **Figure 7. Predicted E2 binding sites by ConDockSite, CASTp, SiteHound,**
434 **Concavity.** Ligand binding sites are colored, predicted by A) ConDockSite, B)
435 CASTp, C) SiteHound, D) Concavity.
436
437
438
439 **Discussion**

440 The ConDockSite scoring method, incorporating information from both surface

441 conservation and docking binding energy, demonstrated high accuracy in predicting ligand

442 binding sites from the crystal structures of two class A GPCRs, the A2A adenosine and β2

443 adrenergic receptors. ConDockSite also successfully predicted the ligand binding sites for many

444 high-quality homology models but failed for models for which loops were modeled poorly and

445 disrupted the ligand binding site. Better homology models with improved loop placement should

446 allow better ConDockSite performance. ConDockSite was also used to predict viable ligand

447 binding sites for four different GPER ligands. In contrast to more typical geometry-based ligand

448 binding site prediction methods, ConDockSite scoring takes advantage of chemistry-specific

449 information about the ligand-receptor interface. The poor performance of SiteHound in

450 predicting ligand binding sites on GPER suggests that a method based only on chemical

451 interactions or docking is highly susceptible to error, most likely due to the inadequate accuracy

452 of homology models. Surface conservation data not only provides orthogonal knowledge but also

453 dampens the influence from the shortcomings of current computational methods in homology

454 modeling, docking, and predicting binding affinity. How best to mathematically combine these

455 multiple data sources has been debated (Capra & Singh, 2007; Capra et al., 2009), but we

456 demonstrate here that a simple product scoring function is already effective. The four GPER

457 ligands studied here differ greatly in chemical structure, but the ConDockSite scoring method

458 predicted that all four bind to the same approximate region, deep in the extracellular cleft of the

459   receptor. Undoubtedly, further refinement of a hybrid scoring function will lead to improved

460   predictions.

461   Earlier GPER modeling studies using molecular dynamics simulations and docking

462   identified different potential binding sites for E2, G1, and G15 near F206 and F208; the

463   interaction with this region was described as driven primarily by π-π stacking interactions

464   (Arnatt & Zhang, 2013; Méndez-Luna et al., 2015). Figure S9 compares the ConDockSite

465   binding site against that predicted in the molecular dynamics simulation and docking study. The

466   ConDockSite binding site is located deeper in the extracellular cleft; the other proposed site

467   involved more surface-exposed loops. It was proposed that Q53, Q54, G58, C205, and H282 all

468   interact with G1 and G15; however, none of these residues are conserved across the six species

469   we analyzed. More recent studies using better homology models and computationally expensive

470   long time-scale molecular dynamics simulations predict E2, G1, and G15 binding sites that

471   approximately match those predicted by ConDockSite (Lappano et al., 2010; Méndez-Luna,

472   Bello & Correa-Basurto, 2016). The ConDockSite binding site predictions can be tested

473   experimentally by performing site-directed mutagenesis and ligand binding assays.

474   In summary, the simple ConDockSite hybrid scoring model predicts physically plausible

475   ligand binding sites by combining information from ligand docking and surface conservation.

476   Using multiple orthogonal sources of information partially avoids errors introduced by modeling

477   (Capra et al., 2009). Given a homology model of modest quality, ConDockSite can sometimes

478   accurately predict ligand binding sites. Using this hybrid method, we identified a site in the

479   extracellular cavity of GPER that has the potential to bind four known GPER ligands. Further

480   optimization of hybrid scoring functions and homology modeling methods should yield

481   significantly improved predictions. Extension of this approach may allow analysis of non-class A

482   GPCRs.

483

484   **Acknowledgments**

485   This work was funded by the Victoria S. and Bradley L. Geist Foundation (H.L.N.), NSF

486   CAREER Award 1833181 (H.L.N.), and the Undergraduate Research Opportunities Program at

487   the University of Hawaii at Manoa (A.R.V., S.M.).

488

489   **Author contributions**

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

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

492

493   **Competing interests**
494
495   The authors have no competing financial or non-financial interests.
496

497   **References**

498   Armon A, Graur D, Ben-Tal N. 2001. ConSurf: an algorithmic tool for the identification of

499   functional regions in proteins by surface mapping of phylogenetic information1. *Journal*

500   *of Molecular Biology* 307:447–463. DOI: 10.1006/jmbi.2000.4474.

501   Arnatt CK, Zhang Y. 2013. G Protein-Coupled Estrogen Receptor (GPER) Agonist Dual

502   Binding Mode Analyses Toward Understanding of Its Activation Mechanism: A

503   Comparative Homology Modeling Approach. *Molecular Informatics* 32:647–658. DOI:

504   10.1002/minf.201200136.

505   Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. 2010. ConSurf 2010: calculating

506   evolutionary conservation in sequence and structure of proteins and nucleic acids.

507   *Nucleic Acids Research* 38:W529-533. DOI: 10.1093/nar/gkq399.

508   Boeckmann B, Blatter M-C, Famiglietti L, Hinz U, Lane L, Roechert B, Bairoch A. 2005.

509   Protein variety and functional diversity: Swiss-Prot annotation in its biological context.

510   *Comptes Rendus Biologies* 328:882–899. DOI: 10.1016/j.crvi.2005.06.001.

511   Bologa CG, Revankar CM, Young SM, Edwards BS, Arterburn JB, Kiselyov AS, Parker MA,

512   Tkachenko SE, Savchuck NP, Sklar LA, Oprea TI, Prossnitz ER. 2006. Virtual and

513   biomolecular screening converge on a selective agonist for GPR30. *Nature Chemical*

514   *Biology* 2:207–212. DOI: 10.1038/nchembio775.

515   Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G,

516   Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J,

517   Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW,

518   Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W,

519   York DM, Karplus M. 2009. CHARMM: the biomolecular simulation program. *Journal*

520   *of Computational Chemistry* 30:1545–1614. DOI: 10.1002/jcc.21287.

521   Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. 2009. Predicting Protein

522   Ligand Binding Sites by Combining Evolutionary Sequence Conservation and 3D

523   Structure. *PLoS Comput Biol* 5:e1000585. DOI: 10.1371/journal.pcbi.1000585.

524   Capra JA, Singh M. 2007. Predicting functionally important residues from sequence

525   conservation. *Bioinformatics* 23:1875–1882. DOI: 10.1093/bioinformatics/btm270.

526   Carmeci C, Thompson DA, Ring HZ, Francke U, Weigel RJ. 1997. Identification of a Gene

527   (GPR30) with Homology to the G-Protein-Coupled Receptor Superfamily Associated

528   with Estrogen Receptor Expression in Breast Cancer. *Genomics* 45:607–617. DOI:

529   10.1006/geno.1997.4972.

530   Chan HCS, Li Y, Dahoun T, Vogel H, Yuan S. 2019. New Binding Sites, New Opportunities for

531   GPCR Drug Discovery. *Trends in Biochemical Sciences* 44:312–330. DOI:

532   10.1016/j.tibs.2018.11.011.

533   Colovos C, Yeates TO. 1993. Verification of protein structures: patterns of nonbonded atomic

534   interactions. *Protein Science: A Publication of the Protein Society* 2:1511–1519. DOI:

535   10.1002/pro.5560020916.

536   Costanzi S, Cohen A, Danfora A, Dolatmoradi M. 2019. Influence of the Structural Accuracy of

537   Homology Models on Their Applicability to Docking-Based Virtual Screening: The β_2

538   Adrenergic Receptor as a Case Study. *Journal of Chemical Information and Modeling*

539   59:3177–3190. DOI: 10.1021/acs.jcim.9b00380.

540   Dennis MK, Burai R, Ramesh C, Petrie WK, Alcon SN, Nayak TK, Bologa CG, Leitao A,

541   Brailoiu E, Deliu E, Dun NJ, Sklar LA, Hathaway HJ, Arterburn JB, Oprea TI, Prossnitz

542   ER. 2009. In vivo effects of a GPR30 antagonist. *Nature Chemical Biology* 5:421–427.

543   DOI: 10.1038/nchembio.168.

544   Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. 2006. CASTp: computed atlas

545   of surface topography of proteins with structural and topographical mapping of

546   functionally annotated residues. *Nucleic Acids Research* 34:W116-118. DOI:

547   10.1093/nar/gkl282.

548   Filardo EJ, Quinn JA, Frackelton AR, Bland KI. 2002. Estrogen Action Via the G Protein-

549   Coupled Receptor, GPR30: Stimulation of Adenylyl Cyclase and cAMP-Mediated

550   Attenuation of the Epidermal Growth Factor Receptor-to-MAPK Signaling Axis.

551   *Molecular Endocrinology* 16:70–84. DOI: 10.1210/mend.16.1.0758.

552   Grosdidier A, Zoete V, Michielin O. 2011a. SwissDock, a protein-small molecule docking web

553   service based on EADock DSS. *Nucleic Acids Research* 39:W270–W277. DOI:

554   10.1093/nar/gkr366.

555   Grosdidier A, Zoete V, Michielin O. 2011b. Fast docking using the CHARMM force field with

556   EADock DSS. *Journal of Computational Chemistry* 32:2149–2159. DOI:

557   10.1002/jcc.21797.

558   Haberthür U, Caflisch A. 2008. FACTS: Fast analytical continuum treatment of solvation.

559   *Journal of Computational Chemistry* 29:701–715. DOI: 10.1002/jcc.20832.

560   Hernandez M, Ghersi D, Sanchez R. 2009. SITEHOUND-web: a server for ligand binding site

561   identification in protein structures. *Nucleic Acids Research* 37:W413–W416. DOI:

562   10.1093/nar/gkp281.

563   Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. 2012. ZINC: A Free Tool to

564   Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*

565   52:1757–1768. DOI: 10.1021/ci3001277.

566 Kalinina OV, Gelfand MS, Russell RB. 2009. Combining specificity determining and conserved

567   residues improves functional site prediction. *BMC Bioinformatics* 10:174. DOI:

568   10.1186/1471-2105-10-174.

569 Kanda N, Watanabe S. 2003. 17β-Estradiol Enhances the Production of Nerve Growth Factor in

570   THP-1-Derived Macrophages or Peripheral Blood Monocyte-Derived Macrophages.

571   *Journal of Investigative Dermatology* 121:771–780. DOI: 10.1046/j.1523-

572   1747.2003.12487.x.

573 Katritch V, Abagyan R. 2011. GPCR agonist binding revealed by modeling and crystallography.

574   *Trends in Pharmacological Sciences* 32:637–643. DOI: 10.1016/j.tips.2011.08.001.

575 Katritch V, Rueda M, Lam PC-H, Yeager M, Abagyan R. 2010. GPCR 3D homology models for

576   ligand screening: lessons learned from blind predictions of adenosine A2a receptor

577   complex. *Proteins* 78:197–211. DOI: 10.1002/prot.22507.

578 Kratochwil NA, Gatti-McArthur S, Hoener MC, Lindemann L, Christ AD, Green LG, Guba W,

579   Martin RE, Malherbe P, Porter RHP, Slack JP, Winnig M, Dehmlow H, Grether U,

580   Hertel C, Narquizian R, Panousis CG, Kolczewski S, Steward L. 2011. G protein-coupled

581   receptor transmembrane binding pockets and their applications in GPCR research and

582   drug discovery: a survey. *Current Topics in Medicinal Chemistry* 11:1902–1924.

583 Kvingedal AM, Smeland EB. 1997. A novel putative G-protein-coupled receptor expressed in

584   lung, heart and lymphoid tissue. *FEBS Letters* 407:59–62. DOI: 10.1016/S0014-

585   5793(97)00278-0.

586 Lai JK, Ambia J, Wang Y, Barth P. 2017. Enhancing Structure Prediction and Design of Soluble

587   and Membrane Proteins with Explicit Solvent-Protein Interactions. *Structure* 25:1758-

588   1770.e8. DOI: 10.1016/j.str.2017.09.002.

589   Lappano R, Rosano C, De Marco P, De Francesco EM, Pezzi V, Maggiolini M. 2010. Estriol

590   acts as a GPR30 antagonist in estrogen receptor-negative breast cancer cells. *Molecular*

591   *and Cellular Endocrinology* 320:162–170. DOI: 10.1016/j.mce.2010.02.006.

592   Levit A, Barak D, Behrens M, Meyerhof W, Niv MY. 2012. Homology model-assisted

593   elucidation of binding sites in GPCRs. *Methods in Molecular Biology (Clifton, N.J.)*

594   914:179–205. DOI: 10.1007/978-1-62703-023-6_11.

595   Li Y, Hou T, Goddard III W. 2010. Computational Modeling of Structure-Function of G Protein-

596   Coupled Receptors with Applications for Drug Design. *Current Medicinal Chemistry*

597   17:1167–1180. DOI: 10.2174/092986710790827807.

598   Lim VJY, Du W, Chen YZ, Fan H. 2018. A benchmarking study on virtual ligand screening

599   against homology models of human GPCRs. *Proteins: Structure, Function, and*

600   *Bioinformatics* 86:978–989. DOI: 10.1002/prot.25533.

601   Madabushi S, Gross AK, Philippi A, Meng EC, Wensel TG, Lichtarge O. 2004. Evolutionary

602   Trace of G Protein-coupled Receptors Reveals Clusters of Residues That Determine

603   Global and Class-specific Functions. *Journal of Biological Chemistry* 279:8126–8132.

604   DOI: 10.1074/jbc.M312671200.

605   Martínez-Muñoz A, Prestegui-Martel B, Méndez-Luna D, Fragoso-Vázquez MJ, García-Sánchez

606   JR, Bello M, Martínez-Archundia M, Chávez-Blanco A, Dueñas-González A, Mendoza-

607   Lujambio I, Trujillo-Ferrara J, Correa-Basurto J. 2018. Selection of a GPER1 Ligand via

608   Ligand-based Virtual Screening Coupled to Molecular Dynamics Simulations and Its

609   Anti-proliferative Effects on Breast Cancer Cells. *Anti-Cancer Agents in Medicinal*

610   *Chemistry* 18:1629–1638. DOI: 10.2174/1871520618666180510121431.

611   Méndez-Luna D, Bello M, Correa-Basurto J. 2016. Understanding the molecular basis of

612   agonist/antagonist mechanism of GPER1/GPR30 through structural and energetic

613   analyses. *The Journal of Steroid Biochemistry and Molecular Biology* 158:104–116.

614   DOI: 10.1016/j.jsbmb.2016.01.001.

615 Méndez-Luna D, Martínez-Archundia M, Maroun RC, Ceballos-Reyes G, Fragoso-Vázquez MJ,

616   González-Juárez DE, Correa-Basurto J. 2015. Deciphering the GPER/GPR30-agonist and

617   antagonists interactions using molecular modeling studies, molecular dynamics, and

618   docking simulations. *Journal of Biomolecular Structure and Dynamics* 33:2161–2172.

619   DOI: 10.1080/07391102.2014.994102.

620 Merz KM. 2010. Limits of Free Energy Computation for Protein−Ligand Interactions. *Journal of*

621   *Chemical Theory and Computation* 6:1769–1776. DOI: 10.1021/ct100102q.

622 O'Dowd BF, Nguyen T, Marchese A, Cheng R, Lynch KR, Heng HHQ, Kolakowski Jr. LF,

623   George SR. 1998. Discovery of Three Novel G-Protein-Coupled Receptor Genes.

624   *Genomics* 47:310–313. DOI: 10.1006/geno.1998.5095.

625 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004.

626   UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of*

627   *Computational Chemistry* 25:1605–1612. DOI: 10.1002/jcc.20084.

628 Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, Chae PS, DeVree BT, Rosenbaum

629   DM, Thian FS, Kobilka TS, Schnapp A, Konetzki I, Sunahara RK, Gellman SH, Pautsch

630   A, Steyaert J, Weis WI, Kobilka BK. 2011. Structure of a nanobody-stabilized active

631   state of the β 2 adrenoceptor. *Nature* 469:175–180. DOI: 10.1038/nature09648.

632 Sanders. 2011. ss-TEA: Entropy based identification of receptor specific ligand binding residues

633   from a multiple sequence alignment of class A GPCRs. *BMC Bioinformatics* 12:332–343.

634   DOI: 10.1186/1471-2105-12-332.

635 Sanders MPA, Verhoeven S, de Graaf C, Roumen L, Vroling B, Nabuurs SB, de Vlieg J, Klomp

636   JPG. 2011. Snooker: A Structure-Based Pharmacophore Generation Tool Applied to

637   Class A GPCRs. *Journal of Chemical Information and Modeling* 51:2277–2292. DOI:

638   10.1021/ci200088d.

639 Segala E, Guo D, Cheng RKY, Bortolato A, Deflorian F, Doré AS, Errey JC, Heitman LH,

640   IJzerman AP, Marshall FH, Cooke RM. 2016. Controlling the Dissociation of Ligands

641   from the Adenosine A2A Receptor through Modulation of Salt Bridge Strength. *Journal*

642   *of Medicinal Chemistry* 59:6470–6479. DOI: 10.1021/acs.jmedchem.6b00653.

643 Shoichet BK, Kobilka BK. 2012. Structure-based drug screening for G-protein-coupled

644   receptors. *Trends in Pharmacological Sciences* 33:268–272. DOI:

645   10.1016/j.tips.2012.03.007.

646 Smith RD, Damm-Ganamet KL, Dunbar JB, Ahmed A, Chinnaswamy K, Delproposto JE,

647   Kubish GM, Tinberg CE, Khare SD, Dou J, Doyle L, Stuckey JA, Baker D, Carlson HA.

648   2016. CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined

649   Computational Protein Design, Docking, and Scoring/Ranking Challenge. *Journal of*

650   *Chemical Information and Modeling* 56:1022–1031. DOI: 10.1021/acs.jcim.5b00387.

651 Souza SA, Kurohara DT, Dabalos CL, Ng HL. 2019. G Protein–Coupled Estrogen Receptor

652   Production Using an Escherichia coli Cell-Free Expression System. *Current Protocols in*

653   *Protein Science* 97:e88. DOI: 10.1002/cpps.88.

654 Tang H, Wang XS, Hsieh J-H, Tropsha A. 2012. Do crystal structures obviate the need for

655   theoretical models of GPCRs for structure-based virtual screening? *Proteins: Structure,*

656   *Function, and Bioinformatics* 80:1503–1521. DOI: 10.1002/prot.24035.

657 Wacker D, Fenalti G, Brown MA, Katritch V, Abagyan R, Cherezov V, Stevens RC. 2010.

658 Conserved Binding Mode of Human β2 Adrenergic Receptor Inverse Agonists and

659 Antagonist Revealed by X-ray Crystallography. *Journal of the American Chemical*

660 *Society* 132:11443–11445. DOI: 10.1021/ja105108q.

661 Wacker D, Stevens RC, Roth BL. 2017. How Ligands Illuminate GPCR Molecular

662 Pharmacology. *Cell* 170:414–427. DOI: 10.1016/j.cell.2017.07.009.

663 Wan S, Knapp B, Wright DW, Deane CM, Coveney PV. 2015. Rapid, Precise, and Reproducible

664 Prediction of Peptide–MHC Binding Affinities from Molecular Dynamics That Correlate

665 Well with Experiment. *Journal of Chemical Theory and Computation* 11:3346–3356.

666 DOI: 10.1021/acs.jctc.5b00179.

667 Wass MN, Sternberg MJE. 2009. Prediction of ligand binding sites using homologous structures

668 and conservation at CASP8. *Proteins: Structure, Function, and Bioinformatics* 77:147–

669 151. DOI: 10.1002/prot.22513.

670 Weiss DR, Bortolato A, Tehan B, Mason JS. 2016. GPCR-Bench: A Benchmarking Set and

671 Practitioners' Guide for G Protein-Coupled Receptor Docking. *Journal of Chemical*

672 *Information and Modeling* 56:642–651. DOI: 10.1021/acs.jcim.5b00660.

673 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure

674 and function prediction. *Nature Methods* 12:7–8. DOI: 10.1038/nmeth.3213.

675 Zhang J, Yang J, Jang R, Zhang Y. 2015. GPCR-I-TASSER: A Hybrid Approach to G Protein-

676 Coupled Receptor Structure Modeling and the Application to the Human Genome.

677 *Structure* 23:1538–1549. DOI: 10.1016/j.str.2015.06.007.

678 Zou, Ewalt, Ng. 2019. Recent Insights from Molecular Dynamics Simulations for G Protein-

679 Coupled Receptor Drug Discovery. *International Journal of Molecular Sciences* 20:4237.

680 DOI: 10.3390/ijms20174237.

Box 1(on next page)

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.

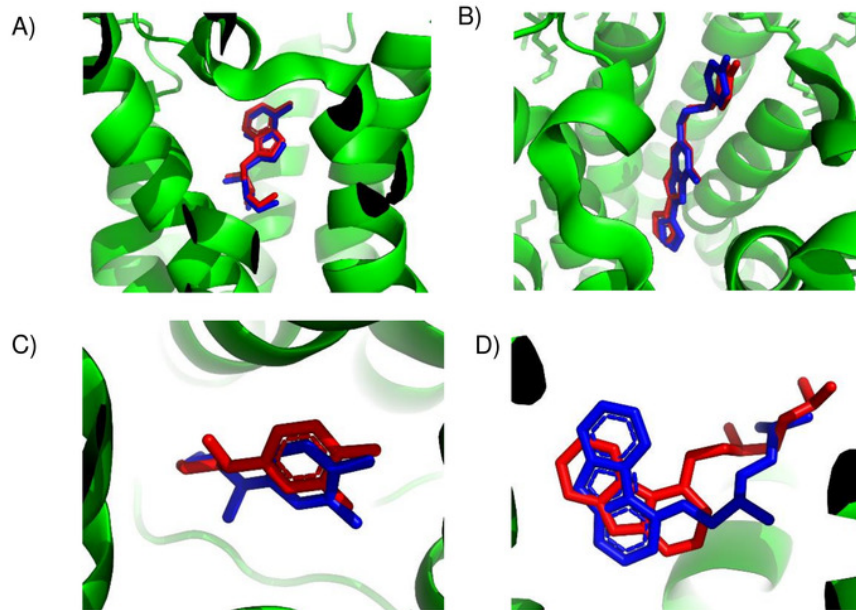
$$\textbf{Combined ConDock Score} = (\text{Conservation Score}) * (\text{Energy Score})$$

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

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

Figure 1

Predicted and experimental ligand binding sites in A2A adenosine and β2 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 β2 adrenergic receptor. D) Carazolol with β2 adrenergic receptor.

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 β2 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.

I)

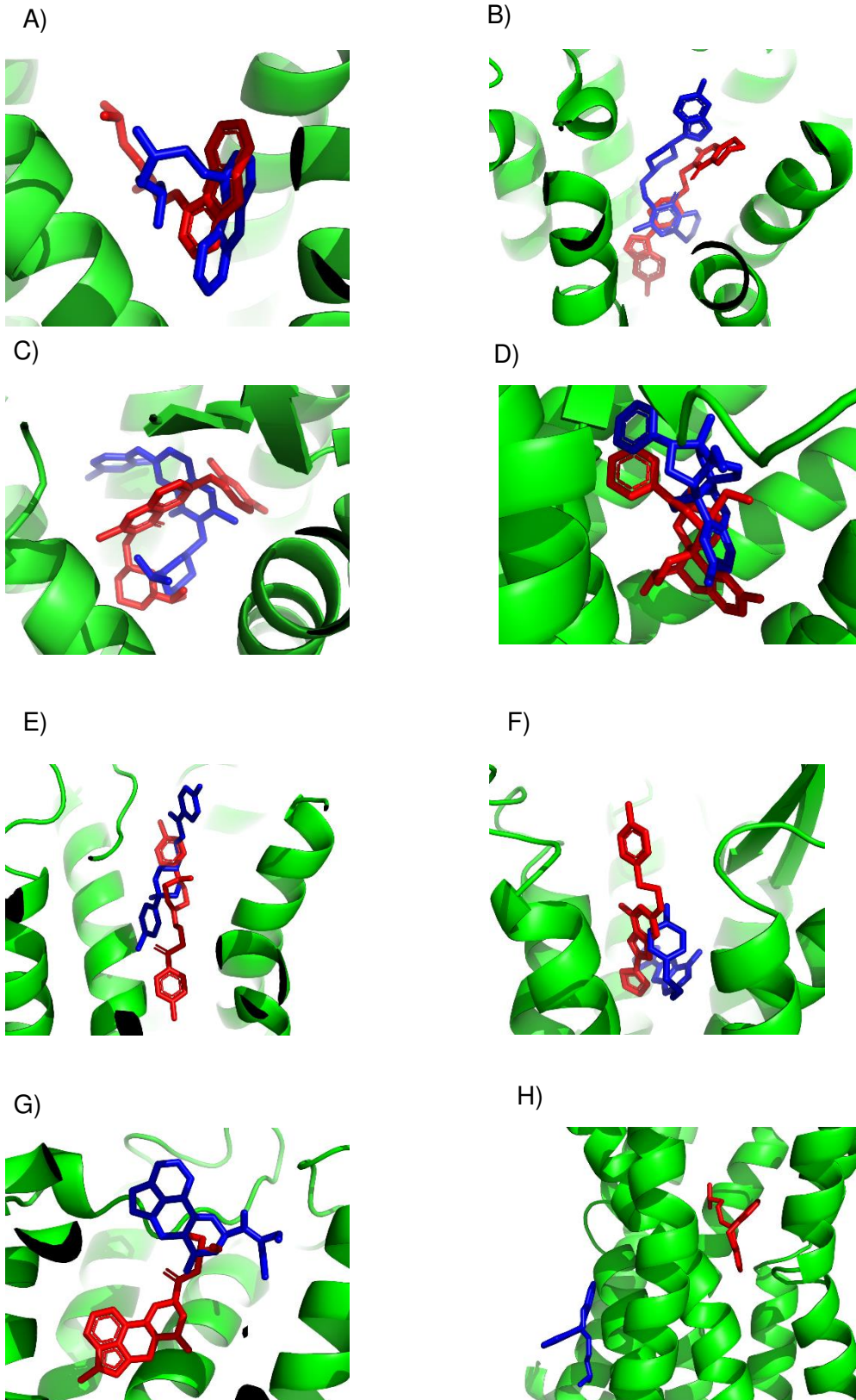
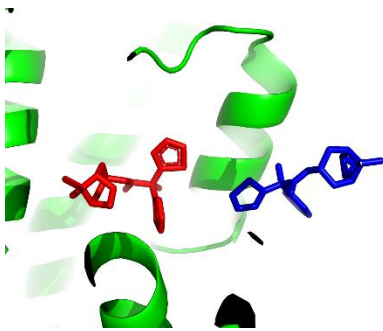



Figure 3

E2 binding sites calculated by SwissDock.

E2 poses are in blue. The top of the figure corresponds to the extracellular face of GPER.

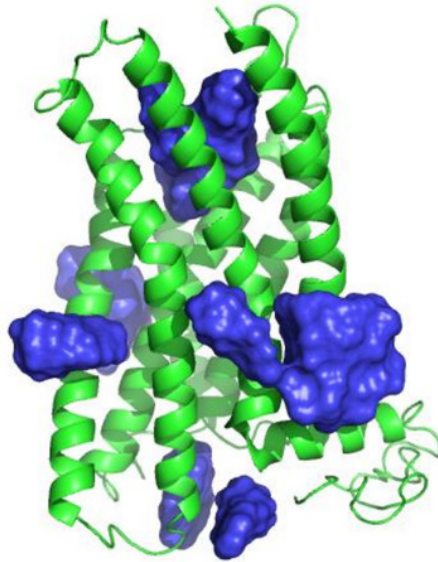



Figure 4

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.

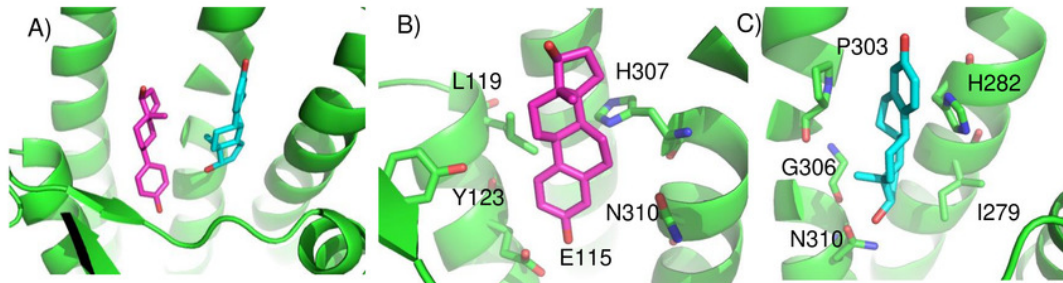



Figure 5

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.

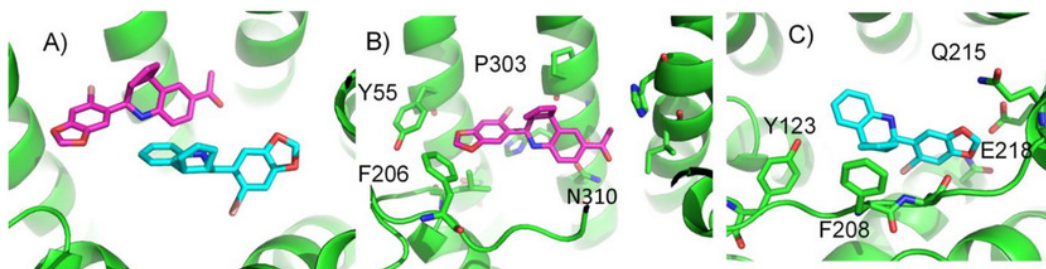



Figure 6

Predicted tamoxifen binding sites in GPER

A) The highest scoring docking poses for tamoxifen, pose 1 (maroon) and pose 2 (cyan).

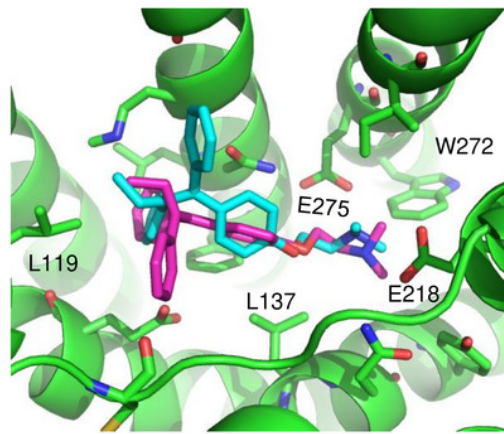



Figure 7

Predicted E2 binding sites by ConDockSite, CASTp, SiteHound, Concavity.

Ligand binding sites are colored, predicted by A) ConDockSite, B) CASTp, C) SiteHound, D) Concavity.