

# SCIENTIFIC REPORTS

OPEN

## Water Pharmacophore: Designing Ligands using Molecular Dynamics Simulations with Water

Sang Won Jung<sup>1,5</sup>, Minsup Kim<sup>1</sup>, Steven Ramsey<sup>2,4</sup>, Tom Kurtzman<sup>2,3,4</sup> & Art E. Cho<sup>1</sup>

In this study, we demonstrate a method to construct a water-based pharmacophore model which can be utilized in the absence of known ligands. This method utilizes waters found in the binding pocket, sampled through molecular dynamics. Screening of compound databases against this water-based pharmacophore model reveals that this approach can successfully identify known binders to a target protein. The method was tested by enrichment studies of 7 therapeutically important targets and compared favourably to screening-by-docking with Glide. Our results suggest that even without experimentally known binders, pharmacophore models can be generated using molecular dynamics with waters and used for virtual screening.

It is a fundamental tenet of drug design that, in order to potentially bind with high affinity to a given protein, a ligand must be complementary to that protein surface by donating and accepting hydrogen bonds and making hydrophobic contacts where appropriate<sup>1,2</sup>. Water molecules solvating a protein surface are complementary to the surface in that they donate and accept hydrogen bonds where appropriate and make corresponding van der Waals contacts with hydrophobic patches of the surface (Fig. 1)<sup>3,4</sup>. In this sense, water on a protein surface mimics the key interactions that a ligand must have in order to bind with high affinity to a targeted protein.

The construction of a pharmacophore is aimed at distilling the essential features that ligands must have to bind to a target<sup>5,6</sup>. The fact that water, when solvating a binding site, has many of these features suggests that a pharmacophore could be constructed based on an analysis of the hydration of an active site alone.

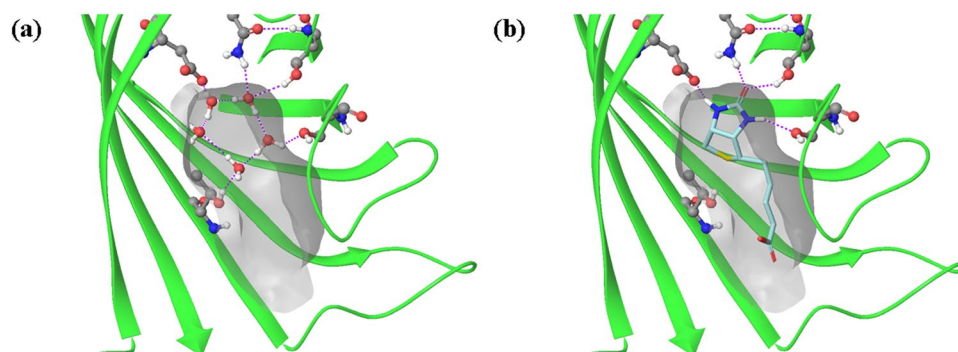
Here, we introduce a pharmacophore generation method that is based solely on the information provided from analysis of water interactions with the protein surface as observed in molecular dynamics simulations. As an initial application this approach was explored against the streptavidin-biotin system which is known to have highly ordered waters that correlate with ligand interactions. In further study, water-based pharmacophore models were constructed and screened against 7 pharmaceutically relevant target proteins from DUD, with use of their test sets<sup>7</sup>. In addition, we compare these results to screening of the same chemical library using conventional docking methodology with Glide<sup>8</sup>.

In this paper, we first detail the methodology used to construct the water-based pharmacophores, which we call simply water pharmacophore (WP), starting from the molecular dynamics simulations on a target protein. Then, we present the resulting pharmacophores for 7 different proteins, and demonstrate the results of screening and enrichment studies (Table 1). Finally, we discuss the results and suggest how the method can be implemented as either a stand-alone technique or can be incorporated into existing drug-discovery schemes in order to improve their efficacy.

### Methods

**Structure selection and preparation.** The X-ray crystal structures of 7 targets - Acetylcholinesterase (AChE)<sup>9</sup>, Androgen receptor (AR)<sup>10</sup>, Glucocorticoid receptor (GR)<sup>11</sup>, Peroxisome proliferator activated receptor gamma (PPAR $\gamma$ )<sup>12</sup>, Poly (ADP-ribose) polymerase (PARP)<sup>13</sup>, Progesterone receptor (PR)<sup>14</sup>, and Retinoid X receptor alpha (RXR $\alpha$ )<sup>15</sup> - were retrieved from the Protein Data Bank (PDB)<sup>16,17</sup>. These 7 targets were selected for this study from the original DUD test set based on their enrichment results in that study, with most systems

<sup>1</sup>Department of Bioinformatics, Korea University, Sejong, 30019, Korea. <sup>2</sup>Department of Chemistry, Lehman College, Bronx, New York, 10468, USA. <sup>3</sup>Ph.D. Program in Chemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York, New York, 10016, USA. <sup>4</sup>Ph.D. Program in Biochemistry, The Graduate Center of the City University of New York, 365 5th Avenue, New York, New York, 10016, USA. <sup>5</sup>Present address: Center for Supercomputing and Big Data, DGIST, Daegu, 42988, Korea. Correspondence and requests for materials should be addressed to T.K. (email: [simpleliquid@gmail.com](mailto:simpleliquid@gmail.com)) or A.E.C. (email: [artcho@korea.ac.kr](mailto:artcho@korea.ac.kr))



**Figure 1.** (a) Water and (b) biotin in the active site of streptavidin. Note how the water molecules and the ligand make similar contacts to the protein surface.

Target	Number of actives	Number of decoys	PDB ID
Acetylcholinesterase (AChE)	107	3892	4EY7
Androgen receptor (AR)	79	2854	1XQ2
Glucocorticoid receptor (GR)	78	2947	1M2Z
Peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ )	85	3127	1ZE0
Poly(ADP-ribose) polymerase (PARP)	35	1351	1EFY
Progesterone receptor (PR)	27	1041	1SR7
Retinoid X receptor alpha (RXR $\alpha$ )	20	750	1MVC

**Table 1.** Targets and DUD sets for enrichment studies.

producing maximum enrichment factors ( $EF_{\max}$ )  $\sim 50$  with one producing a high  $EF_{\max}$  of  $>100$  (RXR $\alpha$ ) and a few producing low  $EF_{\max}$  of  $\leq 5$  (AChE, PPAR $\gamma$ , and PR)<sup>7</sup>. These structures were prepared by protein preparation wizard (PPW) module<sup>18</sup> of Schrödinger suite. PPW adds hydrogen and neutralizes side chains that are neither close to binding cavity nor involved in the formation of salt bridges. In the next step, water molecules are removed and hydrogen atoms are added to the structure, at the most likely positions of hydroxyl and thiol hydrogen atoms. Protonation states and tautomers of His residue and Chi “flip” assignment for Asn, Gln, and His residues are selected during this step as well. Finally, minimization was performed until the average RMSD of non-hydrogen atoms reaches 0.3 Å.

**Molecular dynamics simulations.** Molecular dynamics simulations were performed using AMBER12 software<sup>19,20</sup> with AMBER99SB force field<sup>21</sup> and TIP3P water model<sup>22</sup>. Ligands were not simulated; however, protein bound conformations were restrained during sampling at 2.5 kcal/mol·Å<sup>2</sup> so as to sample water which could be found in a bound state binding site. Protein systems were encased in a cubic of TIP3P water molecules in which each edge had a minimum distance of 10 Å from protein heavy atoms with cubic periodic conditions. First, 10,000 steps of minimization with 100 kcal/mol·Å<sup>2</sup> restraints on all protein heavy atoms was performed. Then, the system was heated with the same restraints for 100 ps until the system reached 300 K in NPT conditions. Several simulation steps were then taken to slowly reduce the restraints on protein heavy atoms from 100 to 10 kcal/mol·Å<sup>2</sup>, these steps total 350 ps simulation time, again in NPT conditions. At this stage another minimization was performed to ensure no bad contacts occurred as the restraint strength has been dropped rapidly, this minimization is performed for 10,000 steps at 10 kcal/mol·Å<sup>2</sup> restraints on heavy atoms. After this several NPT steps were performed to slowly release restraints from 10 kcal/mol·Å<sup>2</sup> to 2.5 kcal/mol·Å<sup>2</sup>, these steps total 1.2 ns of simulation time. Finally, once the system is fully equilibrated, heated, and relaxed a production simulation was conducted under NVT conditions for 10 ns simulation time sampling every 1 ps with 2.5 kcal/mol·Å<sup>2</sup> restraints and 8 Å non-bonding interaction cutoff. Data from this simulation was utilized to produce hydration sites. Langevin dynamics<sup>23</sup> are utilized to stabilize simulation temperatures and NPT simulations were performed with the Berendsen barostat<sup>24</sup>.

**Method for constructing ligand-based pharmacophore.** Ligand-based pharmacophore hypotheses were constructed in this study to serve as a visual comparison to the water-based pharmacophores for each system. We used PHASE module<sup>25,26</sup> of Schrödinger suite with the native ligand poses to generate ligand-based pharmacophore models.

**Hydration site analysis.** Hydration site analysis (HSA) was performed on waters near the binding site in each of the 7 pharmaceutically relevant targets (Tables S1–7). Water was determined to be in the binding pocket by proximity (within 5 Å) to the crystallized ligands. Hydration sites were selected by finding 1 Å spheres which

contain the highest count of oxygen atom of water, these locales are recorded and all waters found within are excluded from becoming unique sites<sup>4</sup>. Selected sites are only considered for analysis if they contain twice bulk density under similar conditions (for 10,000 samples this implies greater than 2,800 oxygen atom of water found). These sites once generated were analyzed for thermodynamic qualities such as energy and hydrogen bonding characteristics each of which would be considered in determining water-based pharmacophore feature viability. Energy was determined by calculating system energy with and without the water found within the site, translational entropy through the histogram method, and orientational entropy through nearest neighbors.

**Pharmacophore feature assignment.** Each identified hydration site could potentially be treated as a WP feature; however, pharmacophore screening tends to produce optimal results when 4–8 feature criteria are utilized. Subsequently, making each hydration site a feature would not be an ideal method. In an effort to design a technique which would identify key contacts without known binders, a method of screening and assigning features to hydration sites was designed so that the method could be utilized blind to a target and produce reasonable results. The detailed criteria are summarized in Fig. S1. This scheme was created by thermodynamic principles and then the detailed parameters were optimized through a trial and error selection trained against our test systems. Starting with differentiation by acceptor and donor ratio, each hydration site is filtered through series of criteria by energy values and acceptor/donor ratio. In case of aromatic ring and hydrophobic features, both hydrogen bond acceptor and donor ratio should be lower than 100%. Then, to differentiate the two features, SiteMap module<sup>27,28</sup> of Schrödinger suite was utilized to determine the surface area of the hydrophobic surface. A site with favorable enthalpy and acceptor/donor ratio was determined to either be a hydrogen bond donor or acceptor site based on its bonding characteristics. Both hydrogen bond donor and acceptor have enthalpic energies values less than  $-8.0$  kcal/mol. In case of hydrogen bond donor, the acceptor and donor ratios should be less than 50% and more than 100%, respectively. Conversely, the acceptor and donor ratios should be more than 100% and less than 50%, respectively, in case of hydrogen bond acceptor. If there are 2 hydrogen bond acceptor features within  $3.5$  Å of each other, they are combined to be a negative feature. To differentiate between hydrogen bond donor and positive features the length of the bond was utilized where greater than  $1.5$  Å describes a donor feature and less than  $1.5$  Å describes a positive feature.

**Pharmacophore model generation.** Pharmacophore features assigned as in the previous section were used to construct a pharmacophore model using the PHASE module of the Schrödinger suite. The detailed criteria are summarized in Fig. S2. This scheme was created to efficiently select essential features and exclude the redundant features, especially hydrophobic features. The algorithm reduces the pharmacophore features to a manageable number typically less than 8 as pharmacophore screening becomes inefficient with more than 8 features. In addition, the scheme optimizes the position of each assigned feature by hydrogen-bond-constrained docking or energy minimization. In case of negative feature, two pharmacophore features are combined to form one negative feature as previously described. A negative feature position is optimized via hydrogen-bond-constraint docking with acetic acid ( $\text{CH}_3\text{COO}^-$ ) as ligand. In case of hydrogen bond donor feature, the position is adjusted via hydrogen-bond-constraint docking with water molecule ( $\text{H}_2\text{O}$ ) as ligand when the enthalpy energy is more than  $-9.0$  kcal/mol. At last, in case of hydrophobic feature, the position is optimized via energy minimization while converting hydration site to methane molecule ( $\text{CH}_4$ ) when the enthalpy energy is more than  $-8.2$  kcal/mol. Among the hydrophobic features, sites with favorable energies are excluded from the energy minimization. Consequently, a water-based pharmacophore model with a set of optimized pharmacophore features is constructed (Tables S1–7).

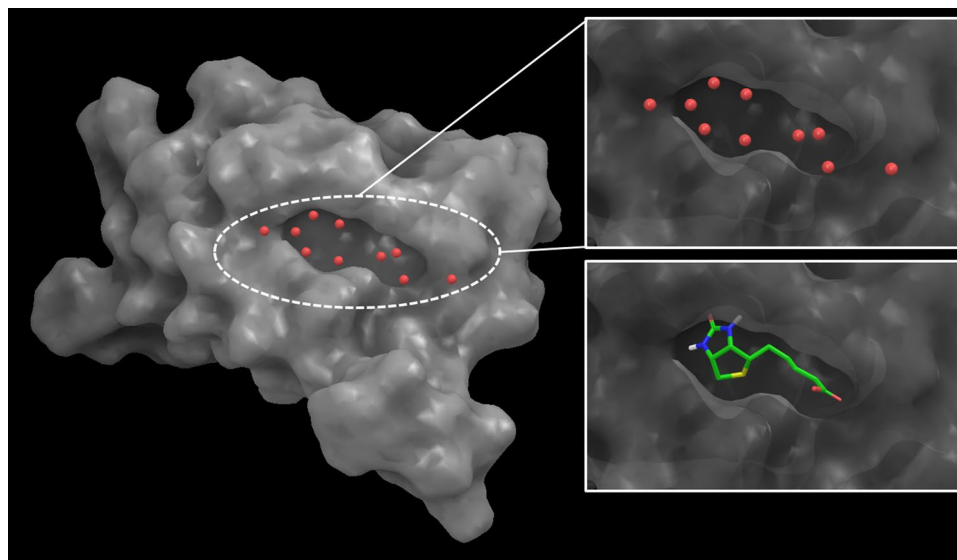
**Pharmacophore screening.** In order to validate our WP model, the PHASE module of Schrödinger suite was utilized for screening against the pharmacophore models constructed for popular targets (Fig. S3). Conformers of the ligands were generated using the ConfGen module<sup>29</sup> of Schrödinger suite for screening. The distance matching tolerance used was  $1.5$  Å and other parameters in the default settings so that hits were rejected if their alignment scores were greater than 1.2, their vector scores were less than  $-1.0$ , or volume scores were less than 0.0, or any combination thereof.

**Molecular docking.** The set of ligands for 7 targets were docked to the binding site of each target proteins and by using the Glide module<sup>8</sup> of Schrödinger suite. Glide uses grids for fast scoring; the grid-generation module was used to generate grids for the binding site of each target proteins. The van der Waals scaling and partial charge cutoff was set to 0.8 and 0.15, respectively. Next, the standard precision (SP) and extra precision (XP) mode of Glide was used to rank the ligands in the order of the lowest docking score.

**Enrichment factor analysis.** Pharmacophore effectiveness was validated via enrichment factor assessment of DUD molecule sets<sup>7</sup>. Enrichment is the likelihood of selecting known active compounds from a portion of highly ranked hits during screening compared to randomly selected from the original molecule set. The enrichment factor is represented by

$$\text{EF} = (\text{Hits}_{\text{sampled}} / \text{N}_{\text{sampled}}) / (\text{Hits}_{\text{total}} / \text{N}_{\text{total}})$$

where EF is enrichment factor,  $\text{Hits}_{\text{sampled}}$  is the number of true hits in the hit list,  $\text{N}_{\text{sampled}}$  is the number of compounds in the hit list,  $\text{Hits}_{\text{total}}$  is the number of hits in the full database,  $\text{N}_{\text{total}}$  is the number of compounds in the full database. In this research, we calculated enrichment factors for the actives found in the top scoring 1%, 5%, and 10% of the total compounds screened.



**Figure 2.** Hydration sites found in the binding site of streptavidin compared to the binding pose of its ligand biotin.

Target	Number of Hydration sites	MODEL	EF 1%	EF 5%	EF 10%
AR*	11	DHHH	15.4	7.1	4.2
PR*	18	ADDHHH	21.6	8.2	4.4
RXR $\alpha$	9	—	—	—	—
GR*	19	DHHH	19.4	6.7	3.3
PPAR $\gamma$	25	DDHHHN	1.2	0.2	0.1
PARP*	36	ADDDR	22.6	5.2	2.6
AChE	43	DDDHHH	3.7	0.7	0.4

**Table 2.** Enrichment analysis results using DUD sets for 7 targets by water-based pharmacophore screening. Enrichment factors are calculated for 1, 5, 10% of original data set. ‘—’ indicates that the model does not generate and evaluate the enrichment analysis. Entries marked with asterisk (\*) are targets in which water-pharmacophore outperformed docking in EF 1%.

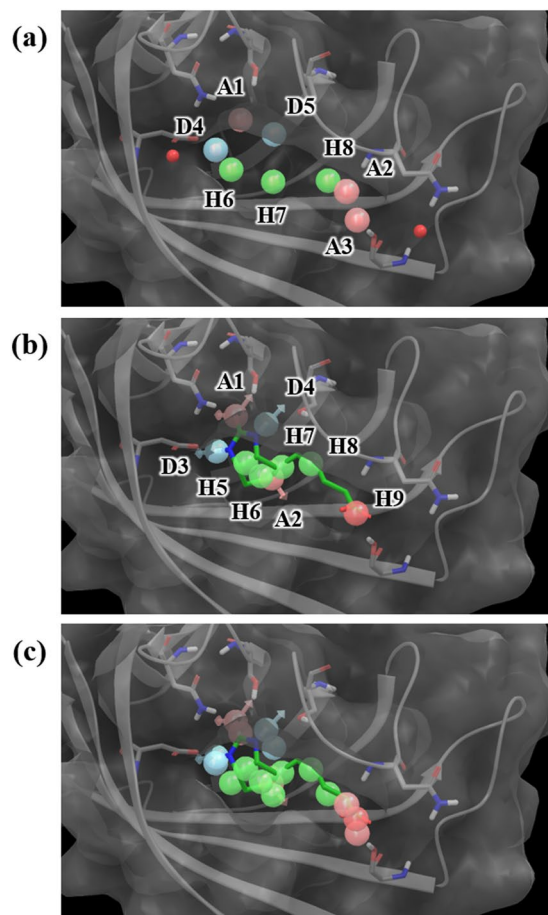
## Results and Discussion

As shown in Fig. 2, 10 high density hydration sites that were generated from the MD simulations of the solvated streptavidin active site. Using our WP algorithm, we assigned pharmacophore feature to each of the 8 hydration sites out of 10, which were selected according to our algorithm. The resulting water-based pharmacophore is shown in Fig. 3a, alongside a ligand based pharmacophore hypothesis constructed off of biotin in Fig. 3b. An overlay of the water-based and ligand-based pharmacophores is shown in Fig. 3c. It appears that the ligand-based and water-based streptavidin pharmacophores overlap to a significant degree.

Using the same process, we then generated water-based pharmacophore and ligand-based pharmacophore for the 7 aforementioned targets.

**Enrichment study results.** Tables 2 and 3 summarizes the results of enrichment study on the 7 targets.

**Androgen receptor (AR).** AR is a nuclear receptor which consists of N-terminal domain, DNA-binding domain (DBD), and ligand-binding domain (LBD). If androgens, such as testosterone and 5 $\alpha$ -dihydrotestosterone, bind to LBD, AR binds with DNA in the form of homodimer and regulate gene expression. AR is involved in a wide variety of physiological processes, especially control of male sexual differentiation. Since abnormality in AR function can lead to prostate cancer or benign prostatic hyperplasia (BHP), AR is an important drug target<sup>30,31</sup>. We used PDB structure 1XQ2, which was crystallized with the ligand metribolone. Metribolone produces 7 total pharmacophore features when run through PHASE (AADHHHH) (Fig. 4a). We identified 11 hydration sites inside the LBD of AR (Fig. S4 and Table S1). Our WP procedure yielded 4 features DHHH (Fig. 4b). The WP features overlap with the ligand pharmacophore features. Water-pharmacophore performed slightly better than docking in enrichment results at 1% and 5%, while docking performed slightly better than WP in enrichment results at 10% (Tables 2 and 3).

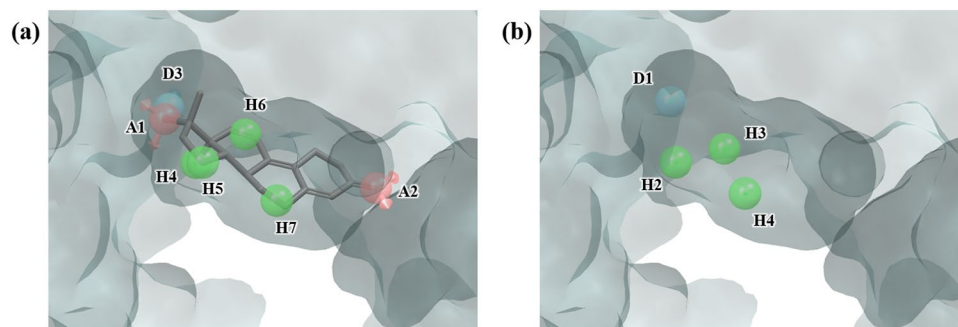


**Figure 3.** (a) Water- and (b) ligand-based pharmacophore model of biotin. (c) The two models overlaid in the binding site of streptavidin. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green; negative (N), red.

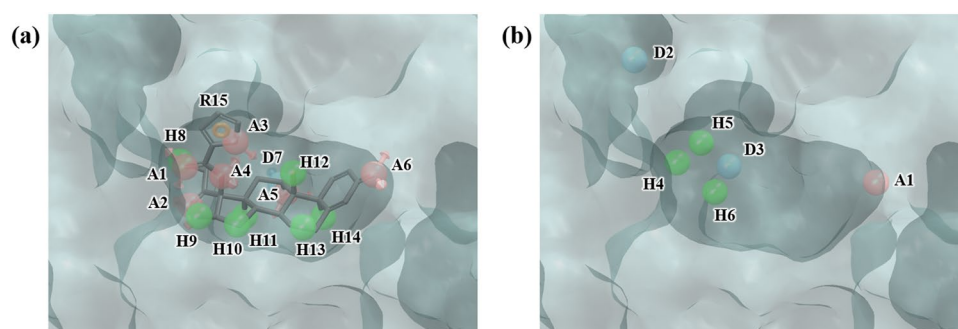
Target	SP			XP		
	EF 1%	EF 5%	EF 10%	EF 1%	EF 5%	EF 10%
AR	11.5	6.6	5.3	10.2	6.6	4.8
PR	14.4	5.2	2.6	10.8	5.2	3.0
RXR $\alpha$	24.1	14.8	8.5	19.3	11.8	8.0
GR	5.2	2.3	1.5	5.2	2.8	2.2
PPAR $\gamma$	23.6	9.7	5.7	27.1	13.7	7.8
PARP	0.0	0.0	0.0	0.0	0.6	0.3
AChE	25.2	8.8	5.5	22.4	9.7	6.1

**Table 3.** Enrichment analysis results using DUD sets for 7 targets by Glide SP and XP docking.

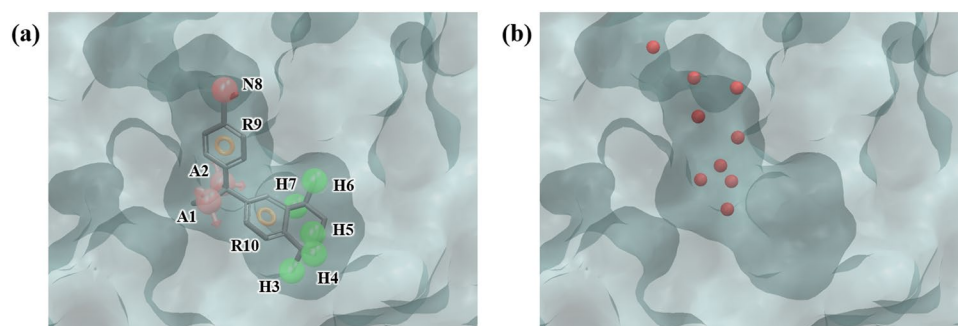
**Progesterone receptor (PR).** PR is another nuclear receptor which when combined with progesterone in cytosol would form a dimer and then regulate gene expression by binding DNA thereby controlling female pregnancy and menstruation. PR is a therapeutic target for miscarriage, uterine cancer, or breast cancer<sup>32–34</sup>. In 1SR7, the PDB structure used for our experiments, mometasone furoate, a potent steroid agonist, is bound in the ligand binding pocket of PR. PHASE produces a total of 15 ligand pharmacophore features (AAAAAADHHHHHHR) for mometasone furoate. Hydration site analysis yielded 18 hydration sites (Fig. S5 and Table S2), some of which coincide with the ligand pharmacophore features (Fig. 5a). The WP generated 6 features, ADDHHH (Fig. 5b). Of these, except for the D2 and H5 features which is located on outside of binding site and on the ring group of the ligand (R15), all the other features overlap with the ligand pharmacophore features. Overall, screening with WP model resulted in better enrichment than screening by docking in enrichment results at 1%, 5%, and 10% (Tables 2 and 3).



**Figure 4.** Ligand- and water-based pharmacophore model of AR. (a) 7 pharmacophore features of the ligand metribolone, AADHHHH, were generated by PHASE. (b) 4 pharmacophore features, DHHH, were generated by water pharmacophore model generation method. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green. To clarify the binding site of AR, the molecular surface is only shown.

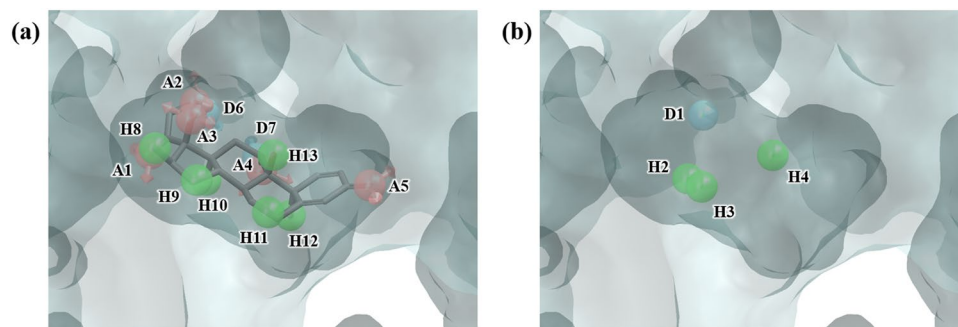


**Figure 5.** Ligand- and water-based pharmacophore model of PR. (a) 15 pharmacophore features of the ligand mometasone furoate, AAAAAADHHHHHHHR, were generated by PHASE. (b) 6 pharmacophore features, ADDHHH, were generated by water pharmacophore model generation method. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green; aromatic ring (R), orange. To clarify the binding site of PR, the molecular surface is only shown.

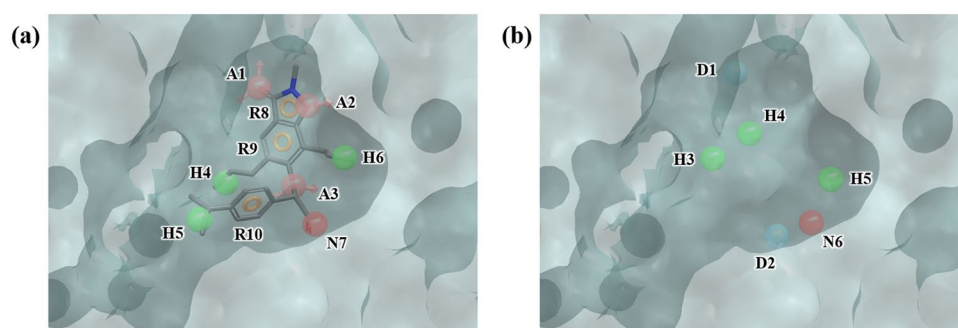


**Figure 6.** Ligand-based pharmacophore model and hydration sites of RXR $\alpha$ . (a) 10 pharmacophore features of the ligand BMS649, AAHHHHNRR, were generated by PHASE. (b) 9 hydration sites generated by MD simulations and hydration site analysis were displayed as red spheres. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrophobic (H), green; negative (N), red; aromatic ring (R), orange. To clarify the binding site of RXR $\alpha$ , the molecular surface is only shown.

**Retinoid X receptor alpha (RXR $\alpha$ ).** RXR $\alpha$  is a nuclear receptor and plays a role as a transcription factor controlling various physiological processes such as cell development, apoptosis, and homeostasis. The natural ligand for RXR $\alpha$  is 9-*cis*-retinoic acid (9-*cis*-RA) and together with synthetic ligands, it is known to be effective in inhibition of tumor formation rendering RXR $\alpha$  to be an anticancer drug target<sup>35,36</sup>. In the PDB structure 1MVC, which is used for our test, a synthetic agonist compound BMS649 is bound to the ligand binding domain (LBD).



**Figure 7.** Ligand- and water-based pharmacophore model of GR. **(a)** 13 pharmacophore features of the ligand dexamethasone, AAAAAADDDHHHHHH, were generated by PHASE. **(b)** 4 pharmacophore features, DHHH, were generated by water pharmacophore model generation method. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green. To clarify the binding site of GR, the molecular surface is only shown.

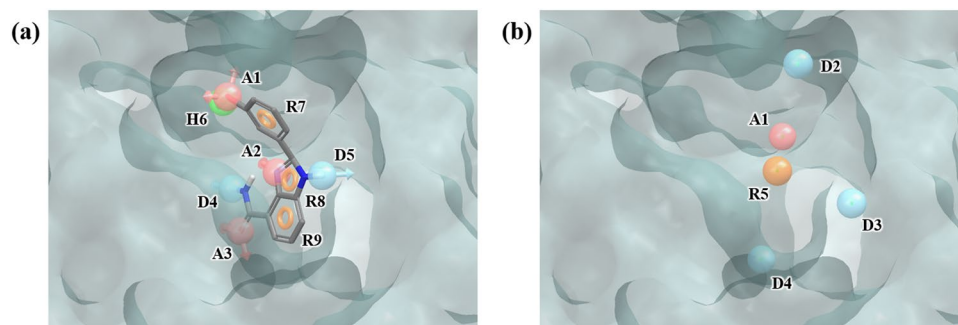


**Figure 8.** Ligand- and water-based pharmacophore model of PPAR $\gamma$ . **(a)** 10 pharmacophore features of the ligand  $\alpha$ -aryloxyphenylacetic acid, AAAHHHNRRR, were generated by PHASE. **(b)** 6 pharmacophore features, DDHHHN, were generated by water pharmacophore model generation method. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green; negative (N), red. To clarify the binding site of PPAR $\gamma$ , the molecular surface is only shown.

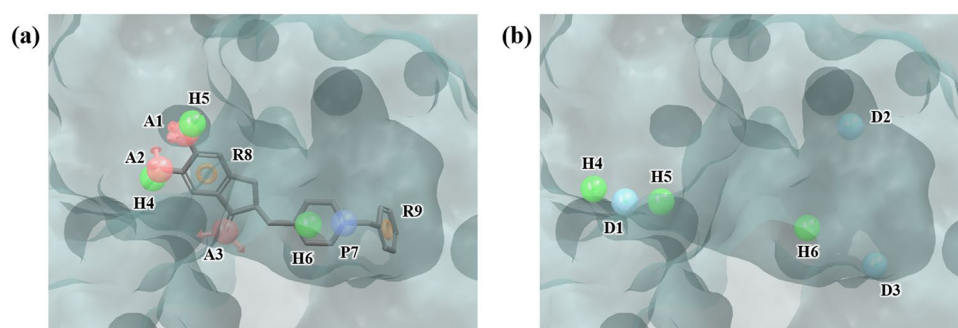
BMS649 produces 10 pharmacophore features when run through PHASE, AAHHHHHNRR (Fig. 6a). Hydration site analysis yielded 9 hydration sites (Fig. S6 and Table S3), but they were all near the opening of the binding site, rather than the LBD (Fig. 6b). This is due to the fact that the binding site is hydrophobic and thus the water molecules cannot penetrate deep inside the pocket prohibiting proper water sampling in the LBD (Fig. S11). By the criteria for WP, no feature was generated, and thus no compound was screened. This case confirms that our method of WP does not suit well the cases in which binding pockets are strongly hydrophobic and water is not sampled thoroughly. On the other hand, docking showed better enrichment results at 1%, 5%, and 10% (Tables 2 and 3).

**Glucocorticoid receptor (GR).** GR is a nuclear receptor which binds DNA when combined with steroid hormones and acts as a transcription factor. It is involved in glucose homeostasis, bone turnover, cell differentiation, lung maturation, and inflammation. Abnormality in GR can cause Cushing's syndrome, autoimmune diseases, and various kinds of cancer and thus has been studied as an important therapeutic target<sup>37–39</sup>. The structure used in our experiment is 1M2Z, in which dexamethasone is bound in the ligand binding pocket of GR. PHASE produces 13 pharmacophore features (AAAAADDDHHHHHH) (Fig. 7a). Hydration site analysis produced 19 hydration sites (Fig. S7 and Table S4), from which a 4-featured WP model, DHHH, was generated (Fig. 7b). In an enrichment study, the WP model screening yielded significantly better result than screening by docking in enrichment results at 1%, 5%, and 10% (Tables 2 and 3).

**Peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ).** PPAR $\gamma$  is a nuclear receptor, which controls fatty acid storage and glucose metabolism. It is related to various diseases, such as obesity, diabetes, atherosclerosis, and cancer<sup>40–42</sup>. The structure used in our experiment 1FM9, is bound with  $\alpha$ -aryloxyphenylacetic acid, which is a PPAR $\gamma$  agonist. PHASE produced 10 pharmacophore features (AAHHHNRRR) (Fig. 8a). Hydration site analysis yielded 25 hydration sites (Fig. S8 and Table S5) and from these 6 WP features were selected, DDHHHN (Fig. 8b). Although WP screening selected a few actives, enrichment factors for WP were minimal, with positive number only for 1% case. In comparison, docking performed well with enrichment factors



**Figure 9.** Ligand- and water-based pharmacophore model of PARP. **(a)** 9 pharmacophore features of the ligand 2-(3-methoxyphenyl)-1H-benzimidazole-4-carboxamide, AAADDHRRR, were generated by PHASE. **(b)** 5 pharmacophore features, ADDDR, were generated by water pharmacophore model generation method. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green; aromatic ring (R), orange. To clarify the binding site of PARP, the molecular surface is only shown.



**Figure 10.** Ligand- and water-based pharmacophore model of AChE. **(a)** 9 pharmacophore features of the ligand donepezil, AAAHHHPRR, were generated by PHASE. **(b)** 6 pharmacophore features, DDDHHH, were generated by water pharmacophore model generation method. The pharmacophore features are numbered to help identification of each site and colored as follows: hydrogen bond acceptor (A), pink; hydrogen bond donor (D), skyblue; hydrophobic (H), green; positive (P), blue; aromatic ring (R), orange. To clarify the binding site of AChE, the molecular surface is only shown.

of more than 10 across 1%, 5%, and 10% (Tables 2 and 3). This case reveals the limitation of WP method. With large binding sites, pharmacophore models, for their imprecision on geometry of the ligands, often fail to filter out ligands that may not have correct interactions with the binding site.

**Poly (ADP-ribose) polymerase (PARP).** PARP is a nuclear enzyme, which plays an important role in DNA repair and is related to cancer. It was found that inhibition of PARP in tumor cells can strengthen the effect of radiotherapy and DNA-targeted chemotherapy<sup>43</sup>. Our HSA yielded 36 hydration sites, which was considerably more than other targets. PHASE produced 9 pharmacophore features on the ligand 2-(3-methoxyphenyl)-1H-benzimidazole-4-carboxamide (AAADDHRRR) (Fig. 9a) and the WP model 5 features of ADDDR (Fig. 9b) was generated from 36 hydration sites (Fig. S9 and Table S6). In enrichment studies, docking failed to select active ligands at all level (Tables 2 and 3). This failure may be explained by observing that the binding site is significantly larger than a usual “fragment-like” ligand and therefore differentiating the actual actives from decoys becomes more difficult as there are much more possibilities of fitting ligands in the binding site. On the other hand, WP performed rather well, achieving enrichment factor of 22.6 at 1%, 5.2 at 5%, and 2.6 at 10%.

**Acetylcholinesterase (AChE).** AChE is an enzyme that hydrolyzes a neurotransmitter acetylcholine. It is involved in synaptic transmission and thus is an important therapeutic target for neurodegenerative diseases such as Alzheimer's disease<sup>44</sup>. The structure 4EY7, used in our experiment, is bound with donepezil, which is known to be effective for neurodegenerative diseases. PHASE produced 9 ligand pharmacophore features for donepezil, AAAHHHPRR (Fig. 10a). HSA produced 43 hydration sites (Fig. S10 and Table S7) and the WP algorithm produced 6 features of DDDHHH (Fig. 10b). In this case, WP did not perform as well as docking at all level (Tables 2 and 3). This poor performance by WP is probably due to the fact that the binding site is large and narrow, which makes pharmacophore model-based filtering process vulnerable to exact geometry of the compound structures. While WP can produce compounds that have relevant interactions with the receptor, these compounds do not necessarily have the right dimensions as pharmacophore filtering do not consider geometries of the ligands being filtered.



Our WP method successfully generated the pharmacophore models derived from the water molecules without known ligands for all 7 targets. When compared with the pharmacophore models derived from known ligands, our pharmacophore models had similar features which are essential for binding target proteins. This shows that even without any knowledge of active ligands, WP can generate a pharmacophore model solely from the binding site structure. An enrichment study was conducted to validate the performance of the method as compared with Glide docking which has been proven to perform well on a wide variety of targets. The results showed that WP outperformed Glide docking in 4 targets out of 7 in enrichment factor at 1%. Time-wise, although WP requires MD simulations beforehand, pharmacophore filtering takes much less than docking. The limitations of WP were also observed when the binding sites were too hydrophobic like RXR $\alpha$  case. Hydrophobic regions in protein binding sites are almost always hydrated and hence WP should be able to capture those points. However, in the case of RXR $\alpha$ , the entrance to the binding site is very narrow and therefore water molecules do not freely venture into the binding site during MD simulations. This fact results in an inadequate sampling of the hydration points, which in turn gives a pharmacophore model that may be non-druglike leading to poor enrichment performance. WP also had problem with large pockets that can accommodate large sized ligands such as PPAR $\gamma$  and AChE (Fig. S11). In these cases, we believe that WP can be combined with docking in order to improve its efficacy. Further work is under way to heighten the usefulness of WP method along this line.

## Conclusions

In this paper, we demonstrated the feasibility of generating pharmacophore models based purely on the receptor structure through probing the protein binding-site surface with explicit water molecular dynamics simulations. We have introduced a method to construct the water-based pharmacophore and demonstrated that such pharmacophore is able to explore chemical space that is explored using more traditional ligand-based approaches. We have further demonstrated that water pharmacophore can be used for virtual ligand screening processes with a performance that compares well with established docking methods shown by enrichment studies. Our WP method selects feature sites based on thermodynamic criteria, therefore selected sites are more likely to be meaningful interaction features than random selection. This technique can be used as a standalone approach such as pharmacophore-based virtual screening when known binder data is lacking. The technique also can be incorporated into the ligand-based approaches. The more important molecular features derived from the known binder data can be selected by the consideration of WP information. In addition, structure-based approaches such as docking can be combined with WP. The two methods can then be applied sequentially in virtual screening. We envision that introducing localized solvation thermodynamics through Grid Inhomogeneous Solvation Theory<sup>45</sup> or hydration site approaches such as WaterMap<sup>46</sup> and STOW<sup>47</sup> could help assign weights to individual pharmacophore sites and help improve searching and scoring schemes.

## References

- Halperin, I., Ma, B. Y., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **47**, 409–443, <https://doi.org/10.1002/prot.10115> (2002).
- Brooijmans, N. & Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Bioph Biom* **32**, 335–373, <https://doi.org/10.1146/annurev.biophys.32.110601.142532> (2003).
- Kim, B. C., Young, T., Harder, E., Friesner, R. A. & Berne, B. J. Structure and dynamics of the solvation of bovine pancreatic trypsin inhibitor in explicit water: A comparative study of the effects of solvent and protein polarizability. *J Phys Chem B* **109**, 16529–16538, <https://doi.org/10.1021/jp051569v> (2005).
- Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *P Natl Acad Sci USA* **104**, 808–813, <https://doi.org/10.1073/pnas.0610202104> (2007).
- van Drie, J. H. Pharmacophore discovery - Lessons learned. *Curr Pharm Design* **9**, 1649–1664, <https://doi.org/10.2174/1381612033454568> (2003).
- Lippert, T., Schulz-Gasch, T., Roche, O., Guba, W. & Rarey, M. De novo design by pharmacophore-based searches in fragment spaces. *J Comput Aid Mol Des* **25**, 931–945, <https://doi.org/10.1007/s10822-011-9473-6> (2011).
- Huang, N., Shoichet, B. K. & Irwin, J. J. Benchmarking sets for molecular docking. *J Med Chem* **49**, 6789–6801, <https://doi.org/10.1021/jm0608356> (2006).
- Friesner, R. A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **47**, 1739–1749, <https://doi.org/10.1021/jm0306430> (2004).
- Cheung, J. *et al.* Structures of Human Acetylcholinesterase in Complex with Pharmacologically Important Ligands. *J Med Chem* **55**, 10282–10286, <https://doi.org/10.1021/jm300871x> (2012).
- He, B. *et al.* Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell* **16**, 425–438, <https://doi.org/10.1016/j.molcel.2004.09.036> (2004).
- Bledsoe, R. K. *et al.* Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110**, 93–105, [https://doi.org/10.1016/S0092-8674\(02\)00817-6](https://doi.org/10.1016/S0092-8674(02)00817-6) (2002).
- Shi, G. Q. *et al.* Design and synthesis of alpha-aryloxyphenylacetic acid derivatives: A novel class of PPAR alpha/gamma dual agonists with potent antihyperglycemic and lipid modulating activity. *J Med Chem* **48**, 4457–4468, <https://doi.org/10.1021/jm0502135> (2005).
- White, A. W. *et al.* Resistance-modifying agents. 9. Synthesis and biological properties of benzimidazole inhibitors of the DNA repair enzyme poly(ADP-ribose) polymerase. *J Med Chem* **43**, 4084–4097, <https://doi.org/10.1021/jm000950v> (2000).
- Madauss, K. P. *et al.* Progesterone receptor ligand binding pocket flexibility: Crystal structures of the norethindrone and mometasone furoate complexes. *J Med Chem* **47**, 3381–3387, <https://doi.org/10.1021/jm030640n> (2004).
- Egea, P. F., Mitschler, A. & Moras, D. Molecular recognition of agonist Ligands by RXRs. *Mol Endocrinol* **16**, 987–997, <https://doi.org/10.1210/me.16.5.987> (2002).
- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242, <https://doi.org/10.1093/nar/28.1.235> (2000).
- Rose, P. W. *et al.* The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res* **45**, D271–D281, <https://doi.org/10.1093/nar/gkw1000> (2017).
- Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aid Mol Des* **27**, 221–234, <https://doi.org/10.1007/s10822-013-9644-8> (2013).
- Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688, <https://doi.org/10.1002/jcc.20290> (2005).

20. Pearlman, D. A. *et al.* Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput Phys Commun* **91**, 1–41, [https://doi.org/10.1016/0010-4655\(95\)00041-D](https://doi.org/10.1016/0010-4655(95)00041-D) (1995).
21. Hornak, V. *et al.* Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **65**, 712–725, <https://doi.org/10.1002/prot.21123> (2006).
22. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **79**, 926–935, <https://doi.org/10.1063/1.445869> (1983).
23. Pastor, R. W., Brooks, B. R. & Szabo, A. An Analysis of the Accuracy of Langevin and Molecular-Dynamics Algorithms. *Mol Phys* **65**, 1409–1419, <https://doi.org/10.1080/00268978800101881> (1988).
24. Allen, M. P. & Wilson, M. R. Computer-Simulation of Liquid-Crystals. *J Comput Aid Mol Des* **3**, 335–353, <https://doi.org/10.1007/Bf01532020> (1989).
25. Dixon, S. L. *et al.* PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aid Mol Des* **20**, 647–671, <https://doi.org/10.1007/s10822-006-9087-6> (2006).
26. Dixon, S. L., Smondryev, A. M. & Rao, S. N. PHASE: A novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des* **67**, 370–372, <https://doi.org/10.1111/j.1747-0285.2006.00384.x> (2006).
27. Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J Chem Inf Model* **49**, 377–389, <https://doi.org/10.1021/ci800324m> (2009).
28. Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* **69**, 146–148, <https://doi.org/10.1111/j.1747-0285.2007.00483.x> (2007).
29. Watts, K. S. *et al.* ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *J Chem Inf Model* **50**, 534–546, <https://doi.org/10.1021/ci100015j> (2010).
30. Mooradian, A. D., Morley, J. E. & Korenman, S. G. Biological Actions of Androgens. *Endocr Rev* **8**, 1–28, <https://doi.org/10.1210/edrv-8-1-1> (1987).
31. Heinlein, C. A. & Chang, C. S. The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions. *Mol Endocrinol* **16**, 2181–2187, <https://doi.org/10.1210/me.2002-0070> (2002).
32. De Vivo, I. *et al.* A functional polymorphism in the promoter of the progesterone receptor gene associated with endometrial cancer risk. *P Natl Acad Sci USA* **99**, 12263–12268, <https://doi.org/10.1073/pnas.192172299> (2002).
33. Gadkar-Sable, S., Shah, C., Rosario, G., Sachdeva, G. & Puri, C. Progesterone receptors: Various forms and functions in reproductive tissues. *Front Biosci* **10**, 2118–2130, <https://doi.org/10.2741/1685> (2005).
34. Knutson, T. P. & Lange, C. A. Tracking progesterone receptor-mediated actions in breast cancer. *Pharmacol Therapeut* **142**, 114–125, <https://doi.org/10.1016/j.pharmthera.2013.11.010> (2014).
35. Germain, P. *et al.* International Union of Pharmacology. LXIII. Retinoid X receptors. *Pharmacol Rev* **58**, 760–772, <https://doi.org/10.1124/pr.58.4.7> (2006).
36. Allenby, G. *et al.* Retinoic Acid Receptors and Retinoid X-Receptors - Interactions with Endogenous Retinoic Acids. *P Natl Acad Sci USA* **90**, 30–34, <https://doi.org/10.1073/pnas.90.1.30> (1993).
37. Rhen, T. & Cidlowski, J. A. Antiinflammatory action of glucocorticoids - New mechanisms for old drugs. *New Engl J Med* **353**, 1711–1723, <https://doi.org/10.1056/NEJMr050541> (2005).
38. Kumar, R. & Thompson, E. B. Gene regulation by the glucocorticoid receptor: Structure: function relationship. *J Steroid Biochem* **94**, 383–394, <https://doi.org/10.1016/j.jsmb.2004.12.046> (2005).
39. Kumar, R. & Thompson, E. B. The structure of the nuclear hormone receptors. *Steroids* **64**, 310–319, [https://doi.org/10.1016/S0039-128x\(99\)00014-8](https://doi.org/10.1016/S0039-128x(99)00014-8) (1999).
40. Willson, T. M., Lambert, M. H. & Kliewer, S. A. Peroxisome proliferator-activated receptor gamma and metabolic disease. *Annu Rev Biochem* **70**, 341–367, <https://doi.org/10.1146/annurev.biochem.70.1.341> (2001).
41. Li, A. C. *et al.* Peroxisome proliferator-activated receptor  $\gamma$  ligands inhibit development of atherosclerosis in LDL receptor-deficient mice. *Journal of Clinical Investigation* **106**, 523 (2000).
42. Savage, D. B. *et al.* Resistin/Fizz3 expression in relation to obesity and peroxisome proliferator-activated receptor-gamma action in humans. *Diabetes* **50**, 2199–2202, <https://doi.org/10.2337/diabetes.50.10.2199> (2001).
43. Jagtap, P. & Szabo, C. Poly(ADP-ribose) polymerase and the therapeutic effects of its inhibitors. *Nat Rev Drug Discov* **4**, 421–440, <https://doi.org/10.1038/nrd1718> (2005).
44. Soreq, H. & Seidman, S. Acetylcholinesterase - new roles for an old actor. *Nat Rev Neurosci* **2**, 294–302, <https://doi.org/10.1038/35067589> (2001).
45. Nguyen, C. N., Young, T. K. & Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit [7] uril. *J. Chem. Phys.* **137**, <https://doi.org/10.1063/1.4733951> (2012).
46. Beuming, T. *et al.* Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization. *Proteins-Structure Function and Bioinformatics* **80**, 871–883, <https://doi.org/10.1002/prot.23244> (2012).
47. Li, Z. & Lazaridis, T. Computing the thermodynamic contributions of interfacial water. *Computational Drug Discovery and Design*, 393–404 (2012).

## Acknowledgements

This work was supported by NIH grant GM095417 (T.K. and S.R.) and NRF grants 2016M3A7B4025405 and 2017R1D1A1B03035870 (A.E.C., S.W.J. and M.K.). We thank Schrödinger LLC (Portland, OR, USA) for the generous supply of the software used for this research.

## Author Contributions

S.W.J., M.K. and S.R. performed simulations and analyzed the data from simulations. T.K. and A.E.C. designed the project. S.W.J., T.K. and A.E.C. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-28546-z>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018