

# Structure-Based Identification of Binding Sites, Native Ligands and Potential Inhibitors for G-Protein Coupled Receptors

Claudio N. Cavasotto,<sup>††</sup> Andrew J.W. Orry,<sup>#</sup> and Ruben A. Abagyan<sup>\*</sup>

*Scripps Research Institute, Department of Molecular Biology (TPC-28), La Jolla, California*

**ABSTRACT** G-protein coupled receptors (GPCRs) are the largest family of cell-surface receptors involved in signal transmission. Drugs associated with GPCRs represent more than one fourth of the 100 top-selling drugs and are the targets of more than half of the current therapeutic agents on the market. Our methodology based on the internal coordinate mechanics (ICM) program can accurately identify the ligand-binding pocket in the currently available crystal structures of seven transmembrane (7TM) proteins [bacteriorhodopsin (BR) and bovine rhodopsin (bRho)]. The binding geometry of the ligand can be accurately predicted by ICM flexible docking with and without the loop regions, a useful finding for GPCR docking because the transmembrane regions are easier to model. We also demonstrate that the native ligand can be identified by flexible docking and scoring in 1.5% and 0.2% (for bRho and BR, respectively) of the best scoring compounds from two different types of compound database. The same procedure can be applied to the database of available chemicals to identify specific GPCR binders. Finally, we demonstrate that even if the sidechain positions in the bRho binding pocket are entirely wrong, their correct conformation can be fully restored with high accuracy (0.28 Å) through the ICM global optimization with and without the ligand present. These binding site adjustments are critical for flexible docking of new ligands to known structures or for docking to GPCR homology models. The ICM docking method has the potential to be used to “de-orphanize” orphan GPCRs (oGPCRs) and to identify antagonists–agonists for GPCRs if an accurate model (experimentally and computationally validated) of the structure has been constructed or when future crystal structures are determined. *Proteins* 2003;51:423–433.

© 2003 Wiley-Liss, Inc.

**Key words:** seven transmembrane receptor; protein–ligand docking; rational drug design; virtual ligand screening; internal coordinate mechanics; ligand-binding pocket prediction; orphan G-protein coupled receptor; flexible ligand receptor docking

## INTRODUCTION

The G-protein coupled receptor (GPCR) superfamily constitutes the largest family of receptors in the cell and is responsible for mediating the effects of over 50% of drugs on the market today.<sup>1</sup> There are ~300 full open reading frames for GPCRs in the public genome databases (excluding ~1200 members in rodents that are taste, vomeronasal, and odorant receptors). Of these 300 GPCRs, ~100 (maybe more, after further analysis of the human genome) are orphan receptors (oGPCRs), which means they are lacking a relevant ligand that will activate the receptor.<sup>2</sup> The challenge is to identify ligands (“de-orphanize”) for these receptors and to find additional ligands (i.e., agonists–antagonists) for known GPCRs (with identified ligands), and hence understand their function and evaluate their pharmacologic potential.

GPCRs are integral membrane proteins characterized by the presence of seven  $\alpha$ -helical hydrophobic domains that range in length from 20 to 25 residues and traverse the cell membrane. GPCRs have an intracellular C-terminal, an extracellular N-terminal, three intracellular loops, and three extracellular loops. Members of the GPCR superfamily are responsible for the transmission of a variety of signals to the interior of the cell and can be activated by a diverse range of small molecules.<sup>3</sup> Activation of the GPCR results in a conformational change followed by a signal cascade that passes information to the inside of the cell via interaction of the receptor with heterotrimeric G-proteins. There are three main classes of GPCRs (A, B, and C) depending on their sequence similarity to rhodopsin (Rho) (class A).<sup>4</sup> Class A GPCRs are the largest group and encompass a wide range of receptors, including receptors for odorants, which bind small ligands; adenosine;  $\beta$ -adrenergic; and Rho. The ligand-binding site is localized within the seven transmembrane (7TM) region.<sup>5</sup>

Grant sponsor: Department of Energy; Grant number: DOE/DE-FG03-00ER6304.

<sup>††</sup>These authors contributed equally to this work.

<sup>†</sup>Present address: Molsoft LLC, 3366 N. Torrey Pines Court, Suite 300, La Jolla, CA 92037. E-mail: claudio@molsoft.com

<sup>\*</sup>Correspondence to: Ruben A. Abagyan, Scripps Research Institute, Department of Molecular Biology (TPC-28), 10550 N. Torrey Pines Road, La Jolla, California 92037. E-mail: abagyan@scripps.edu

Received 5 September 2002; Accepted 24 November 2002

GPCRs are important drug targets; therefore, many attempts have been made to solve their structure by molecular modeling.<sup>6–11</sup> However, to date, bovine rhodopsin (bRho) is the only GPCR to be solved to a high resolution by X-ray crystallography.<sup>12</sup> Electron diffraction and electron microscopy (EM) studies have succeeded in determining the structure of a GPCR-like 7TM protein, bacteriorhodopsin (BR), at medium resolution<sup>13,14</sup> and one GPCR, bRho, at high resolution.<sup>15–17</sup> The X-ray crystallographic structure of BR has also been determined at high resolution from microcrystals grown in a cubic lipidic phase.<sup>18,19</sup> Once three-dimensional structural details are known, they can be used to understand ligand binding and to rationally design novel ligands as prospective drug compounds.

BR and Rho both have a retinal chromophore that is isomerized at the C13=C14 double bond on absorption of a photon. The BR receptor has the retinal chromophore linked via a protonated Schiff base to residue Lys216,<sup>20</sup> and in Rho, retinal is covalently bonded to Lys296.<sup>12</sup> The binding pocket for retinal is in between helices III, IV, V, and VI.

A diverse range of procedures has been undertaken to identify ligands for orphan receptors. These procedures can be summarized into three main approaches: (1) testing ligands for identified GPCRs on oGPCRs with high sequence identity<sup>21–23</sup> and in some cases randomly evaluating oGPCRs against arrayed families of known ligands; (2) identifying relationship between receptor and ligand expression patterns<sup>24,25</sup>; and (3) testing tissue extracts in functional receptor-based assays. Many of these techniques have been successful<sup>21–23,25,26</sup>; however, none of these methods use a rational structure-based approach for identification of ligands.

Other strategies to attempt to understand GPCR ligand binding have included molecular modeling and computational ligand docking in conjunction with mutagenesis experiments (e.g., cannabinoid,<sup>27</sup> serotonin,<sup>28</sup> adrenergic,<sup>29</sup> and adenosine<sup>30</sup>). This approach is a useful tool to determine which residues may be involved in agonist or antagonist binding and enables the corresponding residues to be mutated experimentally; however, the docking approach is generally undertaken manually and without any suitable validation methods. Early GPCR models based on BR are flawed because the arrangement of helices differs in GPCRs.<sup>31–34</sup> In most cases, backbone rearrangement was not included in the homology model or was insufficient.<sup>27–30</sup> In view of the low sequence identity between GPCRs and Rho, this should be considered a serious limitation.

We had three goals in this study relating to 7TM receptors: (1) to test whether a ligand-binding pocket can be automatically identified from the structure without any information about the ligand; (2) to test whether the association of a flexible ligand can be correctly and automatically predicted once the pocket is identified; and (3) to test whether the ligand can be correctly identified from thousands of biologic substrates and then identify potential drug candidates.

As described earlier, a standard experimental method for identifying ligands for new GPCRs is to rigorously test in vitro many hundreds of potential binding compounds. This is extremely time-consuming and as a result of the sheer scale of the task means that many important potential ligands are missed and never tested. Testing is usually limited to compounds with similar structural characteristics; therefore, new potentially structurally diverse ligands will never be identified. The internal coordinate mechanics (ICM)<sup>35</sup> method incorporates a virtual ligand screening (VLS) algorithm that can screen millions of potential ligands by flexible docking them into a crystal structure or accurate model (experimentally and computationally validated) of a GPCR.

In the past, we have successfully used an alliance between state-of-the-art computational techniques and experimental analysis to identify binders for other receptors. The application of this approach has identified antagonists and agonists for the retinoic acid receptor,<sup>36,37</sup> which is implicated in a wide range of diseases, including breast cancer, and we have also discovered agonists and antagonists for the thyroid receptor.<sup>38</sup>

We demonstrate in this article that our docking methodology can accurately predict the binding mode of retinal into the crystal structure of bRho and the GPCR-like 7TM receptor BR. These are currently the only 7TM receptors to have their crystal structures solved to a high resolution. Successful redocking of retinal into bRho can also be achieved when the N-terminus, C-terminus, intracellular loops, and extracellular loops are removed. We also report that the native ligand can be automatically identified via flexible docking and scoring from a database of naturally occurring compounds. Finally, we demonstrate that the conformations of binding-pocket sidechain residues can be accurately predicted starting from a random position, with and without the ligand present. This method can be applied to other GPCRs to predict ligand position and binding-pocket conformations of GPCR homology models or alternative binding modes for different ligands.

The reported procedures enable us to use a receptor-structure-based approach to functional characterization of GPCRs and rational drug design against the GPCR molecular targets.

## MATERIALS AND METHODS

The crystal structures of bRho [Protein Data Bank (PDB) code 1F88: Chain A, 2.88 Å resolution<sup>12</sup>] and BR from *Halobacterium salinarum* (PDB code 1C3W: Chain A, 1.55 Å resolution<sup>20</sup>) have been used in all experiments. The PDB files were prepared by adding hydrogens and missing heavy atoms, and assigning partial charges followed by an energy minimization step, as implemented in ICM version 2.8.<sup>35</sup>

Because the ligands in both crystal structures are covalently bound to the receptor, the chemical groups starting at the  $\epsilon$  carbon atoms of Lys 216 (BR) and Lys 296 (bRho) were removed and replaced by a hydrogen atom to disassociate the covalent bond and allow free, unbiased docking. Two hydrogens were added to the retinal mol-

ecules of the crystal structures at the terminal carbon C15, which makes the covalent bond with lysines in BR and bRho. In this way, the covalent bond is broken but the C15 group is close enough to make a van der Waals contact with the modeled lysines. We refer to this modeled ligand as retinal throughout the text.

### Identification of the Ligand-Binding Pocket

The ICMPocketFinder function was used to identify the retinal binding pocket in the crystal structures of bRho and BR. The algorithm builds a grid map of a binding potential, and the position and size of the ligand-binding pocket are determined based on the construction of equipotential surfaces along those maps.<sup>39</sup>

### Receptor-Ligand Grid Docking and Virtual Ligand Screening

We applied the flexible ligand/grid receptor docking methodology as implemented in ICM using an extension of the Empirical Conformational Energy Program for Peptides 3 (ECEPP/3)<sup>40</sup> force field parameters and Merck Molecular Force Field (MMFF) partial charges.<sup>41</sup> Five potential maps (electrostatic, hydrogen bond, hydrophobic, and two for van der Waals) were calculated, followed by a global optimization of the flexible ligand in the receptor field,<sup>42</sup> so that both the intramolecular ligand energy and the ligand-receptor interaction energy were optimized during the calculation.

The VLS experiments were undertaken with two medium databases containing different types of compounds. We used a database consisting of 7,774 [molecular weight (MW)  $\leq$  1000] varied compounds (metabolic and other compounds, including substrates, products, inhibitors of metabolic pathways, as well as drugs and xenobiotic chemicals) of the LIGAND Chemical Database (Institute for Chemical Research, Kyoto University, Japan, June 22, 2002 edition)<sup>43,44</sup> and also the drug database "CNS" (ChemBridge Corp, San Diego CA) that consists of 11,818 drug compounds (with MW  $\leq$  500).

For each ligand, the best solution is scored by an empirical scoring function based on its fit into the binding pocket of the receptor.<sup>42,45</sup> This score takes into account continuum and discrete electrostatics, and hydrophobic and entropy loss (see review<sup>42</sup>). This methodology allows fast and accurate screening of hundreds of thousands of compounds (the average computing time is approximately 1 min/compound/processor). Each ligand from the database was docked three times and a minimum of the three scores was used. The VLS experiments were undertaken on a state-of-the-art Intel Linux cluster, each 1.33 GHz processor having access to 1 Gb of memory.

### Global Energy Optimization

The global optimization procedure as implemented in ICM<sup>46</sup> was used (1) to undertake an unbiased, all-atom, flexible docking of ligand within the flexible binding pocket, and (2) to predict the sidechain conformation of the binding pocket without the ligand of the bRho crystal structure present. It consists of the following steps: (1) a random conformation change of the free variables

[sidechain torsion angles according to the biased probability Monte Carlo (BPMC) algorithm],<sup>47</sup> torsion, and rotational angles of the ligand; (2) local energy minimization of analytic differentiable terms; (3) calculation of the complete energy, including nondifferentiable terms such as entropy and solvation energy; (4) acceptance or rejection of the total energy based on the Metropolis criterion<sup>48</sup> and return to step 1.

The conformational sampling is based on the BPMC procedure,<sup>47,49</sup> which randomly selects a conformation in the internal coordinate space<sup>46</sup> and then makes a step to a new, random position, independent of the previous one, but according to a predefined continuous probability distribution. It has also been shown that after each random step, full local minimization greatly improves the efficiency of the procedure.<sup>50,51</sup> However, because some energy terms might have no derivatives or might be very expensive to compute, a double-energy Monte Carlo minimization scheme<sup>46</sup> circumvents these obstacles by minimizing the energy with respect to the differentiable terms but calculates the full energy using also the nondifferentiable terms. This double-energy scheme allows for the incorporation of complex energy terms, such as surface-based solvation energy, into the global optimization process.

## RESULTS

### Identification of Ligand-Binding Pocket

The ICMPocketFinder function successfully identified the retinal binding pocket in the crystal structure of bRho and BR [Fig. 1(a and c)]. The BR and Rho pockets have a volume of 392 Å<sup>3</sup> and 481 Å<sup>3</sup>, respectively, which is sufficient to incorporate the retinal ligand. The binding pocket predicted by ICM also encompasses the amino acid required for the covalent bond between the ligand and the receptor (Lys 296 bRho and Lys 216 BR).

### Predicting the Retinal-Ligand Association Mode

We docked the all-trans-retinal and the cis-retinal ligand into the crystal structures of BR and bRho, respectively, using the ICM flexible ligand-grid docking algorithm. The  $\epsilon$  carbon atoms of Lys 216 (BR) and Lys 296 (bRho) were removed and replaced by a hydrogen atom to disassociate the covalent bond and allow free, unbiased docking (see Materials and Methods section).

The predicted structures are shown in Figure 1(b and d). The root-mean-square deviation (RMSD) (for heavy atoms) between the docked ligand and the crystal structure ligand was very low (BR RMSD = 0.14 Å; bRho RMSD = 0.15 Å). The redocked ligand was also in the same orientation (with C15 of the retinal ligand orientated toward the Lys residue forming the covalent bond) as the retinal ligand in the crystal structure. This also constitutes a validation of our modeling of the covalent bond between ligand and receptor (see Materials and Methods section).

A further ligand-docking experiment was undertaken with the bRho and BR structures to determine whether the retinal ligand can be correctly docked into the crystal structure when the loops have been removed. The N-terminus (residues: M1:E33), C-terminus (residues: N310:

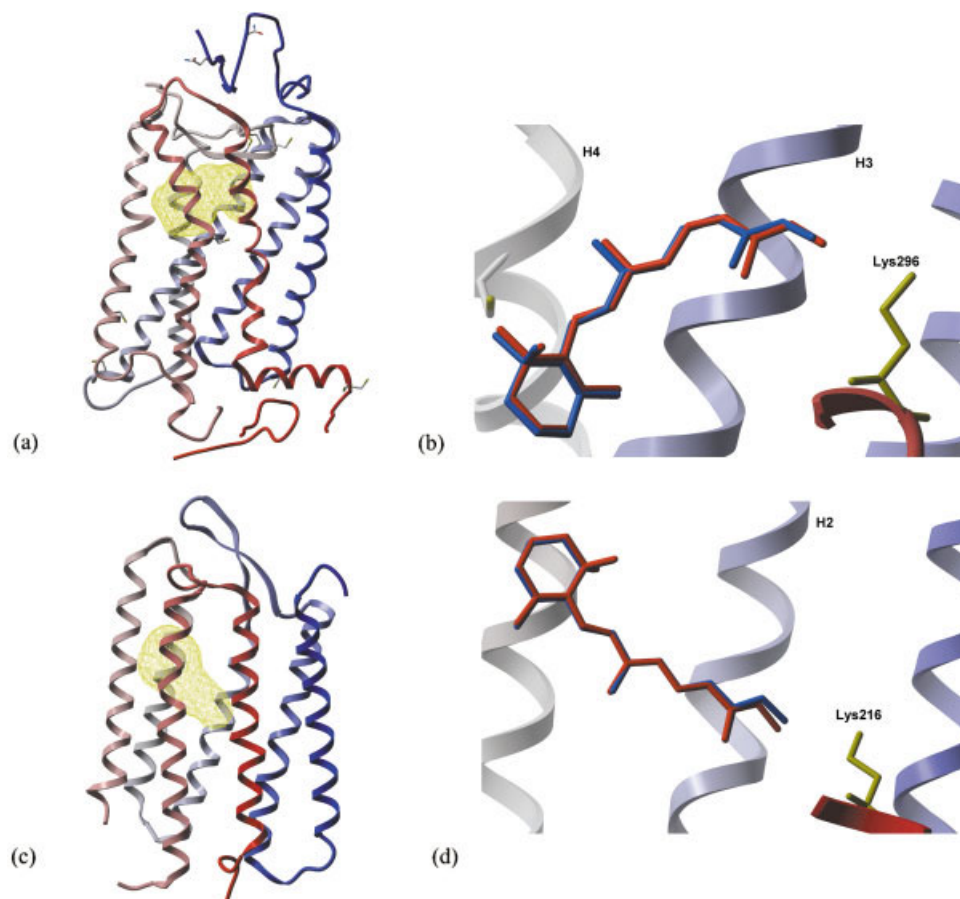


Fig. 1. The correctly identified retinal binding pockets in (a) bRho and (c) BR are shown in yellow. The bound crystal structure retinal (blue) and the redocked retinal (red) are shown for (b) bRho and (d) BR. Pictures were created by ICM, helices (H) VI and VII have been cut away for clarity in (b) and (d), and the extracellular N-terminus is shown in (a) and (d). [Pictures generated by ICM colored from blue (N-terminus) to red (C-terminus).]

A348), three extracellular loops (residues: G101-F105; G174:N199; H278:F284), and three intracellular loops (residues: H65:T70; C140:G149; L226:A246) were deleted from the bRho structure. The N-terminus (residues: T5: E9), C-terminus (residues: S226:G231), three extracellular loops (residues: Y64-Y79; T128:V130; S193:V199), and three intracellular loops (residues: G31:S35; D102:A103; F156:M163) were deleted from the BR structure.

Results achieved under these conditions were similar to those obtained when redocking the retinal ligand into bRho and BR with the extracellular and intracellular regions present.

#### Ab Initio Identification of the Natural GPCR Substrate and Virtual Compound Screening

We screened the LIGAND Chemical Database and the drug database “CNS” (see Materials and Methods section) against bRho and BR.

Figure 2(a and d) show plots of the docking score for the compounds in the LIGAND database screened against both receptors. Each compound is characterized by its molecular weight. Figure 2(b and e) is an enlarge-

ment of the same graph in the region of the best scores with the all-trans retinal and compound highlighted. The retinal docking score was  $-31$  and  $-52$  for bRho and BR, respectively. Other compounds that are closely related to retinal, such as *cis*- and *trans*-retinol, also bound to bRho and BR with a “good” score, but a lower score than the all-trans retinal. A low value represents a good score indicating energetically favorable binding. The low frequency of binders with a good score highlights the ability of ICM VLS to screen out and narrow down the number of ligands needed to test in the laboratory with binding assays [Fig. 2(c and f)]. Similar good docking scores were achieved when screening the “CNS” ChemBridge drug database against the crystal structures of BR and bRho (Fig. 3).

The ICM VLS data present a database enrichment factor of  $\sim 98.5\%$  for bRho and  $\sim 99.8\%$  for the BR receptor when screening against the LIGAND database and the Cambridge CNS drug database. These enrichment factors signify that  $\sim 98.5\%$  of a ligand database can be discarded to identify the native correctly bound ligand for the currently available crystal structures of 7TM receptors.

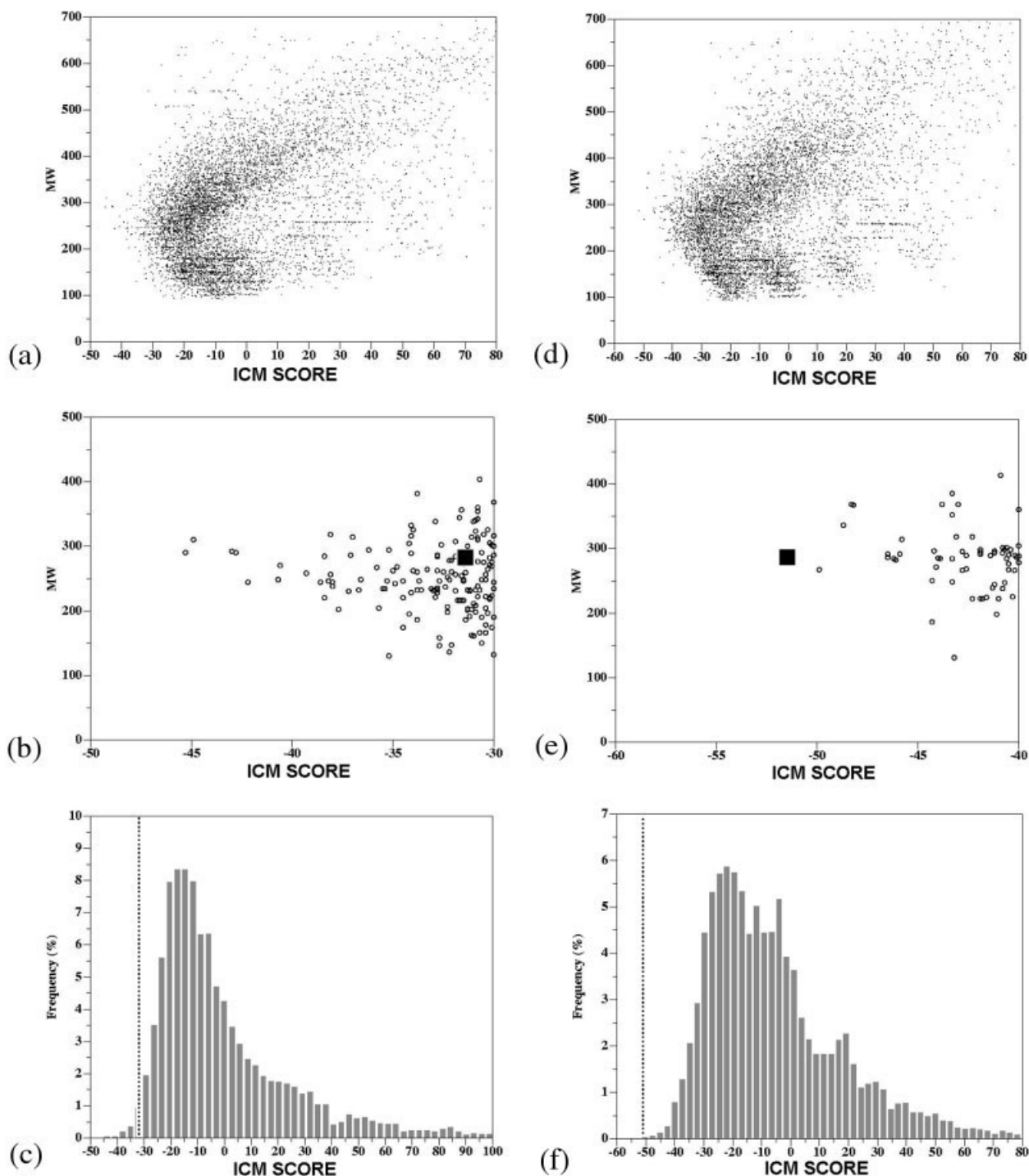


Fig. 2. A VLS plot of docking score for the ligands in the LIGAND database docked into the (a) bRho receptor and (d) BR. Retinal, the known binding ligand to (b) bRho and (e) BR, is highlighted with a square box. Ligands are characterized by their molecular weight (MW). A histogram of frequency percentage of docking score is shown for (c) bRho and (f) BR; the dotted vertical line represents the native ligand-binding threshold limit (i.e., ligands to the left of this line represent the number needed to be tested experimentally to identify the native ligand from the database of compounds).

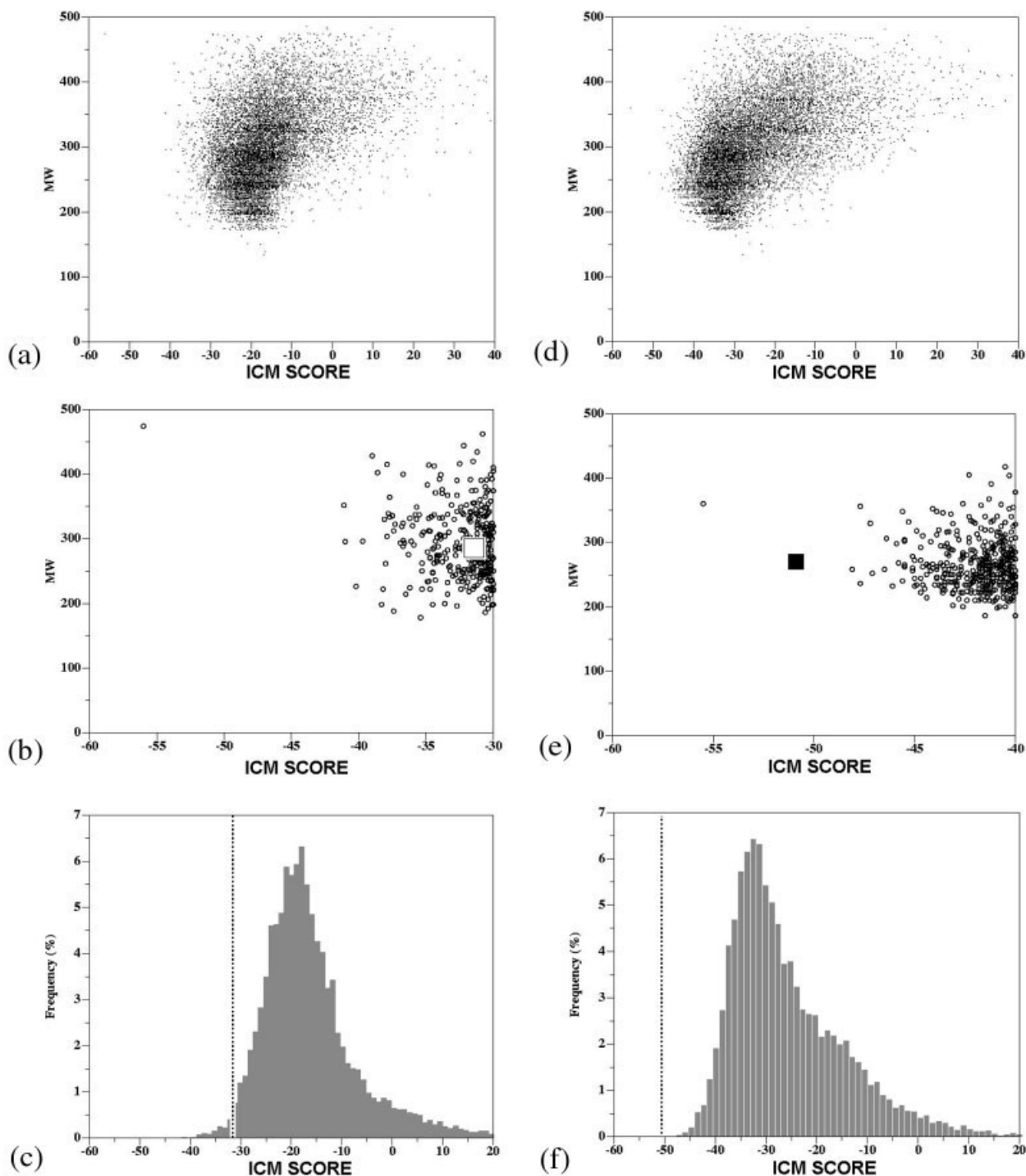


Fig. 3. A VLS plot of docking score for the ligands in the “CNS” ChemBridge drug compound database docked into the (a) bRho receptor and (d) BR. Retinal, the known binding ligand to (b) bRho and (e) BR, is highlighted with a square box in (e) and a white box in (b). Ligands are characterized by their molecular weight (MW). A histogram of frequency percentage of docking score is shown for (c) bRho and (f) BR; the dotted vertical line represents the native ligand-binding threshold limit (i.e., ligands to the left of this line represent the number needed to be tested experimentally to identify the native ligand from the database of compounds).

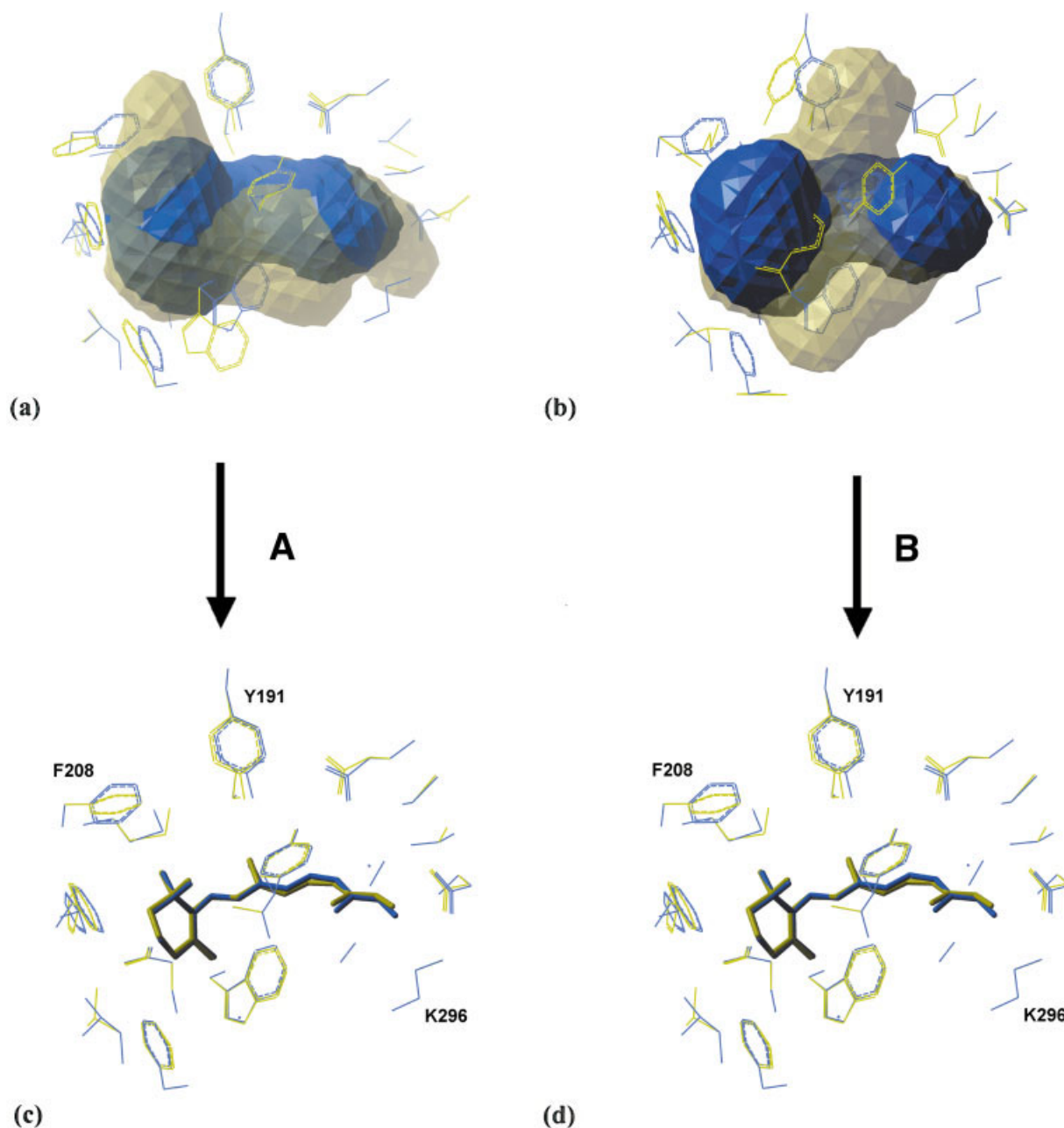


Fig. 4. The shape of the binding pockets (yellow) and the random positions of the sidechains (yellow) for the two different starting conformations (A) and (B) before unbiased global energy optimization are shown in (a) and (b). The position of the sidechains and ligand for the two conformations (A) and (B) after global energy optimization are shown in (c) and (d). The crystal structure-binding pocket, ligand, and sidechains are colored blue. (Pictures generated with the use of ICM.)

### Correct Prediction of Binding-Pocket Geometry

Here, we tested a procedure in which a fully flexible ligand is docked to a flexible pocket. This global energy optimization procedure, with as many as 30–50 essential variables, is therefore very computationally challenging. This sort of procedure is key for predicting the binding mode and sidechain rearrangement for new ligands to the same GPCR, or any ligands to a homology model, provided the backbone changes are small.

The sidechain torsion angles of 21 residues surrounding the binding pocket of bRho (E113, G114, A117, T118, G121, E122, L125, E181, S186, C187, I189, Y191, M207, F208, H211, F212, F261, W265, Y268, A292, K296) were randomized, and two different receptor binding-pocket conformations were generated (A and B) (see Fig. 4 for a comparison of the shape of each of these two pockets compared to the bRho crystal structure pocket). This procedure mimics potential new GPCRs ligand-binding



pockets as a result of the very low sequence identity with respect to the template within this pocket. The three rotational angular variables of the retinal ligand with respect to its center of mass have also been randomized to generate a set of different initial complexes for each of the two new pockets. Each complex underwent a short energy minimization of 5000 iterations prior to any global energy optimization procedure.

We then undertook a global energy optimization on each complex, using the BPMC and the double-energy scheme (see Materials and Methods section) algorithm. The global energy optimization simulation was performed at 600 K with 635,000 steps, which was sufficient to ensure convergence and took approximately 6 h to complete. In the energy evaluation, we considered the following terms: van der Waals (with a separate term for 1–4 interactions), hydrogen bonding, torsion energy, and electrostatic energy with distance-dependent dielectric constant. A separate surface term, which took into account hydrophobic effects, and an entropic term were also included. Because it is usually known which helices define the ligand-binding pockets in GPCRs, a high-energy penalization was included outside a 0.8 Å margin of the box, which contained every atom of the residues of the binding pocket defined above. For each of the two randomly generated starting conformations, we used the final lowest energy structure after the global energy optimization for further RMSD calculations compared to the bRho crystal structure.

We found that our flexible docking procedure was capable of predicting the correct conformation with remarkable accuracy (Fig. 4). The RMSD for ligand heavy atoms in both A and B pockets was 0.28 Å, and the RMSD for the binding-pocket sidechains for A and B was 0.28 Å in both cases.

In a recent article, Lin et al.<sup>52</sup> described a methodology to address receptor flexibility in docking problems. It consists of an initial phase of an extended molecular dynamics (MD) simulation of the unliganded receptor and a second phase in which a small library of candidate binders is docked to the ensemble of the MD-generated structures with the use of a flexible ligand/fixed receptor docking algorithm. This technique was applied successfully to the FK506 binding protein. Our methodology, on the other hand, uses a one-step global energy optimization procedure in which the backbone is fixed but the sidechains are sampled with or without the ligand present.

To further validate our approach, we tried to predict the sidechain conformation of the randomized ligand binding protein (LBP) of bRho using our global energy optimization methodology, but without the ligand inside. We used the same parameters and starting randomized conformations (A and B) as in the previous case. The RMSD for the sidechain residues in both A and B LBPs was also 0.28 Å, thus making the difference in sidechain conformation calculation with and without the ligand negligible. This is especially useful for the “de-orphanization” of oGPCRs.

## DISCUSSION

The GPCR family of membrane proteins is an important target for drug design, because these proteins control a wide range of cellular signals. Therefore, the ability to automatically and accurately dock “drug-like” compounds into a structure of a GPCR receptor will allow a structure-based identification of new candidates for therapeutics. Moreover, this potentially could facilitate the “de-orphanization” of a number of GPCRs by identifying potential natural ligand binders. Here, we present a validation of an ICM method for automated and accurate docking of ligands into GPCRs and the screening of drug-like compound databases.

The ICM procedure accurately identified the ligand binding pocket in both BR and bRho [Fig. 1(a,c)], which is crucial for the structure-based “de-orphanization” of GPCRs and also the identification of other ligand binders to known GPCRs (e.g., antagonists and agonists). These predictions can also be used in conjunction with laboratory-based mutagenesis experiments to identify potential residues to be mutated so as to fully characterize the pocket.

The validation reported here has been successful on the two crystal structures of 7TM proteins that are currently available. The extremely low RMSD values, the correct orientation of the predicted binding mode of the retinal ligand [Fig. 1(b,d)], and the very high database-enrichment factor (Figs. 2 and 3) for ligand and drug databases indicate that this approach can identify potential small-molecule binders that may bind in the transmembrane region of GPCRs and other 7TM proteins.

The low frequency of ligand and drug binders with a high ICM docking score [cutoff shown with black-dashed vertical line in Figs. 2(c,f) and 3(c,f)] highlights the ability of ICM (VLS) to filter and narrow down the number of ligands needed to be tested in the laboratory with binding assays. In the case of BR, the all-trans retinal was the highest scoring ligand when screened against the LIGAND database and second highest against the “CNS” ChemBridge drug database. This, however, was not the case with bRho, but retinal was still ranked in the top 82 highest scoring binders when screening against the LIGAND database and the top 165 ligands against the drug database. This means that, in theory, approximately 1.5% of compounds from a ligand or drug database need to be tested in experimental ligand-binding assays to identify the correctly bound native ligand. Therefore, 98.5% of the database can be screened out by ICM and considered not to be potential binders.

Many of the high-scoring ligands from the VLS data appear interesting, because some are involved in diseases of the eye and further experimental ligand-binding experiments need to be undertaken to test those predictions. The VLS algorithm is capable of screening millions of compounds for potential binding to GPCR structures and narrows down a subset of ligands for further analysis in the laboratory. This is a more rational approach for the identification of potential GPCR binders and permits a broad range of ligands to be tested that might be otherwise



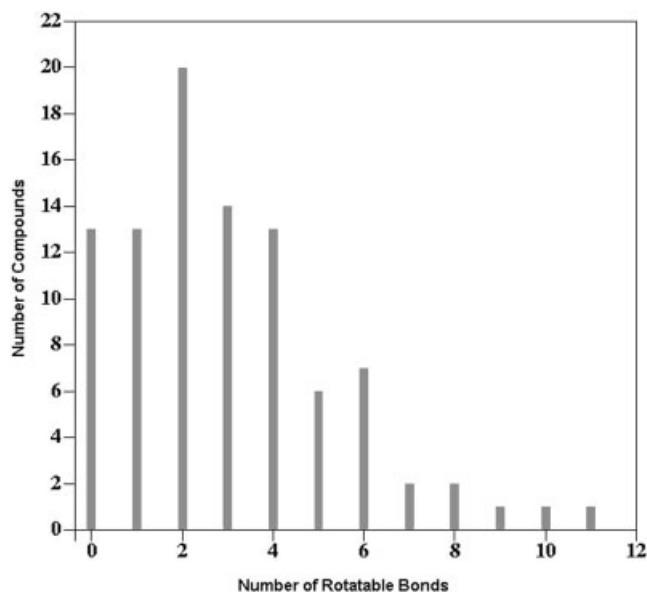


Fig. 5. A comparison of the number of rotatable bonds for the 82 top-scoring ligands after screening the LIGAND database against the bRho receptor; the retinal ligand has six rotatable bonds, which is a higher number and therefore higher entropy compared to the rest of the high-scoring ligands.

overlooked when attempting to identify ligands in the conventional experimental way.

The BR and bRho docking problem shown here is a more complicated scenario than would be expected with other GPCRs that do not rely on isomerization for activation; hence, the excellent validation data are even more remarkable. The docking and VLS are complex for two reasons: First, the ligand is bound by a covalent bond, and second, the ligand has a comparatively high number of rotatable bonds (Fig. 5) and therefore high entropy.

A covalent bond is present between retinal and the lysine residue on helix VII in the receptor for both bRho and BR. It is a challenge to dock ligands that form covalent bonds because of the problems associated with mimicking this covalent bond in the ligand receptor complex. However, most GPCRs do not form covalent bonds with their ligand when bound in the receptor, so we are confident that the docking of ligands and “drug-like” compounds into GPCRs can be achieved with even more accuracy by ICM. The VLS results also showed good ICM docking scores for other retinal isoforms that bind to bRho and BR, such as *cis*- and *trans*-retinal.

Retinal has six rotatable bonds, which is a high number compared with rest of the good scoring ligands in the database. Ligands with a high number of rotatable bonds, such as retinal, lose more entropy upon binding than more rigid ligands. Therefore, a good ICM docking score involving such a large loss of entropy would not be expected, because entropy loss is taken into account in the ICM scoring function. In the particular cases of bRho and BR, the high entropy energy associated with retinal is compensated by the fact that it undergoes chemical modification while bound to the receptor.

Because of the problems involved with the crystallization of membrane proteins, many modeling and EM structural studies<sup>6–11</sup> have been undertaken on GPCRs. Molecular modeling of GPCRs has concentrated on determining the conserved secondary structure of the transmembrane region rather than attempting to model the more variable and intricate loops, N-terminus, and C-terminus regions. This is a sensible approach, especially when modeling class A GPCRs, because the ligand-binding pocket is located in the transmembrane region and, generally, the majority of the mutational data on which the model is built is in this location. Due to the low resolution of EM projection maps, only the transmembrane regions of the receptor can be determined. We have demonstrated that we can accurately dock (low RMSD and correct orientation compared to native ligand) the ligand into the binding pocket even when the N-terminus, C-terminus, and extracellular and intracellular loops are removed. This is achieved with bRho even though the retinal ligand in the crystal structure is in contact with extracellular loop II (ser186, cys187, and ile189); however, the majority of the LBP is deeply buried and well defined, so this result is not too surprising. Even so, this capability can be used for the identification of ligand binders from accurate models and EM projection maps of GPCRs.

Here, we also demonstrate the capability to predict accurately the binding mode of retinal in the binding pocket of bRho starting from a random conformation of sidechains and ligand position using an unbiased flexible, global energy optimization approach (Fig. 4). This approach will allow accurate models of other GPCRs to be constructed and increase the accuracy for the prediction of a docked ligand and conformation of the binding pocket. There is very low sequence identity in the binding pocket of GPCRs compared with bRho, so use of this as a template for modeling other GPCRs is a controversial approach. However, bRho is the only crystal structure currently available, and it provides important information about the position of the helices in the transmembrane section. So by randomizing the sidechains of bRho and the position of the ligand, we have mimicked the situation faced when trying to model the binding pocket of other GPCRs, with in some cases, zero sequence identity. We also undertook the global energy optimization of the randomized LBP sidechains without the ligand present, achieving the same accuracy and conformation as with the previous experiment with the ligand present. This mimics the situation that frequently occurs when trying to model by homology and “de-orphanize” an oGPCR. In the future, when new crystal structures of GPCRs are solved, global energy optimization could be used to identify alternative binding modes for other ligands that may bind in the transmembrane pocket. This validation and optimization of the ligand-binding pocket will enhance the accuracy of a GPCR model, aid mutagenesis experiments, and identify prospective leads for drug discovery.

Recently, in our laboratory, we have implemented this VLS technique in conjunction with experimental binding

assays to identify successfully antagonists to nuclear receptors and the use of thyroid receptor.<sup>36,37</sup> This success and the encouraging validation data reported here lead us to believe that VLS will be a valuable tool for identifying ligands that bind to GPCRs.

The database-enrichment factors achieved in this VLS study and the capability of ICM to predict the binding mode of ligands (with and without the loop regions) into 7TM proteins demonstrates that flexible docking and a grid representation of the receptor molecule can create a model of protein–ligand interaction with an accuracy that can be used for drug-design.

As more accurate models of GPCRs are generated, with the possibility of more crystal structures of GPCRs being determined in the future, the receptor-based flexible docking and VLS technology will be used to identify natural ligands and potential therapeutics for oGPCRs.

### ACKNOWLEDGMENTS

We would like to thank Maxim Totrov for many useful discussions.

### REFERENCES

- Christopoulos A. Allosteric binding sites on cell-surface receptors: Novel targets for drug discovery. *Nat Rev Drug Disc* 2001;1:198–210.
- Howard AD, McAllister G, Feighner SD, Liu QY, Nargund RP, Van der Ploeg LHT, Patchett AA. Orphan G-protein-coupled receptors and natural ligand discovery. *TIPS* 2001;22:132–140.
- Bockaert J, Pin JP. Molecular tinkering of G protein-coupled receptors: An evolutionary success. *EMBO J* 1999;18:1723–1729.
- Probst WC, Snyder LA, Schuster DI, Brosius J, Sealfon SC. Sequence alignment of the G-protein coupled receptor superfamily. *DNA Cell Biol* 1992;11:1–20.
- Wess J. G-protein-coupled receptors: Molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *FASEB J* 1997;11:346–354.
- Baldwin JM, Schertler GFX, Unger VM. An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J Mol Biol* 1997;272:144–164.
- Cronet P, Sander C, Vriend G. Modeling of transmembrane 7 helix bundles. *Prot Eng* 1993;6:59–664.
- Herzyk P, Hubbard RE. Using biophysical and biochemical experimental information to generate 3-D structural models of transmembrane (TM) domains of helical, integral membrane proteins. *Biophys J* 1998;74:A249–A249.
- Orry AJW, Wallace BA. Modeling and docking the endothelin G-protein-coupled receptor. *Biophys J* 2000;79:3083–3094.
- Pogozheva ID, Lomize AL, Mosberg HI. The transmembrane 7-alpha-bundle of rhodopsin: Distance geometry calculations with hydrogen bonding constraints. *Biophys J* 1997;72:1963–1985.
- Sylte I, Edvardsen O, Dahl SG. Molecular-dynamics of the 5-HT(1a) receptor and ligands. *Prot Eng* 1993;6:691–700.
- Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 2000;289:739–745.
- Grigorieff N, Ceska TA, Downing KH, Baldwin JM, Henderson R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J Mol Biol* 1996;259:393–421.
- Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH. Model for the structure of bacteriorhodopsin based on high-resolution electron cryomicroscopy. *J Mol Biol* 1990;213:899–929.
- Unger VM, Hargrave PA, Baldwin JM, Schertler GFX. Arrangement of rhodopsin transmembrane alpha-helices. *Nature* 1997;389:203–206.
- Unger VM, Schertler GFX. 3-Dimensional structure of bovine rhodopsin at 9-angstrom-resolution. *Biophys J* 1994;66:A46–A46.
- Unger VM, Schertler GFX. Low-resolution structure of bovine rhodopsin determined by electron cryomicroscopy. *Biophys J* 1995;68:1776–1786.
- Edman K, Nollert P, Royant A, Belrhali H, Pebay-Peyroula E, Hajdu J, Neutze R, Landau EM. High-resolution X-ray structure of an early intermediate in the bacteriorhodopsin photocycle. *Nature* 1999;401:822–826.
- Luecke H, Richter HT, Lanyi JK. Proton transfer pathways in bacteriorhodopsin at 2.3 angstrom resolution. *Science* 1998;280:1934–1937.
- Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK. Structure of bacteriorhodopsin at 1.55 angstrom resolution. *J Mol Biol* 1999;291:899–911.
- An SZ, Zheng YH, Bleu T. Sphingosine 1-phosphate-induced cell proliferation, survival, and related signaling events mediated by G protein-coupled receptors EDG3 and EDG5. *J Biol Chem* 2000;275:288–296.
- Okamoto H, Takuwa N, Yatomi Y, Gonda K, Shigematsu H, Takuwa Y. EDG3 is a functional receptor specific for sphingosine 1-phosphate and sphingosylphosphorylcholine with signaling characteristics distinct from EDG1 and AGR16. *Biochem Biophys Res Comm* 1999;260:203–208.
- Pyne S, Pyne NJ. Sphingosine 1-phosphate signalling in mammalian cells. *Biochem J* 2000;349:385–402.
- Libert F, Schiffmann SN, Lefort A, Parmentier M, Gerard C, Dumont JE, Vanderhaeghen JJ, Vassart G. The orphan receptor cDNA RDC7 encodes an A1 adenosine receptor. *EMBO J* 1991;10:1677–1682.
- Maenhaut C, Vansande J, Libert F, Abramowicz M, Parmentier M, Vanderhaeghen JJ, Dumont JE, Vassart G, Schiffmann S. RDC8 codes for an adenosine-A2 receptor with physiological constitutive activity. *Biochem Biophys Res Comm* 1990;173:1169–1178.
- Lovenberg TW, Roland BL, Wilson SJ, Jiang XX, Pyati J, Huvar A, Jackson MR, Erlander MG. Cloning and functional expression of the human histamine H-3 receptor. *Mol Pharm* 1999;55:1101–1107.
- Song ZH, Slowey CA, Hurst DP, Reggio PH. The difference between the CB1 and CB2 cannabinoid receptors at position 5.46 is crucial for the selectivity of WIN55212-2 for CB2. *Mol Pharm* 1999;56:834–840.
- Homan EJ, Wikstrom HV, Grol CJ. Molecular modeling of the dopamine D-2 and serotonin 5-HT1A receptor binding modes of the enantiomers of 5-OMe-BPAT. *Bioorg Med Chem* 1999;7:1805–1820.
- Salminen T, Varis M, Nyronen T, Pihlavisto M, Hoffren AM, Lonnberg T, Marjamaki A, Frang H, Savola JM, Scheinin M, Johnson MS. Three-dimensional models of alpha(2A)-adrenergic receptor complexes provide a structural explanation for ligand binding. *J Biol Chem* 1999;274:23405–23413.
- Kim JH, Wess J, Vanrhee AM, Schoneberg T, Jacobson KA. Site-directed mutagenesis identifies residues involved in ligand recognition in the human a(2a) adenosine receptor. *J Biol Chem* 1995;270:13987–13997.
- Findlay J, Eliopoulos E. 3-Dimensional modeling of G-protein-linked receptors. *TIPS* 1990;11:492–499.
- Hibert MF, Trumppkallmeyer S, Bruinvels A, Hoflack J. 3-Dimensional models of neurotransmitter G-binding protein-coupled receptors. *Mol Pharm* 1991;40:8–15.
- Ijzerman AP, Vanderwenden EM, Vangalen PJM, Jacobson KA. Molecular modeling of adenosine receptors—the ligand-binding site on the rat adenosine a(2a) receptor. *Eur J Pharm Mol* 1994;268:95–104.
- Roper D, Jacoby E, Kruger P, Engels M, Grotzinger J, Wollmer A, Strassburger W. Modeling of G-protein coupled receptors with bacteriorhodopsin as a template—a novel approach based on interaction energy differences. *J Rec Res* 1994;14:167–186.
- ICM, version 2.8. San Diego, CA: Molsoft LLC; 2001.
- Schapiro M, Raaka BM, Samuels HH, Abagyan R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc Natl Acad Sci U S A* 2000;97:1008–1013.
- Schapiro M, Raaka BM, Samuels HH, Abagyan R. In silico discovery of novel retinoic acid receptor agonist structures. *BMC Struct Biol* 2001;1:1.
- Schapiro M, Das K, Raaka B, Fan L, Totrov M, Wilson S, Abagyan RA, James TL. Discovery of diverse thyroid hormone receptor antagonists by high-throughput docking. 2002. Submitted for publication.

39. Totrov M, Cavasotto CN, Abagyan RA. A new method for the identification of ligand binding sites. 2003. Manuscript in preparation.
40. Nemethy G, Gibson KD, Palmer KA, Yoon CN, Paterlini G, Zagari A, Rumsey S, Scheraga HA. Energy parameters in polypeptides: Ten improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J Phys Chem* 1992;96:6472–6484.
41. Halgren T. Merck molecular force field I-V. *J Comp Chem* 1995;17:490–641.
42. Totrov M, Abagyan R. Protein-ligand docking as an energy optimization problem. In: Raffa RB, editor. *Drug-receptor thermodynamics: Introduction and experimental applications*. New York: John Wiley & Sons; 2001.
43. Goto S, Nishioka T, Kanehisa M. LIGAND database for enzymes, compounds and reactions. *Nucleic Acids Res* 1999;27:377–379.
44. Goto S, Okuno Y, Hattori M, Nishioka T, Kanehisa M. LIGAND: Database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res* 2002;30:402–404.
45. Totrov M, Abagyan R. Derivation of sensitive discrimination potential for virtual ligand screening. *Proceedings of the Third Annual International Conference on Computational Molecular Biology*. Lyon, France: ACM Press; 1999. p 37–38.
46. Abagyan R, Totrov M, Kuznetsov D. ICM—a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comp Chem* 1994;15:488–506.
47. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol* 1994;235:983–1002.
48. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys* 1953;21:1087–1092.
49. Abagyan RA, Totrov M. Ab initio folding of peptides by the optimal-bias Monte Carlo minimization procedure. *J Comp Phys* 1999;151:402–421.
50. Abagyan R, Argos P. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J Mol Biol* 1992;225:519–532.
51. Li ZQ, Scheraga HA. Monte Carlo minimization approach to the multiple-minima problem in protein folding. *Proc Natl Acad Sci U S A* 1987;84:6611–6615.
52. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: The relaxed complex scheme. *J Am Chem Soc* 2002;124:5632–5633.