Homology Model Versus X-ray Structure in Receptor-based Drug Design: A Retrospective Analysis with the Dopamine D3 Receptor

Nicolas Levoin,* Thierry Calmels, Stephane Krief, Denis Danvy, Isabelle Berrebi-Bertrand, Jeanne-Marie Lecomte, Jean-Charles Schwartz, and Marc Capet

Bioprojet-Biotech, 4 rue du Chesnay-Beauregard, 35762 Saint-Grégoire, France

ABSTRACT: Structure-based design methods commonly used in medicinal chemistry rely on a three-dimensional representation of the receptor. However, few crystal structures are solved in comparison with the huge number of pharmaceutical targets. This often renders homology models the only information available. It is particularly true for G protein-coupled receptors (GPCRs), one of the most important targets for approved medicines and current drug discovery projects. However, very few studies have tested their validity in comparison with corresponding crystal structures, especially in a lead optimization perspective. The recent solving of dopamine D3 receptor crystal structure allowed us to assess our historical homology model. We performed a statistical analysis, by docking our in-house lead optimization library of 1500 molecules. We demonstrate here that the refined homology model suits at least as well as the X-ray structure. It is concluded that when the crystal structure of a given GPCR is not available, homology modeling can be an excellent surrogate to support drug discovery efforts.

KEYWORDS: Dopamine D3 receptor, G protein-coupled receptors, docking, homology model, lead optimization

Structure-based design (SBD) uses the information brought by

a receptor three-dimensional (3D) model to help the discovery and optimization of ligands. The protein 3D representation gives insights into the shape and electrostatics of the binding site, as well as the nature of its constitutive amino acids. Indeed, it gives insights into the possible interactions that ligands could form. SBD is widely used in medicinal chemistry at different stages of the drug discovery process. In the earliest, SBD can be instrumental in selecting new hits from a large library, thus reducing the size of the compounds to be experimentally tested (hit finding). Many examples in the literature report the success of virtual screening studies.^{1,2} SBD has also proved very useful in the following step of lead optimization, notably as a tool to understand and rationalize the structure-activity relationships. 3

This approach commonly starts from a reliable crystal structure of the target receptor. However, in comparison with the huge number of proteins of pharmaceutical interest, the number of available crystal structures is limited. This is particularly true for G protein-coupled receptors (GPCRs), the most important class of drug targets whose ligands represent about 50% of the currently marketed drugs.⁴ To compensate for the scarcity of crystal structures, molecular models of target proteins are often built based on a close structural neighbor (comparative/homology modeling). For GPCRs, bovine rhodopsin has long been the sole experimental structure available. So, a myriad of homology models (HMs) deriving from the breakthrough of Palczewski's work⁵ were built and used in virtual screening experiments.^{1,2}

Companion GPCR structures appeared in past years for opsin,⁶ β -1,⁷ human β -2 adrenergic, $8-10^{-1}$ and A2-adenosine receptor.¹¹

IVERTS)
 Example 18 dx.doi.org/10.1021 and Merican Chemical Chemical Chemical Chemical Chemical Society 2011

The Chemical Society 2012 and Merican Chemical Society 2012

All Chemical Society 2012 and Merican Chemical Very recently, the human chemokine CXCR4 crystal structure was solved in the bound form with small molecule and peptide ligands.¹² Human dopamine D3 receptor (D3R) was the last to be added to the list.¹³ The increasing and accelerating number of experimental GPCR structures is a great opportunity for SBD, mainly for three reasons. First, it enhances the probability of finding a homologue as close as possible to a given GPCR target. It is important since the quality of HM depends on the sequence identity between target and template proteins.¹⁴ Second, additional available scaffolds improve the model building process; indeed, multiple templates provide better models than a single one.¹⁵ Even if the list is growing, it is unlikely that all receptors could be crystallized, so HM might remain an irreplaceable tool. Such an observation brings the third contribution of X-ray structures: It gives the opportunity to test retrospectively the quality of existing HM and, by extrapolation, to lend weight to the models currently in use in many pharmaceutical projects.

Since the cloning of $\overline{D3R}$, ¹⁶ we have been intensively involved in the finding of a drug candidate. Our historical D3R model was built at the time when bovine rhodopsin was the only available template. This HM has been extensively used for the design of potent D3R ligands, resulting in a 1500 molecule library, with K_i values ranging from 0.1 nM to 10 μ M. The library contains very different scaffolds but also very close ligands in a given series

Figure 1. D3R chemical library from Bioprojet. Ligands are represented through 45 fingerprints and physical properties on a principal component space (principal components PC1, PC2, and PC3). On the left, ligands are colored according to their D3R affinity. It can be seen that each of the four chemical clusters contains all of the affinity range (not all potent ligands belong to the same chemical space). On the right, the library is compared to the Merck Index compounds (red spheres), taken as an archetype of chemical diversity.

Figure 2. Side chains contact plot for X-ray structure of D3R (A) and bovine rhodopsin-based HM (B). (C) Superimposition of α -carbon trace of HM (red) and X-ray structure (blue) of the receptor.

(Figure 1). As a consequence, it constitutes a very good and very challenging data set to compare the accuracy of HM and X-ray structures. It has already been suggested that X-ray structures are not a prerequisite for SBD but that HM can be suitable for that purpose.17,18 Recent D3R crystal structures and our D3R library permit us to investigate further the topic and ask the following

question: Does X-ray structure ameliorate profoundly the accuracy of SBD, or does HM perform comparably?

We compared our historical HM and X-ray structure of D3R (PDB: 3PBL) through geometrical as well as statistical analysis. Side chain contact plots for both structures revealed a very similar profile (Figure 2), and superimposition gave the rmsd shown in Table 1. As the two protomers present in the crystal unit cell differ from 1 (α carbons) and 1.7 Å (all heavy atoms), the 2.7 (α carbons) and 3.5 Å (all heavy atoms) between X-ray structure and HM are rather weak. Moreover, if we limit the comparison to the transmembrane domains (TM) and if we disregard the flexible intra- and extracellular loops, the rmsd falls to 1.7 (α carbons) and 2.5 Å (all heavy atoms).

Near the binding site, the most relevant region of the protein for SBD, few amino acid positions deviate significantly (Figure 3). We note that Phe346 is turned toward the binding site in the crystal structure, whereas it was facing out and increased the volume of the cavity in the HM. Trp342 is slightly translated, with centroids being 1.9 Å apart. As Phe346 in TM6, Ser192 and Ser196 of TM5 are also more engaged in the X-ray binding site than in HM (centroids are, respectively, 1.7 and 1.9 Å apart). Interestingly, these three most deviating residues are known to be implicated in the binding of agonists of dopamine D2 or D3 receptors¹⁹⁻²¹ or to be crucial in the conformational transition between inactive and active states of GPCRs.²² D3R was cocrystallized with the antagonist eticlopride. So, it is very likely that the crystal structure represents the inactive conformer. As BP897, 23 most of our ligands are partial agonists; they certainly interact with an (partial) active conformer of the receptor.

Figure 3. Detail of the supposed binding mode of BP897 in HM (carbon atoms in green) and X-ray structure (carbon atoms in gray). Important residues are labeled in the stereoview. A rmsd of 2.2 Å was calculated between BP897 in HM and in X-ray structure. From both complexes, BP897 has been predicted as an active compound $(K_i < 3.5$ nM).

^a The X-ray structure was used as native (raw) or refined after BP897 docking and energy minimization, simulated annealing, or molecular dynamics experiments. ROC AUC summarizes the classification of ligands as active or inactive compounds, with varying threshold. Sensitivity (Se) and specificity (Sp) were calculated at the activity threshold of 3.5 nM.³

So, it is finally not surprising that the residue positions described above are the most divergent.

Docking of eticlopride in our HM proposed a binding mode very similar to that seen in the crystal structure. Interacting residues are the same, and the salt bridge with Asp110 is also present, as well as the π -stacking observed with Phe345.

Hence, both binding sites are very similar. By docking our historical lead BP897, 24 a phenylpiperazine derivative with high potency and moderate selectivity, in both structures, we found identical orientations and conformations of the ligand. Most interacting residues are the same, and interaction types are conserved. The phenylpiperazine is the most buried part of the ligand, located in the middle of the TM section. The corresponding cavity of the receptor is formed by Val111, Cys114, Val189, Ser192, Ser193, His349, and Phe197. The implication of His349 was corroborated by mutagenesis experiments made on D3R²⁵ and that of Phe197 on $D2R^{16,19-21}$ The phenyl of the ligand is engaged in π -stacking with Phe346 of TM6, and the basic amine forms a salt bridge with Asp110 of TM3 (Figure 3).

The naphtyl is just beneath the extracellular loops, in contact with Leu89, Gly94 and Ser366.

In the HM, π -stacking is very probable with His29, Tyr32, and/or Trp370. In the X-ray structure, Tyr365 is a more likely candidate partner. In the HM, the carbonyl forms a hydrogen bond with a backbone amide of Cys181, whereas the nitrogen is H-bonded with backbone amide of Thr369 in the X-ray structure. For both structures, the linker is lined with Phe106, Val107, Tyr365, Thr369, and Thr373, in agreement with mutagenesis data on D3R or other dopamine receptors.²⁶

Both HM and X-ray structure explained very well the observed structure-activity relationships for D3R ligands. Briefly, an aromatic group is very important in the subcavity occupied by the phenyl of BP897. Compounds not able to form the π -stacking interaction with Phe345 are greatly penalized. The salt bridge between a basic

amine and Asp110 is also fundamental for a high affinity ligand. Because of the limited volume in the corresponding cavity, the linker can not be too bulky nor rigid. In agreement with the suggested H-bond with amide of BP897, high affinity ligands conserved this feature. Finally, the cavity embedding the naphtyl is voluminous, accounting for the observed structural tolerance in structureactivity relationships. The corresponding ligand substituent can be of variable lipophilic nature and of variable size without strong influence on the D3R affinity.

To obtain a statistical comparison between the HM and the X-ray structure, we used our "in-house" D3R chemical library through molecular docking. Ligand affinity was estimated by a customized scoring function, 27 defined as a consensus of existing scoring functions and additional terms dealing with lipophilic and conformational energy of the molecules. The accuracy of both models was quantified via receiver operating characteristics area under the curve $(ROC AUC).²⁸ ROC curves are a graphical plot of the sensitivity$ (true positive rate) vs specificity (true negative rate) for a binary classifier system as its discrimination threshold varies. So, it is well suited for evaluating virtual screening performance where one needs to discriminate between active and inactive compounds. According to the quality of the model, AUC values range from 0 to 1, with 0.5 signifying random selection. As seen in Table 2, our HM is very successful, with a ROC AUC of 0.72. This result can be compared to virtual screening experiments conducted with X-ray structures of human immunodeficiency virus HIV protease (AUC = 0.66- 0.69),²⁹ dihydropteroate synthase $(AUC = 0.69 - 0.73)$,³⁰ or a set of other crystal structures $(AUC = 0.55 - 0.77)^{31}$ Surprisingly, the raw X-ray structure appears to be less accurate, with ROC AUC of 0.66. One reason is probably that our HM was subjected to multiple refinements to take into account the knowledge extracted from medicinal chemistry. Hence, during the lead optimization process, the HM has been continuously improved with the ongoing structure-activity relationships. Thus, we tried to slightly relax

ACS Medicinal Chemistry Letters Letters Letters Letters Letters Letters Letters Letters Letters Letters Letters

the crystal structure through different protocols: energy minimization simulated annealing or molecular dynamics experiments in presence of BP897. That way, all ligands can be docked, but accuracy increased only marginally. It seems that a more profound modification of the X-ray structure would be necessary and not only relaxation in presence of one ligand. It is also important to keep in mind that a number of our ligands are partial agonists, and the difficulty of GPCR X-ray structure to deal with agonists has already been reported.³²

The solving of D3R crystral structure by Chien et al.¹³ is a landmark in the future of dopamine receptors ligand design. It is of great pharmaceutical interest given that the cerebral dopaminergic system is implicated in many psychiatric and neurologic disorders, such as Parkinson's disease, schizophrenia, depression, and substance abuse.³⁴ In addition, the X-ray structure of D3R gave us the opportunity to test the validity of a HM that we have built previously based on bovine rhodopsine. This comparison evidenced that we had constructed a 3D model geometrically very close to the experimental one. By docking our historical chemical library and comparing the accuracy estimated affinity, we have also shown that a refined HM is at least as good as the raw X-ray structure. Although medicinal chemistry is still eager to obtain new crystal structures of protein targets, our study with D3R demonstrates that HM is already an excellent substitute. Given the large number of GPCRs currently subject to drug discovery efforts, this study is rather an encouraging account.

AUTHOR INFORMATION

Corresponding Author

*Tel: 02 99 28 04 40. E-mail: n.levoin@bioprojet.com.

REFERENCES

(1) Senderowitz, H.; Marantz, Y. G Protein-Coupled Receptors: Target-based in silico screening. Curr. Pharm. Des. 2009, 15, 4049–4068.

(2) de Graaf, C.; Rognan, D. Customizing G Protein-coupled receptor models for structure-based virtual screening. Curr. Pharm. Des. 2009, 15, 4026–4048.

(3) Levoin, N.; Calmels, T.; Poupardin-Olivier, O.; Labeeuw, O.; Danvy, D.; Robert, P.; Berrebi-Bertrand, I.; Ganellin, C. R.; Schunack, W.; Stark, H.; Capet, M. Refined Docking as a Valuable Tool for Lead Optimization: Application to Histamine H3 Receptor Antagonists. Arch. Pharm. Pharm. Med. Chem. 2008, 341, 610–623.

(4) Lundstrom, K. Latest development in drug discovery on G protein-coupled receptors. Curr. Protein Pept. Sci. 2006, 7, 465–470.

(5) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Trong, I. L.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. Science 2000, 289, 739–745.

(6) Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krausz, N.; Choe, H.-W.; Hofmann, K. P.; Ernst, O. P. Crystal structure of opsin in its G-protein-interacting conformation. Nature 2008, 455, 497–502.

(7) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G. W.; Tate, C. G.; Schertler, G. F. X. Structure of a β 1-adrenergic G-protein-coupled receptor. Nature 2008, 454, 486–491.

(8) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. r. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-Resolution Crystal Structure of an Engineered Human β2-Adrenergic G Protein Coupled Receptor. Science 2007, 318, 1258–1265.

(9) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. r. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Yao, X.-J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. GPCR Engineering Yields High-Resolution Structural Insights into β2-Adrenergic Receptor Function. Science 2007, 318, 1266–1273.

(10) Rasmussen, S. G. F.; Choi, H.-J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R. P.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F. X.; Weis, W. I.; Kobilka, B. K. Crystal structure of the human β 2 adrenergic G-protein-coupled receptor. Nature 2007, 450, 383–387.

(11) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; IJzerman, A. P.; Stevens, R. C. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. Science 2008, 322, 1211–1217.

(12) Wu, B.; Chien, E. Y. T.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. Science 2010, 330, 1066–1071.

(13) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Won Han, G.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. Structure of the Human Dopamine D3 Receptor in Complex with a D2/D3 Selective Antagonist. Science 2010, 330, 1091– 1095.

(14) Forrest, L. R.; Tang, C. L.; Honig, B. On the Accuracy of Homology Modeling and Sequence Alignment Methods Applied to Membrane Proteins. Biophys. J. 2006, 91, 508–517.

(15) Bera, I.; Laskar, A.; Ghoshal, N. Exploring the structure of opioid receptors with homology modeling based on single and multiple templates and subsequent docking: A comparative study. J. Mol. Model. 2010, 1–15.

(16) Sokoloff, P.; Giros, B.; Martres, M.-P.; Bouthenet, M.-L.; Schwartz, J.-C. Molecular cloning and characterization of a novel dopamine receptor (D3) as a target for neuroleptics. Nature 1990, 347, 146–151.

(17) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? Proteins 2003, 50, 5–25.

(18) Costanzi, S. On the Applicability of GPCR Homology Models to Computer-Aided Drug Discovery: A Comparison between In Silico and Crystal Structures of the β2-Adrenergic Receptor. J. Med. Chem. 2008, 51, 2907–2914.

(19) Javitch, J. A.; Ballesteros, J. A.; Weinstein, H.; Chen, J. A Cluster of Aromatic Residues in the Sixth Membrane-Spanning Segment of the Dopamine D2 Receptor Is Accessible in the Binding-Site Crevice. Biochemistry 1998, 37, 998–1006.

(20) Sartania, N.; Strange, P. G. Role of Conserved Serine Residues in the Interaction of Agonists with D3 Dopamine Receptors. J. Neurochem. 1999, 72, 2621–2624.

(21) Cho, W.; Taylor, L. P.; Mansour, A.; Akil, H. Hydrophobic Residues of the D2 Dopamine Receptor Are Important for Binding and Signal Transduction. J. Neurochem. 1995, 65, 2105–2115.

(22) Shi, L.; Liapakis, G.; Xu, R.; Guarnieri, F.; Ballesteros, J. A.; Javitch, J. A. β2 Adrenergic Receptor Activation. J. Biol. Chem. 2002, 277, 40989–40996.

(23) Pilla, M.; Perachon, S.; Sautel, F.; Garrido, F.; Mann, A.; Wermuth, C. G.; Schwartz, J.-C.; Everitt, B. J.; Sokoloff, P. Selective inhibition of cocaine-seeking behaviour by a partial dopamine D3 receptor agonist. Nature 1999, 400, 371–375.

(24) Hackling, A.; Ghosh, R.; Perachon, S.; Mann, A.; Höltje, H.-D.; Wermuth, C. G.; Schwartz, J.-C.; Sippl, W.; Sokoloff, P.; Stark, H. J. Med. Chem. 2003, 46, 3883–3899.

(25) Lundstrom, K.; Turpin, M. P.; Large, C.; Robertson, G.; Thomas, P.; Lewell, X. Q. Mapping of Dopamine D3 Receptor Binding Site by Pharmacological Characterization of Mutants Expressed in Cho Cells with the Semliki Forest Virus System. J. Recept. Signal Transduction 1998, 18, 133–150.

(26) Kortagere, S.; Gmeiner, P.; Weinstein, H.; Schetz, J. A. Certain 1,4-Disubstituted Aromatic Piperidines and Piperazines with Extreme Selectivity for the Dopamine D4 Receptor Interact with a Common Receptor Microdomain. Mol. Pharmacol. 2004, 66, 1491–1499.

(27) We developed a dedicated scoring function (D3R_score) by using a consensus of functions DOCK SCORE, PLP1, and Ludi_1. logD was added to the energy term to account for the cost of ligand desolvation. The principal moment of inertia along Y-axis was also added to penalize folded conformers of ligands:

$\text{D3R_score} = \text{DOCK_SCORE} \times \text{PLP1} \times \text{Ludi_1} \times \text{logD} \times \text{PMI_Y}$

(28) Triballeau, N.; Acher, F.; Brabet, I.; Pin, J.-P.; Bertrand, H.-O. Virtual Screening Workflow Development Guided by the Receiver Operating Characteristic Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. J. Med. Chem. 2005, 48, 2534–2547.

(29) Chang, M. W.; Ayeni, C.; Breuer, S.; Torbett, B. E. Virtual Screening for HIV Protease Inhibitors: A Comparison of AutoDock 4 and Vina. PLoS ONE 2010, 5, e11955.

(30) Hevener, K. E.; Zhao, W.; Ball, D. M.; Babaoglu, K.; Qi, J.; White, S. W.; Lee, R. E. Validation of Molecular Docking Programs for Virtual Screening against Dihydropteroate Synthase. J. Chem. Inf. Model. 2009, 49, 444–460.

(31) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. J. Chem. Inf. Model. 2009, 49, 1455–1474.

(32) Reynolds, K.; Katritch, V.; Abagyan, R. Identifying conformational changes of the $β$ -2 adrenoceptor that enable accurate prediction of ligand/receptor interactions and screening for GPCR modulators. J. Comput.-Aided Mol. Des. 2009, 23, 273–288.

(33) All of the modeling procedures (building and refinement of the homology model, docking, and scoring) followed the same protocols as in ref 3 with histamine H3 receptor. Briefly, after building a raw template-based model by analogy to bovine rhodopsine (Modeller implemented in Discovery Studio, Accelrys, CA), lead receptor complexes were optimized through different protocols, including energy minimization, molecular dynamics, and simulated annealing. As in ref 3, simulated annealing gave the best results (Discover with CFF force field, 6000 steps of minimization followed by 50 iterations of 100 fs drop of temperature from 500 to 300 K, in NVT conditions and temperature controlled by velocity scale. The backbone was kept frozen except for the extracellular loops). The chemical library results from our in-house lead optimization efforts on D3R project. At the K_i threshold of 3.5 nM, it is divided in two equal subpopulations (775 active and 804 inactive ligands). Sensitivity Se = $TP/(TP + FN)$ and specificity Sp = TN/ $(TN + FP)$ were calculated at this threshold. Here, TP stands for true positives, FN for false negatives, TN for true negatives, and FP for false positives. The protonation of ligands was supposed to be the majority state at pH 7. Conformers were generated with the Catalyst module implemented in Discovery Studio (around 250 conformers obtained with the "best" method, with an energy threshold of 20 kcal/mol).

(34) Zhang, A.; Neumeyer, J. L.; Baldessarini, R. J. Recent Progress in Development of Dopamine Receptor Subtype-Selective Agents: Potential Therapeutics for Neurological and Psychiatric Disorders. Chem. Rev. 2006, 107, 274–302.