JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

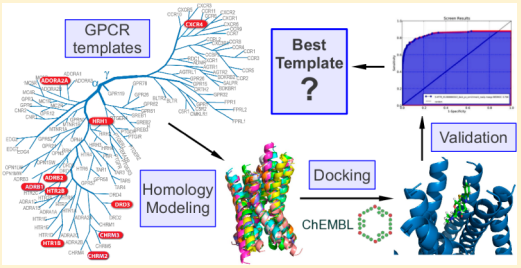
Impact of Template Choice on Homology Model Efficiency in Virtual Screening

Krzysztof Rataj,[†] Jagna Witek,[†] Stefan Mordalski,[†] Tomasz Kosciolek,[†,§] and Andrzej J. Bojarski*[,†]

[†]Department of Medicinal Chemistry, Institute of Pharmacology, Polish Academy of Sciences, 12 Smętna Street, 31-343 Krakow, Poland

Supporting Information

ABSTRACT: Homology modeling is a reliable method of predicting the three-dimensional structures of proteins that lack NMR or X-ray crystallographic data. It employs the assumption that a structural resemblance exists between closely related proteins. Despite the availability of many crystal structures of possible templates, only the closest ones are chosen for homology modeling purposes. To validate the aforementioned approach, we performed homology modeling of four serotonin receptors (5-HT_{1A}R, 5-HT_{2A}R, 5-HT_6R, 5-HT_7R) for virtual screening purposes, using 10 available G-Protein Coupled Receptors (GPCR) templates with diverse evolutionary distances to the targets, with various approaches to alignment construction and model building. The resulting models were further validated in two steps by means of ligand docking and enrichment calculation, using Glide software. The final quality of the models was determined in virtual screening-like experiments by the AUROC score of the resulting ROC curves. The outcome of this research showed that no correlation between sequence identity and model quality was found, leading to the conclusion that the closest phylogenetic relative is not always the best template for homology modeling.



INTRODUCTION

Homology modeling is a popular method of predicting protein's three-dimensional structure in the absence of X-ray crystallography or NMR data. It is based on experimentally determined structures of the target's homologues, according to the assumption that similar protein sequences should result in similar native conformations.[1,2] A three-dimensional structure enables numerous computational studies concerning protein's functionality, stability, interactions, or kinetics, and it often helps to direct *in vitro* research and reduce the time and funds consumed. Among many factors that affect the quality of obtained models, the template choice appears to be one of the most important. Regardless of the applied modeling approach, it is believed that only proteins whose sequence similarity is higher than 50% can provide a reliable model.[2] However, there are exceptions to this rule, such as the case of hemoglobin and myoglobin, that have low sequence similarity (24%) but highly homologous structures (their root-mean-square deviation, RMSD, is only 1.35 Å).

Computational modeling techniques and protein structure prediction methods are continually improving in their predictive capabilities.[3] Although accurate, *de novo* protein folding is extremely time-consuming[4] and is limited to small, single-domain proteins. Therefore, in cases where a homologous protein structure is available, the homology modeling approach is the preferred solution. However, the resulting models bear significant structural resemblance to the template, hence the need for optimal template identification.[3,5,6] There are a number of available automated homology modeling protocols, such as GOMoDo[7] or GPCRM,[8] however such methods rely on substitution matrixes for sequence alignment.

G-protein coupled receptors comprise a transmembrane protein superfamily of great pharmacological interest.[9] GPCRs transduce extracellular stimuli into various signaling pathways that alter cell functions, and their activity can be modulated by exogenous molecules.[10] The GPCR family is divided into six classes (A−F), the largest of which is rhodopsin family A, which accounts for approximately 84% of the entire superfamily. Numerous efforts have been made to model GPCRs on available homologous templates since the crystal structures of bacteriorhodopsin[11] and bovine rhodopsin were determined.[12] Overall, 14 structures of human GPCR-A proteins have been resolved so far,[13−26] which significantly broadens the spectrum of modeling possibilities. Some proteins were cocrystallized as complexes with agonists or antagonists, giving insight into possible conformational changes triggered by ligand binding.

In this study we have focused on modeling four members of class A GPCRs: serotonin receptors 5-HT_{1A}, 5-HT_{2A}, 5-HT_6, and 5-HT_7. These proteins are of great interest to the pharmaceutical industry due to their important roles in mood and behavior regulation.[27−30] These roles make them popular targets for drug research against depression and anxiety. Aside from those aspects, each of the target proteins has additional, specific functions.

Received: January 2, 2014

   dx.doi.org/10.1021/ci500001f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

The initial target of this study, 5-hydroxytryptamine receptor type 6 (5-HT_6R), was identified in 1993[31] and subsequently cloned in 1996.[32] It is almost exclusively expressed in the brain and is considered to be involved in cognitive functions such as learning and memorizing.[33,34]

The 5-hydroxytryptamine receptors of types 1A (5-HT_{1A}R) and 2A (5-HT_{2A}R) were among the earliest serotonin receptors discovered, in 1979[35] and 1957,[36] respectively. 5-HT_{1A}R is largely distributed in the central nervous system (CNS) and has been reported to play a role in neuroendocrine regulation, hyperphagia, and cardiovascular regulation.[37] Meanwhile, the 5-HT_{2A} receptor is considered responsible for smooth muscle tissue contractile responses, such as in the bronchus, uterus, and urinary system. Its activation stimulates hormone secretion, such as corticosterone, oxytocin, or prolactin.[38]

The 5-hydroxytryptamine receptor type 7 (5-HT_7R) is the least homologous serotonin receptor within its family, showing less than 50% homology to other members. The receptor itself was discovered in 1983 as a 5-HT_1-like protein[39] and was cloned and characterized 10 years later.[40] It is involved in thermoregulation and sleep control.[30]

Unfortunately, the three-dimensional conformations of these receptors remain unknown despite numerous attempts to model their structures.[41−51] Most attempts have utilized rhodopsin and beta-2 adrenergic receptor as templates despite the availability of other class A GPCR structures. The key aim of these modeling studies was to reveal the ligand−receptor interaction mechanism with the greatest possible accuracy.

Our research initially focused on modeling 5-HT_6R to acquire a suitable receptor for use in virtual screening (VS). To achieve the best models possible, all template proteins available at that time were used, and different approaches to alignment construction, helix range determination, and model building were employed. We obtained unexpected results and therefore decided to enlarge the scope of our study by including three more serotonin receptors targets and four additional templates.

MATERIALS AND METHODS

Alignments and Construction of Models. The sequences of target human proteins were obtained from the UniProtKB/Swiss-Prot database:

- 5-HT_6R ID: **P50406**
- 5-HT_7R ID: P34969
- 5-HT_{1A}R ID: P08908
- 5-HT_{2A}R ID: P28223

The crystal structures of GPCRs were extracted from the PDB (Table 1), based primarily on the resolution of available structures and secondarily on the type of ligand bound (preferred agonist).

The alignments were performed using Accelrys Discovery Studio.[53] All alignments were constructed manually (no substitution matrix was used). The alignments were generated with special emphasis on the characteristic 5-HTR motifs (Table 2), especially the X.50 amino acid (Ballesteros−Weinstein notation,[54] where X indicates helix number), and avoiding gaps within helices. As transmembrane helices are amphipathic, such gaps could cause a sequence shift and expose a hydrophilic residue to the lipid bilayer. The alignments were validated with the use of mutagenesis data to determine whether functionally relevant residues (whose mutation triggered significant activity changes) were present in the aligned fragments. All information was obtained by mining publications that provide *in vitro* assay data (Table 3).

Table 1. Crystal Structures of Templates Used for Homology Modeling

template	PDB ID	resolution [Å]	publication date
serotonin 1B receptor	4IAR[20]	2.70	13.03.2013
serotonin 2B receptor	4IB4[19]	2.70	13.03.2013
adenosine A2A receptor	3QAK[22]	2.71	09.03.2011
adrenergic beta1 receptor	2Y00[52]	2.50	12.01.2011
adrenergic beta2 receptor	3P0G[21]	3.50	19.01.2011
CXC chemokine receptor type 4	3OE0[24]	2.90	27.10.2010
dopamine 3 receptor	3PBL[23]	2.89	03.11.2010
histamine 1 receptor	3RZE[25]	3.10	15.06.2011
muscarine 2 receptor	3UON[18]	3.00	01.02.2012
muscarine 3 receptor	4DAJ[17]	3.40	22.02.2012

Table 2. Characteristic 5-HT Receptor Sequence Motifs

helix number	position	sequence motif
TM1	1.50	N
TM2	2.50	D
	2.45	S
	2.57−2.59	VMP
TM3	3.50	R
	3.39−3.43	SIXX[L-V]
	3.46−3.54	I[S-A][L-V]DRYXXI
TM4	4.50	W
TM5	5.50	P
	5.47−5.50	F[Y−F]XP
	5.58	Y
TM6	6.50	P
	6.44	F
	6.48−6.52	WXPFF
TM7	7.50	P
	7.40−7.46	W[L-I]GYXXS
	7.49	N
	7.53	Y

Table 3. Site-Directed 5-HT_6 Receptor Mutants with Aberrant Activity

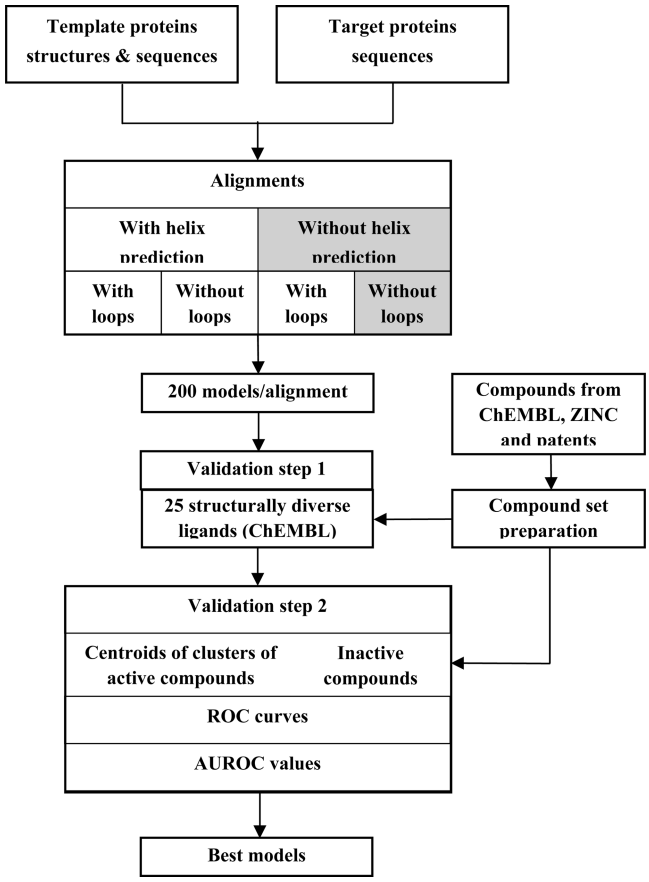
author, publication year	mutation
Harris et al., 2010[55]	D3.32A, C3.36A, S5.43A, T5.46A, W6.48A, F6.52A, N6.55A
Fuente et al., 2010[48]	C3.36A, S5.43A
Purohit et al., 2003[56]	S6.34K

The methodology used for 5-HT_6R modeling (Scheme 1) consisted of two approaches to define helix ranges: (i) assuming that the modeled protein has the same ranges as the template and (ii) prediction of transmembrane helices with the use of metaservers (*PONGO* at pongo.biocomp.unibo.it,[57] *PSIPRED* at bioinf.cs.ucl.ac.uk/psipred/,[58,59] and *ExPASy:PROSITE* at expasy.org/prosite[60]). The output consisted of a consensus helix range, where some of the helices had to be manually altered to retain the X.50 amino acid within the helix range.

Phylogenetic trees of all target proteins and template GPCRs were constructed based on *protdist* and *neighbor* software from Mobyle@Pasteur metaserver[61] and are available in the

dx.doi.org/10.1021/ci500001f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Scheme 1. Workflow for Homology Model Generation and Validation[a]



[a]The approach to alignment generation for 5-HT_{1A}R, 5-HT_{2A}R, and 5-HT_7R is shown in grey.

Supporting Information (Figure S1). Default *Protdist* parameters were used. As GPCRs share a common ancestor,[54] the UPGMA (*Unweighted Pair Group Method using Arithmetic averages*) algorithm was used in *neighbor* to construct a rooted phylogenic tree.

Two methods of model construction were applied: building only transmembrane regions and modeling the entire receptor structure, including loops. Visual inspection of template structures enabled recognition of residues causing helix bending (typically a proline residue). When helix-breaking residues were present, it was necessary to apply additional restraints on predicted secondary structures. If present, the disulfide bridges within loop regions were included in the appropriate templates. For each aligned template, a series of 200 models was produced using Modeller 9v8,[62] employing an approach previously utilized in our laboratory.[44]

The workflow applied to all remaining target proteins was limited to building alignments with only template-determined helix ranges and without loops (Scheme 1), as all other approaches did not show significant changes in the final model's virtual screening (VS) efficiency (see the Results section).

Validation. The modeled structures were validated using a two-step *flexible ligand* docking method (performed using Glide[63] software in XP precision mode) to minimize time consumption and maximize efficacy. Sets of ligands with known activities were acquired from the ChEMBL database (version 11 for 5-HT_6R and 5-HT_7R and version 15 for 5-HT_{1A}R and 5-HT_{2A}R)[64] and patents. The three-dimensional structures of compounds were generated using the LigPrep[65] application from the Schrödinger package. The OPLS-2005 force field was used.

Step 1: For each target protein model, a pretest set of 25 structurally diverse, active (K_i < 10 nM) compounds was docked to reject receptor models binding insufficient ligand number (less than 13) or with too high value of Glide Score (higher than −3). The sets were acquired through hierarchical clustering using the Canvas[66] software (MOLPRINT2D fingerprints, Tanimoto metrics). The number of clusters was manually altered so that, after excluding clusters containing less than three elements, 25 clusters remained. From these, centroids were chosen to form the pretest set.

Step 2: To calculate enrichment factors and receiver operating characteristic (ROC) curves, two sets of ligands were used for each target protein: **Actives**, where for all active (K_i < 100 nM) ligands, Canvas was used to perform hierarchical clustering (see above, with the Complete Cluster Linking Method used). From each cluster containing more than 2 members, the centroid was selected. The final numbers of ligands used are listed in Table 4. **Inactives**, where for each target protein, a collection of inactive compounds from the ChEMBL (K_i > 1000 nM) database was selected. The only exception was for 5-HT_6R, which has few inactive compounds listed in the database. Therefore, a set of ∼1000 randomly extracted compounds from the ZINC[67] database was used (they constituted the so-called assumed inactives).

Table 4. Numbers of Compounds Used in Model Validation

protein	number of active compounds before clustering	number of active compounds used in docking studies	number of inactive compounds
5-HT_{1A}R	3901	229	1194
5-HT_{2A}R	2391	250	2085
5-HT_6R	4298	258	1063
5-HT_7R	855	118	1594

For every model in step 2, a ROC curve was calculated based on the Glide Score values of docked compounds, including undocked actives and inactives as false negatives and true negatives, respectively. The final model quality was determined by the Area Under ROC curve (AUROC).

The electrostatic surfaces were calculated for top ranked models of 5-HT_6R, along with binding cleft analysis performed with the SiteMap[68] module of Schrödinger Suite 2012.

RESULTS

Since predicting transmembrane helix ranges or modeling loop regions for 5-HT_6R (as mentioned in the Materials and Methods section) did not significantly change the quality of the resulting models (AUROC scores varied in range of approximately 10%, whenever the models were available after step 1; Table 5), it is safer to omit their construction while modeling using more distant templates.[69] Therefore, only the limited approach to alignment generation and model construction is discussed further.

Alignments. Transmembrane region alignments are shown in Figure S2 of the Supporting Information. Phylogenetic tree construction based on full protein sequences (Figure S1 in the Supporting Information) allowed identification of the evolutio-

dx.doi.org/10.1021/ci500001f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 5. AUROC Values of 5-HT$_6$R Models Acquired in Different Alignment and Model Generation Approaches

	AUROC score			
	helix prediction		predetermined helix	
template	no loops	with loops	no loops[a]	with loops
adenosine A2A receptor	0.635	n/a[c]	0.693	0.658
adrenergic beta1 receptor	n/a[c]	n/a[c]	n/a[c]	0.714
adrenergic beta2 receptor	**0.777**[b]	n/a[c]	**0.730**[b]	**0.772**[b]
CXC chemokine receptor type 4	0.754	**0.728**[b]	0.718	0.709
dopamine 3 receptor	0.749	0.656	0.689	0.581
histamine 1 receptor	n/a[c]	n/a[c]	0.605	0.609

[a]The modeling method used for the remaining targets. [b]The best AUROC values. [c]The models that did not pass the validation step 1.

narily closest template proteins. Similarity/identity scores for sequence alignments are shown in Table 6.

Modeling and Validation. Modeler generated 200 models per alignment, leading to 8000 GPCR models total. Such an amount of receptor structures provided us with sufficient coverage of the conformational space of crucial residues involved in ligand binding (see Supporting Information for the details). The two-step validation process was introduced to filter out structures with poor ligand binding properties before commencing extensive docking and enrichment calculations. Validation step 1 (see Materials and Methods) resulted in discarding 92% of the initial models (Table 7).

The remaining models were subjected to validation step 2 to assess the model's ability to discriminate between active and inactive compounds. To achieve that, both active and decoy structures were docked, and ROC curves were calculated based on their Glide Score value (Figure 1). The quality of homology models in VS was determined by AUROC values of the curves (Table 8).

DISCUSSION

The results of the two-step VS-like validation process show that, in all cases, the templates that performed best in virtual screening were not the phylogeneticaly closest ones, as the common paradigm states. In the case of 5-HT$_{1B}$R and 5-HT$_{2B}$R templates, which are closely related to 5-HT$_{1A}$ and 5-HT$_{2A}$, respectively, this relation seems to revert, as the most distant protein, CXCR4, proved to be the best template for modeling for VS purposes. This conclusion differs from that of previous

Table 7. Number of Protein Models Fulfilling the Requirements of the First Verification Step

	number of models after validation step 1			
	5-HT$_{1A}$	5-HT$_{2A}$	5-HT$_6$	5-HT$_7$
serotonin 1B receptor	0	0	1	9
serotonin 2B receptor	0	0	0	0
adenosine A2A receptor	36	34	46	8
adrenergic beta1 receptor	1	39	0	24
adrenergic beta2 receptor	14	4	2	4
CXC chemokine receptor type 4	5	33	5	10
dopamine 3 receptor	21	53	4	24
histamine 1 receptor	17	42	7	18
muscarine 2 receptor	2	0	66	17
muscarine 3 receptor	22	0	47	51

reports; however, such outcome could be due to differences in alignment and model building approaches or due to the specificity of the ligands used to validate the models. Other studies built multiple alignments among the target receptor and its homologues, utilizing substitution matrices (e.g., BLOSUM)[45] and choosing one modeling template.[45,48] These approaches revealed conserved sequence motifs for related proteins and appear to be a convenient and fast method of model construction. However, the manual alignment of important regions is more accurate. In other studies, models were constructed with alignment performed only by automatically aligning the target and template protein sequences, without considering conserved residues and the insertion of gaps into helices.[49] We obtained mediocre results when testing a similar procedure for 5-HT$_6$R (Figure S3 of Supporting Information). The most similar approach to the one presented here was performed with alignment construction based on characteristic X.50 motifs, and it avoided the insertion of gaps into helices.[42,70] However, the transmembrane helices ranges were predicted by a hydropathic analysis algorithm.[42] Regardless, it should be noted that for different GPCR targets other alignments may be used to obtain better results. In some cases, even gaps in TM helices are allowed (e.g., μ-opioid receptor[71]).

The results of our study are contradictory to the widespread opinion that the closest related protein is the best modeling template for VS purposes, since no definable correlation between evolutionary distance of target protein to template and its ability to discriminate between active and inactive compounds was found. We examined structural and physical

Table 6. Identity and Similarity Percentages between the Whole Protein Sequences of Templates and Targets[a]

	identity (similarity) [%]			
template	5-HT$_{1A}$	5-HT$_{2A}$	5-HT$_6$	5-HT$_7$
serotonin 1B receptor	**50.34 (57.70)**	33.73 (40.92)	38.86 (47.43)	**36.81 (44.00)**
serotonin 2B receptor	42.80 (51.88)	**47.43 (53.59)**	38.01 (44.34)	30.65 (39.04)
adenosine A2A receptor	42.29 (50.00)	29.62 (37.67)	34.76 (42.29)	30.99 (38.35)
adrenergic beta1 receptor	45.20 (52.22)	34.24 (42.12)	**40.06 (47.77)**	34.07 (40.92)
adrenergic beta2 receptor	42.80 (51.02)	32.53 (39.04)	37.50 (44.52)	32.87 (39.21)
CXC chemokine receptor type 4	33.39 (39.72)	25.51 (32.87)	29.96 (35.78)	24.48 (31.16)
dopamine 3 receptor	45.03 (53.25)	34.93 (41.26)	38.01 (46.23)	34.07 (41.26)
histamine 1 receptor	43.15 (51.36)	31.33 (41.60)	38.18 (45.20)	31.67 (38.86)
muscarine 2 receptor	41.09 (50.00)	29.62 (38.35)	34.93 (42.80)	30.30 (38.18)
muscarine 3 receptor	39.89 (49.31)	30.82 (37.84)	35.10 (42.63)	29.96 (36.98)

[a]Sequences with the highest scores are shown in **bold**.

dx.doi.org/10.1021/ci500001f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

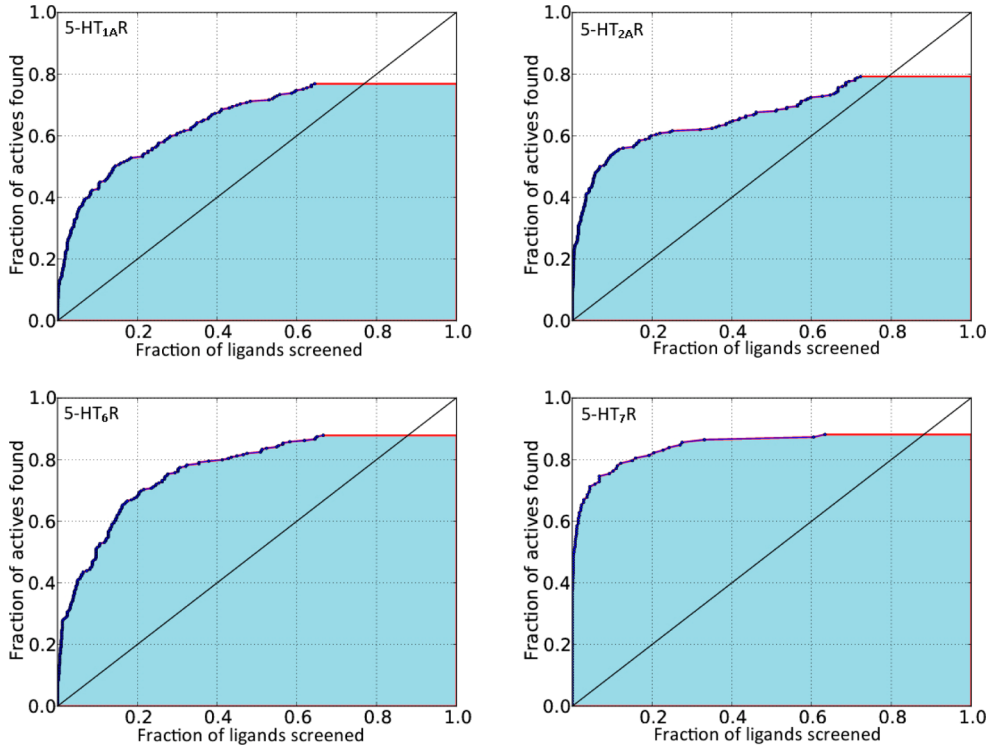


Figure 1. ROC curves for models with the best AUROC scores after second step of validation. Premature ending of the curves is caused by models' inability to dock all of the input active compounds.

Table 8. AUROC Values of the Best Models after Validation Step 2

	AUROC score[a]			
template	5-HT_{1A}	5-HT_{2A}[b]	5-HT_6	5-HT_7
serotonin 1B receptor	n/a[c,d,f]	n/a[c]	0.499	0.441[d,f]
adenosine A2A receptor	0.393	0.620	0.693	0.709
adrenergic beta1 receptor	0.573	0.482	n/a[c,f]	0.786
adrenergic beta2 receptor	0.576	0.541	0.730[e]	0.757
CXC chemokine receptor type 4	0.653[e]	0.681[e]	0.718	0.669
dopamine 3 receptor	0.630	0.611	0.689[d]	0.764
histamine 1 receptor	0.641	0.601	0.605	0.828[e]
muscarine 2 receptor	0.406	n/a[c]	0.639	0.717
muscarine 3 receptor	0.529	n/a[c]	0.661	0.749

[a]The Pearson's Correlation Coefficient between AUROC values and whole protein sequence identities is −0.481. [b]The protein with the closest relatedness and highest sequence identity for 5-HT_{2A} was 5-HT_{2B}, which did not pass the first validation step. [c]The "n/a" means that no respective template passed the validation step 1. [d]The evolutionarily closest template [e]The template with the best AUROC score [f]The template with the highest sequence identity

features that might differ among the templates. The differences in electrostatic surfaces were slight and would not be predicted to affect the final result. The volumes of binding sites, on the other hand, were very diverse, but there was no correlation between their values and the AUROC score. The volumes of the best models varied between 284 $Å^3$ for beta2-AR and 647 $Å^3$ for H_1R (Figure S4 in the Supporting Information). A similar distribution was observed for hydrophobic/hydrophilic surfaces (Figure S5 in the Supporting Information), which leads to the conclusion that there is currently no universal template for GPCR modeling nor any target-specific template.

For VS homology modeling purposes, the three-dimensional structure of the binding pocket should bear more significance than overall protein similarity. Unfortunately, there appears to be no correlation between binding site sequence similarity and the final AUROC value (Table S2 in the Supporting Information), which shows that binding site sequence similarity is not a valid criterion for template choice. Additionally, RMSD values of both binding sites and whole models are not consistent with the differences in overall quality (Table S3 in the Supporting Information). Thus, the optimal VS template/model can be obtained only through extensive template evaluation.

The results show large variation in the quality (assessed by AUROC values) of models built on the same template protein. This variation suggests that, to achieve an optimal structure for VS purposes, the construction of multiple models is necessary, whereas in most studies only a single model was built. The failure of serotonin receptors as template proteins is also surprising, leading to the conclusion that the same natural ligand of the protein may not by an indicator of a proper template protein because X-ray crystal structures are often determined using ligand-bound proteins and, therefore, are limited to only one conformation that is heavily dependent on the ligand type.

It has to be stressed that this research was designed to determine a single model with the best classification properties for VS procedures. The model selected in such a manner can be assumed to be the one most resembling the target protein, especially in the binding pocket region. The GPCR Dock 2010 Assessment states that only closely related proteins are eligible as templates in GPCRs homology modeling,[72] which is in clear contrast to our findings. However, the GPCR Dock study focused on recreating binding modes of ligands in crystal structures and, therefore, cannot be directly related to our

dx.doi.org/10.1021/ci500001f | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

research. Nevertheless, the results of both studies should be taken into consideration, as the validity of models in GPCR Dock was measured by RMSD of only one docked ligand. This result further emphasizes the fact that a single template can generate models of varying predictive quality for active ligand binding.

CONCLUSIONS

The results of our study show that the optimal homology models for use in VS may be based on evolutionarily distant templates. This implicates that the common reflex to use the closest related protein as a template may not be the best approach. However, models constructed and validated in this study are only for use in VS purposes and thus optimized for docking the largest amounts of active compounds and small amounts of decoys.

The prediction of helix ranges yields mediocre results and is not an optimal method for GPCR homology model construction for VS purposes. The addition of intra- and extracellular loops may be beneficial while constructing models on closely related templates, however when more distant proteins are considered, their presence may lower the quality of the resulting models and is not necessary to obtain high quality VS-targeted structures.

To increase the validity of homology modeling, a set of various homologues should be considered, preferably with an extensive library of ligand activities. Such an approach should be routinely implemented, as while it is more time and resource consuming, the current computational capabilities combined with proper scripting should handle it with relative ease, leading to more optimal final results.

ASSOCIATED CONTENT

Supporting Information

Phylogenetic trees for all targets and templates, visualization of alignments used in this study, additional data concerning RMSD and sequence identity of models, the ROC curve of the beta2-AR-based model of 5-HT_6R, and SiteMap calculations of models' binding sites, as well as the explanation to the number of models built in the study. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: bojarski@if-pan.krakow.pl.

Present Address

[§]Institute of Structural and Molecular Biology, University College London, Gower Street, London, WC1E 6BT, United Kingdom

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was partially supported by the project UDA-POIG.01.03.01-12-063/09-00 cofinanced by European Union from the European Fund of Regional Development (EFRD), http://www.prokog.pl

ABBREVIATIONS

A2A, Adenosine A2A receptor; beta1, Beta1 adrenergic receptor; beta2, Beta2 adrenergic receptor; CXCR4, C-X-C chemokine receptor type 4; D3, Dopamine receptor 3; H1, Histamine 1 receptor Muscarine receptor 2 Muscarine receptor 3; 5-HT_6R, 5-HT_6 receptor; 5-HT_7R, 5-HT_7 receptor; 5-HT_{1A}R, 5-HT_{1A} receptor; 5-HT_{2A}R, 5-HT_{2A} receptor; 5-HT_{1B}R, 5-HT_{1B} receptor; 5-HT_{2B}R, 5-HT_{2B} receptor; VS, Virtual Screening; GPCR, G-Protein Coupled Receptor

REFERENCES

(1) Wu, T. T.; Fitch, W. M.; Margoliash, E. The Information Content of Protein Amino Acid Sequences. *Annu. Rev. Biochem.* **1974**, *43*, 539−566.

(2) Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **1986**, *5*, 823−826.

(3) Kryshtafovych, A.; Fidelis, K.; Moult, J. CASP9 results compared to those of previous CASP experiments. *Proteins* **2011**, *79* (Suppl 1), 196−207.

(4) Qian, B.; Raman, S.; Das, R.; Bradley, P.; McCoy, A. J.; Read, R. J.; Baker, D. High-resolution structure prediction and the crystallographic phase problem. *Nature* **2007**, *450*, 259−264.

(5) Mariani, V.; Kiefer, F.; Schmidt, T.; Haas, J.; Schwede, T. Assessment of template based protein structure predictions in CASP9. *Proteins* **2011**, *79* (Suppl 1), 37−58.

(6) Kinch, L.; Yong Shi, S.; Cong, Q.; Cheng, H.; Liao, Y.; Grishin, N. V. CASP9 assessment of free modeling target predictions. *Proteins* **2011**, *79* (Suppl 1), 59−73.

(7) Sandal, M.; Duy, T. P.; Cona, M.; Zung, H.; Carloni, P.; Musiani, F.; Giorgetti, A. GOMoDo: A GPCRs online modeling and docking webserver. *PLoS One* **2013**, *8* (9), e74092.

(8) Latek, D.; Pasznik, P.; Carlomagno, T.; Filipek, S. Towards improved quality of GPCR models by usage of multiple templates and profile-profile comparison. *PLoS One* **2013**, *8* (2), e56742.

(9) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* **2006**, *5*, 993−996.

(10) Xu, H. E.; Xiao, R. A new era for GPCR research: structures, biology and drug discovery. *Acta Pharmacol. Sin.* **2012**, *33*, 289−290.

(11) Grigorieff, N.; Ceska, T. A.; Downing, K. H.; Baldwin, J. M.; Henderson, R. Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J. Mol. Biol.* **1996**, *259*, 393−421.

(12) Palczewski, K. Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor. *Science* **2000**, *289*, 739−745.

(13) Hanson, M. A.; Roth, C. B.; Jo, E.; Griffith, M. T.; Scott, F. L.; Reinhart, G.; Desale, H.; Clemons, B.; Cahalan, S. M.; Schuerer, S. C.; Sanna, M. G.; Han, G. W.; Kuhn, P.; Rosen, H.; Stevens, R. C. Crystal structure of a lipid G protein-coupled receptor. *Science* **2012**, *335*, 851−855.

(14) Wu, H.; Wacker, D.; Mileni, M.; Katritch, V.; Han, G. W.; Vardy, E.; Liu, W.; Thompson, A. A.; Huang, X.-P.; Carroll, F. I.; Mascarella, S. W.; Westkaemper, R. B.; Mosier, P. D.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Structure of the human κ-opioid receptor in complex with JDTic. *Nature* **2012**, *485*, 327−332.

(15) Thompson, A. A.; Liu, W.; Chun, E.; Katritch, V.; Wu, H.; Vardy, E.; Huang, X.-P.; Trapella, C.; Guerrini, R.; Calo, G.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* **2012**, *485*, 395−399.

(16) Zhang, C.; Srinivasan, Y.; Arlow, D. H.; Fung, J. J.; Palmer, D.; Zheng, Y.; Green, H. F.; Pandey, A.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Coughlin, S. R.; Kobilka, B. K. High-resolution crystal structure of human protease-activated receptor 1. *Nature* **2012**, *492*, 387−392.

(17) Kruse, A. C.; Hu, J.; Pan, A. C.; Arlow, D. H.; Rosenbaum, D. M.; Rosemond, E.; Green, H. F.; Liu, T.; Chae, P. S.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Wess, J.; Kobilka, B. K. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **2012**, *482*, 552−556.

(18) Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; Kobayashi, T. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **2012**, *482*, 547−551.

(19) Wacker, D.; Wang, C.; Katritch, V.; Han, G. W.; Huang, X.-P.; Vardy, E.; McCorvy, J. D.; Jiang, Y.; Chu, M.; Siu, F. Y.; Liu, W.; Xu, H. E.; Cherezov, V.; Roth, B. L.; Stevens, R. C. Structural Features for Functional Selectivity at Serotonin Receptors. *Science* **2013**, *340*, 615−619.
(20) Wang, C.; Jiang, Y.; Ma, J.; Wu, H.; Wacker, D.; Katritch, V.; Han, G. W.; Liu, W.; Huang, X.-P.; Vardy, E.; McCorvy, J. D.; Gao, X.; Zhou, E. X.; Melcher, K.; Zhang, C.; Bai, F.; Yang, H.; Yang, L.; Jiang, H.; Roth, B. L.; Cherezov, V.; Stevens, R. C.; Xu, H. E. Structural Basis for Molecular Recognition at Serotonin Receptors. *Science* **2013**, *340*, 610−614.
(21) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. a; Rasmussen, S. 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 beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258−1265.
(22) 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.
(23) Chien, E. Y. T.; Liu, W.; Zhao, Q.; Katritch, V.; Han, G. W.; 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.
(24) 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.
(25) Shimamura, T.; Shiroishi, M.; Weyand, S.; Tsujimoto, H.; Winter, G.; Katritch, V.; Abagyan, R.; Cherezov, V.; Liu, W.; Han, G. W.; Kobayashi, T.; Stevens, R. C.; Iwata, S. Structure of the human histamine H1 receptor complex with doxepin. *Nature* **2011**, *475*, 65−70.
(26) Park, S. H.; Das, B. B.; Casagrande, F.; Tian, Y.; Nothnagel, H. J.; Chu, M.; Kiefer, H.; Maier, K.; De Angelis, A. A.; Marassi, F. M.; Opella, S. J. Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* **2012**, *491*, 779−783.
(27) Kennett, G. A.; Dourish, C. T.; Curzon, G. Antidepressant-like action of 5-HT1A agonists and conventional antidepressants in an animal model of depression. *Eur. J. Pharmacol.* **1987**, *134*, 265−274.
(28) Sato, K.; Yoshida, K.; Takahashi, H.; Ito, K.; Kamata, M.; Higuchi, H.; Shimizu, T.; Itoh, K.; Inoue, K.; Tezuka, T.; Suzuki, T.; Ohkubo, T.; Sugawara, K.; Otani, K. Association between −1438G/A promoter polymorphism in the 5-HT(2A) receptor gene and fluvoxamine response in Japanese patients with major depressive disorder. *Neuropsychobiology* **2002**, *48*, 136−140.
(29) Schechter, L. E.; Lin, Q.; Smith, D. L.; Zhang, G.; Shan, Q.; Platt, B.; Brandt, M. R.; Dawson, L. A.; Cole, D.; Bernotas, R.; Robichaud, A.; Rosenzweig-Lipson, S.; Beyer, C. E. Neuropharmacological profile of novel and selective 5-HT6 receptor agonists: WAY-181187 and WAY-208466. *Neuropsychopharmacology* **2008**, *33*, 1323−1335.
(30) Hedlund, P. B.; Sutcliffe, J. G. Functional, molecular and pharmacological advances in 5-HT7 receptor research. *Trends Pharmacol. Sci.* **2004**, *25*, 481−486.
(31) Ruat, M.; Traiffort, E.; Arrang, J. M.; Tardivel-Lacombe, J.; Diaz, J.; Leurs, R.; Schwartz, J. C. A novel rat serotonin (5-HT6) receptor: molecular cloning, localization and stimulation of cAMP accumulation. *Biochem. Biophys. Res. Commun.* **1993**, *193*, 268−276.
(32) Kohen, R.; Metcalf, M. A.; Khan, N.; Druck, T.; Huebner, K.; Lachowicz, J. E.; Meltzer, H. Y.; Sibley, D. R.; Roth, B. L.; Hamblin, M. W. Cloning, characterization, and chromosomal localization of a human 5-HT6 serotonin receptor. *J. Neurochem.* **1996**, *66*, 47−56.
(33) Perez-García, G.; Meneses, A. Oral administration of the 5-HT6 receptor antagonists SB-357134 and SB-399885 improves memory formation in an autoshaping learning task. *Pharmacol., Biochem. Behav.* **2005**, *81*, 673−682.
(34) King, M. V.; Marsden, C.; Fone, K. C. A role for the 5-HT(1A), 5-HT(4) and 5-HT(6) receptors in learning and memory. *Trends Pharmacol. Sci.* **2008**, *29*, 482−92.
(35) Barnes, N. M.; Sharp, T. A review of central 5-HT receptors and their function. *Neuropharmacology* **1999**, *38*, 1083−1152.
(36) Gaddum, J. H.; Picarelli, Z. P. Two kinds of tryptamine receptor. *Br. J. Pharmacol.* **1957**, *12*, 323−328.
(37) De Almeida, J.; Mengod, G. *J.* Serotonin 1A receptors in human and monkey prefrontal cortex are mainly expressed in pyramidal neurons and in a GABAergic interneuron subpopulation: implications for schizophrenia and its treatment. *Neurochemistry* **2008**, *107*, 488−496.
(38) Van de Kar, L. D.; Javed, A.; Zhang, Y.; Serres, F.; Raap, D. K.; Gray, T. S. *J.* 5-HT2A receptors stimulate ACTH, corticosterone, oxytocin, renin, and prolactin release and activate hypothalamic CRF and oxytocin-expressing cells. *Neuroscience* **2001**, *21*, 3572−3579.
(39) Feniuk, W.; Humphrey, P. P.; Watts, A. D. 5-Hydroxytryptamine-induced relaxation of isolated mammalian smooth muscle. *Eur. J. Pharmacol.* **1983**, *96*, 71−78.
(40) Bard, J. A.; Zgombick, J.; Adham, N.; Vaysse, P.; Branchek, T. A.; Weinshank, R. L. Cloning of a novel human serotonin receptor (5-HT7) positively linked to adenylate cyclase. *J. Biol. Chem.* **1993**, *268*, 23422−23426.
(41) Homan, E. J.; Wikström, H. V.; Grol, C. J. Molecular modeling of the dopamine D2 and serotonin 5-HT1A receptor binding modes of the enantiomers of 5-OMe-BPAT. *Bioorg. Med. Chem.* **1999**, *7*, 1805−1820.
(42) Hirst, W. D.; Abrahamsen, B.; Blaney, F. E.; Calver, A. R.; Aloj, L.; Price, G. W.; Medhurst, A. D. Differences in the Central Nervous System Distribution and Pharmacology of the Mouse 5-Hydroxytryptamine-6 Receptor Compared with Rat and Human Receptors Investigated by Radioligand Binding, Site-Directed Mutagenesis, and Molecular Modeling. *Mol. Pharmacol.* **2003**, *64*, 1295−1308.
(43) Lepailleur, A.; Bureau, R.; Paillet-Loilier, M.; Fabis, F.; Saettel, N.; Lemaître, S.; Dauphin, F.; Lesnard, A.; Lancelot, J.-C.; Rault, S. Molecular modeling studies focused on 5-HT7 versus 5-HT1A selectivity. Discovery of novel phenylpyrrole derivatives with high affinity for 5-HT7 receptors. *J. Chem. Inf. Model.* **2005**, *45*, 1075−1081.
(44) Nowak, M.; Kołaczkowski, M.; Pawłowski, M.; Bojarski, A. J. Homology modeling of the serotonin 5-HT1A receptor using automated docking of bioactive compounds with defined geometry. *J. Med. Chem.* **2006**, *49*, 205−214.
(45) Mosier, P. D.; Kolanos, R.; Roth, B. L.; Glennon, R. A. Binding of Serotonin and N 1 -Benzenesulfonyltryptamine-Related Analogs at Human 5-HT 6 Serotonin Receptors: Receptor Modeling Studies. *J. Med. Chem.* **2008**, *51*, 603−611.
(46) Pecic, S.; Makkar, P.; Chaudhary, S.; Reddy, B. V.; Navarro, H. A.; Harding, W. W. Affinity of aporphines for the human 5-HT2A receptor: insights from homology modeling and molecular docking studies. *Bioorg. Med. Chem.* **2010**, *18*, 5562−5575.
(47) Badarau, E.; Bugno, R.; Suzenet, F.; Bojarski, A. J.; Finaru, A.-L.; Guillaumet, G. SAR studies on new bis-aryls 5-HT7 ligands: Synthesis and molecular modeling. *Bioorg. Med. Chem.* **2010**, *18*, 1958−1967.
(48) De la Fuente, T.; Martín-Fontecha, M.; Sallander, J.; Benhamú, B.; Campillo, M.; Medina, R. a; Pellissier, L. P.; Claeysen, S.; Dumuis, A.; Pardo, L.; López-Rodríguez, M. L. Benzimidazole derivatives as new serotonin 5-HT6 receptor antagonists. Molecular mechanisms of receptor inactivation. *J. Med. Chem.* **2010**, *53*, 1357−1369.
(49) Hao, M.; Li, Y.; Li, H.; Zhang, S. Investigation of the Structure Requirement for 5-HT(6) Binding Affinity of Arylsulfonyl Derivatives: A Computational Study. *Int. J. Mol. Sci.* **2011**, *12*, 5011−5030.
(50) van Loevezijn, A.; Venhorst, J.; Wouter, I.; Bakker, I.; de Korte, C.; de Looff, W.; Verhoog, S.; van Wees, J.; van Hoeve, M.; van de Woestijne, R.; van der Neut, M. A. W.; Borst, A. J. M.; van Dongen, M. J. P.; Kruse, C. G. N0 -(Arylsulfonyl)pyrazoline-1-carboxamidines as Novel, Neutral 5-Hydroxytryptamine 6 Receptor (5-HT6R) Antagonists with Unique Structural Features. *J. Med. Chem.* **2011**, *54*, 7030−7054.

(51) Yap, B. K.; Buckle, M. J. C.; Doughty, S. W. Homology modeling of the human 5-HT(1A), 5-HT (2A), D1, and D2 receptors: model refinement with molecularaz dynamics simulations and docking evaluation. *J. Mol. Model.* **2012**, *18*, 3639−3655.
(52) 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 beta1-adrenergic G-protein-coupled receptor. *Nature* **2008**, *454*, 486−491.
(53) *Discovery Studio Modeling Environment*, Release 3.5; Accelrys Software Inc.: San Diego, CA, 2012.
(54) Ballesteros, J.; Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **1995**, *25*, 366−428.
(55) Harris, R. N.; Stabler, R. S.; Repke, D. B.; Kress, J. M.; Walker, K. A.; Martin, R. S.; Brothers, J. M.; Ilnicka, M.; Lee, S. W.; Mirzadegan, T. Highly potent, non-basic 5-HT6 ligands. Site mutagenesis evidence for a second binding mode at 5-HT6 for antagonism. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3436−3440.
(56) Purohit, A.; Herrick-Davis, K.; Teitler, M. Creation, expression, and characterization of a constitutively active mutant of the human serotonin 5-HT6 receptor. *Synapse* **2003**, *47*, 218−224.
(57) Amico, M.; Finelli, M.; Rossi, I.; Zauli, A.; Elofsson, A.; Viklund, H.; von Heijne, G.; Jones, D.; Krogh, A.; Fariselli, P.; P. L, M.; Casadio, R. PONGO: a web server for multiple predictions of all-alpha transmembrane proteins. *Nucleic Acids Res.* **2006**, *34*, 169−172.
(58) Jones, D. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292*, 195−202.
(59) Buchan, D. W. A.; Minneci, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **2013**, 349−357.
(60) Sigrist, C. J. A.; Cerutti, L.; Hulo, N.; Gattiker, A.; Falquet, L.; Pagni, M.; Bairoch, A. PROSITE: a documented database using patterns and profiles as motif descriptors. *Br. Bioinform.* **2002**, *3*, 265−274.
(61) Néron, B.; Ménager, H.; Maufrais, C.; Joly, N.; Maupetit, J.; Letort, S.; Carrere, S.; Tuffery, P.; Letondal, C. Mobyle: a new full web bioinformatics framework. *Bioinformatics* **2009**, *25*, 3005−3011.
(62) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(63) *Glide*, version 5.7; Schrödinger, LLC: New York, 2011.
(64) ChEMBL_11, ChEMBL-EBI. http://www.ebi.ac.uk/chembldb/index.php (accessed September 1, 2011).
(65) *LigPrep*, version 2.5; Schrödinger, LLC: New York, 2011.
(66) *Canvas*, version 1.4; Schrödinger, LLC: New York, 2011.
(67) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.
(68) *SiteMap*, version 2.6; Schrödinger, LLC: New York, 2012.
(69) Beuming, T.; Sherman, W. Current Assessment of Docking into GPCR Crystal Structures and Homology Models: Successes, Challenges, and Guidelines. *J. Chem. Inf. Model.* **201**, *52*, 3263−3277.
(70) López-Rodríguez, M. L.; Benhamú, B.; de la Fuente, T.; Sanz, A.; Pardo, L.; Campillo, M. A three-dimensional pharmacophore model for 5-hydroxytryptamine6 (5-HT6) receptor antagonists. *J. Med. Chem.* **2005**, *48*, 4216−4219.
(71) Granier, S. Structure of mu and delta opioid receptors. *Med. Sci. (Paris)* **2012**, *28*, 870−875.
(72) Kufareva, I.; Rueda, M.; Katritch, V.; Stevens, R. C. Article Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment. *Structure* **2011**, *19*, 1108−1126.