JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

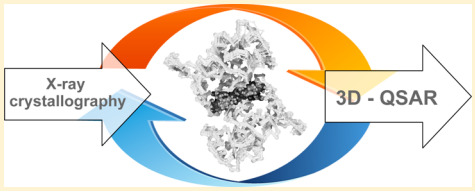
X-ray Crystallographic Structures as a Source of Ligand Alignment in 3D-QSAR

Rafał D. Urniaż and Krzysztof Jóźwiak*

Medical University of Lublin, Laboratory of Medicinal Chemistry and Neuroengineering, Chodźki 4a Street, 20-093 Lublin, Poland

Supporting Information

ABSTRACT: Three-dimensional quantitative structure−activity relationships (3D-QSAR) analyses are methods correlating a pharmacological property with a mathematical representation of a molecular property distribution around three-dimensional molecular models for a set of congeners. 3D-QSAR methods are known to be highly sensitive to ligand conformation and alignment method. The current study collects 32 unique positions of congeneric ligands co-crystallized with the binding domain of AMPA receptors and aligns them using protein coordinates. Thus, it allows for a unique opportunity to consider a ligands' orientation aligned by their mode of binding in a native molecular target. Comparative molecular field analysis (CoMFA) models were generated for this alignment and compared with the results of analogous modeling using standard structure-based alignment or obtained in docking simulations of the ligands' molecules. In comparison with classically derived models, the model based on X-ray crystallographic studies showed much better performance and statistical significance. Although the 3D-QSAR methods are mainly employed when crystallographic information is limited, the current study underscores the importance that the selection of inappropriate molecular conformations and alignment methods can lead to generation of erroneous models and false conclusions.



1. INTRODUCTION

Three-dimensional quantitative structure−activity relationship (3D-QSAR) methods perform an important role in drug discovery and ligand-based design. These methods are used to establish a relationship between molecular properties distributed around molecular models and observed biological effects of a series of congeneric compounds.[1] One of the widely used 3D-QSAR methods is comparative molecular field analysis (CoMFA). It adopts statistical procedures to correlate molecular features, such as steric and electrostatic properties of molecules with their biological activities.[2] The fields are calculated by testing attractive and repulsive interactions between a probe atom and a collection of aligned molecules. The interactions are evaluated for each molecule at regularly spaced intervals on a virtual grid surrounding the molecules. The result is an interactive contour map graphic with indication of statistically significant fields with impact on modeled activity. Such an approach has essentially two main advantages: its ability to represent the information in 3D format, highlighting regions of the molecules with positive and negative effects associated with interactions, and its ability to predict the unknown biological activity of a new compound based on the calculated model.[2] Over the past few decades, CoMFA methods have become widespread in regions of both industrial and academic research. Such approaches are providing not only prediction of specific properties of new compounds but also help elucidate the possible molecular mechanism of the receptor−ligand interactions.[3] Although a high number of QSAR variants exist, they all have general imperfections.[4] As previously reported, one of the most critical steps are a proper structural alignment and modeled conformation of a set of tested compounds. Cramer and his co-workers[2] annotated that "bio-active conformation" and "alignment rules" are major dilemmas in QSAR application. Clark,[5,6] Patterson,[7] and others[8−10] noticed that despite the alignment and conformation criterion, an important aspect is the compounds similarity in context to their interactions with its molecular target. Hence, in order to consider QSAR approaches, molecules have to share the same or similar modes of binding. Despite significant advances in computer-based similarity searching and similarity-based virtual screening and the use of the fundamental similarity principle, which states that similar structures have similar activities,[3,4,8] it is still a challenge to devise meaningful concepts and uses of structural similarity.[9−12] Moreover, studied molecules in the data set should have similar mechanisms of action and common pharmacophores.[3−5] Chemical similarity does not necessarily ensure a common mechanism of action for all congeners.[4,8] Thus, it is generally accepted that 3D-QSAR models are alignment sensitive and selection of the appropriate overlay function is a "bottleneck" of the QSAR methods.[4−6] Additionally, the fundamental assumption of the pharmacophore theory that compounds with the same chemical scaffold share similar orientation in the protein binding site is not always fulfilled.[6−8] For example, IDRA-21 and cyclothiazide are two compounds that were synthesized in the 1980s,[13,14] and based on the pharmacophore theory, the two ligands were thought to have a similar orientation for the benzothiadiazine ring systems interaction with the ligand binding domain. This incorrect assumption was propagated until their protein complex X-ray crystal structure

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

appeared.[15] An overlay of IDRA-21 and cyclothiazide, co-crystallized with the ligand binding domain of the AMPA receptor is shown in Figure 1. A comparison of the ligand–

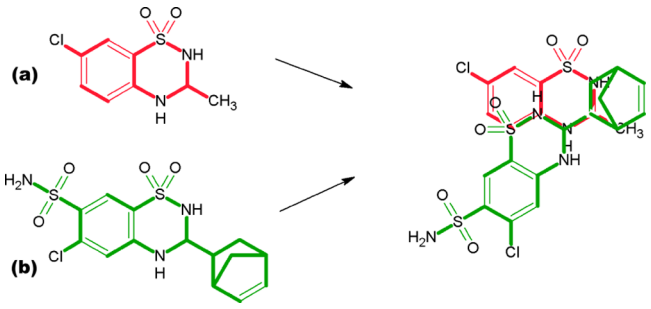


Figure 1. Schematic overlay of positions of two ligand: (a) IDRA-21 (PDB ID: 3IL1) and (b) cyclothiazide (PDB ID: 1LBC) co-crystallized with the ligand binding domain of the AMPA receptor.

receptor complexes obtained in crystallography shows that the two molecules assume significantly distinct orientations, Figure 1; this illustrates the possibility that structurally similar compounds may assume different modes of binding.

The misstatement of the pharmacophore theory assumption in terms of AMPA receptor is not an isolated case and was reported previously.[16] The assumption could propagate some improper or even impossible orientations of ligands within the binding domain. Considering this problem, the authors attempt to confront the prediction abilities of different types of alignment in 3D-QSAR and compare it to the alignment obtained from a set of crystal structures of ligand–receptor complexes. The AMPA receptor co-crystallized with a high number of congeneric compounds was selected for the study. The receptor belongs to the family of ionotropic glutamate receptors and are widespread in the nervous system. The AMPA receptors appear to be critical in mediating synaptic plasticity and long-term potentiation[17] that may control higher-order processes such as learning and memory. Thus, allosteric positive modulation of the receptor is one of the promising directions for Parkinson's and Alzheimer's disease treatments.[17] Therefore, the ligand binding domain of the receptor was co-crystallized with almost 40 different allosteric modulators. It allowed a unique opportunity to evaluate the native orientation and bioactive conformation of a high number of congeneric compounds with similar binding domain conditions. The receptor alignment was obtained as an overlay of the collected crystal structures (co-crystallized with ligands) by the protein scaffold. The main advantage of this alignment method is that it was obtained only by the overlay of crystal structures of ligand–receptor complexes and neither molecular docking nor receptor-based alignment method was applied. The high-resolution X-ray structures[15] provide a rare opportunity to sample ligand orientations and modes of binding for a large set of congeneric ligands.

The AMPA receptor structure is composed of four types of subunits (GluR1–4), which combine to form heterotetramers or according to more recent description a symmetrical system of "dimer of dimers".[18] The ligand binding domain (LBD) is located at the interface of two proteins forming a dimer. In the case of AMPA receptor activation, the conformational changes of the LBD allow the entrance of ligands, such as positive allosteric modulators. The AMPA receptor LBD contains two symmetrical ligand binding subdomains. It allows two different binding modes for positive modulators interacting with these sites: (A) One molecule of bulky nonsymmetrical ligand may bind to (parts of) both regions at the same time. (B) Each binding pocket may adopt one molecule of a small symmetrical ligand.[13] In case of (B), two ligand molecules occupy two symmetrical sites in the X-ray structures. Crystal structures assume reliable orientation of ligands in the binding domain. It is generally accepted that construction of bioactive conformation for a set of molecules can be attained by crystallographic studies. However, the process is expensive and time consuming for a high number of ligands co-crystallized with protein structures. Thus, the presented study compares crystallographic receptor alignment with different types of alignment and is evaluated with respect to the QSAR application. Other investigators suggest applying diverse computational approaches, including molecular docking to predict the bioactive conformation.[5,10,11] Additionally standard structural and pharmacophore alignment based on similarities of a set of ligands and molecular docking were compared to the alignment obtained from the analysis of a set of crystal structures of ligand–receptor complexes. The comparative molecular field analysis (CoMFA) was used to assess predictability of alignment methods.[20,21] Comparison of docking study with crystal structures may indicate that it is a method good enough to reproduce the receptor alignment system in a in silico environment. It can have an impact for future application as a model to activity prediction or to observe a changing of molecular fields in case of new structure design. To our knowledge, this type of alignment containing crystallographic information for all modeled molecules has not been reported previously.[13−15,17−19,22−31]

2. MATERIALS AND METHODS

2.1. Structure Collection. Currently, significant number of crystal structures of AMPAr LBD are deposited in the Protein Data Bank (PDB), most of them were co-crystallized with ligands. The work collects unique and high-resolution structures of the LBD co-crystallized with positive allosteric modulators for which actual and precise activity data were determined (Table S1, Supporting Information). Thirty-two of them were unique and contained 42 different poses of ligands extracted from binding pockets. Structures originally extracted from the crystal structure were verified in terms of chemical structure and type of bonds. All ligand structures were verified in Sybyl-X 2.0,[32] where proper atom hybridization and bond type were assigned. Crystal structures assume reliable orientation of ligands in the ligand binding domain. The conformation of the ligands obtained from the crystallographic studies is also called bioactive conformation due to ligands interactions with its binding pocket in the protein.[2,21] It gives a unique opportunity to compare it to different types of alignment. Thus, the authors considered the orientation of ligands extracted from crystal structure protein scaffold alignment as reference alignment in terms of QSAR conditions.

2.2. Biological Activity. pEC_{50} values used as input experimental data in model constructions were collected from the literature.[13−15,17−19,22−31] All activities were determined using patch-clamp assays, which is a common technique to measure the efficacy ligands on AMPA receptors.[31] Obviously, the data were determined in several laboratories using varied experimental protocols, which may question their comparability; however, some internal standardization can be applied in this case. As most publications report the EC_{50} values for

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

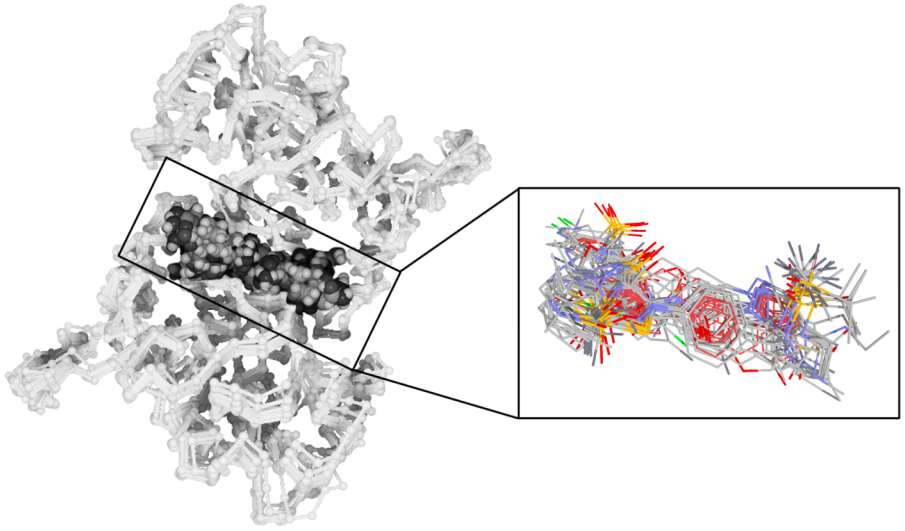


Figure 2. AMPA ligand binding domain of 32 crystal structures (rendered as α carbons) with ligands (rendered as balls) aligned using the MUSTANG algorithm. Inset: details of crystallographic alignment (RA) of all ligands rendered in stick mode.

aniracetam[15] (PDB ID: 2AL4) or cyclothiazide[13] (PDB ID: 1LBC), these compounds can be used as standards to compare values between laboratories. Although the activity of compounds was measured in different laboratories, the value obtained for aniracetam or cyclothiazide were similar.

2.3. Alignments. The main aim of the presented study was to give a retrospective look at the importance of correct alignment and evaluate the new potential directions in drug development based on the molecular fields obtained from the crystal structure receptor alignment. Thus, 11 standard types of alignments were employed: structural alignments (manual structural alignment, flexible ligand alignment), pharamcophore alignments (GASP, GALAHAD, LigandScout, PharmaGist), and receptor-based alignments (docking in Molegro Virtual Docker, Schrödinger Glide, Sybyl-X Surflex-Dock, AutoDock). Additionally receptor alignment was considered as its adaptation and further application. The obtained receptor alignment is a unique opportunity to simulate the native orientation and bioactive conformation of a number of congeneric compounds within LBD.

2.3.1. Receptor Alignment from X-ray Crystallographic Structures (RA). The X-ray structures of positive modulator molecule(s) co-crystallized with LBD of AMPA receptor were obtained from PDB (Table S1, Supporting Information). It should be pointed out that the presented alignment was obtained by the overlay of crystal structures only, and neither molecular docking nor other methods were applied for ligands. In the preparation step, all of LBD crystal structures were aligned using a multiple structural alignment algorithm (MUSTANG) implemented in Yasara 11.6.16.[33] The program constructs a multiple alignment using the spatial information of the α carbon atoms of the protein in the set. MUSTANG reports the multiple sequence alignment and the corresponding superposition of structures. As shown in Figure 2, proteins were overlaid with average RMSD value around 0.40 Å. As a result, unique ligand poses, 22 representing mode of binding (A) and 10 representing mode of binding (B) originally overlaid in crystal structures, comprised the receptor alignment group (RA) (for structures, see Table S1, Supporting Information).

2.3.2. Structural–Pharmacophore Alignment. The pharmacophore alignment is one of the automatic molecular alignment generation methods. In contrast to molecular docking, the method is widely employed in cases when the target structure is not available. It determines the pharmacophore features of the ligands and uses them to align the molecules. The problem of conformational diversity and flexibility may be solved by this type of alignment, and it is widely applied in construction of 3D-QSAR models.[34−37] In the current study, six standard types of alignment methods (GASP, GALAHAD, LigandScout, PharmaGist, flexible ligand alignment, and structural alignment) obtained manually were employed as suggested by other investigators.[40−45] The first four algorithms failed during calculations, and obtained alignment was evidently unacceptable. Despite this fact, alignments were tested; however, the models could not be calculated with a "lack of variation" alert. This problem may occur due to too structurally diverse sets, as was reported elsewhere.[4,7] The methods giving satisfactory results proved to be a flexible ligand alignment (FA) and structural alignment (SA) obtained manually. The alignments were prepared as follows.

FA. Flexible ligand alignment was performed in Schrödinger Maestro 9.3.4. The template molecule (ligand number 3 in Table S1, Supporting Information; PDB ID: 2XHD) was indicated by Schrödinger Strike 2.1 as a scaffold shared by most of molecules in the group. Alignment was generated automatically.

SA. The structural alignment was obtained in the Sybyl-X 2.0 program[32] and defined as structural alignment group (SA). It was a modification of well-known structural alignment based on a molecule scaffold[41] additionally supported by the pharmacophore similarity map suggesting the overlaying queue. Studied cohort of compounds (built as described in section 2.3.3) was primary characterized by similarity map (HQSAR fragment count, default settings)[32] shown in Figure S1 of the Supporting Information. In the applied method, the "seed" molecule has been appointed, assuming the principle "succeeding molecule is overlaying on previous one containing the most similar structure" was restricted. Analysis of the structure similarity map (HQSAR fragment counts) indicated the structure of aniracetam (compound number 2, Table S1, Supporting Information) as a scaffold shared by most of ligands (Figure

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

S1, Supporting Information). The orientation of aniracetam in the crystal structure was taken as a reference position ("seed" molecule) to determine the manual structural alignment. Starting from the "seed" molecule, successive molecules (with high similarity score to previous molecule) have been overlaid as follows (numbers consistent with Table S1, Supporting Information): (a) molecules 1, 3, 4, 5, 6, 7, 8, 9, 10, and 11 on structure number 2, (b) molecule 13 on structure number 6, (c) molecules 12, 14, 15, 16, and 17 on structure number 13, (d) molecule 18 on structure number 17, (e) molecule 19 on structure number 18, (f) molecule 20 on structure number 19, (g) molecule 21 on structure number 18, (h) molecule 22 on structure number 21, (i) molecule 23 on structure number 1, (j) molecules 24, 27−32 on structure number 23, and (k) molecules 25, 26 on structure number 24.

2.3.3. Alignment Based on the Docking to the Receptor. Some investigators[34−37] suggested to employ docking to the receptor as an alternative approach. In this procedure, molecules are positioned to the same region of the active site and then an automatic alignment is generated. In the docking procedure, a ligand is positioned to the functional site of the target for determining the binding mode of the complex as well as generating the conformation of the ligand. It is reported as a successful approach for predicting the bioactive conformation of molecules and alignment derived from the receptor-based approach.[35−37] To generalize the observation, five types of docking procedures, blind (DA 1) and template docking (DA2) in Molegro Virtual Docker, Schrödinger Glide (DG), Sybyl-X Surflex-Dock (DS), and AutoDock (DD), were tested. The compounds were aligned by docking to the receptor as referred in suitable manuals.

Currently, the highest resolution structure of AMPAr LBD is PDB ID: 2XXH (resolution 1.50 Å); this model was used in further simulations. In the protein preparation step, the model was verified in terms of chemical structure and the type of bonds. All cofactors were removed. To evaluate the bioactive conformation prediction ability of the program (as suggested previously[10,16,21]) compounds were built de novo in Spartan'10, and energies were minimized by Hartree−Fock ab initio algorithm with 6-31G* basis set. The compounds were exported as mol2 files and docked into crystal structure as recommended by developers. If not marked, the default settings were applied; the calculations were parametrized as follows.

DA 1. Molegro Virtual Docker 5.0, blind docking to whole binding domain. MolDock Score [GRID] (searching and scoring algorithm); grid resolution 0.3 A. Search area: *x*, 61.74; *y*, 26.90; *z*, 45.05; radius, 17. Number of runs: 1000. Iterations: 1500 max population size 50. Max number of poses return: 10 (the best teen).

DA 2. Molegro Virtual Docker 5.0, docking with clarification the mode of binding. Due to molecules shared, the same scaffold may interact with LBD in a different mode (Figure 1). The clarification of mode of binding was determined as the precise location. The search area depends the mode of binding. Mode of binding A: MolDock Score [GRID] (searching and scoring algorithm); grid resolution 0.3 A. Search area: *x*, 61.99; *y*, 26.90; *z*, 41.05; radius, 15. Number of runs: 1000. Iterations: 1500 max population size 50. Max number of poses return: 10 (the best teen). Mode of binding B: MolDock Score [GRID] (searching and scoring algorithm); grid resolution 0.3 A. Search area: *x*, 61.74; *y*, 20.90; *z*, 45.05; radius: 11. Number of runs: 1000. Iterations: 1500 max population size 50. Max number of poses return: 10 (the lowest teen energies). The lowest energy ligand poses were obtained, and the ligand's best pose considering the most favorable score value and low internal energy was provided to further considerations. Results were evaluated by MolDock score, rerank score, and the protein−ligand interaction score from MolDock [GRID] option.

DG. Schrödinger Glide 5.8, docking procedure was prepared as described in the Glide 5.8 user manual.[38] All of the steps described in the procedure were successfully completed. Preparing the protein: structure was a preprocessed (additionally options Fill and Cap termini were selected) protein and was optimized and minimized under default conditions. Ligand structures for docking were optimized in LigPrep. Generating the Receptor Grid: On the basis of original ligand locations, supplied *X*, *Y*, *Z* coordinates: 30.0, 15.0, 0. Length of cell unit: 11 Å. Docking the ligands: The extra precision variant was selected, and number of runs was increased to 1000.

DS. Sybyl-X 2.0 Surflex-Dock, docking setup files were prepared as described in Surflex-Dock user manual.[32] All of the steps described in the procedure were successfully completed. The protein was prepared in the prepare protein structure module implemented in Sybyl-X 2.0, as the recommended ligand was extracted from the original file to molecular area. The protein structure was analyzed and fixed if necessary. On the basis of ligand location, the protomol file was generated. The ligands were docked under default conditions.

DD. The AutoDock 4, docking setup files were prepared as described in AutoDock user manual.[39] All of the steps described in the procedure were successfully completed using Autodock Tools. Preparing the protein was obtained using options "find missing atom and repairing them" and "add hydrogens". Ligands' charges and torsions were defined in preparing the ligand step, setup files were exported into *.pdbqt file. The search grid and grid maps were calculated running AutoGrid script, and particular regions of the protein (grid volume) were parametrized based on original ligand orientation. Number of runs in the docking parameter file *.dpf was increased to 1000, and the rest of the parameters were left as default.

Results were evaluated by scoring functions implemented in the programs. The lowest energy ligand poses were evaluated, and the ligand's best pose considering the most favorable score value and low internal energy was provided for further considerations.

2.4. 3D-QSAR Studies. CoMFA models were performed using the Sybyl-X 2.0 program package.[32] Because of the specific aim of the study, only atomic charges were calculated using the Gasteiger−Hückel algorithm implemented in the program. In order to not disturb the statistics, in all cases, models were generated with the same default settings of the program. The models have been quantified by the partial-least-squares (PLS) algorithm. To compare the models, cross-validation analysis was carried out using the leave-one-out (LOO) method. The R^2 and Q^2 were calculated according to eqs 1 and 2, respectively.[44,46]

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_c - y_a)^2}{\sum_{i=1}^{n} (y_m - y_a)^2} \quad (1)$$

In eq 1, the squared correlation coefficient is described as: y_a, actual value; y_c, average value of predicted values; y_m, average value of observed activities.

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 1. Predicted and Observed Activities Obtained from the CoMFA Models[a]

		predicted activity[c]							
PDB ID[b]	observed activity[c]	RA	FA	SA	DA1	DA2	DG	DS	DD
1LBC	**5.00**	**4.98**	**4.64**	**4.64**	**5.52**	**4.73**	**4.71**	**4.98**	**5.52**
2AL4	4.50	4.63	5.00	4.36	5.52	5.55	4.4	4.33	5.52
2AL5	3.70	3.73	4.81	4.38	5.47	5.14	3.92	4.34	4.86
2XHD	5.60	5.81	5.39	5.05	5.51	6.00	5.14	5.37	5.19
2XX7	5.00	5.19	5.77	5.19	5.54	5.78	4.78	5.58	5.54
2XX8	4.60	4.96	4.06	5.23	4.39	4.20	4.94	4.96	4.39
2XX9	6.00	6.19	5.53	5.14	5.58	6.00	6.18	5.37	5.58
2XXH	5.50	5.50	5.61	5.46	5.57	5.86	5.38	5.71	5.57
2XXI	6.00	6.15	5.59	5.23	5.60	6.07	5.76	5.4	5.60
3BBR	6.10	6.29	6.74	4.97	5.57	5.82	6.23	6.04	5.57
3H6U	4.58	4.57	4.99	4.64	5.54	4.04	4.22	5.79	5.54
3H6V	4.40	4.42	4.92	4.61	5.53	4.10	4.52	5.26	5.53
3H6W	4.30	4.52	4.36	4.59	5.50	3.92	4.15	5.58	5.50
3IJO	na	4.78	4.84	5.16	5.53	4.12	3.97	4.94	4.78
3IJX	na	5.02	4.59	4.72	5.47	4.13	4.08	3.95	5.02
3IL1	**3.30**	**3.26**	**3.92**	**4.28**	**5.45**	**3.93**	**3.87**	**3.69**	**5.45**
3ILT	na	4.24	4.41	5.05	5.48	4.32	4.2	4.44	5.48
3ILU	na	5.10	4.51	4.75	5.51	4.74	5.5	4.31	5.51
3KGC	5.89	6.01	6.06	5.63	5.55	5.69	6.17	5.69	5.55
3M3L	4.32	4.21	4.62	4.84	5.51	5.65	4.24	5.2	5.51
3O28	6.70	6.61	6.23	6.37	5.59	5.98	6.7	5.76	5.59
3O29	5.20	5.68	5.86	6.32	4.60	5.69	5.28	5.57	4.60
3O2A	6.40	6.44	6.29	6.42	5.58	5.72	6.5	7.21	5.58
3O6G	6.60	6.75	6.62	6.51	5.57	5.68	6.24	6.09	5.57
3O6H	6.10	6.27	4.98	6.39	5.54	5.83	6.29	6.05	5.54
3O6I	6.50	6.70	6.34	6.51	5.58	6.36	6.16	6.23	5.58
3PMV	6.30	6.42	6.52	6.40	5.59	5.94	6.49	5.84	5.59
3PMW	6.30	6.35	6	6.43	5.57	6.10	6.38	5.96	5.57
3PMX	6.00	5.91	5.62	6.46	5.54	5.91	6.17	5.41	5.54
3RN8	6.33	6.29	6.22	6.36	5.58	6.08	6.17	5.61	5.58
3RNN	7.19	7.01	6.58	6.51	5.54	6.19	7.17	5.96	5.20
3TDJ	4.49	4.64	3.89	4.42	5.47	4.47	4.72	3.71	5.47

[a]Calculated values for Idra-21 (PDB ID: 3IL1) and cyclothiazide (PDB ID: 1LBC) are bolded. [b]Ligand structures are available in Table S1 of the Supporting Information. [c]Negative logarithm of the half maximal effective concentration (pEC_{50}).

Table 2. Comparison of Statistics Characterizing Developed CoMFA Models

alignment type	N^a	R^2	F	α	Q^2	SE	$N_{PLS}{}^b$
RA	32	0.9792	1222	44.67	0.407	0.149	2
DG	32	0.9390	400.1	43.19	0.143	0.686	2
FA	32	0.7350	71.51	36.08	0.204	0.545	2
SA	32	0.7232	67.93	35.44	0.306	0.552	2
DA2	32	0.6144	41.43	31.27	0.159	0.649	2
DS	32	0.5627	33.59	29.15	0.124	0.769	2
DD	32	0.0921	2.639[c1]	5.31	0.091	0.818	1
DA1	32	0.0764	2.150[c2]	4.39	−0.144	0.984	1

[a]Number of compounds in the group. [b]Number of components in partial-least-squares (PLS) regression model, optimized automatically by the program. [c]Values is not statistically significant with p-value equal to 0.1164 for C1 and 0.1545 for C2; other values are significant with $p < 0.0001$.

$$Q^2 = 1 - \frac{\sum_{i=1}^{n} (y_a - y_p)^2}{\sum_{i=1}^{n} (y_a - y_m)^2} \quad (2)$$

In eq 2, the predictive squared correlation coefficient is described as: y_a, actual value; y_p, predictive value; y_m, average value of observed activities. The angle of the correlation curve (α) is described in eq 3.

$$\alpha = \arctan(\text{slope}) \quad (3)$$

Results were collected in Table 1. Additionally, the models were compared in case of determination of the CoMFA molecular fields (Figure S2, Supporting Information).

3. RESULTS AND DISCUSSION

For all the studied molecules, a variety of ligand alignments were applied in CoMFA modeling as described in the Materials and Methods section. Primarily, the models were statistically characterized by the most comparable parameters in QSAR study: squared correlation coefficient (R^2) and predictive

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

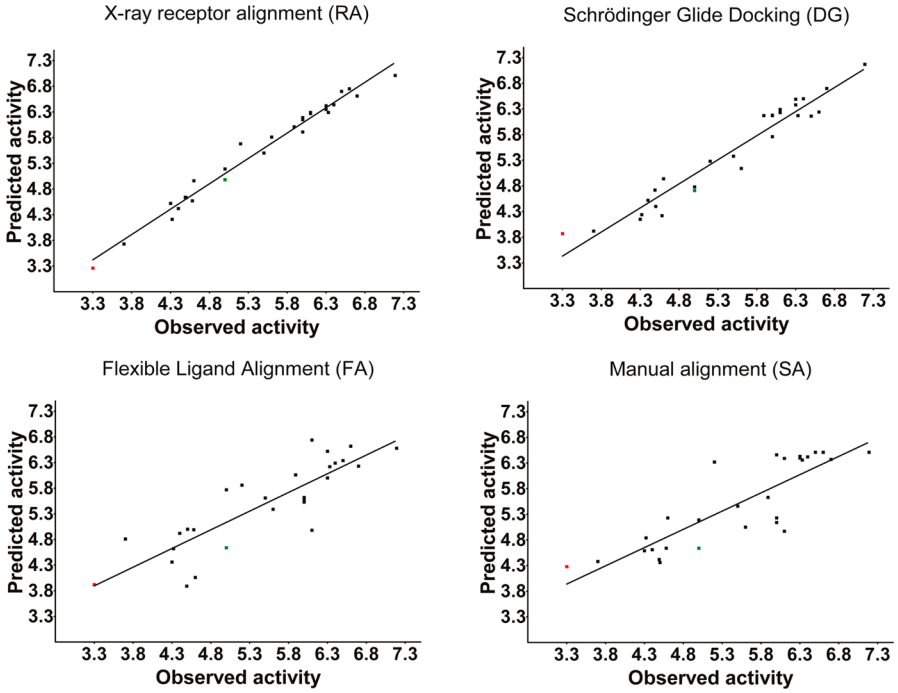


Figure 3. Graph of observed versus predicted activities of the presented models. For details, see description in the text. Red symbols, IDRA-21. Green symbols, cyclothiazide.

squared correlation coefficient (Q^2). The R^2 provides a measure of how well future outcomes are likely to be predicted by the model. In other words, it estimates the goodness-of-fit of the regression model derived from the training set. The Q^2, also known as predictive R^2, refers to the internal robustness of the QSAR model and how reliable the model is for prediction. The models additionally were statistically characterized by the F statistic (p value <0.0001), standard error (SE), and angle of the correlation curve (α). Statistical characterization of models was collected in Table 2.

The general aim is to compare predictability of CoMFA models generated for various alignment sets and evaluate these alignments in QSAR studies. Six of eight presented models (RA, DG, FA, SA, DA2, and DS) gave statistically significant results. In contrast to the first six presented models in Table 2, the models DD and DA1 were statistically insignificant with p-values = 0.1164 and 0.1545 in case of R^2 and 2.150 and 2.639 in case of F statistics, respectively. Moreover DA2 and DS models were weaker than the first four in Table 2. Thus, the results of DA2, DS, DD, and DA1 models were shown in Tables 1 and 2 but will not be discussed. Figure 3 presents correlations between the observed and predicted pEC_{50} values determined in models developed using the RA, DG, FA, and SA approaches. It can be observed that the fit of the linear regression trend line is much better for the RA variant. The squared correlation coefficient is 4.10%, 24.94%, and 26.14% higher for the RA group than for the DG, FA, and SA variants, respectively. The increase in statistical significance of models generated using the RA approach is even more evident when F values are compared. The F value for the RA model is around 3 times higher than for the DG model, 17 times higher than for the FA model, and 18 times higher than for the SA model, with 1222, 400.1, 71.51, 67.93, respectively.

Another parameter for assessing the interpolation of the QSAR model is the slope angle (α) value, which ideally should be around 45 degrees.[1,2] The RA variant is very close to this value. Also the predictive squared correlation coefficient (Q^2) and the standard errors (4.62, 3.66, 3.7 times lower comparing to DG, FA, SA) avail for RA. In our consideration of the problem, we applied CoMFA modeling to the system knowing that structurally similar ligands do not share the same mode of binding (as shown in the Figure 1). This allowed us to confront the quality of the obtained models using the RA, DG, FA, and SA approaches. It can be observed (Figure 3) that the quality of the alignments is strongly determined by two points. The points are corresponding to molecules presented in Figure 1 (IDRA-21, red symbols; cyclothiazide, green symbols). Deeper investigation of the models demonstrates that the removal of values for these two compounds slightly affects the R^2 value (0.0034 for RA, 0.0398 for FA, and 0.0147 for SA; Table 3),

Table 3. Change of Statistics in Terms of Removal of IDRA-21and Cyclothiazide from the Model (details in the text)

alignment type	R^{2a}	F^a
RA	0.9758	969.3
DG	0.9445	408.0
FA	0.6952	54.74
SA	0.7085	58.34

[a]Values are statistically significant with $p < 0.0001$.

although the statistical significance of correlation drops dramatically (20.7% for RA model, 23.45% for FA model, and 14.12% for SA model; Table 3). In case of the DG model, removal of values increase the R^2 value of 0.055 and the p-value of 2%.

Table 1 collects predicted activities of IDRA-21 and cyclothiazide marked in bold. It can be observed that predicted activity values are closer to observed ones only in case of RA variant.

Figure 4 presents the overlay of the IDRA-21 and cyclothiazide as they are in obtained alignments. The most

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

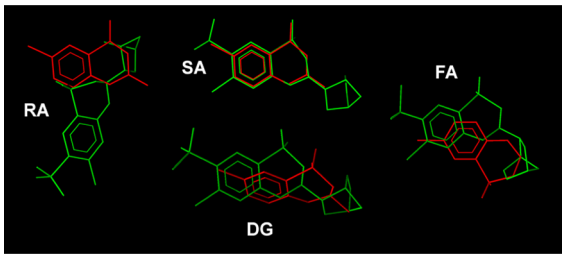


Figure 4. Comparison of positions of two ligands, IDRA-21 (red) and cyclothiazide (green), obtained by different types of alignment. RA: X-ray crystal structure receptor alignment. SA: Manual structural alignment. DG: Alignment obtained as a docking to the receptor binding side obtained in Schrödinger Glide software.

similar orientation (what has also reflection on the predicted activity) has a FA variant. Likewise, as shown in Figure 3, better correlation and fit to the trend line on the scatterplot can be observed in the RA group. The RA model contains more spatially diffused orientations, but their predictability is more accurate than the structural alignment or even docking.

The presented cohort of compounds can be illustrated by comparison of the orientations of the benzene ring systems (yellow line selection on the Figure 5). It is almost precisely

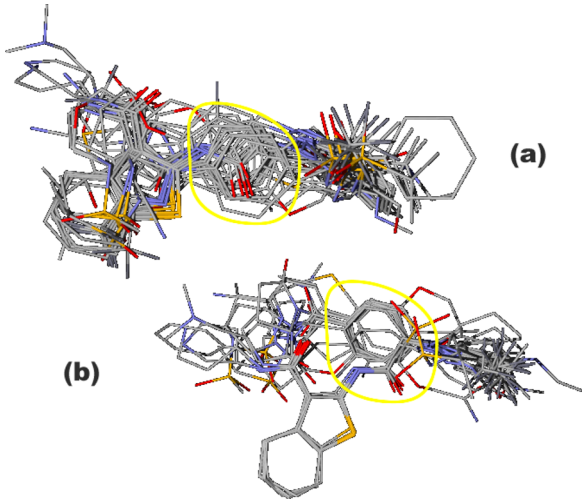


Figure 5. Molecules overlaid with two methods (a) RA and (b) SA. The benzene ring system of the molecules is the main scaffold for alignment in SA, while the same ring in RA (yellow line selection) shows some shift of ligands when adapting to the LBD.

aligned for all compounds in SA (Figure 5b), but the RA approach shows that this assumption is impossible to fulfill in the LBD binding conditions (Figure 5a). Positions of the benzene ring system seem to be more relevantly predicted by the docking study, especially in case of DG.

It is not typical to apply 3D-QSAR modeling for a high number (32) of compounds for which the high resolution crystal structures are available. It allows considering the possibility of reliable reconstruction of the ligand–receptor system applying common molecular modeling techniques and the applicability of the X-ray crystal structure alignment in QSAR studies. As observed, the application of the X-ray crystal structure alignment has potential benefits. The main one is that it composes the reliable arrangement of ligands in the binding domain, and the orientation is obtained in an experimental trial not as a mathematical calculation, so it assumes an actual ligand–receptor interaction. The obtained receptor alignment and the model assume the reliable orientation of the compounds and their bioactive conformation in LBD conditions. No assumptions of the pharmacophore theory has to be applied, as the structures of the complexes were obtained experimentally and not as a result of mathematical approximation. It is well understood that the application of this type of alignment as an activity predictor may be problematic due to the difficulty with alignment of ligands for which the crystal structure is not known. Molecular modeling methods such as docking or molecular dynamics, while useful, are not sufficiently successful for this purpose. It was reported by Tervo and co-workers[47] that the CoMFA models were derived based on docking and manual alignment rules obtained for a set of flexible molecules and suggested that the model with better predictive quality was constructed on the basis of manual alignment rule. Presented findings seem to confirm this observation. The crystallographically derived orientation of a ligand within the binding site enables much higher diversity of points, and it reflects the possibility of mirror arrangements for a subset of ligands, which is ignored by the structural alignment model. It means that for the structural alignment algorithm both mirror orientations of the same structure are equivalent, which may have high interference on the accuracy of the model.

While, almost all of the alignment methods generated statistically significant CoMFA models (although with quite low Q^2 value), none of them gave more reliable results than the RA, even compared to additional CoMFA models constructed with over 100 AMPA receptor positive modulators available in the literature (Section I, Supporting Information). Thus, the CoMFA model obtained from the crystallographic structures can be a source of alignment in 3D-QSAR studies of positive allosteric modulators of AMPA receptor, and even with some of its limitations, this type of alignment can protect against systematic errors. The importance of appropriate alignment is well understood in the 3D-QSAR studies. Different authors indicate that the prediction accuracy is highly dependent on ligands orientation forced by investigators in basic training sets.[4,21,49,50] In front of these findings, the authors recommend to include poses available from the protein structure alignment into the models, especially when 3D-QSAR methods are being used for activity prediction or for the design of new compounds.[51] It could help to exclude some improper or even impossible orientations of ligands within the binding domain. Whereas the concept of collecting positions from the crystal structure is well-known in QSAR studies,[5,6,21,47−50] 3D-QSAR methods are mainly employed when crystal structure is not known. It should be emphasized that this is an approximation method and that the selection of the inappropriate alignment can propagate and lead to incorrect conclusions.

4. CONCLUSIONS

The obtained CoMFA model based on X-ray crystallographic structures employs reliable orientation of the compounds and their bioactive conformation in the LBD conditions. In comparison with classically derived models based on structural or docking alignments, the RA model shows much better performance and statistical significance. The authors understand that application of this kind of analysis for activity prediction is problematic because of the difficulty with alignment of ligands for which a crystal structure is not

dx.doi.org/10.1021/ci400004e | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

known. The analysis, however, underlines the importance of the fact that QSAR modeling should be as much as possible settled in the context of experimental evidence.

■ ASSOCIATED CONTENT

Supporting Information

List of AMPA receptor positive modulators available in the Protein Data Bank, similarity map HQSAR fragment count, and additional CoMFA models of AMPA receptor positive modulators available in the literature. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +4881-7564862. Fax: +4881-7564860. E-mail: krzysztof.jozwiak@umlub.pl.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. All authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work is supported by the Foundation for Polish Science (TEAM Programme 2009−4/5) and using the equipment purchased within the Project "The equipment of innovative laboratories doing research on new medicines used in the therapy of civilization and neoplastic diseases" within the Operational Program Development of Eastern Poland 2007−2013, Priority Axis I Modern Economy, Operations I.3 Innovation Promotion.

■ REFERENCES

(1) Kubinyi, H. Strategies and recent technologies in drug discovery. *Pharmazie* **1995**, *50*, 647−662.
(2) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.
(3) Kubinyi, H. QSAR and 3D QSAR in drug design. I. Methodology. *Drug Discovery Today* **1997**, *2*, 457−467.
(4) Kubinyi, H. QSAR and 3D QSAR in drug design. II. Applications and Problems. *Drug Discovery Today* **1997**, *2*, 538−546.
(5) Clark, R. D.; Norinder, U. Two personal perspectives on a key issue in contemporary 3D QSAR. *WIREs Comput. Mol. Sci.* **2012**, *2*, 108−113.
(6) Clark, R. D.; Abrahamian, E. Using a staged multi-objective optimization approach to find selective pharmacophore models. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 765−771.
(7) Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: A useful concept for validation of "molecular diversity" descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.
(8) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.
(9) Hansch, C.; Fujita, T. ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
(10) Sullivan, D. C.; Martin, E. J. Exploiting structure−activity relationships in docking. *J. Chem. Inf. Model.* **2008**, *48*, 817−830.
(11) Jitender, V.; Vijay, M. K.; Evans, C. C. 3D-QSAR in drug design: A review. *Curr. Top. Med. Chem.* **2010**, *10*, 95−115.
(12) Scior, T.; Medina-Franco, J. L.; Do, Q. T.; Martínez-Mayorga, K.; Yunes Rojas, J. A.; Bernard, P. How to recognize and workaround pitfalls in QSAR studies: A critical review. *Curr. Med. Chem.* **2009**, *16*, 4297−4313.
(13) Hald, H.; Ahring, P. K.; Timmermann, D. B.; Liljefors, T.; Gajhede, M.; Kastrup, J. S. Distinct structural features of cyclothiazide are responsible for effects on peak current amplitude and desensitization kinetics at iGluR2. *J. Mol. Biol.* **2009**, *391*, 906−917.
(14) Ptak, C. P.; Ahmed, A. H.; Oswald, R. E. Probing the allosteric modulator binding site of GluR2 with thiazide derivatives. *Biochemistry* **2009**, *48*, 8594−8602.
(15) Sun, Y.; Olson, R.; Horning, M.; Armstrong, N.; Mayer, M.; Gouaux, E. Mechanism of glutamate receptor desensitization. *Nature* **2002**, *417*, 245−253.
(16) Battisti, U. M.; Carrozzo, M. M.; Cannazza, G.; Puia, G.; Troisi, L.; Braghiroli, D.; Parenti, C.; Jozwiak, K. Molecular modeling studies, synthesis, configurational stability and biological activity of 8-chloro-2,3,5,6-tetrahydro-3,6-dimethyl-pyrrolo[1,2,3- de]-1,2,4-benzothiadiazine 1,1-dioxide. *Bioorg. Med. Chem.* **2011**, *19*, 7111−7119.
(17) Black, M. D. Therapeutic potential of positive AMPA modulators and their relationship to AMPA receptor subunits. A review of preclinical data. *Psychopharmacology* **2005**, *179*, 154−163.
(18) Sobolevsky, A. I.; Rosconi, M. P.; Gouaux, E. X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **2009**, *462*, 745−758.
(19) Mayer, M. L. Glutamate receptor ion channels. *Curr. Opin. Neurobiol.* **2005**, *15*, 282−288.
(20) Kubinyi H. Comparative Molecular Field Analysis (CoMFA). http://www.wiley.com/legacy/wileychi/ecc/samples/sample05.pdf (accessed November 29, 2012).
(21) Zhang, L.; Tsai, K. C.; Du, L.; Fang, H.; Li, M.; Xu, W. How to generate reliable and predictive CoMFA models. *Curr. Med. Chem.* **2011**, *18*, 923−930.
(22) Jin, R.; Clark, S.; Weeks, A. M.; Dudman, J. T.; Gouaux, E.; Partin, K. M. Mechanism of positive allosteric modulators acting on AMPA receptors. *J. Neurosci.* **2005**, *25*, 9027−9036.
(23) Ward, S. E.; Harries, M.; Aldegheri, L.; Andreotti, D.; Ballantine, S.; Bax, B. D.; Harris, A. J.; Harker, A. J.; Lund, J.; Melarange, R.; et al. Discovery of N-[(2S)-5-(6-fluoro-3-pyridinyl)-2,3-dihydro-1H-inden-2-yl]-2-propanesulfonamide, a novel clinical AMPA receptor positive modulator. *J. Med. Chem.* **2010**, *53*, 5801−5812.
(24) Ward, S. E.; Harries, M.; Aldegheri, L.; Austin, N. E.; Ballantine, S.; Ballini, E.; Bradley, D. M.; Bax, B. D.; Clarke, B. P.; Harris, A. J.; et al. Integration of lead optimization with crystallography for a membrane-bound ion channel target: Discovery of a new class of AMPA receptor positive allosteric modulators. *J. Med. Chem.* **2011**, *54*, 78−94.
(25) Kaae, B. H.; Harpsoe, K.; Kastrup, J. S.; Sanz, A. C.; Pickering, D. S.; Metzler, B.; Clausen, R. P.; Gajhede, M.; Sauerberg, P.; Liljefors, T.; et al. Structural proof of a dimeric positive modulator bridging two identical AMPA receptor-binding sites. *Chem. Biol.* **2007**, *14*, 1294−1303.
(26) Ahmed, A. H.; Ptak, C. P.; Oswald, R. E. Molecular mechanism of flop selectivity and subsite recognition for an AMPA receptor allosteric modulator: Structures of GluA2 and GluA3 in complexes with PEPA. *Biochemistry* **2010**, *49*, 2843−2850.
(27) Jamieson, C.; Basten, S.; Campbell, R. A.; Cumming, I. A.; Gillen, K. J.; Gillespie, J.; Kazemier, B.; Kiczun, M.; Lamont, Y.; Lyons, A. J.; et al. A novel series of positive modulators of the AMPA receptor: Discovery and structure based hit-to-lead studies. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 5753−5756.
(28) Jamieson, C.; Maclean, J. K.; Brown, C. I.; Campbell, R. A.; Gillen, K. J.; Gillespie, J.; Kazemier, B.; Kiczun, M.; Lamont, Y.; Lyons, A. J.; et al. Structure based evolution of a novel series of positive modulators of the AMPA receptor. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 805−811.
(29) Timm, D. E.; Benveniste, M.; Weeks, A. M.; Nisenbaum, E. S.; Partin, K. M. Structural and functional analysis of two new positive allosteric modulators of GluA2 desensitization and deactivation. *Mol. Pharmacol.* **2011**, *80*, 267−280.

(30) Krintel, C.; Frydenvang, K.; Olsen, L.; Kristensen, M. T.; de Barrios, O.; Naur, P.; Francotte, P.; Pirotte, B.; Gajhede, M.; Kastrup, J. S. Thermodynamics and structural analysis of positive allosteric modulation of the ionotropic glutamate receptor GluA2. *Biochem. J.* **2012**, *441*, 173−178.
(31) Cotton, J. L.; Partin, K. M. The contributions of GluR2 to allosteric modulation of AMPA receptors. *Neuropharmacology* **2000**, *39*, 21−31.
(32) *SYBYL-X 2.0*, Tripos International: St. Louis, MO.
(33) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. MUSTANG: A multiple structural alignment algorithm. *Proteins* **2006**, *64*, 559−574.
(34) Palomer, A.; Pascual, J.; Cabre, F.; Garcia, M. L.; Mauleon, D. Derivation of pharmacophore and CoMFA models for leukotriene D(4) receptor antagonists of the quinolinyl(bridged)aryl series. *J. Med. Chem.* **2000**, *43*, 392−400.
(35) Wichapong, K.; Lindner, M.; Pianwanit, S.; Kokpol, S.; Sippl, W. Receptor-based 3D-QSAR studies of checkpoint Wee1 kinase inhibitors. *Eur. J. Med. Chem.* **2009**, *44*, 1383−1395.
(36) Huang, H. Q.; Pan, X. L.; Tan, N. H.; Zeng, G. Z.; Ji, C. J. 3D-QSAR study of sulfonamide inhibitors of human carbonic anhydrase II. *Eur. J. Med. Chem.* **2007**, *42*, 365−372.
(37) Ma, X.; Zhou, L.; Zuo, Z. L.; Liu, J.; Yang, M.; Wang, R. W. Molecular docking and 3-D QSAR studies of substituted 2,2-bisaryl-bicycloheptanes as human 5-lipoxygenase-activating protein (FLAP) inhibitors. *Qsar Comb. Sci.* **2008**, *27*, 1083−1091.
(38) Glide 5.8 User Manual. http://www.schrodinger.com/supportdocs/18/5/ (accessedApril 14, 2013).
(39) AutoDock 4 Tutorial. http://autodock.scripps.edu/faqs-help/tutorial/using-autodock-with-autodocktools/UsingAutoDockWithADT_v2e.pdf (accessed April 14, 2013).
(40) Long, W.; Liu, P. X.; Li, Q.; Xu, Y.; Gao, J. 3D-QSAR studies on a class of IKK-2 inhibitors with GALAHAD used to develop molecular alignment models. *QSAR Comb. Sci.* **2008**, *27*, 1113−1119.
(41) Chen, Y. D.; Li, F. F.; Tang, W. Q.; Zhu, C. C.; Jiang, Y. J.; Zou, J. W.; Yu, Q. S.; You, Q. D. 3D-QSAR studies of HDACs inhibitors using pharmacophore-based alignment. *Eur. J. Med. Chem.* **2009**, *44*, 2868−2876.
(42) Narkhede, S. S.; Degani, M. S. Pharmacophore refinement and 3D-QSAR studies of histamine H-3 antagonists. *QSAR Comb. Sci.* **2007**, *26*, 744−753.
(43) Sastry, G. M.; Dixon, S. L.; Sherman, W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. *J. Chem. Inf. Model.* **2011**, *51*, 2455−2466.
(44) Richmond, N. J.; Abrams, C. A.; Wolohan, P. R.; Abrahamian, E.; Willett, P.; Clark, R. D. GALAHAD: 1. pharmacophore identification by hypermolecular alignment of ligands in 3D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567−587.
(45) Richmond, N. J.; Willetta, P.; Clarkb, R. D. Alignment of three-dimensional molecules using an image recognition algorithm. *J. Mol. Graph. Model.* **2004**, *23*, 199−209.
(46) Partha, P. R.; Somnath, P.; Indrani, M.; Kunal, R. On two novel parameters for validation of predictive QSAR models. *Molecules* **2009**, *14*, 1660−1701.
(47) Tervo, A. J.; Nyronen, T. H.; Ronkko, T.; Poso, A. Comparing the quality and predictiveness between 3D QSAR models obtained from manual and automated alignment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 807−816.
(48) Doweyko, A. M. 3D-QSAR illusions. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 587−596.
(49) Cramer, R. D.; Cruz, P.; Stahl, G.; Curtiss, W. C.; Campbell, B.; Masek, B. B.; Soltanshahi, F. Virtual screening for R-groups, including predicted pIC50 contributions, within large structural databases, using Topomer CoMFA. *J. Chem. Inf. Model* **2008**, *48*, 2180−2195.
(50) Akamatsu, M. Current state and perspectives of 3D-QSAR. *Curr. Top Med. Chem.* **2002**, *2*, 1381−1394.
(51) Jozwiak, K.; Woo, A. Y.; Tanga, M. J.; Toll, L.; Jimenez, L.; Kozocas, J. A; Plazinska, A.; Xiao, R. P.; Wainer, I. W. *Bioorg. Med. Chem.* **2010**, *15*, 728−736.