Docking and Three-Dimensional Quantitative Structure−Activity Relationship (3D QSAR) Analyses of Nonsteroidal Progesterone Receptor Ligands

Annu A. Söderholm,* Pekka T. Lehtovuori, and Tommi H. Nyrönen

CSC−Scientific Computing Ltd., P.O. Box 405, FI-02101 Espoo, Finland

Received March 1, 2006

We report a docking and comparative molecular similarity indices analysis (CoMSIA) study of progesterone receptor (PR) ligands with an emphasis on nonsteroids including tanaproget. The ligand alignment generation, a critical part of model building, comprised two stages. First, thorough conformational sampling of docking poses within the PR binding pocket was made with the program GOLD. Second, a strategy to select representative poses for CoMSIA was developed utilizing the FlexX scoring function. After manual replacement of five poses where this approach had problems, a significant correlation ($r^2 = 0.878$) between the experimental affinities and electrostatic, hydrophobic, and hydrogen bond donor properties of the aligned ligands was found. Extensive model validation was made using random-group cross-validations, external test set predictions ($r_{pred}^2 = 0.833$), and consistency check between the CoMSIA model and the PR binding site structure. Robustness, predictive ability, and automated alignment generation make the model a potential tool for virtual screening.

Introduction

Progesterone is an essential hormone for the regulation of female reproductive function. Its central role in women's health establishes several therapeutic uses for synthetic substances that either mimic or counteract the effects of progesterone. The former are referred to as progestins and the latter as antiprogestins. There are many clinical applications for progestins, e.g., in oral contraceptives, hormone-replacement therapies, and treatment of certain reproductive disorders. Thus far, indications for antiprogestin use are rather limited and primarily focus on medical termination of pregnancy although new clinical applications for antiprogestins are emerging.[1]

The biologic effects of progesterone as well as synthetic progestins and antiprogestins are elicited via the progesterone receptor (PR) (Figure 1). PR belongs to the steroid receptor family, which is a member of the nuclear receptor (NR) superfamily of ligand-dependent transcription factors. PR binding compounds initiate their actions by binding to the ligand binding domain (LBD) of the PR three-domain structure, which is common to all NRs. Binding-induced conformational changes in the LBD structure result in recruitment of coregulators leading to an alteration of transcriptional activity. PR ligands (agonists, partial agonists, and antagonists) modulate the transcriptional activity of PR by inducing different conformations of the coregulator binding surface of LBD. The structural modifications taking place in PR LBD upon agonist binding are well documented,[2−5] whereas those for antagonist binding are not resolved. Some compounds, known as the selective progesterone receptor modulators (SPRMs), induce mixed agonist/antagonist effects, which are thought to depend on cell type specific promoter context[6] and coactivator-to-corepressor ratios.[7,8] Binding of such compounds may bring about conformational changes different from pure agonists and antagonists.

Currently all the PR LBD targeted compounds in clinical use are steroids. Drugs with steroidal core often display significant cross-reactivity with closely related steroid receptors, namely, androgen receptor (AR), glucocorticoid receptor (GR), and mineralocorticoid receptor (MR). This functional overlap is partly responsible for the side effects linked with the steroidal drugs. Although the LBDs and the ligand binding pockets (LBPs) of steroid receptors, particularly of PR, GR, and AR, are very similar and share common ligand binding features, there are enough differences that enable discovery of receptor-selective compounds. It is well established that reduced cross-reactivity is most easily attainable with nonsteroidal compounds. Therefore, considerable effort has been placed on identification of nonsteroidal PR binding compounds with improved selectivity profiles.

During the past decade a number of experimental structure−activity relationship (SAR) studies have become available on a few classes of nonsteroidal PR modulators, which have recently been reviewed by Winneker et al.[9] Here, we describe the results of the three-dimensional quantitative structure−activity relationship (3D QSAR) analysis of a fraction of experimental SAR data provided in the literature.[10−14] The primary objective with 3D QSAR modeling was to identify the physicochemical properties that have a substantial effect on the binding affinity of the ligands included in the analysis. An additional goal was to derive a 3D QSAR model of PR ligands that is comparable to our recently published 3D QSAR model of nonsteroidal AR ligands.[15] The modeling procedure was analogous to the previous analysis. Comparative molecular similarity indices analysis (CoMSIA)[16] was applied as the 3D QSAR method, and molecular docking with the program GOLD[17] was used as the method to predict the binding modes of the investigated PR ligands within the LBP and to align the ligands for model building. Combining docking to 3D QSAR analysis is advantageous because it allows direct visualization and interpretation of modeling results within the binding site, thereby revealing the ligand−receptor interactions contributing positively or negatively on binding affinity. These modeling results are discussed thoroughly in relation to the PR LBP structure. The results are also compared with the structure and binding mode of a high-affinity nonsteroidal PR agonist, tanaproget, which has recently been solved in complex with PR LBD.[5] To the

* To whom correspondence should be addressed. Phone: +358-9-4572252. Fax: +358-9-4572302. E-mail: annu.soderholm@csc.fi.

10.1021/jm060234e CCC: $33.50
Published on Web 06/13/2006

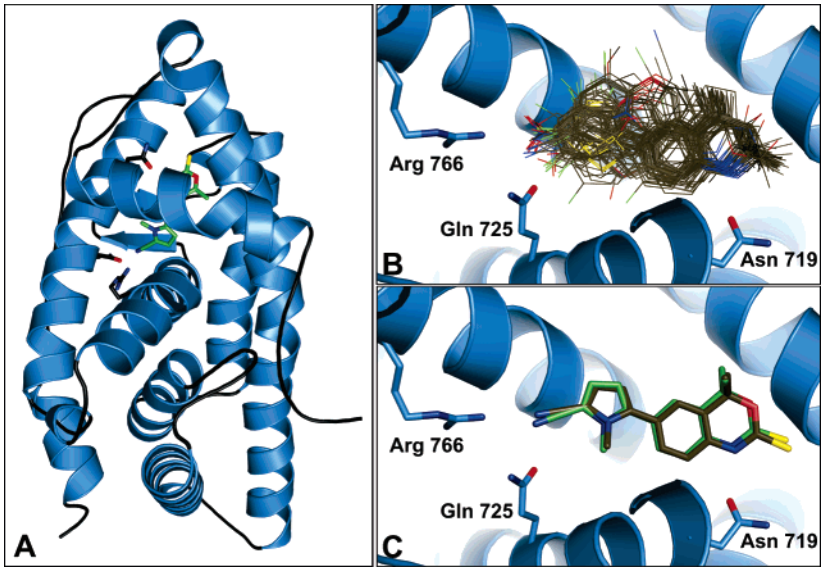


Figure 1. (A) Progesterone receptor ligand binding domain structure in the agonist-bound conformation. The nonsteroidal tanaproget agonist and the three residues in (B) and (C) are shown as sticks. (B) A close-up view of the docking-derived ligand alignment used in 3D QSAR analysis. (C) Superimposition of the cocrystallized tanaproget from the X-ray structure 1zuc (green) and the best docking solution of tanaproget (gray). The picture was generated with PyMOL.[42]

best of our knowledge, this is the first report within PR research where the binding properties of nonsteroidal PR ligands are studied with a 3D QSAR method.

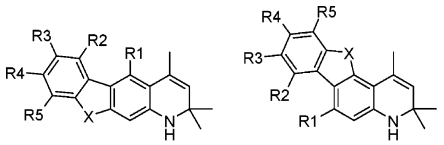
Results and Discussion

Ligand and Protein Structures. We gathered a panel of 74 structurally and pharmacologically diverse PR binding compounds for the 3D QSAR analysis from five publications reported by one laboratory[10−14] (Tables 1−3). Because of similar experimental procedures applied for affinity determination in each publication, the biological data (represented as K_i values) were considered comparable and thus merged into our study. Structurally the compounds are of nonsteroidal nature, except progesterone and medroxyprogesterone acetate. The nonsteroidal compounds represent several different, albeit fairly similar, core structures. Such structural diversity of the ligands used to build the 3D QSAR model is of utmost value if the model is to be used in virtual screening applications, as is our intention. Pharmacologically the majority of the compounds exhibit antagonistic effects of various strengths in cotransfected CV-1 cells, although the entire range of activities from agonists to partial agonists and antagonists are represented in the compound set. Inclusion of all functionalities into model building is motivated by a hypothesis that ligand recognition (binding) depends primarily on the key atomic interactions between the ligand and the receptor.

The mechanism for PR antagonism and the structural modifications taking place upon antagonist binding to PR LBD are incompletely understood.[18−20] Antagonists included in the ligand set do not have structural features that would cause the LBP to undergo major conformational changes, like increase in its volume. All included antagonist and agonist ligands compete against the high-affinity agonist ligand ([^{3}H]progesterone) in the binding assays. The readout from the binding assay reflects the ability of the ligands to replace atomic interactions of progesterone from the agonist bound conformation of PR LBD and therefore does not primarily depend on their pharmacological functionality. As a result, the experimental binding data on agonists and antagonists can be combined and used in the statistical analysis that is dependent on the structural modeling results. This is exactly how we proceeded when modeling the 3D QSAR of AR ligands.[15]

An interesting feature of the nonsteroidal PR antagonists studied here is that they are of equal or even smaller size than the studied PR agonists. The small nonsteroidal PR antagonists fit nicely into the binding pocket of the agonist structure of PR LBD. For these reasons, we used a protein model based on the

Table 1. Structures of Nonsteroidal Ligands **1**−**25**

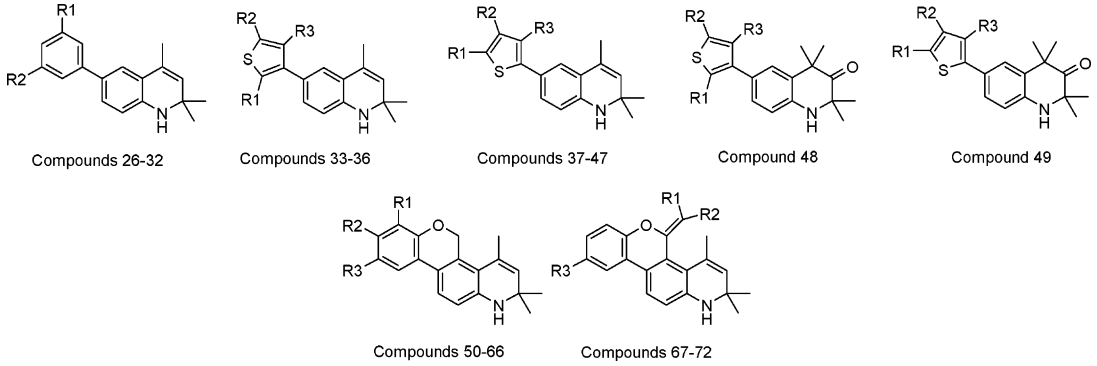


Compounds 1-16 Compounds 17-25

compd	R1	R2	R3	R4	R5	X	pK_{iExp} [b]	pK_{iPred} [c]
1[a,d]	H	H	H	H	H	CH_2	7.85	7.84
2[d]	H	H	H	H	CH_2OH	CH_2	7.91	8.12
3[d]	H	H	H	$COCH_3$	H	CH_2	6.75	6.83
4[d]	H	H	H	NO_2	H	CH_2	8.44	8.64
5[d]	H	H	H	Br	H	CH_2	7.62	7.67
6[d]	H	H	H	Cl	H	CH_2	7.54	7.87
7[d]	H	H	H	F	H	CH_2	8.46	8.44
8[a,d]	H	H	H	H	F	CH_2	8.64	8.25
9[d]	H	F	H	H	H	CH_2	7.95	8.22
10[d]	H	F	H	H	F	CH_2	8.51	8.56
11[d]	H	H	F	NO_2	H	CH_2	8.72	8.93
12[d]	F	H	H	F	H	CH_2	8.82	8.25
13[d]	H	H	H	H	H	O	6.74	7.21
14[d]	H	H	H	H	H	NH	6.95	6.92
15[d]	H	H	H	H	H	C=O	5.45	5.58
16[a,d]	H	F	H	H	H	C=O	7.54	7.21
17[d]	H	H	H	H	H	CH_2	7.89	7.82
18[d]	H	H	H	NO_2	H	CH_2	7.24	7.07
19[d]	H	H	H	Br	H	CH_2	6.7	6.9
20[d]	H	H	H	H	F	CH_2	7.81	7.73
21[d]	H	H	F	NO_2	H	CH_2	7.01	7.34
22[d]	F	H	H	F	H	CH_2	7.68	7.67
23[d]	H	H	H	H	H	O	7.11	7.44
24[d]	H	H	H	H	H	NEt	6.77	6.78
25[d]	H	H	H	H	H	NBu	7.11	7.08

[a] Compound that belongs to the test set. [b] Experimental binding affinity (pK_i). [c] Predicted binding affinity (pK_i). [d] Reference 10.

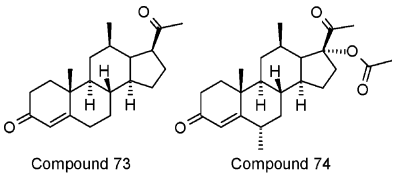
Table 2. Structures of Nonsteroidal Ligands **26**−**72**



compd	R1	R2	R3	p$K_{i\text{Exp}}$ [b]	p$K_{i\text{Pred}}$ [c]	compd	R1	R2	R3	p$K_{i\text{Exp}}$ [b]	p$K_{i\text{Pred}}$ [c]
26[d]	H	H		6.88	7	**50**[a,f]	H	H	H	7.08	6.44
27[d]	F	H		6.74	7.2	**51**[f]	H	H	F	7.71	7.56
28[d]	F	F		8	7.73	**52**[f]	H	H	Cl	7.89	7.68
29[a,d]	CN	H		7.72	7.34	**53**[f]	H	H	Br	7.82	7.61
30[d]	CN	F		8	7.75	**54**[f]	F	H	H	6.61	6.79
31[d]	NO_2	H		7.7	7.76	**55**[a,f]	Cl	H	H	6.56	6.46
32[d]	NO_2	F		8.3	8.18	**56**[f]	H	H	Me	7.35	7.04
33[a,e]	H	CHO	H	7.43	7.08	**57**[f]	H	H	OMe	7.07	7.36
34[e]	H	CN	H	8.46	7.78	**58**[f]	H	H	NO_2	8.17	8.07
35[e]	Me	CN	H	7.51	7.78	**59**[f]	H	OMe	H	6.45	6.24
36[e]	Me	CN	Me	7.44	7.99	**60**[f]	F	H	F	7.87	8.09
37[a,e]	H	H	H	6.63	6.92	**61**[a,f]	F	H	Cl	8.25	7.64
38[e]	H	H	Me	7.6	7.28	**62**[f]	F	H	Br	7.91	7.67
39[e]	H	Br	H	7.49	7.72	**63**[f]	F	H	CN	8.35	8.11
40[e]	Cl	H	H	7.92	7.5	**64**[f]	Me	H	Cl	6.69	7.21
41[e]	Br	H	H	7.51	7.04	**65**[f]	OMe	H	Br	6.32	6.3
42[e]	CHO	H	H	7.1	7.36	**66**[f]	F	CHO	Br	6.54	6.41
43[e]	NO_2	H	H	8.59	8.09	**67**[g]	H	H	F	8.21	8.05
44[e]	CN	H	H	7.26	7.44	**68**[g]	Br	H	F	8.68	8.69
45[e]	CN	H	Br	6.36	6.46	**69**[g]	CN	H	F	7.35	7.49
46[e]	CN	Br	H	6.6	6.3	**70**[g]	*n*-Pr	H	F	8.51	8.58
47[e]	CN	H	Me	7.59	7.67	**71**[a,g]	Et	Me	F	8.1	7.76
48[e]	H	CN	H	7.91	7.84	**72**[g]	Me	Et	F	7.78	7.83
49[e]	CN	H	Me	8.41	8.08						

[a] Compound that belongs to the test set. [b] Experimental binding affinity (pK_i). [c] Predicted binding affinity (pK_i). [d] Reference 11. [e] Reference 14. [f] Reference 12. [g] Reference 13.

Table 3. Structures of Steroidal Ligands **73** and **74**



compd[a]	p$K_{i\text{Exp}}$ [b]	p$K_{i\text{Pred}}$ [c]
73[d]	8.46	8.43
74[d]	9.47	9.53

[a] Neither compound in this table belongs to the test set. [b] Experimental binding affinity (pK_i). [c] Predicted binding affinity (pK_i). [d] Reference 13.

agonist structure of PR LBD[5] in aligning the ligands with docking (Figure 1).

Alignment Generation. There are two stages in our alignment generation process. First, the docking program GOLD[17] was used for conformational sampling of possible ligand binding poses within PR LBP. Second, a scoring function was applied to predict the bioactive conformation of each ligand among the generated docking poses. Prior to docking, a small modification in the LBP structure was made to slightly enlarge the volume of the binding cavity. The change involved replacement of the Met801 side chain conformation observed in the X-ray structure with another rotamer. This alteration allowed inclusion of a few bulkier compounds[13] into the 3D QSAR analysis. The research group, the ligand data of which are used in our study, has reported additional data on nonsteroidal PR modulators[13,21−23] that could not be included in this work because of their large size. More drastic modifications made to the LBP would have rendered the binding site too spacious for the smaller compounds and the docking results less reliable. The structural water molecule, which is found present in nearly every ligand-complexed NR LBD structure, was included as part of the protein structure in the docking simulations.

Generation of the correct spatial alignment of the investigated compounds for 3D QSAR analysis is of vital importance, since the correctness of the analysis is dependent on the quality of the alignment. This challenging step is often impeded by the lack of data on biologically active conformations of the compounds in complex with their target protein. The growing amount of structural data on small compounds cocrystallized with their receptor proteins is, however, becoming available. While we were starting our project, Zhang et al.[5] reported the first cocrystal structure of a nonsteroidal tanaproget agonist with PR LBD, which revealed structural information relevant for modeling the 3D QSAR of nonsteroidal PR ligands.

Tanaproget has a structure related to the structures of many of the nonsteroids in the study and serves as a good indicator for binding interactions of such compounds. Thus, the crystal structure of tanaproget was used to evaluate the quality of

docking results. Tanaproget itself was not included in the 3D QSAR analysis because no comparable binding data were available. Despite the change in Met801 side chain conformation and thus a small increase in LBP volume, GOLD was successful in docking tanaproget close to its crystallographically determined bioactive conformation (Figure 1). The heavy atom root-mean-square deviation (rmsd) for the docking solution closest to the crystal structure of tanaproget was 0.338 Å. Such proximity can be regarded as a good reproduction of the crystal structure. We therefore believe that the binding conformations of the analogous ligands analyzed here are reasonably well predicted by GOLD. Several evaluations of the docking accuracy of commonly used docking programs have concluded that many programs are able to reproduce the crystallographically determined binding modes, GOLD being one of the most reliable docking programs.[24−26]

We aimed at generating the alignment as automatically and objectively as possible by using a single scoring function to select a top-ranked docking pose from all the poses generated for each ligand. However, none of the explored scoring functions described in the methods section resulted directly in an alignment that provided a statistically meaningful correlation between the structural features and the measured binding affinities of the ligands. Even if the docking programs perform well in generating the correct bioactive conformation of a ligand, the current scoring functions are less successful at correctly identifying it. A number of evaluations have been published on the ability of scoring functions to distinguish the correct bioactive conformation from the ensemble of docked poses.[25−27] Major differences in performance have been documented. Availability of data on experimentally resolved protein−ligand interactions within binding sites of different chemical nature sets the limits for the quality of scoring functions that have been created this far. Many of the scoring functions do reasonably well in predicting polar interactions, whereas the hydrophobic interactions are often more problematic. Because the LBP of PR is fairly hydrophobic, we were not surprised that all scoring functions assessed had difficulties in directly producing a ligand alignment that would have resulted in a significant 3D QSAR model. However, FlexX score has been shown to have a higher success rate at identifying the correct ligand−receptor interactions in hydrophobic ligand binding sites than the other scoring functions that we tried.[26]

The determination of the best docking program and scoring function combination for providing the most accurate binding mode prediction is case-specific. We therefore used tanaproget docking results to guide the selection of the most suitable scoring function for our case. We examined the ability of different scoring functions to identify the best tanaproget docking solution with the lowest rmsd value from its crystallographically determined conformation. FlexX[28,29] and PMF[30] scoring functions managed to recognize the lowest rmsd result as the single best docking pose, while XScore[31] identified two other solutions with higher rmsd values that are equally good. From a statistical point of view, the initial alignment based on FlexX selection provided by far the best starting point for 3D QSAR model building. Consequently, we selected the representative docking poses using FlexX scoring function.

Five of the training set compounds (compounds **3**, **36**, **44**, **46**, **74** in Tables 1−3) did not align properly with others in the initial alignment, which is based on the automatic selection of docking poses with the best FlexX scores. However, the differences in the FlexX predicted interaction energies between the top-ranked poses from the best-scoring and the second-best-scoring ensembles of docking poses generated by GOLD for these five compounds were marginal (−2.2 kcal/mol on average) when the inaccuracy of the current scoring functions is considered. By taking those conformations into the model, a significant correlation with the experimental binding affinities was achieved. The replacement of docking poses resulted in a better overall superimposition and prevented the compounds from becoming outliers. As a result, all 64 training set compounds were included in the 3D QSAR model.

3D QSAR Analysis. To examine the structural and chemical features contributing to the biological activity of the studied ligands, the alignment derived from docking simulations was quantitatively analyzed using the CoMSIA procedure.[16] The structure−activity relationships were best explained with the electrostatic, hydrogen bond donor, and hydrophobic properties of the ligands. Electrostatic and hydrogen bond donor fields constitute the most important descriptors to the information content of the CoMSIA model, while the hydrophobic field plays a somewhat smaller role. In fact, the electrostatic field seems to be essential for building a model from this compound set because no statistically significant model could be derived without it. The contributions of the electrostatic, hydrogen bond donor, and hydrophobic fields to the CoMSIA model are 42%, 36%, and 22%, respectively.

To determine the quality of the model, one initially needs to consider the statistical values obtained from the PLS analysis. The statistical quality and the robustness of the model were determined with internal cross-validation procedures. Internal validation using leave-one-out (LOO) cross-validation gave a correlation coefficient q^2_{LOO} of 0.637 and a standard error of prediction ($SDEP_{LOO}$) of 0.480. A more rigorous cross-validation using 10 random groups yielded an average q^2_{10} of 0.601 and $SDEP_{10}$ of 0.501, and use of five random groups yielded an average q^2_5 of 0.563 and $SDEP_5$ of 0.525. The nonvalidated PLS analysis gave a correlation coefficient r^2 of 0.878 and a standard error of estimate (SEE) of 0.278. The q^2 and r^2 values of this magnitude reflect a statistically significant and robust model.

The most rigorous test for the predictive ability of the model was done with the 10 external test set compounds, which were completely excluded from model building but were processed in the same way as the training set compounds. The chosen test set provides both structurewise and activitywise a good representation of the compounds used to build the model. Both high- and low-affinity compounds within the test set were predicted close to their experimentally measured binding affinities, yielding a predictive r^2 of 0.833 with a SEE of 0.294. Hence, besides the good statistical quality, the model also shows excellent predictive properties. Experimental and predicted affinities for all the compounds are presented in Tables 1−3. The predicted affinities are plotted against the experimental affinities in Figure 2.

Visualization and Interpretation of the 3D QSAR Model with Respect to PR LBP. The reliability of our model also necessitates complementarity relative to the binding site structure because the model was derived from a ligand alignment generated with docking. To resolve whether the model derived from the superimposed ligands is in consensus with the receptor structure, the statistically relevant data of the CoMSIA model are visualized as 3D contour maps inside the LBP. Analysis of the contours with respect to the chemical environment of the LBP will then reveal the true quality of the model. The visualization of the CoMSIA model consisting of electrostatic,

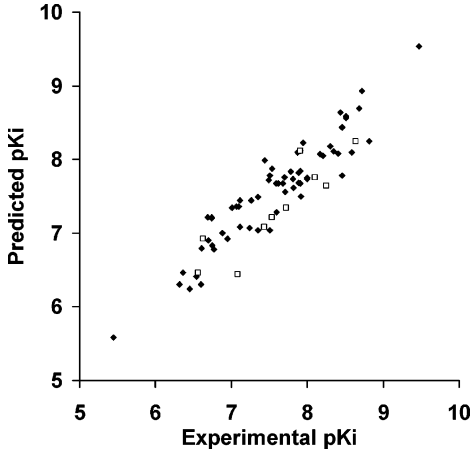


Figure 2. Correlation between the experimental and predicted activities (pK_i) for the training set (◆) and the test set compounds (□).

hydrogen bond donor, and hydrophobic fields is represented with the LBP residues in Figure 3.

Electrostatic Field. Electrostatic field is the major contributor to the information content of our CoMSIA model. Interpretation of the electrostatic fields in relation to the almost entirely hydrophobic PR LBP structure is, however, rather complex. To interpret these fields, one needs to consider, besides their complementarity with the binding site structure, the electrostatic properties (electron distribution) of the molecules within the ligand set.

Next to the structural water molecule at the bottom of the binding pocket there is a small volume that is indicated by the CoMSIA model to be favorable for negative partial charge. This map is in accordance with the chemical environment of the receptor structure. The end of the binding pocket has a minor positive charge due to Arg766 and thus forms constructive interactions with the negatively charged moieties of ligands. This map also agrees with the experimental SAR data, which show a decrease in affinity with decreasing electronegativity of the aromatic ring substituents.[10−12,14] The space behind the volume favoring negative partial charge displays a larger volume favoring positive partial charge. This volume should be interpreted in conjunction with the volume favoring negative partial charge. The binding affinity is enhanced by the presence of strongly electron-withdrawing substituents in the aromatic ring. Correspondingly, the ring structure becomes electron-deficient and is left with a partial positive charge. The more electronegative is the substituent, the more positive is the charge presented by the ring and the stronger is the predicted binding affinity. Together, these two electrostatic fields indicate that a charge polarization (local dipole), which is present at the end of the ligand and positioned at the bottom of the LBP, will enhance binding affinity.

In the middle of the binding pocket there is a small favored volume for negative partial charge. The superimposed ligand set contains a few high-affinity molecules (e.g., **58**, **63**, **68**, and **70** in Table 2) that have an oxygen atom close to this volume. The oxygen exists in a chroman-like ring structure, where it is connected to a six-membered ring system associated with benzene. It is well-known that oxygens present in highly conjugated aromatic systems are poor hydrogen bond acceptors, since the electron density of oxygen can delocalize. Thus, we suspect that oxygens present in such a position in these compounds are of hydrophobic nature and that the favored

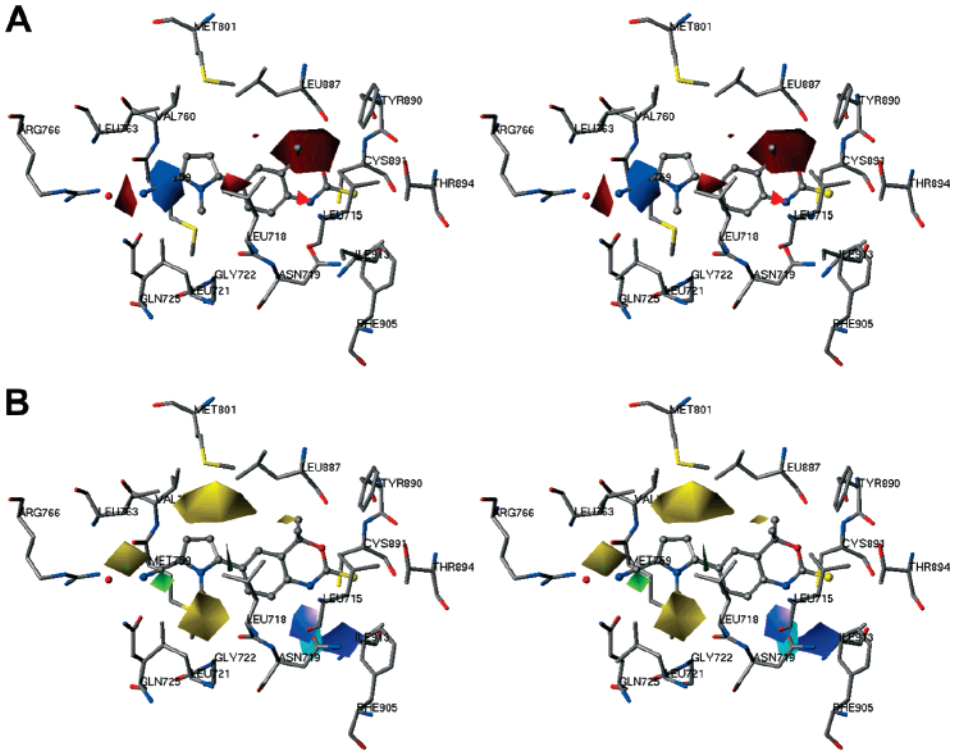


Figure 3. Stereoviews of the 3D contour maps of the CoMSIA model within the PR LBP represented as stdev*coeff plots. The most important LBP residues (sticks) are shown with the crystallographically determined conformation of tanaproget (ball-and-stick). The structural water is shown as a red sphere. (A) The red contours (contoured at −0.04 kcal/mol) indicate volumes where negative potential in ligands increases binding affinity; blue contours (+0.06 kcal/mol) indicate volumes where positive potential increases binding affinity. (B) The cyan contour (+0.15 kcal/mol) represents a region where hydrogen bond donors in ligands enhance affinity; purple contours (−0.05 kcal/mol) represent regions where hydrogen bond donors are detrimental to affinity. The yellow contours (+0.015 kcal/mol) correspond to regions where hydrophobicity in ligands increases binding affinity; green contours (−0.03 kcal/mol) correspond to regions where hydrophilic elements increase binding affinity.

electrostatic field volume shown in the CoMSIA model is an artifact arising from the point charge used for the calculation of electrostatic fields. This conclusion is supported by the protein structure. Within 5−6 Å there are no residues that would benefit from the partial negative charge in this volume. However, this volume has only a small contribution to the prediction of binding affinity. A change of oxygen to carbon, e.g., in compound **70**, causes the predicted activity (pK_i) to drop from 8.58 to 8.39, which is within the margin of error of the model.

At the other end of the binding pocket, next to the electron-deficient side of the Tyr890 phenyl ring, the model displays a favorable volume for partial negative charge. This region is able to interact with electron-rich chemical entities of ligands. The volume is therefore in agreement with the receptor structure. Furthermore, experimental structure data support the presence of this volume because its position encompasses the space occupied by the electron-rich oxygen of the benzoxazine ring of tanaproget.[5]

Hydrogen Bond Donor Field. According to the CoMSIA model, there is one volume in the LBP where the hydrogen bond donor presented by the ligand is favorable for binding affinity. This volume coincides with the region around residues Asn719 and Cys891. The volume agrees with the LBP structure because both residues, particularly Asn719, are willing to accept a hydrogen bond. However, hydrogen-bonding to these residues cannot be considered essential for high-affinity binding because the endogenous ligand, progesterone, shows a K_i value in the nanomolar range[13,14] but according to the X-ray structures does not form such an interaction to either Asn719 or Cys891.[2,3] Nor does the molecular dynamics simulations predict any notable hydrogen-bonding between progesterone and LBP at this location.[32] Yet one can expect hydrogen-bonding to Asn719 or Cys891 to increase ligand binding affinity to PR. This is observed in the binding of tanaproget to PR LBD, where tanaproget offers its benzoxazine NH for hydrogen-bonding with Asn719.[5]

Two unfavorable volumes for donor interactions are displayed close to the favored volume. One of the volumes lies partly on top of the Asn719 side chain and points toward the planes of the peptide bond formed between Leu718 and Asn719 and the backbone hydrogen bond formed between Leu715 and Asn719. The other unfavorable volume for donor interactions extends toward the plane of the aromatic ring of Phe905 and the side chain of Ile913. Both of the binding cavity surfaces surrounding these two volumes are of hydrophobic nature and do not present any good hydrogen bond acceptors to pair with ligand donors. Therefore, a donor group presented to these regions by the ligand will also have an affinity-reducing effect according to the protein structure. Essentially, the positions of these two unfavorable volumes point out the importance of directionality for hydrogen-bonding interactions.

Hydrophobic Field. The CoMSIA model indicates three volumes where hydrophobic elements in the ligands are favored and enhance binding affinity. The first favored volume for hydrophobic interactions is situated at the bottom of the binding pocket, next to the structural water molecule. This volume partly overlaps with the electrostatic volume favoring partial negative charge, suggesting that electron-withdrawing moieties with hydrophobic properties are preferred at this site. This is in line with the experimental affinity measurements.[10] The other two volumes favoring hydrophobicity are located on opposite sides from each other in the central part of the LBP. One of the volumes is lined by side chains of residues Val760, Leu763, Met801, and Leu887, while the side chains of residues Leu718, Leu721, and Met759 and the plane of the peptide bond between Leu721 and Gly722 form the borders of the other volume. All the residues contribute to the formation of hydrophobic surfaces in the middle of the LBP, thus complementing the favored hydrophobic volumes. In the cocrystal structure of tanaproget with PR LBD, the methyl group extending from the cyano-pyrrole ring accommodates the latter volume while Met801, which was modified in our work, extends toward the binding cavity and partly fills the previous volume.[5]

The model also specifies a rather small volume where hydrophobicity is disfavored and hydrophilicity is associated with an increase in binding affinity. The volume favoring hydrophilic elements is found at the bottom of the binding pocket, next to the side chains of Gln725, Arg766, and the structural water molecule. This is one of the most hydrophilic surfaces of the LBP and therefore able to interact with polar parts of ligands.

Conclusions

In our study we have utilized the 3D QSAR method of CoMSIA to explore the receptor−ligand interactions influencing receptor binding affinity of PR ligands, placing an emphasis on the nonsteroidal PR binding compounds. The analysis was performed on a series of 74 publicly available PR binding compounds comprising 72 nonsteroids. The selected series of compounds represents a variety of pharmacological functionalities and several structural scaffolds. A 3D QSAR model generated from such a diverse set of compounds has the capacity for identifying PR binding compounds without discriminating between the different functional activities of the compounds.

The investigated compounds were superimposed for 3D QSAR analysis using molecular docking. By combining docking into 3D QSAR analysis, we connected the PR LBP structure into alignment generation. The alignment, on the other hand, served as the basis for statistical analysis in which the structural differences of the ligands were related to the variations in their experimentally observed binding affinities. The ligand alignment generated within the PR LBP resulted in a statistically significant 3D QSAR model, and the contour maps, which visualize the regions of structural features explaining the variance in the binding affinity, nicely complement the structural elements of PR LBP. Together, these results indicate that the alignment comprises biologically active conformations of the PR ligands, thus confirming the accuracy of the alignment. This in turn suggests that the entire docking procedure, including the protein structure used, is valid.

Besides studying the features affecting the binding affinity of nonsteroidal PR ligands, we aimed to create a model that is methodwise comparable to our previously generated CoMSIA model of nonsteroidal AR ligands. Even if the CoMSIA models of PR and AR ligands were built using analogous modeling procedures, the variance in the experimental binding affinities of the PR and the AR ligands could not be explained with the same molecular field descriptors. Undoubtedly this result is connected to the distinct series of ligands used to build the models, but it also indicates that there are characteristic differences in the molecular properties of the nonsteroidal PR and AR binding compounds, which account for their binding to the associated receptors. A comparison of the two CoMSIA models does not, however, provide any apparent explanation of the receptor−ligand interactions needed for selective ligand binding. Thus, further research is required to reveal the key structural and chemical features behind receptor-specific binding.

Together, the two CoMSIA models and the procedure described are applicable in virtual identification of PR and AR

binding compounds from chemical databases. The models are of particular advantage when used in parallel for screening of receptor-selective nonsteroidal compounds and in the subsequent structure-based optimization of such compounds. In the upcoming virtual screening studies, we intend to apply the derived models in predicting the binding affinities for all the poses generated in docking simulations and use these predictions in guiding the selection of compounds for experimental screening. Because the CoMSIA models are generated specifically for PR and AR from sets of known PR and AR ligands, respectively, and because the models are consistent with their respective LBP structures, they are well suited for binding affinity prediction and can be perceived as receptor-specific scoring functions. The affinity prediction by the CoMSIA models can be used as an extension to the existing scoring functions as the only means for affinity prediction and ranking of compounds from the virtual databases. Thus, the created CoMSIA models are an asset in virtual screening when the limitations of the current scoring functions are known.

Experimental Section

Molecular Modeling. All modeling work was performed using Sybyl, version 7.0,[33] unless otherwise noted.

Protein Data. The human PR LBD structure cocrystallized with the agonist tanaproget[5] was retrieved from the Protein Data Bank[34] (PDB code: 1zuc). From the crystal structure, monomer A was selected for the docking simulations. The side chain conformation of Met801 in the crystal structure was replaced with another one from the rotamer library in the BODIL software[35] to increase the volume of the LBP and to allow docking of ligands larger than the cocrystallized agonist. Tanaproget and all water molecules except the structural water (Wat5 in 1zuc) between the Arg766 and Gln725 were removed from the protein structure. Hydrogens were added to the protein and the water.

Ligand Data. The data for 74 ligands used in the modeling were collected from five publications reported by one laboratory.[10−14] The binding affinity values were represented as K_i values determined in a competitive binding assay. Compounds lacking exact affinity values and defined stereochemistry were excluded from the data set, as were compounds too bulky to fit the model of PR LBP (i.e., compounds **10**, **12**, **18** from Zhi et al.[13]).

The ligands were divided into a training set of 64 compounds and a test set of 10 compounds. The test set was selected to represent the training set as well as possible in terms of structural composition and activity range. Tanaproget was considered an extra test compound used to validate the modeling procedure and to interpret the final model because it could not be included in the model building because of lack of comparable affinity data.

The ligands were converted into 3D structures using the program CORINA, version 3.0.[36,37] CORINA was allowed to generate a maximum of five ring conformations while applying an energy window of 30 kJ/mol between the highest and the lowest energy ring conformations. This yielded three (compounds **1**−**47** and tanaproget), five (**48**−**72**), or eight (**73**, **74**) initial conformations to be used in generating the respective number of ensembles of docking poses in docking simulations. Gasteiger−Hückel charges[38,39] were calculated for each compound.

Alignment Generation. The docking program GOLD, version 2.2,[17] was used to predict the binding conformations and orientations of the compounds within PR LBP. With the default parameters of GOLD, the docking procedure was repeated 10 times for each CORINA-generated conformation. This resulted in 30 (compounds **1**−**47**), 50 (**48**−**72**), or 80 (compounds **73**, **74**) docking solutions for every compound; i.e., 3, 5, or 8 ensembles of docking poses for each ligand were produced. Several scoring functions, including GOLDScore,[17] XScore,[31] and the individual functions integrated within CScore,[40] were attempted to produce a statistically significant alignment. The final alignment was generated by selecting the representative conformation for each compound as the top-ranked pose from the best-scoring ensemble of docking poses according to FlexX scoring function[28,29] for all except five of the ligands (**3**, **36**, **44**, **46**, **74**). For these five compounds the top-ranked pose from the second-best-scoring ensemble of docking poses was included in the alignment to prevent them from being outliers.

3D QSAR Analysis. The 3D QSAR model was built by applying the CoMSIA[16] procedure on the ligand alignment obtained from docking. The CoMSIA molecular descriptor fields were calculated with default parameters and correlated with the variations in the binding affinity data with partial least squares (PLS) analysis.[41] Descriptor fields with standard deviation less than 2 units were filtered out from the PLS analysis. Leave-one-out (LOO) cross-validation was used to determine the optimum number of PLS components. The final model was derived using the electrostatic, hydrophobic, and hydrogen bond donor fields and six principal components in the PLS analysis.

The predictivity of the model was validated using both internal and external methods. LOO and random-group cross-validations with 10 and 5 groups were applied as internal validation methods. Each random group cross-validation was repeated 25 times to obtain the mean values for q^2 and standard error of prediction (SDEP). External validation was performed with the test set of 10 compounds not included in the 3D QSAR model building.

Acknowledgment. This work was funded by the Academy of Finland SYSBIO program and the Finnish Technology Agency (TEKES) Drug2000 program. Computational resources and program licenses were supplied by CSC−Scientific Computing Ltd.

References

(1) Spitz, I. M. Progesterone antagonists and progesterone receptor modulators: an overview. *Steroids* **2003**, *68*, 981−993.
(2) Williams, S. P.; Sigler, P. B. Atomic structure of progesterone complexed with its receptor. *Nature* **1998**, *393*, 392−396.
(3) Matias, P. M.; Donner, P.; Coelho, R.; Thomaz, M.; Peixoto, C.; et al. Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.* **2000**, *275*, 26164−26171.
(4) Madauss, K. P.; Deng, S. J.; Austin, R. J.; Lambert, M. H.; McLay, I.; et al. Progesterone receptor ligand binding pocket flexibility: crystal structures of the norethindrone and mometasone furoate complexes. *J. Med. Chem.* **2004**, *47*, 3381−3387.
(5) Zhang, Z.; Olland, A. M.; Zhu, Y.; Cohen, J.; Berrodin, T.; et al. Molecular and pharmacological properties of a potent and selective novel nonsteroidal progesterone receptor agonist tanaproget. *J. Biol. Chem.* **2005**, *280*, 28468−28475.
(6) Meyer, M. E.; Pornon, A.; Ji, J. W.; Bocquel, M. T.; Chambon, P.; et al. Agonistic and antagonistic activities of RU486 on the functions of the human progesterone receptor. *EMBO J.* **1990**, *9*, 3923−3932.
(7) Giannoukos, G.; Szapary, D.; Smith, C. L.; Meeker, J. E.; Simons, S. S., Jr. New antiprogestins with partial agonist activity: potential selective progesterone receptor modulators (SPRMs) and probes for receptor- and coregulator-induced changes in progesterone receptor induction properties. *Mol. Endocrinol.* **2001**, *15*, 255−270.
(8) Liu, Z.; Auboeuf, D.; Wong, J.; Chen, J. D.; Tsai, S. Y.; et al. Coactivator/corepressor ratios modulate PR-mediated transcription by the selective receptor modulator RU486. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7940−7944.
(9) Winneker, R. C.; Fensome, A.; Wrobel, J. E.; Zhang, Z.; Zhang, P. Nonsteroidal progesterone receptor modulators: structure activity relationships. *Semin. Reprod. Med.* **2005**, *23*, 46−57.
(10) Hamann, L. G.; Winn, D. T.; Pooley, C. L.; Tegley, C. M.; West, S. J.; et al. Nonsteroidal progesterone receptor antagonists based on a conformationally-restricted subseries of 6-aryl-1,2-dihydro-2,2,4-trimethylquinolines. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2731−2736.
(11) Pooley, C. L.; Edwards, J. P.; Goldman, M. E.; Wang, M. W.; Marschke, K. B.; et al. Discovery and preliminary SAR studies of a novel, nonsteroidal progesterone receptor antagonist pharmacophore. *J. Med. Chem.* **1998**, *41*, 3461−3466.
(12) Zhi, L.; Ringgenberg, J. D.; Edwards, J. P.; Tegley, C. M.; West, S. J.; et al. Development of progesterone receptor antagonists from 1,2-dihydrochromeno[3,4-*f*]quinoline agonist pharmacophore. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2075−2078.
(13) Zhi, L.; Tegley, C. M.; Pio, B.; Edwards, J. P.; Jones, T. K. et al. Synthesis and biological activity of 5-methylidene 1,2-dihydrochromeno[3,4-*f*]quinoline derivatives as progesterone receptor modulators. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2071−2074.

(14) Zhi, L.; Tegley, C. M.; Pio, B.; West, S. J.; Marschke, K. B.; et al. Nonsteroidal progesterone receptor antagonists based on 6-thiophene-hydroquinolines. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 415−418.
(15) Söderholm, A. A.; Lehtovuori, P. T.; Nyrönen, T. H. Three-dimensional structure−activity relationships of nonsteroidal ligands in complex with androgen receptor ligand-binding domain. *J. Med. Chem.* **2005**, *48*, 917−925.
(16) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.
(17) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(18) Leonhardt, S. A.; Boonyaratanakornkit, V.; Edwards, D. P. Progesterone receptor transcription and non-transcription signaling mechanisms. *Steroids* **2003**, *68*, 761−770.
(19) Leonhardt, S. A.; Edwards, D. P. Mechanism of action of progesterone antagonists. *Exp. Biol. Med. (Maywood, NJ, U.S.)* **2002**, *227*, 969−980.
(20) Wardell, S. E.; Edwards, D. P. Mechanisms controlling agonist and antagonist potential of selective progesterone receptor modulators (SPRMs). *Semin. Reprod. Med.* **2005**, *23*, 9−21.
(21) Tegley, C. M.; Zhi, L.; Marschke, K. B.; Gottardis, M. M.; Yang, Q.; et al. 5-Benzylidene 1,2-dihydrochromeno[3,4-*f*]quinolines, a novel class of nonsteroidal human progesterone receptor agonists. *J. Med. Chem.* **1998**, *41*, 4354−4359.
(22) Zhi, L.; Tegley, C. M.; Pio, B.; Edwards, J. P.; Motamedi, M.; Jones, T. K.; Marschke, K. B.; Mais, D. E.; Risek, B.; Schrader, W. T. 5-Benzylidene-1,2-dihydrochromeno[3,4-*f*]quinolines as selective progesterone receptor modulators. *J. Med. Chem.* **2003**, *46*, 4104−4112.
(23) Zhi, L.; Tegley, C. M.; Marschke, K. B.; Mais, D. E.; Jones, T. K. 5-Aryl-1,2,3,4-tetrahydrochromeno[3,4-*f*]quinolin-3-ones as a novel class of nonsteroidal progesterone receptor agonists: effect of A-ring modification. *J. Med. Chem.* **1999**, *42*, 1466−1472.
(24) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759−4767.
(25) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47*, 558−565.
(26) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(27) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; et al. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, in press
(28) Rarey, M.; Kramer, B.; Lengauer, T. Docking of hydrophobic ligands with interaction-based matching algorithms. *Bioinformatics* **1999**, *15*, 243−250.
(29) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.
(30) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein−ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791−804.
(31) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.
(32) Hillisch, A.; von Langen, J.; Menzenbach, B.; Droescher, P.; Kaufmann, G.; et al. The significance of the 20-carbonyl group of progesterone in steroid receptor binding: a molecular dynamics and structure-based ligand design study. *Steroids* **2003**, *68*, 869−878.
(33) *Sybyl*, version 7.0; Tripos Inc.: St. Louis, MO.
(34) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(35) Lehtonen, J. V.; Still, D. J.; Rantanen, V. V.; Ekholm, J.; Björklund, D.; et al. BODIL: a molecular modeling environment for structure−function analysis and drug design. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 401−419.
(36) Sadowski, J.; Gasteiger, J. From atoms and bonds to three-dimensional atomic coordinates: model builders. *Chem. Rev.* **1993**, *93*, 2567−2581.
(37) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of automatic three-dimensional model builders using 639 x-ray structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(38) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3222.
(39) Purcell, W. P.; Singer, J. A. A brief review and table of semiempirical parameters used in the Hückel molecular orbital method. *J. Chem. Eng. Data* **1967**, *12*, 235−246.
(40) Clark, R. D.; Strizhev, A.; Leonard, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graphics Model.* **2002**, *20*, 281−295.
(41) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.−Act. Relat.* **1988**, *7*, 18−25.
(42) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.

JM060234E