

# Consensus Adaptation of Fields for Molecular Comparison (AFMoC) Models Incorporate Ligand and Receptor Conformational Variability into Tailor-made Scoring Functions

Benjamin Breu,<sup>†</sup> Katrin Silber,<sup>‡</sup> and Holger Gohlke<sup>\*,†</sup>

Department of Biological Sciences, J.W. Goethe-University, Max-von-Laue-Strasse 9, 60438 Frankfurt, Germany, and Department of Pharmacy, Philipps-University, Marburg, Germany

Received July 13, 2007

Taking into account dynamical behavior and/or structural inaccuracies of receptor–ligand systems becomes increasingly important in structure-based drug design. Here, we describe the development of consensus Adaptation of Fields for Molecular Comparison (AFMoC) (abbreviated as AFMoC<sup>con</sup>) models that account for multiple ligand conformations in an ensemble of protein configurations. Ligand and receptor conformational variability is considered in a “reverse”, protein-based CoMFA-type approach that results in a tailor-made scoring function. As an extension to the current AFMoC approach, AFMoC<sup>con</sup> applies partial-least-squares regression considering multimode binding and a variable influence on the model-based region selection to an extended descriptor matrix. The approach was validated on a dataset of 79 structurally diverse thrombin inhibitors, aligned either in an experimentally determined thrombin structure or superimpositions of three structurally diverse thrombin structures derived by homology modeling. Initially, robust AFMoC models could be obtained for the experimental ( $q^2 = 0.57$ ) and one of the modeled protein structures ( $q^2 = 0.61$ ). However, no relationship between the quality of the homology model and the AFMoC model could be observed, rendering the a priori choice of a single receptor structure a difficult task. Convincingly, a consensus AFMoC model based on the newly developed approach circumvents this problem and shows a comparable internal and external predictivity ( $q^2 = 0.61$ ) like the best model derived from conventional AFMoC. As further advantages, (i) the influence of the single receptor structure–ligand alignment (RSLA) on the AFMoC<sup>con</sup> model can be determined, (ii) there is no principal limitation regarding the number of different RSLA considered, and (iii) AFMoC<sup>con</sup> models can be interpreted in terms of contour plots that aid in proposing variations of the ligand structure to improve binding. We expect the AFMoC<sup>con</sup> approach also to be valuable in those cases where multiple experimental receptor conformations are given.

## INTRODUCTION

Structure-based drug design techniques are increasingly used to identify hits and optimize lead compounds,<sup>1</sup> driven by the insight that biological structures should guide the chemistry in the process of drug discovery.<sup>2</sup> Being able to predict ligand binding affinities accurately (and rapidly) is of central importance in this field.<sup>3–5</sup> Despite considerable efforts recently, however, commonly used scoring approaches still have difficulties in correctly rank-ordering structurally related compounds.<sup>6–8</sup> Among the possible reasons for these failures are, e.g., the inappropriate treatment of solvation effects, unconsidered protein flexibility, and ambiguities in protonation states.<sup>9</sup> One promising way to overcome the scoring problem for particular biological targets is to tailor generally applicable scoring functions by exploiting energetic and structural information about ligands already known to bind to the target. Prominent examples of thus-adapted scoring functions have been obtained by either correlating

force-field contributions or empirical terms with experimentally determined affinities.<sup>10–26</sup>

In the case that no structural information about the receptor is known, the field of drug discovery knows alternative techniques to predict binding affinities within a set of ligands, usually with a surprisingly high predictive power.<sup>27</sup> These are the different variants of three-dimensional quantitative structure–activity relationships (3D QSARs), such as CoMFA<sup>28</sup> or CoMSIA.<sup>29</sup> These approaches are increasingly also applied to cases where the corresponding receptor structure has been determined. Usually, however, protein information is only considered there to generate a meaningful alignment of the ligands<sup>30–32</sup> but not in the model derivation step. In contrast, in the recently developed AFMoC (Adaptation of Fields for Molecular Comparison) approach,<sup>33</sup> knowledge-based pair-potentials of DrugScore<sup>34</sup> are adapted in a protein-specific manner by applying techniques from comparative molecular field analysis. Thus, AFMoC is a truly protein-based or—with respect to the origin of the potential fields—“reverse” CoMFA approach that provides a link between the fields of tailor-made scoring functions and 3D QSARs.<sup>12</sup> Most importantly, AFMoC allows one to gradually move from general knowledge-based potentials to protein-specifically adapted potentials. This has been proven to be of major

\* Author to whom correspondence should be addressed. Phone: (+49) 69 798-29411. Fax: (+49) 69 798-29527. E-mail: gohlke@bioinformatik.uni-frankfurt.de.

<sup>†</sup> Department of Biological Sciences, J.W. Goethe-University, Frankfurt, Germany.

<sup>‡</sup> Department of Pharmacy, Philipps-University, Marburg, Germany.

importance in a recent application of AFMoC that demonstrates the predictive power of the method, even for structurally distinct inhibitors.<sup>35</sup> Furthermore, AFMoC-adapted potential fields have been used as an objective function to improve binding mode predictions in docking.<sup>36</sup>

So far, most of the structure-based or 3D QSAR-related approaches consider static (i.e., single conformation) representations of receptors and ligands. However, the need to account for the dynamical behavior of a receptor–ligand system and/or structural inaccuracies in the case of modeled binding partners has recently gained renewed support.<sup>37–39</sup> As such, in the realm of 3D QSARs, multiple ligand conformations for training sets of structure–activity data are used to incorporate conformational and alignment freedom by ensemble-averaging into the model development. This extends QSAR to four dimensions (4D QSAR).<sup>40–43</sup> Carried one step further, 5D QSAR models are derived if ligand-induced adaptations of the receptor binding site are also taken into consideration.<sup>40</sup> Similarly, in structure-based drug design, ensembles of protein configurations are used to either model protein flexibility in a discrete manner<sup>44,45</sup> or compensate the impact of structural inaccuracies of modeled receptors.<sup>46</sup> That way, multiple protein structures-based methods provide a promising alternative to those approaches that use a single protein structure but in a “soft” representation.<sup>47,48</sup>

In this study, we introduce an extension to the AFMoC approach that takes into account multiple ligand conformations in an ensemble of protein configurations. Hence, our approach implicitly considers receptor and ligand conformational freedom in a tailor-made scoring function. Consensus AFMoC models (AFMoC<sup>con</sup>) are derived from ensembles of modeled receptor structure–ligand alignments (RSLAs) by applying partial least-squares (PLS) analysis, considering multimode binding (MMB),<sup>43,49–51</sup> together with a variable influence on the model (VINFM)-based<sup>52</sup> region selection. Very convincingly, if tested on a dataset of 79 structurally diverse thrombin inhibitors, the consensus model shows internal and external predictivity that is comparable to the best model derived from a single RSLA. The AFMoC<sup>con</sup> approach not only yields accurate binding affinity predictions but also alleviates the need to choose “an appropriate” receptor structure among several alternatives. Thus, it also is expected to be valuable in those cases where multiple *experimental* receptor conformations are available.

## THEORY

**1. AFMoC Analysis.** The AFMoC approach generates protein-specifically adapted potential fields inside binding pockets that can be subsequently used for binding affinity predictions<sup>12,33,35</sup> and as the objective function in docking.<sup>36</sup> The underlying theory has been detailed elsewhere.<sup>33</sup> Here, we describe the approach briefly to introduce the terminology used below. The AFMoC approach consists of four steps: (I) potential field calculation and ligand alignment, (II) interaction field calculations, (III) correlating interaction field values with binding affinities and prediction of binding affinities, and (IV) binding affinity prediction for novel ligands.

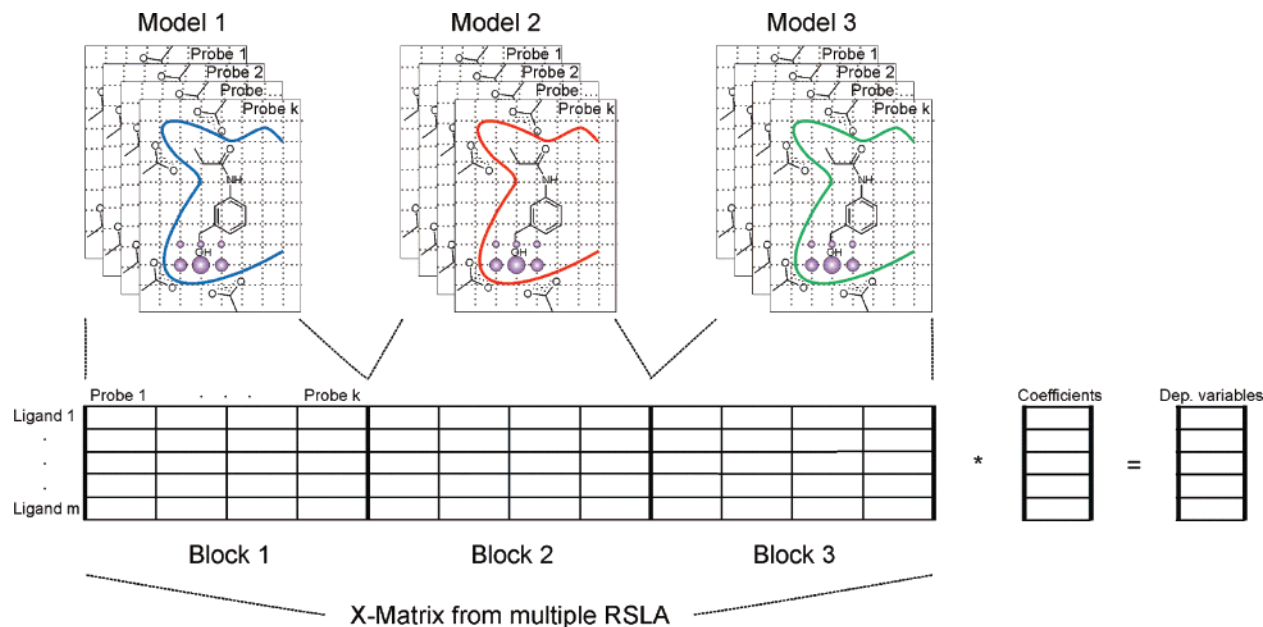
*(I) Potential Field Calculation and Ligand Alignment.* DrugScore<sup>34</sup> distance-dependent pair-potentials between a protein environment and ligand atom probes are mapped onto

the intersections of a regularly spaced grid that is placed inside the binding pocket, resulting in “potential fields”.<sup>53</sup> Although the DrugScore function has been shown to reliably recognize correct binding modes and predict binding affinities,<sup>7,54,55</sup> the underlying distance-dependent pair-potentials only depict an *averaged* representation of protein–ligand interactions derived from a large number of experimentally determined protein–ligand complexes. By including *structural* and *energetic* information about already known ligands of a protein, these generally valid potentials can be tailored in a protein-specific manner. For this, three dimensional (3D) structural protein information and a data set of mutually superimposed ligands inside the binding pocket are required. The ligand superimpositioning can be derived from crystallographically determined complexes,<sup>36</sup> docked ligand geometries,<sup>35</sup> or ligand structures modeled into the binding pocket using reference protein–ligand complexes as templates.<sup>33</sup>

*(II) Interaction Field Calculations.* “Interaction fields” for each ligand are generated from the “potential fields”, considering the structural information about the bound molecule. 3D Gaussian functions placed at the locations of the ligand atoms result in an atom type-specific, distance-dependent contribution of each atom, with respect to each grid point. Multiplying this contribution with the value of the potential field at each grid point finally “embeds” the ligand into the grid. Note that the thus-derived interaction fields are ligand-atom-type specific and mutually orthogonal to each other. This is in contrast to “generic” (e.g., electrostatic or steric) fields used in CoMFA or CoMSIA.<sup>28,29</sup>

*(III) Correlating Interaction Field Values with Binding Affinities and Prediction of Binding Affinities.* Interaction field values are correlated with experimental binding affinities of the embedded ligand using PLS analysis. This results in individual weighting factors for each field value. Together with the standard deviation of each interaction field column, the results of the AFMoC analysis can be displayed as “STDEV\*COEFF” contour maps that help to propose ligand modifications that optimize binding.

*(IV) Binding Affinity Prediction for Novel Ligands.* Finally, binding affinities of novel ligands, which have been treated as described previously in subsection II, can be predicted from a linear combination of the specifically adapted and generally valid interaction fields. Two further remarks are in order here:<sup>33</sup> First, only those interaction fields are subjected to PLS analysis for which ligand atoms of a given type are frequently present in the data set. Conversely, interaction fields for rarely present atom types are excluded in the regression analysis. Because these atoms also contribute to the binding affinities, their contributions are taken into account by subtracting their original DrugScore score ( $pK_i^{\text{Pair}}$ ) from the experimental affinity ( $pK_i$ ), leaving only  $pK_i^{\text{PLS}}$  as the dependent variable for PLS analysis ( $pK_i^{\text{PLS}} = pK_i - pK_i^{\text{Pair}}$ ). Consequently, the statistical results will always be described by two regression coefficients: one for the part of the binding affinity considered in PLS ( $pK_i^{\text{PLS}}$ ) and one for the total affinity ( $pK_i$ ). Second, AFMoC also allows the prediction of affinities of molecules that deviate structurally from those in the training set in two ways: (i) For ligand portions that exceed beyond the scope of the “trained” grid, a score based on the original DrugScore potentials will be used; because the DrugScore potentials



**Figure 1.** Generation of the X matrix that is used as input to the MBPLS regression. First, interaction fields of different ligand atom types (probes) are generated for ligands that are aligned inside the binding pockets of different protein conformations (top part). These fields are converted to one-dimensional vectors that are joined such that each receptor structure–ligand alignment (RSLA) yields one block in the matrix (bottom). Stacking of these vectors for each ligand yields the final X matrix (superblock) that is correlated with experimental binding affinities ( $pK_i$ ).

have an upper bound of 6 Å, parts of the ligand that are further apart than that from the receptor will not be considered. (ii) Similarly, original DrugScore potentials are also used if ligand portions coincide with regions of the grid where the variance of information used for training has been too scarce to be considered in PLS and, hence, no adapted fields exist. Both approaches are based on the assumption that original DrugScore performs better than simply ignoring such situations.

**2. Generating Consensus AFMoC Models. 2.1. Overview.** The aforementioned observations have prompted us to extend the current AFMoC approach in this study to also consider ligand and receptor conformational freedom. As a result, consensus AFMoC models (AFMoC<sup>con</sup>) are derived from interaction fields that have been generated from multiple ligand superimpositions inside conformationally different binding pockets (RSLA, or receptor structure–ligand alignment). For this, a PLS regression analysis considering MMB<sup>43,49–51</sup> was used. Furthermore, a VINFM-based region selection<sup>52</sup> proved valuable to yield models with improved external predictivity. The original interaction fields, together with the specifically adapted ones, can then be used to score novel ligands in the realm of the consensus model. Finally, we describe how the results of the AFMoC<sup>con</sup> analysis can be translated into contour plots that aid in proposing ligand modifications for improved binding affinity. These steps will be detailed in the following paragraphs.

**2.2. X Matrix Generation from Multiple RSLA.** In the final step of the AFMoC analysis (see previous discussion), interaction field value-based descriptors are correlated with experimental binding affinities, which results in individual weighting factors for each field value. Initially, the 3D interaction fields are rearranged as one-dimensional (1D) row vectors, with each vector consisting of  $n = k \times g$  columns (where  $k$  is the number of ligand-atom types used to generate the interaction fields and  $g$  is the number of grid points).

Subsequently, these vectors are used to build a two-dimensional (2D) X matrix, where each row contains the interaction field values of one ligand.

In the case of multiple RSLA, this X matrix generation step is extended in the following way (Figure 1). First, the protein structures are mutually superimposed to put them into a common reference frame. Second, independent ligand alignments are generated inside each binding pocket, and interaction fields are calculated for each of these alignments as in the single RSLA case. The X matrix is then generated by adding, side by side, the row vectors obtained for each RSLA, resulting in  $n' = b \times n$  columns (where  $b$  is the number of different RSLAs). Finally, the extended matrix is used for correlation with experimental binding affinities (see later discussion).

Formulating the problem this way, each grid point of an interaction field of a given ligand-atom type is represented by  $b$  field values. Hence, in the correlation step, a binding pocket region can be characterized by multiple equivalent (yet not identical) descriptors for which individual weighting factors are obtained. This leads to a model that includes the information of all RSLA. Note that there is no conceptual limitation, with respect to the number of different RSLA that can be used to derive the extended X matrix.

**2.3. PLS Regression Analysis Considering Multimode Binding.** Each of the RSLAs describes different binding modes of the ligands in structurally varying receptors, where a binding mode  $j$  of ligand  $i$  is characterized by a partial association constant  $K_{ij}$  (thereby assuming that only one ligand can bind to a binding pocket and that only one class of binding pockets exists). Considering that the total concentration of receptor/ligand complexes is equal to the sum of concentrations of the complexes in which the ligand binds in different binding modes, the overall association constant is<sup>49–51</sup>



$$K_i = \sum_j K_{ij} \quad (1)$$

An intuitive way to correlate experimental binding affinities with the interaction field descriptors then seems to be application of the one-mode AFMoC approach also to the extended  $X$  matrix. This results in a weighted average of the interaction fields of individual RSLA, expressing, according to the classical QSAR equation, the logarithm of the overall association constant  $\ln K_i$  as weighted average of  $\ln K_{ij}$ .<sup>56</sup> Obviously, this yields  $K_i = \prod_j K_{ij}^{w_j}$  (where  $w_j$  represents the weights of the individual modes), which contradicts eq 1 and, thus, lacks a physical basis.<sup>43</sup>

Hence, we followed an alternative route that is based on an approach recently introduced by Balaz and co-workers for considering MMB in CoMFA.<sup>43,57</sup> Here, the partial association constant of a ligand binding in a particular mode is expressed in terms of the interaction field values, according to<sup>28,50</sup>

$$K_{ij} = \exp(c_0 + \sum_{k=1}^{f \times g} c_{jk} X_{ijk}) \quad (2)$$

where  $f$  is the number of field types,  $g$  is the number of grid points,  $c_0$  is the intercept, and  $c_{jk}$  represents the regression coefficients. Equation 2 can be further extended by other descriptors, such as ligand conformational energy or entropy. In the case of one-mode AFMoC, eq 2 is logarithmized, resulting in the classical QSAR equation that is linear in the regression coefficients  $c_{jk}$ . In the case of MMB, however, eq 2 is inserted into eq 1, yielding an equation that is nonlinear with regard to  $c_{jk}$ . This equation can be solved by nonlinear regression as long as the number of regression coefficients is smaller than the number of ligands considered.<sup>57</sup> However, in the case of AFMoC, considering MMB, the number of regression coefficients is much higher than the number of ligands, which requires the use of PLS regression to cope with this situation. For this, we applied a linearized version of eq 1 in association with eq 2:<sup>43</sup>

$$1 - \sum_j \frac{K_{ij}}{K_i} + \sum_j \frac{K_{ij}}{K_i} \times \ln(K_{ij}) = c_0 \sum_j \frac{K_{ij}}{K_i} + \sum_j \frac{K_{ij}}{K_i} \sum_{k=1}^{f \times g} c_{jk} X_{ijk} \quad (3)$$

Because the partial association constants  $K_{ij}$  in eq 3 themselves are dependent on the regression coefficients and the interaction field values, eq 3 is solved in an iterative manner: from an initial set of  $c_{jk}$ ,  $K_{ij}$  are calculated according to eq 2. With these, the variables of eq 3 are updated and a new set of  $c_{jk}$  is calculated from eq 3 by PLS. This procedure is repeated until self-consistency occurs, resulting in a stable set of optimized  $c_{jk}$ . At this stage, the ratio

$$\text{prevalence}_{ij} = \frac{K_{ij}}{K_i} \quad (4)$$

represents the prevalence of the  $j$ th binding mode of ligand  $i$  (i.e., the ratio of the number of ligand molecules bound in that particular mode to the total number of bound ligand

molecules). We note that, in contrast to refs 43 and 57, no variable selection procedure was applied in this step. Instead, the number of optimal components was determined by leave-one-out (LOO) cross-validation runs, applying SAMPLS<sup>58</sup> to eq 3 prior to each PLS analysis.

**2.4. Consensus Model Generation by VINFM-Based Variable Selection.** Initial PLS analyses considering MMB on extended  $X$  matrices (comprising interaction field values of different combinations of RSLA) yielded, in some cases, models with inferior predictive power, compared to AFMoC analyses based on the best single model or the crystal structure (see the Results and Discussion section). Although difficult to quantitate, we conjectured, as a reason for this, that solely combining interaction fields from multiple RSLA decreased the signal-to-noise ratio. A similar explanation was given for the observation that decreasing the grid spacing from 2 Å to 1 Å may yield lower  $q^2$  values in CoMFA analysis, particularly in the case of datasets of diverse structures.<sup>59</sup> Along these lines, although there is a small risk of chance correlation in PLS, it is well-known that including irrelevant variables into the  $X$  matrix can cause detrimental effects on the selection of a 3D QSAR model by PLS.<sup>59</sup> Therefore, variable selection techniques have been devised whose intent is to select only those variables that have a significant effect on the response to be correlated.<sup>31,60,61</sup> So far, however, these techniques have been mostly applied to matrices that do not have a block structure.

In contrast, the extended  $X$  matrix resulting from the combination of interaction fields of multiple RSLA has a meaningful block structure. In particular, the fact that sets  $C_j$  (eq 5) of equivalent columns of the blocks,

$$C_j = \{\text{column}_k \mid k = j + l \cdot n; 1 \leq j \leq n; 0 \leq l \leq b - 1\} \quad (5)$$

describe interaction values at identical points of the grid prompted us to suggest the following three-step variable selection strategy (Figure 2):

(I) A model is derived using all  $n' = b \times n$  variables of the matrix, as described in the previous section.

(II) Considering all sets  $C_j$  of  $b$  equivalent columns in turn, only that column of a set with the largest contribution to the model is retained. The  $(b - 1)$  other columns of the set are discarded.

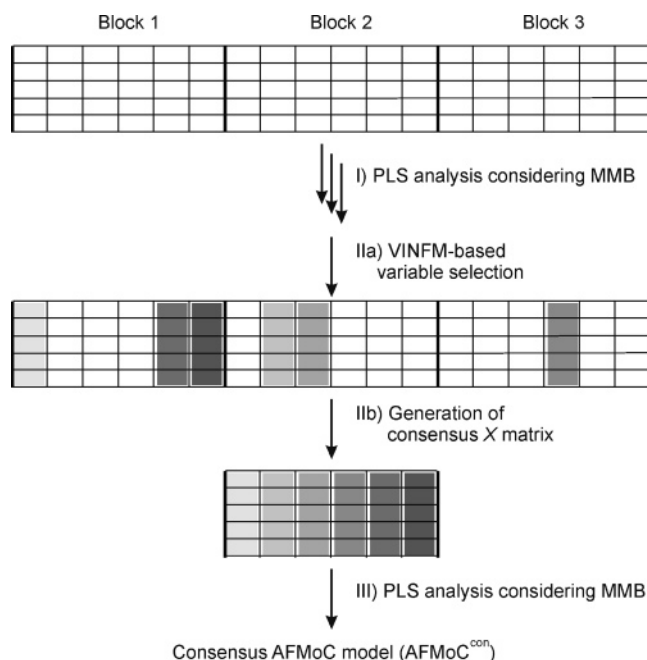
(III) A final model is then derived using only the  $n$  retained variables. Consequently, the final model uses one out of  $b$  interaction field values per ligand atom type per grid point and, hence, can be considered to be a consensus model of the combined blocks (AFMoC<sup>con</sup>).

To measure the contribution of column  $j$  to the model, we apply the variable influence on the model index (VINFM;<sub>j</sub>; eq 6) that was already used previously to reduce the dimensionality of the  $X$  matrix.<sup>52</sup>

$$\text{VINFM}_j = \sum_{i=1}^h \text{ESS}_i \cdot c_{ij}^2 \quad (6)$$

Here, the sum runs over all latent variables (PLS dimensions)  $h$  used,  $\text{ESS}_i$  is the percentage explained sum of squares of dimension  $i$ , and  $c_{ij}$  is the weight of the column.

**2.5. Prediction of Binding Affinities.** To predict the binding affinities of novel ligands, the molecules are docked into all



**Figure 2.** Variation influence on the model (VINFM)-based variable selection: (I) A normal PLS analysis is performed on the full  $X$  matrix. (IIa) Using the VINFM index (eq 6), only that column of a set of  $b$  equivalent columns with the largest contribution to the model is retained. The  $b - 1$  other columns of the set are discarded, resulting in a consensus  $X$  matrix (IIb). (III) A consensus model is then derived using only the  $n$  selected variables (AFMoC<sup>con</sup>).

$b$  binding pockets and interaction fields are calculated following the same protocol as that applied to the training molecules. The (centered and scaled) interaction field values for each molecule are then weighted with the regression coefficients obtained by PLS at every grid point according to eqs 1 and 2, which yields the binding affinity contribution  $pK_i^{\text{PLS}}$  of the protein-specifically adapted fields. This procedure holds, irrespective of whether the full extended  $X$  matrix or the one obtained after VINFM-based variable selection is used.

As in the case of the original AFMoC approach, additional contributions  $pK_i^{\text{Pair}}$  that are due to those ligand-atom types that were not considered in the regression analysis are determined from the original DrugScore methodology. These contributions are obtained separately for each of the  $b$  protein structure–ligand configurations. Weighting each of these contributions by the inverse of the number of RSLA ( $1/b$ ) thus averages over all protein structures/models used and results in a  $pK_i^{\text{Pair}}$  contribution that is consistent with the consensus approach pursued here. Similarly, contributions of ligand portions are averaged based on original DrugScore potentials that exceed beyond the scope of the “trained” grid or that lie in those regions where the variance of information used for training has been too scarce to be considered in PLS. This allows one to predict the affinities of molecules that deviate structurally from those in the training set in AFMoC<sup>con</sup>.

Finally, we note that  $pK_i$  predictions by protein-specifically adapted consensus fields can be linearly “mixed” with those based on original DrugScore ones, as in the original AFMoC approach (see eq 4 in ref 33). This reveals better affinity predictions in those cases where small training sets of closely

related ligands or affinity data not sufficiently spread were used for adaptation. However, this option was not further explored in this study; all predictions are based solely on the adapted fields.

**2.6. Graphical Interpretation of AFMoC<sup>con</sup> Results.** For proposing structural modifications of already known ligands that influence binding, a graphical interpretation of the derived AFMoC<sup>con</sup> model can be obtained in a straightforward manner. Generally, the product of the standard deviation of the considered interaction field multiplied by the coefficient of the QSAR equation (“STDEV\*COEFF”) is calculated at every grid point and displayed as a contour map. Because it is known which column of the consensus  $X$  matrix (obtained after VINFM-based variable selection (Figure 2)) originated from which single RSLA, the STDEV\*COEFF fields can be displayed separately for each of these RSLA in the AFMoC<sup>con</sup> case. This allows one to visualize the influence of each RSLA on the consensus model and, in turn, to relate favorable or unfavorable interaction field contributions to ligand affinities to characteristics of the particular protein structure.

## METHODS

**1. Datasets and Alignments.** Our approach has been validated using a data set of 80 serine protease inhibitors that was used previously to study the selectivity features of this protein class.<sup>62</sup> The data set is an extension to a set used in an earlier study,<sup>63</sup> both in terms of structural diversity and spread in the affinity data. Although all ligands contain the benzamidine moiety, the data set can be divided into 11 different structural classes. As such, ligands that are derived from TAPAP, TIPPS, and NAPAP structures are contained in the set, as are benzamidine, naphthamidine, bisbenzamidine, glycinic acid, butanoic acid, and valeric acid derivatives. In initial tests, one of the ligands in the original data set of 80 notoriously resulted in bad affinity predictions and was discarded for that reason. The dataset was then split into 64 compounds for training and 15 compounds for testing (see Table 1 in the Supporting Information). For the training set, the  $pK_i$  values were spread over a range of 5.5 logarithm units and showed a balanced distribution, albeit the range between  $pK_i = 3.0$  and  $pK_i = 4.5$  was under-represented. The test set was chosen such that each ligand structural class was represented at least once and that the range of  $pK_i$  values covers almost 5.0 logarithm units. For all compounds, standard protonation states are assumed, i.e., carboxylate groups are considered to be deprotonated, and aliphatic amino and benzamidino groups are considered to be protonated. Possible changes in the protonation states upon binding to the protein were neglected. Finally, for CoMFA and CoMSIA analyses, AM1-BCC charges<sup>64</sup> were assigned to each inhibitor, using the antechamber program from AMBER 8.<sup>65,66</sup>

A protein-based alignment was applied in all cases to obtain a consistent superimposition of all inhibitor molecules. Initially, binding-mode templates for some of the ligand classes were obtained from crystallographically determined protein–ligand complexes (Protein Databank (PDB) codes: 1ets, 1ett, 1dwd, 1aht). After superimposing these complexes with respect to the residues of the catalytic triad (His57, Asp102, Ser195) and Asp189 of the specificity pocket, the remaining serine protease inhibitors were constructed and

**Table 1.** Root-Mean-Square Deviations of Homology Models (HMs), with Respect to the Experimental Thrombin Structure and Each Other<sup>a</sup>

	Root-Mean-Square Deviation, rmsd (Å)					
	HM 1		HM 2		HM 3	
model	all <sup>a</sup>	within 6 Å <sup>b</sup>	all <sup>a</sup>	within 6 Å <sup>b</sup>	all <sup>a</sup>	within 6 Å <sup>b</sup>
1ets <sup>c</sup>	2.04	1.48 (2.90)	1.91	1.10 (2.35)	1.97	1.44 (3.37)
HM 1			0.70	1.05 (1.88)	0.42	0.80 (2.37)
HM 2					0.51	0.94 (2.32)

<sup>a</sup> All C<sub>α</sub> atoms were used for the calculation. <sup>b</sup> Only C<sub>α</sub> atoms (non-hydrogen atoms) within 6 Å of the binding pocket were used. <sup>c</sup> The protein structure of PDB entry 1ets was used for comparison.

superimposed manually inside the binding pocket. Finally, all ligands were minimized in the rigid binding pocket of thrombin (protein structure taken from PDB code 1ets), using the MAB force field, as implemented in MOLOC.<sup>67</sup> This superimposition served as a reference to determine the predictive power of AFMoC in the case of a crystallographically determined protein structure (as compared to homology modeling-based structures, see below). Similarly, CoMFA and CoMSIA analyses were performed with the thus-aligned molecules for comparison.

**2. Homology Modeling.** Homology models (hereafter abbreviated as HMs) of thrombin were generated using MODELLER<sup>68</sup> and template structures of protein C (PDB code: 1aut), factor Xa (1hcg), kallikrein (1hia), and  $\beta$ -trypsin (1tp), following ref 46. The crystal structures were resolved to 1.4–2.8 Å. The sequence identity, with respect to thrombin (1ets), amounts to 32%–40% in the case of all residues and 41%–52% in the case of binding site residues only. An initial alignment of the template sequences with that of thrombin was obtained using ClustalW.<sup>69</sup> The alignment was manually corrected such that no gaps occurred in secondary structure regions of the templates. Secondary structure elements were determined with the dssp method.<sup>70</sup> We note that no information about the structure of thrombin was included to obtain the alignment, to comply with a “real-life” scenario.

To model the thrombin structure, all four template structures were considered simultaneously by MODELLER,<sup>68</sup> which allows one to include information about the differing features of the protein family. This has been shown to improve the quality of the models, compared to modeling based on a single template.<sup>46</sup> An initial ensemble of models scored best by MODELLER's energy function was then generated using default settings of MODELLER. That way, different regions of conformational space of the target are explored, and a model that best suits one's needs can be chosen. Although large parts of the thrombin binding pocket are structurally conserved, with respect to the template structures, a particular difficulty arises from modeling the so-called 60-loop that is responsible for the restricted substrate and inhibitor specificity of thrombin. With 8–11 insertion residues, compared to the template structures, the length of the loop is located at the upper bound, up to which loop modeling methods have a reasonable chance of predicting conformations that superimpose well on the native structure.<sup>71</sup> To this end, almost 1000 models were generated by extensively sampling the conformational space of the 60-loop. Because our objective here is to generate models that (i) deviate moderately from the experimental thrombin structure and (ii) structurally vary with respect to each other, three models were finally chosen for the generation of AFMoC models that fulfill these criteria. The root-mean-

square deviation (rmsd) values of these models, with respect to each other and the experimental thrombin structure, are given in Table 1. The generated models were validated using PROCHECK<sup>72</sup> and PROSA.<sup>73</sup> For the backbone torsion angles, Ramachandran plots reveal that the modeled structures have more residues in most favored regions (84%–87%) than the experimental thrombin structure (79%) and only as many residues in disallowed regions (<1.5%). In addition, to distinguish good contacts between amino acids from bad ones, residue interaction energies are calculated based on knowledge-based mean fields.<sup>72</sup> These energy graphs show that all but two regions of the models agree, with respect to the energy values of the experimental structure. With respect to residues 149–159, which reveal marked disfavorable interaction energies, we do not expect any detrimental influence on our results, because this region is not involved in interactions to the ligands. In turn, residues 60B–63, which contribute to ligand binding, show slightly positive interaction energies, which demonstrates the difficulty of correctly modeling this loop region. Nevertheless, we are convinced that the generated models are of sufficient quality to be used in AFMoC analyses.

The HMs were superimposed, with respect to each other (and the experimental thrombin structure), based on the C<sub>α</sub> atoms of residue Asp189 and those of the catalytic triade. The ligand alignment obtained for the experimental thrombin structure (see previous discussion) was then adapted according to the structural requirements of the modeled binding pockets, with a final energy minimization in the rigid pocket of each model as described previously.

**3. AFMoC Analysis.** Interaction fields for the superimposed inhibitors were calculated as described elsewhere<sup>33</sup> using either the protein structure of 1ets or the homology models as references. The size of the grid box was selected in a way that all inhibitor molecules were sufficiently embedded with a margin of at least 4 Å. In all cases, grids of identical locations and dimensions were used. Fields were calculated for six inhibitor-atom types (in SYBYL<sup>74</sup> atom type notation: C.3, C.ar, O.3, O.2, O.co2, N.am) that, to our knowledge, represent the interaction differences of the considered ligands best. In fact, models with inferior quality resulted when different combinations of fields were used (see Table 2 in the Supporting Information). To study the influence of the grid spacing on the quality of the AFMoC models, values of 0.5 Å and 1.0 Å were tested. However, unless stated otherwise, results are reported for a grid spacing of 1.0 Å. To be able to calculate interactions even in the vicinity of protein atoms, where the knowledge-based pair-potentials are not defined due to missing experimental data, an artificial repulsion term based on a Gaussian function was added to each potential (see ref 33 for details). In agreement



with an earlier study,<sup>75</sup> a value of 10 has been selected for the height of the repulsion function at the origin of an atom–atom contact. This value has been adjusted to be of the same order of magnitude as the largest absolute values of the pair-potentials. To estimate  $pK_i^{\text{pair}}$  values, DrugScore pair-potential values were scaled by  $-3.11 \times 10^{-2}$ , as was done in a previous study.<sup>33</sup>

In the case of one-mode binding, interaction field values were correlated with experimental affinities using PLS<sup>76,77</sup> similar to CoMFA.<sup>28</sup> In the case of MMB, interaction field values were correlated iteratively with experimentally determined binding affinities as described previously (see eqs 1–3). The prevalences of  $1/b$  were taken as initial estimates to determine the variables in eq 3. Using various other starting points always resulted in very similar final solutions (data not shown), indicating a large radius of convergence of the procedure. The convergence of the iterative scheme was monitored by the rmsd between  $\ln K_i$  values of  $N$  ligands calculated via eq 1 for subsequent iteration steps  $k$  and  $k - 1$ :

$$\text{convergence} = \sqrt{\sum \frac{(\ln K_i^k - \ln K_i^{k-1})^2}{N}} \quad (7)$$

All statistical analyses were performed using an in-house implementation of the (SAM)PLS algorithms.<sup>58</sup> The columns of the descriptor matrix were centered on input. Because interaction field values are obtained by applying DrugScore pair-potentials, which were already mutually weighted upon derivation of the potentials, neither the columns nor the interaction fields in total were scaled. To determine the optimal number of components (in each step of the iterative scheme), cross-validation runs were performed prior to the PLS analysis by means of the LOO procedure, using the enhanced SAMPLS method, which results in the cross-validated  $r^2$  ( $q^2$ ). Following the recommendations of Wold,<sup>76</sup> Kubinyi et al.,<sup>78</sup> and Bush et al.,<sup>58</sup> the columns of the descriptor matrix were re-centered for every new cross-validation run. The optimal number was determined as the one that resulted in a final increase of the  $q^2$  value by 5%. The final equations are characterized by  $q^2$ ,  $s_{\text{PRESS}}$ ,  $r^2$ ,  $S$ , and *contribution* values (which are computed as defined in ref 28) or *averages over prevalences* for individual RSLA. Statistical results are summarized in Tables 2 and 3.

To apply an even more rigorous statistical test, “leave-multiple-out” (LMO) cross-validations were performed. Thereby, 10 randomly chosen molecules of the training set were excluded from model derivation, the model was derived with the remaining data points, and the binding affinities of the excluded molecules predicted with the thus-derived model. This step is repeated until every element of the dataset was (on average) excluded once, resulting in a  $q_{\text{L100}}^2$  value. The  $q_{\text{L100}}^2$  values are finally averaged over 100 repetitions of the entire LMO procedure in the case of monomode binding. In the case of MMB, only one LMO cross-validation was performed, because of the computational burden (see Table 3 in the Supporting Information). In addition, to detect possible chance correlations, the biological data was randomly scrambled and model calculations were repeated. Only negative  $q^2$  values were obtained in this case (see Table 4 in the Supporting Information), which supports the assump-

**Table 2.** Statistical Results for AFMoC Analyses Using the Experimental Thrombin Structure<sup>a</sup>

parameter	0.5 Å spacing	1.0 Å spacing
$q^2$ <sup>b,c</sup>	0.57 (0.54)	0.57 (0.55)
$s_{\text{press}}$ <sup>d,e</sup>	0.83	0.82
correlation coefficient, $r^2$ <sup>b</sup>	0.78 (0.76)	0.77 (0.76)
$S$ <sup>d,f</sup>	0.60	0.60
Fischer’s $F$ -value <sup>b</sup>	105.1 (97.9)	104.2 (97.1)
number of components	2	2
fraction		
C.3	0.377	0.379
C.ar	0.562	0.559
O.3	0.013	0.013
O.2	0.023	0.024
O.co2	0.003	0.003
N.am	0.021	0.022

<sup>a</sup> Models are based on PLS calculations using a  $\sigma$  value of 0.7 Å and a grid spacing of 0.5 and 1.0 Å, respectively. <sup>b</sup> Values are given considering only  $pK_i^{\text{PLS}}$  or considering  $pK_i^{\text{total}}$  (values in parentheses). <sup>c</sup>  $q^2 = 1 - \text{PRESS}/\text{SSD}$ , as obtained by “leave-one-out” (LOO) cross validation. PRESS equals the sum of squared differences between predicted and experimentally determined binding affinities, SSD is the sum of the squared differences between experimentally determined binding affinities and the mean of the training set binding affinities. <sup>d</sup> In logarithmic units. <sup>e</sup>  $s_{\text{PRESS}} = \sqrt{\text{PRESS}/(n-h-1)}$  as obtained by LOO cross validation.  $n$  equals the number of data points,  $h$  is the number of components. <sup>f</sup>  $S = \sqrt{\text{RSS}/(n-h-1)}$ . RSS equals the sum of squared differences between fitted and experimentally determined binding affinities.

**Table 3.** Statistical Parameters for the Prediction of Binding Affinities of the Test Set Compounds

model	correlation coefficient, $r^2$ <sup>a</sup>	standard deviation, SD <sup>b</sup>
lets <sup>c</sup>	0.65	0.75
HM 1	0.72	0.67
HM 2	0.38	0.99
HM 3	0.36	1.01
HM 1 + 2	0.64	0.77
HM 1 + 3	0.61	0.79
HM 2 + 3	0.41	0.97
HM 1 + 2 + 3	0.64	0.75

<sup>a</sup> Squared correlation coefficient between experimentally determined and predicted  $pK_i$  values of the test-set compounds. <sup>b</sup> Standard deviation between experimentally determined and predicted  $pK_i$  values for the test set compounds. In logarithmic units. <sup>c</sup> AFMoC analysis using the experimental thrombin structure.

tion that meaningful correlations with the biological data are given in the nonscrambled case.

When calculating the interaction fields, protein-to-inhibitor interactions are distributed over neighboring grid points using a distance-dependent Gaussian function. The half-width of this function determines the “local smearing” of the interactions. Smaller  $\sigma$  values result in a more locally restricted mapping, whereas larger  $\sigma$  values have a tendency to average adjacent interactions. To study the influence of  $\sigma$ , values of 0.55, 0.7, ..., 1.15 Å were used to calculate interaction fields.

For comparison, CoMFA and CoMSIA analysis as implemented in SYBYL<sup>74</sup> were performed using the ligand alignment obtained with the crystal structure of thrombin (see Table 4). These calculations were performed with SYBYL standard parameters, using steric and electrostatic fields in the case of CoMFA and, in addition, hydrophobic, H-bond donor and acceptor fields in the case of CoMSIA.

**Table 4.** Statistical Results for CoMFA and CoMSIA Analyses<sup>a</sup>

parameter	CoMFA method	CoMSIA method
$q^2$ <sup>b</sup>	0.58	0.57
$s_{\text{press}}$ <sup>c,d</sup>	0.80	0.82
correlation coefficient, $r^2$	0.86	0.86
$S$ <sup>c,e</sup>	0.47	0.46
Fischer's $F$ -value	120.0	91.1
number of components	3	4
fraction		
steric	0.492	0.157
electrostatic	0.508	0.230
hydrophobic		0.273
H-bond donor		0.139
H-bond acceptor		0.202
$r^2$ <sup>f</sup>	0.78	0.70
standard deviation, SD <sup>g</sup>	0.41	0.42

<sup>a</sup> Models are based on CoMFA and CoMSIA analysis as implemented in SYBYL.<sup>74</sup> SYBYL default settings have been applied in both cases. <sup>b</sup>  $q^2 = 1 - \text{PRESS}/\text{SSD}$  as obtained by "leave-one-out" (LOO) cross validation. PRESS equals the sum of squared differences between predicted and experimentally determined binding affinities, SSD is the sum of the squared differences between experimentally determined binding affinities and the mean of the training set binding affinities. <sup>c</sup> In logarithmic units. <sup>d</sup>  $s_{\text{PRESS}} = \sqrt{\text{PRESS}/(n-h-1)}$  as obtained by LOO cross validation.  $n$  equals the number of data points,  $h$  is the number of components. <sup>e</sup>  $S = \sqrt{\text{RSS}/(n-h-1)}$ . RSS equals the sum of squared differences between fitted and experimentally determined binding affinities. <sup>f</sup> Squared correlation coefficient between experimentally determined and predicted  $\text{p}K_i$  values for the test set compounds. <sup>g</sup> Standard deviation between experimentally determined and predicted  $\text{p}K_i$  values for the test-set compounds. In logarithmic units.

## RESULTS AND DISCUSSION

**1. Datasets and Alignment Procedure.** The consensus AFMoC approach (AFMoC<sup>con</sup>) has been validated on a data set of 79 thrombin inhibitors<sup>62</sup> that was used before by Böhm and Klebe for selectivity studies on serine proteases.<sup>63</sup> The  $\text{p}K_i$  values of the 64 training compounds are spread over a satisfactorily large range of 5.5 logarithm units; the test set is comprised of 15 compounds with a similar  $\text{p}K_i$  range (see Table 1 in Supporting Information). The binding affinities have been determined to have an average error of 10%–15%, with maximal deviations in the range of 25%.<sup>79</sup> All compounds contain the benzamidine moiety that interacts with Asp189 in the S1 pocket. With respect to the molecular substituents that occupy the S1', S2, S3, and S4 pockets of thrombin, the data set can be divided into 11 structural classes. Yet, all ligands are accommodated in the same region of the binding pocket, as indicated by the experimentally observed binding modes that are available for most of the ligand structural classes. Because of the considerable structural heterogeneity and the large range of affinities, we consider this data set to be challenging for the derivation of robust structure–activity relationships.

A relative mutual superimposition of all ligands has been obtained using the protein binding pockets of 1ets or the HMs as constraints (see Figure 3). In many cases, protein structure-based alignments have been shown to produce superior models, compared to purely ligand-based alignments in 3D QSAR applications.<sup>30,80</sup> The aforementioned crystallographically determined protein–ligand complexes were used as binding-mode templates to which the other molecules are aligned manually. A final geometry optimization removes unfavorable intermolecular and intramolecular interactions.

In our hands, alignments obtained this way yielded better models than if ligand geometries that were generated by automated docking had been used.<sup>33</sup>

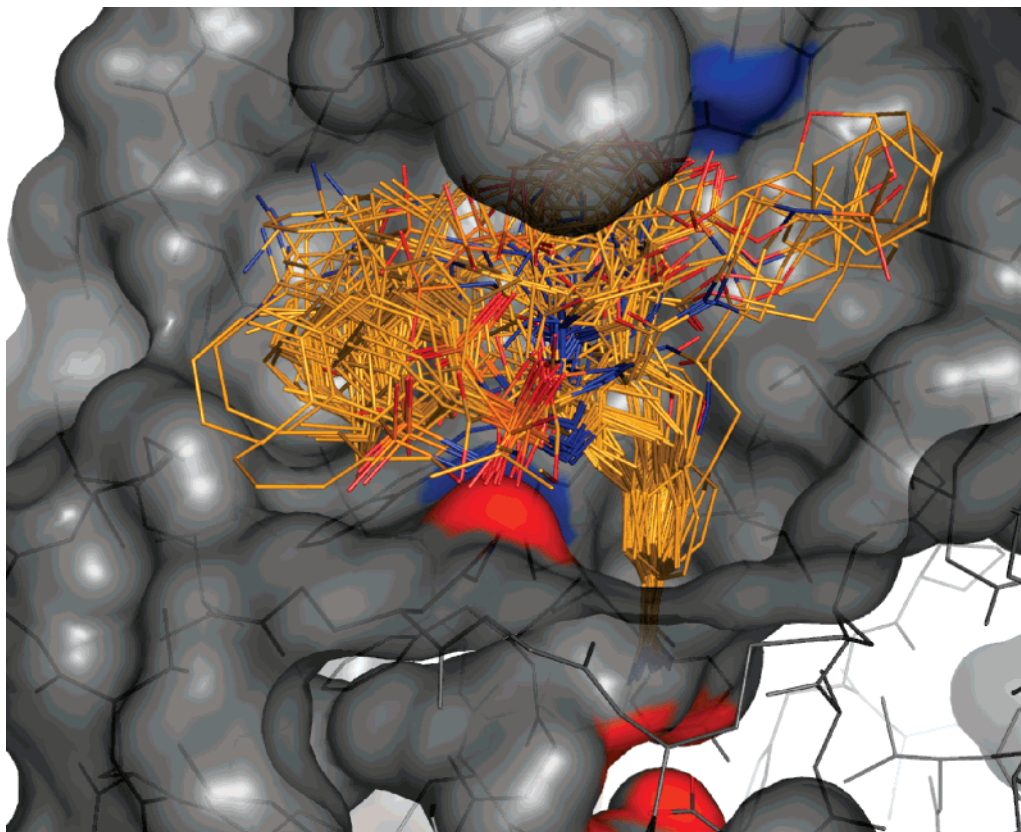
**2. Significance and Robustness of AFMoC Analyses Using the Experimental Protein Structure, and Comparison to CoMFA and CoMSIA Analyses.** In Table 2, the statistical results of AFMoC analyses are summarized with respect to grid spacings of 0.5 and 1.0 Å, using the ligand superimposition obtained in the protein structure of 1ets. Generally, models are accepted to be significant if  $q^2 > 0.5$ , whereas models are considered "good" if  $q^2 > 0.3$ ,<sup>81</sup> although a high value of  $q^2$  seems to be rather a necessary (but not sufficient) condition for a model to have a high predictive power<sup>82,83</sup> (see later discussion). For both grid spacings, a value of  $q^2 = 0.57$  has been obtained for the part of  $\text{p}K_i$  being adapted during PLS analysis, while considering the total binding affinity still results in a value of  $q^2 = 0.54$  or 0.55. The  $r^2$  values (0.78 or 0.77, respectively) are only moderately larger than the  $q^2$  values, providing an indication that the models have not been overfitted. This is corroborated by the small number of PLS components (in both cases, this number is 2) selected that have been chosen such that the addition of the final component to the model still resulted in at least a 5% increase in the  $q^2$  values.

Because the dependency on the grid spacing is negligible, a value of 1.0 Å has been used for further calculations. In this case, for leave-10-out cross validations, a value of  $q_{L100}^2 = 0.56$  was obtained, averaged over 100 repetitions of the procedures (see Table S3 in the Supporting Information). Furthermore, no reasonable models ( $q^2 < 0$ ; Table S4 in the Supporting Information) were obtained when randomly scrambled biological data were used as independent variables. This provides another indication for the significance of the models obtained with the nonscrambled  $\text{p}K_i$  values. The plot of predicted versus actual binding affinities for the PLS analysis (Figure 4) does not reveal any significant overprediction or underprediction across the entire range of activity. Furthermore, no trend can be observed for residuals across different classes of compounds. Thus, both observations suggest that the AFMoC model represents the entire dataset of molecules.

The dependence of the  $q^2$  value on the variation of the  $\sigma$  value that governs the degree of "smearing" protein–ligand interactions across neighboring grid points was investigated. Almost no variation of  $q^2$  is found for  $\sigma$  values in the range of 0.55–1.15 Å (data not shown). Hence, in agreement with recent studies,<sup>33,36</sup> a  $\sigma$  value of 0.7 Å was chosen for all calculations. This value indicates that particularly local interactions are considered by our approach when generating the interaction fields, as an interaction decreases to 21% of its original value across a distance of 1 Å for  $\sigma = 0.7$  Å.<sup>33</sup>

Table 2 lists contributions by interaction fields to explain binding affinity differences. The largest contributions originate from C.ar and C.3 fields, whereas polar interaction fields show contributions of more than an order of magnitude smaller, with O.co2 contributing the least. This finding may be explained by the frequency of occurrence for these atom types in the ligand data set. As tests of the influence of different interaction field combinations on the  $q^2$  value indicate (see Table S2 in the Supporting Information), the combination of six fields (C.3, C.ar, O.3, O.2, O.co2, N.am)





**Figure 3.** Alignment of 79 serine protease inhibitors in the binding pocket of PDB entry 1ets. All figures in this study showing molecular representations were generated with Pymol.<sup>91</sup>

finally chosen to derive the AFMoC models in this study yields the best  $q^2$  value for the part of binding affinity considered in PLS ( $pK_i^{\text{PLS}}$ ).

For comparison, CoMFA and CoMSIA analyses were performed as implemented in SYBYL<sup>74</sup> using the same ligand alignment as that used in the case of the AFMoC analyses. The statistical results summarized in Table 4 show that the  $q^2$  values obtained by both methods are comparable to those computed by AFMoC (see Table 2) but are considerably larger than reported in a previous study.<sup>62</sup>

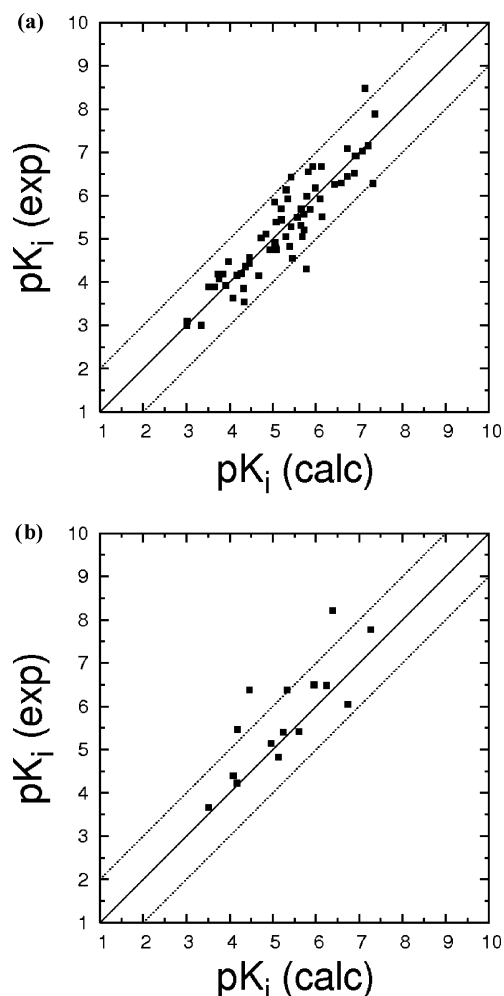
Recently, the sole use of cross-validation to determine the predictive power of QSAR models has been criticized, based on the notion that, for many datasets, no correlation between  $q^2$  values for the training set and the predictive ability for the test set was found.<sup>82,83</sup> Instead, external validation has been considered to be the only way to establish a reliable QSAR model. Accordingly, statistical parameters for the predictions of binding affinities by AFMoC, CoMFA, and CoMSIA models for 15 test-set compounds not considered during training are given in Tables 3 and 4. Convincingly, in the AFMoC case, the squared correlation coefficient ( $r^2$ ) amounts to 0.65, and the deviations between predicted and experimental  $pK_i$  values are below 1 order of magnitude in all but four cases (Figure 4). This indicates that the information contained in the AFMoC fields is sufficiently general to allow for good binding affinity predictions, even for compounds that have not been included in the training set. In the case of CoMFA and CoMSIA, larger  $r^2$  values of 0.78 (4 ligands deviate by more than one logarithm unit in  $pK_i$ ) and 0.70 (5 ligands deviate by more than one logarithm unit in  $pK_i$ ), respectively, are obtained, which points to robust

ligand-based 3D QSAR models that have been derived grounded on a protein-based ligand alignment.

In total, the aforementioned results provide a strong hint to the significance and robustness of the AFMoC model derived for ligands aligned in the experimental thrombin structure. In the following, this model will serve as a “benchmark” for comparisons with AFMoC models derived with homology-modeled protein structures.

**3. Thrombin Structures Obtained by Homology Modeling.** To validate our consensus AFMoC approach, three HMs of thrombin were generated with MODELLER.<sup>68</sup> The rmsd values of these models, with respect to each other, and the experimental thrombin structure are given in Table 1. Figure 5 displays the binding pocket regions of the protein structure of 1ets and the modeled structures. Because the AFMoC analysis focuses on those regions of a protein where ligands interact, we will primarily consider structural differences in the thrombin binding pocket in the following discussion.

As indicated by the rmsd values of all  $C_\alpha$  atoms, the homology-modeled structures roughly agree with each other outside the binding pocket region ( $\text{rmsd} \leq 0.7 \text{ \AA}$ ) and deviate only moderately from the experimental thrombin structure ( $\text{rmsd} \approx 2.0 \text{ \AA}$ ). In turn, if only binding site residues are considered, larger structural deviations are found within the models ( $\text{rmsd}$  of  $C_\alpha$  atoms =  $0.8\text{--}1.0 \text{ \AA}$ ;  $\text{rmsd}$  of heavy atoms =  $1.9\text{--}2.4 \text{ \AA}$ ), whereas the deviations with respect to the experimental structure become smaller ( $\text{rmsd}$  of  $C_\alpha$  atoms:  $1.1\text{--}1.5 \text{ \AA}$ ). The latter is consistent with the finding that binding-site regions are usually more structurally conserved than other protein regions. Analyzing the binding



**Figure 4.** Experimentally determined binding affinities versus fitted predictions obtained by the AFMoC model for the training set of 64 thrombin inhibitors (panel a) or versus predictions for the external test set of 15 thrombin inhibitors (panel b). The inhibitors were aligned in the experimental protein structure. In addition to the line of ideal correlation (straight line), deviations by  $\pm 1$  logarithmic unit are also indicated (dotted lines).

pocket region in more detail (Figure 5) reveals that the structural deviations mostly originate from different conformations of the 60-loop (residues 60, 60A, ..., 60I, 61, 62, 63), while the remaining parts of the binding pocket are rather similar. In this regard, model 2 shows a slightly more narrowed S2 pocket than the experimental thrombin structure, whereas the S2 pocket is almost or completely uncovered in models 1 and 3, respectively.

The thrombin system used here is very well-suited for validating our approach, because of the considerable amount of structural and energetic information available. Yet, we note that no information about the known crystal structure of thrombin was included into the homology-modeling process to comply with conditions comparable to a real-life modeling situation. In addition, we are convinced that the results obtained for the system can be carried over to other targets in structure-based drug design for the following reasons:

(i) The model binding pockets show structurally conserved regions in addition to a less well-defined 60-loop region. This is reminiscent of the dual character of binding sites in terms of the presence of regions with high and low structural stability found in a recent study on 16 proteins.<sup>84</sup>

(ii) The structural deviations within the binding pockets are mostly restricted to a limited set of residues in the 60-loop region. This is consistent with recent findings that, generally, only a small number of residues in a pocket undergo conformational changes.<sup>85</sup>

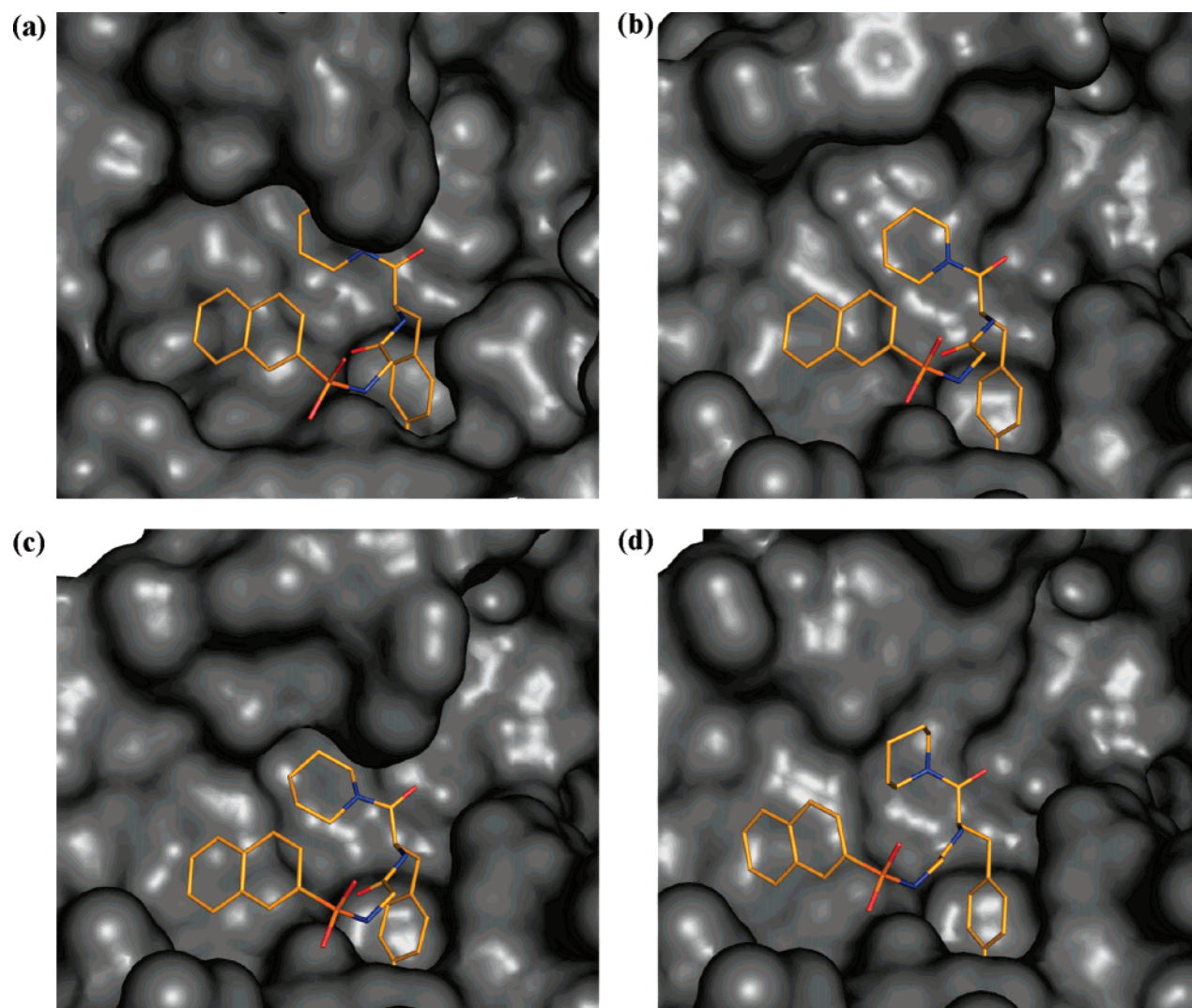
(iii) Yet, the structural deviations of these side chains are prominent and larger than what has been found when analyzing the protein flexibility of typical drug(gable) targets, such as trypsin, ricin, and aldose reductase.<sup>86</sup> In that respect, we think that the HMs are representative for considering the effect of conformational variability on deriving protein-based 3D QSAR models.

**4. AFMoC Analyses Using the Single Homology-Modeled Thrombin Structures.** The HMs were superimposed, with respect to the experimental thrombin structure, based on the  $C_\alpha$  atoms of residues Asp189 and those of the catalytic triade. The ligand alignment obtained for the experimental thrombin structure (see above) was then adapted according to the structural requirements of the modeled binding pockets, with a final energy minimization in the rigid pocket of each model as described previously. That way, a particular ligand superimposition for each binding pocket was generated. Table 5 displays the statistical results of the AFMoC analyses for these HMs, using the same interaction field combination as that used in the case of the experimental thrombin structure (see Table S2 in the Supporting Information).

For HM 1, an AFMoC model of increased predictive power, compared to the case of the experimental thrombin structure, could be derived, as indicated by a  $q^2$  value of 0.61 and a  $r^2$  value for the test-set prediction of 0.72. The robustness of the model is further corroborated by the results of a leave-10-out cross validation (see Table S3 in the Supporting Information). This finding is encouraging, because the structure of HM 1 considerably deviates from the experimental protein structure (the rmsd value of all non-hydrogen atoms in the binding pocket is 2.90 Å) (see Figure 5). However, for HM 2 and HM 3, only weak AFMoC models could be obtained ( $q^2 \approx 0.4$ ) that also show a decreased predictive power, with respect to the external test set ( $r^2 \approx 0.4$ ). Hence, AFMoC models derived from single HM may vary considerably in quality.

In that respect, we note that, although the trend observed in  $q^2$  values parallels that in the  $r^2$  values of the test-set predictions in this case, such a relationship cannot be expected (in general).<sup>82,83</sup> Thus, it may not be straightforward to choose one HM out of the generated ensemble that results in the AFMoC model with the best (external) predictivity. In addition, the “better” HM 2, in terms of deviations from the experimental structure, results in an AFMoC model of the same quality as the “worse” HM 3. This leads to the second implication that no direct correlation between the amount of structural deviation of the modeled proteins and the quality of the AFMoC models can be observed. Instead, the outcome of the AFMoC models seems to be dependent on local structural differences, rather than overall structural deviations in the binding pocket. By including multiple protein structures into a consensus model, we thus expect that the influence of a single receptor structure on the AFMoC results is diminished.

**5. Consensus AFMoC Models.** In the following, the results of AFMoC<sup>con</sup> models will be discussed that are



**Figure 5.** Binding pockets of the (a) crystallographically determined structure (PDB code 1ets), as well as homology models (b) HM 1, (c) HM 2, and (d) HM 3. In addition, ligand **24**, taken from the particular alignment for each pocket, is shown.

**Table 5.** Statistical Results for AFMoC Analyses Using Homology-Modeled Thrombin Structures<sup>a</sup>

parameter	Value		
	HM 1 <sup>b</sup>	HM 2 <sup>b</sup>	HM 3 <sup>b</sup>
$q^{2\ b,c}$	0.61 (0.58)	0.42 (0.41)	0.39 (0.38)
$s_{\text{press}}^{d,e}$	0.79	0.94	0.95
correlation coefficient, $r^{2\ b}$	0.85 (0.84)	0.79 (0.78)	0.51 (0.50)
$S^{d,f}$	0.50	0.58	0.85
Fischer's $F$ -value <sup>b</sup>	112.1 (103.5)	72.8 (70.9)	64.8 (62.4)
number of components	3	3	1
fraction			
C.3	0.365	0.390	0.259
C.ar	0.581	0.519	0.698
O.3	0.014	0.017	0.004
O.2	0.016	0.030	0.019
O.co2	0.004	0.006	0.002
N.am	0.021	0.039	0.018

<sup>a</sup> Models are based on PLS calculations using a  $\sigma$  value of 0.7 Å and a grid spacing of 1.0 Å. <sup>b</sup> Values are given considering only  $pK_i^{\text{PLS}}$  or considering  $pK_i^{\text{total}}$  (values in parentheses). <sup>c</sup>  $q^2 = 1 - \text{PRESS}/\text{SSD}$  as obtained by “leave-one-out” (LOO) cross validation. PRESS equals the sum of squared differences between predicted and experimentally determined binding affinities, SSD is the sum of the squared differences between experimentally determined binding affinities and the mean of the training set binding affinities. <sup>d</sup> In logarithmic units. <sup>e</sup>  $s_{\text{PRESS}} = \sqrt{\text{PRESS}/(n-h-1)}$  as obtained by LOO cross validation.  $n$  equals the number of data points,  $h$  is the number of components. <sup>f</sup>  $S = \sqrt{\text{RSS}/(n-h-1)}$ . RSS equals the sum of squared differences between fitted and experimentally determined binding affinities.

derived from interaction fields generated from multiple ligand superimpositions inside conformationally different binding

pockets. The generation of these AFMoC<sup>con</sup> models is motivated by three observations:



**Table 6.** Statistical Results for AFMoC<sup>con</sup> Analyses Using Homology-Modeled Thrombin Structures<sup>a</sup>

	Value								
	HM 1 + 2		HM 1 + 3		HM 2 + 3		HM 1+ 2+3		
w/o varselect <sup>b</sup>									
$q^2$ <sup>c,d</sup>	0.59 (0.57)		0.45 (0.43)		0.42 (0.41)		0.57 (0.56)		
$s_{\text{press}}$ <sup>ef</sup>	0.81		0.91		0.93		0.81		
correlation coefficient, $r^2$ <sup>c</sup>	0.78 (0.77)		0.55 (0.53)		0.52 (0.51)		0.75 (0.74)		
$S^{e,g}$	0.58		0.83		0.85		0.63		
Fischer's $F$ -value <sup>c</sup>	70.7 (67.2)		74.6 (69.3)		66.3 (64.2)		59.1 (56.7)		
number of components	3		1		1		3		
w/varselect <sup>h</sup>									
$q^2$ <sup>c,d</sup>	0.61 (0.59)		0.56 (0.54)		0.43 (0.42)		0.60 (0.58)		
$s_{\text{press}}$ <sup>ef</sup>	0.78		0.83		0.92		0.79		
correlation coefficient, $r^2$ <sup>c</sup>	0.77 (0.76)		0.78 (0.77)		0.52 (0.51)		0.74 (0.73)		
$S^{e,g}$	0.60		0.59		0.85		0.64		
Fischer's $F$ -value <sup>c</sup>	67.1 (63.6)		69.9 (66.5)		66.0 (63.9)		56.3 (54.0)		
number of components	3		3		1		3		
model	1	2	1	3	2	3	1	2	3
prevalence <sup>i</sup>	0.50	0.50	0.50	0.50	0.51	0.49	0.38	0.33	0.29

<sup>a</sup> Models are based on PLS calculations using a  $\sigma$  value of 0.7 Å and a grid spacing of 1.0 Å. <sup>b</sup> No variable selection was performed. <sup>c</sup> Values are given considering only  $\text{p}K_i^{\text{PLS}}$  or considering  $\text{p}K_i^{\text{total}}$  (values given in parentheses). <sup>d</sup>  $q^2 = 1 - \text{PRESS}/\text{SSD}$  as obtained by “leave-one-out” (LOO) cross validation. PRESS equals the sum of squared differences between predicted and experimentally determined binding affinities, SSD is the sum of the squared differences between experimentally determined binding affinities and the mean of the training set binding affinities. <sup>e</sup> In logarithmic units. <sup>f</sup>  $s_{\text{PRESS}} = \sqrt{\text{PRESS}/(n-h-1)}$  as obtained by LOO cross validation.  $n$  equals the number of data points,  $h$  is the number of components. <sup>g</sup>  $S = \sqrt{\text{RSS}/(n-h-1)}$ . RSS equals the sum of squared differences between fitted and experimentally determined binding affinities. <sup>h</sup> A VINFM-based variable selection was performed. <sup>i</sup> Average over prevalences calculated by eq 4.

(i) During the process of homology modeling, multiple protein structures are often generated to compensate for the impact of structural inaccuracies of each single model.

(ii) Furthermore, the use of ensembles of (modeled or experimentally determined) protein structures allows one to account for protein flexibility in structure-based drug design.<sup>44,45</sup>

(iii) Similarly, multiple ligand conformations incorporate conformational and alignment freedom into the 3D QSAR model derivation process.<sup>40</sup>

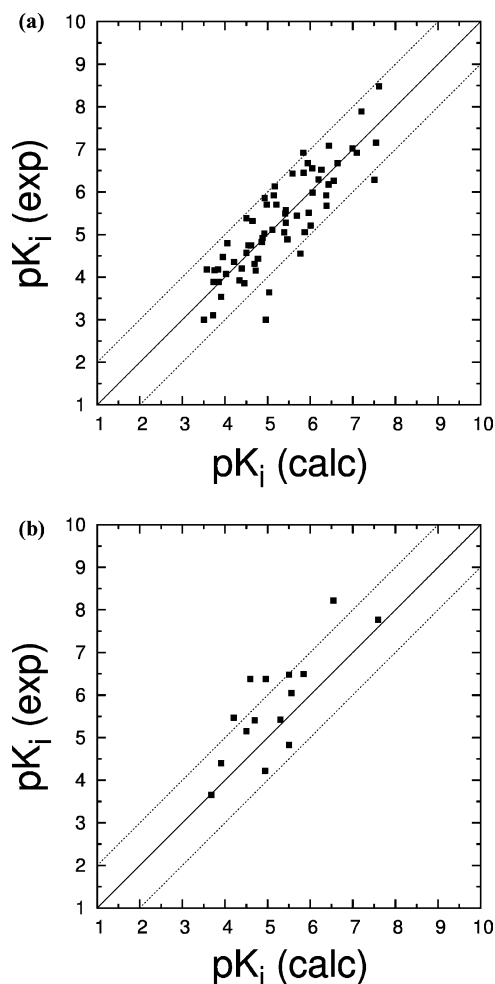
Table 6 displays statistical results for all three pairwise AFMoC<sup>con</sup> models and the one that comprises all three single models. The models are derived by correlating interaction field values iteratively with experimental binding affinities as described previously (eqs 1–3). As exemplarily shown for the HM 1 + 2 + 3 model in Figure S1 in the Supporting Information, the runs converged well within 15 iterations. The results of leave-10-out cross validations are given in Table S3 in the Supporting Information; in all cases, the  $q_{L100}^2$  values parallel the  $q^2$  values obtained by LOO cross-validation. The predicted versus actual binding affinities for the PLS analysis of the consensus model HM 1 + 2 + 3 are depicted in Figure 6.

First, the influence of the VINFM-based region selection (Figure 2) will be investigated. In the case of HM 1 + 3 and HM 1 + 2 + 3,  $q^2$  values are obtained after applying the variable selection, which are 5%–24% larger than the values that have been derived from the unfiltered extended  $X$  matrix (for HM 1 + 3: 0.56 vs 0.45; for HM 1 + 2 + 3: 0.60 vs 0.57). In contrast, no significant improvement is observed in the case of HM 1 + 2 and HM 2 + 3. The same trends are also observed if the  $r^2$  values of the test-set predictions are considered (see Table 3; data for the unfiltered AFMoC<sup>con</sup> models are not shown), as are for the leave-10-out cross validation (see Table S3 in the Supporting Information). Thus, the VINFM-based region selection yields

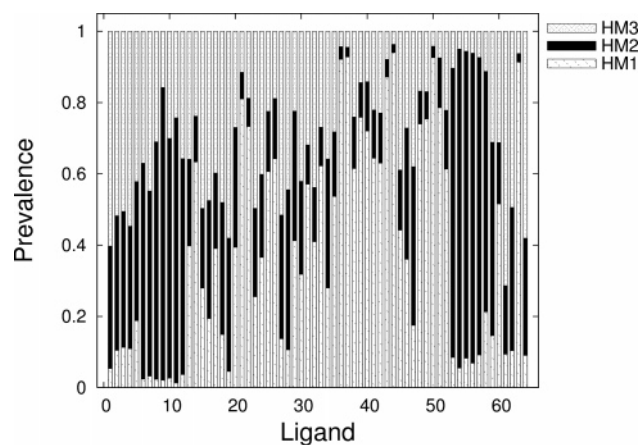
AFMoC<sup>con</sup> models with significantly improved predictive power, particularly in those cases where “good” models are combined with those that show less predictive power. In fact, AFMoC<sup>con</sup> models derived without variable selection are, in many cases, worse than the best single model used in the combinations (see Tables 5 and 6). Hence, the variable selection process apparently distinguishes successfully explanatory information that is contained in the extended  $X$  matrix from noise.

At this point, it must be emphasized that it is still a matter of debate whether an improved predictive power, with respect to new molecules not contained in the training set, is achieved if variable selection techniques are applied.<sup>87,88</sup> Indeed, in many of the techniques, the variable selection process is guided by computing cross-validation performance estimates of the different variable subsets,<sup>60,88</sup> although these internal figures of merit do not necessarily correlate well with the external predictivity of the selected models.<sup>82,87,88</sup> However, we do note that no such “cross-validation guided” selection process is pursued in our case. Instead, by taking into account the block structure of the extended  $X$  matrix, we only select that column out of the set of equivalent columns  $C_j$  (eq 5) that shows the largest contribution to the full model. That way, we expect that the variables selected are of high statistical importance and that the danger of overfitting during variable selection is reduced.

When comparing the AFMoC<sup>con</sup> models after variable selection (Table 6) to the single models (Table 5), it becomes obvious that the  $q^2$  values of the former are comparable to the best single model used in the combination. Intuitively, one would expect that the best model used in the combination also has the greatest contribution to the consensus model. Interestingly, however, as indicated by the average prevalences computed for each model in the extended  $X$  matrix, the single models contribute roughly in equal shares (although the prevalences of single ligands can widely vary



**Figure 6.** Experimentally determined binding affinities versus fitted predictions obtained by the AFMoC model for the training set of 64 thrombin inhibitors (panel a) or versus predictions for the external test set of 15 thrombin inhibitors (panel b). Results are shown for the consensus model generated from all three homology-modeled protein structures (HM 1 + 2 + 3). In addition to the line of ideal correlation (straight line), deviations by  $\pm 1$  logarithmic unit are also indicated (dotted lines).



**Figure 7.** Prevalences (eq 4) of individual binding modes of the training set ligands obtained for the AFMoC<sup>con</sup> model HM 1 + 2 + 3.

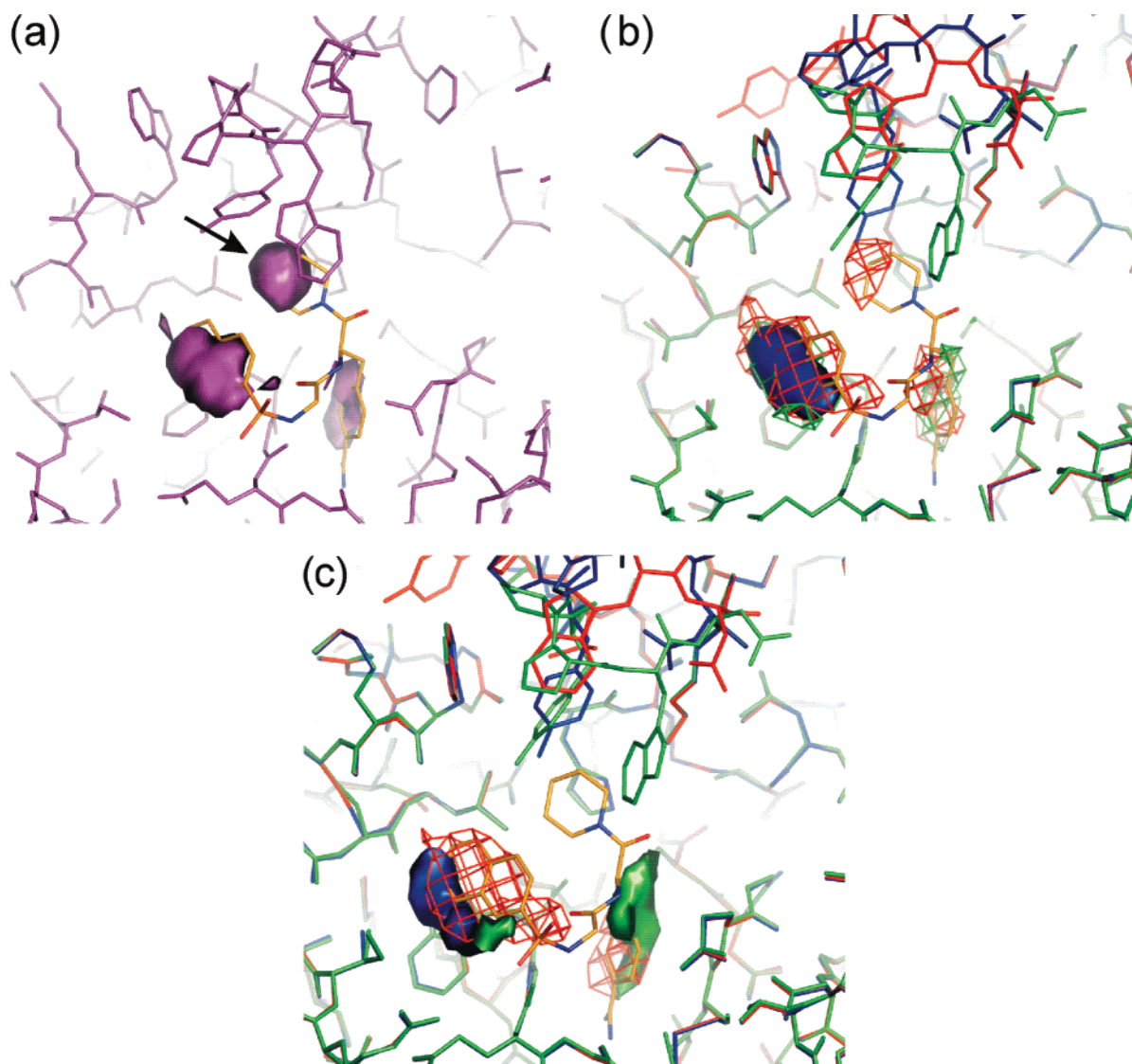
among the models (see Figure 7). Hence, even in models of overall lower quality, regions exist that are of higher statistical importance than the equivalent regions in the better models. Combining such statistically important regions as

in our AFMoC<sup>con</sup> approach then means to pick the best information available across single models to be included into the consensus model. In that sense, our approach differs from the “consensus scoring” approaches that have been recently introduced,<sup>89,90</sup> which combine information from different sources by averaging over them, regardless of the information quality. In fact, consensus models HM 1 + 2 and HM 1 + 2 + 3 show  $q^2$  values of 0.60 and 0.61, respectively, that are significantly higher than that obtained for the AFMoC model using the experimental thrombin structure (0.57; see Table 2).

With respect to the external predictivity (Table 3), the AFMoC<sup>con</sup> models HM 1 + 2 and HM 1 + 2 + 3 show  $r^2$  values of 0.64, which is similar to the value obtained for the experimental thrombin structure ( $r^2 = 0.65$ ). The deviations between predicted and experimental  $pK_i$  values are less than 1 order of magnitude in all but four cases (see Figure 6). Thus, including additional information of HM 2 and HM 3 does not deteriorate the robustness of the consensus model, albeit these single models by themselves only show a weak predictivity. For those cases where multiple RSLA are available, this result is encouraging, because it alleviates the need to address the question of which of the RSLA to choose, which inevitably arises if the standard AFMoC approach is applied.

**Graphical Interpretation of the AFMoC<sup>con</sup> Results.** As an important means for rational drug design, the graphical interpretation of AFMoC<sup>con</sup> models can aid in proposing structural modifications of already known ligands to improve their binding. In this regard, the AFMoC<sup>con</sup> approach reveals interesting features. In Figures 8 and 9, the STDEV\*COEFF fields (which elucidate regions in the binding pocket of thrombin, where the presence of ligand atoms of types “aromatic carbon” and “amide nitrogen” favors binding) are displayed. In both cases, contour maps are given for the AFMoC model obtained with the ligands aligned in the experimental thrombin structure (panel a)), for the AFMoC models obtained with the single HM (panel b)), and for the AFMoC<sup>con</sup> model HM 1 + 2 + 3 (panel c)). In the latter case, the depicted contour maps were obtained by considering separately those regions in the X matrix that have originated from a particular protein in the supermatrix. For reasons of comparison, the same contour level has been applied to display the contour maps for all panels of one figure.

As for the standard AFMoC approach,<sup>33</sup> first, contour maps of the AFMoC<sup>con</sup> model occur at the location of ligand atoms, indicating regions where it is more favorable or unfavorable to place ligand atoms of a given type, with respect to binding affinity. This contrasts to generic property fields (e.g., steric or electrostatic) as obtained by CoMFA or CoMSIA. Second, because of a cutoff distance of 6 Å for the DrugScore potentials used to map protein–ligand interactions onto neighboring grid points, solvent-exposed ligand parts will not be considered in the model derivation. This is in contrast to the ligand-based 3D QSAR techniques CoMFA and CoMSIA, which examine interactions to *all* parts of the ligand. Finally, the contour maps appear contiguous and smoothly connected, which facilitates their interpretation. This is by no means self-evident for the AFMoC<sup>con</sup> case, because, upon variable selection, *no* restraints with regard to preferentially including contiguous grid regions from a single model into the consensus X matrix have been imposed



**Figure 8.** AFMoC STDEV\*COEFF contour maps showing regions in the thrombin pocket where the presence of aromatic carbon in a ligand will enhance binding. (a) The model was derived using the experimentally determined protein structure. (b) The models were derived using either one of the homology-modeled structures (red, HM1; green, HM2; blue, HM3). The colors of the contour maps coincide with those of the protein structures. (c) The model HM 1 + 2 + 3 was derived by applying the consensus approach introduced in this study. The depicted contour maps were obtained by considering separately only those regions in the consensus *X* matrix that have originated from a particular protein in the supermatrix. The colors of the contour maps coincide with those of the protein structures. In all cases, the contour level is  $-0.010$ .

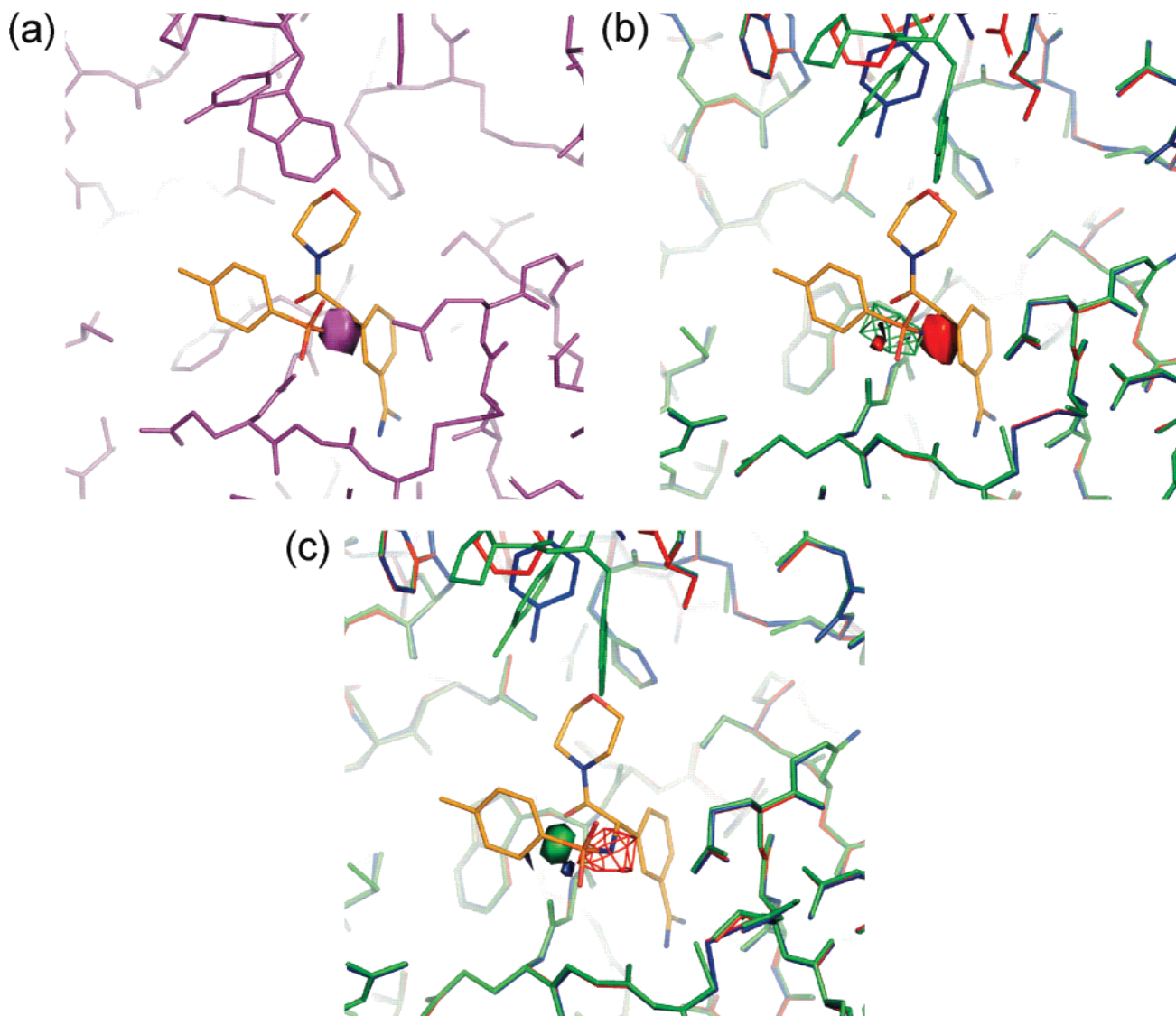
(as opposed to, e.g., in the GOLPE-guided region selection, where the space around the molecules is partitioned after highly informative *X*-descriptors have been extracted<sup>61</sup>). The fact that columns selected from single models originate from connected grid regions nevertheless indicate rather smooth changes of interaction field values in these areas, which is a result of sufficiently smearing particular protein–ligand interactions across neighboring grid points.

In the following, the contour maps as obtained by standard AFMoC, using homology-modeled protein structures (Figures 8b and 9b), and by AFMoC<sup>con</sup> (Figures 8c and 9c) will be compared to those computed from the experimental protein structure (Figures 8a and 9a). In the case of aromatic carbon (Figure 8), not surprisingly, all three HM show isopleths that coincide with those of the model derived from the experimental structure in the parts of the binding pocket that are *structurally conserved* (the S1 and S3/4 pockets). Similarly, these favorable regions are also identified by the AFMoC<sup>con</sup> model. Thereby, non-overlapping isopleths of

comparable sizes are obtained from all three HM in the regions where the binding of aromatic moieties is most favorable: Isopleths from HM 1 appear both in the S1 and S3/4 pockets, whereas those of HM 2 (HM 3) appear mostly in the S1 (S3/4) pocket. This is consistent with the finding that, for HM 1, the largest average over prevalences is found in the consensus model (see Table 6), whereas HM 2 and HM 3 are less prevailing. Nevertheless, the result again also visualizes the fact that regions of higher statistical importance may be even found in models of overall lower quality, compared to equivalent regions in overall better models.

When considering *structurally varying* binding pocket regions, another advantage of AFMoC<sup>con</sup> over standard AFMoC of multiple protein structures becomes obvious. As such, the isopleths depicted at the piperidinyl moiety of ligand **24** in the S2 pocket in Figure 8a) (see arrow) is only found by the AFMoC model derived from HM 1. In the other two cases, modeled amino acid side chains of the 60-loop are either too far apart from or too close to this region to





**Figure 9.** AFMoC “STDDEV\*COEFF” contour maps showing regions in the thrombin pocket where the presence of amide nitrogen in a ligand will enhance binding. (a) The model was derived using the experimentally determined protein structure. (b) The models were derived using either one of the homology-modeled structures (red, HM1; green, HM2; blue, HM3). The colors of the contour maps coincide with those of the protein structures. (c) The model was derived by applying the consensus approach introduced in this study to HM1, HM2, and HM3. The depicted contour maps were obtained by considering separately only those regions in the consensus X matrix that have originated from a particular protein in the supermatrix. The colors of the contour maps coincide with those of the protein structures. In all cases, the contour level is  $-0.037$ .

exhibit favorable interactions to the ligands. Using solely a “wrong” HM in a standard AFMoC approach, one would have missed this information. In the AFMoC<sup>con</sup> model, the S2 pocket is identified again as being favorable for aromatic interactions, albeit less pronounced than in the model of the experimental thrombin structure (i.e., at a lower contour level than that shown in Figure 8c). In a similar way, favorable interactions to amide nitrogens (Figure 9) that are identified by only one AFMoC model in the case of homology-modeled protein structures, when compared to the AFMoC model of the experimental protein structure, are displayed jointly in the case of the AFMoC<sup>con</sup> model. In our opinion, these findings clearly favor the use of AFMoC<sup>con</sup> also for the graphical interpretation of the models when multiple RSLA are available.

Taken together, one can anticipate using AFMoC<sup>con</sup> as an adjustable tailor-made scoring function for the prediction of binding affinities of novel ligands in the following way. Based on lead compounds found with a generally valid

scoring function such as DrugScore and subsequently collected experimental energetic and structural data, an initial AFMoC<sup>con</sup> model is generated. Here, conformational variability of the protein structure and/or structural inaccuracies of modeled structures can already be considered. By interpreting the STDDEV\*COEFF contour maps, binding-enhancing modifications of the current ligands are suggested, based on chemically intuitive atom-type information. Binding affinities of novel ligand candidates are then predicted as described previously, and successful candidates are subsequently tested experimentally. This information is again fed back into the AFMoC<sup>con</sup> model, resulting in an iterative process that steadily adjusts the model according to increasing knowledge about the target under investigation.

## CONCLUSION

We have developed a new approach (consensus adaptation of fields for molecular comparison, AFMoC<sup>con</sup>) that allows one to consider ligand and receptor conformational variability

in a tailor-made scoring function. The development was motivated by the need to account for the receptor flexibility observed in experimental complex structures or structural inaccuracies of modeled receptors in structure-based drug design. Furthermore, the method partially alleviates the “alignment problem” frequently encountered in three-dimensional quantitative structure–activity relationship (3D QSAR) applications. AFMoC<sup>con</sup> makes use of ensembles of (modeled or experimentally determined) receptor structure–ligand alignments (RSLAs). AFMoC<sup>con</sup> models are then derived by adapting knowledge-based interaction fields using structural and energetic information about known receptor–ligand complexes. For this, partial least-squares (PLS) regression analysis considering multimode binding (MMB), together with a variable-influence-on-the-model (VINFM)-based region selection is applied. The AFMoC<sup>con</sup> results can be interpreted in terms of contour plots that aid in proposing variations of the ligand structure to improve ligand binding affinity. Similarly, the adapted interaction fields can be used together with the original ones to score novel ligands in the realm of the consensus model.

We validated our approach on a data set of thrombin inhibitors. Initially, using ligands aligned in an experimental thrombin structure as a benchmark system, a significant adaptation of fields for molecular comparison (AFMoC) model was obtained, with respect to both internal and external predictivity. Convincingly, when ligands are aligned in homology-modeled thrombin structures, a robust AFMoC model also could be obtained in one case. Yet, no relationship between the overall quality of the homology model and the AFMoC model quality could be observed, rendering the a priori choice of a single receptor structure a difficult task. Applying the newly developed AFMoC<sup>con</sup> approach that considers all available RSLAs instead circumvents this problem. Note that, in this manner, AFMoC<sup>con</sup> models were obtained that (i) show a comparable internal and external predictivity, similar to the best single model (derived either from the experimental structure or an homology model); (ii) show contributions to the consensus model by the single models in roughly equal shares; and (iii) allow one to graphically identify important regions in the binding pocket that would have been missed if a “wrong” single model had been used.

As further advantages, no principle limitations, with respect to the number of different RSLAs considered by AFMoC<sup>con</sup> exist, and the approach should be applicable to also tailor objective functions for docking, similar in spirit to the AFMoC<sup>obj</sup> approach,<sup>36</sup> allowing one to account implicitly for protein flexibility during the docking process.

Overall, we are convinced that this method provides an important step toward incorporating ligand and receptor conformational variability into tailor-made scoring and objective functions for structure-based drug design. Particularly, in cases such as G-protein-coupled receptors, AFMoC<sup>con</sup> is expected to be of great value. Here, a plethora of ligand information exists that can be used for training, but experimentally determined protein structures with high sequence similarity to pharmacologically important targets are missing, thus, resulting in rather large uncertainties in homology-modeled structures.

## ACKNOWLEDGMENT

This work was supported by funds from J.W. Goethe-University (Frankfurt, Germany). We are grateful to Domingo González Ruiz and Sebastian Radestock (J.W. Goethe-University, Frankfurt) for critically reading the manuscript.

**Supporting Information Available:** Four tables with (i) thrombin inhibitors used as training and test set compounds; (ii)  $q^2$  values obtained by AFMoC using different combinations of interaction field; (iii) statistical results of leave-10-out cross-validation runs; and (iv)  $q^2$  values for 10 AFMoC models obtained with randomly scrambled affinity data. A figure showing the convergence of a PLS MMB run for the model HM 1 + 2 + 3 also is provided. (PDF.) This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (1) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047–55.
- (2) Drews, J. Drug Discovery: A Historical Perspective. *Science* **2000**, *287*, 1960–1964.
- (3) Gohlke, H.; Klebe, G. Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.
- (4) Muegge, I.; Rarey, M. Small Molecule Docking and Scoring. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2001; Vol. 17, pp 1–60.
- (5) Sotriffer, C.; Stahl, M.; Boehm, H. J.; Klebe, G. Docking and Scoring Functions/Virtual Screening. In *Burger's Medicinal Chemistry and Drug Discovery*; Wiley: New York, 2003; Vol. 1, pp 281–333.
- (6) Pearlman, D. A.; Charifson, P. S. Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System. *J. Med. Chem.* **2001**, *44*, 3417–3423.
- (7) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. Assessing scoring functions for protein–ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3047.
- (8) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluations of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759–4767.
- (9) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem.* **2005**, *48*, 4040–4048.
- (10) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; de-Solms, S. J.; Gaffin, N.; Ghosh, A. K.; Giuliani, E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Darke, P. L.; Zugay, J. A. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* **1995**, *38*, 305–317.
- (11) Murray, C. W.; Auton, T. R.; Eldridge, M. D. Empirical scoring functions. II. The testing of an empirical scoring function for the prediction of ligand-receptor binding affinities and the use of Bayesian regression to improve the quality of the model. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 503–519.
- (12) Matter, H.; Will, D. W.; Nazare, M.; Schreuder, H.; Laux, V.; Wehner, V. Structural requirements for factor Xa inhibition by 3-oxybenzamides with neutral P1 substituents: combining X-ray crystallography, 3D-QSAR, and tailored scoring functions. *J. Med. Chem.* **2005**, *48*, 3290–312.
- (13) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.
- (14) Wang, T.; Wade, R. C. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem.* **2001**, *44*, 961–971.
- (15) Kurinov, I. V.; Harrison, R. W. Prediction of new serine proteinase inhibitors. *Nat. Struct. Biol.* **1994**, *1*, 735–743.
- (16) Kulkarni, S. S.; Kulkarni, V. M. Structure Based Prediction of Binding Affinity of Human Immunodeficiency Virus-1 Protease Inhibitors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1128–1140.
- (17) Rognan, D.; Lauemoller, S. L.; Holm, A.; Buus, S.; Tschinke, V. Predicting Binding Affinities of Protein Ligands from Three-Dimensional Models: Application to Peptide Binding to Class I Major Histocompatibility Proteins. *J. Med. Chem.* **1999**, *42*, 4650–4658.
- (18) Grootenhuys, P. D. J.; van Galen, P. J. M. Correlation of binding affinities with nonbonded interaction energies of thrombin-inhibitor complexes. *Acta Crystallogr., Sect. D* **1995**, *51*, 560–566.

- (19) Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins* **1997**, *29*, 32–58.
- (20) Takamatsu, Y.; Itai, A. A New Method for Predicting Binding Free Energy Between Receptor and Ligand. *Proteins* **1998**, *33*, 62–73.
- (21) Venkatarangan, P.; Hopfinger, A. J. Prediction of Ligand-Receptor Binding Thermodynamics by Free Energy Force Field Three-Dimensional Quantitative Structure–Activity Relationship Analysis: Application to a Set of Glucose Analogue Inhibitors of Glycogen Phosphorylase. *J. Med. Chem.* **1999**, *42*, 2169–2179.
- (22) Viswanadhan, V. N.; Reddy, M. R.; Wlodawer, A.; Varney, M. D.; Weinstein, J. N. An Approach to Rapid Estimation of Relative Binding Affinities of Enzyme Inhibitors: Application to Peptidomimetic Inhibitors of the Human Immunodeficiency Virus Type 1 Protease. *J. Med. Chem.* **1996**, *39*, 705–712.
- (23) Bohacek, R. S.; McMartin, C. Multiple Highly Diverse Structures Complementary to Enzyme Binding Sites: Results of Extensive Application of a *de Novo* Design Method Incorporating Combinatorial Growth. *J. Am. Chem. Soc.* **1994**, *116*, 5560–5571.
- (24) Kasper, P.; Christen, P.; Gehring, H. Empirical Calculation of the Relative Free Energies of Peptide Binding to the Molecular Chaperone DnaK. *Proteins* **2000**, *40*, 185–192.
- (25) Pierce, A. C.; Jorgensen, W. L. Estimation of binding affinities for selective thrombin inhibitors via Monte Carlo simulations. *J. Med. Chem.* **2001**, *44*, 1043–1050.
- (26) Rizzo, R. C.; Tirado-Rives, J.; Jorgensen, W. L. Estimation of Binding Affinities for HEPT and Nevirapine Analogues with HIV-1 Reverse Transcriptase via Monte Carlo Simulations. *J. Med. Chem.* **2001**, *44*, 145–154.
- (27) Cramer, R. D.; Wendt, B. Pushing the boundaries of 3D-QSAR. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 23–32.
- (28) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). I. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959.
- (29) 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.
- (30) Sippl, W. Development of biologically active compounds by combining 3D QSAR and structure-based design methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 825–830.
- (31) Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589–601.
- (32) Wade, R. C.; Henrich, S.; Wang, T. Using 3D protein structures to derive 3D-QSARs. *Drug Discovery Today* **2004**, *1*, 241–246.
- (33) Gohlke, H.; Klebe, G. DrugScore meets CoMFA: Adaptation of Fields for Molecular Comparison (AFMoC) or How to Tailor Knowledge-based Pair-Potentials to a Particular Protein. *J. Med. Chem.* **2002**, *45*, 4153–4170.
- (34) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based Scoring Function to Predict Protein-Ligand Interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- (35) Silber, K.; Heidler, P.; Kurz, T.; Klebe, G. AFMoC enhances predictivity of 3D QSAR: a case study with DOXP-reductoisomerase. *J. Med. Chem.* **2005**, *48*, 3547–3563.
- (36) Radestock, S.; Bohm, M.; Gohlke, H. Improving binding mode predictions by docking into protein-specifically adapted potential fields. *J. Med. Chem.* **2005**, *48*, 5466–5479.
- (37) Carlson, H. A. Protein flexibility and drug design: How to hit a moving target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447–452.
- (38) Teague, S. J. Implications of Protein Flexibility for Drug Discovery. *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.
- (39) Wong, C. F.; McCammon, J. A. Protein Flexibility and Computer-Aided Drug Design. *Annu. Rev. Pharmacol. Toxicol.* **2003**, *43*, 31–45.
- (40) Vedani, A.; Dobler, M. Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quant. Struct. Act. Relat.* **2002**, *21*, 382–390.
- (41) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D QSAR models using the 4D QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
- (42) Pan, D.; Tseng, Y.; Hopfinger, A. J. Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1591–1607.
- (43) Lukacova, V.; Balaz, S. Multimode ligand binding in receptor site modeling: implementation in CoMFA. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2093–2105.
- (44) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
- (45) Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem.* **2000**, *43*, 2100–2114.
- (46) Schafferhans, A.; Klebe, G. Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **2001**, *307*, 407–427.
- (47) Bindewald, E.; Skolnick, J. A scoring function for docking ligands to low-resolution protein structures. *J. Comput. Chem.* **2005**, *26*, 374–383.
- (48) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076–5084.
- (49) Wang, J.; Szewczuk, Z.; Yue, S. Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. Calculation of relative binding free energies and configurational entropies: a structural and thermodynamic analysis of the nature of non-polar binding of thrombin inhibitors based on hirudin55–65. *J. Mol. Biol.* **1995**, *253*, 473–492.
- (50) Balaz, S.; Hornak, V.; Haluska, L. Receptor mapping with multiple binding modes: Binding site of PBC-degrading dioxigenase. *Chemom. Intell. Lab. Syst.* **1994**, *24*, 185–191.
- (51) Hornak, V.; Balaz, S.; Schaper, K. J.; Seydel, J. K. Multiple binding modes in 3D-QSAR: Microbial degradation of polychlorinated biphenyls. *Quant. Struct. Act. Relat.* **1998**, *17*, 427–436.
- (52) Davis, A. M.; Gensmantel, N. P.; Johansson, E.; Marriott, D. P. The use of the GRID program in the 3-D QSAR analysis of a series of calcium-channel agonists. *J. Med. Chem.* **1994**, *37*, 963–972.
- (53) Gohlke, H.; Hendlich, M.; Klebe, G. Predicting Binding Modes, Binding Affinities and “Hot Spots” for Protein-Ligand Complexes using a Knowledge-based Scoring Function. *Persp. Drug Discovery Des.* **2000**, *20*, 115–144.
- (54) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–2303.
- (55) Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein–ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114–2125.
- (56) Cramer, R. D.; Wold, S. Comparative Molecular Field Analysis (CoMFA), 1994.
- (57) Zhang, Y.; Lukacova, V.; Bartus, V.; Balaz, S. Structural determinants of binding of aromates to extracellular matrix: a multi-species multimode CoMFA study. *Chem. Res. Toxicol.* **2007**, *20*, 11–19.
- (58) Bush, B. L.; Nachbar, R. B. Sample-distance Partial Least Squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.
- (59) Kim, K. H.; Greco, G.; Novellino, E. A critical review of recent CoMFA applications. *Persp. Drug Discovery Des.* **1998**, *12*, 257–315.
- (60) Cho, S. J.; Tropsha, A. Cross-validated R2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.
- (61) Cruciani, G.; Clementi, S.; Pastor, M. GOLPE-guided region selection. *Persp. Drug. Discovery Dev.* **1998**, *12/13/14*, 71–86.
- (62) Böhm, M. Ph.D. Thesis, Philipps University, Marburg, Germany, 2002.
- (63) Böhm, M.; Stürzebecher, J.; Klebe, G. Three-dimensional quantitative structure-activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *J. Med. Chem.* **1999**, *42*, 458–477.
- (64) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- (65) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (66) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. B.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; The Scripps Research Institute: La Jolla, CA, 2004.
- (67) Gerber, P. R.; Müller, K. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 251–268.
- (68) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (69) Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T. J.; Higgins, D. G.; Thompson, J. D. Multiple sequence alignment with the Clustal series of programs. *Nucl. Acids Res.* **2003**, *31*, 3497–3500.
- (70) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.



- (71) Krieger, E.; Nabuurs, S. B.; Vriend, G. Homology modeling. In *Structural Bioinformatics*; Bourne, P. E., Weissig, H., Eds.; Wiley-Liss: Hoboken, NJ, 2003; pp 509–524.
- (72) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (73) Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins* **1993**, *17*, 355–362.
- (74) SYBYL *Molecular Modeling Software*, 7.3; Tripos, Inc.: St. Louis, MO, 2006.
- (75) Sotriffer, C. A.; Gohlke, H.; Klebe, G. Docking into knowledge-based potential fields: A comparative evaluation of DrugScore. *J. Med. Chem.* **2002**, *45*, 1967–1970.
- (76) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., II. The Collinearity Problem in Linear Regression. The Partial Least Squares Approach to Generalized Inverses. *SIAM J. Sci. Stat Comput.* **1984**, *5*, 735–743.
- (77) Wold, S.; Johansson, E.; Cocchi, M. PLS—Partial Least Squares Projections to Latent Structures. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993.
- (78) Kubinyi, H.; Abraham, U. Practical Problems in PLS Analyses. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 717–728.
- (79) Stürzebecher, J.; Prasa, D.; Wikstrom, P.; Vieweg, H. Structure–activity relationships of inhibitors derived from 3-amidinophenylalanine. *J. Enzyme Inhib.* **1995**, *9*, 87–99.
- (80) Golbraikh, A.; Bernard, P.; Chretien, J. R. Validation of protein-based alignment in 3D quantitative structure–activity relationships with CoMFA models. *Eur. J. Med. Chem.* **2000**, *35*, 123–136.
- (81) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993.
- (82) Golbraikh, A.; Tropsha, A. Beware of  $q^2$ ! *J. Mol. Graph. Modell.* **2002**, *20*, 269–276.
- (83) Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y. D.; Lee, K. H.; Tropsha, A. Rational selection of training and test sets for the development of validated QSAR models. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 241–53.
- (84) Luque, I.; Freire, E. Structural Stability of Binding Sites: Consequences for Binding Affinity and Allosteric Effects. *Proteins* **2000**, *4*, 63–71.
- (85) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins* **2000**, *39*, 261–268.
- (86) Gunther, J.; Bergner, A.; Hendlich, M.; Klebe, G. Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.* **2003**, *326*, 621–636.
- (87) Reunanen, J. Overfitting in making comparisons between variable selection methods. *J. Machine Learn. Res.* **2003**, *3*, 1371–1382.
- (88) Baumann, K.; von Korff, M.; Albert, H. A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part II. Practical applications. *J. Chemom.* **2002**, *16*, 351–360.
- (89) Charifson, P. S.; Corkerey, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
- (90) Wang, R.; Wang, S. How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1422–1426.
- (91) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: San Carlos, CA 94068.

CI7002472