Journal of Medicinal Chemistry



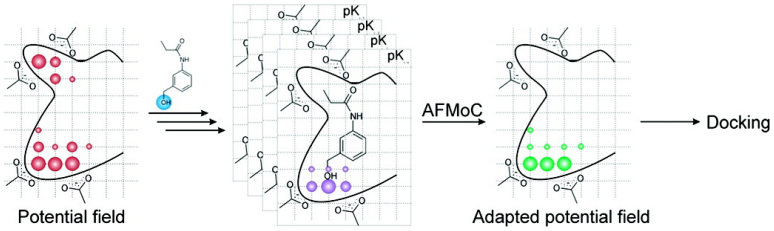
Article

# Improving Binding Mode Predictions by Docking into Protein-Specifically Adapted Potential Fields

Sebastian Radestock, Markus Bhm, and Holger Gohlke

*J. Med. Chem.*, **2005**, 48 (17), 5466-5479• DOI: 10.1021/jm050114r • Publication Date (Web): 26 July 2005






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 8 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML



Journal of Medicinal Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

# Improving Binding Mode Predictions by Docking into Protein-Specifically Adapted Potential Fields

Sebastian Radestock,[§] Markus Böhm,[#] and Holger Gohlke*[,§]

*Department of Biology and Computer Science, J. W. Goethe University, Frankfurt, Germany, and Pfizer Global R&D, Eastern Point Road, Groton, Connecticut 06340*

*Received February 5, 2005*

The development of a protein-specifically adapted objective function for docking is described. Structural and energetic information about known protein−ligand complexes is exploited to tailor knowledge-based potentials using a "reverse", protein-based CoMFA-type (=AFMoC) approach. That way, effects due to protein flexibility and information about multiple solvation schemes can be implicitly incorporated. Compared to the application of AFMoC for binding affinity predictions, a Shannon entropy based column filtering of the descriptor matrix and the capping of adapted repulsive potentials within the binding site have turned out to be crucial for the success of this method. The new developed approach ($AFMoC^{obj}$) was validated on a data set of 66 HIV-1 protease inhibitors, for which experimental structural information was available. Convincingly, for ligands with up to 20 rotatable bonds, in more than 75% of all cases a binding mode below 2 Å rmsd has been identified on the first scoring rank when $AFMoC^{obj}$-based potentials were used as the objective function in AutoDock. With respect to nonadapted DrugScore or AutoDock fields, the binding mode prediction accuracy was significantly improved by 14%. Noteworthy, very similar results were obtained for training and test set compounds, demonstrating the strength and robustness of this method. Implications of our findings for binding affinity predictions and its usage in virtual screening are further discussed.

## Introduction

Driven by the insight that biological space (representing all drug-relevant targets) is considerably smaller than chemical space (representing all synthesizable compounds), it is advisable to let biological structures guide chemistry in the process of drug discovery, not the other way around.[1] Accordingly, structure-based drug design techniques are increasingly used for identifying hits and optimizing lead compounds.[2] Of central importance to the structure-based in silico approaches is the ability to generate and identify relevant binding modes and to accurately predict binding affinities of small molecules to their macromolecular targets by docking techniques.[3−5] Therefore, obtaining "correct" binding modes (i.e., those that are close to the native structure) is a necessary, albeit not sufficient, prerequisite for a reliable estimation of binding affinities.[6,7]

Underlying docking approaches are the principles of molecular recognition and chemical complementarity. Usually, docking solutions are generated by solely considering information about the determinants of binding provided by both binding partners. This allows for exhaustive sampling of the configurational space of the ligand within the binding site, providing a highly diverse set of docking solutions. In addition, because of the lack of any bias, this approach advantageously offers the opportunity of "scaffold hopping"[8] in drug discovery; that is, it allows finding new classes of leads. However, these "pure" docking techniques do not take into account any a priori knowledge about ligands already known to bind to the receptor, neither with respect to the chemical properties of a ligand nor with respect to the location and/or conformation of a ligand in the binding site.

Interestingly, another important principle known in the field of drug design is molecular similarity, where it is exploited that structurally similar molecules should provide similar biological responses. Initially applied to drug design problems where the structure of the target is unknown, the principle has formed the basis of classical QSAR, 3D QSAR,[9] and superimposition techniques,[10] as well as molecular similarity/diversity approaches used to design (virtual) libraries.[11]

Only very recently, both the principles of molecular similarity and molecular recognition have been combined in structure-based drug design applications for the benefit of increased speed and quality of the obtained solutions.[12] As an initial step toward such a strategy, tools for molecular comparison were applied consecutively or iteratively with docking in virtual screening.[13] In addition, ligand information was considered to improve the model building of proteins by homology.[14,15]

In the field of docking, similarity-driven approaches have been introduced that take advantage of additional *structural* information about ligands known to bind to the target, as well as protein-based information, such as receptor-based pharmacophores or "hot-spot" analyses of binding sites.[16,17] Following a classification scheme proposed by Fradera and Mestres,[12] *indirect* approaches incorporate chemical information only implicitly by

* To whom correspondence should be addressed. Address: Marie-Curie-Strasse 9, 60439 Frankfurt, Germany. Phone: (+49) 69 798-29411. Fax: (+49) 69 798-29826. E-mail: gohlke@bioinformatik.uni-frankfurt.de.

§ J. W. Goethe University.

# Pfizer Global R&D.

10.1021/jm050114r CCC: $30.25 
Published on Web 07/26/2005

having an effect on scoring but not on orienting the ligand during sampling, whereby *direct* approaches actively guide the sampling within the binding site.

Accordingly, Fradera et al.[18] have described modifications to the DOCK 4.0 program such that similarity scores derived from small-molecule superpositioning are used to weight the DOCK energy score. Extracting information from experimentally resolved protein−ligand complexes, a (indirect) "similarity-penalized" and a (direct) "similarity-guided" variant have been introduced. In another approach, FlexX-Pharm[19] has been developed as an extended version of the flexible docking tool FlexX, in which interactions or spatial constraints are derived from receptor-based pharmacophore features. These constraints are then used as filters to keep or reject docking solutions, either as a postdocking filter or by look-ahead checks during construction of the ligand in the binding site. Pharmacophore points are also used in a recent study by Daeyaert et al.[20] to initially place pregenerated conformers of small molecules into receptor binding sites. Finally, Wu and Vieth[21] have augmented the CHARMM force field with an additional similarity force derived from the positions of ligands from crystallographic data such that the simulated annealing molecular dynamics based sampling of the ligand is focused on the relevant binding site regions.

One shortcoming of the above methods is that they are restricted to exploit only structural a priori information. Since energetic information is neglected, these methods do not provide an implicit means of weighting the influence of the similarity information with respect to the location in the binding pocket. However, this limitation has been overcome at least qualitatively by defining "essential" and "optional" constraints in the FlexX-Pharm approach.[19] Yet a more fine-grained tuning would be desirable, which can be achieved by combining structural and energetic information about known ligands. In a way, this paradigm has been pursued in so-called tailor-made scoring functions specifically adapted to predict binding affinities with respect to one particular protein.[22−25] Here, structure-based properties (e.g., interaction energy components derived from experimentally determined protein−ligand complexes) are correlated to experimental binding affinities.

In the present study, we introduce a novel approach that exploits structural and energetic information about known protein−ligand complexes in order to tailor an objective function for docking with respect to one particular protein. Particularly, we aim at improving binding mode predictions, motivated by findings from recent studies that (I) accurate predictions of binding affinities critically depend on finding correct binding conformations,[6,7] (II) docking programs need to produce reliable binding modes to obtain good enrichments in virtual screening,[26,27] and (III) docking techniques are increasingly used to support lead optimization efforts,[28] in which docking of analogues can be facilitated by the application of similarity-driven algorithms.

Our strategy rests upon the DrugScore[29] and AFMoC[25] approaches developed by us. DrugScore is a knowledge-based scoring function that has been derived from crystallographically determined protein−ligand complexes. This function was shown to reliably recognize correct binding modes and accurately predict binding affinities.[30−32] The observation that DrugScore also predicts which type of ligand atom will most favorably bind at a given site in a protein binding pocket (resulting in a "hot-spot" analysis)[17] prompted us to apply DrugScore as an objective function in docking optimizations.[33] Thus, the Lamarckian genetic algorithm of AutoDock has been successfully used to search for favorable ligand binding modes, guided by DrugScore potential fields as representations of the binding pocket. In a parallel development, knowledge-based potentials to score binding modes using protein information are combined with techniques from comparative molecular field analyses. As a result, knowledge-based potentials are specifically adapted to a particular protein in a CoMFA-type approach.[25] This protein-based, or with respect to the origin of the fields, "reverse" CoMFA (=AFMoC) allows us to gradually move from general knowledge-based to protein-specifically adapted potentials, depending on the amount of ligand data available for training or the desired degree of generality/specificity for predictions.

Combining both routes as described above naturally leads to integrating the principles of molecular recognition and molecular similarity. Hence, we show here that using potential fields, which have been protein-specifically adapted by a modified AFMoC approach ($AFMoC^{obj}$), as an objective function represents a promising tool for improving docking accuracy with respect to unmodified DrugScore fields or fields generated by the AutoDock function. By the mapping of differences in the energetic information, our approach thereby provides an implicit weighting of the structural information with respect to the location in the binding site.

The methodology, subsequently referred to as $AFMoC^{obj}$-based docking, has been tested on HIV-1 protease inhibitors. In total, 48 crystallographically determined structures of bound HIV-1 protease inhibitors were used to derive AFMoC models and to calculate the adapted potential fields that are then used in docking. To put the results into proper perspective, ligands were also subjected to docking with the regression-based energy function of AutoDock and to docking with DrugScore potentials fields. Encouragingly, $AFMoC^{obj}$-based docking improved binding mode predictions by 14% compared to the nonadapted approaches.

## Methods

**General Strategy.** Here, we outline the general strategy to calculate protein-specifically adapted potential fields pursued in this study (Figure 1). Details of the single steps are given in the paragraphs below. First, atom-type specific potential fields are generated by mapping distance-dependent pair potentials between protein atoms and ligand probe atoms onto a cubic grid located in the binding pocket,[17] applying the DrugScore scoring function[29] (Figure 1, top left). In this step, only structural information of the protein environment is taken into account. Following the AFMoC approach,[25] interaction fields are then calculated for each ligand in the training set by "multiplying" atom-type specific properties of the ligand placed in the binding pocket with neighboring grid values (Figure 1, top right). As a

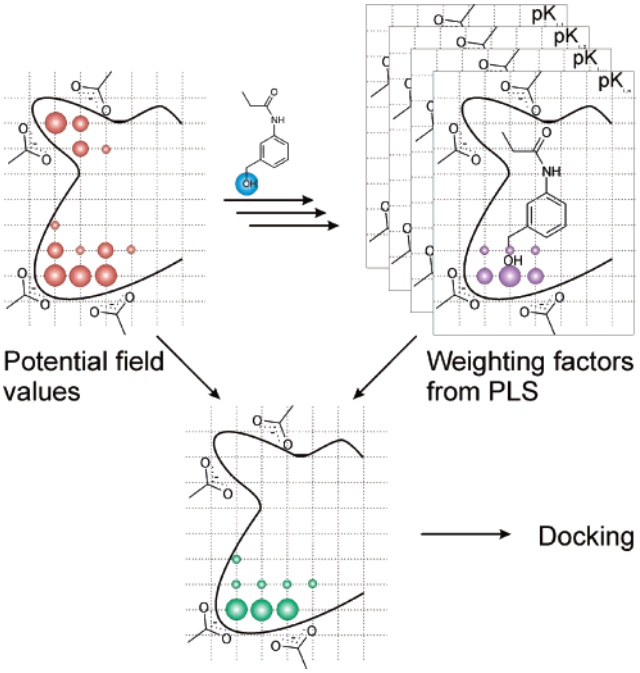



**Figure 1.** Calculation of protein-specifically adapted potential fields. See text for details.

result, this step considers structural information of the ligand molecules. In the next step, interaction field values are correlated to experimentally determined binding affinities ("energetic information") of the training set ligands by means of partial least squares (PLS) analysis, resulting in individual weighting factors for each field position. Applying these weighting factors to the original DrugScore potential fields finally produces protein-specifically adapted potential fields (Figure 1, bottom) that are used as the objective function in docking ($AFMoC^{obj}$).

**Data Sets and Alignments.** Ligand structures from 66 crystallographically determined HIV-1 protease inhibitor complexes were extracted from the Protein Data Bank (PDB). Only complexes with wild-type protein or protein mutants that do not influence the binding mode or the binding affinity were considered.[34−37] The data set comprises 48 compounds for training (Table 1 in Supporting Information) and 18 compounds for testing (Table 2 in Supporting Information). Training compounds with experimentally determined $pK_i$ values were used in generating the AFMoC model. The $pK_i$ values spread over a range of 5.7 log units and showed a balanced distribution, albeit the range between $pK_i$ = 6 and 7 was underrepresented. Compounds for which either no (e.g., 1htf) $pK_i$ value or multiple (e.g., 1aaq) $pK_i$ values were found have been put into the test set. In both data sets, peptidomimetic inhibitors outweigh ligands of other substance classes.

Since the protein structure is considered to be rigid during AFMoC analysis and docking, ligand structures were taken from a structural superimposition of all binding sites using the database ReLiBase+.[38] They were then minimized within a single representative active site to obtain a consistent alignment of all inhibitor molecules using the MAB force field as implemented in Moloc.[39] As a representative active site, the protein structure of the PDB entry 1ajv (resolution: 2 Å) was chosen,[40] which shows minimal conformational changes of binding site residues when compared to the unbound structure (rmsd of binding site residues of ~0.3 Å) or when compared to other bound structures of our data set (average rmsd of binding site residues of ~0.7 Å). The protonation states of the catalytic aspartic acid residues (Asp25 and Asp25′) in the active site were assigned according to a study of Kulkarni and Kulkarni, who found that one of the two catalytic aspartic acid residues is monoprotonated with a proton placed on the outer oxygen of the side chain.[41] Similar results have been obtained by other experimental or theoretical studies.[42−44] Accordingly, in our case the outer oxygen of D25 was protonated. For calculating the DrugScore or AFMoC fields, no further modification of the enzyme structure was necessary. For calculating AutoDock maps, polar hydrogens were added and Kollman united-atom partial charges[45] were assigned.

In general, the inhibitors in our data set represent two structurally distinct classes, peptidomimetic inhibitors and nonpeptidomimetic inhibitors (comprising cyclic ureas, cyclic sulfonamides, hydroxypyrone derivates, and others). While most of the inhibitors of the latter class do not show any specific interactions to structural water molecules, the peptidomimetic inhibitors bind to the so-called flap water. The flap water forms a bridge between the inhibitor and the amide hydrogens of I50 and I50′ in the flap regions of the protease in a tetrahedral arrangement and is considered to be important for inhibitor binding.[46] Hence, peptidomimetic inhibitors were minimized in the presence of the flap water. During this step, the water molecule was allowed to move.

To take into account that inhibitors may bind in two different binding modes due to rotational symmetry,[47] a second data set was created that contained the inhibitors in a "turned" orientation. This data set was obtained by superimposing the representative protein active site residues of chain A onto the corresponding residues of chain B and applying the resulting transformation to the ligands as well.

Since one does not know a priori which of the two symmetry-related orientations of a ligand (i.e., the "turned" or the "nonturned" orientation) should be considered for AFMoC analysis, we embarked to find the superimposition that resulted in the best AFMoC model. Hereby, model quality is measured by means of the cross-validated $r^2$ ($q^2$). In that sense, our approach is similar in spirit to the work of Schaal et al.[48] However, because of the larger number of components, a systematic search over all possible combinations of ligand orientations is computationally intractable. Hence, we applied a greedy optimization procedure. Starting from an arbitrarily selected set of "turned" and "nonturned" orientations of the training set molecules, the orientation of a randomly chosen molecule was flipped. The new superimposed set was kept as starting point for the next cycle only if it resulted in a better model. After optimization of all inhibitors, a new round with the hitherto obtained superimposition was initiated. This step was repeated until no further improvement in the model quality was found.

A literature survey indicated that the binding assays for compounds contained in our data set have been performed at pH values between 4.7 and 6.5. Thus,

standard protonation states were assumed, i.e., carboxylate and phosphate groups were considered to be deprotonated, while aliphatic amino groups were considered to be protonated. Possible changes in the protonation states upon binding to the protein were neglected. Finally, for calculating AutoDock maps, Gasteiger−Marsili charges[49] were assigned to each inhibitor.

**AFMoC Analysis.** Interaction fields for the superimposed inhibitors were calculated as described elsewhere[25] using the protein structure of 1ajv as reference. The size of the grid box was selected such that all inhibitor molecules were sufficiently embedded with a margin of at least 4 Å. Fields were initially calculated for seven inhibitor atom types as stated below. To study the influence of the grid spacing on the quality of the AFMoC models, values of 0.375 and 1.0 Å were tested. However, if not stated differently, results are reported for 0.375 Å grid spacing, which is also recommended for setting up the maps in AutoDock. We further ensured that AFMoC grids had the same dimensions and positions as DrugScore grids and AutoDock maps in order to obtain comparable results. 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 25 for details). In agreement with an earlier study,[33] a multiplication factor 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. Variations of the steepness of the repulsion function did not significantly alter the results (data not shown).

Interaction field values were then correlated with experimentally determined binding affinities using partial least squares regression (PLS)[50,51] similar to CoMFA.[52] Statistical analyses were performed using an in-house implementation of the (SAM)PLS algorithms.[53] The columns of the descriptor matrix have been centered on input. Since 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 check the statistical significance of the obtained PLS models, cross-validation runs were performed by means of the "leave-one-out" (LOO) procedure using the enhanced SAMPLS method, which results in the cross-validated $r^2$ ($q^2$). Following recommendations of Wold,[50] Kubinyi et al.,[54] and Bush et al.,[53] the columns of the descriptor matrix were recentered for every new cross-validation run. The optimal number of components has been determined as the one that resulted in a final increase of the $q^2$ value by 5% (i.e., adding another component increased the $q^2$ by less than 5%). The same number of components was subsequently used to derive the final AFMoC model, applying a "minimum $\sigma$" standard deviation threshold value of 1.

Statistical results are summarized in Tables 1 and 2. The $q^2$, $s_{PRESS}$, $r^2$, $S$, and contribution values were computed as defined in ref 52. To detect possible chance correlations, the biological data were randomly scrambled and model calculations were repeated. Only negative $q^2$ values were obtained in this case (Table 3), supporting the assumption that meaningful correlations with the biological data are given in the nonscrambled case.

Upon calculation of 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 tend to average adjacent interactions. To study the influence of $\sigma$, values of 0.4, 0.55, 0.7, ..., 1.15 Å were used to calculate interaction fields. However, if not stated differently, a $\sigma$ value of 0.7 Å has been applied, in agreement with a recent study.[25]

Since the interaction fields obtained by mapping protein−inhibitor interactions onto neighboring grid

**Table 1.** Statistical Results for AFMoC Analyses[a]

| | spacing[b] | |
|---|---|---|
| | 1.0 | 0.375 |
| $q^2$ [c,d] | 0.66 (0.42) | 0.65 (0.40) |
| $s_{press}$ [e,f] | 1.09 | 1.11 |
| $r^2$ [c,g] | 0.97 (0.94) | 0.97 (0.94) |
| $S$ [e,h] | 0.34 | 0.35 |
| $F$ [c,i] | 319.8 (181.9) | 297.6 (167.8) |
| components[j] | 4 | 4 |
| fraction | | |
| C.3 | 0.356 | 0.356 |
| C.ar | 0.555 | 0.554 |
| O.3 | 0.041 | 0.039 |
| O.2 | 0.017 | 0.019 |
| O.co2 | 0.001 | 0.001 |
| N.am | 0.024 | 0.025 |
| S.3 | 0.006 | 0.007 |

[a] Models are based on PLS calculations using a $\sigma$ value of 0.7 Å and grid spacings of 1.0 and 0.375 Å. [b] In Å. [c] Values are given considering only $pK_{i,PLS}$ or considering $pK_{i,total}$ (values in parentheses). [d] $q^2 = 1 - PRESS/SSD$ as obtained by "leave-one-out" 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. [e] In logarithmic units. [f] $s_{PRESS} = [PRESS/(n - h - 1)]^{1/2}$ as obtained by "leave-one-out" cross-validation. $n$ equals the number of data points. $h$ is the number of components. [g] Correlation coefficient. [h] $S = [RSS/(n - h - 1)]^{1/2}$. RSS equals the sum of squared differences between fitted and experimentally determined binding affinities. [i] Fischer's $F$ value. [j] Number of components.

**Table 2.** Statistical Results for 10 Runs of "Leave-Five-Out" Cross-Validation[a]

| no. of run | $q^2$ [b] | $S_{press}$ [c] | components[d] |
|---|---|---|---|
| 1 | 0.65 (0.40) | 1.1 | 3 |
| 2 | 0.67 (0.43) | 1.08 | 4 |
| 3 | 0.66 (0.41) | 1.08 | 3 |
| 4 | 0.74 (0.55) | 0.96 | 4 |
| 5 | 0.68 (0.45) | 1.05 | 3 |
| 6 | 0.69 (0.46) | 1.04 | 3 |
| 7 | 0.63 (0.35) | 1.15 | 4 |
| 8 | 0.65 (0.39) | 1.12 | 4 |
| 9 | 0.74 (0.55) | 0.94 | 3 |
| 10 | 0.72 (0.51) | 0.99 | 3 |
| LOO[e] | 0.65 (0.40) | 1.11 | 4 |

[a] Models are based on PLS calculations using a $\sigma$ value of 0.7 Å and a grid spacing of 0.375 Å. [b] Values are given considering only $pK_{i,PLS}$ or considering $pK_{i,total}$ (values in parentheses). [c] In logarithmic units. [d] Number of components. [e] For comparison, results of "leave-one-out" cross-validation are shown.

**Table 3.** $q^2$ Values for AFMoC Models Obtained with Randomly Scrambled Affinity Data[a]

| no. of components | $q^2$ [b] |
|---|---|
| 1 | −0.23 (−1.13) |
| 2 | −0.37 (−1.37) |
| 3 | −0.31 (−1.26) |
| 4 | −0.35 (−1.33) |
| 5 | −0.23 (−1.13) |
| 6 | −0.22 (−1.11) |
| 7 | −0.21 (−1.10) |
| 8 | −0.18 (−1.00) |
| 9 | −0.19 (−1.10) |
| 10 | −0.19 (−1.10) |

[a] Models are based on PLS calculations using a $\sigma$ value of 0.7 Å and a grid spacing of 0.375 Å. [b] Values are given considering only $pK_{i,PLS}$ or considering $pK_{i,total}$ (values in parentheses).

**Table 4.** $q^2$ Values Obtained by AFMoC Using Different Combinations of Interaction Fields[a]

| combination of fields[b] | $q^2$ [c] |
|---|---|
| C.3/C.ar/O.3/O.2/O.co2/N.am/S.3 | 0.65 (0.4) |
| C.3/C.ar/O.3/O.2/N.am/S.3 | 0.65 (0.37) |
| C.3/C.ar/O.3/O.2/O.co2/N.am | 0.67 (0.44) |
| **C.3/C.ar/O.3/O.2/N.am** | **0.66 (0.42)** |
| C.3/C.ar/O.3/N.am | 0.68 (0.30) |

[a] Models are based on PLS calculations using a $\sigma$ value of 0.7 Å and a grid spacing of 0.375 Å. [b] The notation of the combination of interaction fields follows the atom type convention of SYBYL.[73] Further details are given in ref 29. [c] Values are given considering only $pK_{i,PLS}$ or considering $pK_{i,total}$ (values in parentheses).

points are ligand atom-type specific, various combinations can be used to build up the descriptor matrix as input for the PLS analysis. Starting with a combination of seven fields, which to our knowledge represent best the interaction differences of the considered ligands, we subsequently reduced the number of fields to obtain a simplified model following the parsimony principle (Table 4). The combination of five fields shown in bold was finally used for all calculations throughout this study.

One of the reasons for the successful application of knowledge-based potentials in docking and recognizing near-native binding modes is their smooth funnel-shape compared to traditional force-fields.[31,32] However, adapted AFMoC fields appear more localized upon visual inspection and, hence, are expected to be rather rugged. To obtain an accurate docking function that still tolerates imperfect ligand placement in a rigid binding pocket, adapted fields and unmodified DrugScore potentials were linearly mixed. In this study, the mixing parameter $\theta$ was set to 0.0, 0.25, 0.5, 0.75, and 1.0, with $\theta$ = 0.0 reflecting "pure" DrugScore and $\theta$ = 1.0 reflecting "pure" AFMoC potentials.

Since PLS analysis relates differences in interaction fields to differences in binding affinity, regression coefficients in the QSAR equation become small for regions within the binding pocket that do not show a significant variation in the interaction field values (e.g., because all ligands have similar substituents). Therefore, these regions will not contribute significantly to the docking function, albeit key interactions may still occur as demonstrated by the similar substitution of all inhibitors. In these cases, using "pure" DrugScore potentials instead of the AFMoC fields may be favorable. To remove those columns from PLS analysis that have little variance, we initially increased the "minimum $\sigma$" standard deviation threshold. However, since our fields have not been scaled and the "minimum $\sigma$" measures the information content of a column in absolute terms, applying an equal $\sigma$ value of the order of 100 to all field types resulted in nonsignificant AFMoC models. Hence, we decided to filter columns according to their Shannon entropy, which provides a relative measure of the information content for each column. Therefore, minimal and maximal field values were determined for each field type. This range was then divided into 15 equally sized bins, and by assignment of column values to these bins, a histogram of each column was obtained. From the normalized frequency distribution of values $p_i$ in bin $i$, the Shannon entropy $S$ [55] of a column was calculated according to

$$S = -\sum_{i=1}^{15} p_i \operatorname{ld} p_i \qquad (1)$$

Columns with Shannon entropies below a given threshold were excluded from PLS analysis. Thresholds ranging from 0.25, 0.275, ..., 0.325, 0.35 times the maximal possible entropy $S_{max} = \operatorname{ld} N$ (with $N$ equaling the number of values in a column) have been tested. To account for the fact that excluded columns still represent interactions of ligand atoms with the protein, the experimentally determined binding affinity ($pK_{i,total}$) was reduced by the contributions of these columns prior to PLS analysis (resulting in $pK_{i,PLS}$). This procedure is similar to the one applied to contributions by ligand atom types that are not included in the finally chosen set of field types.[25] In the prediction step, discarded columns were replaced by DrugScore potential field values.

Correlating binding affinities with interaction fields results in individual weighting factors for each field position. By application of these weighting factors to the original DrugScore potential fields, protein-specifically adapted potential fields are obtained. Some of the regions of these adapted fields favor ligand binding more than the original DrugScore fields; hence, these regions become "attractors" of ligand atoms of a given type in docking. Not unexpectedly, the opposite trend is also observed, resulting in adapted field regions that disfavor ligand binding compared to DrugScore. In some cases, previously favorable regions within the binding site may even become repulsive. While these regions are important for correctly predicting binding affinities based on the adapted fields, initial docking studies revealed that penalizing the placement of ligand substructures that way hampers the prediction of accurate binding *modes*. Hence, we decided to postprocess the AFMoC-based potential fields prior to docking by replacing all columns that (I) have favorable values in the original fields and (II) become repulsive in the adapted fields, with original DrugScore values or with zero. This way it is ensured that previously repulsive field regions (e.g., in the vicinity of the protein) are not changed.

The resulting objective function that is based on the above modifications to the original AFMoC approach will be referred to as AFMoC$^{obj}$.

**Docking.** All docking runs were performed applying the Lamarckian genetic algorithm of AutoDock 3.0.[56] AutoDock maps were generated using the AutoGrid utility from the AutoDock package. The grid spacing

was set to default (i.e., 0.375 Å). DrugScore, AFMoC, and AFMoC$^{obj}$ grids were calculated as follows: First, unweighted potential fields were calculated using a grid spacing of 0.375 Å. In the case of AFMoC, potential field values were multiplied by the coefficients of the QSAR equation. Then, as stated above, DrugScore, AFMoC, and AFMoC$^{obj}$ field values were mixed according to the mixing parameter $\theta$. The "raw" grids were finally transformed into an AutoDock map format. In accordance with earlier studies,[33] a scaling factor of 2.5 $\times$ $10^{-5}$ for DrugScore values was used in all cases.

The standard docking protocol for flexible inhibitor docking consisted of 100 independent runs per inhibitor, using an initial population of 50 randomly placed individuals, a maximum number of 5 $\times$ $10^5$ energy evaluations, a mutation rate of 0.02, a crossover rate of 0.8, and an elitism value of 1. Docked conformations differing by less than 1 Å rmsd from each other were clustered together and represented by the result with the best docking energy. As docking energy, the sum of inter- and intramolecular scores was taken. Of all 66 inhibitors, those with more than 20 rotatable bonds did not result in converged docking runs, as judged from the cluster size distribution. Furthermore, no improvement was observed by increasing the population size and the number of energy evaluations. Thus, only 52 ligands (training set, 34; test set, 18) with less than 20 rotatable bonds were finally used for docking. Binding modes are considered "well-docked" if they have an rmsd value of <2.0 Å with respect to the minimized reference structure. Both the "turned" and "nonturned" reference poses were used for evaluating the rmsd value, and the lower of the two calculated values was finally chosen. The percentage of complexes for which a "well-docked" binding mode is found on the first scoring rank is termed "prediction accuracy" in the following.

## Results and Discussion

**Data Sets and Alignment Procedure.** To validate the AFMoC$^{obj}$-based docking, a data set of 66 HIV-1 protease inhibitors was selected. The data set includes a diverse collection of inhibitors, varying in size, flexibility, and chemical structure. Molecular structures and properties of these compounds are listed in Tables 1 and 2 in Supporting Information. For all selected inhibitors, experimentally determined binding modes were available from the Protein Data Bank (PDB). While these structures provided ligand reference orientations for the validation of our approach, we note that using experimentally determined binding modes is not mandatory in general because significant AFMoC models have also been derived from data sets consisting of modeled ligand poses.[25] To obtain a structural alignment of all ligands, the inhibitors were minimized in the rigid binding pocket of 1ajv after superimposition of the respective protein structures (see Methods). On average, the minimized ligand orientations deviate by 1.28 Å from their corresponding crystal structure except for nine cases (PDB entries: 1dif, 1hpo, 1htf, 1hvi, 1hvj, 1hvk, 1hvl, 1hwx, 3aid) where minimized ligand poses deviate by more than 2 Å. Visual inspection of these rather large and flexible compounds revealed that deviations are predominantly located in the terminal parts of the molecules that lie at the outer regions of the binding pocket. Thus, we decided to retain these compounds in the data set. The p$K_i$ values of the 48 training compounds used for deriving the AFMoC model spread over a satisfactorily large range of more than 5 logarithmic units.[57]

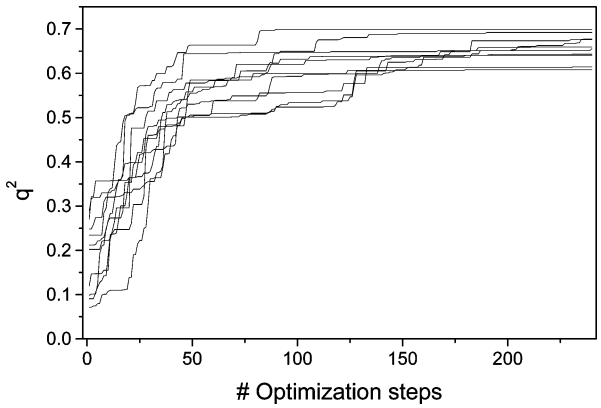



**Figure 2.** Convergence of the greedy optimization procedure to obtain a ligand superimposition. Depicted is the $q^2$ value versus the number of optimization steps. For the sake of clarity, only 10 randomly chosen optimization runs out of 70 are shown.

For AFMoC analyses, the alignment procedure needs to consider the structure of the target protein. The binding pocket of HIV-1 protease shows a 2-fold rotational symmetry, and hence, each inhibitor can bind to the enzyme in two different orientations [47] (denoted "turned" and "nonturned"; see Methods). Thus, to obtain a unique ligand superimposition, one needs to choose one out of the two possible orientations for each ligand.[48] Since in our case a systematic search would have resulted in a computationally intractable number of $2^{48}$ $\approx$ $10^{14.4}$ combinations, a greedy optimization procedure was applied to choose the orientation of each ligand that gives the best AFMoC model with the resulting superimposition. Altogether, five optimization rounds per run were performed, after which the quality of the model has usually converged (Figure 2). The results of about 70 independent runs were analyzed. Clustering of similar superimpositions yielded four highly populated clusters (number of cluster members: 17, 11, 12, and 9), and all corresponding AFMoC models showed a high statistical significance (i.e., with $q^2$ > 0.5; data not shown). The best superimposition among the four clusters was finally selected as the one that resulted in the best AFMoC model (Figure 3).

Optimizing QSAR models by optimizing ligand alignments has been repeatedly criticized.[58,59] In fact, when applying our greedy optimization procedure on the training data with randomly scrambled p$K_i$ values, we were also able to obtain "good" AFMoC models, which is in line with the recent finding that the $q^2$ value may be an inadequate characteristic to assess the predictive ability of models.[60] Nevertheless, these models did not show any predictive power when used as an objective function in docking (data not shown). As judged from the docking results of the AFMoC model obtained with the correct p$K_i$ (see below), we are thus confident that this model is significant. We also emphasize that no new ligand conformations or orientations were generated for the final superimposition, but rather only one out of two given alternatives per ligand was chosen.

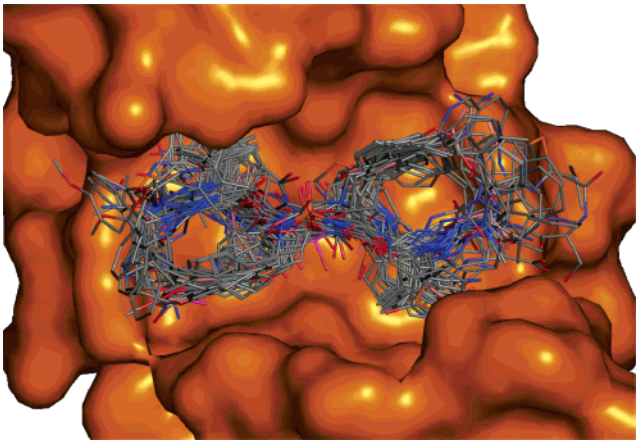


**Figure 3.** Alignment of the HIV-1 protease inhibitors in the binding pocket of 1ajv. Residues of the "flap" regions have been removed for the sake of clarity.

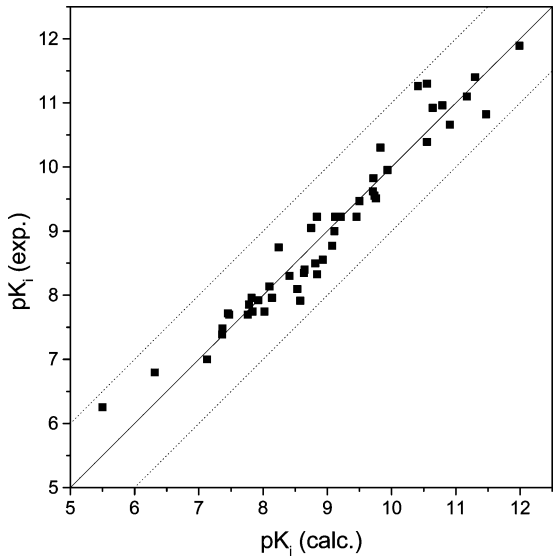



**Figure 4.** Experimentally determined binding affinities versus fitted predictions for the training set of 48 HIV-1 protease inhibitors. In addition to the line of ideal correlation, dotted lines are given that indicate deviations from the actual p$K_i$ value by ±1 logarithmic unit.

**Significance and Robustness of AFMoC Analyses.** Table 1 summarizes the statistical results of AFMoC analyses with respect to grid spacings of 0.375 and 1.0 Å, respectively, using the optimized ligand superimposition. Models with $q^2 > 0.5$ are generally accepted as significant, and models with $q^2 > 0.3$ are considered "good".[61] For both grid spacings, $q^2 > 0.5$ has been obtained for the part of p$K_i$ being adapted during PLS analysis while considering the total binding affinity still results in $q^2 > 0.4$. Hence, the dependence on the grid spacing is negligible, and a value of 0.375 Å has been used for further calculations. In this case, for 10 runs of "leave-five-out" cross-validation, $q^2$ values greater than 0.5 were obtained (Table 2). The plot of predicted versus actual binding affinities for the PLS analysis (Figure 4) does not reveal any significant over- or underprediction across the whole activity range. Furthermore, no trend can be observed for residuals across different classes of compounds. Thus, both observations suggest that the AFMoC model represents the whole

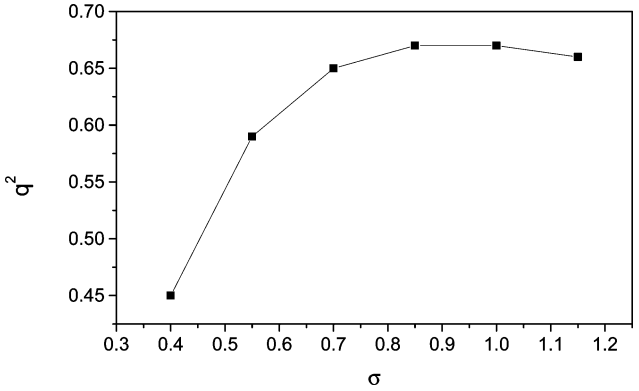



**Figure 5.** Dependence of $q^2$ on the magnitude of $\sigma$ values that govern the degree of "smearing" protein−ligand interactions across neighboring grid points. For $\sigma > 0.55$ Å, significant models are obtained.

data set of molecules. Further indication for the significance of the obtained model is provided by the randomly scrambled data that do not allow one to obtain reasonable models (Table 3). We note that the discussion on the significance of the model could be considerably strengthened by testing the predictive power of the model with respect to binding affinities with an external test set that has not been used during model derivation. However, the test set used in our study consists of all those compounds for which either no p$K_i$ or multiple p$K_i$ values were found (see Methods). Although this choice renders impossible the use for testing affinity predictions, it complies with the main purpose of our study to devise a method to improve binding mode predictions and, at the same time, to include as much structural information as is available from the PDB. Finally, as discussed in more detail below, obtaining very similar results for the training and test sets in the case of binding mode predictions, in our opinion, provides a strong hint to the significance of the model.

Contributions by interaction fields to explain binding affinity differences are listed in Table 1. The fields of atom types C.3 and C.ar contributed considerably more than fields calculated for polar atoms, with O.co2 and S.3 fields showing the least contribution. This finding may be explained by the frequency of occurrence for these atom types in the ligand data set, which led us to test the influence of different interaction field combinations on the $q^2$ value (Table 4). The final model was derived using only five fields (C.3, C.ar, O.3, O.2, N.am) with a $q^2$ value comparable to the seven-field model, both for the part of binding affinity considered in PLS analysis (p$K_{i,PLS}$) as well as the total binding affinity (p$K_{i,total}$).

Figure 5 shows the dependence of the $q^2$ values with respect to variations in the $\sigma$ value that governs the degree of "smearing" protein−ligand interactions across neighboring grid points. Selecting $\sigma < 0.55$ Å results in models with reduced significance, whereas the dependence of $q^2$ on $\sigma$ values between 0.7 and 1.15 Å is small. Hence, in agreement with a recent study,[25] a $\sigma$ value of 0.7 Å was chosen for subsequent calculations. Considering that this $\sigma$ value decreases a particular interaction across a distance of 1 Å to 21% of its original value, it becomes obvious that particularly local interactions are taken into account.

**Table 5.** Docking Results for 52 Ligands Using Objective Functions by AFMoC, AFMoC$^{obj}$, DrugScore, and AutoDock

| | % of complexes with poses exhibiting rmsd of the reference structure | | |
|---|---|---|---|
| method | <1.0 Å[a] | <1.5 Å[a] | <2.0 Å[a] |
| AFMoC [b] | 37.5 (43.8) | 52.1 (66.7) | 60.4 (79.1) |
| AFMoC$^{obj}$ [c] | 40.4 (69.2) | 55.7 (94.2) | 76.9 (100.0) |
| DrugScore | 38.5 (65.6) | 55.7 (94.2) | 67.3 (100.0) |
| AutoDock | 34.6 (53.9) | 48.1 (75.0) | 67.3 (92.3) |

[a] The numbers represent the percentage of cases for which a pose is found on the first scoring rank. Values in parentheses are the percentage of cases for which a pose is found irrespective of the scoring rank. [b] Unmodified AFMoC approach as described in ref 25. [c] Modified AFMoC approach described in this study.

**Modifying AFMoC Fields for Binding Mode Predictions. a. Column Filtering by Shannon Entropy.** Initial docking runs using the protein-specifically adapted potential fields derived from the above AFMoC model yielded less accurate results (proportion of poses found on the first scoring rank with rmsd less than 2.0 Å: 60.4%) than using the unmodified DrugScore potential fields (67.3%) (Table 5). Visual inspection of these potential fields showed particularly that regions in the binding site that were occupied by similar substructures of many ligands (which may be indicative of key interactions occurring with the protein) contributed little to these potential fields (Figure 6a). This can be explained by the fact that PLS analyses relate differences in interaction fields to differences in binding affinities. Thus, for regions in the binding pocket that do not show a significant variance in the interaction field values across all ligands, regression coefficients in the QSAR equation and, hence, the adapted potential fields used for docking become small.

Initially, we sought to overcome this effect by increasing the "minimum $\sigma$" threshold value (in CoMFA-type applications usually applied to remove columns with little variance in the field values), which can then be replaced with original DrugScore values. However, applying an equal $\sigma$ value to all field types resulted in nonsignificant PLS models (data not shown) presumably because the "minimum $\sigma$" is an *absolute* measure of the information content within the columns, and our fields have not been scaled. Instead, we chose to filter columns according to their Shannon entropy (eq 1), which provides a *relative* measure of the information content for each column. For columns that were thus excluded, original DrugScore values were used in the prediction step, assuming that this information is more valuable in docking than close-to-zero field values obtained otherwise (Figure 6b).

Figure 7 shows the dependence of binding mode prediction accuracy on the Shannon entropy threshold values. Although the prediction accuracy with respect to binding modes deviating <2.0 Å from the reference structure varies by approximately 10% over the range of threshold values tested, we note that both curves for training and test set compounds show very similar characteristics. Thus, the optimal Shannon entropy threshold value determined on the training set should also be applicable in the prediction step, which is important for "real-life" scenarios. When a prediction accuracy with respect to binding modes deviating <1.0 Å from the reference structure is used as a more stringent criterion, a maximum is found at a Shannon entropy threshold of 0.325 (data not shown). Thus, this value was chosen in the comparison of docking results (see below).

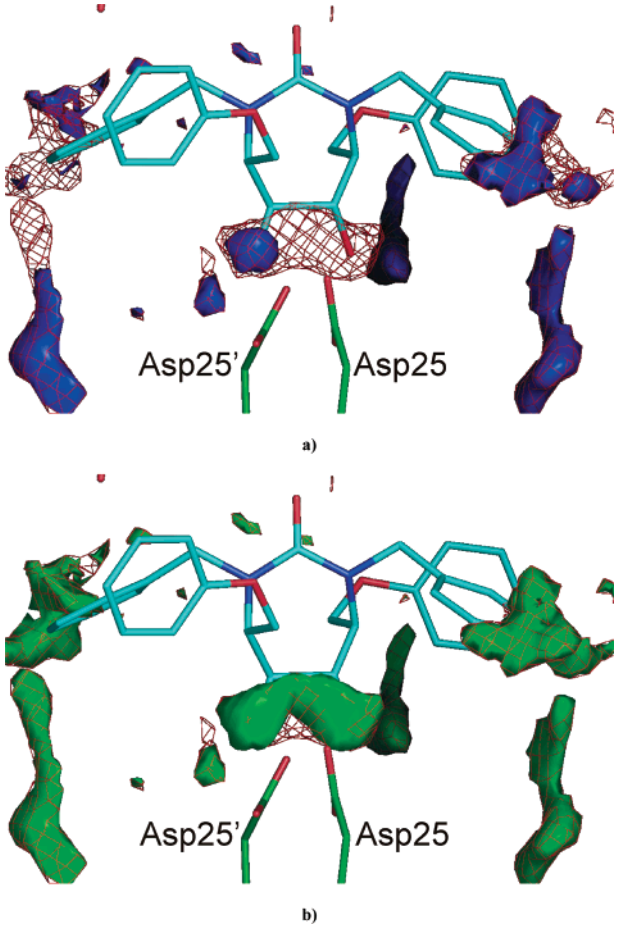



**Figure 6.** Isocontour surfaces (contour level: −0.5) of adapted hydroxyl oxygen (atom type O.3) potential fields calculated with (a) "minimum $\sigma$" (blue) and (b) Shannon entropy (green) threshold during AFMoC analysis, respectively. Isocontour surfaces (contour level: −0.5) of original DrugScore potential fields are shown in mesh representation in both cases. Note that no potential field contributions in the vicinity of one of the hydroxyl groups of the ligand of PDB structure 1ajx (cyan) can be observed at the given contour level in case a; thus, key interactions (indicated by the similar substitution of ligands in this binding pocket region) are neglected. Filtering these regions prior to AFMoC analysis and using original DrugScore values instead alleviates this problem.

**b. Capping Repulsive Regions in AFMoC Fields.** While some regions of adapted potential fields favor ligand binding more than the original DrugScore potential fields, in some cases, previously favorable binding site regions may even become repulsive after adaptation. This is exemplified for adapted fields of aromatic carbon in Figure 8. Although this information is important in the prediction of binding affinities for already superimposed ligands, penalizing ligand substructure placement that way turned out to prevent generation of "well-docked" binding modes in preliminary docking runs. Hence, AFMoC-based potential fields were postprocessed by replacing repulsive regions, which were favorable in the original DrugScore fields, with original DrugScore values. Alternatively, setting these field values to zero resulted in very similar binding mode prediction accuracies.

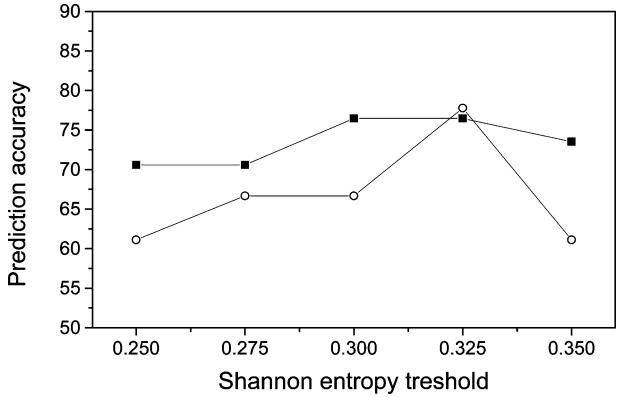



**Figure 7.** Dependence of binding mode prediction accuracy on the Shannon entropy threshold value. Prediction accuracy is measured in terms of the proportion of ligand molecules for which a binding mode below 2 Å rmsd is found on the first scoring rank. Results are shown for both the training set (■) and test set (○) compounds.

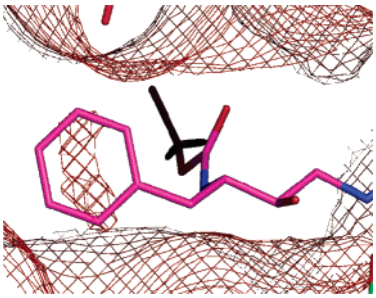


**Figure 8.** Isocontour surfaces of adapted aromatic carbon potential fields (contour level: +0.9) are shown in the binding pocket of HIV-1 protease together with the ligand of PDB structure 1b6m. Note the repulsive field contributions at the location of the phenyl moiety of the ligand.

**c. Linear Mixing of AFMoC and DrugScore Fields.** Visual inspection of pure AFMoC$^{obj}$ fields revealed deep minima in some regions of the binding pocket. In contrast, DrugScore fields are more widespread in the binding site region. Therefore, linear mixing of AFMoC$^{obj}$ fields with DrugScore fields allows varying the amount of specificity and generality contained in the field values. In terms of an objective function for docking, adding contributions by DrugScore fields leads to a smoother potential surface compared to one that originates from AFMoC$^{obj}$ fields alone. That way, the objective function better tolerates imperfectly placed ligand poses, which has been shown to be advantageous for docking.[31,32]

To test the influence of mixing fields on the outcome of the docking runs, mixing parameters $\theta$ of 0.0, 0.25, 0.5, 0.75, and 1.0 have been used, with $\theta = 0.0$ representing pure DrugScore and $\theta = 1.0$ representing pure AFMoC$^{obj}$ fields. Considering the prediction accuracy with respect to binding modes deviating <2.0 Å from the reference structure, Figure 9 reveals that using both pure DrugScore and pure AFMoC$^{obj}$ fields results in inferior prediction accuracy compared to the mixed fields (e.g., in the case of the training set, 65−70% versus a maximum of 80% prediction accuracy). Encouragingly, for $\theta$ values ranging from 0.25 to 0.75, the prediction accuracy changes by less than 5%, which holds for both the training and test sets. Hence, it can be expected that a mixing parameter, which has been optimized for the given protein−ligand system on the training set alone, will also give satisfying results when applied in the prediction step. Using as a more stringent criterion a prediction accuracy with respect to binding modes deviating <1.0 Å from the reference structure results in a maximum at $\theta = 0.25$ (data not shown). Hence, this value was chosen in the comparison of docking results (see below).

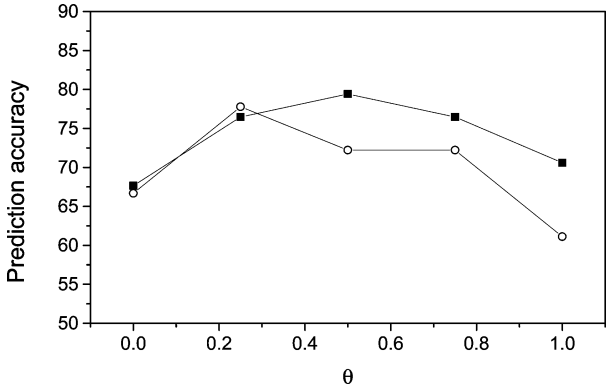



**Figure 9.** Dependence of prediction accuracy on the mixing parameter $\theta$ ($\theta = 0.0$, pure DrugScore fields; $\theta = 1.0$, pure AFMoC$^{obj}$ fields). Prediction accuracy is measured in terms of the proportion of ligand molecules for which a binding mode below 2 Å rmsd is found on the first scoring rank. Results are shown for both the training set (■) and test set (○) compounds.

**Comparison of Docking Results.** In this section, we compare results obtained with the modified AFMoC fields (AFMoC$^{obj}$, using a Shannon entropy threshold value of 0.325, a mixing factor $\theta$ of 0.25, and filtering of repulsive columns) to dockings using original DrugScore fields or the objective function implemented in AutoDock.[62]

Irrespective of their scoring rank, well-docked ligand poses (i.e., those with rmsd less than 2.0 Å) are generated in more than 92% of the cases by all three approaches (Table 5), which indicates an appropriate choice of docking parameters as judged by the overall convergence of the dockings. However, analyzing the size distribution of clusters of similar binding modes (Figure 10) reveals that the AutoDock function produces smaller clusters containing the best-ranked binding mode compared to the objective functions based on DrugScore or AFMoC$^{obj}$ fields. As such, clusters consisting of >70 similar binding modes are found in 10 cases for DrugScore and only 4 cases for AutoDock. Hence, the energy landscape provided by DrugScore appears to be more funnel-shaped compared to that of AutoDock, which favors convergence of independent docking runs to the same final result. This has also been described before.[31,32] Encouragingly, mixing adapted potential fields with DrugScore potentials retains this smoothness, as indicated by 13 clusters consisting of >70 similar binding modes in the case of the AFMoC$^{obj}$-based objective function.

When pure AutoDock or DrugScore fields were used as objective functions, well-docked ligand poses were recognized on the first scoring rank in 67.3% of all cases (Table 5, Figure 11). When AFMoC$^{obj}$ fields were applied instead, well-docked configurations were recognized in 76.9% of all cases on the first rank, which amounts to an improvement of 14% with respect to the other two approaches. Convincingly, very similar results are

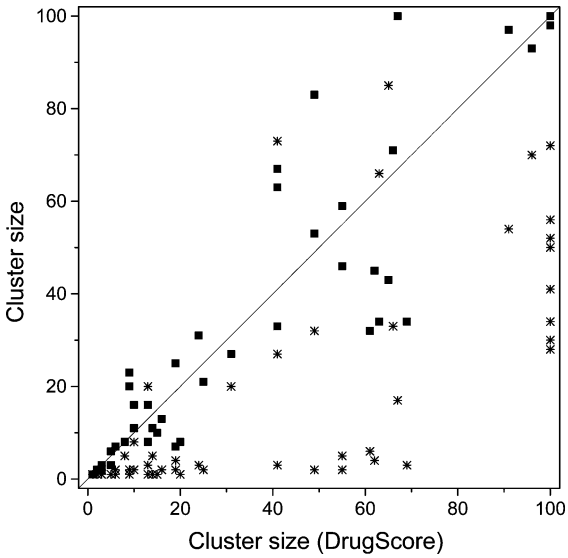



**Figure 10.** Correlation of cluster sizes obtained by AFMoC$^{obj}$- (■) and AutoDock-based (∗) docking with respect to DrugScore-based docking for 52 ligands. In all cases, clusters were chosen that contained the docking solution found on the first scoring rank. While AFMoC$^{obj}$ and DrugScore yield very similar cluster sizes as indicated by the proximity of the data points to the line of ideal correlation, AutoDock-based docking yields clusters of considerably smaller sizes.

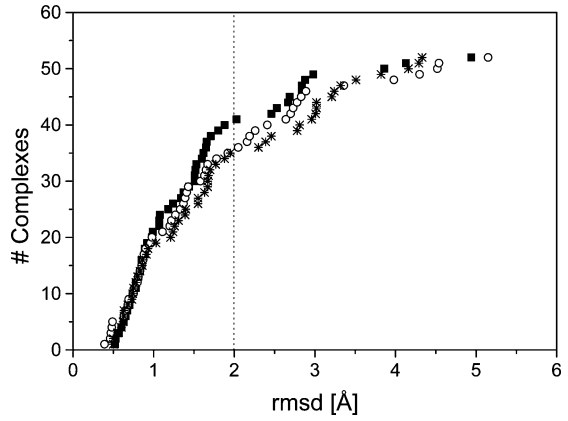



**Figure 11.** Accumulated number of complexes as a function of the rmsd value from the minimized reference structure found for ligand poses on rank 1, using objective functions by AFMoC$^{obj}$ (■), DrugScore (○), and AutoDock (∗).

obtained if only compounds from the training set (76.5%) or the test set (77.8%) are considered. This indicates that the information contained in the AFMoC$^{obj}$ fields is sufficiently general to allow for good docking results, even for compounds that have not been included in the training set. With respect to ligand poses below 1.5 Å, both AFMoC$^{obj}$- and DrugScore-based docking show superior prediction accuracies when compared to using the AutoDock objective function, although the improvement of AFMoC$^{obj}$ over DrugScore in these cases is small.

Figures 12 and 13 exemplify the improvement of ligand placement by using AFMoC$^{obj}$-based docking compared to DrugScore-based docking. In Figure 12, the best-ranked ligand poses of PDB entry 1d4s are depicted for which rmsd values of 4.5 Å (Figure 12a, DrugScore) and 0.8 Å (Figure 12b, AFMoC$^{obj}$) were found. In addition, isocontour plots of potential fields of aliphatic carbon (atom type C.3) are shown (blue, DrugScore-based; red, AFMoC$^{obj}$-based). These isopleths highlight the most pronounced differences in the center region of the figures. In particular, potential fields adapted by AFMoC$^{obj}$ now extend into the S1 and S1′ subsites of the protein. This favors the placement of the C-3α ethyl and phenethyl moieties of the ligand at those sites and, hence, the identification of a pose that is closer to the reference.

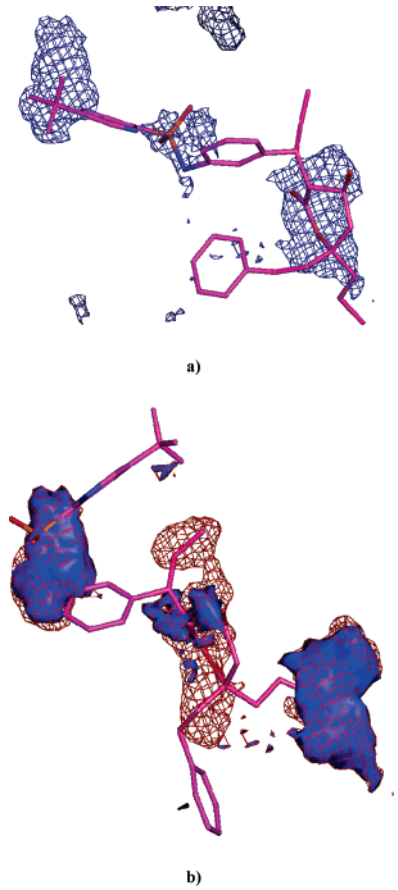



**Figure 12.** Improvement of ligand placement by using AFMoC$^{obj}$-based docking compared to DrugScore-based docking. Best-ranked ligand poses of PDB entry 1d4s are shown, for which rmsd values of 4.5 Å ((a) DrugScore) and 0.8 Å ((b) AFMoC$^{obj}$) were found. In addition, isocontour plots of potential fields for the aliphatic carbon C.3 are depicted (blue, DrugScore-based; red, AFMoC$^{obj}$-based; contour level, −0.7).

In Figure 13, isocontour plots of potential fields of hydroxyl oxygen (atom type O.3) are shown (blue, DrugScore-based; red, AFMoC$^{obj}$-based) together with best-ranked ligand poses of PDB entry 3upj. With respect to the fields of aliphatic carbon in Figure 12, the hydroxyl oxygen fields show a considerably higher spatial restriction. The fact that very similar contour levels were applied in both cases thus indicates the more pronounced orientational dependence of polar interactions compared to nonpolar interactions. Conversely, when the DrugScore potential fields for hydroxyl oxygen are compared to those adapted by AFMoC$^{obj}$, the latter are more expanded in the binding pocket. Thus, the placement of the hydroxyl group of the 4-hydroxycoumarin moiety of the ligand in the vicinity of Asp25 is more favored in the case of AFMoC$^{obj}$-based fields. Although differences in fields of other atom types may also play a role, these findings provide an explanation for the more accurate docking result (rmsd = 1.37 Å)

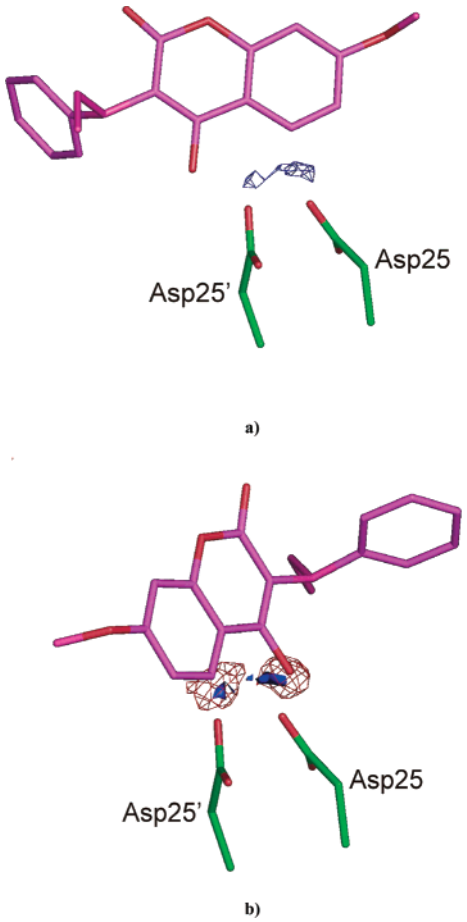



**Figure 13.** Improvement of ligand placement by using AFMoC$^{obj}$-based docking compared to DrugScore-based docking. Best-ranked ligand poses of PDB entry 3upj are shown, for which rmsd values of 2.2 Å ((a) DrugScore) and 1.4 Å ((b) AFMoC$^{obj}$) were found. In addition, isocontour plots of potential fields for the hydroxyl oxygen O.3 are depicted (blue, DrugScore-based; red, AFMoC$^{obj}$-based; contour level, −0.8).

obtained with AFMoC$^{obj}$-based fields compared to using fields based on DrugScore (rmsd = 2.16 Å).

**Assessment of the Approach and Implications for Drug Design.** In a recent study, ligand flexibility has been found to have a greater effect on predicting structures correctly than size or polarity of the compounds.[63] With this in mind, the data set of HIV-1 protease inhibitors provides a challenging test for every docking algorithm because of the high flexibility of the molecules,[33] even if only ligands with <20 rotatable bonds have been used here. This holds true even more when one takes into account that we did not just dock ligands back into their complex protein structure but rather used a single representative protein structure for all ligands. The "consensus-docking" type of approach is more closely related to real-life scenarios but usually leads to lower prediction accuracies compared to "self-docking" because small changes in the receptor conformation can have a large impact on the docking results.[32,63,64] This observation was recently confirmed in a docking study of 21 HIV-1 protease inhibitors with AutoDock. The authors report a prediction accuracy of 62% for cross-docking, which was about 30% lower than in the case of self-docking.[65] Similarly, docking 36 HIV-1 protease inhibitors to a representative protein structure with CDOCKER resulted in a prediction accuracy of 35%, whereas self-docking was successful in 50% of the cases.[63] Another study using the similarity-assisted docking approach SDOCKER showed that for 37 HIV-1 protease inhibitors docked to a single consensus structure only in 30% of the cases a good docking solution was found.[21] Finally, we note that self-docking studies of HIV-1 protease inhibitors have also been reported for other docking programs;[66−72] however, the limited number of complexes used in each case precludes a statistical evaluation of the results. In view of those findings, recognizing in more than 75% of all cases a well-docked solution below 2.0 Å on the first scoring rank using AFMoC$^{obj}$-based docking irrespective of whether training or test set compounds are considered clearly demonstrates the strength and robustness of our approach.

Furthermore, we were pleased to see that the prediction accuracy was particularly increased for peptidomimetic inhibitors, even though they were docked into the protein structure of a complex with a nonpeptidomimetic ligand. This is in contrast to a recent study where it has been found that docking peptidomimetic inhibitors into a cyclic urea inhibitor protein structure was significantly more difficult than the other way around.[70] One may thus anticipate that the protein-specific tailoring of an objective function by considering structural and energetic information about known ligands also accounts implicitly for effects due to protein flexibility. Likewise, no explicit water molecules were considered during the docking. Yet, as mentioned above, improved binding modes are particularly found for peptidomimetic inhibitors, albeit a structural water molecule bound to the protease flaps is known to be important for their binding.[46] Hence, adapted potentials apparently also incorporate implicit information about multiple solvation schemes. While advantageous in our case, it is noted that incorporating different solvation and conformational states in a single objective function may occasionally lead to artifacts such that ligands can simultaneously "experience" mutually exclusive combinations of features from different states.[65]

In this study, we improved binding mode predictions by docking into AFMoC$^{obj}$-based potential fields, which is motivated by the finding that for accurately predicting binding affinities it is a prerequisite to generate and recognize "correct" binding modes.[6,7] In a recent study, we already have demonstrated that the prediction of binding affinities using AFMoC indeed improves significantly if the quality of the docking solutions is increased.[25] Hence, we recommend the following guidelines for ranking different compounds: First, with an AFMoC$^{obj}$-based objective function as described herein, ligand binding modes are generated by docking. As mentioned above, fields based on the original AFMoC approach needed to be modified to be applicable for binding mode prediction (Table 5). Then, in a second step, binding affinities are predicted for ligand geometries found on the first scoring rank. For this, applying unmodified AFMoC fields ($r^2$ = 0.38) outperforms the use of the AFMoC$^{obj}$ objective function ($r^2$ = 0.26) described in this study (Table 6). Mixing unmodified DrugScore fields with the adapted ones results only in marginal improvements in both cases. In turn, both AFMoC approaches show a significantly higher predictive power than the unmodified DrugScore function ($r^2$

**Table 6.** Prediction of Binding Affinities Using Docked Ligand Geometries for the Training Set Compounds[a]

| method | $r^2$ [b] |
|---|---|
| AFMoC [c] | 0.38 |
| AFMoC$^{obj}$ [d] | 0.26 |
| DrugScore | 0.13 |

[a] Binding affinities were predicted for ligand geometries found on the first scoring rank. [b] Squared correlation coefficient of the predicted and experimentally determined binding affinities. [c] Unmodified AFMoC approach as described in ref 25. [d] Modified AFMoC approach described in this study.

= 0.13). We note, however, that the increased predictive power in the latter case may in part be attributed to the fact that the compounds of the training set were used (although for all compounds docked binding geometries that deviate by at least 0.5 Å from the ones used for training are applied in the affinity prediction, lowering the resemblance to the training conditions). In any case, by also taking into account previously reported results for predicting binding affinities of docked geometries,[25] we recommend the application of the unmodified AFMoC approach in this second step.

Recently, a correlation between the capability of a docking tool to generate well-docked ligand geometries and its ability to discriminate known inhibitors from "random" compounds has also been reported.[26,27] In this respect, we expect that obtaining improved binding modes by our approach will eventually lead to better enrichments in virtual screening. Equally important, using an AFMoC$^{obj}$-based objective function comes at no additional cost in the docking step because grid-based potentials need to be evaluated just as in the nonadapted case. In contrast, computational costs in related approaches[18,21] arise because of the additional evaluation of similarity measures between ligand and template molecules.

We have validated our approach only on a single test system for which a considerable amount of experimental structural and energetic information is available. Nevertheless, we are convinced that for the following reasons our approach can be extended to other targets: (I) the HIV-1 protease target is a representative example for proteases, which are among the major targets to which structure-based drug design methods have been applied; (II) the AFMoC approach has been successfully applied in cases for which only little structural information was available such that most of the superimposed ligands were modeled inside the binding pocket;[25] (III) the AFMoC approach yields significantly improved scoring functions even for as few as 15 compounds used for training,[25] allowing its application already in an early phase of drug discovery.

## Conclusion

In the present study, we have developed an approach that exploits structural and energetic information about known protein−ligand complexes to tailor knowledge-based pair potentials toward a protein-specifically adapted objective function to improve binding mode predictions in docking. Our strategy thus combines the principles of molecular recognition (applied in the field of docking) and molecular similarity (applied in the field of 3D QSAR). To achieve this, DrugScore-based potential fields are adapted by a "reverse", protein-based CoMFA-type (=AFMoC) approach, which allows implicit weighting of structural ligand information with respect to the location in the binding site. Compared to using AFMoC for binding affinity predictions, a Shannon entropy-based column filtering of the descriptor matrix and capping of repulsive adapted potentials within the binding site have turned out to be crucial for the success of this method. The resulting modified AFMoC potentials are referred to as AFMoC$^{obj}$.

We validated our approach on a data set of HIV-1 protease inhibitors for which crystallographically determined complex structures were available. HIV-1 protease inhibitors have been repeatedly used in docking studies and represent a challenging system because of the size and flexibility of the ligands as well as the occurrence of large nonpolar regions in the binding pocket. On the basis of leave-one-out and leave-five-out cross-validations, a significant AFMoC model was obtained for a training set of 48 ligands. Noteworthy, when the thus generated adapted potential fields were used as the objective function in AutoDock, in more than 75% of all cases a binding mode below 2 Å rmsd was identified on the first scoring rank for ligands with up to 20 rotatable bonds. With respect to using nonadapted DrugScore or AutoDock fields, AFMoC$^{obj}$-based docking significantly improved binding mode predictions by 14%. Convincingly, very similar prediction accuracies were obtained with training and test set compounds, demonstrating the strength and robustness of the method. Furthermore, our approach allows for gradually moving from generally valid to protein-specifically adapted potentials and, hence, reflecting the amount and degree of diversity available in the training set. Finally, the method is as fast as docking into nonadapted potential fields because no additional CPU cycles are required in this step.

Considering the success of our approach, we expect it to provide a valuable tool for similarity-driven correct binding mode identification, which is a prerequisite for accurate binding affinity prediction and successful virtual screening.

**Acknowledgment.** We are grateful to Katrin Silber (Philipps-University, Marburg) and Domingo Gonzalez Ruiz (J. W. Goethe University, Frankfurt) for critical reading of the manuscript.

**Supporting Information Available:** Two tables with HIV-1 protease inhibitors used as training and test sets. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Drews, J. Drug discovery: A historical perspective. *Science* **2000**, *287*, 1960−1964.
(2) Lyne, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.
(3) Hirst, J. D. Predicting ligand binding energies. *Curr. Opin. Drug Discovery Dev.* **1998**, *1*, 28−33.
(4) 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.
(5) Sotriffer, C.; Stahl, M.; Boehm, H. J.; Klebe, G. Docking and Scoring Functions/Virtual Screening. *Burger's Medicinal Chemistry and Drug Discovery*; Wiley: New York, 2003; pp 281−333.
(6) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.

(7) Kontoyanni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithm. *J. Med. Chem.* **2004**, *47*, 558−565.

(8) Schneider, G.; Neidhardt, W.; Giller, T.; Schmid, G. "Scaffold-hopping" by topological pharmacophore search: A contribution to virtual screening. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894−2896.

(9) Norinder, U. Recent Progress in CoMFA Methodology and Related Techniques. *3D QSAR in Drug Design*; Kluwer Academic Publisher: Dordrecht, The Netherlands, 1998; pp 25−39.

(10) Lemmen, C.; Lengauer, T. Computational methods for the structural alignment of molecules. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 215−232.

(11) Dean, P. M. *Molecular Similarity in Drug Design*; Chapman-Hall: London, 1995.

(12) Fradera, X.; Mestres, J. Guided docking approaches to structure-based design and screening. *Curr. Top. Med. Chem.* **2004**, *4*, 687−700.

(13) Gruneberg, S.; Stubbs, M. T.; Klebe, G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J. Med. Chem.* **2002**, *45*, 3588−3602.

(14) Schafferhans, A.; Klebe, G. Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **2001**, *307*, 407−427.

(15) Evers, A.; Gohlke, H.; Klebe, G. Ligand-supported homology modelling of protein binding sites using knowledge-based potentials. *J. Mol. Biol.* **2003**, *334*, 327−345.

(16) Verdonk, M. L.; Cole, J. C.; Taylor, R. SuperStar: A knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.* **1999**, *289*, 1093−1108.

(17) Gohlke, H.; Hendlich, M.; Klebe, G. Predicting binding modes, binding affinities and "hot spots" for protein−ligand complexes using a knowledge-based scoring function. *Perspect. Drug Discovery Des.* **2000**, *20*, 115−144.

(18) Fradera, X.; Knegtel, R. M.; Mestres, J. Similarity-driven flexible ligand docking. *Proteins* **2000**, *40*, 623−636.

(19) Hindle, S. A.; Rarey, M.; Buning, C.; Lengaue, T. Flexible docking under pharmacophore type constraints. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 129−149.

(20) Daeyaert, F.; de Jonge, M.; Heeres, J.; Koymans, L.; Lewi, P.; et al. A pharmacophore docking algorithm and its application to the cross-docking of 18 HIV-NNRTI's in their binding pockets. *Proteins* **2004**, *54*, 526−533.

(21) Wu, G.; Vieth, M. SDOCKER: a method utilizing existing X-ray structures to improve docking accuracy. *J. Med. Chem.* **2004**, *47*, 3142−3148.

(22) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M.; Vacca, J. P.; et al. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* **1995**, *38*, 305−317.

(23) 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.

(24) 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.

(25) 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.

(26) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57*, 225−242.

(27) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; et al. Virtual screening using protein−ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.

(28) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.

(29) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−356.

(30) 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.

(31) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.

(32) 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.

(33) 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.

(34) Rose, J. R.; Salto, R.; Craik, C. S. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino-acid substitutions. *J. Biol. Chem.* **1993**, *268*, 11939−11945.

(35) Mildner, A. M.; Rothrock, D. J.; Leone, J. W.; Bannow, C. A.; Lull, J. M.; et al. The HIV-1 protease as enzyme and substrate. Mutagenesis of autolysis sites and generation of a stable mutant with retained kinetic-properties. *Biochemistry* **1994**, *33*, 9405−9413.

(36) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; et al. Conserved folding in retroviral proteases. Crystal-structure of a synthetic HIV-1 protease. *Science* **1989**, *245*, 616−621.

(37) Davis, D. A.; Dorsey, K.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J.; et al. Regulation of HIV-1 protease activity through cysteine modification. *Biochemistry* **1996**, *35*, 2482−2488.

(38) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: design and development of a database for comprehensive analysis of protein−ligand interactions. *J. Mol. Biol.* **2003**, *326*, 607−620.

(39) 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.

(40) Backbro, K.; Lowgren, S.; Osterlund, K.; Atepo, J.; Unge, T.; et al. Unexpected binding mode of a cyclic sulfamide HIV-1 protease inhibitor. *J. Med. Chem.* **1997**, *40*, 898−902.

(41) 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.

(42) Wang, Y. X.; Freedberg, D. I.; Yamazaki, T.; Wingfield, P. T.; Stahl, S. J.; et al. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* **1996**, *35*, 9945−9950.

(43) Nam, K. Y.; Chang, B. H.; Han, C. K.; Ahn, S. G.; No, K. T. Investigation of the protonated state of HIV-1 protease active site. *Bull. Korean Chem. Soc.* **2003**, *24*, 817−823.

(44) Chen, X. N.; Tropsha, A. Relative binding free-energies of peptide inhibitors of HIV-1 protease. The influence of the active-site protonation state. *J. Med. Chem.* **1995**, *38*, 42−48.

(45) Weiner, S.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230−252.

(46) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249−284.

(47) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; et al. Structure at 2.5-Å resolution of chemically synthesized human-immunodeficiency-virus type-1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* **1991**, *30*, 1600−1609.

(48) Schaal, W.; Karlsson, A.; Ahlsen, G.; Lindberg, J.; Andersson, H. O.; et al. Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors. *J. Med. Chem.* **2001**, *44*, 155−169.

(49) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity. A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.

(50) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III. The collinearity problem in linear regression. The partial least squares approach to generalized inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735−743.

(51) Wold, S.; Johansson, E.; Cocchi, M. PLS−Partial Least Squares Projections to Latent Structures. *3D QSAR in Drug Design. Theory, Methods and Applications*; ESCOM: Leiden, The Netherlands, 1993.

(52) 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.

(53) 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.

(54) Kubinyi, H.; Abraham, U. Practical Problems in PLS Analyses. *3D QSAR in Drug Design. Theory, Methods and Applications*; ESCOM: Leiden, The Netherlands, 1993; pp 717−728.

(55) Shannon, C. E.; Weaver, W. W. *Mathematical Theory of Communication*; University of Illinois Press: Chicago, IL, 1949.

(56) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.

(57) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; et al. Recommendations to CoMFA Studies and 3D QSAR Publications. *3D QSAR in Drug Design. Theory, Methods and Applications*; ESCOM: Leiden, The Netherlands, 1993; pp 711−716.

(58) Kim, K. Non-linear dependencies in CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 71−82.

(59) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and Limitations. *3D QSAR in Drug Design*; ESCOM: Leiden, The Netherlands, 1993; pp 583−618.
(60) Golbraikh, A.; Tropsha, A. Beware of $q^2$! *J. Mol. Graphics Modell.* **2002**, *20*, 269−276.
(61) Cramer, R. D., III; DePriest, S. A.; Patterson, D. E.; Hecht, P. The Developing Practice of Comparative Molecular Field Analysis. *3D QSAR in Drug Design. Theory, Methods and Applications*; ESCOM: Leiden, The Netherlands, 1993.
(62) Morris, G. M.; Goosell, D. S.; Huey, R.; Hart, W. E.; Belew, R. K.; et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1693−1662.
(63) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45−55.
(64) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547−562.
(65) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.
(66) Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.
(67) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(68) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.
(69) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FlexX incremental construction algorithm for protein−ligand docking. *Proteins* **1999**, *37*, 145−156.
(70) Yang, J. M.; Chen, C. C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* **2004**, *55*, 288−304.
(71) Budin, N.; Majeux, N.; Caflish, A. Fragment-based flexible ligand docking by evolutionary optimization. *Biol. Chem.* **2001**, *382*, 1365−1372.
(72) Cecchini, M.; Kolb, P.; Majeux, N.; Caflish, A. Automated docking of highly flexible ligands by genetic algorithms: A critical assessment. *J. Comput. Chem.* **2003**, *25*, 412−422.
(73) *SYBYL Molecular Modeling Software*, version 6.6; Tripos Inc.: St. Louis, MO.

JM050114R