

3D-QSAR Based on Quantum-Chemical Molecular Fields: Toward an Improved Description of Halogen Interactions

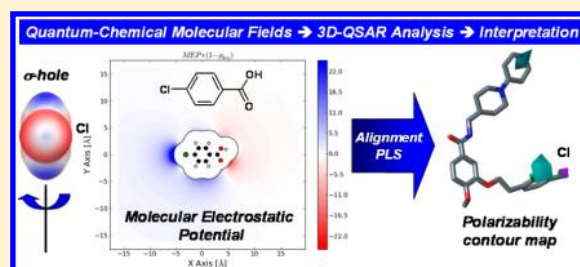
Stefan Güssregen,^{*,†} Hans Matter,[†] Gerhard Hessler,[†] Marco Müller,[†] Friedemann Schmidt,[†] and Timothy Clark^{*,‡,§}

[†]Sanofi-Aventis Deutschland GmbH, R&D, LGCR, Structure, Design and Informatics, Building G 878, 65926, Frankfurt am Main, Germany

[‡]Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nögelsbachstraße 25, 91052 Erlangen, Germany

[§]Centre for Molecular Design, University of Portsmouth, King Henry Building, Portsmouth PO1 2DY, United Kingdom

ABSTRACT: Current 3D-QSAR methods such as CoMFA or CoMSIA make use of classical force-field approaches for calculating molecular fields. Thus, they can not adequately account for noncovalent interactions involving halogen atoms like halogen bonds or halogen- π interactions. These deficiencies in the underlying force fields result from the lack of treatment of the anisotropy of the electron density distribution of those atoms, known as the “ σ -hole”, although recent developments have begun to take specific interactions such as halogen bonding into account. We have now replaced classical force field derived molecular fields by local properties such as the local ionization energy, local electron affinity, or local polarizability, calculated using quantum-mechanical (QM) techniques that do not suffer from the above limitation for 3D-QSAR. We first investigate the characteristics of QM-based local property fields to show that they are suitable for statistical analyses after suitable pretreatment. We then analyze these property fields with partial least-squares (PLS) regression to predict biological affinities of two data sets comprising factor Xa and GABA-A/benzodiazepine receptor ligands. While the resulting models perform equally well or even slightly better in terms of consistency and predictivity than the classical CoMFA fields, the most important aspect of these augmented field-types is that the chemical interpretation of resulting QM-based property field models reveals unique SAR trends driven by electrostatic and polarizability effects, which cannot be extracted directly from CoMFA electrostatic maps. Within the factor Xa set, the interaction of chlorine and bromine atoms with a tyrosine side chain in the protease S1 pocket are correctly predicted. Within the GABA-A/benzodiazepine ligand data set, PLS models of high predictivity resulted for our QM-based property fields, providing novel insights into key features of the SAR for two receptor subtypes and cross-receptor selectivity of the ligands. The detailed interpretation of regression models derived using improved QM-derived property fields thus provides a significant advantage by revealing chemically meaningful correlations with biological activity and helps in understanding novel structure–activity relationship features. This will allow such knowledge to be used to design novel molecules on the basis of interactions additional to steric and hydrogen-bonding features.



1. INTRODUCTION

Today, the introduction of halogen atoms into biologically active lead structures is an important concept in medicinal chemistry in order to modulate specific properties, such as free energy of binding, selectivity, and pharmacokinetics.^{1,2} Consequently a high percentage of structures in virtual screening databases for lead generation and congeneric series for structure–activity relationships (SAR) in lead optimization contain halogen atoms. In the past, the influence of halogen substituents on biological activity has rarely been targeted by structure-based rational design but has generally resulted from systematic SAR investigations after synthesis of appropriate derivatives. In contrast, changes in physicochemical properties on halogenation are predicted reasonably well.³ These modified physicochemical properties influence pharmacokinetics and metabolism significantly. In particular, fluorine is often introduced as a bioisosteric hydrogen replacement to prevent

ligand metabolism by cytochrome CYP P450 enzymes.^{1,2} Adding a halogen atom to a lead structure often increases lipophilicity, which can help improve penetration through lipid membranes and alter the tissue distribution. However, this effect must be weighted against the frequent reduction of aqueous solubility and the undesirable tendency to accumulate in lipid tissue.⁴ Therefore, the balance between lipophilicity and a minimal polarity of the molecule is one of the major challenges in rational drug discovery. Additionally, electronegative halogen substituents are also useful for modulating the pK_a values of basic molecules and for increasing electrophilicity.^{1,2}

Chlorine occupies an intermediate position in the series of halogens with respect to its atomic radius and electronic properties, including its moderate halogen bond donor potential.⁵

Received: June 1, 2012

These electronic features can also modulate molecular recognition of electron-rich aromatic rings.⁶ Chlorine substituents are therefore often used to modulate affinity and lipophilicity. Halogen atoms are frequently part of nonbonded protein–ligand interactions, such as halogen (C–Hal...O) and hydrogen bonds (C–Hal...H), which have different but similar geometrical and energetic characteristics.⁷

It is now commonly recognized that heavier halogen atoms in addition to hypervalent elements like sulfur display a pronounced anisotropy of the electron density that has been named the “sigma hole” (σ -hole).^{8–10} In contrast to fluorine, which is characterized by negative potential over its entire atomic surface, chlorine, bromine, and iodine exhibit a positive electrostatic potential collinear with the C–halogen bond axis and a belt of negative potential orthogonal to it, as shown in Figure 1A. This

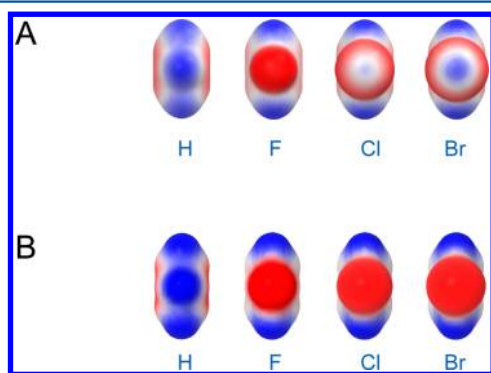


Figure 1. Isocontour plots of benzene (left) or various halobenzenes using B3PW91/cc-pVTZ-PP on the 0.001 au molecular surface (upper panel) or OPLS-AA 2005 force-field derived charges (lower panel). Red indicates negative electrostatic potential; blue refers to positive electrostatic potential.

property of the heavier halogens enables them to form attractive noncovalent interactions such as halogen– π interactions and halogen bonds,^{5,11,12} whose directionality and strength, which is sometimes comparable to classical hydrogen bonding, have recently been exploited in structure-based design of ligands.^{13,14}

However, these important interactions cannot be accounted for in “classical” force fields that are currently in use for small molecule modeling in the pharmaceutical industry, such as the Tripos force field,¹⁵ MMFF94,¹⁶ or OPLS-AA,¹⁷ although specific modifications to account for halogen bonding have very recently been introduced.^{18–20} Whenever a force field employs atom-centered monopoles to represent the molecular electrostatic potential, this σ -hole is ignored²¹ and higher halogens are effectively treated as “large fluorines” (Figure 1B) making a differentiation between halogen atoms impossible.

This has implications for quantitative structure–activity relationships, in particular those that employ field-based 3D-QSAR methods.²² Comparative molecular field analysis (CoMFA)^{23,24} is often used to derive relationships between molecular property fields of aligned compounds and their biological activities. Electrostatic and steric interaction energies are computed between each ligand and probe atoms located on predefined grid points. In this case, the limitations outlined above apply if the molecular electrostatic field is computed based on atom-centered partial charges for individual ligand molecules. Potential interactions involving this σ -hole are not taken into account, which prevents important SAR trends from being recognized.

Quantum mechanics (QM) based methods do not suffer from those limitations, but have not yet found widespread application in 3D-QSAR. A few alternative approaches make use of fields based on conceptual density functional theory (DFT),²⁵ thereby employing electron density, the shape function, the Fukui function, local softness, and the electrostatic potential to generate property fields for 3D-QSAR.^{26,27}

The concept of QM-derived local properties, usually projected onto molecular surfaces, for describing intermolecular interactions has recently been introduced.^{28–30} In addition to the local electron density (ρ) and the molecular electrostatic potential (MEP), which account for the shape and electrostatics of a molecule, further quantum-chemical properties are considered. The concept of using the local ionization energy (IE_L) as a measure of electron-donating capacity, introduced by Sjöberg et al.,³¹ has been extended to use the local electron affinity (EA_L) as the electron acceptor equivalent.^{28,32} The two can be combined to give two further local properties, the local Mulliken electronegativity (χ_L)³³ and the local hardness (η_L).³⁴ Furthermore, the local polarizability (α_L), as defined within NDDO-based semiempirical molecular orbital theory,²⁸ can be used to describe dispersion interactions.²⁹ As these calculations employ semiempirical theory, they can be performed much faster than those using DFT and are much more amenable to an industrial setup. Hence, we were interested in expanding the concept of using local properties from surface to regular lattices and using these as molecular properties fields for 3D-QSAR investigations.

We will first discuss the characteristics of QM-based local property fields and then explore whether they are suitable for statistical analyses and which type of data pretreatment is appropriate. Subsequently we apply the derived method to two reference data sets for inhibitors of the serine protease factor Xa (fXa)³⁵ and GABA-A/benzodiazepine receptors (BzR)³⁶ and compare statistical models and chemical interpretations with results obtained for the original CoMFA method.^{23,24} In particular, we employ the partial least squares (PLS) technique³⁷ to derive a linear relationship between biological activity and molecular property fields, while cross-validation³⁸ serves to check for consistency and predictivity. Finally, we will evaluate whether QM-based property fields allow a consistent chemical interpretation of the 3D-QSAR models, and thereby reveal important SAR trends, possibly modulated by some less-obvious effects. We have therefore selected data sets for which the SAR is dominated by the effects of halogen atoms and electron-withdrawing substituents in modulating biological activity. While the alignment for the fXa data set was guided by the X-ray structure of the protein binding site, the alignment for the BzR data set was obtained by a classical ligand-based approach. This allows us to compare the different alignment strategies.

2. METHODS

2.1. Data Sets. Typically 3D-QSAR methods require alignment of the molecules in the data set. For comparison, we have used the alignment obtained from the authors for both data sets.

For the fXa data set,³⁵ a structure-based alignment resulted from docking 152 3-oxybenzamide ligands into fXa crystal structures using the program QXP³⁹ employing a modified version of the AMBER force field⁴⁰ This was followed by optimization of ligand geometries in the protein–ligand complex using the MMFF94s force field.¹⁶ This workflow

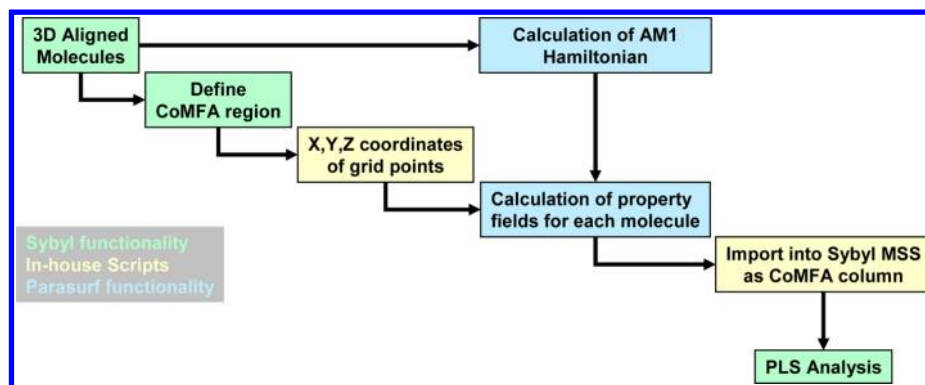


Figure 2. Workflow to calculate ParaSurf quantum-chemical molecular property fields and to apply them to partial least squares (PLS) statistical analysis.

resulted in geometries in accord with six X-ray structures for this chemical series, in particular with regard to relevant halogen atoms. For comparability of methods, we have employed the same alignment as in the original publication.³⁵ Biological data on fXa inhibition were reported as dissociation constants K_i . This data set was divided into a training set of 80 compounds and two test sets of 27 (test set 1) and 72 (test set 2) compounds according to the original publication.³⁵

For the BzR data set,³⁶ a ligand-based alignment was generated for a series of 38 imidazo-1,4-diazepines. Crystal structures of some key ligands served as starting points and were modified accordingly followed by geometry optimization at the MM2⁴¹ and RHF/3-21G⁴² levels while keeping the heterocyclic core fixed. All molecules were used in their neutral form. Biological data were reported as dissociation constants K_i against the “diazepam-insensitive” (DI) and “diazepam-sensitive” (DS) isoforms of the benzodiazepine receptor (BzR).

2.2. Calculation of Molecular Property Fields. The workflow used to calculate molecular property fields for these data sets is shown in Figure 2. A single point calculation using the AM1 Hamiltonian (keywords: AM1 1SCF)⁴³ was performed for each aligned molecule using Mopac⁴⁴ provided by Cepos InSilico Ltd.⁴⁵ This version writes eigenvectors and eigenvalues in a format compatible with all of the following steps. We did not perform semiempirical optimization of the ligands in order to avoid changes to the reference alignment. The ParaSurf10 software⁴⁶ was then used with default settings for calculating fields of different QM-based molecular properties for each molecule at regular grid points similar to CoMFA; those fields will be termed “ParaSurf fields” in this manuscript. The following properties were used: electron density (ρ), molecular electrostatic potential (MEP), local ionization energy (IE_L), local electron affinity (EA_L), local electronegativity (χ_L), local hardness (η_L), and local polarizability (α_L). A grid spacing of 2 Å⁴⁷ was used and the size of the grid was determined such that it extends 4 Å beyond all molecules in each direction. The computed fields were then imported into Sybyl;⁴⁸ further usage was identical to CoMFA fields. The entire process is very fast and requires about 1 s of CPU time on average per druglike molecule on a modern Linux workstation.

Default settings were used for reference CoMFA analysis if not otherwise indicated. Steric and electrostatic energies were calculated at grid points with 2 Å spacing using a positively charged carbon atom as an atomic probe. A distance-dependent dielectric constant was used and molecular atomic charges were taken as obtained from the authors of the original CoMFA studies.

2.3. Statistical Analysis. For all data sets, K_i values were expressed as $-\log(K_i)$ prior to data analysis. Leave-one-out (LOO) cross-validation analyses³⁸ were performed using SAMPLS⁴⁹ up to twelve components. The optimal number of PLS³⁷ components was determined from the first minimum of the standard error obtained for the corresponding leave-one-out PLS models. Equal weights for CoMFA and QM-property fields were assigned using the CoMFA_STD scaling option.⁵⁰ While classical CoMFA columns with a variance smaller than 2.0 were excluded prior to the PLS analysis (*minimum-sigma*), no additional column filtering was used for ParaSurf-fields. The overall quality of PLS analyses was expressed using the cross-validated $r^2(\text{cv})$, defined as

$$r^2(\text{cv}) = (\text{SD} - \text{PRESS})/\text{SD} \quad (1)$$

where SD is the variance of the biological activities around the mean values and PRESS represents the sum of squared differences between predicted and target property values.

3. RESULTS AND DISCUSSION

3.1. Characterization of Local Property Fields. Molecular fields must show certain characteristics to be suitable for statistical analyses.^{26,51} Most importantly, steep fields causing large variances at a small (atomic) scale and fields with singularities should be avoided, as they may prevent SAR trends from being identified. Such unsuitable fields can even compromise the formation of a predictive statistical model altogether. Molecular fields that display variances on a larger scale tend to be more robust with respect to uncertainties in atomic coordinates after alignment.

The Lennard-Jones and Coulomb potentials used in CoMFA exhibit that kind of large fluctuation on a small atomic scale⁵¹ and thus need to be subjected to a suitable pretreatment. CoMFA performs the pretreatment such that the Lennard-Jones potential is arbitrarily truncated at a level of +30.0 kcal·mol⁻¹, and the Coulomb potential is omitted accordingly.²³ Molecular fields with stepless distance dependency can be obtained by using Gaussian-type functional forms as in CoMSIA.⁵¹ Advanced QM-based molecular fields calculated from Fukui functions do not require any pretreatment, as they are computed as finite differences.²⁶ For other fields, such as electron density and electrostatic potential, pretreatments based on multiresolution analysis (MRA) in connection with discrete wavelet transformations (DWT) have been employed.²⁷

To characterize the ParaSurf local property fields, we first performed model calculations using *para*-chlorobenzoic acid in

its neutral form. This simple molecule was placed in the xy -plane of a 0.1 Å grid such that the centroid of the benzene ring located in the origin and the chlorine and the carbon of the carboxyl group are aligned along the x -axis (Figure 3A).

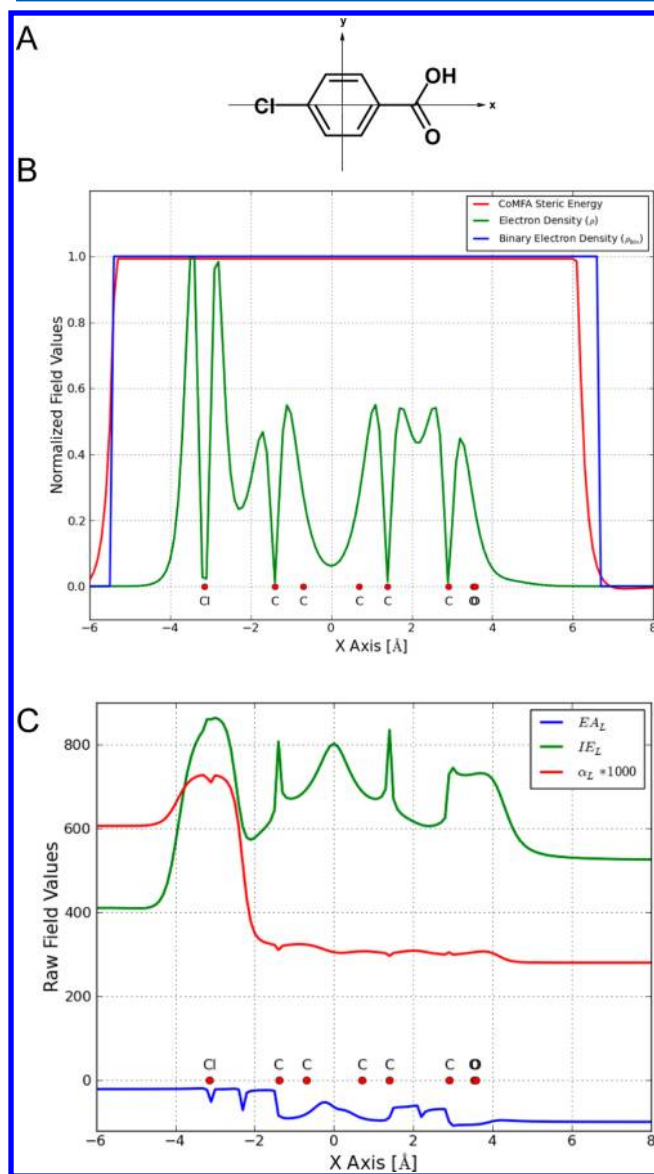


Figure 3. Comparison of different field values along the x -axis ($y = 0$ Å, $z = 0$ Å) for *para*-chlorobenzoic acid. The positions of the atoms as projected onto the x -axis are indicated as red circles in the two lower panels. (A) Orientation of *para*-chlorobenzoic acid on the grid. (B) The normalized field values of CoMFA steric energy are shown in red, the raw electron density (ρ) is shown in green, and the binarized electron density (ρ_{bin}) is shown in blue. (C) The raw field values [kcal·mol^{−1}] of the local electron affinity field (EA_L) is shown in blue and those of the local ionization energy field (IE_L) are in green. The field values of the local polarizability field are given in red multiplied by a factor of 1000.

Figure 3B shows the normalized values of the ParaSurf electron-density field (green) in comparison to those of the CoMFA steric field (red) along the x -axis for $y = 0$ Å and $z = 0$ Å. The electron density field approaches zero outside -5 and $+5$ Å of the x -axis. It rises close to the atomic nuclei and drops to zero at the position of the atoms. Note in this context that NDDO-like semiempirical molecular orbital techniques neglect

core electrons. Thus, the electron-density field exhibits high variance in the vicinity of the atomic nuclei. In contrast, the CoMFA steric field follows a capped Lennard-Jones potential, which is smoother and only exhibits variance on the molecular scale. The values drop continuously until the van-der-Waals minimum is reached and increase between -6 and 7 Å up to the predefined cutoff values of $+30$ kcal·mol^{−1}.

The simple binarization function shown in eq 2 with cutoff values between 10^{-6} and 10^{-4} e[−]·Å^{−3} was found to be a suitable transformation of these property fields, in analogy to CoMFA steric fields, after evaluating different schemes for pretreating the raw electron density fields on both data sets investigated with respect to averaged regression coefficients.

$$\rho_{\text{bin}}(r) = \begin{cases} 0, & \text{if } \rho(r) < 10^{-5}[\text{e}^{-} \cdot \text{\AA}^{-3}] \\ 1, & \text{if } \rho(r) > 10^{-5}[\text{e}^{-} \cdot \text{\AA}^{-3}] \end{cases} \quad (2)$$

A comparison between the CoMFA electrostatic and corresponding ParaSurf MEP field is shown in Figure 4. The white area in the CoMFA electrostatic field (upper panel) close to the nuclei originates from omitted grid field values as the Lennard-Jones potential exceeded the cutoff values at corresponding grid points. Effectively, only the electrostatic potential outside the molecular surface is used in CoMFA. A similar effect is obtained by using the product of ParaSurf MEP with $1 - \rho_{\text{bin}}$, as indicated in Figure 4 (lower panel) using a black isocontour line. Consequently, all MEP field values are set to zero if the electron density exceeds the chosen cutoff value.

A detailed inspection reveals significant differences between the two types of field. In the xy -plane, the CoMFA electrostatic field displays a negative potential near the oxygen and chlorine atoms and a positive potential close to the hydrogen atoms, as expected for a charge model based on atom-centered monopoles. ParaSurf, however, calculates the molecular electrostatic potential in a series expansion up to the quadrupole moment,⁵² thus providing a more accurate view of the molecular electrostatic environment. Here, we found a strong positive electrostatic potential near the chlorine atom predominantly along the C–Cl axis along with a smaller belt in which the electrostatic potential is negative. This reflects the features of the σ -hole as known from quantum mechanical calculations at higher levels (Figure 1). Furthermore, differences in the electrostatic potential near the carbonyl oxygen atom of the carboxyl group are significant. While the CoMFA electrostatic field displays no differences for E - and Z -positions of a potential hydrogen bond donor atom, a more negative electrostatic potential is found in the area toward the other oxygen atom. The electrostatic potential is less negative at the opposite side, reflecting the experimentally proven preference of Z -position in forming hydrogen bonds with carboxyl groups (c.f. the X-ray structure of *para*-chlorobenzoic acid at room temperature⁵³ in the entry CLBZAP02 of the Cambridge Structure Database CSD⁵⁴).

Figure 3C displays raw field values of other local properties local electron affinity (EA_L , blue), local ionization energy (IE_L , green), and local polarizability (α_L , red) calculated along the z -axis of *para*-chlorobenzoic acid for $x = 0$ and $y = 0$. The local electronegativity (χ_L) and local hardness (η_L) fields, which are derived as sum and difference of EA_L and IE_L , respectively,²⁸ are not included in Figure 4C for clarity. In contrast to the electron density and MEP fields, there are no large fluctuations, which allows them to be used unchanged in PLS analyses.

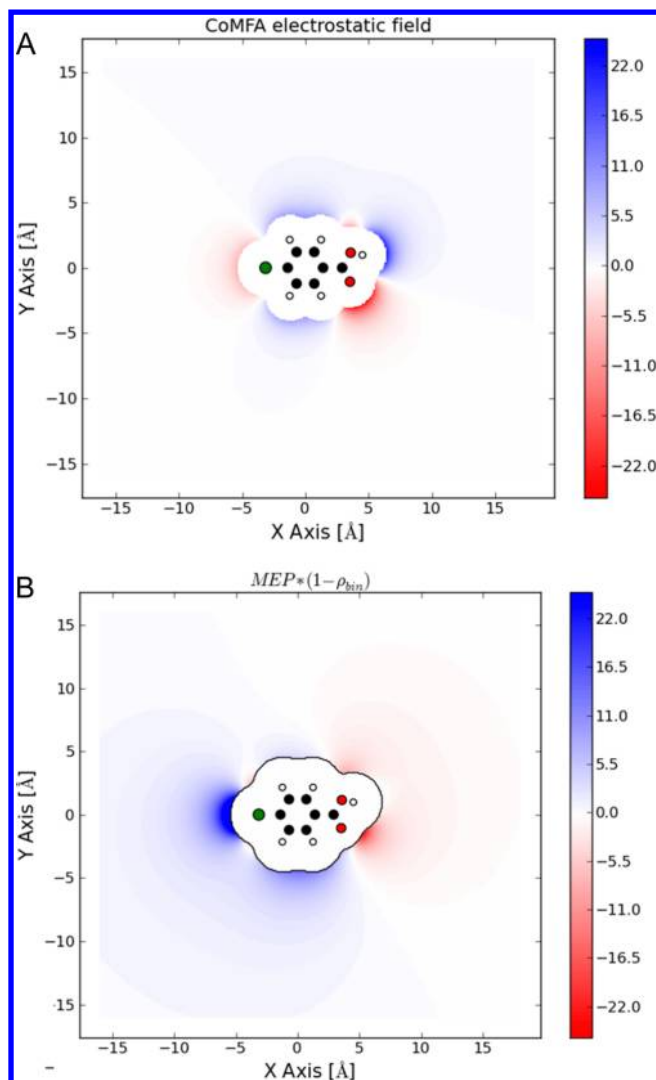


Figure 4. Comparison of CoMFA electrostatic field (upper panel) and ParaSurf local molecular electrostatic potential ($\text{MEP}(1 - \rho_{\text{bin}})$) property field (lower panel) for *para*-chlorobenzoic acid (*xy*-plane).

The IE_L field in Figure 3C (green) is characterized by broad peaks of high ionization energy near the chlorine atom on the left, at the center of the aromatic ring system and around the carboxyl group. Positive spikes appear at the exact location of the atomic nuclei, while minima of ionization energy are located just outside of the aromatic ring system.

The local electron affinity (EA_L) is defined as a density-weighted Koopmans Theorem electron affinity, which in turn is defined as the ionization potential of the reduced species.⁵⁵ Therefore its numerical values are mostly negative, with more positive values indicating high electrophilicity and more negative ones corresponding to low electrophilicity. Hence, areas of low electron affinity are located within the ParaSurf EA_L field in Figure 3C (blue) near the chlorine atom on the left and at the center of the aromatic ring system, while the electron affinity is highest around the oxygen atoms of the carboxyl group on the right. Furthermore, some spikes of high electron affinity are located at the centers of the C–Cl and C–C_{carbonyl} bonds.

The local polarizability (α_L) is governed primarily by shielding of valence electrons from the atomic core by electrons from the inner shells. Thus, a significant difference in the local

polarizability close to third-row atoms as chlorine compared to second-row atoms such as carbon and oxygen is observed, in agreement with this definition. The α_L ParaSurf field in Figure 3C (red) therefore exhibits a single large maximum around chlorine and almost no variance for the rest of the molecule.

3.2. Statistical Performance of Individual Local Property Fields. For a quantitative assessment of the predictive power of local ParaSurf property fields, different statistical measures were employed for each data set. For the fXa data set, the $r^2(\text{cv})$ value and the predictive r^2 values for training and test sets were chosen. Due to their small size ($n = 38$), the BzR data sets could not be split further and were analyzed based on $r^2(\text{cv})$ values. Results obtained with standard CoMFA fields served for benchmarking; they were obtained for models that use steric and electrostatic fields individually or a combination of the two.

The electron density (ρ) and MEP fields constitute the QM-based equivalent to the steric and electrostatic fields in CoMFA. As discussed above, a pretreatment was performed similar to the approach used in CoMFA. No statistically significant models were obtained with raw electron density (ρ) and MEP fields. For all other local molecular property fields, the results obtained with the raw fields are compared with those for which a pretreatment was applied by multiplication with ρ_{bin} . The statistical results for different models and data sets are summarized in Table 1.

3.2.1. Statistical Results for the Factor Xa (fXa) Data Set. For a series of 152 3-oxybenzamide fXa inhibitors, we obtained a $r^2(\text{cv})$ value of 0.712 for the combined steric and electrostatic CoMFA fields, while cross-validated regression coefficients for individual steric ($r^2(\text{cv})$: 0.646) and electrostatic fields ($r^2(\text{cv})$: 0.461) are slightly lower.

The models obtained with ParaSurf electron density and MEP fields are of similar statistical significance as those obtained using CoMFA fields with nearly identical $r^2(\text{cv})$ values and number of PLS components (Table 1). The predicted r^2 values for test sets 1 (27 cpds) and 2 (72 cpds) are even somewhat higher except from the MEP model for test set 2 with a predictive r^2 value of 0.554 vs 0.675 for the conventional CoMFA model.

Using raw field values, all five local ParaSurf molecular property fields resulted in statistically significant models. High $r^2(\text{cv})$ values are observed for IE_L , χ_L , and η_L , while medium to low correlations are found for EA_L and α_L . In spite of that, models of both EA_L and α_L demonstrated high external predictivity despite low $r^2(\text{cv})$.⁵⁶ The highest predictive r^2 values for test set 2 were obtained with α_L , while χ_L performs equally well on test set 1. It is not clear whether the mismatch between $r^2(\text{cv})$ and predictive r^2 for α_L is related to the splitting of training and test sets or the lack of pretreatment.

The application of the binarization function ρ_{bin} resulted in increased $r^2(\text{cv})$ and predictive r^2 values in general. Particularly in comparison to CoMFA, higher predictive r^2 values were found for models of all five ParaSurf molecular property fields. Again, α_L performed best with the highest correlations for test set 1, while only η_L demonstrated a higher predictive r^2 value for test set 2. Collectively these models are comparable or perform slightly better in statistical terms compared to the original CoMFA model, when monitoring internal consistency ($r^2(\text{cv})$) and external predictivity (predictive r^2).

3.2.2. Statistical Results for the Benzodiazepine Receptor (BzR) Data Sets. This data set is comprised of 38 imidazo-1,4-diazepines with inhibitory activity against the “diazepam-insensitive” (DI) and “diazepam-sensitive” (DS) isoforms of

Table 1. Statistical Results Using CoMFA or ParaSulf Molecular Property Fields in Combination with PLS for the fXa and BzR Data Set

method	training set			test set 1	test set 2
	$r^2(\text{cv})$	$c(\text{SE})$	r^2	pred. r^2	pred. r^2
fXa data set ($n = 80/27/72$)					
CoMFA ³⁵	0.741	6	0.947	0.732	0.795
CoMFA	0.712	4 (0.501)	0.897	0.687	0.789
CoMFA steric	0.646	3 (0.552)	0.800	0.719	0.786
CoMFA electrostatic	0.461	5 (0.690)	0.820	0.690	0.675
binary electron density (ρ_{bin})	0.616	3 (0.575)	0.795	0.782	0.823
MEP($1 - \rho_{\text{bin}}$)	0.461	5 (0.690)	0.801	0.757	0.554
local ionization energy (IE_{L})	0.517	3 (0.645)	0.796	0.530	0.727
local electron affinity (EA_{L})	0.271	6 (0.808)	0.831	0.697	0.729
local polarizability (α_{L})	0.312	5 (0.780)	0.637	0.717	0.828
local electronegativity (χ_{L})	0.507	6 (0.664)	0.890	0.719	0.715
local hardness (η_{L})	0.556	3 (0.618)	0.750	0.428	0.710
$\text{IE}_{\text{L}} \cdot \rho_{\text{bin}}$	0.611	4 (0.582)	0.831	0.728	0.827
$\text{EA}_{\text{L}} \cdot \rho_{\text{bin}}$	0.588	3 (0.595)	0.760	0.788	0.816
$\alpha_{\text{L}} \cdot \rho_{\text{bin}}$	0.622	3 (0.571)	0.804	0.809	0.822
$\chi_{\text{L}} \cdot \rho_{\text{bin}}$	0.604	4 (0.587)	0.829	0.735	0.793
$\eta_{\text{L}} \cdot \rho_{\text{bin}}$	0.614	4 (0.580)	0.833	0.728	0.837
DI BzR data set ($n = 38/-/-$)					
CoMFA ³⁶	0.73	7 (0.451)	0.96		
CoMFA	0.737	7 (0.443)	0.959		
CoMFA steric	0.724	8 (0.462)	0.949		
CoMFA electrostatic	0.444	9 (0.666)	0.938		
binary electron density (ρ_{bin})	0.605	6 (0.534)	0.932		
MEP($1 - \rho_{\text{bin}}$)	0.712	6 (0.456)	0.929		
local ionization energy (IE_{L})	0.689	9 (0.498)	0.967		
local electron affinity (EA_{L})	0.728	8 (0.458)	0.943		
DI BzR data set ($n = 38/-/-$)					
CoMFA ³⁶	0.70	11 (0.588)	0.99		
CoMFA	0.497	3 (0.671)	0.806		
CoMFA steric	0.554	3 (0.631)	0.859		
CoMFA electrostatic	0.309	5 (0.810)	0.821		
binary electron density (ρ_{bin})	0.603	6 (0.624)	0.949		
MEP($1 - \rho_{\text{bin}}$)	0.455	3 (0.698)	0.734		
local ionization energy (IE_{L})	0.592	7 (0.643)	0.968		
local electron affinity (EA_{L})	0.643	6 (0.592)	0.902		
local polarizability (α_{L})	0.582	6 (0.641)	0.871		
local electronegativity (χ_{L})	0.648	9 (0.618)	0.983		
local hardness (η_{L})	0.580	7 (0.652)	0.953		
$\text{IE}_{\text{L}} \cdot \rho_{\text{bin}}$	0.750	9 (0.521)	0.976		
$\text{EA}_{\text{L}} \cdot \rho_{\text{bin}}$	0.738	9 (0.534)	0.986		
$\alpha_{\text{L}} \cdot \rho_{\text{bin}}$	0.742	11 (0.550)	0.993		
$\chi_{\text{L}} \cdot \rho_{\text{bin}}$	0.762	10 (0.518)	0.986		
$\eta_{\text{L}} \cdot \rho_{\text{bin}}$	0.770	10 (0.509)	0.979		

the benzodiazepine receptor (BzR).³⁶ Most modifications are introduced at the 7-, 8-, 9-, and 10-positions of a benzene ring annulated to the 1,4-diazepine moiety, including various alkyl, halogen, and electron-withdrawing substituents. In position 3, ester substituents and isosteric heterocyclic rings were varied. Substitutions at the 9- and 10-positions are poorly tolerated in terms of activity for both receptor subtypes. But the structure–activity relationship also differs notably between the data sets, exemplified by ligands that exhibit selectivity for one receptor versus the other.

The BzR (DI) data set is dominated by steric requirements of the receptor binding site. Wong et al.³⁶ reported that substitutions at the 7- or 8-positions are required for high affinity binding to DI BzR. Furthermore, affinity at DI increases with increased steric volume of the ester group at position 3 but drops again, if the substituent is larger than *t*-butyl. Well-defined steric requirements led to a CoMFA model using 7 PLS components with a highly significant reported $r^2(\text{cv})$ value of 0.73.³⁶ We found that the CoMFA steric field exclusively leads to a highly significant model ($r^2(\text{cv})$: 0.724) comparable to the original model (Table 1). In contrast the CoMFA electrostatic model alone is more complex and statistically less significant ($r^2(\text{cv})$: 0.444).

PLS models were also prepared from ParaSulf molecular fields. For the binary electron density field (ρ_{bin}), a model with a $r^2(\text{cv})$ of 0.605 is obtained. The MEP field captures the SAR very nicely, as it provides a more robust model showing a $r^2(\text{cv})$ of 0.712, but with only six PLS components. Each of the raw local ParaSulf property fields led to statistically significant models. The highest $r^2(\text{cv})$ values are obtained for EA_{L} (0.728)

and χ_{L} (0.719) at the same number of PLS components and standard error as observed for the CoMFA steric model. Applying pretreatment to the raw local property fields further reduced the optimum number of PLS components, thus increasing the robustness of the models. Especially, the EA_{L} and α_{L} descriptors provide models with five PLS components and low standard error of predictions.

In contrast to the BzR (DI) data set, the BZR (DS) data set is not simply dominated by steric requirements of the receptor. Wong et al. reported a rather complex PLS model employing 11 components resulting in a $r^2(\text{cv})$ value of 0.70.³⁶ We obtained a CoMFA model of less statistical significance at three optimal PLS components (Table 1). A low $r^2(\text{cv})$ value (0.309) and high standard error is observed for the CoMFA electrostatic field, indicating that the SAR of this series is not fully captured by these models. A better correlation is obtained with PLS using electron density (0.603) and MEP (0.455) at a reasonably low number of optimal PLS components (6 and 3, respectively). EA_{L} and χ_{L} led to improved models with $r^2(\text{cv})$ values of 0.643 and 0.658, respectively. Particularly the raw EA_{L} field descriptor led to a robust model (six components) with the smallest standard error which outperforms the steric and electrostatic fields in CoMFA. Application of pretreatment increased model complexity and led to models with higher $r^2(\text{cv})$ and lower standard error values. Here, the η_{L} model performs best ($r^2(\text{cv})$: 0.770), followed by χ_{L} and EA_{L} .

Overall, the models obtained based on ParaSulf molecular property fields for this BzR data set were found to be predictive and in some cases provide superior statistical quality to capture

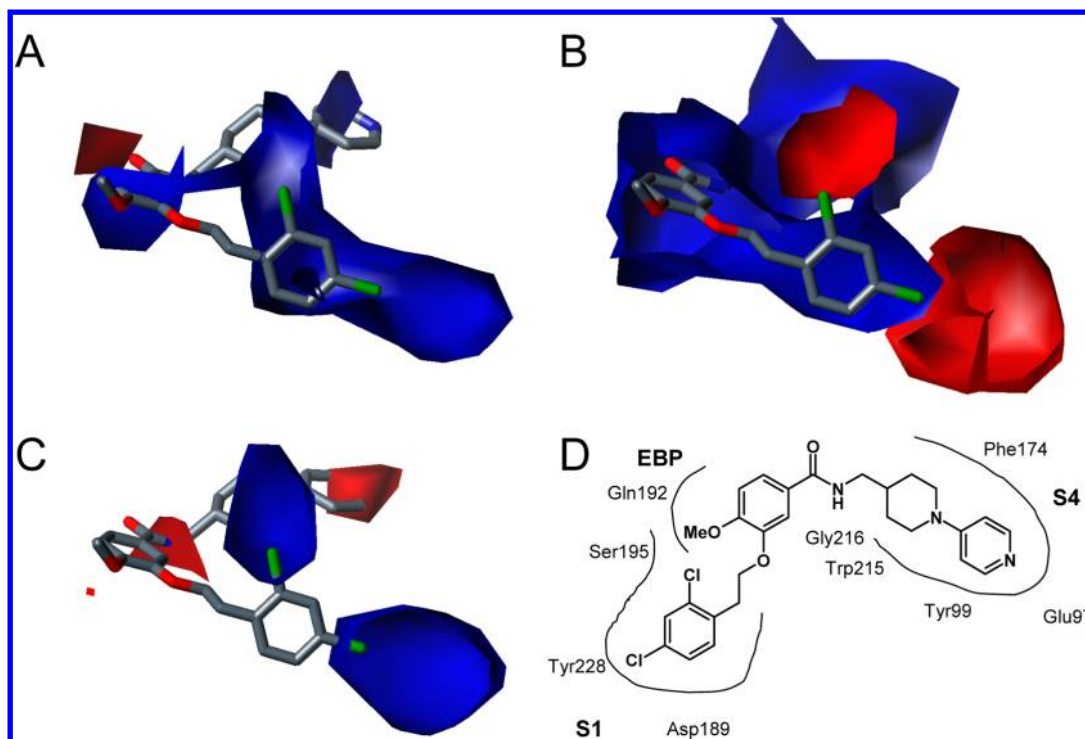


Figure 5. Isocontour plots of actual values of ParaSurf local property fields for the fXa/50 complex ($K_i = 18$ nM).³⁵ Hydrogen atoms are omitted for clarity. (A) Local electron affinity, contour levels: <-110 kcal·mol⁻¹ (red, high electron affinity), >-50 kcal·mol⁻¹ (blue, low electron affinity). (B) Local ionization energy, contour levels: <400 kcal·mol⁻¹ (red, low ionization energy), >500 kcal·mol⁻¹ (blue, high ionization energy). (C) Local polarizability, contour levels <0.27 (red, low polarizability), >0.40 (blue, high polarizability). (D) Two-dimensional representation of essential protein ligand interactions).

SAR trends in these chemical series than the classical CoMFA approach.

3.3. Chemical Interpretation of ParaSurf Derived 3D-QSAR Models. **3.3.1. Binding Mode of Factor Xa Inhibitors.** Chemical interpretability is an important aspect of 3D-QSAR models. Consequently, this improved field-based description can also allow a more consistent chemical interpretation of the 3D-QSAR models, thereby revealing important SAR trends.

The experimental binding mode of inhibitor 50,^{35,57} which has a K_i value of 18 nM, was taken from the X-ray structure in complex with fXa (PDB 2BMG, resolution 2.7 Å) and used as reference for the interpretation of the ParaSurf derived 3D-QSAR models.

While first-generation fXa inhibitors rely on the interaction of a basic moiety with Asp189 at the bottom of the protease S1 pocket, this chemical series is characterized by a favorable nonbasic interaction with the dichlorophenyl group located in S1, as shown schematically in Figure 5D. The *para*-chlorine atom is involved in a direct Cl $\cdots\pi$ contact pointing toward the center of the aromatic ring of Tyr228 with the carbon–chlorine bond directed toward the plane of the Tyr228 ring. Ligands for the protease containing an aryl-chlorine substituent often interact favorably with the aromatic moiety of Tyr228.^{35,58–61} The 4-methoxy group attached to the central oxybenzamide scaffold is situated in the ester binding pocket in contact with the side chains of Glu147 and the flexible Gln192. The benzamide nitrogen interacts with Gly216–CO, while the piperidine in a chair conformation orients the pyridine substituent toward the S4 pocket, stacking the aromatic ring between the aromatic side chains of Tyr99 and Phe174. The pyridine nitrogen is hydrogen bonded to the carbonyl oxygen of Thr98 via a structurally conserved water molecule in S4, which also interacts

with the carbonyl oxygen of Ile175 and the Thr98 hydroxyl group.

3.3.2. Interpretation of ParaSurf Derived 3D Property Fields on fXa Inhibitor. Isocontour plots derived from actual values of the ParaSurf local property fields are shown for the ligand from the fXa/50 X-ray structure in Figures 5A–C to highlight molecular features and potential interactions with the enzyme. Two isocontour levels were chosen for visualization; the first isocontour is located at a manually selected threshold above and the second below the average of the field values. The ParaSurf local electron affinity (EA_L) field is shown in Figure 4A with a contour level <-110 kcal·mol⁻¹ (red) to indicate regions of high electron affinity around the ligand, while a contour level of >-50 kcal·mol⁻¹ (blue) indicates regions of low electron affinity. The dominating high electron affinity region (red) is located close to the central benzamide moiety, while areas of low electron affinity (blue) are located predominantly around the π -systems and chlorine atoms. The ParaSurf local ionization energy (IE_L) field is shown in Figure 4B with a contour level <400 kcal·mol⁻¹ (red) indicating areas of low ionization energy, and >500 kcal·mol⁻¹ (blue) highlighting regions with higher ionization energy. In particular, red contours surround the caps of both chlorine atoms at the distal dichlorophenyl-ring situated in the protease S1 pocket, which is attributed to the nature of higher-row elements with decreasing ionization energy. The ParaSurf local polarizability (α_L) field is shown in Figure 4C. Here, a contour level <0.27 (red) indicates areas of low polarizability, while a level >0.40 (blue) indicates regions of high polarizability. The chemically intuitive concept of hardness and softness³⁴ manifests itself in this field, so the extreme values are located around the “soft” polarizable chlorine atoms at the dichlorophenyl ring in S1 (blue) and around the “hard” nitrogen

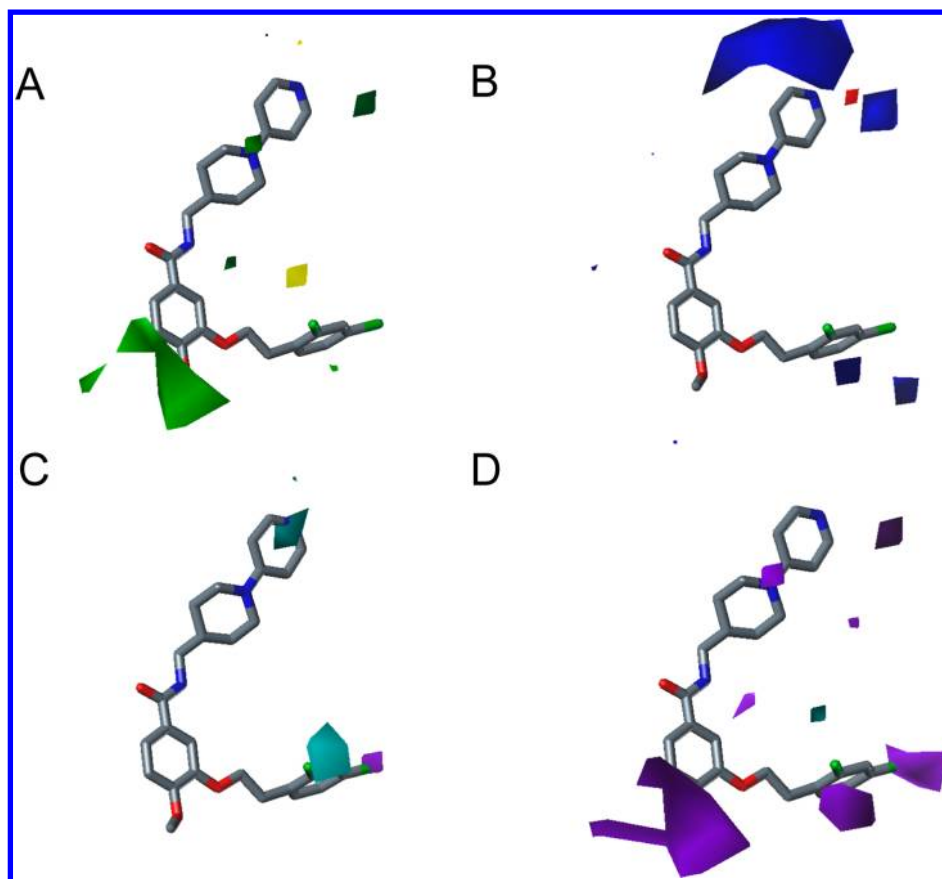


Figure 6. Contour maps from different models in combination with compound **50** ($K_i = 18$ nM).³⁵ Hydrogen atoms are omitted for clarity. (A) Shown is the std*coeff contour map for the binary electron density (ρ_{bin}) model. Green contours (>90% contribution) refer to regions where additional steric bulk is favorable for affinity, while yellow contours (<30% contribution) indicate disfavored areas. (B) Shown is the std*coeff contour map for the MEP model. Blue contours (>80% contribution) refer to regions where positively charged substituents are favorable for affinity. Red contours (<20% contribution) indicate regions where negatively charged substituents are favorable for affinity. (C) Shown is the std*coeff contour map for the local polarizability (α_L) model. Purple contours (>75% contribution) refer to regions where substituents with higher polarizability are favorable for affinity. Cyan contours (<15% contribution) indicate regions where substituents with lower polarizability are favorable for affinity. (D) Shown is the std*coeff contour map for the $\alpha_L \rho_{\text{bin}}$ model. Purple contours (>85% contribution) refer to regions where more steric bulk and polarizability is favorable for affinity. Cyan contours (<20% contribution) indicate regions where less steric bulk and polarizability are favorable for affinity.

atoms at the central amide and distal pyridyl substituent (red). All displayed fields around chlorine in Figure 5 extend significantly beyond the atom's van der Waals radius in the direction of the C–Cl bond vector.

3.3.3. Chemical Interpretation of the fXa Models. A comparison of the ParaSurf and CoMFA countour maps is given in Figure 6 and mapped onto the experimental binding pose of compound **50** in complex with factor Xa.³⁵ The std*coeff contour maps were obtained from the PLS coefficients at each grid point. Although ParaSurf derived 3D-QSAR results are derived from ligand information, these contour maps are consistent with requirements of the fXa binding site.

Figure 6A shows a std*coeff contour map for the binary electron density (ρ_{bin}) model with an $r^2(\text{cv})$ value of 0.616 and three PLS components (Table 1). Green contours (>90% contribution) refer to regions in which additional steric bulk is favorable for affinity, while yellow contours (<30% contribution) indicate disfavored areas. In particular, the green contour region surrounding the *para*-methoxy group attached to the central oxybenzamide scaffold matches the steric requirements in the solvent-exposed ester-binding pocket of factor Xa on top of the Cys191–Cys220 disulfide bridge, limited in size by the

neighboring Gln192 side chain. Although this region opens to the bulk solvent, some interactions with the side chains of Glu143 and Arg222 might be possible. These contours also suggest that steric bulk at positions 4 and 5 of the benzamide ring is favorable for affinity. The optimal fit of the S4 directed 4-piperidyl-pyridyl substituents is indicated by another smaller green contour close to the distal ring. The aromatic portion of this substituent is involved in a favorable π -stacking contact to the aromatic residues of Phe174 and Trp215 defining this pocket. The yellow contour close to the pyridyl ring indicates the boundary of S4. This contour is located at the position of one structurally conserved water molecule within this pocket and suggests that replacement of this hydrogen-bonded water molecule in S4 might not be very favorable for affinity. Hence, extending the size of S4-directed substituents is detrimental for affinity. These interpretations are in good agreement to the interpretations of the CoMFA steric field.³⁵

The chemical interpretation of the MEP model (MEP(1 – ρ_{bin})) is shown as the std*coeff contour map in Figure 6B; here blue contours (>80% contribution) refer to regions in which positively charged substituents are favorable for affinity, while red contours (<20% contribution) indicate those in which

negatively charged substituents are favorable. A red contour region close to the pyridyl-nitrogen atom situated in the S4 pocket indicates favorable interactions in particular with the structurally conserved water molecule in S4. Replacement of this water molecule might not be favorable for affinity for this particular series, while other chemical series that interact with this protease do not exploit this particular interaction,⁵⁹ which is mainly caused by a slightly different topology of the S4 pocket. Blue contours near the dichlorophenyl-ring in S1 indicate that a less electronegative aromatic ring would be favorable for activity, as exemplified in the favorable substitution pattern by electron-withdrawing groups with positive Hammett σ_p values.⁶² This information cannot be extracted directly from the analysis of the corresponding CoMFA electrostatic field.

The chemical interpretation of the local polarizability (α_L) model is shown as the std*coeff contour map in Figure 6C. Purple contours (>75% contribution) refer to regions in which substituents with higher polarizability are favorable for affinity, thereby indicating a preference for the “softness” of functional groups, while cyan contours (<15% contribution) indicate those in which substituents with lower polarizability (“hardness”) are favorable for affinity. The cyan contour surrounding the pyridyl-nitrogen in S4 highlights the preference of a functional group with low polarizability at this position, primarily driven by the favorable interaction to the structurally conserved water molecule in this pocket. This matches perfectly the “hardness” of the aromatic nitrogen in this region. In contrast, the purple contour at the chlorine atom in *para*-position of the dichlorophenyl ring in S1 highlights the favorable influence of higher polarizability (“softness”) to binding affinity in this area. This SAR trend reflects to a certain extent a favorable Cl $\cdots\pi$ interaction to the aromatic ring of Tyr228,⁵⁸ while the effects of replacing a structurally conserved water in this hydrophobic S1 site⁶¹ might also account for some of the affinity increase for fXa. Interestingly, the cyan contour surrounding the *ortho*-chlorine substituent indicates that “harder” substituents might be favorable, which could partially be attributed to interactions pointing toward the active triad Ser195–O γ .

A combined std*coeff contour map is shown in Figure 6D for the model based on the polarizability and the binary electron density field ($\alpha_L \cdot \rho_{\text{bin}}$). This single field combines steric and polarizability influences to explain the experimental SAR. Purple contours in Figure 6D (>85% contribution) refer to regions where more steric bulk and polarizability is favorable for affinity, while cyan contours (<20% contribution) indicate regions where less steric bulk and polarizability are favorable. The favorable effect of a *para*-chlorine atom in S1 on binding affinity due to its higher polarizability plus steric bulk (i.e., partially reflecting the replacement of the structurally conserved water in S1) is captured by a purple contour. The favorable steric requirements of substituents in the ester-binding pocket, here filled with the methoxy-substituent, can be deduced from a purple contour in Figure 6D. Furthermore, favorable contours outline the piperidyl-pyridyl substituent in S4.

None of the discussed electrostatic and polarizability effects can be extracted directly from classical CoMFA electrostatic field maps. Hence, the detailed chemical interpretation of these QM-derived ParaSurf fields provides a significant benefit for identifying chemically meaningful correlations and understanding SAR trends, which might allow these lessons to be used to design novel molecules beyond simply exploring steric and hydrogen-bonding features.

3.3.4. Interpretation of ParaSurf Derived 3D Property Fields on the BzR Data Set. In this section, we will discuss local differences in electron affinity caused by differences in chemical topology of the BzR data sets. Isocontour plots derived from actual values of the ParaSurf local electron affinity fields are shown in Figure 7 for key informative compounds from this data set. Regions of low local electron affinity (EA_L) are shown in blue at a contour level of < -110 kcal·mol⁻¹ while regions of high local electron affinity are shown in red at a contour level of >50 kcal·mol⁻¹.

A key recognition spot is indicated by a central low electron affinity area (blue) located predominantly around the central benzodiazepine moiety. Pronounced differences are, however, found between the chlorine and fluorine substituents. Molecules having chlorine substituents in positions 7 and 8 share large areas of chlorine-induced low electron affinity that are located in different regions in space dependent on the substituent position and that extend beyond the van-der-Waals radius of the chlorine atom (Figures 7B and E). This patch is however not shared with the fluorine substituent (Figure 7D). Here, the isocontour plot for the local electron affinity (EA_L) field very much resembles that of a hydrogen in position 8 (Figure 7C). This picture reminds to the concept of “bioisosterism” between hydrogen and fluorine that is often used in medicinal chemistry, for instance to improve physico-chemical properties of molecules.^{1–3}

3.3.5. Chemical Interpretation of the BzR Models. The local electron affinity (EA_L) field has particular value in the SAR interpretation of the BzR data set (Figure 8). In the context of electron affinity maps,⁵⁵ lower values are typically associated with the presence of substituents, e.g. halogens, and higher values with their absence.

The local electron affinity based models reflect key areas responsible for affinity and selectivity to the DS and DI receptor subtypes. The std*coeff contour map for the DS subtype model (Table 1) is shown in Figure 8 on the left, while the corresponding model for the DI subtype is shown on the right. Purple contours (>80% contribution) in this figure refer to regions where substituents inducing a higher local electron affinity are favorable for affinity, while cyan contours (<20% contribution) indicate regions where substituents inducing a lower local electron affinity are favorable for affinity.

Near the 7-position of the imidazo-1,4-diazepine compound 2⁵⁷ with a value of 0.2 nM for DS and 20 nM for DI, a large favorable region of low electron affinity is observed as purple contour in the DS model on the left of Figure 8. Consequently, halogen atoms are tolerated at this position as they induce lower values of EA_L . This purple contour is surrounding a chlorine substituent in position 7 of this scaffold. Replacing this chlorine by a fluorine atom reduces DS affinity by a factor of 4 (compound 20), while a hydrogen atom at this position results in a lower activity of 1.3 nM for DS (compound 9).

The corresponding purple region is not found in the std*coeff contour map derived from the DI model in Figure 8 on the right. Correspondingly, substitutions at the 7-position do not have a strong effect on affinity for the DI subtype. In contrast to the DI model, a purple contour region is located near the 8-position, while high electron affinity is favored in the DS model at the very same position (Figure 7, left). The absence of halogen substituents improves the DS affinity, while the DI affinity is decreased. This matches well to the observed SAR trend,³⁶ where substitution in both 7- and 8-positions increase affinity for DI. This effect is stronger at the 8-position.

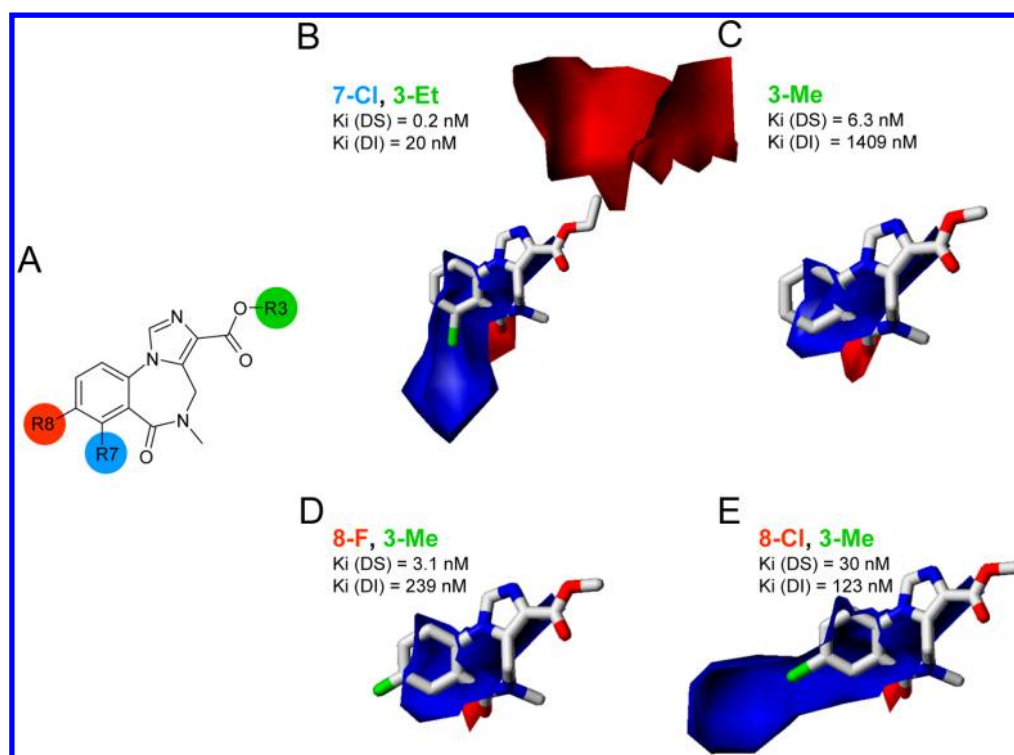


Figure 7. Isocontour plots of actual values of ParaSurf local electron affinity field (EA_L) and activity data for four informative compounds of the BzR data set.^{36,57} The following contour levels were used: <-110 kcal·mol⁻¹ (red, high electron affinity), >-50 kcal·mol⁻¹ (blue, low electron affinity). (A) Schematic description of the imidazo-1,4-diazepine scaffold including substituent positions. (B) Compound 2. (C) Compound 8. (D) Compound 7. (E) Compound 11.

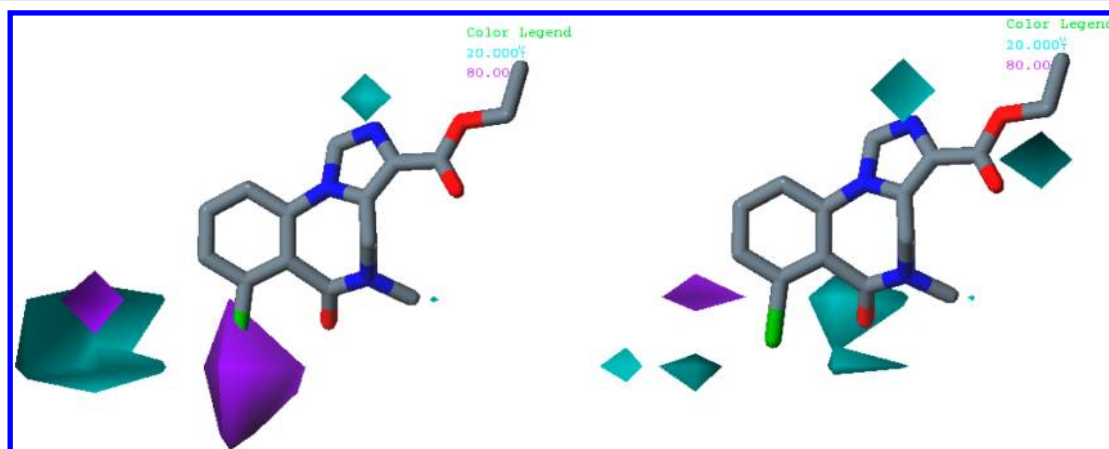


Figure 8. Std*coeff contour maps for the local electron affinity (EA_L) models of DS (left) and DI (right) BzR subtypes in combination with compound 2 (K_i (DS/DI) = 0.2 nM/20 nM).³⁶ Hydrogen atoms are omitted for clarity. Purple contours (>80% contribution) refer to regions where substituents inducing a higher local electron affinity are favorable for affinity. Cyan contours (<20% contribution) indicate regions where substituents inducing a lower local electron affinity are favorable for affinity.

Substitutions in 7-position strongly increase affinity for DS. Furthermore, a region of high favorable affinity is found in the DI model near the 3-position (ester group), revealing that isosteric replacements of the ester group by different oxadiazole moieties are detrimental for affinity. This hotspot is not found in the DS model.

4. CONCLUSIONS

The inability of established current force-field methods to describe molecules with halogen atoms or hypervalent elements like sulfur means that many noncovalent interactions such as halogen-bonds or halogen- π interactions cannot be accounted

for. This limitation is caused by neglecting the anisotropy of the electron density distribution, the σ -hole. The importance of this effect has very recently also prompted several research groups to extend molecular-mechanics force fields to treat the σ -hole by adding explicit point charges adjacent to the center of higher halogen atoms.^{18–20}

Rather than add such ad hoc (and necessarily limited) corrections to atom-monopole based force fields, we have calculated quantum-mechanical (QM) local properties such as the molecular electrostatic potential, local ionization energy, electron affinity, or polarizability on a regularly spaced lattice. Our intention was to produce augmented molecular property

fields that do not suffer from the electronic force-field limitation for 3D-QSAR applications.

In the present study, we have demonstrated the usefulness of these local properties in molecular-field-based 3D-QSAR with data sets of fXa inhibitors and GABA-A/BzR ligands. First we explored the characteristics of QM-based local property fields to demonstrate their suitability for statistical analyses after a pretreatment workflow. A binary truncation function turned out to be most effective to dampen irregular field maxima inside of the solvent accessibility radius of ligands. Property fields were then calculated and analyzed with PLS, resulting in models with equal or even better statistical performance in terms of internal consistency and external predictivity compared to reference CoMFA models. From the current investigations, there is no clear indication that a particular ParaSurf 3D-molecular property field performs consistently superior than all the others. We conclude that the selection of an optimal descriptor combination depends on the investigated data set. Our studies also suggest that a binary normalization (pretreatment) for electron density and MEP fields has a positive influence on statistical performance, which is not a general rule for other molecular property fields, where these options must be evaluated carefully using external test sets.

Finally, we have evaluated whether QM-based property fields allow a consistent chemical interpretation of the resulting 3D-QSAR models, and thereby reveal important SAR trends. None of the observed electrostatic and polarizability effects extracted from improved QM-based property fields discussed above can be extracted from classic CoMFA electrostatic field maps. For both the fXa and GABA-A data sets, additional insights into the structure–activity relationships influenced by a particular substitution with halogen atoms can be extracted from these analyses. Within the fXa data set, protein 3D structural information in complex with members of the training set is available and allowed these SAR trends to be interpreted on a structural basis. The detailed chemical interpretation of these QM-derived ParaSurf fields provides a significant benefit for identifying chemically interpretable correlations. This allows a more detailed analysis of the factors that affect the affinity of novel ligands that goes far beyond the conventional steric and hydrogen-bonding features. This quantitative SAR information from interpreting these QSAR models agrees with all known experimental data and protein binding site topologies, and thus provides design guidelines and reasonable activity estimates to guide further compound optimization.

AUTHOR INFORMATION

Corresponding Author

*E-mail: stefan.guessregen@sanofi.com (S.G.), tim.clark@chemie.uni-erlangen.de (T.C.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) as part of the hpCADD-project (www.hpCADD.com). The authors thank K.-H. Baringhaus, M. Nazaré, D. W. Will, and V. Wehner (all Sanofi) and M. Hennemann (Univ. of Erlangen) for discussions, A. Kugelstadt (Sanofi) for support, and G. Wong and K. Koehler for providing the alignment of the BzR data set.

REFERENCES

- (1) Böhm, H.-J.; Banner, D.; Bendels, S.; Kansy, M.; Kuhn, B.; Müller, K.; Obst-Sander, U.; Stahl, M. Fluorine in medicinal chemistry. *ChemBioChem* **2004**, *5*, 637–643.
- (2) Hagmann, W. K. The Many Roles for Fluorine in Medicinal Chemistry. *J. Med. Chem.* **2008**, *51*, 4359–4369.
- (3) Ismael, F. M. D. Important fluorinated drugs in experimental and clinical use. *J. Fluorine Chem.* **2002**, *118*, 27–33.
- (4) Gerebtzoff, G.; Li-Blatter, X.; Fischer, H.; Frentzel, A.; Seelig, A. Halogenation of drugs enhances membrane binding and permeation. *ChemBioChem* **2004**, *5*, 676–684.
- (5) Politzer, P.; Murray, J. S.; Clark, T. Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7748–7757.
- (6) El Kerdawy, A.; Wick, C. R.; Hennemann, M.; Clark, T. Predicting the Sites and Energies of Noncovalent Intermolecular Interactions Using Local Properties. *J. Chem. Inf. Model.* **2012**, *52*, 1061–1071.
- (7) Murray, J. S.; Riley, K. E.; Politzer, P.; Clark, T. Directional Weak Interactions: σ -Hole Bonding. *Aus. J. Chem.* **2010**, *63*, 1598–1607.
- (8) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. Halogen bonding: the σ -hole. *J. Mol. Model.* **2007**, *13*, 291–296.
- (9) Murray, J. S.; Lane, P.; Clark, T.; Politzer, P. σ -Hole bonding: molecules containing group VI atoms. *J. Mol. Model.* **2007**, *13*, 1033–1038.
- (10) Murray, J. S.; Lane, P.; Clark, T.; Riley, K. E.; Politzer, P. σ -Holes, π -holes and electrostatically-driven interactions. *J. Mol. Model.* **2012**, *18*, 541–548.
- (11) Metrangola, P.; Resnati, G. *Halogen Bonding: Fundamentals and Applications*; Springer: Berlin, 2008.
- (12) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. An overview of halogen bonding. *J. Mol. Model.* **2007**, *13*, 305–311.
- (13) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J.-M.; Hartmann, G.; Banner, D. W.; Haap, W.; Diederich, F. Systematic Investigation of Halogen Bonding in Protein–Ligand Interactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 314–318.
- (14) Hardegger, L. A.; Kuhn, B.; Spinnler, B.; Anselm, L.; Ecabert, R.; Stihle, M.; Gsell, B.; Thoma, R.; Diez, J.; Benz, J.; Plancher, J.-M.; Hartmann, G.; Isshiki, Y.; Morikami, K.; Shimma, N.; Haap, W.; Banner, D. W.; Diederich, F. Halogen Bonding at the Active Sites of Human Cathepsin L and MEK1 Kinase: Efficient Interactions in Different Environments. *ChemMedChem* **2011**, *6*, 2048–2054.
- (15) Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.
- (16) (a) Halgren, T. A. Maximally diagonal force constants in dependent angle-bending coordinates. II. Implications for the design of empirical force fields. *J. Am. Chem. Soc.* **1990**, *112*, 4710–4723. (b) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720–729.
- (17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (18) Ibrahim, M. A. A. Molecular Mechanical Study of Halogen Bonding in Drug Discovery. *J. Comput. Chem.* **2011**, *32*, 2564–2574.
- (19) Kolar, M.; Hobza, P. On Extension of the Current Biomolecular Empirical Force Field for the Description of Halogen Bonds. *J. Chem. Theory Comput.* **2012**, *8*, 1325–1333.
- (20) Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.*, DOI: 10.1021/ct300180w
- (21) Politzer, P.; Murray, J. S.; Concha, M. C. σ -hole bonding between like atoms; a fallacy of atomic charges. *J. Mol. Model.* **2008**, *14*, 659–665.
- (22) *3D-QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden (NL), 1993.

- (23) Cramer, R. D., III; Patterson, D. E.; Bunce, J. E. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (24) Clark, M.; Cramer, R. D., III; Jones, D. M.; Patterson, D. E.; Simeroth, P. E. Comparative Molecular Field Analysis (CoMFA). 2. Towards its use with 3D-Structural Databases. *Tetrahed. Comp. Meth.* **1990**, *3*, 47–59.
- (25) Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793–1874.
- (26) Van Damme, S.; Bultinck, P. 3D QSAR based on conceptual DFT molecular fields: Antituberculous activity. *J. Mol. Struct. THEOCHEM* **2010**, *943*, 83–89.
- (27) Beck, M. E.; Schindler, M. Quantitative structure-activity relations based on quantum theory and wavelet transformations. *Chem. Phys.* **2009**, *356*, 121–130.
- (28) Ehresmann, B.; Martin, B.; Horn, A.H. C.; Clark, T. Local molecular properties and their use in predicting reactivity. *J. Mol. Model.* **2003**, *9*, 342–347.
- (29) Clark, T. QSAR and QSPR based solely on surface properties? *J. Mol. Graphics Model.* **2004**, *22*, 519–525.
- (30) El Kerdawy, A.; Wick, C. R.; Hennemann, M.; Clark, T. Predicting the Sites and Energies of Non-Covalent Intermolecular Interactions Using Local Properties. *J. Chem. Inf. Model.* **2012**, in press.
- (31) Sjöberg, P.; Murray, J. S.; Brinck, T.; Politzer, P. Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity. *Can. J. Chem.* **1990**, *68*, 1440–1443.
- (32) Clark, T. The Local Electron Affinity for Non-Minimal Basis Sets. *J. Mol. Model.* **2010**, *16*, 1231–1238.
- (33) Mulliken, R. S. New electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *J. Chem. Phys.* **1934**, *2*, 782–793.
- (34) Pearson, R. G. Absolute electronegativity and hardness: application to inorganic chemistry. *Inorg. Chem.* **1988**, *27*, 734–740.
- (35) Matter, H.; Will, D. W.; Nazaré, 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*, 3290–3312.
- (36) Wong, G.; Koehler, K. F.; Skolnick, P.; Gu, Z. Q.; Ananthan, S.; Schönholzer, P.; Hunkeler, W.; Zhang, W.; Cook, J. M. Synthetic and computer-assisted analysis of the structural requirements for selective, high-affinity ligand binding to diazepam-insensitive benzodiazepine receptors. *J. Med. Chem.* **1993**, *36*, 1820–1830.
- (37) (a) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbenson, K.; Geladi, P.; Hellberg, S.; Lindberg, W.; Sjöström, M., In *Chemometrics: Mathematics and Statistics in Chemistry*; Kowalski, B., Ed.; Reidel, Dordrecht: The Netherlands, 1984, pp 17–95. (b) Dunn, W. J.; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate Structure-Activity Relationship Between Data from a Battery of Biological Tests and an Ensemble of Structure Descriptors: The PLS Method. *Quant. Struct.-Act. Relat.* **1984**, *3*, 31–137. (c) Geladi, P. Notes on the History and nature of Partial Least Squares (PLS) Modelling. *J. Chemom.* **1988**, *2*, 231–246.
- (38) (a) Wold, S. Cross-Validatory Estimation of the Number of Component in Factor and Principal Component Models. *Technometrics* **1978**, *4*, 397–405. (b) Diaconis, P.; Efron, B. Computer-Intensive Methods for Statistics. *Sci. Am.* **1984**, *116*, 96–117. (c) Cramer, R. D., III; Bunce, J. D.; Patterson, D. E. Crossvalidation, Bootstrapping and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant.-Struct.-Act. Relat.* **1988**, *7*, 18–25.
- (39) McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333–344.
- (40) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- (41) Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.
- (42) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first-row elements. *J. Am. Chem. Soc.* **1980**, *102*, 939–947.
- (43) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- (44) Stewart, J. J. P. MOPAC2000, Fujitsu, Ltd, Tokyo, Japan, 1999. MOPAC 6.0 was once available as Stewart, J. J. P. QCPE no. 455, *Quantum Chemistry Program Exchange*; Bloomsville, IN, 1990.
- (45) MOPAC6; Cepos InSilico Ltd., Bedford MK42 8BQ, UK, 2009.
- (46) *Parasurf10*; Cepos InSilico Ltd, Bedford MK42 8BQ, UK, 2010.
- (47) A grid spacing of 1 Å did not improve the results, data not shown.
- (48) *Sybyl x1.3*; Tripos International, St. Louis, MO, USA, 2010.
- (49) Sheridan, R. P.; Nachbar, R. B.; Bush, B. L. Extending the trend vector: The trend matrix and sample-based partial least squares. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 323–340.
- (50) (a) Thibaut, U.; Folkers, G.; Klebe, G.; Kubinyi, H.; Merz, A.; Rognan, D. Recommendations for CoMFA Studies and 3D QSAR Publications. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 711–717. (b) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and Limitations. In *3D QSAR in Drug Design. Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; pp 583–616. (c) 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; pp 443–485.
- (51) 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.
- (52) Horn, A. H. C.; Lin, J.-H.; Clark, T. Multipole electrostatic model for MNDO-like techniques with minimal valence spd-basis sets. *Theor. Chem. Acc.* **2005**, *114*, 159–168; Erratum: *Theor. Chem. Acc.* **2007**, *117*, 461–465.
- (53) Colapietro, M.; Domenicano, A. Structural studies of benzene derivatives. X. Refinement of the structure of p-chlorobenzoic acid. *Acta Crystallogr. Sect. B: Struct. Crystallogr. Cryst. Chem.* **1982**, *B38*, 1953–1957.
- (54) CSD, version 5.33 (November 2011), available from The Cambridge Crystallographic Data Centre (CCDC): 12 Union Road, Cambridge, CB2, 1EZ, UK.
- (55) IUPAC. *Compendium of Chemical Terminology*, 2nd ed. (the "Gold Book"); McNaught, A. D., Wilkinson, A. Blackwell Scientific Publications: Oxford, 1997. Nic, M.; Jirat, J.; Kosata, B. XML on-line corrected version. <http://goldbook.iupac.org> (2006); Jenkins, A., Ed.; doi: 10.1351/goldbook.
- (56) Golbraikh, A.; Tropsha, A. Beware of q²! *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
- (57) Compound numbering according to the original publication.
- (58) Matter, H.; Nazaré, M.; Güssregen, S.; Will, D. W.; Schreuder, H.; Bauer, A.; Urmann, M.; Ritter, K.; Wagner, M.; Wehner, V. Evidence for C-Cl/C-Br... π Interactions as an Important Contribution to Protein-Ligand Binding Affinity. *Angew. Chem., Int. Ed.* **2009**, *48*, 2911–2916.
- (59) Nazaré, M.; Will, D. W.; Matter, H.; Schreuder, H.; Ritter, K.; Urmann, M.; Essrich, M.; Bauer, A.; Wagner, M.; Czech, J.; Lorenz, M.; Laux, V.; Wehner, V. Probing the Subpockets of Factor Xa Reveals Two Binding Modes for Inhibitors Based on a 2-Carboxyindole Scaffold: A Study Combining Structure-Activity Relationship and X-ray Crystallography. *J. Med. Chem.* **2005**, *48*, 4511–4525.
- (60) (a) Tucker, T. J.; Brady, S. F.; Lumma, W. C.; Lewis, S. D.; Gardell, S. J.; Naylor-Olsen, A. M.; Yan, Y.; Sisko, J. T.; Stauffer, K. J.

Lucas, B. J.; Lynch, J. J.; Cook, J. J.; Stranieri, M. T.; Holahan, M. A.; Lyle, E. A.; Baskin, E. P.; Chen, I. W.; Dancheck, K. B.; Krueger, J. A.; Cooper, C. M.; Vacca, J. P. Design and Synthesis of a Series of Potent and Orally Bioavailable Noncovalent Thrombin Inhibitors That Utilize Nonbasic Groups in the P1 Position. *J. Med. Chem.* **1998**, *41*, 3210–3219. (b) Adler, M.; Kochanny, M. J.; Ye, B.; Rummenik, G.; Light, D. R.; Biancalana, S.; Whitlow, M. Crystal Structures of Two Potent Nonamidine Inhibitors Bound to Factor Xa. *Biochemistry* **2002**, *41*, 15514–15523. (c) Stubbs, M. T.; Reyda, S.; Dullweber, F.; Moller, M.; Klebe, G.; Dorsch, D.; Mederski, W. W. K. R.; Wurziger, H. pH-dependent binding modes observed in trypsin crystals: lessons for structure-based drug design. *ChemBioChem* **2002**, *3*, 246–249. (d) Maignan, S.; Guilloteau, J.-P.; Choi-Sledeski, Y. M.; Becker, M. R.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. Molecular Structures of Human Factor Xa Complexed with Ketopiperazine Inhibitors: Preference for a Neutral Group in the S1 Pocket. *J. Med. Chem.* **2003**, *46*, 685–690. (e) Roehrig, S.; Straub, A.; Pohlmann, J.; Lampe, T.; Pernerstorfer, J.; Schlemmer, K.-H.; Reinemer, P.; Perzborn, E. Discovery of the Novel Antithrombotic Agent 5-Chloro-N-((5S)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolidin-5-yl)methylthiophene-2-carboxamide (BAY 59-7939): An Oral, Direct Factor Xa Inhibitor. *J. Med. Chem.* **2005**, *48*, 5900–5908. (f) Hartshorn, M. J.; Murray, C. W.; Cleasby, A.; Frederickson, M.; Tickle, I. J.; Jhoti, H. Fragment-Based Lead Discovery Using X-Ray Crystallography. *J. Med. Chem.* **2005**, *48*, 403–413. (g) Shi, Y.; Sitkoff, D.; Zhang, J.; Klei, H. E.; Kish, K.; Liu, E.C.-K.; Hartl, K. S.; Seiler, S. M.; Chang, M.; Huang, C.; Youssef, S.; Steinbacher, T. E.; Schumacher, W. A.; Grazier, N.; Pudzianowski, A.; Apedo, A.; Discenza, L.; Yanchunas, J.; Stein, P. D.; Atwal, K. S. Design, Structure-Activity Relationships, X-ray Crystal Structure, and Energetic Contributions of a Critical P1 Pharmacophore: 3-Chloroindole-7-yl-Based Factor Xa Inhibitors. *J. Med. Chem.* **2008**, *51*, 7541–7551. (h) Straub, A.; Roehrig, S.; Hillisch, A. Entering the era of non-basic P1 site groups: discovery of Xarelto (Rivaroxaban). *Curr. Top. Med. Chem.* **2010**, *10*, 257–269.

(61) (a) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831. (b) Abel, R.; Salam, N. K.; Shelley, J.; Farid, R.; Friesner, R. A.; Sherman, W. Contribution of Explicit Solvent Effects to the Binding Affinity of Small-Molecule Inhibitors in Blood Coagulation Factor Serine Proteases. *ChemMedChem* **2011**, *6*, 1049–1066.

(62) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington DC, US, 1995.