JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Predicting the Sites and Energies of Noncovalent Intermolecular Interactions Using Local Properties

Ahmed El Kerdawy,[†] Christian R. Wick,[†] Matthias Hennemann,[†,‡] and Timothy Clark*[,†,‡,§]

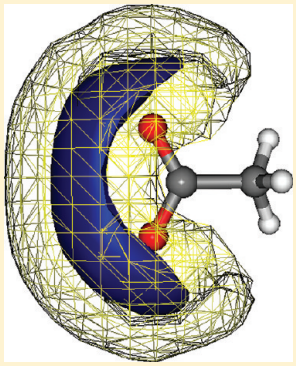
[†]Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraβe 25, 91052 Erlangen, Germany

[‡]Interdisciplinary Center for Molecular Materials, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraβe 49, 91052 Erlangen, Germany

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

Supporting Information

ABSTRACT: Feed-forward artificial neural nets have been used to recognize H-bond donor and acceptor sites on drug-like molecules based on local properties (electron density, molecular electrostatic potential and local ionization energy, electron affinity, and polarizability) calculated at grid points around the molecule. Interaction energies for training were obtained from B97-D and ωB97X-D/aug-cc-pVDZ density-functional theory calculations on a series of model central molecules and H-bond acceptor and donor probes constrained to the grid points used for training. The resulting models provide maps of both classical and unusual H- and halogen-bonding sites. Note that these reactions result even though only classical H-bond donors and acceptors were used as probes around the central molecules. Some examples demonstrate the ability of the models to take the electronics of the central molecule into consideration and to provide semiquantitative estimates of interaction energies at low computational cost.



INTRODUCTION

Noncovalent interactions such as hydrogen (H-)[1−3] or halogen (σ-hole)[4−6] bonding, salt bridges,[7] π-stacking,[8] and cation-π interactions[9] are decisive in natural processes that depend on nonbonding intra- and intermolecular interactions. These include (a) the existence of the liquid phase and related effects such as solvation phenomena, (b) stable secondary-structure motifs of biopolymers such as DNA, RNA, and proteins, where hydrogen bonding controls the backbone structure of protein chains and π-stacking interactions stabilize the double-helix structure of the DNA, (c) molecular recognition processes, for example mutual recognition of bases within DNA, which are important in life processes that ensure an extremely high fidelity in the formation of proteins and RNA,[10−12] or, most predominantly in the context of drug design, protein−ligand interactions.[13] Unlike covalent interactions, which formally involve overlap between unfilled electronic shells of two subsystems at interatomic distances of 2 Å or less, noncovalent interactions can exist at much larger distances, sometimes more than 10 Å,[11] and are generally considerably weaker than covalent bonds. However, in certain cases, H-bonding can be of comparable strength to covalent interactions, as in the case of H-bonding between an anion and a neutral system.[10] Noncovalent interactions are usually a balance between attractive and repulsive forces;[14] the former are usually thought of in terms of four components, Coulomb, polarization, donor−acceptor, and dispersion. The electrostatic (Coulomb) interaction takes place between permanent charges, dipoles, quadrupoles, and higher multipoles of the two subsystems and is the dominant energy term in the gas phase. The polarization (induction) interaction takes place between permanent and induced multipoles and is generally weaker than the Coulomb interaction energy. Donor−acceptor interactions involve transfer of electron density from one partner to the other (although this is impossible to define uniquely), are of intermediate strength, and fall off with the overlap between the interacting orbitals. London dispersion is an instantaneous interaction between induced multipoles and is both weaker and of shorter range than the other interactions but has been shown recently[14] to be comparable to H-bonds for interactions between aromatic systems with delocalized electrons and also becomes increasingly important for large systems. In the other direction, the repulsive force arises mainly from the exchange-repulsion energy, which becomes dominant when the electron-density clouds of two subsystems overlap significantly.

With the exception of dispersion, the above interactions are predominantly local in nature. Specific sites on one subsystem are often considered to interact with complementary ones on the other. Perhaps the most prominent example is the interaction between H-bond donors and acceptors. This locality of interaction sites forms the basis of much of current computer-aided drug design (CADD). The seminal LUDI scoring function[15] and most of its descendants,[16] Vinter's XED technique[17−19] and the concept of pharmacophores[20] all rely on local interaction sites near the surfaces of molecules. However, the choice of interaction sites has largely been based on experience or features extracted from X-ray structures by searching for

Received: February 17, 2012
Published: March 29, 2012

specific combinations of known binding features. This can mean that "unusual" interactions, such as hydrogen bonds to aromatic rings[8,21,22] or halogen bonding,[4−6] are not yet considered by the classical techniques. We have therefore set out to define a systematic protocol for detecting interaction sites of different types in the vicinity of ligands or receptors in order to provide a more consistent and complete picture of the intermolecular binding properties of small molecules and biopolymers. An ultimate goal is clearly to develop algorithms that can describe the binding features of molecules automatically for predictive drug design. Since the points of interaction of molecules lie at or near the molecular surface, surface-based molecular descriptors can be expected to describe them well,[23] as can grid-based techniques such as CoMFA[24] or GRID.[25−29]

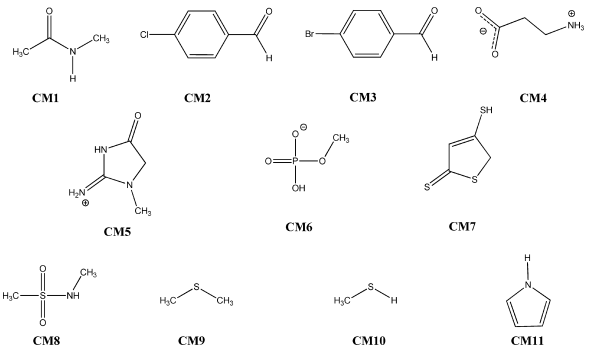
We have used the local molecular properties calculated by ParaSurf'10[30] based on semiempirical molecular-orbital (MO) theory[31] as the basis for our approach.[32,33] These local properties are the molecular electrostatic potential (MEP),[34,35] which describes the Coulomb interaction, the local ionization energy (IE_L),[36−39] and the local electron affinity (EA_L).[33,40] These properties describe the charge-transfer (donor−acceptor) component of the noncovalent interaction, so that to complete the full picture two further properties are necessary. The first is the local polarizability (α_L),[33,41−44] which is a density-weighted property that describes the polarizable nature of the electron density and is therefore relevant for describing the London dispersion interaction, the electrostatic induction and π-stacking.[32,33] Note that charge transfer and polarizability are not separable and that three properties, the local ionization energy, electron affinity, and polarizability, all help describe the effects of what we conventionally understand as charge transfer and polarization. The final property is the electron density, which describes the steric extent of the molecule and therefore also the steric repulsion.[32] The base 10 logarithm of the electron density was used in this work for numerical convenience. This suite of local molecular properties covers the components of noncovalent interactions outlined above. The magnitude and the direction of the gradient (first derivative) for each of these five properties at the point of measurement were also used as descriptors in the models, where such descriptors express the extent to which the local properties are changing locally and so add nonlocal information (analogous to DFT nonlocal terms). Our aim in this work is to construct models for the local energies that use the local molecular properties and the magnitude and direction of their gradients to predict the noncovalent interaction energies at points in the vicinity of the molecule. Noncovalent interactions are distance-dependent; the energies of the interaction are inversely proportional to some power of the distance separating the interacting groups.[45] Thus, we also used the distance between the point being considered and the nearest atom in the molecule as an additional descriptor in order to fine-tune the performance of the models.

METHODS

The training data were calculated by considering the interactions of eight different probe molecules on grids of points around 11 "central" molecules chosen as outlined below. Ideally, we would like to have calculated these interaction energies at definitive levels of theory, but the large number of data points makes this impossible. We have therefore selected two convenient levels of density-functional theory (DFT) to reduce the computational task to a manageable level.

Preparation of the Central and Probe Molecules and Calculating the Descriptors. Eleven central molecules (CM) were selected that are not only rich in chemical features but also chemically diverse so that they represent diverse binding features and thus cover as broad a range of descriptor combinations as possible to give a robust and predictive models (Chart 1). The 11

Chart 1. Central Molecules



molecules are *N*-methylacetamide (**CM1**), *p*-chlorobenzaldehyde (**CM2**), *p*-bromobenzaldehyde (**CM3**), β-alanine zwitterion (**CM4**), creatinine (2-amino-1-methyl-5*H*-imidazol-4-one) cation (**CM5**), methyl hydrogen phosphate anion (**CM6**), 4-mercaptothiophene-2(5*H*)-thione (**CM7**), *N*-methylmethanesulfonamide (**CM8**), dimethylsulfide (**CM9**), methanethiol (**CM10**) and pyrrol (**CM11**).

Initial 3D structures were generated from SMILES strings using CORINA.[46] The geometries of the central molecules were optimized fully with Gaussian09[47] using the B97-D[48] and ωB97X-D[49] functionals, which include an empirical correction for dispersion. These two functionals were chosen because of their reported ability to describe the geometries and energies of the noncovalent complexes with reasonable accuracy at a modest computational cost.[11,48,49] The aug-cc-pVDZ basis set[50−53] was used throughout. Semiempirical MO-calculations with the AM1 Hamiltonian were performed with VAMP10.0.[54] Note, however, that any other common MNDO-like Hamiltonian could have been used. The output of VAMP was used as input for ParaSurf'10,[30] which generated a grid around the molecule and calculated the local properties at each point. All grid points that were inside the vdW volume of the central molecules were removed. Details of the grid generation are given in the Supporting Information. The gradient vector for each local property at each grid point was calculated, and the cosine of the angle between the gradient vector and that connecting the point to the nearest atom in the central molecule ($\cos\theta$) was also calculated to describe the direction of the gradient vector. This procedure gave a set of 16 descriptors at each grid point; the MEP, local ionization energy, local electron affinity, local polarizability, electron density, the magnitudes and the directions ($\cos\theta$) of their gradients, and the distance described below. (Full details of the calculational methods are given in the Supporting Information S1.)

Eight different probe molecules (**Pr**) were selected that can interact noncovalently with the selected central molecules in diverse ways. These are water $\mathbf{H_2O}$ as H-donor and acceptor, ammonia $\mathbf{NH_3}$ as H-donor and acceptor, the ammonium cation $\mathbf{NH_4^+}$, the hydroxyl anion $\mathbf{OH^-}$, formamide as H-donor $\mathbf{HCONH_2}$, and as H-acceptor $\mathbf{HCONH_2}$. The probes were generated and optimized analogously to the central molecules.

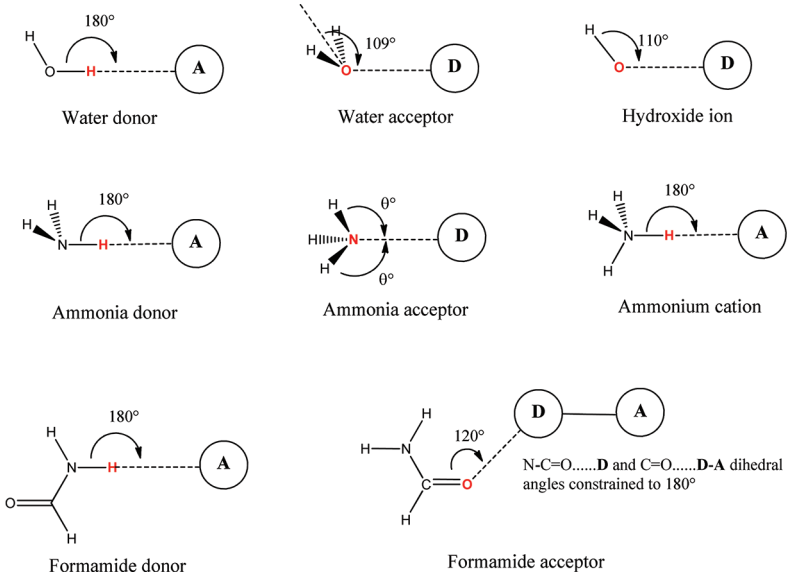
dx.doi.org/10.1021/ci300095x | J. Chem. Inf. Model. 2012, 52, 1061−1071

Table 1. Probe Geometrical Orientation and Constraints Used

probe	interacting atom	probe type	geometrical constraints	reason for constraints
H_2O	H	H-donor	O−H−NA[a] angle = 180°	to place the donor hydrogen atom in the optimal position for H-bonding
H_2O	O	H-acceptor	H−O−NA[a] angle = 109°	to put the acceptor oxygen in the optimal position for H-bonding
NH_3	H	H-donor	N−H−NA[a] angle = 180°	to place the donor hydrogen atom in the optimal position for H-bonding
NH_3	N	H-acceptor	All NA[a]−N−H angles kept equal	to avoid the NH_3 orientation flipping
$HCONH_2$	NH	H-donor	N−H−NA[a] angle = 180°	to place the donor hydrogen atom in the optimal position for H-bonding
$HCONH_2$	O	H-acceptor	(1) C−O−NA[a] angle = 120° (2) C−O−NA[a]−NN[b] and N−C−O−NA[a] dihedral angles = 180°	(1) to put the acceptor oxygen in the optimal position for H-bonding (2) to avoid extra-interactions of the three formamide hydrogens with the central molecule
NH_4^+	H	cationic H-donor	N−H−NA[a] angle = 180°	to place the donor hydrogen atom in the optimal position for H-bonding
OH^-	O	anionic H-acceptor	H−O−NA[a] angle = 110°	to put the acceptor oxygen in the optimal position for H-bonding

[a]NA is the nearest atom of the central molecule to the probe atom on the grid point (the distance is calculated from the grid point to the vdW surface of the central molecule atom). [b]NN is the second nearest atom of the central molecule (this atom is selected arbitrarily to avoid extra interactions between the probe atoms and the central molecule).

Chart 2. Interacting Atoms (Red) and Geometrical Constraints Used for the Probe Molecules



Positioning the Probes on the Grid Points. Each probe was placed on each grid point so that the interacting atom of the probe was fixed on the grid point and the remaining probe atoms were only subject to the defined geometrical constratints. These constraints prevent additional alternative interactions between the probe and the central molecule. The central molecules are kept frozen at their previous fully optimized structures in order to maintain the unique definition of the grid points. Table 1 and Chart 2 show the probe interacting atom and the geometrical constraints used for each probe.

Interaction Energy Calculations. The geometries of the probe molecules constrained on the grid points as described above were optimized at the same DFT levels of theory as used for the central and probe molecules themselves (detailed discussion of this point in Supporting Information S2). The aug−cc−pVDZ basis set[50−53] is not large enough to avoid basis set superposition error (BSSE),[55,56] so that in principle either a larger basis set or a correction for BSSE should be used. The most common BSSE correction uses the counterpoise method proposed by Boys and Bernardi.[57] However, the counterpoise correction and its benefit in calculating interaction energies are controversial. It has recently[58] been shown that the counterpoise correction systematically leads to interaction energies that differ from the CBS-extrapolated (BSSE free) values more strongly than the uncorrected ones. This effect is most significant if only weak noncovalent interactions are analyzed. We have therefore used the uncorrected interaction energies calculated with the extended aug−cc−pVDZ basis set.[50−53]

All grid points that showed steric clashes between the probe and the central molecule were removed, because of their high repulsion energies. Then, the interaction energy was calculated as the difference between the total energy of the geometrically optimized complex (**CM** + **Pr**) and the energies of the individual components **CM** and **Pr**. No zero-point vibrational energy corrections were used (i.e., the calculated and modeled interaction energies are Born−Oppenheimer).

Model Generation. Because the relationship between the 16 descriptors and the calculated interaction energies is likely to

dx.doi.org/10.1021/ci300095x | J. Chem. Inf. Model. 2012, 52, 1061−1071

be very nonlinear (and to involve IF/OR decisions), we used a feed-forward artificial neural network (ANN) to construct the models.[59−61] ANNs have been used widely in computational chemistry and molecular modeling to build models for activity or property qualitative or quantitative prediction.[62−67]

We used three-layer feed-forward neural networks with a sigmoid transfer function. The nets were trained using the back-propagation of errors algorithm (multilayer perceptron) with the RapidMiner 5.0.008 software.[68,69]

We used a 10-fold cross-validation strategy for the ANN models. For each model, the data set was split into ten equal subsets. For each ANN, nine different subsets were selected as training set and the tenth used as the test set, yielding ten training sets of 90% of the points, and their corresponding test sets consisting of 10% of the points. Training and test sets were chosen such that each point appears in the test set for only one network. Then, ten separate ANNs with random starting weights were trained for 1000 cycles, training with different initial weights was repeated several times for each net (data combination), and the net that gave the minimum training RMSE was chosen. The cross-validated result for any given point is then the prediction by the net where it appeared in the test set. This should therefore give a worst case prediction for the given point in this model. Different numbers of hidden nodes (between 10 and 20 nodes) were used for each model generation, and the model generated from the number of hidden nodes which gave the least overall cross-validation RMSE was chosen.

DISCUSSION AND RESULTS

Initial investigations showed that the five local molecular properties described above plus the local hardness and the local electronegativity[33] can be used to classify the interaction energy into binding and nonbinding for some simple systems using decision trees.[70] The next step was therefore to use the same descriptors but with carefully chosen molecules that can represent the different types of noncovalent interactions, including the most important drug−protein binding features. We also extended the model from simple classification to a quantitative model for the interaction energy. However, initial trials with a linear technique, multiple linear regression (MLR)[71] (procedures in Supporting Information S4) gave unsatisfactory results. For example, a water-donor probe model gave only R^2 = 0.26, RMSE = 2.45 kcal mol^{-1}, cross-validated R^2 = 0.25, and cross-validated RMSE = 2.46 kcal mol^{-1}. Thus, the relationship between the descriptors and the interaction energy (if there is one) is not linear. Using the same descriptors for an ANN gave improved results: (water-donor probe model) R^2 = 0.64, RMSE = 1.71 kcal mol^{-1}, cross-validated R^2 = 0.51, cross-validated RMSE = 2.04 kcal mol^{-1}. However, this model gave several strong outliers because of secondary steric clashes between the CM and noninteracting atoms of the probe. All points that showed such secondary steric clashes were removed (see Supporting Information S5 for details). Removing these points improved the performance of the models significantly: (water-donor probe model performance) R^2 = 0.88, RMSE = 0.53 kcal mol^{-1}, cross-validated R^2 = 0.85, cross-validated RMSE = 0.59 kcal mol^{-1}. However, these models could not distinguish between attraction and repulsion at some points because of some residual secondary interactions. We therefore included the distance between the grid point and the nearest atom of the CM as an additional descriptor to provide more information about steric interactions not covered by the local electron density. This once again improved the performance of the models: (water-donor probe model performance) R^2 = 0.93, RMSE = 0.41 kcal mol^{-1}, cross-validated R^2 = 0.92, cross-validated RMSE = 0.43 kcal mol^{-1}. However, because multisite models (e.g., using properties at both the O and H positions of an OH H-bond donor) gave better results, we included the gradient of each property and its direction to add nonlocal information to the single-site models. This gave further improved performance: (water-donor probe model performance) R^2 = 0.97, RMSE = 0.28 kcal mol^{-1}, cross-validated R^2 = 0.96, cross-validated RMSE = 0.30 kcal mol^{-1}. At this stage, we simplified the models to make overtraining less likely and the models more robust by removing the local electronegativity and local hardness from the descriptor pool because these two descriptors are linear combinations of the local ionization energy and the local electron affinity. This did not degrade the performance of the models: (water-donor probe model performance) R^2 = 0.96, RMSE = 0.28 kcal mol^{-1}, cross-validated R^2 = 0.96, cross-validated RMSE = 0.30 kcal mol^{-1}.

The final set of descriptors used in the generation of the models is shown in Table 2, and the statistical performance of the models is shown in Table 3.

Table 2. Descriptors Used for Generating Models

no.	symbol	description
1	MEP	molecular electrostatic potential at the grid point
2	dMEP	gradient magnitude of the MEP at the grid point
3	cos θMEP	gradient direction of the MEP at the grid point
4	EA_L	local electron affinity at the grid point
5	dEA	gradient magnitude of the EA at the grid point
6	cos θEA	gradient direction of the EA at the grid point
7	IE_L	local ionization energy at the grid point
8	dIE	gradient magnitude of the IE at the grid point
9	cos θIE	gradient direction of the IE at the grid point
10	α_L	local polarizability at the grid point
11	dα	gradient magnitude of the α at the grid point
12	cos $\theta\alpha$	gradient direction of the α at the grid point
13	log ρ	electron density common logarithm log D at the grid point
14	d log ρ	gradient magnitude of the log D at the grid point
15	cos θ log ρ	gradient direction of the log D at the grid point
16	R_{min}	distance between the grid point and the NA in the CM

Table 3 shows that 16 ANN models were generated (one for each combination of the eight probes and the two DFT functionals). The models reproduce the interaction energy well. They are unlikely to be overtrained because the number of points used (the minimum is 17 226) in each data set is far larger than the number of descriptors (16) and the total number of nodes (the maximum is 34 nodes). The closeness of RMSE and R^2 found for the training and cross-validation data also indicates that the models are robust.

The models have cross-validated R^2-values between 0.937 and 0.997 for B97-D and between 0.942 and 0.998 for ωB97X-D. Plots of the DFT vs model calculated interaction energies are shown for all the models in Supporting Information S6. The performance of the models depends on the characteristics of the interaction and probes. The MSEs are higher for H-bond acceptors than for donors and also higher for ionic than neutral probes. These values correlate with the range of the interactions, so that the relative performance of the models is comparable, as suggested by the small range of R^2 values.

dx.doi.org/10.1021/ci300095x | J. Chem. Inf. Model. 2012, 52, 1061−1071

Table 3. Performance of the Models (Energies in Kilocalories per Mole)

probe	N^a	NH[b]	cross-validation R^2	RMSE	MUE	MSE	most +ve error	most −ve error
			B97-D/aug-cc-pVDZ					
H_2O H-donor	19232	10	0.961	0.30	0.18	0.01	3.4	−5.0
H-acceptor	17300	15	0.973	0.72	0.30	−0.02	14.3	−22.9
NH_3 H-donor	17808	10	0.937	0.26	0.14	−0.02	4.1	−5.1
H-acceptor	18772	18	0.977	0.81	0.36	0.05	17.0	−26.8
$HCONH_2$ H-donor	19118	16	0.988	0.27	0.16	−0.01	5.2	−8.3
H-acceptor	19008	13	0.968	1.12	0.50	−0.01	24.4	−28.0
$NH4^+$ H-donor	19267	17	0.997	1.31	0.83	0.01	12.6	−11.4
OH^- H-acceptor	19225	11	0.982	4.07	2.75	0.01	52.0	−101.5
			ωB97X-D/aug-cc-pVDZ					
H_2O H-donor	19124	13	0.961	0.31	0.18	−0.02	3.2	−5.9
H-acceptor	17226	14	0.967	0.84	0.37	−0.03	19.8	−22.5
NH_3 H-donor	17674	12	0.942	0.26	0.14	−0.02	4.6	−4.1
H-acceptor	18698	13	0.968	1.05	0.63	−0.06	17.6	−17.3
$HCONH_2$ H-donor	19061	14	0.987	0.29	0.18	−0.00	3.3	−3.7
H-acceptor	18903	18	0.966	1.12	0.51	0.07	27.8	−27.3
$NH4^+$ H-donor	19177	14	0.998	0.97	0.61	0.01	10.4	−10.1
OH^- H-acceptor	19188	17	0.994	2.18	1.25	−0.05	39.4	−41.2

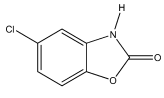
[a]N is the total number of grid points. [b]NH is the number of nodes in the hidden layer.

Table 4. Chlorzoxazone Results (Energies in Kilocalories per Mole)

probe	no. of points	R^2	RMSE	MUE	MSE	most +ve error	most −ve error
			B97-D/aug-cc-pVDZ				
H_2O H-donor	120	0.909	0.31	0.21	−0.18	0.26	−1.2
H-acceptor	119	0.992	1.02	0.43	−0.26	1.85	−8.37
NH_3 H-donor	118	0.880	0.28	0.20	−0.18	0.49	−1.11
H-acceptor	116	0.979	0.49	0.34	0.19	1.99	−1.92
$HCONH_2$ H-donor	121	0.918	0.37	0.25	−0.17	0.78	−2.42
H-acceptor	114	0.960	0.92	0.49	−0.14	2.55	−7.79
$NH4^+$ H-donor	117	0.882	1.42	0.95	−0.41	7.48	−4.41
OH^- H-acceptor	117	0.755	3.32	2.39	0.67	13.02	−13.82
			ωB97X-D/aug-cc-pVDZ				
H_2O H-donor	118	0.919	0.37	0.27	−0.23	0.63	−1.82
H-acceptor	115	0.918	0.74	0.53	0.35	1.76	−4.09
NH_3 H-donor	119	0.950	0.21	0.16	−0.13	0.42	−0.78
H-acceptor	117	0.907	1.07	0.74	0.25	3.37	−3.64
$HCONH_2$ H-donor	118	0.905	0.39	0.24	−0.21	0.28	−2.08
H-acceptor	116	0.807	1.01	0.52	0.16	4.58	−6.66
$NH4^+$ H-donor	124	0.879	1.76	1.11	−0.27	6.16	−6.99
OH^- H-acceptor	117	0.932	2.11	1.31	0.418	7.65	−11.95

The performance of the two models for the two DFT functionals are very similar except for OH^- as probe, which shows much worse performance for the B97-D functional than for ωB97X-D. This may indicate that the performance of B97-D is less consistent, and therefore harder to model, for this probe than that of ωB97X-D.

Model Validation. We selected the feature-rich drug molecule chlorzoxazone as a test for the models.



Chlorzoxazone

Interaction energies were calculated for the different probes using the two DFT functionals at 125 randomly selected grid points (5% of the total grid points) to represent the different environments around the molecule. The points that showed secondary steric clashes after optimization were removed, and the interaction energies at the remaining points were predicted for each probe. Table 4 and Figure 1 show the results of the different models prediction in relation to the calculated DFT interaction energies.

Table 4 shows the models to be robust as they perform comparably or better for chlorzoxazone than for the training data.

Having shown that the models can predict the interaction energies of the different probes with molecules not used in the training set satisfactorily, we move on to test the models for realistic challenges that might arise in a real drug design scenario.

Bioisosterism. The use of isosterism is common in drug design to improve the drug molecules' pharmacokinetic properties or to avoid patent limitations. Nonclassical isosterism (bioisosterism) involves two chemically diverse groups that have the same binding features to the biological target. Perhaps

dx.doi.org/10.1021/ci300095x | J. Chem. Inf. Model. 2012, 52, 1061−1071

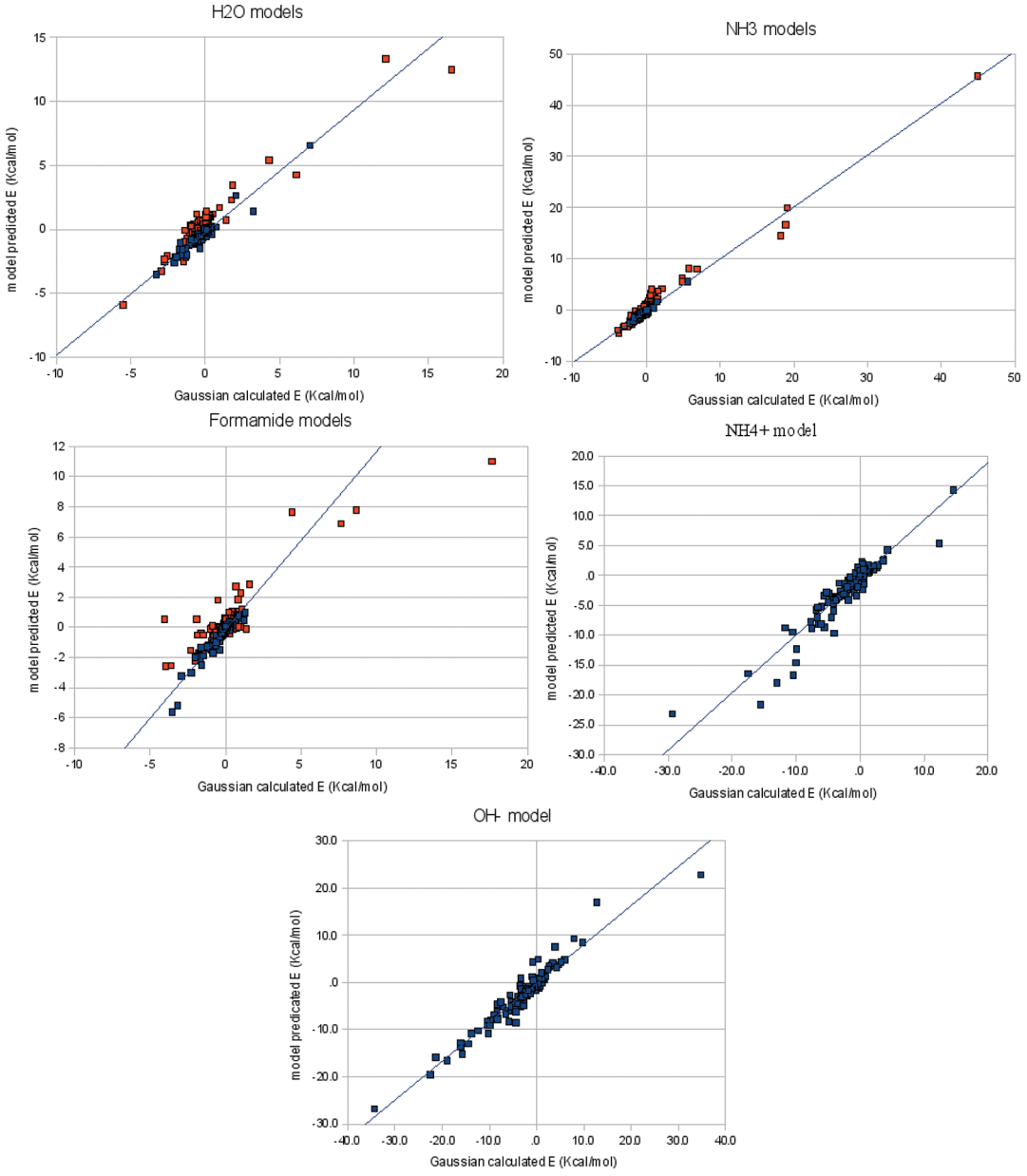


Figure 1. Chlorzoxazone calculated (DFT) vs model predicted interaction energies for the different probes using the ωB97X-D models. The H-bond donor and acceptor models are shown in blue and red, respectively; the line represents perfect agreement.

the best known example of bioisosterism is the relation between carboxylic acids and the corresponding tetrazole derivatives. Using the H_2O H-donor (ωB97X-D functional) model, interaction energy surfaces were constructed for the acetate anion and deprotonated 5-methyltetrazole (which are deprotonated at the physiological pH of 7.4) to test if the model is able to reproduce the near equivalence of the binding features of the two bioisosters.

Figure 2 shows that the −11 kcal mol^{-1} isosurfaces are very similar for the two pharmacophores, so that the model is clearly able to identify the major binding features despite the apparent chemical diversity of the two anions. The solid areas contoured at more negative binding energies (−16.5 kcal mol^{-1} for acetate and −12.5 for the tetrazole) reveal subtle differences. As expected, the acetate, in which the negative charge is formally concentrated on the two oxygen atoms, binds more strongly than the more delocalized tetrazole anion. The contours shown for the acetate are consistent with the known binding mode with water[72] and average interaction energy of −16.5 kcal mol^{-1} agrees well with the reported experimental H-bond energy between the acetate anion and one water molecule of −15.9 kcal mol^{-1}.[73] On the other hand, for the 5-mehyltetrazole anion, the strongest interaction energy positions are distributed around the tetrazole ring and correspond to possible H-bonds to all four nitrogen atoms. The average interaction energy of −12.5 kcal mol^{-1} is lower than that for the acetate, as expected because the charge is more delocalized in the tetrazole anion.

Influence of Different Substituents on an Aromatic H-Acceptor System. One of the major weaknesses of classical, force-field like approaches is that they generally cannot treat electronic effects such as resonance, the inductive effect or hyperconjugation. This deficiency is nowhere more obvious than for substituted aromatic systems, which can vary from electron-rich to electron-poor and also differ at individual ring positions. We therefore selected pyridine as an aromatic H-bond acceptor

dx.doi.org/10.1021/ci300095x | J. Chem. Inf. Model. 2012, 52, 1061−1071

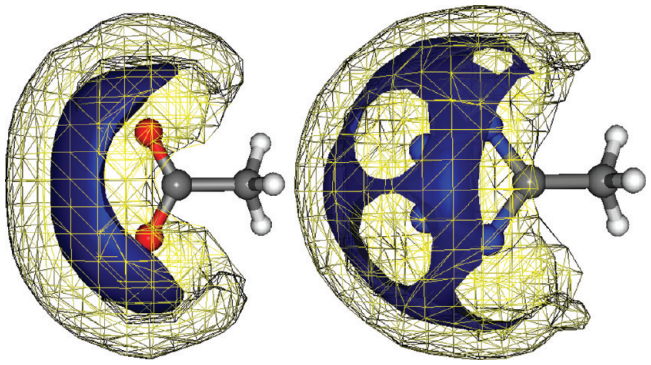


Figure 2. Interaction energy surface at −11 kcal mol^{-1} cutoff level (net) for the acetate anion and deprotonated 5-methyltetrazole. The solid surfaces are contoured at −16.5 and −12.5 kcal mol^{-1} for acetate and tetrazole, respectively.

system and have compared it with *p*-aminopyridine and *p*-chloropyridine as examples of compounds in which the *para*-amino group and the *para*-chlorine donate and accept electrons by the +M and -I effects, respectively. These effects should change the H-bond acceptor characteristics of the ring nitrogen significantly. We therefore used the H_2O H-donor (ωB97X-D functional) model to generate the interaction energy surface for each compound. These surfaces are shown in Figure 3. Both the modeled interaction energies and those calculated for the minimum-energy structures with ωB97X-D/aug-cc-pVDZ correlate with the observed pK_{HB} for these compounds.[74] The data are shown in Supporting Information S7.

Figure 3 clearly shows the ability of the model to take the substituent effects on the H-bond acceptor capability of the pyridine ring into account. This effect is evident in two ways. As expected, the H-bond acceptor strength of the ring nitrogen is

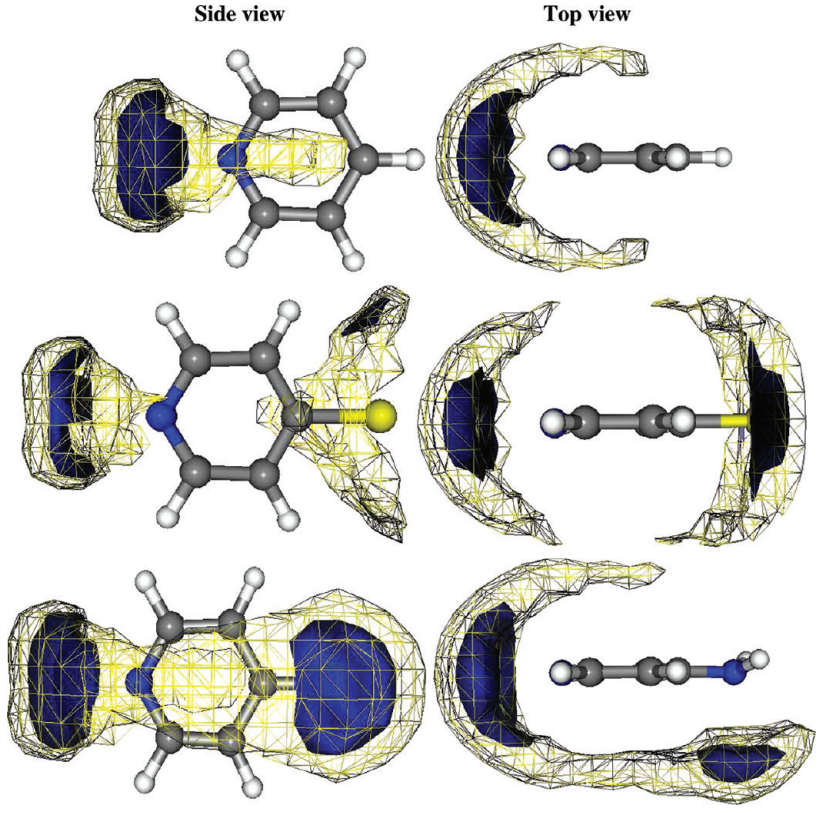


Figure 3. Interaction energy surfaces for (un)substituted pyridine at −2.8 (net) and −4.5 kcal mol^{-1} (solid dark blue) contour levels for pyridine, *p*-chloro pyridine, and *p*-amino pyridine (from top to bottom).

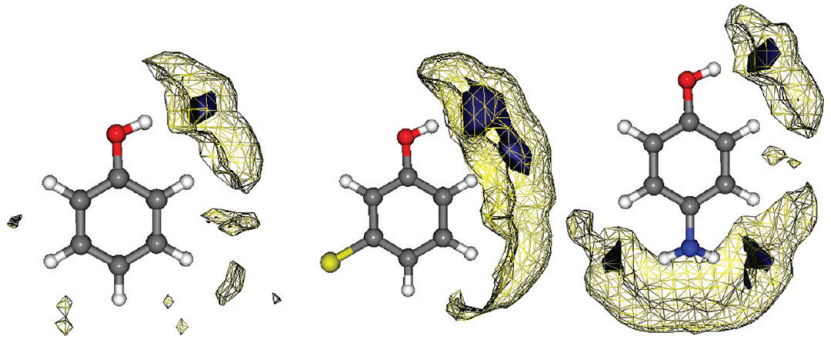


Figure 4. Interaction energy surfaces for (un)substituted phenol at −14.0 (net) and −20.0 kcal mol^{-1} (solid dark blue) contour levels for phenol, *m*-chlorophenol, and *p*-aminophenol.

dx.doi.org/10.1021/ci300095x | *J. Chem. Inf. Model.* 2012, 52, 1061−1071

affected by the substituent, both in the size of the −2.8 kcal mol^{-1} contour and in the depth of the minimum-energy (solid) contours. The maximum well depth for pyridine is −6.36 kcal mol^{-1}, in good agreement with the calculated pyridine-water H-bond energy[75,76] and the Gaussian calculated energy with the ωB97X-D functional according to the constraints used (−6.12 kcal mol^{-1}). As expected this value decreases (to −5.63 kcal mol^{-1}) because of the −I effect of the *p*-chloro substituent, in agreement with the ωB97X-D value of −5.67 kcal mol^{-1}, but remains essentially unchanged (at −6.47 kcal mol^{-1}) by the (purely π) +M effect of the *para*-amino group (ωB97X-D −6.99 kcal mol^{-1}).

The second effect is on the pyridine ring itself as H-bond acceptor. The *para*-amino substituent donates to the aromatic π-system, making the entire ring an H-bond acceptor at the −2.8 kcal mol^{-1} level, whereas the −I effect of the chloro substituent dominates, so that the ring is no longer an acceptor in *p*-chloro-pyridine.

Influence of Different Substituents on an Aromatic H-Donor System. CH-acceptor H-bonds can be surprisingly strong[77] and compete, for instance, with halogen bonds in chlorophenyl moieties. These interactions, which are often ignored in conventional pharmacophore approaches or scoring functions, are not only significant in determining total binding energies, but also depend on the electronic nature of the aromatic system. We have therefore investigated *p*-aminophenol and *m*-chlorophenol as aromatic CH donors. In contrast to the previous case, in *p*-aminophenol, the amino group is expected to decrease the H-bond donor capability of the phenol OH and the chlorine atom in the *m*-chlorophenol will increase its H-bond donor capability. We therefore used the OH^- (ωB97X-D functional) model to calculate the interaction energy surface for each compound, as shown in Figure 4.

Figure 4 demonstrates the ability of the models to reveal substituent effects on the H-bond donor strength of the phenol ring CH-bonds by the differences in the interaction-energy isocontour surfaces between the three compounds. As can be seen, the surface for phenol is intermediate between those of *m*-chlorophenol and *p*-aminophenol with maximum interaction energy of −25.6 kcal mol^{-1} around the OH hydrogen atom (solid) (ωB97X-D −39.1 kcal mol^{-1}, but stabilized by additional secondary interactions). At the −14 kcal mol^{-1} energy contour level, the model shows even more clearly that the ring hydrogens in meta and para positions can interact with the OH^-. *para*-Aminophenol shows the expected smaller interaction energy surface volume with average interaction energy of −24.5 kcal mol^{-1} around the OH hydrogen atom (solid) (ωB97X-D −30.4 kcal mol^{-1}, again with additional secondary interactions). On the other hand, *m*-chlorophenol shows a larger interaction energy surface volume with average interaction energy of −30.7 kcal mol^{-1} around the OH hydrogen atom (solid) (gaussian calculated −41.0 kcal mol^{-1}), reflecting its higher H-bond donor capability due to its lower electron density. Furthermore, the *m*-chlorophenol interaction energy surface at −14 kcal mol^{-1} also shows two important features that show the strength of the models; first, the large interaction energy surface associated with the two ring hydrogen atoms, meta and para to the chlorine atom only, but none for the two ortho hydrogens. This demonstrates the ability of the model to handle and describe the steric effect of the chlorine atom. Second, the appearance of a small interaction energy volume colinear with the carbon-chlorine bond at −14 kcal mol^{-1} cutoff represents the halogen bond interaction (model predicted energy −15.2 kcal mol^{-1}) between the chlorine atom and the OH^-.

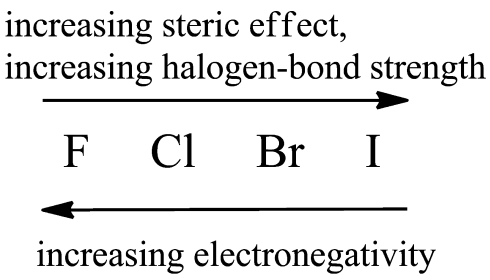


Figure 5. Characteristic properties of the halogens.

Halobenzenes. We chose the halobenzenes to test the models because they represent a 3-fold challenge in which the models must be able to differentiate between the different halogens in their inductive and steric effects and their halogen bond forming propensity. Figure 5 shows the difference between the halogens in these characteristics.

For this test, fluorobenzene, chlorobenzene, and bromobenzene were used with formamide as the H-acceptor (ωB97X-D functional) model. Figure 6 shows the calculated interaction-energy surfaces for the three compounds at −0.7 kcal mol^{-1} cutoff. As expected, fluorobenzene showed no halogen bond interaction because it does not normally act as a halogen donor.[4] In contrast, chlorobenzene shows an intermediate halogen bond interaction in front of the chlorine atom and collinear with the C−Cl axis with a maximum interaction energy of −0.8 kcal mol^{-1} (ωB97X-D −0.3 kcal mol^{-1}), and bromobenzene shows a larger halogen-bond interaction surface volume in front of the equivalent position, indicating a higher propensity to form halogen bonds (more positive σ-hole) with maximum interaction energy of −4.8 kcal mol^{-1} (ωB97X-D −1.0 kcal mol^{-1}). Figure 6 also shows that the models can consider the difference in the steric effect of the halogen atoms, as can be seen from the differences in the interaction energy surface around the ortho hydrogen atoms between the three compounds. Finally, the third characteristic difference that the models have proven to be able to detect is the differing inductive effect of the three halogens atoms. The maximum interaction

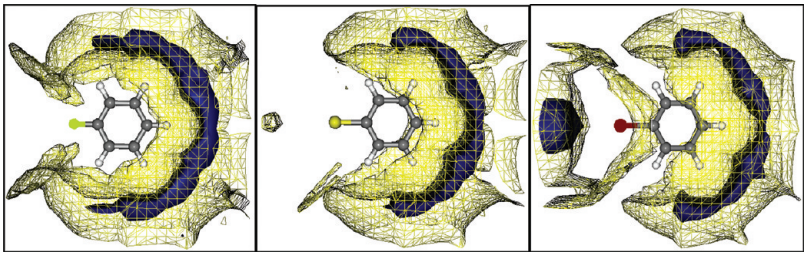


Figure 6. Interaction energy surfaces for halobenzenes at −0.7 (net) and −2.0 kcal mol^{-1} (solid dark blue) contour levels.

dx.doi.org/10.1021/ci300095x | J. Chem. Inf. Model. 2012, 52, 1061−1071

energies predicted around the aromatic ring hydrogen atoms for fluoro-, chloro-, and bromobenzene are −3.0, −3.0, and −2.9 kcal mol^{-1}, respectively, in correspondence to the different halogens' electronegativity (ωB97X-D −2.3, −2.6, and −2.2 kcal mol^{-1}).

These different validation tests show that the models are able to predict the interaction energy between different molecules and the probes satisfactorily and that they can also handle and describe steric interactions, which are an important component of noncovalent interaction that is not described well by most current methods. The models can also predict unusual interactions such as halogen bonds, H-bond involving π-systems as H-acceptors and H-bonds involving C−H as the H-donor. They can also reproduce the effects of small changes in the molecular environment on the interaction energy.

We note that the current models are based on gas-phase Born−Oppenheimer interaction energies, so that they are unlikely to be useful for scoring biological intermolecular interactions. However, they have already proven to be very powerful in detecting local sites of attractive intermolecular interactions, making them ideal for, for instance, pharmacophore and docking applications.

CONCLUSIONS

The technique presented above provides a promising alternative to classical binding-site recognition algorithms for characterizing the specific binding sites around ligands or in the binding sites of proteins. It has the advantage that it does not rely on preconceived ideas of classical binding interactions but attempts to recognize binding sites from their electronic properties. The major disadvantage is that it predicts gas phase interaction energies, so that it requires an accurate solvation modeling technique if it is to be used as a component of scoring functions. Its most obvious application at the moment probably centers on ligand-based (pharmacophore, pattern recognition, or complementarity) activity prediction.

ASSOCIATED CONTENT

Supporting Information

Additional details of the calculations, plots of the results for models not shown in the paper, and a table of experimental pK_{BHX} values and the corresponding calculated interaction energies (19 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Tim.Clark@chemie.uni-erlangen.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Cepos InSilico GmbH and by the grant of a Fellowship to AEK by the *Deutscher Akademischer Austauschdienst*.

REFERENCES

(1) Scheiner, S. *Hydrogen bonding: A theoretical prospective*; Oxford University Press, Inc.: New York, USA, 1997.
(2) Jeffrey, G. A. *An introduction to hydrogen bonding*; Oxford University press, Inc.: New York, USA, 1997.
(3) Desiraju, G.; Steiner, T. *The Weak Hydrogen Bond: In Structural Chemistry and Biology*; Oxford University Press, Inc.: New York, USA, 2001.
(4) Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. An overview of halogen bonding. *J. Mol. Model.* **2007**, *13* (2), 305−311.
(5) Politzer, P.; Murray, J. S.; Clark, T. Halogen bonding: an electrostatically-driven highly directional noncovalent interaction. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7748−7757.
(6) Murray, J. S.; Riley, K. E.; Politzer, P.; Clark, T. Directional Weak Intermolecular Interactions: sigma-Hole Bonding. *Aust. J. Chem.* **2010**, *63* (12), 1598−1607.
(7) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books: Sausalito, CA, 2006.
(8) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem., Int. Ed.* **2003**, *42* (11), 1210−1250.
(9) Ma, J. C.; Dougherty, D. A. The cation-pi interaction. *Chem. Rev.* **1997**, *97* (5), 1303−1324.
(10) Hobza, P.; Zahradnik, R.; Muller-Dethlefs, K. The world of non-covalent interactions: 2006. *Collect. Czech. Chem. Commun.* **2006**, *71* (4), 443−531.
(11) Cerny, J.; Hobza, P. Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* **2007**, *9* (39), 5291−5303.
(12) Gkionis, K.; Hill, J. G.; Oldfield, S. P.; Platts, J. A. Performance of Becke's half-and-half functional for non-covalent interactions: energetics, geometries and electron densities. *J. Mol. Model.* **2009**, *15* (9), 1051−1060.
(13) Chen, K. Kurgan, L. Investigation of Atomic Level Patterns in Protein-Small Ligand Interactions. *PLoS One* **2009**, *4*, 2.
(14) Muller-Dethlefs, K.; Hobza, P. Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.* **2000**, *100* (1), 143−167.
(15) Böhm, H. J. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein Ligand Complex of Known 3-Dimensional Structure. *J. Comput.-Aided Mol. Des.* **1994**, *8* (3), 243−256.
(16) Gohlke, H.; Klebe, G. Statistical potentials and scoring functions applied to protein-ligand binding. *Curr. Opin. Struct. Biol.* **2001**, *11* (2), 231−235.
(17) Vinter, J. G. Extended Electron Distributions Applied to the Molecular Mechanics of Some Intermolecular Interactions. *J. Comput.-Aided Mol. Des.* **1994**, *8* (6), 653−668.
(18) Vinter, J. G. Extended electron distributions applied to the molecular mechanics of some intermolecular interactions 0.2. Organic complexes. *J. Comput.-Aided Mol. Des.* **1996**, *10* (5), 417−426.
(19) Cheeseright, T.; Mackey, M.; Rose, S.; Vinter, A. Molecular field extrema as descriptors of biological activity: Definition and validation. *J. Chem. Inf. Model.* **2006**, *46* (2), 665−676.
(20) Güner, O. F. *Pharmacophore Perception, Development, and Use in Drug Design*, 1st ed.; International University Line: La Jolla, CA, USA, 2000.
(21) Levitt, M.; Perutz, M. F. Aromatic Rings Act as Hydrogen-Bond Acceptors. *J. Mol. Biol.* **1988**, *201* (4), 751−754.
(22) Brinkley, R. L.; Gupta, R. B. Hydrogen bonding with aromatic rings. *AIChE J.* **2001**, *47* (4), 948−953.
(23) Clark, T.; Byler, K. G.; De Groot, M. J. Biological communication via molecular surfaces. In *International Beilstein Workshop*, Bozen, Italy, May 15−19, 2006.
(24) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. Comparative Molecular-Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959−5967.
(25) Goodford, P. J. A Computational-Procedure for Determining Energetically Favorable Binding-Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28* (7), 849−857.
(26) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen-Bond Functions for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known Structure 0.1. Ligand Probe Groups with the Ability to Form 2 Hydrogen-Bonds. *J. Med. Chem.* **1993**, *36* (1), 140−147.
(27) Wade, R. C.; Goodford, P. J. Further Development of Hydrogen-Bond Functions for Use in Determining Energetically

Favorable Binding-Sites on Molecules of Known Structure 0.2. Ligand Probe Groups with the Ability to Form More Than 2 Hydrogen-Bonds. *J. Med. Chem.* **1993**, *36* (1), 148−156.
(28) Boobbyer, D. N. A.; Goodford, P. J.; Mcwhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding-Sites on Molecules of Known Structure. *J. Med. Chem.* **1989**, *32* (5), 1083−1094.
(29) Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen bonding interactions of covalently bonded fluorine atoms: From crystallographic data to a new angular function in the GRID force field. *J. Med. Chem.* **2004**, *47* (21), 5114−5125.
(30) Clark, T.; Lin, J.; Horn, A. H. C. *Parasurf10*; CEPOS Insilico Ltd.: Kempston, Bedford, UK, 2010; http://www.ceposinsilico.de/products/parasurf.htm.
(31) Clark, T.; Stewart, J. J. P. MNDO-like semiempirical molecular orbital theory and its application to large systems. In *Computational methods for large systems: Electronic structure approaches for biotechnology and nanotechnology*, 1st ed.; Reimers, J. R., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2010; pp 259−286.
(32) Clark, T. QSAR and QSPR based solely on surface properties? *J. Mol. Graphics Modell.* **2004**, *22* (6), 519−525.
(33) Ehresmann, B.; Martin, B.; Horn, A. H. C.; Clark, T. Local molecular properties and their use in predicting reactivity. *J. Mol. Model.* **2003**, *9* (5), 342−347.
(34) Murray, J. S.; Politzer, P. Statistical analysis of the molecular surface electrostatic potential: an approach to describing noncovalent interactions in condensed phases. *J. Mol. Struct.: THEOCHEM* **1998**, *425* (1−2), 107−114.
(35) Murray, J. S.; Lane, P.; Brinck, T.; Paulsen, K.; Grice, M. E.; Politzer, P. Relationships of Critical Constants and Boiling Points to Computed Molecular-Surface Properties. *J. Phys. Chem.* **1993**, *97* (37), 9369−9373.
(36) Sjoberg, 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* (8), 1440−1443.
(37) Politzer, P.; Murray, J. S.; Concha, M. C. The complementary roles of molecular surface electrostatic Potentials and average local ionization energies with respect to electrophilic processes. *Int. J. Quantum Chem.* **2002**, *88* (1), 19−27.
(38) Hussein, W.; Walker, C. G.; Peralta-Inga, Z.; Murray, J. S. Computed electrostatic potentials and average local ionization energies on the molecular surfaces of some tetracyclines. *Int. J. Quantum Chem.* **2001**, *82* (4), 160−169.
(39) Murray, J. S.; Abu-Awwad, F.; Politzer, P. Characterization of aromatic hydrocarbons by means of average local ionization energies on their molecular surfaces. *J. Mol. Struct.: THEOCHEM* **2000**, *501*, 241−250.
(40) Clark, T. The local electron affinity for non-minimal basis sets. *J. Mol. Model.* **2010**, *16* (7), 1231−1238.
(41) Schurer, G.; Gedeck, P.; Gottschalk, M.; Clark, T. Accurate parametrized variational calculations of the molecular electronic polarizability by NDDO-based methods. *Int. J. Quantum Chem.* **1999**, *75* (1), 17−31.
(42) Rinaldi, D.; Rivail, J. L. Molecular Polarizability and Dielectric Effect of Medium in Liquid-Phase - Theoretical Study of Water Molecule and Its Dimers. *Theor. Chim. Acta* **1973**, *32* (1), 57−70.
(43) Rinaldi, D.; Rivail, J. L. Calculation of Molecular Electronic Polarizabilities - Comparison of Different Methods. *Theor. Chim. Acta* **1974**, *32* (3), 243−251.
(44) Martin, B.; Gedeck, P.; Clark, T. Additive NDDO-based atomic polarizability model. *Int. J. Quantum Chem.* **2000**, *77* (1), 473−497.
(45) Tsai, C. S. *Biomacromolecules: Introduction to structure, function and informatics*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007; p 5−8.
(46) Sadowski, j.; Schwab, C.; Gasteiger, J. *Corina 3.4*, Molecular Networks GmbH: Erlangen, Germany.
(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02, Gaussian, Inc.: Wallingford CT, 2009.
(48) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27* (15), 1787−1799.
(49) Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615−6620.
(50) Dunning, T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations 0.1. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007−1023.
(51) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron-Affinities of the 1st-Row Atoms Revisited - Systematic Basis-Sets and Wave-Functions. *J. Chem. Phys.* **1992**, *96* (9), 6796−6806.
(52) Woon, D. E.; Dunning, T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations. 3. The Atoms Aluminum through Argon. *J. Chem. Phys.* **1993**, *98* (2), 1358−1371.
(53) Davidson, E. R. Comment on Dunning's correlation-consistent basis sets - Comment. *Chem. Phys. Lett.* **1996**, *260* (3−4), 514−518.
(54) Clark, T.; Alex, A.; Beck, B.; Burckhardt, F.; Chandrasekhar, J.; Gedeck, P.; Horn, A.; Hutter, M.; Martin, B.; Rauhut, G.; Sauer, W.; Schindler, T.; Steinke, T. *VAMP 10.0*; Accelrys Inc.: San Diego, CA, USA, Erlangen, Germany, 2007.
(55) Kestner, N. R. He-He Interaction in the SCF-MO Approximation. *J. Chem. Phys.* **1968**, *48* (1), 252−257.
(56) Liu, B.; Mclean, A. D. Accurate Calculation of Attractive Interaction of Two Ground-State Helium-Atoms. *J. Chem. Phys.* **1973**, *59* (8), 4557−4558.
(57) Boys, S. F.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19* (4), 553−566.
(58) Alvarez-Idaboy, J. R.; Galano, A. Counterpoise corrected interaction energies are not systematically better than uncorrected ones: comparison with CCSD(T) CBS extrapolated values. *Theor. Chem. Acc.* **2010**, *126* (1−2), 75−85.
(59) Müller, B.; Reinhardt, J.; Strickland, M. T. *Neural networks-An introduction*, 2nd updated and corr. ed.; Springer-Verlag: Berlin, Heidelberg, Germany, 2002.
(60) Pao, Y.-H. *Adaptive pattern recognition and neural networks*; Addison-wesely Publishing Co.: Reading, MA, USA, 1989.
(61) Zupan, J.; Gasteiger, J. *Neural networks for chemists: An introduction*; VCH Verlag: Weinheim, Germany, 1993.
(62) Ehresmann, B.; de Groot, M. J.; Alex, A.; Clark, T. New molecular descriptors based on local properties at the molecular surface and a boiling-point model derived from them. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (2), 658−668.
(63) Beck, B.; Breindl, A.; Clark, T. QM/NN QSPR models with error estimation: Vapor pressure and LogP. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (4), 1046−1051.
(64) Chalk, A. J.; Beck, B.; Clark, T. A quantum mechanical/neural net model for boiling points with error estimation. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (2), 457−462.
(65) Göller, A. H.; Hennemann, M.; Keldenich, J.; Clark, T. In silico prediction of buffer solubility based on quantum-mechanical and HQSAR- and topology-based descriptors. *J. Chem. Inf. Model.* **2006**, *46* (2), 648−658.

(66) Garg, P.; Verma, J. In silico prediction of blood brain barrier permeability: An artificial neural network model. *J. Chem. Inf. Model.* **2006**, *46* (1), 289−297.
(67) Talevi, A.; Goodarzi, M.; Ortiz, E. V.; Duchowicz, P. R.; Bellera, C. L.; Pesce, G.; Castro, E. A.; Bruno-Blanch, L. E. Prediction of drug intestinal absorption by new linear and non-linear QSPR. *Eur. J. Med. Chem.* **2011**, *46* (1), 218−228.
(68) *RapidMiner 5.0.008*; Rapid- l GmbH: Dortmund, Germany, 2010.
(69) Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T., YALE: rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*; ACM: Philadelphia, PA, USA, 2006; pp 935−940.
(70) Wick, R. C. Quantenmechanische charakterisierung der wasserstoffbrückenbindung: Wasserstoffbrückendonoren. Bachelor Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, 2009.
(71) Weisberg, S. *Applied Linear Regression*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005.
(72) Görbitz, C. H.; Etter, M. C. Hydrogen-Bonds to Carboxylate Groups - the Question of 3-Center Interactions. *J. Chem. Soc., Perkin Trans. 2* **1992**, No. 1, 131−135.
(73) Mautner, M. M. N.; Elmore, D. E.; Scheiner, S. Ionic hydrogen bond effects on the acidities, basicities, solvation, solvent bridging, and self-assembly of carboxylic groups. *J. Am. Chem. Soc.* **1999**, *121* (33), 7625−7635.
(74) Hennemann, M.; Clark, T. A QSPR-approach to the estimation of the pK(HB) of six-membered nitrogen-heterocycles using quantum mechanically derived descriptors. *J. Mol. Model.* **2002**, *8* (4), 95−101.
(75) Dkhissi, A.; Adamowicz, L.; Maes, G. Density functional theory study of the hydrogen-bonded pyridine-H2O complex: A comparison with RHF and MP2 methods and with experimental data. *J. Phys. Chem. A* **2000**, *104* (10), 2112−2119.
(76) Pápai, I.; Jancsó, G. Hydrogen bonding in methyl-substituted pyridine-water complexes: A theoretical study. *J. Phys.Chem. A* **2000**, *104* (10), 2132−2137.
(77) Scheiner, S.; Kar, T.; Gu, Y. L. Strength of the $C^{\alpha}H..O$ hydrogen bond of amino acid residues. *J. Biol. Chem.* **2001**, *276* (13), 9832−9837.