

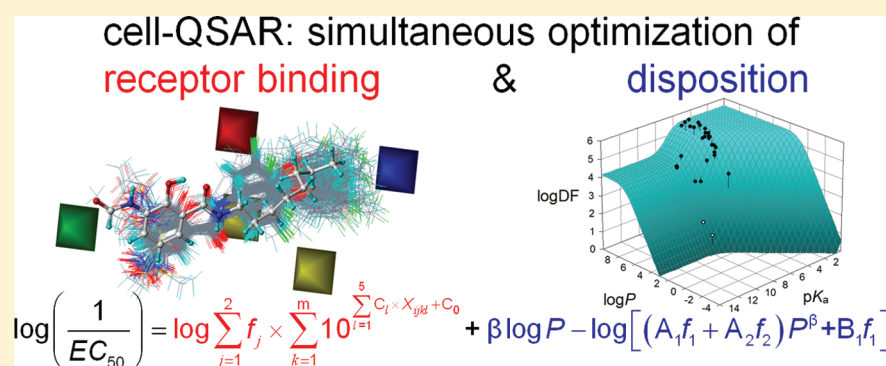
# Cellular Quantitative Structure–Activity Relationship (Cell-QSAR): Conceptual Dissection of Receptor Binding and Intracellular Disposition in Antifilarial Activities of Selwood Antimycins

Senthil Natesan,<sup>†,‡</sup> Tiansheng Wang,<sup>‡</sup> Viera Lukacova,<sup>‡,§</sup> Vladimir Bartus,<sup>‡</sup> Akash Khandelwal,<sup>‡</sup> Rajesh Subramaniam,<sup>†</sup> and Stefan Balaz<sup>\*,†</sup>

<sup>†</sup>Albany College of Pharmacy and Health Sciences, Vermont Campus, Colchester, Vermont 05446, United States

<sup>‡</sup>College of Pharmacy, Nursing and Allied Sciences, North Dakota State University, Fargo, North Dakota 58108, United States

**S** Supporting Information



**ABSTRACT:** We present the cellular quantitative structure–activity relationship (cell-QSAR) concept that adapts ligand-based and receptor-based 3D-QSAR methods for use with cell-level activities. The unknown intracellular drug disposition is accounted for by the disposition function (DF), a model-based, nonlinear function of a drug’s lipophilicity, acidity, and other properties. We conceptually combined the DF with our multispecies, multimode version of the frequently used ligand-based comparative molecular field analysis (CoMFA) method, forming a single correlation function for fitting the cell-level activities. The resulting cell-QSAR model was applied to the Selwood data on filaricidal activities of antimycin analogues. Their molecules are flexible, ionize under physiologic conditions, form different intramolecular H-bonds for neutral and ionized species, and cross several membranes to reach unknown receptors. The calibrated cell-QSAR model is significantly more predictive than other models lacking the disposition part and provides valuable structure optimization clues by factorizing the cell-level activity of each compound into the contributions of the receptor binding and disposition.

## INTRODUCTION

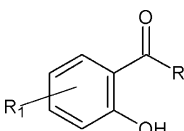
Conceptual methods for the correlation of binding affinities with drug structure (3D-QSARs) are vital to the drug development process. The approaches include receptor-based methods,<sup>1</sup> e.g., free energy perturbation,<sup>2</sup> the linear response method,<sup>3–5</sup> and the mining minima approach,<sup>6–9</sup> and ligand-based methods for unknown receptors, e.g., comparative molecular field analysis (CoMFA).<sup>10</sup> The methods have been developed and perform satisfactorily for the binding data measured with isolated macromolecules. Unfortunately, binding affinities for isolated macromolecules are often unavailable because the receptors have not yet been identified, cannot be isolated without denaturation or dissociation, or do not work in aqueous solutions. In such situations, more complex assays are used, deploying receptors reconstituted in lipid vesicles as the simplest system or utilizing intact cells, tissues, organs, or organisms.

The effective concentrations of the drugs in the receptor surroundings vary among the studied compounds because of interactions with nonreceptor assay constituents. Drug disposition has often been neglected in modeling the bioactivities measured in complex systems. At best, the 1-octanol/water partition coefficients and their squares,<sup>11</sup> and sometimes the dissociation constants,<sup>12</sup> as properties affecting the disposition, have been included in simplified linear forms. In many cell-level studies, 3D-QSAR methods have been applied without any correction for ligand disposition or empirical models with various descriptors have been deployed. Here, we propose a straightforward solution to the problem of QSAR modeling of cell-level data: extend the proven 3D-QSAR methods by accounting for the varying ligand disposition in the receptor surroundings using the *disposition function* (DF). The

**Received:** October 11, 2011

**Published:** April 2, 2012

Table 1. Structures and Properties of Antimycin Analogues



compd no.	R <sub>1</sub>	R <sub>2</sub>	log P <sup>a</sup>	pK <sub>a</sub> <sup>b</sup>	log K <sup>c</sup>	log DF <sup>c</sup>	log(1/EC <sub>50</sub> ) <sup>d</sup>
1	3-NHCHO	NHC <sub>14</sub> H <sub>29</sub>	7.491	7.88	1.181	4.150	5.155
2 <sup>e</sup>	3-NHCHO	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	5.955	7.21	2.307	3.438	5.620
3	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	6.944	5.15	1.710	5.476	7.398
4	5-SCH <sub>3</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	7.402	8.00	2.290	4.086	6.319
5	5-SOCH <sub>3</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	5.722	6.80	1.843	3.535	5.125
6	3-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	6.944	4.60	1.668	5.548	6.824
7	5-CN	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	6.698	5.66	2.534	5.225	7.839
8	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>4</sub> -4-(OC <sub>6</sub> H <sub>4</sub> -4-CF <sub>3</sub> )	6.541	5.15	1.483	5.397	7.022
9	3-SCH <sub>3</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	7.402	7.47	1.931	4.292	6.420
10	5-SO <sub>2</sub> CH <sub>3</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	5.731	6.11	1.681	4.166	6.000
11 <sup>e</sup>	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>4</sub> -4-(OC <sub>6</sub> H <sub>5</sub> )	5.658	5.31	1.450	4.818	6.097
12	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(COC <sub>6</sub> H <sub>4</sub> -4-Cl)	6.264	5.37	1.803	5.184	7.131
13	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>4</sub> -4-(OC <sub>6</sub> H <sub>3</sub> -2-Cl-4-NO <sub>2</sub> )	5.884	5.20	1.936	5.067	6.921
14	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-OCH <sub>3</sub> )	6.150	5.11	1.839	5.271	6.770
15	3-SO <sub>2</sub> CH <sub>3</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(OC <sub>6</sub> H <sub>4</sub> -4-Cl)	5.731	5.10	1.433	5.034	6.301
16	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(SC <sub>6</sub> H <sub>4</sub> -4-Cl)	7.476	5.25	1.554	5.504	7.357
17 <sup>e</sup>	3-NHCHO	NHC <sub>6</sub> H <sub>13</sub>	3.259	7.88	4.018	0.430	<5.000
18	3-NHCHO	NHC <sub>8</sub> H <sub>17</sub>	4.317	7.88	4.091	1.488	5.585
19 <sup>f</sup>	3-NHCOCH <sub>3</sub>	NHC <sub>14</sub> H <sub>29</sub>	7.468	7.88	1.181	4.143	5.097
20 <sup>e</sup>	5-NO <sub>2</sub>	NHC <sub>14</sub> H <sub>29</sub>	8.648	5.83	1.016	5.406	6.893
21	3-NO <sub>2</sub>	NHC <sub>14</sub> H <sub>29</sub>	8.648	5.28	1.293	5.522	6.818
22	3-NO <sub>2</sub> -5-Cl	NHC <sub>14</sub> H <sub>29</sub>	9.419	4.68	1.684	5.564	7.362
23	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>4</sub> -4-C(CH <sub>3</sub> ) <sub>3</sub>	5.386	5.55	1.659	4.369	6.229
24 <sup>e</sup>	5-NO <sub>2</sub>	NHC <sub>12</sub> H <sub>25</sub>	7.590	5.83	1.647	5.361	7.409
25	3-NO <sub>2</sub>	NHC <sub>16</sub> H <sub>33</sub>	9.706	5.28	0.555	5.524	5.959
26	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(NH-C <sub>6</sub> H <sub>4</sub> -4-Cl)	6.900	5.12	0.968	5.477	6.432
27	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>4</sub> -4-(OC <sub>6</sub> H <sub>4</sub> -3-CF <sub>3</sub> )	6.541	5.16	1.675	5.394	7.027
28	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>3</sub> -3-Cl-4-(NHC <sub>6</sub> H <sub>4</sub> -4-SCF <sub>3</sub> )	7.899	5.04	1.711	5.541	7.553
29	5-NO <sub>2</sub>	NH-3-Cl-4-(3-CF <sub>3</sub> C <sub>6</sub> H <sub>4</sub> O)C <sub>6</sub> H <sub>3</sub>	7.114	5.06	1.872	5.510	7.071
30 <sup>e</sup>	5-NO <sub>2</sub>	NHC <sub>6</sub> H <sub>4</sub> -4-(CH(OH)C <sub>6</sub> H <sub>5</sub> )	3.870	5.42	1.197	3.015	<5.000
31 <sup>f</sup>	5-NO <sub>2</sub>	C <sub>6</sub> H <sub>4</sub> -4-Cl	4.329	6.73	0.777	2.221	6.481

<sup>a</sup>ClogP-predicted 1-octanol/water partition coefficient *P*. <sup>b</sup>SPARC-predicted acidity of the phenolic hydroxyl. <sup>c</sup>Affinities and disposition function values predicted from the DF-MSMM CoMFA model. <sup>d</sup>Measured antifilarial activities (molar EC<sub>50</sub>).<sup>36,52</sup> <sup>e</sup>Test set compounds. <sup>f</sup>Compounds excluded before analysis because of singularity.

conceptual *cell*-QSAR approach is based on a new correlation equation combining the 3D-QSAR expression and the DF, whereby the coefficients in both parts can be optimized either simultaneously or separately if the uptake data are available. The exact form of the correlation equation depends on the kinetics of the processes underlying the drug effects and on the complexity of studied compounds (common skeleton, similarity of substituents, ability to ionize and form tautomers), as discussed in the Methods.

The DF describes the kinetics of drug disposition and relates the dose or the initial concentration to the concentration in the receptor surroundings (see eq 2 in the Methods). The first DF forms were peak-shaped dependencies, on logarithmic scales, of the drug concentration inside the receptor compartment, which was separated from the dosing compartment by a few bilayers, on lipophilicity (the reference partition coefficient).<sup>13–16</sup> These dependencies represented the basis of the first QSAR techniques, such as the Hansch approach<sup>13</sup> and related methods. Subsequently, other parameters, including the ionization constant,<sup>17</sup> the exposure time, the metabolic rate

parameters,<sup>18</sup> the ratio of the partition coefficients in alkane/water and 1-octanol/water systems,<sup>19–22</sup> and the polar surface area<sup>23,24</sup> as descriptors of H-bonding ability, the partition coefficient between phosphatidylcholine headgroups and hexadecane as a predictor of bilayer localization and partitioning,<sup>25</sup> the membrane-interaction QSAR parameters,<sup>26,27</sup> the polarizability,<sup>28</sup> the cross-sectional area of the molecule,<sup>29</sup> and numerous others, were added. The DF functional forms developed from the parabolic,<sup>13</sup> bilinear,<sup>15</sup> nonequilibrium,<sup>30</sup> equilibrium,<sup>17</sup> and mixed models to the pseudoequilibrium DF.<sup>31,32</sup> Some of these DFs predicted reliably outside the used property ranges, in contrast to empirical models.<sup>33</sup>

The DF can also be calibrated independently using the data on cellular uptake and disposition, which are obtained by various experimental approaches. High-content screening techniques provide more detailed information than classical uptake experiments,<sup>34</sup> as demonstrated for the disposition of a small library of lipophilic, fluorescent compounds in living HeLa cells.<sup>35</sup> Structure-dependent phenomena, such as active

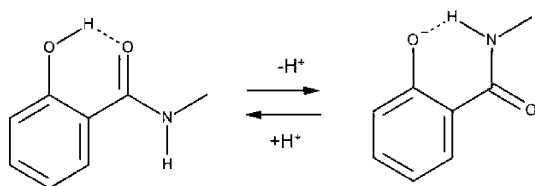
influx or efflux and metabolism, could be captured in this process.

The additional effort to obtain and fit the disposition data can be spared in situations when the disposition is not complicated by structure-dependent phenomena. Then the DF depends on drug properties characterizing the bilayer transport and accumulation, nonspecific protein binding, hydrolysis, and other nonenzymatic reactions. An appropriate DF can be selected from the armoire of available models and can be calibrated simultaneously with the 3D-QSAR model using the cell-level data. This approach is demonstrated here using the Selwood data<sup>36</sup> on filaricidal activities of analogues of antimycin A, which inhibits mitochondrial electron transport and oxidative phosphorylation<sup>37</sup> and induces apoptosis.<sup>38</sup> These data are regarded as difficult to model and are frequently used to validate QSAR methods on the basis of variable selection. The descriptive ability of the generated models did not exceed 60–80% of the explained variance,<sup>36,39–47</sup> so there is space for improvement.

To the best of our knowledge, this cell-QSAR study represents the first attempt at the QSAR modeling of drug effects in complex systems, where both drug–receptor interactions and drug disposition significantly vary among studied compounds and both are conceptually treated using a common correlation equation.

## RESULTS AND DISCUSSION

**Studied Antimycin A Analogues.** The studied compounds, listed in Table 1, have varying substituents,  $R_1$ , on the phenol ring and large lipophilic substructures,  $R_2$ , which replace a complex dilactone moiety of the parent compound. The compounds can be formally divided into three groups: two-ring analogues (23 and 31), three-ring analogues (2–16 and 26–30), and alkyl-chain analogues (1, 17–22, 24, and 25). The salicylamide moiety, present in all compounds except 31, exhibits a peculiar intramolecular H-bonding.<sup>48,49</sup> Two types of intramolecular H-bonds are formed, depending upon ionization of the hydroxyl (Figure 1).



**Figure 1.** Conformational switching of an intramolecular H-bond in the salicylamide ring of antimycin derivatives upon ionization.

The hydroxyl  $pK_a$  values (Table 1) of the studied antimycins were estimated by the SPARC method,<sup>50</sup> which seems to take the intramolecular H-bond into account. The magnitude of the  $pK_a$  values (range 4.6–8.0) is determined mainly by the electron-withdrawing ability of the  $R_1$  substituents on the phenol ring, with the  $R_2$  substituents exhibiting a slighter influence. Under neutral conditions, the compounds are present in the aqueous solution as a mixture of ionized and nonionized species, with the balance shifted toward the ionized species for most compounds, except compounds 1, 2, 4, 9, and 17–19 having  $pK_a > 7$ . Formation of tautomers and carbonyl hydrates was negligible, according to the SPARC estimates; otherwise, they would be included in the model in the same way as ionization. Table 1 also contains the activities spanning

almost 3 orders of magnitude and the 1-octanol/water partition coefficients  $P$  (ClogP estimates<sup>51</sup> for neutral molecules).

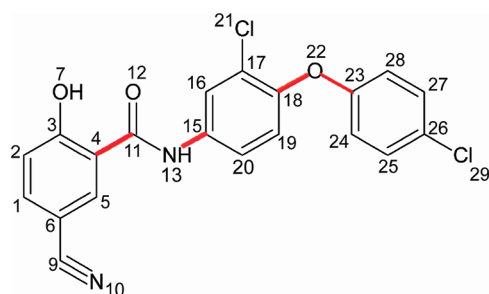
Antimycin A is a known inhibitor of oxidative phosphorylation.<sup>37</sup> Therefore, we examined whether the Selwood data could be explained by our two-decade-old, conceptual QSAR model for uncouplers.<sup>34</sup> The model describes isoeffective concentrations eliciting a given degree of uncoupling as a nonlinear function containing lipophilicity ( $P$ ) and acidity ( $pK_a$ ). Although the surface generated by the fit to the data was reminiscent of the expected shape,<sup>34</sup> the deviations of individual compounds were too large to consider the fit satisfactory. We hypothesized that the deviations were caused by structure-specific interactions with an unknown receptor; therefore, we decided to deploy a ligand-based 3D-QSAR method.

**Setup of the Cell-QSAR Model.** The cell-QSAR correlation equation for the simplest drug effect scenario consists of the product of the drug–receptor association constant  $K$  and the DF (eq 3 in the Methods). The receptor is unknown in this case, so  $K$  was expressed using one of the popular ligand-based methods, CoMFA.<sup>10</sup> The antimycin analogues ionize to different degrees and are flexible; therefore, we applied our multispecies (MS), multimode (MM) modification of CoMFA.<sup>53</sup> The pseudoequilibrium form of the DF (eq 10 in the Methods),<sup>32</sup> treating absorption as an instantaneous process, was chosen, because no lag phase was observed in the time course of toxicity.<sup>36,52</sup> An additional reason for selecting this DF form was the ability of the underlying model<sup>32</sup> to handle speciation of the ligands.

**Ligand Superposition.** Pharmacophores obtained for all species using the Galahad procedure,<sup>54,55</sup> which performs flexible superposition based on the active analogue approach,<sup>56</sup> indicated that the salicylamide moieties are closely superimposed in all cases. This outcome is consistent with the apparent importance of this moiety for inhibition: an analogue capable of H-bond formation (Figure 1) inhibited the cytochrome  $bc_1$  complex activity in submitochondrial particles much stronger than a similar analogue lacking this ability.<sup>57</sup>

A rigid compound that preferably exhibits a high activity would be very useful for guiding the superposition of the ligands. Unfortunately, there is no truly rigid analogue available in the studied series. The three-ring compounds contain more rigid aromatic rings than other compounds, so we selected the most active compound in this group and in the entire series (7) as the template. Because the pharmacophore-generating procedure did not provide clues about the conformation of the lipophilic fragments connected to the salicylamide ring, we used the low-energy conformations of compound 7 as the assumed bound conformations. This step is associated with the risk that the overall spatial organization of the binding site model may not completely correspond to that of the actual receptor. While the binding points will be positioned appropriately around individual rigid rings of the template, the relative positions of the individual groups of points surrounding the rigid fragments may differ from those of the actual receptor because of possible fragment rotations. This drawback would not affect the predictive ability of the model for compounds which can adopt the template conformation without significant conformational strain.

The conformational space of compound 7 is characterized by four torsions (Figure 2). Conformational analysis identified the minimum-energy conformations for the H-bonded forms of both neutral and ionized molecules (Figure 1). The common minimum conformations of ionized and neutral species without



**Figure 2.** Four torsions of compound 7 (Table 1) characterizing the conformations of three-ring analogues displayed in red color. They are  $\Phi_1$  (C3–C4–C11–O12),  $\Phi_2$  (C11–N13–C15–C20),  $\Phi_3$  (C17–C18–O22–C23), and  $\Phi_4$  (C18–O22–C23–C28).

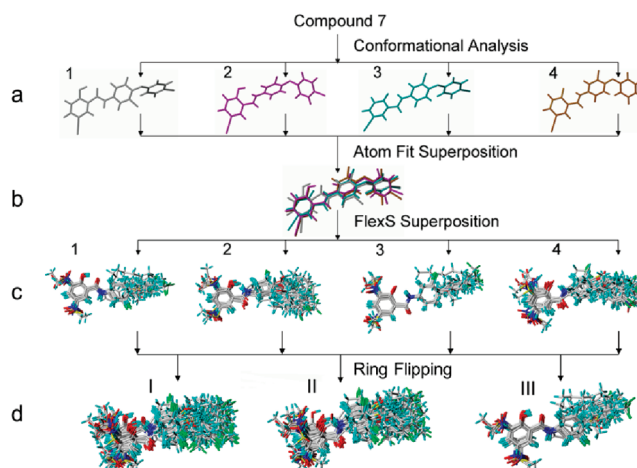
the intramolecular H-bonds were also looked up, for the possibility that the hydroxyl and the amide groups would interact with the receptor instead of forming the intramolecular bond. Two skeleton conformations appear for each of the two species: one *closed* via the intramolecular H-bond and one *open* without the H-bond. The conformations are summarized in Table 2.

It has long been recognized that flipping of symmetrical rings with asymmetrically placed substituents results in similar but distinct bound geometries. If the receptor structure is unknown, preferred geometries cannot be selected a priori and all geometries should be considered in model building.<sup>58–64</sup> A total of 16 of 31 studied compounds (Table 1) have substituents in *meta*-positions on either the middle ring (2–7, 9, 10, 12, 14–16, 26, and 28) or the third ring (13 and 27). For these compounds, rotation of the asymmetrically substituted ring by 180° led to doubling of the existing two modes. Compound 29 with two asymmetrically substituted rings is represented by eight modes.

The process of creating individual superpositions is illustrated in Figure 3. The four basic templates of compound 7 were superimposed by the atom fit method in Sybyl (Figure 3b) to ensure the best overlap of the salicylamide moieties, as suggested by the generated pharmacophore. All other molecules were superimposed on the templates using the FlexS procedure<sup>65</sup> (Figure 3c,d).

**DF-MSMM CoMFA Analysis.** The analysis is based on the correlation equation obtained by taking the logarithm of eq 3 ( $X = 0.5$  because the  $EC_{50}$  values are modeled), with the association constant  $K$  expressed by eq 9, and the DF given by eq 10 (the exposure time  $t = \text{constant}$ ) combined with eqs 11 and 12. For both the receptor-binding and disposition parts, ionization is included and the coefficients are optimized simultaneously.

Optimization examines two forms of the DF, for the aqueous phases (shown in eq 10) and membranes. If the receptors are localized in a membrane and the compounds bind to them



**Figure 3.** Superposition procedure: (a) four basic templates of compound 7, generated by conformational analysis and numbered as in Table 2; (b) superposition of the four basic templates; (c) superpositions of all compounds (Table 1) to the four basic templates in respective species and conformations, numbered as in row a; (d) final superpositions, enriched by flipping the benzene rings by 180° for the MSMM situation (I; 200 molecules = 31 compounds  $\times$  2 species  $\times$  2 skeleton conformations  $\times$  1, 2, or 4 ring-flipping conformations), the MM situation (II; 100 molecules created as in situation I but with 1 species), and the standard one-mode situation (III; 31 molecules).

from the hydrocarbon core of the bilayer, eq 10 needs to be multiplied by  $P^\beta$  ( $P$  is the reference partition coefficient and  $\beta$  is the Collander coefficient, which is one of the optimized regression coefficients). The water and membrane variants of eq 10 are initially used in fully expanded form: each of the coefficients  $A$ – $D$  is replaced by eqs 11 and 12. The optimization proceeds with gradual forward selection of energies in relevant grid points in the CoMFA part, which is done separately for the two forms of the DF. The better DF is selected on the basis of the statistical criteria and further optimized by omitting the parameters with higher standard deviations, if the elimination does not cause a significant decrease in statistical criteria. One, or sometimes both, of the coefficients  $C$  and  $D$ , describing lipophilicity-dependent and lipophilicity-independent elimination, can frequently be omitted. In each pair of coefficients  $A$ – $D$  (eq 11), the term for one species may or may not be significant.

The combinatorial search for the best DF-MSMM CoMFA model was extensive for three reasons: many possible combinations of relevant grid points need to be examined, the DF can have multiple functional forms, and many sets of the initial estimates of coefficients need to be tested to avoid trapping in local minima. The best model was selected on the basis of the  $r^2$  and  $q^2$  (for the leave-one-out cross-validation in the training set, used to eliminate unstable models) values and

**Table 2.** Superposition Templates Based on the Most Active Compound 7 (Table 1)

template no.	species <sup>a</sup>	conformation <sup>b</sup>	intramolecular H-bond	torsion <sup>c</sup> (deg)			
				$\Phi_1$	$\Phi_2$	$\Phi_3$	$\Phi_4$
1	neutral	closed	OH...O=C	4.1	2.8	93.8	178.5
2	neutral	open	none	137.7	–15.7	208.3	225.4
3	ionized	closed	NH...O <sup>–</sup>	180.6	0.8	98.4	174.7
4	ionized	open	none	137.7	–15.7	208.3	225.4

<sup>a</sup>See Figure 1. <sup>b</sup>“Closed” and “open” refer to the intramolecular H-bond being present and absent, respectively. <sup>c</sup>See Figure 2.



the number of optimized coefficients and their standard deviations. For illustration, there were 65 021 models with  $r^2 \geq 0.6$ , 11 142 models with  $r^2 \geq 0.8$  and  $q^2 \geq 0.6$ , and 3752 models with  $r^2 \geq 0.9$  and  $q^2 \geq 0.7$ , of which 25 models had less than 11 coefficients, each with smaller than 60% standard deviation.

The best DF-MSMM CoMFA model contained the DF for the compounds interacting with the receptors in a membrane, with metabolism that was either insignificant or invariant for the studied compounds ( $C = 0$  and  $D = 0$ ), and five required CoMFA grid points. The final, pruned correlation equation, resulting from eq 3 ( $X = 0.5$ ), with the association constant  $K$  expressed by eq 9 (the subscript  $i$  was omitted for fractions  $f$ ), and the DF given by eqs 10–12, was

$$\log\left(\frac{1}{EC_{50}}\right) = \log \sum_{j=1}^2 f_j \sum_{k=1}^m 10^{\sum_{l=1}^5 C_l X_{ijkl} + C_0} + \beta \log P - \log[(A_1 f_1 + A_2 f_2) P^\beta + B_1 f_1] \quad (1)$$

The first term is an MSMM CoMFA expression for the receptor binding ( $\log K$ ), and the remaining terms describe disposition ( $\log DF$ ). The fractions of ionized and nonionized molecules,  $f_1$  and  $f_2$ , respectively, include the  $pK_a$  values and the optimized pH value as shown in eq 12 and are identical in both receptor-binding and disposition terms. The microscopic association constant  $K_{ij}$  for each of the  $m$  considered modes is represented by a  $k$ -summand. Each of the  $k$ -summands contains five identical adjustable coefficients,  $C_1$ – $C_5$ , which, for the selected grid points (coordinates in Table 3), multiply the

**Table 3. Characterization of the Receptor-Binding Part of the DF-MSMM CoMFA Model (Eq 1)**

coefficient	value <sup>a</sup>	coordinates (Å) <sup>b</sup>			energy <sup>c</sup>
		<i>x</i>	<i>y</i>	<i>z</i>	
$C_1$	$0.538 \pm 0.247$	12.66	6.83	4.780	S
$C_2$	$-1.931 \pm 0.711$	6.668	10.836	-5.218	S
$C_3$	$-0.245 \pm 0.108$	2.668	6.836	-3.218	E
$C_4$	$-6.657 \pm 2.873$	4.668	16.863	-5.218	S
$C_5$	$0.116 \pm 0.043$	-3.330	14.836	-7.2184	E

<sup>a</sup>Regression coefficients. <sup>b</sup>Grid point coordinates. <sup>c</sup>Energy types: S, steric, E, electrostatic.

energies  $X_{ijkl}$  for the given modes. The addition of multiple modes does not increase the number of optimized coefficients; it only changes the form of the correlation equation by increasing the number of  $k$ -summands.

Nonlinear regression analysis resulting in eq 1 provided the following outcome for the best DF-MSMM model. For the receptor-binding expression, the results are summarized in Table 3. The optimal values of the disposition-describing coefficients for the DF were  $A_1 = (7.142 \pm 2.213) \times 10^{-5}$ ,  $A_2 = (2.636 \pm 1.254) \times 10^{-6}$ ,  $B_1 = (9.989 \pm 4.233) \times 10^2$ , and pH  $7.56 \pm 0.32$ . The optimization of the Collander coefficient  $\beta$  and the coefficient  $C_0$  did not lead to a sufficient improvement, so their values were fixed as  $\beta = 1$  and  $C_0 = 0$ . The description statistics, calculated for the training set, consists of the sum of squares of errors (SSE), where the error is the difference between calculated and experimental values, and the squared correlation coefficient (the coefficient of determination, equal to the percentage of explained variance)  $r^2 = 1 - SSE/SYY$ , where SYY is the sum of squares of deviations of the experimental  $\log K$  values from their average. The prediction

statistics contains similar indices calculated for the test set that was not used in the development of the model: predictive sum of squares of deviations between predicted and experimental values (PRESS) and the squared correlation coefficient  $q^2 = 1 - PRESS/SYY$ . Table 4 summarizes the descriptive (SSE and  $r^2$  for the training set) and predictive ( $q^2$  and PRESS for the test set) statistical criteria for the DF-MSMM CoMFA model (1) and for the reduced models (2–13). For comparison, the random scrambling of bioactivities (50 runs) produced no significant correlations, with the average values of SSE = 11.12 and  $r^2 = 0.250$ . Optimized eq 1 also provides the affinity ( $\log K$ ) and disposition ( $\log DF$ ) components of bioactivity (Table 1). A test set (Table 1, compounds 2, 11, 17, 20, 24, and 30) covering the entire range of biological activities was selected before optimization. The use of multiple test sets, which would be a preferred way to examine the predictive abilities of the model, was precluded by the extent of the optimization procedure. For compounds 17 and 30 with semiquantitative activities, the  $q^2$  and PRESS values (Table 4) were calculated using the upper limits, i.e.,  $\log(1/EC_{50}) = 5$ . Because only the DF-MSMS CoMFA model correctly predicted  $\log(1/EC_{50}) < 5$  (Figure 4b), the predictive abilities are probably underestimated for the DF-MSMS CoMFA model and overestimated for all other models.

Table 4 demonstrates that all features used in the DF-MSMM CoMFA model are necessary for a good prediction. Omission of the disposition leads to a decline in predictive abilities (PRESS and  $q^2$ ), even if multiple species and/or modes are considered (models 2 and 3). The decline is more severe for standard, one-mode CoMFA models (4–13), including those with the averaged multiple modes (models 6, 9, and 12) and even for the prevalent bound species and modes (model 13), which were selected by DF-MSMM CoMFA model 1. This fact indicates that the one-mode QSAR solution may not exist. Although only this indirect, model-based indication is available for the studied case, multiple modes were experimentally proven to occur in many ligand–macromolecule complexes (see the references in ref 57). Interestingly, the descriptive abilities (SSE and  $r^2$ ) were great in all cases. This may, in part, explain the use of 3D-QSAR methods without any description of disposition for cell-level activities.

It is worth mentioning that an exhaustive examination of modes and species for each compound in a standard CoMFA analysis, seeking a one-mode solution, is technically not feasible. The 23 compounds of the training set, each present in two species and two conformations (for simplicity, the ring-flipping conformations are omitted), would require  $4^{23}$  analyses or about 2.2 million CPU years, if a single one-mode analysis only lasts 1 s. Therefore, the MSMM approach, performing an extensive, albeit not exhaustive search lasting several weeks on a few processors, compares favorably.

The calculated bioactivities for the training set and the predictions for the test set compounds are compared with experimental values in parts a and b, respectively, of Figure 4. The use of multiple species and modes and, especially, of the DF clearly improves predictivity (PRESS and  $q^2$ ). Descriptive statistical indices for calibration display a similar trend, except for the excellent SSE and  $r^2$  values for the standard, one-mode CoMFA model, utilizing the full set of energies. The standard model, however, did not confirm its calibration quality in the prediction exercise (Figure 4b).

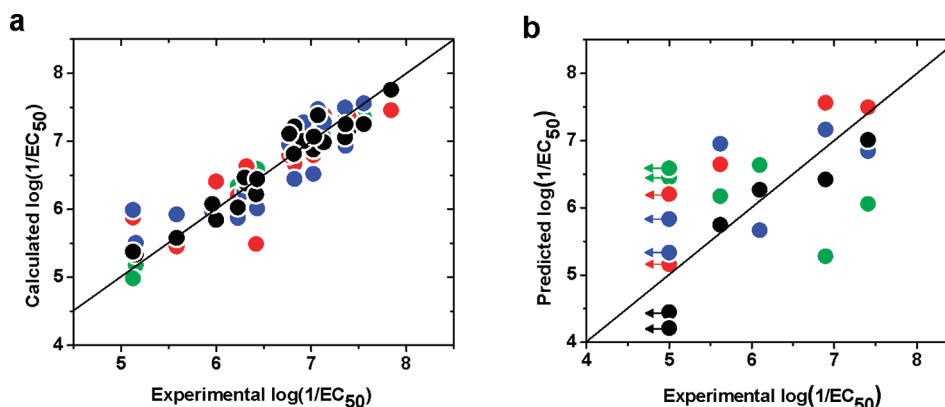
**3D-QSAR Contour Map.** The steric and electrostatic contours for the DF-MSMM CoMFA model with various

Table 4. Statistical Indices Characterizing CoMFA Models of Varying Complexity<sup>a</sup>

CoMFA model				statistical indices			
no.	species	mode	no. of coefficients	descriptive		predictive	
				SSE	$r^2$	PRESS	$q^2$
1	DF-MS	MM	5 + 5	0.902	0.923	1.356	0.815
2	MS	MM	11	2.336	0.801	3.161	0.569
3	SS: major <sup>b</sup>	MM	11	2.476	0.789	2.995	0.592
4	SS: neutral	SM: closed	5 (3568) <sup>c</sup>	0.154	0.987	6.410	0.126
5	SS: neutral	SM: open	5 (3563)	0.361	0.969	6.070	0.173
6	SS: neutral	MM: average <sup>d</sup>	5 (564)	0.400	0.966	6.498	0.114
7	SS: ionized	SM: closed	5 (3569)	0.155	0.987	6.686	0.089
8	SS: ionized	SM: open	5 (3584)	0.196	0.983	5.411	0.262
9	SS: ionized	MM: average <sup>d</sup>	5 (572)	0.580	0.951	6.228	0.151
10	SS: major <sup>b</sup>	SM: closed	5 (3570)	0.154	0.987	6.785	0.075
11	SS: major <sup>b</sup>	SM: open	5 (3583)	0.325	0.972	4.818	0.343
12	SS: major <sup>b</sup>	MM: average <sup>d</sup>	5 (534)	0.064	0.995	5.897	0.139
13	SS: bound <sup>e</sup>	SM: bound <sup>e</sup>	5 (3581)	0.403	0.966	4.455	0.393

<sup>a</sup>The models differ in the use of disposition function (DF), multiple or single species (MS or SS), and multiple or single modes (MM or SM).

<sup>b</sup>Prevailing species in aqueous solution used for each compound. <sup>c</sup>Number of latent variables shown, with the number of all energies given in parentheses. <sup>d</sup>The energies of all considered modes are averaged and analyzed by standard, one-mode CoMFA. <sup>e</sup>Prevalent bound species and mode as determined by the DF-MSMM CoMFA model used for each compound.



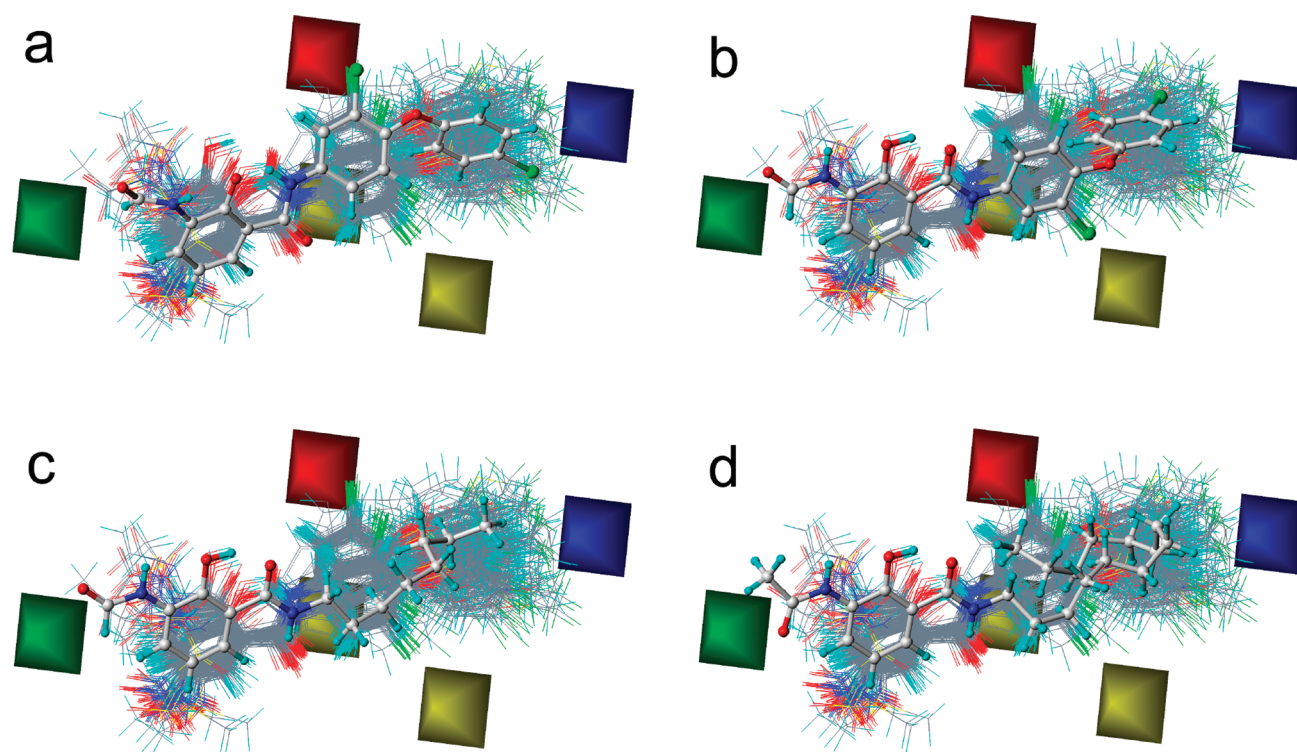
**Figure 4.** Calculated and predicted activities plotted against experimental data for the training set (a) and test set (b), generated with the DF-MSMM (black), MSMM (blue), SSMM (red), and standard (green) CoMFA models (models 1–3 and 11 in Table 4, respectively). In plot b, the two experimental values of  $\log(1/EC_{50}) = 5$  are actually  $< 5$  (Table 1, compounds 17 and 30), as indicated by the arrows.

compounds embedded in the background of overall MSMM superposition (Figure 3d, situation I) are displayed in Figure 5, along with representative ligand conformations. The contour map shows that the binding site model is defined by a sterically favorable region near the salicylamide ring, two sterically unfavorable regions in the midsection of the binding site, an electronegative-favoring region near the second ring of two- or three-ring compounds, and an electropositive-favoring region at the lipophilic end of the molecules. The absence of electrostatic regions around the salicylamide ring may result from small differences among the studied compounds in this substructure.

The binding conformation in which the bulky chlorine atom in the second ring is closer to the salicylamide OH group and far from a sterically unfavorable region (Figure 5a) is preferred in 11 out of 15 compounds. Consequently, the ring-flipped conformation (Figure 5b) is not populated. Compounds with lipophilic alkyl groups tend to have a wide range of activities depending on the length of the alkyl chains. Compounds 17 and 18 (Figure 5c) with six and eight methylene groups, respectively, bind in an extended conformation and show the highest predicted binding affinities (Table 1,  $\log K > 4$ ), indicating the optimal chain length for binding. On the other

hand, compounds with 12–14 methylene groups (Table 1, 1, 19–22) bind in a packed conformation (Figure 5d) that results in low affinities (Table 1,  $\log K < 1.7$ ), probably because the folded chain protrudes into the sterically unfavorable regions. For the same reason, the affinity of compound 25 with 16 methylene groups exhibits a further drop ( $\log K \approx 0.5$ ). This reasoning could also apply to compound 23 showing only average affinity ( $\log K \approx 1.7$ ), although its 4-*tert*-butylphenyl ring is similar in length to the optimal *n*-hexyl chain in the extended conformation (compound 17) but is much bulkier. The red electronegative-favoring region in the midsection of the binding site improves the affinity ( $\log K$ ) of 15 of 31 compounds with an electronegative chlorine atom in the second ring, even though their overall activities differ significantly because of varying disposition. This interaction also causes the compounds with the electronegative oxygen (2–11, 13–15, 27, 29) or carbonyl (12) linkers between the second and third rings to have a higher binding affinity than compounds with other linkers (16, 26, 30).

Two compounds, 19 and 31, were excluded from optimization before its start because of singularities caused by molecular parts protruding into subspaces which were not



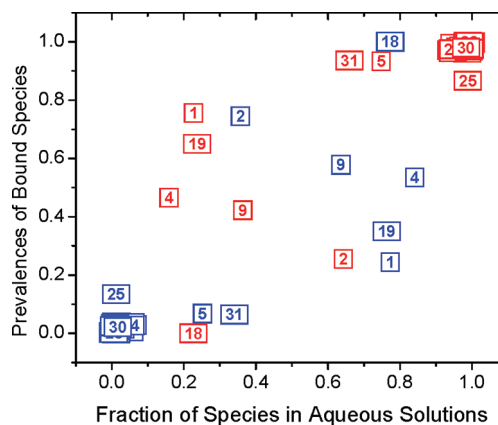
**Figure 5.** Contour maps of the DF-MSMM CoMFA model with embedded superposition of all 200 molecules to indicate the applicability space. Green and yellow contours indicate regions that are sterically favorable and unfavorable, respectively. Blue regions favor electropositive groups, and red regions favor electronegative groups. The following ligands are presented in ball and stick models: (a) compound **2** (Table 1) binding in ionized and open conformation, (b) compound **2** binding in neutral and closed conformation, showing the clash of the 3-Cl in the second ring with one of the sterically unfavorable regions, (c) compound **18** with a shorter alkyl chain, one of the strongest binders, (d) compound **19** with the 3-acetamide group in the salicylamide ring occupying a sterically favorable region, showing clashes of the long chain with one of the sterically unfavorable regions.

reached by other compounds in the series. Singularity of compound **19** is caused by the 3-acetamide group in the salicylamide ring, which reaches a spot in the vicinity of the sterically attractive region (Figure 5d). The singular region may be an open, water-filled space with no effect on affinity because the prediction for compound **19** by the DF-MSMM model ( $\log K + \log DF = 5.324$ , Table 1) is close to the experimental  $\log(1/EC_{50}) = 5.097$ . Compound **31** lacks an amine linker between its rings, and therefore, its molecule extends into uncharted space. Compound **31** has a much lower predicted activity ( $\log K + \log DF = 2.998$ ) than the actual measured  $\log(1/EC_{50}) = 6.461$ . This difference indicates that there may be an attractive spot in the midsection of the binding site which the model does not capture because of the lack of molecules occupying this region.

**Prevalences of Bound Species and Modes.** The prevalences, given by the Boltzmann distribution shown in eq 8 in the Methods, cannot be guessed a priori because they are one of the outcomes of optimization. The prevalences for the DF-MSMM CoMFA model (eq 8), with individual  $K_{ijk}$  values calculated as the decadic terms in eq 1 with optimized values of the coefficients  $C_i$ , are shown in detail in the Supporting Information and summarized here. A total of 16 of 31 compounds (**1**, **2**, **8**, **10**, **11**, **13**, **15–18**, **23**, **25**, **27**, **28**, **30**, **31**) bind predominantly in the closed conformation, maintaining the intramolecular H-bond (Figure 1). The remaining compounds bind in the open conformation, exposing the salicylamide's polar groups, with at least 50% prevalences. Compounds **15**, **17**, **18**, and **25** bind 100% in the closed

conformation, while compound **20** binds 100% in the open conformation.

Most compounds prefer to bind as ionized species. The exceptions are compounds **2**, **4**, **9**, **17**, and **18**, for which the neutral species bind with 74%, 53%, 58%, 100%, and 100% prevalences, respectively. Figure 6 indicates that the species distributions of bound and free molecules are quite similar for most compounds. For neutral species, most compounds have both the fraction in water and binding prevalence less than 10%. For ionized species, most compounds have the fraction in water and binding prevalence as high as 90%. About a quarter



**Figure 6.** Speciation of free and bound molecules according to the DF-MSMM CoMFA model. Neutral and ionized species are shown in blue and red, respectively, as the compound number (Table 1).



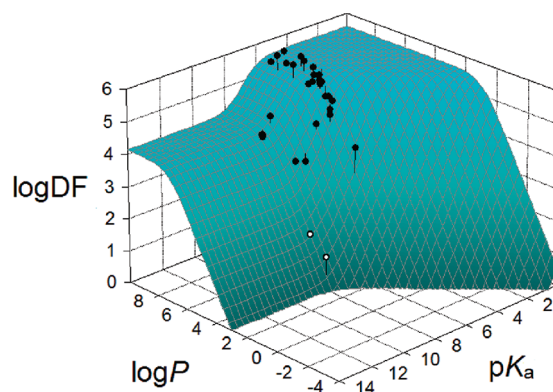
of the compounds (1, 2, 4, 5, 18, 19, 31), however, exhibit significant differences between the fractions in water and binding prevalences.

The predominant binding species and binding conformations are to a large extent affected by the substituents in the salicylamide ring. Strong electron-withdrawing groups promote ionization and also contribute to higher activity: only 3 (4, 9, 31) out of 24 compounds with  $\log(1/EC_{50}) > 6$  have  $pK_a > 6.5$  (Table 1). However, lipophilic moieties including alkyl chains and two or three ring systems also affect the binding as well as disposition, so the structure–activity relationships are complex.

**Binding of Alkyl-Chain Analogues.** The strongest binders, compounds 17 and 18 (Table 1,  $\log K > 4$ ) with a 3-formylamino group in the salicylamide ring predominantly bind as neutral species in the closed conformation. The shorter alkyl chains may allow sufficient conformational freedom for these two compounds to make the best possible interactions in the binding site, leading to high binding affinities. Long-chain analogues (compounds 19–22 and 24, Table 1) predominantly bind in ionized open conformations or sometimes in ionized closed conformations, as seen for compounds 1 and 25. All long-chain analogues exhibit low to average binding affinity, possibly because of steric clashes of long chains in packed conformations (Figure 5d) with the sterically unfavorable regions in the midsection of the binding site.

**Binding of Two- and Three-Ring Analogues.** Compounds with a 2-NO<sub>2</sub> or 5-NO<sub>2</sub> group in the salicylamide ring bind predominantly as ionized species, either in open or closed conformations, depending upon the presence of a 3-Cl substituent in the second ring (the absence of which seems to be compensated by the presence of a 2-Cl or 4-CF<sub>3</sub> substituent in the third ring). Compounds 3, 6, 12, 14, 26, 28, and 29 with a 3-Cl substituent in the second ring mainly bind in the open conformation, although compounds 12 and 29 bind about equally in the closed conformation as well, and compound 28 shows slightly higher preference for the closed conformation. This variation is possibly affected by the presence of a carbonyl linker group between the second and third rings in compound 12 and 4-SCF<sub>3</sub> and 3-CF<sub>3</sub> groups in the third ring of compounds 28 and 29, respectively. Compounds without a 3-Cl substituent in the second ring (8, 11, 13, 23, 27, and 30) bind predominantly as ionized species and show higher preference for the closed conformation. The absence of steric hindrance by a chlorine substituent in the second ring allows these compounds to assume the preferred closed conformation with an intramolecular H-bond between the amide NH and phenoxide ion. Compound 16 (with a 5-NO<sub>2</sub> group in the first ring and a 3-Cl substituent in the second ring) prefers to bind in the closed conformation, possibly because of the presence of a thioether linkage between the second and third rings, resulting in a longer linker. The most active in this series, compound 7 with a 5-CN substituent in the salicylamide ring, binds predominantly as an ionized species in the open conformation.

**Influence of Disposition.** The disposition of antimycins in female worms of *Dipetalonema viteae* after 120 h of exposure<sup>52</sup> can be visualized for individual compounds as the difference between the experimental  $\log(1/EC_{50})$  values and the  $\log K$  values (Table 1) calculated from the affinity part of the DF-MSMM CoMFA model (eq 1). These values are plotted as a function of the drugs' lipophilicity ( $\log P$ ) and acidity ( $pK_a$ ) in Figure 7, along with the surface given by the disposition part of the DF-MSMM CoMFA model (eq 1) with optimized



**Figure 7.** Dependence of the pseudoequilibrium, membrane-based DF on lipophilicity and acidity, as given by the disposition part of eq 1 with the optimized coefficient values listed in the text. The data for best binders with poor disposition (compounds 17 and 18, Table 1) are shown as open points.

coefficients. The shape of  $\log DF(P, K_a)$  is remarkably nonlinear: it is essentially composed of interconnected planes positioned at varying angles. No further curvatures are expected, according to the model, as the planes extend in all four directions to the limit property values, so this conceptual function safely predicts outside of the property ranges.<sup>33</sup> Disposition significantly modulates the bioactivity of antimycins by  $\sim 5 \log(1/EC_{50})$  units. Using the correct form of the DF for the correlations is essential, because its peculiar shape (Figure 7) is difficult to emulate by polynomials, which are too flexible to approximate the planar areas.<sup>33</sup>

The resulting fit (eq 1, Table 3 and accompanying text) can be interpreted as follows. The variation in disposition is caused by pseudoequilibrium partitioning of ionizable compounds. The hypothetical receptors are localized in a membrane, with the drugs accessing the binding site (the CoMFA model in Figure 5) from within the bilayer. The optimized coefficients are associated with the processes that determine drug disposition:  $A$ , with bilayer accumulation and nonspecific protein binding;  $B$ , with accumulation in aqueous phases, with subscripts 1 and 2 indicating the relatedness to the nonionized and ionized species, respectively. The coefficient  $B_2$  was not necessary for a good fit, indicating that accumulation of ionized molecules in the aqueous compartments did not cause significant variation in overall disposition. The exponent, coming from eq 10 and containing coefficients  $C$  and  $D$  describing metabolism, need not be used in eq 1, meaning that  $C_1 = C_2 = D_1 = D_2 = 0$ . The model indicates that the pseudoequilibrium disposition of the studied compounds was not affected by metabolism, which could be either negligible or similar for all compounds.

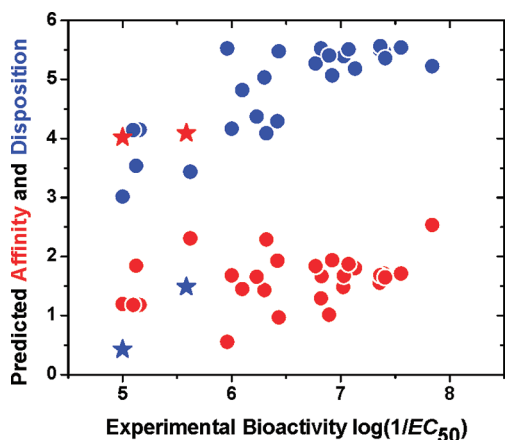
The internal pH value in terms  $f_1$  and  $f_2$  (eq 12) was optimized as  $pH 7.56 \pm 0.32$ . This high value can normally be encountered only in one part of the cell: in the matrix of mitochondria. Therefore, the fit indicates that the hypothetical receptors are localized in the bilayer that is in contact with the mitochondrial matrix. This indication agrees with the current knowledge about antimycin A's mechanism of action, which focuses on inhibition of electron transport.<sup>57,66–69</sup> Antimycin A binds at the ubiquinone binding site  $Q_i$  of the complex that is located on the matrix side of the cristae membrane.<sup>70</sup>

Notably, the disposition function provides clear guidelines for the optimization of the drug structure for the best



disposition. The top plateau in Figure 7 indicates the range of  $\log P$  and  $pK_a$  values that lead to the maximum drug disposition in the receptor surroundings. The region of the maximum disposition can be defined as  $pK_a \leq \log P$  for  $\log P \leq 6$  and  $pK_a \leq 6$  for  $\log P > 6$ . The knowledge of suitable combinations of  $\log P$  and  $pK_a$  values provides unique clues for optimization of drug structure, also because the lower  $\log P$  and  $pK_a$  values will lead to increased aqueous solubility.

**Factoring Bioactivity into Affinity and Disposition Components.** Equation 1 defines two contributions to the affinity:  $\log K$  (the first term) and disposition ( $\log DF$ ; the remaining terms). The two contributions were enumerated from the calibrated DF-MSMM CoMFA model (eq 1 and the accompanying text, Table 3), summarized in Table 1, and plotted against the experimental antifilarial activities (Table 1) in Figure 8.



**Figure 8.** Dissection of affinity ( $\log K$ ; red symbols) and disposition ( $\log DF$ ; blue symbols) contributions to overall bioactivity by the calibrated DF-MSMM CoMFA model. Stars mark the best binders 17 and 18 with poor disposition. All data are given in Table 1.

Figure 8 and Table 1 show that two weakly active compounds (17 and 18) are actually the best binders, which are held back by poor disposition. Their disposition can be, in principle, increased by 4–5 logarithmic units, leading to overall bioactivities approaching  $\log(1/EC_{50}) \approx 9.5$ , which is almost 2 orders of magnitude better than that of the most active compound 7. Disposition is easier to optimize than binding because the former depends on lipophilicity and acidity, which are well understood in terms of drug structure. The DF-MSMM CoMFA model, and cell-QSAR models in general, provides unique clues for optimization of the compounds.

## CONCLUSIONS

The cell-QSAR concept adapts ligand-based and receptor-based 3D-QSAR approaches for use with the bioactivities measured in cells or more complex systems. The adaptation consists of the extension of the correlation equation specific for the given 3D-QSAR method by the DF. The DF describes changes in the concentration of compounds in the receptor surroundings originating from membrane accumulation, binding to non-receptor proteins, hydrolysis, and possibly other processes. The cell-QSAR concept was used here with the popular ligand-based CoMFA method. A conceptual combination of the DF with our MSMM extension of the CoMFA model was applied to antifilarial activities of antimycin A analogues, representing a

part of the difficult-to-model Selwood data set that is often used as a test case for assessment of new variable-selection algorithms. The fitting results are very satisfactory and show that each component of the model (DF, MS, MM) improves the predictive ability. The optimized form of the disposition function indicates that the compounds bind to the hypothetical, membrane-bound receptor in the cristae bilayer of mitochondria and the differences in their metabolic rates are negligible. The perceived complexity of the Selwood data set may be explained by the nonlinearity of the model in lipophilicity and acidity in combination with the lack of proper descriptors for acidity in the original set, the peculiar intramolecular H-bonds in the salicylamide moiety that differ for neutral and ionized species, and the structure-specific receptor binding. The results also illustrate the utility of the cell-QSAR approach in medicinal chemistry. Specifically, the cell-QSAR approach delineates the affinity and disposition contributions to bioactivity, which provide for a straightforward, separate structure optimization for affinity and disposition. Although further testing is necessary, the cell-QSAR concept seems to be a promising direction in the structure-based or ligand-based predictions of bioactivities in cellular or more complex systems and drug structure optimization. For larger and more diverse data sets, better superposition techniques and methods for prediction of species fractions will be necessary. The model-based nature opens the cell-QSAR concept for incorporation of any future improvements in 3D-QSAR techniques and the disposition function, as well as new pharmacology and pharmacodynamic scenarios.

## METHODS

**Pharmacokinetics (PK), Pharmacology (PL), and Pharmacodynamics (PD) Phases of the Drug Effect in QSAR.** The PK phase is described by the DF. The simplest PL/PD scenario consists of a situation when the PL phase is represented by the 1:1, fast, and reversible ligand–receptor interaction and the PD phase describes the biological effect, which is a direct and immediate consequence of the drug–receptor interaction and is proportional to the fraction of occupied receptors. The ligand–receptor association constant  $K$  is defined using the concentrations (denoted by square brackets) of the ligand (L), receptor (R), and complex (LR) as  $K = [LR]/([L][R])$ . In a cellular system, the free ligand concentration  $[L]$  varies because of the disposition, and consequently, all three concentrations are time-dependent. Usually, the concentrations are sufficiently low to approximate activities. If the concentration of the free receptors  $[R]$  is expressed as the difference between the total receptor concentration  $[R]_0$  and  $[LR]$ , the fraction of occupied receptors is given as  $[LR]/[R]_0 = K[L]/([L] + 1) = E/E_{max}$ . The latter equality comes from the simplest PD phase, describing the primary effect  $E$  as being proportional to the fraction of occupied receptors,  $[LR]/[R]_0$ . Secondary effects lacking this simple tie to the receptor modification require more complex PD models.<sup>71</sup> Nevertheless, the above scenario applies to many drug effects caused by competitive enzyme inhibition or receptor binding. Since the receptor-bound drug represents only a small drug fraction in the body, the disposition is not significantly affected by the receptor binding, and the ligand concentration in the receptor surroundings,  $[L]$ , can be expressed using the DF with properties  $p$  and time  $t$  as variables and the initial ligand concentration  $c_0$ .<sup>72</sup>

$$[L] = DF(p, t)c_0 \quad (2)$$

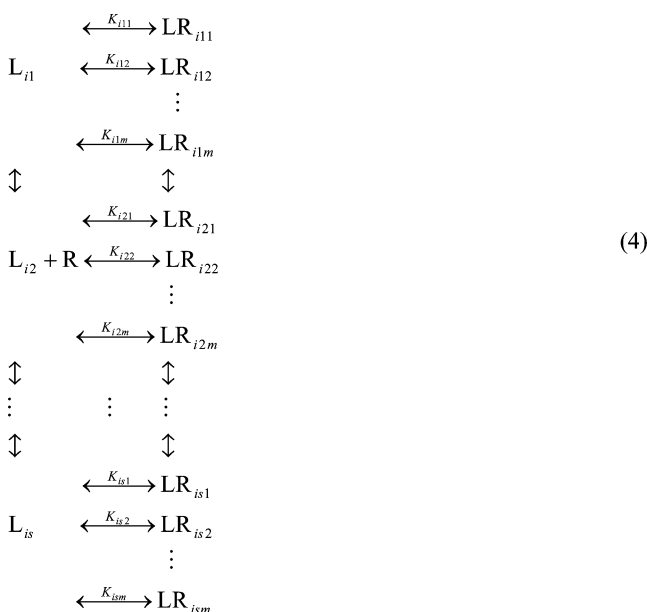
To obtain a PK/PL/PD expression,  $[L]$  in the above expression for  $E/E_{max}$  is replaced by the product  $DF(p, t)c_X(t)$ , where  $c_X(t)$  are the isoeffective drug concentrations eliciting the effect, which is equal to the fraction  $X$  of the maximum effect. With  $E/E_{max} = X$ , the bioactivity, characterized by  $c_X(t)$ , is described as<sup>72</sup>

$$1/c_X = K[\text{DF}(p, t)]X/(1 - X) \quad (3)$$

Equivalent PK/PL/PD dependencies were derived for the irreversible drug–receptor interactions, either single or preceded by a fast, reversible step.<sup>72</sup>

Equation 3 and equivalent PK/PL/PD relationships provide the blueprint for a conceptual combination of the disposition function with any ligand-based or receptor-based 3D-QSAR method for prediction of binding affinities. The  $K$  term in eq 3 needs to be replaced by a suitable conceptual 3D-QSAR correlation equation, and the disposition function can be selected from the available repertoire.<sup>30,32</sup> This cell-QSAR concept represents the recipe for the formulation of conceptual QSARs for cell-based assays and other complex systems.

**Multispecies Multimode Equilibria.** Studied antimycin analogues ionize under the conditions of the experiments, and each species forms different intramolecular H-bonds. The molecules are flexible, and each species can bind to the receptor R in different orientations or conformations (modes). The generalized multispecies multimode binding equilibria are outlined in the following equation:



Each binding equilibrium is characterized by the microscopic association constant  $K_{ijk}$  (eq 5) and represents the receptor binding of the  $i$ th ligand in the  $j$ th species and the  $k$ th binding mode. In principle, the number of species  $s$  and the number of binding modes  $m$  can differ for individual ligands. In the studied case, there are only two species ( $s = 2$ , nonionized and ionized, with charges 0 and  $-1$ , respectively) for each ligand (Table 1). The number of modes is unknown, since there is no experimental information available on either the receptor or the complex, and can differ for individual ligands. In the general expressions,  $s$  and  $m$  represent the maximum numbers of species and binding modes for the tested ligands. The missing species or modes of some ligands will have zero fractions  $f_{ij}$  or zero association constants  $K_{ijk}$  (see below).

The microscopic association constant for the  $i$ th molecule's  $j$ th species in the  $k$ th binding mode is

$$K_{ijk} = \frac{[\text{LR}_{ijk}]}{[L_{ij}][R]} \quad (5)$$

Usually, only the total association constant of the  $i$ th ligand ( $K_i$ ) is determined, although the microscopic association constants  $K_{ijk}$  are the relevant quantities to be correlated with the structure. To obtain a conceptual correlation equation for  $K_i$ , it must be expressed as a function of  $K_{ijk}$ .

The experimentally measured  $K_i$  reflects the total concentration of the ligand/receptor complexes in any species and any binding mode. A series of rearrangements to the definition of  $K_i$  provides the desired

$$\begin{aligned}
 K_i &= \frac{[\text{LR}_i]}{[L_i][R]} \\
 &= \sum_{j=1}^s \frac{[\text{LR}_{ij}]}{[L_i][R]} \\
 &= \sum_{j=1}^s f_{ij} \frac{[\text{LR}_{ij}]}{[L_{ij}][R]} \\
 &= \sum_{j=1}^s f_{ij} \sum_{k=1}^m \frac{[\text{LR}_{ijk}]}{[L_{ij}][R]}
 \end{aligned} \quad (6)$$

function of  $K_{ijk}$ , shown in eq 6: (1) the total concentration of the complex  $[\text{LR}_i]$  is expressed as the sum of bound concentrations of individual species, (2) each summand is formally multiplied by  $[L_{ij}]/[L_i]$  to introduce the fraction of each species,  $f_{ij} = [L_{ij}]/[L_i]$ , and (3) the bound concentration of each species  $[\text{LR}_{ij}]$  is broken down to individual binding modes. The relationship between the microscopic and total association constants is obtained by expressing the  $m$ -summands in eq 6 using eq 5:

$$K_i = \sum_{j=1}^s f_{ij} \sum_{k=1}^m K_{ijk} \quad (7)$$

Here  $f_{ij}$  is the fraction of the  $j$ th molecular species in the  $i$ th compound in solution. For compounds present in the solution as one species, the right-hand side of eq 7 is equal to the  $k$ -summation. In this case, eq 7 is in accordance with published analyses of formally analogous situations: the statistical thermodynamic<sup>73</sup> and equilibrium<sup>58,59</sup> treatment of multimode binding in ligand/protein interactions and kinetic analyses of a reversible unimolecular reaction leading to different products<sup>74</sup> or isomers.<sup>75</sup> The expressions for the fractions  $f_{ij}$  of the  $j$ th species can be derived from the definition of the ionization constants.

Only one molecule of a compound, in any species and any mode, can be bound in each complex. The distribution of species and modes in the bound state is characterized by their prevalences. For the  $k$ th mode of the  $j$ th bound species for the  $i$ th compound, the prevalence is

$$\frac{[\text{LR}_{ijk}]}{[\text{LR}_i]} = \frac{K_{ijk}[L_{ij}][R]}{K_i[L_i][R]} = \frac{K_{ijk}f_{ij}}{K_i} \approx \frac{K_{ijk}f_{ij}}{\sum_{j=1}^s f_{ij} \sum_{k=1}^m K_{ijk}} \quad (8)$$

The numerator and denominator come from eqs 5 and 6, respectively. The third quasi-equality ensures that the sum of all  $s \times m$  prevalences equals unity. The prevalence of the  $j$ th bound species in all modes and the prevalence of the  $k$ th bound mode in all species can easily be calculated using eq 8.

For the application in CoMFA, eq 7 needs to be logarithmized and the  $\log K_{ijk}$  values must be replaced by the correlation equation for CoMFA:

$$\log K_i = \log \sum_{j=1}^s f_{ij} \sum_{k=1}^m 10^{\log K_{ijk}} = \log \sum_{j=1}^s f_{ij} \sum_{k=1}^m 10^{\sum_{l=1}^{2g} C_l X_{ijkl} + C_0} \quad (9)$$

Here  $C_l$  are the regression coefficients characterizing the weights of the ligand–probe interaction energies  $X$  in individual cross-sections of a grid surrounding the superimposed ligands. The subscripts indicate the relatedness to the  $i$ th compound,  $j$ th species,  $k$ th mode, and  $l$ th grid point, and  $g$  is the total number of grid points. Other contributions to the binding free energy (internal ligand energy, binding entropy, desolvation) can be added to the summation in the exponents. In this case, we only used the CoMFA energies because the size of the used

data set did not allow a rigorous optimization of more than a dozen coefficients.

**Intracellular Disposition.** The pseudoequilibrium DF expresses the time course of the concentration of nonionized ligand molecules in the aqueous phases of the studied biosystem during the elimination period:

$$DF(p, t) = \frac{1}{AP^\beta + B} e^{-(CP^\beta + D)t/(AP^\beta + B)} \quad (10)$$

Here  $P$  is the reference partition coefficient,  $\beta$  is the empirical Collander coefficient, and  $t$  is the exposure time. The terms  $A$ ,  $B$ ,  $C$ , and  $D$  characterize the key fate-determining processes the ligand molecules undergo in the biological system:  $A$ , accumulation in membranes/protein binding;  $B$ , distribution in aqueous phases;  $C$ , lipophilicity-dependent elimination;  $D$ , lipophilicity-independent elimination. For the receptors localized in the bilayer, eq 10 needs to be multiplied by  $P^\beta$ .

If compounds are present in the aqueous phase in the receptor surroundings in  $s$  species created by ionization, tautomerism, or carbonyl hydration, the model suggests that each of the parameters  $A$ ,  $B$ ,  $C$ , and  $D$  (represented by  $Y$ ) from eq 10 be expanded as<sup>76</sup>

$$Y = \sum_{j=1}^s f_j Y_j \quad (11)$$

The fractions  $f$  of individual species are the same quantities as in eq 9, except the subscripts  $i$ , indicating the compound, are omitted. The subscripts  $j$  denote, as in eq 9, the quantities associated with the  $j$ th species. The fractions of nonionized, neutral molecules (subscript  $j = 1$ ) and ionized molecules with an overall charge of  $-1$  (subscript  $j = 2$ ) are given by the definition of the dissociation constant  $K_a$  for the given pH of the medium and can be written, respectively, as

$$f_1 = \frac{10^{pK_a - \text{pH}}}{1 + 10^{pK_a - \text{pH}}} \quad \text{and} \quad f_2 = \frac{1}{1 + 10^{pK_a - \text{pH}}} \quad (12)$$

Generally, on the basis of the probable intracellular localization of the receptors, a reasonable intracellular pH range is 5–8, the lower limit being observed in lysosomes and the upper limit in the mitochondrial matrix. The pH value of the cultivation medium could affect these values, but the exact pH of the assay was not published.<sup>52</sup> For this reason and because the receptor localization is unknown, the pH value in eq 12 was treated as one of the adjustable coefficients.

**Data Set.** The Selwood data set includes in vitro filaricidal activities against *D. viteae* of 31 antimycin analogues (Table 1), which were determined by monitoring the leakage of incorporated radiolabeled adenine from intact female *D. viteae* after 120 h of exposure.<sup>36,52</sup> The EC<sub>50</sub> values for adenine leakage were calculated after 120 h of exposure to a range of drug concentrations at 37 °C under neutral conditions.<sup>52</sup>

The 1-octanol/water partition coefficients  $P$  were calculated by the ClogP program in Bio-Loom<sup>51</sup> software for neutral molecules and are summarized in Table 1. The values differ from the original data<sup>77</sup> because of the developments in the ClogP software. The differences are within  $\pm 0.25$  log unit, except for compounds **6** (Table 1,  $\log P_{\text{old}} - \log P_{\text{new}} = 1.526$ ), **10** ( $-0.770$ ), **12** ( $-0.638$ ), **14** ( $-1.654$ ), and **26** ( $0.325$ ). The  $\log P$  values were used with the belief that they properly reflect the relative lipophilicity of the studied compounds, although the values of  $\log P > 7$  were marked in the ClogP output as unrealistically high. Ionization decreases the  $\log P$  values by several units, and most compounds were present mainly as ionized species under physiologic conditions.

Our multispecies approach can take into account tautomers and hydrates of aldehydes and ketones; hence, formation of these species was examined using the SPARC Web server<sup>50</sup> and was found negligible for all compounds on the basis of the following reasoning.

As shown in eq 7, the binding contribution of each species is given by the product of the association constant  $K_{ij}$  (represented by the sum of the association constants  $K_{ijk}$  for individual modes<sup>61</sup>) and the species fraction  $f_{ij}$ . The fraction limit for a species to be included in the correlation was set to  $10^{-4}$  on the basis of the following considerations.

The interaction of a rare species with the receptor can release about 2–3 kcal/mol more free energy than that of the prevalent species because both species bind by weak interactions and differ in a single, well-defined structural feature. For this situation, the product  $f_{ij}K_{ij}$  for the rare species would be less than 1% of the same quantity for the prevalent species.

**Superposition.** Pharmacophore models were built using the Galahad procedure<sup>54,55</sup> in Sybyl-X. The ligands are first aligned flexibly to each other in torsional space, and then the conformations produced are rigidly aligned in Cartesian space. Pharmacophore and steric bitmaps (TUPLET<sup>78</sup>) representing H-bonding and steric terms, respectively, along with the total ligand energy, serve as input to a multiobjective fitness function. Produced pharmacophores indicated that the salicylamide moieties of all compounds in all species are superimposed in all cases. In the absence of structural and other information on the receptor, the most active and, at the same time, one of most rigid analogues, compound **7** (Figure 2, Table 1), was selected as the template for superposition.

Conformational analysis was performed using the Tripos force field in Sybyl<sup>79</sup> with an increment of  $10^\circ$  for all considered torsions. When the molecule forms an intramolecular H-bond (Figure 1), either as a neutral species (OH...O=C) or as an ionized species (NH...O<sup>-</sup>), the first torsion  $\Phi_1$  is fixed and only  $\Phi_2$ ,  $\Phi_3$ , and  $\Phi_4$  are allowed to rotate in conformational analysis. In the opposite case, when analogues have no intramolecular H-bond, all four torsions in Figure 2 are considered rotatable. The approximate minimum conformations were further optimized and Mulliken charges were calculated in Jaguar<sup>80</sup> using the DFT/B3LYP method with a 6-31G\*\* basis set. Charge and spin multiplicity were entered manually for different protonation states. In this way, the conformations of the template for the superposition were obtained. Open conformation, the mode without the intramolecular H-bond has the same optimal geometries for both neutral and ionized species.

In the MM approach, the user is allowed to enter several possible bound conformations of the template in one optimization run. The MSMM approach extends this option to all species which are present in the solution surrounding the receptor. The four basic template molecules of compound **7** were superimposed by the atom fit method in Sybyl (Figure 3a). All studied compounds do not share a common skeleton; therefore, for each compound, both species in all modes were superimposed on the templates using the FlexS procedure<sup>65</sup> as incorporated in Sybyl (Figure 3c). FlexS combines conformational search with alignment in two steps. First, a base fragment of the ligand is selected and placed on the template. Second, as the rest of the ligand parts are incrementally built, torsion angles are perturbed to generate new conformations, and each conformation is scored for superposition quality. The obtained conformations of all compounds were minimized by Jaguar using the DFT/B3LYP method<sup>81</sup> with a 6-31G\*\* basis set<sup>82,83</sup> with all torsions fixed. The Mulliken charges<sup>84</sup> were calculated for the optimized geometries. The superposition was repeated, and the resulting conformations were used in the CoMFA analyses.

**Interaction Energies.** Steric and electrostatic interaction energies  $X_{ijkl}$  with an sp<sup>3</sup> carbon probe with a +1 charge were calculated for each compound (subscript  $i$ ) in each species (subscript  $j$ ) and each binding mode (skeleton conformations  $\times$  ring-flipping conformations, subscript  $k$ ). The probe was placed at each lattice intersection (subscript  $l$ ) of a 3D grid (2 Å in each coordinate direction). The grid has the following coordinates:  $-7$  to  $+18$  Å along the  $x$ -axis,  $1$ – $21$  Å along the  $y$ -axis, and  $-15$  to  $+11$  Å along the  $z$ -axis. An energy cutoff of 30 kcal/mol was used to cap the maximum used energies.

**Regression Analysis.** The correlation equation is obtained by taking the logarithm of eq 3, in which the association constant  $\log K$  is replaced by eq 9 and the disposition function  $DF(p,t)$  is given by eq 10, as combined with eq 11, in which the fractions are expressed by eq 12. The receptor-binding and disposition components have, from the optimization viewpoint, quite different properties: eq 9 has only one form, and the descriptive variables need to be selected from a large pool, while eqs 10–12 contain only two variables, which are organized in one of two possible functional forms, depending upon the location



of the receptors either in the aqueous phase or in membranes. For the receptor-binding part (eq 9), the coefficients  $C_i$  are associated with the grid points and are independent of the species and modes. Therefore, the final set of optimized coefficients is independent of the numbering of individual species and modes. Notably, the inclusion of multiple species and multiple modes does not increase the number of optimized regression coefficients  $C_i$ .

The correlation equation is nonlinear in the coefficients  $C_i$  in eq 9 and the coefficients  $A$ – $D$ ,  $\text{pH}$  and  $\beta$  in eqs 10–12. All statistical procedures<sup>61</sup> were written in C programming language and integrated in Sybyl through the Sybyl Programming Language (SPL) scripts. The choice of one of the two implemented optimization methods depends on the total number ( $N$ ) of coefficients  $C$  in the calibrated model and on the number of compounds ( $n$ ). Nonlinear regression analysis is applied when  $N \leq n/2$ . Otherwise, the developed software automatically switches to the PLS analysis with a linearized form<sup>61</sup> of eq 9. The second option was not used for optimization of the best models. Both optimization methods are sensitive to good initial estimates. Therefore, a forward-selection procedure, conceptually similar to that used before for multimode analysis,<sup>61</sup> was a rational choice.

Before analysis, the  $X_{jkl}$  variables (energy columns) were sorted using the following criteria: (1) high and sustained variability, (2) the minimal number of colinear columns with similar information, and (3) even distribution of selected grid points around the ligands. Sustained variability means that the energy values cover the whole interval more or less evenly and are not clustered. An extreme case is the situation when all energies in a given column are equal (usually zero) and only one mode has a different energy. These singularities originate when only one molecule protrudes into a subspace and other molecules do not affect the probe interaction energies in that subspace. Singular molecules need to be eliminated from the analysis because they do not provide statistically significant information about the poorly covered subspace.

The model calibration can be steered by user-defined inputs, for which the default values are shown in parentheses. The process starts with a quick scan through the models consisting of a few (four) variables which are randomly selected from the top variables (10%). The initial values of the coefficients ( $\pm 0.1$ ) are systematically evaluated by a grid search. In this step, no optimization is applied and the correlation equation is evaluated for the initial coefficients. For the best models (top 10%  $r^2$ ), the coefficients are optimized. At each of the following steps, a reduction of the number of coefficients is attempted: the coefficients, which do not lead to a decrease in the  $r^2$  value, are eliminated. The best models (top 5%  $r^2$ ) are developed by a gradual addition of groups of a few (three) variables until all variables from the selected set (top 10%) are used. The procedure is finalized by fine-tuning of the best models (top 5%  $r^2$ ) by addition of single variables. The best model is selected on the basis of the following criteria:  $N$ ,  $r^2$ ,  $q^2$  for the leave-one-out cross-validation (applied to the training set only to eliminate unstable models), and the standard errors of the parameters.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Prevalences of all species and modes as described in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: stefan.balaz@acphs.edu. Phone: (802) 735-2615. Fax: (802) 654-0716.

### Present Address

§Simulations Plus, Inc., Lancaster, CA 93534.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported in part by NIH NIGMS Grant R01 GM80508, NIH NCRG Grants P20 RR 15566 and P20 RR 16471, the NSF EPSCoR EPS-0814442 program, and access to Teragrid computation resources for Projects MCB100078 and MCB110017.

## ■ ABBREVIATIONS USED

$A$ , disposition function term for accumulation in lipids and protein binding;  $\beta$ , Collander coefficient;  $B$ , disposition function term for accumulation in aqueous phases;  $C$  and  $D$ , disposition function terms for elimination;  $c_0$ , initial concentration;  $c_X$ , isoeffective concentration;  $C_i$ , regression coefficients in the CoMFA expression;  $DF$ , disposition function;  $E$ , drug effect;  $E_{\text{max}}$ , maximum drug effect;  $f$ , species fraction;  $g$ , number of grid points;  $K$ , association constant;  $L$ , ligand;  $LR$ , ligand–receptor complex;  $m$ , number of modes;  $MM$ , multimode;  $MS$ , multispecies;  $n$ , number of compounds;  $N$ , number of coefficients;  $P$ , partition coefficient;  $PD$ , pharmacodynamics;  $PK$ , pharmacokinetics;  $PL$ , pharmacology;  $PRESS$ , predictive sum of squares of deviations;  $q^2$ , squared predictive correlation coefficient;  $r^2$ , squared descriptive correlation coefficient;  $R$ , receptor;  $s$ , number of species;  $SM$ , single mode;  $SS$ , single species;  $SSE$ , sum of squares of errors;  $SY$ , sum of squares of deviations from the average;  $X$ , interaction energy

## ■ REFERENCES

- (1) Gilson, M. K.; Zhou, H. X. Calculation of Protein-Ligand Binding Affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (2) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (3) Aqvist, J.; Medina, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. *Protein Eng.* **1994**, *7*, 385–391.
- (4) Hansson, T.; Aqvist, J. Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137–1144.
- (5) Aqvist, J. Calculation of Absolute Binding Free Energies for Charged Ligands and Effects of Long-Range Electrostatic Interactions. *J. Comput. Chem.* **1996**, *17*, 1587–1597.
- (6) Gilson, M. K.; Given, J. A.; Head, M. S. A New Class of Models for Computing Receptor-Ligand Binding Affinities. *Chem. Biol.* **1997**, *4*, 87–92.
- (7) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047–1069.
- (8) Head, M. S.; Given, J. A.; Gilson, M. K. Mining Minima: Direct Computation of Conformational Free Energy. *J. Phys. Chem. A* **1997**, *101*, 1609–1618.
- (9) Chang, C. E.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156–13164.
- (10) Cramer, R. D., III; 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*, 5959–5967.
- (11) Thomas, B. F.; Compton, D. R.; Martin, B. R.; Semus, S. F. Modeling the Cannabinoid Receptor: A Three-Dimensional Quantitative Structure-Activity Analysis. *Mol. Pharmacol.* **1991**, *40*, 656–665.
- (12) McFarland, J. W. Comparative Molecular Field Analysis of Anticoccidial Triazines. *J. Med. Chem.* **1992**, *35*, 2543–2550.
- (13) Hansch, C.; Fujita, T.  $\rho$ – $\sigma$ – $\pi$  Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.

- (14) Higuchi, T.; Davis, S. S. Thermodynamic Analysis of Structure-Activity Relationships of Drugs: Prediction of Optimal Structure. *J. Pharm. Sci.* **1970**, *59*, 1376–1383.
- (15) Kubinyi, H. Quantitative Structure–Activity Relationships. 7. The Bilinear Model, a New Model for Nonlinear Dependence of Biological Activity on Hydrophobic Character. *J. Med. Chem.* **1977**, *20*, 625–629.
- (16) Dearden, J. C.; Townend, M. S. Digital Computer Simulation of the Drug Transport Process. In *Quantitative Structure-Activity Analysis*; Franke, R., Oehme, P., Eds.; Akademie-Verlag: Berlin, 1978; pp 387–393.
- (17) Martin, Y. C.; Hackbarth, J. J. Theoretical Model-Based Equations for the Linear Free Energy Relationships of the Biological Activity of Ionizable Substances. 1. Equilibrium-Controlled Potency. *J. Med. Chem.* **1976**, *19*, 1033–1039.
- (18) Balaz, S.; Sturdik, E.; Rosenberg, M.; Augustin, J.; Skara, B. Kinetics of Drug Activities As Influenced by Their Physico-Chemical Properties: Antibacterial Effects of Alkylating 2-Furylethylenes. *J. Theor. Biol.* **1988**, *131*, 115–134.
- (19) Seiler, P. Interconversion of Lipophilicities from Hydrocarbon/Water Systems into the Octanol/Water System. *Eur. J. Med. Chem.* **1974**, *9*, 473–479.
- (20) Burton, P. S.; Conradi, R. A.; Hilgers, A. R.; Ho, N. F. H.; Maggiora, L. L. The Relationship between Peptide Structure and Transport across Epithelial Cell Monolayers. *J. Controlled Release* **1992**, *19*, 87–97.
- (21) Chikhale, E. G.; Ng, K. Y.; Burton, P. S.; Borchardt, R. T. Hydrogen Bonding Potential as a Determinant of the *In Vitro* and *In Situ* Blood-Brain Barrier Permeability of Peptides. *Pharm. Res.* **1994**, *11*, 412–419.
- (22) Caron, G.; Ermondi, G. Calculating Virtual log *P* in the Alkane/Water System (log  $P_{\text{alk}}^{\text{N}}$ ) and Its Derived Parameters  $\Delta \log P_{\text{oct-alk}}^{\text{N}}$  and log  $D_{\text{alk}}^{\text{PH}}$ . *J. Med. Chem.* **2005**, *48*, 3269–3279.
- (23) Van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Chretien, J. R.; Raevsky, O. A. Estimation of Blood-Brain Barrier Crossing of Drugs Using Molecular Size and Shape, and H-Bonding Descriptors. *J. Drug Targeting* **1998**, *6*, 151–165.
- (24) Egan, W. J.; Lauri, G. Prediction of Intestinal Permeability. *Adv. Drug Delivery Rev.* **2002**, *54*, 273–289.
- (25) Lukacova, V.; Peng, M.; Tandlich, R.; Hinderliter, A.; Balaz, S. Partitioning of Organic Compounds in Phases Imitating the Headgroup and Core Regions of Phospholipid Bilayers. *Langmuir* **2006**, *22*, 1869–1874.
- (26) Kulkarni, A. S.; Hopfinger, A. J. Membrane-Interaction QSAR Analysis: Application to the Estimation of Eye Irritation by Organic Compounds. *Pharm. Res.* **1999**, *16*, 1245–1253.
- (27) Iyer, M.; Tseng, Y. J.; Senese, C. L.; Liu, J.; Hopfinger, A. J. Prediction and Mechanistic Interpretation of Human Oral Drug Absorption Using MI-QSAR Analysis. *Mol. Pharmaceutics* **2007**, *4*, 218–231.
- (28) Gratton, J. A.; Abraham, M. H.; Bradbury, M. W.; Chadha, H. S. Molecular Factors Influencing Drug Transfer across the Blood-Brain Barrier. *J. Pharm. Pharmacol.* **1997**, *49*, 1211–1216.
- (29) Fischer, H.; Gottschlich, R.; Seelig, A. Blood-Brain Barrier Permeation: Molecular Parameters Governing Passive Diffusion. *J. Membr. Biol.* **1998**, *165*, 201–211.
- (30) Balaz, S. Subcellular Pharmacokinetics and Drug Properties: Numerical Simulations in Multicompartment Systems. *Quant. Struct.-Act. Relat.* **1994**, *13*, 381–392.
- (31) Balaz, S.; Wiese, M.; Seydel, J. K. A Time Hierarchy-Based Model for Kinetics of Drug Disposition and Its Use in Quantitative Structure-Activity Relationships. *J. Pharm. Sci.* **1992**, *81*, 849–857.
- (32) Balaz, S. Lipophilicity in Trans-Bilayer Transport and Subcellular Pharmacokinetics. *Perspect. Drug Discovery* **2000**, *19*, 157–177.
- (33) Balaz, S.; Lukacova, V. Subcellular Pharmacokinetics and Its Potential for Library Focusing. *J. Mol. Graphics Model.* **2002**, *20*, 479–490.
- (34) Balaz, S.; Sturdik, E.; Durcova, E.; Antalik, M.; Sulo, P. Quantitative Structure-Activity Relationship of Carbonylcyanide Phenylhydrazones as Uncouplers of Mitochondrial Oxidative Phosphorylation. *Biochim. Biophys. Acta* **1986**, *851*, 93–98.
- (35) Chen, V. Y.; Khersonsky, S. M.; Shedden, K.; Chang, Y. T.; Rosania, G. R. System Dynamics of Subcellular Transport. *Mol. Pharmaceutics* **2004**, *1*, 414–425.
- (36) Selwood, D. L.; Livingstone, D. J.; Comley, J. C. W.; O'Dowd, A. B.; Hudson, A. T.; Jackson, P.; Jandu, K. S.; Rose, V. S.; Stables, J. N. Structure–Activity Relationships of Antifilarial Antimycin Analogues: A Multivariate Pattern Recognition Study. *J. Med. Chem.* **1990**, *33*, 136–142.
- (37) Plaisance, I.; Duthe, F.; Sarrouilhe, D.; Herve, J. C. The Metabolic Inhibitor Antimycin A Can Disrupt Cell-to-Cell Communication by an ATP- and  $\text{Ca}^{2+}$ -Independent Mechanism. *Pfluegers Arch.* **2003**, *447*, 181–194.
- (38) King, M. A. Antimycin A-induced Killing of HL-60 Cells: Apoptosis Initiated from within Mitochondria Does Not Necessarily Proceed Via Caspase 9. *Cytometry, A* **2005**, *63A*, 69–76.
- (39) Kubinyi, H. Variable Selection in QSAR Studies. I. An Evolutionary Algorithm. *Quant. Struct.-Act. Relat.* **1994**, *13*, 285–294.
- (40) Kubinyi, H. Variable Selection in QSAR Studies. II. A Highly Efficient Combination of Systematic Search and Evolution. *Quant. Struct.-Act. Relat.* **1994**, *13*, 393–401.
- (41) McFarland, J. W.; Gans, D. J. On Identifying Likely Determinants of Biological Activity in High Dimensional QSAR Problems. *Quant. Struct.-Act. Relat.* **1994**, *13*, 11–17.
- (42) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure–Activity Relationships and Quantitative Structure–Property Relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
- (43) So, S. S.; Karplus, M. Evolutionary Optimization in Quantitative Structure–Activity Relationships: An Application of Genetic Neural Networks. *J. Med. Chem.* **1996**, *39*, 1521–1530.
- (44) Waller, C. L.; Bradley, M. P. Development and Validation of a Novel Variable Selection Technique with Application to Multidimensional Quantitative Structure–Activity Relationship Studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 345–355.
- (45) Cho, S. J.; Hermsmeier, M. A. Genetic Algorithm Guided Selection: Variable Selection and Subset Selection. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 927–936.
- (46) Saxena, A. K.; Prathipati, P. Comparison of MLR, PLS and GA-MLR in QSAR Analysis. *SAR QSAR Environ. Res.* **2003**, *14*, 433–445.
- (47) Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. Detecting “Bad” Regression Models: Multicriteria Fitness Functions in Regression Analysis. *Anal. Chim. Acta* **2004**, *515*, 199–208.
- (48) Jabalpurwala, K. E.; Venkatachalam, K. A.; Kabadi, M. B. Proton-Ligand Stability Constants of Some Ortho-Substituted Phenols. *J. Inorg. Nucl. Chem.* **1964**, *26*, 1011–1026.
- (49) Tosi, C. Ultraviolet Spectrum of Salicylanilide. *Spectrochim. Acta, A* **1968**, *24*, 743–746.
- (50) Hilal, S. H.; Karickhoff, S. W.; Carreira, L. A. Estimation of Microscopic, Zwitterionic Ionization Constants, Isoelectric Point and Molecular Speciation of Organic Compounds. *Talanta* **1999**, *50*, 827–840.
- (51) *Bio-Loom*; BioByte Corp.: Claremont, CA, 2006.
- (52) Comley, J. C. W.; Rees, M. J.; O'Dowd, A. B. Leakage of Incorporated Radiolabelled Adenine—a Marker for Drug-Induced Damage of Macrophilic In Vitro. *Trop. Med. Parasitol.* **1988**, *39*, 221–226.
- (53) Natesan, S.; Wang, T.; Lukacova, V.; Bartus, V.; Khandelwal, A.; Balaz, S. Rigorous Treatment of Multispecies Multimode Ligand–Receptor Interactions in 3D-QSAR: CoMFA Analysis of Thyroxine Analogs Binding to Transthyretin. *J. Chem. Inf. Model.* **2011**, *51*, 1132–1150.
- (54) Richmond, N. J.; Willett, P.; Clark, R. D. Alignment of Three-Dimensional Molecules Using an Image Recognition Algorithm. *J. Mol. Graphics Modell.* **2004**, *23*, 199–209.

- (55) Richmond, N. J.; Abrams, C. A.; Wolohan, P. R. N.; Abrahamian, E.; Willett, P.; Clark, R. D. GALAHAD: 1. Pharmacophore Identification by Hypermolecular Alignment of Ligands in 3D. *J. Comput.-Aided Mol. Des.* **2006**, *20*, 567–587.
- (56) Marshall, G. R. Binding-Site Modeling of Unknown Receptors. In *3D-QSAR in Drug Design: Theory, Methods, and Applications*; Kubinyi, H., Ed.; Escom: Leiden, The Netherlands, 1993; pp 80–116.
- (57) Miyoshi, H.; Tokutake, N.; Imaeda, Y.; Akagi, T.; Iwamura, H. A Model of Antimycin A Binding Based on Structure-Activity Studies of Synthetic Antimycin A Analogues. *Biochim. Biophys. Acta, Bioenerg.* **1995**, *1229*, 149–154.
- (58) Balaz, S. Quantitative Structure-Time-Activity Relationships for Single and Continuous Dose. *Chemom. Intell. Lab. Syst.* **1994**, *24*, 177–183.
- (59) Hornak, V.; Balaz, S.; Schaper, K. J.; Seydel, J. K. Multiple Binding Modes in 3D-QSAR: Microbial Degradation of Polychlorinated Biphenyls. *Quant. Struct.-Act. Relat.* **1998**, *17*, 427–436.
- (60) Clare, B. W. QSAR of Benzene Derivatives: Comparison of Classical Descriptors, Quantum Theoretic Parameters and Flip Regression, Exemplified by Phenylalkylamine Hallucinogens. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 611–633.
- (61) Lukacova, V.; Balaz, S. Multimode Ligand Binding in Receptor Site Modeling: Implementation in CoMFA. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2093–2105.
- (62) Supuran, C. T.; Clare, B. W. Quantum Theoretic QSAR of Benzene Derivatives: Some Enzyme Inhibitors. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 237–248.
- (63) Clare, B. W.; Supuran, C. T. A Physically Interpretable Quantum-Theoretic QSAR for Some Carbonic Anhydrase Inhibitors with Diverse Aromatic Rings, Obtained by a New QSAR Procedure. *Bioorg. Med. Chem.* **2005**, *13*, 2197–2211.
- (64) Clare, B. W.; Supuran, C. T. Predictive Flip Regression: A Technique for QSAR of Derivatives of Symmetric Molecules. *J. Chem. Inf. Model.* **2005**, *45*, 1385–1391.
- (65) Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: A Method for Fast Flexible Ligand Superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
- (66) Ramachandran, S.; Gottlieb, D. Mode of Action of Antibiotics. II. Specificity of Action of Antimycin A and Ascocin. *Biochim. Biophys. Acta* **1961**, *53*, 396–402.
- (67) von Jagow, G.; Bohrer, C. Inhibition of Electron Transfer from Ferrocycytochrome B to Ubiquinone, Cytochrome C1 and Duroquinone by Antimycin. *Biochim. Biophys. Acta* **1975**, *387*, 409–424.
- (68) Kim, H.; Esser, L.; Hossain, M. B.; Xia, D.; Yu, C. A.; Rizo, J.; van der Helm, D.; Deisenhofer, J. Structure of Antimycin A1, a Specific Electron Transfer Inhibitor of Ubiquinol–Cytochrome *c* Oxidoreductase. *J. Am. Chem. Soc.* **1999**, *121*, 4902–4903.
- (69) Tokutake, N.; Miyoshi, H.; Nakazato, H.; Iwamura, H. Inhibition of Electron Transport of Rat-Liver Mitochondria by Synthesized Antimycin A Analogs. *Biochim. Biophys. Acta, Bioenerg.* **1993**, *1142*, 262–268.
- (70) von Jagow, G.; Link, T. A.; Ohnishi, T. Organization and Function of Cytochrome B and Ubiquinone in the Cristae Membrane of Beef Heart Mitochondria. *J. Bioenerg. Biomembr.* **1986**, *18*, 157–179.
- (71) Kenakin, T. *Pharmacologic Analysis of Drug-Receptor Interaction*; Lippincott-Raven Publishers: Philadelphia, PA, 1997.
- (72) Balaz, S.; Sturdik, E.; Tichy, M. Hansch Approach and Kinetics of Biological Activities. *Quant. Struct.-Act. Relat.* **1985**, *4*, 77–81.
- (73) Wang, J.; Szewczuk, Z.; Yue, S. Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. Calculation of Relative Binding Free Energies and Configurational Entropies: A Structural and Thermodynamic Analysis of the Nature of Non-Polar Binding of Thrombin Inhibitors Based on Hirudin55-65. *J. Mol. Biol.* **1995**, *253*, 473–492.
- (74) Jullien, L.; Proust, A.; LeMenn, J. C. How Does the Gibbs Free Energy Evolve in a System Undergoing Coupled Competitive Reactions? *J. Chem. Educ.* **1998**, *75*, 194–199.
- (75) Smith, W. R.; Missen, R. W. *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*; John Wiley and Sons: New York, 1982; pp 1–364.
- (76) Balaz, S.; Cronin, M. T. D.; Dearden, J. C. Relationship between Structure of Ionizable Bioactive Compounds and Their in Vitro Biological Activity Measured at Varying pH of the Medium. *Pharm. Sci. Commun.* **1993**, *4*, 51–58.
- (77) Kubinyi, H. Evolutionary Variable Selection in Regression and PLS Analyses. *J. Chemom.* **1996**, *10*, 119–133.
- (78) Abrahamian, E.; Fox, P. C.; Narum, L.; Christensen, I. T.; Thogersen, H.; Clark, R. D. Efficient Generation, Storage, and Manipulation of Fully Flexible Pharmacophore Multiplets and Their Use in 3-D Similarity Searching. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 458–468.
- (79) *Sybyl 7.2*; Tripos Inc.: St. Louis, MO, 2009.
- (80) *Jaguar*; Schrödinger LLC: Portland, OR, 2003.
- (81) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (82) Hay, P. J.; Wadt, W. R. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Scandium to Mercury. *J. Chem. Phys.* **1985**, *82*, 270–283.
- (83) Wadt, W. R.; Hay, P. J. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for Main Group Elements Sodium to Bismuth. *J. Chem. Phys.* **1985**, *82*, 284–298.
- (84) Mulliken, R. S. Electronic Population Analysis on LCAO-MO (Linear Combination of Atomic Orbital-Molecular Orbital) Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833–1840.