A New Approach To Predict the Biological Activity of Molecules Based on Similarity of Their Interaction Fields and the log*P* and log*D* Values: Application to Auxins

Branimir Bertoša,[†] Biserka Kojić-Prodić,[†] Rebecca C. Wade,[‡] Michael Ramek,[§] Stavroula Piperaki,[||] Anna Tsantili-Kakoulidou,[||] and Sanja Tomić*,[†]

Ruđer Bošković Institute, P.O.B. 180, HR-10002 Zagreb, Croatia, European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany, Institut für Physikalische und Theoretische Chemie, Technische Universität Graz, A-8010 Graz, Austria, and Department of Pharmaceutical Chemistry, School of Pharmacy, University of Athens, Panepistimiopolis, Zografou, GR-157 71, Athens, Greece

Received April 7, 2003

The activity of a biological compound is dependent both on specific binding to a target receptor and its ADME (Absorption, Distribution, Metabolism, Excretion) properties. A challenge to predict biological activity is to consider both contributions simultaneously in deriving quantitative models. We present a novel approach to derive QSAR models combining similarity analysis of molecular interaction fields (MIFs) with prediction of log*P* and/or log*D*. This new classification method is applied to a set of about 100 compounds related to the auxin plant hormone. The classification based on similarity of their interaction fields is more successful for the indole than the phenoxy compounds. The classification of the phenoxy compounds is however improved by taking into account the influence of the log*P* and/or the log*D* values on biological activity. With the new combined method, the majority (8 out of 10) of the previously misclassified derivatives of phenoxy acetic acid are classified in accord with their bioassays. The recently determined crystal structure of the auxin-binding protein 1 (ABP1) enabled validation of our approach. The results of docking a few auxin related compounds with different biological activity to ABP1 correlate well with the classification based on similarity of MIFs only. Biological activity is, however, better predicted by a combined similarity of MIFs + log*P*/log*D* approach.

INTRODUCTION

Among the five known plant hormones,[1] auxins were the first identified and are the most widely investigated. The effects of auxins at the cellular level and their consequent influence on overall plant development are well established. However, the mechanisms of auxin activity at the molecular level are poorly understood. Studies on auxin transport[2] and auxin function in the degradation of regulated proteins[3] indicate that the auxin response pathway is much more complicated than expected and that it probably embodies a complex network rather than a linear pathway (see below). Immediate responses to auxins include changes in protoplast electrophysiology, guard cell gating, and early-response-gene induction. Longer-term responses include cell elongation, cell reorganization/division, and cell differentiation. There is strong evidence that auxin-binding protein 1 (ABP1), a high affinity auxin receptor, is a component of the auxin response pathway, which mediates cell expansion. On the other hand, cell division is stimulated via a different, as yet unidentified auxin pathway, possibly mediated by a low-affinity auxin receptor.[4] The AUX/IAA transcriptional regulators (comprising a large family of proteins) synthesized shortly after auxin treatment[3] repress the auxin response pathway. Recently Gray et al.[5] have proposed that auxins unblock the pathway by promoting the binding of the AUX/IAA transcriptional regulators to the ubiquitin-ligase SCF^{TIR1} and related SCF complexes (which are named after their main components, Skip I, Cullin, and an F-box protein), and consequently leading to the degradation of the transcriptional regulators.

The importance of auxins, the recent findings relating to their response pathway, and new facts revealed by the structure of ABP1[6] motivated us to investigate computationally this chemically diverse class of compounds. We have previously developed a robust QSAR method for predicting the biological activity of compounds based on the similarity of their molecular interaction fields and used it to perform a classification of about 70 auxin-related compounds into four activity classes.[7,8] Although the method was successful for the majority of compounds, the metabolic stability and the bioavailability of the molecules were not considered in that work. In this work, we combined our previously derived QSAR method with lipophilicity, log*P*, and apparent lipophilicity expressed as log*D* to develop a new activity prediction method. For this purpose, different log*P* (or log*D*) prediction methods have been incorporated, based on traditional fragmental procedures, electrotopological state indices, and the recently developed computational procedure, Volsurf. Volsurf has been developed by Cruciani et al.[9,10] as a tool to predict molecular ADME properties, by combining calculation of one-dimensional and two-dimensional molecular descriptors from the molecular interaction fields determined by GRID[11] with chemometric procedures. For auxin-related compounds, the log*P* values predicted by this procedure are mostly represented by four descriptors related to the size and shape of a molecule (volume, surface, ratio

* Corresponding author phone: +385 1 4561025; fax: +385 1 4680245; e-mail: sanja.tomic@irb.hr.

[†] Ruđer Bošković Institute.
[‡] European Media Laboratory.
[§] Technische Universität Graz.
[||] University of Athens.

10.1021/ci034063n CCC: $25.00
Published on Web 08/05/2003

of volume and surface, and globularity) and are conformation-dependent.

Waterbeemd et al.[12] have considered hydrogen bonding capacity and molecular size to be the most important properties for absorption and permeability. They prompted us to explore the use of volume and hydrogen bonding capacity descriptors for prediction of the permeability of auxin-like molecules.

The final MIF/log*P* (or log*D*) combined method turned out to be more reliable for predicting the biological activity, especially that of the phenoxy compounds. It was used to improve classification of the auxin-like compounds previously classified and also to classify additional 30 diverse compounds with measured auxin activity. Furthermore, this method permits simultaneous consideration of specific and nonspecific influences on the auxin-response pathway.

MATERIAL AND METHODS

Molecular Modeling. For each compound, two low-energy conformers, with the carboxylate side chain either approximately tilted ('T') or coplanar ('P') to the conjugated ring plane, were modeled using the INSIGHT[13] and DISCOVER[14] programs. The choice of low-energy conformer was based on ab initio quantum mechanics calculations for a series of indole-3-acetic acid derivatives.[15−20] Geometry optimization was performed by steepest descent and conjugate gradients protocols (until a final energy derivative ≤ 0.42 kJ/(mol·Å) (0.1 kcal/(mol·Å)) was achieved) using the CVFF[21] force field. Relatively mild optimization during molecular mechanics calculations was used in order to avoid the compound conformation to escape from its local minimum defined by the position of the carboxyl chain toward the conjugated ring plain.

MIF Calculation. The GRID[11] program was used to calculate molecular interaction fields (MIFs). For each MIF, the interaction energy between a molecule of known three-dimensional structure ("target") and a small chemical group ("probe") was computed for a set of positions of the probe around the target. The selection of probes was carried out so as to mimic functional groups that may be present in an auxin receptor and/or transport-protein binding site. Altogether, four probes were used: H_2O, $-NH_2^+$, $-CH_3$, and =O probe. For each probe, four sets of MIFs were calculated: at the nodes of the three regular three-dimensional grids (see the following chapter) and one at the molecular surface defined by a 2 Å radius probe. A grid with 1 Å spacing (for Volsurf, 0.5 Å spacing) was used. The size of the grid was adjusted to extend at least 4 Å in all directions from the surface of the largest molecule in the set when this molecule was placed at the center of the grid. A dielectric constant of 4.0 was assigned to the whole grid (i.e. inside and outside the molecule). This choice was made considering that our calculations represent the interaction of the auxin-related compound with its protein receptor or a cell membrane. The calculated MIFs (computed with a maximum energy cutoff of +125.6 kJ/mol (30 kcal/mol)) contain information about the physical and chemical properties of auxin-related compounds. They were further used to classify a series of auxin-related compounds as well as to derive a model for log*P* prediction.

Alignment of Molecules. To use the MIFs calculated at points on a three-dimensional grid in similarity analysis, the molecules had to be aligned. The alignment of the molecules is important for MIF similarity index (SI) analysis, since it should correspond with the orientation of the molecules in the binding site of the auxin receptor. In this work, we used two alignment procedures, one based on the similarity of the MIFs at the predefined molecular surface and the other based on the similarity of atomic properties. In the first procedure, which was developed by us,[7] the similarity of the MIFs calculated at the surfaces of the target molecule (2-Me-4-Cl-PAA) and the molecule to be aligned was optimized. The second procedure is that one implemented in the SEAL program.[22] For this latter alignment, partial atomic charges from the CFF91 force field parameter set[23] were used. Two different alignments were performed with the SEAL program: (a) electrostatic (E) and (b) steric-electrostatic (SE). In the E alignment, the carboxyl groups of the main pharmacophore in auxin molecules tend to superimpose well. In both alignments, 4-Cl-IAA was used as the target compound, on which all the other compounds were aligned (for details see ref 7).

Table 1. Compounds Representing Auxin-Activity Classes

class	kernel compounds[a]
1	1-NAA
1	2-Me-4-Cl-PAA
1	IAA
1	4-Cl-IAA
2	4,7-Cl_2-IAA
2	3-IPA
3	BA
4	4-Cl-PIBA
4	5,7-Cl_2-IIBA

[a] PAA: phenoxy acetic acid; NAA: naphthalene acetic acid; IAA: indole-3-acetic acid; BA: benzoic acid; 3-IPA: 3-indole-3-propionic acid; PIBA: phenoxy isobutyric acid; IIBA: indole-3-isobutyric acid.

For each conformer, MIFs were calculated on three different sets of grid points.

Classification. Earlier, we developed a classification method based on similarity of molecular interaction fields.[7] For each of 43 auxin-related molecules in the initial data set, we calculated MIFs using the program GRID. From the $(43 \times 43)/2$ pairwise similarity indices calculated for different probes at each set of points (see the previous paragraphs: "Alignment of molecules", "MIF calculation"), the compounds were classified into four classes. Inspection of the biological activity of the compounds within the individual classes revealed that molecules in the same class had similar biological activity. Therefore, the four classes of auxin-related molecules were defined according to the results of the similarity based classification and the measured biological activities. The four classes are as follows: class 1 consists of compounds active over a wide range of concentrations ($10^{-8}-10^{-4}$ mol/dm^3); class 2 includes compounds with very low activity, which at some concentrations (within the range of concentrations widely used in biological tests) show weak antiauxin behavior; class 3 consists of the compounds inactive over a wide range of concentrations; and class 4 consists of compounds with pronounced inhibitory characteristics.

To facilitate the classification of further auxin-related molecules, we represented classes by their minimum representative sets, class kernels (Table 1). The kernel compounds

Table 2. Log*P* Values Predicted by the Conformationally Dependent Volsurf Models, Pallas 2.1 and IA for the Compounds with Available Experimental Values

code[a]	compound[b]	log*P*[h 30]	log*P* model 1[c]	log*P* model 2[d]	log*P* CDR[e]	log*P* IA[f]
N1	3-n-But-PAA	3.18	2.94	3.27	3.34	3.13
N2	3,4-$(CH_2)_4$-PAA	2.67	2.67	2.83	5.41	2.40
N3	3,4-$(CH_2)_3$-PAA	2.33	2.33	2.33	4.37	1.95
N4	3-C_6H_5-PAA	3.18	3.44	3.24	3.21	2.93
N5	3-OCF_3-PAA	2.48	2.22	2.21	2.25	1.96
N6	4-Cl-PAA	1.97	2.20	1.86	2.00	1.64
N7	4-F-PAA	1.41	1.56	1.43	1.48	0.96
N8	3-Br-PAA	2.22	2.16	2.03	2.21	2.00
N9	3-Cl-PAA	2.03	1.89	1.86	2.00	1.62
N10	3-C_2H_5-PAA	2.24	2.07	2.20	2.30	2.10
N11	3-CH_3-PAA	1.78	1.58	1.66	1.78	1.58
N12	4-OCH_3-PAA	1.23	1.16	1.42	1.33	1.39
N13	3-$COCH_3$-PAA	0.98	1.45	1.37	0.85	1.05
N14	3-OH-PAA	0.76	0.92	1.03	0.73	0.79
N15	2-F-PAA	1.26	1.29	1.38	1.48	0.86
N16	3-COOH-PAA	1.11	0.92	0.79	0.98	1.12
N17	2-NAOA	2.54	2.58	2.48	2.53	2.28
N18[g]	3-CF_3-PAA	2.36	2.39	2.52	2.41	1.71
N19[g]	3-F-PAA	1.40	1.27	1.44	1.48	0.93
N20[g]	PAA	1.27	1.04	1.20	2.34	1.09

[a] The compound identifier number. [b] PAA: phenoxy acetic acid; NAOA: 2-naphthoxyacetic acid. [c] Values predicted by Volsurf model obtain using tilted conformations. [d] Values obtained by Volsurf model obtained using planar conformations. [e] Values obtained by Interactive Analysis.[27] [f] Reference not provided. [g] Compounds used for external prediction. [h] Experimentally determined log*P* values.[30]

were chosen to span the largest possible volume in the multidimensional space of each class with the smallest number of compounds. Class 1 (the most numerous and diverse class) was represented by the four kernel compounds, while for the other classes, one or two kernel compounds was sufficient. This choice of class kernels was based on reliability of the model to correctly predict classification of the already classified compounds.[7]

Subsequently, the model was used successfully to classify a set of other auxin-related compounds (see refs 7, 8, 24, and this work). For each of the 16 MIFs (for the four probes at the four sets of points), the Carbo (C)[25] and Hogkin (H)[26] similarity indices (SI) were calculated between the kernel and the compound to be classified. A molecule was classified to the class on the basis of the largest similarity index to its kernel.

Estimation of log*P* and log*D* Values. The log*P* and log*D* values were calculated for the auxin-related compounds to describe their permeability and to improve the MIFs based classification method. For this purpose, different procedures were applied and compared.

The conformationally dependent log*P* models were derived using the Volsurf program[9,10] as described in detail below. Interactive Analysis, a program available on the Internet (www.logp.com), was also used for log*P* predictions. Interactive Analysis employs electrotopological state indices using neural networks in order to predict log*P*.[27] The software Pallas 2.1 (Pallas 2.1 CompuDrug Chemistry Limited) was used to calculate log*P* by Rekker's traditional fragmental procedure[28] as encoded in PrologP/CDR option. Since the molecules under study possess at least one acidic function, the module PrologD implemented in the same software was also used to derive apparent lipophilicity expressed as log*D*.[29] To simulate the plant physiological conditions, log*D* calculations were performed at pH 5 and 7 at low ionic strength

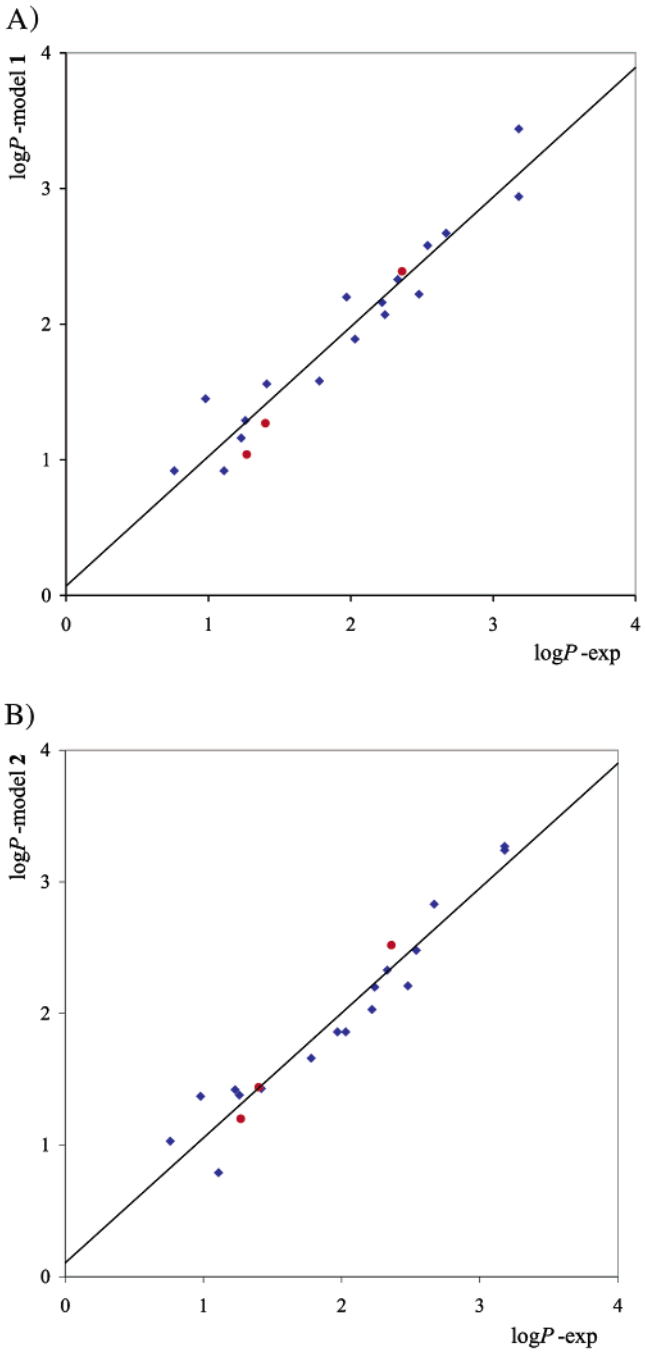

Figure 1. Predicted vs measured log*P* values, compounds from the training set (◆), compounds from the 3-object test set (●). (A) Model derived for the 'T'-set of conformers. (B) Model derived for the 'P'-set of conformers.

(0.001 M) and in the presence of 0.15 M KCl as well as for completely ionized compounds. All those calculated log*P* and log*D* values were used as a correction for the classification based on similarity of MIFs.

Development of Volsurf Models for log*P* Prediction. Using the Volsurf[9,10] program, a QSPR (quantitative structure property relationship) model for log*P* prediction was built for 17 compounds with measured log*P* values[30] (N1−N17, Table 2). For this purpose, about 100 descriptors calculated by the Volsurf program were utilized. Among them, the chemometric PLS (Partial Least Square) analysis revealed volume, surface, ratio of volume and surface, and globularity as the most important.

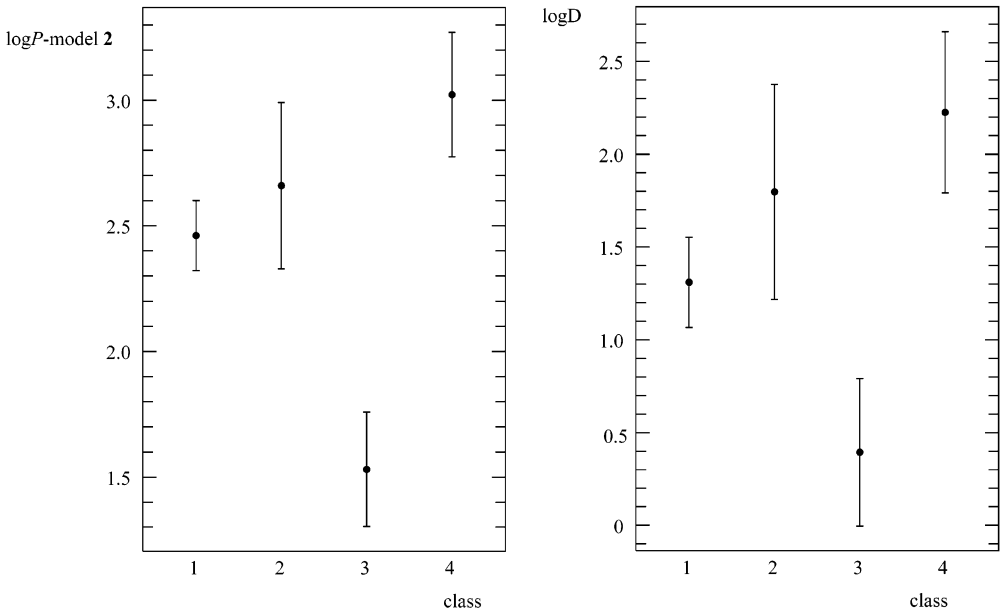

Figure 2. Distribution of the log*P*/log*D* values for the four auxin activity classes: (A) log*P* values, predicted by model 2, (B) log*D* values, calculated at pH = 5 and 0.15 M KCl. The figure was made using the Interactive statistical calculation pages, available at the Internet (www.statistics.com/content/javastat.html).

Since some of the Volsurf descriptors are conformation dependent, two groups of prediction models were derived, one for tilted and one for extended conformers. Each group comprises several different models based on interaction maps for different probes.

To extract the most predictive X variables, X variable selection was performed by a fractional factorial design (FFD) strategy.[31] For each model, one to three cycles of FFD variable selection, depending on the improvement of predictive ability (internal and external), were performed. In each cycle, 25% of the variables were assigned as dummy variables. The ratio of the number of combinations of variables used to derive a model vs the number of variables was three. Uncertain variables were retained. The model derived using autoscaling of X-variables resulted in systematic overestimation of log*P* values with respect to the values calculated by the Interactive Analysis (IA) and the PrologP programs. Consequently, no scaling procedure was applied to X-variables in the final model.

The quality of the models was evaluated by checking their predictive abilities. For this purpose, internal and external validations were performed. For internal, "random groups cross-validation" using 3 random groups and 20 randomizations were performed. The predictive ability of a model was quantified by the SDEP (Standard Deviation of Error of Prediction)[31]

$$\mathrm{SDEP} = \sqrt{\sum \frac{(Y - Y')^2}{N}} \tag{1}$$

and the correlation between the experimental and predicted log*P* values, Q^2

$$Q^2 = 1 - \frac{\sum (Y - Y)^2}{\sum (Y - \langle Y \rangle)^2} \tag{2}$$

where Y is the experimental, Y' is the predicted, and $\langle Y \rangle$ is the average experimental log*P* value.

Final Classification based on log*P* + MIF Similarity. For the final classification, the similarity indices of a compound with the classes (class kernels) were modified proportionally to the deviation of its log*P* value from the log*P* range optimal for auxins (a justification for such corrections is outlined in Results, see also Figure 2). Thus, the similarity indices were corrected as follows, resulting either in an increase or a decrease of the original values (SI_n^C = Carbo similarity indices of a molecule with the n class kernel):

$$\mathrm{SI}_1^{\mathrm{C}} = \mathrm{SI}_1(1 + f(\log P)) \tag{3}$$

$$\mathrm{SI}_2^{\mathrm{C}} = \mathrm{SI}_2(1 + f(\log P)) \tag{4}$$

$$\mathrm{SI}_3^{\mathrm{C}} = \mathrm{SI}_3(1 - f(\log P)) \tag{5}$$

$$\mathrm{SI}_4^{\mathrm{C}} = SI_4(1 - f(\log P)) \tag{6}$$

With the $f(\log P)$ defined (as a bottom flat function) in the following way:

$f(\log P) = 0$, and

$$f(\log P) = \frac{(\log P^H - \log P)(\log P - \log P^L)}{(\log P^H - \log P^L + 3\bar{\sigma})^2} \tag{7}$$

for $\log P^{\mathrm{L}} < \log P < \log P^{\mathrm{H}}$, and for $\log P \leq \log P^{\mathrm{L}}$ and $\log P \geq \log P^{\mathrm{H}}$, respectively.

The higher ($\log P^{\mathrm{H}}$) and the lower ($\log P^{\mathrm{L}}$) border of the optimal log*P* range for auxin activity correspond to the mean log*P* values in the fourth and the third class, respectively. Such borders were utilized since almost all log*P* values for the strong auxins (class 1) are within this range, see Results and Discussion. $\bar{\sigma}$ is the mean of the standard deviations of the log*P* values within the four auxin classes.

The same SI correction procedure was performed for each set of log*P* and log*D* values. Since the log*P* and the log*D*

values depend on the method and/or the conditions used for their evaluation, new borders, logP^H (logD^H) and logP^L (logD^L), were calculated for each set of values.

The new class similarity index was calculated for each MIF, and a compound was classified according to the population of the corrected similarity indices. Using the additive models 3−6, we ensure that classification of a compound with log*P* (log*D*) slightly out of the optimal range will be affected only if its similarity indices with classes 1 and 2 are similar to its similarity index with either class 3 or class 4.

Docking. The recently determined crystal structure of the auxin-binding protein 1 (ABP1),[6] the best candidate for an auxin receptor, and its complex with the 1-NAA were used as initial structures for the docking study. The study was performed for IAA, 5-OH-IAA, 4-Cl-IAA, 2,4-Cl_2-PAA, 3-CF_3-PAA, and 3-O-CF_3-PAA.

Ligands were built and manually docked into the ABP1 active site using 1-NAA as a template. Hydrogens were added using the Biopolymer module in the program insightII.[13] Parametrization was performed within the AMBER force field 96.[32] Parameters for the Zn ion from our earlier work[33] were used. The geometry optimization of the complex was accomplished through several cycles using the three algorithms (steepest decent, Polak-Ribier modified Conjugate gradient[34] and, only for the first two cycles, Truncated Newton modified Conjugate gradient[35]) with different constraints. In the first cycle all heavy atoms were fixed using the frozen atom mode available in MacroModel. In the second cycle the protein backbone atoms and zinc were frozen, while the other atoms were free to move. During the third cycle the constraint on the backbone atoms was subsequently decreased from 100 to 10 kJ/(mol·Å^2) with the zinc remained fixed. Energy minimizations were carried out until the energy gradient became less than 0.1 kJ/(mol·Å).

The structure of the ABP1 in the optimized complex ABP1-1-NAA was used in the docking studies performed for the following auxin-related compounds: IAA, 5-OH-IAA, 4-Cl-IAA, 2,4-Cl_2-PAA, 3-CF_3-PAA, and 3-O-CF_3-PAA. For the docking study each substrate was manually inserted into the ABP binding site, and the search for possible binding modes was performed using the programs MacroModel[36] and DISCOVER[14] as follows. The initial set of binding modes was generated during 1000 steps of a multiple minimum Monte Carlo search (MCMM)[37,38] combined with rotation and translation of the ligand (option MOLS). All atoms within 10 Å of the carboxyl C-atom of the ligand in the starting structure were free to move, and those at a larger distance were frozen. Each conformer of the ABP1-auxin related molecule complex within 30 kJ/mol of the global minimum was further subjected to geometry optimization (the Polak-Ribier modified Conjugate gradient method was used[35]) until a gradient of 0.1 kJ/(mol·Å^2) was achieved). A few selected conformers that we considered as representatives of different binding modes were subjected to MD simulation (DISCOVER[14]) to explore their stability and possible conformational transitions. After 5 ps of initialization at 250 K, a molecule was simulated at temperatures increasing from 250 to 350 K in intervals of 50 ps, with a time step of 1 fs.

The binding pocket in ABP1 is hydrophobic and entirely buried; there are no water molecules in the binding pocket[6] in the crystal structure of the ABP1-1-NAA complex, and thus explicit water molecules were not included in the calculations. The electrostatic interaction screening due to the protein environment was modeled by the distance dependent dielectric constant $\epsilon = 4$.

RESULTS AND DISCUSSION

Classification based on MIF Similarity Only. Previously, we developed a method to classify auxin-related molecules that was based on computation of the similarity of interaction fields of a molecule with the interaction fields of predefined classes represented by minimum representative sets (see Methods). The method was successfully applied to a set of about 70 auxin-related compounds.[7,8,24] About half of these molecules were indole compounds, one-tenth phenoxy acetic acid (PAA) derivatives, two-tenths benzoic acid derivatives, and the rest phenoxy propionic acid and naphthalene derivatives. Although the majority of the molecules (almost 90%) were properly classified, we noticed that among the phenoxy compounds, incorrectly classified molecules were more frequent than among the indole analogues.[7,8]

In the present work, the MIFs-based classification method was applied to a new set of 30 auxin-related compounds, the majority of which are PAA derivatives. The classification was performed considering the deprotonated species of the carboxyl group, which with a pK_a value around 4.8,[39] is predominantly ionized at physiological pH. Further support for this ionization state comes from the recently determined crystal structure of the complex of ABP1 with 1-NAA[6] which revealed binding of an ionized auxin molecule in the active site.

MIFs were calculated at the four sets of predefined points using four probes (see Methods), and, for each set, similarity with the four class-kernels was evaluated. The results of the calculations averaged over different probes and sets of points are summarized in Table 3. For 16 of the 30 compounds, the classification was consistent with the measured biological activity. A high proportion of the wrongly classified compounds (11 out of 14) are phenoxy derivatives.

Comparison of log*P* and log*D* Values. The Volsurf models with the best predictive performances are the models 1 and 2, for 'T' and for 'P' conformers, respectively (Table 4). Predicted and experimental values are presented in Table 2, log*P* values predicted by Interactive Analysis (logP_{IA}) and PrologP (logP_{CDR}) are also included for comparison. External validation of the models derived for the 17-object training set was performed on a three-object test set, N18−N20, belonging to the same set of experimental data. Both models correctly predicted the log*P* values for the three objects in the test set belonging to the same set of experimental data (Table 4, Figure 1) and were used to predict log*P* value for the remaining (about 80) compounds.

Two sets of log*P* values derived by Volsurf models 1 and 2 are highly correlated, more moderate correlations were found with the log*P* values determined by Interactive Analysis and PrologP, indicating that the different log*P* expressions do not encode identical information. The largest deviation between the PrologP and the log*P* values predicted by the other methods was determined for two two-alkyl-PAA derivatives (N2, N3). PrologD incorporates the ionization of molecules; therefore, no good correlation is expected

Table 3. Classification of the Auxin-Related Compounds by Different Classification Methods

compound ID[a]	compound[b]	MIF, only[c]	MIF+ log*P*[d]	MIF+ log*D*[e]	Exp[f]
N1	3-n-Bu-PAA	1,2	2,4		4
N2	3,4-Et_2-PAA	2		4	2,4
N3	3,4-Me_2-PAA	1		4	2,4
N4	3-C_6H_5-PAA	1,2	1		4
N5	3-O-CF_3-PAA	1			4
N6	4-Cl-PAA	1,2			1
N7	4-F-PAA	1,2	1,2,3	1,2	1
N8	3-Br-PAA	1		1	1
N9	3-Cl-PAA	1			1
N10	3-Et-PAA	1			1
N11	3-Me-PAA	1	1	1	1
N12	4-OCH_3-PAA	2	2,4	2	1,3, (4)[g]
N13	3-$COCH_3$-PAA	1	1,3	3	1,3
N14	3-OH-PAA	1	3	3	3
N15	2-F-PAA	2	3	3	3
N16	3-COOH-PAA	1	3	3	3
N17	2-NAOA	2			1
N18	3-CF_3-PAA	1			1
N19	3-F-PAA	1	3	3	3
N20	PAA	1	3		3
N21	2,3-Cl_2-PAA	1			1
N22	2-Cl-PAA	1			1
N23	3,4-Cl_2-PAA	1			1
N24	5-Cl-2-Me-IAA	1	1		1
N25	7-aza-IAA	1	1,3	1,2,3	1
N26	7-Br-IAA	2	2		2
N27	5-Cl-7-Me-IAA	2	2		2
N28	2-Me-7-Cl-IAA	2			2
N29	5-OH-IAA	1	1		2,3
N30	N-Me-IAA	1	1	1	2,3

[a] The compound identifier number. [b] PAA: phenoxy acetic acid; NAA: naphthalene acetic acid; IAA: indole-3-acetic acid. [c] Classification based on similarity analysis for 16 MIFs (for the four probes at the four sets of points). [d] Classification obtained by taking into account the log*P* values + MIFs. [e] Classification obtained by taking into account the log*D* values + MIFs. [f] Classification based on experimental measurements.[30,40−46] [g] Experimental data are not clear enough to make straightforward classification.

Table 4. Volsurf Models for log*P* Prediction

model	probe[a]	LV[b]	*V*[c]	R^2	SDEC	Q^2 [d]	SDEP	SDEP[e]
1	H_2O, $-CH_3$, =O, DRY	3	35	0.96	0.195	0.76	0.356	0.151
2	H_2O, $-CH_3$, =O, DRY	3	30	0.97	0.191	0.78	0.341	0.101

[a] Selection of probes used with the GRID program. [b] Number of latent variables. [c] Number of variables. [d] Q^2 is the three random groups cross-validated coefficient. [e] Calculated for the three compounds of the external validation test set.

with the corresponding log*P* values. Interestingly, the Volsurf derived log*P* values correlate better with the PrologD than with PrologP values (Table 5).

Independently of the method used (and conditions, in the case of the log*D* calculation, see Materials and Methods), the distributions of the log*P* values as well as the log*D* values among the auxin classes differ significantly, see, for example, Figure 2. The log*P* (and the log*D*) values of the inactive (class 3) and the antiauxin compounds (class 4) are lower and higher, respectively, than those of the active auxins (class 1). The distribution of the log*P* values in class 2 is approximately between distributions in classes 1 and 4. This is in accord with the biological activity of compounds belonging to that class.

Table 5. Correlation Coefficients Calculated for the Different Sets of log*P* Values

model	log*P* model 1	log*P* model 2	log*P* Interactive analysis	log*P* PrologP	log*D*[a]
log*P* model 1	1	0.91	0.84	0.65 (0.76)[b]	0.82
log*P* model 2		1	0.86	0.75 (0.81)[b]	0.77
log*P* IA			1	0.78 (0.85)[b]	0.77
log*P* PrologP				1	0.69
log*D*[a]					1

[a] This is for log*D* calculated for pH = 5 and c(KCl) = 0.15 mol/dm^3. However, all six sets of the log*D* values are highly correlated ($R > 0.95$), and their correlation with the log*P* values are similar to the correlations of this one. [b] Correlation coefficient obtained omitting outliers N2 and N3.

Classification based on MIF Similarity + log*P*. The final classification combines the classification procedure based on the similarity of MIFs with the log*P* (log*D*) values. Since it is hard to predict which set of the log*P* and the log*D* values is the most reliable and considering that each set encodes somewhat different information, all sets were considered for the final classifications.

Similarity indices calculated for the MIFs evaluated on three different grids (12 MIFs per compound, the correction was not performed for the MIFs calculated on the molecular surface) were corrected using eqs 3−6 (see Methods). The similarity indices were corrected using the four different log*P* and the six different log*D* sets of values (setting the different borders for each set). Finally, a compound is classified to the class with the highest frequency of choice over the set of SIs. For example, N14 was put into class 1 16 times by the MIFs-only similarity analysis. Considering the log*P* correction, using the values predicted by the Volsurf model 1 as well as using those calculated by Interaction Analysis, the same compound was classified to class 3 nine times and to class 4 three times. Using the log*P* values predicted by model 2, it was classified to class 3 nine times, to class 4 twice, and to class 1 once. Using the PrologP log*P* values, it was put into class 3 12 times. Overall, the MIFs + log*P* procedure classified N14 to class 3, in accord with the experimental results. Similar corrections were obtained with the log*D* values, so by the MIFs + log*D* approach, N14 is also classified to class 3.

The results of the classification based on MIFs only and those of combined classifications are shown in Tables 3 and 6. The log*P* and the log*D* values influence the final classification similarly. For example, compounds N14−N16, N19, and N20 were initially classified to class 1, but the new combined approaches, using either the log*P* or log*D* correction, put them into class 3, consistently with experimental results.[30] In a few cases, however, the classifications based on the log*P* and log*D* corrections differ: using the log*D* correction, compounds N2 and N3 were classified as antiauxins (class 4), while the log*P* correction did not influence their classification to classes 2 and 1, respectively, as determined by similarity of MIFs. From the experimental measurements,[30] it is not clear whether these two compounds belong to class 2 or 4.

An example of the influence of log*P* on the classification of auxins is shown in Figure 3. The mean N14 SIs, calculated for the four sets of MIFs, are greater for classes 1 and 2

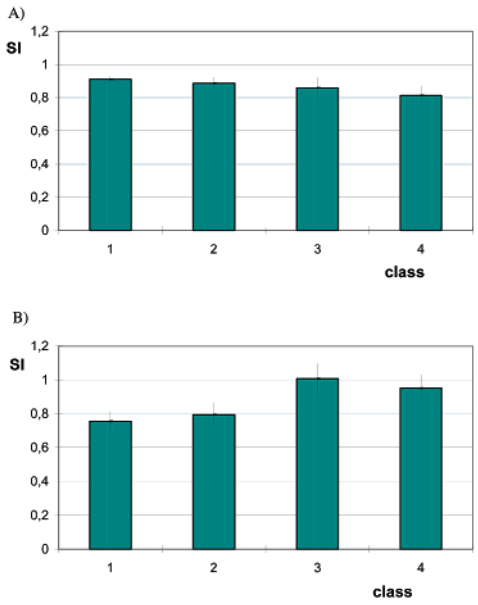

Figure 3. Similarity indices for N14 (3-OH-PAA) averaged over values calculated for the four probes and all 12 sets of MIFs, (A) before and (B) after the log*P* correction; standard deviations are given as bars.

than for classes 3 and 4 (Figure 3A). However, after the log*P* correction (Figure 3B) and in agreement with experimental measurements, the class 3-similarity index is the largest.

Compounds N1−N4 showed antiauxin behavior at the concentration of about 10^{-4} mol/dm^3.[30] So, they might be considered either as antiauxins or as weak auxins with antiauxin behavior. From MIF similarity, they were classified into classes 1 and 2, see Table 3. Consistent with experiment, the log*P* correction shifted N1 toward the antiauxin class, and the log*D* corrected classification put N2 and N3 in class 4.

For eight out of 10 of the phenoxy compounds, which were misclassified by the similarity of MIFs only analysis the log*P* and/or the log*D* corrections improved classification. However, classification of the phenoxy compounds N4 and N5 as well as of the indoles N29 and N30 was not affected by these corrections, because they have mean log*P* and log*D* values below the means determined for auxins, and they remained misclassified. However, using the PrologP log*D* for completely ionized compounds (c(KCl) = 0.015 mol/dm^3), N30 was classified, more consistently with experiment, as inactive compound (class 3).

For N5 inhibitory behavior was determined at a concentration of 10^{-6} mol/dm^3,[3,30] by our method it was classified to class 1. It is interesting to notice that a very similar compound, N18, is a strong auxin. We performed docking for the compounds N5 and N18 and found that they both bind into the ABP1 active site but with different distributions of binding modes (for details, see Results of docking study).

From the biological tests made for compounds N30[40,41] and N29,[41,42] it was not clear whether they should be placed in classes 2 or 3. Their inhibitory effect was not measured, and the auxin activity of those compounds was weak. Our classification procedure(s) put them into class 1. By the MIF-only based approach N25 was correctly classified to class 1;[41] however, the two log*P* corrections shifted it toward inactive compounds [interactive analysis (log*P* = 0.83) and PrologP/CDR (log*P* = 0.86)]. For the compounds N30 and N25, experiments had not been performed on Avena but on pea and tomato.[40,41] In our earlier investigations, we noticed that the auxin activities on Avena and maize are similar and differ from that of pea. This discrepancy might point either to a different auxin receptor in these plants or differences in transport. Obviously, the experimental data are not always accurate enough to enable straightforward evaluation of the classification results. A further example of the later is N12 for which weak auxin activity was measured[30] only at high concentrations of about 10^{-4} mol/dm^3, while measurements of the antiauxin activity had not been performed. Our procedure using the MIF-only classification assigned to class 2, whereas the combined classification shifted it toward antiauxins.

Apart from N25 (see the text above), the log*P* and the log*D* correction also slightly worsened the classification for N7. According to experiments, N7[30] is a strong auxin, from the similarity of its MIFs, it is at the borderline between strong and weak auxins, and the log*P* and the log*D* corrections shifted it toward inactive compounds.

According to the experimental data,[30] N6 belongs to class 1. By our classification, it was put at the borderline of classes 1 and 2. The difference between its class 1 and class 2 similarity indices (SI_1 and SI_2) is for most sets of MIFs extremely low (lower than 1%).

The new, combined classification method was also applied to a set of about 70 auxin-related compounds that were previously classified according to their MIF similarity.[7,8,24] The majority of these compounds have log*P* and log*D* values such that the lipophilicity correction does not influence their classification, i.e., either they have lipophilicity values within the range considered as optimal or the correction confirms the similarity of MIF-only based classification. However, in a few cases, the classification changed. For example O41, O44, and O49 could not be unambiguously classified by similarity of their interactions with the individual class kernels (previously they were screened out as compounds not belonging to auxins only after using combined class 1 and class 2 kernels).[7] Taking into account the lipophilicity of these compounds, TFIBA (O41) was classified as an antiauxin (class 4), and O44 and O49 were classified as inactive compounds (Table 6). This brought the predictions for O41 and O49 into agreement with experiment but not that of O44. However, O44 has very high mean log*P* and log*D* values (4.23 and 2.36, respectively) and, according to the distribution of lipophilicity among the classes, it should belong to the antiauxin class. The correction failed in this case since the similarity index with class 3 is larger than the similarity index with class 4, and the same correction is applied to both.

The lipophilicity correction changed the initial, similarity of MIFs-only based classification of the compounds O59 and O60 into the class 1 and shifted them toward antiauxins. These two compounds have log*P* and log*D* values above the higher limit of the range considered as optimal; however, according to experimental measurements and the classification based on the similarity of the MIFs-only they belong to auxins.

Beside log*P*/log*D* correction of SIs, we have also tried to improve classification procedure by using other descriptors related to permeability of the compounds, such as volume and hydrogen bonding capacity. We found that the hydrogen

Table 6. Results of Different Classification Methods for the Set of the Auxin-like Compounds, that We Analyzed in Our Earlier Publications[7,8,24]

compound ID[a]	compound[b]	MIF only[c]	MIF+ log*P*[d]	MIF+ log*D*[e]	Exp[f]
O1	1-NAA	1			1
O2	2-Me-4-Cl-PAA	1			1
O3	2,5-Cl_2-PAA	1			1
O4	2,4-Cl_2-PAA	1			1
O5	2,5-Cl_2-PA2PA	1			1
O6	2,4-Cl_2-PA2PA	1			1
O7	IAA	1			1
O8	4-Cl-IAA	1			1
O9	5-Cl-IAA	1			1
O10	6-Cl-IAA	1			1
O11	5,6-Cl_2-IAA	1,4	1,4		1
O12	6-F-IAA	1			1
O13	4-F-IAA	1			1
O14	4-Me-IAA	1			1
O15	2,3,4-Cl_3-PAA	1			1
O16	5-F-IAA	1	1		1
O17	4-Et-IAA	1	1		1
O18	5-Me-IAA	1			1
O19	7-F-IAA	2 (1)[g]	2		1
O20	4,6-Cl_2-IAA	1	1,4		1
O21	2,4,5-Cl_3-PAA	1	1		1
O22	6,7-Cl_2-IAA	1,2	1,2		1
O23	4-IBA	1,2			1
O24	2,5-Me_2-PA2PA	1			1
O25	7-Cl-IAA	2			2
O26	4,7-Cl_2-IAA	2			2
O27	3-IPA	2			2
O28	5,7-Cl_2-IAA	2			2
O29	2-NAA	1,2			2
O30	2-Cl-BA	3			3
O31	2-F-BA	3			3
O32	2-I BA	3			3
O33	2,6-Cl_2-BA	3			3
O34	BA				3
O35	4-Cl-BA	1,3			3
O36	β-naphthoic acid	3			3
O37	2-Me-BA	3			3
O38	4-Cl-PIBA				4
O39	3-IIBA	4			4
O40	5,7-Cl_2-IIBA				4
O41	TFIBA	1,4	4		4
O42	3-IBA	1,4			4
O43	2,4-Cl_2-PIBA	4	4	4	4
O44	3,4,5-I_3-BA	1,4	3	3	4
O45	3,4-Cl_2-BA	4			4
O46	2-Me-IAA	1			1
O47	2,5-Me_2-PAA	1			2
O48	2,3,5-Cl_3-BA	1	1,3		3
O49	2,4,6-Cl_3-PAA	3,2	3		3
O50	2,3,5-I_3-BA	3	3	3	4
O51	3,5-Cl_2-PA2PA	4			4
O52	5-Et-IAA	1	1		1
O53	6-Et-IAA	1	1		1
O54	7-Et-IAA	1			1
O55	6-Me-IAA	1			1
O56	7-Me-IAA	1			1
O57	2-Et-IAA	1			1
O58	5-Br-IAA	1	1		1
O59	5-n-Bu-IAA	1	4	4	1
O60	5-n-Pr-IAA	1	4	1	1
O61	4,5-Cl_2-IAA	1	1		1
O62	5-Methoxy-IAA	1			1

[a] The compound identifier number. [b] PAA: phenoxy acetic acid; PA2PA: phenoxy-2-propionic acid; IAA: indole-3-acetic acid; BA: benzoic acid; 3-IPA: 3-indole-3-propionic acid; PIBA: phenoxy isobutyric acid; 4-IBA: 4-indole-3-butyric acid; 3-IBA: 3-indole-3-butyric acid; IIBA: indole-3-isobutyric acid; TFIBA: 4,4,4-trifluoro-3-indole-3-butyric acid. [c] Classification based on similarity analysis for 16 MIFs (for the four probes at the four sets of points). [d] Classification obtained by taking into account log*P* correction of SIs. [e] Classification obtained by taking into account log*D* correction of SIs. [f] Classification made from experimental measurements (see refs 7 and 8 and references therein). [g] Difference between the class 1 and class 2 SIs is extremely low (from 1−2%).

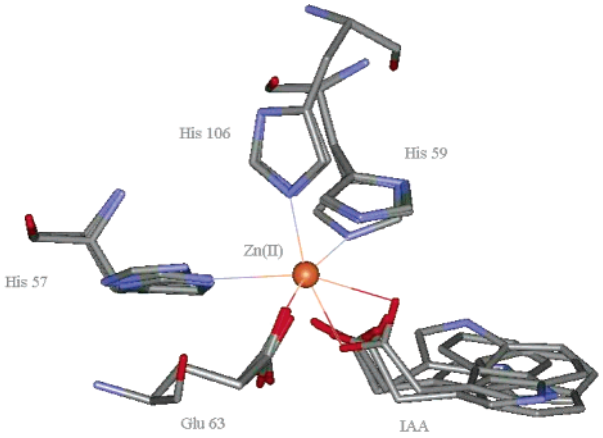

Figure 4. Superposition of the IAA productive binding modes (carboxyl group bidentately coordinate Zn^{2+}) with the Zn ion (represented by a sphere) and the amino acid residues that take part in its coordination displayed.

bonding capacity did not influence classification, whereas volume improved classification in a few cases. However, the correction achieved was not as good as when log*P*/log*D* values were used so we did not pursue application of these parameters further.

Results of Docking Study. In the crystal structure of the ABP1-1-NAA complex 1-NAA is bound in the hydrophobic pocket and bidentately coordinates zinc ion.[6] Besides coordination by the carboxyl group of 1-NAA, the zinc ion is coordinated by three histidines (His 57, His 59, His 106) and Glu 63. According to the results of the docking study performed for the four IAA and the three PAA derivatives (see Materials and Methods), auxins can bind into the ABP1 active site in a few different but energetically similar orientations and conformations (with the carboxyl side chain inclination to the conjugated ring plane ranging from about 10° to 100°). In some of these binding modes, auxin molecules coordinate the Zn ion monodentately and in some, bidentately. Based on the crystal structure of the ABP1-1-NAA complex, we considered the latter modes as productive. For IAA, five such modes were found within 20 kJ/mol of the global minimum (Figure 4). Similar binding modes, spanning approximately the same energy range, were found for 4-Cl-IAA and 5-OH-IAA. However, for 5-OH-IAA, the energy difference between the lowest and the second lowest productive binding modes is larger (about 11 kJ/mol, compared to 8 kJ/mol for IAA and 4-Cl-IAA). The reason for this is the hydrogen bond between the hydroxyl group and the carbonyl oxygen of Glu 46 that decreases the energy of the global minimum.

The similarity of the binding modes of IAA, 4-Cl-IAA, and 5-OH-IAA is in accord with the classification based on the similarity of the MIFs-only which put these compounds to class 1. However, their biological responses differ significantly, as do their log*P* and log*D* values. A decrease in their biological activity (4-Cl-IAA > IAA > 5-OH-IAA) correlates with a decrease of their lipophilicity, i.e., the mean log*P* value of IAA and 5-OH-IAA is about 0.6 and 1.1 lower, respectively, than that of 4-Cl-IAA ($<\log P$(4-Cl-IAA)$> = 2.47$, $<\log P$(IAA)$> = 1.90$, $<\log P$ 5-OH-IAA$> = 1.42$).

The experimental measurements of Rescher et al.[47] fit well with our proposal that the similarity of MIFs provides

information about the possibility of a molecule to bind into the ABP1 active site but that biological activity depends on both binding and lipophilicity. They found that, at the concentration of 10^{-6} mol/dm^3 (which is the optimal concentration for auxin binding), IAA binds slightly better than 4-Cl-IAA, but 4-Cl-IAA produces a stronger biological response, namely elongation of coleoptil in maize.

The docking study revealed an explanation for the inhibitory activity of 3-O-CF_3-PAA which is in contrast to the auxin activity of 3-CF_3-PAA. Both 3-O-CF_3-PAA and 3-CF_3-PAA can bind into the ABP1 binding site in a few different modes. The most favored is, for both compounds, the mode in which the substrate carboxyl group coordinates Zn monodentately. However, the productive binding mode of 3-CF_3-PAA is only 7 kJ/mol above the global minimum, whereas that of 3-O-CF_3-PAA is 25 kJ/mol above the global minimum. Besides, the bulky OCF_3 group at position 3 of the phenyl ring decreases the rotational entropy of the neighboring amino acid residue side chains and hinders the conformational change of Trp151 that, according to Woo et al.,[6] might be connected with the signal transfer.

CONCLUSION

In this work, we have combined the QSAR method based on the similarity of molecular interaction fields that we developed earlier with the theoretically determined lipophilicity given by log*P* and log*D* values. Log*P* and log*D* values of a compound are correlated not only with its membrane permeability but also with its possibility to fit a receptor binding site.

Analysis of the distribution of log*P* and log*D* values among auxin activity classes (*t*-test and variance analysis) has revealed significant difference in the values determined for auxin, antiauxin, and inactive compounds, i.e., compounds in classes 1, 4, and 3. In general, the inactive auxin-like compounds have significantly lower log*P* and log*D* values than auxins, whereas these values are significantly higher for the compounds with antiauxin behavior. These findings are similar to that of Hansch et al.[30] who, for a set of phenoxy acetic derivatives, found an equation that correlates auxin activity with the Hammett substitution constant and the log*P* value. For the diverse set of auxin-related compounds considered by us, such an equation is not possible. Besides, differently from Hansch et al., we distinguish inactive compounds from those with inhibitory characteristics. However, by using the log*P* and log*D* values to modify the similarity indices calculated for the molecular interaction fields, the classification improved. The greatest improvement was observed for phenoxy compounds, which in general have lower log*P* and log*D* values than IAA derivatives (confirmed by *t*-test). Since the docking study indicated that PAA compounds can bind to the ABP1 in a similar way to IAA derivatives, it can be concluded that their activity is in many cases affected by low bioavailability.

The docking study that we performed for the auxin-related molecules with different log*P* and log*D* values proved that similarity of MIFs is a good method to predict binding of a molecule into the receptor active site. However, inclusion of lipophilicity enables better modeling of biological activity. Apparently, extreme log*P* and log*D* values can be correlated with loss of auxin activity. Due to an inappropriate lipophilicity, a compound has either low bioavailability and does not reach the auxin receptor (class 3) or binds in an inappropriate manner and does not produce the expected biological response (class 4). According to the results of our docking analysis, antiauxins are able to bind into the ABP1 active site, but it seems that they are inefficient in regulating signal transfer i.e., the steps in the auxin response pathway that follow its binding to the auxin receptor are probably hindered.

The method presented in this work should be most useful for cases in which the biological response of compounds is the result of the combination of their passive transport through cell membranes with their affinity toward their receptor. Recent results, however, indicate that auxin-like molecules are transported among the cells predominantly by uptake and efflux carrier proteins.[48,49] In this case lipophilicity might be considered as ability of a ligand to bind to these proteins either specifically or unspecifically, while the similarity of MIFs deal predominately with the specific binding to the ABP1.

In this work, we have started the analysis of auxin binding into the active site of the ABP1. We are continuing this study and are planning to carefully examine possible binding modes for a set of auxin related molecules with different biological activity. Besides we will investigate conformational changes of the receptor upon auxin binding that might be relevant for the signal transfer.

ACKNOWLEDGMENT

We are grateful to Dr. V. Magnus for valuable discussions. S. Tomić gratefully acknowledges continuing financial support from the Alexander von Humboldt Foundation. The authors are grateful for support of this work by the Croatian Ministry of Science, the Klaus Tschira Foundation, and the German Ministry of Science (BMBF) through bilateral projects KRO HRV 98/006 and HRV 01/010. It is also part of the Austrian-Croatian cooperation (Österreich-Kroatien WTZ, Projekt Nr 15/2002).

REFERENCES AND NOTES

(1) *Plant Hormones: Physiology, Biochemistry and Molecular Biology*; Davies, P. J., Ed.; Kluwer Academic Publishers; New York, 1995.
(2) Geldner, N.; York-Dieter, F.; Stierhof; Gerd Jürgens; Palme, K. Auxin transport inhibitors block PIN1 cycling and vesicle trafficking. *Nature* **2001**, *413*, 425−428.
(3) Abel, S.; Nguyen M. D.; Theologist A. ThePS-IAA4/5-like Family of Early Auxin-inducible mRNAs in *Arabidopsis thaliana*. *J. Mol. Biol.* **1995**, *251*, 533−549.
(4) Chen, J. G. J. Dual Auxin Saignaling Pathways Control Cell Elongation and Division. *Plant Growth Reg.* **2001**. *20*, 255−264.
(5) Gray, W. M.; Kepinski, S.; Rouse, D.; Leyser, O.; Estelle, M. Auxin regulates SCF^{TIR1}-dependent degradation of AUX/IAA proteins. *Nature* **2001**, *414*, 271−276.
(6) Woo, E.-J.; Marshall, J.; Bauly, J.; Chen, J.-G.; Venis, M.; Napier, R. M.; Pickersgill, W. Crystal structure of auxin-binding protein 1 in complex with auxin. *EMBO J.* **2002**, *21*, 2877−2885.
(7) Tomić, S.; Gabdoulline, R.R.; Kojić-Prodić, B.; Wade, R. C. Classification of auxin plant hormones by interaction property similarity indices. *J. Comput. Aided Mol. Design.* **1998**, *12*, 63−79.
(8) Tomić, S.; Gabdoulline, R. R.; Kojić-Prodić, B.; Wade, R. C. Classification of auxin related compounds based on similarity of their interaction fields: Extension to a new set of compounds. *Internet J. Chem.* **1998**, *26*, 1. URL <http://www.ijc.com/articles/1998v1/26/> ISSN 1099.
(9) Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *J. Mol. Struct. (THEOCHEM)* **2000**, *503*, 17−30.

(10) Cruciani, G.; Pastor, M.; Guba, W. VolSurf: A New Tool for Pharmacokinetic Optimization of Lead Compounds. *Eur. J. Pharm. Sci.* **2000**, *11* (Suppl. 2), S29−S39.
(11) Goodford, P. J. Computational procedure for determining energetically favourable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(12) Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. Property-Based Design: Optimization of Drug Absorption and Pharmacokinetics. *J. Med. Chem.* **2001**, *44*, 1313−1333.
(13) INSIGHT, v. 2000.1, Accelrys Inc., San Diego, CA, 2002.
(14) DISCOVER, v. 2.98, Accelrys Inc., San Diego, CA, 2002.
(15) Ramek, M.; Tomić, S.; Kojić-Prodić, B. Systematic Ab Initio SCF Conformational Analysis of Indol-3-ylacetic Acid Phytohormone (Auxin): Comparison with Experiment and Molecular Mechanics Calculations. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1995**, *22*, 75−81.
(16) Ramek, M.; Tomić, S.; Kojić-Prodić, B. Comparative Ab Initio SCF Conformational Study of 4-Chloro-indole-3-acetic Acid and Indole-3-acetic Acid Phytohormones (Auxins). *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1996**, *23*, 3−9.
(17) Ramek, M.; Tomić, S. Ab initio RHF investigation of mono- and dichalogenated indole-3-acetic acid (IAA) phytohormones. *J. Mol. Struct.* (*THEOCHEM*) **1998**, 454, 167−173.
(18) Ramek, M.; Tomić, S. RHF Conformational Analysis of the Auxin Phytohormones n-Ethyl-Indole-3-Acetic Acid (n = 4, 5, 6). *Int. J. Quantum Chem.: Quantum Biol. Symp.* **1998**, *70*, 1169−1175.
(19) Tomić, S.; Ramek, M.; Kojić-Prodić, B. Combined ab initio SCF and molecular mechanics study of propionic and isobutyric acids and their biologically active derivatives related to the phytohormone auxin (indole-3-acetic acid). *Croat. Chem. Acta* **1998**, *71*, 511−525.
(20) Ramek, M.; Tomić, S. Quantum chemical conformational analysis of the auxin phytohormone 4-methyl-3-indole acetic acid. *Int. J. Quantum Chem.* **1999**, *75*, 1003−1008.
(21) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T. A Structure and Energetics of Ligand Binding to Proteins: E. Coli Dihydrofolate Reductase-trimethhoprim, a Drug-Receptor System. *Proteins Struct. Funct. Genet.* **1988**, *4*, 31−47.
(22) Kearsley, S. K.; Smith, G. M. An Alternative Method for the Alignment of Molecular Structures: Maximizing Electrostatic and Steric Overlap. *Tetrahedron Comput. Methodol.* **1990**, *3*, 615−633.
(23) Maple, J. R.; Thacher, T. S.; Dinur, U.; Hagler, A. T. Biosym force field research results in new techniques for the extraction of inter- and intramolecular forces. *Chem. Design Autom. News.* **1990**, *5*(9), 5−10.
(24) Antolić, S.; Dolušić, E.; Kožić, E.; Kojić-Prodić, B.; Magnus, V.; Ramek, M.; Tomić, S. Auxin activity and molecular structure of 2-alkylindole-3-acetic acids. *Plant Growth Regul.* **2003**, *39*. 235−252.
(25) Carbo, R.; Calabuig, B. Molecular Quantum Similarity Measures and N-Dimensional Representation of Quantum Objects. *Int. J. Quantum Chem.* **1992**, *42*, 1681−1693.
(26) Hodgkin, E. E.; Richards, W. G. Molecular Similarity Based on Eletrostatic Potential and Electric Field. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 105−110.
(27) www.logp.com, M. Parham, Interactive Analysis.
(28) Rekker, R. F.; Mannhold, R.; *Calculation of Drug Lipophilicity*; VCH: Weinheim, 1992.
(29) Csizmadia, F.; Tsantili-Kakoulidou, A.; Panderi, I.; Darvas, F. Prediction of Distribution Coefficients from Structure. 1. *Estimation Methodol. J. Pharm. Sci.* **1997**, *86*, 865−871.
(30) Hansch, C.; Muir, M. R.; Fujita, T.; Maloney, P. P.; Geiger, F.; Streich, M. The Correlation of Biological Activity of Plant Growth-Regulators and Chloromycetin Derivatives with Hammett Constants and Parititon Coefficients. *J. Am. Chem. Soc.* **1963**, *85*, 2817−2824.
(31) Cruciani, G.; Clementi, S.; Baroni, M. Variable Selection in PLS Analysis. In *3D QSAR. Drug Design Theory Methods and Applications*; Kubinyi, H., Ed.; ESCOM Publishers: Leiden, 1993.
(32) Cornell, W. D.; Cieplak, P.; Payly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(33) Tomić, S.; Nilsson, L.; Wade, R. C. Nuclear receptor-DNA binding Specificity: A Combine and Free-Wilson QSAR Ananlysis. *J. Med. Chem.* **2000**, *43*, 1780−1792.
(34) Polak, E.; Ribiere, G. *Revue Francaise Inf. Rech. Oper.* **1969**, *16-r1*, 35.
(35) Ponder, J. W.; Richards, F. M. An Efficinent Newton-likeMethoid for Molecular Mechanics Energy Minimization of Largee Molecules. *J. Comput. Chem.* **1987**, *8*, 1016−1024.
(36) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Molecular Mechanics with MacroModel. *J. Comput. Chem.* **1990**, *11*, 440−467.
(37) Chang, G.; Guida W. C.; Still, W. C. An Internal Coordinate Monte Carlo Method for Searching Conformational Space. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.
(38) Saunders: M.; Houk, K. N.; Yun-Dong Wu, Still, W. C.; Lipton, M.; Chang, G.; Guida W. C. Conformations of Cycloheptadecan. A Comparison of Methods for Conformational Searching. *J. Am. Chem. Soc.* **1990**, *112*, 1419−1427.
(39) Pine, S. H.; Hendrickson, J. B.; Cram, D. J.; Hammond, G. S. *Organic Chemistry*; McGraw-Hill Kogakusho, LTD Publishers: Tokyo, 1980.
(40) Hoffmann, A. L.; Fox, S. W.; Bullock, M. W. Auxin-like activity of systematically substituted indoleacetic acid. *J. Biol. Chem.* **1952**, *196*, 437−441.
(41) Porter, W. L.; Thimann, K. V. Molecular Requirements for Auxin Action-I. Halogenated Indoles and Indoleacetic Acid. *Phytochemistry* **1965**, *4*, 229−243.
(42) Bottger, M.; Engvild, K. C.; Soll, H. Growth of Avena Coleoptiles and pH Drop of Protoplast Suspensions Induced by Chlorinated Indoleacetic Acids. *Planta* **1978**, *140*, 89−92.
(43) Katekar, G. F.; Geissler, A. E. Structure−Activity Differences Between Indoleacetic Acid Auxins on Pea and Wheat. *Phytochemistry* **1983**, *22*, 27−31.
(44) Katekar, G. F. Auxins: on the Nature of the Receptor Site and Molecular Requirements for Auxin Activity. *Phytochemistry* **1979**, *18*, 223−233.
(45) Pybus, M. F.; Wain, R. L.; Wightman. F. New Plant Growth-Substance with Selective Herbicidal Activity. *Nature* **1958**, *182*, 1094−1095
(46) Wain, R. L.; Wightman. F. Studies on Plant Growth-Regulating Substances. *Ann. Appl. Biol.* **1953**, *43*, 244−249.
(47) Rescher, U.; Walther, A.; Schiebl, C.; Klambt, D. In Vitro Binding Affinities of 4-Chloro-, 2-Methyl-, 4-Methyl-, and 4-Ethylindoleacetic Acid of Auxin-binding Protein 1 (ABP1) Correlate with Their Growth-Stimulting Activities. *J. Plant Growth. Regul.* **1996**, *15*, 1−3.
(48) Palme K.; Gälweiler L. PIN-pointing the molecular basis of auxin transport. *Curr. Opinion Plant Biol.* **1999**, *2*, 375−381.
(49) Muday, G. K.; Murphy, A. S. An emerging model of auxin transport regulation. *Plant Cell* **2002**, *14*, 293−299.

CI034063N